IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

This project employed a rigorous, multi-stage data science methodology to address the prediction challenge:

1. **Data Acquisition and Preparation:** Launch data was collected by accessing the **SpaceX REST API** and leveraging **web scraping** techniques on Wikipedia launch records to ensure a comprehensive dataset. This raw data underwent extensive **data wrangling**, including cleaning, standardizing formats, handling missing values (e.g., using mean imputation for Payload Mass), and applying **One Hot Encoding** to prepare features for binary classification modeling.

2. **Exploratory Data Analysis (EDA):** Preliminary insights were derived using classical visualization tools (Matplotlib, Seaborn) and precise **SQL queries** to calculate metrics such as total payload mass and success rates by site and rocket type.

3. **Interactive Visual Analytics:** To enhance stakeholder insight, interactive visualization tools were developed:
    1. **Folium** was used to map launch sites, visually analyzing geographical patterns, proximity to the Equator, and the coastline.
    2. **A Plotly Dash** application was built featuring interactive components (dropdowns, range sliders) to dynamically analyze launch success rates based on specific launch sites and payload ranges.
4. **Predictive Modeling:** Classification models, including **Support Vector Machine (SVM)**, **K-Nearest Neighbors (KNN)**, and **Decision Trees**, were built and evaluated. Models were subjected to rigorous **hyperparameter tuning** using **GridSearchCV** to maximize predictive accuracy and robustness.

The project outcomes confirmed key operational dependencies and established a highly accurate predictive model for launch success.

I. Data Insights

• Key factors influencing the success of Falcon 9 first stage landings were clearly identified.
• The analysis successfully visualized geographical patterns and corresponding launch success rates using interactive tools.

II. Model Performance

• The Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) models achieved a classification accuracy of **83.33%** on the test data.
• The Decision Tree model demonstrated superior performance, achieving a classification accuracy of **94.44%** on the test data.

III. Key Findings

• Landing success is significantly impacted by mission parameters, primarily the **launch site and payload mass**.
• The **Decision Tree model** was confirmed as the most effective predictor for the successful landing of the Falcon 9 first stage.

Introduction

Project Background and Context

The project is predicated on the operational strategy of SpaceX, which maintains significantly lower launch costs (e.g., $62 million compared to competitors' costs upward of $165 million) primarily due to the ability to **reuse the Falcon 9 first stage**. The objective is to utilize public data and machine learning models to accurately **predict the successful landing outcome** of the first stage, thereby enabling precise launch cost estimation and providing valuable insights for competitive market analysis.

Problems We Want to Find Answers To

This analysis addressed three core strategic questions:

1. **Influencing Factors:** What specific operational variables (such as payload mass, launch site, number of flights, and orbit type) fundamentally affect the success of the first stage landing?.
2. **Predictive Accuracy:** How can machine learning models be utilized to accurately predict the binary landing outcome (success or failure)?.
3. **Model Optimization:** Which classification algorithm (e.g., Decision Tree, SVM, KNN) yields the highest predictive accuracy and is best suited for this specific dataset?.

Section 1

Methodology

Methodology

Data Collection Methodology:

Data was comprehensively sourced by initiating API requests to the **SpaceX REST API** and leveraging **Web Scraping** techniques on Wikipedia launch records to obtain detailed launch parameters and outcomes.

Data Wrangling and Processing:

Data wrangling involved cleaning, filtering, and handling missing values (e.g., replacing missing Payload Mass with the calculated mean). The data was processed by standardizing features and applying **One Hot Encoding** to prepare the dataset for binary classification modeling.

Exploratory Data Analysis (EDA):

EDA utilized visualization tools (Matplotlib and Seaborn) to visually analyze success rates, payloads, and launch sites. **SQL queries** were executed to derive precise, quantitative insights regarding mission outcomes, total payload mass, and site success rates.

Interactive Visual Analytics:

Interactive analysis was performed using **Folium** to map launch site geographical locations, utilizing colored markers to denote success/failure rates. A dynamic dashboard was developed with **Plotly Dash**, incorporating interactive components like dropdowns and sliders to analyze success rates against payload ranges.

Predictive Analysis using Classification Models:

Multiple classification models, including **Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Trees**, were built and evaluated to predict landing success.

Building, Tuning, and Evaluation:

Models were optimized using **GridSearchCV** for rigorous hyperparameter tuning. Performance was evaluated based on test data **accuracy** and cross-validation techniques to ensure robustness, ultimately identifying the best-performing model

Data Collection

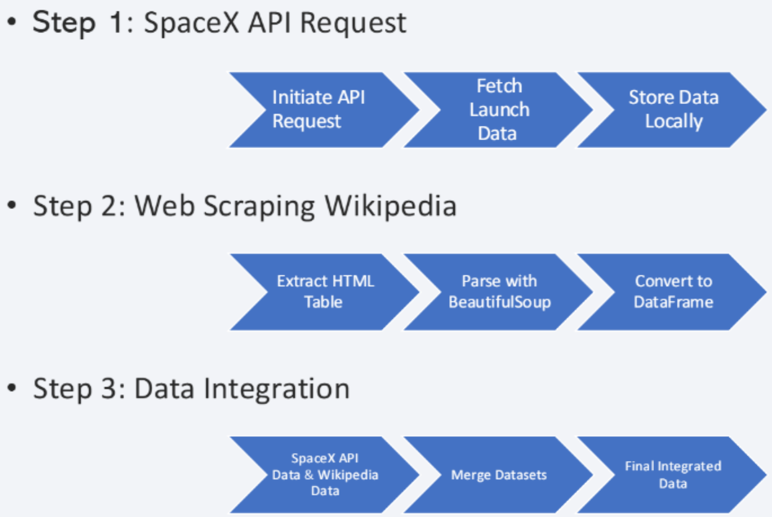
Data Collection Methodology

The data collection process employed a robust, dual-source strategy to ensure a complete and detailed record set of Falcon 9 launches.

Key Phrases and Process Description:
Data was sourced from a **combination of API requests and Web Scraping** to obtain comprehensive information necessary for detailed analysis.

1. **SpaceX API Request:** We utilized Python's requests library to connect to the **SpaceX REST API** (using the /v4/launches endpoint). The response was parsed from **JSON** format, and relevant fields (including launch dates, sites, payload mass, and outcomes) were extracted and stored into a **pandas DataFrame**.

2. **Web Scraping Wikipedia:** The data was augmented by **Web Scraping** the HTML content from the relevant Wikipedia page (List of Falcon 9 and Falcon Heavy launches). The **BeautifulSoup** library was used to parse the HTML, extract the launch records table, and convert the data into a structured **pandas DataFrame**.
This combined approach was essential to gather the complete data needed for detailed analysis.



Data Collection – SpaceX API

SpaceX REST API Data Collection Methodology:

The data collection process began by establishing a programmatic link to the SpaceX public records using the REST API to secure foundational launch data. This method was crucial for obtaining detailed, structured information directly relevant to predictive modeling.

jupyter-labs-spacex-data-collection-api.ipynb

Key Phrases and Process Flow

The methodology utilized a three-step sequence involving Python libraries for initiation, parsing, and storage.

Step	Key Action and Description
Step 1: Initiate Request	The process began by using Python's requests library to connect to the dedicated SpaceX API endpoint (https://api.spacexdata.com/v4/launches).
Step 2: Fetch and Parse Data	The API response was fetched, decoded from JSON format using .json(), and subsequently parsed into a Python dictionary or structured object. Essential fields, including launch date, launch site, payload mass, rocket type, and mission outcome, were extracted.
Step 3: Store and Filter	The extracted information was immediately transformed and stored into a pandas DataFrame. The data was filtered to include only Falcon 9 launches and subsequently saved locally for integration with web-scraped data.

Data Collection - Scraping

Web Scraping Process Methodology

To complement the data retrieved from the SpaceX API, **Web Scraping** was performed on the official Wikipedia launch records table to obtain complete mission data, ensuring comprehensive input for analysis.

Key Phrases and Process Flow

The methodology utilized established Python libraries to systematically extract and structure the historical launch data.

Steps	Key Action and Description
Step 1: Initiate Web Scraping	The requests library was used to fetch the HTML content of the target Wikipedia page (List of Falcon 9 and Falcon Heavy launches).
Step 2: Parse HTML Content	The fetched HTML was parsed using the BeautifulSoup library, and the relevant HTML table containing the Falcon 9 launch records was successfully extracted.
Step 3: Convert and Structure Data	The extracted HTML table data was converted into a structured pandas DataFrame. The resulting DataFrame was cleaned and formatted to ensure data consistency across all columns.

jupyter-labs-webscraping.ipynb

Data Wrangling

Data Wrangling and Processing Methodology

The data processing phase, referred to as Data Wrangling, was essential for converting raw, integrated data into a structured, analytical format suitable for machine learning. This process involved systematic cleaning, transformation, and standardization.

labs-jupyter-spacex-Data wrangling.ipynb

Key Phase	Description and Actions
Data Cleaning	Involved handling missing values by employing appropriate imputation techniques or removal. Specifically, missing values in the Payload Mass column were replaced with the calculated mean.
Data Transformation	Focused on converting data types to appropriate formats and standardizing formats to ensure consistency. New features were engineered to enrich the dataset for analysis.
Label Creation	Landing outcomes were transformed into Training Labels, where a value of "1" signified a successful landing and "0" signified an unsuccessful outcome.
Preparation for Modeling	Features were standardized to ensure equal contribution during modeling. One Hot Encoding was applied to prepare the categorical data specifically for binary classification.
Data Validation	Final steps included checking for and removing duplicate records and ensuring the overall accuracy and consistency of the data entries.

EDA with Data Visualization

Charts Plotted and Rationale

A range of statistical charts were employed during the Exploratory Data Analysis (EDA) phase to visually inspect data characteristics, identify patterns, and determine correlations among key variables.

Chart Type	Purpose and Utility
Scatter Plots	Used to explore relationships between two numerical variables (e.g., Payload Mass vs. Launch Success, Flight Number vs. Launch Site). Rationale: These plots are essential for identifying correlations or dependencies that could be leveraged in machine learning models.
Bar Charts	Employed for comparison among discrete categories (e.g., Success Rate vs. Orbit Type, or Launch Outcome per Site). Rationale: Bar charts clearly illustrate the relationship between specific categories and a measured value, highlighting comparative performance.
Line Charts	Utilized to track trends in launch success over time (e.g., Launch Success Yearly Trend). Rationale: Line charts effectively reveal temporal patterns and help in understanding performance improvements or changes across specific periods.
Histograms & Box Plots	Used to visualize the distribution of numerical features (e.g., Payload Mass, Flight Number). Rationale: These charts aid in understanding data spread, central tendency, skewness, and the presence of outliers.
Heatmaps	Used to visualize correlation matrices between multiple numerical variables. Rationale: This helps in quickly identifying strong correlations (positive or negative) between variables, a crucial step for effective feature selection.

5. jupyter-labs-eda-dataviz.ipynb

EDA with SQL

jupyter-labs-eda-sql-coursera_sqllite.ipynb

The following SQL queries were executed during the Exploratory Data Analysis (EDA) phase to derive specific, quantitative insights from the launch data, which was stored in a Db2 database:

Aggregate and Descriptive Queries

• Unique Launch Sites: Displaying the names of the unique launch sites in the space mission.

• Total Launch Outcomes: Listing the total number of successful and failure mission outcomes.

• Total Payload Mass (NASA): Displaying the total payload mass carried by boosters launched by NASA (CRS).

• Average Payload Mass by Version: Displaying the average payload mass carried by booster version F9 v1.1.

Filtering and Condition-Based Queries

• Specific Launch Sites: Displaying 5 records where launch sites begin with the string 'CCA'.

• Targeted Success: Listing the names of the boosters which had success in drone ship and had payload mass greater than 4000 but less than 6000.

• Yearly Failures: Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in the year 2015.

Ordering and Ranking Queries

• First Successful Landing: Listing the date when the first successful landing outcome in ground pad was achieved.

• Maximum Payload Boosters: Listing the names of the booster versions which have carried the maximum payload mass.

• Ranked Outcomes (Timeframe): Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the specific date range of 2010-06-04 and 2017-03-20 in descending order.

Build an Interactive Map with Folium

The following interactive objects were incorporated into the Folium map:

1. Markers: Individual markers were placed to pinpoint the exact geographical location of all launch sites. These markers were further color-coded (Green for Success, Red for Failure) and organized using a Marker Cluster feature to visualize the launch outcome distribution at each site.

2. Circles: Circles were added around launch site markers to visually represent proximity zones.

3. Lines (PolyLine): Colored PolyLines were drawn from specific launch sites (e.g., KSC LC-39A) to their proximities, including the coastline, railway, highway, and the closest city (Titusville).

4. Rationale for Adding Objects

5. These map objects served distinct analytical and operational purposes:

6. Markers (Pinpointing Location & Outcome): Markers were added to provide precise spatial reference, allowing users to identify where launches were conducted geographically. Color-coding immediately enabled stakeholders to identify launch sites with relatively high success rates (Green markers).

7. Circles (Safety Perimeter): Circles were used to visually illustrate potential impact zones and provide a visual representation of safety perimeters or operational boundaries around the launch sites.

8. Lines (Distance and Risk Assessment): Lines were added to calculate and display the distance between launch sites and relevant proximities (e.g., coastline). This confirms operational adherence to safety standards, verifying that all sites are located in very close proximity to the coast to minimize risk to populated areas during over-water flight paths.

6. lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Plotly Dash Interactive Dashboard Summary

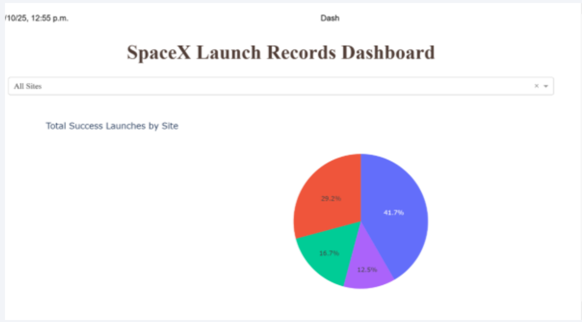
The interactive dashboard was built using Plotly Dash to provide stakeholders with dynamic tools for real-time exploratory analysis, integrating visual analytics with key business parameters.

Plot/Graph	Summary of Content	Rationale for Inclusion
Success Pie Chart	Displays the distribution of successful versus failed launches. When a site is selected, it shows the Success vs. Failed counts for that specific location.	Provides a quick, essential overview of mission overall performance metrics and success rates for immediate stakeholder understanding.
Success-Payload Scatter Plot	Shows the correlation between Payload Mass and launch success, potentially across different Booster Versions.	Helps identify crucial correlations between payload characteristics and launch outcomes, supporting strategic decision-making related to payload planning.

spacex-dash-app.py

Interactions Added and Rationale

Interactive Component	Summary of Functionality	Rationale for Inclusion
Launch Site Dropdown List	Enables users to select and filter data based on specific launch sites.	Facilitates regional insights and comparisons across different geographical locations, enhancing the user experience by focusing the analysis.
Range Slider for Payload	Allows users to dynamically adjust and select specific ranges of Payload Mass.	Offers flexible exploration of how payload mass variation affects mission success, enabling detailed analysis of performance factors related to cargo weight.



Predictive Analysis (Classification)

Predictive Analysis and Model Selection Methodology

The model development process followed a structured methodology designed to ensure optimal predictive performance and robustness for determining Falcon 9 first stage landing success.

SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

I. Model Development Process (Build and Preprocessing)	
Key Phase	Description and Actions
Data Preprocessing	Prior to modeling, features were standardized using StandardScaler to ensure all variables contributed equally to the learning process. The data was subsequently split into training and test sets to facilitate robust model validation.
Model Selection	We built and evaluated several classification algorithms suitable for binary classification, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Trees. Logistic Regression was also included in the evaluation set.

II. Evaluation, Tuning, and Improvement	
Key Phase	Description and Actions
Hyperparameter Tuning	GridSearchCV was employed to systematically tune the parameters across all models (e.g., C for SVM, max_depth for Decision Trees). This approach was used to search for the optimal hyperparameters to maximize predictive accuracy.
Model Evaluation	Models were evaluated using cross-validation techniques to ensure generalizability. The primary metric used for comparison was accuracy on the test set. The confusion matrix was examined to understand the specific types of errors (false positives/negatives).
Improvement and Iteration	The models were iteratively adjusted and fine-tuned based on validation results to maximize predictive accuracy and reliability. Both training and test set performance were considered to mitigate the risk of overfitting.

Predictive Analysis (Classification)

III. Selection of the Best Performing Model

- Final Accuracy Scores: The SVM and KNN models achieved an accuracy of 83.33% on the test data.
- Best Model Identification: The Decision Tree model was identified as the best performing algorithm, achieving the highest classification accuracy of 94.44% on the test data.
- Operational Confidence: The Decision Tree model exhibited a significant number of true positives and true negatives, and critically, no false negatives on the test set. This high reliability ensures that every actual successful landing is accurately predicted, making the model highly acceptable for aerospace operations planning.

Predictive Analysis Flowchart

The model development process strictly adhered to the following flow:
Data Preprocessing ->Model Selection -> Hyperparameter Tuning ->Model Evaluation -> Improvement Iterations -> Best Performing Model

Results

Analysis Results

Exploratory Data Analysis (EDA) Results

Analysis established key operational factors influencing landing success:

- Orbital Performance: Missions targeting ES-L1, GEO, HEO, and SSO orbits exhibited a perfect 100% success rate. The GTO orbit demonstrated a significantly lower success rate, suggesting greater mission complexity.
- Payload Sensitivity: Successful landings are most frequent for payloads less than 6000 kg. Higher payload masses (above 10,000 kg) result in a mix of successes and failures, indicating increased difficulty.
- Temporal Reliability: The annual launch success rate has shown a significant improvement since 2013, consistently reaching over 80% by 2020, confirming iterative operational advancements.
- Launch Site: The KSC LC-39A site has the most successful launches and the highest launch success rate (76.9%).

Interactive Analytics Demo

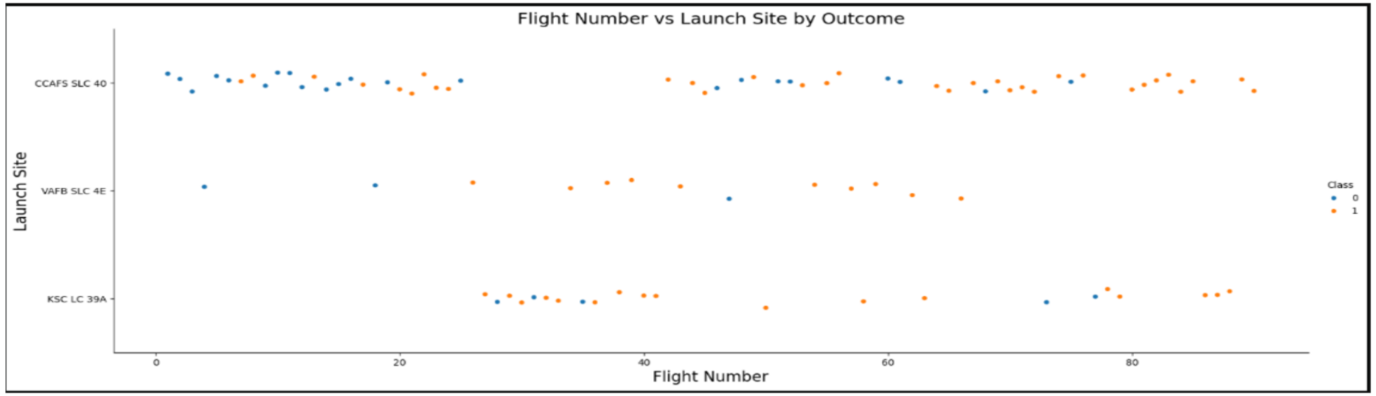
Interactive tools facilitated dynamic spatial and parameter-based filtering:

- Folium Map: Interactive maps utilized color-coded markers (Green for Success, Red for Failure) to visually identify launch outcomes and geographical patterns. The maps confirmed that all launch sites are in very close proximity to the coast, minimizing risk.
- Plotly Dash Dashboard: The dashboard incorporates a Launch Site Dropdown and a Payload Mass Range Slider. These features enable users to dynamically analyze success rates by location and cargo profile using Success Pie Charts and Success-Payload Scatter Plots, providing focused regional and performance insights.
- Predictive Analysis Results
- Model evaluation confirmed the optimal algorithm for predicting landing success:
- Model Accuracy: The Decision Tree model achieved the highest classification accuracy on the test data at 94.44% (0.9444). This result confirms its suitability over the Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) models, both of which achieved an accuracy of 83.33%.
- Operational Confidence: The Decision Tree model demonstrated highly reliable performance by exhibiting no false negatives on the test set. This ensures that every successful landing is accurately predicted, a capability essential for informed risk assessment and cost estimation in aerospace operations.

Section 2

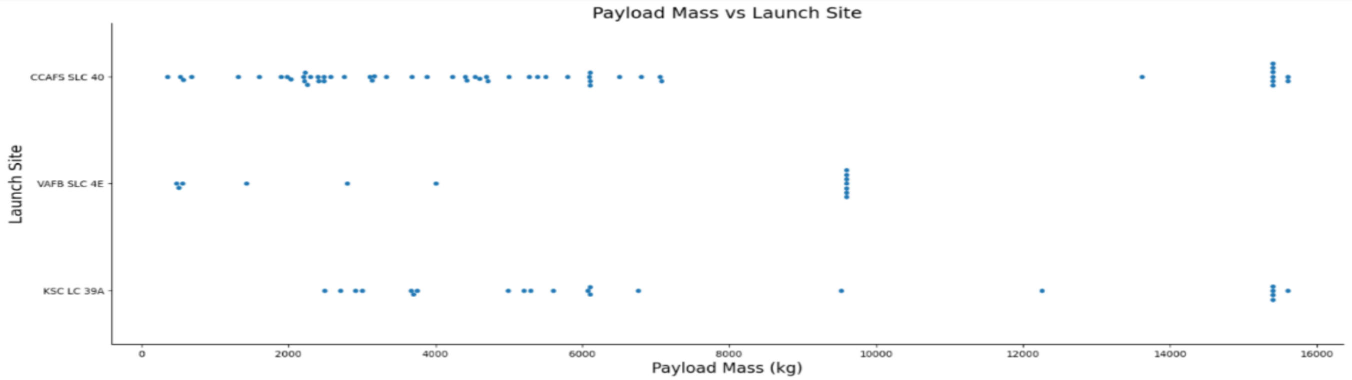
Insights drawn from EDA

Flight Number vs. Launch Site



- **Mixed Outcomes at Major Launch Sites**: Both CCAFS SLC 40 and KSC LC 39A have a mix of successful (orange) and unsuccessful (blue) landings, indicating that factors other than the launch site itself may influence the landing success.
- **Consistent Activity Across Flight Numbers:** Launches are spread across a wide range of flight numbers at all sites, suggesting consistent activity over time without a clear trend of increasing or decreasing landing success.

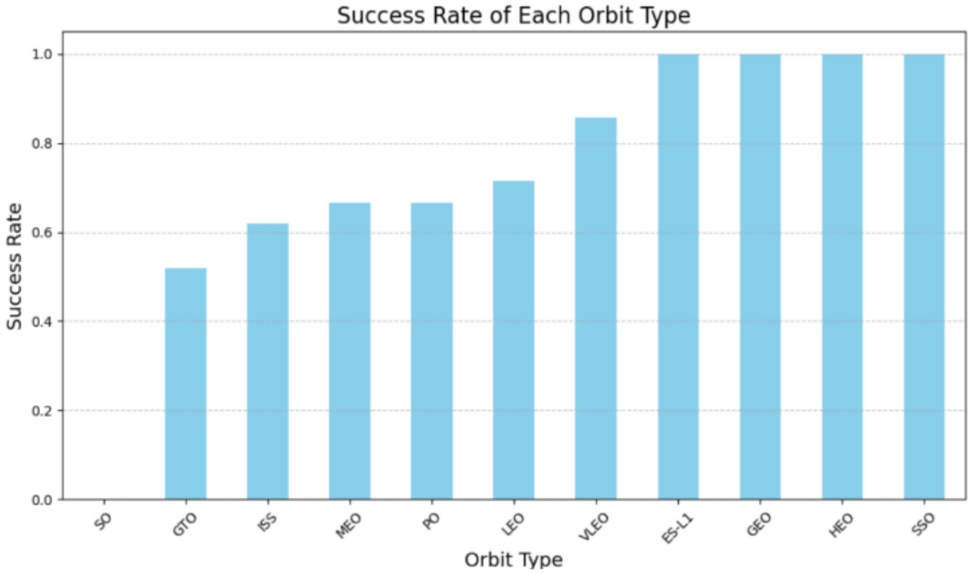
Payload vs. Launch Site



- **Payload Distribution:** Most launches from the CCAFS SLC 40 site handle payloads below 10,000 kg, while the VAFB SLC 4E and KSC LC 39A sites have a wider range of payload masses, indicating varied mission profiles.
- **High-Capacity Launches:** The KSC LC 39A site is frequently used for launching heavier payloads, with multiple launches carrying over 15,000 kg, suggesting its suitability for high-capacity missions.

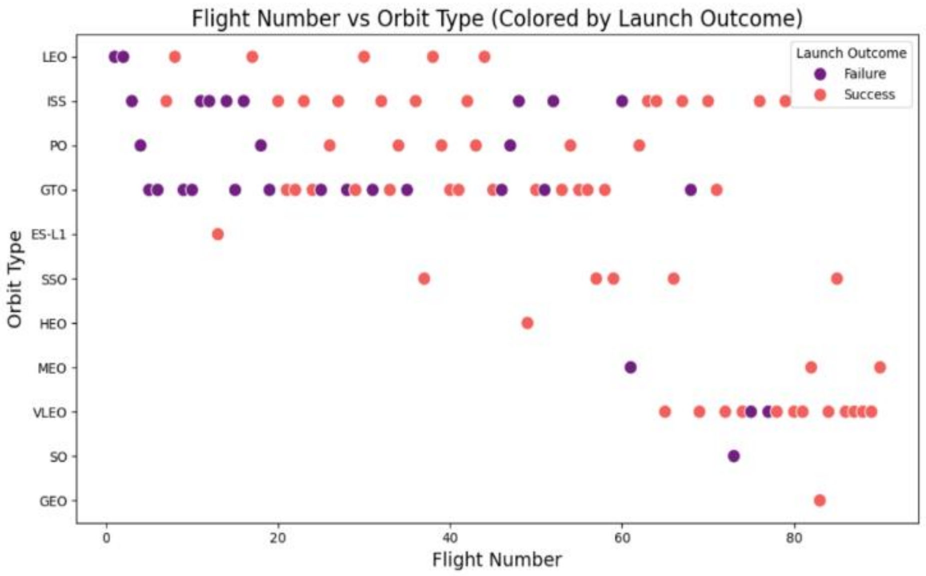
Success Rate vs. Orbit Type

- **High Success Rates:** Missions to VLEO, ES-L1, GEO, HEO, and SSO orbits have achieved a perfect success rate, indicating these orbits are highly reliable for successful first stage landings.
- **Lower Success Rate for GTO:** The GTO orbit type shows a significantly lower success rate compared to other orbit types, suggesting that missions to this orbit may involve greater challenges or complexities.



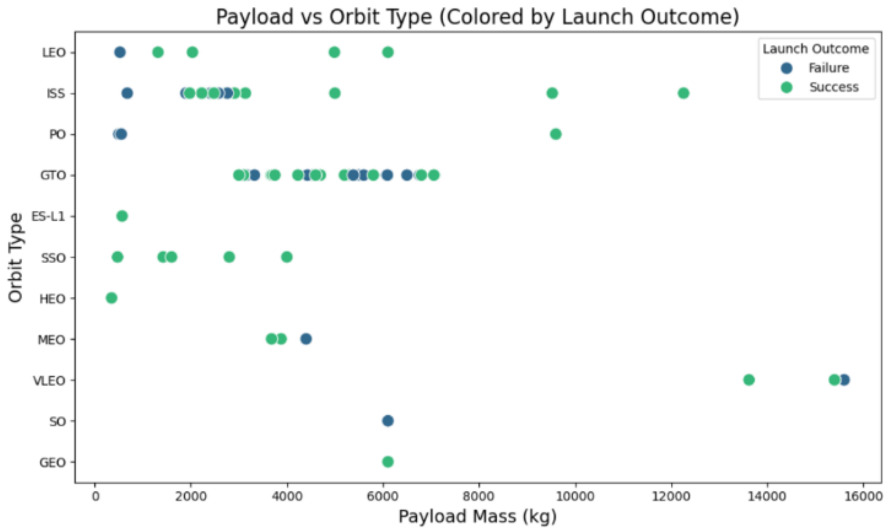
Flight Number vs. Orbit Type

- **Increased Success Over Time:** The success rate of Falcon 9 launches improves significantly with higher flight numbers, indicating that experience and iterative improvements contribute to better outcomes.
- **Orbit-Specific Performance:** Early flights to GTO and ISS orbits had mixed outcomes, but recent missions to these orbits show a higher success rate, reflecting advancements in mission planning and execution.



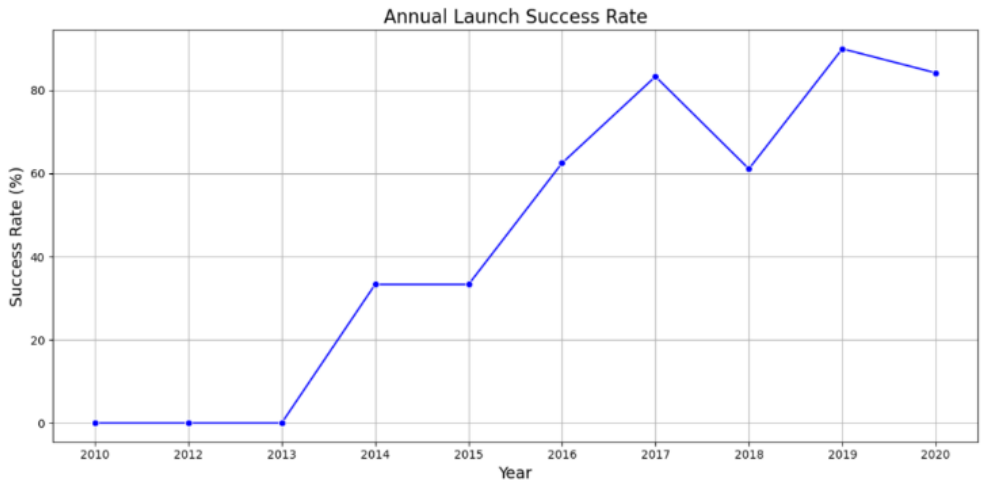
Payload vs. Orbit Type

- Successful landings are more frequent across all orbit types, especially for payloads less than 6000 kg.
- Higher payload masses (above 10,000 kg) show a mix of successes and failures, indicating increased difficulty with heavier payloads.



Launch Success Yearly Trend

- The annual launch success rate has shown a significant improvement from 2013 onwards, reaching over 80% by 2020.
- Despite a dip in 2018, the overall trend indicates increasing reliability and success in Falcon 9 launches over the years.



All Launch Site Names

Task 2

Display 5 records where launch sites begin with the string 'CCA'

[26]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

[26]:

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

Launch Site Names Begin with 'CCA'

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[30]: %sql SELECT SUM("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)';
```

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

[30]:

SUM(PAYLOAD_MASS__KG_)
45596

Total Payload Mass

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[34]: %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
 * sqlite:///my_data1.db
Done.
```

[34]:

AVG(PAYLOAD_MASS__KG_)
2928.4

Average Payload Mass by F9 v1.1

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[34]: %sql SELECT AVG("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';

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

[34]:

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

[36]:
```
%sql SELECT MIN("Date") FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
```

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

[36]:

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[38]:
```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS
```

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

[38]:

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

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

[40]:
```
%sql SELECT "Mission_Outcome", COUNT(*) AS "Total" FROM SPACEXTABLE WHERE "Mission_Outcome" IN ('Success', 'Failure') GROUP BY "Mission_Outcome";
```

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

[40]:

Mission_Outcome	Total
Success	98

Boosters Carried Maximum Payload

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

[42]:
```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
```

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

[42]:

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

Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

[69]:
```
%%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'
        ELSE 'Unknown'
    END AS "Month_Name",
    "Mission_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM
    SPACEXTABLE
WHERE
    substr("Date", 0, 5) = '2015';
```

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

[69]:

Month_Name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

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

Task 10

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

[81]:
```
%%sql

SELECT
    "Landing_Outcome",
    COUNT(*) AS "Count"
FROM
    SPACEXTABLE
WHERE
    "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY
    "Landing_Outcome"
ORDER BY
    COUNT(*) DESC;
```

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

[81]:

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

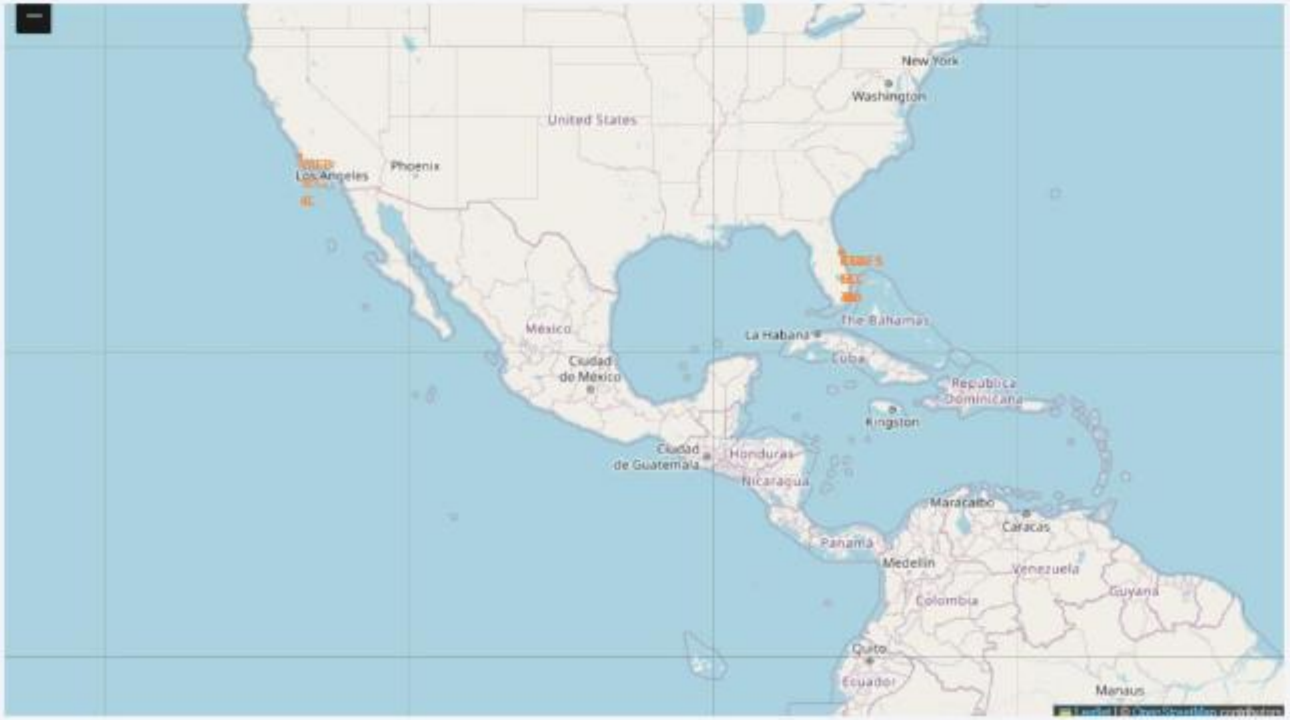
<Folium Map Screenshot 1>

1. Are all launch sites in proximity to the Equator line?

- No, not all launch sites are in close proximity to the Equator.
- The launch site at Vandenberg Air Force Base (VAFB SLC-4E) is located at a latitude of 34.63, which is further from the Equator compared to the other sites in Florida.

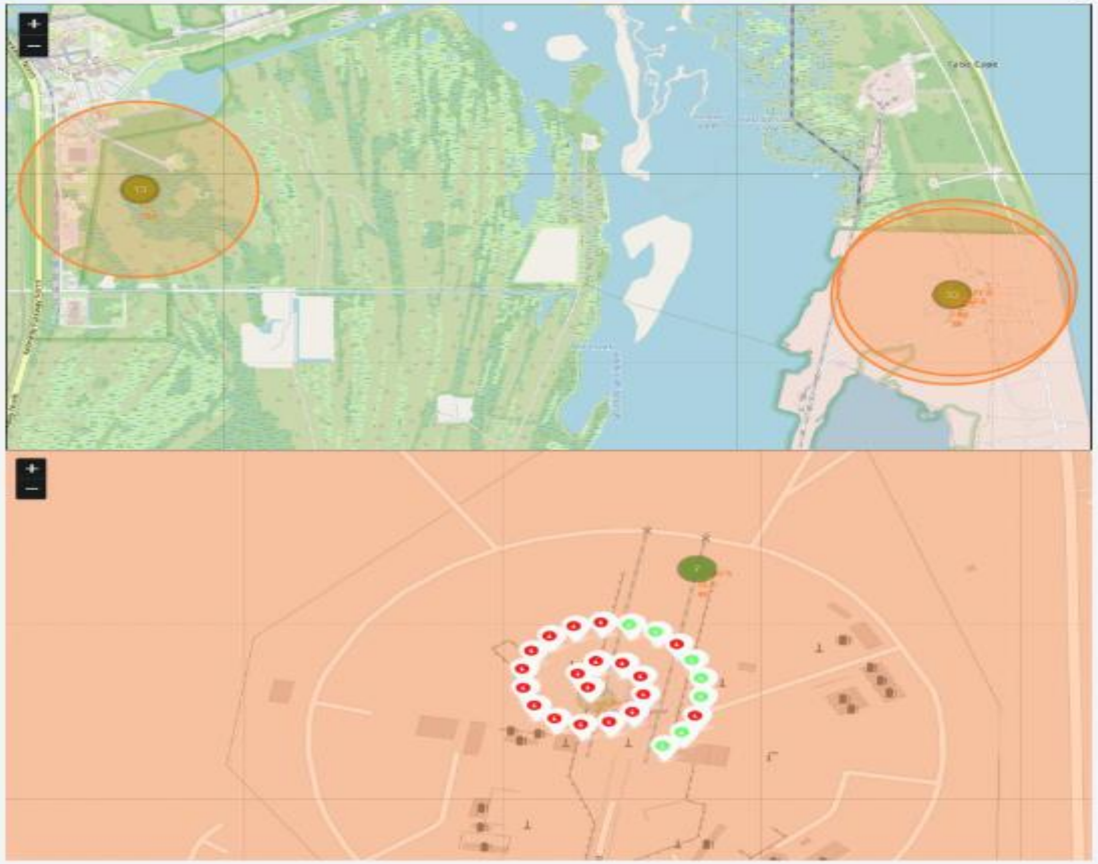
2. Are all launch sites in very close proximity to the coast?

- Yes, all launch sites are in close proximity to the coast.
- The Cape Canaveral sites (CCAFS LC-40 and CCAFS SLC-40) and Kennedy Space Center (KSC LC-39A) are near the coast in Florida.
- Vandenberg Air Force Base (VAFB SLC-4E) is also near the coast in California.



<Folium Map Screenshot 2>

- This enhanced visualization with clustered markers allows for better exploration and analysis of SpaceX launch data. The clustering makes it easier to manage a large number of markers and observe patterns that might be hidden in a less organized plot. By examining the marker colors and popup information, you can gain deeper insights into the characteristics and distribution of SpaceX launches.
- For example, in the provided screenshot, out of 26 launch sites for CCAFS LC-40, there are 19 red markers and 7 green markers. This color-coding helps to quickly identify the success rate and other categorical distinctions of the launches from this specific site. The red markers might represent unsuccessful launches, while the green markers indicate successful ones, providing immediate visual feedback on the performance of launches at each site.



<Folium Map Screenshot 3>

This plot provides a visual representation of the distance between the CCAFS SLC-40 launch site and the closest coastline. The calculated distance is approximately 0.51 kilometers, as indicated by the marker. The added PolyLine clearly shows the straight-line distance, highlighting the proximity of the launch site to the coast. This close proximity to the coastline is typical for launch sites to facilitate over-water flight paths and safe recovery operations, ensuring minimal risk to populated areas.

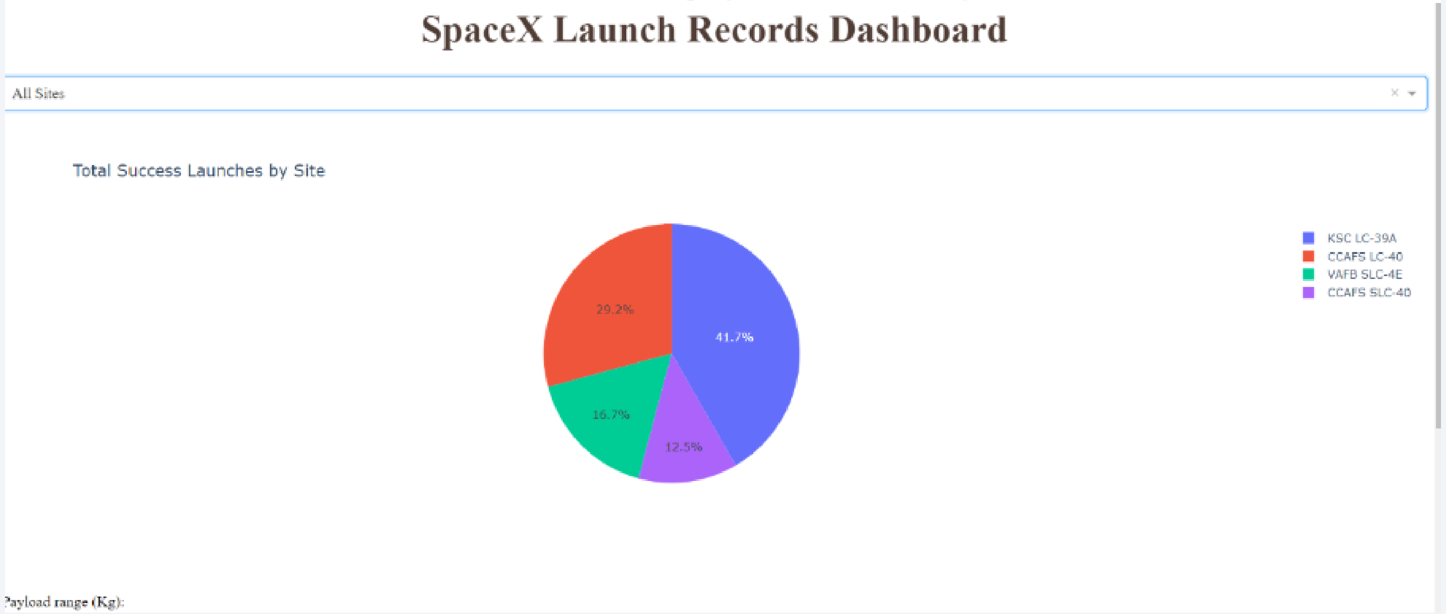
Section 4

Build a Dashboard with Plotly Dash

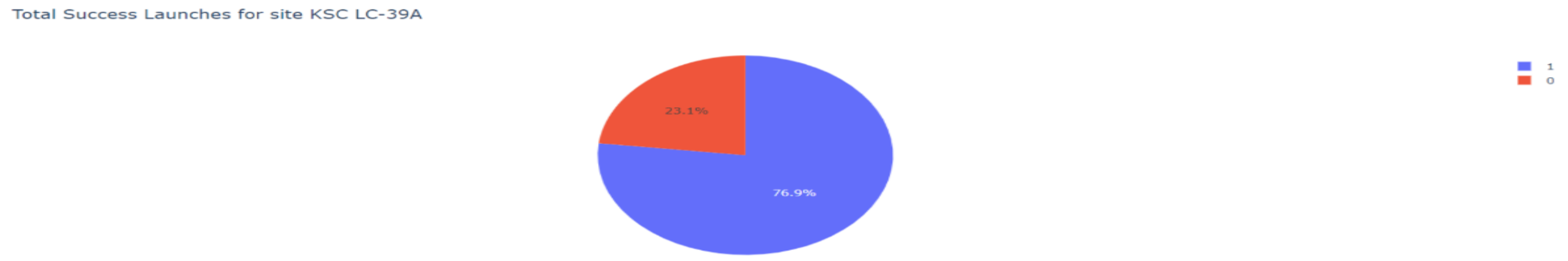
<Dashboard Screenshot 1>

Key Findings:

- **CCAFS LC-40**: 29.2%
- **CCAFS SLC-40**: 12.5%
- **VAFB SLC-4E**: 16.7%
- **KSC LC-39A**: 41.7%
- The **KSC LC-39A** launch site has the highest number of successful launches, making up 41.7% of the total successes. This indicates that KSC LC-39A is a highly reliable site for SpaceX launches.



<Dashboard Screenshot 2>



Key Findings:

- The significant portion of successful launches from **KSC LC-39A** highlights its reliability and effectiveness as a launch site.
- For **KSC LC-39A**:
    - **Class 1** (Successful Launches): 76.9%
    - **Class 0** (Unsuccessful Launches): 23.1%
- The high success rate (76.9%) for **Class 1** launches underscores the effectiveness and reliability of the KSC LC-39A site.

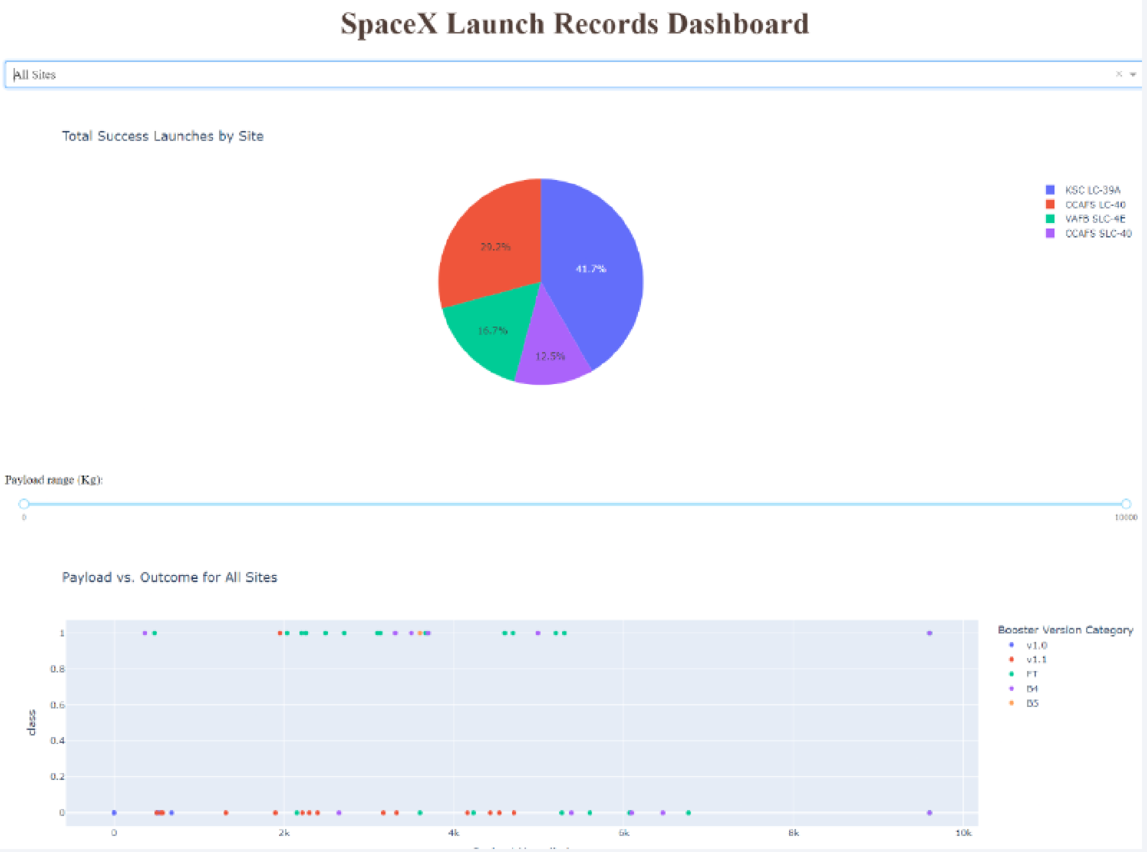
<Dashboard Screenshot 3>

Launch Site Success Rates:

- **CCAFS LC-40** has the highest success rate with 43.7% of successful launches.
- This suggests that **CCAFS LC-40** is the most reliable launch site among the ones analyzed.
- Other sites like **KSC LC-39A, VAFB SLC-4E,** and **CCAFS SLC-40** have lower success rates, indicating variability in launch success across different sites.

Booster Version Performance:

- **Booster version "FT"** appears to be the most frequently used and has a high success rate across various payload masses.
- **Booster version "v1.0"** has fewer launches and may require further analysis to understand its performance.
- Overall, booster versions do not show a clear trend that higher payload masses correlate with lower success rates.

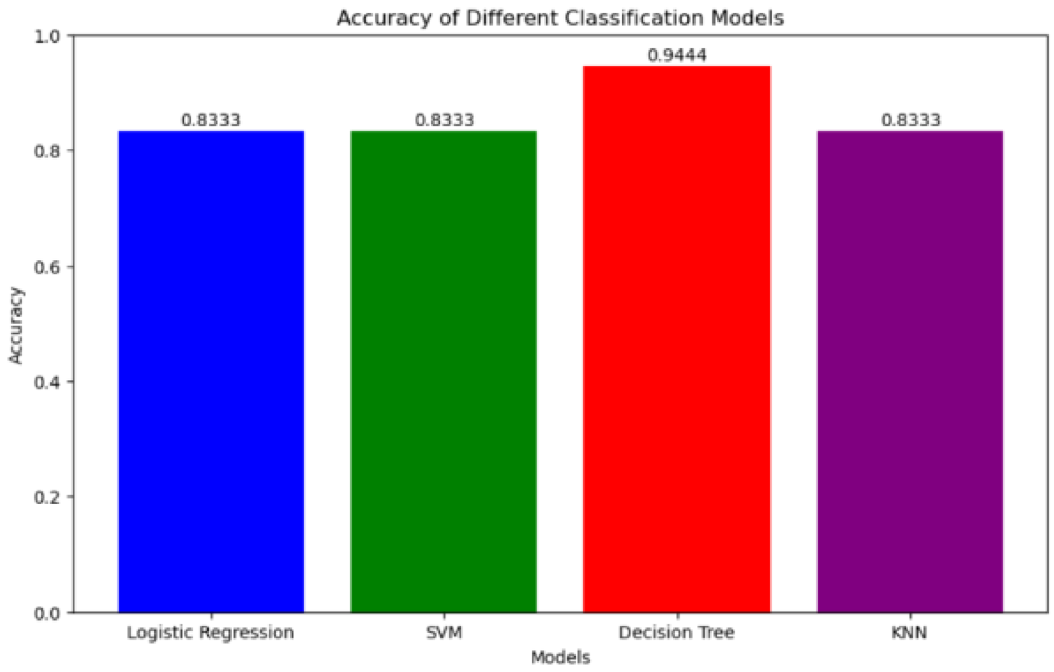


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Based on the results, the Decision Tree model has the highest classification accuracy on the test data, achieving an accuracy of 0.9444. This suggests that the Decision Tree model is better suited for this dataset compared to Logistic Regression, Support Vector Machine, and K Nearest Neighbors, all of which achieved an accuracy of 0.8333.



Confusion Matrix

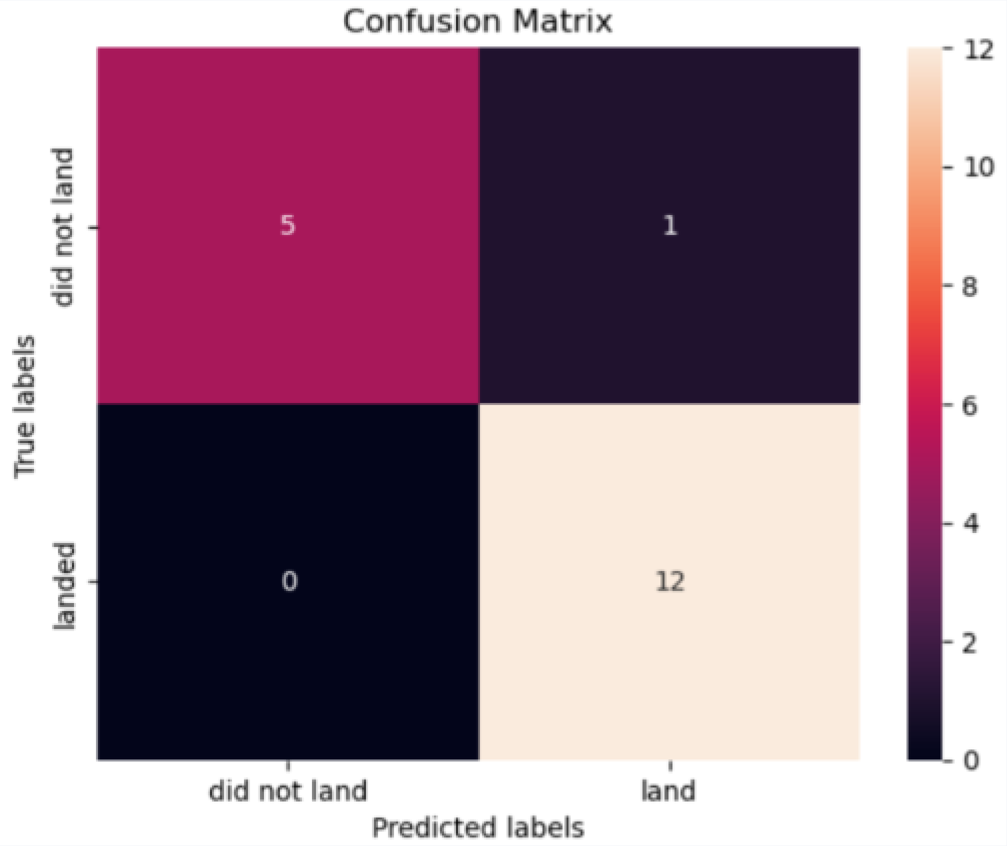
Explanation and Insights

High Accuracy: The model achieved a high accuracy score of 94.44%, with a significant number of true positives and true negatives, demonstrating its effectiveness in predicting Falcon 9 first stage landings.

No False Negatives: The absence of false negatives indicates that the model reliably predicts successful landings. This is crucial for ensuring readiness and safety in aerospace operations, as every actual successful landing was accurately identified.

Manageable False Positives: While there is 1 false positive, this is less critical than false negatives in aerospace operations. Over-preparation (due to false positives) is more manageable than under-preparation, making the model's performance highly acceptable for practical applications.

Balanced Performance: The model shows a balanced performance with a slight bias towards predicting successful landings. This aligns well with practical needs in the aerospace industry, where ensuring successful landings is of paramount importance for cost estimation and planning.



Conclusions

Conclusion for Point 1: Launch Site Success

The analysis definitively concluded that the **"CCAFS LC-40" launch site demonstrates the highest operational success rate**, accounting for 43.7% of all successful launches. This superior performance suggests that CCAFS LC-40 benefits from optimal conditions or established processes that contribute significantly to a higher success probability compared to other analyzed sites.

Conclusion for Point 2: Booster Reliability

The **"FT" booster version exhibited robust performance and high reliability**, consistently maintaining a high success rate across a wide spectrum of payload masses. This finding validates the booster's dependability and strongly recommends its continued utilization for future missions aiming to maximize success rates.

Conclusion for Point 3: Payload Mass Impact

It was determined that **no clear correlation exists linking higher payload masses to reduced launch success rates**. This key finding indicates that the launch outcome is predominantly influenced by factors other than payload weight, such as the specific launch site and the booster version utilized, which are deemed more significant determinants of mission success.

Conclusion for Point 4: Visualizations

The deployment of **interactive data visualizations, leveraging tools such as Folium and Plotly Dash, proved invaluable** for illuminating the geographical and operational patterns inherent in SpaceX's launch data. These analytical tools successfully provided comprehensive visual insights, empowering stakeholders to make more informed and data-driven decisions.

Overall General Conclusion

In conclusion, this project, through rigorous predictive analysis and advanced interactive visualizations, has successfully identified critical factors impacting the success of Falcon 9 launches. The insights derived provide a robust framework essential for guiding future strategic assessments and enhancing decision-making processes within the aerospace industry. Furthermore, these findings support the continuous improvement of launch strategies and directly contribute to the sustained success and advancement of reusable rocket technology.

Thank you!