CITY OF DETROIT
DEPARTMENT OF INNOVATION AND TECHNOLOGY

COLEMAN A. YOUNG MUNICIPAL CENTER
2 WOODWARD AVE., SUITE 1212
DETROIT, MICHIGAN 48226
PHONE: 313-224-2900
WWW.DETROITMI.GOV

MEMORANDUM

TO: **Hon. Scott Benson, City Council District 3 THRU: Hon. Council President Mary Sheffield**

FROM: **Art Thompson, Chief Information Officer & Kat Hartman, Chief Data Officer**

DATE: **August 15th, 2024**

RE: **Data Science Utilization in Municipal Operations**

Thank you for your interest in how Data Strategy & Analytics Team (DS&A) is partnering with departments to improve City services through data science. DS&A was established in April 2020 to lead data governance efforts and enable advanced analytics. We treat this work as a holistic strategy encompassing both technical and process solutions, people and hardware. The arrival at advanced analytics does not ensure their utility. We believe in order to fully utilize our data assets as a jurisdiction, we must build real institutional knowledge and collaborative processes that cross divisions and departments.

Data is the new electricity: Residents and departmental colleagues require data to inform City operations and decisions on a daily and sometimes hourly or minute basis. Data is a utility, and the hardware required is infrastructure. ARPA funds allowed DS&A to grow capacity and build an enterprise data warehouse (EDW—more detail in response to question 4), deferring costs on this essential infrastructure for a number of years. Centralizing clean and standardized data in a well-structured warehouse is a necessary preliminary step for making data science possible at a meaningful scale and cadence.

Please find answers to your questions below:

1. **Current Utilization of Data Science:**
   a.) Is the City of Detroit actively using data science to deploy municipal resources and address citywide problems? If so, please provide an overview of how data science is being integrated into our operations.

   We want to highlight some key data science and analytics projects over 4 areas: Complex citywide problems, tracking ARPA outcomes, interdepartmental projects requests, and publicly available tools. Firstly, we have provided improvements to data addressing complex citywide problems that lack clear departmental ownership:
   - For Census purposes, we developed multiple algorithms that improved identification of properties that have likely gone from uninhabitable to inhabitable, allowing us to challenge the base housing unit count the Census uses to determine official population estimates These algorithms greatly reduced the time spent gathering evidence through manual review of street view imagery for Census Challenge submissions. **This work led directly to the Census Bureau acknowledging Detroit's first population growth since 1957.** This will result in increased federal dollars, estimated in the Millions, flowing into Detroit over the next decade.
   - We recently received an NSF Grant with MIDAS (University of Michigan's center of Data Science excellence) that we will use to explore ways artificial intelligence (AI) can **safely** improve delivery of City services.

Secondly, we provide the Mayor and OCFO with quarterly and an annual report of ARPA projects that includes research briefs, performance indicators, and outcomes/outputs. This is used both to evaluate particular programs as well as for managing the overall success of the entire ARPA opportunity.

- Network analysis, a set of algorithms that analyze relationships between things, was able to show that **849 unique households have participated in more than one ARPA program**, indicating which ARPA programs best promote wraparound services. We found that Renew Detroit has the most multiprogram households, and that Summer Youth Employment Program and Detroit Housing Services were key programs facilitating multiprogram household participation. This allows us to identify programs that may warrant more funding due to their effectiveness at providing a holistic approach to mitigating structural inequities.
- Network analysis was also able to show that **42.5% of ARPA projects involve interdepartmental collaborations**. This shows movement toward a less siloed approach to municipal service delivery. It also stresses the long term need to invest in enterprise data infrastructure that centralizes and standardizes City data.
- Additionally, basic statistical analysis of the effect of the Downpayment Assistance Program (DPA) on mortgage denials using Home Mortgage Disclosure Act data found that, **at the City level, DPA decreased mortgage denials by 13.8%**.

Thirdly, we have also generated interdepartmental analytics to facilitate coordinated responses by departments addressing connected issues or cross functional services:

- Climate Vulnerability Index: This data science tool is specific to Detroit's context and is being used by the Office of Sustainability and other departments for targeted actions to address climate change's impacts. This index incorporates multiple risk variables, weighting accordingly, into a more digestible decision-making tool.
- Statistical analysis of USPS deliverability data: Used to provide insight into population estimates and occupancy patterns for multiple departments to make strategic decisions about where to invest resources. Detroit housing patterns and complex and very specific to our locale. Many regional and national estimates have high margins of error at neighborhood and block level.
- Neighborhood Economic Development (NED) dashboard: Allows departments to analyze how developers are behaving in the City across their different holding companies and with consideration to geography, parcel, and building characteristics. This tool makes it easier for departments to see through the fog of holding companies and LLCs to property owners and how they are acting in the City.
- Community Risk Reduction Tool: At the request of DFD, we created a tool that uses building and neighborhood data to identify areas most at risk of future fires. As a public tool it also provides residents with information to better understand their own risks.
- Scooter Dashboard: Provides summary statistics and location data about scooter use throughout the City. This is used by OMI to understand how this new transportation technology is being used. It also allows them to track violations of the use agreements of the approved scooter companies. These massive datasets require sophisticated data science techniques to make them manageable and meaningful for decision-making.
- Operational dashboards: Using software such as Tableau and Apache Superset, we have provided a variety of data solutions to departments to help track and visualize metrics and summary statistics related to their operational goals.
- Data profiles: When preparing to ingest new datasets into the warehouse, we often create a visual of summary statistics that gives the department insight into its structure and any potential integrity issues.

Additionally, we manage the Open Data Portal. Through this platform we provide multiple public facing data science tools and analytics resources.

COLEMAN A. YOUNG MUNICIPAL CENTER
2 WOODWARD AVE., SUITE 1212
DETROIT, MICHIGAN 48226
PHONE: 313-224-2900
WWW.DETROITMI.GOV

- Analytics Portal: Launched this past year, this subportal makes open datasets more accessible to the public by visualizing them and providing basic summary statistics. In addition to make open data more accessible and our work more transparent, it also facilitates non-governmental data innovations.
- Crime Viewer: Developed with DPD, this tool utilizes an algorithm to anonymize victim data to provide residents with important data about crimes in their communities. Data science is a key tool for balancing competing governmental priorities: protecting privacy and promoting transparency.
- Land Value Tax Estimator: This tool allowed residents to look up their addresses and see how the proposed land value tax model would impact them. Our earlier work developing the Base Units (mentioned above) was critical for the success of this tool, as was our ability to develop and support complex data analytics on the backend to provide accurate estimations.
- Base Units: Additionally, every open dataset with a location component is standardized using the Base Unit database. We attach additional columns of location information allowing the public to more easily join these disparate datasets together for analysis.

2. **Team Composition and Compensation:**

a.) How many people are currently on the data science team?

6 FTE and 24 ARPA-funded TASS. There are 3 major projects they support:

- ARPA Project Data Management & Reporting (part of admin support for the entire ARPA portfolio),
- Digital Inclusion Data Operation,
- Enterprise Data Warehouse.

b.) What are the individual staff titles?

Led by our Chief Data Officer, we have the following titles: Deputy Director of Research & Outcomes, Deputy Director of Enterprise Data, GIS Analyst, Data Engineer, Data Librarian, Program Analyst, Analytics Engineer, Community Engagement Specialist, Policy Analyst, Data & Outcomes Manager, Business Analyst, and Civic Data Designer.

c.) What is the total budget, including compensation, for the data science team?

Current: Our General fund budget is $754,558 for DS&A (Data Strategy & Analytics). This is not including software utilized by the team. Starting salaries are in the low six figures. The current FTEs on the team are all Detroit Residents. With the ARPA funds we were able to hire an additional 10 Detroit Residents. Between $5-8M will have been invested primarily on staffing over the course of the ARPA grant period for the 3 major projects they support.

Future: When ARPA ends, we are seeking to reduce staff to a sustainable level, but above the original General Fund staffing. In order to maintain base infrastructure, we will need to maintain at least a 10 person team. Additional positions, specifically that of Data Scientist, on top of those 10, will be required to fully utilize data science effectively and see the citywide cost savings we are striving towards. We utilize Open Source software whenever possible in order to keep costs low. Investing in staff, over proprietary software, allows us to remain nimble and flexible, enabling more customized departmental solutions when warranted.

3. **Key Performance Indicators (KPIs) and Metrics:**

a.) What are the key performance indicators (KPIs) and metrics used to measure the success or failure of the data science team's initiatives and projects?

KPIs are determined in consultation with the director or lead for the project we are collaborating on. This allows for the measures to reflect both the real-world context and goals of the program as well as our ability to collect data that accurately measure those achievements. For instance, when DS&A was developing the

Paratransit Performance Tracker with LEAN and DDOT, it was critical for accurately reporting the contractors' performance that we understand what meaningful measures were as well as how the data reflecting those performance metrics would be provided.

b.) How is performance evaluated and what benchmarks are set to ensure accountability and continuous improvement?

Team performance is generally evaluated on four benchmarks—cost-savings, time-savings, improved resident outcomes or service delivery, and expanded use of our data tools and services. Our work costs 50% less than contractors by a conservative estimate; onboarding and procurement delays do not need to be factored in. We come with a wealth of institutional and regional knowledge. As a result, we are able to execute on custom departmental requests within weeks rather than months and the end-product is often better. Our three Business Analysts facilitate our continuous improvement by regularly interacting with departments to seek feedback as we incrementally deliver goals for the project, beginning with a thorough intake process.

4. Data Management

a.) How and where is the data collected for the data science team managed and stored?

Data is collected and stored in our new Enterprise Data Warehouse (EDW). We named it Topsoil, to highlight the tangible role we feel it should play in growing the City's capacity and feeding its knowledge, while increasing equity and efficiency. The EDW is a digital store of the City's information that can be accessed, compiled, and analyzed quickly to make informed decisions. Data flows into the EDW from multiple departmental systems, and other sources on a regular cadence. DSA staff clean and standardize the data using various software as it is imported and then store it for later use in the central warehouse.

Promoting data standardization around city data is one way that data management and governance paves the way to data science. We have developed 4 data standards thus far, including:

- Address data standard: The Base Units database and toolbox is a data product maintained by DS&A. Via the standardization of address data it allows for automated data cleaning and address verification. It also **provides the foundation for data integration and analytics across city systems via standardized reference locations within the City by address, building, street or parcel.** In addition to being the backbone of the successful census challenge, it enables departments and the public to conduct their own data science and analytics projects by joining disparate datasets.
- Disaggregated demographic data standard: This governance tool allows programs and services to adopt a standard way of tracking different demographic details on participants when appropriate. **This allows for an equity lens to be incorporated into operational decisions**, but also balances the need for privacy by providing strict guidance on data structure and management.

Many data science projects are considerably slowed by a lack of available data or data that requires significant cleaning and prep work before analysis can begin. Data management and governance are important investments and necessary preliminary steps for making data science possible at a meaningful scale and cadence. As we evolve, departmental self-service and more sophisticated algorithms and modeling will emerge. However, its still worth highlighting some examples of the immediate benefits of good data management, prior to any data science applications:

Energy Benchmarking Ordinance: We ingested data from two utility companies and then used our Base Units to match that data to the correct municipal buildings to satisfy the requirements of this new ordinance. It is important for this cleaning and standardization be completed early to be prepared for the ordinance's scaled requirements over time, let alone create the opportunity for accurate and effective use in data science.

Detroit Restaurant Inspections: The legacy data system used to track and manage restaurant inspections has a very complex and large data schema. Providing basic information about inspections takes a very complex set of queries and joins. We have this data available on the open data portal as well as in a more user-friendly app. We are now prepared for the potential need to further publish or summarize inspection data as required by city council.

In addition, we work with interested departments to improve their data gathering processes and configuration of information systems to further streamline data cleaning and centralization. **We often partner closely with the Continuous Improvement team.** Here are some example collaborations:

- Visualizing CERA data to improve the distribution of rental assistance leading to over $262 Million dispersed to renters and landlords and over 32,000 approved applications.
- Developing platform integrations to improve the NEZ Homestead Tax Abatement process for appraisers, identifying 30,000 eligible homes and improving the experience for 9,000 households to apply and take advantage of the program.
- Designing and maintaining a dashboard to support track case closure metrics on the 125 monthly cases the Detroit Board of Police Commissioners receives, ensuring that investigators are held accountable for timely investigation completion.

b.) How is the data protected?
All data ingested into the EDW is protected to extent required by laws, regulations, and City policies. The specific protections implemented will depend on how the data is classified via our data classification policy. The most sensitive data is protected through its storage in on-prem servers. Most EDW data is housed in Snowflake; a secure and flexible cloud-based data warehouse platform used by many government entities including City of Denver, City of San Francisco, and NASA. It has multiple types of secure data warehouse tools for querying and analyzing data. DS&A has partnered with our Cyber Security team to ensure there is adequate protection for all steps in these processes.

5. Departmental Support and Requests:

a.) What is the process for other departments within the city to request support or collaboration from the data science team?
Primarily, departments will contact the Data Strategy and Analytics Team (DSA) to request support for a specific project. They work closely with a Business Analyst on the team to identify technical requirements, which are then delegated to engineers, architects and analysts on the team to execute.

We also conduct monthly data stewardship meetings that are open to all data professionals in the City. Over 100 employees have gone through our workforce trainings and have been introduced to the EDW and importance of data standards and data management. These meetings highlight new opportunities and developments in the City's data science ecosystem and provide a less formal mechanism to request support or collaboration as well as enable self-service analytics. This approach reflects our understanding of the most effective data infrastructure for the City being like a wheel. **DSA provides a strong, central hub of clean data and the tools and expertise to use it, while also encouraging the development of robust spokes connecting departmental data stewards to that hub.**

b.) Are there specific protocols or criteria that departments must follow to leverage the expertise of the data science team?
We want to engage any and all customers where they are at. The history of data in the City is largely one of departmental silos and bespoke solutions to fit the capabilities of legacy systems and the specific needs of the moment. Every project is the chance to build trust and ideally, over time, change the data culture from one

focused on data ownership to data stewardship, so that valuable data assets managed by one department can be easily utilized for the good of the whole and the benefit of ever resident.

CC: Hon. Latisha Johnson, Chair, Internal Operations Standing Committee
Hon. Janice Winfrey, City Clerk
Malik Washington, Mayor's Liaison