IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Johar Ali
Mar 04, 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- In this applied data science capstone project, we will make prediction about the spacex Falcon 9 first stage whether it will land successfully or not using some machine learning classification algorithms.
- The process in this project used the following steps.
  - Data collection through API and web scrapping, data wrangling and data formatting.
  - Exploratory data analysis.
  - Data visualization.
  - Prediction about the success.
- The graph shows that there is correlation between the rocket launches and launches outcome.
- The results shows that to predict if the Falcon 9 first stage the decision tree is the best ML algorithms.

Introduction

- In this capstone project, our aim is to predict the successful landing of the Falcon 9 first stage. SpaceX prominently features Falcon 9 rocket launches on its website, pricing them at 62 million dollars each, significantly lower than other providers whose costs can reach up to 165 million dollars per launch. A key factor in this cost reduction is SpaceX's ability to reuse the first stage of the rocket. Therefore, by accurately predicting whether the first stage will successfully land, we can estimate the overall cost of a launch. This information holds strategic value, especially if competing companies seek to bid against SpaceX for a rocket launch contract.
- Many unsuccessful landings of Falcon 9 first stages are intentional. On certain occasions, SpaceX opts for a controlled landing in the ocean.
- Our primary objective is to determine, based on various features associated with a Falcon 9 rocket launch such as payload mass, orbit type, launch site, among others, whether the first stage of the rocket will achieve a successful landing.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from the website of spacex through API and web scrapping.
- Perform data wrangling
  - Data was processed and cleaned. Missing values were removed.
- 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 collection and preprocessing, Splitting data, Choosing appropriate model, Building the initial model, Cross validation, Model evaluation, Final evaluation, Iterative improvement and Deployment.

Data Collection – SpaceX API

- The provided API, accessible at https://api.spacexdata.com/v4/rockets/, furnishes a broad spectrum of information concerning various rocket launches conducted by SpaceX. In this context, the focus narrows down to Falcon 9 launches exclusively. To address any absence of data, each missing value within the dataset is substituted with the mean of its respective column. Consequently, the dataset comprises 90 rows or instances and 17 columns or features. The initial rows of the dataset are depicted in the accompanying image.

- Applied-Data-Science-Capstone/jupyter-labs-spacex-data-collection-api.ipynb at main · diamond25587/Applied-Data-Science-Capstone (github.com).

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857

Data Collection - Scraping

- Data is gathered through web scraping from the URL https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922. Specifically, the website exclusively presents information pertaining to Falcon 9 launches. Following the extraction process, the resulting dataset comprises 121 rows or instances and 11 columns or features. A visual representation of the initial rows of the dataset is provided in the image below.
- Applied-Data-Science-Capstone/jupyter-labs-webscraping.ipynb at main · diamond25587/Applied-Data-Science-Capstone (github.com).

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Data Wrangling

- Subsequent to data processing, any missing entries are addressed, and categorical features undergo one-hot encoding for enhancement. Additionally, an additional column labeled 'Class' is introduced to the data frame. This 'Class' column denotes 0 for failed launches and 1 for successful ones. Consequently, the dataset is refined to encompass 90 rows or instances and expands to 83 columns or features.

- Applied-Data-Science-Capstone/labs-jupyter-spacex-Data wrangling.ipynb at main · diamond25587/Applied-Data-Science-Capstone (github.com).

EDA with Data Visualization

- Utilizing functions from the Pandas and NumPy libraries, essential insights are derived regarding the collected data. This encompasses determining:
  - The count of launches conducted at each launch site.
  - The frequency of occurrences for each orbit.
  - The tally and distribution of mission outcomes.
- Applied-Data-Science-Capstone/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb at main · diamond25587/Applied-Data-Science-Capstone (github.com).

EDA with SQL

- SQL queries are employed to extract insights from the dataset, addressing various inquiries, such as:
    - Identifying the distinct launch sites involved in space missions.
    - Calculating the total payload mass transported by boosters launched under NASA's CRS program.
    - Determining the average payload mass transported by booster version F9 v1.1.
- Applied-Data-Science-Capstone/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · diamond25587/Applied-Data-Science-Capstone (github.com).

Predictive Analysis

- We utilize functions from the Scikit-learn library to develop our machine learning models. The prediction phase of our machine learning process involves the following steps:
- Standardizing the data.
- Dividing the data into training and test sets.
- Constructing machine learning models, specifically:
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree
  - K Nearest Neighbors (KNN)
- Fitting the models to the training set.
- Identifying the optimal combination of hyper parameters for each model.
- Evaluating the models based on their accuracy scores and confusion matrices.

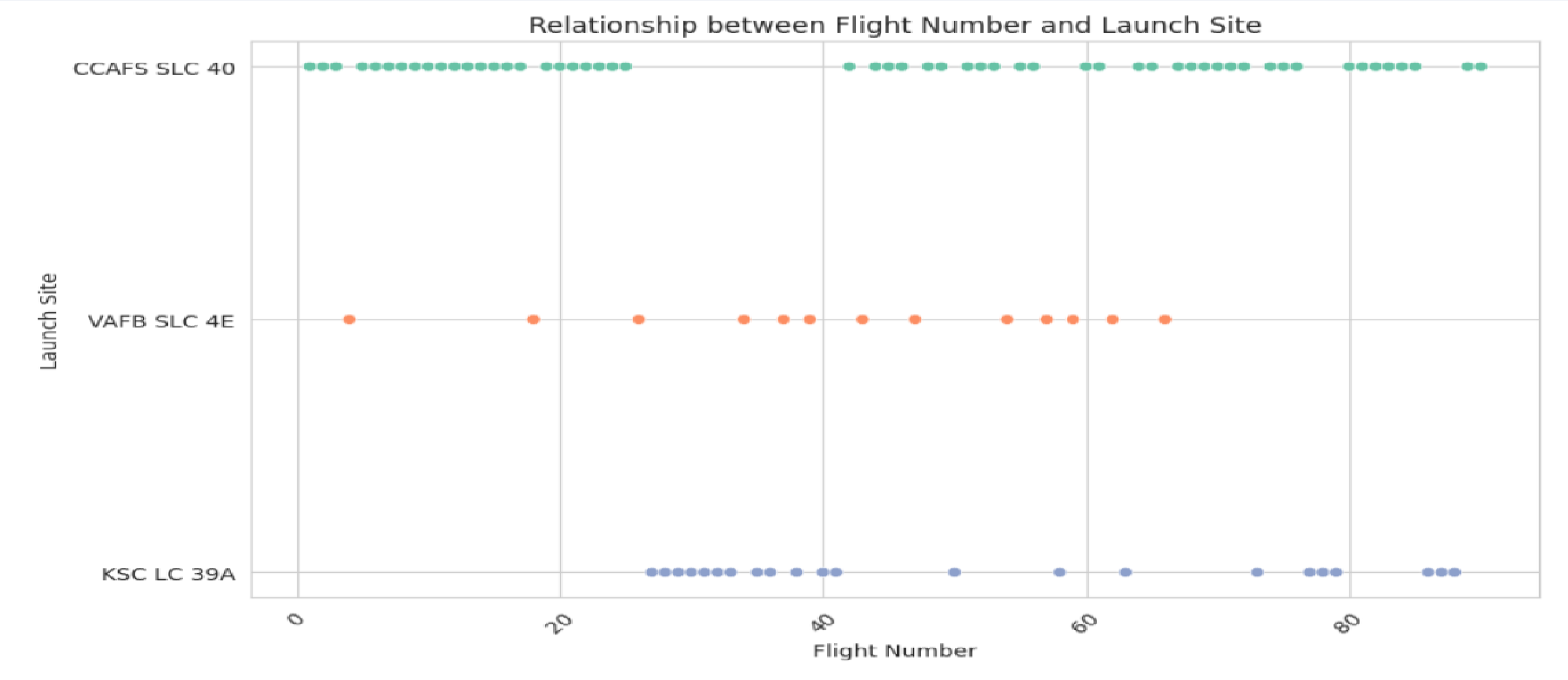
Results

- Scatter plot the flight number vs payload.
- Relationship between flight number and launch site through scatter plot.
- Relationship between payload and launch site through scatterplot.
- Relationship between success rate and orbit through bar chart.
- Relationship between flight number and orbit type through scatter plot.
- Relationship between payload and orbit type trough box plot.
- Success rate for years through line chart.

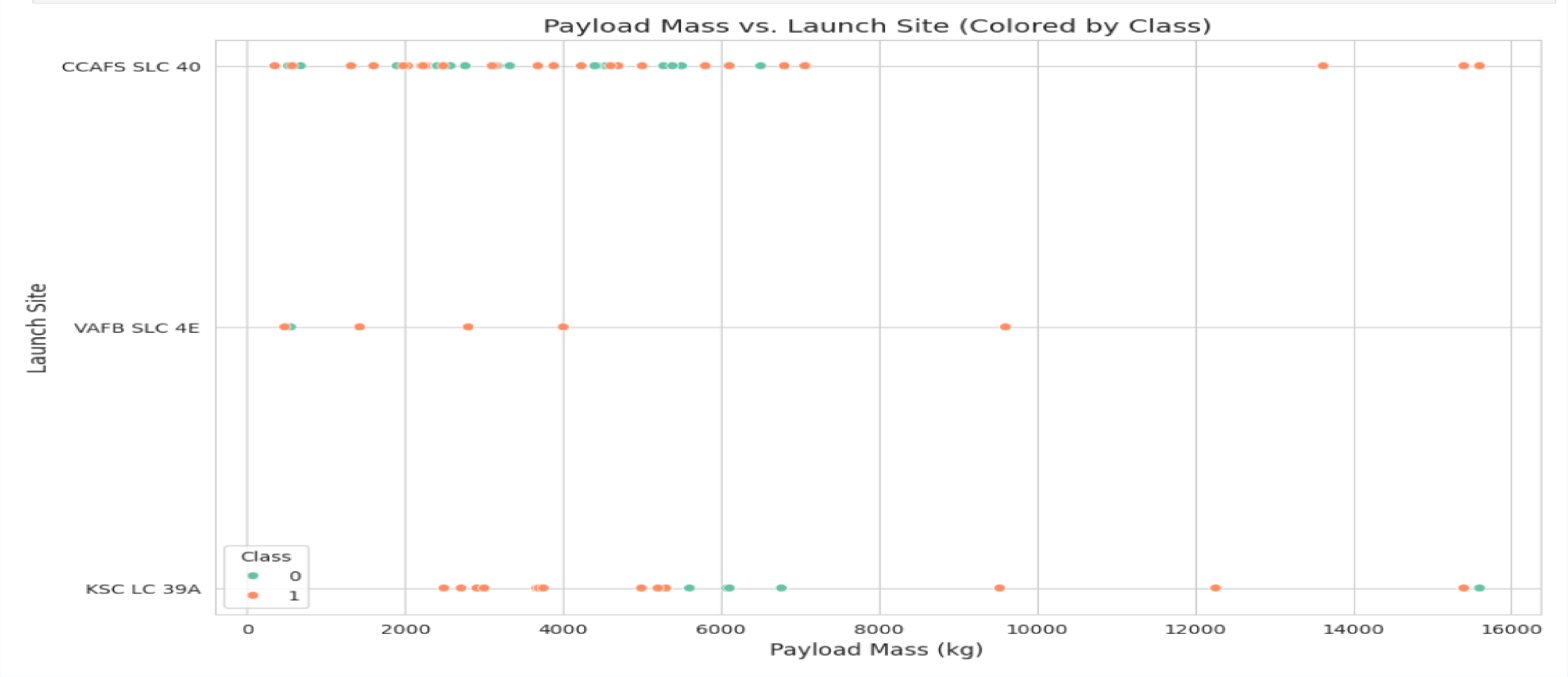
Section 2

Insights drawn from EDA

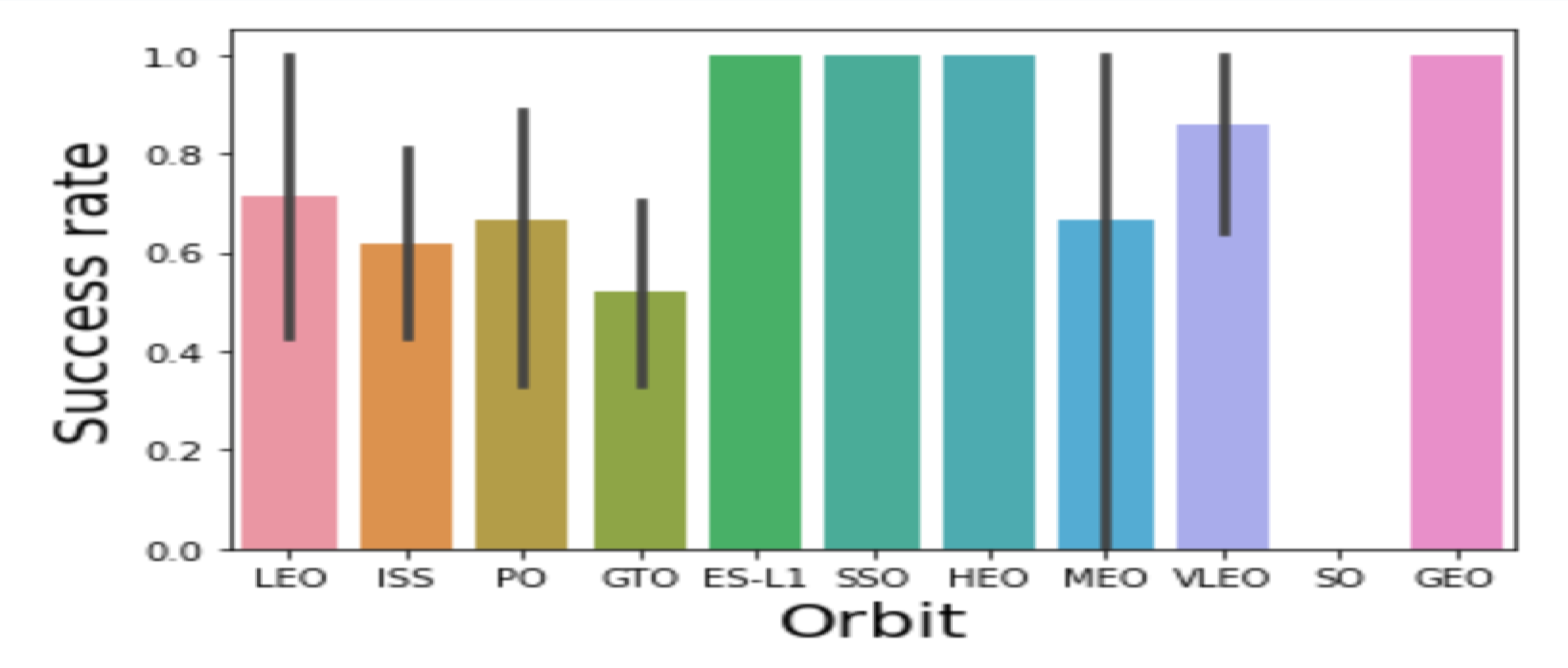
Flight Number vs. Launch Site



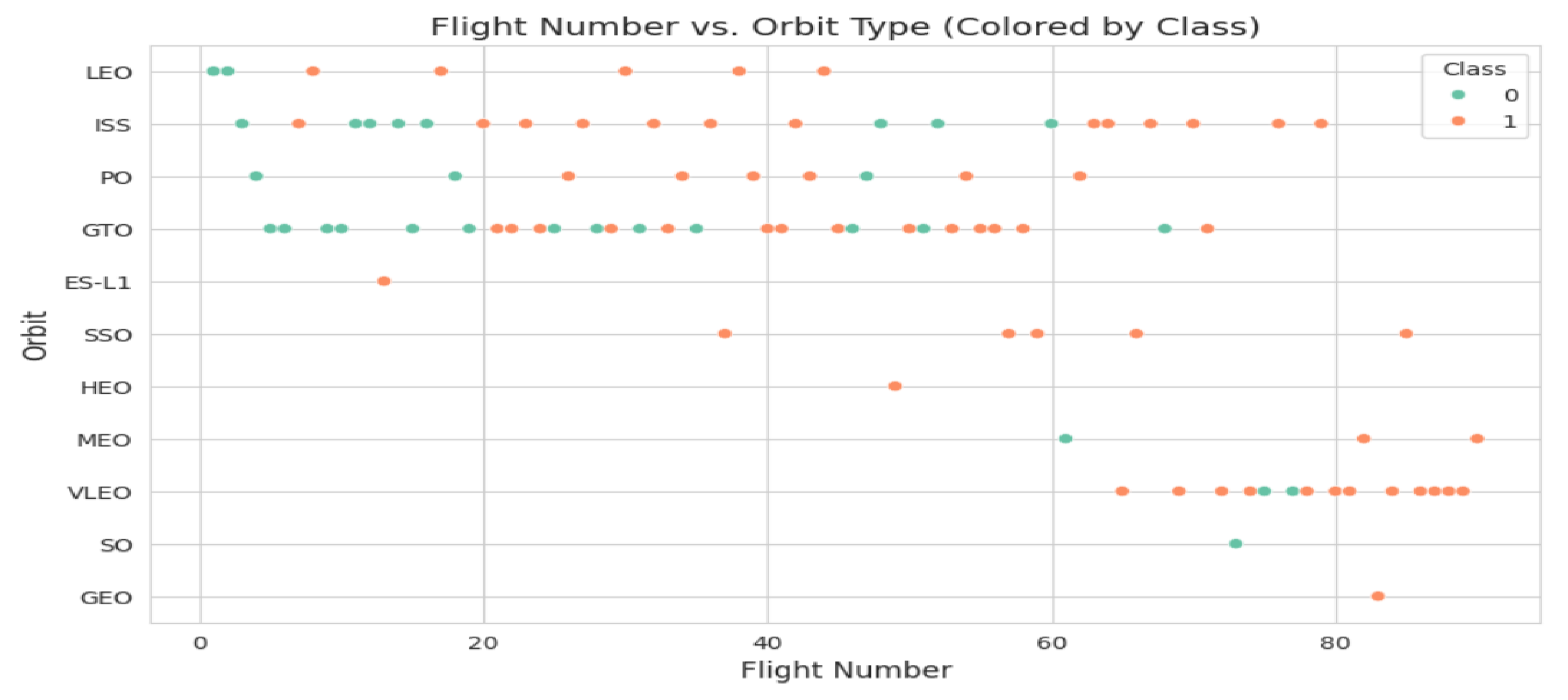
Payload vs. Launch Site



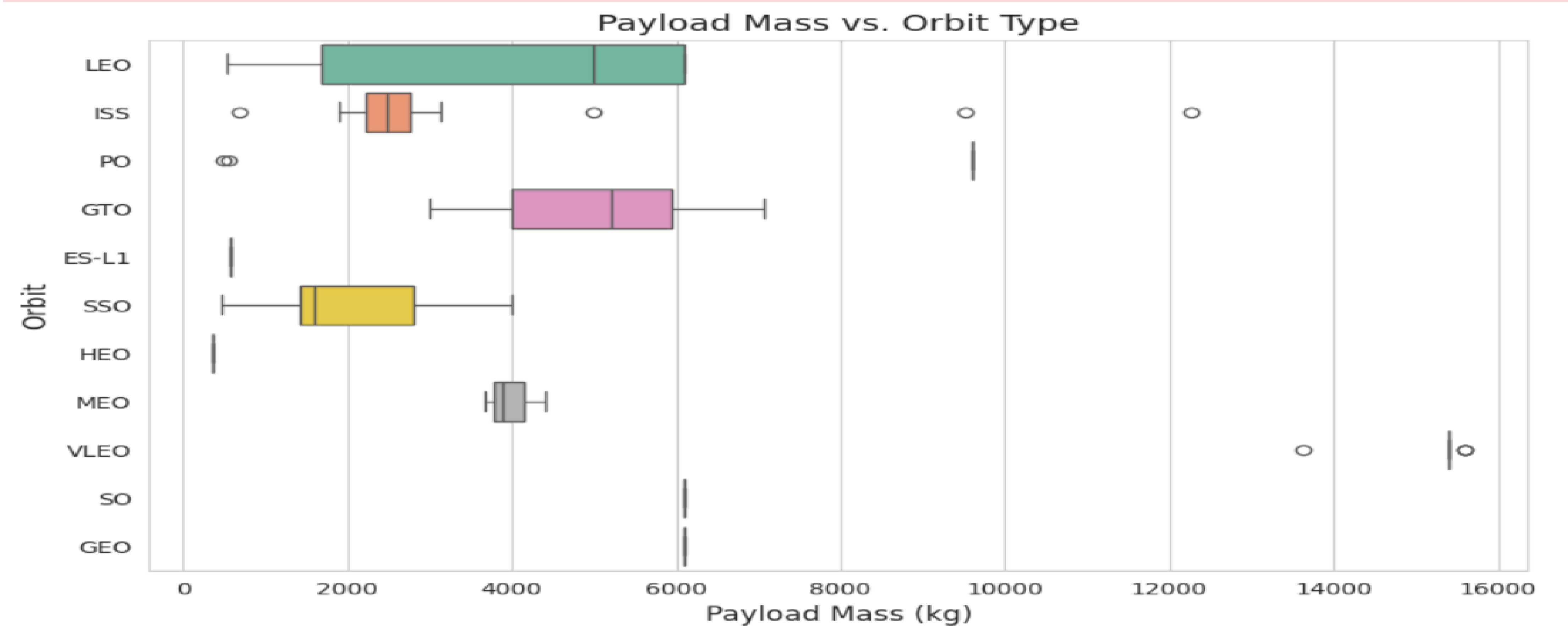
Success Rate vs. Orbit Type



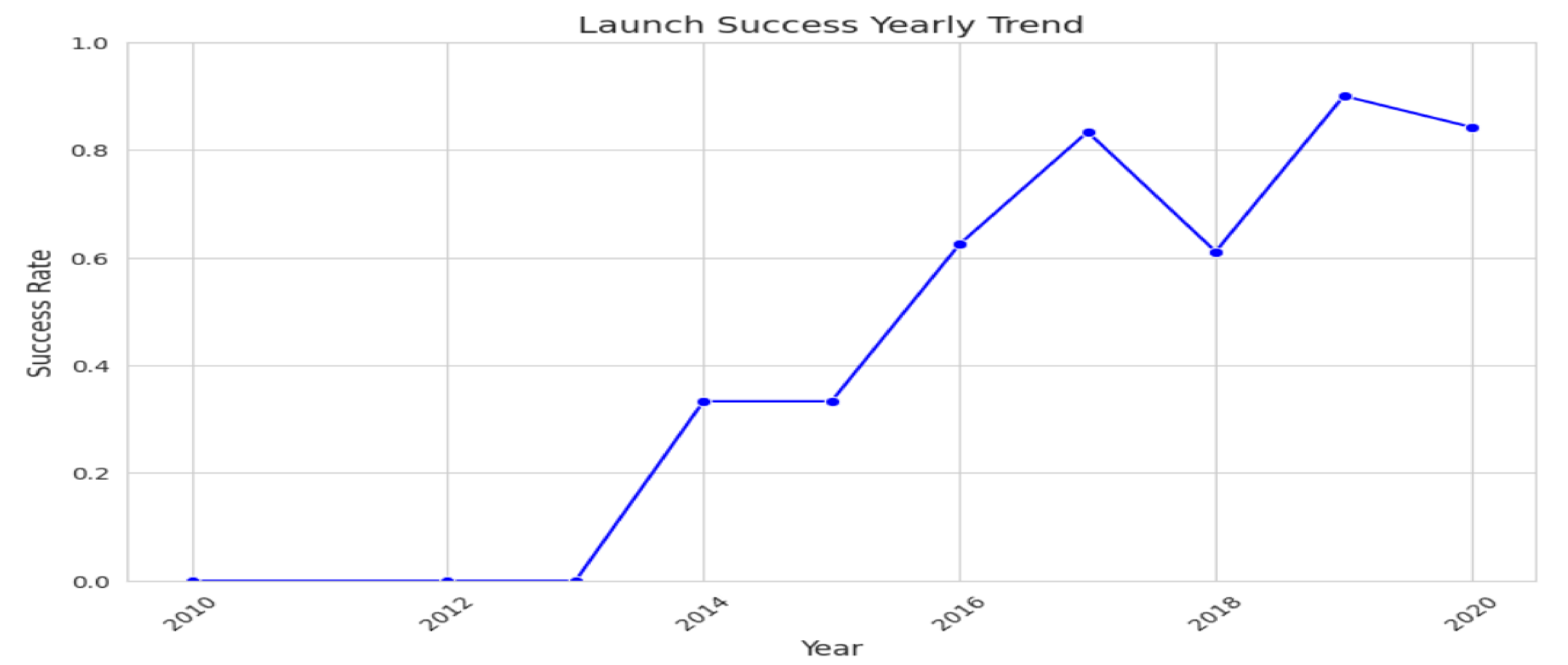
Flight Number vs. Orbit Type



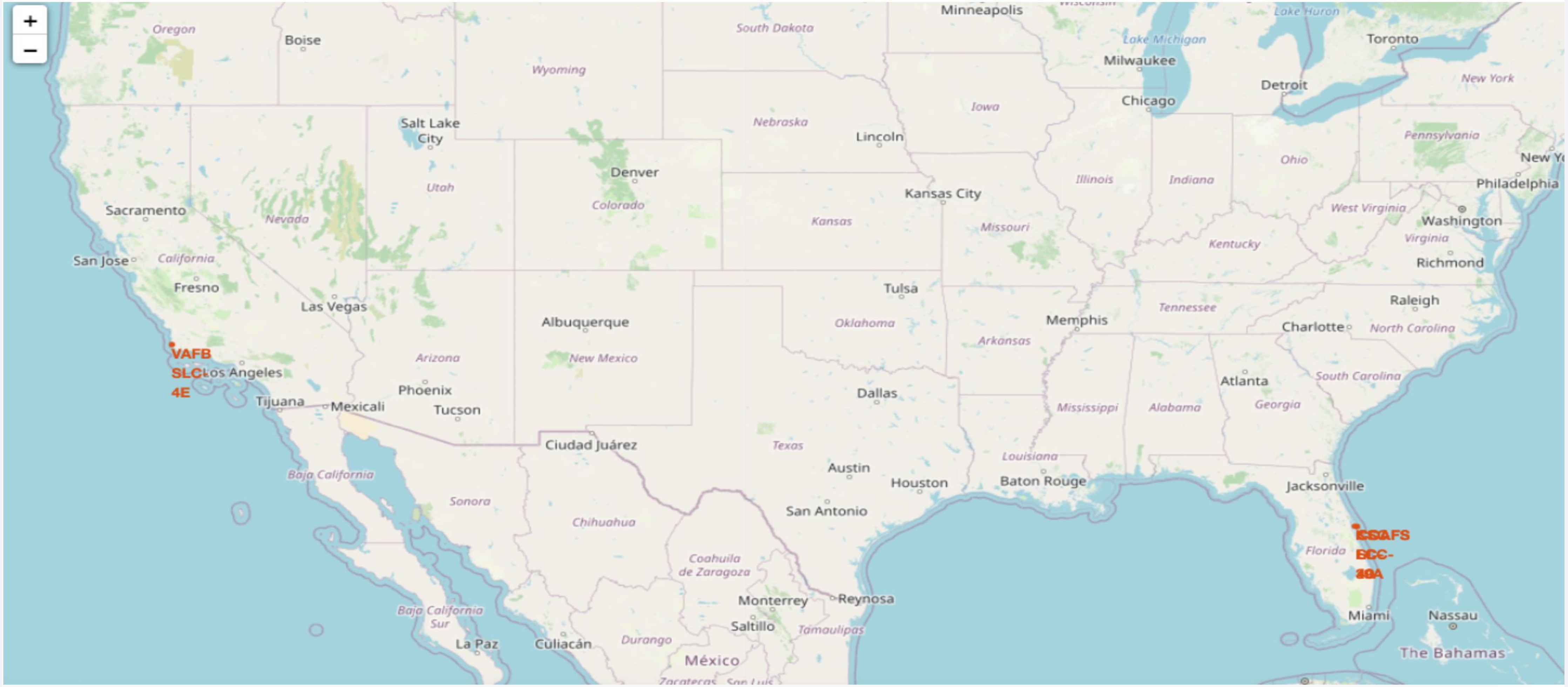
Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names



Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites 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 attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Total payload mass by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

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

Average payload mass by Booster Version F9 v1.1
2928

First Successful Ground Landing Date

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

Date of first successful landing outcome in ground pad
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

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

number_of_success_outcomes	number_of_failure_outcomes
100	1

Boosters Carried Maximum Payload

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

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
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

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

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

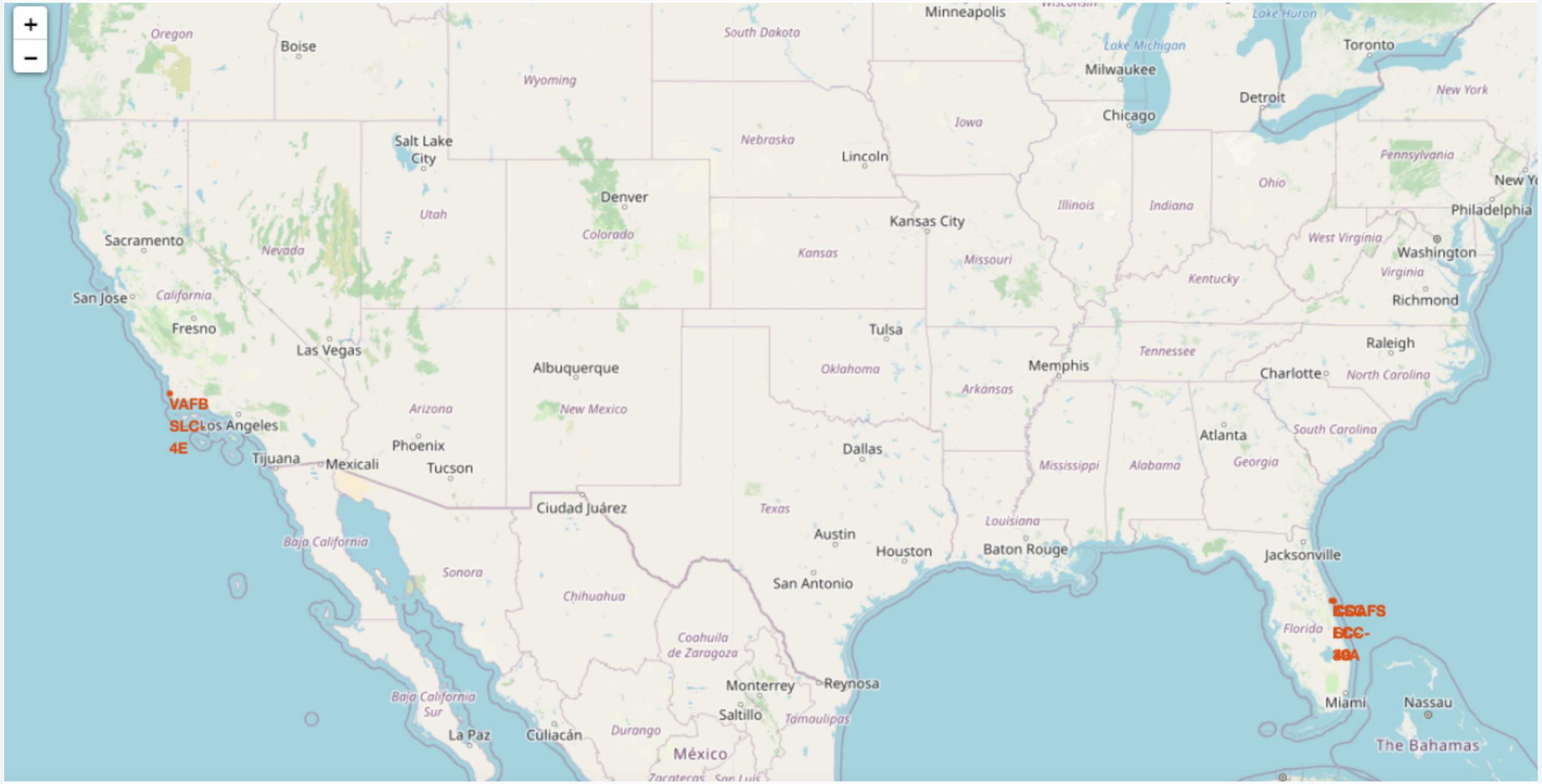
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

landing__outcome	landing_count
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

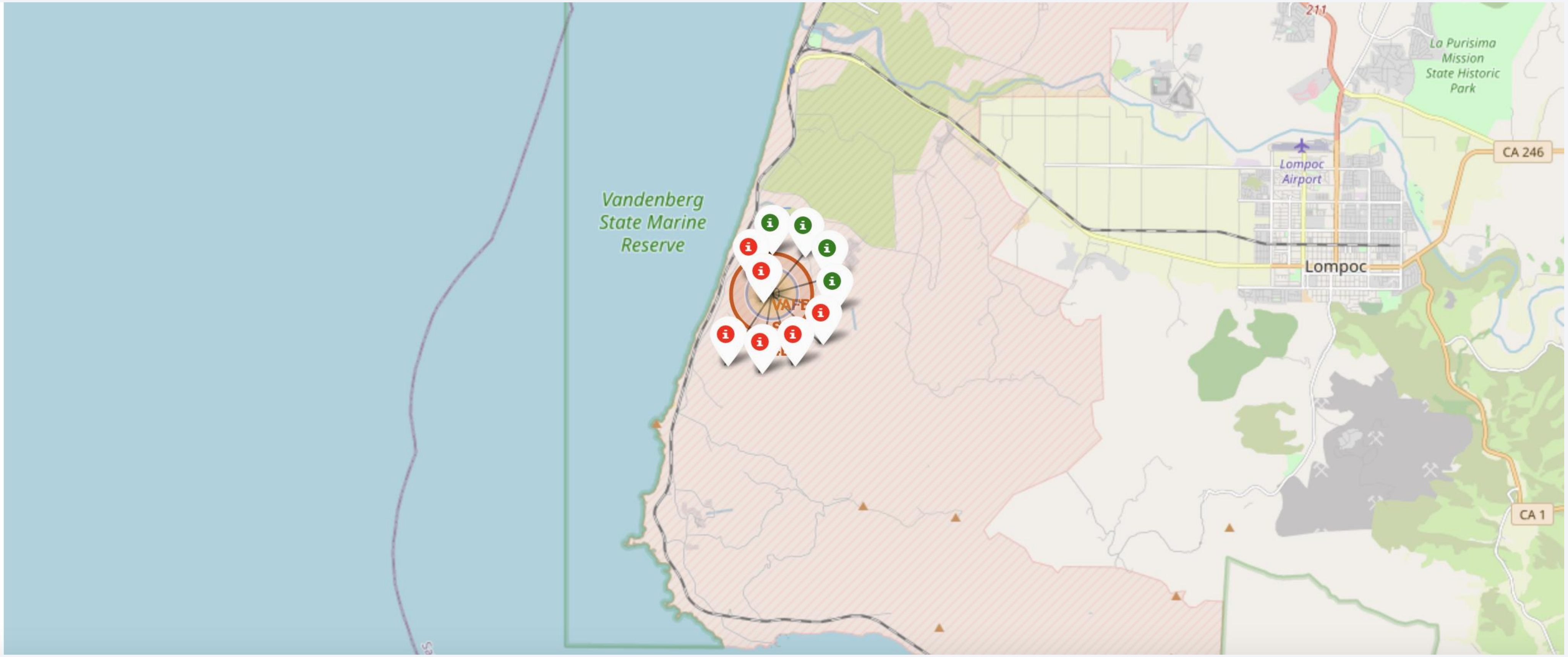
Section 3

Launch Sites Proximities Analysis

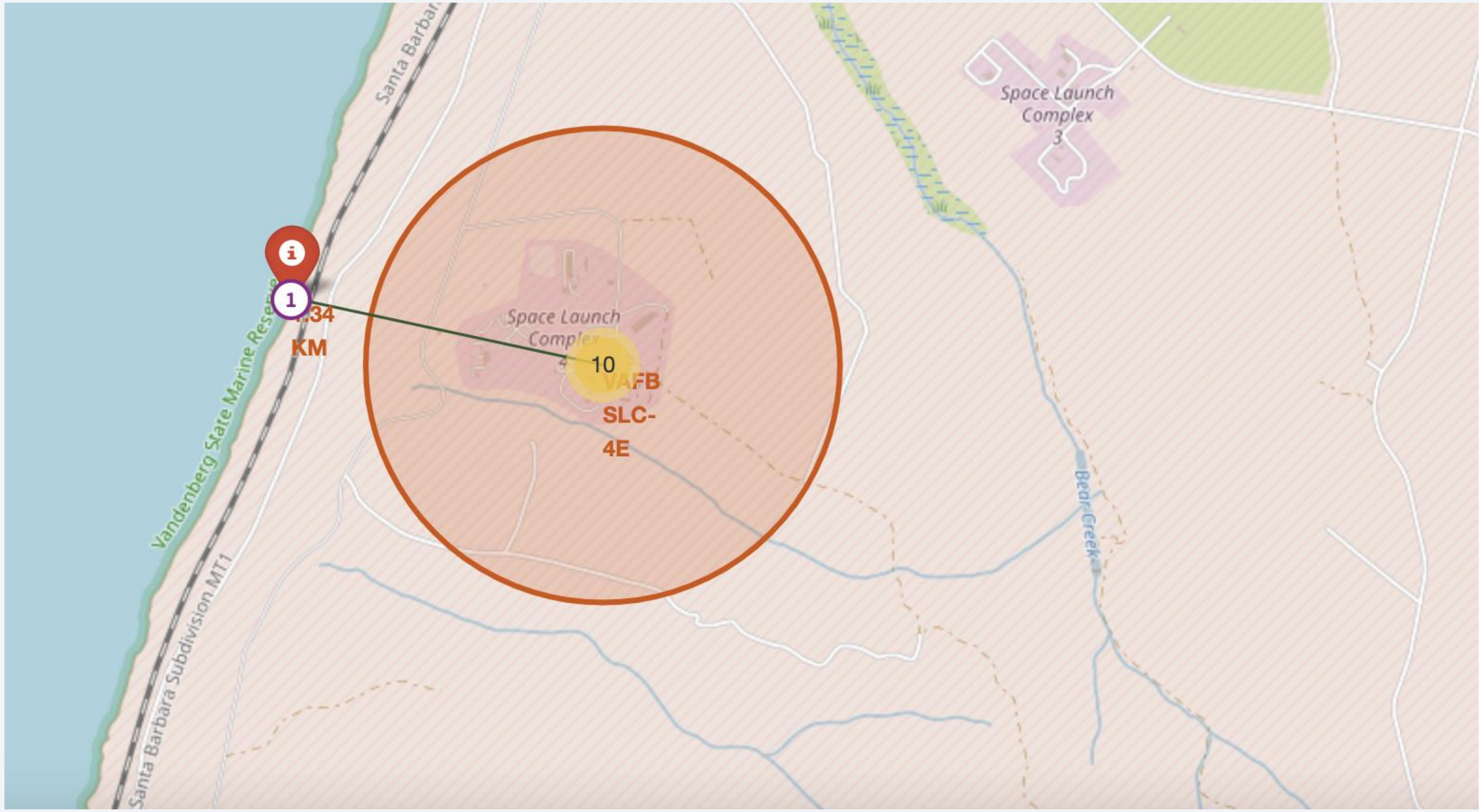
<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



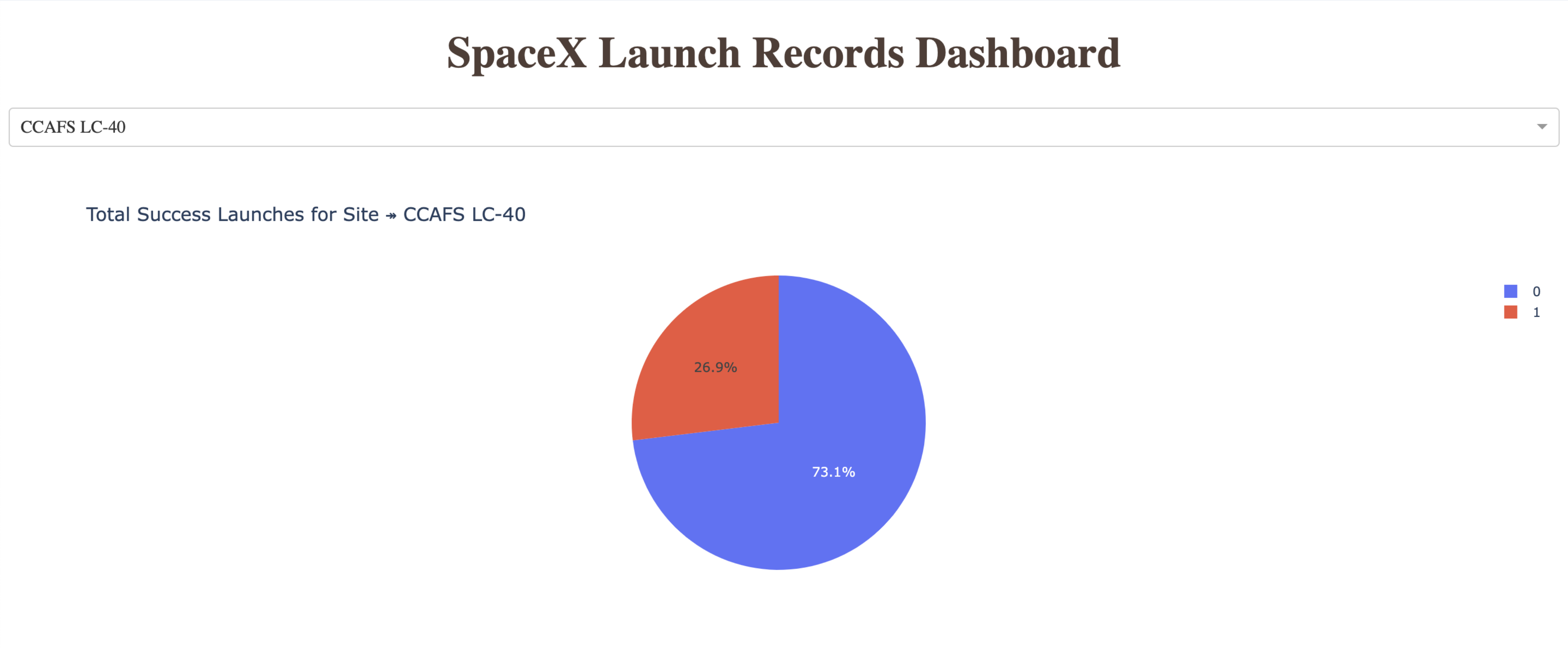
<Folium Map Screenshot 3>



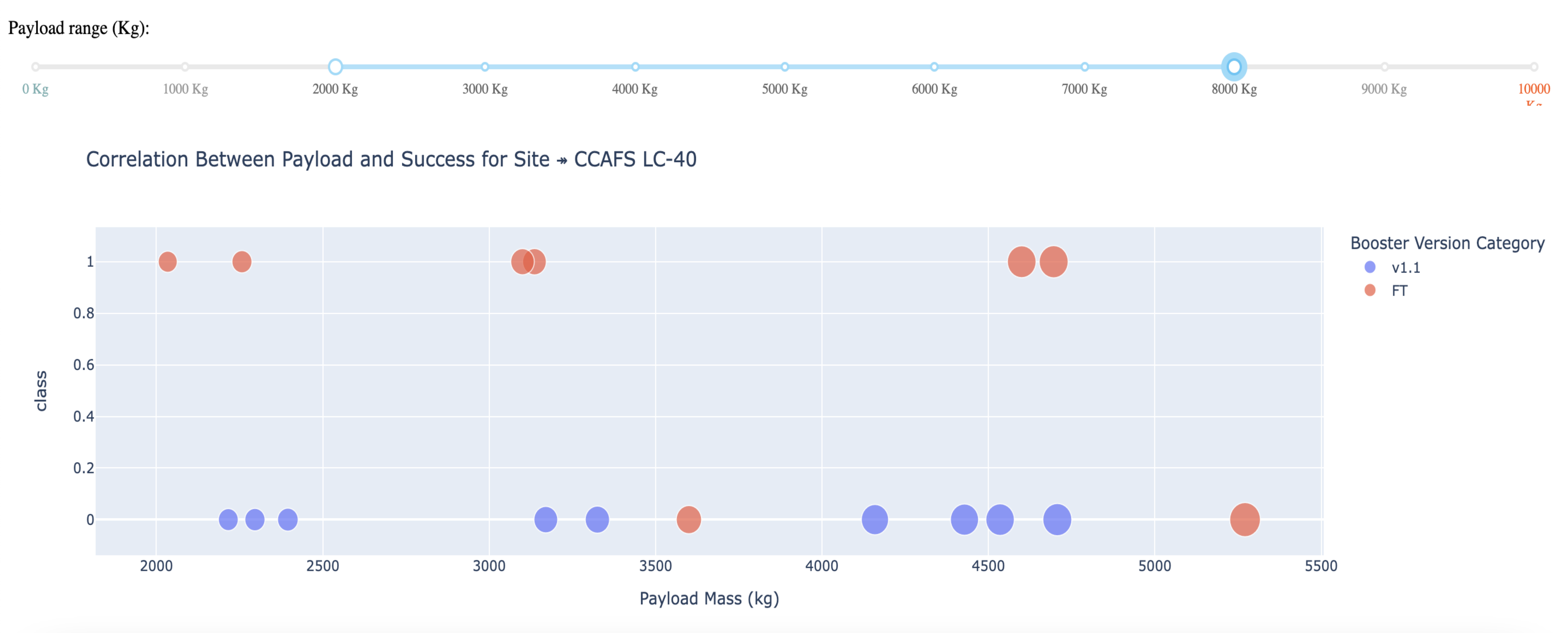
Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



<Dashboard Screenshot 2>



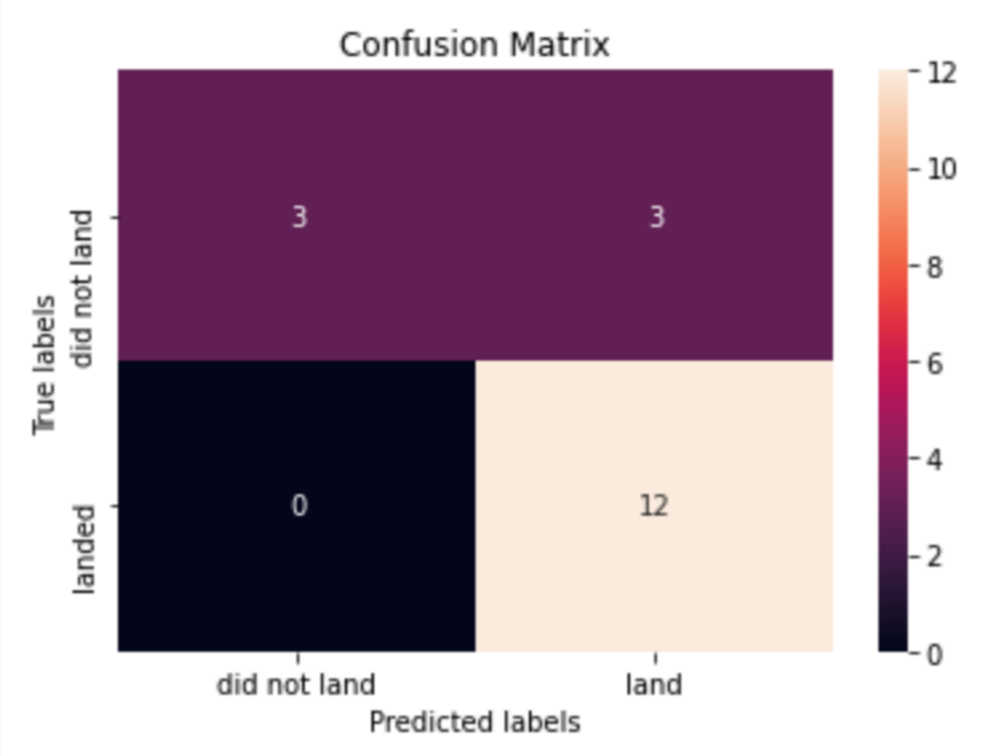
<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

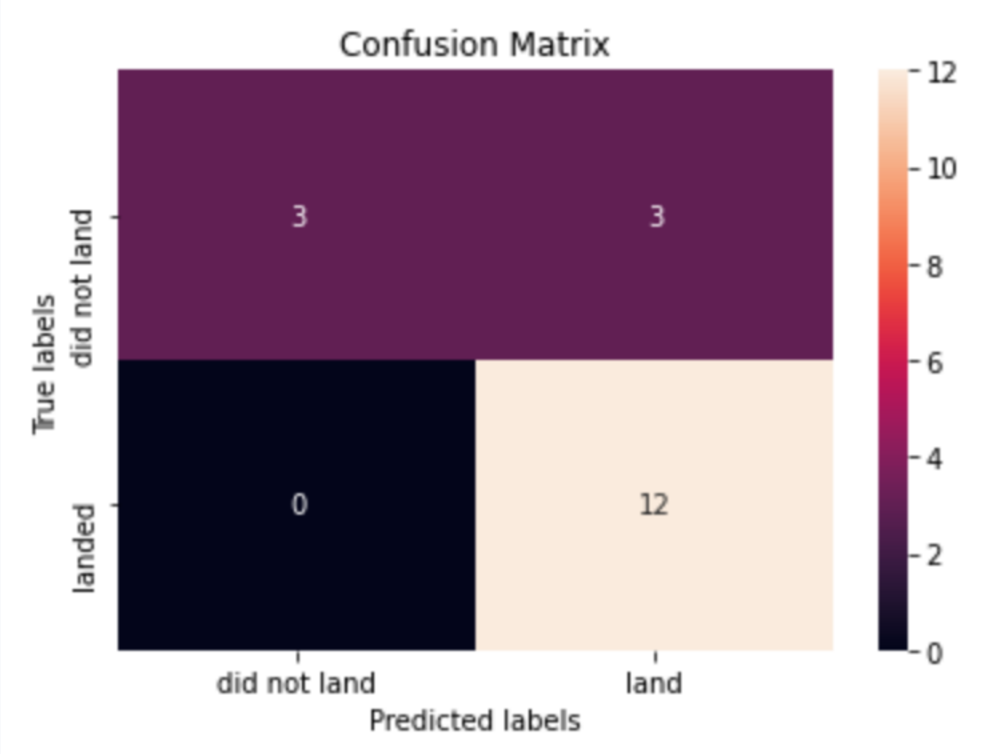
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- In this endeavor, we aim to forecast whether the initial phase of a specific Falcon 9 launch will successfully land, thereby influencing the overall cost of the launch. Various factors associated with a Falcon 9 mission, such as payload mass or orbit type, are likely to impact its outcome. Through the utilization of multiple machine learning algorithms, we analyze historical Falcon 9 launch data to develop predictive models capable of anticipating mission results.
- Among the four machine learning algorithms utilized, the decision tree algorithm yielded the most promising performance, demonstrating its efficacy in discerning patterns within the data and producing accurate predictions for Falcon 9 launches.

Thank you!