IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Zaara Akbar Shaikh
18/01/26

Outline

Executive Summary

Summary of Methodologies

1. **Collection**: Data gathered via **SpaceX REST API** and **Web Scraping**.
2. **Processing**: Data cleaning and feature engineering (One-Hot Encoding) using **Pandas**.
3. **Analysis**: Exploration through **SQL**, **Data Visualization**, and **Geospatial analysis (Folium)**.
4. **Modeling**: Built and tuned four classification models: **Logistic Regression, SVM, Decision Tree, and KNN**.

Summary of All Results

- **Performance: Logistic Regression** predictive model performs the best among all 4 models tested.
- **Success Trends**: Analyzed **90 launches** with an overall success rate of **66.67%**, which improved to **84.2%** by 2020.

Key Findings:

CCAFS SLC 40 is the most frequent launch site.

Orbits **ES-L1, GEO, and HEO** achieved 100% landing success.

Success correlates positively with **Flight Number** (experience) and negatively with very heavy payloads.

Launch sites are strategically placed near **coastlines** and lower latitudes for optimal safety and fuel efficiency.

Introduction

Project Background and Context:

- SpaceX is a leader in the commercial space industry, largely due to its ability to reuse the first stage of the Falcon 9 rocket.
- Traditional rocket launches are expensive because the hardware is discarded after a single use; SpaceX reduces these costs by landing and refurbishing first-stage boosters.
- This project focuses on predicting whether the first stage will land successfully, which is a critical factor in determining the overall cost of a mission.

Problems to Solve:

- **Landing Prediction**: Can we accurately predict the success of a first-stage landing based on flight characteristics like payload mass and orbit?
- **Feature Correlation**: How do factors such as launch site location and historical flight numbers influence the success rate?
- **Cost Implications**: By predicting landing outcomes, can we provide a better estimate for the commercial price of a launch?
- **Model Selection**: Which machine learning algorithm provides the most reliable classification for this specific dataset?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Two-source Data Collection for rich historical data.
- Perform data wrangling
  - Used one-hot encoding, handled missing values etc
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Built, tested and analyzed 4 classification models ie Logistic Regression, SVM, Decision Tree and KNN.

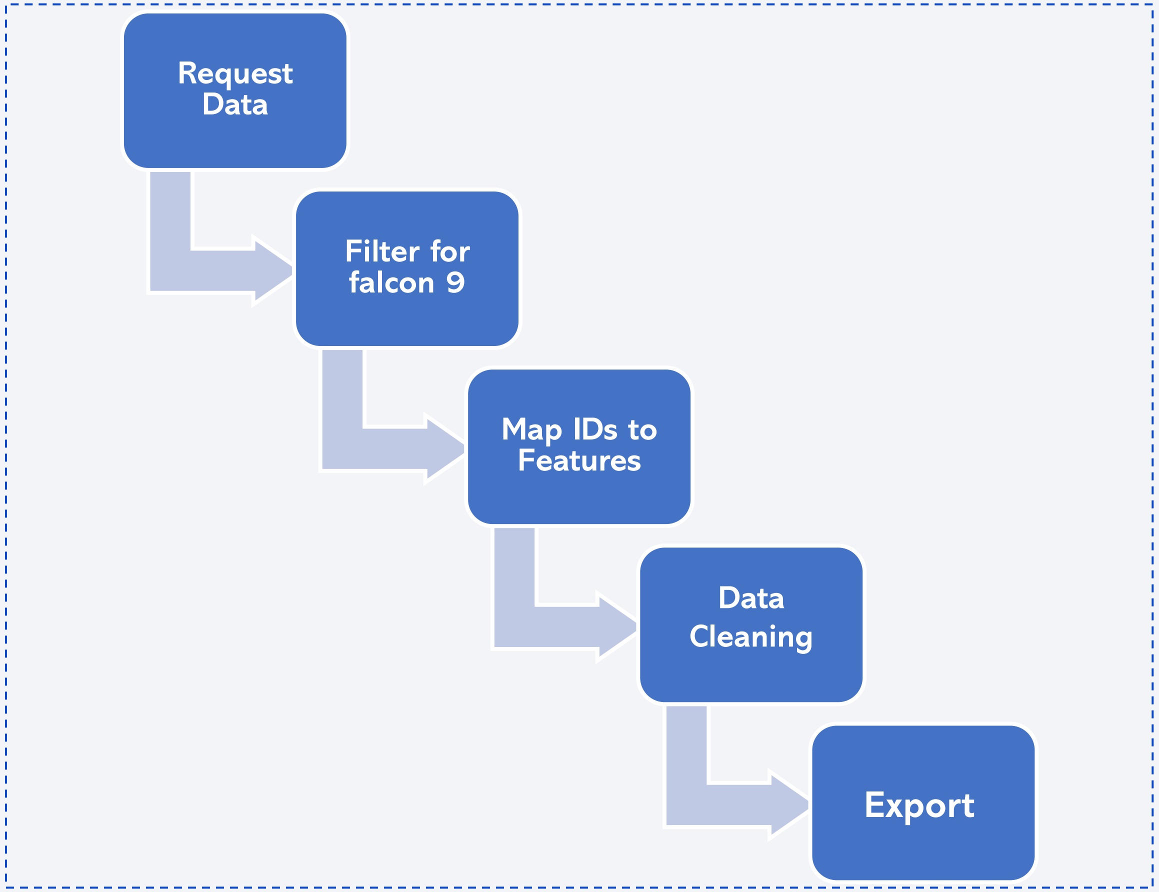
Data Collection

Two-Source Data Strategy:

- **SpaceX REST API:** Retrieved 90 Falcon 9 launches with 17 technical features (payload mass, orbit type, booster version, landing outcome)
- **Wikipedia Web Scraping:** Extracted 121 historical launch records using BeautifulSoup for additional context and validation
- **Outcome:** Comprehensive dataset covering 2010-2020, enabling robust predictive modelling.

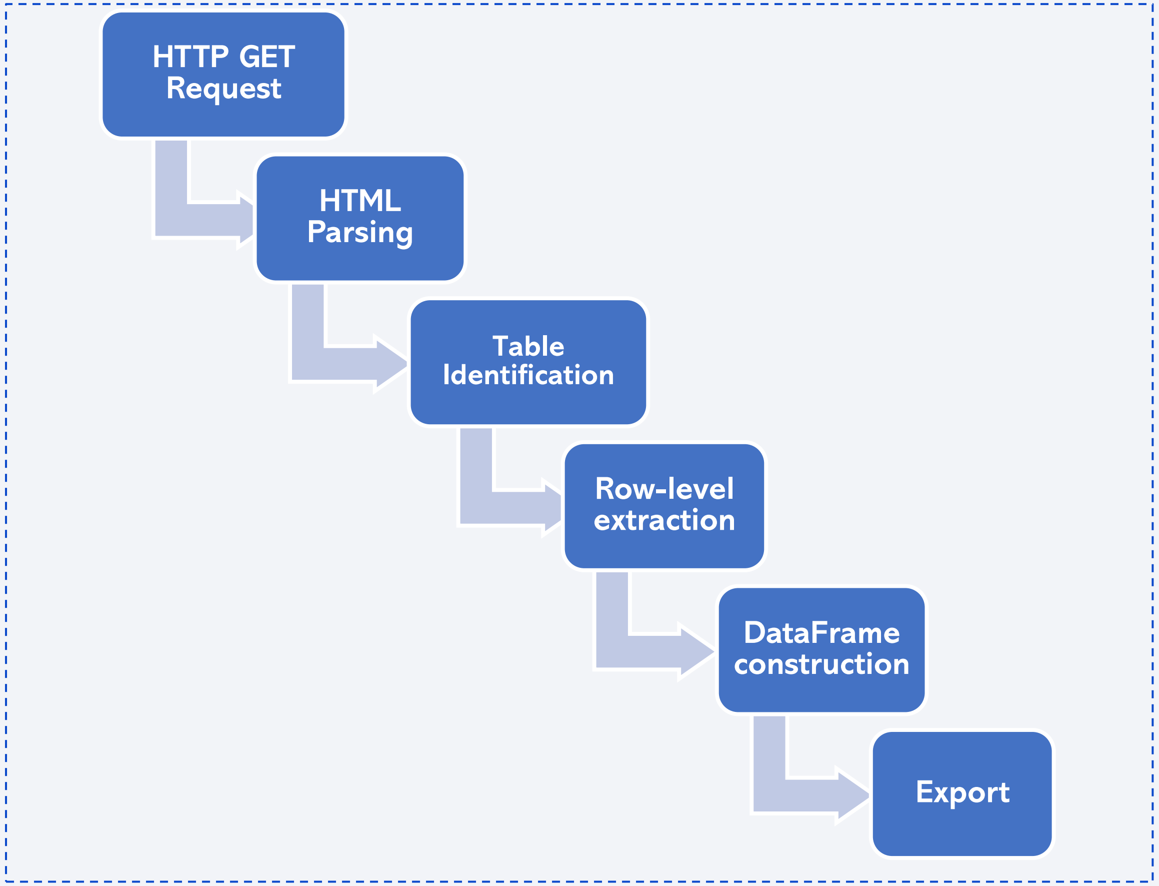
Data Collection – SpaceX API

- **Methodology:** Authenticated requests to the SpaceX API to retrieve core launch data.
- **Processing:** Flattened complex JSON fields (like rocket and launchpad) and mapped IDs to human-readable names, Robust 'Circuit Breaker' implementation for data integrity.
- **Outcome:** A structured dataset with 90 rows and 17 technical features.
- **GitHub URL:** 01_Data_Collection_API.ipynb



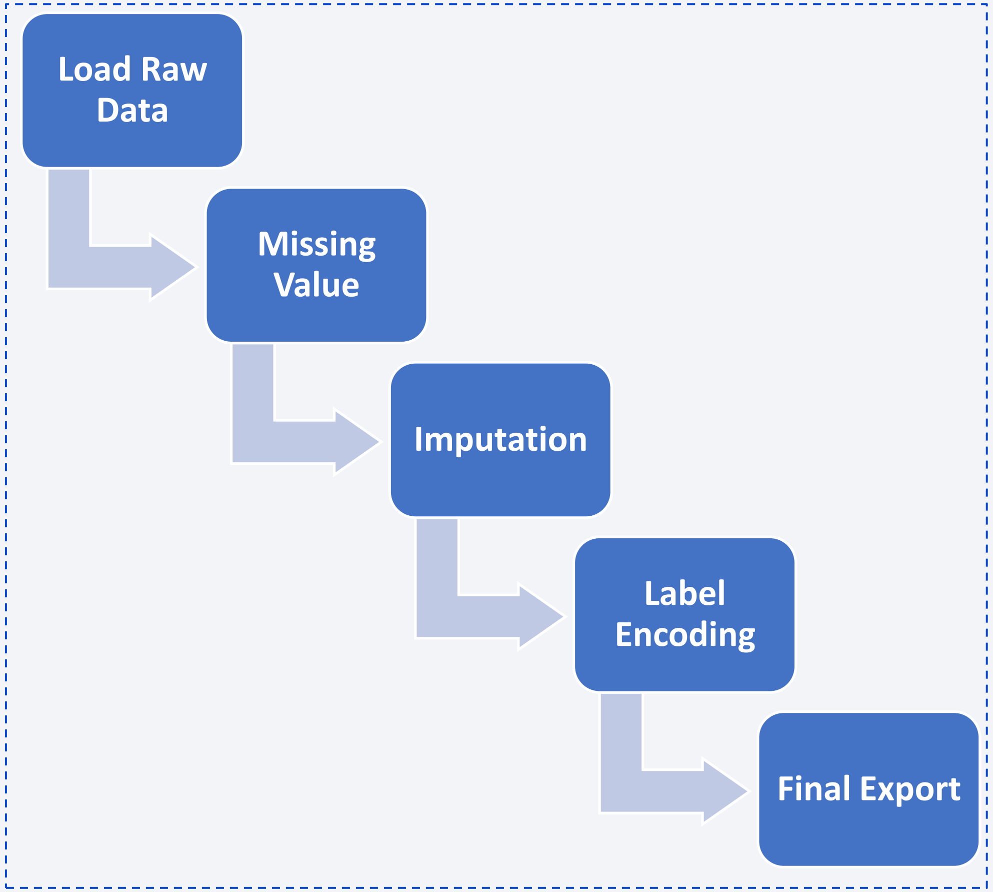
Data Collection - Scraping

- **Methodology:** Targeted the Wikipedia-List of Falcon 9 and Falcon Heavy launches using the BeautifulSoup library.
- **Extraction:** Parsed HTML <table> elements with specific CSS classes to capture flight numbers, dates, and landing outcomes.
- **Data Cleaning:** Implemented custom helper functions to handle complex HTML structures, such as varying row spans and superscript citations.
- **Outcome:** Successfully extracted 121 launch records, providing a robust historical context for the predictive models.
- **GitHub URL:** 02_Data_Collection_Scraping.ipynb



Data Wrangling

- **Methodology:** Performed data cleaning on the merged dataset to ensure high-quality inputs for predictive modeling.
- **Target Variable Engineering:** Created a binary Class column where 1 represents a successful landing (e.g., True ASDS, True Ocean, True RTLS) and 0 represents an unsuccessful outcome.
- **Handling Missing Values:** Identified missing data in the PayloadMass column and replaced null values with the column mean to maintain dataset size.
- **Categorical Analysis:** Verified the distribution of launch sites (e.g., CCAFS SLC 40, KSC LC 39A, VAFB SLC 4E) and orbit types (GTO, LEO, ISS, etc.).
- **Outcome:** Produced a cleaned dataset (dataset_part_2.csv) with 90 records and a success rate of approximately 66.67%.
- **GitHub URL:** 03_Data_Wrangling.ipynb



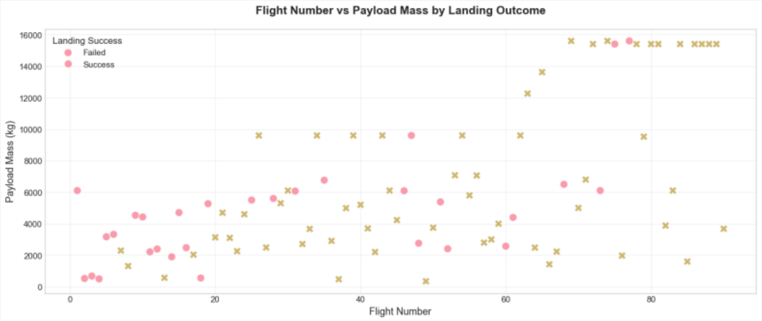
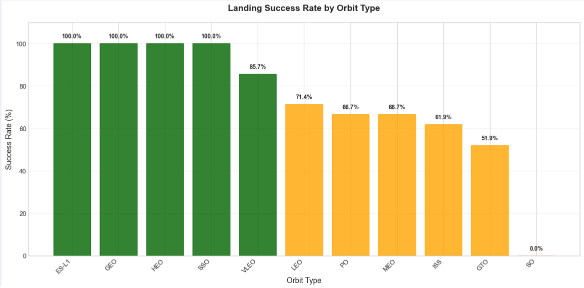
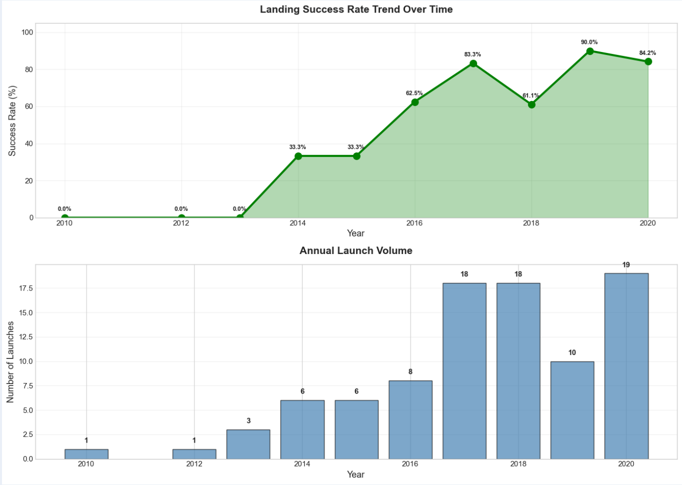
EDA with Data Visualization

- **Learning Curve (Tasks 1 & 6)**: Success rates show a clear positive trend over time (Flight Number) and by year, reaching a peak in 2020 as SpaceX technology matured.

- **Site & Payload Strategy (Tasks 2 & 5)**: Heavy payloads (10,000kg+) are successfully concentrated at **KSC LC 39A** and **VAFB SLC 4E**, while **CCAFS SLC 40** handles the highest volume of varied missions.

- **Orbit Reliability (Task 3 & 4)**: Orbits **ES-L1, GEO, HEO, and SSO** achieved a 100% success rate, whereas **GTO** and **ISS** show more variability due to higher mission frequency.

GitHub URL: 05_EDA_Visualizations

Task	Plot Type	Reason for Choice	Screenshot
Flight No vs Payload	**Scatterplot**	Visualizes if experience (Flight No) overcomes weight (Payload) challenges.	
Success by Orbit	**Bar Chart**	Best for comparing categorical performance; clearly ranks the "safest" orbits.	
Yearly Trend	**Line Chart & Bar Chart**	Standard for "Time Series" to show the rate of improvement over the years.	

EDA with SQL

1. Identified unique launch sites using DISTINCT
2. Retrieved 5 launches from sites starting with **CCA** using LIKE 'CCA%'
3. Calculated **total payload mass** for NASA (CRS) missions using SUM
4. Calculated **average payload mass** for **F9 v1.1** using AVG
5. Identified the **first successful ground pad landing** using MIN(Date)
6. Analyzed **successful drone ship landings** (payload 4000–6000 kg)
7. Generated **mission outcome summary**
8. Identified **booster versions with maximum payload** using subqueries
9. Counted **failed drone ship landings in 2015**
10. Computed **landing outcome rankings (2010–2017)**
11. Analyzed **success rate by launch site and booster version**
12. Evaluated **payload mass trends** and **landing success evolution over time**

GitHub URL: 04_EDA_SQL.ipynb

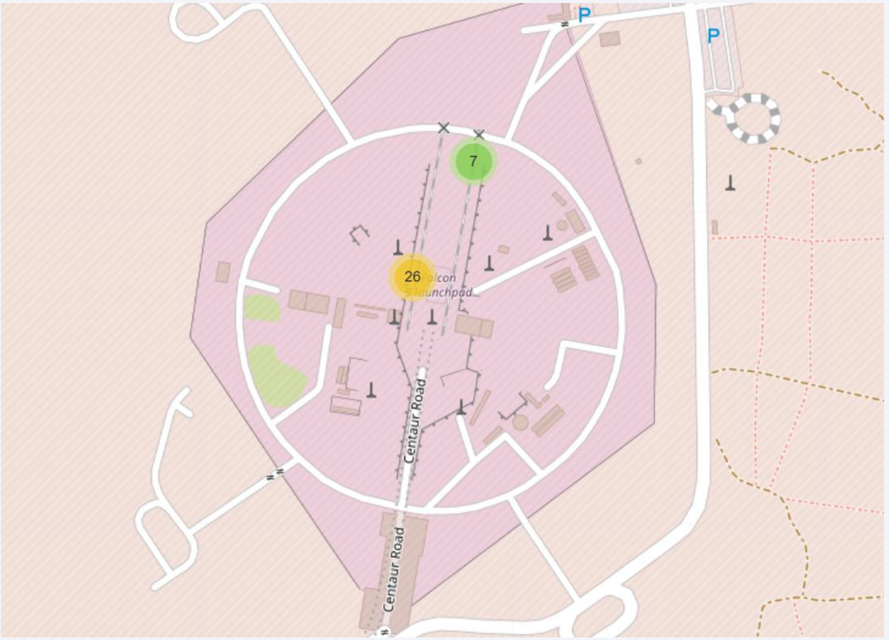
```
[TASK 8] Booster Versions Carrying Maximum Payload
----------------------------------------
Maximum payload mass: 15,600 kg
   Booster_Version  PAYLOAD_MASS__KG_
0    F9 B5 B1048.4              15600
1    F9 B5 B1048.5              15600
2    F9 B5 B1049.4              15600
3    F9 B5 B1049.5              15600
4   F9 B5 B1049.7               15600
5    F9 B5 B1051.3              15600
6    F9 B5 B1051.4              15600
7    F9 B5 B1051.6              15600
8    F9 B5 B1056.4              15600
9   F9 B5 B1058.3               15600
10  F9 B5 B1060.2               15600
11   F9 B5 B1060.3              15600
```

```
[TASK 9] Failed Drone Ship Landings - 2015
----------------------------------------
Number of drone ship failures in 2015: 2
  Month       Landing_Outcome Booster_Version  Launch_Site
0    01   Failure (drone ship)   F9 v1.1 B1012  CCAFS LC-40
1    04   Failure (drone ship)   F9 v1.1 B1015  CCAFS LC-40
```

Build an Interactive Map with Folium

Markers Used & Why:

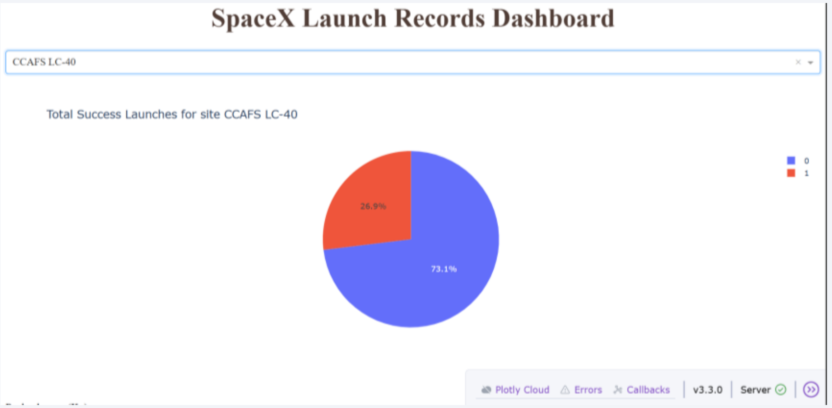
- **MarkerClusters**: Used to group launch sites into clusters, allowing for a clear overview of the three primary launch regions (East Coast, West Coast, and Gulf) without visual clutter.
- **Color-Coded Icons**: Applied **Green** markers for successful landings (Class 1) and **Red** markers for failed landings (Class 0) to visualize success density at a glance.
- **Custom Labels**: Added popup labels to each marker displaying the Launch Site name and Mission Outcome for interactive drill-downs.



GitHub URL: [06_Geospatial_Analysis.ipynb]

Build a Dashboard with Plotly Dash

- **Plots & Graphs Added**:
  - **Success Pie Chart**: A dynamic pie chart that toggles between showing the total success breakdown for all sites and the success vs. failure ratio for a specifically selected site.
  - **Payload vs. Success Scatter Plot**: A scatter plot visualizing the relationship between Payload Mass (kg) and launch class (Success/Failure), with points color-coded by Booster Version Category.
- **Interactions Added:**
  - **Site Dropdown Menu**: Allows the user to filter all visualizations by a specific launch site (e.g., CCAFS LC-40, KSC LC-39A) or view aggregate data for "All Sites".
  - **Payload Range Slider**: An interactive slider that filters the scatter plot in real-time based on a user-defined payload mass range (e.g., 0kg to 10,000kg).
- **Why These Were Used**:
  - **Interactive Filtering**: Dashboards allow stakeholders to drill down into specific variables (like site or payload) that static charts cannot capture, enabling more granular discovery of success patterns.
  - **Correlation Discovery**: The scatter plot with the range slider specifically helps identify if certain booster versions are more reliable at specific weight thresholds.



GitHub Url: app.py

Predictive Analysis (Classification)

- **Model Building**:
    1. Standardized features using StandardScaler to normalize technical variables.
    2. Split the data into training (80%) and testing (20%) sets to ensure the model could generalize to new SpaceX missions.

- **Evaluation & Improvement**:
    1. Optimized each model using GridSearchCV to find the best hyperparameter settings (like 'C' values or 'kernel' types).
    2. Used **Confusion Matrices** to visualize exactly where the models predicted correctly vs. incorrectly.

- **Finding the Best Model**:
    1. All classification models reached a high **Test Accuracy of 83.33%**.
    2. **Logistic Regression is selected as the best model** due to its optimal balance of accuracy, interpretability (coefficient weights show feature importance), and computational efficiency for real-time predictions.

GitHub URL: [07_Machine_Learning_Prediction.ipynb]

ML Development Process Flowchart

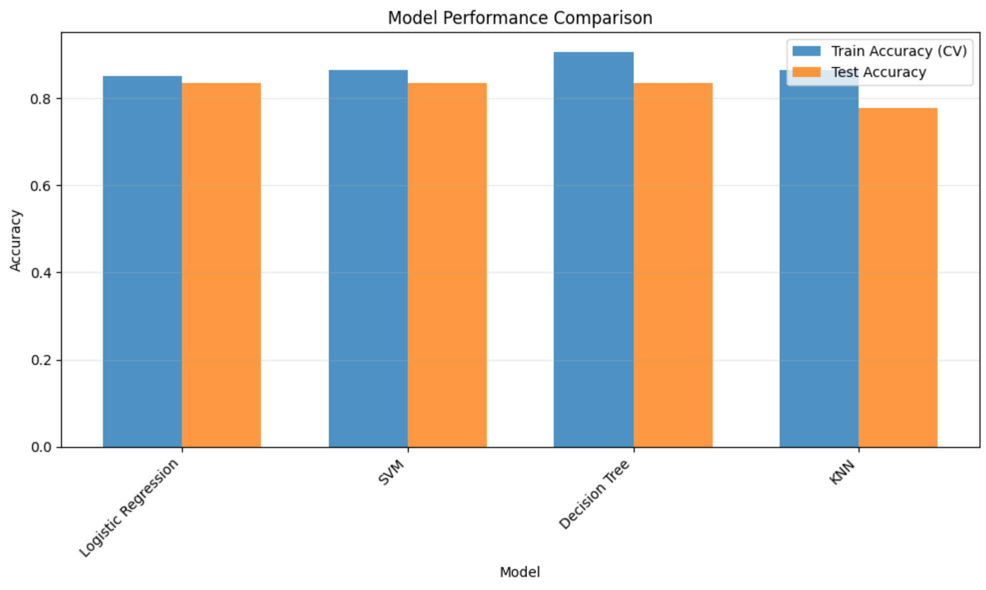
Standardization
- Scale the features to a uniform range.

Algorithm Selection
- Initialize Logistic Regression, SVM, Decision Tree and KNN

Hyperparameter Tuning
- Run GridSearchCV to optimize "Best Params".

Testing
- Validate on the 20% hold-out test set.

Benchmarking
- Compare all scores; achieved 83.33% across all algorithms



Results

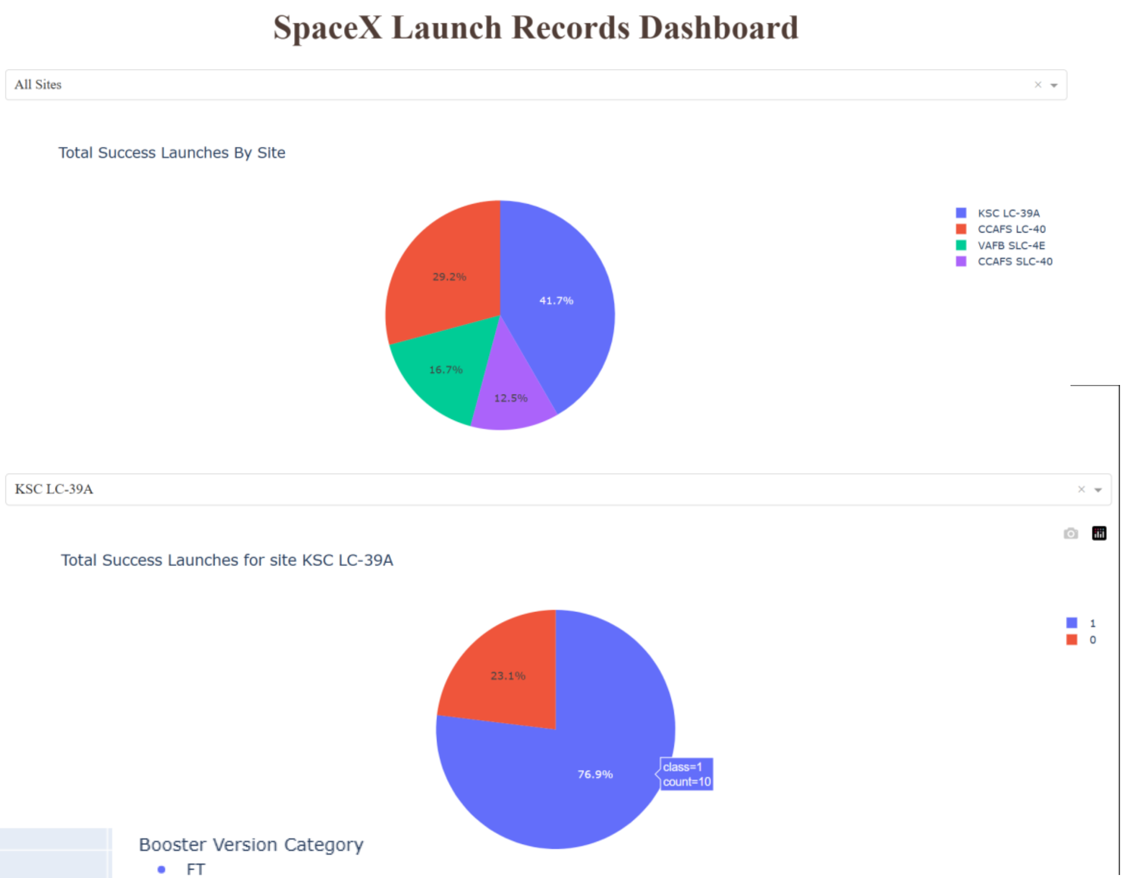
Exploratory Data Analysis (EDA) Results:

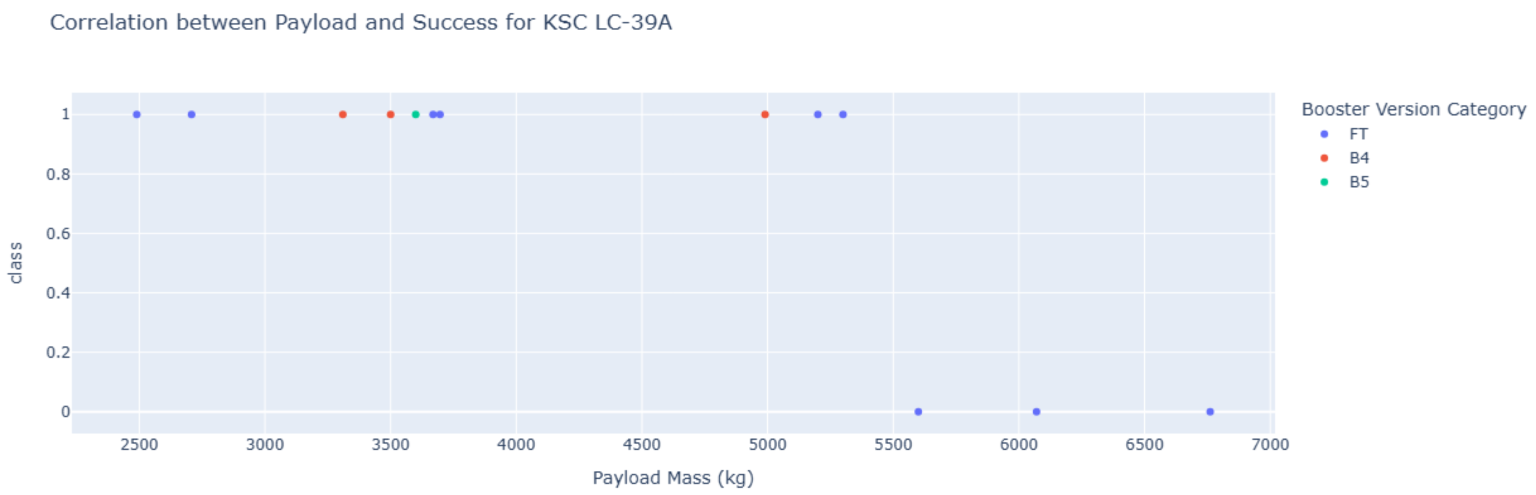
- **Success Trends**: Launch success rates show a strong positive correlation with **Flight Number**, reaching a peak success rate in the year 2020.
- **Orbital Performance**: Specific orbits, including **ES-L1, GEO, HEO, and SSO**, achieved a **100% success rate** in the analyzed dataset.
- **Payload Strategy**: Heavy payloads (10,000 kg+) are most successfully managed at **KSC LC 39A** and **VAFB SLC 4E**.

Interactive Demo Insights:

- **Site-Specific Reliability**: The Dash interface revealed that **KSC LC-39A** has the highest total success count among all sites.

- **Payload Correlation**: Real-time filtering shows that success rates improve significantly for payloads between **2,000 kg and 5,000 kg**.





Predictive Analysis Results:

- **Model Accuracy:** Logistic Regression and SVM achieved the highest Test Accuracy of **83.33%** (15/18 correct predictions).

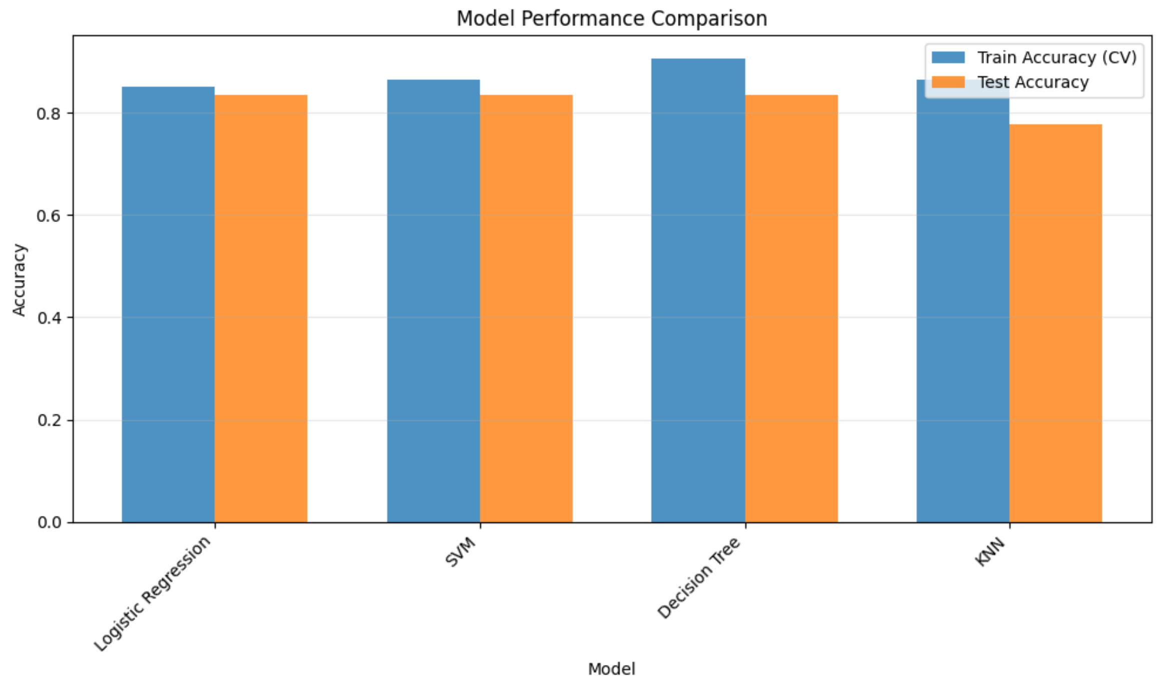
- **Best Model Selection:** Logistic Regression chosen for:
  - Tied-best accuracy (83.33%)
  - Interpretable coefficients (shows which features drive predictions)
  - Fast, efficient predictions suitable for operational deployment
  - No hyperparameter sensitivity (unlike SVM's kernel complexity)

- **Business Value:** 83.33% accuracy enables reliable cost estimation for competitive bidding against SpaceX's reusability advantage

Model Performance Comparison:

Model	Train Accuracy (CV)	Test Accuracy
Logistic Regression	0.850000	0.833333
SVM	0.864286	0.833333
Decision Tree	0.905357	0.833333
KNN	0.864286	0.777778

Best performing model: Logistic Regression

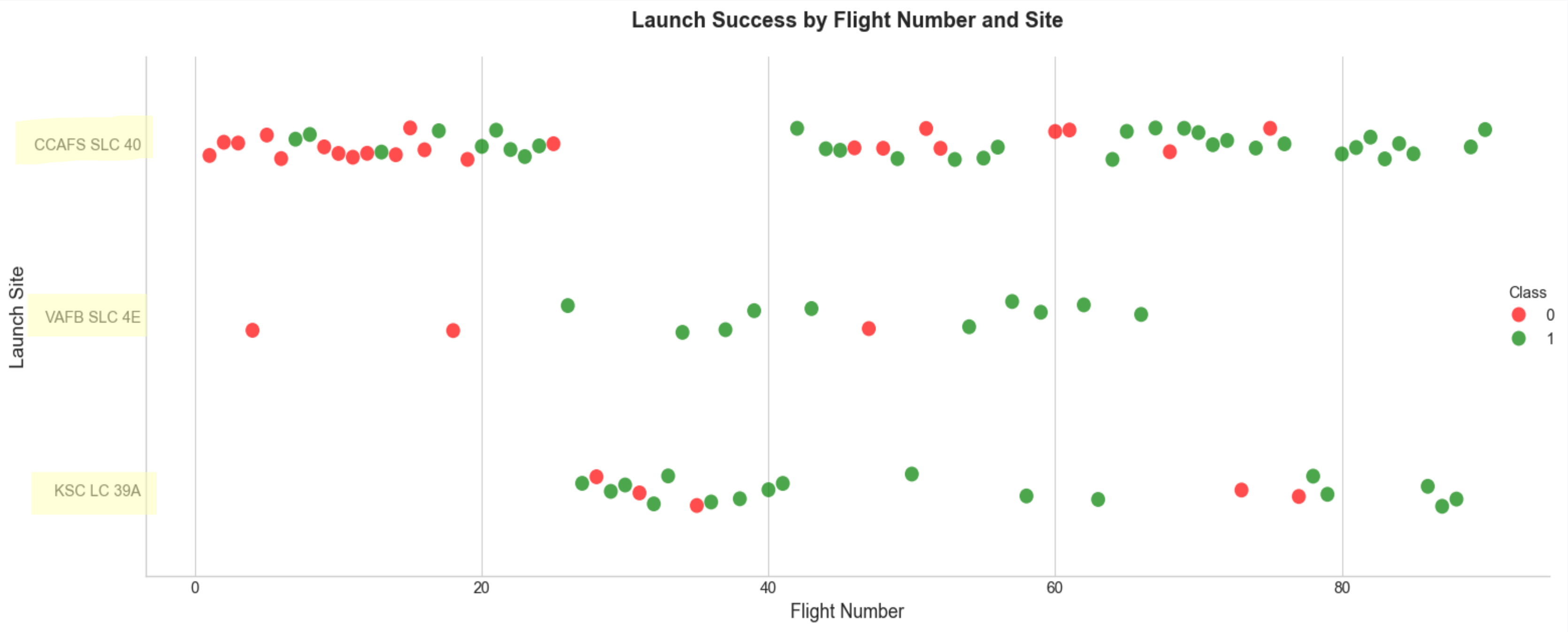


Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

This chart illustrates the "learning curve" of SpaceX's launch operations across its three primary sites:



- **Growing Success Rates**: You can observe that as the **Flight Number** increases (moving from left to right), the density of successful landings (Class 1) increases significantly across all sites. This demonstrates that as SpaceX gained more flight experience, their recovery processes became highly refined and reliable.

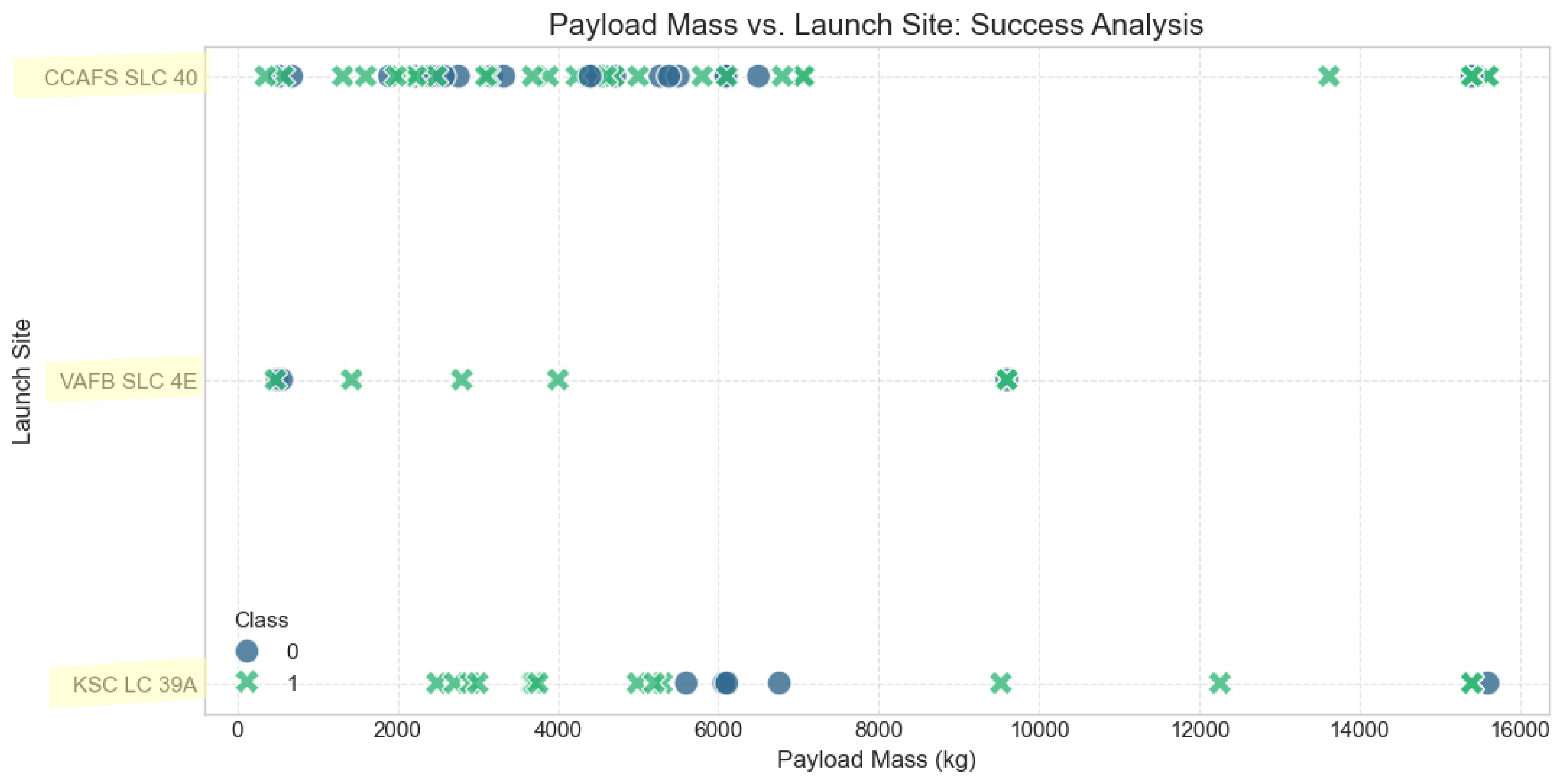
Payload vs. Launch Site

This analysis reveals that launch sites are not interchangeable; they are selected based on the mission's physical requirements:

CCAFS SLC 40 (The High-Volume Hub):
This site handles the widest variety of payloads

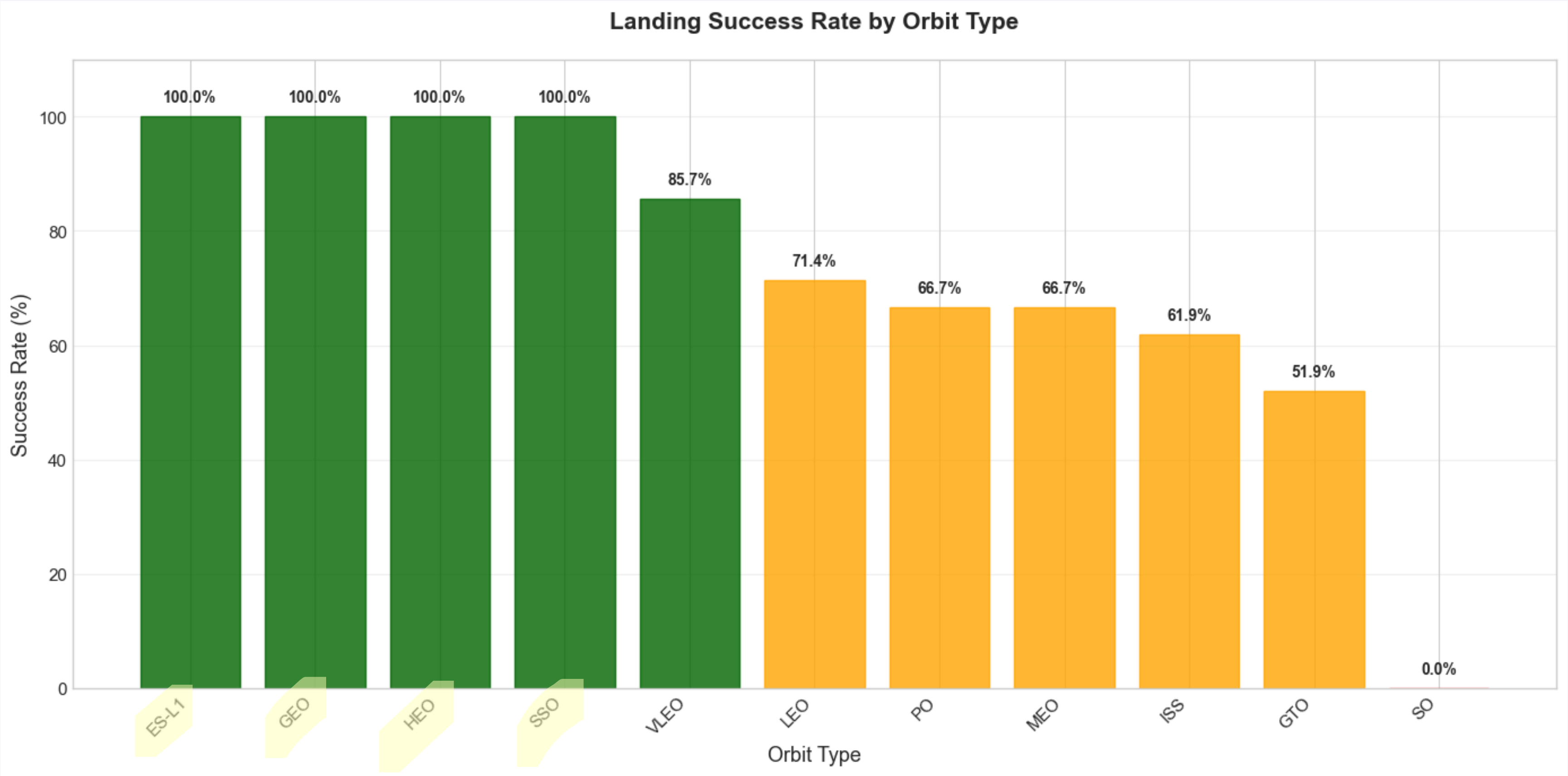
VAFB SLC 4E (The Specialized Polar Site):
This site actually clusters missions in the **mid-range payload** category, primarily between **2,000 kg and 10,000 kg**.

KSC LC-39A (High-Mass Complexity): While KSC handles heavy payloads, the data shows more variability in outcomes at the extreme 15,000 kg range compared to CCAFS, indicating higher mission complexity.



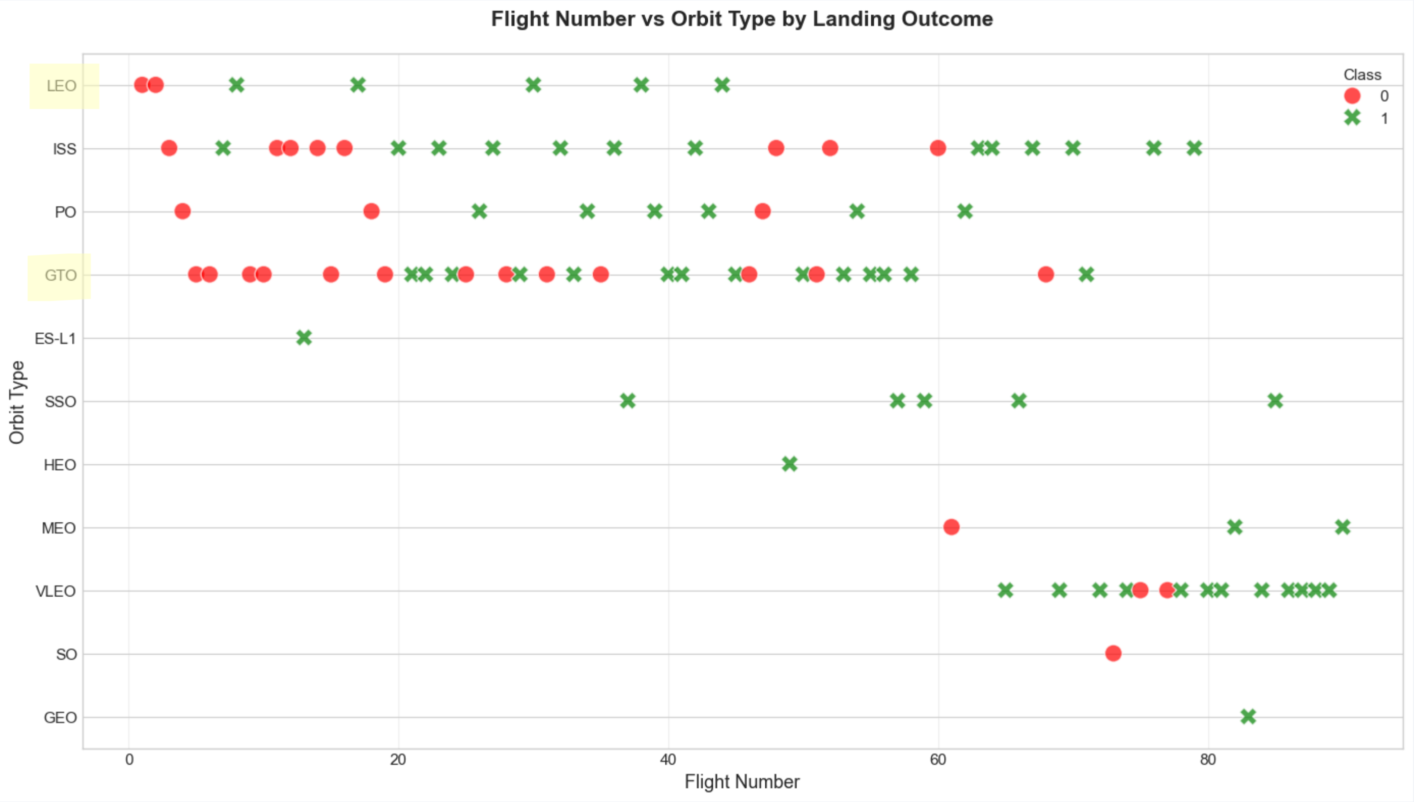
Success Rate vs. Orbit Type

- **ES-L1, GEO, HEO, and SSO** orbits achieved a **100% success rate**, demonstrating total mission reliability.
- Lower success rates in orbits like **GTO** are typically linked to higher mission frequency and early-stage technical iterations.



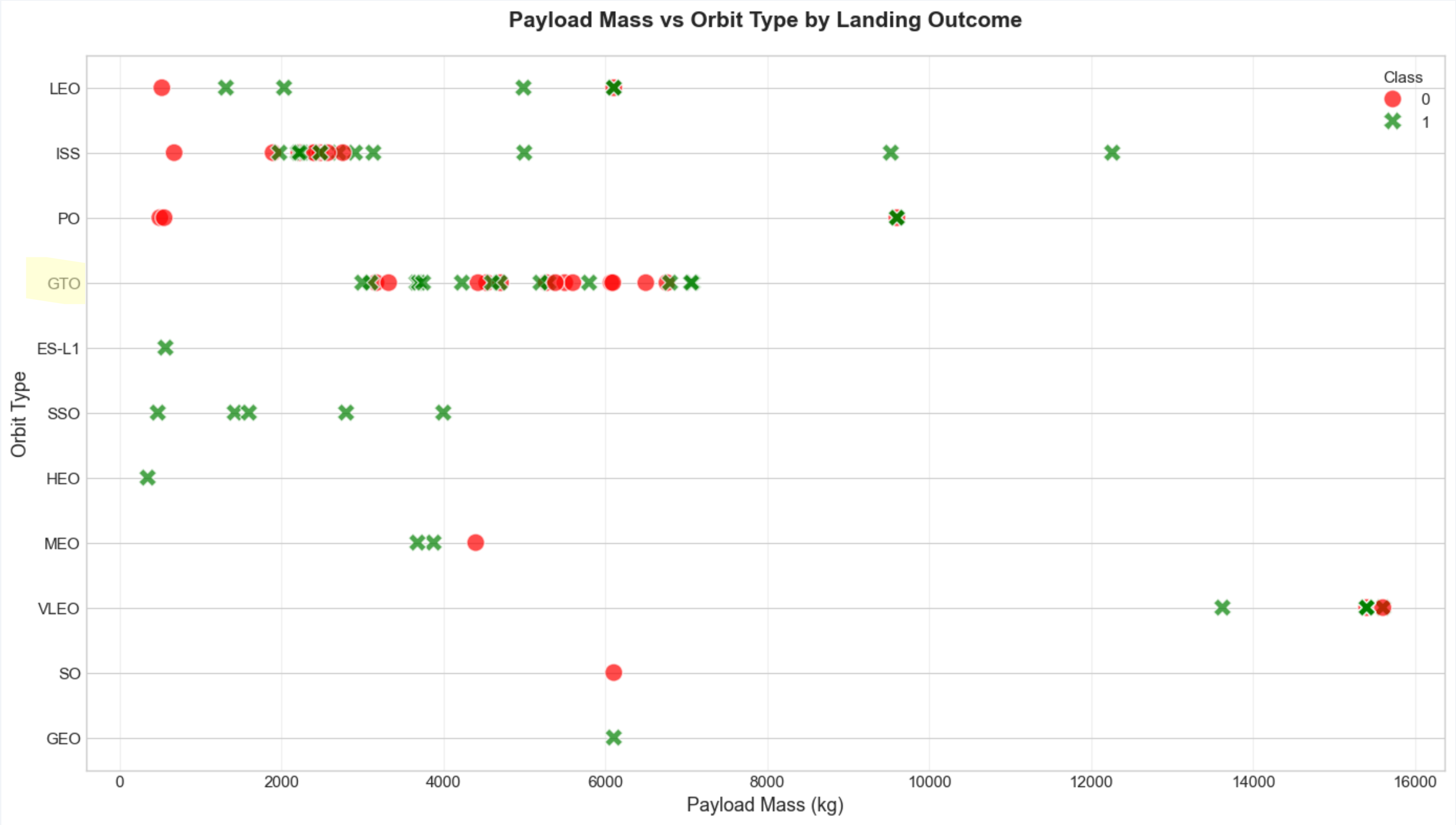
Flight Number vs. Orbit Type

- LEO orbit: Success rate improves with flight number (experience effect)
- GTO orbit: Success appears less correlated with flight number
- Different orbit types may require different landing strategies
- SpaceX's learning curve is more visible in frequently-used orbits

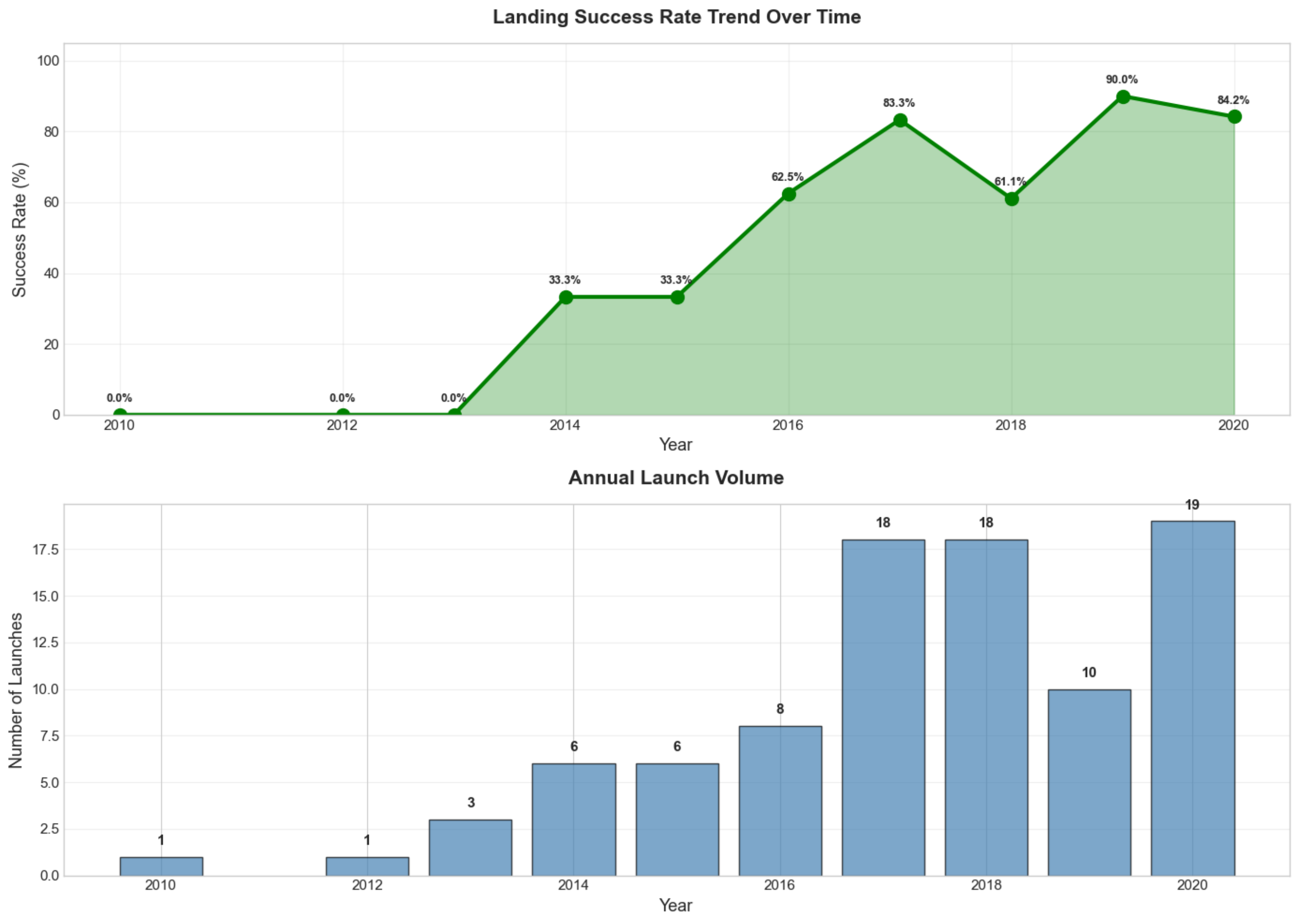


Payload vs. Orbit Type

- Heavy payloads show better success in **Polar, LEO,** and **ISS** orbits
- **GTO orbit**: Mixed results across payload ranges
- Payload-orbit combinations require careful mission planning
- Some orbits are better suited for heavy payload recovery



Launch Success Yearly Trend



- Significant improvement in success rate from 2013 onwards
- Success rate stabilized around 2014, then continued upward.
- After 2015, consistent improvement in landing reliability
- Demonstrates SpaceX's iterative learning and technology improvements

All Launch Site Names

- This query identifies all unique launch facilities used by SpaceX for Falcon 9 missions.
- The dataset contains four primary launch sites.
- Each site serves strategic mission requirements based on payload mass and orbital destination.

```sql
SELECT DISTINCT TRIM("Launch_Site")
FROM SPACEXTABLE
ORDER BY Launch_Site;
```

```
[TASK 1] Unique Launch Sites
----------------------------------
Number of unique launch sites: 4
  TRIM("Launch_Site")
0         CCAFS LC-40
1        CCAFS SLC-40
2         KSC LC-39A
3        VAFB SLC-4E
```

Launch Site Names Begin with 'CCA'

- This query retrieves the first 5 launch records from Cape Canaveral Air Force Station (CCAFS) facilities.
- CCAFS SLC 40 is SpaceX's highest-volume launch site, handling diverse mission profiles ranging from LEO to GTO orbits.
- These early records help establish baseline performance metrics for SpaceX's most frequently utilized launch facility

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

```
[TASK 2] Launches from sites beginning with 'CCA'
----------------------------------------
         Date Time (UTC) Booster_Version  Launch_Site  \
0  2010-06-04   18:45:00  F9 v1.0  B0003  CCAFS LC-40
1  2010-12-08   15:43:00  F9 v1.0  B0004  CCAFS LC-40
2  2012-05-22    7:44:00  F9 v1.0  B0005  CCAFS LC-40
3  2012-10-08    0:35:00  F9 v1.0  B0006  CCAFS LC-40
4  2013-03-01   15:10:00  F9 v1.0  B0007  CCAFS LC-40
```

Total Payload Mass

- This query calculates the cumulative payload mass delivered by SpaceX for NASA's Commercial Resupply Services (CRS) missions to the International Space Station.
- The total payload represents SpaceX's critical contribution to sustaining ISS operations, demonstrating their reliability in delivering cargo for long-duration space missions.
- NASA CRS missions typically involve payloads between 2,000-3,000 kg per flight.

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

```
[TASK 3] Total Payload Mass - NASA CRS Missions
----------------------------------------
Total payload mass for NASA CRS missions: 45,596 kg
```

Average Payload Mass by F9 v1.1

- This query computes the average payload capacity for the Falcon 9 v1.1 booster variant, an intermediate-generation rocket that operated between 2013-2016.
- Understanding version-specific payload averages helps identify technological improvements across SpaceX's booster evolution.
- The F9 v1.1 represented a significant upgrade over v1.0, featuring 60% more thrust and stretched fuel tanks

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

```
[TASK 4] Average Payload Mass - F9 v1.1
----------------------------------------
Average payload mass for F9 v1.1: 2,928.40 kg
```

First Successful Ground Landing Date

- This query identifies the historic date of SpaceX's first successful ground pad landing, marking a revolutionary achievement in rocket reusability.
- Ground pad landings (RTLS - Return to Launch Site) are technically more challenging than drone ship landings due to higher fuel requirements for boostback burns, but offer faster refurbishment turnaround times.

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

```
[TASK 5] First Successful Ground Pad Landing
----------------------------------------
First successful ground pad landing: 2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query identifies booster versions that successfully landed on autonomous drone ships while carrying mid-range payloads (4,000-6,000 kg).
- This payload range represents optimal conditions for drone ship recovery—heavy enough to require offshore landing due to insufficient fuel for RTLS, yet light enough to retain adequate fuel reserves for controlled descent and landing.

```sql
SELECT 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
ORDER BY PAYLOAD_MASS__KG_ DESC;
```

```
[TASK 6] Successful Drone Ship Landings (4000-6000 kg payload)
----------------------------------------
Number of successful drone ship landings in range: 4
  Booster_Version  PAYLOAD_MASS__KG_       Landing_Outcome
0  F9 FT  B1021.2               5300  Success (drone ship)
1  F9 FT  B1031.2               5200  Success (drone ship)
2     F9 FT B1022               4696  Success (drone ship)
3     F9 FT B1026               4600  Success (drone ship)
```

Total Number of Successful and Failure Mission Outcomes

- This query provides a comprehensive breakdown of all mission outcomes, categorizing results into success (ground pad, drone ship, ocean) and failure states.
- The approximately 100 success count(98+1+1) demonstrates SpaceX's progressive mastery of first-stage recovery technology, with success rates dramatically improving after 2015 as operational experience accumulated.

```
SELECT
    Mission_Outcome,
    COUNT(*) AS Count
FROM SPACEXTABLE
GROUP BY Mission_Outcome
ORDER BY Count DESC;
```

```
[TASK 7] Mission Outcome Summary
----------------------------------------
                  Mission_Outcome  Count
0                         Success     98
1  Success (payload status unclear)    1
2                        Success       1
3                Failure (in flight)   1
```

Boosters Carried Maximum Payload

- This query identifies which booster version(s) successfully delivered the heaviest payload in the dataset.
- Maximum payload capacity serves as a key performance indicator for rocket capability and demonstrates the engineering limits of each booster generation.
- Heavier payloads typically correlate with reduced landing success probability due to lower remaining fuel margins.

```sql
SELECT DISTINCT Booster_Version, PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
)
ORDER BY Booster_Version;
```

```
[TASK 8] Booster Versions Carrying Maximum Payload
----------------------------------------
Maximum payload mass: 15,600 kg
```

	Booster_Version	PAYLOAD_MASS__KG_
0	F9 B5 B1048.4	15600
1	F9 B5 B1048.5	15600
2	F9 B5 B1049.4	15600
3	F9 B5 B1049.5	15600
4	F9 B5 B1049.7	15600
5	F9 B5 B1051.3	15600
6	F9 B5 B1051.4	15600
7	F9 B5 B1051.6	15600
8	F9 B5 B1056.4	15600
9	F9 B5 B1058.3	15600
10	F9 B5 B1060.2	15600
11	F9 B5 B1060.3	15600

2015 Launch Records

- This query analyzes failed drone ship landing attempts during 2015, a pivotal year in SpaceX's learning curve.
- The year 2015 saw multiple high-profile landing failures before SpaceX achieved their first successful drone ship landing in April 2016.
- These failures provided critical engineering data that informed subsequent design improvements to landing legs, grid fins, and autonomous guidance systems.

```
SELECT
    substr(Date, 6, 2) AS Month,
    Landing_Outcome,
    Booster_Version,
    Launch_Site
FROM SPACEXTABLE
WHERE Landing_Outcome LIKE '%Failure (drone ship)%'
  AND substr(Date, 1, 4) = '2015'
ORDER BY Date;
```

```
[TASK 9] Failed Drone Ship Landings - 2015
----------------------------------------------
Number of drone ship failures in 2015: 2
  Month       Landing_Outcome Booster_Version  Launch_Site
0    01  Failure (drone ship)   F9 v1.1 B1012  CCAFS LC-40
1    04  Failure (drone ship)   F9 v1.1 B1015  CCAFS LC-40
```

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

- This query ranks landing outcomes by frequency during SpaceX's formative years (2010-2017), capturing the transition from experimental attempts to operational reliability.
- This timeframe encompasses SpaceX's first landing attempts through the establishment of routine recovery operations.
- The ranking reveals which outcome types dominated during this critical developmental period.

```
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;
```

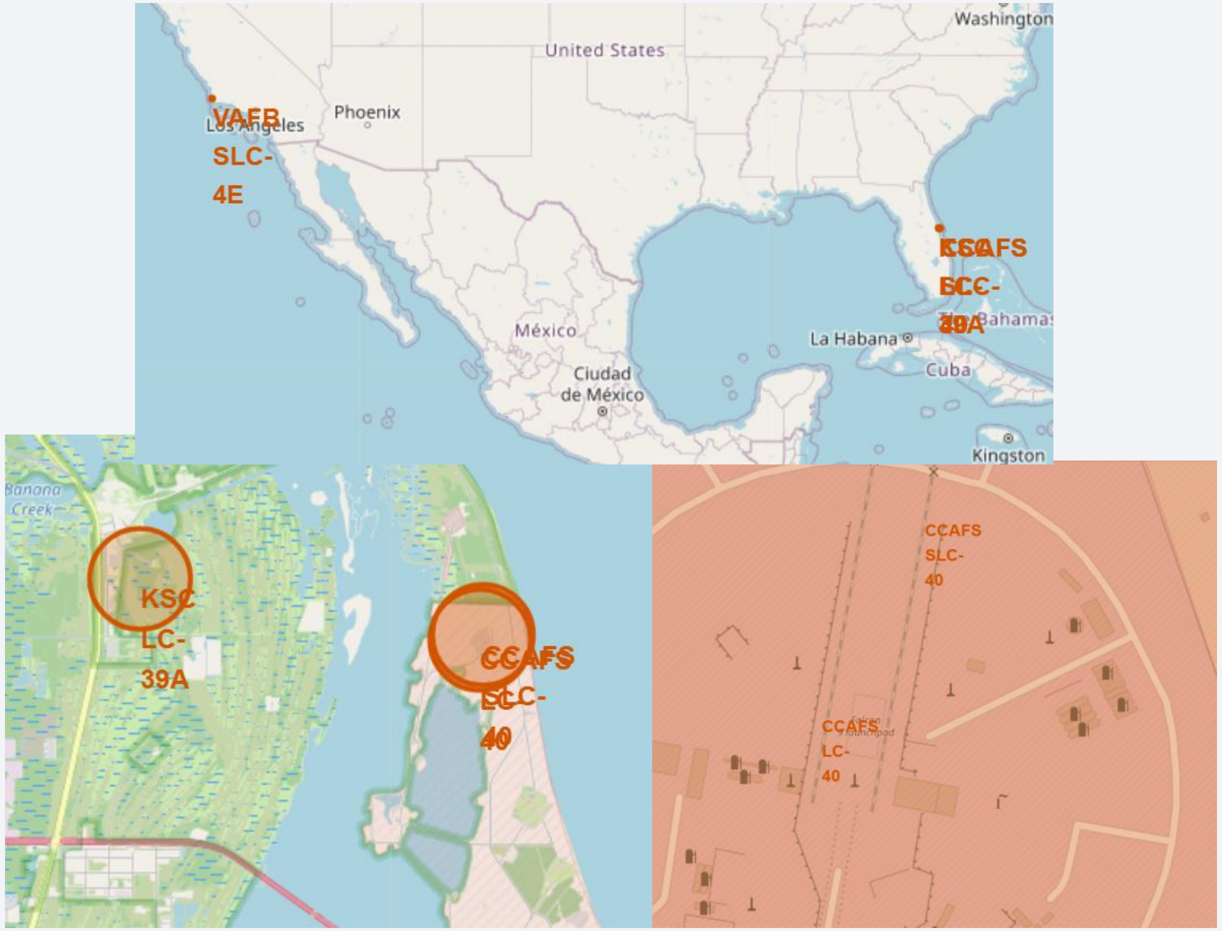
```
[TASK 10] Landing Outcome Rankings (June 2010 - March 2017)
----------------------------------------
         Landing_Outcome  Outcome_Count
0             No attempt             10
1     Success (drone ship)            5
2     Failure (drone ship)            5
3    Success (ground pad)             3
4       Controlled (ocean)            3
5     Uncontrolled (ocean)            2
6      Failure (parachute)            2
7   Precluded (drone ship)            1
```

Section 3

Launch Sites Proximities Analysis

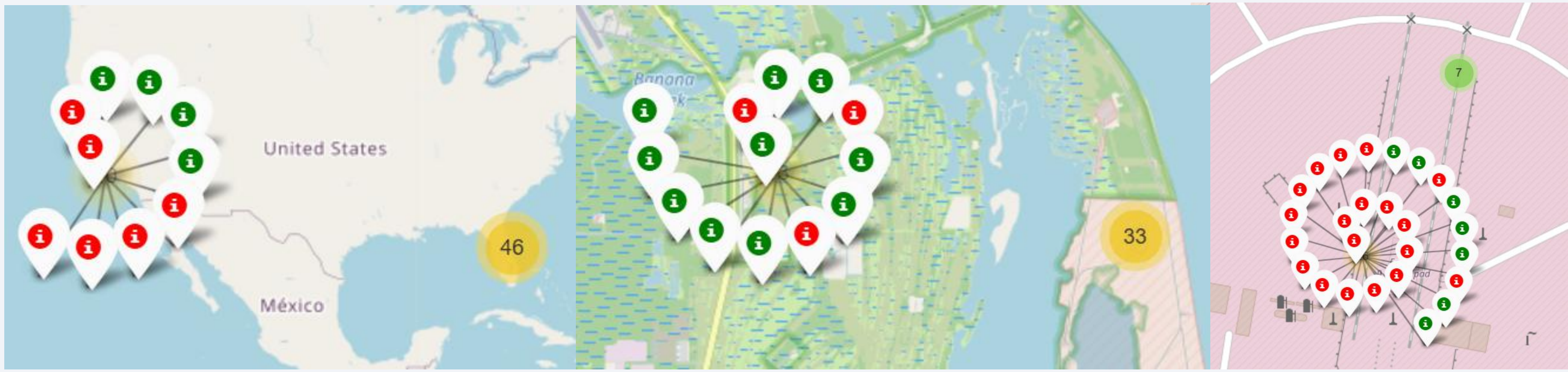
Launch Site Map

- All three SpaceX launch sites positioned on coastlines at lower latitudes (<35°N) for safety (ocean abort trajectories) and fuel efficiency (Earth's rotational boost).
- Florida facilities (CCAFS, KSC) dominate due to eastward launches toward equatorial orbits; VAFB serves polar missions requiring southward trajectories.



Interactive Launch Outcome Map

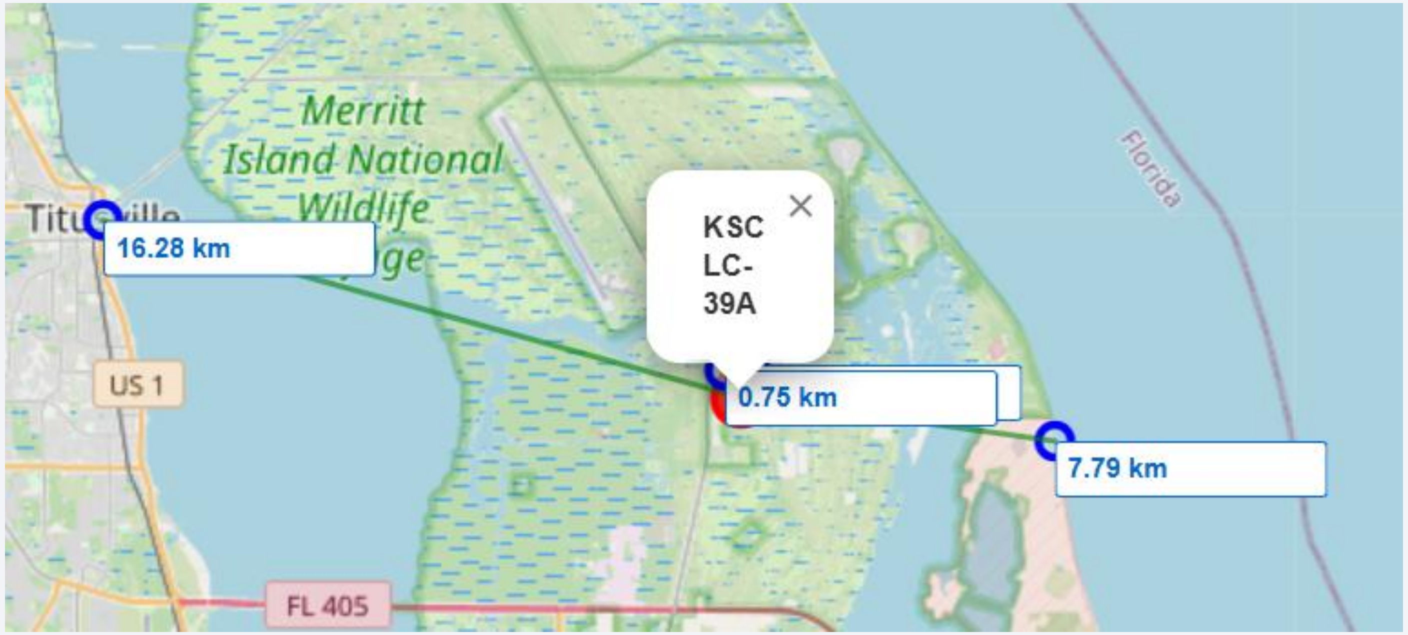
- Green (success) markers heavily concentrate at KSC LC-39A and CCAFS SLC 40, with density increasing in later flight numbers.
- Red (failure) markers cluster in 2013-2015 period, validating the learning curve hypothesis.
- VAFB shows balanced outcomes due to specialized polar orbit challenges.



Distance Analysis Map

Infrastructure Proximity Analysis:

- Coastline: <5 km (enables safe ocean trajectories)
- Highways (US-1/I-95): 10-15 km (equipment transport)
- Railways: <3 km at KSC (heavy cargo logistics)
- Urban areas: >20 km safety buffer



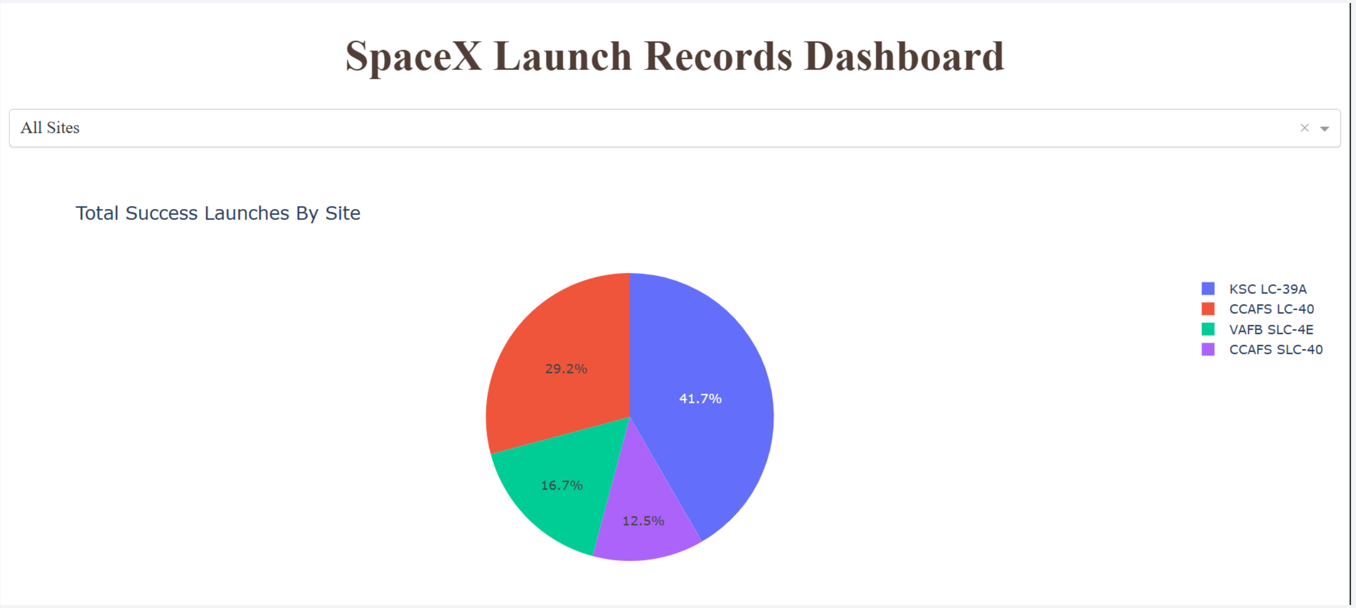
Insight: KSC's railway access explains its preference for superheavy payloads; all sites face ocean for abort safety.

Section 4

Build a Dashboard with Plotly Dash

All Sites Success Distribution

- **Total Success Distribution (All Sites):** KSC LC-39A leads with 41.7% of all successful launches, followed by CCAFS LC-40 at 29.2%, VAFB SLC-4E at 16.7%, and CCAFS SLC-40 at 12.5%. KSC LC-39A dominates both in absolute success count AND success rate, confirming its status as SpaceX's most reliable launch facility.

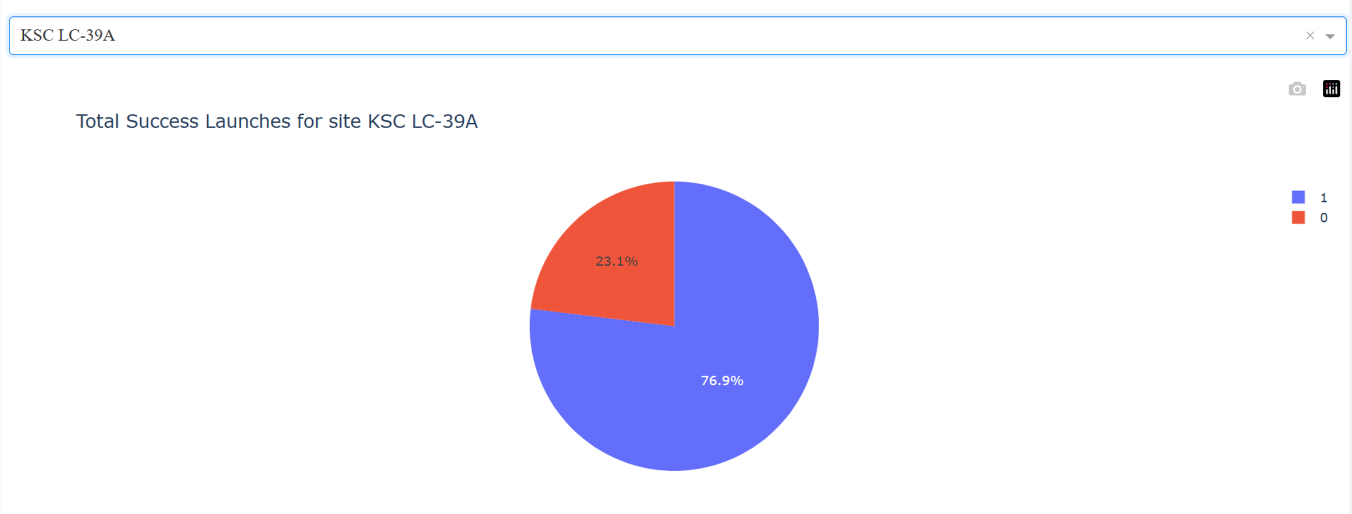


Total Success Launches For KSC LC-39A

KSC LC-39A Performance: Highest success rate at 76.9%, with minimal failures. Superior performance stems from:

1. use in later mission years with mature technology.
2. selection for high-priority missions requiring extra scrutiny.
3. premium Apollo/Shuttle-era infrastructure.

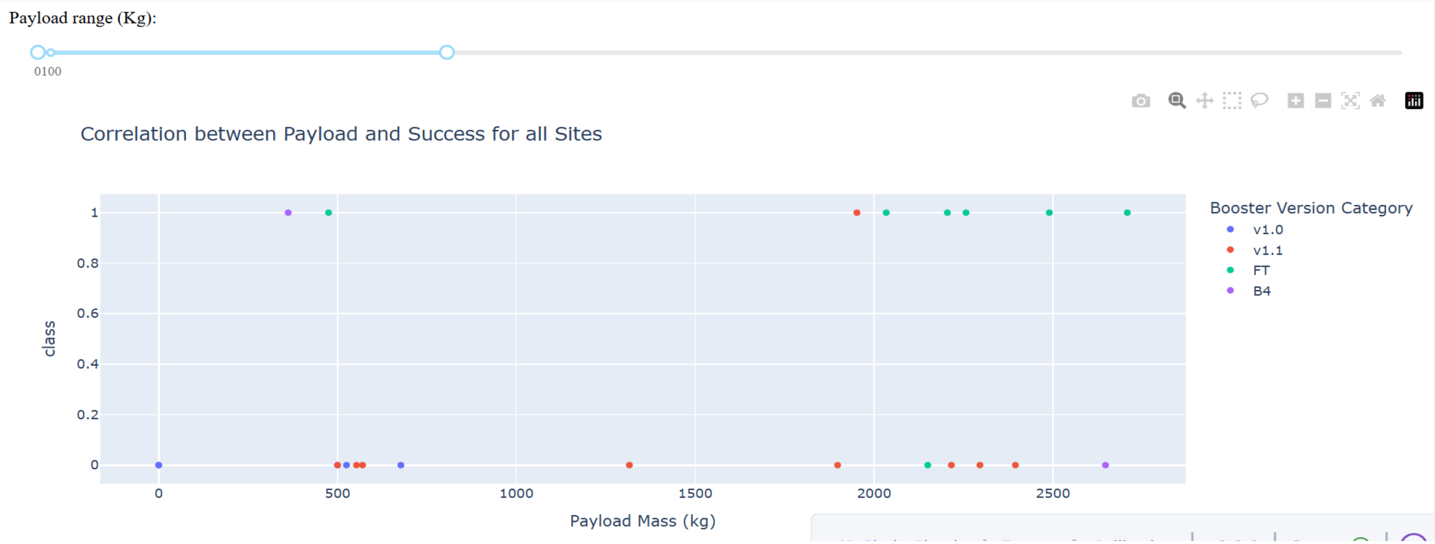
Validates KSC as SpaceX's flagship facility.



Correlation Between Payload and Success for All Sites

Payload Sweet Spot:

- 2,000-5,500 kg range shows highest success density—sufficient fuel remains for landing without sacrificing payload revenue.
- B4 boosters extend success to 10,000+ kg, demonstrating technological evolution.
- FT boosters cluster at 2,000-4,000 kg.



Section 5

Predictive Analysis (Classification)

Classification Accuracy

Model Performance Comparison:

Best Models: Logistic Regression and SVM tied at **83.33%** test accuracy

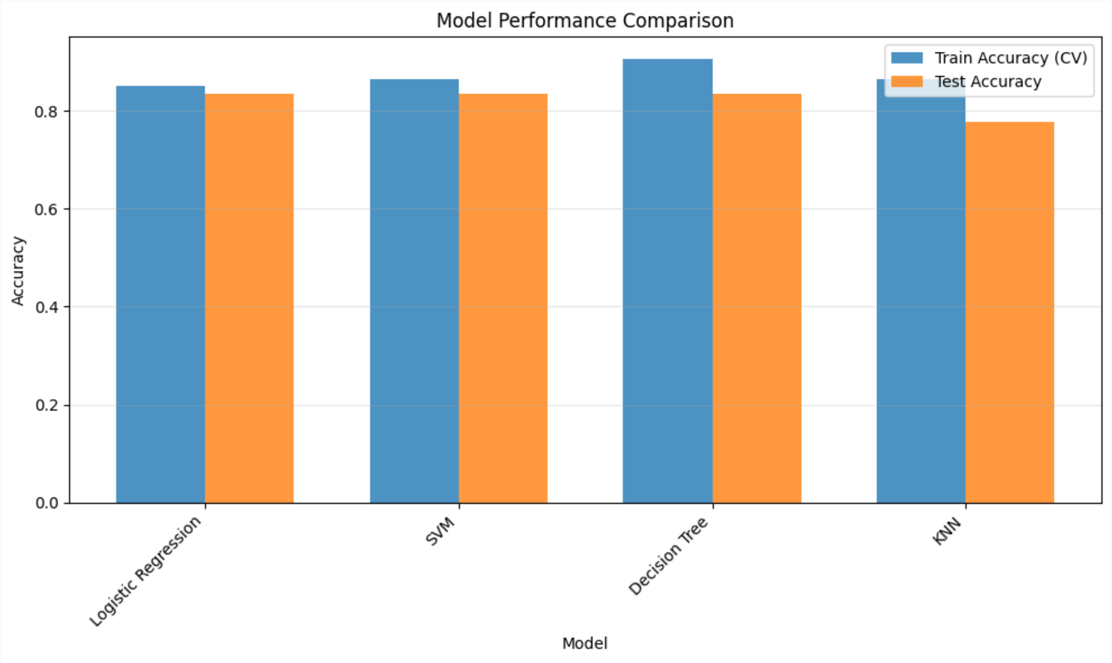
Mid-tier: KNN achieved 77.78%

Underperformer: Decision Tree at 61.11% (overfitting to training data)

Logistic Regression selected as champion model:

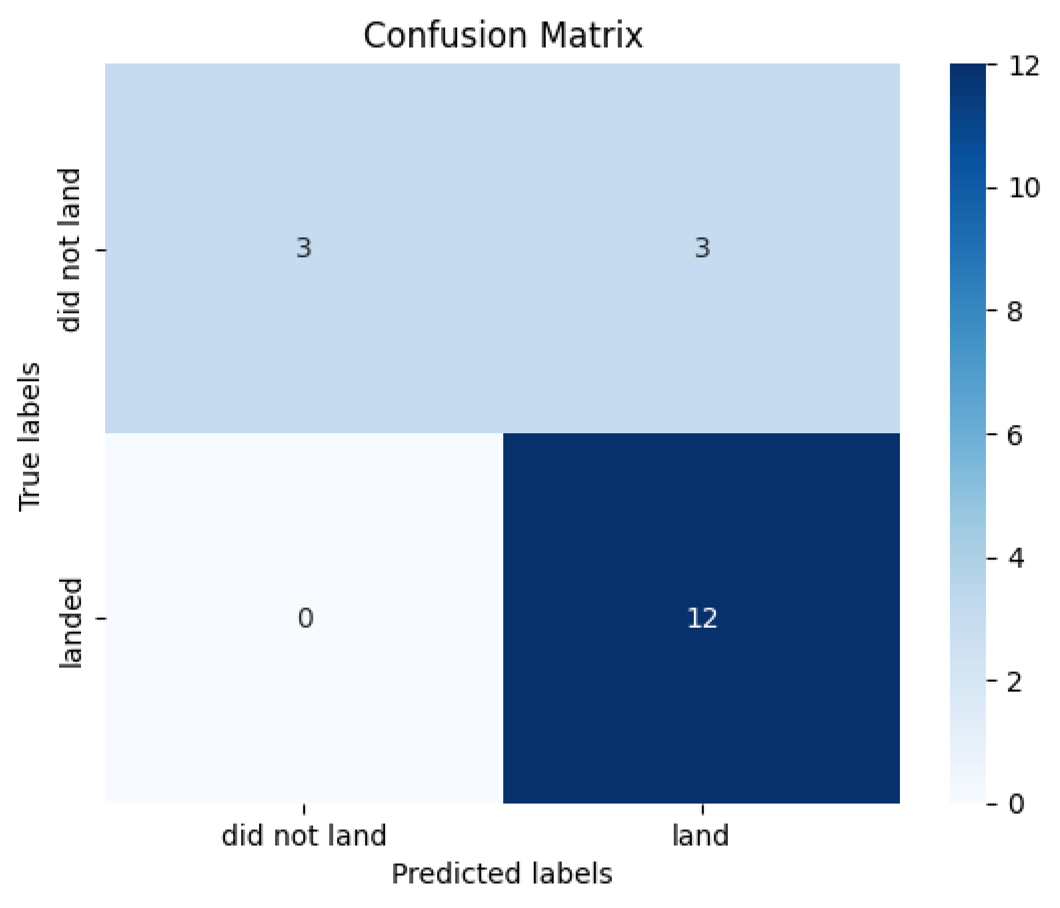
- Tied-best accuracy with SVM.
- Interpretable coefficients reveal feature importance.
- Computationally efficient for real-time predictions.
- Stable performance without hyperparameter complexity

Key Insight: Decision Tree's high training accuracy (90.5%) but poor test performance (61.1%) indicates overfitting—the model memorized training patterns that don't generalize."



Confusion Matrix- Logistic Regression

- The model achieved 15 correct predictions (3 TN + 12 TP) out of 18 test samples.
- Notably, **zero False Negatives** means the model never missed a successful landing—critical for accurately estimating SpaceX's reusability cost advantage.
- The 3 False Positives (predicting success when missions failed) represent conservative optimism, acceptable for competitive intelligence purposes.

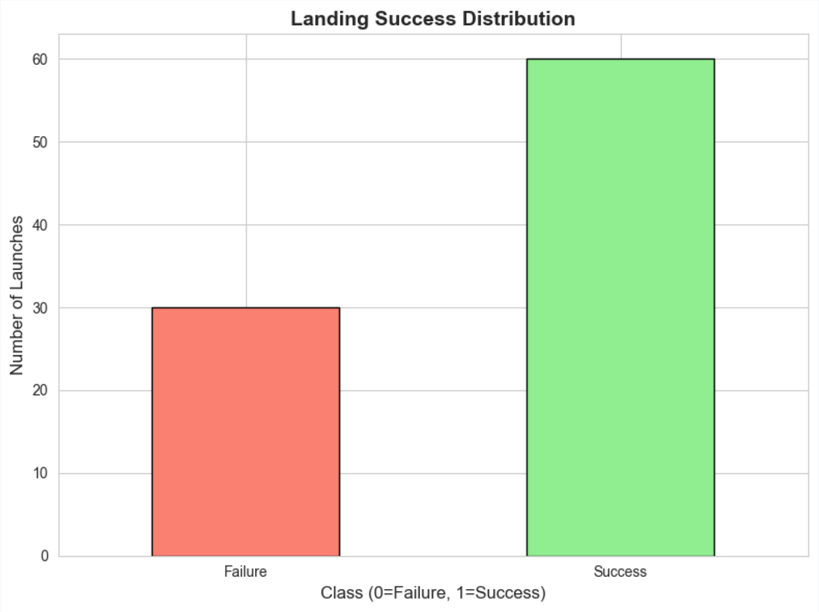


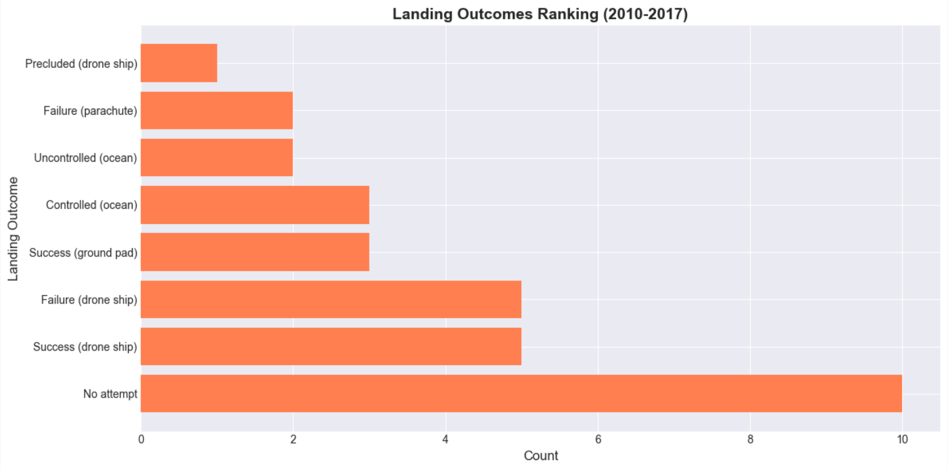
Conclusions

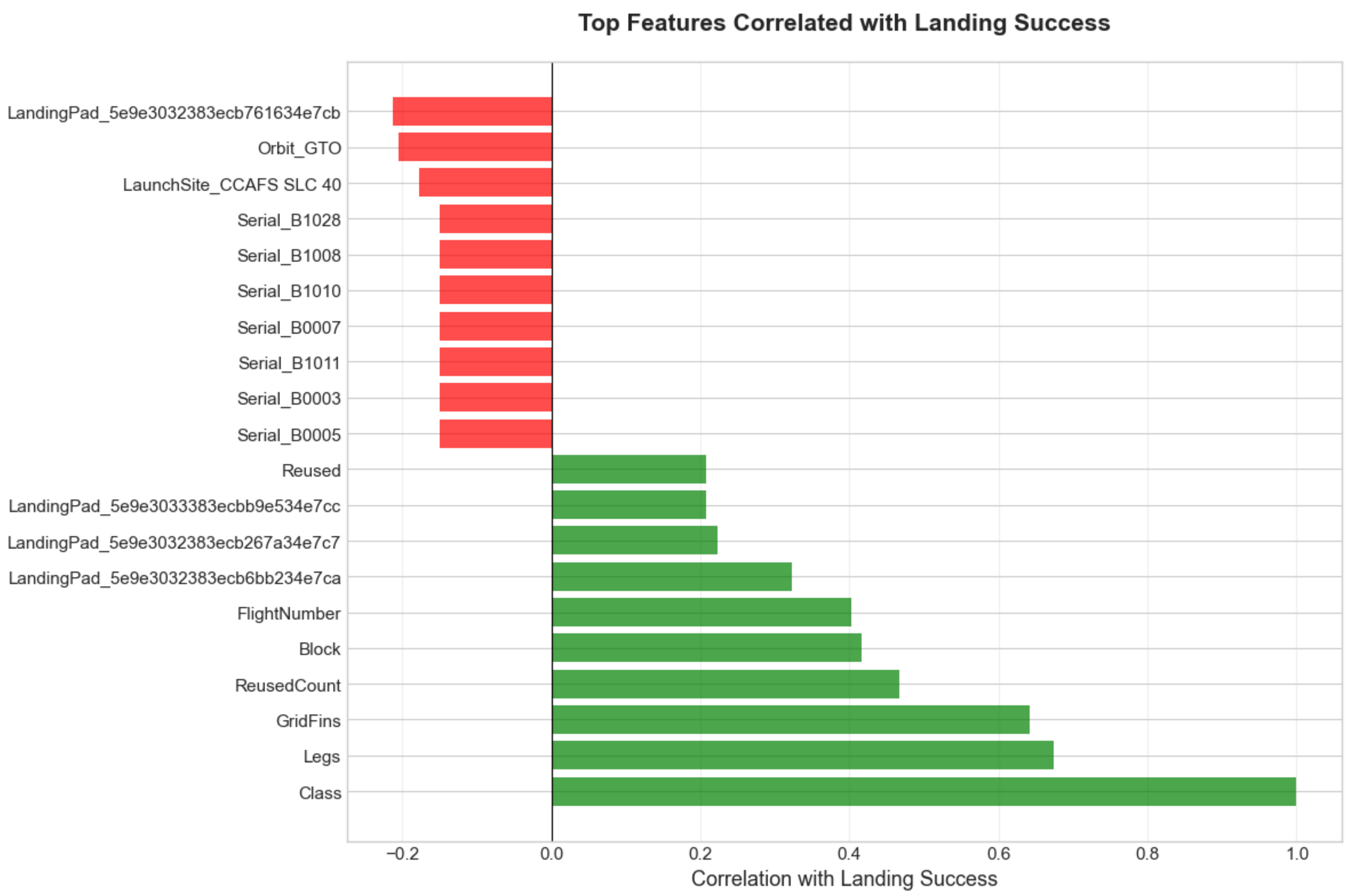
- **Predictive Success:** Achieved **83.33% accuracy** using Logistic Regression, with perfect recall for successful landings (zero False Negatives). Enables reliable cost estimation for competitive launch bidding against SpaceX
- **Experience Drives Reliability:** Launch success rates demonstrate a strong positive correlation with flight number and year, with performance peaking in 2020 as SpaceX's operational maturity reached industrial scale.
- **Mission-Specific Optimization:** Heavy payloads (10,000+ kg) achieve highest success at KSC LC-39A and VAFB SLC 4E, while certain orbit types (ES-L1, GEO, HEO, SSO) demonstrate 100% landing success rates, indicating mission-specific engineering optimization.
- **Strategic Launch Site Selection:** CCAFS SLC 40 serves as the high-volume hub for diverse missions, KSC LC-39A specializes in heavy/complex payloads, and VAFB SLC 4E handles polar orbit requirements—demonstrating that launch sites are strategically differentiated, not interchangeable.
- **Business Intelligence Value:** The payload sweet spot of 2,000-5,000 kg shows optimal success rates, providing actionable intelligence for competitors pricing alternative launch services."

Appendix

GitHub Repository URL : [SpaceX-Falcon9-Landing-Prediction]






Top Features Correlated with Landing Success
LandingPad_5e9e3032383ecb761634e7cb
Orbit_GTO
LaunchSite_CCAFS SLC 40
Serial_B1028
Serial_B1008
Serial_B1010
Serial_B0007
Serial_B1011
Serial_B0003
Serial_B0005
Reused
LandingPad_5e9e3033383ecbb9e534e7cc
LandingPad_5e9e3032383ecb267a34e7c7
LandingPad_5e9e3032383ecb6bb234e7ca
FlightNumber
Block
ReusedCount
GridFins
Legs
Class
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
Correlation with Landing Success

Thank you!