IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Pradeep Kumar J K
22 June 2023

Outline

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

Executive Summary

Summary of methodologies:

- Publicly available data has been collected through online resources, namely ,Wikipedia and the SpaceX API.
- The data has been pre-processed before being converted into a Pandas Dataframe for further manipulation.
- Missing data has been replaced by the mean value
- Exploratory data analysis using SQL has been carried out to find out important summaries
- Data Visualisation was used to discover important variables that could determine the landing success of a launch
- These important variables were then used as input to several Machine Learning models: Logistic Regression, k-Nearest Neighbours, Decision Tree and Support Vector Machine
- The models were fit to the training data and a comparison was made between the accuracy on the test data

Summary of all results

- All models used were comparable in their accuracy on the test data. The accuracy in predicting the landing success rate was around 83%

Introduction

Project Background and Context

- This project is a comprehensive study of SpaceX’s launch data for the Falcon 9 rocket.
- The variable of interest is the landing success rate of the first stage.
- The success of the landing determines the price of a launch

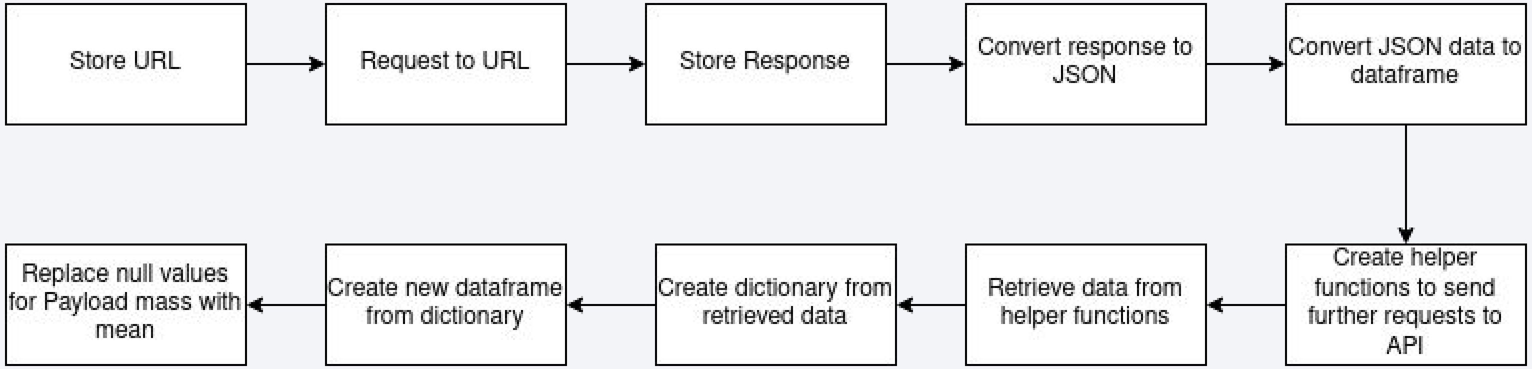
Problems you want to find answers

- The project aims to discover the independent variables that allow a machine learning model to predict the landing outcome
- A comparison has to be made between different model to determine the suitable one for future predictions
- By determining the success rate and, subsequently, the price of the launch, the model will allow for the better management of future launches to maximise success rates.

Section 1

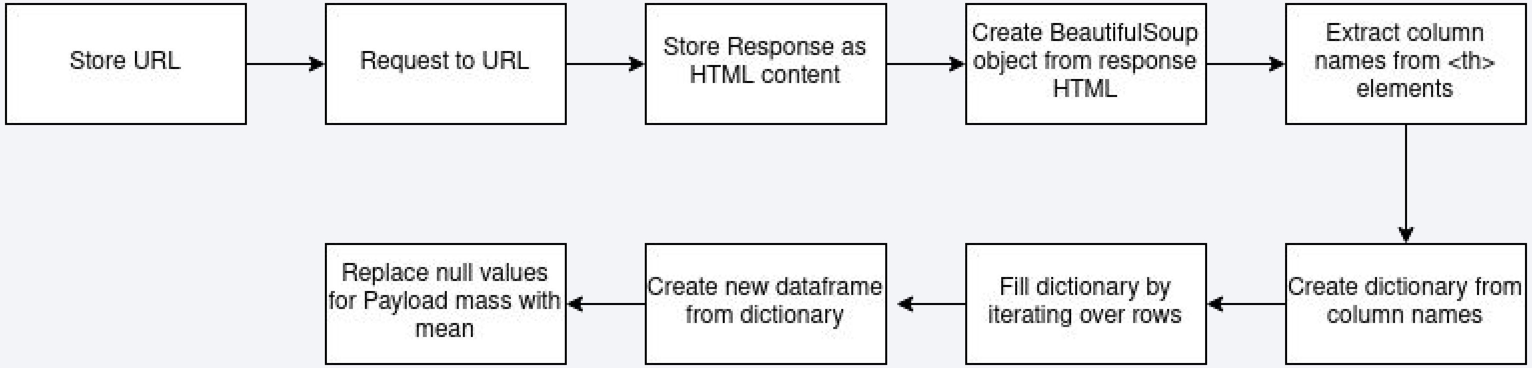
Methodology

Data Collection – SpaceX API



- GitHub URL of the completed SpaceX API calls :
- https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/Data%20Collection%20Api%20.ipynb

Data Collection - Scraping



- GitHub URL of the completed web scraping notebook:
  - https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- Irrelevant columns containing ID numbers were either dropped or helper functions were utilised to retrieve information using these ID values.
- The data collected was robust.
- There were only a few missing values for the Payload Mass (kg) column
- These values were replaced by the mean of the Payload Mass (kg) column

EDA with Data Visualization

- The following charts were plotted to discover relationships between variables:
    - Flight Number vs Launch Site with landing result overlayed
    - Payload vs Launch Site with landing result overlayed
    - Success Rate vs Orbit Type
    - Flight Number vs Orbit Type with landing result overlayed
    - Payload vs Orbit Type with landing result overlayed
    - Launch Success Yearly Trend - Line Plot

- GitHub URL of completed EDA with data visualization notebook:
    - https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/EDA%20with%20Visualization.ipynb

EDA with SQL

- Data retrieved using SQL:
  - Names of all launch sites
  - Total payload carried by boosters launched by NASA
  - Average payload carried by booster version F9
  - Date of first successful landing
  - Boosters with success in drone ship and payload mass between 4000 and 6000 kg
  - Total number of successful and failed mission outcomes
  - Booster versions that have carried the maximum payload mass
  - Landing Outcome categories ranked by total for a given time period
  - failed landing outcomes in drone ship for the year 2015
- GitHub URL of completed EDA with SQL notebook
  - https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Folium was used to find out some geographical insights about the data
- Launch sites were marked on the map
- The successful and failed launches for each launch site were also marked
- Proximities between launch site and features such as the coast and equator line were found to discover any impact that these may have on the success

- Circles, Markers and MArker Clusters were used for the tasks

- GitHub URL of completed interactive map with Folium map:

    - https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium-checkpoint.ipynb

Build a Dashboard with Plotly Dash

- The following plots/graphs were added to the interactive dashboard
  - Pie chart showing success rate for each site
  - Pie chart showing Success and Failure show for a specific site
  - Payload vs Launch Outcome scatter plot for all sites
- GitHub URL completed Plotly Dash lab:
  - https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/Data%20wrangling%20.ipynb

Predictive Analysis (Classification)

- The following methodology was used in predictive analysis:
    - Separate Independent variables from Dependent Variable (Class)
    - Split into training and testing data
    - define parameters dictionary
    - Initialise each model, create GridSearchCV object to run the model using all specified parameters
    - find the best parameters
    - Calculate accuracy of prediction on test data

- GitHub URL of completed predictive analysis lab:
    - https://github.com/pradeepjk28/Applied-Data-Science-Capstone/blob/main/Machine%20Learning%20Prediction-checkpoint.ipynb

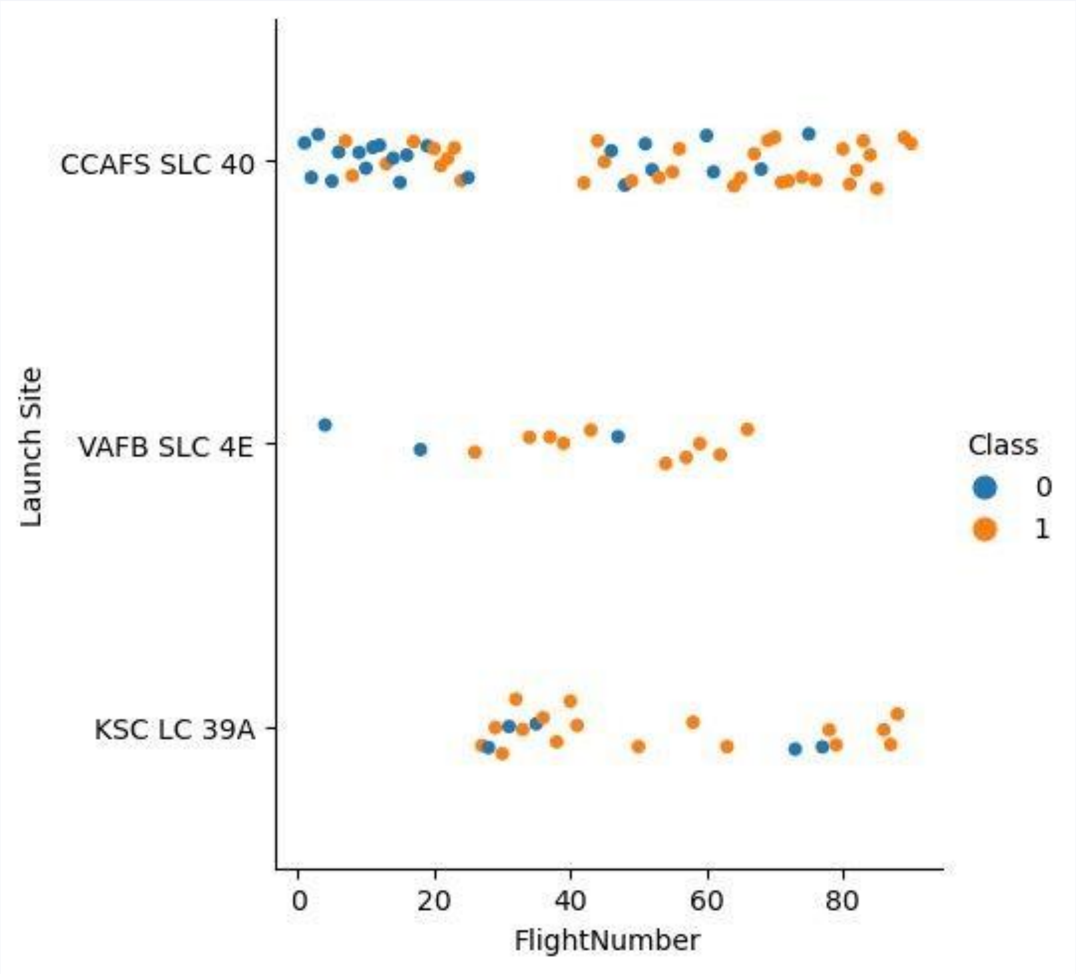
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site

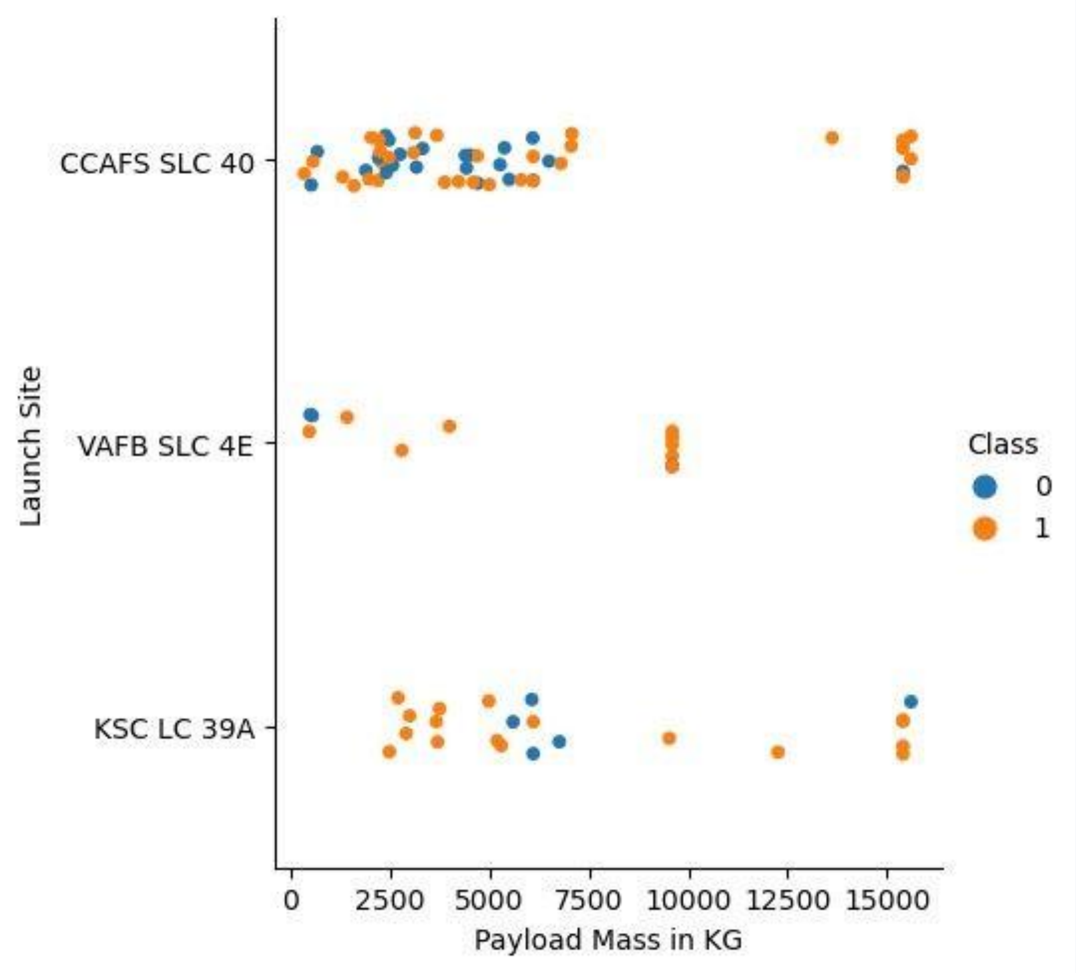
- Outcome is overlayed



Payload vs. Launch Site

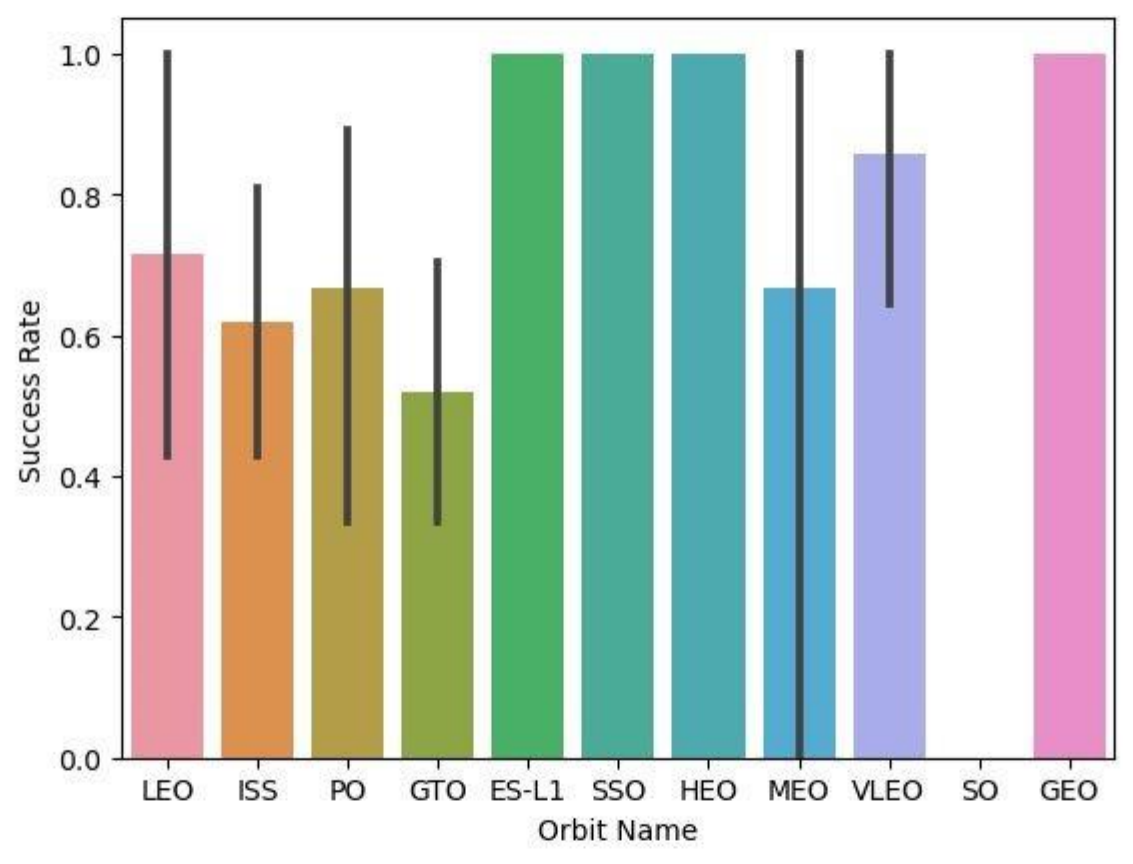
- Scatter plot of Payload vs. Launch Site

- Outcome is overlayed



Success Rate vs. Orbit Type

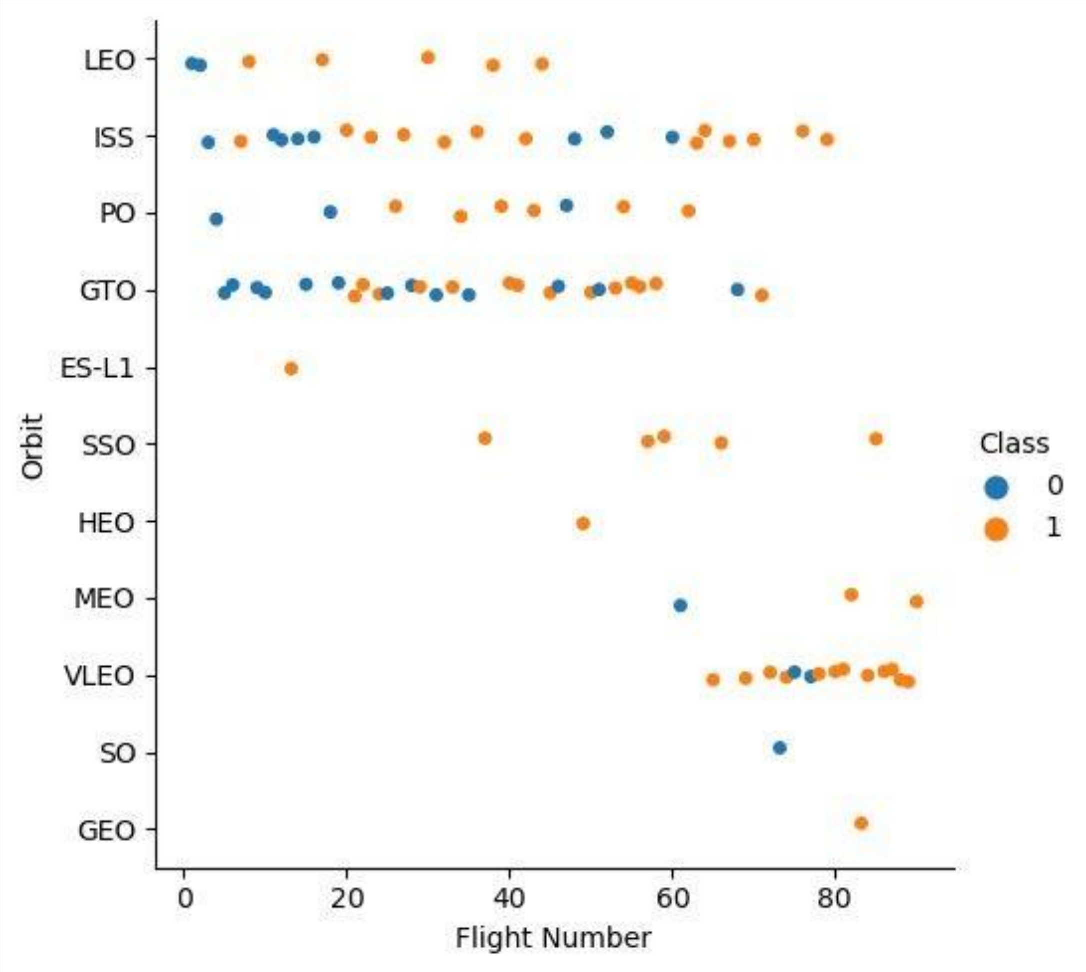
- Bar chart for the success rate of each orbit type



Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type

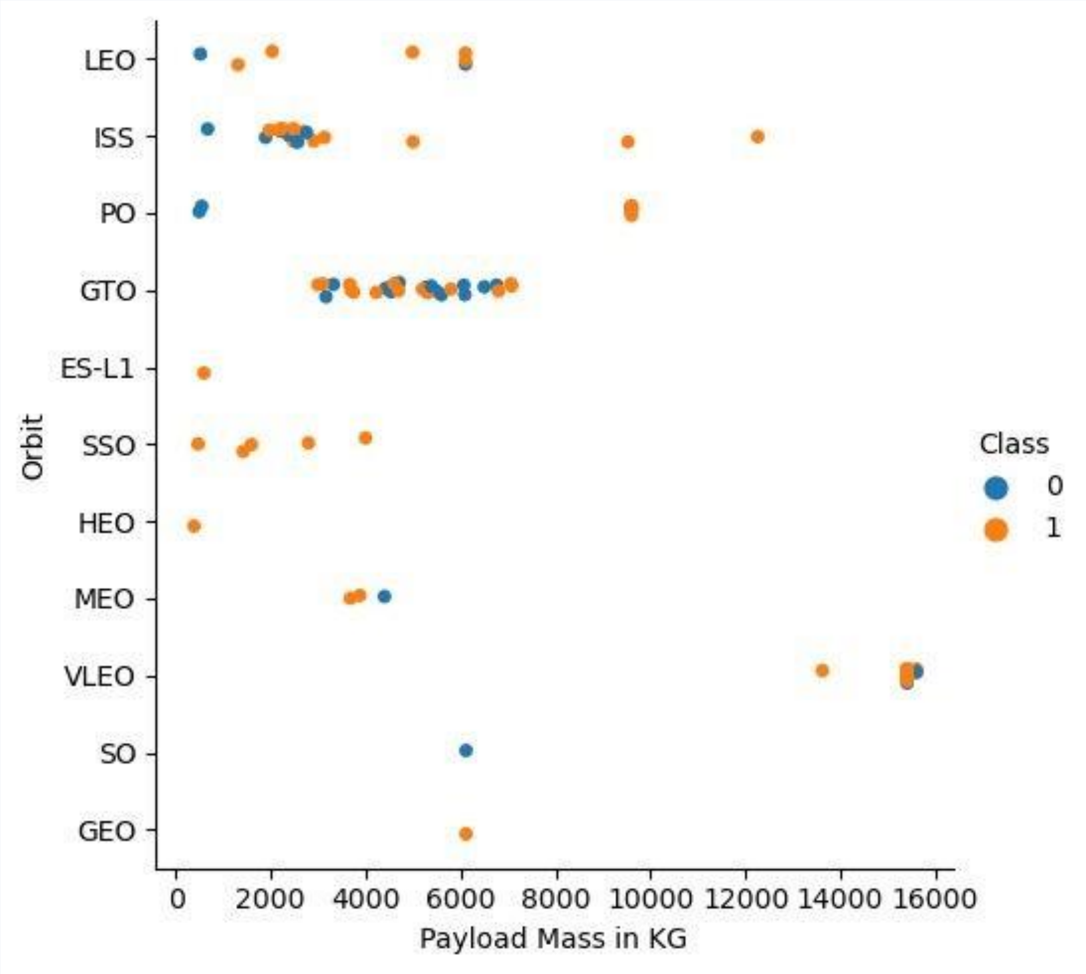
- Outcome is overlayed



Payload vs. Orbit Type

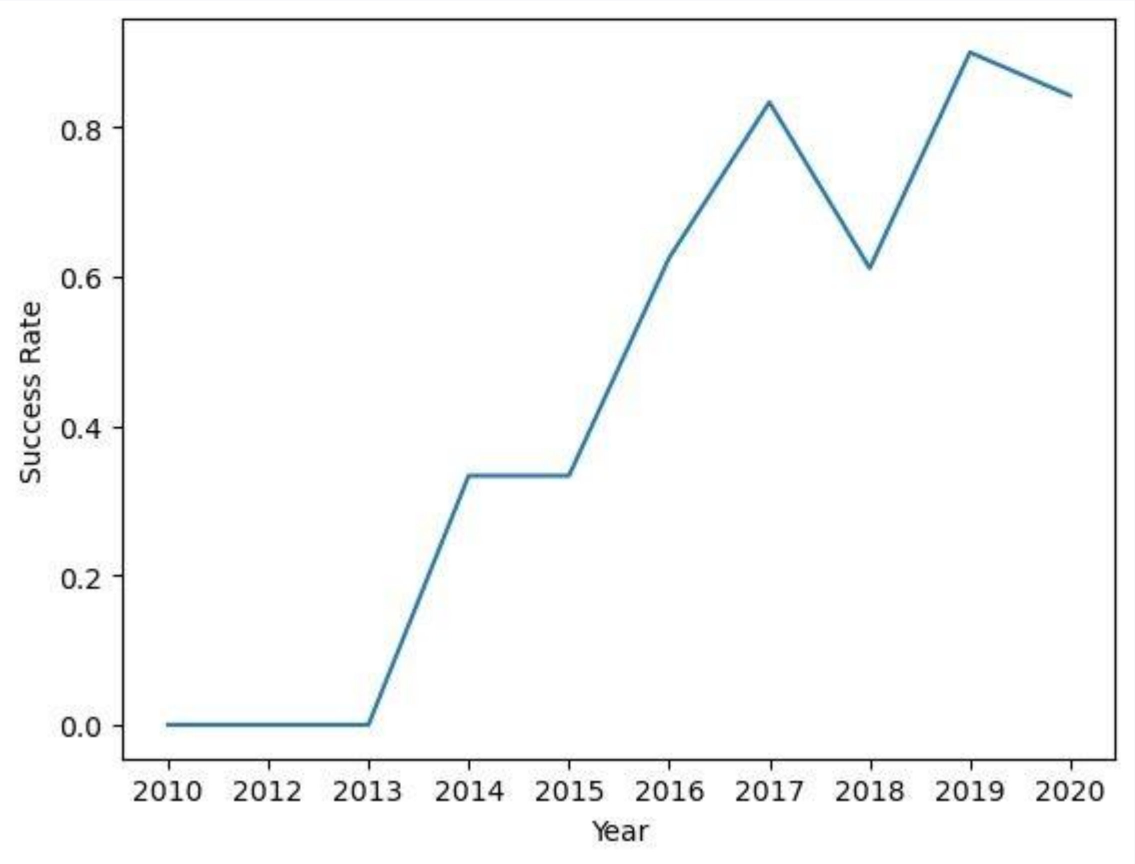
- Scatter point of payload vs. orbit type

- Outcome is overlayed



Launch Success Yearly Trend

- Line chart of yearly average success rate



All Launch Site Names

- Find the names of the unique launch sites

Task 1

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

In [8]:
```
%%sql

select distinct "Launch_Site" from spacextbl
```

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

Out[8]:

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`

Task 2

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

In [47]:
```
%%sql

select * from spacextbl where "Launch_Site" like "CCA%" limit 5
```

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

Out[47]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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

Task 3

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

In [43]:
```
%%sql

select sum("PAYLOAD_MASS__KG_") from spacextbl where "Customer" like "%NASA%"
```

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

Out[43]:

sum("PAYLOAD_MASS__KG_")
107010

Average Payload Mass by F9 v1.1

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

Task 4

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

In [45]:
```
%%sql
select avg ("PAYLOAD_MASS__KG_") from spacextbl where "Booster_Version" like "F9%"
```

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

Out[45]:

avg ("PAYLOAD_MASS__KG_")
6138.287128712871

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

In [75]:
```
%%sql

select "Date" from spacextbl where "Landing _Outcome" == "Success (ground pad)"
```

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

Out[75]:

Date
22-12-2015
18-07-2016
19-02-2017
01-05-2017
03-06-2017
14-08-2017
07-09-2017
15-12-2017
08-01-2018

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship

Task 6

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

```
In [67]: %%sql

select "Booster_Version", "PAYLOAD_MASS__KG_"  from spacextbl where "Landing _Outcome" == "Success (drone ship)" and ("PAYLOAD_MASS__KG_" < 6000 and "PAYLOAD_MASS__KG_" > 4000 )

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

Out[67]:

Booster_Version	PAYLOAD_MASS__KG_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

In [69]:
```
%%sql
select "Mission_Outcome", count (*) as "total" from spacextbl group by "Mission_Outcome"
```

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

Out[69]:

Mission_Outcome	total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [72]:
```
%%sql
select distinct "Booster_Version", "PAYLOAD_MASS__KG_" from spacextbl where "PAYLOAD_MASS__KG_" == (select max("PAYLOAD_MASS__KG_") from spacextbl)
```

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

Out[72]:

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

2015 Launch Records

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

Task 5

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

Hint:Use min function

In [75]:
```
%%sql

select "Date" from spacextbl where "Landing _Outcome" == "Success (ground pad)"
```

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

Out[75]:

Date
22-12-2015
18-07-2016
19-02-2017
01-05-2017
03-06-2017
14-08-2017
07-09-2017
15-12-2017
08-01-2018

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

Task 5

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

Hint:Use min function

In [75]:
```
%%sql

select "Date" from spacextbl where "Landing _Outcome" == "Success (ground pad)"
```

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

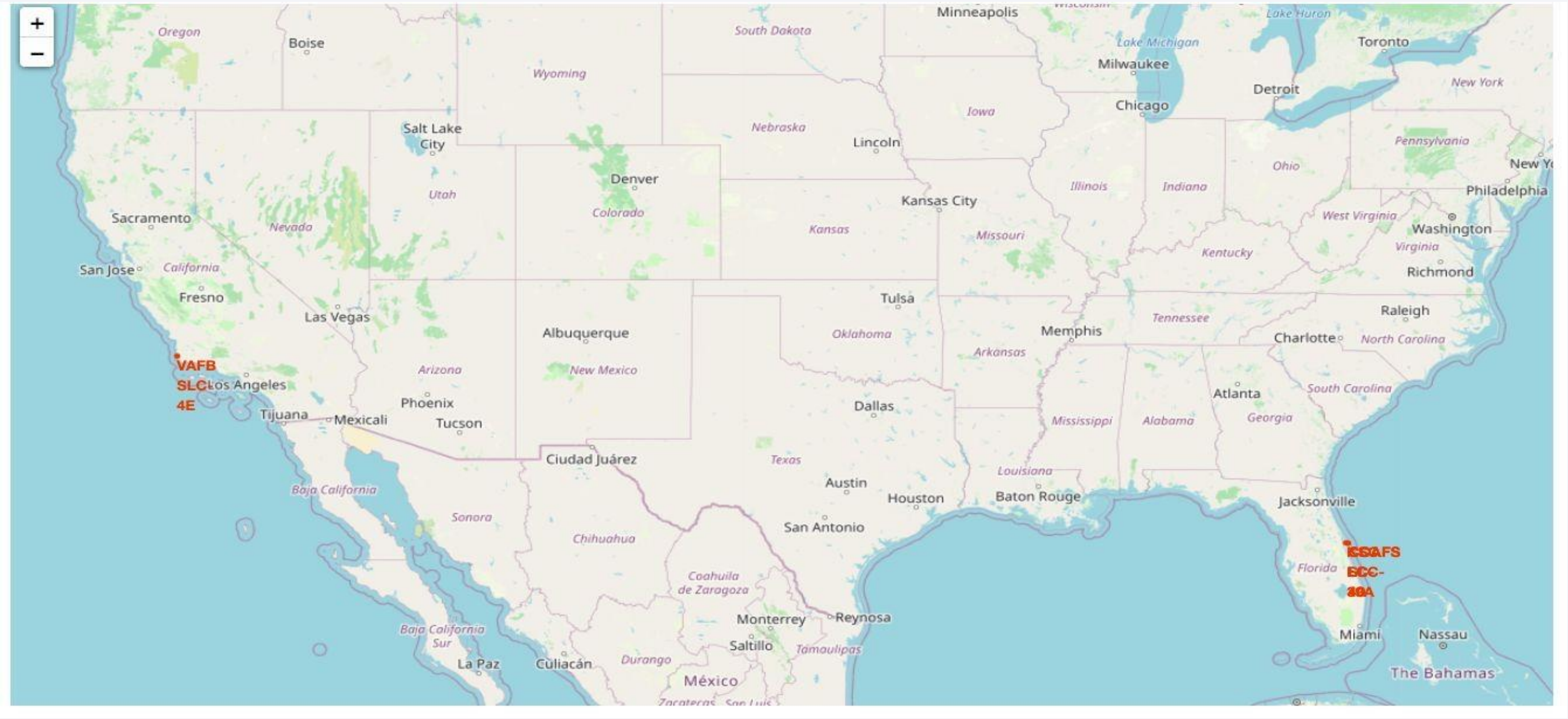
Out[75]:

Date
22-12-2015
18-07-2016
19-02-2017
01-05-2017
03-06-2017
14-08-2017
07-09-2017
15-12-2017
08-01-2018

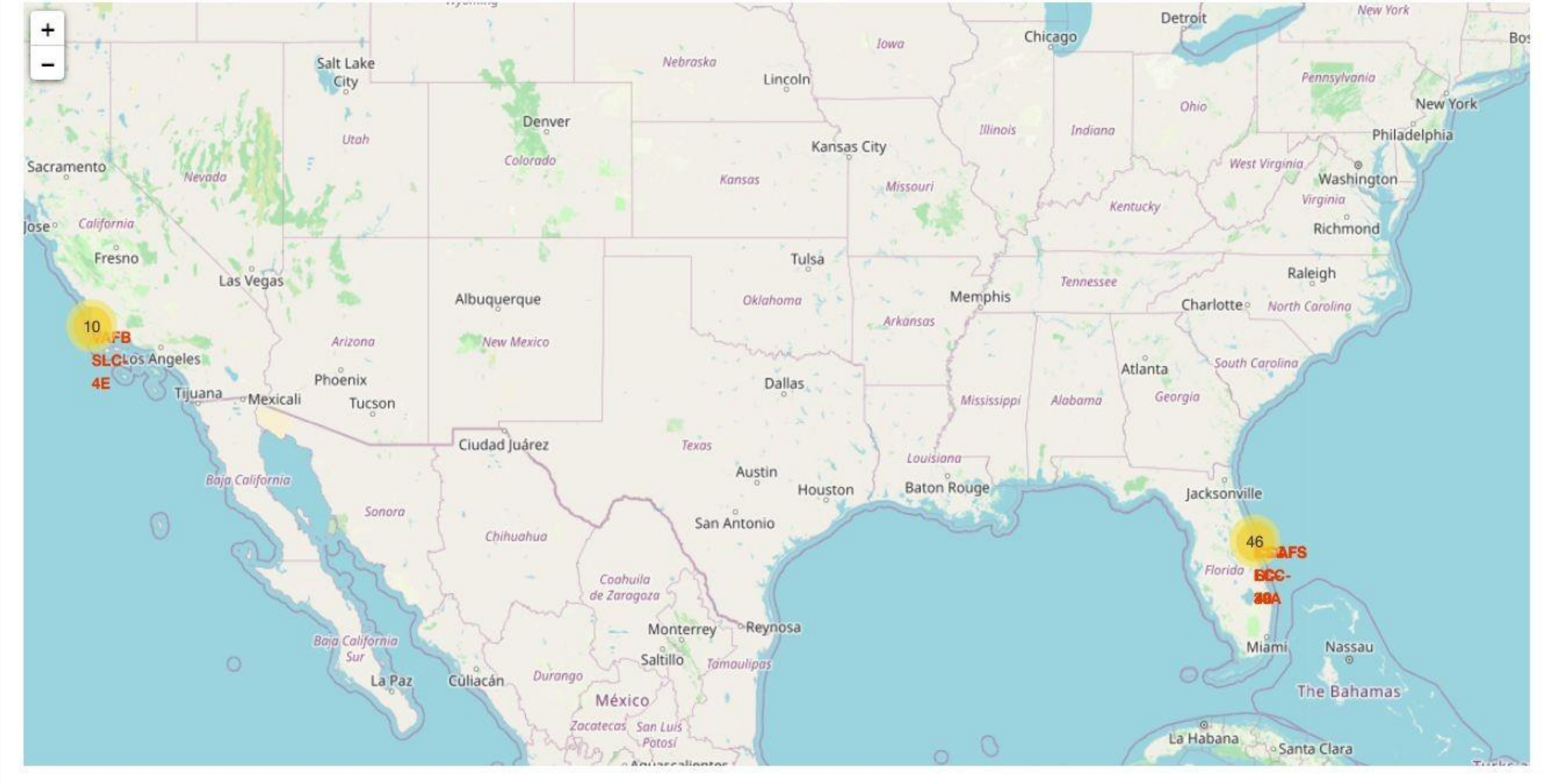
Section 3

Launch Sites Proximities Analysis

Launch Sites on Map



Launch Outcomes

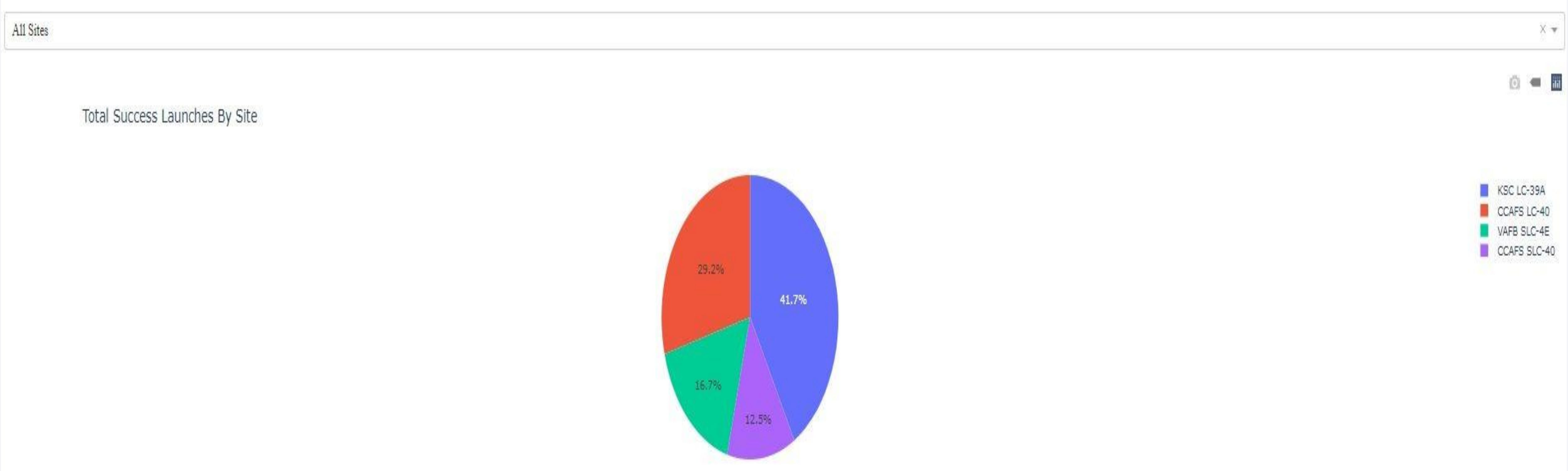


Proximities to Launch Site

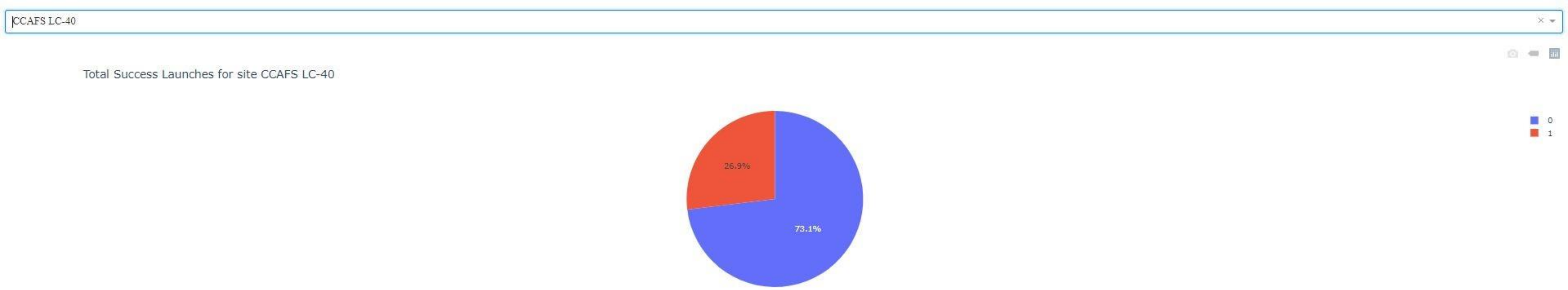
Section 4

Build a Dashboard with Plotly Dash

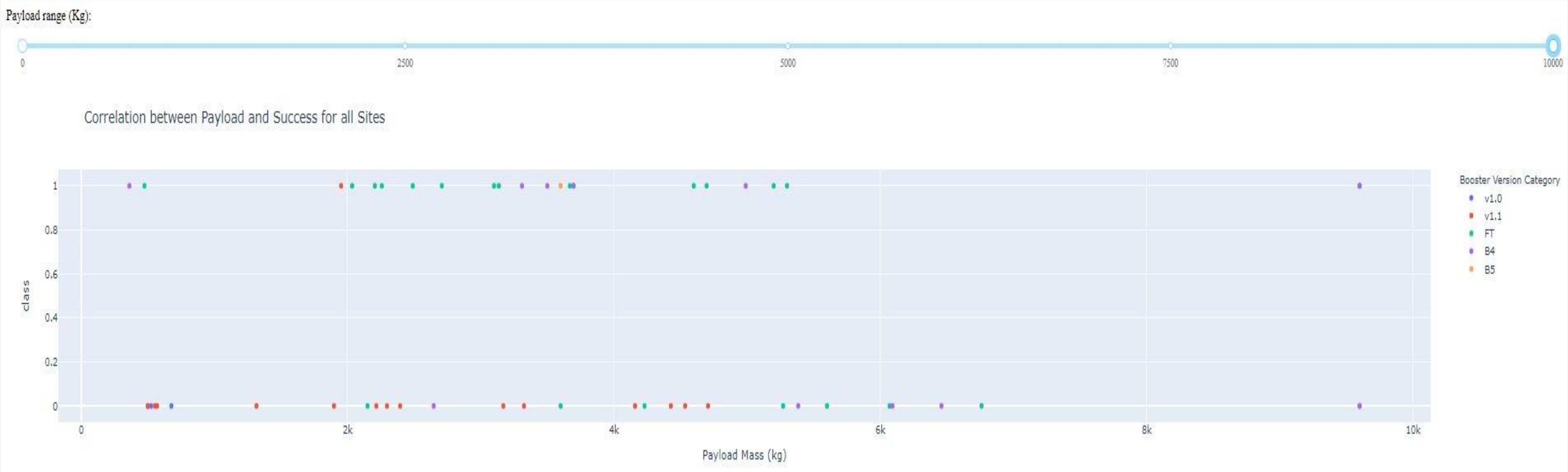
Launch Success - All Sites



Success Ratio for Most Successful Site



Payload vs Launch Outcome

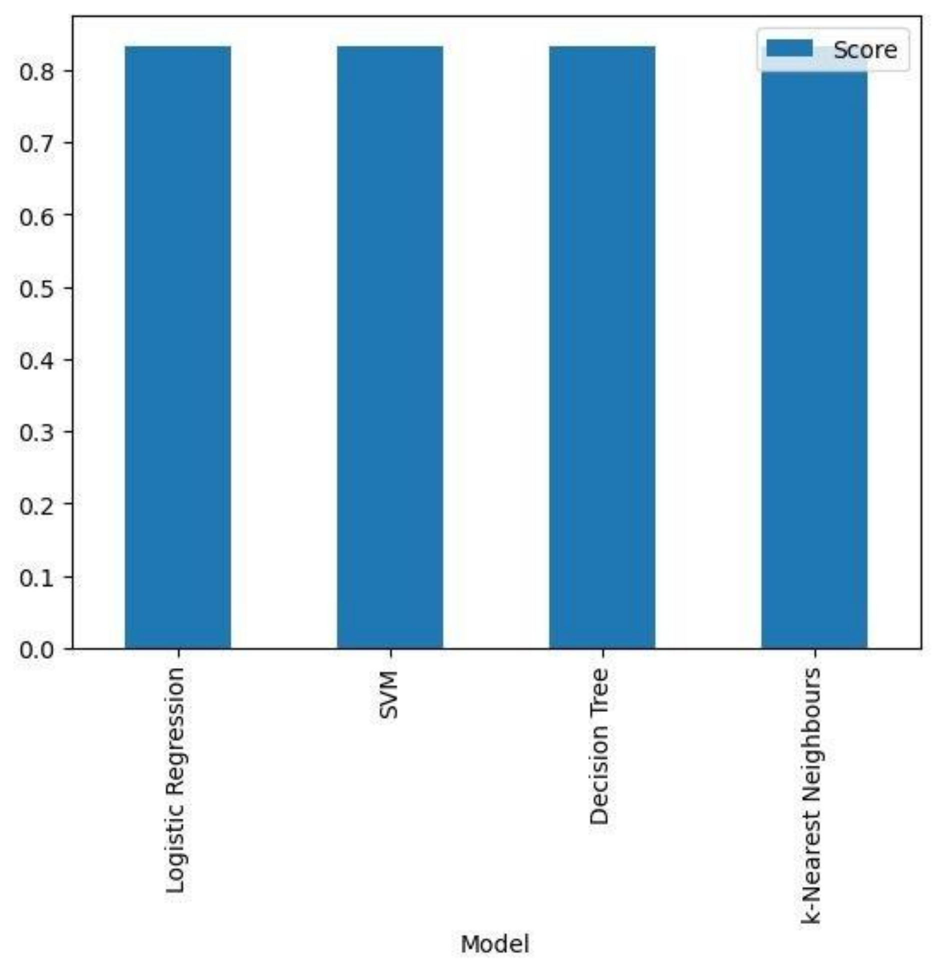


Section 5

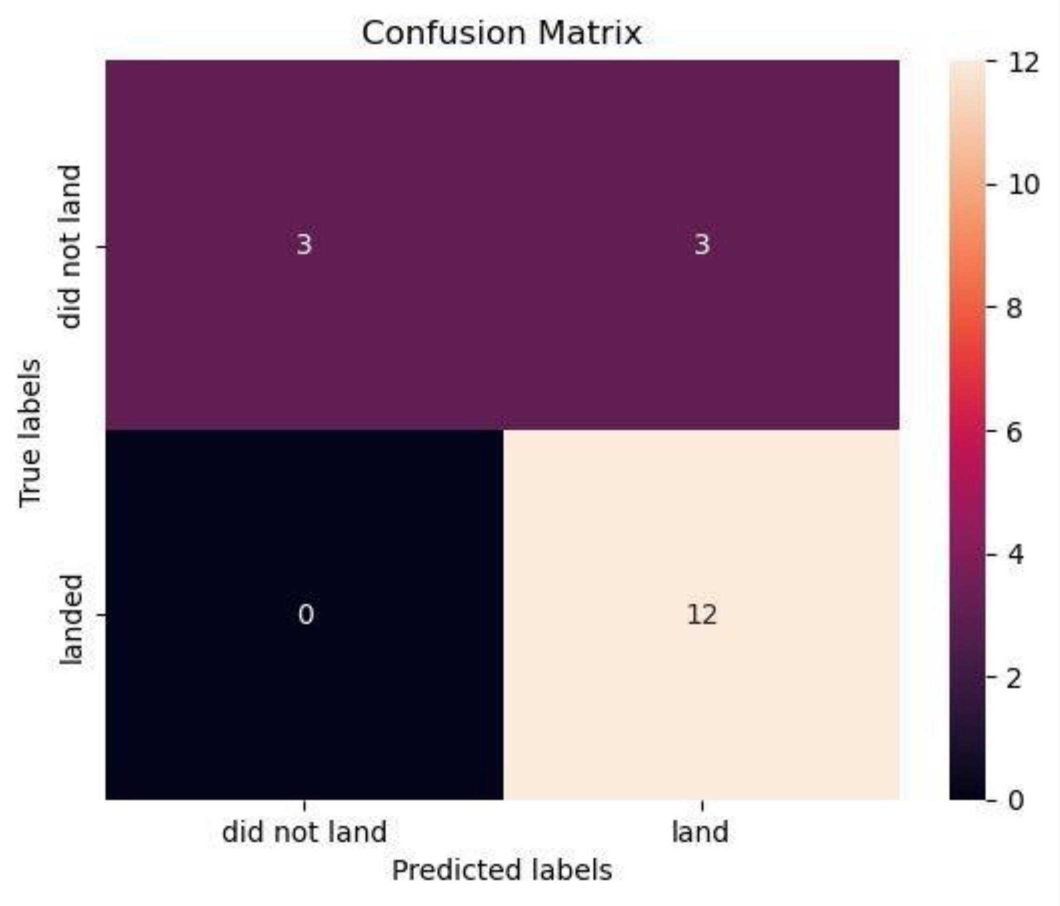
Predictive Analysis (Classification)

Classification Accuracy

- Built model accuracy for all built classification models, in a bar chart



Confusion Matrix



Conclusions

- The variables that have best allowed us to predict launch outcome are:
  - 'Flight Number'
  - 'PayloadMass'
  - 'Orbit'
  - 'LaunchSite'
  - 'Flights'
  - 'GridFins'
  - 'Reused'
  - 'Legs'
  - 'LandingPad
  - 'Block'
  - 'ReusedCount'
  - 'Serial'
- The suitable machine learning method for prediction is classification
- All classification models used on the data had similar accuracy
- With the given data, the accuracy of predicting the landing outcome of a launch is around 83.3%

Thank you!