
# Winning Space Race with Data Science

Sherwina Augustine
04/09/22

# Outline

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

# Executive Summary

## Summary of Methodologies

Data Collection

Data Wrangling

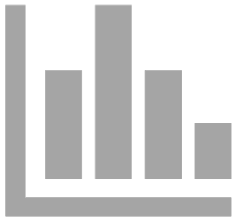
EDA with Data Visualizations

EDA with SQL

Interactive map with Folium

Dashboard with Plotly Dash

Machine Learning Predictions



## Summary of all results

EDA Results

Interactive Analysis

Predictive Analysis Results

# Introduction

SpaceX was founded to revolutionize space technology towards expanding the life on multi-planets. SpaceX is the world's leading provider of launch services and is one of the first company to complete an all-civilian crewed mission to orbit.

Using this information from the Falcon 9 Data, will help companies predict what will lead to a successful Launch

Section 1

# Methodology

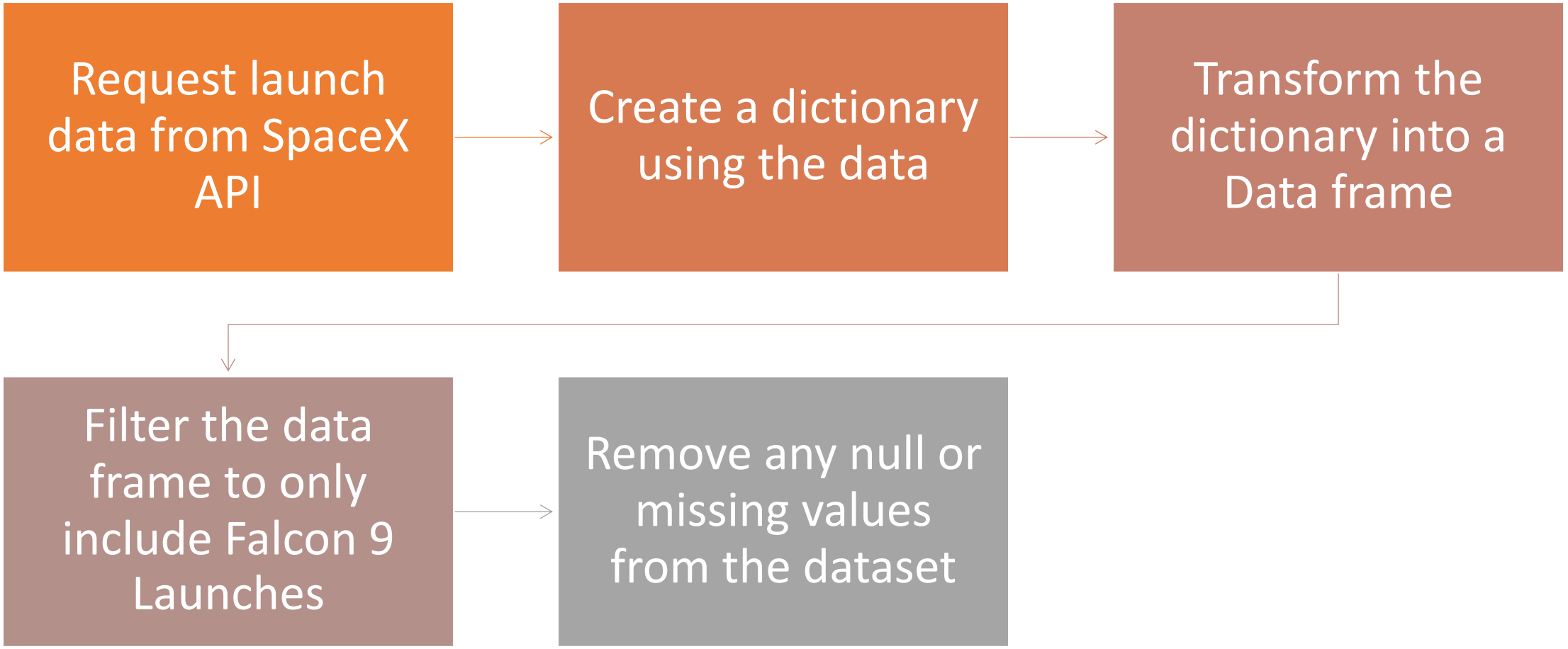
# Methodology

- Data collection methodology:
  - *Data was collected using a REST API gathered from the SpaceX launch data.*
  - *Web scrapping from Wikipedia*
- Perform data wrangling
  - *Transform data for Machine learning*
- Perform exploratory data analysis (EDA) using visualization and SQL
  - *Analyze results by using Scatter and Bar graphs, to detect any relationship*
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - *How to build, tune, evaluate classification models*

# Data Collection

- The SpaceX launch data was collected using SpaceX REST API
- Use the API to analyze the information in the data such as, rockets used, mass payloads, flight number, date and the launch specifications.
- Requesting the Falcon9 Launch data from Wikipedia, using the beautifulsoup.

# Data Collection – SpaceX API




Data Collection Github Link

# Data Collection - Scraping

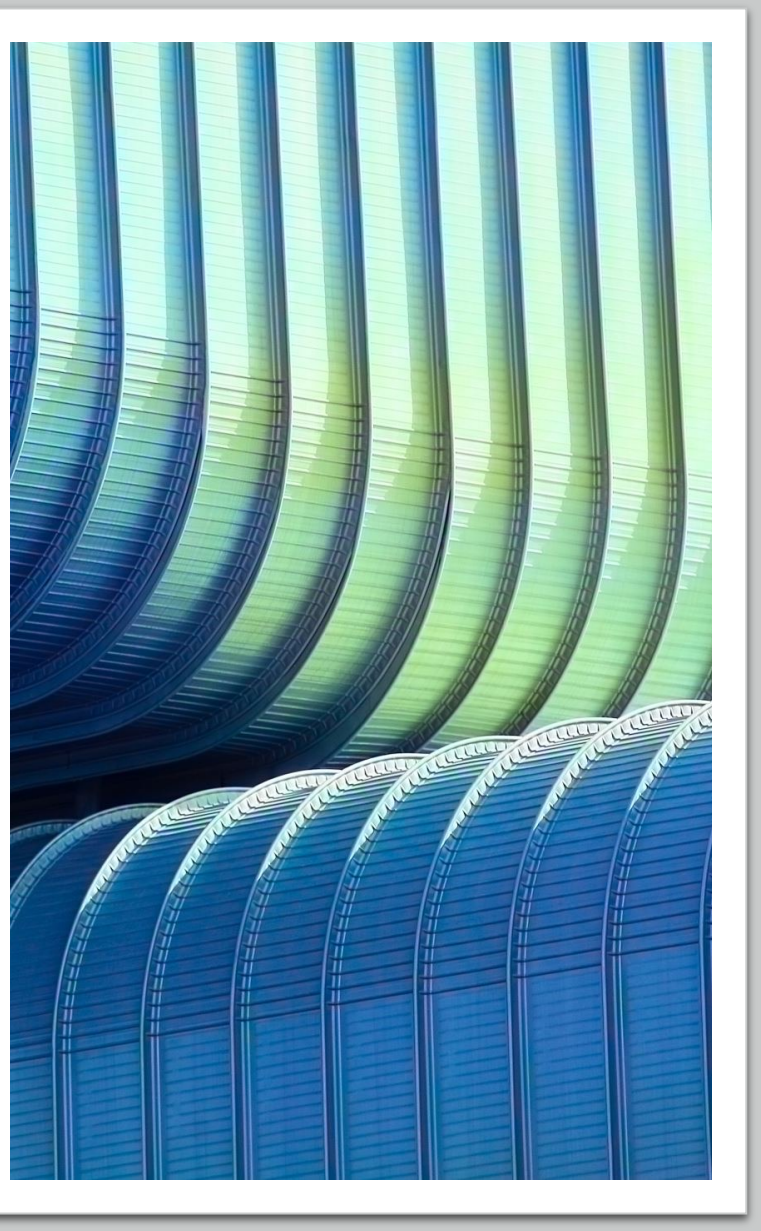
Web Scrapping Github Link

Request Falcon 9 Launch Wiki page

Create a BeautifulSoup from HTML

Add the data to a dictionary

Convert the dictionary to a Dataframe



# Data Wrangling

- Describe how data were processed
  - During this process we analyze the success rate of boosters landing successfully.
  - Training Labels use were 0 means it was unsuccessful and 1 for successful.

- There are several outcomes which can describe the landing attempt:
  - True ASDS → Successful Landing on a drone ship
  - False ASDS → Unsuccessful Landing on a drone ship
  - True RTLS → Successful Landing on a ground pad
  - False RTLS → Unsuccessful Landing on a ground pad
  - True Ocean → Successfully landed to a specific region of the ocean
  - False Ocean → Unsuccessfully Landed to a specific region of the ocean

- Data Wrangling Github Link

# EDA with Data Visualization

Different types of graphs were using to analyze the relationship between the variables.

- Flight Number Vs. Launch Site (Cat Plot)
- Flight Number Vs. Orbit Type (Scatter Plot)
- Payload Vs. Launch Site (Cat Plot)
- Payload Vs. Orbit Type (Scatter Blot)
- Success Rate Vs Orbit Type (Bar plot)
- Launch yearly trend (Line Plot)

- Data Visualization Github Link

# EDA with SQL

- Using SQL, we were able to process basic queries to get more insights within the data. In addition, we were able to analyze any relationship between different variables such as
  - Launch Sites
  - Payload Mass (kg)
  - Mission Outcome
  - Booster Version
  - Date

- EDA with SQL Github Link

# Build an Interactive Map with Folium

- Using the package Folium within Python, we were able to create an interactive map where we were able to visualize the following:
  - Falcon 9 Launch Sites, which were represented with a red circle
  - The number of launches occurred at each site; this was represented with green markers for those that were successful and red markers for those that were not.
  - Determine distances between any railways, highways and coastline. This was represented by a blue line connecting the locations.

- Folium Github Link

# Build a Dashboard with Plotly Dash

- With Plotly Dash, we were able to complete a Dashboard that shows:
  - Pie Chart, which demonstrates the success/Unsuccessful rates of all the Boosters. We can view each Booster separately using the dropdown menu.
  - Scatter Plot, which shows the range of the payload Mass from 0- 10000 kg.
- The Dashboard provides visualize insights on the launches and the Size of the Payload Mass; This will provide us evidence of any Relationships.

# Predictive Analysis (Classification)

Standardize the data

Split the data into training and test data

Create a Logistic Regression, Support Vector, Decision Tree, KNN

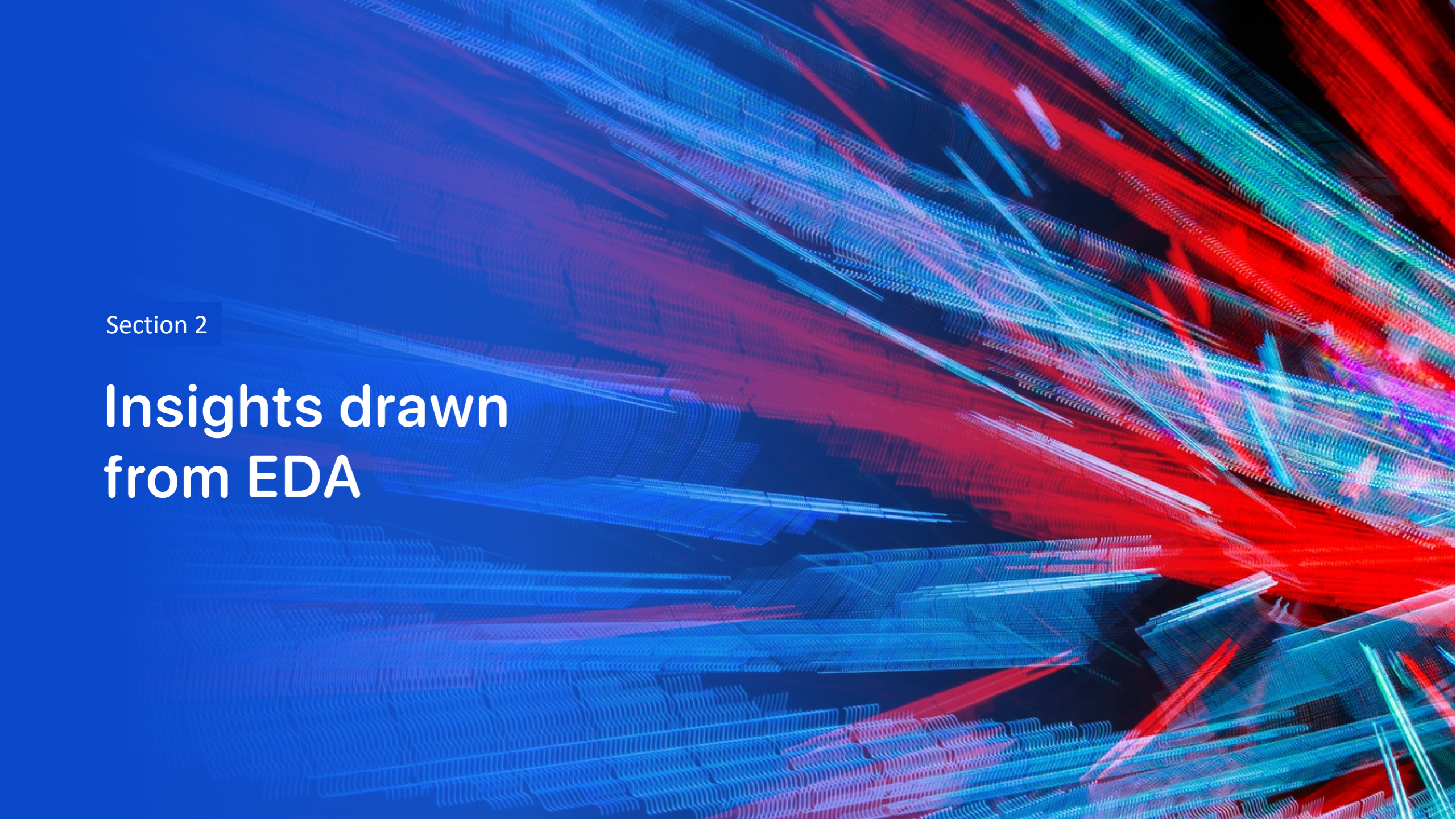
Create a GridSearchCV to find the best parameters

Test Models using testing data, Confusion Matrix

[Machine Learning Github Link]

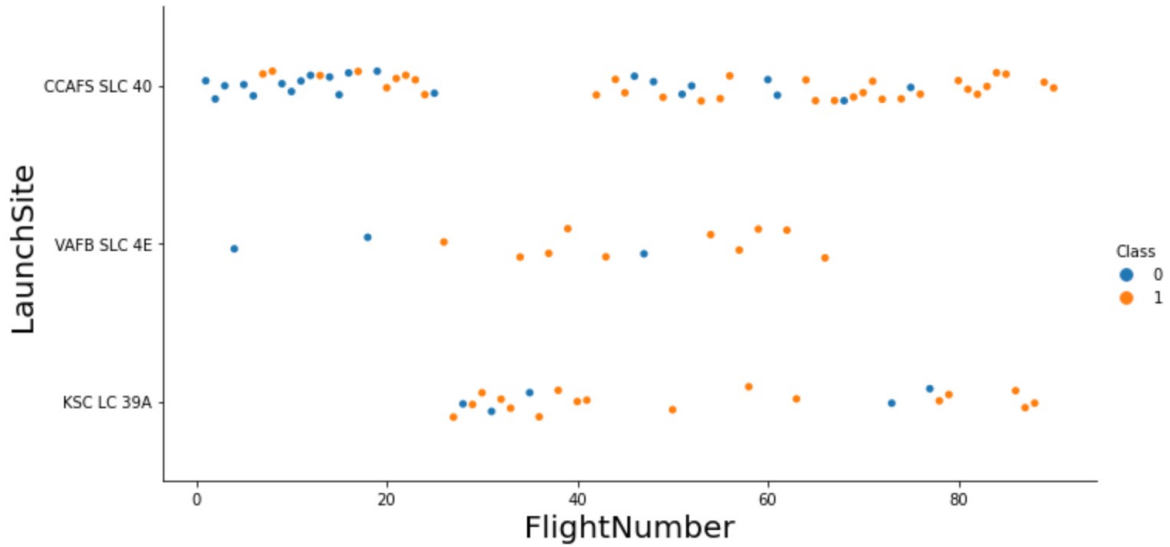
# Results

- Exploratory data analysis results
  - Higher success rate when carrying Payload Mass between 2000 – 4000 kg.
  - KSC LC-39A has the highest success rate with a rate of 76.9%
- Interactive analytics demo in screenshots
  - Screenshots are provided on Slide 34.
- Predictive analysis results
  - 3 Models Accuracy percent were very similar, with the Accuracy rate of 83.33%
  - The Decision Tree, However was less at around 77.78%


Section 2
Insights drawn from EDA

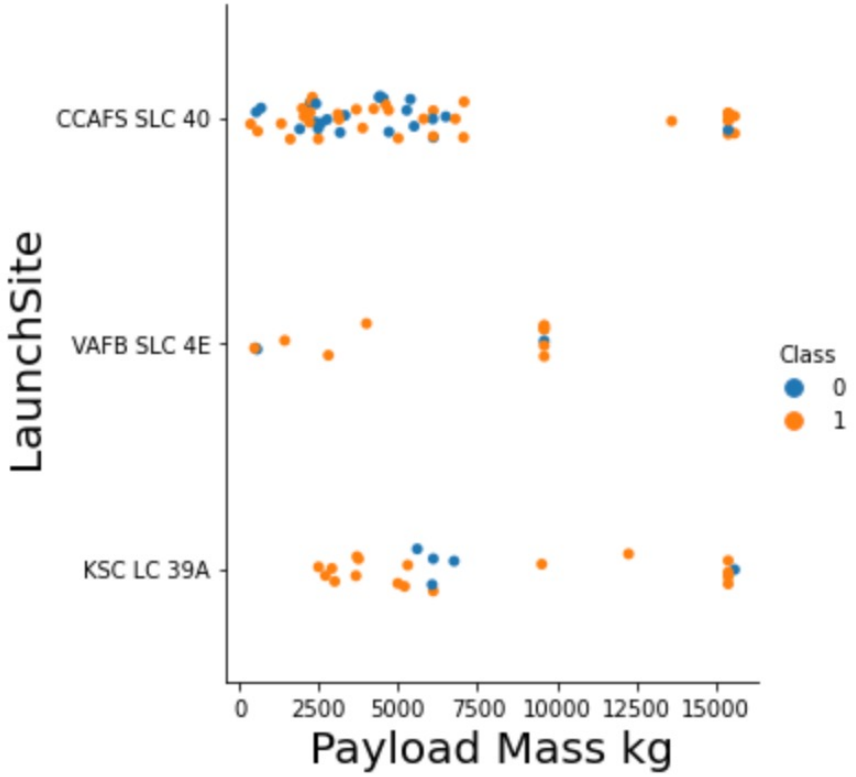
# Flight Number vs. Launch Site

- Between the 3 launch sites, CCAFS SLC-40, conducted more test.
- VAFB SLC-4E, conducted fewer test, but received a high success rate, with very few failures.

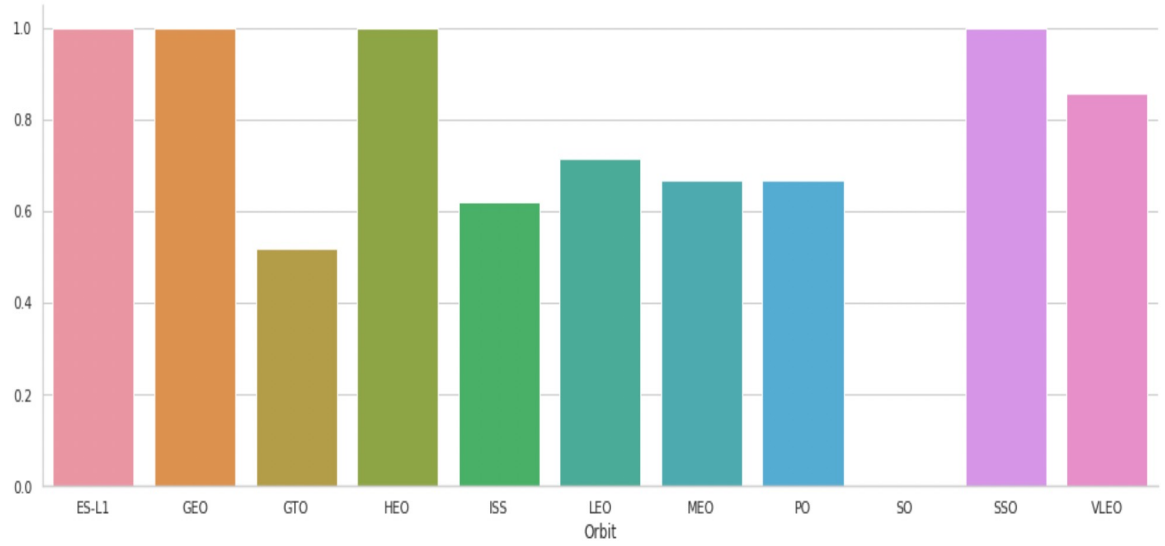


# Payload vs. Launch Site

- Payloads less that 6000 kg, seem to have a higher success rate



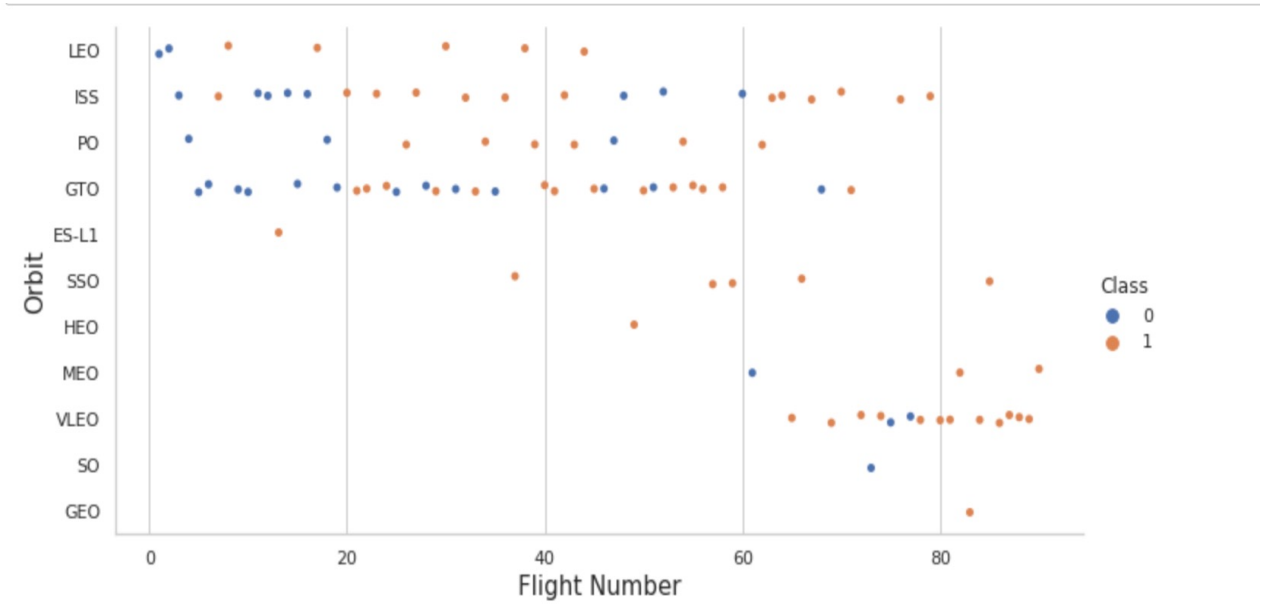
# Success Rate vs. Orbit Type




There are 4 Orbit with high success rate: ES-L1, GEO, HEO and SSO

We also notice, there is one orbit without any success: SO

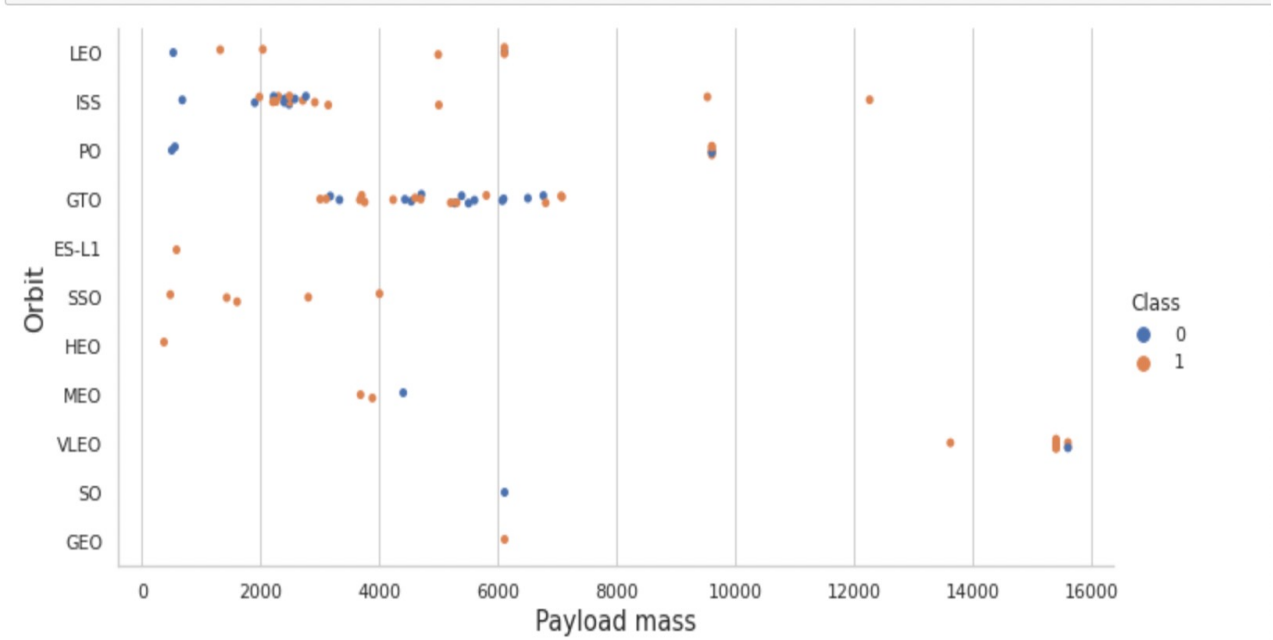
# Flight Number vs. Orbit Type




- GEO, HEO, and ES-L1, shows high reliability, but may be skewed since they only conducted one launch.
- LEO, VLEO Shows high success rate and have multiple successful launches.

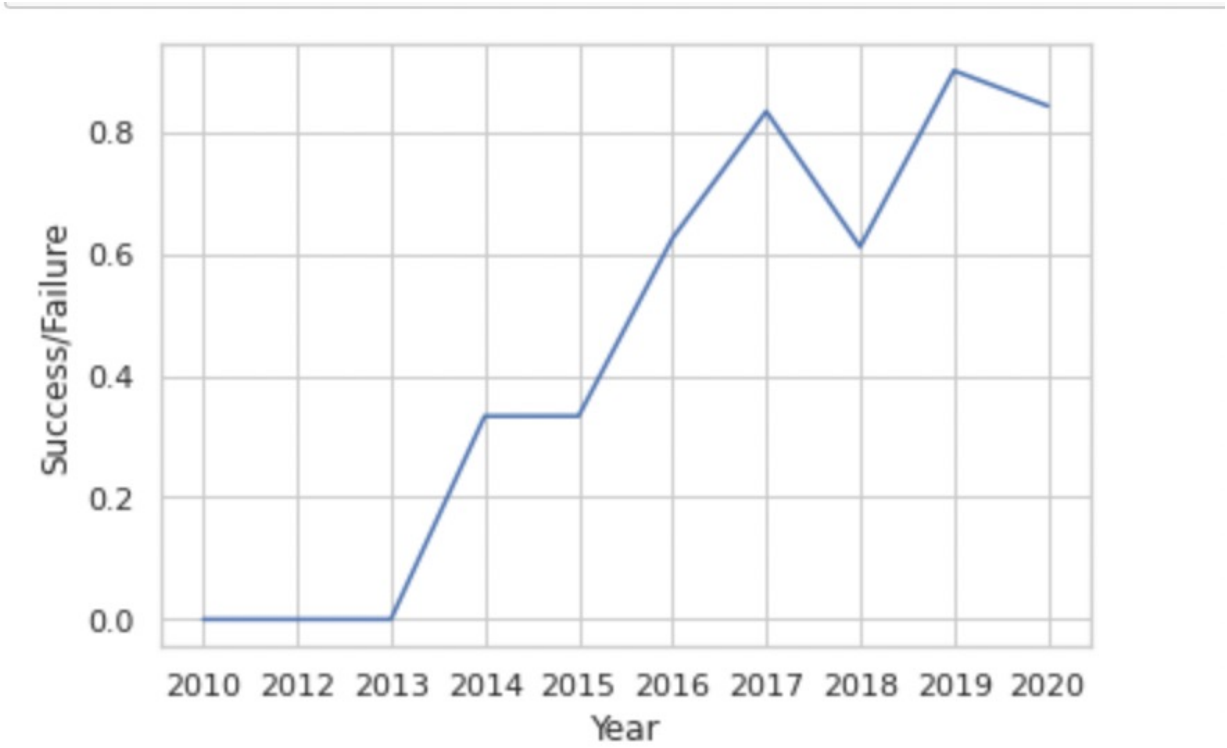
# Payload vs. Orbit Type

- We noticed that small Payload Mass did very well with the success of the launches. SSO and LEO orbit seem to perform better with lighter Payload Mass.



# Launch Success Yearly Trend

- As show by the chart, we notice an incline in the success trend from year 2013 and 2017.
- Shows that there was a slight decline in success between 2017-2018.



# All Launch Site Names

- There are 4 Unique Launch Sites

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

# Launch Site Names Begin with 'CCA'

- 5 records where launch sites beginning 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

- Calculate the total payload carried by boosters from NASA
  - The Total payload mass carried by boosters

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
    - Average payload mass carried by booster version F9 v1.1 is 2,928.

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - The first successful landing outcome on the ground pad was (12-22-2015)

# Successful Drone Ship Landing with Payload between 4000 and 6000

- 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

- Calculate the total number of successful and failure mission outcomes
  - We notice that success rate is high, with only one failure

| mission_outcome | COUNT |
|---|---|
| Failure (in flight) | 1 |
| Success | 99 |
| 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 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

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

| landing__outcome | booster_version | launch_site |
|---|---|---|
| Failure (drone ship) | F9 v1.1 B1012 | CCAFS LC-40 |
| Failure (drone ship) | F9 v1.1 B1015 | CCAFS LC-40 |

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

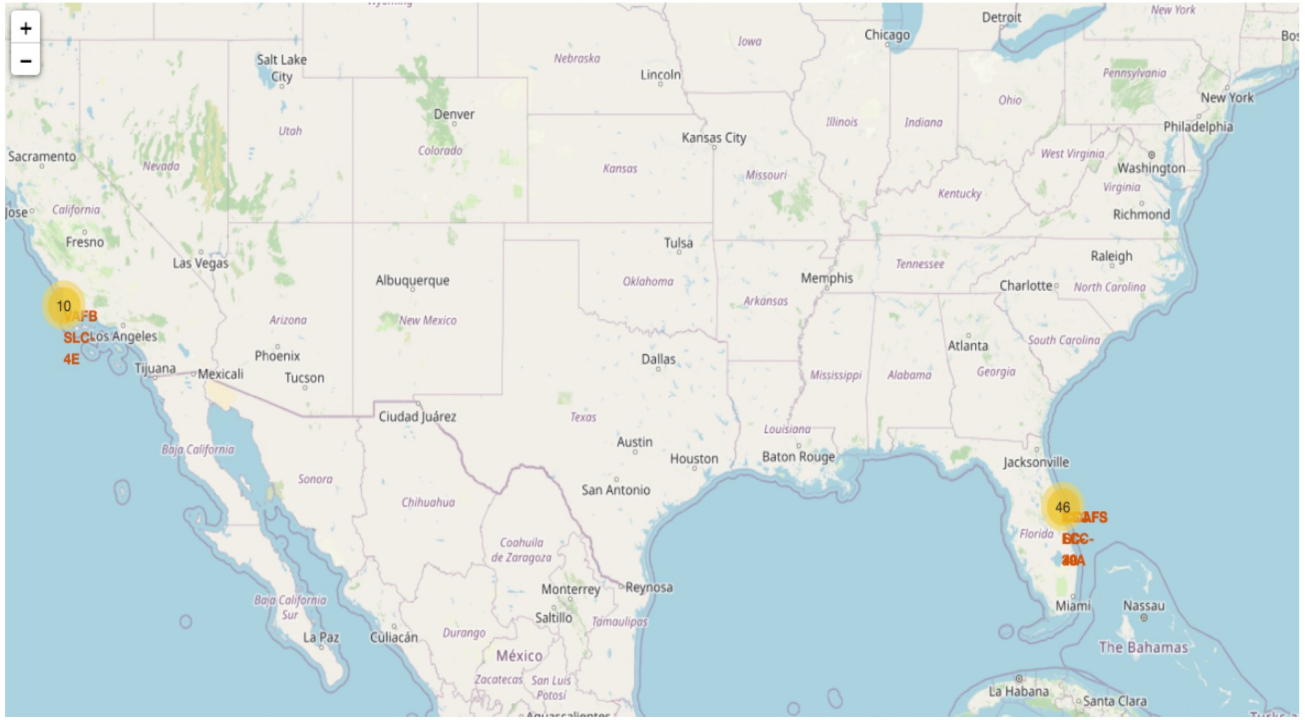
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

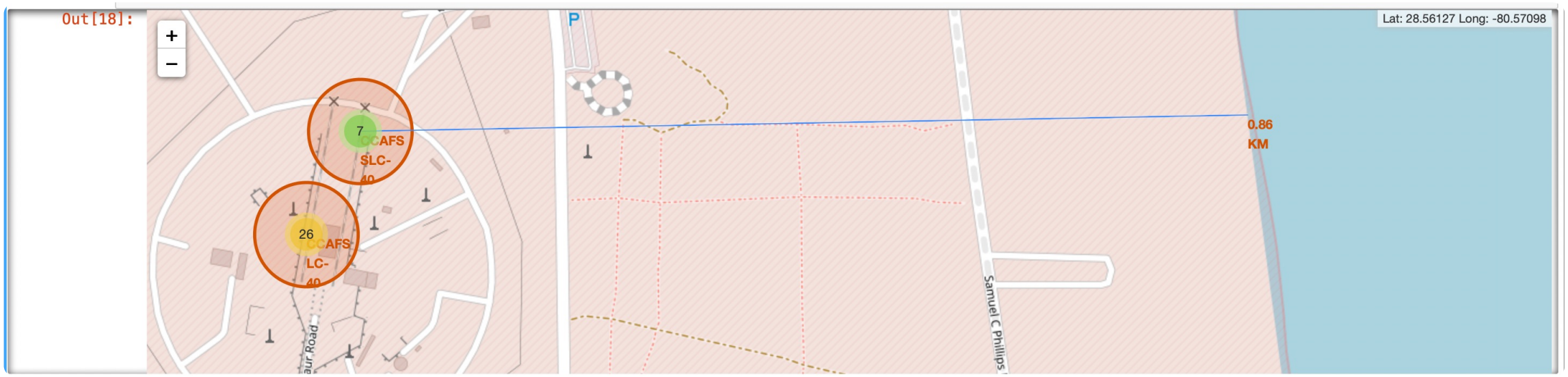
| landing__outcome | 2 |
|---|---|
| No attempt | 10 |
| Failure (drone ship) | 5 |
| Success (drone ship) | 5 |
| Controlled (ocean) | 3 |
| Success (ground pad) | 3 |
| Failure (parachute) | 2 |
| Uncontrolled (ocean) | 2 |
| Precluded (drone ship) | 1 |


Section 3
Launch Sites
Proximities Analysis

# Launch Site Locations

- SpaceX Has launch sites on two sides of the country. One located on the west coast, and multiple in the southeast of the country in Florida.

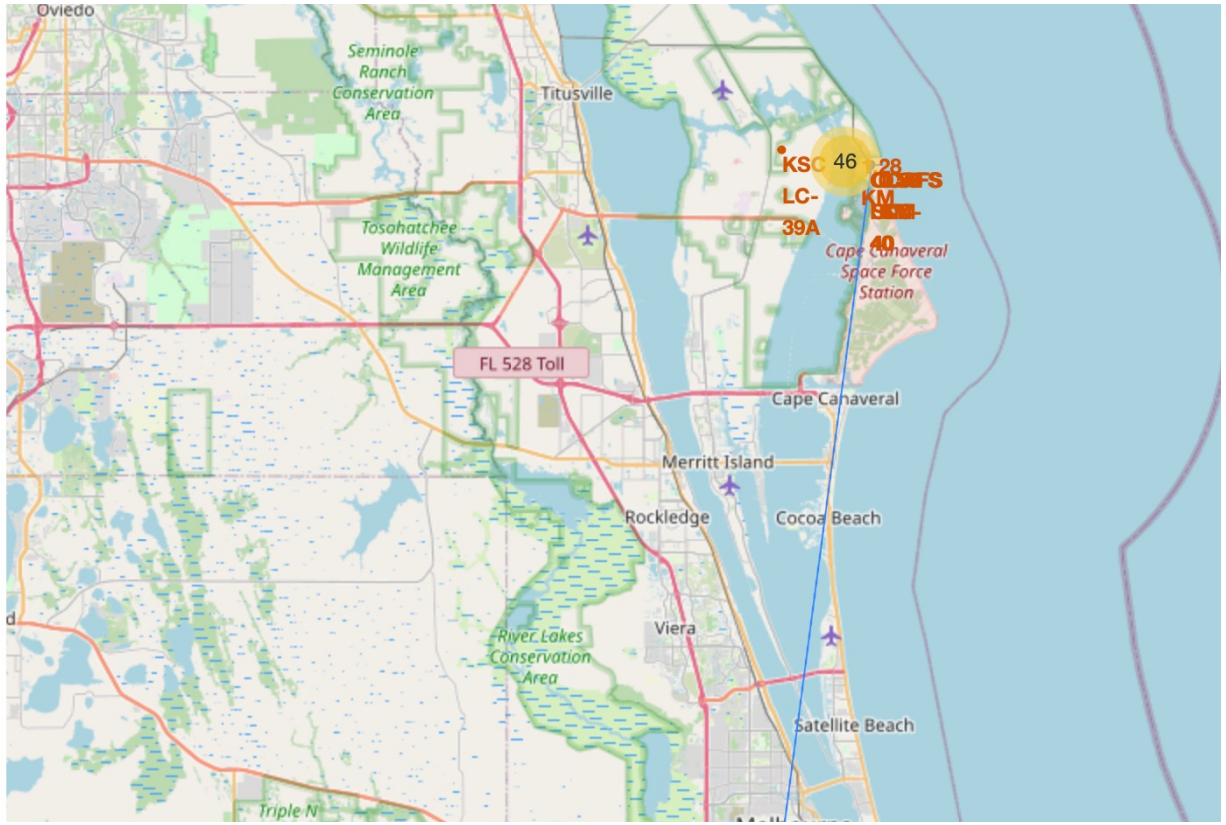


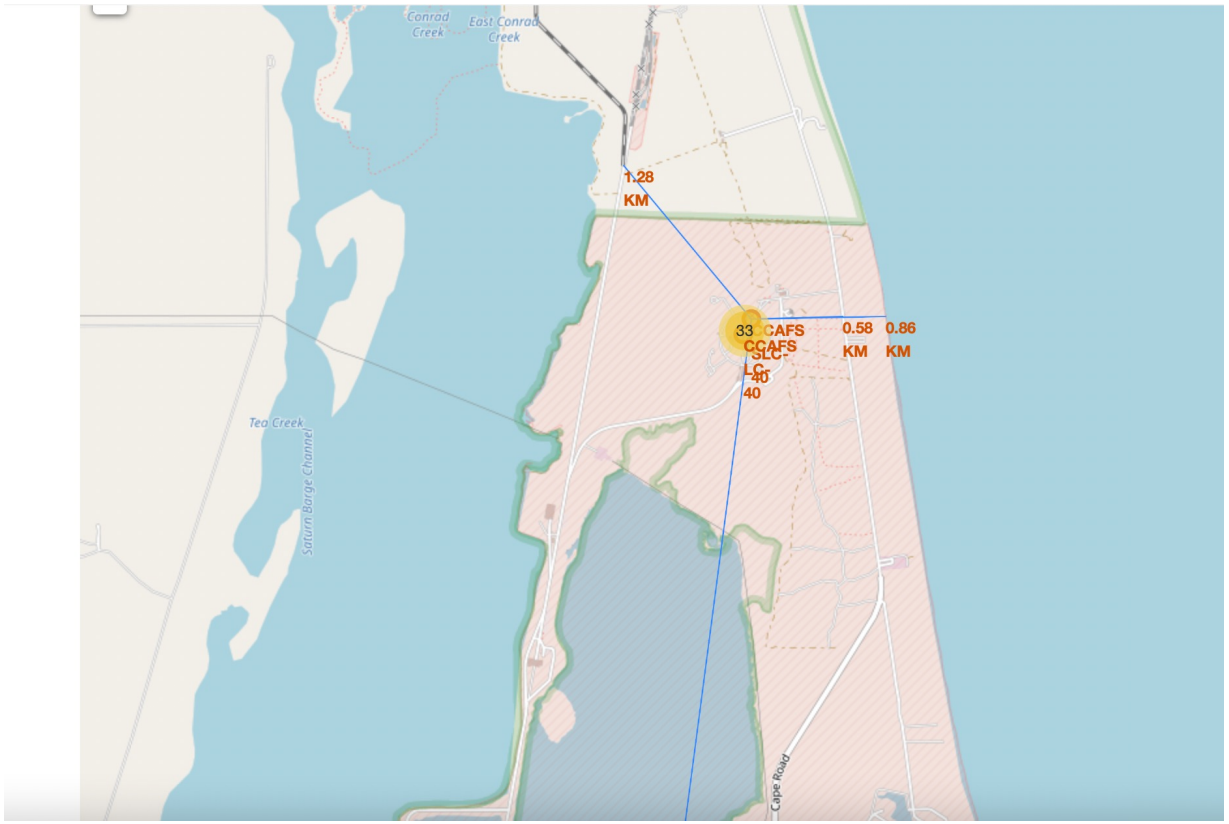


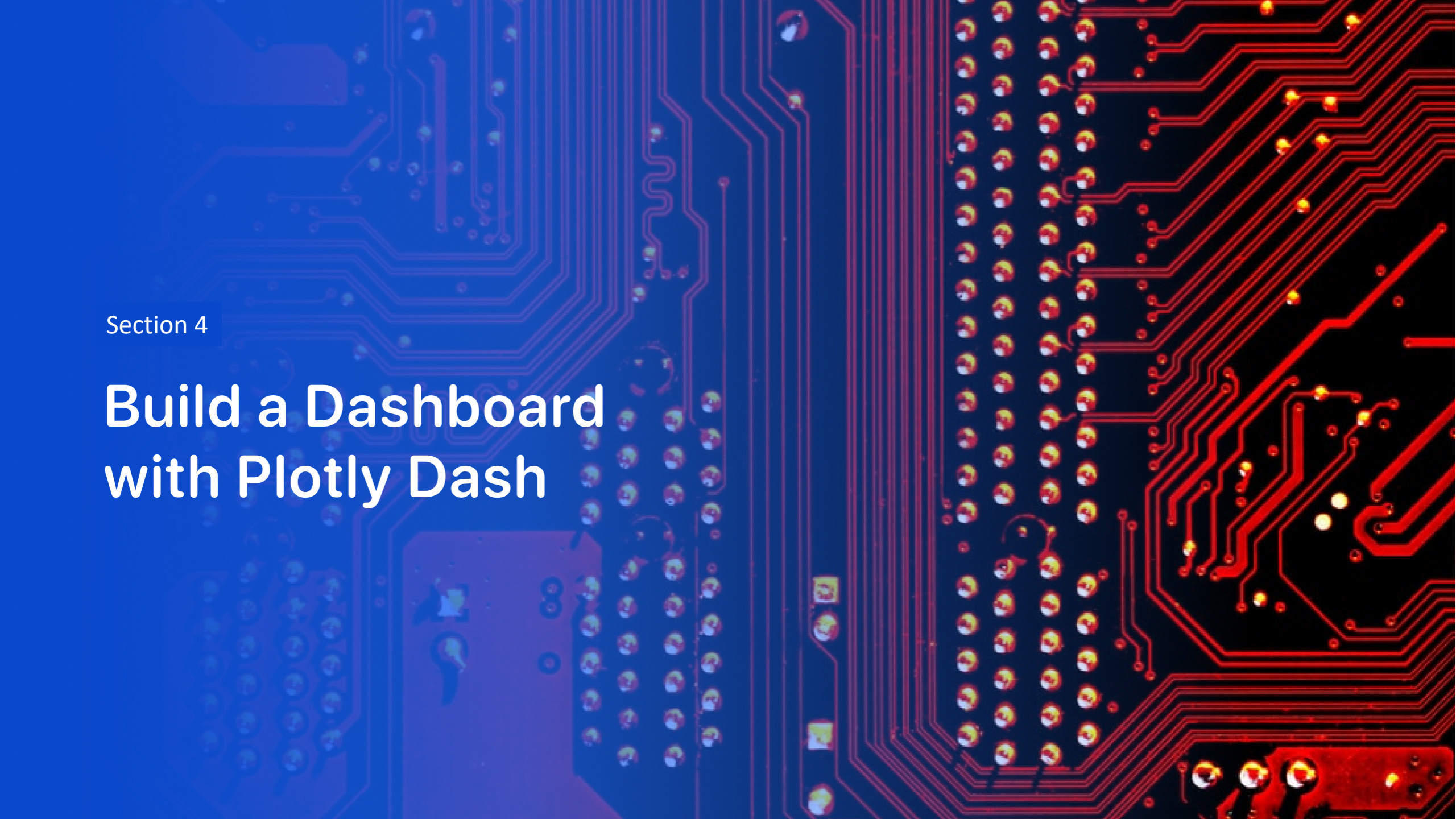

# Launch Sites

- Blue line represents the distance to the nearest:
  - Coastline
  - City/Town
  - Railway
  - Highway

# Launch Sites Cont.

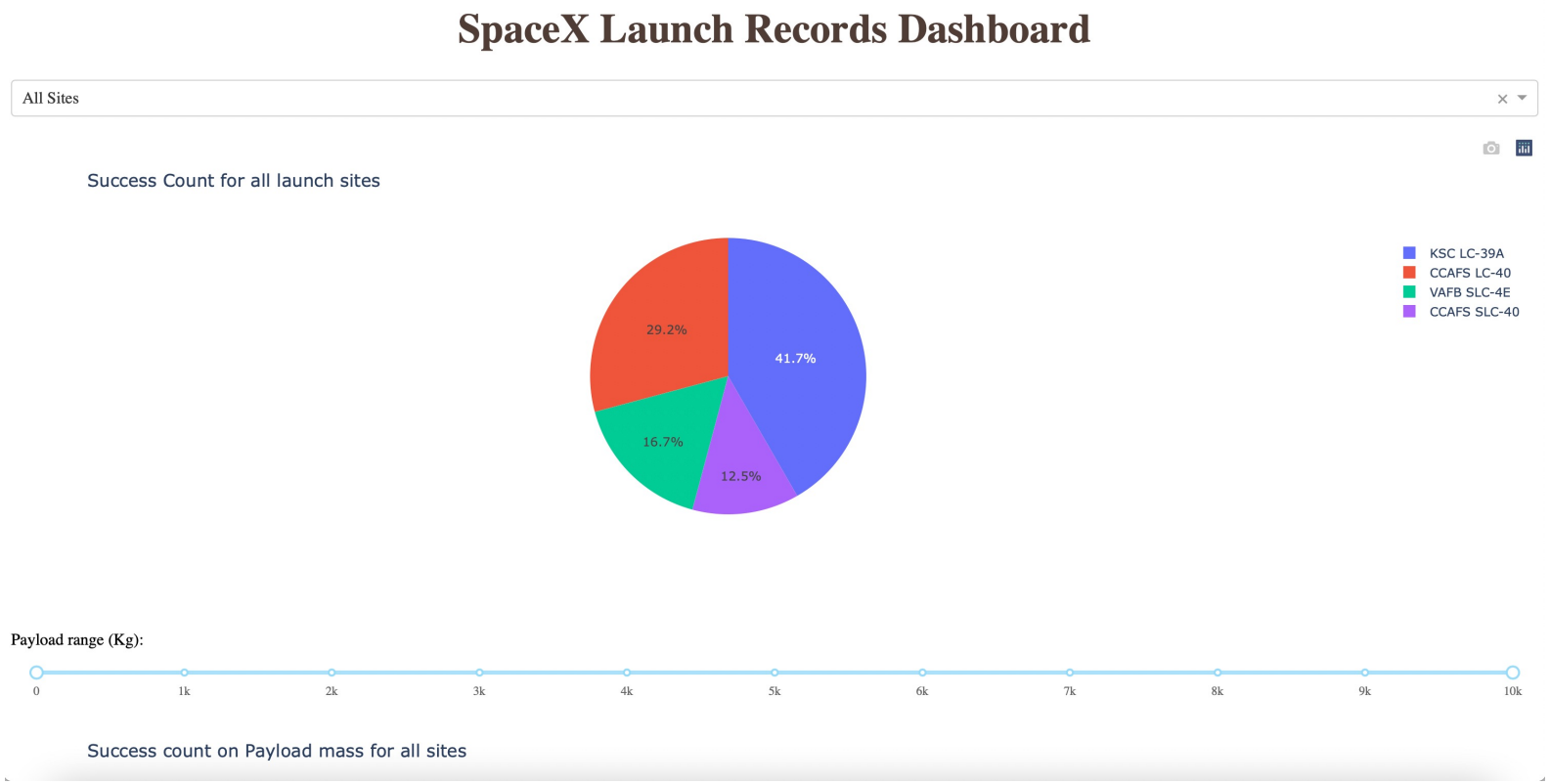



Section 4
Build a Dashboard with Plotly Dash

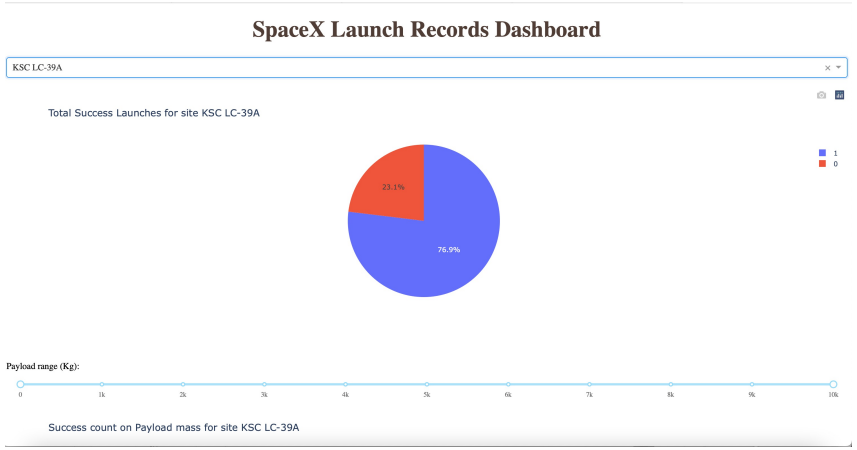
# SpaceX Launch Sites Success

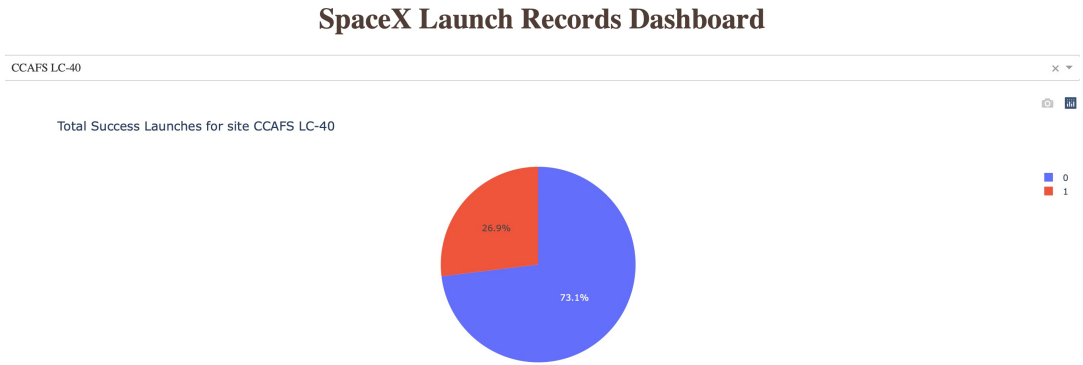
KSC LC-39A has the most successful Launches



# Determine the Most Successful Launch

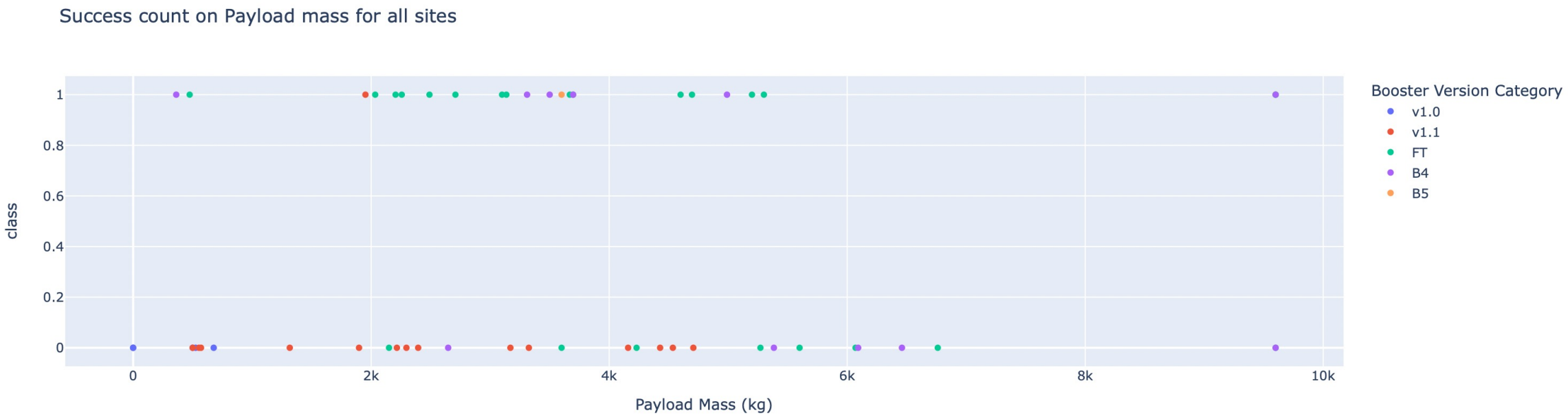
- CCAFS LC-40, conducted the most launches, and has a high success rate, in respects to the number of launches conducted.
- The Success Rate for KSC LC-39A, conducted fewer launches but have a higher success.



# Payload Mass Vs. Class

- The most successful payload mass range appears to be between 2K and 4K kg.
- FT shows to be the most successful booster .
- From results we noticed that V1.1 Failed the most With various Payload mass.



Section 5

# Predictive Analysis (Classification)

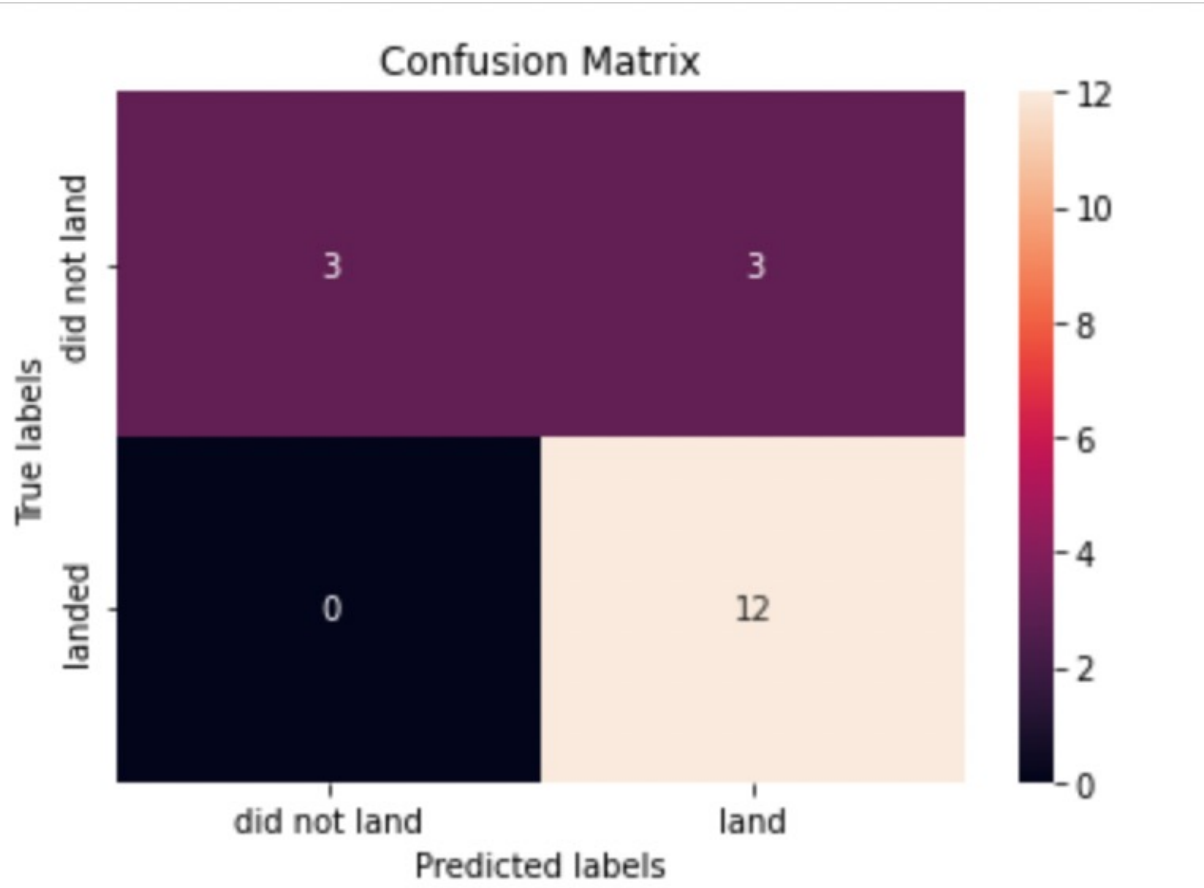
# Classification Accuracy

- As the Bar plot displays, three of the model's accuracy score is around 83.33%.

- However, It shows that the Decision Tree is slightly less accurate with only 77.78%

# Confusion Matrix

- We notice that within the Confusion Matrix, there were many false positives.
- There were 3 landings that were predicted as successful when they were



# Conclusions

- The Logistic Regression, KNN, and SVM models, shown to be the most accurate for the dataset; based on Prediction methods.
- The Success of launches have increased throughout the years, starting from 2015.
- KSC LC-39A is considered the most successful out of all 4 launches. However, in terms of success with the number of launches conducted CCAFS LC-40 is close second.
- Payloads with a mass of 2000-4000 kg considered to be more successful.
- The most Successful Orbits are ES-L1, GEO, HEO and SSO.

# Thank you!