**Amrita School Of Artificial Intelligence**

**B.Tech CSE-AI**



**19MNG338 – Operations Research**

**19MEE346 – Managerial Statistics**

**Professor:** **Dr. Shyam A V**

**Semester-7 Project Report**

**Project Title:-** Optimized Budget Allocation and Analysis Using Boston Operating Budget Dataset

**Batch-A Team-06**

|  |  |
| --- | --- |
| Name | Roll No |
| B. Harish Balaji | CB.EN.U4AIE21007 |
| Challa Yoganandha Reddy | CB.EN.U4AIE21008 |
| M. Mohit Sasank Reddy | CB.EN.U4AIE21031 |
| P. Sai Vijay Kumar | CB.EN.U4AIE21040 |
| Y. Sri Sathwik Reddy | CB.EN.U4AIE21077 |

**ABSTRACT**

This project explores the Boston Operating Budget dataset to analyze budgetary trends and develop optimized allocation strategies. Data is stored in MySQL, analyzed in Python, and visualized through an HTML template. Statistical methods, including time series analysis, correlation matrices, and distribution analysis, reveal financial patterns, while optimization scenarios simulate practical decision-making under constraints. By addressing budget limits, departmental spending caps, and key program funding requirements, this project demonstrates how data-driven strategies can enhance fiscal discipline, ensure equity, and prioritize critical services. The outcomes highlight the potential for informed financial planning in organizational and governmental contexts.

**INTRODUCTION**

Effective fiscal management is crucial for ensuring that public funds are allocated judiciously across various administrative entities and programs. With increasing financial complexities and resource constraints, organizations must adopt data-driven methods to optimize budget distribution and track spending patterns. This project leverages the Boston Operating Budget dataset to analyze historical financial trends, identify inefficiencies, and propose optimized budget allocations. Using advanced statistical techniques, visualization tools, and optimization models, we aim to provide actionable insights that improve decision-making and promote equitable resource distribution.

**PROBLEM STATEMENT**

Budget allocation in large organizations is often challenged by competing priorities, resource constraints, and the need to maintain fiscal discipline. Traditional methods of budget planning may result in inefficiencies, inequitable resource distribution, or underfunding of critical programs. The Boston Operating Budget dataset provides an opportunity to explore these issues by analyzing historical data and applying optimization techniques.

**TOOLS**

* MySQL: For storing and managing the Boston Operating Budget dataset.
* MySQL Connector: To connect Python with the MySQL database for data retrieval.
* Python: For data analysis, visualization, and optimization.
* OR-Tools: For solving optimization problems efficiently using constraint programming.
* PuLP: A Python library for linear programming and optimization modeling.
* HTML Templates: For presenting analysis and optimization results interactively.

**DATASET DESCRIPTION**

**Dataset Description:** The dataset represents a comprehensive financial overview of an organization's operating budget spanning multiple fiscal years. It details budget allocations, actual expenditures, and appropriations across various administrative entities, including cabinets, departments, and programs. The data categorizes expenses into distinct types, such as personnel costs, operational expenses, or contractual services, and tracks changes in financial planning over time. This information provides insights into financial trends, resource allocation, and spending patterns, making it valuable for analysis, optimization, and strategic planning in budgeting processes.

**Data Source:-** [**Click here to see data**](https://data.boston.gov/dataset/operating-budget/resource/8f2971f0-7a0d-401d-8376-0289e3b810ba)

**DATABASE INTEGRATION**

**Database Integration and Data Retrieval:**

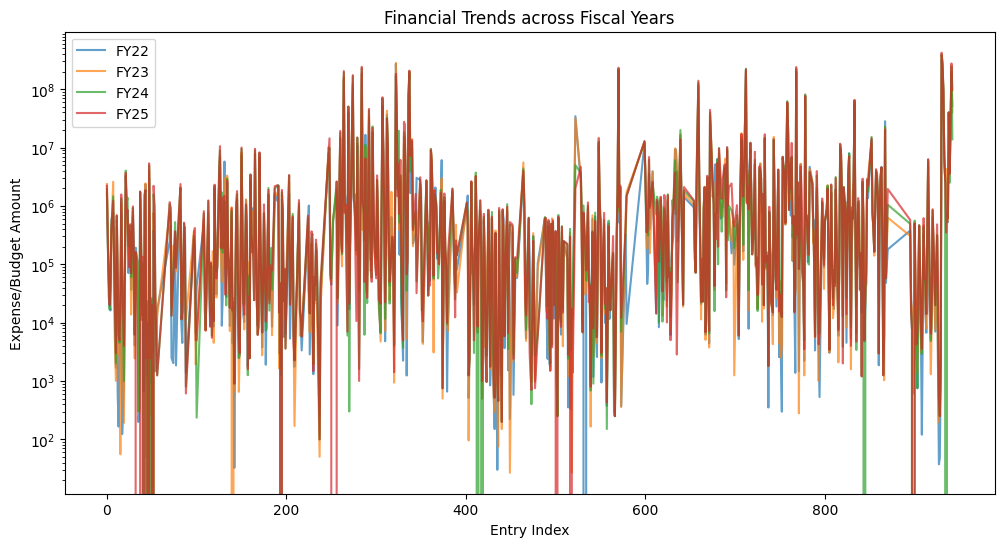
To efficiently manage and query the dataset, we integrated MySQL into the project. The workflow involved the following key steps:

1. **Loading Data into MySQL Database:** The dataset, originally stored as a .csv file, was imported into a MySQL database to allow for structured storage and efficient querying. This process ensured that the data could be easily accessed and manipulated in a scalable manner, especially when dealing with large datasets.
2. **Loading Data into MySQL Database:** The dataset, originally stored as a .csv file, was imported into a MySQL database to allow for structured storage and efficient querying. This process ensured that the data could be easily accessed and manipulated in a scalable manner, especially when dealing with large datasets.
3. **Data Retrieval:** Once the connection was established, we used SQL queries to extract the relevant data from the database. The retrieved data was then loaded into a Pandas Data-Frame, which facilitated easier manipulation and analysis for optimization tasks.
4. **Data Visualization:** After retrieving the data, we utilized Python libraries such as matplotlib and seaborn to visualize the dataset.

By storing and retrieving data through MySQL, we ensured efficient data management, enabling better performance when working with large datasets and improving the overall scalability of the project.

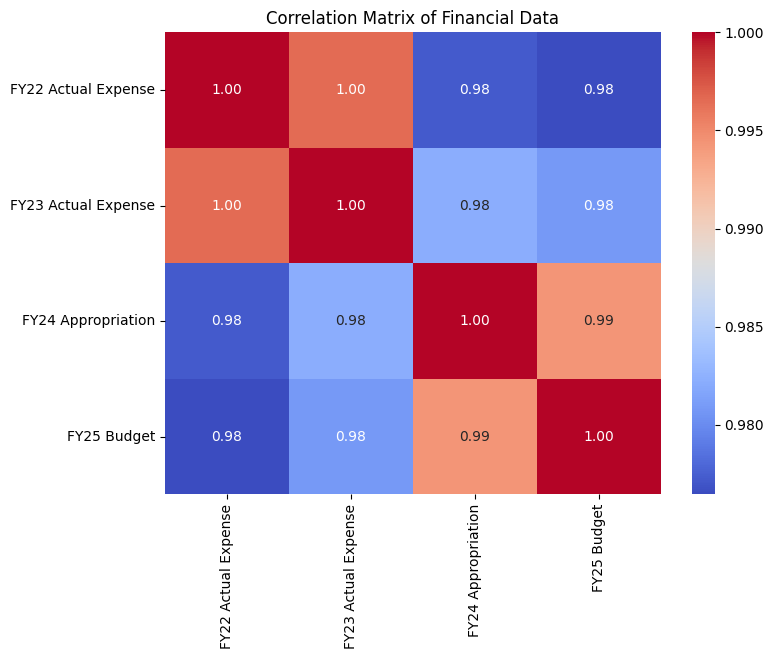
**STATISTICAL METHODS**

1. **Time Series Visualization of nature of expenses (Plot 1)**

* Method: Line Plot
* Purpose:
* About to demonstrate variations in expenses and the corresponding budget for the fiscal years FY22 to FY25.
* Data for each fiscal year is graphed separately in order to compare with other years. Statistical
* Insights:
* Analysing and getting insights of time series concerns itself with recognising patterns, deviation or trend within a certain period of time.
* The relative variability of the data is managed by using a logarithmic scale (yscale = ‘log’) which effectively allows numbers to be comparable but maintains their proportions.
* Strengths:
* Enables one to identify results in relation to the previous year’s results.
* Avoids skewness issues in data, but at the cost of data changing scale with logarithmic transformation.
* Limitations:
* Lacks the capability of dealing with monthly or weekly patterns and is unaffected by external factors.

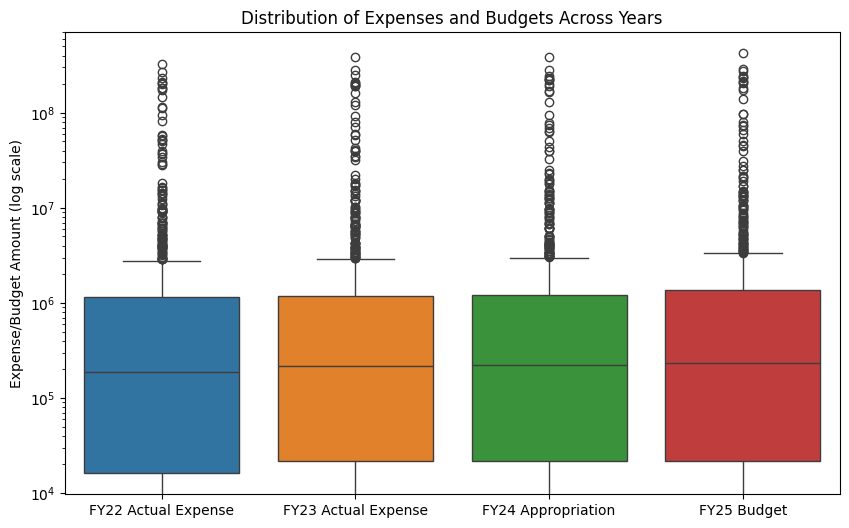
1. **Correlation Analysis (Plot 2)**

* Method: Heatmap of Correlation Matrix Coefficient of correlation is done using the Pearson method by default (.corr()).
* Purpose:
* To examine the extent of the relationship between fiscal year variables that are on a linear scale.
* Identifies correlation coefficients and specifies the form of relations (the values vary between -1 and 1).
* Statistical Insights:
* Positive correlation mean the variables rise or fall together while negative correlation mean the two variable move in opposite directions.
* Values near 0 indicate that there is no linear association existing in between the two variables.
* Interpretation:
* Of particular interest when assessing the extent of duplications or possible interdependence of fiscal information.
* Strengths:
* Easy and straightforward software to represent correlations.
* Limitations:
* While Pearson correlation coefficient is easy to calculate and used only for linear relationship it is also affected by outliers.



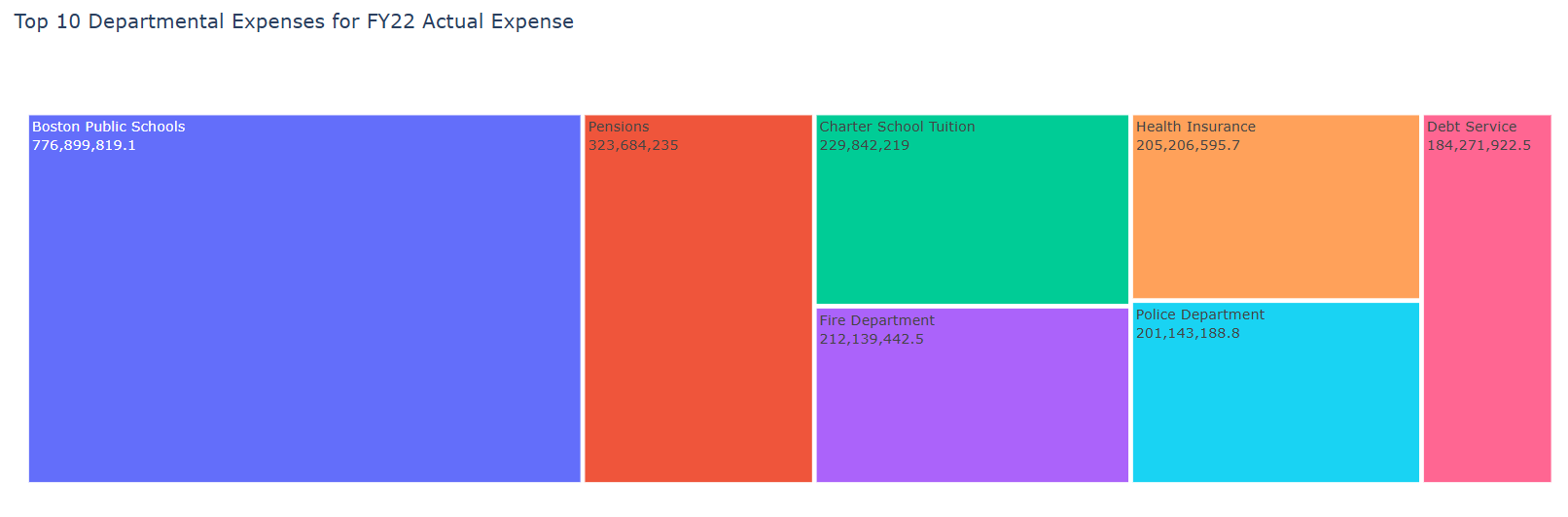
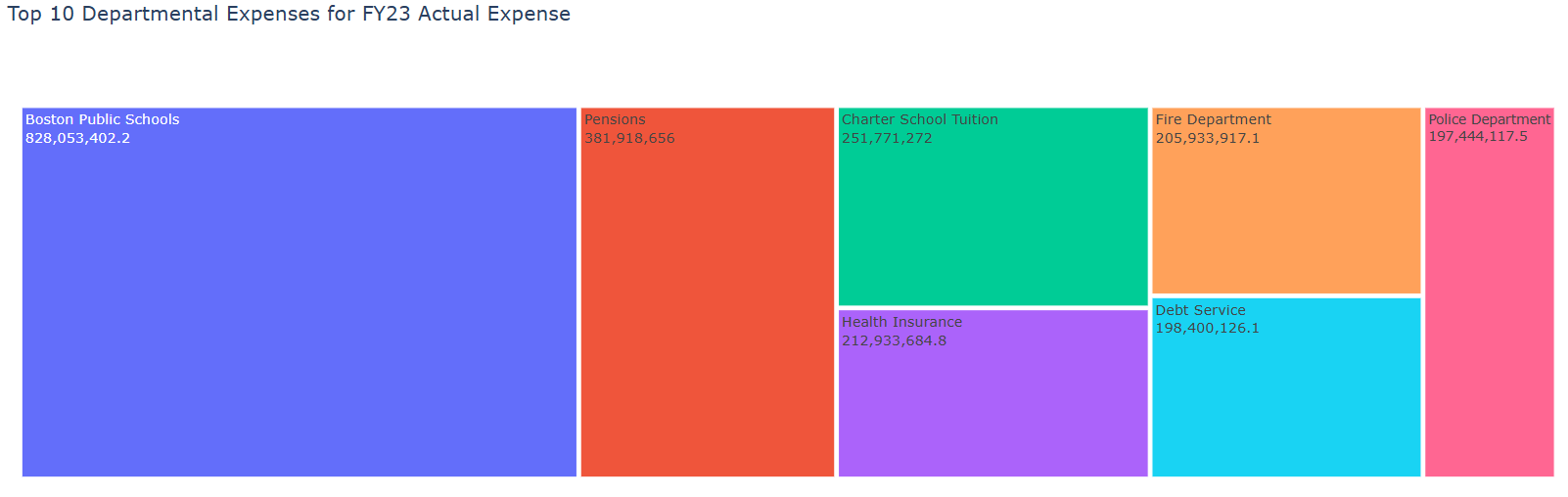
1. **BOX PLOT For Distribution Analysis (PLOT 3)**

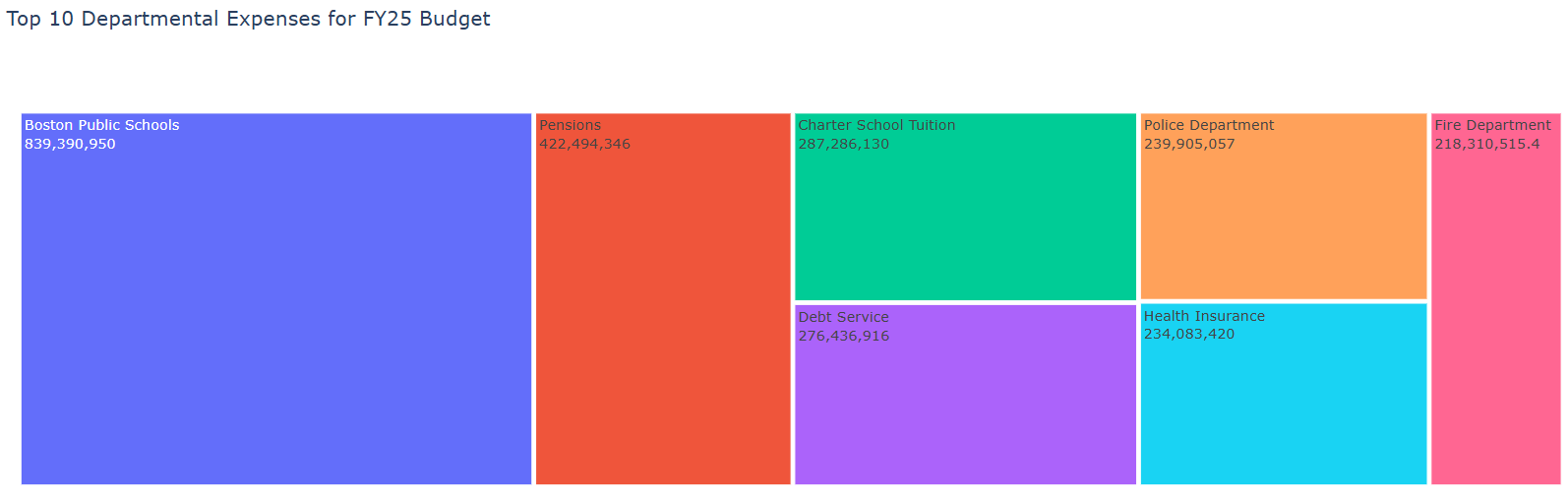
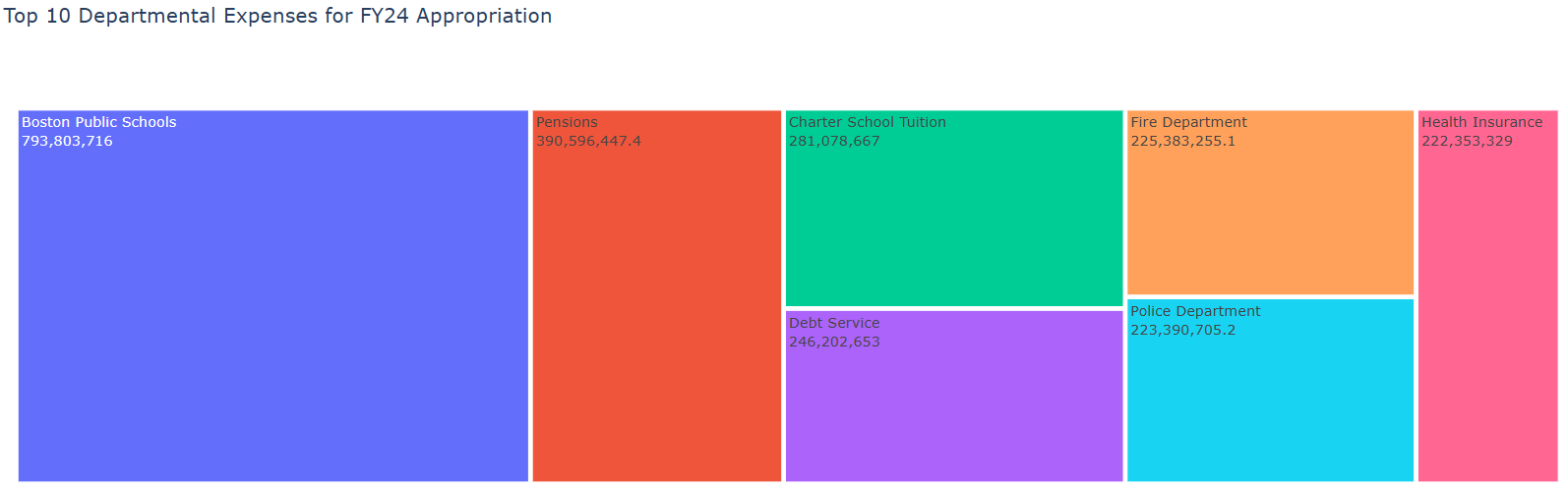
* Method: Box Plot is also referred to as the Box-and-Whisker Plot.
* Purpose:
* This page consolidates all the expenses and budgets for every fiscal year.
* Displays key statistical properties:
* Median: Central tendency of the data.
* Interquartile Range (IQR): Mean Plus/Minus One standard deviation spread of the data.
* Whiskers: Estimation of data variation or fluctuation or dispersion.
* Outliers: Crosses beyond whiskers are signs of unusual values.
* Statistical Insights:
* When making the y-axis logarithmic, as yscale= ‘log’, distributions that have a high variance or are skewed, becomes easier to analyze.
* Emphasizes variation and points at outliers (for example, remarkably high or low designations in a budget plan).
* Strengths:
* Tells the user a brief detail on how the values is distributed in the data set. It also is useful for stereotypically identifying potential outliers.
* Limitations:
* They do not preserve the detailed shape of the distribution characteristics such as modality.



1. **Treemap Visualization**

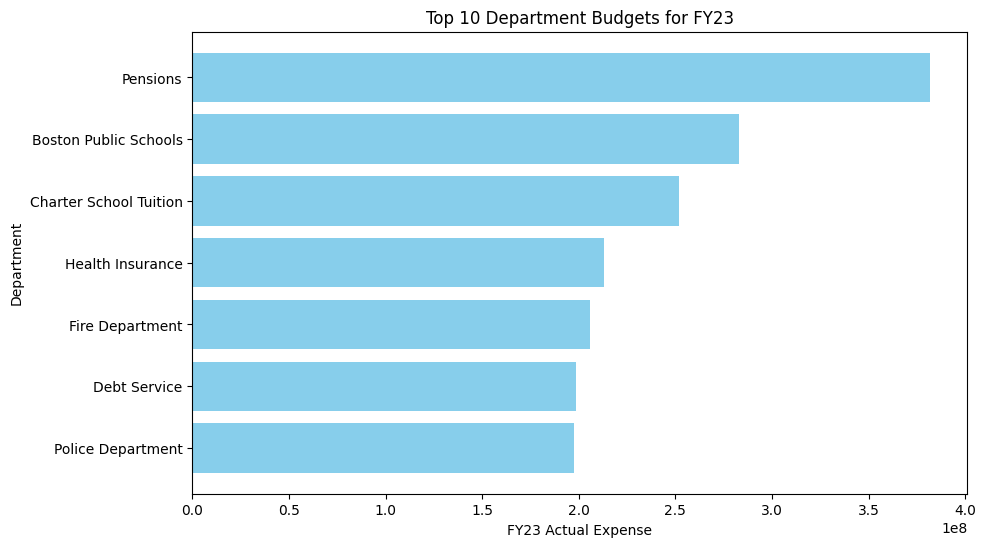
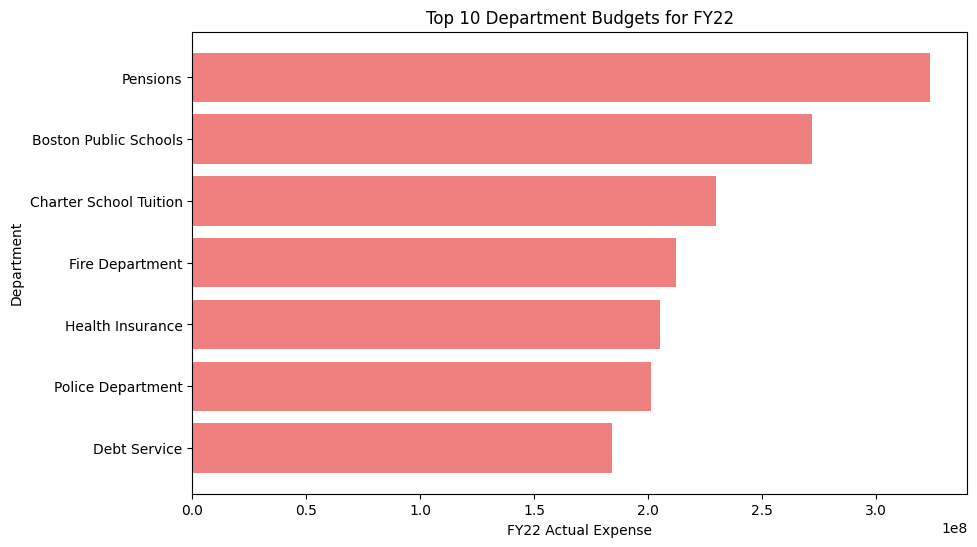
* Method: Treemap
* Purpose:
* Reads data in terms of rectangles which are scaled with the selected value being ‘FY22 Actual Expense, FY23 Actual Expense, FY24 Actual Expense, FY25 Actual Expense’ Yields both ratio-analytical in rectangles size and nominal in the department names.
* Statistical Insights:
* Illustrates how the top departments incurred the FY22 expenses, FY23 Actual Expense, FY24 Actual Expense, FY25 Actual Expense. Facilitate a quick comparison of the proportions of contribution between departments as well.
* Design Features:
* Path: The path=['Dept’ option means that the hierarchy is constructed based on departments.
* Values: The values parameter set to ‘FY22 Actual Expense, FY23 Actual Expense, FY24 Actual Expense, FY25 Actual Expense’ proactively occurred in the choosing of rectangle sizes to animate.
* Textinfo: Appending “label+value” ensures that both department names and expense values are depicted right on the plot which increases the plot’s readability.
* Strengths:
* A dense and ‘‘low-profile’’ form of graphic display for emphasizing quantitative differences and similarities.
* Stresses on pro rata representation of the expenses of each department compared to the company’s average.
* Limitations:
* It becomes difficult to read treemaps if it makes use of many categories or if the values in the categories are quite close to one another.
* It lacks some descriptive statistics such as variance or distribution.

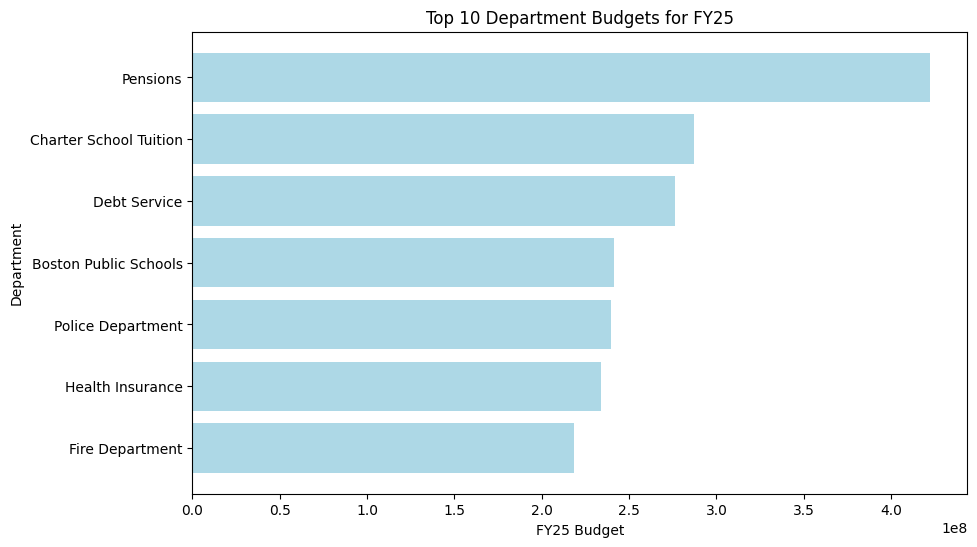
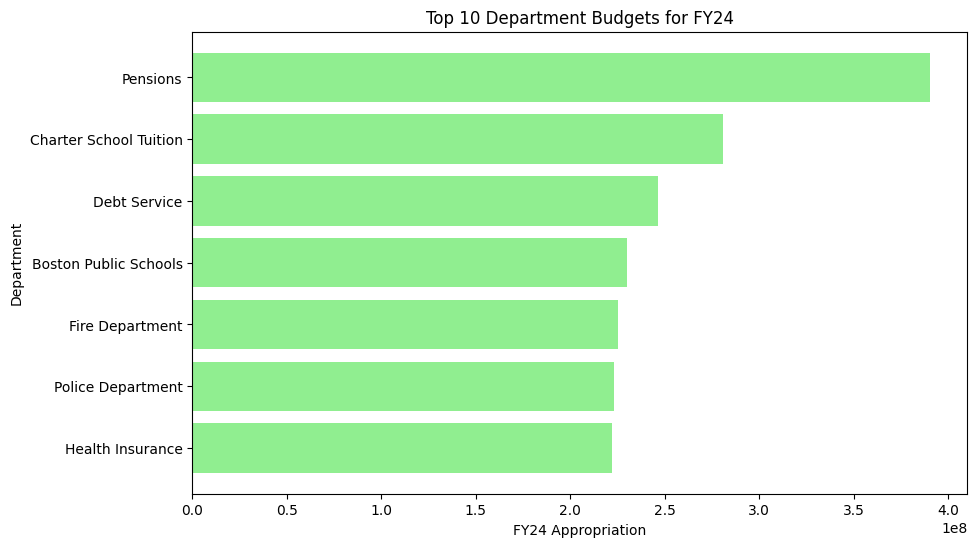




1. **Analysis of the data through the creation of a Horizontal Bar Chart Visualization**

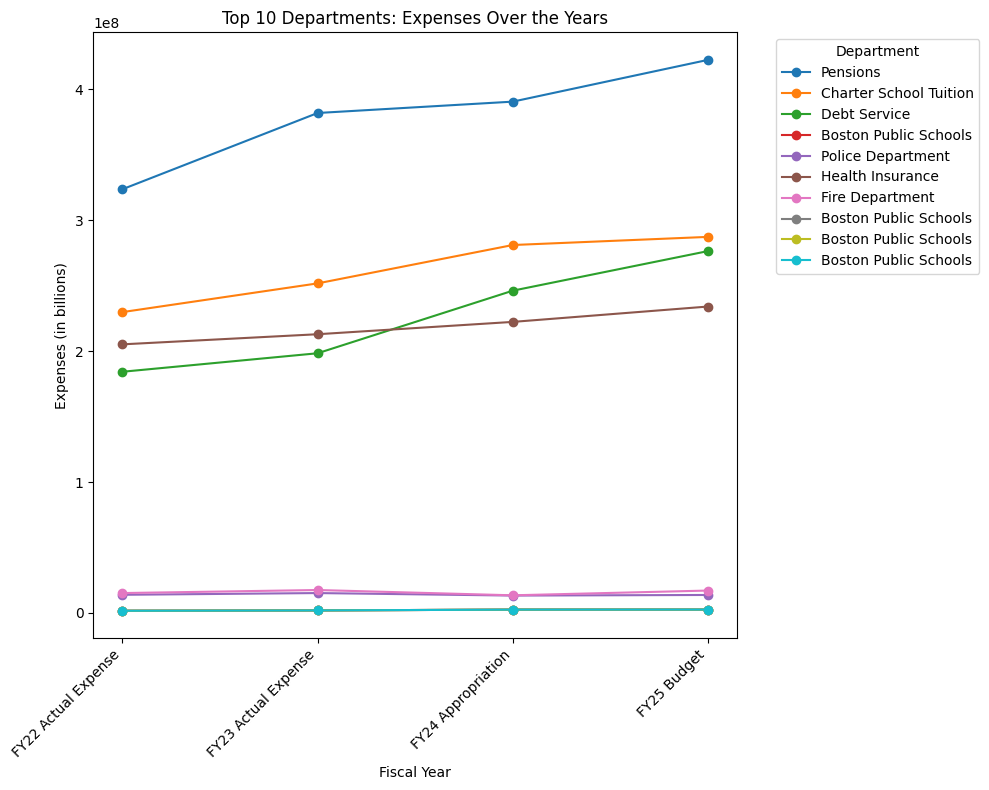
* Method: Horizontal Bar Chart (plt.barh) is the fourth method that gives visualization for two datasets using bar plot with horizontal orientation.
* Purpose:
* Summarizes FY22 Actual Expense, FY23 Actual Expense, FY24 Actual Expense, FY25 Actual Expense for the ten largest departments to facilitate paring them against one another.
* Statistical Insights:
  + The length of every bar indicates the extent of expenses for a department.
  + The alignment used is hierarchical with regard to descending order of the expenses making the department with the highest expenditure stand out.
* Design Features:
  + Color Scheme: Simplicity is achieved while avoiding visual confusion through the use of lightcoral as the color for the headings.
  + Axis Customization: To make the figure consistent with other graphs that demonstrate easy readability from top down, the y-axis can be inverted (gca().invert\_yaxis()).
  + Labels: Adding axis titles (FY22 Actual Expense, FY23 Actual Expense, FY24 Actual Expense, FY25 Actual Expense and Department) directly enhance interpretability.
* Strengths:
  + Suitable for making comparisons where categories are a few in number.
  + Offers a firsthand graphical view of the departments based on magnitude of expense.
* Limitations:
  + Exploitation of such kind of segmentation is not suitable in datasets with many categories so that they do not overcrowd.
  + Lacks modification or variation within an individual category or in quantitative distribution.

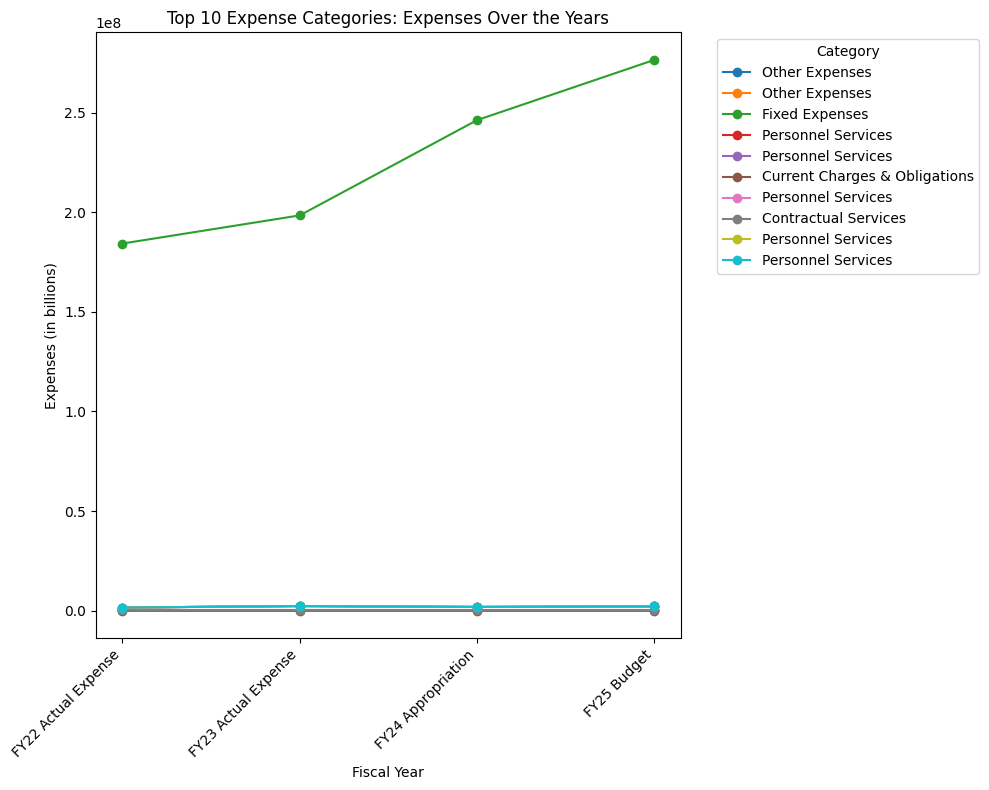




1. Multi-Line

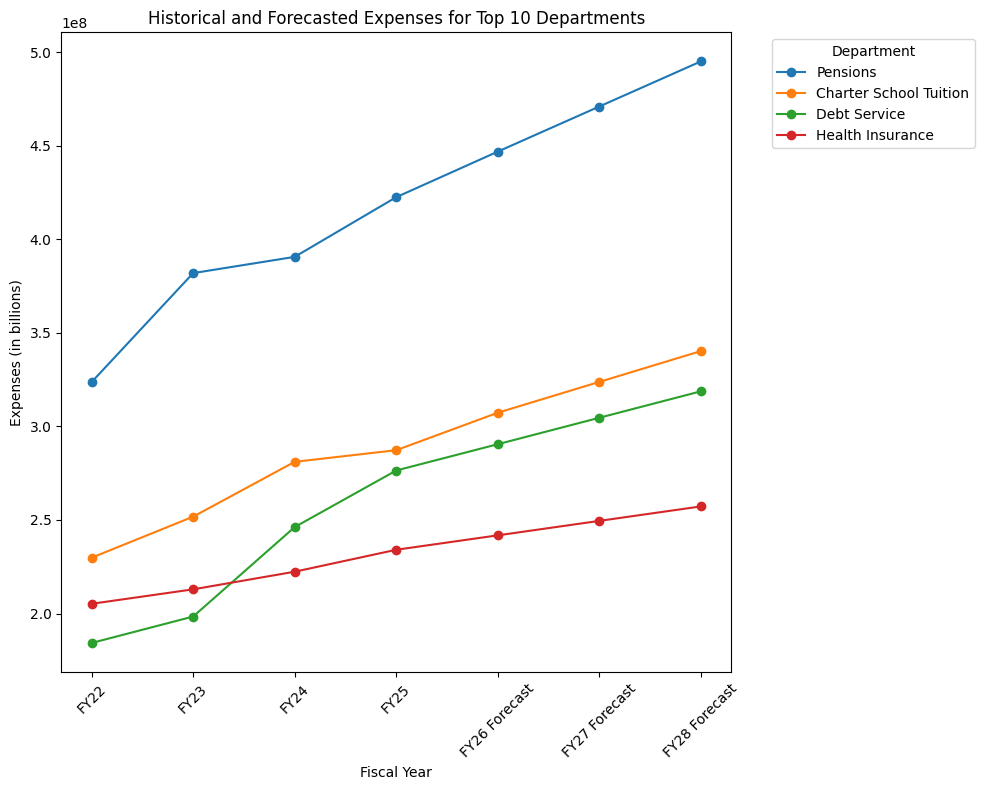
* Plot Method:
  + Line Plot for Trends What is displayed in each department is a line linking values in different fiscals years.
* Purpose:
  + Highlights fluctuation and changes in expense and budget proportions for major departments in the long run.
* Statistical Insights:
  + Is able to predict whether some of the expenses in the departments are on the increase or decrease or remain stagnant.
  + Some aspects to be noted inclusions aspects related to the budgetary arrangements among different departments of an organization and also differences in their expenditure trends.
* Design Features:
  + X-Axis: Nominal data that indicates levels of erector functionality throughout the fiscal years (FY22–FY25).
  + Y-Axis: Amounts, referred to as expense and budget to facilitate interpretation.
  + Markers: Placed for every point (marker=’o’) in order to enhance the contrast and separate the data points and see the values being represented individually.
  + Legend: A line legend is provided to enable easy identification of trends for each particular department.
* Strengths:
  + Managing temporal changes and comparisons across different departments have been achieved effectively.
* Limitations:
  + Parallel lines complicate the detection of trends for similar data sets.





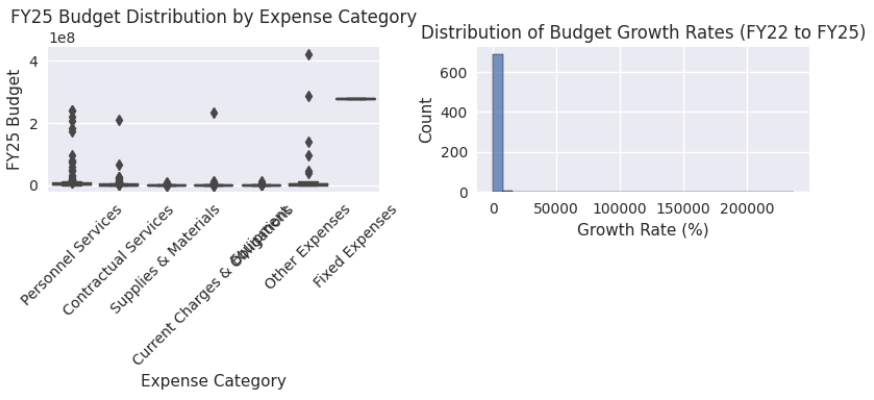
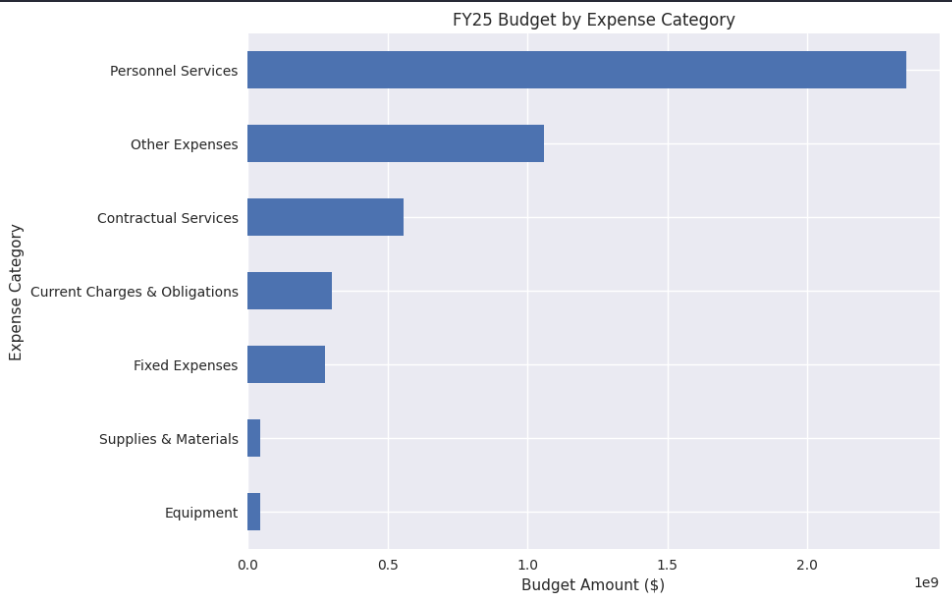
1. Exponential Smoothing, otherwise known as the Holt-Winters Method

* Method:
  + Additive Trend Model Under the additive trend option they make use of changes in the expense by modeling the trend as a linear function of time.
  + The model thus represents the magnitude or mean, the variability, and the rate of change of value over every other period of time.
* Purpose:
  + The reason is to determine future expenses for top departments using historical data.
* Statistical Insights:
  + The method tends to extrapolate the recent past activity data in the future, and therefore should be used where departments are characterised by linear activity trends.
* Strengths:
  + Easy to handle for short term planning and forecasting.
  + Gains or loses occurring at a linear path may be captured in this data.
* Limitations:
  + This data does not show cyclical trends since it is an average of data over an unspecified period.
  + Based on linear trends while this may not be the case for all the departments. It is better to have not less than a few actual figures/numbers (e.g., FY22 – FY25 figures) for forecasting purposes.



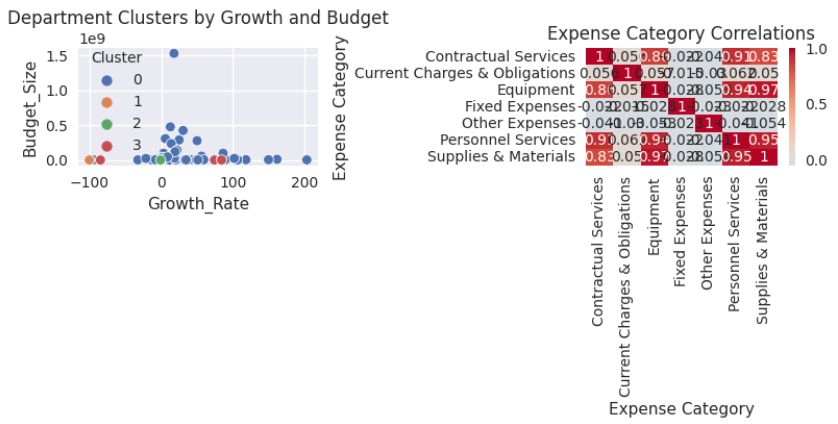
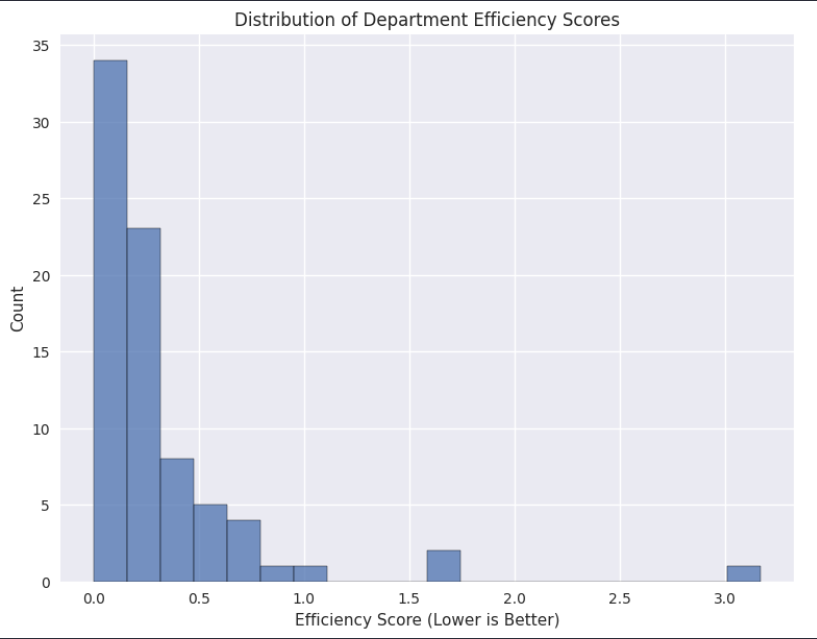
1. Summary Report

* Provides:
  + Annual expense statement for each year of the current strategic plan for Years: FY22, FY23, FY24, and FY25.
  + YoY changes as percentages.
  + Top 5 departments by number of budget in FY 25 budget.
* Recommendations
  + Customized recommendations are based on:
  + A continuous and progressive boost of the budget for the academic year as mentioned earlier, for instance, £710,000 to £890,500.
  + Fluctuations in between YoY.
  + In other words, large variations from trends observed in other years but not strongly associated with either party.



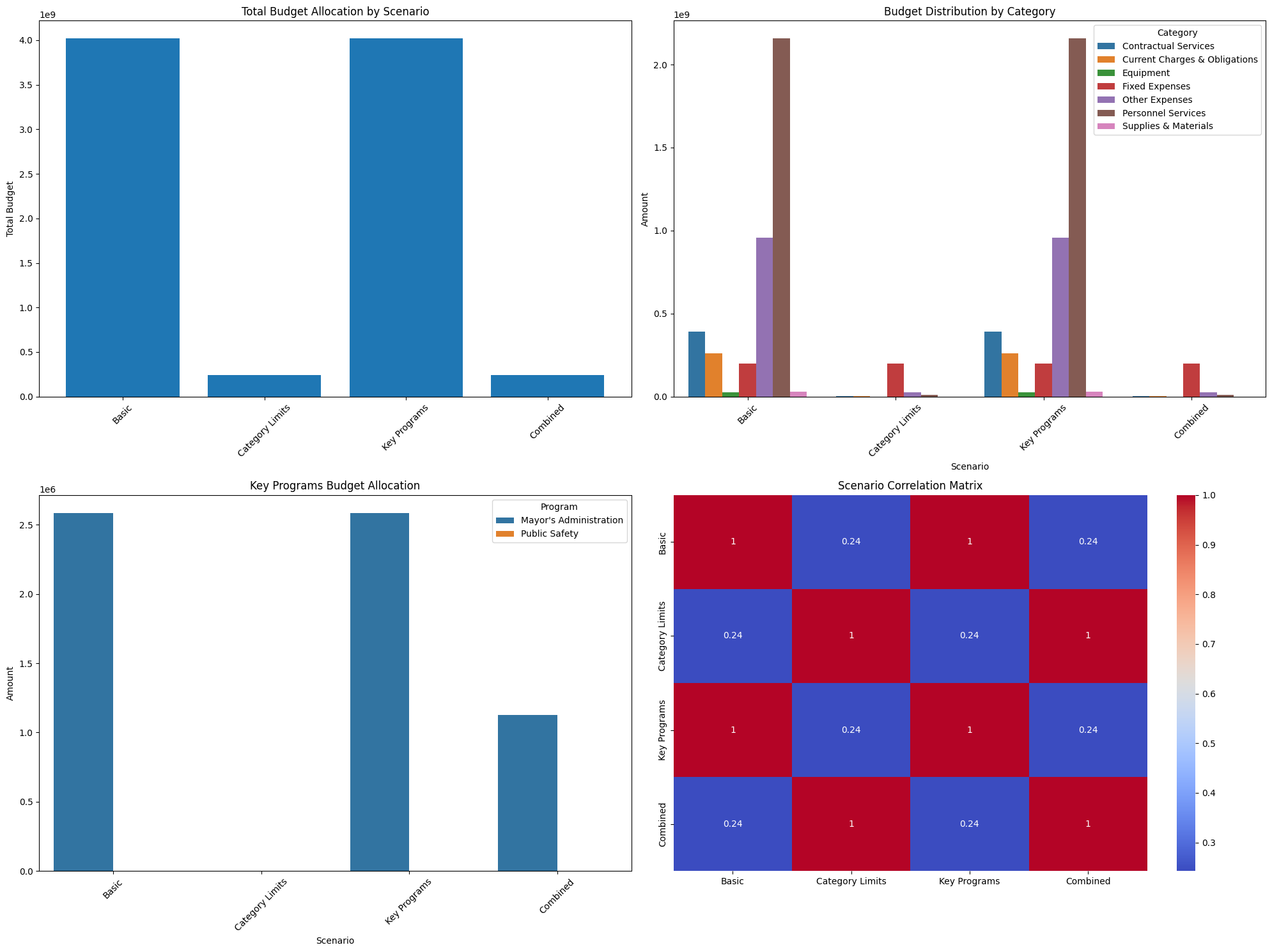
1. Summary Report

* Provides:
  + Efficiency score distribution
  + Budget scenarios comparison
  + Clustering results of departments
  + Correlation heatmap between expense categories



1. Summary:

* Total Budget Allocation: This will show how the total budget changes across the different scenarios.
* Budget Distribution by Category: A bar plot will show how the budget is distributed across different categories in each scenario.
* Key Programs Budget Allocation: A bar plot will show the budget allocation for specific key programs (like Mayor's Administration, Public Safety) in each scenario.
* Scenario Correlation Heatmap: A heatmap will show how similar the department-wise budget allocations are between different scenarios.



**OPTIMIZATION SCENARIOS**

This section explains the objectives and constraints for each scenario defined in the budget allocation optimization problem. These scenarios simulate practical decision-making in fiscal planning for an organization, aiming to achieve balanced and fair resource distribution across programs and expense categories.

**Scenario-1: Total FY25 Budget + Departmental Limits**

**Objective:** Maximize the total allocation of the FY25 budget while adhering to constraints on departmental spending limits and total budget limits.

**Constraints:**

1. **Total FY25 Budget Constraint**:

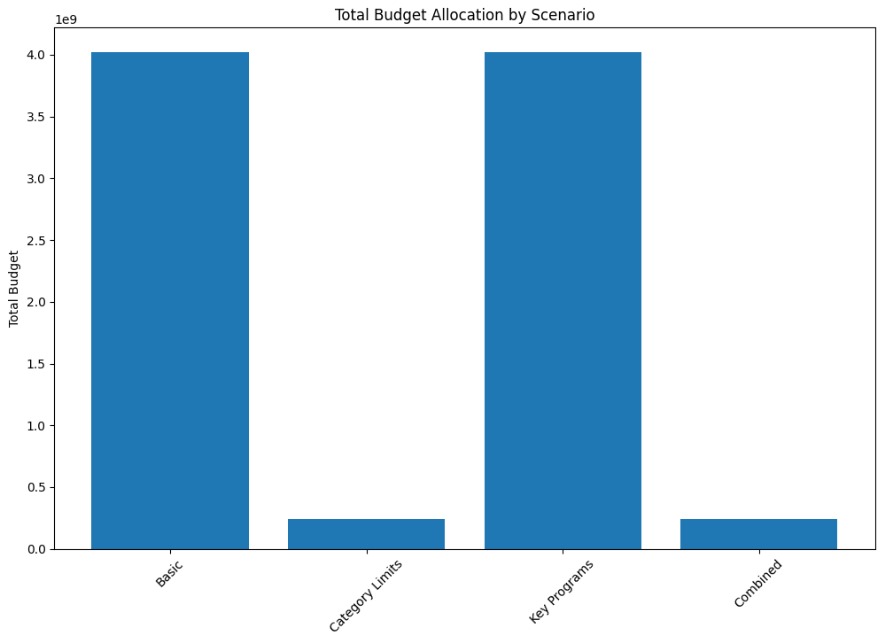
* The sum of all allocations cannot exceed the total FY25 budget available.
* **Reason**: Ensures that the total spending aligns with the approved FY25 budget, avoiding overspending.

1. **Departmental Budget Limits**:

* Each department's allocations cannot exceed its actual expenses in FY23.
* **Reason**: Maintains fiscal responsibility by restricting departments to historical spending levels to prevent over-allocation or underutilization.

**Why These Constraints?**

This scenario balances maximizing resource allocation while preventing overspending and ensuring departments operate within historically established budgets.



**Scenario-2: Total FY25 Budget + Departmental Limits + Category Spending Limit**

**Objective:** Maximize the total allocation of the FY25 budget while incorporating constraints on departmental limits and expense category limits to maintain fairness across categories.

**Constraints:**

1. **Total FY25 Budget Constraint**:

* The total allocation across programs must not exceed the total FY25 budget.
* **Reason**: Maintains fiscal discipline.

1. **Departmental Budget Limits**:

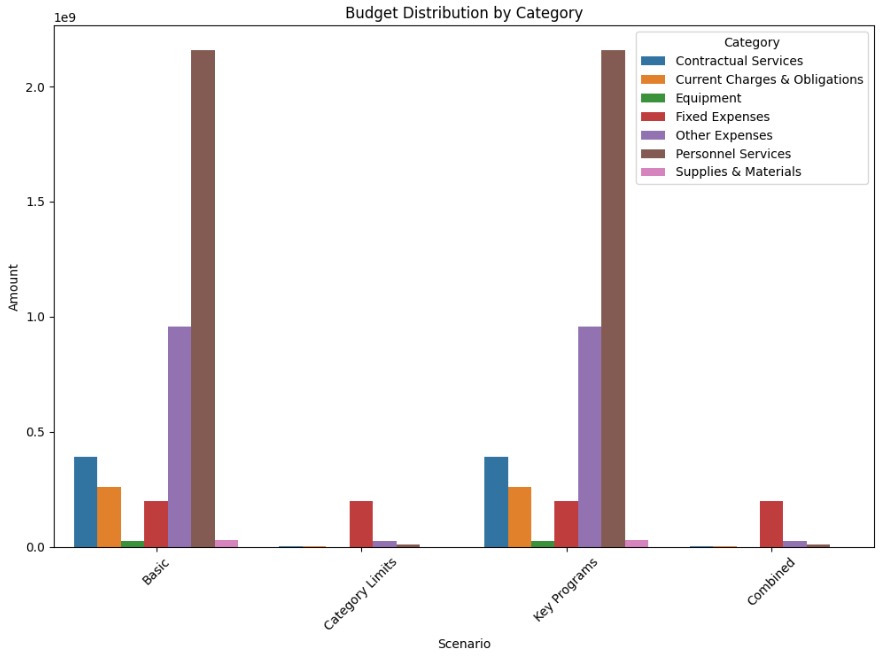
* Each department’s total allocation is capped by its actual expenses in FY23.
* **Reason**: Prevents departments from exceeding their historical spending patterns.

1. **Category Spending Limit**:

* The allocation for each expense category is capped by its average FY23 actual expense.
* **Reason**: Ensures equitable distribution across expense categories and prevents disproportionate funding to specific categories.

**Why These Constraints?**

This scenario adds an additional layer of fairness by capping allocations per expense category, ensuring that resources are distributed equitably across the organization.



**Scenario-3: Total FY25 Budget + Departmental Limits + Minimum Funding for Key Programs**

**Objective:** Maximize the total allocation of the FY25 budget while prioritizing critical programs and ensuring departments remain within their limits.

**Constraints:**

1. **Total FY25 Budget Constraint**:

* The total allocation cannot exceed the FY25 budget.
* **Reason**: Enforces fiscal responsibility.

1. **Departmental Budget Limits**:

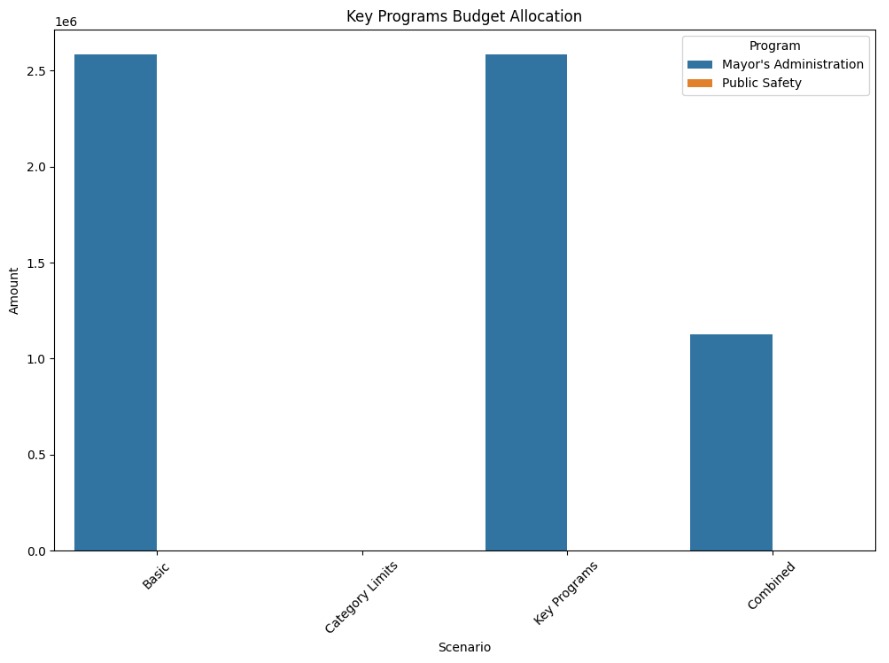
* Allocations for each department are capped at their FY23 actual expenses.
* **Reason**: Encourages departments to operate within historical spending limits.

1. **Minimum Funding for Key Programs**:

* Critical programs (e.g., Public Safety, Mayor’s Administration) must receive at least 50% of their FY24 appropriation.
* **Reason**: Protects essential services and high-priority programs from underfunding.

**Why These Constraints?**

This scenario prioritizes essential programs that are critical to the organization’s functioning, ensuring they receive sufficient funding while still adhering to overall fiscal discipline.



**Scenario-4: Total FY25 Budget + Departmental Limits + Category Spending Limit + Minimum Funding for Key Programs**

**Objective:** Maximize the total allocation of the FY25 budget while maintaining equity across categories, protecting critical programs, and ensuring departments stay within historical spending levels.

**Constraints:**

1. **Total FY25 Budget Constraint**:

* The total allocation must not exceed the FY25 budget.
* **Reason**: Adheres to the approved FY25 fiscal plan.

1. **Departmental Budget Limits**:

* Allocations for each department are limited to their FY23 actual expenses.
* **Reason**: Maintains spending discipline by capping allocations to historical levels.

1. **Category Spending Limit**:

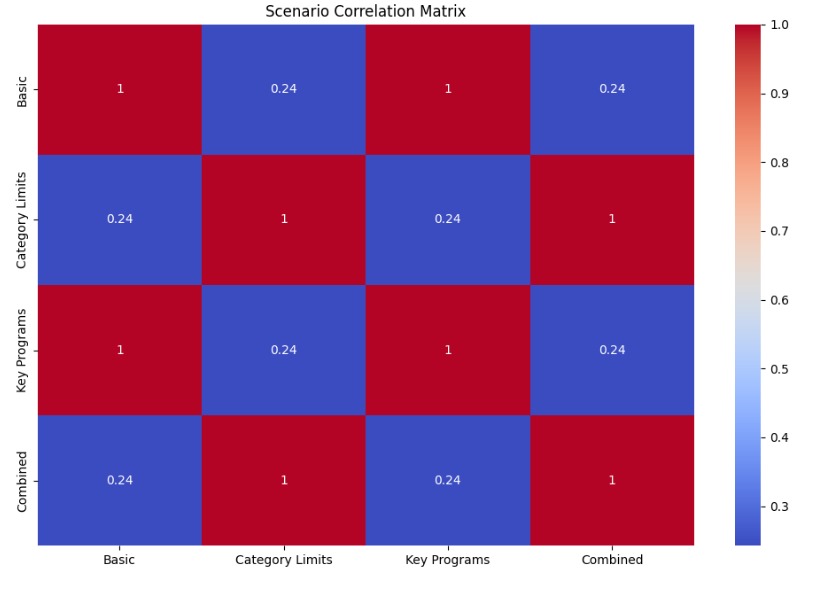
* Allocations for each expense category are limited to the average FY23 actual expense for that category.
* **Reason**: Prevents disproportionate allocation to any single expense category, ensuring fairness.

1. **Minimum Funding for Key Programs**:

* Key programs (e.g., Public Safety, Mayor’s Administration) receive at least 50% of their FY24 appropriation.
* **Reason**: Ensures critical programs have the resources to operate effectively.

**Why These Constraints?**

This scenario combines all previous constraints for a comprehensive fiscal planning strategy. It ensures overall budgetary discipline, protects high-priority programs, and maintains fairness across categories and departments.



**Scenario-5: Total FY25 Budget + Proportional Allocation Based on FY23 Expenses**

**Objective**: Maximize the total allocation of the FY25 budget by distributing funds proportionally across programs and categories, based on their FY23 expenses.

**Constraints**:

1. **Total FY25 Budget Constraint**:

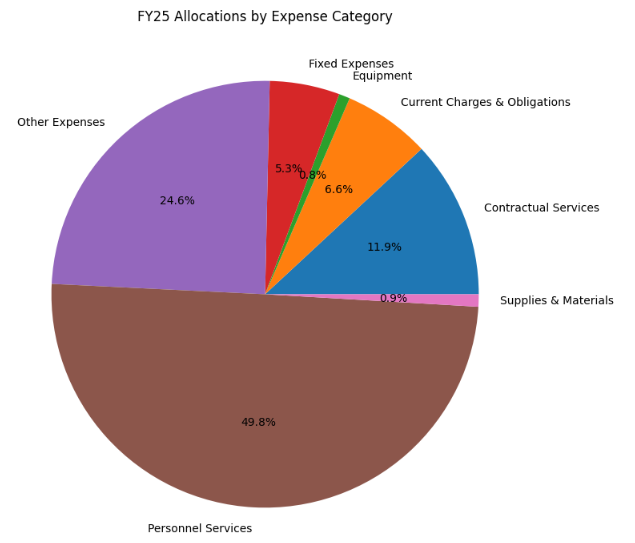
* The total allocation must not exceed the FY25 budget.
* **Reason**: Ensures the budget remains within the allocated resources, respecting the fiscal plan for FY25.

1. **Proportional Allocation Based on FY23 Expenses**:

* Allocations for each program and category are proportional to their FY23 actual expenses.
* **Reason**: Distributes the FY25 budget based on past spending patterns, which reflects the program and category priorities in the previous year.

**Why These Constraints?**

This scenario is designed to distribute the FY25 budget fairly, based on historical spending. The proportional allocation ensures that programs and categories with larger FY23 expenses are allocated more resources in FY25, maintaining continuity and consistency in funding distribution across programs.



**Scenario-6: Total FY25 Budget + Critical Program Funding + Spending Caps for Non-Critical Programs**

**Objective**: Maximize the total allocation of the FY25 budget while ensuring that critical programs are funded adequately and non-critical programs are kept within reasonable spending caps.

**Constraints**:

1. **Total FY25 Budget Constraint**:

* The total allocation must not exceed the FY25 budget.
* **Reason**: Adheres to the approved FY25 fiscal plan, ensuring the allocation remains within the available resources.

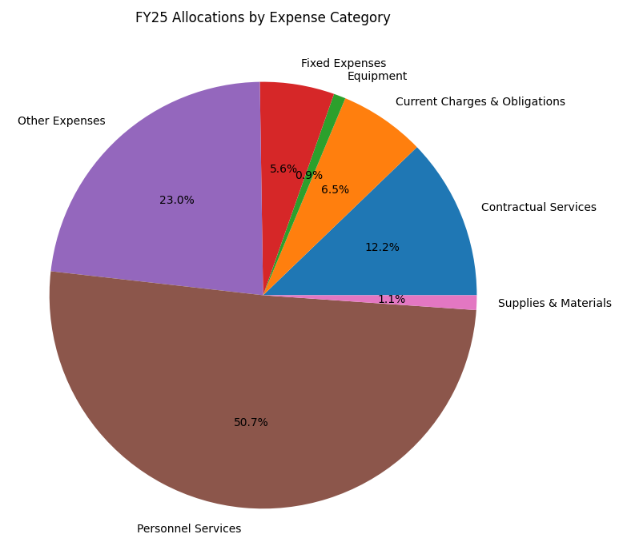
1. **Critical Program Funding (Minimum 70% of FY24 Appropriation)**:

* Critical programs (e.g., Public Safety, Health Services) must receive at least 70% of their FY24 appropriation.
* **Reason**: Ensures that essential services continue to be adequately funded and do not face significant budget cuts that could impact their operations.

1. **Spending Caps for Non-Critical Programs (Max 50% of FY24 Appropriation)**:

* Non-critical programs must receive no more than 50% of their FY24 appropriation.
* **Reason**: Prevents non-essential programs from receiving too much funding, allowing more resources to be directed to critical programs.

**Why These Constraints?** This scenario prioritizes critical programs by ensuring they receive a minimum level of funding while imposing stricter caps on non-critical programs to ensure fiscal discipline. The combination of these constraints ensures that limited resources are allocated where they are most needed without exceeding the overall FY25 budget.



**QUESTION & ANSWERS**

**1. How does your optimization approach balance equity and efficiency in budget allocation?**

**Purpose:** To assess your understanding of trade-offs between resource fairness and maximizing allocation objectives.

**Answer:** Our approach uses constraints like proportional allocation and category spending caps to ensure equitable distribution, while maximizing the FY25 budget allocation. Efficiency is achieved by prioritizing critical programs and adhering to historical spending patterns, ensuring resource utilization aligns with organizational goals.

**2. How do you address the challenge of interdependencies between expense categories in your optimization model?**

**Purpose:** To explore how well you considered complex financial relationships in your modelling.

**Answer:** We analysed correlations between categories using Pearson correlation matrices to identify interdependencies. While the optimization model does not explicitly include dependency variables, insights from correlation analysis guided constraints to prevent overfunding or underfunding interrelated categories.

**3. What are the implications of using historical spending data as the basis for your constraints?**

**Purpose:** To evaluate your critical thinking on the reliability and biases of historical data.

**Answer:** Historical spending data reflects past priorities but may perpetuate biases or inefficiencies. While useful for establishing realistic caps, it risks overlooking evolving needs or strategic shifts. This limitation underscores the importance of periodically updating constraints and incorporating stakeholder input.

**4. How scalable and generalizable is your optimization model for different organizational contexts?**

**Purpose:** To assess the adaptability of your methodology to varying datasets and requirements.

**Answer:** Our model is scalable due to its modular constraints and parameterized budget allocation process. Generalizability is high, as the methodology can adapt to different fiscal structures by adjusting constraints (e.g., departmental limits, category caps) and optimization objectives.

**5. What trade-offs are involved in choosing linear versus non-linear optimization for your scenarios?**

**Purpose:** To probe the mathematical underpinnings of your model.

**Answer:** Linear optimization ensures computational efficiency and straightforward implementation but oversimplifies relationships, especially for non-linear dependencies or diminishing returns. Non-linear models could capture complex interactions but at the cost of higher computational demand and potential overfitting to specific datasets.

**6. How do your visualizations enhance interpretability for stakeholders, particularly non-technical users?**

**Purpose:** To understand your approach to bridging technical analysis and actionable insights.

**Answer:** Interactive HTML templates allow stakeholders to explore trends and scenarios intuitively. For instance, tree maps and bar charts provide clear visual hierarchies and comparisons, enabling non-technical users to grasp budget allocations without delving into raw data or mathematical models.

**7. What justifies your choice of constraints across different scenarios? How did you validate them?**

**Purpose:** To assess your justification for constraint design and validation approach.

**Answer:** Constraints like departmental spending caps and proportional allocations were based on historical spending and organizational policies. Validation involved analysing their impact on outcomes through scenario testing, ensuring realistic and balanced allocations that meet fiscal and strategic goals.

**8. How could external factors like inflation or policy changes impact your optimization results?**

**Purpose:** To evaluate your model’s robustness against real-world uncertainties.

**Answer:** External factors could significantly alter spending priorities and available resources. To mitigate this, the model can integrate adjustable parameters for inflation or contingency reserves. Incorporating predictive models for policy and economic trends would further enhance robustness.

**9. What advanced techniques could be used to refine your optimization model, and why weren’t they included?**

**Purpose:** To probe your knowledge of cutting-edge methods and project scope limitations.

**Answer:** Advanced techniques like multi-objective optimization or stochastic programming could account for uncertainty and competing priorities. These were not included due to the complexity of implementation and focus on developing a baseline model with interpretable constraints.

**10. How do you ensure the ethical use of data in your analysis and optimization process?**

**Purpose:** To understand your awareness of ethical considerations in data-driven projects.

**Answer:** We ensured ethical use by relying on publicly available data and avoiding manipulations that could misrepresent trends. Decisions were guided by fairness principles, like equitable resource distribution, and transparent documentation of assumptions to promote accountability.

**CONCLUSION**

This project successfully analyzed and optimized the Boston Operating Budget using advanced statistical methods and optimization techniques. By leveraging tools like MySQL, Python, OR-Tools, and PuLP, we efficiently stored, processed, and modeled budget data to reveal financial trends and propose equitable resource allocations. Visualization techniques, implemented through HTML templates, allowed stakeholders to intuitively explore the results and gain actionable insights. The project highlights the potential of data-driven approaches in fiscal planning, enabling better decision-making and promoting transparency, fairness, and fiscal discipline.