IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kimberley Usher

24 February 2026

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- This project uses SpaceX launch data to predict Falcon 9 first-stage landing success - a key driver of the company's cost advantage.

- Using data collected from the SpaceX API and Wikipedia, four classification models were built and compared.

- The **Decision Tree model** achieved 94.4% test accuracy, correctly identifying all 12 successful landings in the test set

Introduction

- SpaceX offers Falcon 9 launches at approximately $62M, significantly below competitors charging $165M or more, largely due to its ability to recover and reuse the first-stage booster.
- Landing success is therefore a critical cost driver and a key determinant of launch economics.
- This project leverages publicly available SpaceX launch data to model first-stage landing outcomes as a binary classification problem.
- Specifically, it evaluates whether launch parameters such as payload mass, orbit type, launch site, and booster version can reliably predict landing success, and compares multiple machine learning algorithms to determine which achieves the strongest classification performance.
- The sections that follow walk through how that data was collected, explored, and ultimately used to answer this question.

Section 1

Methodology

Methodology

Launch data was obtained from two primary sources: the SpaceX REST API (v4) for structured mission records and Wikipedia via web scraping for historical Falcon 9 launch tables The datasets were cleaned and merged into a consolidated dataset comprising 90 Falcon 9 launches. The project followed a four-stage pipeline, each building on the last.

Data wrangling

- Missing values in the PayloadMass variable were imputed using the column mean.
- Falcon launches were excluded to maintain model consistency.
- A binary target variable, Class, was engineered where 1 represents a successful landing and 0 an unsuccessful outcome.

Exploratory data analysis (EDA) using visualization and SQL

- Python visualizations, including scatter plots and bar charts, were used to examine relationships between payload mass, orbit type, launch site, and landing outcome.
- SQL queries were applied to extract summary statistics and validate distributional patterns within the dataset.

Interactive visual analytics using Folium and Plotly Dash

- Folium interactive maps visualized launch site geography, outcome distribution, and proximity to key infrastructure.
- A Plotly Dash dashboard enabled dynamic filtering by launch site and payload range to support interactive pattern exploration.

Predictive analysis using classification models

- Four supervised classification algorithms were implemented: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN).
- Models were trained on labeled data and optimized using GridSearchCV with 10-fold cross-validation for hyperparameter tuning.
- Performance was evaluated on a held-out test set.
- The Decision Tree model achieved the highest test accuracy (94.44%), correctly predicting all 12 successful landings with one misclassification.

Data Collection

Data underpins everything that follows. Two sources were used to build the most complete picture of SpaceX launches possible:

1. The SpaceX REST API (v4) provided structured launch records including rocket type, payload, launch site, and landing outcome.
2. Wikipedia provided historical Falcon 9 launch tables collected via web scraping.

Both datasets were merged and cleaned to produce a unified dataset of 90 Falcon 9 launches used for analysis and modeling.

Data collection process

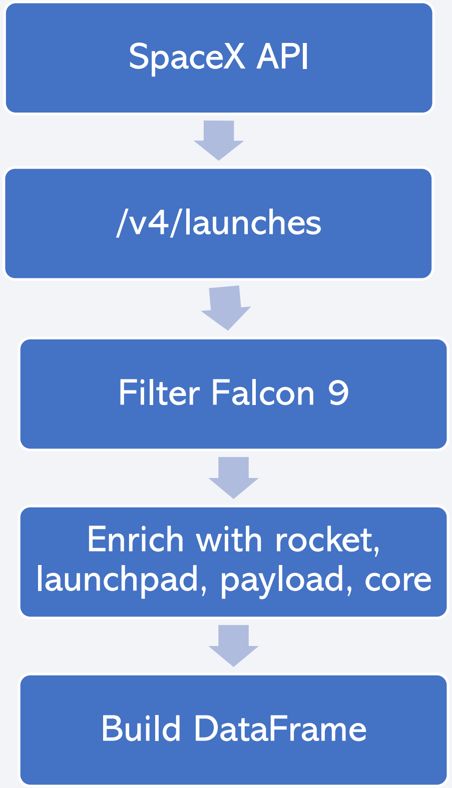
SpaceX API → Web Scraping → Merge & Clean → Final Dataset

Key fields collected:

- Flight Number,
- Date,
- Booster Version,
- Payload Mass (kg),
- Orbit,
- Launch Site,
- Landing Outcome,
- Mission Outcome.

Data Collection - SpaceX API

- Step 1: GET /v4/launches - retrieved all Falcon 9 launch records.
- Step 2: Used launch IDs to call helper endpoints: /v4/rockets/{id} for booster name, /v4/launchpads/{id} for site name and coordinates, /v4/payloads/{id} for mass and orbit, /v4/cores/{id} for landing type and outcome.
- Step 3: Compiled all fields into a structured Pandas DataFrame with 90 rows and 18 columns.
- https://github.com/kimannu/kimcapstone/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb



Data Collection - Scraping

Step 1: Identified Wikipedia's Falcon 9 launch history table as the scraping target.

Step 2: Used Python requests and BeautifulSoup to parse the HTML table.

Step 3: Extracted columns including Flight No., Launch Site, Payload, Orbit, Customer, Launch Outcome, and Booster Landing result.

Step 4: Cleaned and exported data to CSV for merging with the API dataset.

https://github.com/kimannu/kimcapstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Filtered dataset to include only Falcon 9 launches (excluding Falcon 1).
- Handled missing PayloadMass values by replacing with column mean.
- Created a binary target variable ‘Class’: 1 = successful landing, 0 = unsuccessful.
- Verified no duplicate records and confirmed final dataset of 90 launches.
- With a clean, consistent dataset of 90 Falcon 9 launches and a clearly defined target variable, exploratory analysis could begin

Raw API Data → → Filter Falcon 9 → Fill missing PayloadMass (mean) → Encode landing outcomes → Create Class column (0/1) → Final dataset: 90 rows, 18 features

https://github.com/kimannu/kimcapstone/blob/main/labs-jupyter-spacex-Data%20wrangling%20.ipynb

EDA with Data Visualization

Before modelling, exploratory analysis was used to understand which features were likely to influence landing outcome using the following charts:

Flight Number vs. Launch Site scatter plot

- Revealed that CCAFS SLC-40 had the most launches.

Payload Mass vs. Orbit chart

- Showed heavier payloads tend to go to LEO.

Flight Number vs. Orbit

- Showed success rates improved over time for ISS missions.

Launch success rate by orbit

- Showed ES-L1, GEO, HEO, and SSO all achieved 100% success.

Bar chart of launch site success rates

- Showed KSC LC-39A had the highest success ratio.

https://github.com/kimannu/kimcapstone/blob/main/edadataviz.ipynb

EDA with SQL

SQL queries were used to validate and extend these findings numerically

Summary of SQL queries performed:

- Queried total payload mass carried by NASA (CRS) missions.
- Identified average payload mass by booster version.
- Listed launch sites with at least 10 successful ground landings.
- Counted total successful and failed mission outcomes.
- Queried landing outcomes between 2010-06-04 and 2017-03-20 ranked by date.

Together, the visual and SQL-based analysis revealed clear patterns across launch sites, payload ranges, and booster versions that would later inform feature selection for the classification models.

https://github.com/kimannu/kimcapstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(2).ipynb

Build an Interactive Map with Folium

Interactive geospatial mapping revealed that all SpaceX launch sites are coastal, a deliberate design choice for trajectory safety and booster recovery.

The following objects were created and added to a folium map:

Polylines: Connected sites to landmarks to visualise prCircles (1 km radius):

- Added to each SpaceX launch site (CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E) to highlight their exact locations.
- Popups were included to display the site name when clicked.

Text Markers (DivIcon):

- Placed at each site to label the locations directly on the map for easy identification.

MarkerCluster with Colour-Coded Markers:

- Used to group launch outcomes.
- Green markers represent successful launches (class = 1) and red markers represent failed launches (class = 0), enabling visual comparison of success rates.

MousePosition Plugin:

- Added to capture coordinates while hovering over the map, helping identify nearby landmarks.

Markers for Nearby Features:

- Used to display the coastline, highway, railway, and nearest city, including calculated distances in kilometres.

Polylines:

- Drawn to connect launch sites to nearby landmarks, visually showing proximity and highlighting strategic site placement near coasts and transport routes while remaining distant from populated areas.

https://github.com/kimannu/kimcapstone/blob/main/lab_jupyter_launch_site_location%20(2).ipynb

Build a Dashboard with Plotly Dash

The Dash dashboard was built to make the patterns from EDA interactively explorable allowing any stakeholder to filter by site and payload range without needing to run code.

Plots Added:

Pie Chart

- Shows total successful launches by site, or success vs failure rate for a selected site.
- This was chosen because it quickly and clearly shows proportions, making it easy to see which site performs best at a glance.

Scatter Plot

- Shows correlation between payload mass and launch outcome, color-coded by booster version.
- This was chosen because it is ideal for revealing patterns between two variables, helping identify whether heavier payloads or certain boosters affect mission success.

Interactions Added

Dropdown Menu

- Allows user to select All Sites or a specific launch site, updating both charts automatically.
- This was added so users can drill down into individual sites rather than being overwhelmed by all the data at once.

Range Slider

- Filters launches by payload mass (0 -10,000kg), updating the scatter plot in real time.
- This was added because payload mass varies widely, so filtering by range helps isolate patterns that might not be visible when viewing all data together.

https://github.com/kimannu/kimcapstone/blob/main/spacex_dash_app%20(1).py

Predictive Analysis (Classification)

With the data thoroughly explored, the question became: can these features reliably predict landing outcomes? Four classification models were trained to find out.

- **Build**

Trained 4 classifiers: Logistic Regression, SVM, Decision Tree, KNN

- **Evaluate**

Used confusion matrices and accuracy scores on an 18-sample test set

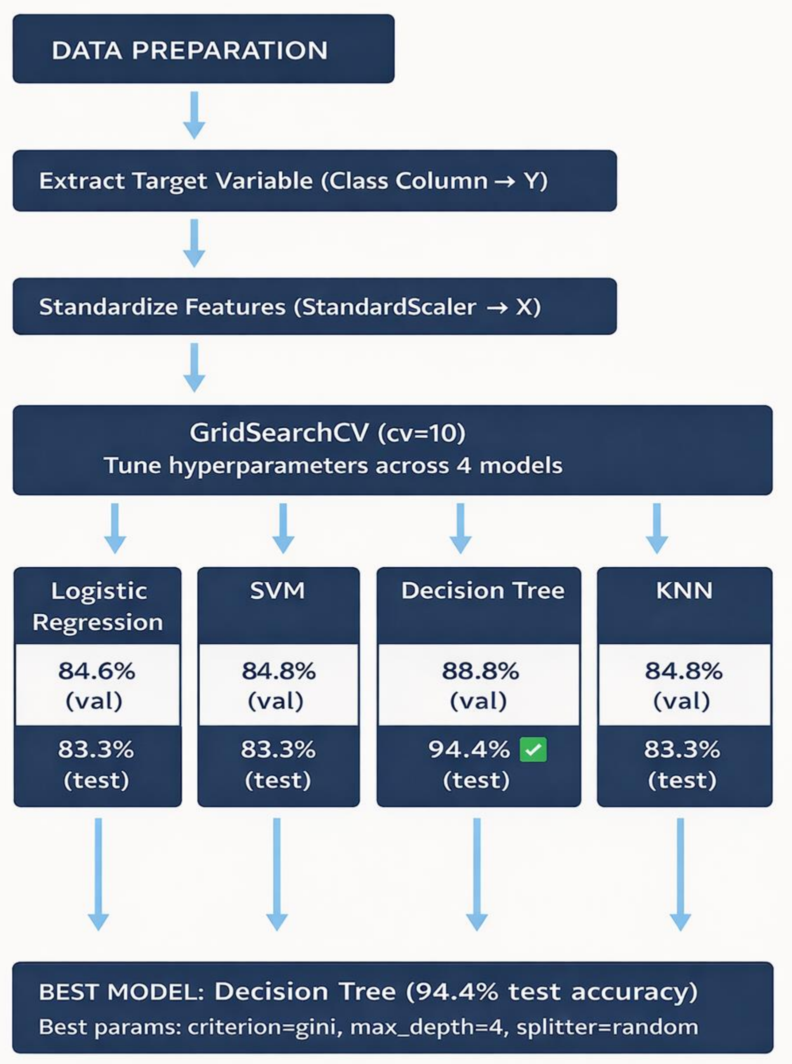
- **Improve**

Applied GridSearchCV with 10-fold cross-validation to tune hyperparameters for each model

- **Best Model**

Decision Tree achieved the highest test accuracy at **94.4%**, outperforming all other models which plateaued at 83.3%

https://github.com/kimannu/kimcapstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

The following summarizes the key findings from each stage of the analysis (which is explored in detail later).

Exploratory Data Analysis

- **KSC LC-39A** has the largest number of successful launches
- **KSC LC-39A** also has the highest launch success rate at **76.9%**
- Success rates have **improved over time** as SpaceX refined their technology
- Heavier payloads (5,000–10,000 kg) showed **higher success rates** in later years
- Launches to **LEO orbit** had the highest success rate across all sites

Predictive analysis

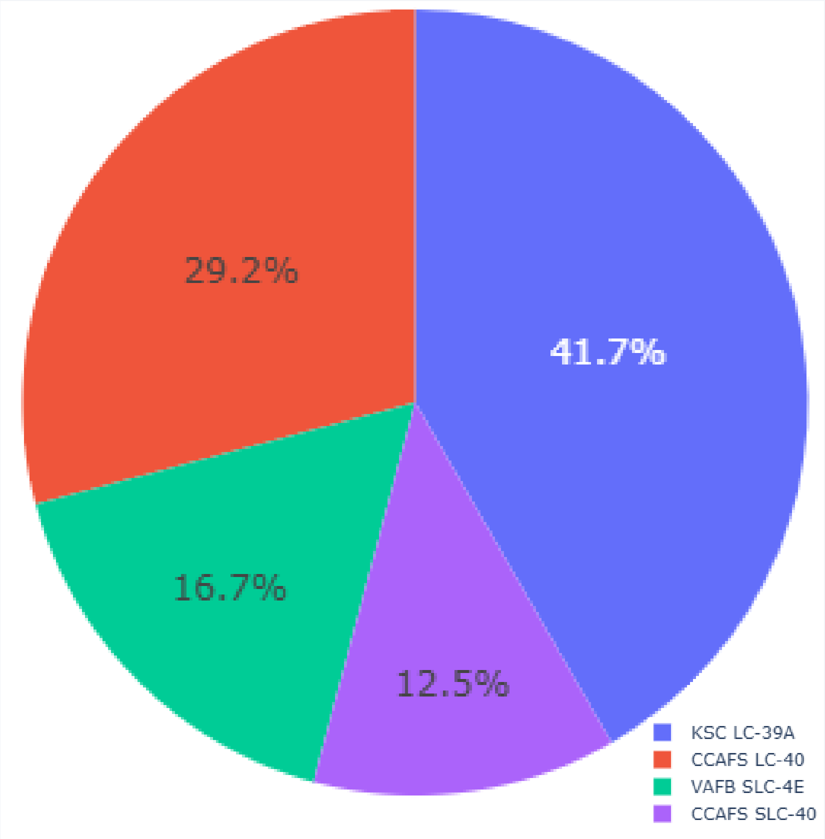
Model	Validation Accuracy	Test Accuracy
Logistic Regression	84.6%	83.3%
SVM	84.8%	83.3%
Decision Tree	**88.8%**	**94.4%** ✅
KNN	84.8%	83.3%

- **Decision Tree** was the best performing model with **94.4% test accuracy.**
- Best hyperparameters: criterion=gini, max_depth=4, splitter=random.
- The model successfully distinguishes between successful and failed landing.

Results

Dashboard analysis reinforced the EDA findings that payload range and booster version were consistent predictors of success.

- Pie chart showed KSC LC-39A contributes 41.7% of all successful launches across all sites
- Payload range 2,0r plot revealed that FT Booster version 00–5,500 kg showed the highest launch success rate
- Payload range 6,000+ kg showed the lowest launch success rate
- Scatter plot revealed that FT Booster version has the highest launch success rate across all payload ranges
- Confusion matrices and accuracy charts from the classification models confirmed the Decision Tree as the clear best performer



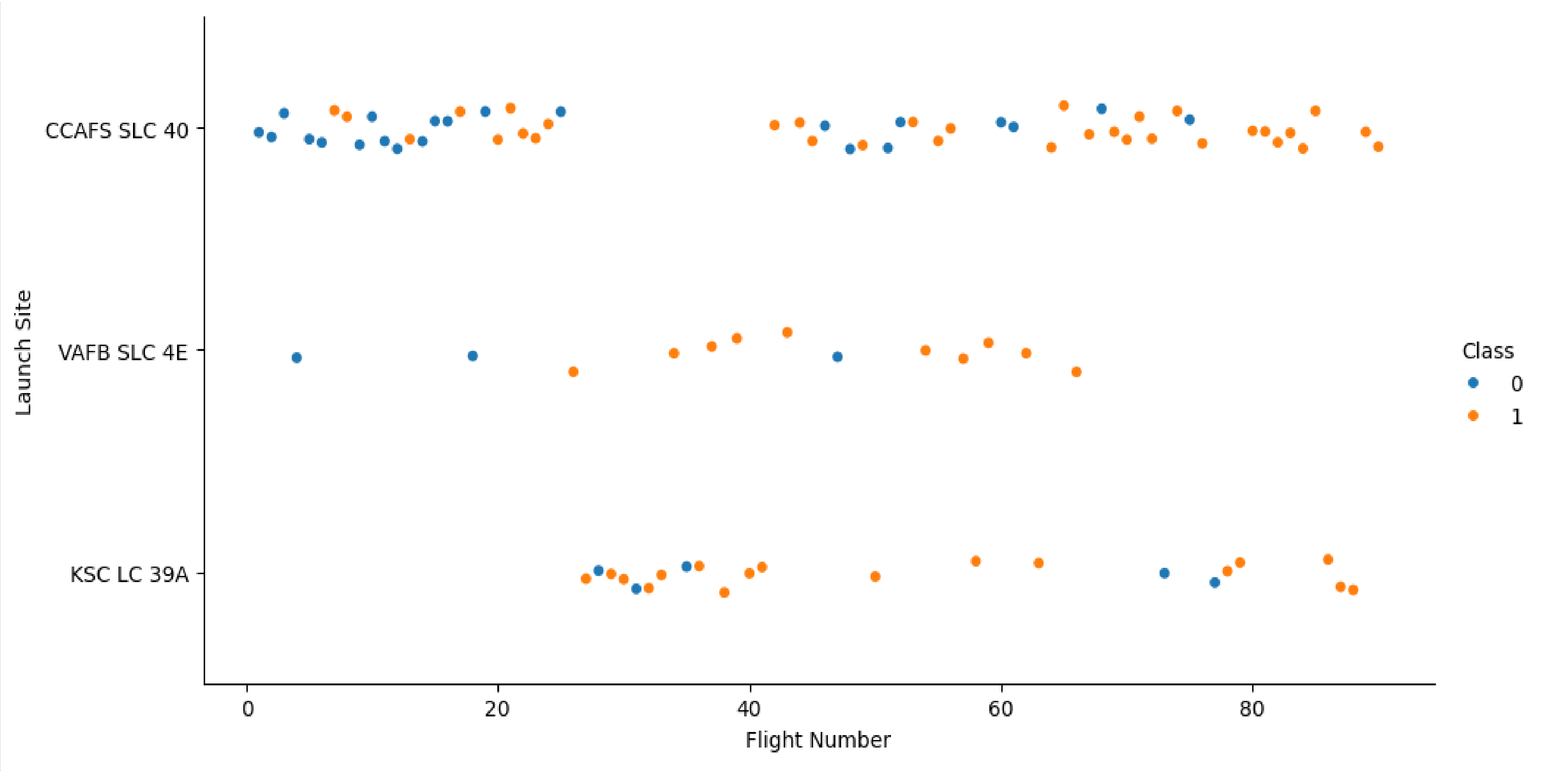
Results

Payload range (Kg):
0
9600
0
2500
5000
7500
10000
Correlation between Payload and Success for all Sites
class
1
0.8
0.6
0.4
0.2
0
2k
4k
6k
8k
10k
Payload Mass (kg)
Booster Version Category
v1.0
v1.1
FT
B4
B5

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

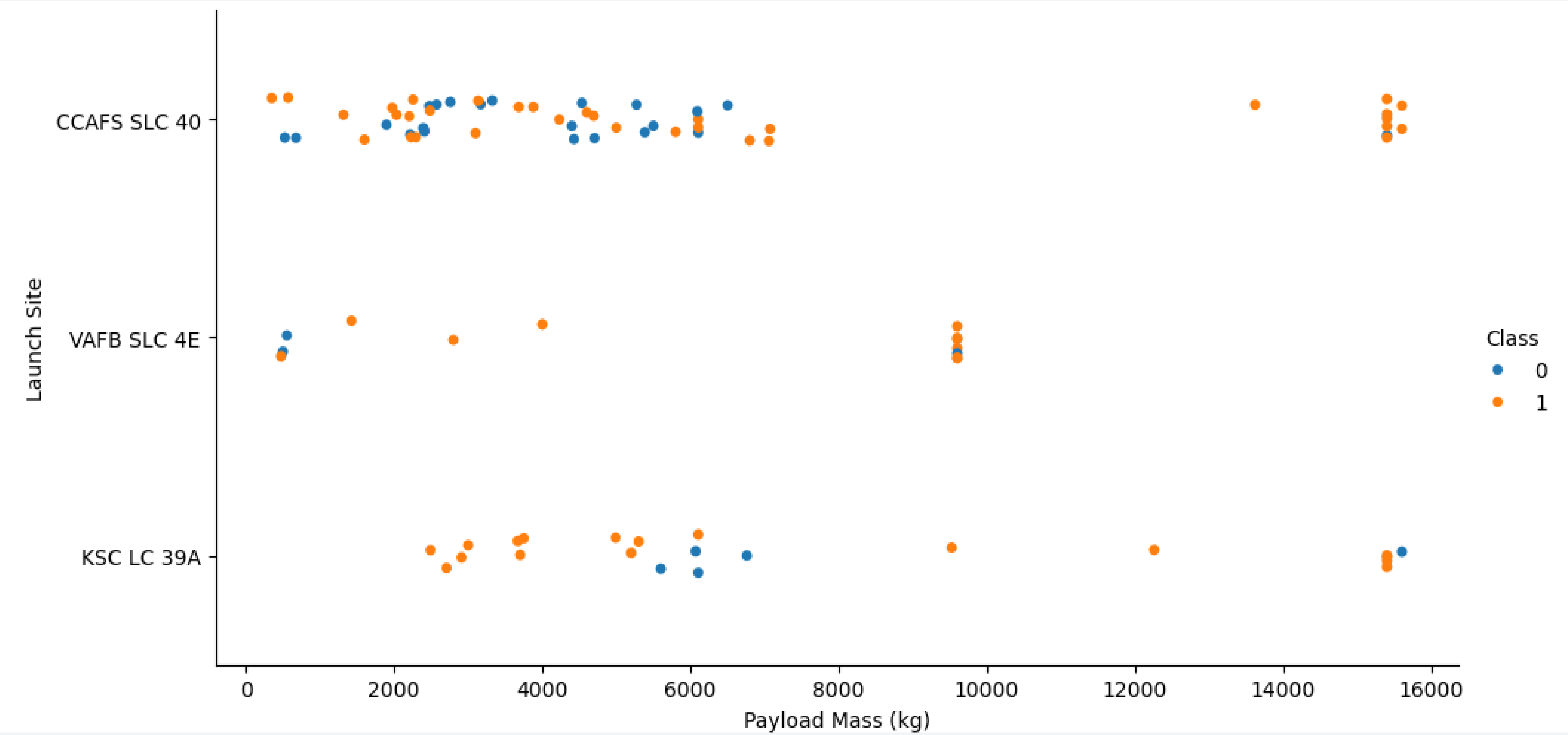


This scatter plot shows which launch site was used for each flight, colored by mission success (orange = success, blue = failure).

It shows that:

- **CCAFS SLC-40** was by far the most frequently used site, active across nearly all flight numbers from the very beginning to the end of the dataset.
- Success rate visibly improves at higher flight numbers across all sites, showing that as SpaceX gained experience, failures became increasingly rare.

Payload vs. Launch Site



This scatter plot shows the payload mass carried per mission at each launch site, colored by mission success (orange = success, blue = failure).

It reveals that:

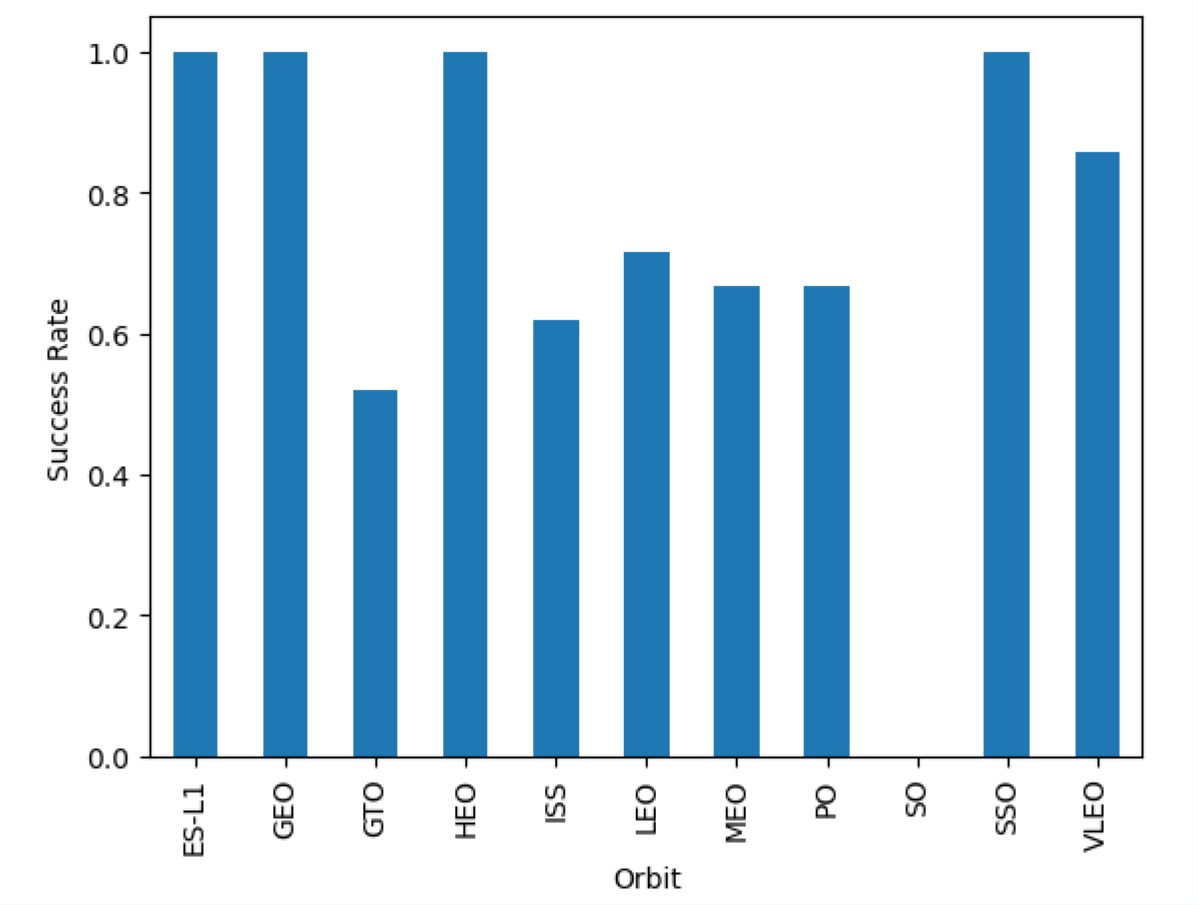
- **CCAFS SLC-40** handles the widest range of payload sizes, from very light to extremely heavy (~15,500 kg), making it SpaceX's most versatile launch site.
- Failures (blue) are scattered across all payload ranges with no clear pattern, suggesting payload mass alone wasn't the determining factor in mission success or failure.

Success Rate vs. Orbit Type

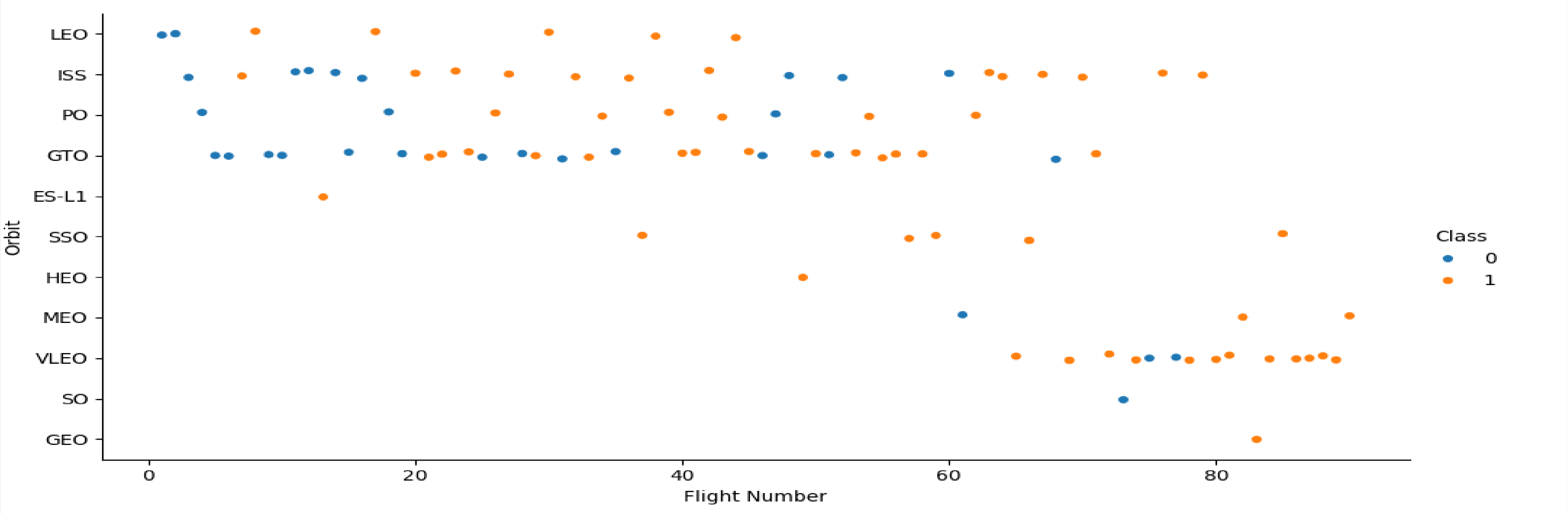
This bar chart shows the mission success rate for each orbit type SpaceX targeted.

It reveals:

- **ES-L1, GEO, HEO, and SSO** all achieved a perfect **100% success rate**, indicating SpaceX had fully mastered these orbit types by the time of the dataset.
- **GTO has the lowest success rate (~52%)**, likely because geostationary transfer orbits require maximum fuel, leaving little margin for booster recovery and increasing mission complexity.



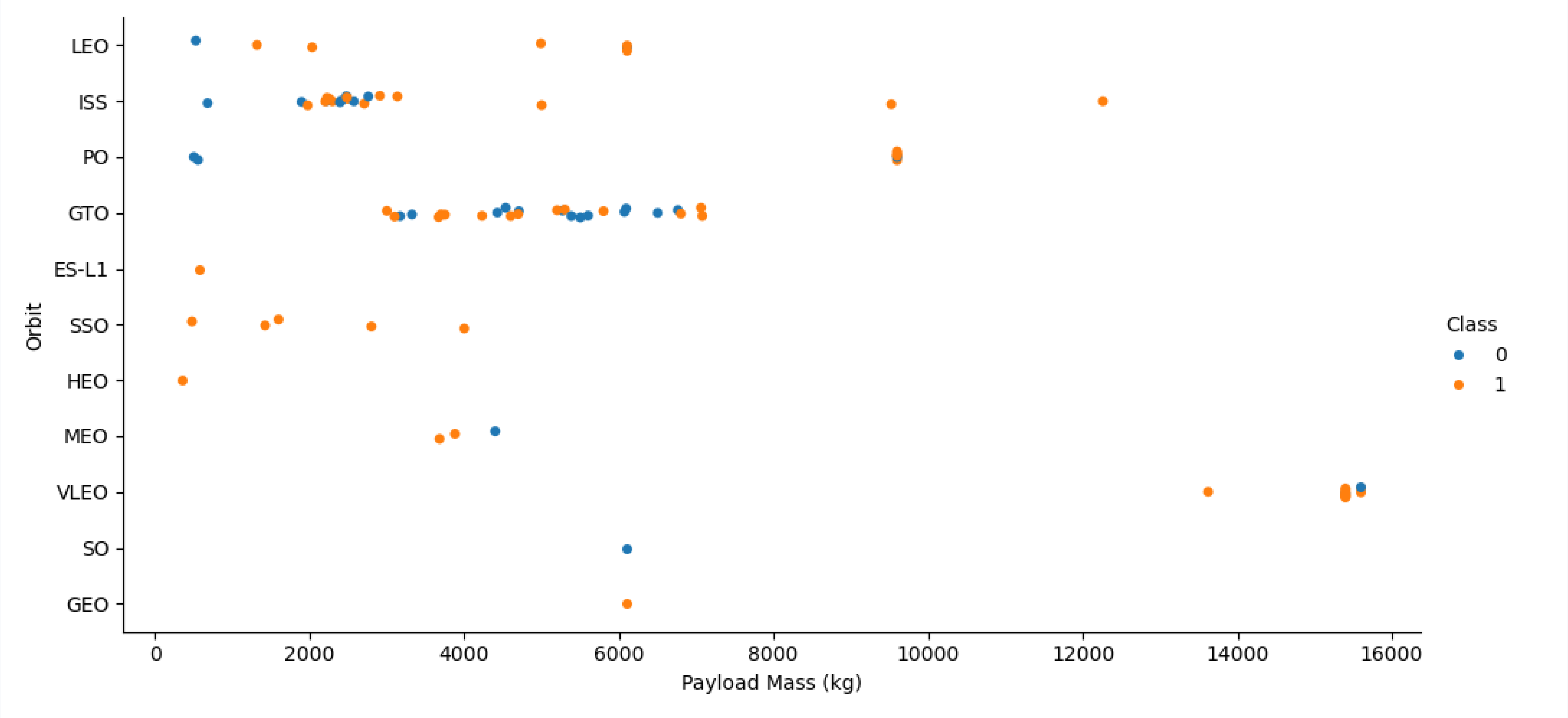
Flight Number vs. Orbit Type



This scatter plot shows which orbit type was targeted on each flight, colored by mission success (orange = success, blue = failure). It reveals that:

- **GTO and ISS missions dominate the early flights** (low flight numbers) and show the most failures, while newer orbit types like VLEO and SSO appear only at higher flight numbers with mostly successful outcomes.
- The shift toward more orange dots at higher flight numbers across all orbit types confirms SpaceX's growing reliability and expanding mission portfolio over time.

Payload vs. Orbit Type



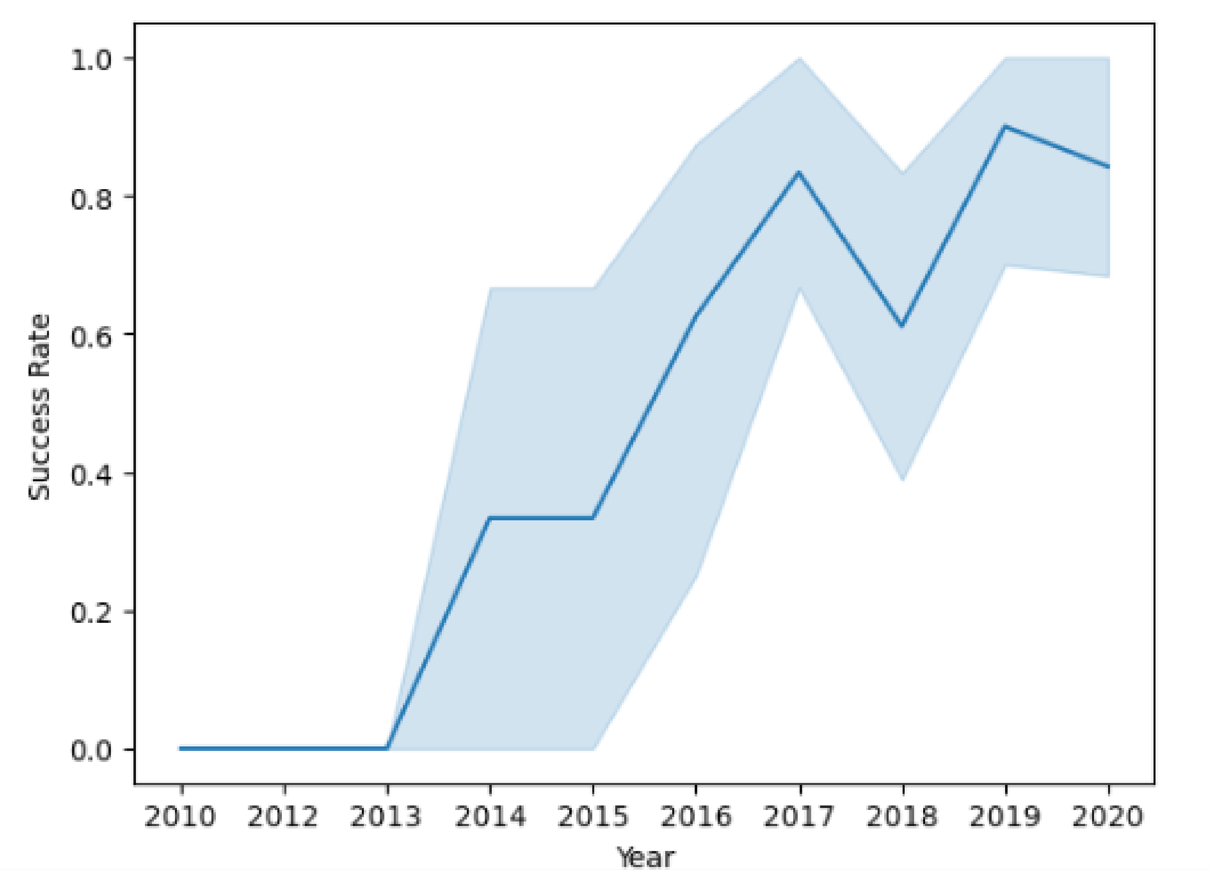
This scatter plot shows the payload mass carried for each orbit type, colored by mission success (orange = success, blue = failure). It revels that:

- **GTO missions cluster tightly between 3,000–6,500 kg** and show the most failures (blue), reinforcing that heavy geostationary payloads are the most challenging for SpaceX to execute successfully.
- **VLEO missions carry the heaviest payloads (13,000–15,500 kg)** and are almost entirely successful, reflecting the Starlink era where SpaceX had already mastered high-mass, low-orbit deployments

Launch Success Yearly Trend

This line chart showing SpaceX's average mission success rate per year, with a shaded confidence interval band.

- The trend tells a clear story of growth from **0% success in 2010–2013** to consistently above **80% from 2017 onwards**, reflecting how rapidly SpaceX matured as a launch provider.
- The **wide confidence band in the middle years (2015–2018)** indicates variability in results during that period, which narrows toward 2020 as successes became the norm rather than the exception.



All Launch Site Names

This query retrieves all unique launch sites used in SpaceX missions, returning 4 distinct locations:

- **CCAFS LC-40 -** Cape Canaveral
- **CCAFS SLC-40** - Cape Canaveral
- **VAFB SLC-4E** - Vandenberg Air Force Base.
- **KSC LC-39A** - Kennedy Space Centre

```
%%sql
SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

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 dataset shows the earliest SpaceX Falcon 9 launches (2010–2013), all from **CCAFS LC-40**, carrying Dragon spacecraft and NASA CRS missions with zero or low payloads, where landing was either failed via parachute or not attempted at all - reflecting SpaceX's early experimental phase before reusable landing was developed.

Total Payload Mass

Total Payload carried by boosters from NASA = 45596

```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass
FROM SPACEXTABLE
WHERE Customer = 'NASA (CRS)';
```

This query calculates the total payload mass (in kg) across all SpaceX launches where NASA was the customer under the Commercial Resupply Services (CRS) program, by summing the `PAYLOAD_MASS__KG_` column and filtering rows where the Customer is exactly `'NASA (CRS)'`.

Average Payload Mass by F9 v1.1

This query calculates the average payload mass carried by the F9 v1.1 booster version, returning 2928.4 kg representing the typical cargo weight this early Falcon 9 variant was carrying per mission.

```
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS Avg_Payload_Mass
FROM SPACEXTABLE
WHERE Booster_Version = 'F9 v1.1';
```

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

Avg_Payload_Mass
2928.4

First Successful Ground Landing Date

This query finds the earliest date of a successful Falcon 9 first-stage landing on a ground pad, revealing that December 22, 2015 was the historic milestone when SpaceX first successfully landed a booster back on solid ground.

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

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

First_Successful_Ground_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

This query identifies the **booster versions** that successfully landed on a drone ship while carrying a payload between 4,000 - 6,000 kg, returning 4 boosters:

B1022

B1026

B1021.2

B1031.2

All from the F9 FT (Full Thrust) generation, which was SpaceX's workhorse during their drone ship landing development phase.

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

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

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

Total Number of Successful and Failure Mission Outcomes

This query groups and counts missions by their outcome, showing that SpaceX had an overwhelming **98 successes** against just **1 in-flight failure**, with 1 additional ambiguous success reflecting an exceptionally high mission success rate across the dataset.

```
%%sql
SELECT Mission_Outcome, COUNT(*) AS Count
FROM SPACEXTABLE
GROUP BY Mission_Outcome;
```

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

Mission_Outcome	Count
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

This query uses a subquery to find the **maximum payload mass** in the dataset, then lists all boosters that carried that exact weight - showing that **12 F9 B5 boosters** all carried the maximum payload, highlighting that the latest Block 5 generation was SpaceX's most capable and frequently used heavy-lift variant.

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

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

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

This query uses substr() to extract the month and year from the date (since SQLite lacks date functions), revealing that:

- In 2015 SpaceX had 2 drone ship landing failures in January and April
- Both used the F9 v1.1 booster from CCAFS LC-40
- These failures represented their early and unsuccessful attempts to master controlled ocean landings.

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

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

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

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

This query counts and ranks all landing outcomes between June 2010 and March 2017, showing SpaceX's full learning curve

- Between June 2010 and March 2017, SpaceX conducted 31 landing attempts across 8 distinct outcomes, painting a clear picture of a company rapidly evolving its reusable rocket technology.
- The dominance of "No Attempt" (10 missions) reflects SpaceX's earliest flights, where recovery wasn't yet a priority - the focus was simply getting payloads to orbit. As confidence grew, SpaceX began experimenting with ocean splashdowns, first uncontrolled (2), then controlled (3), using the ocean as a safe testing ground before committing to full landings.
- The earliest recovery method - parachutes (2 failures) - was quickly abandoned after both attempts failed in 2010, leading SpaceX to develop the propulsive, powered landing system that defines them today.
- By 2015, drone ship and ground pad landings were being attempted regularly. Drone ship results were evenly split - 5 successes vs. 5 failures - capturing the intense trial-and-error phase of mastering ocean landings, with the 2015 failures in January and April eventually giving way to consistent successes. Ground pad landings (3 successes) proved more reliable once achieved, with the historic first on December 22, 2015 marking a turning point for the entire space industry.
- A single precluded attempt serves as a reminder that even beyond technical challenges, logistics and weather played a role in SpaceX's journey.

Taken together, the data tells a story of deliberate, methodical progress - from parachutes to precision landings - that ultimately revolutionized the economics of space travel.

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

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

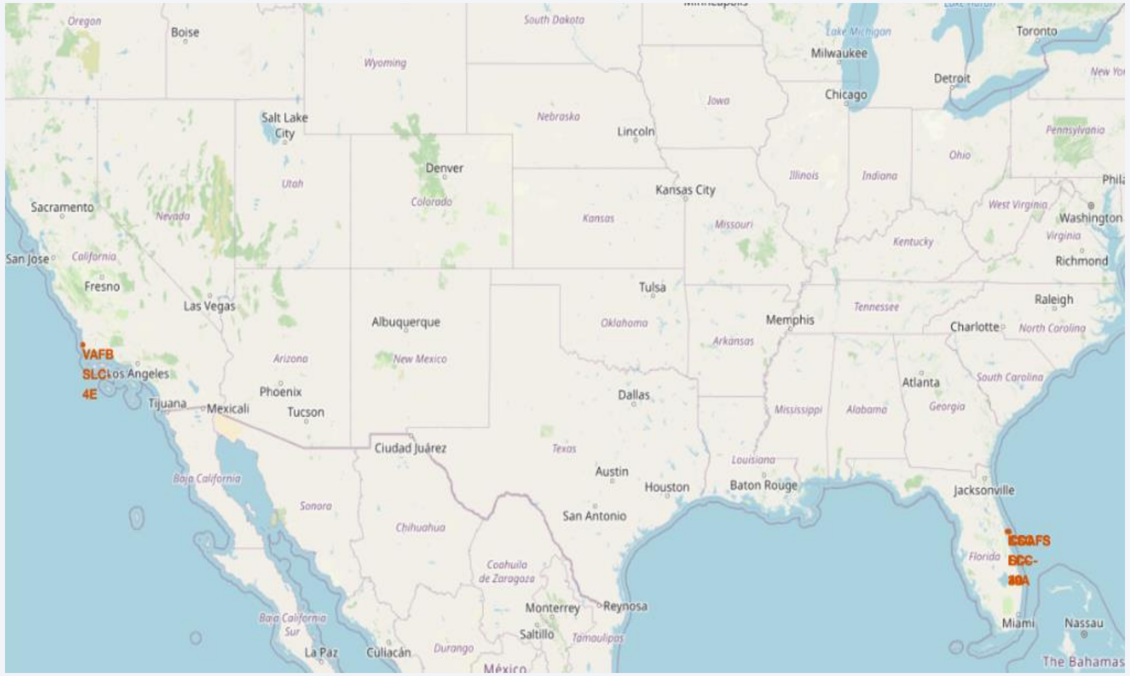
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

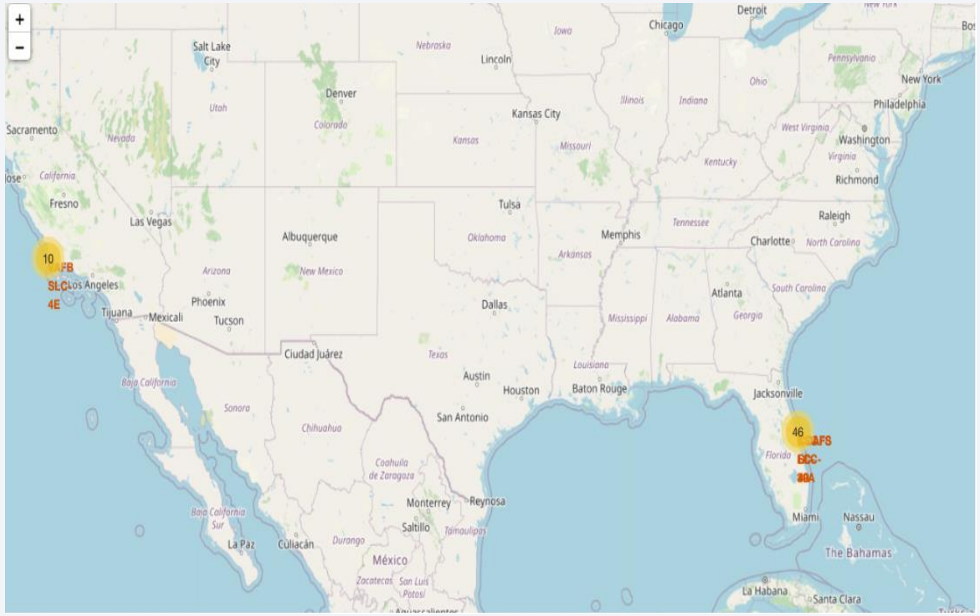
Global SpaceX Launch Site Locations

- The global map displays all 4 SpaceX launch sites marked with circle icons: CCAFS SLC-40 and KSC LC-39A in Florida, VAFB SLC-4E in California, and CCSFS SLC-40.
- All sites are coastal, confirming that launch sites are deliberately located near the ocean for safety and trajectory requirements.
- All launch sites are located along coastlines, enabling safe downrange trajectories over water.
- KSC and CCAFS in Florida support eastward equatorial launches; VAFB in California supports polar orbit launches.



Launch Outcomes by Site - Color-Coded Map

- Each launch marker is color-coded by outcome: green for successful landing, red for failed landing.
- KSC LC-39A shows a high concentration of green markers, indicating it has the best success rate.
- Earlier launches (lower flight numbers) show more red markers, reflecting the learning curve in SpaceX’s landing technology.
- Color-coded outcomes confirm that landing success improved significantly after 2017, with KSC LC-39A achieving the highest success rate among all sites.
- Green markers cluster at KSC LC-39A; red markers are more frequent at CCAFS SLC-40, especially in earlier flight numbers



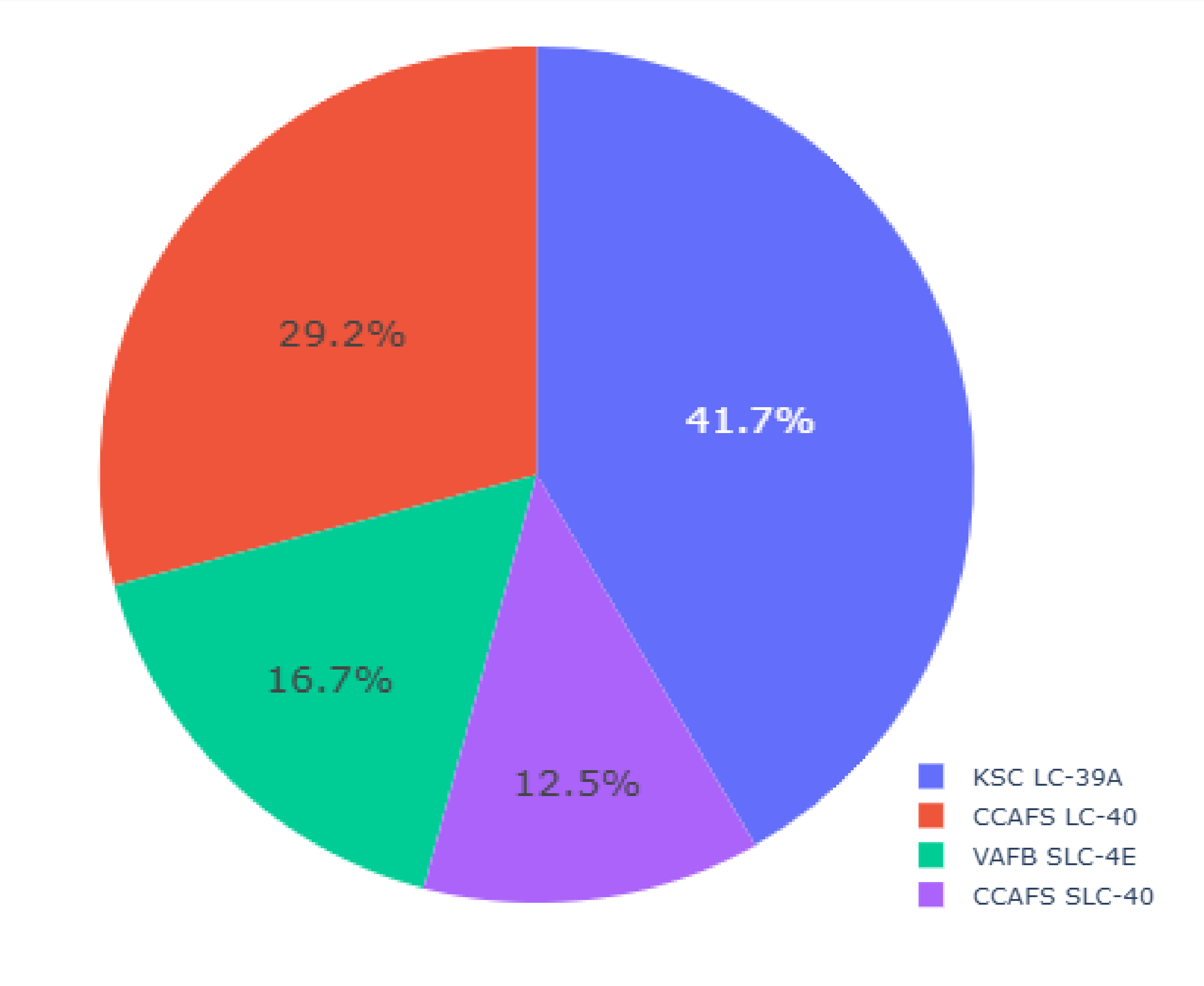
Launch Site Proximity Analysis - KSC LC-39A

- The proximity map for KSC LC-39A shows distances to nearby infrastructure.
- The site is approximately 1.6 km from the coastline, 6.5 km from the nearest highway, and 18 km from the nearest railway.
- KSC LC-39A's proximity to the coast minimizes safety exclusion zone requirements and provides direct ocean access for booster recovery operations.

Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site

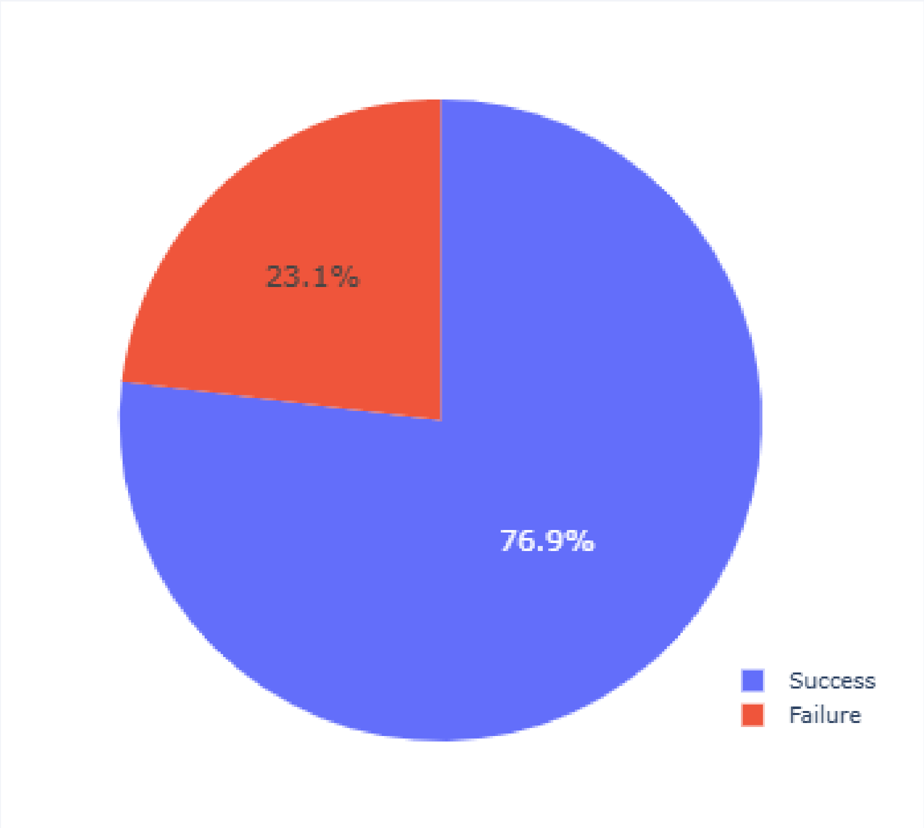


The pie chart shows the total launch success count across all sites.

- **KSC LC-39A-** the most reliable site - leads with **41.7%** of all successful launches.
- **CCAFS LC-40** - one of SpaceX's earliest active sites - is second at **29.2%.**
- **VAFB SLC-4E** - mainly used for polar orbit missions - contributes **16.7%**
- **CCAFS SLC-40** contributes the least at **12.5%**

Highest Success Rate Launch Site- KSC LC-39A

- KSC LC-39A achieved a success rate of approximately 76.9%, the highest among all Falcon 9 launch sites in the dataset.
- KSC LC-39A historically hosted NASA's Apollo and Space Shuttle programs, giving it a legacy infrastructure advantage that contributes to its high launch success reliability.
- KSC LC-39A's combination of infrastructure maturity and favorable launch geometry makes it SpaceX's most reliable site



Interactive Payload Analysis

•The following 4 charts explore the relationship between payload mass and launch outcomes for all of the sites.

•Each view filters a different payload range using an interactive slider

•The scatter plot shows payload mass (kg) on the x-axis and launch outcome on the y-axis

•Color coding: Failure = 0, Success = 1

0kg -10000kg (Full Payload)

The Full payload scatterplot shows a baseline view showing all launches across every site.

- CCAFS SLC-40 dominates in launch frequency across all payload sizes
- Failures are scattered with no clear payload-specific pattern at this level

0kg - 2000kg (Low Payload)

Zooming into lighter payloads reveals early mission behavior and site preferences.

- Lighter missions were mostly early flights with lower success rates
- VAFB SLC-4E features more prominently here with smaller, polar orbit payloads

2000kg - 5000kg (Medium Payload)

The mid payload range is the most competitive and mixed range where most GTO missions fall.

- Highest concentration of failures, particularly for GTO-bound missions
- B5 booster begins appearing here with noticeably better success rates

5000kg -10000kg (High Payload)

The heaviest payloads tell a story of maturity and mastery.

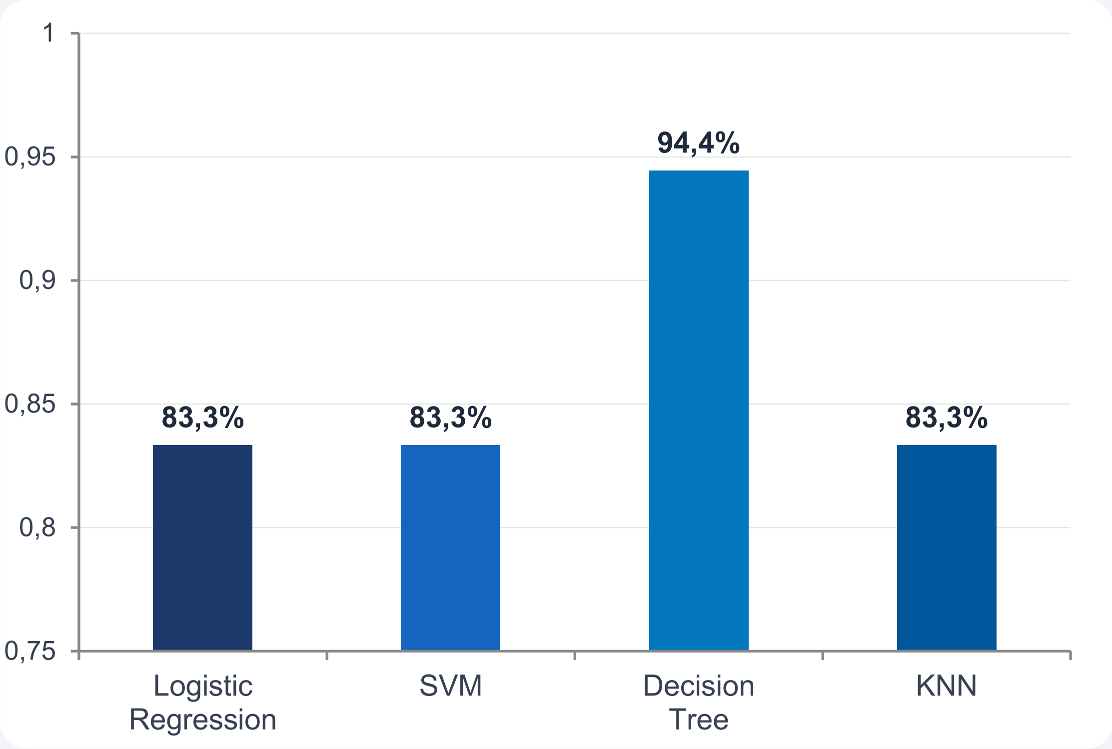
- Almost exclusively successful, dominated by Starlink/VLEO missions on the B5 booster
- KSC LC-39A handles the bulk of heavy payload launches with near-perfect reliability

Section 5

Predictive Analysis (Classification)

Classification Accuracy

Four classification models were trained and evaluated using GridSearchCV with 10-fold cross-validation.



Best Model

Decision Tree

94.44% Test Accuracy

Logistic Regression	83.33%
SVM	83.33%
KNN	83.33%

- The Decision Tree classifier achieved the highest test accuracy at 94.44%, outperforming Logistic Regression, SVM, and KNN which all scored 83.33%.
- The Decision Tree model was selected as the final model for its superior accuracy and interpretability, making it well-suited for binary classification of landing outcomes.

Confusion Matrix

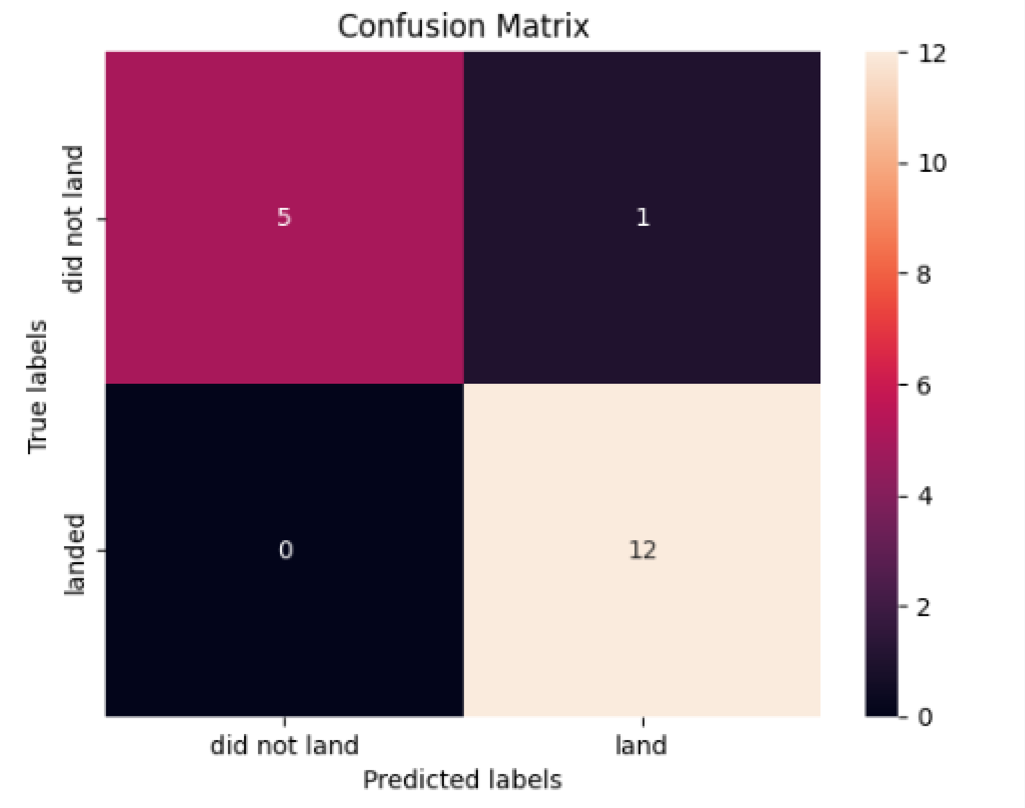
The Decision Tree model's confusion matrix on the test set shows:

- True Negatives (did not land, correctly predicted): 5
- False Positives (did not land, predicted land): 1
- False Negatives (landed, predicted did not land): 0
- True Positives (landed, correctly predicted): 12.

The model correctly identified all 12 successful landings with zero misses (0 false negatives).

Only 1 non-landing was incorrectly classified as a landing.

This gives a precision of 92.3%, recall of 100%, and F1-score of 96.0% for the positive class.



Conclusions

- Returning to the question posed at the outset - **yes**, launch parameters can reliably predict Falcon 9 landing success.
- The Decision Tree classifier achieved the highest accuracy (94.44%) in predicting Falcon 9 first-stage landing success, outperforming Logistic Regression, SVM, and KNN (all 83.33%).
- KSC LC-39A is the most successful Falcon 9 launch site with a ~76.9% landing success rate, while CCAFS SLC-40 has the highest number of total launches.
- Payload mass between 2,000–6,000 kg and the Falcon 9 FT booster version are associated with the highest landing success rates across all launch sites.
- Landing success rates have improved significantly over time - earlier missions (pre-2017) had much higher failure rates, demonstrating SpaceX's rapid iteration and refinement of reusable rocket technology.
- Orbits such as ES-L1, GEO, HEO, and SSO achieved 100% mission success, while GTO and SO missions showed more variability, suggesting orbit type is an important predictor of landing outcome.
- These findings demonstrate that publicly available mission data alone contains sufficient signal to support operational launch planning and cost forecasting.

Thank you!