IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kevin Bernardo
Sunday, 17 September 2022

Outline

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

Executive Summary

- This project tried to predict, if the Falcon 9 first stage will land successfully or not.
- The data used were collected, cleaned, analyzed using EDA, visualized, and also used for the prediction using different machine learning methods.
- The results were divided into the importance of each features for the machine learning methods, visuals including maps and dashboard, and prediction model.

Introduction

Background

- SpaceY wants to compete with the SpaceX for price bidding of a rocket launch.
- Price of the launch is essential factor for winning the price bidding.
- To determine the price of a launch, we must predict, if the first stage of the first stage of the launch will land or not, since parts from this stage can be reused.

Question

- Which features influence the success of the launch?
- Which machine learning model should be used to predict the launch success?

Section 1

Methodology

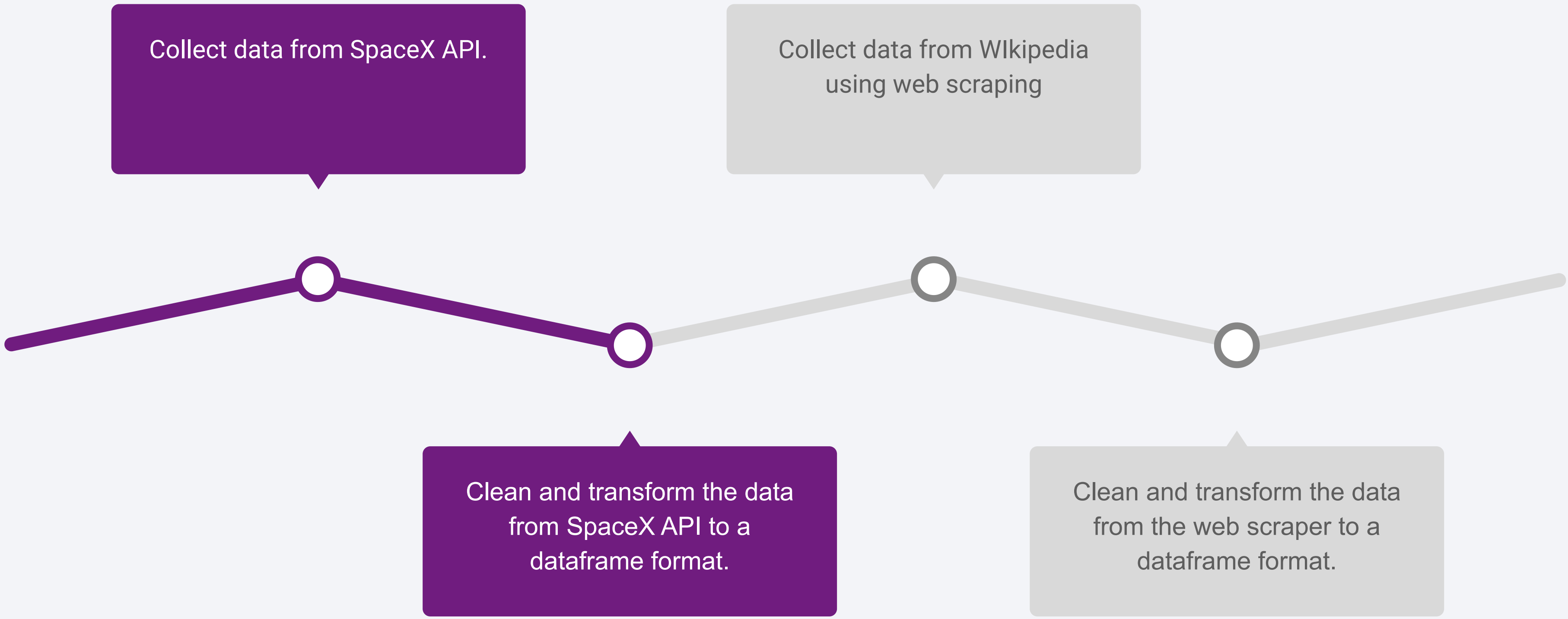
Methodology

Executive Summary

- Data collection methodology:
  - The data were gathered from the SpaceX API and Web scraping from Wikipedia.
- Perform data wrangling
  - Changing the feature values, select features, and filling the missing values.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Train test split, using GridSearchCV for hyperparameter tuning, and using accuracy to find the best model.

Data Collection

- The dataset was collected from the SpaceX API mainly, with more information about the Falcon 9 historical launch records from Wikipedia using web scraping



Data Collection – SpaceX API

Request and parse data → **Filter data** → **Save data**

Request and parse data

- Use GET request to get the data.
- Take a subset of the data such as rockets, payloads, launchpads, cores, flight number, and date.
- Get more information about the booster, launch site, payload, and core data.
- Save the data in dataframe.

Filter data

- Get data only from Falcon 9 booster.

Save data

- Save the data in CSV format.

[Link to the Jupyter notebook.]()

Data Collection - Scraping

Request and extract data	Parse data	Save data
• Request the Wikipedia table using BeautifulSoup. • Extract the column name of the HTML table.	• Parse each row of the HTML table into a dataframe.	• Save the data in CSV format.

Link to the Jupyter notebook.

Data Wrangling

Fill missing values	Count feature values	Create new column

Fill missing values

- Fill missing values in the payload mass column using the mean of the data.

Count feature values

- Calculate number of launches on each launch site.
- Calculate number and occurence of each orbit.
- Calculate the number and occurence of mission outcome.

Create new column

- Create new column to determine the outcome of the launch.

[Link to the Jupyter notebook.]

EDA with Data Visualization

- Using categorical plot and scatter plot from seaborn to find the distribution and correlation of the data between flight number, payload, launch site, and orbit type.
- Using bar plot to find the success rate of the launch based on orbit type.
- Using line plot to find the yearly trend of the launch success rate.
- Link to the Jupyter notebook.

EDA with SQL

- Display data about the launch sites, booster versions, and launching outcome.
- Calculate total and average payload mass of certain booster versions.
- Link to the Jupyter notebook.

Build an Interactive Map with Folium

- Using circle and marker map object to show and annotate the launch sites on the map.
- Using marker cluster map object to group multiple launches for a launch site with colors to determine the success of the launch.
- Using polyline map object to draw lines from the launch site to its proximities.
- Link to the Jupyter notebook.

Build a Dashboard with Plotly Dash

- Using dropdowns to select the intended launch sites for the visualization.
- Using pie chart to show the percentage of the success and fail launches for a certain or all launch sites.
- Using scatter plot to show the relationship between the payload mass and the launch outcome.
- Using range slider to determine the range of the payload mass shown in the scatter plot.
- Link to the python file.

Predictive Analysis (Classification)

Scale and create train and test data

Tune model hyperparameter

Evaluate model

- Scale the data using standard scaler.
- Divide the dataset into train and test data with 8:2 ratio.

- Use GridSearchCV to find the best hyperparameter for logistic regression, SVM, decision tree, and KNN.

- Calculate the accuracy of the prediction for each model.
- Draw the confusion matrix of the prediction for each model.
- Determine the best model between logistic regression, SVM, decision tree, and KNN.

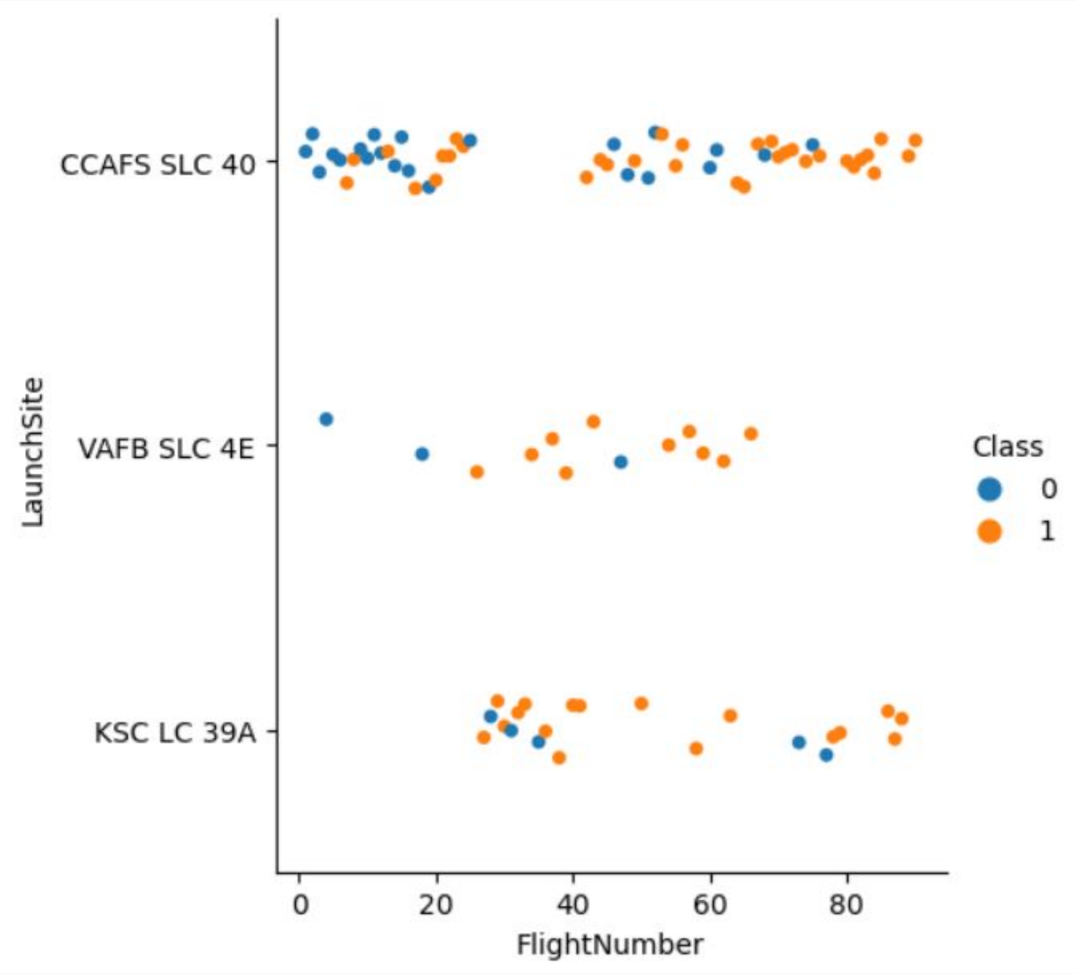
[Link to the Jupyter notebook.]

Section 2

Insights drawn from EDA

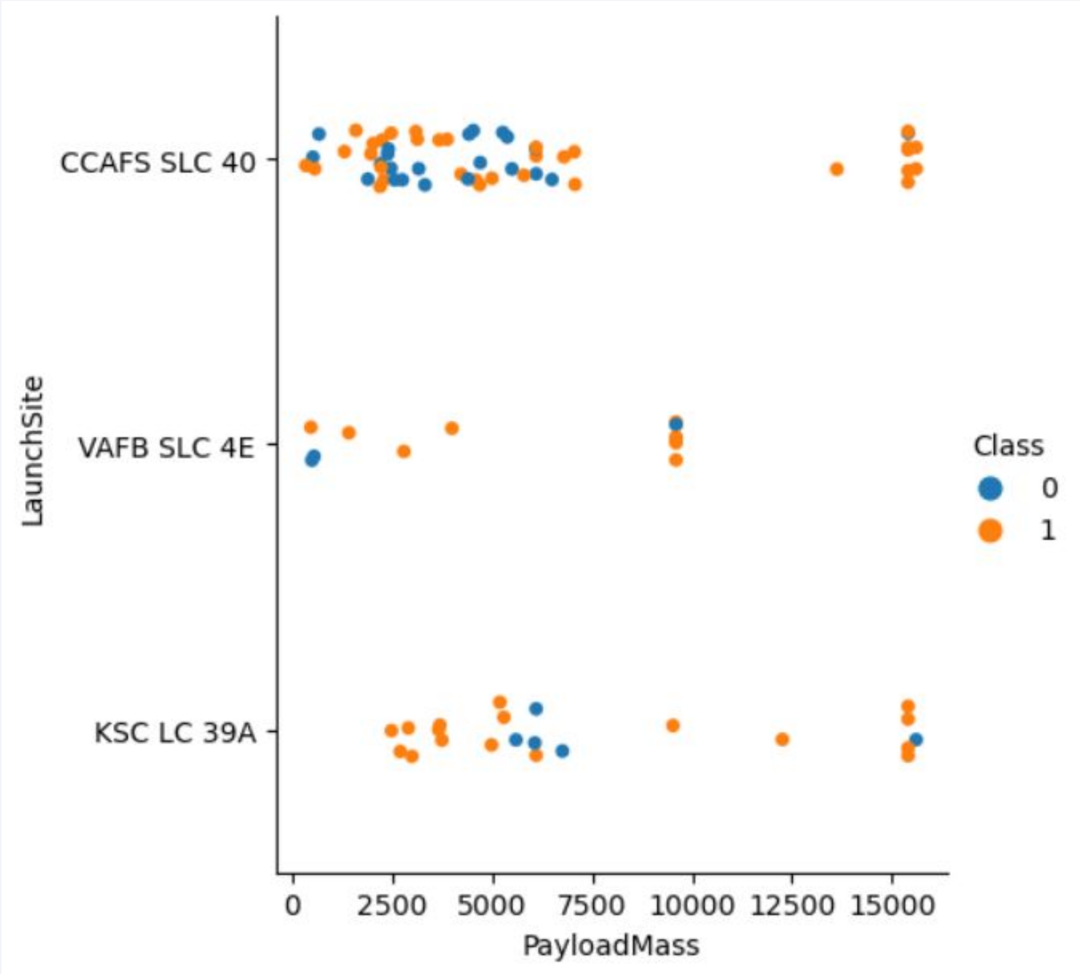
Flight Number vs. Launch Site

- Most of the launches were done in CCAFS SLC 40, and it also has the most fail launches.
- VAFB SLC 4E has the least amount of launches.



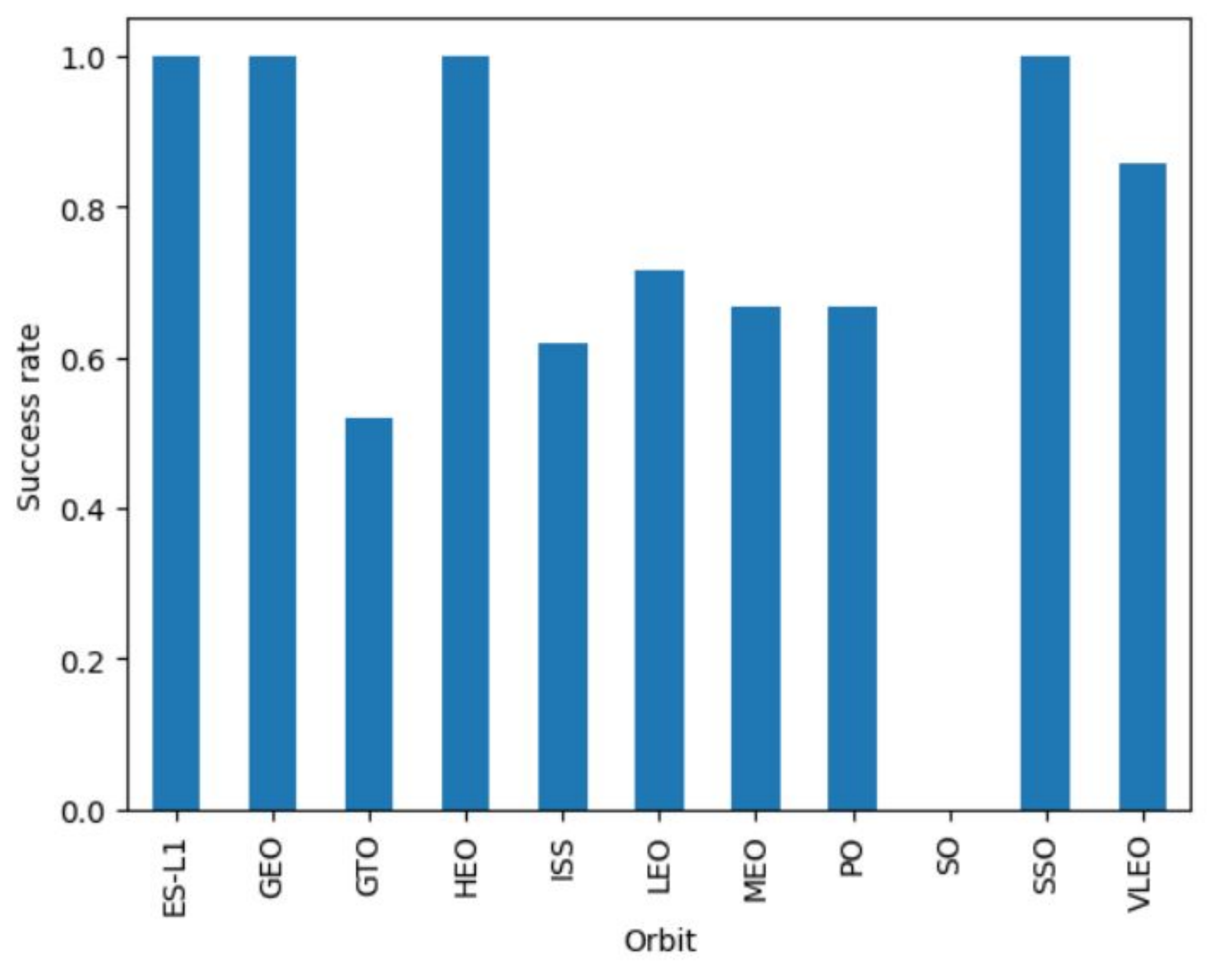
Payload vs. Launch Site

- VAFB SLC 4E has launches with payload lighter than 10000 kg.
- KSC LC 39A and CCAFS SLC 40 and KSC LC 39A have heavy launches (>10000 kg), and most of them were successful.



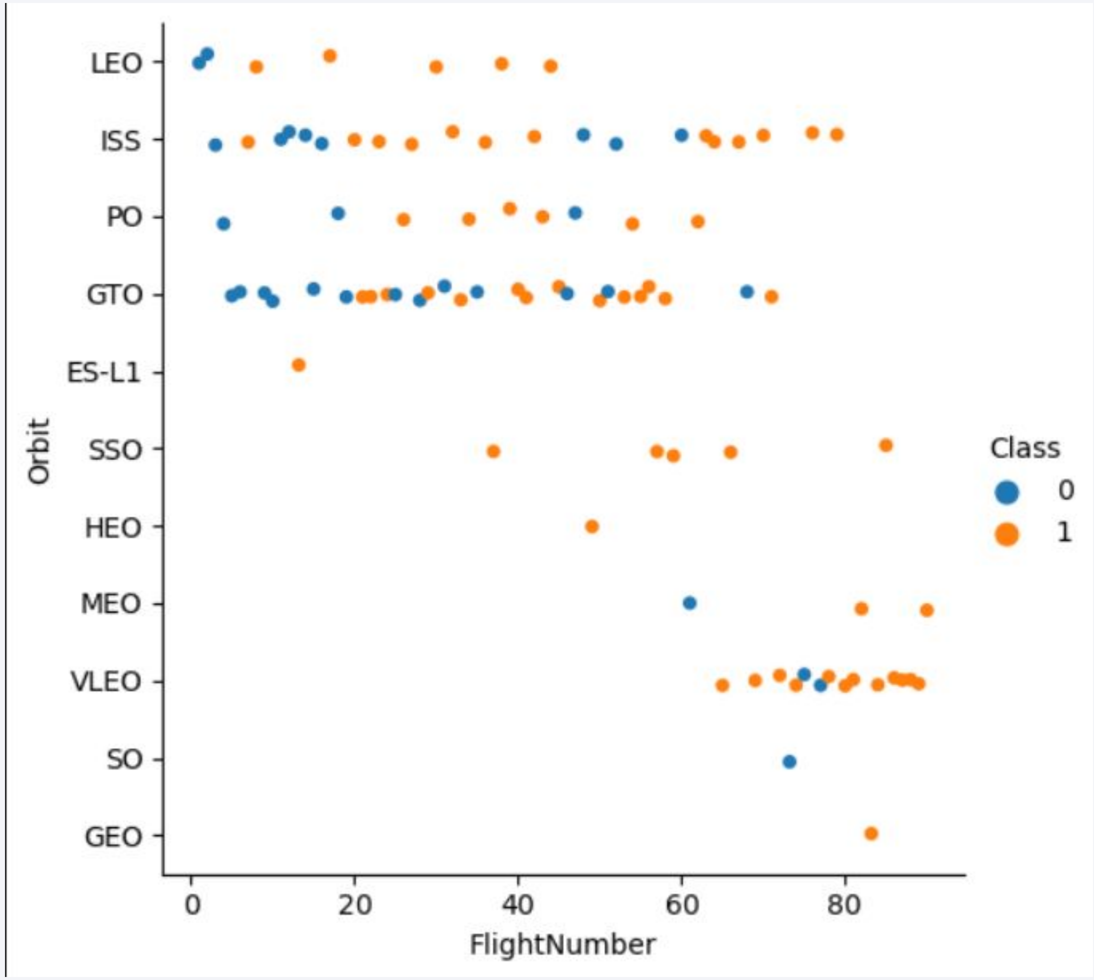
Success Rate vs. Orbit Type

- The launch to SO was unsuccessful.
- Most of the launches have success rate more than 50%.
- Launches to ES-L1, GEO, HEO, and SSO have the 100% success rate.



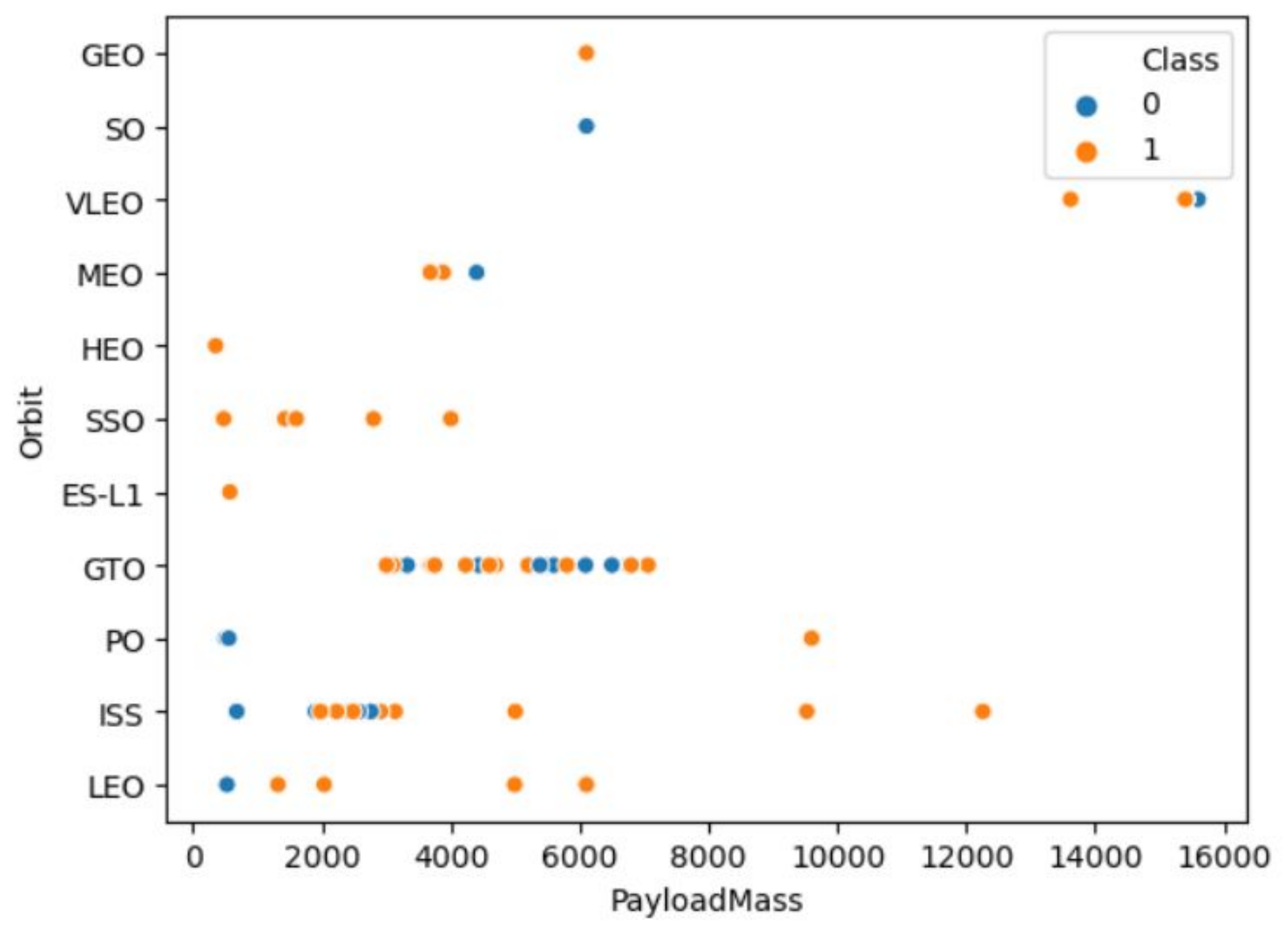
Flight Number vs. Orbit Type

- SO, ES-L1, HEO, and GEO had only one launch.
- GTO and ISS had the most launches.



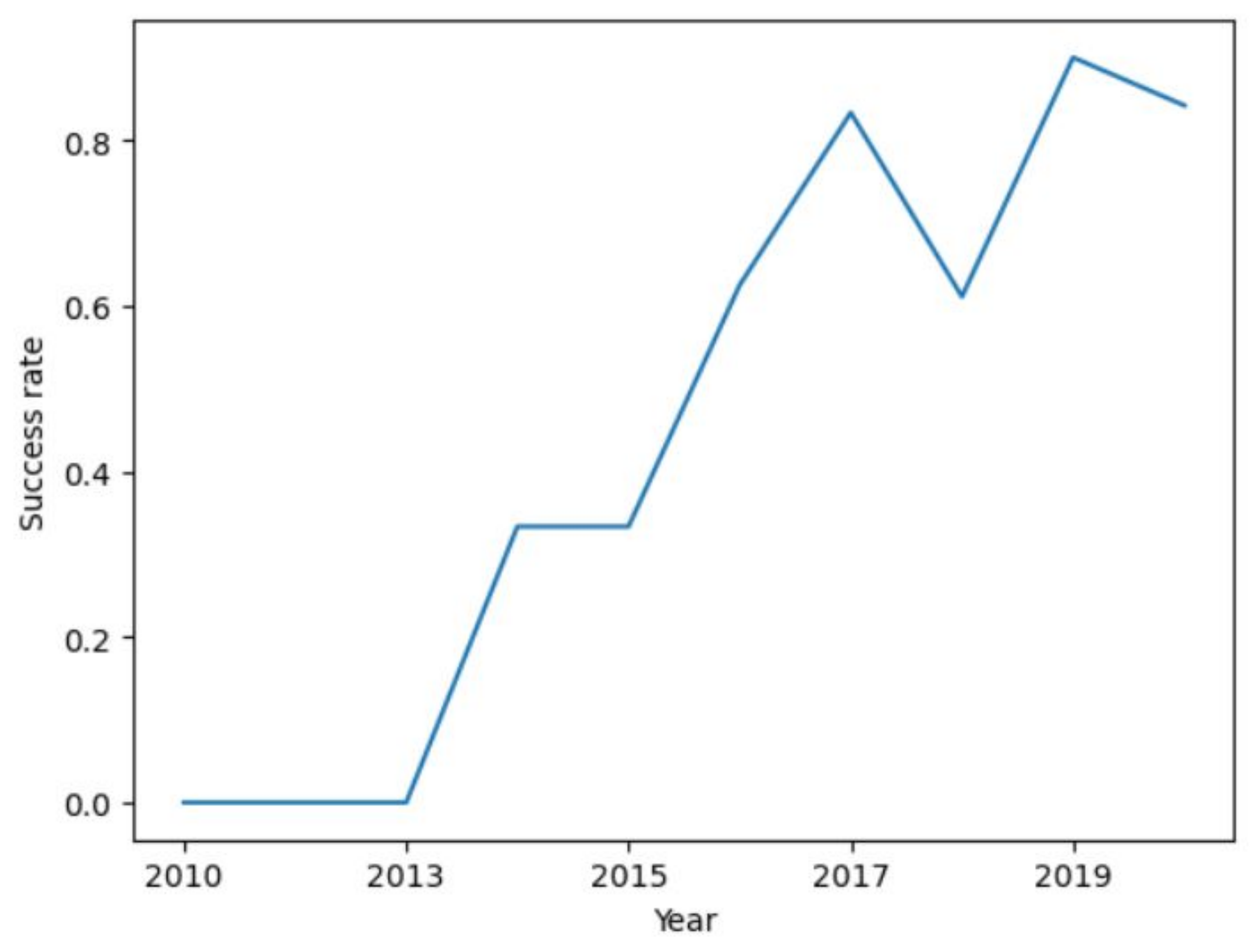
Payload vs. Orbit Type

- VLEO had the heaviest launches in comparison to all other orbit.
- Most of ISS payloads were light.
- HEO, SSO, ES-L1, and LEO have light payloads.



Launch Success Yearly Trend

- The launch success increases since 2013 to 2020.



All Launch Site Names

- Names of the unique launch sites.

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

Launch Site Names Begin with 'CCA'

- 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

- Total payload carried by boosters from NASA is 45596 kg.

1
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is 2928 kg.

1
2928

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad is 22 December 2015.

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- 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

- Total number of successful and failure mission outcomes is 71.

1
71

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	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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. Most of them are 'No attempt'.

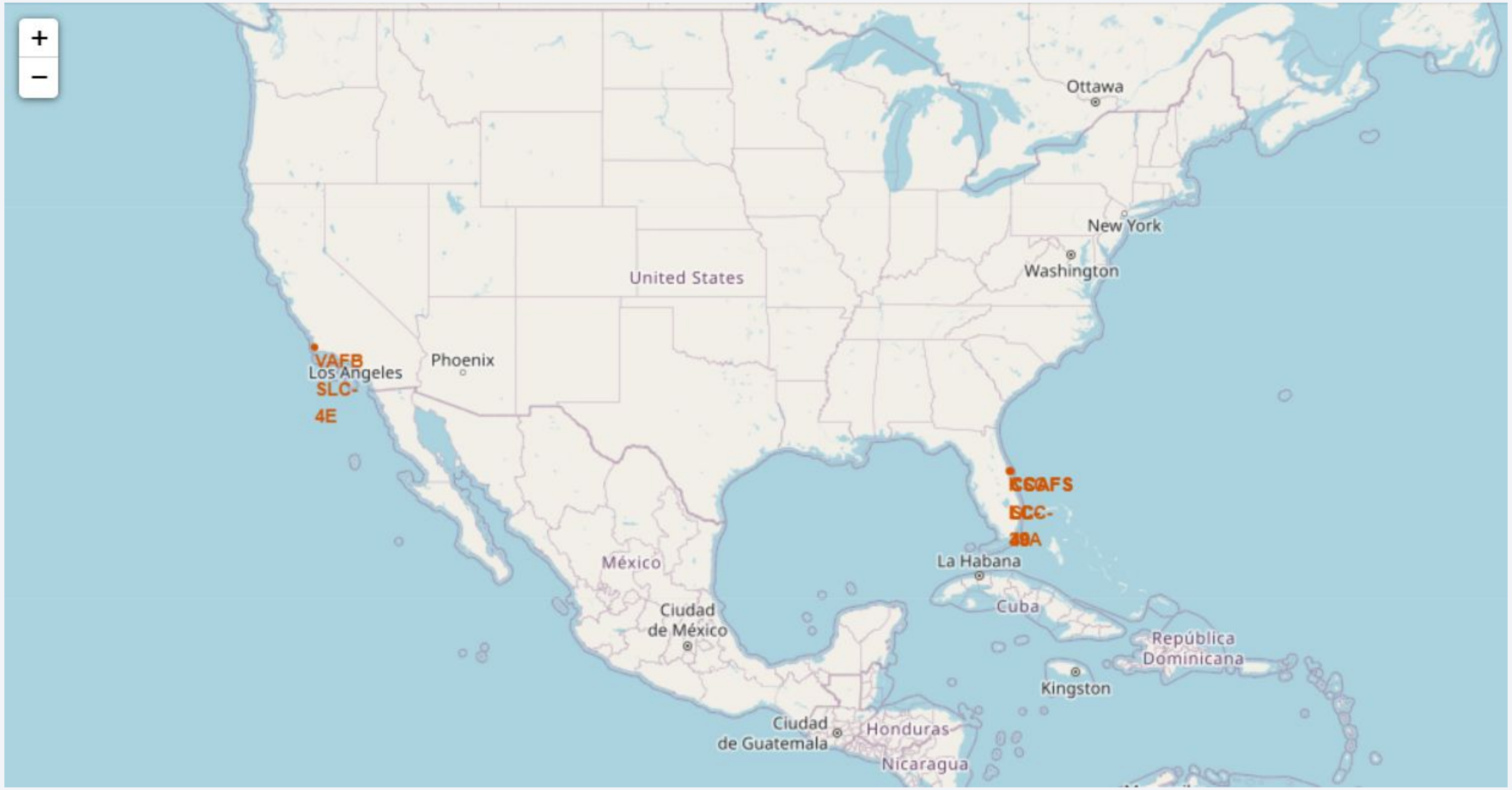
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Sites on Map

The launch sites are located in the east and west coast, close to the equator.

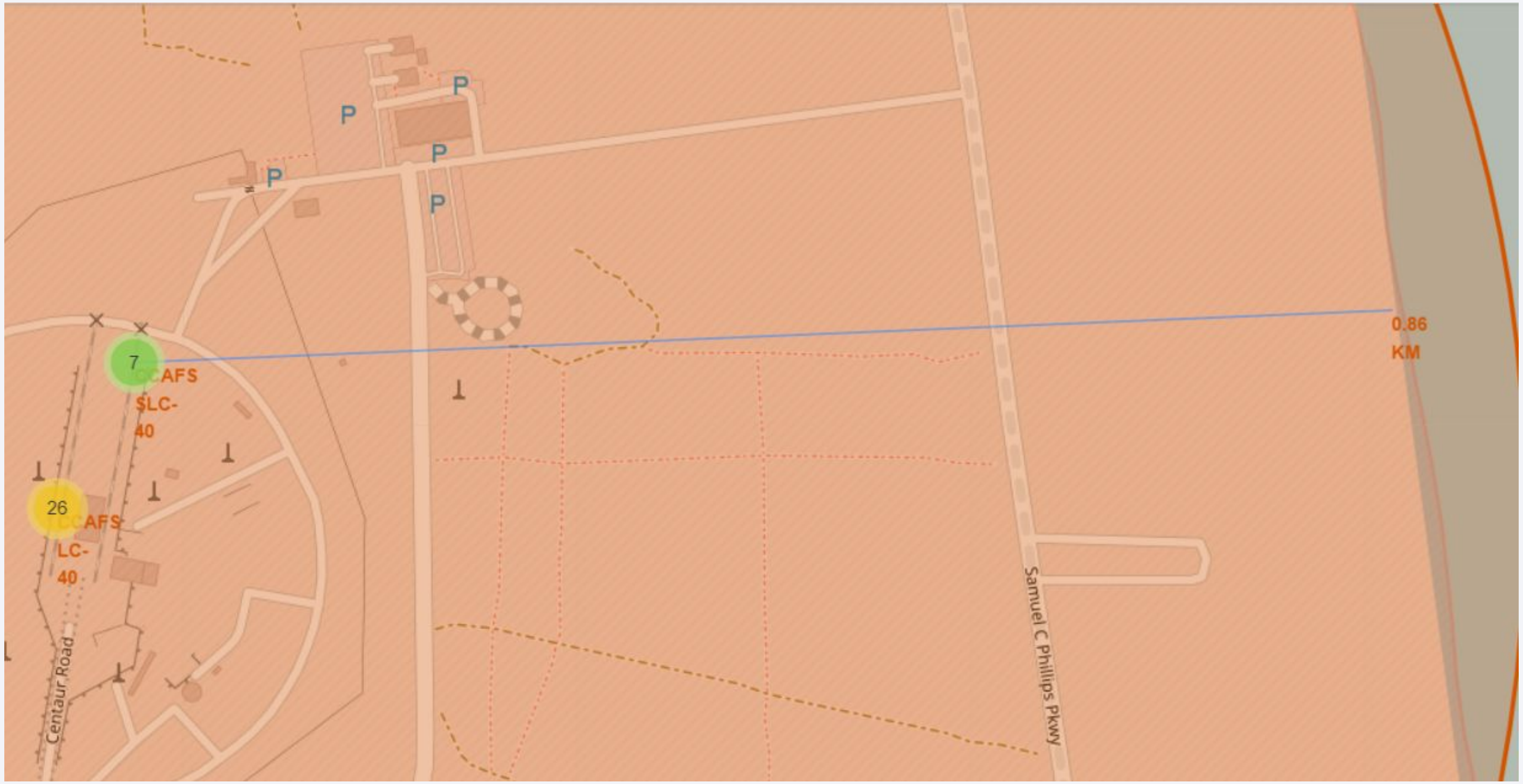


Launch Outcomes on Map

The marker color shows the launch outcome in each launch site. The screenshot shows the markers from launch site CCAFS LC-40

Launch Site Proximities on Map

CCAFS SLC-40 is 860 m away from the coast.

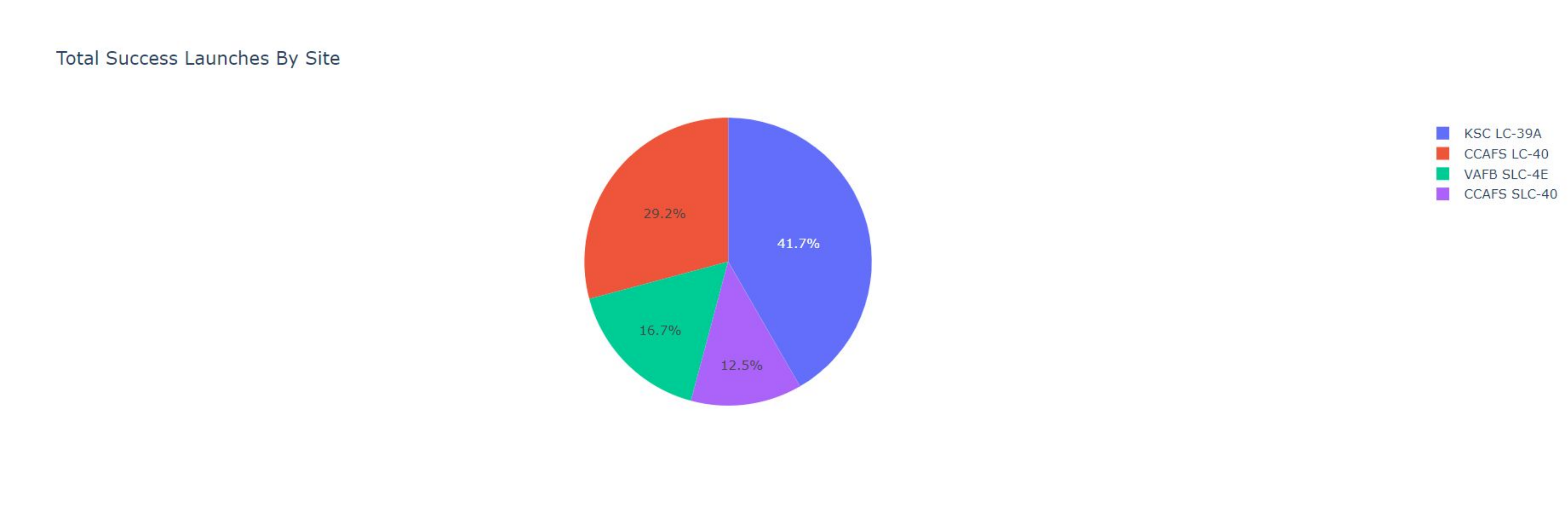


Section 4

Build a Dashboard with Plotly Dash

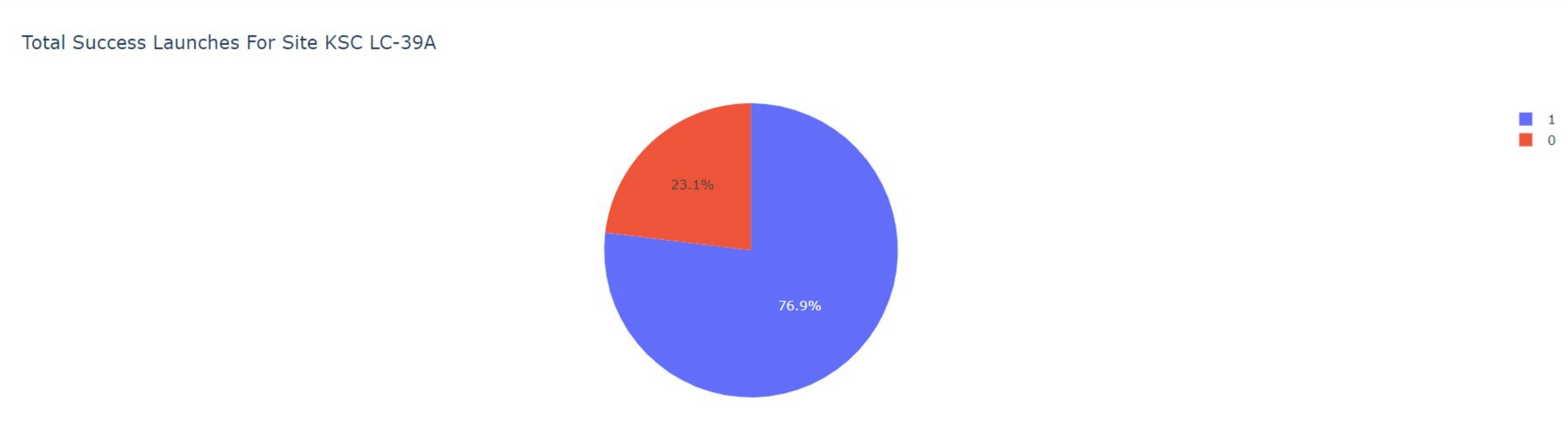
Success Rate of Launch Sites

KSC LC-39A has the most successful launches in comparison to other sites.



Success Rate of KSC LC-39A

76.9% of the launches from KSC LC-39A are successful.



Payload and Success

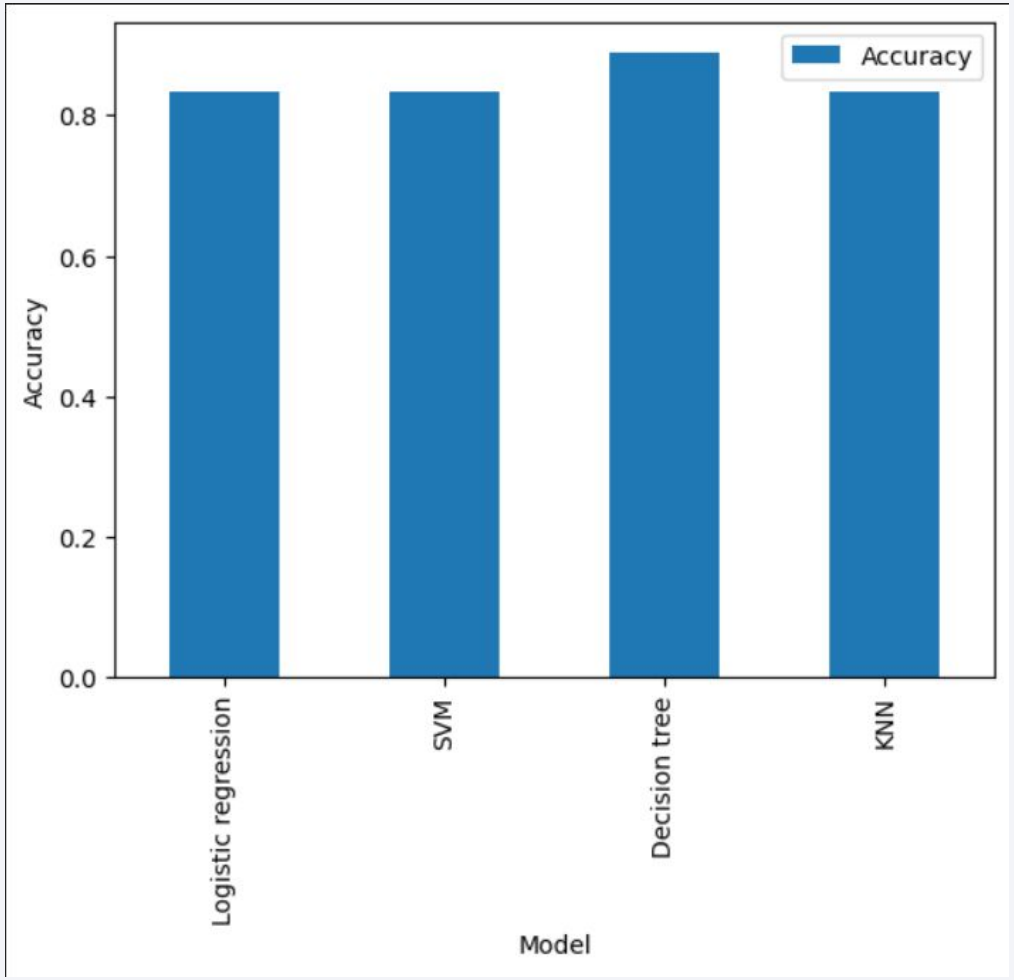
There is no direct correlation between the payload mass and the launch outcome.

Section 5

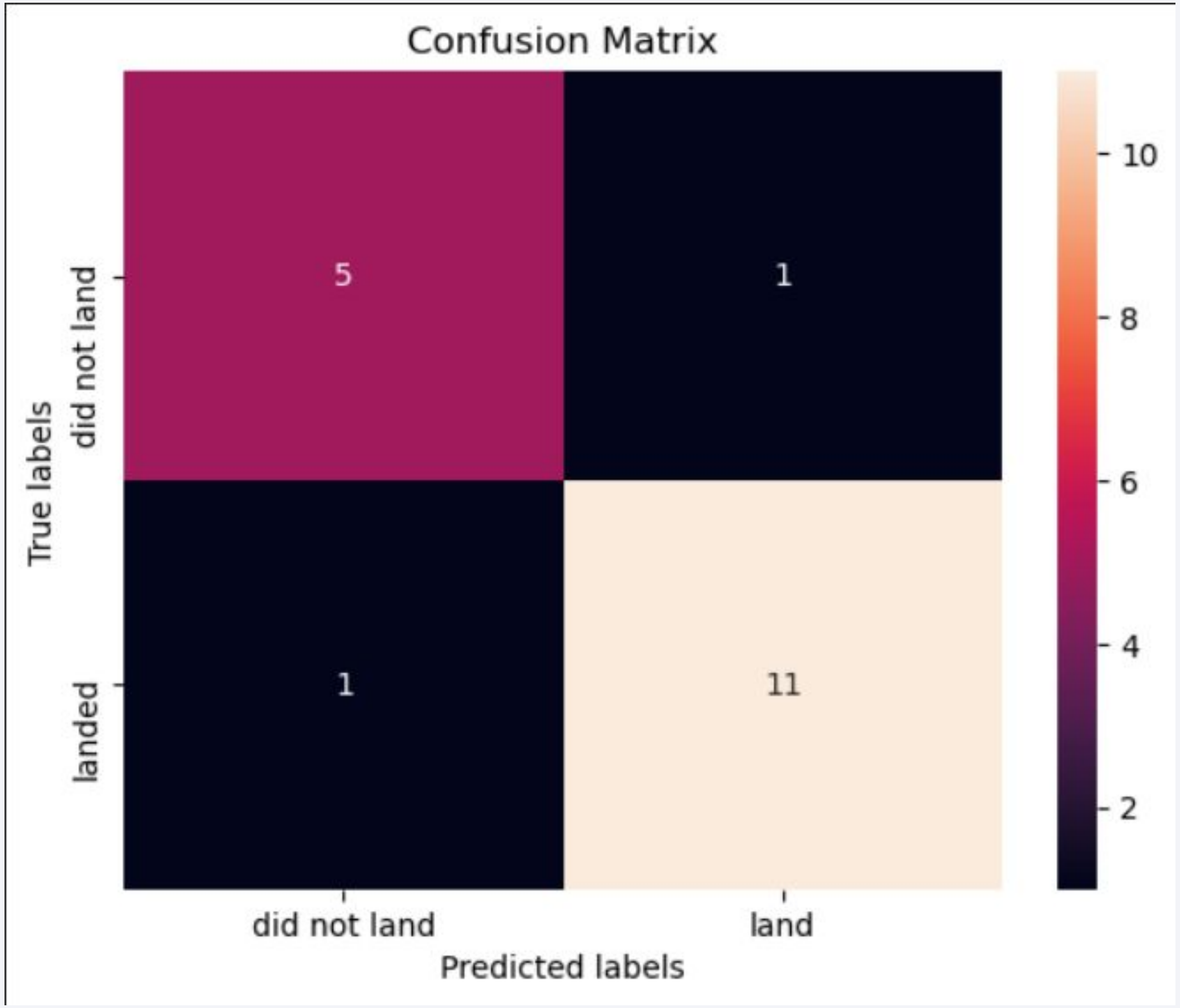
Predictive Analysis (Classification)

Classification Accuracy

- Decision tree model has the best accuracy (88.89%).



Confusion Matrix of Decision Tree



Conclusions

- Orbit type and launch site has an influence on the launch success.
- Decision tree has the best accuracy, and therefore the best model for the prediction.

Thank you!