IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

LEBEDIEVA OLEKSANDRA
23.11.2024

# Outline

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

# Executive Summary

- Summary of methodologies
  - Data collection
  - Data Wrangling
  - EDA with data visualization
  - EDA with SQL queries
  - Building interactive maps with Folium
  - Building a DashBoard with Plotly Dash
  - Predictive Analysis (Classification)
- Summary of all results
  - EDA results
  - Interactive analysis
  - Results of building classification models

# Introduction

## Project background and context:

One of SpaceX's Falcon 9 rocket's  key benefits is its reusability. The first stage of the Falcon 9 can land back on Earth, be refurbished, and reused for future missions. This significantly reduces launch costs. By reusing rockets, SpaceX can drive down the cost per launch, making space access more affordable and opening the door for more frequent missions. Falcon 9 has a high success rate, with numerous successful missions since its debut. This reliability makes it a trusted option for both government and commercial launches. It can carry large payloads to orbit, including satellites and supplies to the International Space Station (ISS), making it versatile for a wide range of missions. Falcon 9 is known for its fast turnaround times, as it can be reused quickly and efficiently between launches. SpaceX continually improves the Falcon 9 with upgraded engines, enhanced performance, and the integration of new technologies like the "Starlink" payload.

These advantages make Falcon 9 a leading option in the commercial space industry.

## Business problem:

Improving the prediction of a successful landing for Falcon 9 brings several key benefits:

More accurate landing predictions lead to higher success rates, allowing the rocket's first stage to be reused more frequently. This reduces the cost of launching and accelerates mission timelines.

Reliable predictions reduce the risk of landing failures, which could result in the loss of expensive rocket stages. Better predictions enable cost-efficient recovery operations, minimizing financial losses.

With improved landing success prediction, Falcon 9 can be prepared for reuse more quickly, reducing the time between launches and increasing the overall frequency of missions. Accurate landing forecasts allow for more complex and diverse mission planning, as the rocket's stage can be confidently landed and reused in different locations, including on autonomous droneships at sea/

**In essence, improved prediction of successful landings strengthens the overall efficiency, affordability, and reliability of SpaceX's operations.**

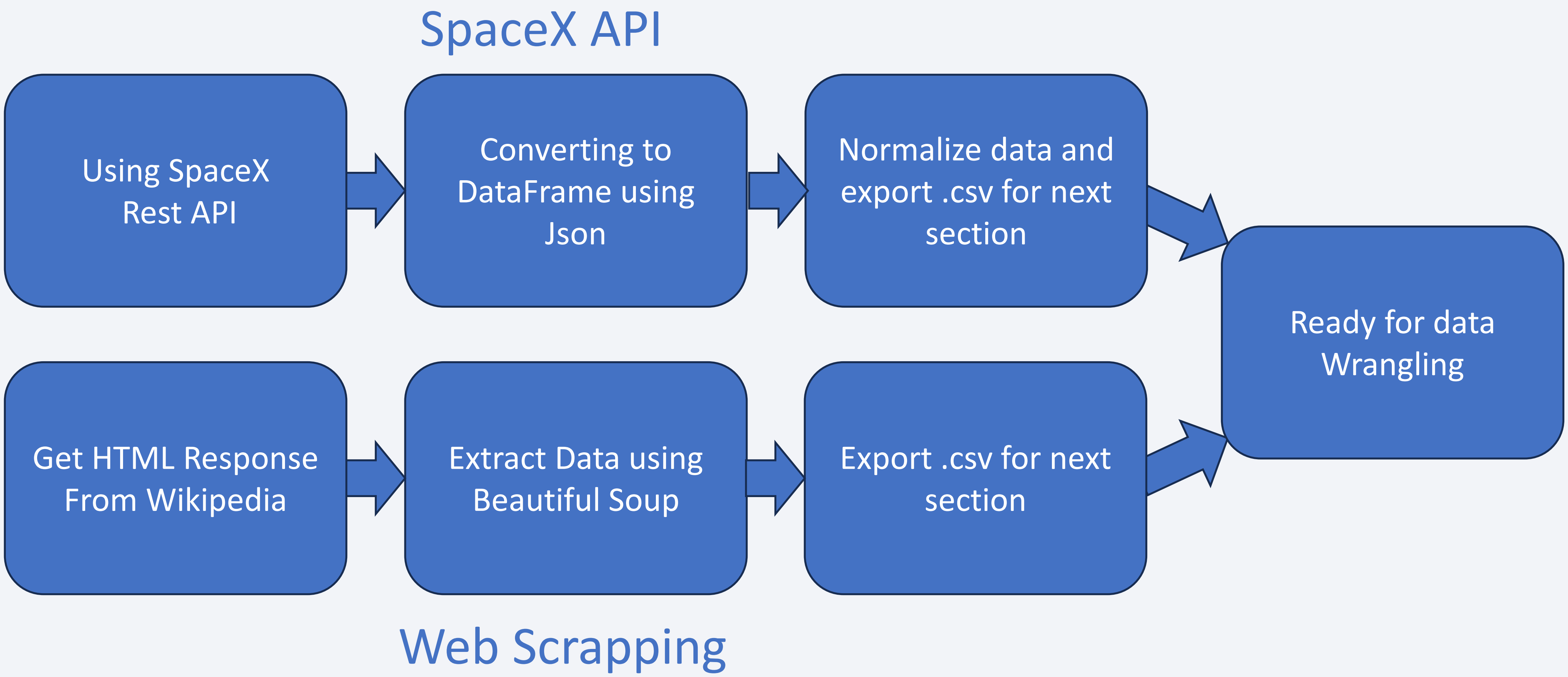
Section 1

# Methodology

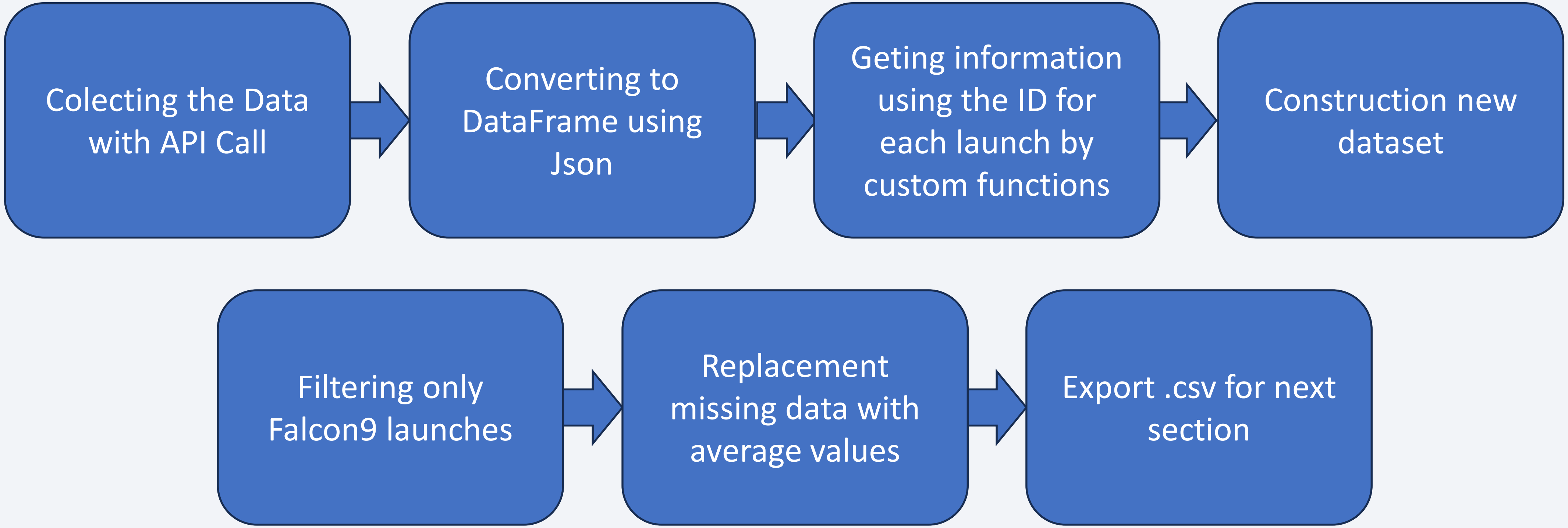
# Methodology

- Data collection methodology:
  - SpaceX Open-Source Rest API
  - Web Scraping from Wikipedia
- Perform data wrangling
  - Removing empty or irrelevant information
  - Transforming categorical data using One Hot Encoding for machine learning algorithms
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, K-Nearest Neighbors, Support Vector Machines and Decision Tree models have been developed to determine the most effective classification method.

# Data Collection

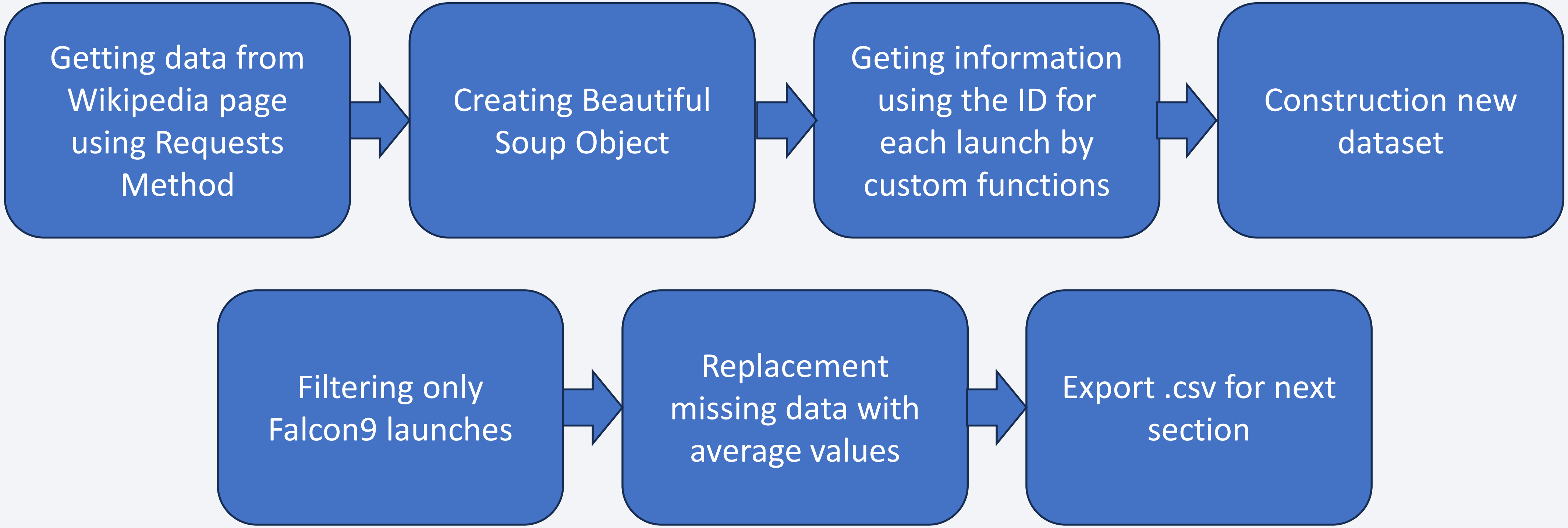


# Data Collection – SpaceX API



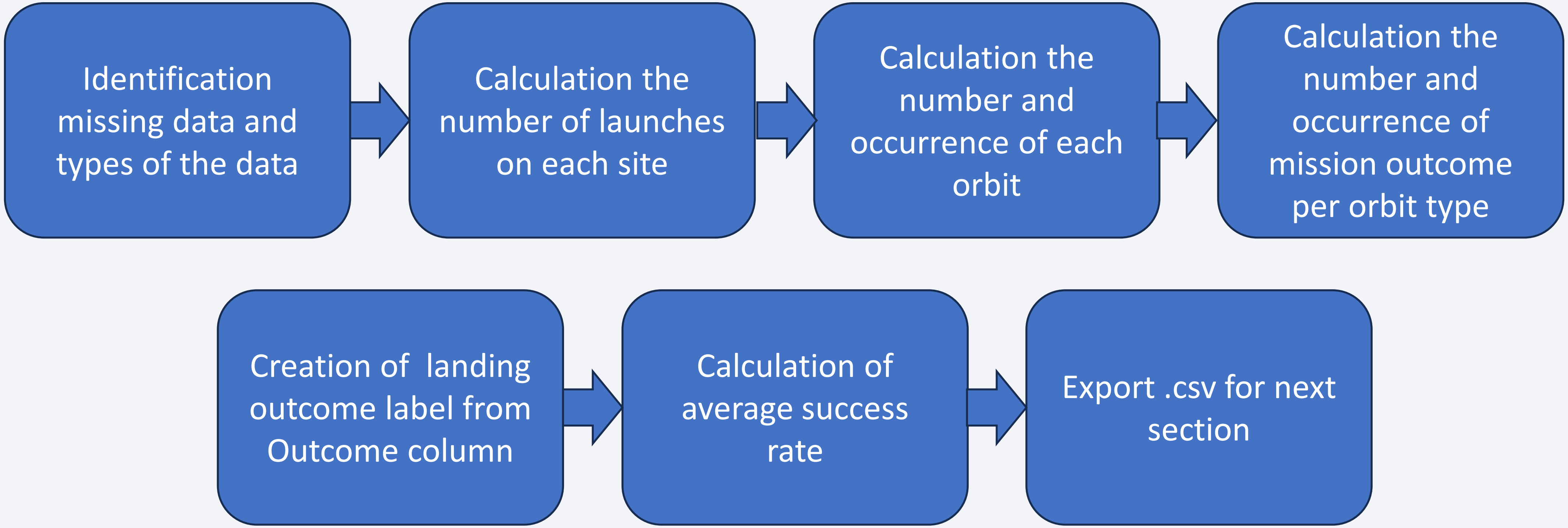

GitHub URL: [Collection_the_data.ipynb]

# Data Collection - Scraping




GitHub URL: [Webscraping.ipynb]

# Data Wrangling

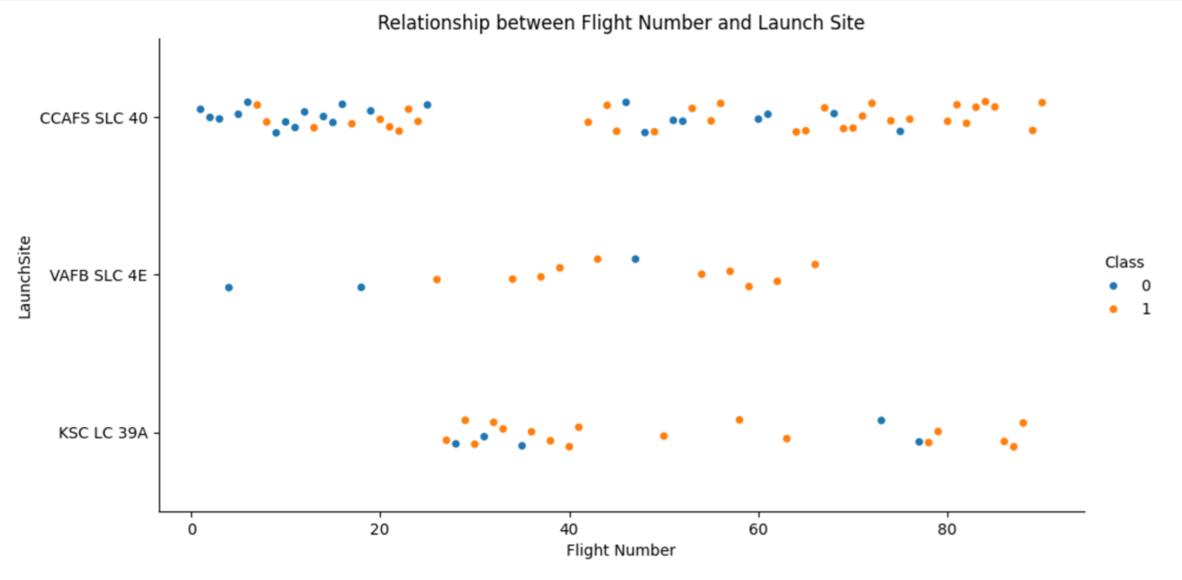



GitHub Link: [Data Wrangling.ipynb]()

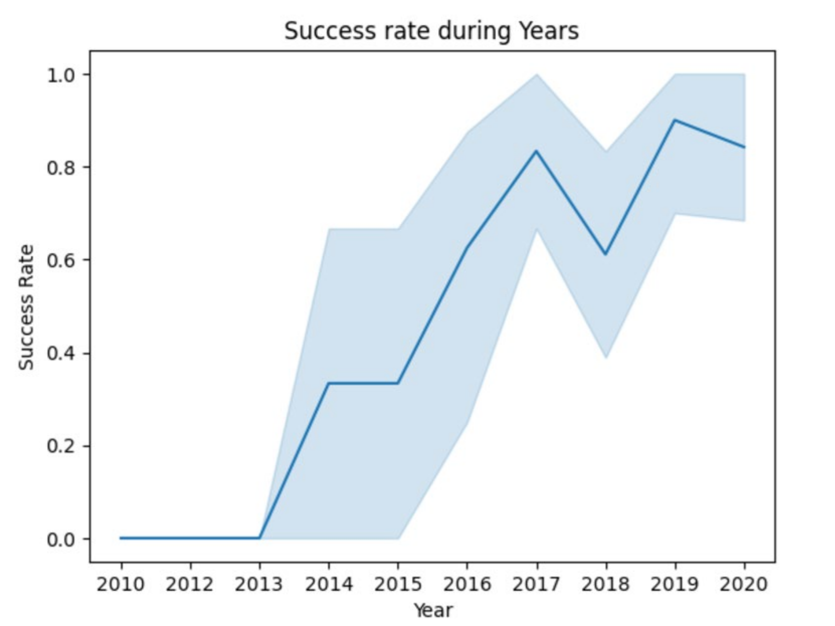
# EDA with Data Visualization

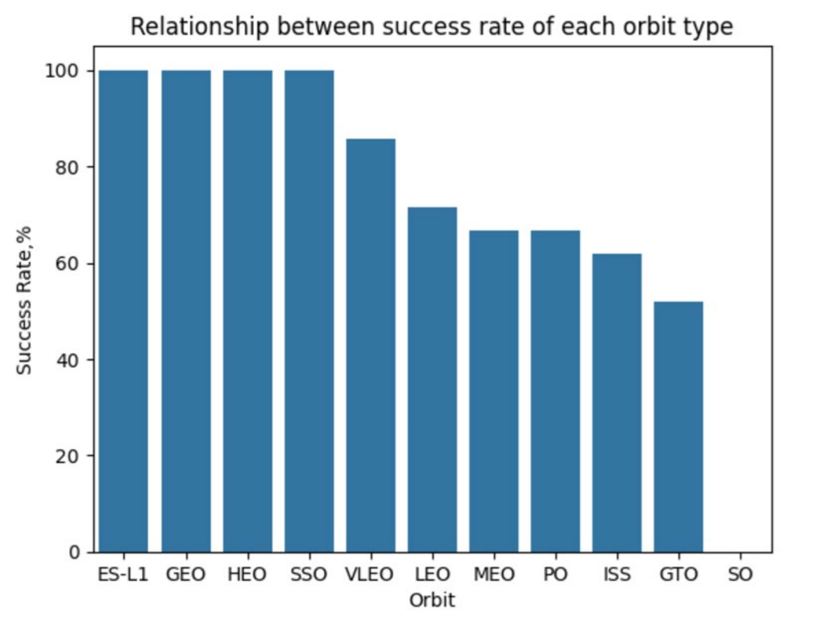
The relationships between Success Rate and other variables were considered. For better highlight were used:

- scatter plots
- bar chart
- line chart




GitHub Link: [Exploring and Preparing Data.ipynb]()



# EDA with SQL

## SQL queries performed include:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string 'CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved.
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster_versions which have carried the maximum payload mass.
- Listing the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Ranking 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

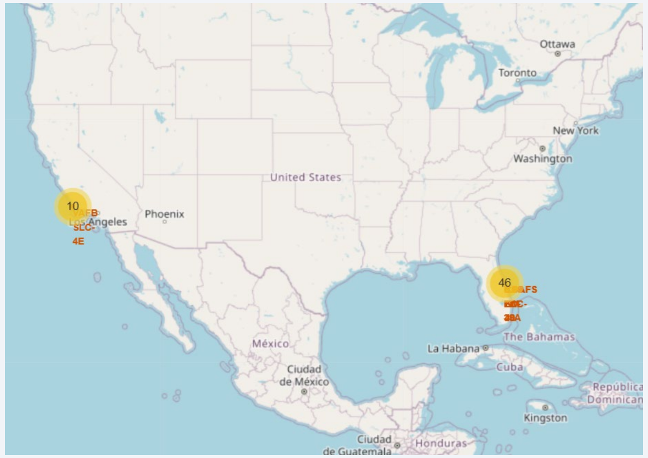


GitHub Link: [EDA with SQL.ipynb]()

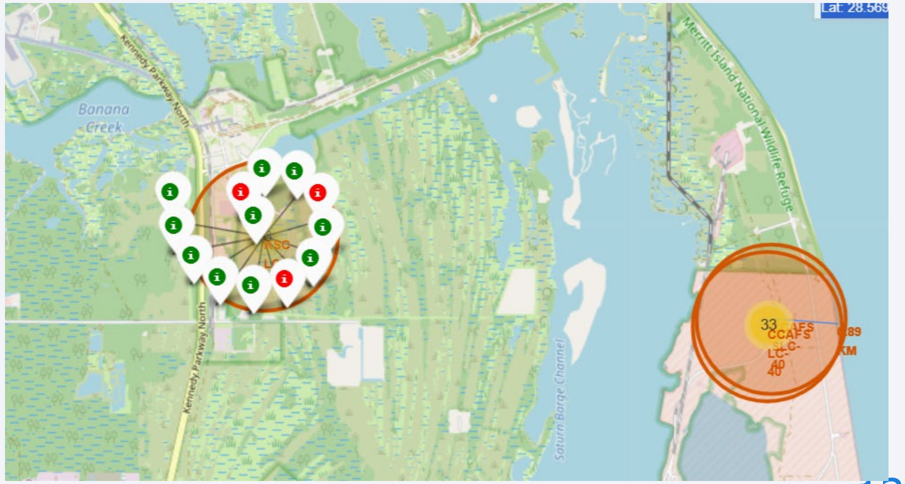
# Build an Interactive Map with Folium

Map objects were used:

- Marker to mark the location of Launch Sites, Nasa Space Center and calculated distances on the map
- Circle for circle above the markers to show number of launches from the site
- Icon for icon of successful (green) or unsuccessful(red) launches
- Cluster to group icons and simplify the map
- Polyline between selected points and Launch Site to show direction

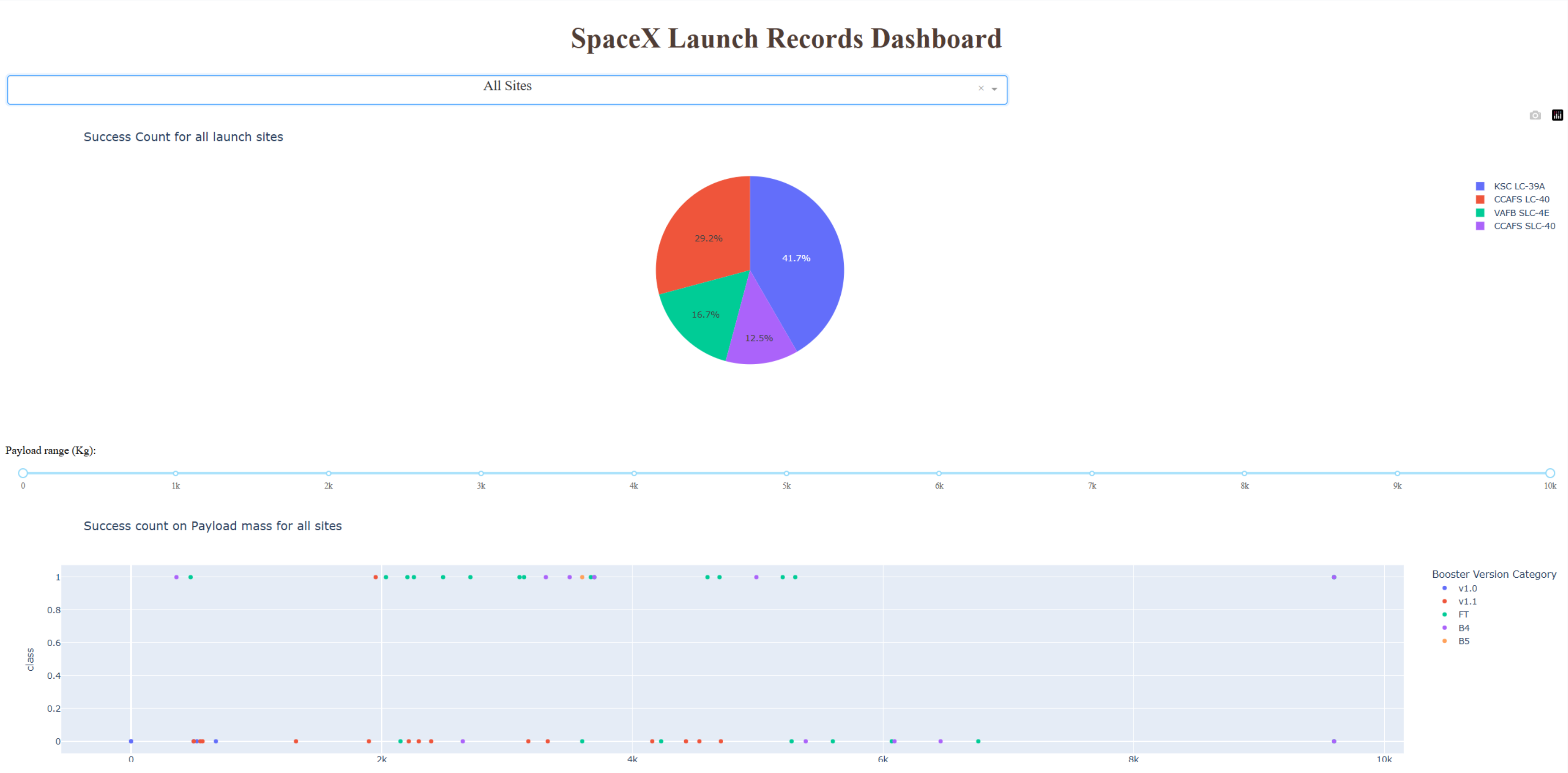







GitHub Link: [Launch_site_location.ipynb]

# Build a Dashboard with Plotly Dash

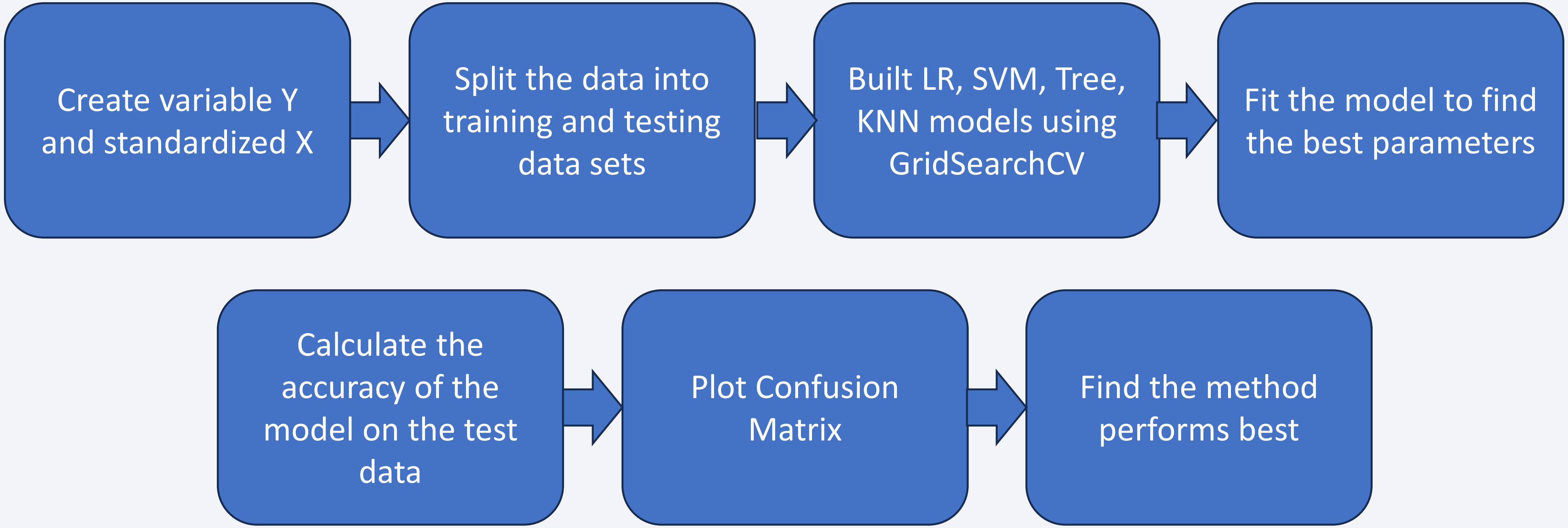



Dropdown option of pie chart created to show the success launches of all/ each Site

Scatter Plot with payload range slider to show the success launches of all / each site by payload mass

GitHub Link: [Dash app.ipynb]

# Predictive Analysis (Classification)




GitHub Link: [Machine Learning Prediction.ipynb]

# Results

- Exploratory data analysis results
  - We have successfully investigated the relationships between launch parameters and its success rate
- Interactive analytics demo in screenshots
  - The locations of Launch Sites and the distances from proximities were put on an interactive map and analyzed
  - Using Dash total success rate and ratio for each Launch Site were examined. The quantity and success of different Payload Masses were investigated
- Predictive analysis results
  - LR, SVM, Tree and KNN models were built with best hyperparameters

Section 2

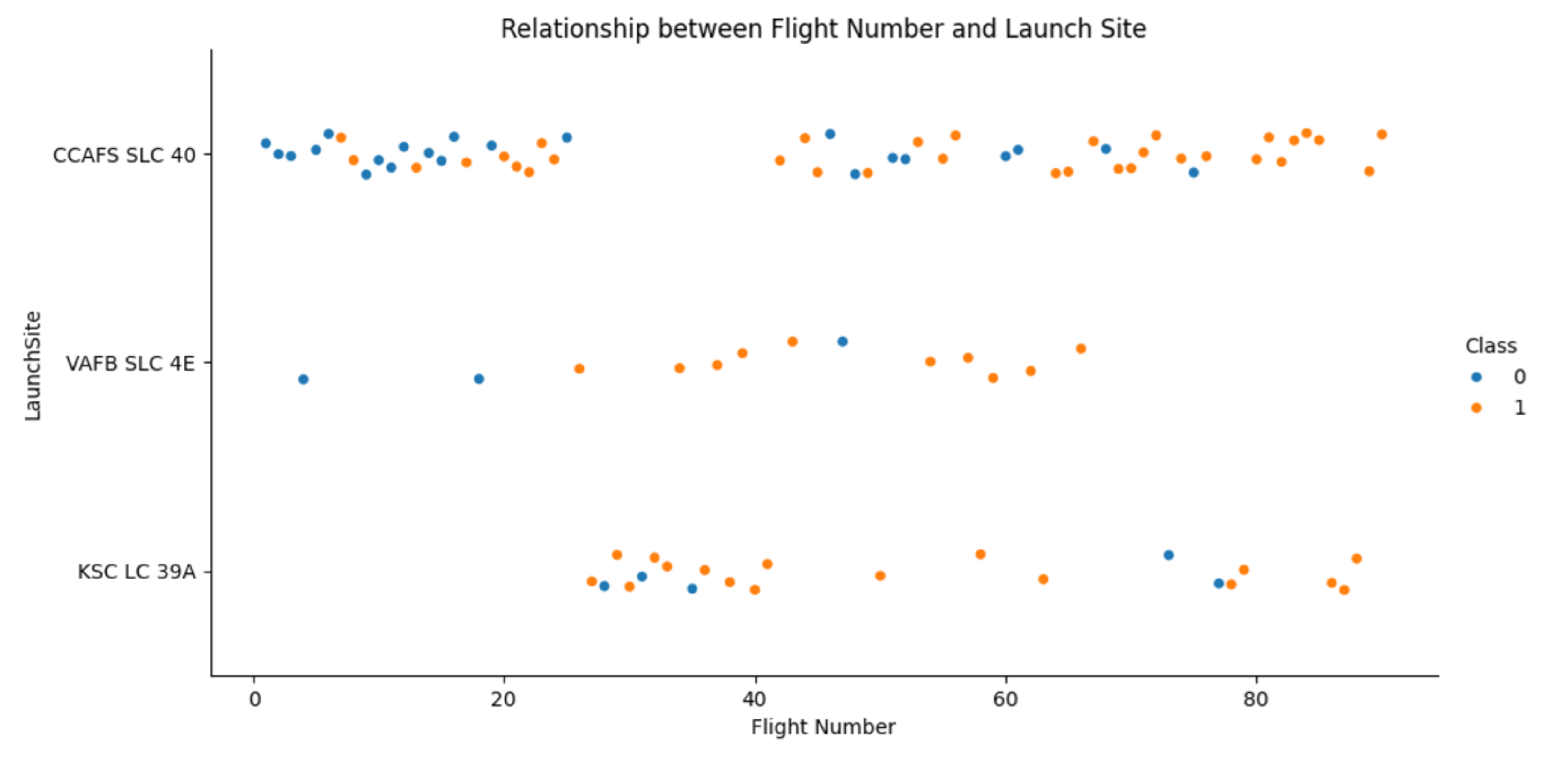
# Insights drawn from EDA

# Flight Number vs. Launch Site

Total Number of launches from CCAFS SLC-40 is significantly higher than from other launches.

First Launches were made from CCAFS SLC-40 and VAFB SLC 4E

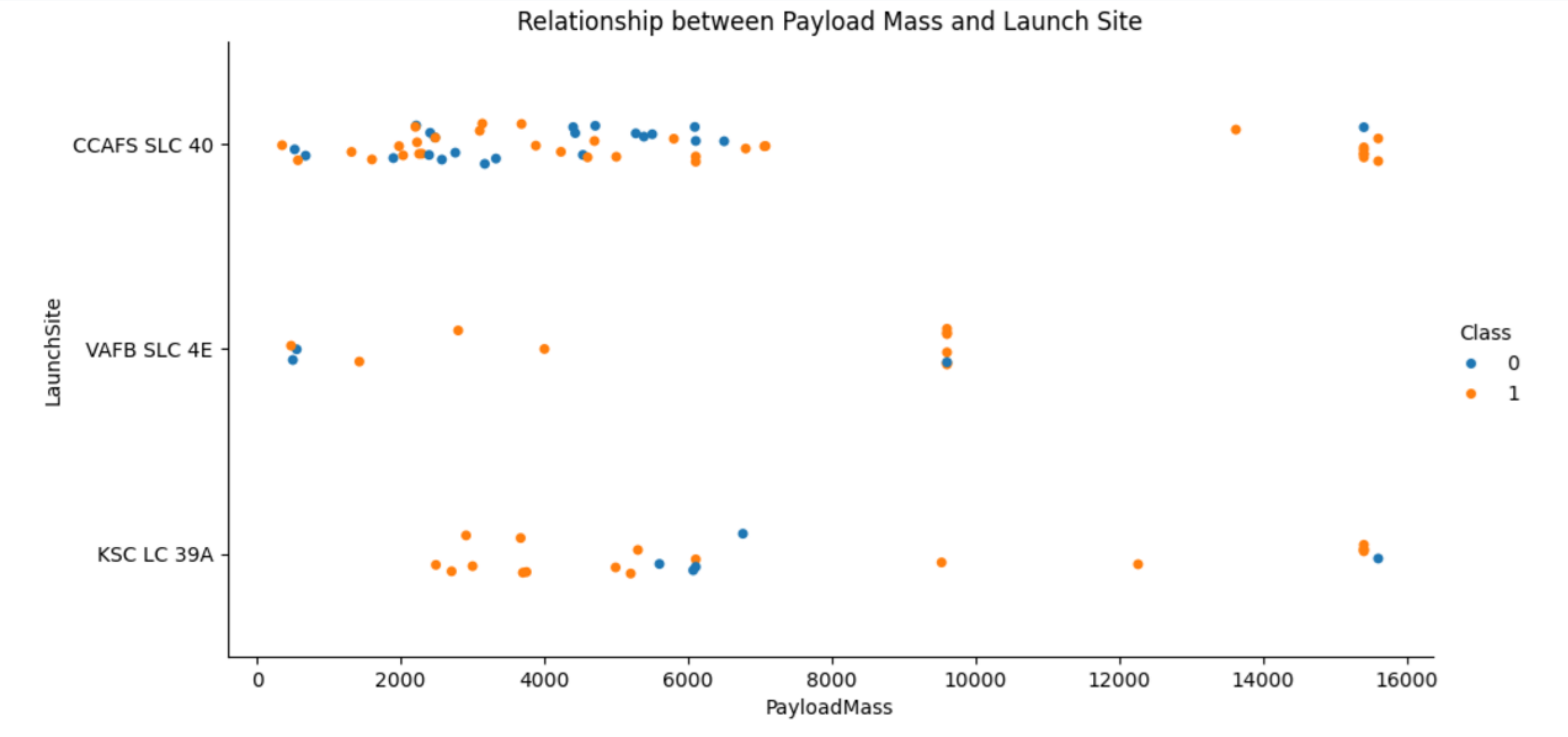
The Success Rate is increasing with increase of flight numbers.



# Payload vs. Launch Site

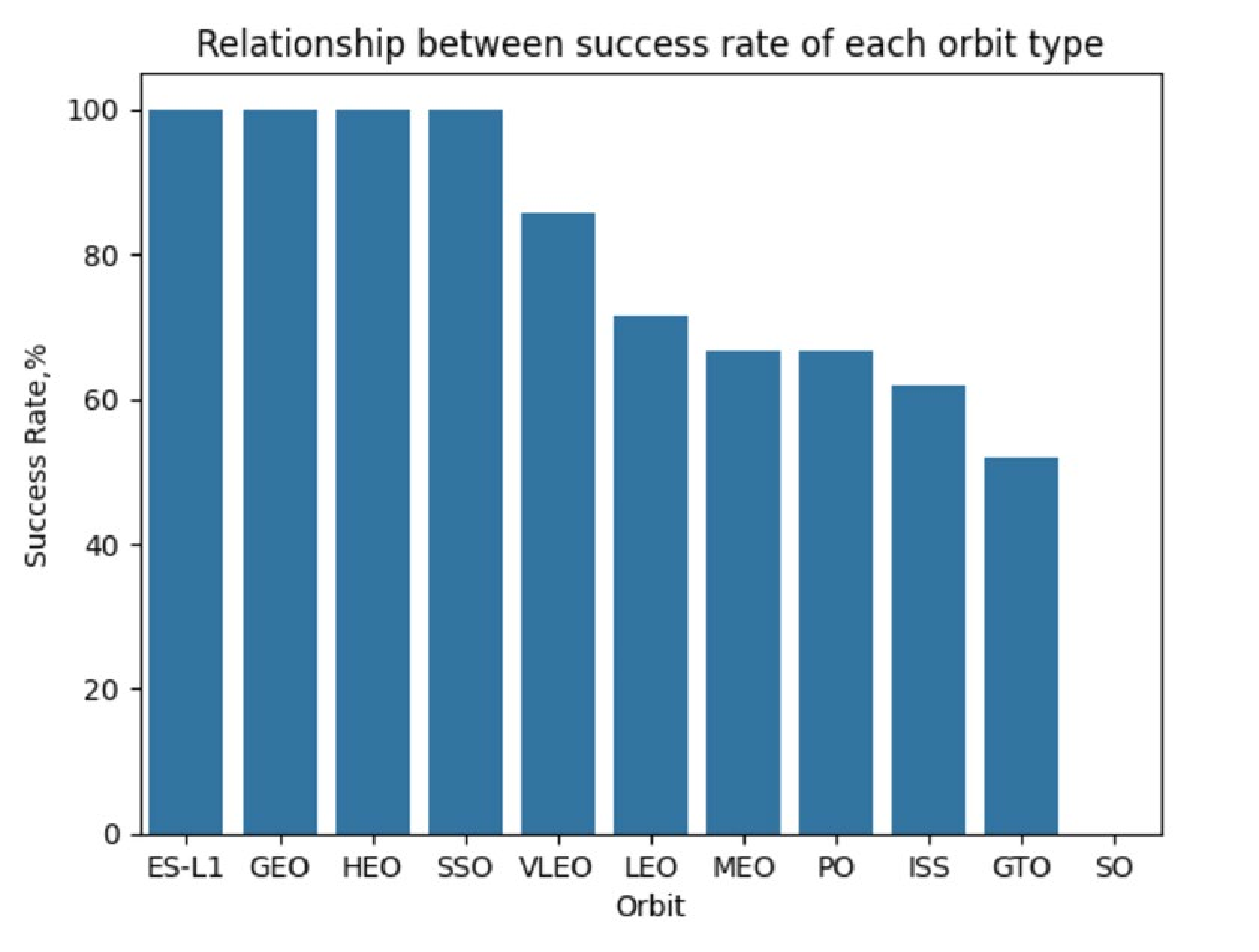
There are no rockets launched for heavy payload mass (greater than 10000) from the VAFB-SLC launch site.

Payloads with lower mass have more launches compared to those with higher mass across all three launches.



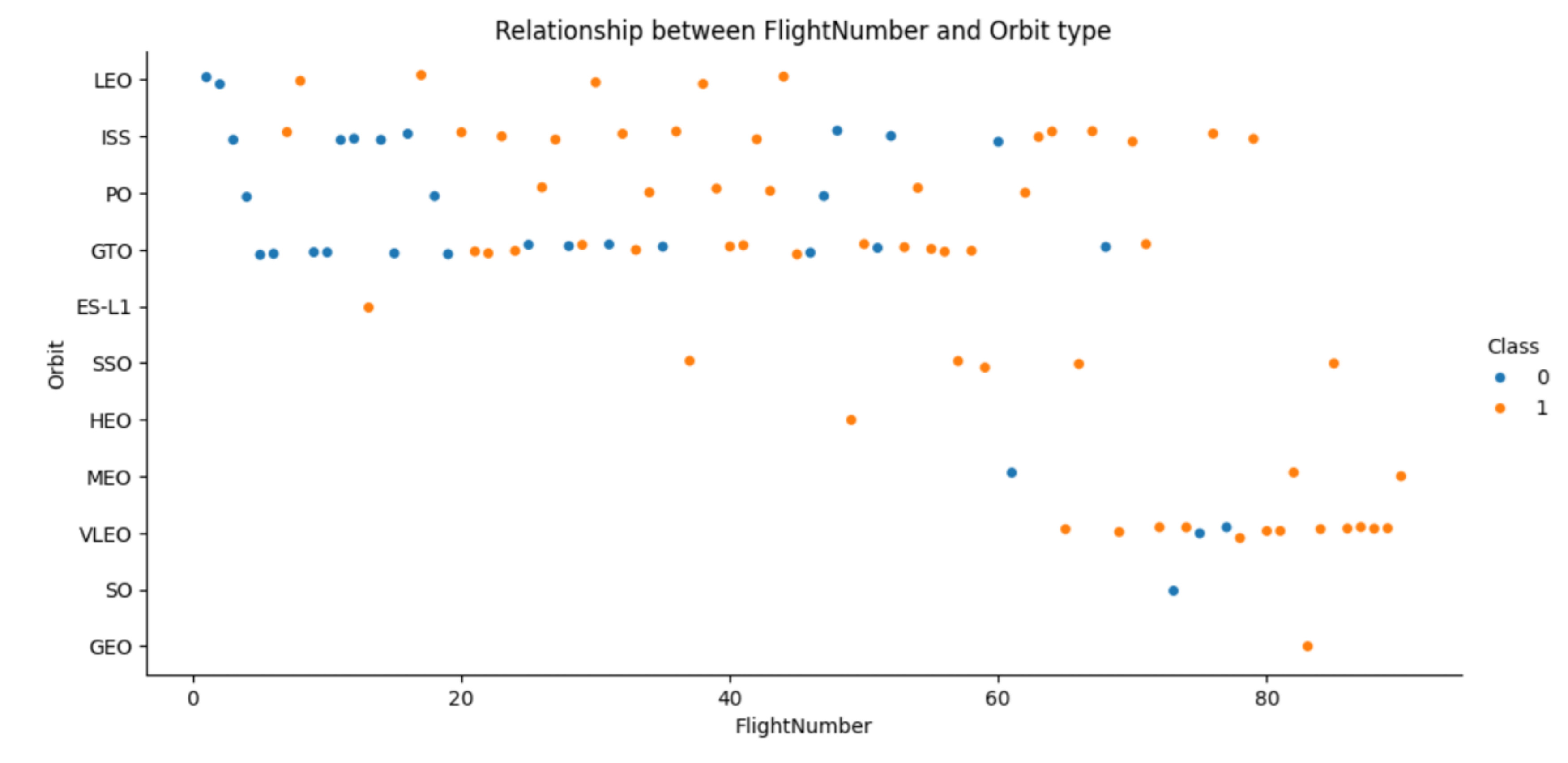
# Success Rate vs. Orbit Type

The rare used orbits ES-L1(1 launch), GEO (1 launch), HEO (1 launch) and SSO (5 launches) have the highest success rate 100%, meanwhile the most frequently used GTO (27 launches) and ISS (21 launches) have the lowest rates. 1 launch attempt on SO orbit wasn’t successful.



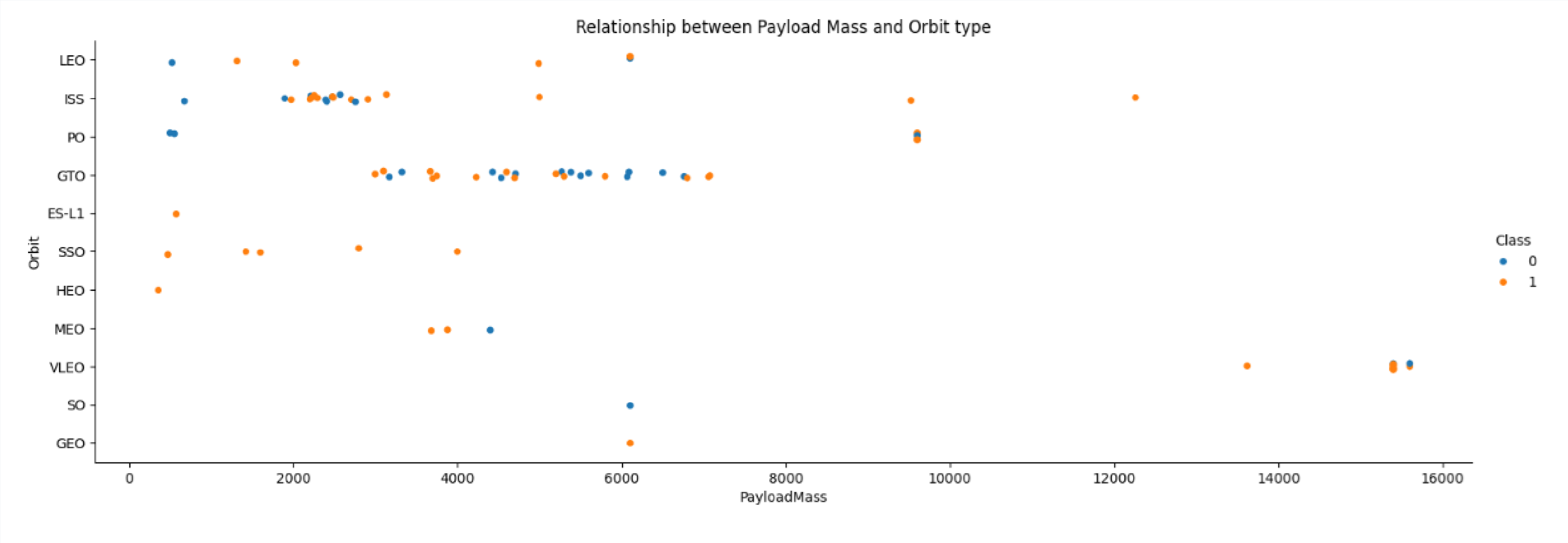
# Flight Number vs. Orbit Type

In the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.



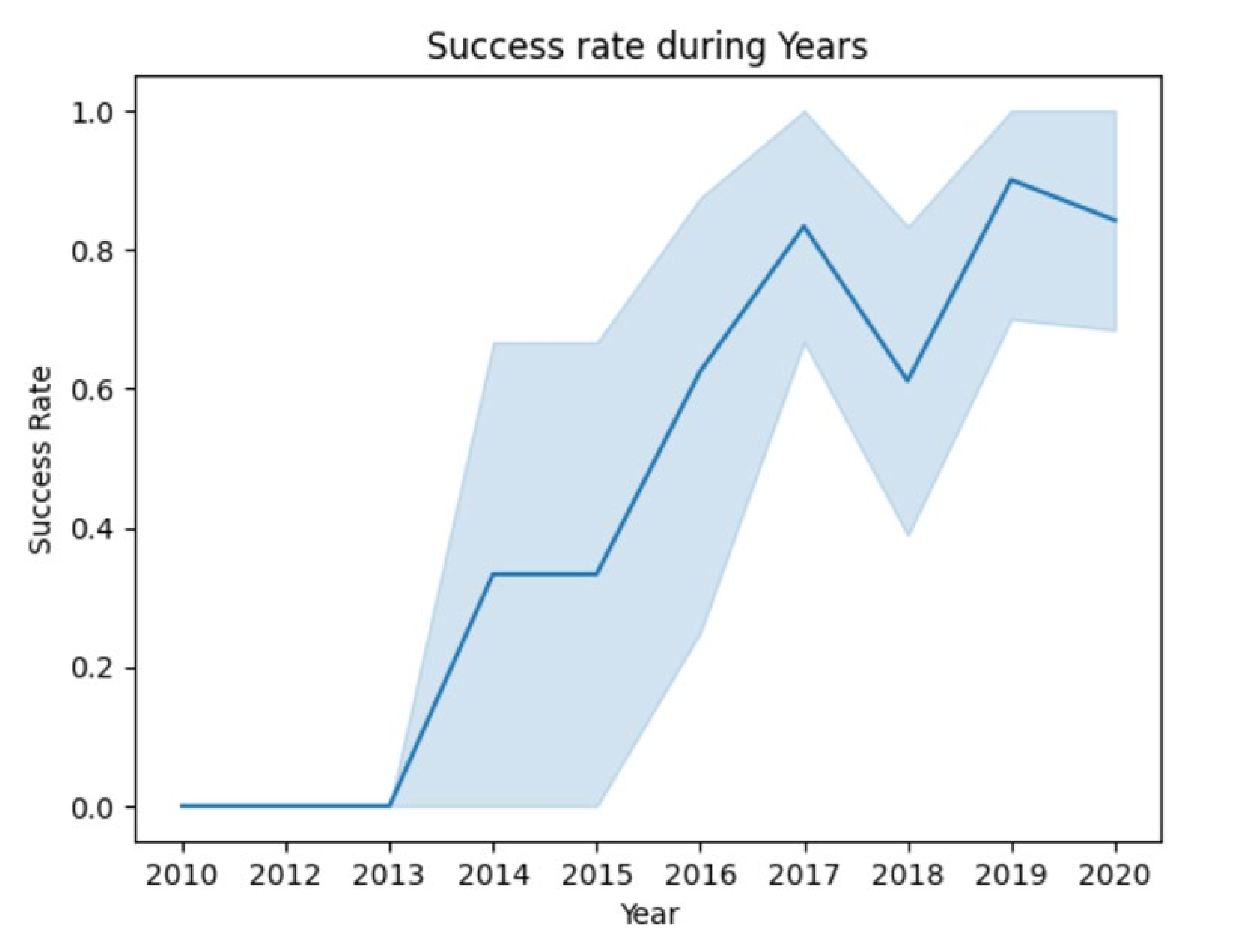
# Payload vs. Orbit Type

- Launches with heavy payloads were made only for VLEO and ISS orbit.
- However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present



# Launch Success Yearly Trend

The Success rate since 2013 kept increasing till 2020, possibly due to technology advancement and experience.



# All Launch Site Names

Performed  an SQL query to obtain all launches sites names.

```
%sql SELECT "Launch_Site"       FROM SPACEXTABLE GROUP BY "Launch_Site";
```

```
 * sqlite:///my_data1.db
Done.
```

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

# Launch Site Names Begin with 'CCA'

Performed an SQL query to obtain 5 records where launch sites begin with `CCA`

```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

| 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 |

# Total Payload Mass carried for NASA

Performed an SQL query to calculate the total payload carried by boosters from NASA

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS "TOTAL PAYLOAD MASS FOR NASA(CRS), KG" FROM SPACEXTABLE WHERE "Customer"= "NASA (CRS)";
```

```
%sql SELECT SUM("PAYLOAD_MASS__KG_") AS "TOTAL PAYLOAD MASS FOR NASA(CRS), KG" FROM SPACEXTABLE WHERE "Customer"= "NASA (CR
```

```
 * sqlite:///my_data1.db
Done.
```

| TOTAL PAYLOAD MASS FOR NASA(CRS), KG |
| --- |
| 45596 |

# Average Payload Mass by F9 v1.1

Performed an SQL query to calculate the average payload mass carried by booster version F9 v1.1

```
%sql SELECT AVG("PAYLOAD_MASS__KG_") AS "AVG PAYLOAD MASS carried by booster version F9 v1.1, KG" FROM SPACEXTABLE WHERE "Booster_Version"= "F9 v1.1";
```

```
SELECT AVG("PAYLOAD_MASS__KG_") AS "AVG PAYLOAD MASS carried by booster version F9 v1.1, KG" FROM SPACEXTABLE WHERE "Booster_
```

```
 * sqlite:///my_data1.db
Done.
```

| AVG PAYLOAD MASS carried by booster version F9 v1.1, KG |
| --- |
| 2928.4 |

# First Successful Ground Landing Date

Performed  an SQL query to find the date of the first successful landing outcome on ground pad

```
%sql SELECT MIN("Date") AS "First succesful landing" FROM SPACEXTABLE WHERE "Mission_Outcome"="Success";
```

```
 * sqlite:///my_data1.db
Done.
```

| First succesful landing |
| --- |
| 2010-06-04 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

Performed an SQL query to obtain a list the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)" AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000;
```

```
%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Success (drone ship)" AND "PAYLOAD_MASS__KG_" BETW
```

```
 * sqlite:///my_data1.db
Done.
```

| Booster_Version |
|---|
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

# Total Number of Successful and Failure Mission Outcomes

Performed an SQL query to calculate the total number of successful and failure mission outcomes

```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total Number" FROM SPACEXTABLE GROUP BY "Mission_Outcome";
```

```
 * sqlite:///my_data1.db
Done.
```

| Mission_Outcome | Total Number |
|---:|---:|
| Failure (in flight) | 1 |
| Success | 98 |
| Success | 1 |
| Success (payload status unclear) | 1 |

# Boosters Carried Maximum Payload

Performed an SQL query to obtain a list the names of the booster which have carried the maximum payload mass

%sql SELECT "Booster_Version", "PAYLOAD_MASS__KG_" AS "Maximumum Payload mass,kg" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_"=(SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);

```
%sql SELECT "Booster_Version", "PAYLOAD_MASS__KG_" AS "Maximumum Payload mass,kg" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_'
```

```
 * sqlite:///my_data1.db
Done.
```

| Booster_Version | Maximumum 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 |

# 2015 Launch Records

Performed  an SQL query to obtain a list the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql SELECT substr(Date, 6,2) AS Month,"Booster_Version","Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Failure (drone ship)" AND substr(Date,0,5)='2015'
```

```
In [48]: %sql SELECT substr(Date, 6,2) AS Month,"Booster_Version","Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = "Failure (
```

```
 * sqlite:///my_data1.db
Done.
```

Out[48]:

| Month | Booster_Version | Launch_Site |
|---|---|---|
| 01 | F9 v1.1 B1012 | CCAFS LC-40 |
| 04 | F9 v1.1 B1015 | CCAFS LC-40 |

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

Performed an SQL query to 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 "Count" FROM SPACEXTABLE WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20" GROUP BY "Landing_Outcome"\ **ORDER BY** "Count" **DESC**;

```
%sql SELECT "Landing_Outcome", COUNT(*) AS "Count" FROM SPACEXTABLE  WHERE "Date" BETWEEN "2010-06-04" AND "2017-03-20" GROU
    ORDER BY "Count" DESC;
```

```
 * sqlite:///my_data1.db
Done.
```

| 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 |

Section 3

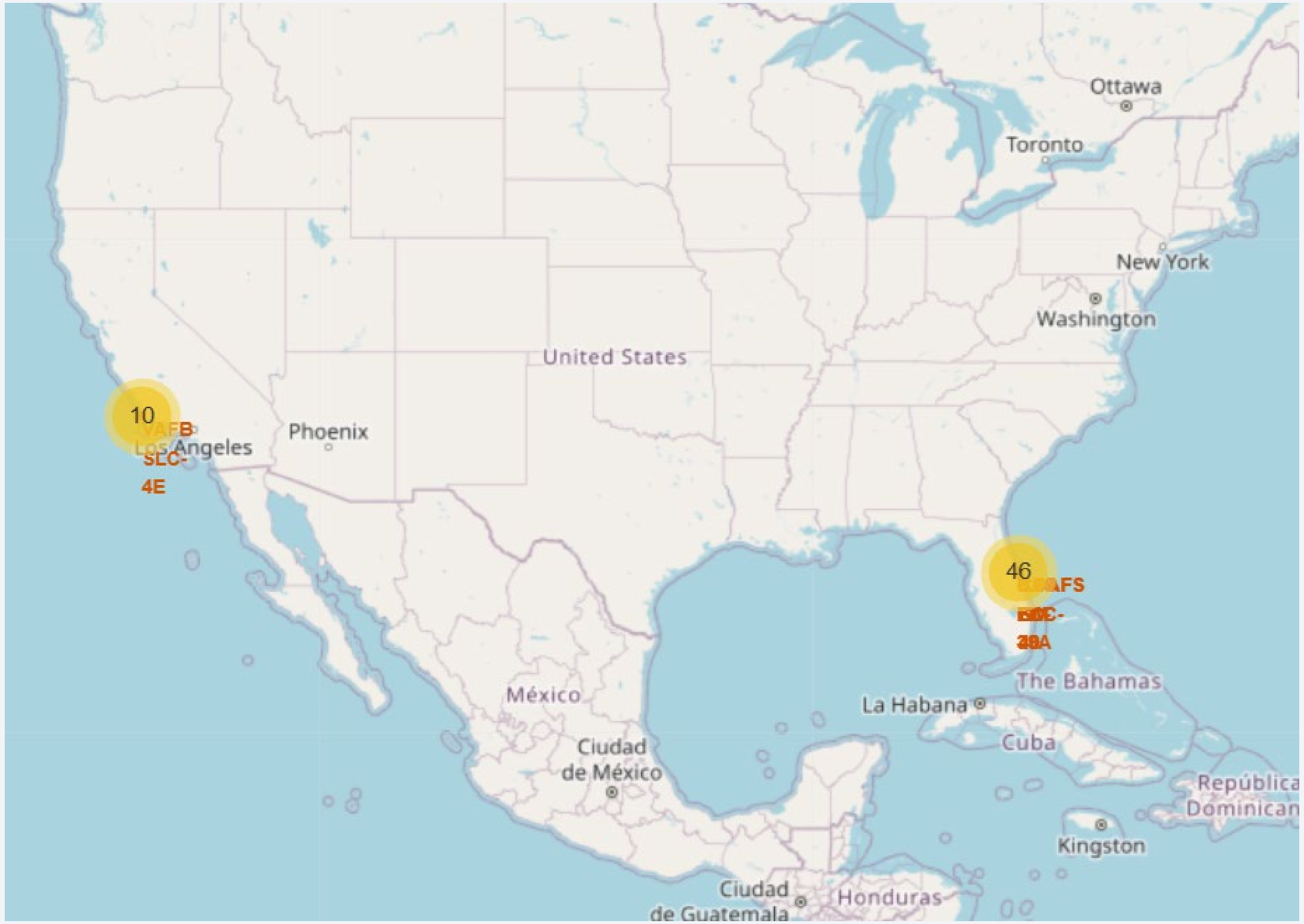
# Launch Sites Proximities Analysis

# All Launch Sites on a Map with number of Launches

The locations of all Launch Sites were put on a map with proper names and quantity of launches from each.

As you can see all of them are in proximity of Equator because it can take optimum advantage of the Earth's substantial rotational speed and help to economy the fuel.

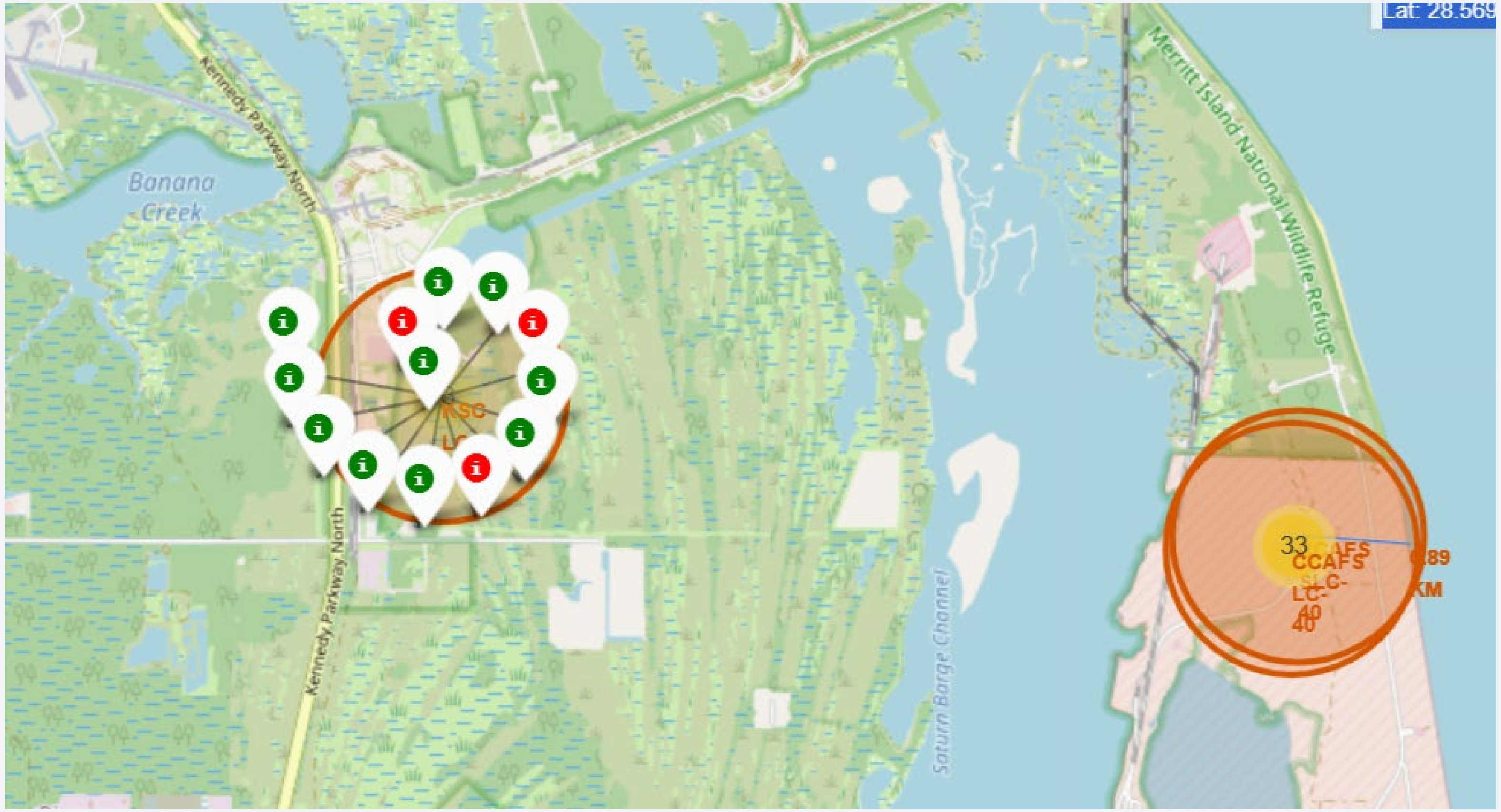
Similarly all launch sites are in very close proximity to the coast for safety reasons, to make the parts of rocket's falling to the ocean in case of unsuccessful launch



# Success / Failure Launches marked on the map

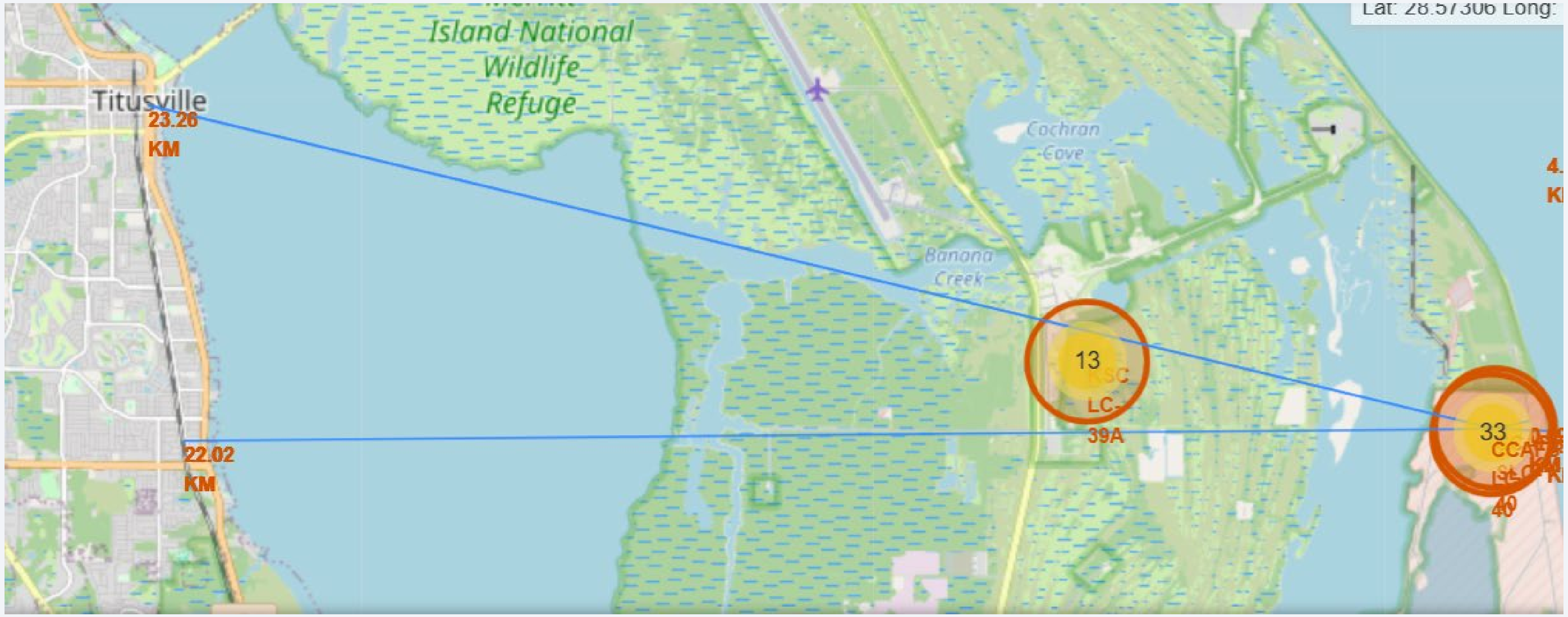
Launch outcome were put on a map as colorful labels (green for success/ red for failure) and grouped in clusters for each launch site.

Was discovered that the best success rate was accomplished on Launch Site KSC LC-39.



# Distance between Launch site and its proximities

The distances from CCAFS SLC-40 to nearest proximities (nearest coast line, railway, town, highway) were calculated, marked on the map and directions were drawn by PolyLines.

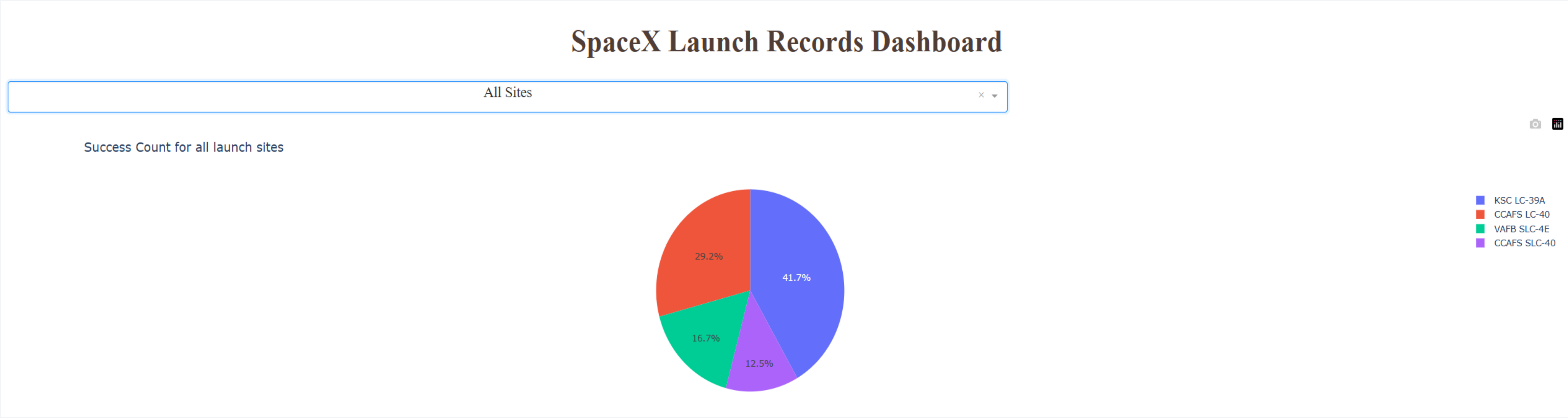



Was discovered that Launch Site is located near coast line (0.89 km) and highway (0.59 km) but distantly from towns (23.26 km) and railways (22.02 km) because of safety reasons.

Section 4

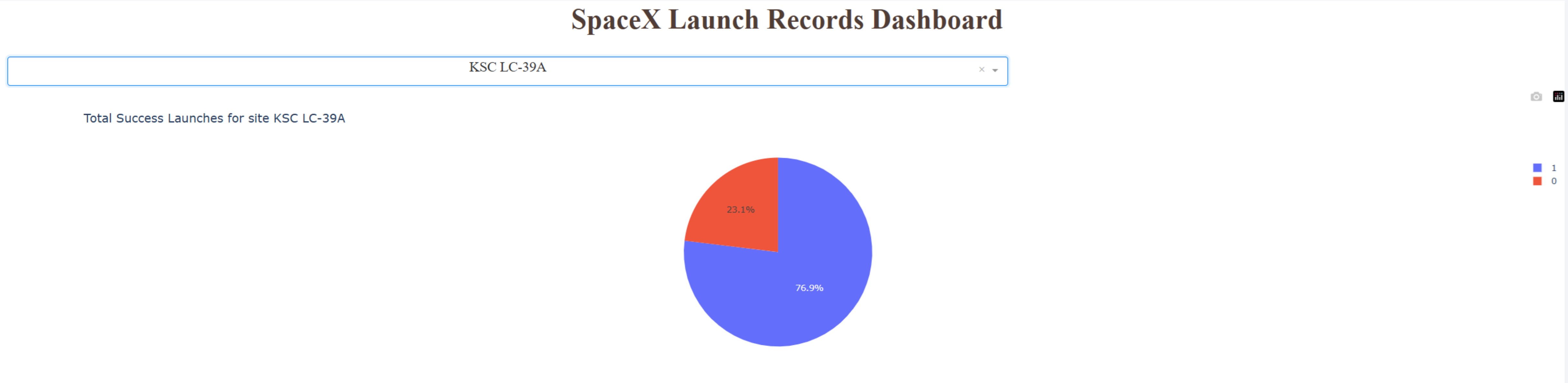
# Build a Dashboard with Plotly Dash

# Total success launches of all Launch Sites




Was discovered that KSC LC-39A had the biggest number of successful launches with 41.7% from the entire record whereas CCAFS SLC-40 had the lowest 12.5%

# Launch Site with highest launch success ratio

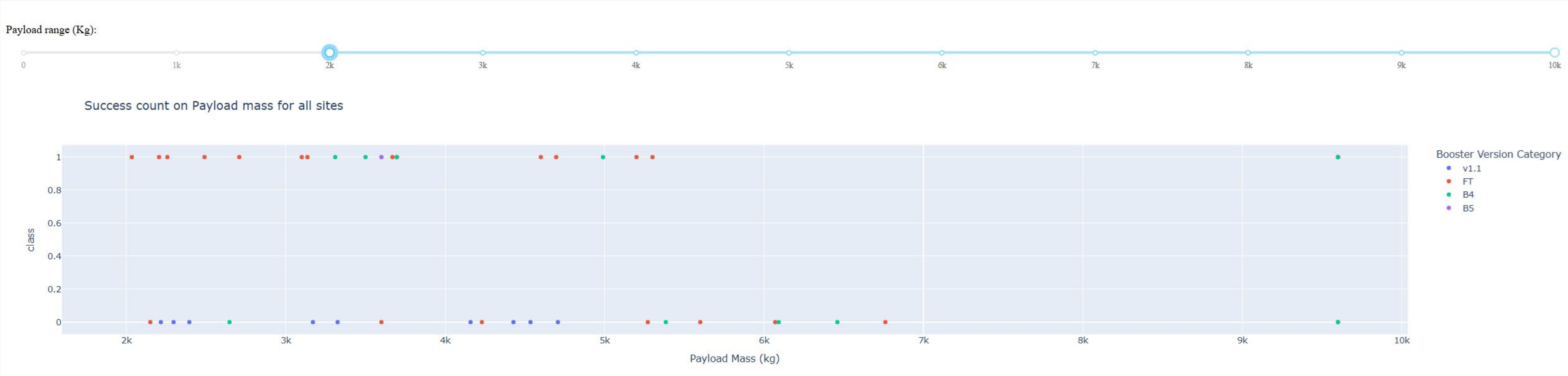


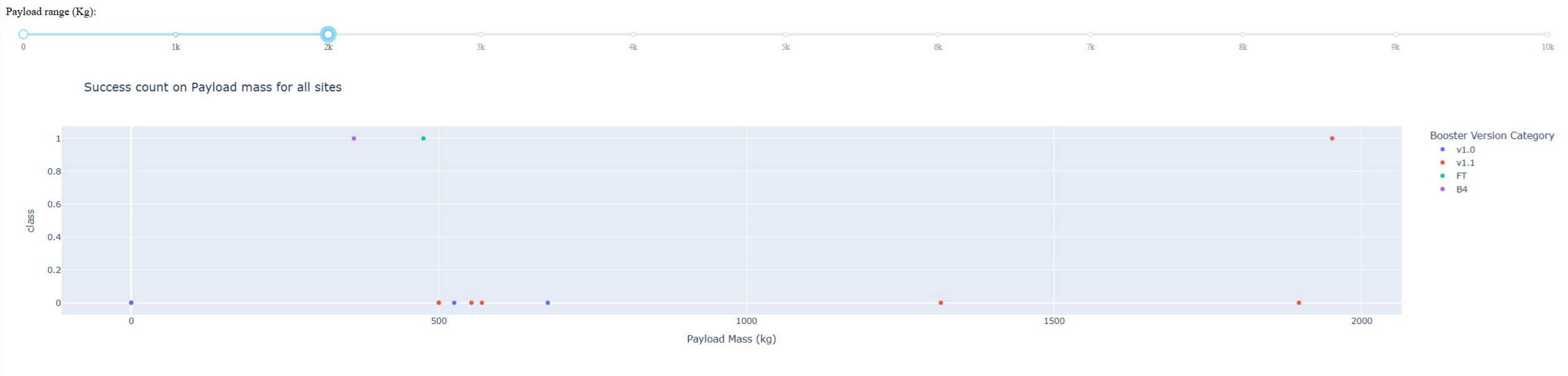

Launch Site KSC LC-39A achieved 76.9% success rate

# Relation between Payload Mass and Success

Launches with Payload mass between 2000 kg and 6000kg had the highest success rate than others.

Booster Version FT was the most successful.



Section 5

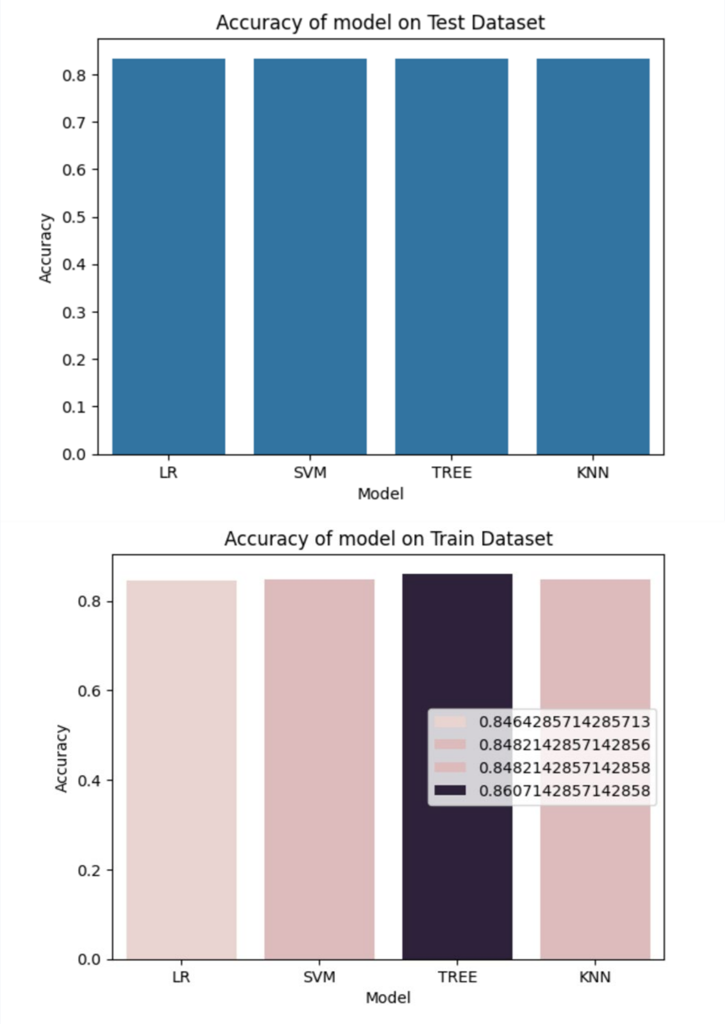
# Predictive Analysis (Classification)

# Classification Accuracy

Different classification models, such as Logical Regression, Support Vector Machine, Decision Tree and k-nearest neighbors were built.

All of them gave us the same result as accuracy rate 0.833 on test set.

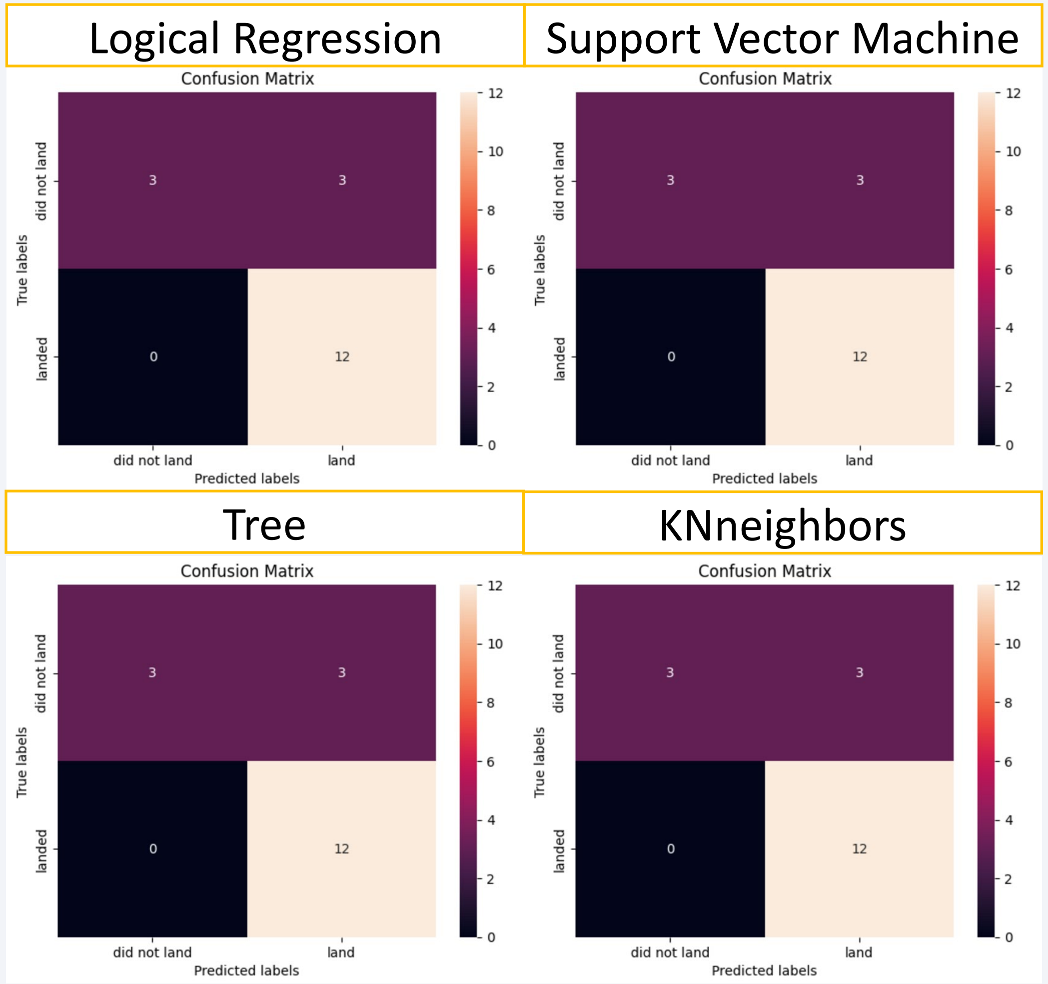
But on the Train Dataset Decision Tree gave better result as accuracy 0.86.



# Confusion Matrix

We can make conclusion that based on the accuracies of the Test Set, we can not confirm which method performs best.

Same Test Set scores may be due to the small test sample size (18 samples).



# Conclusions

- First Launches were made from CCAFS SLC-40 and VAFB SLC 4E
- Total Number of launches from CCAFS SLC-40 is significantly higher than from other launches, but only 57.1% of them were successful and only 12.5% of all successful launches were made from this Launch Site.
- Launch Site KSC LC-39A achieved the highest 76.9% success rate and 41.7% of all successful launches were made there
- The Success rate since 2013 kept increasing till 2020, possibly due to technology advancement and experience
- The average Payload Mass was 2928.4 kg and total Payload Mass carried for NASA was 45 596 kg.
- Launches to rarely used orbits ES-1, GEO, HEO, SSO had a 100% success rate, but to the most frequently used ones GTO and ISS only 50-65%.
- Launches with Payload mass between 2000 kg and 6000kg had the highest success rate than others.
- The Decision Tree Model is the best for prediction of success outcome.

# Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

# Thank you!