IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

This project aimed to predict the success of the Falcon 9 first-stage landings, crucial for SpaceX's cost reduction strategy. We followed four key steps:

- Data Collection & Preparation:
    - Data was gathered via APIs and web scraping, then cleaned and formatted into a Pandas DataFrame for analysis.
- Exploratory Data Analysis (EDA):
    - We analyzed patterns in launch success rates, finding key factors like payload mass and launch sites influencing outcomes.

Executive Summary

(continue)

- **Interactive Dashboard & Mapping**:
    - Iinteractive Plotly Dash dashboard with visualizations (pie charts, scatter plots)
    - Folium map to explore launch site proximity, enabling dynamic analysis of launch records.
- Predictive Modeling:
    - Using machine learning (Logistic Regression, SVM, Decision Trees), we predicted landing success. Decision Tree was the top-performing model with over 87.5% accuracy.
    - This tool helps in predicting successful landings, enabling better cost management for SpaceX and providing insights for competitive bidding in the aerospace industry.

Introduction

- **Project Background and Context:** SpaceX's Falcon 9 rocket has revolutionized space travel by significantly reducing costs, primarily through the reuse of its first stage. A successful first-stage landing allows SpaceX to cut launch costs from the industry average of $165 million to just $62 million. This project aims to predict the success of these first-stage landings, providing crucial information for businesses evaluating SpaceX's cost advantage. Accurate predictions can also be used by competitors to optimize bids for future rocket launches.

- **Problems to Answer:**

1. **Can we predict whether the Falcon 9 first stage will land successfully?**
    1. Understanding the key factors influencing landing outcomes, such as launch site, payload mass, and orbit type.
2. **Which machine learning model can provide the most accurate predictions?**
    1. By training multiple models, we aim to identify the best approach for predicting landing success.

- These insights will help stakeholders make data-driven decisions, optimize costs, and evaluate the feasibility of competing with SpaceX.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API
  - Web scraping
- Perform data wrangling
  - Mission outcomes are encoded as 1 for success and 0 for failure
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

/

Data Collection – SpaceX API

- https://api.spacexdata.com/v4/rockets/
- https://api.spacexdata.com/v4/launchpads/
- https://api.spacexdata.com/v4/payloads/
- https://api.spacexdata.com/v4/past/

- GitHub URL of the notebook https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/spacex_data_collection_api_v2.ipynb

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- BeautifulSoup

- GitHub URL of the web scraping notebook: https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/webscraping.ipynb

Place your flowchart of web scraping here

Data Wrangling

- Mission outcomes are encoded as 1 for success and 0 for failure
- https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/spacex_Data_wrangling_v2.ipynb

EDA with Data Visualization

- Scatter point charts where plotted to demonstrate:
  - Flight number vs. Payload mass
  - Flight number vs launch site
  - Payload mass vs launch site
- Line chart to demonstrate success rate over the years
- https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/eda_dataviz_v2.ipynb

EDA with SQL

- SQL queries performed to:
  - 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
  - List the date when the first succesful landing outcome in ground pad was acheived.
  - 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

- https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/eda_sql_coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Added colored markers for all launch records (green for success and red for failed)
- Added circles on:
  - NASA Johnson Space Center
  - Launch Sites
- https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/launch_site_location_v2.ipynb

Build a Dashboard with Plotly Dash

- Added a pie chart to show the total successful launches count for all sites and for each one seperately
- Add a scatter chart to show the correlation between payload and launch success
- https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Tested Models:
  - Logistic Regression
  - Support Vector Machine Classifier
  - Dessicion Tree Classifier
  - K Nearest Neighbours
- Splitted data to train/test sets for better model evaluation and GridsearchCV for hyperparameter tuning
- https://github.com/arieg88/IBM-Applied-Data-Science-Capstone/blob/main/Learning_Prediction_Part_5_v1.ipynb

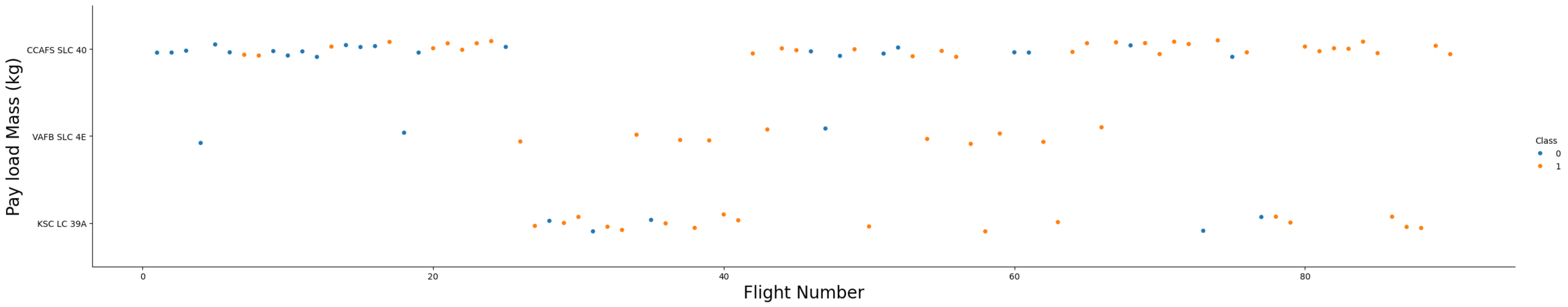
Results

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

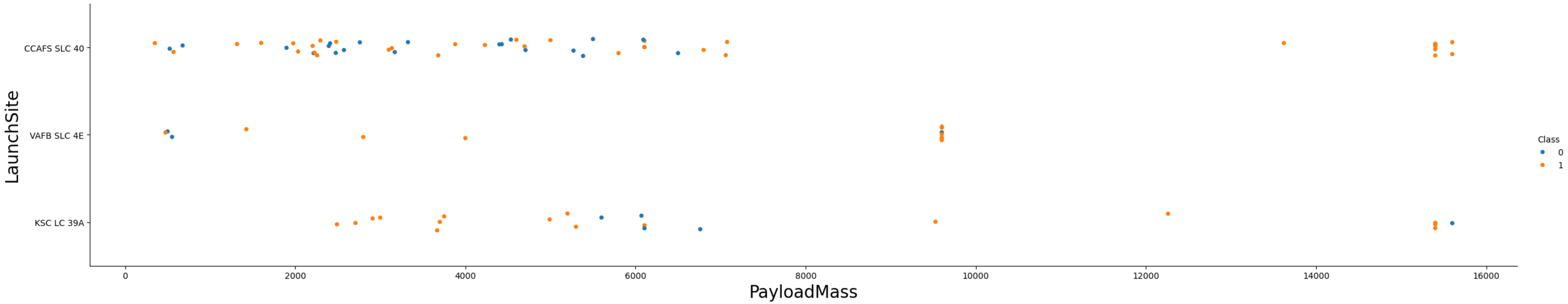
Section 2

Insights drawn from EDA

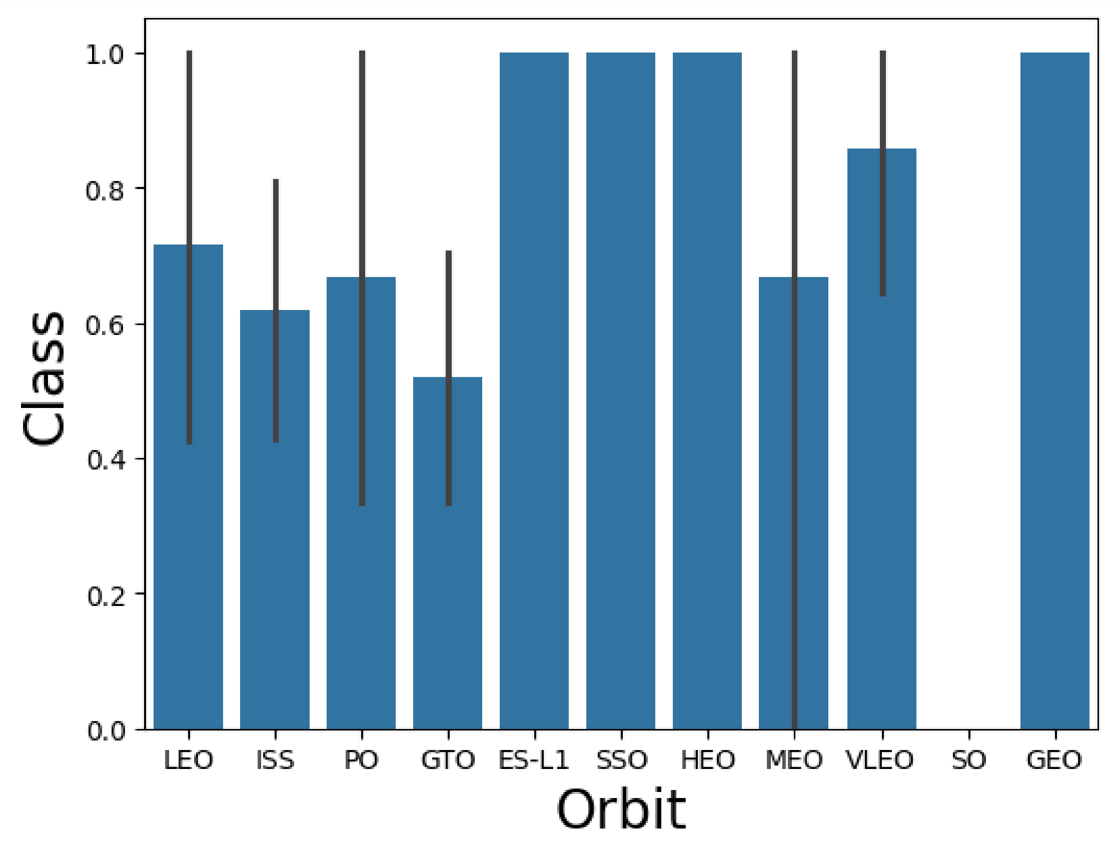
Flight Number vs. Launch Site



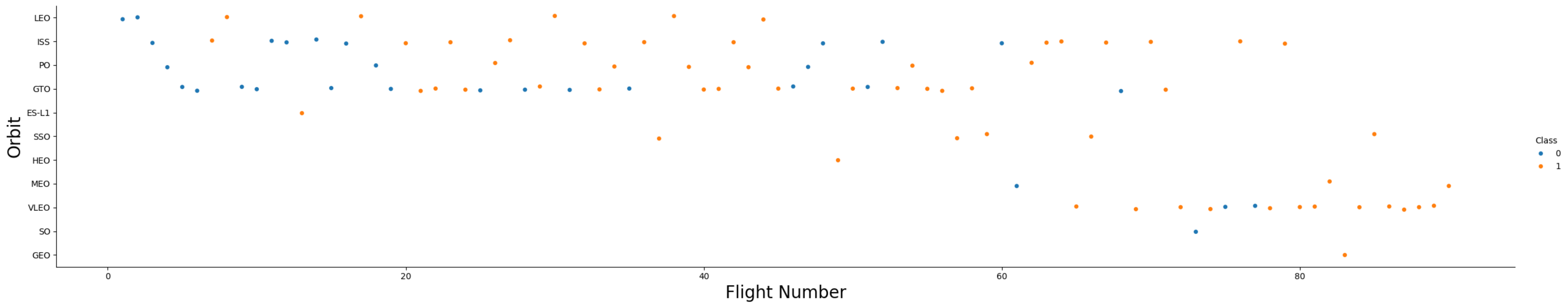
Payload vs. Launch Site



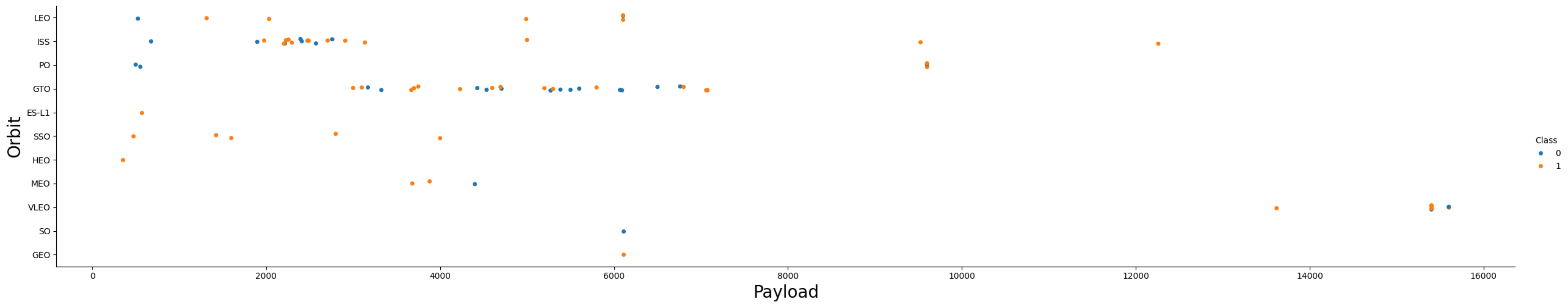
Success Rate vs. Orbit Type



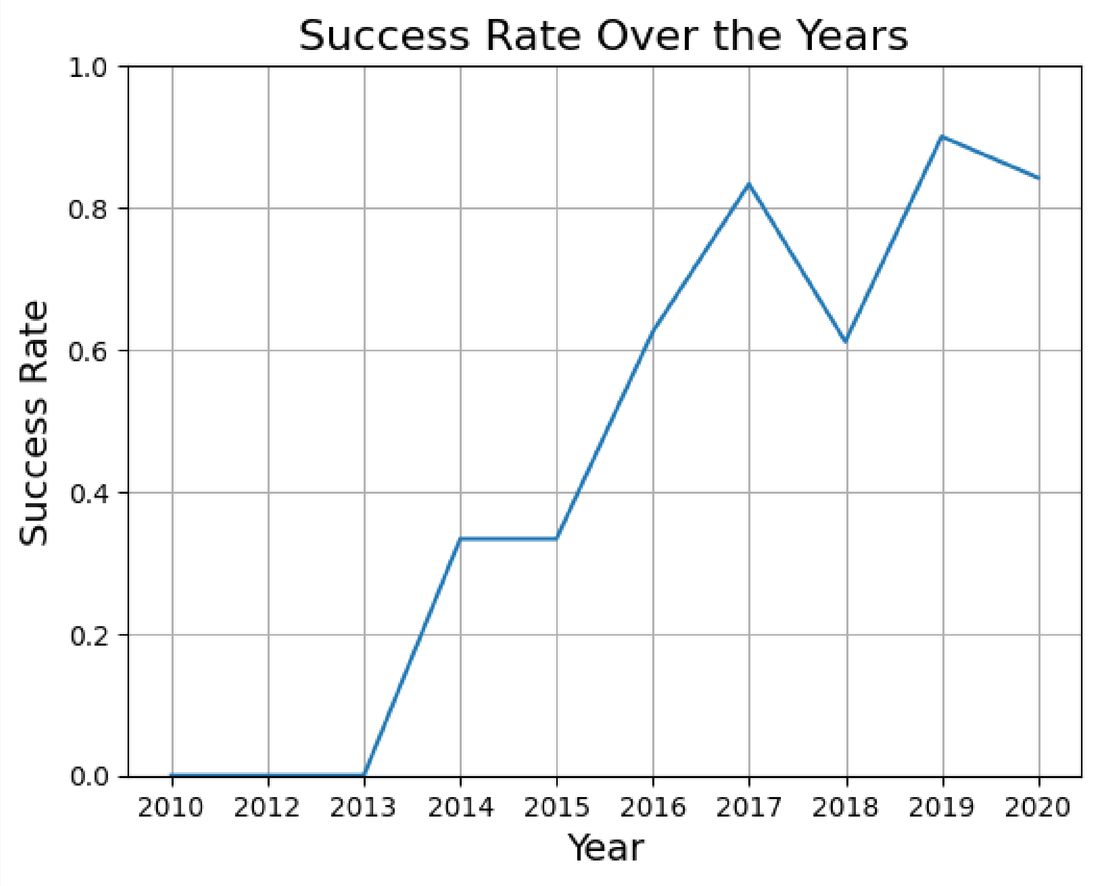
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- Distinct will show us only the distinct names in the launch site column

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

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

```
 * 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'

- Used the LIKE keyword to search for launch sites that start with cca, and the LIMIT keyword to show only 5 results.

```
%%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

```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") as totalPayload
FROM SPACEXTABLE
WHERE Customer = "NASA (CRS)";
```

totalPayload
45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG("PAYLOAD_MASS__KG_")
FROM SPACEXTABLE
WHERE Booster_Version LIKE "F9 v1.1";
```

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

AVG("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

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

Hint:Use min function

```
%%sql
SELECT MIN(Date)
FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Success%';
```

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

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT DISTINCT Booster_Version
FROM SPACEXTABLE
WHERE Mission_Outcome = 'Success'
  AND 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

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

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

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

Boosters Carried Maximum Payload

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

```
 * 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

```
%%sql
SELECT
    CASE SUBSTR(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month_Name,
    Booster_Version,
    Launch_Site,
    Landing_Outcome
FROM SPACEXTABLE
WHERE Landing_Outcome='Failure (drone ship)' AND
      SUBSTR(Date, 0, 5) = '2015';
```

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

Month_Name	Booster_Version	Launch_Site	Landing_Outcome
January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

```
%%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.
```

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

<Folium Map Screenshot 1>

- Replace <Folium map screenshot 1> title with an appropriate title

- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

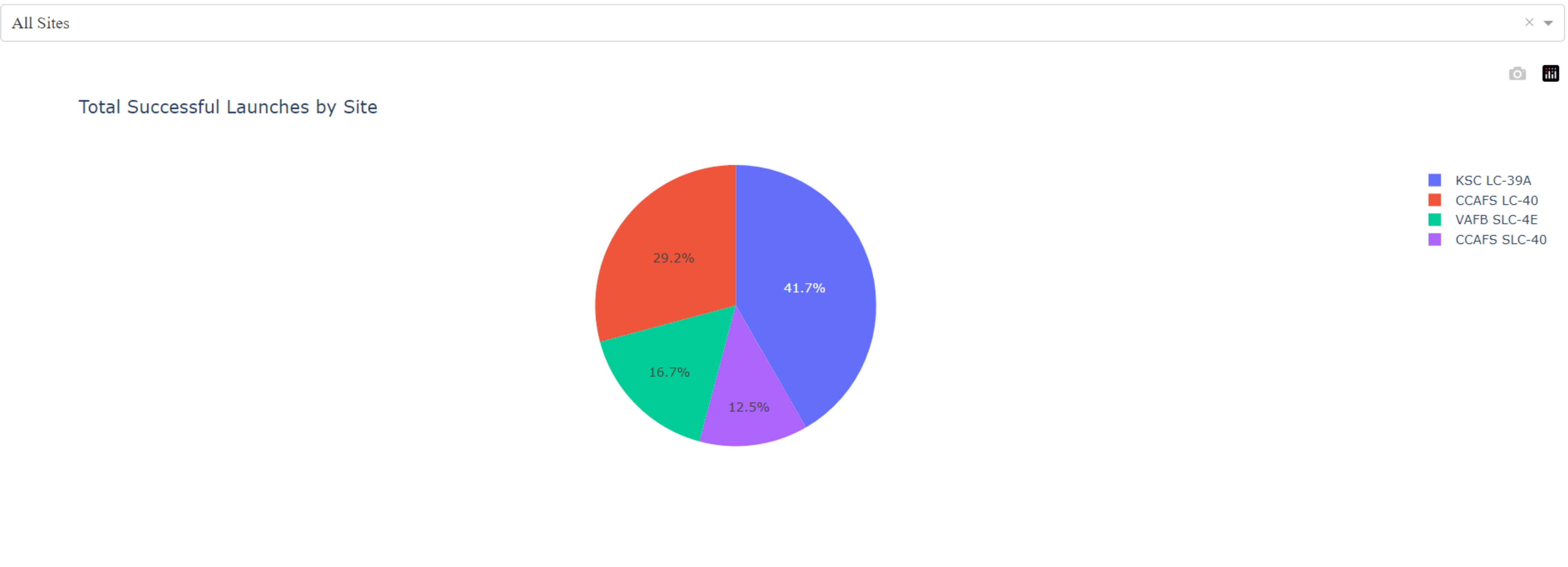
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

Section 4

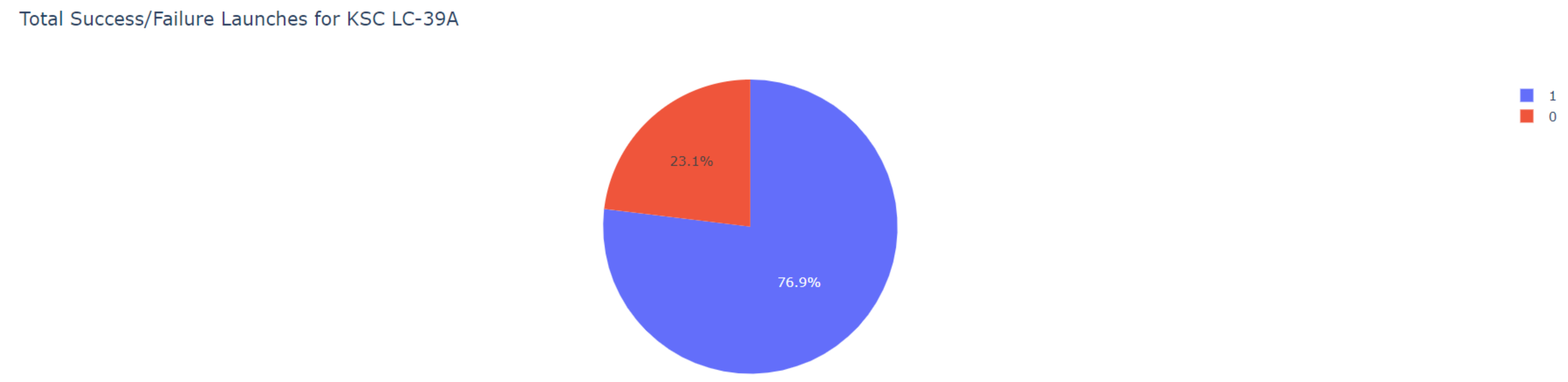
Build a Dashboard with Plotly Dash

Total Successful Launches By Site



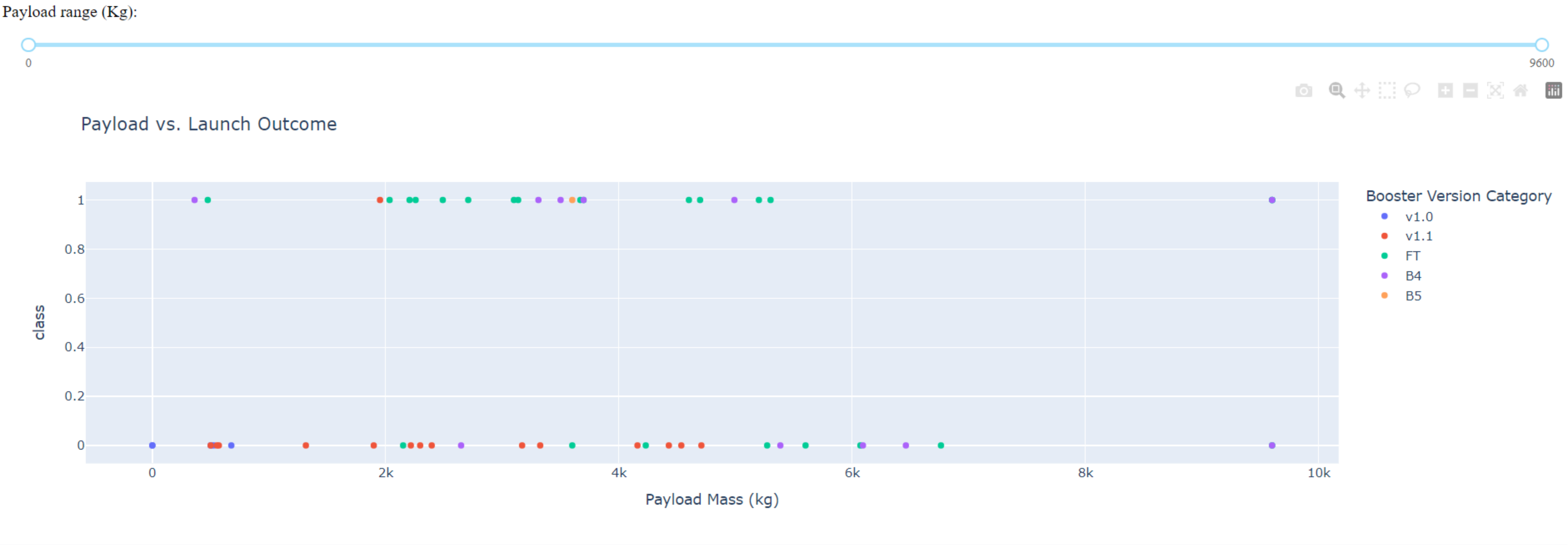
Launch Site With Highest Success Rate

KSC LC-39A

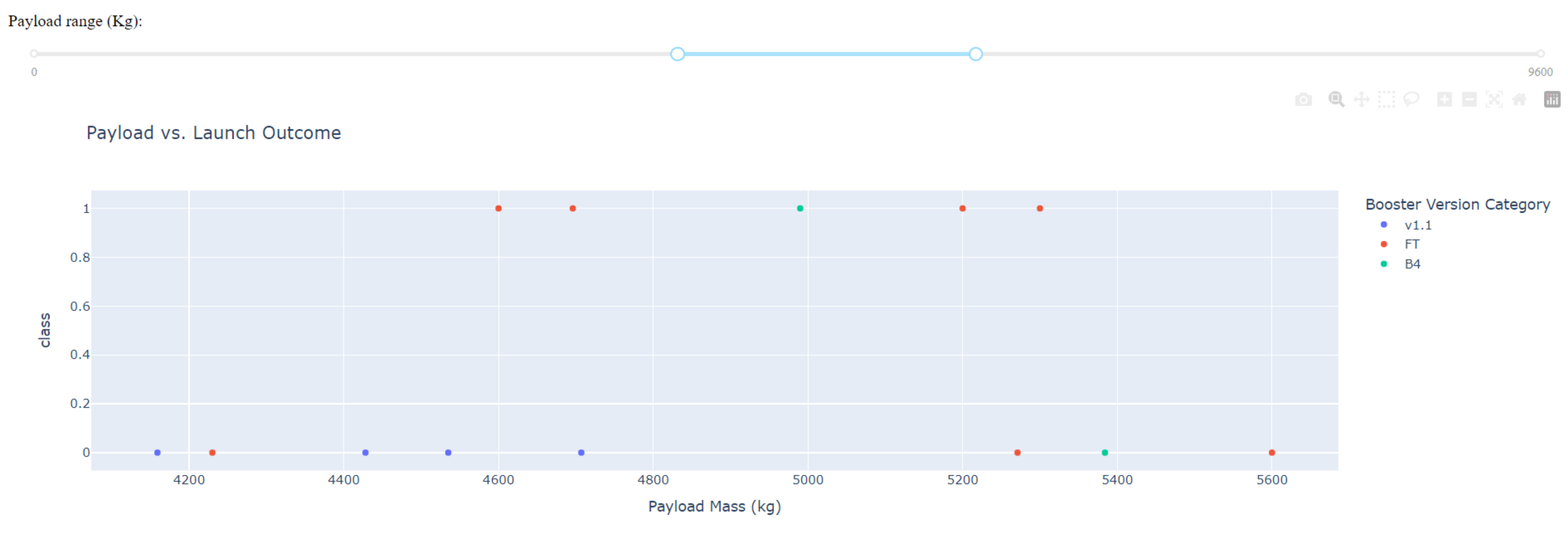


Payload range (Kg):

Payload vs. Launch Outcome scatter plot for all Sites



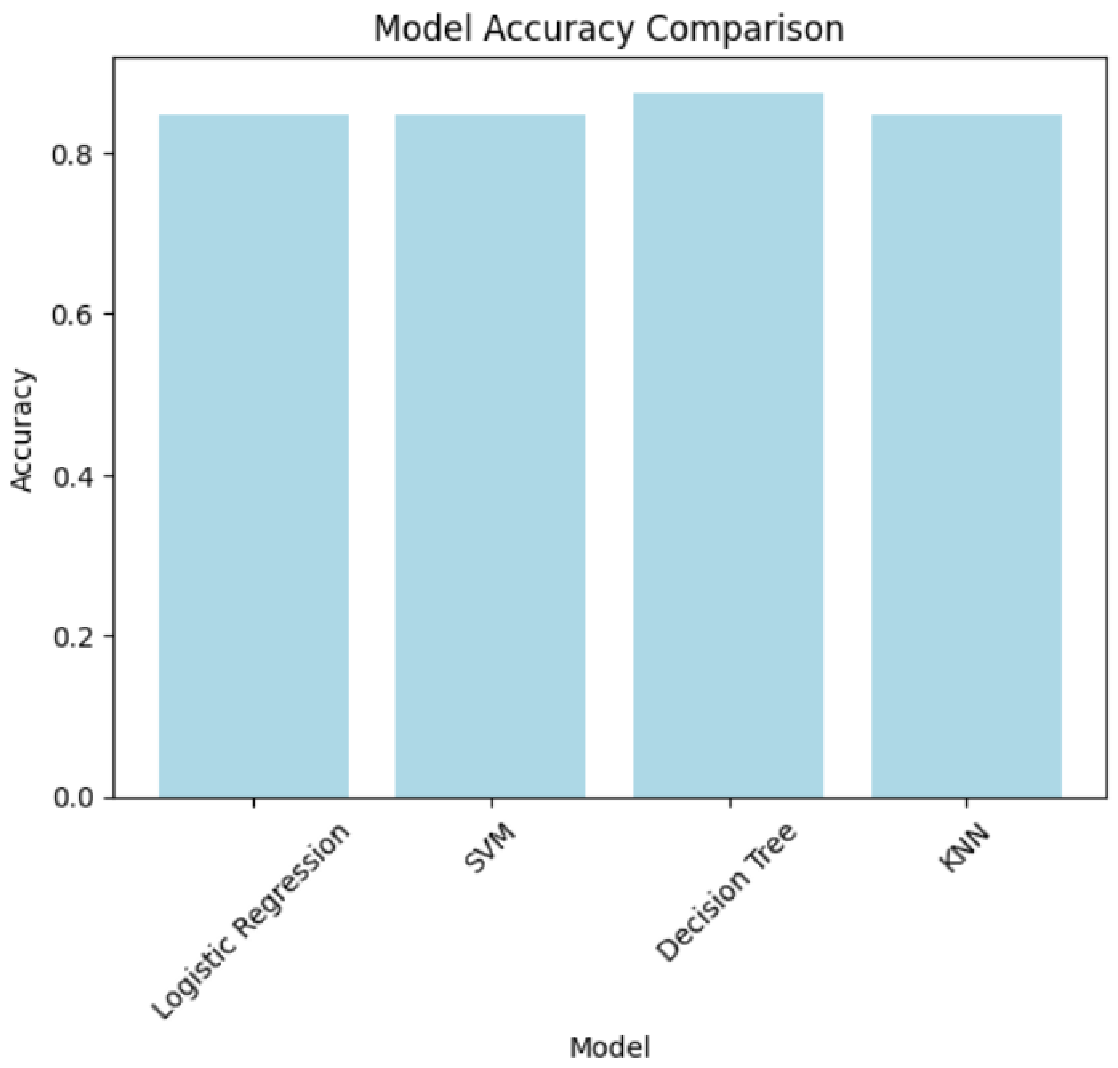
Payload vs. Launch Outcome scatter plot for all Sites



Section 5

Predictive Analysis (Classification)

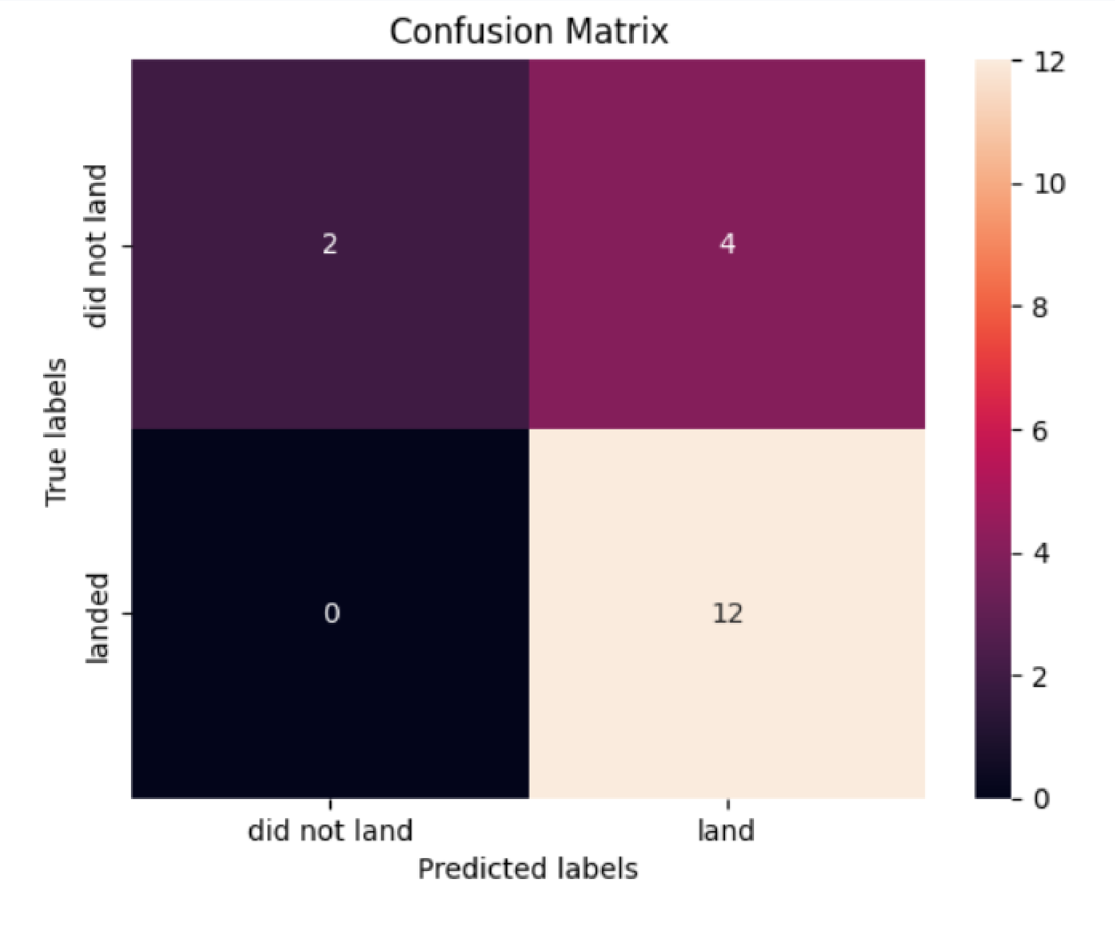
Classification Accuracy




```
Model with highest accuracy: Decision Tree with an accuracy of 0.8750
```

Confusion Matrix

- The model correctly predicted 12 successful landings true positives and 2 non-landings true negatives, but it falsely predicted 4 landings that did not occur false positives and had no missed landings false negatives.



Conclusions

- Our analysis successfully predicted Falcon 9 first-stage landings with high accuracy. By leveraging machine learning models and interactive data visualizations, we:
    - Identified key factors like **payload mass** and **launch site** that influence landing success.
    - Found **Tree Classifier** as the best-performing model with over **87.5% accuracy**.
    - This project provides a valuable tool for assessing launch costs, aiding decision-makers in the space industry to evaluate and optimize their strategies.

Thank you!