IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Rizky Fibryan Surya Nadhief
27 Januari 2023

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data Collection using API and Webscrapping
  - EDA using Visualization and SQL
  - Interactive Visual Analysis using Folium and Plotly Dash
  - Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Visualization
  - Interactive Map with Folium
  - Dashboard using Plotly
  - Predictive Analysis using Machine Learning

# Introduction

- Project background and context
  - The aerospace now a day is likely going more commercial rather than before, the problem for this was the expensive cost that needed. SpaceX one of the most successful aerospace provider is more efficient and inexpensive rather than other provider, also SpaceX is can recover the first stage, so the first stage can be reused.
  - As a data scientist, we will analyze the SpaceX data of Falcon 9 and evaluate the finding, so we can determine the new aerospace provider SpaceY is capable to be the competitor of SpaceX
- Problems you want to find answers
  - What the key features to determine price in each launch ?
  - What the things that affect the success rate of SpaceX launch ?
  - What the best site to be used for launch process ?
  - Can we predict the outcome of every launch using machine learning ?

Section 1

# Methodology

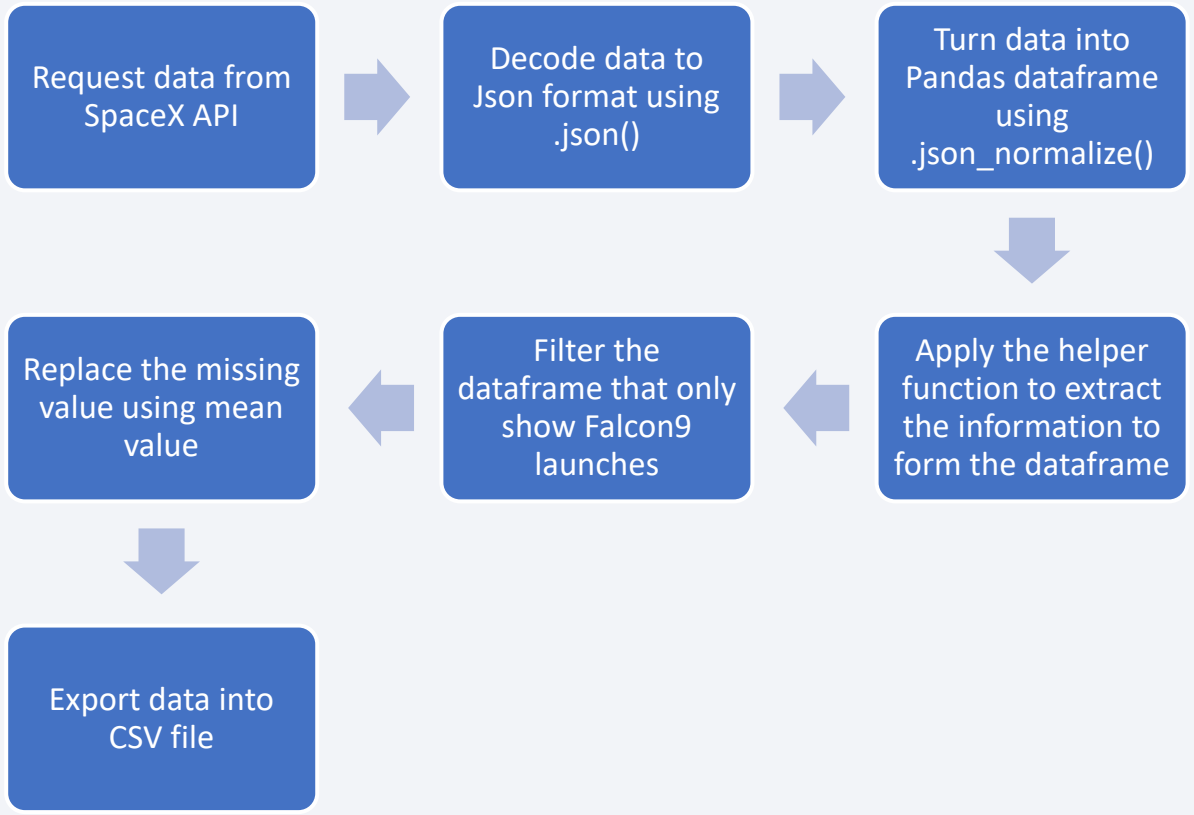
# Methodology

## Executive Summary

- Data collection methodology:
  - Request the data from SpaceX API
  - Webscrapping the data from Wikipedia
- Perform data wrangling
  - Filter the data to only Falcon9 launches data
  - Clean the data to fix the missing value
  - Create a label for landing outcome
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Transform the data set and split between train and test data
  - Train the data using different models to find the best parameter and score

# Data Collection – SpaceX API

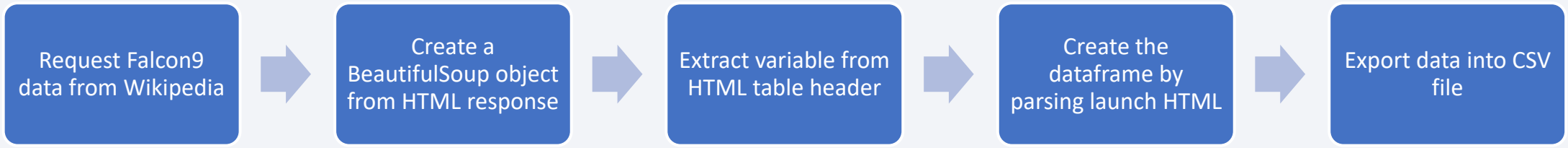
- Performing the data collection process by requesting the data from SpaceX API
- The data that has been collected was getting filtered to only show the Falcon9 launches record




Data Collection - SpaceX API

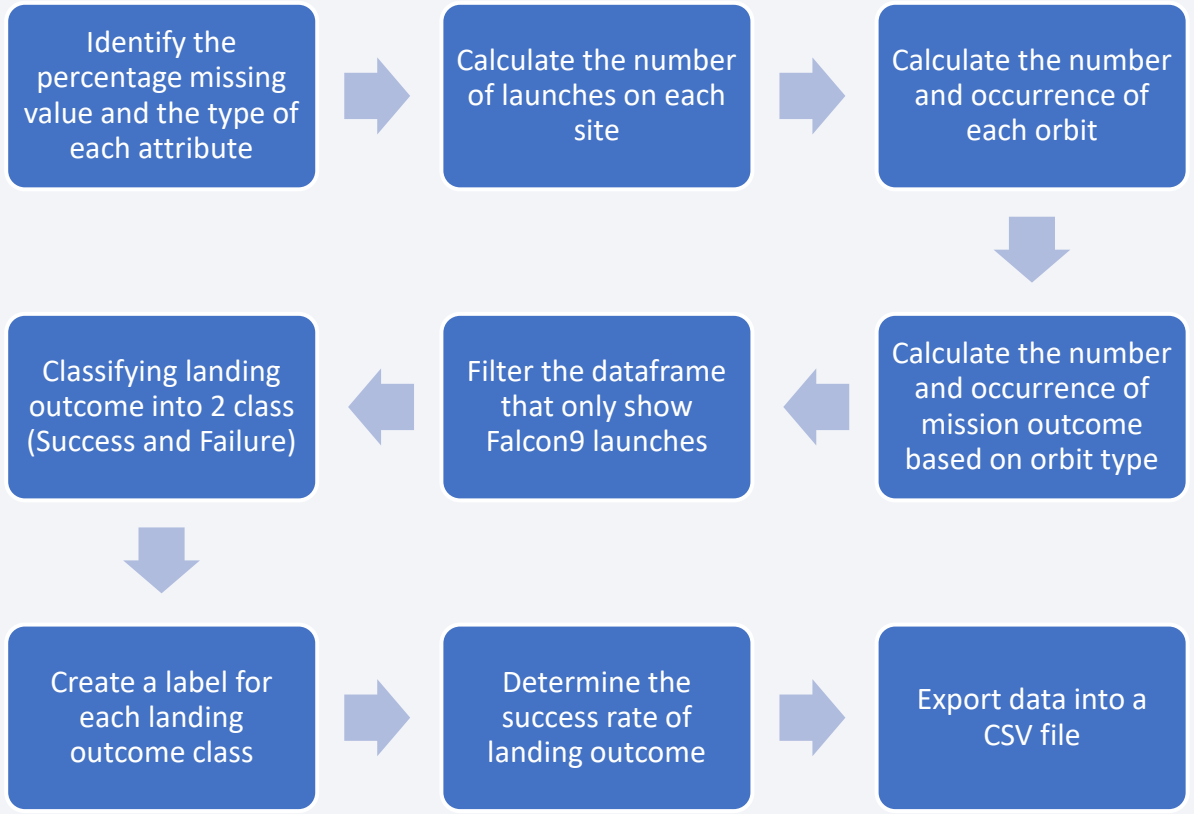
# Data Collection - Scraping

Performing the data collection process using the webscrapping method from Falcon9 Wikipedia page



# Data Wrangling

Identify the percentage of missing value in each attribute, and which column are numerical or categorical, also create a landing outcome classification class.



# EDA with Data Visualization

- Use a scatter chart to plotting out :
  - Payload Mass vs Launch Site
  - Payload Mass vs Orbit
  - Flight Number vs Launch Site
  - Flight Number vs Orbit
- Use a bar chart to visualize :
  - Relationship between success rate for each orbit type
- Use a line chart to plotting out :
  - Trends of success rate over the year
- Create a dummy variable for each categorical columns

# EDA with SQL

Executing the SQL queries to :

- Displaying the names of the unique launch sites
- Displaying 5 records of the launch site location that start with ‘CCA’
- Displaying a total payload mass that carried by booster launched by NASA(CRS)
- Displaying the average payload mass carried by booster version F9 v1.1
- List the date when the first successful ground pad landing was achieved
- List the name of booster that have success landing in drone ship with payload mass between 4000 and 6000 kg
- List the total number of successful and failure missions
- List the booster version that carried the maximum payload mass
- List the record month name, landing outcome, booster version and launch site of failure landing in 2015
- Rank the successful landing between the date of 04-06-2010 to 20-03-2017 in descending order

# Build an Interactive Map with Folium

Performing an interactive map for visual analytics using folium :

- Create a map visual object using folium
- Adding the site location in the map using folium.circle and folium.map.marker
- Create a cluster to divide whether the launch success or fail for each sites on the map using MarkerCluster()
- Mark a few proximity points and calculate the distance of launch site to the proximity points
- Draw a PolyLine between launch site and proximity point

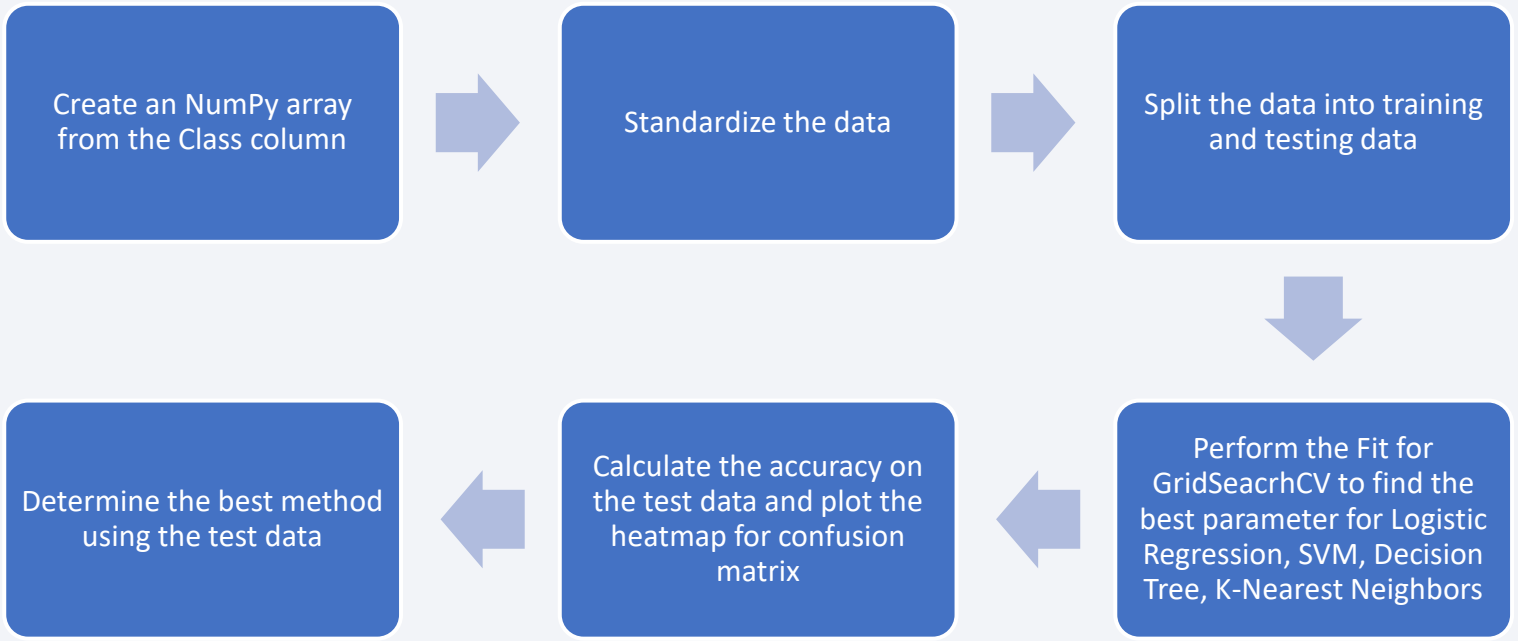
# Build a Dashboard with Plotly Dash

Build a dashboard application to perform interactive real-time visual analytics on SpaceX launch data :

- Create a dropdown menu for choosing the different launch site data that want to displayed
- Create a pie chart to visualize a success rate of launch in each launch site
- Create a slider to easily choose the different range of payload mass that want to be displayed
- Plotting a scatter plot to visualize relationship between payload mass and booster version to the missions outcomes from each launch site

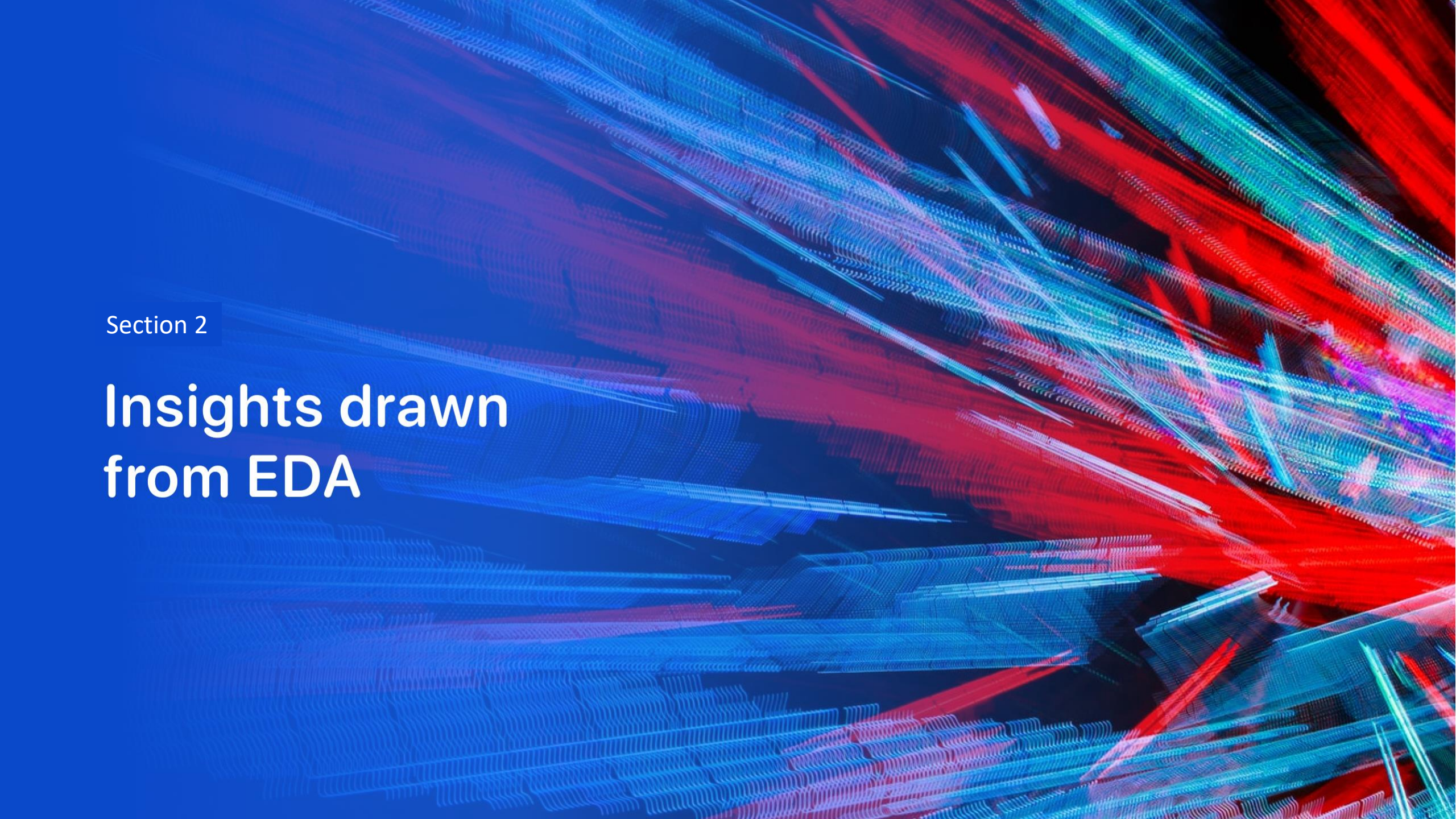
# Predictive Analysis (Classification)

Create a machine learning pipeline to predict whether the first stage lauch success or fail using few method



# Results

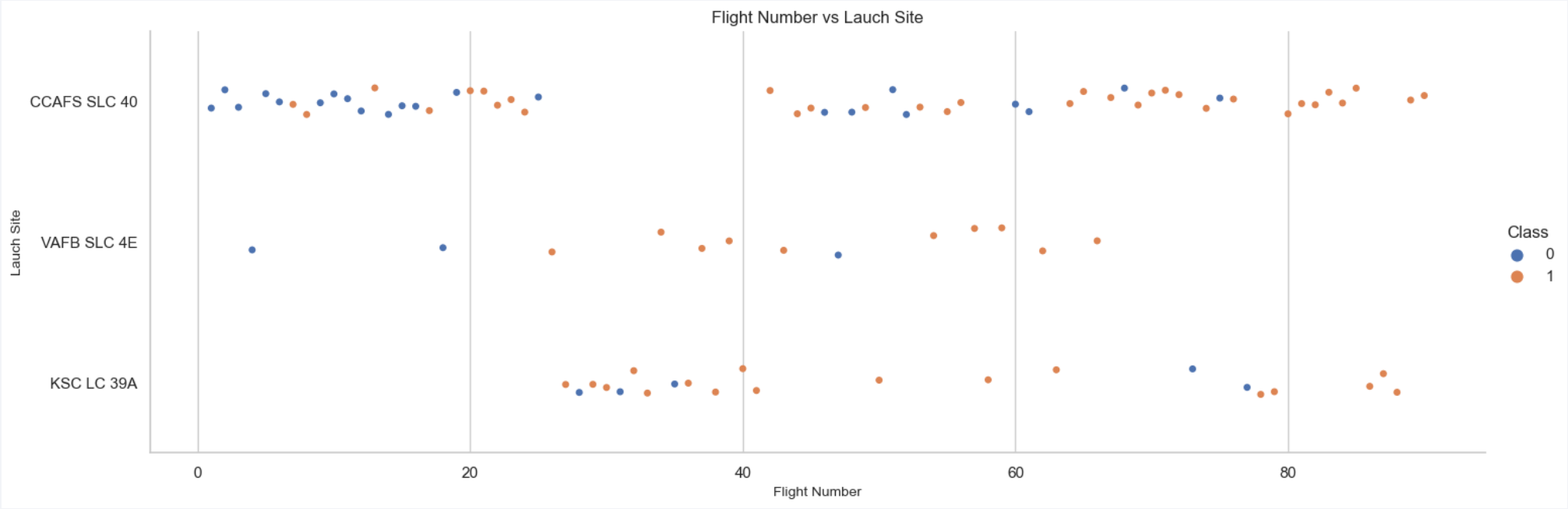
- Insight drawn form EDA
- Launch site proximity analysis
- Build a Dashboard with Plotly Dash
- Predictive Analysis


Section 2
Insights drawn from EDA

# Flight Number vs. Launch Site

Explanation :

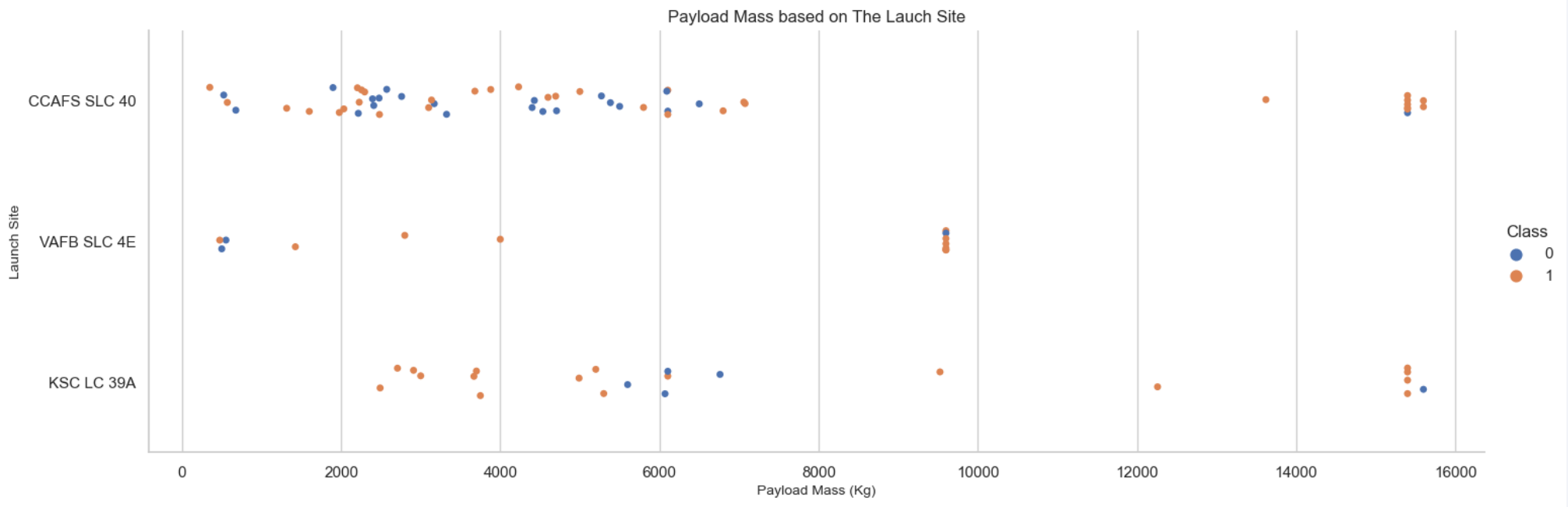
- CCAFS SLC-40 is the launch site with the most used launch site, however the success rate is not really high.
- As the flight number increasing, the landing of the first stage is likely successful.



# Payload vs. Launch Site

Explanation :

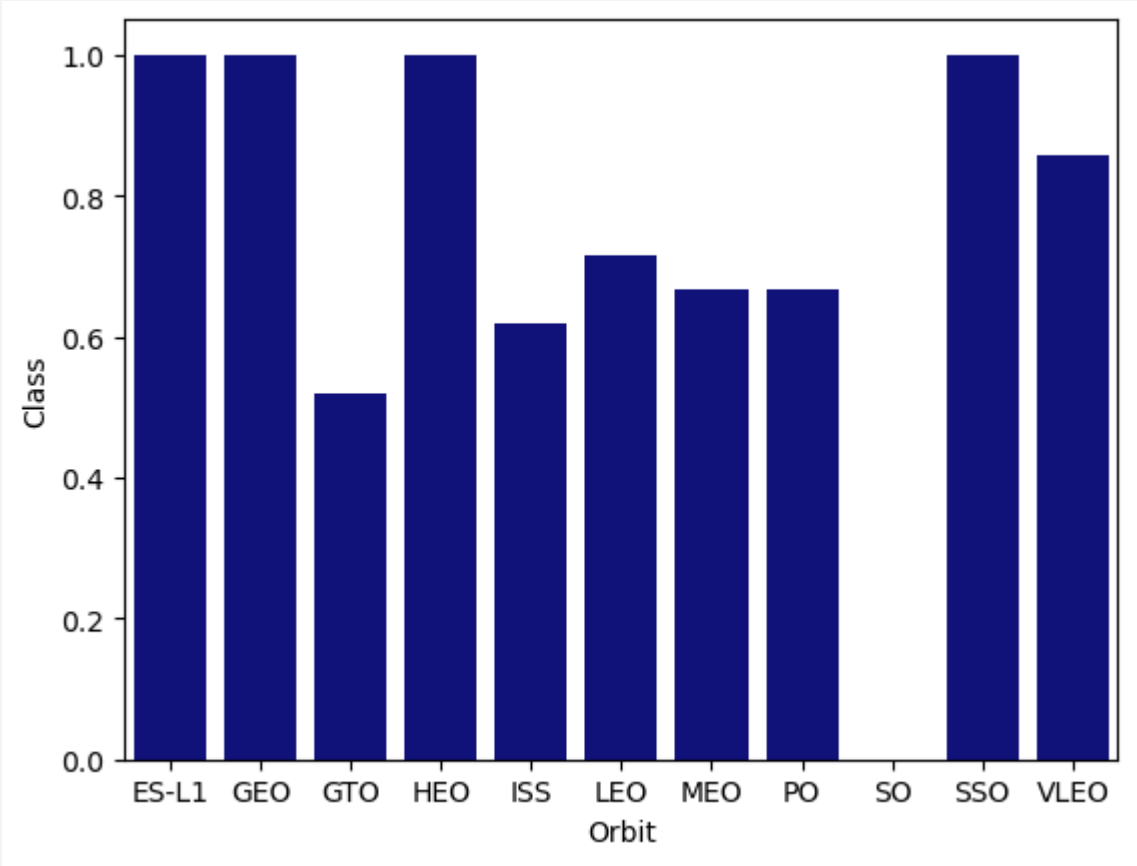
- Most of the rocket was carrying a payload under 8.000 kg.
- The success rate when rocket carrying a payload more than 8.000 kg is more higher than under 8.000 kg.
- There are no rocket launched in VAFB SLC-4E with payload more than 10.000 kg.



# Success Rate vs. Orbit Type

Explanation :

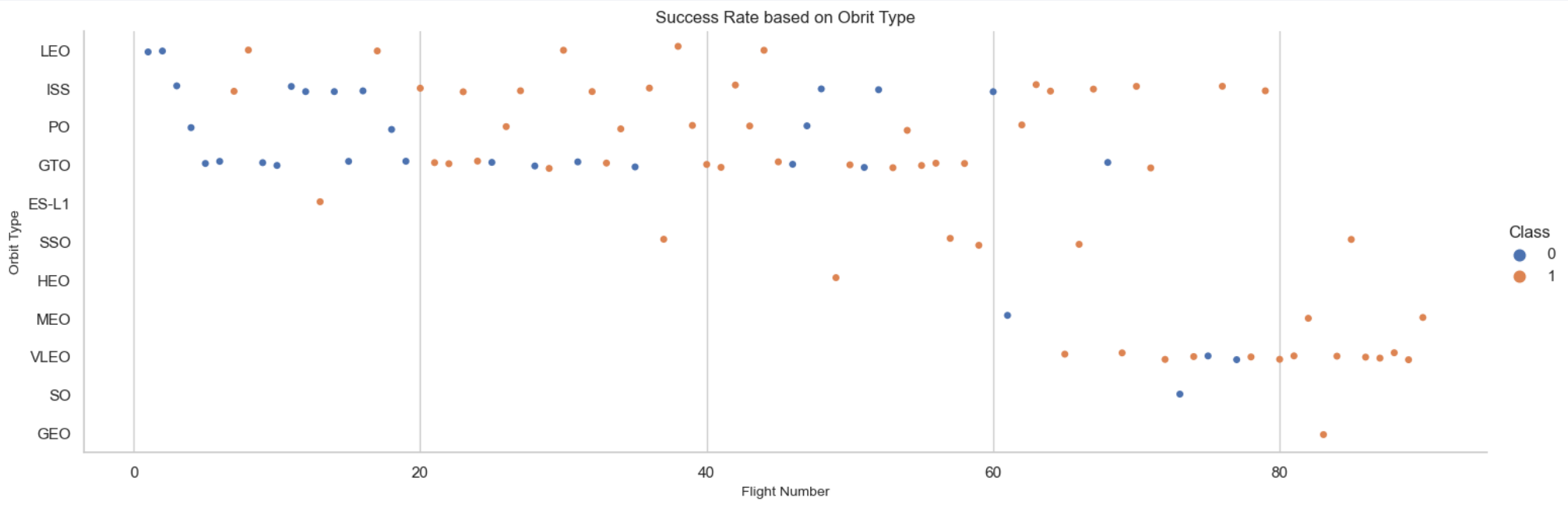
- ES-L1, GEO, HEO, and SSO orbit have highest success rate with 100% of success rate.
- SO orbit is having the lowest success rate with 0% success rate.
- The majority of orbit have success rate above 60%, except GTO and SO orbit.



# Flight Number vs. Orbit Type

Explanation :

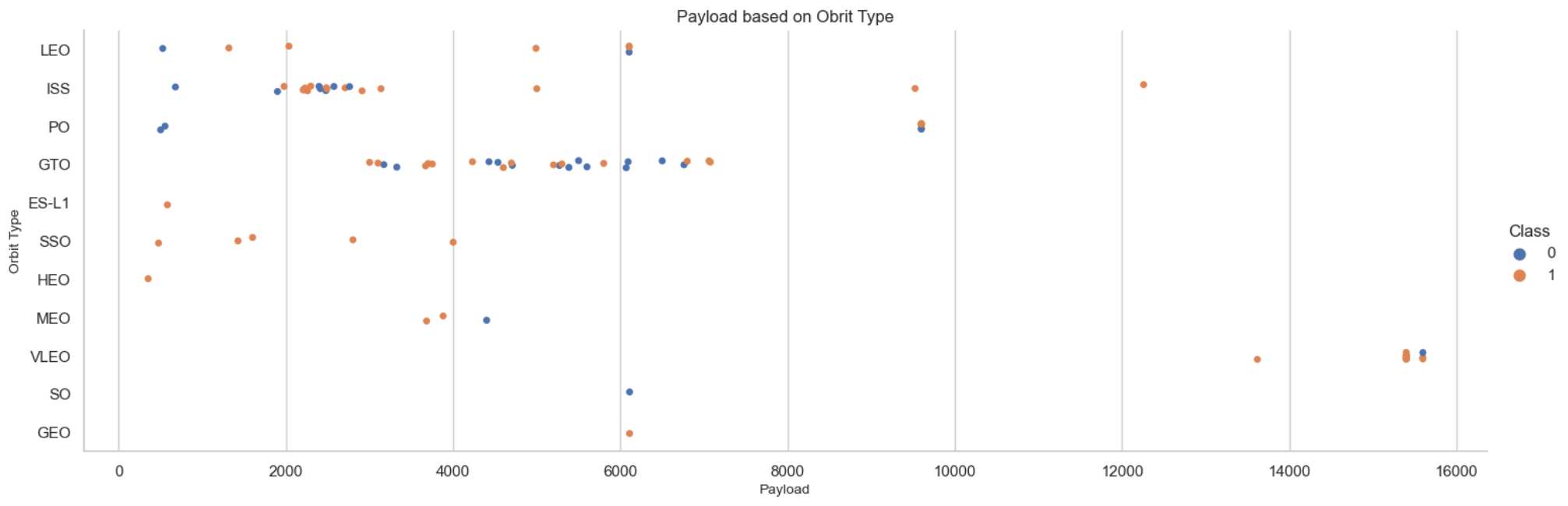
- Same as the launch site, as the flight number increase the first stage likely land successfully. Except the ISS orbit
- The LEO orbit success related to the number of flight
- In GTO orbit seems there is no relationship with the flight number



# Payload vs. Orbit Type

Explanation :

- With the heavy payload the successful landing rate is higher in ISS, PO, and VLEO orbit
- In SSO orbit the successful landing rate is 100% but the payload is less than 5.000 kg
- In GTO orbit, success or failure cannot be distinguished properly because both successes and failures are here



# Launch Success Yearly Trend

Explanation :

- The trend of success rate is increasing over the year from 2013 to 2019
- But the trend have decreased in 2018 and 2020

# All Launch Site Names

There is a 4 launch site that used to launch the rocket.

```
%sql SELECT distinct LAUNCH_SITE FROM SPACEXTBL
```

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

# Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Explanation :
  - All of the 5 records from the launch sites, LEO(ISS) is the common orbit that used to launch the rocket
  - Even though, the success rate of the mission is 100%, but the first stage is having 0% on landing success rate, neither fail or not having an attempt.

```
%sql select * from SPACEXTBL where LAUNCH_SITE like '%CCA%' LIMIT 5
```

| 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

- Calculate the total payload carried by boosters from NASA
- Explanation :
  - The total payload mass that carried by booster launched from NASA(CRS) is 48.213 Kg

```
%sql select sum(PAYLOAD_MASS__KG_) as 'Total Payload Mass Carried by Boosters Launched by NASA (CRS)' from SPACEXTBL where CUSTOMER like '%NASA (CRS)%'
```

| Total Payload Mass Carried by Boosters Launched by NASA (CRS) |
| --- |
| 48213 |

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Explanation :
  - The average payload mass that carried by booster version F9 v1.1 is 2.534,66 Kg

```
%sql select avg(PAYLOAD_MASS__KG_) as 'Average Payload Mass Carried by Booster Version F9 v1.1' from SPACEXTBL where BOOSTER_VERSION like '%F9 v1.1%'
```

| Average Payload Mass Carried by Booster Version F9 v1.1 |
| --- |
| 2534.6666666666665 |

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Explanation :
  - 1 May 2017 is the first time that first stage is successfully landed on a ground pad

```
%sql select min(DATE) as 'First Successful Landing' from SPACEXTBL where LANDING__OUTCOME like '%Success%'
```

| First Successful Landing |
| --- |
| 01-05-2017 |

# 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
- Explanation :
  - There are 4 booster version that successfully landed on a drone ship when carrying payload mass between 4.000 to 6.000 kg

```
%sql select BOOSTER_VERSION from SPACEXTBL where LANDING__OUTCOME like '%Success (drone ship)%' and PAYLOAD_MASS__KG_ between 4000 and 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
- Explanation :
  - The data have a total 101 outcome, with 1 failure and rest was success. But there are 1 mission outcome was success but the payload status was unclear.

```
%sql select MISSION_OUTCOME, count(MISSION_OUTCOME) as 'Number of Outcome' from SPACEXTBL GROUP BY MISSION_OUTCOME
```

| Mission_Outcome | Number of Outcome |
|---|---|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Explanation :
  - There was 12 booster version that can carried the maximum payload mass.

```
%%sql select BOOSTER_VERSION as 'Booster Version', PAYLOAD_MASS__KG_ as 'Payload Mass' from SPACEXTBL
        where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

| Booster Version | Payload Mass |
|---|---|
| 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 |

# 2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Explanation :
  - The was 2 failed landing at drone ship in 2015, that happens in January and April.

```
%%sql select substr(DATE, 4, 2) as 'Month', LANDING__OUTCOME as 'Status', BOOSTER_VERSION as 'Booster Version', LAUNCH_SITE as 'Launching Site' from SPACEXTBL
      where substr(DATE, 7, 4) = '2015' and LANDING__OUTCOME like '%Failure%'
```

| Month | Status | Booster Version | Launching 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

```
%%sql select LANDING__OUTCOME, count( * ) as 'Total' from SPACEXTBL
      where DATE between '04-06-2010' and '20-03-2017' group by LANDING__OUTCOME order by Total DESC;
```

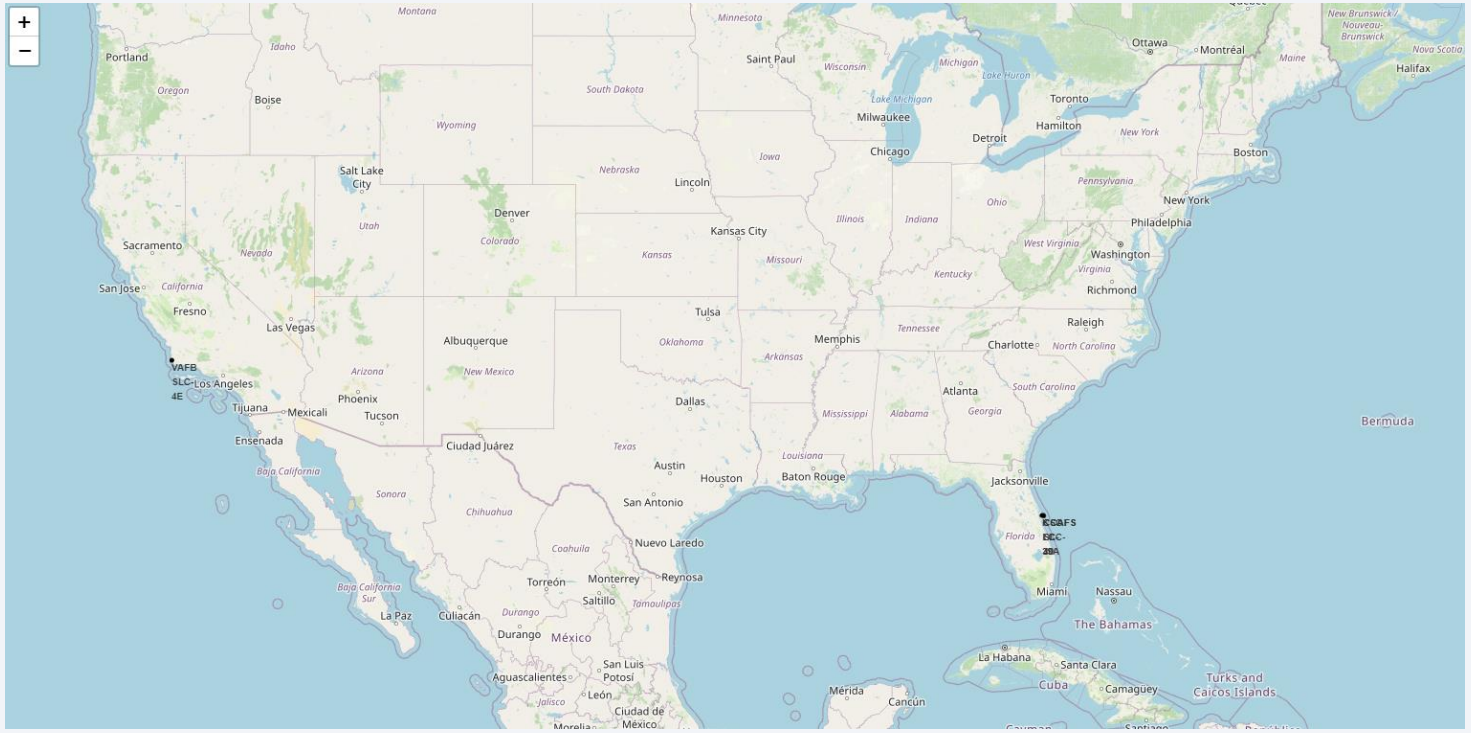
| Landing__Outcome | Total |
|---|---|
| Success | 20 |
| No attempt | 10 |
| Success (drone ship) | 8 |
| Success (ground pad) | 6 |
| Failure (drone ship) | 4 |
| Failure | 3 |
| Controlled (ocean) | 3 |
| Failure (parachute) | 2 |
| No attempt | 1 |

Section 3

# Launch Sites Proximities Analysis

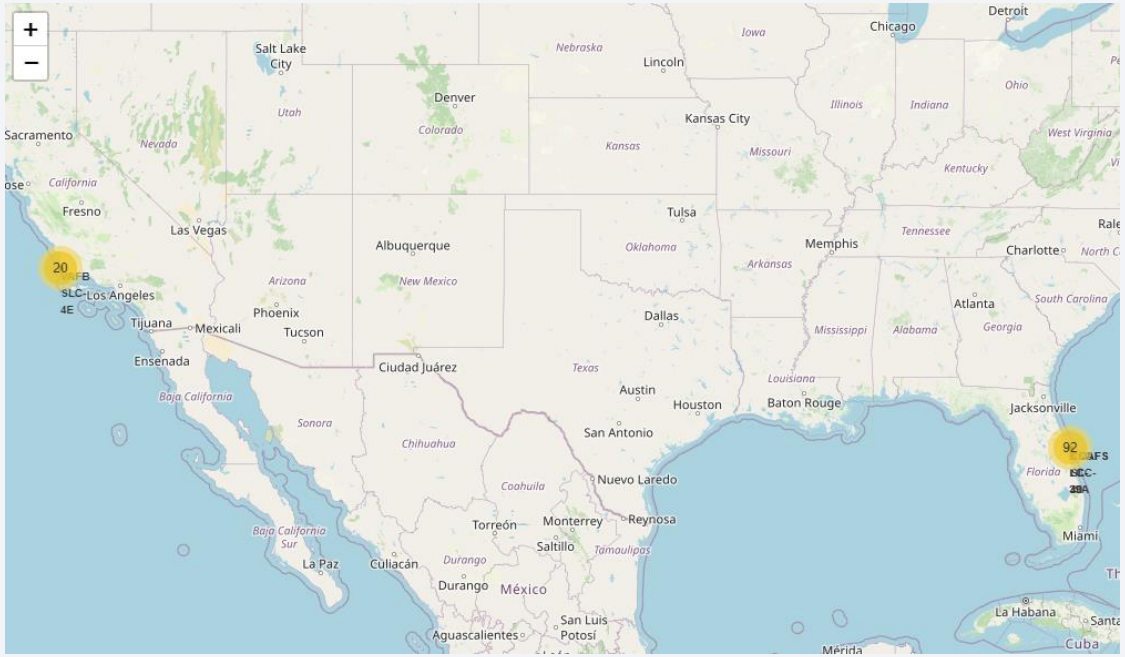
# Launch Site Location Map

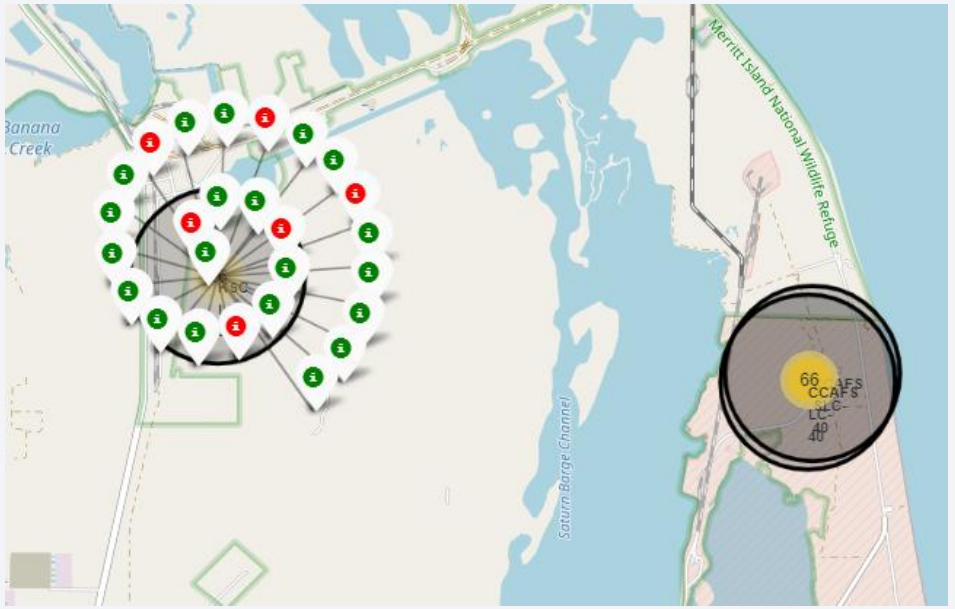
- All the launch site was located near to the coast.
- Most of the lauch site was located in the east coast of U.S.A.
- All the lauch site is proximity located close to the equator line.



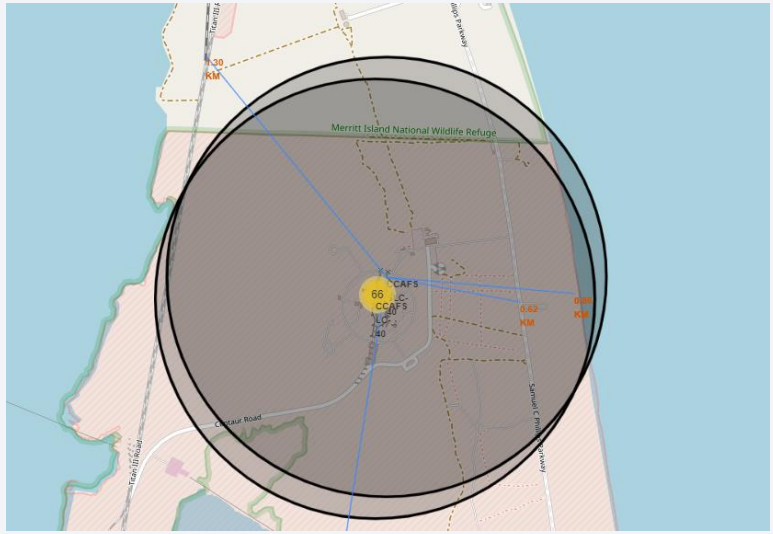
# Launches Status for Each Site on Map

- Color-labeled marker is used to determine the status : Green – Successful, Red – Failure
- CCAFS LC-40 is launch site with most of rocket launches
- KSC LC-39A is the launch site with relatively highest rate of successful launch

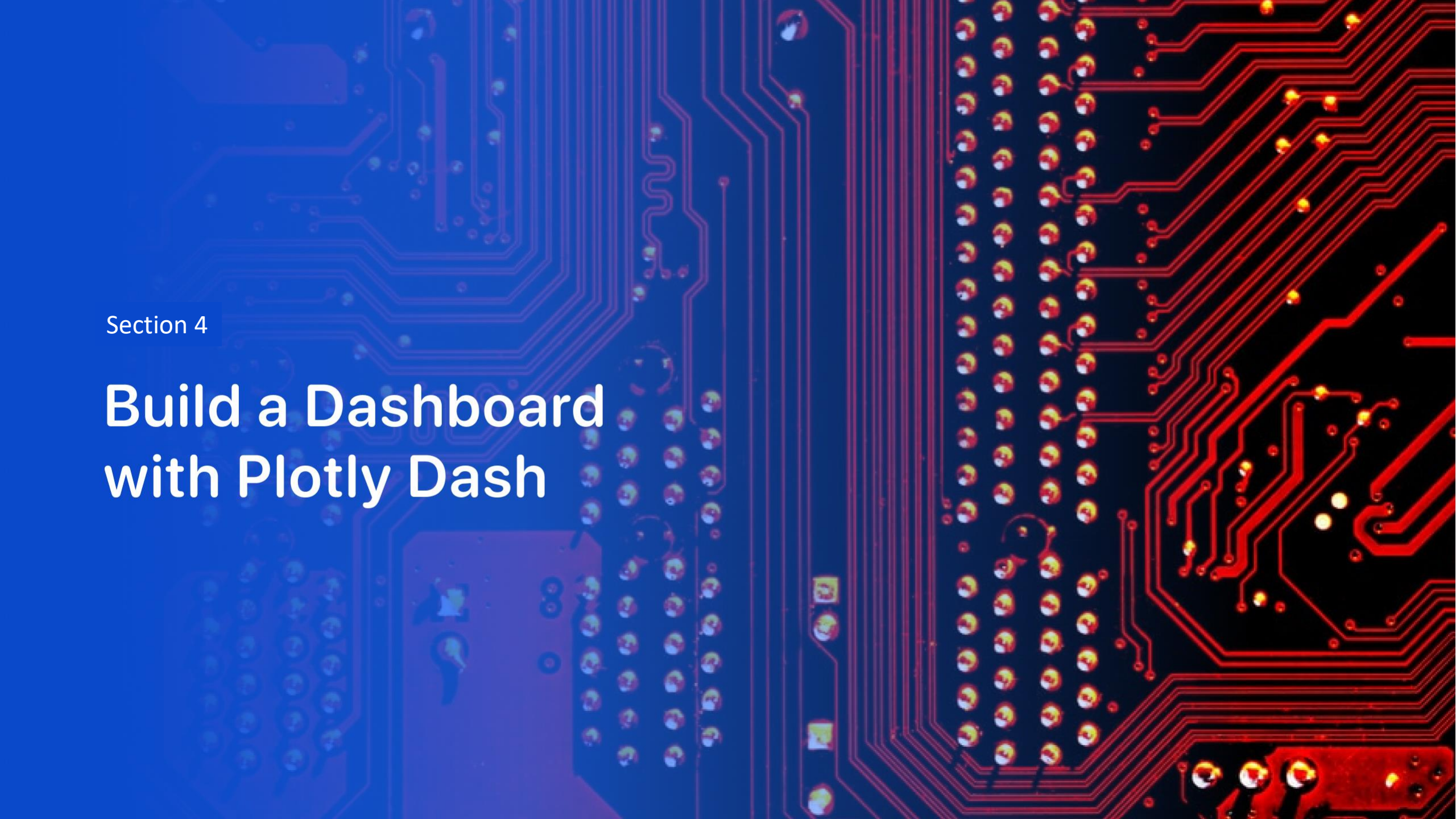


# Distance Between Launch Site to Closest Proximity Point

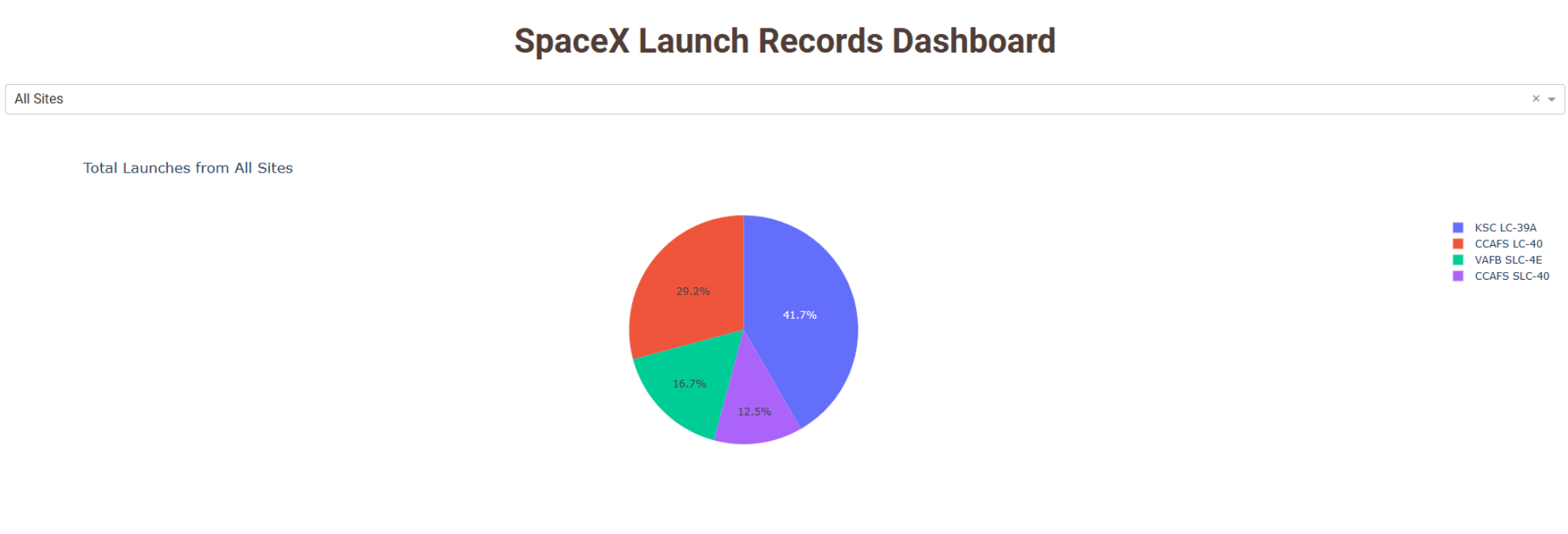



- The proximity distance between launch site to closest highway, railway, and coast line are within 1.5 km, it was fit for convenient transportation
- The closest city to CCAFS LC-40 is Cape Canaveral, with 19.14 km to the south of launch site. That could be dangerous if the launches of the rocket failed.


Section 4
Build a Dashboard with Plotly Dash

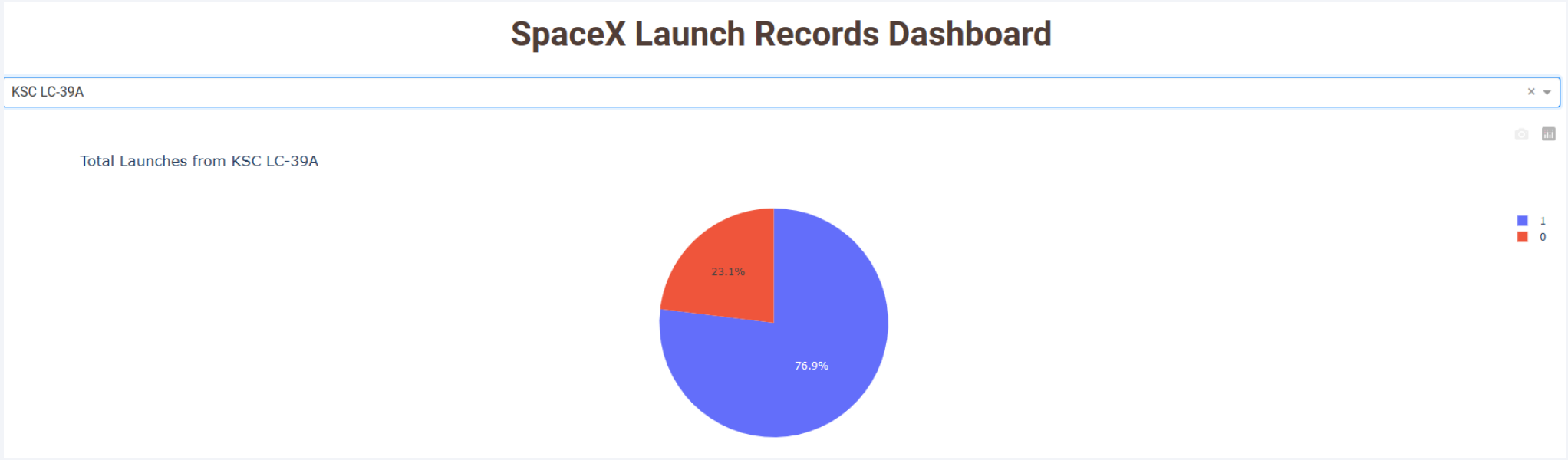
# Total Success Launches from All Sites

- KSC LC-39A is a launch site with the highest success rate among all the launch site.
- CCAFS SLC-40A is a launch site with the lowest success rate.



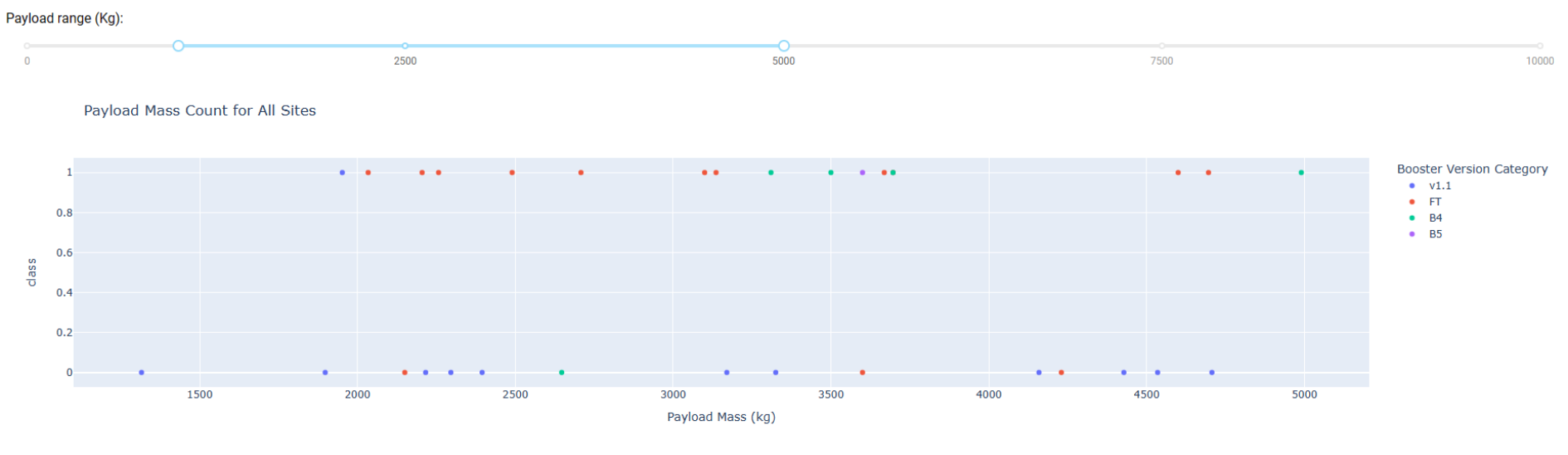
# Launch Site with Highest Successful Ratio

- KSC LC-39A was the highest success rate launch site with the success ratio 76.9%



# Payload vs Launch Outcome

- Most of the launches was carrying payload mass between 1.000 – 5.000 kg
- FT version is have the highest successful rate within payload range between 1.000 – 5.000 kg, and also have the most launches
- V1.1 version is have the most launches failed within payload range between 1.000 – 5.000 kg



# Heavy Payload vs Launch Outcome

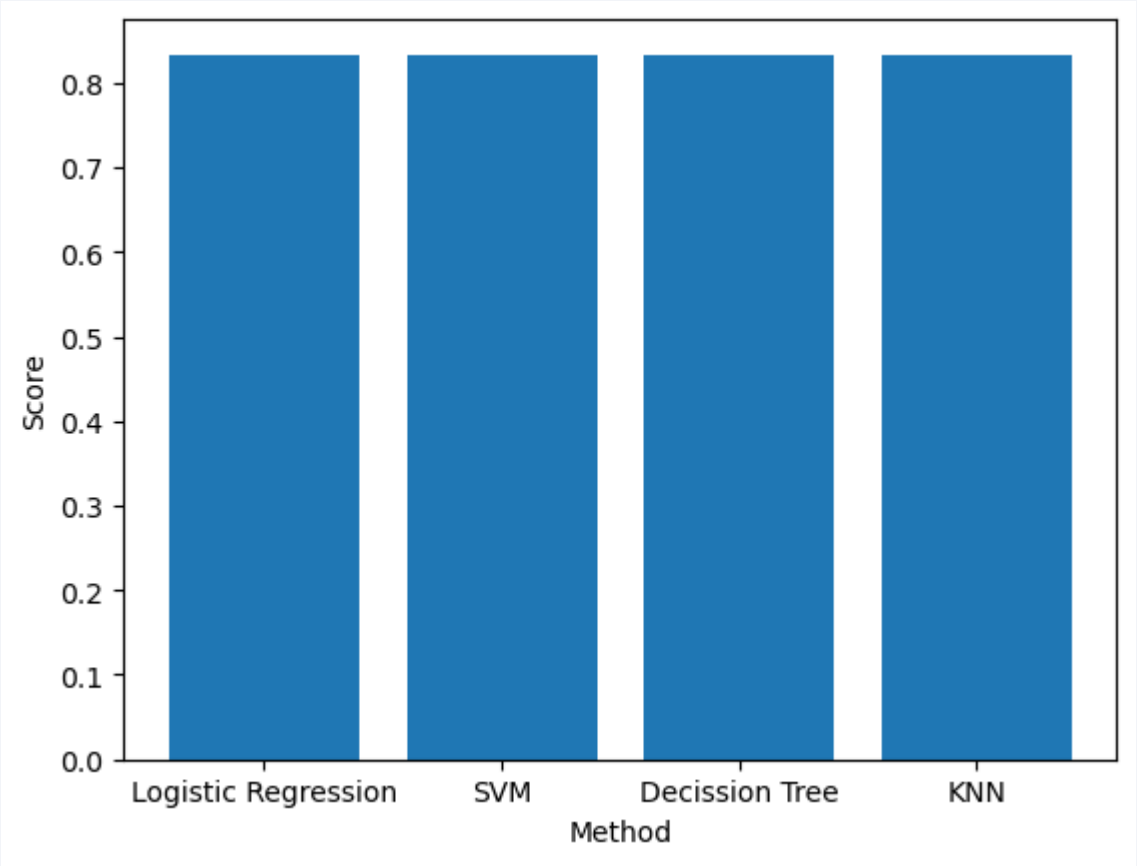
- Most of the launches was failed when carrying payload mass higher than 5.000 kg
- Only FT and B4 booster version that launch with the heavy payload mass higher than 5.000 kg
- Only B4 booster version that can successfully land with payload higher than 9.000 kg

Section 5

# Predictive Analysis (Classification)

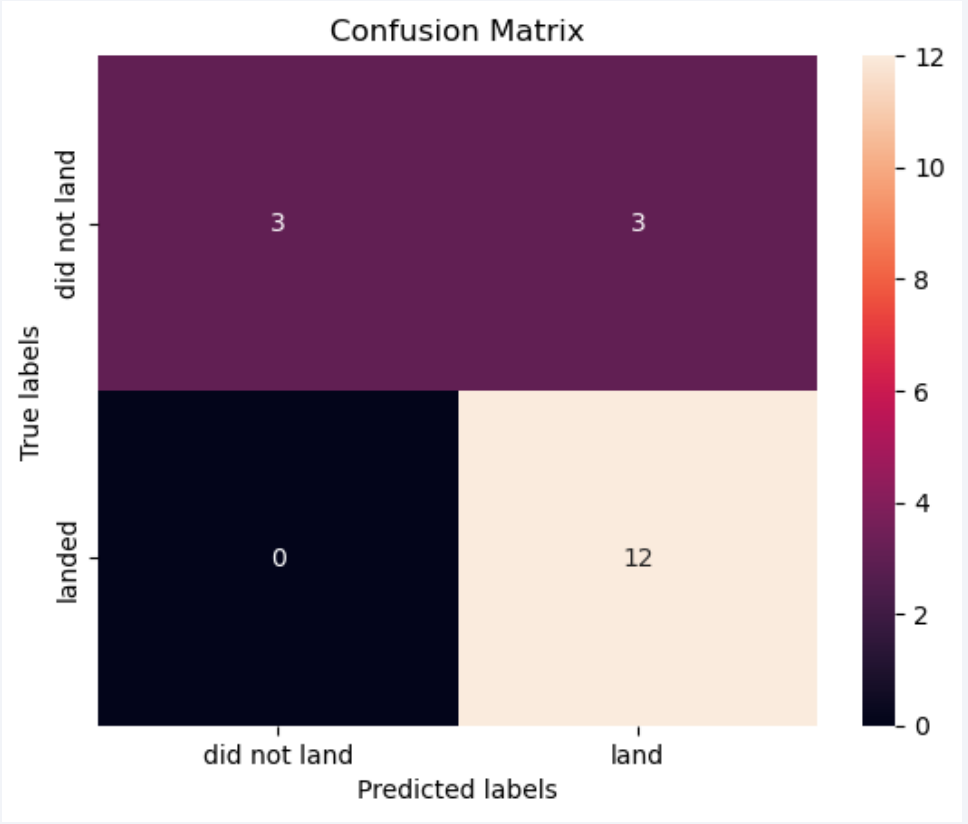
# Classification Accuracy

- The data was split between training and testing data, with the parameter test size was set to 0.2 and random state to 2.
- The method that used is logistic regression, SVM, Decision Tree, and K-Nearest Neighbors.
- Based on the bar chart, it was hard to find the best method to predict the target, because the sample data that used is to small.



# Confusion Matrix

- The built classification model have similar confusion matrix based on the test data.
- The major issue was the false positive, that located in the left part of confusion matrix.
- To gain the better result on predicting the result is needed the more bigger data set.



# Conclusions

- To determine price cost for every launch it was needed some of the key feature such as payload mass, launch site, orbital type, and booster version.
- The success rate of successful landing increase, as the number of flight increasing
- CCAFS SLC-40 is the most used launch site, but doesn't have a good successful rate
- Either, KSC LC-39 has the highest successful rate among other site
- ES-L1, GEO, HEO, and SSO orbit have highest success rate with 100% of success rate, with the mostly of orbit have success rate above 50% except SO orbit.
- FT booster version has the highest success rate
- The trend of success rate is increasing over the year

# Appendix

- Github Repository Link for This Project :

  Data Science Capstone  - SpaceX Falcon9 First Stage Landing Prediction

# Thank you!