IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Lander Laparra
2025/06/26

Outline

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

Executive Summary

Summary of Methodologies:

- **Data Collection:**
  The dataset for SpaceX launches was collected using a combination of API calls and web scraping techniques. This data includes detailed information about launch sites, payload mass, orbit type, launch outcomes, and rocket specifications.

- **Data Cleaning:**
  Data preprocessing was performed to handle missing values, remove irrelevant records, and filter the dataset to focus solely on Falcon 9 launches. Missing payload mass values were imputed with the mean, and categorical variables were standardized using one-hot encoding.

- **Exploratory Data Analysis (EDA):**
  EDA was conducted to understand relationships between key variables such as payload mass, launch site, and landing success. Visualizations such as histograms, scatter plots, and interactive maps created with Folium were used to analyze the proximity of launch sites and landing success.

- **Predictive Modeling:**
  Various machine learning models, including Logistic Regression, SVM, Decision Trees, and K-Nearest Neighbors (KNN), were used to predict landing success. Hyperparameters for these models were optimized using GridSearchCV, and model performance was evaluated based on accuracy, precision, and recall.

- **SQL Queries:**
  SQL was employed to perform advanced data extraction and aggregation, answering critical questions such as the success rate of landings at each launch site.

Executive Summary

Summary of All Results:

- **Key Findings:**
  - **Launch Site:** The analysis revealed that the success rate of Falcon 9 landings varied across different sites. For example, **KSC LC 39A** and **VAFB SLC 4E** showed a success rate of approximately **77%**, whereas **CCAFS LC-40** had a lower success rate of **60%**.
  - **Payload Mass:** A payload mass **greater than 10,000 kg at CCAFS LC-40** was associated with a **100% success** rate in landings.
- **Best Model:**
  The **Support Vector Machine (SVM)** with a **linear kernel** provided the best results for predicting landing success. It outperformed other models in terms of **accuracy**.
- **Prediction Accuracy:**
  The **Logistic Regression** model achieved reliable performance, showing **high accuracy and precision in predicting landing outcomes** (success or failure).
- **SQL Analysis:**
  SQL queries were used to uncover key trends, such as the number of successful landings at each launch site and the relationship between landing outcomes and orbit types. It was found that **ISS orbit had a higher success rate for landin**gs.
- **Interactive Analysis:**
  Folium maps were utilized to visualize the geographical locations of the launch sites and their proximities. These maps revealed patterns in the distribution of launch sites and success rates.

Introduction

Project Background and Context:

SpaceX has revolutionized the space industry with its ability to reuse rocket components, specifically the Falcon 9's first stage. Reusing the first stage significantly reduces the cost of launches, making space exploration more affordable. However, predicting the likelihood of successful landings is still a challenge. The goal of this project is to develop a predictive model that can forecast whether the first stage of a Falcon 9 rocket will land successfully.

This analysis is crucial not only for optimizing costs but also for improving the efficiency and reliability of rocket launches. By analyzing past launches and identifying patterns, SpaceX can refine its operations and make more informed decisions about future missions.

Introduction

Problems to Find Answers:

1. **What factors influence the success or failure of Falcon 9 landings?**
    - Does the launch site impact landing success?
    - How does payload mass affect landing outcomes?
2. **Can we predict the likelihood of a successful landing?**
    - Can machine learning models, such as Logistic Regression, SVM, and Decision Trees, accurately predict landing success based on various features?
3. **Which machine learning model performs the best for predicting landing success?**
    - How can we optimize model hyperparameters to achieve the highest accuracy?
4. **How do geographical factors play a role in landing success?**
    - Can we identify geographical patterns using maps to determine optimal landing sites?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Data Collection Process:

1. **API Data Collection:**
   The launch data was collected using the SpaceX API. The data included information such as the **launch site, payload mass, orbit type, landing outcomes**, and **rocket details** (e.g., rocket type, booster version).
   - API Endpoint: Data was pulled from the SpaceX API's launch data endpoint.

2. **Web Scraping:**
   Web scraping was employed to gather additional **data from Wikipedia**. We specifically targeted the **list of SpaceX Falcon 9 and Falcon Heavy launches** for more detailed records on landing outcomes and launch specifics.
   - **Scraping Method: BeautifulSoup** and requests were used for extracting tables from the web page.

3. **Database Queries (SQL):**
   A **SQL database was used** to filter and aggregate launch data. We performed complex **SQL queries to analyze trends**, such as the number of successful landings per site or payload mass distribution.

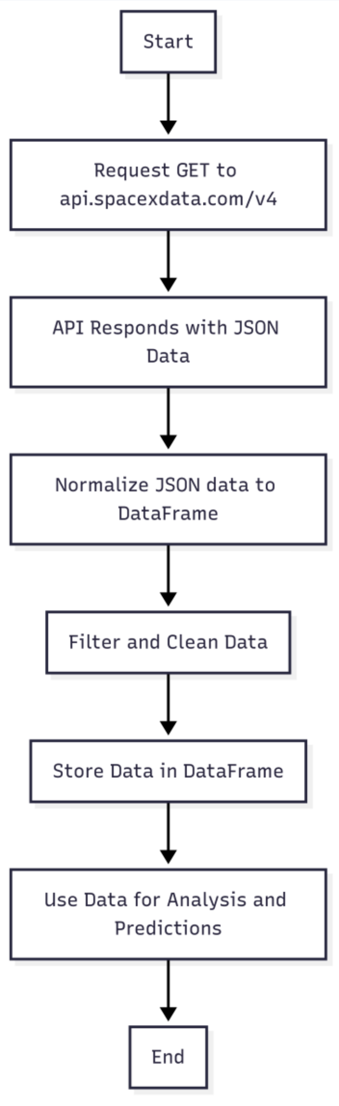
Data Collection

Key Points:

- **Data Sources:**
  SpaceX API, Web Scraping, SQL Database.
- **Tools Used:**
  Python (requests, BeautifulSoup, pandas, SQLAlchemy), PowerBI (for visualizations).
- **Data Sources Outcome:**
  Detailed launch data including both successful and failed landings.

Data Collection – SpaceX API

Notebook Link



Data Collection – SpaceX API

API Endpoints:

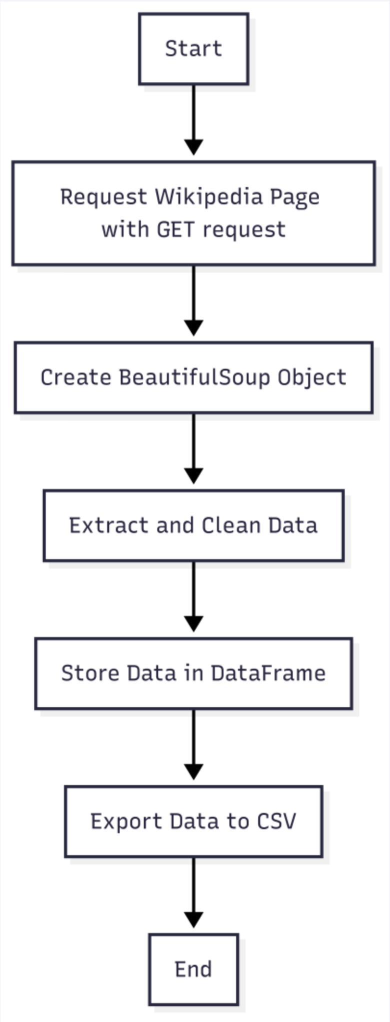
- Request to SpaceX Launches (Past Data):
  - Endpoint: /v4/launches/past
  - Purpose: This endpoint is used to pull historical SpaceX launches data including information like the launch site, payload mass, orbit type, landing outcomes, and rocket details.
- Request to SpaceX Rocket Data:
  - Endpoint: /v4/rockets/{rocket_id}
  - Purpose: To gather detailed information about the rocket (booster version) used in each launch.
- Request to SpaceX Launchpad Data:
  - Endpoint: /v4/launchpads/{launchpad_id}
  - Purpose: To get information about the launch site location, including longitude, latitude, and name.
- Request to SpaceX Payload Data:
  - Endpoint: /v4/payloads/{payload_id}
  - Purpose: To collect details about the payload (payload mass, orbit type).
- Request to SpaceX Core Data:
  - Endpoint: /v4/cores/{core_id}
  - Purpose: To retrieve data about the core, including landing success, type of landing, number of flights, and whether the core was reused.

Data Collection - Scraping

[Notebook Link]

Key Steps:

- HTTP GET Request to Wikipedia page using requests.
- Parse the page using BeautifulSoup.
- Extract, clean, and process the table rows and columns.
- Store the clean data in a Pandas DataFrame.
- Export the DataFrame to a CSV file.



Data Wrangling

Notebook Link

Data Wrangling

- **Handling Missing Data:**
  - The missing values in the payload mass column were replaced with the mean value for consistency.
  - For other missing data, such as in the launch site or customer columns, appropriate handling was applied (e.g., replacing None with specific categories or values).
- **Filtering Irrelevant Records:**
  - Non-Falcon 9 launches were filtered out to focus on the relevant dataset.
  - Records with inconsistent or incomplete data were either cleaned or removed as necessary.
- **Converting Data Types:**
  - Categorical variables, such as launch sites and orbit types, were encoded using one-hot encoding to prepare them for analysis and machine learning.
  - Numeric data, including payload mass and flight numbers, were ensured to be in the correct data type (e.g., float64, int64).
- **Handling Outliers and Irregular Data:**
  - Outliers in the payload mass column were identified and handled accordingly (e.g., trimming or capping).
- **Training Labels:**
  - Landing outcomes were transformed into binary labels where 1 represents a successful landing and 0 represents a failed landing.

EDA with Data Visualization

[Notebook Link]

Charts Plotted:

- **Flight Number vs Launch Site:** Scatter plot showing no clear relationship between flight number and launch site success.
- **Payload Mass vs Launch Site:** Scatter plot revealing heavy payloads are launched from CCAFS SLC 40.
- **Success Rate by Orbit Type:** Bar chart highlighting ISS orbit as the highest success rate, with GTO lower.
- **Flight Number vs Orbit Type:** Scatter plot showing success rate related to flight number, particularly for LEO.
- **Payload Mass vs Orbit Type:** Scatter plot identifying heavy payloads in Polar, LEO, and ISS orbits.
- **Launch Success Rate Over Years:** Line plot showing increasing success rate from 2013 to 2020.

EDA with SQL

[Notebook Link]

SQL Queries Performed:

- **Count of Successful Landings Per Launch Site:**
  Calculated the number of successful launches per site to analyze patterns in landing success across different locations.

- **Landing Success Rate by Payload Mass:**
  Investigated if payload mass correlates with landing success using aggregate SQL queries to group by payload mass and calculate success rates.

- **Average Payload Mass by Launch Site:**
  Aggregated data to calculate the average payload mass for each launch site, providing insights into payload characteristics by site.

- **Filter Data for Successful Launches:**
  Used SQL WHERE clauses to filter records for successful landings, focusing the analysis only on successful missions.

- **Identify Key Trends in Landing Outcomes:**
  Analyzed trends in landing outcomes, including identifying landing success rates across different orbit types.

Build an Interactive Map with Folium

[Notebook Link]

Map Objects Created:

- **Markers:**
  - Added markers for each launch site to represent the geographic location of SpaceX launch sites on the map.
  - Each marker is interactive, displaying relevant information when clicked, such as the site name and its geographical coordinates.
- **Circles:**
  - Used circles to highlight launch site locations with a radius proportional to the number of launches at each site.
  - This visually indicates the density of launches per site, helping to emphasize the importance of more frequently used sites.
- **Lines:**
  - Drawn lines between key launch sites to show proximity and provide spatial relationships between different sites.
  - This was done to visualize the proximity of launch sites in relation to each other, especially for geographical analysis.

Build an Interactive Map with Folium

Reason for Adding Map Objects:

- **Markers** allow users to visually pinpoint launch sites on the map and interact with them for detailed information.
- **Circles** help to quickly assess which launch sites are used more frequently, providing a visual cue about site activity.
- **Lines** aid in understanding the spatial relationships between launch sites, enhancing geographical context.

Build a Dashboard with Plotly Dash

[Dash Source Link]

Overview of the Dashboard Features:

- **Dropdown Selector:** Allows dynamic filtering by launch site, enabling focused analysis on specific locations or aggregated view.
- **Range Slider:** Enables users to interactively filter payload mass (0–10,000 kg) to explore its effect on success rate.
- **Pie Chart:** Visualizes success vs failure counts, adapting dynamically to the selected site.
- **Scatter Plot:** Shows the correlation between payload mass and mission outcome, color-coded by booster version category.

Purpose & Insights:

These interactive elements help users explore launch patterns, identify high-performing sites (e.g., CCAFS LC-40), and observe success trends across different payload ranges and booster versions.

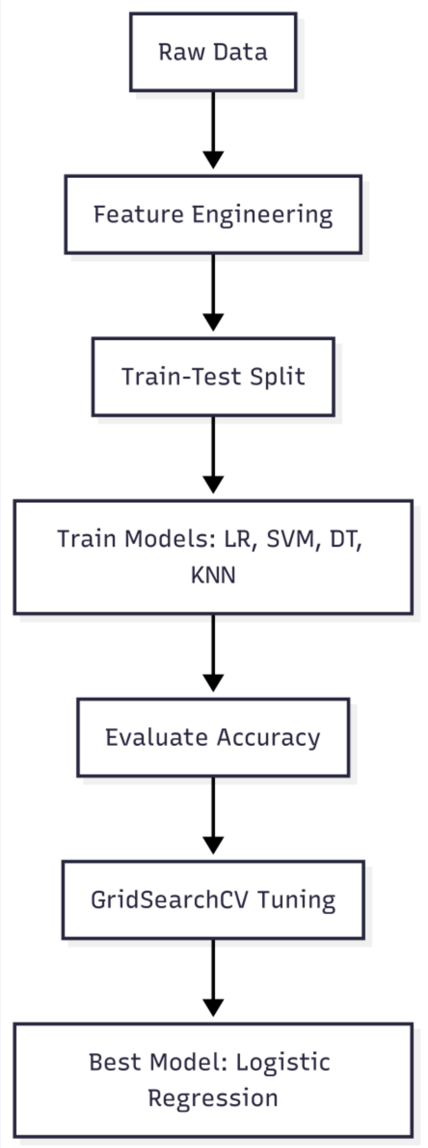
Predictive Analysis (Classification)

[Notebook Link]

Model Development Summary:

- Data Preparation:
  - Features selected: Payload Mass, Orbit, Launch Site, GridFins, Reused, Legs, etc.
  - Applied one-hot encoding to categorical variables and normalization to numeric features.
- Model Building:
  - Trained and evaluated:
  - Logistic Regression
  - SVM
  - Decision Tree
  - K-Nearest Neighbors
- Model Evaluation:
  - Accuracy measured for each model using a train-test split.
  - Logistic Regression achieved the highest baseline performance.
- Hyperparameter Tuning:
  - GridSearchCV used to fine-tune hyperparameters of all models.
  - Final comparison showed Logistic Regression as the best performing model.
- Result:
  - Best model: Logistic Regression
  - Final accuracy: 0.8333.

Predictive Analysis (Classification)



Results

Exploratory Data Analysis (EDA):

- Launch **success rate is correlated** with both **payload mass and orbit type**.
- Certain launch **sites (e.g., KSC LC 39A) show higher success rates**.
- **Success rates increased steadily over time**, reaching over 90% by 2019–2020.
- Orbits like SSO, HEO, and ES-L1 had near-perfect success.

Interactive Analytics (Dashboard):

- Dropdowns and sliders allow filtering by launch site and payload range.
- Pie charts summarize success vs failure counts interactively.
- Scatter plots show correlation between payload and success per orbit/site.
- Enables dynamic, real-time insights for business and technical audiences.

Predictive Analysis:

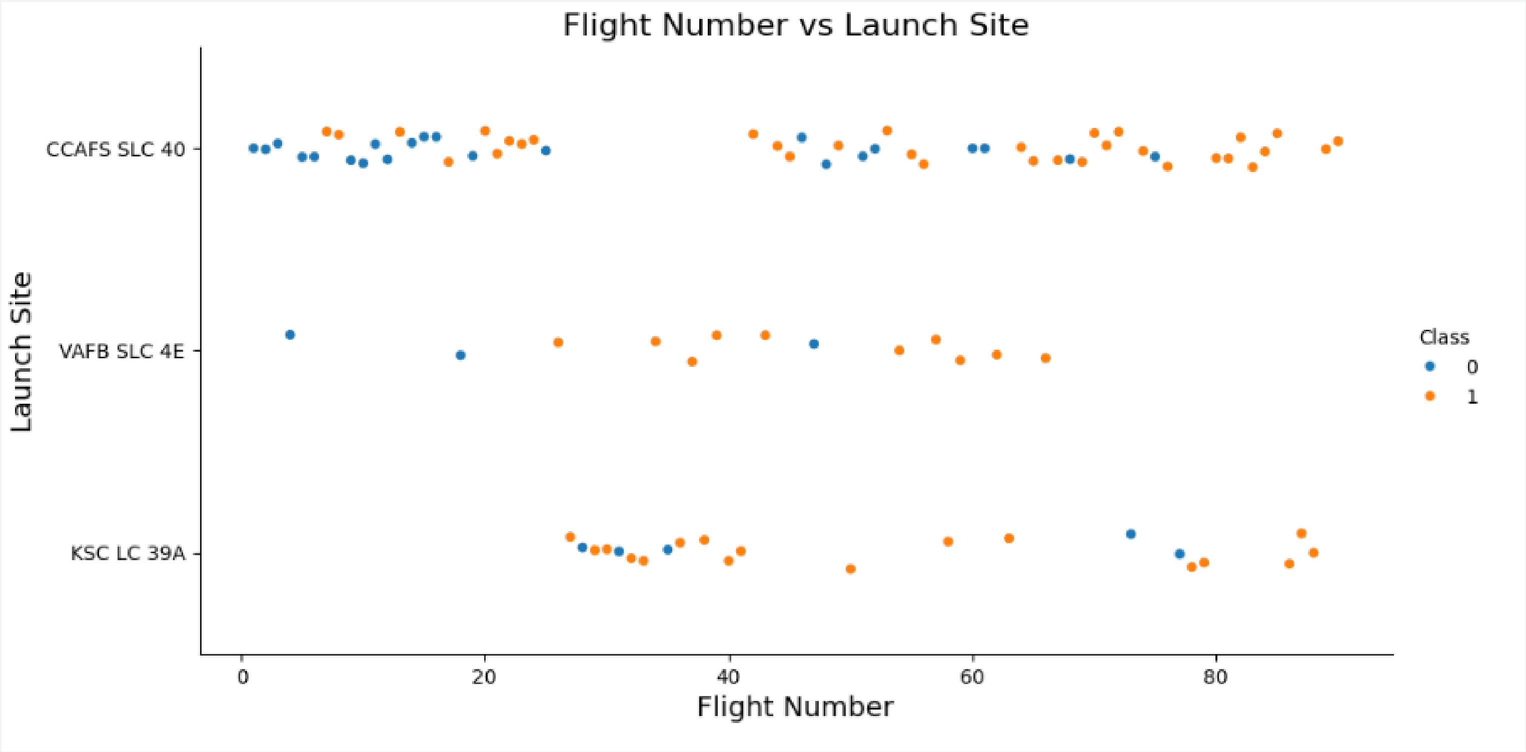
- **Logistic Regression** selected as **best model with accuracy = 0.8333**.
- Models trained: Logistic Regression, SVM, Decision Tree, KNN.
- Features used: Payload mass, Orbit, Launch site, Booster reuse indicators.
- Data was preprocessed, encoded, scaled and fine-tuned using GridSearchCV

Section 2

Insights drawn from EDA

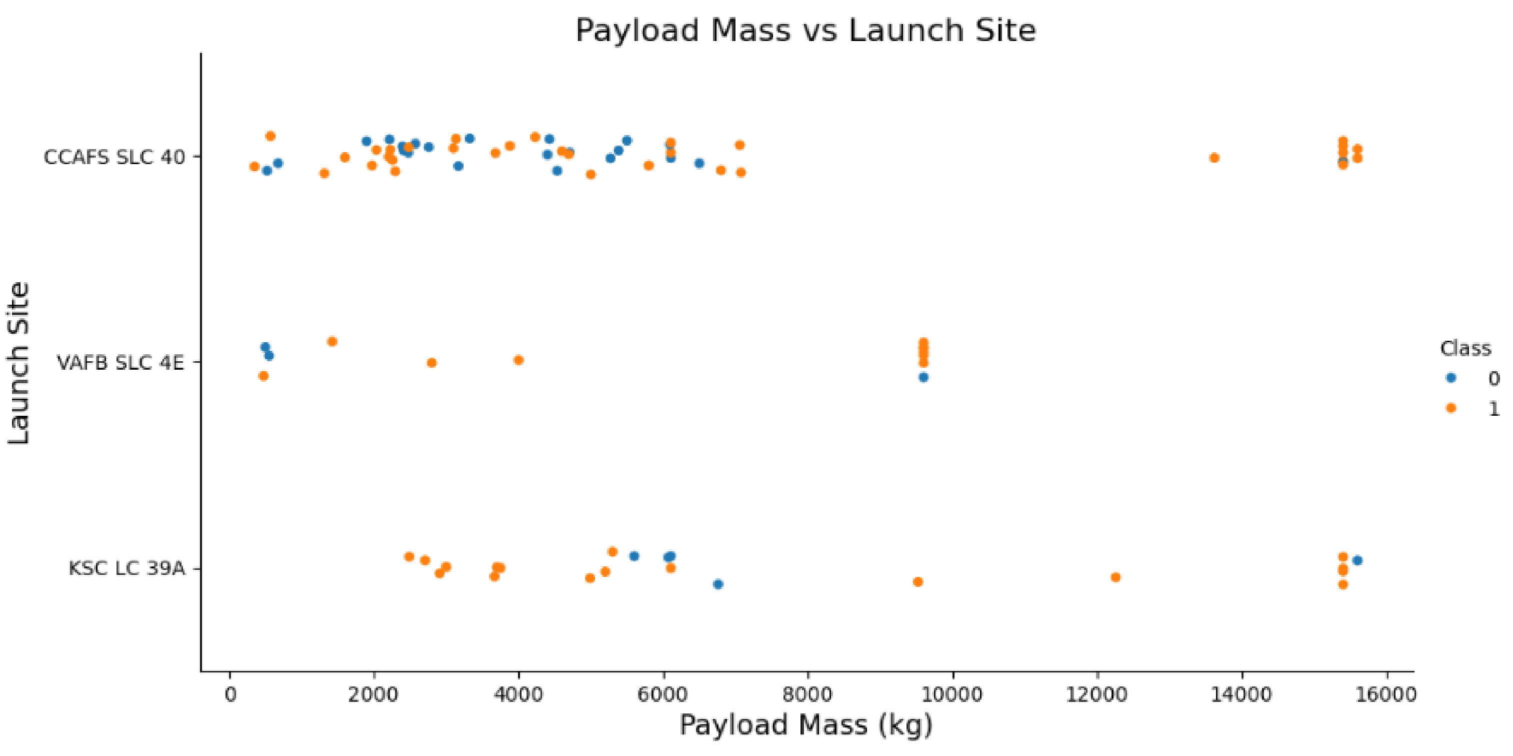
Flight Number vs. Launch Site

- **Objective:** Analyze whether the number of flights (experience) affects launch success per site.
- **Visualization:** Scatter plot of Flight Number vs. Launch Site, colored by success class
  (0 = Failure, 1 = Success).
- **Insights:**
  - Launches at **KSC LC 39A** and **VAFB SLC 4E** show more consistent success rates in later flights.
  - **CCAFS SLC 40** has a mix of failures and successes across many flight numbers.
  - There's a visible trend where **higher flight numbers correlate with more successes**, especially at KSC LC 39A.
- **Interpretation:** Experience and infrastructure maturity (represented by higher flight numbers) contribute to better outcomes.



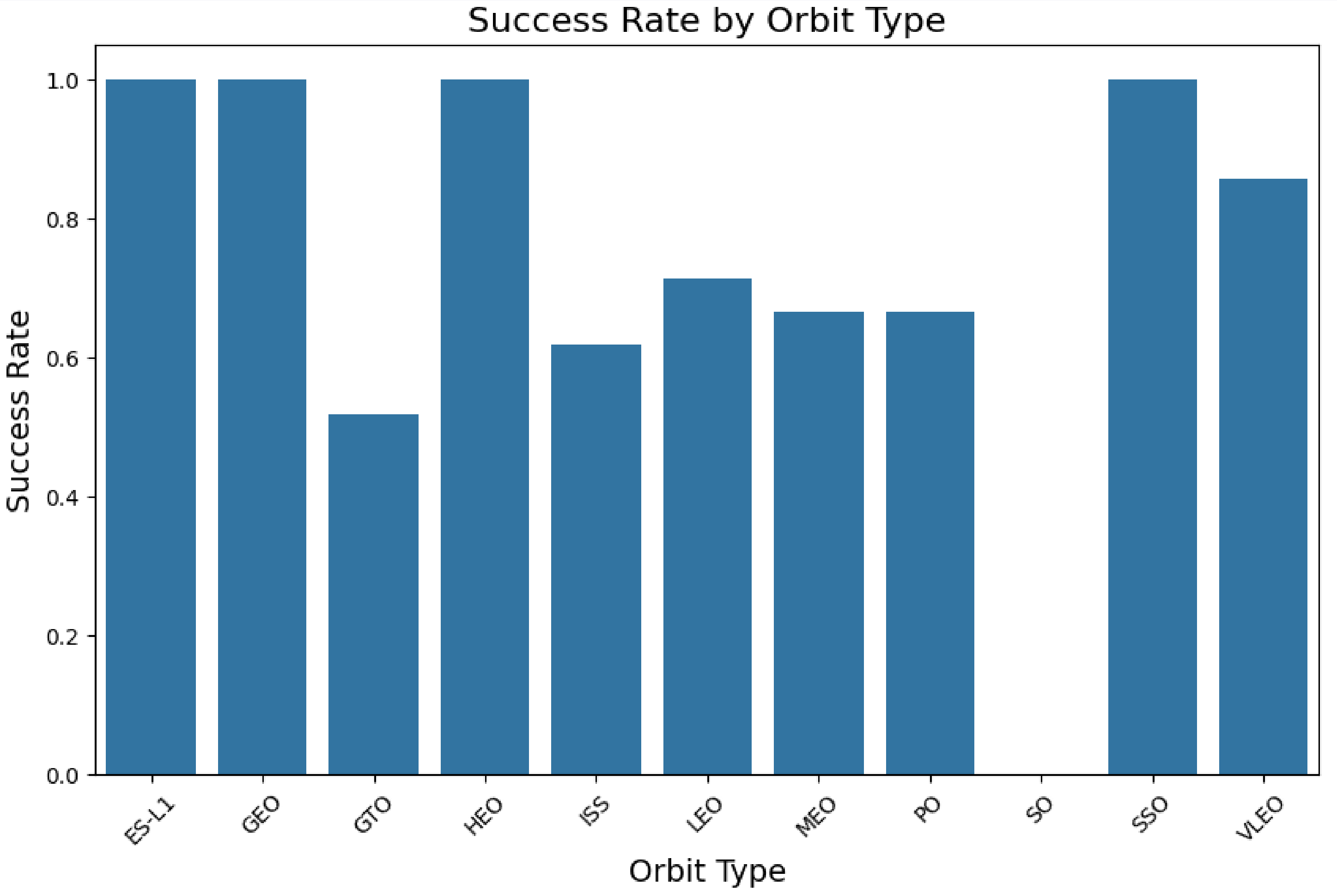
Payload vs. Launch Site

- **Objective:** Explore whether the payload mass influences the outcome depending on the launch site.
- **Visualization:** Scatter plot of Payload Mass (kg) vs. Launch Site, colored by class (0 = Failure, 1 = Success).
- **Insights:**
    - **VAFB SLC 4E** has no launches with payloads above 10,000 kg.
    - Both **CCAFS SLC 40** and **KSC LC 39A** supported heavy payloads.
    - Successful landings are more prevalent for moderate payloads across sites.
- **Interpretation:** Heavy payloads seem less frequent and more successful from advanced launch facilities like KSC LC 39A.



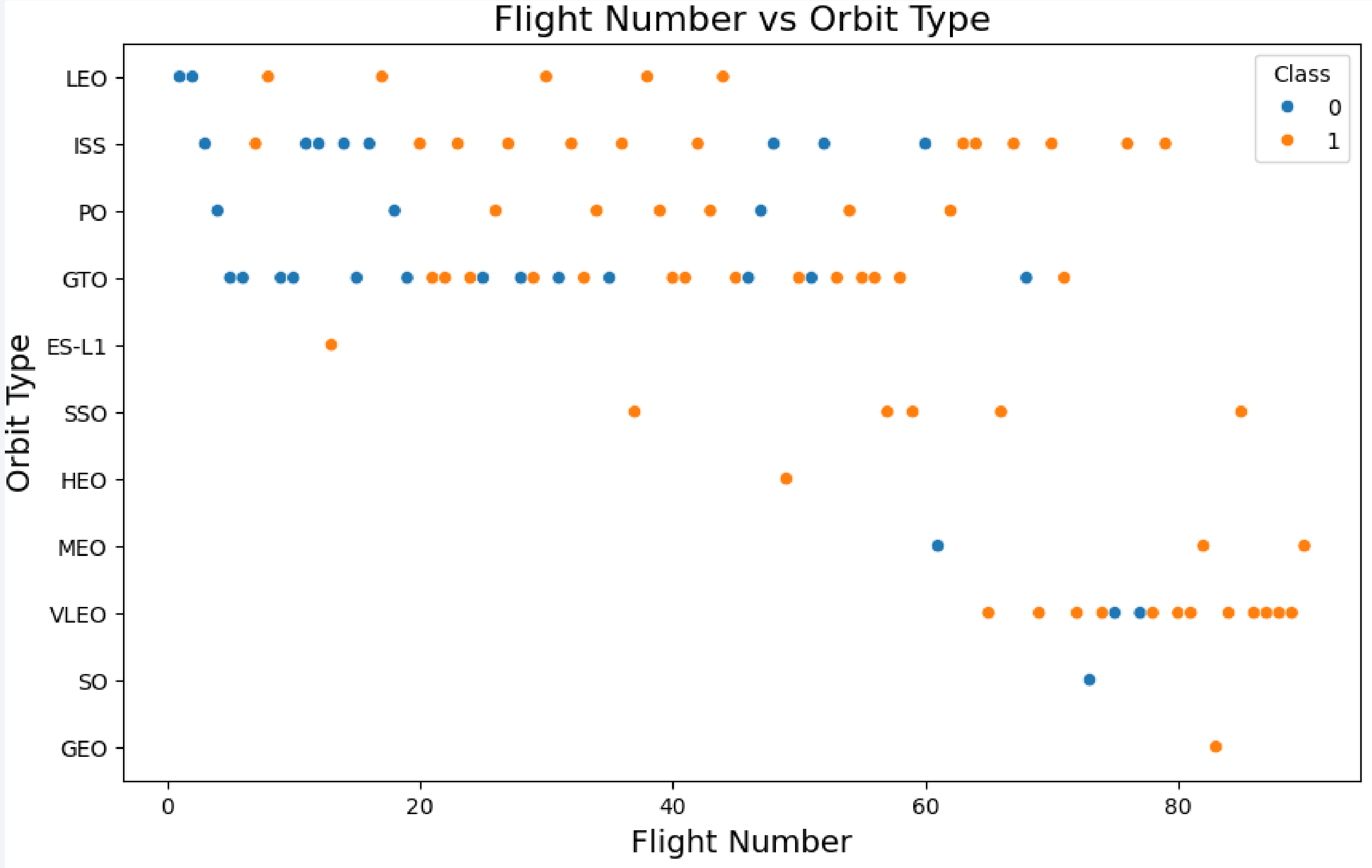
Success Rate vs. Orbit Type

- **Objective:** Analyze how different orbit types relate to mission success rates.
- **Visualization:** Bar chart showing the average Class (success rate) per Orbit.
- **Insights:**
  - Highest success rates observed in orbits ES-L1, GEO, HEO, and SSO.
  - Orbits like **GTO** and **ISS** show comparatively lower success rates.
  - Some orbits (e.g., **SO**) show no successful missions.
- **Interpretation:** The orbit type can influence mission complexity and landing success, with specialized orbits showing better reliability.



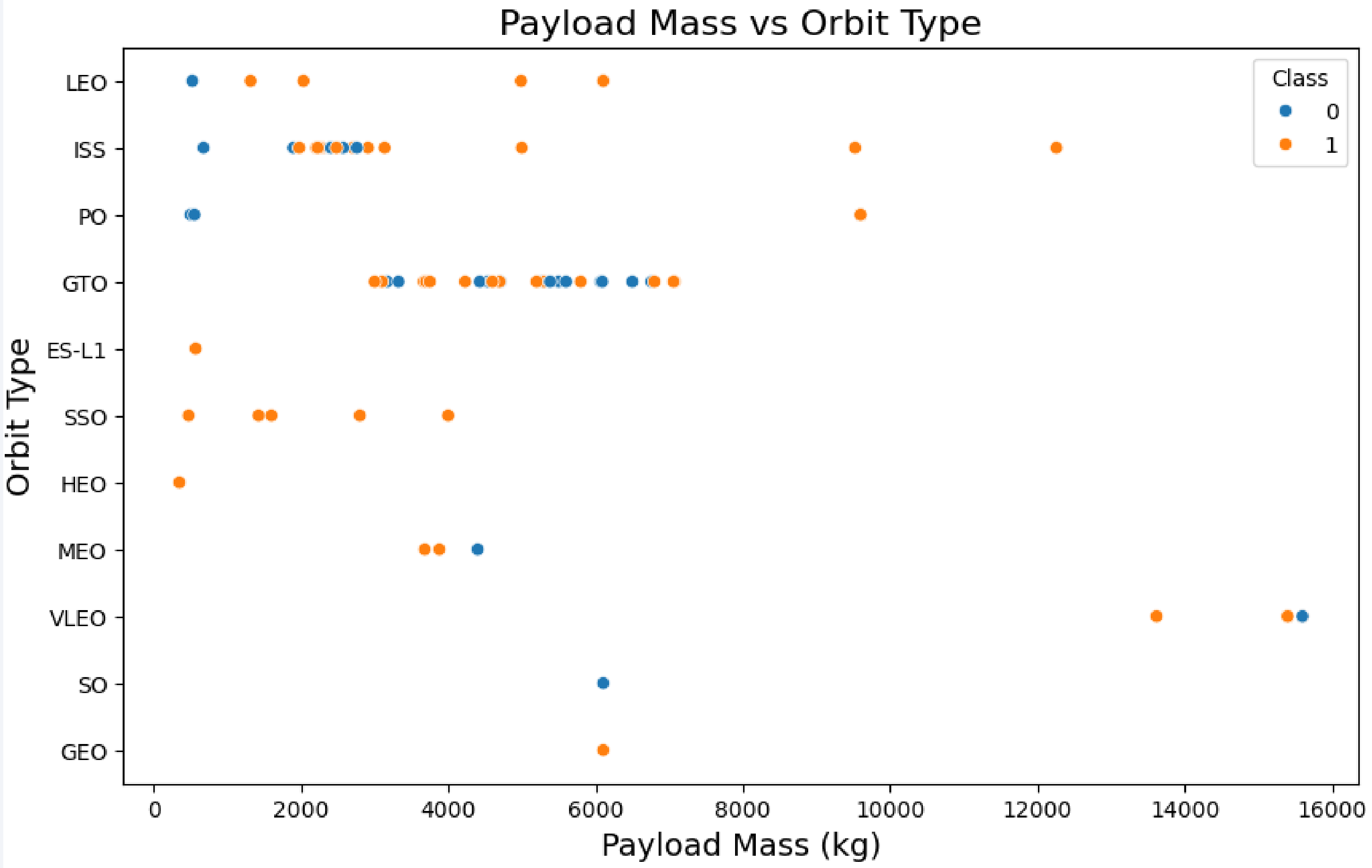
Flight Number vs. Orbit Type

- **Objective:** Examine whether mission experience (flight number) influences success across orbit types.
- **Visualization:** Scatter plot with Flight Number on the x-axis, Orbit on the y-axis, and Class as hue (success/failure).
- **Insights:**
  - In **LEO**, success improves with flight number — possibly due to learning effects and process refinement.
  - In **GTO**, no clear relationship is observed between flight number and mission outcome.
  - Some orbits show limited data, making patterns inconclusive.
- **Interpretation:** The impact of launch experience (flight number) varies across orbit types, indicating some missions are more sensitive to operational maturity.



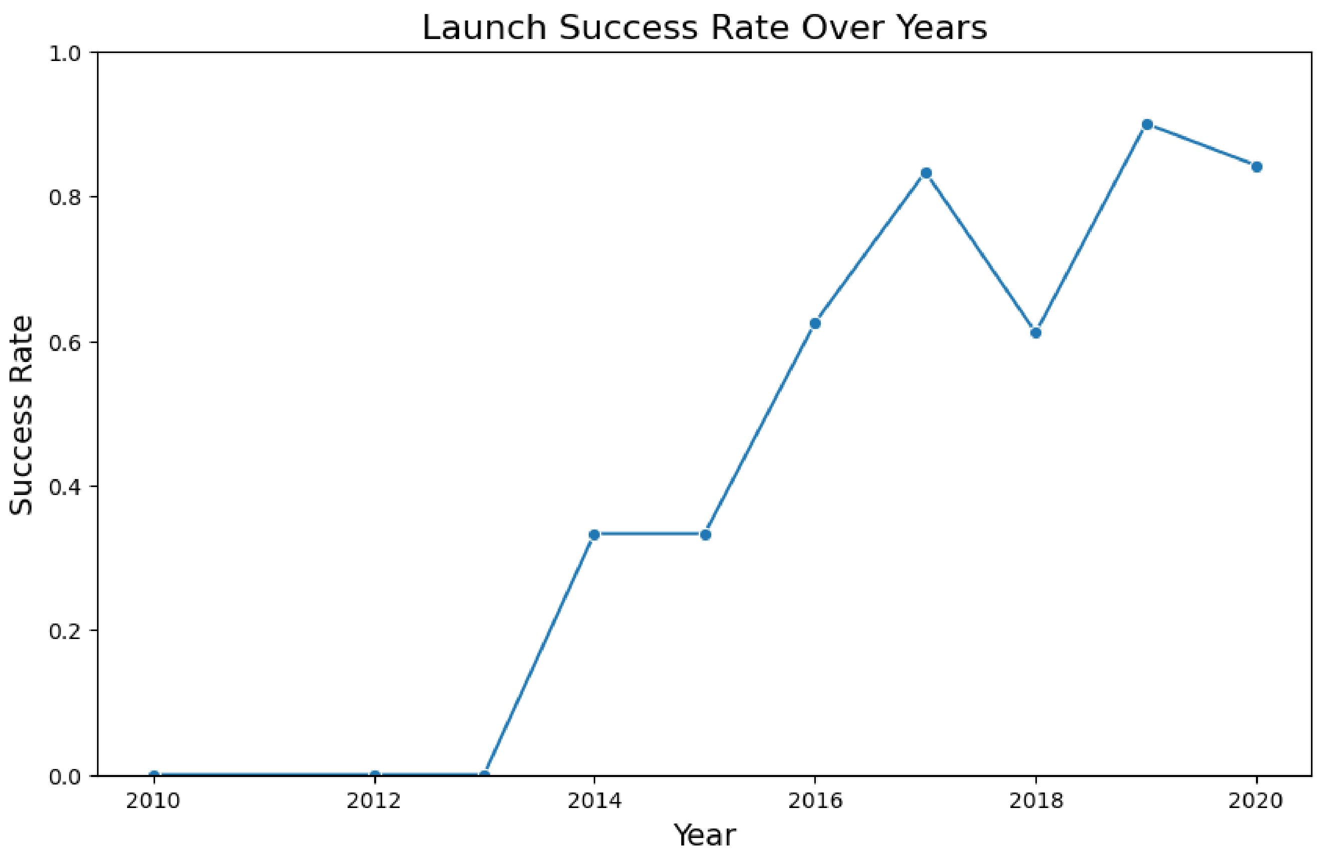
Payload vs. Orbit Type

- **Objective:** Explore how payload mass influences success rates across different orbit types.
- **Visualization:** Scatter plot with Payload Mass (kg) on the x-axis, Orbit Type on the y-axis, and Class (success/failure) as hue.
- **Insights:**
  - Higher payloads to **LEO**, **Polar**, and **ISS** orbits tend to correlate with successful landings.
  - In **GTO**, both successful and unsuccessful landings are observed across payloads — no clear trend.
  - Several orbit types show consistent success regardless of payload weight.
- **Interpretation:** Mission design and orbit type appear more predictive of outcome than payload mass alone in some orbits.



Launch Success Yearly Trend

- **Objective:** Analyze the trend of launch success over time to assess improvements in SpaceX's booster landing performance.
- **Visualization:** A line chart was plotted to show the yearly average success rate, using the launch year on the x-axis and the mean of the Class label on the y-axis. This visualization highlights how SpaceX's success rate evolved over time.
- **Insights:**
  - **2010–2013:** No successful landings recorded (success rate = 0).
  - **2014–2017:** Rapid growth in landing success.
  - **2018:** Slight dip, followed by continued improvements.
  - **2019–2020:** Consistently high performance, success rate near 0.9.
- **Interpretation:** This trend demonstrates the increasing reliability of SpaceX's landing technology. The sharp increase after 2014 aligns with the company's investment in reusable rocket stages, validating the effectiveness of iterative development and real-world testing.



All Launch Site Names

- **Query Executed**
  - SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
- **Result**
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
- **Explanation**
  - The query retrieves distinct values from the Launch_Site column, identifying the three unique locations used by SpaceX for launches in this dataset. These results establish the basis for further site-specific performance analysis.

Launch Site Names Begin with 'CCA'

- **Query Executed**
  - SELECT * FROM SPACEXTBL WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
- **Result**

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

- **Explanation**
  - This query filters for launches from Cape Canaveral (CCAFS), a key SpaceX site. Reviewing these entries helps analyze patterns in payloads, mission types, and landing outcomes from this site.

Total Payload Mass

- **Query Executed**
  - SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM SPACEXTBL WHERE "Customer" = 'NASA (CRS)';
- **Result**
  - Total Payload Mass: **45,596 kg**
- **Explanation**
  - This query calculates the cumulative payload mass delivered by SpaceX boosters under NASA's Commercial Resupply Services (CRS) program.
    It highlights the significant logistical role of CRS missions in supporting the International Space Station.

Average Payload Mass by F9 v1.1

- **Query Executed**
  - SELECT AVG("PAYLOAD_MASS__KG_") AS avg_payload_mass FROM SPACEXTBL WHERE "Booster_Version" = 'F9 v1.1';
- **Result**
  - Average Payload Mass: **2,928.4 kg**
- **Explanation**
  - This query determines the average payload mass carried by Falcon 9 version 1.1 boosters.
    It provides insight into the typical cargo capacity of this specific rocket variant, useful for performance benchmarking.

First Successful Ground Landing Date

- **Query Executed**
  - SELECT MIN("Date") AS first_successful_landing_date FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';
- **Result**
  - Date: **2015-12-22**
- **Explanation**
  - This query retrieves the earliest date when a successful landing on a ground pad occurred.
    The result highlights a major milestone in SpaceX’s reusability efforts: the first successful Falcon 9 booster recovery on solid ground.

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Query Executed**
  - SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
- **Result**
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2
- **Explanation**
  - This query filtered for booster versions that successfully landed on a drone ship and carried payloads between 4000 and 6000 kg.
    The result identifies specific missions that achieved a balance between heavy payload delivery and recovery success on sea platforms.

Total Number of Successful and Failure Mission Outcomes

- **Query Executed**
  - SELECT "Landing_Outcome", COUNT(*) AS outcome_count FROM SPACEXTBL GROUP BY "Landing_Outcome";
- **Result Summary**

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

- **Explanation**
  - This query grouped all landing attempts by outcome and counted occurrences of each. It shows that out of the documented missions, 61 ended in successful landings and 10 resulted in failures. This breakdown highlights SpaceX's progress and growing reliability in booster recovery.

Boosters Carried Maximum Payload

- **Query Executed**
  - SELECT "Booster_Version" FROM SPACEXTBL WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTBL);
- **Result Summary**
  - Multiple entries match the maximum value due to repeated use of the same payload mass across different flights or boosters.

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

- **Explanation**
  - This subquery identifies the highest payload mass in the entire dataset and then filters the boosters that achieved this value.
    The result highlights a cluster of F9 B5 series boosters, demonstrating their enhanced payload capacity compared to earlier versions.

2015 Launch Records

- Query Executed
  - SELECT CASE strftime('%m', "Date") WHEN '01' THEN 'January'... END AS "Month", "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE "Landing_Outcome" = 'Failure (drone ship)'AND strftime('%Y', "Date") = '2015';
- Result Summary
  - The query retrieved **2 failed drone ship landings in 2015**, specifically in **January** and **April**.
  - Both launches occurred at **CCAFS LC-40**.
  - The booster versions involved were **F9 v1.1 B1012** and **F9 v1.1 B1015**.
- Explanation
  - This query demonstrates how time-based filtering and pattern matching help isolate mission failures by month, which is crucial for identifying temporal patterns or issues in recovery processes.

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

- **Query Summary**
  - SELECT "Landing_Outcome", COUNT(*) AS outcome_count FROM SPACEXTBL WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY outcome_count DESC;
- **Result Summary**

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

- **Explanation**
  - This query ranks the different types of landing outcomes to evaluate the frequency of successful recoveries versus other landing events. It reveals early trends in recovery strategy and operational maturity before 2017.

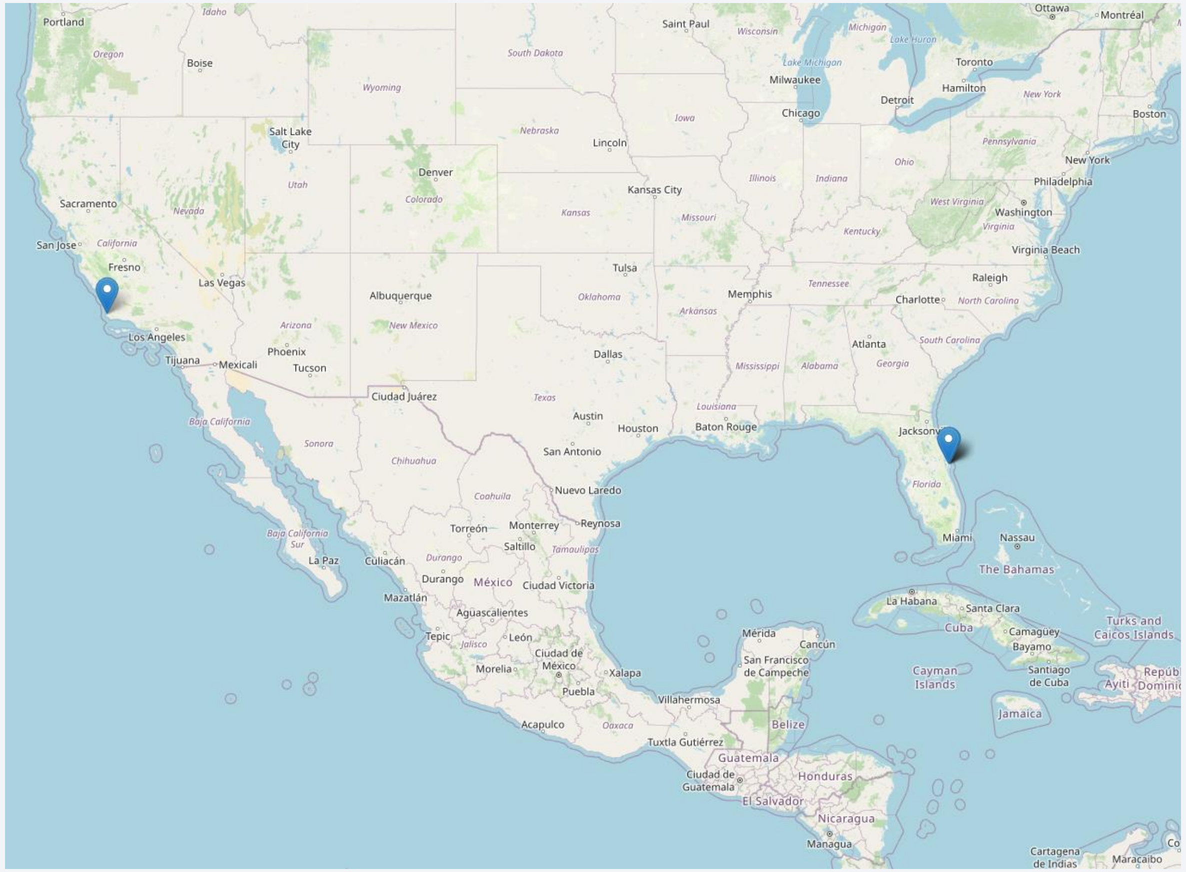
Section 3

Launch Sites Proximities Analysis

All Launch Sites - Global Geolocation

The screenshot displays a Folium-generated interactive map that marks all SpaceX launch site locations across the United States.

- **Key Elements**
  - Blue location markers identify launch sites.
  - Sites are clustered along the U.S. coastlines: one on the West Coast (VAFB SLC-4E, California) and multiple on the East Coast (KSC LC-39A and CCAFS LC-40, Florida).
  - Zoom level adjusted to show continental context, highlighting how launch site geography may affect mission planning and logistics.
- **Findings**
  - The East Coast is more densely populated with active launch sites, reflecting its logistical and orbital advantages (e.g., proximity to equator for geostationary launches).
  - The geographic spread supports diverse launch trajectories and recovery operations, which are critical for reusable rockets.



Launch Outcomes by Color

Objective: Visualize the outcome of Falcon 9 booster landings across different launch sites using color-coded markers.

Visualization: A Folium map is generated with clustered markers representing individual launch events. Each marker's color reflects the outcome of the landing:

- Green for successful landings
- Red for failures
- Yellow for no attempt or unknown outcome

Insights:

- The map clearly shows that the majority of launches occurred in Florida (CCAFS and KSC).
- Most successful landings are concentrated at specific sites, particularly KSC LC-39A.
- California's VAFB site appears to have fewer launch attempts, and a mix of outcomes.

Interpretation: The spatial distribution of landing outcomes reveals that SpaceX's reliability has improved over time at its main operational sites, with a high cluster of green markers indicating success. This visual summary supports temporal and spatial pattern analysis in SpaceX operations.

Launch Site Proximity Analysis – Coastline Distances

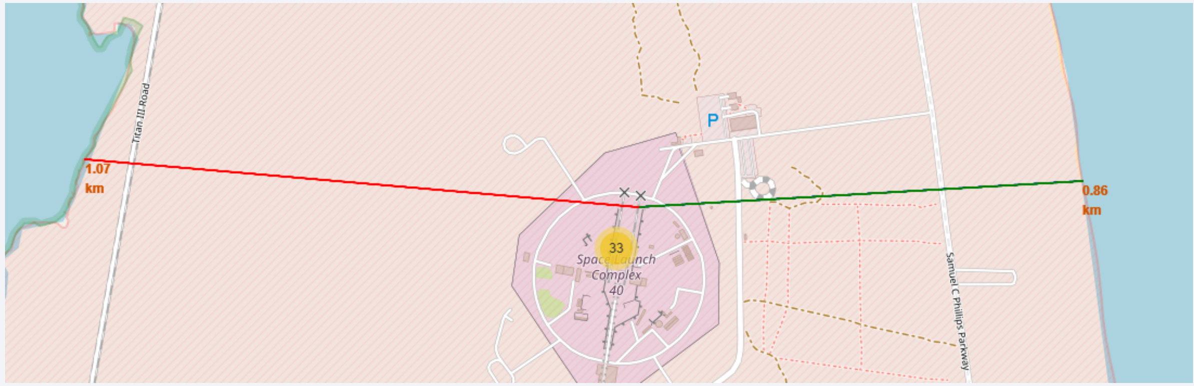
The folium map displays Launch Complex 40 (SLC-40) with lines pointing to the nearest coastal access points:

- East Coast (green line)
- West Coast (red line)

Key Elements:

- Two labeled lines connecting the launch site to:
  - The Atlantic coastline to the east (0.86 km)
  - The inland coastal boundary to the west (1.07 km)
- Color-coded markers and distances aid visual comparison.

- **Insights:**
  - The launch site is closer to the eastern coastline, a strategic advantage for eastward launches over the ocean.
  - The measured distances help assess potential risk zones and logistic paths for transport or recovery.
- **Interpretation:**
  - Including these proximity measures supports operational planning for launch logistics, safety buffers, and recovery strategies, especially in coastal launch facilities like SLC-40.



Section 4

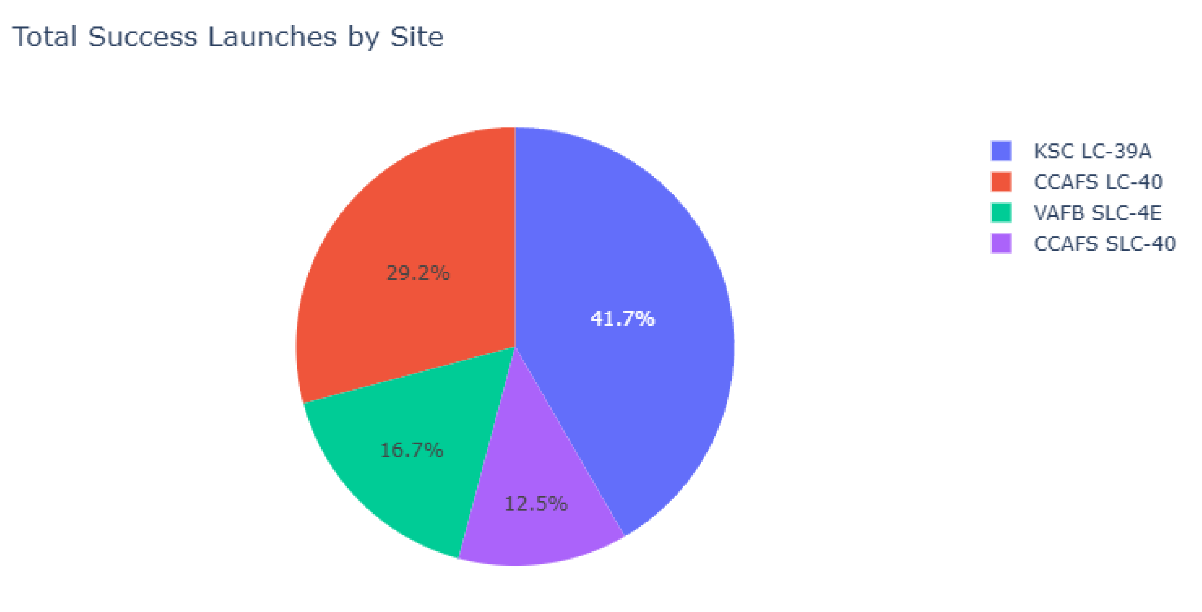
Build a Dashboard with Plotly Dash

Launch Success Distribution Across All Sites

This pie chart displays the proportion of successful launches per launch site. It reveals that:

- **KSC LC-39A** leads with **41.7%** of total successful launches.
- Followed by **CCAFS LC-40** with **29.2%.**
- **VAFB SLC-4E** and **CCAFS SLC-40** contribute **16.7%** and **12.5%**, respectively.

These proportions highlight the relative operational success across sites and indicate that KSC LC-39A is both highly active and reliable.

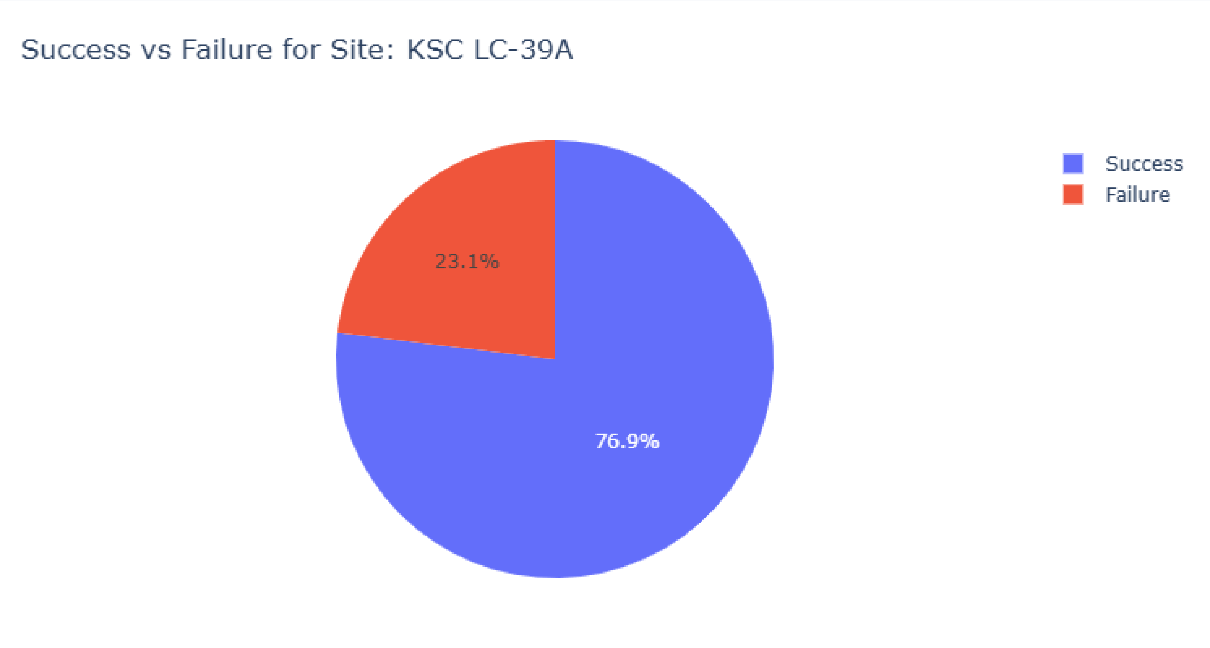


Highest Launch Success Rate: KSC LC-39A

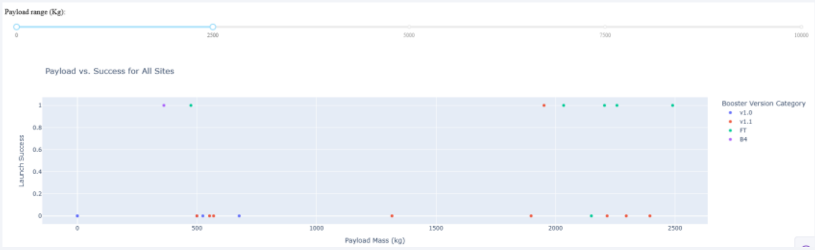
This pie chart shows the launch success distribution for the Kennedy Space Center Launch Complex 39A (**KSC LC-39A**). With a success rate of approximately **76.9%**, it is the highest among all launch sites in the dataset.

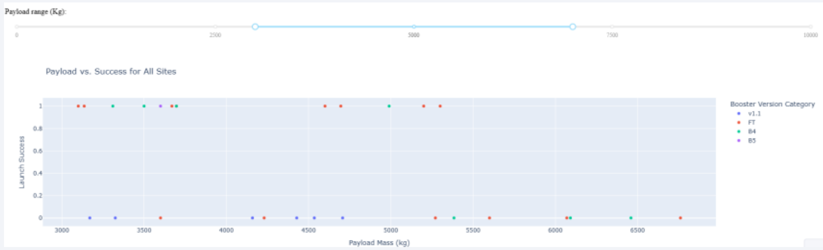
The chart highlights:

- A dominant share of successful missions (blue).
- A smaller proportion of failures (red), indicating a high mission reliability.
- Useful insight for identifying optimal launch sites in terms of performance.



Payload Mass vs. Launch Outcome Across Ranges





- **Insights:**
  - Shows all launch outcomes across the full payload mass spectrum (0–10,000 kg).
  - Booster versions B4, B5, and FT demonstrate strong success performance (class = 1).
  - FT boosters appear the most consistent in achieving successful launches across different payloads.

- **Insights:**
  - This range shows a high frequency of failures, especially for v1.0 and v1.1 boosters.
  - Only a few successful launches observed in this low-payload range.
- **Conclusion:**
  - Lower payload mass does not necessarily lead to higher success.
  - Booster version appears to be a more critical factor than payload weight in early flights.

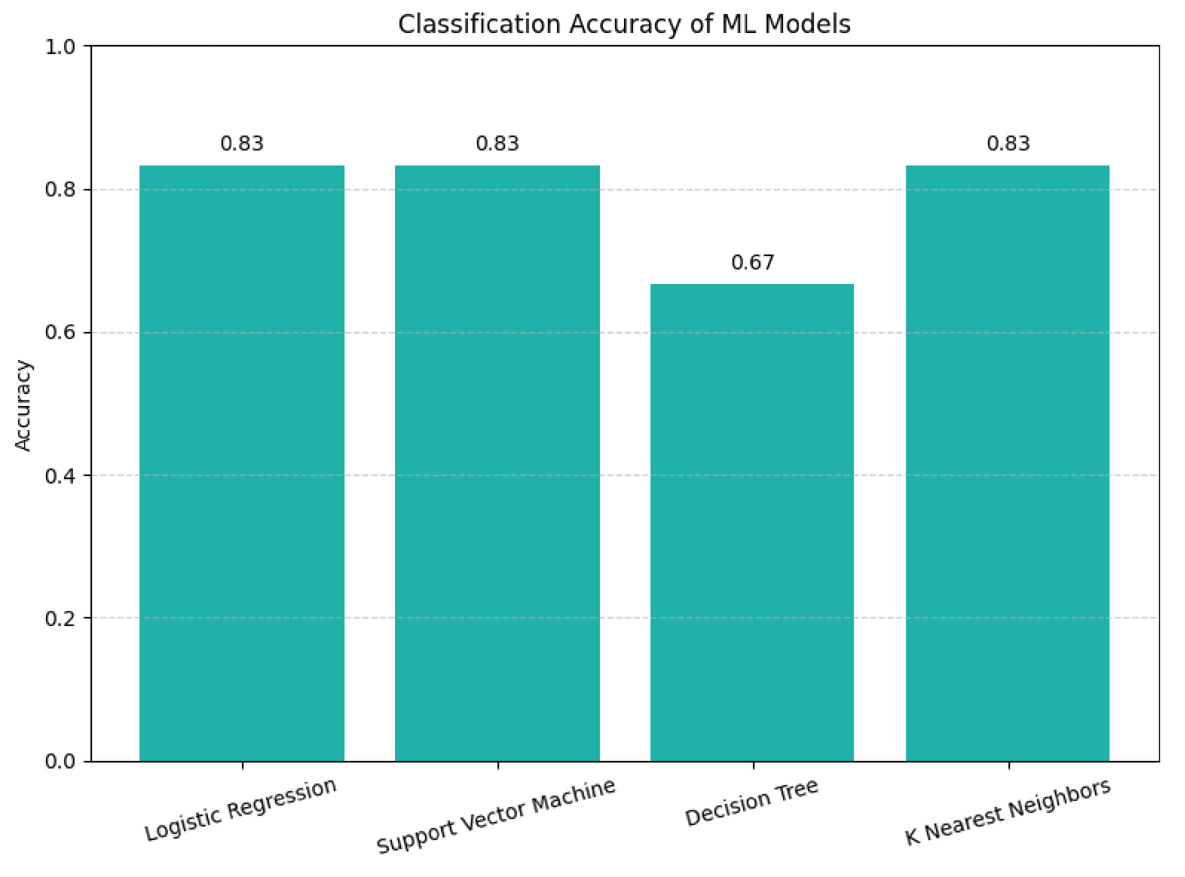
- **Insights:**
  - This range shows a high density of successful launches, particularly with FT, B4, and B5 boosters.
  - Successful launches are more concentrated in the 3000–6000 kg range.
- **Conclusion:**
  - Medium payloads (3000–6000 kg) combined with newer boosters yield the highest success rates.
  - Booster innovation over time has contributed to improved mission outcomes.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- **Objective:**
  - Compare the performance of all trained classifiers.
- **Visualization:**
  - Bar chart showing the accuracy scores of:
  - Logistic Regression
  - SVM
  - Decision Tree
  - K-Nearest Neighbors (KNN)
- **Insight:**
  - While Logistic Regression, SVM, and KNN all achieved an accuracy of 0.83, the model with the highest accuracy returned by the selection logic was Logistic Regression. This suggests that **Logistic Regression was the most robust choice** among the evaluated models for this classification task. The Decision Tree performed notably worse with an accuracy of 0.67.



Confusion Matrix

Objective:

Evaluate the classification performance of the best model, Logistic Regression, by visualizing its confusion matrix.

Visualization:

The confusion matrix summarizes the model's predictions on the test set:

- **True Positive (TP):** 12 — correctly predicted landings
- **False Positive (FP):** 3 — predicted landing, but did not land
- **True Negative (TN):** 3 — correctly predicted failures to land
- **False Negative (FN):** 0 — no landings missed by the model

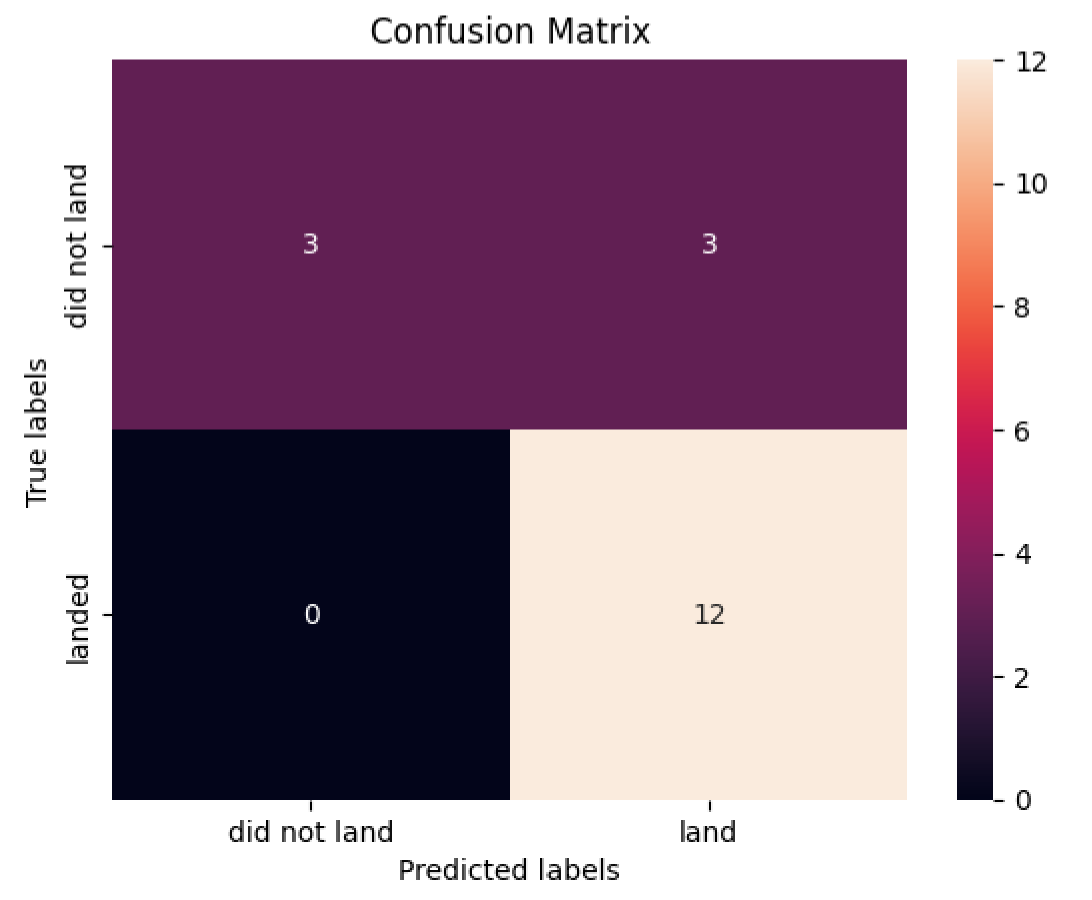
Insight:

Logistic Regression shows strong predictive performance with:

- **High recall** for the 'landed' class (no false negatives)
- **Some false positives**, indicating occasional overprediction of landings
- **Balanced classification capability**, with the main issue being limited false alarms

Interpretation:

The model effectively distinguishes between successful and unsuccessful landings. Its perfect recall and balanced accuracy (0.833) justify its selection as the top performer.



Conclusions

- **Launch Success Improved Over Time**
  SpaceX has significantly increased its launch success rate since 2013, reaching over 85% by 2020, indicating continuous operational improvement.
- **Payload Mass and Launch Site Are Related to Success**
  Missions with heavier payloads tend to have higher success rates when launched from KSC LC-39A and CCAFS SLC-40, suggesting these are better equipped for complex missions.
- **Orbit Type Influences Success**
  Orbits like LEO, ISS, and SSO show higher success rates than GTO, possibly due to differing technical and risk profiles.
- **Flight Number Matters in Certain Orbits**
  In the LEO orbit, a higher flight number (i.e., more experience) correlates with higher success, emphasizing learning effects and system reliability.
- **Proximity to Infrastructure is Strategic**
  Launch sites are typically located close to coastlines, roads, and railways, showing clear logistical planning in SpaceX's infrastructure decisions.
- **Logistic Regression Is a Strong Predictor**
  Among the ML models tested, Logistic Regression achieved the best predictive performance (83.3% accuracy) for mission success, proving effective in this binary classification task.
- **Data-Driven Decisions Are Feasible**
  The project demonstrated that success probability can be reasonably predicted from structured mission data, supporting future mission planning or resource allocation.

Appendix

- Python Libraries:
  - pandas, numpy, matplotlib, seaborn – for data wrangling and plotting
  - folium – for interactive geographic maps
  - plotly, dash – for interactive dashboards
  - sklearn – for building and evaluating ML models
  - sqlite3, sqlalchemy – for SQL-based analysis
- Visualization Types:
  - Scatter plots, bar charts, line charts
  - Interactive maps with markers, polylines, and popups
  - Confusion matrix and model accuracy comparison chart
- Other Tools:
  - GitHub – project version control and code sharing
  - Jupyter Notebooks – interactive coding and visualization
  - PowerPoint – final project storytelling

Thank you!