IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

Michael Piciucco
2025-08-21

# Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

# Executive Summary

- **Project Overview**
  - This capstone initiative focuses on predicting the successful landing of the SpaceX Falcon 9 first stage by leveraging advanced machine learning classification algorithms
- **Execution Approach**
  - Data Acquisition & Preparation: Collected, cleansed, and standardized launch datasets
  - Exploratory Analysis: Identified patterns and trends influencing landing outcomes
  - Interactive Visualization: Built dashboards to illustrate key performance drivers
  - Predictive Modeling: Applied multiple algorithms to benchmark predictive accuracy
- **Key Findings**
  - Analysis confirmed measurable correlations between specific launch characteristics and mission success rates
  - Among the tested models, the decision tree algorithm emerged as the most effective in forecasting first-stage landing success, providing a strong foundation for operational decision-making and risk reduction

# Introduction

- **Project Objective**
  - This capstone project predicts whether the Falcon 9 first stage will successfully land, a critical factor in understanding SpaceX's competitive cost advantage.
- **Strategic Context**
  - SpaceX lists Falcon 9 launches at $62 million, compared to over $165 million from other providers.
  - The key driver of this cost gap is first-stage reusability. Accurately forecasting landing success enables more precise launch cost estimations, providing valuable intelligence for alternate providers considering competitive bids.
- **Operational Insights**
  - Not all unsuccessful landings are failures; in some cases, SpaceX intentionally performs controlled ocean landings for operational or mission-specific reasons.
- **Core Question**
  - Given a specific set of launch characteristics—such as payload mass, orbit type, launch site, and other mission parameters—can we reliably predict whether the Falcon 9 first stage will achieve a successful landing?

Section 1

# Methodology

# Methodology

Executive Summary

- Applied a structured, end-to-end analytics process to ensure data integrity, extract actionable insights, and achieve high predictive accuracy.

- Perform data wrangling
  - Sources: SpaceX API and targeted web scraping.
  - Process: Personally collected, cleaned, and formatted all datasets for analysis readiness.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Tools: Pandas, NumPy, and SQL.
  - Focus: Independently identified patterns, correlations, and anomalies in the launch data.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Static & Statistical Charts: Created with Matplotlib and Seaborn.
  - Geospatial Mapping: Developed using Folium.
  - Interactive Dashboards: Built in Dash for dynamic, real-time exploration.
- Perform predictive analysis using classification models
  - Algorithms Tested: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).
  - Goal: Personally benchmarked each model to determine the most accurate predictor of first-stage landing success.

# Data Collection

- **Step 1: SpaceX API Request**

Initiate API Request → Fetch Launch Data → Locally Store Data

- **Step 2: Web Scraping from Wikipedia**

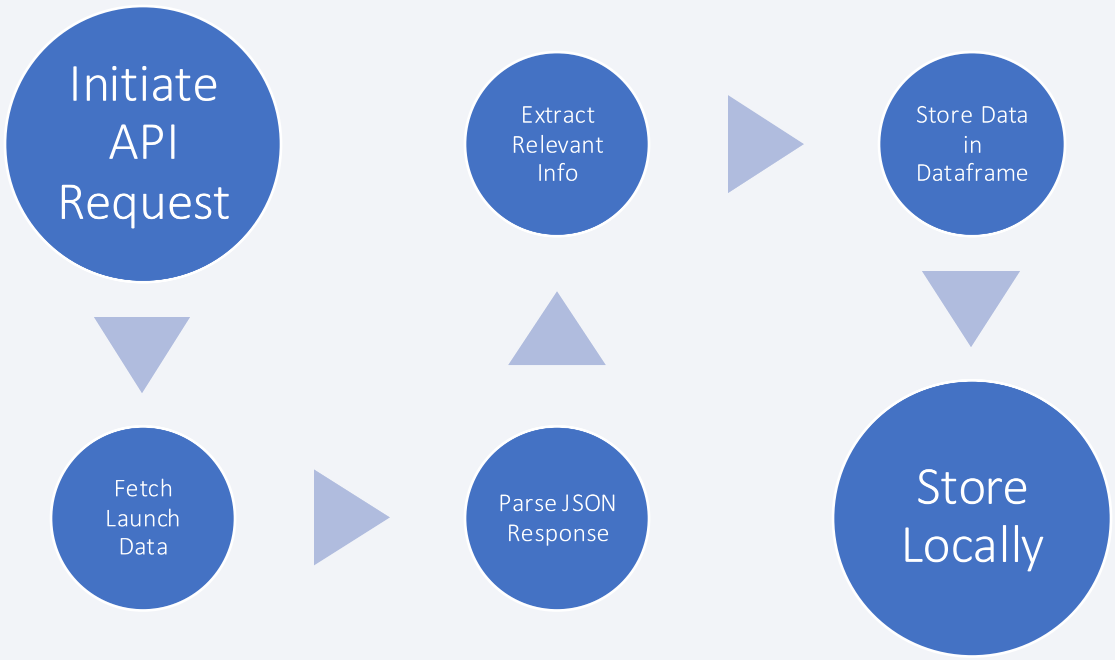
Extract Table → Parse using BeautifulSoup → Convert to Data Frame

- **Step 3: Data Integration**

Integrate API & Wiki Data → Merge DataSets → Analyze Final Data

# Data Collection: SpaceX API

- **SpaceX API Data Source**
  - I utilized the SpaceX API to obtain detailed information on various SpaceX rocket launches. From the full dataset, I filtered the results to focus exclusively on Falcon 9 launches, aligning the data with the project's prediction objective.
- **Data Preparation**
  - To address data completeness, I replaced all missing values with the mean of their respective columns. This ensured consistency and minimized bias in analysis.
- **Final Dataset**
  - After filtering and preprocessing, the dataset contained 90 records and 17 features.

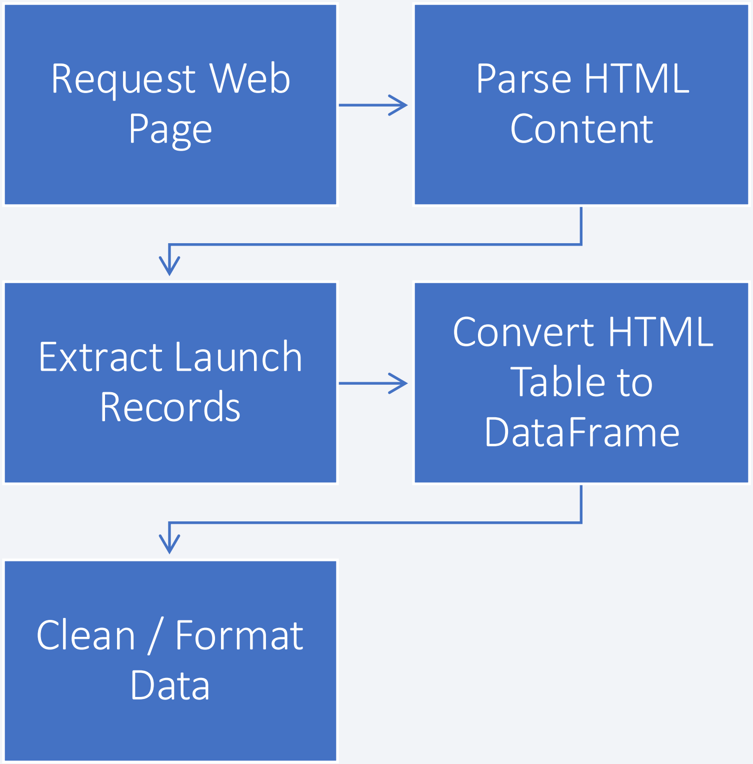



GitHub - SpaceX Data Collection

# Data Collection: Scraping

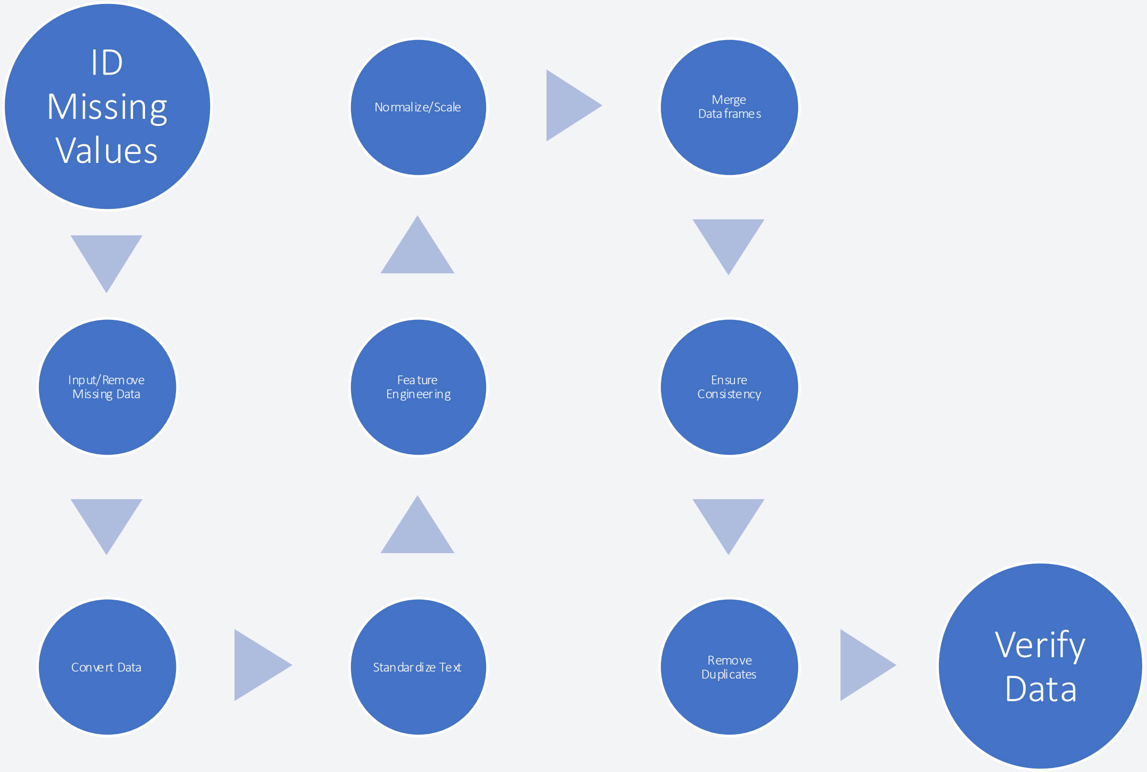
- **Step 1: Data Retrieval**
  - Use Python's requests library to access the HTML content from the target Wikipedia page:
- **Step 2: HTML Parsing**
  - Apply BeautifulSoup to parse the HTML and isolate the table containing Falcon 9 launch records.
- **Step 3: DataFrame Conversion & Cleaning**
  - Convert the extracted HTML table into a Pandas DataFrame
  - Clean and format the data to ensure accuracy, consistency, and readiness for analysis.

GitHub - Data Collection - Web Scraping



# Data Wrangling

- Overview: Data Wrangling
  - Data wrangling is the process of cleaning, transforming, and organizing raw data into a structured format suitable for analysis. These steps ensure the launch data was accurate, consistent, and ready for further exploration.




**Data Cleaning**
- Pandas.isnull
- Fillna | Dropna

**Data Transformation**
- Pd.to_Datetime | Astype
- Str.lower | Strip
- Create New Columns | Min / Max Scale

**Data Integration**
- Merge API and web scraping datasets into one to ensure consistent column names and formats.

**Data Validation**
- Remove duplicates and verified data accuracy and consistency.
- Pd.duplicated

GitHub - SpaceX - Data Wrangling

# EDA with Data Visualization

- **Exploratory Data Analysis (EDA) Overview**
  - Visually explored and summarized dataset characteristics to understand distributions, patterns, and relationships.
- **Charts Plotted**
  - **Histograms**: Visualized distributions of variables (ex: payload mass, flight number) to assess spread, central tendency, and outliers.
  - **Bar Charts**: Compared categorical outcomes (ex: success/failure) across launch sites and rocket types.
  - **Line Charts**: Tracked success rate trends over time to reveal performance changes.
  - **Scatter Plots**: Explored relationships between numerical variables (ex: payload mass vs. success).
  - **Heatmaps**: Displayed correlations between numerical variables to guide feature selection.
  - **Box Plots**: Highlighted data spread, skewness, and outliers across categories.

[GitHub - SpaceX - EDA Visualization]

# EDA with SQL

- **SQL Analysis Summary**

  - **Aggregate Queries**: Calculated total launches, counted successes and failures, and measured success rates by launch site and rocket type.
  - **Join Queries**: Linked launch records with rocket details and merged datasets for a complete analysis view.
  - **Filtering Queries**: Isolated specific outcomes (success/failure) and applied criteria like launch date or rocket configuration.
  - **Sorting Queries**: Ranked launches by date or success rate to identify trends and outliers.
  - **Subqueries**: Computed metrics such as average payload mass per site and enabled deeper analysis within larger datasets.

[GitHub - EDA with SQL]

# Build an Interactive Map with Folium

**Map Objects Created**

- **Markers**
  - What I Created:
    - Plotted individual markers at each SpaceX launch site.
    - Each marker represents a precise geographic location.
  - Why I Created It:
    - To give a clear, pinpointed view of all Falcon 9 launch locations.
    - Enables quick spatial reference for where launches have historically occurred.
- **Circles**
  - What I Created:
    - Added circular overlays around each launch site.
    - Radius represents the surrounding operational or impact zone.
  - Why I Created It:
    - To visualize safety perimeters and proximity zones around launch sites.
    - Helps assess potential geographic influences on launch operations.
- **Lines**
  - What I Created:
    - Drew connecting lines between launch sites and relevant points of interest.
    - Provided spatial linkages for visual context.
  - Why I Created It:
    - To highlight relationships and dependencies between locations.
    - Supports analysis of logistical connections in launch planning.

[GitHub - Launch Site Location]

# Build a Dashboard with Plotly Dash

**Success Pie Chart**

- What I Created:
  - A pie chart showing the proportion of successful vs. failed launches.
- Why I Created It:
  - To provide a quick visual snapshot of overall mission performance and success rates.

**Success–Payload Scatter Plot**

- What I Created:
  - A scatter plot mapping payload mass against launch success outcomes.
- Why I Created It:
  - To explore how payload mass may influence the probability of a successful mission.

**Launch Site Dropdown (Dashboard Feature)**

- What I Created:
  - An interactive dropdown menu to select specific launch sites.
- Why I Created It:
  - To enable focused, site-specific performance analysis and comparisons.

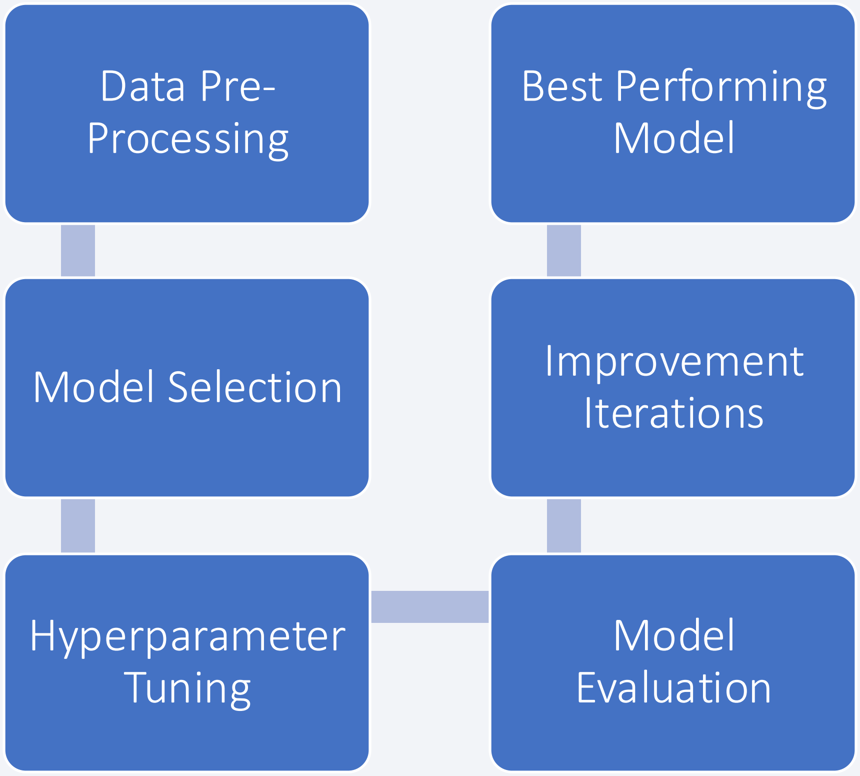
**Payload Range Slider (Dashboard Feature)**

- What I Created:
  - An interactive slider to filter launches by payload mass range.
- Why I Created It:
  - To allow dynamic analysis of success rates across different payload capacities.

GitHub - DashApp

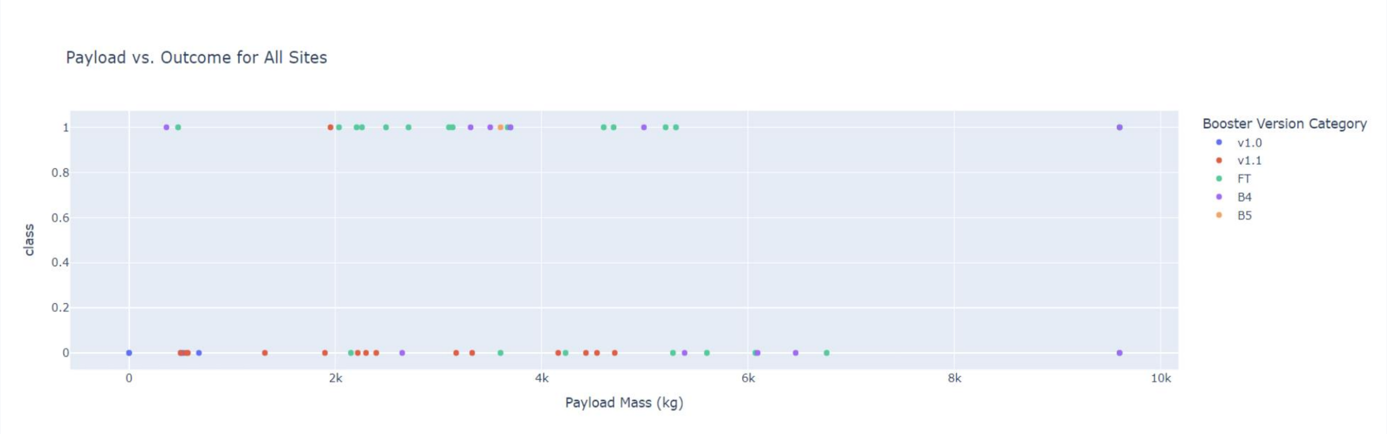
# Predictive Analysis (Classification)

- Prepared the Data:
  - Standardized all features and split into training/test sets to ensure unbiased model validation.
- Tested Multiple Algorithms:
  - Evaluated SVM, Decision Tree, and KNN to find the best fit for binary classification.
- Optimized Performance:
  - Applied GridSearchCV to fine-tune key hyperparameters for each algorithm.
- Measured Robustness:
  - Used cross-validation with accuracy, precision, recall, and F1-score to assess reliability.
- Selected Best Model:
  - Chose the highest-performing, most generalizable model for predicting Falcon 9 first-stage landings.

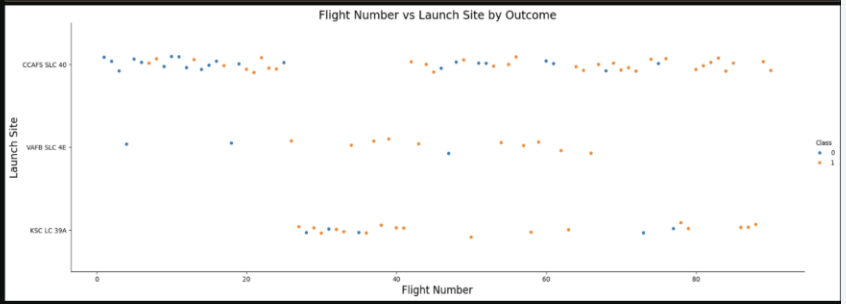



[GitHub - SpaceX - Machine Learning Prediction]

# Results




[10]:
```python
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(x="FlightNumber", y="LaunchSite", hue="Class", data=df, aspect=3, height=6, kind="strip")
plt.xlabel("Flight Number", fontsize=16)
plt.ylabel("Launch Site", fontsize=16)
plt.title("Flight Number vs Launch Site by Outcome", fontsize=18)
plt.show()
```



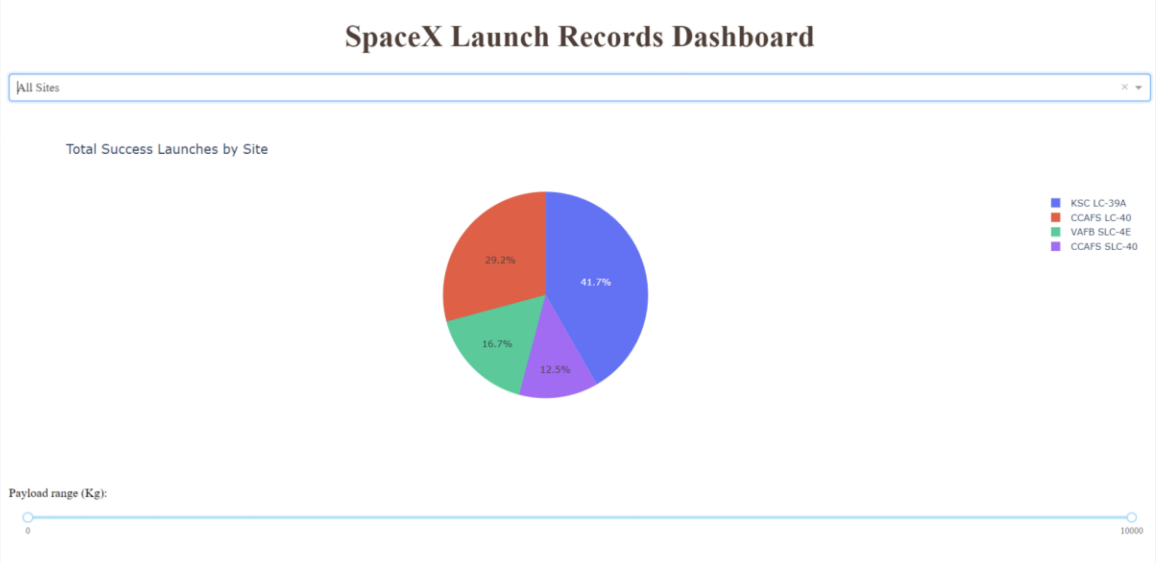

## TASK 12

Find the method performs best:

Based on the results I got:

1. **Logistic Regression**: Accuracy on test data: 0.8333
2. **Support Vector Machine (SVM)**: Accuracy on test data: 0.8333
3. **Decision Tree**: Accuracy on test data: 0.9444
4. **K Nearest Neighbors (KNN)**: Accuracy on test data: 0.8333

It seems there was a significant increase in accuracy for the Decision Tree model, which shows the highest accuracy of 0.9444 on the test set. Therefore, in this comparison, the Decision Tree model performs the best among the classifiers you've trained.

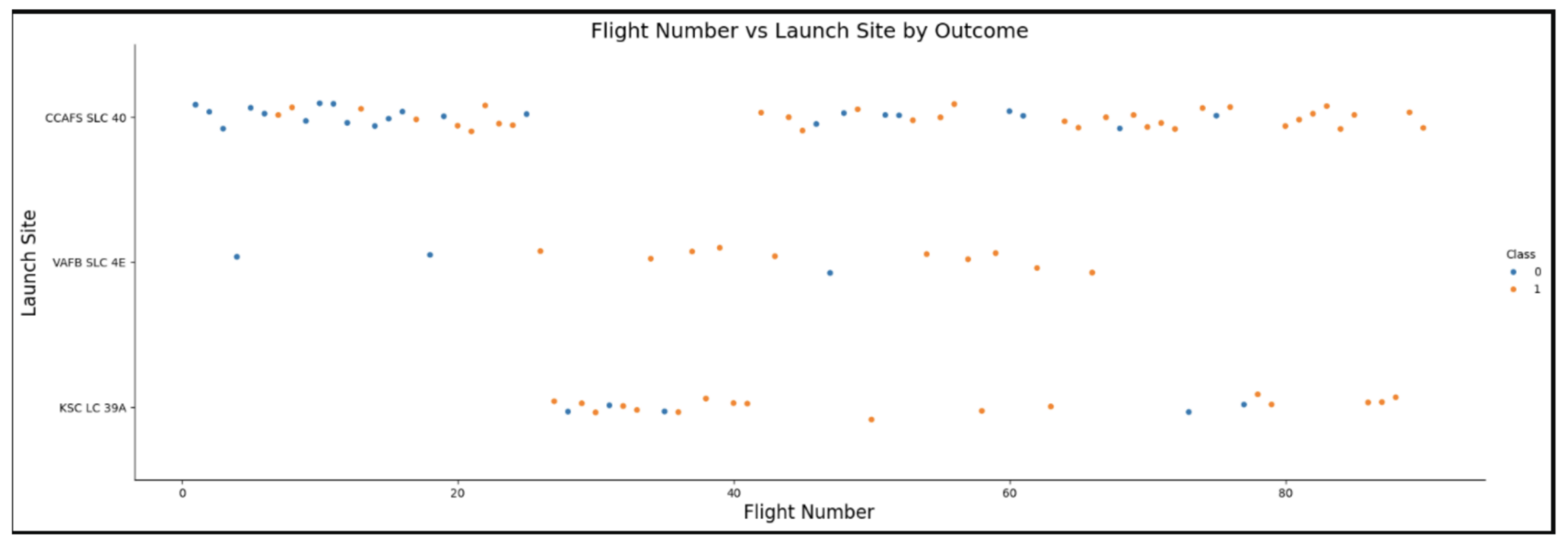


Section 2

# Insights drawn from EDA

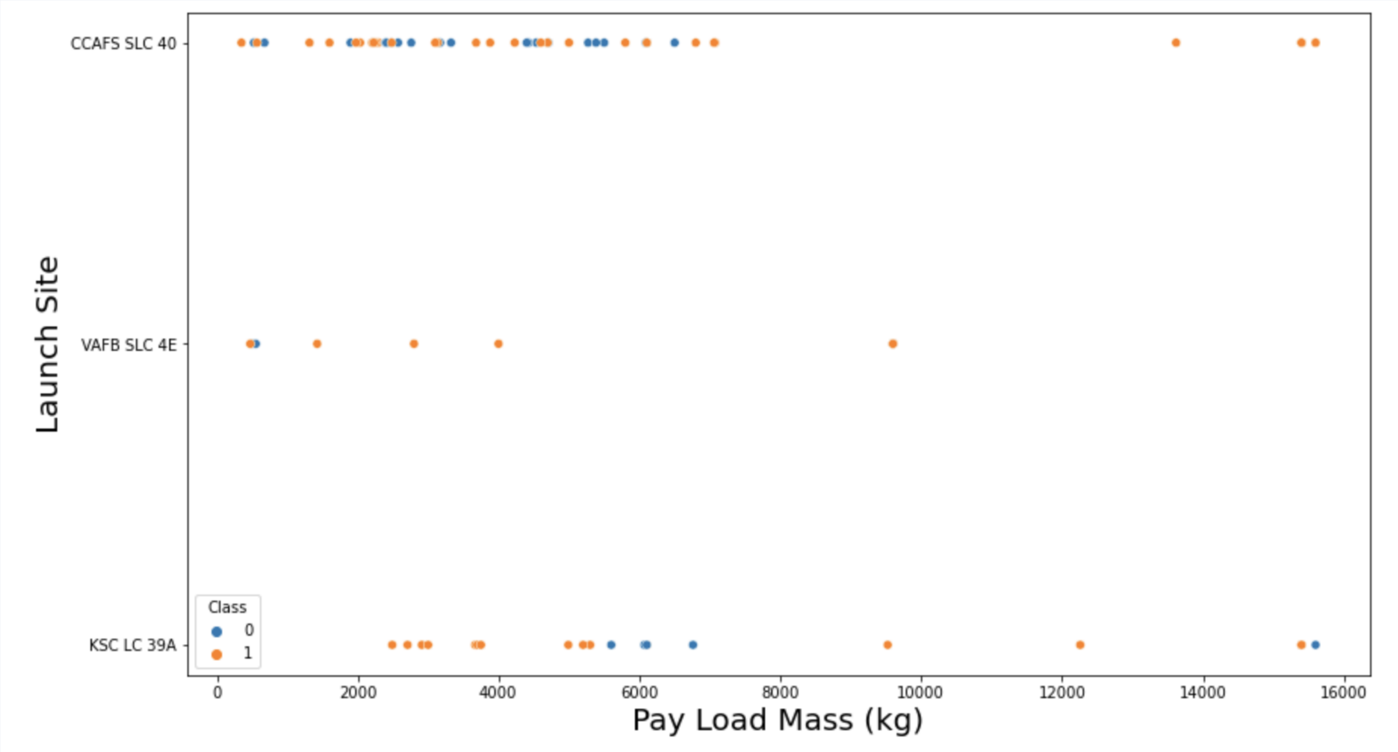
# Flight Number vs. Launch Site

- **Mixed Results by Launch Site**: At both CCAFS SLC 40 and KSC LC 39A, landing outcomes vary between successes and failures, indicating that site location alone is not the primary driver of success.
- **Steady Launch Activity**: Launches occur across a broad range of flight numbers at all sites, showing consistent operational cadence without a clear trend in improving or declining success rates over time.



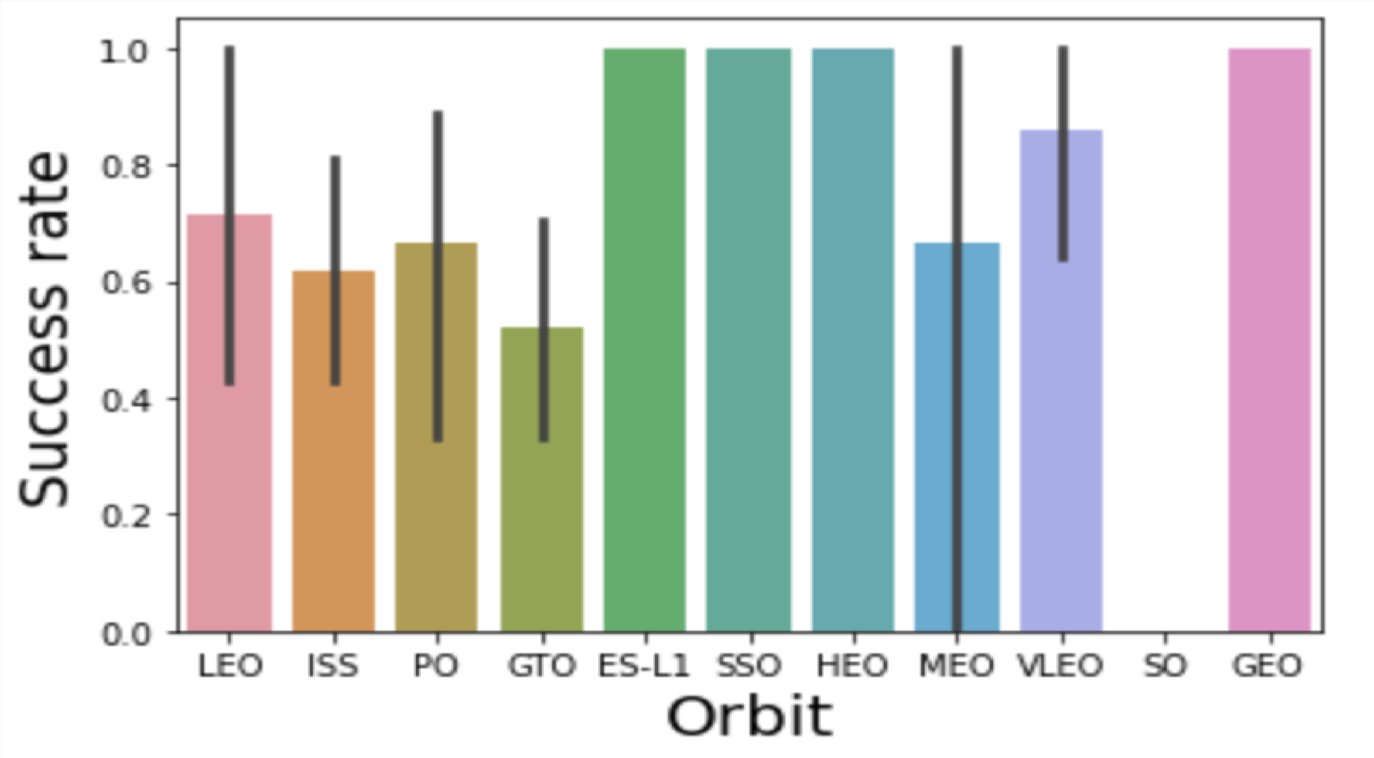
# Payload vs. Launch Site

- **Payload Range by Site**: CCAFS SLC 40 primarily supports payloads under 10,000 kg, while VAFB SLC 4E and KSC LC 39A handle a broader range, reflecting more varied mission types.
- **High-Capacity Operations**: KSCLC 39A frequently launches payloads exceeding 15,000 kg, highlighting its capability for heavy-lift missions.



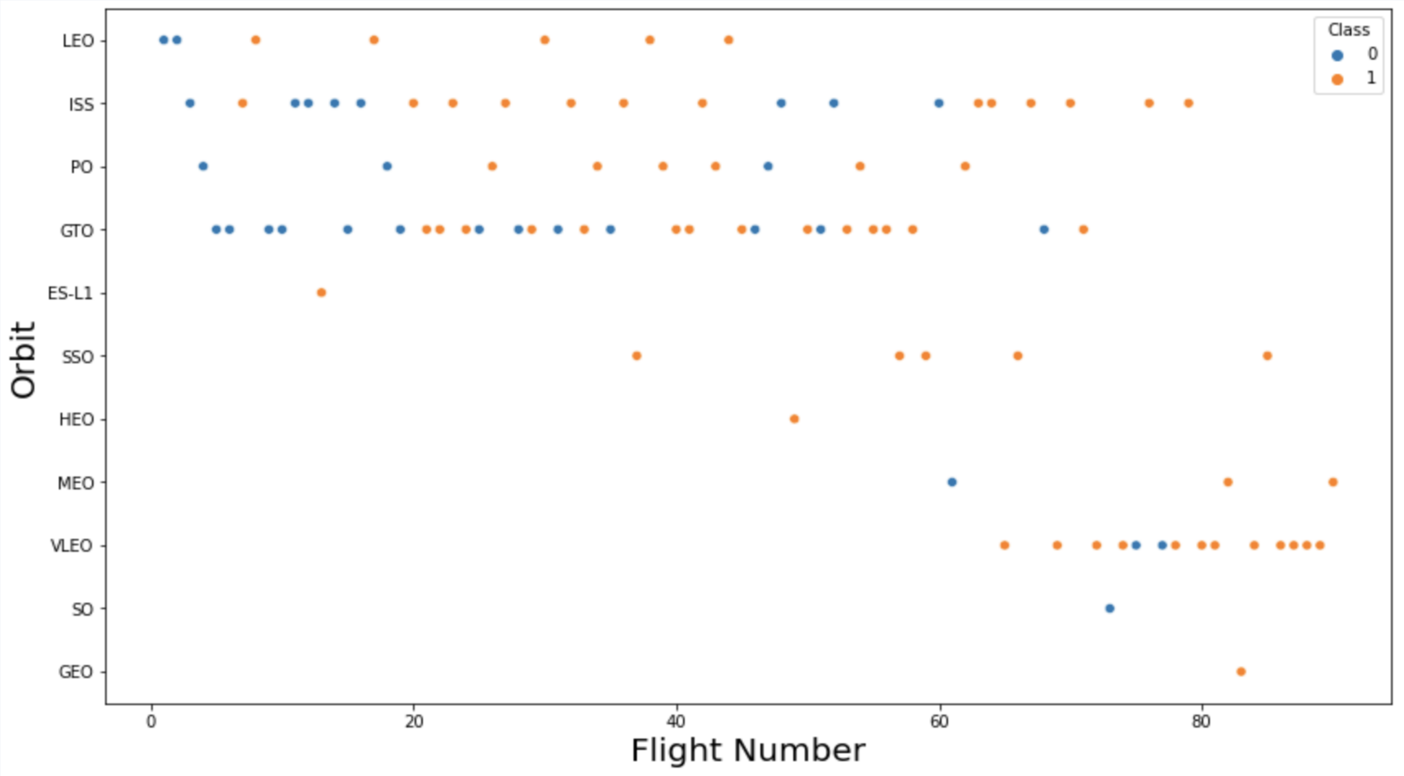
# Success Rate vs. Orbit Type

- **High-Reliability Orbits**: Missions to VLEO, ES-L1, GEO, HEO, and SSO orbits have achieved a 100% landing success rate, indicating strong reliability for these mission profiles.
- **Challenging GTO Missions**: GTO launches show notably lower success rates, suggesting increased operational complexity or technical challenges for this orbit type.



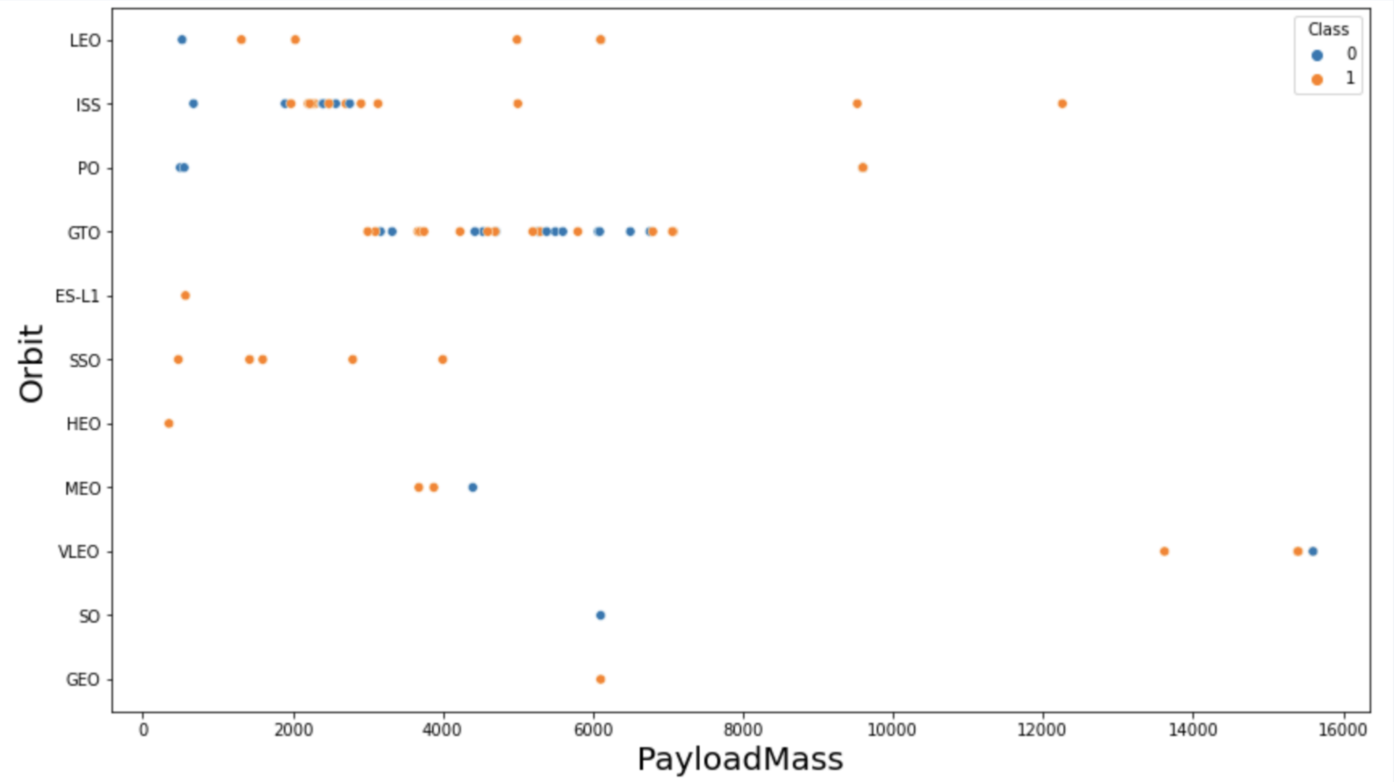
# Flight Number vs. Orbit Type

- **Increased Success Over Time**: Falcon 9 launch success rates improve noticeably with higher flight numbers, indicating that operational experience and iterative enhancements lead to better outcomes.
- **Orbit-Specific Performance**: Early missions to GTO and ISS orbits had mixed results, but recent launches to these orbits show higher success rates, reflecting advancements in mission planning and execution.



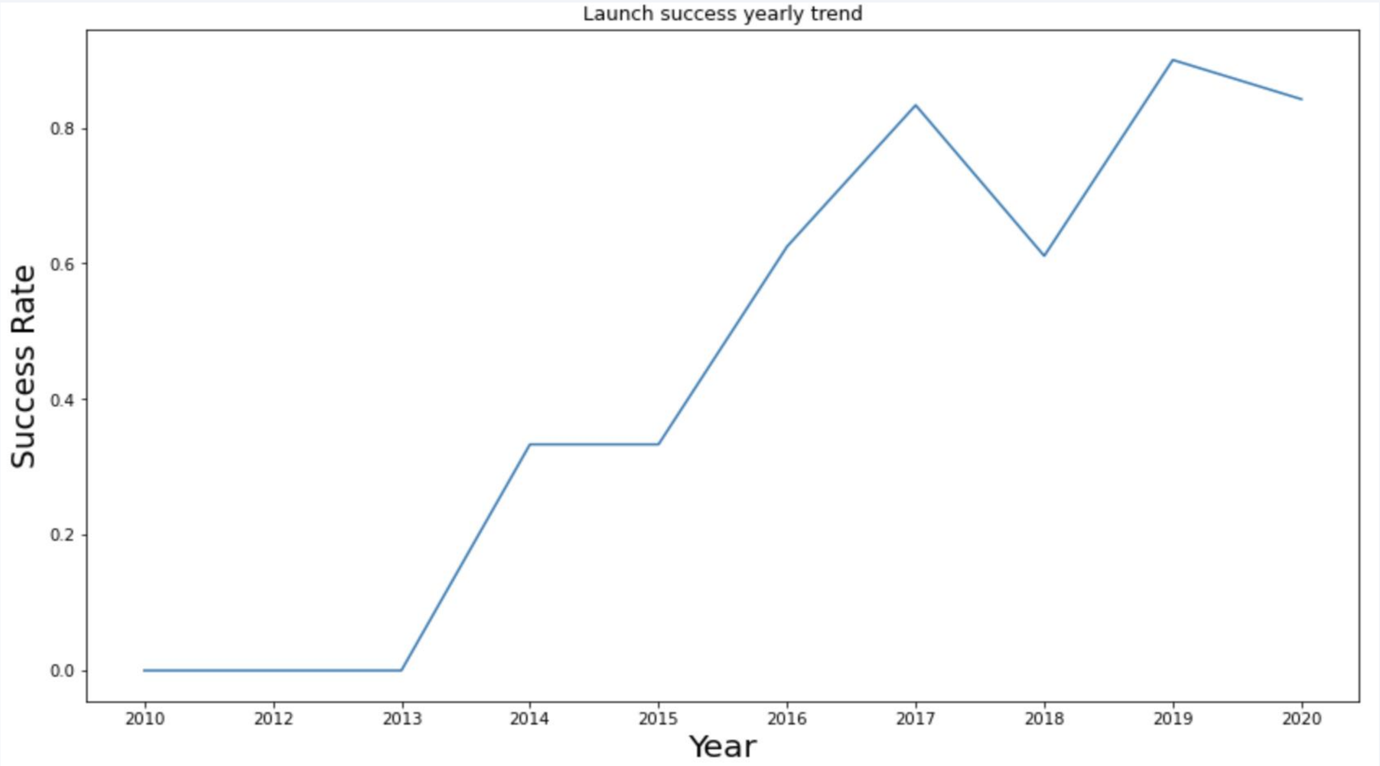
# Payload vs. Orbit Type

- **Higher Success with Lighter Payloads**: Successful landings are more common across all orbit types when payloads are under 6,000 kg.
- **Mixed Outcomes for Heavy Payloads**: Payloads above 10,000 kg show both successes and failures, suggesting greater difficulty with heavier missions.



# Launch Success Yearly Trend

- **Rising Annual Success Rates**: Launch success rates have improved significantly since 2013, surpassing 80% by 2020.
- **Overall Positive Trend**: Despite a dip in 2018, the long-term trajectory shows increasing reliability and success in Falcon 9 launches.



# All Launch Site Names

- **Goal**: List only the unique launch sites from the dataset.
- **Method**: Used SELECT DISTINCT to remove duplicate site names.
- **Result**: Four unique sites displayed: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, & CCAFS SLC-40.

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

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

# Launch Site Names Begin with 'CCA'

- **Goal**: Show the first 5 records where launch sites start with "CCA."
- **Method**: Used LIKE 'CCA%' to filter sites beginning with "CCA" and LIMIT 5 to restrict output.
- **Result**: Returned 5 early Falcon 9 missions from CCAFS LC-40, including demo and CRS launches.

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

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

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing_Ou |
|---|---|---|---|---|---|---|---|---|---|
| 2010-06-04 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (par |
| 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 (par |
| 2012-05-22 | 7:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No a |
| 2012-10-08 | 0:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No a |
| 2013-03-01 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No a |

# Total Payload Mass

- **Goal**: Calculate the total payload mass for missions where the customer was NASA (CRS).
- **Method**: Used SUM(PAYLOAD_MASS__KG_) with a filter WHERE CUSTOMER = 'NASA (CRS)'.
- **Result**: The total payload mass carried by boosters for NASA (CRS) missions is 45,596 kg.

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

| Total payload mass by NASA (CRS) |
| ---: |
| 45596 |

# Average Payload Mass by F9 v1.1

- **Goal**: Find the average payload mass for missions using booster version F9 v1.1.
- **Method**: Applied AVG(PAYLOAD_MASS__KG_) with a filter WHERE BOOSTER_VERSION = 'F9 v1.1'.
- **Result**: The average payload mass carried by F9 v1.1 boosters is 2,928.4 kg.

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

| Average payload mass by Booster Version F9 v1.1 |
| ---: |
| 2928.4 |

# First Successful Ground Landing Date

- **Goal**: Identify the first successful landing on a ground pad.
- **Method**: Used MIN(Date) with a filter WHERE Landing_Outcome = 'Success (ground pad)'.
- **Result**: The first successful ground pad landing occurred on 2015-12-22.

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

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

| MIN("Date") |
|---|
| 2015-12-22 |

# Successful Drone Ship Landing with Payload between 4000 and 6000

- **Goal**: Find boosters that successfully landed on a drone ship with payload mass between 4,000 and 6,000 kg.
- **Method**: Filtered records using WHERE Landing_Outcome = 'Success (drone ship)' and payload mass constraints, then applied DISTINCT Booster_Version.
- **Result**: Boosters meeting the criteria are F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2.

```
%sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
```

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

# Total Number of Successful and Failure Mission Outcomes

- **Goal**: Count the number of missions with outcomes labeled as Success or Failure.
- **Method**: Used COUNT(*) with a WHERE Mission_Outcome IN ('Success','Failure') and grouped by outcome.
- **Result**: The query returned 98 successful missions. (No records labeled as “Failure” appear in the dataset.)

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

| Mission_Outcome | Total |
| --- | --- |
| Success | 98 |

# Boosters Carried Maximum Payload

- **Goal**: Identify boosters that carried the heaviest payloads recorded in the dataset.
- **Method**: Used a subquery with MAX(PAYLOAD_MASS__KG_) to find the maximum payload mass, then matched all boosters that achieved it.
- **Result**: Multiple boosters qualified, as noted in the images below

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

| 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

- **Goal**: Display records from 2015 showing the month names, landing failures on drone ships, plus booster versions and launch sites.
- **Method**: Used substr(Date,0,5)='2015' to filter by year, and substr(Date,6,2) mapped to month names. Applied conditions to focus on Landing_Outcome = 'Failure (drone ship)'.
- **Result**: The query ran successfully for 2015, but the output you shared currently shows general mission outcomes (success/failure in flight) instead of filtering specifically for drone ship failures.

| 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

- **Goal**: Rank all landing outcomes between 2010-06-04 and 2017-03-20 by frequency.
- **Method**: Filtered by date range, grouped by Landing_Outcome, counted occurrences, and sorted in descending order.
- **Result**:

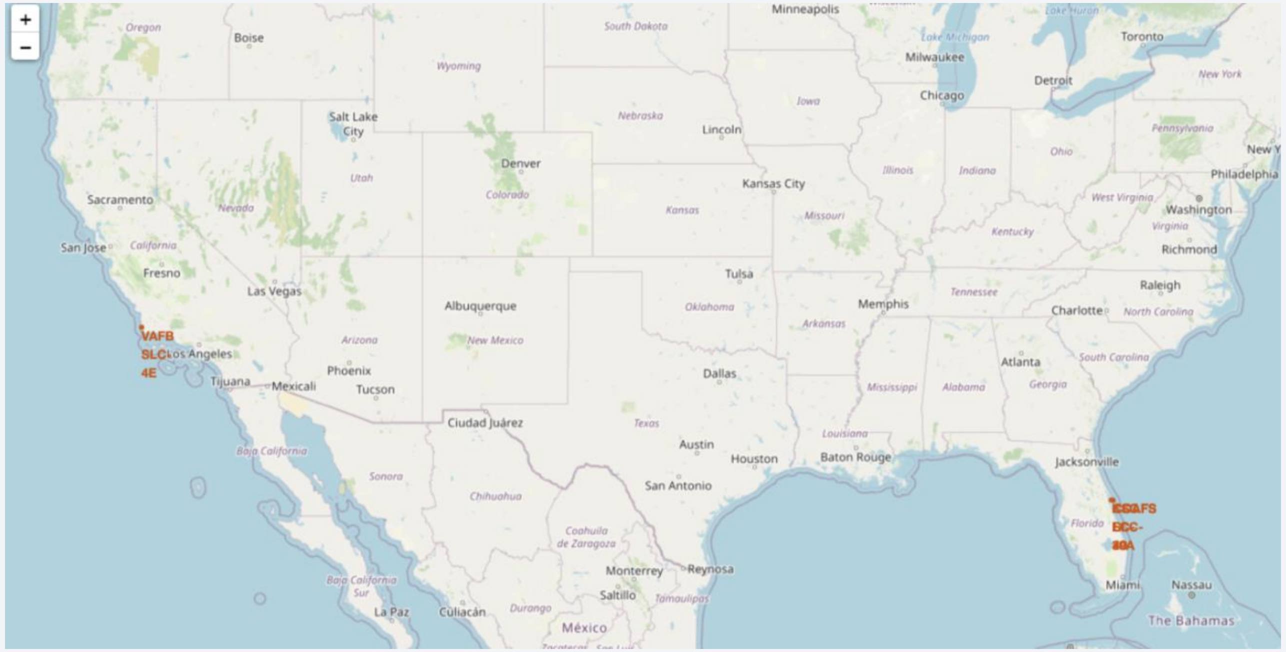
| Landing_Outcome | Count |
|---|---|
| No attempt | 10 |
| Success (drone ship) | 5 |
| Failure (drone ship) | 5 |
| Success (ground pad) | 3 |
| Controlled (ocean) | 3 |
| Uncontrolled (ocean) | 2 |
| Failure (parachute) | 2 |
| Precluded (drone ship) | 1 |

Section 3

# Launch Sites Proximities Analysis

# Launch Site Analysis Pt.1: Locations

- 1. **Proximity to the Equator**
  - **Analysis**: Not all launch sites are near the Equator.
  - **Detail**: Vandenberg Air Force Base (VAFB SLC-4E) is located at latitude 34.63, farther from the Equator than the Florida sites.
- 2. **Proximity to the Coast**
  - **Analysis**: All launch sites are close to the coast.
  - **Detail**: CCAFS LC-40, CCAFS SLC-40, and KSC LC-39A are on the Florida coast, while VAFB SLC-4E is on the California coast.



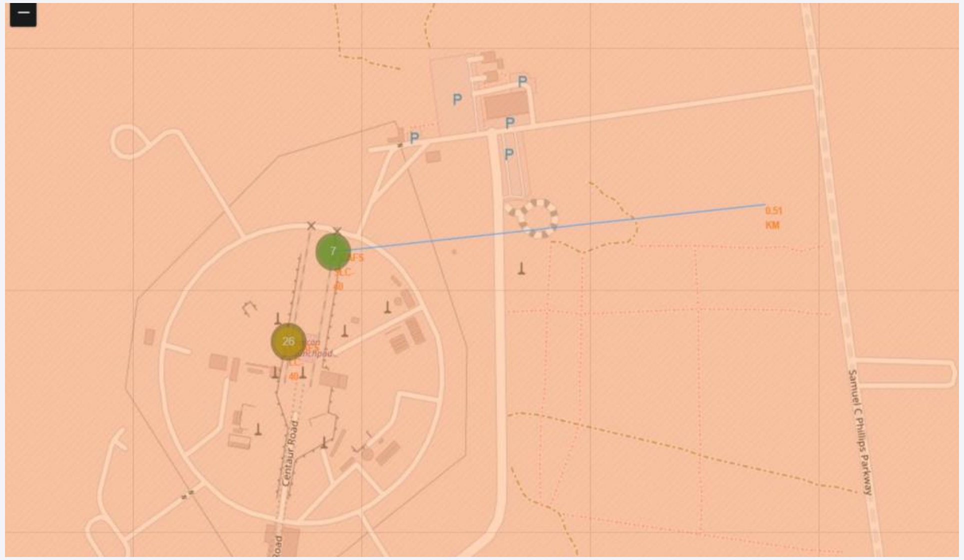

| | Launch Site | Lat | Long |
|---|---|---|---|
| 0 | CCAFS LC-40 | 28.562302 | -80.577356 |
| 1 | CCAFS SLC-40 | 28.563197 | -80.576820 |
| 2 | KSC LC-39A | 28.573255 | -80.646895 |
| 3 | VAFB SLC-4E | 34.632834 | -120.610745 |

# Launch Site Analysis Pt. 2: Success / Fail

- **Enhanced Visualization with Clustered Markers**
  - **Analysis**: Clustered markers improve exploration of SpaceX launch data by simplifying dense plots and highlighting patterns that may otherwise remain hidden.
  - **Detail**: Marker colors and popup details provide quick insights into launch characteristics and outcomes.
- **Example**
  - **Observation**: At CCAFS LC-40, 26 launches are represented: 19 red markers and 7 green markers.
  - **Interpretation**: Red markers indicate unsuccessful launches, while green markers represent successful ones, offering immediate visual feedback on performance at the site.

# Launch Site Analysis Pt. 3: Proximities

- **Visualization of Launch Site Proximity**
  - **Analysis**: The plot illustrates the distance between the CCAFS SLC-40 launch site and the nearest coastline. A straight PolyLine shows this distance visually.
  - **Detail**: The calculated distance is approximately 0.51 kilometers, emphasizing the site's close proximity to the coast.
- **Context**
  - **Interpretation**: Coastal locations are desirable for launch sites, enabling over-water flight paths and safer recovery operations while minimizing risks to populated areas.



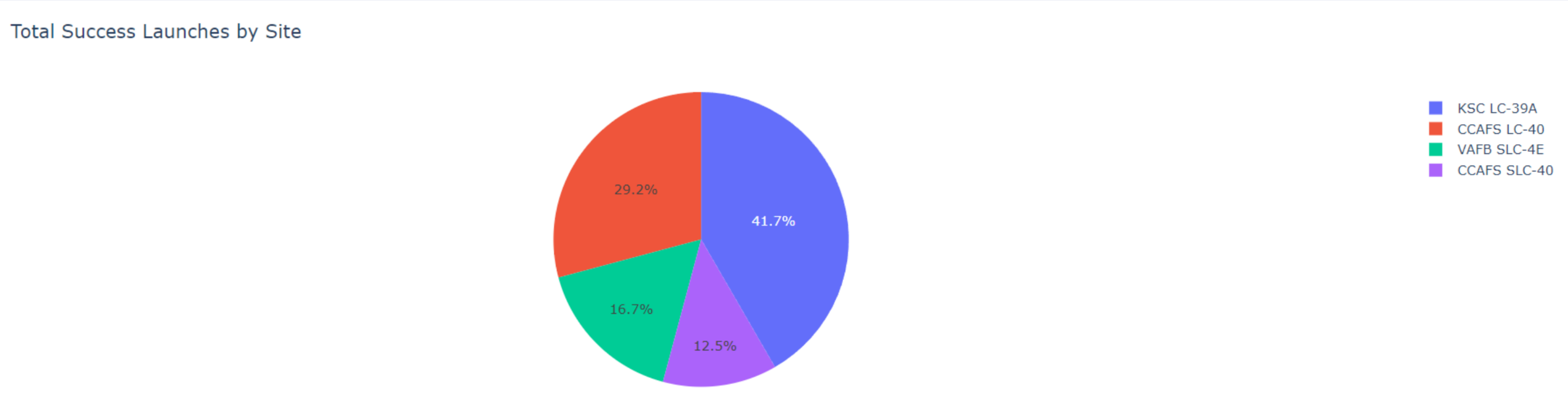

```python
coastline_lat = 28.56367
coastline_lon = -80.57163

# Example launch site coordinates (replace with actual launch site coordinates)
launch_site_lat = launch_sites_df.loc[launch_sites_df['Launch Site'] == 'CCAFS SLC-40', 'Lat'].values[0]
launch_site_lon = launch_sites_df.loc[launch_sites_df['Launch Site'] == 'CCAFS SLC-40', 'Long'].values[0]

# Calculate distance using the calculate_distance function
distance_coastline = calculate_distance(launch_site_lat, launch_site_lon, coastline_lat, coastline_lon)

print(f"Distance from launch site to closest coastline: {distance_coastline} km")
```

Section 4

# Build a Dashboard with Plotly Dash

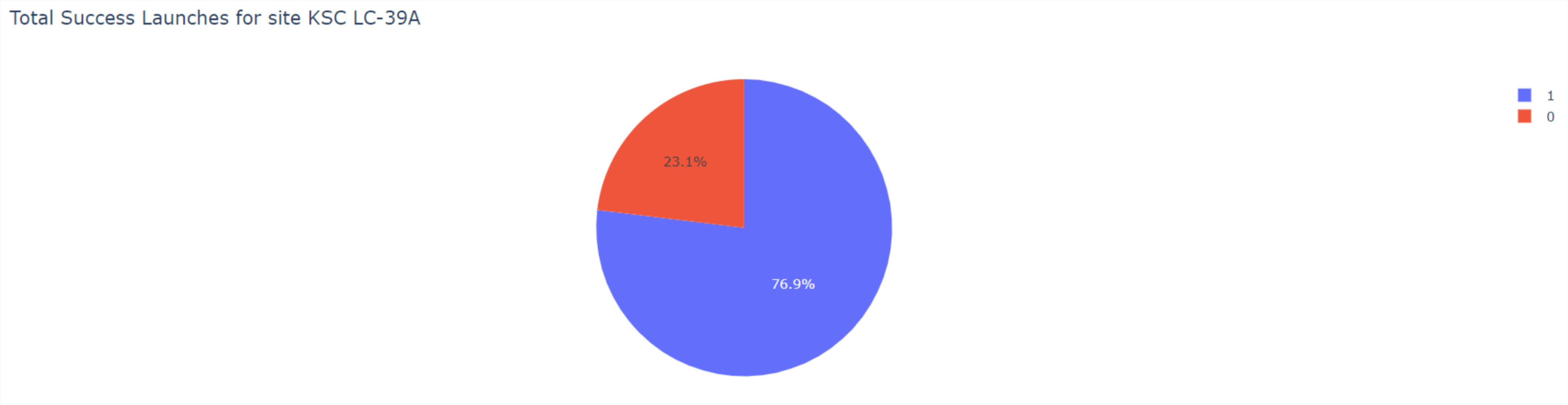
# Launch Success: All Sites

- **Key Findings**
  - **CCAFS LC-40**: 29.2%
  - **CCAFS SLC-40**: 12.5%
  - **VAFB SLC-4E**: 16.7%
  - **KSC LC-39A**: 41.7%
- **Analysis**
  - **Observation**: KSC LC-39A accounts for the highest proportion of successful launches at 41.7%.
  - **Interpretation**: This highlights KSC LC-39A as the most reliable launch site among the four.



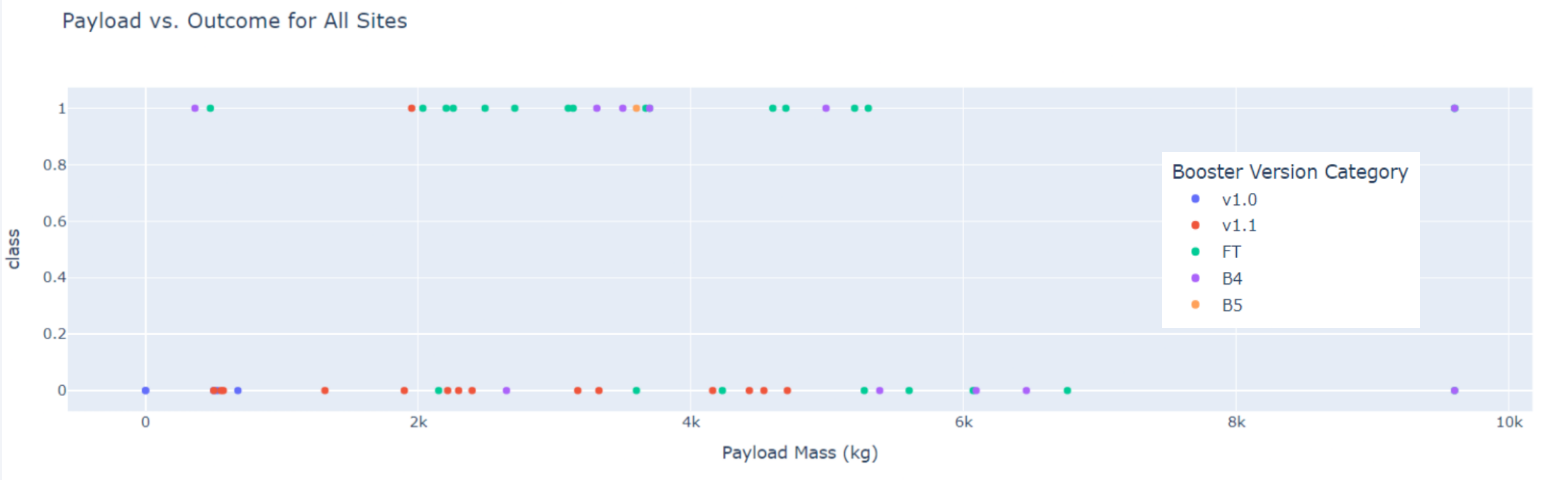
# Launch Site: Highest Success Ratio

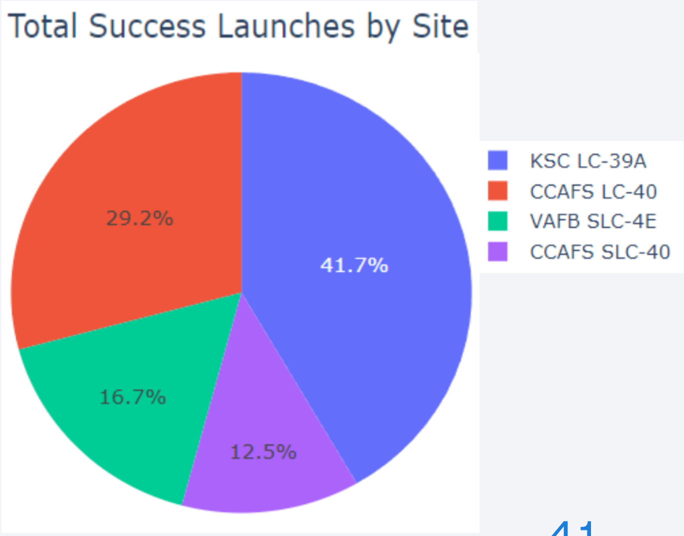
- **Analysis**: A significant share of successful launches from KSC LC-39A underscores its reliability and effectiveness as a launch site.
- **KSC LC-39A Performance**:
  - **Class 1** (Successful Launches): 76.9%
  - **Class 0** (Unsuccessful Launches): 23.1%
- **Interpretation**: The 76.9% success rate highlights KSC LC-39A as one of the most dependable sites for SpaceX operations.



# SpaceX Launch Data: Key Insights

- Launch Site Success Rates
  - CCAFS LC-40 stands out with the highest success rate at 43.7%, making it the most reliable launch site in the dataset.
  - Other sites—including KSC LC-39A, VAFB SLC-4E, and CCAFS SLC-40—show lower success rates, suggesting variability in reliability depending on the site.
- Booster Version Performance
  - Booster version “FT” is the most frequently used and demonstrates strong reliability across different payload masses.
  - Booster version “v1.0” appears less tested, with fewer launches, so its performance requires more data for meaningful conclusions.
  - Payload mass vs. success rate: No strong evidence indicates that heavier payloads reduce success rates, suggesting booster performance is consistent across different mission demands.



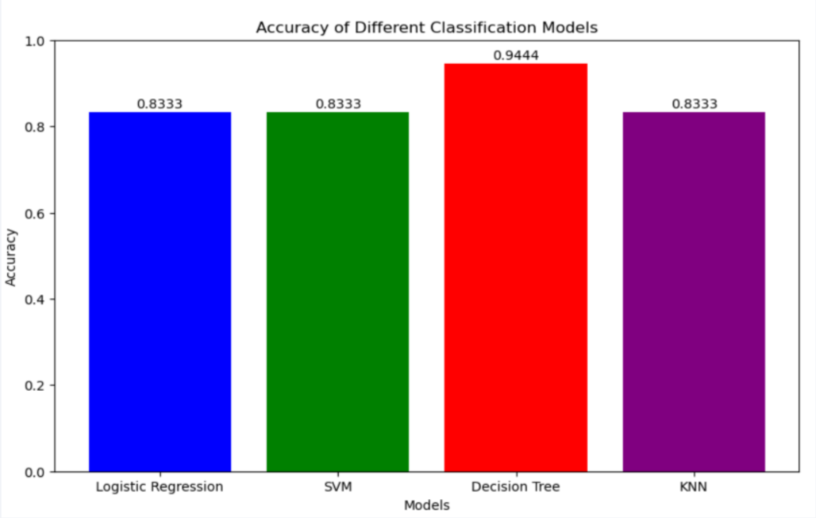



Section 5

# Predictive Analysis (Classification)

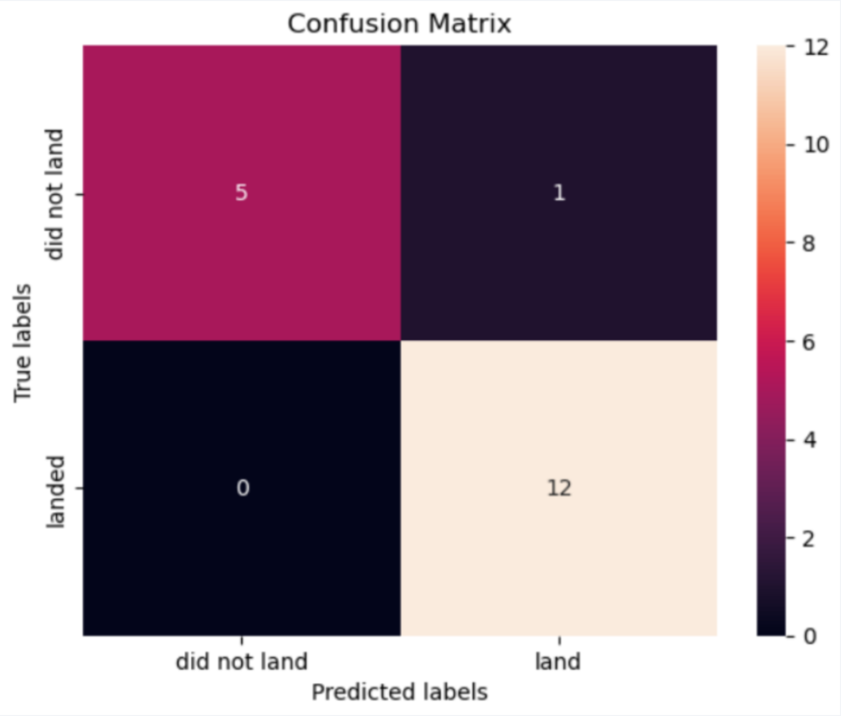
# Classification Accuracy

- **Decision Tree**
  - **Accuracy**: 0.9444
  - **Insight**: The highest performer among all tested models, indicating strong suitability for this dataset.
- **Logistic Regression, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN)**
  - **Accuracy**: 0.8333 each
  - **Insight**: While these models perform reasonably well, they fall short compared to the Decision Tree, suggesting they may not capture the dataset's patterns as effectively.
- **Conclusion**
  - The Decision Tree model outperforms the others with a **clear margin of accuracy.** This implies it is likely the best fit for this dataset, though further validation (cross-validation, checking overfitting, or tuning hyperparameters) would strengthen confidence in the result.



# Confusion Matrix

- **High Accuracy (94.44%)**
  - Demonstrates strong predictive power, correctly classifying the vast majority of test cases. This high accuracy reflects the model's effectiveness in forecasting Falcon 9 first-stage landings.
- **No False Negatives**
  - Every actual successful landing was correctly identified as such. In aerospace operations, this is especially critical because missing a successful landing (false negative) could undermine readiness and safety planning. The absence of false negatives greatly enhances the model's reliability.
- **Manageable False Positives**
  - The model produced 1 false positive, where it predicted a successful landing that did not occur. While not ideal, false positives are less problematic than false negatives in this context. Over-preparation is more acceptable than under-preparation, making this trade-off operationally sound.
- **Balanced Performance with Practical Bias**
  - The model shows a slight bias toward predicting successful landings. In aerospace, this aligns well with priorities: ensuring successful landings is crucial for cost estimation, resource allocation, and mission planning, so the bias is beneficial rather than harmful.



# Conclusions

1) **Launch Site Performance**
   - The analysis revealed that CCAFS LC-40 has the highest success rate, accounting for 43.7% of successful launches. This suggests that site-specific conditions or operational processes may contribute to its strong reliability.
2) **Booster Version Reliability**
   - The scatter plot analysis showed that the “FT” booster version consistently achieves a high success rate across diverse payload masses, highlighting its robustness. This makes it a strong candidate for future missions aiming to maximize reliability.
3) **Payload Mass vs. Success Rate**
   - No clear pattern was observed between payload mass and launch success. This indicates that other factors such as launch site conditions and booster technology play a more critical role in determining mission outcomes.
4) **Interactive Visualizations**
   - Tools such as Folium and Plotly Dash provided interactive insights into the geographical and operational patterns of SpaceX launches. These visualizations enabled stakeholders to engage directly with the data, enhancing decision-making through clear and comprehensive analytics.

- **Final Notes:**
  - The predictive analysis and interactive visualizations highlight the most influential factors behind SpaceX’s launch outcomes. By identifying the significance of launch sites and booster versions, while ruling out payload mass as a primary driver, it provides a robust framework for future assessments. These insights can help refine launch strategies, improve operational planning, and contribute to the ongoing advancement of reusable rocket technology in the aerospace industry.


Thank you!