IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Elena Sánchez-Laulhé Dégano
01/07/2026

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA (Exploratory Data Analysis)
  - Plotly Dash Dashboard
  - Predictive Analysis
- Summary of all results
  - EDA (Exploratory Data Analysis)
  - Interactive Dashboard
  - Predictive Analysis

Introduction

- Project background and context

SpaceX has revolutionized the space launch industry by developing reusable rocket technology, significantly reducing the cost of space missions. The Falcon 9 launch system plays a key role in this strategy, where successful first-stage landings enable booster reuse and operational efficiency.

Understanding the factors that influence launch success is critical for improving mission planning, reducing risks, and optimizing launch operations. This project uses historical Falcon 9 launch data to analyze mission outcomes and identify patterns associated with successful launches.

- Problems you want to find answers
  - Which factors have the strongest impact on Falcon 9 launch success?
  - How do launch site, payload mass, orbit type, and booster version relate to mission outcomes?
  - Are there specific launch sites or payload ranges associated with higher success rates?
  - Can a classification model accurately predict whether a launch will be successful based on historical data?
  - What are the limitations of predictive modeling given the available dataset?

Section 1

Methodology

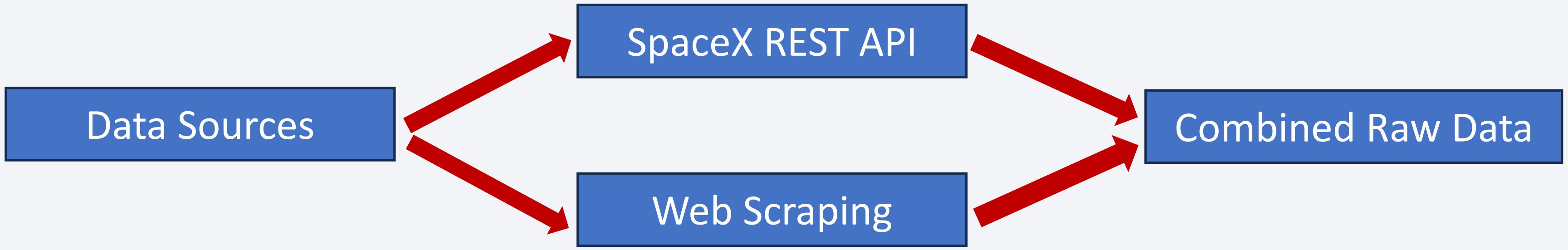
Methodology

Executive Summary

This capstone project analyzes SpaceX Falcon 9 launch data to identify the key factors that influence mission success and to build a predictive classification model for launch outcomes. The dataset was collected using the SpaceX REST API and web scraping from Wikipedia, followed by data wrangling, exploratory data analysis (EDA), and feature engineering. Exploratory analysis using data visualization and SQL revealed strong relationships between launch success and variables such as launch site, payload mass, orbit type, and booster version. Interactive visual analytics were developed using Folium to analyze geographical patterns and Plotly Dash to explore launch success trends by site and payload range. Several classification models, including Logistic Regression, Support Vector Machine, K-Nearest Neighbors, and Decision Tree, were trained and evaluated. All models achieved similar performance, with an accuracy of approximately 83.33%, indicating comparable predictive capability across methods. Due to the limited dataset size, model results are sensitive to data splitting, and performance differences should be interpreted with caution. Future work could include expanding the dataset, applying cross-validation, and incorporating additional operational features to improve model generalization.

Data Collection

- Data were collected from two sources:
  - SpaceX REST API (primary): multiple endpoints were queried to build the main analytical dataset (launch outcome, payload, orbit, booster, and launch site).
  - Wikipedia (supplementary): web-scraped tables were used to validate and complement historical context.
  - The primary dataset generated from the API lab was then used for data wrangling, EDA, and modeling.



Data Collection – SpaceX API

- Launch data was collected using the SpaceX REST API through HTTP GET requests.
- Relevant endpoints were queried to retrieve information about launches, payloads, rockets, landing, outcomes and launch sites.
- Nested JSON responses were parsed and converted into structured tabular format for analysis.
- The API data served as the primary source for launch outcome and mission related variables.

[GitHub Notebook]

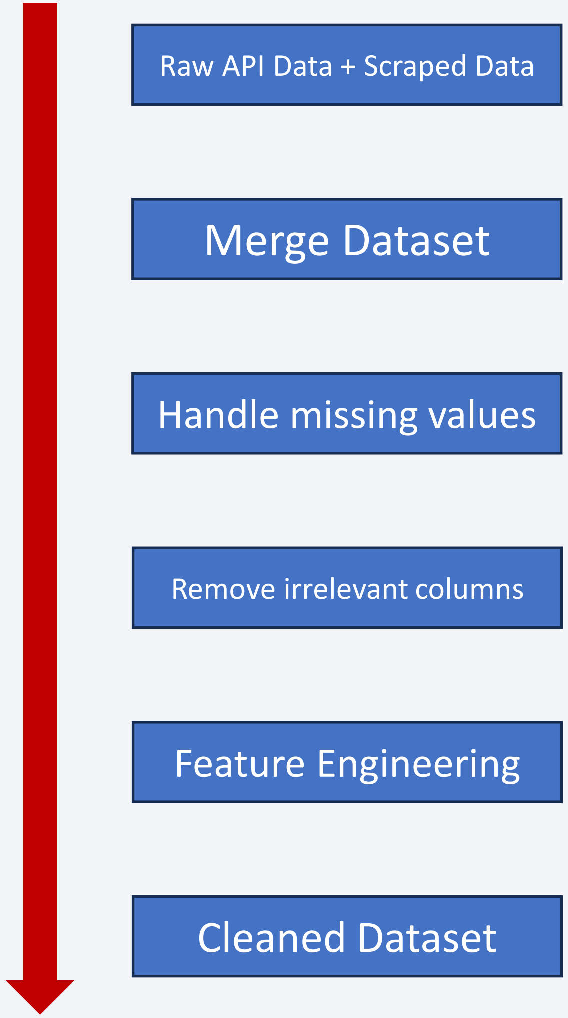
Data Collection - Scraping

- Web scraping was used to complement API data with additional historical and descriptive information.
- Launch related tables were extracted from Wikipedia using HTML parsing techniques.
- Relevant rows and columns were filtered and cleaned to ensure data consistency.
- Scraped data was merged with API data to create a unified dataset.

[GitHub Notebook]

Data Wrangling

- The analytical dataset (dataset_part_1) produced in the API lab was loaded for cleaning and feature preparation.
- Missing values were identified and handled.
- Irrelevant and redundant columns were removed.
- Categorical variables were encoded and numerical features were standardized.
- A binary target variable (class) representing launch success was created for predictive analysis.



[GitHub Notebook]

EDA with Data Visualization

- Scatter plots were used to analyze the relationship between payload mass, flight number, launch site and orbit type with launch success.
- Bar charts were created to compare success rates across different orbit types.
- Line charts were used to analyze the historical trend of launch success over time.
- These visualizations were selected to identify patterns, correlations and trend that could influence mission outcomes.

GitHub Notebook

EDA with SQL

- Retrieved the list of all unique launch sites.
- Identified launch sites starting with specific prefixes.
- Calculated total and average payload mass by booster version.
- Retrieved the first successful landing dates.
- Identified boosters with successful landings under specific payload conditions.
- Calculated the total number of successful and failed missions.
- Ranked landing outcomes over a specific period.

GitHub Notebook

Build an Interactive Map with Folium

- Markers were added to present the geographical locations of SpaceX launch sites.
- Color coded markers were used to distinguish successful and failed launch outcomes.
- Circles and distance lines were added to analyze proximity to infrastructure such as coastlines, highways and railways.
- These map objects were used to explore geographical patterns and potential environmental and logistical factors affecting launch success.

[GitHub Notebook]

Build a Dashboard with Plotly Dash

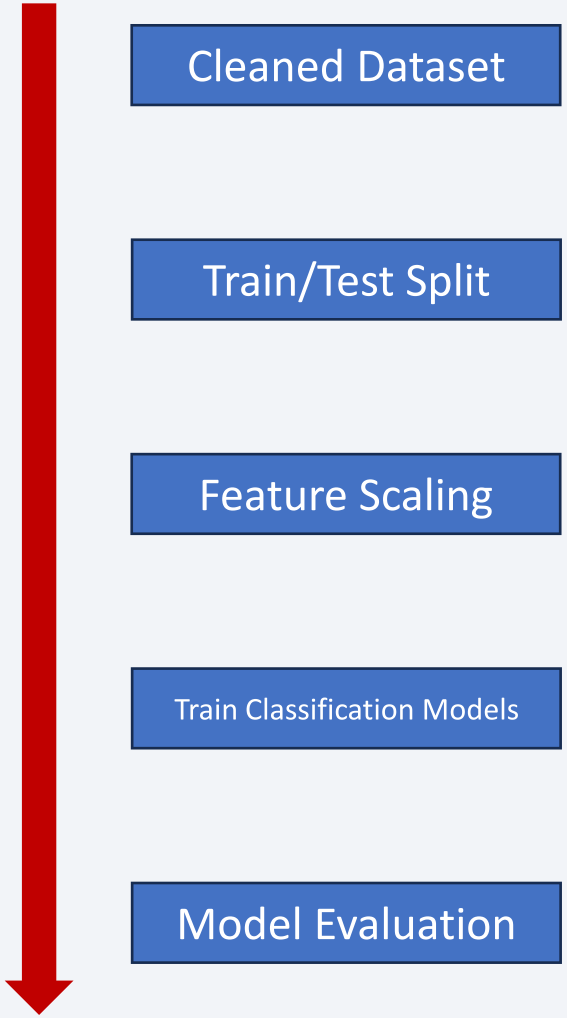
- A pie chart was created to display the total number of successful launches by site.
- A second pie chart shows the success ratio for the launch site with highest success rate.
- A scatter plot of payload mass versus launch outcome was added with an interactive range slider.
- These interactive components allow users to explore how payload range and launch site affect mission success.
- The dashboard distinguishes between total successful launches (counts) and success ratio (within a selected site).

GitHub Notebook

Predictive Analysis (Classification)

- The dataset was split into training and testing sets
- Several classification models were built, including Logistic Regression, Support Vector Machine, K-Nearest Neighbors and Decision Tree.
- Feature scaling and hyperparameter tuning were applied to improve model performance.
- Models were evaluated using accuracy scores and confusion matrices to compare performance.

[GitHub Notebook]



Results

The results section presents insights derived from exploratory data analysis, interactive visual analytics, and predictive modeling.

Exploratory analysis highlights key relationships between launch success and variables such as launch site, payload mass, orbit type, and mission chronology.

Interactive dashboards and maps provide deeper insight into spatial and operational patterns.

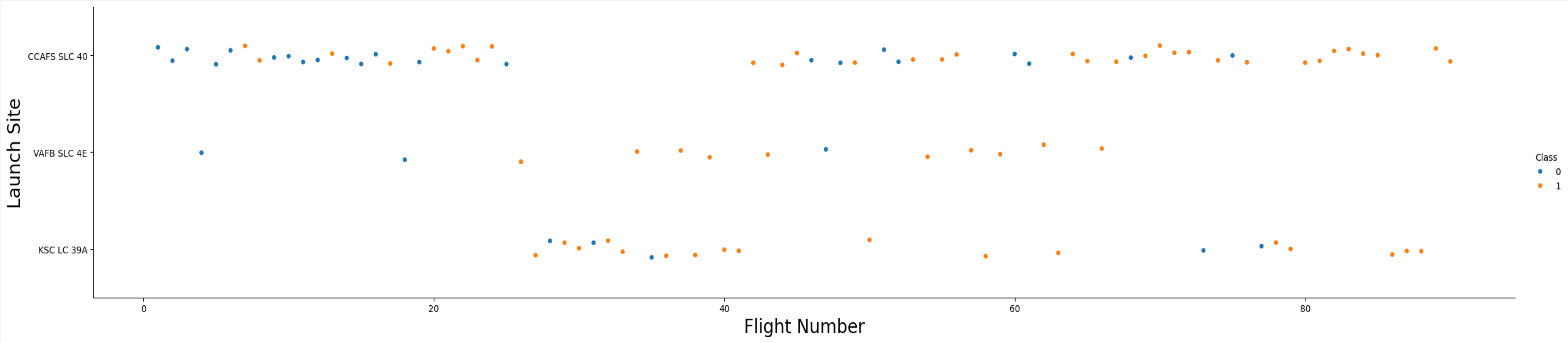
Predictive analysis evaluates the extent to which launch outcomes can be inferred from historical mission data.

Section 2

Insights drawn from EDA

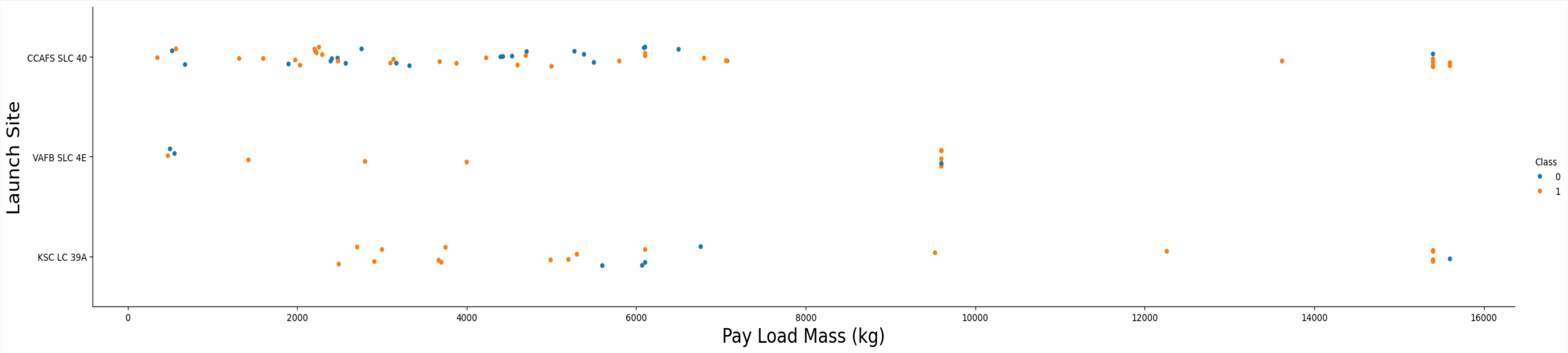
Flight Number vs. Launch Site

- Early Falcon 9 launches show a higher frequency of failures across all launch sites.
- As the flight number increases, successful launches become more frequent, indicating a clear improvement over time.
- CCAFS SLC 40 and KSC LC 39A show a strong increase in successful launches in later missions.
- This trend suggests that launch success is positively correlated with operational experience and technological maturity rather than launch site alone.



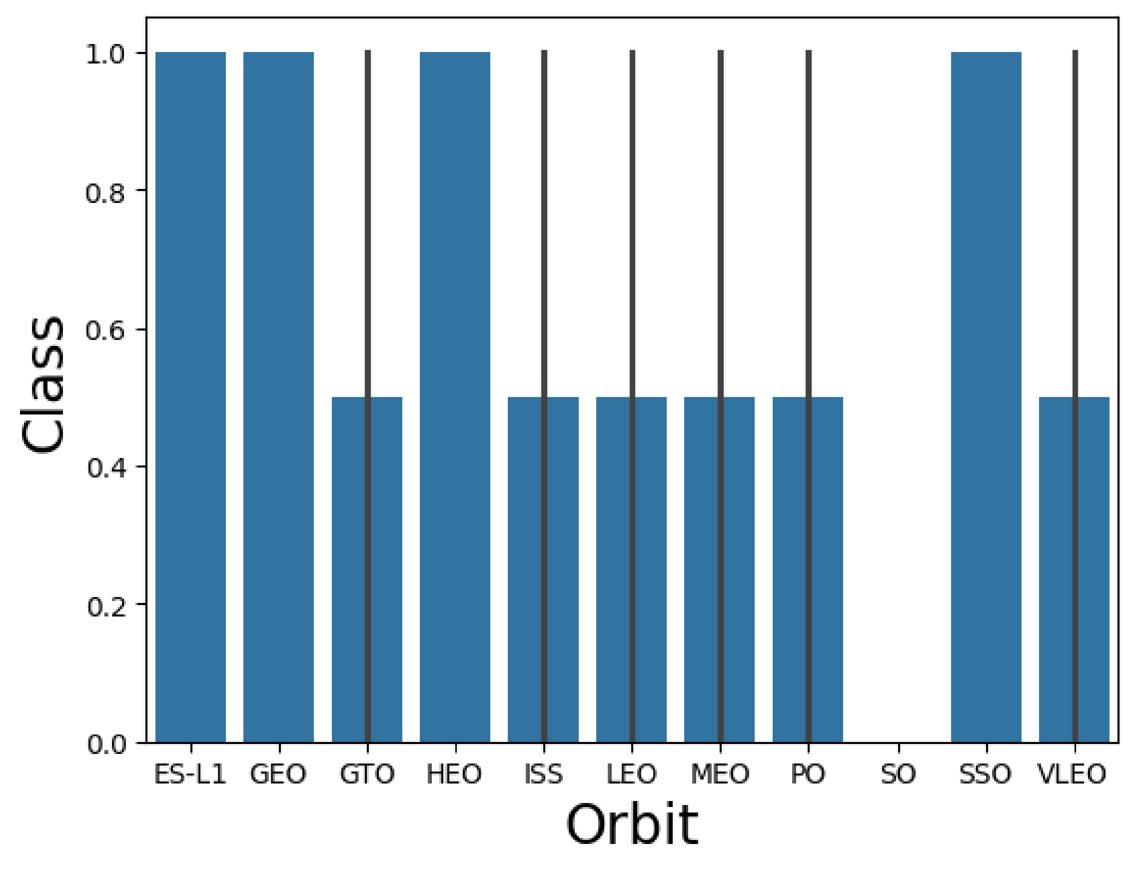
Payload vs. Launch Site

- Successful launches are observed across a wide range of payload masses at all launch sites.
- CCAFS SLC 40 handles a broad distribution of payloads, including several high-payload missions with successful outcomes.
- KSC LC 39A is associated with medium to high payload missions, showing a high proportion of successful launches.
- VAFB SLC 4E mainly supports lower to medium payload missions, with fewer launches at very high payload masses.
- Overall, payload mass alone does not determine launch success; its impact varies depending on the launch site and mission profile.



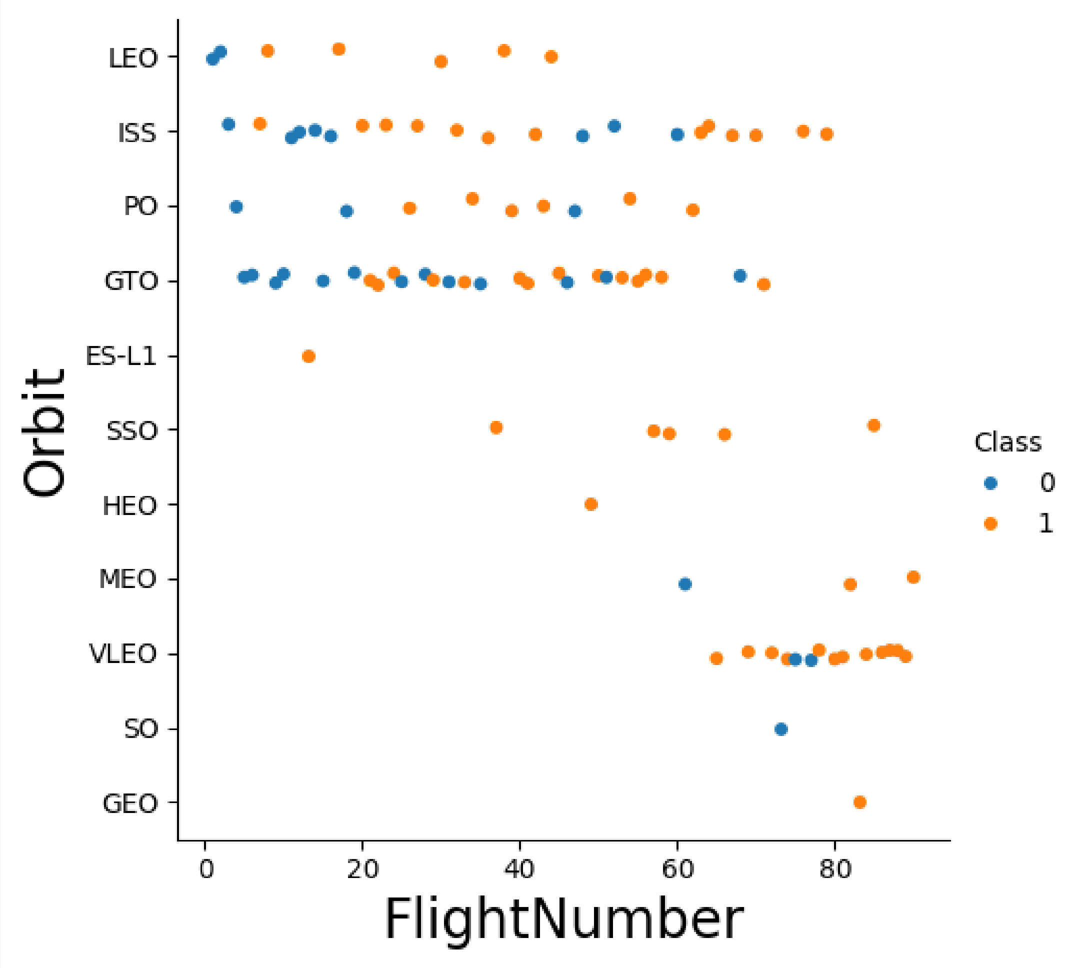
Success Rate vs. Orbit Type

- Certain orbit types, such as GEO, HEO, ES-L1, and SSO, show higher success rates compared to others.
- Orbits like LEO, ISS, MEO, and PO present more mixed outcomes, indicating higher variability in mission success.
- Differences in success rates across orbit types suggest that mission complexity and orbital requirements influence launch outcomes.
- Orbit type appears to be a relevant factor in launch success, although it should be considered together with other variables such as payload mass and booster version.



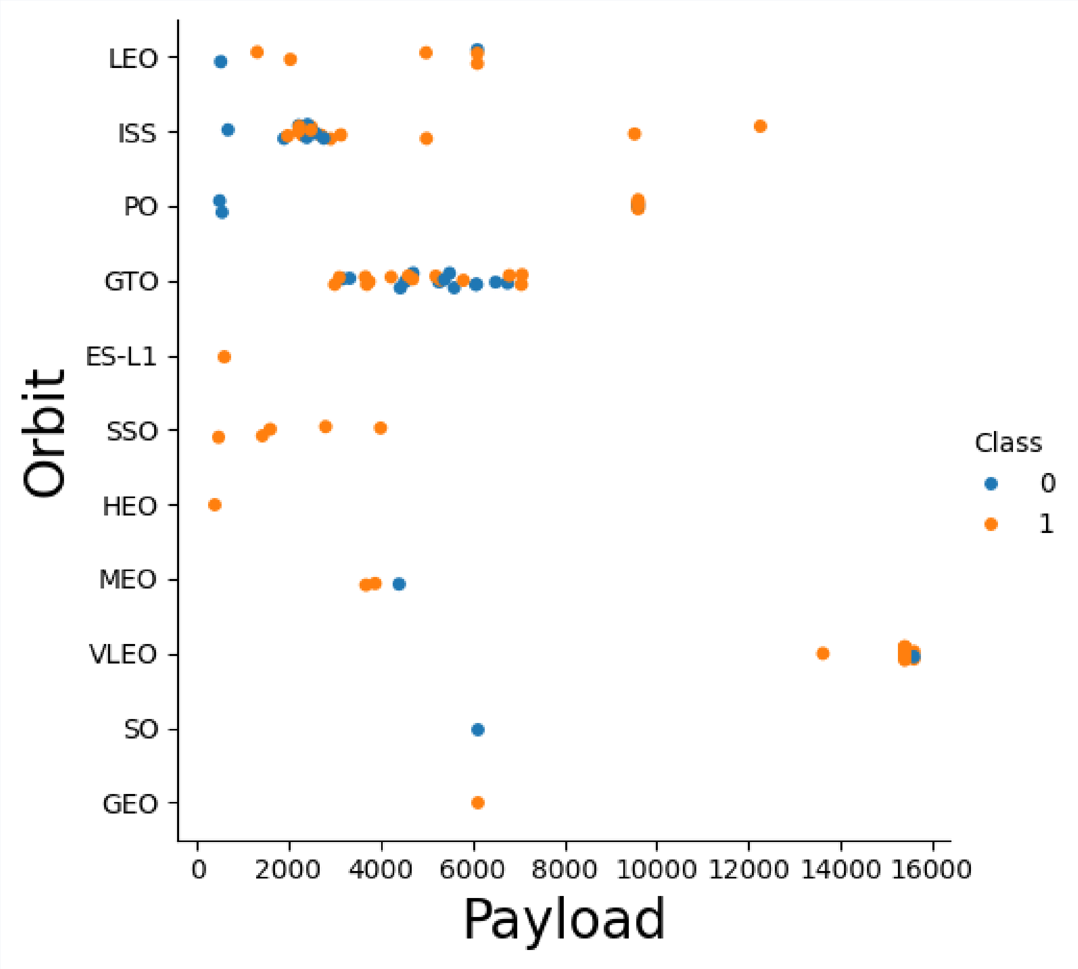
Flight Number vs. Orbit Type

- Early Falcon 9 missions are concentrated in a limited number of orbit types, primarily LEO, ISS, and GTO.
- As the flight number increases, SpaceX missions expand to a wider range of orbit types, including more complex orbits such as SSO, HEO, and VLEO.
- Later missions show a higher proportion of successful outcomes across multiple orbit types.
- This trend suggests that increased operational experience enables SpaceX to reliably support more diverse and complex mission profiles over time.



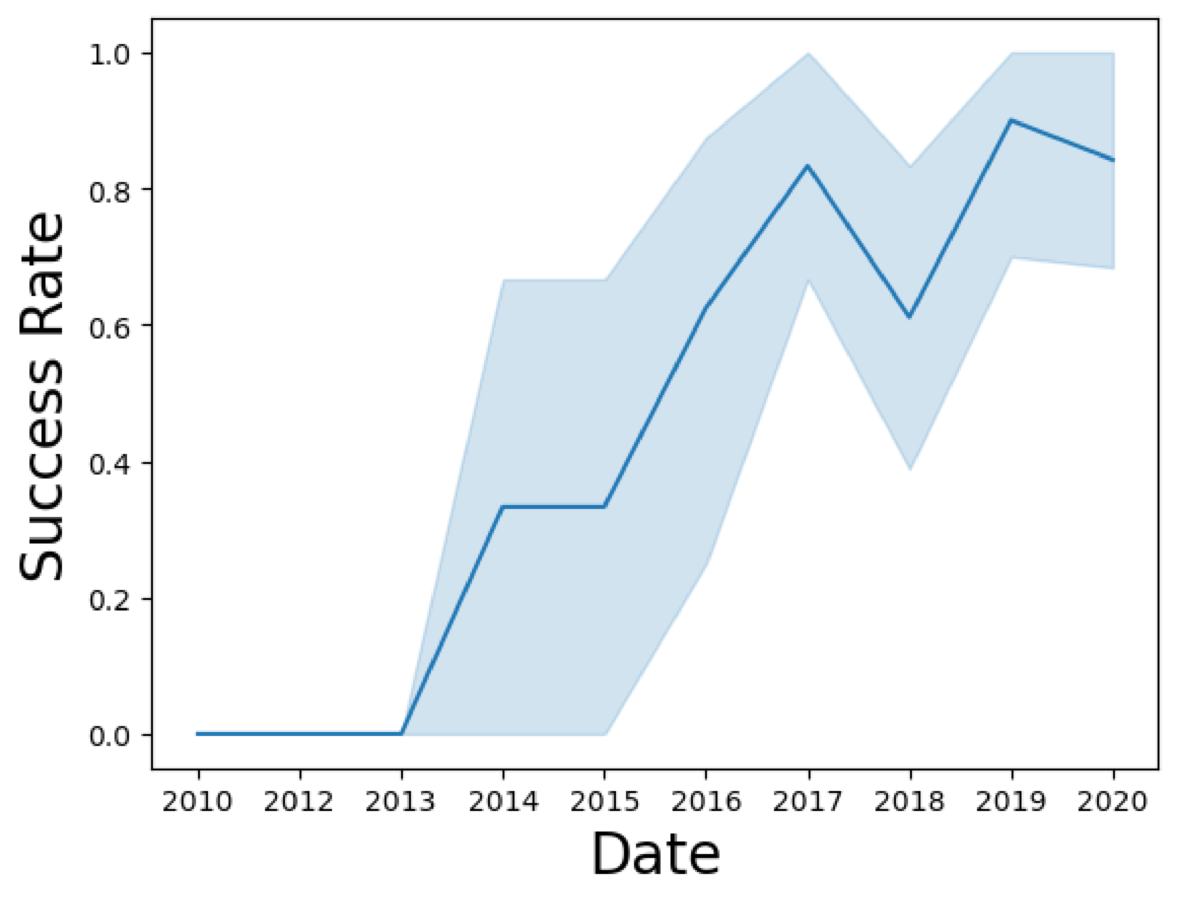
Payload vs. Orbit Type

- Payload mass distributions vary significantly across orbit types.
- LEO, ISS, and GTO missions are associated with a wide range of payload masses, including many successful launches.
- More specialized orbits such as SSO, HEO, and ES-L1 generally involve lower payload masses.
- VLEO missions are associated with very high payload masses and show a high proportion of successful outcomes.
- These patterns indicate that payload requirements are strongly influenced by orbital characteristics, and payload mass should be evaluated together with orbit type when assessing launch success.



Launch Success Yearly Trend

- The yearly average launch success rate shows a clear upward trend over time.
- Early years are characterized by low or zero success rates, reflecting the initial development and testing phase.
- From 2014 onwards, launch success improves steadily, reaching consistently high levels in later years.
- Temporary fluctuations are observed, but the overall trend indicates continuous improvement in launch reliability.
- This pattern highlights the impact of technological maturity, operational learning, and process optimization on mission success.



All Launch Site Names

- Query Result
  - The dataset contains four unique launch sites
    - CCAFS LC-40
    - VAFB SLC-4E
    - KSC LC-39A
    - CCAFS SLC-40
- Explanation
  - These launch sites represent the complete set of locations used by SpaceX for the missions included in the dataset.

Launch Site Names Begin with 'CCA'

- Query Result
  - Launches Sites beginning with CCA:
    - CCAFS LC-40
    - CCAFS SLC-40
- Explanation
  - Filtering launch sites by the “CCA” prefix isolates missions launched from Cape Canaveral–related complexes.
  - This allows focused analysis on launches originating from this region, which hosts multiple SpaceX launch pads.
  - Such filtering is useful for regional performance comparison and launch-site-specific analysis.

Total Payload Mass

- Query Result
  - Total payload mass carried by NASA missions: 45596 kg
- Explanation
  - The query aggregates the total payload mass associated with NASA missions, providing a cumulative measure of payload delivery.
  - This result reflects the overall contribution of SpaceX launches to NASA-related missions within the dataset.
  - It also serves as an indicator of SpaceX's role in supporting government and institutional payload requirements.

Average Payload Mass by F9 v1.1

- Query Result
  - Average payload mass for Falcon 9 v1.1: 2928.4 kg
- Explanation
  - The calculated average payload mass represents the typical payload capacity carried by the Falcon 9 v1.1 booster.
  - This value helps characterize the operational profile of this booster version.
  - Understanding average payload capacity is useful when comparing different booster generations or evaluating performance evolution.

First Successful Ground Landing Date

- Query Result
  - First successful ground pad landing date: 2015-12-22
- Explanation
  - This query identifies the earliest date on which a successful ground pad landing was recorded.
  - The result marks a key technological milestone in SpaceX’s reusability program.
  - Ground landings represent a significant advancement in recovery reliability compared to earlier landing attempts.
  - This metric reflects cumulative payload delivered for NASA-related missions in the dataset (not all SpaceX history).

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query Result
  - Boosters that successfully landed on a drone ship with payloads between 4000 and 6000 kg:
    - F9 FT B1022
    - F9 FT B1026
    - F9 FT B1021.2
    - F9 FT B1031.2
- Explanation
  - The query isolates missions that successfully landed on a drone ship while carrying medium-range payloads.
  - This demonstrates SpaceX's capability to recover boosters even under non-trivial payload conditions.
  - These results highlight the feasibility of offshore recovery for operational missions, not just experimental launches.

Total Number of Successful and Failure Mission Outcomes

- Query Result
  - Total successful missions: 99
- Explanation
  - This query summarizes mission outcomes by counting successful and failed launches.
  - The result provides a high-level view of overall mission performance across the dataset.
  - Such summary statistics are essential for understanding class balance prior to predictive modeling.

Boosters Carried Maximum Payload

- Query Result
  - Boosters that carried the maximum payload mass:
    - F9 B5 B1048.4
    - F9 B5 B1049.4
    - F9 B5 B1051.6
    - F9 B5 B1060.3
    - F9 B5 B1049.7
- Explanation
  - The query identifies the boosters associated with the highest payload mass recorded in the dataset.
  - These boosters represent peak payload capability within the analyzed missions.
  - Identifying maximum payload cases helps assess performance limits and operational extremes.

2015 Launch Records

- Query Result
  - Failed drone ship landings in 2015:
    - Flight 01 – Booster F9 v1.1 B1012, Launch Site CCAFS LC-40
    - Flight 04 – Booster F9 v1.1 B1015, Launch Site CCAFS LC-40
- Explanation
  - The query retrieves failed drone ship landing attempts during 2015, including booster versions and launch sites.
  - These records reflect early challenges in booster recovery operations.
  - Analyzing these failures provides historical context for later improvements in landing success rates.

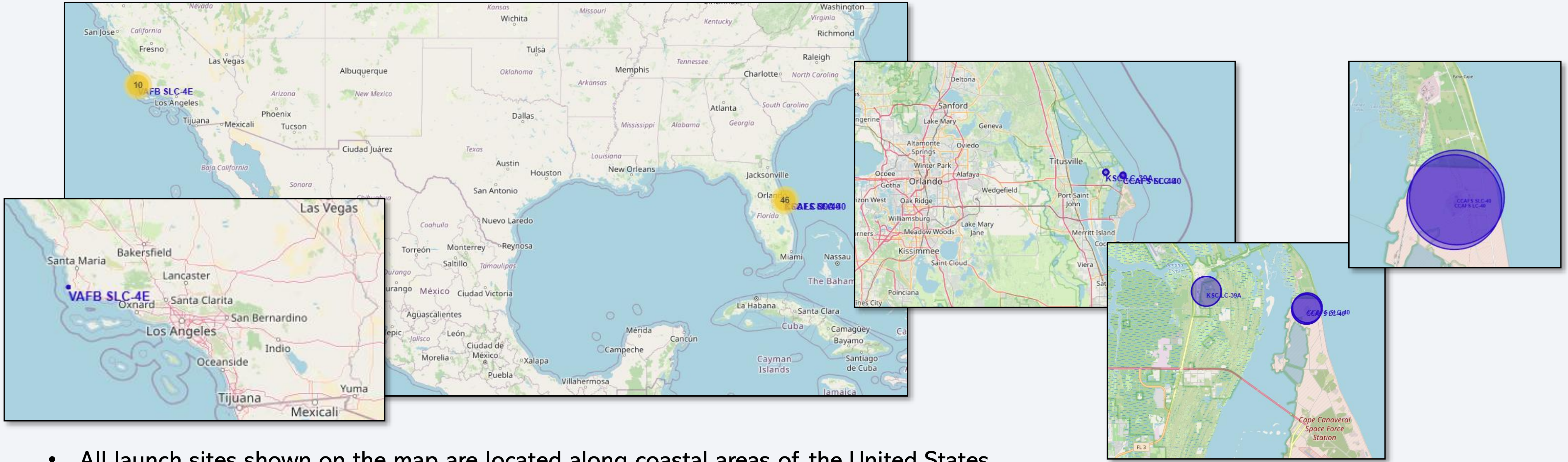
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Query Result
  - Landing outcome ranking:
    - Success (drone ship): 5
    - Success (ground pad): 3
- Explanation
  - The ranking shows the relative frequency of different landing outcomes within the specified period.
  - Successful drone ship landings appear more frequently than successful ground pad landings during this timeframe.
  - This result illustrates the growing reliance on offshore recovery methods as part of SpaceX's operational strategy.

Section 3

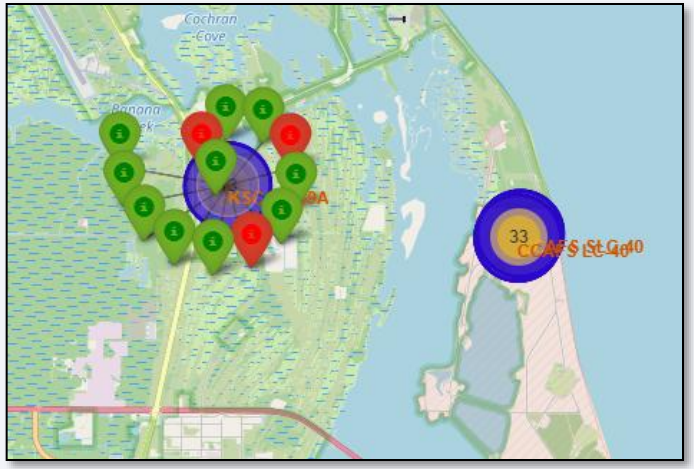
Launch Sites Proximities Analysis

Global Distribution of SpaceX Launch Sites



- All launch sites shown on the map are located along coastal areas of the United States.
- The spatial distribution indicates a clear preference for coastal locations, which is consistent with launch safety requirements and trajectory constraints.
- The launch sites are geographically separated, suggesting independent operational zones rather than clustered facilities.

Launch Outcomes by Site



- Both successful and failed launches are observed across multiple launch sites.
- No single launch site is exclusively associated with failures.
- Successful launches appear repeatedly at the same locations over time.
- This suggests that launch success is not determined solely by geographic location but likely influenced by operational and technological factors.

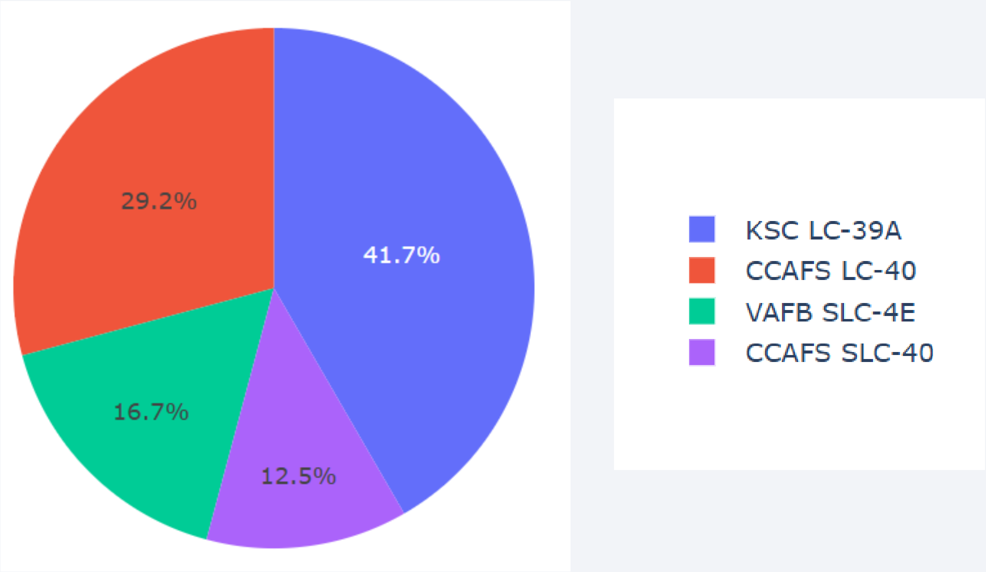
Proximity of Launch Site to Infrastructure and Coastline

- The selected launch site (CCAFS SLC-40) is located very close to the coastline, allowing launch trajectories over open water.
- Distance indicators show short distances to major infrastructure, including highways and nearby facilities (less than 1 km)
- The proximity to transportation routes supports efficient logistics for equipment transport and personnel access.
- Coastal proximity also facilitates booster recovery operations, particularly for drone ship landings offshore.

Section 4

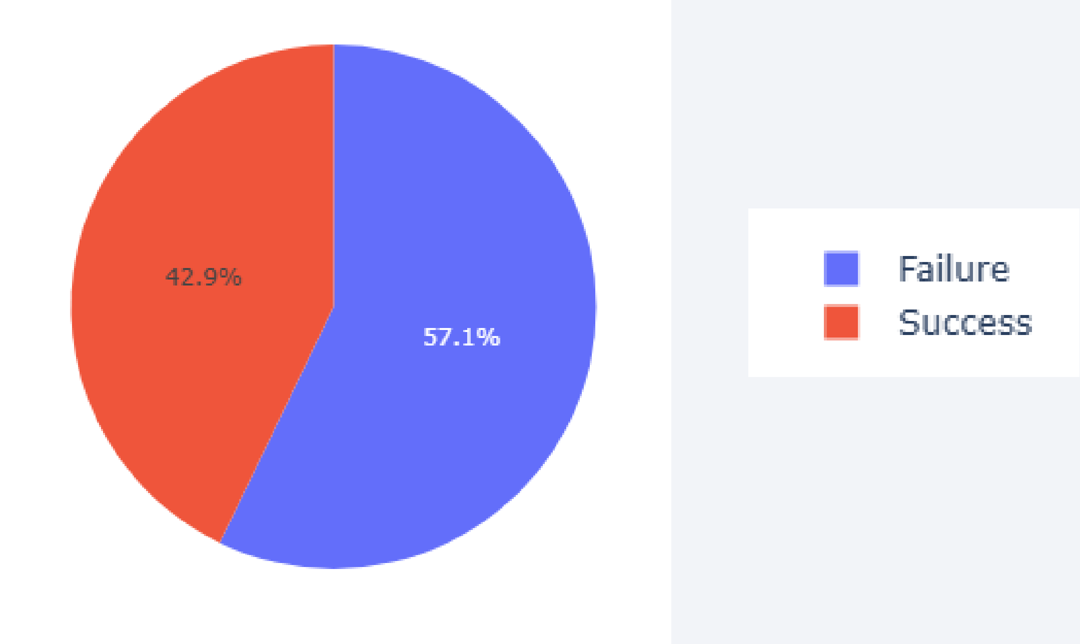
Build a Dashboard with Plotly Dash

Total Successful Launches by Site



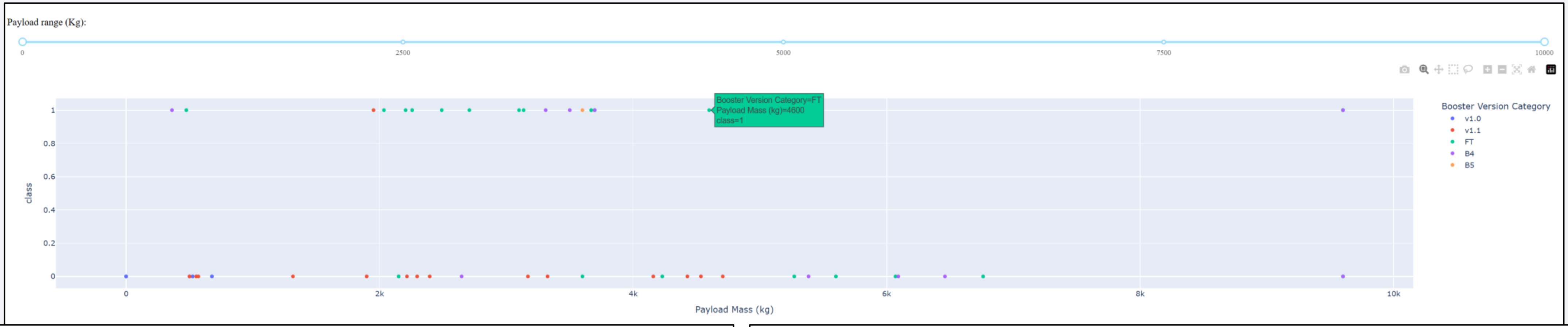
- The pie chart displays the distribution of total successful launches across all launch sites.
- KSC LC-39A contributes the largest proportion of successful launches, accounting for approximately 41.7% of total successes.
- CCAFS LC-40 is the second most successful site with roughly 29.2%, followed by VAFB SLC-4E (16.7%) and CCAFS SLC-40 (12.5%).
- This indicates that KSC LC-39A is currently the most productive and reliable launch site.

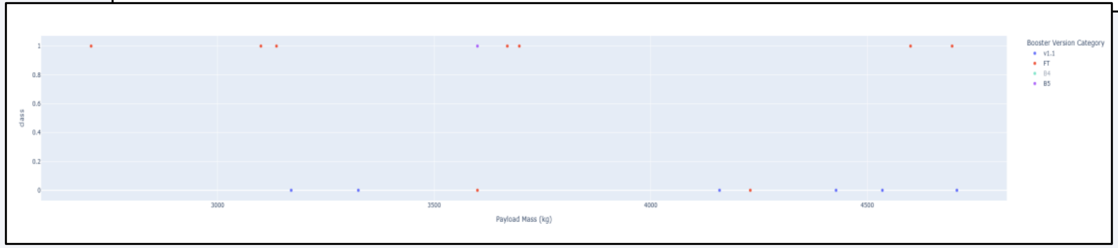
Success vs Failure for selected Launch Site (CCAFS SLC-40)

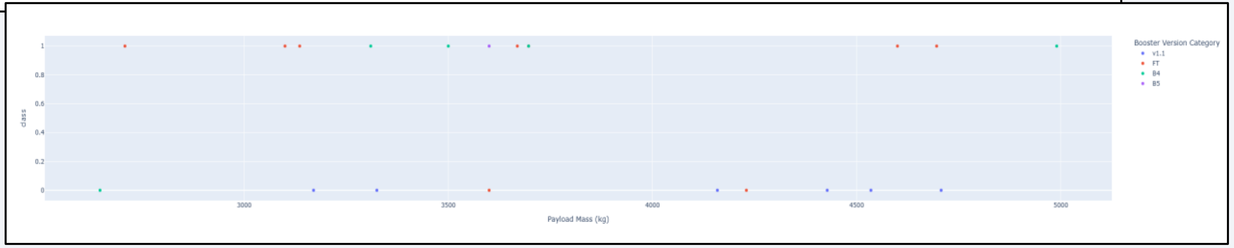


- When filtering the dashboard to CCAFS SLC-40, the pie chart shows a dominance of successful launches.
- Approximately 42,9% of launches from this site are successful, while 57,1% resulted in failure.
- This success ratio highlights CCAFS SLC-40 as the most mature and operationally stable launch site.
- The visualization confirms that site-specific analysis provides more accurate insights than global aggregation.

Payload Mass vs Launch Success







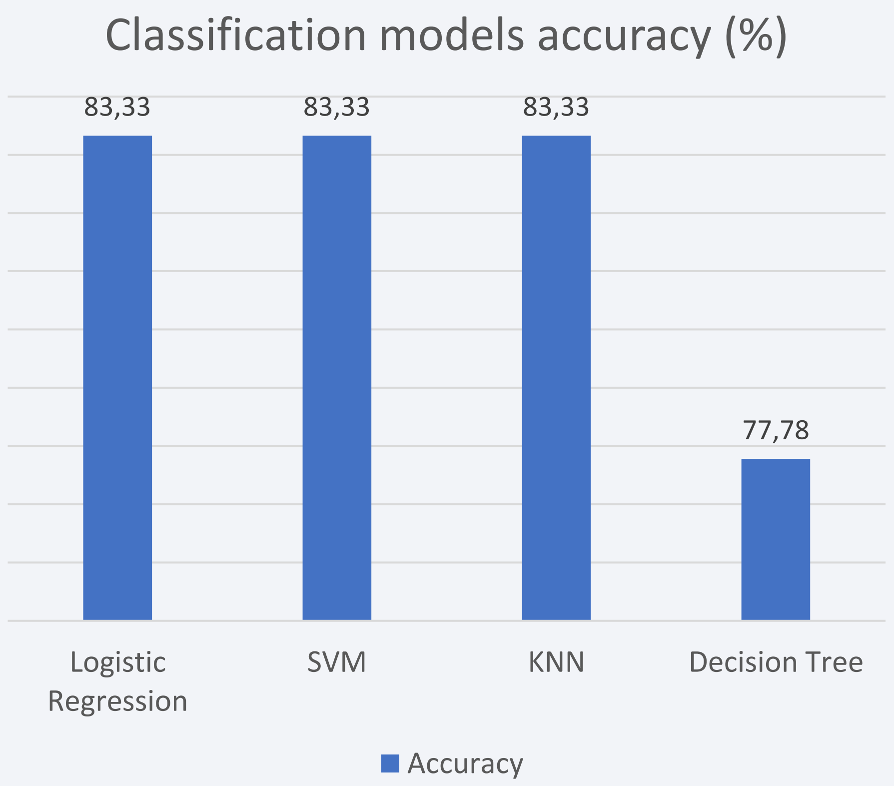
- The scatter plot shows the relationship between payload mass and launch success, with color indicating booster version category.
- Successful launches (class = 1) are more frequent in the payload range between 2,000 kg and 6,000 kg.
- Certain booster versions (notably FT and B4) demonstrate higher success consistency across wider payload ranges.
- Failures are more scattered at very low and very high payload masses, suggesting operational limits and early-stage testing effects.
- The interactive payload slider reveals that payload mass and booster version together strongly influence launch outcomes.

Section 5

Predictive Analysis (Classification)

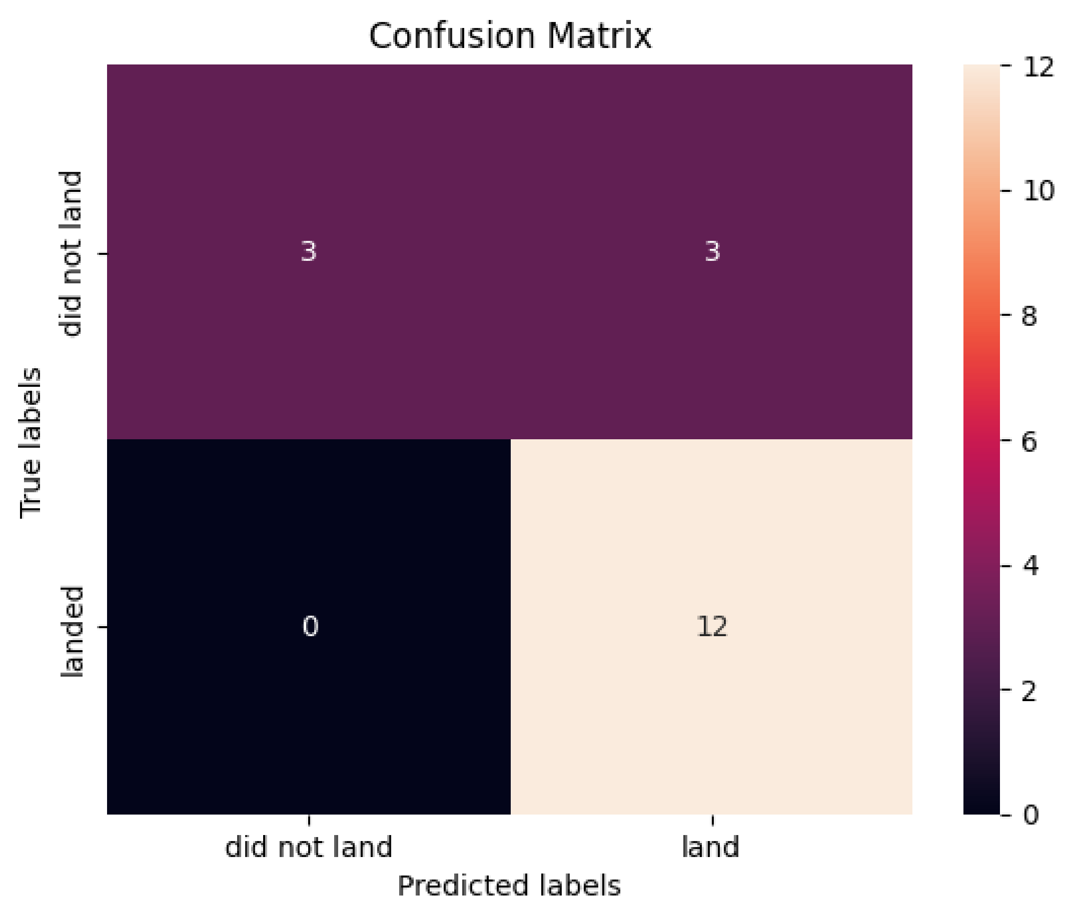
Classification Accuracy

- Test set accuracy comparison
  - Logistic Regression: 83.33%
  - SVM: 83.33%
  - KNN: 83.33%
  - Decision Tree: 77.78%
- Key findings
  - Logistic Regression, SVM, and KNN show equivalent test performance.
  - Decision Tree performs slightly worse, likely due to higher sensitivity to data splitting.
  - Accuracy values are computed on the held-out test set.
  - With a relatively small test set, identical accuracy scores across models suggest limited sensitivity to model choice.



Confusion Matrix – Logistic Regression

- Given similar accuracy across models, Logistic Regression was selected due to its interpretability and stable performance.
- The model correctly predicts most successful launches.
- Some failed launches are misclassified as successful, reflecting overlap in feature space.
- This confirms that while the model generalizes reasonably well, classification is challenging for borderline cases.



Conclusions

- Launch site performance differs significantly: KSC LC-39A and CCAFS LC-40 concentrate the highest number of successful launches, while VAFB SLC-4E shows fewer total missions, mainly associated with specific orbit types.
- Payload mass influences launch success: Medium payload ranges (approximately 2,000–6,000 kg) show a higher concentration of successful launches, whereas very low and very high payloads present more failures.
- Orbit type impacts mission outcomes: LEO missions present the highest success rates, while more complex orbits such as GTO and polar orbits show higher variability in outcomes.
- Technological evolution improves success rates: Launch success rates increase over time, reflecting improvements in booster versions and operational experience.
- Predictive models achieve moderate accuracy: All tested classification models show similar performance (~83% test accuracy), indicating that launch success is predictable to a certain extent but still influenced by complex, non-linear factors.

Appendix

- Data Sources and Resources
  - GitHub Repository
    - https://github.com/esanchezlaulhe1312/ibm-datascience-capstone-spacex.git
  - Primary Data Sources
    - SpaceX REST API:
      - https://api.spacexdata.com/v4/rockets/
      - https://api.spacexdata.com/v4/launchpads/
      - https://api.spacexdata.com/v4/payloads/
      - https://api.spacexdata.com/v4/cores/
  - Supplementary Data
    - Wikipedia: List of Falcon 9 and Falcon Heavy launches: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
  - Course Dataset
    - IBM Skills Network: Falcon 9 launch dataset: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json


Thank you!