IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Gina Ohmann
August 7, 2025

# Outline

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

# Executive Summary

SUMMARY OF METHODOLOGIES

SUMMARY OF ALL RESULTS

# Introduction

- Space X 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 Space X can reuse the first stage
- If we can determine if the first stage will land, we can determine the cost of a launch to bid against space X for a rocket launch
- Determine the most important factors that impact launch success

Section 1

# Methodology

# Methodology

Executive Summary

Data collection methodology:

- SpaceX launch data was obtained from the SpaceX Rest API in json format
- Web scraping of HTML tables and data parsed into a Pandas dataframe

Perform data wrangling

- We combined API and webscraping data, filtered on booster type, and calculated and replaced the null values with the mean of the PayloadMass

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 prepared for machine learning modeling to determine predict if the first stage will successfully land

# Data Collection

SpaceX REST API

Webscraping

Wrangling

# Data Collection – SpaceX API

- Using SpaceX REST API

- SpaceY/jupyter-labs-spacex-data-collection-api.ipynb at main · gohmann/SpaceY

Used the url = "https://api.spacexdata.com/v4/launches/past"

Collected data: rocket used, payload delivered, launch specifications, landing specifications, and landing outcome

Performed a 'get' request

response = requests.get(url)

Obtained response in form of json objects

# Data Collection - Scraping

- SpaceY/jupyter-labs-webscraping.ipynb at main · gohmann/SpaceY

Web scraping HTML tables containing Falcon 9 Launch Records

Web scraping with the Python BeatifulSoup Package

Parse data in those tables and convert them into a Pandas data frame for further visualization and analysis

# Data Wrangling

Converted API json objects to a flat data table.   data = pd.json_normalize(response.json())

Transformed raw data: in some of the columns, like rocket, we have an identification number, not actual data

Used the API again to gather data for each ID number for: Booster, Launchpad, payload, and core. Stored data in lists.

Filtered/sampled data to remove Falcon 1 launches to focus on the Falcon 9 launches.

Replaced null values of PayloadMass with the mean.

Left LandingPad with NULL values & dealt with these using one hot encoding later in the process.

- SpaceY/labs-jupyter-spacex-Data wrangling.ipynb at main · gohmann/SpaceY

# EDA with Data Visualization

- Plots were made to observe relationships between factors to select the features to be used for launch success prediction.

  Flight Number & Launch Site overlaying the outcome of the launch – scatterplot

  Payload Mass & Launch Site – scatterplot

  Success Rate & Orbit Type – bar chart

  Flight Number & Orbit Type – scatterplot

  Payload Mass & Orbit Type – scatterplot

  Launch Success Yearly Trend – Line plot

- SpaceY/edadataviz.ipynb at main · gohmann/SpaceY

# EDA with SQL

- SQL Queries were performed to explore the data
- Launch Site Names
- Payload Mass Totals & Averages
- Dates of first successful landings
- Names of boosters with success in payload mass specific range
- Totals of successful & failure mission outcomes
- List of boosters versions that carried the maximum payload mass
- Selected dates and ranked landing outcomes within a specific time frame
- SpaceY/jupyter-labs-eda-sql-coursera_sqllite (1).ipynb at main · gohmann/SpaceY

# Build an Interactive Map with Folium

- Folium circles, markers, marker clusters, and color-labeled markers were added to the map
- These markers identify locations, manage areas where there were a high number of markers, and allow for ease of identification of launch sites with high success rates.

- SpaceY/lab_jupyter_launch_site_location (1).ipynb at main · gohmann/SpaceY

# Build a Dashboard with Plotly Dash

- Created a Dashboard including a Dropdown, Pie chart, Range slider, and Scatter plot
- For Users to perform interactive visual analytics on SpaceX launch data in real-time:
  - Select the launch site (Drop-down)
  - Observe success/failure of launches at the site (Pie Chart)
  - Select the payload (Range Slider)
  - Render a Success/Payload Chart (Scatter Plot)
- SpaceY/spacex-dash-app.py at main · gohmann/SpaceY

# Predictive Analysis (Classification)

Created a NumPy array with column Class. Applied the method to_numpy() assigned to the variable Y with output as a Pandas series

Standardized & transformed data in X using
transform = preprocessing.StandardScaler()
X = transform.fit_transform(X)

Used function train_test_split to split the data X and Y into training and test data. Parameter test_size to 0.2 and random_state to 2. The training data and test data assigned to the labels: X_train, X_test, Y_train, Y_test

Created objects:
Logisic Regression
Support Vector Machine SVM
Decision Tree Classifier
K Nearest Neighbor KNN

Created GridSearch CV object (cv=10)

Fit the object to find best dictionary parameters

Calculated accuracy on the test data using Score method

Produced a confusion matrix

- SpaceY/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · gohmann/SpaceY

# Results

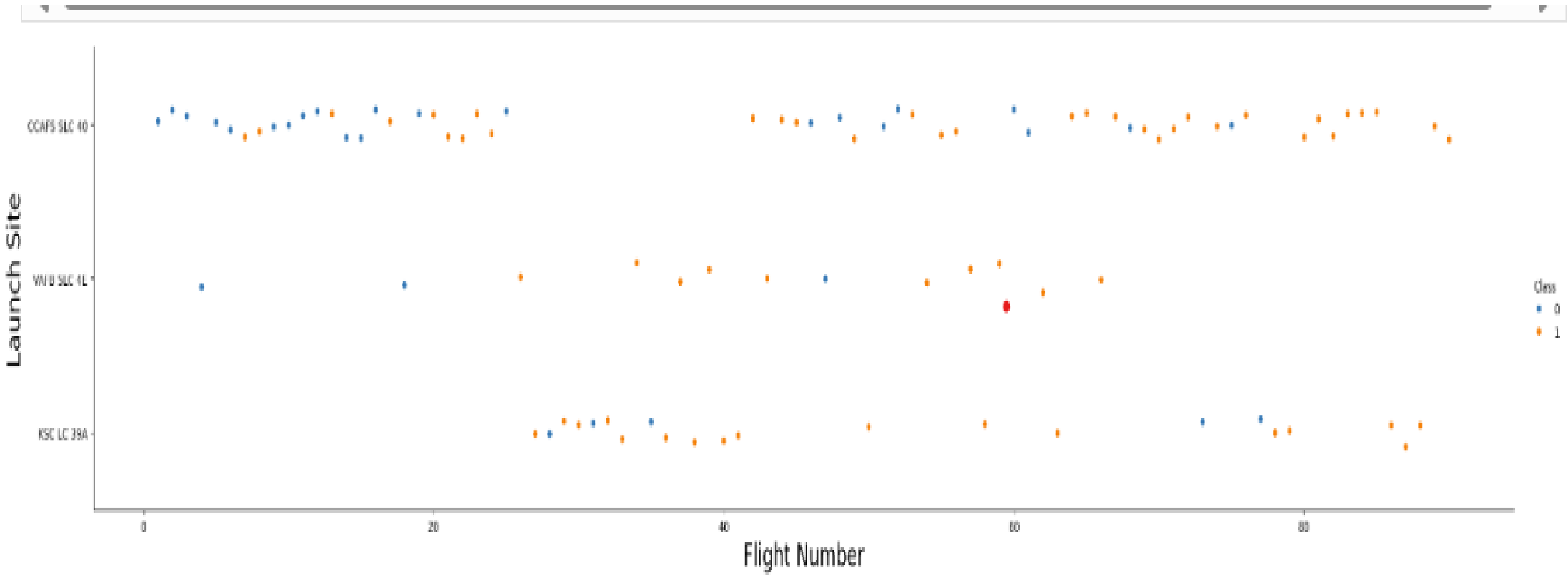
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

# Insights drawn from EDA

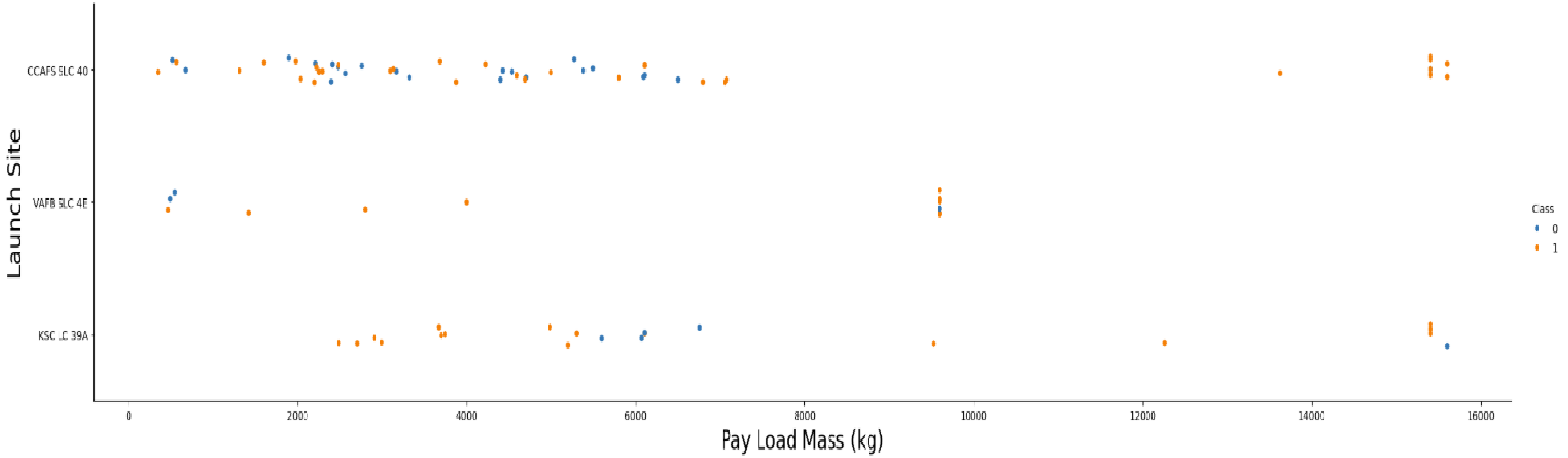
# Flight Number vs. Launch Site

- CCAFS SLC 40 had the highest number of launches overall with 33 of 55 successful.
- KS LL39A had the second highest number of successful launches with 17 of 22 successful.



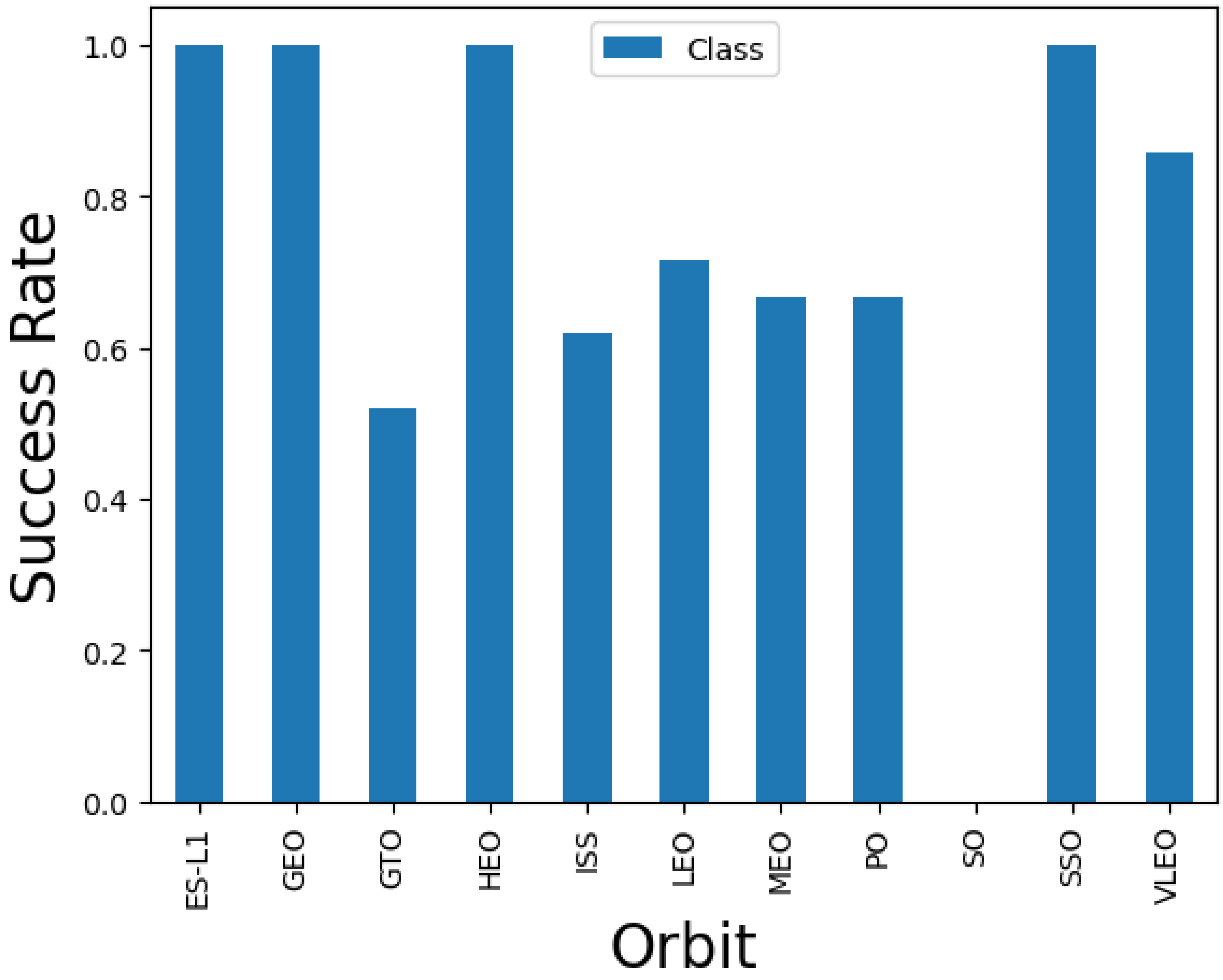
# Payload vs. Launch Site

- Most successful launches were spread across a range for CCAFS SLC 40 & KSC LC-39A
- VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000kg)



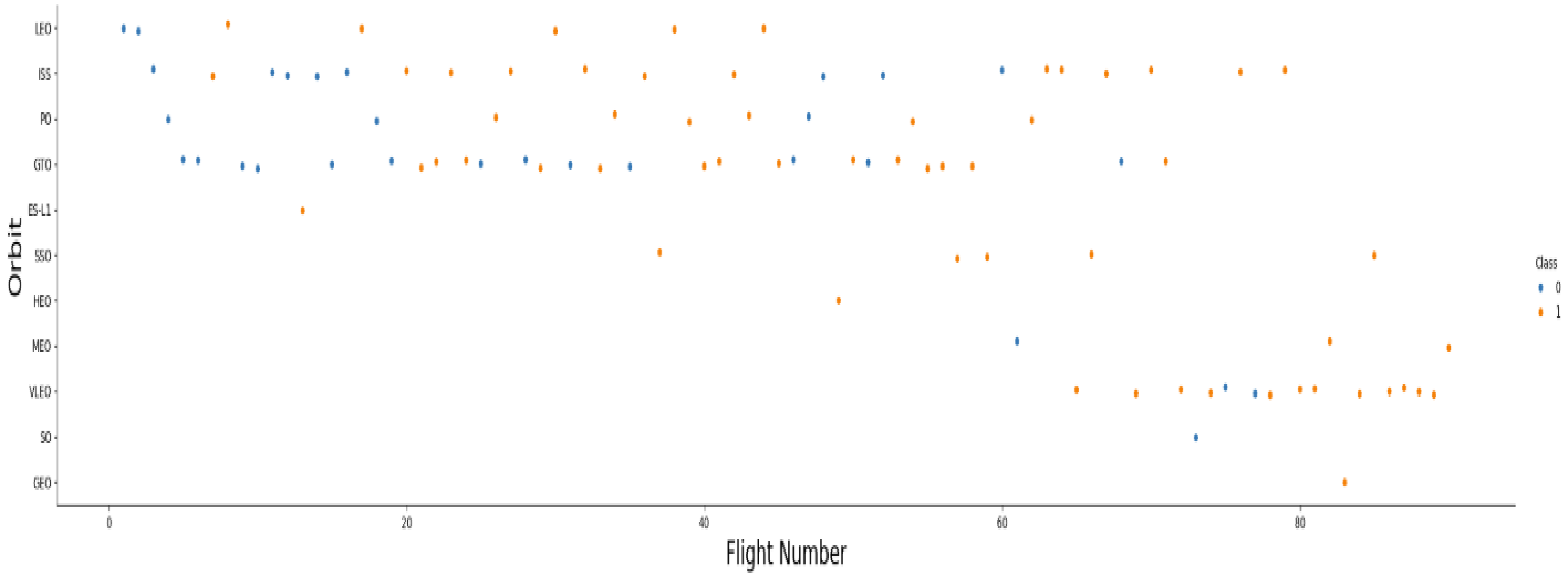
# Success Rate vs. Orbit Type

Orbit Types: ES-L1, GEO, HEO, & SSO had the highest landing success rate



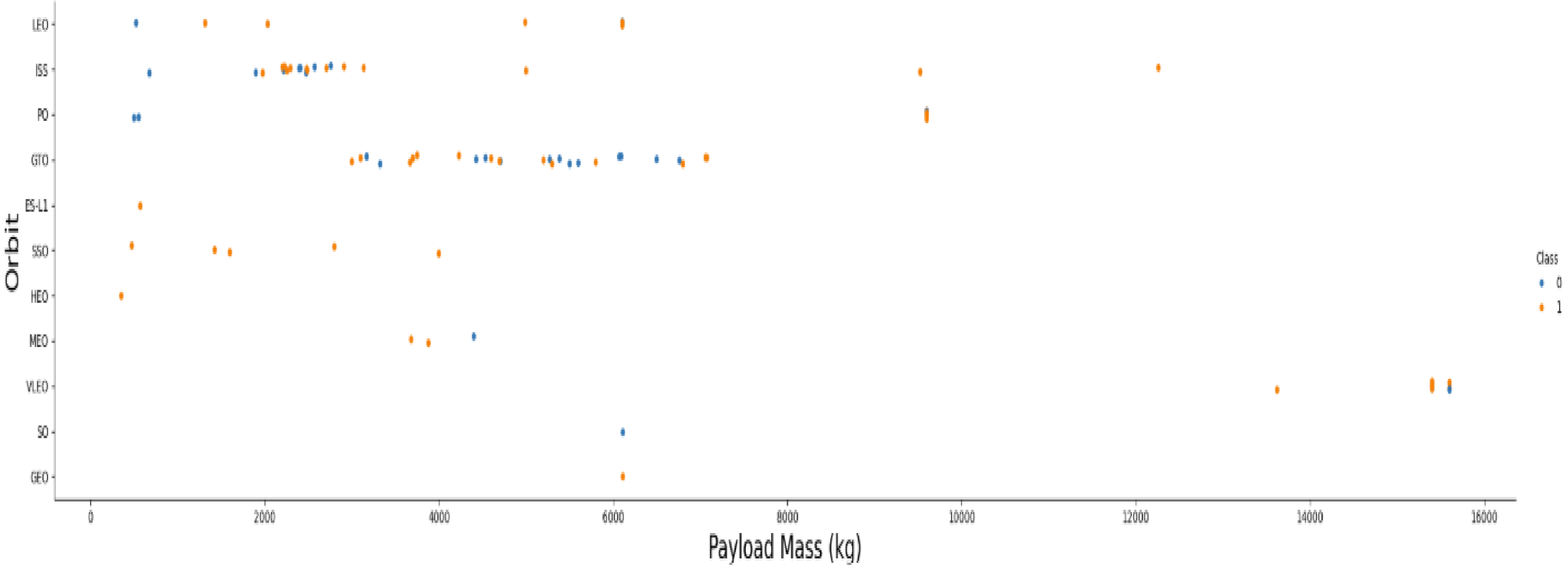
# Flight Number vs. Orbit Type

- Success in the LEO orbit appears to be related to the number of flights
In the GTO orbit, there appears to be no relationship between flight number and success



# Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.



# Launch Success Yearly Trend

The success rate since 2013 kept increasing until 2020

# All Launch Site Names

The unique launch site names were identified using:

Query:

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

# Launch Site Names That Begin with 'CCA'

| Launch_Site |
| --- |
| CCAFS LC-40 |
| CCAFS LC-40 |
| CCAFS LC-40 |
| CCAFS LC-40 |
| CCAFS LC-40 |

Find 5 records where launch sites begin with `CCA`

%sql SELECT "Launch_Site" FROM SPACEXTABLE WHERE "Launch_Site" like '%CCA%' LIMIT 5;

# Total Payload Mass

The total payload mass carried by boosters launched by NASA (CRS):

48213 kg

Query used:

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer LIKE '%NASA (CRS)%';
```

# Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1:

2534.66 kg

Used query:

%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE '%F9 v1.1%';

# First Successful Ground Landing Date

The date when the first successful landing outcome in ground pad was achieved:

2015-12-22

Used query:

%sql SELECT MIN(Date) from SPACEXTABLE WHERE Landing_Outcome == 'Success (ground pad)';

# 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 |

Query Used:

```
%sql SELECT Booster_Version FROM SPACEXTABLE WHERE
Landing_Outcome == 'Success (drone ship)' AND PAYLOAD_MASS__KG_
BETWEEN 4000 AND 6000;
```

# Total Number of Successful and Failure Mission Outcomes

**Successful mission outcomes: 100**

Used query:

```
%sql SELECT COUNT("Misson_Outcome") FROM
SPACEXTABLE WHERE "Mission_Outcome" LIKE '%Success%';
```

**Failure mission outcomes: 1**

Used query:

```
%sql SELECT COUNT("Misson_Outcome") FROM
SPACEXTABLE WHERE "Mission_Outcome" LIKE '%Failure%';
```

# Boosters Carried Maximum Payload

- %sql SELECT Booster_Version, PAYLOAD_MASS__KG_ FROM SPACEXTABLE GROUP BY Booster_Version ORDER BY PAYLOAD_MASS__KG_ DESC LIMIT 12

| Booster_Version | PAYLOAD_MASS__KG_ |
|---|---|
| F9 B5 B1060.3 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1049.7 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1049.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1048.4 | 15600 |

# 2015 Launch Records

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.

Query: %sql SELECT substr(Date, 6,2), Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE WHERE Landing_Outcome=='Failure (drone ship)' AND substr(Date,0,5)='2015';

| substr(Date, 6,2) | Landing_Outcome | Booster_Version | Launch_Site |
|---|---|---|---|
| 01 | Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| 04 | Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

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

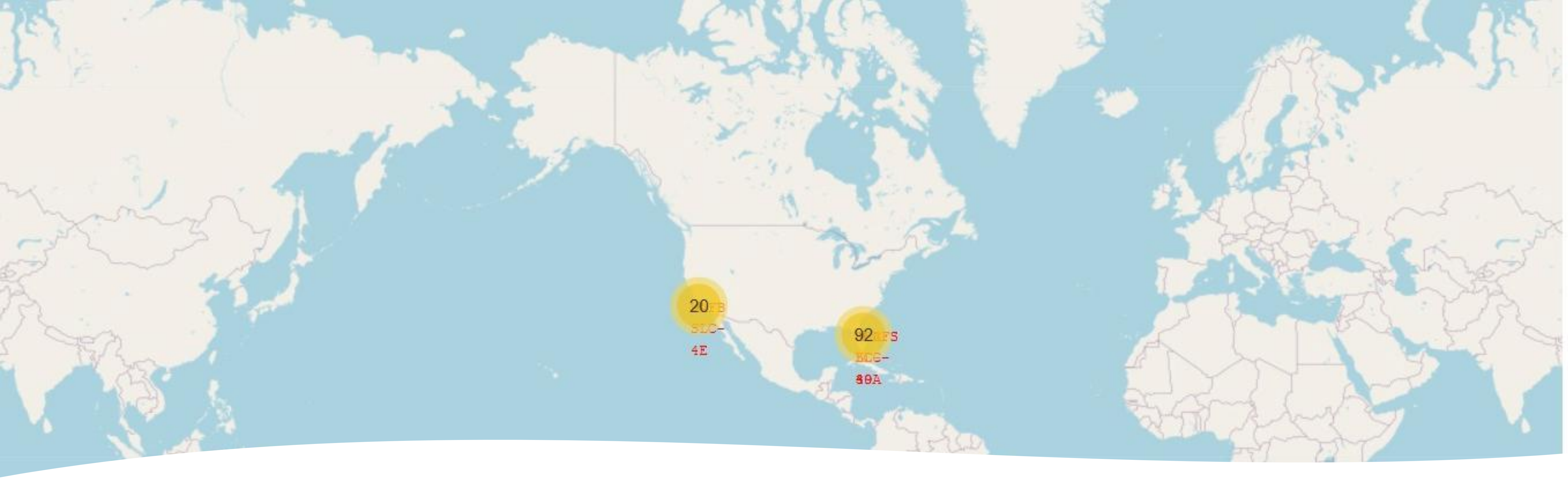
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

Query: %sql SELECT Landing_Outcome, COUNT(*) AS Outcome_Count, RANK() OVER (ORDER BY COUNT(*) DESC) AS Outcome_Rank FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome;

| Landing_Outcome | Outcome_Count | Outcome_Rank |
|---|---|---|
| No attempt | 10 | 1 |
| Success (drone ship) | 5 | 2 |
| Failure (drone ship) | 5 | 2 |
| Success (ground pad) | 3 | 4 |
| Controlled (ocean) | 3 | 4 |
| Uncontrolled (ocean) | 2 | 6 |
| Failure (parachute) | 2 | 6 |
| Precluded (drone ship) | 1 | 8 |

Section 3

# Launch Sites Proximities Analysis




# Worldwide SpaceX Launch Sites

- Launch sites proximities:
  - 0.6 to 1.3 km from railways
  - 0.6 to 8.8 km from highways
  - 0.8 to 1.3 km from coastlines
  - 12 to 52km from cities

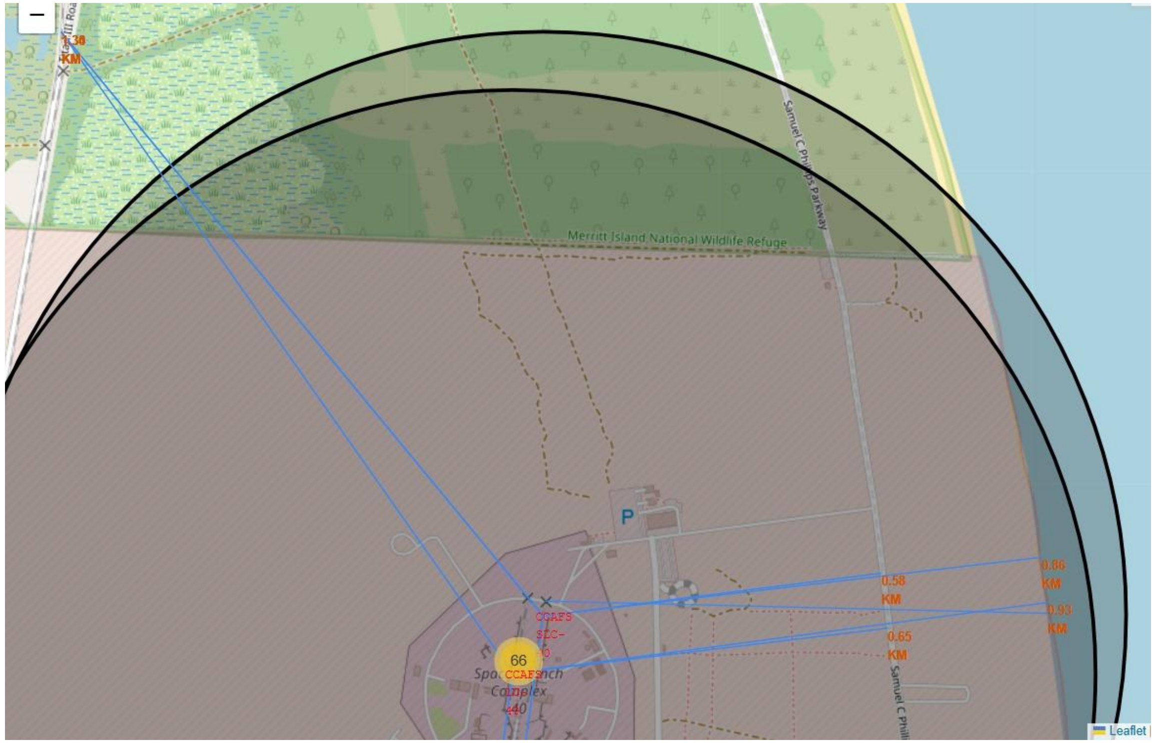
# Successful Launch Outcomes

- CCAFS LC-40 had highest success
- Successful launches in green
- Unsuccessful launches in red

# CCAFS LC-40

## Proximities

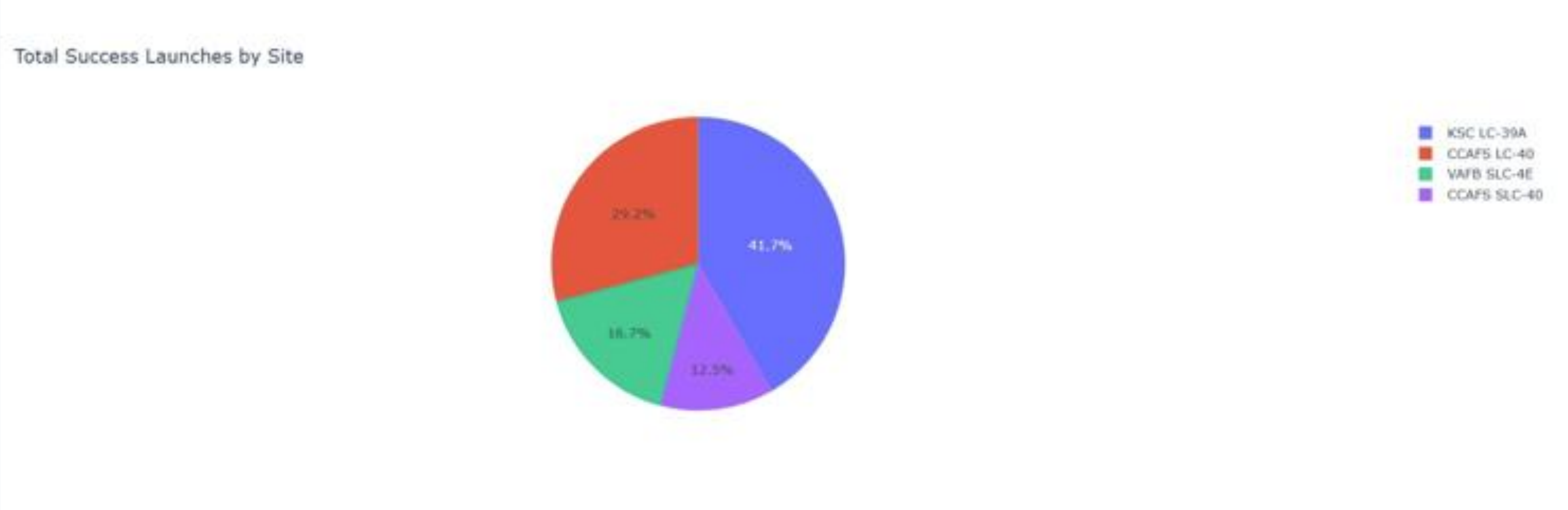
- 52.82 km from a public airport
- 0.65 km from a parkway
- 0.93 km from the coastline
- 1.38 km from railway



Section 4

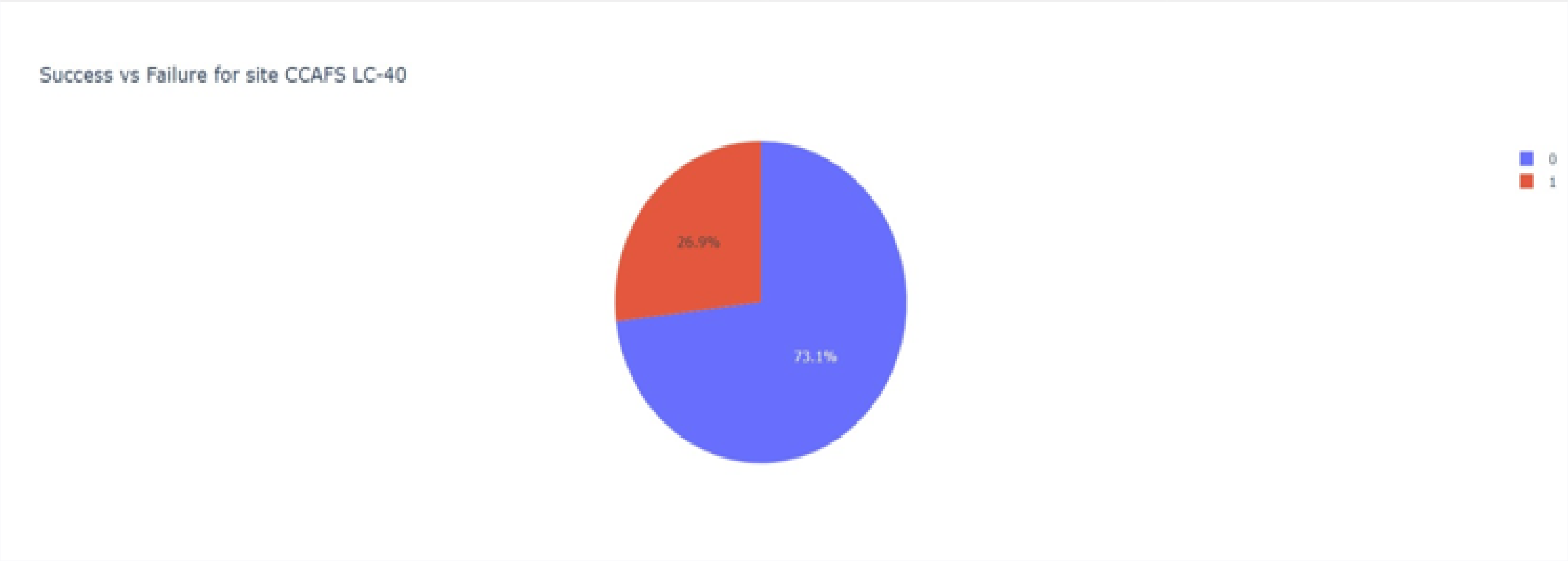
# Build a Dashboard with Plotly Dash

# Total Successful Launches By Site



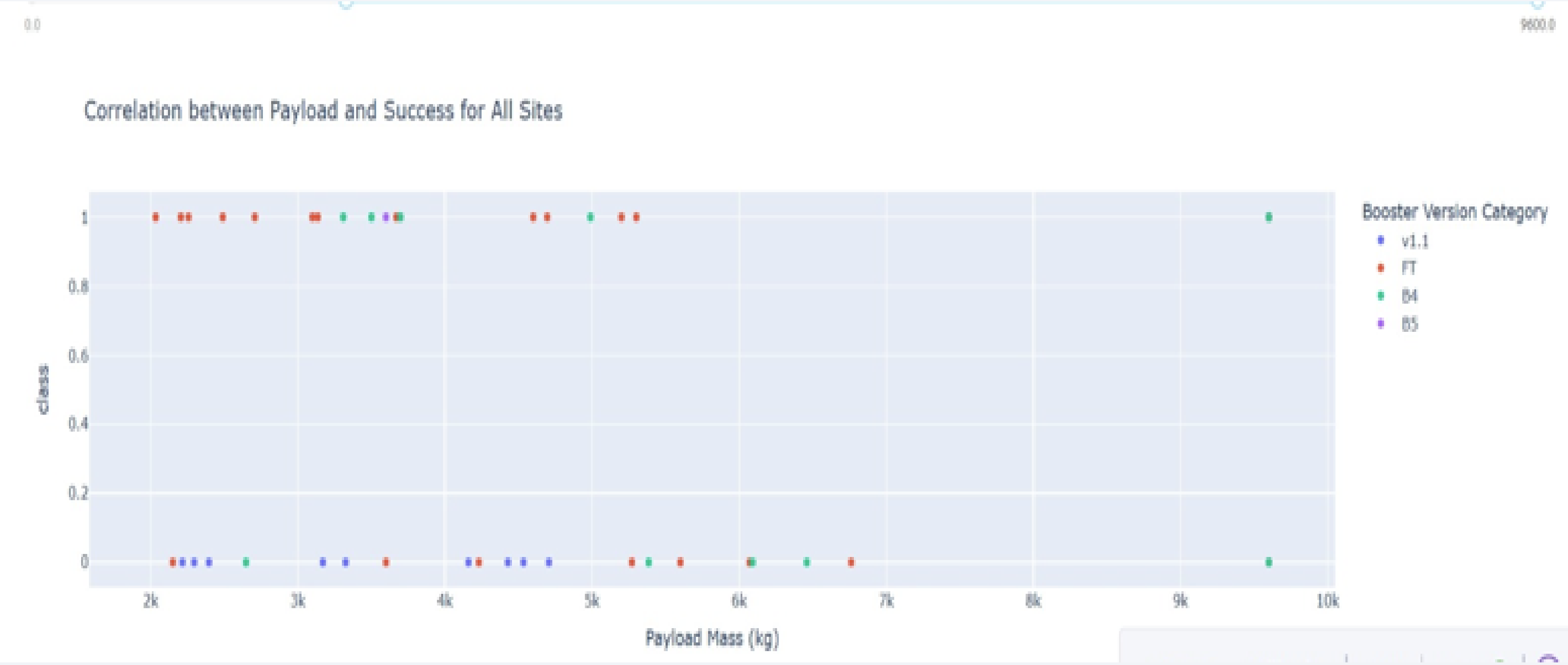

When adding the total number of successful launches by site, KSC LC 39A had the highest number

# Site with the Highest Launch Success Ratio




CCAFS LC-40 had the highest ratio of launch success

# Payload vs. Launch Outcome Scatter Plot All Sites

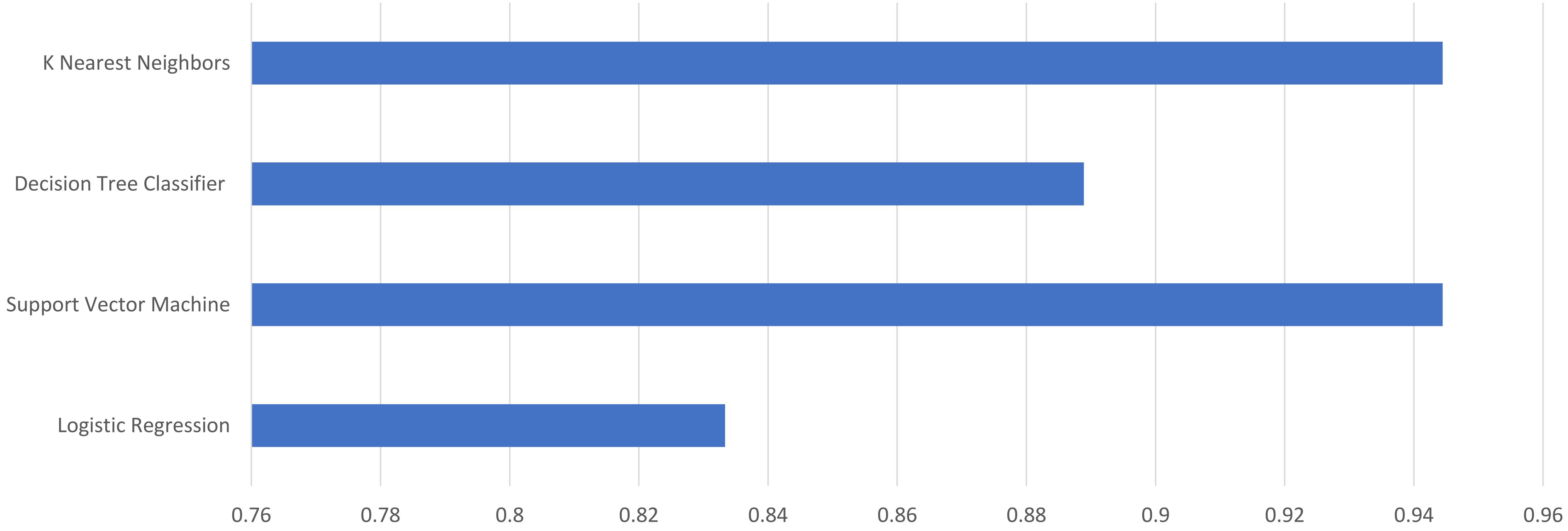



- The Payload Range with the highest launch success rate was: 2034 kg to 3696 kg
- The FT Version of the F9 Booster has the highest launch success rate

Section 5

# Predictive Analysis (Classification)

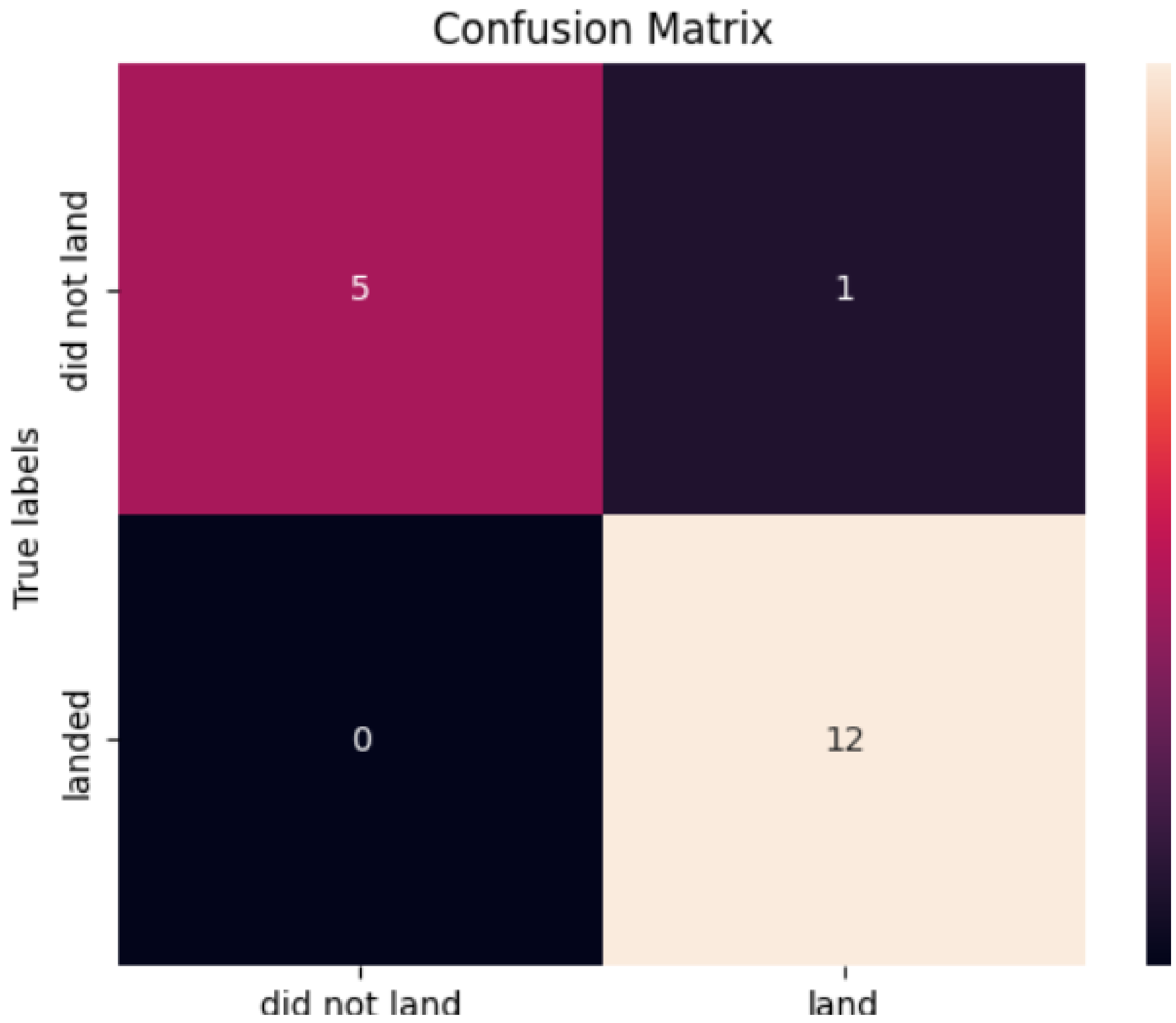
# Classification Accuracy – Tie between SVM & KNN




Both SVM & KNN models had an accuracy score o 94.4%

# Confusion Matrix

Both the SVM & KNN Models performed the best on the test data with an accuracy of  94.4% and identical Confusion Matrices.



# Conclusions

## Selection of launch site

- KSC LC-39A launch site had the largest number of successful launches
- CCAFS LC-40 had the overall highest success rate.

## Payload Mass Range

- Successful launches had payloads of 2034 kg to 3696 kg
- Payloads above or below that mass had lower successful launches

## Booster selection is key

- The F9 Booster version with the highest launch success rate is FT

# Appendix

Jupytr Notebooks, Python files, Screenshots, and other supporting documentation related to this Power Point presentation located at:

gohmann/SpaceY: Space Y Capstone Project

# Thank you!