IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Prince Appiah
February 23, 2025

Outline

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

Executive Summary

Summary of methodologies

- Data Collection & Wrangling – Preprocessed and cleaned raw data for analysis.
- Exploratory Data Analysis (EDA) – Applied statistical & visual techniques for insights.
- SQL-Based EDA – Queried and analyzed data using SQL.
- Interactive Visual Analytics – Developed dynamic visualizations for deeper exploration.
- Interactive Mapping – Used Folium for geospatial data visualization.
- Plotly Dash Dashboard – Designed an interactive dashboard for results presentation.
- Predictive Analysis (Classification) – Built and evaluated machine learning models.

Executive summary

Summary of all results

- **Exploratory Insights:** Identified key patterns and trends through EDA and SQL analysis
- **Geospatial Analysis:** Developed an interactive Folium map to visualize location-based trends.
- **Dashboard Visualization:** Created an interactive Plotly Dash dashboard showcasing key findings.
- **Predictive Modeling:** Achieved **83.33% test accuracy** across Logistic Regression, SVM, Decision Tree, and KNN, indicating potential data-related factors. Further evaluation with precision, recall, and F1-score is needed.

Introduction

Project background and context

- This capstone project aims to predict whether the **Falcon 9 first stage** will land successfully. SpaceX advertises Falcon 9 rocket launches at **$62 million**, significantly lower than competitors, who charge upwards of **$165 million**. A key factor in this cost reduction is SpaceX's ability to **reuse the first stage**.
- By predicting the success of first-stage landings, we can estimate launch costs and provide valuable insights for potential competitors bidding against SpaceX. This project explores the problem, methodologies, and tools required to achieve this goal.

Introduction

Problems to answer

- Data Manipulation: Develop Python code to process and analyze data using Pandas.
- Data Conversion: Convert JSON files into Pandas DataFrames for structured analysis.
- Notebook & Collaboration: Create and share Jupyter notebooks via GitHub.
- Problem Definition: Apply data science methodologies to define and formulate a real-world business problem.
- Data Analysis: Load, clean, and extract meaningful insights from datasets.

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

The dataset was collected using two primary methods: API data retrieval and web scraping.

API Data Retrieval (SpaceX REST API)

- Collected launch data from api.spacexdata.com/v4/launches/past using requests.
- Extracted rocket specs, payload, launch & landing details.
- Converted JSON responses into a Pandas DataFrame.

Web Scraping (Wikipedia)

- Used **BeautifulSoup** to extract Falcon 9 launch data from HTML tables.

Data Collection – SpaceX API

Key Phrases

1. **API Call to SpaceX REST API** – Requesting launch data from http://api.spacexdata.com/v4/launches/past

2. **Receive JSON Response** – Extract structured launch data.

3. **Data Transformation** – Convert JSON to Pandas DataFrame using json_normalize().

4. **Additional API Calls** – Fetch Booster, Launchpad, Payload, and Core details.

5. **Data Cleaning** – Handle missing values and filter only **Falcon 9** launches.

View the notebook on GitHub

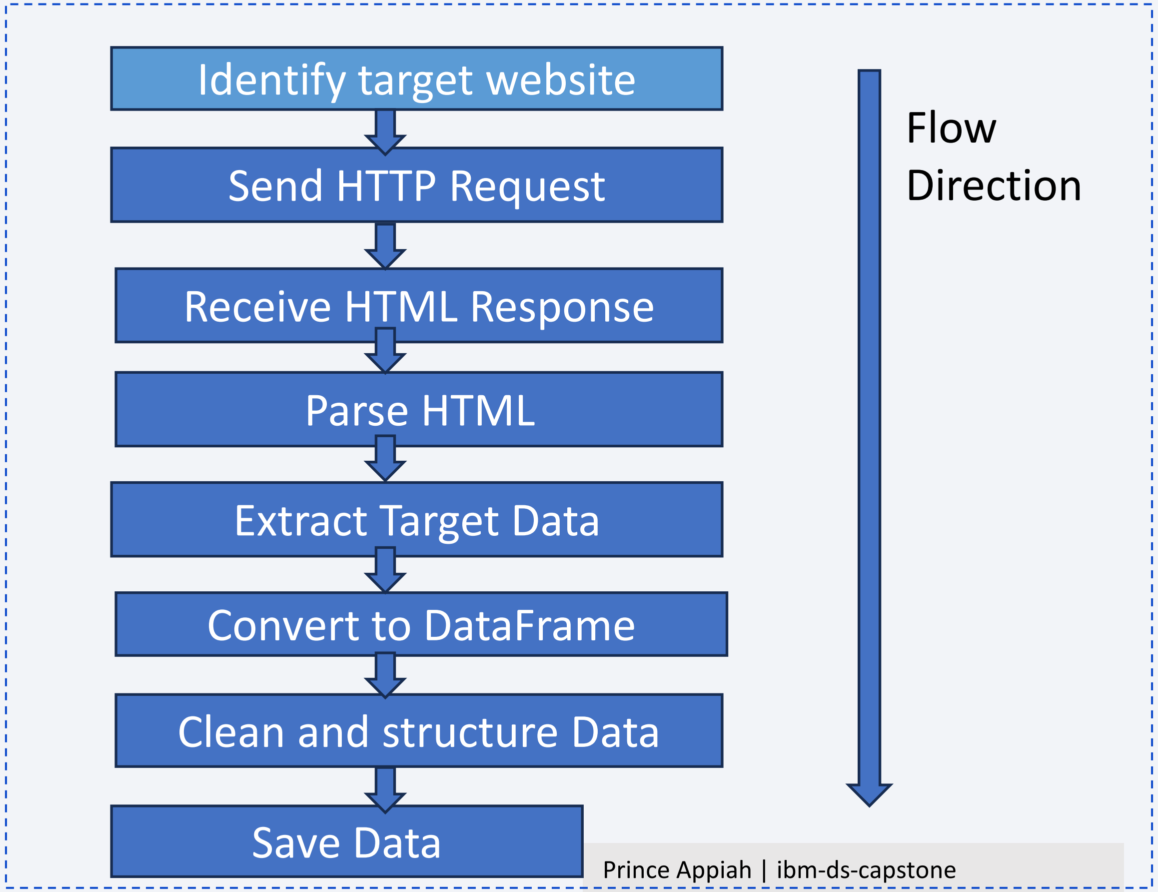
Prince Appiah | ibm-ds-capstone

Data Collection - Scraping

Key Phrases for Web Scraping Process

1. **Identify Target Website** – Wikipedia Falcon 9 launch records.
2. **Send HTTP Request** – Use requests library to fetch the webpage.
3. **Receive HTML Response** – Extract webpage content.
4. **Parse HTML** – Use BeautifulSoup to extract tables.
5. **Extract Target Data** – Find and clean relevant launch details.
6. **Convert to DataFrame** – Store extracted data in Pandas DataFrame.
7. **Clean & Structure Data** – Handle missing values, format dataset.
8. **Save Data** – Store for further analysis & modeling.

[web_scraping - GitHub]



Prince Appiah | ibm-ds-capstone

Data Wrangling

The data were processed as follows:

- **Feature Selection:** Extracted key attributes (Flight Number, Payload Mass, Orbit, Launch Site, Outcome, etc.).
- **Target Variable:** Converted **Outcome** into binary classes:

1 = Successful Landing, **0** = Unsuccessful Landing

- **Data Wrangling:** Cleaned missing values, encoded categorical features, and normalized numerical data for analysis.

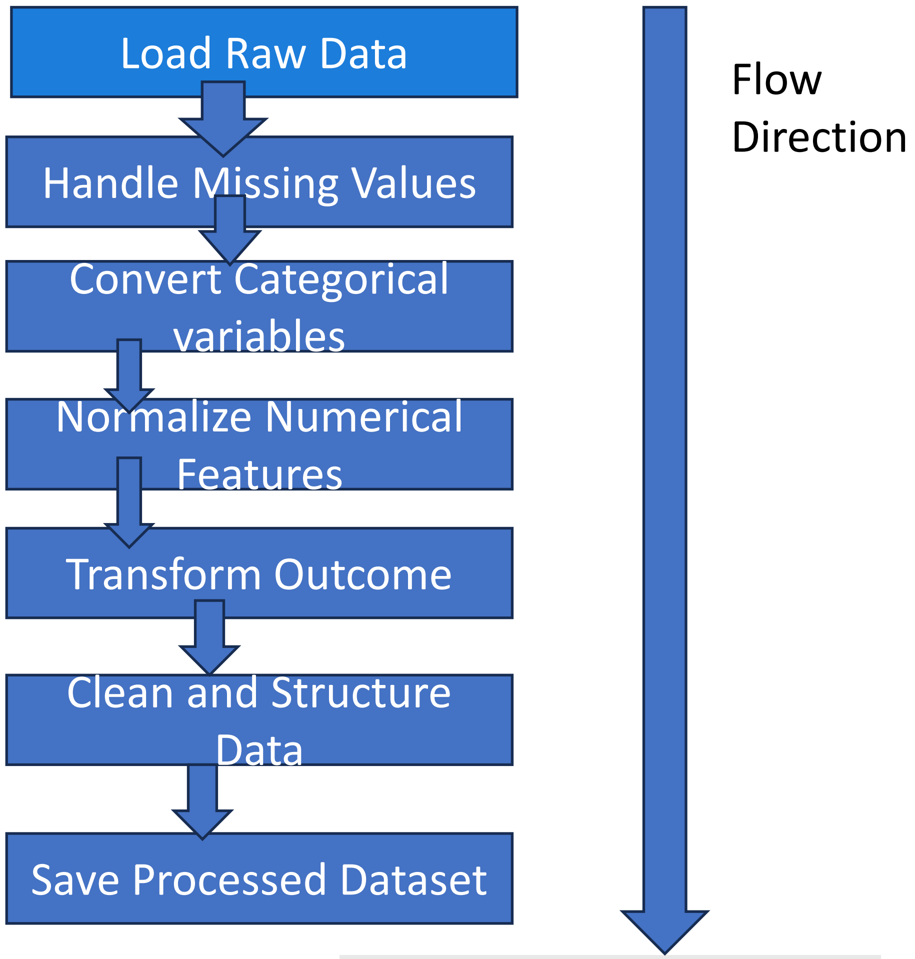
Data Wrangling 1 - GitHub

Data Wrangling 2 - GitHub

Data Wrangling

Key Phrases for Data Wrangling Process

1. **Load Raw Data** – Import dataset (API/Web Scraping).
2. **Handle Missing Values** – Fill, remove, or impute NULL values.
3. **Convert Categorical Variables** – Encode launch sites, orbits, and landing outcomes.
4. **Normalize Numerical Features** – Scale payload mass, longitude, latitude, etc.
5. **Transform Outcome Column** – Convert landing success to binary (0 = Failure, 1 = Success).
6. **Clean & Structure Data** – Ensure data consistency and proper format.
7. **Save Processed Dataset** – Store cleaned data for analysis & modeling.



EDA with Data Visualization

Summary of charts plotted and why we used those charts

Bar Chart – Compared success rates and orbit types to understand frequency distributions.

Line Chart – Visualized yearly launch success trends, highlighting changes in success rates over time.

Scatter Plot – Analyzed relationships, like FlightNumber and Orbit type, to identify correlations.

These charts provided clear insights into launch performance, trends, and key factors affecting Falcon 9 landings.

EDA with Data Visualization - GitHub

EDA with SQL

Summary of the SQL queries you performed

- ***Display the names of the unique launch sites in the space mission***

%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;

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

%sql SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM SPACEXTBL WHERE customer = 'NASA (CRS)';

- ***List the total number of successful and failure mission outcomes***

%sql SELECT Mission_Outcome, COUNT(*) AS total_count FROM SPACEXTBL GROUP BY Mission_Outcome ORDER BY total_count DESC;

[EDA SQL - GitHub]()

Build an Interactive Map with Folium

Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map

- **Markers** – Plotted launch site locations to visually identify where SpaceX missions took off.
- **Circles** – Represented impact areas or launch site zones, providing a clearer view of the **geographical region** around each launch site.
- **Lines (Polylines)** – Connected launch sites to landing locations, showing the flight path and distance between launch and landing sites.

Explain why you added those objects

- **Markers** help pinpoint exact launch site locations.
- **Circles** give a visual boundary around important areas.
- **Lines** illustrate trajectories and landing success patterns.

[Interactive map with Folium - GitHub]

Build a Dashboard with Plotly Dash

Plots/Graphs Added:

- **Pie Chart** – Displays total successful launches for all sites or Success vs. Failure counts for a selected site.
- **Scatter Plot** – Shows correlation between payload mass and launch success.
- **Interactions Added:**
- **Dropdown Filter** – Allows users to select a specific launch site or view data for all sites.

Why these were added.

- **Pie Chart** – Provides a quick summary of mission success rates.
  Scatter Plot – Helps analyze how payload mass impacts landing success.
  Dropdown Filter – Enhances interactivity and focused analysis.

[Plotly Dash - GitHub]

Predictive Analysis (Classification)

Model Development Process

Step 1: Data Preprocessing

- Standardized continuous features using StandardScaler before splitting the dataset.
- Applied stratified train-test split to maintain class balance.

Step 2: Model Selection & Hyperparameter Tuning

- Trained four models: Logistic Regression, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN)
- Used GridSearchCV (cv=10) to optimize hyperparameters.

Step 3: Model Evaluation

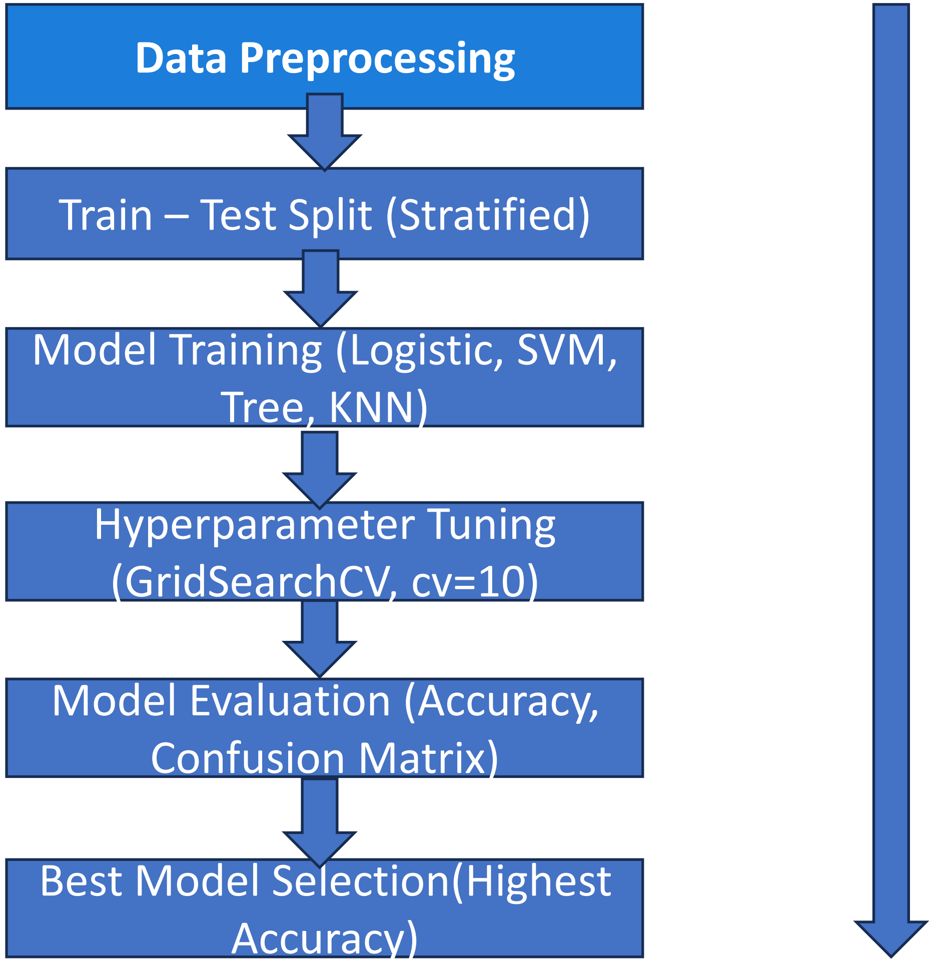
- Calculated test accuracy for each model.
- Analyzed confusion matrices to assess classification performance.

Step 4: Identifying the Best Model

Decision Tree had the highest accuracy.

Predictive Analysis(Classification) - GitHub

Predictive Analysis (Classification)



Flow Direction

Results

Exploratory data analysis results

1. Landing Success Trends

Early flights had more failures, but success rates improved significantly over time.

A major breakthrough occurred in 2015, when SpaceX successfully landed a Falcon 9 booster for the first time.

By 2019, success rates reached ~90%, reflecting technological advancements.

2. Launch Site Performance

Most launches occurred from CCAFS SLC 40 and KSC LC 39A, with increasing success rates over time.

VAFB SLC 4E had fewer launches but a high success rate.

3. Orbital Success Rates & Challenges

ES-L1, GEO, HEO, and SSO achieved a 100% success rate, indicating reliable landings.

GTO had the lowest success rate, suggesting greater landing challenges.

LEO and ISS missions had moderate to high success rates.

Results

Exploratory data analysis results

4. Payload & Landing Success

Higher payload missions were mostly successful, showing improved booster performance over time.

Lower payload missions had a mix of successes and failures, particularly in LEO and GTO.

5. Booster Performance & Reusability

Falcon 9 Block 5 became the most powerful and reusable booster, handling the highest payloads.

Drone ship landings were used for high-velocity missions, increasing reusability.

SpaceX achieved 98 successful missions, demonstrating high reliability.

6. Impact on SpaceX's Mission Success

NASA's CRS missions successfully delivered 45,596 kg of cargo to the ISS.

SpaceX's reusable rocket technology has significantly reduced launch costs and improved efficiency.

Results

Interactive Analytics Demo in Screenshots

Launch Sites & Frequencies

- Major launch sites include KSC LC-39A, CCSFS LC-40 (East Coast), and VAFB SLC-4E (West Coast), strategically chosen for optimal launch trajectories.
- Florida sites (KSC LC-39A & CCSFS LC-40) have the highest launch concentration.

Launch Success Trends

- 42.9% of launches were successful, showing improvements in rocket recovery.
- KSC LC-39A had the highest success rate (76.9%), indicating refined landing technology.

Results

Interactive Analytics Demo in Screenshots

Booster & Payload Impact:

- Older boosters (v1.0, v1.1) had lower success rates, especially for low payloads.
- Newer boosters (FT, B4, B5) significantly improved success rates, even for heavier payloads.

Key Takeaways:

- SpaceX's technology advancements have led to improved landing success.
- Launch sites are strategically positioned for logistics and operational efficiency.
- Newer booster versions and optimized landing strategies have increased mission reliability.

Results

Predictive Analysis

Best Performing Model:

- The Decision Tree model achieved the highest accuracy of 0.8333.

Confusion Matrix Insights:

- The model correctly identified all successful landings (12/12 correct).
- False Positives: 3, meaning it predicted landings that did not actually occur.
- False Negatives: 0, indicating 100% sensitivity in detecting actual successful landings.

Key Takeaways:

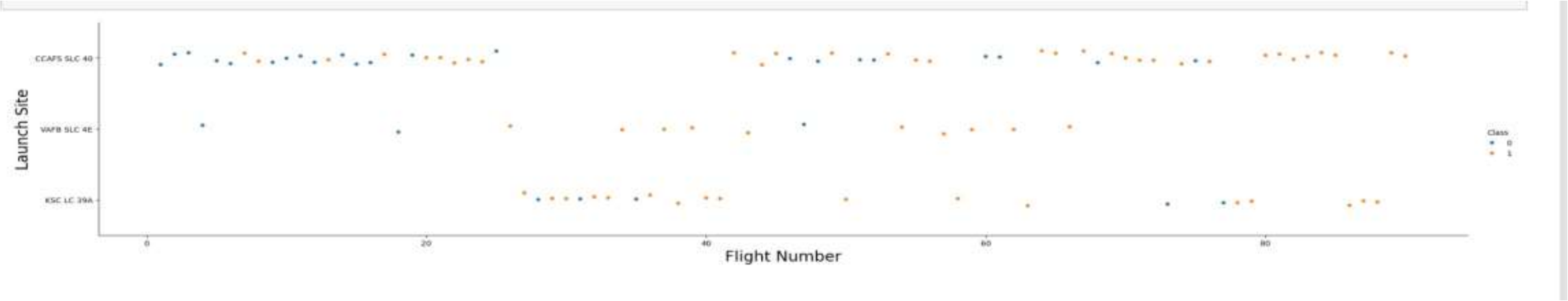
- The model is highly effective in identifying successful landings.
- False positives remain a challenge, suggesting potential for further refinement.
- The absence of false negatives makes it a reliable choice for predicting successful landings.

Section 2

Insights drawn from EDA

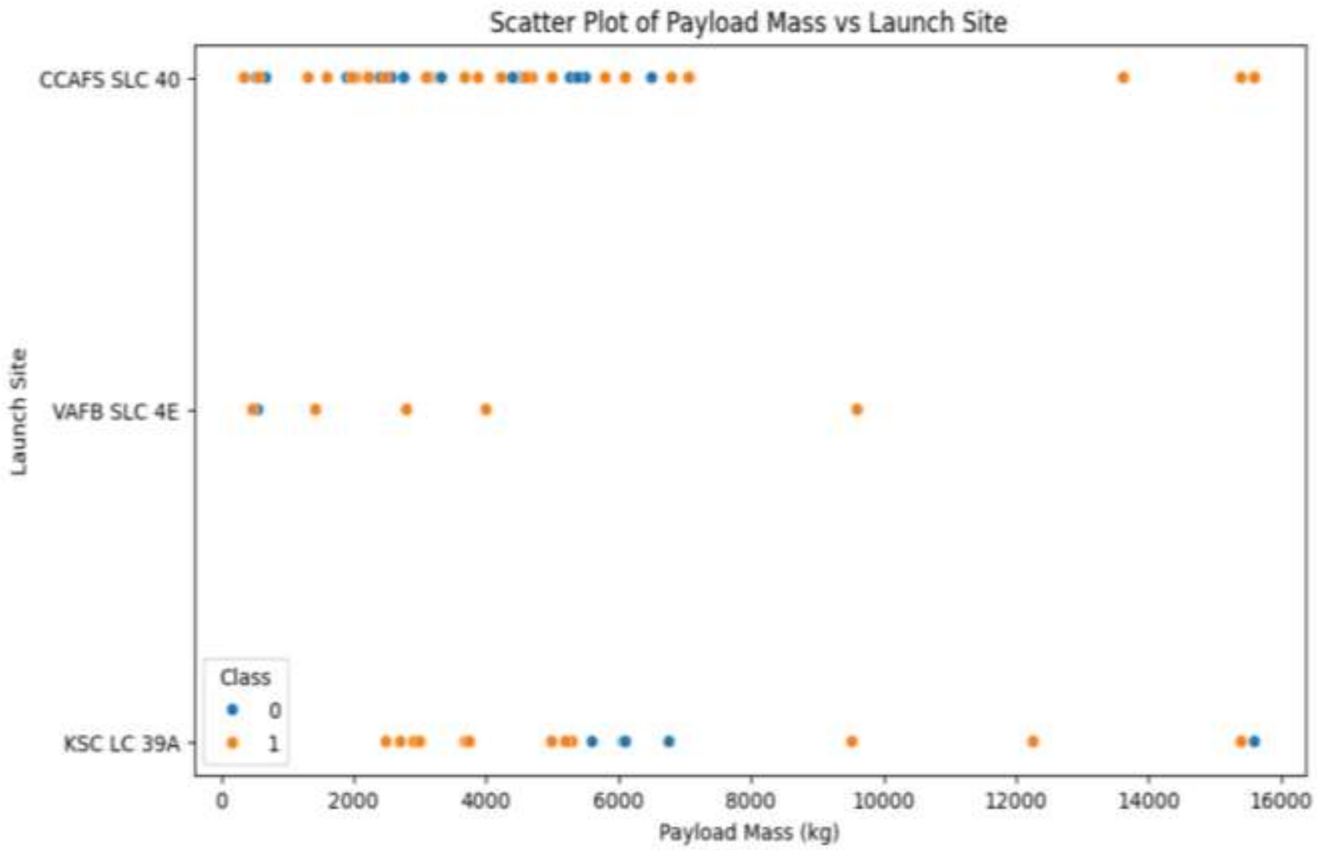
Flight Number vs. Launch Site

- Each dot represents a rocket launch, with color indicating successful (orange) vs. unsuccessful (blue) landings.
- Early flights had more failures (blue dots), indicating improvements over time.
- CCAFS SLC 40 and KSC LC 39A have more launches, with a higher success rate in later flights.
- VAFB SLC 4E has fewer launches overall, but most were successful.



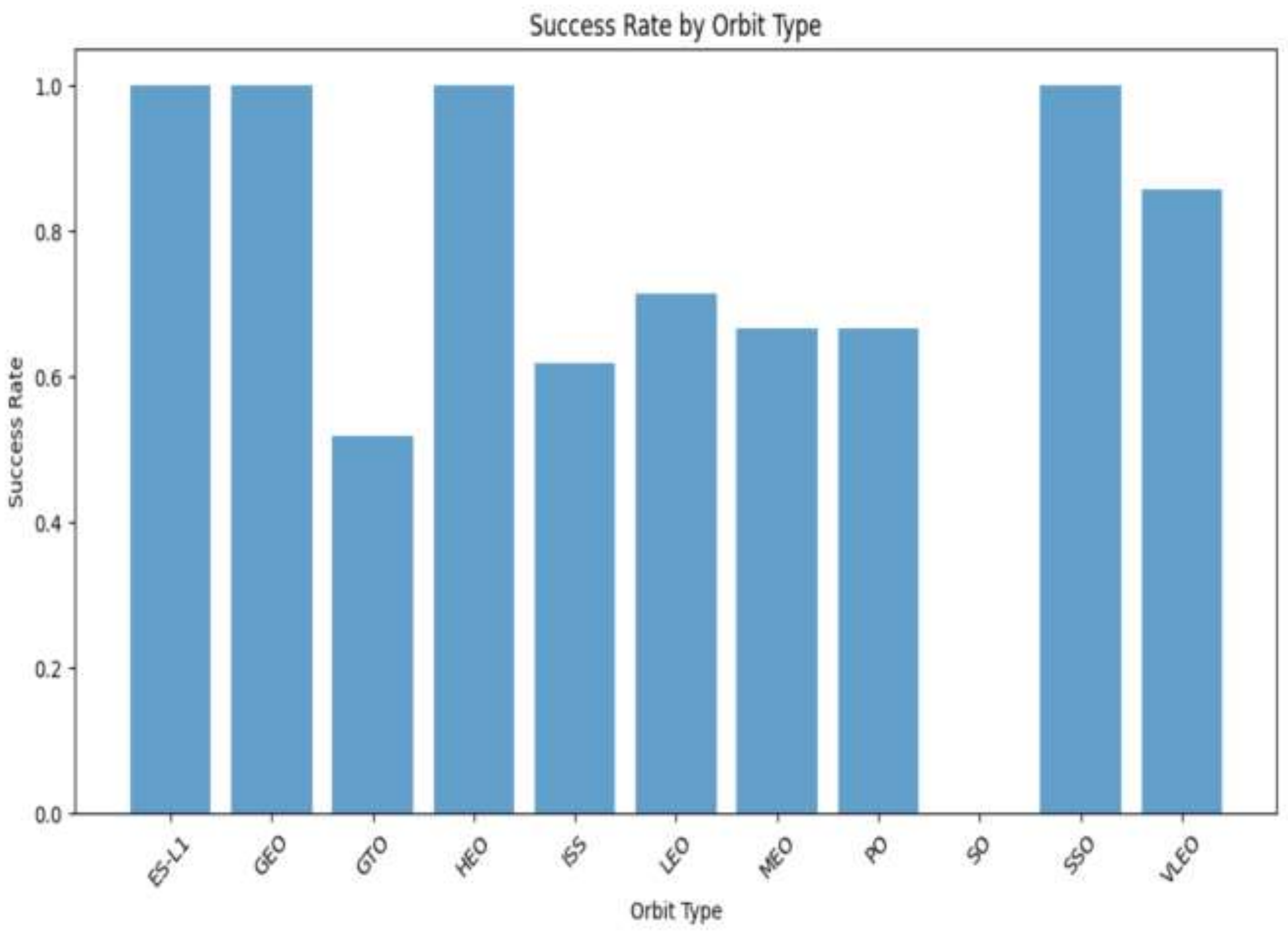
Payload vs. Launch Site

- Color-coded points indicate successful (orange) vs. unsuccessful (blue) landings.
- Most launches from KSC LC 39A and CCAFS SLC 40 resulted in successful landings (orange dots).
- Few unsuccessful landings (blue dots), concentrated around low to mid-range payload masses.
- VAFB SLC 4E has fewer launches overall, but most were successful.



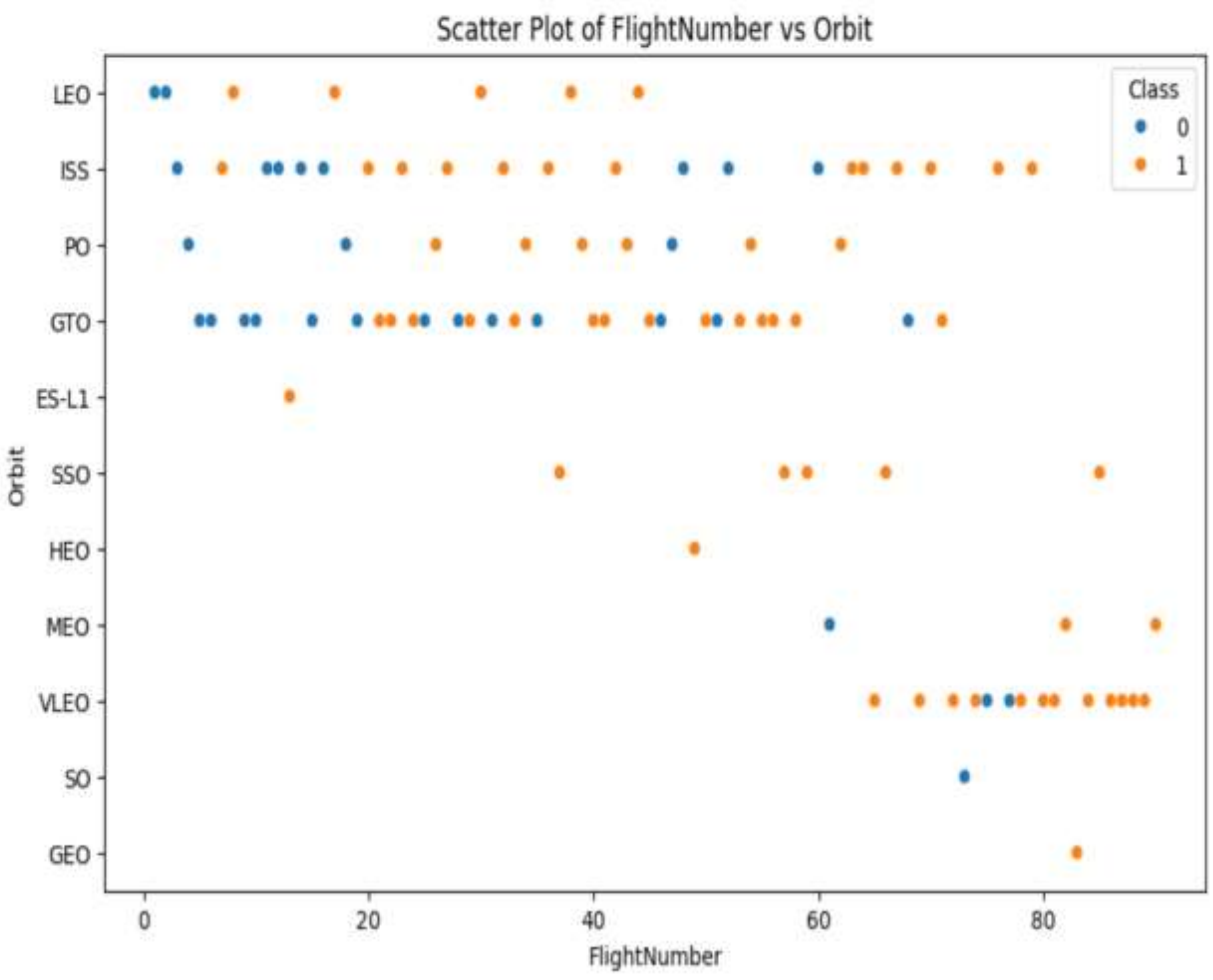
Success Rate vs. Orbit Type

- ES-L1, GEO,HEO and SSO have a 100% success rate, indicating reliable performance for landings.
- GTO has the lowest success rate, suggesting it may pose more challenges for landings.
- ISS, LEO, MEO, and PO have moderate success rates, highlighting some variability in landing success.



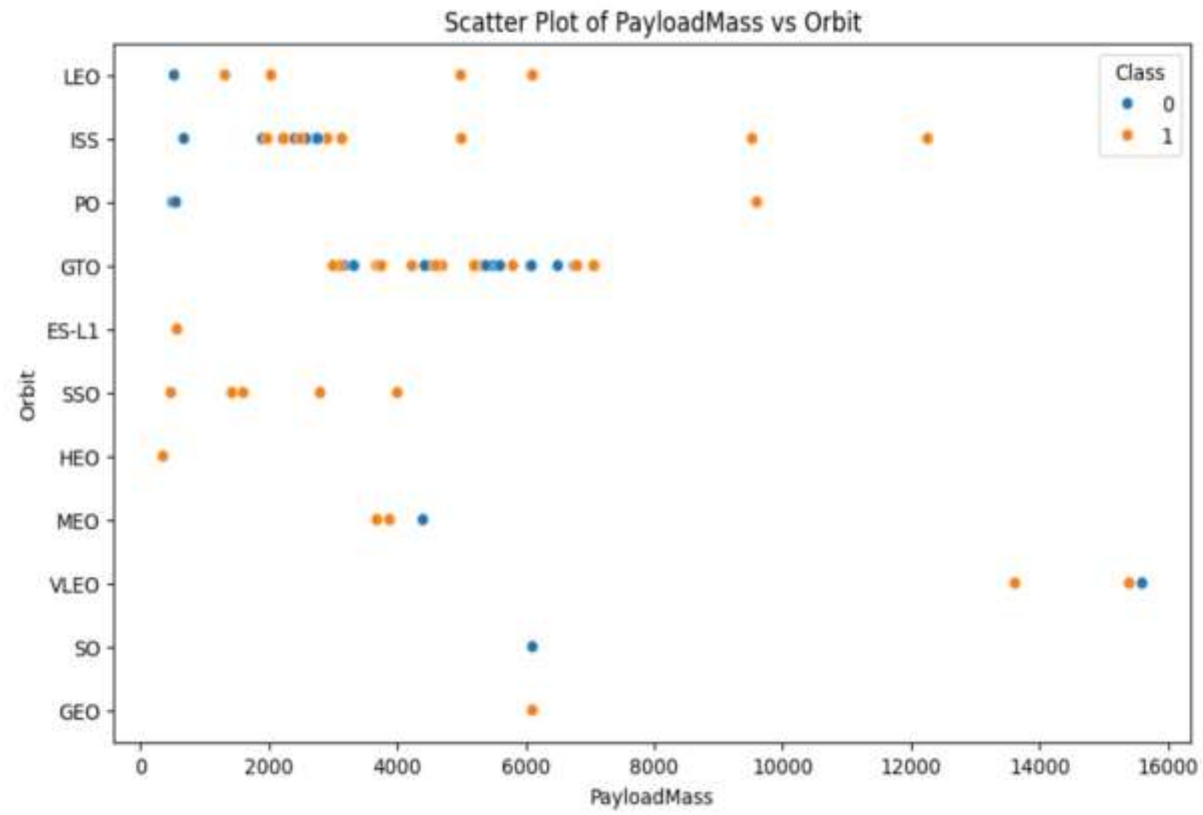
Flight Number vs. Orbit Type

- Early flights had more failures, suggesting improvements over time.
- LEO, ISS, and SSO show high success rates, while GTO has a mix of successes and failures.
- Recent missions have a higher proportion of successful landings, reflecting technological advancements.
- The trend suggests mission success has improved with experience. Certain orbits, like GTO, may be more challenging for successful landings.



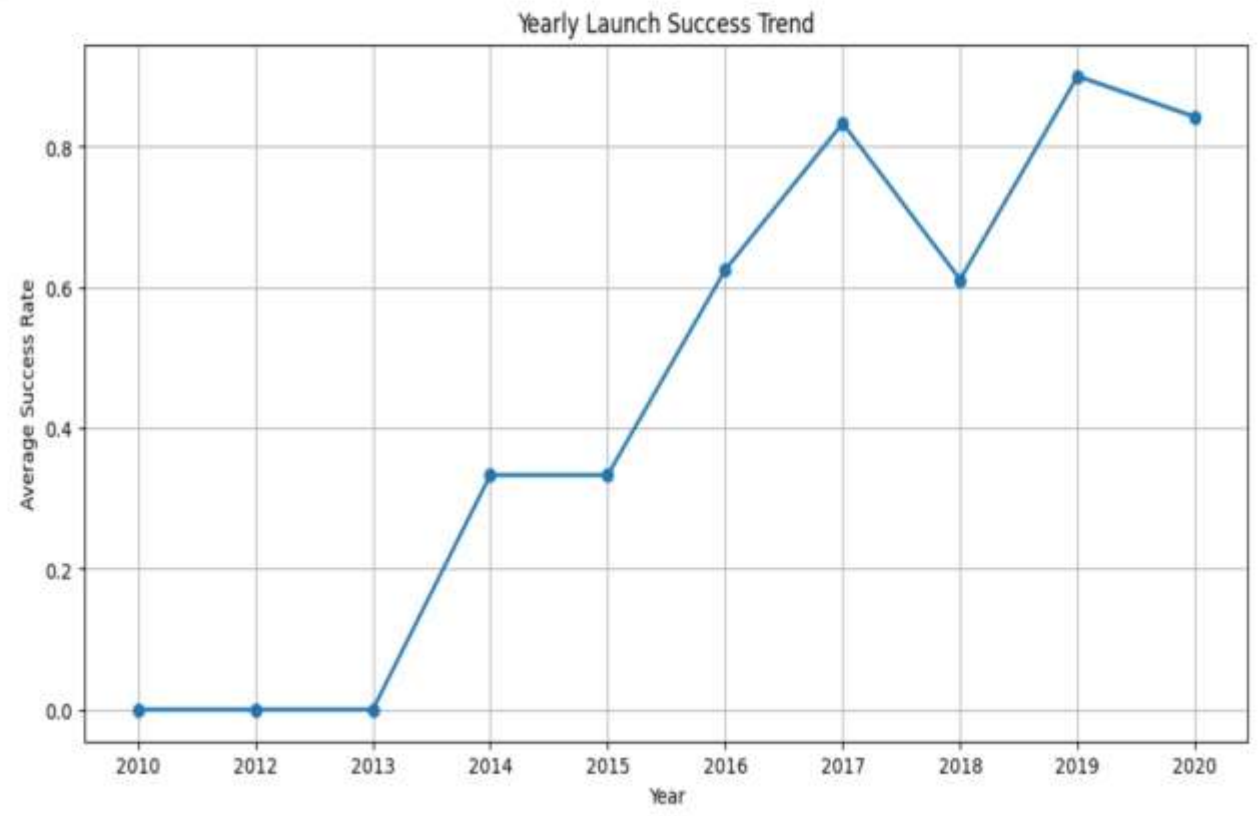
Payload vs. Orbit Type

- Lower payload masses tend to have a mix of successes and failures, particularly in LEO and GTO.
- Higher payloads are mostly successful, suggesting better handling of heavier payloads over time.
- GTO has a notable mix of successes and failures, indicating greater difficulty in landing for this orbit.
- SSO and ES-L1 show a high success rate, likely due to optimized launch parameters.



Launch Success Yearly Trend

- From 2010 to 2013, there were no successful landings, reflecting the early testing phase.
- Starting in 2014, success rates began to improve, likely due to advancements in technology and landing strategies.
- A steady increase is observed from 2015 to 2017, with some fluctuations.
- The highest success rate (~90%) was achieved in 2019, showing significant progress.
- A slight dip in 2020 suggests external factors may have impacted landings.
- The data suggests that landing technology and operational efficiency improved over time.



All Launch Site Names

Launch Sites Identified:

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

- The presence of CCAFS LC-40 and CCAFS SLC-40 suggests Cape Canaveral Air Force Station (CCAFS) has multiple launch complexes.
- KSC LC-39A is part of Kennedy Space Center (KSC), historically used for major space missions.
- VAFB SLC-4E (Vandenberg Air Force Base) is typically used for polar orbits and military payloads.
- The variety of launch sites indicates that different locations are used based on mission requirements, such as orbit type, payload weight, and mission objective.

Launch Site Names Begin with 'CCA'

- Launch Site: All missions in this table were launched from CCAFS LC-40.
- Booster Versions: Different booster versions (F9 v1.0 B0003 - B0007) were used.
- Payloads: Various payloads, including Dragon spacecraft and cargo for NASA missions, were launched.
- Orbit: All missions targeted LEO (Low Earth Orbit), specifically the ISS for cargo resupply.
- Mission Outcome: Every mission listed was a success.
- Landing Outcome:

- Early missions had failures (e.g., parachute failures).

- Some missions had "No attempt" for landing, indicating that early launches did not include booster recovery.

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

Total Payload Mass

Total Payload Mass: 45,596 kg

Explanation

- NASA's CRS (Commercial Resupply Services) missions are primarily designed to deliver cargo to the International Space Station (ISS).
- The total payload mass of 45,596 kg represents the sum of all payloads launched under NASA CRS contracts.
- These missions utilize Falcon 9 boosters, reflecting SpaceX's role in ISS resupply efforts.
- The substantial mass transported highlights the efficiency and reliability of Falcon 9 for repeated space deliveries.

Average Payload Mass by F9 v1.1

Average Payload Mass: 2534.67 kg

- The F9 v1.1 booster version was an upgraded Falcon 9 model, featuring increased thrust and better payload capacity.
- An average payload mass of 2534.67 kg indicates that most missions using F9 v1.1 carried mid-range payloads.
- This booster was commonly used for low-Earth orbit (LEO) and ISS resupply missions, balancing efficiency and reusability.

First Successful Ground Landing Date

Date: 2015-12-22

Explanation

- This was the first time SpaceX successfully landed a Falcon 9 booster on a ground pad.
- The achievement marked a major milestone in rocket reusability, proving that boosters could be recovered and reused.
- This event paved the way for cost-effective space travel, reducing launch costs significantly.

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster Versions Identified

- F9 FT B1022
  F9 FT B1026
  F9 FT B1021.2
  F9 FT B1031.2
- These boosters successfully landed on drone ships, showcasing SpaceX's advancements in reusable rocket technology.
- Their payload masses ranged between 4000 and 6000 kg, which indicates moderate payload missions.
- Drone ship landings are typically used for high-velocity missions, where boosters do not have enough fuel to return to a ground pad.
- These missions contribute to SpaceX's cost reduction strategy, allowing boosters to be reused for multiple launches.

Total Number of Successful and Failure Mission Outcomes

Mission Outcome	Total Count
Success	98
Success(Payload status unclear)	1
Success	1
Failure(in flight)	1

- 98 missions were successful, demonstrating high reliability in SpaceX launches.
- One mission had an unclear payload status, meaning the rocket landed successfully, but there may have been issues with the payload.
- One additional success is listed, possibly a duplicate or different classification of mission success.
- One mission failed in flight, indicating an anomaly that led to mission failure.
- Overall, SpaceX has achieved a high mission success rate, with very few failures. This reflects continuous technological improvements and operational efficiency.

Boosters Carried Maximum Payload

Booster Versions Identified:

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

- All listed boosters are from the Falcon 9 Block 5 (F9 B5) series, SpaceX’s most powerful and reusable booster version.
- These boosters carried the highest payload masses, showcasing their capability for heavy-lift missions.

2015 Launch Records

Month	Launching Outcome	Booster Version	Launch Site
01	Failure(drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure(drone ship)	F9 v1.1 B1015	CCAFS LC-40

- In January and April 2015, two Falcon 9 v1.1 boosters (B1012 & B1015) attempted landings on drone ships but failed.
- These failures were part of early booster recovery tests, leading to refinements in landing technology and precision control.
- CCAFS LC-40 was the launch site for both missions, highlighting its frequent use for SpaceX launches.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20 in descending order.

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

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20 in descending order.

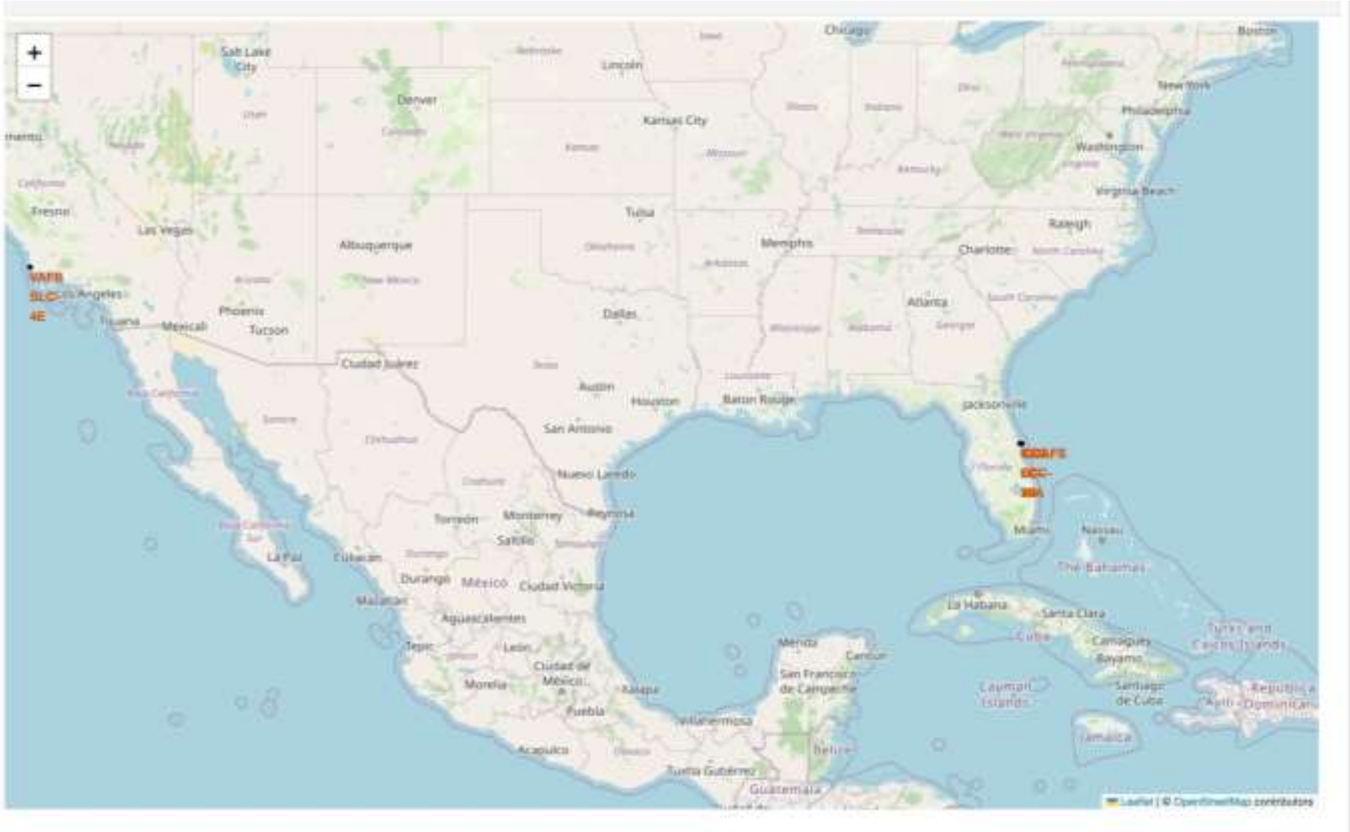
- 10 launches had "No attempt" for landing, indicating early missions did not focus on booster recovery.
- Drone ship landings had a 50% success rate (5 successes, 5 failures), highlighting the challenges of ocean-based recovery.
- Ground pad landings were fewer (3) but consistently successful, showcasing the increased precision of land-based recovery.
- Controlled and uncontrolled ocean landings (5 total) indicate SpaceX’s early-stage recovery efforts.
- Parachute failures (2) suggest earlier recovery attempts relied on different techniques.

Section 3

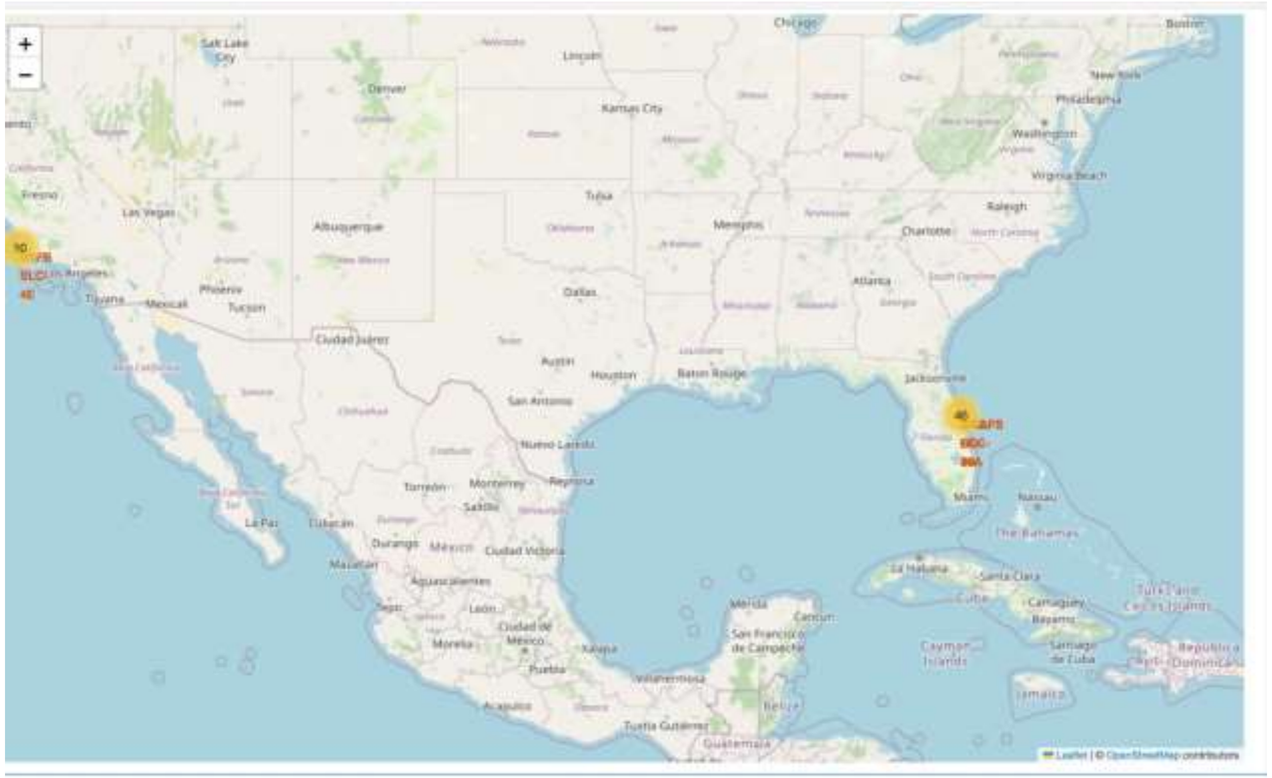
Launch Sites Proximities Analysis

Launch Sites Overview

- This map provides a global view of SpaceX launch sites across the United States.
- The major launch sites are labeled, including:
- **Vandenberg Space Force Base (VAFB SLC-4E)** on the West Coast.
- **Cape Canaveral Space Force Station (CCSFS LC-40)** and **Kennedy Space Center (KSC LC-39A)** on the East Coast.
- These locations are strategically chosen for optimal launch trajectories.



Mapping Launch Success and Failures Across Sites



- The heatmap visualizes the frequency of launches from different sites. The size and color of the circles indicate the number of launches:
- **Larger and more intense color circles** represent higher launch counts.
- **Florida launch sites** (CCSFS LC-40 and KSC LC-39A) have the highest concentration of launches.
- **Vandenberg (VAFB SLC-4E)** shows fewer launches compared to Florida sites.

Proximity Analysis of a Launch Site

Key Elements:

- Distance markers show the launch pad's location relative to key infrastructure:
  - **Highway:** ~0.60 KM away.
  - **Railway:** ~0.51 KM away.
- Color-coded lines illustrate connections to transportation routes.

Key Takeaways:

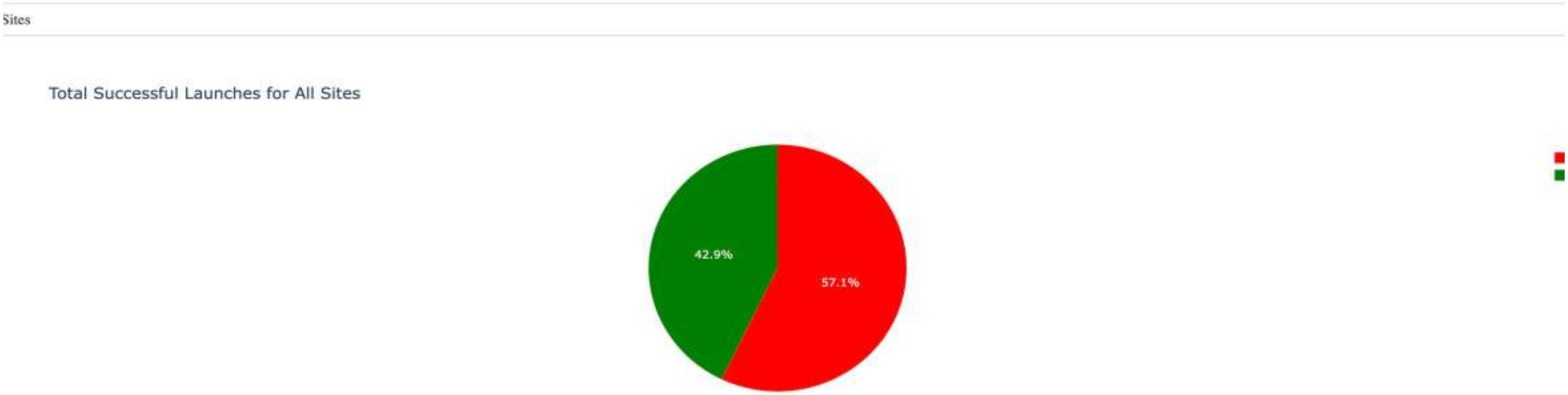
- Launch sites are located near key transportation networks for logistics and emergency access.
- Proximity to infrastructure is essential for efficient operations and support.

Section 4

Build a Dashboard with Plotly Dash

Overall Launch Success Rate for All Sites

- Legend:0 (Red) → Failed launches. 1 (Green) → Successful launches.
- More than half (57.1%) of the launches were unsuccessful, indicating early challenges in landing attempts.
- 42.9% of the launches were successful, reflecting significant improvements in rocket recovery technology over time.
- The data highlights SpaceX’s transition toward achieving higher landing success rates through continuous technological advancements.

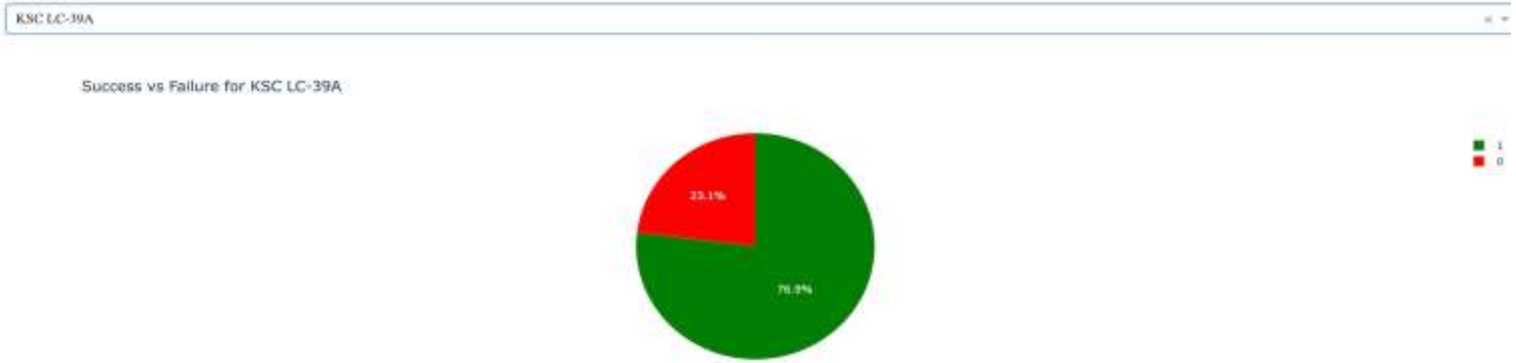


Success Rate Analysis for the Top-Performing Launch Site

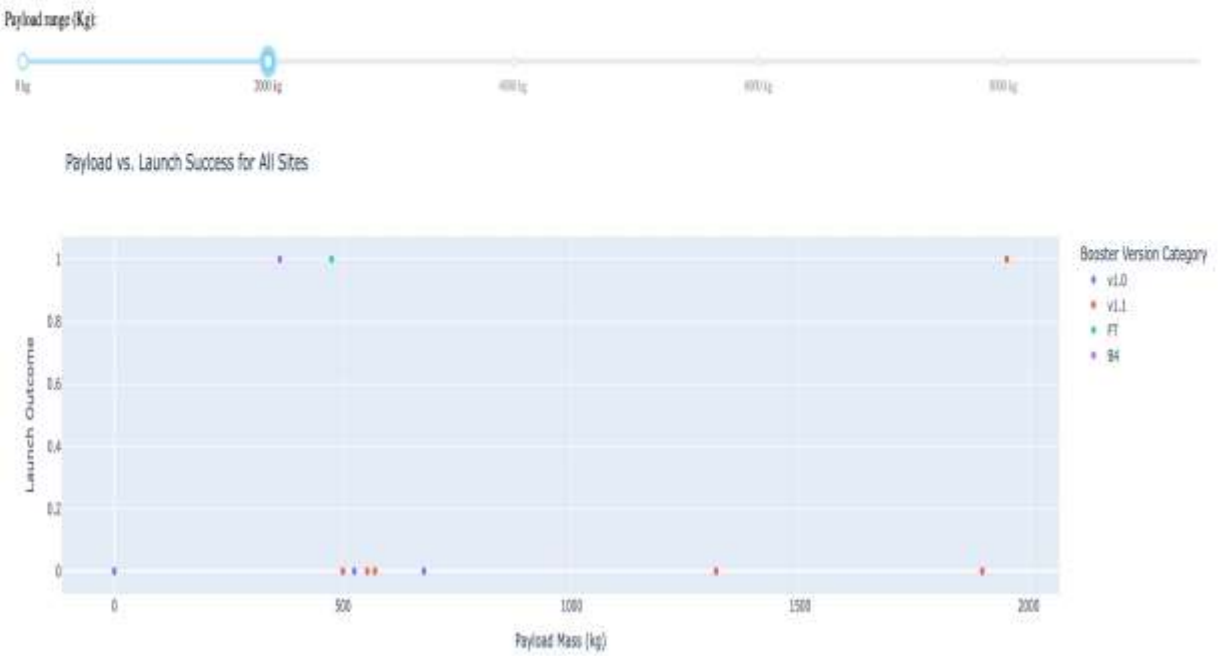
- KSC LC-39A has a high success rate (76.9%), indicating that most missions launched from this site were successful.

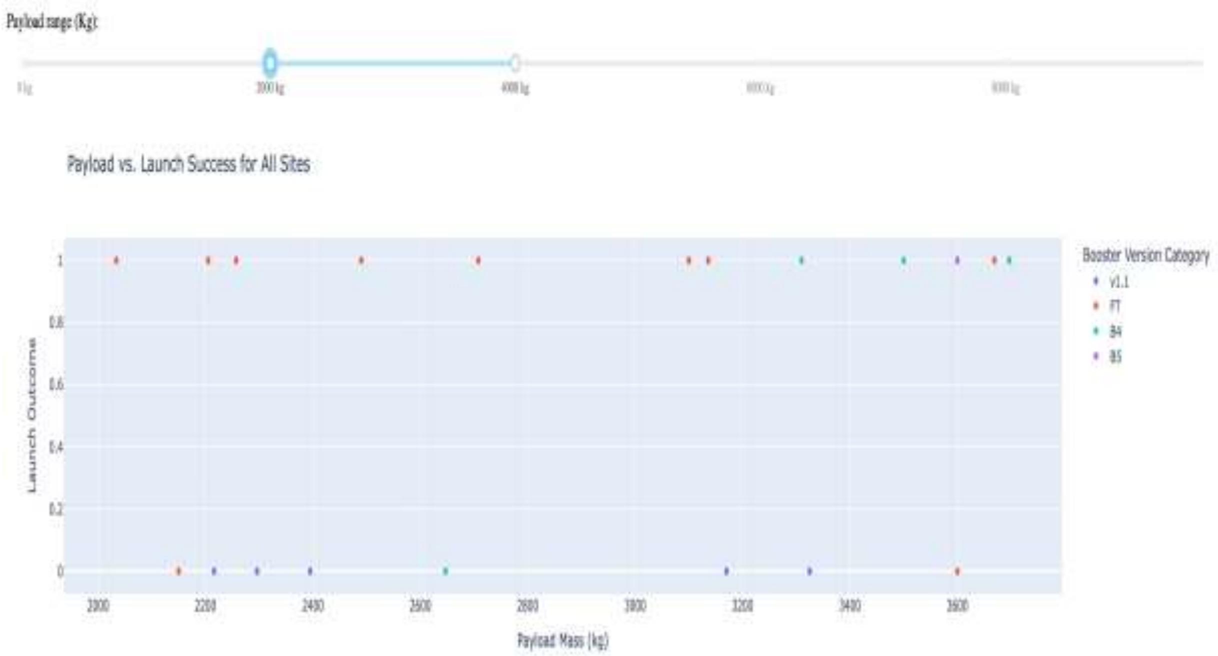
- Only 23.1% of launches failed, suggesting improvements in landing technology and mission execution.

- Compared to other launch sites, KSC LC-39A is a historically significant site, often used for major missions, including crewed flights.

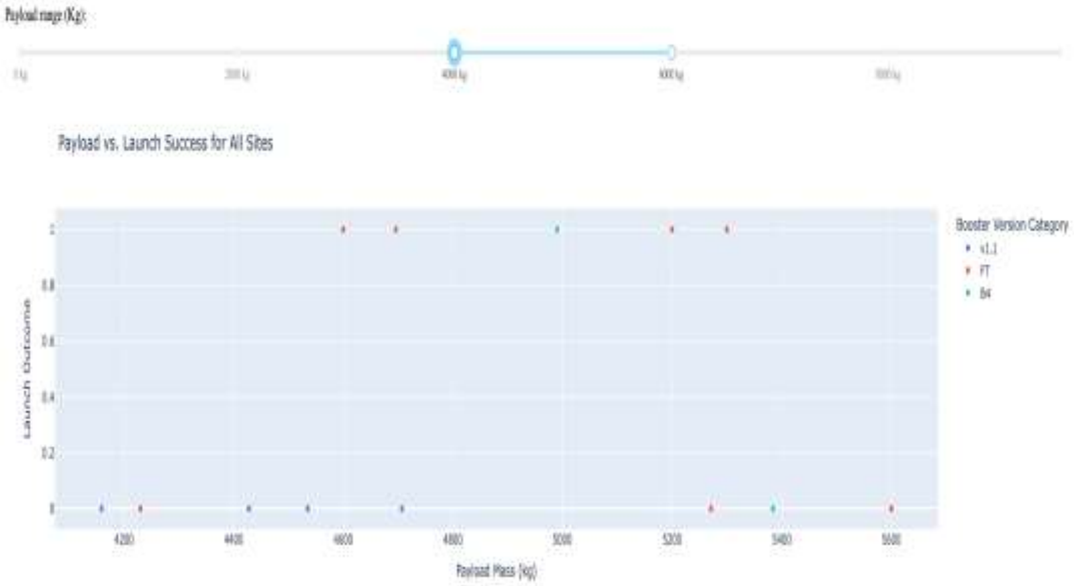


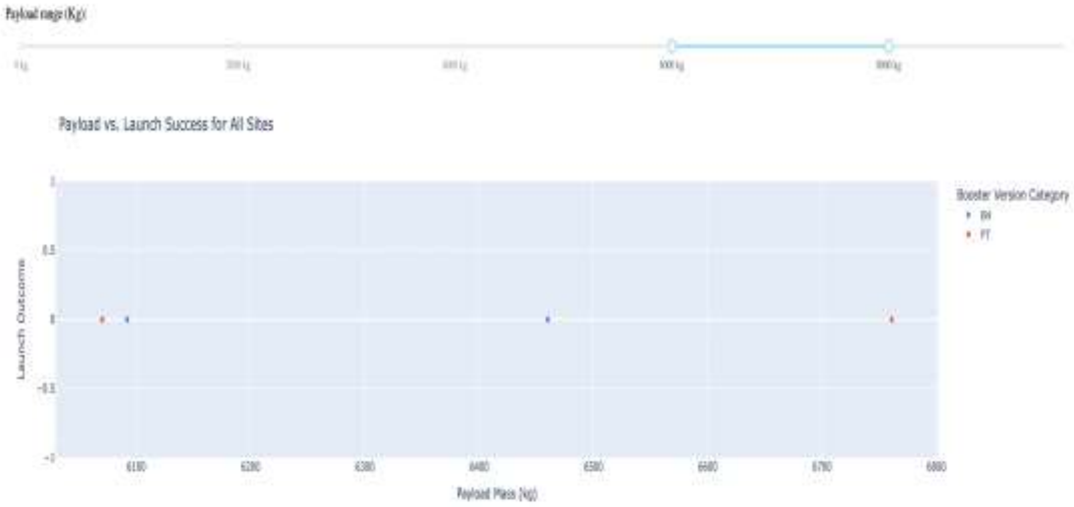
Impact of Payload and Booster Version on Launch Success





Impact of Payload and Booster Version on Launch Success





Impact of Payload and Booster Version on Launch Success

Key Findings from Payload vs. Launch Outcome Scatter Plots

- **Booster Version Impact:**
  - Older versions (**v1.0, v1.1**) had **lower success rates**, especially in low payload ranges.
  - Newer boosters (**FT, B4, B5**) **significantly improved launch success rates**, even for higher payloads.
- **Payload Mass vs. Success:**
  - Low Payloads (<2000 kg): Mixed results with **some failures** for older boosters.
  - Medium Payloads (2000–4000 kg): Success rate increases as newer boosters dominate.
- **High Payloads (>4000 kg): Predominantly successful launches** with FT, B4, and B5 boosters.

Technological Evolution:

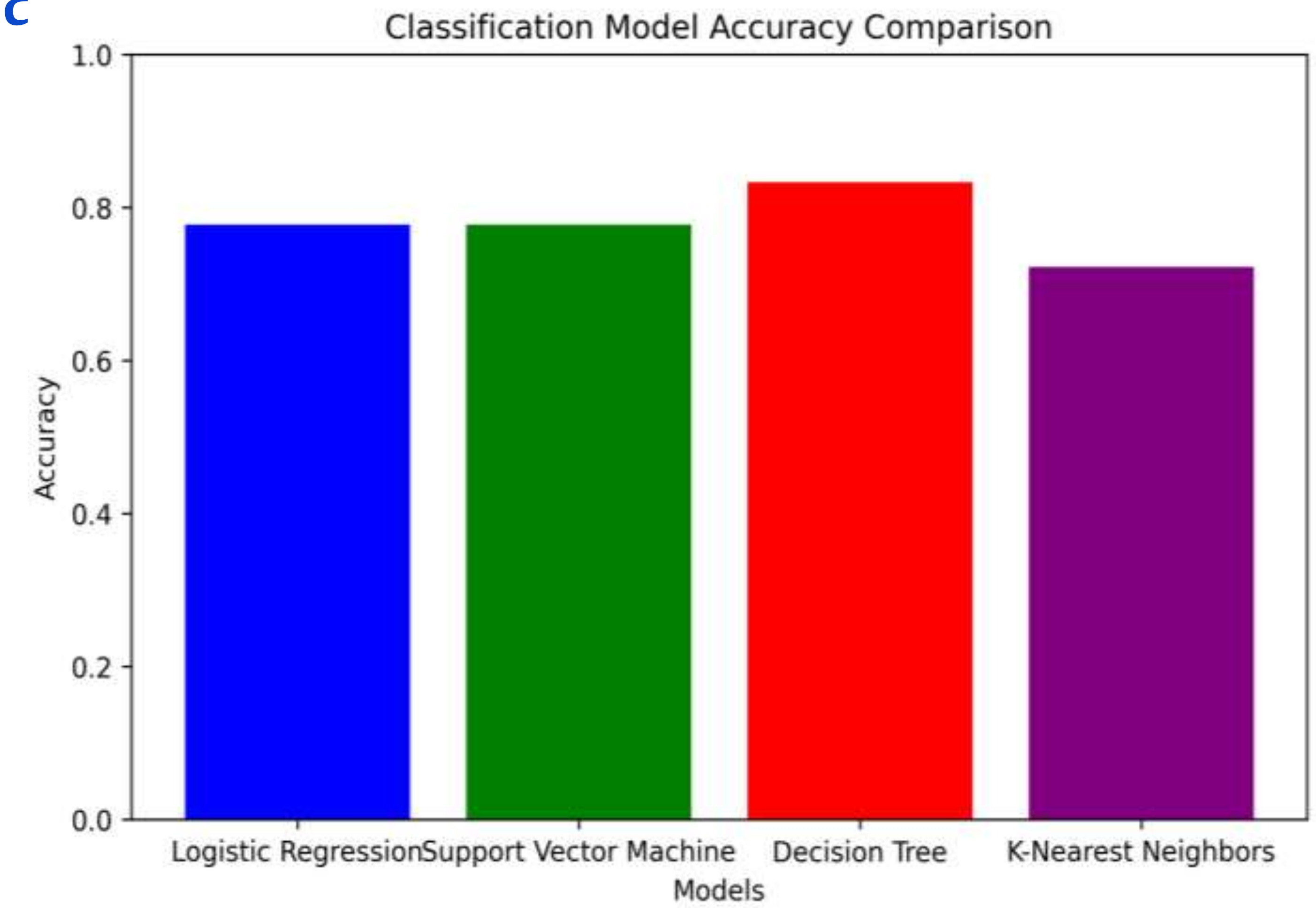
- Over time, **SpaceX's booster upgrades improved launch reliability**, especially for **heavier payloads**.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

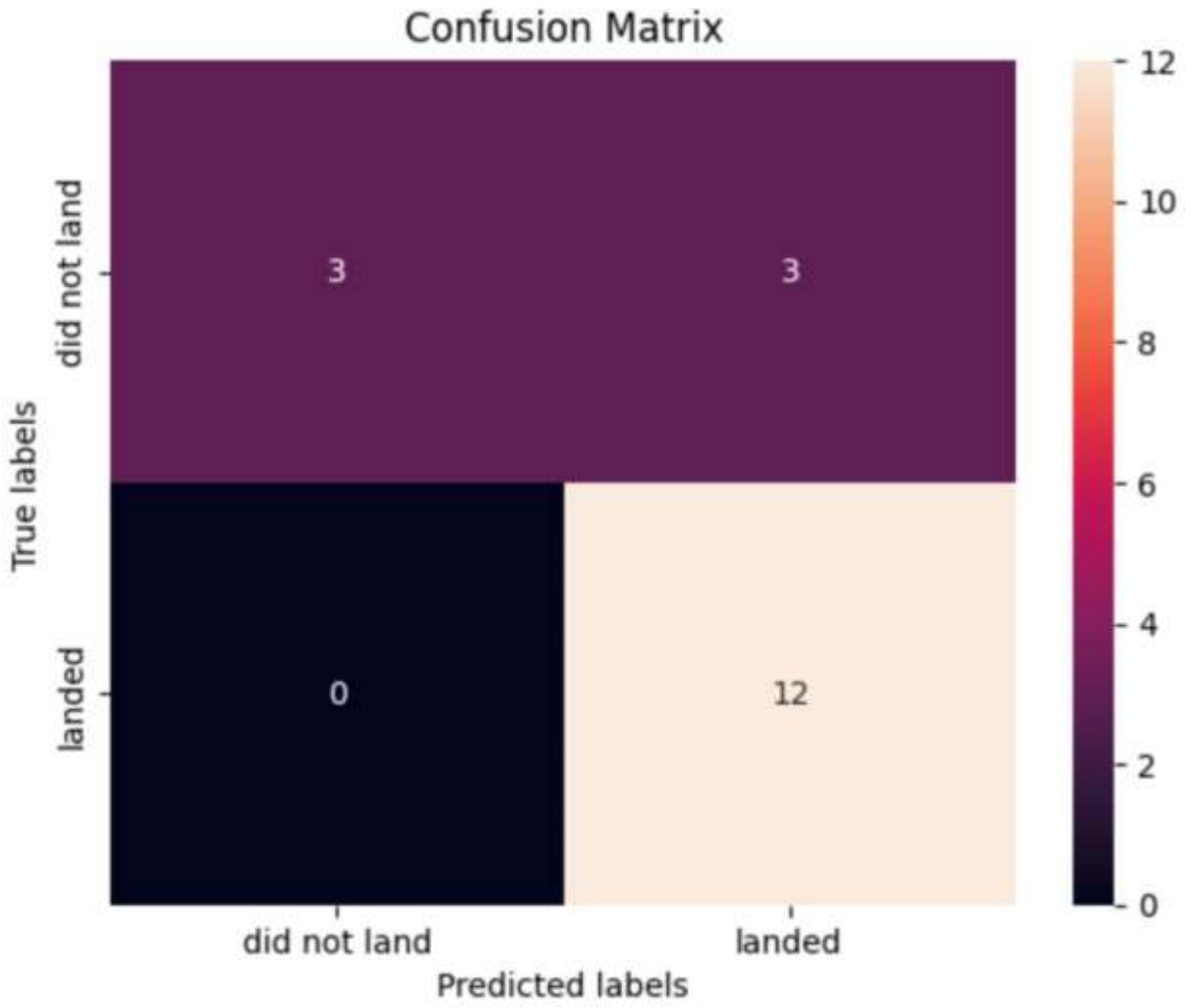
The model with the highest accuracy is Decision Tree with 0.8333 accuracy



Confusion Matrix

Interpretation of the confusion matrix

1. The model is **highly accurate** in identifying landings (12/12 correct).
2. However, it made **3 false positive errors**, predicting landings that did not occur.
3. The **absence of false negatives (0 FN)** indicates that **every actual landing was correctly classified**, making this model highly sensitive to detecting successful landings.



Conclusions

Successful Prediction of Falcon 9 Landings:

The project effectively developed a predictive model to determine whether the Falcon 9 first stage will successfully land, which is crucial for cost estimation and mission planning.

Significance of Reusability in Cost Reduction:

SpaceX's ability to reuse the Falcon 9 first stage significantly reduces launch costs compared to competitors, making reusability a key factor in pricing rocket launches.

Insights from Predictive Analysis:

The Decision Tree model achieved the highest accuracy (0.8333), correctly identifying all successful landings while showing some false positives, indicating room for optimization.

Conclusions

Strategic Implications for Competitors:

The findings provide valuable insights for alternate companies aiming to compete with SpaceX, helping them assess launch costs and landing success probabilities when bidding against SpaceX.

Future Improvements and Applications:

Further refinements in predictive modeling, including more advanced machine learning techniques, could enhance accuracy.

The methodology can be extended to optimize rocket design, improve landing success rates, and support broader space industry advancements.

Appendix

1. GitHub Repositories & Notebooks

Below are the GitHub links to all notebooks used in this project, along with their descriptions:

- **Data preprocessing and Cleaning 1 , 2** – Data extraction, missing value handling, and feature engineering.
- **Exploratory Data Analysis 1** , 2, Interactive Analytics 1 , 2 - Visualization of launch trends, success rates, and feature distributions.
- Machine Learning Models Model training, evaluation, and selection of the best predictive model.
- **Predictive Analysis & Performance Metrics** – Confusion matrix interpretation, accuracy evaluation, and key insights.

2. Data Source & Description

- **Dataset 1** , 2 , 3 , 4 SpaceX Falcon 9 launch data.
- **Features:** Launch site, payload mass, booster version, orbit, Landing Pad, Flight Information and landing outcome.
- **Preprocessing:** Missing value handling, categorical encoding, and feature scaling.

Thank you!