Will It Land?

Using Machine Learning to Predict Falcon 9 Success Rates

Presented by: Michael Pereira

Context: IBM Data Science Capstone Project

Date: January 2026

OUTLINE

1. Executive Summary
2. Introduction
3. Methodology
4. Visual Analytics
5. Predictive Analysis
6. Conclusion

EXECUTIVE SUMMARY

Methodologies

Data Pipeline: Automated collection via API and web scraping.

Analysis: Cleaned and explored data to identify success factors.

Visualization: Built interactive dashboards to map launch sites and payloads.

Modeling: Trained machine learning classifiers to predict landing outcomes.

Results

Predictive Accuracy: Achieved 83% accuracy across models.

Key Drivers: Validated that Payload Mass and Orbit Type dictate success.

Launch Sites: Certain sites show consistently higher success rates.

INTRODUCTION

Background & Context

The New Era:

SpaceX is revolutionizing the aerospace industry by making rocket launches affordable through reusability.

The Cost Factor:

The Falcon 9 First Stage is the most expensive component of the rocket.

The Goal:

Landing and reusing the first stage saves millions of dollars per launch, making space accessible.

Problem Statement

The Core Question:

Can we use historical data to predict if theFirstStage will land successfully?

Key Variables to Investigate:

Does Payload Mass impact landingsuccess?

Which Launch Sites have the highestsuccess rates?

How do different Orbit Types (LEO, GTO,ISS) affect the outcome?

Methodology

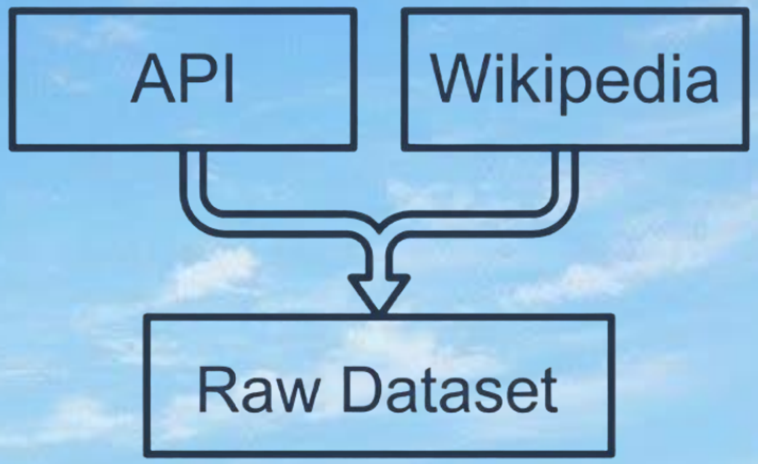
Data Science Pipeline

- **Data Collection** (API & Scraping)
- **Data Wrangling** (Cleaning & Encoding)
- **Exploratory Analysis** (SQL & Visualization)
- **Interactive Analytics** (Folium & Dash)
- **Predictive Modeling** (Classification)

Dual-Source Data Collection Strategy

- **Source 1: The Official Record.**
  - Used the SpaceX REST API to retrieve technical launch specifications.

- **Source 2: The Historical Context.**
  - Web Scraped Wikipedia public records to supplement API gaps.



API Extraction Process

The Source

Target Endpoint:
api.spacexdata.com/v4/launches/past

Objective: Gather historical launch data to predict future landing attempts.

The Process

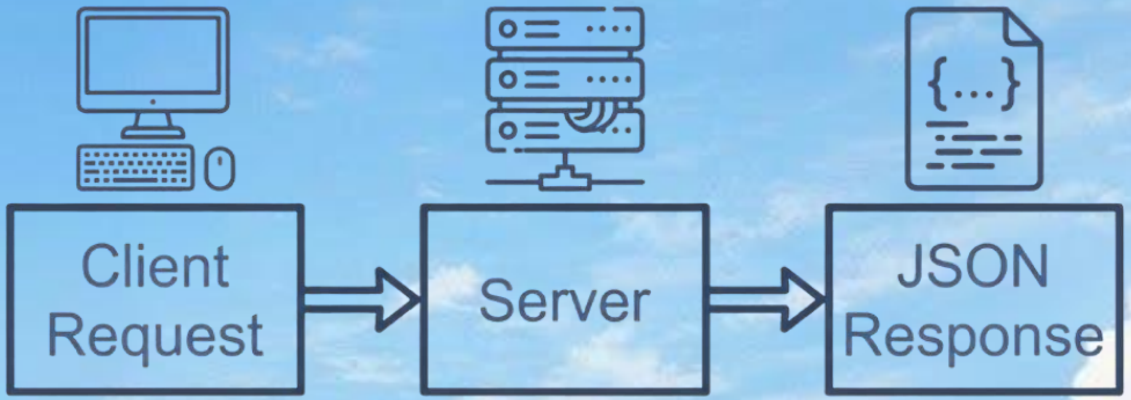
Request: Executed a GET request using the Python requests library.

Response: Received a list of JSON objects, where each object represents a single launch.

Transformation

Normalization: Utilized pd.json_normalize() to convert the nested JSON structure into a flat Pandas DataFrame.

Outcome: Successfully structured raw JSON into a tabular format containing rocket, payload, and landing specifications.



https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Web Scraping Wikipedia

Objective:

Recover historical launch data not fully detailed in the v4 API.

Tools:

BeautifulSoup for HTML parsing.

Process:

Request HTML from Falcon 9 Wiki page.

Parse <table> elements for launch records.

Extract specific columns (Date, Payload, Outcome).

[hide] Flight No.	Date and time (UTC)	Version, booster[j]	Launch site	Payload[k]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
286	January 3, 2024 03:44[23]	F9 B5 B1082-1	Vandenberg, SLC-4E	Starlink: Group 7-9 (22 satellites)	~16,800 kg (37,000 lb)	LEO	SpaceX	Success	Success (OCISLY)
	Launch of 22 Starlink v2 mini satellites, including the first six to feature direct-to-cell connectivity, to a 525 km (326 mi) orbit at an inclination of 53° to expand internet constellation.								
287	January 3, 2024 23:04[24]	F9 B5 B1076-10	Cape Canaveral, SLC-40	Ovzon-3	1,800 kg (4,000 lb)	GTO	Ovzon	Success	Success (LZ-1)
	Broadband internet provider satellite.[25] First Falcon 9 launch to GTO with a return-to-launch-site (RTLS) landing. First commercial satellite with Roll Out Solar Array that was deployed on January 10, 2024.[26][27]								
288	January 7, 2024 22:35[28]	F9 B5 B1067-16	Cape Canaveral, SLC-40	Starlink: Group 6-35 (23 satellites)	~17,100 kg (37,700 lb)	LEO	SpaceX	Success	Success (ASOG)
	Launch of 23 Starlink v2 mini satellites to a 530 km (330 mi) orbit at an inclination of 43° to expand internet constellation. Falcon record for total time from hangar rollout to launch at 6 hours, 33 minutes.[29]								

https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling: Preprocessing & Feature Engineering

1. Data Expansion (API Recursion)
   - **Problem:** Initial dataset contained only ID numbers for Rocket, Launchpad, and Core.
   - **Solution:** Targeted specific API endpoints to extract actual attributes (Booster version, Longitude/Latitude, Reused count) for every ID.
2. Data Filtering
   - **Target:** Isolated the study to Falcon 9 boosters only.
   - **Action:** Removed all Falcon 1 rows to ensure analysis relevance.
3. Handling Missing Values (Nulls)
   - **Payload Mass:** Calculated the mean of the column and replaced all NULL values with that mean.
   - **Landing Pad:** Retained NULL values to explicitly represent launches where no landing pad was used (e.g., ocean crashdown).
4. Outcome Classification (The Target Variable)
   - **Raw Data:** contained 8 distinct string labels (e.g., "True ASDS" for successful drone ship landing).
   - **Transformation:** Converted all outcomes into a binary Class variable:
     - 1 (Success): The booster landed successfully.
     - 0 (Failure): The booster crashed or did not attempt landing.

https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Exploratory Data Analysis (Visualization)

Payload & Experience Analysis (Scatter Plots)

Plotted Flight Number vs. Payload Mass.

To determine if the likelihood of a successful landing is impacted by the rocket's workload (mass) or the company's experience (flight number).

Site Distribution Analysis (Catplots)

Visualized Flight Number vs. Launch Site.

To analyze the frequency of launches at each site and see if certain sites are prioritized as the program matures.

Orbital Risk Assessment (Bar Charts)

Aggregated Success Rate vs. Orbit Type.

To identify which specific orbits (e.g., LEO, GTO, ISS) pose the highest technical challenge for stage recovery.

Temporal Trends (Line Chart)

Plotted Yearly Success Rate (2013–2020).

To observe the historical trend and validate if SpaceX's reliability has improved over time.

https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Exploratory Data Analysis (SQL)

Site & Payload Aggregation

Identified Unique Sites: Used DISTINCT to isolate all launch locations (CCAFS, VAFB, KSC).

Payload Statistics: Calculated the Total Payload Mass for specific customers (e.g., NASA CRS) and the Average Mass for specific booster versions (F9 v1.1).

Max Capacity: Located the booster version carrying the Maximum Payload Mass in the dataset.

Outcome & Milestone Analysis

First Success: Used MIN(Date) to pinpoint the exact date of the first successful ground pad landing.

Success vs. Failure: Executed sub-queries to compare total "Success" counts against "Failure" counts.

Targeted Filtering: Queried for specific success scenarios (e.g., Drone Ship landings with payload between 4,000–6,000kg).

Temporal Trends

Outcome Ranking: Used GROUP BY and ORDER BY to rank the most frequent landing outcomes between 2010 and 2017.

Failure Analysis: Extracted and named months (using CASE statements) to analyze failure patterns in specific years (e.g., 2015).

[19]:
```
%%sql
SELECT
    "Landing_Outcome",
    COUNT(*) as Outcome_Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
  AND "Landing_Outcome" IS NOT NULL
GROUP BY "Landing_Outcome"
ORDER BY Outcome_Count DESC;
```

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

[19]:

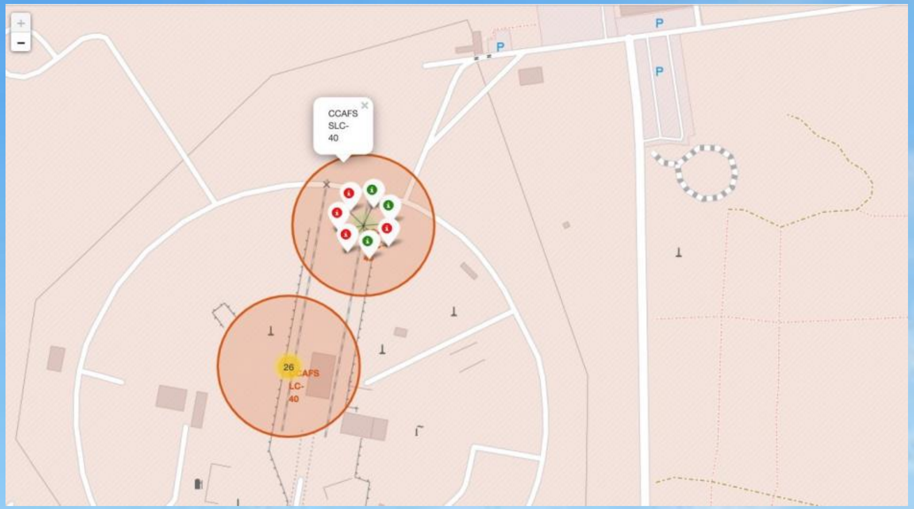
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/edadataviz.ipynb

Built an Interactive Map with Folium

Launch Site & Outcome Visualization

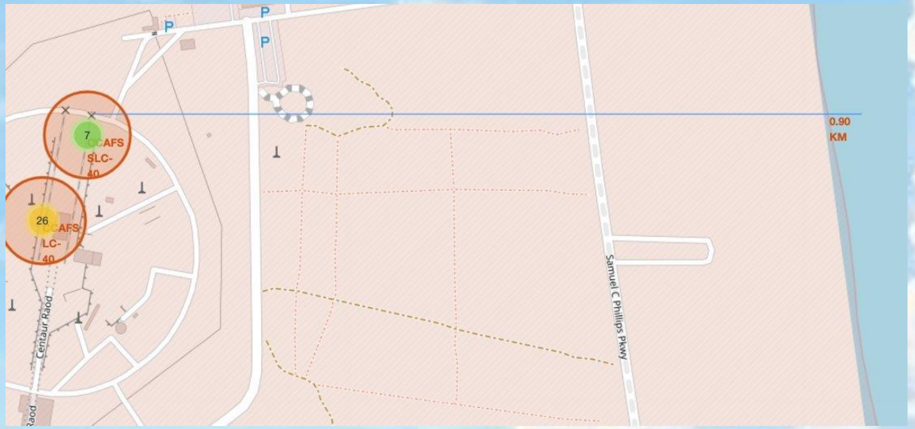
Added Circles to **mark site locations** and Marker Clusters to group individual launches.



Distance & Infrastructure Analysis

Implemented MousePosition for coordinate extraction and PolyLines to measure distances.

Calculated proximity to critical infrastructure to validate safety protocols.



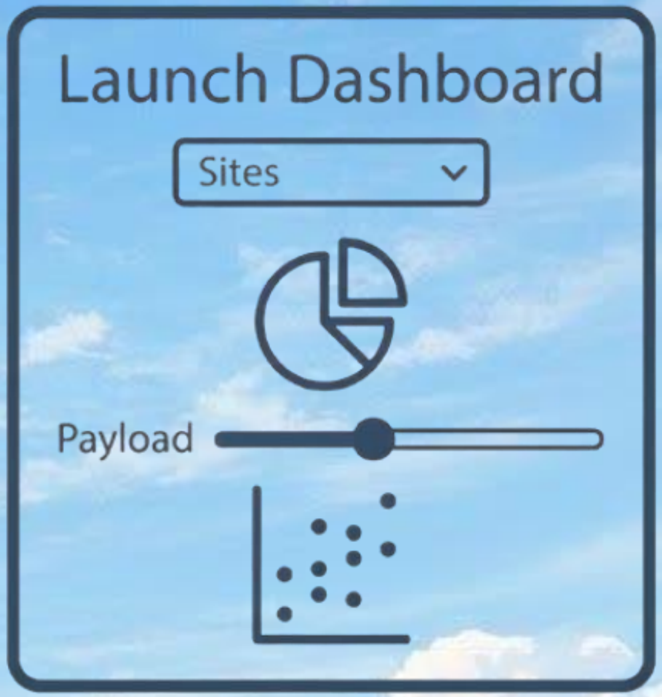
https://nbviewer.org./github/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/lab_jupyter_launch_site_location.ipynb

Built a Dashboard with Plotly Dash

Launch Site Dropdown & Pie Chart

A Dropdown Menu linked to a Pie Chart.

To allow users to select a specific launch site (e.g., KSC LC-39A). The Pie Chart instantly updates to show the Success vs. Failure percentage for that specific location, identifying the most reliable sites.



Payload Slider & Scatter Plot

A Range Slider (0–10,000 kg) linked to a Scatter Plot.

To enable filtering by payload mass. The Scatter Plot renders specific missions within that weight range, color-coded by Booster Version, revealing which boosters are best suited for heavy vs. light payloads.

https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

1. Data Preparation (Build)

Standardization: Applied StandardScaler to normalize data features (crucial for SVM and KNN algorithms).

Splitting: Divided the dataset using train_test_split into Training (80%) and Testing (20%) sets.

2. Model Optimization (Improve)

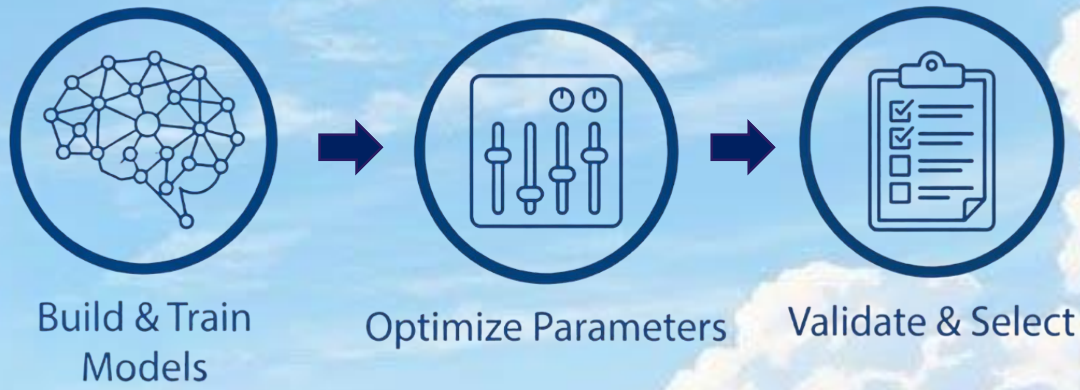
Hyperparameter Tuning: Utilized GridSearchCV with 10-fold Cross-Validation (cv=10) to exhaustively search for optimal settings.

Objective: Systematically identified the best parameters (e.g., C-value, Kernel, Max Depth) to maximize validation accuracy.

3. Evaluation & Selection (Find)

Candidate Models: Trained four distinct classifiers: Logistic Regression, SVM, Decision Tree, and K-Nearest Neighbors (KNN).

Metric: Selected the best performing model based on Test Data Accuracy and Confusion Matrix performance (minimizing False Positives).



https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

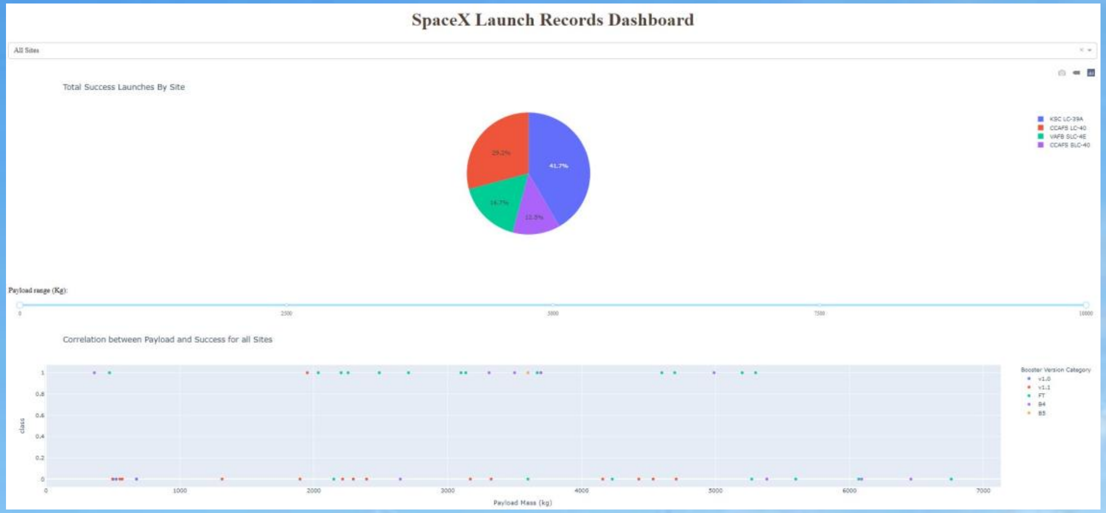
Results

Exploratory Data Analysis (EDA) Results

Trend Insight: Launch success rates have consistently increased annually from 2013 to 2020, showing a clear learning curve.

Site Insight: VAFB-SLC is exclusively used for lighter payloads (<10,000 kg), while KSC handles heavier missions.

Orbit Insight: Heavy payloads have high success rates in LEO, ISS, and Polar orbits, while GTO missions remain mixed.



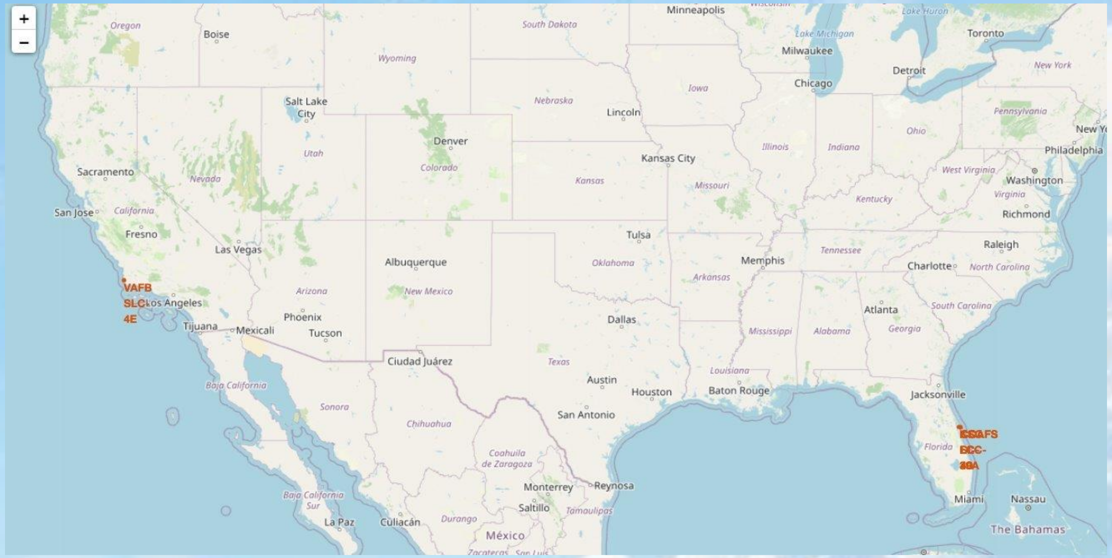
Predictive Analysis Results

Model Convergence: All four models (Logistic Regression, SVM,

Decision Tree, KNN) converged at approximately 83.33% Test Accuracy.

Data Limitation: The identical scores are due to the limited size of the test dataset.

Verdict: Despite the sample size, the high accuracy and low False Positive rate confirm the models are reliable predictors of landing success.



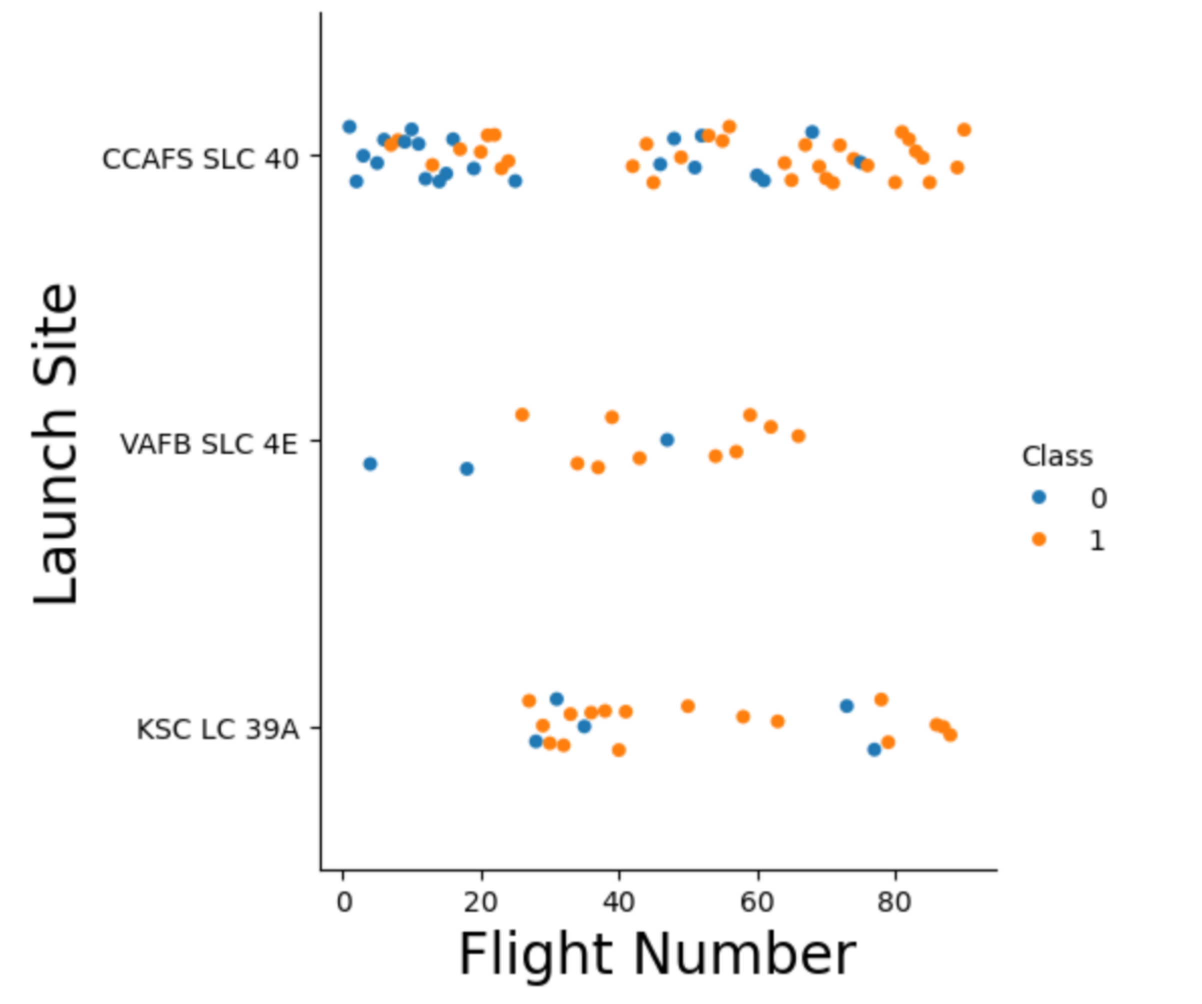
Insights drawn from EDA

Flight Number vs. Launch Site

Experience Drives Reliability: There is a clearcorrelation between higher flight numbers andsuccess rates. As the flight number increases(moving right), the density of Orange dots(Success) significantly increases, validating alearning curve.

Site Usage & History: CCAFS SLC 40 is the"workhorse" site with the longest history(lowest flight numbers) and the highestvolume of launches. It bears the brunt of earlyfailures (Blue dots) but shows a strongsuccess streak in later flights.

Legacy Benefit: VAFB SLC 4E and KSC LC 39Acame online later in the program (higher flightnumbers). They show consistently high successrates from the start, likely benefiting from thetechnical maturity gained at CCAFS.

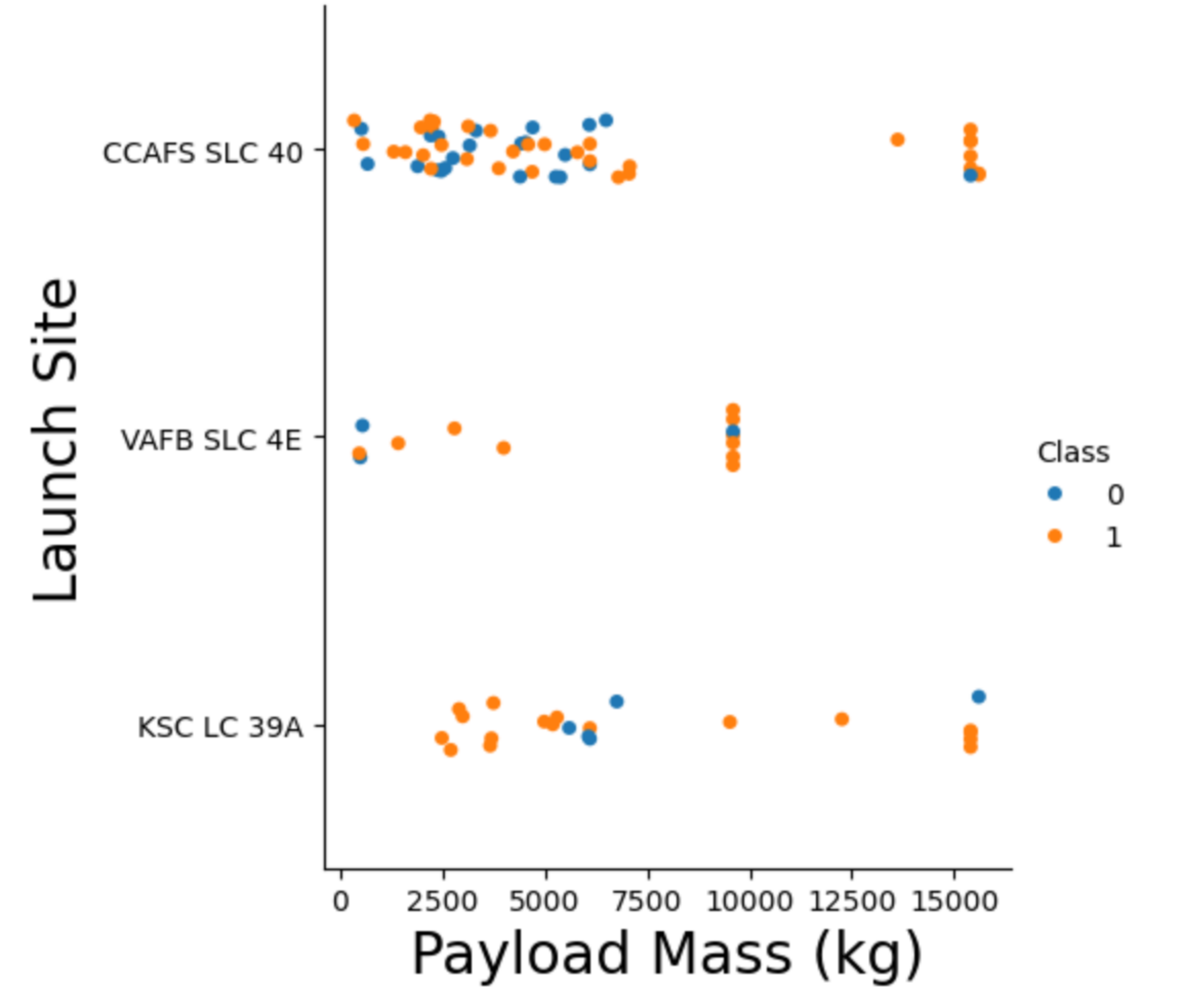


Payload vs. Launch Site

VAFB-SLC Limitation: This site shows aclear cutoff. There are no launches with apayload mass greater than 10,000 kg atVandenberg (VAFB SLC 4E). It isexclusively used for light-to-medium polarorbit missions.

Heavy Payload Success: Contrary to theassumption that heavier rockets areharder to land, the heaviest payloads (farright, >15,000 kg) at CCAFS SLC 40 andKSC LC 39A show a very high success rate(mostly Orange dots).

Site Specialization: While CCAFS SLC 40handles the widest range of payload masses(from light to very heavy), KSC LC 39Aappears to handle a more distributed mix ofmedium-to-heavy payloads compared toVAFB.

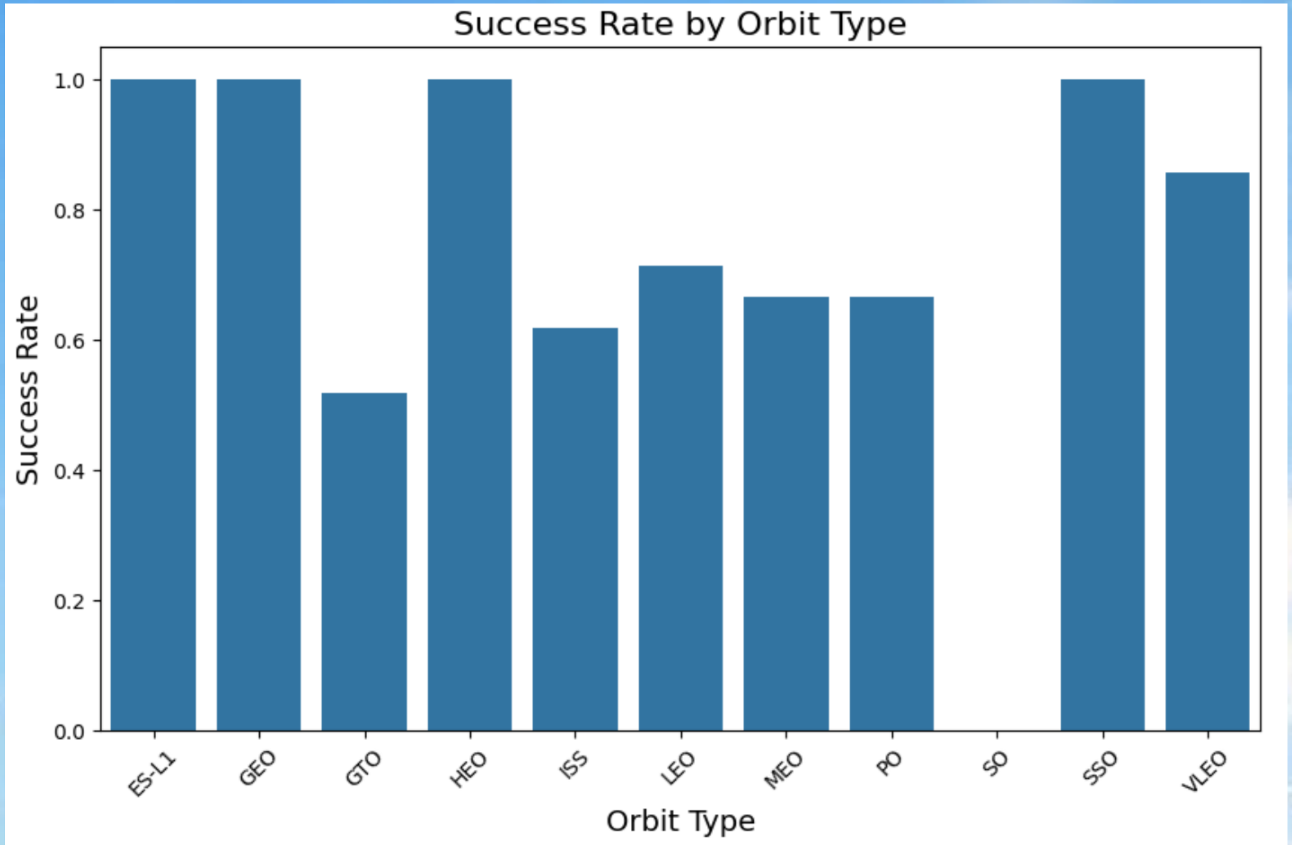


Success Rate vs. Orbit Type

Perfect Track Records: Four specificorbits—ES-L1, GEO, HEO, and SSO—achieved a 100% success rate for first-stage landings. This suggests thatmissions to these specific orbits in thisdataset were executed flawlessly.

The "GTO" Challenge: GTO (GeostationaryTransfer Orbit), which is a commondestination for commercial satellites, hasa success rate of only ~50%. Thisindicates it is a technically challengingorbit for recovery, likely due to the higherfuel requirements leaving less margin forthe landing burn.

The "SO" Outlier: The SO orbit shows a 0%success rate (no bar visible), identifying it asthe riskiest or least successful orbit type inthe dataset for landing attempts.

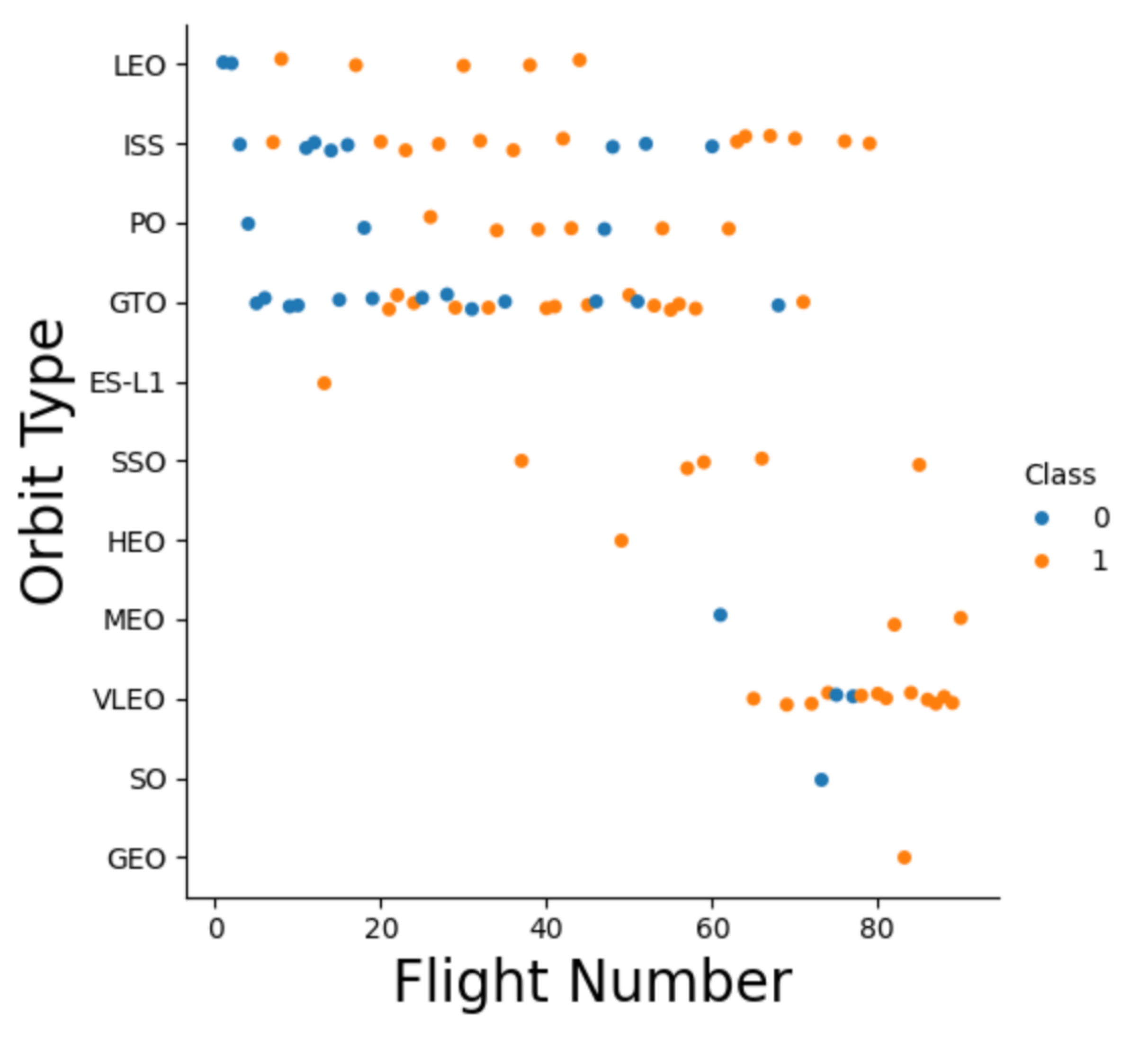


Flight Number vs. Orbit Type

The VLEO Shift (Starlink Era): There is amassive cluster of launches to VLEO (VeryLow Earth Orbit) appearing only in the laterstages (Flight Numbers > 60). These show anear-perfect success rate (almost allOrange), reflecting the maturity of theprogram during the Starlink deploymentphase.

GTO Consistency & Challenge: GTO missionsappear consistently throughout the entiretimeline (from Flight 0 to 90). However,unlike other orbits, GTO retains a mix offailures (Blue dots) even in the middle flightnumbers, reinforcing that these high-energymissions remain technically demanding.

ISS Reliability Growth: Missions to the ISSspan most of the timeline. You can visuallytrace the improvement: early flights showsome failures/mixed results, but later flightsto the ISS become consistently successful.

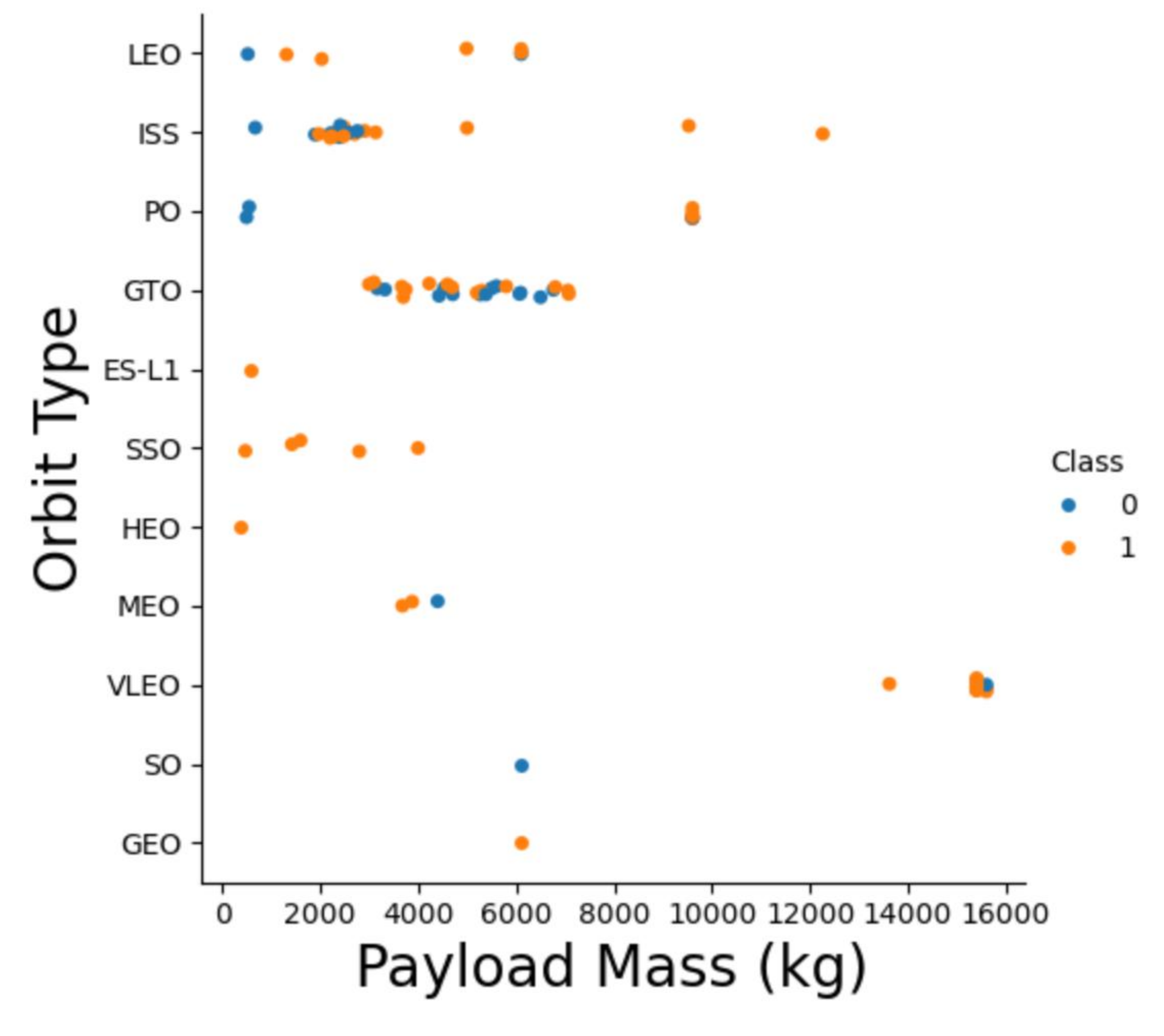


Payload vs. Orbit Type

Heavy Payloads & VLEO: The heaviestpayloads (>14,000 kg) are almostexclusively targeting VLEO (Very LowEarth Orbit). These missions (likelyStarlink) show a near-perfect successrate (Orange dots), indicating that heavymass does not negatively impact landingreliability for this orbit.

The GTO Cluster: The 2,000 kg – 6,000 kgrange is dominated by GTO missions.This specific mass/orbit combinationshows the highest variance in outcomes(mixed Blue and Orange), reinforcing thatmid-weight GTO missions remain acomplex sector for recovery.

Polar/SSO Constraints: Missions to Polar(PO) and Sun-Synchronous (SSO) orbits arestrictly limited to lighter payloads(generally < 5,000 kg), and they displaygenerally high success rates.

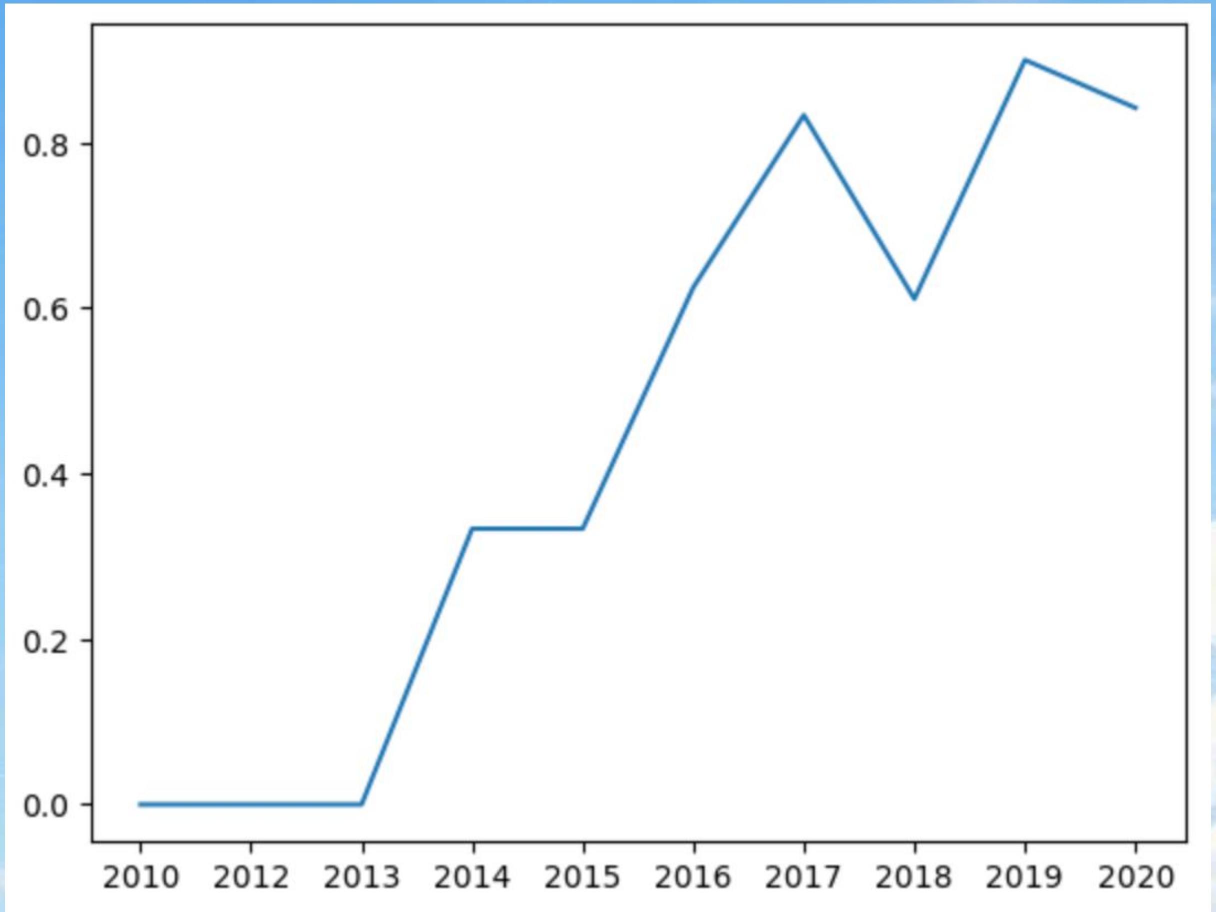


Launch Success Yearly Trend

Rapid Maturity: The chart visualizes a massive "learning curve." SpaceX went from a 0% landing success rate in 2013 to over 90% reliability by 2019. This confirms the technology matured incredibly fast.

The Turning Point (2013): The sharp rise starting in 2013 marks the transition from purely "expendable" launches to the first successful recovery attempts.

Stabilization: Despite a dip in 2018 (likely due to new hardware iterations or difficult mission profiles), the success rate has stabilized above 80% in the most recent years (2019–2020), proving the system is now operationally robust.



All Launch Site Names

Query Objective: Executed SELECT DISTINCT to isolate unique launch locations from the raw dataset and verify data consistency.

Key Findings: The query confirmed that SpaceX operates primarily from three major spaceports: Cape Canaveral (CCAFS), Vandenberg Space Force Base (VAFB), and Kennedy Space Center (KSC).

Data Note: The output reveals slight naming variations (e.g., CCAFS LC-40 vs. CCAFS SLC-40), highlighting the need for data standardization during the cleaning phase.

```
[10]: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
 * sqlite:///my_data1.db
Done.
```

[10]:

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

Launch Site Names Begin with 'CCA'

Query Objective: Executed a pattern-matching query (LIKE 'CCA%') to isolate and inspect the first 5 records for Cape Canaveral launch sites.

Data Validation: Verified the integrity of the dataset columns (Date, Payload Mass, Orbit) and confirmed that early missions (2010–2013) often resulted in "Failure (parachute)" or "No attempt," providing a baseline for measuring future success.

[11]:
```
%%sql
SELECT *
FROM SPACEXTABLE
WHERE "Launch_Site" LIKE 'CCA%'
LIMIT 5;
```

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

[11]:

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

Total Payload Mass

Query Objective: Utilized the SUM() function combined with a WHERE clause to calculate the total payload mass specifically for NASA (CRS) missions.

Key Finding: The query revealed that SpaceX has delivered 45,596 kg of cargo for NASA's Commercial Resupply Services in this dataset.

This validates the ability to extract specific quantitative metrics for key stakeholders from the raw data.

[12]:

```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") as Total_Payload_Mass_KG
FROM SPACEXTABLE
WHERE "Customer" LIKE 'NASA (CRS)';
```

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

[12]:

Total_Payload_Mass_KG
45596

Average Payload Mass by F9 v1.1

Query Objective: Calculated the average payload mass (AVG) specifically for the F9 v1.1 booster version to establish a performance baseline.

Key Finding: The query determined that the F9 v1.1 typically carried an average payload of approximately 2,928 kg.

This metric provides a benchmark for comparing against newer booster versions to quantify improvements in lift capacity.

[13]:
```
%%sql select "Booster_Version", avg(PAYLOAD_MASS__KG_) as AVG_PAYLOAD
from SPACEXTBL
where "Booster_Version"='F9 v1.1'
```

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

[13]:

Booster_Version	AVG_PAYLOAD
F9 v1.1	2928.4

First Successful Ground Landing Date

Query Objective: Executed MIN(Date) to pinpoint the exact date of the very first successful landing on a ground pad.

Key Finding: The query returned December 22, 2015.

This marks a major turning point in the program (the Orbcomm OG2 mission), proving for the first time that a Falcon 9 booster could successfully return to the launch site.

```
[14]: %%sql select min("Date") as Date, "Landing_Outcome", "Mission_Outcome"
          from SPACEXTBL
      where Landing_Outcome like '%ground%' and Mission_Outcome='Success'

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

[14]:

Date	Landing_Outcome	Mission_Outcome
2015-12-22	Success (ground pad)	Success

Successful Drone Ship Landing with Payload between 4000 and 6000

Query Objective: Executed a multi-conditional query to isolate boosters that achieved "Success (drone ship)" while carrying a payload specifically between 4,000 kg and 6,000 kg.

Key Finding: identified four specific F9 FT (Full Thrust) boosters (e.g., B1022, B1026) that met these strict operational parameters.

This demonstrates the capability to perform granular analysis, identifying the specific hardware used for mid-weight, high-difficulty sea landings.

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

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

[15]:

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Query Objective: Calculated the aggregate counts for "Success" vs. "Failure" mission outcomes to establish a baseline reliability metric.

Key Finding: The dataset is overwhelmingly positive, containing 98 Successful Missions and only 1 Failed Mission.

This extreme class imbalance (98:1) highlights the high reliability of the Falcon 9 vehicle but also signals a challenge for predictive modeling (potential bias toward the majority class).

```
[16]: %%sql select (select count(*) from SPACEXTBL where "Mission_Outcome"='Success') as "Successful Missions",
      (select count(*) from SPACEXTBL where "Mission_Outcome" like '%Failure%') as "Failed Missions"
      from SPACEXTBL
      limit 1

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

[16]:

Successful Missions	Failed Missions
98	1

Boosters Carried Maximum Payload

Query Objective: Used a subquery (SELECT MAX...) to identify which specific booster versions have carried the heaviest payloads in the dataset.

Key Finding: The query revealed that multiple Falcon 9 Block 5 (B5) boosters share the record for the maximum payload of 15,600 kg.

This confirms 15,600 kg as the operational peak payload for this dataset (likely Starlink missions).

[17]:
```
%%sql select "Booster_Version", "PAYLOAD_MASS__KG_"
from SPACEXTBL
where PAYLOAD_MASS__KG_ in (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

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

[17]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

Query Objective: Utilized substr filtering to isolate specific landing failures on drone ships occurring strictly within the year 2015.

Key Finding: The query identified exactly two failures: January (Booster B1012) and April (Booster B1015), both launching from CCAFS LC-40.

This granular view helps correlate specific hardware versions (F9 v1.1) with landing anomalies during the program's critical developmental phase.

[18]:

```
%%sql
SELECT
    CASE substr(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
        ELSE 'Unknown'
    END as Month_Name,
    "Landing_Outcome",
    "Booster_Version",
    "Launch_Site"
FROM SPACEXTABLE
WHERE substr(Date, 1, 4) = '2015'
  AND "Landing_Outcome" = 'Failure (drone ship)';
```

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

[18]:

Month_Name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Query Objective: Ranked the frequency of landing outcomes during the program's developmental phase (2010–2017) to identify common failure modes.

Key Finding: "No attempt" was the most common result (10 counts), reflecting the early era where recovery was not yet a primary mission objective.

Reliability Snapshot: The data reveals the learning curve for sea landings, showing an exact 50/50 split between "Success (drone ship)" (5) and "Failure (drone ship)" (5) during this period.

[19]:
```
%%sql
SELECT
    "Landing_Outcome",
    COUNT(*) as Outcome_Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
  AND "Landing_Outcome" IS NOT NULL
GROUP BY "Landing_Outcome"
ORDER BY Outcome_Count DESC;
```

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

[19]:

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

Launch Sites Proximities Analysis

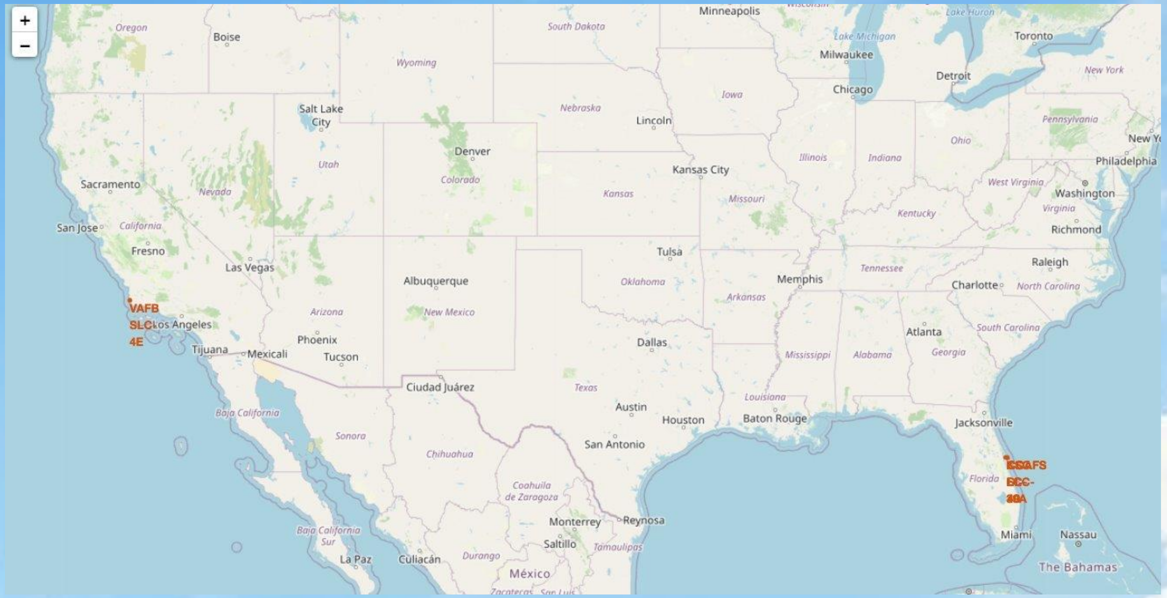
Strategic Coastal Launch Locations

West Coast (VAFB): The single marker in California is Vandenberg Space Force Base. It is positioned for launches requiring Polar Orbits, allowing rockets to fly south over the ocean.

East Coast Cluster (CCAFS & KSC): The two overlapping markers in Florida are Cape Canaveral and Kennedy Space Center. They are adjacent to each other and serve as the primary hub for Equatorial and ISS missions.

Key Finding: Proximity to Coastlines

Safety Protocol: All launch sites are located immediately next to the ocean. This ensures that flight paths and falling debris (stages/fairings) remain over unpopulated water, minimizing risk to populated areas.



Site-Specific Performance & Reliability

<u>Color-Coded Indicators</u>

Green Markers: Represent Successful landings.

Red Markers: Represent Failed landing attempts.

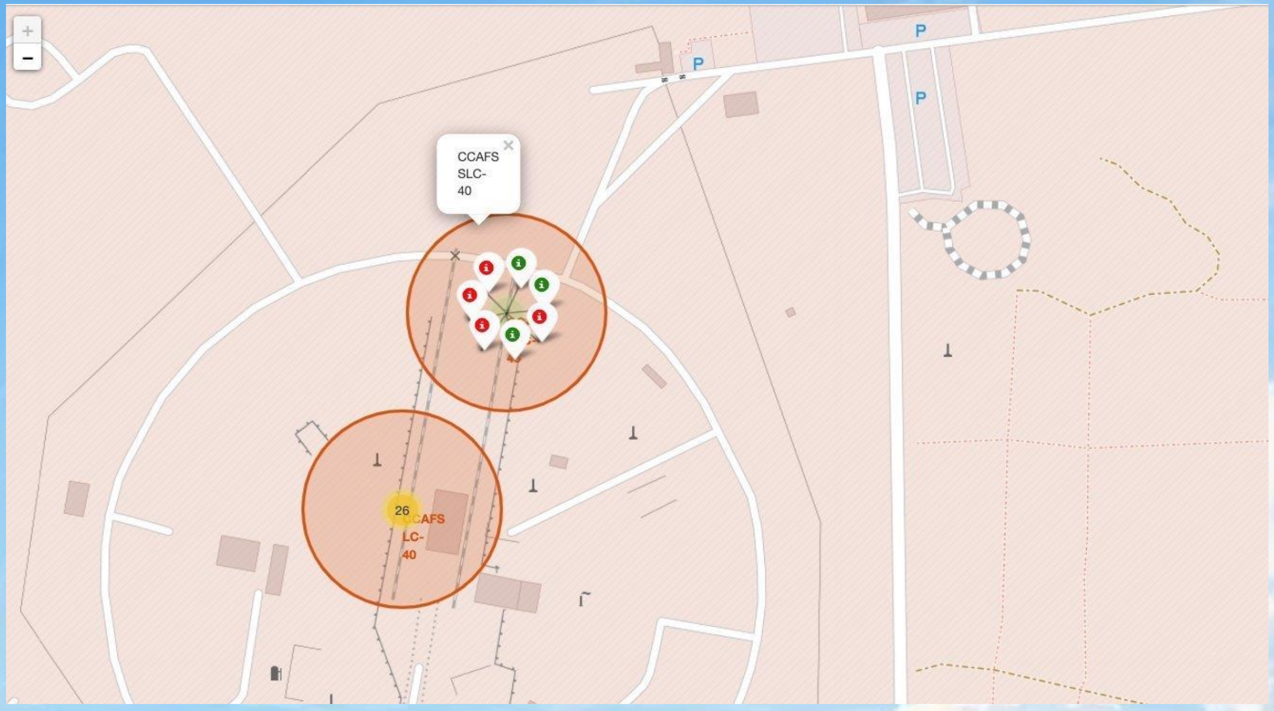
This allows stakeholders to instantly assess the historical reliability of a specific launch site just by glancing at the color ratio.

<u>High-Density Activity</u>

Marker Clusters: The yellow circle with the number "26" indicates a cluster of 26 launches at this location. This visually confirms CCAFS SLC-40 as the program's "workhorse" site with the highest mission volume.

<u>Site Analysis (CCAFS SLC-40)</u>

Mixed History: The visible presence of both Red and Green markers reflects this site's long operational history, capturing both the early experimental failures and the consistent successes of the modern Block 5 era.



Balancing Logistics & Safety

Infrastructure & Logistics

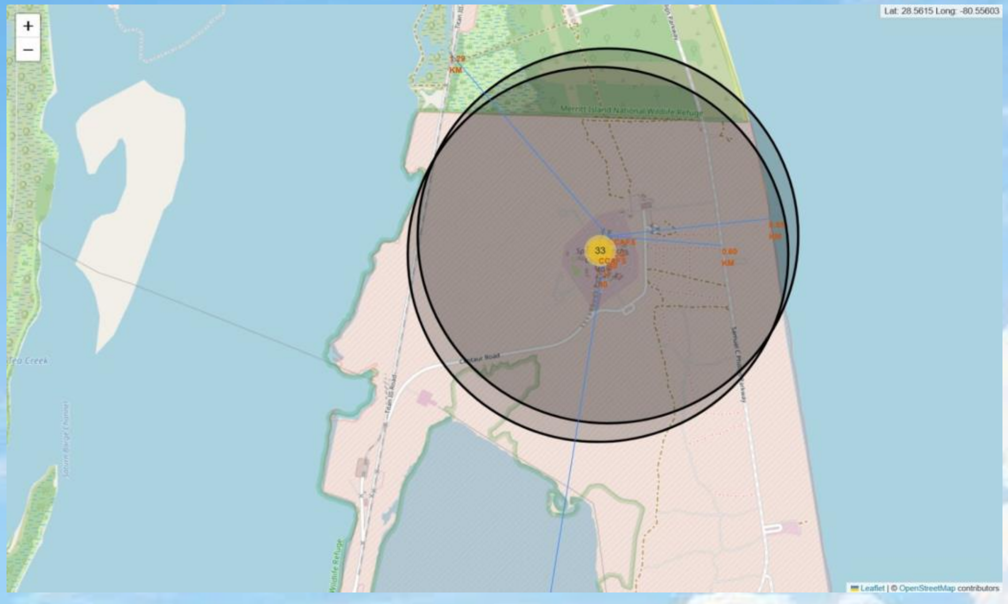
Transport Access: The blue measurement lines confirm the site is located within 1 km of both Railways and Highways. This close proximity is essential for transporting heavy rocket stages and equipment to the pad.

Safety Buffers

Population Isolation: The analysis shows the site is significantly distant from the nearest major city (Cape Canaveral). This safety buffer protects populated areas from noise and potential launch hazards.

Visual Tool

Interactive Markers: The map creates a visual "fencing" zone, allowing analysts to instantly verify that the launch site meets all regulatory safety and logistical distance requirements.



Build a Dashboard with Plotly Dash

Contribution to Mission Success by Launch Site

Leading Success Contributor

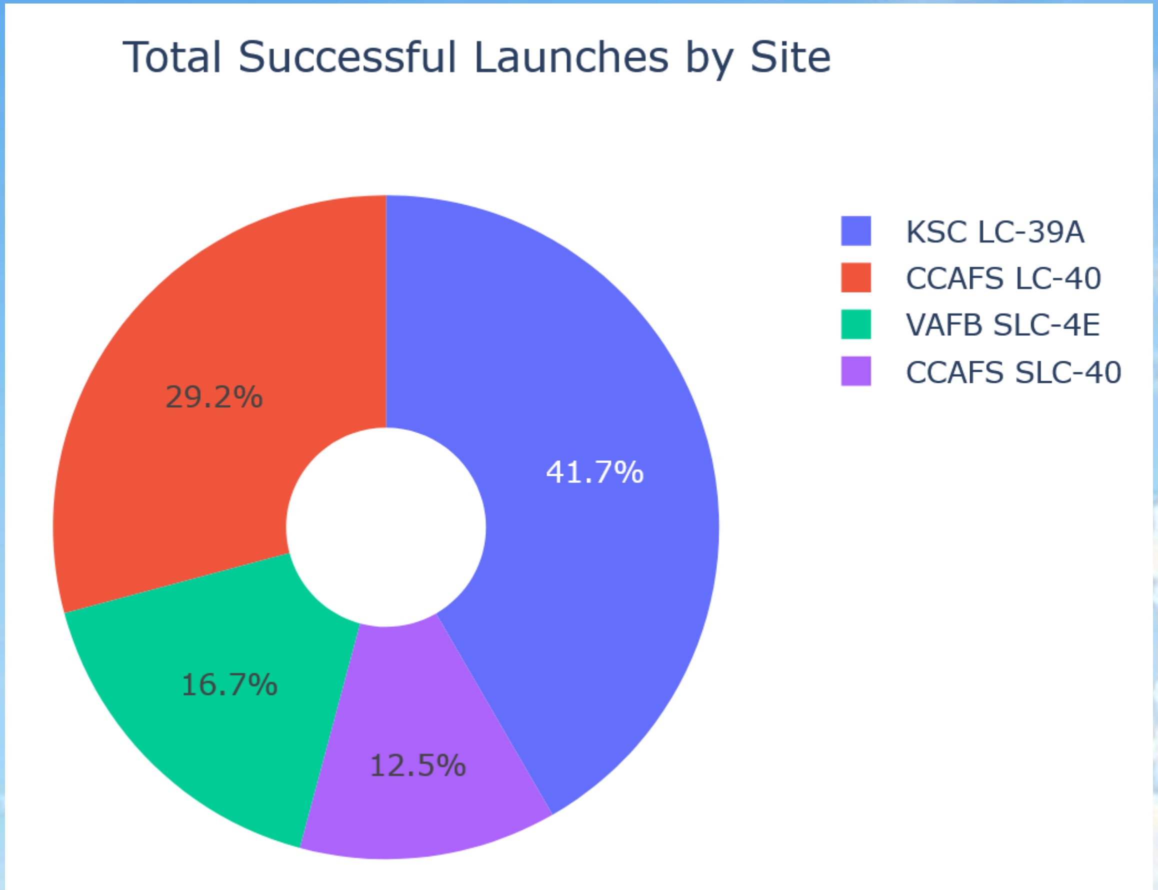
Kennedy Space Center (KSC LC-39A) is the single largest contributor to mission success in this dataset, accounting for 41.7% of all successful landings.

Cape Canaveral Operations

The chart displays two distinct slices for Cape Canaveral (CCAFS LC-40 at 29.2% and CCAFS SLC-40 at 12.5%) due to historical naming conventions. Combined, the Cape Canaveral complex accounts for 41.7% of successes, effectively tying with KSC.

West Coast Contribution

VAFB SLC-4E contributes 16.7% of the total successes. This smaller share reflects its specialized nature for Polar orbits rather than a lack of reliability.



Operational Performance: KSC LC-39A

High Success Rate

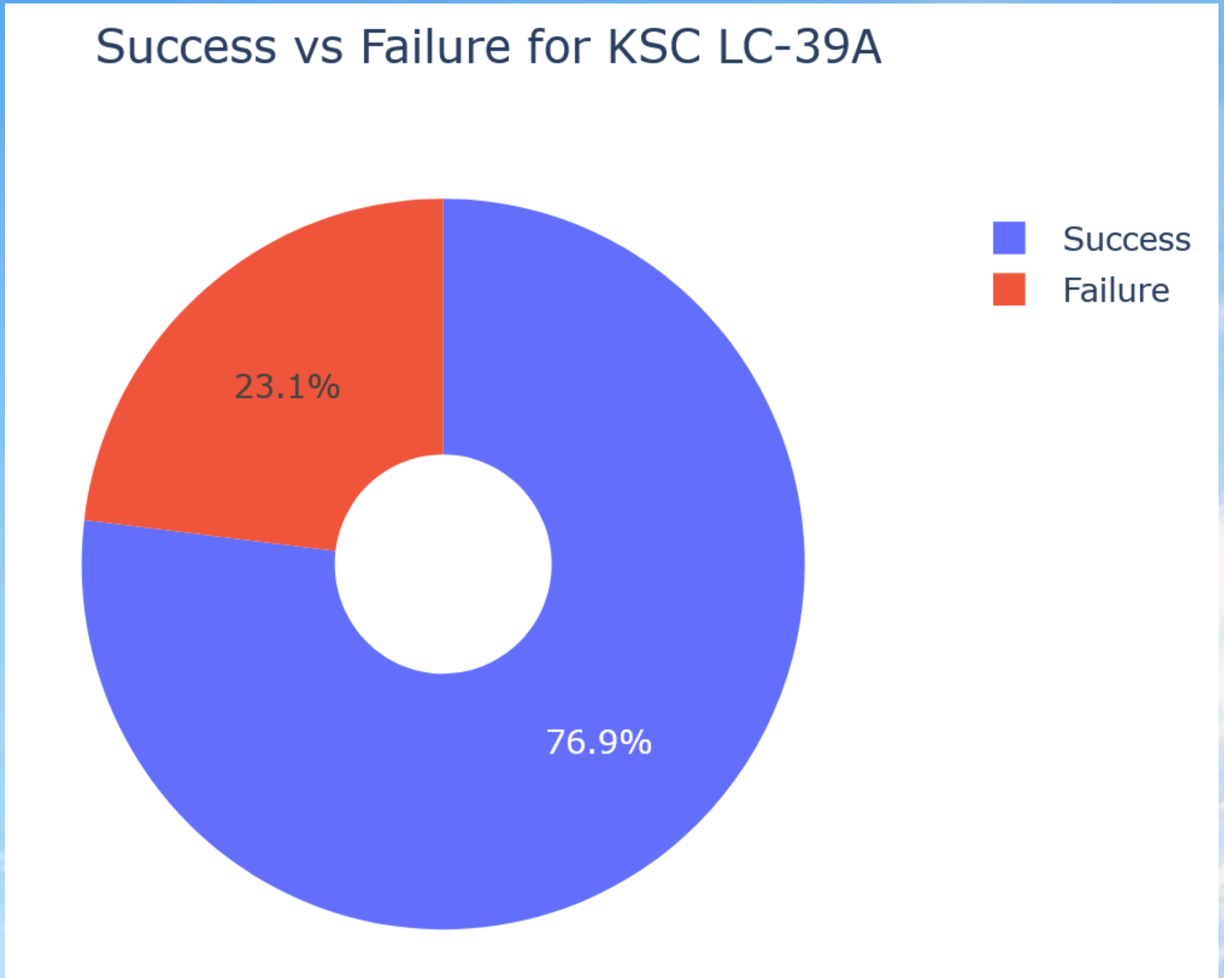
The chart reveals that Kennedy Space Center (KSC LC-39A) achieves a high success rate of 76.9% (Blue area). This confirms it is not only the busiest site but also highly reliable for landing recovery.

Failure Analysis

The 23.1% failure rate (Red area) represents historical data that includes early experimental landings. This "risk margin" is critical data for predictive models to understand where and why failures occurred at this specific location.

Strategic Importance

With over three-quarters of its missions resulting in a successful landing, KSC LC-39A serves as the performance benchmark against which other launch sites are compared.



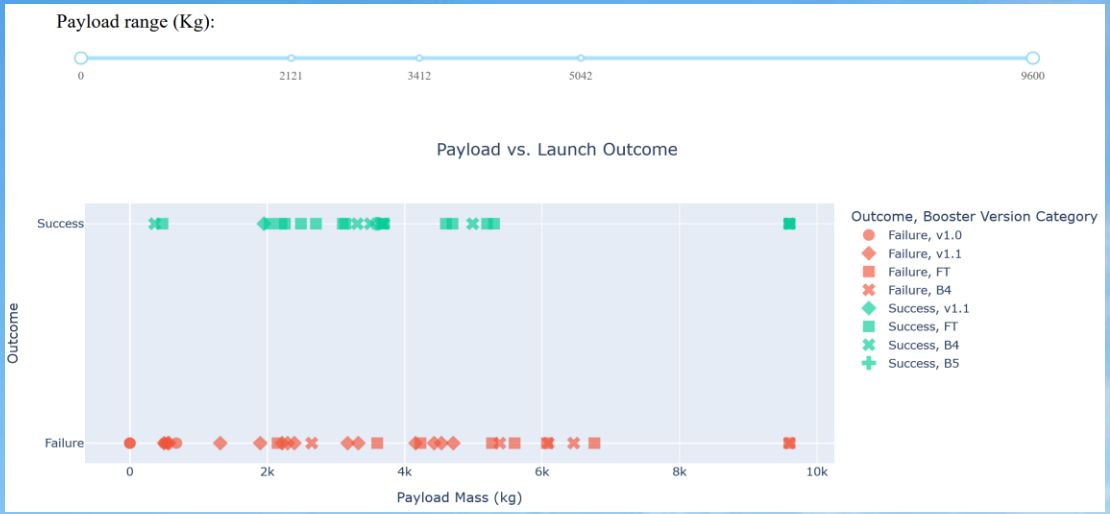
Payload Mass vs. Launch Outcome

The "Sweet Spot" (2,000kg – 5,500kg)

The filtered view confirms that the payload range between 2,000 kg and 5,500 kg represents a "high reliability zone." This window contains the highest density of successful landings (Green markers).

The "Game Changer": Falcon 9 FT

The Falcon 9 FT (represented by Diamonds) dominates this chart. While it had a few early failures as SpaceX pushed boundaries, the overwhelming number of Green Diamonds proves this was the version that made booster recovery routine.



Predictive Analysis (Classification)

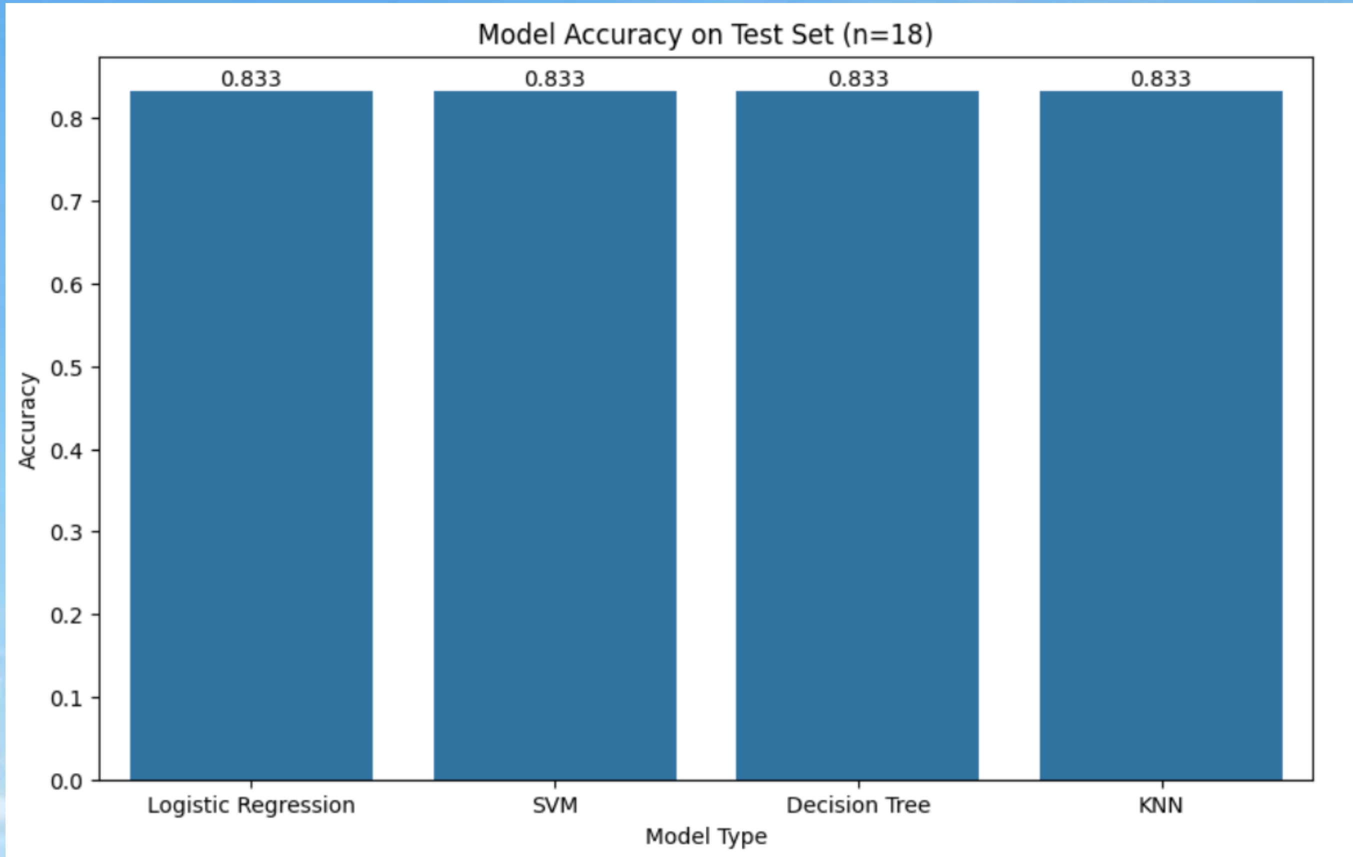
Classification Accuracy

Uniform Accuracy

All four classification models—Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN)—achieved an identical accuracy score of 83.3% on the test dataset.

Model Robustness

The fact that parametric models (like Logistic Regression) and non-parametric models (like KNN/Decision Trees) performed equally suggests that the features extracted are robust and the decision boundary is relatively stable regardless of the mathematical approach used.



Confusion Matrix

Perfect Recall for Successes

No Missed Opportunities: The model correctly identified 12 out of 12 successful landings (Bottom Right).

Zero False Negatives: The value of 0 in the Bottom Left quadrant confirms that the model never predicted a crash for a mission that actually succeeded.

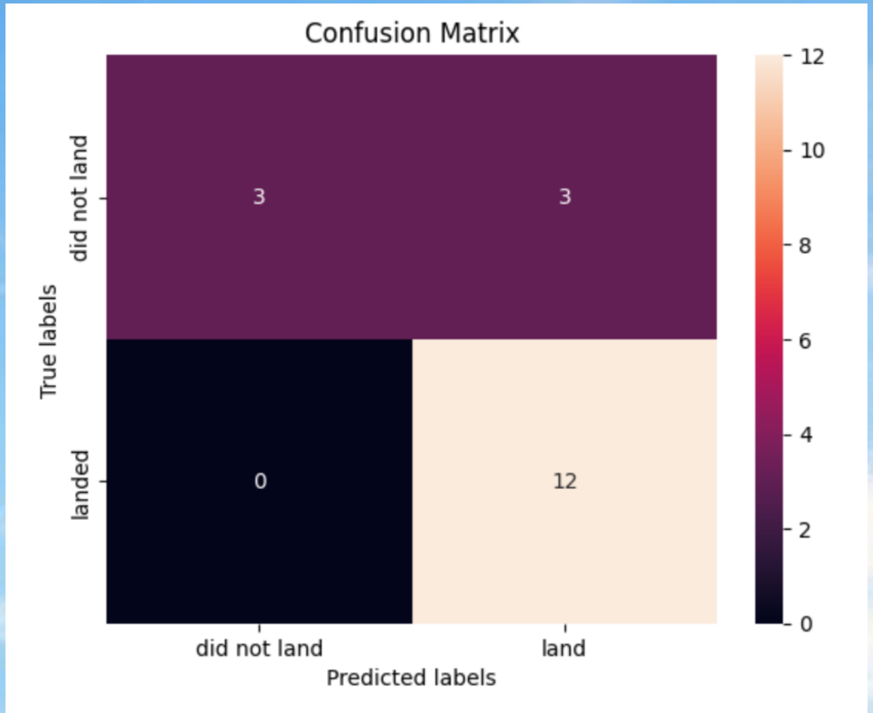
The Problem Area: False Positives

Optimism Bias: The main source of error is the 3 False Positives (Top Right). The model predicted these 3 missions would "land," but they actually "did not land."

Risk Implication: This indicates the model is slightly "optimistic." In a real-world scenario, this is the riskier type of error (predicting a safe landing when a crash is imminent).

Root Cause

Class Imbalance: This error pattern (predicting the majority class "Success" too often) reinforces the earlier finding that the dataset is heavily skewed toward successful missions. The model struggles to learn the subtle patterns of failure because it has seen so few examples of them.

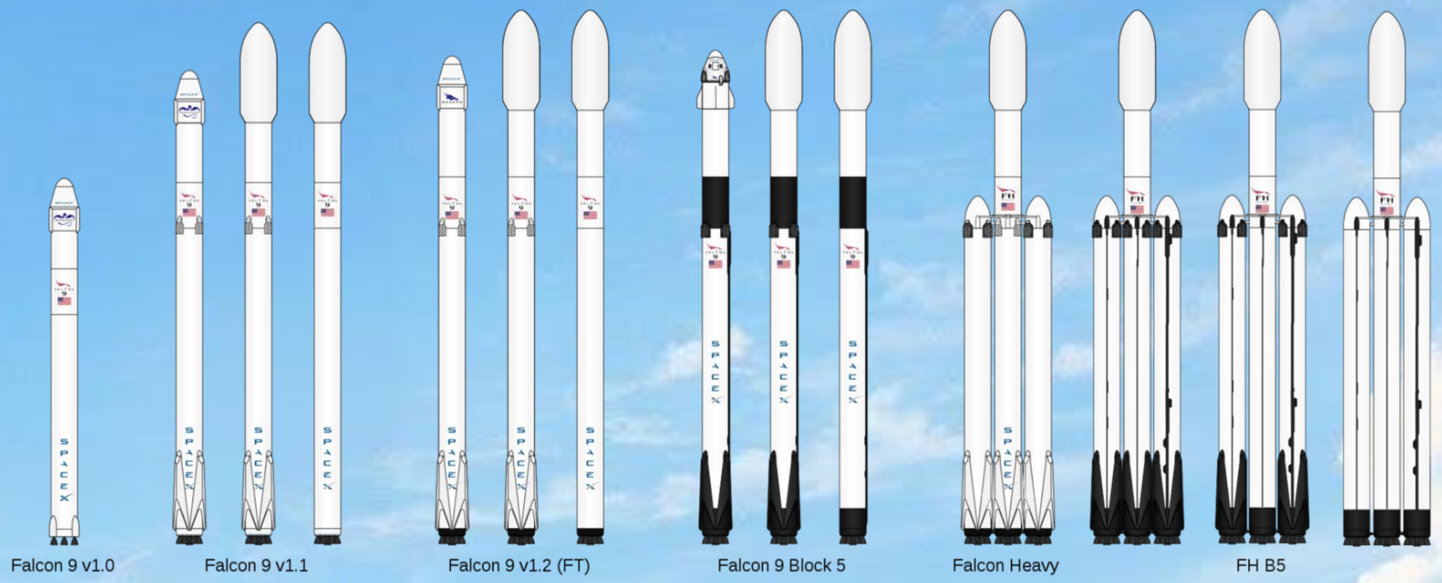


Conclusions

Validated Reusability: The trend analysis confirms that SpaceX has successfully transitioned from an experimental phase (2013) to a highly stable operational phase (2020), validating the economic viability of rocket reusability.

Strategic Optimization: The clear segmentation of launch sites—using VAFB for lighter polar missions and KSC for heavy commercial payloads—demonstrates highly optimized logistical planning that maximizes success rates.

Predictive Baseline: Achieving 83.33% accuracy across all models establishes a strong baseline. It proves that landing success is not random, but is statistically predictable based on specific flight parameters (Orbit, Payload, and Site).



Appendix

https://github.com/eatcodedesign/Data-Science-Machine-Learning-Project/tree/main

Thank you!