Winning the Space Race with Data Science

Vitalii Hryhoryshen
July 27, 2024

Outline

1. Executive Summary
2. Introduction
3. Methodology
4. Results
5. Conclusion
6. Appendix

Executive Summary

This presentation predicts the success of Falcon 9 first stage landings, leveraging SpaceX's cost-effective launch model. SpaceX's reuse of the first stage significantly reduces launch costs to $62 million, compared to $165 million for other providers. Forecasting landing success allows us to estimate launch costs and provide strategic insights for competitors bidding against SpaceX.

Methodologies Employed:

- **Data Collection:** Using SpaceX’s API and web scraping to gather comprehensive launch data.
- **Data Wrangling:** Preprocessing and cleaning the collected data.
- **Exploratory Data Analysis (EDA):** Using SQL for data querying and visualization tools for in-depth analysis.
- **Interactive Visual Analytics:** Using Plotly Dash and Folium to map and explore data spatially.
- **Machine Learning Prediction:** Developing models to predict the success of first-stage landings.

Executive Summary

Summary of Results:

- **Exploratory Data Analysis:** Identified key features influencing launch success.
- **Interactive Visual Analytics:** Provided dynamic insights, enhancing understanding of data patterns.
- **Predictive Analysis:** Determined the most effective machine learning model, highlighting critical features for predicting landing success. This model forecasts launch costs accurately, offering a competitive edge to companies aiming to challenge SpaceX’s market position.

Introduction

Project Background and Context:

SpaceX has revolutionized the commercial space industry by making space travel more affordable. The company advertises Falcon 9 rocket launches at a cost of $62 million, significantly lower than the $165 million charged by other providers. This cost reduction is largely due to SpaceX’s ability to reuse the first stage of its rockets. Predicting the success of the first-stage landing is crucial for estimating the overall cost of a launch. By leveraging public data and machine learning models, we aim to forecast whether the Falcon 9 first stage will land successfully. This information is invaluable for competitors seeking to bid against SpaceX for rocket launches.

Problems to be Answered:

- What factors determine the success of the Falcon 9 first stage landing?
- How do variables such as payload mass, launch site, number of flights, and orbits impact the landing success?
- Has the success rate of first-stage landings improved over time?
- What is the most effective machine learning algorithm for predicting the success of the first-stage landing?

SECTION 1

Methodology

Methodology Overview:

Data Collection:

- Data obtained from two primary sources:
    1. SpaceX API (https://api.spacexdata.com/v4/rockets/)
    2. Web scraping from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Wrangling:

- Filtered and cleaned data to handle missing values.
- Applied one-hot encoding to prepare categorical data for binary classification.

Exploratory Data Analysis (EDA):

- Conducted EDA using SQL and various data visualization techniques to uncover insights and patterns.

Interactive Visual Analytics:

- Utilized Folium and Plotly Dash for dynamic and interactive data visualizations.

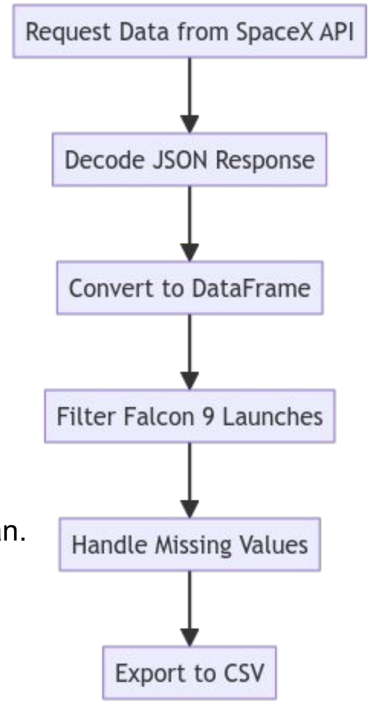
Predictive Analysis:

- Developed, tuned, and evaluated classification models to predict the success of the Falcon 9 first stage landing, ensuring optimal performance and accuracy.

Data Collection Process

API Requests from SpaceX REST API

1. **Request Data:**
   - Utilize the SpaceX API to gather rocket launch data.
   - Decode the JSON response and convert it into a DataFrame using `.json_normalize()`.
2. **Create DataFrame:**
   - Extract relevant details: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude.
   - Filter the DataFrame to include only Falcon 9 launches.
3. **Clean Data:**
   - Handle missing values, replacing missing PayloadMass values with the calculated mean.
4. **Export Data:**
   - Export the cleaned DataFrame to a CSV file for further analysis.

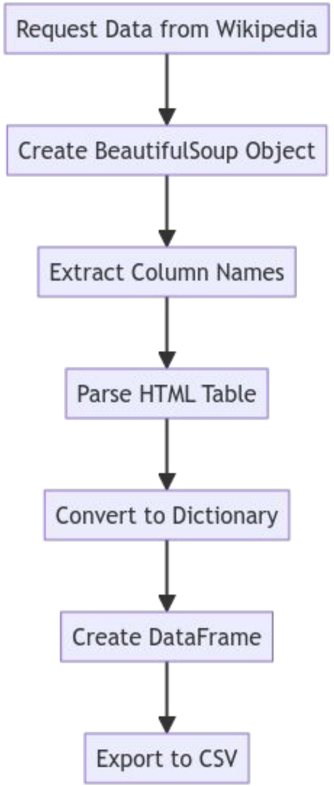


For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/1_data-collection-api.ipynb

Data Collection Process

Web Scraping from Wikipedia

1. **Request Data:**
    - Perform web scraping on the Wikipedia page containing Falcon 9 launch records.
    - Use BeautifulSoup to create an object from the HTML response.
2. **Extract Data:**
    - Extract column names from the HTML table header.
    - Parse the HTML table to collect data.
3. **Create DataFrame:**
    - Convert the parsed table data into a dictionary.
    - Create a DataFrame from the dictionary, including columns like Flight No., Launch Site, Payload, PayloadMass, Orbit, Customer, Launch Outcome, Version Booster, Booster Landing, Date, and Time.
4. **Export Data:**
    - Export the DataFrame to a CSV file for integration with the API data.



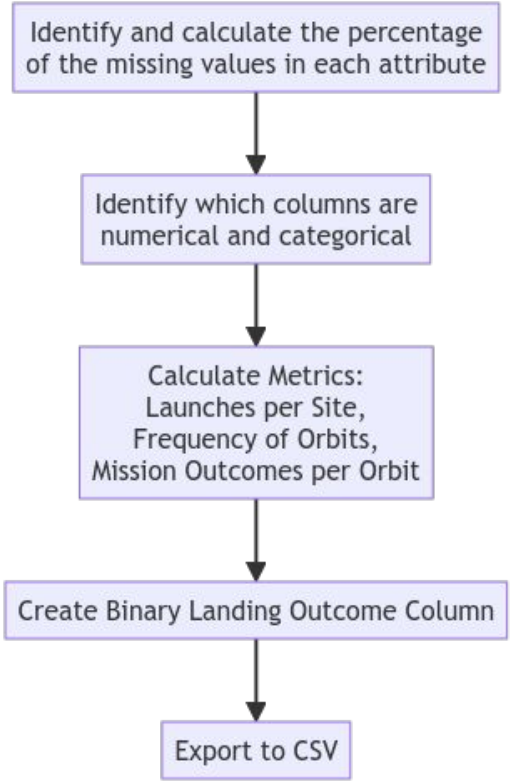
For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/1_data-collection-with-web-scraping.ipynb

Data Wrangling

Steps

1. **Perform Exploratory Data Analysis (EDA):**
    - Identify and calculate the percentage of the missing values in each attribute.
    - Identify which columns are numerical and categorical.
2. **Calculate Key Metrics:**
    - Number of launches per site.
    - Frequency and occurrence of different orbits.
    - Mission outcomes for each orbit type.

…



For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/2_data-wrangling.ipynb

Data Wrangling

Steps

3. **Create Binary Landing Outcome Column:**
    - Define landing outcomes:
        - True Ocean: Successful landing in the ocean.
        - False Ocean: Unsuccessful landing in the ocean.
        - True RTLS (Return to Launch Site): Successful ground pad landing.
        - False RTLS: Unsuccessful ground pad landing.
        - True ASDS (Autonomous Spaceport Drone Ship): Successful drone ship landing.
        - False ASDS: Unsuccessful drone ship landing.
    - Convert these outcomes into binary labels:
        - 1: Successful landing.
        - 0: Unsuccessful landing.
4. **Export Processed Data:**
    - Export the final processed DataFrame to a CSV file for further analysis and model training.

For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/2_data-wrangling.ipynb

Exploratory Data Analysis with Data Visualization

Charts Plotted and Reasons

1. **Flight Number vs. Payload Mass:**
   - **Scatter Plot:**
     - Visualizes the relationship between flight number and payload mass.
     - Identifies trends and outliers in payloads over different flights.
2. **Flight Number vs. Launch Site:**
   - **Scatter Plot:**
     - Shows distribution of flight numbers across different launch sites.
     - Understands the frequency of launches from each site over time.
3. **Payload Mass vs. Launch Site:**
   - **Scatter Plot:**
     - Compares average payload mass across different launch sites.
     - Highlights which sites handle heavier payloads.
4. **Orbit Type vs. Success Rate:**
   - **Bar Chart:**
     - Compares landing success rates across different orbit types.
     - Identifies orbit types more likely to result in successful landings. …

For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/3_eda-with-visualization.ipynb

Exploratory Data Analysis with Data Visualization

Charts Plotted and Reasons

5. **Payload Mass vs. Orbit Type:**
    - **Scatter Plot:**
        - Illustrates distribution of payload masses across various orbit types.
        - Indicates payload capacity requirements for different orbits.
6. **Flight Number vs. Orbit Type:**
    - **Scatter Plot:**
        - Displays relationship between flight numbers and orbit types.
        - Shows trends in orbit type selection over time.
7. **Success Rate Yearly Trend:**
    - **Line Chart:**
        - Tracks yearly trend of landing success rates.
        - Reveals improvements or declines in success rates over time.

Insights: These visualizations provide insights into the key factors influencing the success of Falcon 9 first-stage landings, helping identify features for predictive modeling.

For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/3_eda-with-visualization.ipynb

Exploratory Data Analysis with SQL

Key SQL Queries:

1. **Unique Launch Sites:**
   - Display unique launch site names.
2. **Launch Sites Starting with 'CCA':**
   - Display 5 records where launch site names begin with 'CCA'.
3. **Total Payload Mass by NASA (CRS):**
   - Display total payload mass carried by NASA (CRS) launches.
4. **Booster Version F9 v1.1:**
   - Display average payload mass for booster version F9 v1.1.
5. **First Successful Ground Pad Landing:**
   - List date of first successful ground pad landing.

…

For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/3_eda-with-sql.ipynb

Exploratory Data Analysis with SQL

Key SQL Queries:

6. **Boosters with Successful Drone Ship Landings:**
    - List boosters with successful drone ship landings and payloads between 4000-6000 kg.
7. **Mission Outcomes:**
    - List total number of successful and failed mission outcomes.
8. **Maximum Payload Mass:**
    - List booster versions carrying maximum payload mass.
9. **Failed Drone Ship Landings in 2015:**
    - List failed drone ship landings in 2015 with booster versions and launch sites.
10. **Count of Landing Outcomes:**
    - Rank landing outcomes between 2010-06-04 and 2017-03-20 in descending order.

For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/3_eda-with-sql.ipynb

Build an Interactive Map with Folium

Map Features:

1. **Launch Site Markers:**
   - Red circle markers at launch site coordinates with popup labels.
2. **Launch Outcome Markers:**
   - **Successful Launches:** Green markers.
   - **Failed Launches:** Red markers.
3. **Proximity Distance Lines:**
   - **Example:** CCAFS SLC-40 to nearby features (coastline, railway, highway, nearest city).

Purpose:

- **Geographic Distribution:**
  - Visualize launch site locations.
- **Launch Success Rates:**
  - Color-coded markers for quick success/failure identification.
- **Proximity to Infrastructure:**
  - Assess site logistics and accessibility.

For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/4_interactive-visual-analytics-with-folium.ipynb

Build a Dashboard with Plotly Dash

Interactive Elements:

1. **Launch Sites Dropdown List:**
   - Filter data by launch site or view all sites.
2. **Pie Chart Showing Launch Outcomes:**
   - Display total successful and unsuccessful launches; updates per selected site.
3. **Payload Mass Range Slider:**
   - Filter data based on payload mass range.
4. **Scatter Chart of Payload Mass vs. Success Rate:**
   - Show relationship between payload mass and success rates, categorized by booster versions.

Purpose:

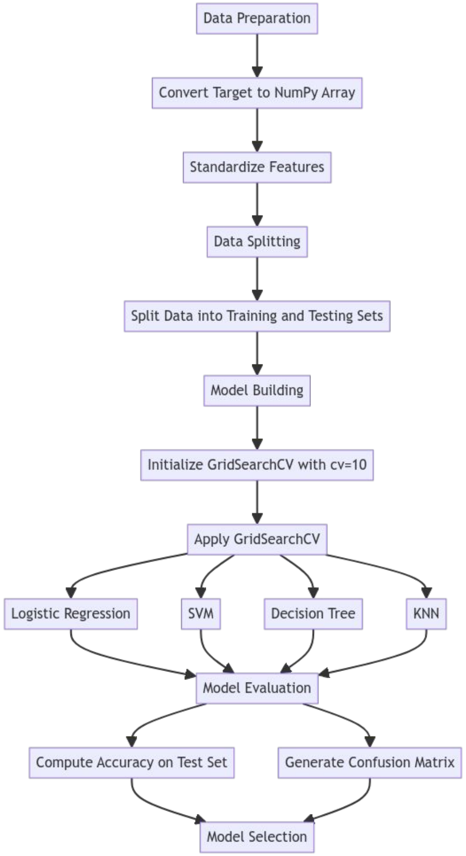
- **Site-specific and Overall Data:**
  - Flexible data viewing options.
- **Performance Visualization:**
  - Quick grasp of launch site performance and payload mass influence.

For more details, see the completed source code on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/4_interactive-dashboard-with-plotly-dash.py

Predictive Analysis (Classification)

Model Development Process:

1. **Data Preparation and Standardization:**
    - Convert target to NumPy array.
    - Standardize features with StandardScaler.
2. **Data Splitting:**
    - Split data into training and testing sets.
3. **Model Building and Hyperparameter Tuning:**
    - Use GridSearchCV with 10-fold cross-validation for:
        - Logistic Regression
        - SVM
        - Decision Tree
        - KNN
4. **Model Evaluation:**
    - Evaluate models using accuracy score and confusion matrices.
5. **Model Selection:**
    - Compare models using accuracy score.



For more details, see the completed notebook on GitHub:
https://github.com/vithryh/winning-the-space-race-with-data-science/blob/main/5_machine-learning-prediction.ipynb

Results

- **Exploratory Data Analysis:**
  - Summary of key findings.
- **Interactive Analytics:**
  - Screenshots of interactive visualizations.
- **Predictive Analysis:**
  - Performance and insights from the best classification model.

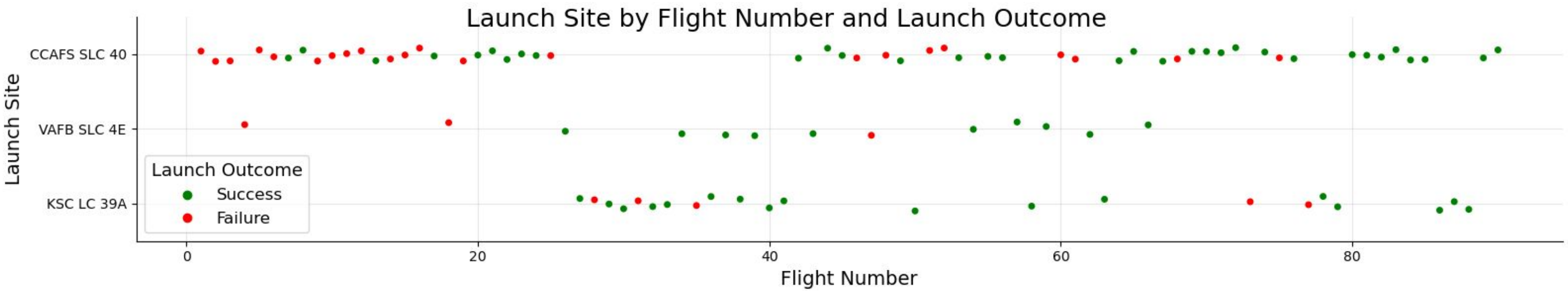
SECTION 2

Insights from EDA

Launch Sites and Outcomes Across Flight Numbers

Key Observations:

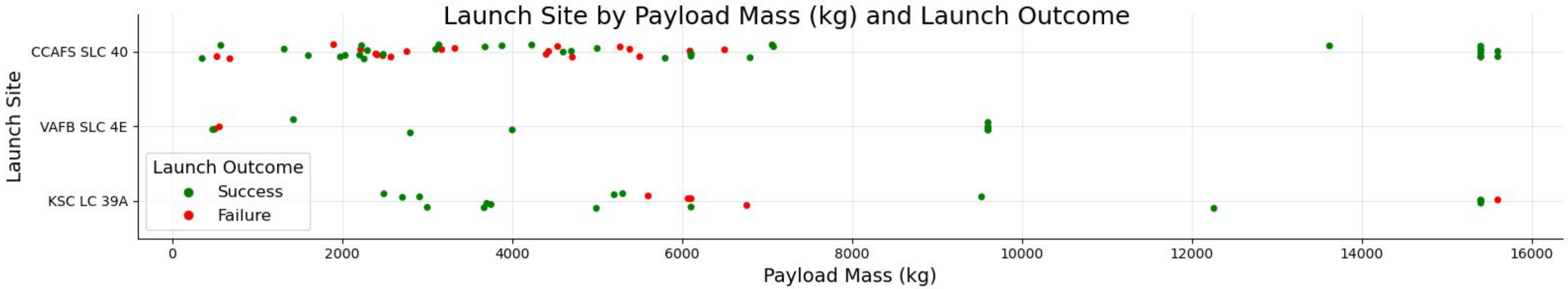
- **Success Rate Over Time:**
  - Early flights (blue) had lower success rates, while recent flights (orange) show improvement.
- **Launch Site Distribution:**
  - About a half of the launches came from CCAFS SLC 40.
  - Nearly all of the first twenty launches were from CCAFS SLC-40, and most were unsuccessful.
- **Trends:**
  - Increasing success rate over time, indicating technological and process improvements.



Launch Sites and Outcomes Across Payload Masses

Key Observations:

- **Payload Mass & Success Rate:**
  - Higher payload masses generally correlate with higher success rates.
  - Payloads >9,500 kg mostly successful.
- **Launch Site Performance:**
  - KSC LC 39A: 100% success for payloads <5,400 kg.
  - VAFB SLC 4E: No payloads >9,600 kg.
- **Insights:**
  - KSC LC-39A has achieved 10 successful launches, with payloads ranging from 2,490 kg to 5,300 kg.

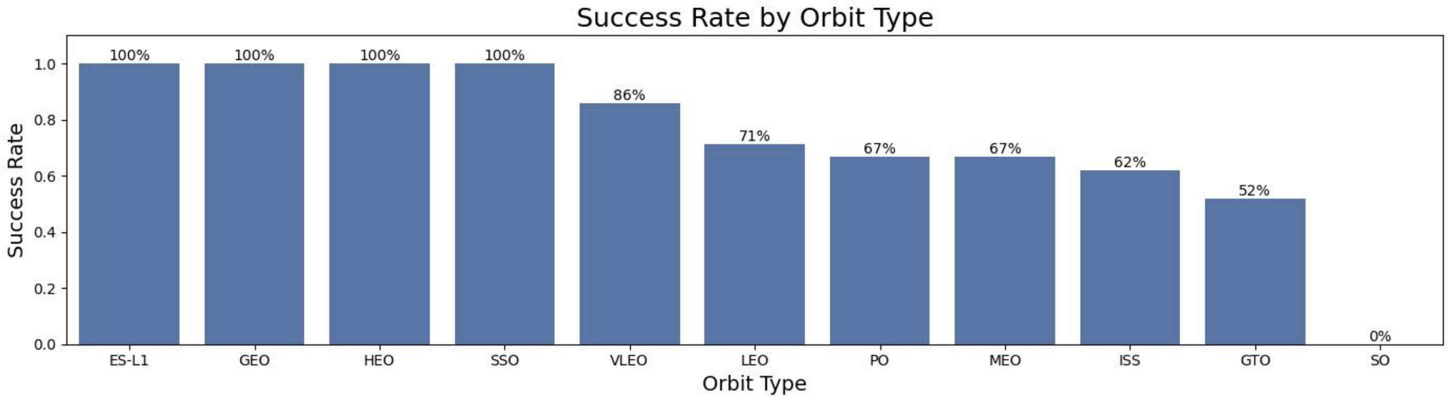


Success Rates Across Orbit Types

Key Observations:

- **High Success Rates (100%):**
  - ES-L1, GEO, HEO, and SSO.
- **Moderate Success Rates (50%-75%):**
  - GTO, ISS, LEO, MEO, PO.
- **No Success (0%):**
  - SO orbit.

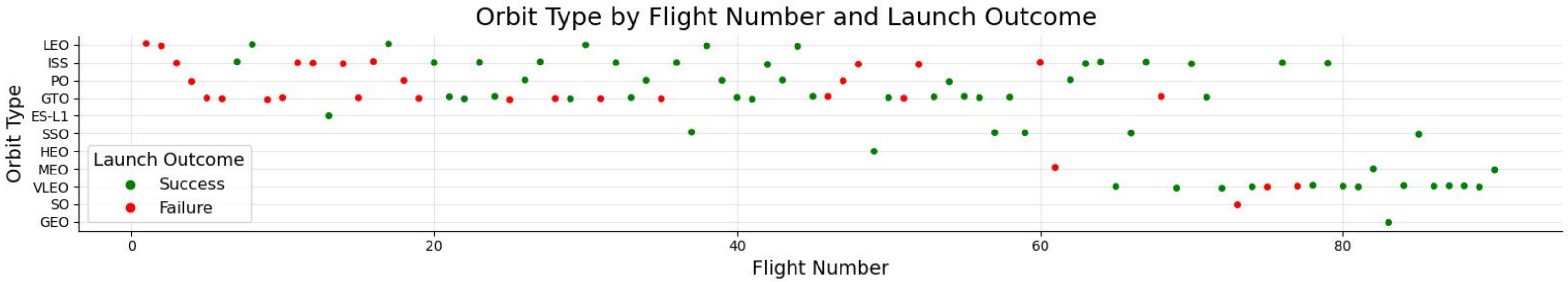
- **Summary:**
  - Perfect success in orbits like ES-L1, GEO, HEO, SSO.
  - Common orbits (LEO, GTO) show moderate success, reflecting launch frequency and challenges.



Orbit Types and Outcomes Across Flight Numbers

Key Observations:

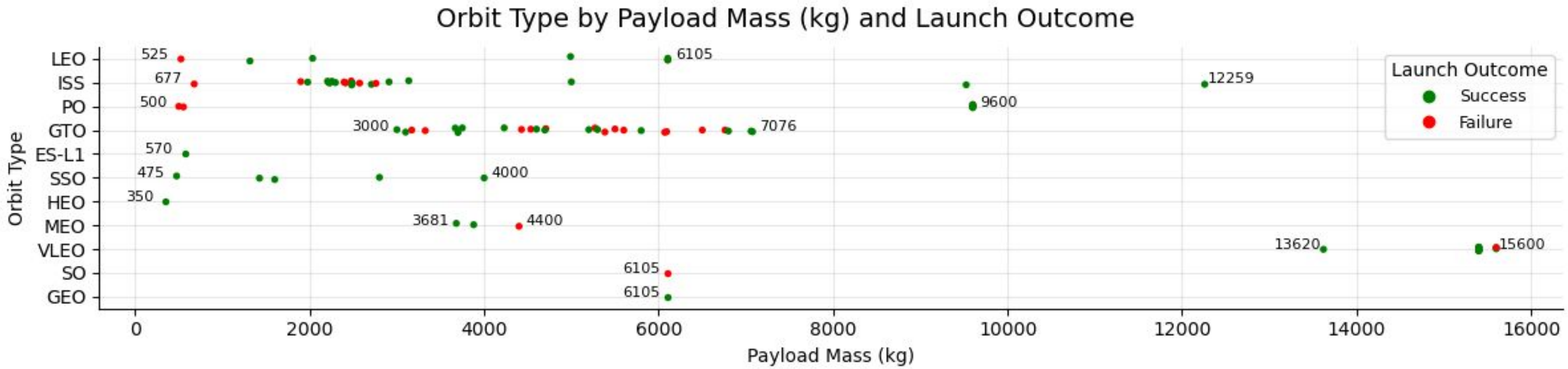
- **Trends in Specific Orbits**:
  - In the first two-thirds of the flights, no missions were launched to orbits such as Medium Earth Orbit (MEO), Very Low Earth Orbit (VLEO), Suborbital (SO), or Geostationary Orbit (GEO), and in the last one-third of the flights, almost no missions were launched to Low Earth Orbit (LEO), Polar Orbit (PO), or Geostationary Transfer Orbit (GTO).
  - All five flights to Sun-Synchronous Orbit (SSO) and the only flight to the highly Elliptical Orbit (HEO) and the only flight to the Geostationary Orbit (GEO) were successful.
  - The last six successful launches to the ISS were preceded by three failures.



Orbit Types and Outcomes Across Payload Masses

Key Observations:

- **Payload Performance by Orbit:**
  - Heavier payloads have higher success rates in LEO, ISS, and PO orbits.
- **GTO Orbit Trends:**
  - Mixed success with heavier payloads, indicating variable impact of payload weight.



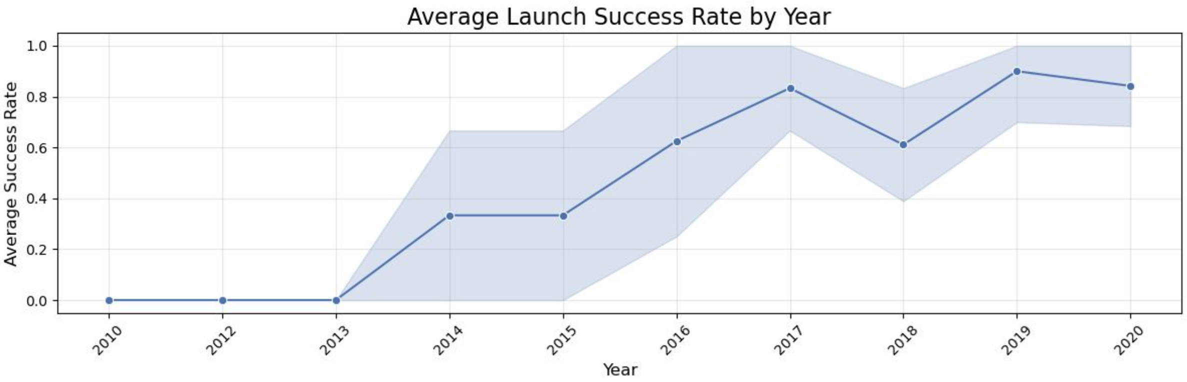
Launch Success Yearly Trend

Highlights:

- **Overall Trend:**
  - Success rate improved from 2013 to 2020.
- **Yearly Fluctuations:**
  - Increased success rates: 2013-2017, 2018-2019.
  - Dips observed: 2017-2018, 2019-2020.

Key Takeaway:

- Despite fluctuations, a steady improvement in launch success since 2013.



All Launch Site Names

Result Description:

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

Purpose:

- Understanding the geographical distribution and usage frequency of each site.

```
%%sql
SELECT
    distinct("Launch_Site")
FROM
    SPACEXTBL;
```

Running query in 'sqlite:///data/my_data1.db'

[8]:

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

Launch Site Names Begin with 'CCA'

Result Description:

- **Records:**
    - First five launches from CCAFS LC-40.

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

Running query in 'sqlite:///data/my_data1.db'

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

Result Description:

- **Calculation:**
  - Total payload mass for NASA CRS missions: 45,596 kg.

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

Running query in 'sqlite:///data/my_data1.db'

[11]:

sum("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

Result Description:

- **Average Payload Mass:**
  - Falcon 9 v1.1 boosters carry an average of 2,928 kg.

Insight:

- Typical payload capacity of the Falcon 9 v1.1 booster.

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

Running query in 'sqlite:///data/my_data1.db'

[13]:

avg("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

Result Description:

- **Date of First Successful Landing:**
    - 22 December 2015

Significance:

- Milestone in SpaceX's history of achieving ground pad landings.

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

Running query in 'sqlite:///data/my_data1.db'

[15]:

min("Date")
2015-12-22

Successful Drone Ship Landings (4000-6000 kg)

Result Description:

- **Boosters:**
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Insight:

- Specific booster versions capable of handling moderate payloads with successful drone ship landings.

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

Running query in 'sqlite:///data/my_data1.db'

[16]:

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

Total Number of Mission Outcomes

Result Description:

- **Mission Outcomes:**
  - Failure (in flight): 1
  - Success: 99
  - Success (unclear payload status): 1

Overview:

- SpaceX's mission success rate and outcome frequencies.

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

Running query in 'sqlite:///data/my_data1.db'

[16]:

Mission_Outcome	count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters with Maximum Payload

Result Description:

- **Boosters:**
  - 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

Insight:

- Maximum payload capacity achieved by these booster versions.

```sql
%%sql
SELECT
    "Booster_Version"
FROM
    SPACEXTBL
WHERE
    "PAYLOAD_MASS__KG_" = (SELECT max("PAYLOAD_MASS__KG_")
                           FROM SPACEXTBL);
```

Running query in 'sqlite:///data/my_data1.db'

[18]:

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

Result Description:

- **Failed Drone Ship Landings:**
  - January: F9 v1.1 B1012 at CCAFS LC-40
  - April: F9 v1.1 B1015 at CCAFS LC-40

Insight:

- Detailed information about unsuccessful landings in 2015.

```
%%sql
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,
    Booster_Version,
    Launch_Site,
    Landing_Outcome
FROM
    SPACEXTBL
WHERE
    substr(Date, 0, 5) = '2015'
    AND Landing_Outcome LIKE '%Failure (drone ship)%'
ORDER BY
    Date;
```

Running query in 'sqlite:///data/my_data1.db'

[19]:

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)

Rank Landing Outcomes (2010-06-04 to 2017-03-20)

Result Description:

- **Landing Outcome Rankings:**
    - 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

Insight:

- Distribution and frequency of landing outcomes over time.

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

Running query in 'sqlite:///data/my_data1.db'

[20]:

Landing_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

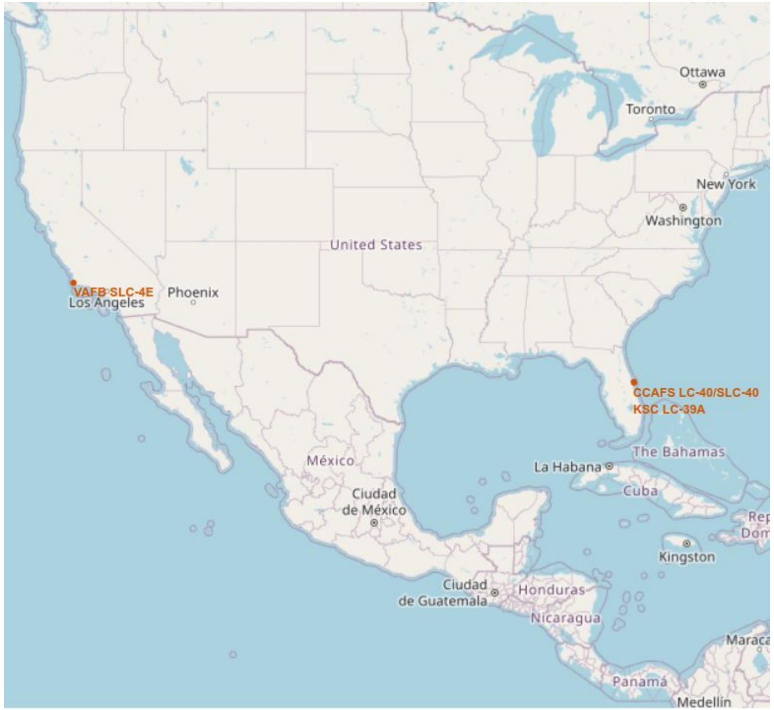
SECTION 3

Launch Sites Proximities Analysis

Launch Sites with Global Markers

Explanation:

- **Equatorial Proximity:** Launch sites near the equator benefit from the Earth's rotational speed, allowing for more efficient achievement of orbital velocity.
- **Coastal Locations:** Launch sites are located near coastlines to minimize the risk to populated areas in the event of launch failures.

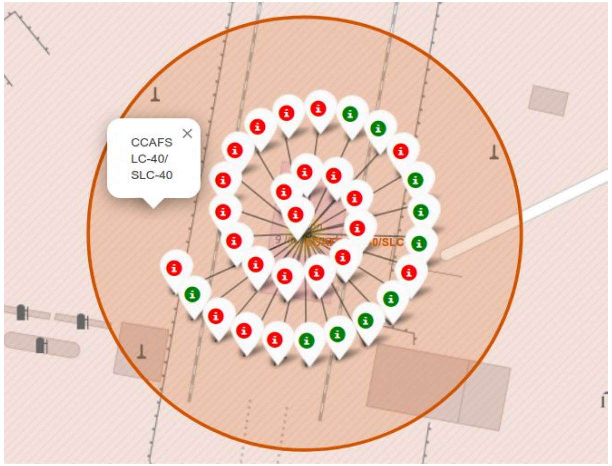


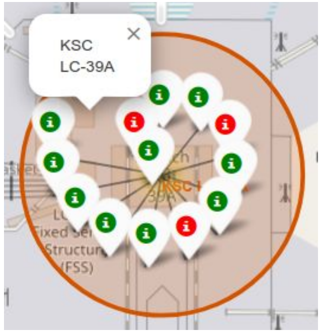
Color-Coded Launch Outcomes

Explanation:

- **Marker Colors:**
  - Green: Successful launch
  - Red: Failed launch

	Success	Fail	Total
CCAFS LC-40/SLC-40	10	23	33
KSC LC-39A	10	3	13
VAFB SLC-4E	4	6	10
Total	24	32	56

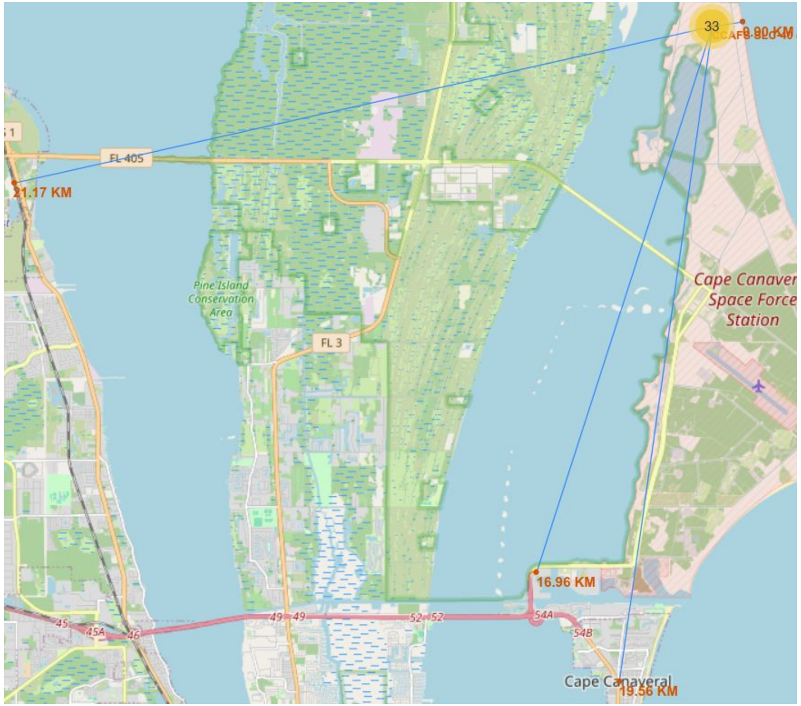




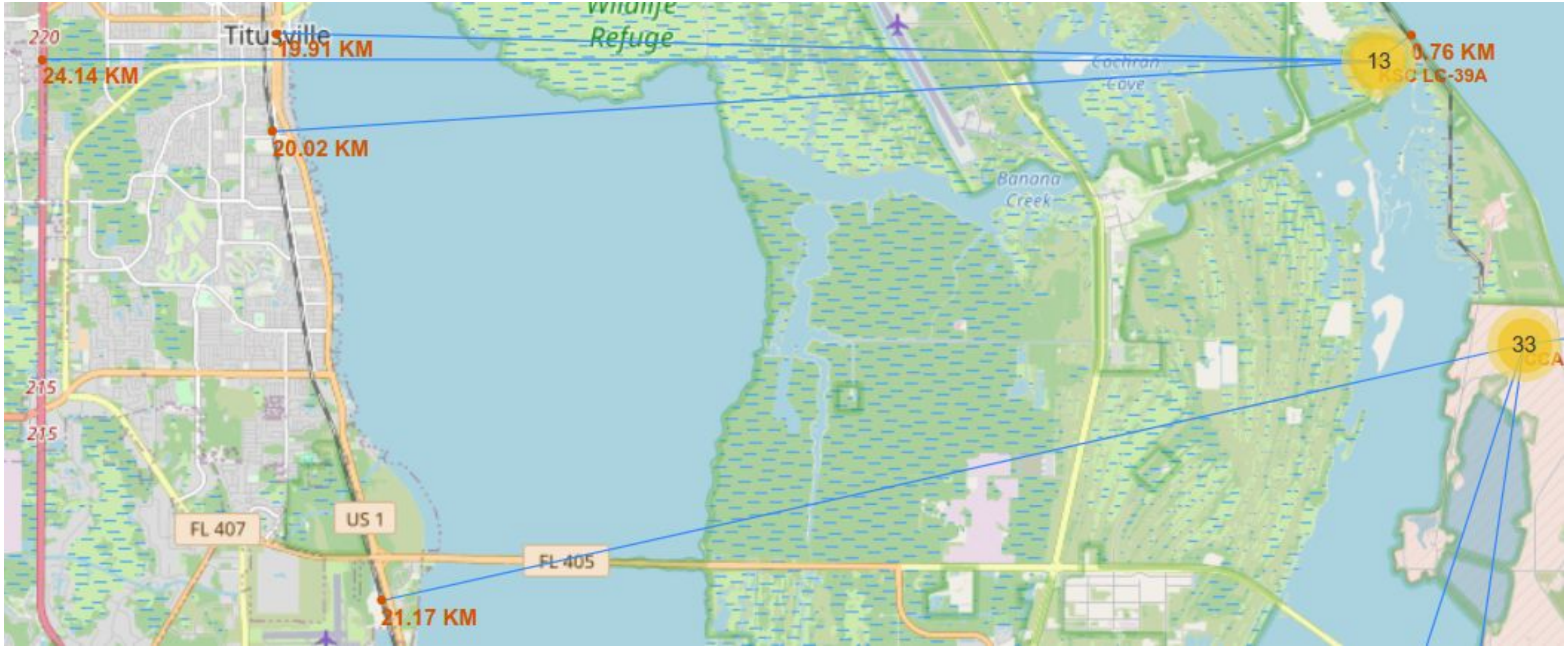
Launch Site Proximities and Distances

Explanation:

- **Proximities to Key Locations:** Distances from a selected launch site to significant locations.
- **Example - CCAFS LC-40/SLC-40:**
    - Coastline: 0.90 km
    - Highway (SR 401): 16.96 km
    - Railway (FEC): 21.17 km
    - City (Cape Canaveral): 19.56 km
- **Example - KSC LC-39A:**
    - Coastline: 0.76 km
    - Highway (SR 401): 24.14 km
    - Railway (FEC): 20.02 km
    - City (Titusville): 19.91 km
- **Importance:** These proximities are crucial for operational efficiency and safety considerations.



Launch Site Proximities and Distances



SECTION 3

Build a Dashboard with Plotly Dash

Launch Success by Site

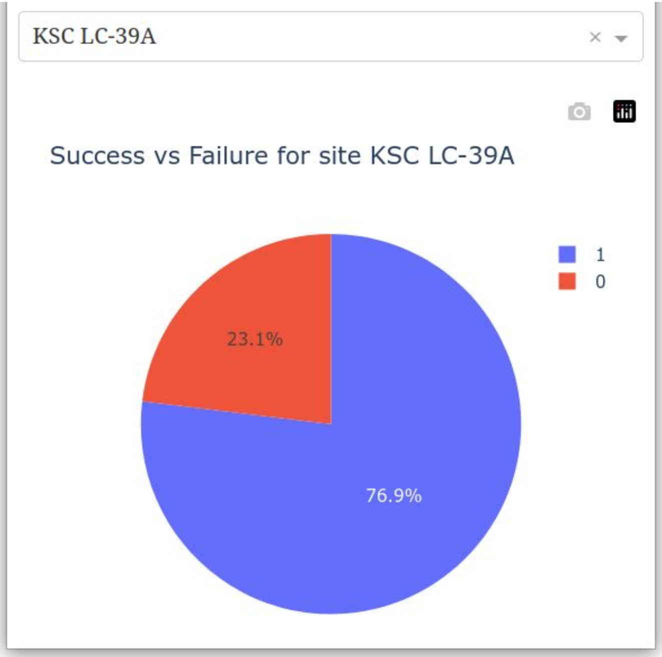
Explanation:

- **Distribution:** Represents successful launches across different sites.
- **Key Findings:**
  - KSC LC-39A and CCAFS LC-40/SLC-40 each account for the highest percentage of successful launches at 41.7%, demonstrating their critical roles.
  - VAFB SLC-4E contributes significantly, with 16.7% of the total successful launches, though this is lower compared to KSC LC-39A and CCAFS LC-40/SLC-40.

Highest Launch Success Ratio

Explanation:

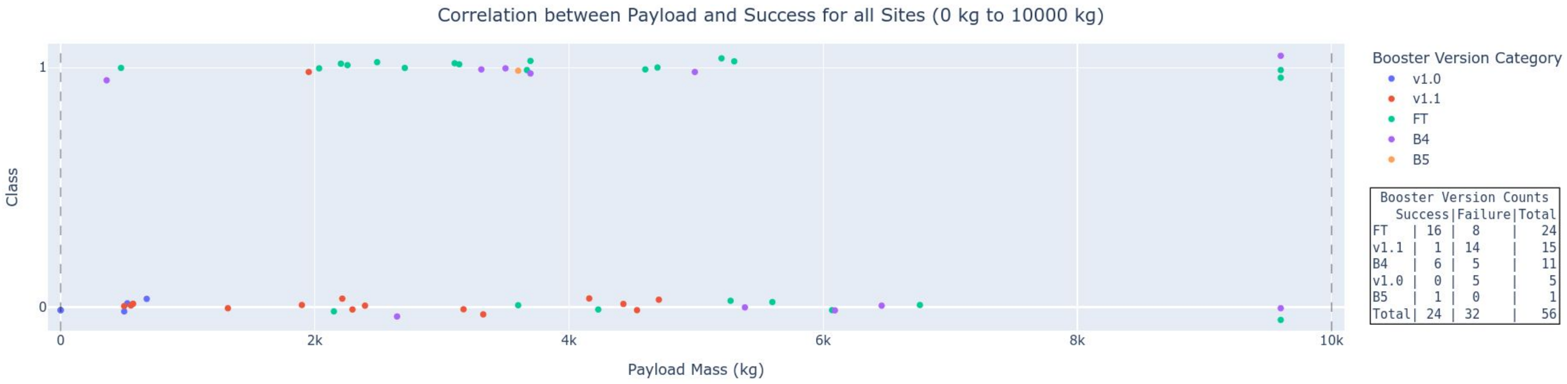
- **Success Rate:** Highlights the site with the highest success ratio.
- **Key Findings:**
    - **KSC LC-39A:** Highest success rate at 76.9% (10 successes, 3 failures).
    - **Significance:** Indicates reliability and efficiency at KSC LC-39A.



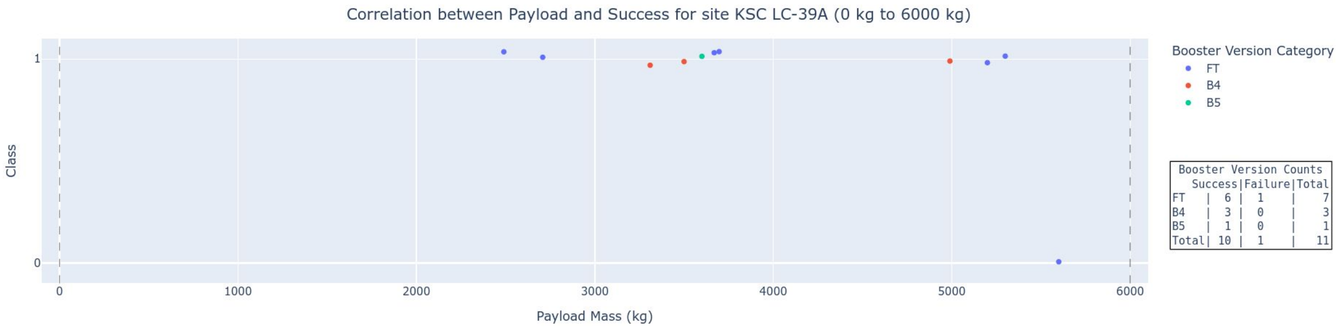
Payload Mass vs. Launch Outcome

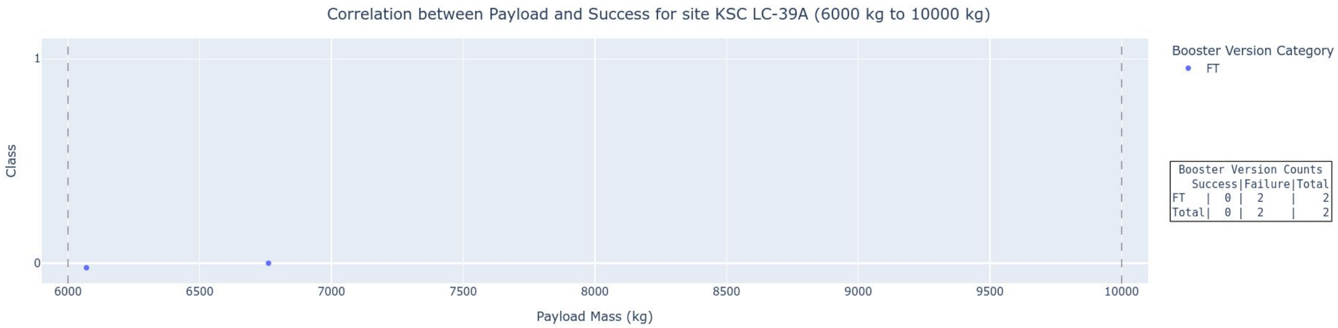
Explanation:

- **Key Findings:**
  - **Payload Range:** 2,000 kg to 5,400 kg shows highest success rate.
  - The **v1.0** version has not achieved any successes, indicating possible reliability problems.
  - The **v1.1** version's high failure rate suggests potential issues that need to be addressed.
  - The **FT** version demonstrates strong performance in absolute terms despite having some failures.
  - The **B5** booster version stands out with a perfect success rate, having achieved 1 success and 0 failures.



Payload Mass vs. Launch Outcome





SECTION 4

Predictive Analysis (Classification)

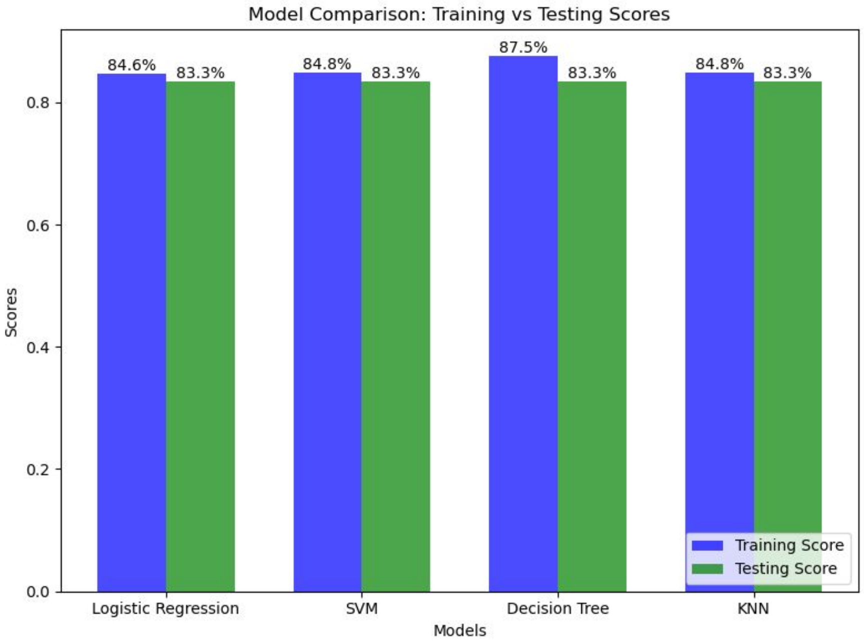
Classification Accuracy

Model Performance Summary

All models achieved the same highest testing accuracy of 83.3%.

- **Decision Tree** had the highest training accuracy at 87.5%.

Note: The uniformity in testing accuracy may be due to a small dataset.



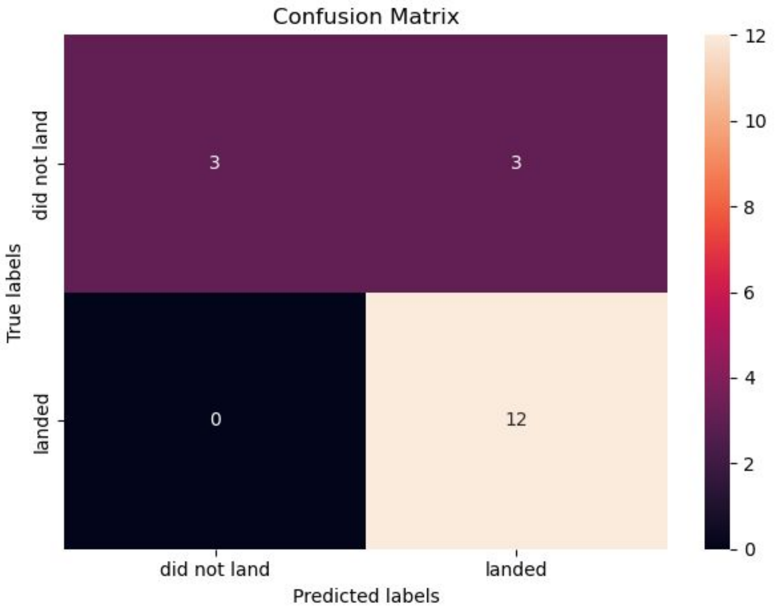
Confusion Matrix

Decision Tree Model Performance Insights:

- **Accuracy**: The model correctly classified the landing status in most cases, with a high accuracy rate.
- **Error Analysis**: The primary errors are false positives, where the model predicts a successful landing when there has been a failure. There are no false negatives, indicating strong sensitivity in detecting actual successful landings.

Conclusion:

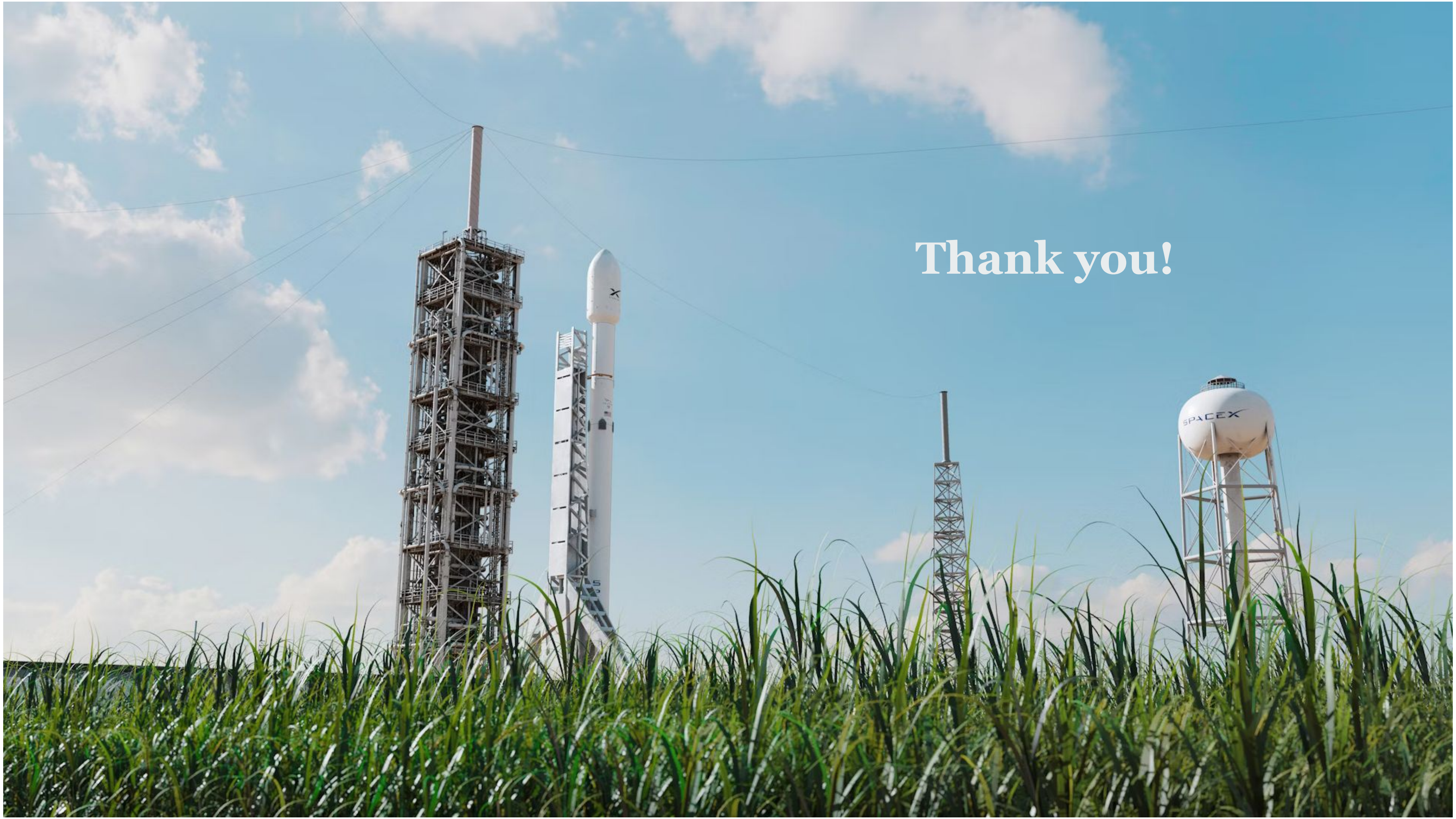
The decision tree classifier, optimized with GridSearchCV, shows robust performance in predicting rocket landings, with perfect sensitivity and a minor rate of false positives. This indicates high reliability for operational use, with further tuning potentially reducing false positives.



Conclusions

Summary of Key Insights:

- **Top Launch Site:** KSC LC-39A has the highest success rate at **76.9%**.
  - **10 successes:** Payload mass ranged from 2,490 kg to 5,300 kg.
  - **3 failures:** Payload mass ranged from 5,600 kg to 6,761 kg.
- **Booster Performance:**
  - **v1.0:** No successes, reliability issues (0 successes, 5 failures).
  - **v1.1:** High failure rate, needs improvement (1 success, 14 failures).
  - **FT:** Strong performance overall, despite some failures (16 successes, 8 failures).
  - **B5:** Perfect success rate (1 success, 0 failures).
- **Model Performance:** All models achieved the same highest testing accuracy of **83.3%** (perhaps due to a small dataset)**.**
- **Success Trends:** Launch success rates have steadily improved over the years.
- **Orbit Success Rates:** ES-L1, GEO, HEO, and SSO achieved a 100% success rate. GTO, ISS, LEO, MEO, and PO achieved a 50%-75% success rate.
- **Equatorial Advantage:** Sites near the equator leverage Earth's rotational speed, optimizing fuel usage.
- **Coastal Proximity:** Sites are close to coastlines to reduce debris risks.


Thank you!
SPACEX