IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ana Belén Robles
November 27, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with Data Visualization
  - ExploratoryData Analysis with SQL
  - Interactive Visual Analytics with Folium
  - Predictive Analysis with Machine Learning
- Summary of all results
  - Exploratory Data Analysis
  - Interactive Analytics in Screenshots
  - Predictive Analytics Results

Introduction

- Project background and context

  SpaceX offers Falcon 9 rocket launches on its website at a cost of $62 million, compared to other providers whose prices can exceed $165 million per launch. Much of this cost reduction is due to SpaceX's ability to reuse the rocket's first stage. Therefore, if we can determine whether the first stage will successfully land, we can also estimate the cost of the launch. This information is crucial for other companies looking to compete against SpaceX in bidding for space launches. The primary goal of this project is to develop a machine learning pipeline to predict whether the first stage will land successfully.

- Problems you want to find answers
  - What factors determine whether a rocket will land successfully?
  - How do the various features interact to influence the success rate of landings?
  - What operating conditions need to be fulfilled to ensure a successful landing program?

Section 1

Methodology

Methodology (I)

Executive Summary

- Data collection methodology:
  - SpaceX API: https://api.spacexdata.com/v4/rockets/
  - Webscraping from Wikipedia: https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_l
- Perform data wrangling

The data underwent a comprehensive wrangling process to ensure it was ready for analysis. One-hot encoding was applied to transform categorical features into numerical format, facilitating their integration into machine learning models. Additionally, the dataset was enriched by creating a new landing outcome label. This label was derived from outcome data after summarizing and analyzing key features, enabling a more structured representation of the target variable for predictive modeling.

Methodology (II)

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

The collected data was preprocessed through normalization to ensure consistent feature scaling. The dataset was then divided into training and test subsets to evaluate model performance effectively. Four distinct classification models were built, each tested and tuned using various combinations of hyperparameters to optimize performance. Model evaluation was conducted based on accuracy metrics, providing insights into the predictive capabilities and reliability of each model under different configurations.

Data Collection

- Describe how data sets were collected.

The datasets were gathered using a combination of API requests and web scraping techniques. Data from the SpaceX API (https://api.spacexdata.com/v4/rockets/) was collected via GET requests, with the response content decoded as JSON using the .json() function and subsequently transformed into a pandas DataFrame using .json_normalize. Additionally, launch records were extracted from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches) through web scraping with BeautifulSoup. The HTML tables containing launch information were parsed and converted into pandas DataFrames for further analysis. During this process, the data was cleaned, missing values were identified and filled where necessary, ensuring the datasets were complete and ready for analysis.

Data Collection – SpaceX API

Data Collection Overview

- The SpaceX API (https://api.spacexdata.com/v4/rockets/) was utilized to collect structured data on Falcon 9 launches.
- The process involved sending **GET requests** to retrieve data in JSON format.
- The response was parsed using Python and converted into a **pandas DataFrame** for further analysis.

Steps Performed

- **API Call:** Utilized the requests library to send GET requests to the SpaceX REST API.
- **Data Parsing:** Decoded the JSON response using .json() and normalized nested data with pandas.json_normalize().
- **Data Cleaning:** Ensured data completeness by identifying and filling missing values.

External Reference
For complete code and results, refer to the GitHub notebook: https://github.com/abrob-hub/MyCapstone/blob/main/jupyter-labs-spacex-data-collection-api-resuelto.ipynb

Access SpaceX API Endpoint

Send JSON Request

Receive JSON Response

Parse JSON Response

Normalize JSON Data

Save Data to a Pandas DataFrame

Sort Data in Pandas DataFrame

Export Data to CSV

Data Collection - Scraping

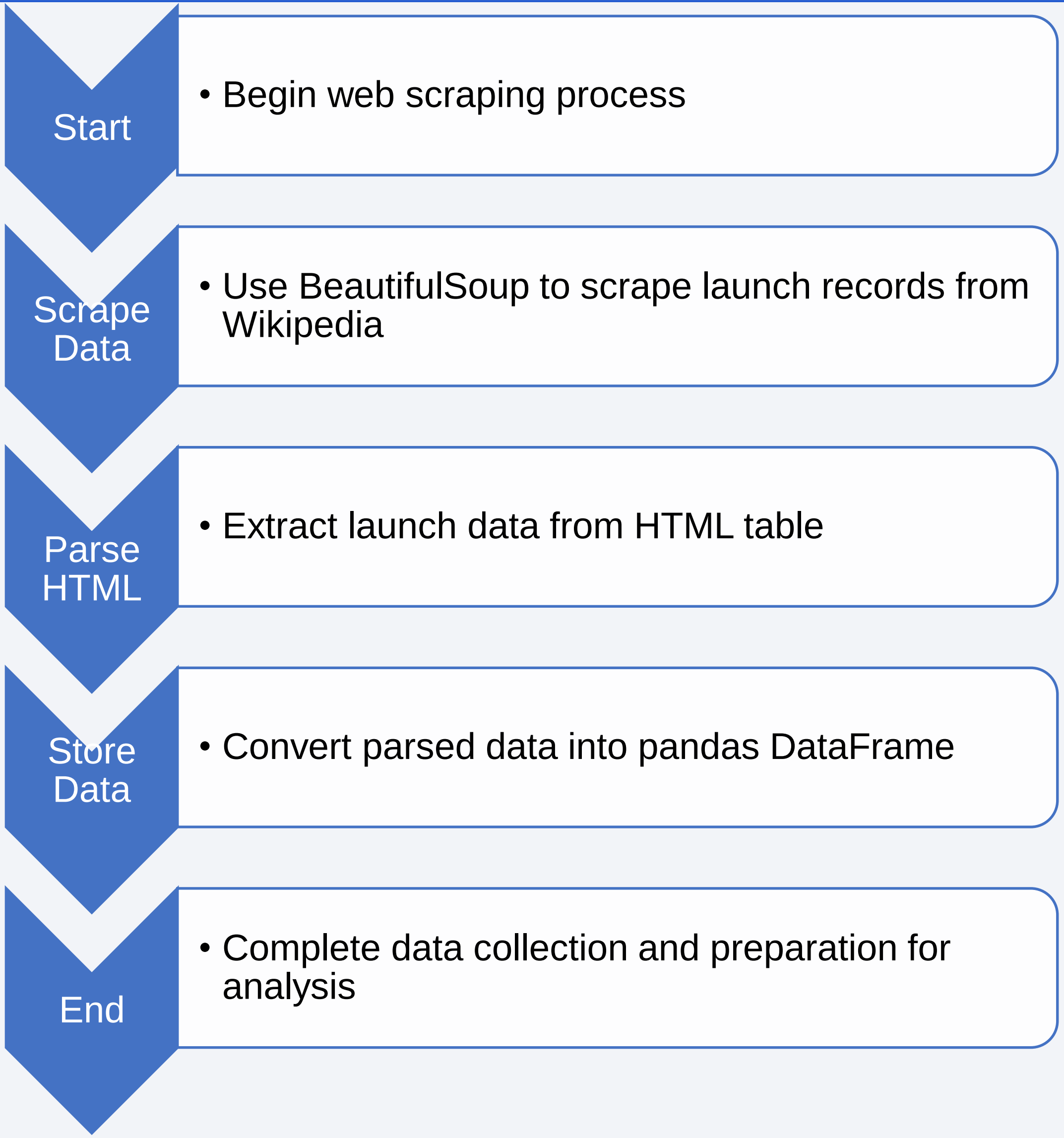
Step 1: Web scraping was conducted on Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches) using BeautifulSoup to collect Falcon 9 launch records.

Step 2: The scraped HTML table was parsed to extract the relevant launch data.

Step 3: The data was then structured and converted into a pandas DataFrame for future analysis.

External Reference
For complete code and results, refer to the GitHub notebook

https://github.com/abrob-hub/MyCapstone/blob/main/jupyter-labs-webscraping-resuelto.ipynb



Data Wrangling

Overview

- We began with Exploratory Data Analysis (EDA) to understand the dataset's structure. Key insights were gathered, including the number of launches per site, the occurrence of each orbit type, and mission outcomes by orbit category.
- A new feature, the "Landing Outcome" label, was created from the existing "Outcome" column to classify launches as successful or unsuccessful.
- Finally, the cleaned data, including the new labels, was exported to CSV for further analysis and model development.

External Reference
For complete code and results, refer to the GitHub notebook

https://github.com/abrob-hub/MyCapstone/blob/main/labs-jupyter-spacex-Data%20wrangling-resuelto.ipynb

Exploratory Data Analysis (EDA)

Count launches per site

Analyze occurrences of each orbit type

Analyze mission outcomes by orbit type

Create Landing Outcome label

Export cleaned data to CSV

EDA with Data Visualization

Plotted charts

Type of Chart	EDA Goal
Scatterplot Chart	Visualize the relationship between Flight Number and Launch Site
Scatterplot Chart	Visualize the relationship between Payload Mass and Launch Site
Bar Chart	Visualize the relationship between success rate of each orbit type
Scatterplot Chart	Visualize the relationship between FlightNumber and Orbit type
Scatterplot Chart	Visualize the relationship between Payload Mass and Orbit type
Line Chart	Visualize the launch success yearly trend

External Reference
For complete code and results, refer to the GitHub notebook

https://github.com/abrob-hub/MyCapstone/blob/main/jupyter-labs-eda-dataviz-v2-resuelto.ipynb

EDA with SQL

SQL queries performed

- Retrieve the unique names of launch sites
- Display 5 records where launch sites start with "CCA"
- Calculate the total payload mass carried by NASA (CRS) launchers
- Calculate the average payload mass carried by the F9 v1.1 rocket version
- List the date of the first successful landing on a ground platform
- List rockets with successful landings on drone ships and payload mass between 4000 and 6000 kg
- Count the total number of successful and failed mission outcomes
- List rockets that carried the maximum payload using a subquery
- List records with months, failed drone ship landing outcomes, rocket versions, and launch sites for 2015
- Rank the number of landing outcomes between 2010-06-04 and 2017-03-20 in descending order

In addition to these queries, tasks such as loading the dataset into an SQLite table, cleaning blank rows, and executing the queries in the context of SpaceX space data analysis are performed.

External Reference
For complete code and results, refer to the GitHub notebook

https://github.com/abrob-hub/MyCapstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite-resuelto.ipynb

Build an Interactive Map with Folium

Following map objects were created using Folium:

Adding these objects makes the map more interactive and analytical, helping to explore relationships between launch sites and their outcomes.

Markers and Circles: Improve visualization and allow exploration of the distribution of launch sites.

- **Markers**: Used to pinpoint the launch sites, displaying the site name on the map for easy identification.
- **Circles**: Added around each site with a 1000-meter radius, highlighting the area and improving visualization, along with a popup displaying the site name.

Marker Clusters: Manage marker density, improving map readability.

- To group markers and prevent overlap when multiple launches have the same coordinates. Marker colors (green for success and red for failure) indicate the launch outcome.
- Helps maintain map clarity by grouping markers, avoiding visual clutter.

Popup and Information

- **Launch Outcome**: Each marker has a popup showing the launch site name and the launch result (success or failure), making the map more interactive and informative.

Mouse Position Plugin: Facilitates exploring geographic features and measuring distances.

- Displays the latitude and longitude of any point on the map when the mouse hovers over it, useful for exploring distances and nearby geographical features.

External Reference
For complete code and results, refer to the GitHub notebook

https://github.com/abrob-hub/MyCapstone/blob/main/lab_jupyter_launch_site_location-resuelto.ipynb

Build a Dashboard with Plotly Dash

Here is a summary of the plots/graphs and interactions added to the dashboard, along with the explanation of why they were added:

Dropdown for Launch Site Selection

- Allows users to select a specific launch site or view data from all sites, providing flexibility in exploring the dataset and visualizing the launches of interest.

Pie Chart for Success vs Failure Count

- This plot visualizes the proportion of successful and failed launches. It helps the user quickly assess the launch success rate either globally (for all sites) or per site (based on the dropdown selection).

Payload Range Slider

- Provides an interactive way for users to filter the data by payload mass. This allows users to focus on launches with specific payload sizes, offering better control over the data analysis.

Scatter Plot for Payload vs Launch Success

- Visualizes the relationship between payload mass and launch success. It allows users to analyze how payload size affects the success or failure of launches. Additionally, this plot is dynamic and updates based on the selected launch site and payload range.

These elements make the dashboard interactive and informative, allowing users to explore the SpaceX launch data in an engaging way, tailored to their specific interests and needs.

External Reference
For complete code and results, refer to the GitHub code

https://github.com/abrob-hub/MyCapstone/blob/main/spacex_dash_app-resuelto.py

Predictive Analysis (Classification)

Key Steps in the Model Development Process:

Data Preprocessing:

- **Exploratory Data Analysis (EDA)**: The first step is to explore the dataset to understand the structure, detect any missing values, and visualize distributions. This gives insight into what transformations may be required.
- **Label Creation**: A new target variable (Class) is created, indicating whether the rocket stage landed or not.
- **Standardization**: The features (X) are standardized using StandardScaler to ensure all variables are on the same scale, which is important for algorithms like SVM and Logistic Regression.
- **Train-Test Split**: The dataset is split into training and test sets using train_test_split, with 80% used for training and 20% for testing.

Model Training:

- Four classification models were trained:
  - **Logistic Regression**: A simple and fast model used as a baseline.
  - **Support Vector Machine (SVM)**: A powerful model that tries to find a hyperplane separating the classes with the largest margin.
  - **Decision Tree**: A non-linear classifier that creates rules based on feature splits.
  - **K-Nearest Neighbors (KNN)**: A simple, instance-based learning algorithm that classifies based on the nearest data points.

Each model was optimized using **GridSearchCV** to find the best hyperparameters. This process involves:

- Defining a dictionary of hyperparameters to search over.
- Using cross-validation to evaluate different hyperparameter combinations.
- Finding the best-performing parameters for each model.

Predictive Analysis (Classification)

Model Evaluation:

- The accuracy of each model is calculated on the test set using the score method.
- A **confusion matrix** is used to visualize the performance, showing the true positives, false positives, true negatives, and false negatives.
- The **accuracy score** is computed for each model on the test data.

Model Comparison:

- The accuracies of each model are compared, and the model with the best accuracy is selected.
- This process allows you to determine which algorithm performs the best on this specific task.

Predictive Analysis (Classification)

Flowchart for Model Development:

This process ensures that the models are optimized, evaluated, and compared systematically to select the best one for predicting the landing of the SpaceX Falcon 9 rocket stage.

Data Preprocessing
- EDA & Data Cleaning
- Create Labels
- Standardize Data (X)
- Split Data into Train-Test (80-20)

Model Training
- Logistic Regression (GridSearchCV)
- SVM (GridSearchCV)
- Decision Tree (GridSearchCV)
- KNN (GridSearchCV)

Model Evaluation
- Calculate Accuracy on Test Data
- Plot Confusion Matrix

Model Comparison
- Compare Accuracy of Models
- Select Best Performing Model

External Reference
For complete code and results, refer to the GitHub notebook

https://github.com/abrob-hub/MyCapstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5-resuelto.ipynb

Results

- Exploratory data analysis results
  - SpaceX operates from four distinct launch sites.
  - Initial launches were conducted for SpaceX itself and NASA.
  - The average payload capacity of the Falcon 9 v1.1 booster is 2,928 kg.
  - The first successful booster landing occurred in 2015, five years after SpaceX's initial launch.
  - Several versions of the Falcon 9 booster achieved successful landings on drone ships, with payloads exceeding the average capacity.
  - Nearly 100% of SpaceX's mission outcomes have been successful.
  - In 2015, two Falcon 9 v1.1 booster versions, specifically B1012 and B1015, failed to land successfully on drone ships.
  - Over time, the success rate of booster landings has significantly improved.

Results

- Interactive analytics demo
  - ✓ Launch sites are located in close proximity to coastlines.
  - ✓ Most launches take place at locations along the East Coast.
  - ✓ The launch site with the highest success rate is KSC LC-39-A.
  - ✓ The launch site with the highest failure rate is CCAFS LC-40.
  - ✓ Launch sites are near railways and highways while maintaining a certain distance from urban areas.

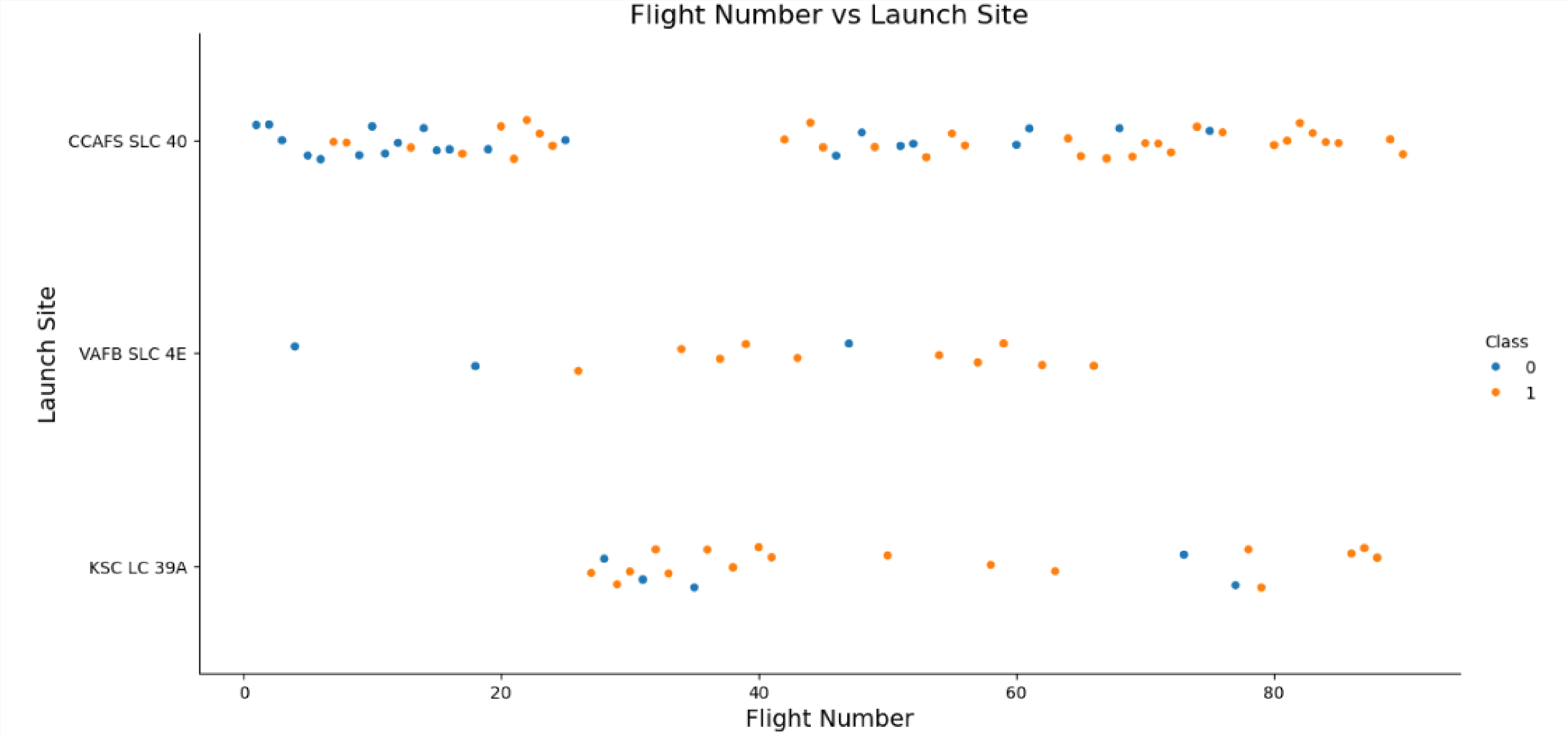
Results

- Predictive analysis results
  - ✓ Key metric: accuracy on the test dataset.
  - ✓ The Decision Tree model is discarded due to its lower accuracy on the test dataset.
  - ✓ The Logistic Regression, SVM, and KNN models exhibit the same accuracy, as shown in the confusion matrices, where the same number of true positives (12) and true negatives (3) are generated.
  - ✓ Given the observations for each model, the Logistic Regression model is recommended due to its simplicity, high accuracy, and ability to distinguish between classes effectively.

Section 2

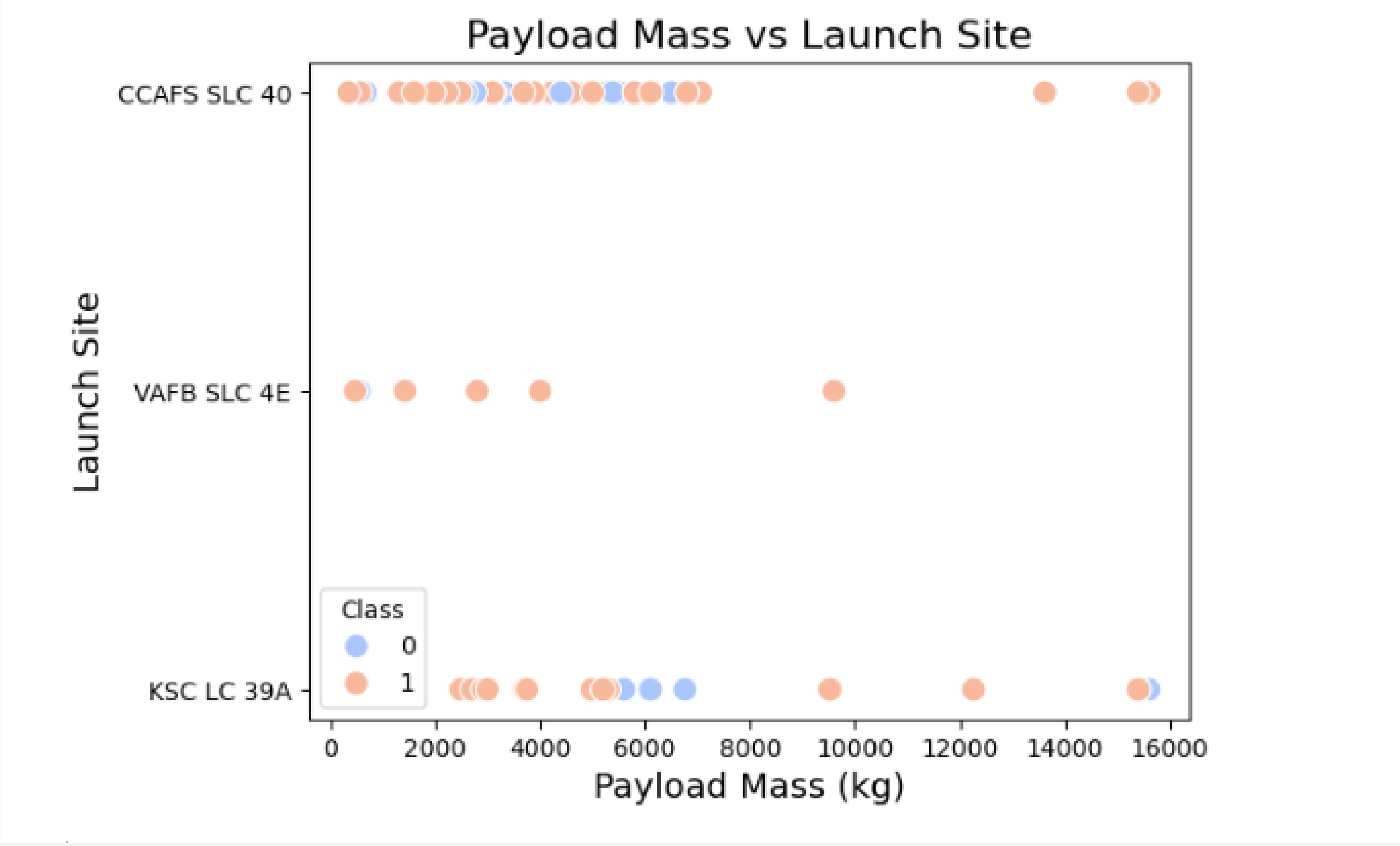
Insights drawn from EDA

Flight Number vs. Launch Site



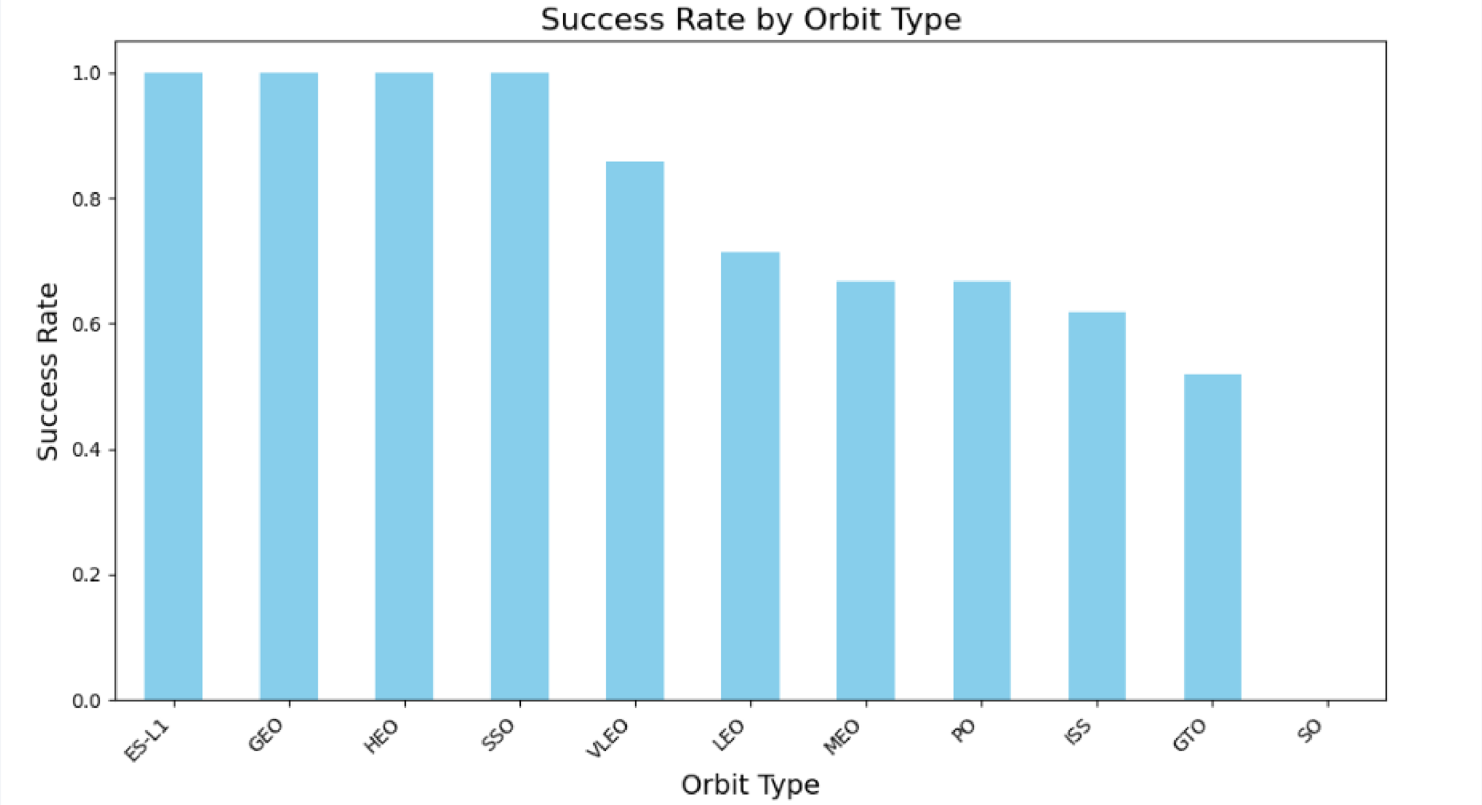
As the Flight Number increases, it can be observed that the number of successful launches also increases at each launch site.
It must be taken into account that Flight Number is correspondent to Launching Date.

Payload vs. Launch Site



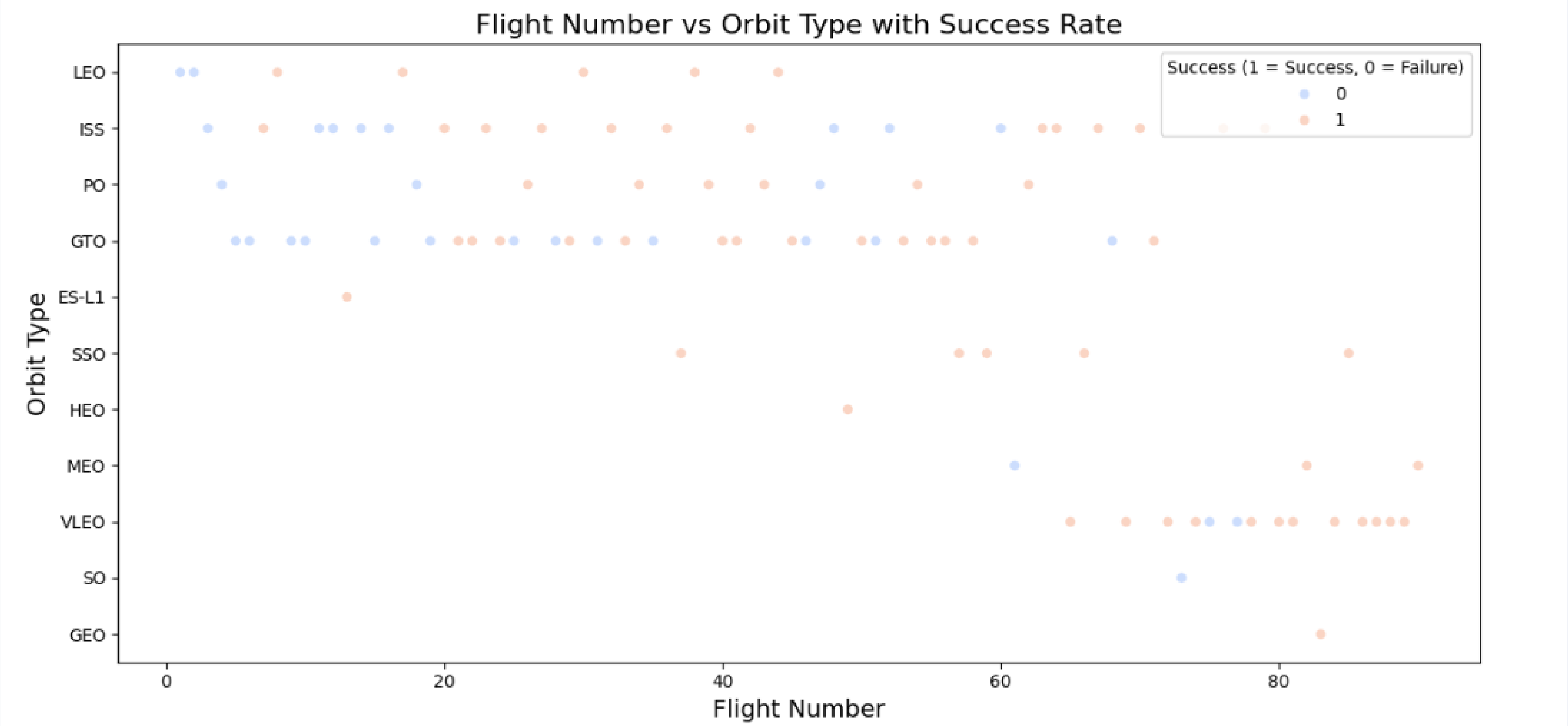
For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).
For the CCAFS SLC 40 the launch of heavy payload mass rockets (greater than 8000) has been always succesful.

Success Rate vs. Orbit Type



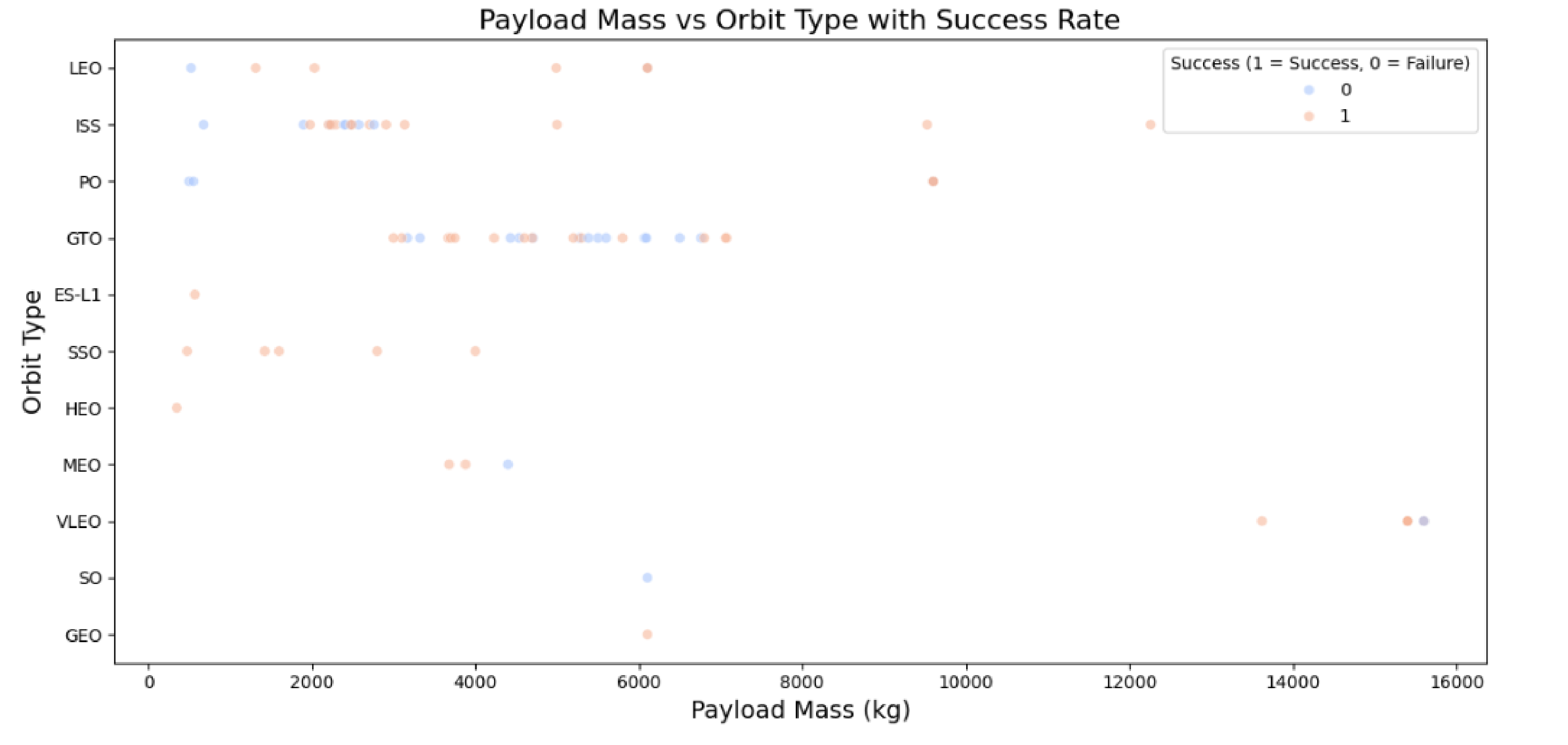
ES-L1, GEO, HEO, SSO are the orbits that have the highest success rates

Flight Number vs. Orbit Type



You can observe that 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.

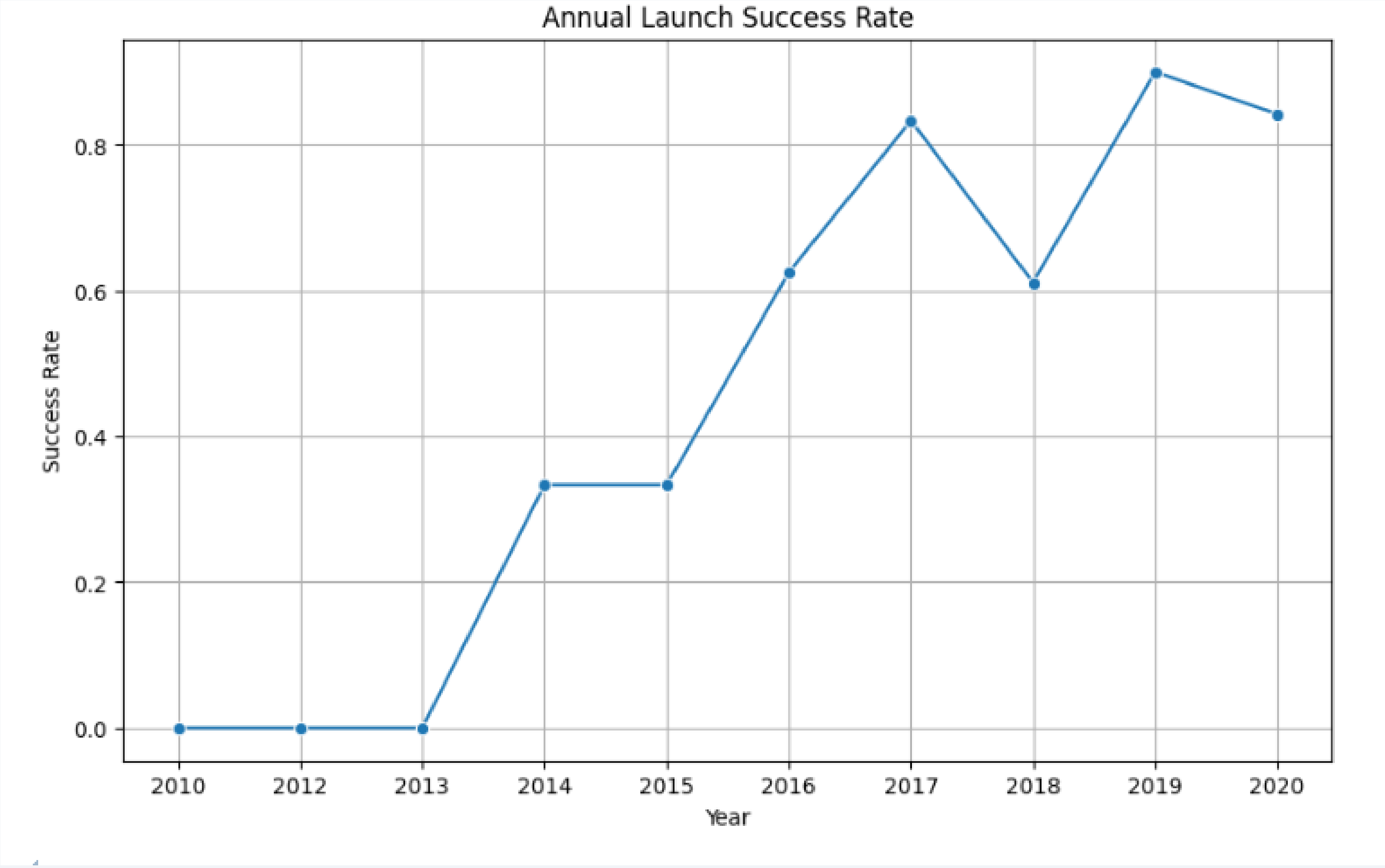
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

You can observe that the sucess rate since 2013 kept increasing till 2020



All Launch Site Names

- There are four launch sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

Query

Task 1

Display the names of the unique launch sites in the space mission

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

Results

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

Task 2

Display 5 records where launch sites begin with the string 'CCA'

Query

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

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

Results

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 carried by boosters from NASA: 48213

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

Query

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

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

Results

Total_Payload_Mass
48213

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1: 2928.4

Task 4

Display average payload mass carried by booster version F9 v1.1

Query

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

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

Results

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad: 2015-12-22

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

Query

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

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

Results

First_Successful_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000: F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

Query

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

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

Results

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful mission outcomes: 100
- Total number of successful mission outcomes: 1

Task 7

List the total number of successful and failure mission outcomes

Query

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

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

Results

Mission_Outcome	Outcome_Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

Query

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

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

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

Results

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

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015here

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

Query

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

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

Results

Month	Landing_Outcome	Booster_Version	Launch_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

Task 10

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.

Query

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

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

Results

Landing_Outcome	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

Launch sites location

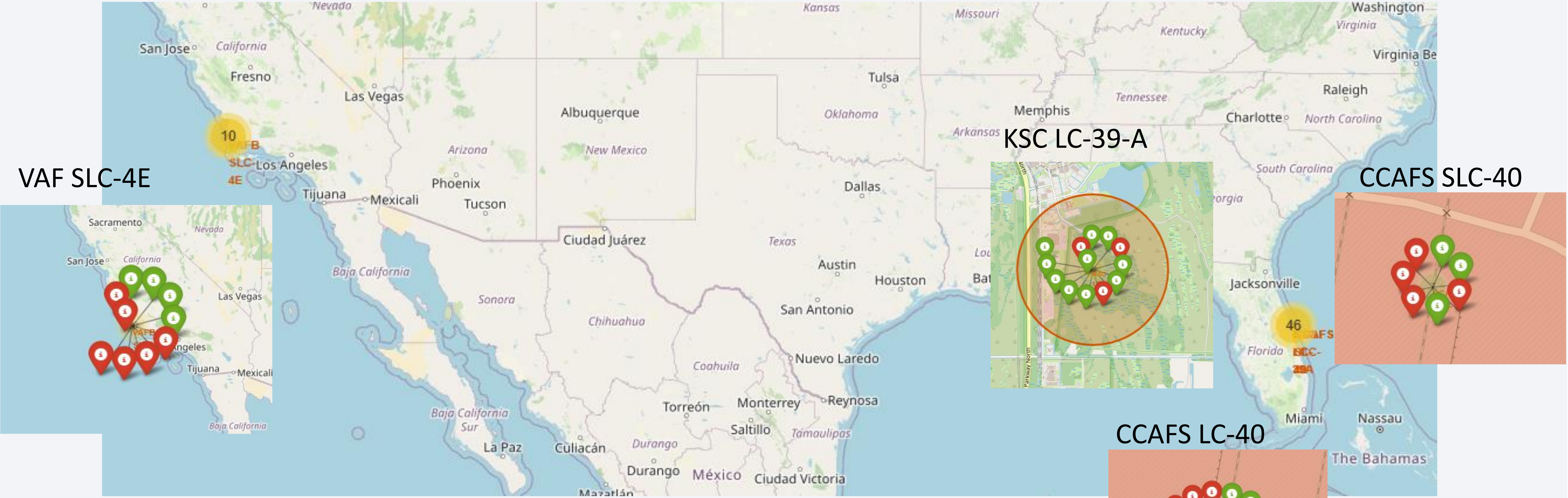
- Explain the important elements and findings on the screenshot

SpaceX launch sites are located in USA, in close proximity to Florida and California coastlines

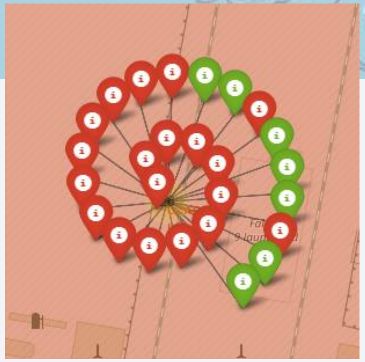
Launch sites location

Most launches take place at locations along the East Coast

Successful / Failure launches

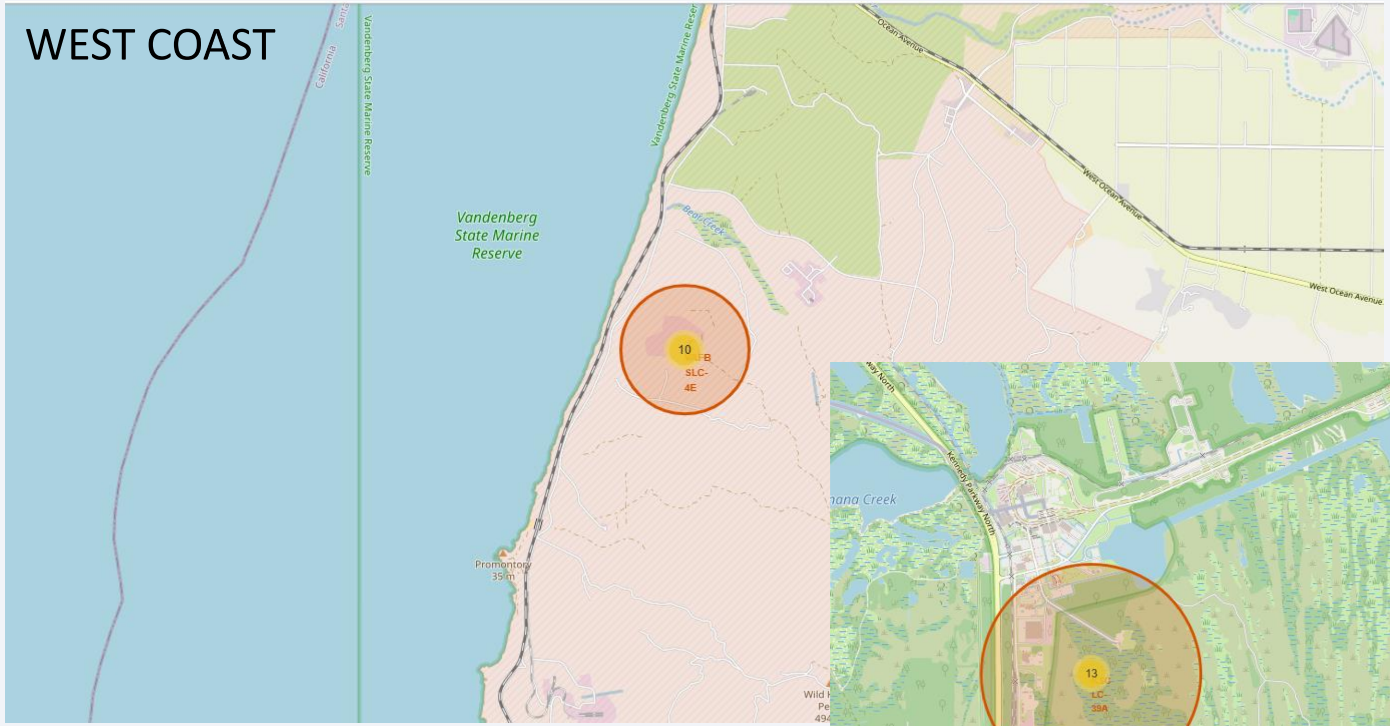


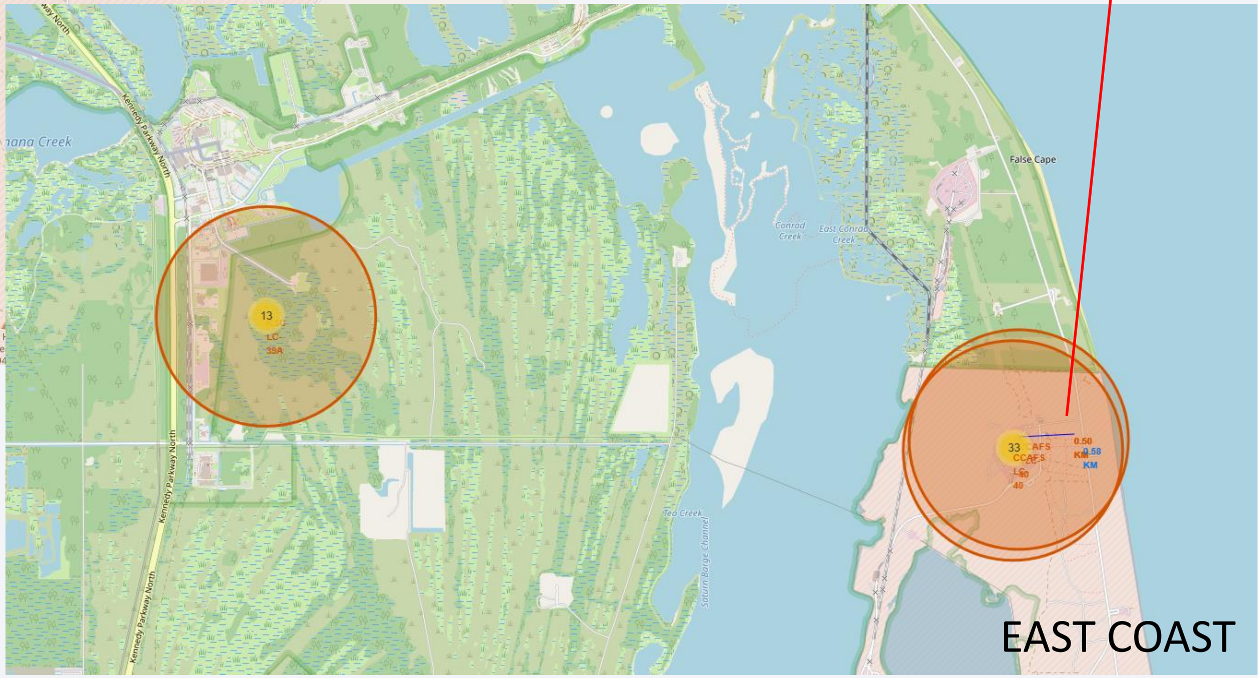
CCAFS LC-40



- ✓ The launch site with the highest success rate is KSC LC-39-A.
- ✓ The launch site with the highest failure rate is CCAFS LC-40.

Launch sites strategic location



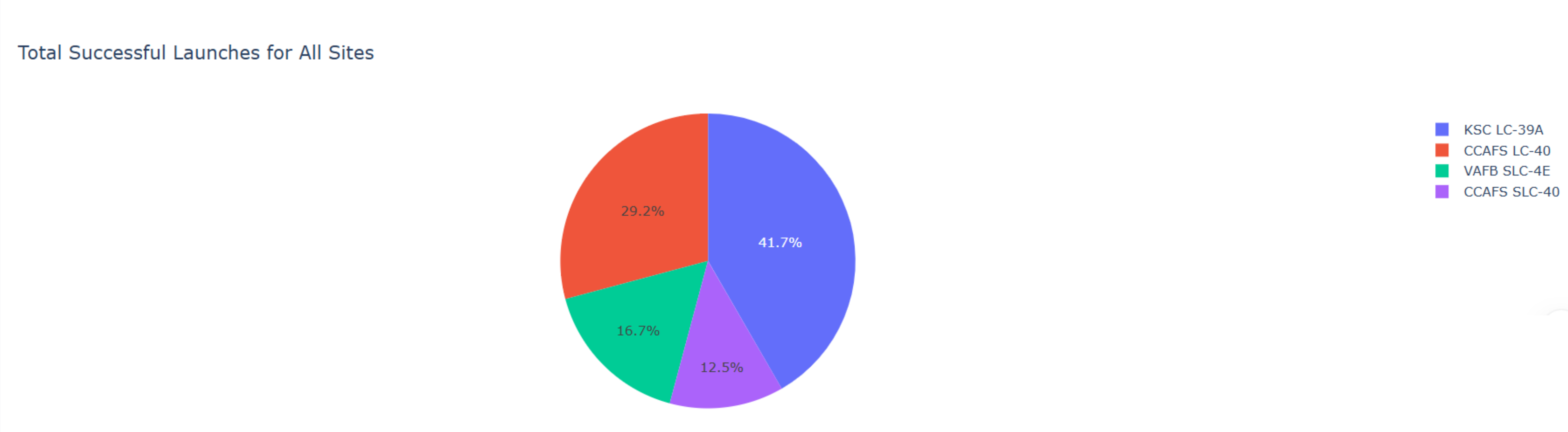


Launch sites are near railways and highways while maintaining a certain distance from urban areas.

Section 4

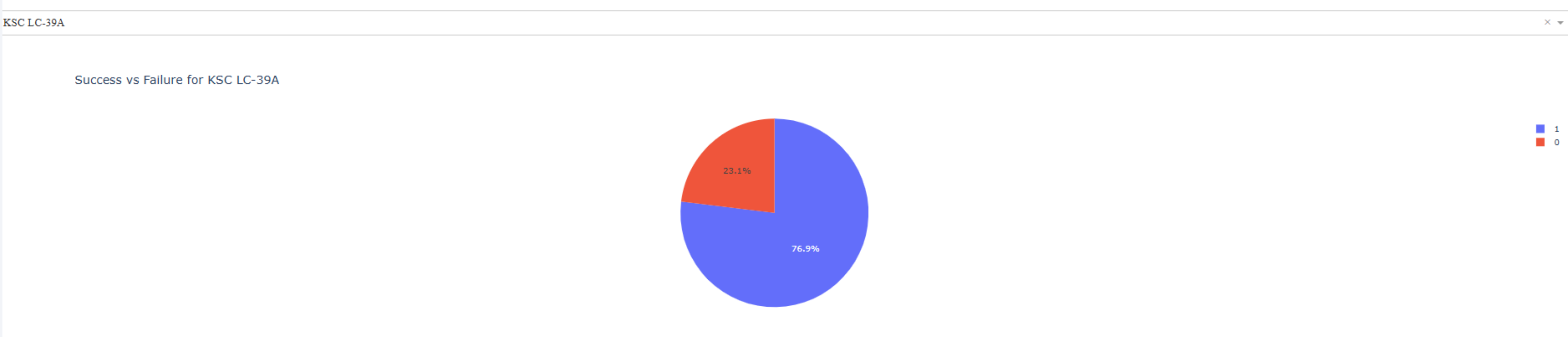
Build a Dashboard with Plotly Dash

Total successful launches for all sites



KSC LC-39A is the launch site with the highest success rate of launches

Launch site with highest success ratio

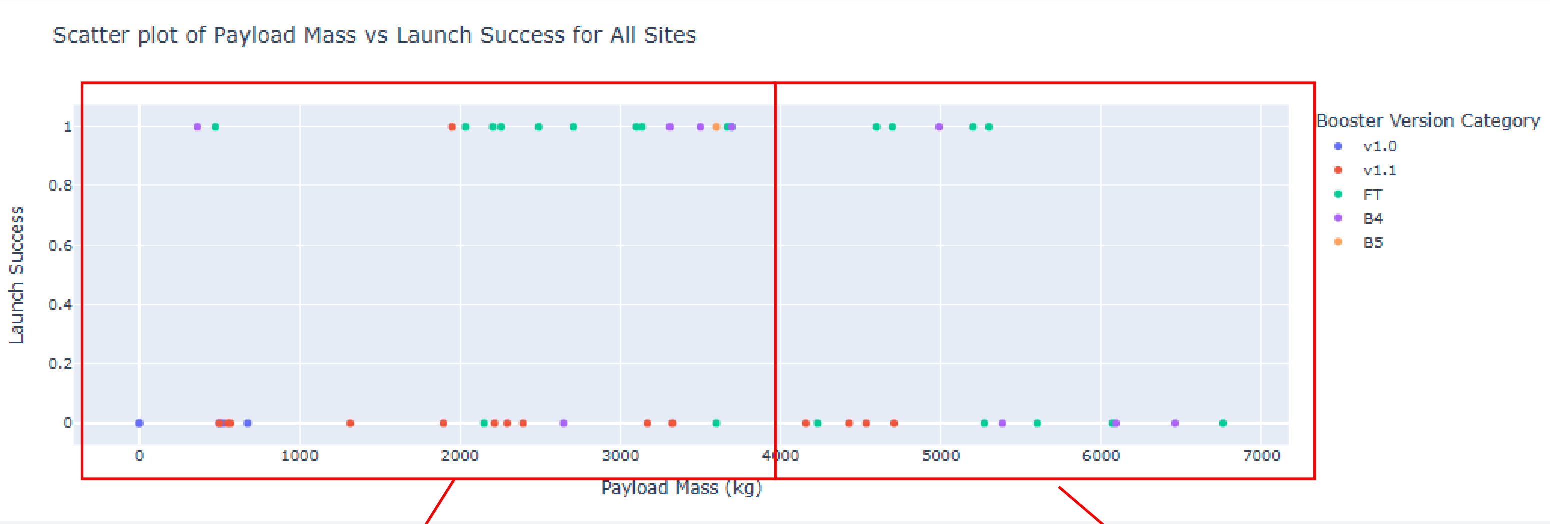


KSC LC-39A lauch site

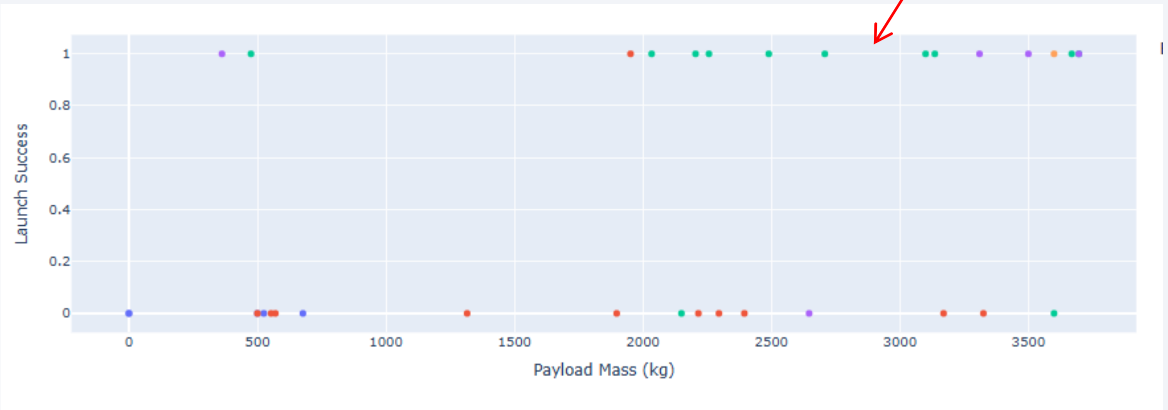
Success rate: 76.9 %

Failure rate: 23.1 %

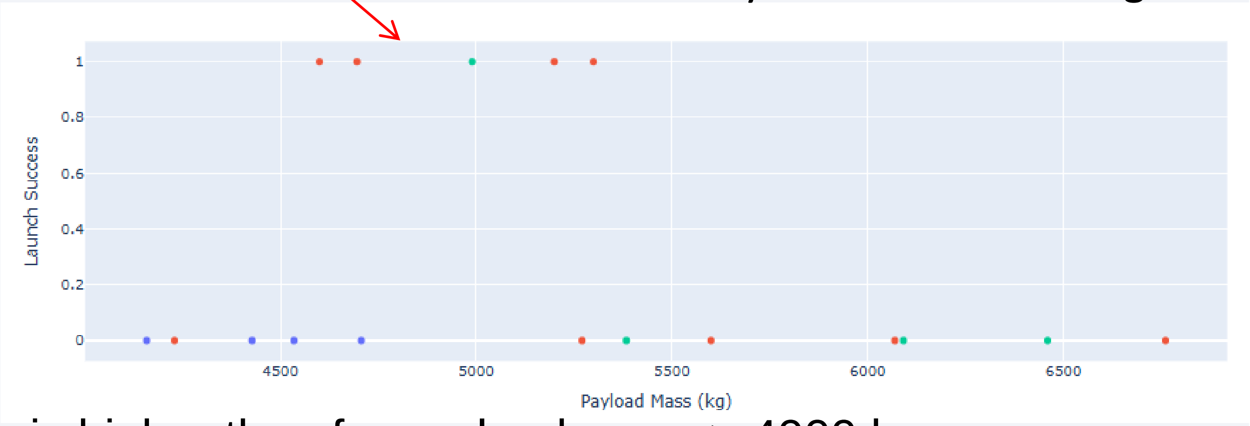
Payload Mass vs Launch Success



Payload mass <= 4000kg



Payload mass > 4000kg



- ✓ Success rate for payload mass < = 4000kg is higher than for payload mass > 4000 kg
- ✓ The payload range with highest succes rate is 3000-4000 kg
- ✓ The payload range with lowest success rate is 0-1000 kg
- ✓ FT is the booster version with highest success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

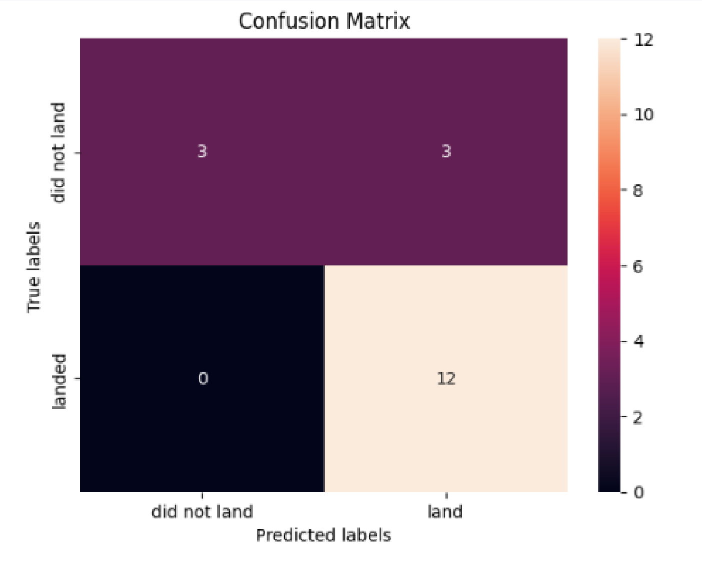
Machine learning models to predict whether the first stage of SpaceX's Falcon 9 rocket will successfully land.

Key metric: accuracy on the test dataset.

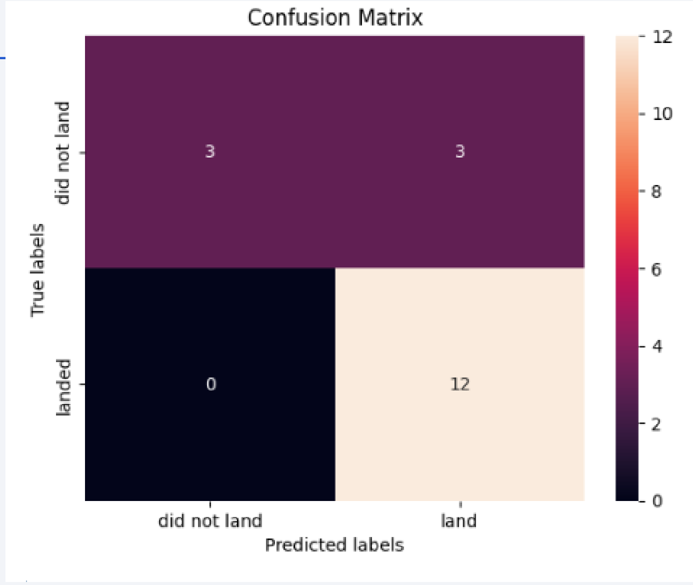
MODEL	ACCURACY ON TEST DATA	OBSERVATIONS
Logistic Regression	0.8333333333333334	Best at avoiding false positives.
Support Vector Machine (SVM)	0.8333333333333334	Requires further tuning to improve performance.
Decision Tree	0.6111111111111112	Simple but less robust.
K-Nearest Neighbors (KNN)	0.8333333333333334	Sensitive to training data.

Confusion Matrix

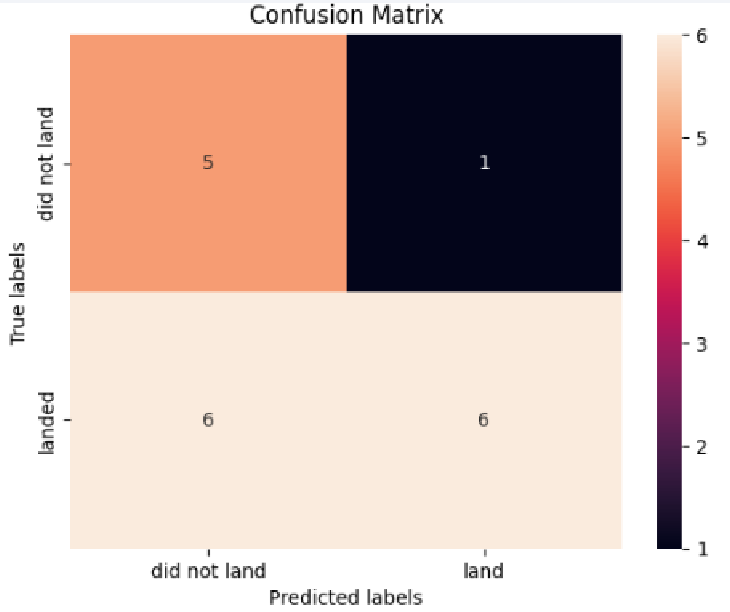
Logistic Regression



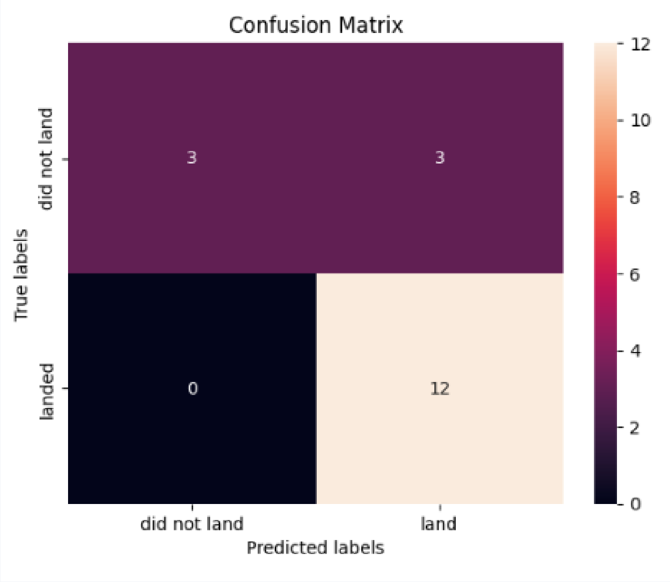
Support Vector Machine (SVM)



Decision Tree



K-Nearest Neighbors (KNN)



Conclusions

- ✓ The Decision Tree model is discarded due to its lower accuracy on the test dataset.
- ✓ The Logistic Regression, SVM, and KNN models exhibit the same accuracy, as shown also in the confusion matrices, where the same number of true positives (12) and true negatives (3) are generated.
- ✓ Given the observations for each model, the Logistic Regression model is recommended due to its simplicity, high accuracy, and ability to distinguish between classes effectively.

Appendix

- Code repository including all code files at:

https://github.com/abrob-hub/MyCapstone

Thank you!