IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yuhe Zhang
06/21/2023

Outline

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

Executive Summary

- Summary of Methodologies
  - Data Collection: SpaceX API, Webscraping from SpaceX Wikipedia
  - Data Wrangling: Missing Values Replaced by Mean Values for Payload Mass
  - Exploratory Data Analysis:
    - Analyzing Landing Outcomes by Orbit Types, Payload Mass, Booster Versions with SQL
    - Visually Analyzing Landing Outcomes with Charts by Orbit Types, Payload Mass, Launch Sites, Booster Versions
  - Interactive Plotly Dashboard: Analysis by Launch Sites, Payload Mass, Booster Versions
  - Predictive Analysis with Machine Learning: Logistic Regression, SVM, Decision Tree, KNN
- Summary of All Results
  - Launch Success Rate Increases Over Time
  - Higher Success Rate for Higher Orbits and Lower Payload Mass
  - Higher Success Rate for Kennedy Space Center and Recent Launches at Cape Canaveral
  - Low Success Rate for Booster Versions V1.0, V1.1 and High for FT, B4, B5

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars
  - Other providers cost upward of 165 million dollars each
  - Much of the savings is because SpaceX can reuse the first stage
- Problems we want to find answers
  - If we can determine if the first stage will land, we can determine the cost of a launch
  - This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

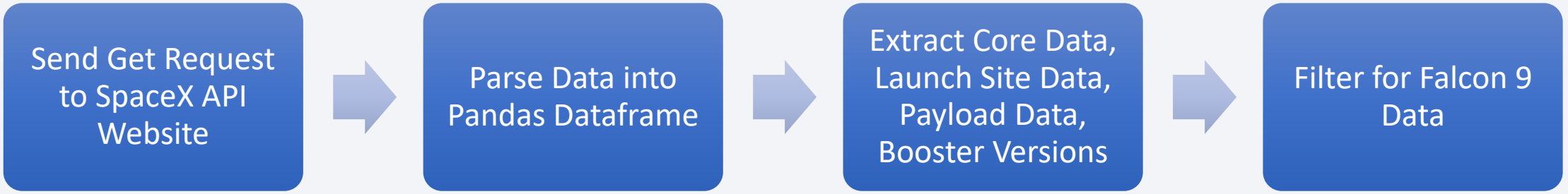
- Data collection methodology
  - SpaceX API and Webscraping from SpaceX Wikipedia
- Perform data wrangling
  - Missing Values Replaced by Mean Values for Payload Mass
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Analyzing Landing Outcomes by Orbit Types, Payload Mass, Booster Versions with SQL
  - Visually Analyzing Landing Outcomes with Charts by Orbit Types, Payload Mass, Launch Sites, Booster Versions
- Perform interactive visual analytics using Folium and Plotly Dash
  - Visual Analysis by Launch Sites with Folium Maps
  - Visual Analysis by Launch Sites, Payload Mass, Booster Versions with Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, SVM, Decision Tree, KNN
  - Parameter Tuning with GridSearch

Data Collection

- SpaceX API
  - RESTful Interface
  - Get Core Data
  - Get Launch Site Data
  - Get Booster Version
  - Get Payload Data
- Webscraping from SpaceX Wikipedia Page
  - HTML Requests
  - Python BeautifulSoup Package
  - Extract column names from HTML table header

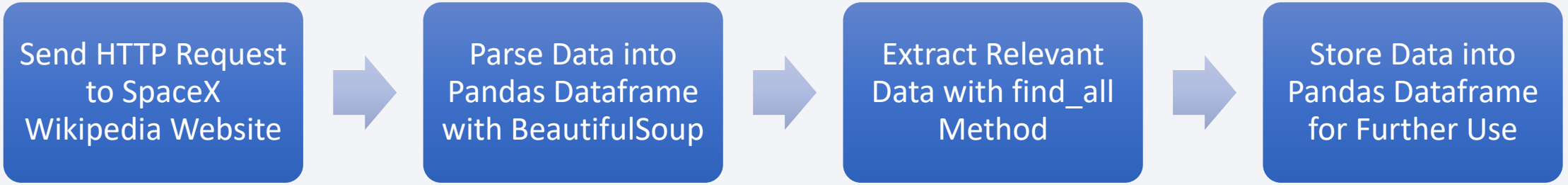
Data Collection – SpaceX API

Data Collection - API



Data Collection – Webscraping

Data Collection - Webscraping



Data Wrangling

Data Wrangling

Replace missing values for Payload Mass data with mean values

Create a landing outcome column "Class" for the dataframe

EDA with Data Visualization

EDA with Visualization

Charts

- Flight Number vs. Launch Site: Success rate of each launch site over time
- Payload Mass vs. Launch Site: Success rate of increase in payload mass over time
- Orbit Type vs. Success Rate: Success rate of each orbit type
- Flight Number vs. Orbit Type: Success rate with the development of orbit types over time
- Payload Mass vs. Orbit Types: Success rate of each orbit type with different payload mass
- Launch Success Yearly Trend: Success rate development over time

EDA with SQL

EDA with SQL

SQL Queries

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass
- List the records which will display the month names, failure landing outcomes in drone ship, booster versions, launch site for the months in year 2015
- Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order

Build an Interactive Map with Folium

Launch Site Locations Analysis with Folium

Map Objects

- Edged circles with radius = 1000: Space launch sites
- Markers: Labeling all objects
- Marker Cluster: Creating a bunch of markers around space launch sites to indicate success (green) or failure (red) of Falcon 9 first stage landing
- Lines: Measuring the distance between the launch site and the next coastline or city.

Build a Dashboard with Plotly Dash

Interactive Plotly Dashboard

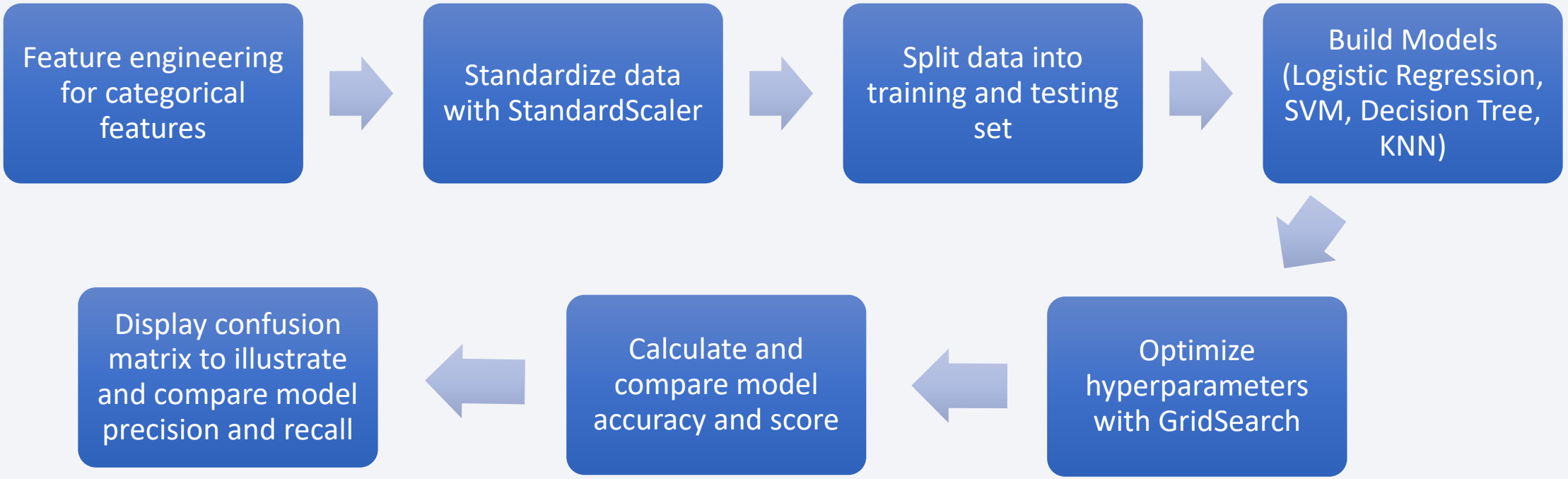
- Input Elements
  - Dropdown list for the launch sites
  - Range Slider: Selecting the payload mass
- Output Elements
  - Pie Chart: Showing the success rate of each launch site
  - Scatterplot: Showing the landing outcomes by payload mass and booster versions

Predictive Analysis (Classification)

Machine Learning Prediction

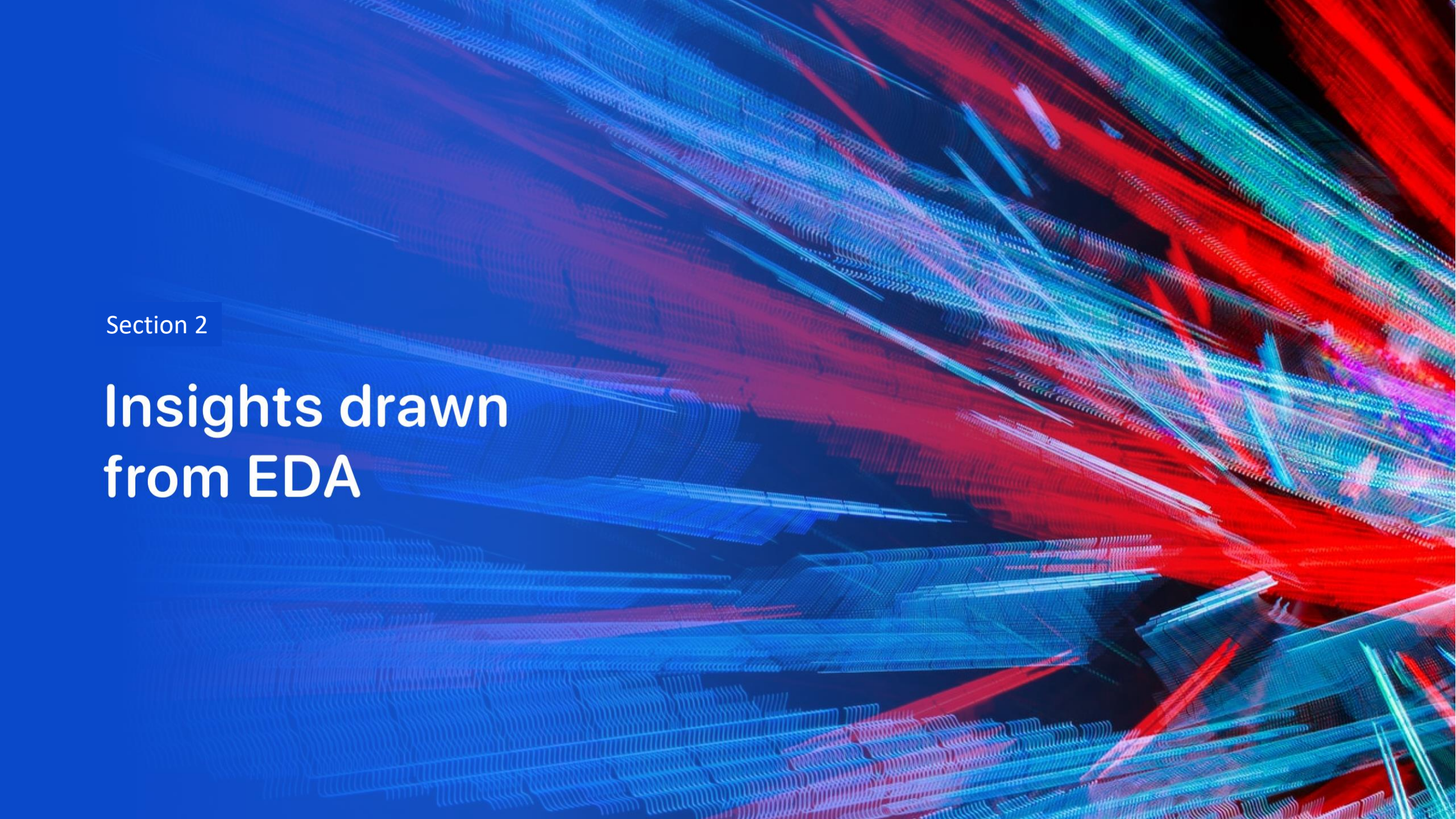
- Preprocessing
  - One-Hot-Encoding for categorical features
  - Standardize data with StandardScaler
  - Split data into independent/dependent variables and training and testing set
- Model Building
  - Logistic Regression, SVM, Decision Tree, KNN
- Optimization
  - Optimizing the models based on their hyperparameters using GridSearch
- Evaluation
  - Select the best tuned hyperparameters
  - Use Score and Accuracy to compare among models
  - Display confusion matrix to illustrate and compare the precision and recall of each model

Predictive Analysis (Classification)

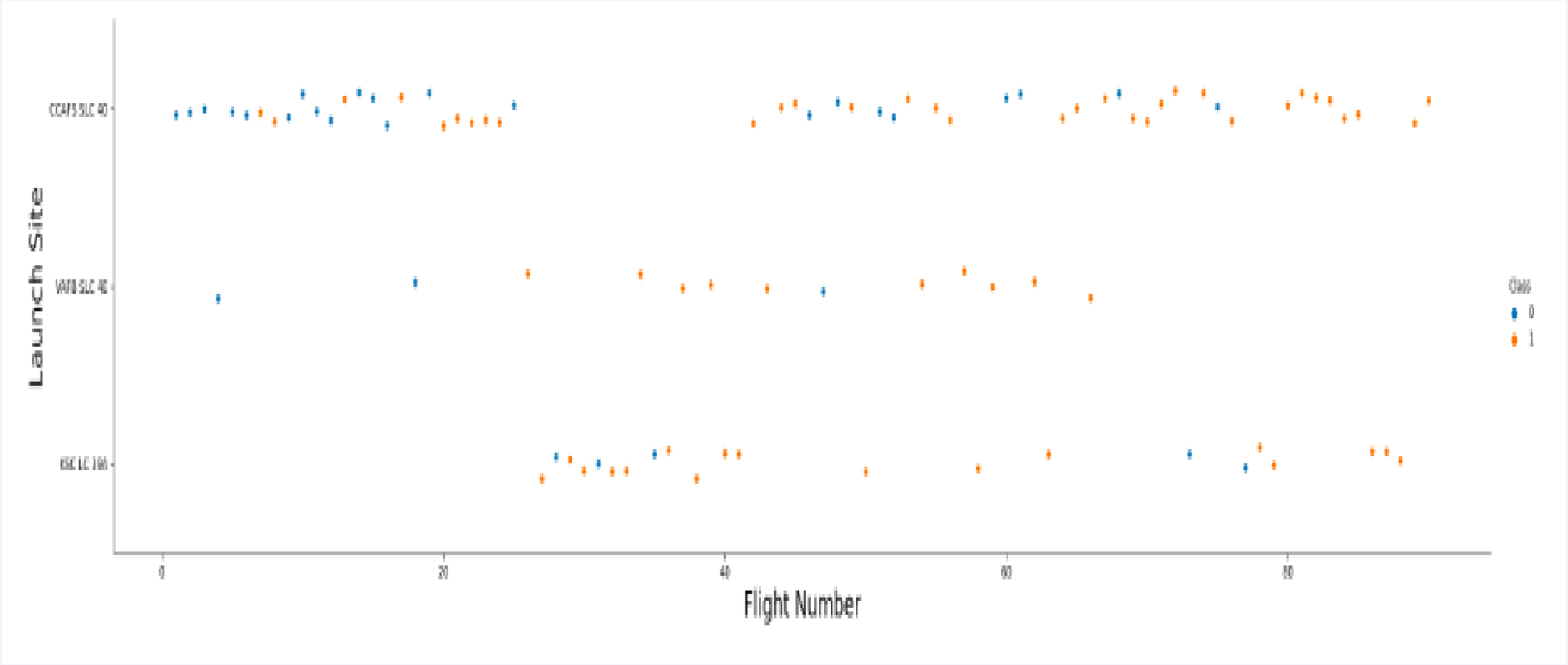


Results

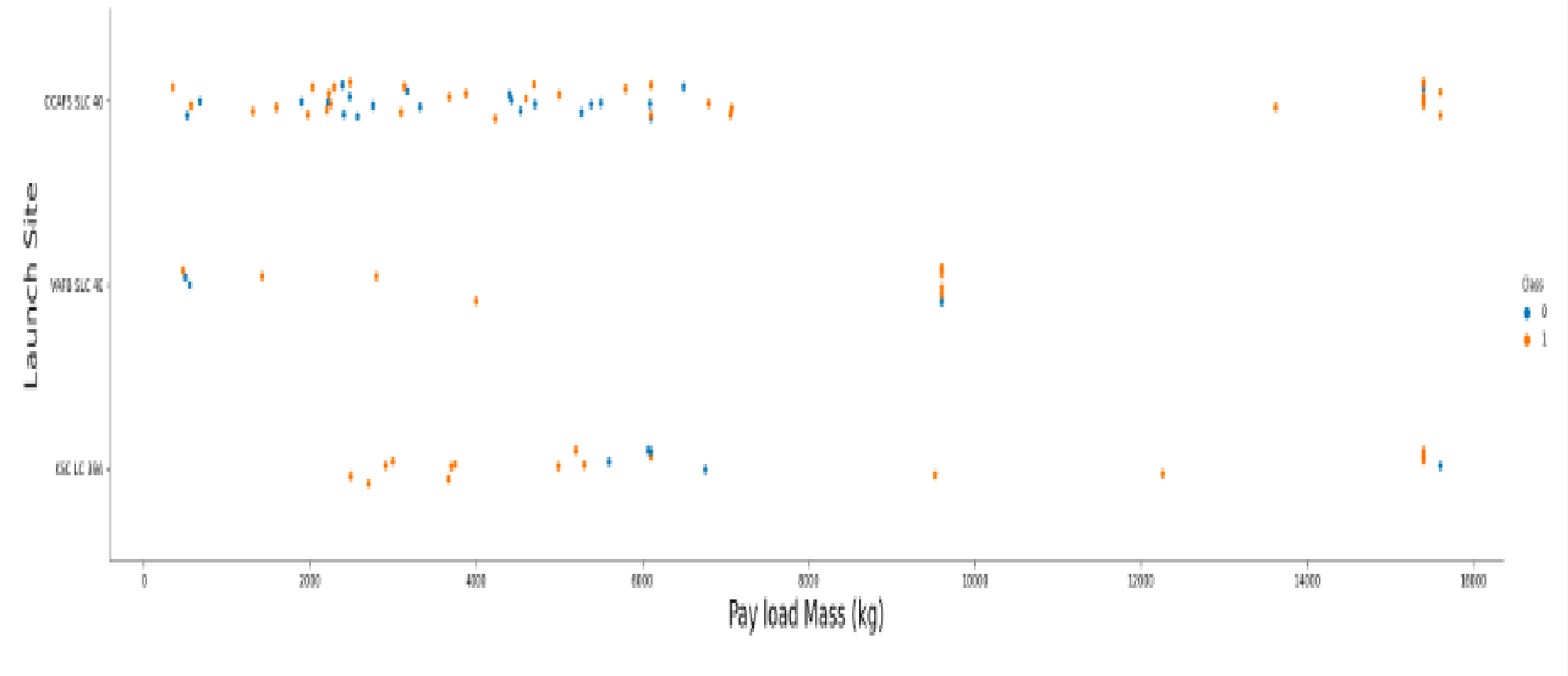
- Exploratory data analysis results
  - Launch success rate increases over time
  - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
  - Higher success rate for lower payload mass
  - Higher success rate for Kennedy Space Center and recent launches at Cape Canaveral
  - Low Success Rate for Booster Versions V1.0, V1.1 and High for FT, B4, B5
- Predictive analysis results
  - Best prediction by Decision Tree model


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

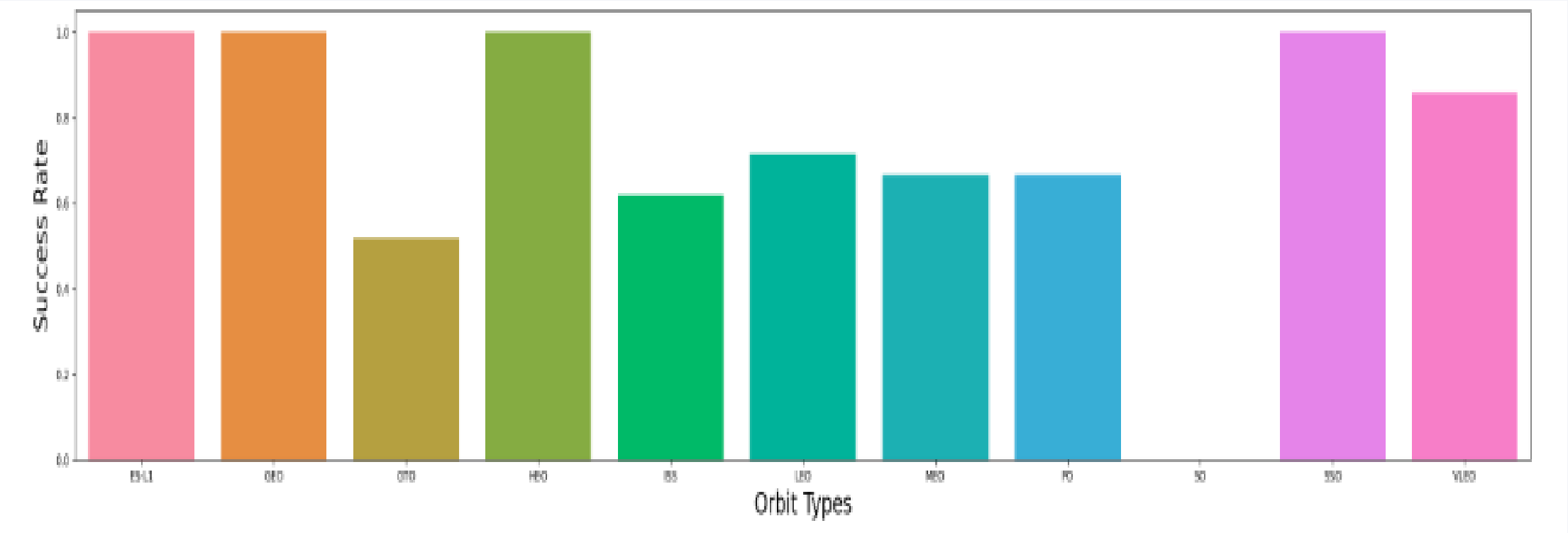


Payload vs. Launch Site

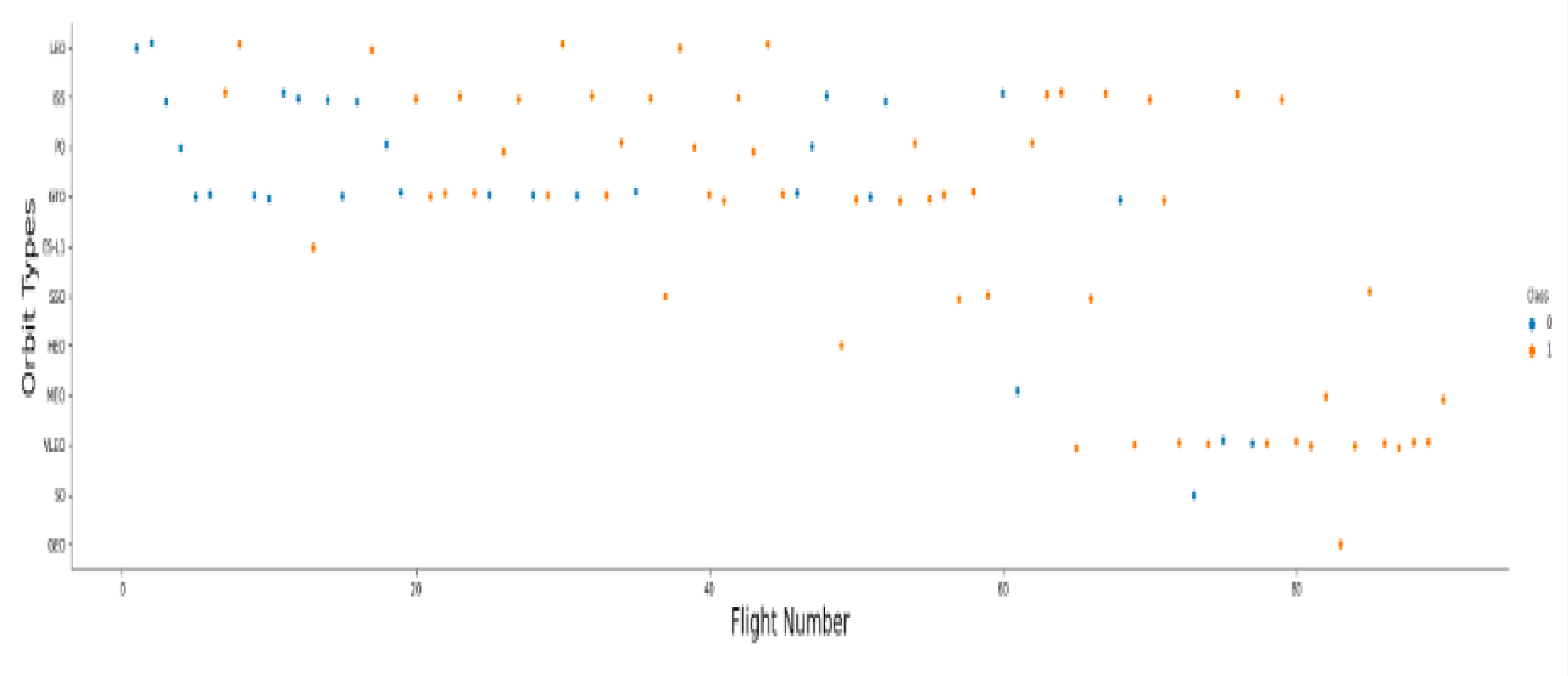


Success Rate vs. Orbit Type

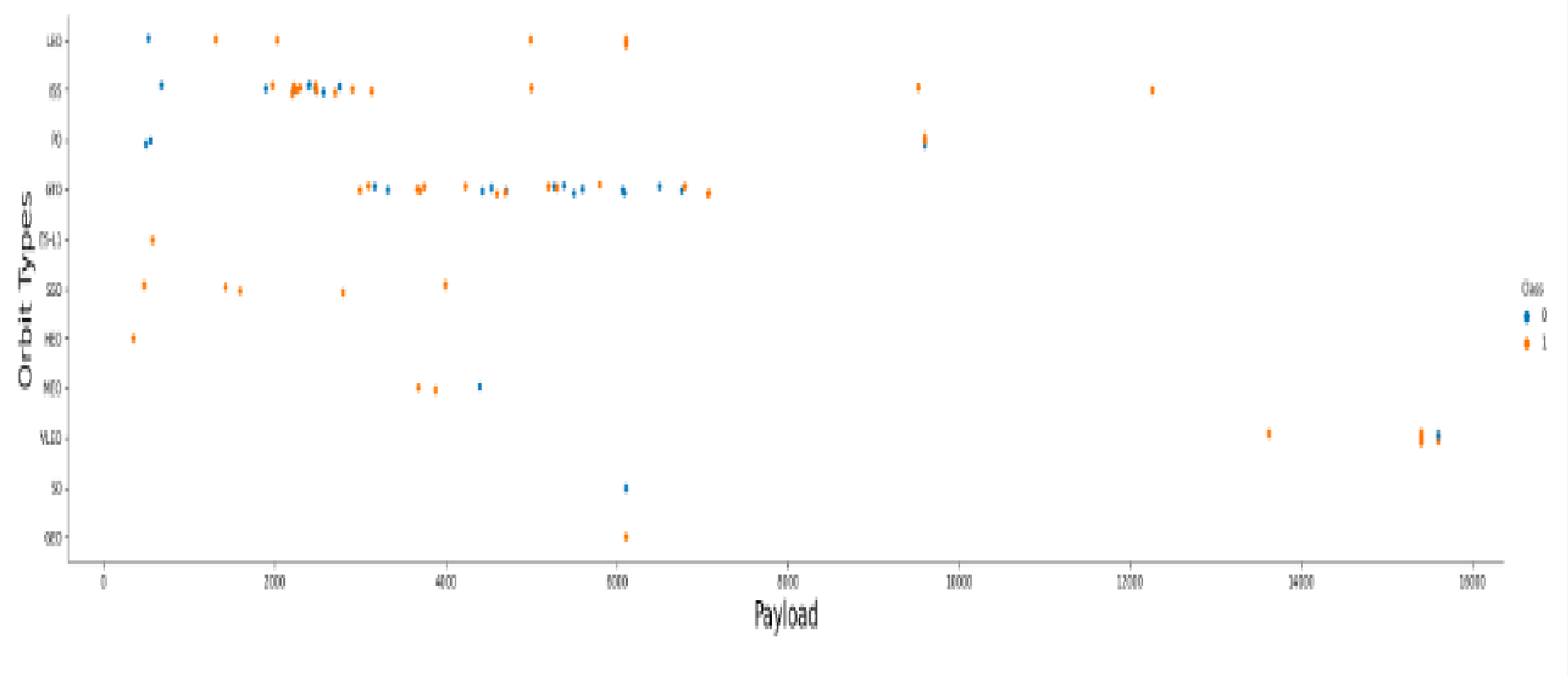
High Orbits: ES-L1, GEO, HEO, SSO
Low Orbits: GTO, ISS, LEO, MEO, PO, VLEO



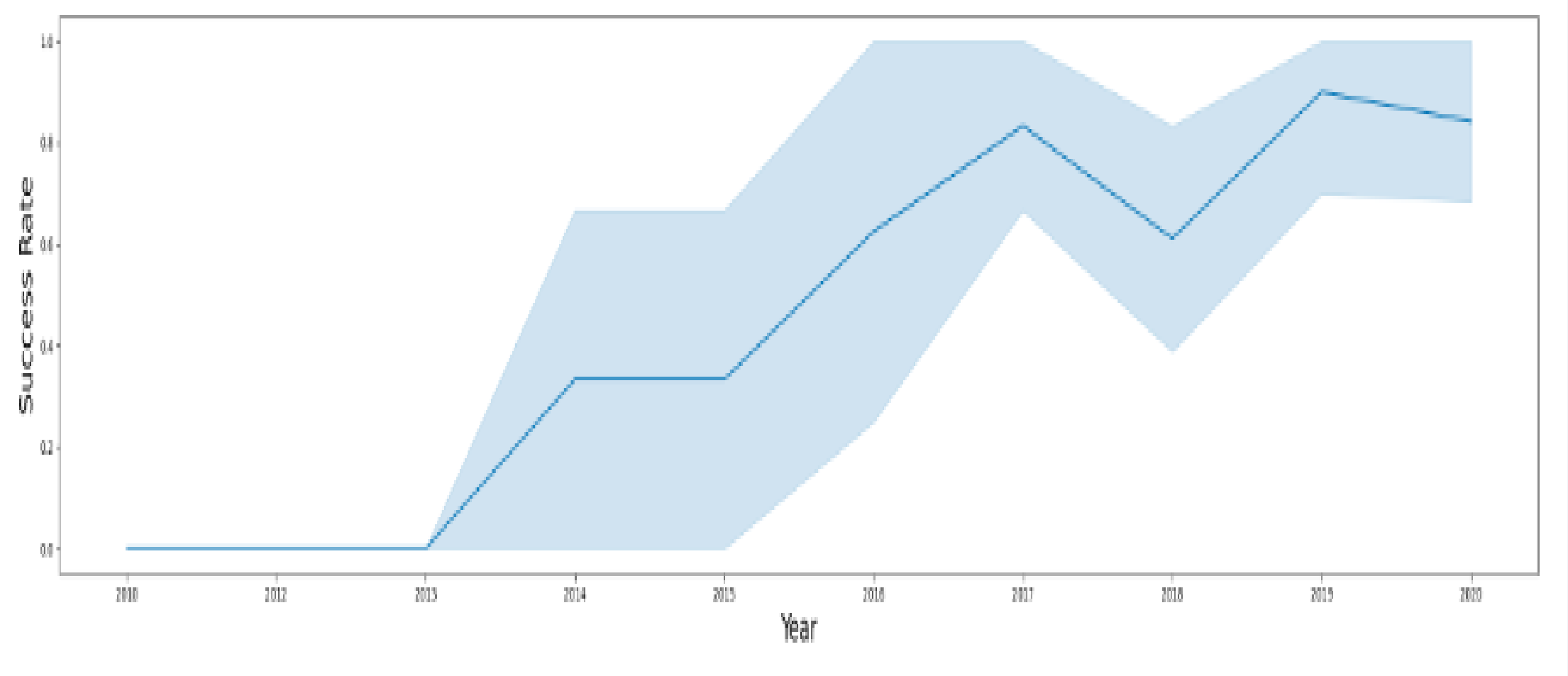
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

CCAFS: Cape Canaveral Air Force Station

VAFB: Vandenburg Air Force Base

KSC: Kennedy Space Center

Launch Site Names Begin with 'CCA'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

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

Total_Payload_Mass_KG
45596.0

Average Payload Mass by F9 V1.1

Average payload mass carried by booster version F9 V1.1

AVG_Payload_MASS_KG
2928.4

First Successful Ground Landing Date

Date of the first successful landing outcome in ground pad was achieved

MAX(Date)
22/12/2015

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

Mission_Outcome	Outcome
None	0
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

Names of the booster which have carried the maximum payload mass

2015 Launch Records

Records of the failed landing outcomes in drone ship, their booster versions, and launch site names in 2015

month	year	Booster_Version	Launch_Site	Landing_Outcome
10	2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

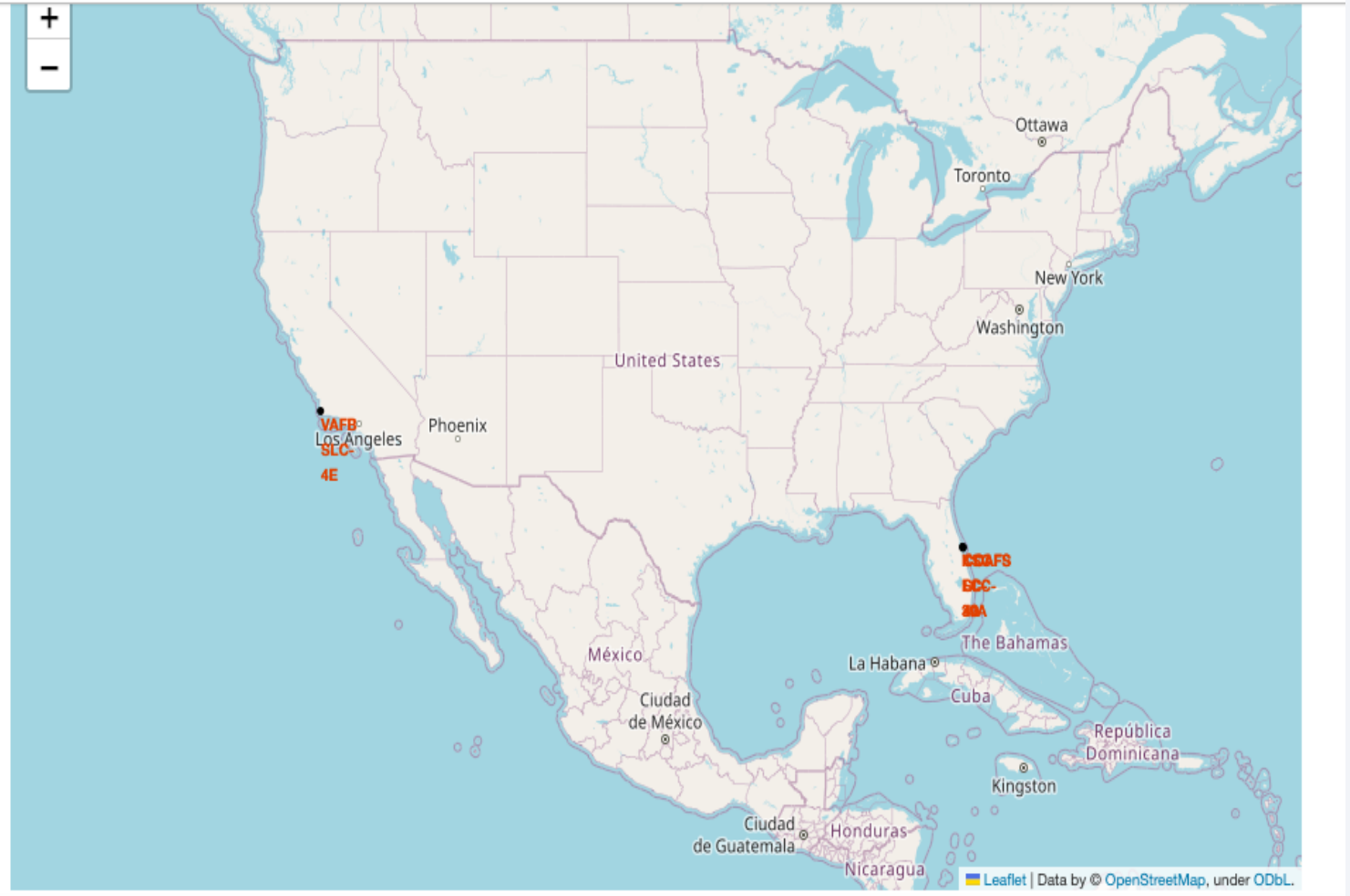
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Ranking the count of successful landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

Landing_Outcome	count_outcomes
Success	20
Success (drone ship)	8
Success (ground pad)	7


Section 3
Launch Sites
Proximities Analysis

All Launch Site Locations

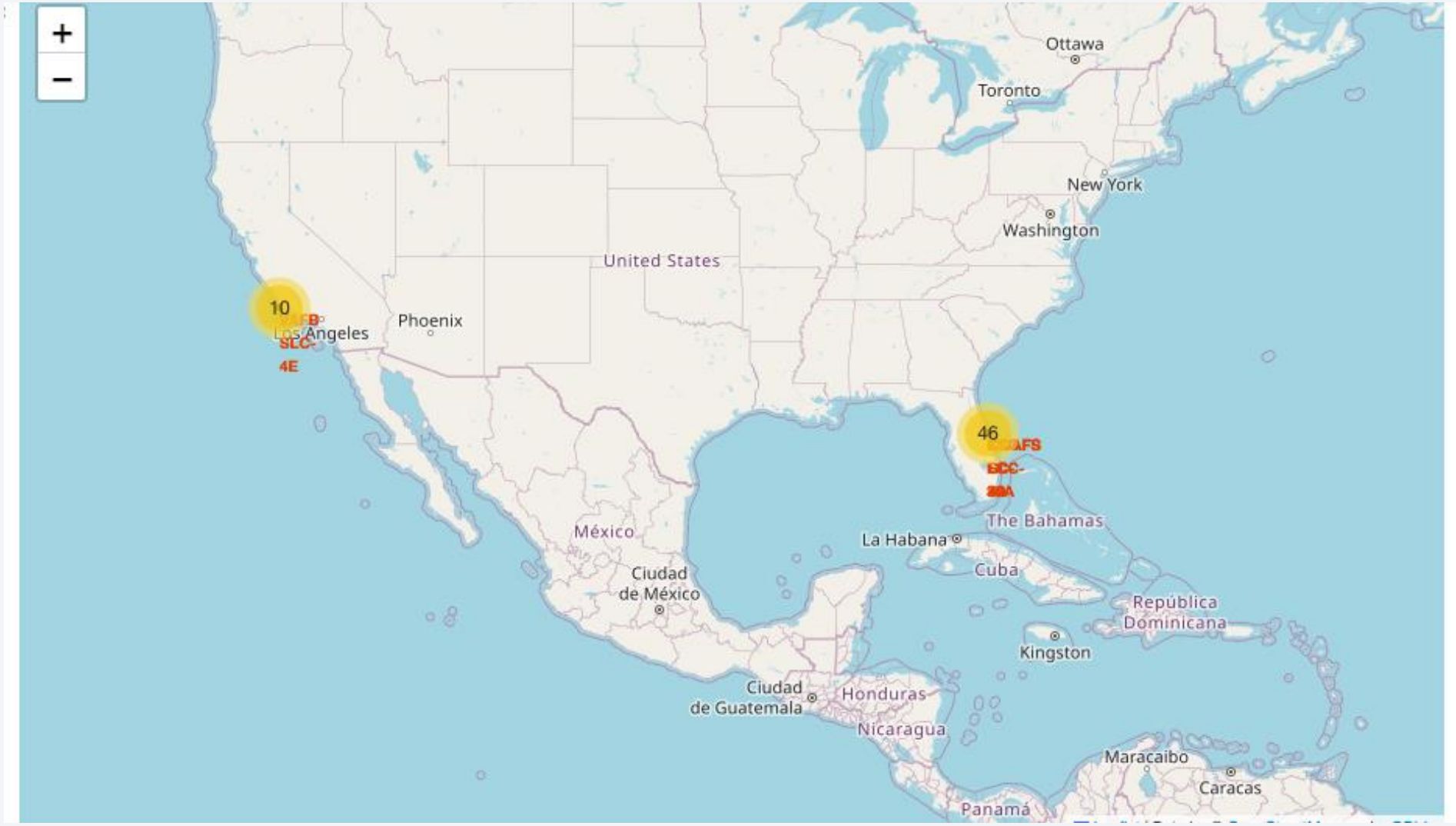


VAFB Location: California Co

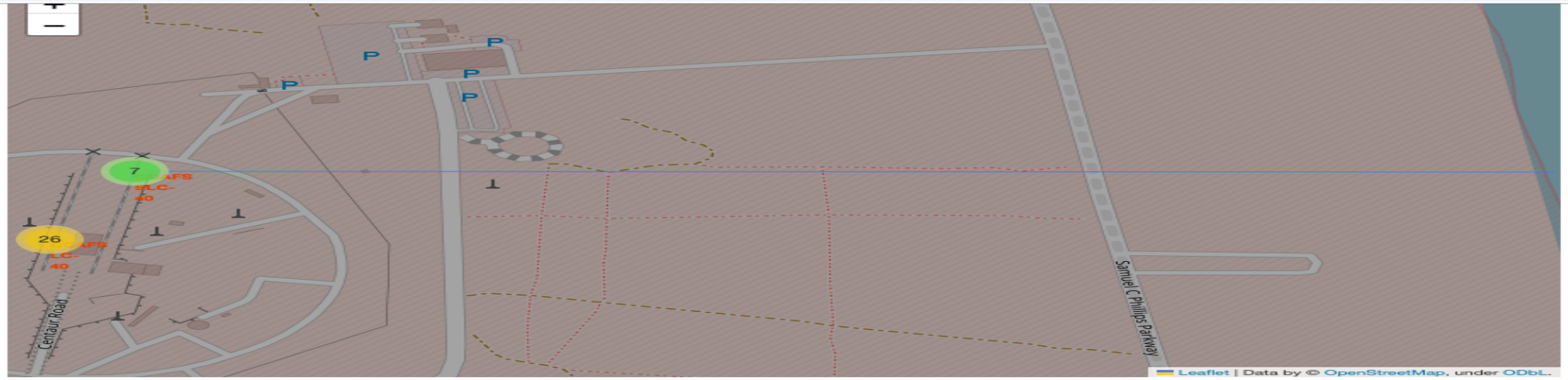
KSC Location: Florida Coast

CCAFS Location: Florida Coa

Success vs. Failed Launches For each Site on a Map



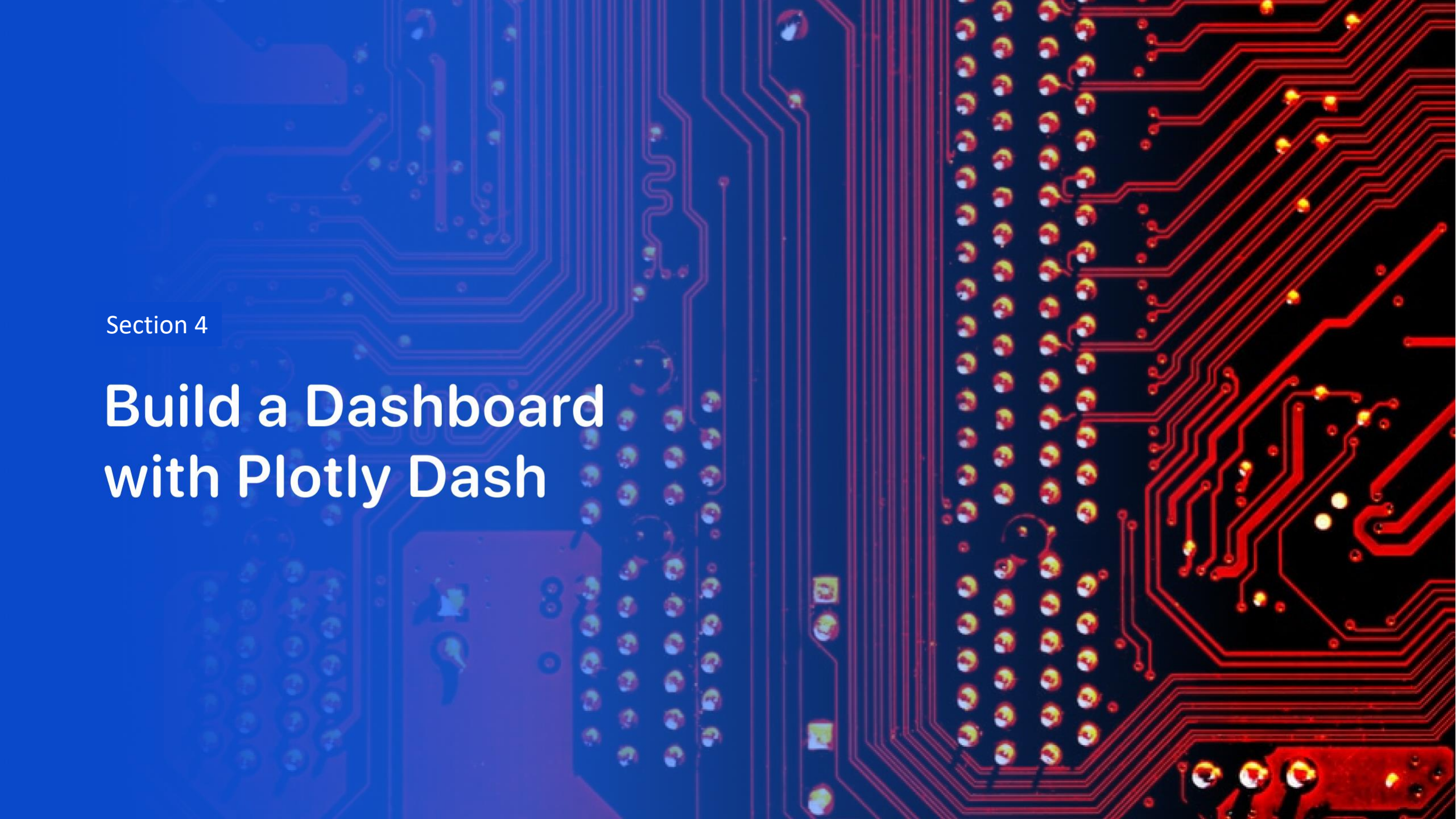
CCAFS Location Distance to Florida Coast




```
# find coordinate of the closet coastline
# e.g.,: Lat: 28.56367  Lon: -80.57163
launch_site_lat=28.563197
launch_site_lon=-80.576820
coastline_lat=28.56319
coastline_lon = -80.56785
distance_coastline = calculate_distance(launch_site_lat, launch_site_lon, coastline_lat, coastline_lon)
distance_coastline
```

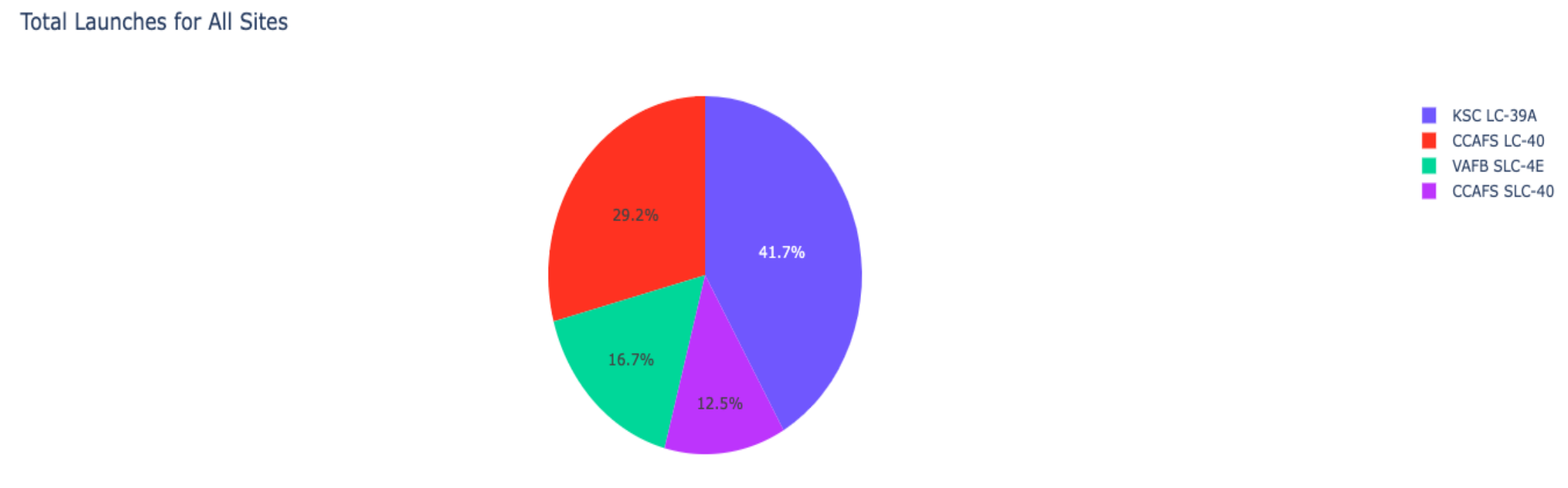
```
0.8762983388668404
```

```
# Create and add a folium.Marker on your selected closest coastline point on the map
# Display the distance between coastline point and launch site using the icon property
# for example
distance_marker = folium.Marker(
    location = [coastline_lat, coastline_lon],
    icon=DivIcon(
        icon_size=(20,20),
        icon_anchor=(0,0),
        html='<div style="font-size: 12; color:#d35400;"><b>%s</b></div>' % "{:10.2f} KM".format(distance_coastline),
        )
    )
```


Section 4
Build a Dashboard with Plotly Dash

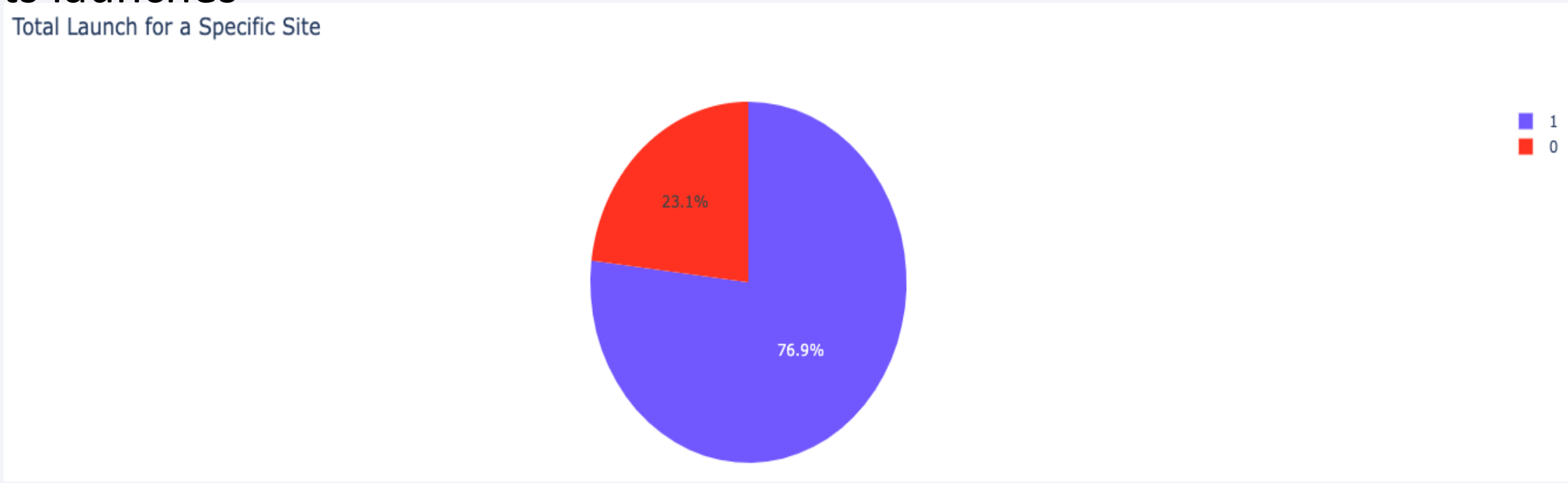
Success Percentage Composition by Each Launch Site

KSC LC-39A has the highest success percentage of 41.7% among all launch si



KSC LC-39A Success Rate

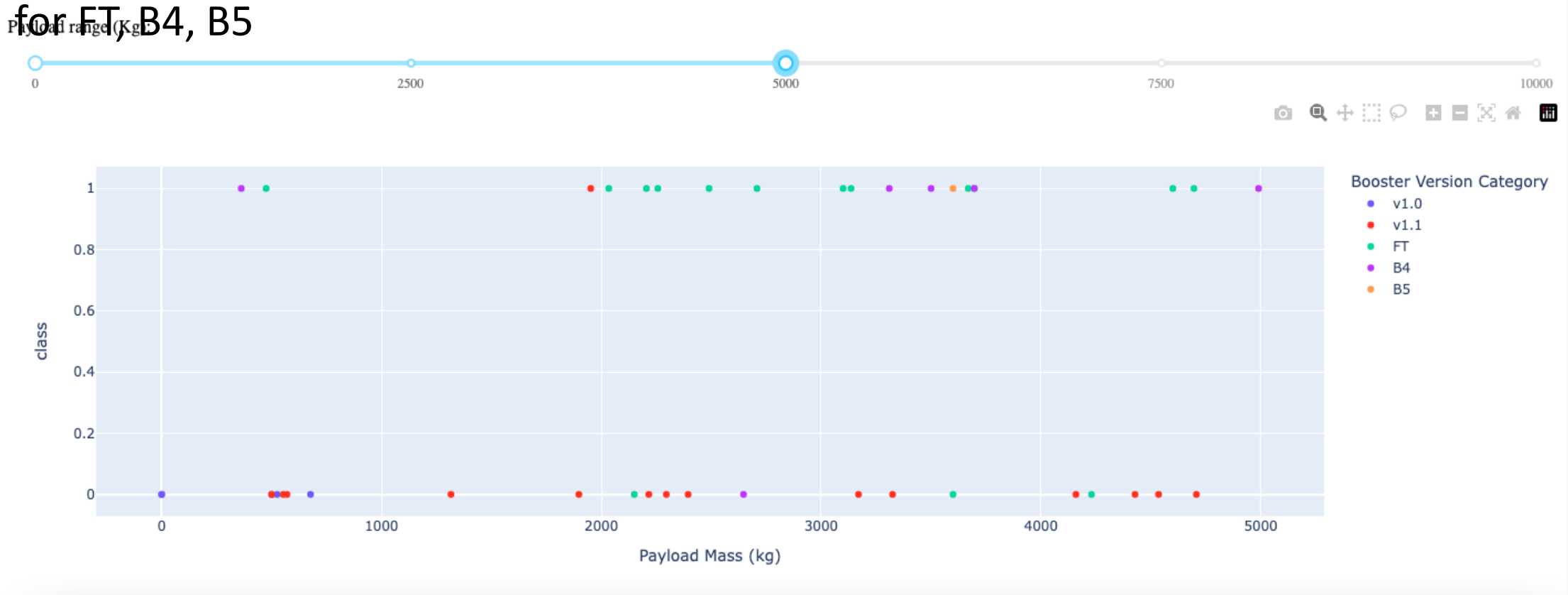
KSC LC-39A achieved a success rate of 76.9% and a failure rate of 23.1% for its launches



Payload Mass vs Launch Outcome for All Sites: 0 – 5000kg Range

Higher success rate for lower payload mass

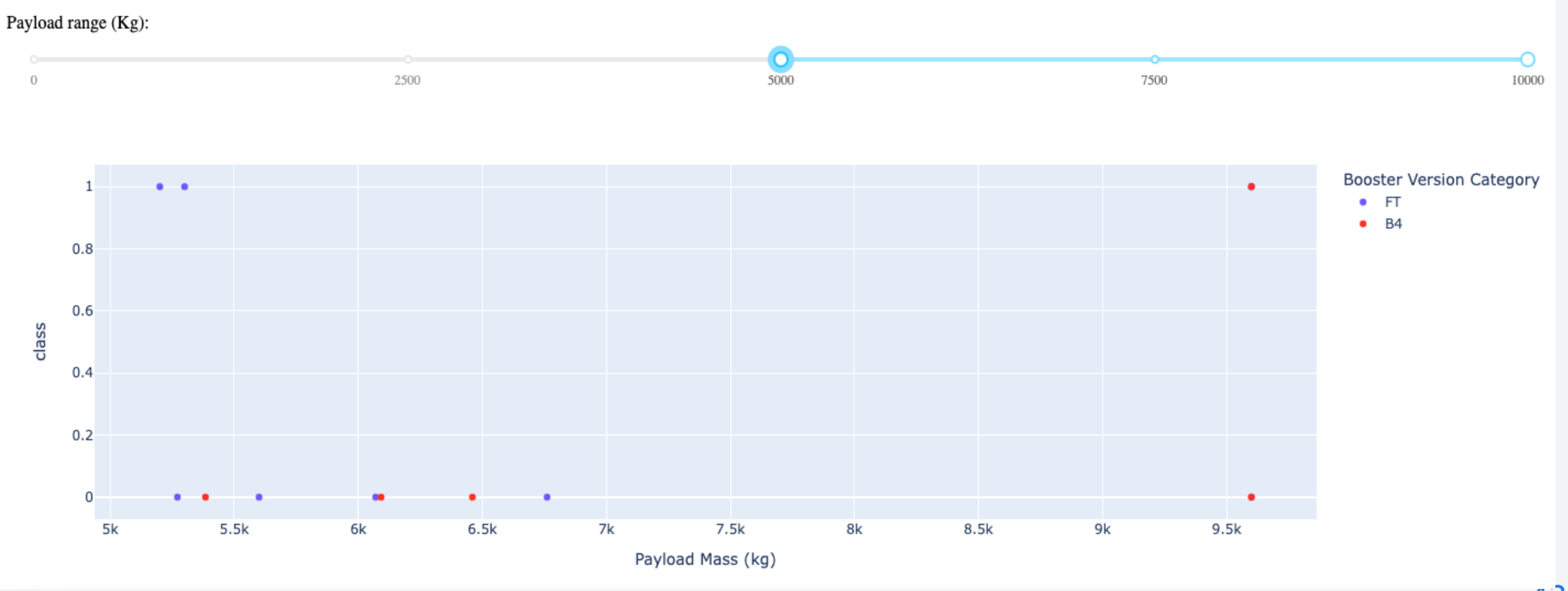
Low success rate for Booster Versions V1.0, V1.1 and higher for FT, B4, B5



Payload Mass vs Launch Outcome for All Sites: 5000 – 10000kg Range

Lower success rate for higher payload mass

Low success rate for Booster Versions V1.0, V1.1 and higher for FT, B4, B5



Section 5

Predictive Analysis (Classification)

Classification Accuracy

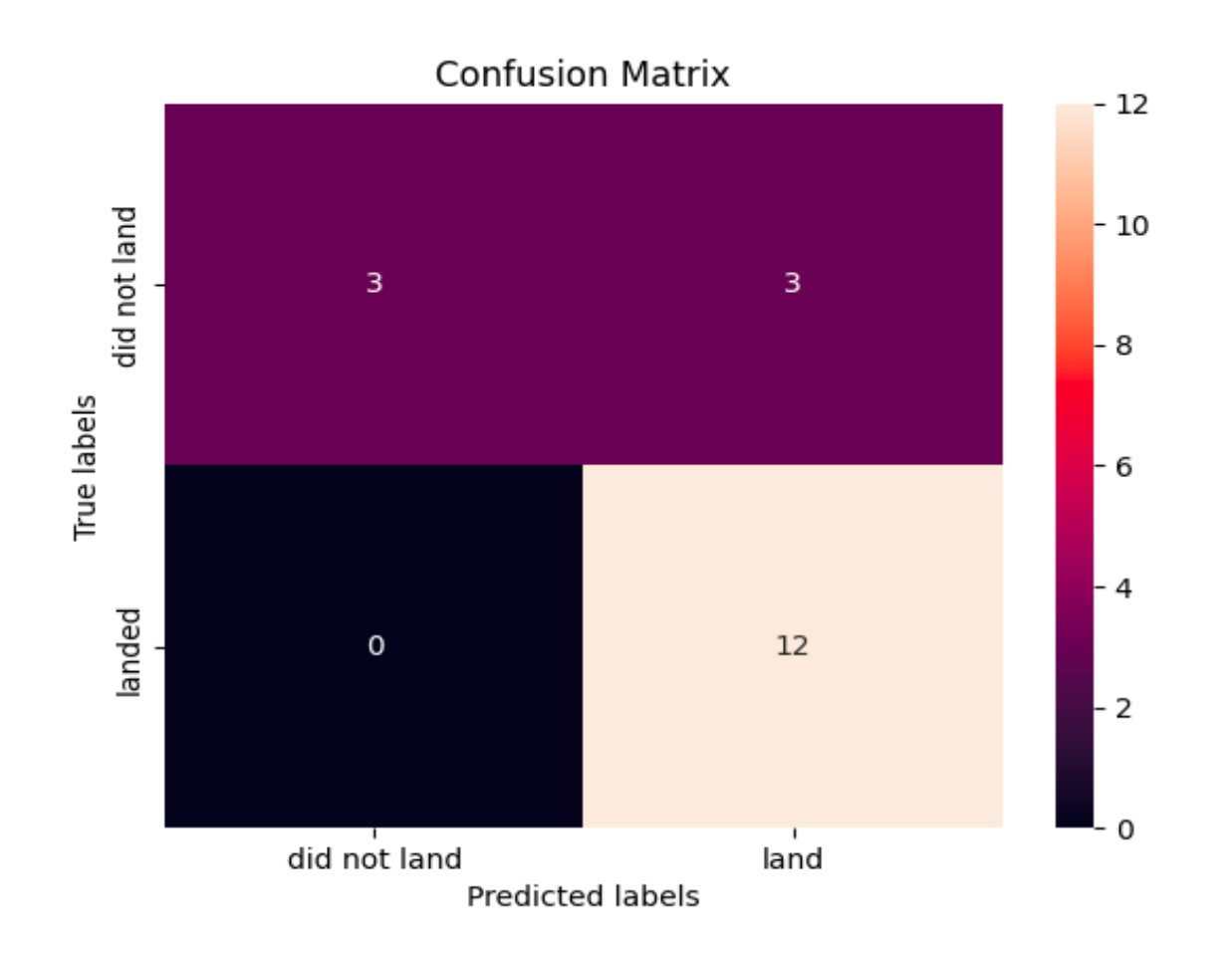
Find the method performs best:

```
predictors = [logreg_cv, svm_cv, tree_cv, knn_cv]
for predictor in predictors:
    print(predictor.best_score_)
```

```
0.8464285714285713
0.8482142857142856
0.8875
0.8482142857142858
```

The Decision Tree model has the highest accuracy of 88.75% across 4 models

Confusion Matrix



The decision tree model shows high precision in the way it can distinguish between the different classes. The major problem lie in the low recall rate in terms of false positives. In other words, the model falsely predicts unsuccessful landing as successful landing for 50% of the testing set

Conclusions

- Success rate increases with flight number increase at each launch site
- Launch success rate started to increase in 2013 till 2020.
- Higher orbits such as ES-L1, GEO, HEO, SSO have the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

Thank you!