IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Chetan Sai Abhishek Injavarapu
02/20/2026

Outline

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

Executive Summary

- **Summary of Methodologies**
- This project implemented a complete end-to-end data science workflow to predict the successful landing of SpaceX Falcon 9 first-stage boosters.
- **Module 1 – Data Collection:**
  Launch data was collected using the SpaceX REST API, and historical landing outcomes were extracted through web scraping from Wikipedia. The raw JSON and HTML data were transformed into structured Pandas Data Frames for analysis.
- **Module 2 – Data Wrangling & Preparation:**
  The dataset was cleaned by removing irrelevant variables and handling missing values. Categorical features such as launch site and orbit were encoded, and a binary target variable was created to represent landing success (1 = success, 0 = failure).
- **Module 3 – Exploratory Data Analysis & SQL:**
  SQLite was used to perform structured queries on launch counts, payload distributions, and success rates. Visual and statistical analysis identified key relationships between payload mass, orbit type, launch site, and landing outcomes.
- **Module 4 – Predictive Modeling:**
  Multiple classification models — Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) — were developed and evaluated. Model validation was conducted using train-test split and cross-validation, and hyperparameter tuning was performed to improve performance and select the best model.

Executive Summary

- **Summary of All Results**
- Landing success rates improved significantly over time.
- Certain launch sites demonstrated higher landing reliability.
- Payload mass showed an inverse relationship with landing success probability.
- Orbit type influenced landing outcomes.
- Among all models tested, **Support Vector Machine (SVM)** achieved the highest predictive accuracy (~84%).
- Machine learning models successfully demonstrated that booster landing outcomes can be predicted using historical launch data.

Introduction

- SpaceX has revolutionized the aerospace industry by developing reusable Falcon 9 rockets. The ability to successfully land and reuse the first-stage booster significantly reduces launch costs and increases operational efficiency.
- However, not all landings are successful. Landing outcomes depend on multiple technical and operational factors such as payload mass, launch site, orbit type, and mission profile.
- Understanding and predicting landing success is critical for:
- Estimating mission reliability
- Assessing cost efficiency of reusability
- Supporting data-driven aerospace decision-making
- This project leverages historical launch data to analyze these factors and build predictive models for landing outcomes.
- **Problems to Find Answers**
- What factors most influence Falcon 9 first-stage landing success?
- How does payload mass affect the probability of a successful landing?
- Do launch sites and orbit types impact landing outcomes?
- Has landing success improved over time?
- Can machine learning models accurately predict landing success based on historical data?

Section 1

Methodology

Methodology

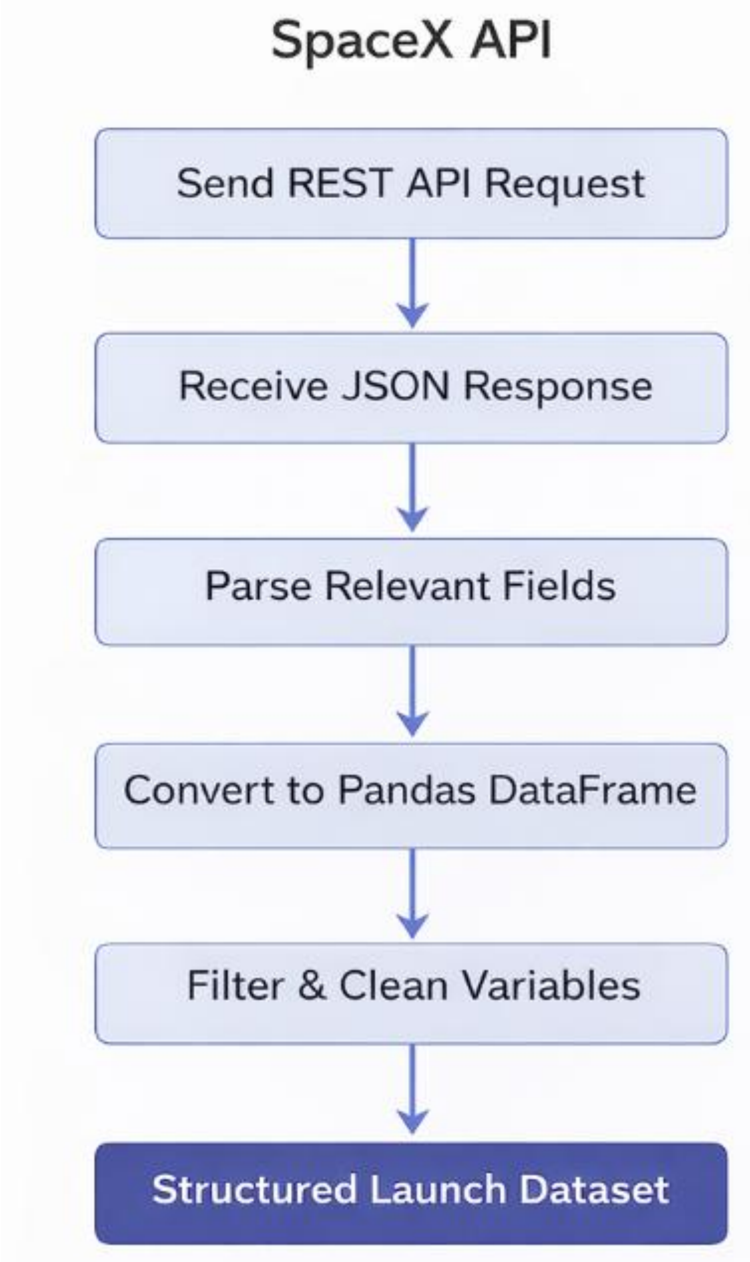
- **Executive Summary**
- A structured data science workflow was followed: data collection, preprocessing, exploratory analysis, interactive visualization, and predictive modeling to estimate Falcon 9 landing success.
- **Data Collection**
- Launch data was gathered using the SpaceX REST API and historical landing outcomes were extracted through web scraping. The data was converted into structured Pandas Data Frames and merged into a unified dataset.
- **Data Wrangling & Processing**
- The dataset was cleaned, missing values were handled, categorical variables were encoded, and a binary landing success target variable was created.
- **Exploratory Data Analysis (EDA)**
- SQL queries and visualizations were used to analyze launch trends, payload impact, orbit influence, and success rates over time.
- **Interactive Visual Analytics**
- Folium and Plotly Dash were used to create interactive maps and dashboards for launch site analysis and success rate visualization.
- **Predictive Modeling**
- Classification models (Logistic Regression, SVM, Decision Tree, KNN) were built, tuned, and evaluated using train-test split and cross-validation to identify the best-performing model.

Data Collection

- **Data Sources**
- **SpaceX REST API**
  - Launch details
  - Payload mass
  - Orbit type
  - Launch site
  - Booster version
- **Wikipedia Web Scraping**
  - Historical landing outcomes
  - Booster reuse information

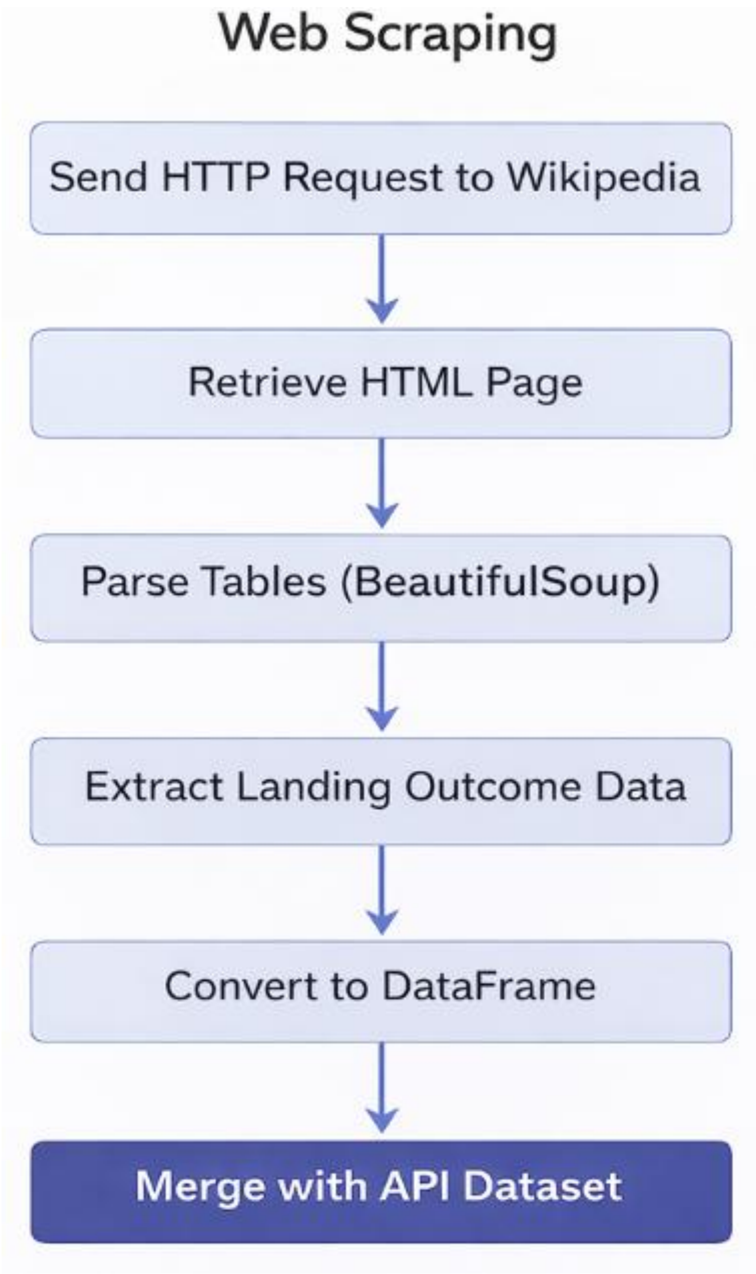
Data Collection – SpaceX API

- Sent HTTP GET requests to SpaceX public REST API endpoints
- Retrieved structured launch records in JSON format
- Parsed nested JSON objects to extract mission-specific variables
- Selected relevant fields: flight number, payload mass, orbit, launch site, booster version, landing outcome
- Flattened nested JSON structures into tabular format
- Used **pandas.json_normalize()** to normalize hierarchical JSON into a structured DataFrame
- Filtered required columns and removed unnecessary attributes
- Generated a clean, structured launch dataset ready for preprocessing
- GitHub Url: https://github.com/chetan-957/IBM-Capstone/blob/main/jupyter-labs-spacex-data-collection-api%20with%20answers.ipynb



Data Collection – Web Scraping

- Sent HTTP request to retrieve Wikipedia launch history page
- Extracted HTML content using requests
- Parsed webpage structure using BeautifulSoup
- Identified and extracted specific launch history tables
- Iterated through table rows and columns to extract landing outcome data
- Cleaned irregular characters and formatting inconsistencies
- Converted extracted table data into structured Pandas DataFrame
- Standardized column names and data types
- Merged scraped landing outcome data with API dataset
- GitHub Url: https://github.com/chetan-957/IBM-Capstone/blob/main/jupyter-labs-webscraping%20Answers.ipynb



Data Wrangling

- Removed irrelevant and redundant variables
- Filtered only Falcon 9 launches
- Handled missing values in payload mass and landing outcome
- Converted categorical variables (orbit, launch site, booster version) into encoded features
- Standardized column names and data types
- Created binary target variable:
    - 1 → Successful landing/0 → Failed / Unsuccessful landing
- Verified dataset consistency and removed duplicates
- Prepared final feature matrix for modeling
- **Technical Processing Details**
- Used drop() to remove unnecessary columns
- Applied filtering conditions to isolate relevant missions
- Used .isnull() and conditional logic for missing value handling
- Converted data types using .astype()
- Applied one-hot encoding using pd.get_dummies()
- Created target variable using logical condition mapping

GitHub Url:https://github.com/chetan-957/IBM-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling%20Answers.ipynb

EDA With Data Visualization

- The exploratory data analysis focused on understanding relationships between mission features and landing success.
- **Launch Site vs Flight Count (Bar Chart)**
- Showed distribution of launches across sites
- Helped identify high-frequency launch locations

Launch Site vs Success Rate (Bar Chart)

- Compared landing success percentages by site
- Identified which sites had higher reliability

Payload Mass vs Landing Outcome (Scatter Plot)

- Examined impact of payload weight on landing success
- Used to observe potential inverse relationship

Payload Mass vs Flight Number (Scatter Plot)

- Analyzed performance trends as SpaceX gained experience
- Helped visualize improvement over time

Orbit Type vs Landing Success (Bar Chart)

- Compared success rates across different orbit categories
- Identified orbit types with higher landing probability

Success Rate Over Time (Line Chart)

- Showed improvement trend in landing performance by year
- Demonstrated operational learning and technological maturity

Why These Charts?

Bar charts → Best for comparing categorical variables (launch site, orbit type)
Scatter plots → Best for analyzing relationships between continuous variables (payload mass vs outcome)
Line charts → Best for trend analysis over time
These visualizations helped identify key predictors for modeling.

GitHub Url:https://github.com/chetan-957/IBM-Capstone/blob/main/eda-dataviz%20Answers.ipynb

EDA With SQL

- **Summary of SQL Queries Performed**
- Created and connected to SQLite database containing SpaceX launch dataset
- Queried total number of launches in the dataset
- Retrieved all launches from specific launch sites (e.g., CCAFS SLC-40)
- Counted number of successful landings
- Calculated landing success rate using conditional aggregation
- Retrieved distinct booster versions used in missions
- Identified booster versions used for specific payload ranges
- Queried launches with payload mass greater than a specified threshold
- Calculated maximum payload mass for successful landings
- Retrieved missions in specific orbit categories (e.g., GTO, ISS)
- Counted total launches per orbit type
- Analyzed landing outcomes grouped by launch site
- Ordered launches chronologically to examine performance progression
- Used GROUP BY, COUNT(), MAX(), DISTINCT, WHERE, ORDER BY clauses for structured analysis

Key SQL Concepts Used

SELECT statements for data retrieval
WHERE filters for conditional querying
GROUP BY for aggregation analysis
COUNT() and MAX() for summary statistics
DISTINCT for unique value extraction
ORDER BY for sorting results

Insights Derived from SQL Analysis

Certain launch sites had higher success rates
Payload mass influenced landing outcomes
Booster versions varied across orbit types
Landing success improved over time

GitHub Url: https://github.com/chetan-957/IBM-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20answers.ipynb

Build an Interactive Map with Folium

- **Map Objects Created:**
- Base map centered on NASA launch regions
- Circles and text labels to identify each launch site
- MarkerCluster to group multiple launch events and avoid clutter
- Color-coded markers (Green = Successful landing, Red = Failed landing)
- Polylines connecting launch sites to nearest coastline, highway, railway, and city
- Distance labels (in KM) calculated using geospatial distance formula
- **Purpose:**
- Visually analyze geographic distribution of launch sites
- Compare landing success vs. failure spatially
- Evaluate proximity to logistics infrastructure
- Assess safety buffer from populated cities
- **Key Insight:**
  Launch sites are strategically located near transportation networks and coastline for operational efficiency, while maintaining substantial distance from major cities to enhance safety.

GitHub Url: https://github.com/chetan-957/IBM-Capstone/blob/main/lab_jupyter_launch_site_location%20with%20answers.ipynb

Build a Dashboard with Plotly Dash

- **Dashboard Components Added:**
- **Dropdown (Launch Site Selector)**
  Allows users to select *All Sites* or a specific launch site.
- **Pie Chart (Success Distribution)**
  - For *All Sites*: Shows total successful launches by site.
  - For a selected site: Shows Success vs Failure breakdown.
- **Range Slider (Payload Mass Filter)**
  Enables interactive filtering of payload mass range.
- **Scatter Plot (Payload vs Launch Outcome)**
  Displays correlation between payload mass and launch success.
  Colored by **Booster Version Category**.

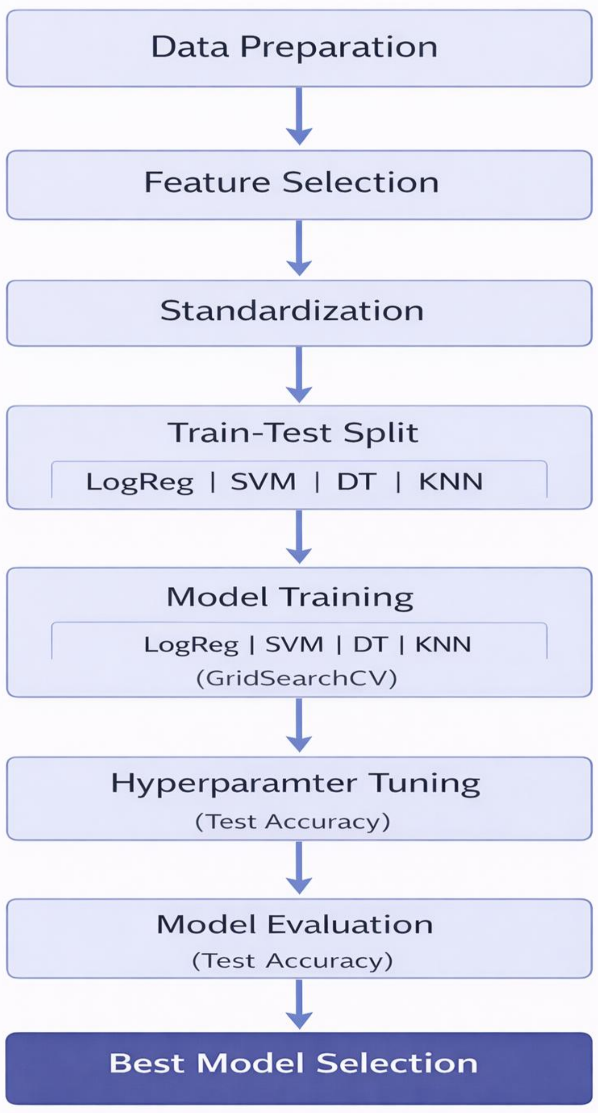
Why These Were Added:

- The **dropdown** enables site-level comparison and filtering.
- The **pie chart** quickly highlights success distribution patterns.
- The **range slider** allows dynamic payload-based analysis.
- The **scatter plot** helps identify relationships between payload weight and launch success probability.
- Together, these components provide interactive, multi-dimensional analysis of SpaceX launch data.

GitHub Url: https://github.com/chetan-957/IBM-Capstone/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)

- **Approach**
- Defined target variable (Landing Success: 0/1)
- Standardized features (StandardScaler)
- Split data into Train (80%) and Test (20%)
- **Models Built**
- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- K-Nearest Neighbors (KNN)
- **Evaluation & Improvement**
- Applied 10-fold Cross-Validation (GridSearchCV)
- Tuned hyperparameters
- Compared test accuracy across models
- **Best Model**
- **Support Vector Machine (Highest Test Accuracy)**
- Strong generalization performance



GitHub Url: https://github.com/chetan-957/IBM-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5%20Answers.ipynb

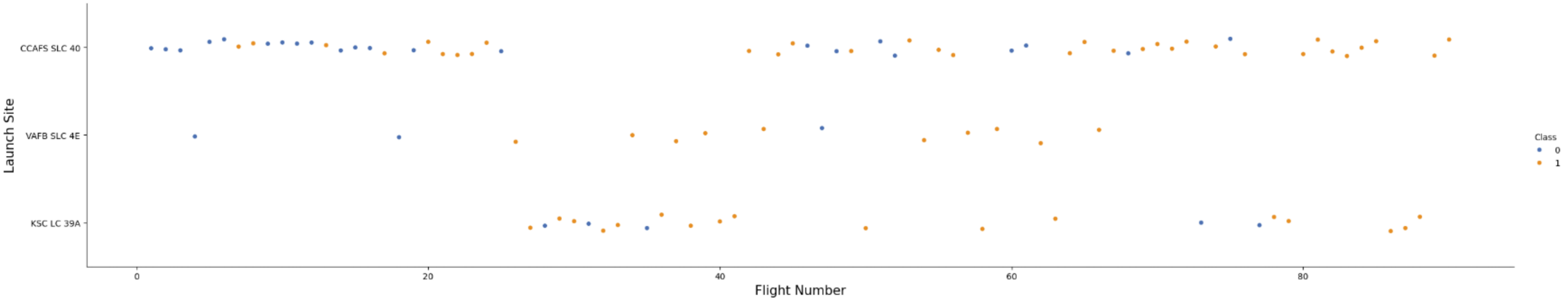
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

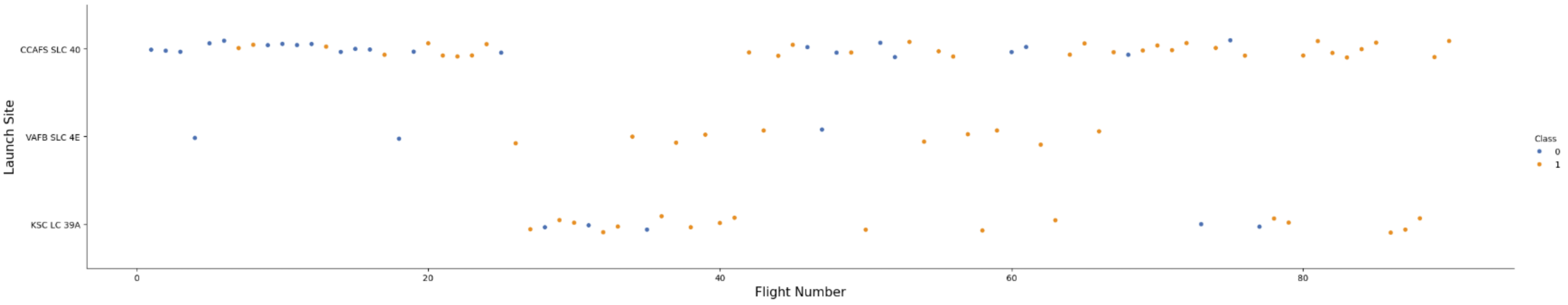
Insights drawn from EDA

Flight Number vs. Launch Site



•**Strong Learning Curve Effect:**
The increase in successful landings over flight number indicates operational improvement and engineering refinement.
•**Experience Improves Reliability:**
Later missions have significantly fewer failures, suggesting process stabilization and better booster reuse strategy.
•**Launch Site Impact:**
CCAFS handled early experimental phases, which explains more early failures.
KSC LC-39A appears during later, more mature stages, showing consistently strong success.
•**Performance Evolution:**
This pattern confirms that landing success is time-dependent and supports including Flight Number as a predictive feature.

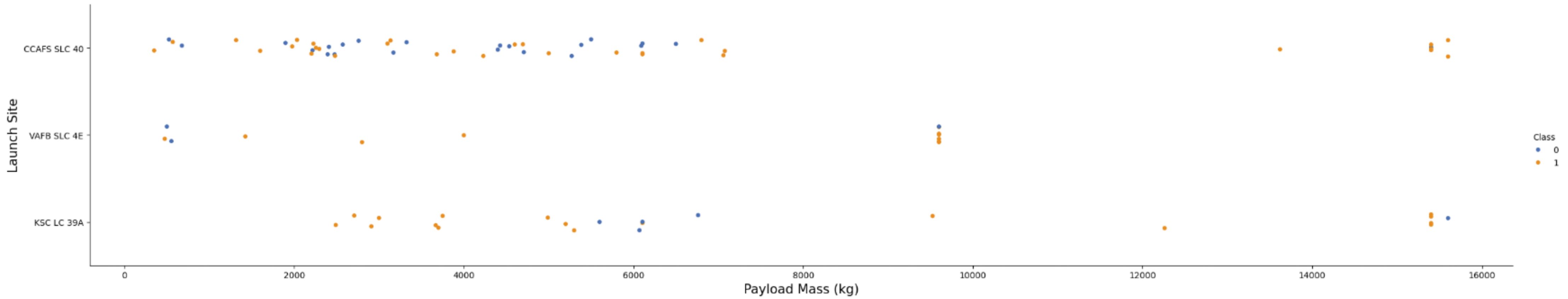
Flight Number vs. Launch Site



Conclusion

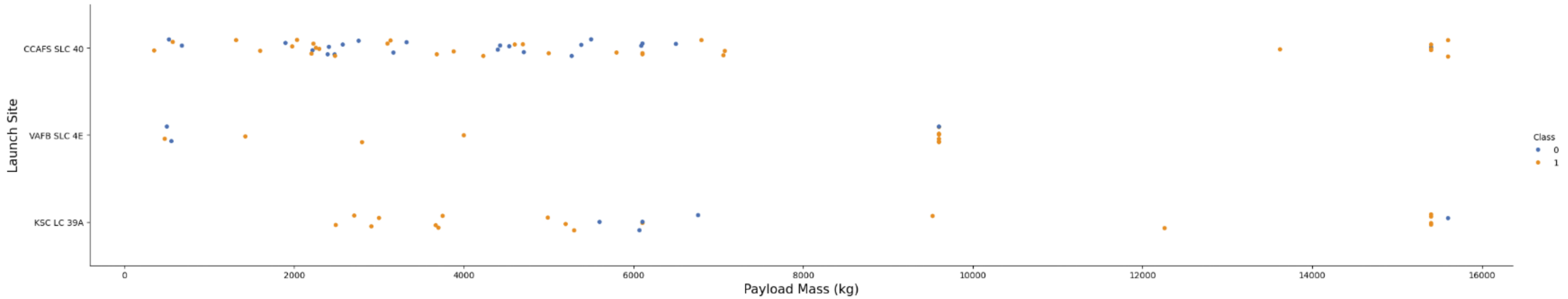
- Landing success improved significantly over time across all launch sites.
  The pattern clearly demonstrates technological advancement, operational learning, and increased reliability in later missions.

Payload vs. Launch Site



•**Launch Site Specialization:**
VAFB appears focused on lighter missions, possibly polar or specific orbital trajectories, which explains absence of heavy payload launches.
•**Heavy Payload Capability:**
KSC LC-39A and CCAFS SLC-40 handle heavier payloads, indicating stronger infrastructure and higher thrust mission profiles.
•**Payload vs Success Pattern:**
There is no clear linear decline in success with increasing payload mass.
In fact, later heavy payload missions show high success rates, suggesting improved booster performance.
•**Operational Maturity Effect:**
High-mass missions mostly appear in later flights, when landing technology had already matured — explaining strong success rates even for heavy payloads.

Payload vs. Launch Site

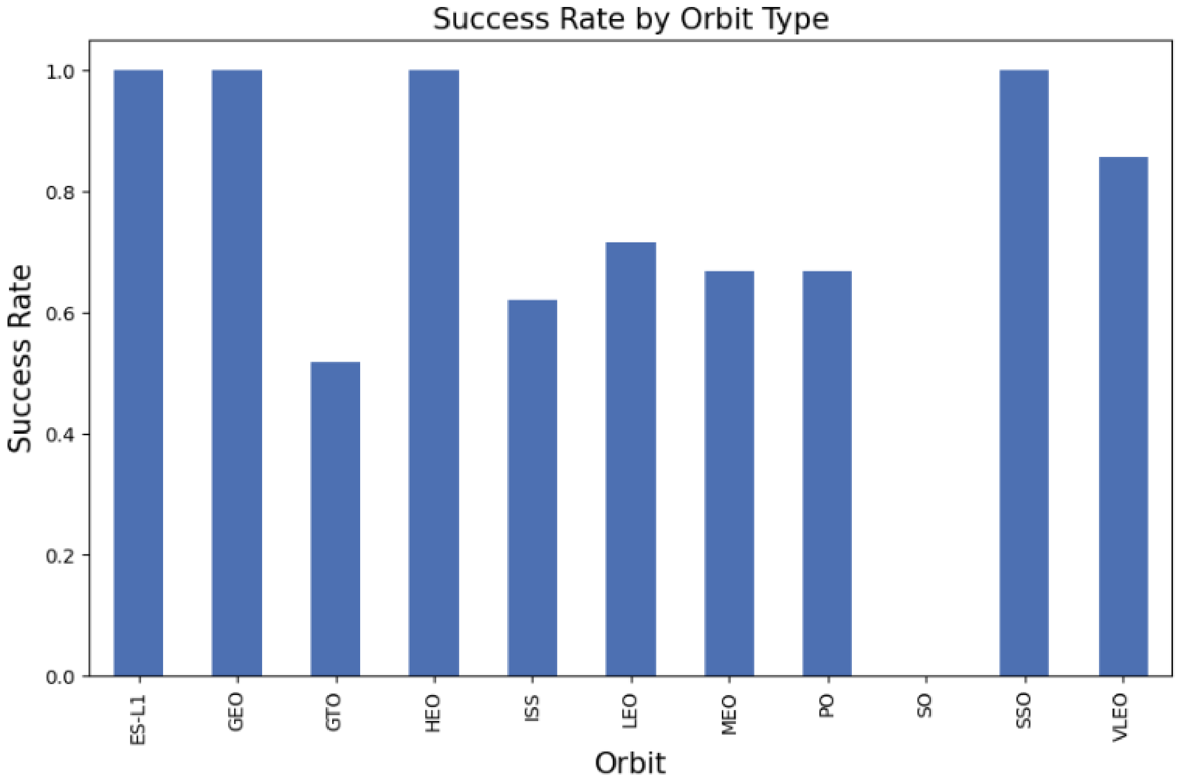


Conclusion

- Payload mass alone does not determine landing failure.
  Instead, landing success appears more strongly influenced by mission maturity and technological evolution rather than payload weight alone.

Success Rate vs. Orbit Type

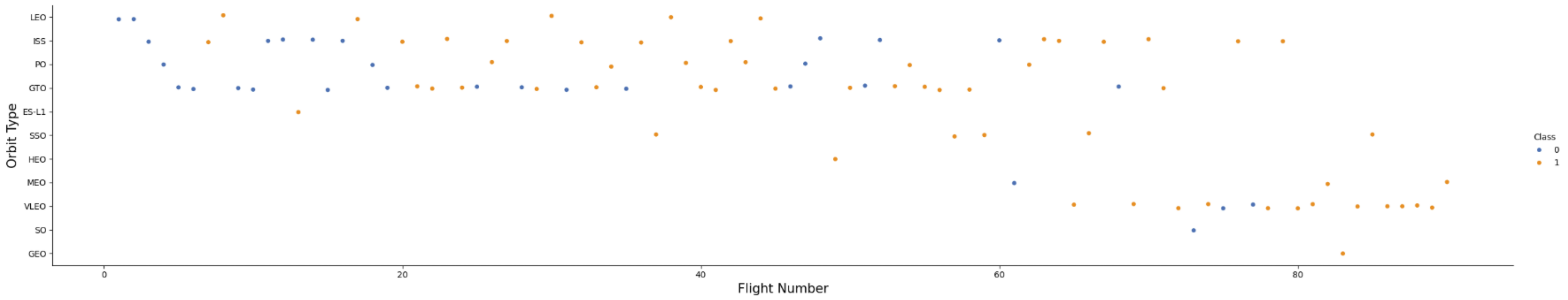
•**Orbit Complexity Matters:**
GTO missions typically require higher energy and more complex flight profiles. The lower success rate reflects higher landing difficulty.
•**Mature Mission Profiles:**
GEO and SSO show very high success rates, suggesting operational stability and well-optimized landing procedures for those trajectories.
•**Operational Learning Curve:**
More common orbits (LEO, ISS) show solid but not perfect success — indicating improvement over time but with earlier failures included in the data.
•**Sample Size Warning:**
Orbits like ES-L1 and HEO may show 100% success, but this likely reflects very few missions. We cannot over-generalize from small counts.



Conclusion:

- Orbit type clearly influences landing success probability.
  Higher-energy orbits (like GTO) introduce greater landing risk, making Orbit a strong predictive feature for classification modeling.

Flight Number vs. Orbit Type



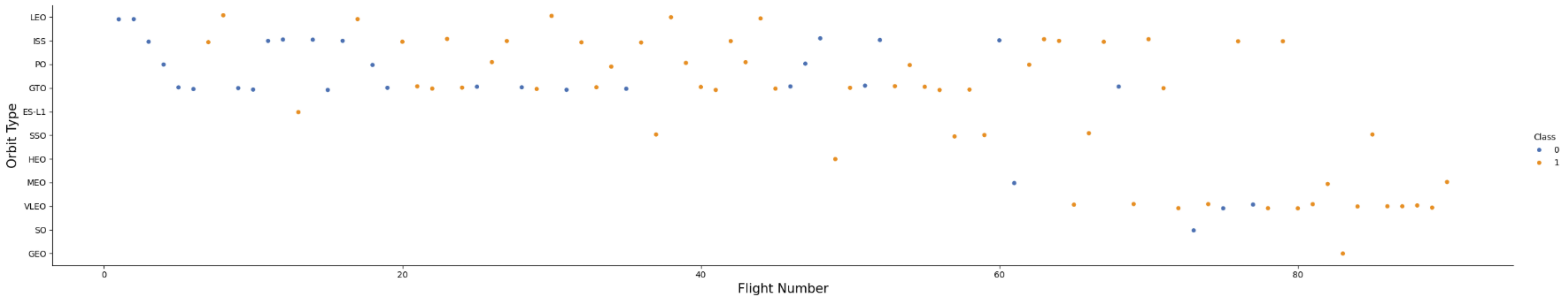
•**Strong Learning Effect in LEO:**
As flight number increases, success rate improves — clear evidence of operational maturity and landing optimization over time.

•**Early Program Volatility:**
Lower flight numbers show more failures across orbit types, consistent with early-stage testing and experimentation.

•**Orbit-Specific Difficulty:**
GTO does not show a strong upward trend in success over time, suggesting intrinsic mission complexity rather than just learning effects.

•**Expansion into New Orbits:**
Later flight numbers include newer orbit categories (VLEO, MEO) with higher success rates, indicating technological advancement and confidence.

Flight Number vs. Orbit Type



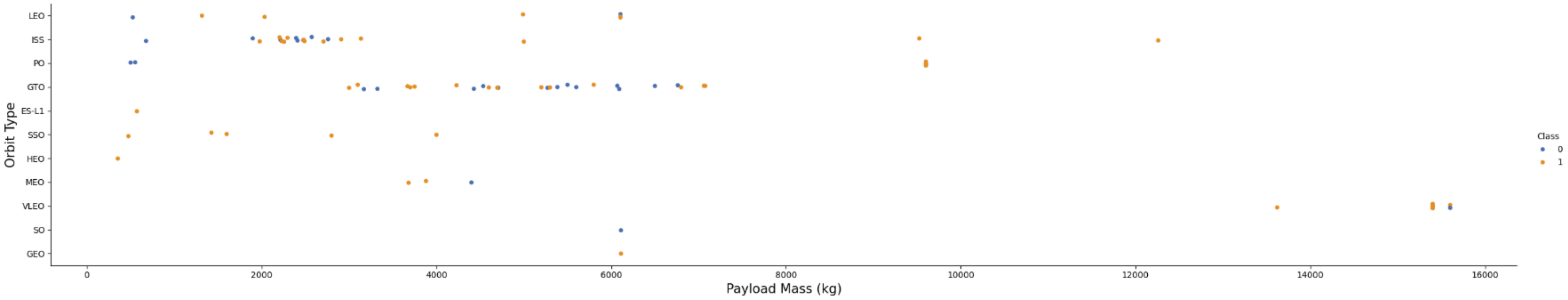
Conclusion:

•Flight number acts as a proxy for experience.

Landing success improves over time, particularly for common orbits like LEO.

This confirms that historical sequence (experience) is an important predictive feature for modeling success.

Payload vs. Orbit Type



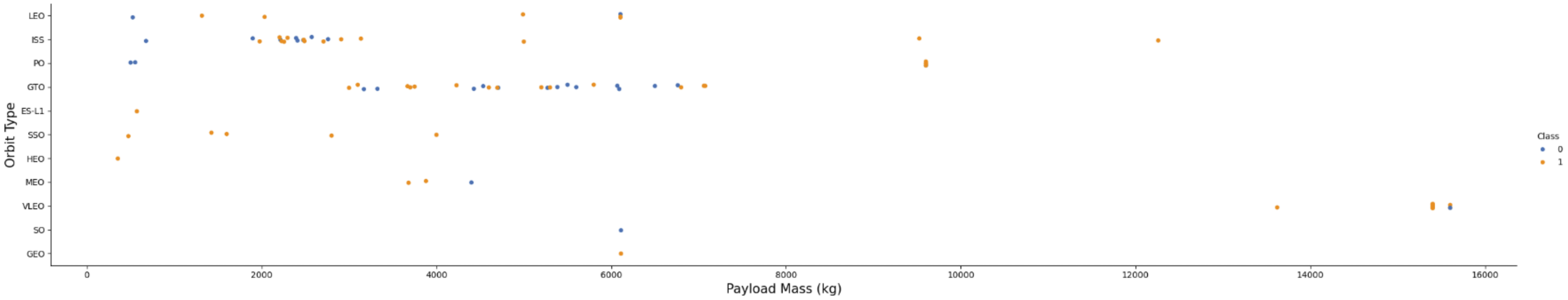
•**Payload Alone Is Not Enough:**
Success does not depend purely on payload mass — orbit type clearly influences landing outcome.

•**LEO & ISS Are Operationally Stable:**
Even as payload mass increases, success remains consistently high — indicating optimized recovery procedures.

•**GTO Is More Challenging:**
At similar payload ranges, GTO shows both success and failure — suggesting mission complexity impacts landing reliability.

•**Heavy Payload Success in Later Missions:**
Very high payload missions appear later in the program and mostly succeed — again showing learning + technological advancement.

Payload vs. Orbit Type



Conclusion:

There is an interaction effect between **Payload Mass and Orbit Type**.

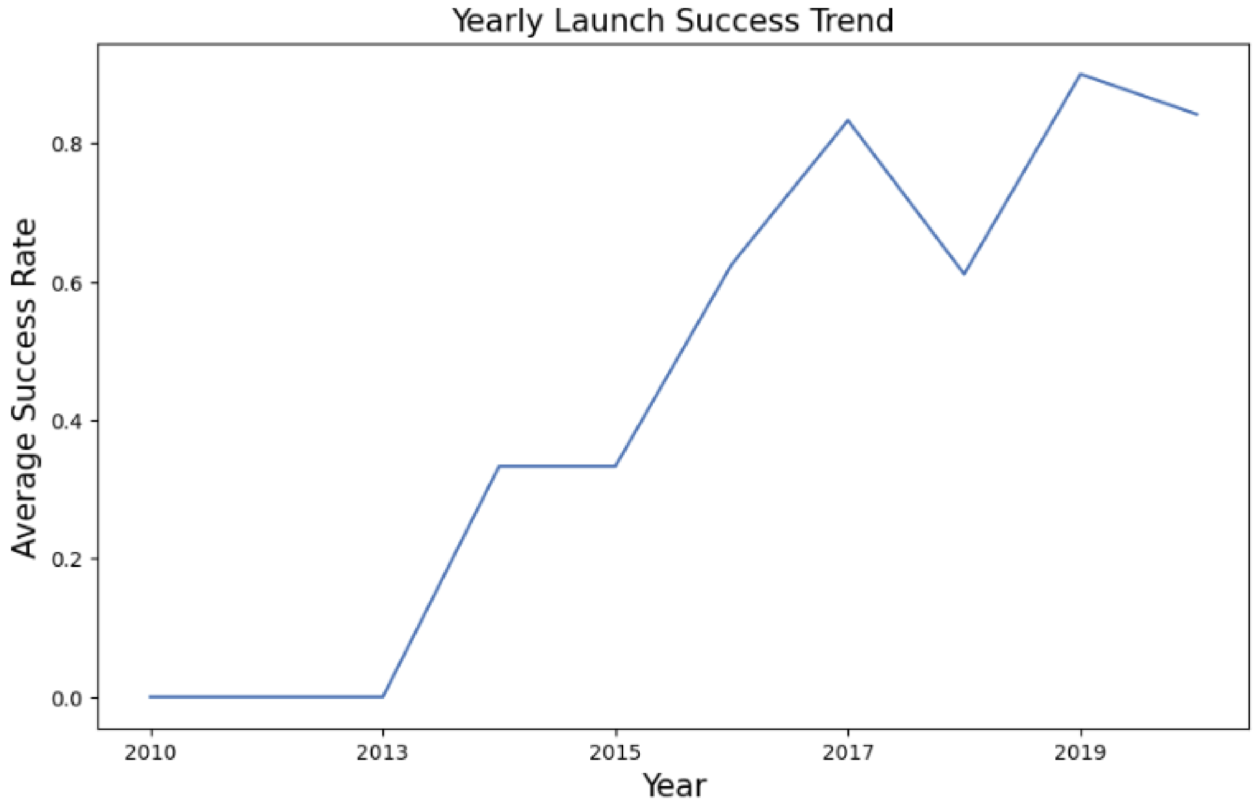
This justifies including both variables in the predictive model instead of treating them independently.

Launch Success Yearly Trend

- **Clear Learning Curve Effect:**
  SpaceX significantly improved landing reliability over time.
- **Operational Maturity After 2015:**
  Post-2015 missions show consistent upward trend, indicating improved engineering and recovery systems.
- **Temporary Dip in 2018:**
  Slight drop suggests experimentation phase or mission complexity, but recovery was quick.
- **High Stability in Recent Years:**
  2019–2020 demonstrates mature reusable rocket technology.

Conclusion:

Time (Year / Flight Number) is a **strong predictive feature**.
Modeling should account for program maturity since later missions have higher probability of success.



All Launch Site Names

In [11]:

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

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

Out[11]:

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

Unique Launch Sites Identified

•The dataset contains **four distinct launch sites**:
CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, and CCAFS SLC-40.

•This confirms SpaceX operations during the analyzed period were concentrated across **Florida (Cape Canaveral & Kennedy Space Center)** and **California (Vandenberg Air Force Base)**.

•The limited number of sites indicates launch activity is geographically focused, which helps simplify comparative performance analysis across locations.

Launch Site Names Begin with 'CCA'

In [13]:

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

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

Out[13]:

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

Records Where Launch Site Begins with “CCA”

•The query returned **5 launch records from CCAFS LC-40**, confirming that launches starting with “CCA” correspond to Cape Canaveral Air Force Station.

•All five missions show **successful mission outcomes**, indicating strong early operational reliability at this site.

•Most early launches (2010–2013) carried **low payload masses** and targeted **LEO / LEO (ISS)**, reflecting initial Dragon cargo and qualification missions.

This suggests CCAFS LC-40 played a key role in SpaceX’s early commercial and NASA cargo operations.

Total Payload Mass

In [14]:
```
%%sql
SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass
FROM SPACEXTABLE
WHERE "Customer" LIKE '%NASA%'
  AND "Customer" LIKE '%CRS%';
```

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

Out[14]:

total_payload_mass
48213

Total Payload Carried by NASA (CRS) Missions

•The total payload mass carried by boosters for **NASA CRS missions** is **48,213 kg**.

•This reflects cumulative cargo transported primarily to the **International Space Station (ISS)** under NASA's Commercial Resupply Services program.

•The large total payload indicates that NASA CRS missions represent a **major portion of SpaceX's operational workload**, highlighting the long-term strategic partnership between NASA and SpaceX.

Average Payload Mass by F9 v1.1

In [16]:
```
%%sql
SELECT AVG("PAYLOAD_MASS__KG_") AS avg_payload_mass
FROM SPACEXTABLE
WHERE "Booster_Version" = 'F9 v1.1';
```

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

Out[16]:

avg_payload_mass
2928.4

Average Payload Mass – Booster Version F9 v1.1

- The **average payload mass** carried by booster version **F9 v1.1** is **2,928.4 kg**.
- This indicates that F9 v1.1 was primarily used for **medium-weight payload missions**, rather than very heavy satellite deployments.
- Compared to later Falcon 9 versions, this reflects the **earlier developmental stage** of the rocket, before major performance upgrades increased payload capacity.

First Successful Ground Landing Date

```
In [18]:
%%sql
SELECT MIN("Date") AS first_success_ground_pad
FROM SPACEXTABLE
WHERE "Landing_Outcome" = 'Success (ground pad)';

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

Out[18]:

first_success_ground_pad
2015-12-22

Query Result:

The first successful landing on a ground pad occurred on **2015-12-22**.

Explanation:

•This date represents the earliest mission where the landing outcome was recorded as **"Success (ground pad)"**.

•It marks a major milestone in SpaceX's reusable rocket program.

•After this date, ground pad landings became increasingly frequent, showing rapid improvement in landing technology and operational reliability.

Successful Drone Ship Landing with Payload between 4000 and 6000

In [19]:

```
%%sql
SELECT DISTINCT "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.
```

Out[19]:

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

Query Result:

The boosters that successfully landed on a drone ship with payload mass between **4000 kg and 6000 kg** are:

•**F9 FT B1022**

•**F9 FT B1026**

•**F9 FT B1021.2**

•**F9 FT B1031.2**

Explanation:

•These booster versions achieved **“Success (drone ship)”** landings within the specified payload range.

•This indicates that SpaceX was capable of reliably landing boosters on drone ships even with medium-to-heavy payloads.

•The presence of multiple booster variants in this range reflects growing operational maturity and improved landing consistency.

Total Number of Successful and Failure Mission Outcomes

In [21]:

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

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

Out[21]:

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

Query Result:

•**Successful Missions:** 99
(98 “Success” + 1 additional “Success” record)

•**Failures:** 1 (“Failure (in flight)”)

Explanation:

•The dataset shows an overwhelmingly high mission success rate, with only one recorded in-flight failure.

•This indicates strong launch reliability overall.

•The presence of a “Success (payload status unclear)” entry suggests a minor reporting variation but does not change the overall success dominance.

Boosters Carried Maximum Payload

In [23]:

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

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

Out[23]:

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

•All boosters listed belong to the **Falcon 9 Block 5 (B5)** generation.

•This confirms that Block 5 variants handled the heaviest payload missions, reflecting their upgraded thrust capacity and structural improvements.

•It highlights the technological advancement of the Block 5 series compared to earlier Falcon 9 versions.

2015 Launch Records

In [24]:

```
%%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)';
```

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

Out[24]:

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

Query Result (Year 2015 – Drone Ship Failures):

•**January (01)** – Booster: *F9 v1.1 B1012* – Launch Site: *CCAFS LC-40*

•**April (04)** – Booster: *F9 v1.1 B1015* – Launch Site: *CCAFS LC-40*

Explanation:

•In 2015, there were **two drone ship landing failures**, both from Falcon 9 v1.1 boosters.

•Both launches occurred at **CCAFS LC-40**, indicating early-stage drone ship recovery challenges at that site.

•These failures happened before SpaceX achieved consistent drone ship landing reliability later in the program.

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

In [25]:

```
%%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;
```

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

Out[25]:

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

Explanation:

•Early in the program, **“No attempt”** dominates, reflecting a phase before active landing recovery efforts began.

•Drone ship landings show equal counts of success and failure (5 each), indicating an experimental period with rapid learning.

•Ground pad successes (3) mark the beginning of reliable reusable booster operations during this timeframe.

Section 3

Launch Sites Proximities Analysis

Global Distribution of SpaceX Launch Sites

Important Elements

The map displays **all SpaceX launch site markers** plotted using Folium on a global geographic layout.

•Three primary locations are visible:

- **CCAFS LC-40** (Florida)
- **KSC LC-39A** (Florida)
- **VAFB SLC-4E** (California)

•The markers clearly show operations concentrated in **two U.S. coastal regions**: East Coast (Florida) and West Coast (California).

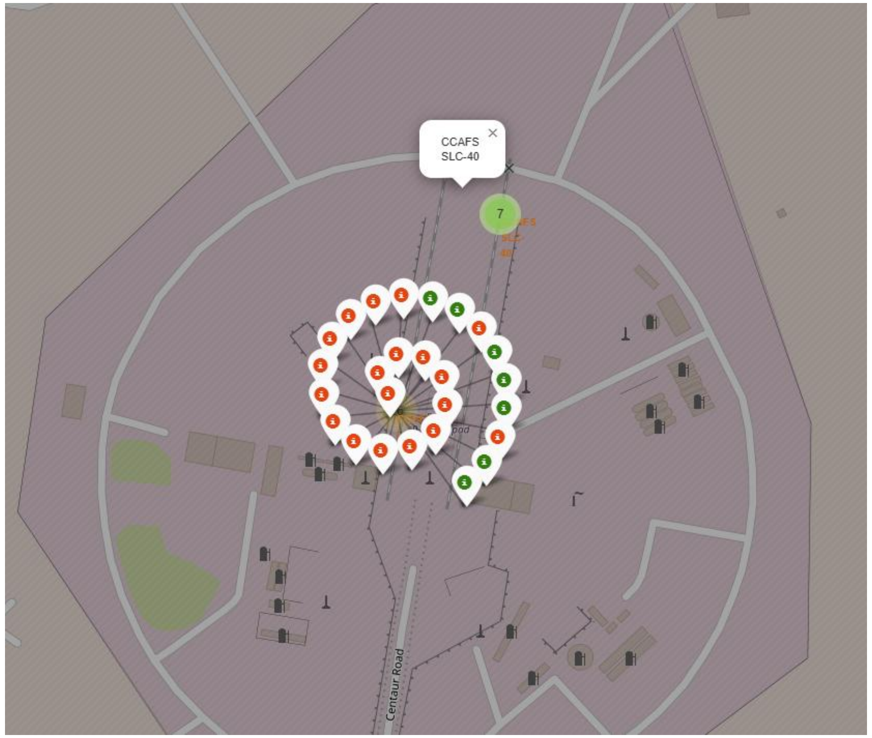
Key Findings

•**Bi-Coastal Strategy:** SpaceX operates from both Atlantic and Pacific coasts, allowing flexibility in orbital trajectories (LEO, ISS, polar orbits).

•**Florida Dominance:** Two launch pads are located in Florida, indicating it as the primary launch hub.

•**Strategic Geography:** Coastal locations reduce risk by allowing rocket stages to travel over ocean rather than populated areas.

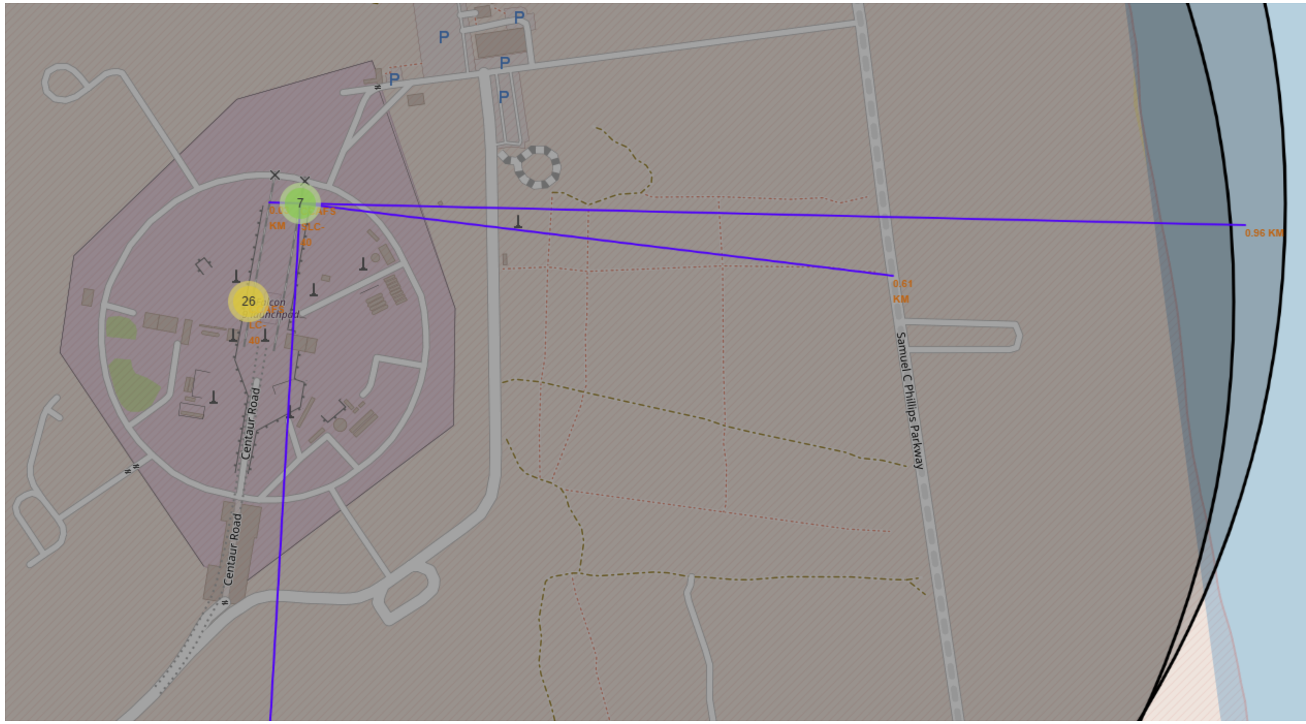
Color-Labeled Launch Outcomes at CCAFS LC-40



Important Elements
Each marker represents a single launch attempt at **CCAFS LC-40**.
•**Green markers = Successful Launch**
•**Red markers = Failed Launch**
•Marker clustering is enabled (numbered circle), showing high launch density at this site.

Key Findings
•Early launches show a higher concentration of **red markers**, indicating initial launching failures during the experimental phase.
•As missions progress, **green markers increase**, showing clear improvement in launching reliability.
•The dense clustering confirms that **CCAFS LC-40 is a primary operational hub**, playing a major role in SpaceX’s learning curve and recovery optimization.

Proximity Analysis of Launch Site – CCAFS LC-40



Important Elements

The **launch site marker** is displayed with connecting blue distance lines to nearby infrastructure.

•Distances are clearly labeled on the map:

- **Railway ≈ 0.3 km**
- **Highway ≈ 0.6 km**
- **Coastline ≈ 0.9–1.0 km**

•The surrounding urban reference (Melbourne) is located much farther away (≈ **53.4 km**).

Key Findings

•**Close to Railway:**

The site's proximity (~0.3 km) supports efficient transportation of rocket components and heavy equipment.

•**Near Major Highway:**

Being ~0.6 km from a highway enables smooth logistics, workforce mobility, and emergency access.

•**Coastal Location:**

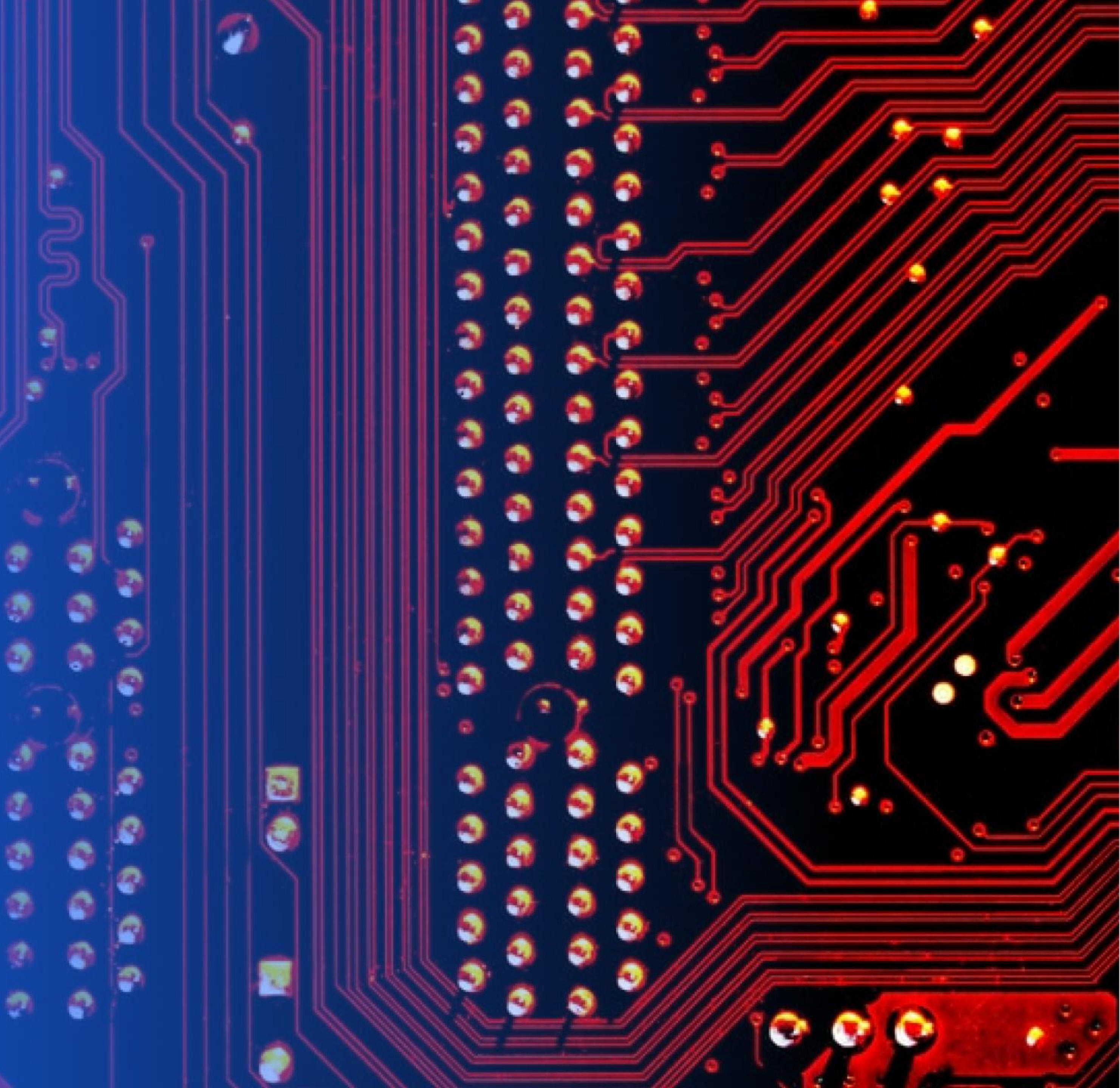
At under 1 km from the coastline, launches can safely occur over the ocean, minimizing risk to populated land areas.

•**Far from Major Cities:**

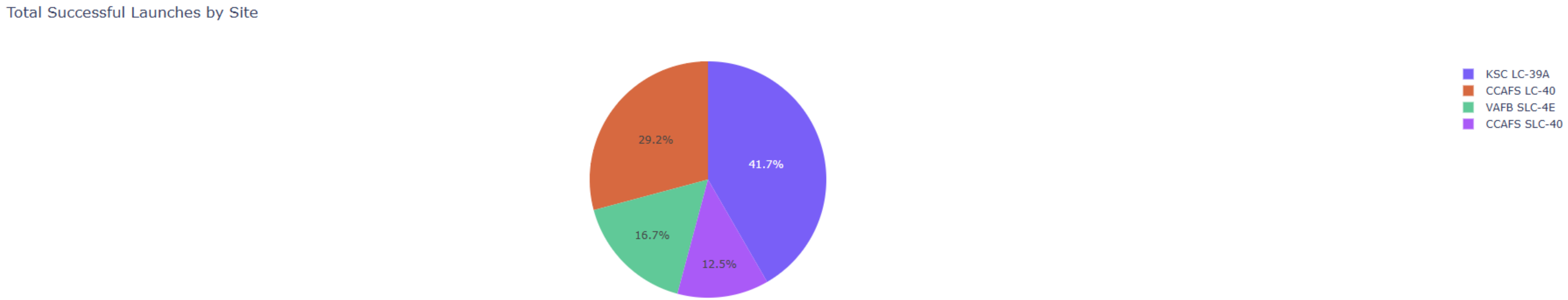
The ~53 km distance from Melbourne confirms deliberate placement away from dense population centers for safety.

Section 4

Build a Dashboard with Plotly Dash



Launch Success Distribution Across All Sites



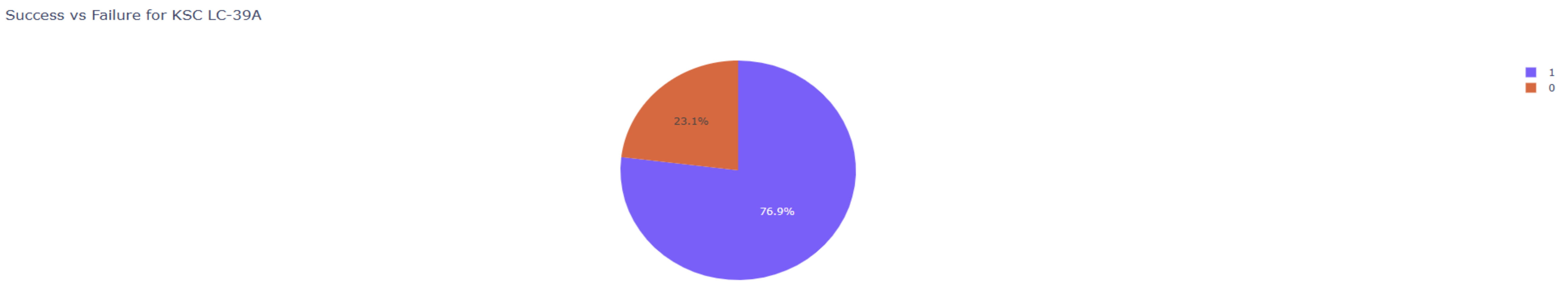
Key Highlights

- **KSC LC-39A leads** with the highest share of successful launches (~41.7%).
- **CCAFS LC-40** is second (~29.2%), showing strong performance.
- **VAFB SLC-4E** and **CCAFS SLC-40** contribute smaller portions (~16.7% and ~12.5%).
- Success is concentrated mainly in two major launch sites.

Insight

KSC LC-39A is the dominant launch site in terms of total successful missions, indicating higher operational activity and strong performance compared to other sites.

KSC LC-39A – Highest Launch Success Ratio



Key Highlights

•**Success Rate: ~76.9%**

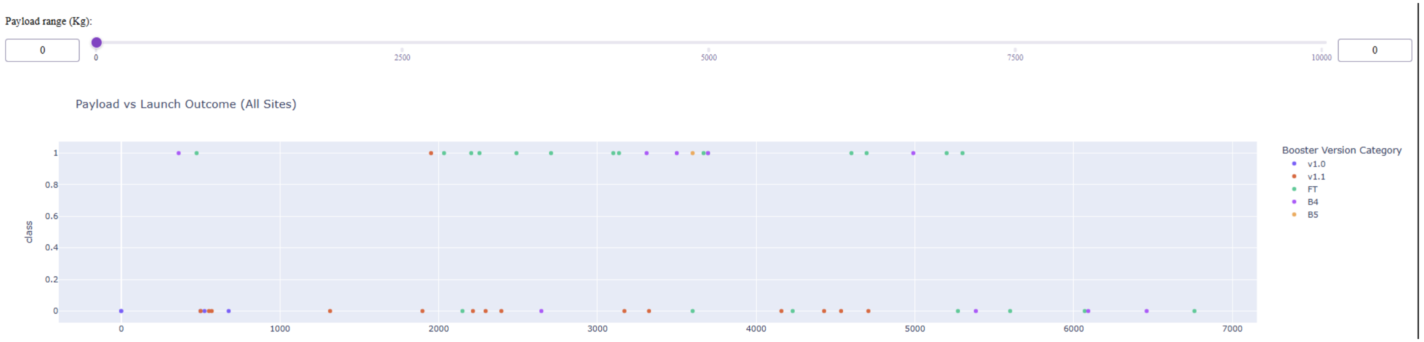
•**Failure Rate: ~23.1%**

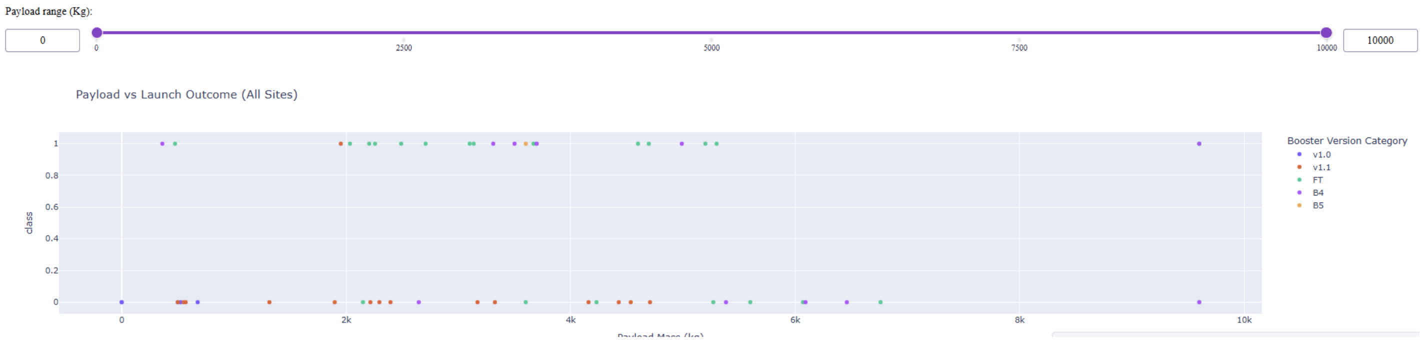
•Majority of launches from KSC LC-39A are successful (Class = 1 dominates the chart).

Insight

KSC LC-39A not only has the highest number of successful launches but also the **highest success ratio**, making it the most reliable and best-performing launch site among all locations.

Payload Mass vs Launch Outcome Across All Sites





Overall Insights

•**Mid-range payloads (3000–6000 kg) have the highest success concentration.**

•**FT and B4 boosters demonstrate stronger performance compared to v1.0/v1.1.**

•Technological upgrades in booster versions correlate with improved success rates.

Screenshot 1: Narrow Payload Range (Low Range Selected)Key Observations:

•Lower payload ranges (0–3000 kg approx.) show **mixed outcomes**.

•Earlier booster versions (**v1.0 and v1.1**) show more failures (Class = 0).

•Success rate is less consistent in lower payload bands.

Screenshot 2: Full Payload Range (0–10000 kg Selected)Key Observations:

•Mid to higher payload ranges (~3000–6000 kg) show **more consistent success (Class = 1)**.

•**FT and B4 booster versions** dominate successful launches.

•Higher payloads (>7000 kg) appear limited but mostly successful.

•Earlier versions (v1.0, v1.1) show comparatively lower success rates.

Section 5

Predictive Analysis (Classification)

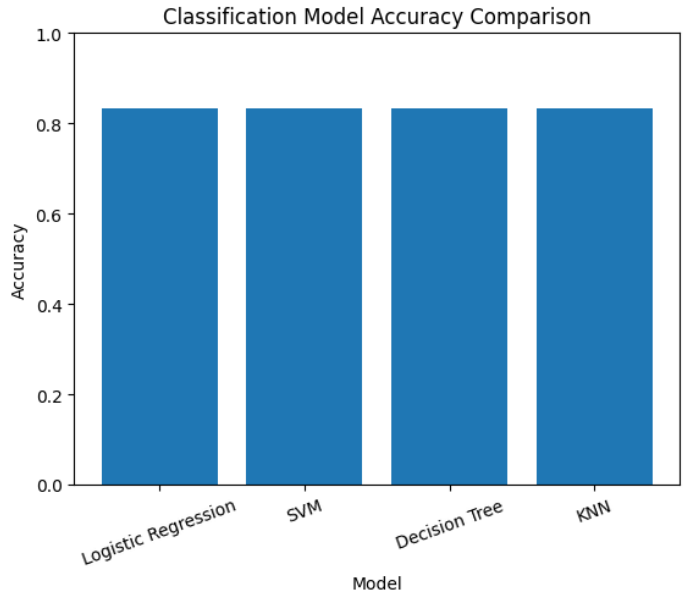
Classification Accuracy

Objective

- To compare the performance of multiple classification models built to predict Falcon 9 first-stage landing success.

Models Evaluated

- Logistic Regression
- Support Vector Machine (SVM)
- Decision Tree
- K-Nearest Neighbors (KNN)



Results (Test Accuracy)

- All four tuned models achieved a similar test accuracy of approximately **83.33%**.

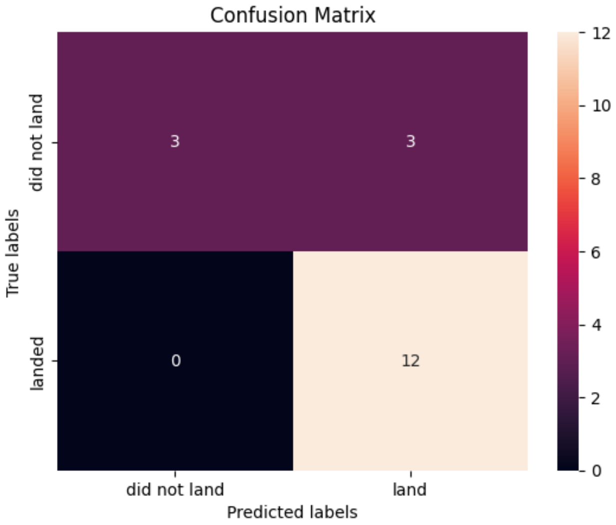
Best Performing Model

- Although test accuracy was identical across models, the **Decision Tree model achieved the highest cross-validation accuracy (~88.9%) during hyperparameter tuning**.
- Therefore, the **Decision Tree was selected as the best-performing model** based on overall training performance and validation stability.

Conclusion

- The models demonstrate comparable generalization performance; however, the **Decision Tree** shows stronger learning capability during cross-validation and is considered the most reliable model for this prediction task.

Confusion Matrix



Performance Insight

•The model is **very strong at predicting successful landings** (100% recall for landed cases).

•Slight weakness exists in distinguishing failed launches (some false positives).

•Overall test accuracy ≈ **83.33%**.

•The model correctly identified **all successful landings** (Recall = 1.00), meaning no successful missions were missed.

•Some failed launches were misclassified as successful (FP = 3), reducing precision to 0.80.

•The high F1 score (0.89) indicates strong overall balance between precision and recall.

•The model performs very well at detecting successful landings but is weaker at distinguishing failures.

This supports selecting the Decision Tree model as the best-performing classifier in this analysis.

Confusion Matrix Values

•True Positives (TP) = 12

•True Negatives (TN) = 3

•False Positives (FP) = 3

•False Negatives (FN) = 0

Total test samples = 18

The **Decision Tree** model performs reliably in predicting landing success, with strong success detection capability and moderate error in identifying failures. This makes it suitable for estimating mission landing outcomes while minimizing missed successful landings.

Conclusions

Key Analytical Findings

- **Launch Site Impact**
  - KSC LC-39A had both the highest number of successful launches and the highest success ratio.
  - Launch performance varies significantly across sites.
- **Payload Influence**
  - Mid-range payloads (approximately 3000–6000 kg) showed higher success consistency.
  - Very low payloads had more mixed outcomes.
- **Booster Version Evolution**
  - Later booster versions (FT, B4, B5) demonstrated improved success rates compared to earlier versions (v1.0, v1.1).
  - Technological improvement clearly correlates with higher landing reliability.

Model Performance Conclusion

Four classification models were built and tuned:

- Logistic Regression
- SVM
- Decision Tree
- KNN

All achieved similar test accuracy (~83%).

However, the **Decision Tree model achieved the highest cross-validation accuracy (~88.9%)** and demonstrated strong recall (1.00) and F1-score (~0.89).

Therefore, **Decision Tree was selected as the best-performing predictive model**.

Overall Conclusion

The analysis confirms that:

- Landing success is strongly influenced by booster version and payload characteristics.
- SpaceX's technological evolution significantly improved landing reliability.
- Machine learning models can effectively predict landing outcomes with high reliability.
- This project demonstrates how data analytics and machine learning can support aerospace operational decision-making and performance optimization.

Thank you!