IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jericho Ross R. Operio
22 Dec 2025

Outline

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

Executive Summary

- Launch data was collected through web scraping and API access, then cleaned and prepared through structured data wrangling
- Preliminary exploratory data analysis was conducted using SQLite to examine trends and relationships within the dataset
- Further EDA and visualization were performed to identify key patterns related to landing outcomes
- Interactive visual analytics were developed using Folium and Dash to explore spatial and operational insights
- Predictive analysis was conducted using machine learning classification models, including Logistic Regression, SVM, Decision Tree, and KNN

Executive Summary

- Logistic Regression, SVM, and KNN achieved the highest test accuracy (83.3%) with identical confusion matrices, indicating consistent and stable predictions
- The Decision Tree model showed lower performance (77.8%), suggesting reduced generalization on the available dataset
- Logistic Regression was selected as the final model due to its strong performance, simplicity, and interpretability
- Overall, the project demonstrates that Falcon 9 first-stage landing success can be reliably predicted, supporting data-driven cost estimation and competitive launch analysis

Introduction

- SpaceX's Falcon 9 launch system significantly reduces launch costs through first-stage reusability
- Falcon 9 launches are advertised at approximately $62 million, compared to $165 million or more from competing providers
- The primary source of cost savings is SpaceX's ability to successfully land and reuse the first-stage booster
- Accurately predicting first-stage landing success enables estimation of true launch costs
- Such predictions provide valuable insights for competing aerospace companies when bidding against SpaceX for launch contracts
- This capstone project applies data analytics and machine learning techniques to predict Falcon 9 first-stage landing outcomes

Introduction

- Since Falcon 9’s cost advantage relies heavily on successful first-stage recovery, accurately assessing landing outcomes is essential for understanding true launch costs
- Landing success depends on multiple mission and booster-related factors, making outcomes non-trivial to predict
- Using historical Falcon 9 launch data, this capstone project seeks to:

1. Predict whether the first-stage booster will land successfully
2. Evaluate the effectiveness of machine learning classification models in making this prediction
3. Compare model performance to identify the most reliable approach

- The resulting predictions can support data-driven cost estimation and inform competitive decision-making in the commercial launch market

Section 1

Methodology

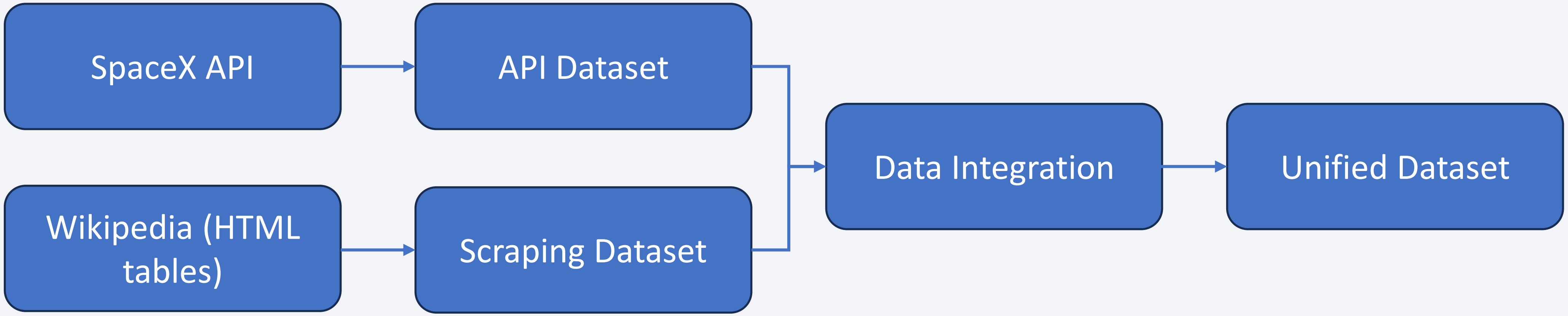
Methodology

Executive Summary

- Data collection methodology:
  - Collected Falcon 9 launch data through web scraping and public APIs
- Perform data wrangling
  - Cleaned and processed raw data to handle missing values and inconsistencies
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built and evaluated classification models to predict first-stage landing success
  - Tuned model hyperparameters using cross-validation
  - Compared model performance to identify the most effective approach

Data Collection

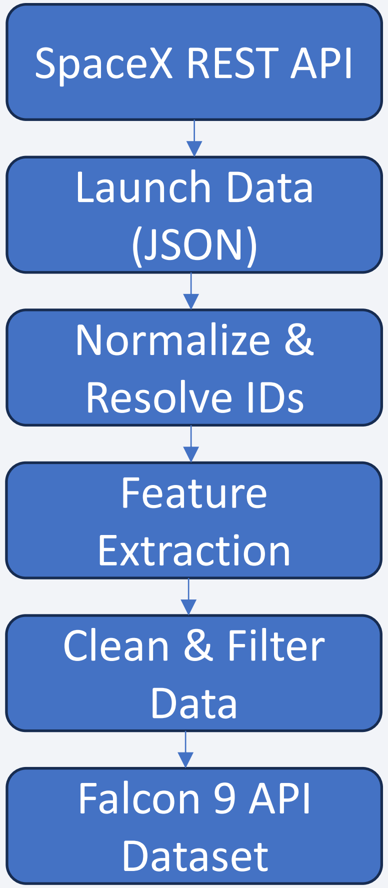
- Collected launch and mission data from the SpaceX REST API and cleaned the responses
- Web-scraped Falcon 9 launch records from Wikipedia using BeautifulSoup
- Extracted and parsed HTML tables into Pandas Dataframes
- Integrated API and scraped datasets into a single unified dataset for analysis and modeling



Data Collection – SpaceX API

- Retrieved launch data via SpaceX REST API (GET requests)
- Parsed and normalized JSON responses into tabular format
- Extracted booster, payload, launch site, and landing attributes
- Filtered Falcon 9 launches and handled missing values

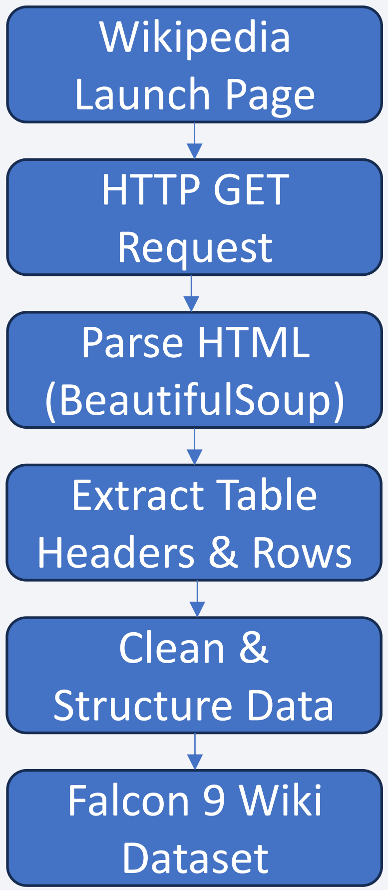
https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

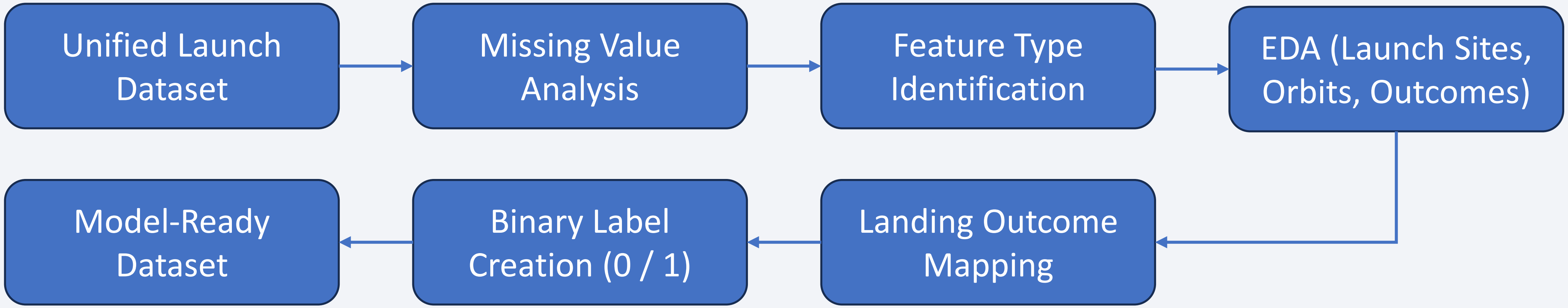
- Scraped Falcon 9 launch records from Wikipedia using BeautifulSoup
- Retrieved HTML page via HTTP GET request
- Identified and extracted the launch records table
- Parsed table headers and rows into structured data
- Converted cleaned records into a Pandas DataFrame

https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/jupyter-labs-webscraping.ipynb



Data Wrangling

- Loaded unified Falcon 9 launch dataset
- Analyzed missing values
- Identified numerical and categorical features
- Performed exploratory analysis on launch sites, orbits, and outcomes
- Converted landing outcomes into binary training labels
- Prepared final dataset for machine learning



https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Key Visualizations Used:

- **Bar charts**

1. Launch success vs. failure
2. Number of launches by launch site
3. Distribution of orbit types

- **Scatter plots**

1. Payload mass vs. landing success
2. Flight number vs. landing outcome

- **Categorical plots**

1. Landing outcome by orbit type
2. Landing outcome by launch site

https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/edadataviz.ipynb

Why These Visualizations

- Identify **patterns and trends** affecting first-stage landing success
- Examine relationships between **mission parameters** (orbit, payload, site) and outcomes
- Detect **imbalances or biases** in categorical variables
- Support **feature selection** for predictive modeling

Key Insights

- Landing success improves with **increasing flight number**
- Certain **orbits and launch sites** show higher success rates
- Higher payload mass tends to reduce landing success probability

EDA with SQL

SQL Analysis Tasks Performed:

- Queried unique launch sites and filtered launches by site name patterns
- Retrieved specific launch records using conditional string matching
- Aggregated total and average payload mass by customer and booster version
- Identified the earliest successful ground-pad landing using date aggregation
- Filtered launches by landing outcome, payload mass range, and booster type
- Counted and compared successful vs. failed mission outcomes
- Used subqueries to identify boosters carrying maximum payload mass
- Extracted time-based insights (year and month) from launch dates
- Ranked landing outcome frequencies over a specified historical period

https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Map Objects Created

- Folium Map centered at NASA Johnson Space Center for geographic context
- Circle markers to highlight and label each SpaceX launch site
- Standard markers to pinpoint exact launch site locations
- Color-coded markers (green/red) to represent successful vs. failed launches
- Marker clusters to manage overlapping launch records at the same site
- MousePosition plugin to dynamically display geographic coordinates
- Polylines to visualize distances between launch sites and nearby features (e.g., coastline)

https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/lab_jupyter_launch_site_location.ipynb

Why These Objects Were Used

- Visualize the global and regional distribution of Falcon 9 launch sites
- Compare launch success rates by location using interactive, color-coded markers
- Reduce visual clutter and improve interpretability with marker clustering
- Analyze geographic proximity to coastlines, highways, and railways
- Support exploration of spatial factors influencing launch site selection and success

Key Insights

- Launch sites are generally close to coastlines, supporting safe recovery operations
- Sites maintain distance from dense urban areas, reducing operational risk
- Geographic location appears to influence launch success patterns

Build a Dashboard with Plotly Dash

Plots and Interactions Added

- **Launch Site Dropdown**

 - Allows selection of all sites or a specific launch site

- **Success vs. Failure Pie Chart**

 - Displays total successful launches for all sites

 - Shows success vs. failure breakdown for a selected site

- **Payload Range Slider**

 - Enables dynamic filtering of launches by payload mass

- **Payload vs. Launch Outcome Scatter Plot**

 - Visualizes relationship between payload mass and launch success

 - Color-coded by Falcon 9 booster version

https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/spacex-dash-app.py

Why These Plots and Interactions Were Used

- Enable interactive exploration of launch performance by site
- Compare success rates across launch locations
- Examine how payload mass influences mission success
- Identify booster versions associated with higher success rates
- Support data-driven insights without static assumptions

Key Insights

- Identification of launch sites with highest success counts and rates
- Payload ranges associated with higher and lower success probabilities
- Booster versions demonstrating stronger performance trends

Predictive Analysis (Classification)

Model Development Process

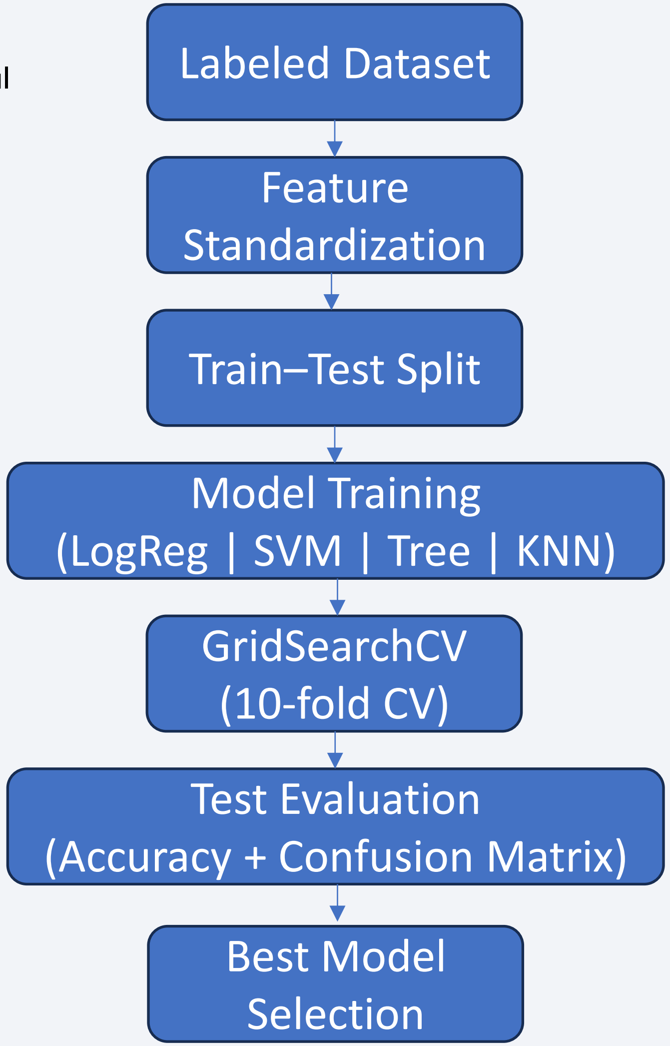
- **Label definition:** Converted landing outcomes into binary class(1 = successful landing, 0 = unsuccessful landing)
- **Feature scaling:** Standardized input features using **StandardScaler**
- **Data splitting:** Train–test split (80% training, 20% testing)
- **Model training:**

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree Classifier
4. K-Nearest Neighbors (KNN)

- **Hyperparameter tuning:** Used GridSearchCV (cv = 10) to optimize each model
- **Model evaluation:**

1. Test accuracy
2. Confusion matrix analysis

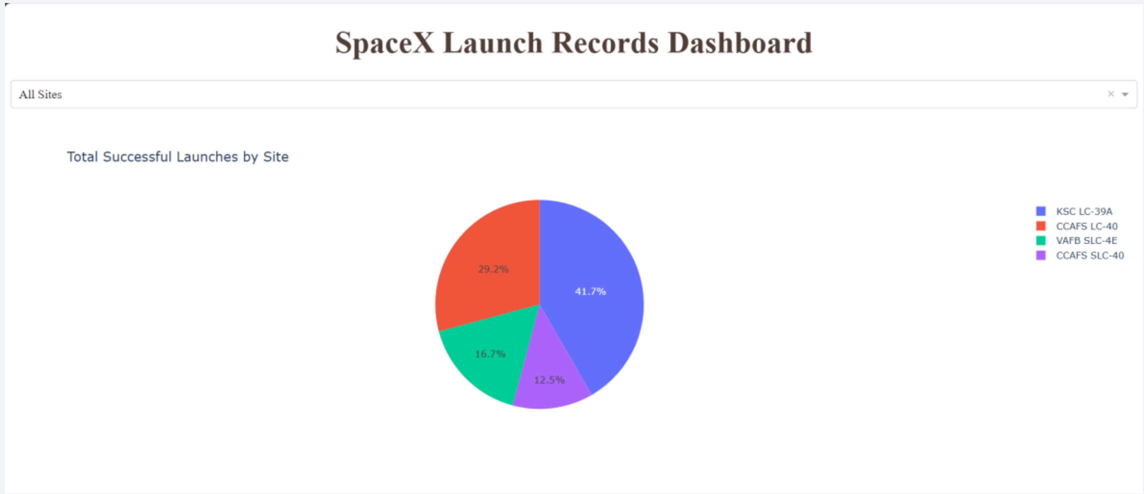
- **Model selection:** Selected best-performing model based on accuracy, robustness, and interpretability



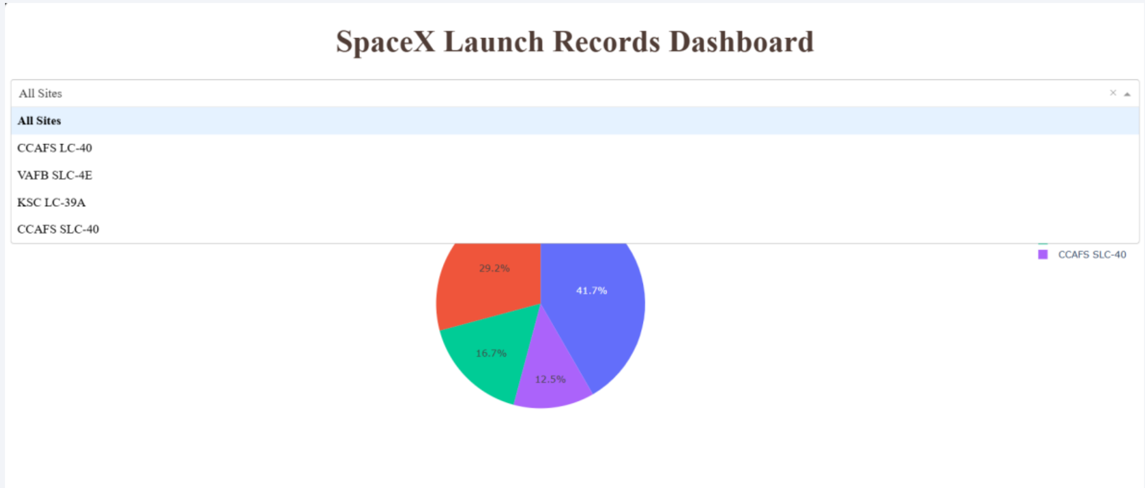
https://github.com/operiojrr/IBM-Data-Science-Track-Outputs/blob/main/Applied%20Data%20Science%20Capstone%20Project/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

A



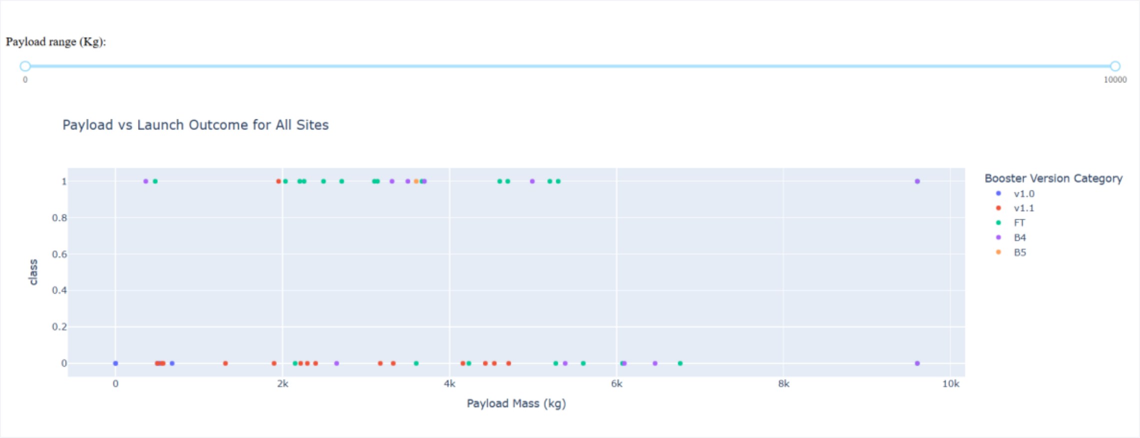
A. Pie chart to show the total successful launches count for all sites



B

B. Dropdown list to enable Launch Site selection

C



C. Slider to select payload range

Results

Exploratory Data Analysis Results

- Launch success rate increased significantly over time, reflecting SpaceX's iterative improvements in booster recovery.
- Payload mass showed a strong relationship with landing success: heavier payloads were more likely to result in controlled ocean landings rather than successful recoveries.
- Certain orbits (e.g., LEO and ISS missions) exhibited higher landing success rates compared to high-energy orbits.
- Ground pad landings achieved higher success rates than drone ship landings, indicating operational constraints in maritime recoveries.
- Launch site analysis revealed performance differences across sites, influenced by mission profiles and orbital requirements.

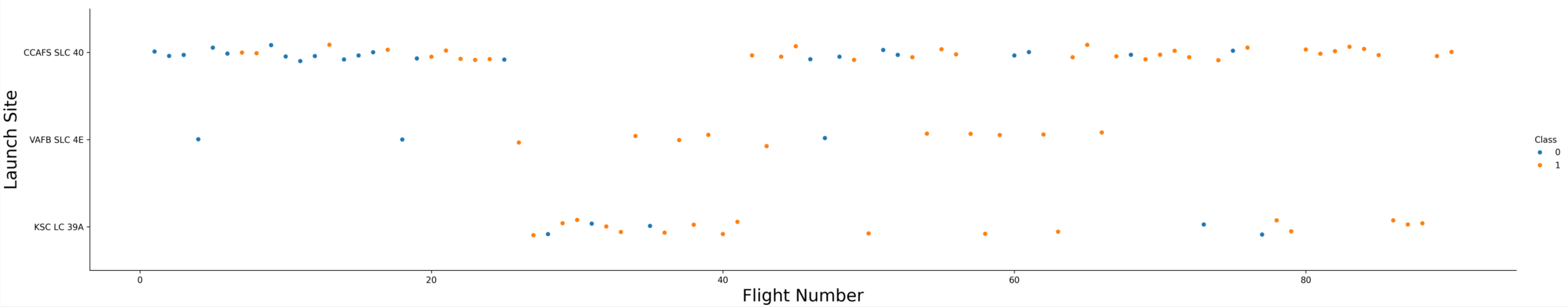
Predictive Analysis Results

- Multiple classification models—Logistic Regression, SVM, Decision Tree, and KNN—were trained and optimized using GridSearchCV.
- Logistic Regression, SVM, and KNN achieved the highest test accuracy (83.3%), while Decision Tree showed slightly lower performance (77.8%).
- Identical confusion matrices for Logistic Regression, SVM, and KNN suggest consistent decision boundaries and similar model behavior on the dataset.
- Misclassifications were primarily false positives, indicating overlap in feature space rather than model inadequacy.
- Logistic Regression was selected as the final model due to its strong performance, interpretability, and computational efficiency.

Section 2

Insights drawn from EDA

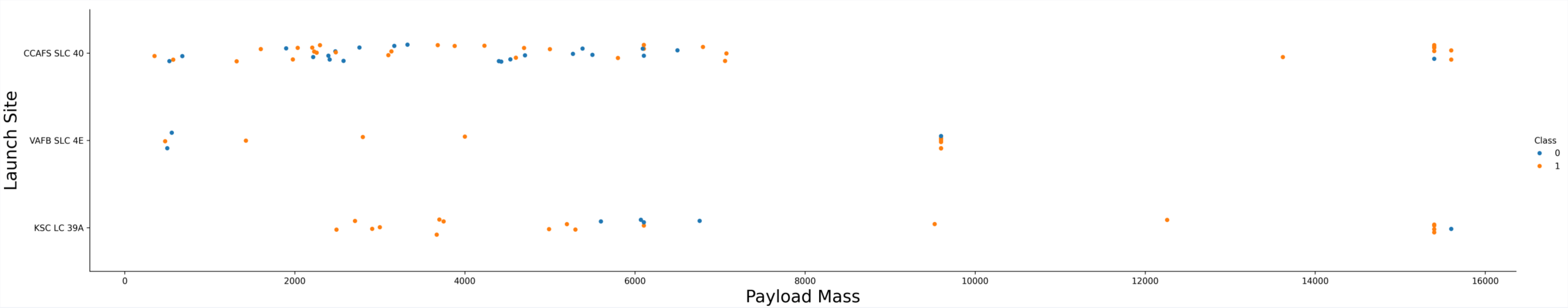
Flight Number vs. Launch Site



This scatter plot tracks the chronological progress of launches at three primary sites: CCAFS SLC 40, VAFB SLC 4E, and KSC LC 39A.

- Success Trends: Success (represented by Class 1, orange) becomes significantly more frequent as flight numbers increase, particularly after Flight 40.
- Site Usage: CCAFS SLC 40 handles the highest volume of early flights, while KSC LC 39A shows a high density of successful launches in the mid-to-later stages of the program.

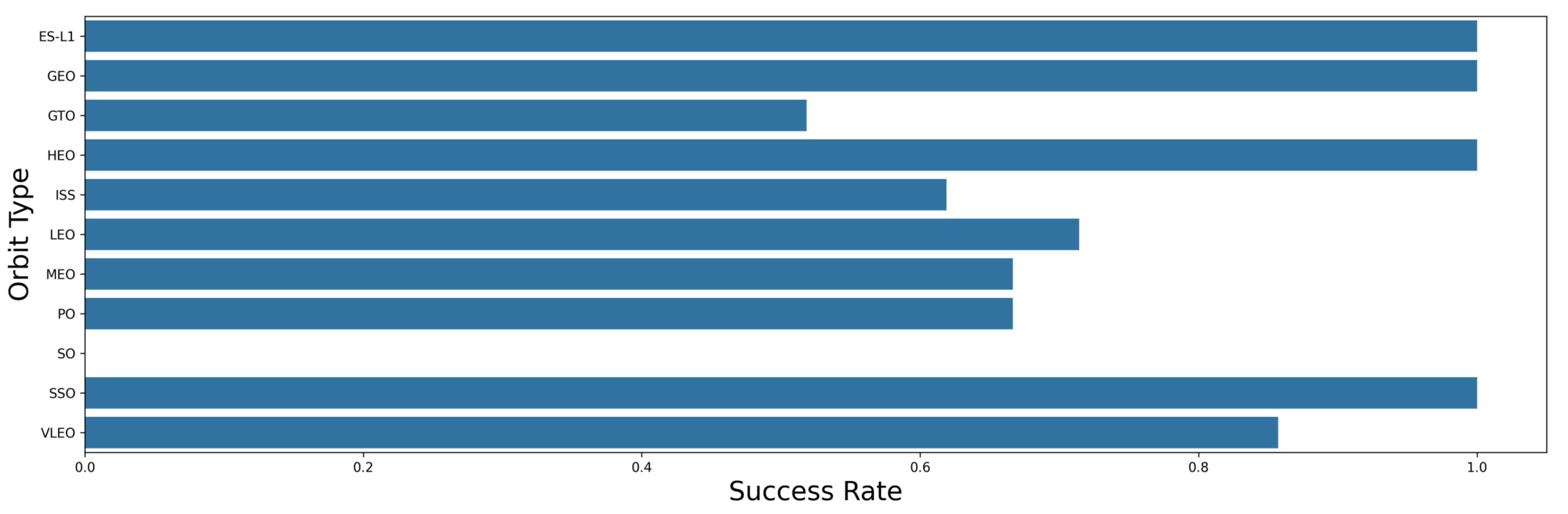
Payload vs. Launch Site



This plot illustrates how payload weight correlates with landing success at specific sites. If you observe Payload Mass Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavy payload mass (greater than 10000).

- Heavy Payloads: There is a distinct cluster of heavy payloads (approx. 15,000 kg) across all sites, most of which resulted in successful landings.
- VAFB SLC 4E: This site appears to handle a specific range of payloads, with a noticeable gap between light (under 2,000 kg) and heavy (around 10,000 kg) missions.

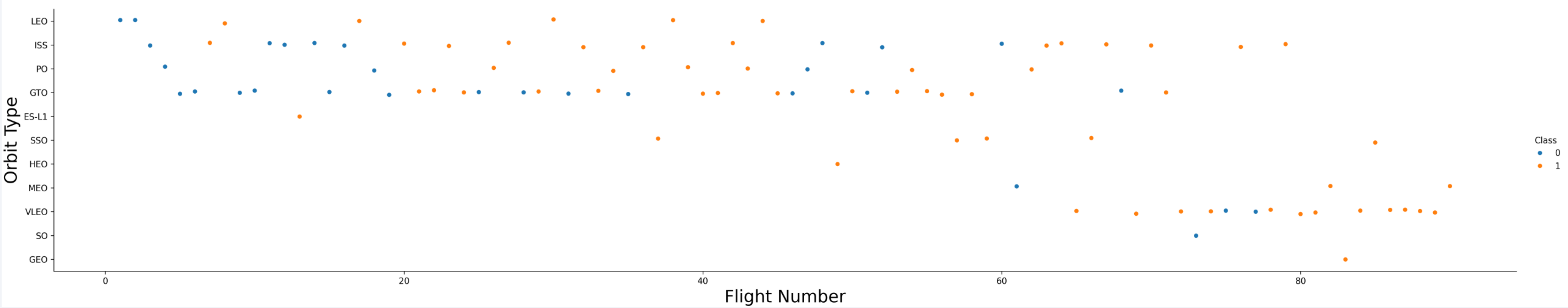
Success Rate vs. Orbit Type



This bar chart provides a clear comparison of reliability across different orbital destinations.

- Top Performers: Orbits like ES-L1, GEO, HEO, and SSO show a 100% success rate.
- Underperformers: The SO (Sun-Synchronous Orbit) shows a 0% success rate in this dataset, and GTO (Geostationary Transfer Orbit) sits at roughly 50%.

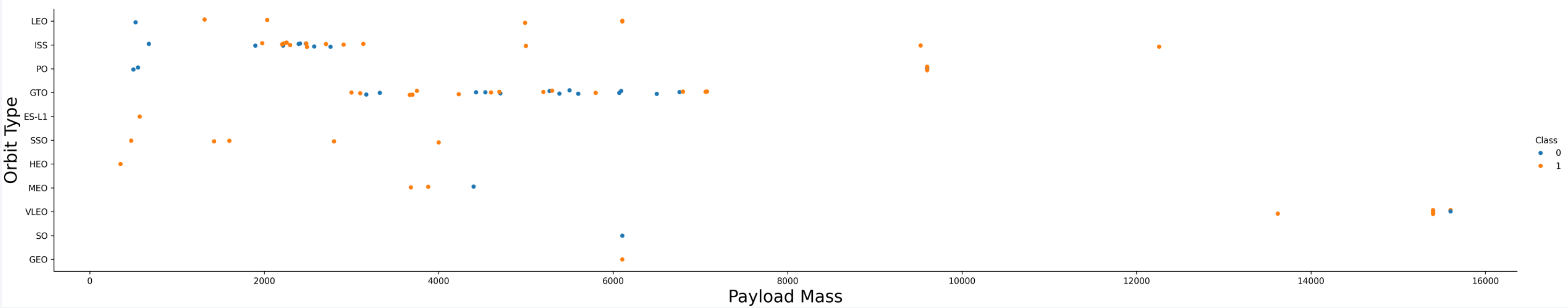
Flight Number vs. Orbit Type



This scatter plot maps the chronological transition of missions across various orbits. You can observe that in the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

- Early vs. Late: Early flights (1–20) were heavily focused on LEO, ISS, and GTO.
- Success Improvement: Similar to the launch site plot, success rates across almost all orbit types improve as flight numbers increase, suggesting a maturation of the landing technology over time.

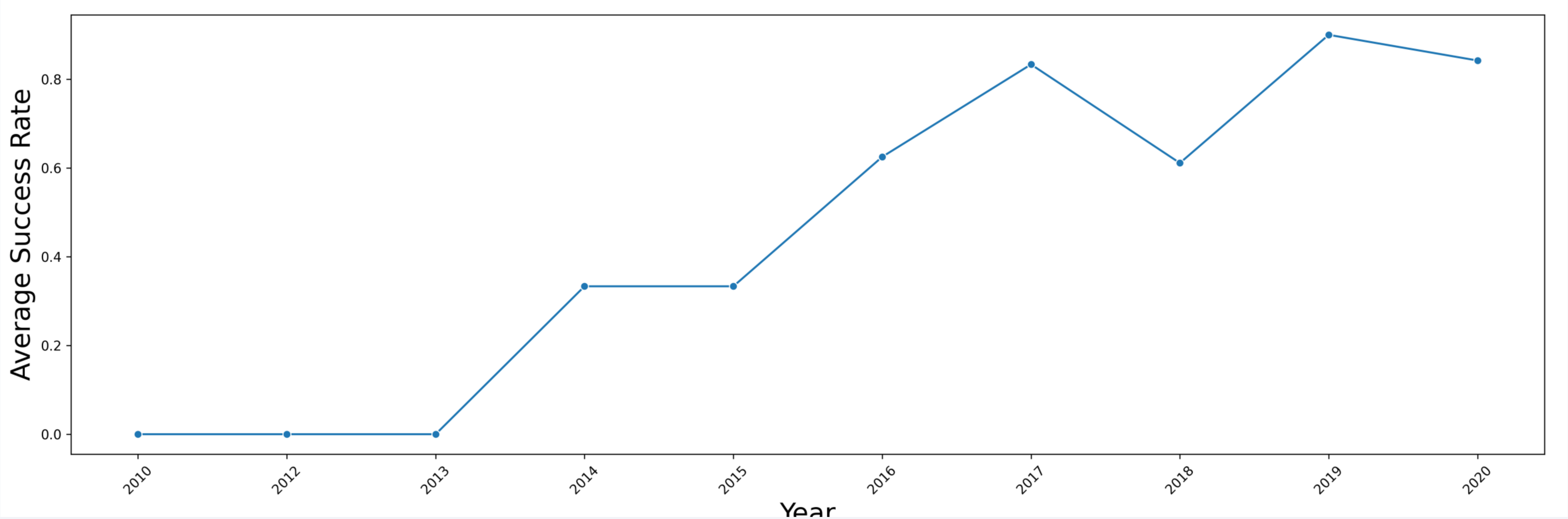
Payload vs. Orbit Type



This plot analyzes whether payload weight influences success for specific orbits. With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. However, for GTO, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

- LEO/ISS: These orbits show a mix of success and failure across lower payload masses.
- High-Mass Success: Missions to VLEO (Very Low Earth Orbit) involve some of the highest payload masses in the dataset (around 15,000 kg).

Launch Success Yearly Trend



This line graph tracks the overall performance of SpaceX launches by year (2010–2020). You can observe that the success rate since 2013 kept increasing till 2020

- Initial Phase (2010–2013): The success rate remained at 0, reflecting the experimental phase of booster recovery.
- Growth Phase: A sharp upward trajectory began in 2013, peaking in 2019 at roughly 90%.
- Dip in 2018: There is a slight drop in the success rate in 2018 before it rebounded to its highest point.

All Launch Site Names

Launch Sites
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

- The dataset contains four unique Falcon 9 launch sites
- Each site supports different orbital inclinations and mission profiles
- Launch site diversity reflects SpaceX's operational flexibility

Launch Site Names Begin with 'CCA'

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

- The query filters launch records where the launch site name begins with “CCA”, referring to Cape Canaveral Air Force Station
- Multiple launches were conducted from CCAFS LC-40, indicating its frequent operational use
- This site is primarily associated with low-Earth orbit (LEO) and ISS resupply missions

Total Payload Mass

Total Payload Mass = *45596 Kg*

- Computed the total payload mass delivered by Falcon 9 boosters for NASA Commercial Resupply Services (CRS) missions
- Used the SUM aggregation function on PAYLOAD_MASS__KG_
- Applied a customer-level filter to isolate NASA CRS launches only

Insights

- Indicates the overall contribution of Falcon 9 missions to NASA resupply operations
- Highlights SpaceX's role in delivering large cumulative payloads to orbit
- Reinforces the importance of booster reusability in reducing long-term launch costs for government missions

Average Payload Mass by F9 v1.1

Average Payload Mass = *2928.4 Kg*

- Used the AVG aggregation function to compute the mean payload mass
- Filtered records to include only Falcon 9 v1.1 launches
- Result reflects the typical payload capacity of the F9 v1.1 booster

Insights

- Provides a baseline reference for comparing payload capabilities across Falcon 9 variants
- Helps assess performance evolution of SpaceX launch vehicles

First Successful Ground Landing Date

First Successful Ground Landing Date: *December 22, 2015*

- Used the MIN() aggregation function to identify the earliest date
- Filtered launches with a successful ground pad landing outcome
- Result represents the first recorded successful ground landing by SpaceX

Insights

- Marks a key technological milestone in Falcon 9 reusability
- Demonstrates SpaceX's early success in controlled booster recovery on land

Successful Drone Ship Landing with Payload between 4000 and 6000

- Filtered launches with successful drone ship landings
- Restricted payload mass to the medium-heavy range (4000–6000 kg)
- Used DISTINCT to list only unique booster versions meeting the criteria

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

Insights

- Demonstrates that drone ship recovery is reliable even at higher payload masses
- Identifies booster variants capable of balancing payload capacity and landing success
- Supports the operational strategy of using drone ships for energy-intensive missions where ground landings are not feasible

Total Number of Successful and Failure Mission Outcomes

- Grouped all launches by Mission Outcome
- Counted the number of records for each outcome using COUNT(*)
- Provides an overall distribution of successful vs. failed missions

Insights

- The results show that successful missions significantly outnumber failures, reflecting SpaceX's high operational reliability
- The imbalance toward success supports the feasibility of booster reusability and cost reduction
- Failure records remain important for identifying edge cases and risk factors used later in predictive modeling

Mission_ Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Identified the maximum payload mass using a subquery with MAX(PAYLOAD_MASS__KG_)
- Retrieved the booster version(s) that carried this maximum payload
- Used DISTINCT to avoid duplicate booster entries

Insights

- The booster version(s) listed represent SpaceX's highest payload-capable configurations
- These boosters are typically associated with later-generation Falcon 9 variants, reflecting engineering improvements over time
- High payload capability is often linked with more demanding missions, which may influence landing success probability and recovery strategy

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

Month	Landing_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

- Filtered launch records to year 2015 using substr(Date,0,5)
- Selected only failed drone ship landings
- Extracted month, booster version, and launch site to analyze temporal and operational context

Insights

- The failures occurred during early Falcon 9 reusability attempts, when drone ship landings were still experimental
- Affected booster versions reflect earlier Falcon 9 configurations, which had lower landing reliability compared to later versions
- Certain launch sites appear more frequently, suggesting site-specific recovery challenges such as weather, trajectory, or range safety constraints
- These 2015 failures highlight the learning curve that preceded SpaceX's later high success rates in drone ship recoveries

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

- Filtered launch records between June 4, 2010 and March 20, 2017
- Grouped launches by Landing_Outcome
- Counted occurrences of each outcome and ranked them in descending order

Insights

- Failure-related outcomes dominate the early years, reflecting SpaceX's initial development and testing phase for reusable boosters
- Ground pad successes appear earlier and more consistently than drone ship successes, indicating that land-based recovery was operationally simpler in the early stages
- Drone ship successes increase closer to 2017, signaling significant improvements in guidance, control, and landing precision
- The ranking clearly illustrates SpaceX's progression from experimental landings to reliable recovery, supporting the trend observed in the EDA and predictive modeling stages

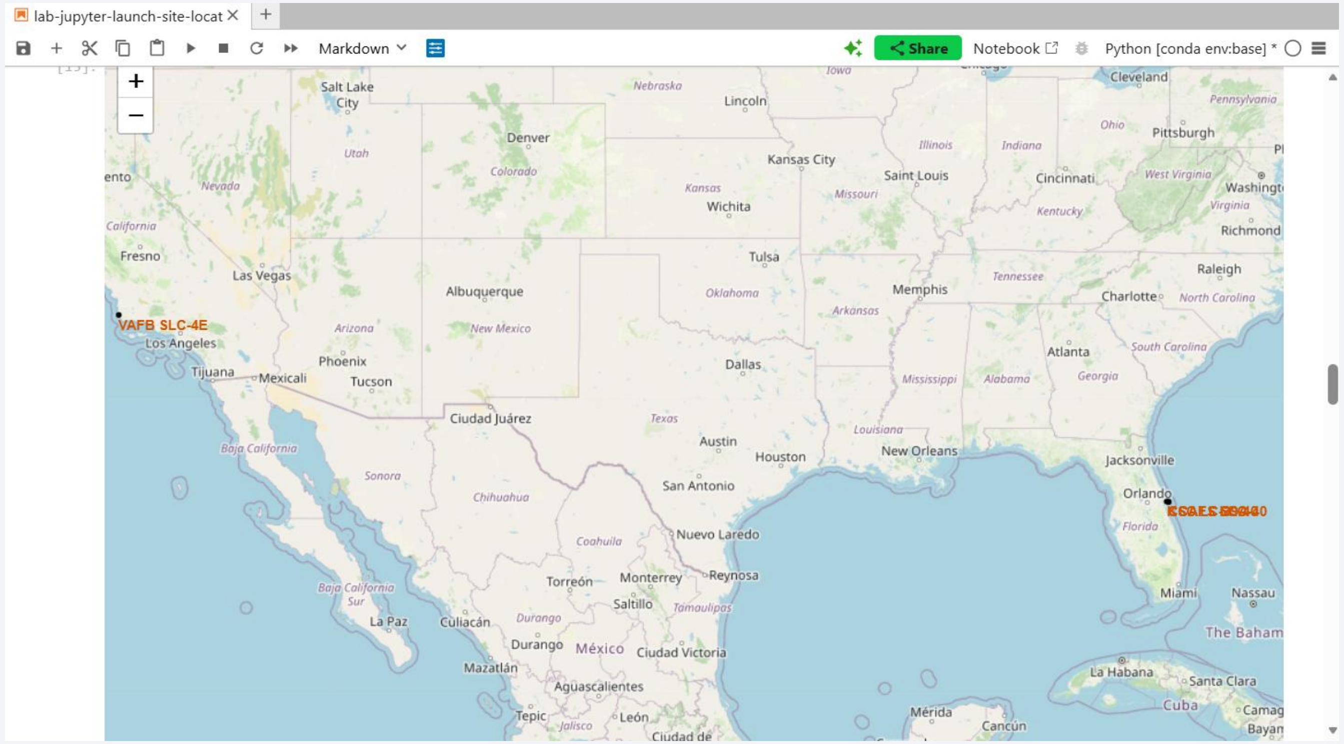
Landing Outcome	Outcome Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Geographical Distribution and Strategic Positioning of SpaceX Launch Sites

- Most major SpaceX launch sites are located in the southern regions of the United States to be as close to the Equator as possible.
- Sites like CCAFS SLC-40 and KSC LC-39A in Florida are at approximately 28° N latitude. While not at the Equator, they are in the southernmost part of the continental US to maximize the "boost" from the Earth's rotational speed, which is highest at the Equator. VAFB SLC-4E in California is further north at approximately 34° N.
- All three primary launch sites are located directly on the coastline.
- VAFB SLC-4E is on the Pacific coast of California while the other three are on the Atlantic coast of Florida. Coastal locations allow rockets to launch over open water, ensuring that spent stages or debris from a failed launch fall into the ocean rather than populated areas.



Visualizing Mission Outcomes: Success Rates Across Primary Launch Sites

- CCAFS LC-40 has the lowest success rate, with a significantly higher number of failed landings (red markers), particularly in early missions, as shown in the screenshot.

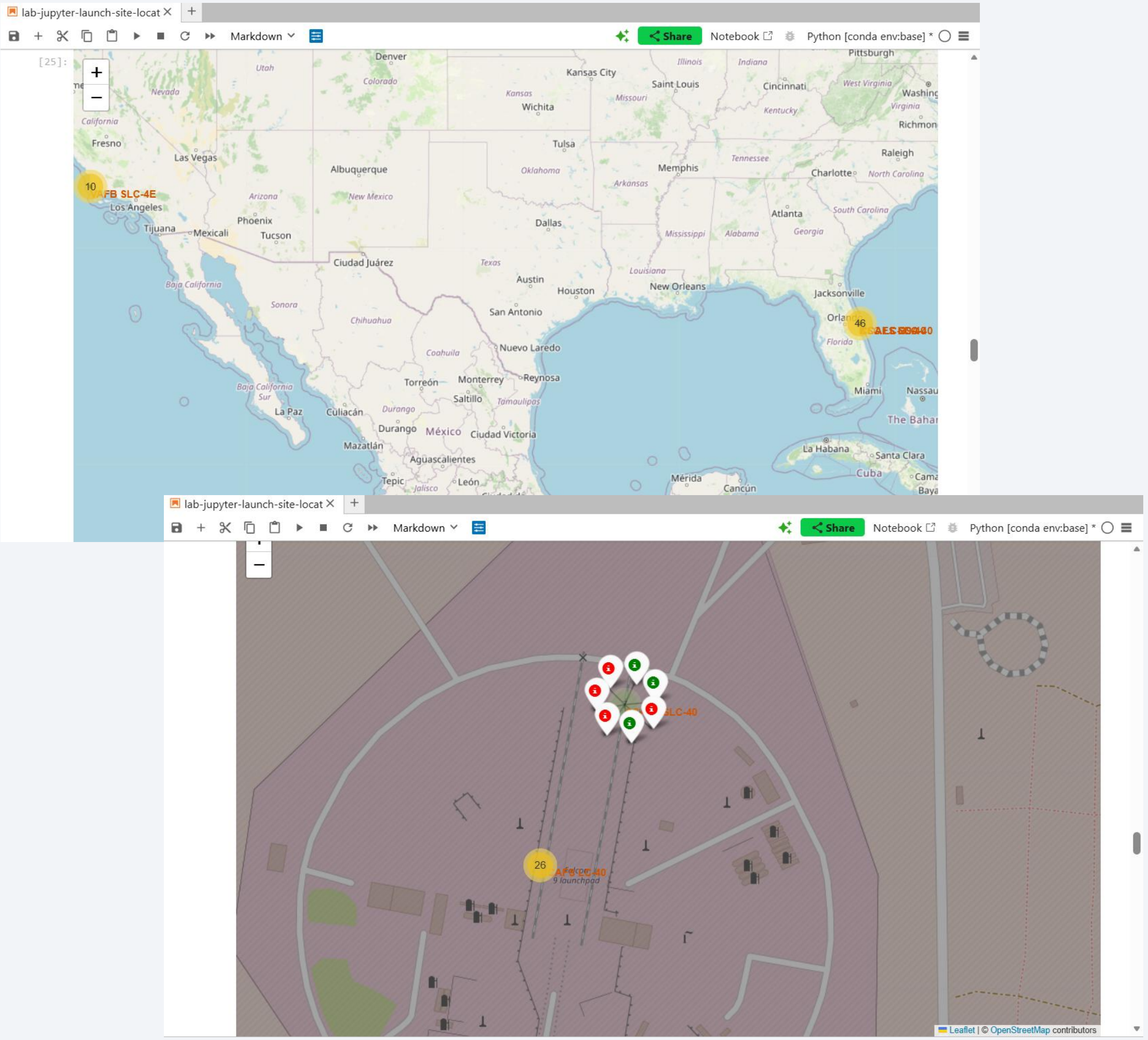
- KSC LC-39A shows the highest success rate, with a clear majority of launches resulting in successful landings (green markers dominate).

- CCAFS SLC-40 also demonstrates a relatively high success rate, especially in later launches, indicating operational improvement over time.

- VAFB SLC-4E exhibits a more balanced mix of successes and failures, resulting in a moderate overall success rate compared to Florida-based sites.

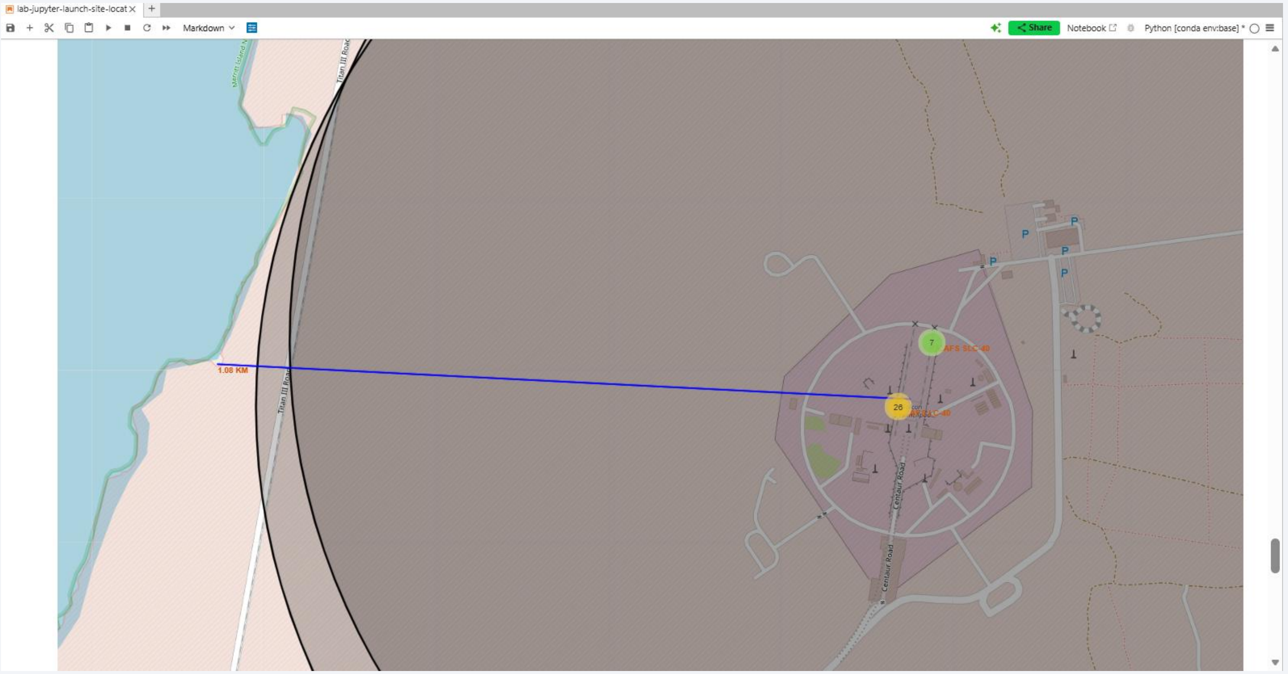
- Overall, Florida launch sites (KSC LC-39A and CCAFS SLC-40) outperform Vandenberg (VAFB SLC-4E) in landing success, suggesting advantages related to mission profile, recovery strategy, or operational maturity.

*for the remaining launch sites, please refer to the interactive map included in the notebook for this activity, accessible via the GitHub link provided in the previous section.



Launch Site Proximity

- The launch sites are intentionally placed as close to the coastline as possible. Such an example is the CCAFS LC 40, located just 1.08 km away from the coastline.
- Launch sites are located near railways to facilitate the delivery of heavy equipment and rocket stages that are too large for standard road transport. The closest railway to CCAFS LC 40 is just 1.69 km away.
- Close access to major highways (such as Florida State Road A1A or CA-1) is essential for the daily commute of thousands of employees and the logistics of smaller hardware components. The closest highway to CCAFS LC 40 is just 2.15 km away.
- Launch sites maintain a significant "buffer zone" from major population centers. This safety margin protects residents from the acoustic impact (noise/vibration) of launches and mitigates risks in the event of a flight anomaly or explosion. Although the closest city is only 1.08 km away, the nearest neighborhood is much farther away.

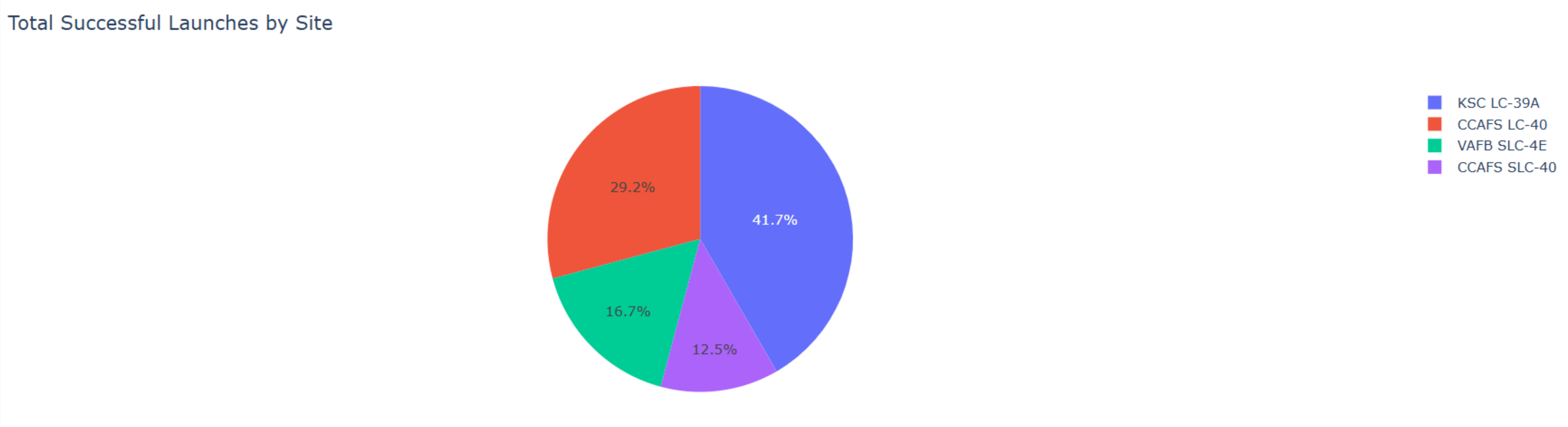


*Additional interactive maps are available in the notebook for this part of the project uploaded at GitHub (the link is given at the previous section.

Section 4

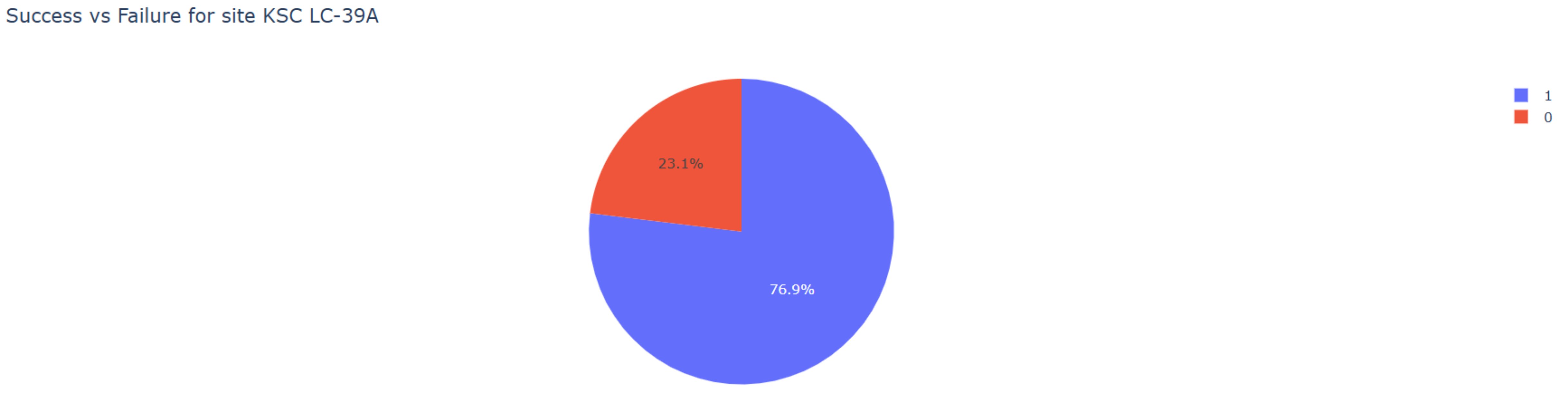
Build a Dashboard with Plotly Dash

Aggregate Mission Outcomes Across SpaceX Launch Facilities



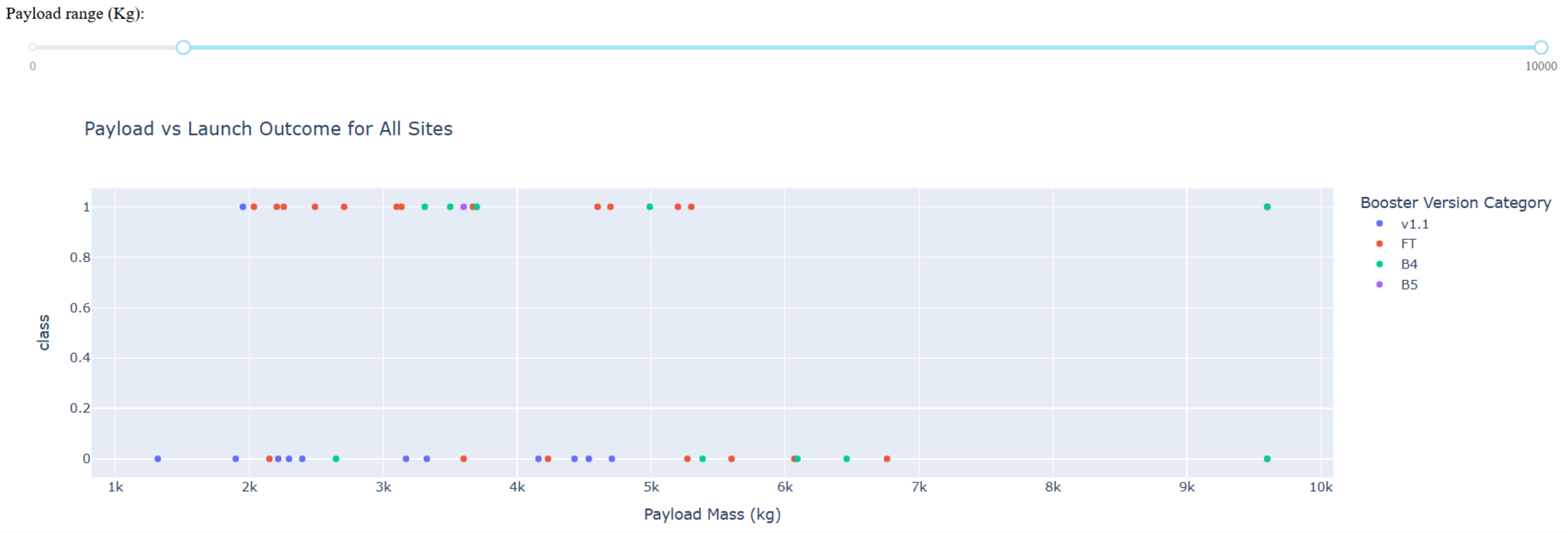
This pie chart breaks down the total volume of successful launches (Class 1) contributed by each site. KSC LC-39A typically accounts for the largest "slice" of total successes in these datasets, reflecting its role as a primary high-cadence site for SpaceX. CCAFS SLC-40 often shows the highest total mission count, but because it handled many early development flights, its share of the total "success pie" may be lower proportionally compared to the more modern operations at KSC.

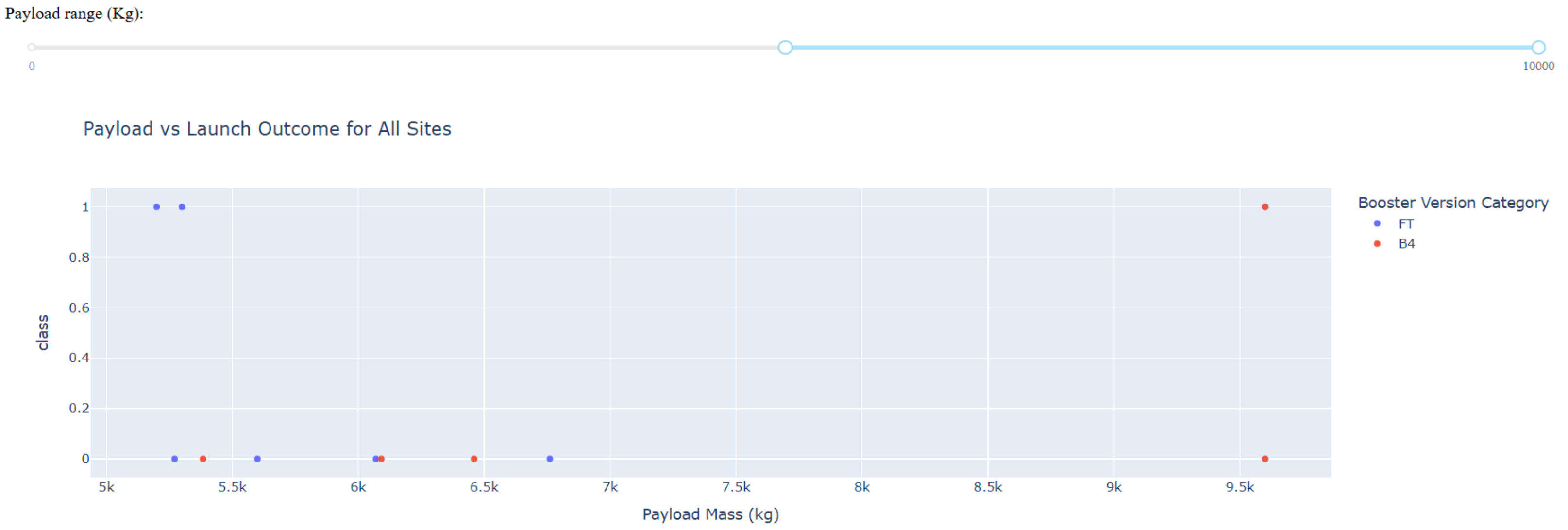
Top Performance Leader: Success vs. Failure Ratio at KSC LC-39A



This chart focuses exclusively on the site with the best performance (KSC LC-39A), showing a direct comparison between successful (Class 1) and failed (Class 0) outcomes. For this specific site, the chart is dominated by success. Data indicates that KSC LC-39A has achieved success rates as high as 77% depending on the specific flight window analyzed, significantly outperforming earlier sites like SLC-40.

Correlation Between Payload Mass, Booster Version, and Mission Success





These scatter plots use a range slider to filter missions by payload mass (kg) while color-coding results by Booster Version (e.g., v1.1, FT, B4, B5).

- The Block 5 (B5) booster version has the highest success rate.
- Success is consistently high for heavy payloads particularly for missions to Very Low Earth Orbit (VLEO) using the Block 5 boosters.
- Earlier versions like v1.1 show a higher concentration of "Class 0" failures, primarily at lower payload ranges representing the developmental phase of booster landing technology.

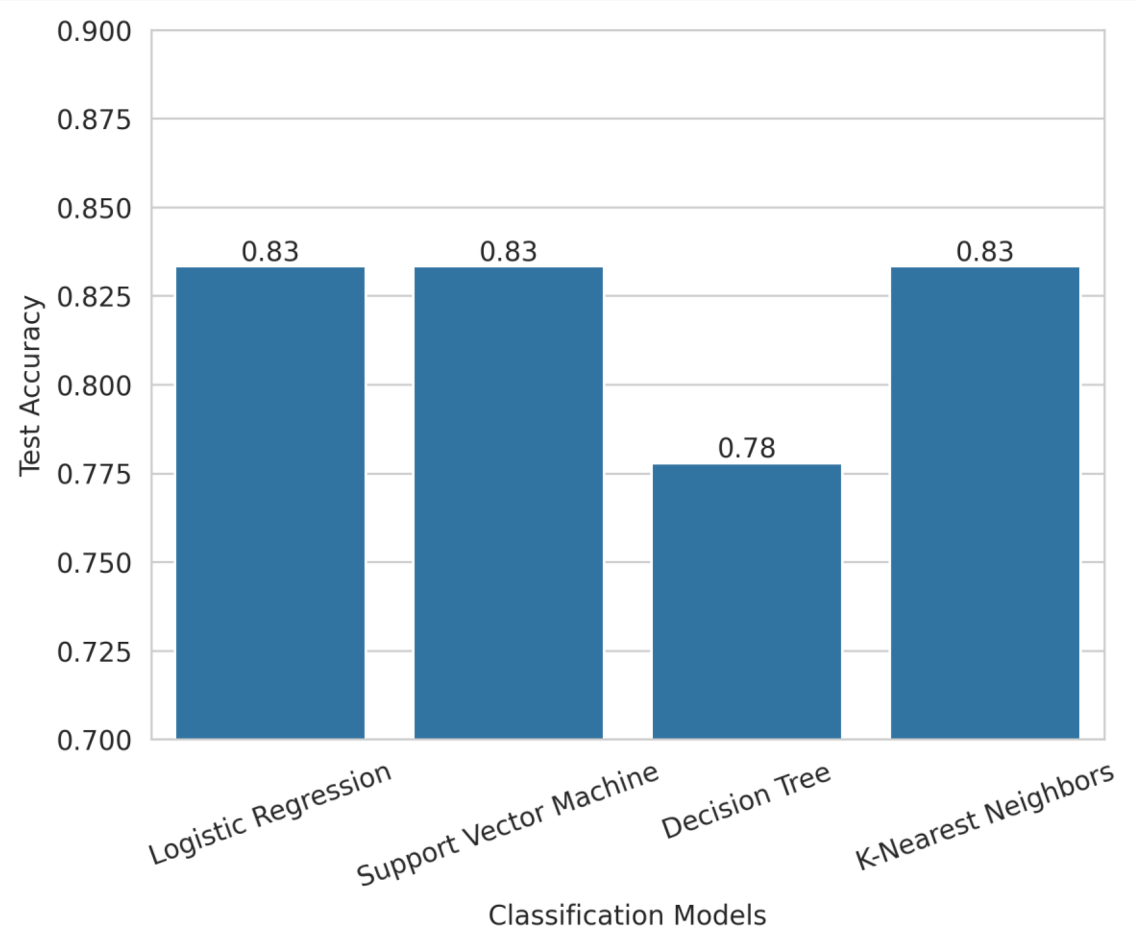
Section 5

Predictive Analysis (Classification)

Classification Accuracy

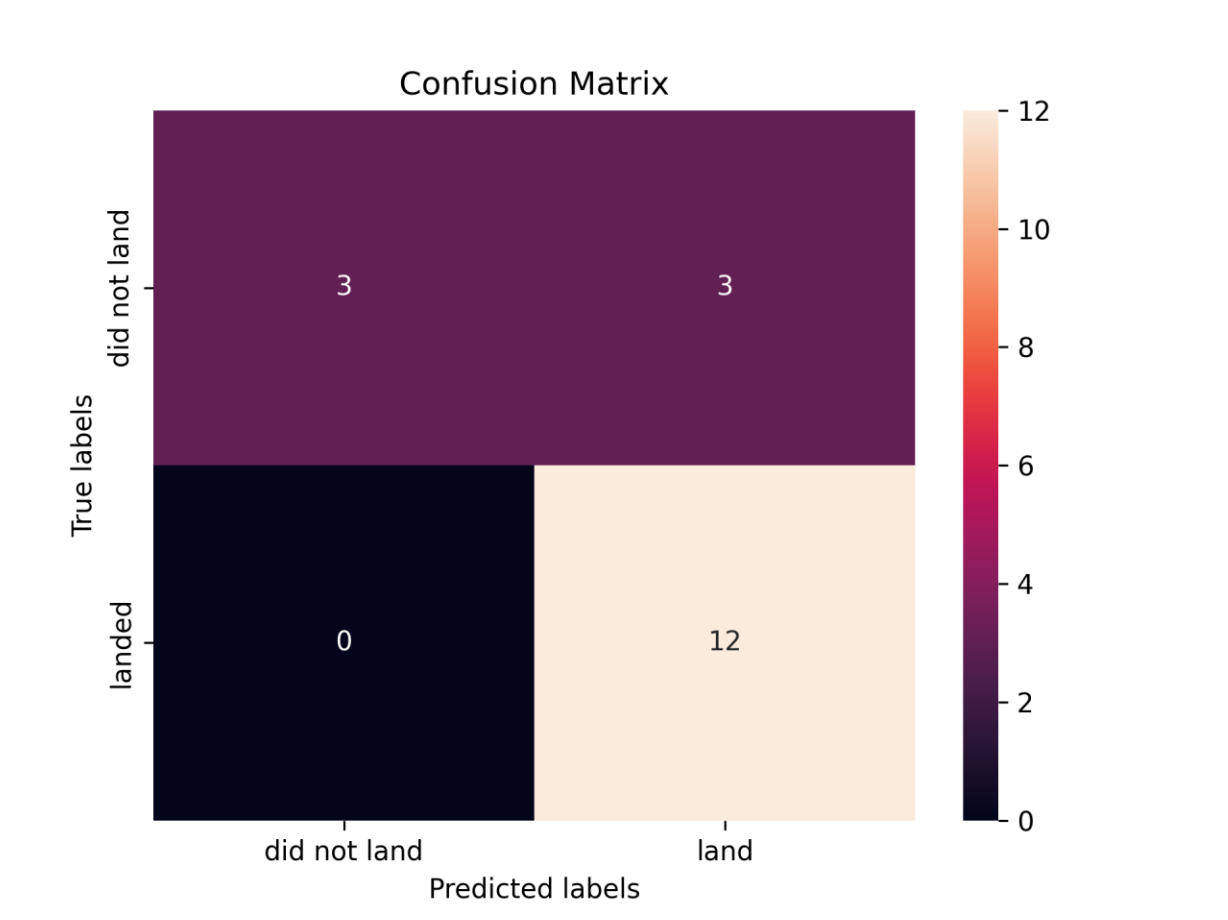
- Logistic Regression, SVM, and KNN achieved the highest accuracy (83.3%); The Decision Tree model only scored 77.8% accuracy.
- Logistic Regression was selected as the final model due to:

1. Comparable accuracy
2. Simplicity and interpretability
3. Lower computational complexity



Confusion Matrix

- 12 successful landings were correctly predicted as landed (True Positives).
- 3 failed landings were correctly predicted as failures (True Negatives).
- 3 failed landings were misclassified as successful (False Positives).
- No successful landings were misclassified as failures (False Negatives = 0).
- The models demonstrate excellent recall for successful landings, meaning they consistently identify successful outcomes.
- There is some confusion in predicting failures, where a few failed landings are predicted as successful.



Conclusions

- Launch success is strongly influenced by launch site location, payload mass, and booster version.
- KSC LC-39A and VAFB SLC-4E show higher success rates compared to other launch sites.
- Classification models (LogReg, SVM, and KNN) achieved high accuracy, with excellent detection of successful landings.
- Confusion matrix results show zero missed successful landings, indicating strong recall for landing success.
- Minor misclassification occurs for failed landings, suggesting room for improvement in failure prediction.

Appendix

882b8ce8773a","id":"62a9f89a20413d2695d8871a"},{"fairings":{"reused":null,"recovery_attempt":null,"recovered":null,"ships":
[]},"links":{"patch":{"small":"https://images2.imgbox.com/a9/9a/NXVkTZCE_o.png","large":"https://images2.imgbox.com/e3/cc/hN9
6PmST_o.png"},"reddit":{"campaign":"https://www.reddit.com/r/spacex/comments/jhu37i/starlink_general_discussion_and_deploymen
t_thread/","launch":"https://www.reddit.com/r/spacex/comments/xd8vhj/rspacex_starlink_434_launch_discussion_and/","media":nul
l,"recovery":"https://www.reddit.com/r/spacex/comments/k2ts1q/rspacex_fleet_updates_discussion_thread/"},"flickr":{"small":
[],"original":[]},"presskit":null,"webcast":"https://youtu.be/ZlQHF_yBkMQ","youtube_id":"ZlQHF_yBkMQ","article":null,"wikiped
ia":null},"static_fire_date_utc":null,"static_fire_date_unix":null,"net":false,"window":null,"rocket":"5e9d0d95eda69973a809d1
ec","success":true,"failures":[],"details":null,"crew":[],"ships":[],"capsules":[],"payloads":["63161699ffc78f3b85670719"],"l
aunchpad":"5e9e4501f509094ba4566f84","flight_number":185,"name":"Starlink 4-34 (v1.5)","date_utc":"2022-09-17T01:05:00.000
Z","date_unix":1663376700,"date_local":"2022-09-16T21:05:00-04:00","date_precision":"hour","upcoming":false,"cores":[{"cor
e":"60b800111f83cc1e59f16438","flight":6,"gridfins":true,"legs":true,"reused":true,"landing_attempt":true,"landing_success":t
rue,"landing_type":"ASDS","landpad":"5e9e3033383ecbb9e534e7cc"}],"auto_update":true,"tbd":false,"launch_library_id":"9ba04064
-c329-40bf-b477-ff468d7d8058","id":"63161329ffc78f3b8567070b"},{"fairings":{"reused":null,"recovery_attempt":null,"recovere
d":null,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/a9/9a/NXVkTZCE_o.png","large":"https://images2.imgb
ox.com/e3/cc/hN96PmST_o.png"},"reddit":{"campaign":"https://www.reddit.com/r/spacex/comments/jhu37i/starlink_general_discussi
on_and_deployment_thread/","launch":"https://www.reddit.com/r/spacex/comments/xn028t/rspacex_starlink_435_launch_discussion_a
nd/","media":null,"recovery":"https://www.reddit.com/r/spacex/comments/k2ts1q/rspacex_fleet_updates_discussion_thread/"},"fli
ckr":{"small":[],"original":[]},"presskit":null,"webcast":"https://youtu.be/VVu2bSJJhgI","youtube_id":"VVu2bSJJhgI","articl
e":null,"wikipedia":null},"static_fire_date_utc":null,"static_fire_date_unix":null,"net":false,"window":null,"rocket":"5e9d0d
95eda69973a809d1ec","success":true,"failures":[],"details":null,"crew":[],"ships":[],"capsules":[],"payloads":["631616a7ffc78
f3b8567071a"],"launchpad":"5e9e4501f509094ba4566f84","flight_number":186,"name":"Starlink 4-35 (v1.5)","date_utc":"2022-09-24
T23:30:00.000Z","date_unix":1664062200,"date_local":"2022-09-24T19:30:00-04:00","date_precision":"hour","upcoming":false,"cor
es":[{"core":"627843d657b51b752c5c5a53","flight":4,"gridfins":true,"legs":true,"reused":true,"landing_attempt":true,"landing_
success":true,"landing_type":"ASDS","landpad":"5e9e3033383ecbb9e534e7cc"}],"auto_update":true,"tbd":false,"launch_library_i
d":"1c903b65-6667-4fd5-944d-296c5f13e01f","id":"63161339ffc78f3b8567070c"},{"fairings":null,"links":{"patch":{"small":"http
s://images2.imgbox.com/eb/d8/D1Yywp0w_o.png","large":"https://images2.imgbox.com/33/2e/k6VE4iYl_o.png"},"reddit":{"campaign":
null,"launch":"https://www.reddit.com/r/spacex/comments/xvm76j/rspacex_crew5_launchcoast_docking_discussion_and/","media":nul
l,"recovery":null},"flickr":{"small":[],"original":[]},"presskit":null,"webcast":"https://youtu.be/5EwW8ZkArL4","youtube_i
d":"5EwW8ZkArL4","article":null,"wikipedia":"https://en.wikipedia.org/wiki/SpaceX_Crew-5"},"static_fire_date_utc":null,"stati
c_fire_date_unix":null,"net":false,"window":null,"rocket":"5e9d0d95eda69973a809d1ec","success":true,"failures":[],"details":n
ull,"crew":["62dd7196202306255024d13c","62dd71c9202306255024d13d","62dd7210202306255024d13e","62dd7253202306255024d13f"],"shi
ps":[],"capsules":["617c05591bad2c661a6e2909"],"payloads":["62dd73ed202306255024d145"],"launchpad":"5e9e4502f509094188566f8
8","flight_number":187,"name":"Crew-5","date_utc":"2022-10-05T16:00:00.000Z","date_unix":1664985600,"date_local":"2022-10-05T
12:00:00-04:00","date_precision":"hour","upcoming":false,"cores":[{"core":"633d9da635a71d1d9c66797b","flight":1,"gridfins":tr
ue,"legs":true,"reused":false,"landing_attempt":true,"landing_success":true,"landing_type":"ASDS","landpad":"5e9e3033383ecbb9
e534e7cc"}],"auto_update":true,"tbd":false,"launch_library_id":"f33d5ece-e825-4cd8-809f-1d4c72a2e0d3","id":"62dd70d5202306255

A snippet of the JSON file containing rocket launch data from SpaceX API

Appendix

```
<table class="wikitable plainrowheaders collapsible" style="width: 100%;">
<tbody><tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="List of Falcon 9 first-stage boosters">Version,<
br/>Booster</a> <sup class="reference" id="cite_ref-booster_11-0"><a href="#cite_note-booster-11"><span class="cite-bracket">
[</span>b<span class="cite-bracket">]</span></a></sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-Dragon_12-0"><a href="#cite_note-Dragon-12"><span class="cite-brac
ket">[</span>c<span class="cite-bracket">]</span></a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falcon 9 first-stage landing tests">Booster<br/>lan
ding</a>
</th></tr>
<tr>
<th rowspan="2" scope="row" style="text-align:center;">1
</th>
<td>4 June 2010,<br/>18:45
</td>
<td><a href="/wiki/Falcon_9_v1.0" title="Falcon 9 v1.0">F9 v1.0</a><sup class="reference" id="cite_ref-MuskMay2012_13-0"><a h
ref="#cite_note-MuskMay2012-13"><span class="cite-bracket">[</span>7<span class="cite-bracket">]</span></a></sup><br/>B0003.1
<sup class="reference" id="cite_ref-block_numbers_14-0"><a href="#cite_note-block_numbers-14"><span class="cite-bracket">[</s
pan>8<span class="cite-bracket">]</span></a></sup>
</td>
```

A snippet of the HTML table produced during web scraping via BeautifulSoup

Appendix

Out[15]:

s	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
2	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
0	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
0	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
0	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Parts of dataframe showing created landing outcome labels during data wrangling

Appendix

```
%sql SELECT substr(Date,6,2) AS Month, \
            "Landing_Outcome", \
            "Booster_Version", \
            "Launch_Site" \
     FROM SPACEXTABLE \
     WHERE substr(Date,0,5) = '2015' \
       AND "Landing_Outcome" = 'Failure (drone ship)';
```

Month	Landing_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

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

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

Some SQL commands and their appropriate response used for EDA

Appendix

```python
# Create an app layout
app.layout = html.Div(children=[html.H1('SpaceX Launch Records Dashboard',
                                        style={'textAlign': 'center', 'color': '#503D36',
                                               'font-size': 40}),
                                # TASK 1: Add a dropdown list to enable Launch Site selection
                                # The default select value is for ALL sites
                                dcc.Dropdown(id='site-dropdown',
                                             options=[
                                                 {'label': 'All Sites', 'value': 'ALL'}
                                             ] + [
                                                 {'label': site, 'value': site}
                                                 for site in spacex_df['Launch Site'].unique()
                                             ],
                                             value='ALL',
                                             placeholder='Select a Launch Site here',
                                             searchable=True
                                             ),
                                html.Br(),

                                # TASK 2: Add a pie chart to show the total successful launches count for all sites
                                # If a specific launch site was selected, show the Success vs. Failed counts for the site
                                html.Div(dcc.Graph(id='success-pie-chart')),
                                html.Br(),

                                html.P("Payload range (Kg):"),
                                # TASK 3: Add a slider to select payload range
                                dcc.RangeSlider(id='payload-slider',
                                                min=0,
                                                max=10000,
                                                step=1000,
```

```python
Add a callback function for `site-dropdown` as input, `success-pie-chart` as output
pp.callback(
  Output(component_id='success-pie-chart', component_property='figure'),
  Input(component_id='site-dropdown', component_property='value')
  )
f get_pie_chart(entered_site):

  if entered_site == 'ALL':
      # Count total successful launches per site
      success_df = spacex_df[spacex_df['class'] == 1]
      fig = px.pie(
          success_df,
          names='Launch Site',
          title='Total Successful Launches by Site'
      )
      return fig
  else:
      # Filter data for the selected site
      site_df = spacex_df[spacex_df['Launch Site'] == entered_site]
      outcome_counts = site_df['class'].value_counts().reset_index()
      outcome_counts.columns = ['class', 'count']

      fig = px.pie(
          outcome_counts,
          values='count',
          names='class',
          title=f'Success vs Failure for site {entered_site}'
      )
      return fig
```

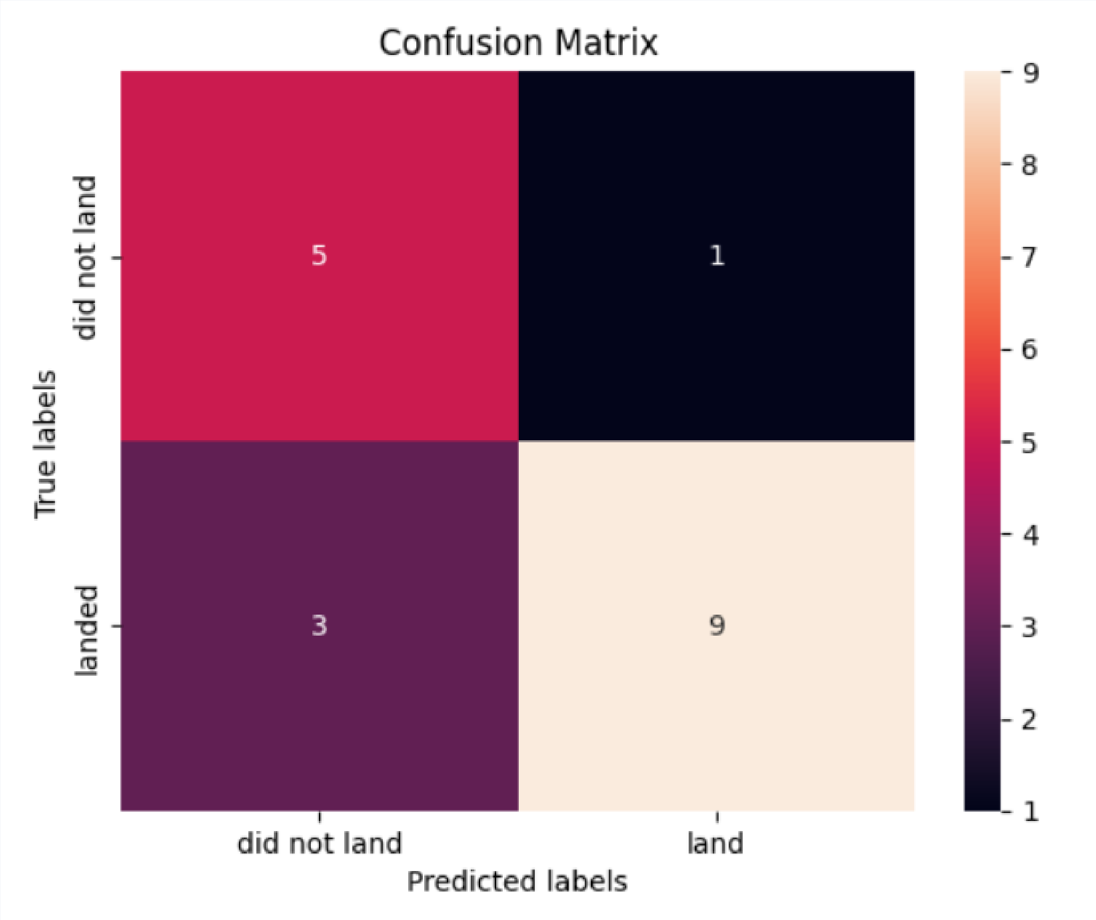
```python
@app.callback(
    Output(component_id='success-payload-scatter-chart', component_property='figure
    [
        Input(component_id='site-dropdown', component_property='value'),
        Input(component_id='payload-slider', component_property='value')
    ]
)
def update_scatter_plot(selected_site, payload_range):

    # Filter by payload range first
    low, high = payload_range
    filtered_df = spacex_df[
        (spacex_df['Payload Mass (kg)'] >= low) &
        (spacex_df['Payload Mass (kg)'] <= high)
    ]

    # If ALL sites are selected
    if selected_site == 'ALL':
        fig = px.scatter(
            filtered_df,
            x='Payload Mass (kg)',
            y='class',
            color='Booster Version Category',
            title='Payload vs Launch Outcome for All Sites'
        )
        return fig
```

Snippets of the code used to build the SpaceX dashboard using Plotly and Dash

Appendix



The confusion matrix of the Decision Tree model, which garnered the lowert test accuracy among the models

Thank you!