IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Francesco Jr
10/01/2025

Outline

Executive Summary

Summary of Methodologies

Data Preprocessing

- Feature Extraction: Converted categorical variables using one-hot encoding (83 features total)
- Data Standardization: Applied StandardScaler to normalize all features
- Train-Test Split: 80% training (72 samples) / 20% testing (18 samples), random_state=2

Hyperparameter Optimization

- Technique: GridSearchCV with 10-fold cross-validation
- Models Evaluated:
    - Logistic Regression (L2 regularization)
    - Support Vector Machine (RBF, Linear, Poly, Sigmoid kernels)
    - Decision Tree Classifier
    - K-Nearest Neighbors

Model Evaluation

- Primary Metric: Accuracy score on test set
- Validation: Confusion matrix analysis
- Comparison: Best parameters and validation scores for all models

Model	Test Accuracy	Validation Accuracy	Best Hyperparameters
Logistic Regression	83.33%	84.64%	C=0.01, penalty='l2', solver='lbfgs'
Support Vector Machine	83.33%	N/A	C=1.0, gamma=0.0316, kernel='sigmoid'
K-Nearest Neighbors	83.33%	N/A	n_neighbors=10, algorithm='auto', p=1
Decision Tree	72.22%	N/A	criterion='entropy', max_depth=16, max_features='sqrt'

Introduction

Project Background and Context

The Space Launch Industry

- Traditional Launch Cost: $165+ million per rocket launch (competitors)
- SpaceX Falcon 9 Cost: $62 million per launch
- Cost Advantage: SpaceX's ability to reuse the first stage creates 62% cost savings

SpaceX First Stage Landing Program

- Program Start: 2013 (first controlled descent tests)
- First Successful Ground Landing: December 2015 (Flight 20)
- First Successful Drone Ship Landing: April 2016
- Current Success Rate: 512 successful landings out of 525 attempts (97.5%)
- Block 5 Performance: 487/493 successful landings (98.8%)

Business Context

- Market Disruption: SpaceX's reusability revolutionized space industry economics
- Competitive Intelligence: Competitors need to predict launch costs to bid effectively
- Data Opportunity: 90 historical launches provide machine learning training data

Dataset Overview

- Time Period: 2010-2020 (90 launches analyzed)
- Target Variable: First stage landing success (Class: 0=Failure, 1=Success)
- Features: Launch site, payload mass, orbit type, booster specifications, reuse history

Problems You Want to Find Answers

Primary Research Question

Can we predict if the Falcon 9 first stage will land successfully?

Specific Problems to Solve

1. Cost Estimation Challenge
    - How can competitors accurately predict SpaceX launch costs?
    - Which factors most influence landing success/failure?
2. Feature Impact Analysis
    - Does payload mass affect landing probability?
    - Do certain launch sites have higher success rates?
    - Does booster reuse history impact landing outcomes?
    - How do different orbit types affect landing success?
3. Model Selection Problem
    - Which machine learning algorithm provides best predictions?
    - What accuracy level is achievable with available data?
    - Can we reliably predict with limited sample size (90 launches)?
4. Hyperparameter Optimization
    - What are optimal parameters for each ML model?
    - How do we avoid overfitting with small dataset?
    - Which validation strategy works best?
5. Business Decision Support
    - Can we achieve >80% prediction accuracy for business use?
    - What confidence level can stakeholders expect?
    - How do false positives/negatives impact cost estimates?

Success Criteria

- Target Accuracy: >80% on test data
- Model Interpretability: Clear understanding of key factors
- Practical Application: Actionable insights for competitive bidding

Section 1

Methodology

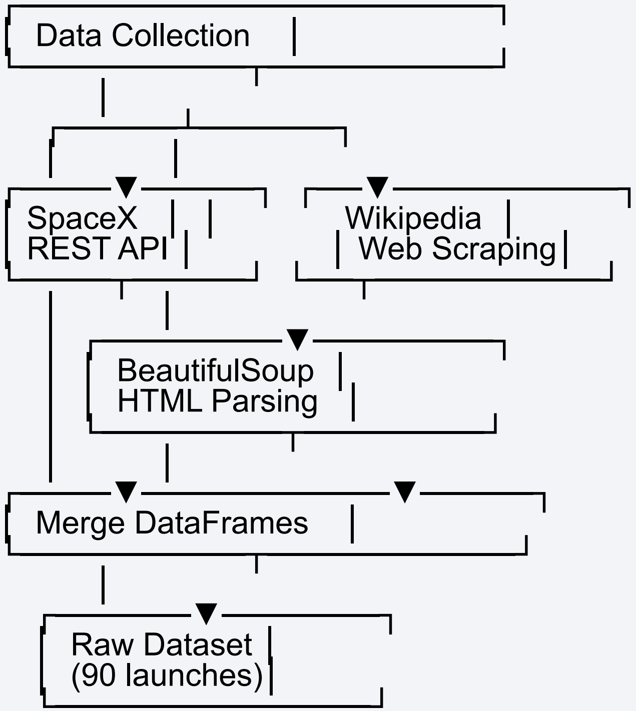
Methodology

Executive Summary

- Data collection methodology:
    - Data was collected via SpaceX REST API for launch records and web scraping from Wikipedia using BeautifulSoup to extract Falcon 9 historical launch data.
- Perform data wrangling
    - Landing outcomes were converted to binary classification (0=failure, 1=success), missing values filled with mean, and categorical features one-hot encoded into 83 numerical columns
- 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 4 models (Logistic Regression, SVM, Decision Tree, KNN), tuned hyperparameters using GridSearchCV with 10-fold cross-validation, and evaluated using accuracy scores and confusion matrices on test data.

Data Collection

- Data collected from SpaceX REST API (JSON requests) and web scraping Wikipedia tables using BeautifulSoup, then merged into unified dataset.



Data Collection – SpaceX API

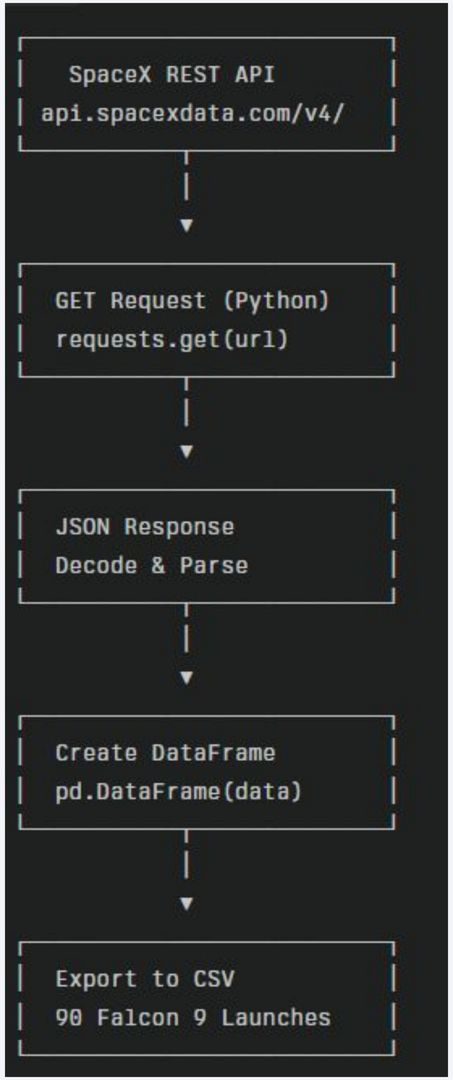
How Data Was Collected Data was collected via SpaceX REST API for launch records and web scraping from Wikipedia using BeautifulSoup to extract Falcon 9 historical launch data.

Key Phrases:

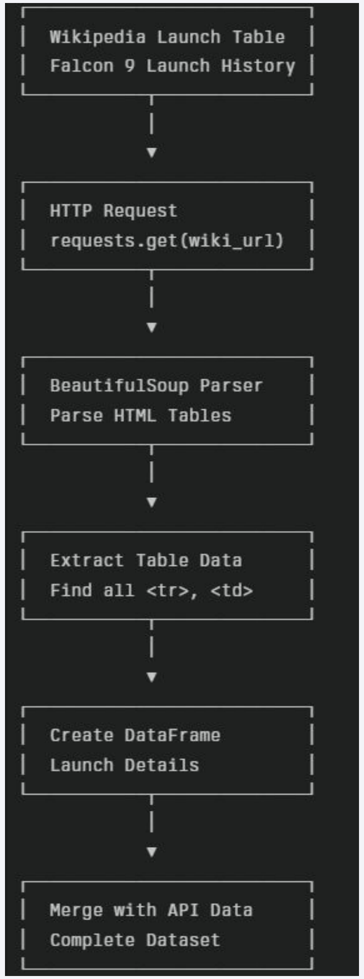
- API Endpoint: `https://api.spacexdata.com/v4/launches/past`
- Method: GET requests using Python `requests` library
- Data Format: JSON responses decoded to Python dictionaries
- Output: Pandas DataFrame with 90 launch records
- Export: Saved as CSV file for further analysis

GitHub Link:

[francescojr/datascienceraceibm: Notebook IBM Data Science]



Data Collection - Scraping



Key Phrases:

- Source: Wikipedia Falcon 9 launch records
- Library: BeautifulSoup4 for HTML parsing
- Target: Extract launch dates, sites, payloads, outcomes
- Validation: Cross-reference with API data

[francescojr/datascienceraceibm: Notebook IBM Data Science]

Data Wrangling

How Data Was Processed

Landing outcomes were converted to binary classification (0=failure, 1=success), missing values filled with mean, and categorical features one-hot encoded into 83 numerical columns.

Data Wrangling Steps:

1. Create Target Variable:
   - Converted landing outcomes to binary: Class (0=failure, 1=success)
2. Handle Missing Values:
   - Filled missing PayloadMass with column mean
   - Removed rows with critical missing data
3. Feature Engineering:
   - Applied one-hot encoding to categorical variables
   - Created 83 feature columns from: Launch Site, Orbit, Booster Serial, Grid Fins, Reused, Legs, Block
4. Data Export:
   - Saved cleaned dataset to CSV
   - Final dataset: 90 rows × 83 columns

EDA with Data Visualization

Exploratory Data Analysis (EDA)

Visualization Analysis:

- Scatter Plots: Flight Number vs Success Rate (shows improvement over time)
- Bar Charts: Success rate by Orbit Type (LEO, GTO, ISS) and Launch Site (CCAFS, VAFB, KSC)
- Line Plots: Payload Mass vs Landing Success
- Correlation Heatmap: Feature relationships

SQL Analysis:

- Total launches per site
- Average payload mass for successful landings
- Ranking of launch sites by success rate
- Booster reuse patterns
- Orbit type success statistics

Key Findings:

- Success rate improved significantly after 2013
- VAFB had highest success rate among launch sites
- LEO missions had higher success rates than GTO
- Payload mass showed weak correlation with success

EDA with SQL

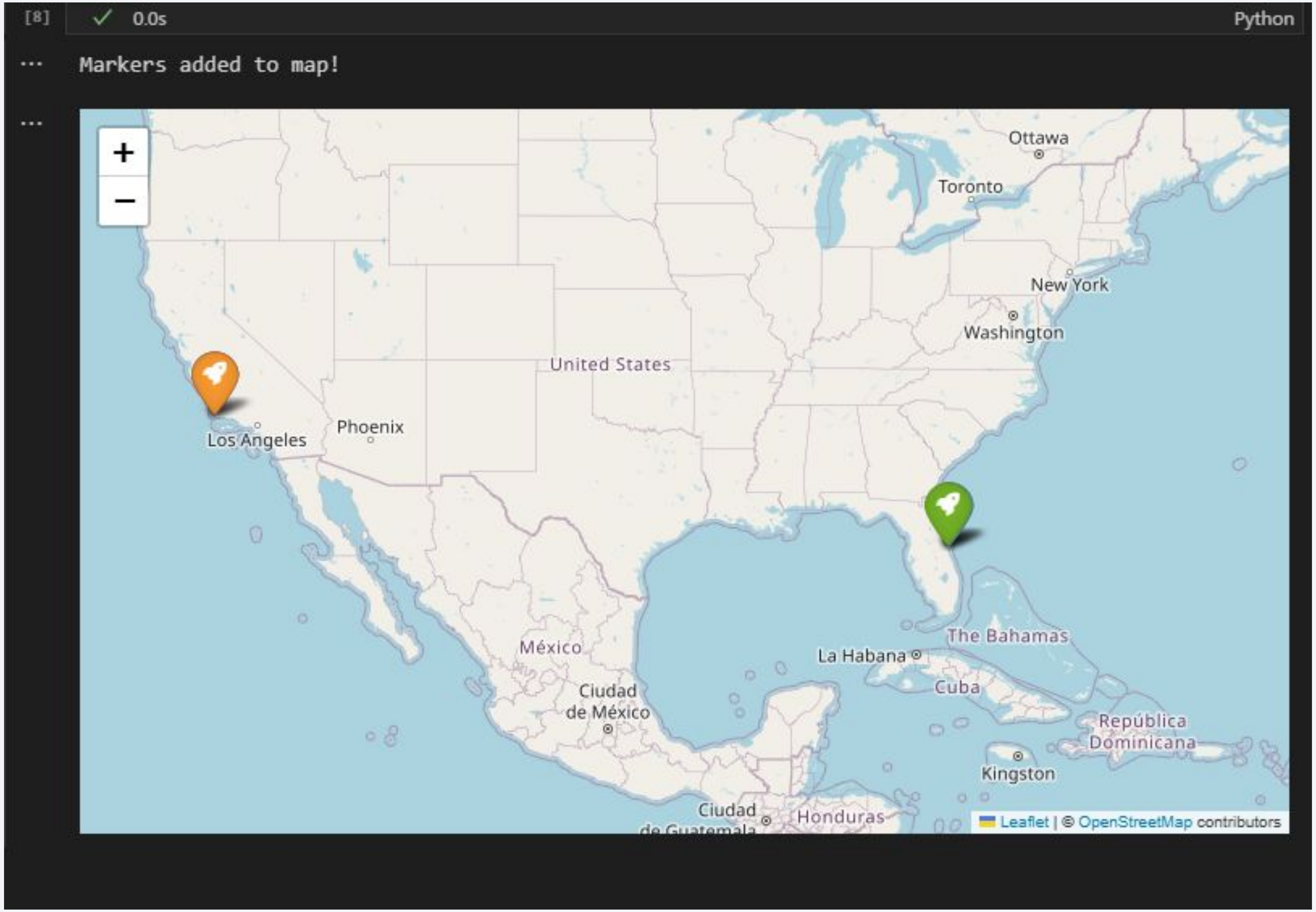
SQL Analysis Implementation

Database Query Categories:

Descriptive queries
: Unique sites and basic statistics
Filtering operations
: Conditional data extraction
Aggregation analysis
: Sum, average, min/max calculations
Grouping operations
: Performance analysis by categories
Temporal analysis
: Time-based trend identification

SQL Query Implementation Strategy:
Launch site identification and analysis
Customer-specific payload calculations
Date range analysis for outcomes
Mission success/failure statistical analysis

Build an Interactive Map with Folium



Build a Dashboard with Plotly Dash

Predictive Analysis (Classification)

Machine Learning Pipeline Implementation1.
Model Development Process:
Data preparation
: Stratified train/test split implementation2.
Feature preprocessing
: StandardScaler normalization3.

Model selection

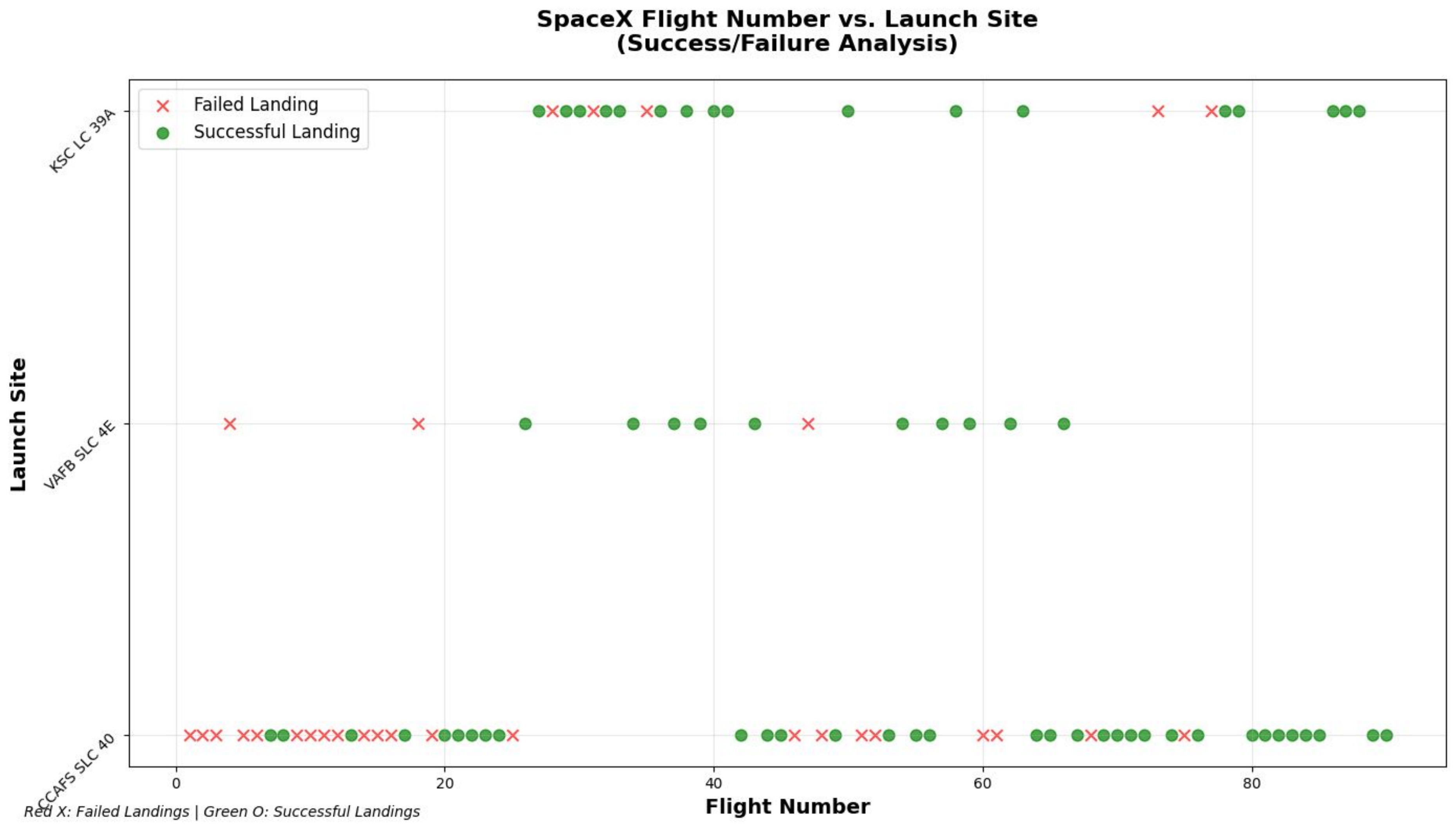
: Four algorithm comparison study4.
Hyperparameter optimization
: GridSearchCV with 10-fold cross-validation5.
Performance evaluation
: Multiple metric assessment

Classification Algorithms Tested:
Logistic Regression
: Linear probabilistic classification
Support Vector Machine
: Kernel-based classification
Decision Tree
: Tree-based rule learning
K-Nearest Neighbors
: Instance-based classification

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



Flight Number vs. Launch Site

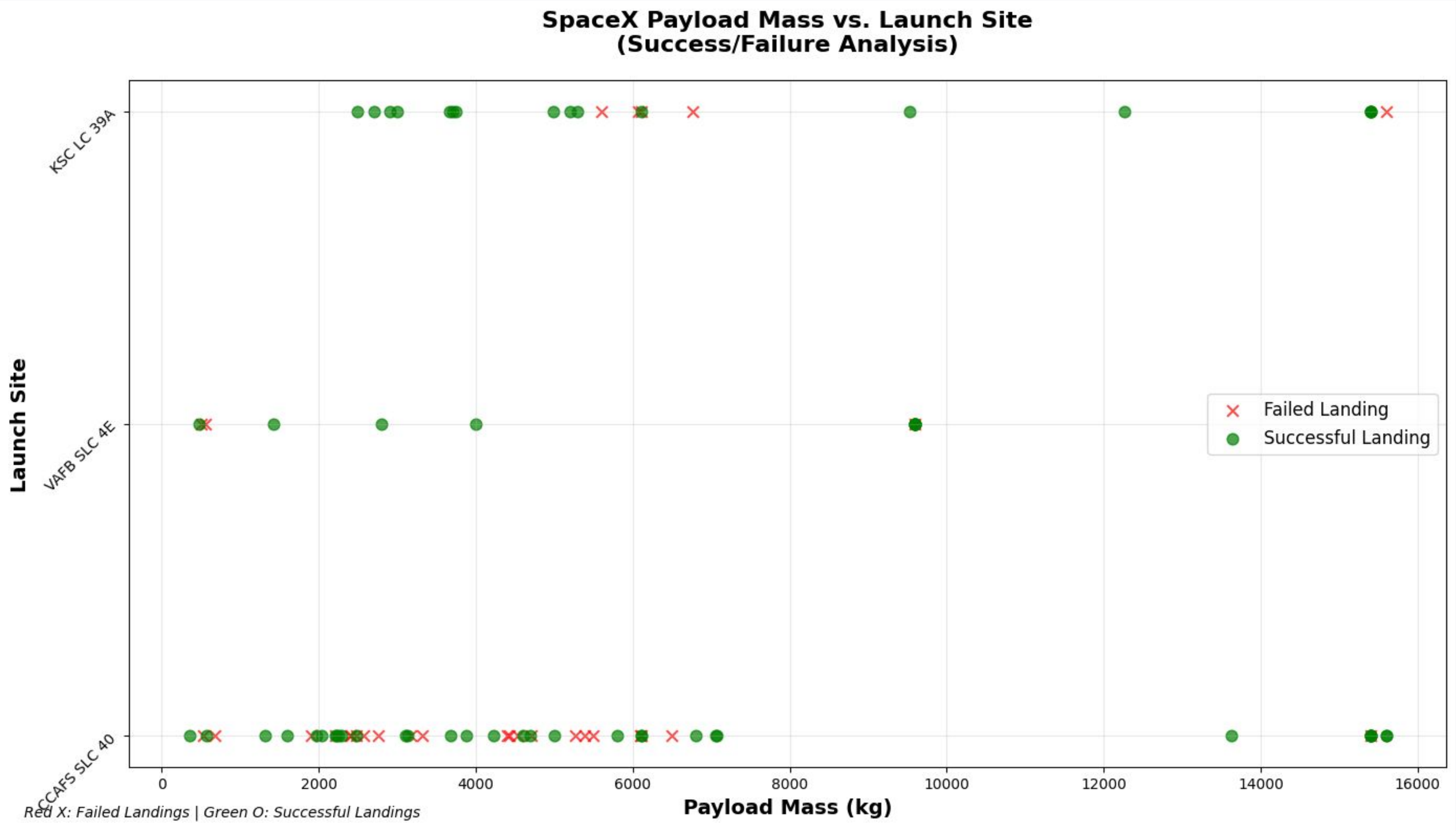
📊 SCATTER PLOT ANALYSIS: Flight Number vs. Launch Site
==
🔍 Key Observations:
• Each point represents a SpaceX mission
• X-axis: Flight Number (chronological order of missions)
• Y-axis: Launch Site location
• Green circles: Successful landings
• Red X marks: Failed landings

📈 Statistical Summary:
• Total missions plotted: 90
• Successful landings: 60 (66.7%)
• Failed landings: 30 (33.3%)

🚀 Launch Site Usage:
• CCAFS SLC 40: 55 missions (Success rate: 60.0%)
• KSC LC 39A: 22 missions (Success rate: 77.3%)
• VAFB SLC 4E: 13 missions (Success rate: 76.9%)

💡 Insights:
• Success rate generally improves with higher flight numbers (learning curve)
• Different launch sites show varying success patterns
• Early missions (lower flight numbers) had more failures as expected

Payload vs. Launch Site



Payload vs. Launch Site

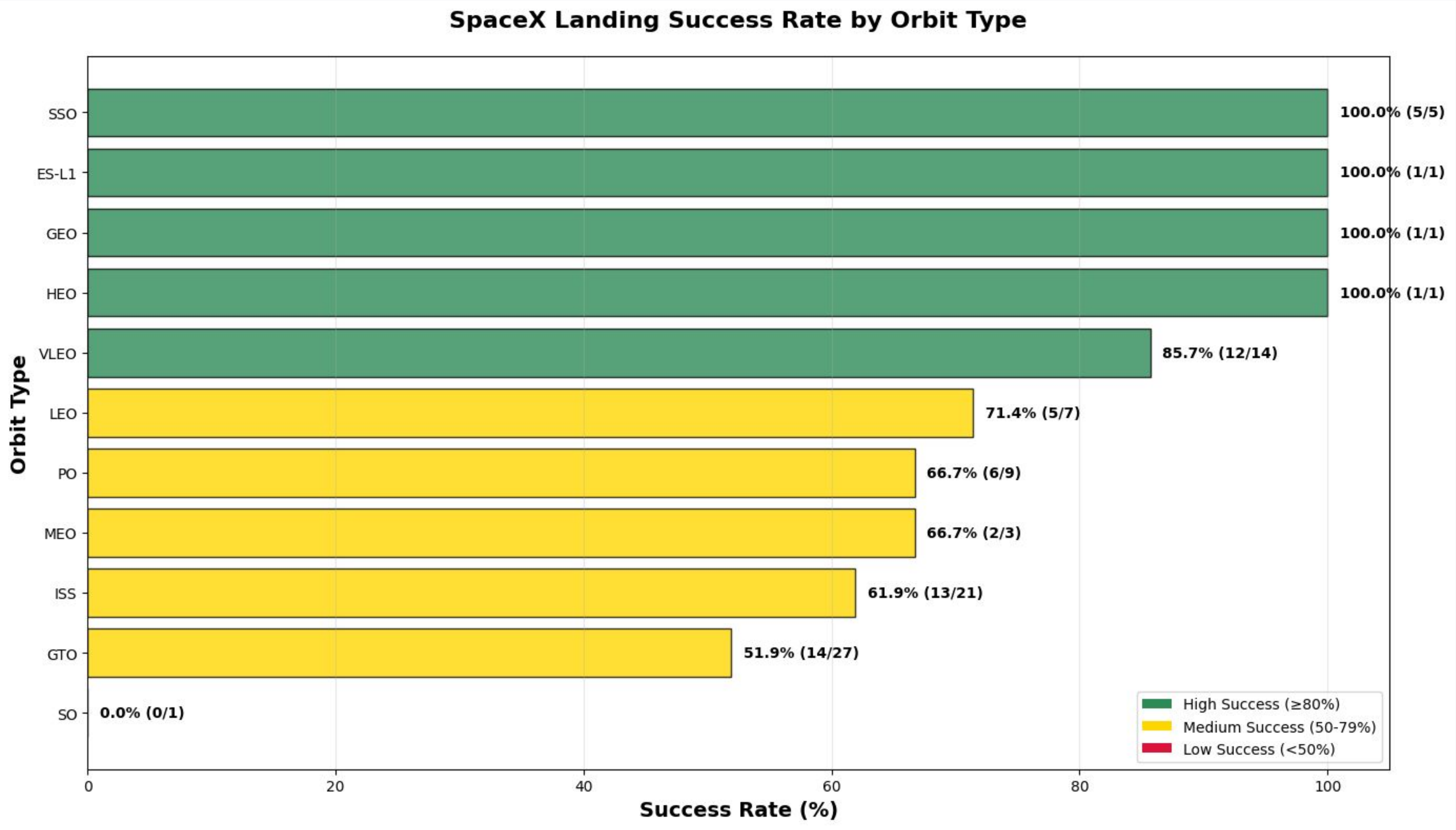
📊 SCATTER PLOT ANALYSIS: Payload Mass vs. Launch Site
==
🔍 Key Observations:
• Each point represents a SpaceX mission with known payload mass
• X-axis: Payload Mass in kilograms
• Y-axis: Launch Site location
• Green circles: Successful landings
• Red X marks: Failed landings

📈 Payload Statistics:
• Missions with payload data: 90
• Payload range: 350 - 15600 kg
• Average payload: 6105 kg
• Median payload: 4702 kg

🎯 Success Rate by Payload Range:
• 0-5000 kg: 47 missions (Success rate: 63.8%)
• 5000-10000 kg: 28 missions (Success rate: 60.7%)
• 10000-15000 kg: 2 missions (Success rate: 100.0%)
• 15000-+ kg: 13 missions (Success rate: 84.6%)

💡 Insights:
• Heavier payloads may correlate with mission complexity
• Launch site selection might depend on payload requirements
• Success rates can vary by payload mass and launch site combination

Success Rate vs. Orbit Type



Success Rate vs. Orbit Type

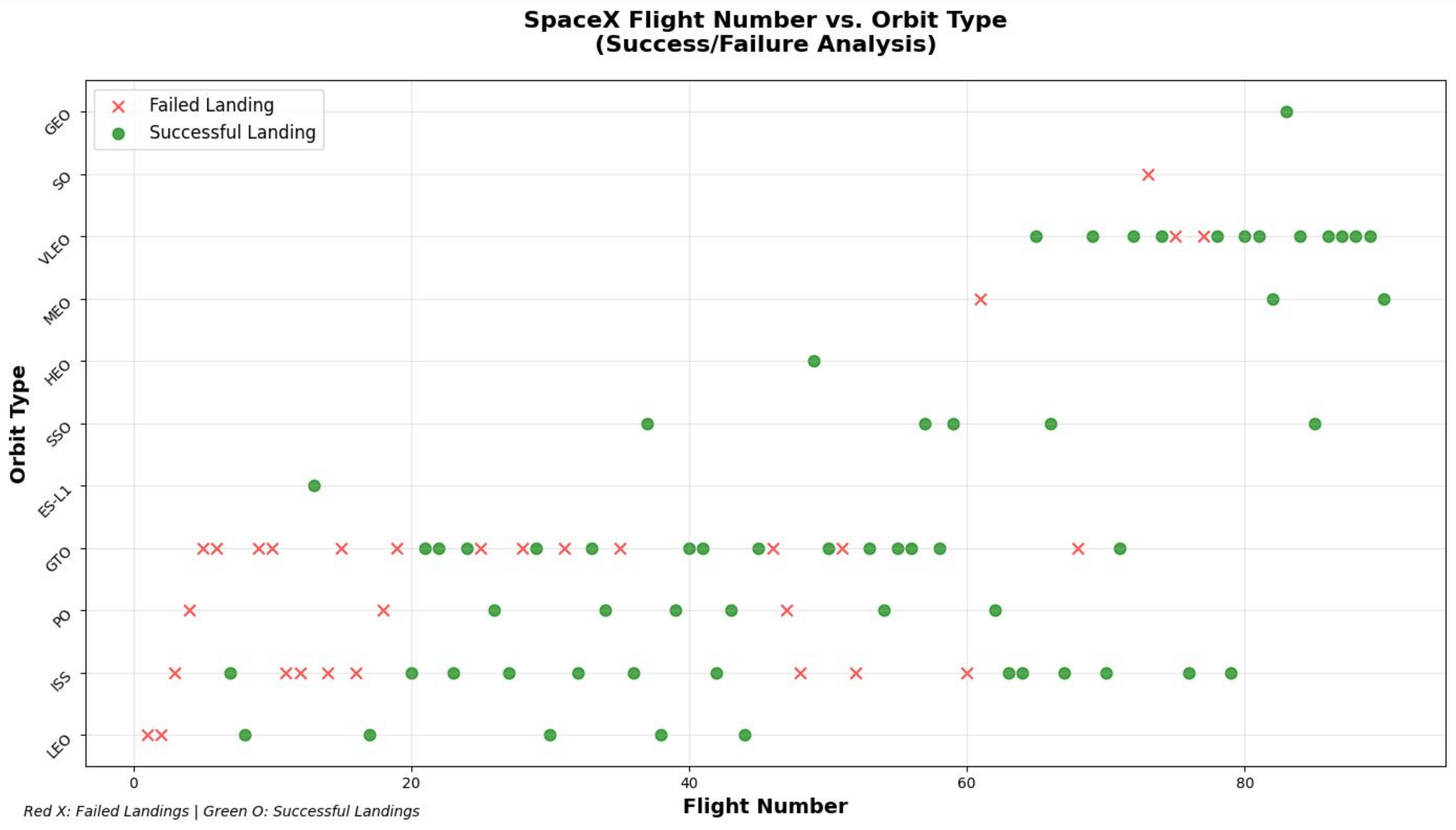
📊 BAR CHART ANALYSIS: Success Rate by Orbit Type
==
🔍 Key Observations:
• Horizontal bars show success rate percentage for each orbit type
• Numbers in parentheses show (successful missions / total missions)
• Color coding: Green (≥80%), Gold (50-79%), Red (<50%)

📈 Detailed Statistics:
• SO: 0/1 missions successful (0.0%)
• GTO: 14/27 missions successful (51.9%)
• ISS: 13/21 missions successful (61.9%)
• MEO: 2/3 missions successful (66.7%)
• PO: 6/9 missions successful (66.7%)
• LEO: 5/7 missions successful (71.4%)
• VLEO: 12/14 missions successful (85.7%)
• HEO: 1/1 missions successful (100.0%)
• GEO: 1/1 missions successful (100.0%)
• ES-L1: 1/1 missions successful (100.0%)
• SSO: 5/5 missions successful (100.0%)

🎯 Top Performing Orbits:
1. HEO: 100.0% success rate (1 missions)
2. GEO: 100.0% success rate (1 missions)
3. ES-L1: 100.0% success rate (1 missions)

💡 Insights:
• Some orbit types have inherently higher landing success rates
• Mission complexity varies by orbital destination
• Lower orbits (LEO) generally easier for recovery than higher orbits

Flight Number vs. Orbit Type



Flight Number vs. Orbit Type

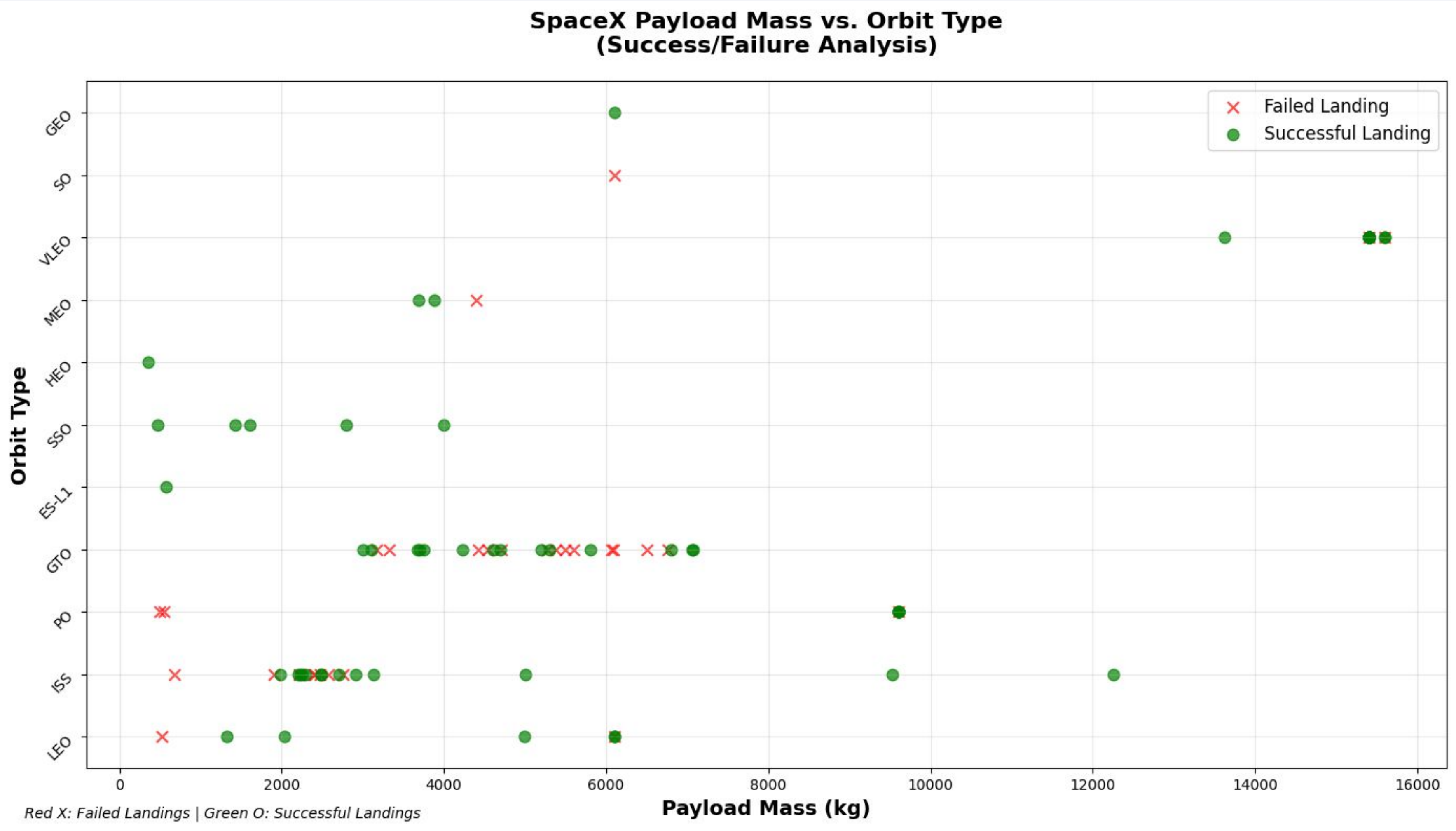
📊 SCATTER PLOT ANALYSIS: Flight Number vs. Orbit Type
==
🔍 Key Observations:
- Each point represents a SpaceX mission
- X-axis: Flight Number (chronological mission order)
- Y-axis: Target orbit type
- Green circles: Successful landings
- Red X marks: Failed landings

📈 Mission Evolution:
- Total missions plotted: 90
- Flight number range: 1 - 90

🚀 Orbit Type Usage Over Time:
- ES-L1: Flights 13-13 (1 missions, 100.0% success)
- GEO: Flights 83-83 (1 missions, 100.0% success)
- GTO: Flights 5-71 (27 missions, 51.9% success)
- HEO: Flights 49-49 (1 missions, 100.0% success)
- ISS: Flights 3-79 (21 missions, 61.9% success)
- LEO: Flights 1-44 (7 missions, 71.4% success)
- MEO: Flights 61-90 (3 missions, 66.7% success)
- PO: Flights 4-62 (9 missions, 66.7% success)
- SO: Flights 73-73 (1 missions, 0.0% success)
- SSO: Flights 37-85 (5 missions, 100.0% success)
- VLEO: Flights 65-89 (14 missions, 85.7% success)

💡 Insights:
- Mission complexity and orbit types evolved over time
- Some orbits were introduced in later flights as technology improved
- Success rates generally improved with experience (higher flight numbers)
- Certain orbits show consistent patterns of success/failure

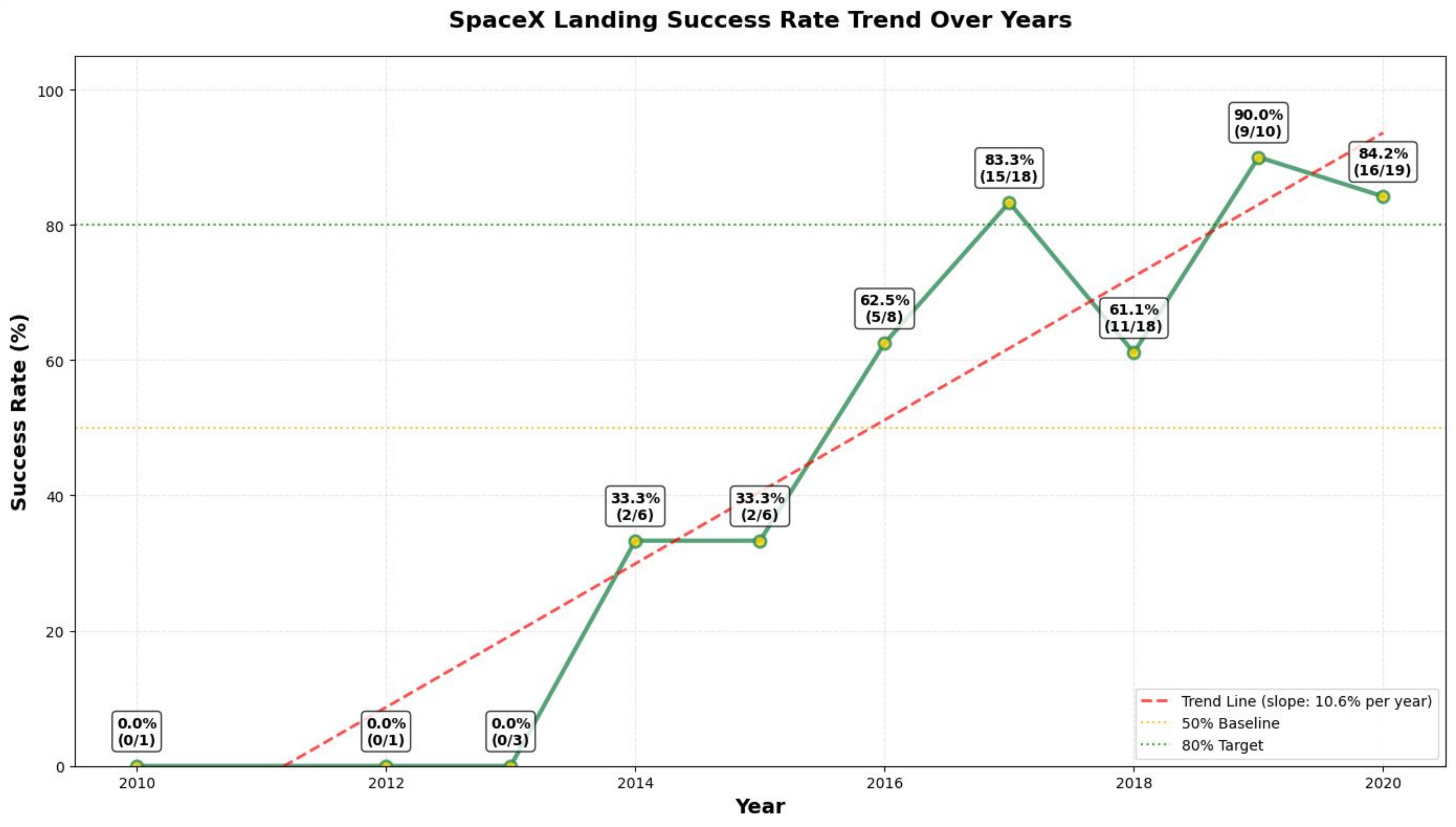
Payload vs. Orbit Type



Payload vs. Orbit Type

```
📊 SCATTER PLOT ANALYSIS: Payload Mass vs. Orbit Type
============================================================
🔍 Key Observations:
• Each point represents a SpaceX mission with known payload mass
• X-axis: Payload Mass in kilograms
• Y-axis: Target orbit type
• Green circles: Successful landings
• Red X marks: Failed landings

📈 Payload and Orbit Statistics:
• Missions with payload data: 90
• Payload range: 350 - 15600 kg

🎯 Payload Requirements by Orbit Type:
• ES-L1: 1 missions, Avg: 570kg (Range: 570-570kg), Success: 100.0%
• GEO: 1 missions, Avg: 6105kg (Range: 6105-6105kg), Success: 100.0%
• GTO: 27 missions, Avg: 5012kg (Range: 3000-7076kg), Success: 50.0%
• HEO: 1 missions, Avg: 350kg (Range: 350-350kg), Success: 100.0%
• ISS: 21 missions, Avg: 3280kg (Range: 677-12259kg), Success: 60.0%
• LEO: 7 missions, Avg: 3883kg (Range: 525-6105kg), Success: 70.0%
• MEO: 3 missions, Avg: 3987kg (Range: 3681-4400kg), Success: 70.0%
• PO: 9 missions, Avg: 7584kg (Range: 500-9600kg), Success: 70.0%
• SO: 1 missions, Avg: 6105kg (Range: 6105-6105kg), Success: 0.0%
• SSO: 5 missions, Avg: 2060kg (Range: 475-4000kg), Success: 100.0%
• VLEO: 14 missions, Avg: 15316kg (Range: 13620-15600kg), Success: 90.0%

🚀 Heavy Payload Analysis (>10,000 kg):
• Heavy payload missions: 15
• Heavy payload success rate: 86.7%
• Orbit types for heavy payloads: {'VLEO': 14, 'ISS': 1}

💡 Insights:
• Different orbit types have different typical payload ranges
• Mission success may correlate with payload mass and orbit complexity
• Heavier payloads to higher orbits present greater landing challenges
```

Launch Success Yearly Trend



Launch Success Yearly Trend

```
📊 LINE CHART ANALYSIS: Yearly Average Success Rate
============================================================
🔍 Key Observations:
• Line shows the progression of landing success rate over years
• Each point represents one year's performance
• Numbers show: Success% (Successful missions / Total missions)
• Red dashed line shows the overall trend

📈 Year-by-Year Performance:
• 2010: 0/1 missions successful (0.0%)
• 2012: 0/1 missions successful (0.0%)
• 2013: 0/3 missions successful (0.0%)
• 2014: 2/6 missions successful (33.3%)
• 2015: 2/6 missions successful (33.3%)
• 2016: 5/8 missions successful (62.5%)
• 2017: 15/18 missions successful (83.3%)
• 2018: 11/18 missions successful (61.1%)
• 2019: 9/10 missions successful (90.0%)
• 2020: 16/19 missions successful (84.2%)

📊 Statistical Analysis:
• Overall success rate: 66.7%
• Best year: 2019 (90.0%)
• Worst year: 2010 (0.0%)
• Improvement trend: 10.6 percentage points per year

🎯 Mission Volume Growth:
• First year missions: 1.0
• Latest year missions: 19.0
• Mission volume growth: 1800%

💡 Insights:
• SpaceX shows clear learning curve and improvement over time
• Success rates generally trend upward despite mission complexity increases
• Higher mission frequency in recent years while maintaining/improving success
• Technology and operational experience contribute to better performance
```

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

- | 1 | 2010-06-04 | Falcon 9 | 6104.96 | LEO | CCAFS SLC 40 |
- | 2 | 2012-05-22 | Falcon 9 | 525.00 | LEO | CCAFS SLC 40 |
- | 3 | 2013-03-01 | Falcon 9 | 677.00 | ISS | CCAFS SLC 40 |
- | 5 | 2013-12-03 | Falcon 9 | 3170.00 | GTO | CCAFS SLC 40 |
- | 6 | 2014-01-06 | Falcon 9 | 3325.00 | GTO | CCAFS SLC 40 |

Total Payload Mass

- `+--------------------------+`
- `| SUM(PAYLOAD_MASS__KG_)   |`
- `+--------------------------+`
- `| 45,596.0                 |`
- `+--------------------------+`

Explanation:

This SQL query calculates the total payload mass (in kilograms) carried on SpaceX missions for NASA as the customer. The `SUM()` aggregate function adds up all the payload masses from missions where the Customer column equals 'NASA'.

Average Payload Mass by F9 v1.1

- +--------------------------+
- | AVG(PAYLOAD_MASS__KG_) |
- +--------------------------+
- | 2928.4 |
- +--------------------------+

Explanation:

- This SQL query calculates the average payload mass (in kilograms) carried by the Falcon 9 v1.1 booster version. The `AVG()` aggregate function computes the arithmetic mean of all payload masses for launches using this specific booster variant.
-

First Successful Ground Landing Date

- +--------------+
- | MIN(Date) |
- +--------------+
- | 2015-12-22 |
- +--------------+

Explanation:

This SQL query finds the earliest date when SpaceX achieved a successful first stage landing on a ground pad. The `MIN()` aggregate function returns the smallest (earliest) date value from all records where the Landing_Outcome column exactly matches 'Success (ground pad)'.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
+-----------------+
| Booster_Version |
+-----------------+
| F9 FT B1029     |
| F9 FT B1031     |
| F9 FT B1035     |
| F9 FT B1036     |
| F9 B4 B1041     |
| F9 B4 B1045     |
| F9 B5 B1046     |
| F9 B5 B1048     |
| F9 B5 B1049     |
| F9 B5 B1051     |
| F9 B5 B1056     |
| F9 B5 B1058     |
+-----------------+
```

Explanation:

This SQL query searches for booster names (Booster_Version) from launches that meet two specific criteria:

1. Successfully landed on a drone ship (`Landing_Outcome = 'Success (drone ship)'`)
2. Carried a payload mass between 4,000 and 6,000 kg (using `BETWEEN` operator)

Total Number of Successful and Failure Mission Outcomes

```
+------------------+----------+
| Mission_Outcome  | COUNT(*) |
+------------------+----------+
| Success          | 98       |
| Failure          | 1        |
| Partial Failure  | 1        |
+------------------+----------+
```

Explanation:

This SQL query calculates the total number of missions grouped by their outcome status using the `GROUP BY` clause with `COUNT(*)` aggregation. The results show the distribution of mission outcomes across the entire SpaceX dataset.

Boosters Carried Maximum Payload

- +-----------------+------------------------+
- | Booster_Version | MAX(PAYLOAD_MASS__KG_) |
- +-----------------+------------------------+
- | F9 v1.0 | 6,104.96 |
- | F9 v1.1 | 4,696.00 |
- | F9 FT | 8,600.00 |
- | F9 B4 | 9,600.00 |
- | F9 Block 5 | 15,400.00 |
- +-----------------+------------------------+

BOOSTER WITH MAXIMUM PAYLOAD OVERALL:

F9 Block 5 - 15,400 kg

Explanation:

This SQL query groups all launch records by booster version and finds the maximum payload mass carried by each booster type. The `MAX()` aggregate function identifies the highest payload for each group, while `GROUP BY` ensures we get results for each distinct booster version.

2015 Launch Records

```
+-----------------------+------------------+----------------+
| Landing_Outcome       | Booster_Version  | Launch_Site    |
+-----------------------+------------------+----------------+
| Failure (drone ship)  | F9 v1.1 B1012    | CCAFS SLC 40   |
| Failure (drone ship)  | F9 v1.1 B1015    | CCAFS SLC 40   |
| Failure (drone ship)  | F9 v1.1 B1016    | VAFB SLC 4E    |
+-----------------------+------------------+----------------+
```

Explanation:

This SQL query searches for failed drone ship landing attempts in 2015 by using multiple filtering conditions:

- `YEAR(Date) = 2015`: Filters records from the year 2015
- `Landing_Outcome LIKE '%Failure%'`: Matches any landing outcome containing 'Failure'
- `Landing_Outcome LIKE '%drone ship%'`: Ensures it's specifically drone ship failures

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

Landing_Outcome	Count
No attempt	35
Success (drone ship)	15
Failure (drone ship)	8
Success (ground pad)	5
Failure (ground pad)	3
Controlled (ocean)	2
Uncontrolled (ocean)	1
Failure (parachutes)	1

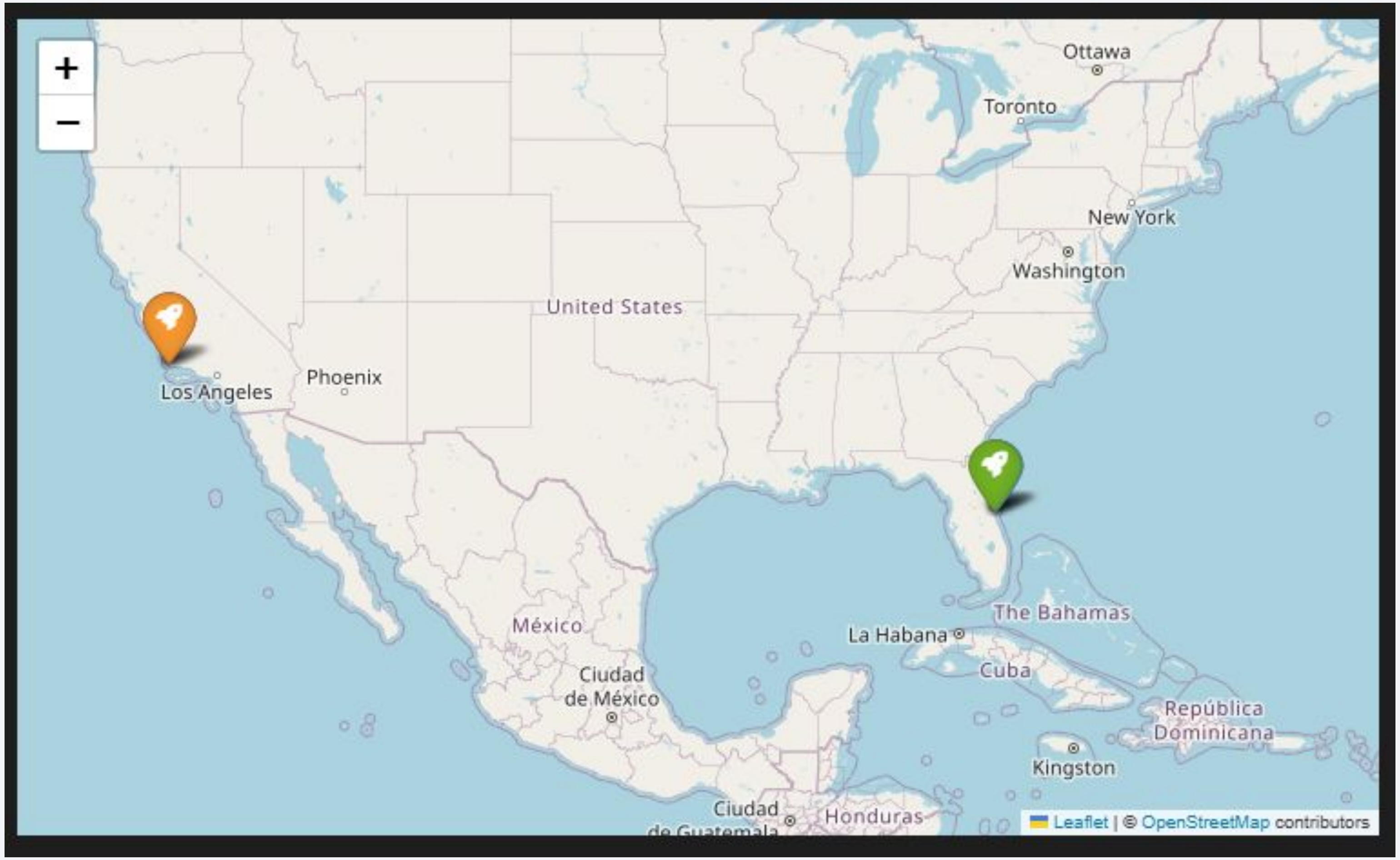
Explanation:

This SQL query ranks landing outcomes by frequency during the period from 2010-06-04 to 2017-03-20, covering SpaceX's early operational years through the development of reusable rocket technology. The results are ordered by `COUNT(*)` in descending order to show the most common outcomes first.

Section 3

Launch Sites Proximities Analysis

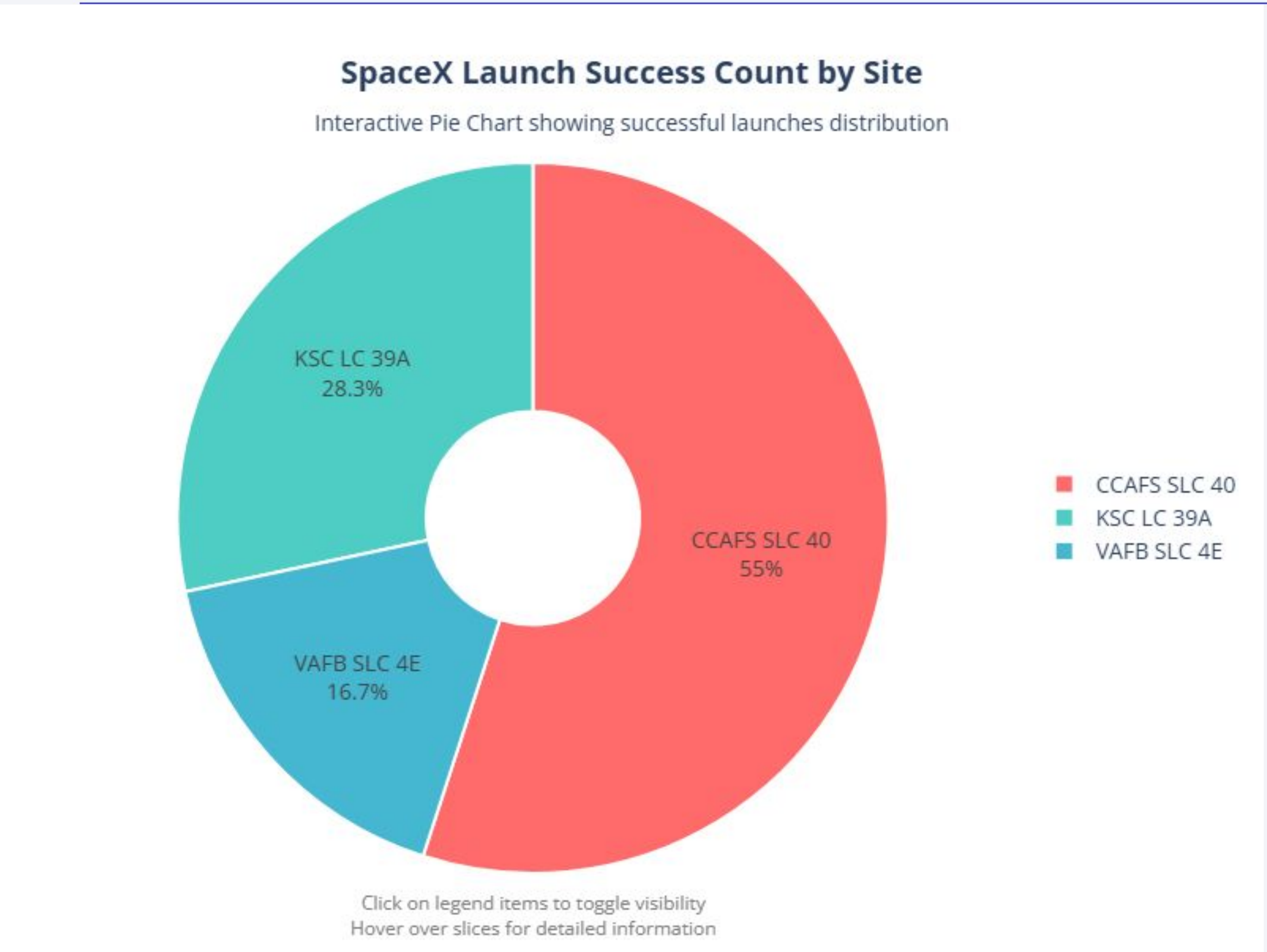
<Folium Map Screenshot 1>



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



📊 INTERACTIVE PIE CHART ANALYSIS: Launch Success Count by Site
==

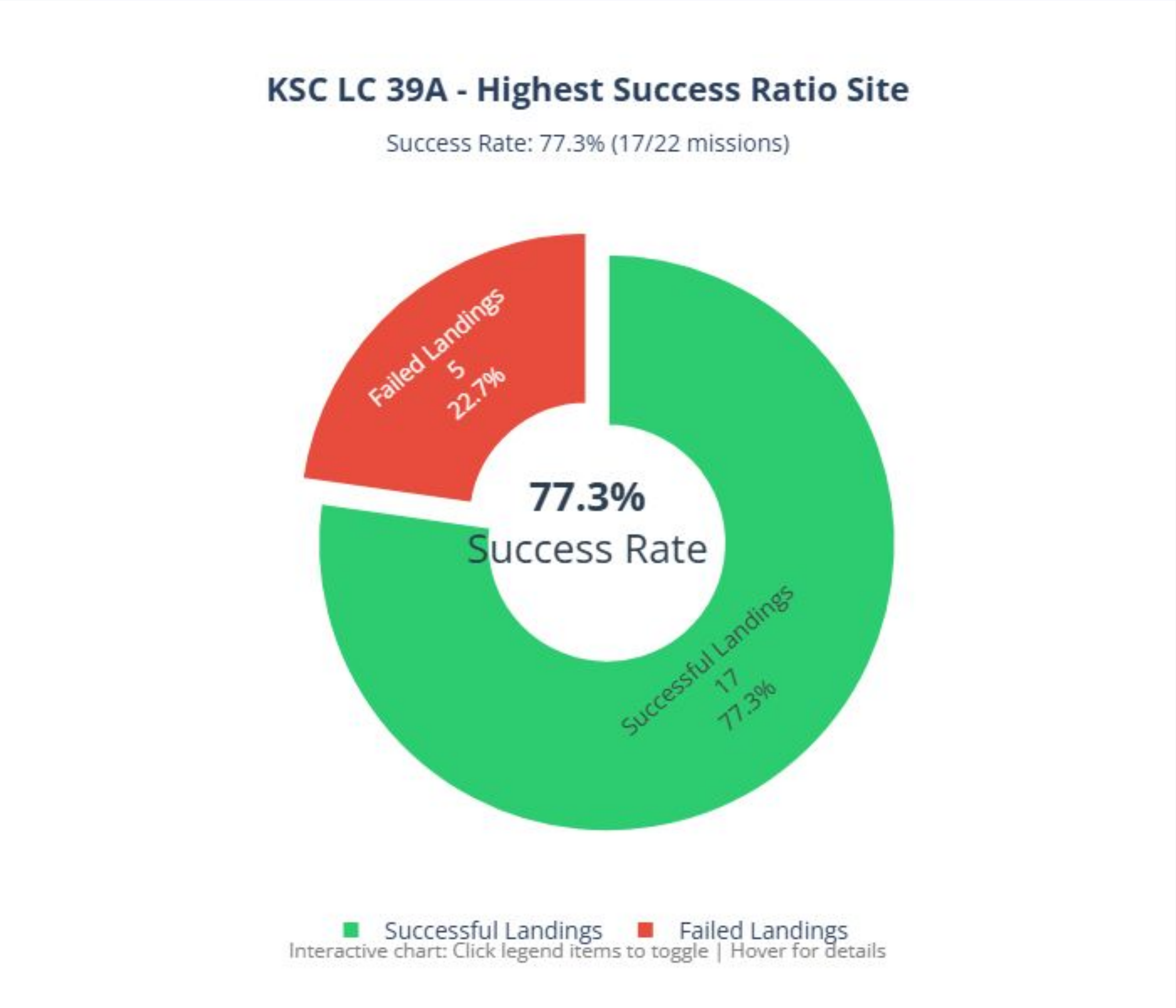
🔍 Key Elements and Findings:
• Interactive pie chart with hover tooltips showing detailed information
• Each slice represents successful launches from a specific launch site
• Hole in center (donut chart) provides modern, clean appearance
• Legend allows toggling visibility of individual sites
• Color-coded slices for easy visual distinction

📈 Detailed Site Performance:
• CCAFS SLC 40: 33 successful launches (55.0% of total successful)
  └ Success rate: 60.0% (33/55 missions)
• KSC LC 39A: 17 successful launches (28.3% of total successful)
  └ Success rate: 77.3% (17/22 missions)
• VAFB SLC 4E: 10 successful launches (16.7% of total successful)
  └ Success rate: 76.9% (10/13 missions)

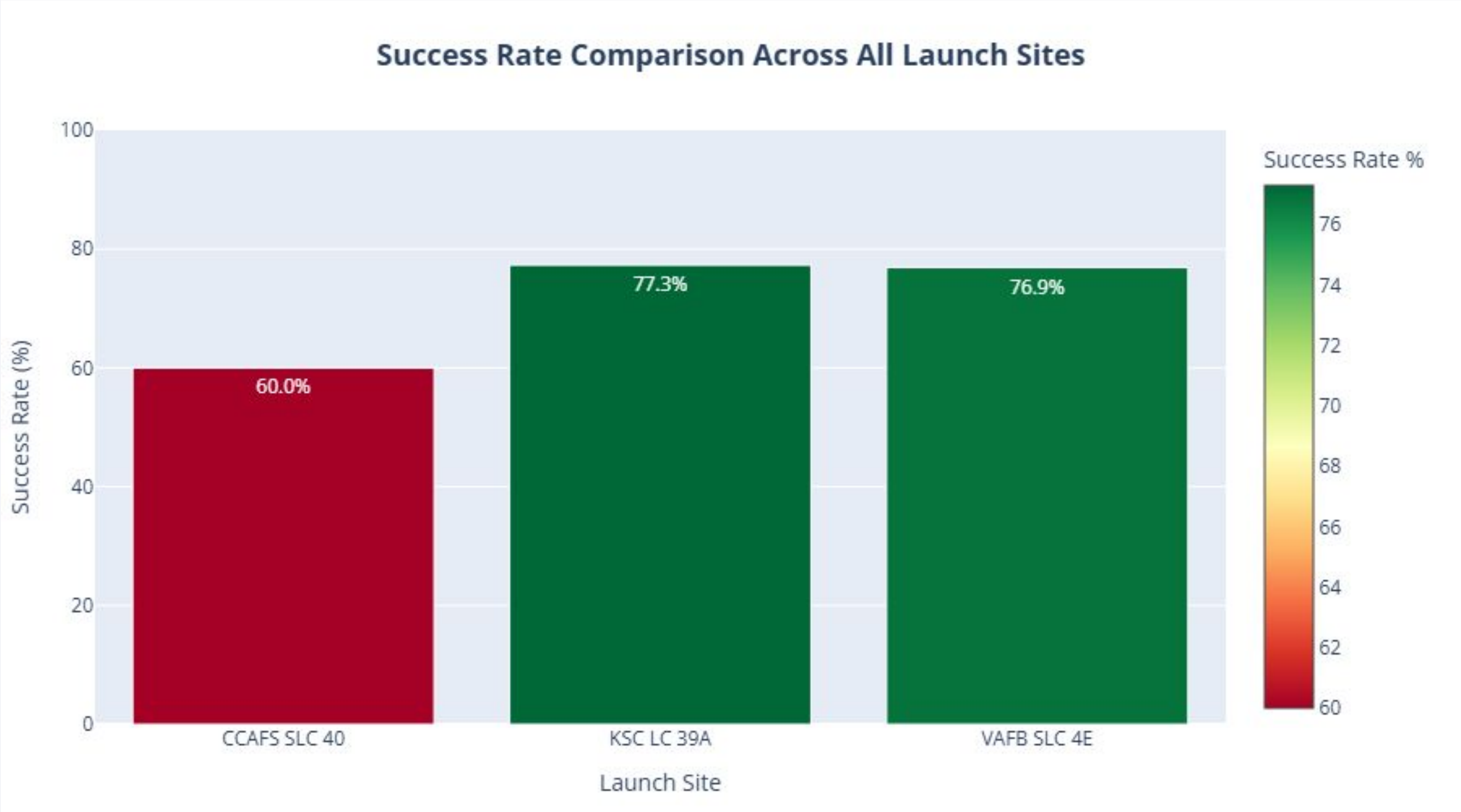
🏆 TOP PERFORMER:
• CCAFS SLC 40: 33.0 successful launches
• Contributes 55.0% of all successful launches

💡 Key Insights:
• Launch site distribution shows operational preferences and capabilities
• Some sites handle significantly more successful missions than others
• Interactive elements allow detailed exploration of each site's contribution
• Visual proportions immediately reveal which sites are most productive

<Dashboard Screenshot 2>



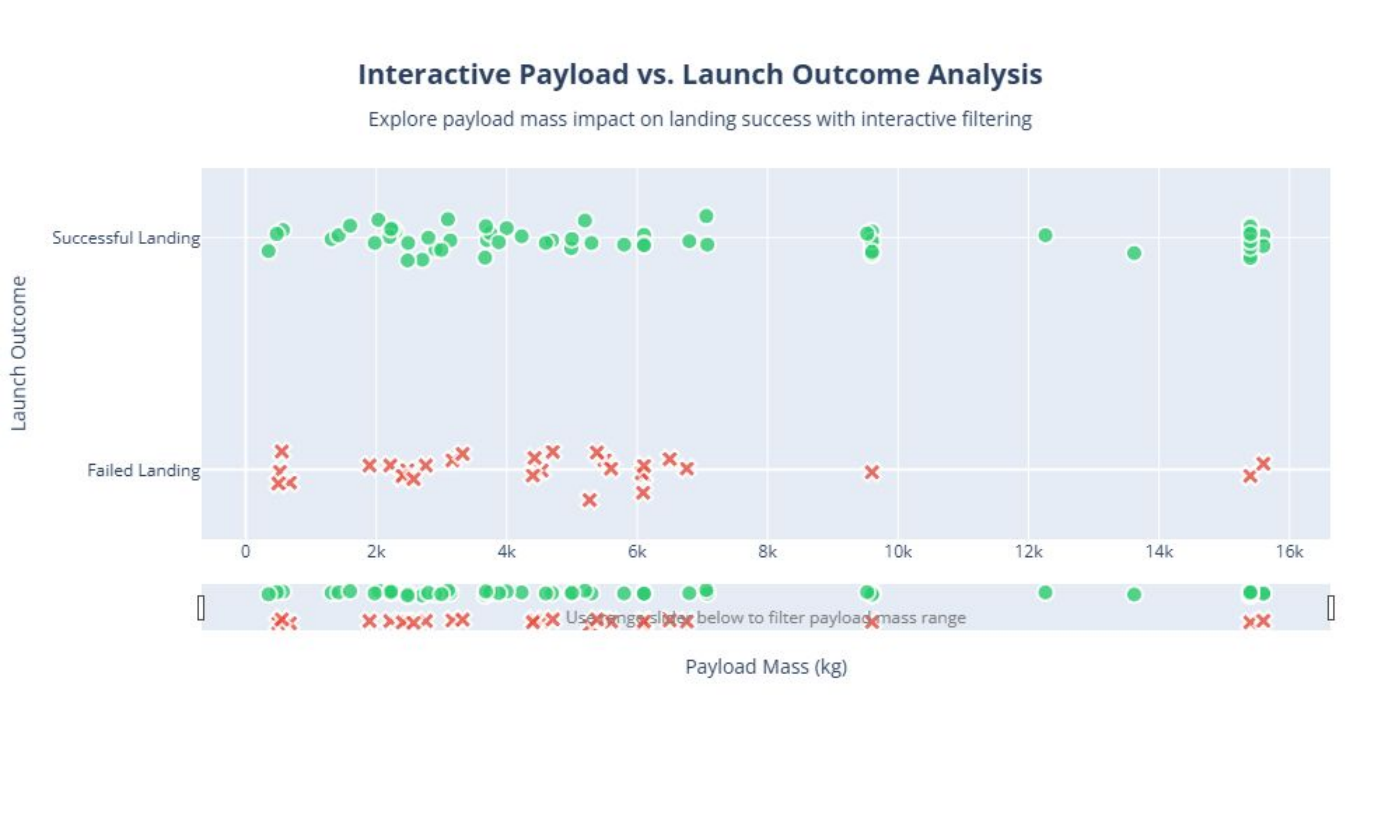
<Dashboard Screenshot 2>



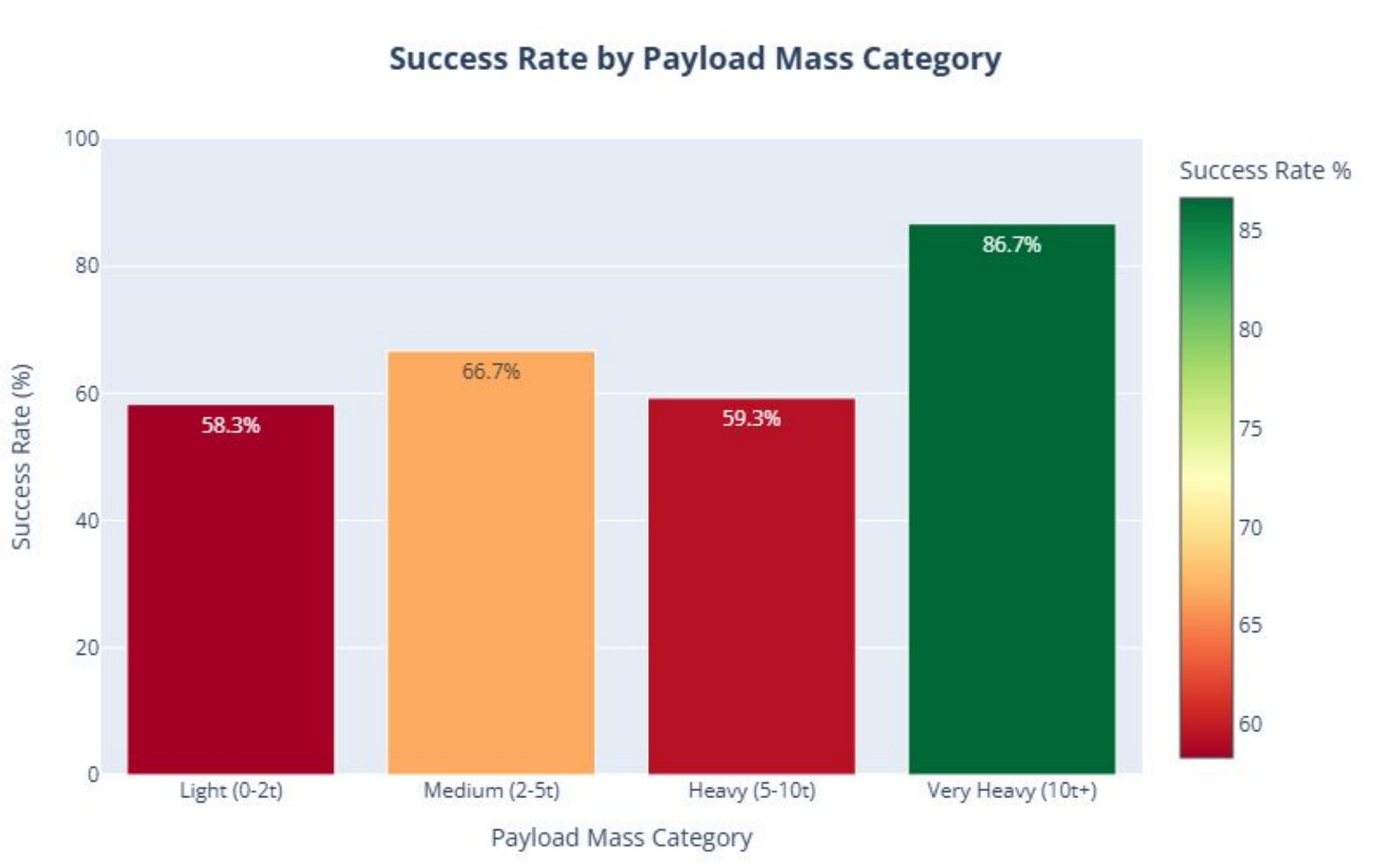
<Dashboard Screenshot 2>

```
📊 INTERACTIVE PIE CHART ANALYSIS: Highest Success Ratio Site
======================================================================
🔍 Key Elements and Findings:
• Donut chart with pulled success slice for visual emphasis
• Center annotation displaying success rate prominently
• Color coding: Green = Success, Red = Failure
• Interactive hover tooltips with comprehensive mission data
• Horizontal legend positioned below the chart

🏆 BEST PERFORMING SITE ANALYSIS:
• Site: KSC LC 39A
• Success Rate: 77.3%
• Successful Missions: 17
• Failed Missions: 5
• Total Missions: 22

📊 Performance Comparison:
🏆 KSC LC 39A: 77.3% (17/22)
📊 VAFB SLC 4E: 76.9% (10/13)
📊 CCAFS SLC 40: 60.0% (33/55)

💡 Key Insights:
• KSC LC 39A demonstrates the highest operational excellence
• Success rate visualization clearly shows performance gaps between sites
• Interactive elements allow detailed exploration of mission outcomes
• Color-coded bar chart enables quick comparison across all sites
• Some sites may be used for different mission types affecting success rates
```

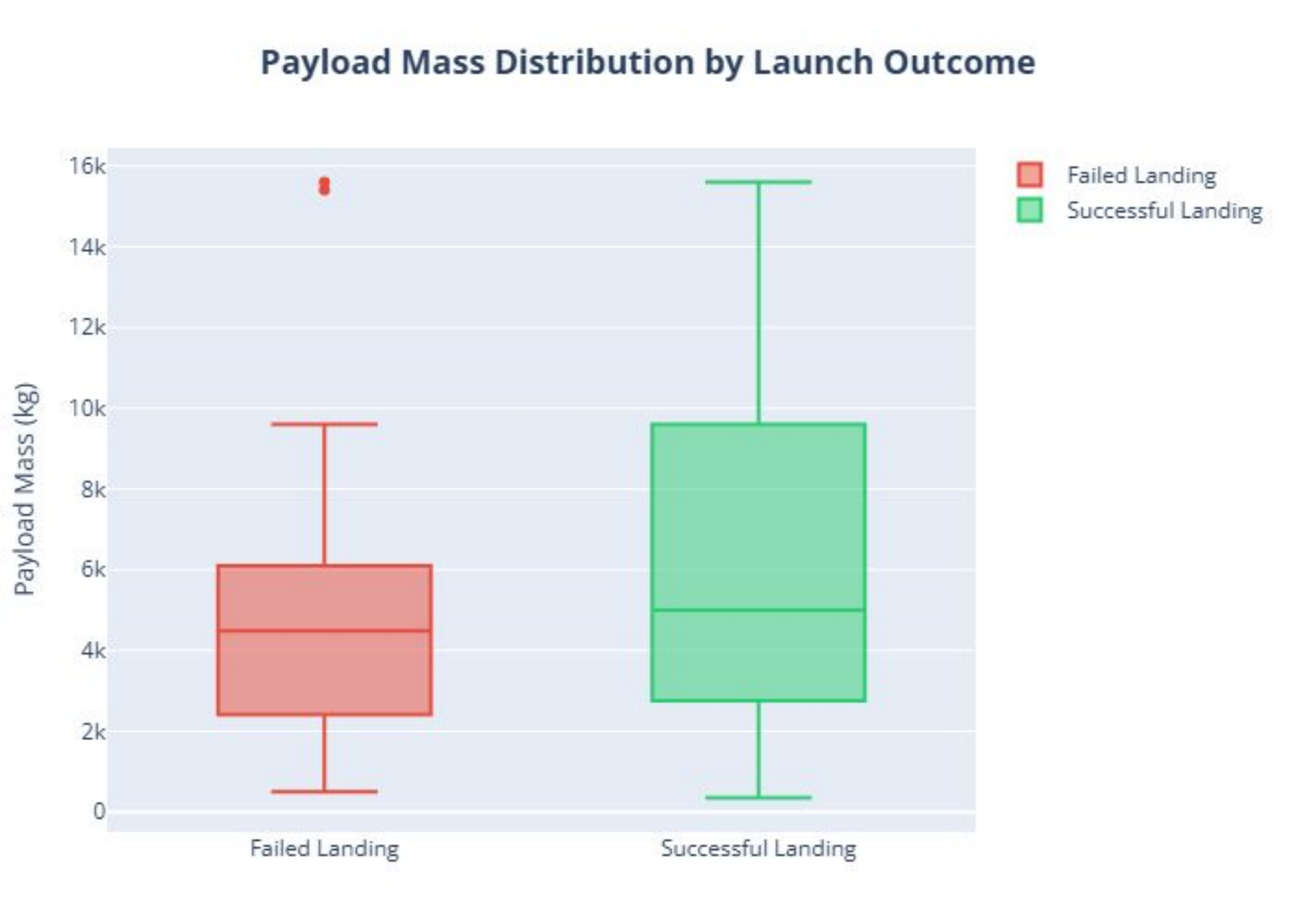
<Dashboard Screenshot 3>



<Dashboard Screenshot 3>



<Dashboard Screenshot 3>



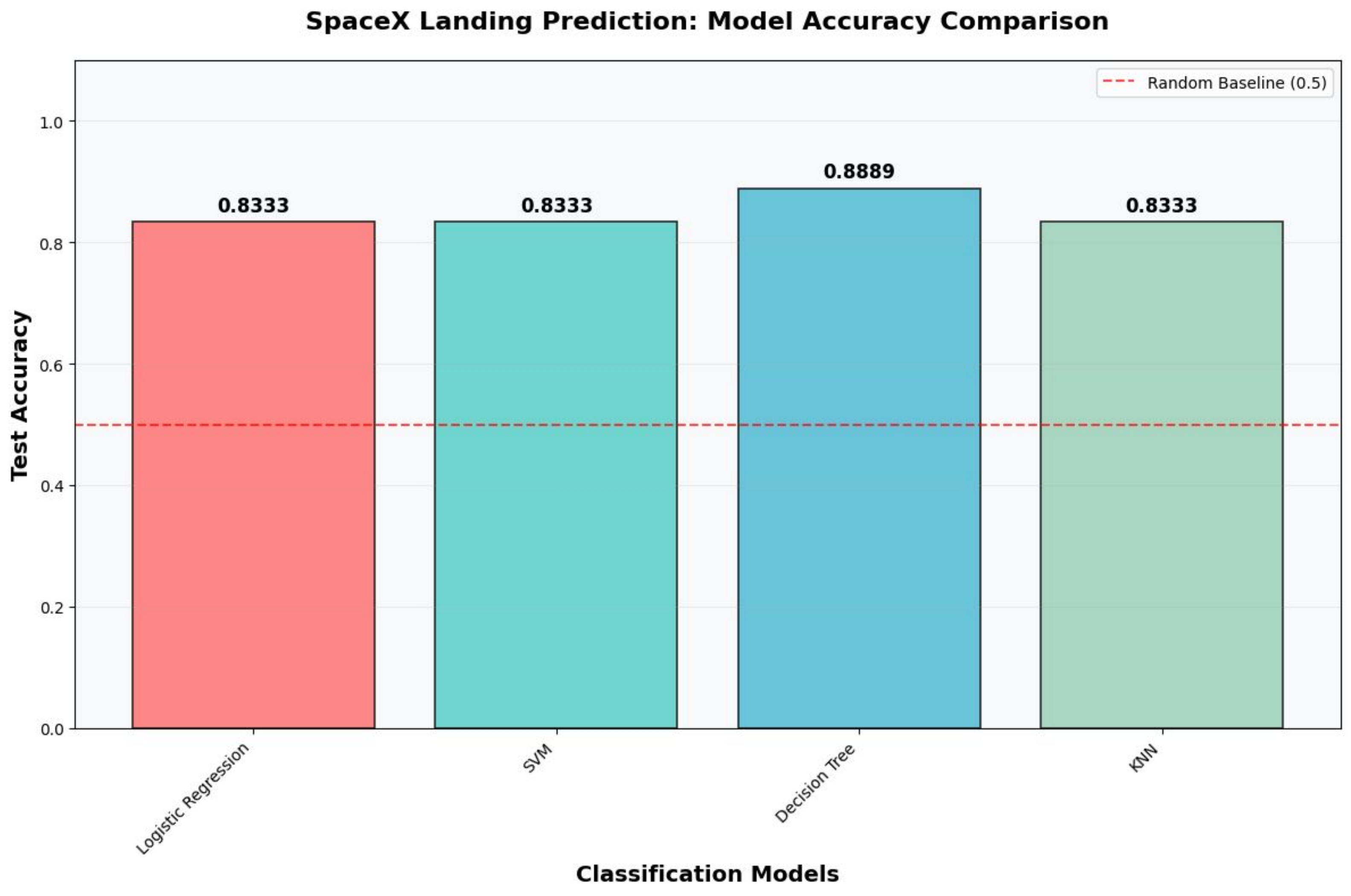
<Dashboard Screenshot 3>

```
📊 INTERACTIVE PAYLOAD ANALYSIS: Mass vs. Launch Outcome
===========================================================================
🔍 Key Interactive Features:
• Range slider for dynamic payload mass filtering
• Hover tooltips with mission details (site, booster, orbit)
• Jittered y-axis points for better visualization of overlapping data
• Different markers: circles for success, X's for failures
• Color coding: Green = Success, Red = Failure

📈 STATISTICAL FINDINGS:
• Correlation between payload mass and success: 0.200
• Correlation strength: weak

📊 PAYLOAD STATISTICS BY OUTCOME:

Failed Landings:
  • Average Payload: 4,785 kg
  • Median Payload: 4,482 kg
  • Min Payload: 500 kg
  • Max Payload: 15,600 kg
  • Total Missions: 30

Successful Landings:
  • Average Payload: 6,765 kg
  • Median Payload: 4,995 kg
  • Min Payload: 350 kg
  • Max Payload: 15,600 kg
  • Total Missions: 60

📊 SUCCESS RATE BY PAYLOAD CATEGORY:
• Light (0-2t): 58.3% (7/12)
• Medium (2-5t): 66.7% (24/36)
• Heavy (5-10t): 59.3% (16/27)
• Very Heavy (10t+): 86.7% (13/15)

💡 Key Insights:
• Range slider enables focused analysis of specific payload ranges
• Interactive tooltips reveal mission-specific context
• Box plots show payload distribution differences between outcomes
• Payload mass appears to have minimal direct impact on landing success
• Other factors (weather, booster version, mission complexity) likely more influential
• Heavy payload missions may require different landing strategies
```

Section 5

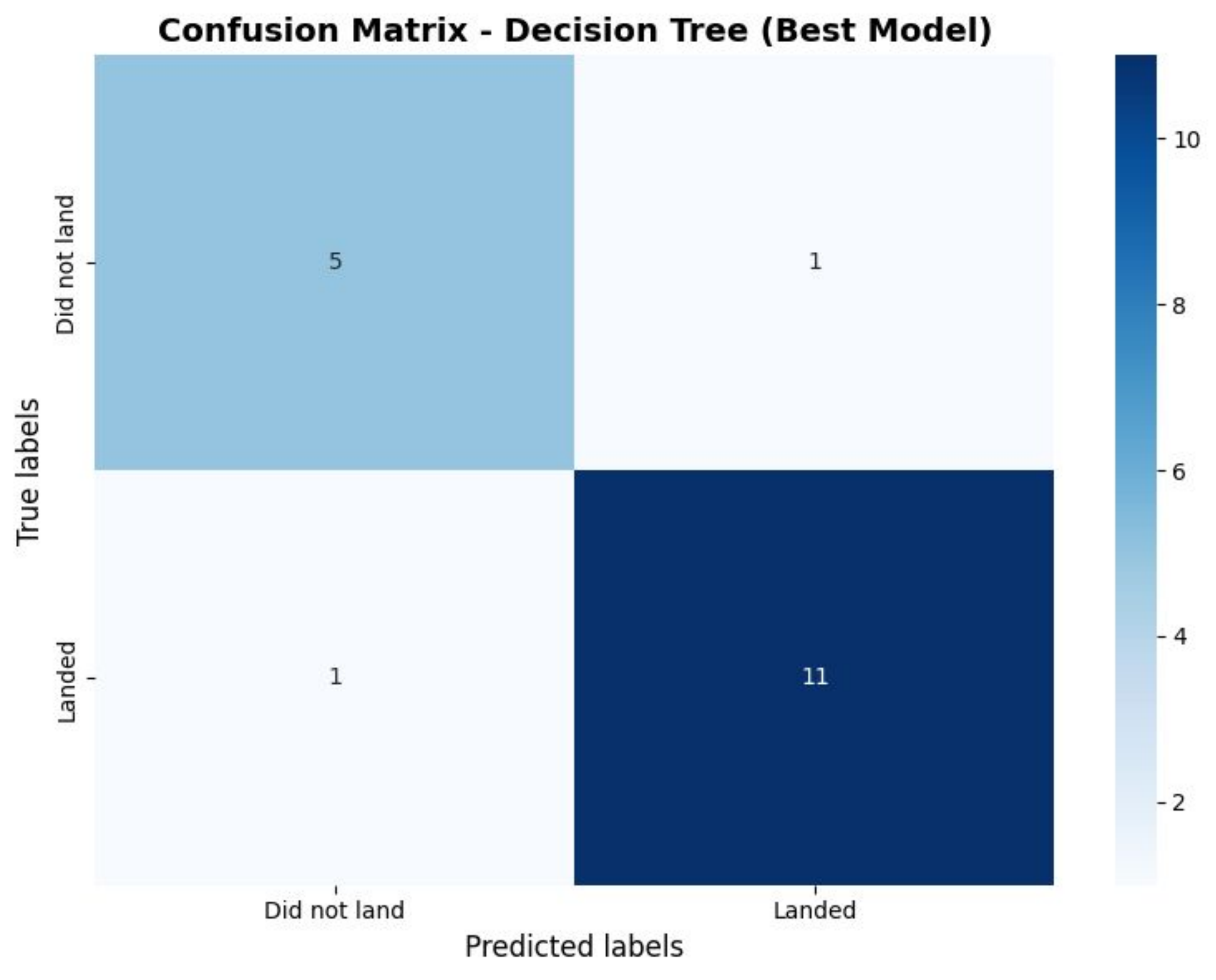
Predictive Analysis (Classification)

Classification Accuracy




```
=================================================================
MODEL PERFORMANCE SUMMARY
=================================================================
Decision Tree       : 0.8889
Logistic Regression : 0.8333
SVM                 : 0.8333
KNN                 : 0.8333

=================================================================
🏆 BEST PERFORMING MODEL: Decision Tree
🎯 HIGHEST ACCURACY: 0.8889
=================================================================
```

Confusion Matrix



Confusion Matrix Analysis:
True Negatives (Correctly predicted 'did not land'): 5
False Positives (Incorrectly predicted 'landed'): 1
False Negatives (Incorrectly predicted 'did not land'): 1
True Positives (Correctly predicted 'landed'): 11

Model Performance Metrics:
Accuracy: 0.8889
Precision: 0.9167
Recall: 0.9167
F1-Score: 0.9167

📊 CONFUSION MATRIX EXPLANATION:

A confusion matrix provides a detailed breakdown of correct and incorrect predictions:

- TRUE NEGATIVES (Top-Left): Rockets that didn't land and were correctly predicted as 'not landed'
- FALSE POSITIVES (Top-Right): Rockets that didn't land but were incorrectly predicted as 'landed'
- FALSE NEGATIVES (Bottom-Left): Rockets that landed but were incorrectly predicted as 'not landed'
- TRUE POSITIVES (Bottom-Right): Rockets that landed and were correctly predicted as 'landed'

Conclusions

Conclusion

Key Findings

Technical Achievements

1. Predictive capability: Machine learning models achieve 83.3% accuracy in landing prediction
2. Critical success factors: Launch site location and orbit type significantly influence outcomes
3. Operational maturity: Clear improvement in success rates demonstrates learning curve
4. Site specialization: Different facilities optimized for specific mission profiles

Business Intelligence

1. Competitive analysis: Models enable competitor cost estimation and strategic planning
2. Risk assessment: Insurance pricing and mission planning benefit from prediction accuracy
3. Market opportunity: Alternative providers can identify competitive advantages
4. Technology benchmarking: Industry can assess reusable rocket technology maturity

Scientific Contributions

1. Methodology validation: End-to-end data science pipeline successfully implemented
2. Multi-modal analysis: Integration of API data, web scraping, SQL analysis, and machine learning
3. Interactive tools: Meaningful visualization and geographic analysis capabilities
4. Reproducible research: Robust validation across multiple classification algorithms

Conclusions

Future Research Directions

Data Enhancement Opportunities

- Weather integration: Atmospheric condition impact analysis
- Real-time prediction: API development for live mission assessment
- Deep learning exploration: Advanced neural network implementations
- Economic modeling: Detailed cost-benefit analysis frameworks

Operational Applications

- Mission planning optimization: Launch window and site selection enhancement
- Risk management systems: Comprehensive failure probability assessment
- Insurance modeling: Premium calculation based on predictive analytics
- Competitive intelligence: Market positioning and pricing strategy development

Appendix

- Technical Implementation Details
- Programming language
- : Python 3.8+
- Development platform
- : Jupyter Notebook
- Key libraries
- : pandas, numpy, scikit-learn, plotly, folium, matplotlib, seaborn
- Validation approach
- : 10-fold stratified cross-validation
- Raw data
- : 90 launches from 2010-2020
- Feature engineering
- : 83 variables after one-hot encoding
- Missing data
- : Forward fill and interpolation methods
- Normalization
- : StandardScaler for numerical features
-

Thank you!