IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ibrahim George Foday
6th August 2023

Outline

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

Executive Summary

Methodology Overview

- This research focuses on discerning the key factors contributing to a successful rocket landing. To achieve this goal, a combination of methodologies has been employed:

Data Collection:

Gathering data through the SpaceX REST API and employing web scraping techniques.

Data Wrangling:

Organizing and refining the collected data to create an outcome variable differentiating between successful and unsuccessful landings.

Data Exploration:

Utilizing data visualization methods to examine specific factors, including payload, launch site, flight number, and yearly trends.

Data Analysis:

Employing SQL for thorough data analysis, involving calculations such as total payload, payload range for successful launches, and the total count of both successful and failed outcomes.

Launch Site Assessment:

Investigating launch site success rates and their proximity to geographical markers.

Visualization of Successful Launches:

Creating visualizations that highlight launch sites with the highest success rates and successful payload ranges.

Predictive Model Development:

Constructing predictive models, including logistic regression, support vector machine (SVM), decision tree, and K nearest neighbor (KNN), to forecast landing outcomes.

Executive Summary

Results

Exploratory Data Analysis:

- Over time, there is an observable enhancement in launch success rates.

- The launch site designated as KSC LC 39A demonstrates the most elevated success rate among all landing sites.

- Certain orbits, such as ES L1, GEO, HEO, and SSO, exhibit a flawless 100% success rate.

Executive Summary:

Visualization and Analytics:

- The majority of launch sites are positioned close to the equator and adjacent to coastlines.

Predictive Analytics:

- All the developed models exhibited comparable performance on the test set.

- The decision tree model exhibited a slightly superior performance compared to the other models.

Introduction

Project Background and Context

- The commercial space age has opened up opportunities for various companies to enter the space industry and provide affordable space travel solutions.
- SpaceX stands out as one of the most successful companies in the field, and one key advantage SpaceX has over others is the relatively low cost of its rocket launches.
- Space Y aims to compete with SpaceX by providing affordable and reliable space travel

Problems to Find Answers

- Gather Information about SpaceX's Rocket Launch Costs
- Predicting First Stage Reusability
- Determining the Price of Each Launch
- Identify areas where Space Y can improve and innovate

Section 1

Methodology

Methodology

Executive Summary

- **Data Collection:** The data was collected using two methods: the SpaceX REST API and web scraping techniques. The data was filtered to remove any irrelevant or incomplete data, and one-hot encoding was used to convert categorical data into numerical data.
- **Data Wrangling:** The data was wrangled to prepare it for analysis and modeling. This included filtering the data, handling missing values, and applying one-hot encoding.
- **Exploratory Data Analysis:** The data was explored using SQL and data visualization techniques. This helped to identify any trends or patterns in the data.
- **Visualization:** The data was visualized using Folium and Plotly Dash. This helped to make the data more understandable and to identify any trends or patterns.
- **Modeling:** Models were built to predict landing outcomes using classification models. The models were tuned and evaluated to find the best model and parameters.

Data Collection – SpaceX API

- **Utilizing APIs and web scraping techniques to gather SpaceX launch data.**
- **Converting JSON to Data Frame:**

➢ Processing and converting the collected data from JSON format into a structured DataFrame for analysis.

- **Data Wrangling and Cleaning:**

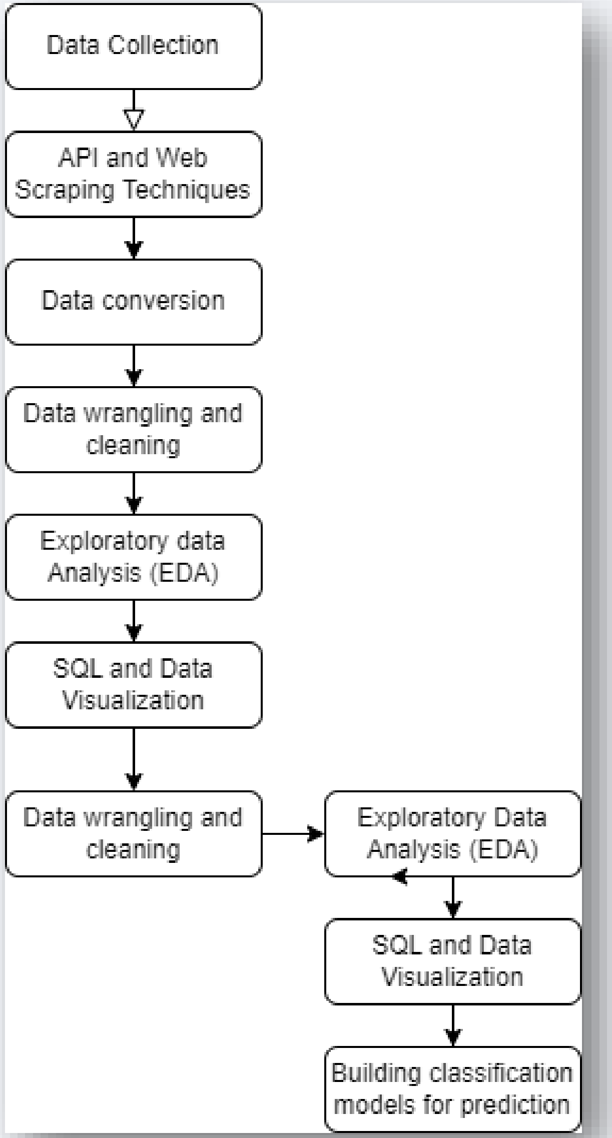
➢ Identifying and handling NULL values to ensure the data is clean and ready for analysis.

- **Exploratory Data Analysis (EDA):**

➢ Conducting in-depth EDA using SQL to extract insights from the data.

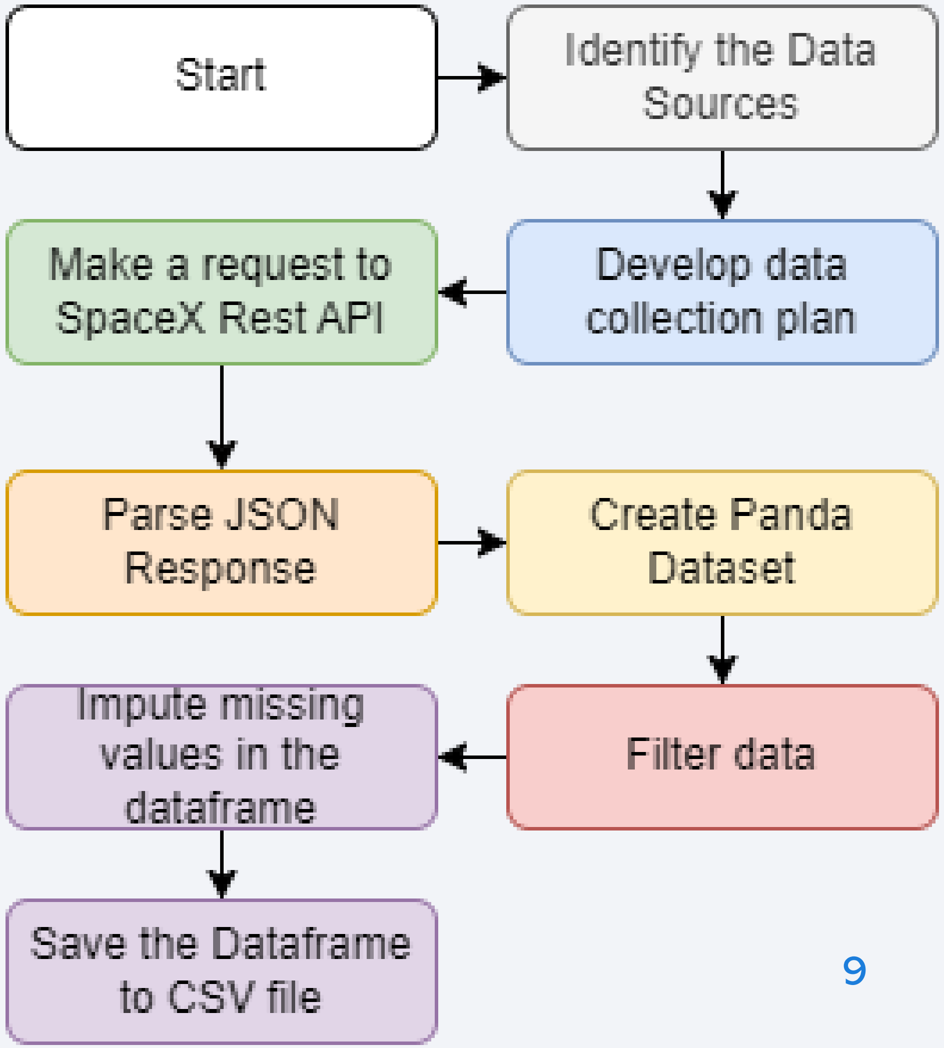
- **Model Building for Predicting Landing Outcomes:**

➢ Constructing classification models to predict landing outcomes based on various features.



Data Collection – SpaceX API

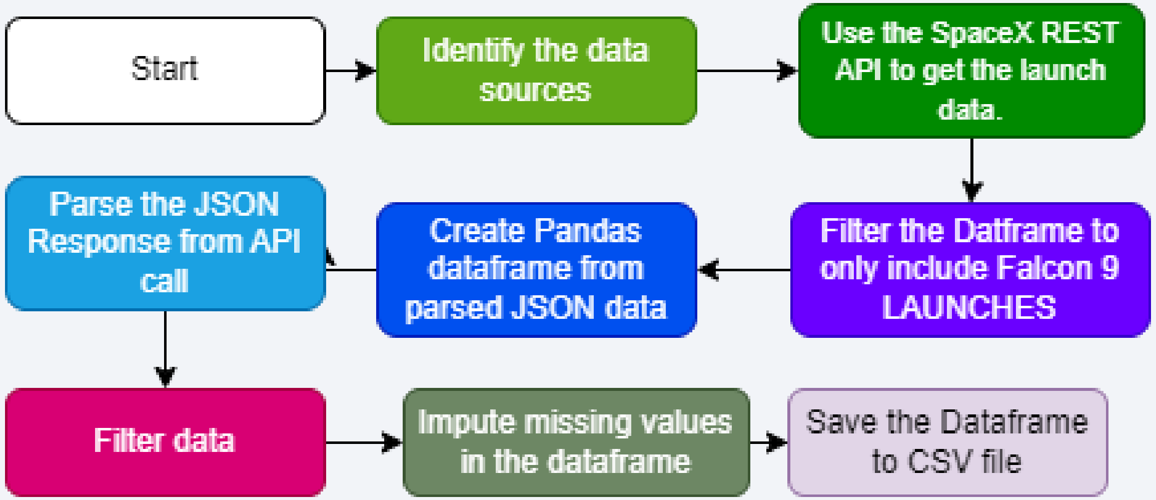
- **Identify the data sources** SpaceX REST API.
- **Develop a data collection plan.** The data collection plan should include the following steps:
- Make a **REQUEST** to the SpaceX REST API to get the launch data.
- Parse the **JSON** response from the API call.
- Create a **PANDAS DATAFRAME** from the parsed JSON data.
- **FILTER** the DataFrame
- **IMPUTE** missing values in the DataFrame.
- Save the DataFrame to a CSV file.



https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.1.SpaceX_Data_Collection_API.ipynb

Data Collection – Web scraping

- ❑ **Identify the data sources.** Identify the SpaceX REST API that contain the data to be scraped.
- ❑ Determine the data to be collected.
- ❑ Make **REQUESTS** to the URLs or SpaceX REST API.
- ❑ **STORE** response data and store.
- ❑ **PARSE** the JSON response.
- ❑ Create a **PANDAS DATAFRAME**.
- ❑ Filter the DataFrame to only include Falcon 9 launches.
- ❑ Impute missing values in the DataFrame.
- ❑ Save the DataFrame to a CSV file.



- https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.2_SpaceX_Web_Scraping.ipynb

Data Wrangling

- The data wrangling process can be summarized as follows:
  - Importing of necessary libraries.
  - **Loading dataset.**
  - **Identifying and calculating the percentage of the missing values in each attribute**.
    - Calculate the number of launches on each site.
    - Calculate the number and occurrence of each orbit.
    - Calculate the number and occurrence of mission outcome per orbit type.
  - **Create a landing outcome** label from Outcome column.
  - **Export** to a CSV file.
  - https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.3-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

Landing was not always a success:

TRUE OCEAN: Outcome was successful.

FALSE OCEAN: Outcome was unsuccessful, and the first stage failed to land in a specific region of the ocean.

TRUE RTLS: Outcome was successful, and the first stage was landed successfully on a ground pad.

FALSE RTLS: Outcome was unsuccessful, and the first stage failed to land on a ground pad.

TRUE ASDS: Outcome was successful, and the first stage was landed successfully.

FALSE ASDS: Outcome was unsuccessful, and the first stage failed to land.

None ASDS and None None: These represent a failure to land,

EDA with Data Visualization

The following charts were plotted:

- ➢ Flight Number vs. Payload Mass and overlay the outcome of the launch.
- ➢ Payload vs. Launch Site.
- ➢ A bar chart for the success rate of each orbit type.
- ➢ A scatter point chart with x axis to be Flight Number and y axis to be the Orbit, and hue to be the class value.
- ➢ A scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value.
- ➢ A line chart with x axis to be the extracted year and y axis to be the success rate.

These charts help us to understand the **relationship between different variables** and the **success rate** of the SpaceX Falcon 9 first stage landing.

https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.4%20jupyter-labs-eda-dataviz.ipynb

EDA with SQL

SQL queries performed:

Display Queries

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1

https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.5%20jupyter-labs-eda-sql-coursera_sqllite.ipynb

List Queries

- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing outcomes in drone ship.

Build an Interactive Map with Folium

Map objects created and added to a folium map:

Folium.Circle to add highlighted circle areas with a text label on a specific coordinate.

- Eg: Added a circle on **NASA Johnson Space Center's** coordinate with a popup label showing its name.

Folium.Marker to add markers for each launch site in data frame launch sites.

Additionally added a popup label to each marker showing the launch site name.

MarkerCluster to simplify a map containing many markers having the same coordinate.

- Eg: Added a MarkerCluster to the **sitemap** and added all the markers in **spaceX df data frame to the MarkerCluster.**

MousePosition to get the coordinate (Lat, Long) for a mouse over a point on the map.

These objects were added to the folium map to visualize the launch sites and their proximities.

https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.6%20lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Dropdown list with launch sites

- Allows the user to select all launch sites or a specific launch site.

Slider of payload mass range

- Allows the user to select a range of payload masses.

Pie chart showing successful launches

- Shows the percentage of successful launches out of the total number of launches.

Scatter chart showing payload mass vs. success rate by booster version

- Shows the relationship between payload mass and launch success rate for each booster version.

Adding these plots and interactions to the dashboard to makes it more interactive and informative.

https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/1.7_SpaceX_Interactive_Visual_Analytics_Plotly.py

Predictive Analysis (Classification)

- Imported the necessary libraries and defined some auxiliary functions.
- Loaded the data from two files into two **DataFrames,** data and X.
- Created a **NumPy** array from the column Class in data.
- Standardized the data in X using the **preprocessing.StandardScaler()** function.
- Split the data into training and test data using the **train_test_split()** function.
- Created three classification models: **logistic regression**, **support vector machine**, and **decision tree**.
- Used a **GridSearchCV** object to find the best hyperparameters.
- Evaluated the accuracy of each model on the test data.
- Plotted the confusion matrices for all three models to visualize the results.

- https://github.com/ibrahimgeorgefoday/IBM-APPLIED-DATA-SCIENCE-FINAL-CAPSTONE-PROJECT./blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_1.8%20SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Section 2

Insights drawn from EDA

Results Summary

- After conducting Exploratory Data Analysis (EDA) on the launch data, several key insights were discovered:
- **Launch Success Improvement**
- Over time, the launch success rate has shown improvement, indicating advancements in space technology and operational efficiency.
- **Highest Success Rate Landing Site**
- **Kennedy Space Center Launch Complex 39A** has the highest success rate among all the landing sites, suggesting that it is a well-established and reliable location for space launches.
- **Successful Orbits**
- Certain orbital destinations, such as ES-L1, GEO (Geostationary Orbit), HEO (Highly Elliptical Orbit), and SSO (Sun-Synchronous Orbit), have achieved a 100% success rate, indicating their successful track record for missions to those specific orbits.

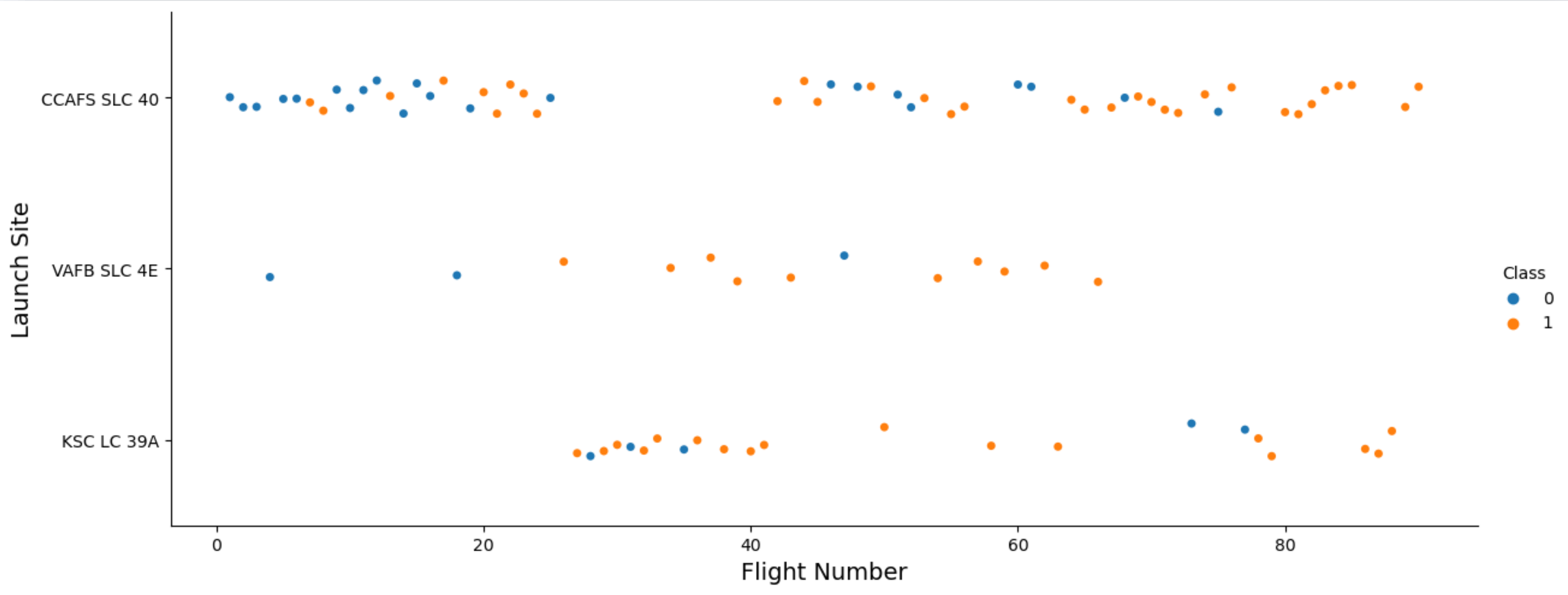
Visual Analytics

- **Geographical Locations:** Most launch sites are situated near the equator,
- **Safety Measures:** The launch sites have been strategically located far enough away to minimize the potential damage in case of a failed launch. Results Summary for Predictive Analytics:
- **Best Predictive Model:** The Decision Tree model was identified as the most effective predictive model for the dataset.

Flight Number vs. Launch Site

Exploratory Data Analysis reveals:

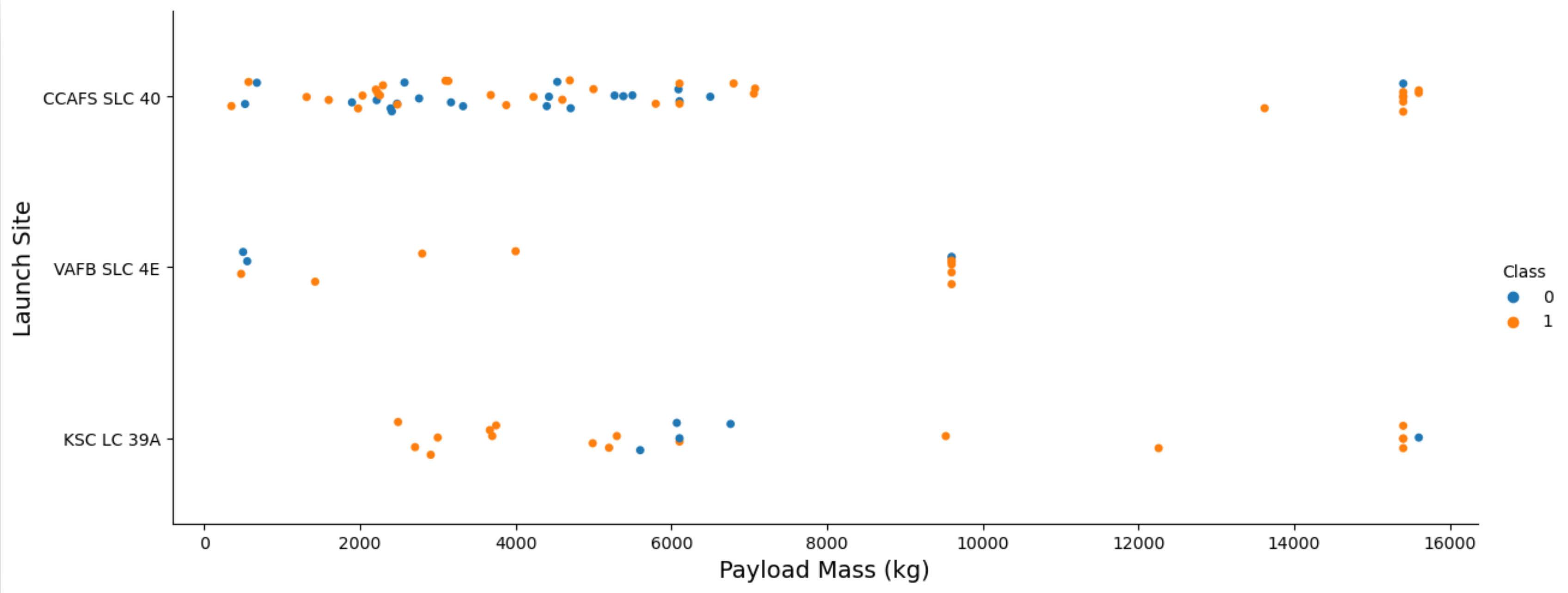
- Earlier flights exhibited a lower success rate, indicated by the **blue** color denoting "fail."
- Later flights showed a higher success rate, represented by the orange color denoting "success."
- Approximately half of the launches originated from the **CCAFS SLC 40** launch site.
- **VAFB SLC 4E** and **KSC LC 39A** demonstrated higher success rates.
- It can be inferred that new launches tend to have a higher success rate.



Payload vs. Launch Site

Exploratory Data Analysis:

- Success rate is positively correlated with payload mass (kg).
- Launches with payload greater than 7,000 kg tended to be successful.
- **KSC LC 39A** achieved a 100% success rate for launches with payload less than 5,500 kg.
- **VAFB SKC 4E** has not conducted any launches with a payload greater than approximately 10,000 kg.

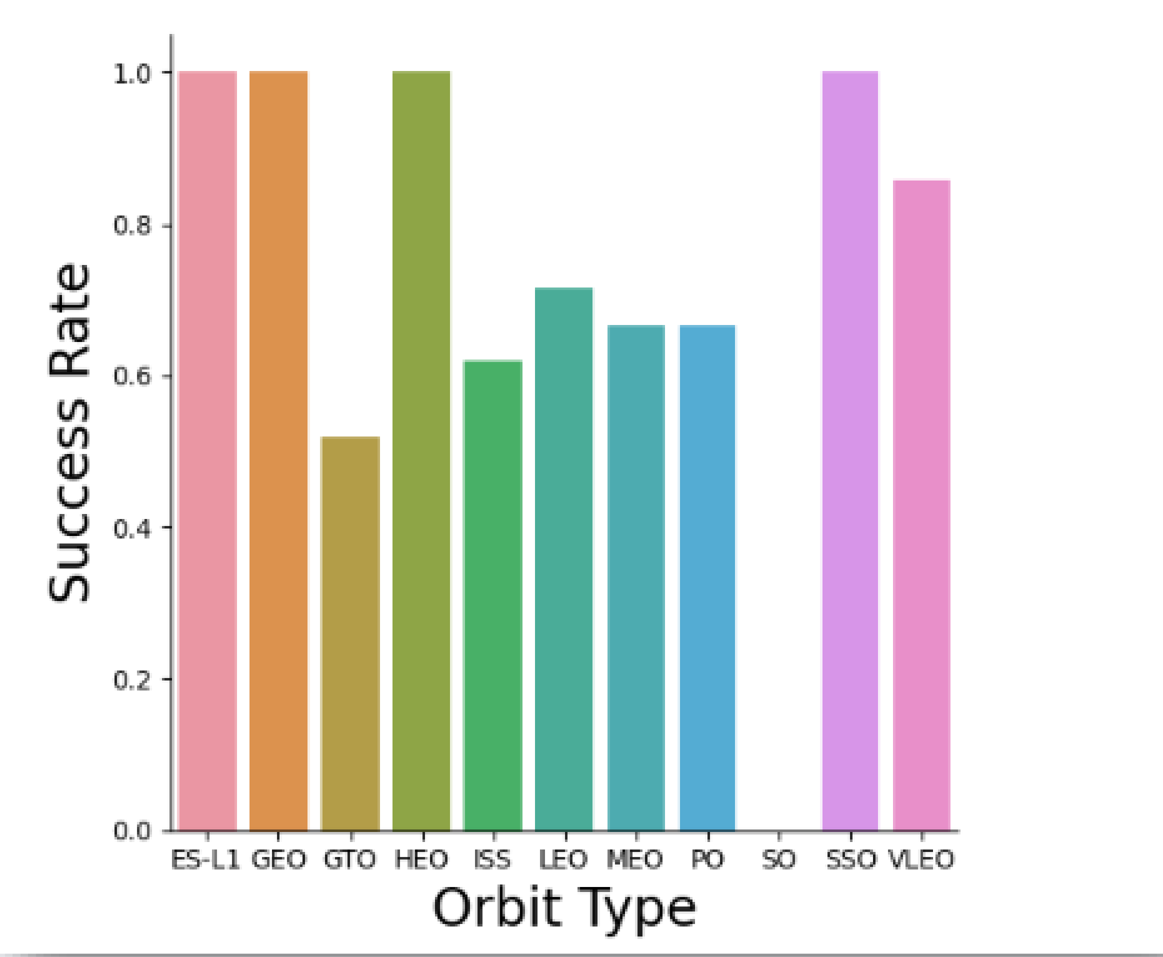


Class 0 = Fail
Class1=Success

Success Rate vs. Orbit Type

Exploratory Data Analysis

- 100% Success Rate: ES-L1, GEO, HEO and SSO
- 50%-80% Success Rate: GTO, ISS, LEO, MEO, PO
- 0% Success Rate: SO

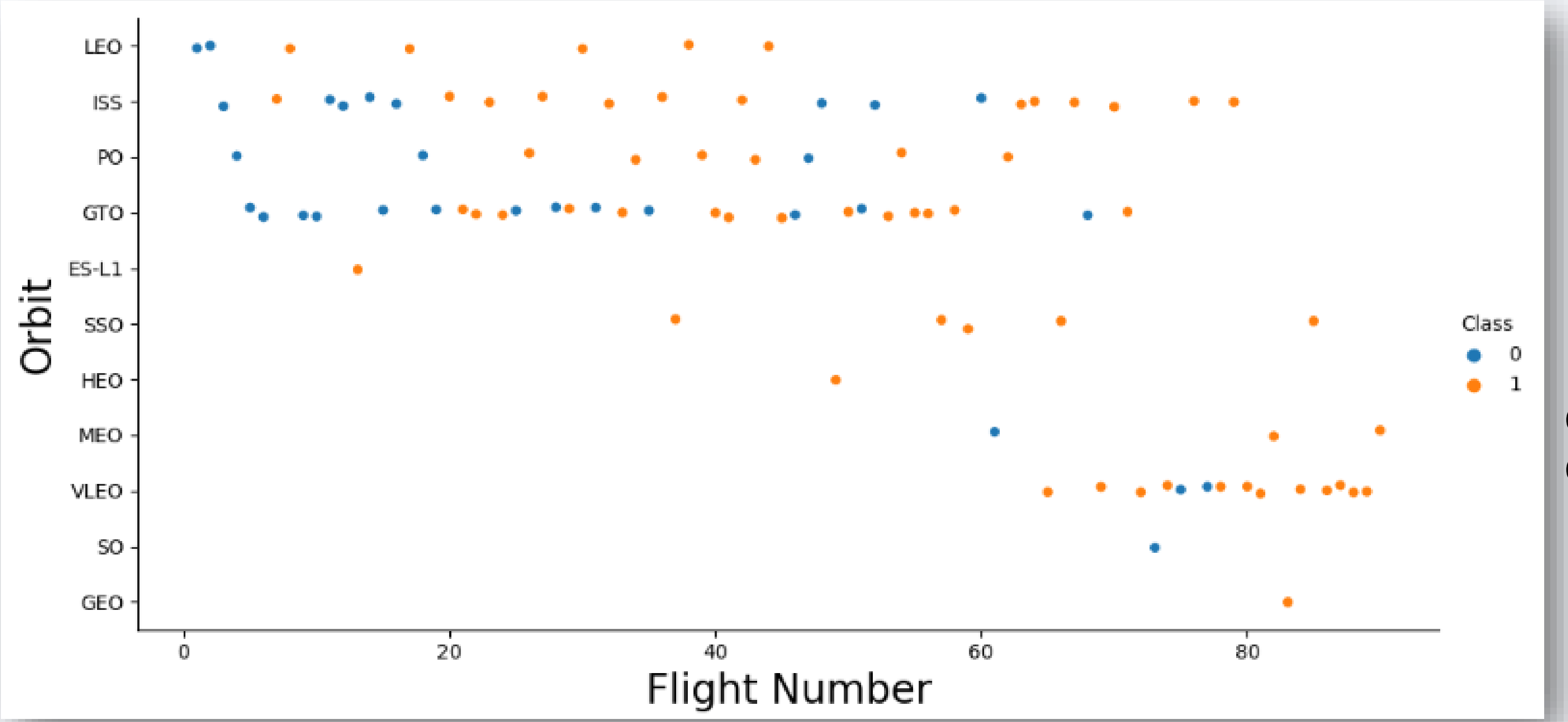


Flight Number vs. Orbit Type

✓Success rate generally rises with the number of flights per orbit.

✓This trend is particularly noticeable for the LEO (Low Earth Orbit).

✓However, the GTO (Geostationary Transfer Orbit) does not exhibit the same pattern.

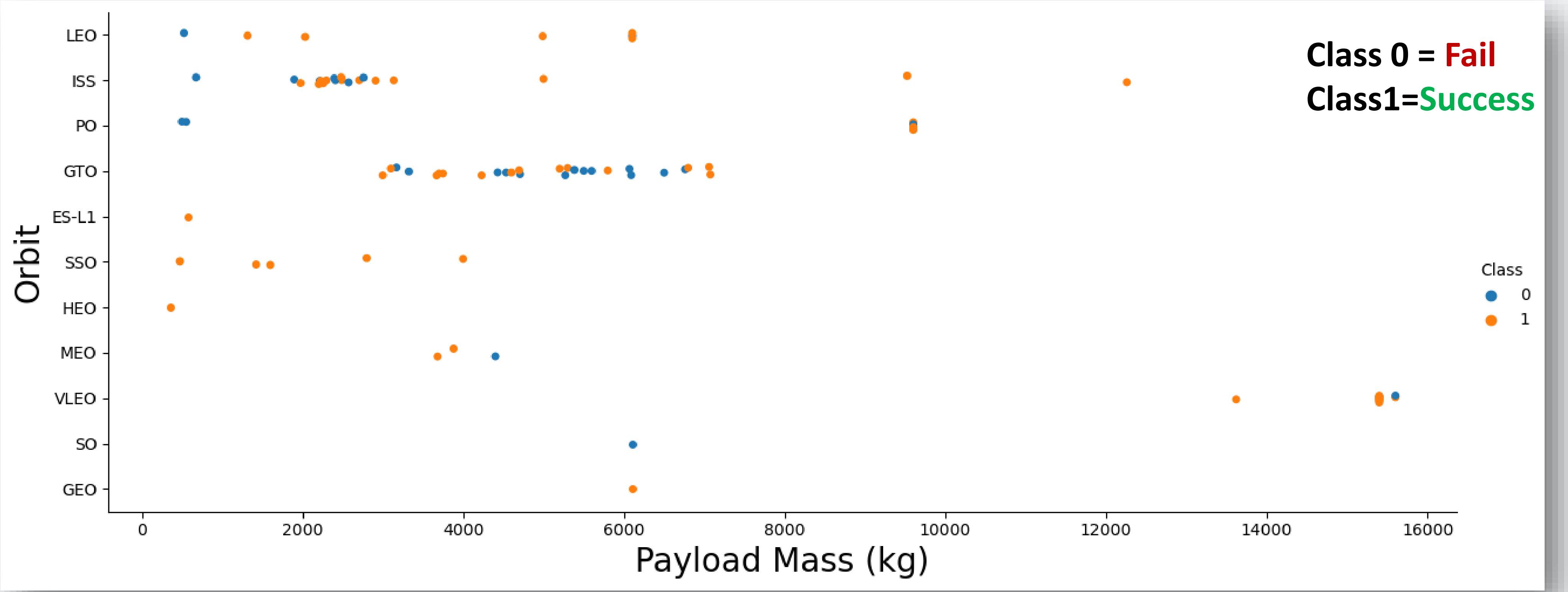


Class 0 = Fail
Class1=Success

Payload vs. Orbit Type

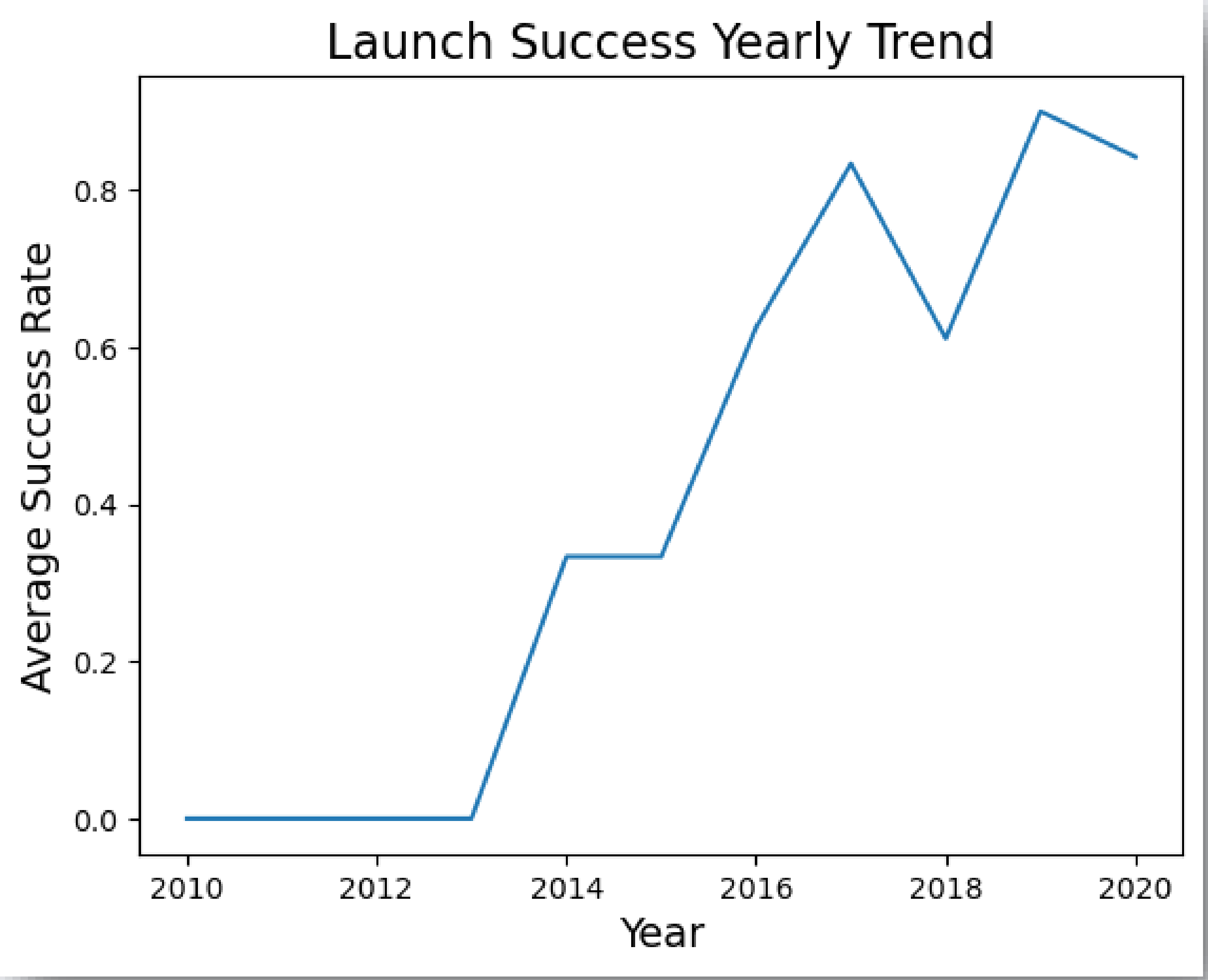
Exploratory Data Analysis:

- **LEO (Low Earth Orbit), ISS (International Space Station),** and **PO (Polar Orbit)** show higher success rates for heavy payloads.
- **GTO (Geostationary Transfer Orbit)** exhibits mixed success when dealing with heavier payloads.



Launch Success Yearly Trend

- Success rate trend:
  - Improved from **2013** to **2017** and **2018** to **2019.**
  - Decreased from **2017** to **2018** and from **2019** to **2020.**
- Overall, the success rate has shown improvement since **2013.**



All Launch Site Names

Launch Site Names

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

Landing Outcome.

```
[36]: %sql SELECT DISTINCT "LAUNCH_SITE" FROM SPACEXTABLE;

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

[36]:

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

Launch Site Names Begin with 'CCA'

- The results provide specific information about successful launches from the launch site codes starting with 'CCA.'

```
]: %sql SELECT * \
     FROM SPACEXTBL \
     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
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- All the boosters launched by NASA (CRS) combined carried a total of 45596 kilograms of payload mass.

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) \
    FROM SPACEXTBL \
    WHERE CUSTOMER = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by the booster version "F9 v1.1" is approximately 2928.4 kilograms.

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

```
]: %sql SELECT AVG(PAYLOAD_MASS__KG_) \
       FROM SPACEXTBL \
       WHERE BOOSTER_VERSION = 'F9 v1.1';
```

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

]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

December 22nd 2015

```
%sql SELECT MIN(DATE) \
FROM SPACEXTBL \
WHERE LANDING__OUTCOME = 'Success (ground pad)'
```

```
 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b
   sqlite:///my_data1.db
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The booster mass is between **4,000 and 6,000.**

- Missions with this booster mass range are
  - JSCAT 14,
  - JSCAT 16,
  - SES 10, and
  - SES 11 / EchoStar 105.

```
%sql SELECT PAYLOAD \
FROM SPACEXTBL \
WHERE LANDING__OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

```
 * ibm_db_sa://yyy33800:***@1bbf73c5-d84a-4bb0-85b9-
   sqlite:///my_data1.db
Done.
```

payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

- 1 Failure in Flight
- 99 Success
- 1 Success (payload status unclear)

```
[51]: %sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY MISSION_OUTCOME;

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

[51]:

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

Boosters Carried Maximum Payload

Booster version Carrying Max Payload

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

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

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

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

- In the year 2015, the following information is displayed, including the month, date, booster version, launch site, and landing outcome.

```
%sql SELECT substr(Date,4,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [Landing _Outcome] \
FROM SPACEXTBL \
where [Landing _Outcome] = 'Failure (drone ship)' and substr(Date,7,4)='2015';
```

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

month	Date	Booster_Version	Launch_Site	Landing _Outcome
01	10-01-2015	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	14-04-2015	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- **Ranked in Descending Order**

Count of landing outcomes between 2010-06-04 and 2017-03-20 in descending order

```
%sql SELECT [Landing _Outcome], count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '04-06-2010' and '20-03-2017' group by [Landing _Outcome] order by count_outcomes DESC;
```

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

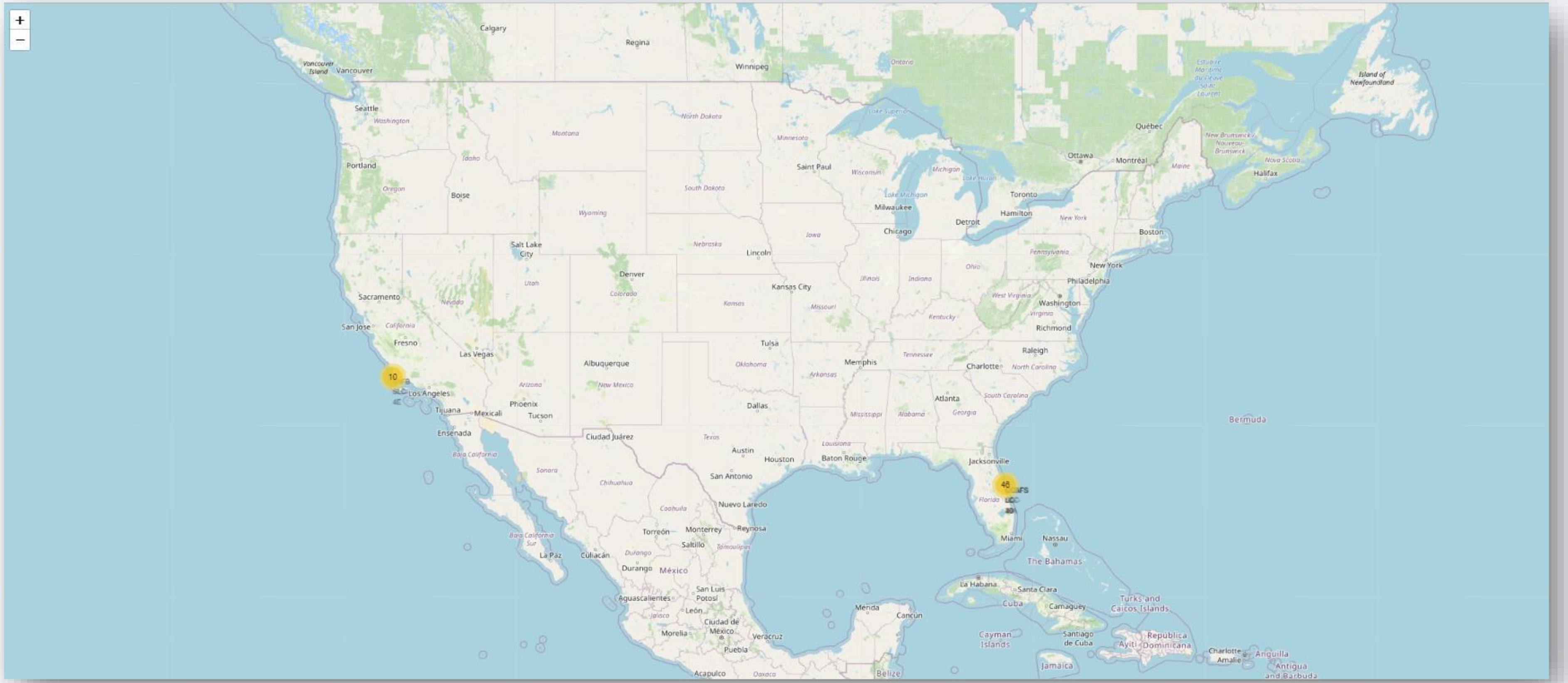
Landing _Outcome	count_outcomes
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

Section 3

Launch Sites Proximities Analysis

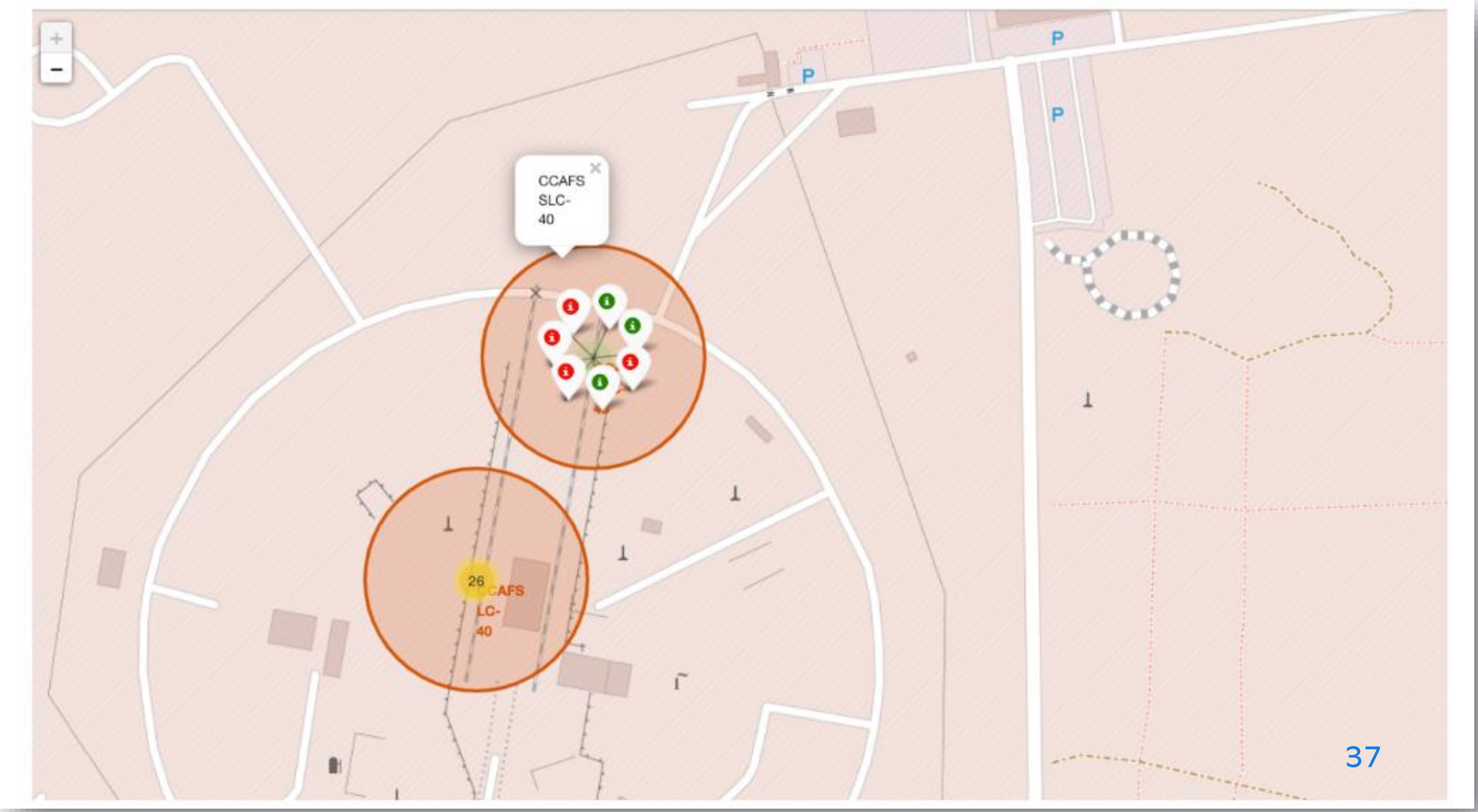
Launch Sites

Near Equator:
- Launch sites closer to the equator benefit from Earth's rotation during a prograde orbit.
- Being close to the equator makes it easier to launch payloads into equatorial orbits.
- Rockets launched from equatorial sites receive a natural boost from Earth's rotational speed.
- This natural boost helps reduce the need for additional fuel and boosters, resulting in cost savings.



Launch Outcomes

- **At Each Launch Site**
- **Outcomes**:
- Greenmarkers for successful launches
- Redmarkers for unsuccessful launches
- Launch site **CCAFS SLC-40** has a **3/7 success rate** (42.9%)



Distance Proximities

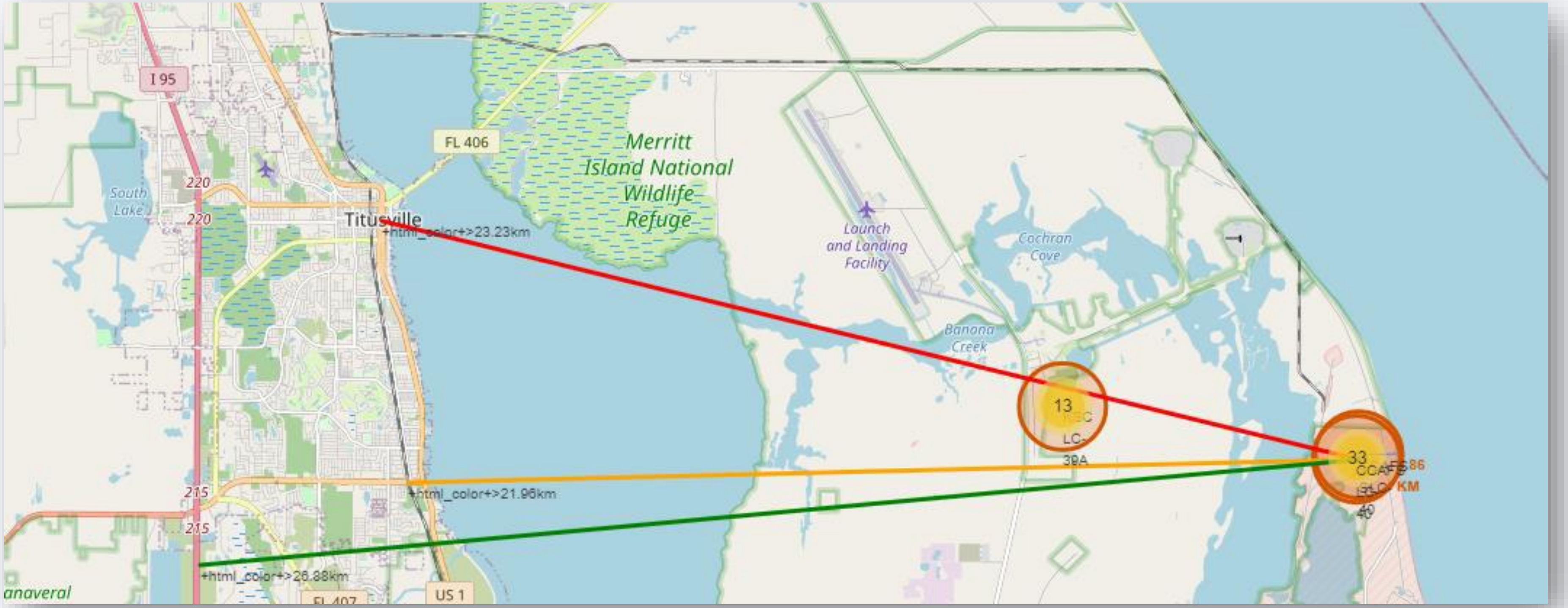
- The distance to proximities from CCAFS SLC 40 is as follows:

✓ **.86** km from the nearest coastline

✓ **21.96** km from the nearest railway

✓ **23.23** km from the nearest city

✓ **26.88** km from the nearest highway

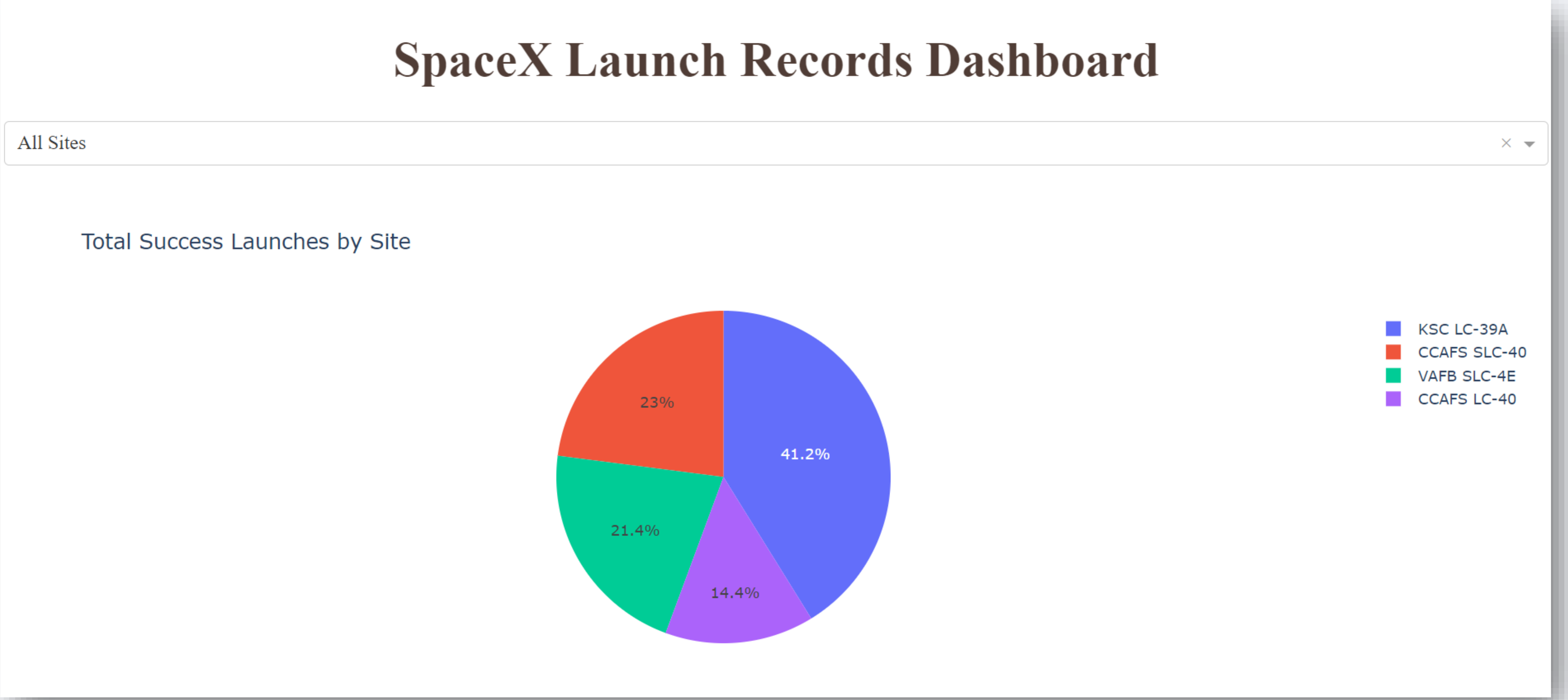


Section 4

Build a Dashboard with Plotly Dash

Launch Success By Site

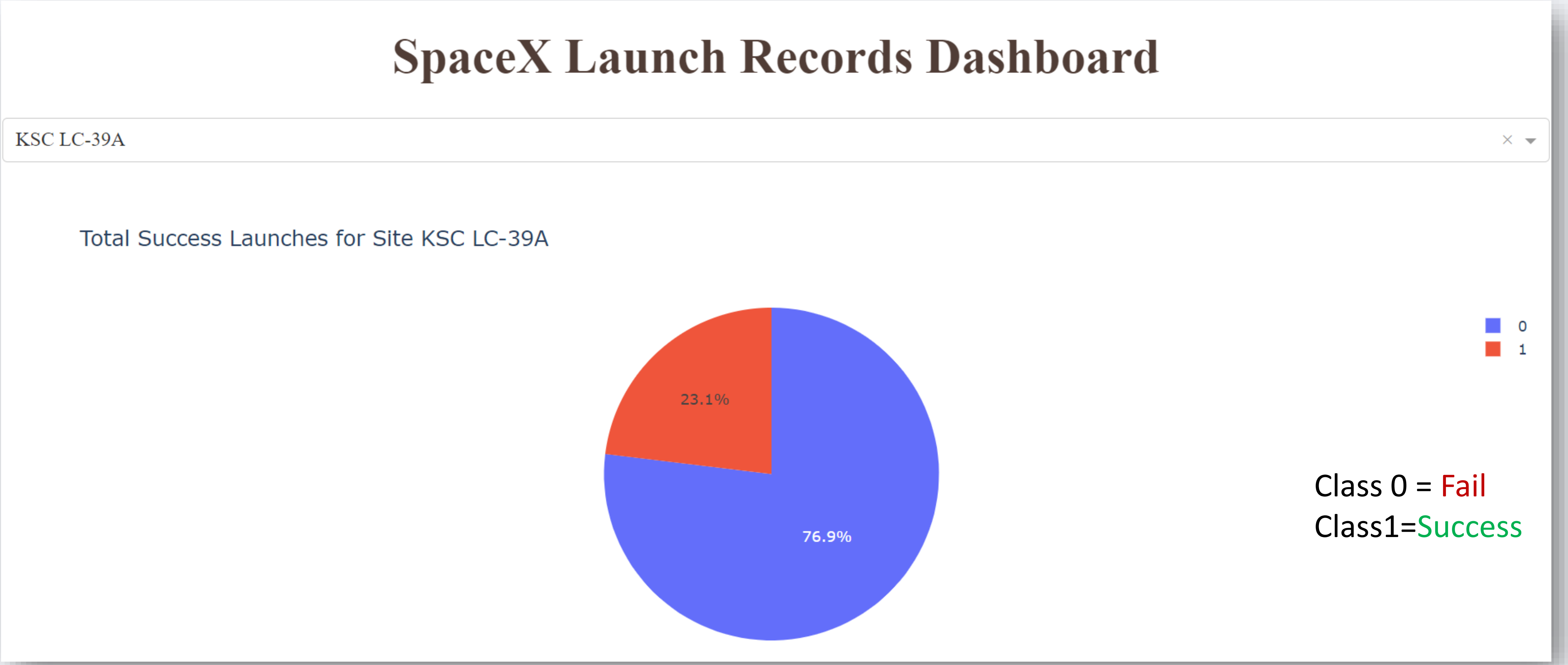
- The percentage of successful launches, relative to the total number of launches, is highest at KSC LC 39A, with a rate of 41.2%.



Launch Success (KSC LC 39A)

Among all the launch sites, KSC LC 39A has the highest success rate, which is approximately 76.9%.

It achieved 10 successful launches and experienced 3 failed launches.



Payload Mass and Success

By Booster Version:

- Payloads with masses between 2,000 kg and 5,000 kg show the highest success rate.

- In the context of outcomes, a "1" represents a successful outcome, while a "0" indicates an unsuccessful outcome.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Accuracy:

All the models exhibited **similar performance** with **identical scores** and accuracy. This similarity is likely attributed to the small dataset used for training.

Among the models, the **Decision Tree model** slightly **outperformed** the others when considering the **`.best_` attribute**.

The **`.best_score_`** attribute represents the average performance across all cross-validation folds for a specific combination of the model's parameters.

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.833333	0.833333

```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgorithm = max(models, key=models.get)
print('Best model is', bestalgorithm,'with a score of', models[bestalgorithm])
if bestalgorithm == 'DecisionTree':
    print('Best params is :', tree_cv.best_params_)
if bestalgorithm == 'KNeighbors':
    print('Best params is :', knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best params is :', logreg_cv.best_params_)
if bestalgorithm == 'SupportVector':
    print('Best params is :', svm_cv.best_params_)
```

```
Best model is DecisionTree with a score of 0.9017857142857142
Best params is : {'criterion': 'gini', 'max_depth': 16, 'max_features': 'auto', 'min_samples_leaf': 4, 'min_samples_split': 10, 'splitter': 'random'}
```

Confusion Matrix

Summary of Performance Metrics:

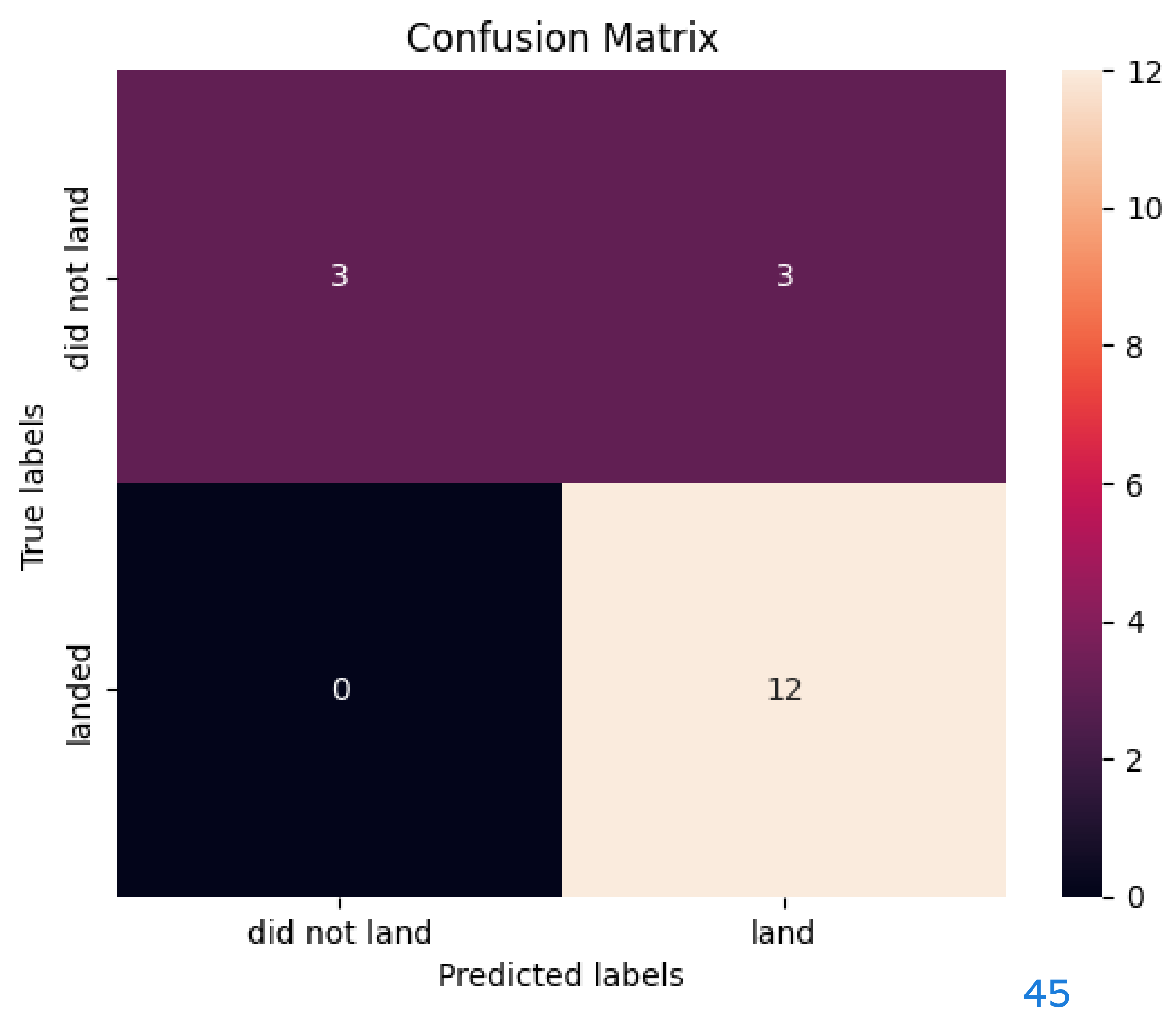
- ➢ The performance of a classification algorithm can be evaluated using a confusion matrix, which summarizes its predictions.
- ➢ In this case, all the confusion matrices were identical. However, the presence of false positives (Type 1 errors) is undesirable in the classification results.
- ➢ The confusion matrix outputs for the evaluation are as follows:

- **True Positives (TP): 12**
- **True Negatives (TN): 3**
- **False Positives (FP): 3**
- **False Negatives (FN): 0**

- ➢ Various performance metrics can be derived from the confusion matrix:

- ✓ Precision: TP / (TP + FP) = 12 / 15 ≈ **0.80**
- ✓ Recall (also known as Sensitivity or True Positive Rate): TP / (TP + FN) = 12 / 12 = **1**
- ✓ F1 Score: 2 * (Precision * Recall) / (Precision + Recall) = 2 * (0.8 * 1) / (0.8 + 1) ≈ **0.89**
- ✓ Accuracy: (TP + TN) / (TP + TN + FP + FN) = (12 + 3) / (12 + 3 + 3 + 0) ≈ **0.833**

These metrics help in assessing the algorithm's performance in terms of its predictive accuracy, precision, recall, and overall effectiveness in handling both positive and negative cases.



Conclusions

Findings:

- **Model Performance:** The models tested showed similar performance on the test set, with the decision tree model slightly outperforming the others.
- **Equator:** Launch sites are strategically located near the equator to take advantage of the Earth's rotational speed, which provides a natural boost and reduces the need for additional fuel and boosters, thus saving costs.
- **Coast:** All launch sites are situated close to the coast.
- **Launch Success:** The overall launch success rate has shown an increasing trend over time.
- **KSC LC 39A:** Among all launch sites, KSC LC 39A has the highest success rate. Specifically, it achieves a 100% success rate for launches with payloads less than 5,500 kg.
- **Orbits:** Launches to orbits such as ES L1, GEO, HEO, and SSO have a 100% success rate.
- **Payload Mass:** Across all launch sites, there is a positive correlation between the payload mass (in kg) and the success rate; higher payload masses tend to result in higher success rates.

Conclusions

Recommendation

Based on the findings, here are some recommendations for action:

Decision Tree Model Implementation: Given that the decision tree model slightly outperformed the others in terms of performance, it's recommended to consider implementing the decision tree model for future predictions or decision-making processes related to space missions.

Strategic Launch Site Selection: The knowledge that launch sites near the equator offer natural advantages due to Earth's rotational speed should guide future launch site selection.

Emphasize Coastal Launch Sites: Given that all launch sites are situated close to the coast, this trend should be continued.

Recommendation for Future Research:

Here are some recommendations for future research:

Factors Affecting Launch Success: Conduct an in-depth analysis to identify the specific factors contributing to the increasing trend in launch success rates over time.

Optimal Payload Mass Consideration: Explore the relationship between payload mass and launch success rates in more detail. Investigate the upper limits of payload mass beyond which success rates might start to decline, as this could provide insights into optimizing payload planning.

Performance of Different Orbits: Further investigate the factors that contribute to the 100% success rates for launches to specific orbits such as ES L1, GEO, HEO, and SSO.

Extended Payload Mass Analysis: Expand the analysis of payload mass to delve deeper into the effects of varying payload masses on launch success rates.

Appendix

- Import Libraries and Define Auxiliary Functions
- Python

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

def plotter(df, x, y, hue, title):
    plt.figure(figsize=(10, 6))
    sns.catplot(x=x, y=y, hue=hue, data=df, kind='scatter', aspect=2.5)
    plt.xlabel(x, fontsize=14)
    plt.ylabel(y, fontsize=14)
    plt.title(title, fontsize=16)
    plt.show()
```

- Exploratory Data Analysis
- Python

```
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate
df['Year'] = df['Date'].apply(lambda date: int(date.split('-')[0]))
launch_success_yearly_trend = df.groupby('Year')['Class'].mean()
plt.plot(launch_success_yearly_trend.index, launch_success_yearly_trend.values)
plt.xlabel('Year')
plt.ylabel('Average Success Rate')
plt.title('Launch Success Yearly Trend')
plt.show()
```

```
# Read the SpaceX dataset into a Pandas dataframe
df = pd.read_csv('https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv')

# Plot FlightNumber vs. PayloadMass and overlay the outcome of the launch
plotter(df, 'FlightNumber', 'PayloadMass', 'Class', 'FlightNumber vs. PayloadMass')

# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value
plotter(df, 'PayloadMass', 'LaunchSite', 'Class', 'PayloadMass vs. LaunchSite')

# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value
plotter(df, 'FlightNumber', 'Orbit', 'Class', 'FlightNumber vs. Orbit')

# Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
plotter(df, 'PayloadMass', 'Orbit', 'Class', 'PayloadMass vs. Orbit')
```

Thank you!