ELDER RESEARCH
DATA SCIENCE & PREDICTIVE ANALYTICS

Evaluation of the EEOC's Data Analytics Activities
Final Report
OIG Report Number 2017-02-EOIG

September 5, 2018

National Capital Region
2107 Wilson Boulevard, Suite 850
Arlington, VA 22201
855.973.7673

Headquarters
300 W. Main Street, Suite 301
Charlottesville, VA 22903
434.973.7673 | fax 434.973.7875
www.elderresearch.com

Friendship Landing Office
839 Elkridge Landing, Suite 215
Linthicum, MD 21090
855.973.7673

TABLE OF CONTENTS

1.0 EXECUTIVE SUMMARY

This analytics evaluation of the Equal Opportunity Employment Commission (EEOC) was conducted by Elder Research on behalf of the EEOC Office of Inspector General (EEOC OIG). The three primary objectives of the assessment were to:

1) Assess the strengths and weaknesses of the EEOC's data analytics culture, strategy, tactics, and capabilities (people, processes, technologies, financial resources, etc.).
2) Assess EEOC's strategies for ensuring the validity and accuracy of critical databases.
3) Identify improvements, opportunities, and best practices regarding EEOC's data analysis and predictive analytics activities.

During this engagement, the assessment team relied on interviews/walkthroughs that were supplemented, as needed, by EEOC strategic plans, reports, and reviews. The evaluation focused on data flows and usage of data within the organization to guide decision-making processes. The evaluation team conducted 26 meetings with a diverse group of stakeholders at EEOC headquarters and two district offices (Charlotte, Chicago) to inform its evaluation[1].

The following table outlines the five assessment areas and a brief summary of findings for each area:

[1] The analyses, conclusions, and subsequent recommendations within this report are based on entirely upon data and information gathered during the fieldwork phase of this review. Fieldwork started with first stakeholder meeting on 2 November 2017 and concluded with final district office interview on 27 February 2018.

Table 1-1: Summary of Assessment Areas and Corresponding Findings

Area	Area Overview	Finding Summary
Culture	The extent the EEOC's culture views data as a core asset and the extent analytics benefits from executive leadership, awareness and vision related to analytics, and environments that foster both collaboration and objective evaluation.	Stakeholders within the EEOC are largely unaware of the differences between reporting and predictive analytics and therefore are unaware of the value that can be unlocked by treating data as a strategic asset.
People	The extent the analytics team(s) understand organizational needs, creatively approach problems, and effectively utilize available tools.	The EEOC lacks an enterprise-scope analytics team devoted to addressing a variety of organizational challenges.
Process	The extent the analytics team(s) have clearly-defined processes to gain business understanding and implement repeatable processes to derive data-oriented insights to address organizational needs.	The EEOC lacks an enterprise-scope analytics team to perform data analytics, so therefore lacks a clearly defined process for such a team.
Analytics Capability	The extent analytics team(s) demonstrate appropriate technical sophistication of analytical products, incorporate feedback into model evaluation and management processes, and evaluate new techniques and technologies.	While existing areas of reporting and analysis use appropriate levels of sophistication to adequately address case-specific goals, the EEOC lacks a general-purpose analytics team to evaluate general analytics capabilities.
Infrastructure	The extent the IT infrastructure fosters data collection and analysis from disparate data sources and enables delivery of effective reporting and analytic products.	The EEOC lacks key, foundational components of infrastructure to support both reporting and data analytics initiatives.

A review of the above table should not be discouraging. Rather, the EEOC should be encouraged by the opportunity that lies ahead: effective implementation of the recommendations contained within this report hold the potential to unlock substantial value and significantly improve the EEOC's ability to accomplish its core mission. This report aims to provide a practical roadmap for the agency to progressively move towards a well-functioning analytics program that empowers individuals to more efficiently and effectively accomplish tasks and make decisions.

This report covers the cultural assessment area first because it provides the foundation necessary to guide the formation of an effective analytics program that provides value to the organization. Although the EEOC may face challenges in securing the resources needed to invest in analytic capability and infrastructure, the cultural recommendations are low-cost and are designed to provide the foundation needed to fully realize long-term value. In addition, the appendix contains a prioritized list of low cost, high impact analytics projects that are aimed to quickly demonstrate the value that can be unlocked from an effective analytics program. These suggested projects provide the EEOC with an actionable starting point to facilitate some "quick wins" without large investments of resources. The EEOC should evaluate such "quick win" projects for returns on investment, providing a baseline from which the recommended governance bodies can build upon in prioritizing projects and allocating resources to best address the EEOC's ongoing needs.

This report contains detailed descriptions of what an effective implementation in each area looks like. These, along with the associated recommendations to achieve greater organizational maturity in each area, provide long-term goals and vision to guide both the formation and evolution of an effective analytics program.

Each recommendation is mapped to one or more responsible parties who will be detailed in the body of this report. The list of responsible parties, and associated abbreviations, are:

- AC: Analytics Champion
- APMO: Analytics Program Management Office
- CDO: Chief Data Officer
- CIO: Chief Information Officer
- DGC: Data Governance Committee
- EDAB: Executive Data Analytics Board
- OCH: Office of the Chair

The appendices to this report contain a summary of all findings (Section 9.1), some example analytics projects the EEOC may consider when evolving its analytics capabilities (Section 9.2), and EEOC response and Elder Research comments to the draft report (Section 9.3). The assessment team thanks all participants at both the EEOC and the EEOC OIG for their time and energy in support of this assessment.

The following table provides a summary of all the recommendations within the report, categorized by each of the five assessment areas:

Table 1-2: Summary of Recommendations

Assessment Area	Section	Section Description	Responsible Party	Brief Overview
Culture	4.1	Shared Vision for Analytics	EEOC OCH, EEOC EDAB	Establish data analytics governance infrastructure.
Culture	4.2	Executive Leadership	EEOC OCH, EEOC EDAB, EEOC AC	Establish tone advocating for analytics in strategic planning and reviewing recommendations of data analytics governance bodies.
Culture	4.3	Culture of Evaluation and Improvement	EEOC EDAB	Invest in the generation of new metrics that quantify opportunity costs and corresponding benefits of data collection and data assurance.
Culture	4.4	Collaborate Environment	EEOC OCH	Engender trust in enterprise-wide steering committees and governance boards.
Culture	4.5	Continued Education and Learning	EEOC OCH, EEOC AC	Designate an analytics champion to foster and evaluate cultural awareness of analytics.
People	5.1, 5.2, 5.3, 6.2, 6.4	Understanding Business Needs, Technological Breadth	EEOC EDAB, EEOC CDO	Establish a centralized, enterprise-wide analytics team or Analytics Center of Excellence.
Analytic Capability	6.3	Modeling Process, Evaluation, and Management	EEOC APMO	Adopt proven modeling approaches and model management techniques.
Process	7.2	Process	EEOC EDAB, EEOC APMO, EEOC CDO	Support analytics projects through governance of the Analytics Center of Excellence, promoting awareness of iterative analytical project processes, and promoting usage of Agile-friendly project management tools.
Infrastructure	8.1	IT Infrastructure and Data Storage	EEOC CIO, EEOC EDAB	Consider new approaches, such as web-enabled and cloud-based solutions, to support expanding IT infrastructure needs of both the analytics team as well as users of analytical products.
Infrastructure	8.2	Data Availability and Transformability	EEOC EDAB, EEOC CIO	Establish a data warehouse to address data retention, versioning, and reporting needs.
Infrastructure	8.3	Visualization and Delivery	EEOC CIO, EEOC EDAB	Invest in modern reporting and visualization tools that allow for automated, customizable, visualization-enhanced reporting that effectively leverage a data warehouse.

2.0 INTRODUCTION

Data analytics is a process of inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information to support the decision-making process.[2] Every mission-oriented organization stands to benefit from an effective data analytics program that leverages data-driven insights to assess and advise how that organization can efficiently and effectively accomplish its mission. At the Equal Employment Opportunity Commission (EEOC), the mission is clear: to stop and remedy unlawful employment discrimination in both the public and private sectors.[3] An effective analytics program must always strive to serve this core mission, empowering both decision-makers as well as those on the front lines who work daily to accomplish this mission.

2.1 ASSESSMENT GOALS

An EEOC data analytics program should address current organizational challenges as well as detect and advise on the emerging challenges the EEOC may face in the near future. To this end, the analytics program should address each operational area. The analytics program itself should be regularly evaluated to measure the extent the data analysis and insights remain effective. To accomplish this, the analytics program must be:

1) **Measurement-oriented:** Analytics must leverage data to generate quantitative metrics that objectively evaluate progress towards clearly-defined, measurable goals.
2) **Evidence-based:** Metrics must capture results of organizational efforts, fostering an evaluation of both those results and the costs of achieving them.
3) **Integrated into Decision-Making Process:** Evaluate evidence regularly to determine whether processes are optimal to achieve desired results. This requires a culture that embraces honest evaluation of decisions, associated outcomes, and a willingness to pivot decisions in the directions where data indicates the most promise in achieving desired goals.

This analytics assessment reflects that every organization is unique. Specifically, the EEOC seeks ways to expand its use of data to more efficiently and effectively accomplish its mission, including:

- Estimating the level of employment discrimination at the national level, including assessments of how discrimination changes over time.

[2] Deal, Jeff, et al. *Mining Your Own Business: A Primer for Executives on Understanding and Employing Data Mining and Predictive Analytics*. Data Science Publishing, 2016. (Chapter 2)

[3] https://www.eeoc.gov/eeoc/internal/eeo_policy_statement.cfm

- Assessing the information received from employers to determine the merits of expanding information received from employers.
- Developing statistics related to the number of pending charges and complaints at a specified point in time, broken out by priority.
- Developing performance measures based on outcomes of charges and cases.
- Providing easy access to information related to outcomes, easily broken down or visualized by characteristics such as priority level, industry, or other key characteristics of charging parties.
- Reviewing the latest performance information on both process and outcome measures, including but not limited to Strategic Enforcement Plan progress reports.
- Identifying emerging trends from charge data as well as other sources.

This report first outlines the methodology and approach employed for this assessment, defining five distinct assessment areas. The subsequent five sections provide details, including findings and recommendations, for each specific assessment area. Recommendations will range from quickly actionable items requiring minimal resources to longer-term, strategic initiatives to foster a sustainable and effective enterprise analytics program.

2.2 GUIDE TO INTERPRETING RESULTS

The findings and recommendations in this report recognize that building an effective analytics strategy for the long-term does not happen overnight. In fact, long-term cultural acceptance of emerging analytics programs are often greatest when the program's earliest stages focus on maximizing the effectiveness of less sophisticated analytical techniques. Put simply, this means priority should be placed on automating workflows, improving data quality, and on the creation of interactive reports and visualizations of data. Once solutions to these basic needs prove to be effective, the ability and willingness of end-users to digest further insights should increase. This serves to increase the rate in which outputs from more sophisticated methods translate into organizational benefits.

In keeping with the idea of prioritizing recommendations in the areas that hold the greatest initial prospective benefits, this report covers the cultural assessment area first. This is due to the fact that EEOC is early in the process of identifying the multitude of specific areas where an enterprise analytics program can provide value. In order to realize such value, stakeholders at all levels of the organization need to adjust work patterns and decision-making processes in order to integrate data-based insights. While interactive reports and visualizations can help, all recommendations in the cultural assessment area are assessed as "high priority" and should be in-progress before implementation of advanced levels of analytics are attempted.

This report provides some recommendations where analytical methods can substantially improve workflow and data quality. Recommendations covering these areas provide good candidates for simple, limited scope initial analytical efforts that can be implemented within a pilot program concurrently with implementation of cultural recommendations. These initial pilots should pay close attention to feedback of end-users to assess their effectiveness, delaying widespread implementation until feedback reaches acceptable levels. If initial analytical efforts are perceived to be less effective than existing methods, trust in the analytics program risks erosion. Conversely, demonstrable improvements in end-user workflow and data quality will serve to increase the benefits of future analytical efforts due to a higher acceptance rate of data-driven insights.

Recommendations range from smaller, immediately actionable items to long-term evolutionary guidelines to build analytic capability. The long-term goal is to foster analytics that effectively predicts and advises on emerging organizational needs over time. Because of the variation in costs and benefits of the recommendations, the EEOC should apply a data-driven approach in creating its strategy: EEOC should regularly assess costs versus benefits of specific efforts to increase analytical capabilities over subsequent evaluation periods. This report contains an appendix highlighting areas where adopting new technologies, such as cloud computing and open source software, hold promise to maximize benefits while controlling costs. However, these highlights are not intended to be prescriptive: the EEOC should evaluate many potential options, considering items such as privacy and security that are outside the scope of this report, when determining specifications designed to address this report's recommendations.

3.0 METHODOLOGY AND ASSESSMENT AREAS

3.1 METHODOLOGY

The objective of this engagement was to assess EEOC's knowledge of data analytics, strategies, and capabilities within mission-critical activities. To accomplish this objective, the assessment team relied on interviews/walkthroughs that were supplemented, as needed, by EEOC strategic plans, reports, and reviews. The evaluation did not include a thorough examination of underlying data, but instead focused on data flows and usage of data within the organization to guide decision-making processes.

This analytics assessment involved three distinct phases, each outlined in its own subsection to follow.

3.1.1 Introduction

The assessment team began the engagement with an entrance conference with multiple stakeholders at EEOC Headquarters in Washington, DC. At the entrance conference, the assessment team described the five areas of assessment (see Section 3.2) and communicated the scope of the engagement: data and processes that are related to the EEOC's core mission. During this presentation, the assessment team provided a list of specific questions to attendees to offer both guidance and structure to the evaluation and subsequent report.

To conclude the introduction phase, the assessment team held a follow-up meeting with multiple stakeholders to allow individuals representing offices within the EEOC to describe their involvement with data, be it through data collection, data generation, or data analysis and reporting. This larger meeting provided individuals located within various offices at EEOC Headquarters a chance to discuss, at a high level, their office's role with regard to data collection, reporting, and analysis. The evaluation team used this information to formulate a list of offices with whom to conduct additional, more targeted interviews.

3.1.2 Fieldwork

The assessment team then entered the fieldwork phase which was characterized by a series of targeted stakeholder interviews on 2 November 2017. These interviews, both at EEOC Headquarters as well as two EEOC District offices, provided the evaluation team an understanding of:

- Each office's primary operational processes,
- Where data is either used or generated by these processes, and
- The effectiveness of tools, technology, and analysis of data in bringing insights to both processes and resourcing decisions related to the EEOC's mission.

Stakeholder interviews were primarily structured in a question and answer format, with minutes compiled from each meeting. The number of targeted stakeholder meetings each office participated in was commensurate with the number of ways that office generates or uses data that, in the opinion of the evaluation team, held potential for data analytics to further the EEOC's mission. Although the evaluation team identified and requested most of the targeted stakeholder meetings, several meetings were held at the request of individual offices.

In addition to discussions related to collection and usage of data, the evaluation team also sought to gain knowledge of inefficiencies in current data collection, processing, or analysis activities. The assessment team supplemented its interviews with observation of the usage of such systems, such as the use of the Integrated Mission System platform (IMS) in the Intake Process that occurs in District Offices.

The Fieldwork phase concluded with the final district office meeting on 27 February 2018.

3.1.3 Reporting

Based upon the minutes and observations taken from all targeted stakeholder interviews and observational sessions, the evaluation team mapped stakeholder comments as well as its own evaluative comments to the five areas of assessment. This mapping provided the basis to create a findings outline, which was provided to the EEOC OIG.

The assessment team provided a follow-up presentation to EEOC stakeholders to outline the findings from the assessment. Based upon the comments received from this presentation as well as the EEOC OIG, the evaluation team drew upon its extensive experience in data mining and predictive analytics to develop the recommendations contained within this report.

3.2 Assessment Areas

Successful analytics programs not only demonstrate technical excellence, they must also deliver on the promise of providing actionable analysis and insights to best support the decisions and activities needed to further organizational mission. This analytics assessment covers five capability areas that are necessary to achieve analytics excellence:

Table 3-1: Analytics Assessment Focus Areas

Assessment Area	Summary of Items Covered	Why Important
Culture	The extent analytics benefit from: • Executive leadership • Collaborative environment • Culture of evaluation • Vision and awareness	A pervasive culture understanding analytics and emphasizing the acceptance of metrics to guide all levels of decisions must exist before the full benefits of analytical methods can be realized.
People	The extent the analytics team: • Understands organizational needs • Creatively approaches problems • Utilizes effective tools	The analytics team must understand and creatively approach organizational problems utilizing a breadth of knowledge and techniques in order to find optimal solutions to problems.
Process	The extent the analytics team: • Has a clearly defined process to understand a problem and derive actionable insights • Is able to repeat this process in an iterative fashion, offering improvements with each iteration	A process that takes time to understand core organizational problems, derive insights, receive feedback, and assess further refinements is needed to ensure analytic products remain relevant in addressing organizational needs.
Analytics Capability	The extent the analytics team demonstrates technical prowess to: • Apply the appropriate levels of sophistication to address organizational needs • Incorporate feedback into model evaluation and management • Prototype new technologies, creating metrics to evaluate potential new investments in tools and IT infrastructure.	The analytics team must not only understand how to implement the appropriate level of analytical sophistication to meet today's needs, but demonstrate ability to evaluate products and technologies to effectively manage capabilities to ensure lasting relevance and effectiveness.
Infrastructure	The extent the IT infrastructure: • Supports aggregation of data from different data sources • Has sufficient storage, memory, processing power • Can deliver analytic products / visualizations that are easily digested by end-users	The information technology infrastructure must be able to collect data from different sources, ensure the integrity of the data, process the data, and support effective delivery mechanisms (reports and visualizations) that foster end-user adoption of analytics in their decision-making processes.

4.0 CULTURE

A healthy, data-driven culture and mindset is the single best indicator of future analytics success. Though analytics is largely a technical discipline, technology alone is insufficient to drive organizational maturity and growth through analytic findings and outcomes. In short, fully realizing benefits from analytical investments requires:

- Awareness of how analytics can help solve agency problems,
- Trust that the analytical results are both valid and valuable, and
- Leadership to drive organizational change.

This assessment covers five specific cultural components necessary to achieve awareness and trust:

Table 4-1: Five Cultural Components of this Assessment

Cultural Sub-Area	Why Important	Summarization of Core Finding
Shared Vision for Analytics	Awareness of analytics, and understanding of the types of challenges it can help address, is needed before benefits can be fully realized. Vision establishes the future position of organization's analytic capabilities and provides a directed, fostering environment in which capabilities can evolve over time.	EEOC Executive leadership lacks awareness of the many ways an analytics program can help the EEOC better achieve its mission, resulting in an incomplete vision for enterprise-wide analytics. See Section 4.1 for details.
Executive Leadership	Executives establish funding, resource allocation, and priorities of an analytics group. Leadership also establishes "tone from the top" that promotes and endorses the use of analytics in making decisions.	EEOC Executive leadership lacks awareness of the many ways analytics can benefit EEOC, resulting in historical lack of strategic level executive endorsements. See Section 4.2 for details.
Culture of Evaluation and Improvement	Being aware of the benefits analytics can bring to each organizational challenge is only the beginning: the culture must be such that decision-makers are willing to try new approaches and follow the insights the data-based analyses provide. This also requires investment of time and resources to measure and assess the effectiveness decisions, pivoting to new approaches when evidence shows potential for improvement.	The EEOC should consider the large quantity of data in its possession to be a strategic asset that enables, rather than encumbers, its ability to accomplish its mission. See Section 4.3.
Collaborative Environment	Most organizations have either data or process silos, and some organizations have both. This often results in members from different departments preventing sharing of lessons learned from their past endeavors. Collaboration, including a willingness to openly share both data and insights with others, is a prerequisite for analytical results to leverage organizational knowledge.	Years of dwindling resources has resulted in a culture where each office prioritizes its own objectives instead of trusting and supporting enterprise-wide solutions. See Section 4.4.
Continued Education and Learning	Analytics itself is not a static discipline—it constantly uses data to assess its effectiveness, evolving to leverage both past insights and new technologies. Analytics, including artificial intelligence, will be able to address even more problems in the future than is the case today. A high-level awareness of new data sources, new technologies, and new techniques ensures the organization's vision can be refreshed periodically to incorporate new components.	The EEOC lacks a centralized analytics team, governing board, or analytics champion to foster or evaluate awareness of analytics. See Section 4.5.

4.1 Shared Vision for Analytics

Before embarking on any journey, it is useful to establish the destination as well as the direction in which to travel to get there. The same holds true for analytics: each organization's needs are different, and therefore each organization must create its own vision of how analytics best serves its operational and decision-making processes. A shared vision established by executive leadership should provide details on what an effective analytics program would look like within the EEOC.

4.1.1 Finding

EEOC Executive leadership lacks awareness of the many ways an analytics program can help the EEOC better achieve its mission, resulting in an incomplete vision for enterprise-wide analytics. A complete vision would encompass an understanding of various analytic techniques and their potential positive impact upon agency programs and processes. Without a full understanding, two distinct difficulties will arise when articulating vision:

1) **Commitment to reasonable growth in analytic maturity:** Growth in analytic maturity comes with costs and benefits. The costs are the time and resources needed to develop capability; the benefits come from awareness and acceptance of analytical results coupled with corresponding impact on decisions and processes. Analytic maturity models help establish boundaries that can help contain costs while fostering increased acceptance of analytical results throughout the organization.
2) **Definition of scope and type of analytical products desired:** Analytics covers a broad range of data and techniques, resulting in a wide variety of end-products for decision-makers to use. The scope and type of products is sufficiently broad that it can be difficult for decision-makers to understand and effectively utilize. Establishing a vision involving a specific maturity level with specific use cases helps focus awareness and acceptance of analytical results, facilitating their integration into the organizational processes that stand to benefit.

None of the enterprise-wide analytics endeavors encountered during the assessment focused on predictive analytics. Only a small number of office-specific, small-scope initiatives were noted to have any components that would be considered a more technically and statistically advanced use of analytics. Of these, none had fully functional models designed to persist over time. Rather, they were designed to address specific, one-time questions and, once answered, the effort was closed and models no longer used.

4.1.2 Recommendation

The EEOC Office of the Chair should establish and lead an Executive Data Analytics Board that recognizes and treats data as a strategic organizational asset. Adoption of this paradigm will foster projects and ongoing activities to be implemented by other bodies but governed by this board. As such, this Board will need to:

- Make/approve investment and resourcing decisions to support strategic initiatives
- Prioritize and review strategic and/or infrastructure-oriented initiatives
- Ensure program management activities are appropriately aligned with vision
- Ensure business units are engaged and supportive of data governance activities
- Foster a culture that values integrating data and model insights into decision-making processes.

To implement this vision, the Executive Data Analytics Board would create one or more governance bodies. These bodies would have specific roles within a data governance framework. This recommendation is intended to provide the EEOC flexibility in how best to implement the vision articulated by the Executive Data Analytics Board. However, if the EEOC is interested in an example of how this could be accomplished, the next section offers an example structure to accomplish these objectives.

Please note: the specific governance bodies created to fulfill this recommendation would inherit responsibility to implement many of the recommendations in this report. The recommendations in this report will reflect the specific names of the data governance bodies that are suggested in the following section. If the EEOC chooses a different structure or nomenclature for the components to implement the strategic vision, the responsible party names will need to be mapped to the EEOC's planned governance structure.

4.1.3 Implementation Guidance: Data and Analytics Governance

There are many different ways the above recommendation for a data governance framework can be achieved. This section provides an example of how two distinct governing bodies that report to the Executive Data Analytics Board may best serve the EEOC's needs:

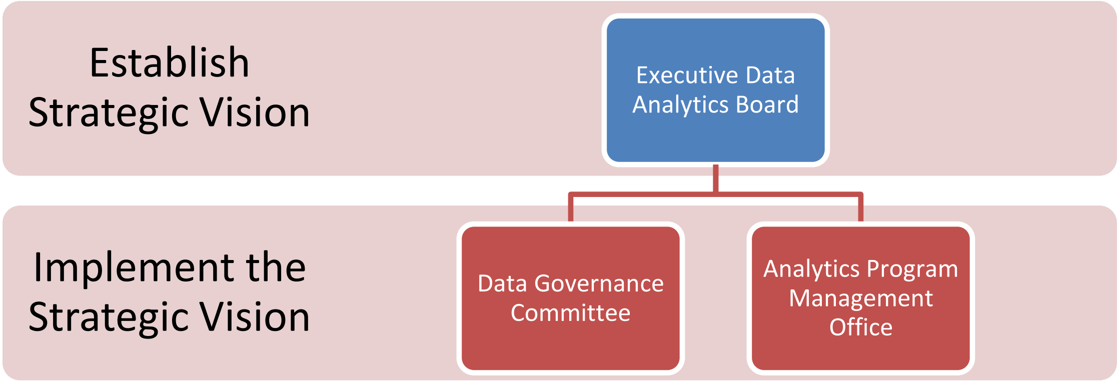

Establish Strategic Vision
Executive Data Analytics Board
Implement the Strategic Vision
Data Governance Committee
Analytics Program Management Office

Figure 4-1: Example Data Analytics Governance Structure

The table below provides details on the role of each body in this example Data Analytics Governance Structure:

Table 4-2: Data Governance Implementation Guidance

Governing Body Name	Area of Focus	Potential Members
Executive Data Analytics Board	A group that is responsible for articulating a strategic vision that recognizes and **treats data as a strategic asset** to the entire agency. This Board will need to: • Make/approve investment and resourcing decisions to support strategic initiatives • Prioritize and review strategic and/or infrastructure-oriented initiatives • Ensure program management activities are appropriately aligned with vision • Ensure business units are engaged and supportive of data governance activities • Foster a culture that values integrating data and model insights into decision-making processes. This board should also: • Designate liaisons between the other governance bodies • Designate a Data Champion to communicate its vision throughout the agency.	The EEOC Chair would be the leader of this governing body along with leadership from other offices, which should include, but not be limited to: • Chief Data Officer • Chief Information Officer • Chief Information Security Officer • Data Champion • Liaisons to other data governance bodies
Data Governance Committee	A group that is responsible for implementing the points of the strategic vision articulated by the Executive Data Analytics Board. This may involve strategic implementation of a diverse set of data governance objectives, including but not limited to: • **Data inventory and ownership designation:** establish processes to locate and inventory data assets, establishing ownership for each dataset. Data owners are responsible for understanding characteristics and use cases of their data. • **Data collection:** review infrastructure and processes (EEO surveys, portals, federal data, and integration of public data sources) • **Data generation:** explore ways to improve capturing of data from key business processes, such as intake/case management (IMS). • **Data augmentation:** explore ways to enrich data assets through augmentation with external public and/or private datasets. • **Data retention:** implement ways to store and track changes in data over time to promote consistency in reporting and analysis (example: data warehousing). • **Data security:** provide appropriate level of assurance regarding the confidentiality, integrity, and availability of data, in light of requirements posed by analytics team(s) and regulations.	Permanent members: • Chief Data Officer • Chief Information Officer • Chief Information Security Officer • Data Champion Other members should include one or more representatives from additional offices that are involved in collection or usage of data (i.e. data owners).

Governing Body Name	Area of Focus	Potential Members
Analytics Program Management Office	A group responsible for providing cross-project, program-level support for data reporting and analytics initiatives. Governance activities should provide project-based support, including: • **Tactical project prioritization and resourcing requirements**: review potential analytics projects and prioritize based on project feasibility and return on investment. • **In-Process project review:** review status of projects in-progress and ensure project outputs remain aligned with organizational goals and appropriate levels of sophistication and model management are employed. Intervene on blocked projects to minimize downtime. • **Post-completion project review:** review completed projects by engaging stakeholders to assess project outcomes. Considers stakeholder ability to leverage analytic product(s) received within operational or decision-making processes.	Permanent members: • Chief Data Officer • Analytics Team Manager • Data Champion Persons involved with currently active projects should also be involved for duration of their projects: • Project managers / leads • Project stakeholders • Data owners

This design addresses planning, resourcing, and implementation of strategic data initiatives set forth in the recommendation stated in Section 4.1.2. After each data governance body is established, specific activities to implement the Executive Data Analytics Board's vision can occur concurrently. Please note that the implementation of an analytics program need not wait for completion of all activities under the purview of the Data Governance Committee.

The recommendations in this report will use the governance body names outlined in this example structure. If the EEOC chooses a different structure or nomenclature for the entities responsible for implementing the strategic vision, the responsible party names will need to be mapped to the EEOC's planned governance structure.

4.2 Executive Leadership

As with most endeavors, success in analytics starts with endorsement from the highest levels of leadership within an organization. This is especially important in analytics: large potential benefits often lead to large expectations. Reality is that in most analytical endeavors, initial results may fall short of hyped expectations due to cultural or technical barriers that must be overcome. It is only through persistence in iterative processes that incorporate an honest, retrospective look of where results fall short can significant benefits be realized.

Persistence requires active management of expectations and allocation of resources not only to start projects, but to see them through what may be multiple iterations before benefits are fully realized. To be successful, such management needs to be backed by executive leadership, establishing a "tone from the top" related to analytics. Often, the designation of an Analytics Champion to communicate and implement executive endorsements is the most effective way to accomplish this.

4.2.1 Finding

EEOC Executive leadership lacks sufficient awareness of the ways analytics can benefit the EEOC's ability to achieve operational, tactical, and strategic goals. Because of this, analytic initiatives at EEOC have not benefitted from:

1) **Inclusion in Strategic Plan:** Despite a reference in the most recent Strategic Plan[4], support for an enterprise-wide approach to analytics has largely been absent in past EEOC strategic plans.
2) **A Champion to Spearhead Enterprise-Wide Analytics Program:** Such a champion, when backed by executive leadership, would foster better awareness of analytics, including both its benefits and its limitations.
3) **Sufficient Resource Allocation:** Executive leadership is required to allocate sufficient resources to both start and maintain an analytics program.
4) **Long-Term Analytics Strategy:** Executive leadership is required to endorse a long-term analytics strategy or vision.

4.2.2 Recommendation

The EEOC Office of the Chair should:

1) Review this report, in addition to other reports from applicable data governance bodies, to understand both current and potential uses for analytics within the EEOC.
2) Provide leadership, guidance, and resources to the Executive Data Analytics Board in assessing and prioritizing analytical projects, advocating for the resources needed to support prioritized projects by demonstrating improved effectiveness and/or efficiencies in achieving EEOC's mission.
3) Designate an Analytics Champion to spearhead adoption of analytics, including fostering awareness of analytics and management of both resources and expectations throughout projects involving the development and implementation of analytical initiatives.

[4] https://www.eeoc.gov/eeoc/plan/strategic_plan_18-22.cfm

4) Advocate for greater inclusion of analytics in future updates of the EEOC Strategic Plan as well as progress within reports sent to the Executive and Congressional branches of government.

In short, the EEOC should establish a “tone from the top” that designates enterprise analytics as an organizational priority and conduct steps to ensure adequate resources, including an analytics champion, are allocated to see prioritized projects through to fruition. This tone should establish mindset that potential reporting and analytics projects (i.e. projects designed to leverage data to enhance the EEOC’s ability to efficiently and effectively accomplish its mission) are prioritized based on quantifiable cost-benefit analyses along with expectation that results will be regularly re-assessed to determine continued efficacy.

4.3 Culture of Evaluation and Improvement

With respect to analytics, a culture of evaluation and improvement has four salient characteristics:

1) **Willingness to try new approaches:** From decision-makers down to each employee who supports the organizational mission, analytics holds potential to challenge commonly held assumptions. By definition, the process of “gaining insights” implies reaching a deeper understanding than was had previously. Willingness to follow these insights is required to fully realize the benefits of data analytics.

2) **Data is considered a valuable asset:** Data analytics requires data, either generated internally or obtained through outside sources. The insights that analytical endeavors provide spring forth from the available data, with insights being only as useful as the data used to generate them. Success stems from cultures that demonstrate the “data is a core asset” mindset by prioritizing data collection and quality assurance.

3) **Value is placed on numerical measurement:** A culture of evaluation and improvement allows sufficient resources to be dedicated to accurately create metrics designed to measure effectiveness in a manner that is free of bias or pre-conceived notions.

4) **Pivoting decisions:** Each decision made during operations represents selecting one path amongst multiple available options. Decisions are either validated or refuted by measurements of their effectiveness. A culture of evaluation and improvement must be willing to admit a decision, or even a data model, is sub-optimal when sufficient data suggests this is the case. Such a culture is willing to pivot to another decision in as part of a process designed to foster continual improvement.

4.3.1 Finding

The EEOC should consider the large quantity of data in its possession to be a core asset that enables more efficient and effective approaches in accomplishing its mission. Instead, the prevailing mindset is that that data collection and quality assurance is "an extra task" that detracts from mission-accomplishing activities. For example, several staff members from Office of Field Personnel (OFP) explained that the IMS pop-ups designed to ensure data quality from the intake process had become so numerous they are often ignored. In addition, a manager from Office of Federal Operations (OFO) reported that despite sufficient quantity of data available for the federal sector, the team lacked tools to quickly explore relationships between datasets, making such explorations a burdensome process. Such examples demonstrate a preference towards conducting operational activities that has led to an underinvestment in infrastructure to support automated processes around data.

A further consequence of underinvestment is that it results in further increases to the reporting and data collection burden within the organization's front lines. This negative feedback loop fosters a culture that believes the costs of evaluation are significant and rarely worthwhile. This leads to a lack of quantitative metrics that demonstrates the value and efficiencies that can be unlocked through new approaches and better infrastructure.

4.3.2 Recommendation

Fostering a culture that prioritizes evaluation will involve reducing the opportunity cost, real or perceived, associated with measurement, data collection, and quality assurance activities. In short, the EEOC Executive Data Analytics Board should conduct activities to demonstrate that increases in efficiencies will ultimately *reduce* burden of workers in the long-term.

To accomplish this, the EEOC Data Analytics Board should consider following a multi-step process that invests in the generation of new metrics that quantify the real opportunity costs and corresponding benefits of data collection and quality assurance activities. This process should also demonstrate gains in efficiencies resulting from remediation efforts in order to engender further cultural acceptance of evaluative, metric-driven approaches. The steps needed to start this process include:

1) **Assess opportunity costs of current measurement methods:** Current methods of reporting progress on certain objectives at the EEOC are only partially automated and require substantial time to generate reports. An agency-wide initiative to capture the amount of time spent reporting progress and status towards goals would uncover where inefficiencies exist.

2) **Quantify inefficiencies and target investments to address them:** Once these inefficiencies are quantified, a true cost-benefit analysis of infrastructure investments

designed to automate core reporting processes can occur. Armed with measurements, these investments can not only be targeted, the improvement in efficiency can be quantified and translated into quarterly or annual cost savings.

3) **Foster culture of evaluation:** Leveraging the quantified efficiency gains above, the EEOC should maintain momentum by having the Data Governance Committee study other areas where inefficiencies may result. While this may continue to involve reporting, it should also involve data collection activities, both externally sourced (EEO surveys and corporate RFIs) and internally-sourced (especially data generated from intake and investigation processes). Create metrics to evaluate the efficiency of the infrastructure utilized by personnel to achieve their tasks.

4) **Foster culture of improvement:** Measure and quantify inefficiencies, conduct cost-benefit analyses, and prioritize specific investments to address the most pressing inefficiencies. Utilize pilot programs to test the effectiveness of planned improvements in order to engender buy-in from front-line workers on new ways of accomplishing tasks. As management demonstrates increases in efficiencies from such efforts, a positive feedback loop will form: past successes engender a culture that fosters further ideas to increase efficiency and effectiveness across the organization.

4.4 Collaborative Environment

Data analytics requires collection and sharing of data sources that are rich with patterns and, in the case of supervised approaches, outcomes and/or lessons learned. In order to extract insights from data that can have an enterprise-wide positive impact, the data used in the analysis would optimally have an enterprise-wide scope. Although many organizations may have data or process silos, the ability to create broad data sets and share insights between offices stems from shared desire of each office to address common challenges in implementing a shared organizational mission.

4.4.1 Finding

Years of dwindling resources has resulted in a culture where each office prioritizes its own objectives instead of trusting and participating in enterprise-wide solutions. During the assessment, stakeholders from multiple offices recounted past enterprise IT initiatives that created working groups or steering committees that collected and published requirements, only to lose the resources needed to complete the effort. This resulted in the working groups eliminating large groups of requirements, often with impacted offices receiving little to no communication of such decisions.

For some offices, the impact of these decisions were quite high. In light of this environment, multiple offices have created stopgap solutions to address their needs. These stopgap solutions are often in the form of spreadsheets to track operational items for small groups of people, such as in-person interviews of potential charging parties conducted at district offices. Although these stopgap solutions are necessary to achieve specific office objectives, they neglect to adequately:

1) **Leverage data and resources from other offices:** Office-specific workarounds are created by people within specific offices, and therefore contain only the ingredients their creators are aware of.
2) **Create insights that are shared with other offices:** Results are utilized to achieve only office-specific goals and are not shared with other offices because of disincentives, as performance metrics evaluate each office on getting its job done, not how much it invests time to share creative approaches that can be customized and applied elsewhere. The consequence of this is that Office A may have found a way to collect and leverage data to address an operational need that Office B also shares, but since Office B lacks the data or resources needed to devise a solution, Office B's needs remain unmet. Thus, Office A operates with the problem solved while Office B suffers the inefficiencies associated with an unsolved problem. For example, both districts visited by the assessment team reported that for most action items, IMS does not provide reminders or other "to-do" lists associated with each person's backlog. Each district has devised its own unique spreadsheet system, which covers different action items for different types of charges, in order to address only the most pressing needs of each office.

Because of the natural alignment of shared objectives in district and field offices, collaboration within those offices was found to be more prevalent than within headquarters or across district offices.

4.4.2 Recommendation

The EEOC Office of the Chair should consider ways to engender trust in enterprise-wide steering committees and governance boards. Failure to engender this trust creates a high risk that any enterprise-level IT or analytics initiative will lack sufficiently diverse stakeholder participation in solutions. Some potential remedies may include, but are not limited to, the following:

1) **Improved communication of goals and achievability:** Participants in requirements-gathering committees want to know early in the process the extent to which their desired outcomes can be achieved so they can gauge the return on their time investment in participation in the group. This thought process should be *encouraged* as it is an example of a data-driven decision. However, the decision to participate must be driven by good data on what is possible. Committees must therefore communicate realistic goals to stakeholders at the outset, and these goals should be backed by the highest levels of leadership within the organization.

2) **Improved communication of deprioritized requirements:** If the environment in which the committee operates changes, such as a reduction in funding, the committee should quickly and clearly communicate to each stakeholder which requirements are deprioritized and why. This allows better management of expectations and prevents trust erosion associated with disappointing, unrealized aspirations after long waiting periods with little to no updates provided.
3) **Establish incentives to share workarounds:** Some office-specific problems have already been solved using specific subsets of data that simply need scaling to benefit other offices within the organization. Foster incentives for such offices to share their home-grown solutions, coupled with a promise that once shared, that solution will not be replaced until another solution can demonstrably meet or exceed those specific needs.

In short, the EEOC must study and assess ways to engender trust in enterprise IT and analytics initiatives to engender the sufficiently diverse participation needed to ensure the ultimate success of those initiatives.

4.5 Continued Education and Learning

Analytics is a rapidly evolving discipline. Although originally rooted in reporting and descriptive statistics, analytics now involves statistical analysis, data mining, simulation, and machine learning/artificial intelligence. As this discipline grows to employ more advanced approaches and techniques, the analytics team must embrace continued education and learning to properly assess applicability of these advances in solving specific organizational problems. This awareness helps foster a culture where stakeholders can identify challenges and effectively communicate their need for a solution to an enterprise analytics team or governing board.

Going forward, assessment of the effectiveness of this cultural component is best performed by the analytics team or governing board. This team is in the best position to assess whether people throughout the organization are able to "ask the right questions" to facilitate actionable responses from the analytics team or governing board.

4.5.1 Finding

The EEOC lacks a centralized analytics team, governing board, or analytics champion to foster or evaluate awareness of analytics. The effect of this condition is that most EEOC employees interviewed in this assessment did not initially understand the benefits and efficiencies that can be unlocked through an enterprise analytics program.

4.5.2 Recommendation

The EEOC Office of the Chair should designate an Analytics Champion to foster awareness and education of the ways analytics can address inefficiencies, solve problems, and unlock hidden value in data. This champion should work to bridge the gap between end-user requests and technical requirements of an analytic teams, fostering a culture where data is viewed as a core asset that can better enable more efficient and effective processes to accomplish the organizational mission.

5.0 PEOPLE

This assessment has two evaluation areas that are specific to the analytics team(s) that exist within the organization:

1) **People:** Focuses on the analytics team's interaction of business units to gain understanding of organizational problems and essential components that must be addressed in solutions. Also focuses on the team's ability to leverage creativity in the problem-solving process, applying a breadth of knowledge along with the appropriate level of analytics in solving those problems.
2) **Analytics Capability:** Focuses on the technical acumen, including analytical sophistication, evaluation approach, model management processes, and tools/technology available to the analytics team both for analysis of data as well as delivery and/or visualization of results. Can be thought of as "hard, technical" capabilities of the team, assessing the maximum capabilities of the team even if those capabilities have not yet been brought to bear to solve a problem.

This section of the assessment report covers the "People" evaluation area. In short, this area evaluates the extent to which the analytics team(s) successfully apply established problem-solving processes while leveraging both creativity and wide breadth of knowledge to find optimal, effective solutions.

The "People" evaluation area covers four specific components of the analytics team:

Table 5-1: Four People Components of this Assessment

People Sub-Area	Why Important	Summarization of Core Finding
Understanding of Business Needs	Analytics is first and foremost focused on leveraging data to solve problems that can improve operations and decision-making capability. To be effective, the analytics must address the right questions and understand the context in which the questions are asked.	EEOC lacks a centralized analytics team that focuses on understanding processes and gathering requirements from multiple stakeholders. See Section 5.1 for details.
Technological Breadth	Many tools, techniques, and technologies exist to analyze data and provide insights. This evaluates the extent that the analytics team demonstrated the proper understanding and application of the available tools and techniques that exist within the EEOC analytics toolkit.	EEOC lacks a centralized analytics team to assess application tools and techniques within an enterprise-wide framework. See Section 5.1 for details.
Creativity	The problems that are presented to analytics teams are often the hardest within the organization, otherwise departments/offices would have already solved them internally. Difficult problems often require novel, creative approaches to foster breakthroughs.	EEOC lacks a centralized analytics team to assess its creativity. See Section 5.2 for observations on the creativity employed by analysts in specific areas.
Analytical Knowledge and Skills	Analytics teams need to be able to demonstrate success applying appropriate levels of analytics in solving organizational problems, including communication of the assumptions and limitations inherent within the techniques employed. Correct application also requires staying abreast of new techniques that may be more effective and/or have fewer assumptions and limitations.	EEOC lacks a centralized analytics team to assess its application of knowledge and techniques. See Section 5.3 for discussion on the application of knowledge and skills found in specific areas.

5.1 Understanding Business Needs, Technological Breadth

Analytics teams must be able to engage people of various levels across the organization in order to gain understanding of the problems the organization faces. In many regards, the extent to which an analytics project can succeed is based on the analytics team's ability to empathize with stakeholders, truly understand their needs, and assess ways of solving those needs which are consistent with the "big picture" needs of the organization as a whole. Analytic products need to address all requirements unless technical limitations prevent this from occurring.

Similarly, analytics teams need to demonstrate understanding and proper application of the tools and techniques available to them to address organizational needs. This involves understanding and application of available software tools, including commercial off-theshelf software (COTS), customized software, and open source platforms and packages. This also involves demonstrated understanding and application of appropriate levels of analytics (see Section 6.2 for details) needed to address the organization's most pressing analytical problems.

5.1.1 Finding

EEOC lacks a centralized analytics team that focuses on understanding requirements and delivering analytical solutions to multiple stakeholders throughout the enterprise.
EEOC does have some problem-specific analytics capability:

1) **Office of Research and Information Planning (ORIP):** The Program Research Branch employs highly trained professionals, including Labor Economists, Statisticians, and Survey Specialists, who assist with gathering data and developing statistically testable "theories" for Systemic Investigators. The techniques employed here are statistical in nature and are among the most advanced analytics currently performed in the agency. However, the nature of this work is both specialized (designed specifically for a targeted group of stakeholders) and is addressed on a case-by-case basis. Because of this, the concept of gathering business requirements from multiple stakeholders across the organization for a persistent analytical product does not apply.
2) **Office of General Counsel (OGC):** OGC has a small team of dedicated analysts within the Research and Analytic Services (RAS) group. This group provides expertise and analysis specifically on cases that are sent to EEOC legal teams. Similar to the ORIP Program Research Branch, the analysis activities of this group are amongst the most advanced within EEOC but are largely focused on addressing case-specific needs. Although some RAS analytical products persist longer, these address specific needs of the legal department that are outside the scope of the services a centralized analytics team would provide. Because of this, the concept of gathering business requirements from multiple stakeholders across the organization for a persistent analytical product does not apply.

Similarly, these two groups have specific analytics toolkits designed to address their specific use cases. Both of the above groups have access to newer computer hardware and software resources, such as SAS, R, and other statistical tools, that the standard EEOC employee lacks access to. This assessment found no evidence that these groups lacked breadth and understanding of their available toolkits.

5.1.2 Recommendation

The EEOC Executive Data Analytics Board should work with a high-ranking executive or Office Director, such as the Chief Data Officer, to establish a centralized analytics team that is available to all offices across the organization to address unmet strategic data analytics and reporting/visualization needs. A dedicated analytics team would allow the EEOC to:

1) **Address problems strategically:** The EEOC can leverage a dedicated team to grow beyond the case-by-case approach for analysis to address needs strategically and proactively.

2) **Prioritize and address problems with "big picture" organizational knowledge:** The advantage of an enterprise-wide analytics team is the enterprise-wide, big picture perspective on current and emerging problems. This comes naturally from investing the time needed to gain understanding of the needs of a diverse set of stakeholders and affords data governance bodies additional inputs to better prioritize projects.

3) **Maintain consistency in solutions and on-going evaluations:** A centralized analytics group can better understand enterprise-wise requirements and can customize consistent user interfaces that members of one office can easily learn upon moving to another office. Similarly, a centralized group can foster consistency in the creation of bias-free metrics to evaluate efficacy of analytical solutions.

4) **Hire appropriate skillsets:** Most enterprise-wide analytical needs would not require such specific expertise, but would require cross-discipline thinking that more generalized data science and analytics training and experience provides. A centralized analytics group should consist of members with a technical background, such as Data Scientists, Statisticians, and Computer Scientists, who demonstrate the ability to empathize and communicate well with stakeholders to gather requirements, both spoken and implied, translating those requirements into solutions that best address organizational needs. A centralized analytics team may also facilitate hiring the appropriate technical skill sets.

5) **Leverage economies of scale:** Some of the tools, both software and hardware, utilized by analytical groups can be expensive. A centralized analytics team would allow the EEOC to better leverage software and hardware investments to realize benefits that enjoy an enterprise-wide scope.

In short, a centralized analytics team, would effectively leverage limited infrastructure and resources to address a wide range of currently unmet needs. Furthermore, this structure provides the added benefit of increasing capabilities without reorganizing the two case-specific analytical teams who are already well-suited to address their specific cases.

Note: The two case-specific analytical groups, ORIP and OGC, are staffed by specialists suited to meet their specific needs. As such, these groups are out-of-scope of this recommendation and should continue to function as currently structured as long as they continue to meet the needs of their specific stakeholders.

5.2 CREATIVITY

Although analytics teams need to have a deep understanding of statistics, data science, and computer science to be effective, excellence requires leveraging creative approaches in solving problems. This is a reflection of the fact that data science does not exist in a vacuum—some approaches will work better than others for certain types of stakeholders and certain types of problems. The wider the breadth of problems the analytics team is expected to address, the more important creativity becomes. The ability to effectively apply creativity is a key differentiator between generalized data science and specific use-case statistical analysis.

Each step CRISP-DM process outlined in Section 7 requires at least some degree of creativity to be performed effectively for the desired use case. Examples of where creativity is important in the data science process include:

1) **Business Understanding:** Over time, a centralized analytics team that is allowed to address problems throughout the organization will not only learn subject matter expertise, but be able to make creative connections between similar problems solved in the past.
2) **Data collection:** Models are only as good and complete as the data used to build them. Creativity may be required to determine which data sets can be brought to bear to solve a problem, as certain types of data may need to undergo a series of intermediate joins before it can be aligned to the problem at hand.
3) **Data preparation and imputation:** Reality is that most data sets have imperfections, such as missing data, potentially erroneous data, and data in a format that is not conducive to analysis. Creativity is often required to effectively impute missing data, assess whether outlier points are valid or in error, and finesse the data into a format appropriate for the problem at hand.
4) **Feature engineering:** A key component to technical success of a model is creation of features that directly highlight patterns in the data. This serves to simplify models, making them more effective and interpretable, by leveraging domain-specific knowledge. The process of creating features that add value often requires creativity and application of subject matter expertise.
5) **Model selection and interpretability of results:** Understanding and creatively guiding solutions that recognize the tradeoffs between accuracy and interpretability of results can go a long way to creating analytic products that are actionable by end-users. This includes selection of the appropriate level of analytics needed to address the business problem at hand.

6) **Evaluation:** Creating metrics that fairly and succinctly capture multiple aspects of model performance in an appropriate format for the business problem often requires creativity. This is especially true when the cost of incorrect results, such as false positives and false negatives, are high.

7) **Deployment and visualization of results:** Creativity may be needed to capture complex insights into results that the consumers find simple and intuitive to understand. Creativity can be leveraged to creative striking, intuitive visualizations; lack of creativity leads to spreadsheet results that users must read lengthy instructions to interpret. When creativity is not leveraged here, results are often ignored or misinterpreted.

5.2.1 Finding

EEOC lacks a centralized analytics team to assess its creativity. Where analytics capability does exist, i.e. ORIP Program Research Branch and the OGC Research and Analytic Services (RAS), use cases are domain-specific and therefore do not require the same level of creativity that a generalized, enterprise data analytics team would require. Because of this, assessment of the creativity of these groups was deemed out of scope for this assessment which is focused on evaluating enterprise-wide analytical solutions.

5.2.2 Recommendation

The EEOC Executive Data Analytics Board should work with a high-ranking executive or Office Director, such as the Chief Data Officer, to establish a centralized analytics team. This team can include individuals with experience in, but not necessarily limited to, data science, operations research, and statistics. Such individuals should be able to demonstrate problem-solving capabilities with easily interpreted results for a wide range of problems. Candidates working in this capacity should be able to demonstrate creative approaches to solving challenges that arise within the data science process, such as the seven areas outlined in Section 5.2.

Note: The two case-specific analytical groups, ORIP and OGC, are staffed by specialists suited to meet their specific needs. As such, these groups are out-of-scope of this recommendation and should continue to function as currently structured as long as they continue to meet the needs of their specific stakeholders.

5.3 Analytical Knowledge and Skills

Analytics has many moving parts, and to be effective, analytics teams must apply knowledge and skills appropriately across each step of the process. Examples include:

1) **Business understanding:** Successful analytics stems from the first step of the analytics process: gaining understanding of the business need to address. Effective analytics teams demonstrate a willingness to engage stakeholders many times throughout an analytical project, seeking to gain insight both on stakeholder needs and subject matter expertise.

2) **Data understanding, preparation, and feature engineering:** Proper analysis of data requires a deep understanding of the data. In fact, upwards of 80% of the time and effort spent on data science project involves these important steps. Effective analytics teams demonstrate ability to not only interpret data, but also to apply subject matter expertise to engineer features that highlight patterns in the data and systematic ways to evaluate feature importance.

3) **Model selection and interpretability of results:** Many problems lend themselves to more than one solution. Within the realm of analytics, multiple approaches can be employed, each with its own benefits and drawbacks, to gain insight or solve a problem. Unlocking value therefore depends on being able to recognize and select amongst candidate solutions, taking into account which solution "best" meets the needs of stakeholders. To accomplish this, effective analytics teams understand and stay abreast on all available approaches, demonstrating ability to select the correct type and level of models that provide appropriate combination of complexity, interpretability, and accuracy. All of this must be done while avoiding common mistakes that undermine the effectiveness of results.

4) **Evaluation:** Evaluation of model performance, both before and after deployment, is critical to ensure success of data-driven approaches. Effective analytics teams ensure models are cross-validated, contain signal and are not overly sensitive to random patterns, are tuned in accordance with costs of errors, and remain so after deployment. When models begin to erode in performance or no longer meet expectations, model management techniques are followed to ensure continued efficacy of solutions.

5) **Deployment and visualization of results:** To bring value to the organization, models must be usable by their stakeholders. Effective analytics teams demonstrate creative ways to best package results in a fashion easily consumed by stakeholders while effectively communicating the assumptions and limitations of the models.

5.3.1 Finding

EEOC lacks a centralized analytics team in which to assess its application of knowledge, skills, and techniques. Where analytics capability does exist, i.e. ORIP Program Research Branch and the OGC Research and Analytic Services (RAS), analysis is performed on a case-by-case basis that is often focused on validating a hypothesis. Although the above steps still apply for this type of analysis, the intensity in which they must be adhered to matters less since those results are utilized only a small number of times. Because of this, an evaluation of the analytical knowledge and skills of these groups was deemed out of scope for this assessment which is focused on evaluating enterprise-wise analytical solutions with models designed to persist as long as they continue to demonstrate effectiveness.

5.3.2 Recommendation

The EEOC Executive Data Analytics Board should work with a high-ranking executive or Office Director, such as the Chief Data Officer, to establish a centralized analytics team that is available to all offices across the organization to address strategic data analytics and reporting/visualization needs. This team should be staffed with talented individuals experienced in data science and risk management of data science projects. That can include, but is not necessarily limited to, data scientists, operations research professionals, and statisticians who are able to demonstrate appropriate application of knowledge and skills that address technical challenges that arise within the data science process, such as the areas outlined in Section 6.

6.0 ANALYTIC CAPABILITY

This analytics assessment is concerned with organizational capabilities related to the utilization of data to understand current status and make better-informed decisions related to the organizational mission. This assessment area contains five sub areas as summarized in the table below:

Table 6-1: Five Analytic Capability Areas of this Assessment

Analytics Capability Sub-Area	Description	Summarization of Core Finding
Analytic Sophistication	Evaluation of the analytic team's knowledge of different levels of analytical sophistication, including the team's capability to effectively map diverse organizational needs to the appropriate level of analysis.	EEOC lacks a centralized analytics team that focuses on understanding processes and gathering requirements from multiple stakeholders. See Section 6.2 for details.
Modeling Process	Evaluation of the analytic team's capabilities in translating the business problem to one or more competing models that can be packaged to appropriate respond to end-user needs.	EEOC lacks a centralized analytics team to modeling process within an enterprise-wide framework. See Section 6.3 for details.
Evaluation Approach	Evaluation of the analytic team's capabilities in quantifying quality of results and evaluating which of the competing candidate models are best suited to address each business problem.	EEOC lacks a centralized analytics team to assess its evaluation approach. See Section 6.3 for details.
Model Management	Evaluation of the analytic team's capabilities in versioning models, monitoring model effectiveness, and applying lifecycle concepts to a model.	EEOC lacks a centralized analytics team or data governance board to assess model management techniques. See Section 6.3 for details.
Tools and Technology	Evaluation of the analytic team's implementation of the tools and technologies available to it for creation of prototypes, pilots, and production models. Includes communication of emerging technology needs with Information Security and OIT and an evaluation of exploration of new technologies via prototypes and experiments.	EEOC lacks a centralized analytics team to assess implementation of tools and technologies used in its analytic process. See Section 6.4 for details.

6.1 Background

This analytics assessment is concerned with organizational capabilities related to utilizing data to understand status and guide the decision-making process. This belies two distinct activities[5]:

1) **Reporting:** The process of organizing and combining data to accurately describe the organization's current status. Report products compile and present information for delivery in a static format (paper or PDF) or dynamic/interactive format (drill-down menus, interactive visualizations, etc.).
2) **Data Analytics:** The process of inspecting, cleaning, transforming, and creating data-based models with the goal of discovering useful information and insights that directly support decision-making activities. Delivery can be static (spreadsheet or simple graphic) or dynamic (interactive visualization, scenario-based updates, etc.).

While reporting is often a prerequisite for organizations to understand their status and perform more advanced data analytics, this and subsequent sections of this report focus on data mining, predictive analytics, and beyond. In short, analytics capability evaluates the capability of the organization to utilize data mining and predictive analytics in a manner that best empowers end-users to understand and act upon insights from data.

6.1.1 Background: Levels of Data Analytics

Regardless of delivery method, data analytics itself comes in different flavors that are associated with different use cases. The main types of analysis include[5]:

1) **Statistics:** The use of deterministic methods and formulas to calculate statistical measures, often to prove or disprove a testable hypothesis.
2) **Data Mining:** The use of inductive processes that detect hidden/unknown patterns within data from which insights can be inferred. This is beyond the more formulaic, deterministic capabilities associated with spreadsheets, often employing a form of machine learning to derive insights. There are two main types of machine learning:
    a. **Unsupervised Learning:** Explores relationships between observations and features within data without utilizing labeled responses. Sample uses include, but are not limited to, clustering, anomaly/change detection, and dimensionality reduction.

[5] Deal, Jeff, et al. *Mining Your Own Business: a Primer for Executives on Understanding and Employing Data Mining and Predictive Analytics*. Data Science Publishing, 2016. (Chapter 2)

b. **Supervised Learning:** This form of machine learning leverages targets that denote "outcomes" and utilizes techniques to learn which characteristics are most closely associated with the various outcomes.

3) **Predictive Analytics:** Leverages statistics, data modeling, data mining, and sometimes human-expertise to predict outcomes designed to facilitate decision-making. Predictive analytics aggregates results from the statistics and data mining steps utilizing well-defined model(s) that address model testing and validation, fine-tuning of thresholds based on the cost of errors for the intended application, and creation of a human-interpretable result (such as a score). The salient knowledge sought is a probability of a defined outcome, sometimes denoted as risk.

4) **Prescriptive Analytics[6]:** Also known as Uplift Modeling, this leverages results from predictive analytics, attempting to "prescribe" possible actions related to a decision. The salient knowledge sought is impact of the treatment, not the estimate of the outcome. As this requires assessing effects that are not directly measurable, it employs highly sophisticated techniques.

6.1.2 Background: Pattern Recognition vs. Knowledge Discovery

Data mining and statistics are focused on creation of metrics to describe data, through summarization, classification, hypothesis, or probability of a specific outcome. These methods provide insight by recognizing distributions or patterns within data.

Conversely, predictive and prescriptive analytics are focused on knowledge discovery. This may leverage codified human expertise (as available) along with pattern recognition from one or more statistical and data mining steps. Output of predictive analytics products should not be merely a raw number, category, or ranking, but rather be tailored to inform business decisions. Within knowledge discovery, the salient knowledge characteristic differs between predictive analytics and prescriptive analytics:

1) **Predictive Analytics:** Based on combinations of past patterns in data and codified human expertise, a quantified probability or human-readable score is created to provide classification, ranking, or scoring of all data points on the assessed criteria.

2) **Prescriptive Analytics:** Demonstrates the impact of alternative treatments by comparing the confidence intervals of the quantified predicted results of each outcome (step 1), then comparing amongst those outcomes to find which are significantly different than others (step 2).

[6] https://www.elderresearch.com/hubfs/Whitepaper_Uplift-Modeling-Making-Predictive-Models-Actionable.pdf

6.2 ANALYTIC SOPHISTICATION

Although some levels of analytics and their associated techniques are more sophisticated than others, higher sophistication does not automatically imply higher levels of effectiveness. What matters most is whether the analytics team(s) understand and know how to apply the various levels of analysis as described in Section 6.1. Specifically, can the analytics team(s) effectively prioritize and map the business problem to the appropriate level of analytics given the available data and organization's requirements for actionable insights?

The below table provides some examples demonstrating the type of question each level of analysis can answer as well as example assertions made at each level, providing use cases that could be applicable to the EEOC:

Table 6-2: Types of Analytics plus Example Questions and Hypothetical Statements

Type of Analysis	Example Question This Analysis can Answer	Hypothetical Assertion that can be Made from this Analysis Type
Statistics	Is company X's age distribution of employees statistically different than its industry peers?	Data indicates that at the 95% confidence level, it is appropriate to reject the null hypothesis that Company X's age distribution is not different than its industry peers. Therefore, we assert that age distribution is significantly different.
Data Mining: Unsupervised Learning	What types (clusters) of potential charging parties are there and what are their salient features?	According to the clustering model, there are 4 major types of charging parties. Type A predominately has these characteristics, Type B these other characteristics, etc.
Data Mining: Supervised Learning	What is the expected pay range of an employee with a bachelor's degree and 12+ years of experience in industry X?	According to the regression model, such a person should have an expected pay range of $65,478 - $86,261, coming from a base salary plus adjustments for bachelor degree, age, industry, and other characteristics. (each value quantified)
Predictive Analytics	Given the company information available for all pending charges, what is the risk score for company X that it exhibits systemic issues related to equal opportunity employment?	Given the information from the EEO-1 survey plus additional information received from the company, company X has a risk score of 955 of 1000, placing it in the top 5th percentile of all companies in such assessed risk.
Prescriptive Analytics / Uplift Modeling	What features related to a charge should I mention to company X to maximize their likelihood of participating in ADR?	When speaking with a company about ADR options, companies like company X exhibit a statistically significant increase (average 20% uplift at the 90% confidence level) in participating in ADR if I mention HR policy issues than if policy issues are not mentioned.

6.2.1 Finding

Numerous process and decision-making areas within the EEOC lack access to an analytics team. This deprives the organization of numerous opportunities to systematically collect data and apply advanced analytics to address business problems. These lost opportunities to identify and analyze problem areas means that many organizational needs remain unknown and therefore unmet. The sample projects listed in appendix 2 (Section 9.2) provide several examples of such unmet needs.

Two specific-use-case analytics teams were found to exist within EEOC. However, in both cases, these teams are comprised of personnel with specialized skillsets who are given specific, limited scope problems. These use cases do not reflect the enterprise-wide scope of organizational problems this assessment aims to evaluate. Recommendations associated with this finding are not intended for these groups, which handle their specific functions effectively. For completeness, a description of these teams follows:

1) **ORIP Program Research Branch:** This analytics group focuses on application of statistics to test one or more hypotheses. Although case-by-case assessments are conducted to formulate and test hypotheses, this team's analytics experience is largely limited to the application of statistics. Special projects requiring the use of more sophisticated levels of analytics sometimes exist, and when they occur, personnel involved appropriately employ data mining techniques. This group mainly employs statisticians or graduate-level social scientists with demonstrated experience in statistics, appropriate for the scope of questions addressed.
2) **OGC Research Analytic Services:** This analytics group focuses on application of statistics, data mining, and predictive analytics in support of specific legal cases. This group appeared to appropriately utilize the most sophisticated analytical techniques the evaluation team observed within EEOC. However, given RAS's business function and scope of only supporting efforts related to the legal function of EEOC, RAS lacks the enterprise-wide scope this assessment aims to evaluate and was therefore deemed out-of-scope for assessment of analytical capability. This group mainly employs attorneys and other legal-oriented professionals, appropriate for the scope of questions addressed.

6.2.2 Recommendation

The EEOC Executive Data Analytics Board should work with a high-ranking executive or director, such as the Chief Data Officer, to establish a centralized analytics team. This team should not supplant the existing ORIP or OGC analytics capabilities, which should remain focused on their areas to continue leveraging the legal and social scientist skills those domain-specific problems require. Because of broader scope, personnel in a centralized analytics team need not have specialized social scientist or legal training. Rather, members of a centralized analytics team should have demonstrated experience in predictive analytic techniques and tools as well as the ability to effectively communicate and empathize with stakeholders, demonstrating ability to understand and devise solutions to business needs. Such personnel should work closely with the Executive Data Analytics Board or Data Governance Committee to advise on data quality and data analysis needs.

6.3 Modeling Process, Evaluation, and Management

Once the analytics team understands how to assess business problems and available data, mapping solutions to the appropriate level of analytics, the team must go about the process of building models. This sub-area evaluates the processes utilized to guide the project from idea to deployment. In the realm of predictive and prescriptive analytics, this involves:

1) **Data Discovery and Feature Engineering:** The process of collecting data, exploring data, performing quality control on the data, placing the data into the appropriate format for analysis, and creating features to enhance pattern recognition ability of data mining or machine learning techniques.

2) **Encoding Business Understanding and Feature Selection:** The process of integrating human expertise (if available or desired) and assessing which features are appropriate to include in various models to maximize their effectiveness.

3) **Evaluating Effectiveness of Competing Models:** The process of using training, validation, and test sets with appropriate data quantities to assess the error rate of the model. May also include combining multi-step models, creating model collections (ensembles), and applying human-interpretable scores. Evaluation of model effectiveness should be approached in the context of the intended use case and anticipated future data inputs, evaluating accuracy, interpretability, and limitations of analysis.

4) **Deploy Models:** The process of "productionizing" the model, with consideration given to the consumption of the analytic product's results, including scoring, visualization, and collection of end-user feedback.

5) **Model Management Techniques:** All models degrade in effectiveness over time as the world around them evolves. Are mechanisms in place to monitor long-term performance of such models and alert if performance either suddenly drops or drops below a specific threshold?

In short, to what extent does the analytics team demonstrate the capability to apply technical knowledge to implement the concepts in the Process section of this assessment?

6.3.1 Finding

EEOC's existing analytic solutions are primarily focused on solving problems on a case-by-case basis through the use of statistical analysis or basic data mining techniques. Analytic outputs are typically embedded in a report or a spreadsheet and are typically used only once to make a case-specific decision. For example, ORIP analysts work with Systemic Investigators to develop a case-specific hypothesis and then work to obtain the data from the organization in question to statistically test the hypothesis. As a result, models are rarely saved and updated for use more than once, creating a situation where many of the above steps are unnecessary overhead.

6.3.2 Recommendation

The EEOC Analytics Program Management Office should encourage a centralized analytics team to adopt proven modeling approaches and model management processes. This will enable the EEOC to move beyond the mindset of creating single-use, disposable models. In these new areas, the EEOC should aim to capture and leverage longer-term organizational knowledge, past results, and experiences by developing model frameworks designed to persist over designated periods of time. Such endeavors would allow the EEOC to utilize data mining and predictive analytics to address a much wider range of business problems.

New analytic products should not only help guide important decisions and establish operational priorities, but also increase efficiency of operations through collection of user experience data within EEOC applications by utilizing feedback loops. Feedback loops should foster creation of quantitative measures to more accurate gauge where future improvements and upgrades should be targeted with respect to processes and systems.

Longer-term, an outside entity not involved in the creation and deployment of analytic products should regularly evaluate the modeling process, model evaluation, and model management approach utilized by analytics team(s). This evaluation should include the extent to which the analytics team is exposed to new technologies and techniques and utilizes mechanisms, such as proof-of-concepts or limited scope pilot programs, to learn how to properly apply those technologies and techniques into production-quality analytic products using the processes outlined here.

6.4 Tools and Technology

An evaluation of analytics capability is not complete with a review of the utilization of tools and technologies by the analytics team(s) within the organization. This capability subsection is distinguished from the Infrastructure section of the analytics assessment as outlined in the below table:

Table 6-3: Differences between Tools & Technology vs. Infrastructure Sections

Evaluation Area	Scope
Analytics Capability: Tools and Technology	Focus on the software, tools, and technologies utilized strictly by the analytics team: • Does the hardware and software available to the team effectively meet the levels of analytic sophistication needed to address business problems? • Does the analytics team have the ability to communicate with IT and Information Security functions on its current and emerging software and hardware needs? • Does the analytics team have an organized code repository that supports versioning and documentation efforts? In short, this subsection evaluates whether the infrastructure used by the analytic team supports or inhibits effective data analysis and development.
Infrastructure (entire section)	Encompasses a broader, organization-wide view of IT infrastructure beyond that of the analytics team: • Organization-wide evaluation of data storage and warehousing capabilities, including the ability of different data systems to interface with each other. • Organization-wide evaluation of available compute power to analytic product end-users. In short, this subsection evaluates whether organization-wide infrastructure enables or prohibits effective data collection and effective delivery of analytic products.

Analytics teams need to have access to the tools and technologies needed to access data, assess its quality, prepare analytics base tables with engineered features, build and assess multiple types of candidate models, and prototype packaged analytic products. This includes components traditionally assigned to categories of "software" and "hardware":

1) **Analysis Infrastructure:** Is the technological infrastructure appropriate for the levels of analytics needed at the appropriate stages of analytical projects? The "software" component includes ability to evaluate and effectively utilize both commercial and open-source frameworks to maximize the diversity of candidate models created. The "hardware" component includes the memory, network capacity, and computational power needed to explore and prepare data, create features to enhance pattern recognition of data mining or machine learning techniques, train candidate models, evaluate models, and create ensembles (when appropriate).

2) **Planning for Future Analytical Needs:** Does the analytics team, by itself or in conjunction with the Executive Data Analytics Board, have representation with both

> Information Technology (IT) and Information Security (IS) planning functions at both the tactical (mid-term time frame for projects) and strategic (long-term planning) levels? A well-defined process of receiving software authorization to operate (ATO)[7] is needed for production environments to ensure potential solutions are not encumbered by delays inherent from software and hardware approval processes.

In short, to what extent does the analytics team leverage its available tools and technologies, communicating both current and emerging needs, to ensure efficient and effective data analysis and analytic product development?

6.4.1 Finding

EEOC lacks a centralized analytics team having the ability to communicate directly with the Chief Information Officer (CIO) and Chief Information Security Office (CISO) regarding analytics tools and technologies. Of the two special-purpose analytics teams found, ORIP Program Research Branch and OGC Research Analytic Services, analysts either had more computational power available in their workstations or were able to utilize centralized services that leveraged greater processing power, either in-house or within a cloud environment. In addition, these teams often had access to data mining tools based on SAS, R, and Python.

6.4.2 Recommendation

The EEOC Executive Data Analytics Board should work with a high-ranking executive or Office Director, such as the Chief Data Officer, to establish a centralized analytics team. This team should work with the Chief Data Officer to address resourcing needs and the Analytics Program Management Office to incorporate new tools and technologies into its body of accepted analytics tools and techniques. To address unmet resourcing needs, Chief Data Officer would make a request to the Executive Data Analytics Board to fund new tools and techniques that could have a positive, strategic impact to the entire organization.

This recommendation is intended to implement evaluation and governance processes, not to make specific recommendations related to the use of commercial vs. open-source software or cloud-based vs. on premise servers. The governance process itself should evaluate such decisions based on communication of stakeholder needs and consideration of constraints, including security requirements and resources available for such investments.

[7] A general overview of ATO can be found at: https://www2a.cdc.gov/cdcup/library/pmg/implementation/ato_description.htm

6.4.3 Notes on Recommendation

If the EEOC implements the above recommendation to further increase its analytic capabilities, the EEOC should also consider the following two observations made by the assessment team:

1) **Scope:** While there are two existing groups within the EEOC that do provide some advanced analytics, the broad data governance principles of a Data Governance Board should support, not supplant, these case-specific initiatives: OGC RAS, ORIP Program Research. These case-specific initiatives are best to remain decentralized because of the high-level of domain knowledge required for their specific use cases.

2) **Investment in Data Warehouse:** Given that the EEOC lacks an existing data warehouse, the EEOC would initially need to undertake substantial effort to plan and oversee implementation work for an effective data warehouse. Although the assessment team recognizes this effort may be substantial, it provides a framework through which all stakeholders, including members of the analytics team who will ultimately utilize the data warehouse, can guide the design of an analytics-enabling schema. This approach will pay long-term dividends as a correct design could permanently reduce one of the most time-costly components of analytics projects: getting the data into the appropriate format for analysis.

7.0 PROCESS

7.1 BACKGROUND

In practice, data analytics, which at its heart involves the steps needed to extract potentially useful information and insights from data, is both an art and a science. It is a science in that it involves observations, measurements, testable hypotheses, and is concerned with repeatability of results. It is an art in that there are hundreds of approaches, techniques, and toolkits needing varying level of domain subject matter expertise that must be meshed with data sets that themselves have hundreds, if not thousands, of differing characteristics. Knowledge and experience are frequently the best guides to navigate through the sheer number of possible approaches to generate effective solutions that are both consumable and actionable by stakeholders.

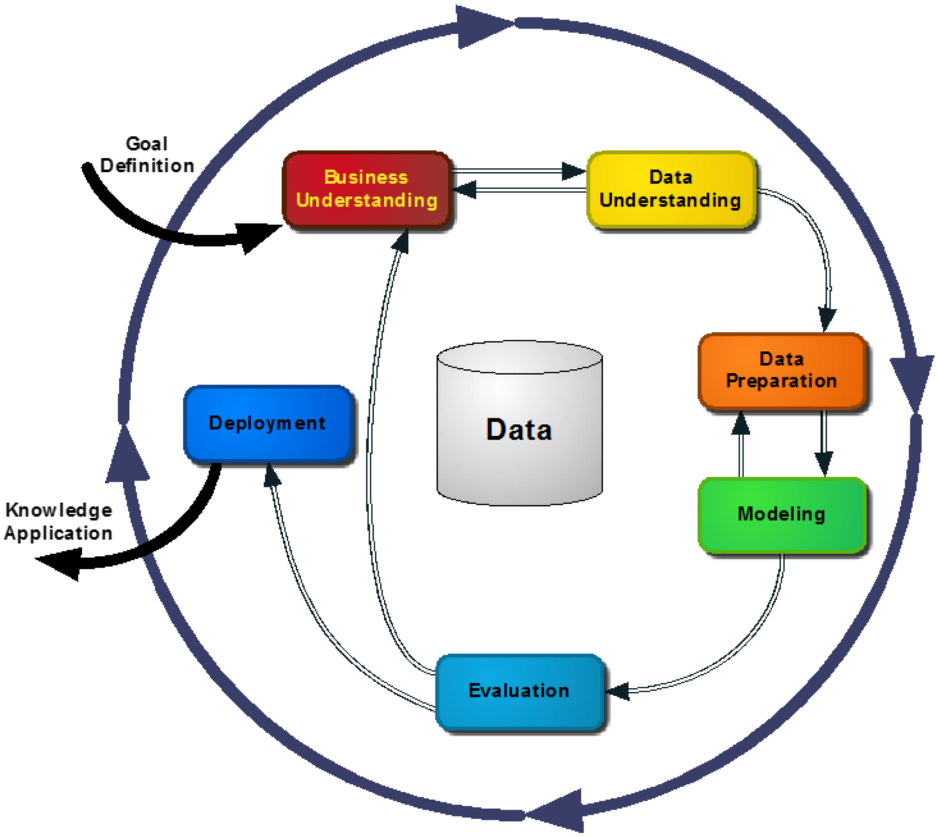
The sheer number of possibilities gives rise for the need for a process framework—key steps that provide basic outline of what needs to be accomplished and the order in which these steps should be executed. Frameworks also provide guidance on when it is appropriate to switch from one step to another, revisit and refining results from previous steps or to move forward in the process. Such guidance enables consistency both within and across organizations and is helpful in demonstrating that all appropriate steps were taken in the development of data-based insights. This consistency is helpful when situations arise when data-based insights are counterintuitive or indicate that there is little to no signal in the data from which to build any insights upon.

7.2 CRISP-DM

Data science endeavors are often project-oriented, meaning there is a defined start and end to each endeavor. CRISP-DM, which stands for "CRoss-Industry Standard Process for Data Mining," is a high-level, extensible process framework that is effective in guiding most types of data science projects. Similar to other iterative development process frameworks, such as the Agile software development process, CRISP-DM requires that key steps in the process be revisited multiple times during a project. This act of revisiting steps allows for information to be considered in context of lessons learned from other project activities, allowing the project team to identify useful information and bring it into better focus before moving onward with the project.

The below graphic shows the six main steps within the CRISP-DM process framework:

Figure 7-1: The CRISP-DM Process Framework



CRISP-DM specifies six steps of a data science project. Although the steps below will be shown in the general order they occur, it is important to note that CRISP-DM is an iterative process, meaning each step should be revisited as many times as needed to refine understanding and results. The six specified steps are:

1) **Business Understanding:** A solid understanding of the problem to solve is a prerequisite for success for most data science projects. First and foremost, this step involves gaining understanding of the problem and explicitly defining “success criteria” in meeting with stakeholders and domain subject matter experts. Stakeholders benefit from being able to set realistic expectations on the scope and magnitude of the benefits they may enjoy from the project. The analytics team benefits by knowing what a “good” model must look like and can therefore devise evaluative metrics to assess models.

2) **Data Understanding:** This step may contain many sub-steps, which may include data acquisition, data integration, data description, and data quality assessment. The key theme is gaining an understanding of the quality and applicability of the data to address the objectives defined in the business understanding phase. This may require

adjustment of expectations, either positively or negatively, with stakeholders depending on the quality and perceived applicability of the provided data. This step should also encompass a review of publicly available data to assess whether external data sources can enhance results. It is highly recommended that the Business Understanding step be revisited one or more times while Data Understanding processes are in progress.

3) **Data Preparation:** This step involves all the processes required to access, transform, and condition available data so that it is in a format suitable for the types of analysis to be performed. This step frequently involves a significant portion of time and resources needed for data science projects. Processes involved in this step include, but are not limited to, data cleansing, data imputation, data transformation, data weighting and/or balancing, data abstraction, feature engineering, and evaluation of feature importance. It is frequently in this step where the "art" of data science becomes most prevalent.

4) **Modeling:** Modeling is the act of creating a representation of an object, system, or business process containing an optimal mix of core features relevant to the desired use cases. Models are typically used for classification or predictive purposes. This step calls for trying multiple types of models that are appropriate for the stated project goals, letting the models compete later in the Evaluation stage. Determination of the "appropriate" model approaches as well as the "optimal" mix of core features must be guided by the requirements set forth in the Business Understanding step, weighing benefits across multiple dimensions that can include, but are not necessarily limited to, simplicity vs. complexity and accuracy vs. interpretability.

5) **Evaluation:** Multiple competing models must be evaluated to determine which model (or ensemble) is "best" in addressing stated business objectives. Evaluation is highly dependent on the definition of success gathered in Business Understanding step. The goal is to create quantitative metrics and evaluate the performance of each model in light of intended usage, which includes items such as the costs of errors (i.e. false positives and false negatives). This evaluation will determine not only which model(s) are best, but also which thresholds (or sensitivity levels) are most appropriate. Once evaluation results are available, communicate the results with stakeholders while revisiting Business Understanding or other previous steps. Revisiting these steps will better establish end-user expectations while communicating assumptions and limitations of the recommended approach. Output of the Evaluation step should also include a business case for future studies of this problem that build upon merits of the present study.

6) **Deployment:** Deployment focuses on how to best make results actionable and easy to understand by the end users of the analytic product. This should be driven by the "success criteria" established in Business Understanding step, answering questions such as how best to display model results (spreadsheet, visualization, interactive dashboard) and educating end users on how to properly interpret the insights. This step must also involve communication to the end users of the assumptions and limitations of the data and techniques employed in building the analytic product.

Note that deployment may require both IT and Information Security involvement: new software may require authorization to operate (ATO) and IT infrastructure may need upgrades to support the optimal end-user delivery mechanisms. As such, it is important the analytics team interfaces with IT and Information Security frequently throughout the life cycle, but especially during deployment, to ensure deployment itself can effectively proceed.

In addition to the CRISP-DM steps, analytics teams, or the appropriate data governance body, should also ensure that deployed solutions benefit from:

1) **Feedback Loops:** No model or analytic product is 100% perfect, nor will a model that is effective today remain equally effective as time passes. It is critical to record the true outcomes of as many recommendations or predictions as possible in order to better incorporate such changes in new or refreshed models to be deployed in the future. This recording of outcomes (such as correct/incorrect) is referred to as a Feedback Loop. When effective feedback loops exist, analytical techniques can be used to determine when models are losing their effectiveness so that action can be taken. Feedback loops serve as essential inputs into a proper model governance process.

2) **Model Governance:** This involves a separate set of processes designed to properly manage models and other analytical products using a lifecycle mindset. This involves regular evaluation of the effectiveness of analytic products, utilizing feedback loops and risk management processes, to ensure continued effectiveness and appropriate usage of such products. Model governance involves establishing criteria to initiate analytics project when analytic projects no longer meet desired effectiveness or when business needs have evolved to the point that previous "definitions of success" have become insufficient.

7.2.1 Finding

EEOC lacks a centralized analytics team and a corresponding data governance group that is necessary to fully implement the CRISP-DM and Model Governance processes as described.

7.2.2 Recommendation

The EEOC Executive Data Analytics Board should support data reporting and analytic projects through the approval of the following initiatives delegated to other governance bodies:

1) **Establish a Centralized Analytics Team:** The Executive Data Analytics Board should work with a high-ranking executive or Office Director, such as the Chief Data Officer, to establish a centralized analytics team. This team can build experience in solving the EEOC's problems utilizing well-defined process frameworks, such as CRISP-DM, and allows for organizational knowledge and experience to be leveraged in the creation of analytical products.
2) **Establish Governance and Support the Centralized Analytics Team:** EEOC can be most effective in model governance and risk management by utilizing a high-level governing body with authority to set evaluation criteria designed to measure the extent in which EEOC's analytical solutions are effective at meeting organizational needs. The Analytics Program Management Office should meet regularly to conduct necessary processes that include, but not necessarily limited to, assessing potential new data sets, analytical product risks, functionality of feedback loops, and analytical product effectiveness.
3) **Conduct Awareness Training of Iterative Processes:** Awareness training of iterative processes, such as CRISP-DM and/or Agile, should be provided to select individuals, starting with members of a Data Governance Committee and members of a centralized analytics team. This training should also be provided for key stakeholders and business process owners at the start of their first project engagement with the centralized analytics team. Such training would help set expectations and preempt stakeholder questions on why items such as "definition of success" are frequently revisited multiple times during the lifetime of an analytics project.
4) **Provide Access to Project Management Tools:** Because analytical product creation is predominately project-oriented, each member of a centralized analytics team should have access to computerized project management tools (as already utilized within OIT) and be versed in usage of such tools.

Note: As this recommendation is meant to facilitate processes designed to address the unmet analytical needs of the EEOC, the OGC RAS and ORIP Program Research Groups, who are already effectively meeting limited-scope needs, are out-of-scope for this recommendation. As long as those specific stakeholder needs continue to be effectively met, those groups may continue to function as currently structured.

8.0 INFRASTRUCTURE

Organizations that embrace analytics devote planning and allocate resources to analytics team(s). Sadly, this planning is often disproportionally focused on software-based tools of the analytics team, ignoring the organization-wide infrastructure needed to support effective reporting and predictive analytics. Only through the identification of specific infrastructure needs can a systematic approach be employed to select appropriate tools and technologies to further enable reporting, data mining, and analytic product delivery.

This analytics assessment evaluates the information technology infrastructure of the EEOC within three specific areas:

Table 8-1: Infrastructure Assessment Areas

Infrastructure Area	Why Important	Summarization of Core Finding
IT Infrastructure and Data Storage	IT departments often must manage competing demands: address compliance requirements, support public-facing services, maintain old/comfortable user interfaces, and enable transformative technologies, all within a shrinking budget footprint. Processing and data storage infrastructure must continually adapt and scale to meet evolving requirements in all of these areas.	EEOC Office of Information Technology is in process of deploying new infrastructure, but in order to effectively prepare for adoption of transformative technologies, OIT should find ways to foster ongoing flexibility and scalability. See Section 8.1 for details.
Data Availability and Transformability	As IT systems evolve over many years, data created at specific points in time are often stored in differing ways on differing systems. Analytics requires mining of vast amounts of historical data in order to identify patterns and provide insights. Analytics teams must be able to readily access and transform historical data into formats that allow for consistent analysis of data across time.	EEOC lacks a data warehouse that enables versioning and consistent storage of cleansed data across time, creating substantial impediments for both reporting and analysis of historical data and trends. See Section 8.2 for details.
Visualization and Delivery	Humans are well suited to find patterns and spot trends in images. In order to leverage this innate ability, IT systems must be able to support complex visualizations of data. Effective analytics is rarely deployed via a spreadsheet: analytic products maximize usability of results via effective visualization utilizing context-dependent delivery mechanisms.	The EEOC's lack of a modern reporting platform that enables report customization and automates the data gathering process, results in reporting largely based on static requirements and requires substantial effort to maintain. This inhibits both reporting variety and frequency, leading to unmet organizational needs. See Section 8.3 for details.

Important note: this analytics assessment does not include an extensive inventory of IT systems within the EEOC. Instead, results are based on interviews with dozens of stakeholders, both within EEOC headquarters and in two EEOC district offices. As such, the findings are designed to capture only high-level observations and recommendations are geared towards creation of enterprise analytic products.

8.1 IT Infrastructure and Data Storage

For purposes of this assessment, IT infrastructure and data storage encompasses the following components as related to implementation of analytics and analytic products:

1) **Processing Power:** This involves the implementation of central processing units (CPUs) and graphics processing units (GPUs) throughout the organization to provide sufficient processing power and display capabilities to load data, transform data, and display/report analysis results. In general, greater amounts of data tend to better leverage modern machine learning and other more sophisticated analysis techniques. As more data becomes available, from public datasets as well as internally-generated data, the need for processing power will continue to increase.
2) **Memory:** This involves the various types of random-access memory (RAM) that CPUs and GPUs rely upon to store and temporarily hold intermediate results. For analytics teams, insufficient RAM can slow performance by several orders of magnitude even when sufficient processing power is available. Similar to processing power, as more data becomes available, memory needs will continue to increase.
3) **Disk Storage:** This involves the various forms of long-term storage of either structured or unstructured data, often utilizing a file system. There are many types of technologies, such as magnetic platters or solid-state devices, included in this category. Disk storage requirements are directly proportional to the amount of data that must be stored for analysis. As more data becomes available, the need for more disk storage will continue to increase.
4) **Network Capacity:** This involves the end-to-end communication pathways between the locations where items are stored and consumed. Network capacity needs are highly dependent on the relative locations of clusters of data storage, clusters of processing power, and where end-users are located. In traditional client-server environments, this frequently involves transferring data from clusters of storage to individual workstations where processing power resides. In a cloud environment, this involves securely transferring results over an Internet connection from scalable, high-density clusters of storage and processing to local end-users or developers. Similar to the other components, as more data becomes available, the network capacity will increase.

The effectiveness of analytics is limited to the amount and quality of the data used as its inputs. In most cases, a commitment to continual improvement in analytics therefore implies a commitment to improving both the quantity and quality of data available for analysis. This, in turn, implies a commitment to an ever-increasing need for more processing power, more memory, greater amounts of disk storage, and higher network capacity. Analytics teams should therefore be able to advise IT departments on emerging needs in each of these areas, and in turn, IT departments that can quickly scale/reallocate resources where needed can provide demonstrable returns on investments in support of analytic endeavors.

8.1.1 Finding

The EEOC Office of Information Technology (OIT) is in the process of deploying new hardware to increase the processing power and memory of a wide swath of end-users. Because multiple end users reported latency and reliability problems with existing applications, this hardware investment was designed primarily to address operational computing needs. Although additional infrastructure will likely be needed for analytics teams, the evaluation team anticipates the hardware upgrades in progress will largely address the initial analytics and reporting needs for most end-users.

In the past, the EEOC OIT has been project-oriented: it rolled out new systems, trained end-users as appropriate, and then moved those systems into ongoing maintenance mode. With respect to hardware, this has resulted in fixed hardware capacity tied to upgrade cycles that are lengthy. This traditional approach has resulted in a lack of flexibility and responsiveness in addressing in infrastructure needs, creating an environment that more frequently inhibits, rather than fosters, utilization of transformative technologies such as analytics and visualization.

The EEOC OIT is aware that broader adoption of data mining and predictive analytics will place increased demands on IT infrastructure, both for the analytics team(s) as well as for analytic product end-users. To this end, the EEOC OIT is actively seeking new ways of operating that can maximize its ability to remain responsive to the infrastructure needs given its operating constraints.

8.1.2 Recommendation

The EEOC OIT should continue with existing plans to upgrade hardware infrastructure to address current needs. However, if resources will remain limited in the foreseeable future, the EEOC Chief Information Officer should work with the EEOC Executive Data Analytics Board to consider the feasibility of adopting technologies that can provide increased responsiveness and flexibility of supporting the infrastructure requirements of improved reporting and analytics initiatives:

1) **Cloud-based data storage and analysis:** Most of the processing power, memory, and disk storage requirements in analytics is used by the analytics team, and not by end users of analytic products. To this end, EEOC analytics teams may be better served utilizing cloud-based or similar services that allow for dynamic allocation of processing power, memory, and disk storage in various forms. This not only allows for better project-based tracking of costs, as such services are often pay-for-use models, but allows for on-the-fly infrastructure updates without significant outlays of monetary or personnel resources. This will help address ongoing needs of the centralized analytics team, allowing them to both evaluate and utilize new infrastructure-related technologies in a cost-effective manner.

2) **Lightweight delivery of analytics products:** Analytics products are often delivered to end-users either via a software-based client or via a web-based interface. Software-based clients, especially if closely integrated with existing software in use, generally allow for greater flexibility and customization of reports. However, web-based interfaces, especially modern dashboards and visualization tools, are not far behind in functionality. Either approach offers substantial improvements to the EEOC's existing reporting/visualization delivery mechanisms. Approaches that emphasize reduced IT infrastructure needs may better serve the EEOC in the long-run.

There are multiple providers who offer solutions to address both of these needs. The intent of this recommendation is to maximize the long-term flexibility to respond to infrastructure needs, not to recommend any particular software package or cloud service provider. Although this recommendation is aimed primarily towards delivery of reporting and predictive analytic products, this approach may also be helpful in other OIT support areas.

8.2 Data Availability and Transformability

As IT systems evolve over many years, data is often stored in many different places and in many different formats. It is quite common for organizations that have been around more than 5-10 years to have disparate data assets in varied formats. These are commonly referred to as "data stovepipes" or "data silos." With data in this state, extra work is required to unlock the insights that may be lurking within such data.

For purposes of this assessment, data availability and transformability refers to the extent in which the organization is able to utilize infrastructure to access legacy data silos and transform the data contained within to a common format that allows for consistent, across-time analysis. The ability to accomplish these tasks has both technical and administrative challenges related to IT infrastructure. From the technical standpoint, legacy data silos may require inefficient allocation of storage space and other system resources to maintain any level of access to the data. Administratively, legacy data silos may require inordinate amount of IT staff resources to maintain, resulting in required maintenance of outdated skillsets within IT staff. This results in increased organizational risk if the expert of a legacy system were to leave or retire.

To combat these issues, many organizations implement what is known as a data warehouse. Put simply, a data warehouse stores data from multiple, disparate systems in a common format in a single place, often automating the extra steps referenced above related to unlock insights from historical data. This provides multiple benefits, which can include:

1) **Reduction in Legacy Systems:** Once legacy system data is transferred to a data warehouse and is verified, there is more flexibility to take legacy systems offline, freeing up system resources and reducing need of IT staff to maintain outdated skillsets.

2) **Better versioning of data:** Even with currently production-use systems, correctly designed data warehouses can provide improved ability to utilize point-in-time snapshots of data. This can greatly simplify the tasks involved with reporting and verification of results, as naturally occurring live system data updates create a moving target that cause data quality and other complex cross-check methods to fail.

3) **Better detection of data changes:** Depending on how the data warehouse is implemented, it can allow for detection of changes of the same data reporting elements over time. This can be useful when looking for data restatements or when there is a need to ensure that data aggregation processes are accurate.

4) **Automation of data cleaning processes:** Most data sources have some quality issues, especially data that is manually entered as well as survey data. Many organizations using such data sources spend a significant amount of time "cleaning" the data to ensure it provides an accurate representation of reality. Instead of handling quality assurance on a quarterly or annual basis, automated data warehousing forces data quality assurance / cleaning steps to be codified and largely automated during a data ingestion process. While this may require significant up-front investment for some data sources, in the long-term, it frees personnel resources from having to repetitiously perform data cleaning.

5) **Freeing System Resources:** With increased usage of enterprise analytics program, there are increased demands on databases. It is important to ensure that analytics teams have minimal impact on production databases where data entry takes place. Data warehouses help accomplish this goal by keeping analysis and reporting system demands off of the live database system.

These benefits do not come without costs—such systems require an investment and often a substantial amount of development time to set up. However, once data warehouses properly ingest the organization's data, the organization will reap long-term benefits associated with the organized maintenance of data assets in a single location. Newly created systems in the future should have data warehouse integration efforts built into initial development and system lifecycle. This serves to reduce the burden of data warehouse integration of new systems over time.

8.2.1 Finding

The EEOC lacks an enterprise data warehouse used as a central location to store cleaned, versioned data from disparate sources. This has led to several organizational challenges:

1) **Consistent reporting:** Because the EEOC lacks data versioning often provided with an effective data warehouse, responding to public information requests entails unnecessary challenges in finding consistent snapshots of data to report from. For example, the EEOC recently had to involve multiple offices for help in deciphering

internal data to address media information requests regarding workplace harassment. This example demonstrated challenges that result in reporting inconsistencies for same time period across different report compilation times. The versioning of data in a correctly designed data warehouse offers an effective way to address many of these challenges.

2) **Persistence of verified data:** The EEOC Office of Research Information and Planning (ORIP) has a team of dedicated professionals who perform data quality control to produce cleansed, verified datasets. In the case of survey data, the effort expended by this group is substantial. However, because the EEOC lacks a good place to store and retain these cleaned datasets, only the most recent versions of these cleaned datasets remain available for use and then are purged. The purging of this cleansed data inhibits the ability of the EEOC to analyze many years' worth of data to identify trends. A data warehouse would provide a natural location for such cleansed data to persist over time.

3) **Capturing knowledge to verify data:** The same team of dedicated professionals who perform data quality control currently perform their work repetitiously and manually. Although code is available to leverage knowledge gained, there has not been a systematic effort to automate large portions of this process. While it may be true that data quality control of survey data may never be able to be fully automated, the pervasive lack of automation in current processes has led to delays in data availability as well as increased risks to the organization if experts within this team were to leave or retire from the EEOC. A properly designed data warehouse utilizing extract-transform-load (ETL) processing would help codify this organizational knowledge, potentially increasing efficiencies and timeliness of the data quality control team.

8.2.2 Recommendation

The EEOC Executive Data Analytics Board should work with EEOC Chief Information Officer to investigate investments in a data warehouse to address its long-term data storage, versioning, and analysis needs. While this recommendation likely requires the most substantial investment of all the recommendations in this assessment, it also offers the greatest long-term benefit to the organization as a whole. Without such an investment, the EEOC will continue to encounter challenges and inefficiencies related to reporting and long-term trend analysis.

A data warehouse initiative should take into account the needs of multiple stakeholders. To the extent possible, the development of data warehouse architecture and associated ETL processes should include oversight from multiple stakeholders, including a centralized analytics team, who can benefit from appropriately optimized design.

Please note that this assessment did not entail a deep dive into the different types of data warehouses and the extent each could address EEOC needs. As such, this recommendation is not prescriptive on the type of data warehouse and where data should be persisted (locally or in a cloud environment). The Executive Data Analytics Board should work closely with EEOC OIT to evaluate options and determine how to implement an infrastructure that addresses the current, cumbersome state of reporting and analysis with investments that lead to a new state that facilitates these activities.

8.3 Visualization and Delivery

Data visualization and analytic product delivery are critical to the success of any analytics program. Without these components, end-users and decision-makers are not empowered to adequately digest insights and integrate them into their workflows or decisions. Effective visualization and analytic product delivery should engage users' most pressing questions, enabling them to develop new questions and engender desire for more.

While both visualization and analytic product delivery are important, each has a distinct focus:

1) **Visualization:** Human beings are well-suited to intuitively analyze complex relationships in large amounts data when shown in an effective, visual fashion. Visualizations leverage this innate ability while creating an avalanche of numbers in a spreadsheet or report inhibits this ability. To be effective, visualizations must be accurate, readable, interactive, customizable, and accessible[8].
2) **Analytic product delivery:** Analytics, by its very nature, deals with large amounts of data and employs complex techniques to create models that are reflective of reality. Analytic product delivery involves integrating new data points, allowing the user to see model results within the context of the business problem and be able to understand the salient characteristics that contributed to those results.

Visualizations can be a component of analytic product delivery, but effective analytic products do not stop there. Data mining and predictive analytics seeks patterns within the data, and thereby identifies certain portions and characteristics of the data that are more pertinent than others in leading to model results. Analytic product delivery must integrate new data with the aim of making models "alive" to users by providing insight on the "drivers" that led to particular results. This can be quite important to many end users of analytic products because those users may have no other means to access or understand the

[8] https://www.elderresearch.com/company/blog/5-keys-to-powerful-data-visualizations

underlying data, especially in a manner that considers the context of the problem they seek to address.

A well-known example of how these two concepts support each other comes from credit scores. In this use case, analytic product delivery would be focused on integrating all of a person's data into a credit-worthiness model with results codified into a numerical credit score. That score may additionally be separated into different components, providing insights into the drivers behind the overall credit score. Visualization further enhances communication of insights by allowing for comparisons between entities, comparisons across time, and so forth. Well-designed visualization and delivery mechanisms should be able to proactively address the most common end-user questions about the provided results.

8.3.1 Finding

The EEOC's existing systems are primarily focused on delivering raw data or basic, tabular reports. Requirements for these reports, such as the "396 Report", are largely static, often exhibiting only slow evolution over the years. As a result, most reports focus on data aggregation and simple statistical reporting, with little to no delivery of predictive analytic products.

Consequently, many reports within the EEOC, including the "396 Report", are delivered in either a PDF or a spreadsheet format that lacks the ability to support customized visualizations. Despite the prevalence of static report formats, some reports require substantial amount of manual effort to create. Even though these reports could provide management insight on the effectiveness of new initiatives, the burden of producing the reports inhibits high frequency report generation that would be needed to make frequent assessments.

In short, because of the lack of a modern platform that enables report customization and automates the data gathering process, reporting within the EEOC is largely based on static requirements, often requiring substantial effort that inhibits reporting variety and frequency needed to effectively address organizational needs.

8.3.2 Recommendation

The EEOC Executive Data Analytics Board should work with the Chief Information Officer to investigate investments in modern report delivery tools that:

1) **Automate the reporting process:** Reporting and data analytics should not be a "chore" whose creation consumes time and resources from mission-oriented activities. Rather, reporting process should be automated as much as possible, incorporating the latest data points. This allows end users to understand the evolving effectiveness of recent initiatives and decisions. Appropriately customized modern reporting tools can help accomplish this.

2) **Customize the reports:** The environment in which any organization operates will evolve over time, and this evolution impacts the pertinence of old questions and gives rise to new questions. A reporting and analytic product delivery tool should reduce the burden of updating reports, possibly through drill-down lists or interactive dashboards. This allows reports to be more reflective of the evolving environment in which the EEOC operates within, allowing people to ask a greater quantity of questions that are pertinent to mission-oriented activities.

3) **Provide effective visualizations:** Reporting of results can be made more concise and more effective by leveraging the innate human ability to visually identify complex relationships. Optimally, visualizations themselves can be customized to be more reflective of the emerging environment in which the EEOC operates.

4) **Leverage data warehouses:** While most reporting is focused on the present time period, it can be instructive to provide historical comparisons. A modern reporting platform should utilize the versioned, quality-controlled data that resides within a data warehouse to foster these types of comparisons.

Effective reporting that accurately reflects the organization's current status is a necessary prerequisite to more advanced stages of analytics. If users are not engaged and actively asking questions about the present, they are less likely to understand or care about predictive analytic results related to the future. Hence, this recommendation focuses first and foremost on the effective application of modern reporting tools through which the centralized analytics team can leverage in delivering analytic results to users.

Longer-term, the centralized analytics team must engage end-users of analytic products to determine which delivery mechanisms foster clear, actionable results. At first, determinations should be project-specific, occurring within the Business Understanding phase of the project (see Section 7). After a sufficient number of projects have been successfully completed and evaluated retrospectively, constituent patterns of effective delivery mechanisms should emerge and can become part of the default framework from which future analytic product can be customized. This recommendation is therefore not prescriptive on what this may look like—it should be part of an integrated analytic product delivery process and be based on data collected from end users during the planning phase of each project.

9.0 APPENDICES

This section of the report summarizes in a tabular format the specific recommendations and other insights gained by the assessment team during this engagement:

1) **Appendix 1:** Contains a table of all unique recommendations, providing a suggested order for implementation. This unique order addresses the redundancy of recommendations that were one-to-one mapped with findings within the body of this report.
2) **Appendix 2:** Contains a listing of lower-resource, actionable project examples customized for the EEOC, with the goal of demonstrating how improvements in reporting and data analytics can be integrated into workflows and improve efficiency and/or effectiveness of meeting the organization's mission.
3) **Appendix 3:** Contains the EEOC's response to the draft report as well as Elder Research comments.

Please note: The table containing the summary of recommendations in Appendix 1 uses abbreviations to denote responsible parties. The list of responsible parties, and associated abbreviations, are:

- AC: Analytics Champion
- APMO: Analytics Program Management office
- CDO: Chief Data Officer
- CIO: Chief Information Officer
- DGC: Data Governance Committee
- EDAB: Executive Data Analytics Board
- OCH: Office of the Chair

9.1 Appendix 1: Table of All Recommendations

Table 9-1: Summary of all Recommendations

Phase	Report Section	Section Description	Responsible Party	Brief Overview
1	4.1	Shared Vision for Analytics	EEOC OCH, EEOC EDAB	Establish data analytics governance infrastructure.
1	4.4	Collaborate Environment	EEOC OCH	Engender trust in enterprise-wide steering committees and governance boards.
2	4.2	Executive Leadership	EEOC OCH, EEOC EDAB, EEOC AC	Establish tone advocating for analytics in strategic planning and reviewing recommendations of data analytics governance bodies.
2	8.1	IT Infrastructure and Data Storage	EEOC CIO, EEOC EDAB	Consider new approaches, such as web-enabled and cloud-based solutions, to support expanding IT infrastructure needs of both the analytics team as well as analytical product users.
3	5.1, 5.2, 5.3, 6.2, 6.4	Understanding Business Needs, Technological Breadth	EEOC EDAB, EEOC CDO	Establish a centralized, enterprise-wide analytics team or Analytics Center of Excellence.
3	4.5	Continued Education and Learning	EEOC OCH, EEOC AC	Designate an analytics champion to foster and evaluate cultural awareness of analytics.
3	8.3	Visualization and Delivery	EEOC CIO, EEOC EDAB	Invest in modern reporting and visualization tools that allow for automated, customizable, visualization-enhanced reporting that effectively leverage a data warehouse.
3	8.2	Data Availability and Transformability	EEOC EDAB, EEOC CIO	Establish a data warehouse to address data retention, versioning, and reporting needs.
4	7.2	Process	EEOC EDAB, EEOC APMO, EEOC CDO	Support analytics projects through governance of the Analytics Center of Excellence, promoting awareness of iterative analytical project processes and usage of Agile-friendly project management tools.
4	4.3	Culture of Evaluation and Improvement	EASC EDAB	Invest in the generation of new metrics that quantify opportunity costs and corresponding benefits of data collection and data assurance.
4	6.3	Modeling Process, Evaluation, and Management	EEOC APMO	Adopt proven modeling approaches and model management techniques.

9.2 Appendix 2: Example Projects to Jumpstart Analytics

This analytics assessment aims primarily to provide high-level objectives and associated recommendations to achieve those objectives. However, the assessment team realizes that in the real world, a long list of do-items may seem intimidating and insurmountable. The below table provides specific examples of lower-cost projects that the EEOC can implement to generate some "quick wins" in data analytics to demonstrate how analytics can enable the EEOC to more efficiently accomplish its mission.

Please note that the entries in the below table are not listed in any particular order. With that said, the evaluation team believes additional ability to deliver analytic product results to those actively using the Integrated Mission System platform (IMS) for intake or charge/case management holds the greatest potential for immediate improvements in efficiency for a large population of users. These efficiency improvements can be quantified and serve as a baseline for return on investment for other potential analytics projects of similar scope.

Table 9-2: Example Analytics and Reporting Projects

Potential Target Area	Brief Description	Why Important, Evaluation Metric
IMS Company Name	Company names entered into IMS are not necessarily reflective of the entire company's history of the EEOC, due to misspellings and other various ways a name can be entered (doing business as, with or without "Inc.", etc.). Entity resolution can leverage existing data to connect different entries to address this issue, providing a list of potential company names with a likelihood score denoting the likelihood the returned result is the entity the user seeks.	Almost all EEOC processes of allegations, charges, and cases depend on the quality of data from the intake process. Making the intake process more efficient and more accurate holds potential to save time at all levels of processing. Evaluate by number of new insights uncovered (i.e. multiple names representing same corporate entity) as well as reduction in time of intake process itself.

Potential Target Area	Brief Description	Why Important, Evaluation Metric
Federal Data	For federal sector data, automate methods to prepare data for analysis.	Because of the level of detail and completeness of data related to employment in federal agencies, this data set is already sufficiently rich to learn many insights and trends in the federal sector. Many of these are already reported on, but effort required to do access and prepare the data is significant. Automation of any portions of data access and preparation would save time and enable more advanced analytics with this already rich dataset. Measure effectiveness by number of hours saved and number of new analytical initiatives that can be launched on these rich datasets.
IMS Text Search, Text Analytics	Stage 1: Many charges can be best identified by certain key terms, but the system lacks a way to search free-form text fields, either within a specific charge or system-wide. Stage 2 (a later project, requiring a data warehouse): Text analytics can be applied to automate detection of certain characteristics of each charge, allowing for analysis of the salient textual characteristics of charges that are either settled or adjudicated in the charging party's favor.	Any additional capabilities in text-based searching and/or analysis unlocks the value that currently resides in unstructured text. Simple search unlocks the ability of users to manually find currently relationships between charges/cases. Evaluate by number of new insights uncovered as well as reduction in time of intake process itself. Full text analytics is a larger project that simple text search would serve as a starting point. Such a larger project would reduce the burden on intake staff, investigators, and others who currently have to check certain boxes to associate records with certain types of cases. Evaluate by number of new insights uncovered as well as reduction in time of intake process itself.

Potential Target Area	Brief Description	Why Important, Evaluation Metric
Data Collection Improvements	The EEOC has various methods of collecting data on outside entities, including the EEO surveys and the various online portals where individuals and companies respond to information requests. Stage 1 (low effort): A project to solicit comments from end-users, intake personnel, and investigators about the efficacy of portal design to provide data to guide improvements to portal design or password dissemination methods. Stage 2 (large effort): A project to explore which portions of private industry respondents utilize cloud-based HR services. EEOC can then gather knowledge about available fields and design specific schema to automate the import process of provided data.	Efficient and effective usage of these mechanisms is critical to ensure EEOC receives data in electronic format that can be analyzed. Current systems are properly aligned with this goal, but anecdotal evidence suggests effectiveness is encumbered by implementation issues. Evaluate effectiveness by the reduction in the number of support requests for portals (stage 1) and the hours of time saved in importing structured data associated with RFI responses (stage 2).
Reporting	Current District and Field Office Reporting, such as the "396 Report" is important to measure and assess effectiveness of operations and progress towards established goals. However, the report takes significant time and effort to run and detracts from operational efforts of those running them, which can include attorneys and members of management.	Evaluating and measuring the burden of specific components of the reporting process can allow targeted investments to be made to automate processes designed to reduce reporting burden. Measure effectiveness via two metrics: 1) Number of personnel hours saved by reducing this burden, either through automation or via use of interactive dashboards with drill-down capability. 2) Quantify value of more immediate feedback to district management who lack resources to run reports on a more frequent basis to assess progress of new initiatives.
Systematic Data Imputation and Augmentation	Create a framework in which categorical data can be imputed or augmented by publicly-available data sets. Example: imputing race of an individual when only name, age, and address are known.	Currently, there are places within the EEOC's processes where data is needed and is either unavailable or not collected at all. This results in personnel having to create case-specific ways to impute data to provide the EEOC the ability to answer questions or test hypotheses. Measure effectiveness by the number of hours saved by automated methods.

Potential Target Area	Brief Description	Why Important, Evaluation Metric
Emerging Trend Identification	As an ongoing project that can start small and grow over time, the EEOC can utilize publicly-available social media data, such as message feeds and other posts, to look for messages that appear to be related to equal employment opportunity risks. This would involve a mixture of text analytics, trend analysis, and change detection.	Congress expects the EEOC to have a proactive stance in identifying emerging threats to equal opportunity. While the EEOC would not take specific actions on insights learned, it can use the insights learned to predict emerging risks. This allows for customization of training programs, addition of new EEO survey questions, recommended legislation, and proactive reporting to Congress on emerging issues. Measure effectiveness based on feedback of stakeholders who receive customized products.
Mine User Behavior and Support Requests to Guide System Improvements	Stage 1: Apply text-mining techniques to semi-structured support requests for EEOC applications to determine areas/processes that seem to present the biggest issues to users. Stage 2: Perform user-behavior case studies during pilots of new or redesigned internal application to generate quantifiable metrics of areas users run into trouble. Stage 3: Embed new production EEOC applications (such as new versions of IMS) with buttons and other features to collect user behavior data and solicit user feedback, with the goal of creating metrics to guide specific updates.	To demonstrate willingness to become a data-driven organization, EEOC should monitor and evaluate effectiveness of its in-house applications for the intended users. This data should then be used to quantify costs of current problem areas, better enabling a return-on-investment approach to updates to internal applications. Because these applications can have hundreds of users, even incremental increases in efficiency of a few percentage points can have a significant overall impact. Evaluate effectiveness by number of hours (or dollars) saved by addressing the most prevalent problem areas.

9.3 APPENDIX 3: EEOC RESPONSE AND ELDER RESEARCH COMMENTS

Section 9.3.1 of this appendix contains the agency's response (a memo and table) along with some additional comments from Elder Research. With the exception of Elder Research's comments located within an additional column inserted within the table, all content contained within Section 9.3.1 is the EEOC's response to the draft report.

Elder Research believes the actions outlined within the agency's response memo, including the creation of the Chief Data Officer position late in 2017, are consistent with the recommendations within the evaluation report. Please note that actions undertaken by the EEOC after fieldwork ended on 27 February 2018 were not reflected in the main report. As such, Section 9.3.1 serves to document the EEOC's responses to this report as well as detail the agency's actions that commenced after conclusion of fieldwork.

9.3.1 Agency Response and Elder Research Comments

U.S. EQUAL EMPLOYMENT OPPORTUNITY COMMISSION
Washington, D.C. 20507

Office of the Chair

June 22, 2018

MEMORANDUM

To: Milton A. Mayo, Jr.
Inspector General

From: Victoria A. Lipnic
Acting Chair

Subject: Comments and Responses to Elder Research "Evaluation of the EEOC's Data Analytics Activities" Draft Report
OIG Report Number 2017-02-EOIG

Thank you for the opportunity to provide comments to the above captioned report. This is a timely review as the agency is already in multiple stages of investments in personnel and in significant hardware and software upgrades crucial to effective data governance and the growing field of analytics. The Recommendations provide important touch points as we build the organizational infrastructure to enable leaders and front line staff to develop long term goals and vision for the formation and evolution of an effective analytics program.

In anticipation of the need for culture change and to move the agency toward embracing data driven decision making and the use of data analytics to enhance mission effectiveness, I initiated a number of actions to start the process toward better data governance and analytics:

- November 2017 – hired EEOC's first Chief Data Officer
- November 2017 – initiated the reorganization of the Office of Research, Information, and Planning (ORIP) into the Office of Enterprise Data and Analytics (OEDA)
- February 2018 – released EEOC Strategic Plan for Fiscal Years 2018-2022 committing to the expanded use of data and technology to support, evaluate, and improve the Agency's programs and processes (Strategy III.B.2).

- April 2018 – chartered the Data Governance Board (DGB) therein creating the executive leadership team and infrastructure to address the items in the Summary of all Recommendations at Table 9-1, Appendix 1 of the report.
- May 2018 – approved the official reorganization of ORIP into OEDA, after formal internal agency review process, and announced such to the agency June 2018.

It is with this background we are beginning to vision the many ways an analytics program can help the EEOC better achieve its mission. I expect the DGB and ultimately the entire leadership team will work to address each of the identified focus areas, findings and recommendations in the report as the agency builds an enterprise wide analytics program with appropriate products for our various internal users. It is anticipated the scope and type of products will include key items such as dashboards and predictive tools to better manage workloads and resources.

One final note as to process with the contractor. On March 29, 2018, the OIG invited leadership and staff from the various incumbent EEOC offices to a briefing about the draft report. I attended that briefing, along with, among others, the Chief Operating Officer (COO), the Deputy COO, and the agency's new Chief Data Officer (CDO). It was pointed out to the Elder Research staff at that meeting that many of the recommendations they were making were already well-underway or happening in real-time at the EEOC. It is surprising then that the written draft reviewed here (dated May 24, 2018) reflects no change from that briefing, not even an acknowledgement that the EEOC had already created and installed its first Chief Data Officer. Notwithstanding that oversight, it was a useful and valuable briefing and I appreciate having that opportunity to be briefed by the contractor at that draft stage afforded to the agency.

The Appendix 1: Table of All Recommendations offers a number of sound suggestions which the agency can use as it charts a path forward. Some suggestions have already been adopted and are being implemented. Several action items had to be adapted to the personnel and infrastructure limitations of a small agency where leaders and staff often fill multiple roles in pursuit of organizational excellence. The table of recommendations is annotated below to reflect actions already taken and anticipated as the Office of Enterprise Data and Analytics is staffed and comes online. The Data Governance Board includes Directors or Deputy Directors from each headquarters program and administrative office – inclusive of the Office of Inspector General. The Board will also include a Senior Executive Level representative from the field. The core responsibilities of the Data Governance Board include ongoing development and oversight of the agency enterprise data management strategies and practices, collaboration with the Information Technology Investment Review Board (ITIRB) which sets funding priorities for data and IT resources, assessing the analytic and reporting needs of the agency and insuring future data analytics investments align with the agency mission and strategic objectives.

Appendix 1: Table of all Recommendations *(Imported from the draft report and edited. Responsive comments are in red for contrast.)*

Phase	Report Section	Section Description	Responsible Party	Brief Overview	Elder Research Comments (in blue)
1	4.1	Shared Vision for Analytics	EEOC OCH, EEOC EDAB	Establish data analytics governance infrastructure. The EEOC Data Governance Charter was signed April 19, 2018 creating the Data Governance Board (DGB). The appointment of a Chief Data Officer and organization of OEDA was shared at the earlier briefing. The CIO has been a champion of this initiative and the promise of enhanced effectiveness through the power of analytics has been a consistent message from the Acting Chair to the Board.	
1	4.4	Collaborate Environment	EEOC OCH	Engender trust in enterprise-wide steering committees and governance boards. The DGB is enterprise wide inclusive of all organizational components. The Board is empowered to sponsor or create steering committees, boards or other working groups as needed in support of its mission.	
2	4.2	Executive Leadership	EEOC OCH, EEOC EDAB, EEOC AC	Establish tone advocating for analytics in strategic planning and reviewing recommendations of data analytics governance bodies. The Acting Chair memo to all staff announcing creation of OEDA on June 12, 2018 emphasized the commitment to "… develop an enterprise-wide data analytics strategy which not only supports the mission of the EEOC, but also makes our data readily available and easily accessible to those within the agency, as well as the public." As noted earlier, this has been an ongoing concern and priority in the Strategic Plan and across agency leadership.	The creation of the Office of Enterprise Data Analytics (OEDA) gives the EEOC an organizational structure capable of advocating for an organization-wide approach to data and analytics.

Phase	Report Section	Section Description	Responsible Party	Brief Overview	Elder Research Comments (in blue)
2	8.1	IT Infrastructure and Data Storage	EEOC CIO, EEOC EDAB	Consider new approaches, such as web-enabled and cloud-based solutions, to support expanding IT infrastructure needs of both the analytics team as well as analytical product users. Starting in FY'2016 the agency began acquisition and deployment of an advanced analytics application in a government community cloud. Almost all new infrastructure services since that time have been upgraded or developed in a government community cloud largely replacing the local server infrastructure. The agency is currently testing a cloud based enterprise analytics toolset. Managers and staff have been united in a call for dashboards for workload management and visualization tools to enhance data analysis and efficiency in decision-making.	As the EEOC considers new approaches, including web-enabled and cloud-based solutions, the new EEOC Data Governance Board (DGB) can advise on infrastructure capable of meeting the reporting and dashboard needs to end users throughout the agency.
3	5.1, 5.2, 5.3, 6.2, 6.4	Understanding Business Needs, Technological Breadth	EEOC EDAB, EEOC CDO	Establish a centralized, enterprise- wide Analytics Center of Excellence. The DGB led by the CDO with collaboration from the CIO will be the Agency Analytics Center of Excellence. The structure will provide ample opportunity for input and collaboration by analytics staff currently staffed in OGC, OFO and OEDA.	Upon clarification from the agency, the proposed structure meets the intent of the recommendation to establish an enterprise-wide analytics team.
3	4.5	Continued Education and Learning	EEOC OCH, EEOC AC	Designate an analytics champion to foster and evaluate cultural awareness of analytics. Analytics Champion will be a shared responsibility among leaders on the DGB.	Upon clarification from the agency, the EEOC's plans to foster and evolve one or more persons within the agency to serve as analytics champion meets the intent of this recommendation.

Phase	Report Section	Section Description	Responsible Party	Brief Overview	Elder Research Comments (in blue)
3	8.3	Visualization and Delivery	EEOC CIO, EEOC EDAB	Invest in modern reporting and visualization tools that allow for automated, customizable, visualization-enhanced reporting that effectively leverage a data warehouse. Although the agency is at the early stages of development, substantial resource investments have already been made. As noted above, the agency is testing a cloud based enterprise analytics toolset. This holds the promise of significantly expanding the use of dashboards and access to real time operational information across the agency. With the introduction of data warehousing, the agency expects both delivery and visualization capabilities to be significantly improved for all users.	
3	8.2	Data Availability and Transformability	EEOC EDAB, EEOC CIO	Establish a data warehouse to address data retention, versioning, and reporting needs. Data collection, warehousing, versioning and access for reporting is a core responsibility of the DGB. The Charter anticipates this group will plan and provide oversight for all phases of the data life cycle from creation to destruction.	
4	7.2	Process	EEOC EDAB, EEOC APMO, EEOC CDO	Support analytics projects through governance of the Analytics Center of Excellence, promoting awareness of iterative analytical project processes and usage of Agile-friendly project management tools. It is agreed that analytics projects should be closely planned and monitored. The DGB or its subset are the appropriate holder of this responsibility.	

Phase	Report Section	Section Description	Responsible Party	Brief Overview	Elder Research Comments (in blue)
4	4.3	Culture of Evaluation and Improvement	EASC EDAB	Invest in the generation of new metrics that quantify opportunity costs and corresponding benefits of data collection and data assurance. This enterprise is a relatively new journey for the agency. Agency leaders agree there is a great deal of work remaining to raise everyone's understanding of the value to be gained from this process in comparison to the effort expended to capture and manage quality data. The agency has already identified numerous reports and data analysis functions that should be automated. We will continue to invest human and capital resources, gather and evaluate feedback and make appropriate adjustments to assure we are building the best model possible for data governance and analytics.	The recommendations related to the Culture of Evaluation and Improvement are designed to help the EEOC identify areas within its own processes that result in inefficiencies. The EEOC Data Governance Board (DGB) is structured to sponsor projects that involve the collection of such data that will enable analyses to prioritize and guide subsequent treatments to address inefficiencies.
4	6.3	Modeling Process, Evaluation, and Management	EEOC APMO	Adopt proven modeling approaches and model management techniques. The DGB will seek industry Best Practices to guide all phases of the data governance and analytics process.	The EEOC may find value in encouraging participation of its analysts in analytics training, conferences, and inter-agency government working groups to ensure its teams remain abreast of potentially valuable approaches and techniques.

Two of our key organizational components that will have significant oversight and input throughout the process largely support the recommendations in the report. OEDA and OIT will be charged with significant oversight in the planning and execution of the agency analytics program. As noted in the chart comments, they have already begun work to help the agency acquire and deploy some of the tools we will need for the program. They will continue work on several of the projects identified at Appendix 2, Table 9-2, to generate "wins" for the program and build agency wide enthusiasm and support for the added value it can bring all agency users and our customers.

The recommendations provide useful insights that we will utilize to map a course forward toward a fully integrated analytics program.

Thank you for the opportunity to review the draft report.