
# Winning Space Race with Data Science

Cristina Dragomir
30.06.2022

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data Collection using Web Scraping and API calls
  - Exploratory Data Analysis (EDA) – data wrangling, data visualizations, interactive dashboards
  - Machine Learning Prediction
- Summary of all results
  - Data collection using public sources
  - Data wrangling and feature engineering
  - Trained and Tested multiple Machine Learning Models

# Introduction

- Project background and context:
  - The Objective of the current Project was to identify the possibility for our company, Space Y, to compete with the well-known Space X company
- Problems we managed to identify answers to:
  - Predict the successful landing of the first stage of the rockets
  - Identify best place to launch that could ensure successful landing

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Public API from SpaceX
  - WebScraping from Wikipedia
- Perform data wrangling
  - Data was analyzed, enriched and labelled
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - 4 Machine Learning Models were trained, and their best parameters identified; the models have been tested and accuracy has been established for each of the models

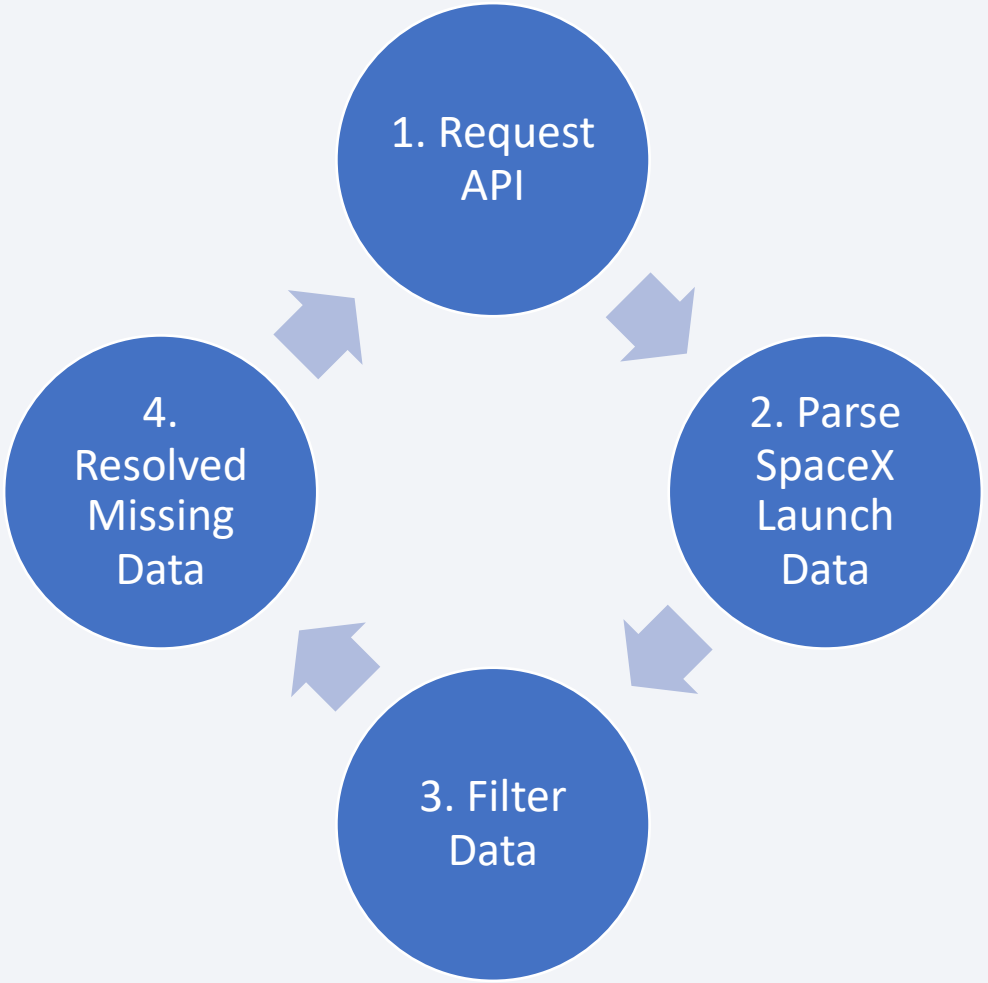
# Data Collection

Two Public Sources for Data Collection have been identified as follows:

1. https://api.spacexdata.com/v4 from where multiple information has been extracted and added into a Pandas DataFrame
2. https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches from where we have perfomed WebScraping in order to get the Falcon 9 historical launch records

# Data Collection – SpaceX API

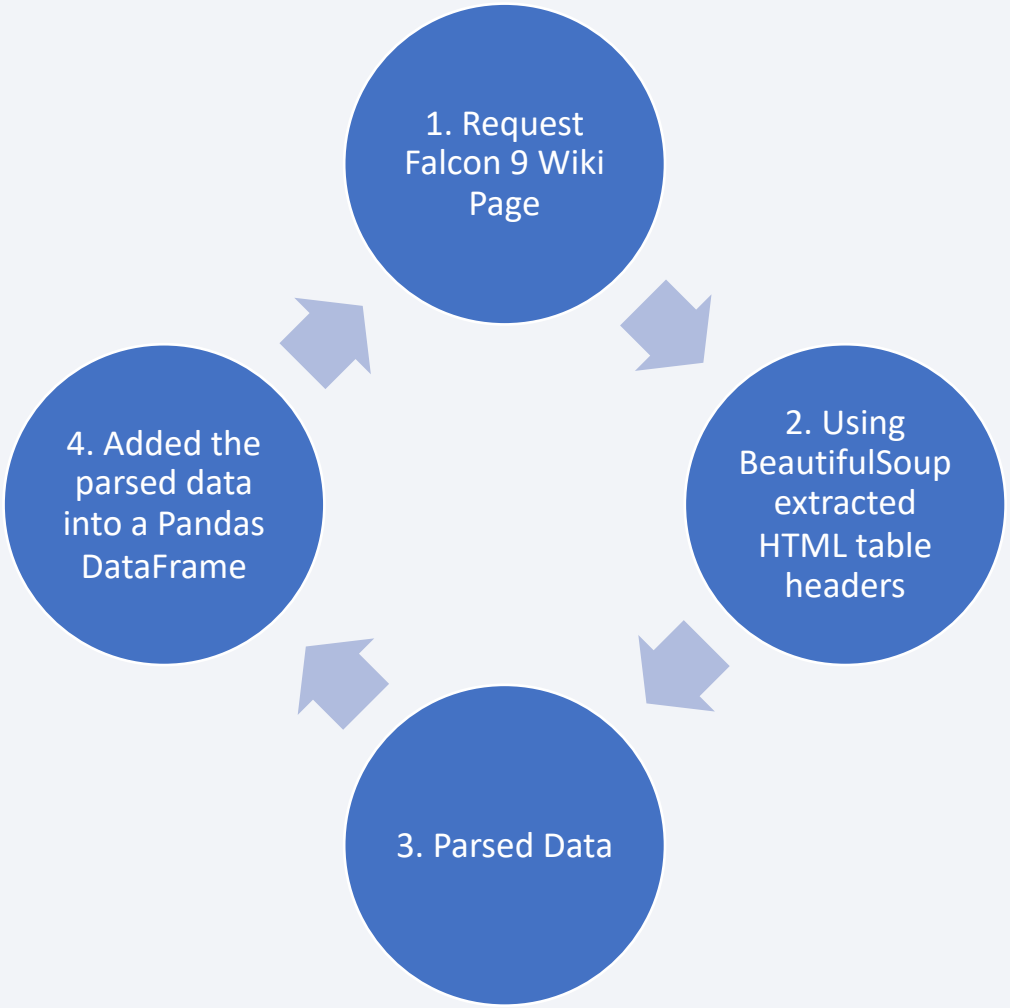
https://api.spacexdata.com/v4 from where multiple information has been extracted and added into a Pandas DataFrame




- For detailed data collection code, please click here

# Data Collection - Scraping

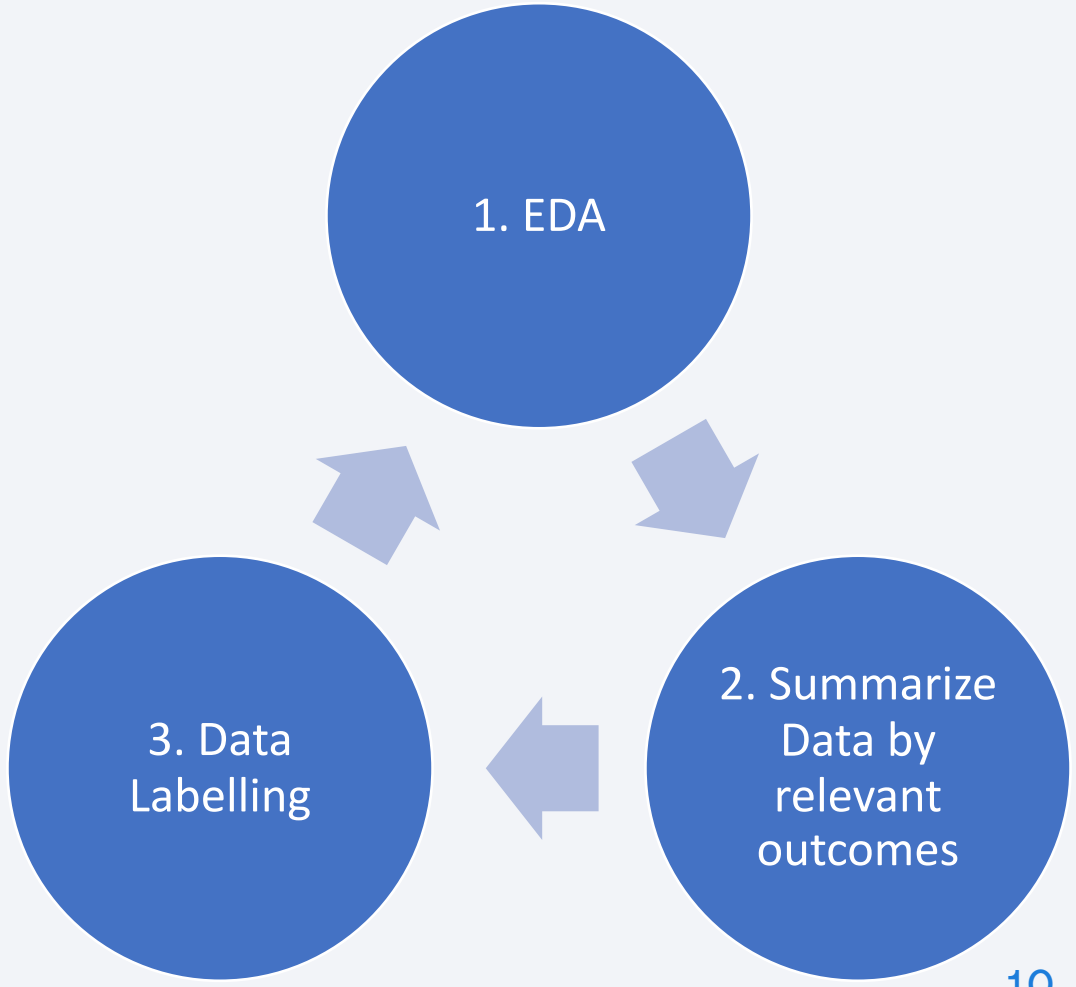
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches from where we have perfomed WebScraping in order to get the Falcon 9 historical launch records
- For detailed WebSraping code, please click here



# Data Wrangling

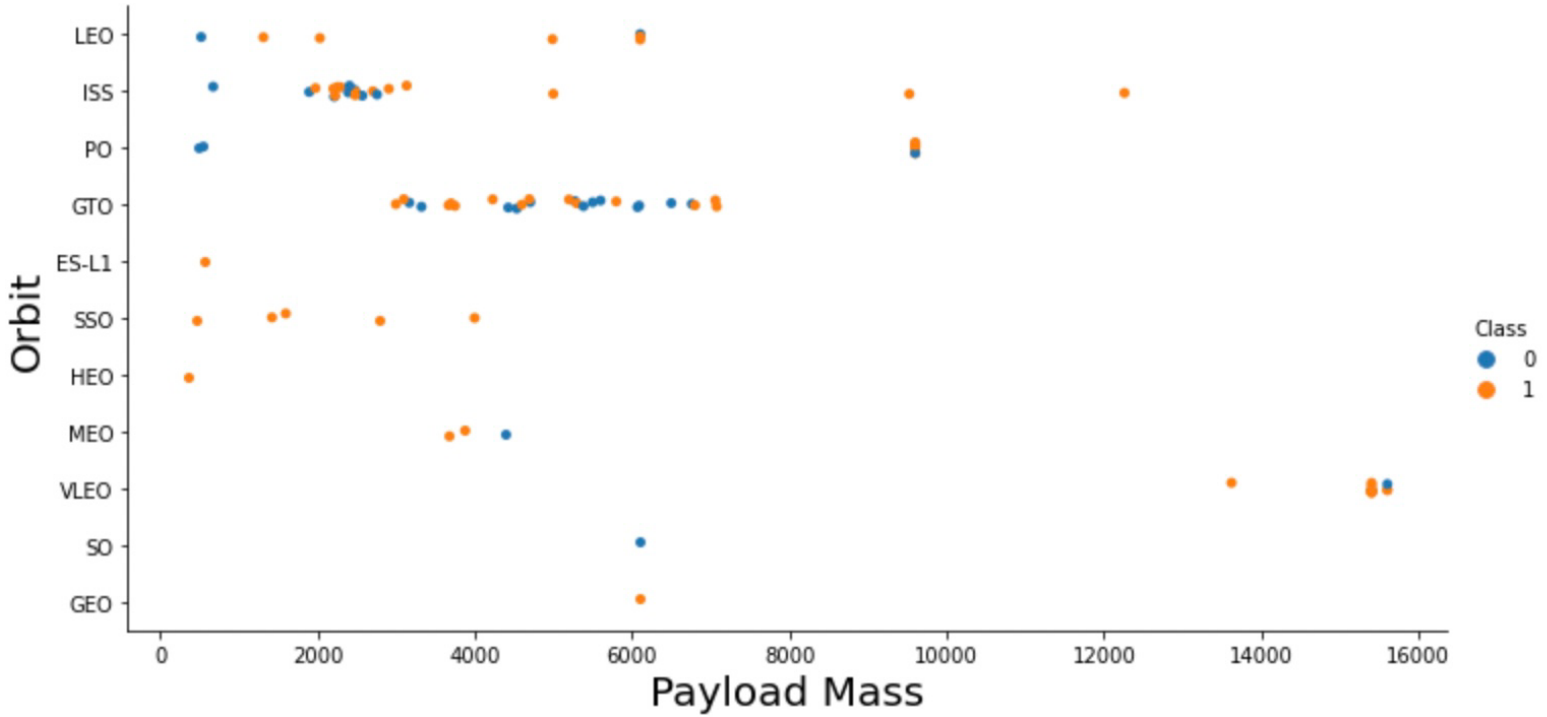
- After performing the EDA on the data, the following steps have been performed:
  - Summarize the total launches per Site
  - Occurrences of each Orbit
  - Mission Outcomes by Orbit Type
  - Outcome Label has been added

For full data wrangling code, please click here



# EDA with Data Visualization

- In order to perform exploratory data analysis, we have plotted bar plots as well as scatterplots that enabled an overview of the relationships between the different pairs of features, like Payload Mass vs Flight Number, Launch Site vs Payload Mass, Payload by Orbit etc




- [For full EDA with Data Vizualization code, please click here]

# EDA with SQL

- EDA has been performed using also a suite of SQL queries as follows:
  - **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 succesful landing outcome in ground pad was acheived.**
  - **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. Use a subquery**
  - **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.**

- For the entire set of SQL queries, please click here

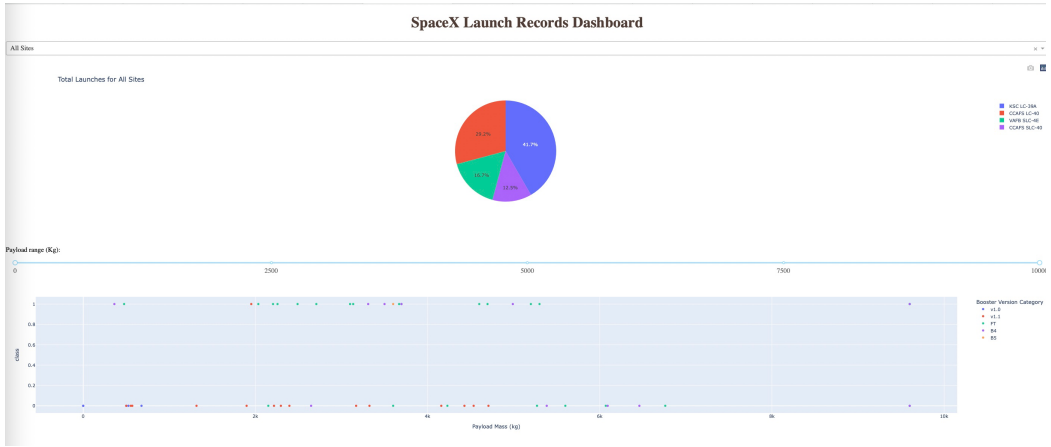
# Build an Interactive Map with Folium

- In order to build an Interactive Map we have used objects such as markers, circles, lines, as well as marker clusters
- Each of the added objects helped in identifying the following:
  - Using Markers we have indicated points like the Launch Sites
  - Using Circles we have highlighted the areas around specific coordinates
  - Using Marker Clusters we have indicated groups of events for each coordinate
  - Lines were used to indicate distances between important points and the Launch Sites

- For the interactive maps code and visualizations, please click here

# Build a Dashboard with Plotly Dash

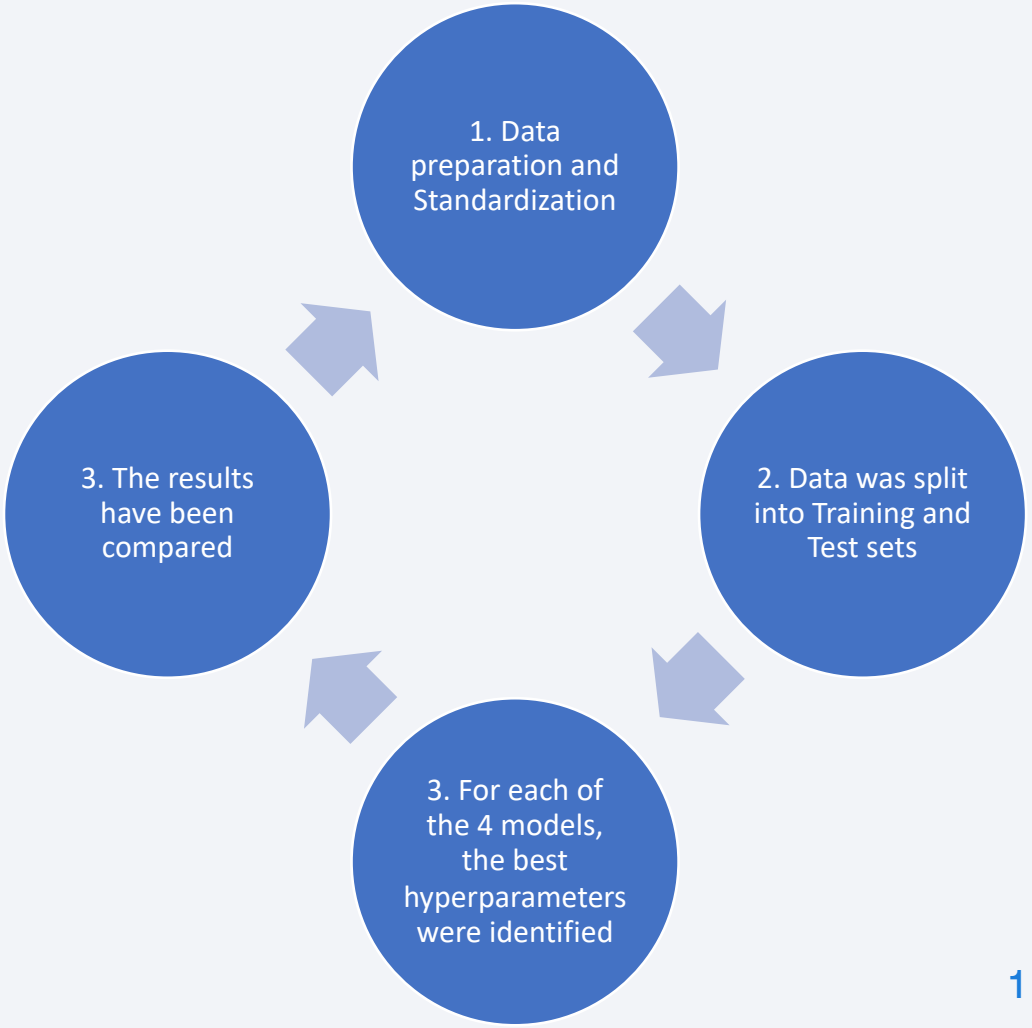
- The Interactive dashboard contains the following plots:
  - A PieChart that shows the total successful launches
  - A ScatterChart that shows the correlation between payload and launch success
  - A Slider to select the Payload range
- Adding these graphs helped quickly analyze the relationships between payloads and launch sites and consequently identify the best launch site
- To review the code, please click here



# Predictive Analysis (Classification)

- We have tested four classification models:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbors

- For the code, please click here



# Results

- Exploratory data analysis results
  - The SpaceX company uses 4 sites for the launches
  - Different launch sites have different success rates
  - Success rate increased starting with 2013 and it continued to increase until 2020
  - Launch Sites, Orbit and Payload Mass seem to be good predictors to the mission success

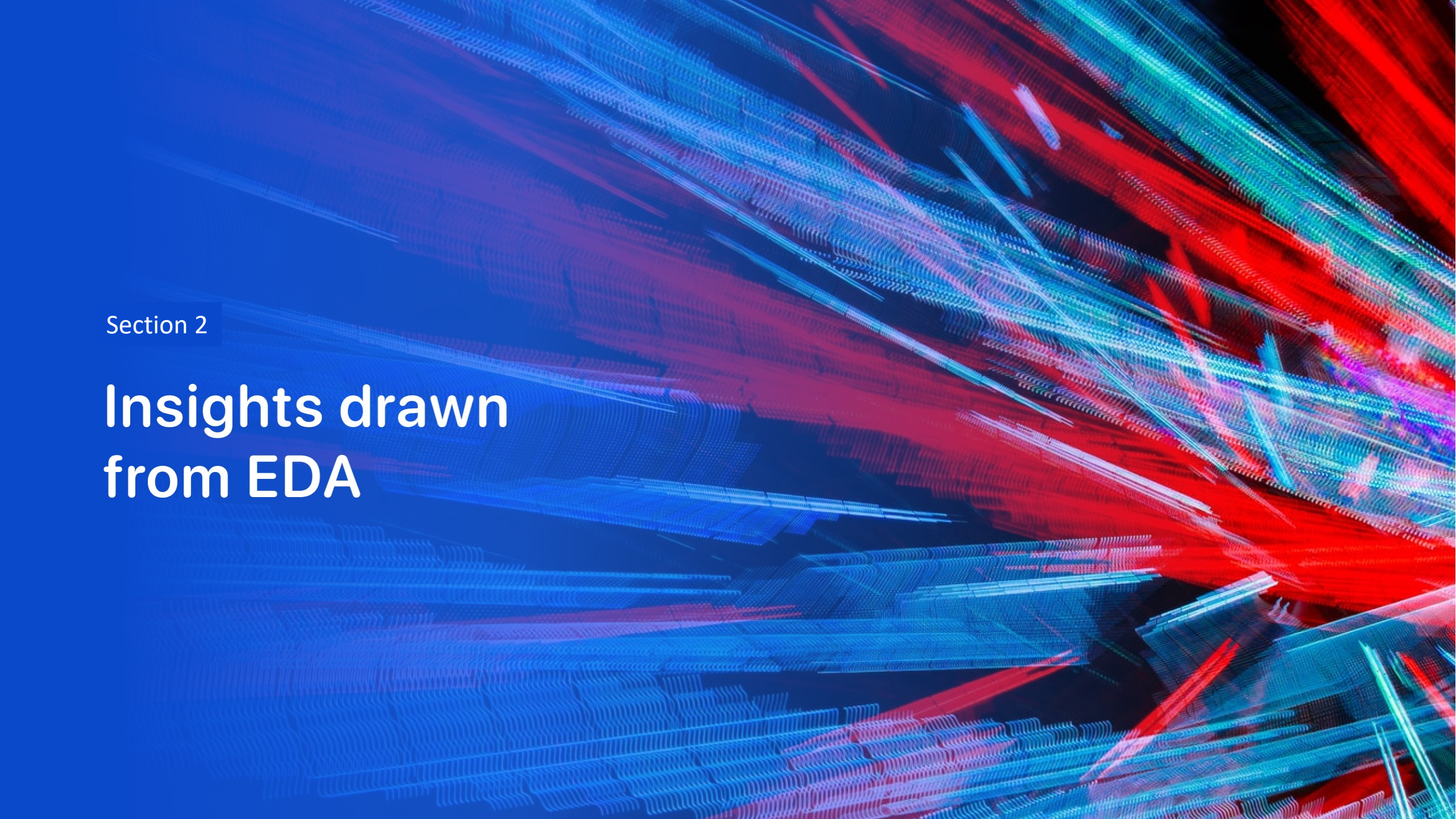


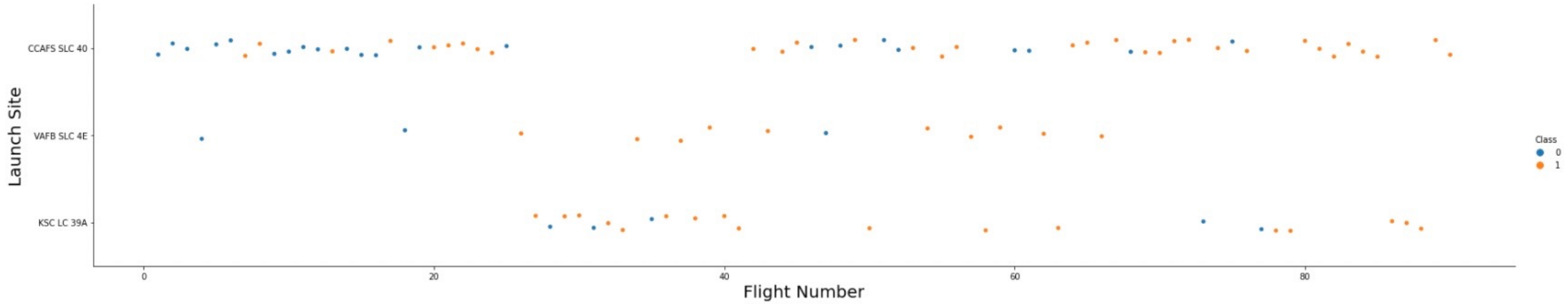
# Results

- Using the interactive analytics, we managed to understand that launch sites are usually placed near the Equator line and on the Coast
- Also, we understood that all sites should be in the railways, highways proximity and they need to have a certain distance from the cities

# Results

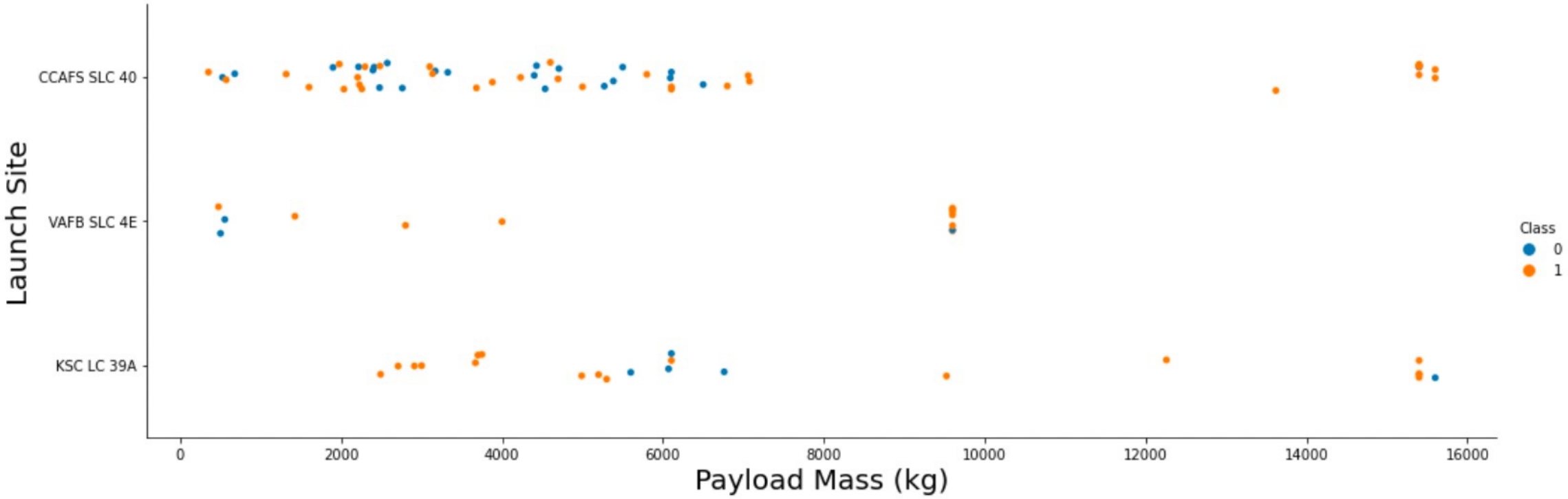
- Decision Tree Classifier showed the best results after the hyperparameters tunning with and accuracy score of 87.5% on the train set and 83% in the test set


Section 2
Insights drawn from EDA



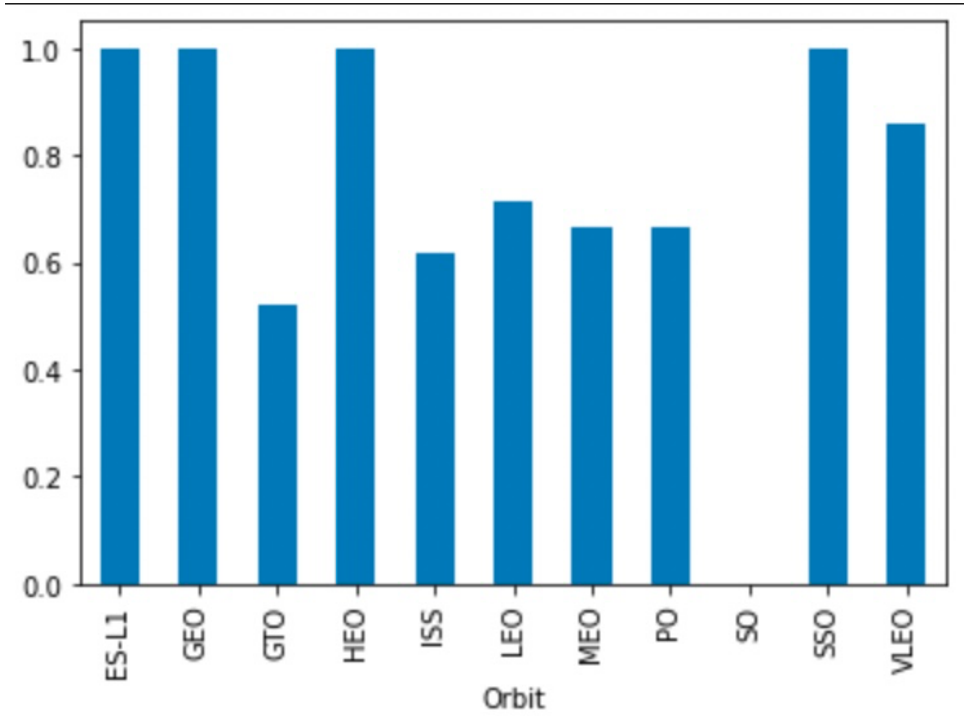

# Flight Number vs. Launch Site

- From the above scatterplot it may be observed that CCAF5 SLC 40 has the best success rate
- It can be observed that general success rate improved over time



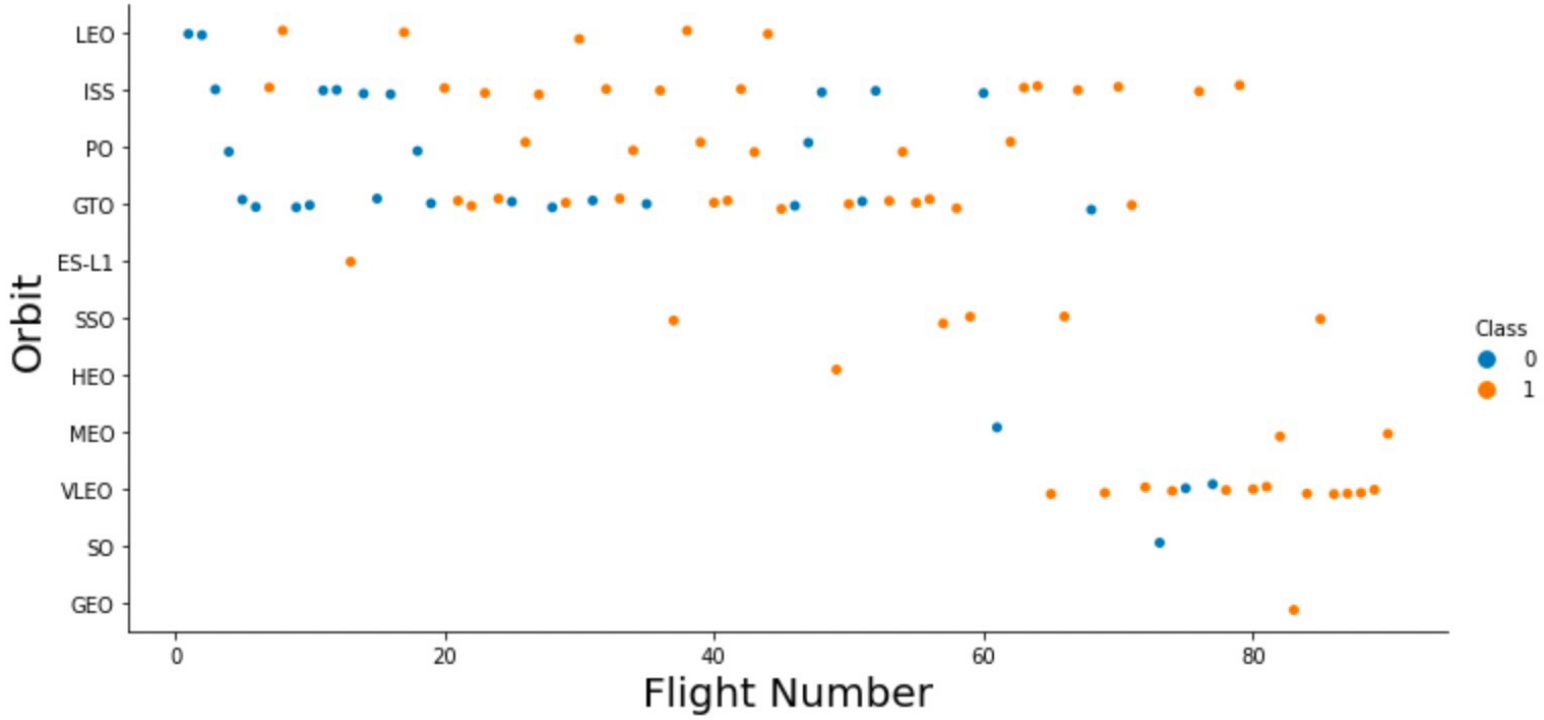

# Payload vs. Launch Site

- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000)
- Payloads over than 9000kg have good success rate
- There are only two sites from where massive payloads are being launched



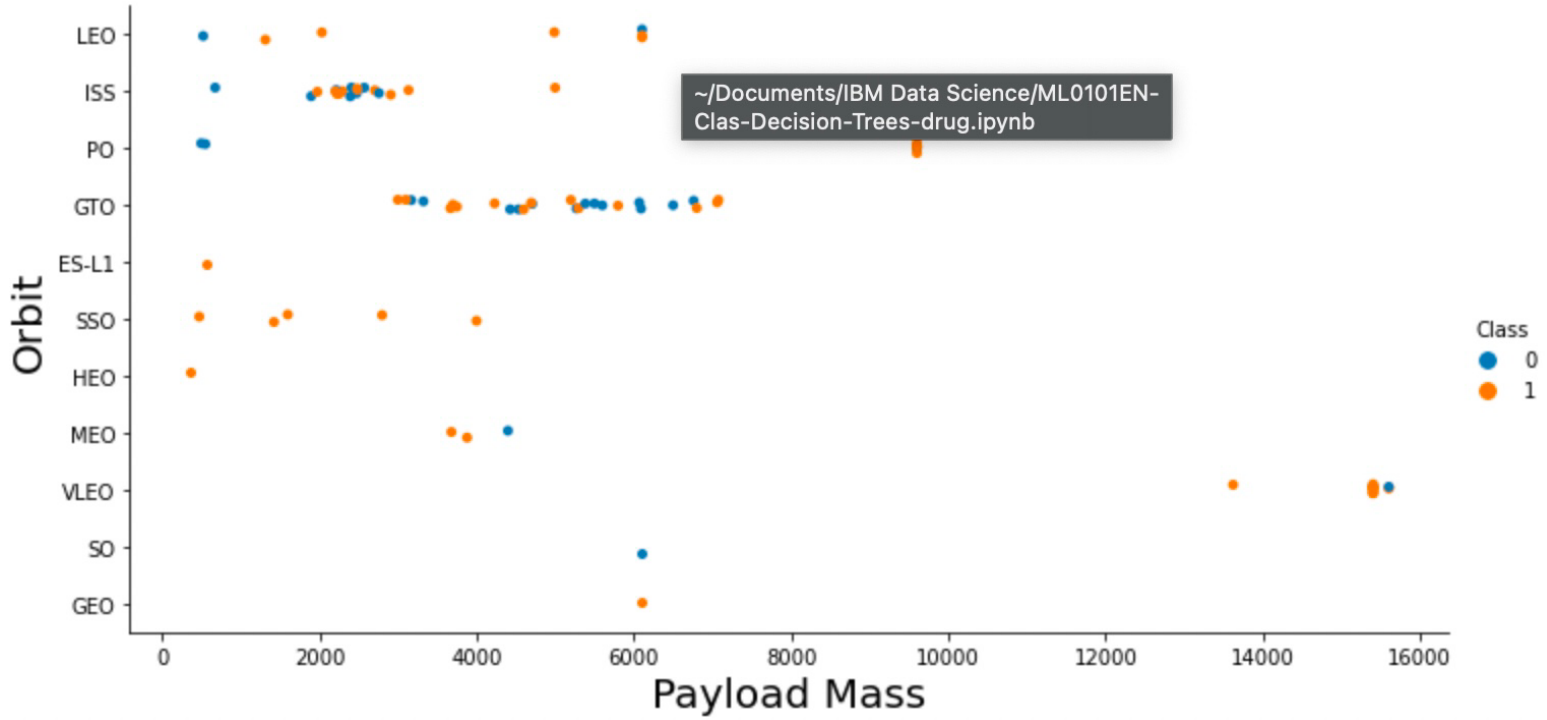

# Success Rate vs. Orbit Type

- There are 4 Orbits that have 100% (or very close) success rates



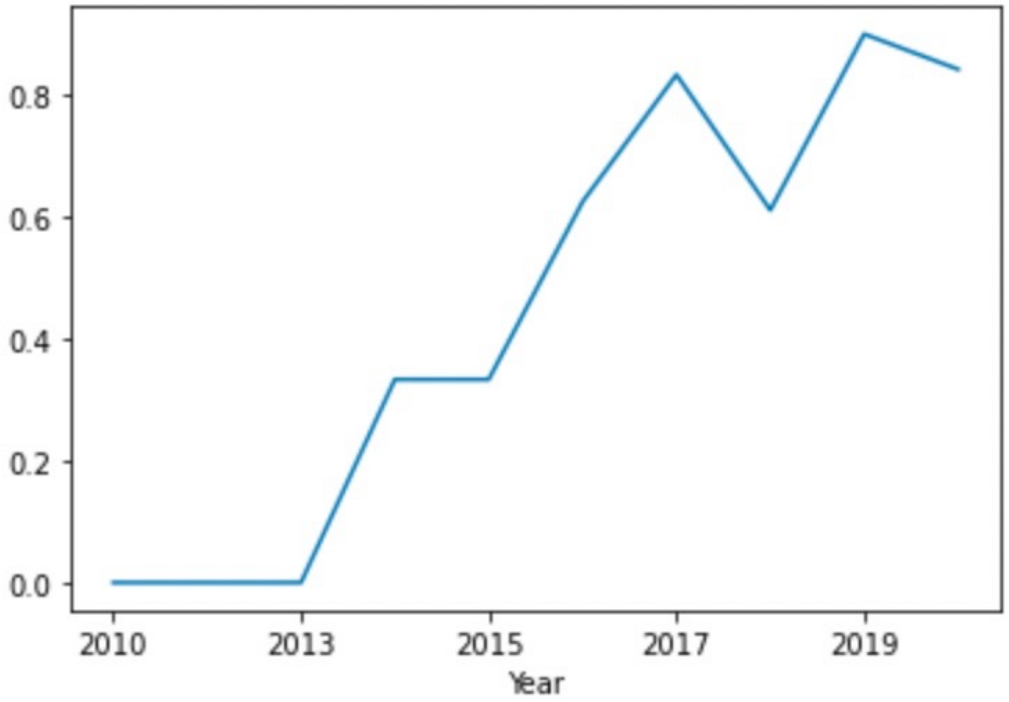

# Flight Number vs. Orbit Type

- Success rate seem to have been improved over time for all the orbits




# Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

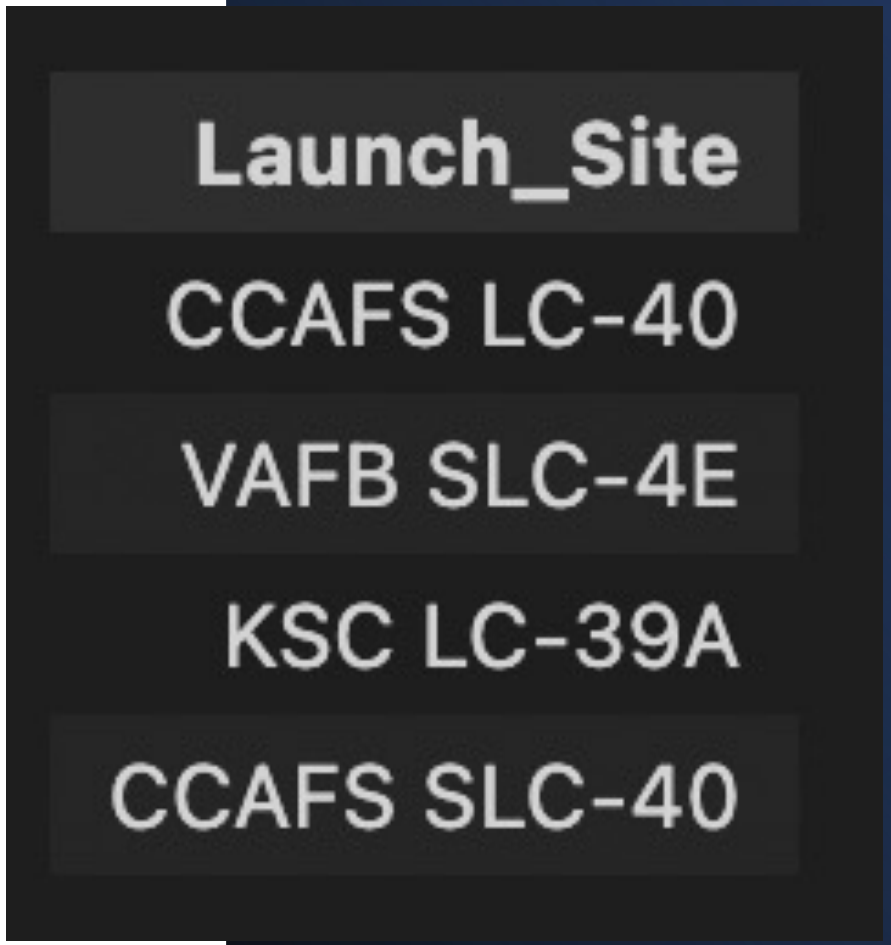



# Launch Success Yearly Trend

- It can be observed that the success rate since 2013 kept increasing till 2020

# All Launch Site Names

- In the SQL results we may observe that there are 4 launch site names

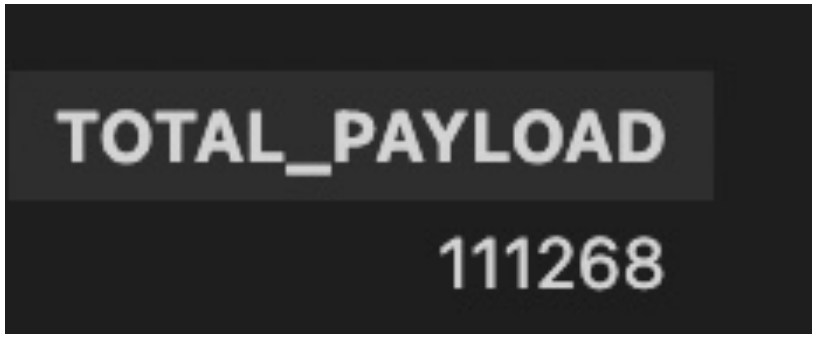


# Launch Site Names Begin with 'CCA'

- There are 5 site names that begin with CCA, please see them in the below screenshot “Launch_Site” column

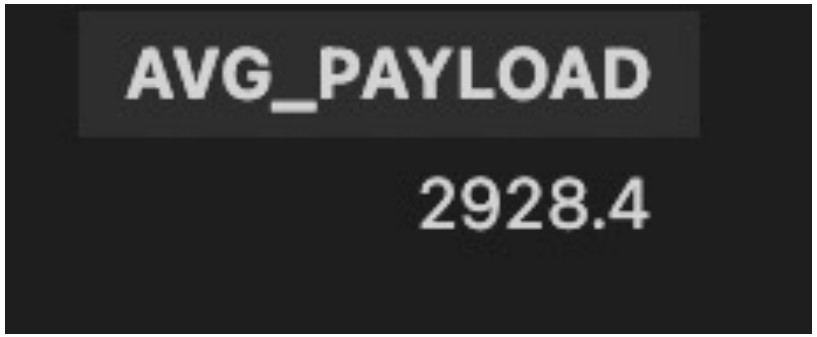
| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing _Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 08-12-2010 | 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) |
| 22-05-2012 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 08-10-2012 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 01-03-2013 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

# Total Payload Mass



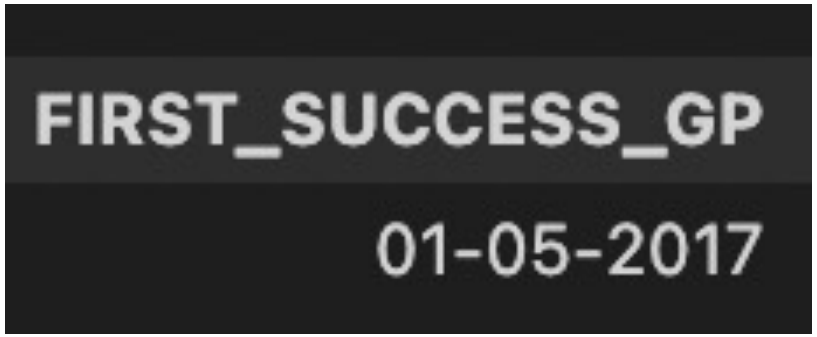

- The calculated total payload carried by boosters from NASA is 111,268

# Average Payload Mass by F9 v1.1



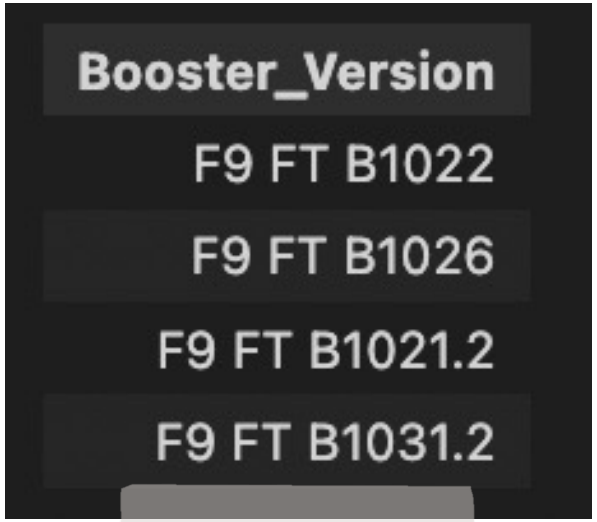

- The calculated the average payload mass carried by booster version F9 v1.1 is 2,928.4

# First Successful Ground Landing Date




- The first successful landing outcome on ground pad was on the 1st of May 2017

# Successful Drone Ship Landing with Payload between 4000 and 6000




- In the results screen shot you may observe the list of the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

# Total Number of Successful and Failure Mission Outcomes

- The screenshot presents the total number of Successful and Failure Mission Outcomes

| Mission_Outcome | QTY |
|---|---|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

- In the list, you may find the names of the boosters which have carried the maximum payload mass

| MONTH | 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 |

# 2015 Launch Records

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

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

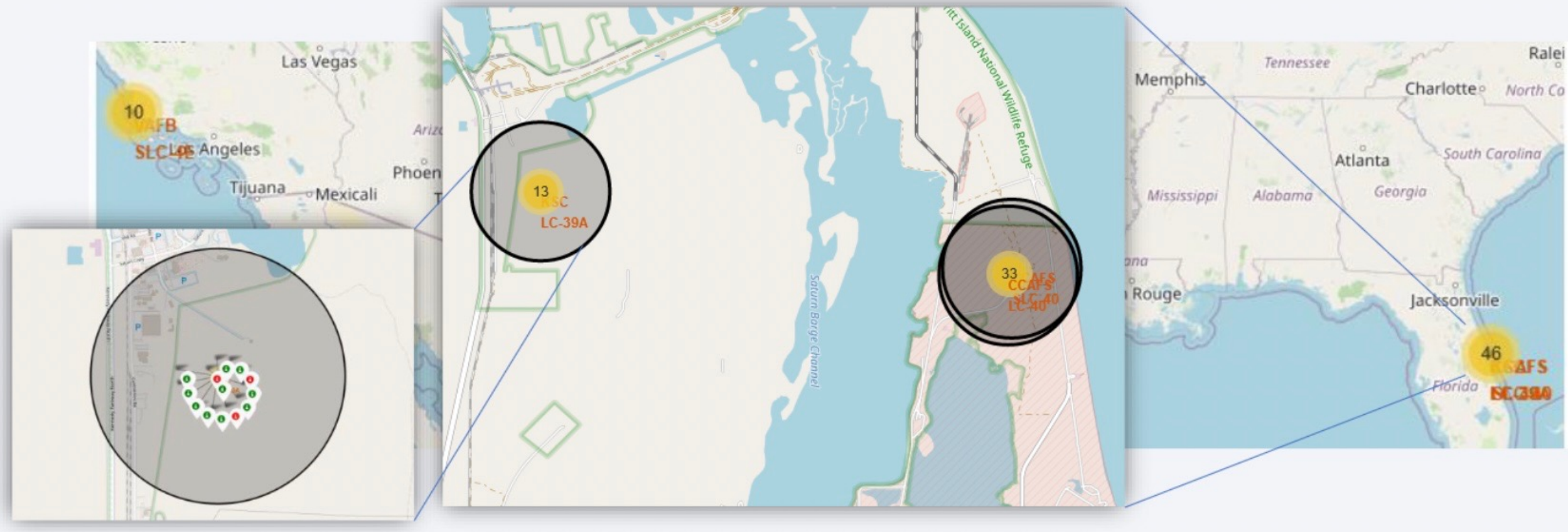
- You may see also the rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

| Landing _Outcome | QTY |
|---|---|
| Success | 20 |
| Success (drone ship) | 8 |
| Success (ground pad) | 6 |


Section 3
Launch Sites
Proximities Analysis

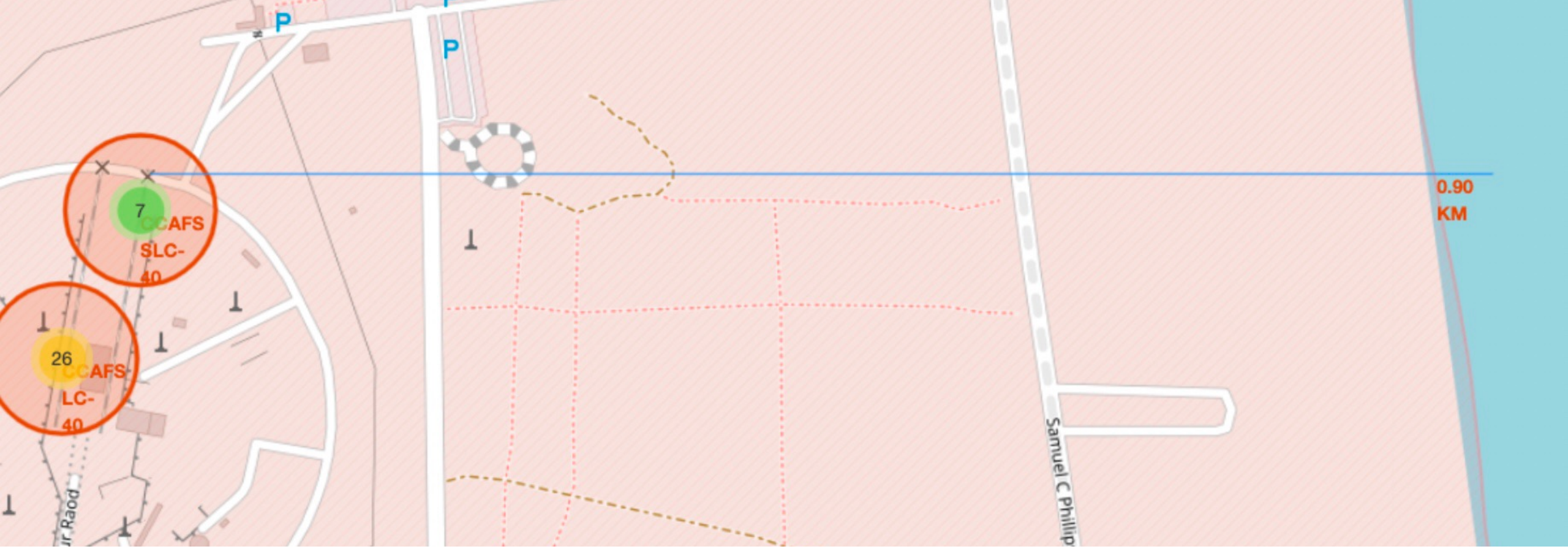
# All Launch Sites


- From the map, you may notice that all Launch Sites are on the coast



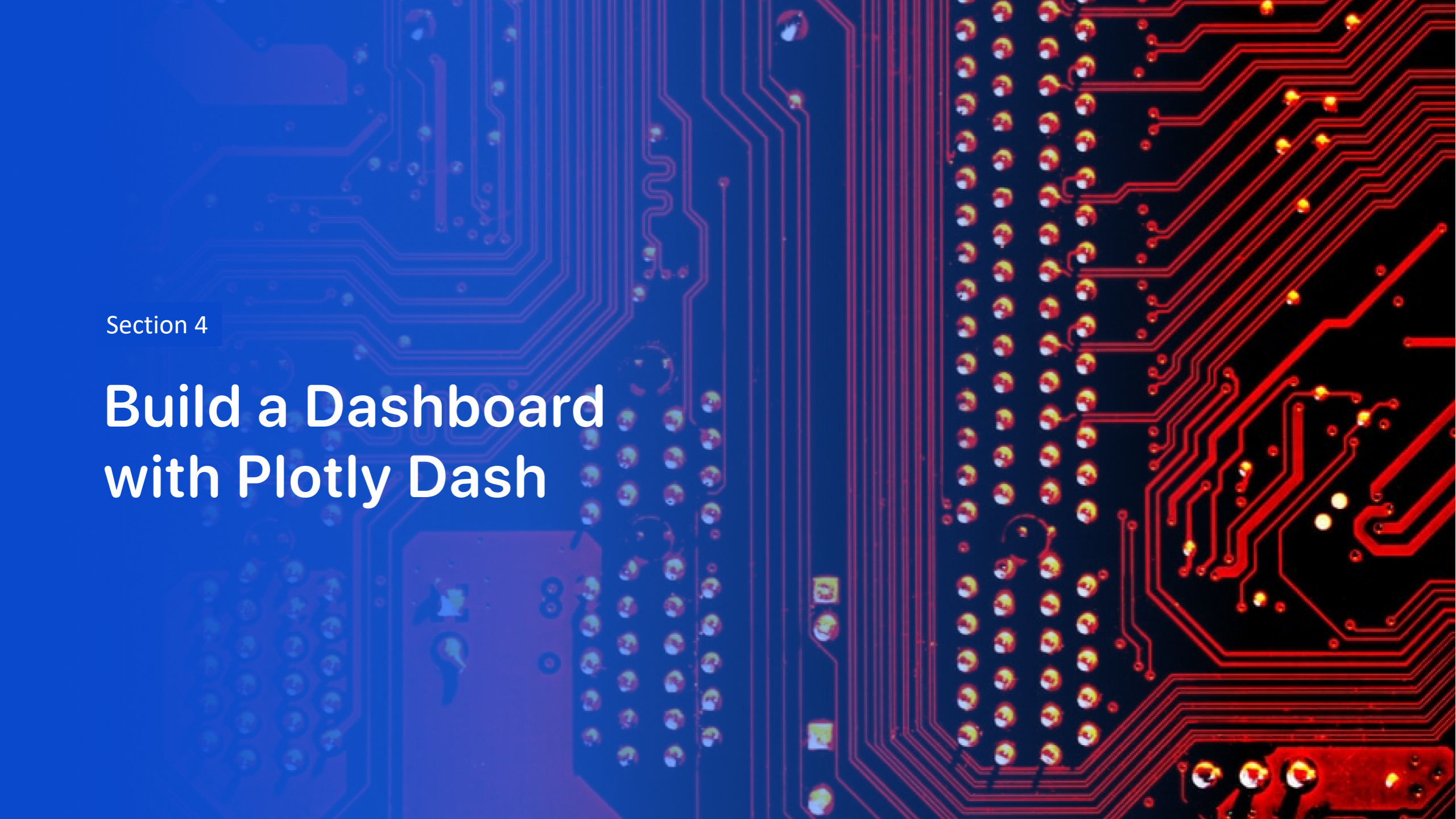

# Launch Outcomes by Site

In the screen shot you may notice that using the interactive map you can check the launch outcomes for each site; in our example the green dots are successful launches, and the red ones are unsuccessful



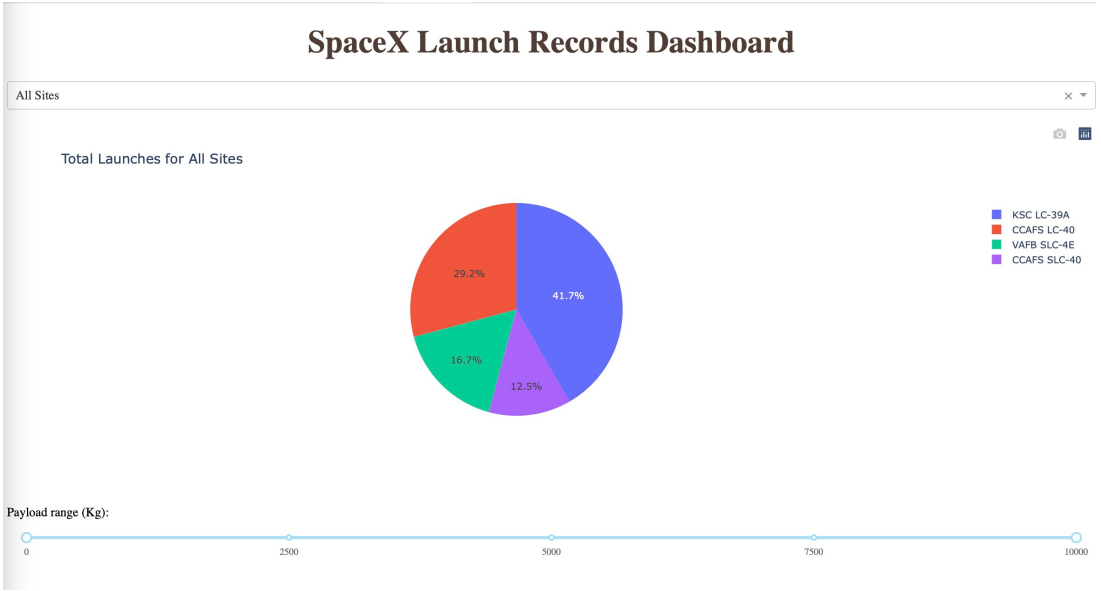

# Distances

- In the screen shot, the blue line represents the distance from the launch site to the sea

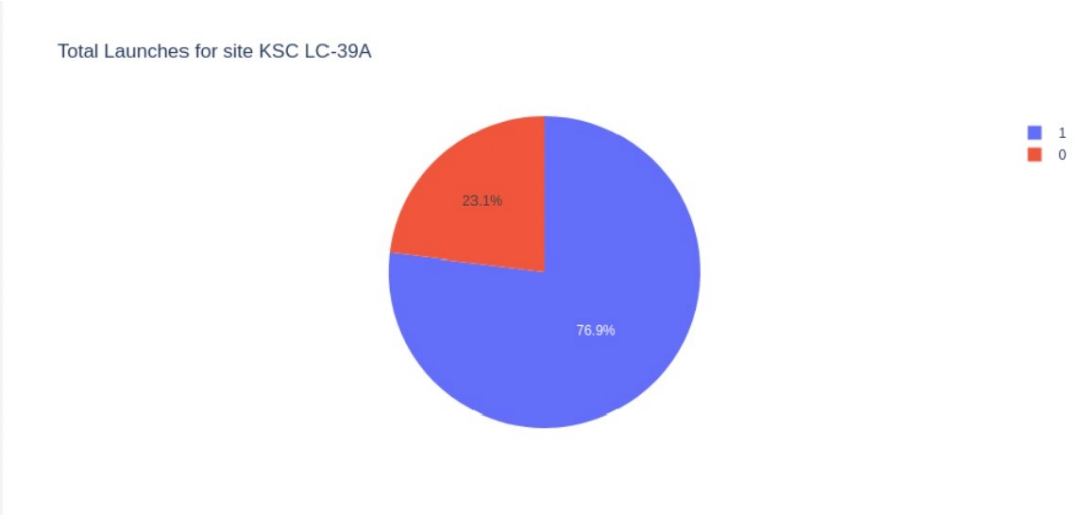

Section 4
Build a Dashboard with Plotly Dash

# Launch Success counts

As presented in the dashboard, you may notice that the launch site plays an important role in the mission success

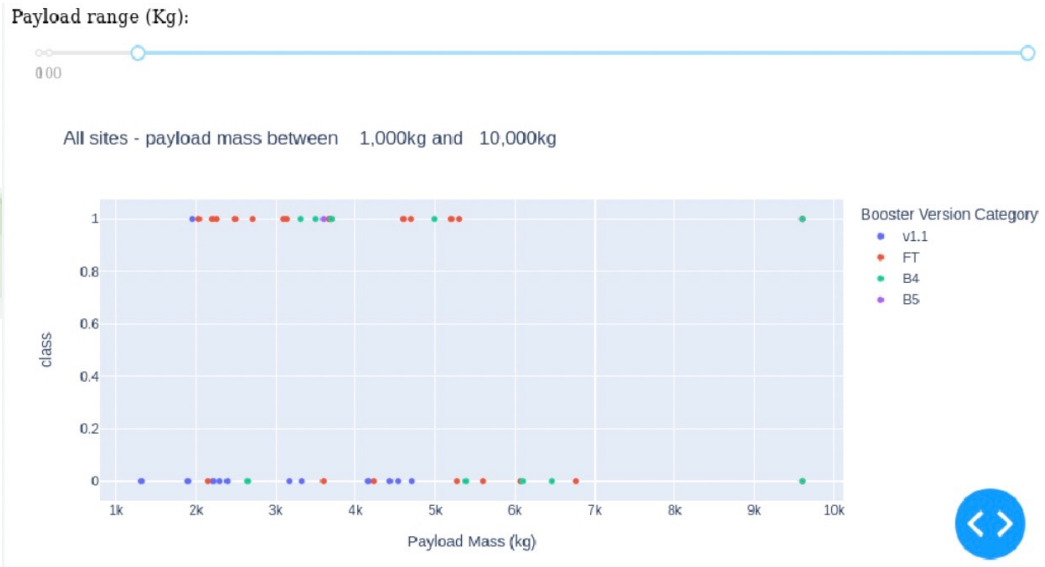


# Highest Launch Success Ratio




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

# Payload vs Launch Outcome

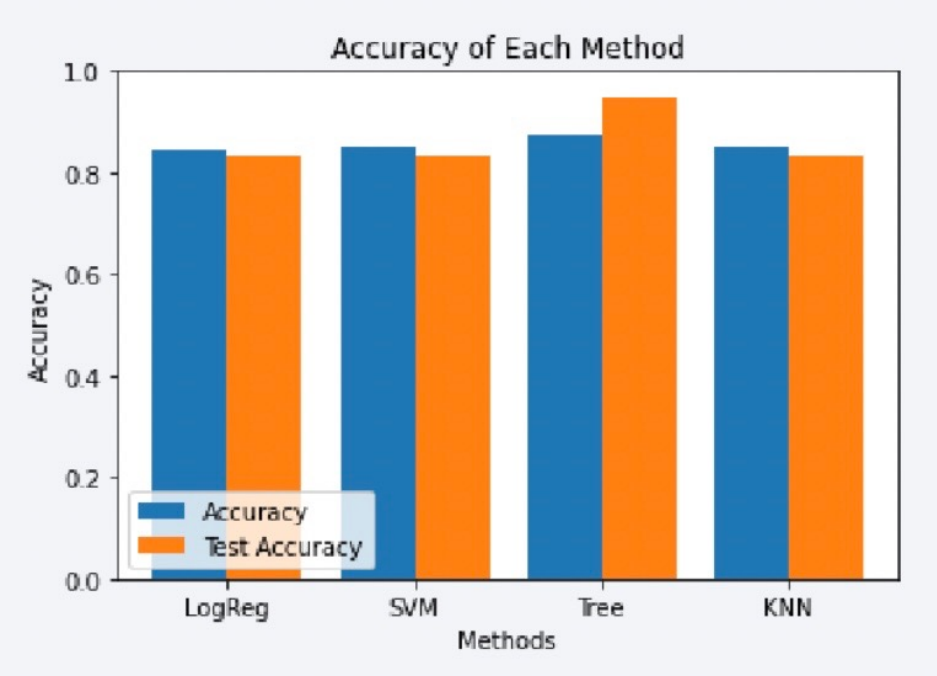



- Payloads under 6000 kg and FT boosters are the most successful combination

Section 5

# Predictive Analysis (Classification)

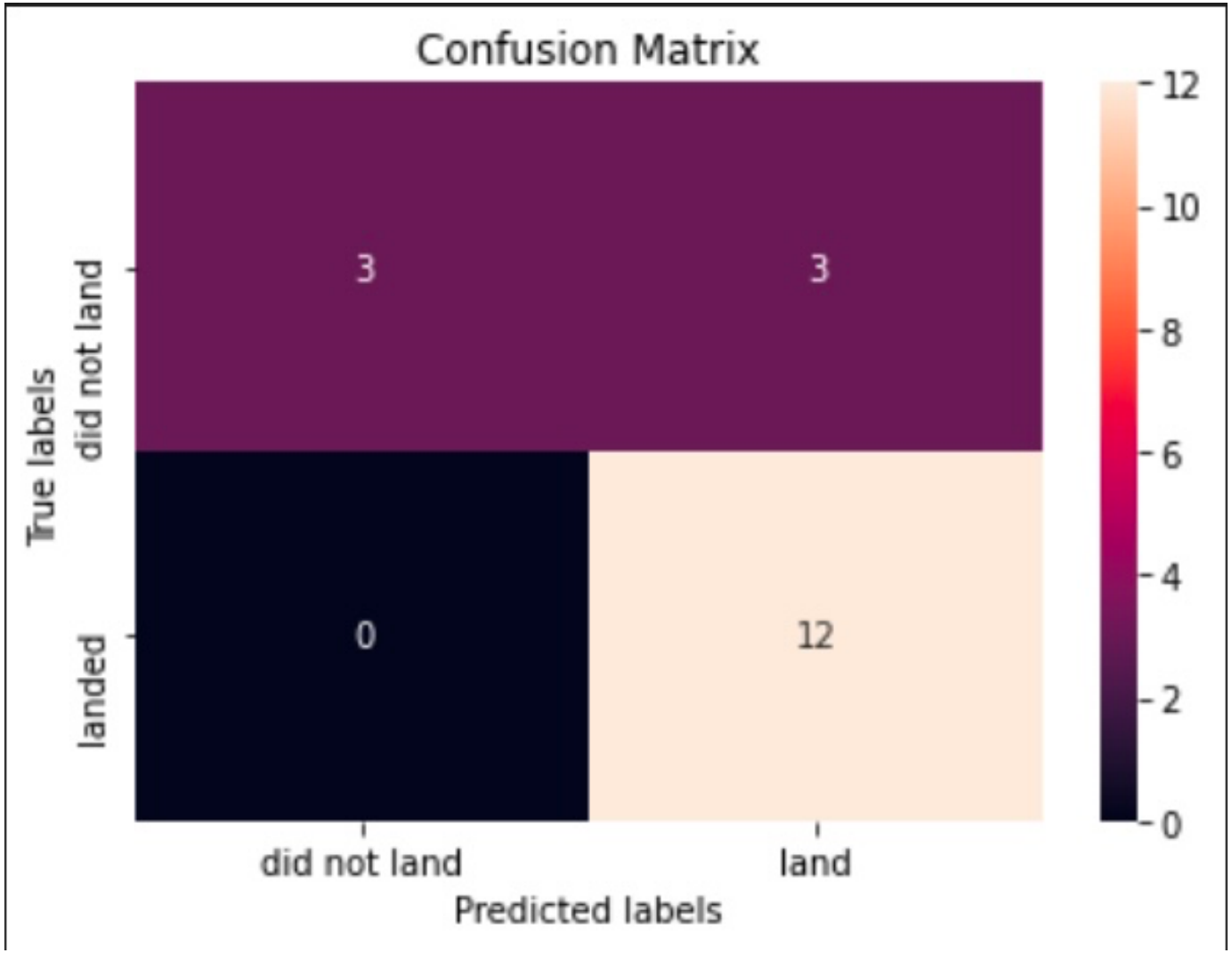
# Classification Accuracy




- The model with the highest accuracy is the Decision Tree Model

# Confusion Matrix

- The confusion matrix of the Decision Tree proves its accuracy by the 12 true positives out of a total of 18 observations



# Conclusions

- The best Launch Site seems to be KSC LC-39A
- Launches above 7000 kg are less risky
- Successful landing seem to be improving over time
- Decission Tree classifier may be used to predict successful landing

# Appendix

- To remove the multiple warnings that I was getting in the environment, I have used the warnings library

```
import warnings
warnings.filterwarnings("ignore")
```

# Thank you!