IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rita Ramos
03/08/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Colletion
  - Data Wrangling
  - Explonatory Analysis
  - Data Visualizations and analytics
- Summary of all results
  - Predictive Analysis (Classification)
  - Results

Introduction

- Project background and context

  In this project, we will predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 SpaceX for a rocket launch.

- Problems you want to find answers

  1. Identifying Critical Factors for Successful Rocket Landings
  2. Understanding Interactions Among Features
  3. Defining Optimal Operating Conditions

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from 2 sources:
    - SpaceXAPI(https://api.spacexdata.com/v4/rockets/)
    - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - The data was cleaned by replacing missing values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

Methodology

Executive Summary

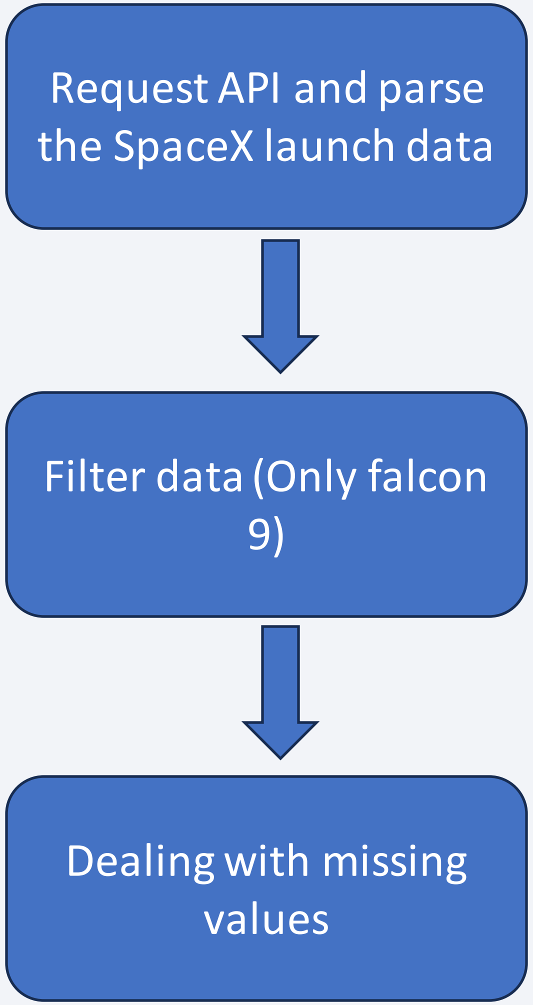
- Perform predictive analysis using classification models
    - With the predicted data set, 4 machine learning models were used to discover which was best to solve the problem.

Data Collection

- Datasets were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

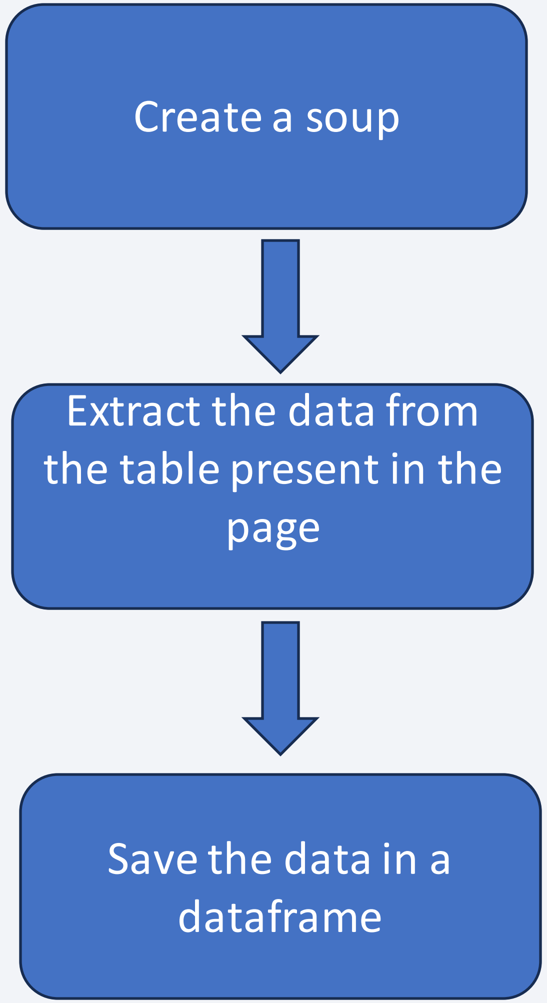
Data Collection – SpaceX API

- GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/1)jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/2)jupyter-labs-webscraping.ipynb

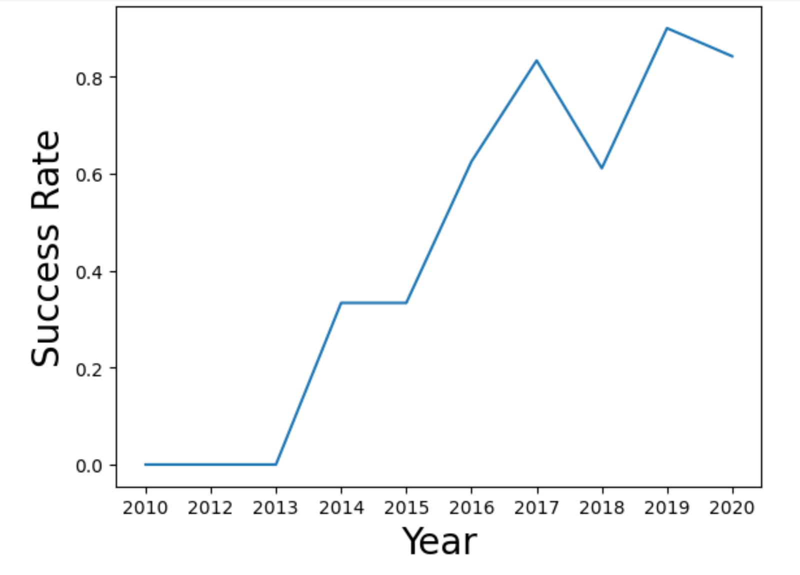


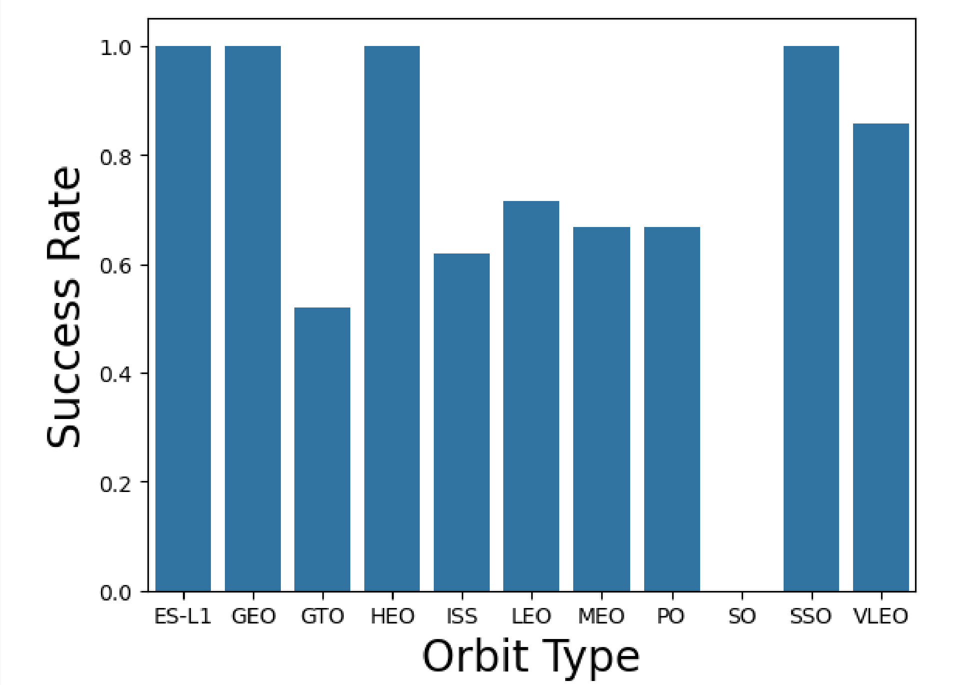
Data Wrangling

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurence of mission outcome of the orbits
- Create a landing outcome label from Outcome column

GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/3)labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization





- GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/5)jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

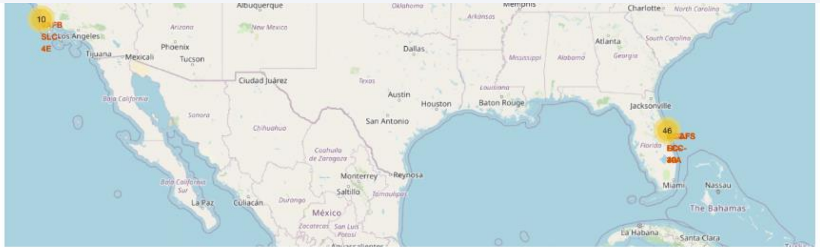
EDA with SQL

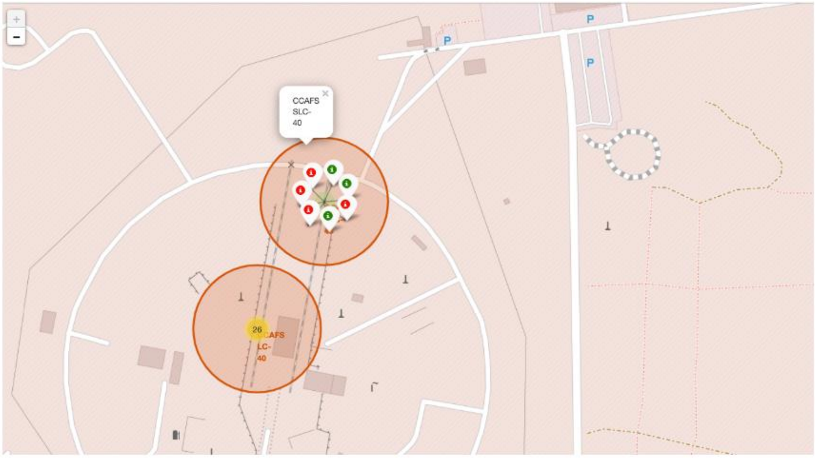
- The following SQL queries were performed:
  - Names of the unique launch sites in the space mission;
  - Top 5 launch sites whose name begins with the string 'CCA’;
  - Total pay load mass carried by boosters launched by NASA (CRS);
  - Average payload mass carried by booster version F9 v1.1;
  - Date when the first successful landing outcome in ground pad was achieved;
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
  - Total number of successful and failure mission outcomes;
  - Names of the booster versions which have carried the maximum payload mass;
  - Failed landing out comes in droneship, their booster versions, and launch site names for in year 2015; and
  - Rank of the count of landing outcomes (such as Failure (droneship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.
- GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/4)jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities

GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/6)lab_jupyter_launch_site_location.jupyterlite.ipynb





Build a Dashboard with Plotly Dash

- An interactive dashboard with Plotly dash with pie charts showing the total launches by a certain sites and scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.
- GitHub URL:https://github.com/Lancome25/capstone_coursera/blob/main/7)spacex_dash_app.py

Predictive Analysis (Classification)

- Create a NumPy array from the column Class
- Standardize the data in X
- Use the function train_test_split to split the data X and Y into training and test data
- Create 4 different models and evaluate
    - Logistic regression model
    - Support vector machine
    - Decision tree
    - Knearest neighbors

GitHub URL: https://github.com/Lancome25/capstone_coursera/blob/main/8)SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- Exploratory data analysis results
  - Space X uses 4 different launch sites
  - The average payload of F9 v1.1 booster is 2,928 kg
  - The number of landing outcomes became as better as years passed.

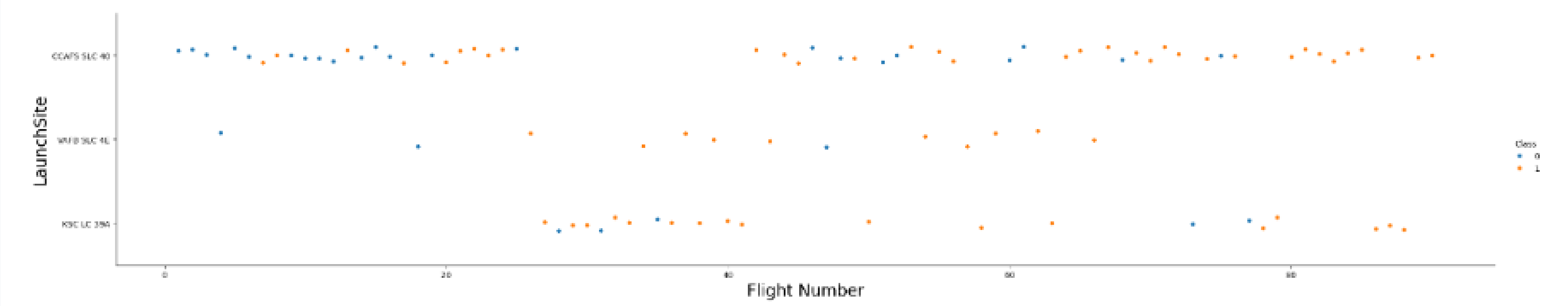
- Predictive analysis results
  - Accuracy for Logistics Regression method: 0.8333333333333334
  - Accuracy for Support Vector Machine method: 0.8333333333333334
  - Accuracy for Decision tree method: 0.7222222222222222
  - Accuracy for K nearst neighbors method: 0.8333333333333334

Section 2

Insights drawn from EDA

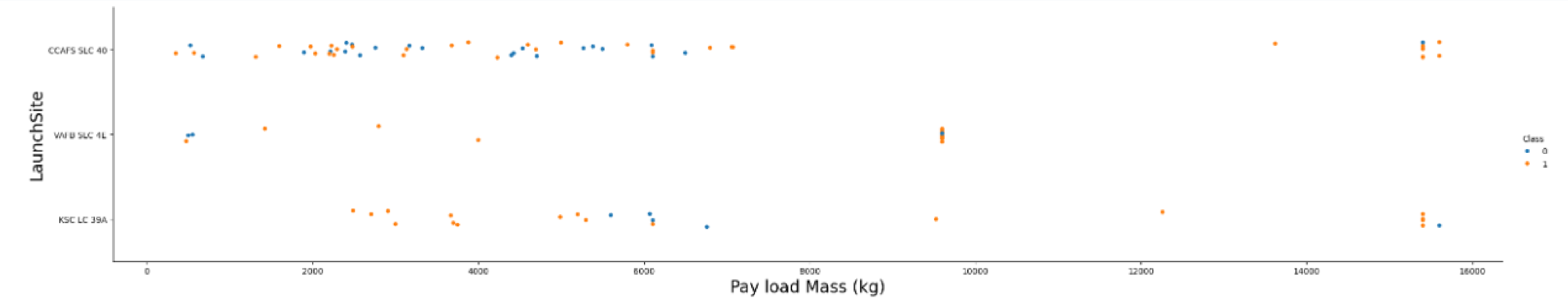
Flight Number vs. Launch Site

- Plot of Flight Number vs. Launch Site



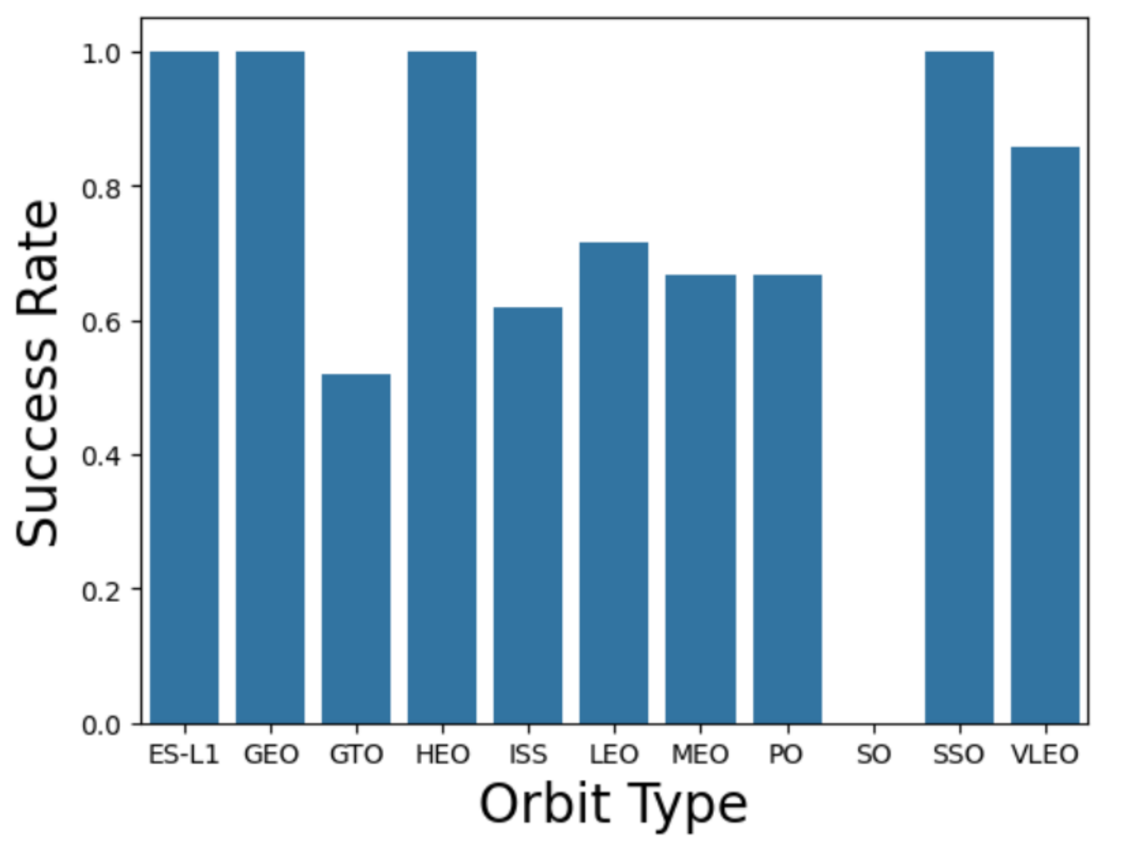
Payload vs. Launch Site

- Plot of Payload vs. Launch Site



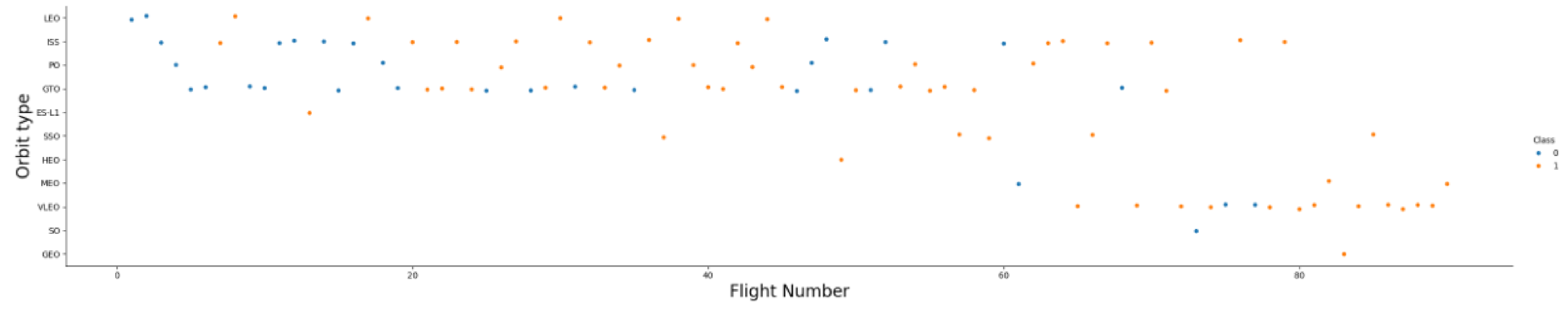
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type



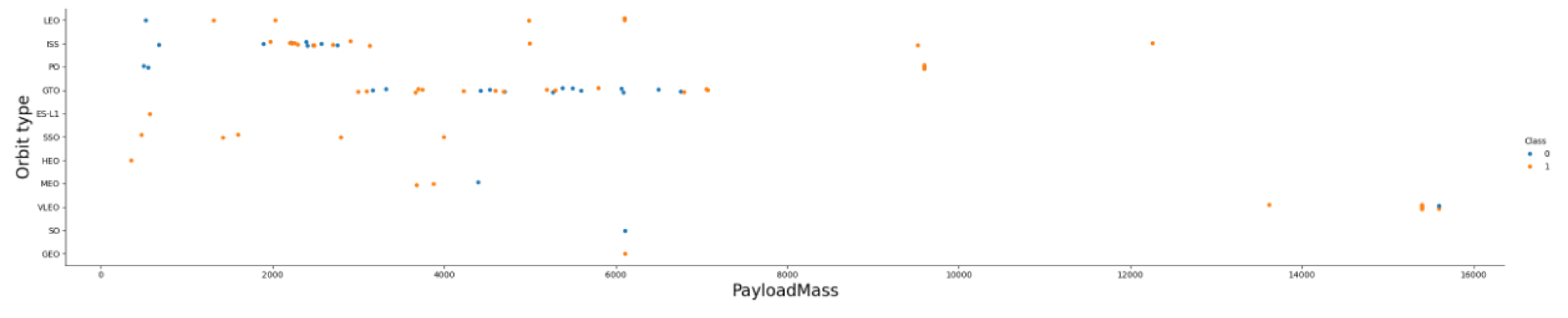
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type



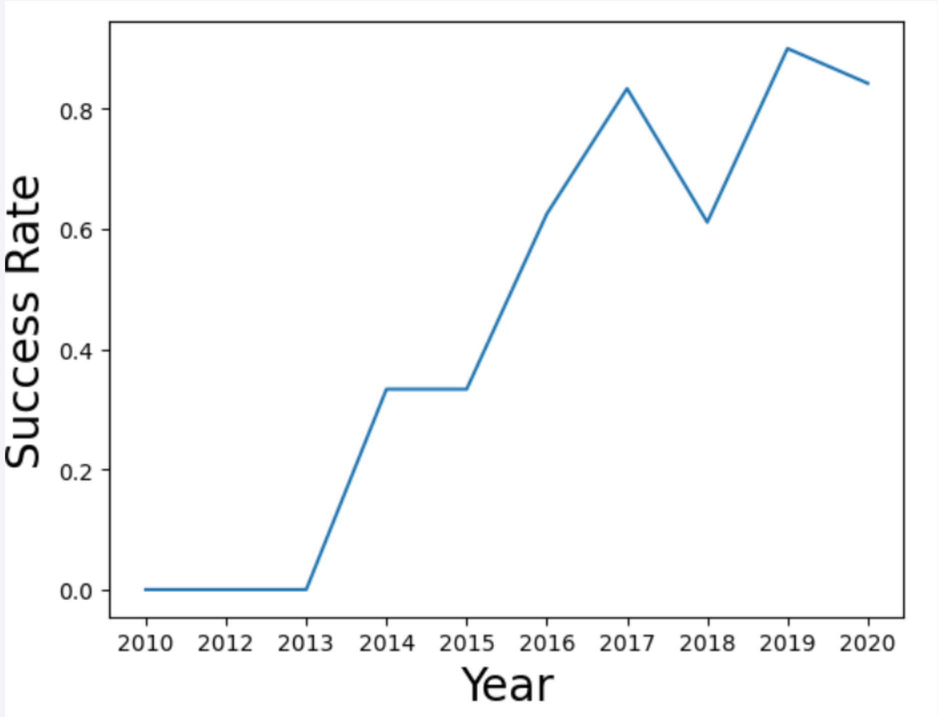
Payload vs. Orbit Type

- Scatter point of payload vs. orbit type



Launch Success Yearly Trend

- Line chart of yearly average success rate



All Launch Site Names

- Find the names of the unique launch sites

```
In [13]: %sql SELECT DISTINCT Launch_Site from SPACEXTBL

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

Out[13]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%sql select * from SPACEXTBL where Launch_Site LIKE 'CCA%' LIMIT 5
```

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

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

- Calculate the total payload carried by boosters from NASA

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

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer='NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version='F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
]: %sql select min(Date) from SPACEXTBL where Landing_Outcome='Success (ground pad)'

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

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql select Booster_Version from SPACEXTBL WHERE Landing_Outcome='Success (drone ship)' and  PAYLOAD_MASS__KG_ between 4000
```

```
 * 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

- Calculate the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(*) FROM SPACEXTBL WHERE MISSION_OUTCOME LIKE '%Success%' OR MISSION_OUTCOME LIKE '%Failure%'
```

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

COUNT(*)
101

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

Booster_Version
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

2015 Launch Records

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

```
%sql select substr(Date, 6,2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL where Landing_Outcome=
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%sql SELECT "Date", Landing_Outcome, COUNT(Landing_Outcome) as COUNT FROM SPACEXTBL \
    WHERE "Date" BETWEEN '2010-06-04' and '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(Landing_Outcome) DESC
```

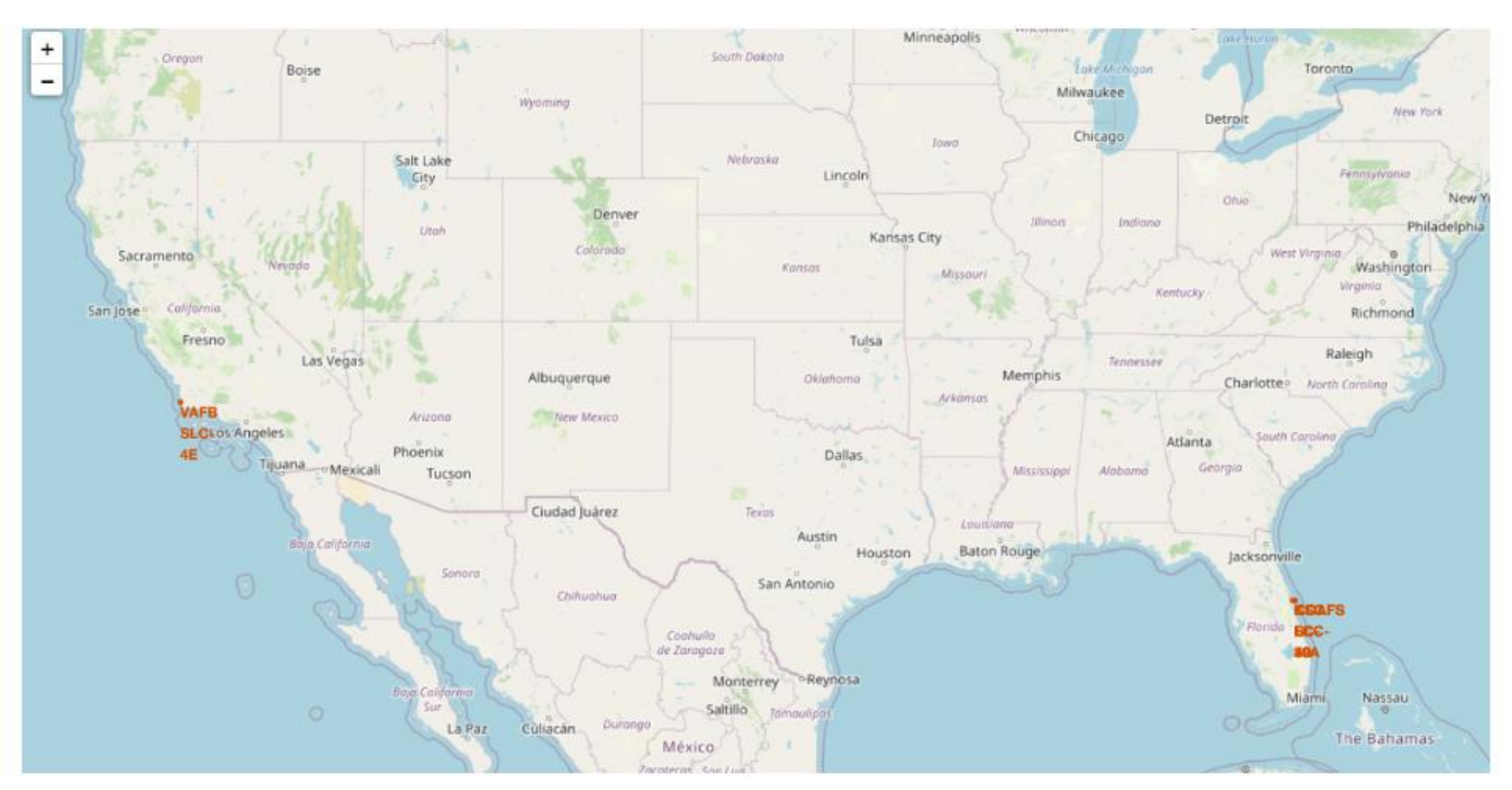
```
 * sqlite:///my_data1.db
Done.
```

Date	Landing_Outcome	COUNT
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2010-06-04	Failure (parachute)	2
2015-06-28	Precluded (drone ship)	1

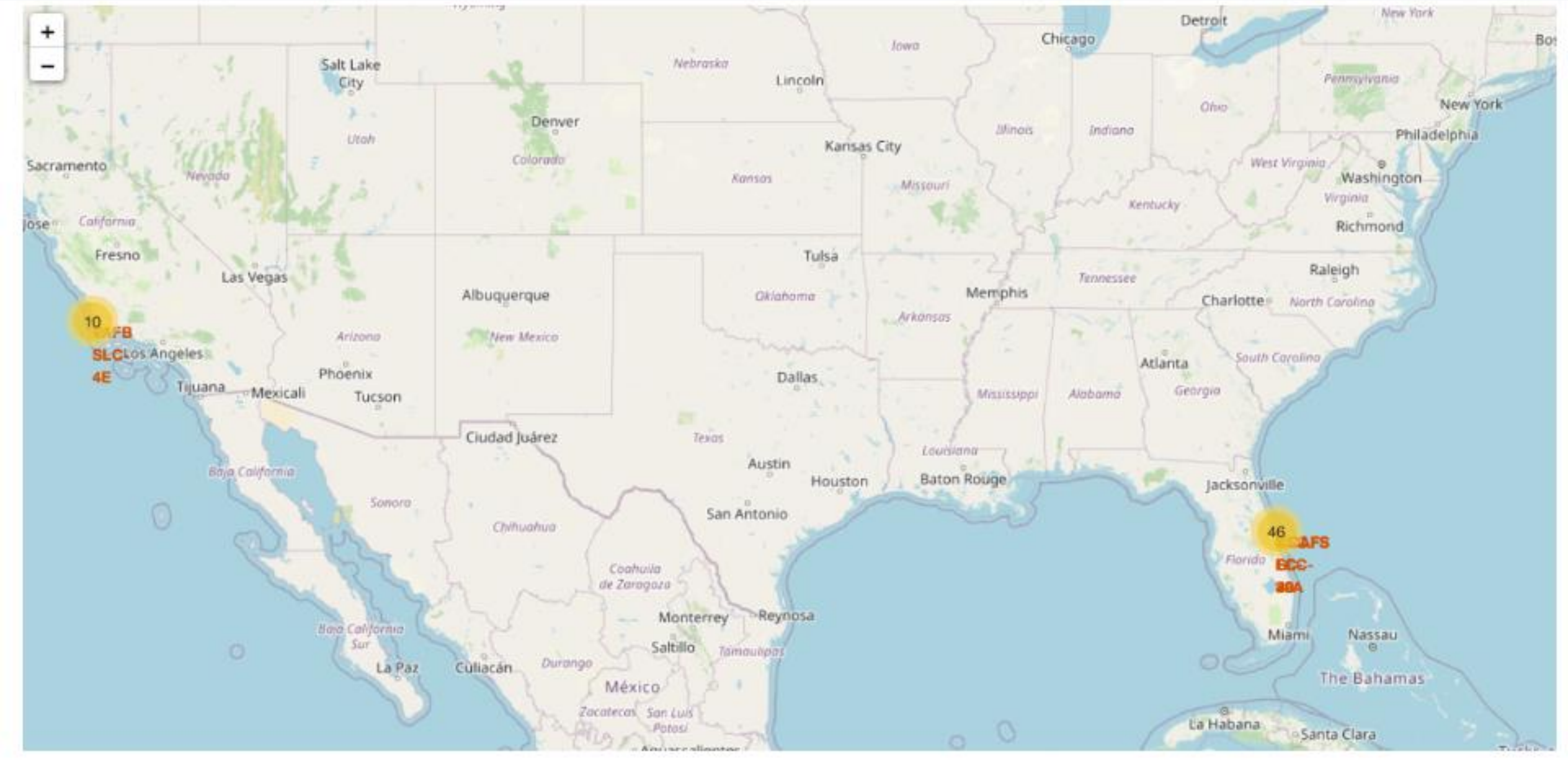
Section 3

Launch Sites Proximities Analysis

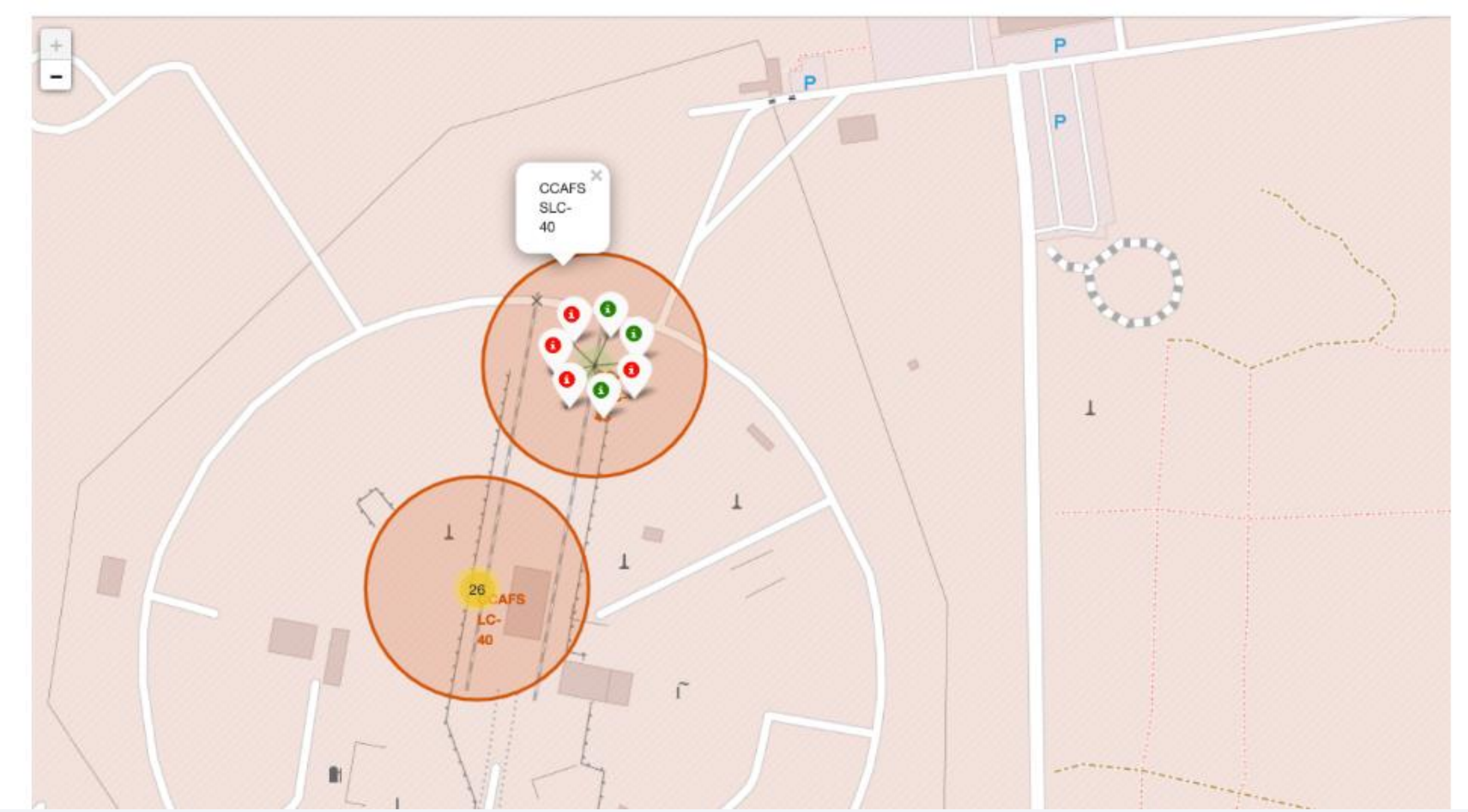
All launch sites global map markers



Marker cluster in the site map



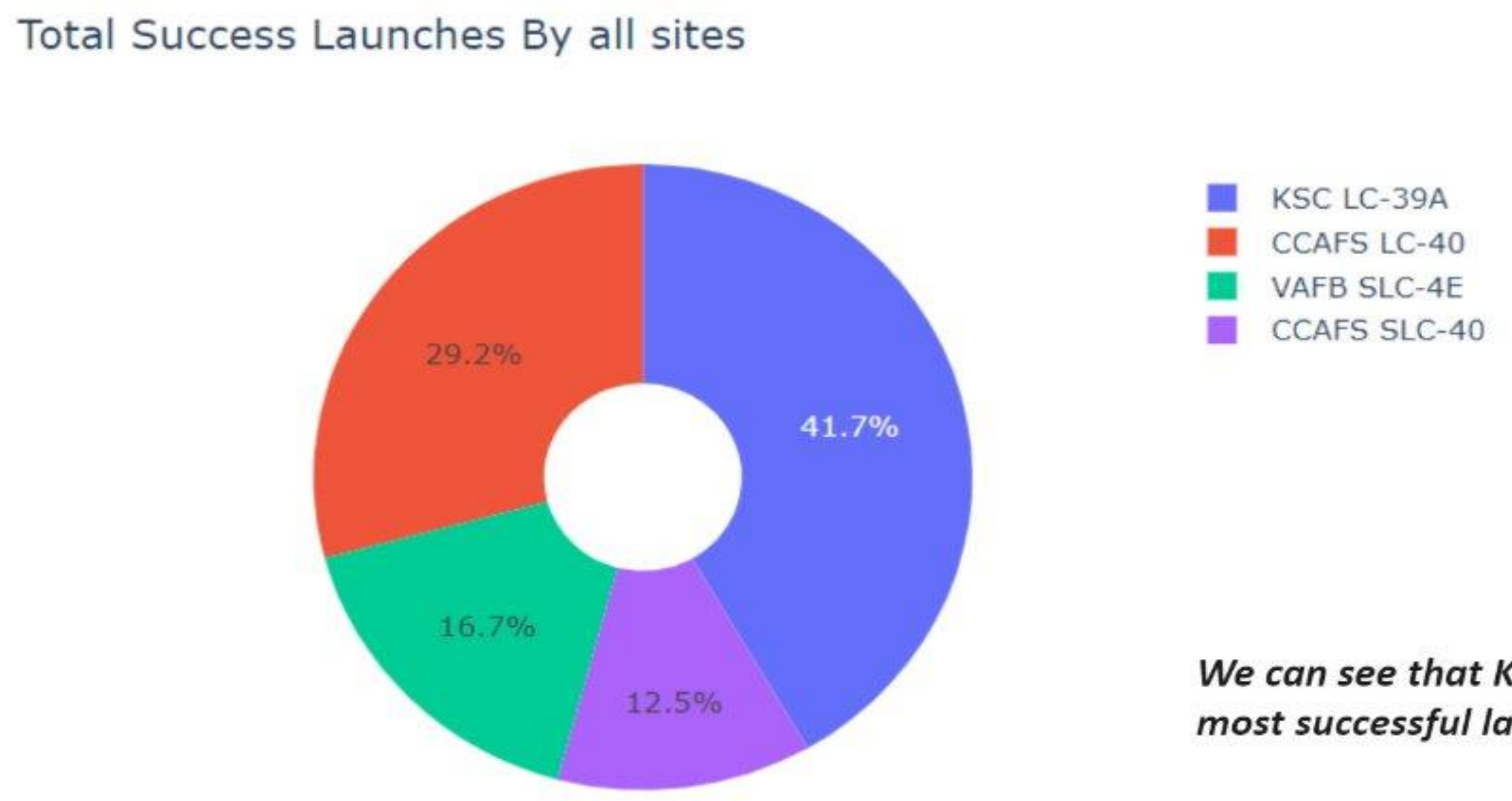
Marker cluster in the site map Green and Red (Success and Failure)



Section 4

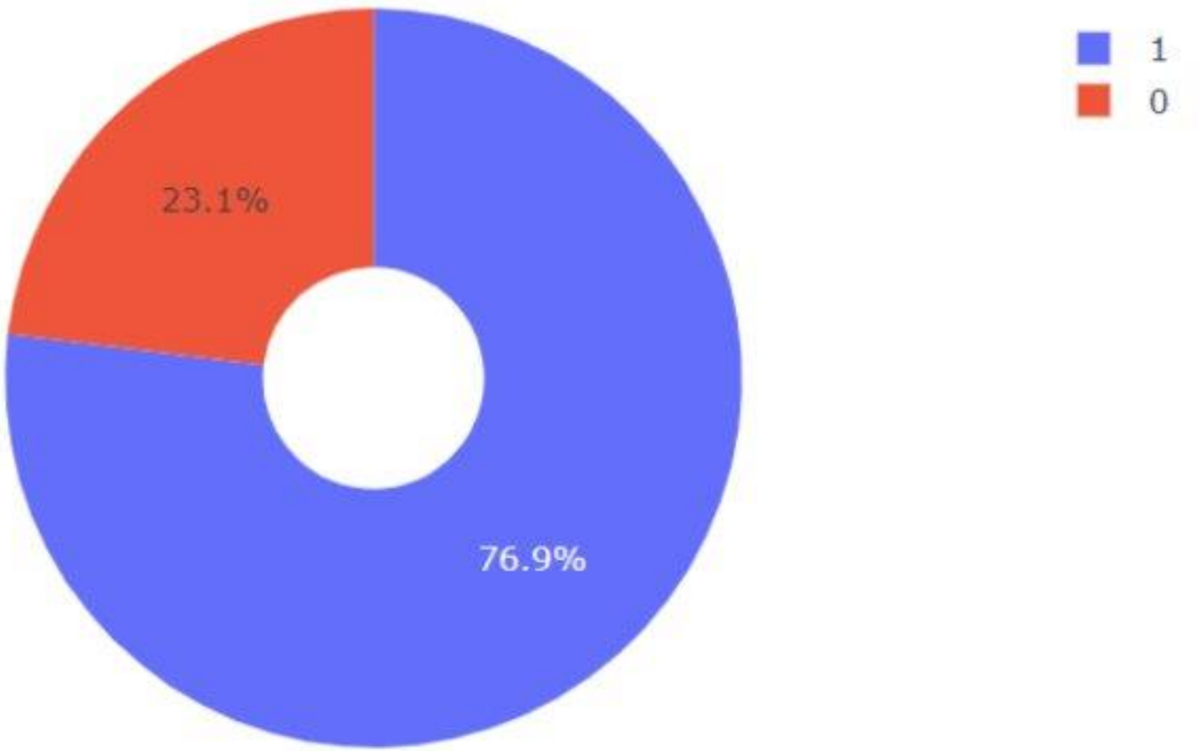
Build a Dashboard with Plotly Dash

Total Success Launches by all sites



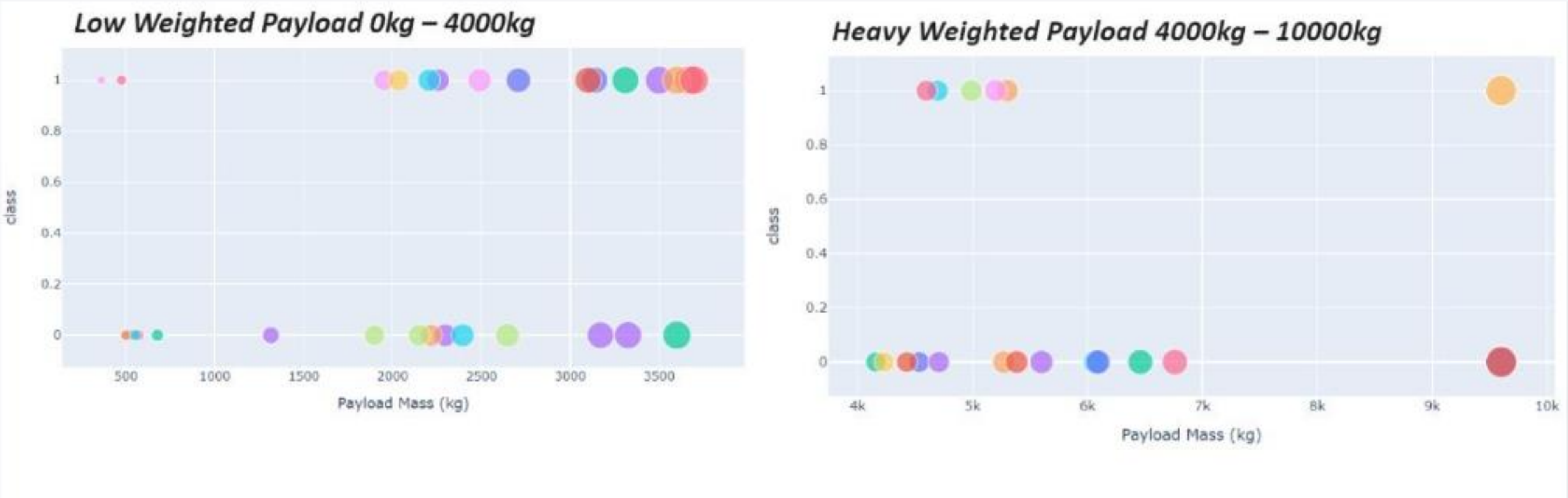
We can see that KSC LC-39A had the most successful launches from all the sites

Sucess Ratings



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

Payload Mass vs Class



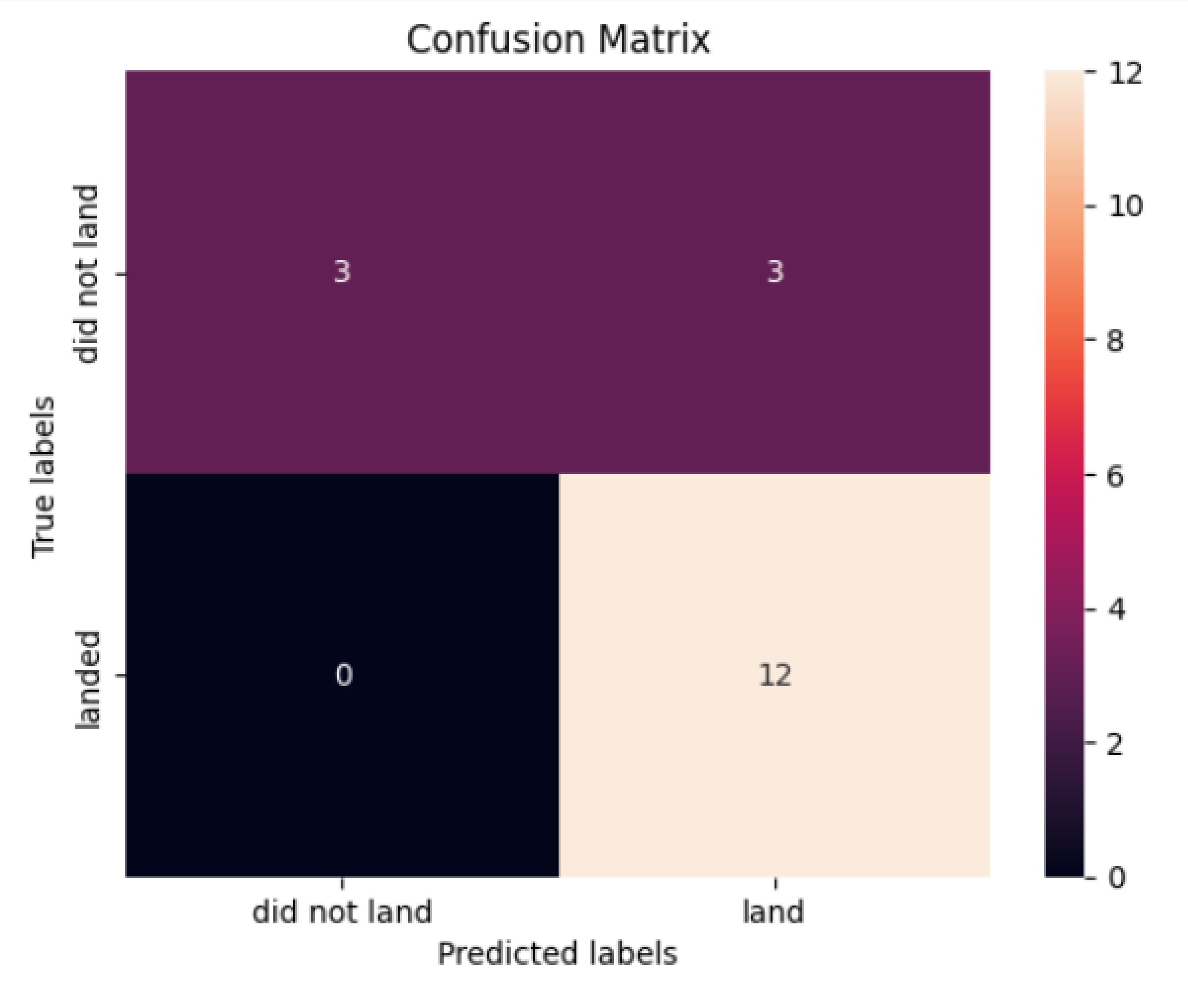
Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.7222222222222222
Accuracy for K nearsdt neighbors method: 0.8333333333333334
```

Confusion Matrix



Conclusions

- It possible to conclude that using any of these 3 models (Logistic regression, SVM, K nearest) we can predict with an accuracy of 83% if the spacex land successfully.

Thank you!