IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Cornelius Enetomhe

11th July 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

- Project background and context

  Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch. This goal of the project is to predict with the aid of machine learning pipeline if the first stage will land successfully.

- Problems you want to find answers

  - What factors determine if the rocket will land successfully?
  - The relationship amongst various features that determine the success rate of a successful landing.
  - What conditions ensures a successful landing.

Section 1

Methodology

Methodology

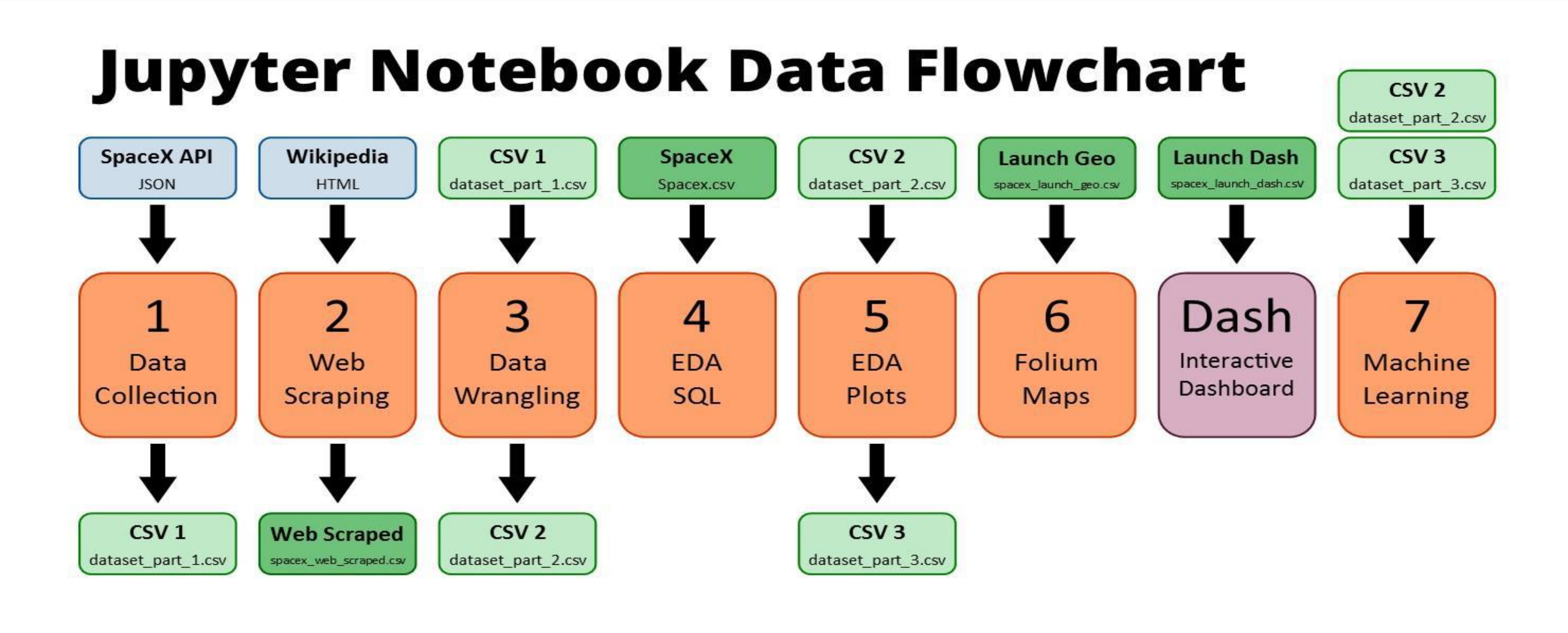
Executive Summary

- **Data collection methodology:**
  - With Rest API and Web scraping .
- **Perform data wrangling**
  - One-hot encoding was applied to categorical features
- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - Different models were built and evaluated to get best classifier

Data Collection

▪ The data sets were collected from:

- An IBM copy of a call to the publically accessible SpaceX API with launch data in JSON format.
- A permanently linked Wikipedia page with launch data in HTML tables .
- Further data sets were provided. See darker green .csv files in top row of diagram below.

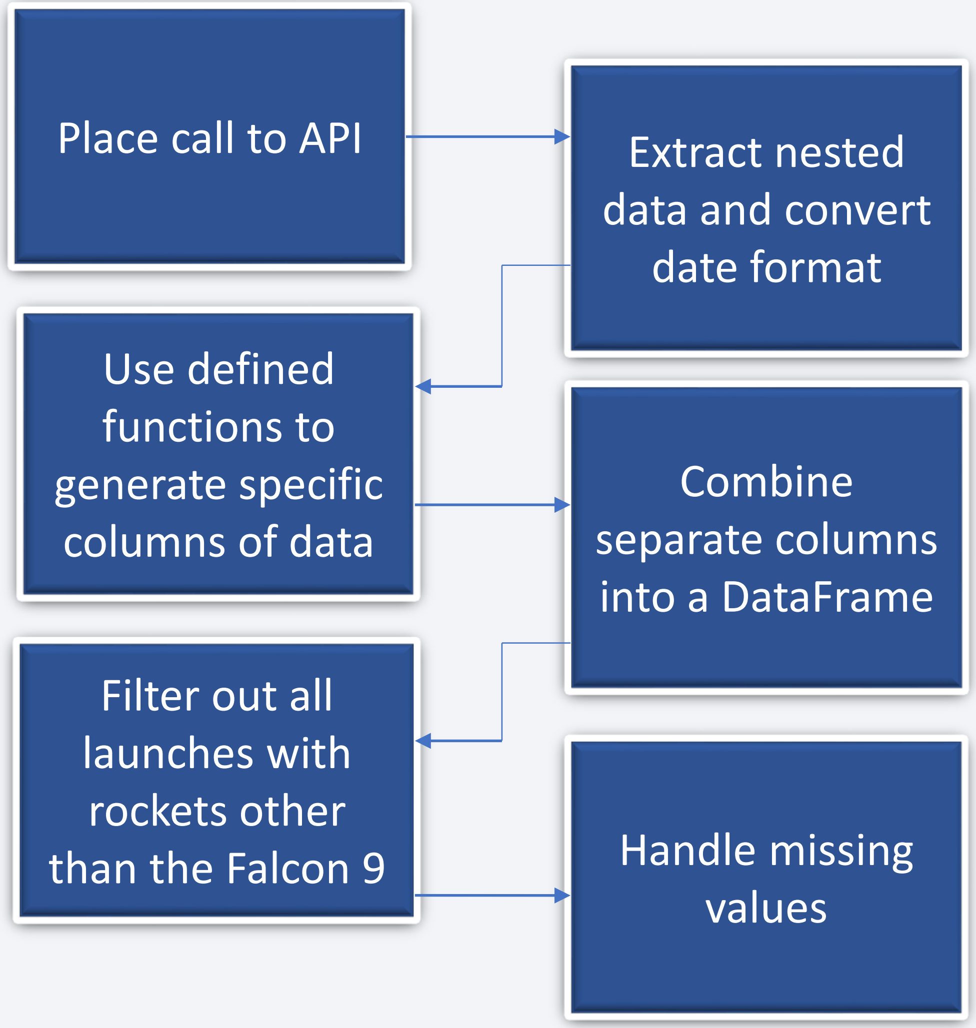


Data Collection – SpaceX API

- We used the get request to the SpaceX API to collect data and placed it into a Pandas dataframe for further analysis

- The GitHub link to the notebook is https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/spacex-data-collection-api.ipynb

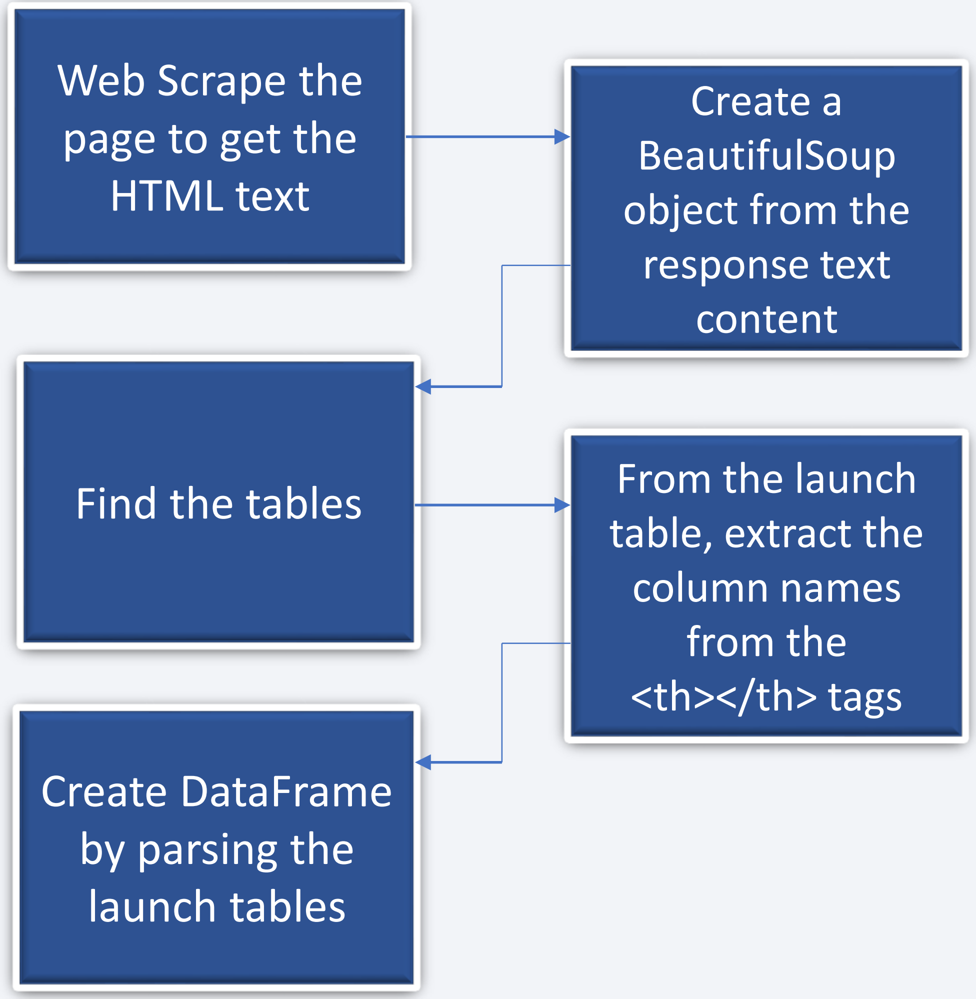
Flowchart of SpaceX API Calls



Data Collection - Scraping

- We applied web scraping to webscrap Falcon 9 launch records with BeautifulSoup from a Wikipedia page.
- We parsed the table and converted it into a pandas dataframe for further analysis.
- The GitHub link to the notebook is https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceXwebscraping.ipynb

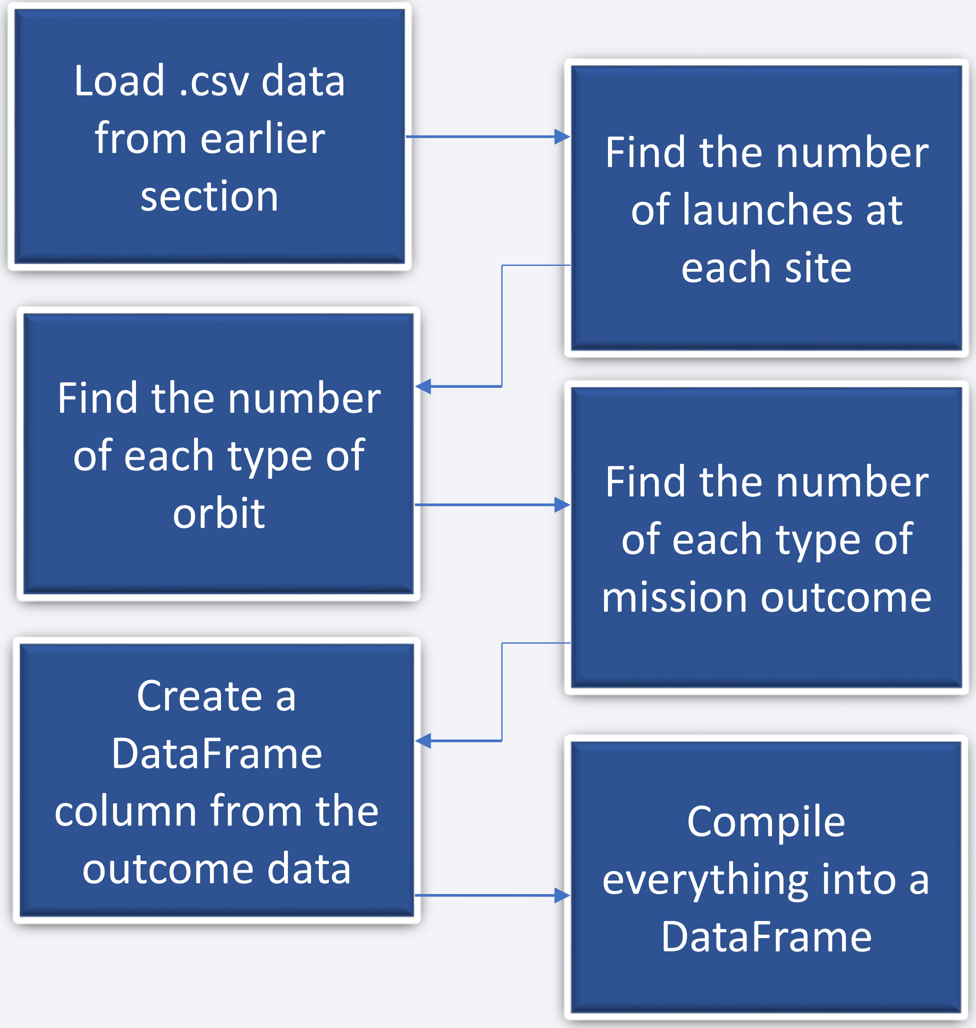
Flowchart of SpaceX Web Scraping



Data Wrangling

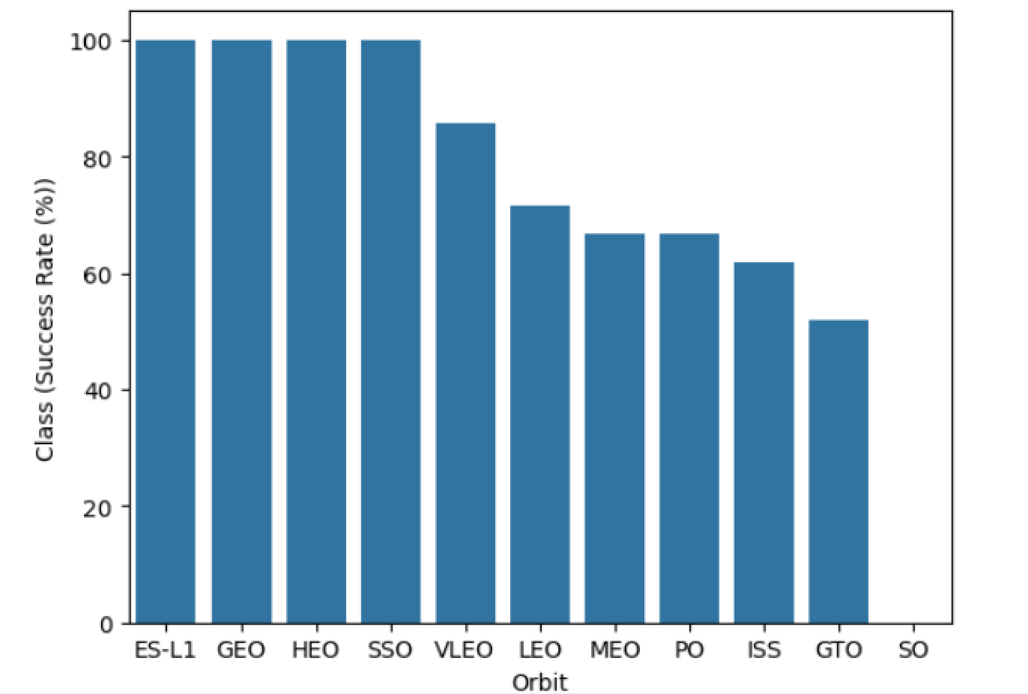
- We loaded the .csv files from earlier sections and cleaned them.
- We calculated the number of launches at each site, and the number and occurrence of each orbits
- The handful of mission outcome types were converted to a binary classification where 1 means that the Falcon 9 first stage landing was a success and 0 means that it was a failure.
- Added the classification to the dataframe.
- The GitHub link to the notebook is https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/spacex-Data%20wrangling.ipynb

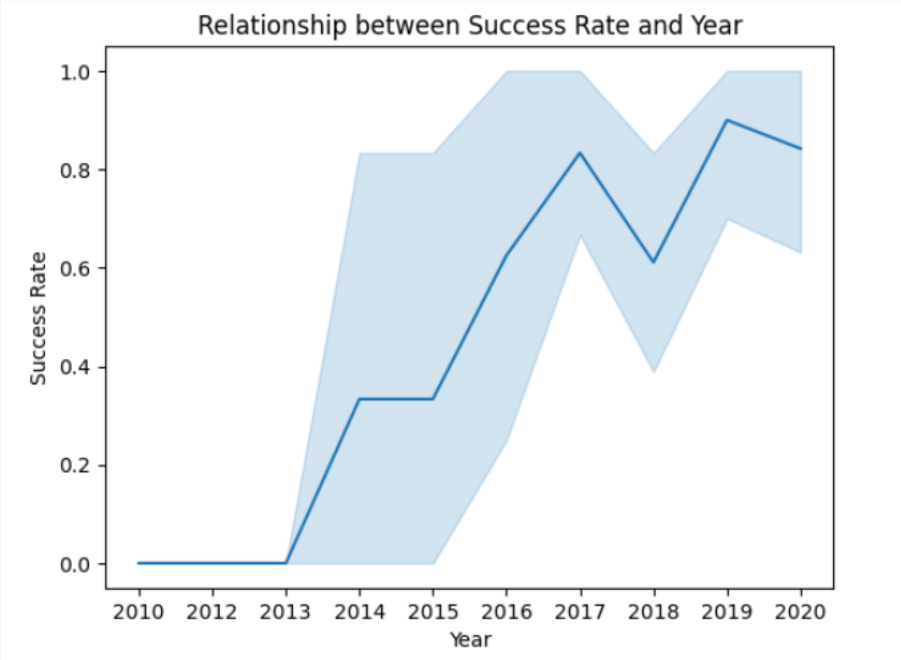
Flowchart of SpaceX Data Wrangling



EDA with Data Visualization

- We explored the data by visualizing the relationship between flight number and launch Site, payload and launch site, success rate of each orbit type, flight number and orbit type, the launch success yearly trend.





- The GitHub link to the notebook is https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX%20Exploratory%20Analysis.ipynb

EDA with SQL

- We applied EDA with SQL to get insight from the data. We wrote queries to find out for instance:
  - The names of unique launch sites in the space mission.
  - The total payload mass carried by boosters launched by NASA (CRS)
  - The average payload mass carried by booster version F9 v1.1
  - The total number of successful and failure mission outcomes
  - The failed landing outcomes in drone ship, their booster version and launch site names.
  - Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20
- The GitHub link to the notebook is

https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Markers were added for launch sites and for the NASA Johnson Space Center
  - Circles were added for the launch sites.
  - Lines were added to show the distance to the nearby features
    - Distance from CCAFS LC-40 to the coastline
    - Distance from CCAFS LC-40 to the rail line
    - Distance from CCAFS LC-40 to the perimeter road
- The GitHub link to the notebook is

https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_launch_site_location.ipynb

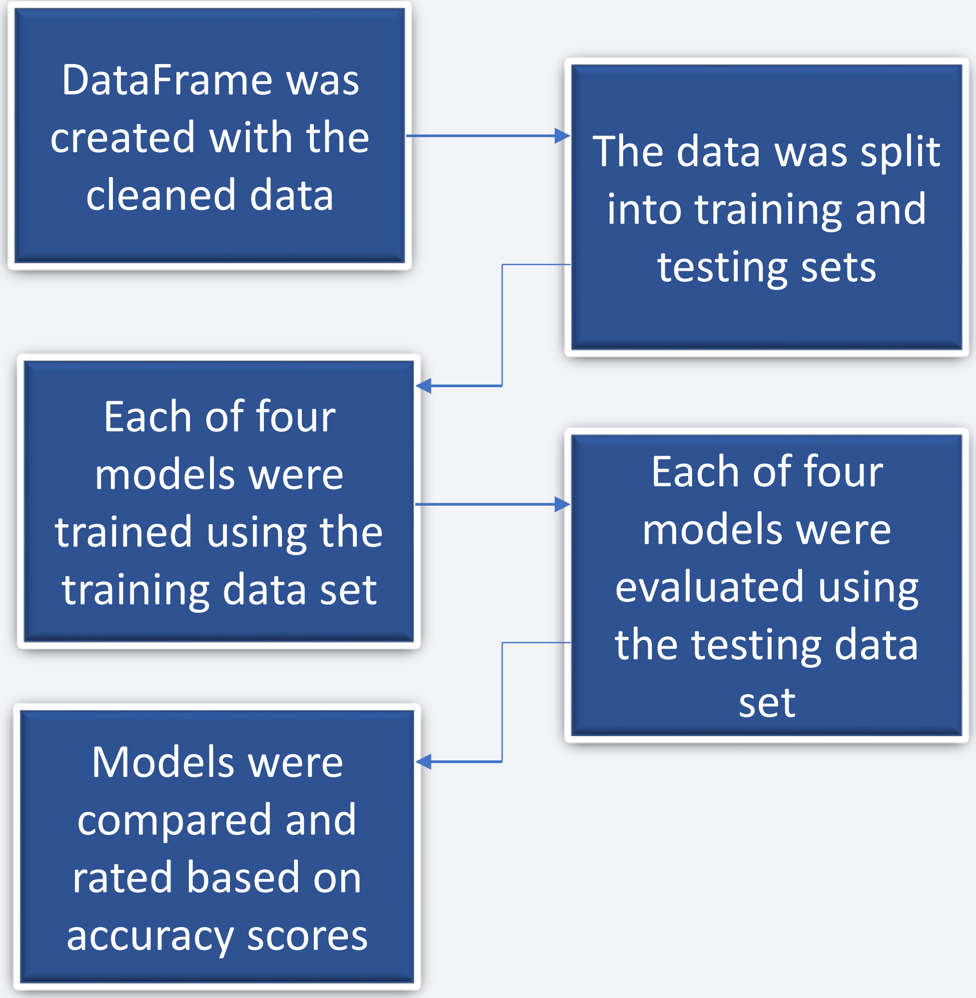
Build a Dashboard with Plotly Dash

- The input dropdown is used to select one or all launch sites for the pie chart and scatterplot.
- The pie chart displays one of two things:
  - For All Sites –the distribution of successful Falcon 9 first stage landings between the sites
  - For One Site –the distribution of successful and failed Falcon 9 first stage landings for that site
- The input slider is used to filter the payload masses for the scatterplot.
- The scatterplot displays the distribution of Falcon 9 first stage landings split by payload mass, mission outcome and by booster version category

Predictive Analysis (Classification)

- The dataset was split into training and testing sets.
- Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) machine learning models were trained on the training data set.
- Hyper-parameters were evaluated using GridSearchCV() and the best was selected using '.best_params_'.
- Using the best hyper-parameters, each of the four models were scored on accuracy by using the testing data set.
- The GitHub link to the notebook is

https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

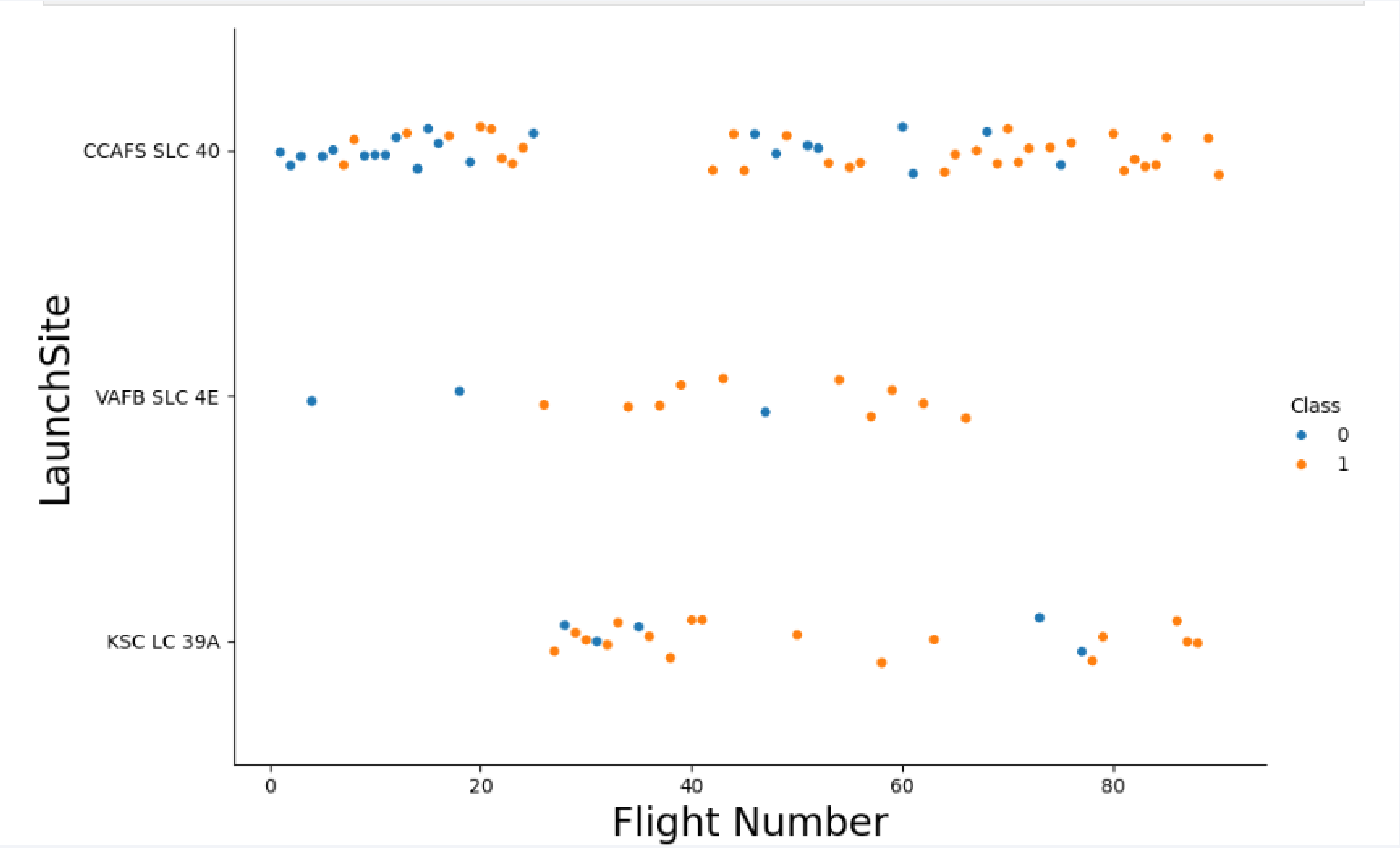
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

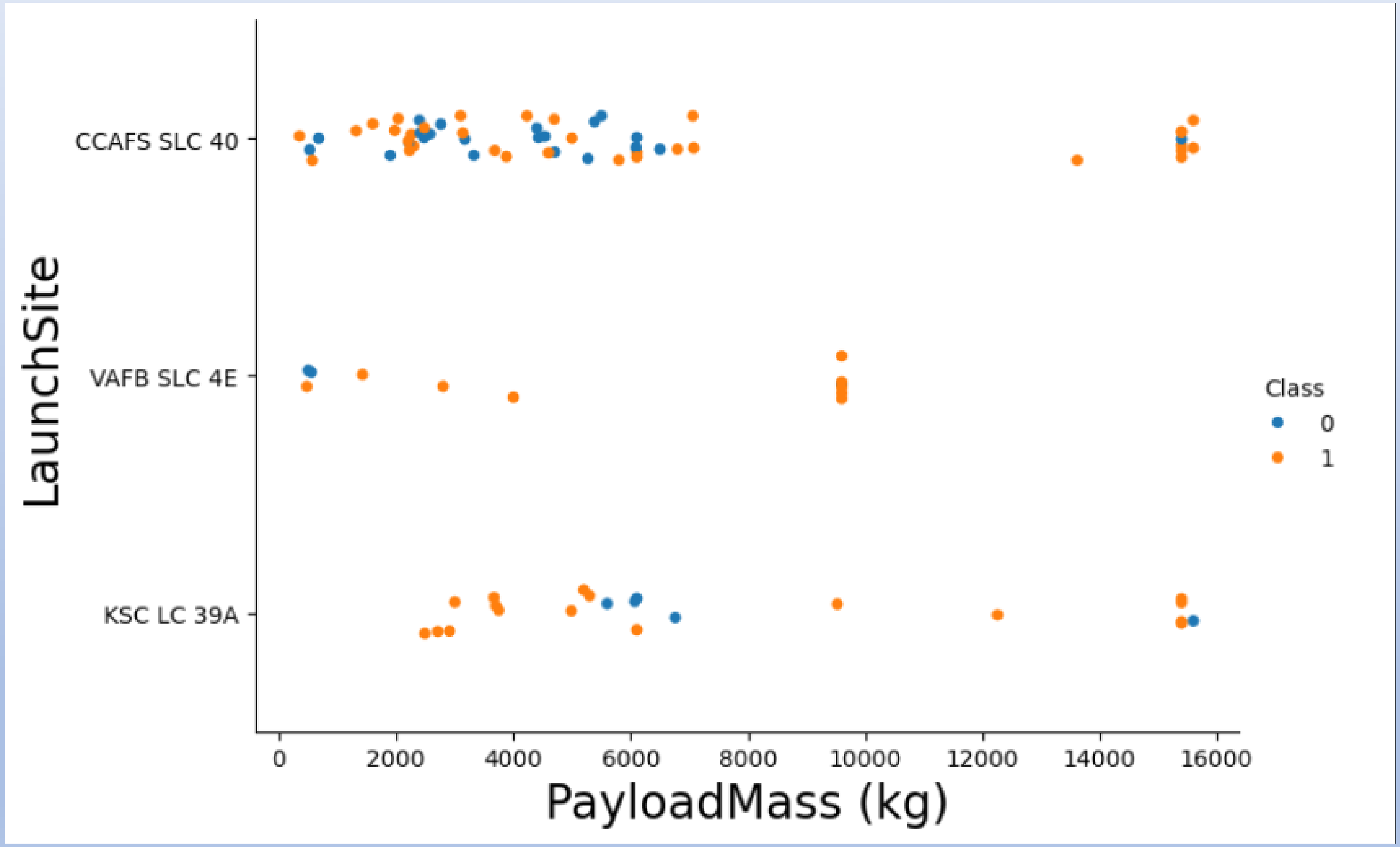
Flight Number vs. Launch Site

- Successful Falcon 9 first stage landings appear to become more prevalent as the flight number increases in each launch site.



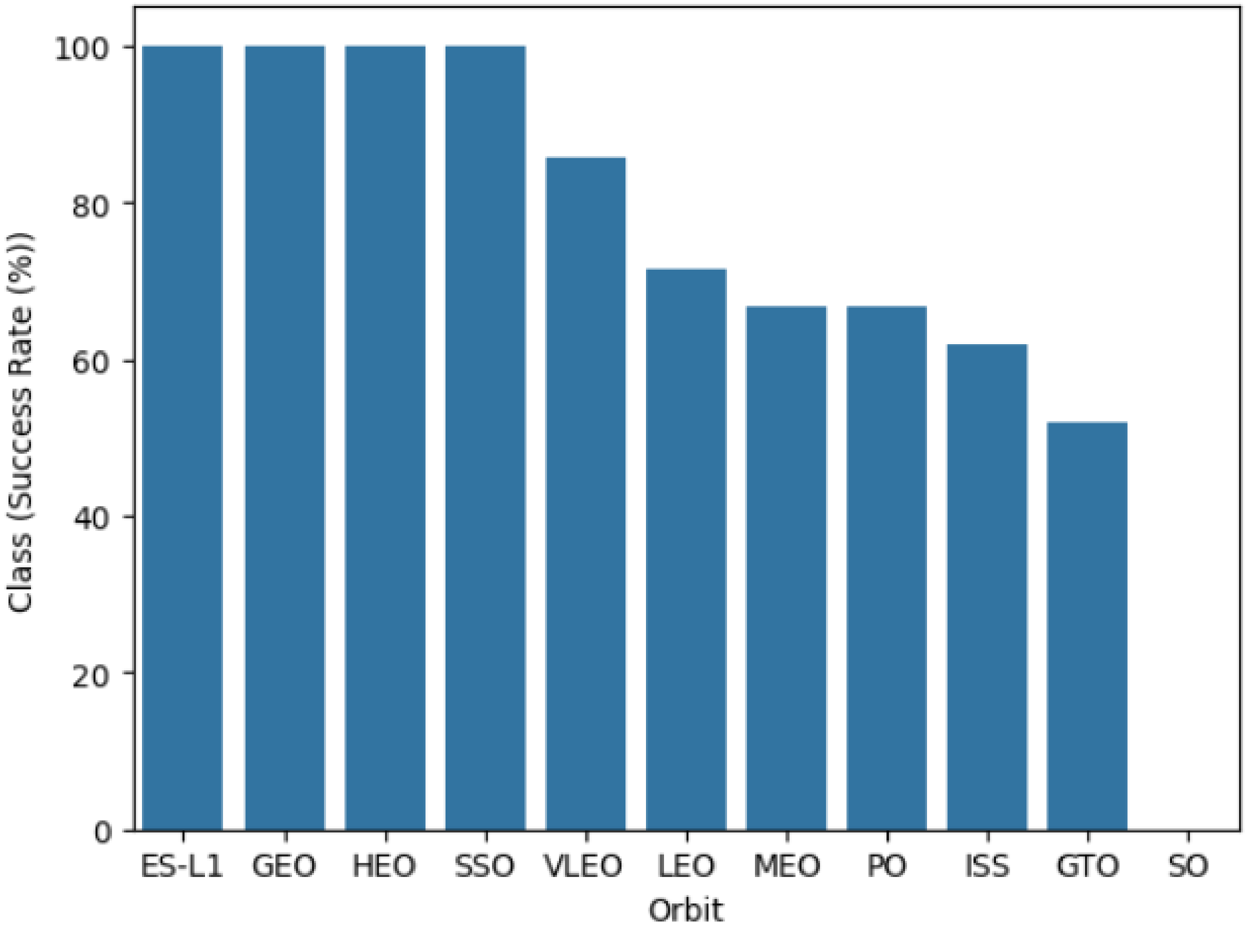
Payload Vs. Launch Site

- For the CCAFS SLC 40 launch site, the payload mass and the landing outcome appear to not be strongly correlated.
- The failed landings at the KSC LC 39A launch site are all grouped around a narrow band of payload masses.



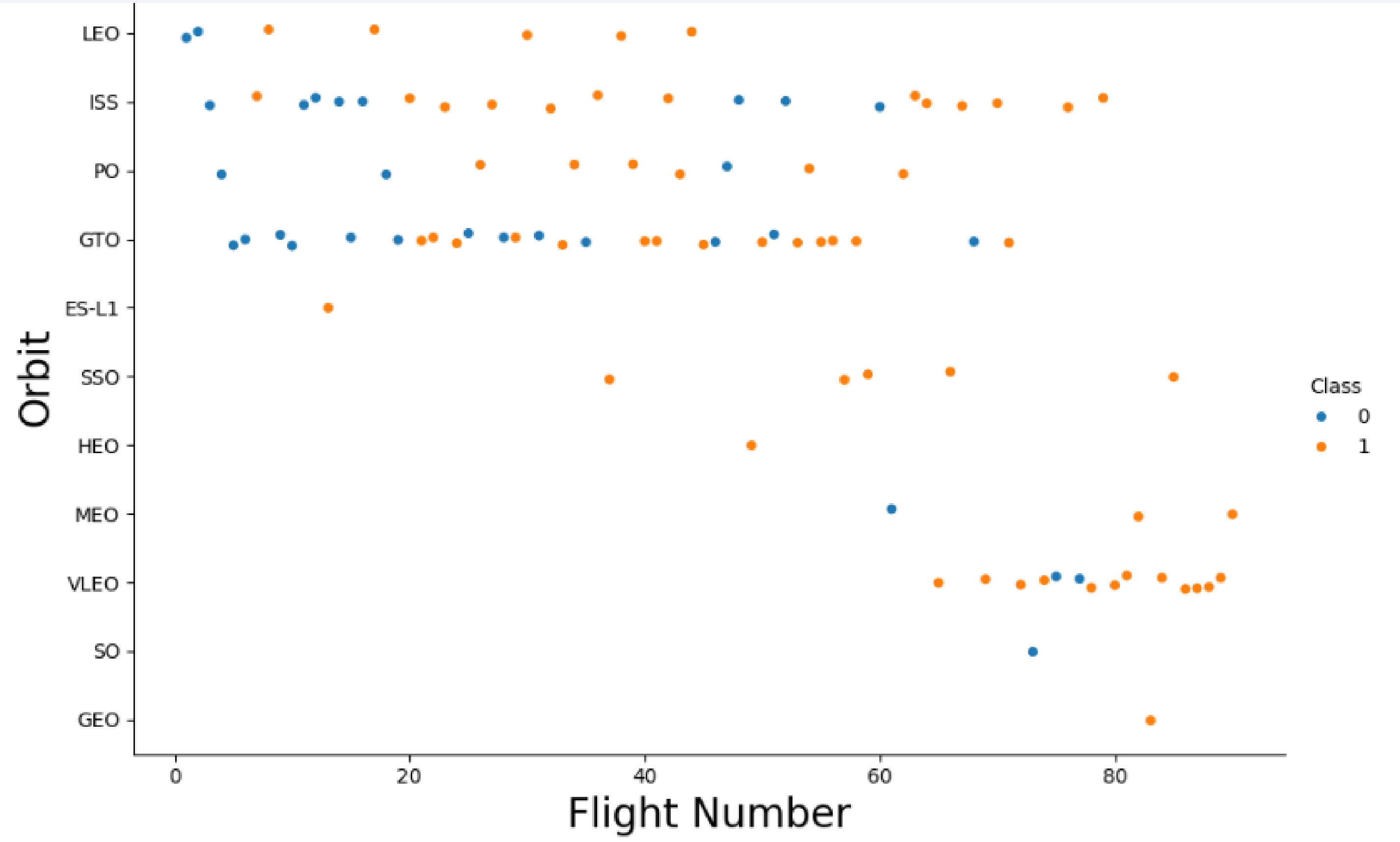
Success Rate vs. Orbit Type

- ES-L1, SSO, HEO and GEO orbits have no failed first stage landings.
- SO orbits have no successful first stage landings.



Flight Number vs. Orbit Type

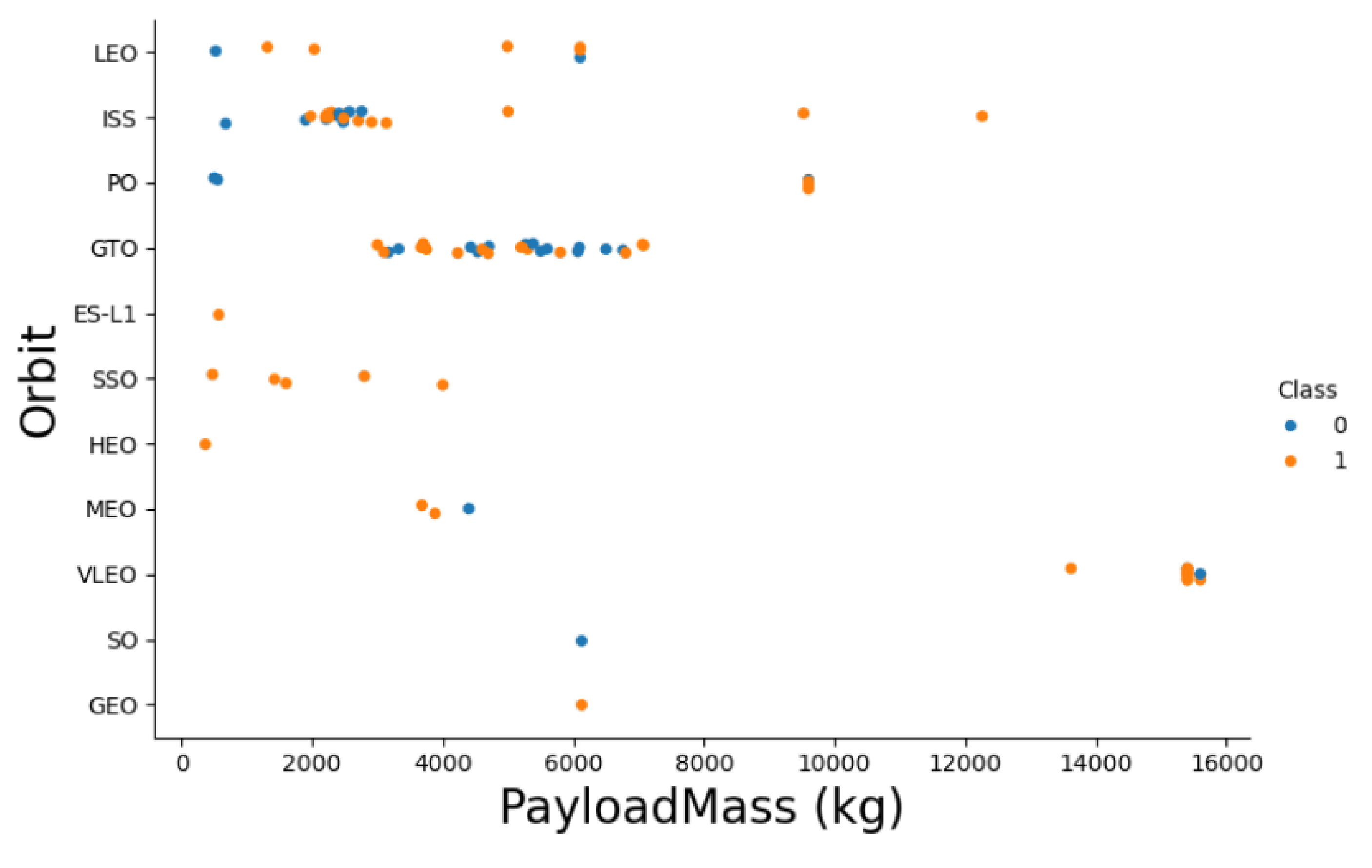
- There is a correlation between flight number and success rate with larger flight numbers being associated with higher success rates.
- Although we observe that in the GTO orbit, there is no relationship between flight number and the orbit.



Payload vs. Orbit Type

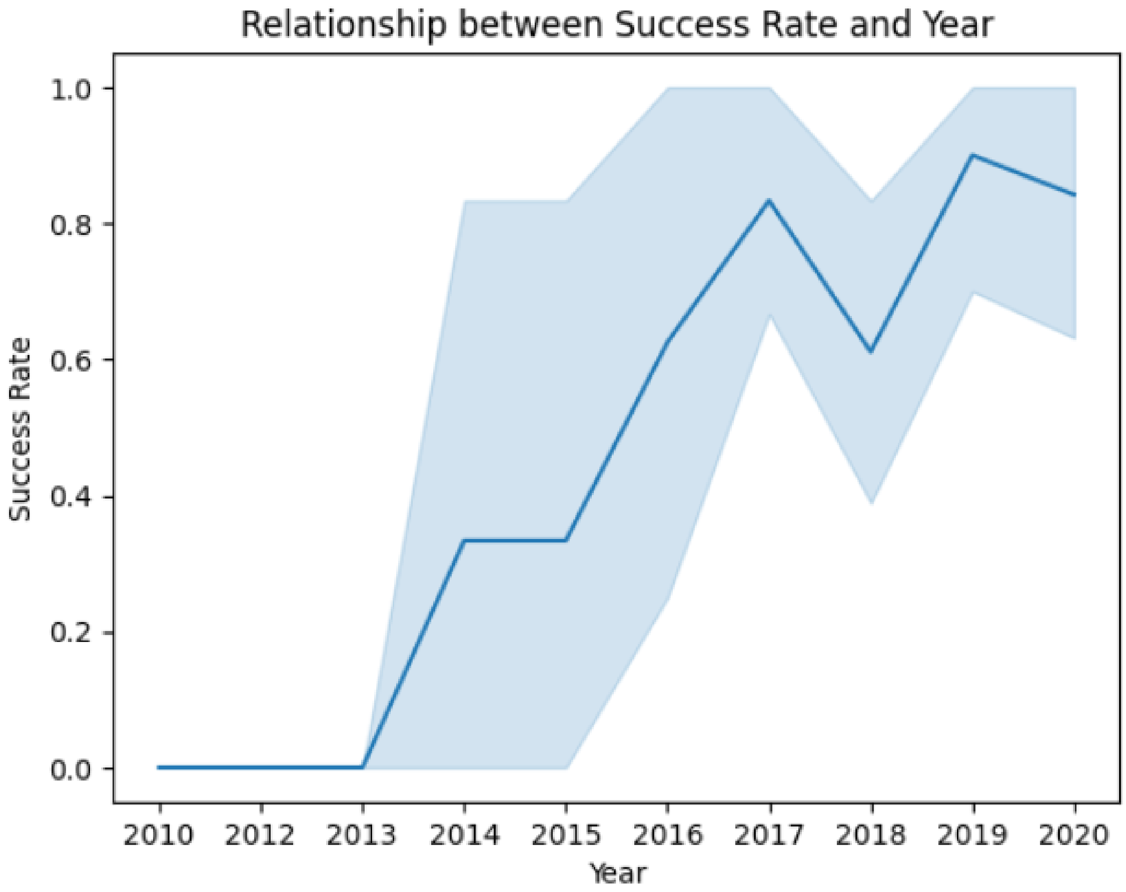
- We can observe that with heavy payloads, the successful landing are more for PO, LEO and ISS orbits.

- Success rate appears to have no obvious correlation with payload mass as some have better success rates than others.



Launch Success Yearly Trend

- We can observe that success rate since 2013 has been on the increase



All Launch Site Names

- We used the key word **DISTINCT** to show only unique launch sites from the SpaceX data.
- There are four launch sites

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT (LAUNCH_SITE) FROM SPACEXTABLE;
```

```
 * sqlite:///my_data1.db
Done.
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Launch Site Names Begin with 'CCA'

- We used the query to display 5 records where launch sites begin with `CCA`

Display 5 records where launch sites begin with the string 'CCA'

In [9]:
```
%sql SELECT * \
    FROM SPACEXTABLE \
    WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[9]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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 attempt

Total Payload Mass

- We calculated the total payload carried by boosters from NASA as 45596 using the query below

Display the total payload mass carried by boosters launched by NASA (CRS)

```
In [10]: %sql SELECT SUM(PAYLOAD_MASS__KG_) \
              AS Total_Payload_mass \
              FROM SPACEXTABLE \
              WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[10]:

Total_Payload_mass
45596

Average Payload Mass by F9 v1.1

- We calculated the average payload mass carried by booster version F9 v1.1 as 2928.4

Display average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) \
     AS Average_F9_Payload_mass \
     FROM SPACEXTABLE \
     WHERE Booster_Version = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

Average_F9_Payload_mass
2928.4

First Successful Ground Landing Date

- We observed that the dates of the first successful landing outcome on ground pad was 22nd December 2015

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT MIN(Date) \
     AS First_Date \
     FROM SPACEXTABLE \
     WHERE Landing_Outcome = 'Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

First_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- We used the **WHERE** clause to filter for boosters which have successfully landed on drone ship and applied the **AND** condition to determine successful landing with payload mass greater than 4000 but less than 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
]: %sql SELECT Booster_Version \
        FROM SPACEXTABLE \
        WHERE Landing_Outcome = 'Success (drone ship)'\
           AND PAYLOAD_MASS__KG_ > 4000  AND PAYLOAD_MASS__KG_ < 6000
```

```
 * sqlite:///my_data1.db
Done.
```

]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- We used wildcard like ‘%’ to filter for **WHERE** MissionOutcome was a success or a failure.

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(Mission_Outcome) AS SuccessOutcome\
        FROM SpaceXTABLE\
        WHERE Mission_Outcome LIKE 'Success%'
```

```
 * sqlite:///my_data1.db
Done.
```

SuccessOutcome
100

```
%sql SELECT COUNT(Mission_Outcome) AS FailureOutcome\
        FROM SpaceXTABLE\
        WHERE Mission_Outcome LIKE 'Failure%'
```

```
 * sqlite:///my_data1.db
Done.
```

FailureOutcome
1

Boosters Carried Maximum Payload

- We determined the booster that have carried the maximum payload using a subquery in the **WHERE** clause and the **MAX()** function.
- The maximum payload mass carried in this dataset is 15,600 kg. Twelve (12) separate Falcon 9 boosters carried this amount of payload mass.

[16]:

```
%sql SELECT Booster_Version,PAYLOAD_MASS__KG_ \
     FROM SPACEXTABLE \
     WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)\
    ORDER BY Booster_Version;
```

```
 * sqlite:///my_data1.db
Done.
```

[16]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

- We used a combinations of the **WHERE** clause, **LIKE**, **AND**, and **BETWEEN** conditions to filter for failed landing outcomes in drone ship, their booster versions, and launch site names for year 2015

- There were two failed landing outcomes with a drone ship in 2015. Both launched from CCAFS LC-40. One occurred in January and the other in April

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [18]:

```
task_9 = '''
        SELECT BoosterVersion, LaunchSite, LandingOutcome
        FROM SpaceX
        WHERE LandingOutcome LIKE 'Failure (drone ship)'
            AND Date BETWEEN '2015-01-01' AND '2015-12-31'
        '''
create_pandas_df(task_9, database=conn)
```

Out[18]:

	boosterversion	launchsite	landingoutcome
0	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- We found out that the most common landing outcome was ‘not attempted’ and least was ‘Failure (Parachute).

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad))

In [19]:
```
task_10 = '''
        SELECT LandingOutcome, COUNT(LandingOutcome)
        FROM SpaceX
        WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
        GROUP BY LandingOutcome
        ORDER BY COUNT(LandingOutcome) DESC
        '''
create_pandas_df(task_10, database=conn)
```

Out[19]:

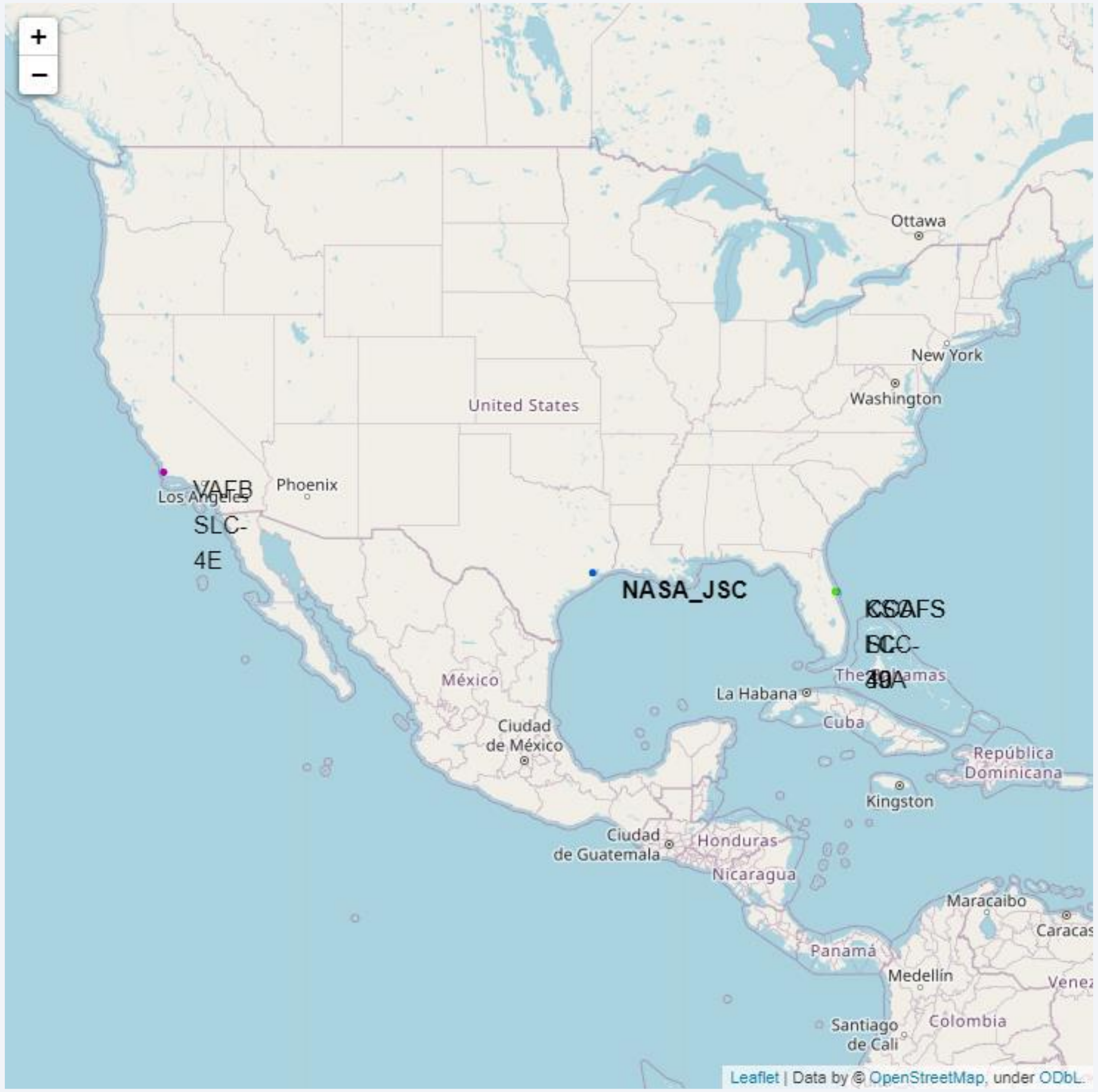
	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

Section 4

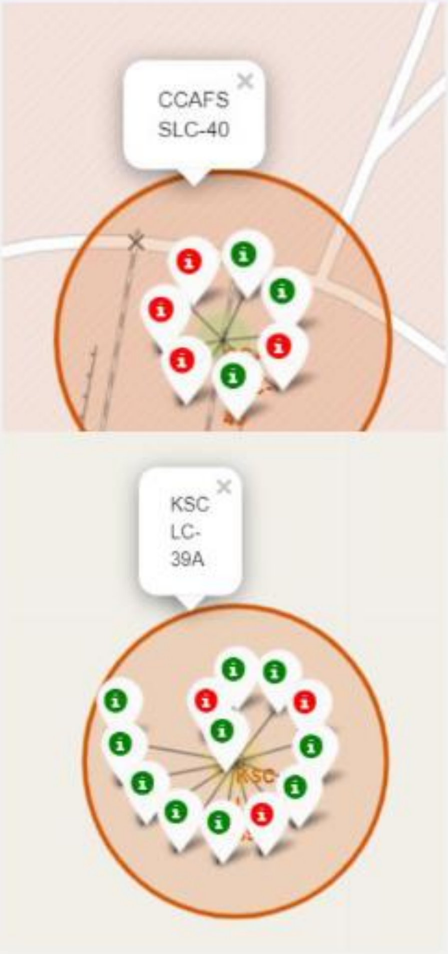
Launch Sites Proximities Analysis

All launch sites global map markers

- VAFB SLC-4E (California, USA)
  - Vandenberg Air Force Base Space Launch Complex 4E
- KSC LC-39A (Florida, USA)
  - Kennedy Space Center Launch Complex 39A
- CCAFS LC-40 (Florida, USA)
  - Cape Canaveral Air Force Station Launch Complex 40
- CCAFS SLC-40 (Florida, USA)
  - Cape Canaveral Air Force Station Space Launch Complex 40



Markers showing launch sites with color labels



Florida Launch Sites

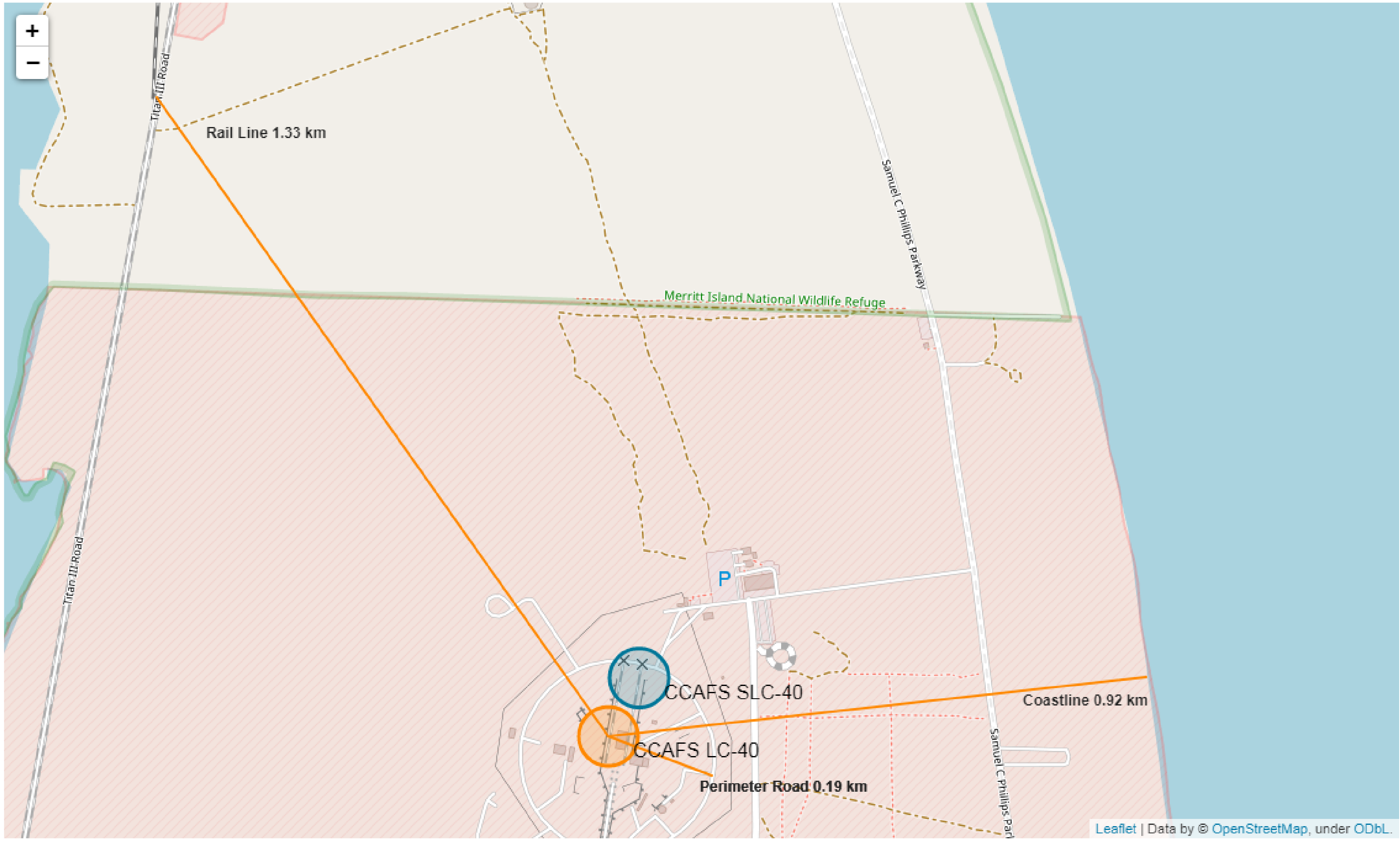
Green Marker shows successful Launches and Red Marker shows Failures

California Launch Site

37

Launch Site distance to landmarks

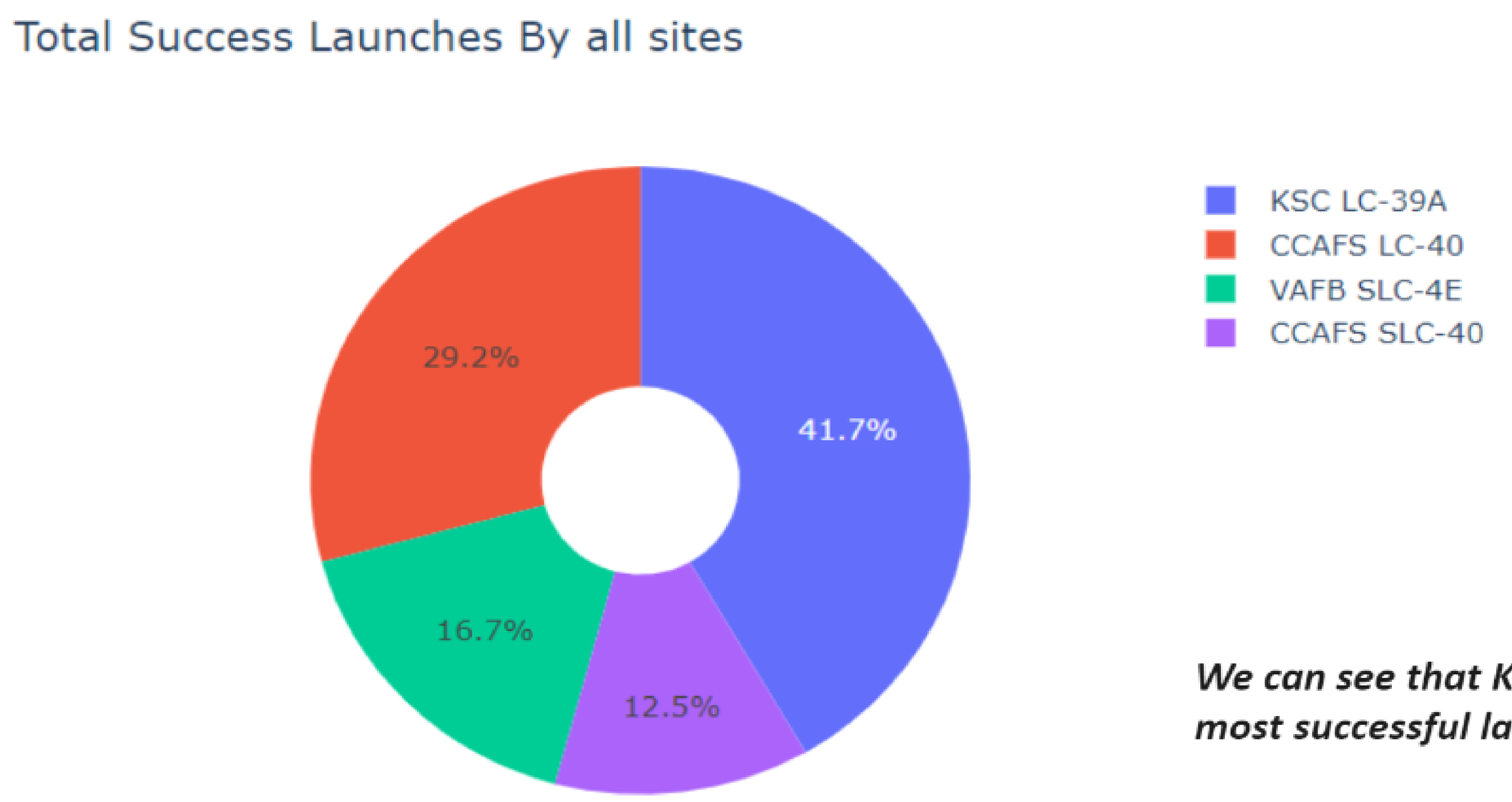
- The CCAFS LC-40 and CCAFS SLC-40 launch sites have coordinates that are close to being, but are not exactly, right on top of each other.
- The perimeter road around CCAFS LC-40 is 0.19 km away from the launch site coordinates.
- The coastline is 0.92 km away from CCAFS LC-40.
- The rail line is 1.33 km away from CCAFS LC-40.



Section 5

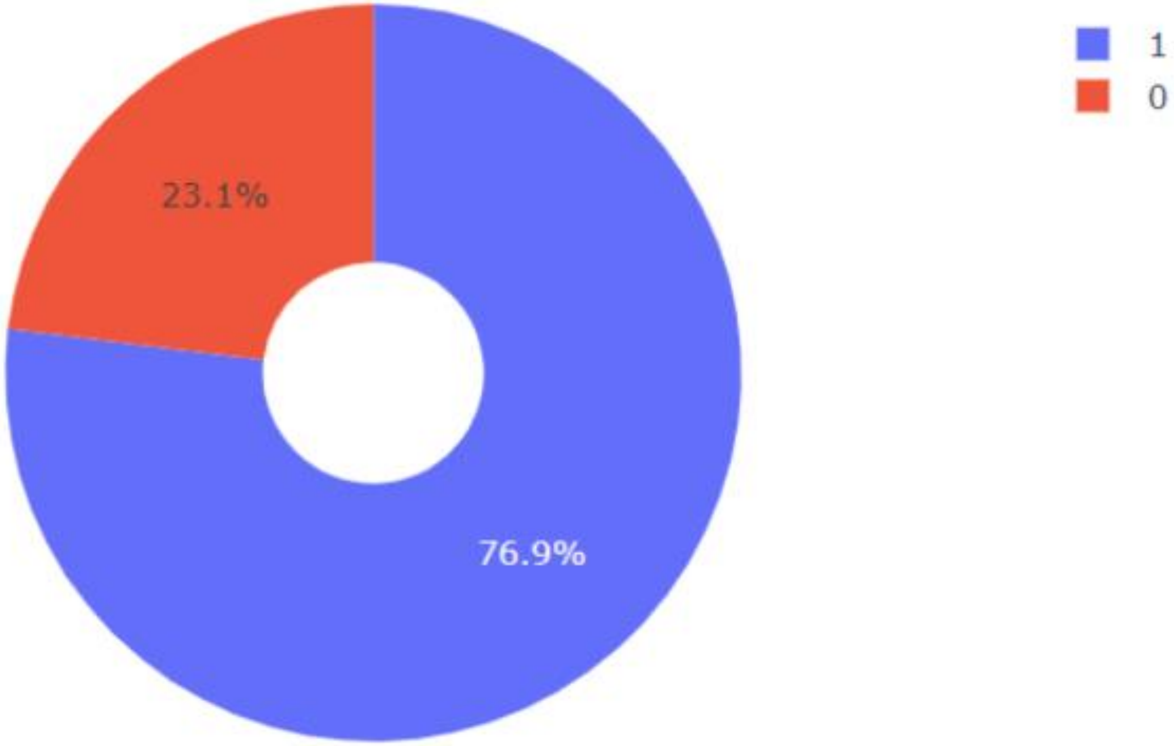
Build a Dashboard with Plotly Dash

Pie chart showing the success percentage achieved by each launch site



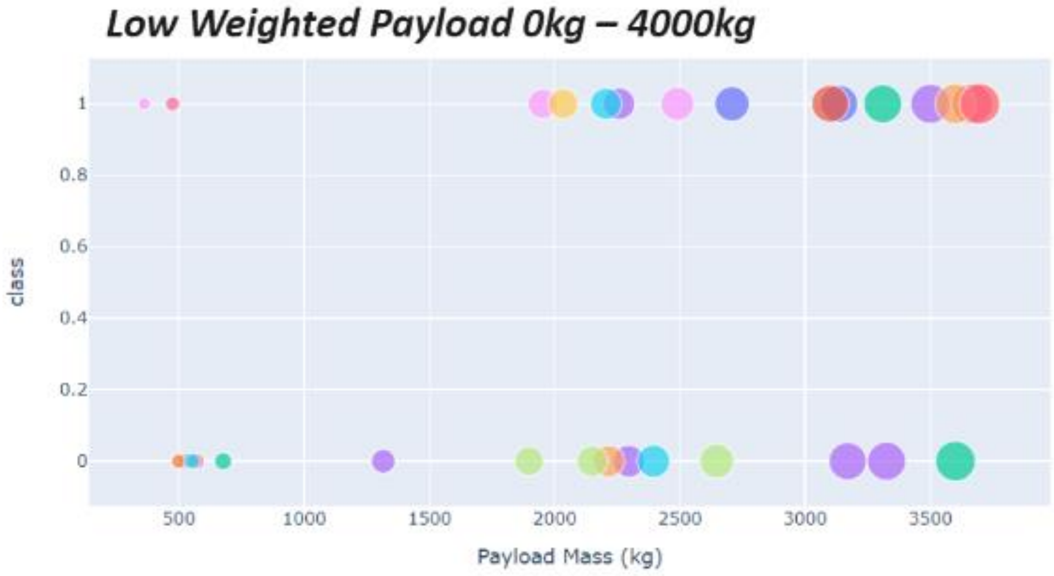
We can see that KSC LC-39A had the most successful launches from all the sites

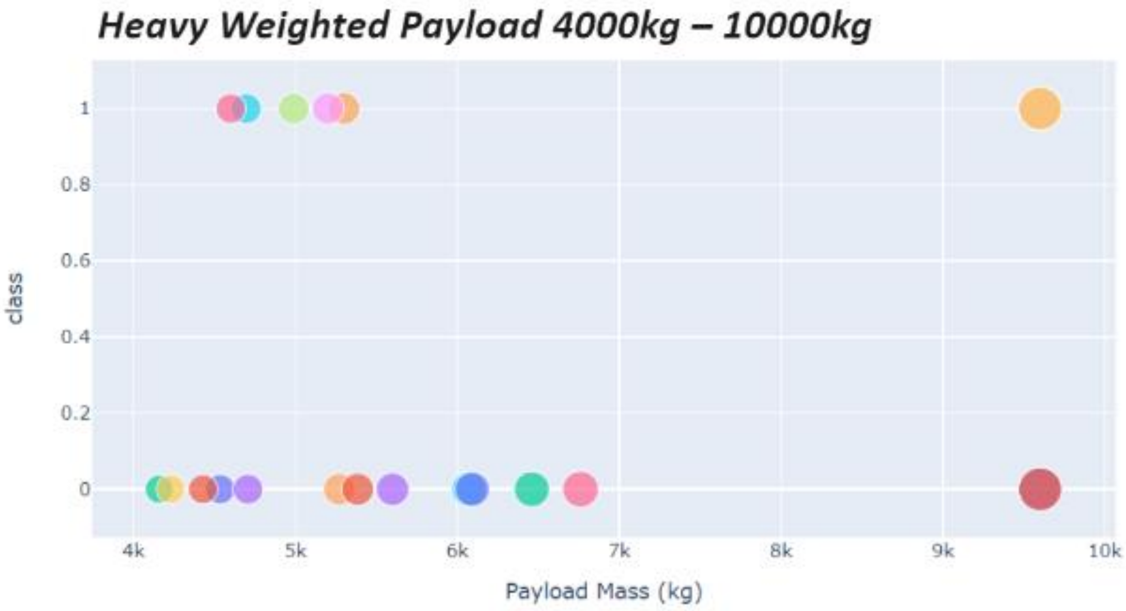
Pie chart showing the Launch site with the highest launch success ratio



KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Scatter plot of Payload vs Launch Outcome for all sites, with different payload selected in the range slider





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 6

Predictive Analysis (Classification)

Classification Accuracy

- All models performed equally well

Find the method performs best:

[41]:

```python
Results = pd.DataFrame({'Method' : ['Test Data Accuracy']})

knn_accuracy=knn_cv.score(X_test, Y_test)
Decision_tree_accuracy=tree_cv.score(X_test, Y_test)
SVM_accuracy=svm_cv.score(X_test, Y_test)
Logistic_Regression=logreg_cv.score(X_test, Y_test)

Results['Logistic_Reg'] = [Logistic_Regression]
Results['SVM'] = [SVM_accuracy]
Results['Decision Tree'] = [Decision_tree_accuracy]
Results['KNN'] = [knn_accuracy]


print('They all have similar test results', Results.transpose())
```

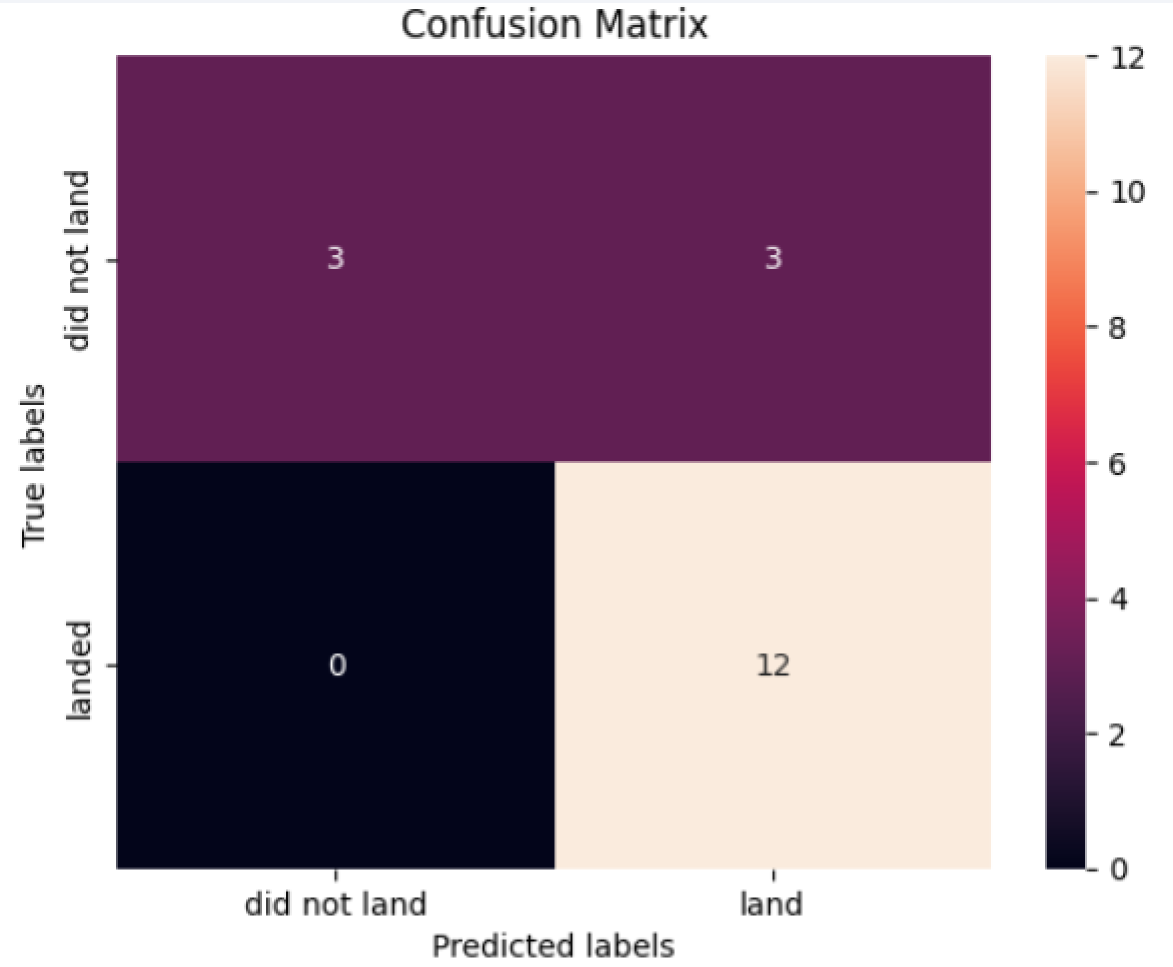
```
They all have similar test results                                   0
Method              Test Data Accuracy
Logistic_Reg                  0.833333
SVM                           0.833333
Decision Tree                 0.833333
KNN                           0.833333
```

Confusion Matrix

- The confusion matrix for the classifiers shows that they can distinguish between the different labels. The major problem is the false positives .i.e., unsuccessful landing marked as successful landing by the classifier.
- Confusion matrices can be read as:

True Negative	**False Positive**
False Negative	**True Positive**

- Prediction Breakdown:
  - 12 True Positives and 3 True Negatives
  - 3 False Positives and 0 False Negatives



Conclusions

We can conclude that:

- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- SpaceX does not have a perfect track record of Falcon 9 first stage landing outcomes.
- SpaceX's Falcon 9 first stage landing outcomes have been trending towards greater success since 2013 as more launches are made.
- The machine learning models can be used to predict future SpaceX Falcon 9 first stage landing outcomes.
- The larger the flight amount at a launch site, the greater the success rate at a launch site.

Appendix

▪Initial Data Sets

- **SpaceX API (JSON):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- **Wikipedia (Webpage):** https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- **SpaceX (CSV):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01
- **Launch Geo (CSV):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
- **Launch Dash (CSV):** https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv

Appendix

- Jupyter Notebooks and Dashboard Python File
  - **GitHub URL (Data Collection):** https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/spacex-data-collection-api.ipynb
  - **GitHub URL (Web Scraping):** https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceXwebscraping.ipynb
  - **GitHub URL (Data Wrangling):** https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/spacex-Data%20wrangling.ipynb
  - **GitHub URL (EDA with SQL):** https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/eda-sql-coursera_sqllite.ipynb
  - **GitHub URL (EDA with Data Visualization):** https://https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX%20Exploratory%20Analysis.ipynb
  - **GitHub URL (Folium Maps):** https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_launch_site_location.ipynb
  - **GitHub URL (Machine Learning):** https://github.com/jamesenet/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Thank you!