IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

AGNI WIGUNA
18 AUGUST 2024

# Outline

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

# Executive Summary

- Summary of methodologies
  - Collecting data using webscrapping on Wikipedia and SpaceX API
  - Exploratory Data Analysis (EDA) : Data wrangling and Data Visualization
  - Machine Learning Prediction
- Summary of all results
  - We can get valuable information using public API and website
  - Using EDA we can new insight about data
  - Machine learning can help us to get new information and opportunity

# Introduction

- Project background and context
  - Space X is a large rocket launch company. SpaceX aims to conduct space tours. SpaceX has stated that it can reduce the cost of launching rockets by analyzing the first stage. If we can calculate that the first stage can land, we can calculate the cost required for the next project. From the existing public information, we will create machine learning to predict whether SpaceX can use its first stage..
- Problems you want to find answers
  - How do we look at the variables that can determine the success of a first stage landing?
  - What is the best algorithm to perform this analysis?

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Using SpaceX open API
  - Using web Scrapping from Wikipedia
- Perform data wrangling
  - Filtering the data
  - Handling Missing Value
  - Apply one hot encoding for prediction preparation
- Perform exploratory data analysis (EDA) using visualization and SQL

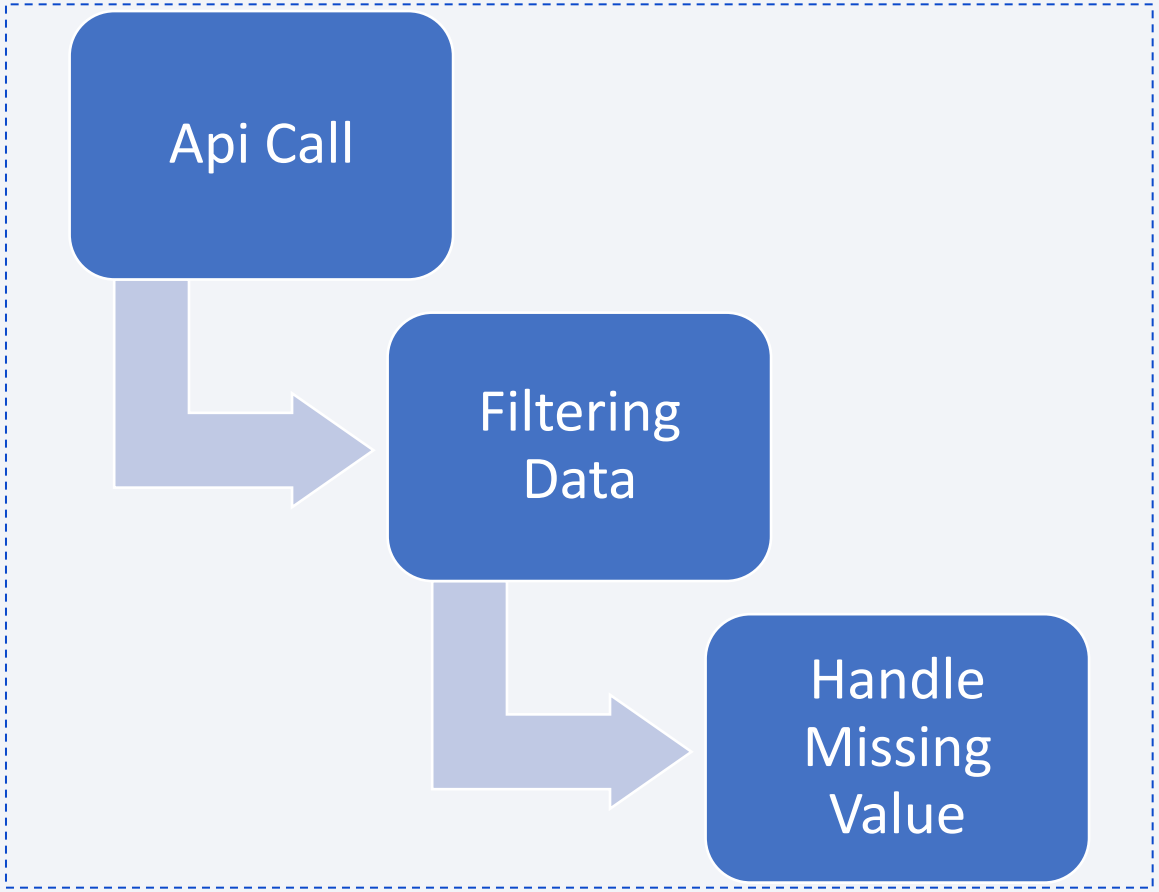
# Data Collection

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - From the data that has been prepared, classification is carried out with various models to determine which model is most suitable.

# Data Collection – SpaceX API

- Data taken from spaceX Open API
- The API will be utilized as per the flowchart on the side

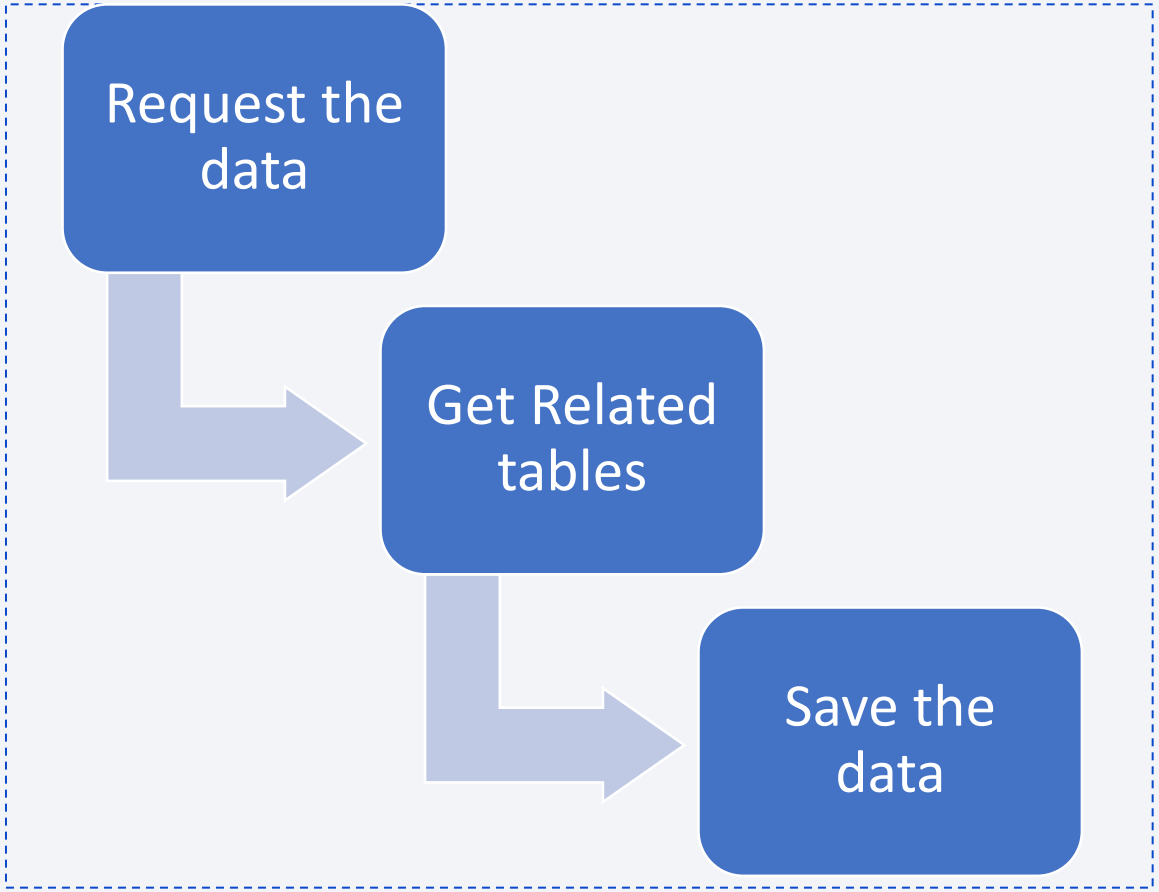
Source Code : [IBM-Applied-Data-Science-Capstone/Data Collection API.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)]



# Data Collection - Scraping

- SpaceX rocket launch data can also be obtained from wikipedia
- Data is taken by scraping using BS4 and retrieved related tables

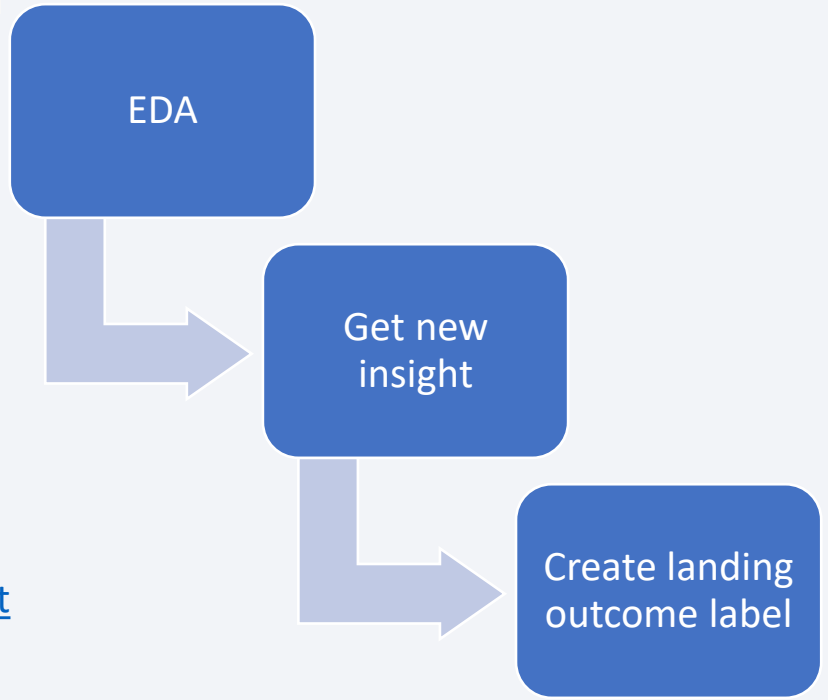
Source Code : [IBM-Applied-Data-Science-Capstone/Data Collection with Web Scrapping.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)]



# Data Wrangling

- In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; for example, `True Ocean` means the mission outcome was successfully landed to a specific region of the ocean while `False Ocean` means the mission outcome was unsuccessfully landed to a specific region of the ocean. `True RTLS` means the mission outcome was successfully landed to a ground pad `False RTLS` means the mission outcome was unsuccessfully landed to a ground pad.`True ASDS` means the mission outcome was successfully landed on a drone ship `False ASDS` means the mission outcome was unsuccessfully landed on a drone ship.
- In this lab we will mainly convert those outcomes into Training Labels with `` `1` `` means the booster successfully landed `` `0` `` means it was unsuccessful.

Source Code : [IBM-Applied-Data-Science-Capstone/Data Wranggling.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)]



# EDA with Data Visualization

- Charts were plotted:
  - Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Tren.
  - Scatter plots show the relationship between variables.
  - Bar charts show comparisons among discrete categories.
  - Line charts show trends in data over time (time series)

Source Code : IBM-Applied-Data-Science-Capstone/Eda Data Visualization.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)

# EDA with SQL

## Performed SQL Query:

- Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA';
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015;
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

Source Code : IBM-Applied-Data-Science-Capstone/EDA with SQL.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)

# Build an Interactive Map with Folium

- Markers of all Launch Sites:
  - Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location.
  - Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.
- Coloured Markers of the launch outcomes for each Launch Site:
  - Added coloured Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.
- Distances between a Launch Site to its proximities:
  - Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City

Source Code : IBM-Applied-Data-Science-Capstone/Visual Analytics with Folium.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)

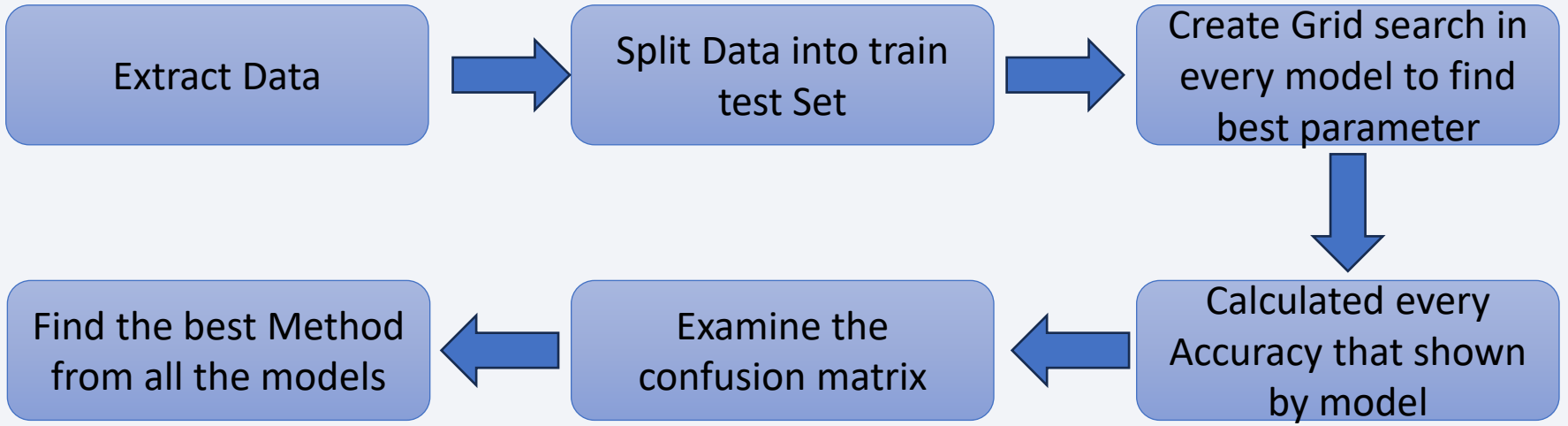
# Build a Dashboard with Plotly Dash

- Launch Sites Dropdown List
  - Added a dropdown list to enable Launch Site selection.
- Pie Chart showing Success Launches (All Sites/Certain Site)
  - Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected
- Slider of Payload Mass Range
  - Added a slider to select Payload range
- Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions:
  - Added a scatter chart to show the correlation between Payload and Launch Success

Source Code : IBM-Applied-Data-Science-Capstone/spacex_dash_app.py at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)

# Predictive Analysis (Classification)

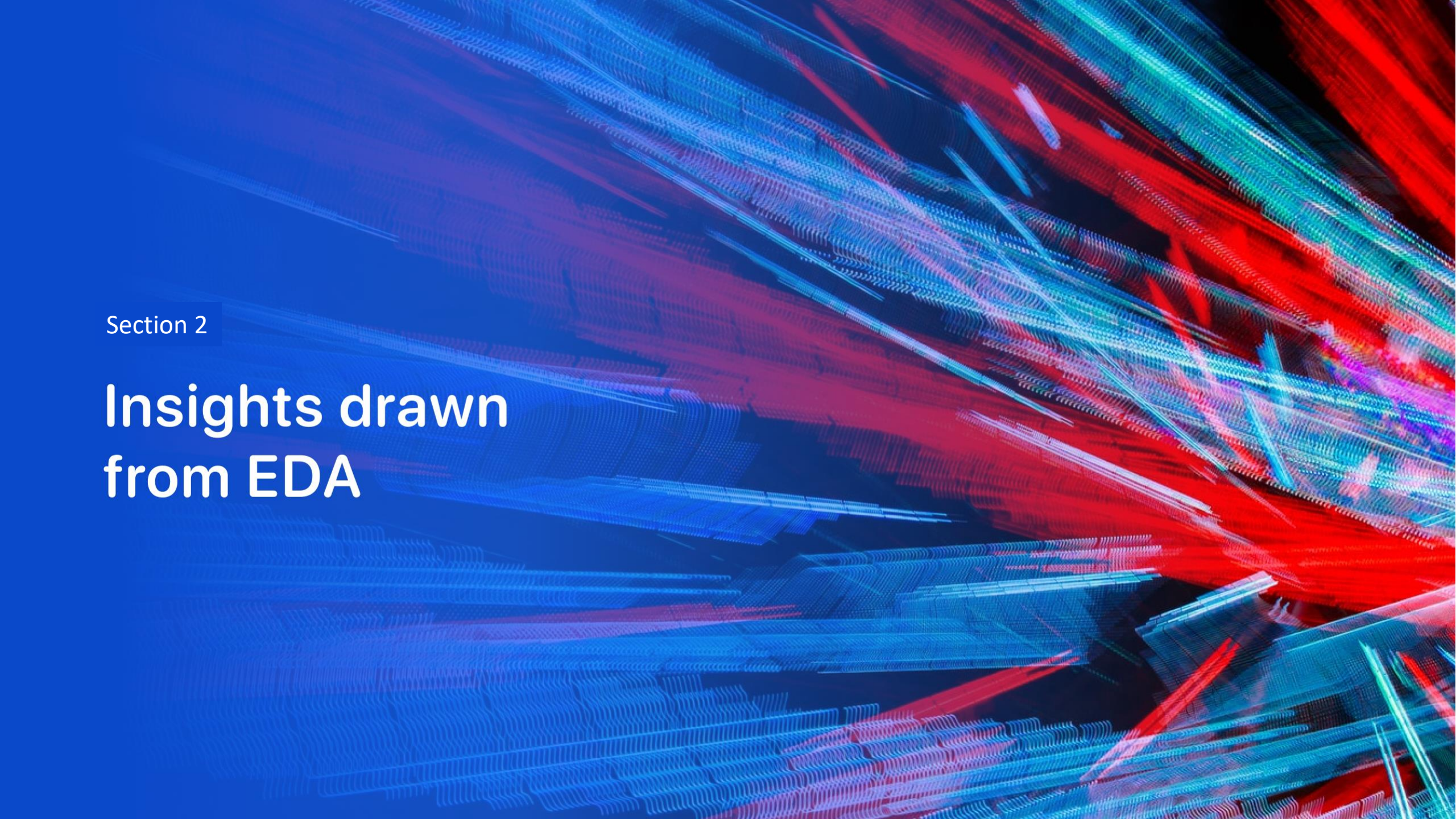
- The steps I applied from start to finish were data extraction, data transformation, data sharing, finding the best parameters, model evaluation, and finding the best model.




Source Code : [IBM-Applied-Data-Science-Capstone/Predictive Analyze.ipynb at main · agniwii/IBM-Applied-Data-Science-Capstone (github.com)]

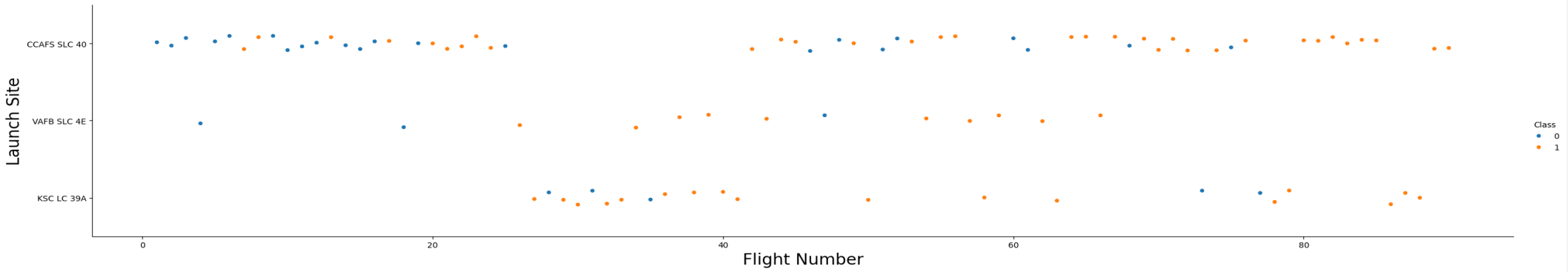
# Results

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


Section 2
Insights drawn from EDA

# Flight Number vs. Launch Site

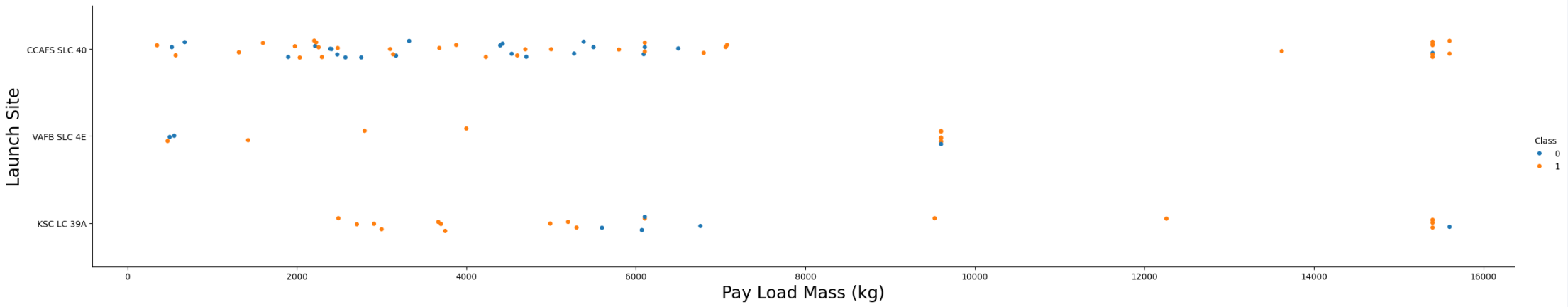
- A scatter plot of Flight Number vs. Launch Site




- Explanations
  - There are some gaps in the flight number sequence, particularly noticeable for VAFB SLC-4E and KSC LC-39A.
  - CCAFS SLC-40 is the most frequently used launch site, with points clustered densely at the top of the chart.
  - Most of the earliest flight are not success
  - Lastest flight have more successfull rate

# Payload vs. Launch Site

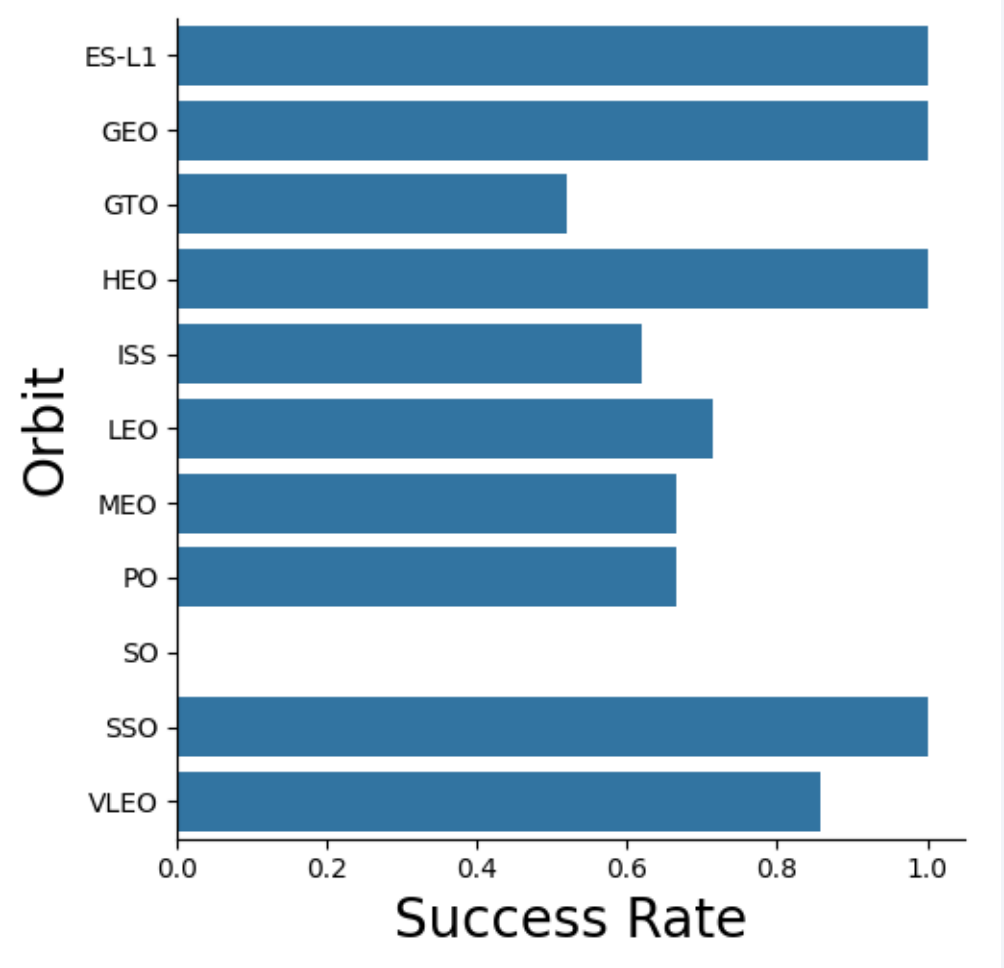
- A scatter plot of Payload vs. Launch Site




- Explanations
  - Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

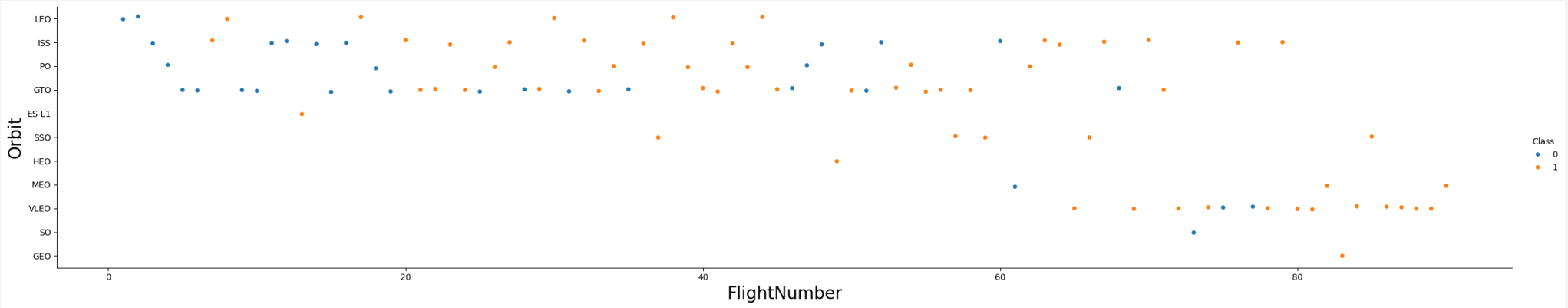
# Success Rate vs. Orbit Type

- A bar chart for the success rate of each orbit type
- Explanation
  - 1. ES-L1, GEO, HEO, and SSO have the highest success rates, both at or very close to 100%.
  - 2. VLEO has a success rate of about 85-90%.
  - 3. LEO , ISS , MEO, and PO all have success rates between 70-80%.
  - 4. GTO has the lowest success rate among the orbits shown, at around 50-60%.



# Flight Number vs. Orbit Type

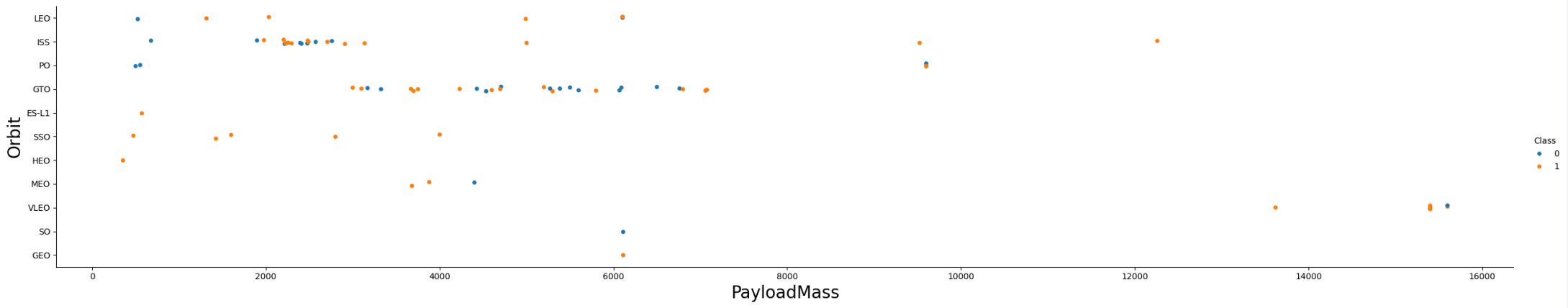
- A scatter point of Flight number vs. Orbit type




- Explanations
  - You should see that in the 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

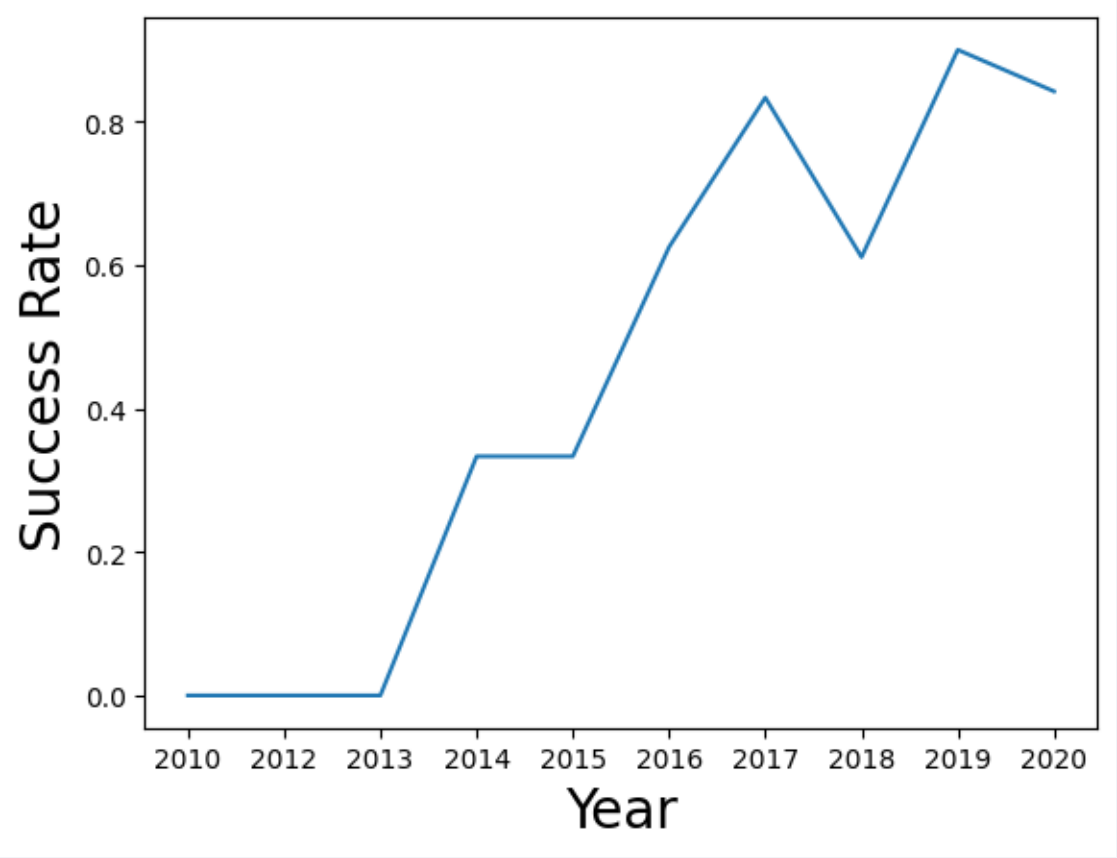
- Show a scatter point of payload vs. orbit type




- Explanations
  - 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

- Aline chart of yearly average success rate
- Explanations
  - You can observe that the success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.



# All Launch Site Names

- Find the names of the unique launch sites

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

- It retrieved all unique launch site using distinct function

# Launch Site Names Begin with 'CCA'

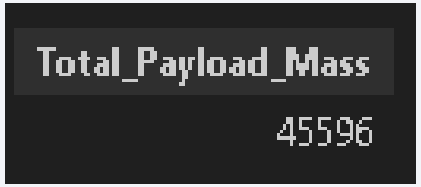
- Find 5 records where launch sites begin with `CCA`

| 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 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 2012-10-08 | 0:35:00 | 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 |

- Retrieve only 5 record that launchsite begin with CCA using like and limit method

# Total Payload Mass

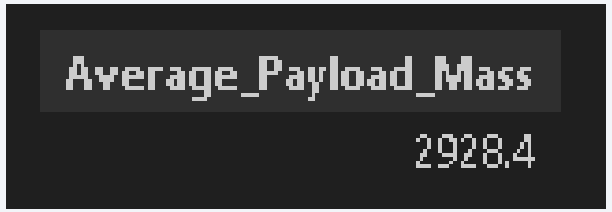
- Calculate the total payload carried by boosters from NASA




- Sum all total payload mass on the table

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1




- Calculate all average data if booster version F9 v1.1

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad


- Get minimum date when landing outcome on ground pad

# Successful Drone Ship Landing with Payload between 4000 and 6000

- List the 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 |

- Include all booster_version when they had payload mass greater than 4000 but less than 6000

# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

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

- Count every mision outcome that have been grouping before

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

| 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 |

- Select all booster version when it has same maximum mass

# 2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

| month | Date | Booster_Version | Launch_Site | Landing_Outcome |
|---|---|---|---|---|
| None | 2015-01-10 | F9 v1.1 B1012 | CCAFS LC-40 | Failure (drone ship) |
| None | 2015-04-14 | F9 v1.1 B1015 | CCAFS LC-40 | Failure (drone ship) |

- Shown landing outcome that failure in 2015

# 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

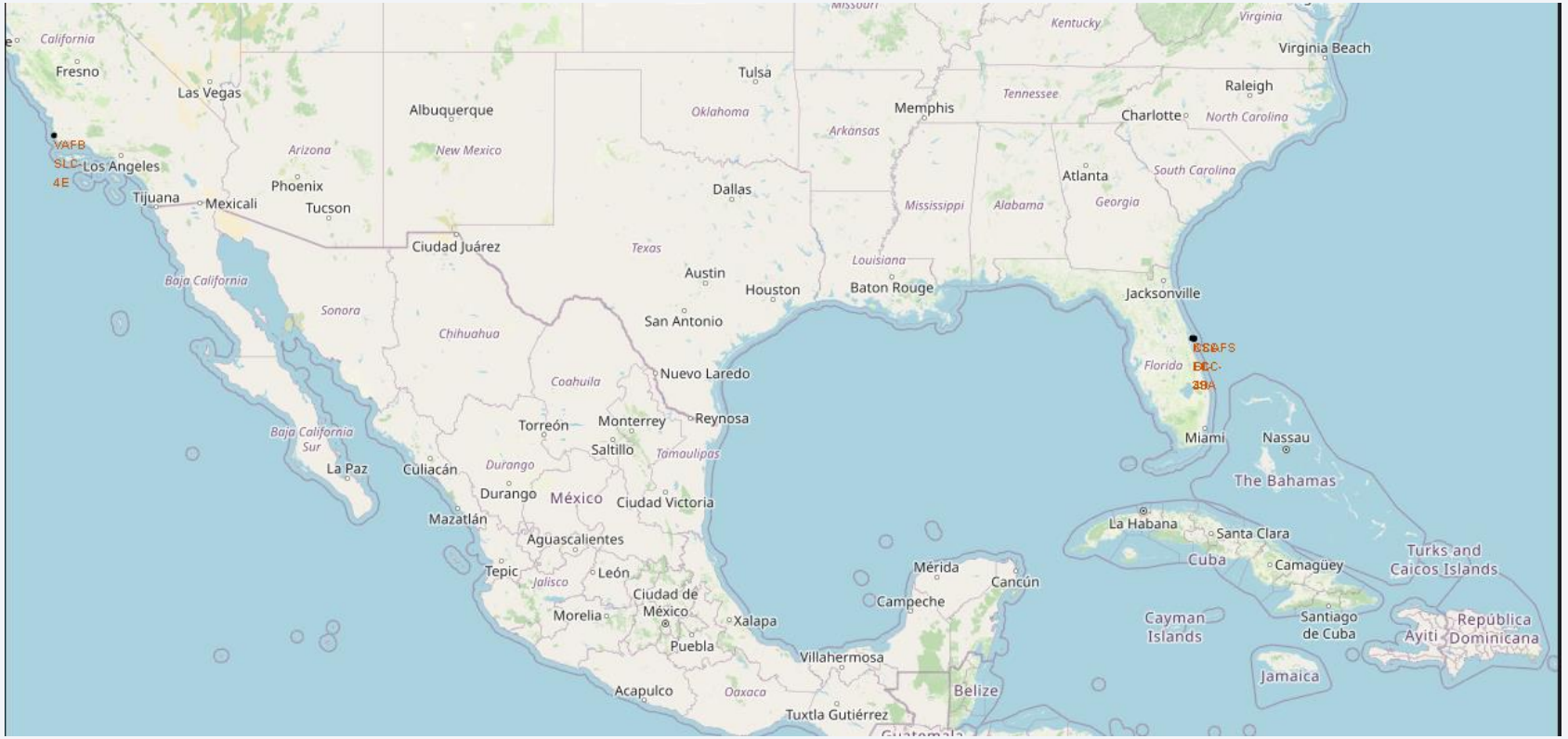
| Landing_Outcome | Count |
|---|---|
| No attempt | 10 |
| Success (drone ship) | 5 |
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |
| Controlled (ocean) | 3 |
| Uncontrolled (ocean) | 2 |
| Failure (parachute) | 2 |
| Precluded (drone ship) | 1 |

- This view of data alerts us that “No attempt” must be taken in account

Section 3

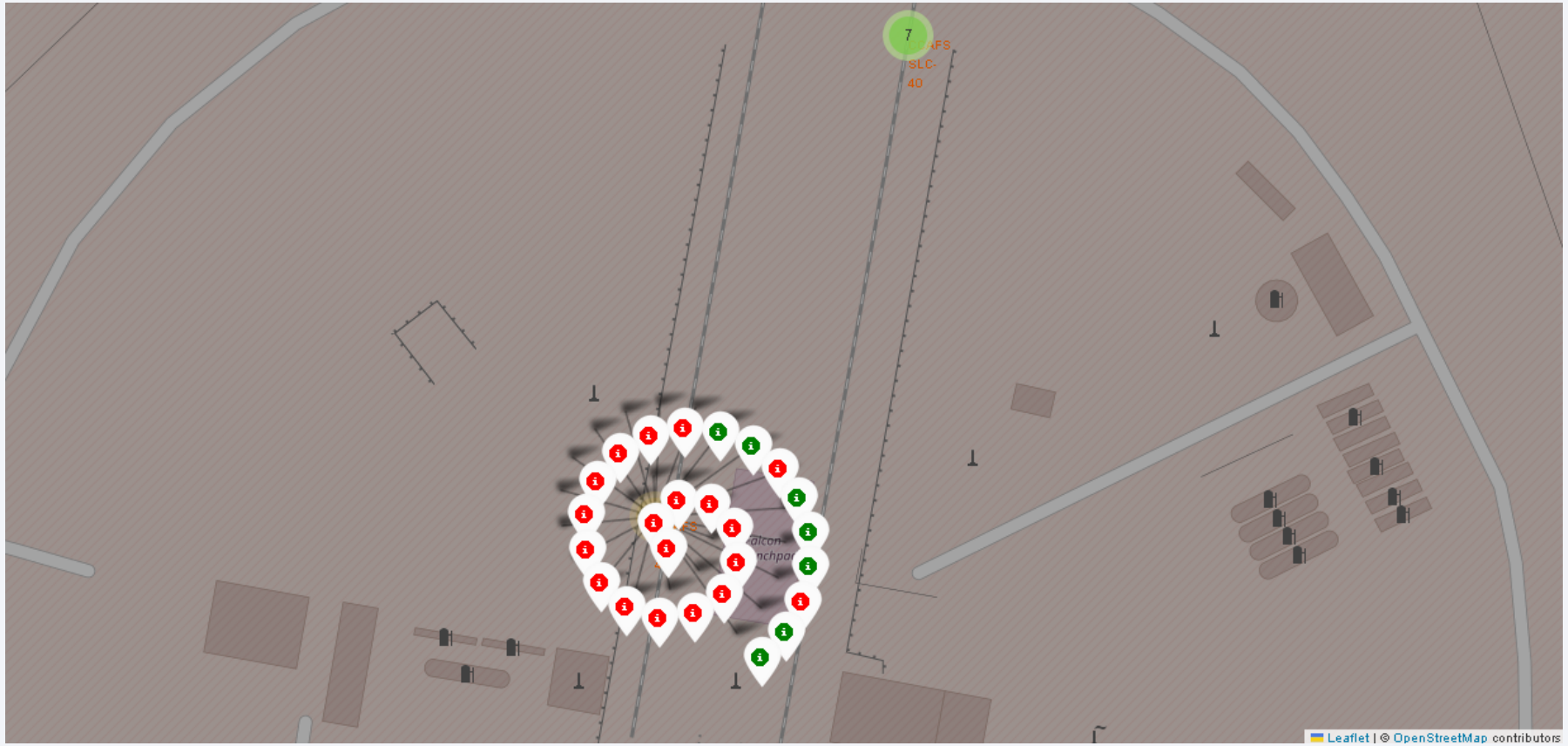
# Launch Sites Proximities Analysis

# All Launch Site




- Those Launch site located near sea

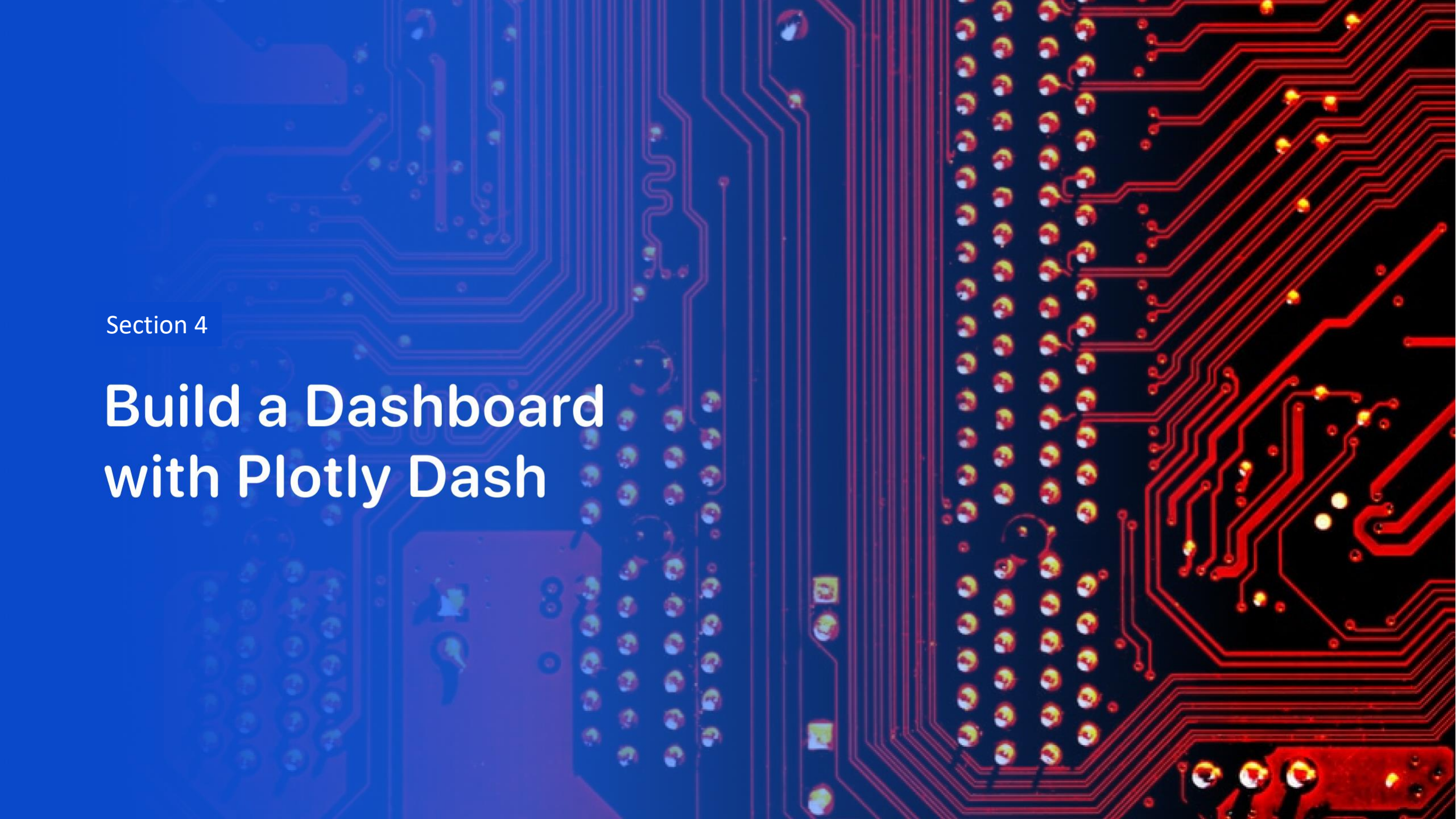
# Success and Failure Launch



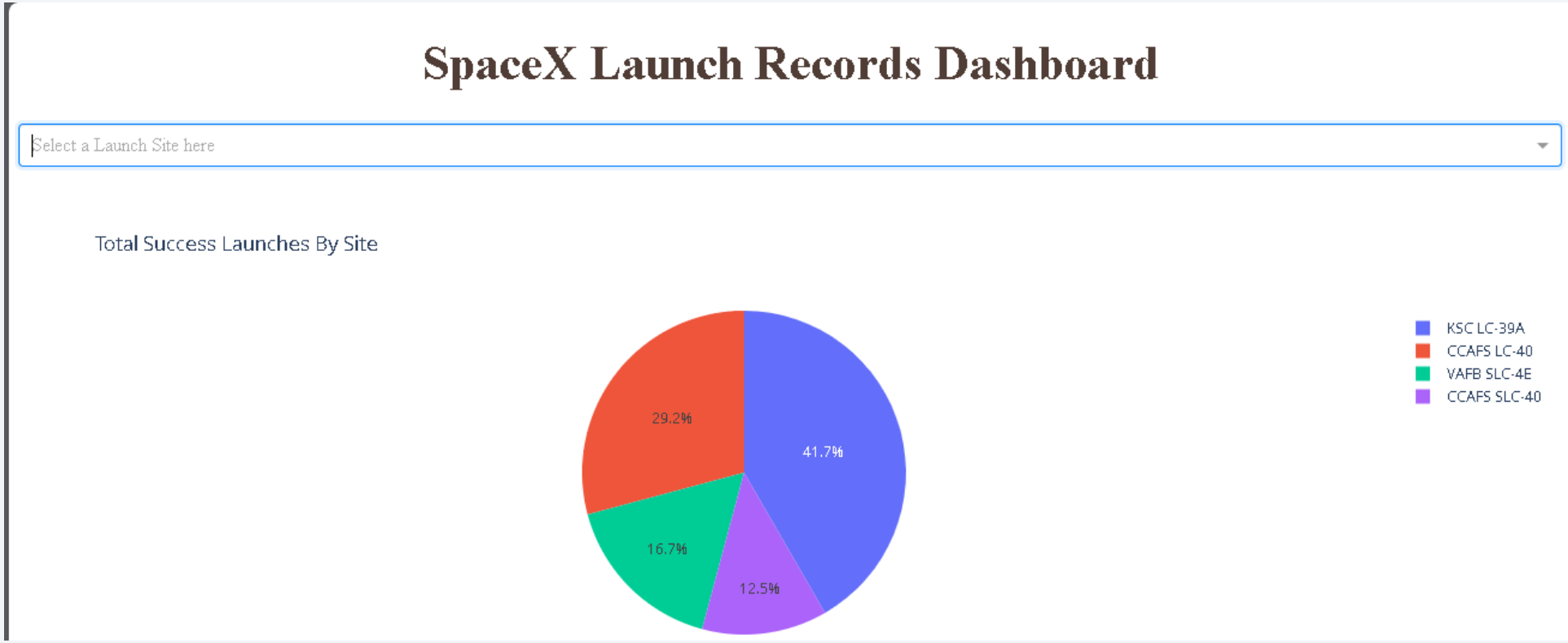

- Add marker color to see the spread data in map

# Distance to Railway and the other


- Using harvenstein distance to calculate distance in spherical object

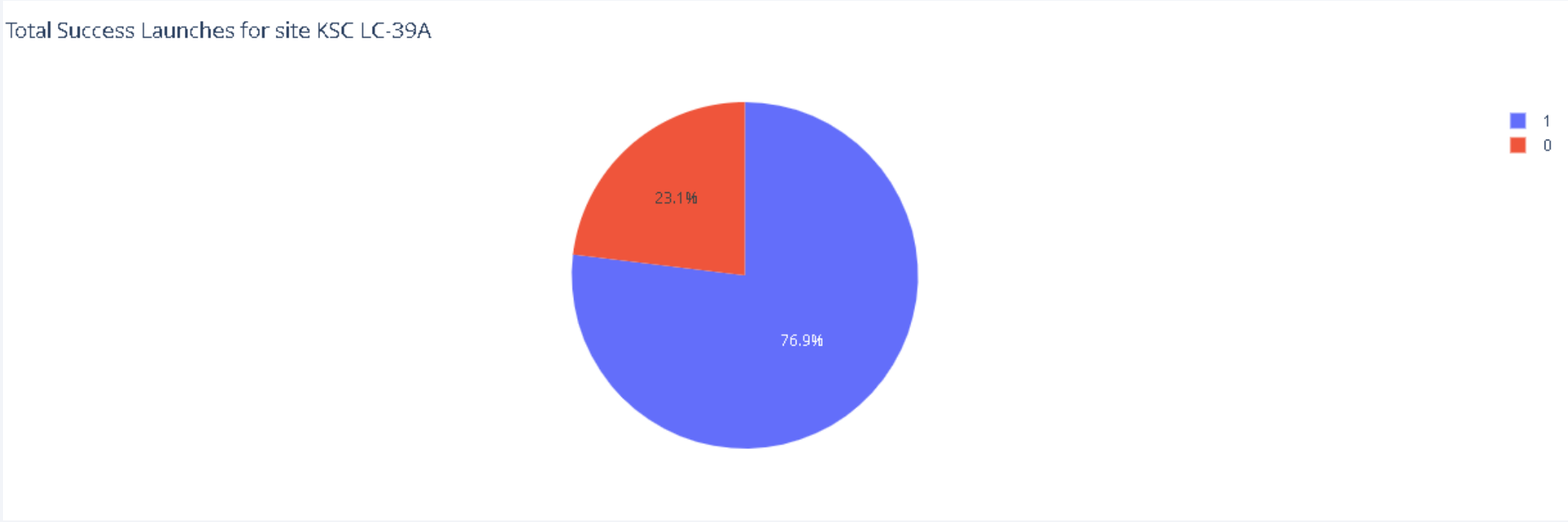

Section 4
Build a Dashboard with Plotly Dash

# Successfull Launch by launch site



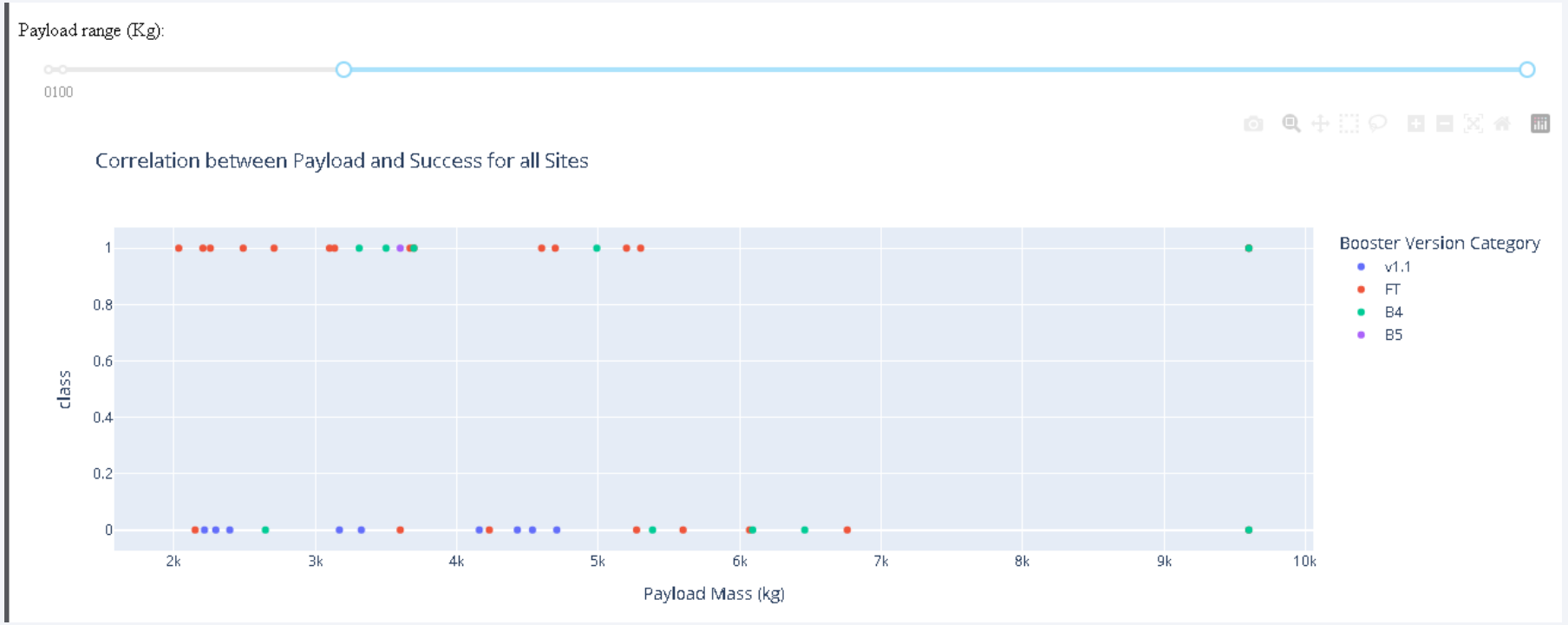

- Its important to know launch success distributed

# Highest Launch Ratio




- KSC LC-39 A has highest launch rasio : 76.9%

# Payload vs. Launch Outcome



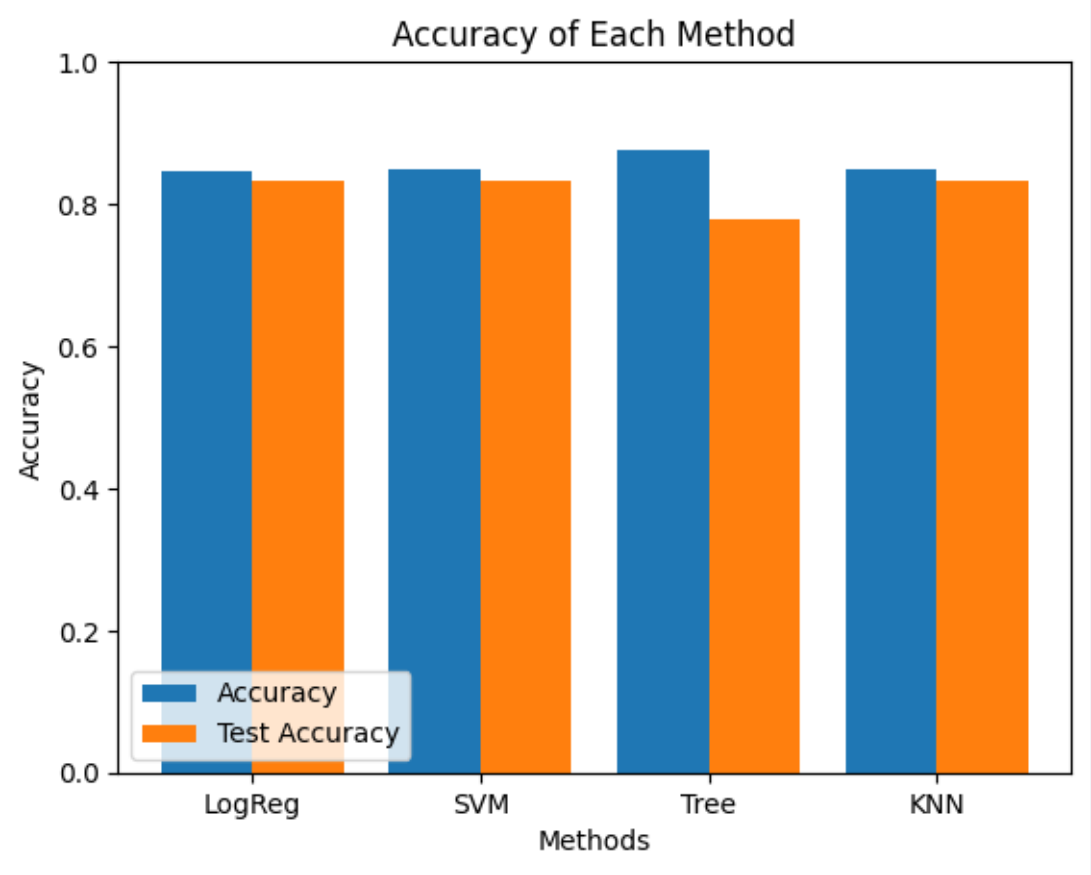

- Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

# Predictive Analysis (Classification)

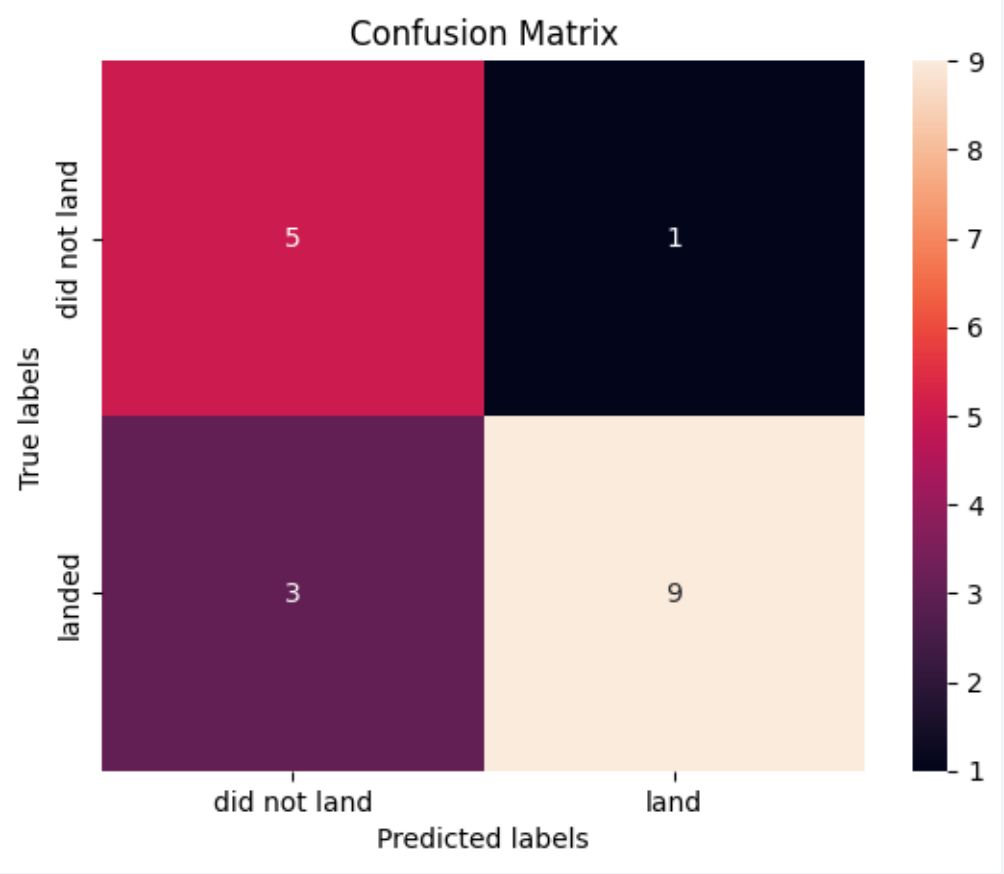
# Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%



# Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



# Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets
- Set Parameter on model is key to obtain best result and decision tree have the best result over the other model

# Appendix

- Some of code need to explore by ourshelf

# Thank you!