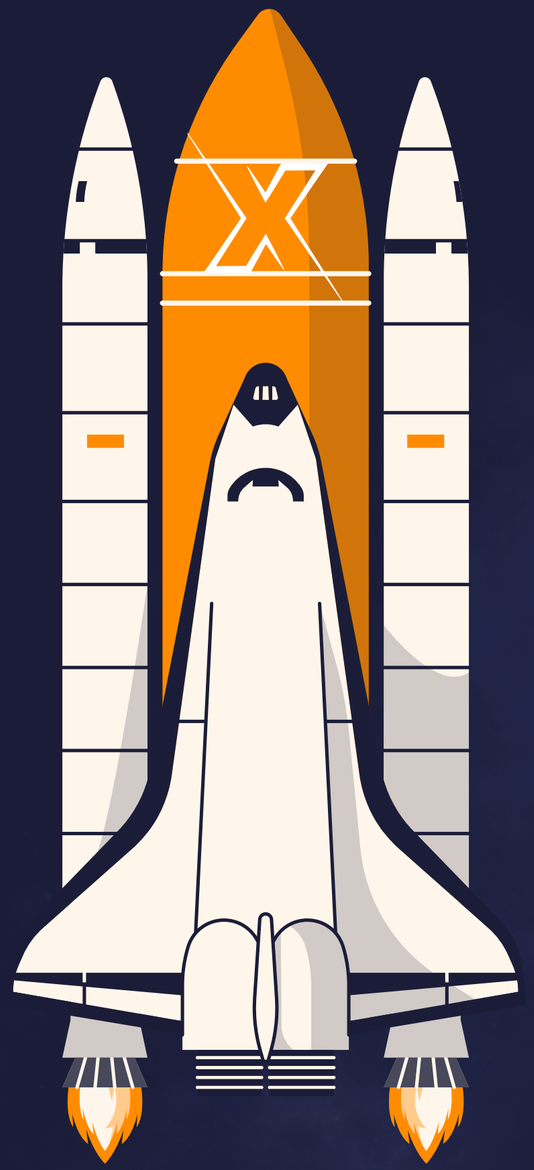


SpaceX Falcon 9 First stage Landing Prediction

Hary Yusuf

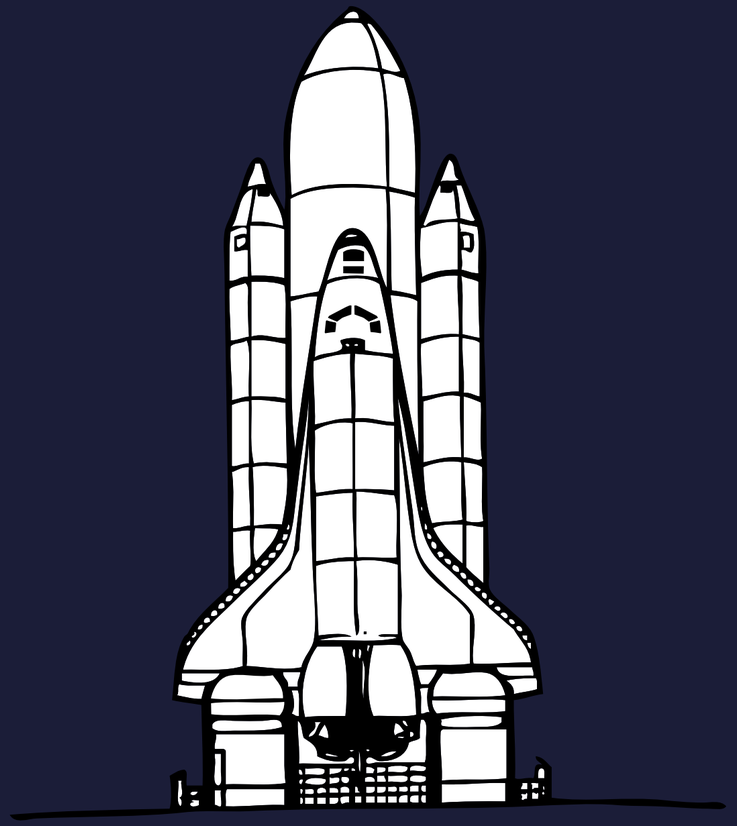
June 24, 2024

Data Science Capstone Project

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

Executive Summary

We collect SpaceX data using **APIs** and **web scraping**. Then, we apply **exploratory data analysis** to explore and visualize the data. Finally, we use the **Grid Search** method to **identify** the **best machine learning model** for **predicting** the next **landing's classification.**



Introduction

SpaceX's ability to reuse the first stage of its Falcon 9 rocket cuts launch costs to 62 million dollars, while other providers may charge up to 165 million dollars. If we can predict whether the first stage will land successfully, we can estimate the launch costs, which is useful for companies competing with SpaceX.

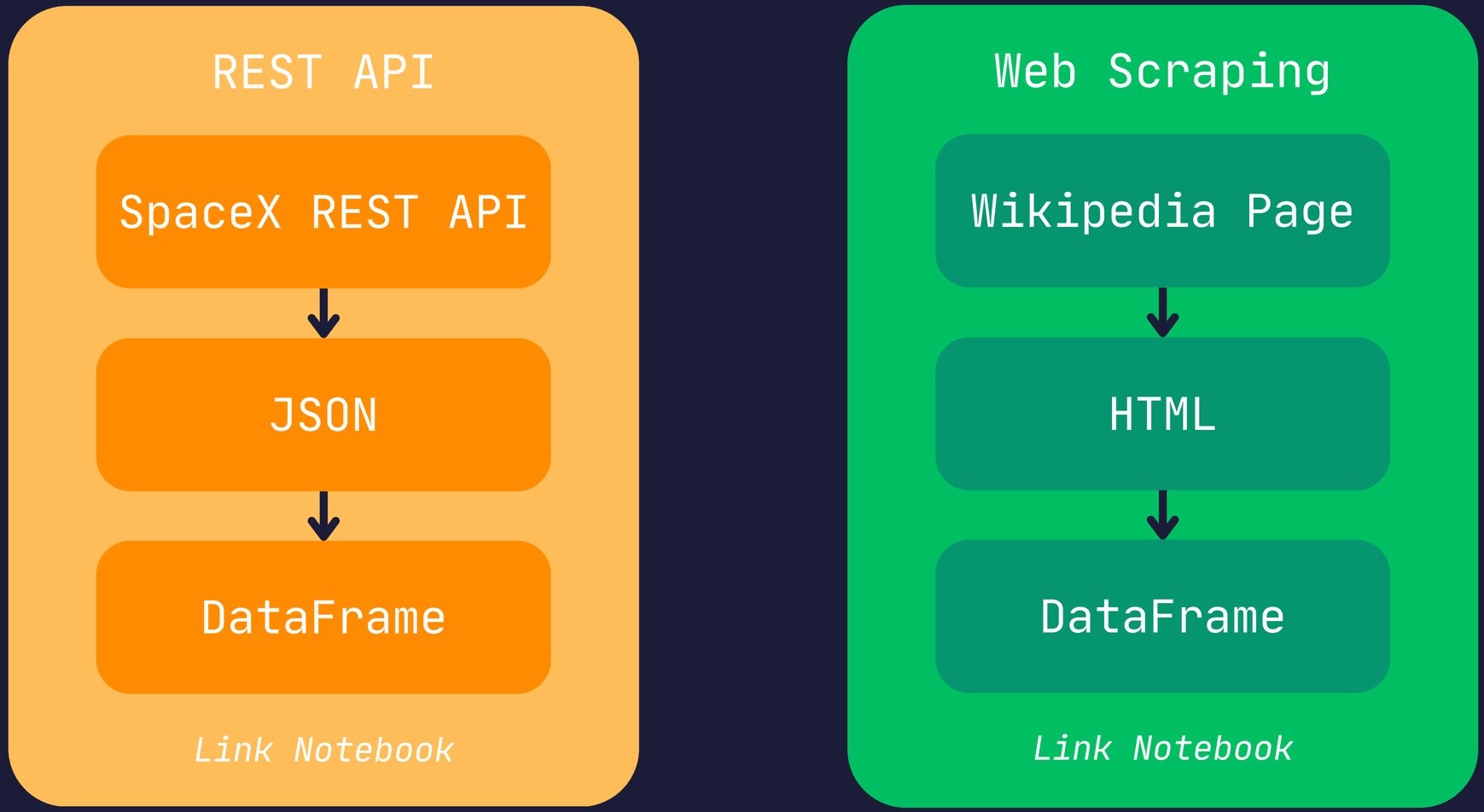
Key Questions:

- Can we predict the success of the first stage landing for new launches based on historical data?
- What factors give the best chance for a successful launch?

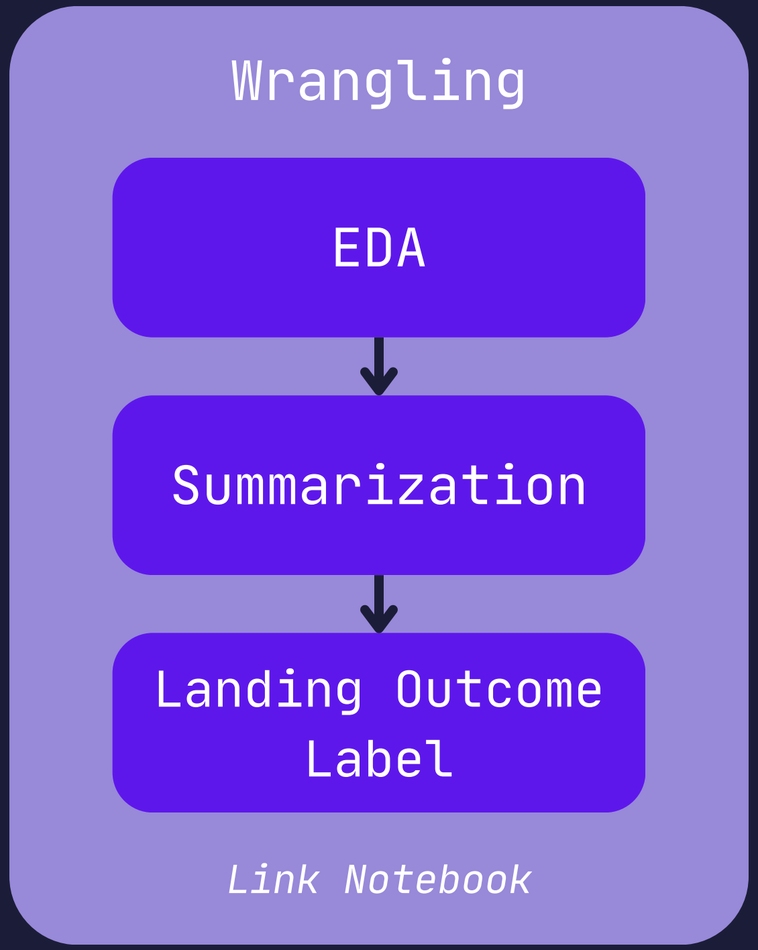
Methodology

- Data collection:
  - SpaceX REST **API**
  - **Web Scraping** (Wikipedia)
- Data wrangling
  - **Generate landing class** from outcome column
- Exploratory Data Analysis (EDA) using **visualization** and **SQL**
- Interactive visual analytics using **Folium** and **Plotly Dash**
- Predictive analysis using **classification models**:
  - Using **GridSearch** to find best fit model

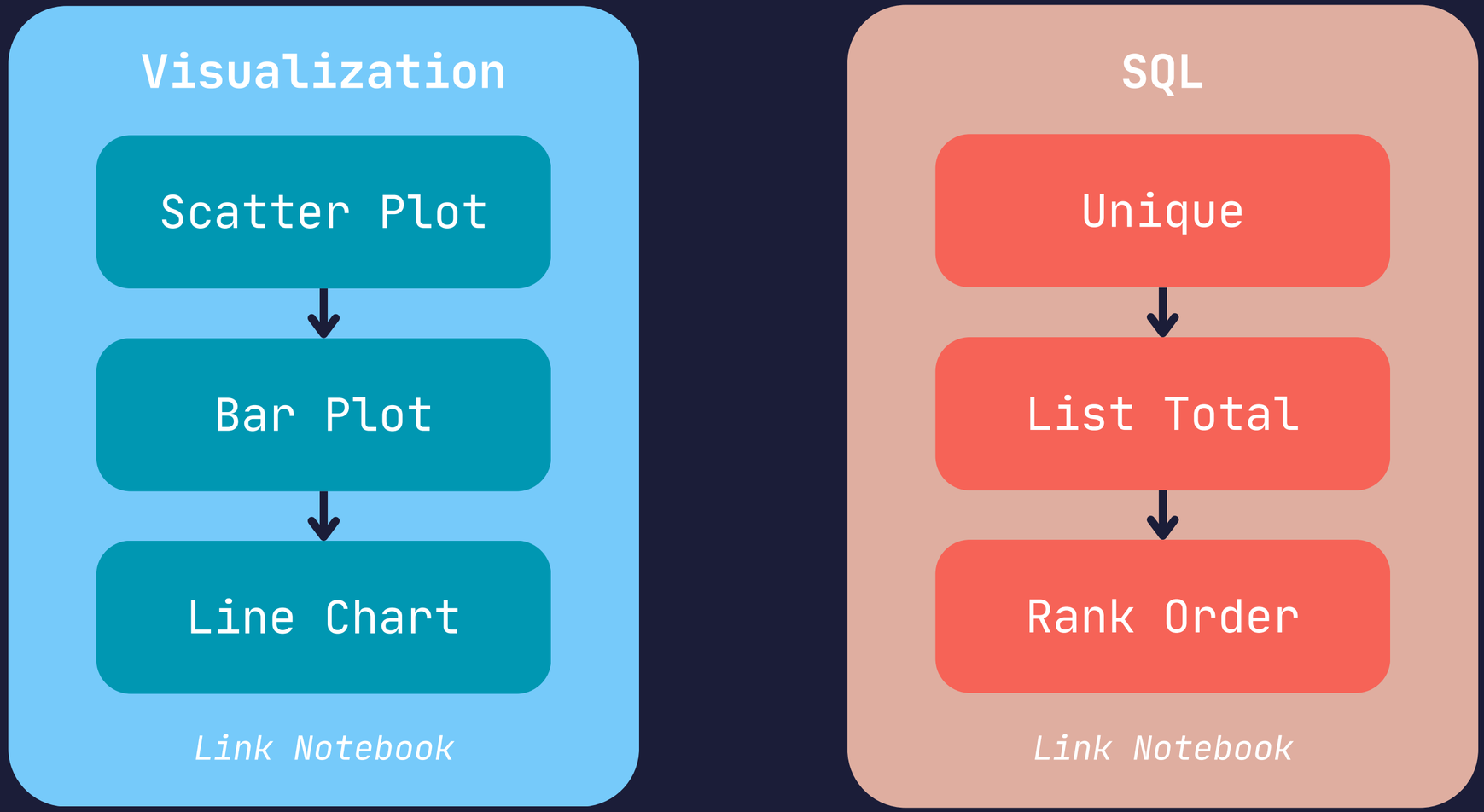
Data Collection



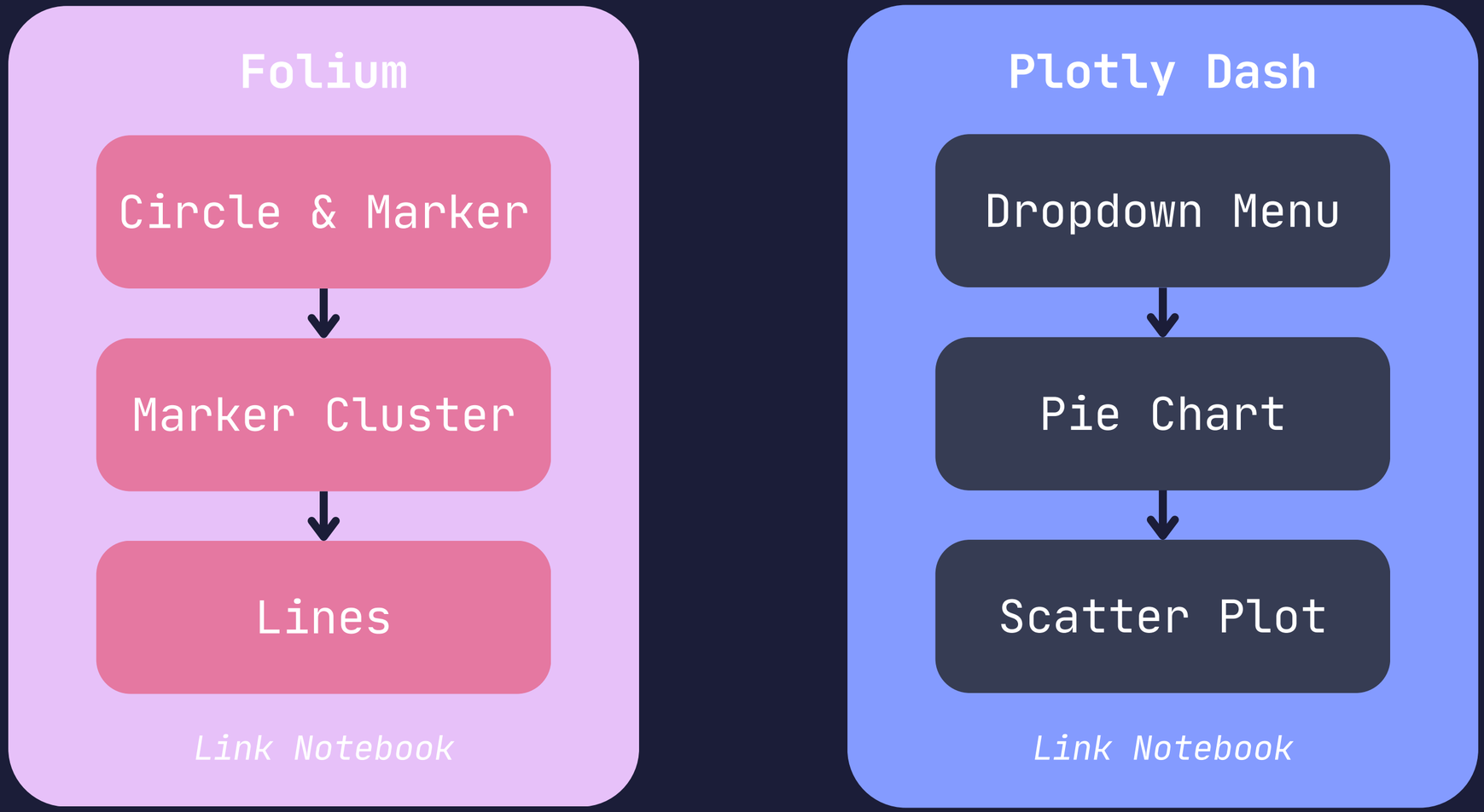
Data Wrangling



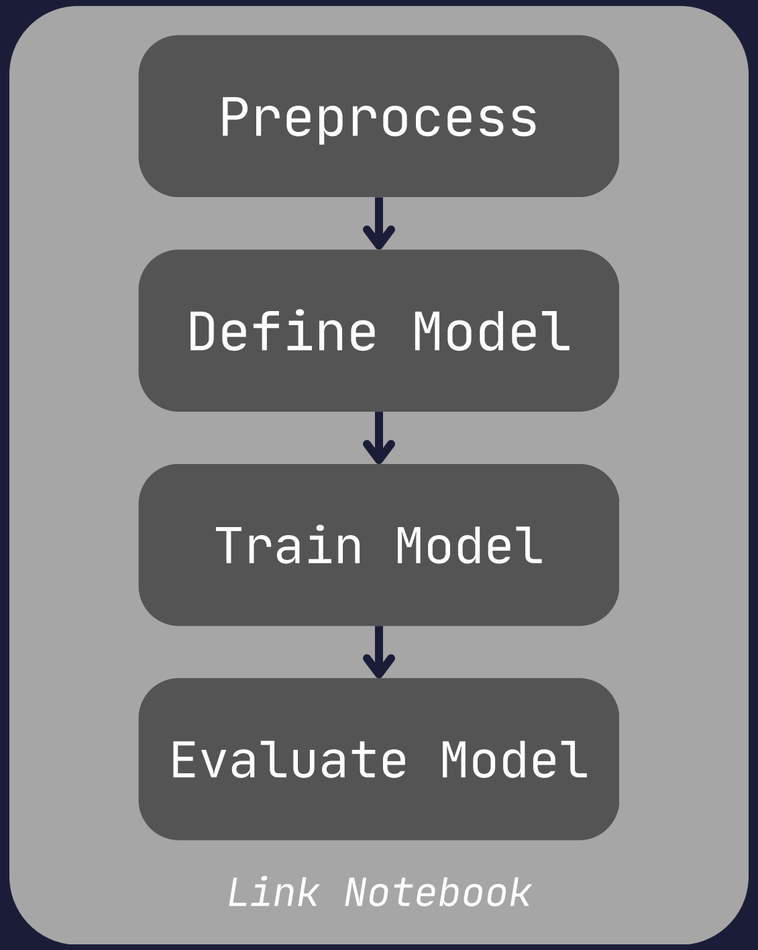
Exploratory Data Analysis



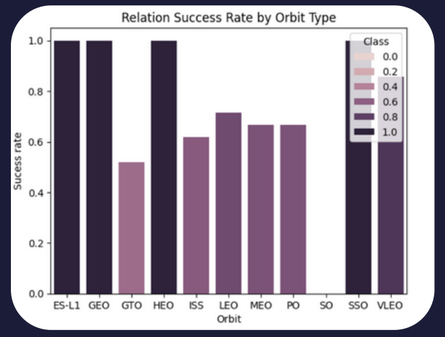
Interactive Map

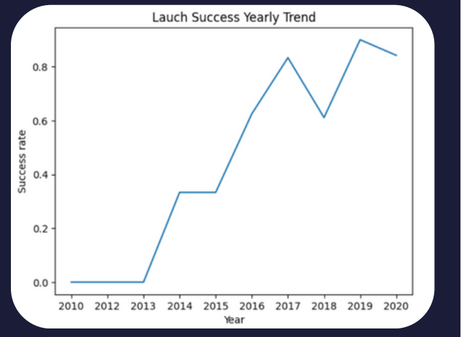


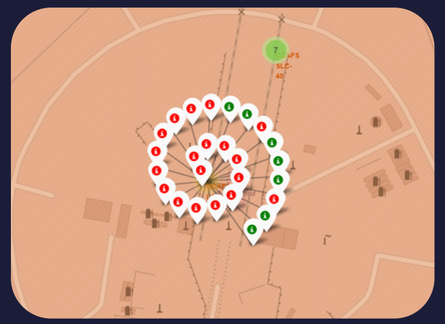
Predictive Analysis



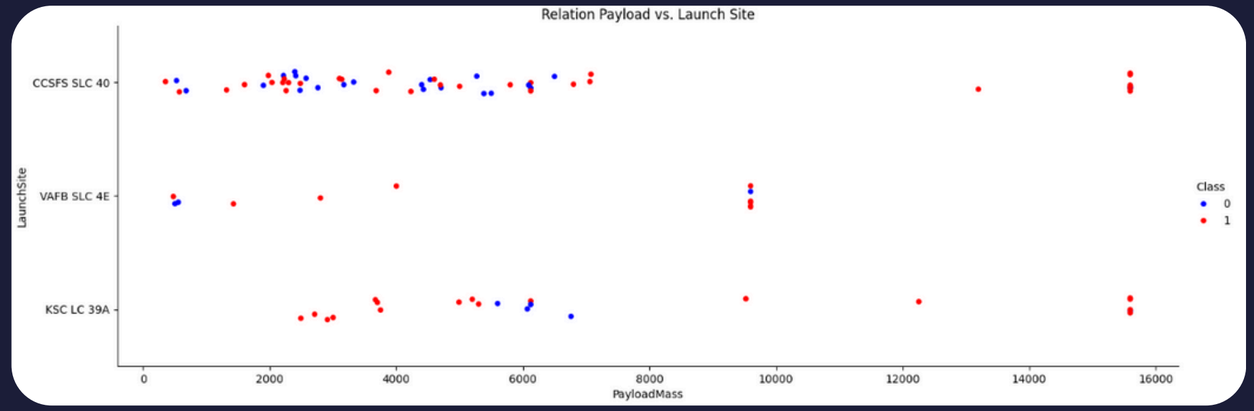
Result

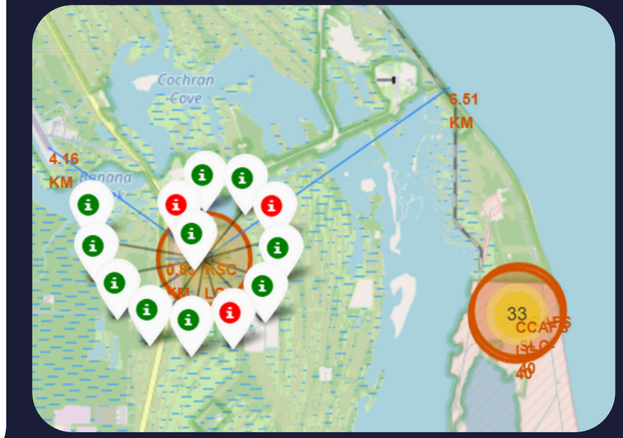






Booster_Version
varchar
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7
12 rows





	train accuracy	test accuracy
lr	0.777778	0.944444
svm	0.844444	0.944444
tree	0.855556	0.944444
knn	0.766667	0.944444

Conclusion

The dataset has 90 rows and 83 columns. By splitting it 80/20, we get 72 rows for training and 18 for testing.

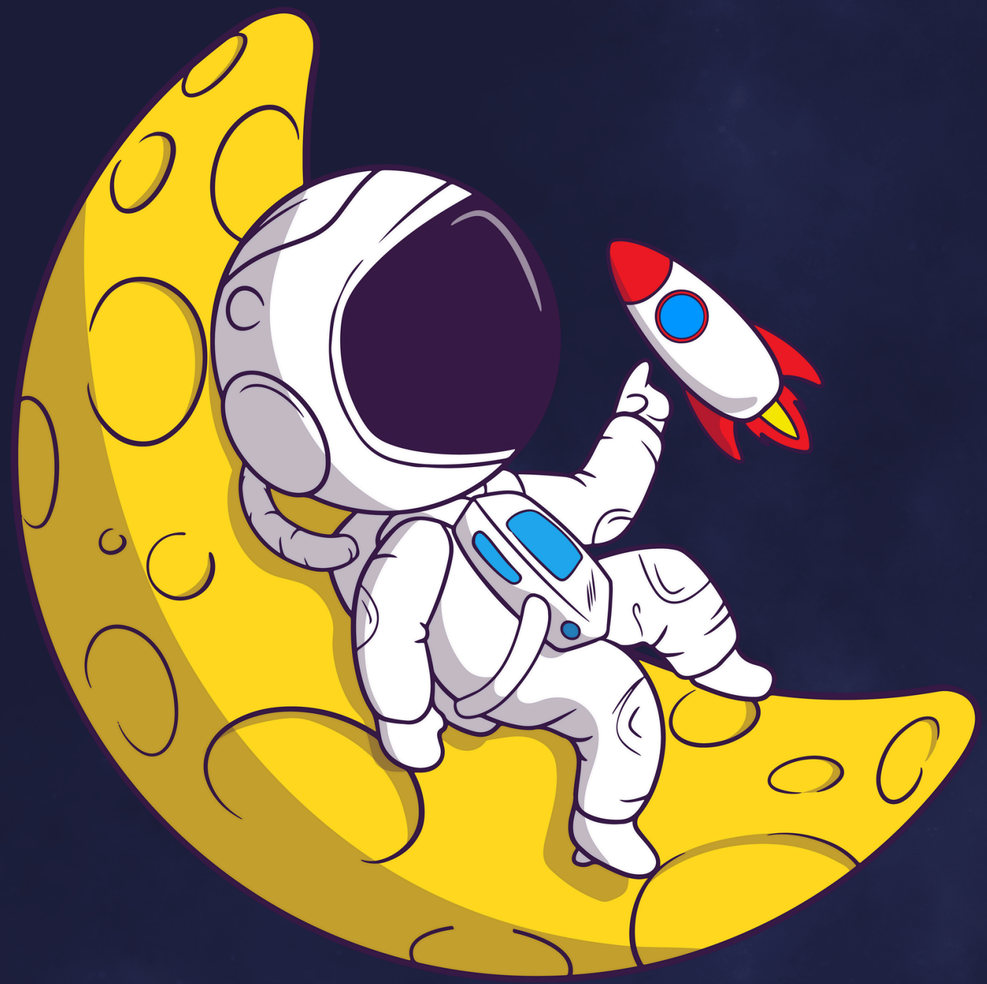
Key insights:

- The best launch site is **KSC LC-39A**.
- Orbits **ES-L1**, **GEO**, **HEO**, and **SSO** have a **perfect success rate**.
- The **Decision Tree** model is the **most effective** for this dataset.

Appendix

Link

- *GitHub Repository*
- *IBM Data Science Professional Certificate*



Thank You