IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Serna H.
07/29/24

Outline

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

Executive Summary

- Summary of methodologies

  The first step was performing data collection and data wrangling, later, a data analysis was performed, by plotting certain variables to search relation to the success of landing missions, as well as building maps with markers that indicate the outcome of missions. Also, a dashboard was built to present de success rates and some machine learning algorithms to predict whether the landings are going to be successful or not.

- Summary of all results

  Al the procedures showed the importance of some variables like payload mass to the success of the mission, and the machine learning algorithms had high accuracy to predict the outcome of the mission.

Introduction

The objective is to predict if the Falcon 9 first stage will land successfully. This stage can be reused wich saves SpaceX a lot of money. Thru data analysis and machine learning, the prediction of the outcome of missions can be possible, and this will provide an advantage to Aerospatiale enterprises.

Section 1

Methodology

Methodology

Executive Summary

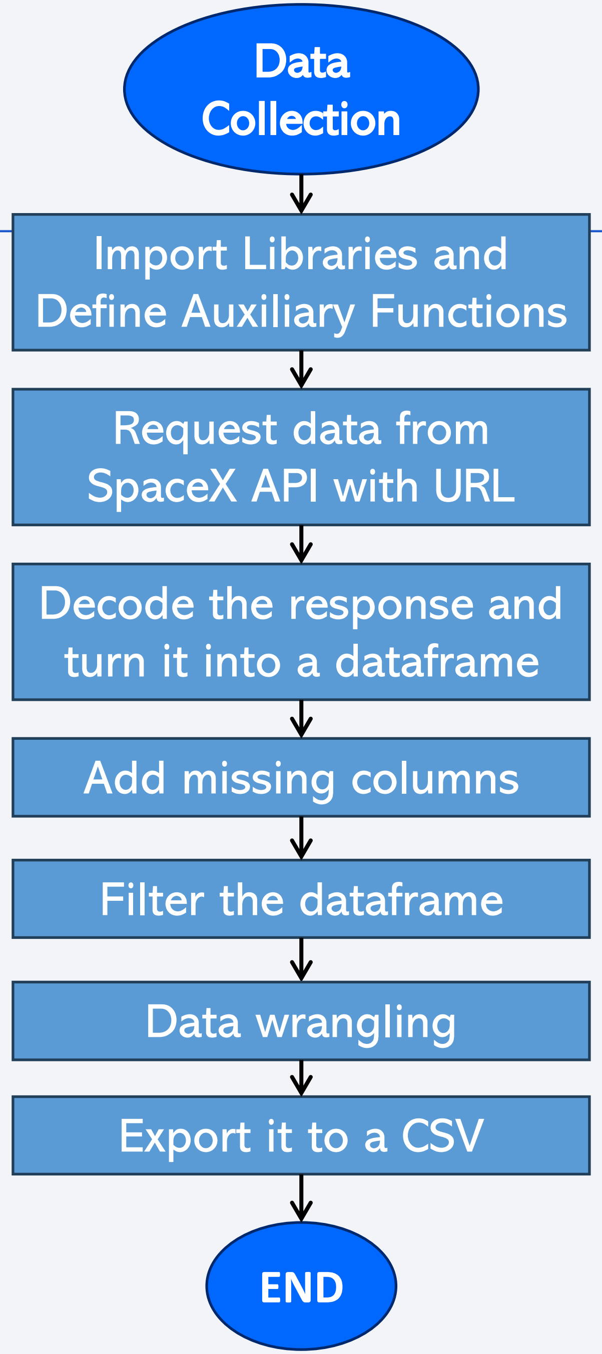
- Data collection methodology:
  - Loading data using python libraries.
- Perform data wrangling
  - Dealing with missing values and classifying the data.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building several machine learning models and testing them.

Data Collection – SpaceX API

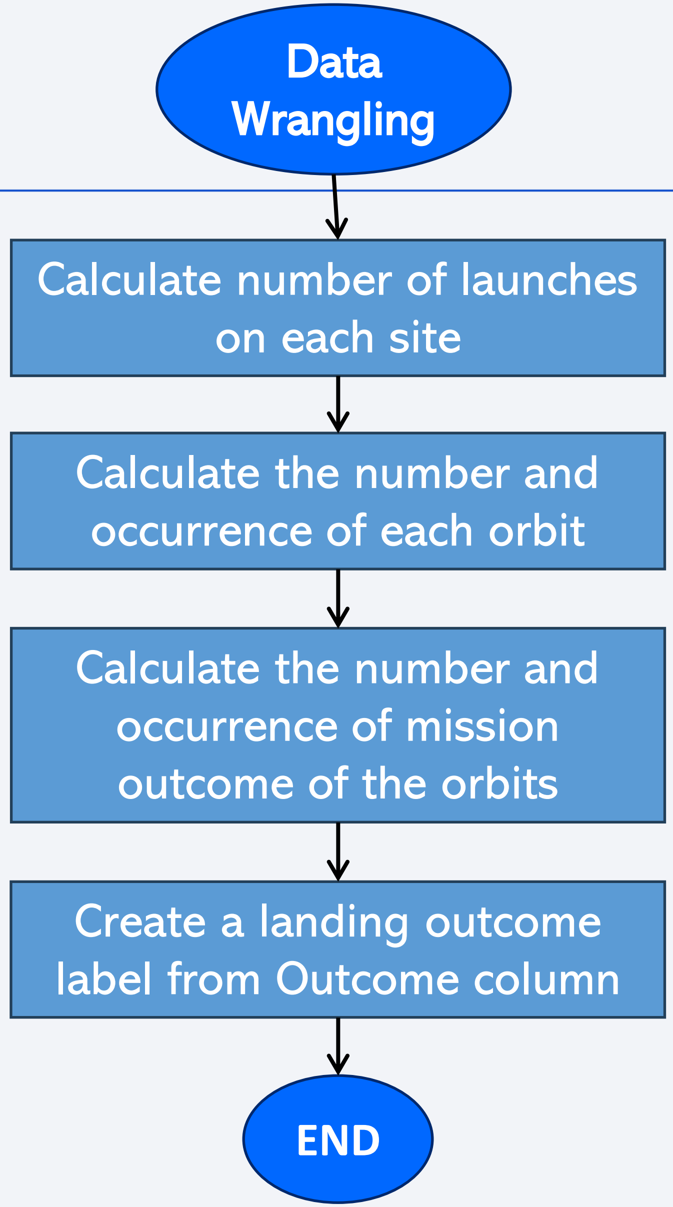
The data collection process is shown in the flow chart

[GitHub URL of the completed SpaceX API calls notebook]

The result dataframe has 90 rows and 17 columns



Data Wrangling



Once missing values and types of data were identified, there were several steps to obtain a dataframe for or data analysis, those steps are shown in the flowchart.

[GitHub URL of completed data wrangling notebooks]

EDA with Data Visualization

Charts used for exploratory data analysis:

- Scatter plot. Used to visualize relationship between variables such as Launch Site, Payload Mass, etc., using the Class variable to observe the success and failed landings.
- Bar plot. Used to visualize the success rate of the orbit types.
- Line plot. Used to observe the launch success yearly trend.

[GitHub URL of completed EDA with data visualization notebook]

EDA with SQL

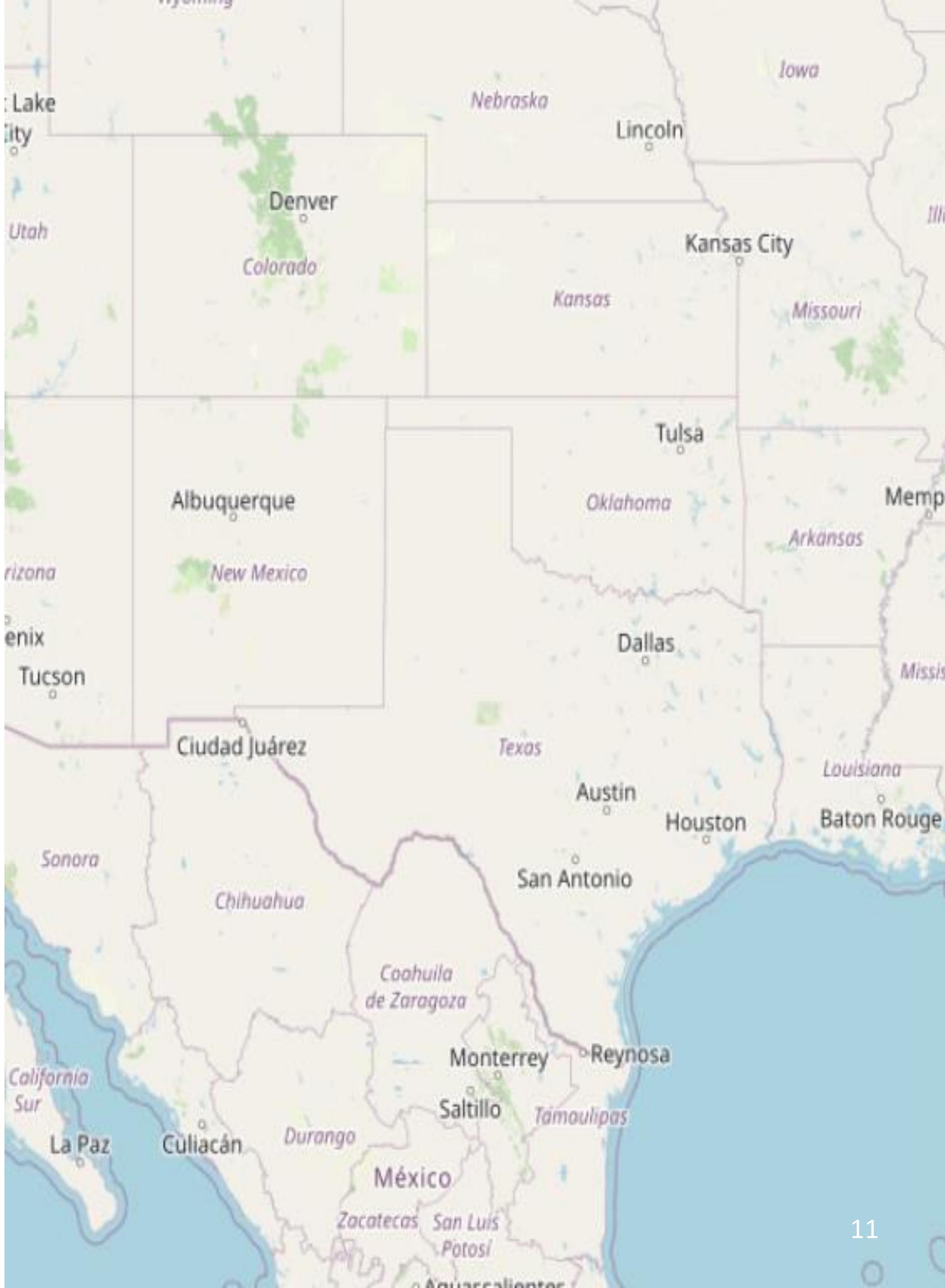
- SQL queries performed:
  - Display unique launch sites
  - Display specific launch sites
  - Display payload mass (total, average)
  - List dates of successful and failure landings
  - List names of boosters
- GitHub URL of completed EDA with SQL notebook

Build an Interactive Map with Folium

Using the dataframe with latitude and longitude for each launch site, the next folium tools were applied:

- Circles: to identify the launch sites area, as well as NASA Johnson Space Center
- Markers: to indicate the specific launch sites and identify if they were successful using color markers.
- Lines: to mark distance between launc sites and other places of interest.

[GitHub URL of completed interactive map with Folium map]()



Build a Dashboard with Plotly Dash

The dashboard allows the user to select all launch sites or only one, and gives a pie chart and a scatter plot.

- The pie chart shows the success and failure rate of the launch sites
- The scatter plot comes with a slider that allows to select a Payload Mass range and see the launch attempts in different color according to the success of the mission.

GitHub URL of completed Plotly Dash lab

Predictive Analysis (Classification)

The variables X and Y were prepared and used the train-test split method, and the train set was used to create the next estimators:

- Logistic regression
- Support vector machine (SVM)
- Decision tree classifier
- K-nearest neigbors

Next, using Grid Search CV the best parameters for each estimator were tested and the accuracy of the model was tested using de test set and method score, as well as the confusion matrix.

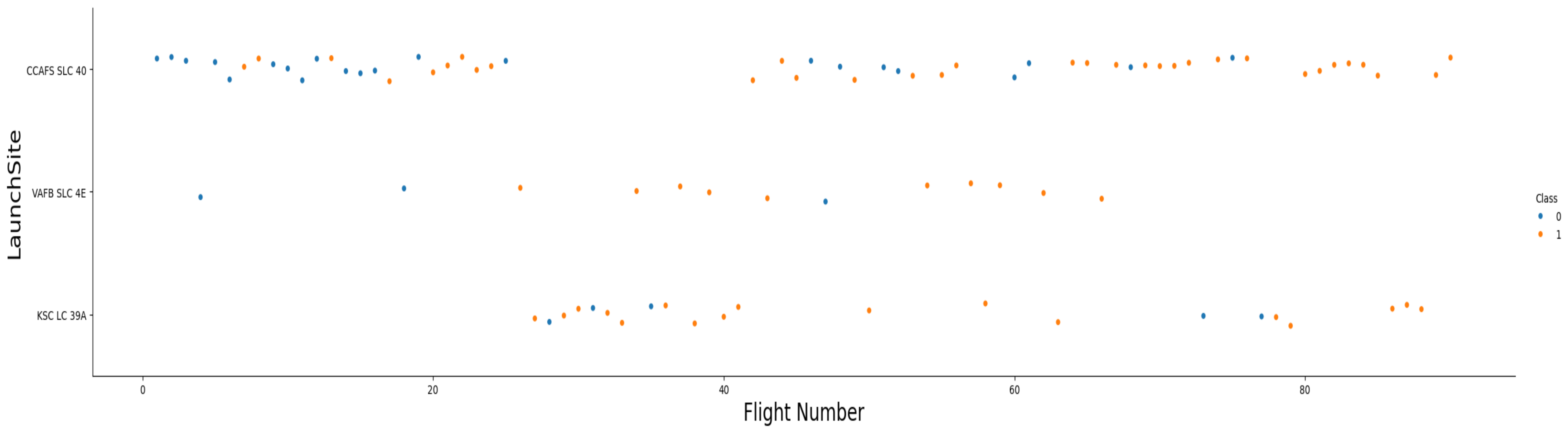
GitHub URL of completed predictive analysis lab

Section 2

Insights drawn from EDA

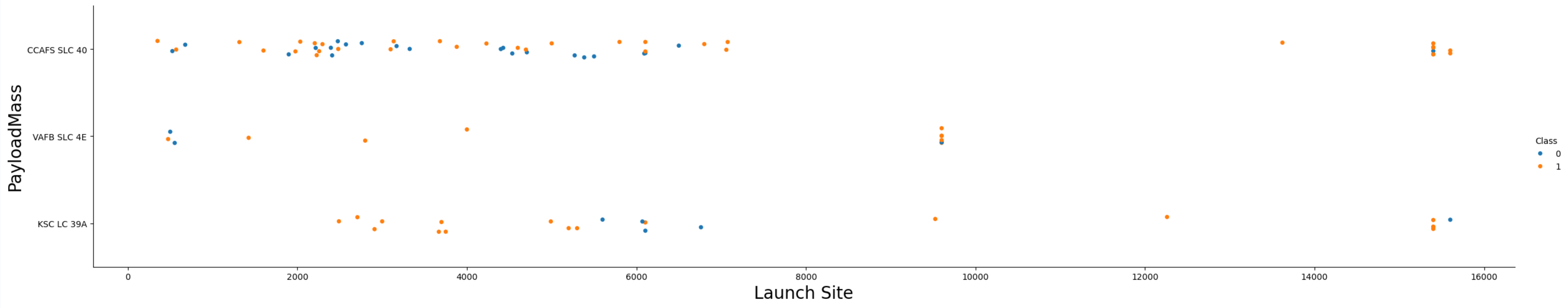
Flight Number vs. Launch Site

Relationship between flight number and launch site, showing the successful (class 1) and failed (class 0) attempts.



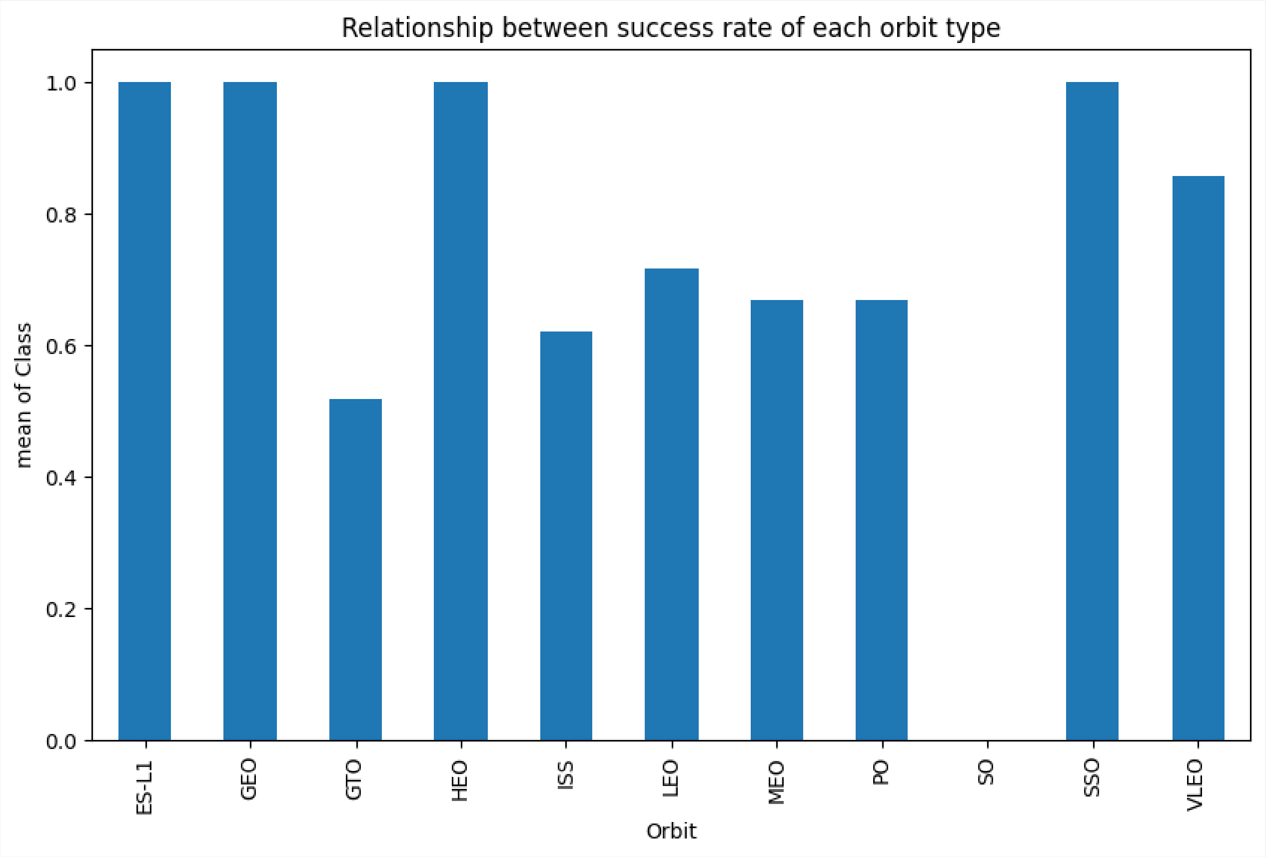
Payload vs. Launch Site

Relationship between payload mass (Kg) and launch site, showing the successful (class 1) and failed (class 0) attempts.



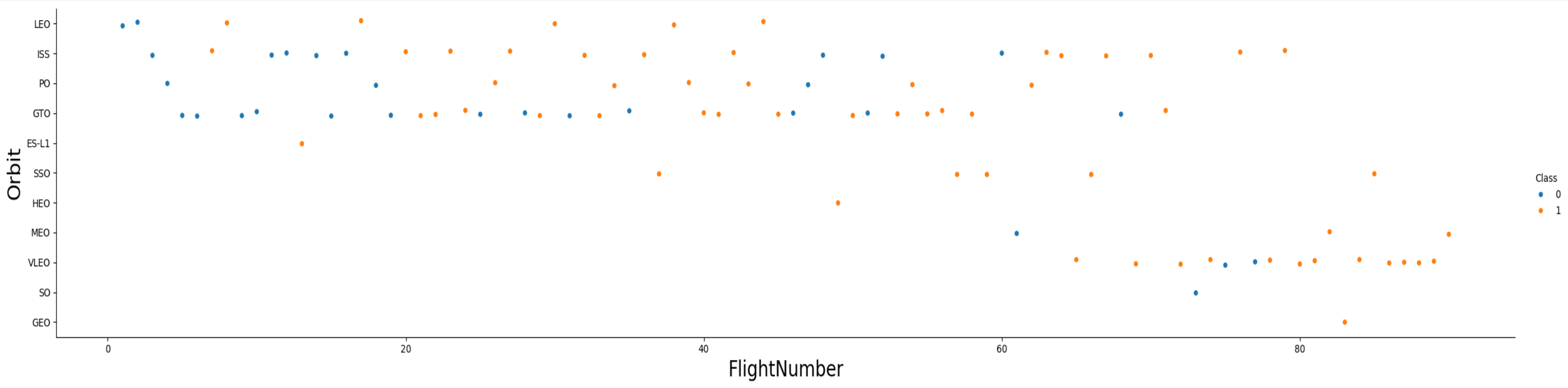
Success Rate vs. Orbit Type

Bar chart for the success rate of each orbit type



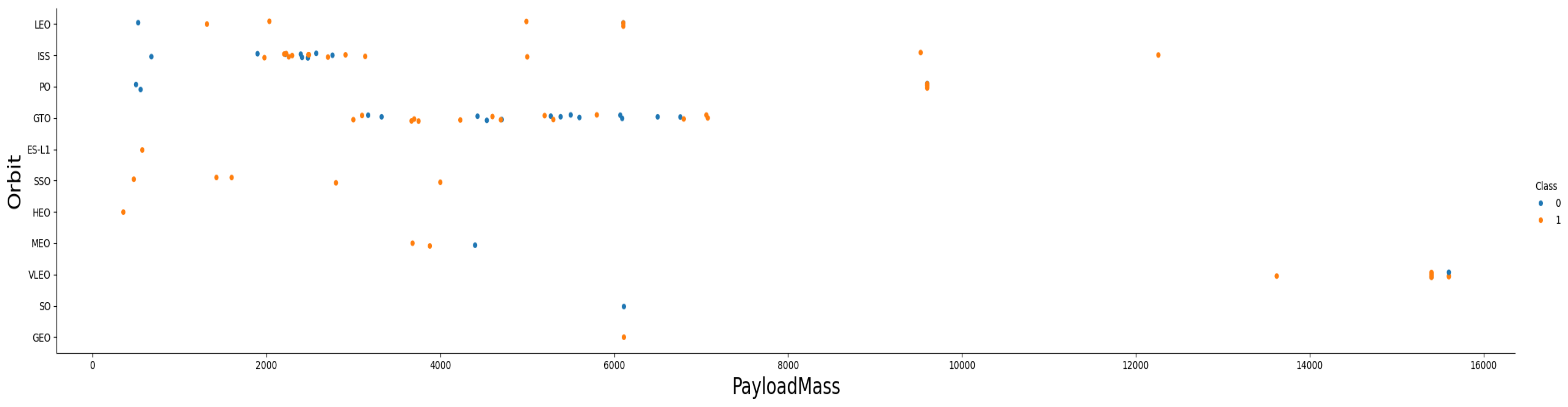
Flight Number vs. Orbit Type

Scatter point of Flight number vs. Orbit type



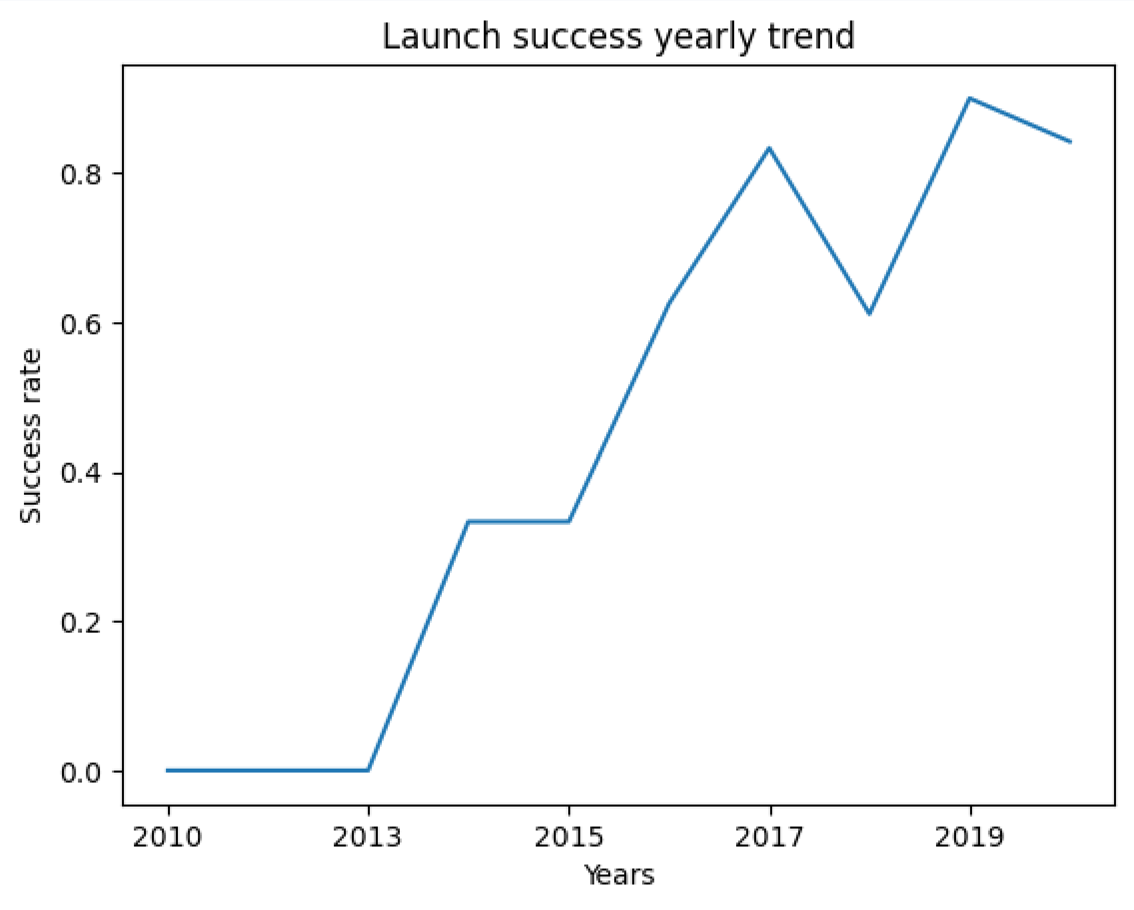
Payload vs. Orbit Type

Scatter point of payload vs. orbit type



Launch Success Yearly Trend

Line chart of yearly average success rate



All Launch Site Names

The names of the unique launch sites using DISTINCT

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

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA` using LIKE

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

Total payload carried by boosters from NASA using SUM

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

Average payload mass carried by booster version F9 v1.1 using AVG and WHERE

First Successful Ground Landing Date

Date of the first successful landing outcome on ground pad using WHERE and LIMIT

Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 kg, using WHERE and conditions.

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes

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

Boosters Carried Maximum Payload

Names of the booster which have carried the maximum payload mass

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

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 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

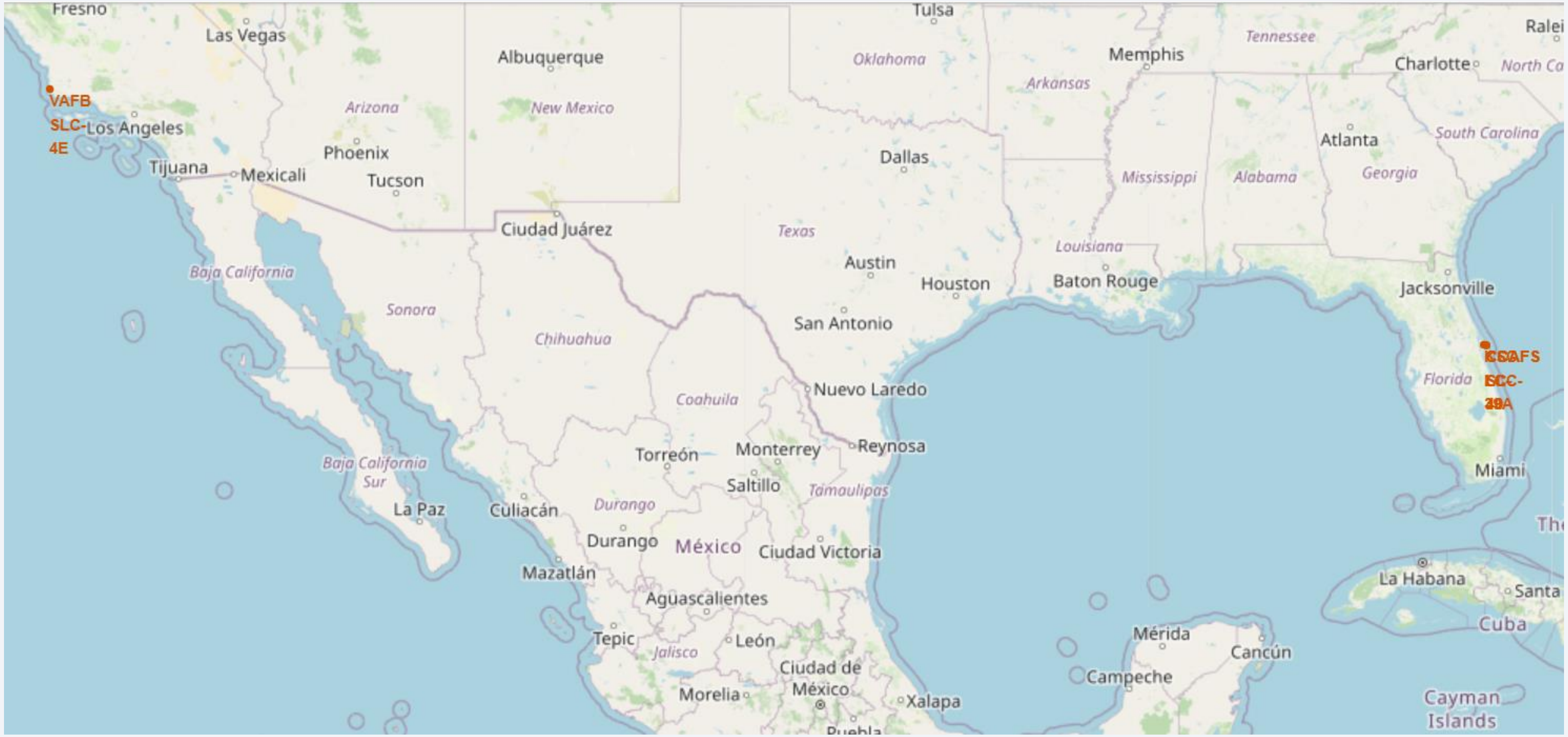
Date	Landing_Outcome	COUNT
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2010-06-04	Failure (parachute)	2
2015-06-28	Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

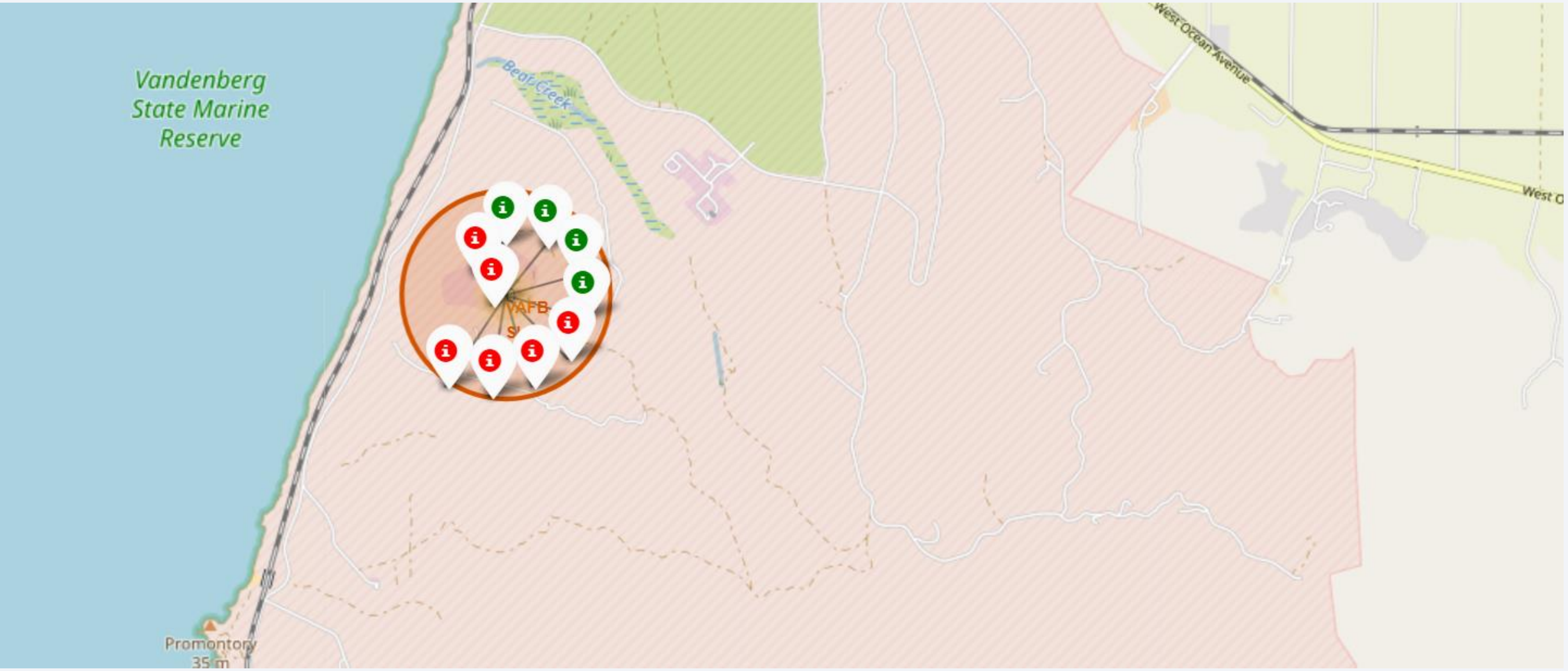
Launch location sites

All launch sites' location markers on a global map



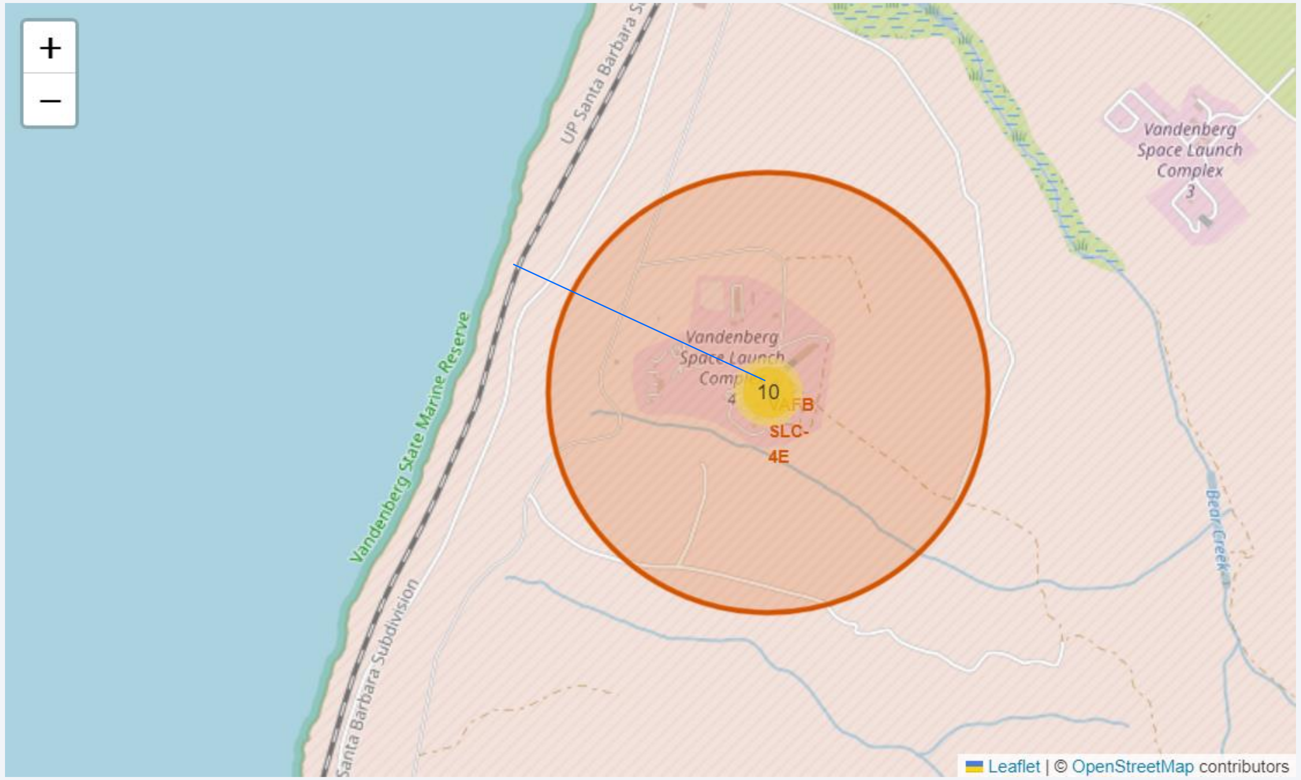
Color labeled launch outcomes in VAFB SLC-4E

Color-labeled launch outcomes on the map



Launch site proximity to railway

Launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

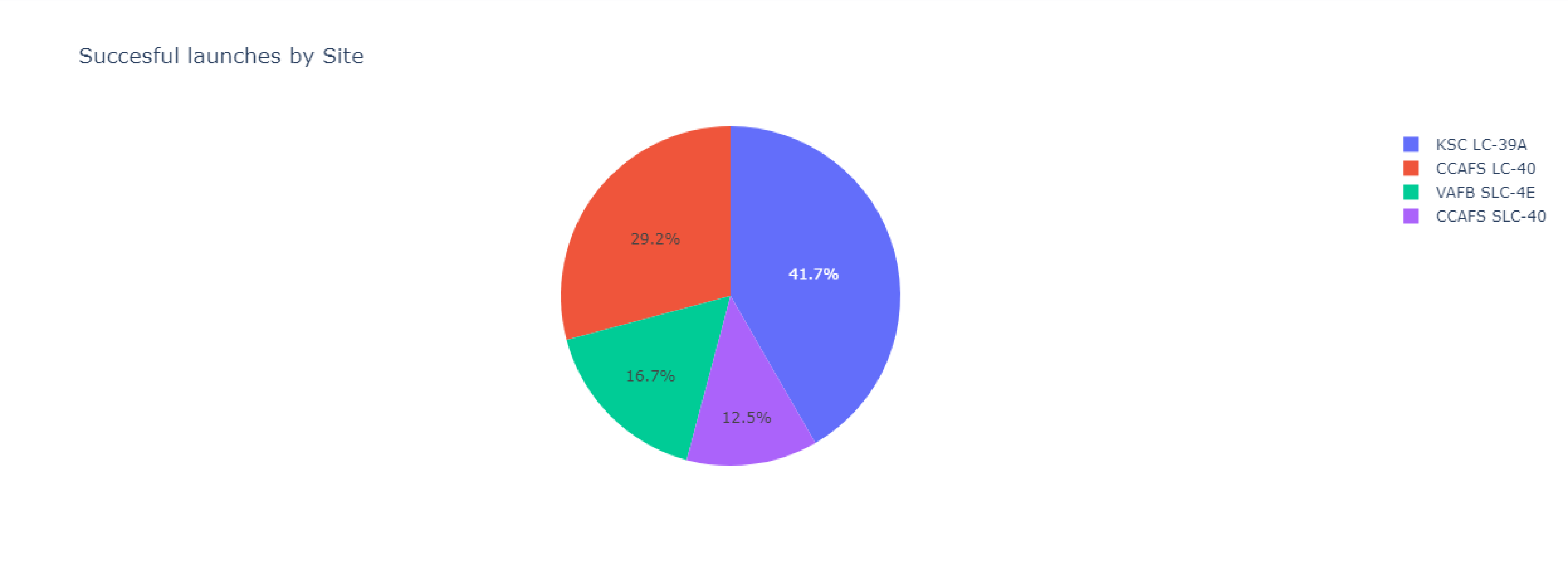


Section 4

Build a Dashboard with Plotly Dash

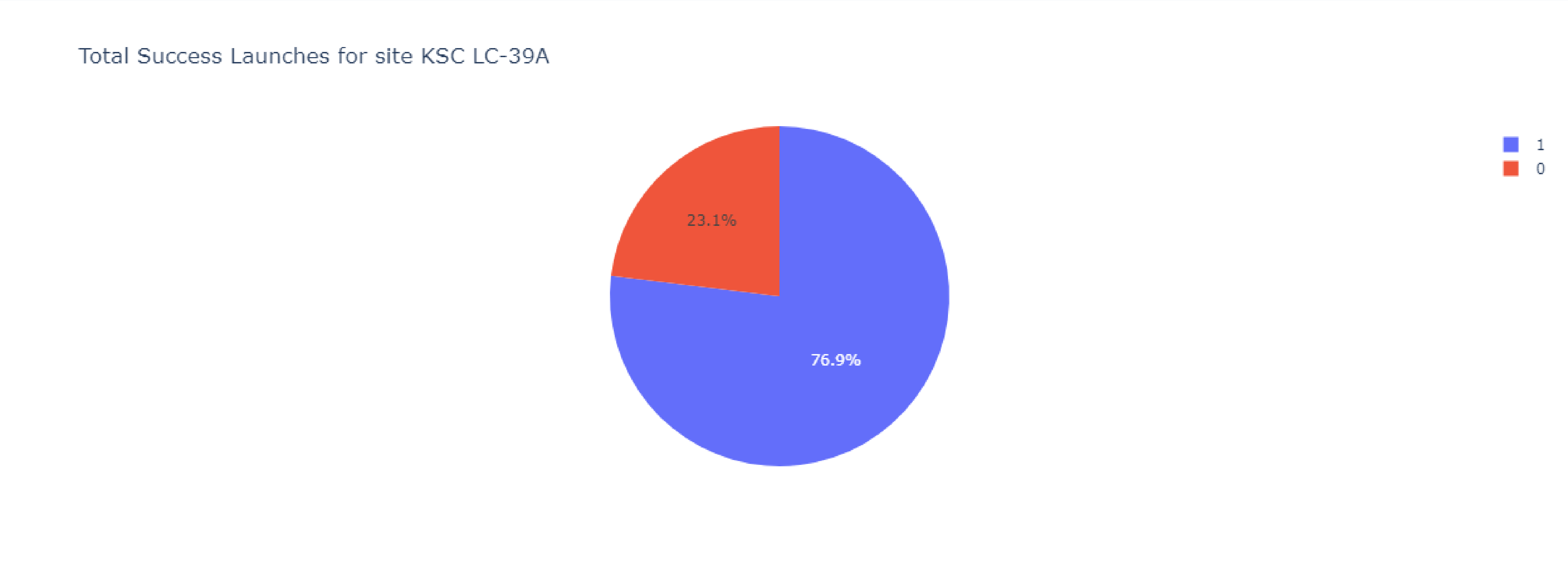
Successful launches by Site

Launch success count for all sites, in a piechart



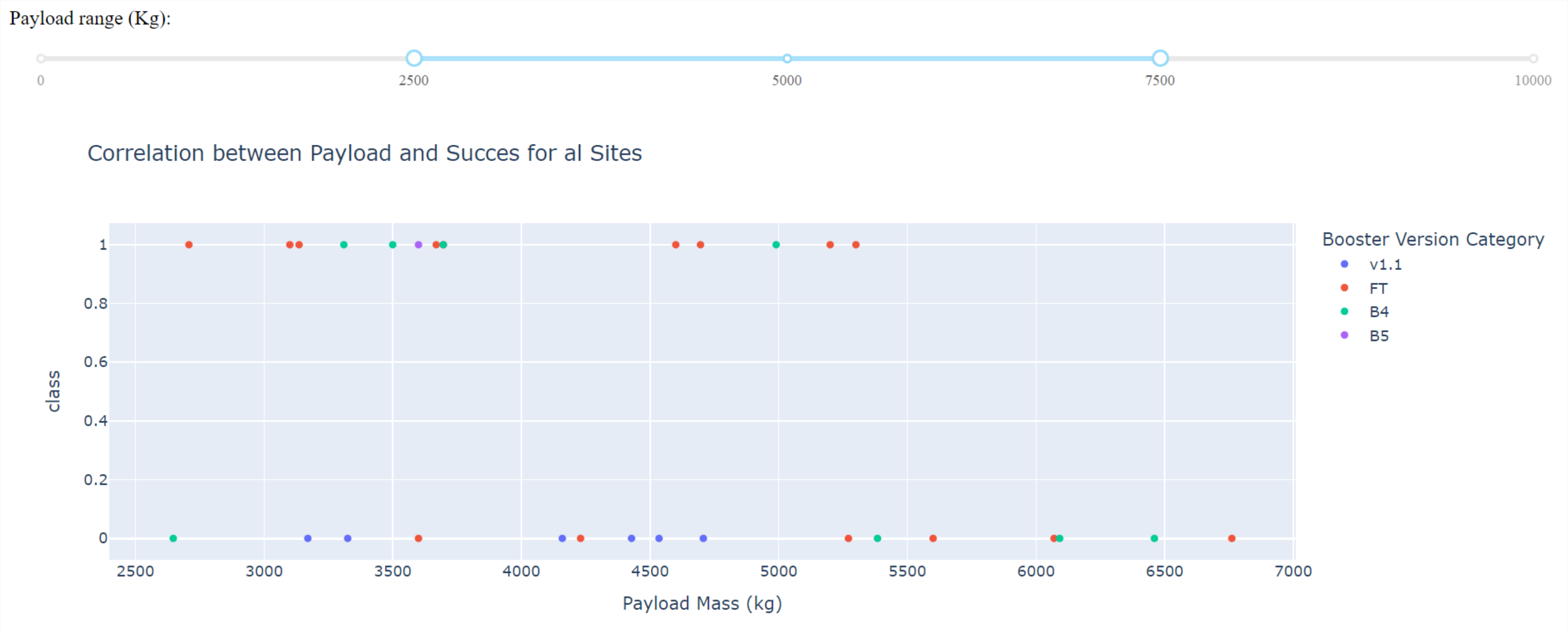
Total success launches for site KSC-LC-39A

Piechart for the launch site with highest launch success ratio



Correlation between Payload and Success for all sites

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



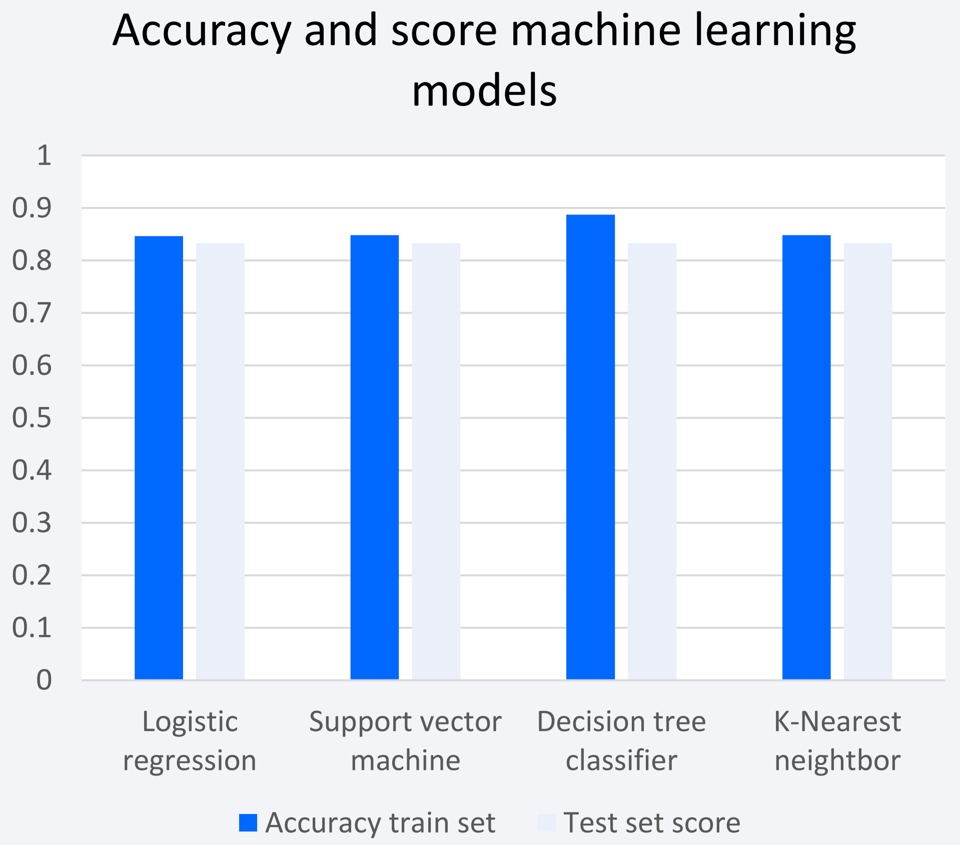
Section 5

Predictive Analysis (Classification)

Classification Accuracy

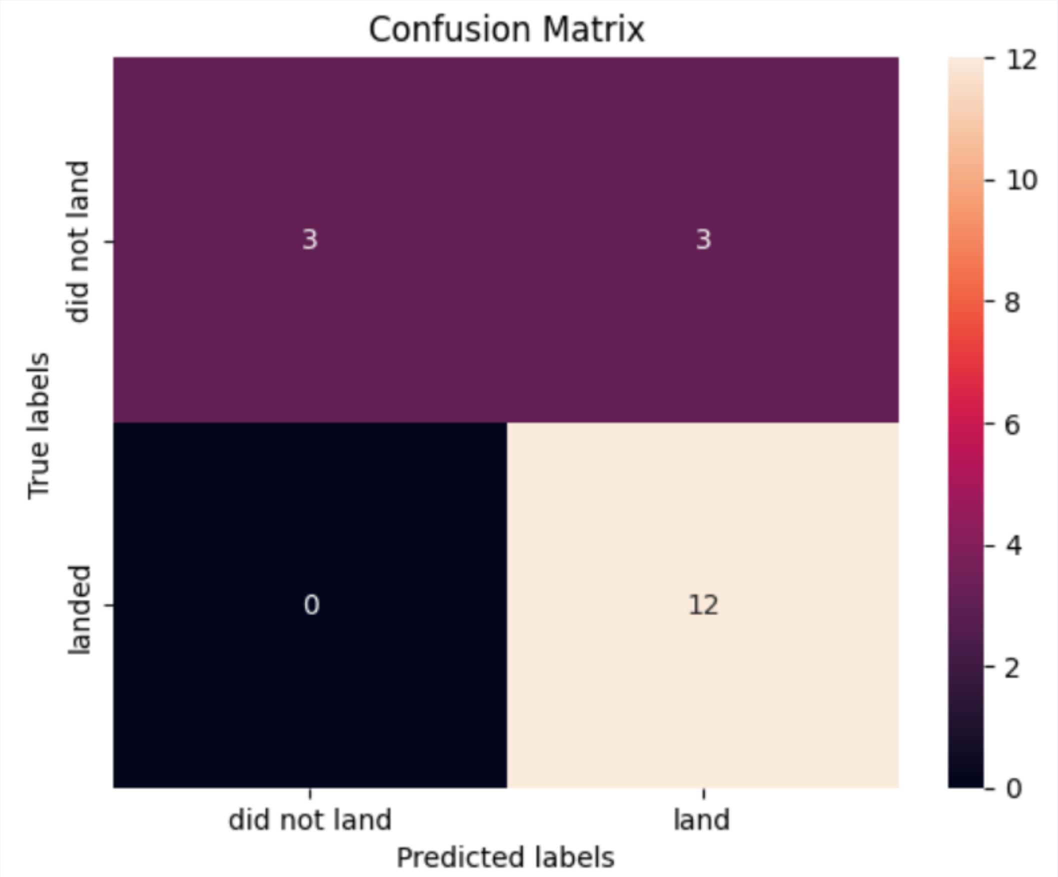
According to the bar plot, all models have the same test set score, but decision tree classifier had the highest accuracy using the train set.

	Accuracy train set	Test set score
Logistic regression	0.8464	0.8334
Support vector machine	0.8482	0.8334
Decision tree classifier	0.8875	0.8334
K-Nearest neightbor	0.8482	0.8334



Confusion Matrix

All models had the same confusion matrix performance, which showed a problem fit false positives.



Conclusions

- During the process of data analysis, it was found that the payload mass had an effect in the outcome of the landing missions.
- In the yearly success rate, it was shown that the outcome of missions improved with time.
- The KSC-LC-39A launch site showed a higher success than the other sites.
- During the machine learning process there were 4 models, all of them had the same behavior with the test set and showed the same confusion matrix.

Thank you!