

# Winning Space Race with Data Science

Harsh Karekar  
July 29, 2025



# Outline

---

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

# Executive Summary

---

- **Objective:** Develop predictive models to determine first stage landing success rates using machine learning classification techniques.
- **Research Process:**
  - a. Information gathering and data preprocessing
  - b. In-depth analytical exploration
  - c. Interactive visualization development
  - d. Predictive algorithm implementation
- **Research Outcomes:** Visual analysis reveals significant correlations between launch parameters and mission success rates.
- **Recommended Model:** Decision tree classification demonstrates superior performance for Falcon 9 first stage landing predictions.

# Introduction

---

- **Project Background:** This capstone research focuses on developing predictive models to determine the successful recovery of SpaceX Falcon 9 first stage boosters. By leveraging machine learning classification algorithms, we aim to forecast landing outcomes based on various mission parameters and characteristics.
- **Problems We Solve:**
  1. How do variables such as payload mass, launch site, and orbit affect landing success?
  2. Does the success rate improve over time?
  3. What is the optimal algorithm for this binary classification problem?
- **Strategic Value:** This research analyzes public data to predict SpaceX's reusability patterns, enabling cost estimation and competitive analysis in the commercial space industry.

Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Using SpaceX Rest API
  - Using Web Scrapping from Wikipedia
- Perform data wrangling
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning and evaluation of classification models to ensure the best results

# Data Collection

---

Our data acquisition strategy employed a dual-source approach combining SpaceX REST API calls and Wikipedia web scraping to ensure comprehensive launch information for thorough analysis.

This hybrid collection method was essential to obtain complete datasets, as neither single source provided all required variables for detailed predictive modeling.

## **Data Sources & Variables:**

### **1. SpaceX REST API Data Fields:**

- FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite
- Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block
- ReusedCount, Serial, Longitude, Latitude

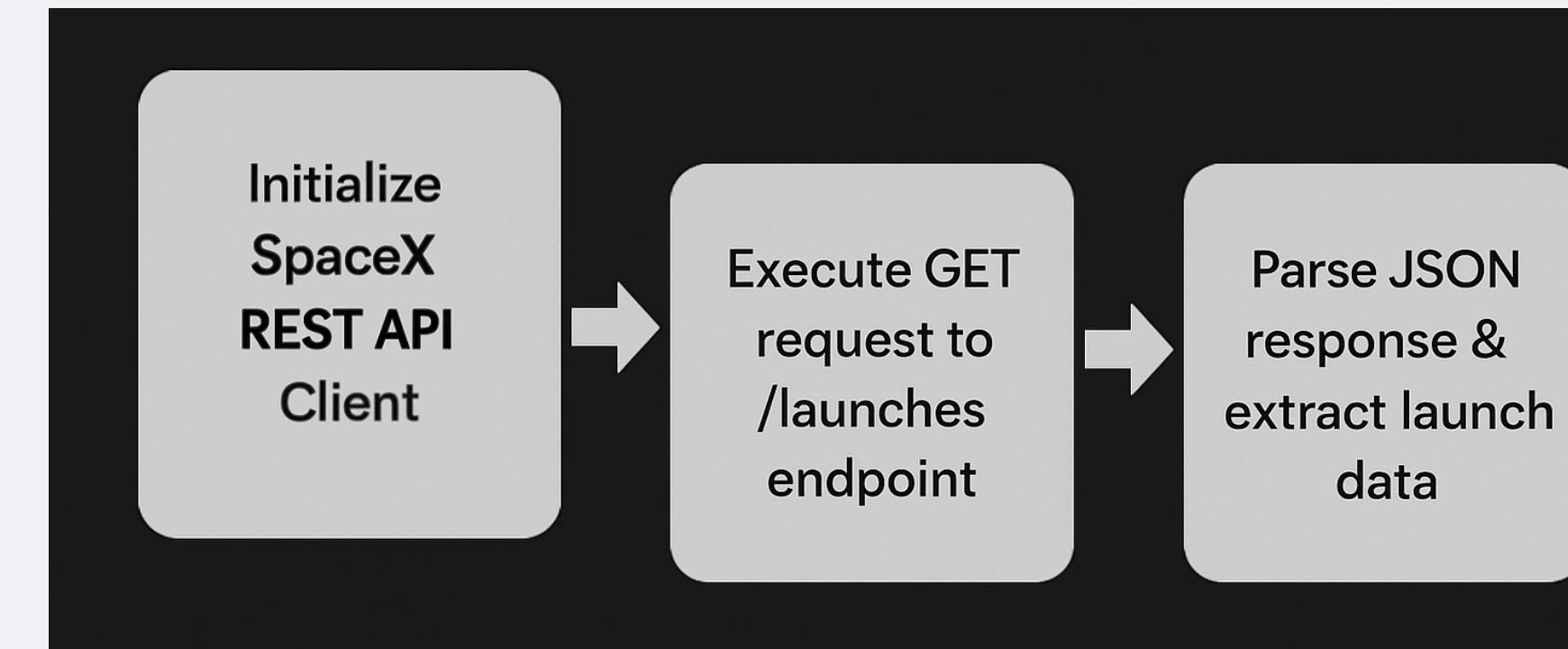
### **2. Wikipedia Web Scraping Data Fields:**

- Flight No., Launch site, Payload, PayloadMass, Orbit, Customer
- Launch outcome, Version Booster, Booster landing, Date, Time

# Data Collection – SpaceX API

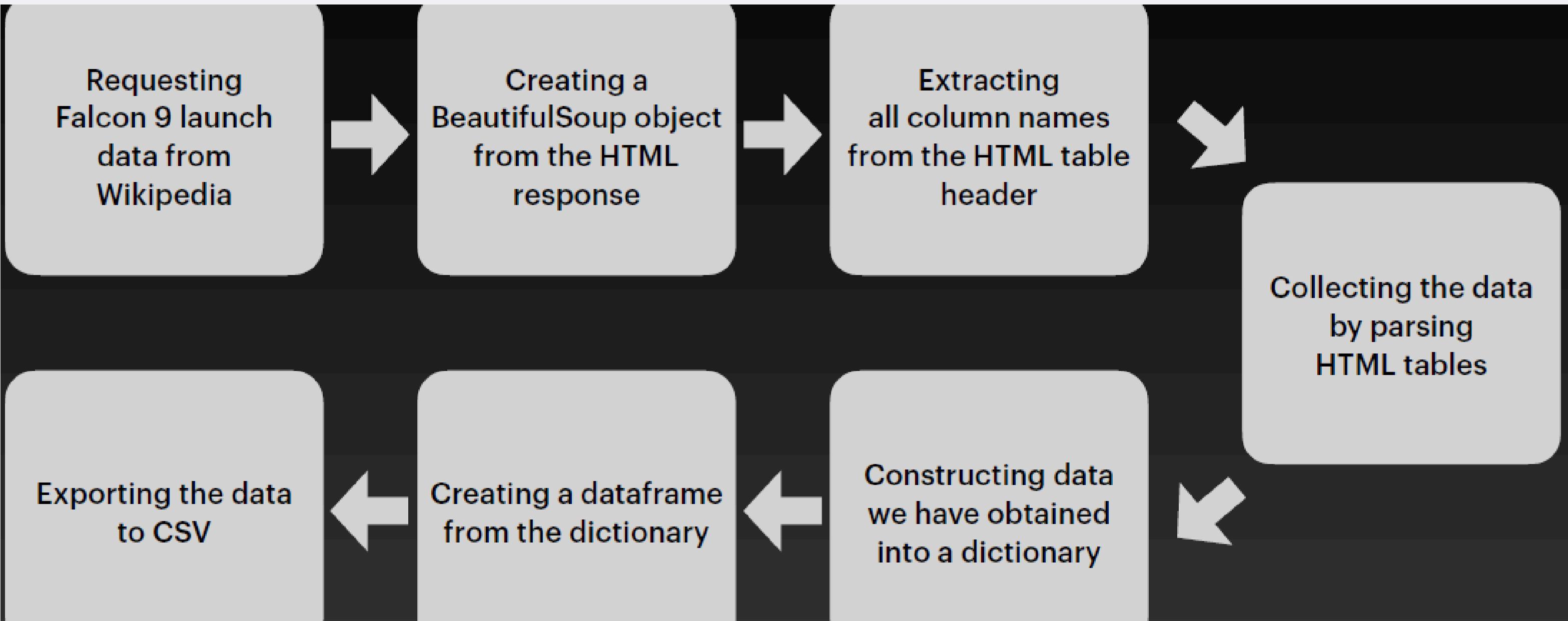
## Key API Data Collection Phrases:

- Endpoint Access: GET  
<https://api.spacexdata.com/v4/launches>
- Authentication: Public API - no authentication required
- Response Format: JSON structured data objects
- Data Extraction: Parse launch records for mission parameters
- Field Mapping: Convert API response to analysis-ready variables



[Github link](#)

# Data Collection - Scraping



[Github link](#)

# Data Wrangling

Dataset contains multiple SpaceX booster landing scenarios:

Landing Types & Success Indicators

Ocean Landing: Targeted water recovery

- True Ocean = Successful ocean landing
- False Ocean = Failed ocean landing

RTLS (Return to Launch Site): Ground pad recovery

- True RTLS = Successful ground pad landing
- False RTLS = Failed ground pad landing

ASDS (Autonomous Spaceport Drone Ship): Sea platform recovery

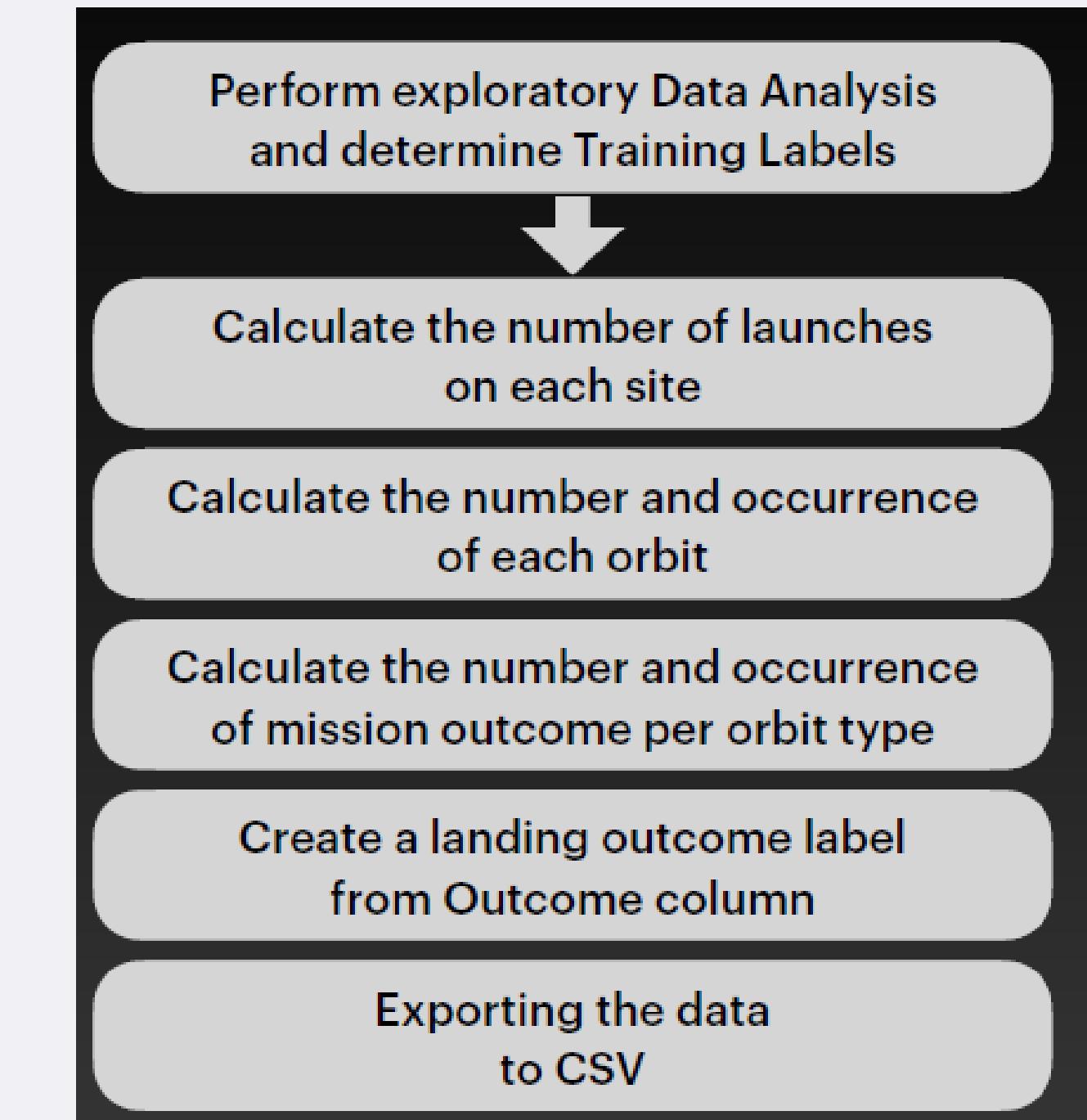
- True ASDS = Successful drone ship landing
- False ASDS = Failed drone ship landing

Binary classification system:

- "1" = Successful landing (any type)
- "0" = Unsuccessful landing (any type)

*This standardization enables machine learning model training for landing success prediction*

[Github Link](#)



# EDA with Data Visualization

---

- The file presents an exploratory data analysis of SpaceX launch data using three main charts. A category plot was used to display the relationship between Flight Number and Payload Mass, with color indicating the outcome of each launch. This visualization helped illustrate how booster landing success rates changed with increasing flight numbers and with varying payload sizes.
- Next, a scatter plot of Flight Number versus Launch Site, also colored by outcome, was employed to compare the distribution and trends of successful and failed landings at different launch locations across time.
- Lastly, a line plot was created to show the trend in launch success rates by year, which provided an overarching view on how reliability has evolved annually. Together, these charts were selected to comprehensively explore how booster recovery outcomes are influenced by mission characteristics, launch site, and temporal progression within the SpaceX launch program.

# EDA with SQL

---

Identified and displayed all distinct launch sites available in the dataset.

- Viewed launches with detailed fields
  - Selected and displayed key columns (e.g., date, booster version, launch site, payload, mission outcome) for initial inspection of mission records.
- Calculated total payload mass delivered for NASA (CRS) missions
  - Filtered launches for NASA Commercial Resupply Services (CRS) missions and summed total payload mass sent.
- Computed average payload mass by booster version
  - Aggregated launches by booster version (e.g., F9 v1.1) and calculated the average payload mass for each.
- Found the date of first successful ground pad landing
  - Queried for the earliest date with a landing outcome classified as successful on a ground pad.
- Analyzed successful and failed landings by booster version
  - Grouped launches by booster version to count and compare the number of successful and unsuccessful landing outcomes.
- Retrieved launches by date, booster version, and launch site for 2015
  - Listed selected details (date, booster, site) for launches carried out in the year 2015.
- Summarized landing outcome counts
  - Counted and displayed the number of missions for each landing outcome type (e.g., no attempt, success, failure), providing a breakdown of how often each scenario occurred.

# Build an Interactive Map with Folium

---

## Map Objects Created:

- Color-coded markers - Added individual markers at each SpaceX launch site using latitude/longitude coordinates with green markers for successful landings (class = 1) and red markers for failed landings (class = 0)<sup>1</sup>
- Launch site location markers - Created markers for four main SpaceX launch sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E to serve as geographic reference points<sup>1</sup>
- Binary color system - Implemented green/red color scheme for immediate visual identification of landing success patterns across different launch locations<sup>1</sup>

## Purpose & Analysis:

- The interactive geospatial visualization revealed location-based patterns in SpaceX booster landing success rates, enabling analysis of whether certain launch sites had higher success rates and identifying geographic factors influencing mission outcomes<sup>1</sup>. This approach allowed for quick visual identification of success clusters and failure patterns at specific launch facilities.

[Github Link](#)

# Build a Dashboard with Plotly Dash

---

## Plots/Graphs Added

- Pie Chart - Shows total successful launches by site (all sites view) or success vs. failure breakdown (specific site view)<sup>1</sup>
- Scatter Plot - Displays correlation between payload mass and launch success with color-coded booster versions and interactive sizing<sup>1</sup>

## Interactive Components

- Launch Site Dropdown - Allows filtering between "All Sites" view or specific launch sites (CCAFS, KSC, VAFB)<sup>1</sup>
- Payload Range Slider - Enables filtering data by payload mass range (0-10,000 kg) with 1,000 kg increments<sup>1</sup>

## Purpose & Rationale

- Pie Chart Purpose - Provides quick visual comparison of success rates across launch sites and identifies which sites perform better for mission planning<sup>1</sup>
- Scatter Plot Purpose - Reveals payload-success relationships to determine optimal payload ranges and identifies booster version performance patterns<sup>1</sup>
- Dropdown Interaction - Enables site-specific analysis for targeted performance evaluation and comparison studies<sup>1</sup>
- Slider Interaction - Allows payload-based filtering to analyze success rates within specific weight categories for mission optimization<sup>1</sup>

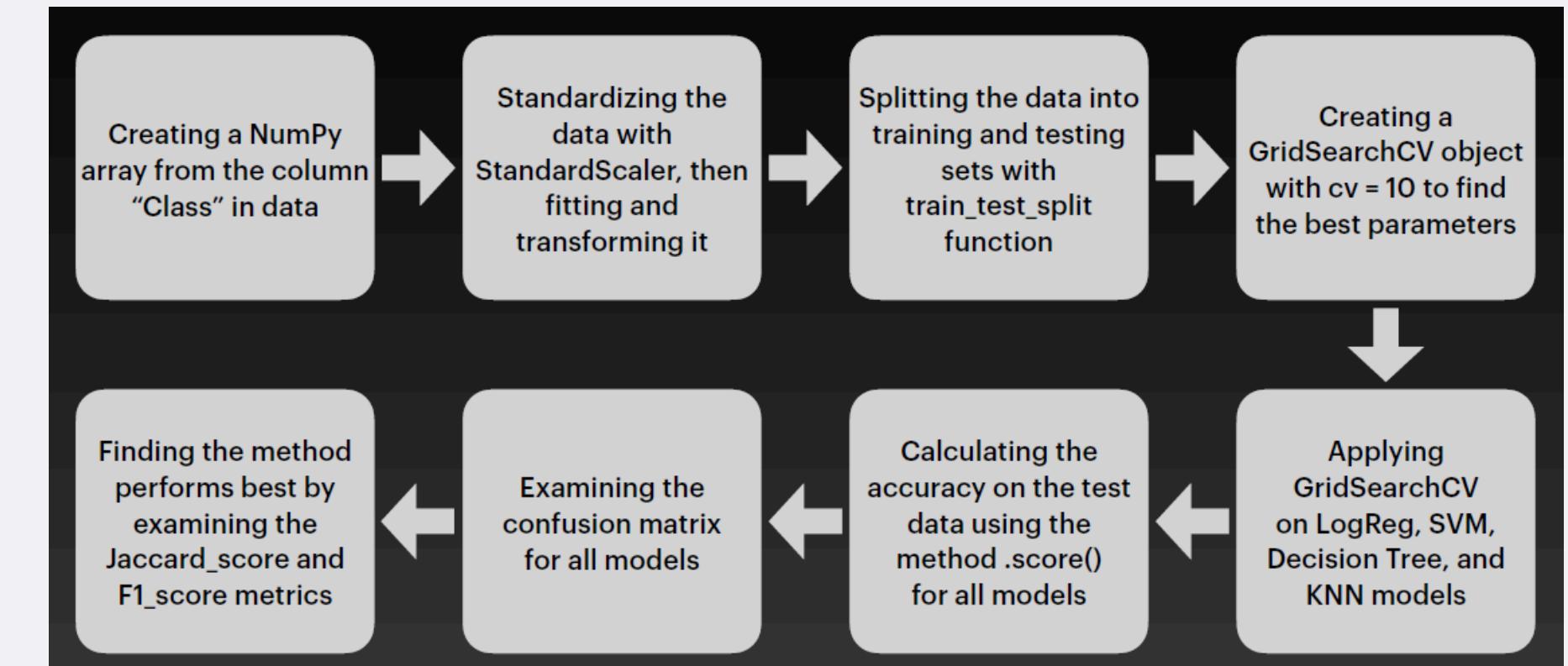
*These interactive visualizations enable comprehensive analysis of SpaceX launch performance across multiple dimensions simultaneously*

[Github Link](#)

# Predictive Analysis (Classification)

## Key Steps Taken

- Data Preparation
  - Selected features and target label (Class)
  - Applied standardization to feature data
- Train-Test Split
  - Split dataset into training and test sets (80/20)
- Model Selection & Tuning
  - Built four classifiers: Logistic Regression, SVM, Decision Tree, KNN
  - Used GridSearchCV for hyperparameter optimization (10-fold cross-validation per model)
- Model Evaluation
  - Evaluated best-found models by test accuracy and confusion matrix on hold-out data
- Model Comparison & Selection
  - Compared models using validation and test set accuracy
  - Selected best-performing model: Decision Tree (highest validation score)



## Outcome:

Successfully developed and compared four ML classifiers. After comprehensive tuning and evaluation, the Decision Tree showed the highest performance and was chosen as the best model for booster landing prediction1.

[Github Link](#)

# Results

---

## Key Findings

- Success Rate Improvement: Significant increase in landing success over time with newer booster versions
- Payload Impact: Heavier payloads correlate with lower landing success rates
- Site Performance: Launch sites show varying success rates based on mission types
- Best Model: Decision Tree classifier achieved highest accuracy for landing prediction

## Business Impact: Cost Savings-

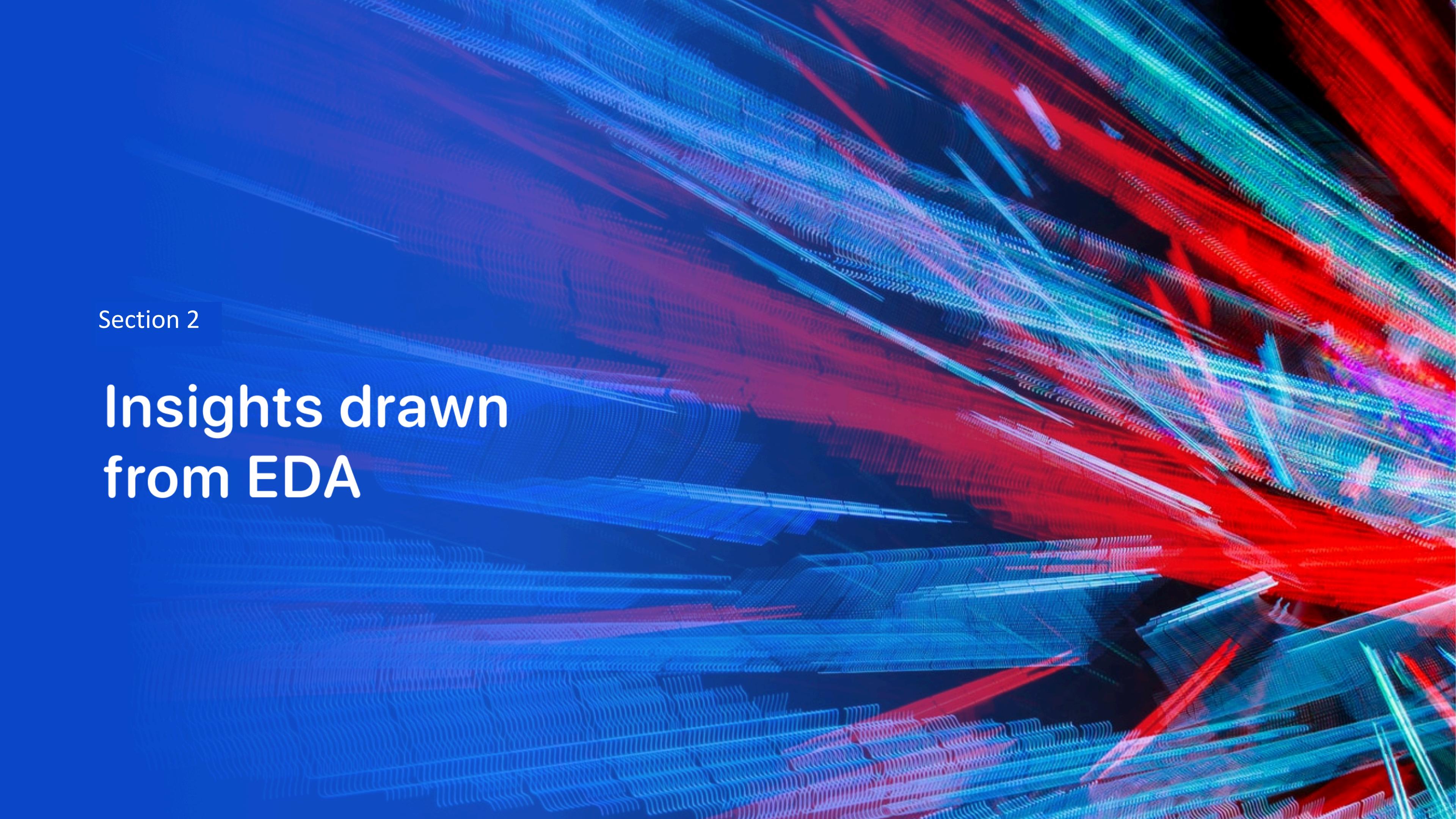
- \$50-60M saved per successful booster recovery
- Data-driven payload optimization and launch site selection

## Dashboard Features-

- Interactive filtering by launch site and payload range
- Real-time success rate visualization
- Mission planning insights

## Conclusion

- Successfully developed ML pipeline predicting Falcon 9 landing outcomes with high accuracy. Decision Tree model provides interpretable predictions for mission planning and cost optimization, delivering complete end-to-end solution from data extraction to deployment-ready model.

The background of the slide features a complex, abstract pattern of colored lines. These lines are primarily blue, red, and green, creating a sense of depth and motion. They appear to be wavy and layered, resembling a microscopic view of a neural network or a complex signal processing system.

Section 2

## Insights drawn from EDA

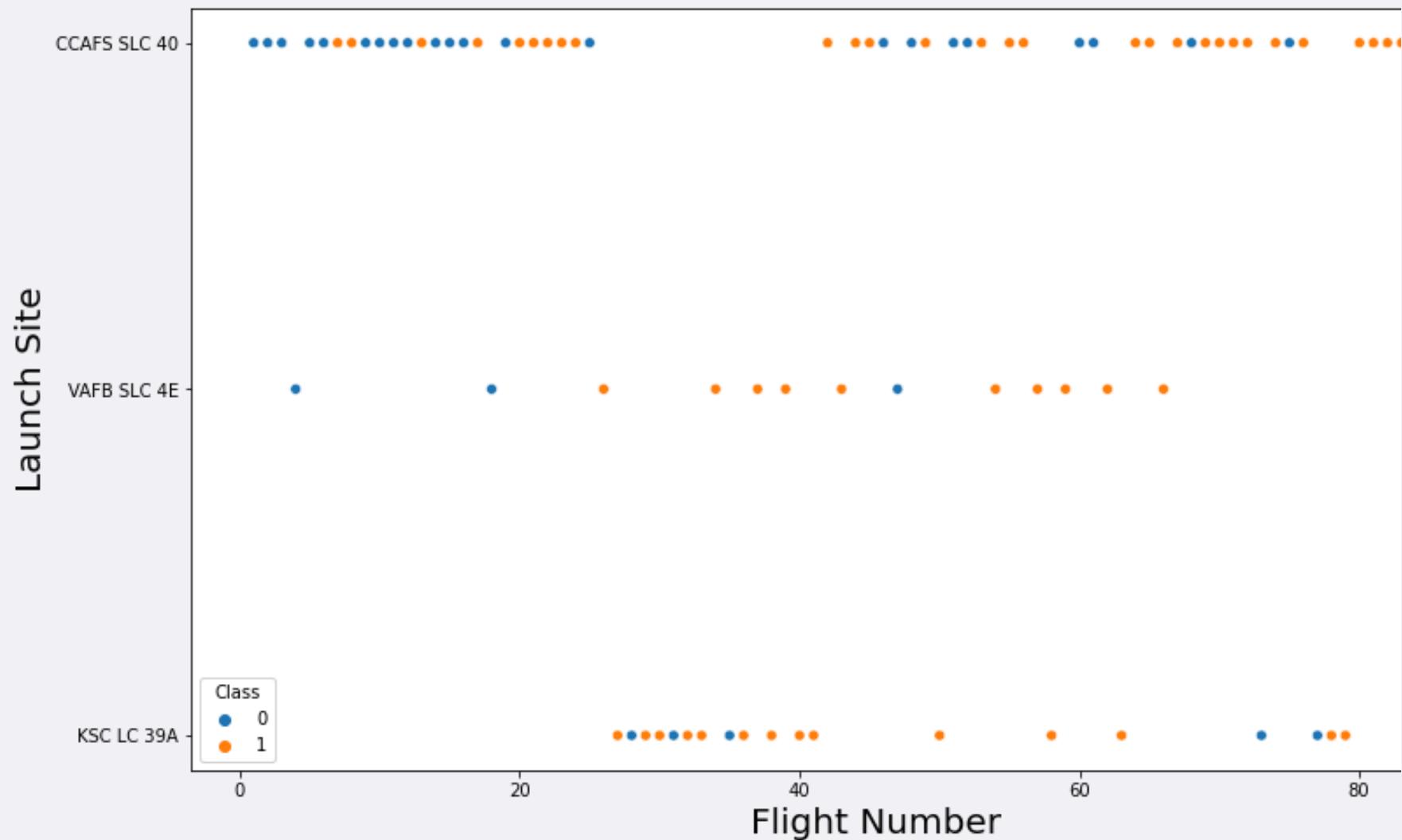
# Flight Number vs. Launch Site

## Key Observations:

- CCAFS SLC 40: Shows the highest launch frequency with missions distributed throughout the flight sequence, displaying a mix of successful and failed landings
- VAFB SLC 4E: Limited number of launches concentrated in the middle flight range, with mixed success rates
- KSC LC 39A: Launches appear later in the flight sequence with relatively good success rates

## Analysis Purpose:

- This visualization helps identify patterns in landing success across different launch sites and over time, enabling analysis of whether certain facilities have better infrastructure or conditions that contribute to higher booster recovery success rates. The temporal aspect (flight number progression) also reveals how SpaceX's landing capabilities evolved across their launch program.



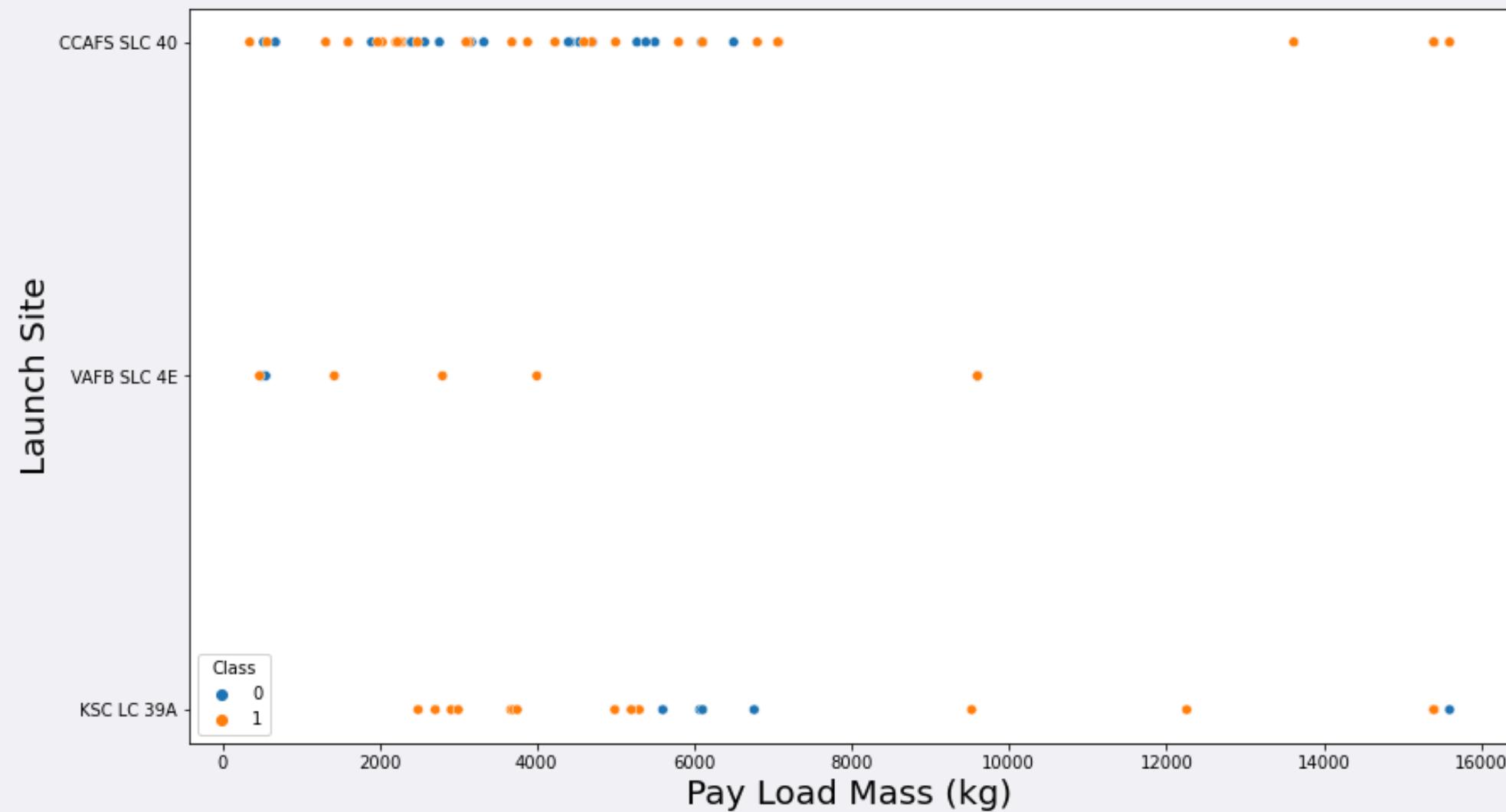
# Payload vs. Launch Site

## Key Observations:

- CCAFS SLC 40: Handles the widest range of payload masses (0-16,000 kg) with the highest launch frequency, showing mixed success rates across different payload weights
- KSC LC 39A: Launches concentrated in lower to mid-range payload masses with relatively consistent performance
- VAFB SLC 4E: Limited launches with lighter payloads, showing sporadic success patterns

## Analysis Insights:

- The visualization reveals that heavier payloads (above 10,000 kg) appear to correlate with certain launch sites and success patterns. CCAFS SLC 40 demonstrates the most operational flexibility, handling both light and heavy payloads, while other sites show more specialized payload ranges. This data helps identify optimal launch site selection based on mission payload requirements and expected success rates.

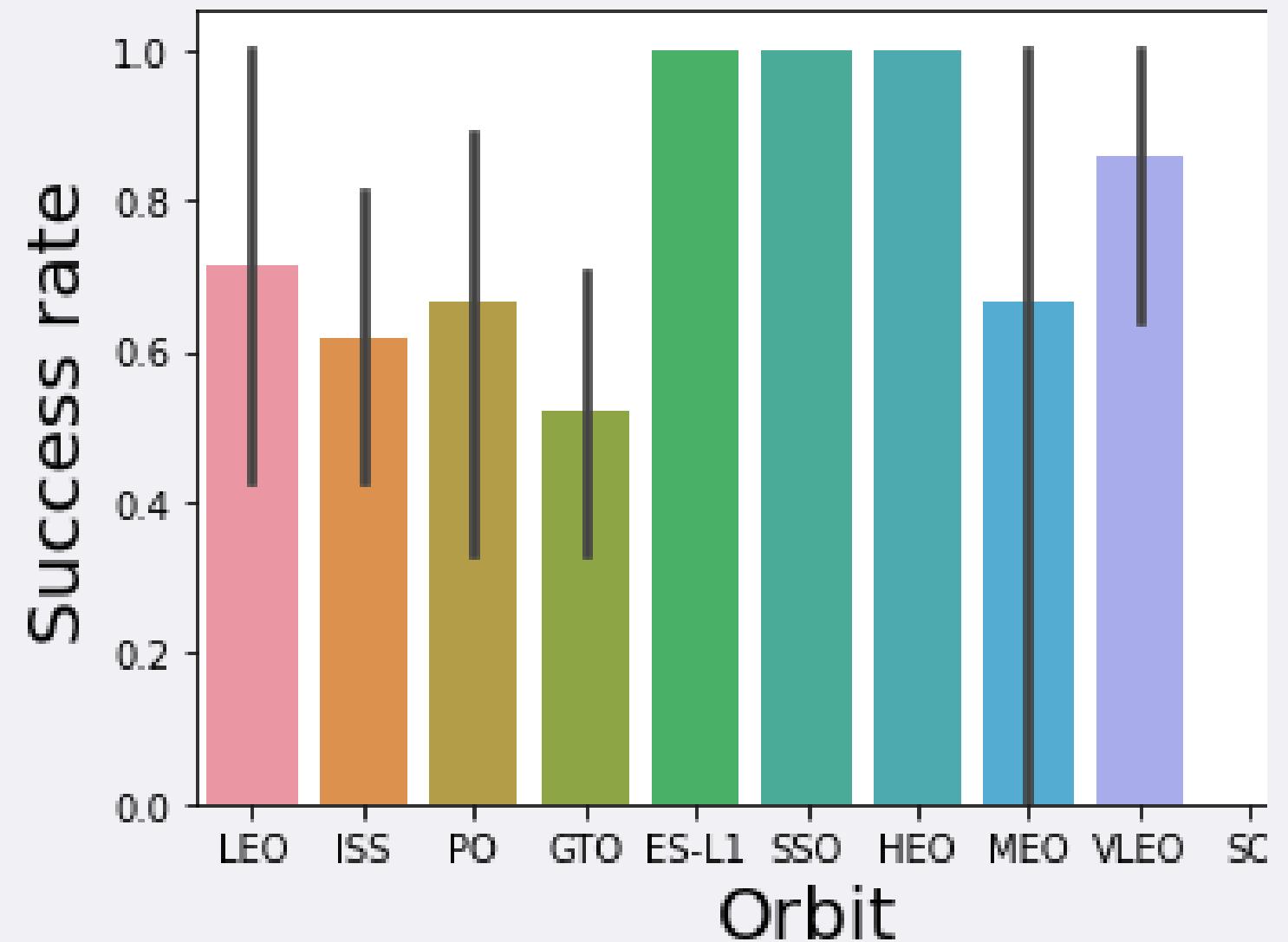


# Success Rate vs. Orbit Type

- This bar chart shows booster landing success rates for different orbital destinations. Each bar represents the success rate (0-1.0 scale) for missions targeting specific orbits like LEO (Low Earth Orbit), ISS, GTO (Geostationary Transfer Orbit), etc.

## Key Insights:

- SSO, HEO, MEO show near-perfect success rates (~1.0)
- LEO and ISS missions have moderate success rates (~0.7)
- GEO missions show the highest success rate
- Error bars indicate variability in performance for each orbit type
- The chart reveals that orbital destination significantly impacts landing success, likely due to varying fuel requirements and mission complexity for different orbital trajectories.



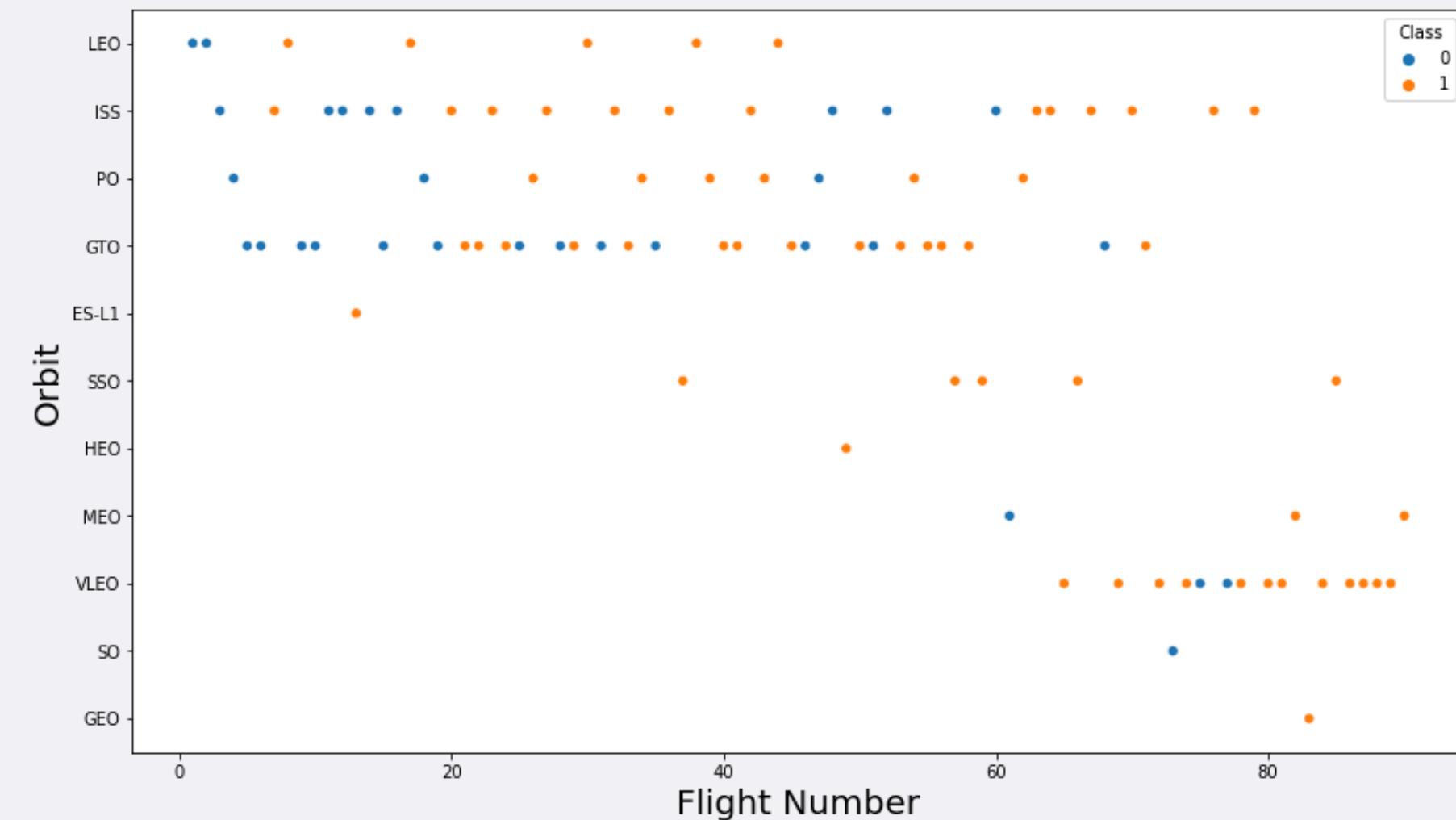
# Flight Number vs. Orbit Type

This scatter plot shows SpaceX launch outcomes across flight numbers (mission sequence) categorized by orbit type (LEO, ISS, PO, GTO, etc.). Blue dots represent Class 0 (one outcome type) and orange dots represent Class 1 (another outcome type).

## Key Observations:

- LEO missions dominate early flight numbers with mixed success patterns
- ISS missions cluster in mid-range flight numbers showing consistent performance
- GTO missions appear throughout the timeline with varying outcomes
- Later flight numbers show more diverse orbit types, indicating SpaceX's expanding mission capabilities

The visualization demonstrates how SpaceX's mission portfolio diversified over time while revealing orbit-specific success patterns across their launch history.



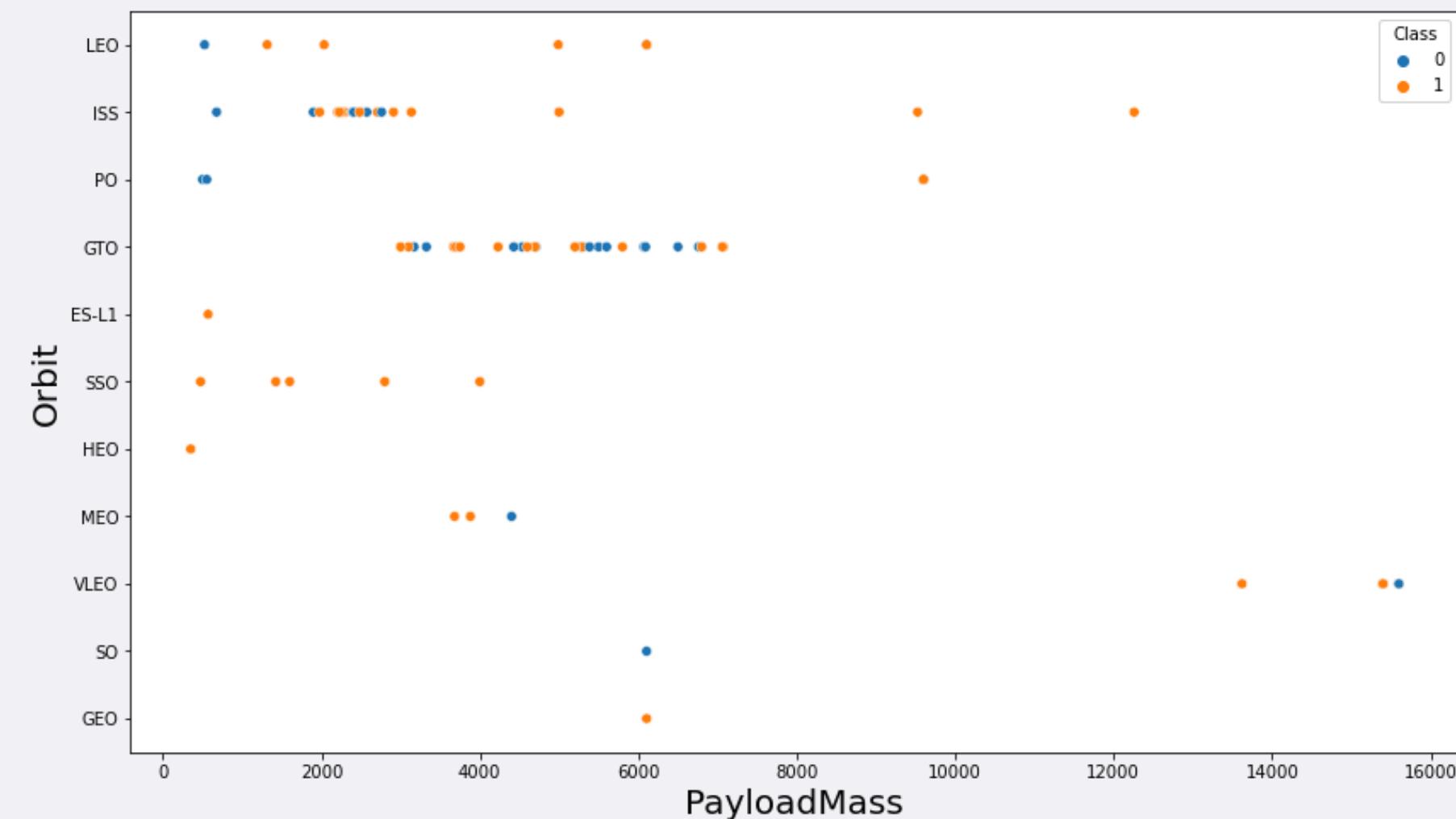
# Payload vs. Orbit Type

This scatter plot displays the relationship between Payload Mass (kg) and Orbit Type for SpaceX missions, with color coding showing landing outcomes (blue for Class 0, orange for Class 1).

## Key Observations:

- GTO missions handle the heaviest payloads (up to ~15,000 kg) with mixed success rates
- LEO missions show the widest payload range distribution with varied outcomes
- ISS missions cluster around lighter payloads (~2,000-6,000 kg) with consistent performance
- SSO, PO, ES-L1 missions handle lighter payloads with generally good success rates

The chart reveals that heavier payloads and certain orbit types (like GTO) present greater challenges for successful booster recovery, likely due to increased fuel consumption requirements.



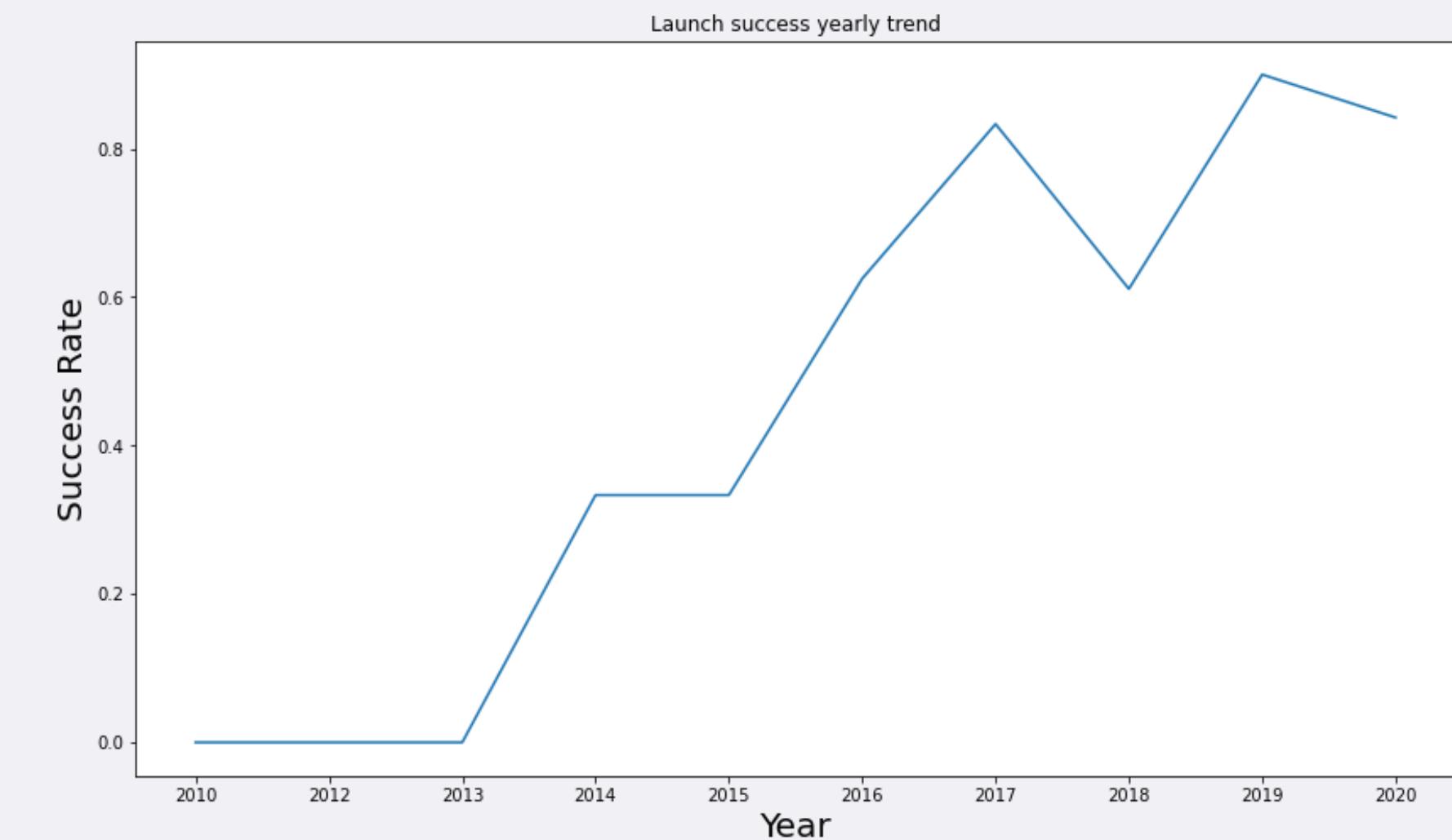
# Launch Success Yearly Trend

This line chart shows the evolution of SpaceX booster landing success rates over time from 2010 to 2020.

## Key Observations:

- 2010-2013: Zero success rate - no successful landings during early program years
- 2013-2014: First breakthrough achieving ~35% success rate
- 2015: Maintained steady ~35% success rate
- 2016-2017: Dramatic improvement reaching peak of ~83% success
- 2018: Temporary dip to ~62% success rate
- 2019: Recovery to highest point at ~95% success rate
- 2020: Slight decline to ~85% but still maintaining high performance

The chart demonstrates SpaceX's remarkable improvement in landing technology over the decade, transitioning from complete failure to achieving consistently high success rates above 80% in recent years.



# All Launch Site Names

---

This SQL query extracts all unique launch site locations from the SpaceX database table (SPACEEXTBL).

Query Breakdown:

- **SELECT DISTINCT**: Returns only unique values, eliminating duplicate entries
- **LAUNCH\_SITE**: The database column containing launch facility names
- **as "Launch\_Sites"**: Renames the output column for clearer presentation
- **FROM SPACEEXTBL**: Specifies the SpaceX data table as the source

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

# Launch Site Names Begin with 'CCA'

This SQL query retrieves the first 5 launch records from Cape Canaveral Air Force Station facilities.

## Query Breakdown:

- **SELECT \*:** Returns all columns from the table
- **FROM SPACEXTBL:** Queries the SpaceX database table
- **WHERE LAUNCH\_SITE LIKE 'CCA%':** Filters for launch sites starting with "CCA" (Cape Canaveral Air Force Station)
- **LIMIT 5:** Restricts output to only the first 5 matching records

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

# Total Payload Mass

---

SQL Query Explanation: NASA CRS Total Payload Mass

This SQL query calculates the total payload mass delivered for NASA Commercial Resupply Services (CRS) missions.

Query Breakdown:

- `SELECT SUM(PAYLOAD_MASS__KG_)`: Adds up all payload masses in kilograms
- `AS "Total payload mass by NASA (CRS)"`: Renames the output column for clarity
- `FROM SPACEXTBL`: Queries the SpaceX database table
- `WHERE CUSTOMER = 'NASA (CRS)'`: Filters only for NASA Commercial Resupply Service missions

Total payload mass by NASA (CRS)
45596

# Average Payload Mass by F9 v1.1

---

This SQL query calculates the total payload mass delivered for NASA Commercial Resupply Services (CRS) missions.

Query Breakdown:

- `SELECT AVG(PAYLOAD_MASS__KG_)`: Averages up all payload masses in kilograms
- `AS "Avrage payload mass by NASA (CRS)"`: Renames the output column for clarity
- `FROM SPACEXTBL`: Queries the SpaceX database table
- `WHERE CUSTOMER = 'NASA (CRS)'`: Filters only for NASA Commercial Resupply Service missions

Average payload mass by Booster Version F9 v1.1
2928

# First Successful Ground Landing Date

---

SQL Query Explanation: First Successful Ground Pad Landing

This SQL query finds the earliest date when SpaceX achieved their first successful booster landing on a ground pad.

Query Breakdown:

- `SELECT MIN(DATE)`: Returns the minimum (earliest) date value from the filtered results
- `AS "Date of first successful landing outcome in ground pad"`: Renames the output column for clarity
- `FROM SPACEXTBL`: Queries the SpaceX database table
- `WHERE LANDING_OUTCOME = 'Success (ground pad)'`: Filters only for missions with successful ground pad landings

Date of first successful landing outcome in ground pad

2015-12-22

# Successful Drone Ship Landing with Payload between 4000 and 6000

---

This SQL query identifies booster versions that successfully landed on drone ships while carrying medium-weight payloads.

Query Breakdown:

- `SELECT BOOSTER_VERSION`: Returns the booster version identifiers (e.g., F9 v1.1, F9 FT, F9 B4)
- `FROM SPACEXTBL`: Queries the SpaceX database table
- `WHERE LANDING_OUTCOME = 'Success (drone ship)'`: Filters for successful ASDS (Autonomous Spaceport Drone Ship) landings
- `AND PAYLOAD_MASS_KG_ BETWEEN 4000 AND 6000`: Further filters for payloads in the 4,000-6,000 kg range

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

# Total Number of Successful and Failure Mission Outcomes

---

This SQL query provides a comprehensive mission outcome comparison by counting successful and failed SpaceX missions side-by-side.

Query Breakdown:

- Two Subqueries with Cross Join:
  - Success Subquery: COUNT(\*) WHERE MISSION\_OUTCOME LIKE 'Success%' - counts all missions with outcomes starting with "Success"
  - Failure Subquery: COUNT(\*) WHERE MISSION\_OUTCOME LIKE 'Failure%' - counts all missions with outcomes starting with "Failure"
- Pattern Matching: LIKE 'Success%' and LIKE 'Failure%' capture variations like "Success (drone ship)", "Failure (parachute)", etc.
- Cross Join: Combines both counts into a single row result

	number_of_success_outcomes	number_of_failure_outcomes
	100	1

# Boosters Carried Maximum Payload

This SQL query identifies which booster version(s) carried the heaviest payload in SpaceX's launch history.

Query Breakdown:

- `SELECT DISTINCT BOOSTER_VERSION`: Returns unique booster version identifiers, eliminating duplicates
- `FROM SPACEXTBL`: Queries the SpaceX database table
- `WHERE PAYLOAD_MASS_KG_ =`: Filters for records matching a specific payload mass
- `(SELECT MAX(PAYLOAD_MASS_KG_) FROM SPACEXTBL)`: Subquery that finds the maximum payload mass in the entire dataset

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

# 2015 Launch Records

---

This SQL query retrieves specific mission details for SpaceX launches in 2015 that failed to land successfully on drone ships.

Query Breakdown:

- `SELECT DATE, BOOSTER_VERSION, LAUNCH_SITE`: Returns three key mission identifiers
- `FROM SPACEXTBL`: Queries the SpaceX database table
- `WHERE year(DATE) = '2015'`: Filters for missions launched in 2015 using date extraction function
- `AND LANDING_OUTCOME = 'Failure (drone ship)'`: Further filters for failed ASDS (drone ship) landing attempts

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

This SQL query analyzes SpaceX landing attempt results during their early development period, ranking outcomes by frequency.

Query Breakdown:

- **SELECT LANDING\_\_OUTCOME, COUNT(LANDING\_\_OUTCOME) AS Landing\_Count:**  
Returns outcome types and their occurrence counts
- **FROM SPACEXTBL:** Queries the SpaceX database table
- **WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20':**  
Filters for ~7-year period covering early SpaceX missions
- **GROUP BY LANDING\_\_OUTCOME:** Groups records by landing outcome type for counting
- **ORDER BY COUNT(LANDING\_\_OUTCOME) DESC:** Sorts results from most to least frequent outcomes

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

The background of the slide is a nighttime satellite photograph of Earth. The curvature of the planet is visible against the dark void of space. City lights are scattered across the continents, appearing as glowing yellow and white dots and streaks along coastlines and major rivers. The atmosphere is visible as a thin blue layer at the top of the image.

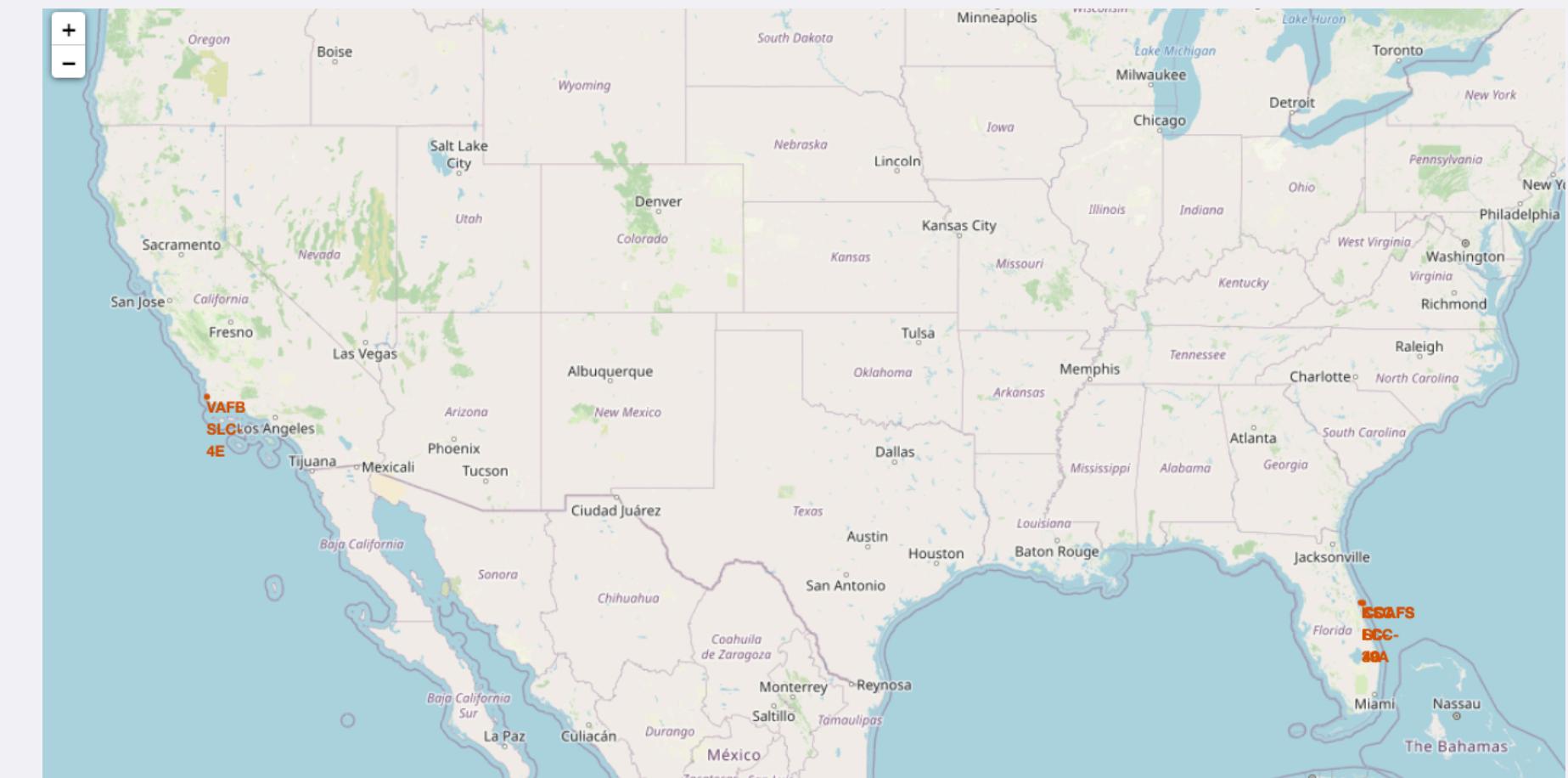
Section 3

# Launch Sites Proximities Analysis

# Map with Marked Launch Sites

## Explanation:

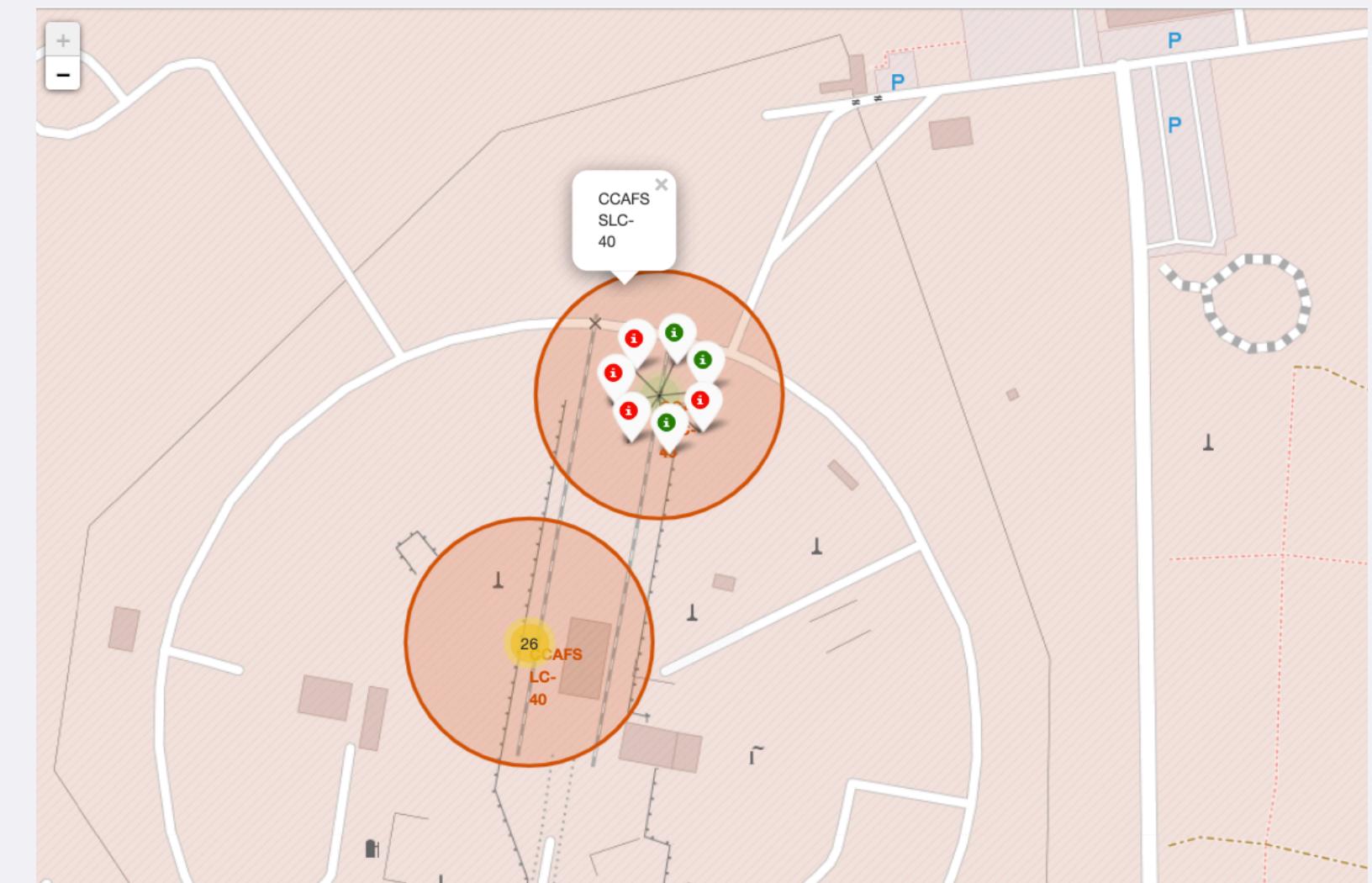
- Most of Launch sites are in proximity to the Equator line. The land is moving faster at the equator than any other place on the surface of the Earth. Anything on the surface of the Earth at the equator is already moving at 1670 km/hour. If a ship is launched from the equator it goes up into space, and it is also moving around the Earth at the same speed it was moving before launching. This is because of inertia. This speed will help the spacecraft keep up a good enough speed to stay in orbit.
- All launch sites are in very close proximity to the coast; while launching rockets towards the ocean it minimizes the risk of having any debris dropping or exploding near people.



# Mark the Success/Failed Launches for Each Site on the Map

## Explanation:

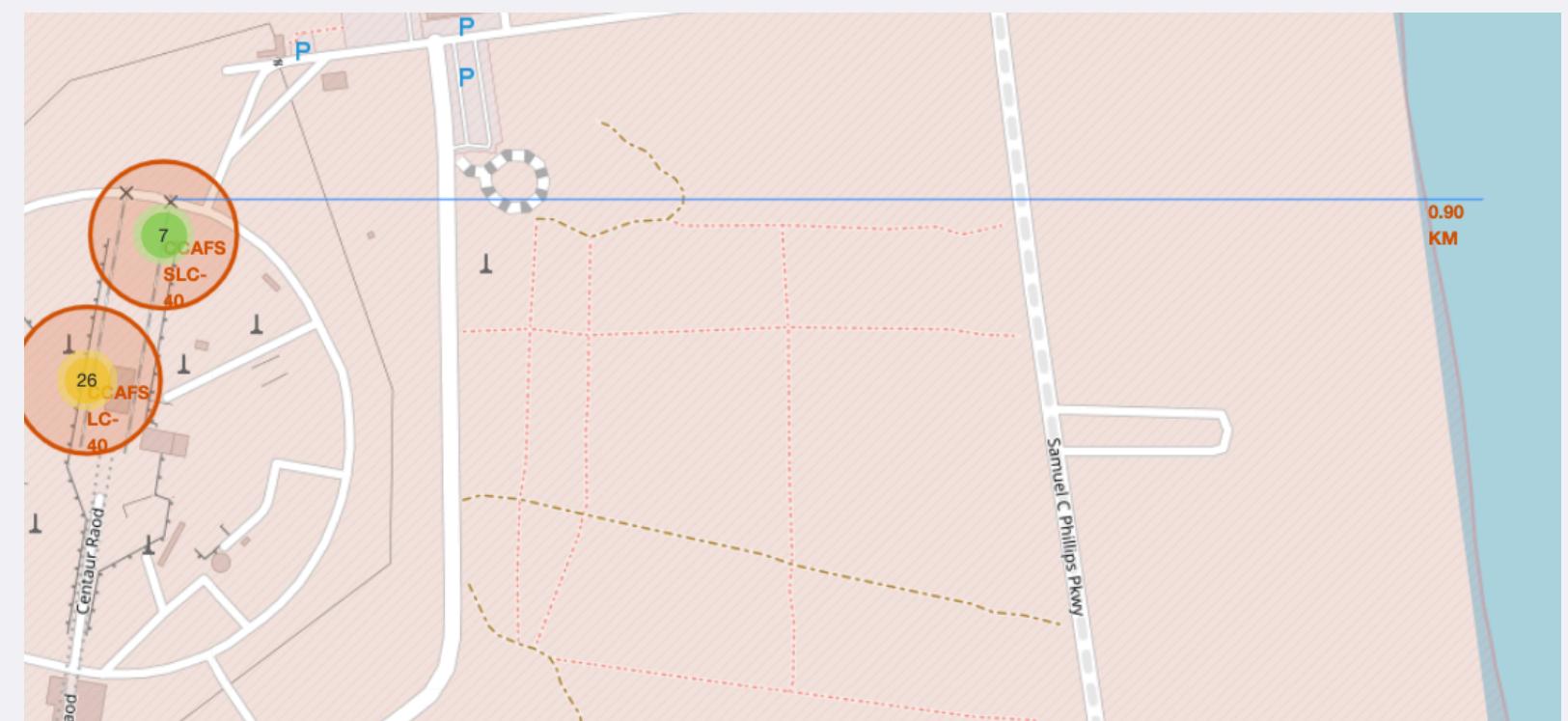
- From the color-labeled markers we should be able to easily identify which launch sites have relatively high success rates.
  - Green Marker = Successful Launch
  - Red Marker = Failed Launch
  - Launch Site KSC LC-39A has a very high Success Rate.



# Distance from the Launch Site KSC LC-39A to its Proximities

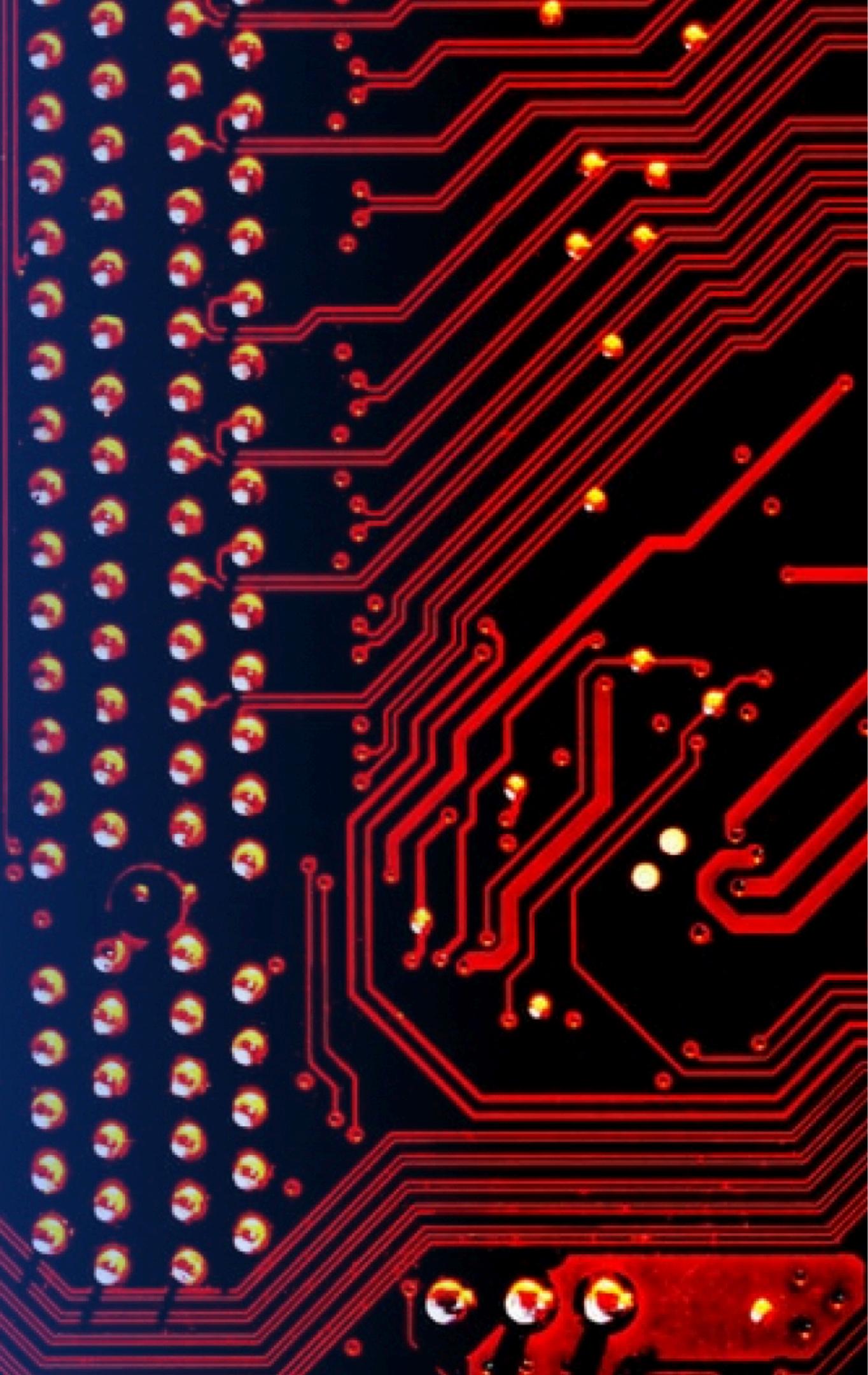
## Explanation:

- From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:
  - relatively close to railway (15.23 km)
- relatively close to highway (20.28 km)
  - relatively close to coastline (14.99 km)
- Also, the launch site KSC LC-39A is relatively close to its closest city Titusville (16.32 km).
- Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas.

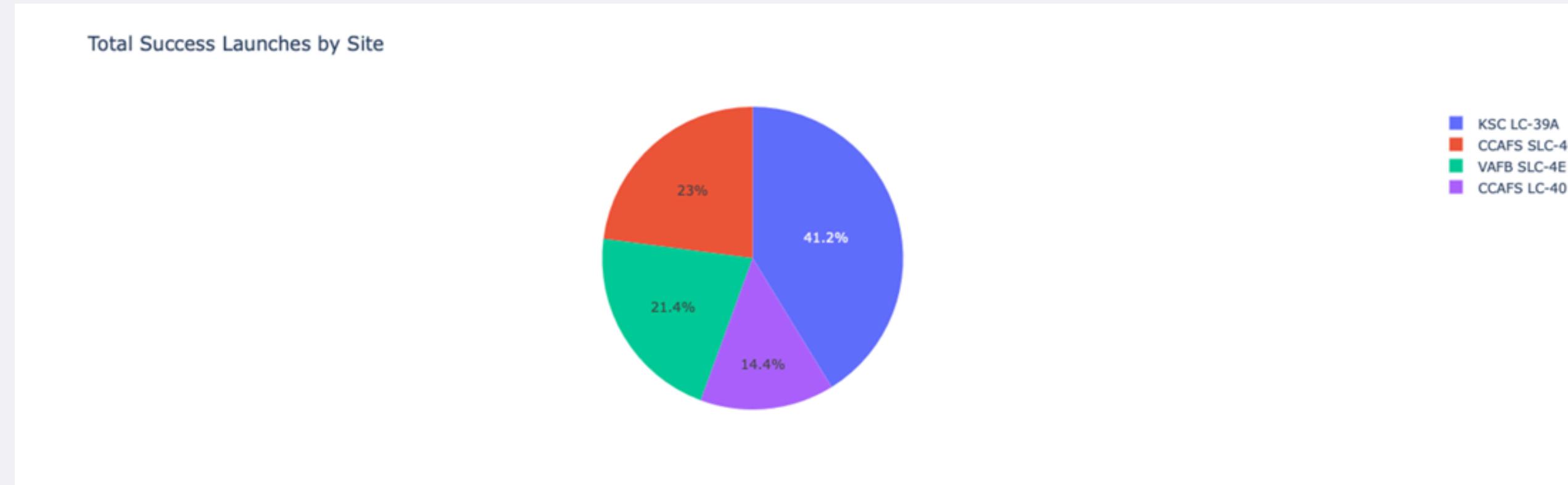


Section 4

# Build a Dashboard with Plotly Dash



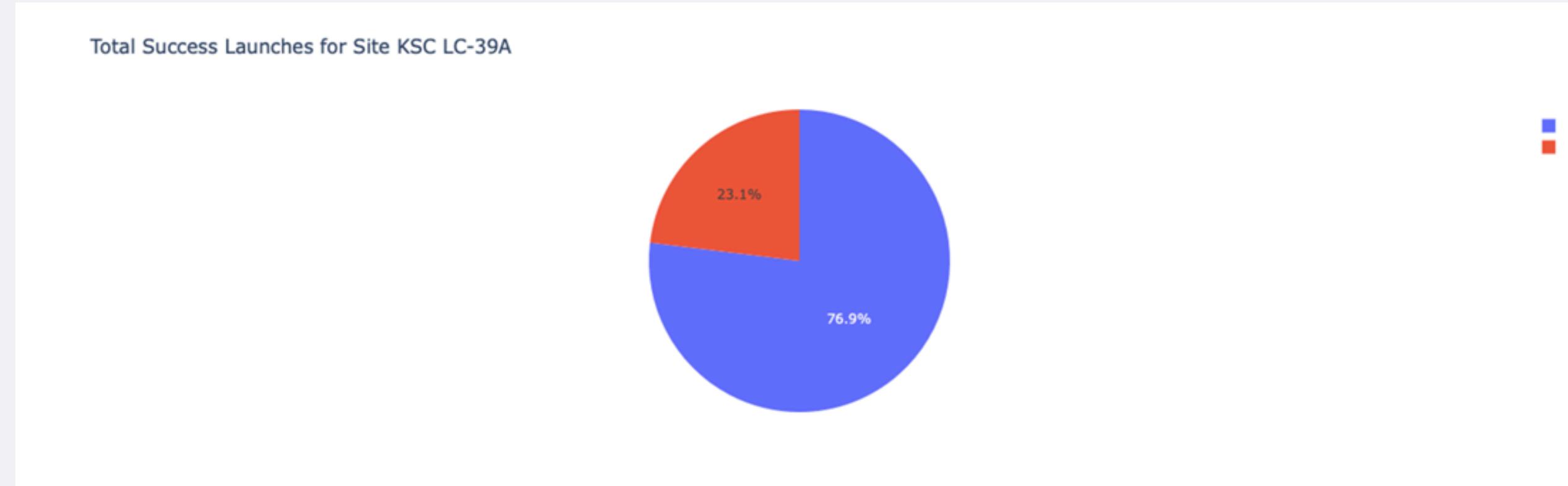
# Launch success count for all sites



Explanation:

- The chart clearly shows that from all the sites, KSC LC-39A has the most successful launches.

# Launch Site with Highest Launch Success Ratio



Explanation:

- KSC LC-39A has the highest launch success rate (76.9%) with 10 successful and only 3 failed landings.

# Payload Mass vs. Launch Outcome for All Sites

Explanation:

- The charts show that payloads between 2000 and 5500 kg have the highest success rate.



The background of the slide features a dynamic, abstract design. It consists of several thick, curved lines that transition from a bright yellow at the top right to a deep blue at the bottom left. These lines create a sense of motion and depth, resembling a tunnel or a stylized landscape. The overall effect is modern and professional.

Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

---

## Explanation:

- Based on the scores of the Test Set, we can not confirm which method performs best.
- Same Test Set scores may be due to the small test sample size (18 samples). Therefore, we tested all methods based on the whole Dataset.
- The scores of the whole Dataset confirm that the best model is the Decision Tree Model. This model has not only higher scores, but also the highest accuracy.

## Scores and Accuracy of the Test Set

	<b>LogReg</b>	<b>SVM</b>	<b>Tree</b>	<b>KNN</b>
<b>Jaccard_Score</b>	0.800000	0.800000	0.800000	0.800000
<b>F1_Score</b>	0.888889	0.888889	0.888889	0.888889
<b>Accuracy</b>	0.833333	0.833333	0.833333	0.833333

## Scores and Accuracy of the Entire Data Set

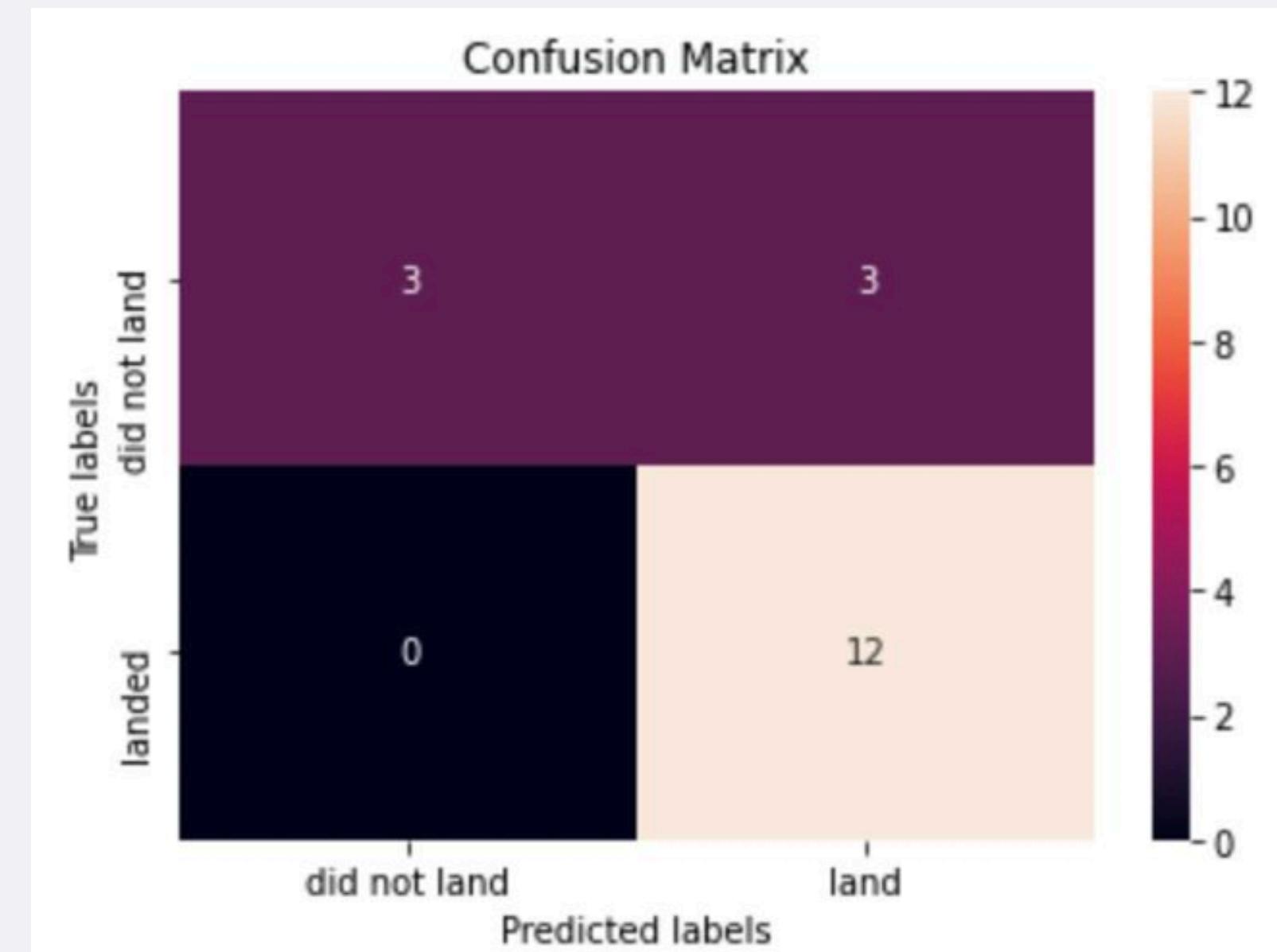
	<b>LogReg</b>	<b>SVM</b>	<b>Tree</b>	<b>KNN</b>
<b>Jaccard_Score</b>	0.833333	0.845070	0.882353	0.819444
<b>F1_Score</b>	0.909091	0.916031	0.937500	0.900763
<b>Accuracy</b>	0.866667	0.877778	0.911111	0.855556

# Confusion Matrix

Explanation:

- Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

		Predicted Values	
		Negative	Positive
Actual Values	Negative	TN	FP
	Positive	FN	TP



# Conclusions

---

- Decision Tree Model is the best algorithm for this dataset.
- Launches with a low payload mass show better results than launches with a larger payload mass.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- The success rate of launches increases over the years.
- KSC LC-39A has the highest success rate of the launches from all the sites.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.

# Appendix

---

Special Thanks to:

- Instructors
- Coursera
- IBM

Thank you!

