Winning Space Race with Data Science

Rihards Rusins
09.07.2023.

Outline

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

Executive Summary

- Data analysis involved the utilization of the following methodologies:
  - Data Collection using web scraping and SpaceX API;
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics;
  - Machine Learning Prediction.
- Summary of all results
  - It was possible to collected valuable data from public sources;
  - EDA allowed to identify which features are the best to predict success of launchings

Introduction

- The aim is to assess the potential of the newly established company, Space Y, in terms of its competitiveness against Space X.

- By accurately predicting the successful landings of the first stage of rockets, one can determine the most effective approach for estimating the overall cost of launches and find best place for launch sites.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from:
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - To enhance the collected data, a landing outcome label was generated by summarizing and analyzing the features of the outcome data.
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

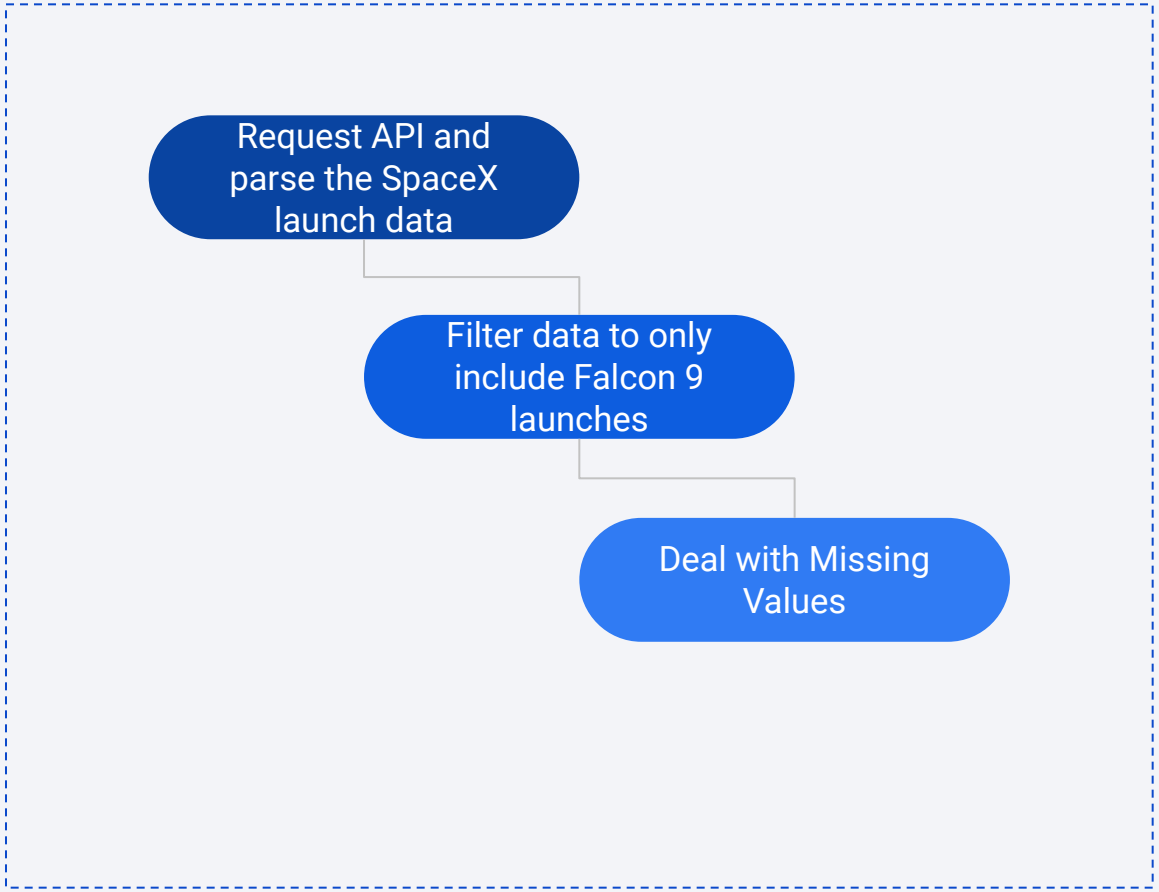
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches),

Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained and then used;
- This API was used according to the flowchart beside and then data is persisted

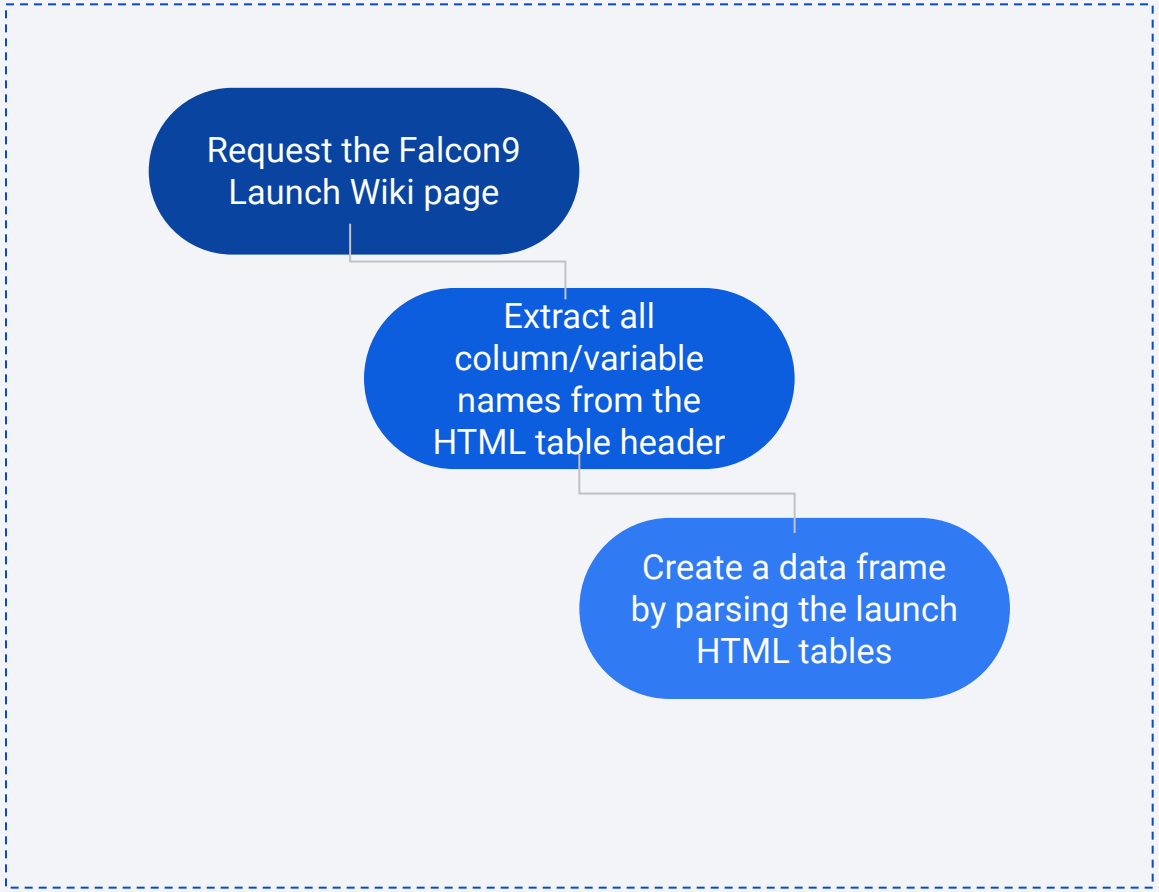


https://github.com/rick-russ/applied-data-science-capstone/blob/main/Data%20Collection%20API.ipynb

Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia

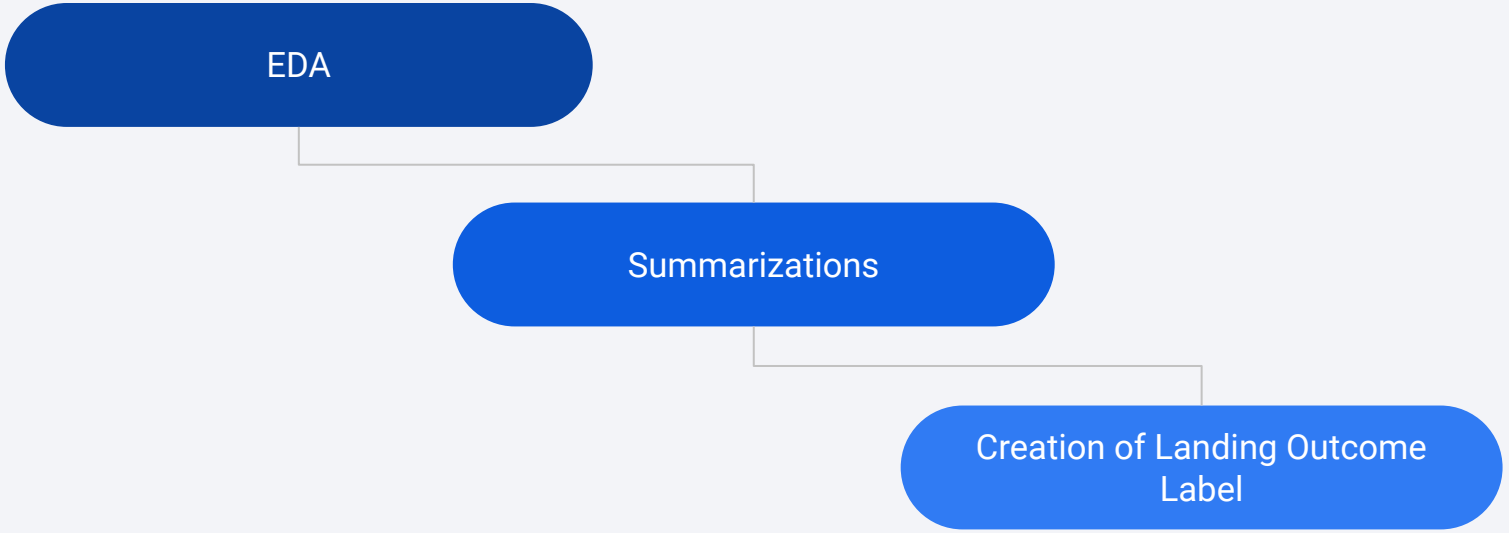
- Data are downloaded from Wikipedia according to the flowchart and then persisted.



https://github.com/rick-russ/applied-data-science-capstone/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

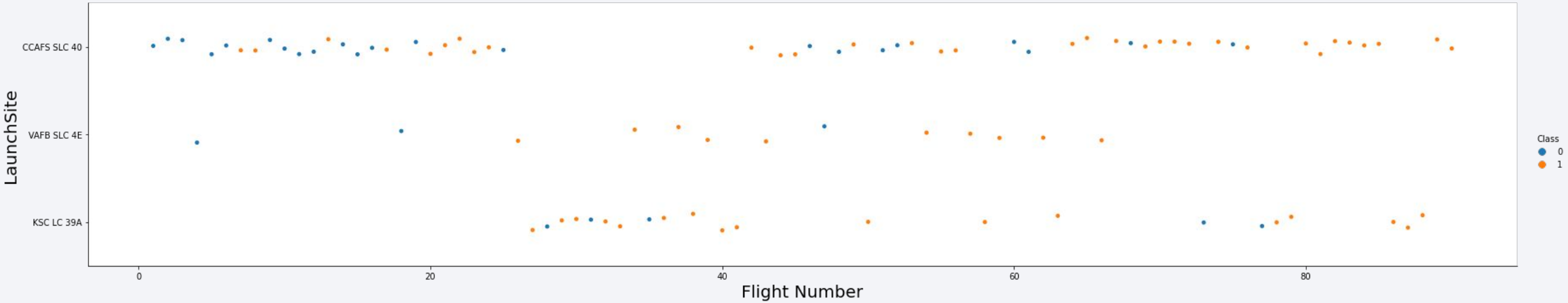
- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.



https://github.com/rick-russ/applied-data-science-capstone/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- Scatterplots and barplots were used to visualize the relationship between pair of features:
  Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



https://github.com/rick-russ/applied-data-science-capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA';
- Total payload 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 outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

https://github.com/rick-russ/applied-data-science-capstone/blob/main/EDA.ipynb

Build an Interactive Map with Folium

Markers, circles, lines, were created and added to a folium map

- Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates
- Marker clusters indicates groups of events in each coordinate
- Lines are used to indicate distances between two coordinates.

https://github.com/rick-russ/applied-data-science-capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

The following graphs and plots were used to visualize data

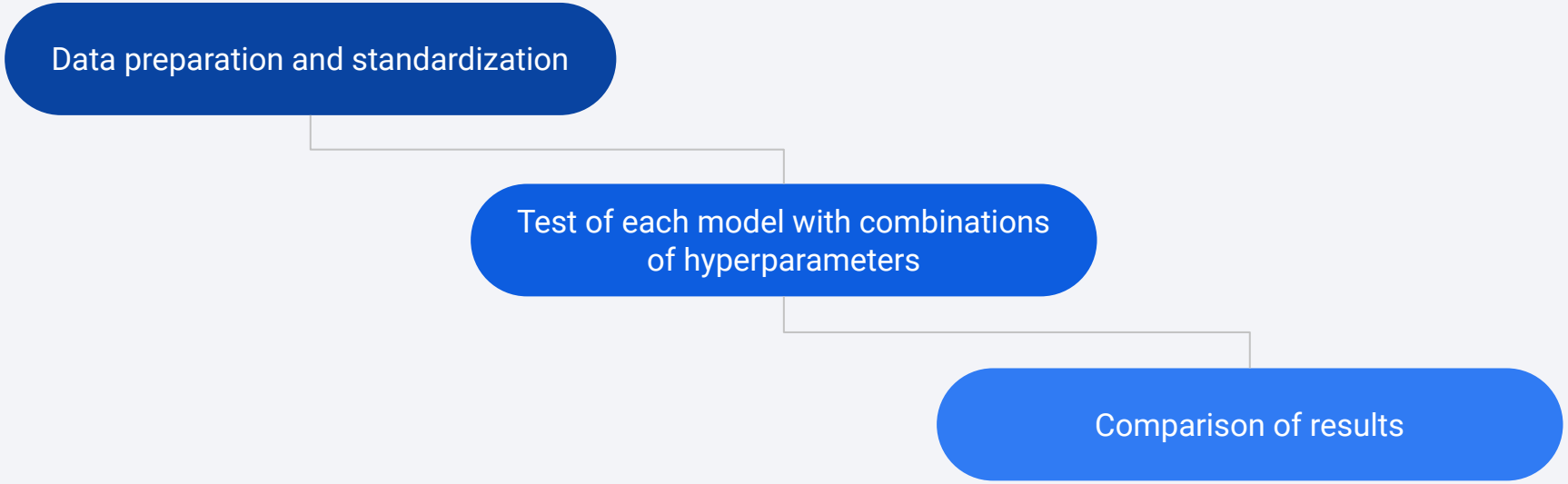
- Percentage of launches by site
- Payload range

By utilizing this combination, it became possible to rapidly examine the correlation between payloads and launch sites. This analysis greatly aided in identifying the optimal launch locations based on the characteristics of the payloads.

https://github.com/rick-russ/applied-data-science-capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

Four classification models, namely logistic regression, support vector machine, decision tree, and k nearest neighbors, were subjected to a comparative analysis.



https://github.com/rick-russ/applied-data-science-capstone/blob/main/Machine%20Learning%20Prediction.ipynb

Results

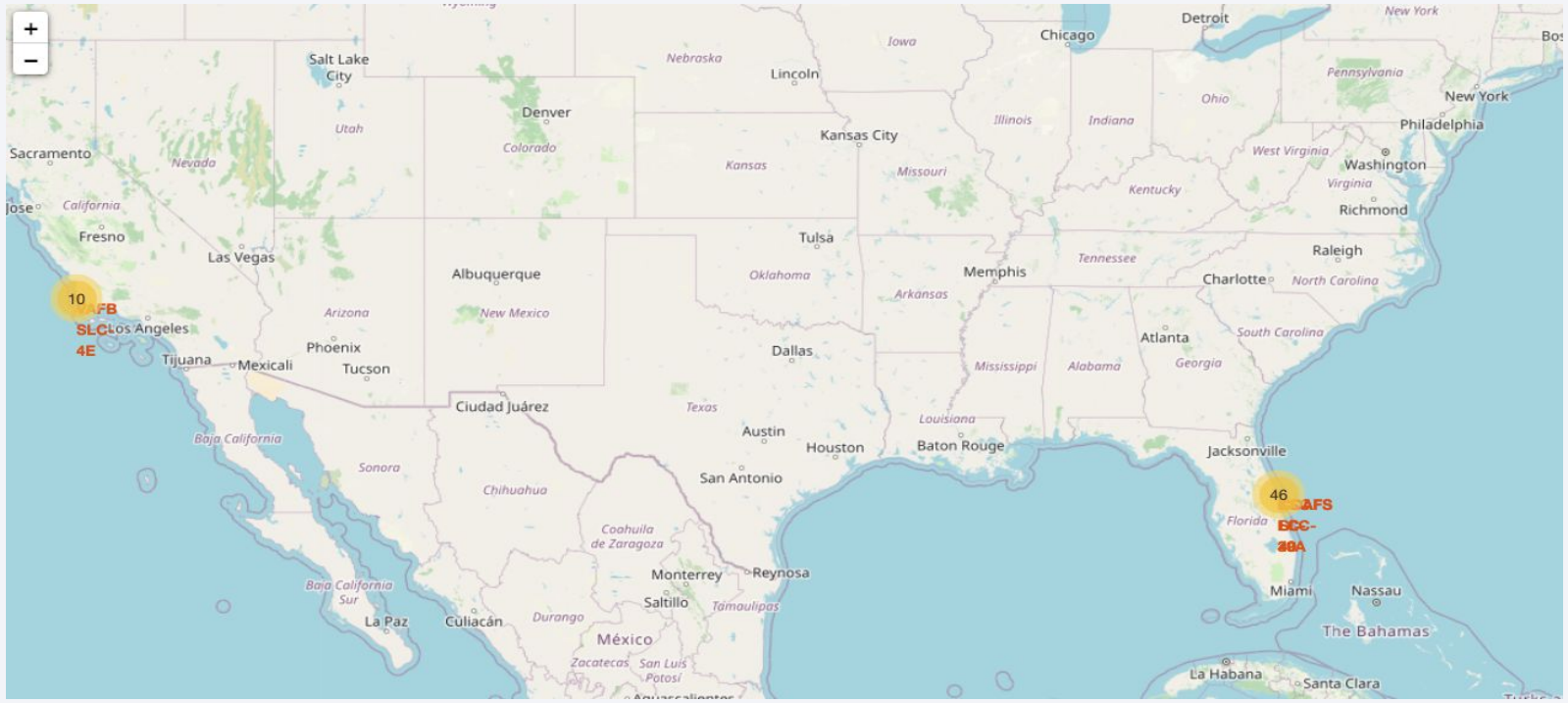
Exploratory data analysis results

- Space X uses 4 different launch sites;
- The first launches were done to Space X itself and NASA;
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The number of landing outcomes became as better as years passed.

Results

Interactive analytics

Most of the launches happens at US East coast launch sites.

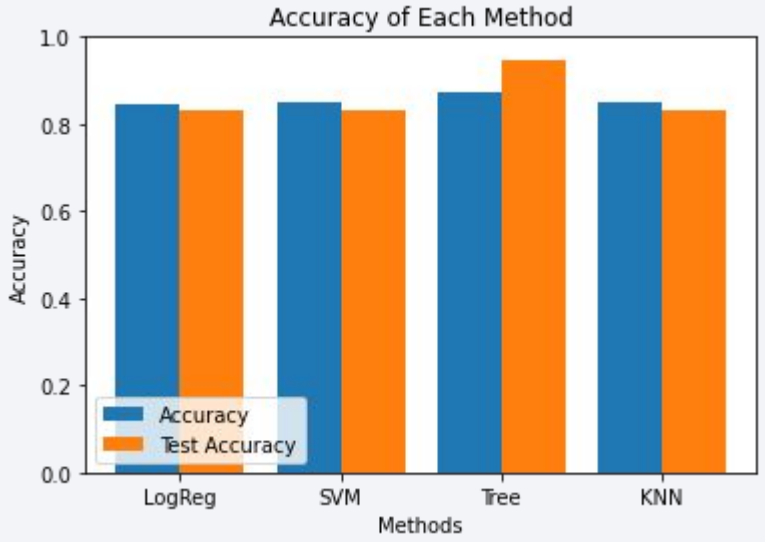


Results

Predictive analysis results

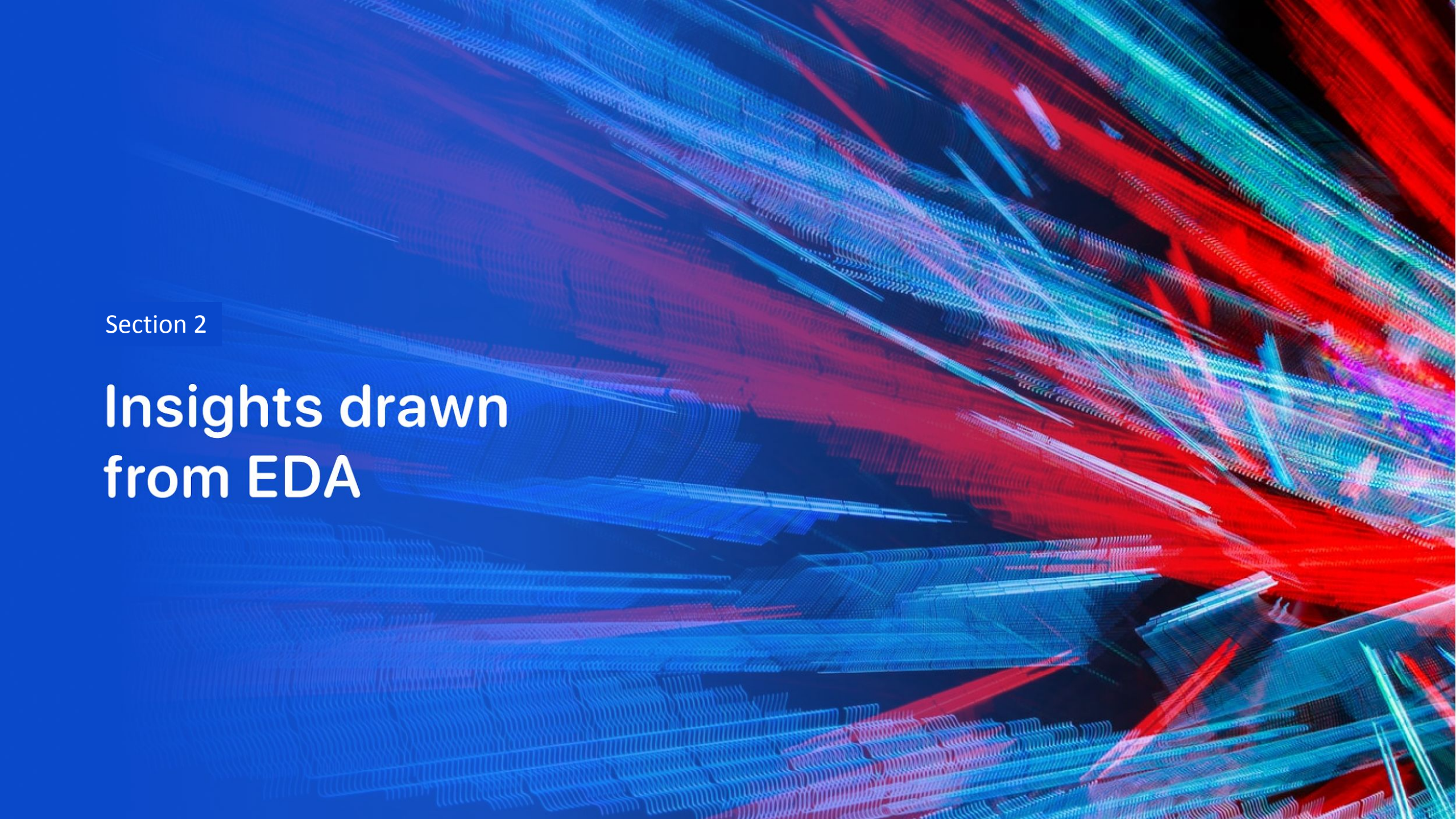
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	**0.875**	**0.94444**
KNN	0.84821	0.83333

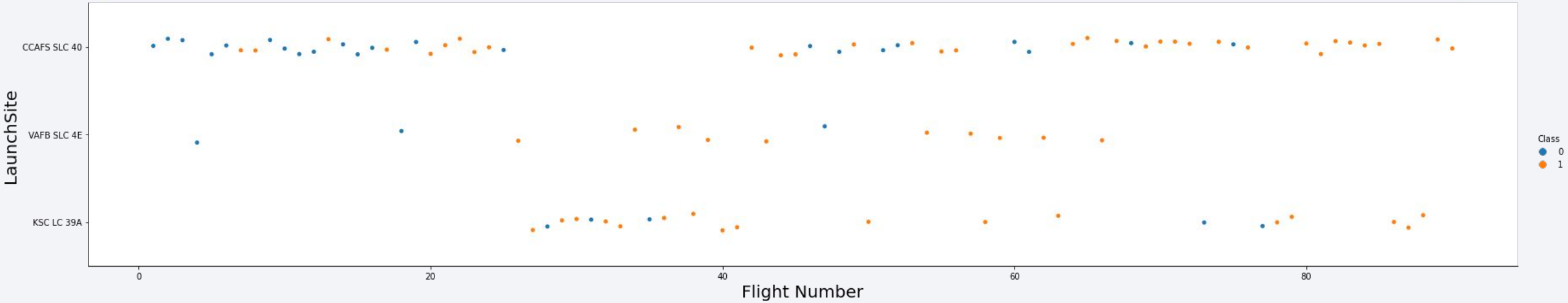


Section 2

Insights drawn from EDA

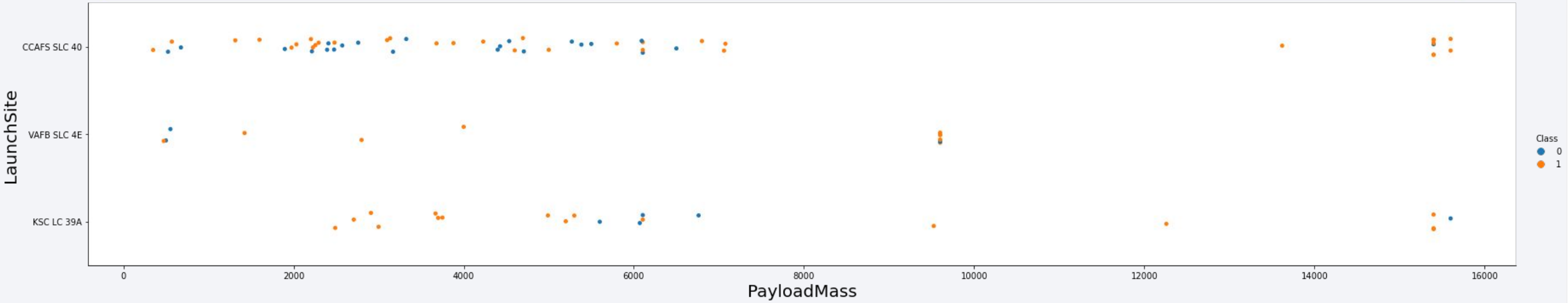


Flight Number vs. Launch Site



- Based on the plot, it is evident that the current preferred launch site is CCAF5 SLC 40, as it exhibits the highest success rate among recent launches.

Payload vs. Launch Site

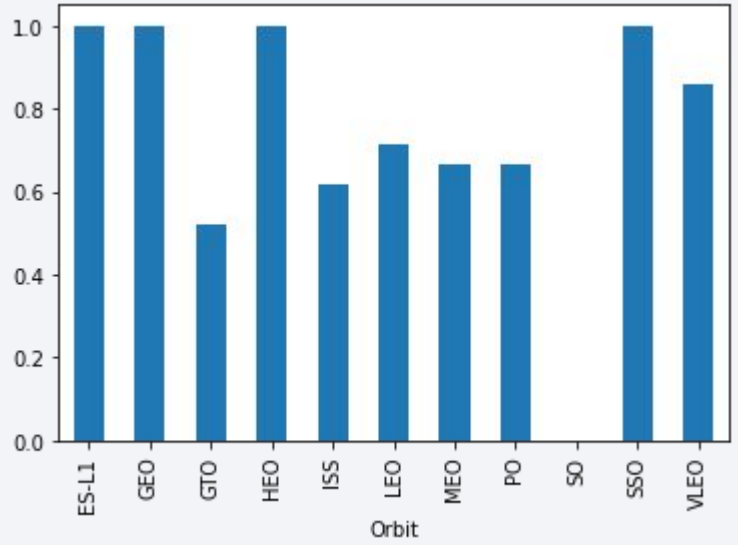


- Most launches with payload mass under 10,000 kg are from any launch site, but heavier ones happens mainly at CCAFS SLC 40 and KSC LC 39A

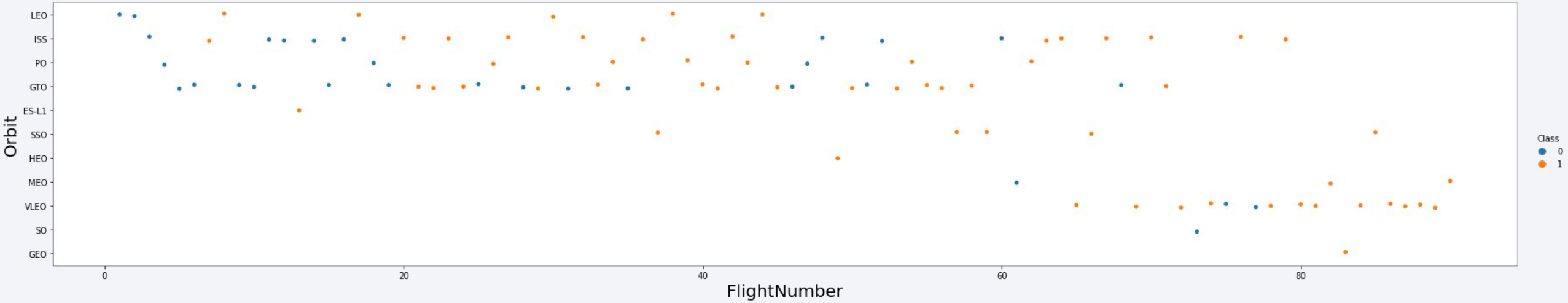
Success Rate vs. Orbit Type

The highest success rates to orbits:

- ES-L1;
- GEO;
- HEO;
- SSO.

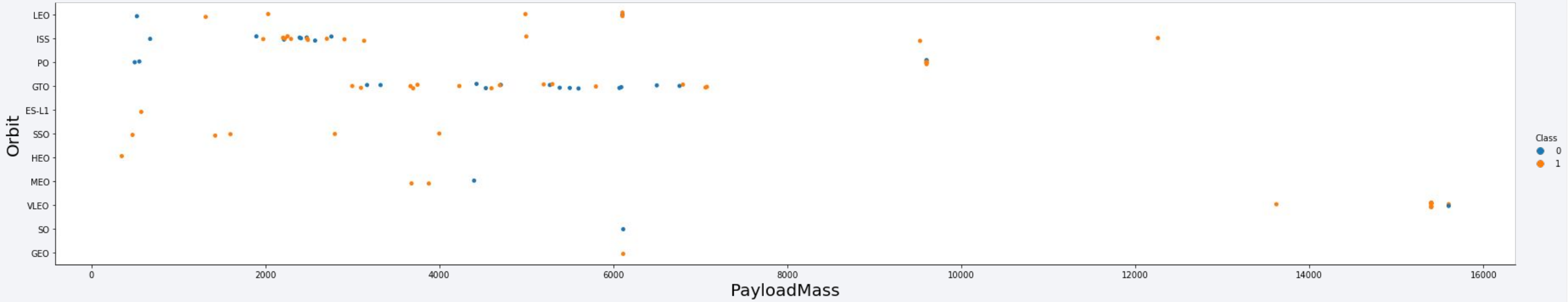


Flight Number vs. Orbit Type



- LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

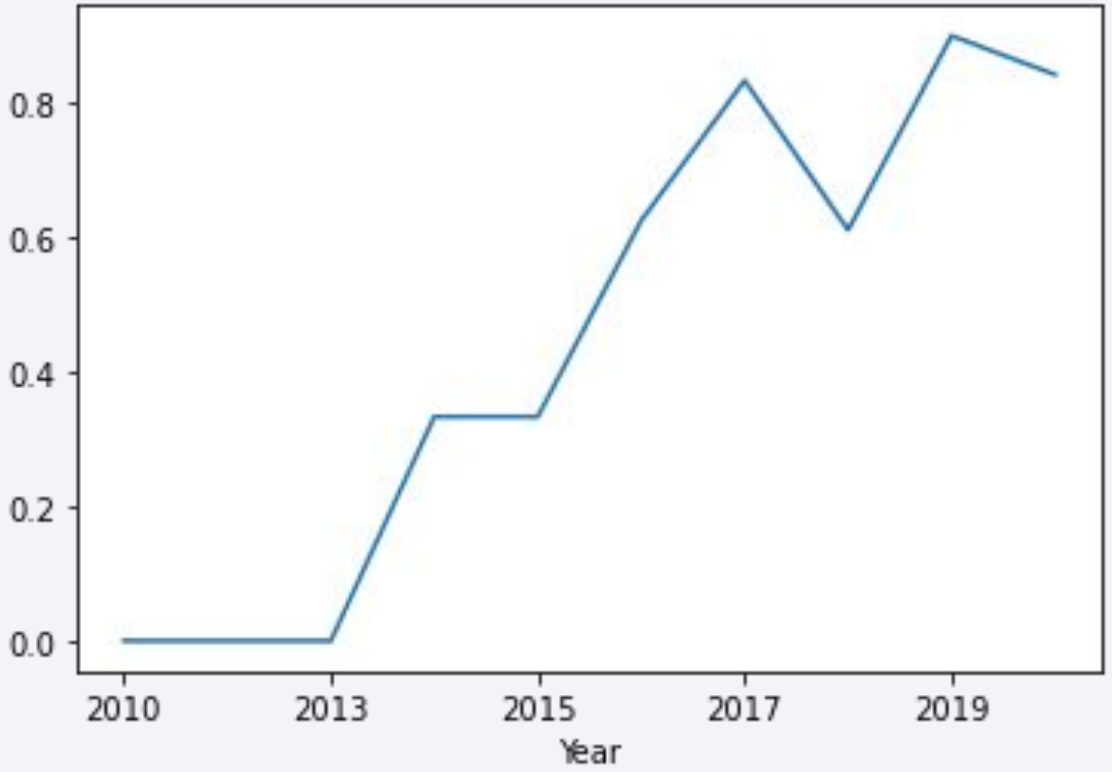
Payload vs. Orbit Type



- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2019



Launch Site Names 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

- 5 sites with launches from Cape Canaveral site.

Total Payload Mass

Total payload mass carried by boosters launched by NASA (CRS)

total_payload
111268

Average Payload Mass by F9 v1.1

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

avg_payload
2928

First Successful Ground Landing Date

- The first successful landing outcome in ground pad was acheived on 22nd of December in 2015.

first_success_gp
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List of 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 B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	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 of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

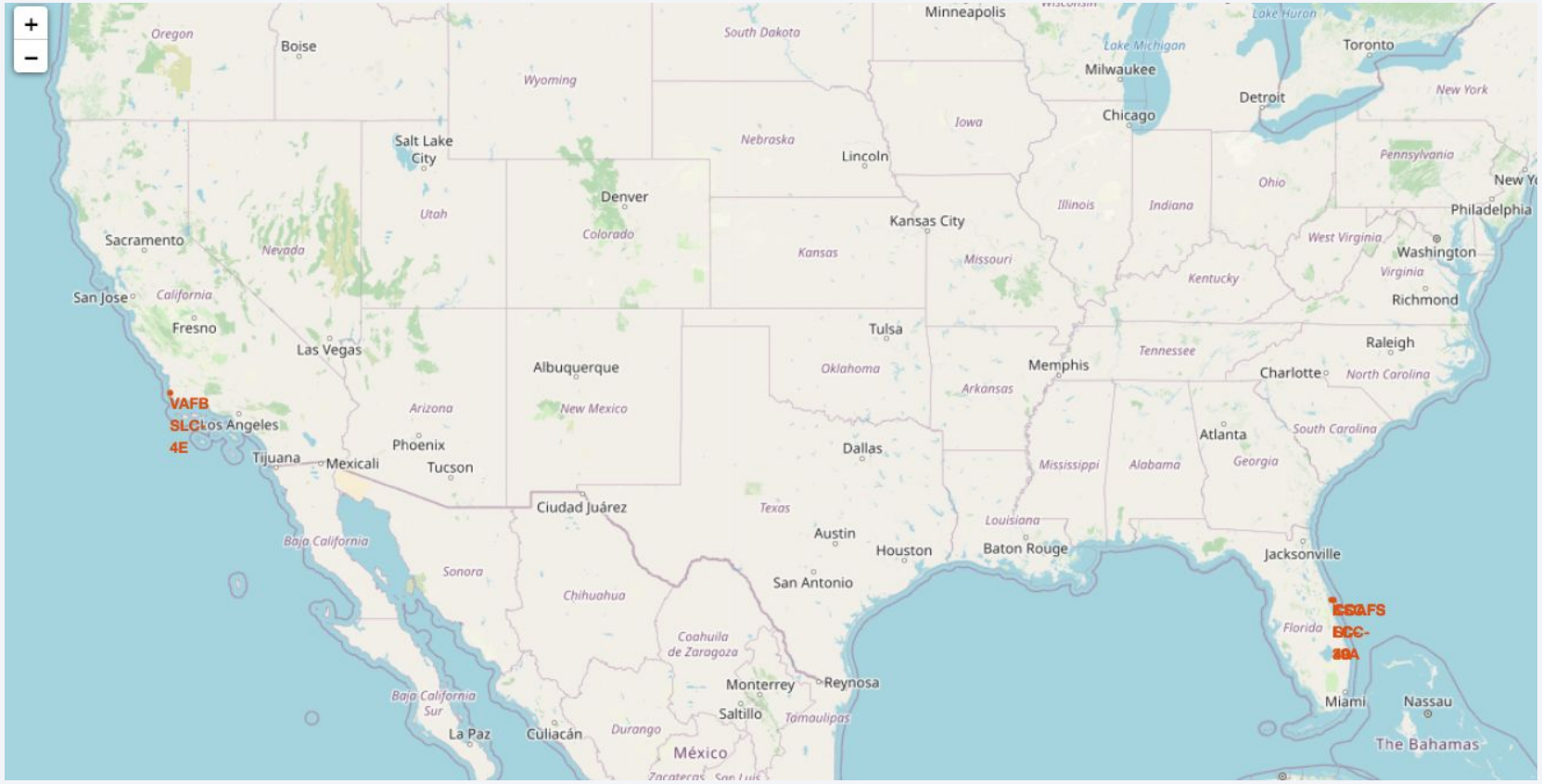
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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	qty
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

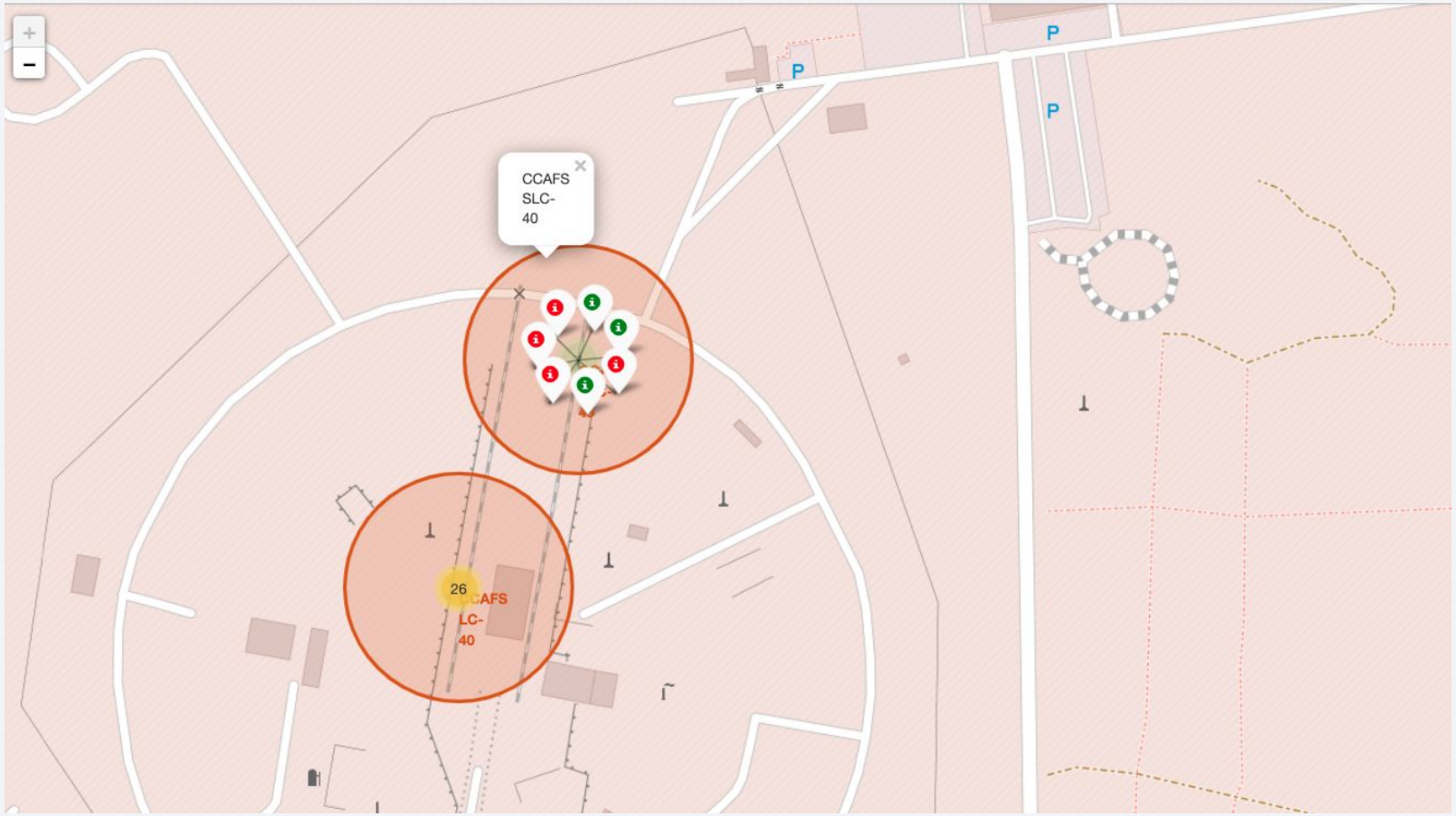

Section 3
Launch Sites
Proximities Analysis

Launch sites



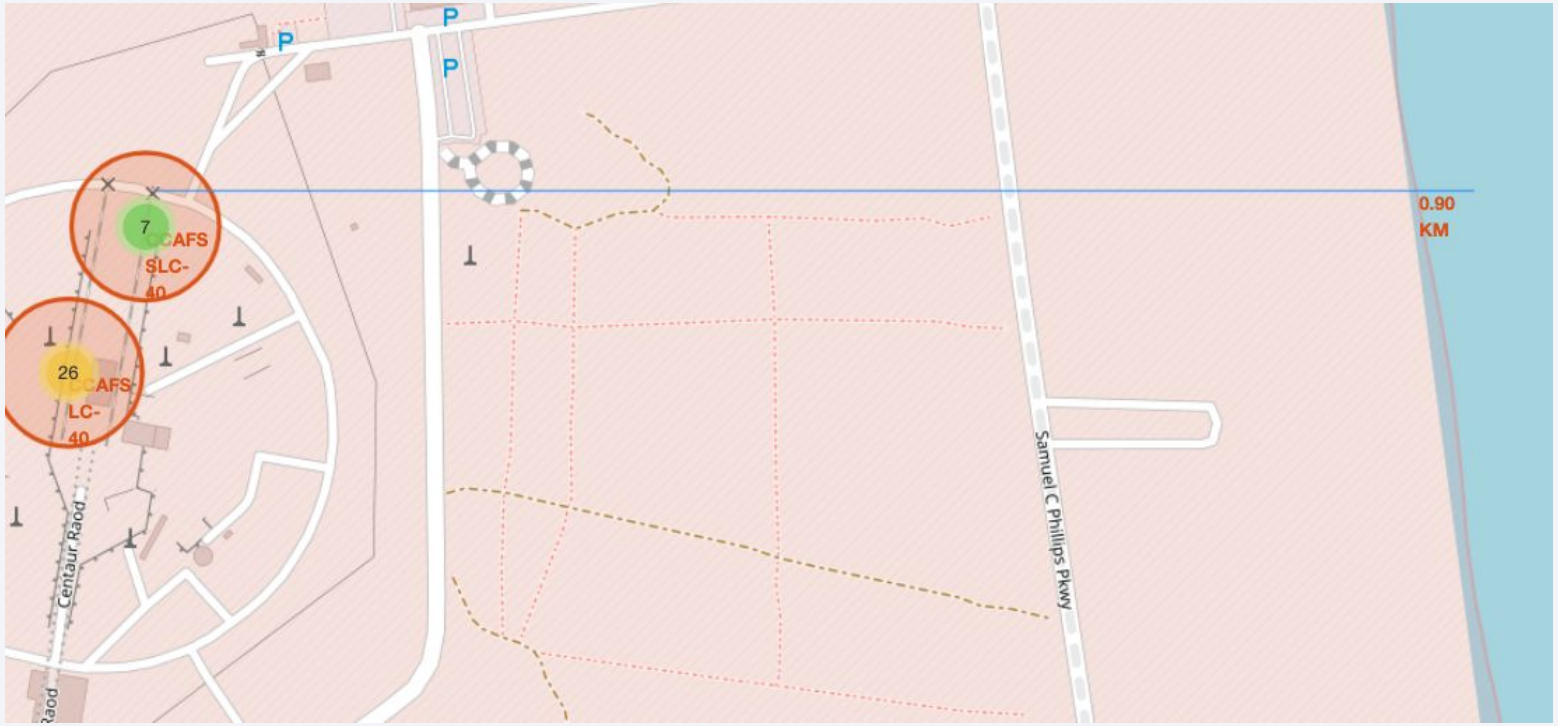
- Launch site proximity to Equator line and very close proximity to the coast

Success rate

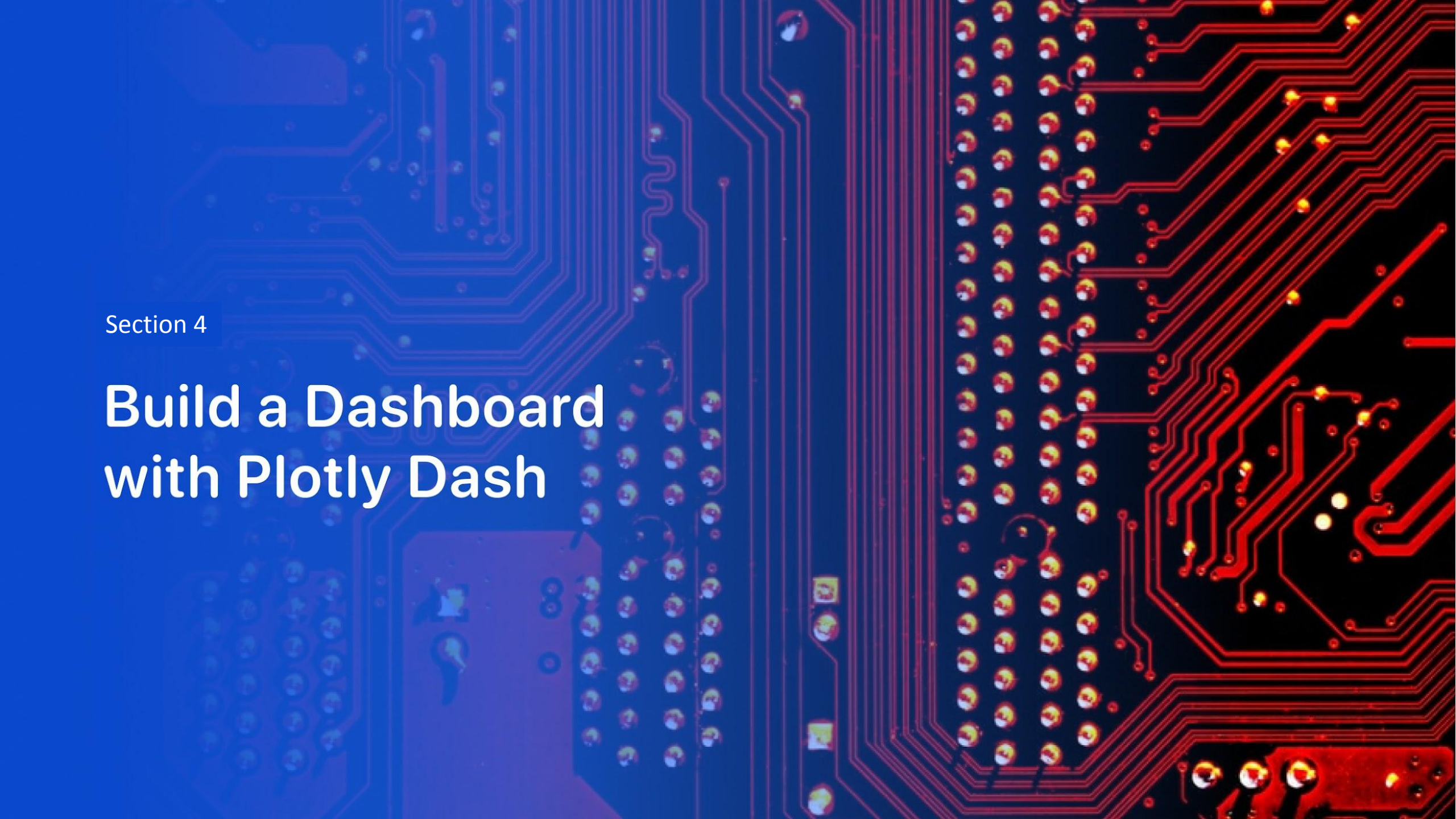


- From the color-labeled markers in marker clusters, it is able to identify which launch sites have relatively high success rates.

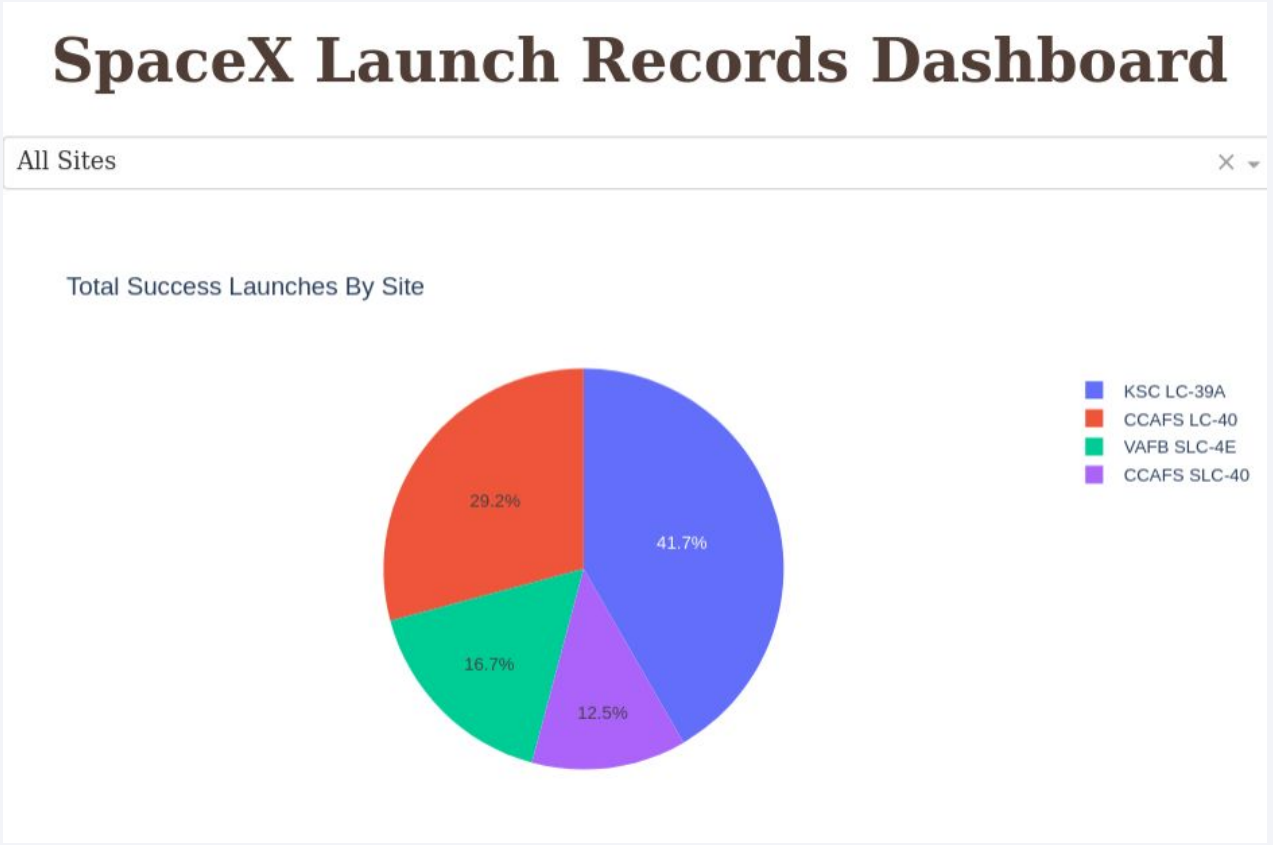
Proximity



- Launch sites are close to coastline for safety reasons.

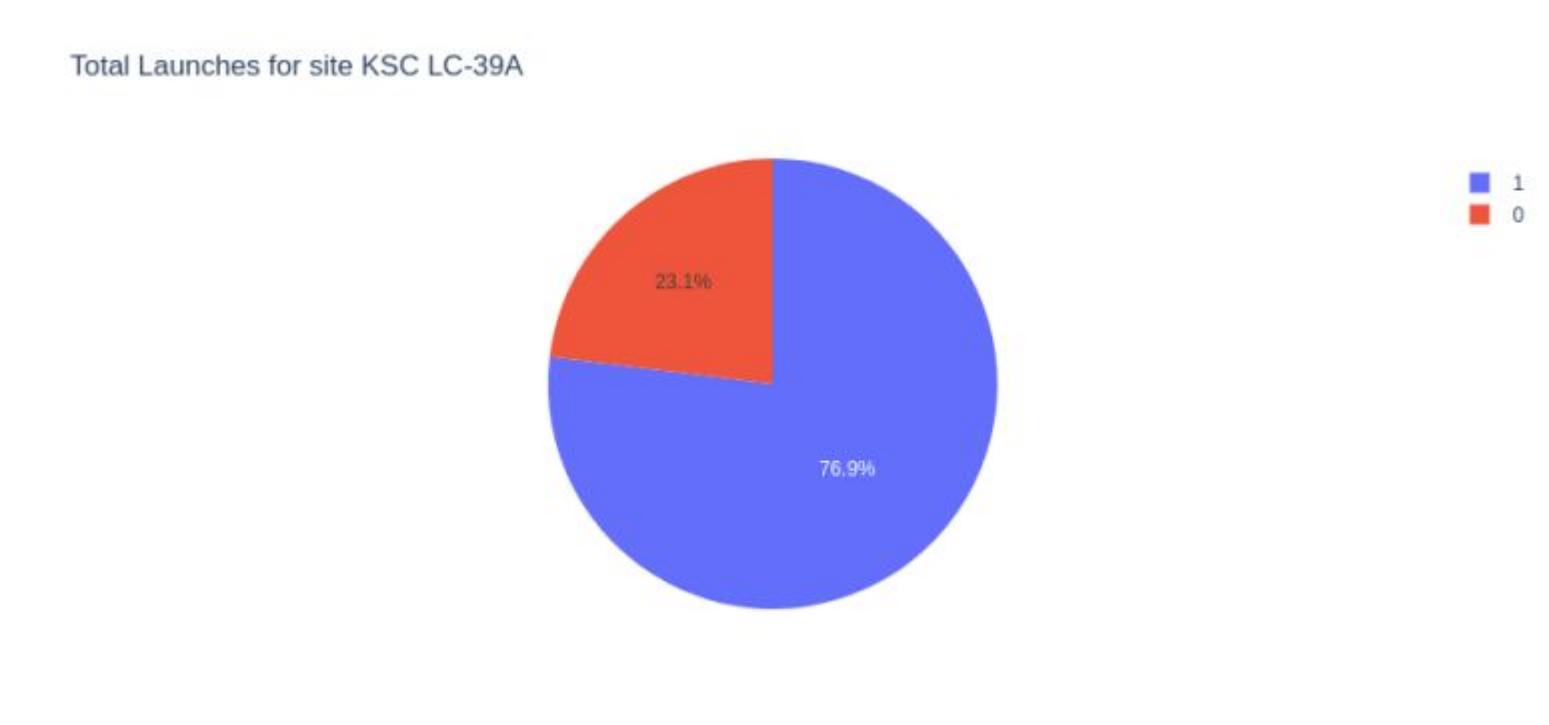

Section 4
Build a Dashboard with Plotly Dash

SpaceX successful launches



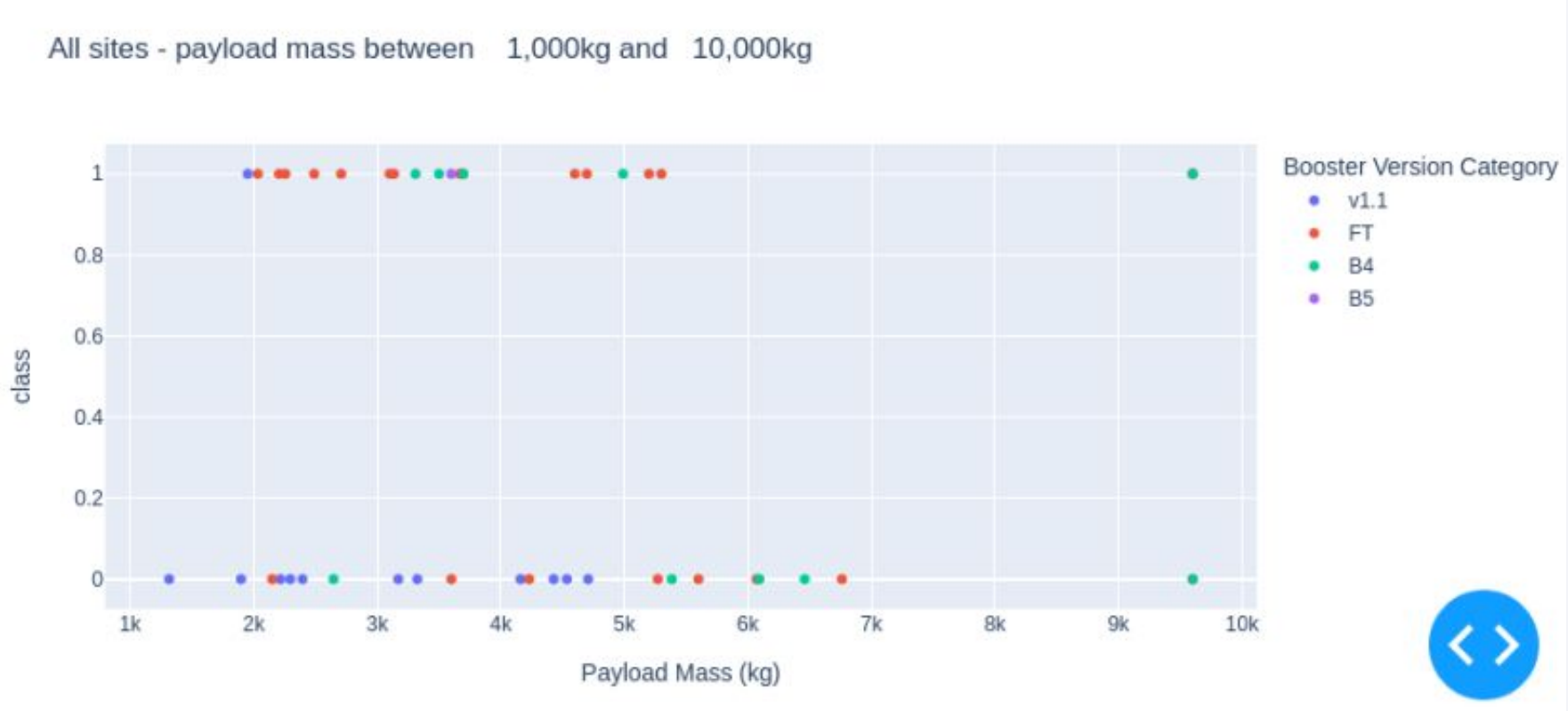
- KSC LC-39A and CCAFS LC-40 launch sites has highest total success rate

Success ratio for KSC LC-39A launch site



- Success ratio are 76.9%

Payload and launch outcome

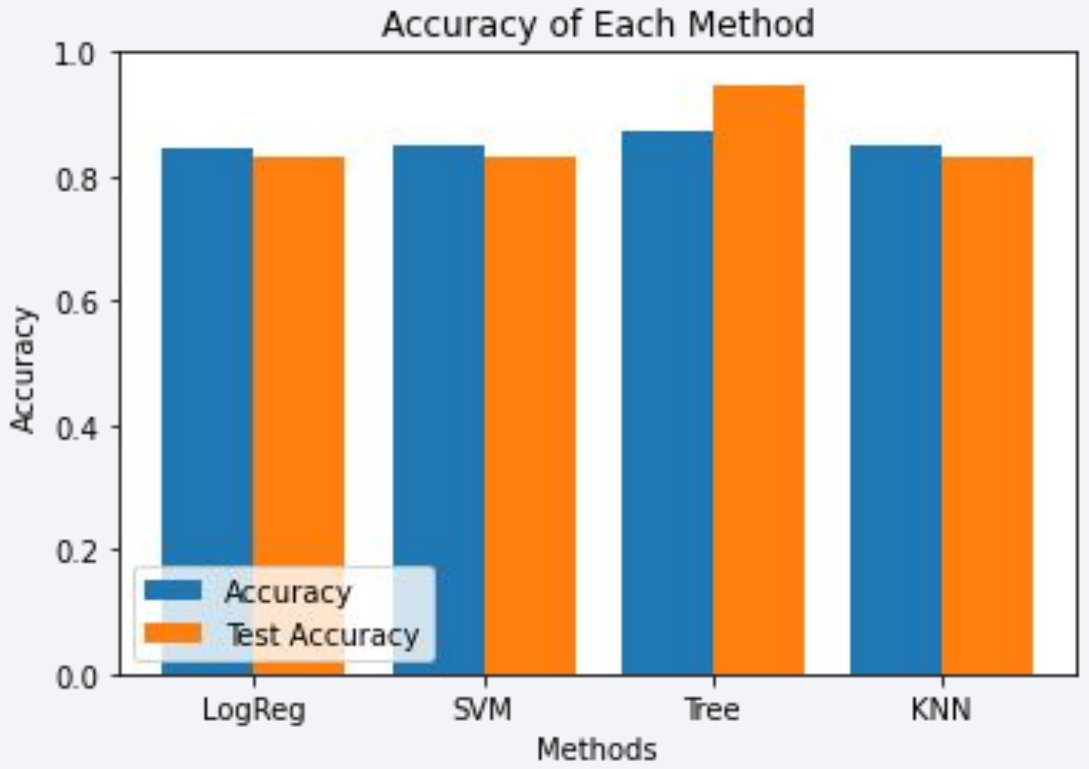


Section 5

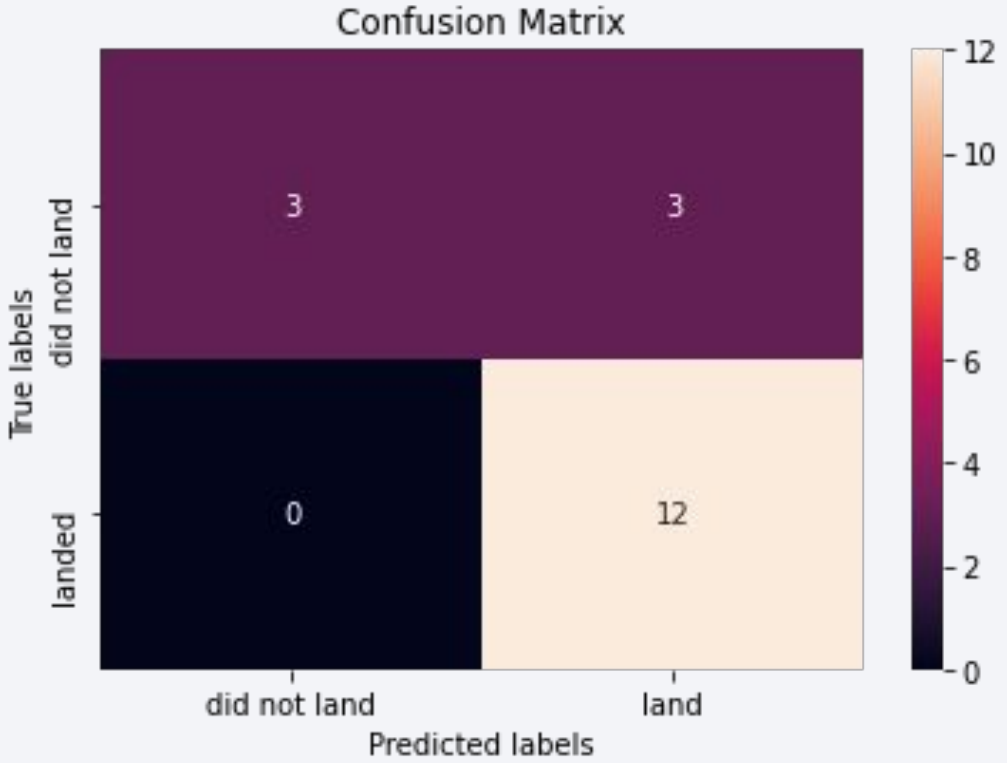
Predictive Analysis (Classification)

Classification Accuracy

- The highest classification accuracy has Decision Tree Classifier with 87.5% accuracy



Confusion Matrix



A confusion matrix of a decision tree classifier is a table that summarizes the performance of the classifier by showing the counts of true positive, true negative, false positive, and false negative predictions.

Conclusions

- Throughout the process, various data sources were analyzed, contributing to the refinement of conclusions.
- KSC LC-39A is the optimal launch site
- As processes and rockets evolve, there appears to be an improvement in successful landing outcomes over time, despite the overall high rate of successful mission outcomes

Thank you!