IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mohamed Salem BERRADA
09 /10 / 2025

Outline

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

Executive Summary

Data Analysis Methodology

The analysis of launch data employed a structured, three-phase approach:

1. **Data Acquisition:** Data was collected from public sources using **web scraping** techniques and the **SpaceX API**.
2. **Exploratory Data Analysis (EDA):** This phase involved **data wrangling** to clean and structure the data, followed by **data visualization** and **interactive visual analytics** to uncover patterns and relationships.
3. **Machine Learning (ML) Prediction:** A variety of machine learning models were applied to the processed data to predict launch outcomes.

Executive Summary

Key Findings and Results

Data Acquisition and Quality

- **Valuable data** relevant to launch characteristics and outcomes was successfully gathered from public sources (web scrapes and API calls).

Exploratory Data Analysis (EDA)

- The EDA phase was instrumental in **identifying the key features** (variables) that are most predictive of **successful launchings**.

Machine Learning Prediction

- The ML models identified the **optimal predictive model** for determining which launch characteristics are most important for driving mission success, leveraging the entirety of the collected data.

Introduction

Evaluation of Space Y's Viability Against Space X

This evaluation determines the competitive strength of **Space Y** by focusing on two key economic and operational pillars inherited from the Space X model:

1. **Cost Estimation (Reusability Model):**
   - **Deliverable:** Define the optimal method for estimating total launch costs by accurately predicting the success rate and frequency of **first-stage rocket landings**. This is critical for maximizing savings from booster reusability.
2. **Launch Location Strategy:**
   - **Deliverable:** Identify the **best geographical location** for Space Y's launch operations, balancing factors like orbital access, efficiency (Earth's rotation boost), and downrange safety to ensure maximum cost-effectiveness.

Section 1

Methodology

Methodology

Data Collection and Processing Methodology

1. **Data Acquisition:** Launch data was sourced from the **SpaceX API** (https://api.spacexdata.com/v4/rockets/) and supplemented with historical information obtained through **web scraping** of the **List of Falcon 9 and Falcon Heavy launches** (Wikipedia).
2. **Data Wrangling:** The combined dataset underwent a rigorous cleaning and structuring process.
3. **Feature Engineering:** A key enhancement was the creation of a definitive **landing outcome label**, which was derived through summary and analysis of existing raw outcome features.

Methodology

Predictive Analysis

- **Models:** Used four distinct classification models.
- **Data Prep:** Normalized data and split into training/test sets.
- **Evaluation:** Model **accuracy** was optimized by tuning and testing different parameter combinations.

Data Collection

- Launch data for SpaceX was acquired from two sources: the official SpaceX API (https://api.spacexdata.com/v4/rockets/) and the Wikipedia List of Falcon 9 and Falcon Heavy launches, via web scraping.

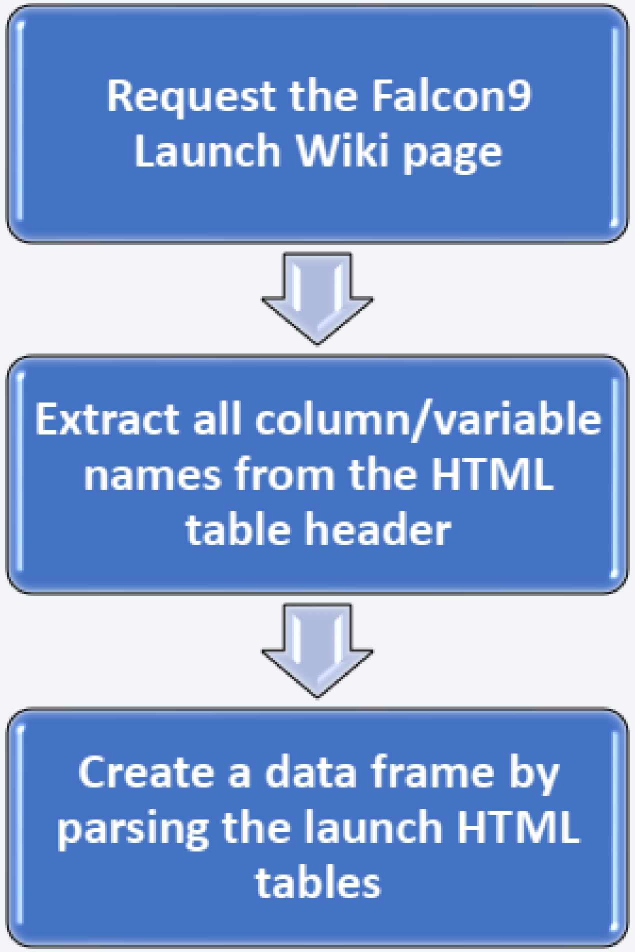
Data Collection – SpaceX API

- The public **SpaceX API** provided the data, which was retrieved via the process shown in the flowchart and then persisted.

Source code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- SpaceX launch data was also downloaded from Wikipedia using web scraping techniques, following the accompanying flowchart, and then persisted.



Source Code :
https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- EDA was performed, including summaries for launches per site and mission outcomes per orbit.
- The landing outcome label was created from the raw Outcome column.

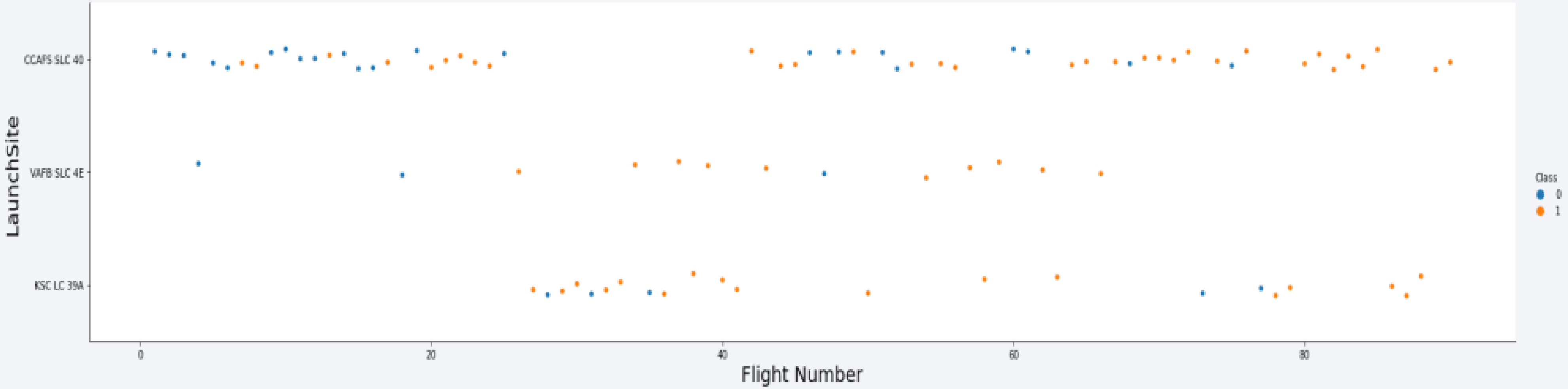
Source Code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Exploratory Data Analysis (EDA) was performed using **scatter plots** and **bar plots** to visualize the relationships between the following key feature pairs:

- Payload Mass vs. Flight Number
- Launch Site vs. Flight Number
- Launch Site vs. Payload Mass
- Orbit vs. Flight Number
- Payload Mass vs. Orbit

EDA with Data Visualization



Source code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-eda-dataviz.ipynb

EDA with SQL

The following ten analytical and statistical tasks were performed on the space mission database:

- **Identify** all **unique launch site names**.
- **Retrieve** the **top 5 launch sites** whose names start with 'CCA'.
- **Calculate** the **total payload mass** for missions launched by 'NASA (CRS)'.
- **Determine** the **average payload mass** carried by the 'F9 v1.1' booster version.
- **Find** the **date of the first successful landing** on a 'ground pad'.
- **List** the **names of successful boosters** with a 'drone ship' landing outcome and a payload mass between 4,000 and 6,000 kg.

EDA with SQL

The following ten analytical and statistical tasks were performed on the space mission database:

- **Count** the **total number of successful and failure mission outcomes**.
- **Identify** the **booster versions** that carried the **maximum payload mass**.
- **Retrieve** the **booster versions and launch sites** for all 'Failure (drone ship)' landing outcomes in the year **2015**.
- **Rank** the **counts of all landing outcomes** (e.g., 'Failure (drone ship)', 'Success (ground pad)') within the date range of 2010-06-04 to 2017-03-20.

Source Code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Folium Maps were utilized with the following elements to present geographical data:

- **Markers:** Used to pinpoint specific locations, such as **launch sites**.
- **Circles:** Highlighted defined **areas around coordinates**, like the NASA Johnson Space Center.
- **Marker Clusters:** Grouped multiple **events (launches)** occurring at the same coordinate (e.g., a single launch site).
- **Lines:** Indicated the **distance between two coordinates**.

Source Code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

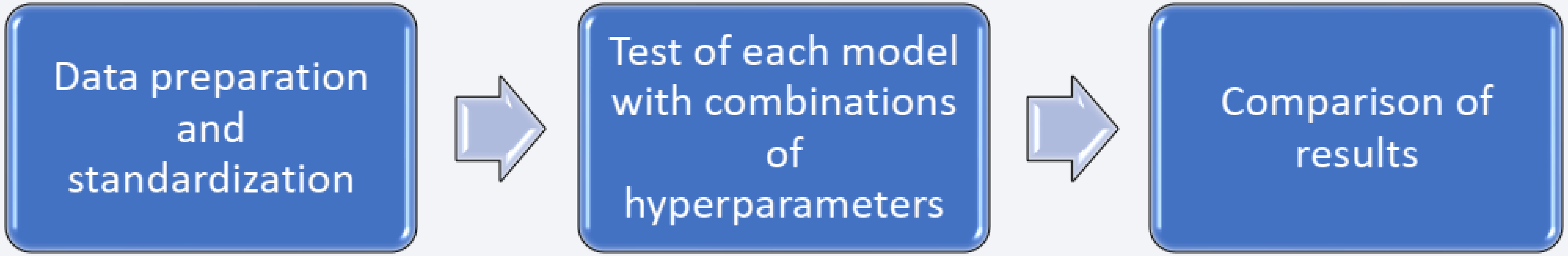
Data visualization centered on two key plots:

- **Percentage of Launches by Site**
- **Payload Mass Range Distribution**
- This combination of visualizations allowed for a rapid analysis of the relationship between **payload mass** and **launch sites**, thereby facilitating the determination of the **optimal launch site** based on mission payload requirements.

Source Code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/sapcex-dash-app.py

Predictive Analysis (Classification)

- Four classification models were selected and compared: **Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)**



Source Code : https://github.com/BMSalem/IBM-Applied-Data-Science-Capstone/blob/main/jupyter-lab-SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

Key Findings from Exploratory Data Analysis

Operational Overview

- **Launch Sites:** SpaceX utilizes a total of **four different launch sites**.
- **Early Missions:** The company's initial launches were primarily dedicated to internal SpaceX needs and contracts with **NASA**.
- **Mission Success:** The overall mission success rate is exceptionally high, approaching **100% of all mission outcomes**.

Results

Key Findings from Exploratory Data Analysis

Booster Performance and Reusability

- **Average Payload:** The average payload mass for the **F9 v1.1 booster** version is **2,928 kg**.
- **First Success:** The first successful landing outcome was achieved in **2015**, five years after the first launch attempt.
- **Drone Ship Success:** Numerous **Falcon 9 booster versions** have successfully landed on drone ships, typically carrying payloads **above the average mass**.
- **2015 Failures:** Two specific F9 v1.1 booster versions (**B1012** and **B1015**) recorded failed landings on drone ships during the year **2015**.
- **Reusability Trend:** The frequency and reliability of successful landing outcomes have **steadily improved** over time.

Results

Launch Site Geography and Infrastructure

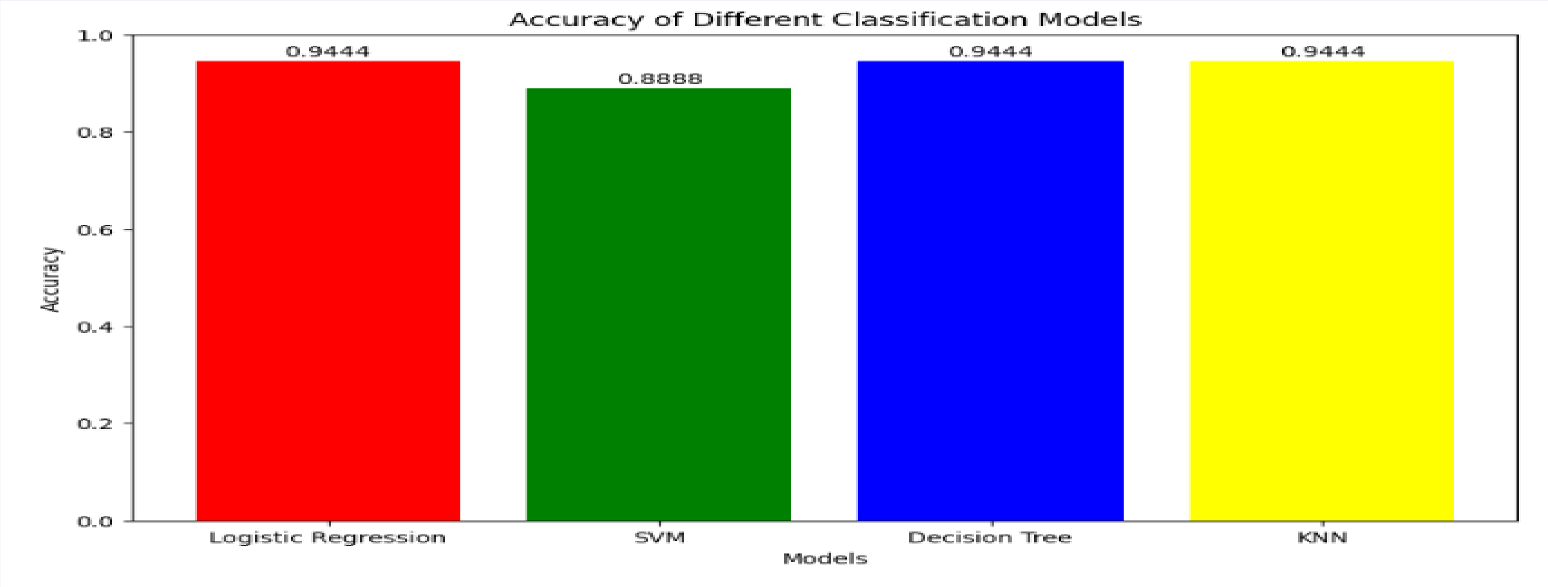
Interactive visual analytics confirmed that launch sites are strategically located in areas offering both **safety** and **strong logistical support**.

- **Safety & Location:** Sites are typically situated near the **sea** to ensure downrange safety over unpopulated areas.
- **Launch Preference:** The majority of launches are executed from **East Coast launch sites**.

Results

Predictive Analysis Results

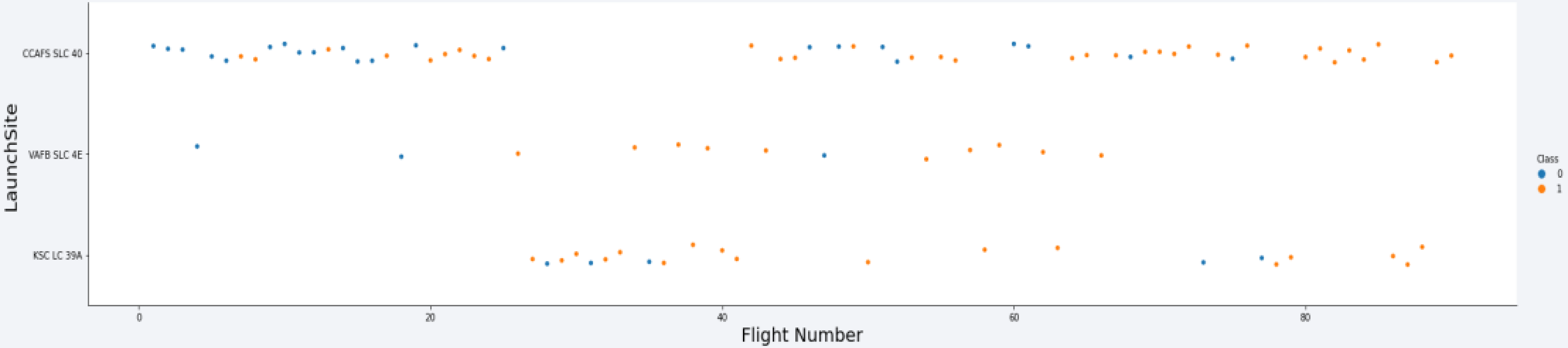
- The **SVM** was identified as the least-performing model for predicting successful rocket landings. It achieved a strong **accuracy over 83%** on the training data and an even higher **accuracy of over 88%** on the test data.



Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

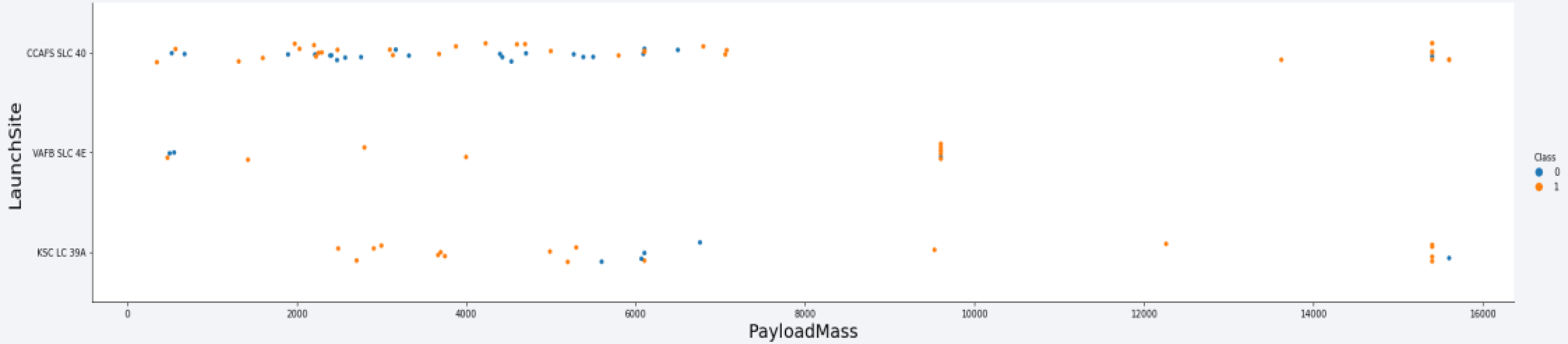


Analysis of recent launch data indicates a clear hierarchy of performance among launch sites:

- **1st Place: CCAF SLC 40** is currently the best site, having hosted the majority of recent successful launches.
- **2nd Place: VAFB SLC 4E**.
- **3rd Place: KSC LC 39A**.

Overall, the plot confirms a trend of **improving general success rate over time**.

Payload vs. Launch Site



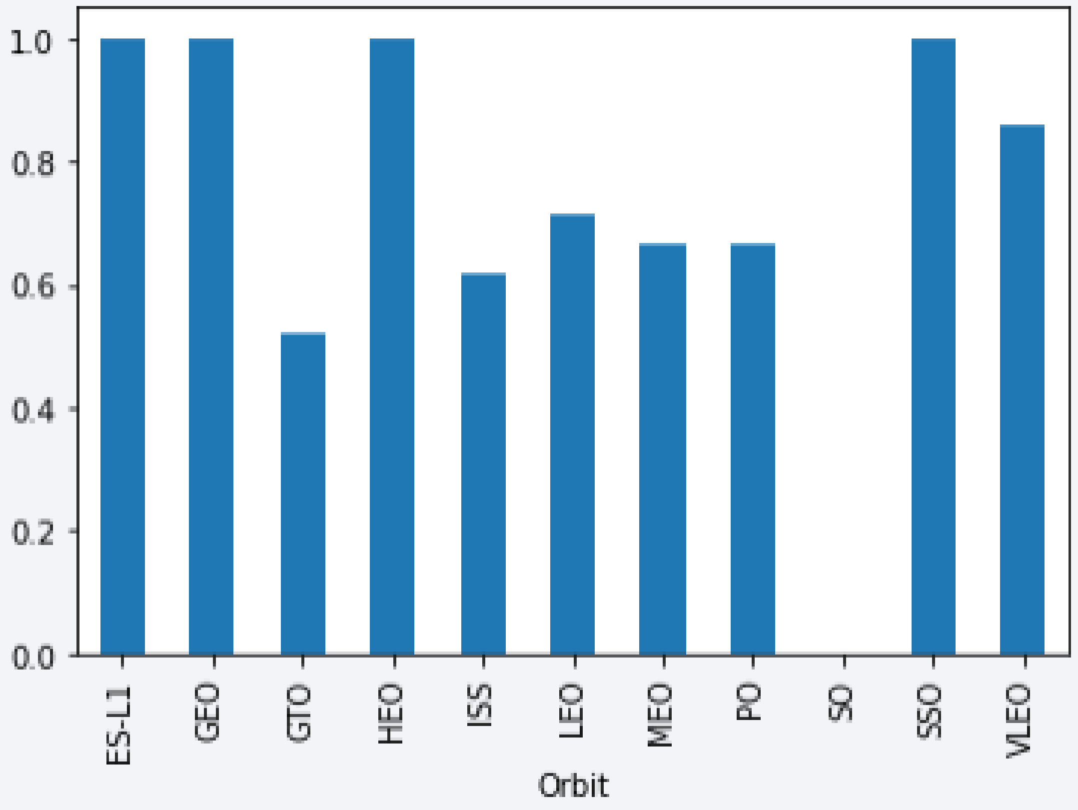
Payload Success and Site Limits

- Launches carrying payloads exceeding **9,000 kg** (a heavy lift class) demonstrate an **excellent success rate**.
- The heaviest payloads, those **over 12,000 kg**, appear to be restricted to only two launch sites: **CCAFS SLC 40** and **KSC LC 39A**.

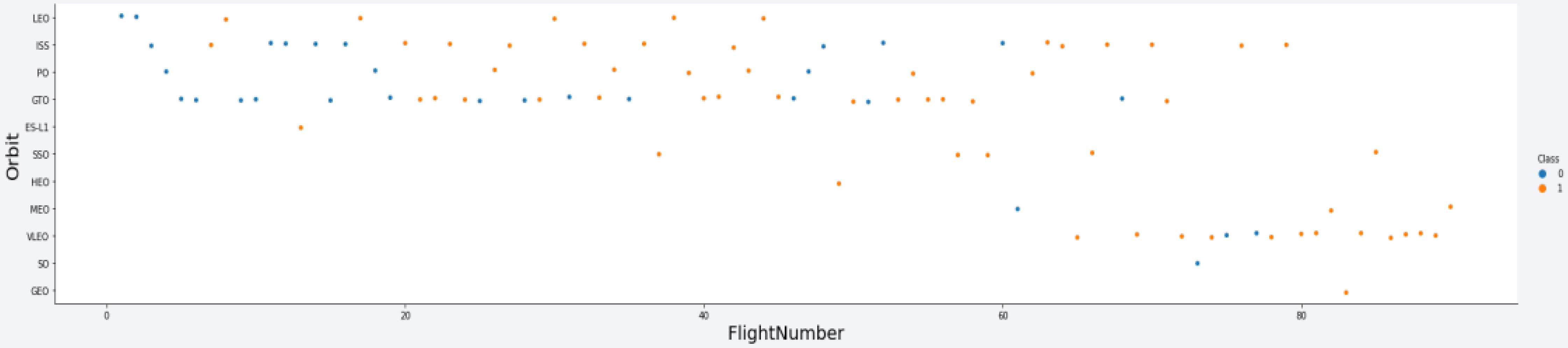
Success Rate vs. Orbit Type

Orbital Success Rates

- The highest mission success rates were observed for launches targeting the **ES-L1**, **GEO**, **HEO**, and **SSO** orbits.
- These were closely followed by:
- **VLEO** (above 80% success)
- **LFO** (above 70% success)



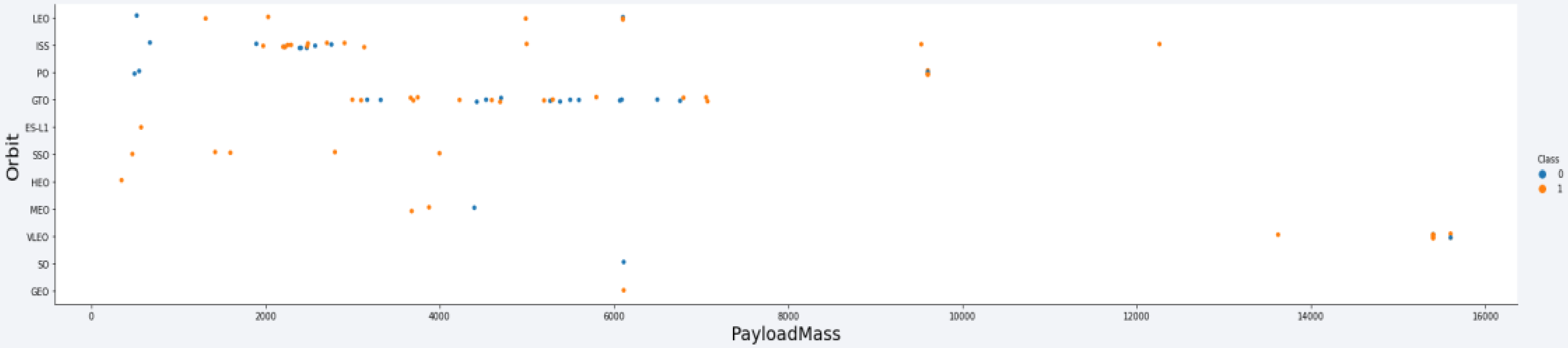
Flight Number vs. Orbit Type



Orbital Trends and Opportunities

- The **success rate** appears to have **improved over time** across all target orbits.
- The **VLEO** orbit represents a promising **new business opportunity** due to its recent and notable increase in launch frequency.

Payload vs. Orbit Type



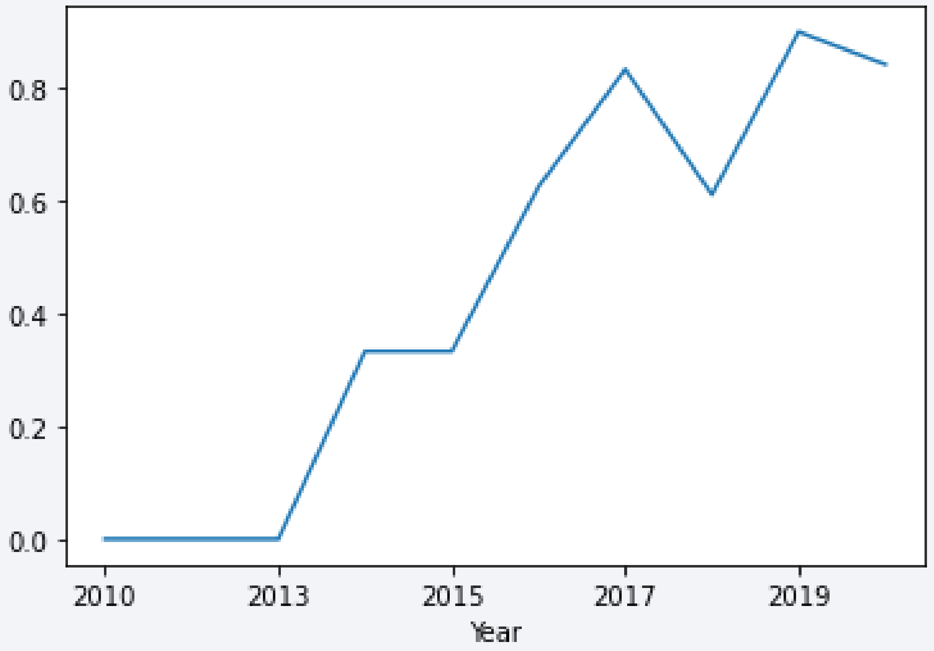
Orbit-Specific Payload and Success Observations

- **GTO Orbit:** No apparent correlation was found between **payload mass** and the mission's success rate for launches targeting the Geostationary Transfer Orbit (**GTO**).
- **ISS Orbit:** The International Space Station (**ISS**) orbit demonstrates both the **widest range of payloads** carried and a consistently **good success rate**.
- **Launch Volume:** The **SO** and **GEO** orbits have seen only a **few launches** overall.

Launch Success Yearly Trend

Success Rate Timeline

- The mission success rate began a sustained period of increase starting in **2013** and continued to improve through **2020**. This suggests that the initial three years of operation were primarily a period dedicated to **adjustments** and **technological refinement**.



All Launch Site Names

- According to data, there are four launch sites:

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

- They are obtained by selecting unique occurrences of “launch_site” values from the dataset.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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 attemp

- Here we can see five samples of Cape Canaveral launches.

Total Payload Mass

- Total payload carried by boosters from NASA:

Total Payload (kg)
111.268

- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2.928

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

Min Date
2015-12-22

- By filtering data by successful landing outcome on ground pad and getting the minimum value for date it’s possible to identify the first occurrence, that happened on 22 – 12 - 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes:

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- These are the boosters which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- The list above has the only two occurrences.

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

- Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20:

Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This view of data alerts us that “No attempt” must be taken in account.

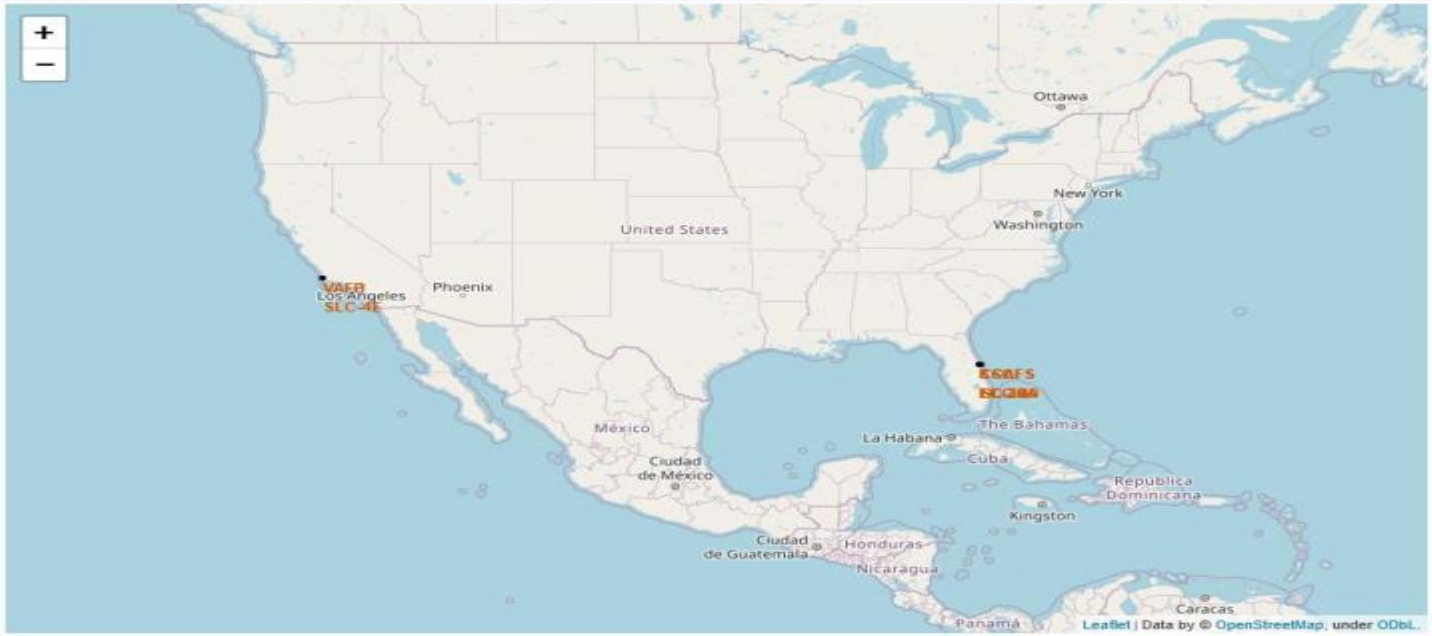
Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

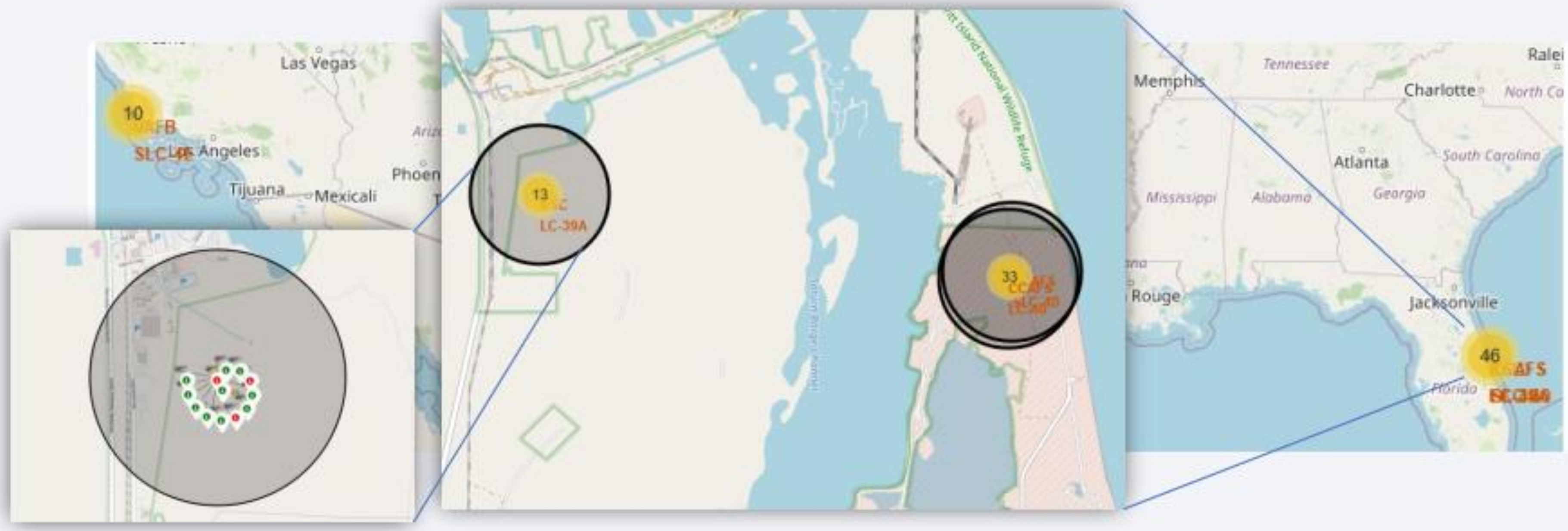
Launch Site Location Rationale

Launch sites are strategically located near the **sea**, primarily for **safety** (downrange clearance), yet remain close to essential **road and rail infrastructure** to ensure efficient logistics and transportation.



<Folium Map Screenshot 2>

- Example of KSC LC-39A launch site launch outcomes

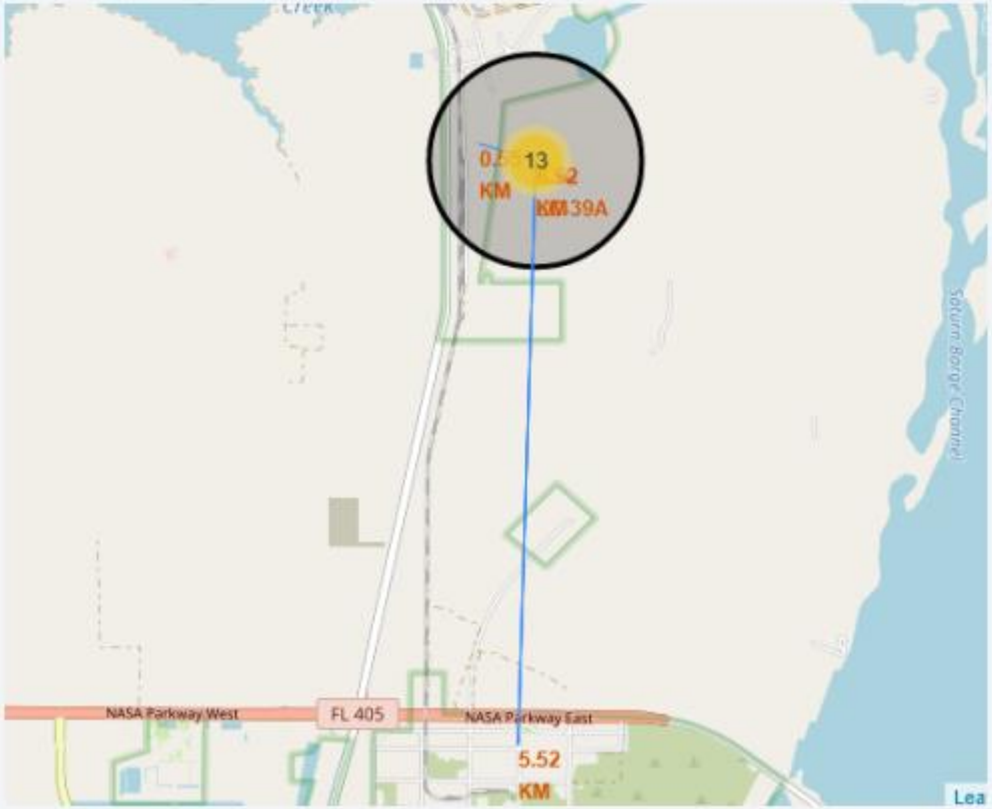


- Green markers indicate successful and red ones indicate failure.

<Folium Map Screenshot 3>

KSC LC-39A Launch Site Attributes

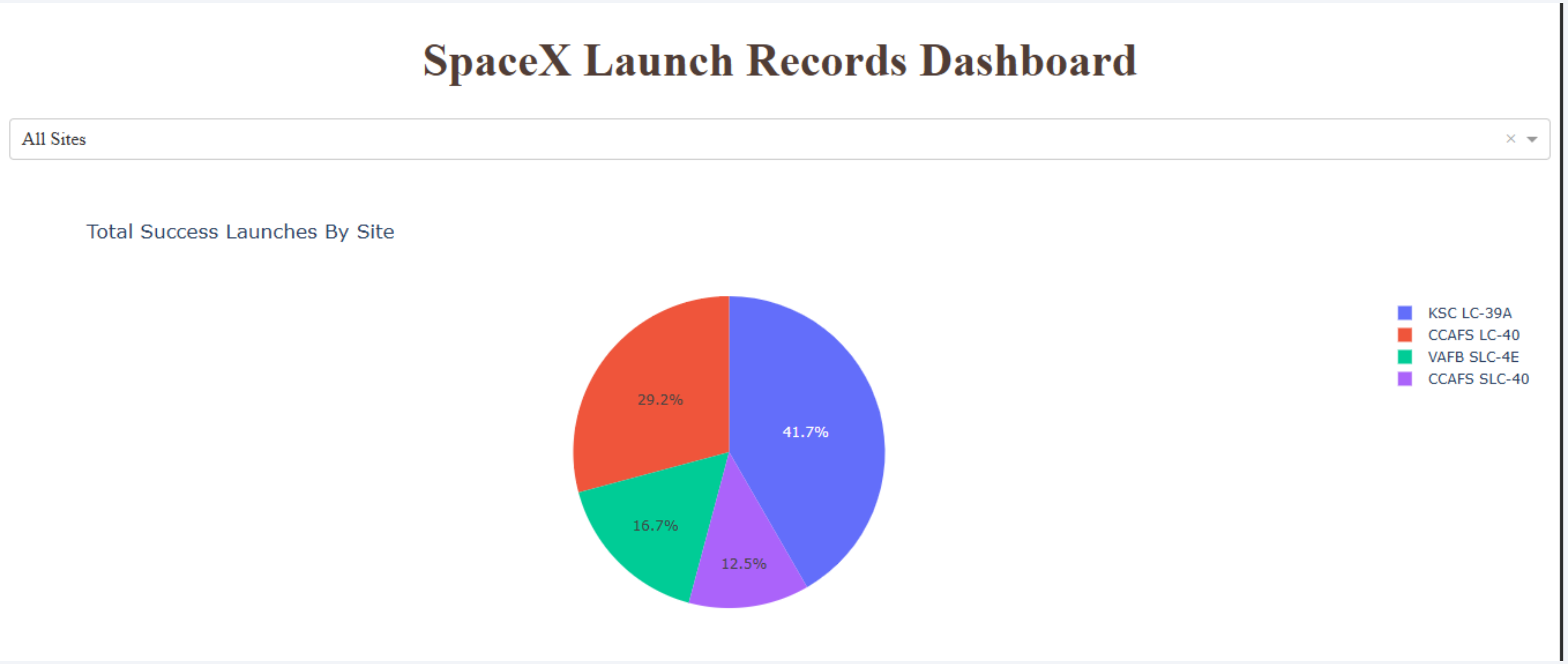
- Launch Complex 39A (KSC LC-39A) possesses strong logistical advantages, including proximity to **railroad and road infrastructure**, while maintaining necessary safety separation by being **relatively far from inhabited areas**.



Section 4

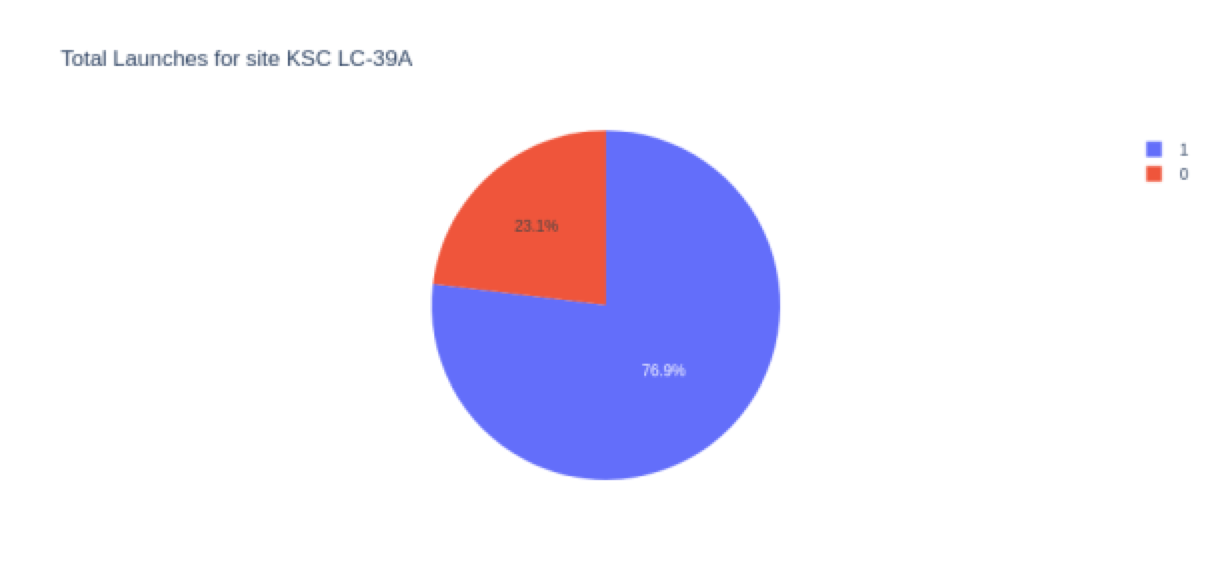
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



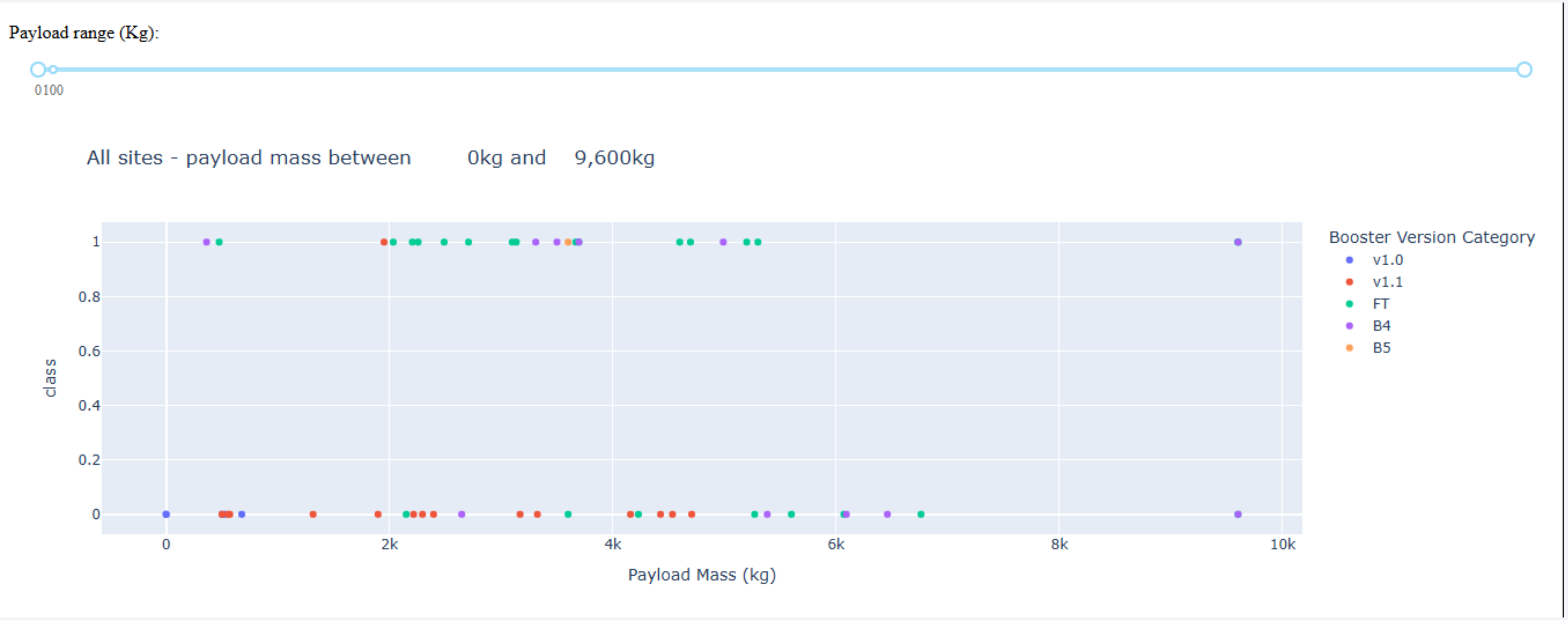
The **launch site location** is a critical factor strongly correlated with the overall **success of missions**.

<Dashboard Screenshot 2>



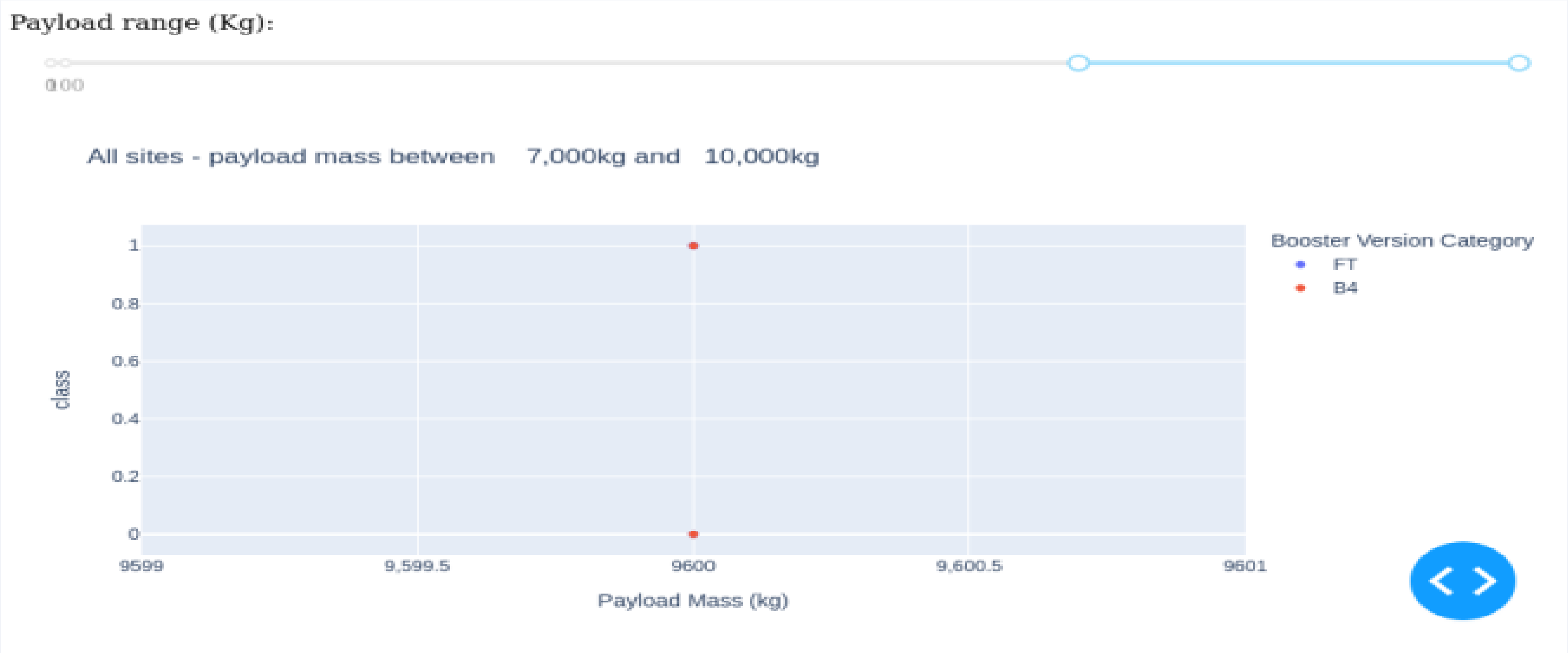
- 76.9% of launches are successful in this site.

<Dashboard Screenshot 3>



The combination of **Payloads under 6,000 kg** and the **FT (Full Thrust) boosters** represents the most successful pairing for missions.

<Dashboard Screenshot 3>



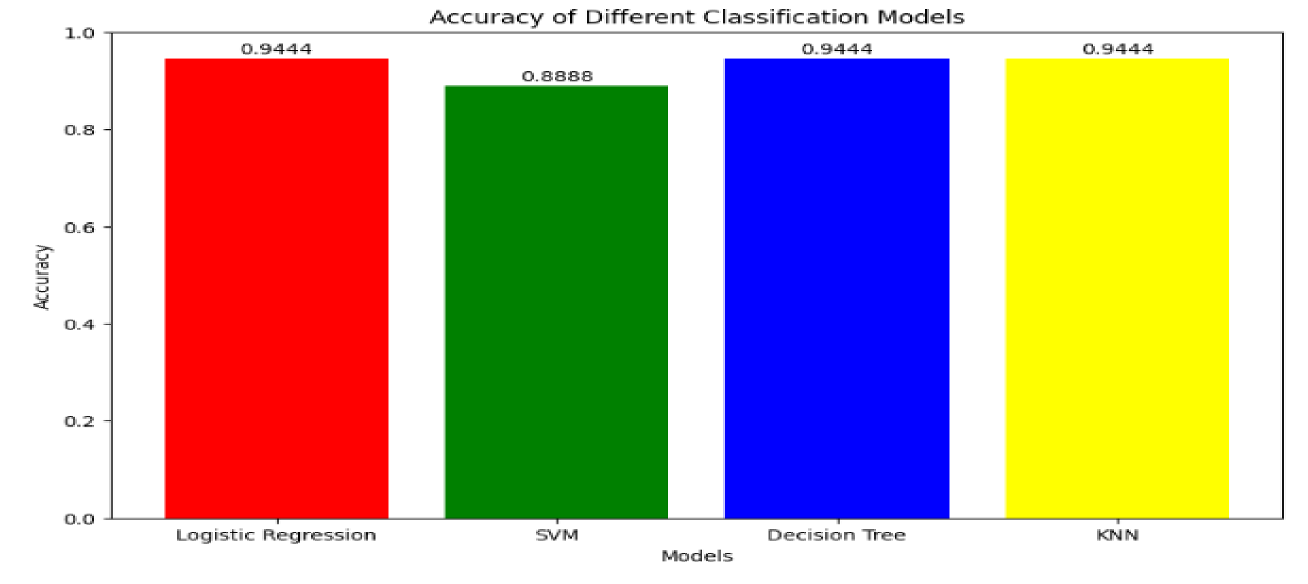
There’s not enough data to estimate risk of launches over 7,000kg

Section 5

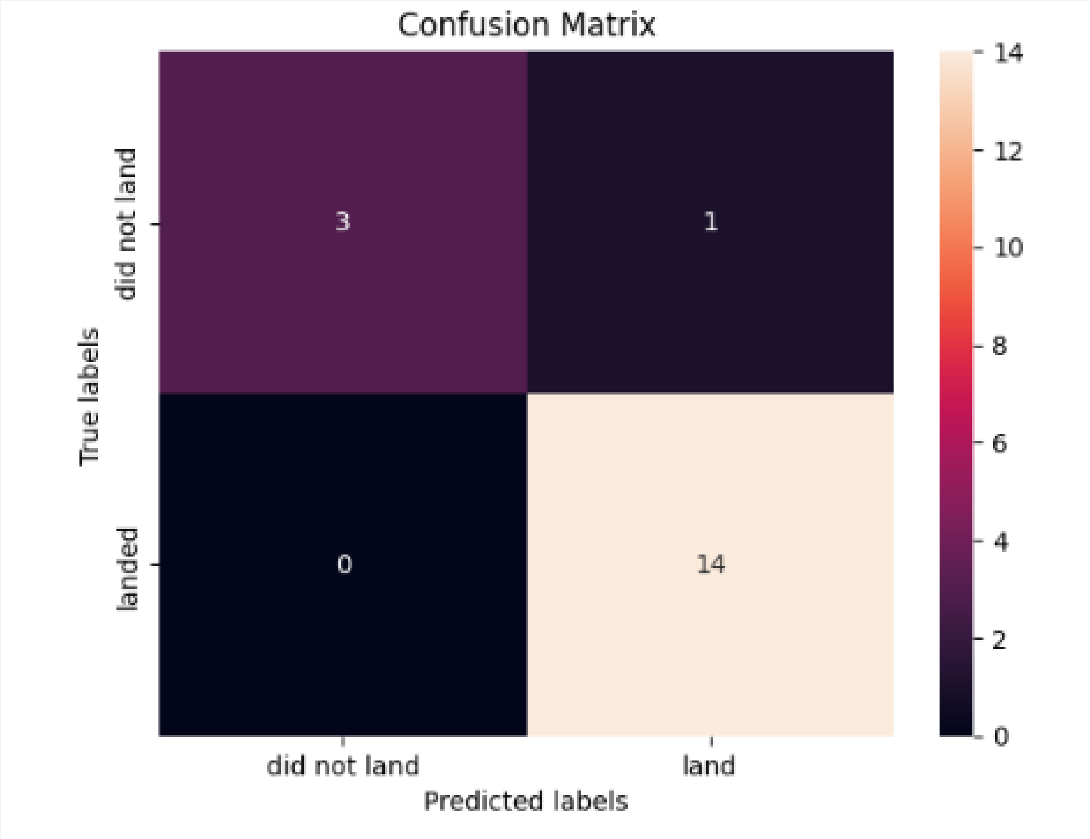
Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted beside;
- The model with the lowest classification accuracy is Decision Tree Classifier, which has accuracies over than 88%



Confusion Matrix



Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

Conclusions

Summary of Key Conclusions and Recommendations

- After analyzing and refining data from various sources, the following conclusions were reached:
- **Optimal Launch Site: KSC LC-39A** is identified as the best launch site overall.
- **Payload Risk:** Launches with payloads **above 7,000 kg** exhibit lower risk.
- **Reusability Trend:** While most mission outcomes are successful, the success rate of **rocket landings** continues to improve, reflecting positive evolution in processes and booster technology.
- **Predictive Model:** The **Decision Tree Classifier** is recommended for predicting successful landings, a capability that can be used to significantly **increase profitability**.

Appendix

Technical Notes and Limitations

- **Model Reproducibility:** To ensure the reproducibility of the predictive analysis, a fixed value should be set for the **np.random.seed** variable in future model tests.
- **Visualization Workaround:** Due to the inability of **Folium maps to render directly on GitHub**, screenshots were used as an alternative for map visualization in the final presentation.

Thank you!