Winning Space Race with Data Science

Varatharaj Kannan S
19-October-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
    - Utilized SpaceX public API and web scraping to gather data from Wikipedia
    - Data includes variables such as flight number, launch date, payload mass, orbit type, and mission outcomes
  - Data Preparation
    - Data wrangling was performed to clean and prepare the dataset.
    - Created a binary outcome variable representing success or failure of the first-stage rocket landing
  - Exploratory Data Analysis (EDA)
    - Visualized factors such as launch site, payload mass, and flight trajectory using static and interactive plots
    - SQL queries were used to compute key statistics like total payload and success rates.
  - Predictive Modeling
    - Built multiple machine learning models (LR, SVM, Decision tree, KNN) to predict landing success.
    - Model performance was evaluated using accuracy, precision, recall and F1 score
  - Data Visualization
    - Plotted geographical launch site proximity to coastlines, success rates, and payload ranges using Folium
    - Created an interactive dashboard to explore trends and patterns in launch outcomes using Plotly Dash

Executive Summary

- Summary of results
  - Key Insights:
    - Improved Success Over Time: SpaceX's landing success rate has steadily increased.
    - Launch Site Performance: KSC LC-39A was identified as the most successful launch site.
    - Orbit Type Analysis: Specific orbits (ES-L1, GEO, HEO, SSO) showed 100% success in landings.
    - Payload Impact: Payload mass influences landing outcomes, with certain ranges leading to higher success rates.
  - Model Performance:
    - All models performed similarly, with Logistic Regression, SVM, and KNN showing competitive results.
    - Achieved an accuracy rate of 85% in predicting landing outcomes
    - .Important features for prediction included launch site, orbit type, and payload mass.
  - Business Implications:
    - The models provide actionable insights for competitors in the space launch industry.
    - Competitors can use these predictions to make informed, cost-effective bids against SpaceX for rocket launch contracts.

Introduction

Project Background

SpaceX offers rocket launches at a cost of $62 million, significantly lower than competitors charging over $165 million. This cost reduction is largely due to their ability to reuse the first stage of the Falcon 9 rocket. We are tying to predict the success of the Falcon 9 first-stage landing using public data and machine learning models. Accurately predicting landing success can help estimate launch costs and give a competitive advantage to rival companies.

Key Problems to Solve

- Can We Predict Successful Landings?
- What Factors Influence Success? Which Model Performs Best?
- Can we analyze how success rates have improved?
- Analyze how landing success rates have improved over time, and which launch sites perform best.

Solutions Proposed

- Build models to predict if the Falcon 9 first stage will land successfully
- Explore how payload mass, launch site, orbit type, and number of flights affect landing success.
- Identify the best machine learning model (Logistic Regression, SVM, Decision Tree, KNN) for predicting landing success.
- Analyze how landing success rates have improved over time, and which launch sites perform best.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API (open source)
  - Web scraping Techniques (Wikipedia page ‘List of Falcon 9 and Falcon Heavy Launches)
- Perform data wrangling
  - Transforming categorical data using One Hot Encoding for ML algirithms, filtering unnecessary information and handling missing values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Models like Logistic Regression, SVM, Decision Tree, and KNN were built, with the data split into training and testing sets. The models were tuned using hyperparameter optimization techniques such as Grid Search, and their performance was evaluated using metrics like accuracy, F1 score, and cross-validation to select the best model for predicting outcomes.

Data Collection – SpaceX API

Request Data from SpaceX API:

- Used the SpaceX REST API to gather data on Falcon 9 launches.
- The API provides detailed information on rocket specifications, payloads, and landing outcomes.

Data Transformation:

- After making a GET request, the API returns data in JSON format.
- The JSON data is then normalized using json_normalize() to create a structured Pandas Dataframe for further analysis.

Custom Functions:

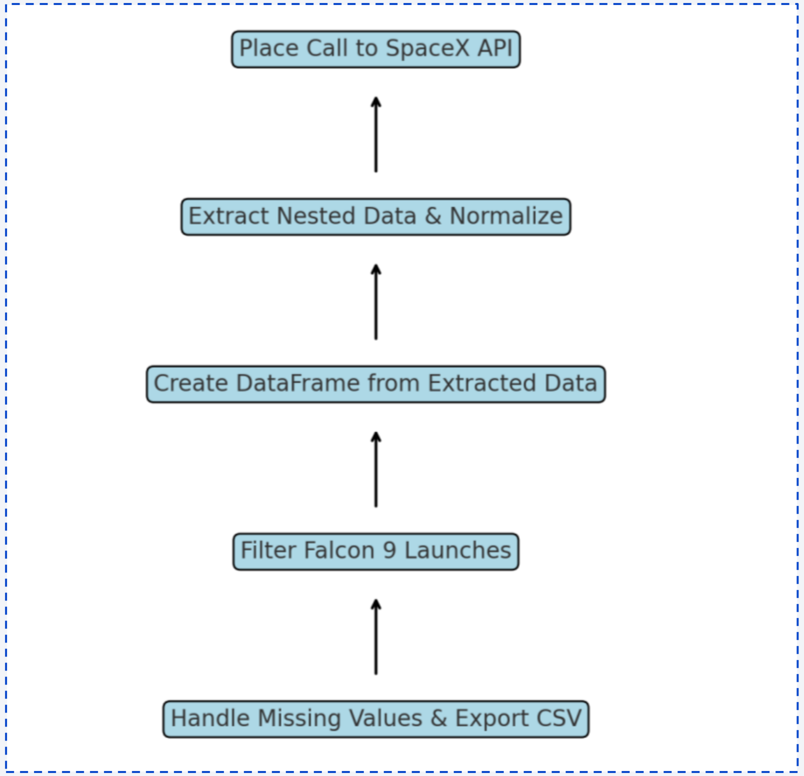
- Implemented functions to extract key launch details such as launch site, payload mass, and landing success.

Data Cleaning:

- Filtered the data to include only Falcon 9 launches.
- Missing values in columns like Payload Mass were handled by calculating and imputing the mean.

Export Process:

- The cleaned data was exported into CSV format for further analysis and modeling.



Github Link:
[Space X Data Collection API Notebook]

Data Collection – Web scraping from Wikipedia

Request HTML from Wikipedia:

- Utilized BeautifulSoup to scrape HTML tables from Wikipedia containing Falcon 9 launch data.

Data Parsing:

- Extracted relevant tables from the HTML response, identifying columns such as flight number, payload, launch site, and landing outcome

Data Cleaning:

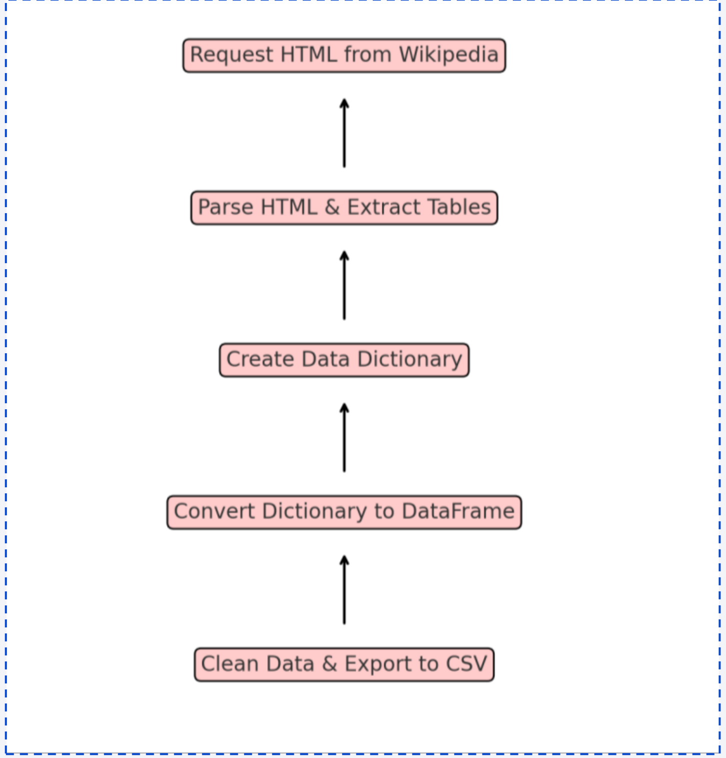
- Cleaned the extracted tables by removing irrelevant columns and rows.
- Created a dictionary from the clean data and assigned column names

Data Wrangling:

- Converted the dictionary to a Pandas DataFrame for structured analysis.
- Missing values were handled using appropriate methods such as mean imputation for numerical data.

Export Process:

- Exported the final cleaned data into CSV format for further analysis and modeling.



Github Link:
Space X Webscraping Notebook

Data Wrangling

Data Cleaning and Outcome Analysis

1. Handling Missing Data: Checked for missing values in columns like LandingPad, handled missing data appropriately.
2. Launch Site Analysis: Identified launch sites using value_counts() for LaunchSite.
3. Orbit Analysis: Calculated the number of occurrences for each orbit type from the Orbit column.
4. Mission Outcome Analysis: Determined landing outcomes using the Outcome column. Defined bad outcomes that include failed or no landing attempts.

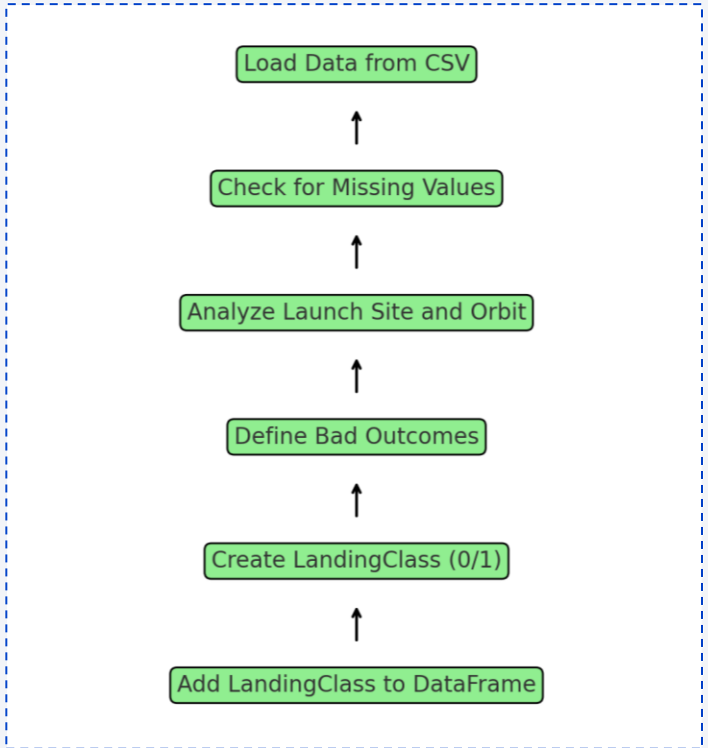
Labeling Mission Outcomes:

- Converted mission outcomes to a binary classification:
  - 1: Successful landings (e.g., "True ASDS", "True RTLS", "True Ocean").
  - 0: Failed or no landing (e.g., "False ASDS", "None ASDS", "False Ocean").
- Added the new LandingClass column to the dataset.

Final Results:

- Success Rate Calculation: The success rate of the landings was calculated using binary classifications.
  - Total number of landings: total_landings
  - Number of successful landings: successful_landings
  - Success Rate: success_rate

- Showcased the unique mission outcomes along with their corresponding LandingClass values to visualize landing success and failure patterns.



Github Link:
[Space X Data Wrangling Notebook]

EDA with Data Visualization

Visualizations and Their Purpose:

- Scatterplots:
  - Launch Site vs. Payload: Visualizes the distribution of payload mass across different launch sites
  - Launch Site vs. Flight Number: Shows the relationship between flight count and landing success at each launch site.
  - Orbit Type vs. Payload: Highlights the payload mass distribution across various orbit types.
  - Flight Number vs. Orbit: Reveals how flight numbers align with orbit types and mission outcomes.
- Bar Chart:
  - Orbit Type vs. Landing Success Rate: Compares success rates across different orbit types, illustrating their influence on landing outcomes.
- Line Plot:
  - Success Rate Over Time: Tracks improvements in Falcon 9 landing success rates over the years, reflecting technological advancements.

Analysis:

- Scatterplots: Help explore relationships between variables like launch site, orbit type, and payload, offering insights for predictive analysis.
- Bar Charts: Show categorical comparisons like success rates for different orbit types, making trends easier to interpret.
- Line Plot: Demonstrates temporal trends, highlighting the improvements in landing success over time.

Github Link:
[Space X - Data Visualization with EDA]

EDA with SQL

SQL Queries Performed:

Display:

- Unique launch site names.
- Five records where the launch site begins with 'CCA'.
- Total payload mass carried by boosters launched by NASA.
- Average payload mass for booster version F9 v1.1

List:

- Date of the first successful ground pad landing.
- Boosters with successful drone ship landings and payload between 4,000 and 6,000 kg.
- Total successful and failed missions.
- Booster versions with the maximum payload.
- Failed landings on drone ships for the year 2015.
- Count of landing outcomes (success/failure) between 2010-06-04 and 2017-03-20.

Github Link:
[SpaceX - SQL with EDA]

Build an Interactive Map with Folium

Map Objects Added to the Folium Map:

Markers: Added markers for launch sites and NASA Johnson Space Center to indicate important locations. The markers provide an easy way to visualize launch points on the map and offer additional information via pop-ups.

Circles: Created circles around each launch site to highlight specific regions, such as the NASA Johnson Space Center. Circles provide a visual cue for proximity and help differentiate the launch areas.

Lines (Polylines): Added lines connecting launch sites to nearby features like coastlines, rail lines, and perimeter roads. These lines illustrate distances to key geographical features and provide insight into the launch site's surroundings.

Why These Objects Were Added:

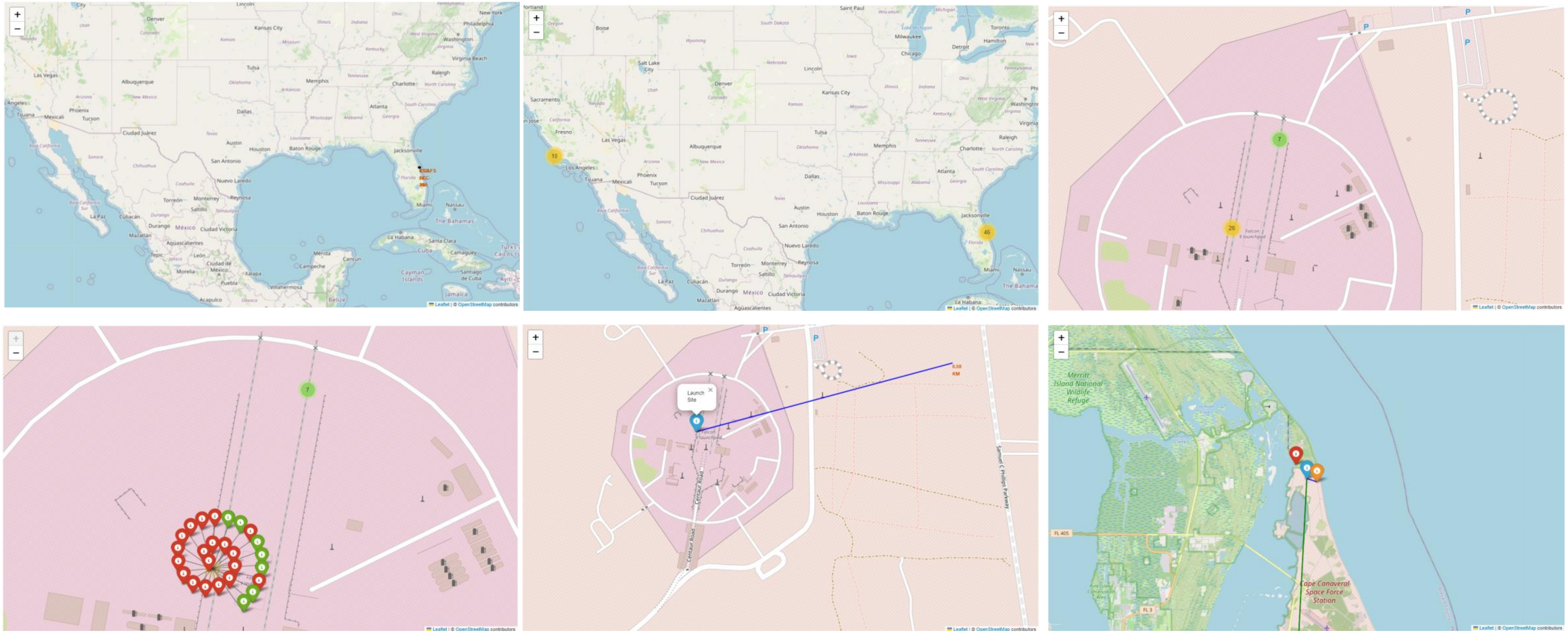
Markers provide a clear and interactive way to identify specific locations on the map. By clicking the markers, users can obtain additional details about each location, making it more engaging.

Circles help indicate the areas around launch sites and make it easier to focus on each launch site's region.

Lines visually represent the distance between launch sites and important geographic features like railways or coastlines, helping users understand the site's logistics and infrastructure proximity.

Github Link:
SpaceX - Interactive Map with Folium

Build an Interactive Map with Folium - Samples



Build a Dashboard with Plotly Dash

Plots/Graphs and Interactions Added to the Dashboard

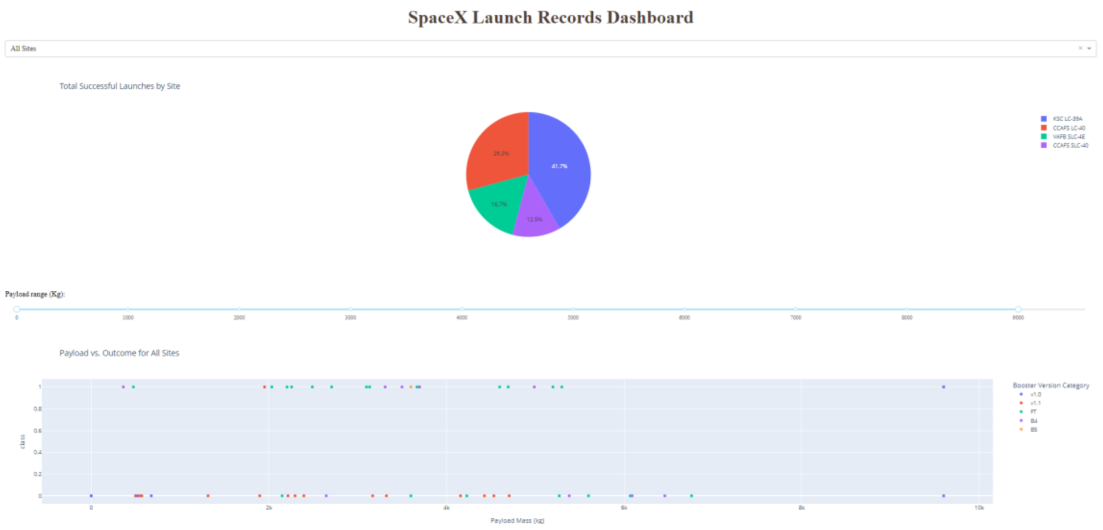
- Dropdown for Launch Site Selection:
    - This allows users to filter the data by different launch sites or view data from all sites. By selecting a specific site, users can see the launch outcomes relevant to that location.
- Pie Chart (Total Successful Launches by Site):
    - Displays the success rates across all launch sites or for a specific site. This gives an overview of how well launches have performed in terms of success and failure.
- Slider for Payload Range Selection:
    - Allows users to filter the data by payload mass. Users can explore how different payload sizes are associated with successful and failed launches.
- Scatter Plot (Payload vs. Success):
    - This graph shows the relationship between payload mass and the outcome (success or failure) for each launch. It also distinguishes between booster versions to add context.

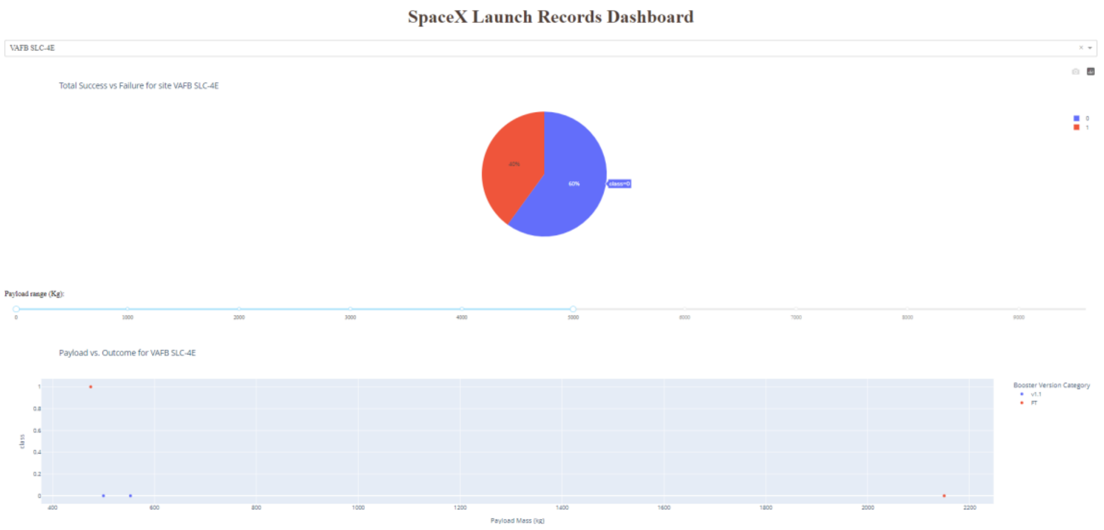
Reason for Adding These Components

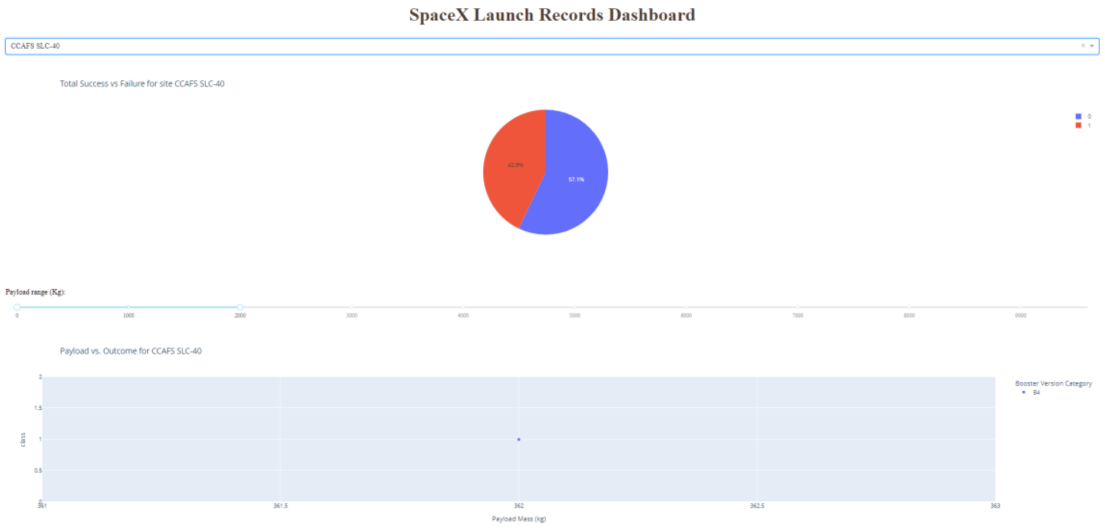
- Dropdown:
  To enable a user-friendly way of exploring different launch sites and their associated data without cluttering the view.
- Pie Chart:
  Provides a visual summary of the overall success and failure rates across different launch sites, allowing for easy comparison.
- Slider:
  Allows dynamic filtering of payload data so users can see how payload mass affects the success rate of launches.
- Scatter Plot:
  Helps to identify trends and correlations between payload mass and launch outcomes, providing deeper insights into the data.

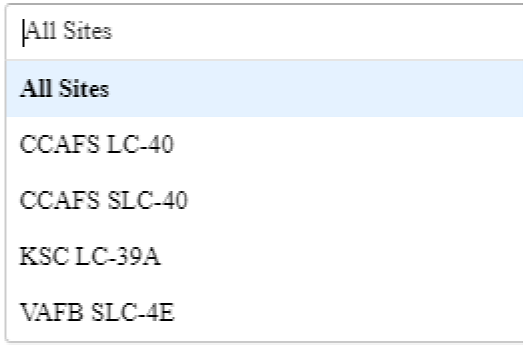
Github Link:
[SpaceX - Interactive Map with Folium]

Build a Dashboard with Plotly Dash - Samples









Predictive Analysis (Classification)

- Create a Numpy array from the column Class in data
- Standardize the data in X. train_test_split to split the data X and Y
- Create logistic regression object, train and test
- Create support vector machine object, train and test
- Create decision tree classifier object, train and test
- Create KNN object, train and test
- Find maximum accuracy to determine which method performs best

S No	Steps	Activities
1	Data Preprocessing	Collect and clean the dataset; Handle missing values and standardize features; Create Numpy array from the Class column
2	Stadardisation	Standardise the data with StandardScaler. Fit and transform the data
3	Train – Test Split	Split the data into training and test sets (e.g., 80% training, 20% testing)
4	Model Selection	Logistic Regression, SVM, Decision Tree, and KNN are chosen as candidate models
5	Hyperparameter Tuning	Use GridSearchCV for each model to find the best parameters using cross-validation (e.g., 10-fold)
6	Model Training	Train each model on the training data using the best parameters from GridSearchCV
7	Model Evaluation	Evaluate each model on the test data by calculating accuracy using score() and other metrics like confusion matrix
8	Model Comparison	Compare the accuracy scores of all models and choose the best-performing model
9	Result Conclusion	The best model is selected, and conclusions are drawn based on performance metrics (Jaccard_Score, F1_Score, Accuracy

Github Link:
SpaceX - ML Prediction

https://github.com/varatharajs/SpaceX-Learning/blob/main/SpaceX_ML_Prediction.ipynb

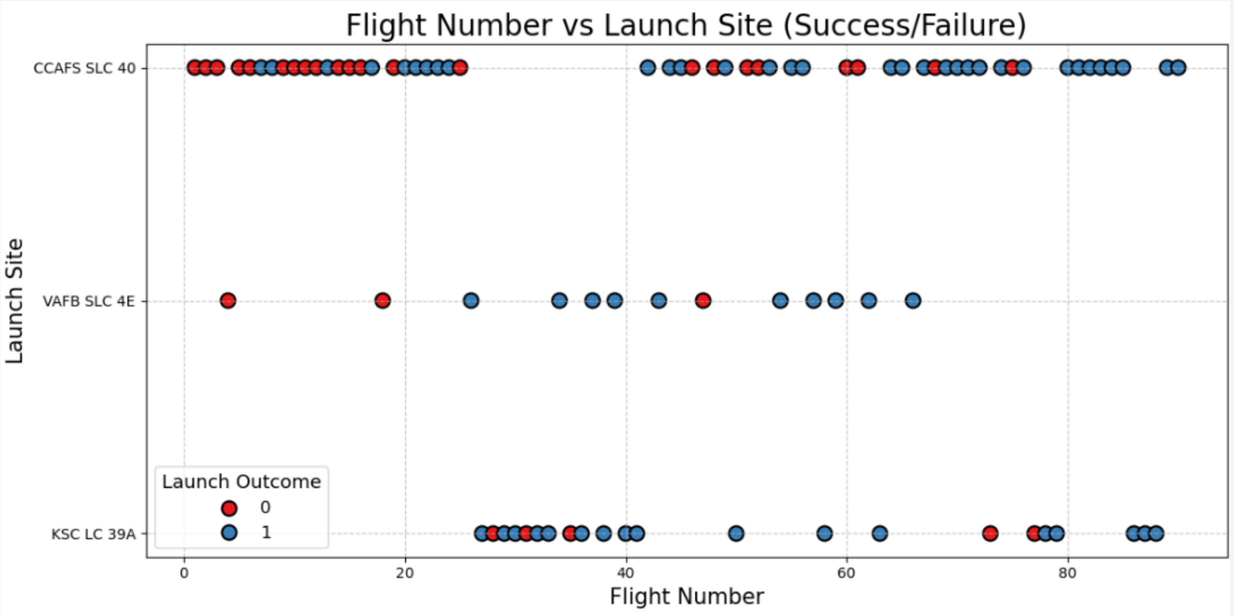
Results

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

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Launch Sites Overview:

- CCAFS SLC 40: Most frequently used.
- KSC LC 39A: Usage increases over time.
- VAFB SLC 4E: Fewer launches but high success rates.

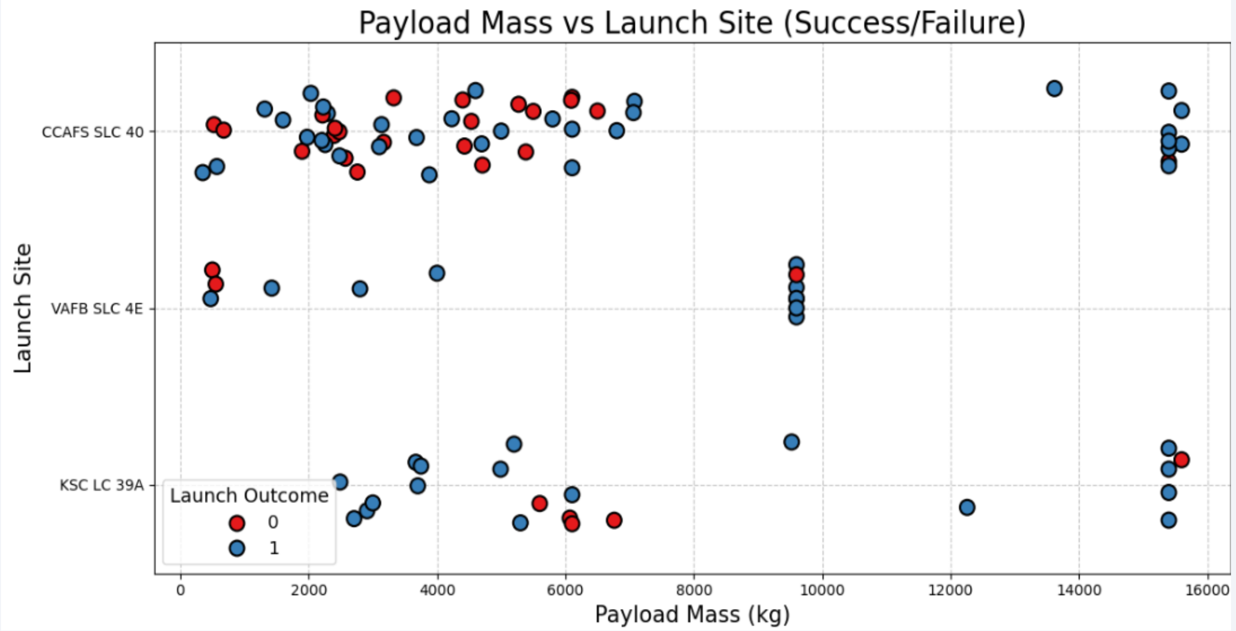
Success and Failure Distribution:

- Success rates improve over time, with early failures decreasing in later flights.
- CCAFS SLC 40
  - shows mixed results, with improvement in later flights.
- KSC LC 39A
  - exhibits a clear upward trend in success over time.
- VAFB SLC 4E
  - shows high performance despite fewer launches.

Conclusion:

SpaceX's success rates improve over time, with fewer failures in later flight numbers, especially at KSC LC 39A.

Payload vs. Launch Site

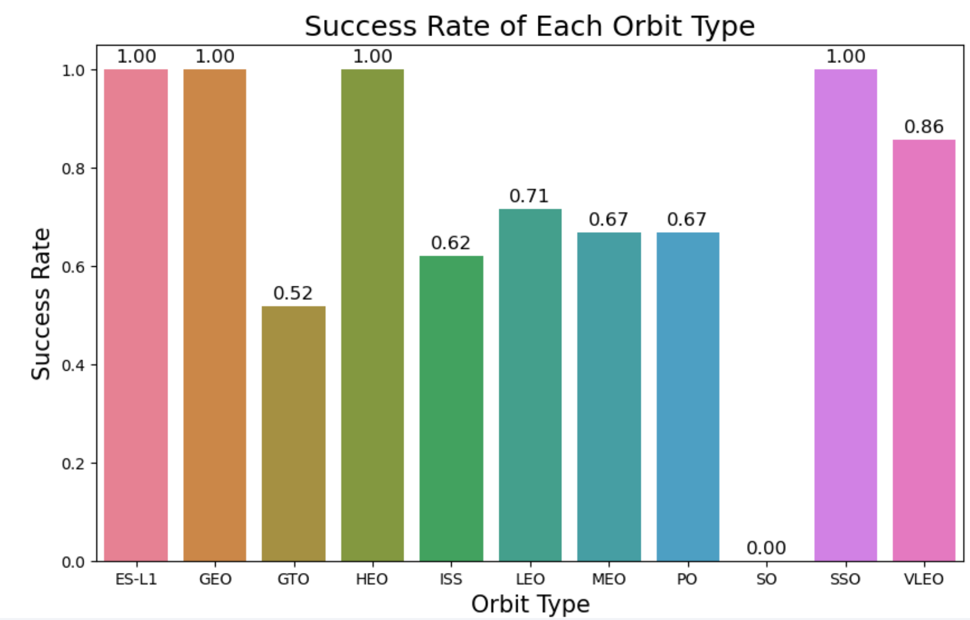


- CCAFS SLC 40:
  - This site has a wide range of payload masses, with payloads from 0 to 16,000 kg.
  - A mix of successes (blue) and failures (red) can be seen across the entire payload range.
  - Wide payload range (0–16,000 kg), with mixed success/failure for payloads below 6000 kg.
- VAFB SLC 4E:
  - Fewer launches, mostly successful for payloads below 4000 kg.
  - The majority of launches at this site were successful (blue), with very few failures.
- KSC LC 39A:
  - Larger payloads (4000–16,000 kg), showing increasing success rates for higher payloads, especially above 6000 kg

Summary:

- CCAFS SLC 40: Most frequently used with a wide payload range; success rate improves as payload mass increases.
- VAFB SLC 4E: Few launches, mostly successful, with payloads below 4000 kg.
- KSC LC 39A: Larger payloads, with a clear trend of higher success rates for larger payloads.

Success Rate vs. Orbit Type

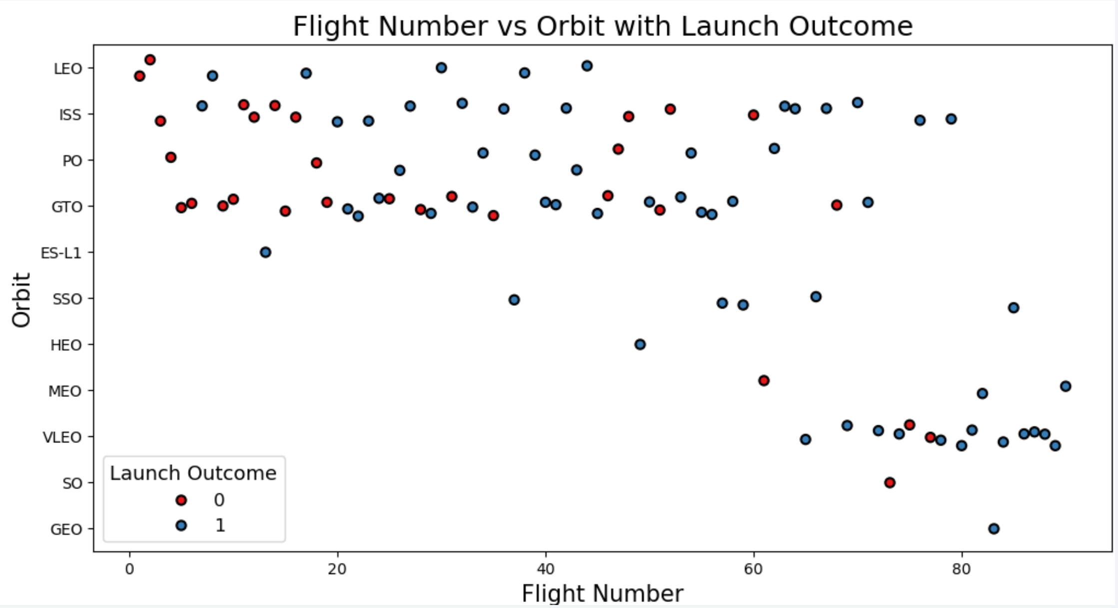


Summary

- SpaceX demonstrates strong success in most orbits, with perfect success in ES-L1, GEO, HEO, and SSO.
- GTO and SO require attention to improve mission reliability.

- Perfect Success (100%):
  - ES-L1, GEO, HEO, SSO orbits had a 100% success rate, indicating flawless performance for missions to these orbits.
- High Success Rates:
  - VLEO had an 86% success rate, showing strong reliability.
  - LEO showed 71% success, which is solid but with some room for improvement.
- Moderate Success Rates:
  - MEO and PO had 67% success, indicating reasonably consistent mission success with potential for improvement.
  - ISS had 62% success, which suggests the need for further refinement, especially for critical missions to the ISS.
- Low Success Rate:
  - GTO had a 52% success rate, indicating challenges in reaching this orbit. This is a key area for improvement.
- No Success:
  - SO had a 0% success rate, highlighting significant challenges or limited experience with missions to this orbit.

Flight Number vs. Orbit Type



Summary
The overall observation suggests that SpaceX has made substantial improvements in success rates, particularly for commonly used orbits like LEO and GTO.

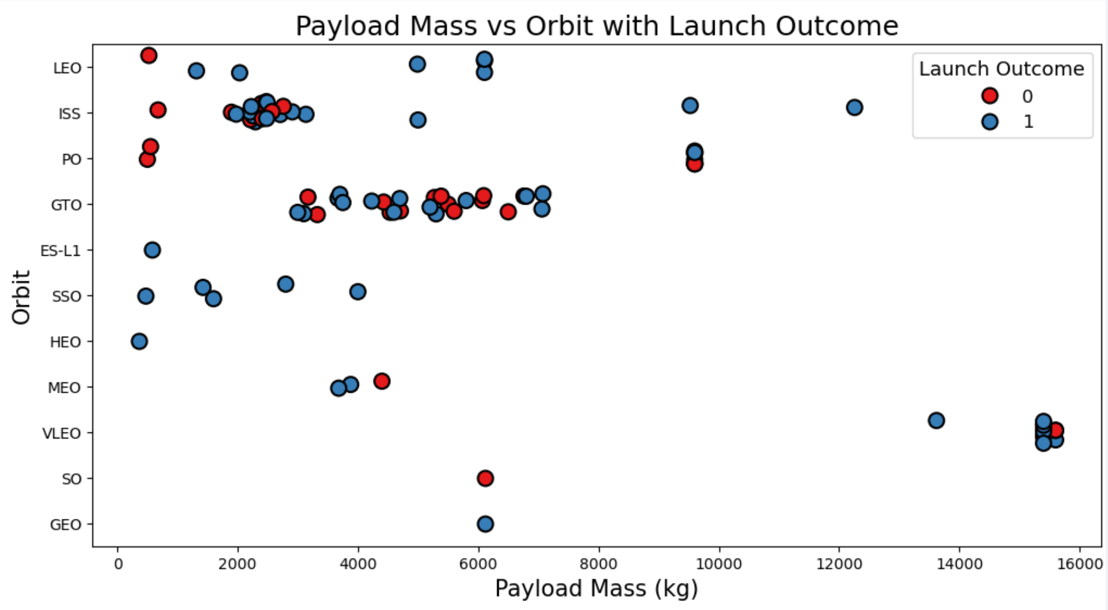
- LEO and ISS Orbits: Consistently used across a wide range of flight numbers. These orbits show a mix of successes and failures, with an increasing trend of success as flight numbers increase.

- GTO and PO Orbits: These orbits are also frequently used. GTO, in particular, shows a more even distribution of successes and failures over time.

- Other Orbits (SSO, VLEO, GEO, etc.): These orbits have fewer launches overall. Notably, VLEO has an excellent success rate, while SO has no successful launches.

- Trends Over Time: For most orbits, especially LEO and ISS, failures tend to be clustered toward earlier flight numbers, with a higher proportion of successes in later flights, indicating a learning curve and improvement over time.

Payload vs. Orbit Type



Key Insight:

- Higher payloads (above 6000 kg) seem to have a higher chance of success, particularly in certain orbit types like VLEO, SO, and GTO.
- Orbit type and payload mass do not consistently determine success, but certain orbits with higher payloads show more success (e.g., SSO, GTO).

LEO:

- Launches cover a wide range of payloads, with a mix of successes and failures.
- No clear pattern regarding payload size and launch success.

ISS:

- Similar to LEO, with mostly smaller payloads (under 6000 kg) and both success and failure outcomes.

PO & GTO:

- Both show a broad range of payloads from around 2000 to 6000 kg.
- The success rate is more evenly distributed between successes and failures.

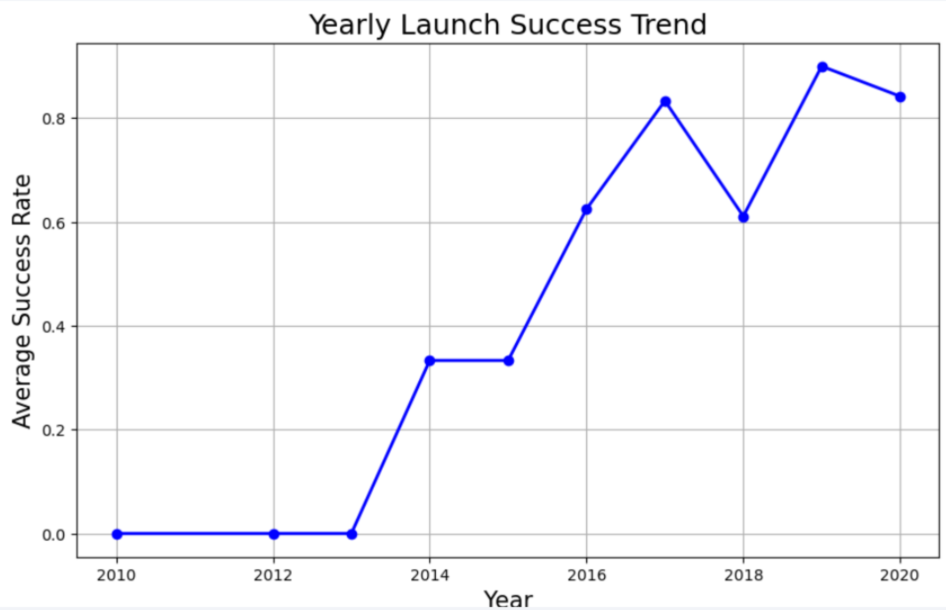
HEO & SSO:

- Fewer launches but high success rates, particularly with heavier payloads (5000 kg and above).

VLEO, GEO, and SO Orbits:

- Represent fewer launches with mixed results, especially at higher payload levels.

Launch Success Yearly Trend



Key Insight:

- Steady Improvement: The graph shows a clear trend of increasing launch success rates over time, with a few fluctuations.
- Recent Years: Since 2018, SpaceX has maintained a high success rate, nearing 90%.

Early Years (2010-2013):

- The success rate remains at 0% during the initial years, indicating no successful launches in this period.

2014:

- The success rate starts to rise, reaching around 40%. This marks the beginning of SpaceX's successful launches.

2015-2016:

- There is a steep increase in success rate, going from 40% to 80% by 2016, indicating a significant improvement in launch performance over these years.

2017:

- A temporary dip occurs, with the success rate dropping back to 60%, suggesting some launch failures during this period.

2018-2020:

- The success rate rises again, peaking at around 90% in 2019. In 2020, the success rate remains high, although with a slight decrease compared to 2019.

All Launch Site Names

- Unique Launch Sites:
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40

```
query_1 = 'SELECT DISTINCT "Launch_Site" FROM SPACEXTBL'
unique_launch_sites = run_query(query_1)
print(unique_launch_sites)
```

```
    Launch_Site
0   CCAFS LC-40
1   VAFB SLC-4E
2    KSC LC-39A
3  CCAFS SLC-40
```

Explanation:

- CCAFS LC-40 and CCAFS SLC-40 are two distinct launch sites at Cape Canaveral Air Force Station, with different designations (LC for Launch Complex, SLC for Space Launch Complex).
- VAFB SLC-4E is located at Vandenberg Air Force Base.
- KSC LC-39A refers to Launch Complex 39A at Kennedy Space Center, which has been used by SpaceX for several high-profile launches.
- These launch sites represent the primary locations SpaceX uses to launch their missions.

Launch Site Names Begin with 'CCA'

te	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	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 attempt

Explanation:

- The above table displays 5 records where the Launch Site starts with "CCA".
- All launches were conducted at CCAFS LC-40, with booster versions ranging from F9 v1.0 B0003 to B0007.
- The Mission Outcome for all 5 records was successful, though the Landing Outcome varied from failure to no attempt.

```
query_2 = '''
SELECT * FROM SPACEXTBL
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;
'''

cca_launch_sites = run_query(query_2)
print(cca_launch_sites)
```

```
         Date Time (UTC) Booster_Version  Launch_Site  \
0  2010-06-04   18:45:00  F9 v1.0  B0003  CCAFS LC-40
1  2010-12-08   15:43:00  F9 v1.0  B0004  CCAFS LC-40
2  2012-05-22    7:44:00  F9 v1.0  B0005  CCAFS LC-40
3  2012-10-08    0:35:00  F9 v1.0  B0006  CCAFS LC-40
4  2013-03-01   15:10:00  F9 v1.0  B0007  CCAFS LC-40

                                             Payload  PAYLOAD_MASS__KG_  \
0               Dragon Spacecraft Qualification Unit                  0
1  Dragon demo flight C1, two CubeSats, barrel of...                  0
2                              Dragon demo flight C2                525
3                                       SpaceX CRS-1                500
4                                       SpaceX CRS-2                677

       Orbit         Customer Mission_Outcome      Landing_Outcome
0        LEO           SpaceX         Success  Failure (parachute)
1  LEO (ISS)  NASA (COTS) NRO         Success  Failure (parachute)
2  LEO (ISS)      NASA (COTS)         Success           No attempt
3  LEO (ISS)       NASA (CRS)         Success           No attempt
4  LEO (ISS)       NASA (CRS)         Success           No attempt
```

Total Payload Mass

Total Payload Mass

48,213 kg

Explanation:

- The query calculates the total payload mass carried by boosters for NASA (CRS) missions.
- The total payload mass amounts to 48,213 kg.
- This gives an overview of how much payload mass SpaceX has delivered specifically for NASA's Cargo Resupply Services (CRS) missions.

```
query_3 = '''
SELECT SUM("Payload_Mass__kg_") AS Total_Payload_Mass
FROM SPACEXTBL
WHERE "Customer" LIKE '%NASA (CRS)%';
'''
nasa_payload_mass = run_query(query_3)
print(nasa_payload_mass)
```

```
   Total_Payload_Mass
0               48213
```

Average Payload Mass by F9 v1.1

Average Payload Mass

2,928.4 kg

Explanation:

- The query calculates the average payload mass carried by the booster version F9 v1.1.
- The average payload mass delivered by this booster is 2,928.4 kg, showcasing the payload capacity of this specific version of the Falcon 9 rocket during its missions.

```
query4 = '''
SELECT AVG("Payload_Mass__kg_") AS Average_Payload_Mass
FROM SPACEXTBL
WHERE "Booster_Version" = 'F9 v1.1';
'''

# Run the query and load the result into a DataFrame
average_payload_mass = run_query(query4)

# Display the result
print(average_payload_mass)
```

```
   Average_Payload_Mass
0                2928.4
```

First Successful Ground Landing Date

First successful landing outcome on ground pad

2015-12-22

This insight is significant because it marks a major milestone in SpaceX's mission to reuse rockets by successfully landing boosters back on ground pads, contributing to cost savings and technological advancement.

```
query5 = '''
SELECT MIN("Date") AS First_Successful_Landing
FROM SPACEXTBL
WHERE "Landing_Outcome" LIKE '%Success%ground%';
'''

# Run the query and load the result into a DataFrame
first_successful_landing = run_query(query5)

# Display the result
print(first_successful_landing)
```

```
  First_Successful_Landing
0               2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

#	Booster Version
1	F9 FT B1022
2	F9 FT B1026
3	F9 FT B1021.2
4	F9 FT B1031.2

These boosters show consistency in achieving successful drone ship landings with medium payloads, reinforcing their reliability for specific mission payload capacities.

```
query6 = '''
SELECT "Booster_Version"
FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Success (drone ship)'
AND "Payload_Mass__kg_" > 4000
AND "Payload_Mass__kg_" < 6000;
'''

# Run the query and load the result into a DataFrame
successful_drone_ship_boosters = run_query(query6)

# Display the result
print(successful_drone_ship_boosters)
```

```
  Booster_Version
0     F9 FT B1022
1     F9 FT B1026
2  F9 FT  B1021.2
3  F9 FT  B1031.2
```

Total Number of Successful and Failure Mission Outcomes

Landing_Outcome	Outcome_Count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

```
query7 = '''
SELECT "Landing_Outcome", COUNT(*) AS Outcome_Count
FROM SPACEXTBL
GROUP BY "Landing_Outcome";
'''

# Run the query and load the result into a DataFrame
mission_outcomes = run_query(query7)

# Display the result
print(mission_outcomes)
```

```
           Landing_Outcome  Outcome_Count
0       Controlled (ocean)              5
1                  Failure              3
2     Failure (drone ship)              5
3      Failure (parachute)              2
4               No attempt             21
5              No attempt               1
6   Precluded (drone ship)              1
7                  Success             38
8     Success (drone ship)             14
9     Success (ground pad)              9
10    Uncontrolled (ocean)              2
```

- There are 38 successful landings, 14 successful drone ship landings, and 9 successful ground pad landings, making a total of 61 successful missions.
- There are 3 failed landings, 5 drone ship failures, 2 parachute failures, and other failures, making a total of 10 failure outcomes.

Boosters Carried Maximum Payload

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

```
query8 = '''
SELECT "Booster_Version"
FROM SPACEXTBL
WHERE "Payload_Mass__kg_" = (
    SELECT MAX("Payload_Mass__kg_")
    FROM SPACEXTBL
);
'''

# Run the query and load the result into a DataFrame
max_payload_boosters = run_query(query8)

# Display the result
print(max_payload_boosters)
```

```
   Booster_Version
0    F9 B5 B1048.4
1    F9 B5 B1049.4
2    F9 B5 B1051.3
3    F9 B5 B1056.4
4    F9 B5 B1048.5
5    F9 B5 B1051.4
6    F9 B5 B1049.5
7   F9 B5 B1060.2
8   F9 B5 B1058.3
9    F9 B5 B1051.6
10   F9 B5 B1060.3
11  F9 B5 B1049.7
```

These booster versions successfully carried the highest recorded payload mass among all SpaceX launches. Each of these boosters has made significant contributions to the payload transportation capabilities of SpaceX.

2015 Launch Records

Month	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)

- In 2015, two landing failures occurred on drone ships. Both failures involved the booster version F9 v1.1 and took place at the CCAFS LC-40 launch site.
- These landings were intended to be on drone ships, but they did not succeed.

```
query9 = '''
SELECT
    CASE
        WHEN substr("Date", 6, 2) = '01' THEN 'January'
        WHEN substr("Date", 6, 2) = '02' THEN 'February'
        WHEN substr("Date", 6, 2) = '03' THEN 'March'
        WHEN substr("Date", 6, 2) = '04' THEN 'April'
        WHEN substr("Date", 6, 2) = '05' THEN 'May'
        WHEN substr("Date", 6, 2) = '06' THEN 'June'
        WHEN substr("Date", 6, 2) = '07' THEN 'July'
        WHEN substr("Date", 6, 2) = '08' THEN 'August'
        WHEN substr("Date", 6, 2) = '09' THEN 'September'
        WHEN substr("Date", 6, 2) = '10' THEN 'October'
        WHEN substr("Date", 6, 2) = '11' THEN 'November'
        WHEN substr("Date", 6, 2) = '12' THEN 'December'
    END AS Month_Name,
    "Booster_Version",
    "Launch_Site",
    "Landing_Outcome"
FROM SPACEXTBL
WHERE "Landing_Outcome" = 'Failure (drone ship)'
AND substr("Date", 0, 5) = '2015';
'''

# Run the query and load the result into a DataFrame
failure_drone_ship_2015 = run_query(query9)

# Display the result
print(failure_drone_ship_2015)
```

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

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

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

- The highest number of landing outcomes during this period was "No attempt" with 10 occurrences.
- Both "Success (drone ship)" and "Failure (drone ship)" occurred 5 times each, showing a balance between successes and failures on drone ships.
- "Success (ground pad)" had 3 occurrences, indicating successful landings on ground pads.

```
query10 = '''
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;
'''

# Run the query and load the result into a DataFrame
ranked_landing_outcomes = run_query(query10)

# Display the result
print(ranked_landing_outcomes)
```

```
          Landing_Outcome  Outcome_Count
0              No attempt             10
1    Success (drone ship)              5
2    Failure (drone ship)              5
3    Success (ground pad)              3
4      Controlled (ocean)              3
5    Uncontrolled (ocean)              2
6     Failure (parachute)              2
7  Precluded (drone ship)              1
```

Section 3

Launch Sites Proximities Analysis

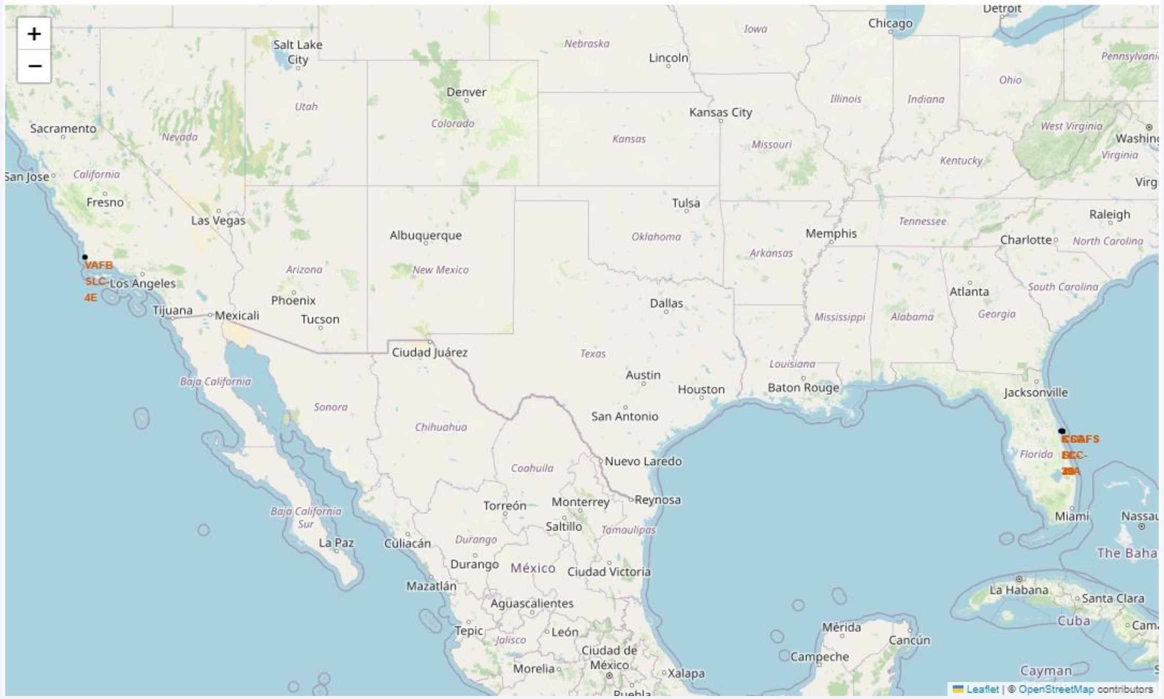
Launch Sites

Proximity to the Equator:

- The launch sites are not near the Equator. Vandenberg (34° N) and Cape Canaveral (28° N) are located at higher latitudes, showing that not all launch sites are close to the Equator.

Proximity to the Coast:

- All launch sites are near the coast (Pacific and Atlantic Oceans). Coastal locations are chosen for safety and to take advantage of the Earth's rotation for efficient launches.



Near Equator: the closer the launch site to the equator, the easier it is to launch to equatorial orbit, and the more help you get from Earth's rotation for a prograde orbit. Rockets launched from sites near the equator get an additional natural boost - due to the rotational speed of earth - that helps save the cost of putting in extra fuel and boosters.

Launch Outcomes

At Each Launch Site:

Green markers are successful launches

Red markers for unsuccessful launches

Launch site CCAFS SLC-40 has a 3/7 success rate (42.9%)

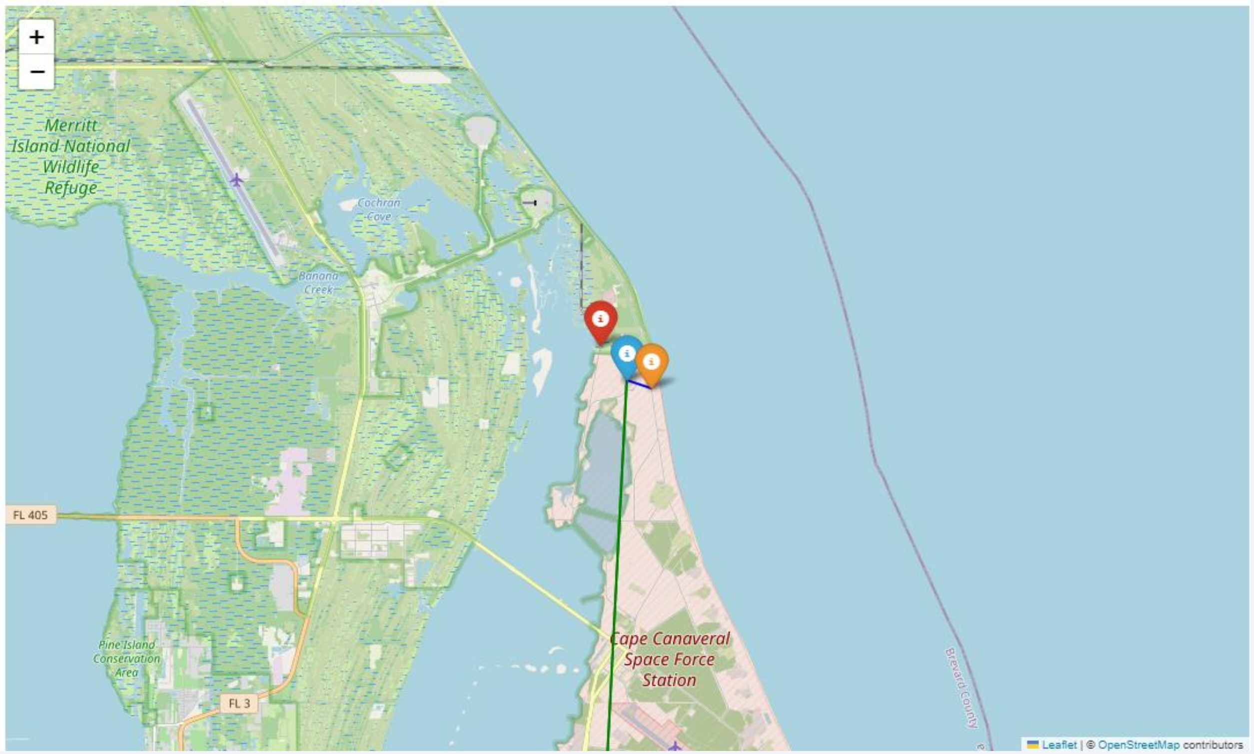
Distance to Proximities

CCAFS SLC-40

• Coasts: help ensure that spent stages dropped along the launch path or failed launches don't fall on people or property.

• Safety/ Security: Need to be an exclusion zone around the launch site to keep unauthorized people away and keep people safe.

• Transportation/ Infrastructure and Cities: need to be away from anything a failed launch can damage, but still close enough to roads/rails/docks to be able to bring people and material to or from it in support of launch activities.



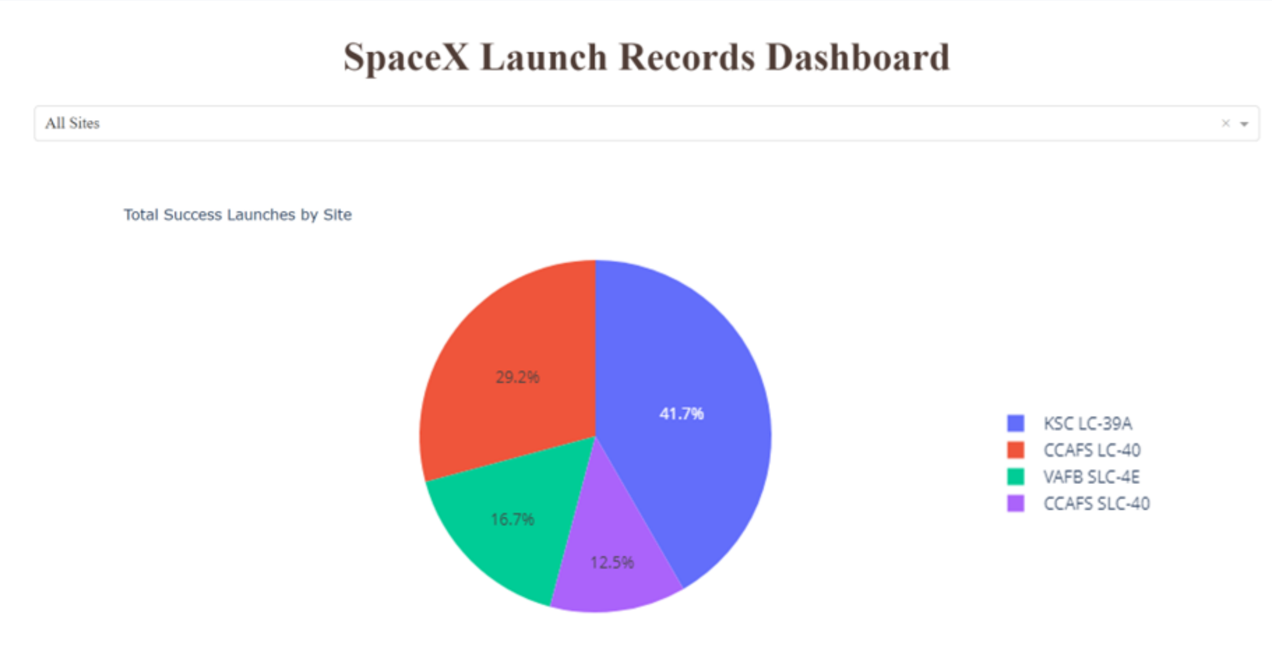
- .86 km from nearest coastline
- 21.96 km from nearest railway
- 23.23 km from nearest city
- 26.88 km from nearest highway

Section 4

Build a Dashboard with Plotly Dash

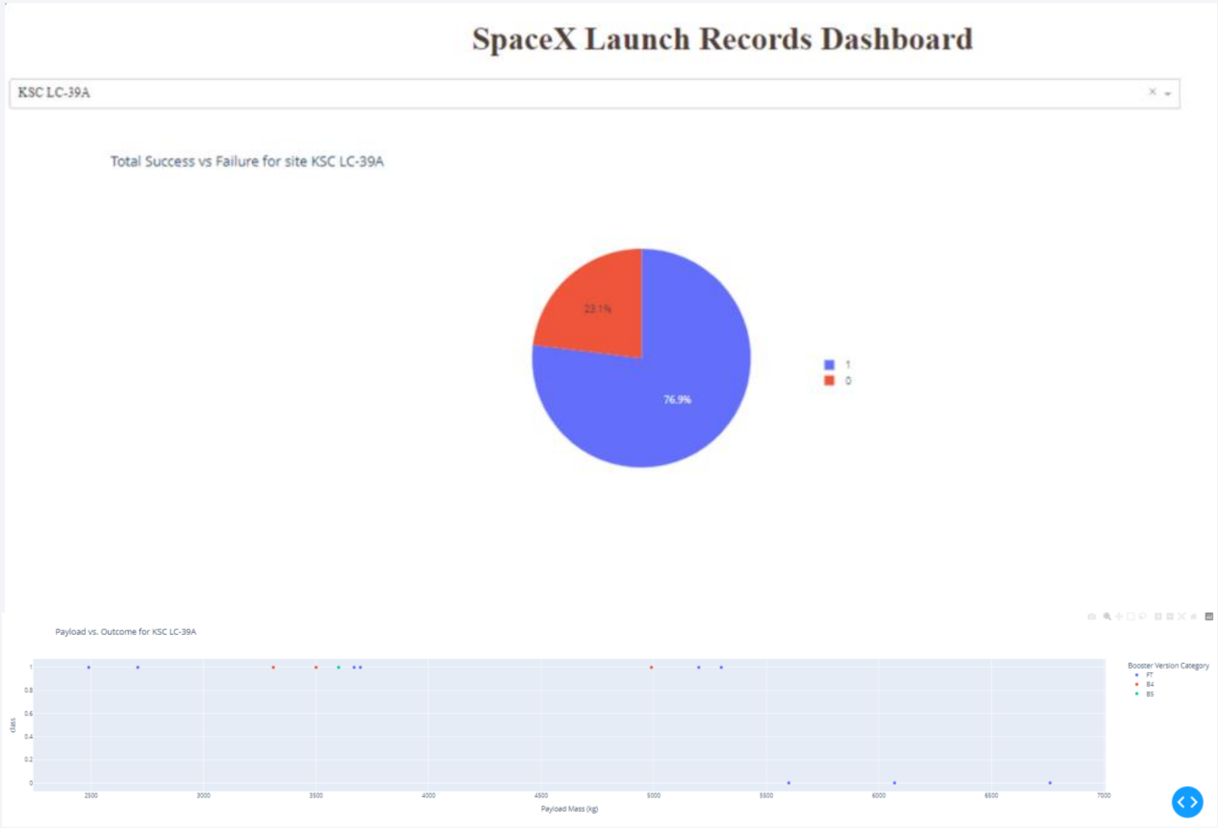
Launch Success by Site

- KSC LC-39A accounts for the largest portion of successful launches, contributing to 41.7% of the total.
- CCAFS LC-40 is the second-largest contributor, with 29.2% of successful launches.
- VAFB SLC-4E accounts for 16.7% of the total successful launches.
- CCAFS SLC-40 has the smallest share, with 12.5% of successful launches.

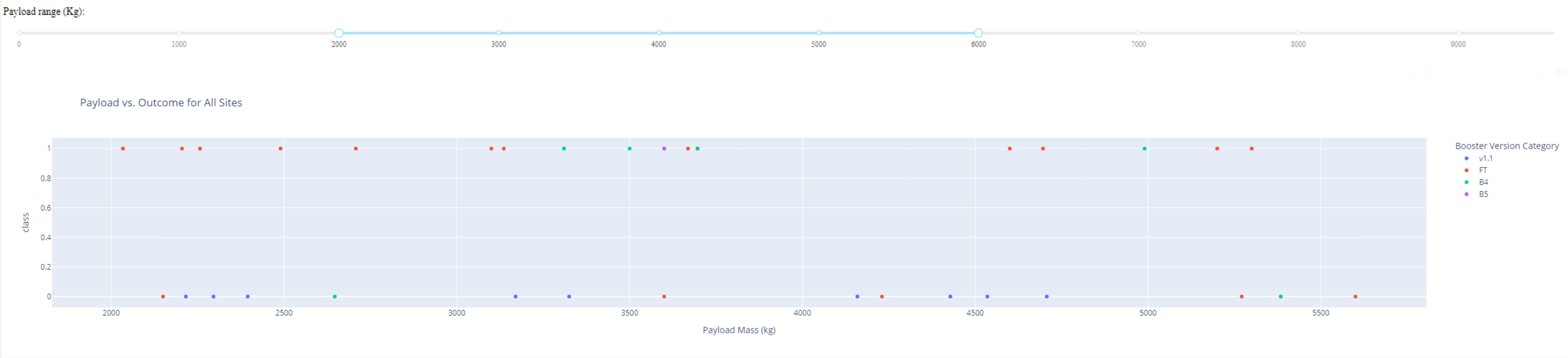


KSC LC-39A – Site with Highest Success Launch

- For KSC LC-39A, we have 10 successful launches accounting to 76.9% and only 3 failed launches which is 23.1%
- This is the highest successful site among 4 sites in use



Payload vs. Outcome for All Sites

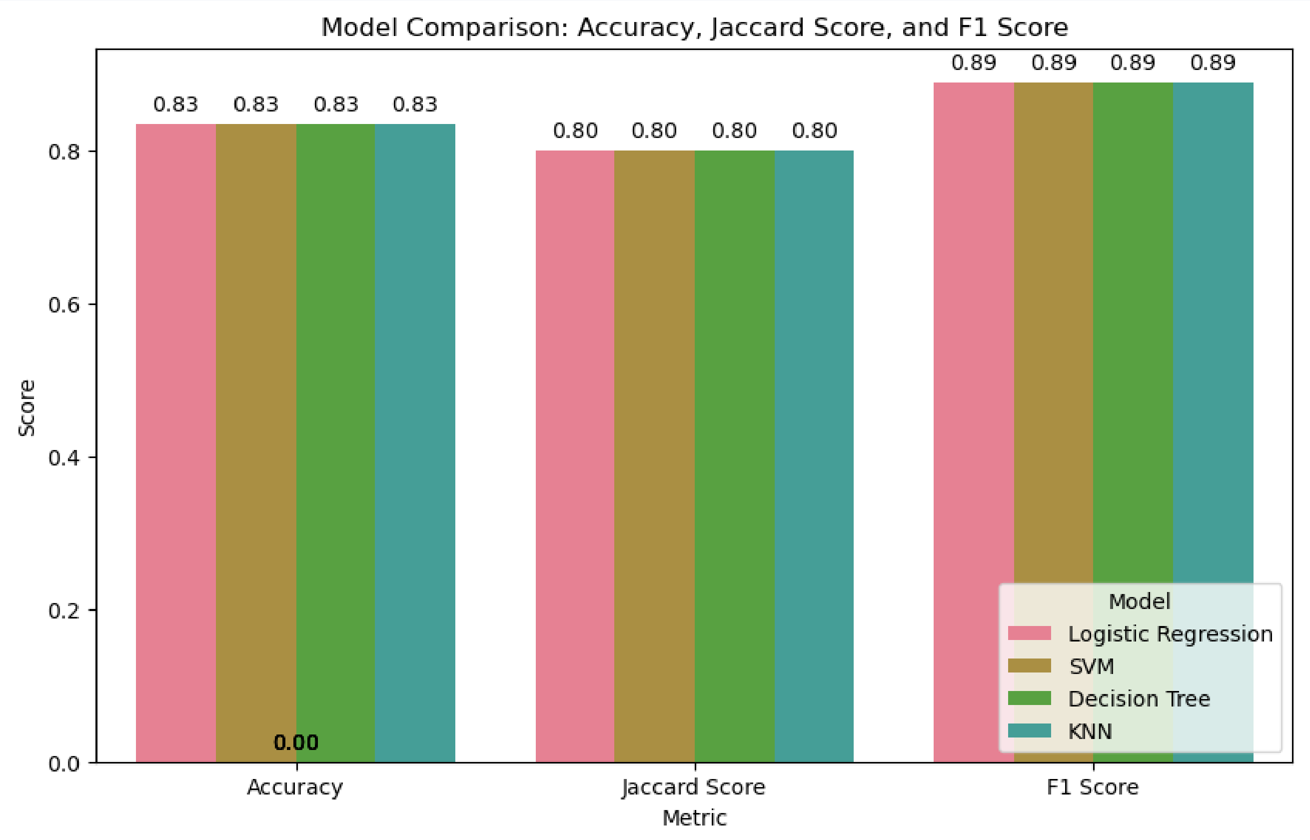


- We have 4 Booster versions used for Payload range of 2000 to kg to 6000 kg
- Payload between this range has the highest success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy



- All models perform in same level and have same accuracy and same level of scores.
- This is due to small dataset size
- Accuracy is in range of 83.33% for all models
- Jaccard Score is in range of 80%
- F1 Score is in 89%

Confusion Matrix

A confusion matrix summarizes the performance of a classification algorithm

Confusion matrix of all models are identical

Type 1 Error (False positives are not acceptable)

Confusion Matrix Outputs:

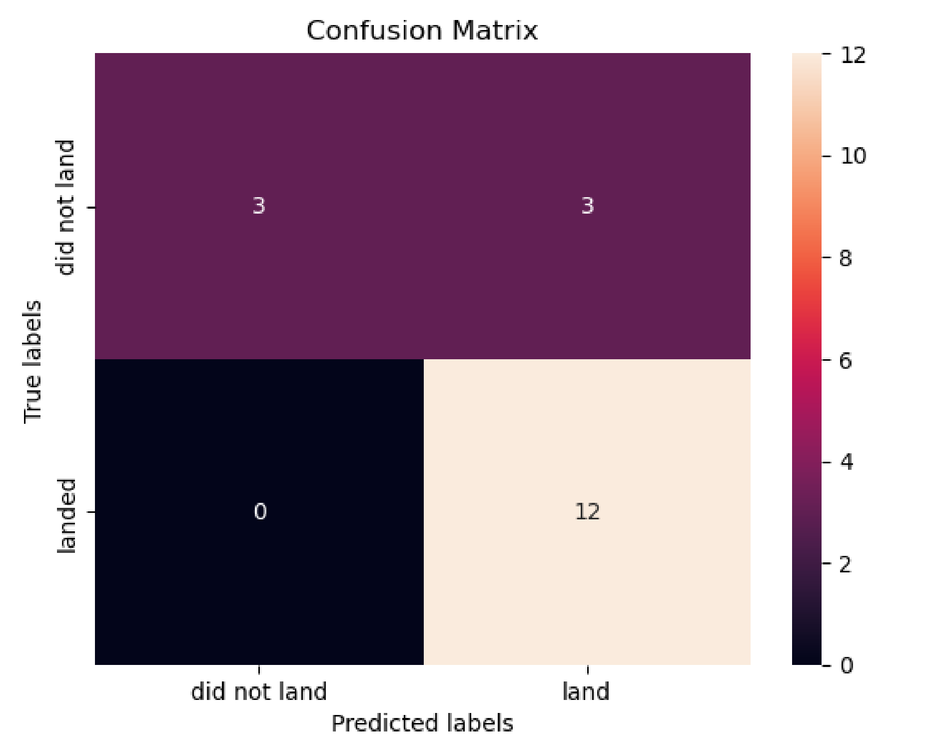
12 True Positive | 3 True Negative | 3 False Positive | 0 False Negative

Precision = TP / (TP + FP) = 12/15 = 0.80

Recall = TP / (TP + FN) = 12/12 = 1

F1 Score = 2 * (Precision * Recall) / (Precision + Recall) = 0.89

Accuracy = (TP + TN) / (TP + TN + FP + FN) = 0.8333



Conclusions

- **Model Performance:** All 4 models exhibited similar performance on the test data.
- **Proximity to the Equator:** Many of the launch sites are located closer to the equator, which provides a natural boost due to Earth's rotational speed. This helps reduce fuel consumption and the need for additional boosters.
- **Coastal Locations:** All launch sites are situated near coastal areas, allowing for safer launches over the ocean and reducing the risks associated with failures.
- **Launch Success Over Time:** The success rate of launches has increased consistently over time.
- **KSC LC-39A:** This site has the highest success rate among all launch locations, achieving a 100% success rate for payloads weighing less than 5,500 kg.
- **Orbital Success:** Orbits such as ES-L1, GEO, HEO, and SSO show a 100% success rate.
- **Payload Mass:** Higher payload masses (kg) across all launch sites tend to be associated with higher success rates.

Appendix

Web based Data Sources:

- SpaceX API (JSON): https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- Wikipedia: https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- Rest of the data sources are uploaded into the below github repository

Github Repository

- Github Public Link: https://github.com/varatharajs/SpaceX-Learning

Thank you!