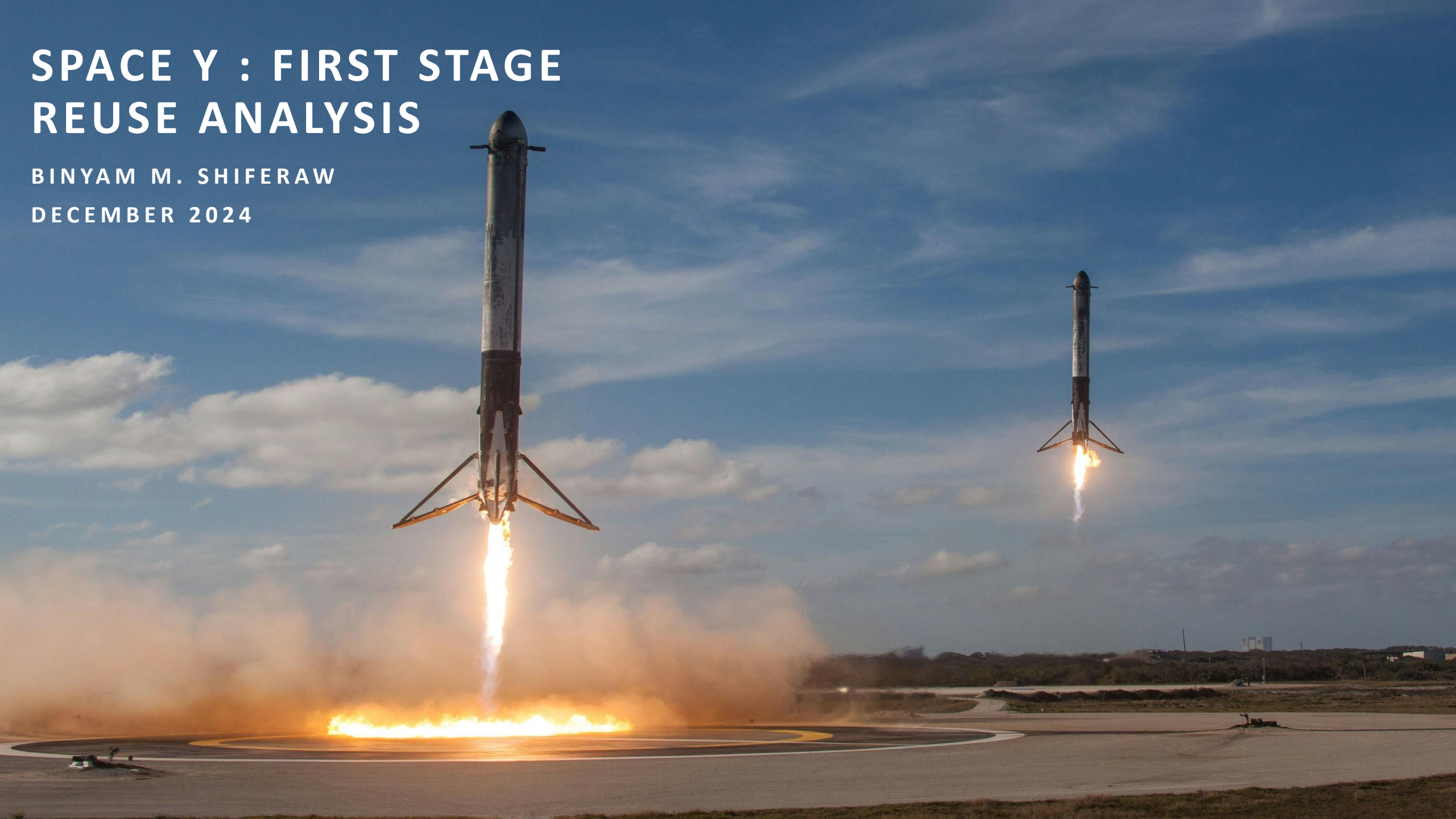
SPACE Y : FIRST STAGE REUSE ANALYSIS

BINYAM M. SHIFERAW

DECEMBER 2024



AGENDA

- EXECUTIVE SUMMARY
- INTRODUCTION
- METHODOLOGY
- RESULTS
- CONCLUSION
- ANNEX

EXECUTIVE SUMMARY

METHODOLOGIES

The study focuses on uncovering the key factors contributing to a successful rocket landing. The following approaches were employed to achieve this objective:

- **Data Collection:** Information was retrieved using the SpaceX REST API and supplemented with web scraping methods.
- **Data Preparation:** The dataset was processed to generate a variable representing success or failure outcomes.
- **Exploratory Data Analysis:** Data visualization techniques were applied to investigate various factors such as payload, launch site, flight sequence, and yearly trends.
- **Data Analysis:** SQL was utilized to calculate key metrics, including total payload, payload ranges for successful launches, and the number of successful versus failed outcomes.
- **Launch Site Evaluation:** Success rates for different launch sites were assessed, alongside their proximity to geographic features.
- **Visualization:** Charts and maps highlighted the most successful launch sites and payload ranges associated with successful outcomes.
- **Predictive Modeling:** Multiple machine learning models, including logistic regression, support vector machines (SVM), decision trees, and K-nearest neighbors (KNN), were developed to forecast landing outcomes.

KEY RESULTS

- **Exploratory Data Analysis:** Launch success has improved over time, with KSC LC-39A achieving the highest success rate among sites and specific orbits (ESL1, GEO, HEO, SSO) showing 100% success rates.
- **Visualization/Analytics:** Launch sites are strategically located near the equator and close to the coast for optimal efficiency and safety.
- **Predictive Analytics:** All predictive models performed comparably, with the SVM model showing a slight edge in test set performance.

INTRODUCTION

CONTEXT

SpaceX is revolutionizing space exploration by making it more accessible and cost-effective. Its groundbreaking achievements include transporting cargo to the International Space Station, deploying a satellite network for global internet coverage, and executing manned space missions. The key to SpaceX's affordability lies in its innovative approach to rocket reuse. While SpaceX's Falcon 9 launches cost approximately $62 million due to reusable first stages, competitors relying on disposable rockets face costs exceeding $165 million per launch. Predicting the likelihood of a successful first-stage landing is crucial to estimating launch costs, enabling SpaceX to maintain its competitive edge.

OBJECTIVES

- Investigate the impact of payload mass, launch sites, flight frequency, and orbital parameters on first-stage landing success.
- Analyze trends in successful landings over time.
- Identify the most effective machine learning model for predicting first-stage landing success (binary classification).

METHODOLOGY

METHODOLOGY

- **Data Collection:** Data was gathered using SpaceX's REST API and supplemented with web scraping techniques to ensure comprehensive coverage.
- **Data Preparation/Wrangling:** The dataset was cleaned and preprocessed by filtering relevant information, addressing missing values, and implementing one-hot encoding to make it suitable for analysis and modeling.
- **Exploratory Data Analysis (EDA):** SQL queries and visualization techniques were used to analyze the data and uncover meaningful patterns.
- **Data Visualization:** Tools like Folium and Plotly Dash were utilized to create interactive and geographical visualizations for deeper insights.
- **Model Development:** Classification models were built to predict first-stage landing outcomes. Models were fine-tuned and evaluated to identify the best-performing algorithm and optimal parameters.

DATA COLLECTION – API

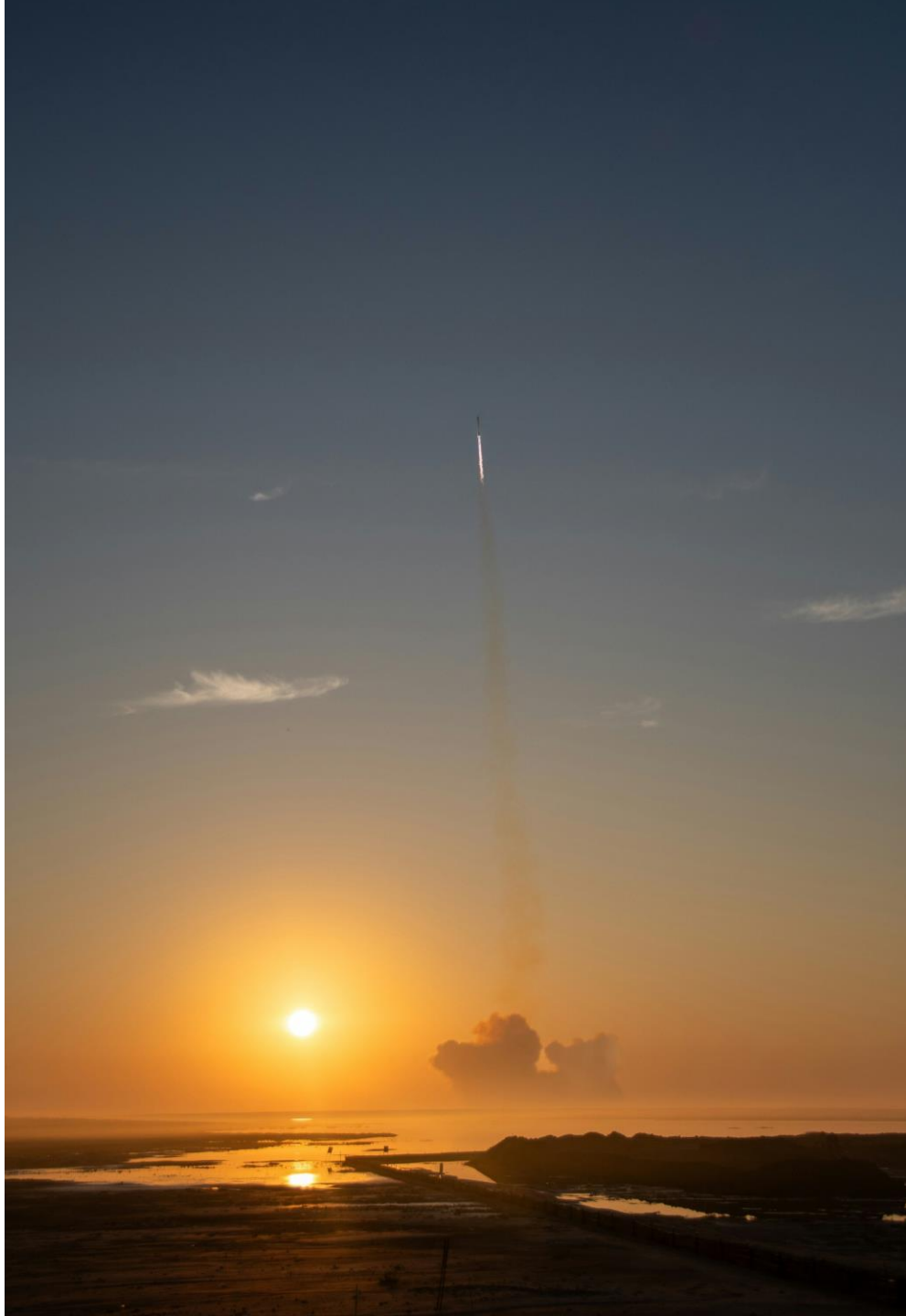
- **Retrieve Data:** Accessed SpaceX's API to request detailed rocket launch information.
- **Process Response:** Parsed the API responses using .json() and converted the data into a structured format using .json_normalize.
- **Custom Functions:** Leveraged custom-built functions to fetch additional launch details from the SpaceX API.
- **Data Structuring:** Organized the fetched data into a dictionary and subsequently transformed it into a Data Frame.
- **Filter Criteria:** Refined the Data Frame to include only Falcon 9 launches.
- **Handle Missing Values:** Addressed missing payload mass values by replacing them with the dataset's calculated mean.
- **Save Data:** Exported the cleaned and filtered dataset to a CSV file for further analysis.

DATA COLLECTION – WEB SCRAPING

- **Data Retrieval:** Retrieved Falcon 9 launch data directly from Wikipedia.
- **HTML Parsing:** Constructed a BeautifulSoup object to parse the HTML content of the webpage.
- **Column Extraction:** Identified and extracted table headers to define the column names.
- **Data Parsing:** Processed the HTML tables to gather the relevant launch data.
- **Data Structuring:** Organized the extracted data into a dictionary and converted it into a Data Frame.
- **Save Data:** Exported the resulting Data Frame to a CSV file for subsequent analysis.

DATA PREPARATION/WRANGLING

- **Assess Missing Data:** Calculated the percentage of missing values for each attribute in the dataset.
- **Categorize Columns:** Classified columns as either numerical or categorical based on their data types.
- **Analyze Launch Sites:** Examined the frequency of launches conducted at each launch site.
- **Orbit Distribution:** Visualized the distribution of orbit types present in the dataset.
- **Outcome Analysis:** Investigated launch outcomes and grouped them into binary categories (success or failure).
- **Target Variable Creation:** Created a new column labeled "Class" to represent the binary outcome, which serves as the target variable for model training.



EDA WITH DATA VISUALIZATION

- **Visualize Relationships:** Explored key variable relationships to uncover insights:
  - Analyzed the correlation between flight number and launch outcome.
  - Examined how flight number varies by launch site.
  - Investigated the relationship between payload mass and launch site.
  - Studied the connection between orbit type and launch outcomes.
  - Explored the relationship between flight number and orbit.
  - Assessed the interaction between payload mass and orbit type.
  - Tracked the yearly success rate to identify trends over time.
- **Categorical Expansion:** Transformed categorical variables into multiple "dummy" columns to facilitate analysis.
- **Data Type Conversion:** Converted numerical columns into the float64 data type for compatibility with analysis tools and models.

EDA WITH SQL

- **Unique Launch Sites:** Queried the dataset to list all distinct launch sites.
- **Launch Site Filter:** Retrieved 5 records where the launch site names start with 'CCA'.
- **Total Payload by NASA (CRS):** Calculated the total payload mass for boosters launched by 'NASA (CRS)'.
- **Average Payload for Falcon 9 v1.1:** Determined the average payload mass carried by the Falcon 9 v1.1 booster.
- **First Ground Landing Success:** Identified the date of the first successful ground landing outcome.
- **Drone Ship Success with Payload Range:** Listed booster versions that had successful landings on drone ships with payloads between 4000kg and 6000kg.
- **Mission Outcomes:** Counted the number of successful and failed mission outcomes.
- **Boosters with Max Payload:** Listed all booster versions that carried the maximum payload mass.
- **Failure Outcomes in 2015:** Extracted details (month name, outcome, booster version, and launch site) for missions in 2015 that failed to land on a drone ship.
- **Outcome Distribution:** Analyzed the distribution of mission outcomes between June 4, 2010, and March 20, 2017.

INTERACTIVE MAP WITH FOLIUM

Markers Indicating Launch Sites:

- Placed a blue circle marker at the coordinates of NASA Johnson Space Center, with a popup displaying its name using its latitude and longitude.
- Added red circle markers at all launch site coordinates, each featuring a popup with the respective launch site names.

Colored Markers for Launch Outcomes:

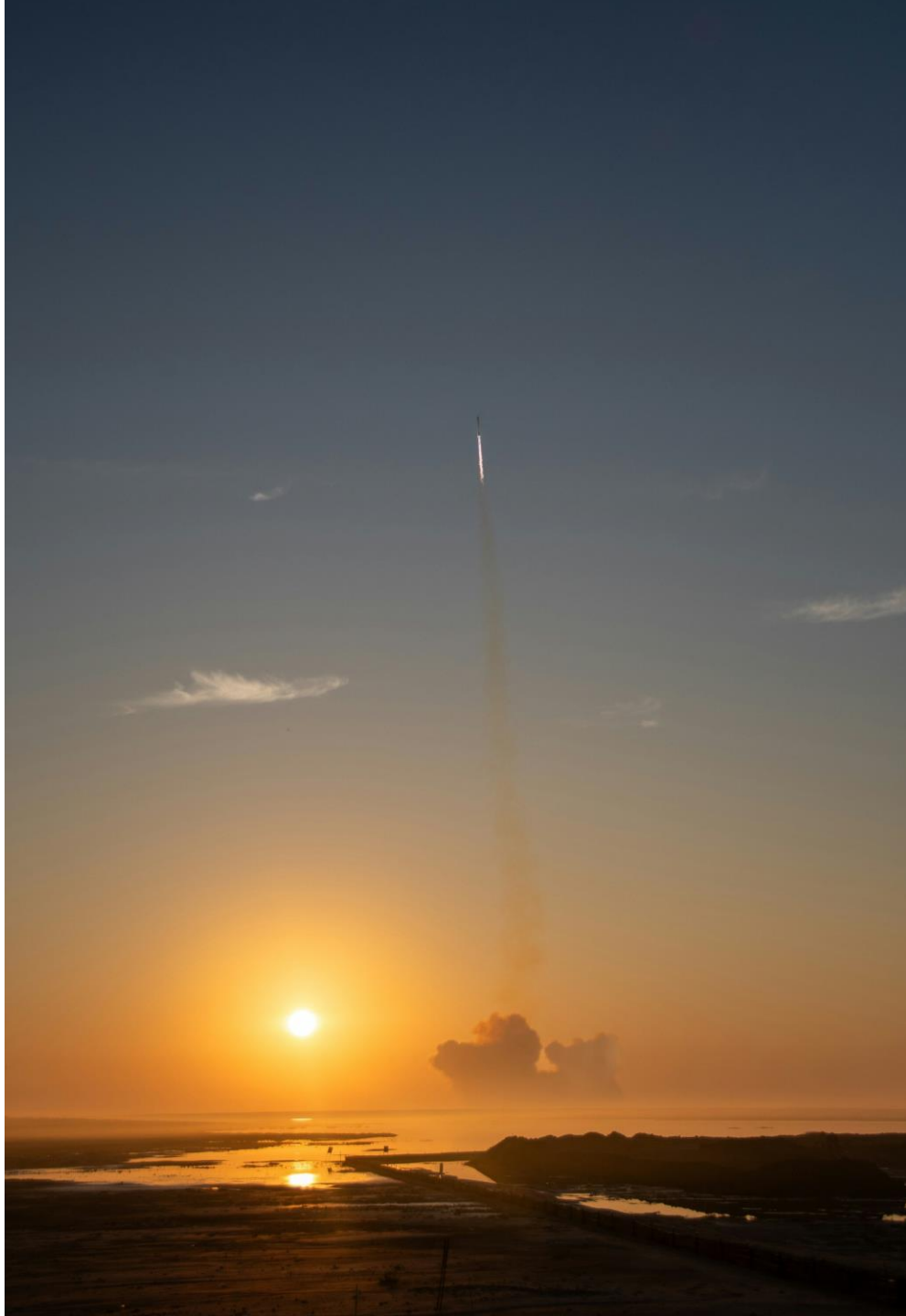
- Displayed green markers for successful launches and red markers for unsuccessful launches at each launch site, highlighting sites with higher success rates.

Distances to Proximities:

- Drew colored lines to indicate the distances from the CCAFS SLC-40 launch site to nearby landmarks, including the nearest coastline, railway, highway, and city.

DASHBOARD WITH PLOTLY DASH

- **Dropdown for Launch Sites:** Enables users to select either all launch sites or a specific one for analysis.
- **Payload Mass Range Slider:** Allows users to filter and adjust the payload mass range dynamically.
- **Pie Chart for Launch Outcomes:** Displays the percentage of successful vs. unsuccessful launches as part of the total.
- **Scatter Plot for Payload Mass and Success Rate:** Visualizes the correlation between payload mass and launch success by booster version.



PREDICTIVE ANALYSIS (CLASSIFICATION)

- **Data Preparation:** Created a NumPy array from the "Class" column and standardized the dataset using *StandardScaler* for fitting and transforming the data.
- **Data Splitting:** Split the dataset into training and testing sets using *train_test_split.*
- **Model Optimization:** Created a *GridSearchCV* object with cross-validation (cv=10) for hyperparameter tuning.
- **Algorithm Application:** Applied *GridSearchCV* on various algorithms, including Logistic Regression (*LogisticRegression()*), Support Vector Machine (*SVC()),* Decision Tree (*DecisionTreeClassifier()*), and K-Nearest Neighbors (*KNeighborsClassifier()*).
- **Performance Evaluation:** Calculated accuracy for each model on the test data using the *.score()* method and assessed confusion matrices.
- **Model Selection:** Identified the best-performing model based on metrics such as Jaccard Score, F1 Score, and Accuracy.

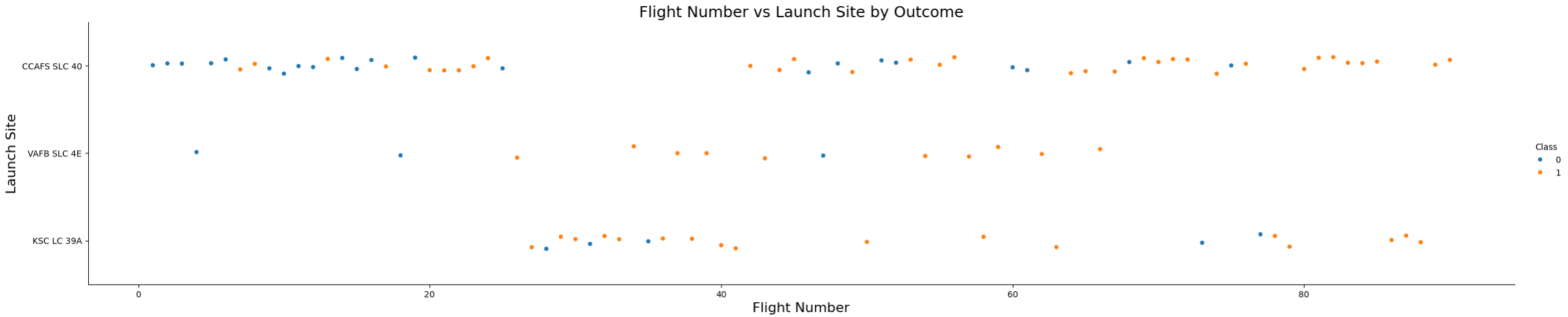
RESULTS

RESULTS SUMMARY

- **Exploratory Data Analysis:** Launch success has steadily improved over time, with **KSC LC-39A** achieving the highest success rate among all launch sites. Specific orbits such **as ESL1, GEO, HEO**, and **SSO** demonstrated a 100% success rate.
- **Visual Analytics:** Most launch sites are strategically located near the equator and along the coast, ensuring safety from potential failed launches while remaining accessible for logistical support.
- **Predictive Analytics:** The **SVM model** emerged as the best-performing predictive model for the dataset, although performance difference between the various models is not large.

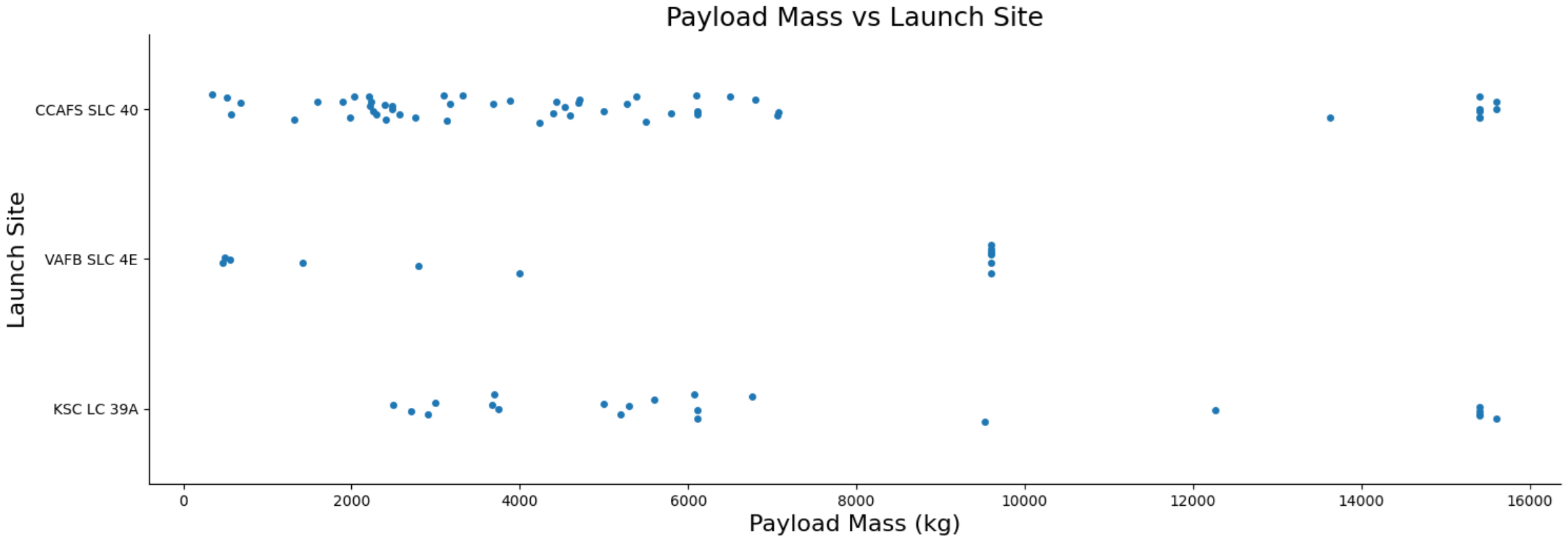
FLIGHT NUMBER VS LAUNCH SITE

- **Success Rates:** Early flights had lower success rates (blue = fail), while later flights were more successful (orange = success).
- **Launch Sites:** About half of the launches were from CCAFS SLC-40, with VAFB SLC-4E and KSC LC-39A showing higher success rates.
- **Trend:** Newer launches demonstrate improved success rates.



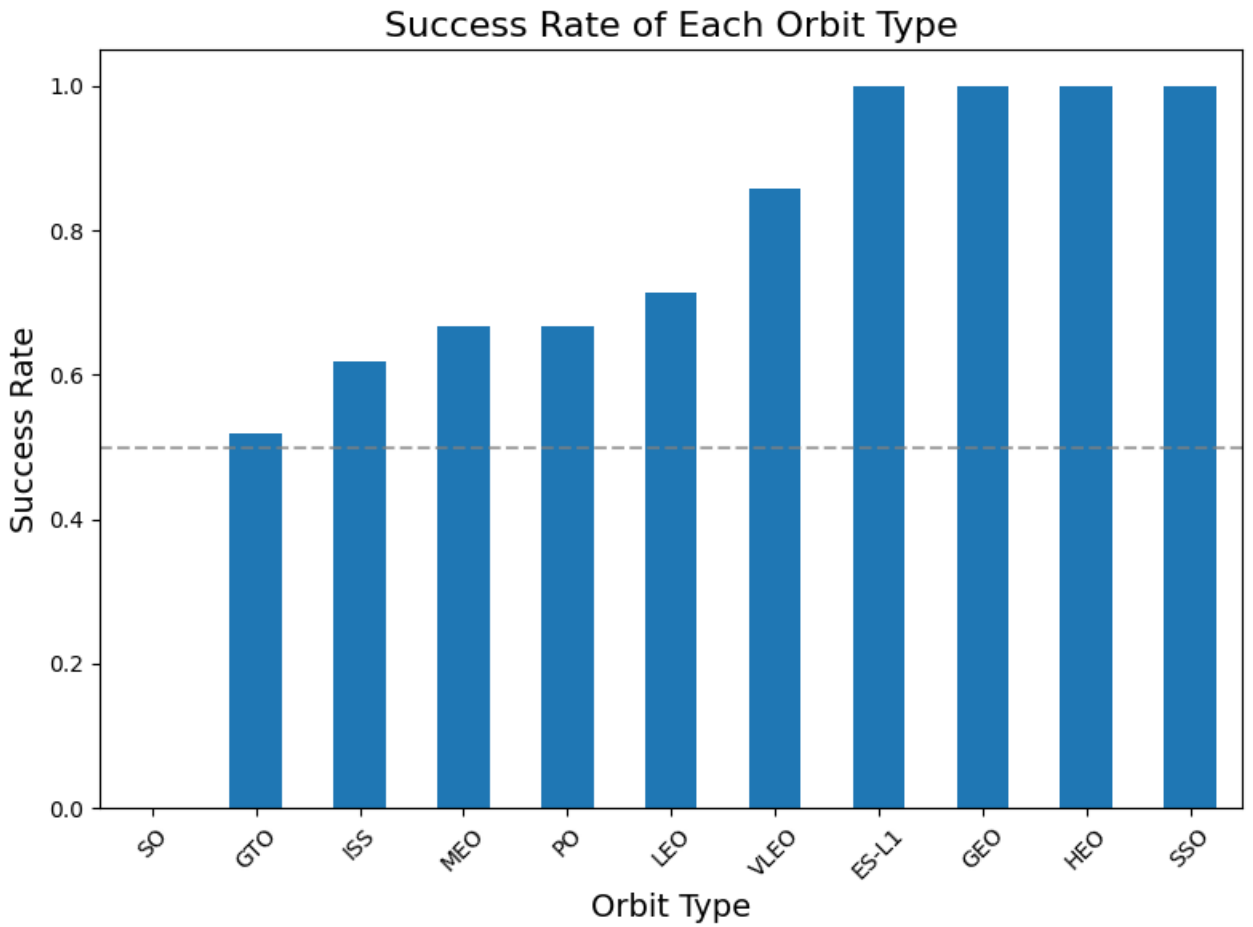
PAYLOAD VS LAUNCH SITE

- **Payload and Success:** Higher payload masses generally correlate with higher success rates.
- **Success Threshold:** Most launches carrying payloads over 7,000 kg were successful.
- **Site-Specific Performance:** KSC LC-39A: Achieved a 100% success rate for payloads under 5,500 kg. VAFB SLC-4E: Has not launched payloads exceedingly approximately 10,000 kg.



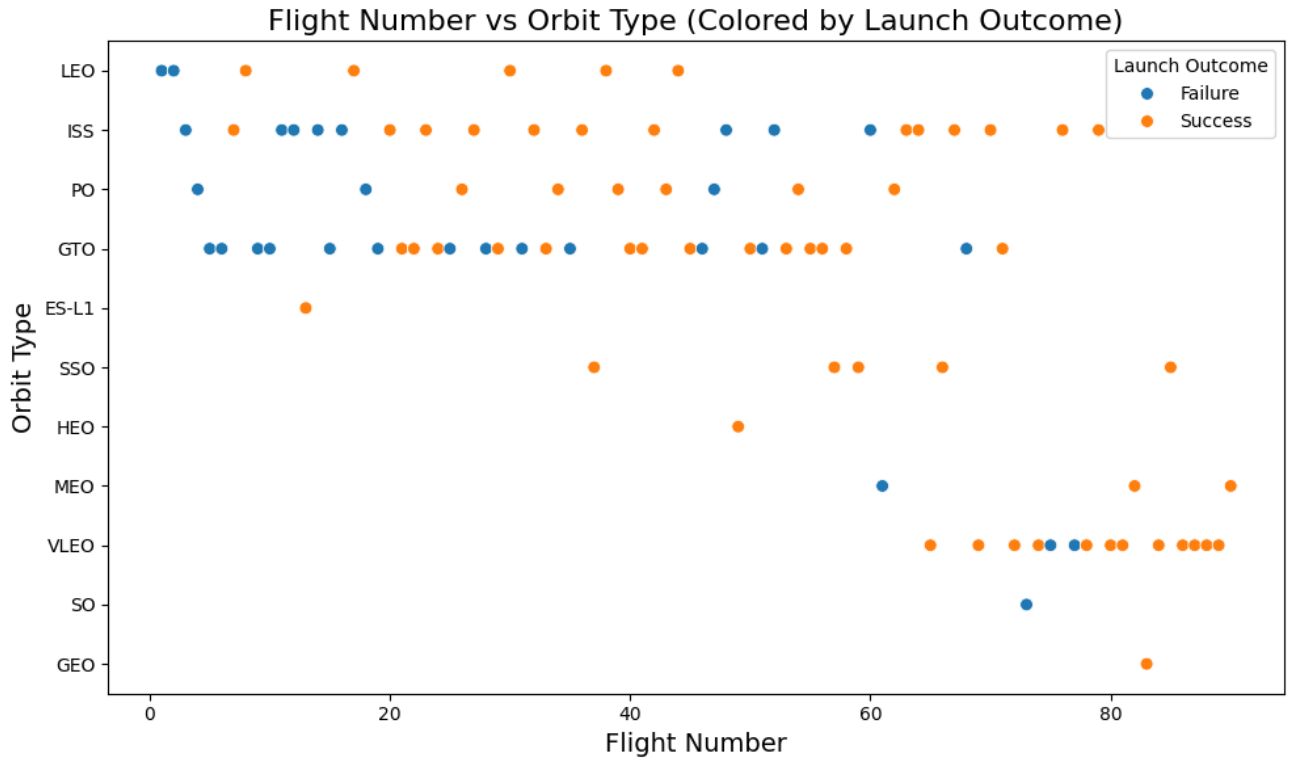
SUCCESS RATE BY ORBIT

- **100% Success:** Orbits like ESL1, GEO, HEO, and SSO achieved perfect success rates.
- **Moderate Success (50–80%):** Orbits including GTO, ISS, LEO, MEO, and PO showed varied success rates.
- **No Success:** Orbit SO recorded a 0% success rate.



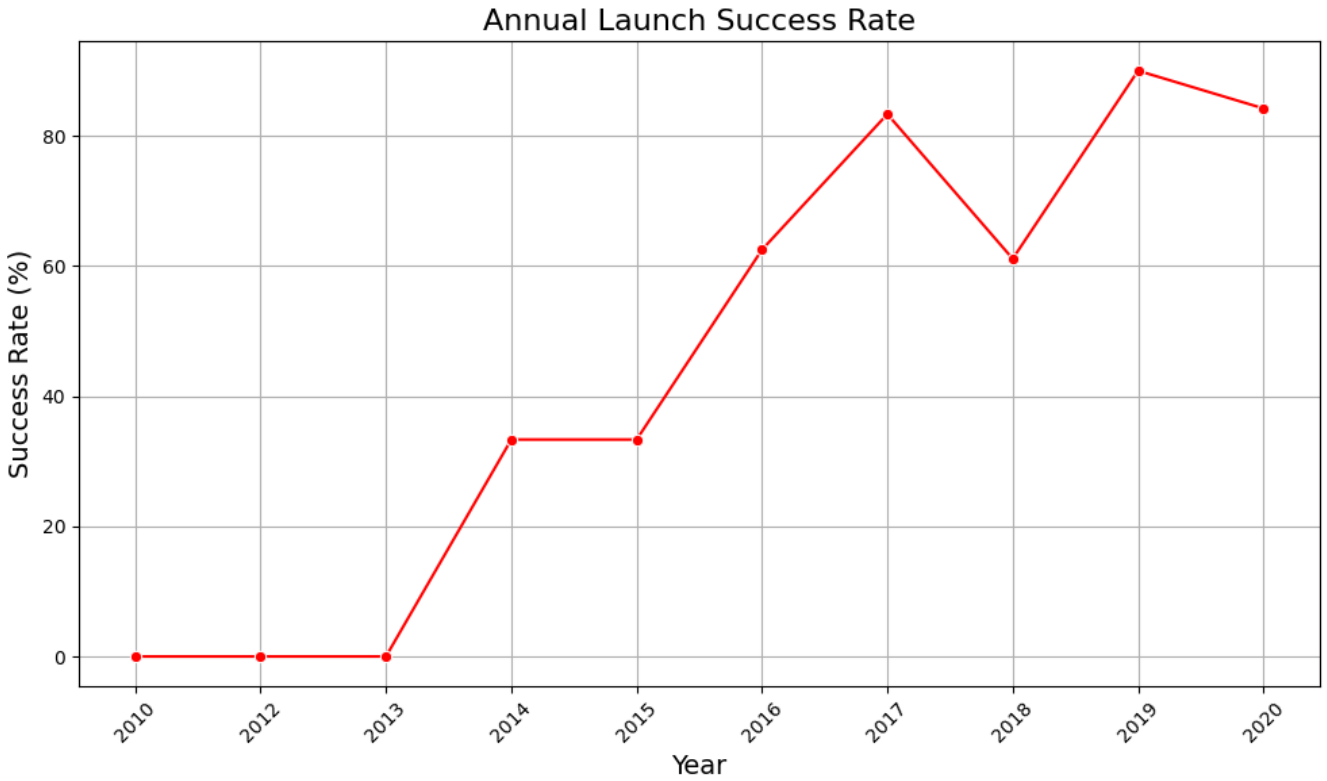
FLIGHT NUMBER VS ORBIT

- **Overall Trend:** Success rates generally improve with an increasing number of flights for most orbits.
- **LEO Orbit:** Displays a strong correlation between flight number and higher success rates.
- **GTO Orbit:** Deviates from the trend, showing inconsistent success rates.



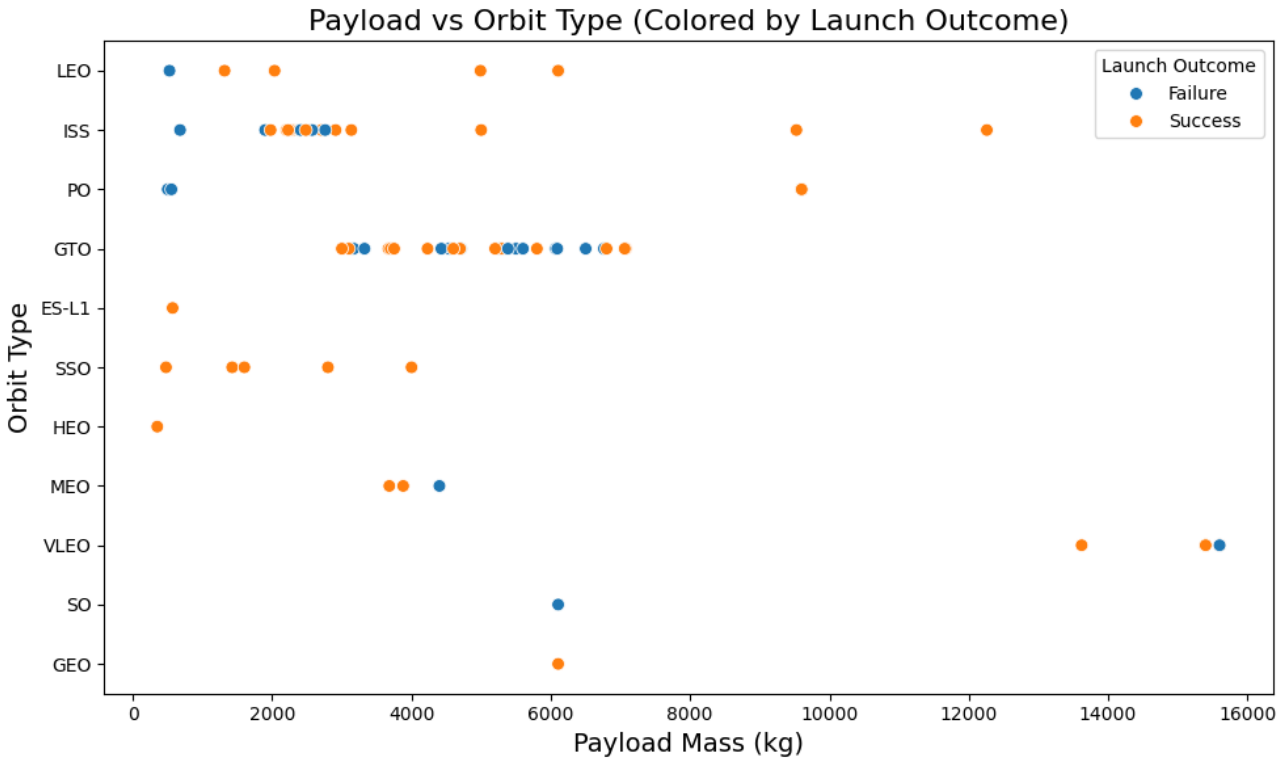
LAUNCH SUCCESS OVER TIME

- **Improvement Periods:** Success rates significantly improved between 2013–2017 and 2018–2019.
- **Decline Period**s: Observed decreases in success rates from 2017–2018 and 2019–2020.
- **Overall Trend:** A clear improvement in launch success rates since 2013.



PAYLOAD VS ORBIT

- **Improvement Periods: Success rates significantly improved between 2013–2017 and 2018–2019.Decline Periods: Observed decreases in success rates from 2017–2018 and 2019–2020.Overall Trend: A clear improvement in launch success rates since 2013.**



LAUNCH SITE NAMES

Launch Site Names

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

```
1 %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

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

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

Launch Site Names Begin with 'CCA'

```
1 %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

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

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

PAYLOAD MASS

- The **total payload** carried by boosters from NASA (CRS) is **45,596kg**.

```
1 %sql SELECT SUM(PAYLOAD_MASS__KG_) AS "Total payload mass by NASA (CRS)" FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

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

Total payload mass by NASA (CRS)
45596

- The **average payload** carried by booster version F9 v1.1 is **2,928.4 kg**

```
1 %sql SELECT AVG(PAYLOAD_MASS__KG_) AS "Average payload mass by Booster Version F9 v1.1" FROM SPACEXTBL WHERE BOOSTER_VERSIC
```

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

Average payload mass by Booster Version F9 v1.1

LANDING AND MISSION INFORMATION

- The first successful landing outcome on ground pad occurred on **December 22nd, 2015**

```
1 %sql SELECT MIN(DATE) AS "Date of first successful landing outcome in ground pad" FROM SPACEXTBL WHERE Landing_Outcome = 'S
```

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

Date of first successful landing outcome in ground pad
2015-12-22

The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2

```
1 %sql SELECT Booster_Version FROM SPACEXTBL WHERE Landing_Outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 400
```

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

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

LANDING AND MISSION INFORMATION (CONTD.)

- Total Number of Successful and Failed Mission Outcomes are 100 and 1 respectively.

```
1 %sql SELECT number_of_success_outcomes, number_of_failure_outcomes FROM (SELECT COUNT(*) AS number_of_success_outcomes FROM
```

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

number_of_success_outcomes	number_of_failure_outcomes
100	1

LANDING AND MISSION INFORMATION (CONTD.)

```
1 %sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

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

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

The boosters that carried **maximum payload** are:

- 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

LANDING AND MISSION INFORMATION (CONTD.)

- List of failed landing outcomes in drone ship, their booster versions, and launch site for the year **2015**:

```
%sql SELECT CASE substr(Date, 6, 2) WHEN '01' THEN 'January' WHEN '02' THEN 'February' WHEN '03' THEN 'March' WHEN '04' THE
```

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

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

- Ranks of the count of landing outcomes, such as "Failure (drone ship)" or "Success (ground pad)," between the dates **2010-06-04 and 2017-03-20** in descending order.

```
%sql SELECT Landing_Outcome, COUNT(Landing_Outcome) AS Landing_Count FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2
```

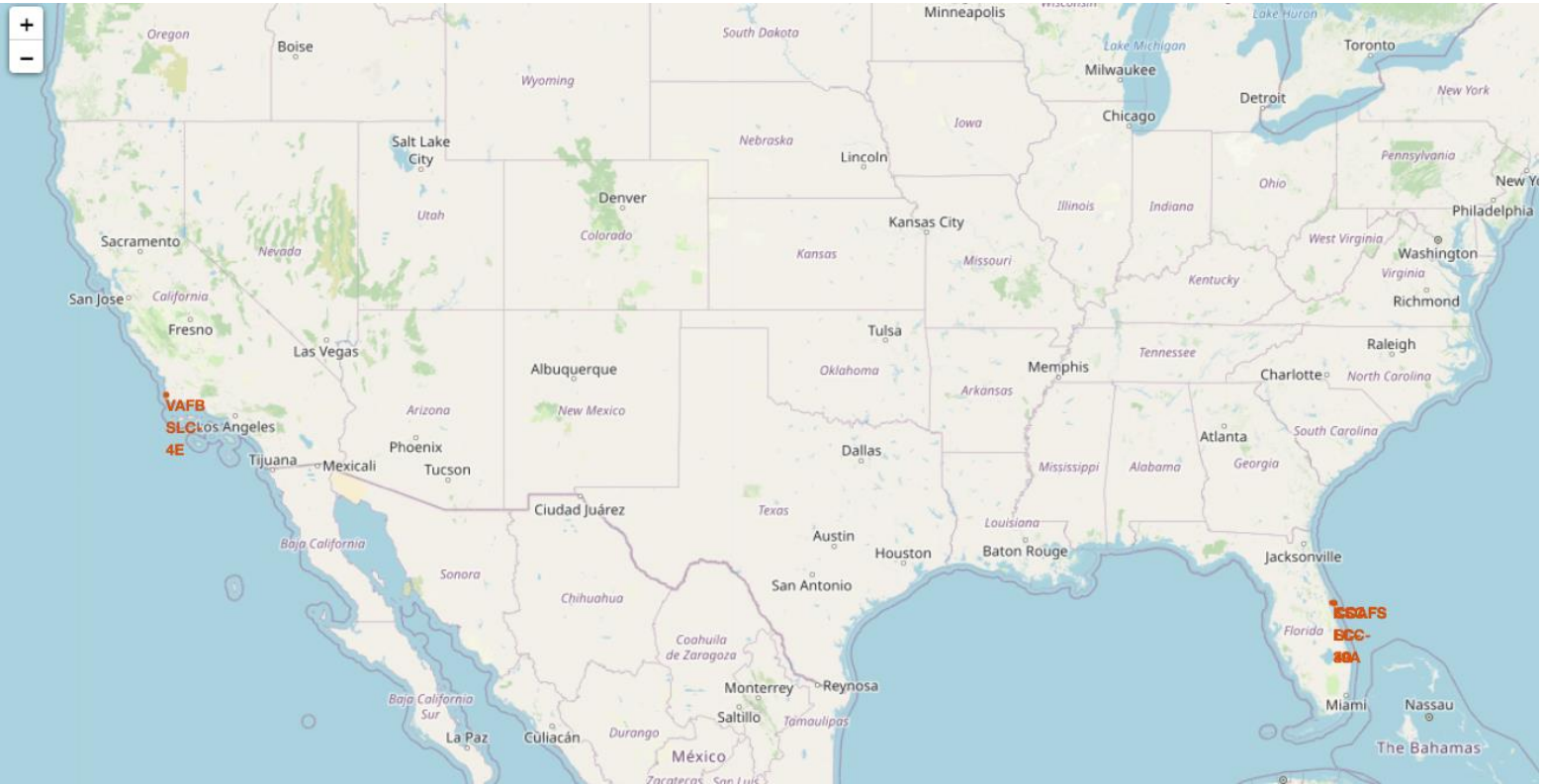
```
 * sqlite:///my_data1.db
Done.
```

Landing_Outcome	Landing_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

LAUNCH SITE ANALYSIS

LAUNCH SITES

- **Proximity to the Equator:** Launch sites located closer to the equator benefit from Earth's rotational speed, providing a natural boost for prograde orbits. This advantage makes it easier to reach equatorial orbits while reducing the need for additional fuel and boosters, ultimately saving costs.



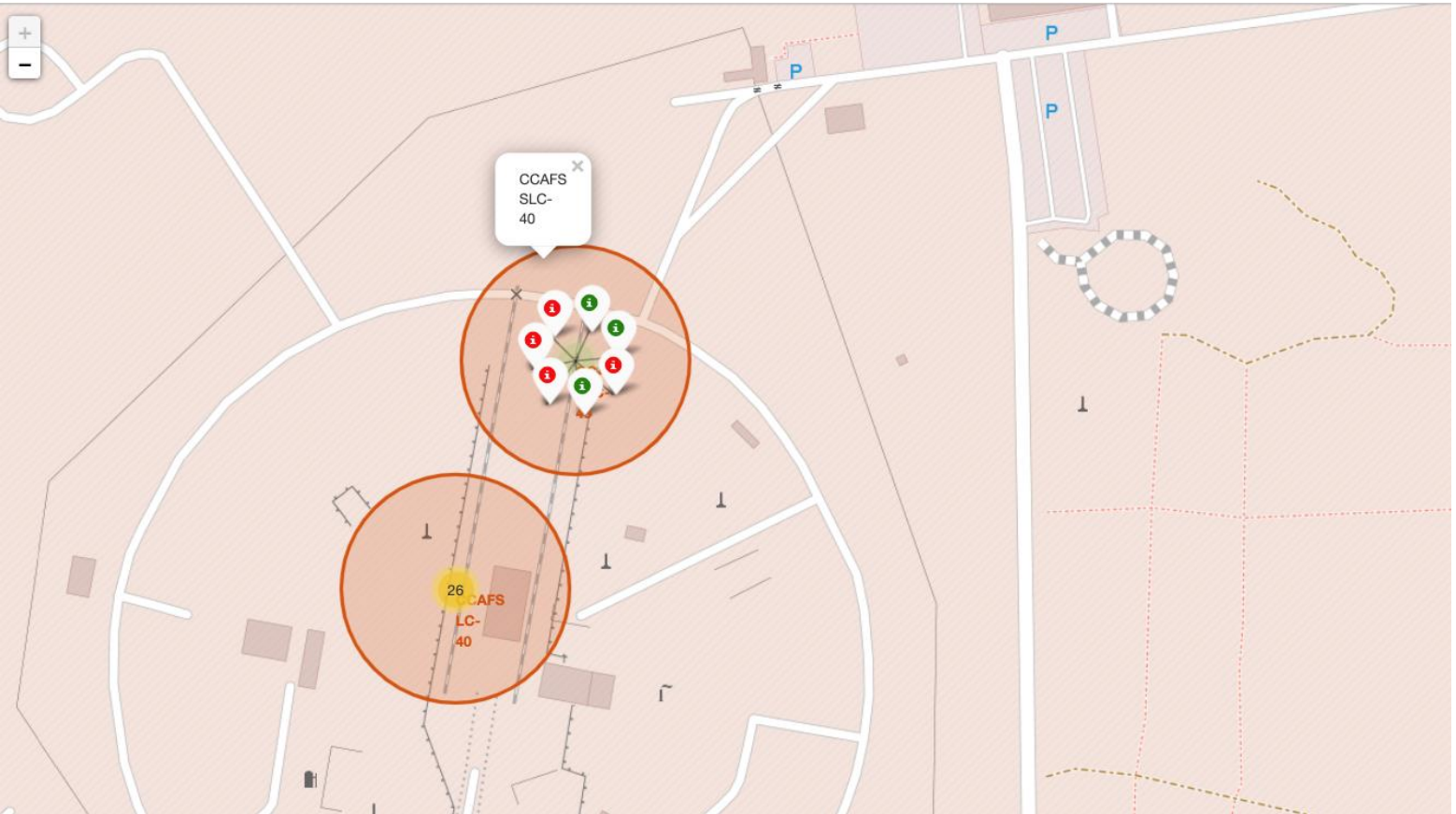
LAUNCH OUTCOMES

Outcomes:

- **Green** markers represent successful launches.
- **Red** markers indicate unsuccessful launches.

Launch Site Performance:

- **CCAFS SLC-40** achieved a success rate of 3 out of 7 launches (42.9%).



PROXIMATE LOCATIONS

CCAFS SLC-40 Proximities

- Distance to Nearest Coastline: **0.86 km**
- Distance to Nearest Railway: **21.96 km**
- Distance to Nearest City: **23.23 km**
- Distance to Nearest Highway: **26.88 km**

DASHBOARD WITH PLOTLY DASH

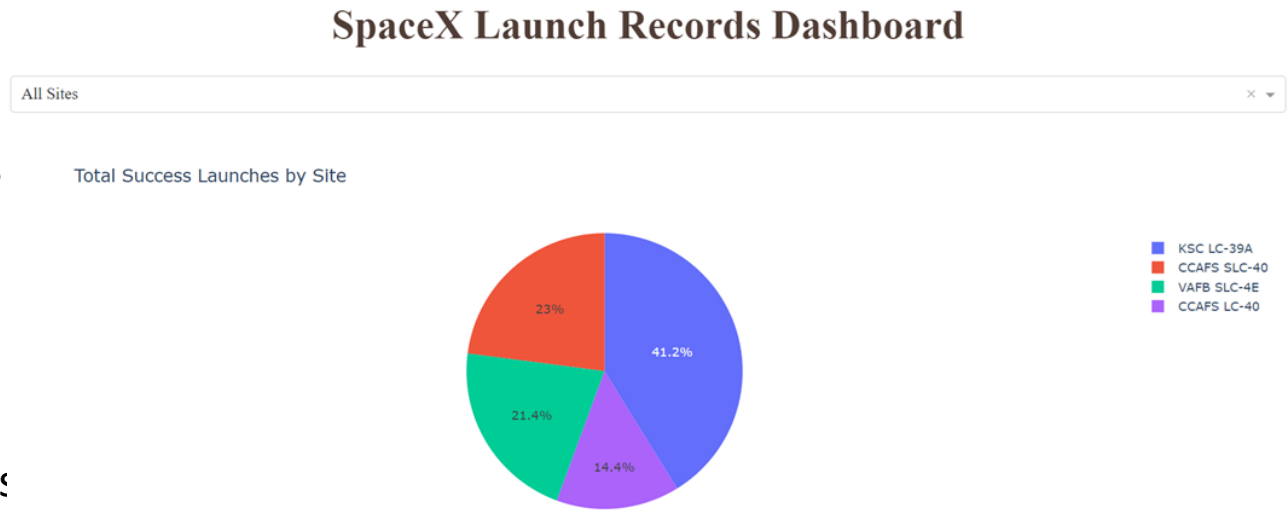
LAUNCH SUCCESS BY SITE

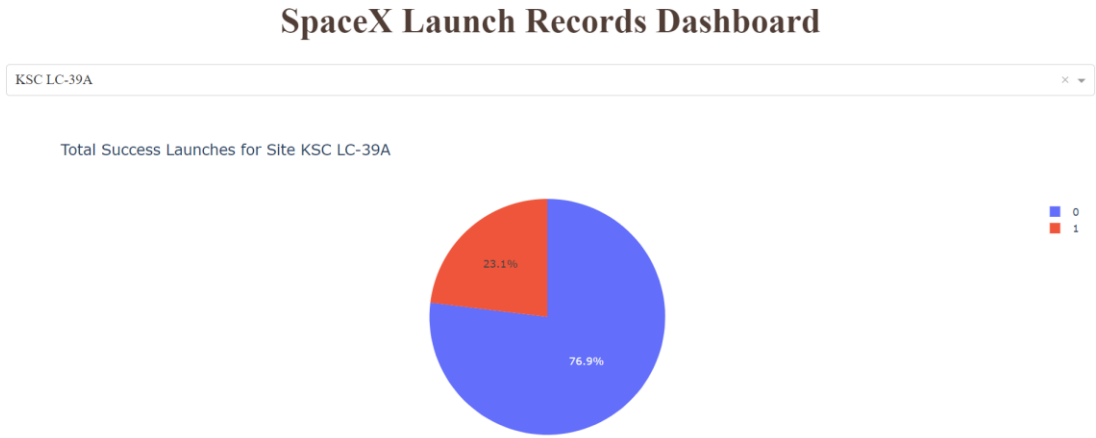
Success as Percent of Total:

- **KSC LC-39A** has the highest percentage of successful launches among all sites, achieving **41.2%** success.

Success as Percent of Total:

- **KSC LC-39A** boasts the highest success rate among all launch sites at 76.9%.
- This includes **10 successful launches** and **3 failed launches**.





PAYLOAD MASS AND SUCCESS

Success by Booster Version:

- **Payload Range Success:** Payloads between 2,000 kg and 5,000 kg achieve the highest success rates.
- **Outcome Representation:** A value of 1 represents a successful outcome, while 0 indicates an unsuccessful outcome.

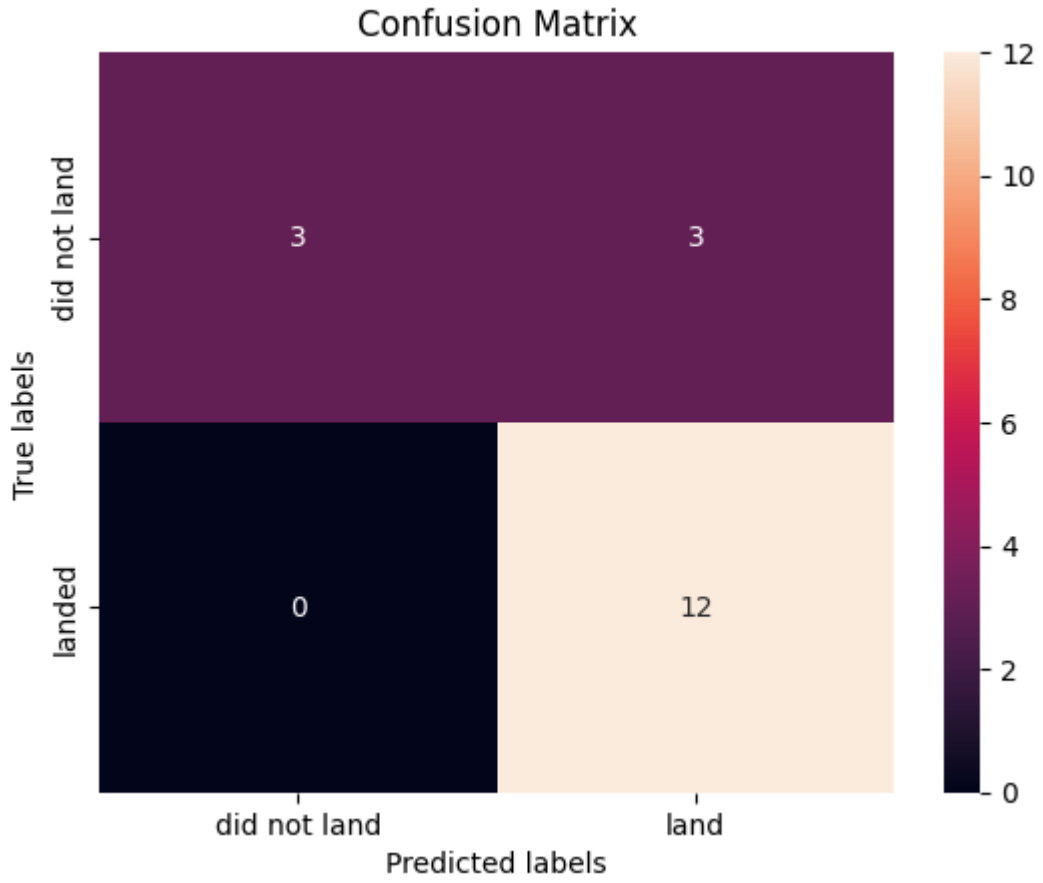
PREDICTIVE ANALYTICS (CLASSIFICATION)

CONFUSION MATRIX - SVM

The Confusion Matrix Analysis for SVM Model shows:

- **True Positives (Landed Successfully):** 12 landings were correctly predicted as successful.
- **True Negatives (Did Not Land):** 3 instances were correctly predicted as failures.
- **False Positives (Predicted Landed but Did Not Land):** 3 instances were incorrectly classified as successful landings.
- **False Negatives (Predicted Did Not Land but Landed):** None of the landings were incorrectly predicted as failures.

This indicates that the **SVM model** demonstrates strong performance with no false negatives, though there is room for improvement in minimizing false positives.



CONCLUSION

- **SVM Model Performance**: As seen above, the SVM model demonstrated the best performance among all models, achieving the highest accuracy, Jaccard Score, and F1 Score.
- **Equatorial Advantage**: Launch sites near the equator benefit from Earth's rotational speed, reducing the need for additional fuel and boosters.
- **Coastal Proximity**: All launch sites are strategically located near the coast for logistical efficiency and safety.
- **Launch Success Trends**: Success rates have steadily increased over time, reflecting ongoing improvements in operations and technology.
- **Top Launch Site**: KSC LC-39A emerged as the most successful launch site, achieving a 100% success rate for payloads under 5,500 kg.
- **Orbit Performance**: Specific orbits such as ES L1, GEO, HEO, and SSO consistently achieved 100% success rates.
- **Payload Impact**: Higher payload masses generally correspond to higher success rates across all launch sites.
- **Predictive Insights**: Launch site and orbit type were highly predictive factors for success, with SVM excelling in modeling outcomes accurately.

ANNEX

Link to GitHub Repository:

- https://github.com/binyammesfin/IBM-Applied-Data-Science-Capstone/upload/main


THANK YOU