IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Collection of data using Web Scraping and API
  - Exploratory Data Analysis (EDA) using data wrangling, data visualization, and the use of a dashboard to display information
  - Machine Learning
- Summary of all results
  - Valuable and interesting data is contained within open sources to the public
  - Through EDA, we can identify the best tools to calculate our predictions
  - Machine Learning allows us to predict the success of a landing

Introduction

- Can SpaceY feasibly compete with SpaceX?
- We can make a prediction based on the estimation of cost for launches and the percent chance of a successful landing of the first stage rocket

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was obtained from SpaceX API and web scraping
- Perform data wrangling
  - Created a landing outcome label to better analyze data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was collected and then normalized. Then the data was split into train and test groups to be evaluated by four different models

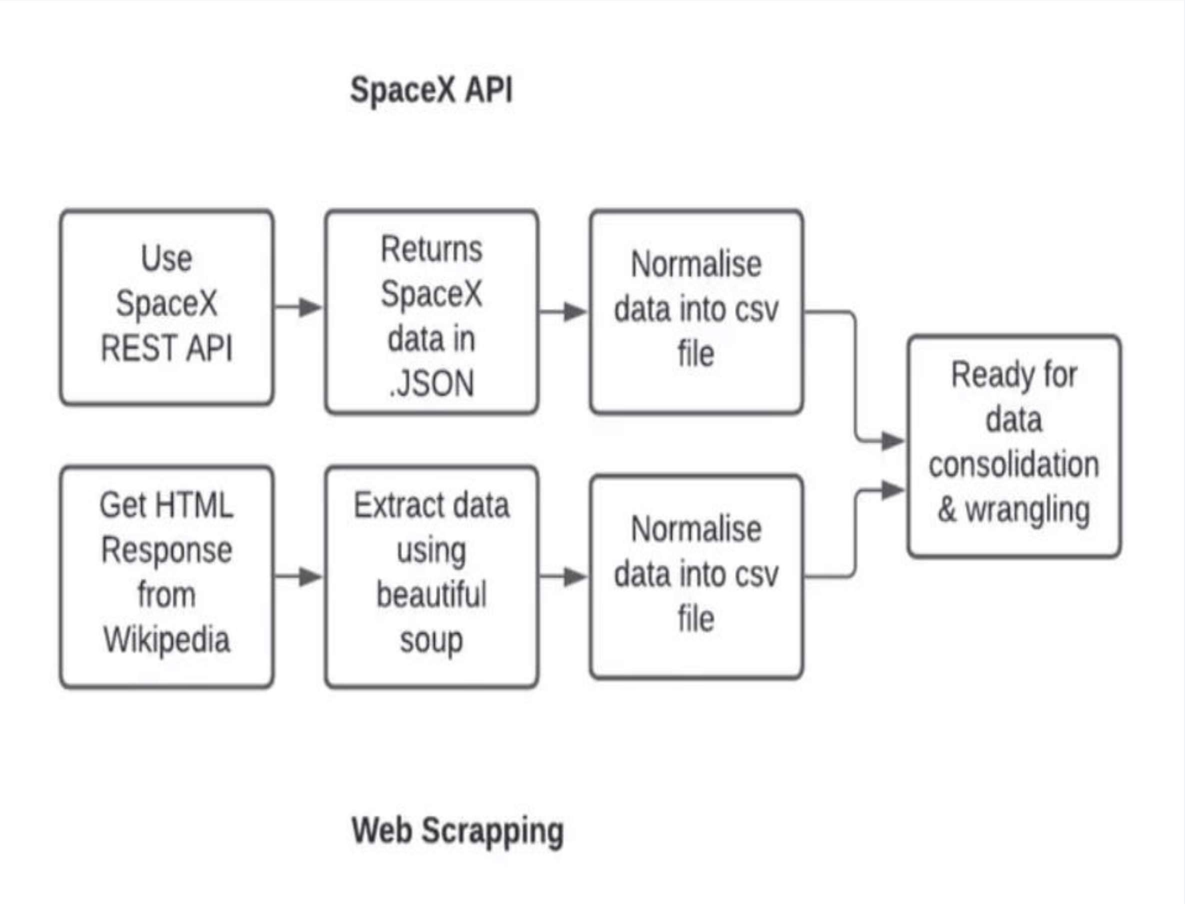
Data Collection

- Data was collected using SpaceX API and web scraping

Data Collection – SpaceX API

- Was collected from a public API through SpaceX

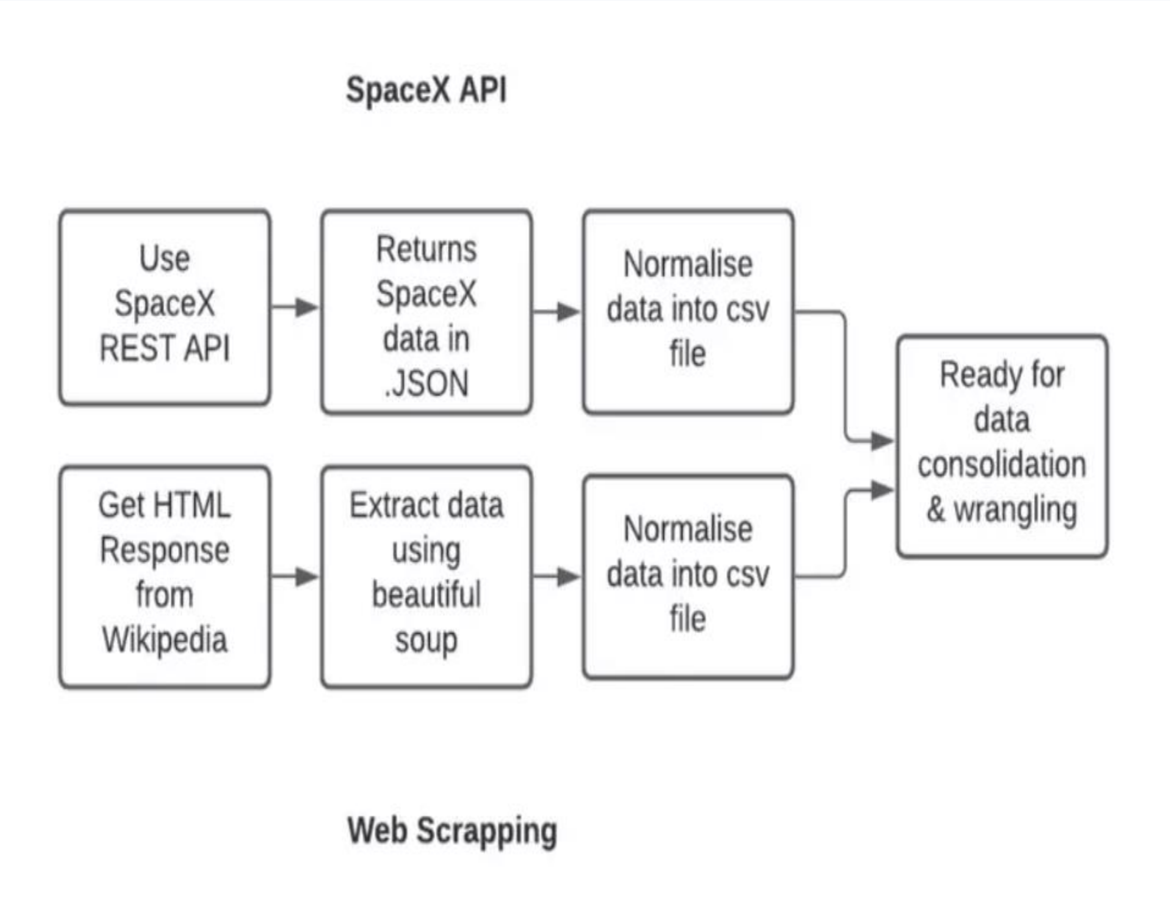
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/Data%20Collection%20API.ipynb



Data Collection - Scraping

- Obtained from API can also be obtained through web scraping
- In this case, Wikipedia was used

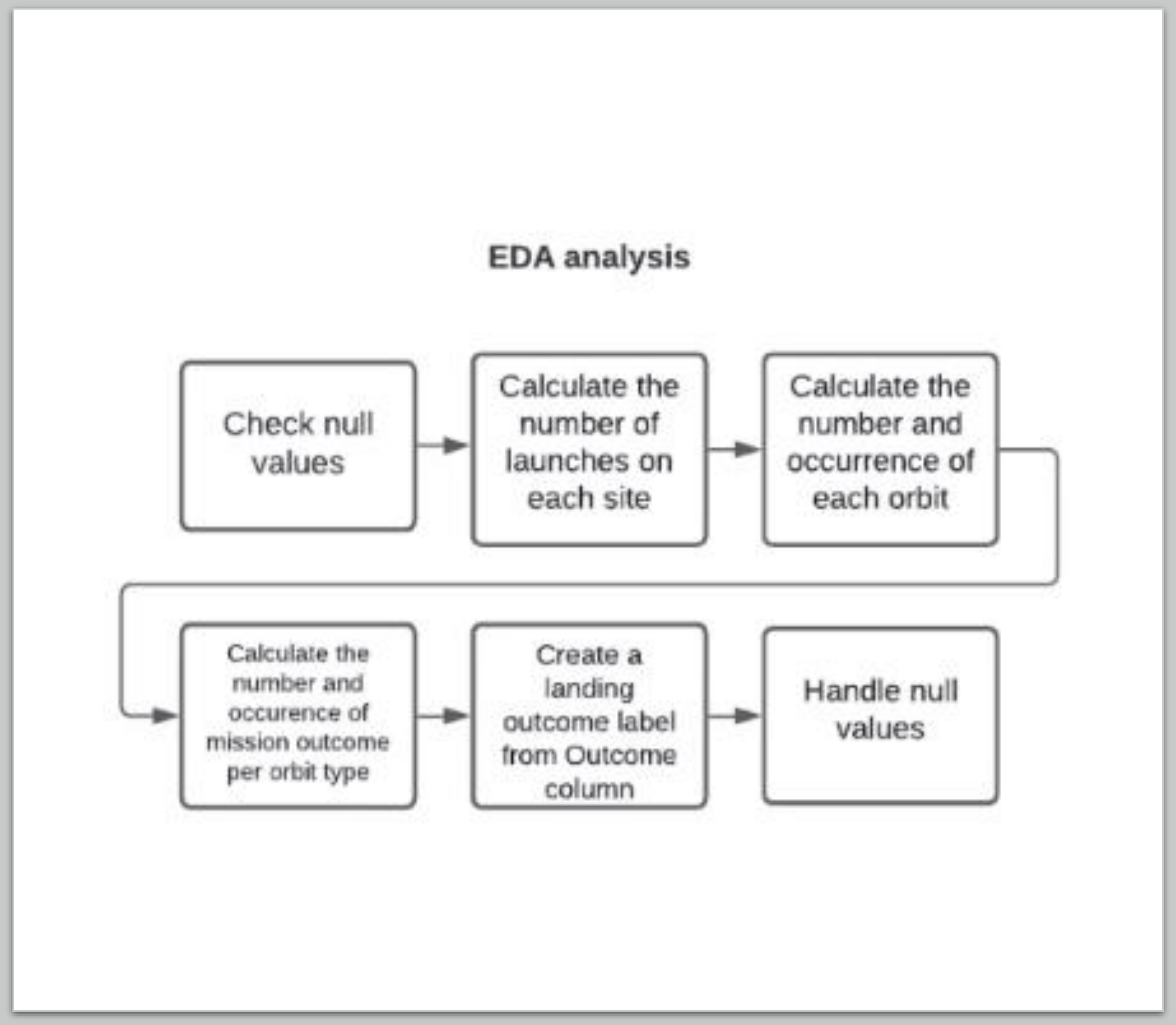
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Performed EDA on dataset
- Used this data to calculate number of launches, outcome of launches, creation of the outcome label, and the elimination of null values

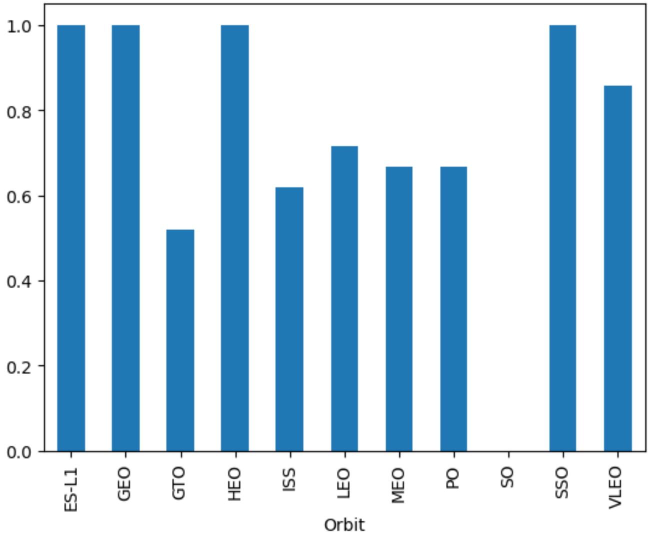
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/EDA%20Data%20Wrangling.ipynb



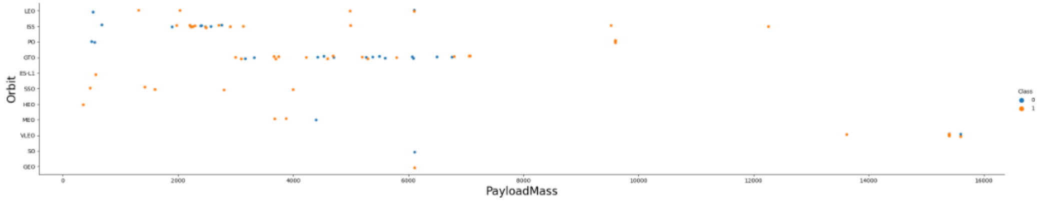
EDA with Data Visualization

- Used scatter and bar graphs to visualize collected data
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/EDA%20with%20Visualization.ipynb

```
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hu
sns.catplot(y="Orbit", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("FlightNumber",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```




```
# Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```

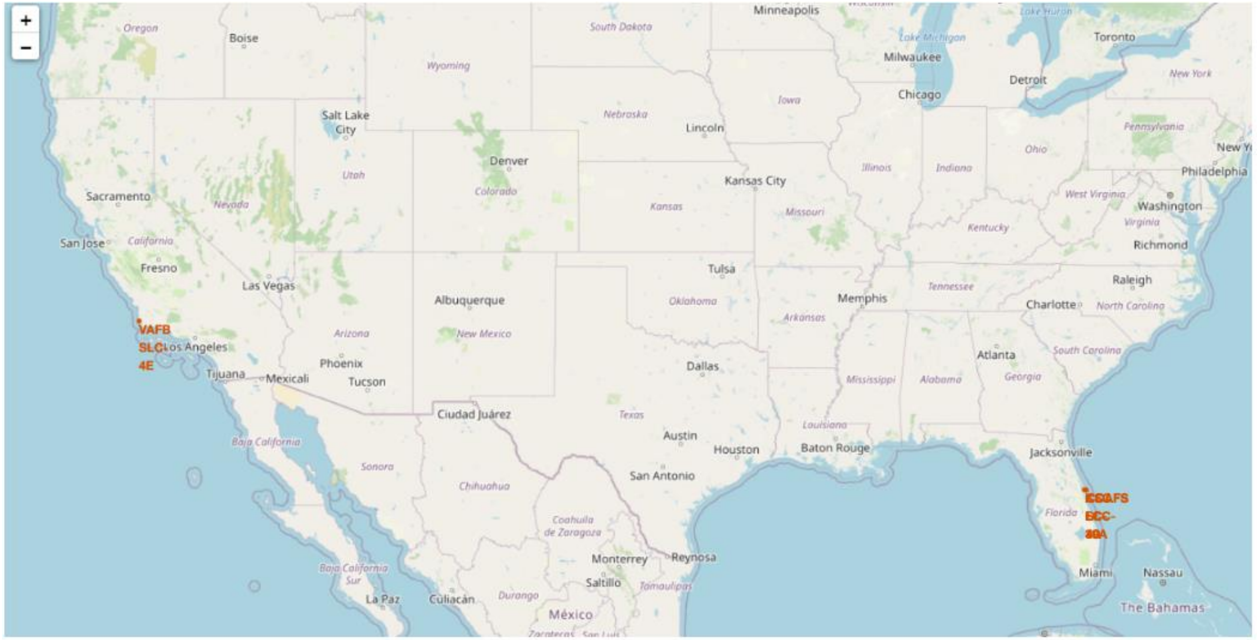


EDA with SQL

- Display unique launch sites
- First 5 launch sites beginning with CCA
- Total payload carried by NASA boosters
- Average payload carried by F9
- Date of first successful landing
- Name of boosters to successfully carry payloads between 4000-6000 kg
- Total number of successes and failures
- Name of booster that carried max payload
- Failed landings in 2015
- Ranking of outcomes between 2010-2017
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/EDA%20with%20SQL%20Data%20Wrangling.ipynb

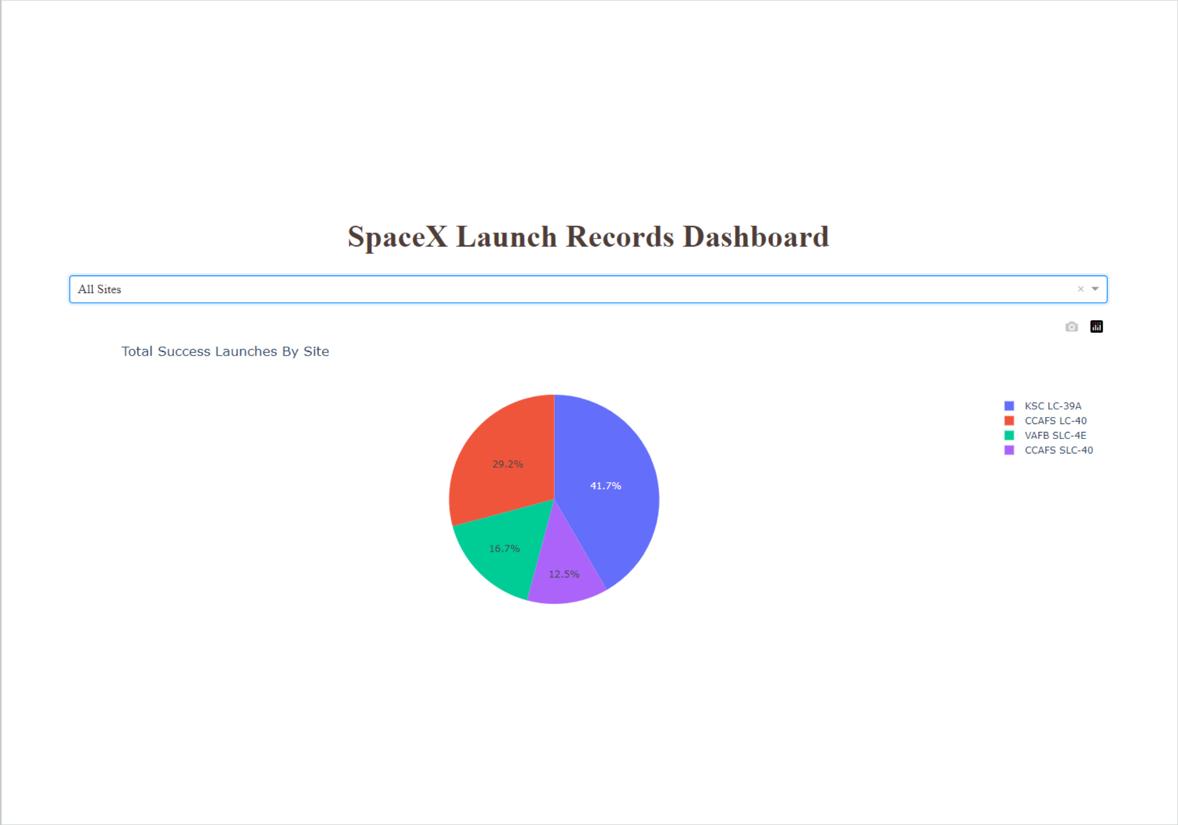
Build an Interactive Map with Folium

- Markers: Launch Sites
- Circles: Highlighted Areas at specific coordinates
- Marker Clusters: Groups of events
- Lines: Distance between coordinates
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/Interactive%20visual%20Analytics%20with%20Folium.ipynb



Build a Dashboard with Plotly Dash

- Utilized pie charts and scatter plots to visualize payload range and percentages of launches by site
- https://reedcahq-8050.theiadocker-0-labs-prod-theiak8s-4-tor01.proxy.cognitiveclass.ai/
- https://github.com/coreyreed/Data-Science-Capstone/blob/master/Dashboard%20Application%20with%20Plotly%20Dash.ipynb



Predictive Analysis (Classification)

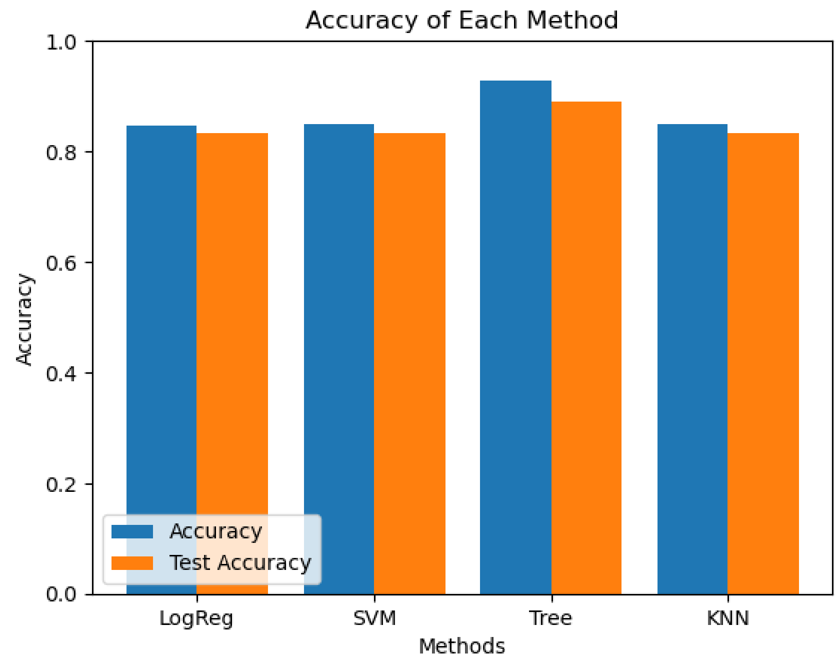
- Four models to make predictions: logistic regression, SVM, decision tree, and k nearest neighbor
- Prepare and Normalize data, Train and test date, make calculations and compare results

Prepare and Normalize data → Train and test data → Make calculations and compare results

https://github.com/coreyreed/Data-Science-Capstone/blob/master/Machine%20Learning%20Prediction.ipynb

Results

- Almost all launches were successful
- Average payload is 2928 kg
- The number of successful landings continued to increase after 2015, the year of the first success
- Most launches were on the east coast
- Decision Tree shown to have the best accuracy and test accuracy

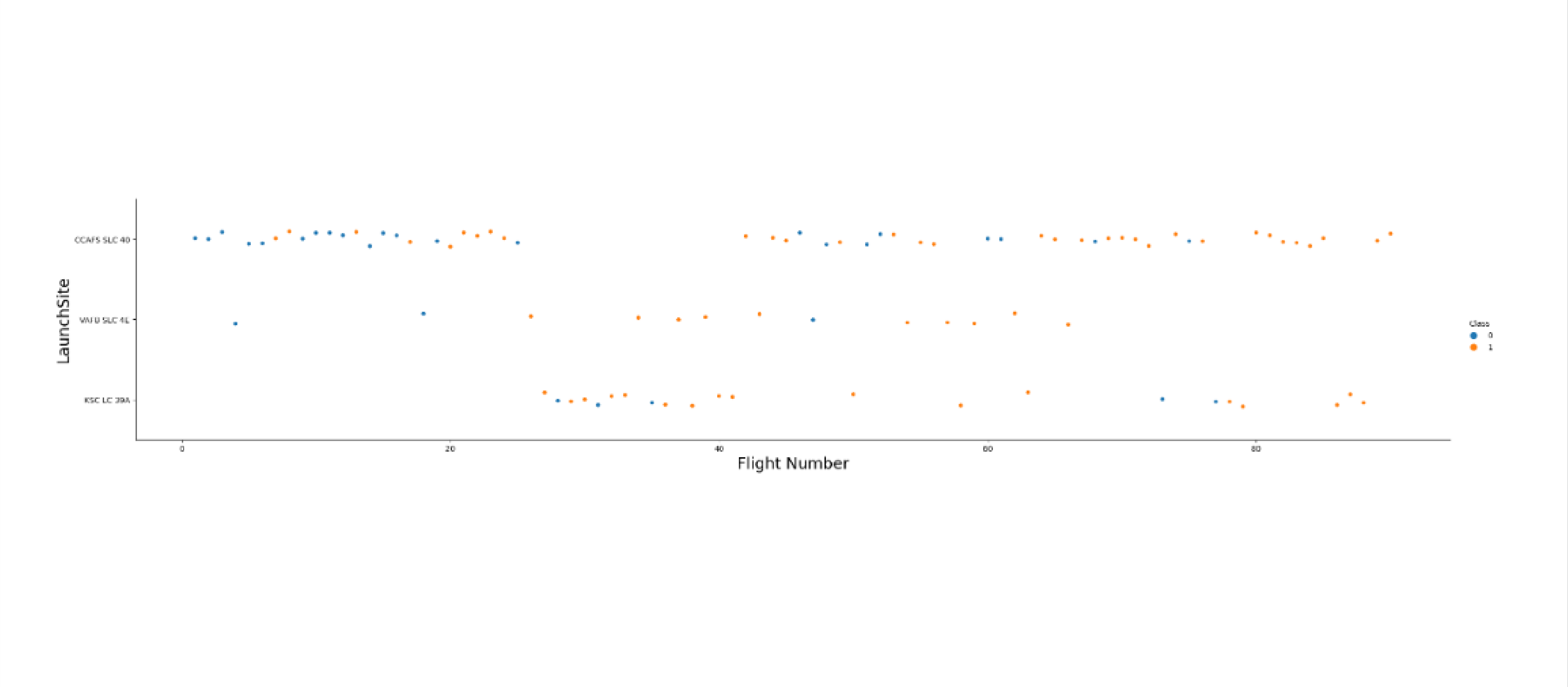


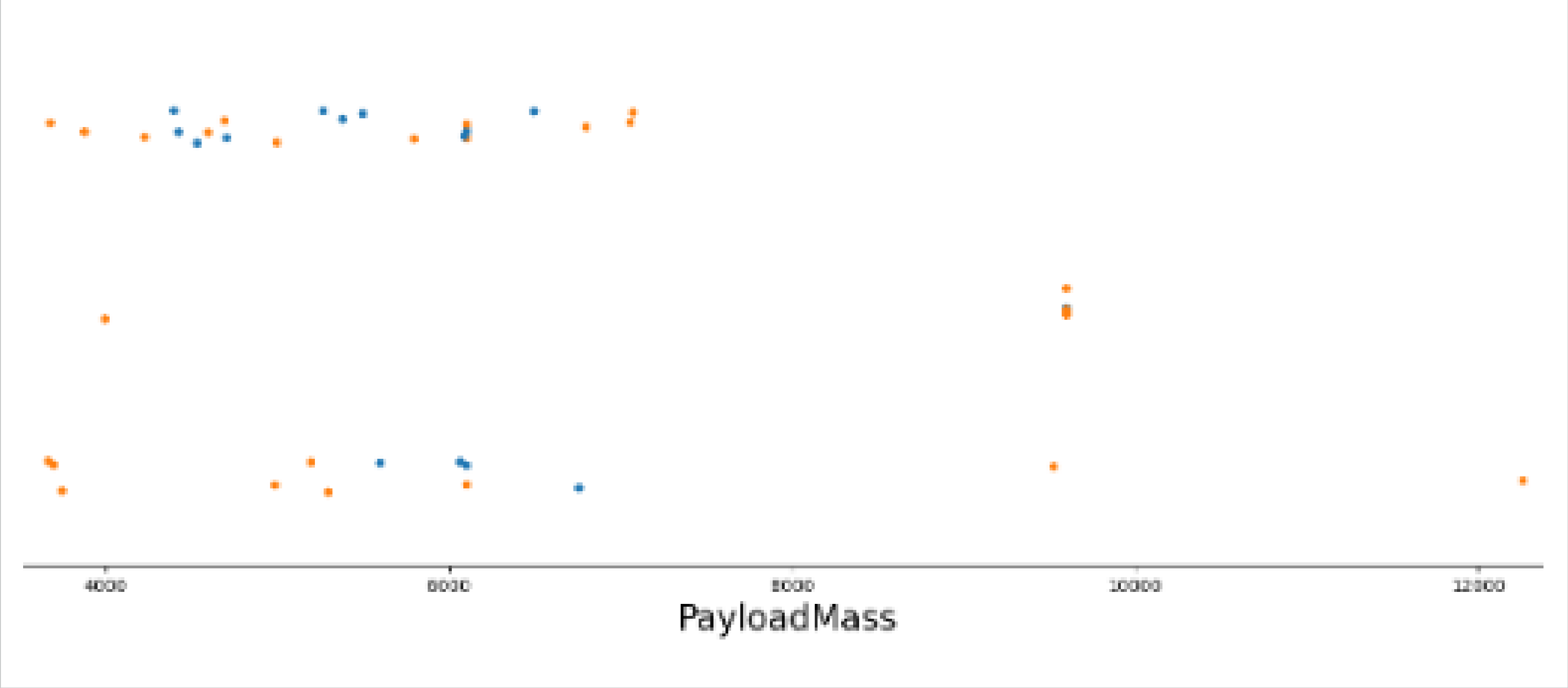
Section 2

Insights drawn from EDA

Flight Number vs Launch Site

- CCAF5 SLC 40 had the most successes and is where most launches occur
- Success rate increases over time



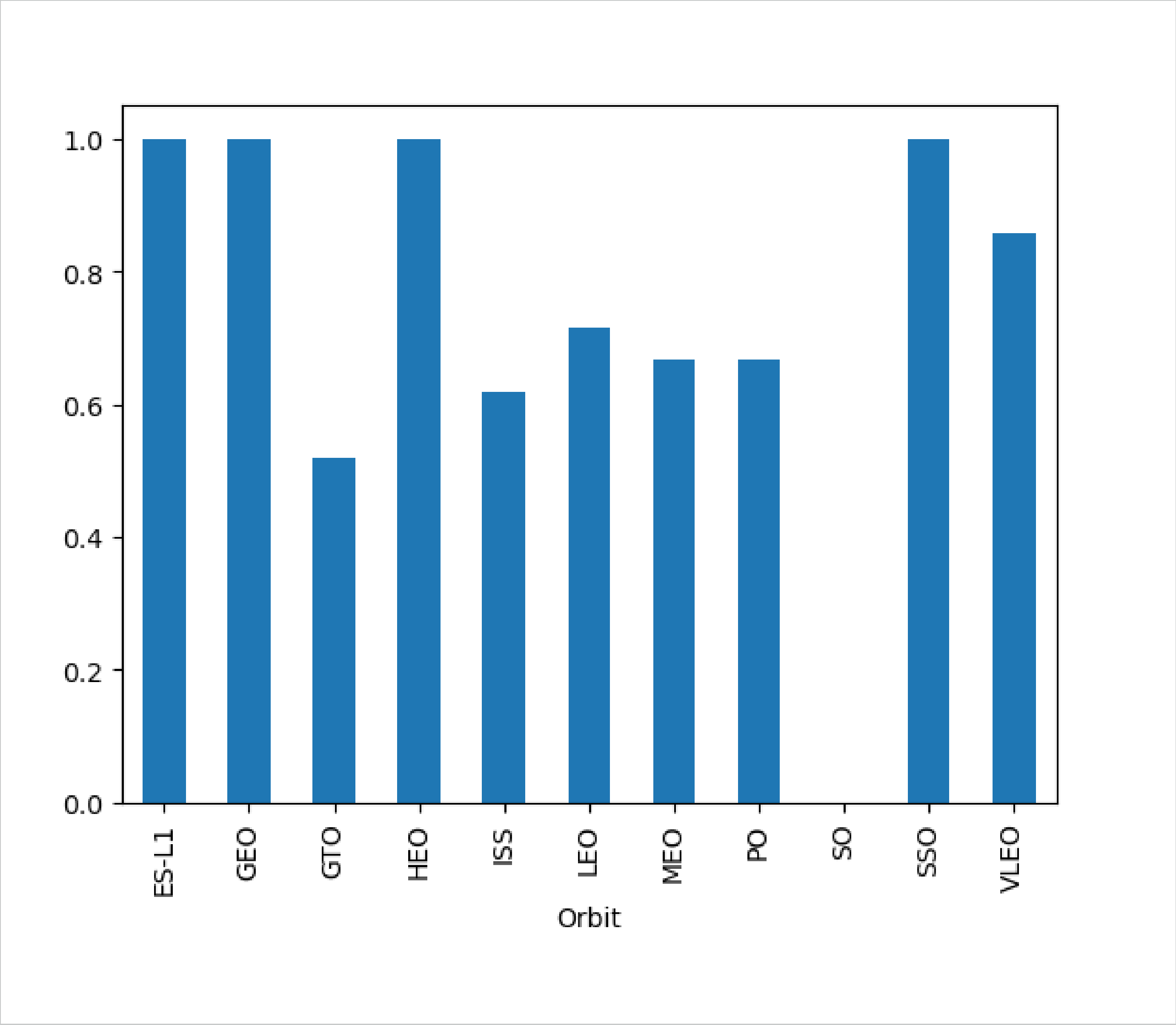


Payload vs. Launch Site

- Payloads over 9000 kg have excellent success

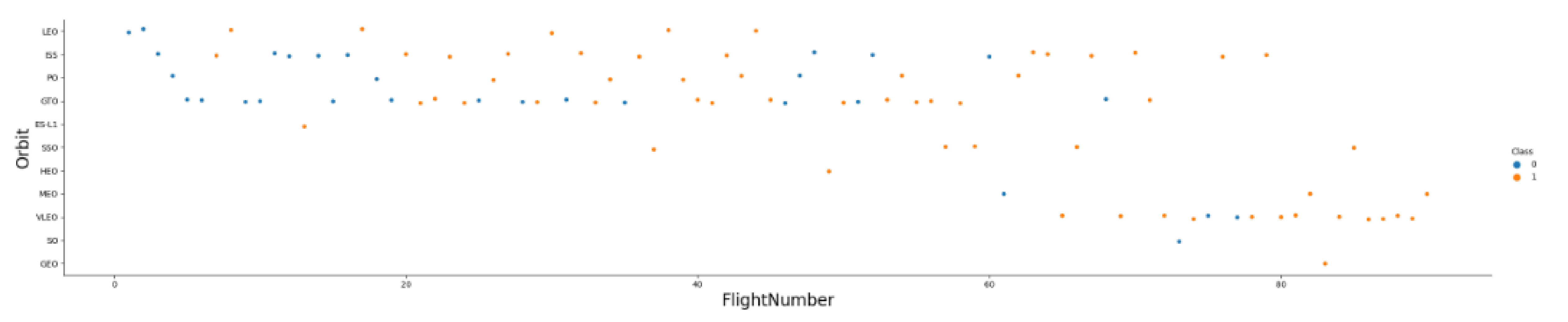
Success Rate vs. Orbit Type

- ESL-L1, GEO, HEO, and SSO have the highest rate of success



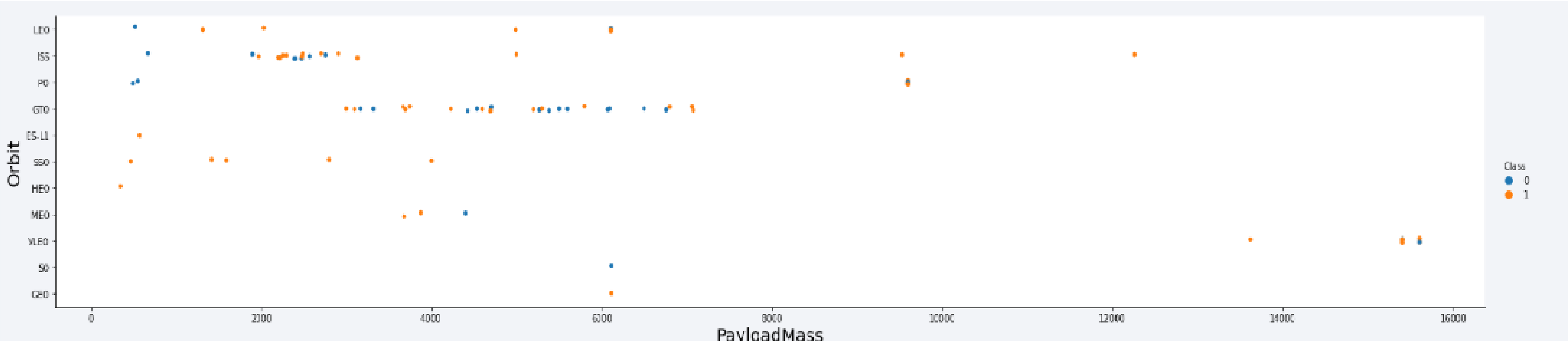
Flight Number vs. Orbit Type

- Rate of success increased in all orbits



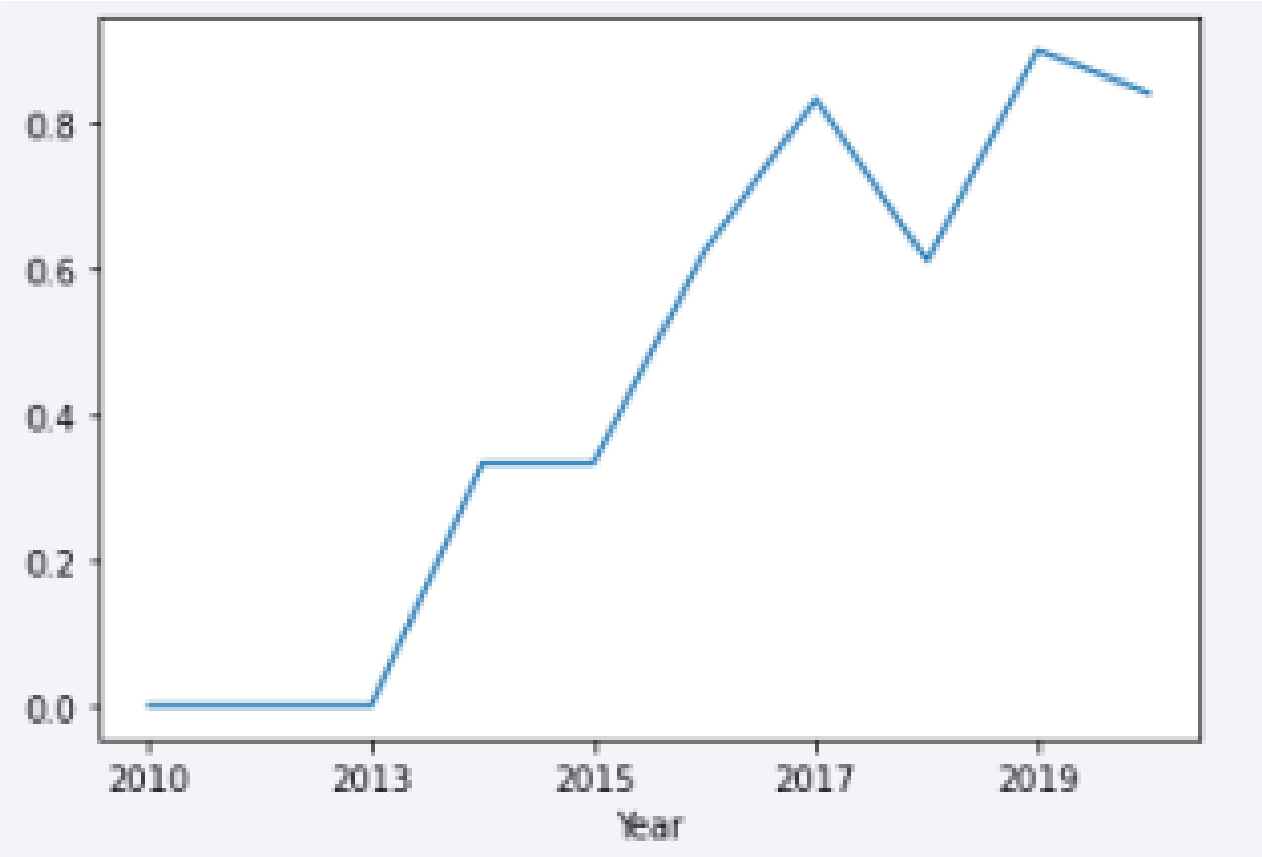
Payload vs. Orbit Type

- ISS uses different size payloads and high success rate
- Few SO and GEO launches



Launch Success Yearly Trend

- Success increased from 2013-2020
- 2010-2013 saw little success



All Launch Site Names

- Use launch_site query to find results

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

Date	Time UTC	Booster Version	Launch Site	Payload	Payload Mass kg	Orbit	Customer	Mission Outcome	Landing Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

- Sum all payloads beginning with CRS

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Avg Payload (kg)
2.928

First Successful Ground Landing Date

- Filtered data to find min value

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters that have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Filter the results

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- Use code to find number of occurrences of each group

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

Boosters Carried Maximum Payload

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- Only two failed landings

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Ranking of all landing outcomes

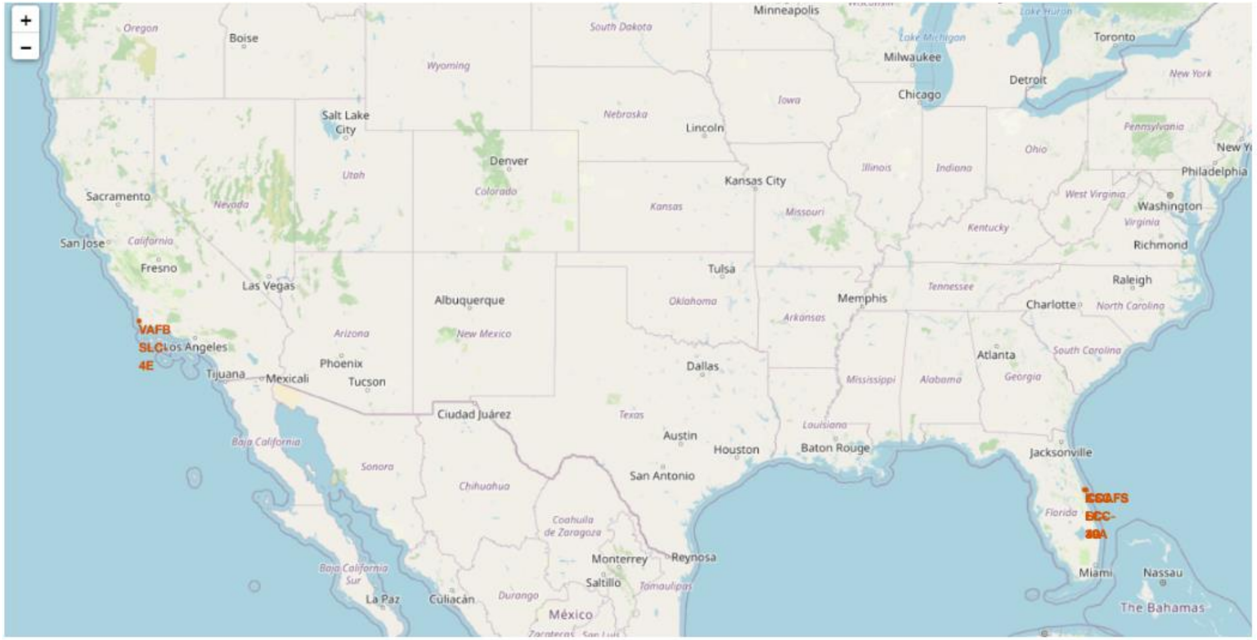
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

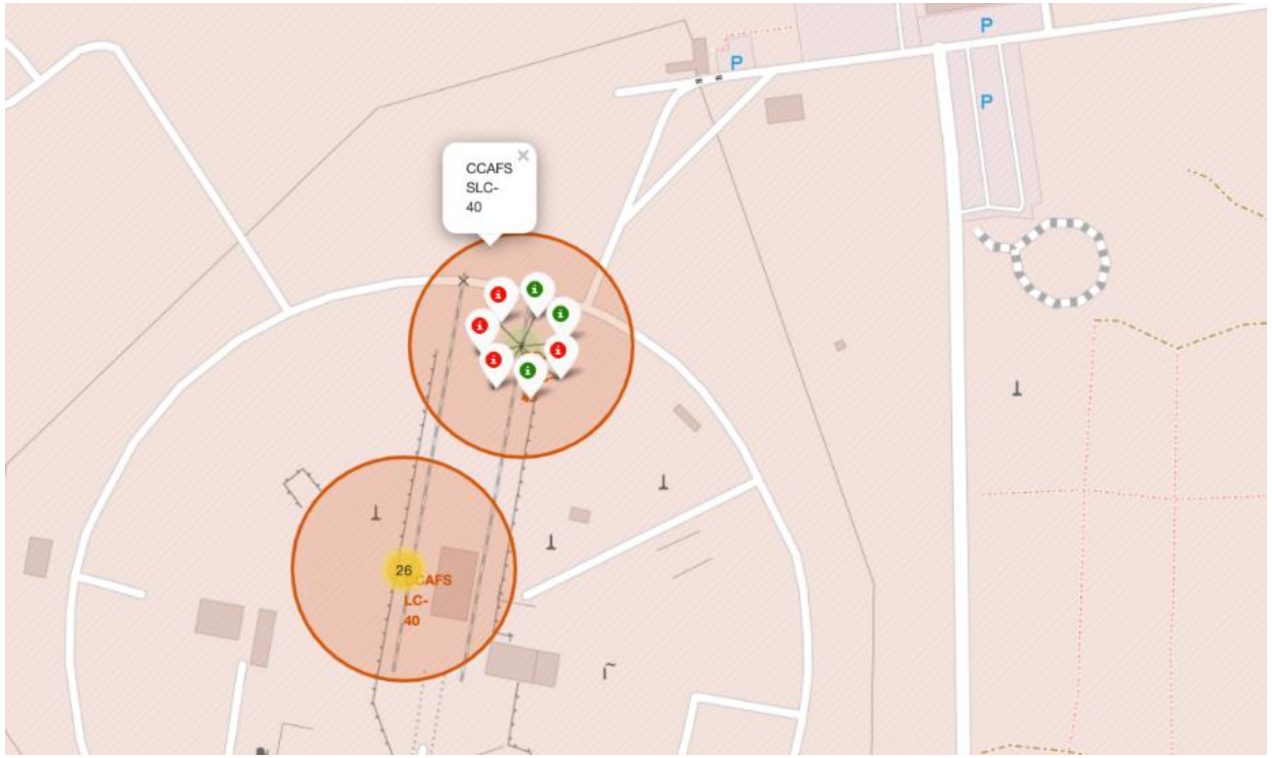
All Launch Sites

- Launch sites are all on the coasts and near roads



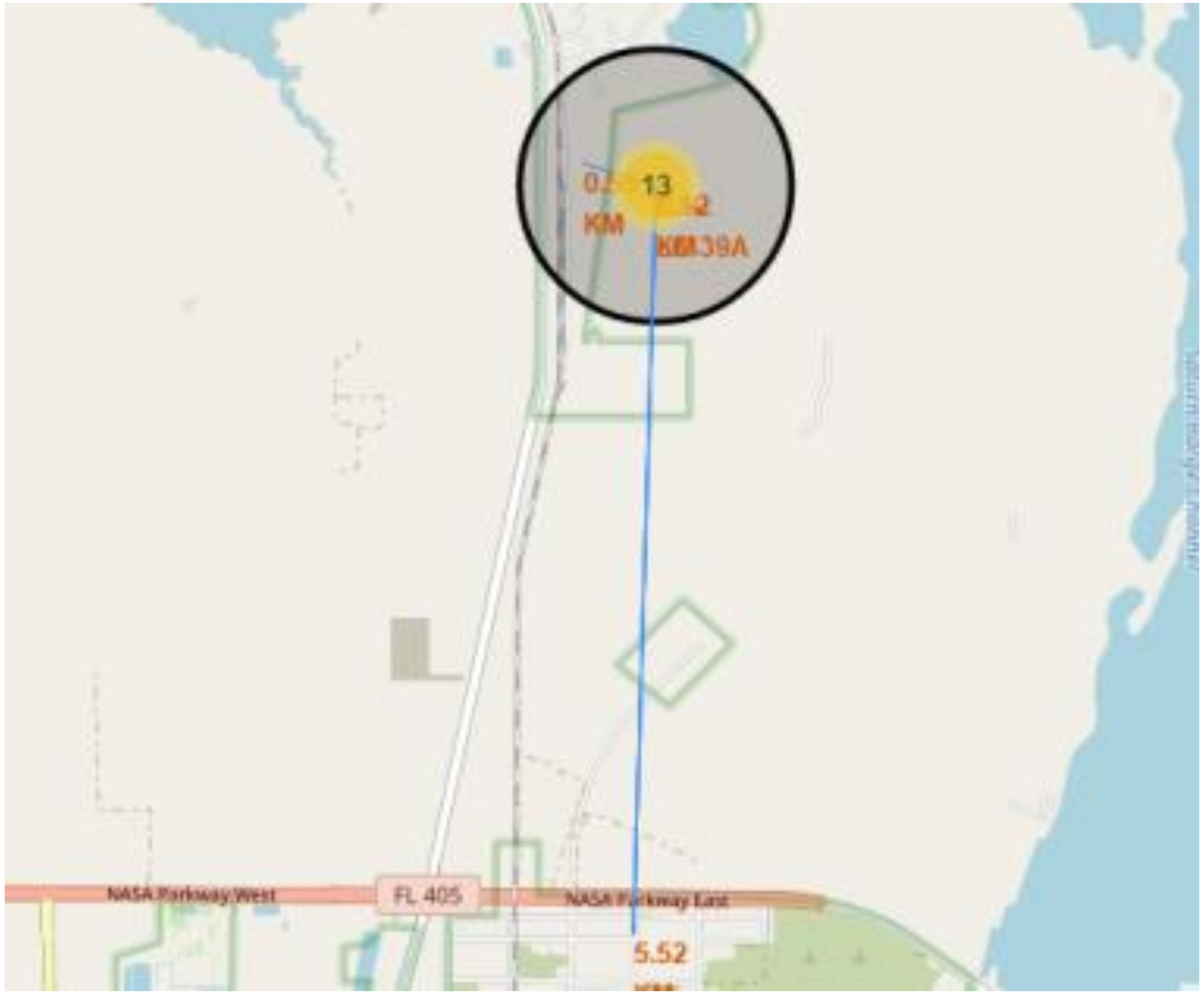
Outcomes by Site

- Green indicates success; red indicates failure



Distance to Roads

- All launch sites are within reasonable distance to roads for safety purposes

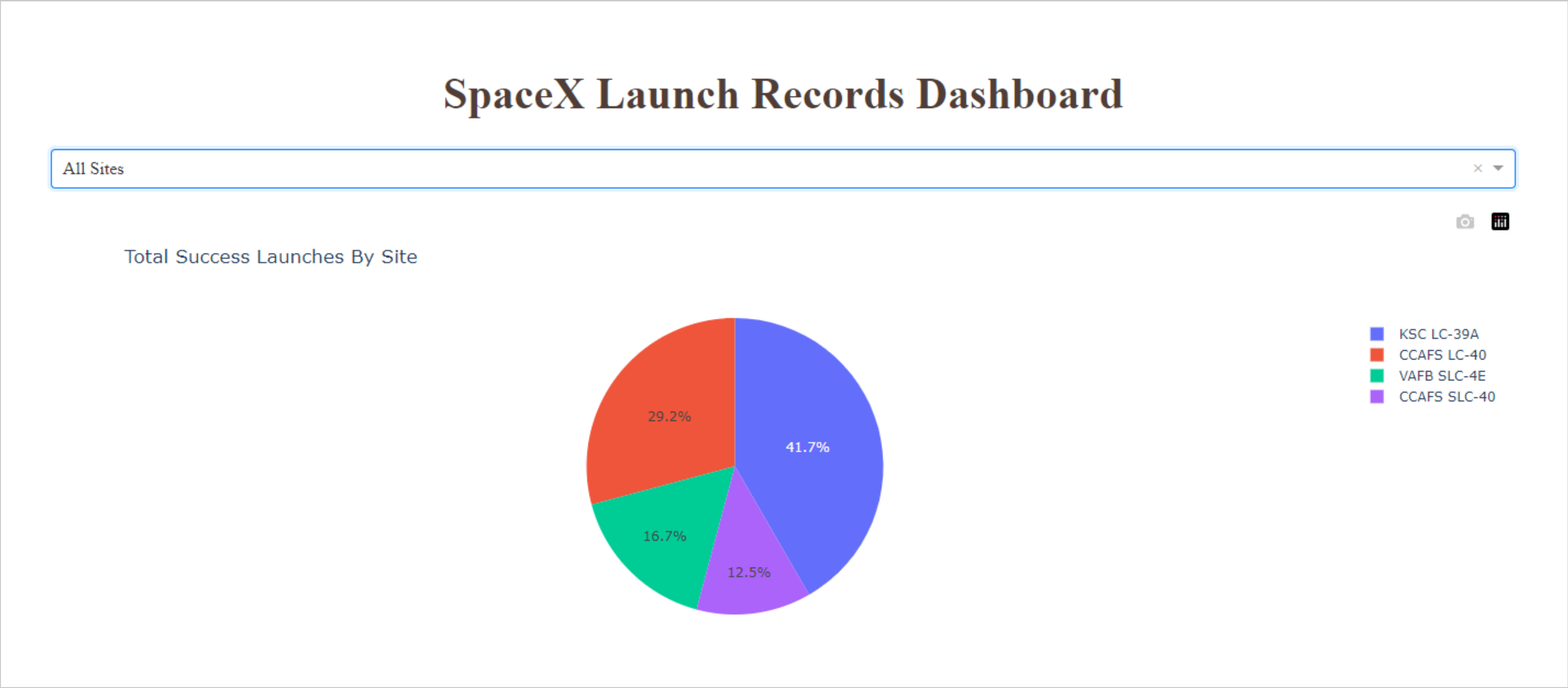


Section 4

Build a Dashboard with Plotly Dash

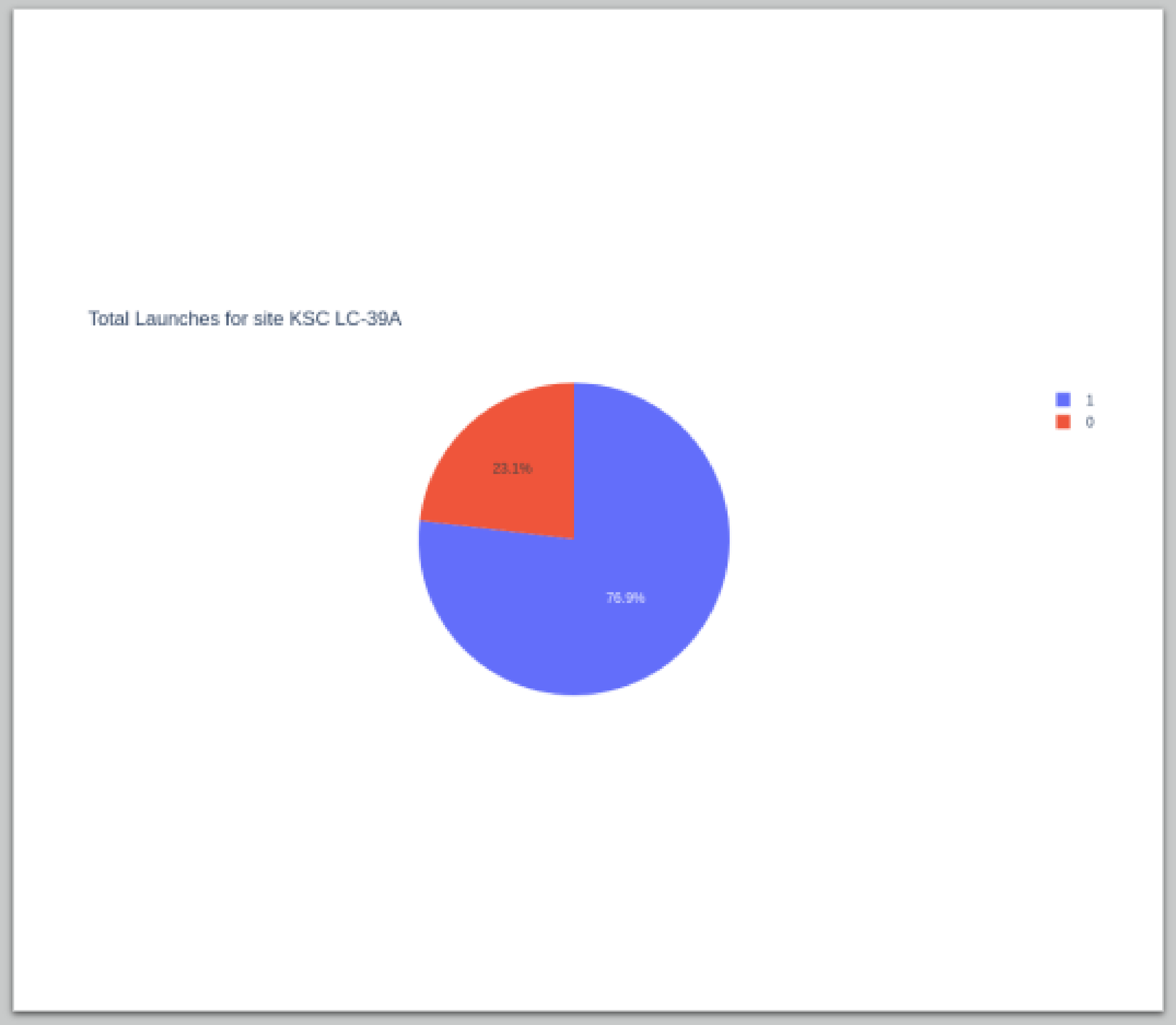
Successful Launches by Site

- From the chart, the location of each launch has an impact on the success rate



Most Successful Launch Site

- KSC LC-39A has a 76.9% success rate



Payload vs Launch Orbit

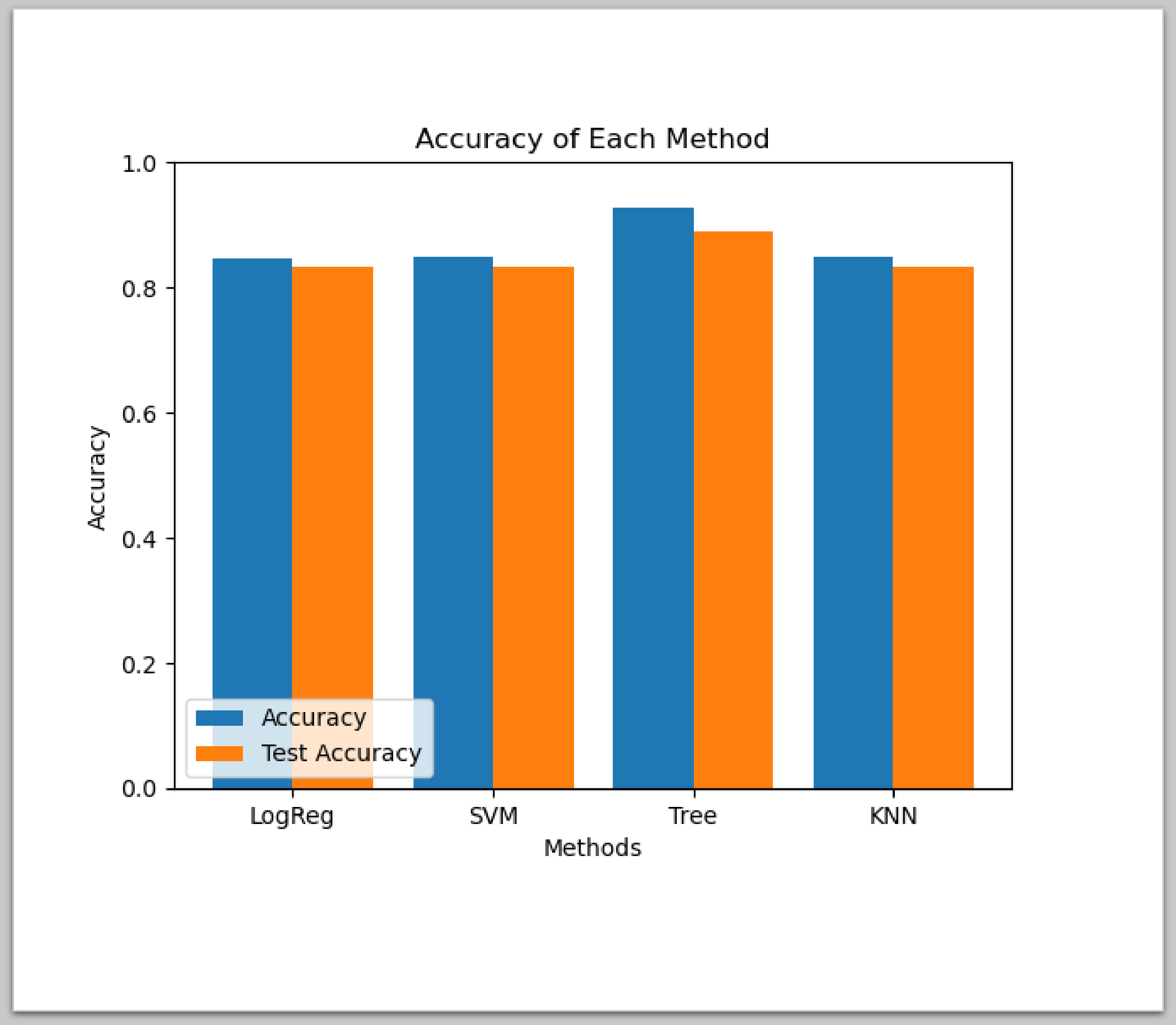
- Payloads > 6000 kg and FT boosters have the highest success rate

Section 5

Predictive Analysis (Classification)

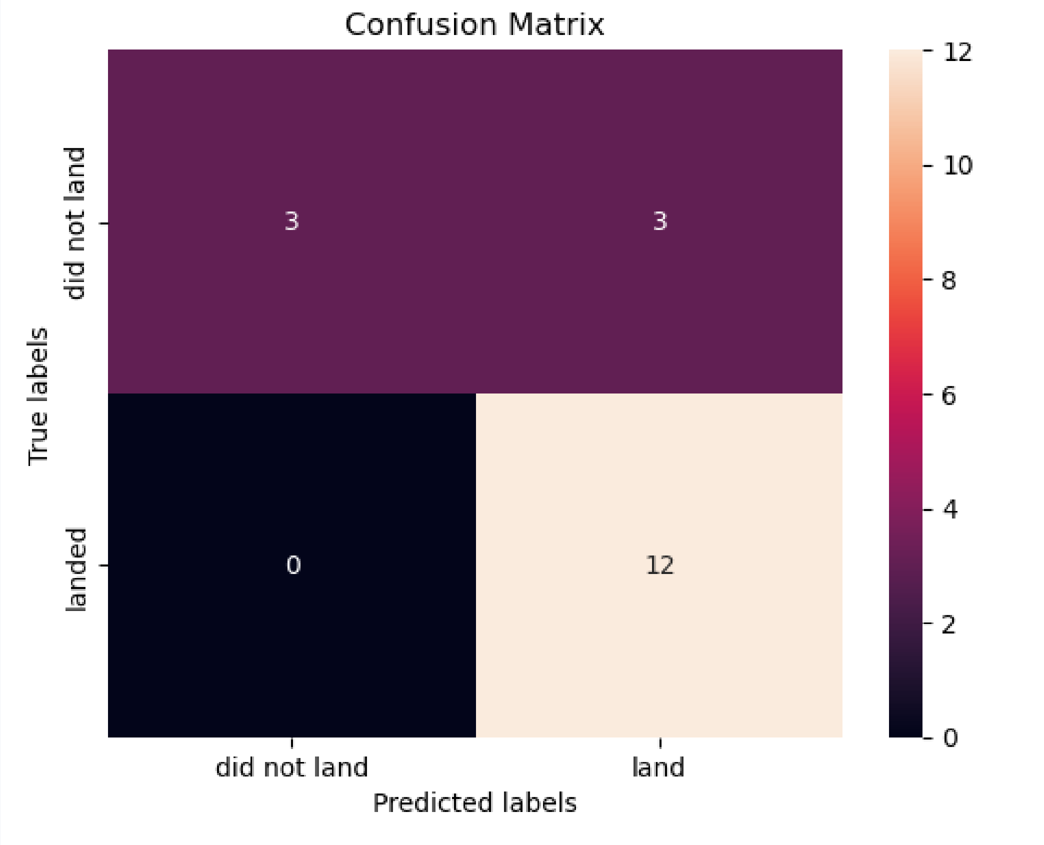
Classification Accuracy

- Four Models were tested
- Decision Tree shown to have the most accuracy and test accuracy



Confusion Matrix

- X of Decision tree shows the accuracy of the model



Conclusions

- The best launch site is KSC LC-39A
- Ideal payload mass is greater than 7000 kg, preferably about 9000 kg
- Decision Trees are very accurate models
- Most launches were successful from the beginnings, but they continued to improve over time
  - Possibly work done from 2010-2013 allowed for this high success rate

Appendix

- Some charts and images do not show up on GitHub, so screenshots were taken from Watson Studio

Thank you!