Winning Space Race with Data Science

Name: Vanshika Raj
Date: 01-06-2024

Outline

- Executive Summary

- Introduction

- Methodology

- Results

- Conclusion

- Appendix

Executive Summary

Summary of Methodologies

Data Wrangling: Identification and imputation of missing values, particularly in the 'PayloadMass' column using mean calculation, followed by data export to CSV for easy access in later stages.

EDA and Visualization: Exploratory data analysis using SQL, and interactive visual analytics with Folium and Plotly Dash.

Predictive Analysis: Building and evaluating classification models such as Logistic Regression, SVM, Decision Tree, and KNN using GridSearchCV for parameter tuning.

Model Comparison: Assessing model performance with accuracy metrics and confusion matrices to determine the best performing model.

- ## Summary of all results

The project involves a comprehensive data science workflow from preprocessing to model evaluation. The first successful landing outcome on a ground pad occurred on 12/22/2015. The average payload mass for the least successful Booster Version, F9 v1.1, is 2928.4kg.

The launch success rate has been increasing since 2013 until 2020, with the KSC LC-39A launch site having the highest success rate at 76.9%. The payload mass for most successful launches is between 2000 – 6000kg.

With heavy payloads, the successful landing or positive landing rate is higher for Polar, LEO, and ISS Orbits. Both CCAFS SLC 40 and KSC LC 39A Payload Masses are mostly under 8000kg, with CCAFS SLC 40 having significantly more successful launches. VAFB SLC 4E has the least number of launches.

Introduction

- **Background**: SpaceX, a private American aerospace manufacturer and space transportation company, has revolutionized space travel with its Falcon 9 rockets. A significant factor contributing to their success and cost-effectiveness is the reusability of their rockets, specifically the Falcon 9's first stage. This reusability drastically reduces the cost of space travel, making SpaceX a highly competitive player in the market.

Problem:
The cost of a Falcon 9 launch is advertised on SpaceX's website as 62 million dollars, significantly lower than other providers who charge upward of 165 million dollars. A key factor in this cost difference is the reusability of the Falcon 9's first stage. If we can predict whether the first stage will land successfully, we can estimate the cost of a launch. This information could be invaluable for alternate companies looking to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

- ## Data collection methodology:

  - The project will involve data collection, preprocessing, exploratory data analysis, feature selection, model building, and validation. Various machine learning algorithms will be explored to find the most accurate model.

- ## Perform data wrangling

  - Identification of Missing Values: The initial step in the data processing was to identify any missing values within the dataset. Missing values can lead to inaccurate analysis and predictions, hence it's crucial to handle them appropriately.

  - Mean Calculation: The mean (average) value of the 'PayloadMass' column was calculated. The mean is a central tendency measure and provides a reasonable estimate that can be used to fill in missing values.

  - Imputation of Missing Values: All the missing values in the 'PayloadMass' column were replaced with the mean value. This technique, known as mean imputation, is a common strategy for handling missing data. It helps to maintain the overall distribution and variance of the dataset.

  - Data Export: After handling the missing values (except for in the 'LandingPad' column), the cleaned dataset was exported to a CSV file. This file serves as a checkpoint, allowing for the data to be easily accessed in subsequent stages of the project without needing to repeat the preprocessing steps.

  - Data Consistency: To ensure consistency in the analysis, a decision was made to provide data in a pre-selected date range in the next lab. This helps to standardize the results and makes it easier to compare findings across different analyses.

Methodology

- **Perform exploratory data analysis (EDA) using visualization and SQL**

- **Perform interactive visual analytics using Folium and Plotly Dash**

- **Perform predictive analysis using classification models**

    - Data Loading: Two datasets were loaded from CSV files into pandas DataFrames.

    - Data Preparation: The 'Class' column was converted to a NumPy array and assigned to 'Y'. The data in 'X' was standardized.

    - Data Splitting: The data was split into training and test sets, with 20% of the data reserved for testing.

    - Model Building: Four different models were built - Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K-Nearest Neighbors. For each model, a GridSearchCV object was fitted to find the best parameters from a predefined dictionary of parameters.

    - Model Evaluation: The accuracy of each model was calculated on the test data. A confusion matrix was also plotted for each model to visualize its performance.

    - Best Performing Model: The performance of all the models was compared.
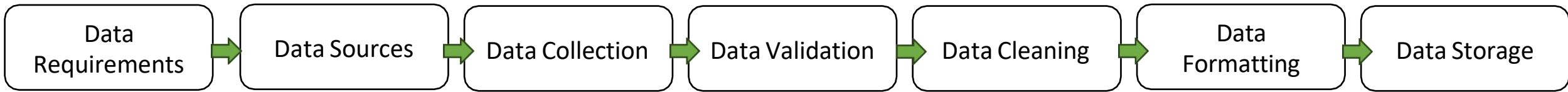
Data Collection

Start by importing necessary libraries and setting up pandas options. Define helper functions to extract information from the SpaceX API using identification numbers in the launch data. These functions help learning various details about the rocket, launch site, payload, and cores.

Make a request to the SpaceX API to get data about past launches, decode the response as a JSON, and convert it into a pandas DataFrame. Prepare the data by taking a subset of the DataFrame, performing some cleaning operations, and using the helper functions to extract more data from the API.

Create a new DataFrame from the dictionary built with the extracted data. Display the summary of the DataFrame to understand its structure and contents. Filter the DataFrame to only include Falcon 9 launches and save the filtered data to a new DataFrame. After filtering the data, reset the Flight Number column to reflect the new row indices.

Observe that some rows in the DataFrame have missing values. Calculate the mean for the Payload Mass and replace the missing values in the data with this mean value. Check the number of missing values in the Payload Mass column again to ensure they have been replaced.

Data Requirements	→	Data Sources	→	Data Collection	→	Data Validation	→	Data Cleaning	→	Data Formatting	→	Data Storage

1. Identifying Data Requirements: The first step in the data collection process involves determining the necessary data. This step requires understanding the problem to be solved or the question to be answered.

2. Determining Data Sources: Once the required data is known, the next step is to identify where this data can be obtained. Possible sources could include internal databases, public data sets, APIs, web scraping, surveys, and so on.

3. Data Collection: This is the actual process of gathering the data from the identified sources. The method of collection will depend on the nature of the source.

4. Data Validation: After the data is collected, it's important to validate it to ensure its accuracy and relevance. This could involve checking for missing values, outliers, or incorrect entries.

5. Data Cleaning: This step involves cleaning the data by handling missing values, removing duplicates, and correcting errors.

6. Data Formatting: The collected data may need to be converted or formatted into a suitable format for analysis. This could involve converting data types, restructuring data, or encoding categorical variables.

7. Data Storage: Finally, the collected data is stored for future use. This could be in a database, a CSV file, a data warehouse, and so on.
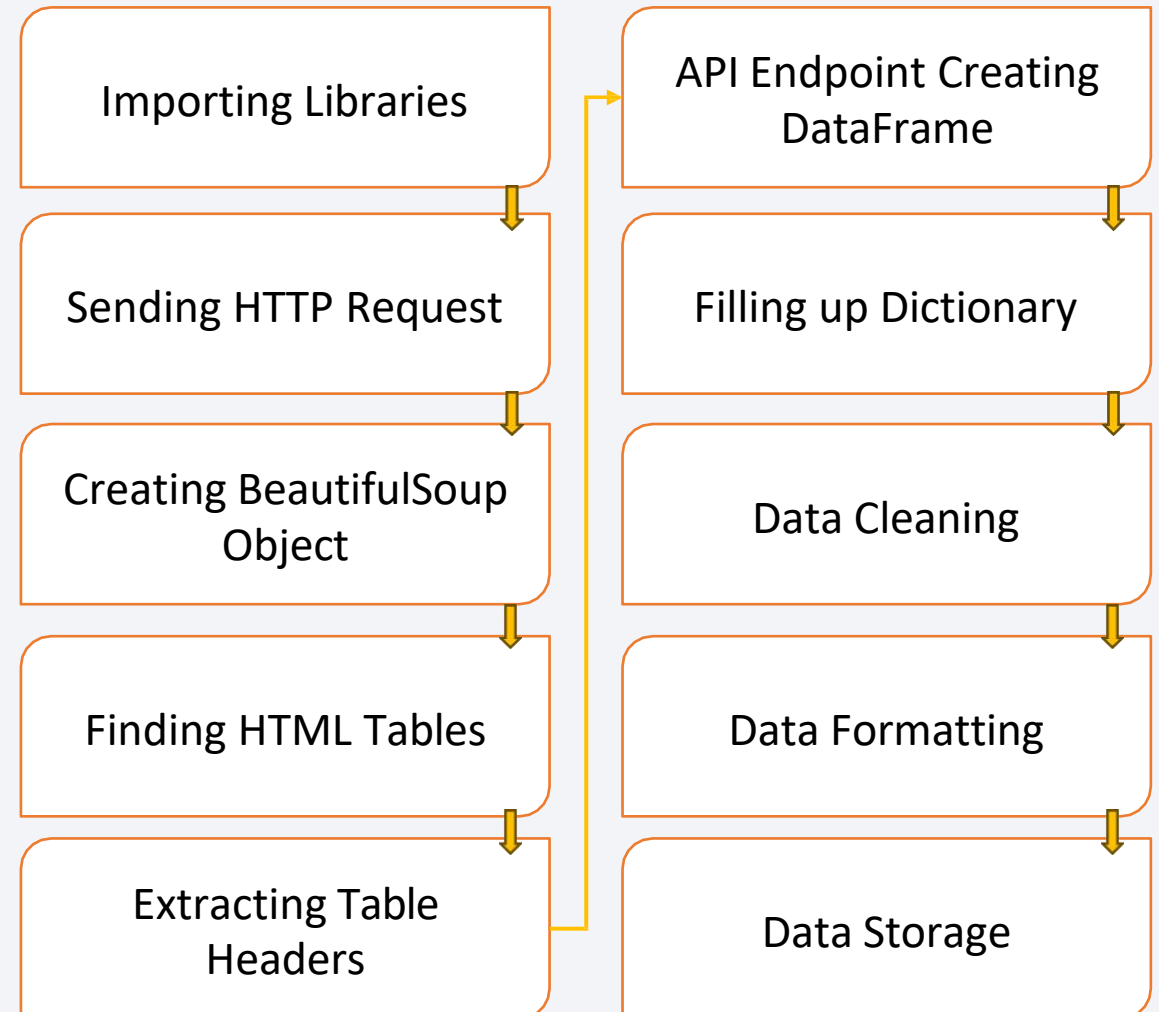
Data Collection – SpaceX API

- API Endpoint Identification: Determine the specific SpaceX API endpoints that will provide the required data.

- REST Call Execution: Use HTTP methods such as GET to make calls to the SpaceX API endpoints.

- Data Retrieval: Extract the data returned by the API in response to the REST calls.

- Data Processing: Process the retrieved data as needed, which may include parsing JSON responses and transforming data into a desired format.

- Data Analysis: Analyze the processed data to extract insights or make decisions based on the data collected from the SpaceX API.

API Endpoint Identification

↓

REST Call Execution:

↓

Data Retrieval

↓

Data Processing

↓

Data Analysis

10
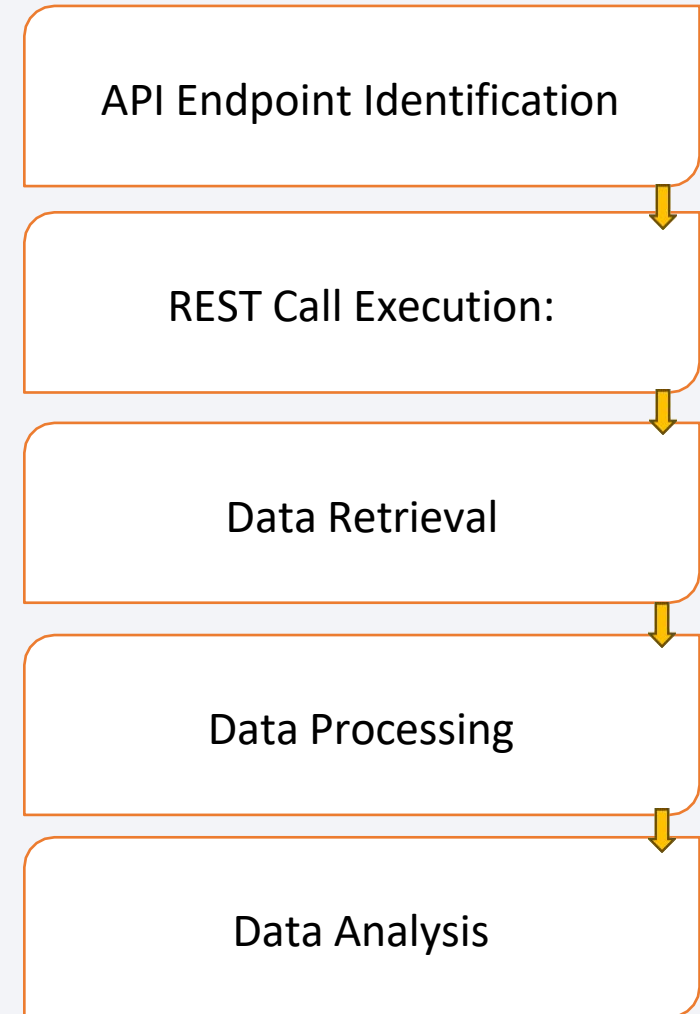
Data Collection - Scraping

- **Importing Libraries**: The process begins by importing the necessary libraries such as requests, BeautifulSoup, re, unicodedata, and pandas.

- **Sending HTTP Request:** An HTTP GET request is sent to the target URL (in this case, a Wikipedia page) using the requests.get() method. The server responds to the request by returning HTML content of the webpage.

- **Creating BeautifulSoup Object**: The HTML content returned by the server is then parsed and converted into a BeautifulSoup object. BeautifulSoup is a Python library that is used for web scraping purposes to extract the data from HTML and XML documents.

- **Finding HTML Tables**: All tables on the webpage are found using the find_all function in BeautifulSoup with the element type table.

- **Extracting Table Headers**: The column names are extracted from the HTML table header elements <th>.

- **Creating DataFrame**: An empty dictionary is created with keys from the extracted column names. This dictionary is later converted into a pandas DataFrame.

- **Filling up Dictionary with Launch Records**: The dictionary is filled up with launch records extracted from table rows. An incomplete code snippet is provided to fill up the dictionary, which needs to be completed with TODOs.

- **Data Cleaning**: The data is cleaned by handling missing values, removing duplicates, correcting errors, etc.

- **Data Formatting**: The collected data may need to be converted or formatted into a suitable format for analysis. This could involve converting data types, restructuring data, or encoding categorical variables.

- **Data Storage:** Finally, the collected data is stored for future use. This could be in a database, a CSV file, a data warehouse, etc.

Importing Libraries → Sending HTTP Request → Creating BeautifulSoup Object → Finding HTML Tables → Extracting Table Headers → API Endpoint Creating DataFrame → Filling up Dictionary → Data Cleaning → Data Formatting → Data Storage

Data Wrangling

- **Data Acquisition**: The SpaceX dataset was loaded from a CSV file hosted online.

- **Data Inspection**:

The first 10 rows of the dataset were displayed to understand the data structure and content.

Missing values were identified and calculated as a percentage of the total dataset for each attribute.

Data types for each column were identified as numerical or categorical.

- **Data Transformation:**

The number of launches per launch site was determined using the value_counts() method on the LaunchSite column.

The number and occurrence of each orbit type were calculated using value_counts() on the Orbit column.

The number and occurrence of mission outcomes were analyzed using value_counts() on the Outcome column.

- **Feature Engineering:**

A new categorical feature, Class, was created based on the Outcome column.

Class was assigned a value of 0 if the Outcome was a "bad outcome" (i.e., failed landing), and 1 if it was a successful landing.

A set of "bad outcomes" was defined to identify unsuccessful landing scenarios.

- **Data Export:**The modified dataset, including the Class feature, was exported to a new CSV file for future analysis.

Data Acquisition

Data Inspection

Data Transformation

Feature Engineering

Data Export

EDA with Data Visualization

- 4 different Charts, scatterplot, catplot, bar chart and line chart are used in for Data Visualization

1. **Scatterplot:** is used to display the relationship between two numerical variables. Each dot on the scatterplot represents an observation in the dataset and the position of the dot represents its values on the two variables.

2. **Categorical plot**: is used to visualize the distribution of categorical data. It can show the counts of observations in each categorical bin using bars. A catplot can also support higher-dimensional visualizations, allowing to visualize the effect of additional variables on the distribution of categories.

3. **Barchart:** is used to display and compare the quantity, frequency, or other measure (like mean) for different categories or groups. The length of the bars represents the measure.

4. **Line Chart**: is used to represent the change in a numerical variable based on another continuous variable, often time. Each point on the line represents the value of the variable at a certain point of time or a certain condition, and the line segments between the points show the changes in the variable.

EDA with SQL

- **Connect to Database**: Established a connection with the SQLite database and loaded the SpaceX dataset into a table named SPACEXTABLE, removing blank rows.

- **Unique Launch Sites**: Displayed the names of unique launch sites from the dataset.

- **Records Starting with 'CCA'**: Retrieved 5 records where launch sites start with 'CCA'.

- **Total Payload Mass by NASA (CRS)**: Calculated the total payload mass for boosters launched by NASA (CRS).

- **Average Payload Mass by Booster Version 'F9 v1.1'**: Computed the average payload mass carried by the booster version 'F9 v1.1'.

- **First Successful Ground Pad Landing**: Listed the date of the first successful landing outcome on a ground pad.

- **Successful Drone Ship Landings with Specific Payload Mass**: Identified booster versions that successfully landed on a drone ship with payload mass between 4000 and 6000 kg.

- **Count of Successful and Failed Mission Outcomes**: Summarized the total number of successful and failed mission outcomes.

- **Booster Versions with Maximum Payload Mass**: Listed booster versions that carried the maximum payload mass.

- **2015 Records with Failed Drone Ship Landings**: Displayed records for 2015 showing month names, booster versions, launch sites, and failed landing outcomes on drone ships.

- **Ranked Count of Specific Landing Outcomes**: Ranked the count of specific landing outcomes (Failure on drone ship and Success on ground pad) between specified dates in descending order.

Build an Interactive Map with Folium

- **Marker**: Represent NASA Johnson Space Center, success/failed launches for each site. Point of interest near the launch site

- **Circles:** Represent Launch results from the launch sites and point of interest.

- **Line**s: Represent the distances and direction between the launch sites and point of interest.

Build a Dashboard with Plotly Dash

- Scatterplot and pie graph are added to a dashboard. The Scatterplot shows the payload mass and the correlation between payload and launch success.

- Pie chart shows the total successful launches count for all sites or each of the site.

Predictive Analysis (Classification)

Data Loading:

• Loaded datasets from URLs into Pandas DataFrames.

• Displayed the first few rows to understand the structure and content.

Data Preprocessing:

• Extracted target variable Class into a NumPy array Y.

• Standardized features in X using StandardScaler.

Data Splitting:

• Split data into training and test sets with train_test_split (80% training, 20% testing).

Model Selection and Hyperparameter Tuning:

1. **Support Vector Machine (SVM):**

   • Defined an SVM model.

   • Specified hyperparameter grid for tuning (C, gamma, kernel).

   • Used GridSearchCV with 5-fold cross-validation to find the best parameters.

   • Fitted the model to training data and made predictions on test data.

   • Achieved performance insights via output logs.

2. **Logistic Regression:**

   • Defined a Logistic Regression model.

   • Specified hyperparameter grid (C, penalty, solver).

   • Used GridSearchCV with 10-fold cross-validation to find the best parameters.

   • Evaluated and printed the best parameters and validation accuracy.

   • Calculated and printed test accuracy.

   • Plotted and analyzed the confusion matrix to understand prediction errors.

3. **Additional SVM Tuning:**

   • Created another SVM model with a refined hyperparameter grid.

   • Used GridSearchCV with 10-fold cross-validation to further optimize.

Model Evaluation:

• Evaluated models on test data using accuracy metrics and confusion matrices.

• Identified false positives and other errors to assess model performance.

Best Model Identification:

• Determined the best-performing model based on validation accuracy and test accuracy scores.

Data Loading

Data Preprocessing

Data Splitting

Hyperparameter Tuning

Model Evauluation

Best Model Identification

❑ [GitHub URL of completed predictive analysis](#)

17
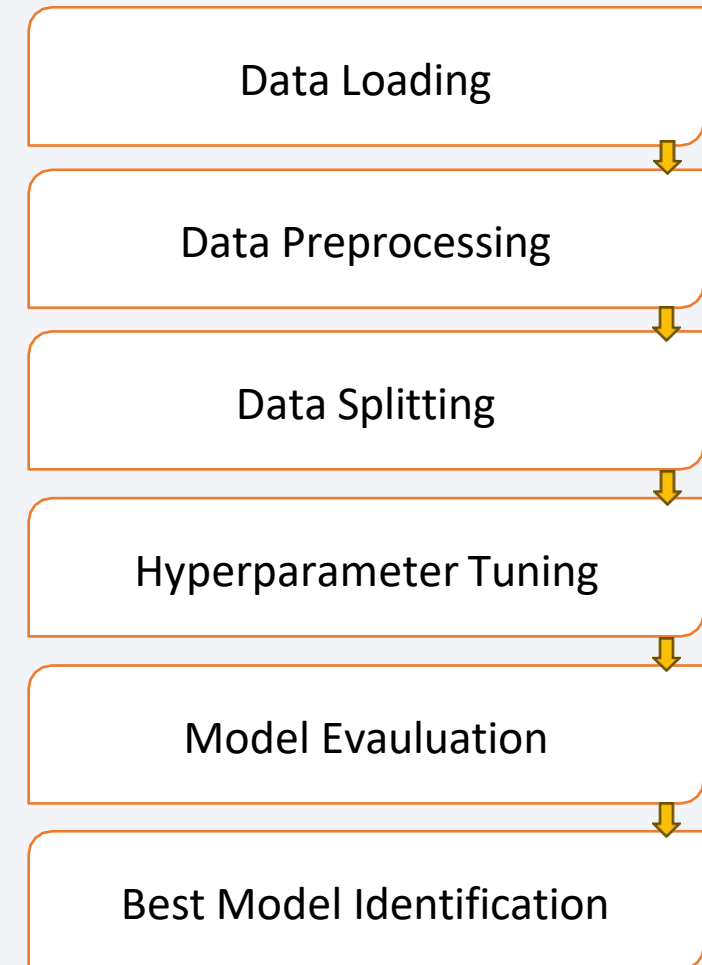
Results

- **Exploratory data analysis results:**

Flight numbers and Payload variables would affect the launch outcome. As the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

Majority of the payload are less than 8000kg across all launch sites.

Orbit ES-L1, GEO, HEO AND SSO have success rate of 100%.

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.

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.

The launch success rate since 2013 kept increasing till 2017

- **Predictive analysis results:**

Logistic Regression accuracy: 0.8333
SVM accuracy: 0.8333
Decision Tree accuracy: 0.7777
KNN accuracy: 0.8333
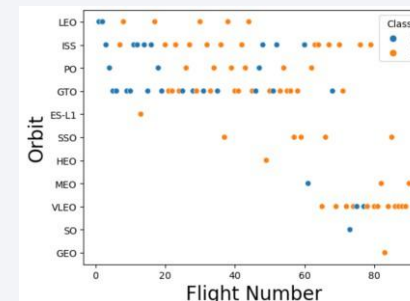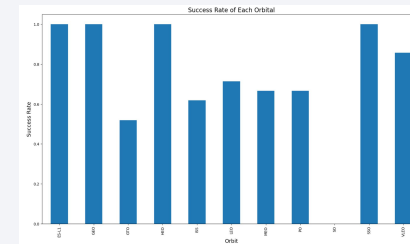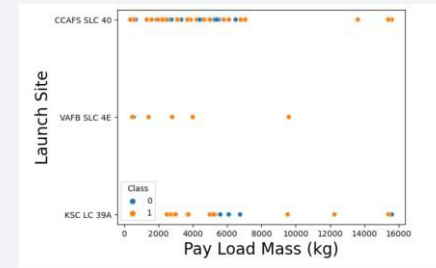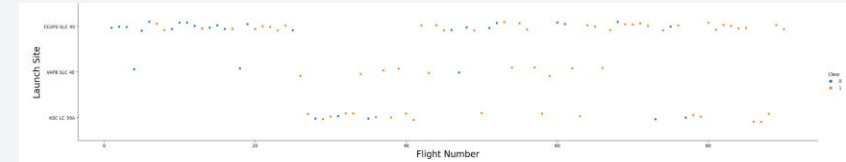
18

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site
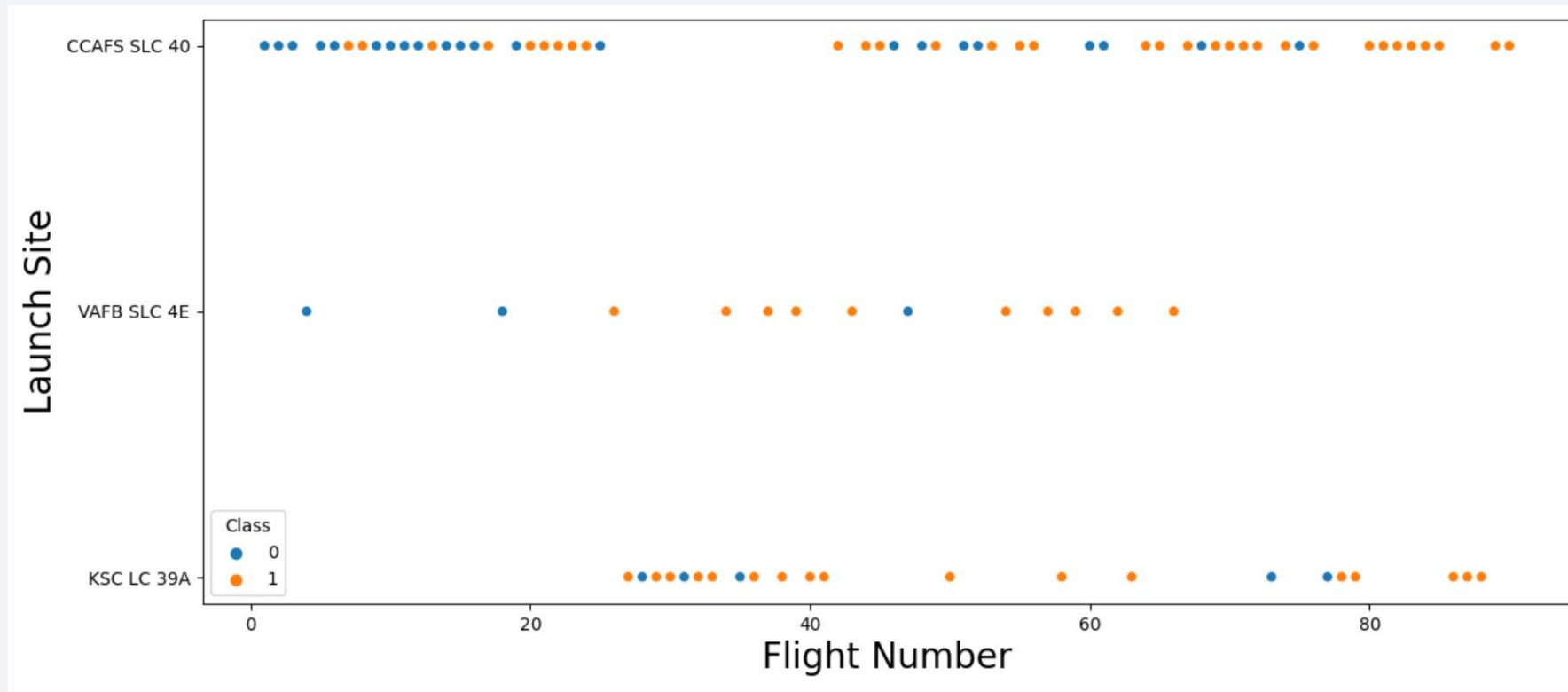
- There are much more successful launches at CCAFS SLC 40.

- VAFB SLC 4E has the least amount of launches.

- KSC LC 39A begins its launch after flight number 20.

Payload vs. Launch Site
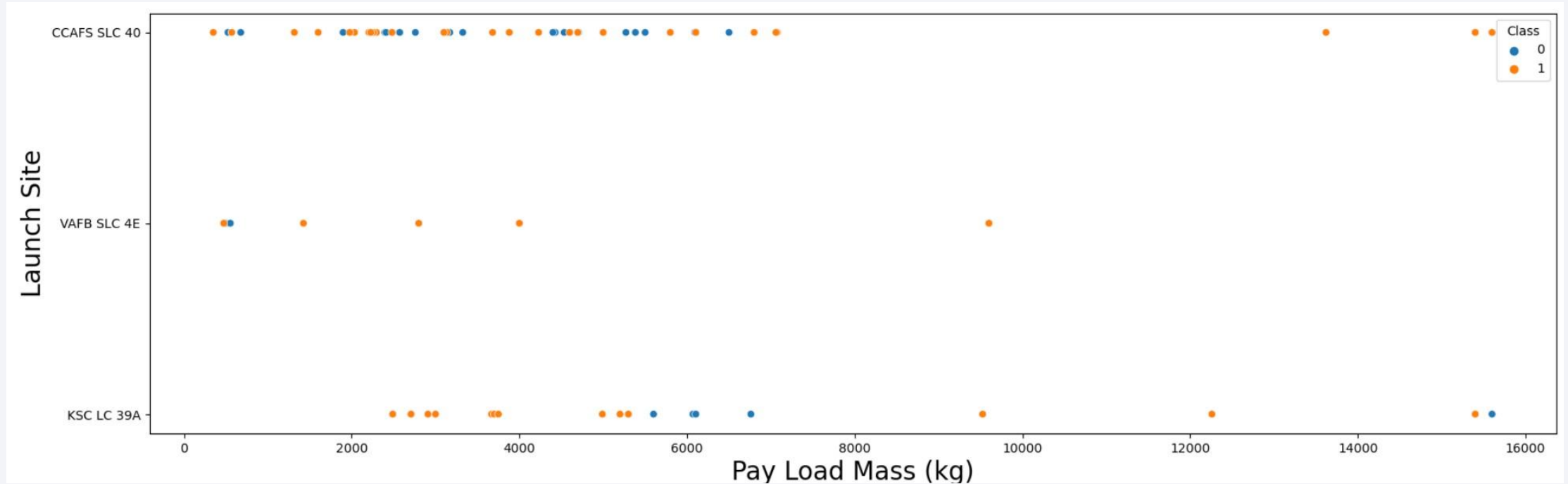
- For VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000).

- CCAFS SLC 40 has the most total Payload Mass.

- Both CCAFS SLC 40 and KSC LC 39A Payload Mass are mostly under 8000kg.

Success Rate vs. Orbit Type

- Orbit ES-L1, GEO, HEO AND SSO have success rate of 100%.

- Majority of other orbit types have a success rate between 50-70%

Success Rate of Each Orbital
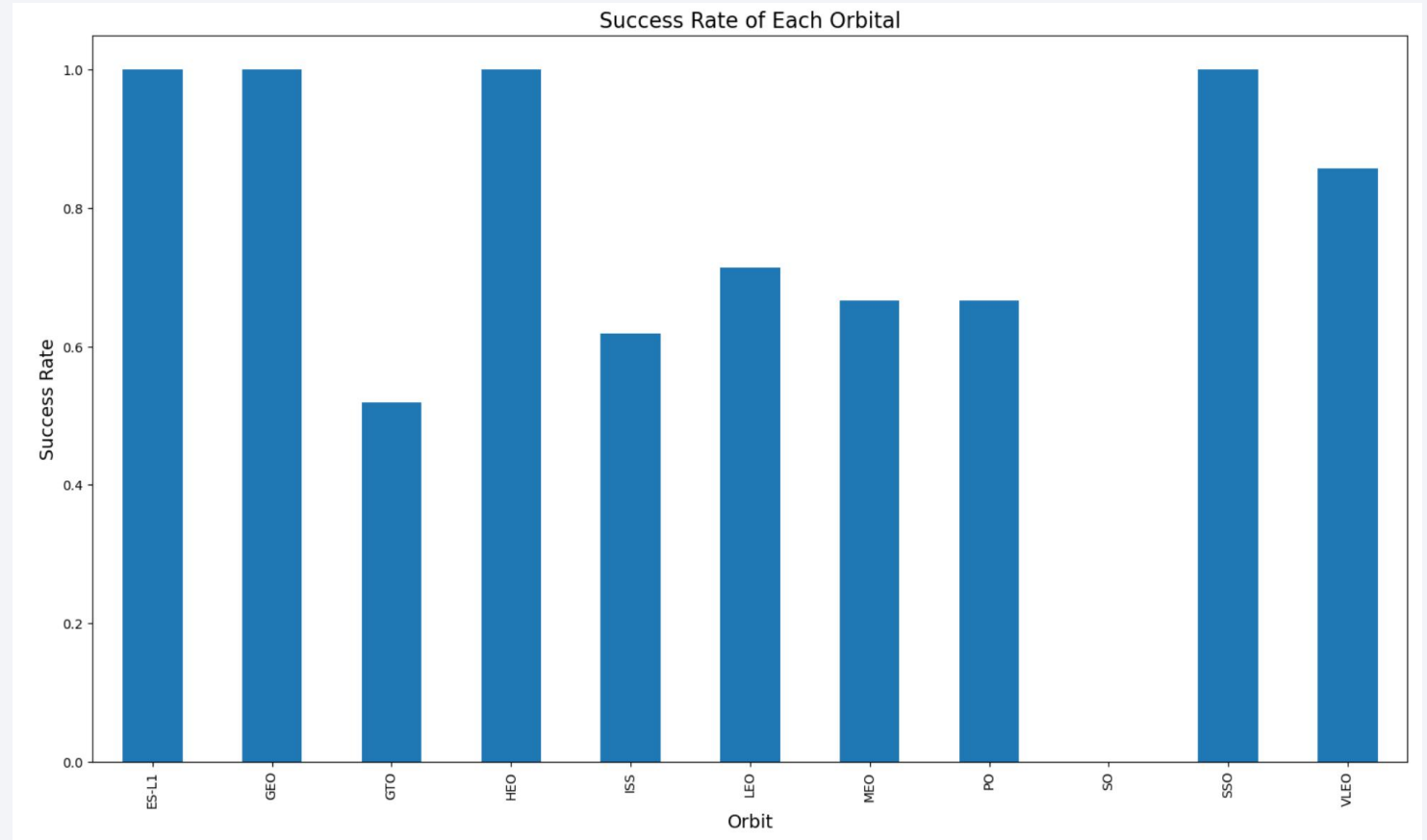
Flight Number vs. Orbit Type
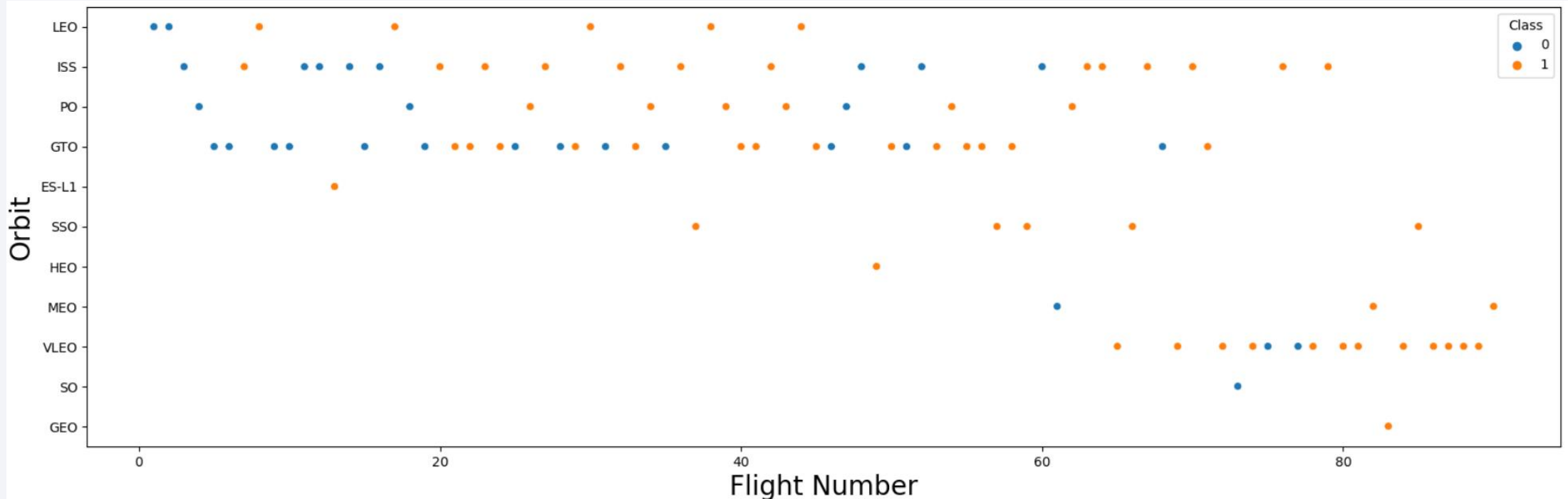
- The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

- GEO, ES-L1, SO, HEO orbits have the least amount of flight numbers.

Payload vs. Orbit Type
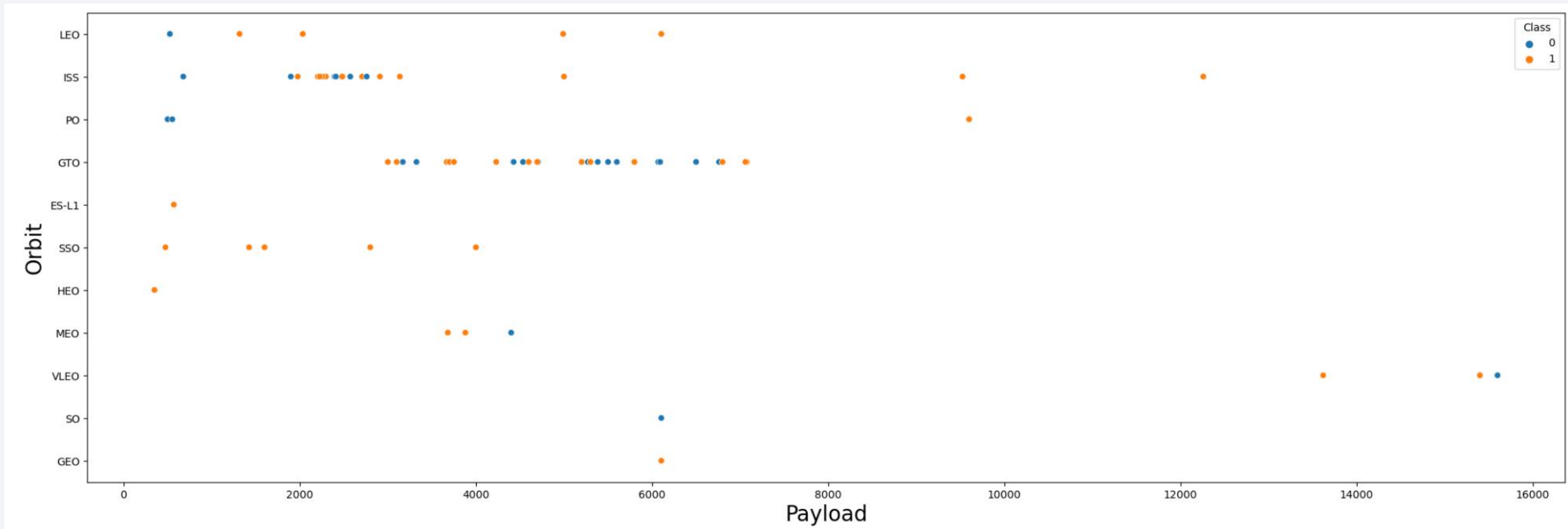
- With heavy payloads, the successful landing or positive landing rate are more for Polar, LEO and ISS.

- However for GTO we cannot distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are both there here.

- Majority of the Payload is between 2000-6000.

- ISS and GTO orbits has the most amount of Payload.

Launch Success Yearly Trend
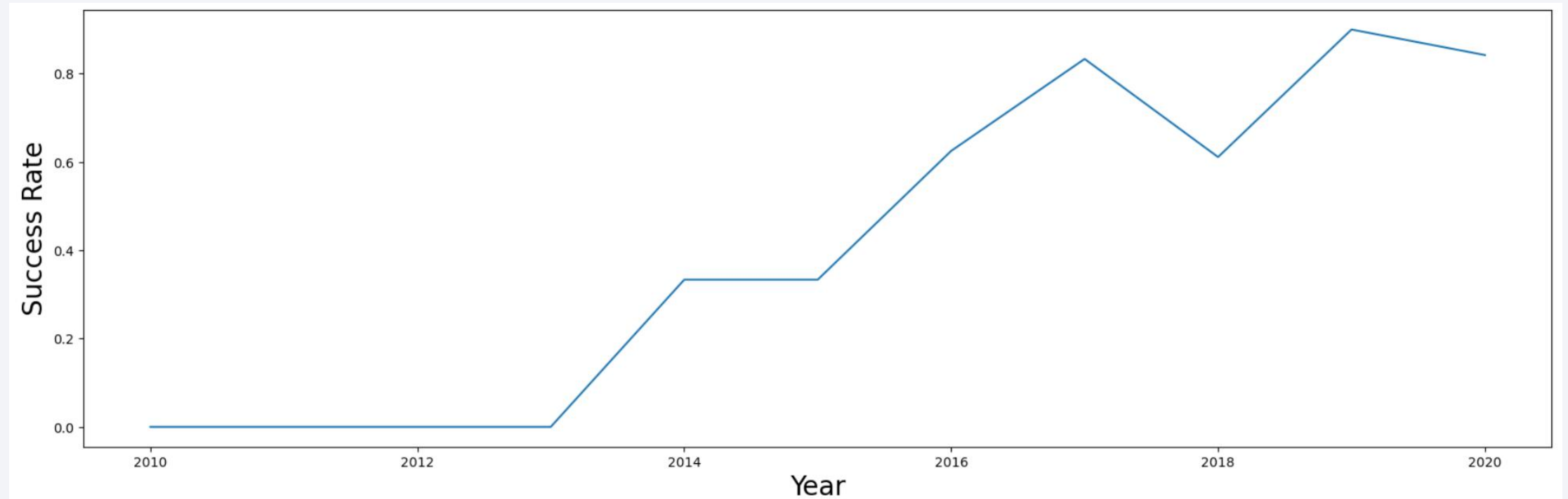
- The success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing.

All Launch Site Names

SQL queries to extract info for launch sites.

%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;

- There are four Launch Sites:

- CCAFS SLC 40

- KSC LC 39A

- VAFB-SLC

- VAFB SLC-4E

Launch Site Names Begin with 'CCA'

SQL result shows 5 records begin with CCA.

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

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

Total Payload Mass carried by Boosters launched by NASA (CRS)

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTABLE WHERE Customer LIKE 'NASA (CRS)%';
```

- Using the SQL query above we have the total Payload Mass carried by Boosters launched by NASA (CRS) is 48213kg.

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTABLE WHERE BOOSTER_VERSION = 'F9 v1.1';
```

- Using the SQL query above we have the **average payload mass of 2928.4**

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';
```

- Using the SQL query above we have the first successful landing outcome in ground pad on **12/22/2015.**

Successful Drone Ship Landing with Payload between 4000 and 6000

%sql SELECT Booster_Version FROM SPACEXTABLE WHERE Landing_Outcome='Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4001 AND 5999;

- Using the SQL query above we discovered 4 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

```sql
%%sql SELECT
SUM(CASE WHEN MISSION_OUTCOME LIKE 'Success%' THEN 1 ELSE 0 END) AS Successful_Missions,
SUM(CASE WHEN MISSION_OUTCOME LIKE 'Failure%' THEN 1 ELSE 0 END) AS Failed_Missions
FROM SPACEXTABLE;
```

- Using the SQL query above we have a total number of 100 successful and 1 failure mission outcomes

Successful_Missions	Failed_Missions
100	1

Boosters Carried Maximum Payload

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE);
```

Booster Versions
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

- Using the SQL query above we have a list of names of the booster versions which have carried the maximum payload mass

2015 Launch Records

```
%%sql SELECT
SUBSTR(Date, 6, 2) AS Month,
BOOSTER_VERSION,
LANDING_OUTCOME,
LAUNCH_SITE
FROM SPACEXTABLE
WHERE SUBSTR(Date, 0, 5) = '2015'
AND LANDING_OUTCOME LIKE 'Failure (drone ship)%';
```

- Using the SQL query on the left we have a list of launch_site for the months in year 2015.

Month	Booster_Version	Landing_Outcome	Launch_Site
01	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
04	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40

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

```sql
%%sql
SELECT LANDING_OUTCOME, COUNT(*) as Count
FROM SPACEXTABLE
WHERE (Date >= '2010-06-04' AND Date <= '2017-03-20')
AND (LANDING_OUTCOME = 'Failure (drone ship)' OR LANDING_OUTCOME = 'Success (ground pad)')
GROUP BY LANDING_OUTCOME
ORDER BY Count DESC;
```

- Using the SQL query above we have a Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20

Landing Outcome	Count
Failure (drone ship)	5
Success (ground pad)	3

Section 3

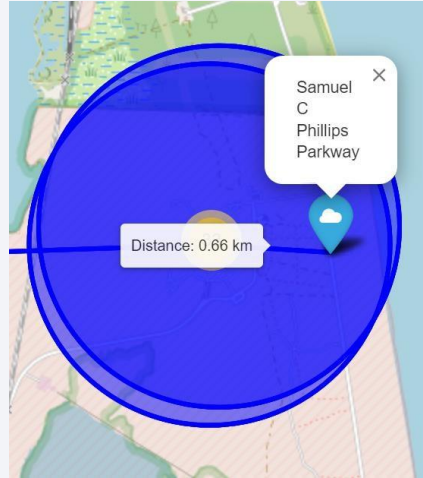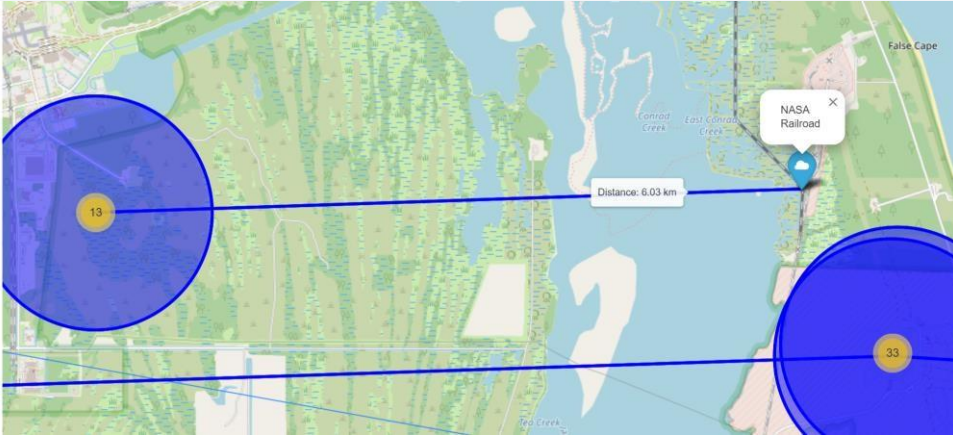# Launch Sites Proximities Analysis

Launch sites map
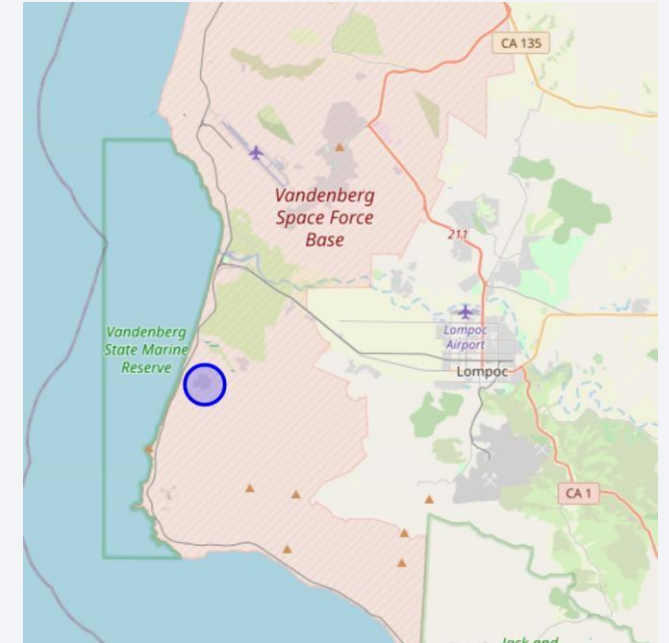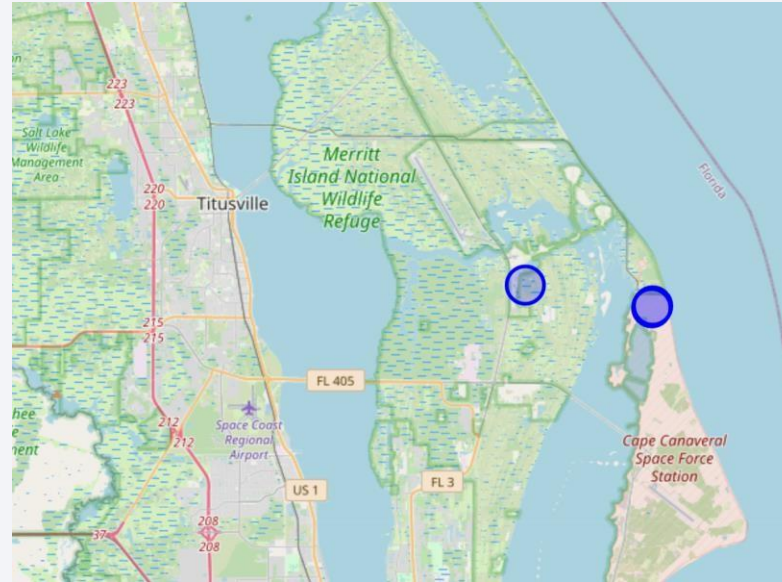
- The launch sites are mainly located in two states, Los Angles and Florida.

Map of Success/Failed Launches
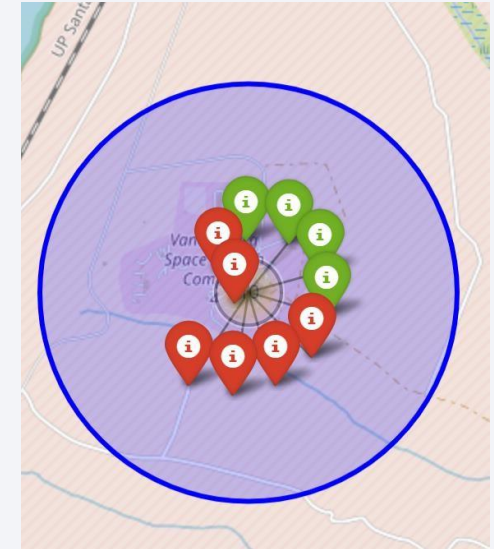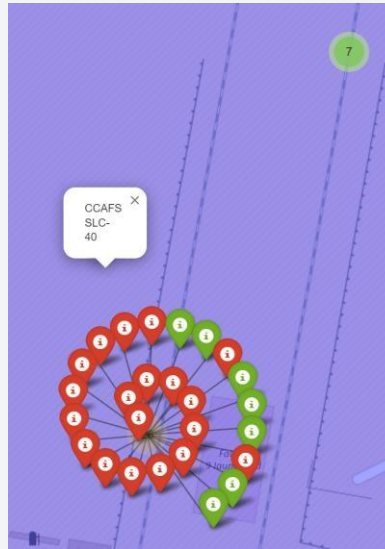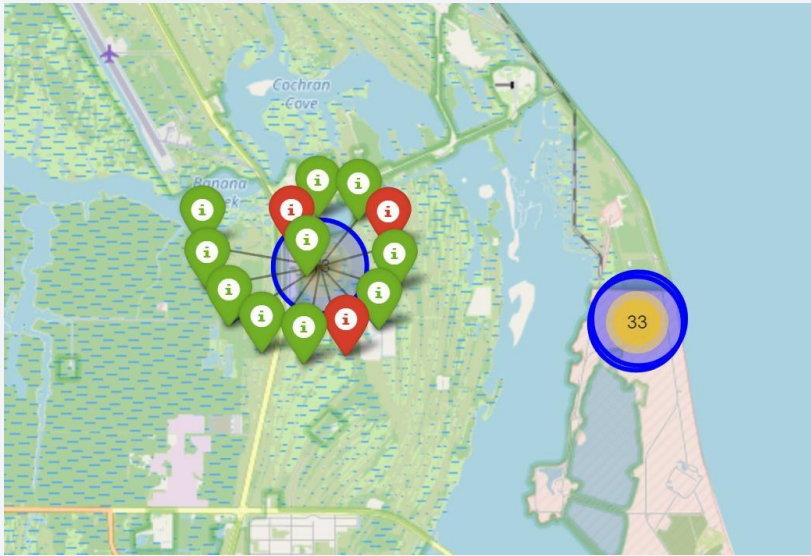
- Green colored markers show Success Launches.
- Red colored markers show Failed Launches.

Launch Site Proximities

The distance between CCAFS LC-40 and NASA Railroad is 1.61km.

The distance between CCAFS LC-40 and Samuel C Phillips Parkway is 0.66km.

The distance between CCAFS LC-40 and Orlando is 78.41km.

The distance between CCAFS SLC-40 and NASA Railroad is 1.55km.

The distance between CCAFS SLC-40 and Samuel C Phillips Parkway is 0.62km.

The distance between CCAFS SLC-40 and Orlando is 78.47km.

The distance between KSC LC-39A and NASA Railroad is 6.02km.

The distance between KSC LC-39A and Samuel C Phillips Parkway is 7.56km.

The distance between KSC LC-39A and Orlando is 71.67km.

The distance between VAFB SLC-4E and NASA Railroad is 3825.82km.

The distance between VAFB SLC-4E and Samuel C Phillips Parkway is 3827.67km.

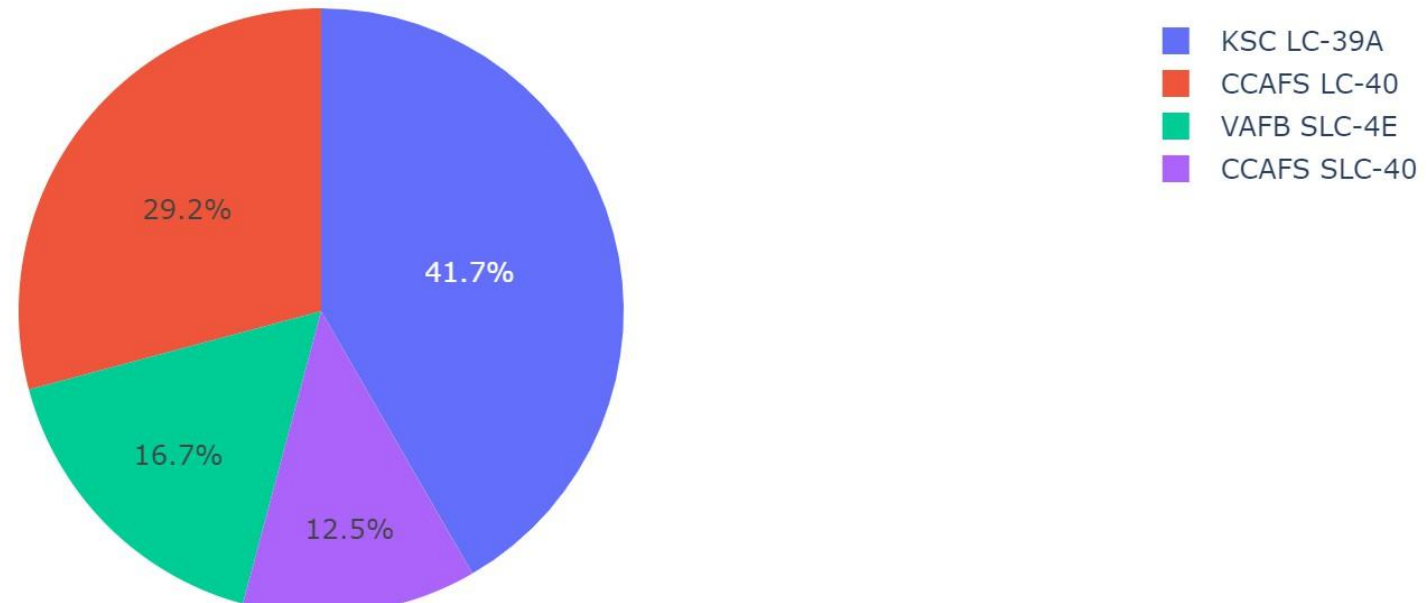The distance between VAFB SLC-4E and Orlando is 3754.54km.
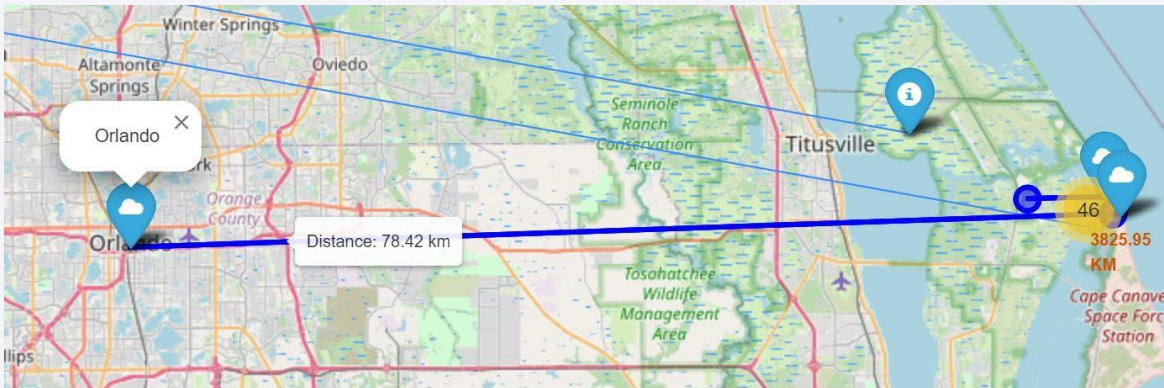
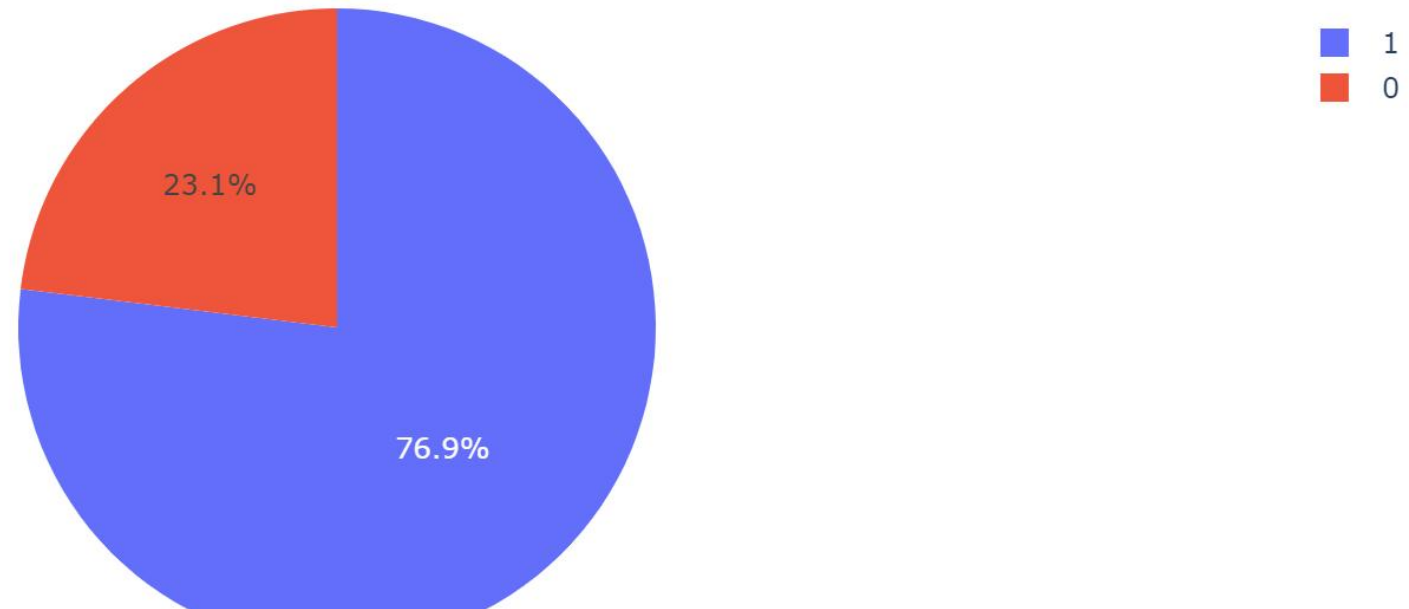# Build a Dashboard with Plotly Dash

Success Launches By Site

Total Success Launches By Site

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

41.7%
29.2%
16.7%
12.5%

- KSC LC-39A has the most successful launches and CCAFS SLC-40 has the least.

Launch Site with Highest Launch Success

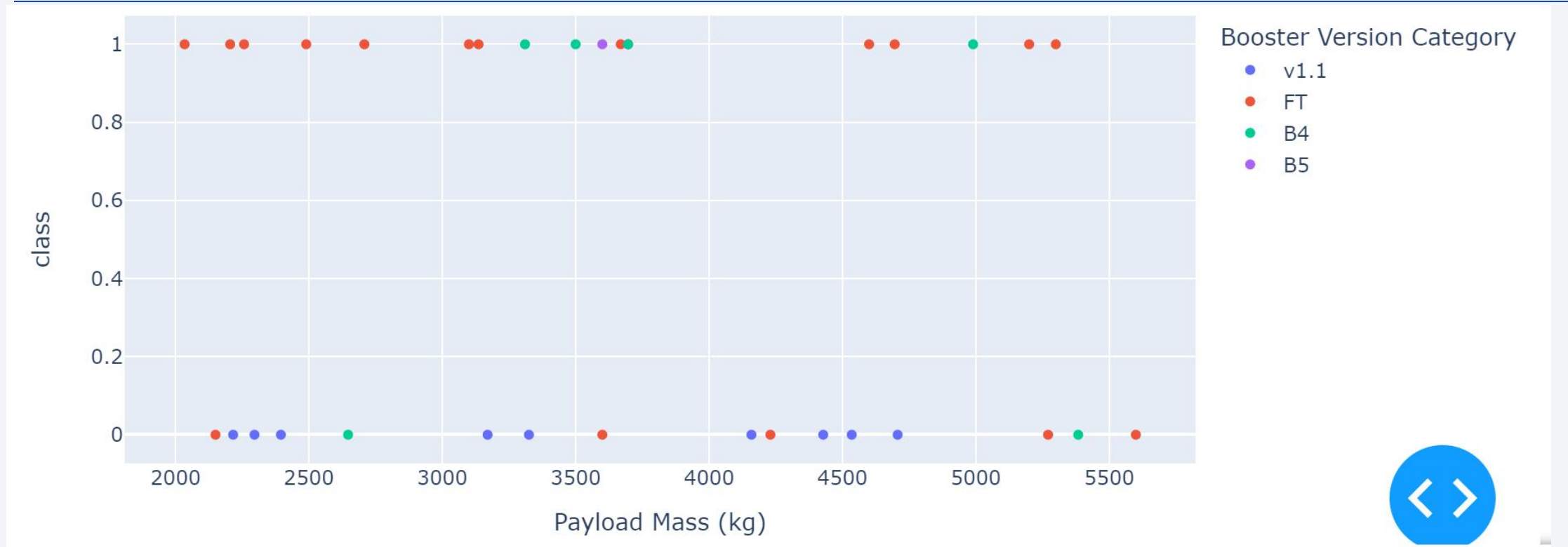Total Launches for site KSC LC-39A

- KSC LC-39A launch site has the highest launch success at 76.9%.

All Launch Sites Outcome

- Booster Version FT has the most success but v1.1 has the least successful launches.
- Most of the successful launches are between 2000 – 6000kg Payload Mass.
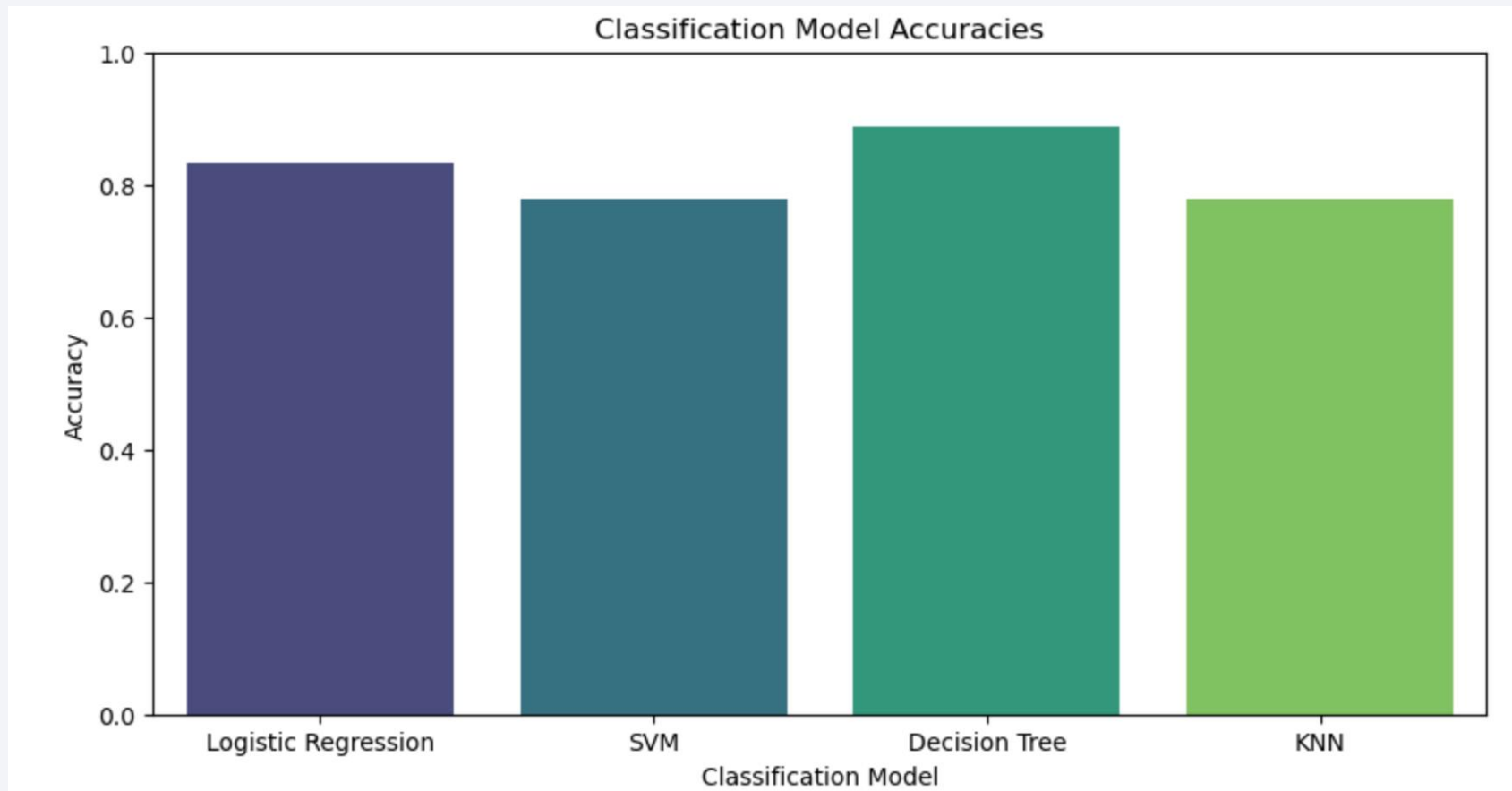
Different Class of Payload Mass.

Payload Mass between 2000 – 6000kg.

- In this category of payload mass, booster version v1.1 continue to experience difficulties.
- Booster Version FT and B4 have enjoyed success.
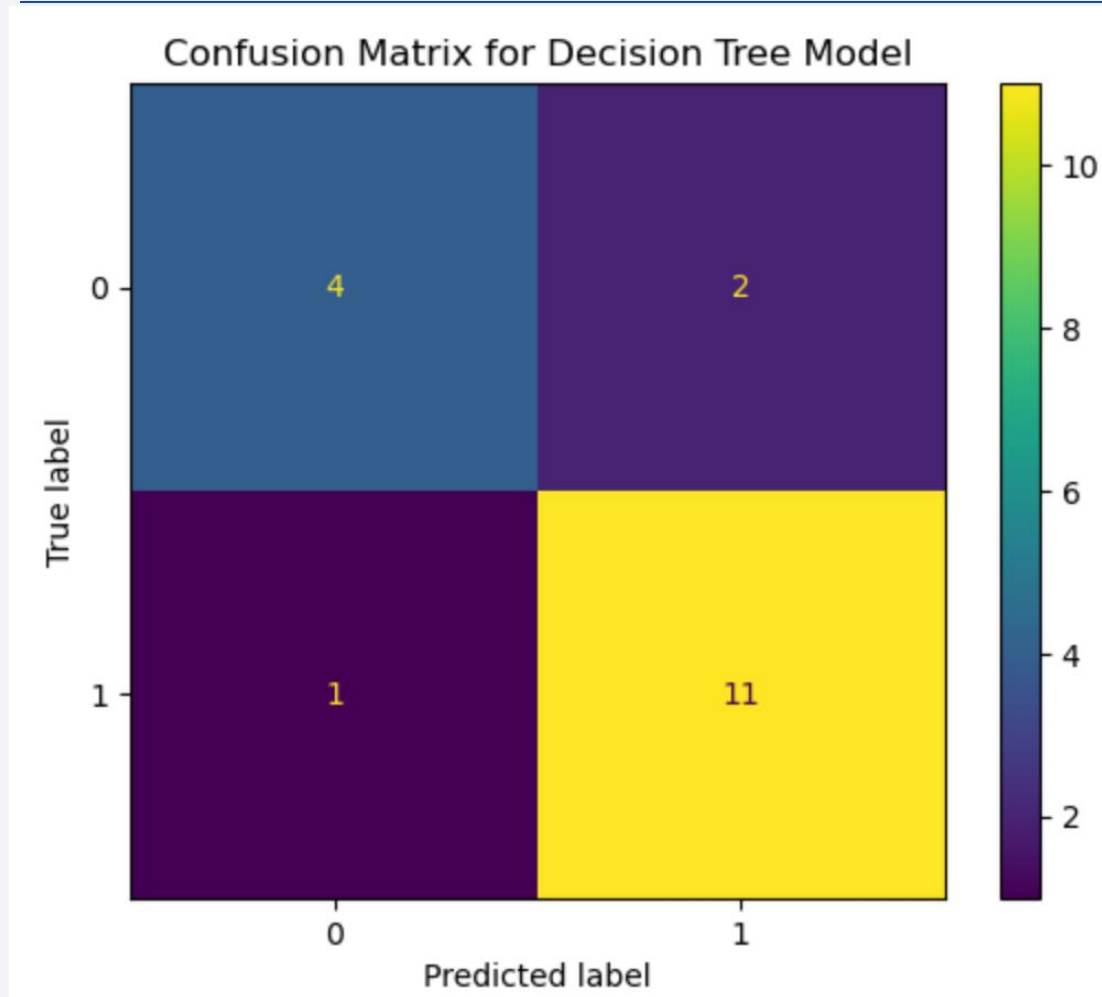
Section 5

Predictive Analysis (Classification)

Classification Accuracy

Classification Model Accuracies

- The model with the highest accuracy is Decision Tree with an accuracy of 0.89.

Confusion Matrix

Confusion Matrix for Decision Tree Model

Confusion Matrix Explanation:	
True Negatives (TN):	4
False Positives (FP):	2
False Negatives (FN):	1
True Positives (TP):	11

Conclusions

- There are four Launch Sites:
    1. CCAFS SLC 40
    2. KSC LC 39A
    3. VAFB-SLC
    4. VAFB SLC-4E
- Booster Version FT has the most success but v1.1 has the least successful launches.
- The model with the highest accuracy is Decision Tree with an accuracy of 0.89
- Most of the successful launches are between 2000 – 6000kg Payload Mass.
- KSC LC-39A launch site has the highest launch success at 76.9%.
- The launch sites are mainly located in two states, Los Angles and Florida
- There are total number of 100 successful and 1 failure mission outcomes.
- The first successful landing outcome in ground pad was on 12/22/2015.
- The average payload mass for the least successful Booster Version, F9 v1.1, is 2928.4kg.
- Total Payload Mass carried by Boosters launched by NASA (CRS) is 48213kg
- The launch success rate since 2013 kept increasing till 2020.
- With heavy payloads, the successful landing or positive landing rate are more for Polar, LEO and ISS Orbit.
- ISS and GTO orbits has the most amount of Payload.
- GEO, ES-L1, SO, HEO orbits have the least amount of flight numbers.
- Orbit ES-L1, GEO, HEO AND SSO have success rate of 100%.
- Majority of other orbit types have a success rate between 50-70%
- CCAFS SLC 40 has the most total Payload Mass.
- Both CCAFS SLC 40 and KSC LC 39A Payload Mass are mostly under 8000kg.
- There are much more successful launches at CCAFS SLC 40.
- VAFB SLC 4E has the least amount of launches.

Appendix

- Serisouly import pandas as pd

- Why import numpy as np

- This import matplotlib.pyplot as plt

- Is import seaborn as sns

- Here from sklearn import preprocessing

- I mean from sklearn.model_selection import train_test_split

- For thisfrom sklearn.tree import DecisionTreeClassifier

- Project from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay, accuracy_score?

Thank you!