NYC Expense Budget

Group 4

Jianfei Shi, Leyi Shi, Yumeng Tian, Ruitong Yin, Ziyue Pan

Background

> "
>
> **The expense of NYC government is highly scattered; it can be levitated to mainly two cases in our project study.**

Business Use Case

Leveraging NYC's expense budget analysis, this study aims to uncover opportunities for potential savings and efficiency improvements. Targeted at city planners, policy makers, and civic tech organizations, the analysis compares the **Adopted and Financial Plan expense data** to assess budget execution efficiency. By identifying potential areas of overspending and cost-saving initiatives, the findings shed light on the underlying motivations driving budget allocations. The study further explores data insights to propose rational budget adjustments and cost-saving strategies, paving the way for **budget optimization** and **resource allocation improvements**.

Analytics Use

Leveraging advanced data analytics, we can uncover valuable insights by identifying trends, anomalies, and patterns within NYC's expense budget.

Expense Trends: By meticulously analyzing historical data, we can discern expense trends and inconsistencies between planned budgets and actual expenditures. Such insightful analysis empowers decision-makers to make informed choices for the future, optimizing resource allocation and budget planning.

Data Specification

Data Source

The origin of the dataset is NYC Open Data - Expense Budget dataset from https://opendata.cityofnewyork.us/data/ run by the NYC government.

We can access the New York City expense budget dataset through the Mayor's Office of Management & Budget section. The dataset was created in 2016. It is presented by the government and updated every 4 months. The most recent update is on July 7, 2023.

Data

Procurement Details

The data is publicly available and can be accessed via API or downloaded as CSV, JSON, etc.

We can directly access the open source through the link https://data.cityofnewyork.us/City-Government/Expense-Budget/mwzb-yiwb

Project Framework

Step 1: ETL Pipeline - Extraction

Extraction To obtain the most accurate and up-to-date data, we will directly access the NYC Open Data API in JSON format. This ensures that our analysis is based on real-time information, enabling us to make informed decisions and draw valuable insights from the expense budget data.

Step 2: ETL Pipeline - Transform & Load

In **PostgreSQL**, we transform JSON data into data frames. After selecting specific rows, we perform data cleaning by removing duplicate values. Then, we proceed to insert the cleaned data into our normalized tables, ensuring a structured and organized representation of the information.

MongoDB is an ideal choice for managing unstructured data, like text or mixed data. It automatically drops invalid rows and efficiently inserts the data into the database, streamlining the process and ensuring data integrity.

Step 3: Front-end API

Flask Use this lightweight web framework to create an API for data retrieval. Flask is easy to use and can be deployed on various platforms. The API will allow users to query the data in the PostgreSQL and MongoDB databases

PostgreSQL (SQL Database)

Key Advantages

- **Data Accuracy Assurance**

- **Compelling Performance**

- **Enhanced Efficiency**

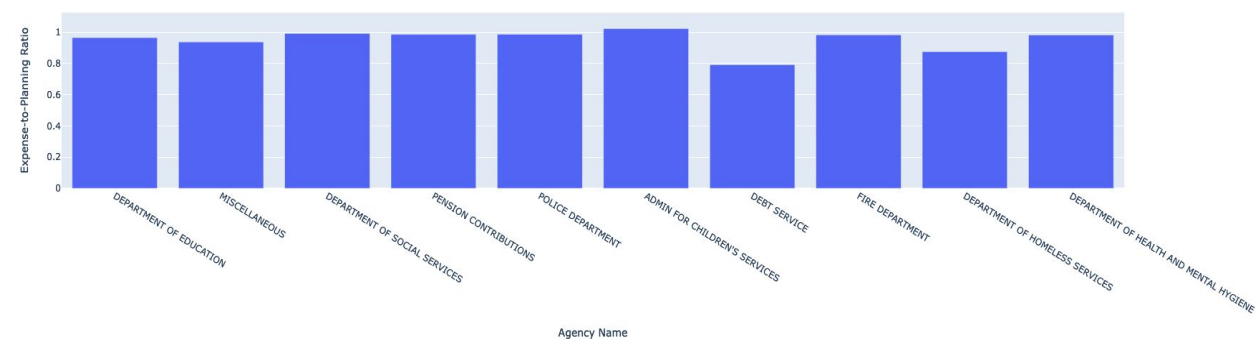- **Facilitating Data Interpretation**
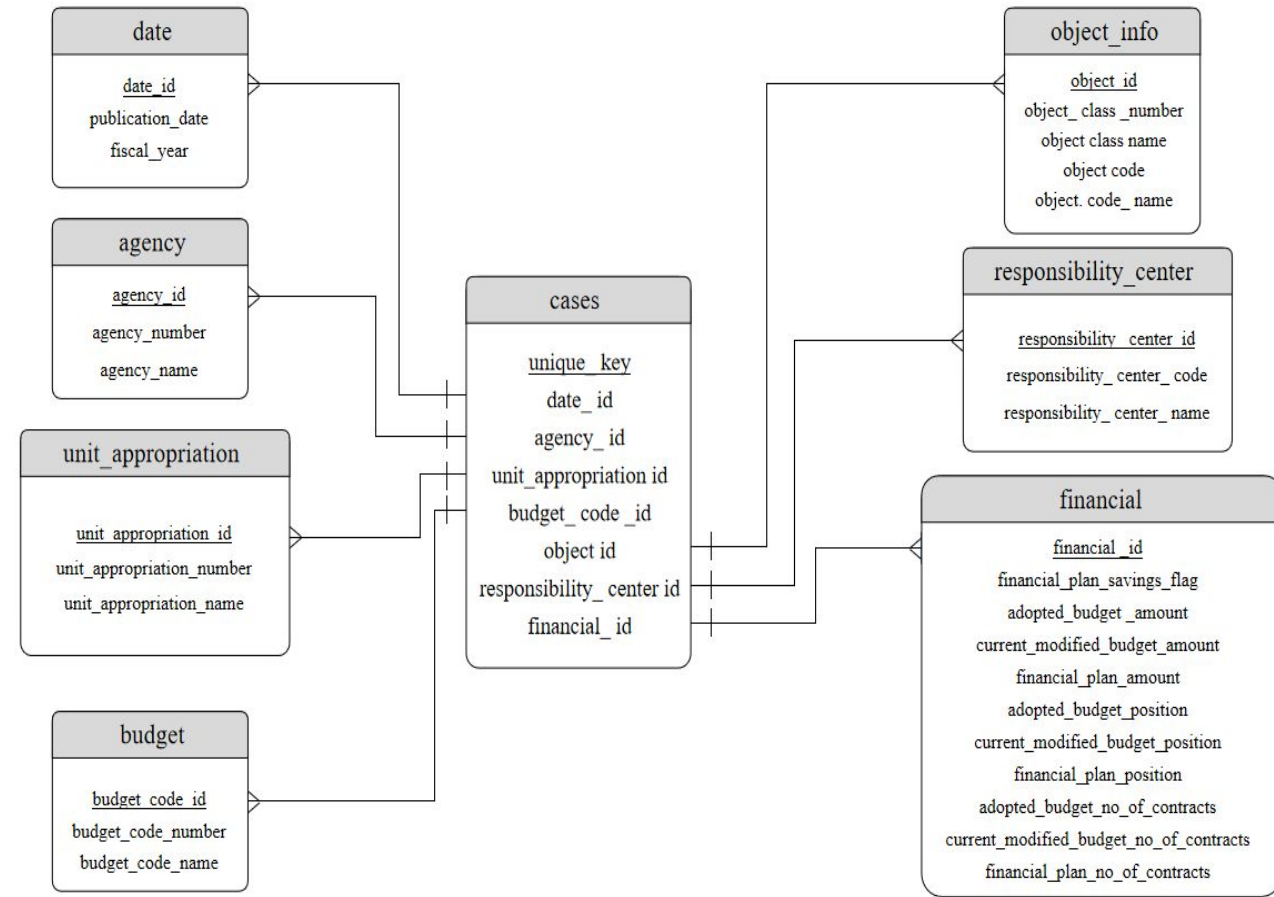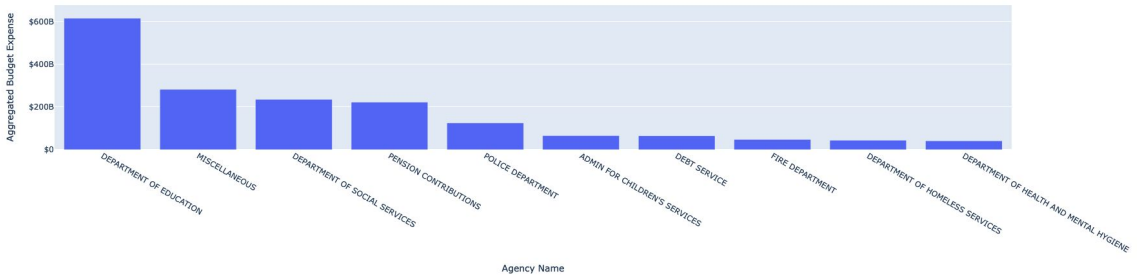
MongoDB (NoSQL Database)

Unstructured or Semi-Structured Data

Flexible and scalable data storage, such as web and mobile applications, content management systems, and real-time Analysis
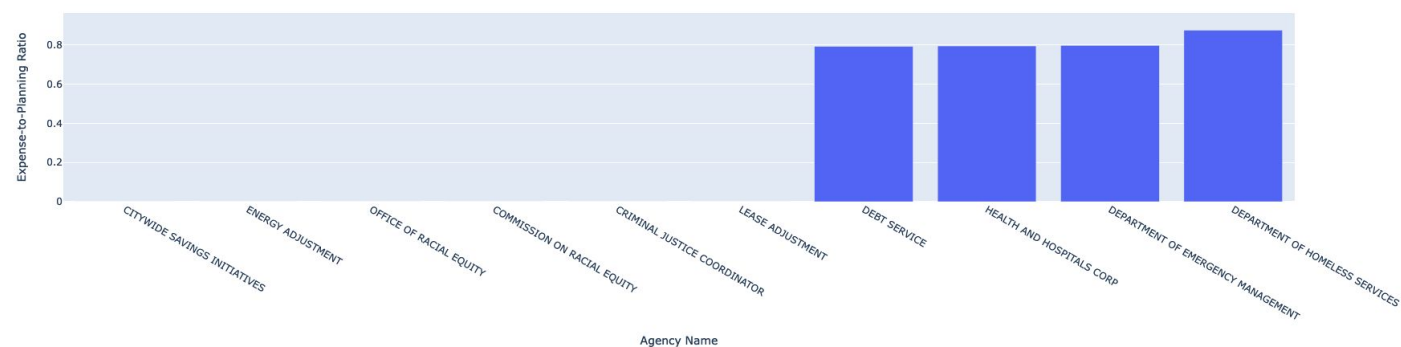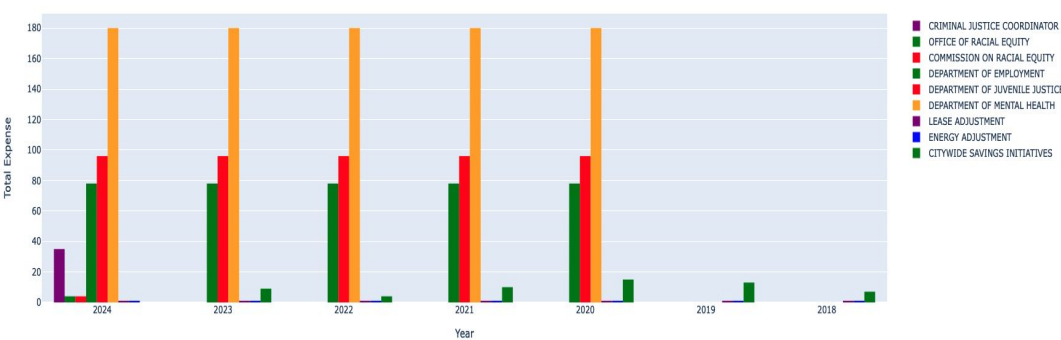
Flask

Flask is simple, quick and has interactive capabilities make it an ideal api for users to handle data retrieval, analysis, and visualization

Easy to Use

Designed to be simple and easy to use, making it accessible for developers and users. Its lightweight nature allows quick setup and deployment, reducing development time and resource consumption.

Fast Data Retrival

Its seamless integration with PostgreSQL, enabling efficient data retrieval and handling. Users can easily query the dataset by various criteria, to facilitate their targeted analysis without overwhelming the system.

Interactive Data Visualization

Flask's integration with MongoDB enables the access to interactive data visualization. Users can access charts, graphs, and visual representations made by plotly to comprehend expense patterns and identify trends.

Scalability and Cost Implications

Scalability

The chosen technologies have been carefully designed to efficiently handle large datasets, ensuring seamless scalability. Notably, both PostgreSQL and MongoDB possess the capability to be readily scaled up to accommodate any future increase in data volume, if needed.

Cost

PostgreSQL, MongoDB, and Flask are all open-source and available for free usage. However, expenses may arise when hosting the databases and the Flask API on a cloud platform, such as AWS, Google Cloud, or Azure. The costs incurred will be contingent on the scale of data and the volume of API requests.

Applicable Data Quality Dimensions

Completeness

We have imputed the missing values and kept the dataset as complete as optimal.

Uniqueness

We have removed duplicate values and kept only meaningful ones.
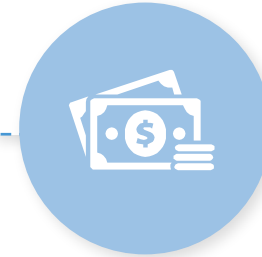
Timeliness

The Expense Budget Dataset from NYC OpenData is lastly updated on July 7, 2023.

Validity

We have corrected the data format, type, and range for specific use of project.

Accuracy

We have removed inaccurate entries and outliers to obtain precise outcome.

Conclusion

Licensing
The NYC Open Data platform operates under the NYC Open Data Terms of Use. The Expense Budget dataset is provided by Mayor's Office of Management & Budget.

Data-driven Decision Making
The project aims to use data analytics to drive decision-making around NYC's expense budget, potentially leading to cost savings and efficiency improvements.

Forecasting
From the budget allocation, data analyst may offer strategic advice based on expense trend in the past and forecast investment in future.

Recommendation
- Develop a variety of analysis such as cost and sensitivity analysis based on our project outcome.
- Adding more advanced analytical sessions like predictive modeling will enhance overall comprehension and consistency of project scope