IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Younes AJEDDIG
18/05/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data was sourced from the SpaceX public API and publicly available information on Wikipedia. The data preparation phase involved extracting details about launch outcomes to be used as the target variable in machine learning models.
  - **Analysis and Visualization**: Employed SQL queries and developed various visualizations including static plots, interactive maps, and a dynamic dashboard to uncover insights and address specific questions about the dataset.
  - **Predictive Modeling**: Conducted predictive analysis using multiple machine learning algorithms, including Logistic Regression, Support Vector Machine (SVM), Decision Tree, and k-Nearest Neighbors (KNN) to forecast launch outcomes.
- Summary of all results
  - **Launch Data**: Includes details such as flight number, launch date, payload mass, orbit type, launch site, mission outcome, and other relevant variables.
  - **Machine Learning Performance**: Logistic Regression, Support Vector Machine (SVM), and k-Nearest Neighbors (KNN) models all showed comparable performance when applied to this dataset.

Introduction

- **Competitor Analysis**: A rival rocket launch company aims to predict the success or failure of SpaceX Falcon 9 rocket first stage landings.
- **Data Scope**: What information is available regarding SpaceX Falcon 9 first stage landings, and how extensive is this dataset?
- **Model Selection**: Which machine learning model demonstrates the highest accuracy for predicting the outcome of a Falcon 9 first stage landing?
- **Prediction Inquiry**: Will an upcoming Falcon 9 first stage landing be successful?

Section 1

Methodology

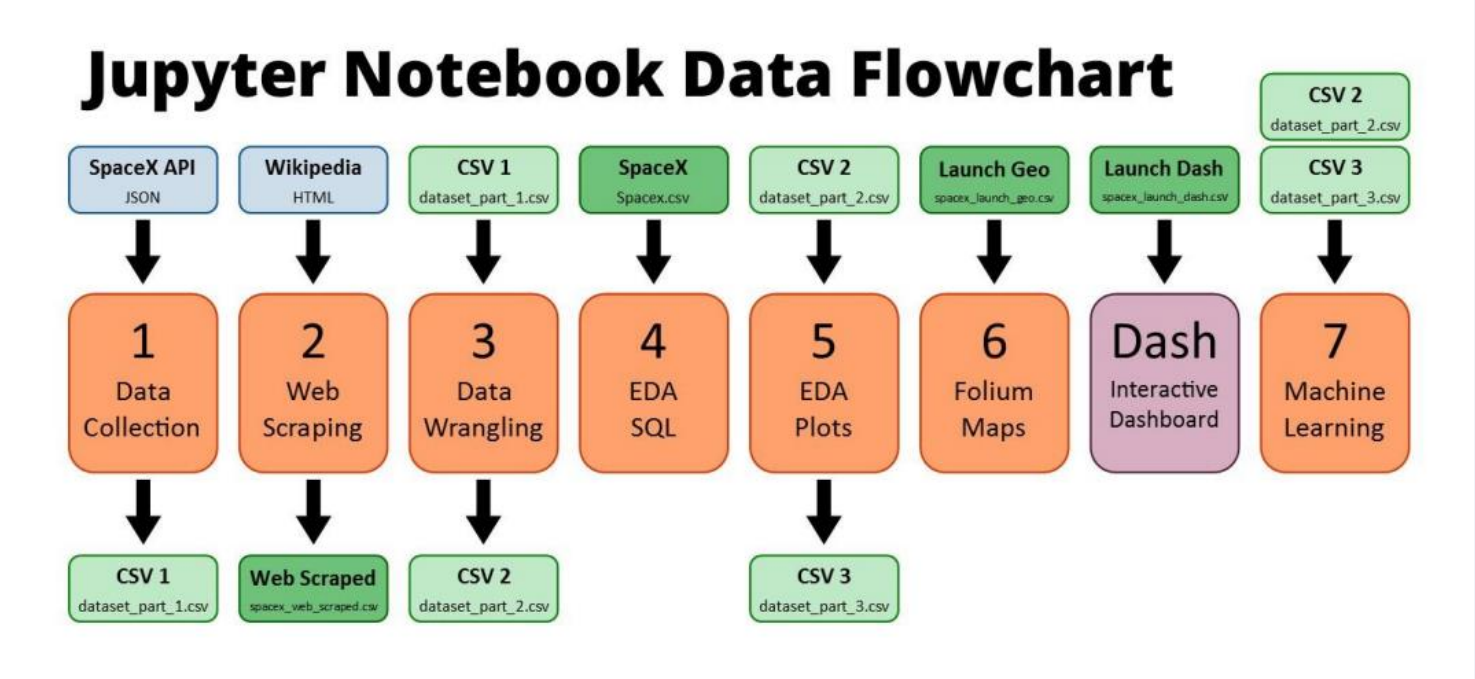
Methodology

Executive Summary

- Data Collection Methodology:
  - Collected data from SpaceX API and Wikipedia launch table
- Perform Data Wrangling:
  - Cleaned data for visualizations, queries, and machine learning model creation
- Perform Exploratory Data Analysis (EDA) Using Visualization and SQL:
  - Conducted exploratory data analysis (EDA) using visualization and SQL
- Perform Interactive Visual Analytics Using Folium and Plotly Dash:
  - Developed interactive visual analytics using Folium and Plotly Dash
- Perform Predictive Analysis Using Classification Models:
  - Implemented predictive analysis using classification models
- How to Build, Tune, Evaluate Classification Models:
  - Built, tuned, and evaluated classification models for prediction

Data Collection

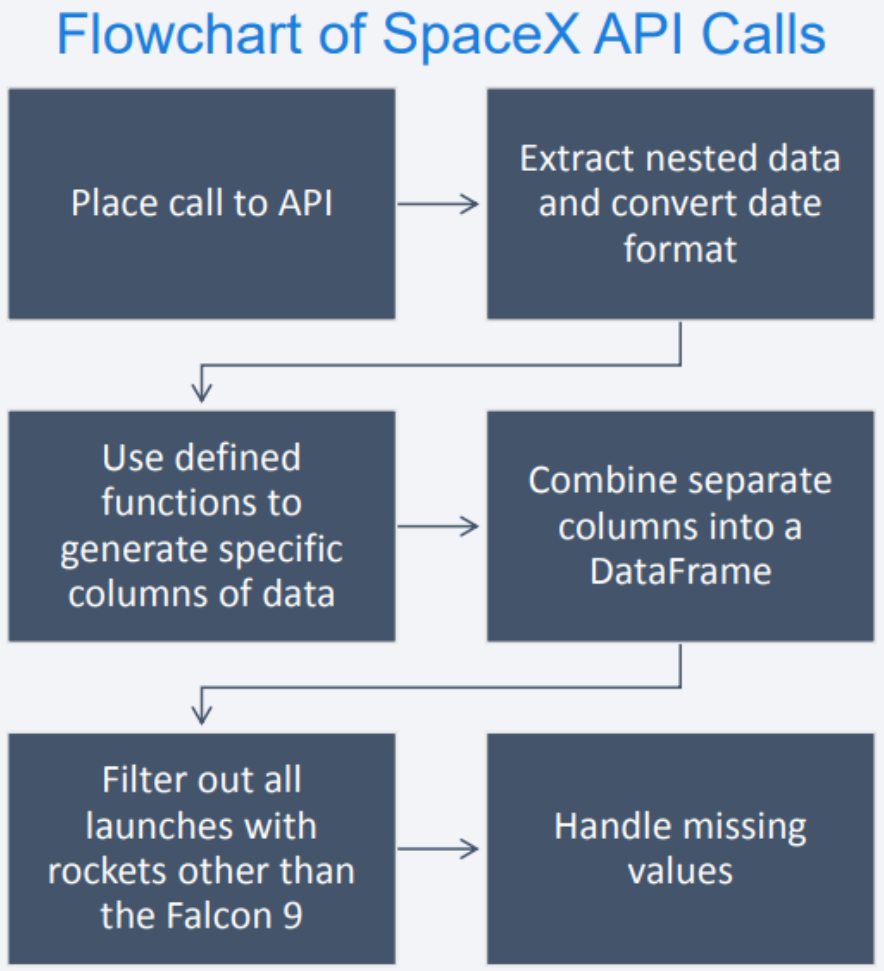
- SpaceX API
- Wikipedia page (9 June 2021)
- Provided data from IBM course



Data Collection – SpaceX API

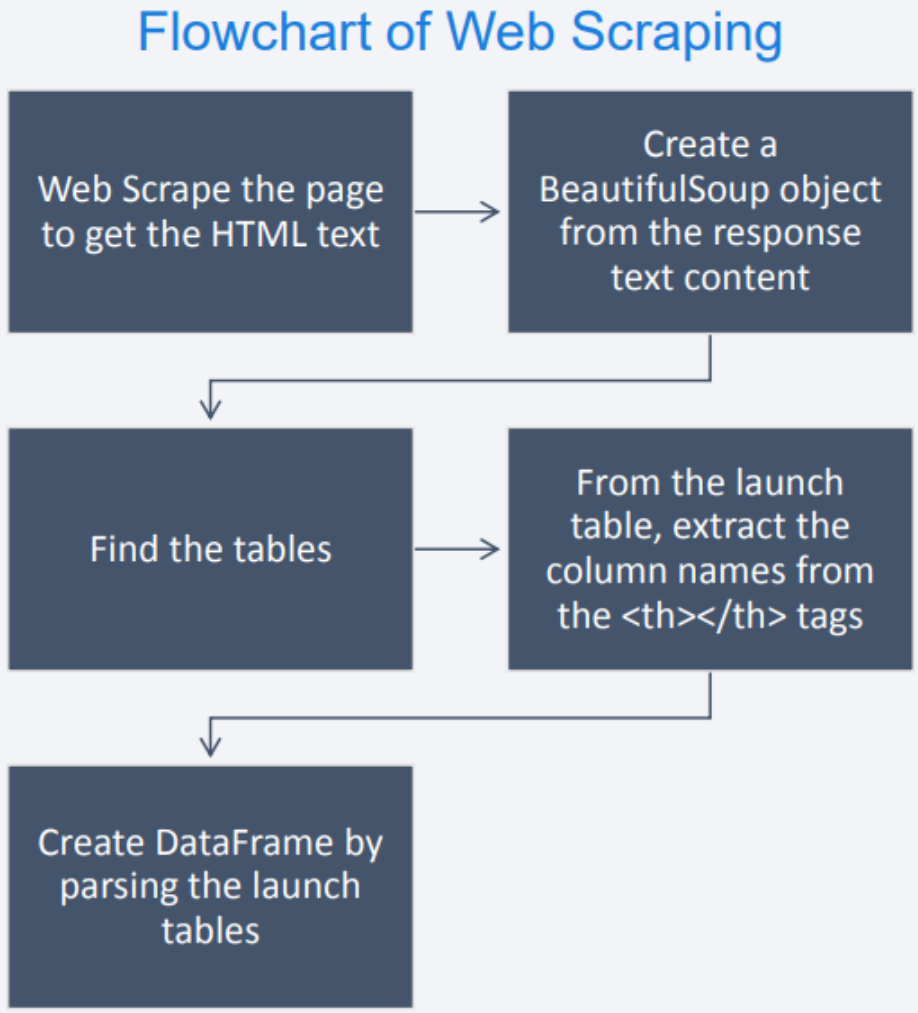
- SpaceX API available publically
- Get request ➔ Pandas dataframe

- GitHub URL : https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week1/jupyter-labs-spacex-data-collection-api.ipynb



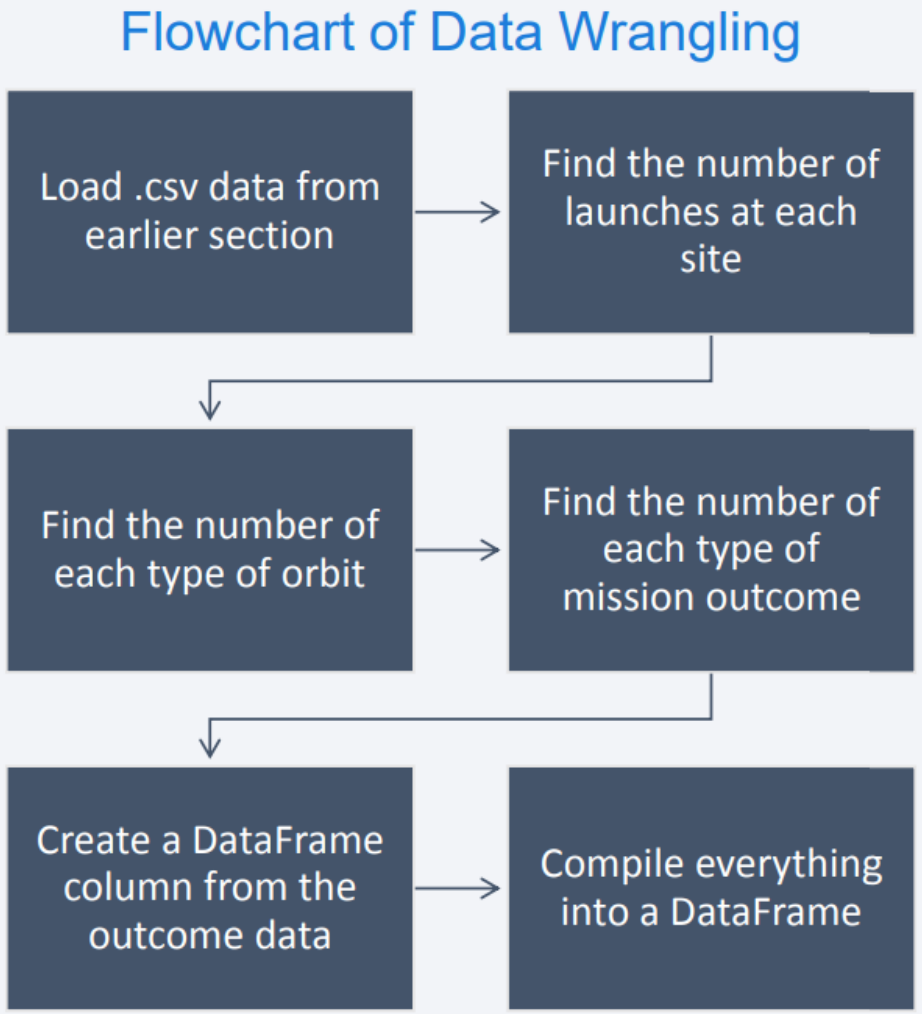
Data Collection - Scraping

- A Wikipedia page contains tables with information about SpaceX launches.
- These tables can be scraped to gather launch data, which can then be compiled into a Pandas DataFrame for further examination.
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week1/jupyter-labs-webscraping.ipynb



Data Wrangling

- The .csv file from the initial section contains data requiring cleaning.
- Launch sites, orbit types, and mission outcomes were refined.
- Mission outcome types were simplified into a binary classification: 1 for successful Falcon 9 first stage landing and 0 for failure.
- This new classification was incorporated into the DataFrame for further analysis.
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week1/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Charts for Launch Site Trends:
  - Scatterplot showing mission outcomes in relation to Launch Site and Flight Number.
  - Scatterplot showing mission outcomes in relation to Launch Site and Payload.
- Charts for Orbit Type Trends:
  - Bar chart displaying mission outcomes in relation to Orbit Type.
  - Scatterplot showing mission outcomes in relation to Orbit Type and Flight Number.
  - Scatterplot showing mission outcomes in relation to Orbit Type and Payload.
- Chart for Time-Based Trends:
  - Line plot illustrating mission outcome trends by year.
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week2/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Queries were written to extract information about:
  - Launch sites
  - Payload masses
  - Dates
  - Booster types
  - Mission outcomes
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week2/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

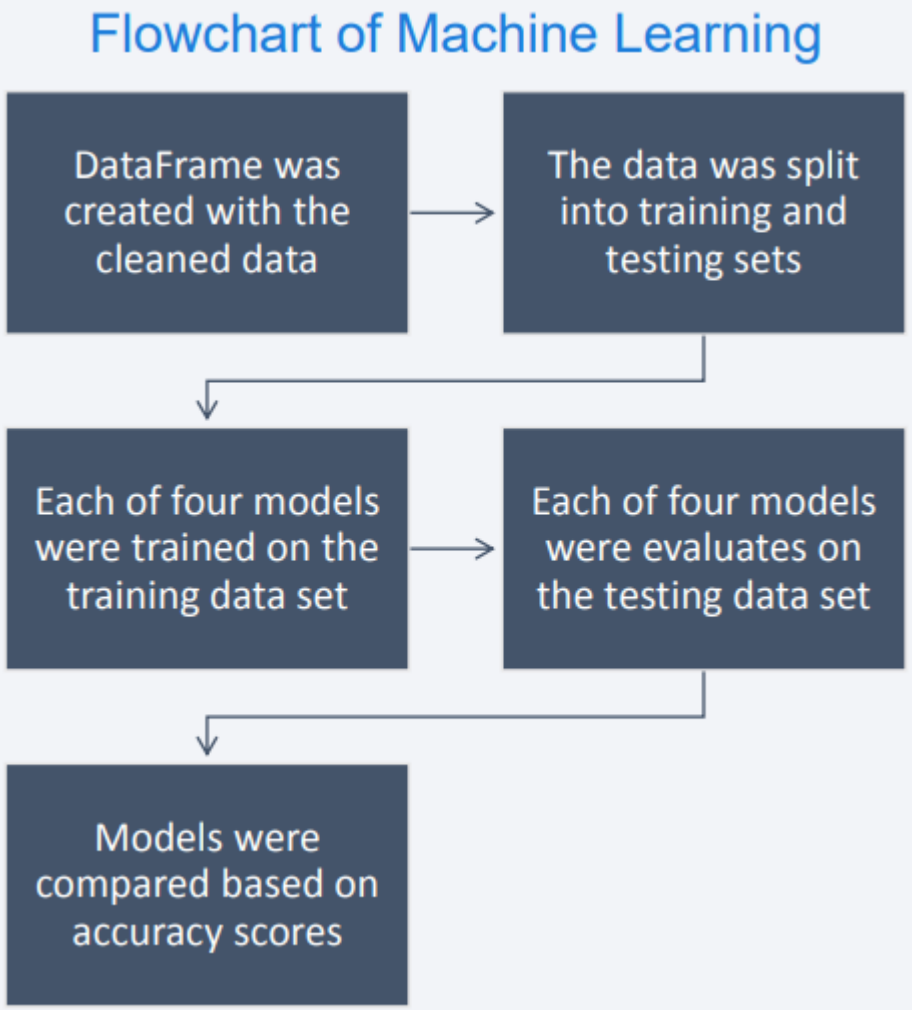
- **Markers**:
  - Added markers for launch sites.
  - Added markers for the NASA Johnson Space Center.
- **Circles**:
  - Added circles to represent the launch sites.
- **Lines**:
  - Added lines to illustrate distances to nearby features:
    - Distance from CCAFS LC-40 to the coastline.
    - Distance from CCAFS LC-40 to the rail line.
    - Distance from CCAFS LC-40 to the perimeter road.
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week3/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- **Input Dropdown**:
  - Used to select one or all launch sites for the pie chart and scatterplot.
- **Pie Chart**:
  - Displays one of two things:
    - For All Sites: The distribution of successful Falcon 9 first stage landings among the sites.
    - For One Site: The distribution of successful and failed Falcon 9 first stage landings for that specific site.
- **Input Slider**:
  - Used to filter the payload masses for the scatterplot.
- **Scatterplot**:
  - Displays the distribution of Falcon 9 first stage landings split by payload mass, mission outcome, and booster version category.
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week3/Build_a_Dashboard_Application_with_Plotly_Dash.pdf

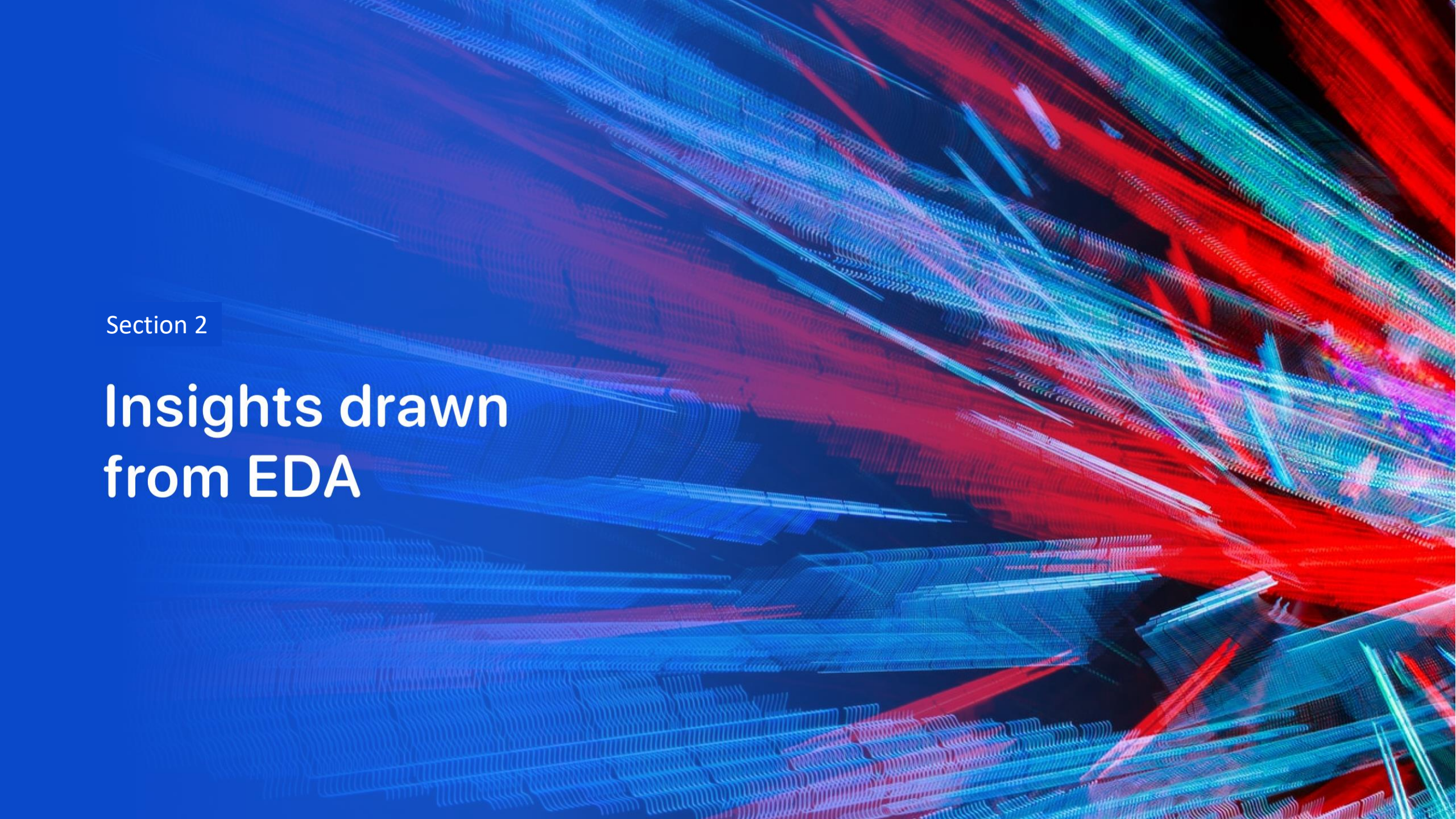
Predictive Analysis (Classification)

- The dataset was divided into training and testing sets.
- Machine Learning Models:
  - Logistic Regression, SVM (Support Vector Machine), Decision Tree, and KNN (k-Nearest Neighbors) models were trained on the training data.
- Hyper-parameter Tuning:
  - Hyper-parameters were evaluated using GridSearchCV(), and the best parameters were selected with .best_params_.
- Model Evaluation:
  - Using the best hyper-parameters, each of the four models was scored on accuracy using the testing data set.
- https://github.com/claquettes-sandales/DataScience_Capstone/blob/main/Week4/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb



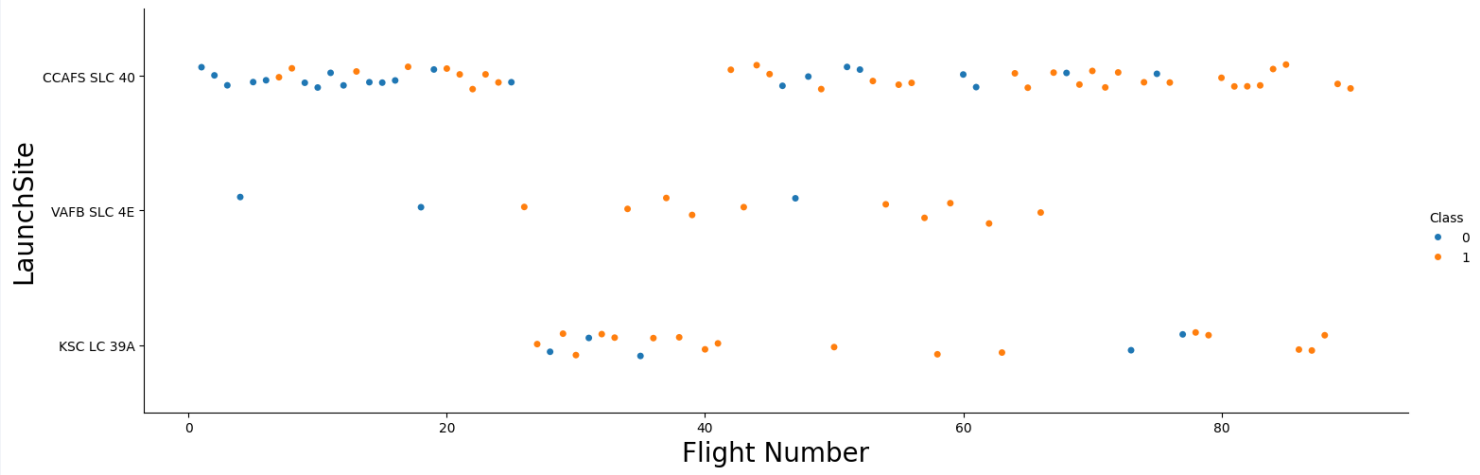
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

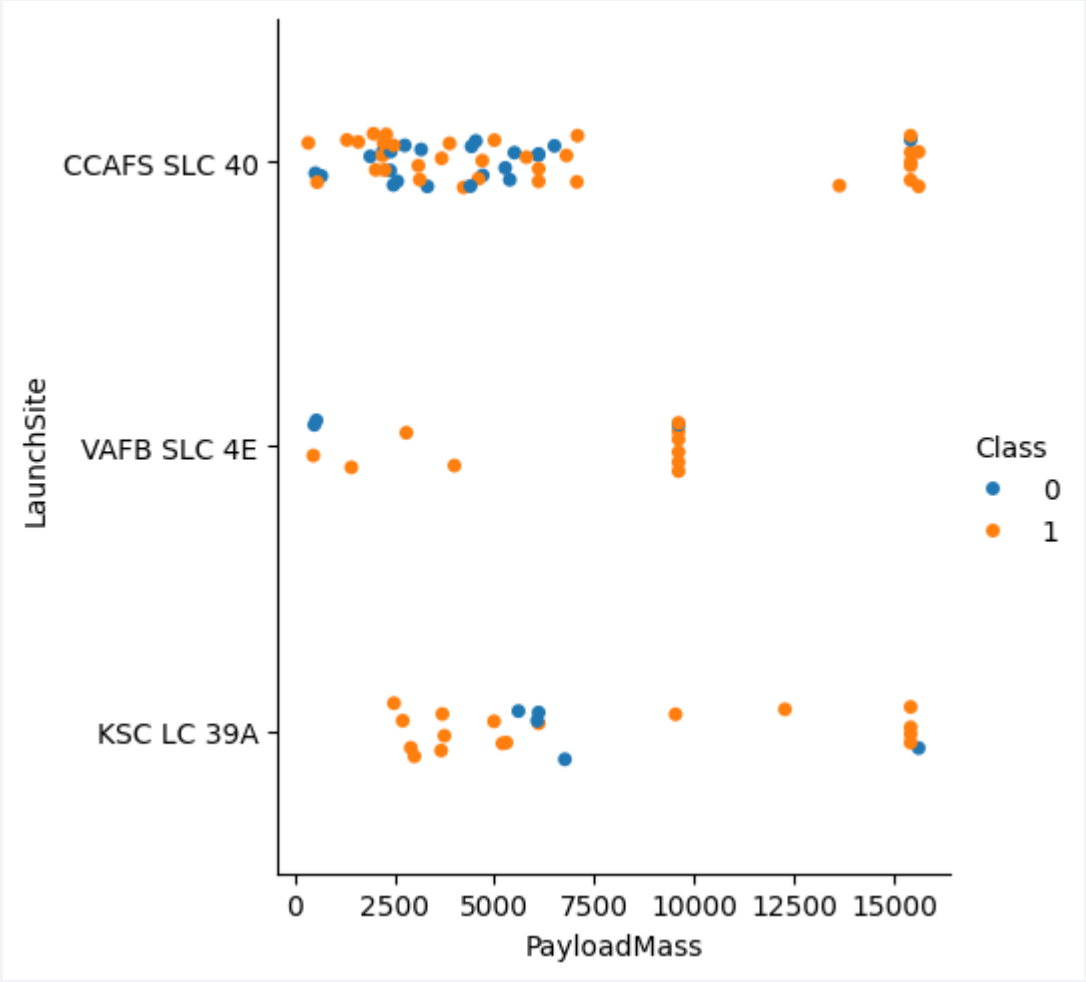
- Launch Site Analysis:
  - Success rate varies noticeably with different launch sites.
- Flight Number Analysis:
  - Successful Falcon 9 first stage landings appear to become more prevalent as the flight number increases.



Falcon 9 first stage failed landing are indicated by 0 class and success by 1 class

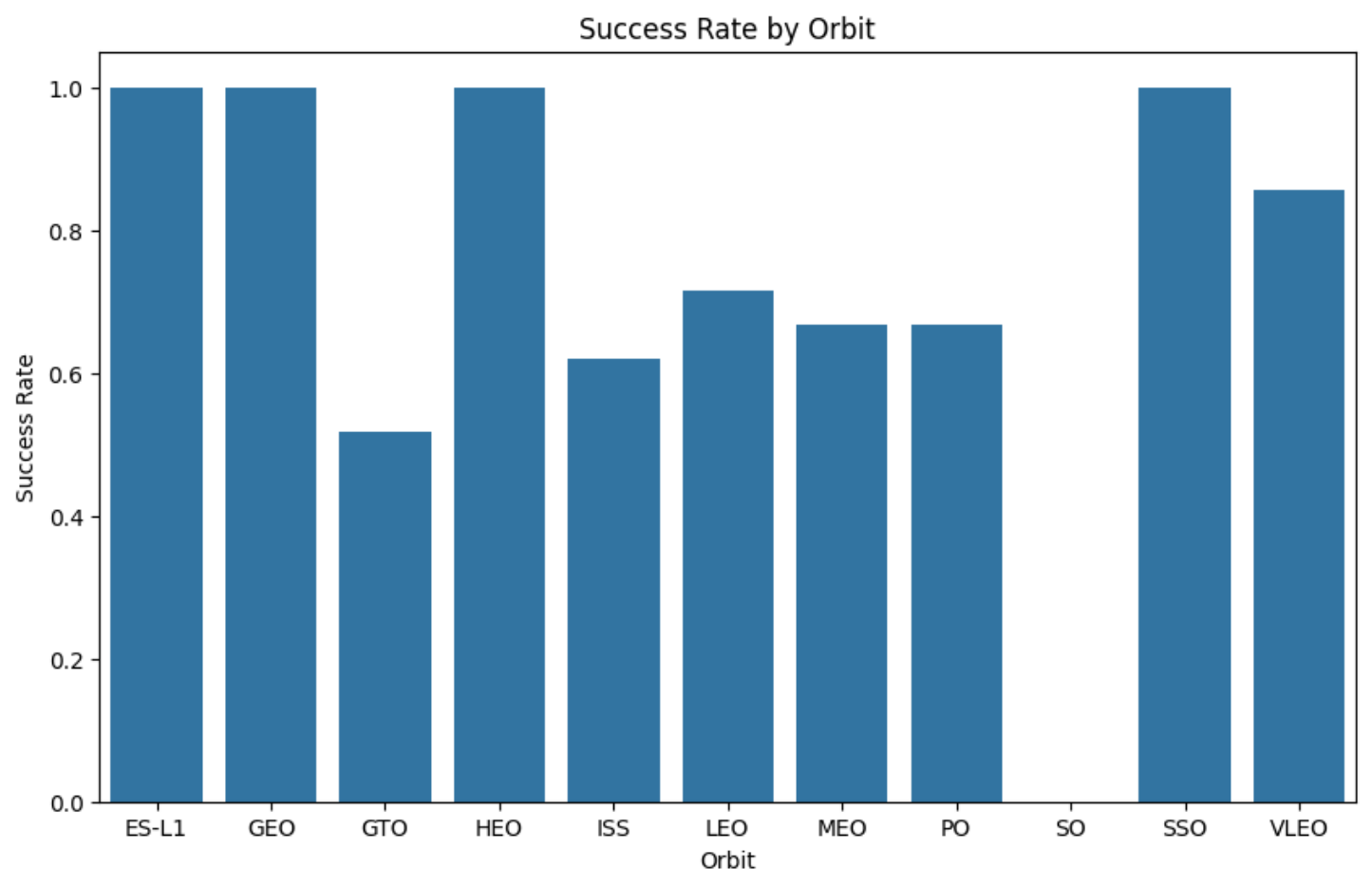
Payload vs. Launch Site

- **CCAFS SLC-40 Launch Site**:
  - The payload mass and the landing outcome do not appear to be strongly correlated.
- **KSC LC-39A Launch Site**:
  - The failed landings are all grouped around a narrow band of payload masses.



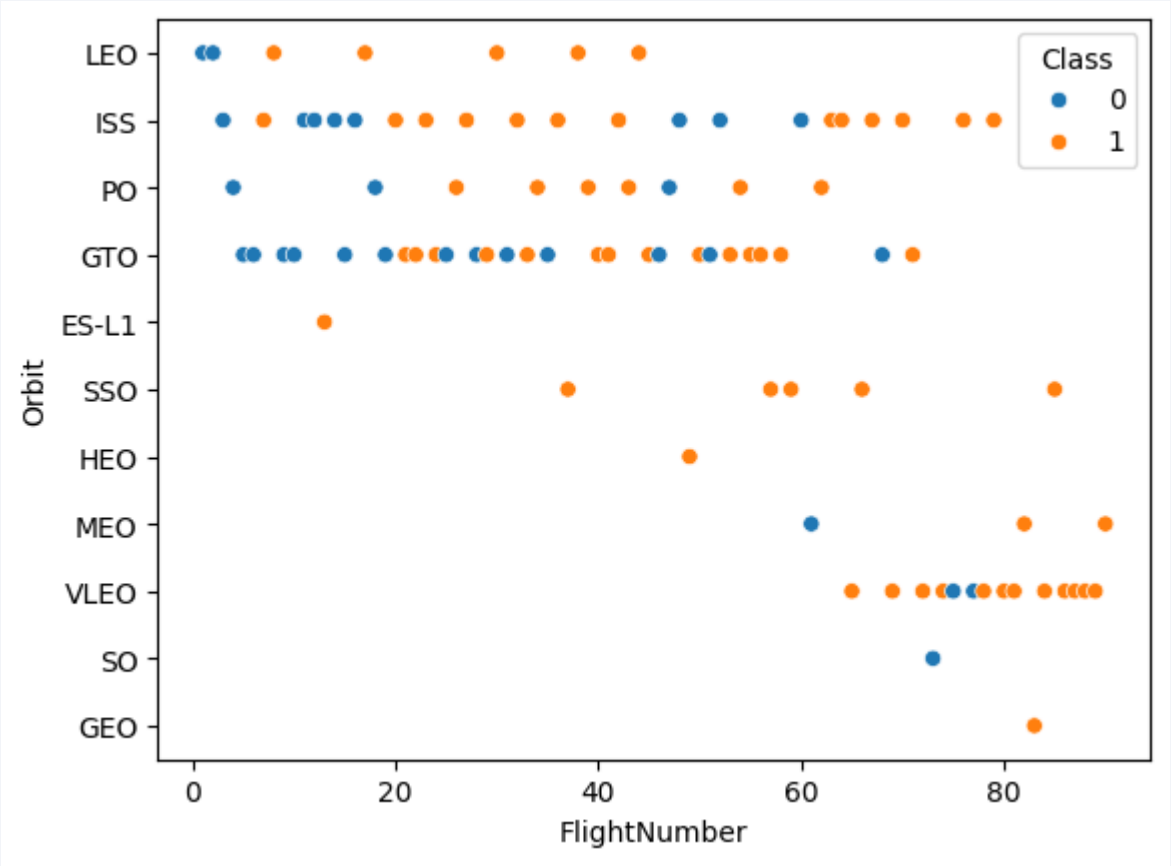
Success Rate vs. Orbit Type

- **Orbits with No Failed First Stage Landings**:
  - ES-L1, SSO, HEO, and GEO orbits.
- **Orbits with No Successful First Stage Landings**:
  - SO orbits.



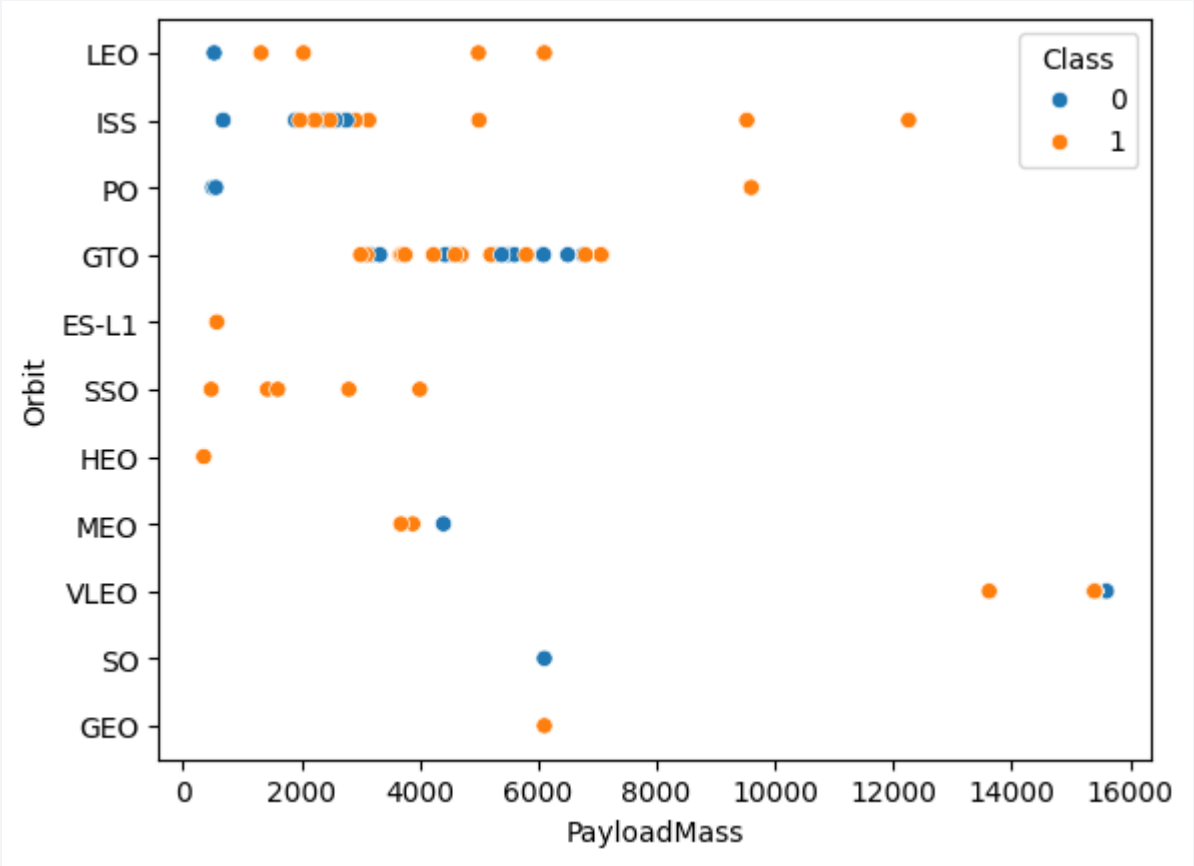
Flight Number vs. Orbit Type

- Flight number and Success rate are somewhat correlated



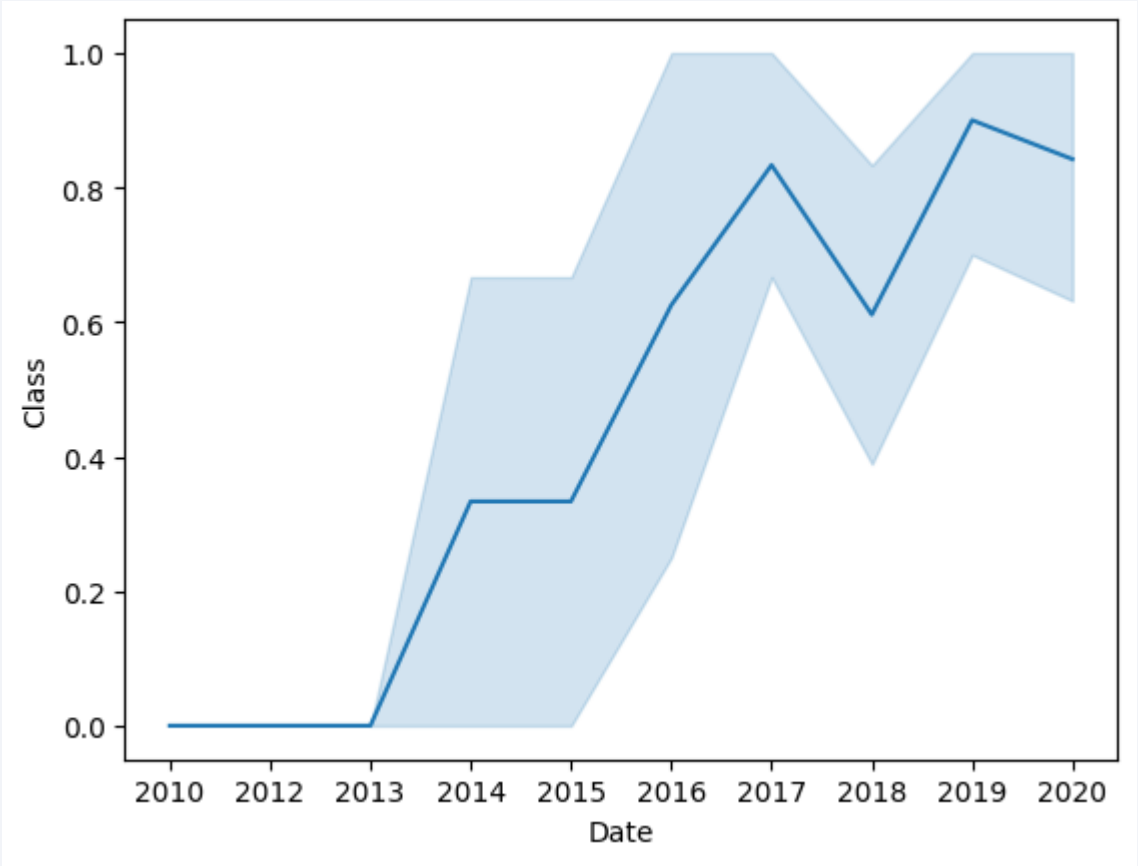
Payload vs. Orbit Type

- Some orbit types have better success rates than others.
- Success rate appears to have no obvious correlation with payload mass



Launch Success Yearly Trend

- Success rate increase significantly over the years



All Launch Site Names

Task 1

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

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

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

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

Launch Site Names Begin with 'CCA'

Task 2

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

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

Python

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

Task 3

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

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

31]

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

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

Task 4

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

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

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

Average_Payload_Mass
2928.4

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

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

3]

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

First_Successful_Landing_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

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

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

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

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

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

Boosters Carried Maximum Payload

Task 8

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

```
%%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

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.

```
%%sql
SELECT SUBSTR("Date", 6, 2) AS Month,
       "Landing_Outcome",
       "Booster_Version",
       "Launch_Site"
FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE 'Failure (drone ship)%'
  AND SUBSTR("Date", 1, 4) = '2015';
```

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

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

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.

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

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

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

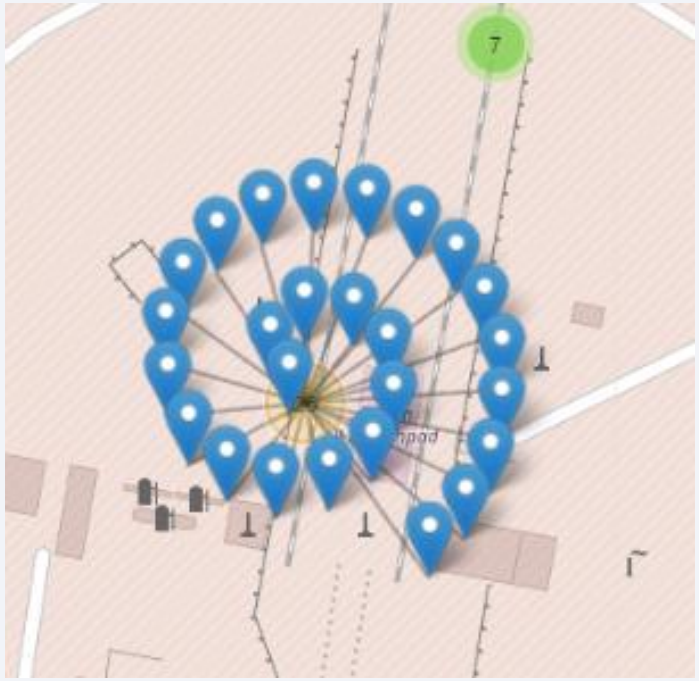
Launch Sites Proximities Analysis

Falcon 9 Launch Site Locations

- VAFB SLC-4E (California, USA)
  - Vandenberg Air Force Base Space Launch Complex 4E
- KSC LC-39A (Florida, USA)
  - Kennedy Space Center Launch Complex 39A
- CCAFS LC-40 (Florida, USA)
  - Cape Canaveral Air Force Station Launch Complex 40
- CCAFS SLC-40 (Florida, USA)
  - Cape Canaveral Air Force Station Space Launch Complex 40

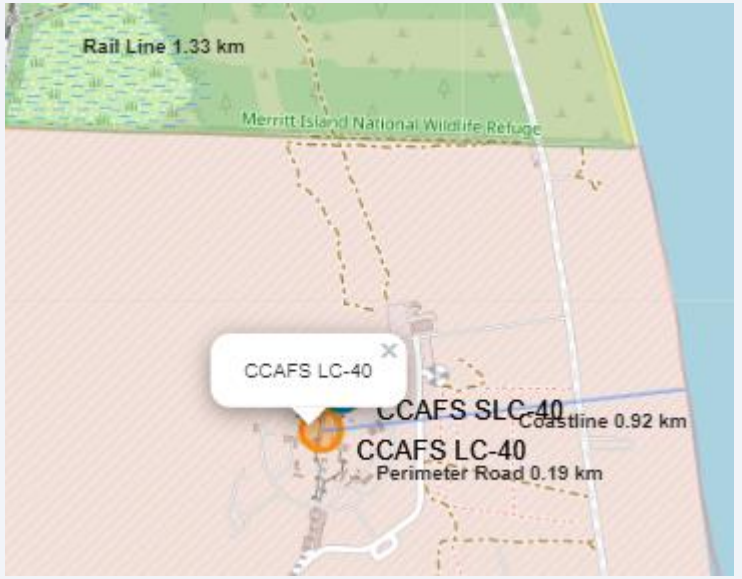
Map Markers of Success/Failed Landings

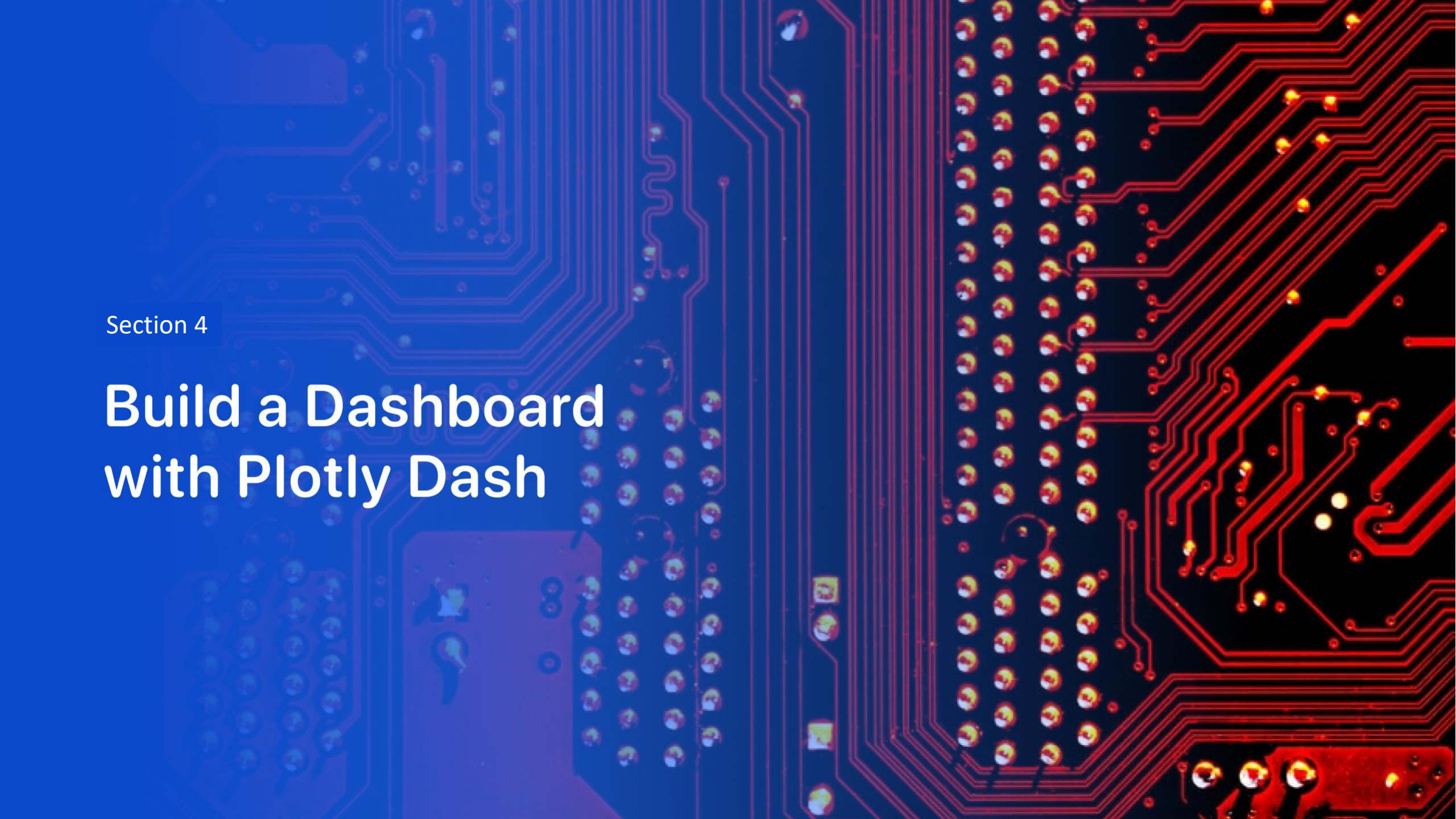
- The markers display the mission outcomes (Success/Failure) for Falcon 9 first stage landings. They are grouped on the map to be associated with the geographical coordinates for the launch site.
- A sense of a launch site's success rate for Falcon 9 first stage landings can be gleaned from the relative number of green success markers to red failure markers.



Distance from Launch Site to Proximities

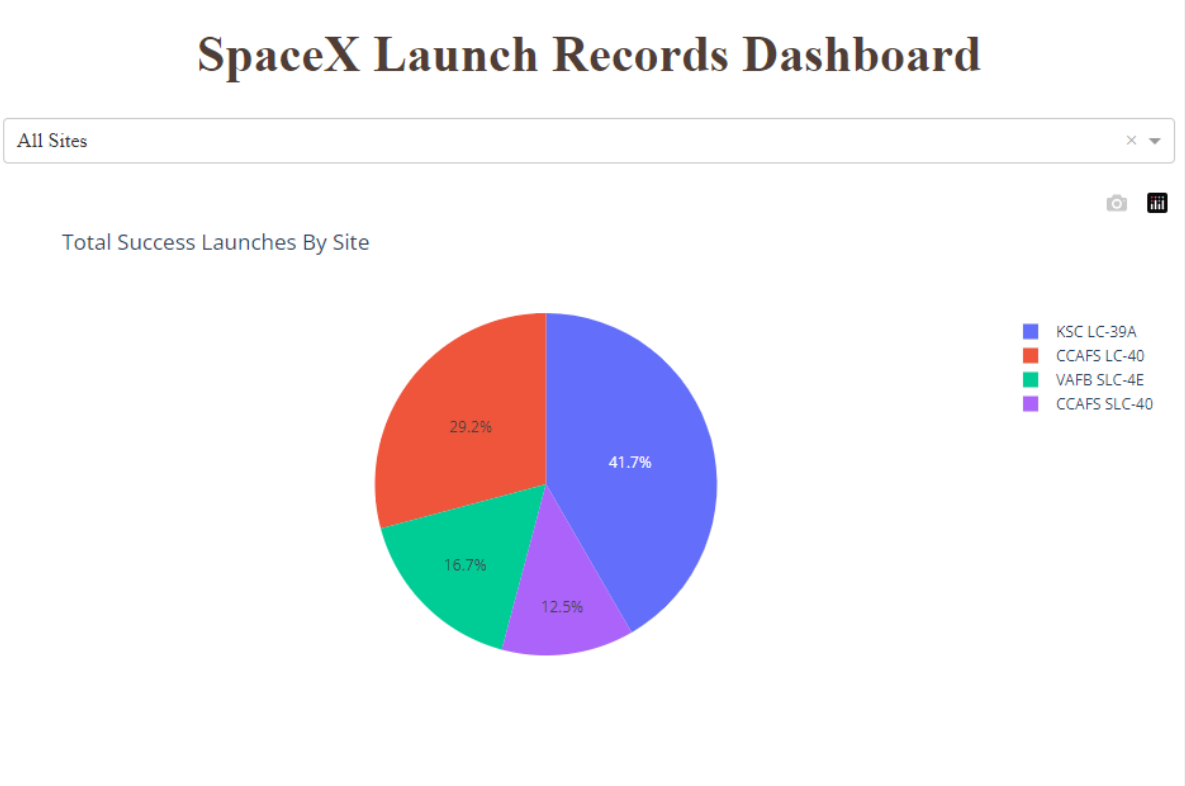
- **CCAFS Launch Sites**:
  - CCAFS LC-40 and CCAFS SLC-40 have coordinates that are close but not exactly on top of each other.
- **Proximity to CCAFS LC-40**:
  - The perimeter road is 0.19 km away.
  - The coastline is 0.92 km away.
  - The rail line is 1.33 km away.




Section 4
Build a Dashboard with Plotly Dash

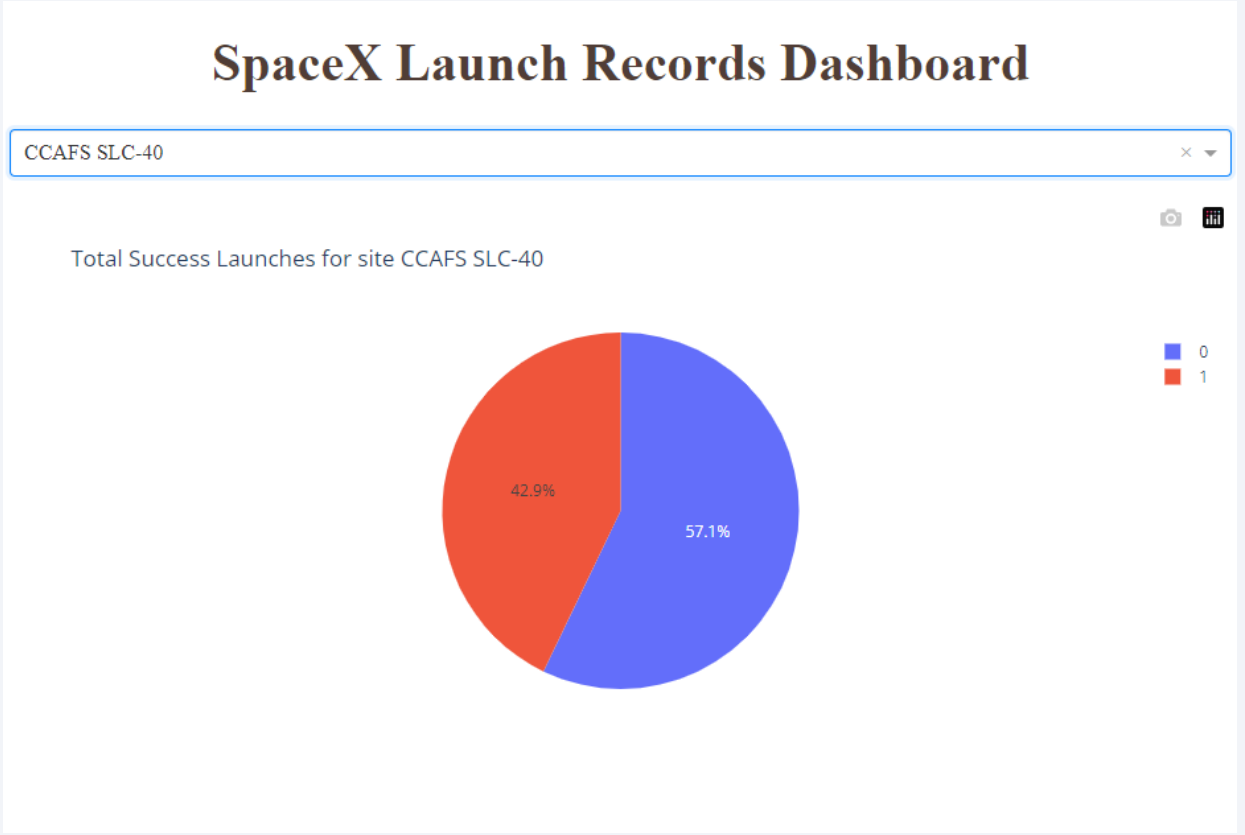
Launch Success Count for All Sites

- **Dropdown Menu Functionality**:
  - Allows selection of one or all launch sites.
- **Pie Chart Display**:
  - When all launch sites are selected, it shows the distribution of successful Falcon 9 first stage landing outcomes across different launch sites.
- **Success Rate Highlight**:
  - The highest share of successful Falcon 9 first stage landings (41.7% of the total) occurred at KSC LC-39A.



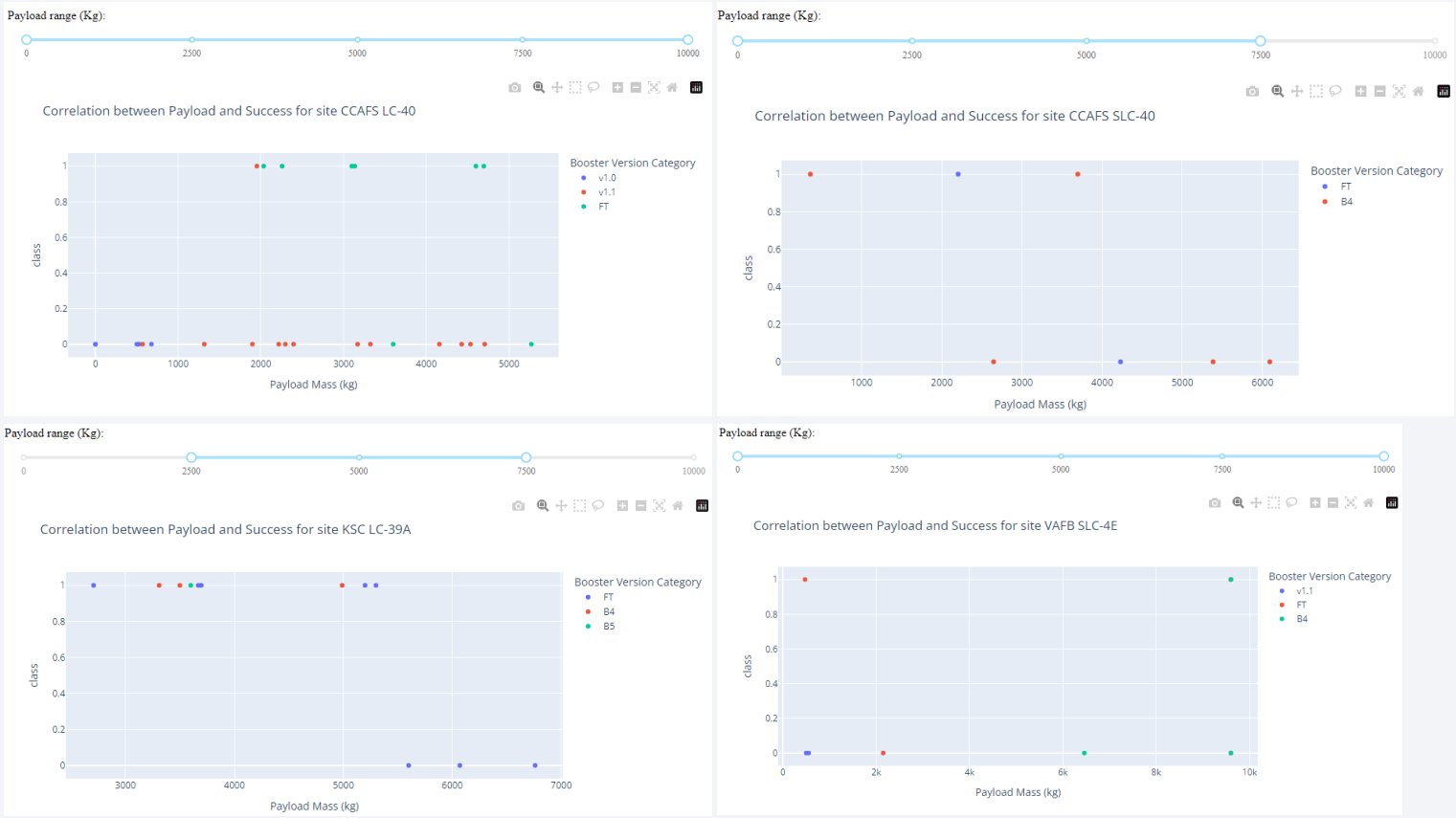
Launch Site with Highest Launch Success Ratio

- **Landing Outcomes in Pie Chart**:
  - Failed landings are represented by the '0' Class (blue wedge).
  - Successful landings are represented by the '1' Class (red wedge).
- **Highest Success Rate**:
  - CCAFS SLC-40 had the highest Falcon 9 first stage landing success rate at 42.9%.



Payload vs. Launch Outcome

- **Payload vs. Launch Outcome Scatter Plots**:
  - Screenshots depict scatter plots for all sites, with different payload ranges selected in the slider.
- **Payload Range**:
  - The payload range from approximately 2,000 kg to 5,000 kg has the highest success rate.
- **Booster Version Category**:
  - The 'FT' booster version category has the highest success rate.

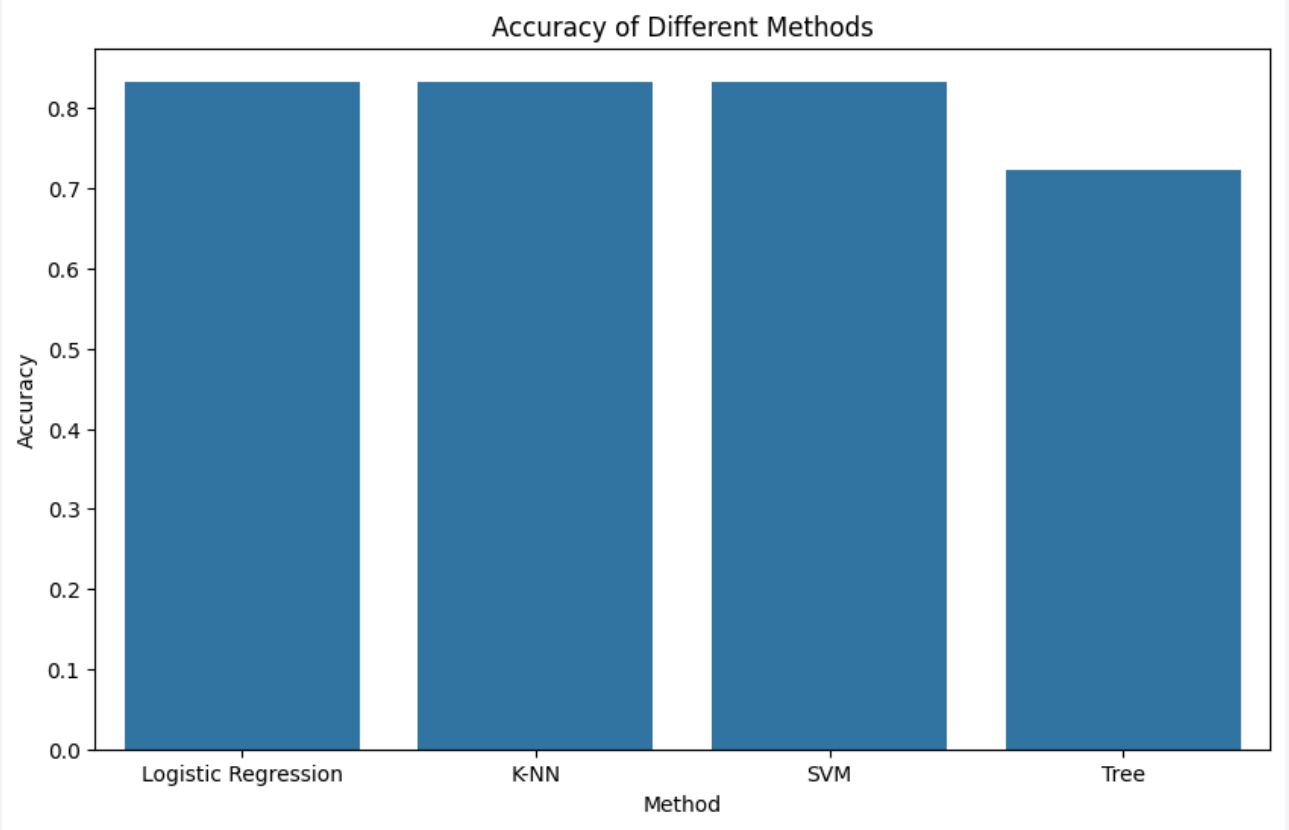


Section 5

Predictive Analysis (Classification)

Classification Accuracy

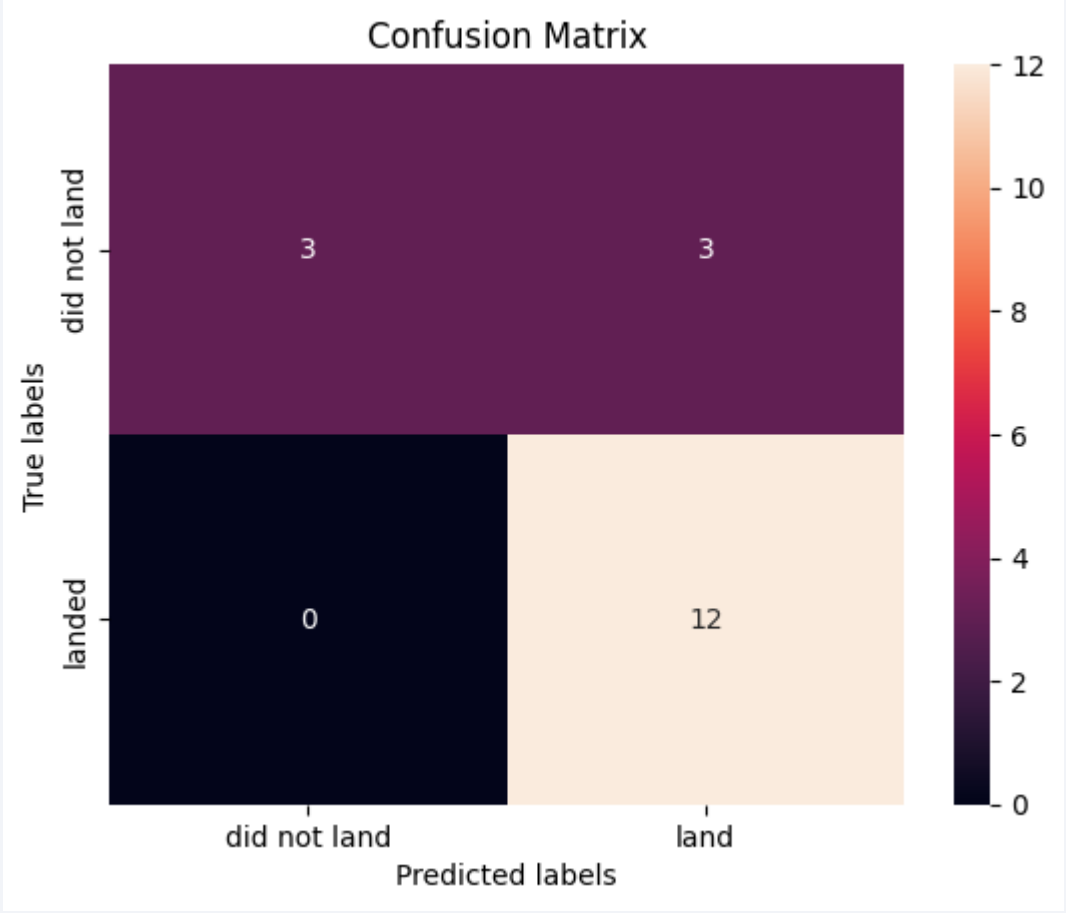
All models performed equally well but the random tree



Confusion Matrix

- Here the confusion matrix of the KNN estimator

- 12 True Positives // 3 True Negatives
- 3 False Positive // 3 False Negative



Conclusions

- SpaceX does not have a flawless record of Falcon 9 first stage landing outcomes.
- Falcon 9 first stage landing outcomes have shown a trend towards greater success with increasing launch numbers.
- Machine learning models can be utilized to predict future Falcon 9 first stage landing outcomes.

Appendix

- https://github.com/claquettes-sandales/DataScience_Capstone

Thank you!