SpaceX Falcon9 Landing Prediction

A Data Science Capstone Project with IBM Course

Amanpreet Kaur

Sep 06, 2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Introduction
- Data Collection and Wrangling
- Exploratory Data Analysis
- Predictive Analysis
- EDA results
  - With matplotlib and seaborn
  - With SQL
  - With Folium
- Dashboard with Plotly
- Appendix

INTRODUCTION

- This capstone project's main aim is to predict whether a falcon9 launch will be successful
- Falcon9 launched 260 times since 2010 and the data used here are publicly available on wikipedia.
- A total of 14 features were used for this work including payload mass, orbit, location

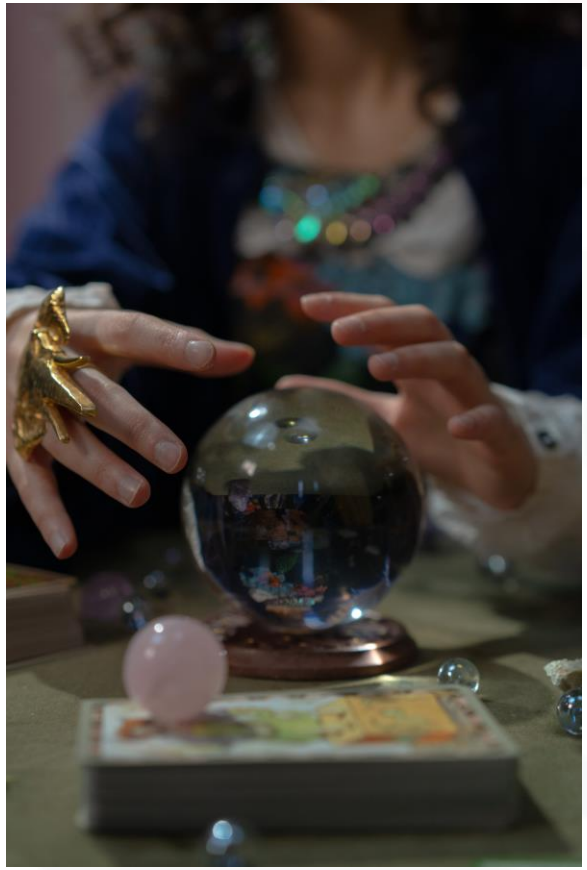
Step 1: Data Collection and Wrangling

- Webscraped to download the relevant data from Wikipedia (link) using python library, beautifulsoup
- Performed Exploratory Data Analysis to:
  - Find patterns in the data
  - Remove Null Values and clean data for the next step
  - Create labels for the training data (landed or failed)

Step 2: Exploratory Data Analysis

- The cleaned data were explored visually using python packages like seaborn, matplotlib to
  - find patterns among features
  - Create labels to determine whether the landing of falcon9 was a success or a failure

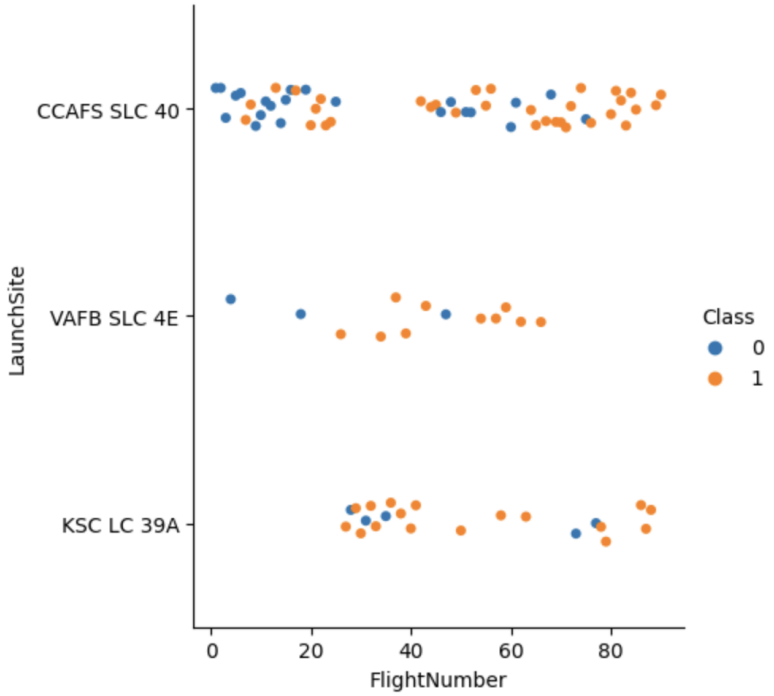
Step 3: Predicitive Analysis

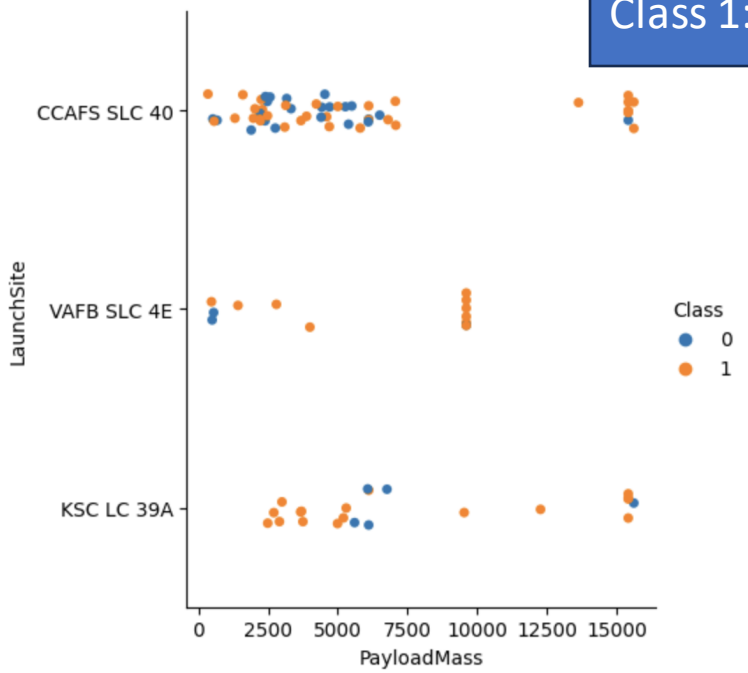


- Performed Feature Engineering on all the features
- The categorical variables were converted to numerical ones using one-hot encoding method
- Utilized Decision Tree, SVM, KNN and Logistic Regression for the prediction for this work.

Exploratory Data Analysis Results-1

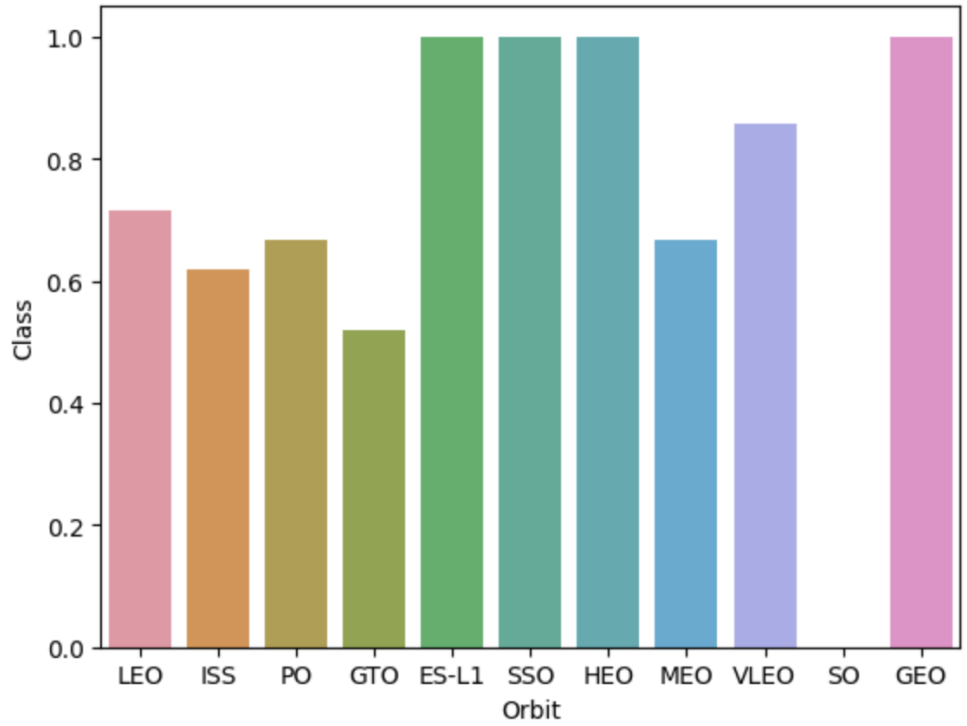
Class 0: Failed to land
Class 1: Successfully landed

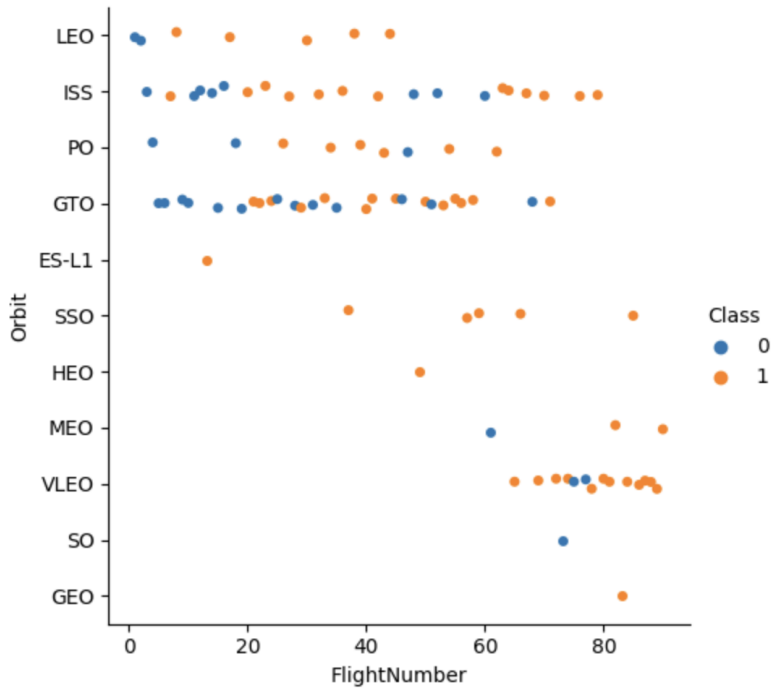




- Flight number has no association the launch Site
- Payloads with weight (5000-7500) seem to fail more
- SLC 4E has been the most successful.

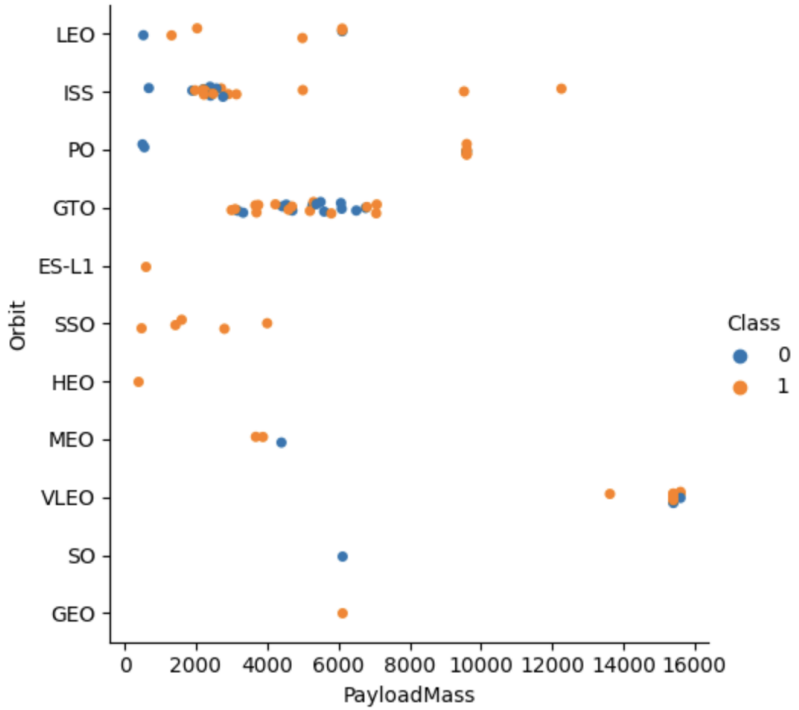
Exploratory Data Analysis Results-2

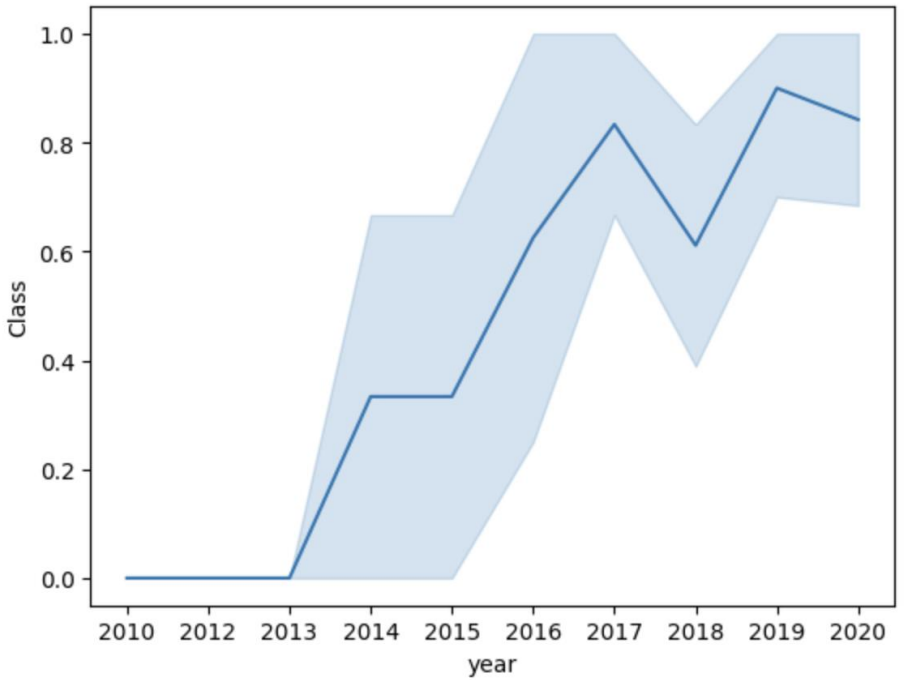




- The type of orbits contribute to a success or failure; e.g., SO orbit returned falcon has always failed
- Flight number and orbit have no correlation

Exploratory Data Analysis Results -3s





- The success rate of landing has gone up over the last 10 years
- Very light payloads failed and particularly the GTP, PO, and ISS ones failed more than others

EDA with SQL

A few insights:

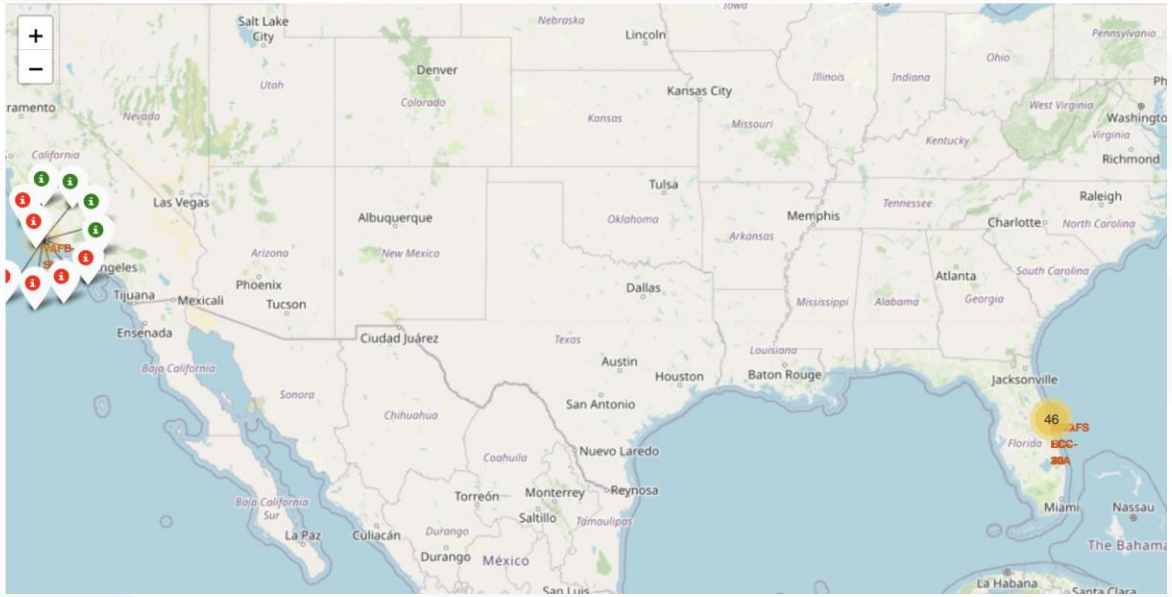
- ❑ Launch Sites: 4
- ❑ First successful landing: 2015-12-22
- ❑ Total missions: 101 (for this project)

More detailed insights in the notebook here:
https://github.com/amanpkaur/IBM-course/blob/main/DS_capstone/EDA_overview/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with Folium

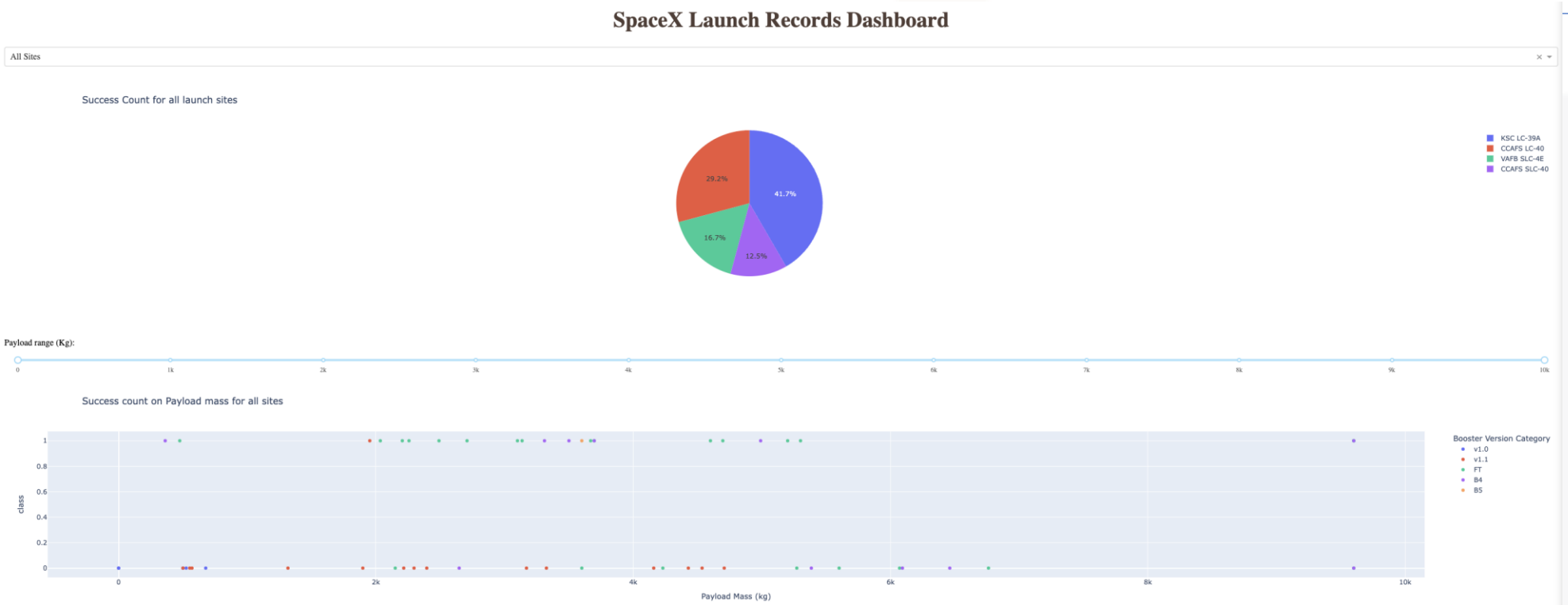
An example EDA plot using Folium:

All the launch sites are marked with success and failure landings marked with green and red colors, respectively.

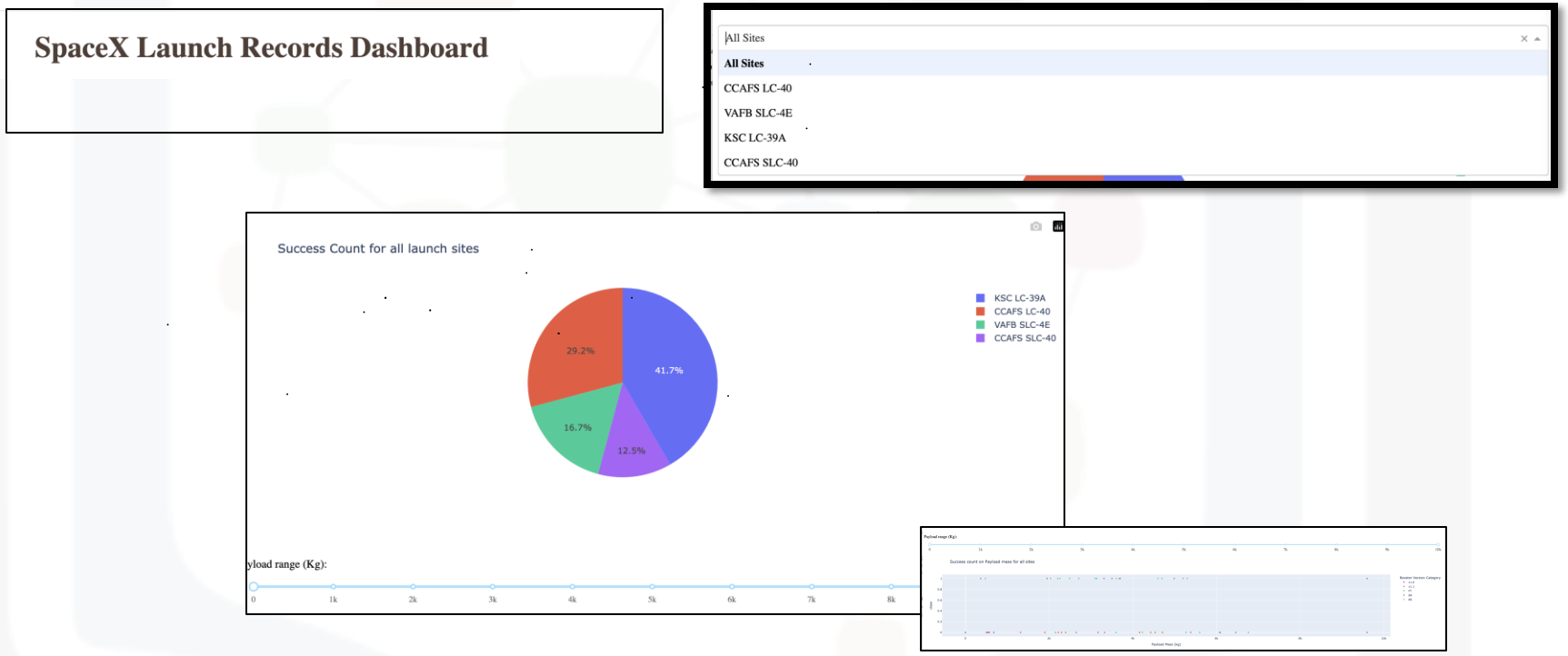


More detailed insights in the notebook here:
https://github.com/amanpkaur/IBM-course/blob/main/DS_capstone/EDA_overview/lab_jupyter_launch_site_location.ipynb

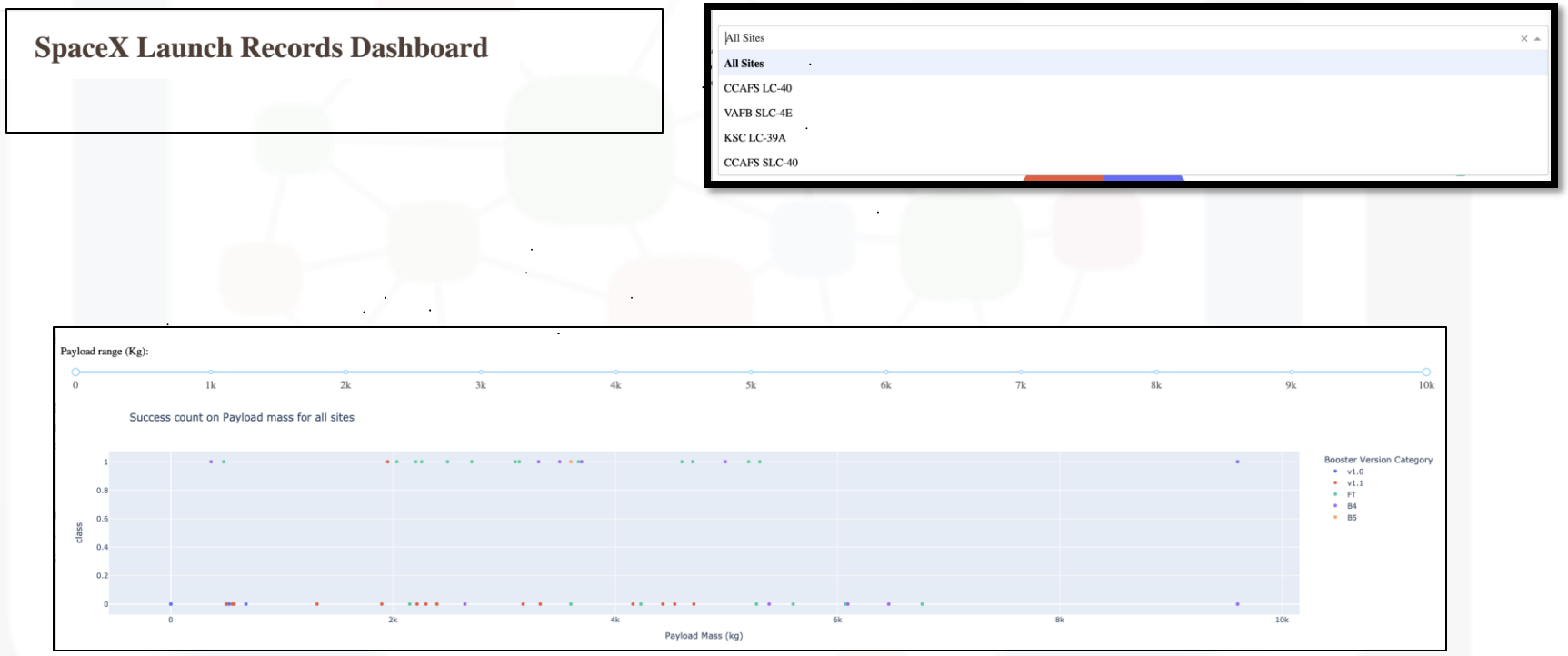
Dashboard with Plotly and Dash (Full Image)



Dashboard with Plotly and Dash (zoomed in-1)



Dashboard with Plotly and Dash (zoomed in-2)



Classification Results

```
df=pd.DataFrame({"method":["Tree","SVM","KNN","LR"],"Score":[accuracy_Tree,accuracy_SVM,accuracy_KNN,accuracy_LR]})
df
#print(f"The method that performs the best is '{df.method[df.Score==max(df.Score)]}'")
```

✓ 0.0s

	method	Score
0	Tree	0.944444
1	SVM	0.833333
2	KNN	0.833333
3	LR	0.833333

Decision Tree returns the highest accuracy among all the models.

More detailed insights are here: https://github.com/amanpkaur/IBM-course/blob/main/DS_capstone/predictive_analysis/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

DISCUSSION and CONCLUSIONS

- Falcon 9 landings have seen more success in the last few years.
- More features like weather, technology improvements could aid in a much better classification
- In this scenario, a simpler decision tree model provided better accuracy than the more complex models! Therefore, simplicity is sufficient to answer complicated questions