IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Filip Németh
8.9.2023

# Outline

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

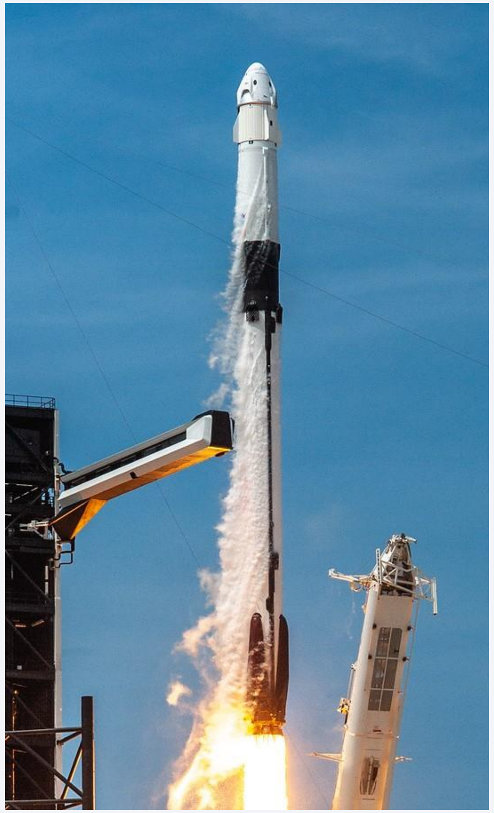
# Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

# Introduction

### Optimizing Rocket Launch Costs with Predictive Analytics

- In this capstone project, we aim to predict the successful landing of the Falcon 9 first stage during rocket launches
- SpaceX's cost-effective approach to rocket launches, driven by reusable first stages, has disrupted the industry, with savings of up to $100 million per launch
- Predictive analytics plays a pivotal role in optimizing decisions and costs in the aerospace industry



Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Data were collected from an API  to ensure its accuracy and reliability.
- Perform data wrangling
  - Historical Falcon 9 launch records through web scraping from the Wikipedia page titled 'List of Falcon 9 and Falcon Heavy launches
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - We used classification models like Logistic regression, SVM, KNN and Decision trees

# Data Collection

- We collect data from an API to ensure its accuracy and reliability
- Ensuring the data is in the correct format is essential for our predictive models

Request and parse the SpaceX launch data using the GET request → Filter the dataframe to only include `Falcon 9` launches → Dealing with Missing Values

Github code

# Data Wrangling

- Collecting historical Falcon 9 launch records through web scraping from the Wikipedia page titled 'List of Falcon 9 and Falcon Heavy launches,' enabling us to access valuable data for our analysis.

Request the Falcon9 Launch Wiki page from its URL → Extract all column/variable names from the HTML table header → Create a data frame by parsing the launch HTML tables

Github code

Section 2

# Insights drawn from EDA

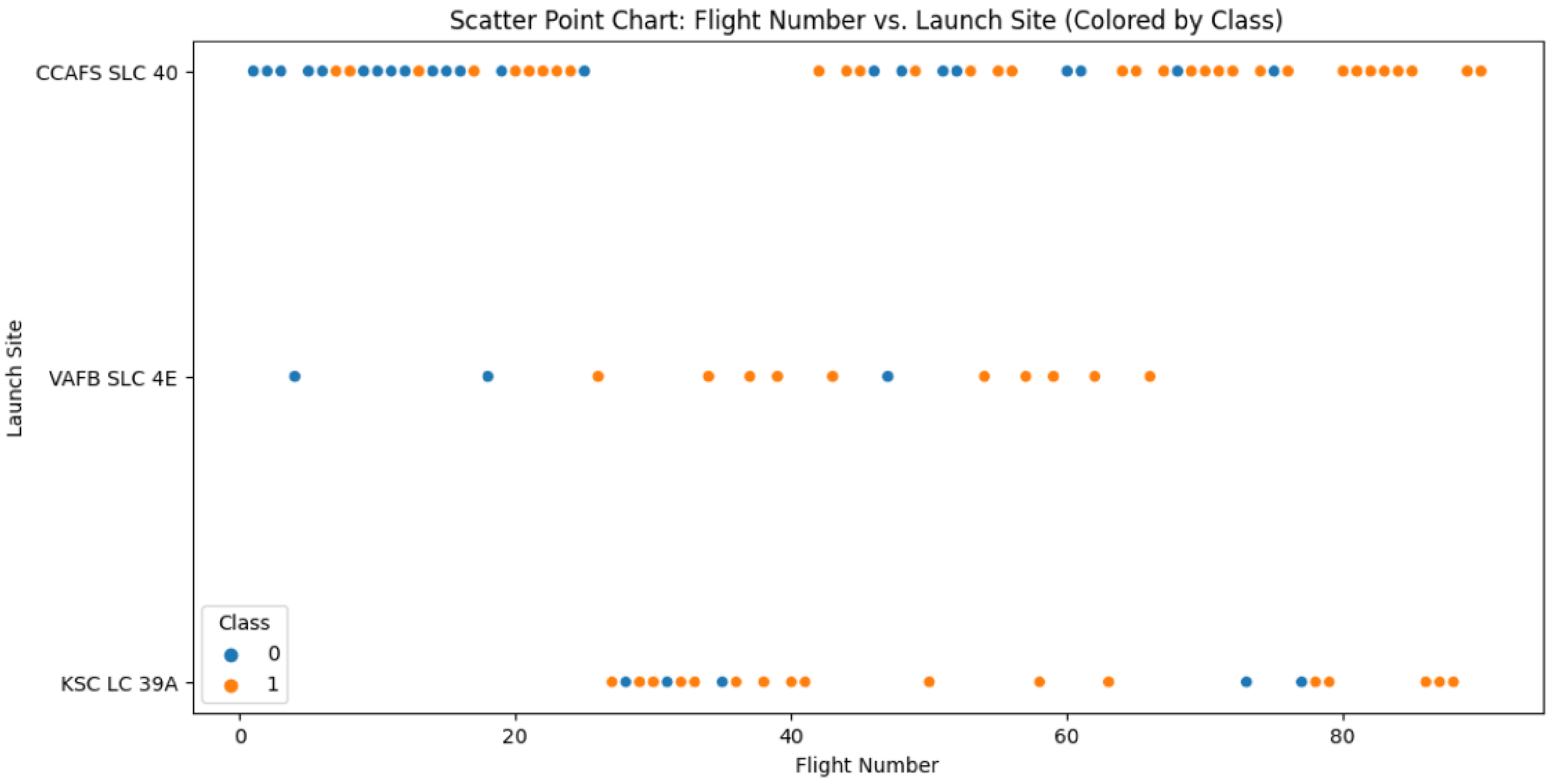
# EDA with Data Visualization

- Visualizations can help us gain insights into the Falcon 9 launches
- Performing EDA using **Pandas** and **Matplotlib**


Github code

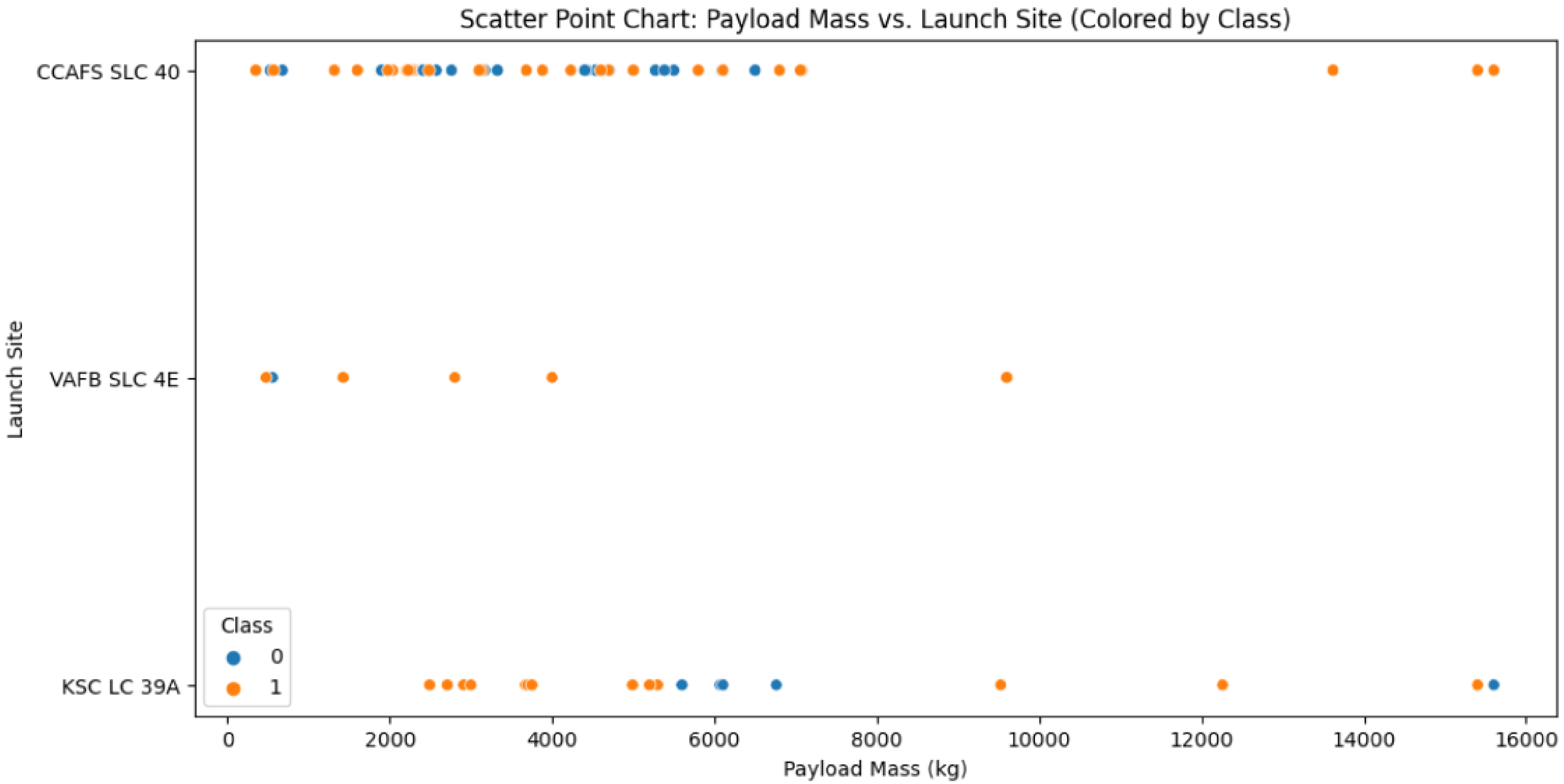
# Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site
- Most LauchSite are from CCAFS SLC 40 in both classes



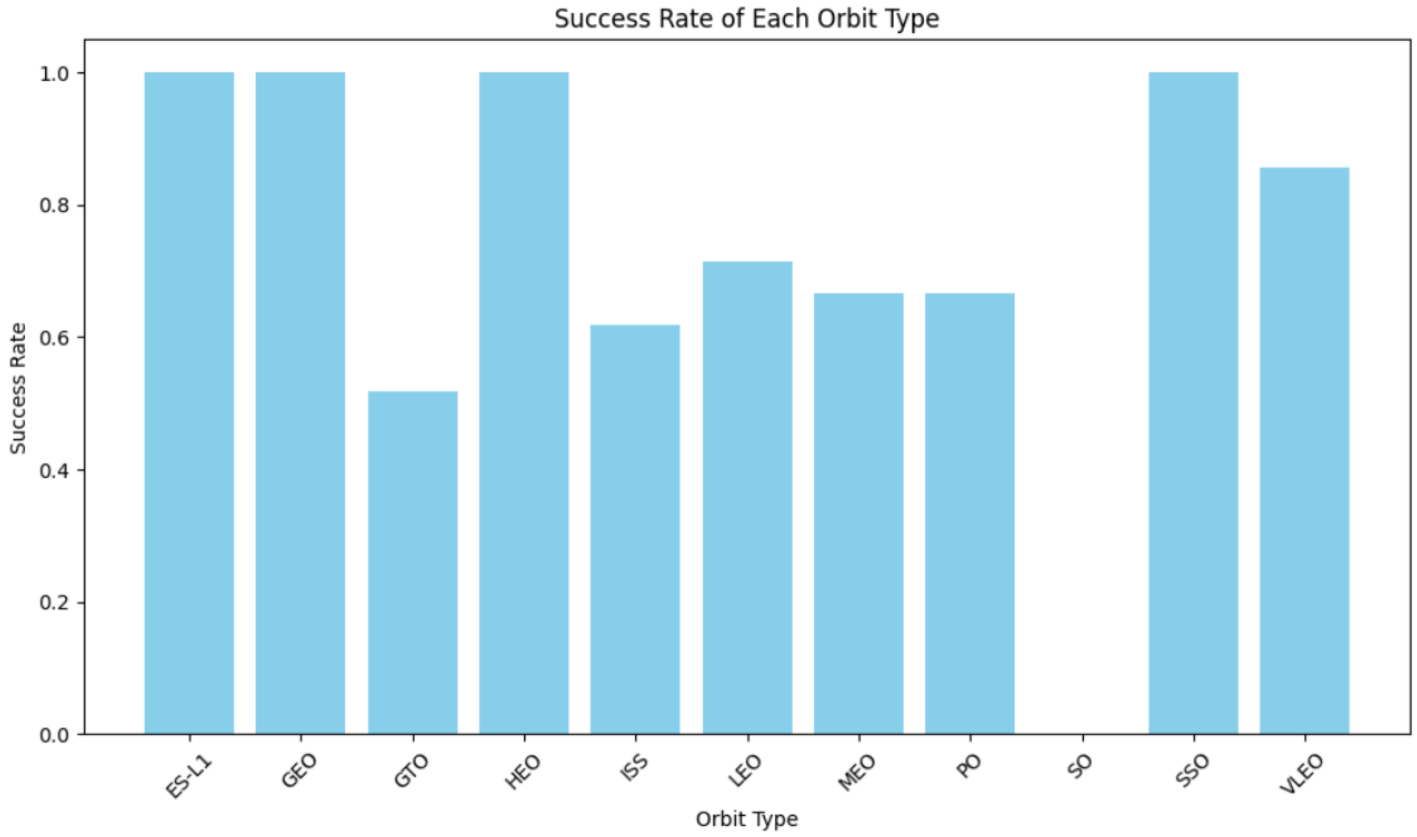
# Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- Launch Site CCAFS SLC 40 is more effective with Payload less than 8000 kg



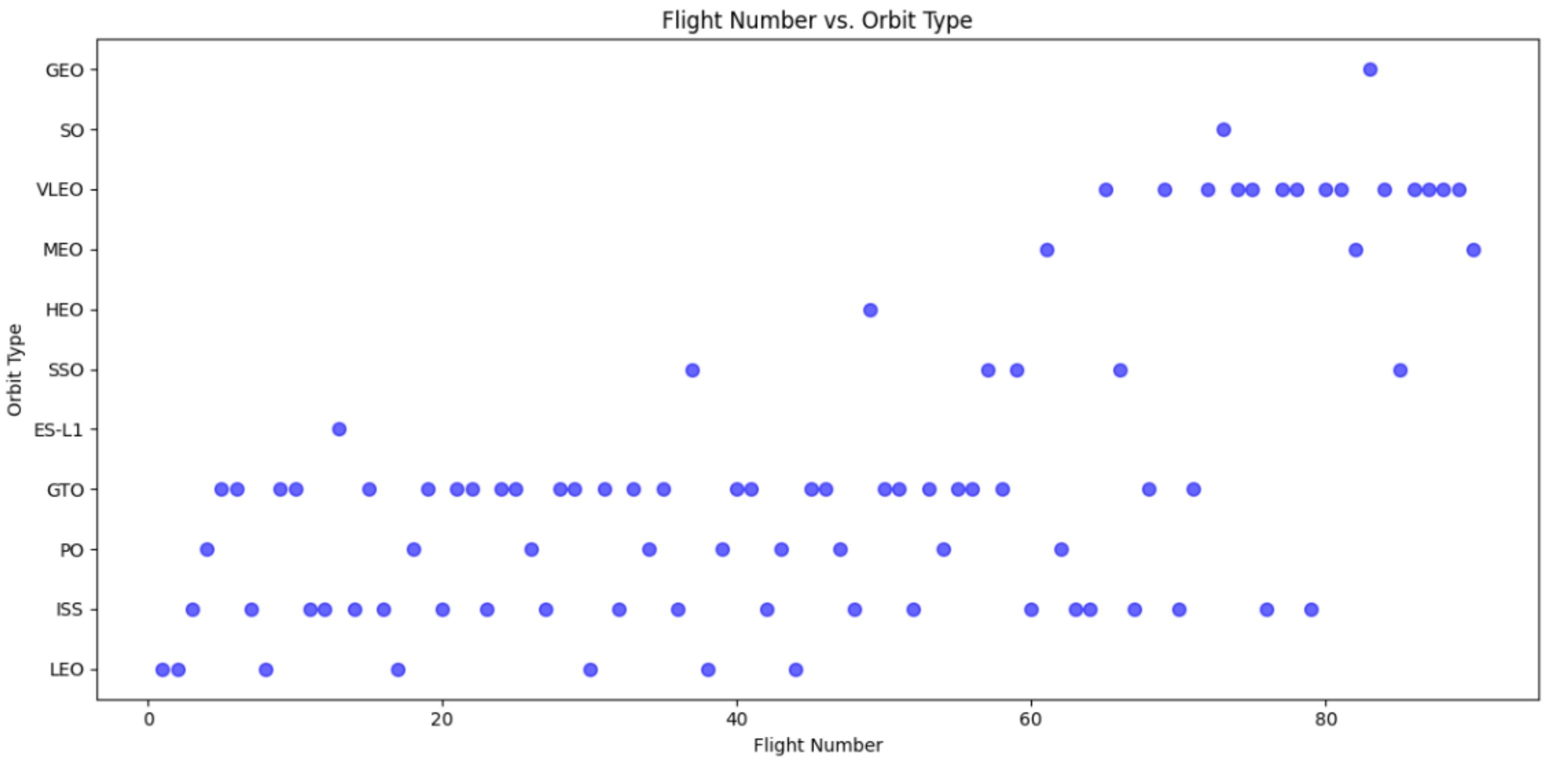
# Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type
- 4 of orbit type has successs rate 100 %
- Orbit type SO is useless



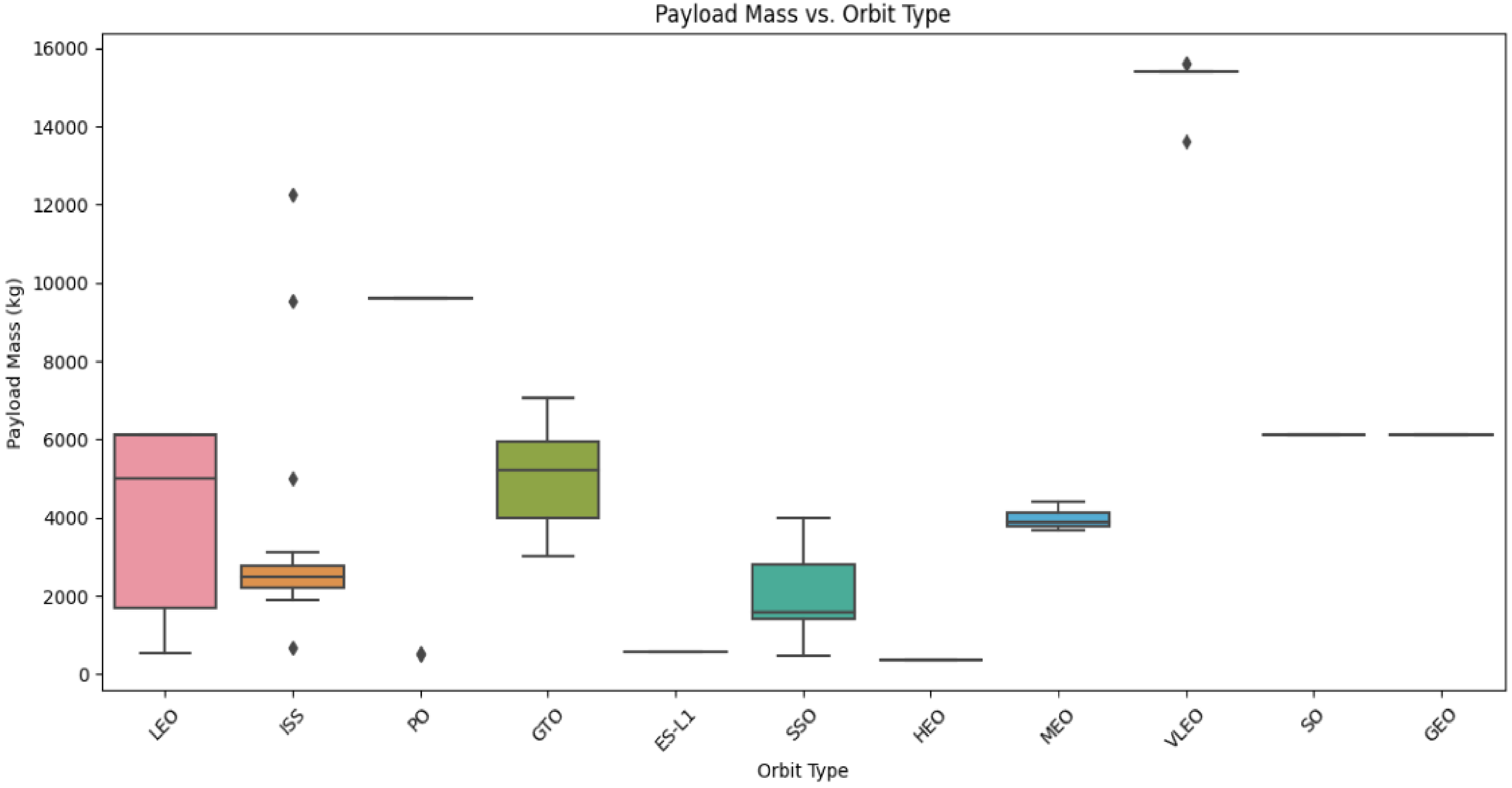
# Flight Number vs. Orbit Type

- Scatter point of 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

- Scatter point of payload vs. orbit type
- LEO has best interval of payload mass



# EDA with SQL

- Using SQL queries we performed this tasks

1. Names of the unique launch sites  in the space mission
2. 5 records where launch sites begin with the string 'CCA‘
3. Total payload mass carried by boosters launched by NASA (CRS)
4. Average payload mass carried by booster version F9 v1.1
5. The date when the first succesful landing outcome in ground pad was achieved
6. Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. Total number of successful and failure mission outcomes
8. Names of the booster_versions which have carried the maximum payload mass
9. Records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
10. 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.

# All Launch Site Names

- Names of the unique launch sites

1. **CCAFS LC-40:** Cape Canaveral Air Force Station Launch Complex 40. This is a launch site located at Cape Canaveral Space Launch Complex 40 in Florida, USA.
2. **VAFB SLC-4E:** Vandenberg Air Force Base Space Launch Complex 4E. This launch site is located at Vandenberg Air Force Base in California, USA.
3. **KSC LC-39A:** Kennedy Space Center Launch Complex 39A. This launch site is located at the Kennedy Space Center in Florida, USA. It's a historic launch site used for both crewed and uncrewed missions.
4. **CCAFS SLC-40:** Another reference to Cape Canaveral Air Force Station Launch Complex 40, indicating the same launch site as the first entry.

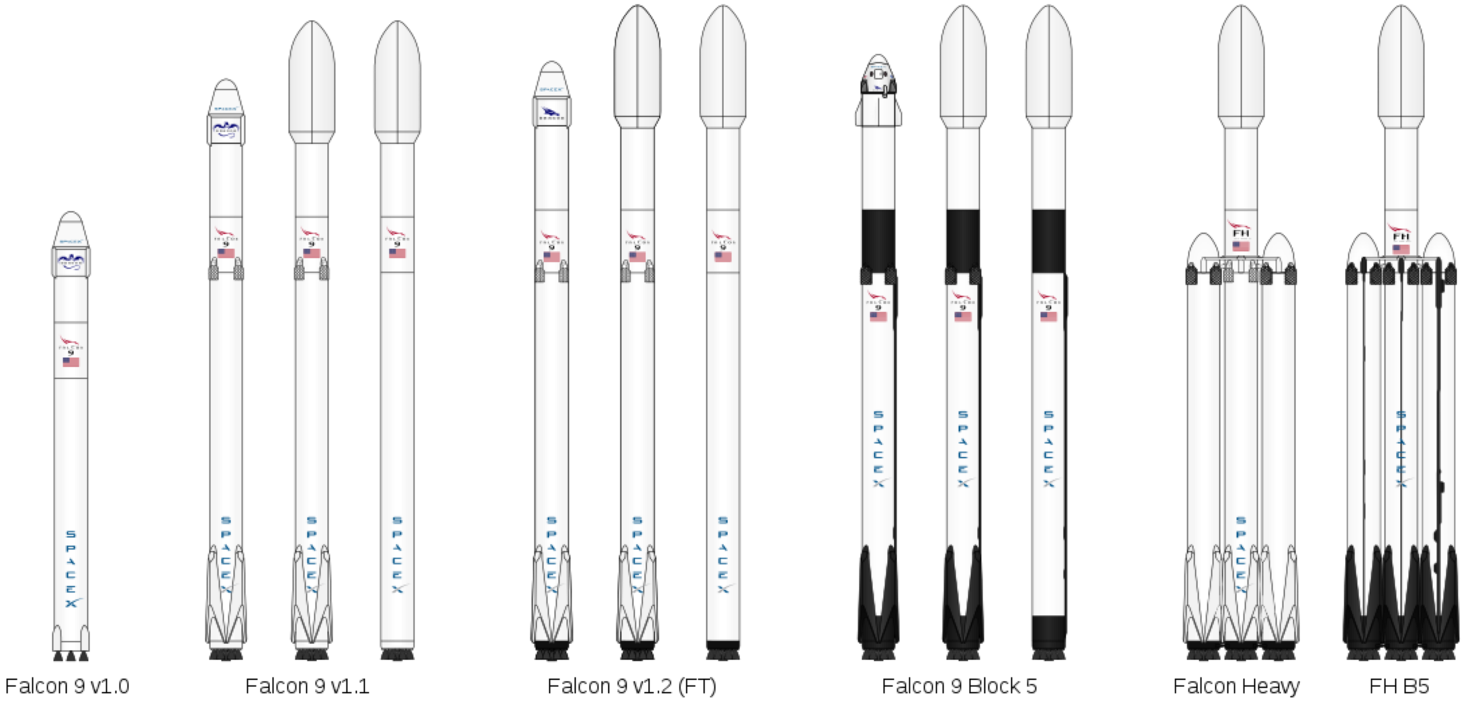
# Launch Site Names Begin with 'CCA'

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

- All Launch Sites which contain CCA were successful from year 2010 to 2013.
- Maximum payload was 677 kilogram and all flew in same orbit with abbreviation LEO

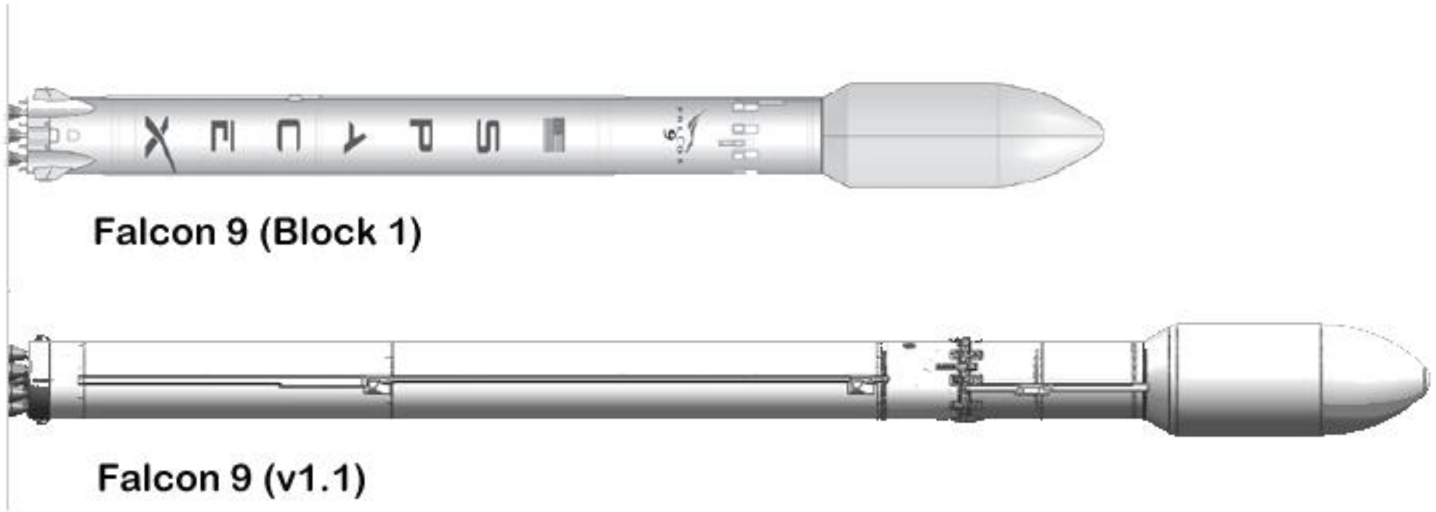
# Total Payload Mass

- Total payload carried by boosters from NASA is **45 596 kilogram**



# Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is **2534.67 kg**



# First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad is **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

- The total number of successful and failure mission outcomes

| Successful |
| --- |
| 100 |

| Failure |
| --- |
| 1 |

# Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass
- Different Booster Version has same maximum Payload mass

| Booster_Version | Payload mass (kg) |
|---|---|
| F9 B5 B1048.4 | 15600 |
| F9 B5 B1049.4 | 15600 |
| F9 B5 B1051.3 | 15600 |
| F9 B5 B1056.4 | 15600 |
| F9 B5 B1048.5 | 15600 |
| F9 B5 B1051.4 | 15600 |
| F9 B5 B1049.5 | 15600 |
| F9 B5 B1060.2 | 15600 |
| F9 B5 B1058.3 | 15600 |
| F9 B5 B1051.6 | 15600 |
| F9 B5 B1060.3 | 15600 |
| F9 B5 B1049.7 | 15600 |

# 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
- The more landing outcome are without the attempt with score of 10 very slighty behind is success in ground pad with score of 5

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

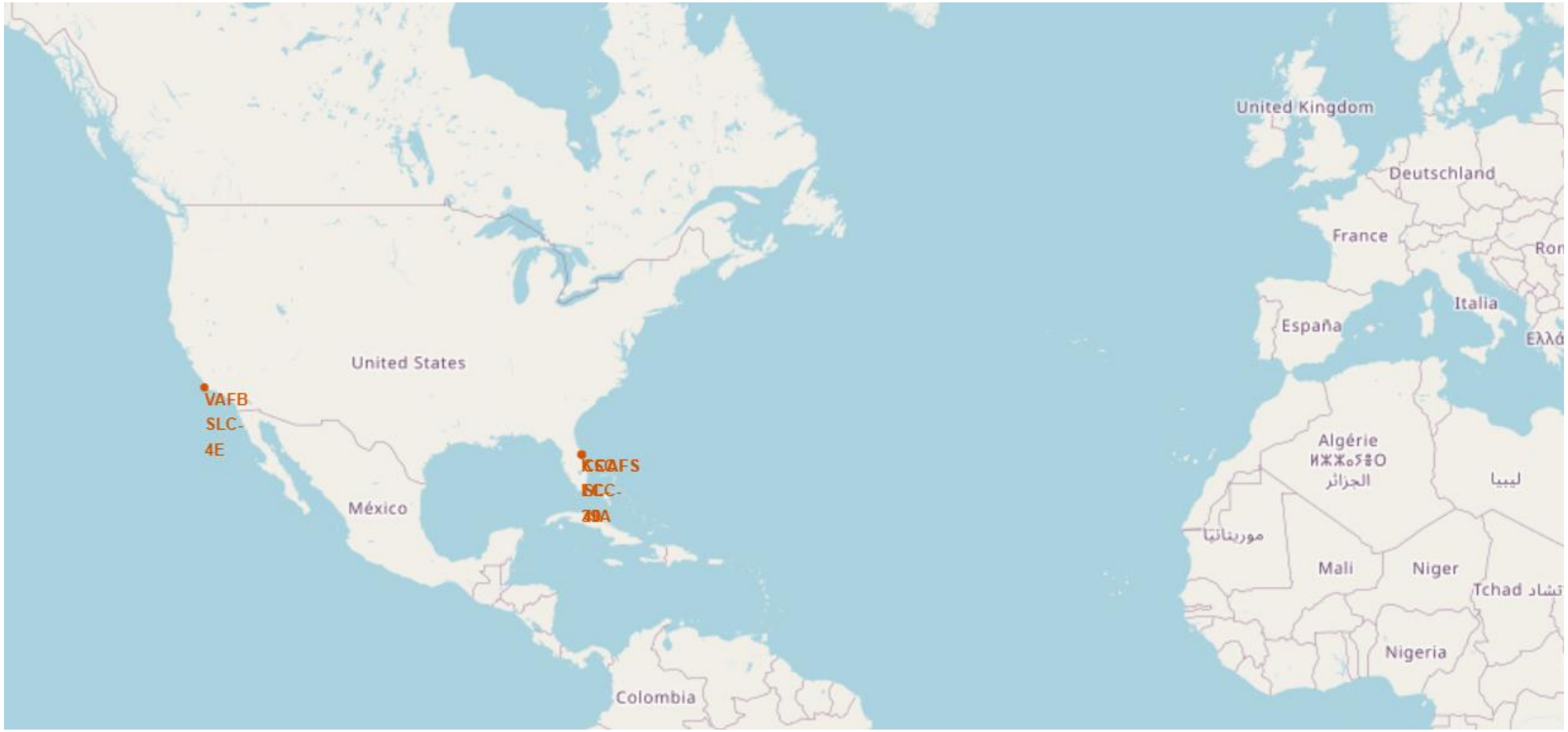
Section 4

# Launch Sites Proximities Analysis

# Build an Interactive Map with Folium

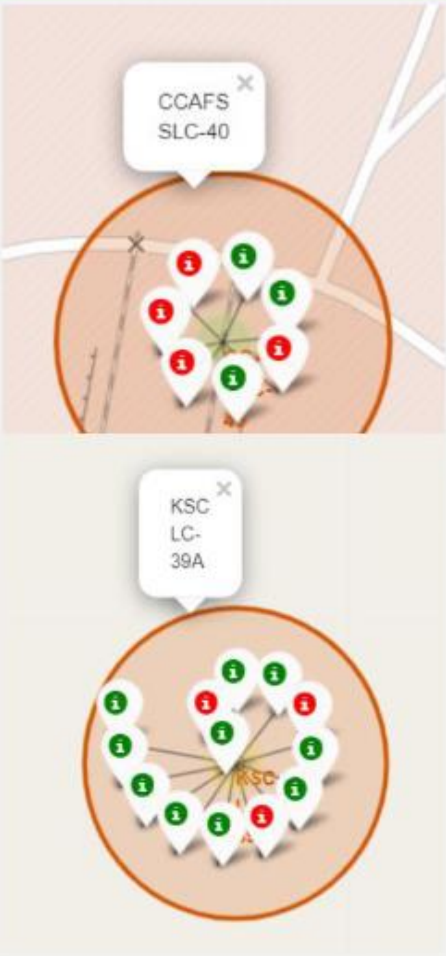
- Markers are added to highlight specific points of interest or to provide location-based information to users.
- Circles are used to visualize coverage areas, proximity zones, or radii around specific points.
- Lines are employed to show routes, connections, or pathways between locations.
- Polygons are utilized to define and display areas or regions of significance.

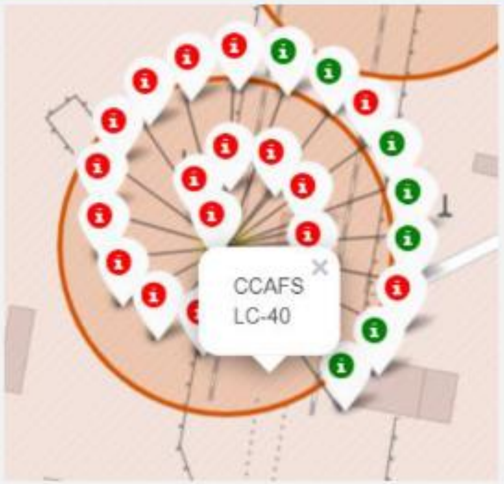
# Global map markers with all launch sites




[Github code](#)

# Markers showing launch sites with color labels




***Florida Launch Sites***

*Green Marker shows successful Launches and Red Marker shows Failures*


***California Launch Site***

37

# Launch Site distance to landmarks


- •Are launch sites in close proximity to railways? No
- •Are launch sites in close proximity to highways? No
- •Are launch sites in close proximity to coastline? Yes
- •Do launch sites keep certain distance away from cities? Yes

Section 5

# Build a Dashboard with Plotly Dash

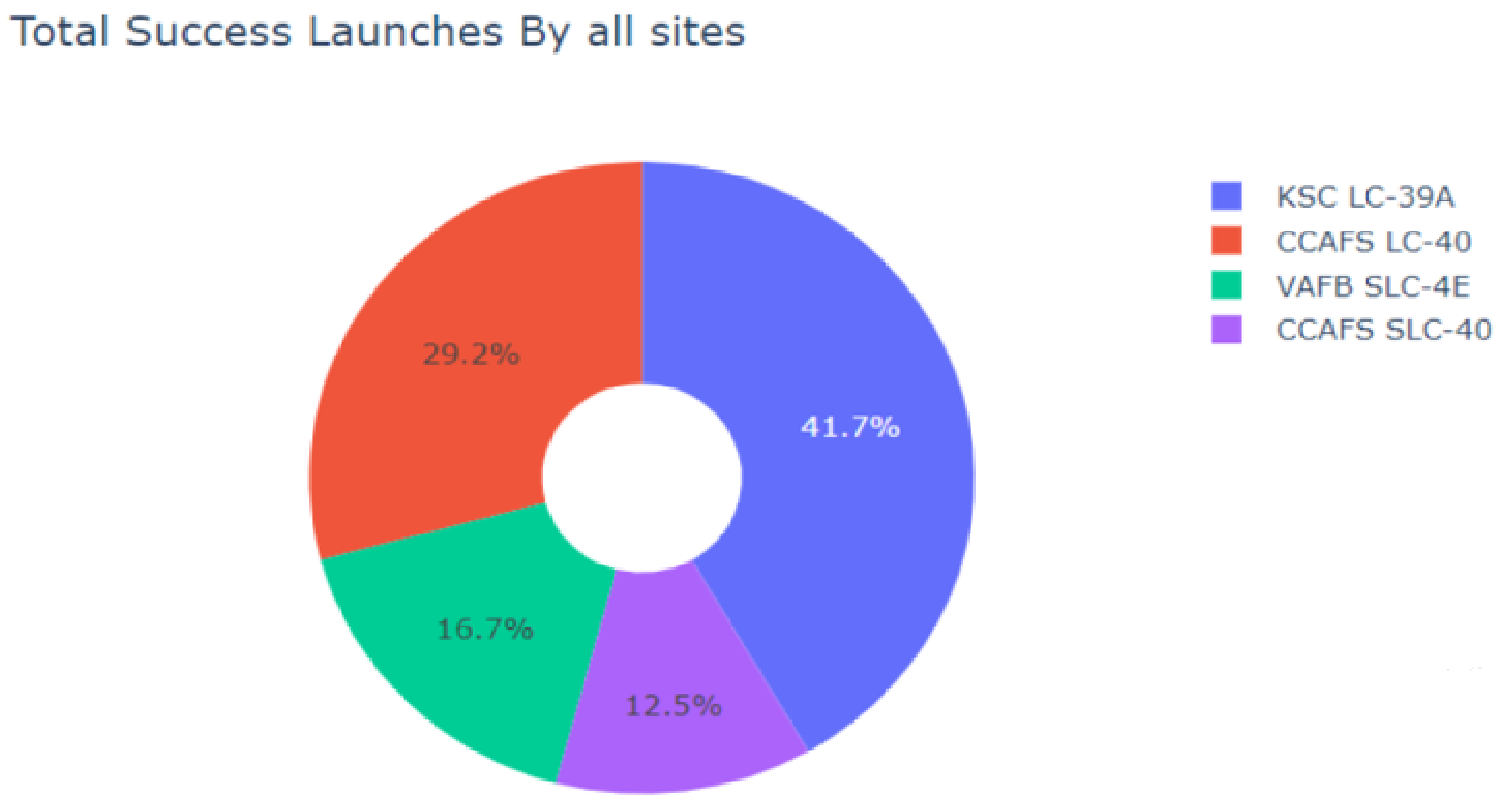
# Build a Dashboard with Plotly Dash

- We added graphs with many purposes

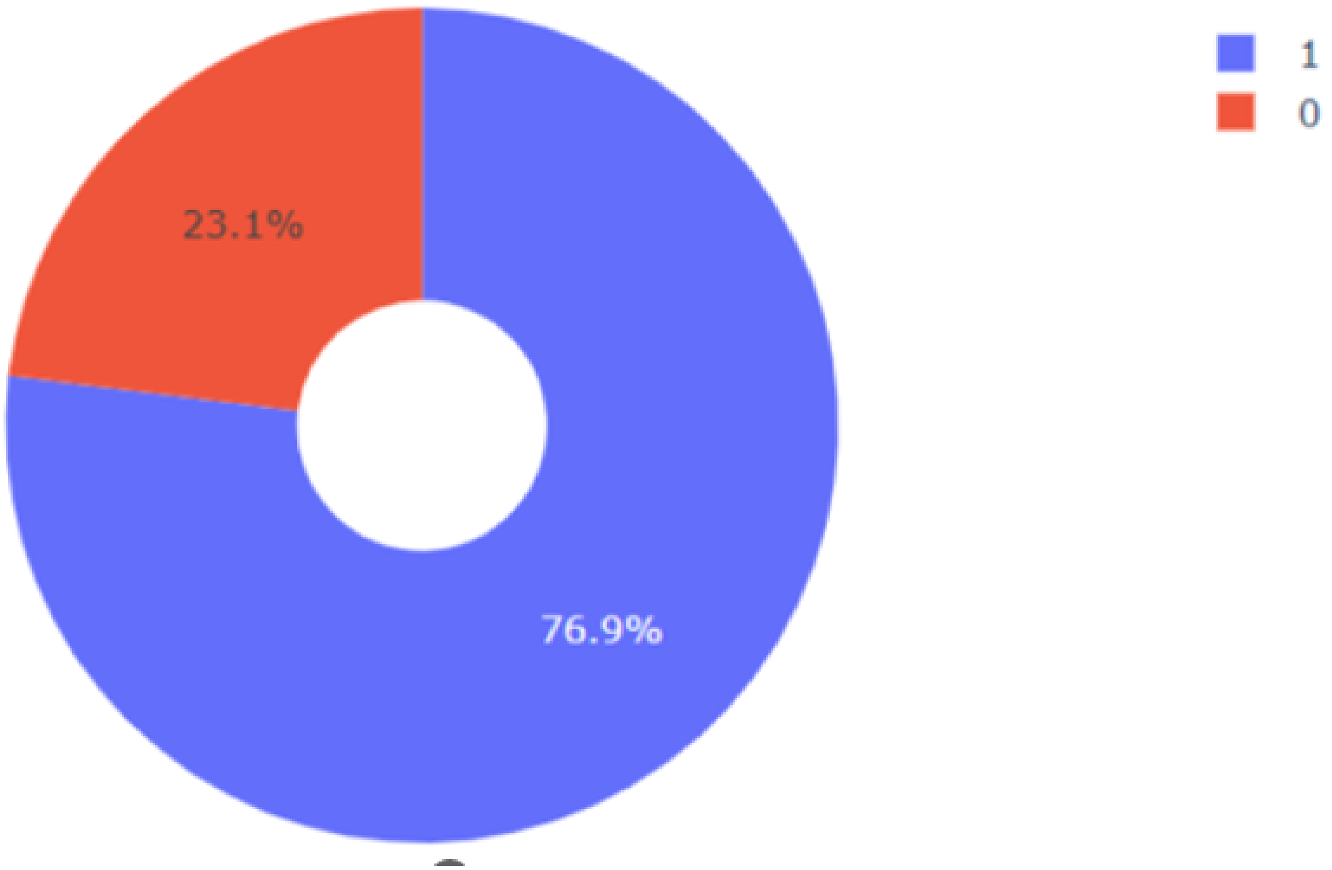
1. Pie chart showing the success percentage achieved by each launch site
2. Pie chart showing the Launch site with the highest launch success ratio
3. Scatter plot of Payload vs Launch Outcome for all sites, with different payload selected in the range slider

[Github code]()

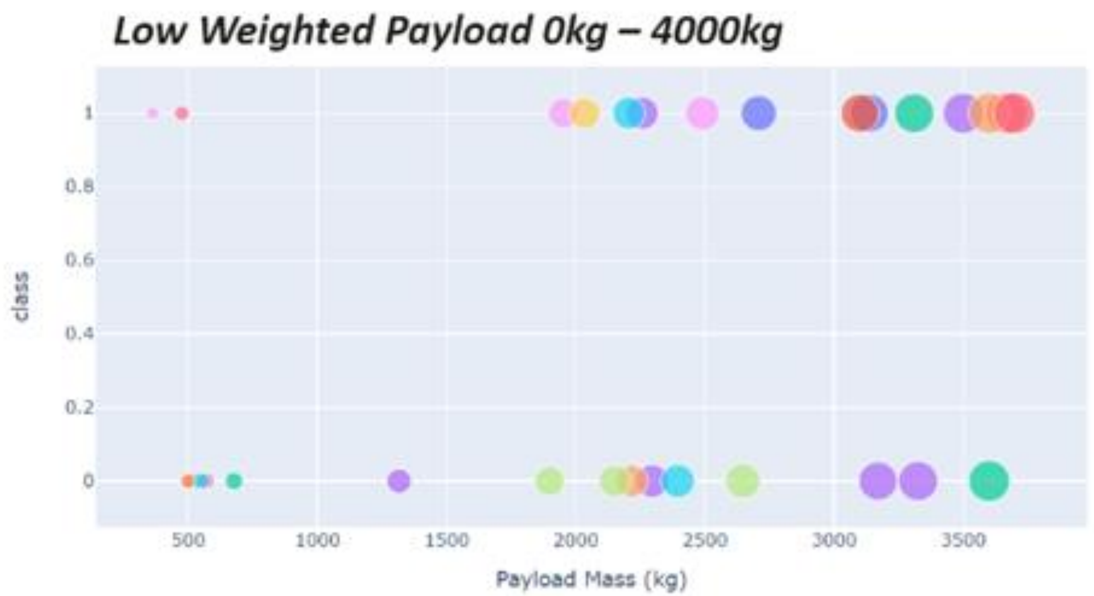
# Pie chart showing the success percentage achieved by each launch site

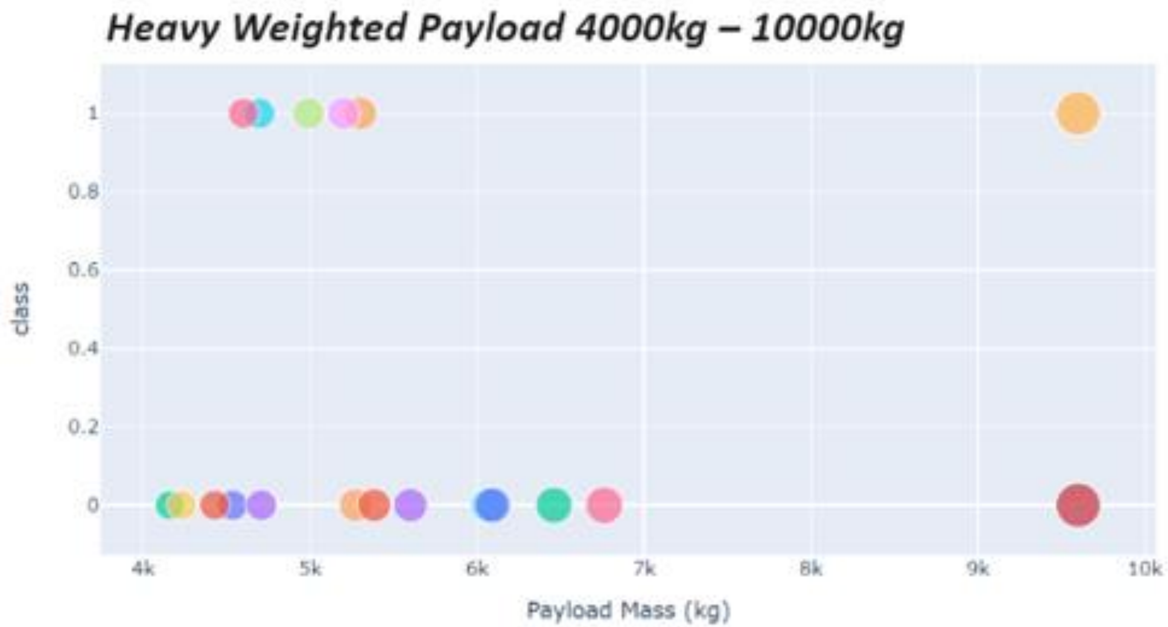


# Pie chart showing the Launch site with the highest launch success ratio



# Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider



Section 6

# Predictive Analysis (Classification)

# Predictive Analysis (Classification)

- The decision tree classifier's confusion matrix indicates its ability to differentiate between various classes. The primary issue lies in the occurrence of false positives, meaning that the classifier incorrectly identifies unsuccessful landings as successful landings.

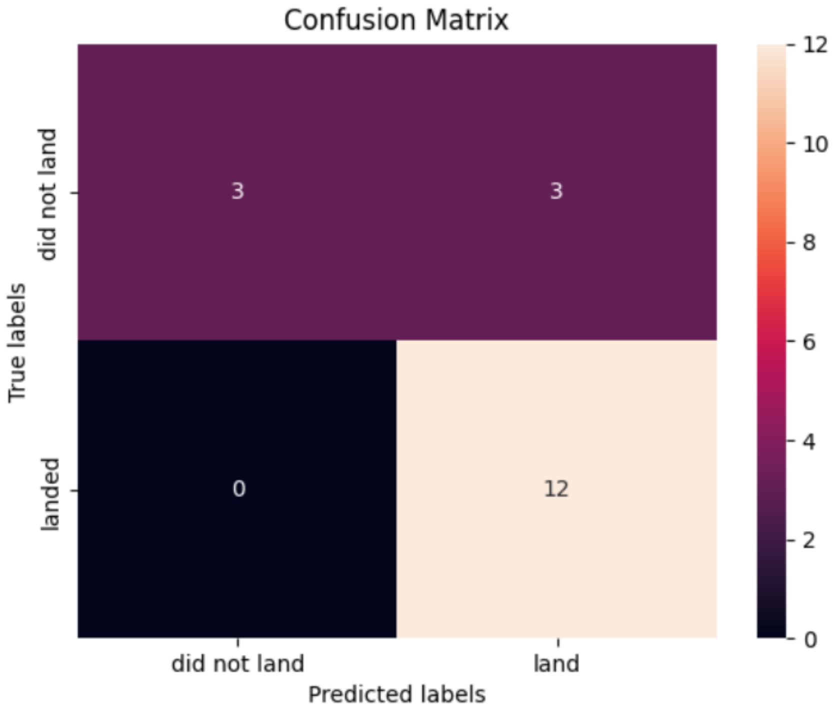
Github code

# Classification Accuracy

- The decision tree classifier's confusion matrix indicates its ability to differentiate between various classes. The primary issue lies in the occurrence of false positives, meaning that the classifier incorrectly identifies unsuccessful landings as successful landings.
- Best model is **Decision Tree** with a score of 0.90

# Confusion Matrix

- It is a matrix that summarizes the number of correct and incorrect predictions made by the model, broken down by class or category. The confusion matrix typically consists of four important values.



# Conclusions

- There is a positive correlation between the number of flights conducted at a launch site and its success rate. More launches tend to result in higher success rates at a site.
- The trend of launch success rates has been on the rise from 2013 to 2020.
- The orbits labeled ES-L1, GEO, HEO, SSO, and VLEO experienced the highest success rates among all orbit types.
- Among all launch sites, KSC LC-39A stands out with the highest number of successful launches.
- The Decision tree classifier appears to be the most effective machine learning algorithm for this specific task, based on its performance and results.

# Thank you!