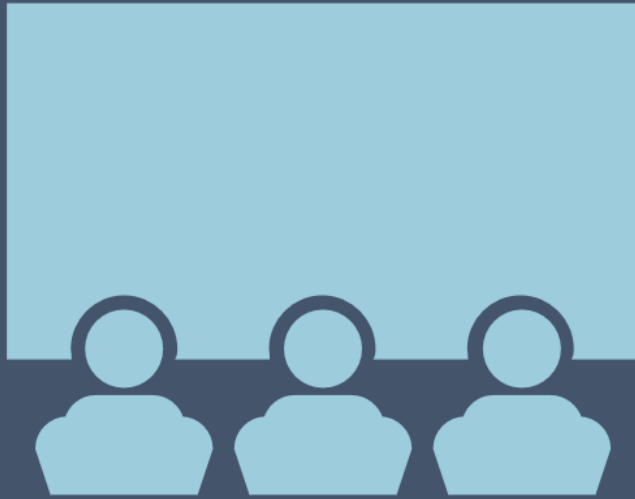


Job Report Analysis

SUJITHRA E

19.08.2021



Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- This report shows the detailed job analysis conducted to determine the cost allotted for each job or task assigned in the company. Recommendations are also included in the report.



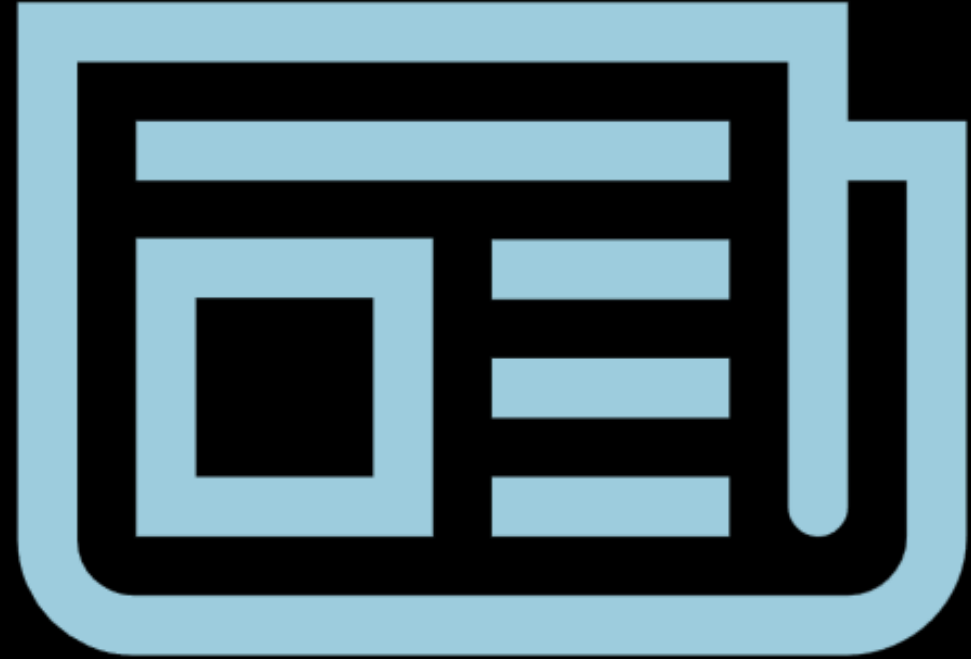
Introduction



- The report centralized on the study on varying tasks which corresponded to the execution of the costs and strategies from the third quarter of 2025.

Methodology

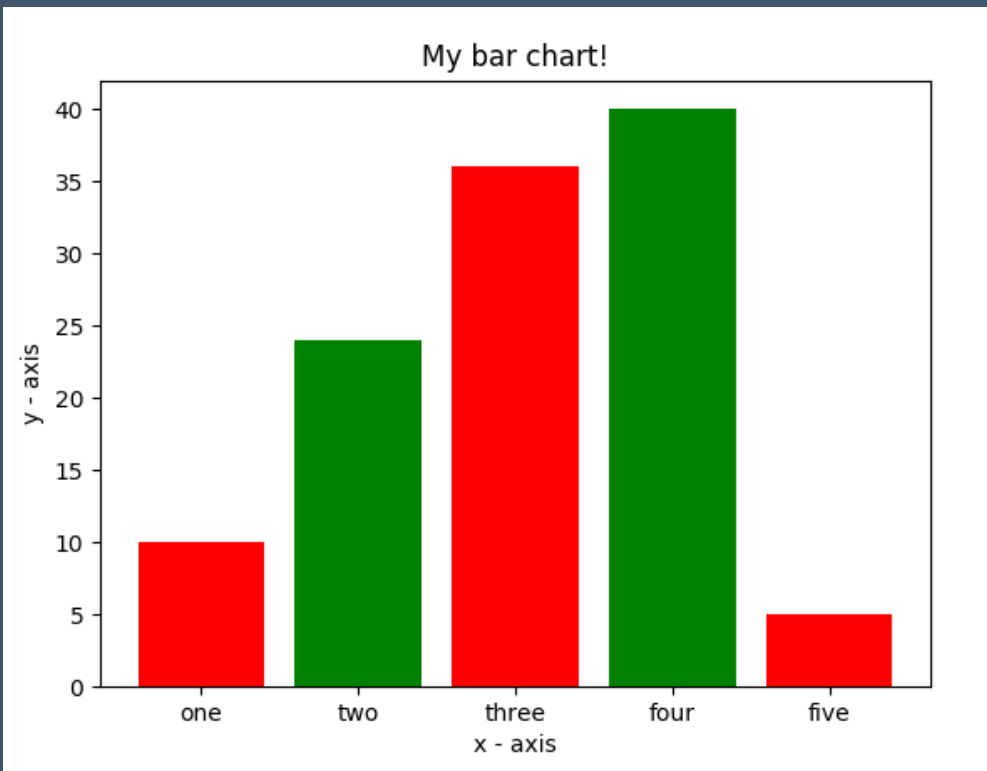
- Data collection methodology:
 - Describe how data were collected
- Perform data wrangling
 - Describe how data were processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - How to build, tune, evaluate classification models



Data Collection

- For job analysis there is the need of data and it can be collected by different methods. They are the following:
- **Company records**
- **Personal interview**
- **Observation method**
- **Diary or log of job incumbent**
- **Recording**

Data collection SpaceX api



The github url:

[https://github.com/sujithra1997/coursera-projects/blob/main/Applied%20ata%20science%20capstone/week1/datacollection.ipynb](https://github.com/sujithra1997/coursera-projects/blob/main/Applied%20data%20science%20capstone/week1/datacollection.ipynb)

Data collection – Web scraping

- The git url :
- <https://github.com/sujithra1997/coursera-projects/blob/main/Applied%20data%20science%20capstone/week1/datawrangling.ipynb>

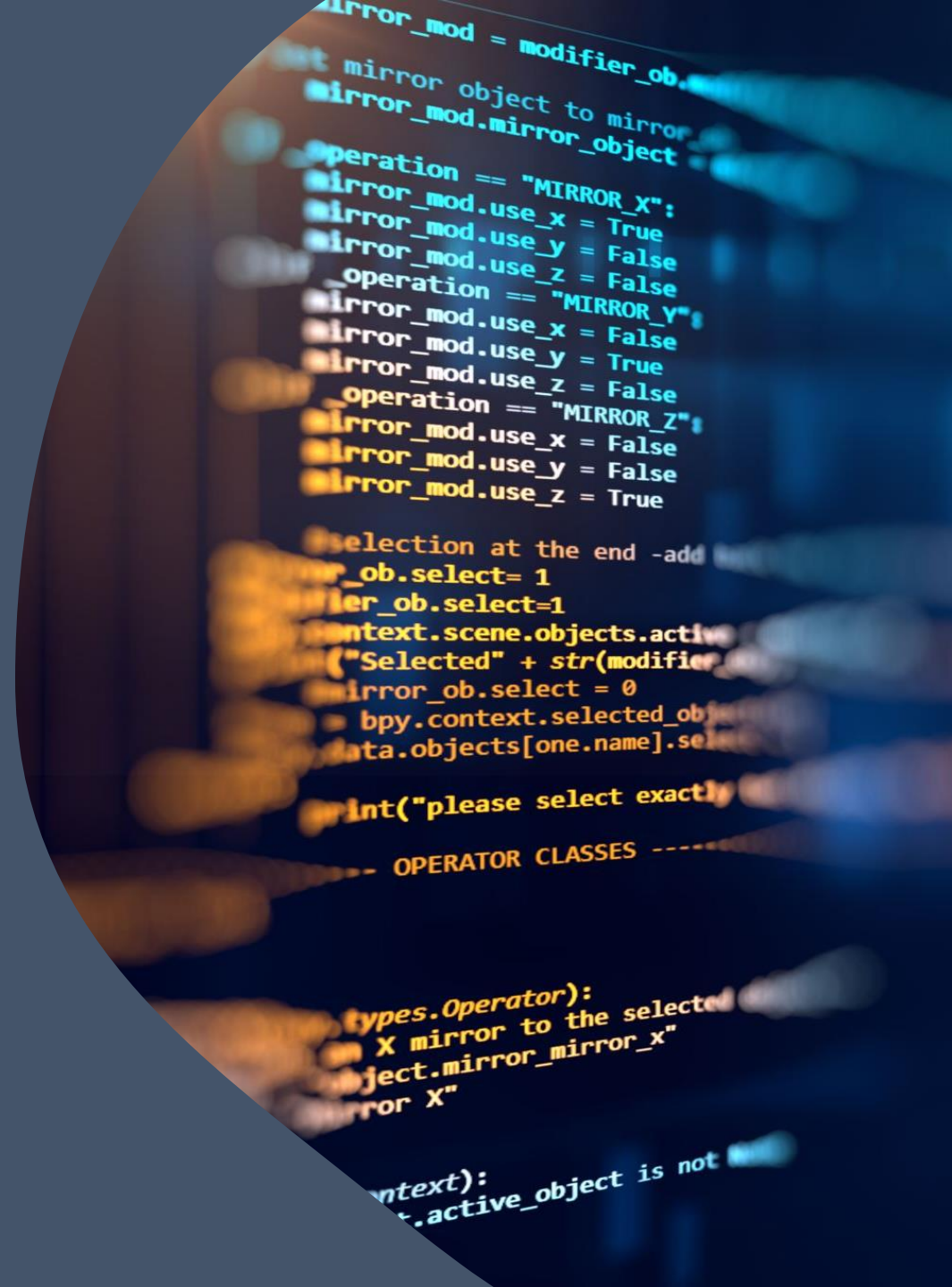
EDA with data visualization

- We can view the job vacancy analysis using data visualization in real time process.
- https://github.com/sujithra1997/coursera-projects/blob/main/Applied%20data%20science%20capstone/week2/exploratory_analysis.ipynb



EDA with SQL

- We can use SQL queries to count the available job vacancy as well as max,min number of job vacancy per day or year wise.
- <https://github.com/sujithra1997/coursera-projects/blob/main/Applied%20data%20science%20capstone/week2/EDA-SQL.ipynb>



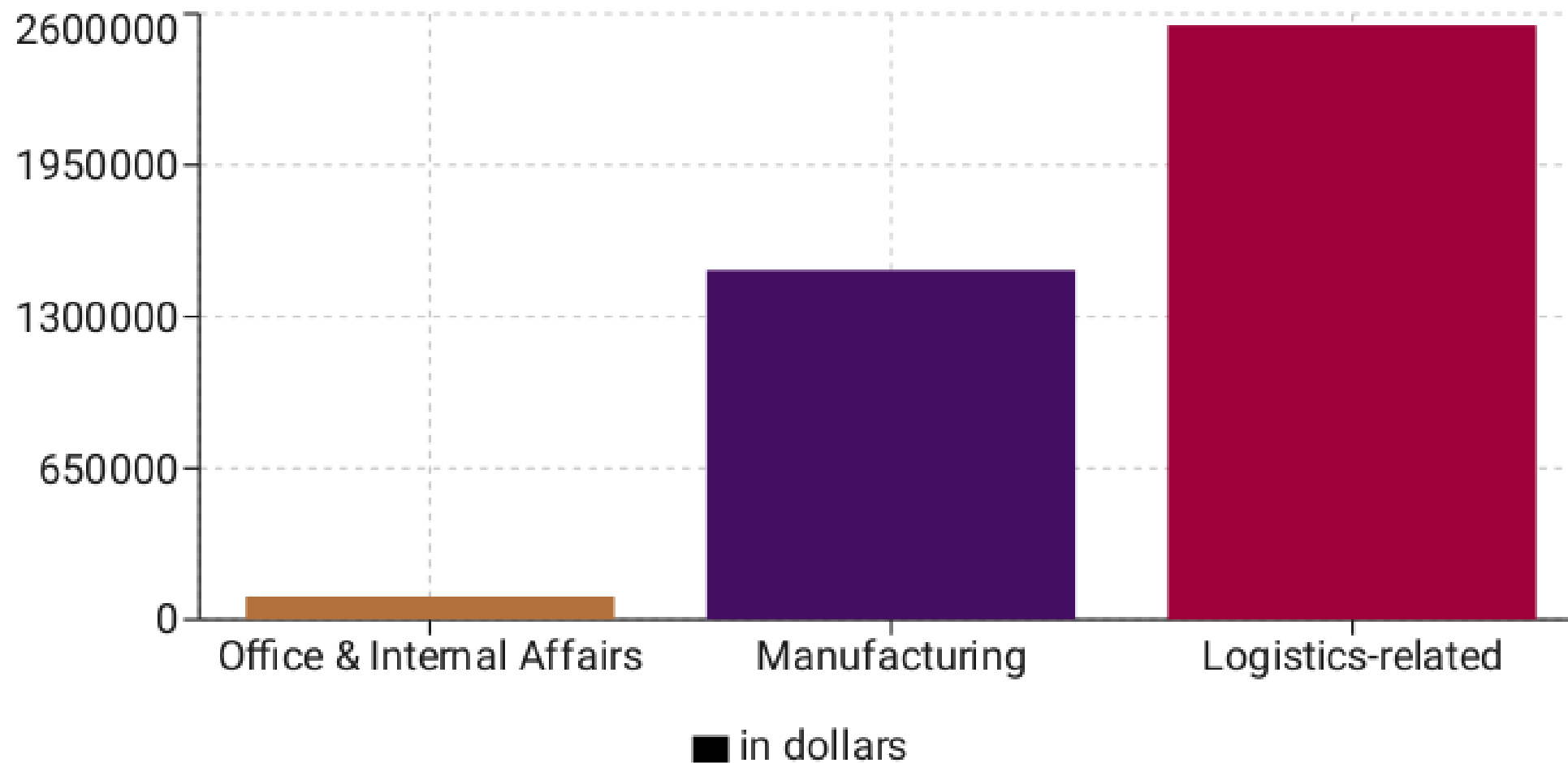
Predictive Analysis Classification

COSTS

COSTS	PREVIOUS COSTS	NEW COSTS
Office & Internal Affairs Tasks	\$50,000.00	\$100,000.00
Manufacturing Tasks	\$1,000,000.00	\$1,500,000.00

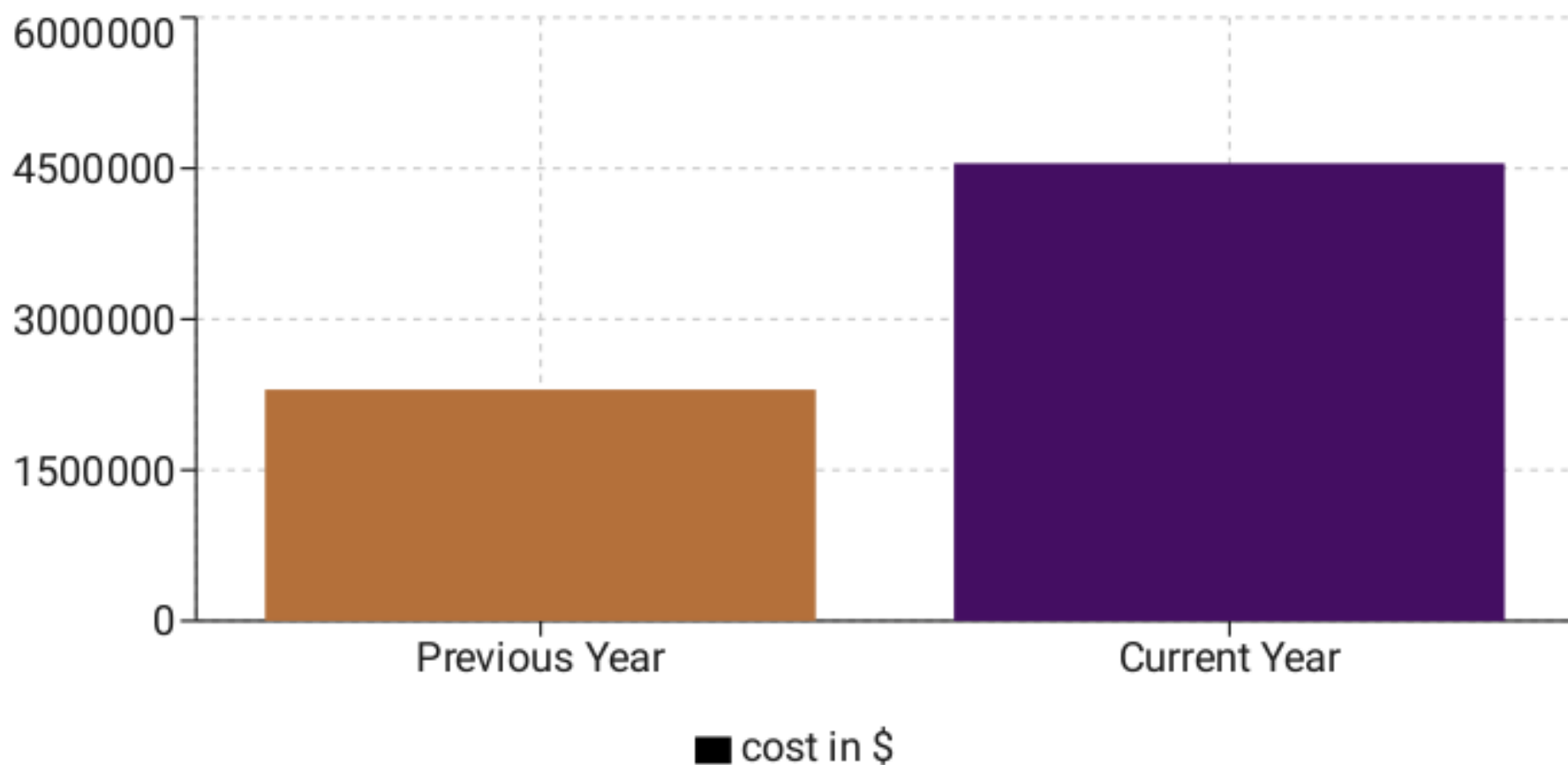
Finding 1

Based on the study and evaluation, logistics-related tasks consume more of the company's budget with \$2,550,000.00 in total as compared to the office and internal affairs-task and manufacturing tasks.



Finding 2

Based on the study and evaluation, the overall costs of the current year are higher than the previous year, amounting to \$4,550,000.00 as opposed to last year's \$2,300,000.00.



The background is a complex, abstract digital visualization. It features a series of concentric circles and arcs that create a sense of depth and rotation. The color palette is primarily blue and teal, with some white and light gray elements. The patterns within the circles resemble data visualizations, such as bar charts or histograms, where the height of the bars varies across the segments. The overall effect is a high-tech, futuristic aesthetic that suggests data analysis and visualization.

EDA with Visualization

Flight Number vs Launch site

All

Search

Upgrade

🔍

🔔

sujithra E's Account

science captstone / expdv

✎

🔒

🔗

📁

⌵

⬇️

🔔

🕒

💬

TASK 1: Visualize the relationship between Flight Number and Launch Site

Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`, set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

```
5]: # Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number", fontsize=20)
plt.ylabel("Launch Site", fontsize=20)
plt.show()
```

Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

Payload mass vs launchsite

All

Search

Upgrade

sujithra E's Account

science captstone / expdv

Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

TASK 2: Visualize the relationship between Payload and Launch Site

We also want to observe if there is any relationship between launch sites and their payload mass.

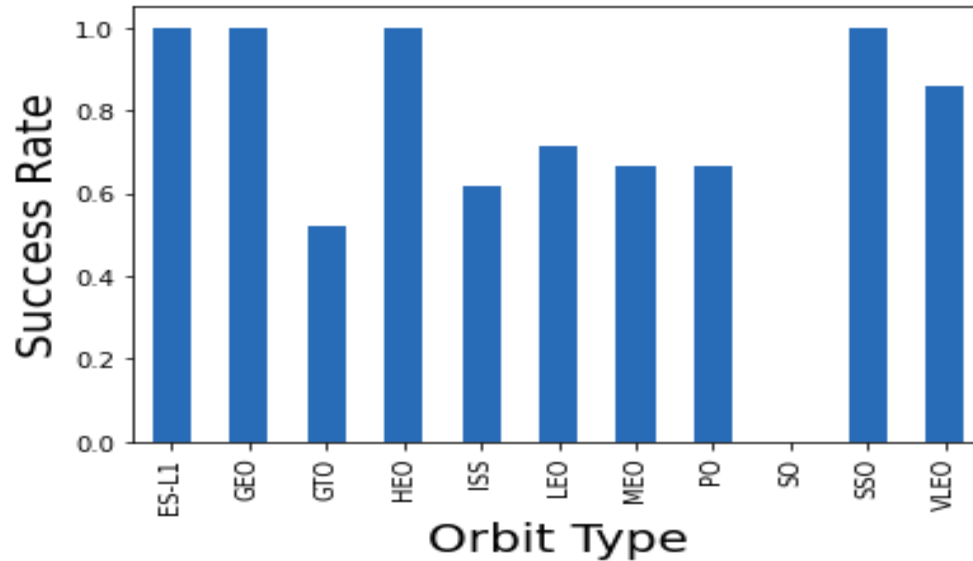
```
[8]: # Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```

Now try to explain any patterns you found in the Payload Vs. Launch Site scatter point chart.

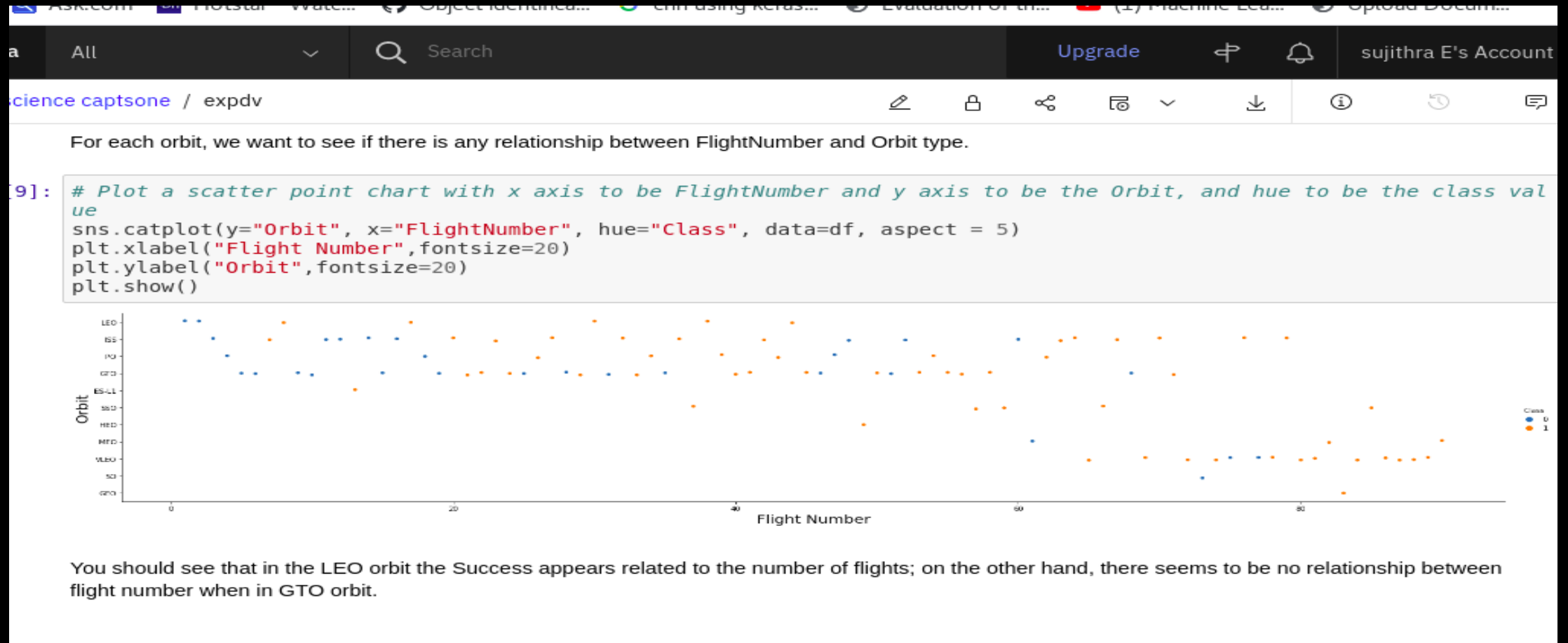
Success rate vs Orbit type

Let's create a bar chart for the success rate of each orbit

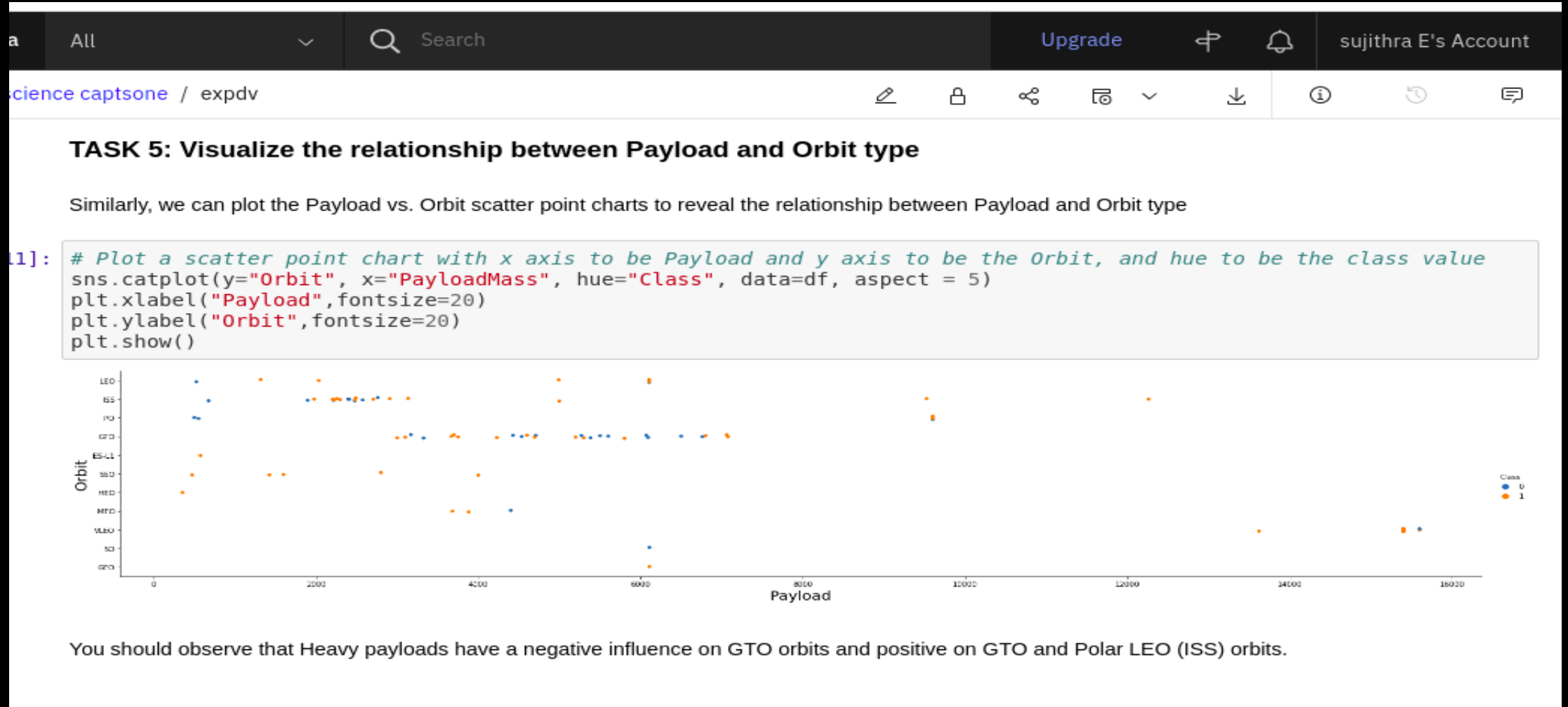
```
In [7]: # HINT use groupby method on Orbit column and get the mean of Class column
df.groupby(['Orbit']).mean()['Class'].plot(kind='bar')
plt.xlabel("Orbit Type", fontsize=20)
plt.ylabel("Success Rate", fontsize=20)
plt.show()
```



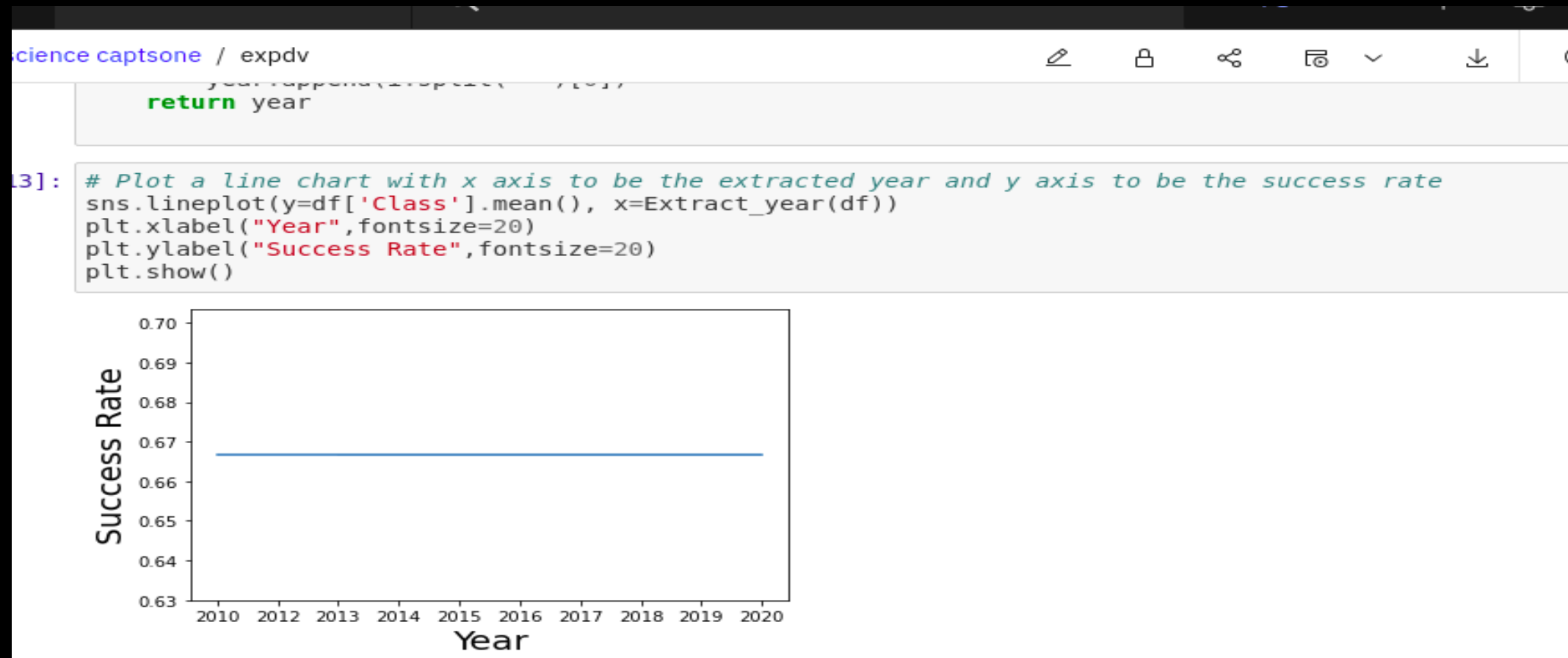
Flight Number vs Orbit Type



Payload vs Orbit Type



Launch Success yearly Trend



Conclusion

- Although the costs for the jobs or tasks increased in the current year, the budgets were necessary for the improvement of the company's services, both for the clients and internally.

Appendix



It is suggested that the company conduct a quarterly job analysis.

With the increased costs, the company should acquire more investors, at least 5, for the next year.