IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

ADHAM HAMED

2025-03-21

Outline

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

Executive Summary

🔧 Methodology Overview

•**Data Sources**:
- SpaceX REST API , Web-scraped Wikipedia launch records

•**Data Processing**:
- Filtered for **Falcon 9** launches ,Merged data from multiple API endpoints

•**Exploratory Data Analysis (EDA)**:
- Identified key success correlations , Visualized trends over time and feature relationships

•**Visualization Tools**:
- **Folium** for interactive launch site mapping and **Plotly Dash** for dashboard with dropdowns, sliders, pie & scatter charts

•**Machine Learning Models Tested**:
- Logistic Regression,Support Vector Machine,Decision Tree,K-Nearest Neighbors,Used **Grid Search** for hyperparameter tuning

Executive Summary

📊 Summary of Results

- **EDA Highlights**:
  - Launch success rate improved significantly **after 2013**
  - **KSC LC-39A** and **VAFB SLC 4E** had ~77% success
  - Payloads over **10,000 kg** at CCAFS LC-40 → 100% success

- **Best Performing Model**:
  - Final model selected based on **test accuracy and confusion matrix** analysis

- **Interactive Dashboard Insights**:
  - Users can explore **success rates by site**, **payload impact**, and **outcome distributions**
  - Dashboard enables **real-time exploration** for stakeholders

Introduction

📌 Project Background & Context

• The **commercial space industry** is expanding rapidly with key players like **SpaceX**, **Blue Origin**, and **Rocket Lab**
• **SpaceX** leads the market, offering cost-effective launches through **reusable first-stage rockets (Falcon 9)**
• Reusability is a major factor in reducing launch costs (from ~$165M to ~$62M)
• This project simulates a competitor analysis from a fictional rival company, **Space Y**, founded by *Allon Musk*

🧠 Problem Statements

• Can we predict if the Falcon 9's first stage will land successfully?
• What features most influence successful landings?
• How can interactive analytics help stakeholders explore launch data effectively?
• Which machine learning model provides the most accurate predictions?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX REST API , Web-scraped Wikipedia launch records
- Perform data wrangling
  - Filtered for Falcon 9 launches
  - Merged data from multiple API endpoints (rocket, payload, core, launch site)
  - Handled missing values and one-hot encoded categorical features
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression,Support Vector Machine,Decision Tree,K-Nearest Neighbors,Used Grid Search for hyperparameter tuning

Data Collection

Data Sources

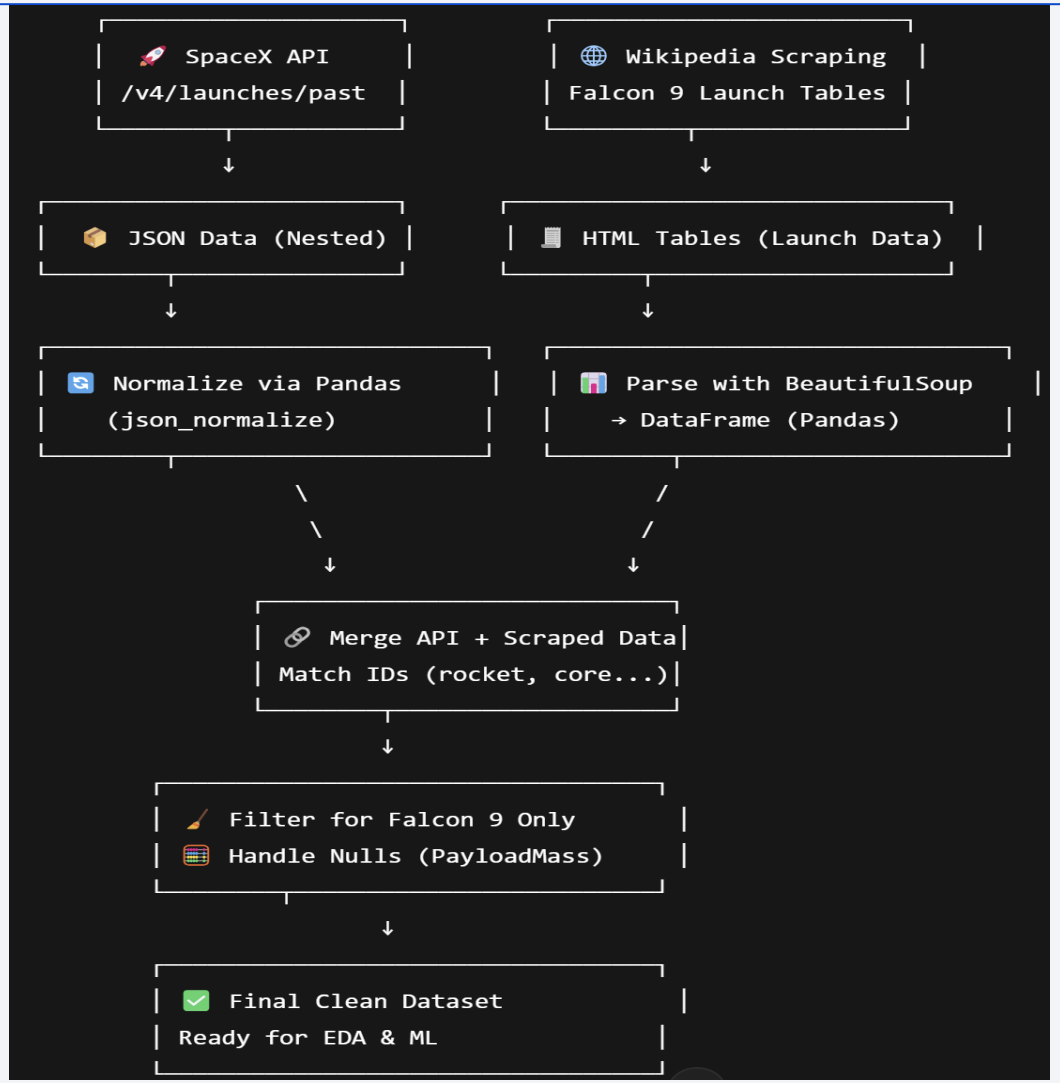
- SpaceX REST API

- Web Scraping

Data Collection – SpaceX API & Web Scraping

Data Sources

- SpaceX REST API
  - Endpoint used: /v4/launches/past
  - Supplemented with: /rockets, /payloads, /cores, /launchpads
- Web Scraping
  - Wikipedia launch records using BeautifulSoup
  - Extracted Falcon 9 launch tables for cross-verification and enrichment
  - Link to SpaceX API
  - Link to Web Scraping



Data Wrangling

🔧 Key Data Processing Steps

- ✅ Filtered for Falcon 9 launches (excluded Falcon 1 data)
- 🔗 Merged datasets from multiple API endpoints using ID fields
- 🧮 Handled Null Values:
  - Replaced PayloadMass nulls with column mean
  - Left LandingPad nulls for one-hot encoding later
- 🏷️ One-Hot Encoding of categorical features:
  - Launch Site, Orbit, Landing Outcome, etc.
- ⚖️ Feature Standardization applied to numerical columns
- 🧼 Cleaned and structured final dataset for EDA and ML
- Link to Data Wrangling

EDA with Data Visualization

🧩 Summary of Plotted Charts

Chart Type	Features Visualized	Purpose
Bar Chart	Launch Site vs. Landing Outcome Success Rate	Compare performance of different launch sites
Pie Chart	Overall Landing Outcome Distribution	Visualize class imbalance between successful vs. failed landings
Scatter Plot	Payload Mass vs. Landing Outcome (colored by site)	Identify payload ranges related to landing success
Line Plot	Launch Success Rate over Time / Flight Number	Show trend of increasing success over years
Heatmap (Correlation Matrix)	Numerical feature correlations (Payload, Reuse Count, Block)	Detect which variables are strongly associated with success
Histogram	Distribution of Payload Mass	Understand range and skew of payload weights
Box Plot	Payload Mass grouped by Orbit	See how payload varies with orbit types

EDA with SQL

✅ **Filtered successful drone ship landings**
- Identified distinct booster versions with payloads between **4,000–6,000 kg**
- WHERE Landing_Outcome = 'Success (drone ship)' AND PayloadMass BETWEEN 4000 AND 6000

📊 **Summarized mission outcomes**
- Counted total launches grouped by Mission_Outcome
- Helped visualize success/failure distribution

🚀 **Identified heaviest payload booster**
- Selected booster version that carried the **maximum payload mass**

🗓️ **Analyzed failed drone ship landings in 2015**
- Extracted **month, booster version, and launch site** for failed landings in 2015
- Useful for identifying seasonal or version-specific failure trends

📈 **Tracked landing outcomes over a time range**
- Counted landing outcomes between **2010-06-04** and **2017-03-20**
- Ordered by frequency to detect most common outcomes

EDA With SQL Link

Build an Interactive Map with Folium

Launch Site Markers

Added popups displaying the **launch site names**

Circles for Launch Sites

Marker Clusters for Launches

Distance Markers

Polylines for Distance Analysis

Link to Folium

🎯 **Why These Objects Were Added**

✓ **Enhance Visual Storytelling** – Makes it easy to **see** key geographic insights
✓ **Improve Data Exploration** – Users can **interact** and zoom in/out to discover patterns
✓ **Identify Optimal Locations** – Helps analyze launch proximity to coastlines & landing zones
✓ **Compare Success vs. Failure Rates** – Color-coded markers highlight performance

Build a Dashboard with Plotly Dash

📌 Plots & Graphs Added

1. 📈 **Pie Chart – Total Successful Launches by Site**
    1. Displays **success rate** per launch site
    2. Helps compare **which sites perform best**
    3. **Color-coded** by site for clarity

2. 📊 **Scatter Plot – Correlation Between Payload Mass & Landing Success**
    1. X-axis: **Payload Mass (kg)**
    2. Y-axis: **Landing Outcome (1 = success, 0 = failure)**
    3. Shows **how payload mass affects success rates**
    4. Color-coded by **booster version**

3. 🎛️ **Interactive Controls**
    1. **Dropdown Menu** – Selects specific launch sites for filtering
    2. **Range Slider** – Adjusts payload mass range dynamically
    3. **Hover & Click** – Allows deeper exploration of data points

🎯 **Why These Visuals Were Added**

✓ Helps identify best-performing launch sites

✓ Shows relationships between payload and landing success

✓ Interactive elements improve user engagement

✓ Enables real-time exploration instead of static charts

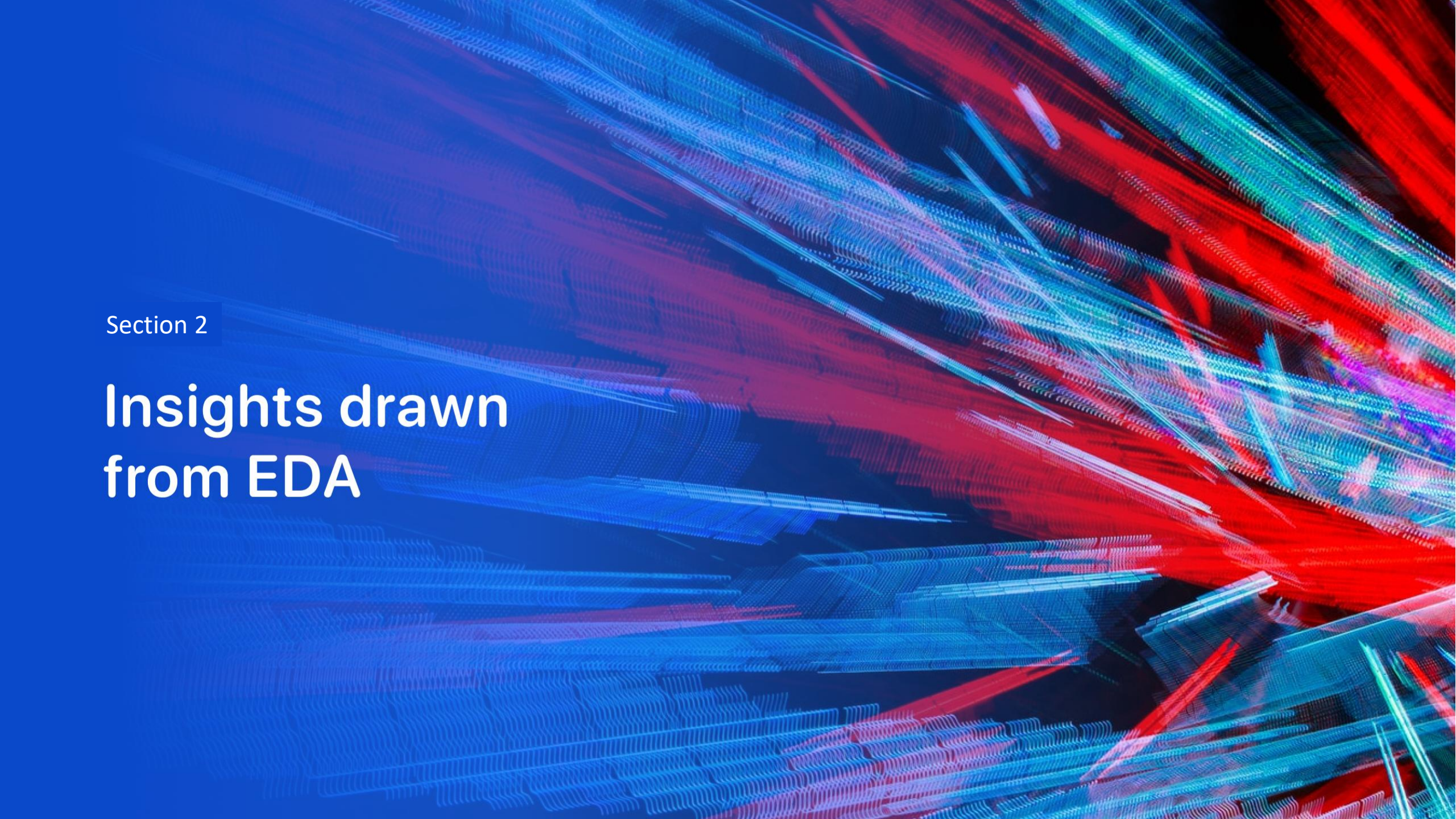
Plotly Link

Predictive Analysis (Classification)

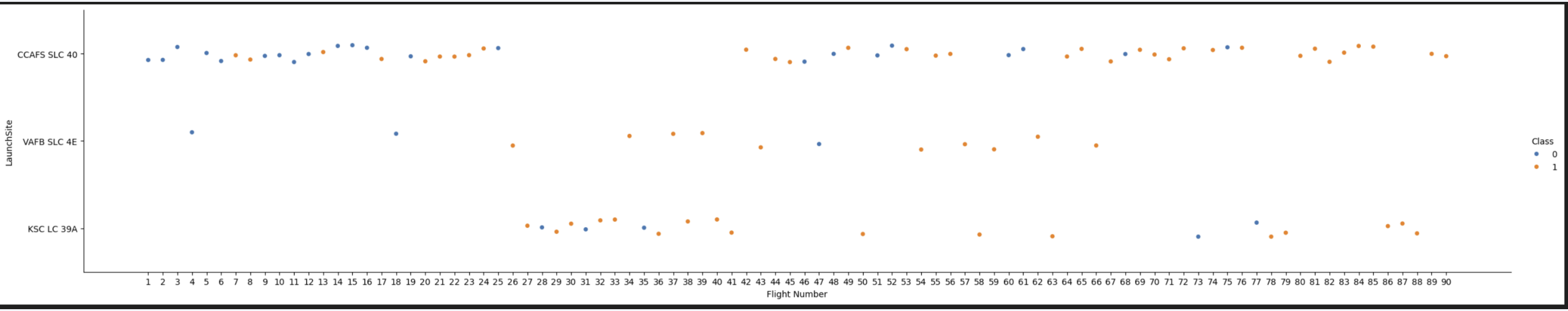
1. 📦 **Data Preprocessing**
   - Standardized numerical features using **Sklearn Preprocessing**
   - Applied **One-Hot Encoding** to categorical features
2. 🔀 **Train-Test Split**
   - Divided dataset into **80% training** and **20% testing**
   - Used train_test_split(X, y, test_size=0.2, random_state=2)
3. 📊 **Models Evaluated**
   - **Logistic Regression**
   - **Support Vector Machine (SVM)**
   - **Decision Tree Classifier**
   - **K-Nearest Neighbors (KNN)**
4. ⚙️ **Hyperparameter Tuning with Grid Search**
   - Used GridSearchCV to **find the best parameters** for each model
   - Tuned parameters like **C (regularization), kernel, tree depth, and K value**
5. 📈 **Model Performance Evaluation**
   - **Accuracy** scores compared across models
   - **Confusion Matrices** plotted to analyze **false positives & false negatives**
   - **Best Model Selected** based on **highest accuracy & lowest error rate**
   - Link to Model

Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



📈 **Key Insights**

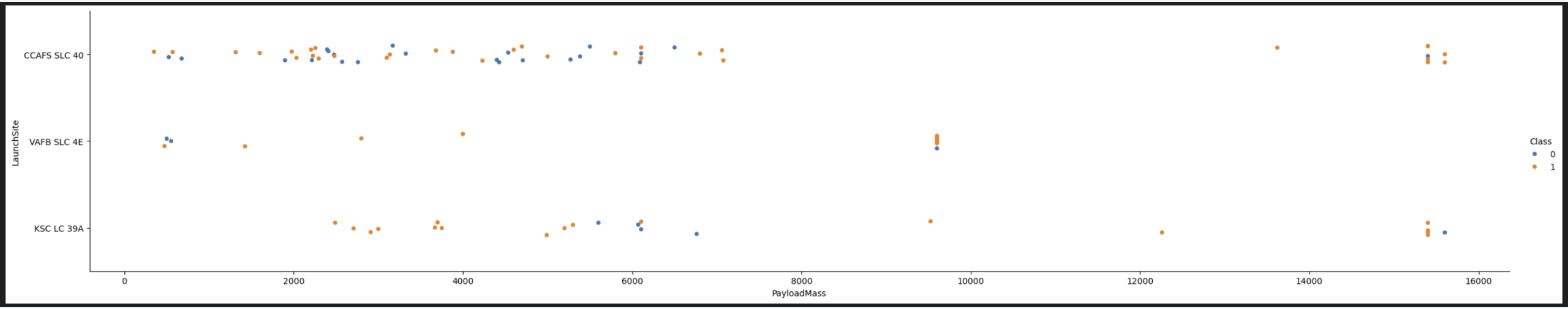
✓ Early launches were **mostly from CCAFS LC-40**

✓ **KSC LC-39A and VAFB SLC-4E** became more active after **2017**

✓ As the **flight number increases**, launch sites are used more consistently

✓ Helps in **understanding launch site success trends**

Payload vs. Launch Site



📈 **Key Insights**

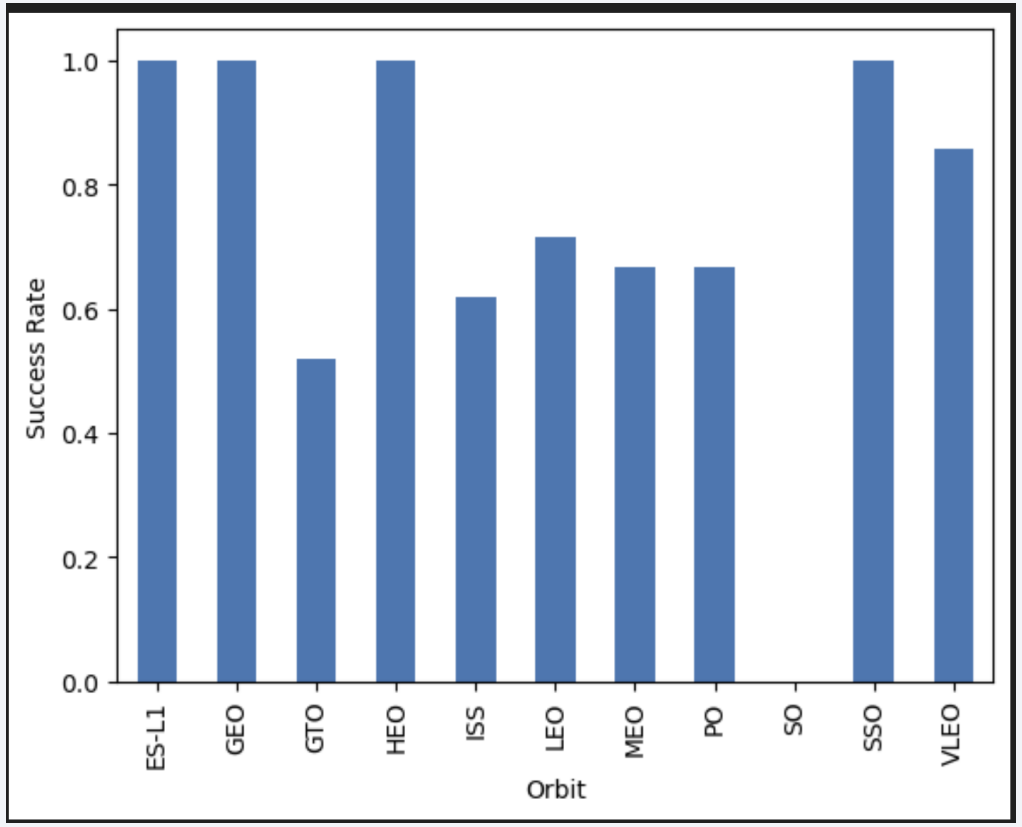
✓ **KSC LC-39A** and **CCAFS LC-40** handled a **wide range** of payload masses

✓ **VAFB SLC-4E** primarily launched **lighter payloads**

✓ **Higher payload masses (>10,000 kg)** were **mostly launched from KSC LC-39A**, showing its capability for **heavy missions**

✓ Useful for **predicting launch site selection** based on payload requirements

Success Rate vs. Orbit Type



📈 **Key Insights**

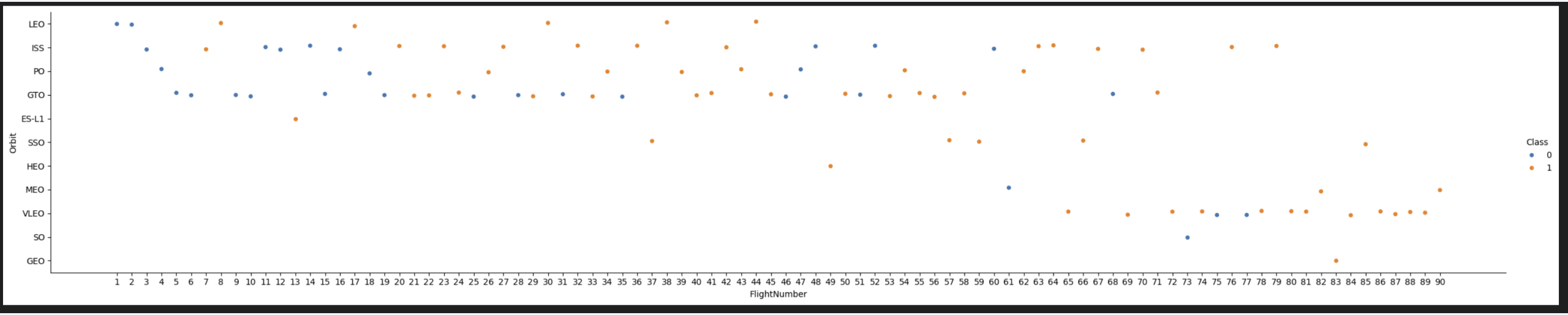
✓ **Low Earth Orbit (LEO)** had the **highest success rate**, as it requires **less energy** for delivery and return

✓ **Geostationary Transfer Orbit (GTO)** showed a **lower success rate**, as missions to GTO require **higher velocity and fuel consumption**, reducing landing feasibility

✓ **Polar and Sun-Synchronous Orbits (SSO)** had **moderate success rates**, indicating variable mission complexity

✓ Useful for **predicting landing feasibility** based on mission orbit

Flight Number vs. Orbit Type



📈 **Key Insights**

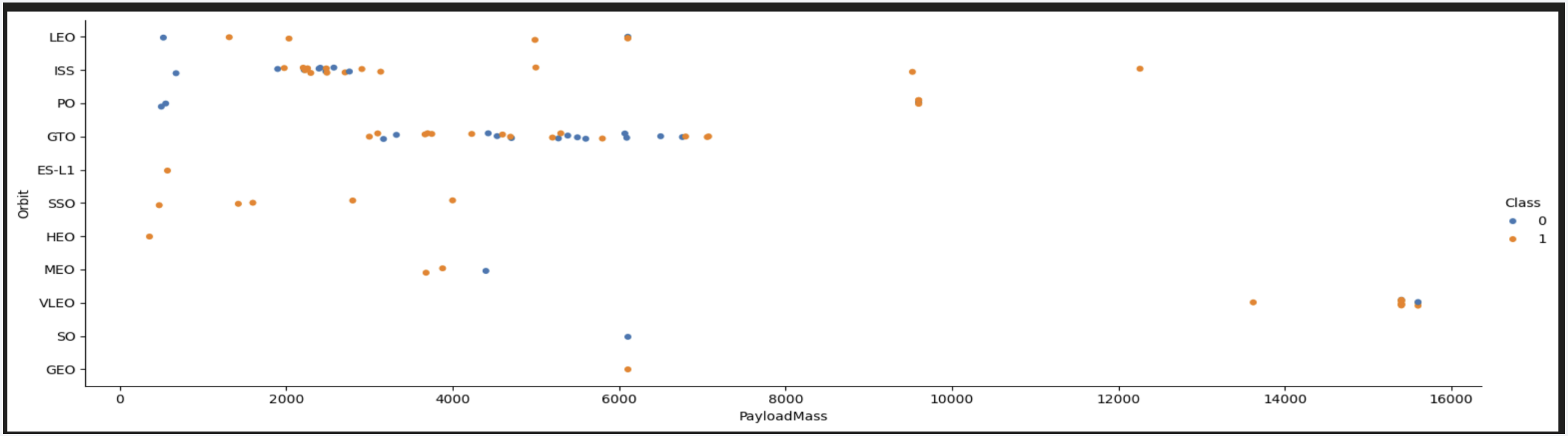
✓ **Early missions** were primarily to **LEO (Low Earth Orbit)**, suggesting initial focus on **shorter-range missions**

✓ **As flight numbers increased**, more missions targeted **GTO (Geostationary Transfer Orbit)**, requiring **higher fuel efficiency and precision**

✓ **SSO (Sun-Synchronous) and Polar orbits** were used less frequently but saw **steady use in later flights**

✓ Helps **predict SpaceX's future mission trends** based on flight experience

Payload vs. Orbit Type



📈 **Key Insights**

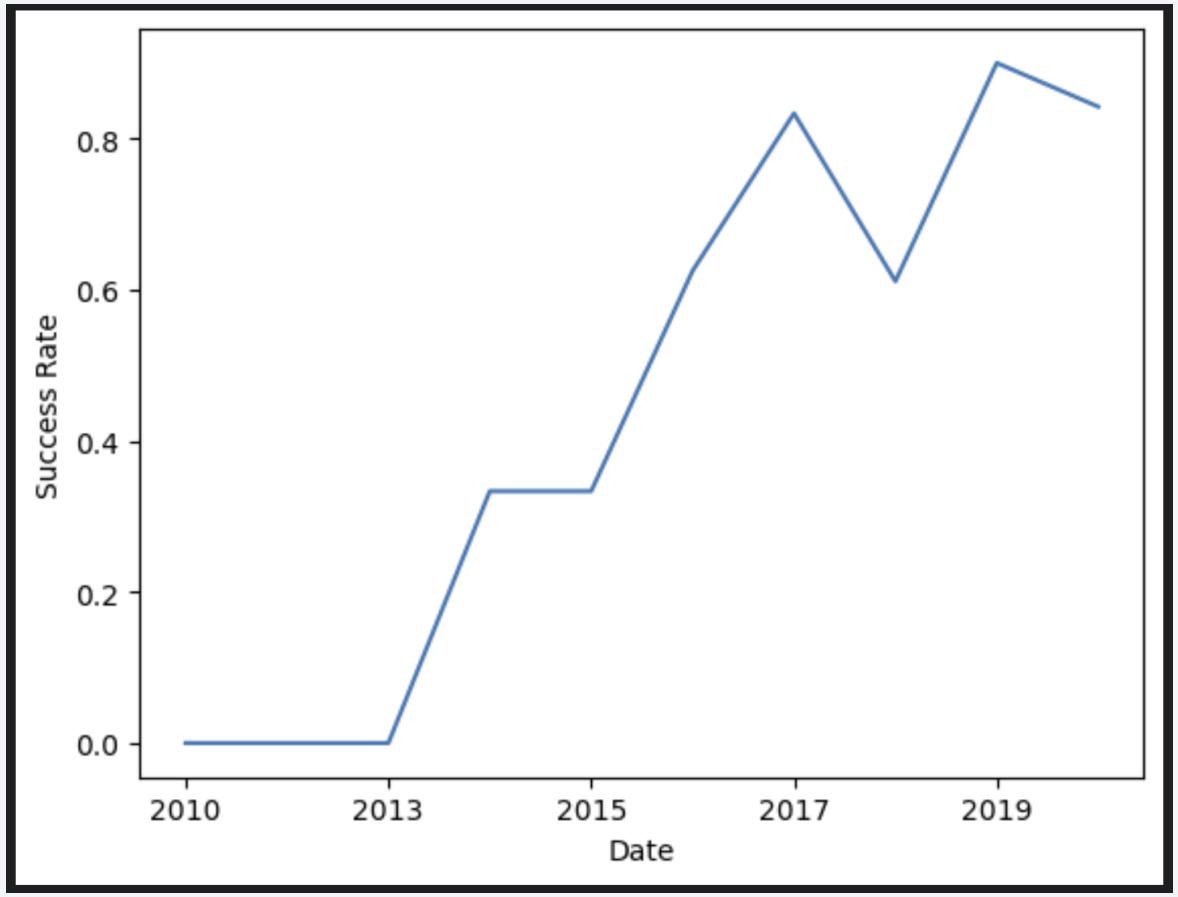
✓ **LEO (Low Earth Orbit)** missions had a **wide range of payloads**, from light to heavy

✓ **GTO (Geostationary Transfer Orbit)** missions typically had **medium to heavy payloads**, indicating **higher energy requirements**

✓ **SSO (Sun-Synchronous Orbit) and Polar orbits** carried **lighter payloads**, suggesting their focus on **small satellite missions**

✓ Helps predict **future launch needs** based on orbit selection

Launch Success Yearly Trend



📈 **Key Insights**

✓ **Early years (2010-2015)** had **lower success rates**, indicating initial testing and failures

✓ **Post-2015**, success rates **steadily improved**, showing advancements in reusability and landing precision

✓ **Recent years (2018-Present)** show **near-perfect success rates**, proving system maturity

✓ Helps **forecast SpaceX's future reliability trends**

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

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

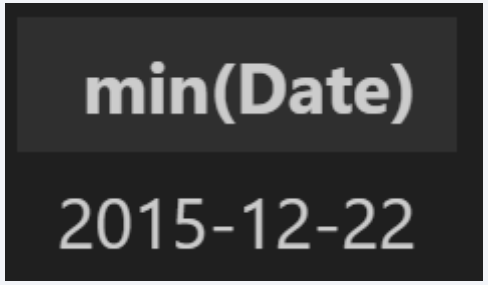
Total Payload Mass

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date



Successful Drone Ship Landing with Payload between 4000 and 6000

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

Total Number of Successful and Failure Mission Outcomes

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

Boosters Carried Maximum Payload

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

2015 Launch Records

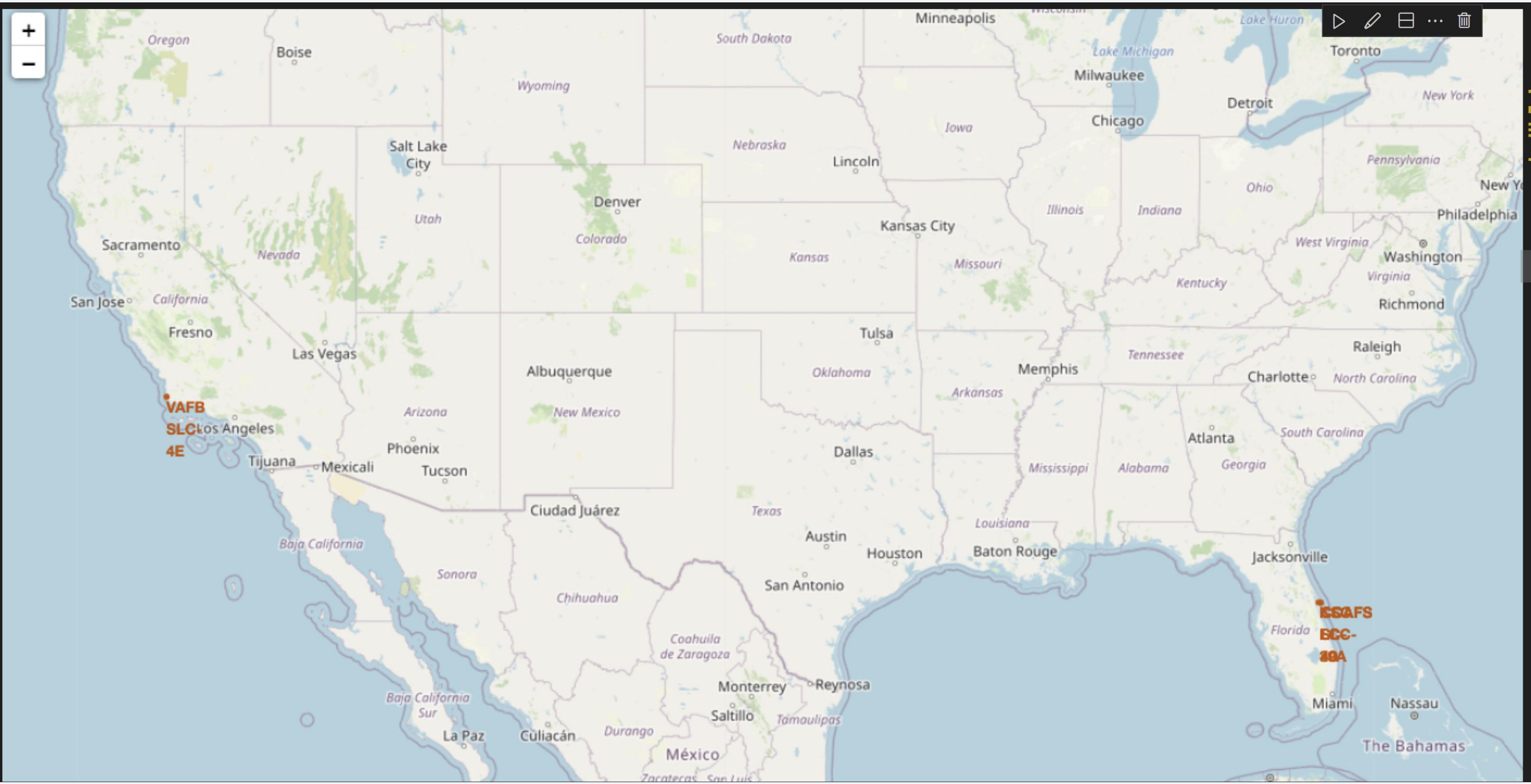
month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

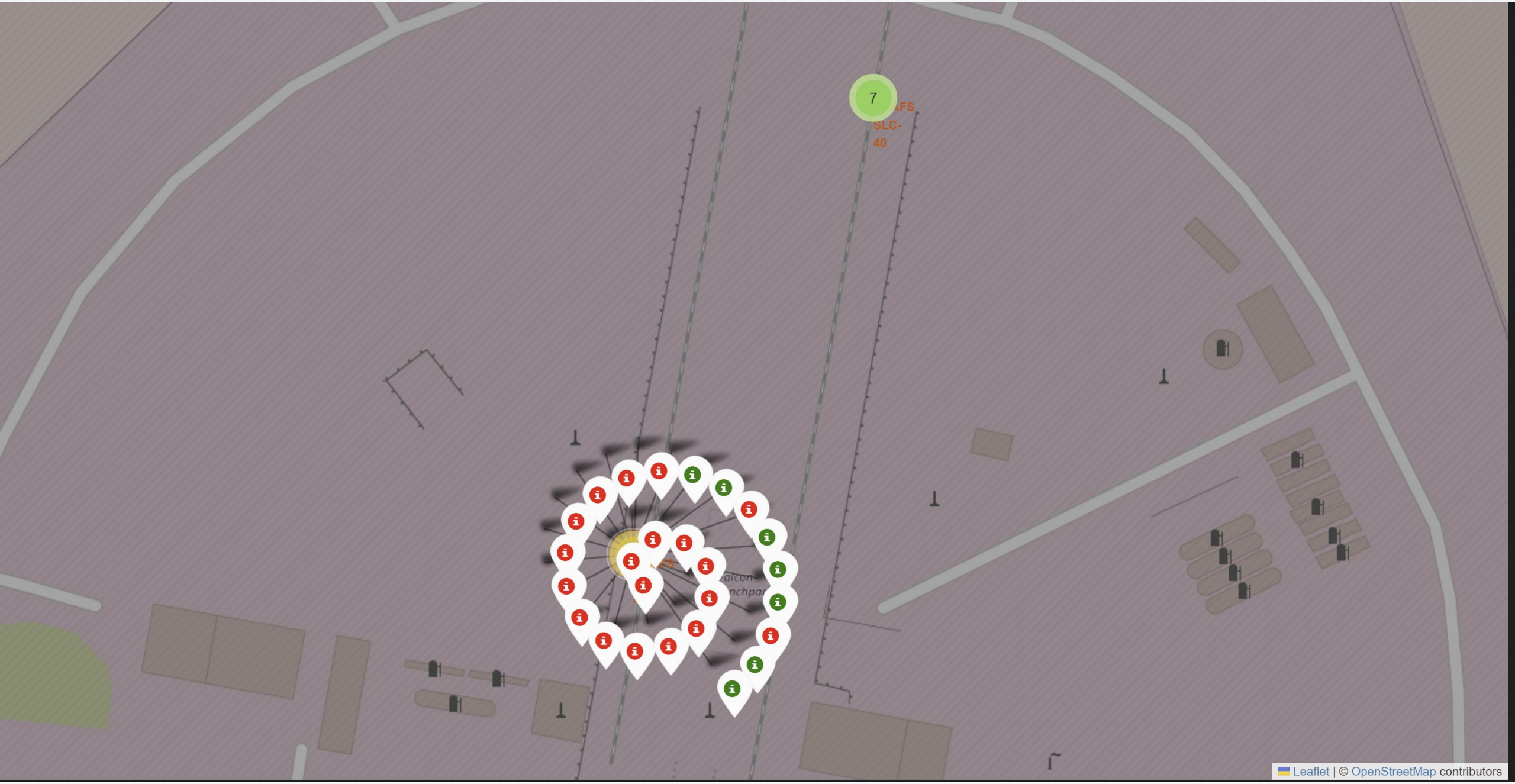
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

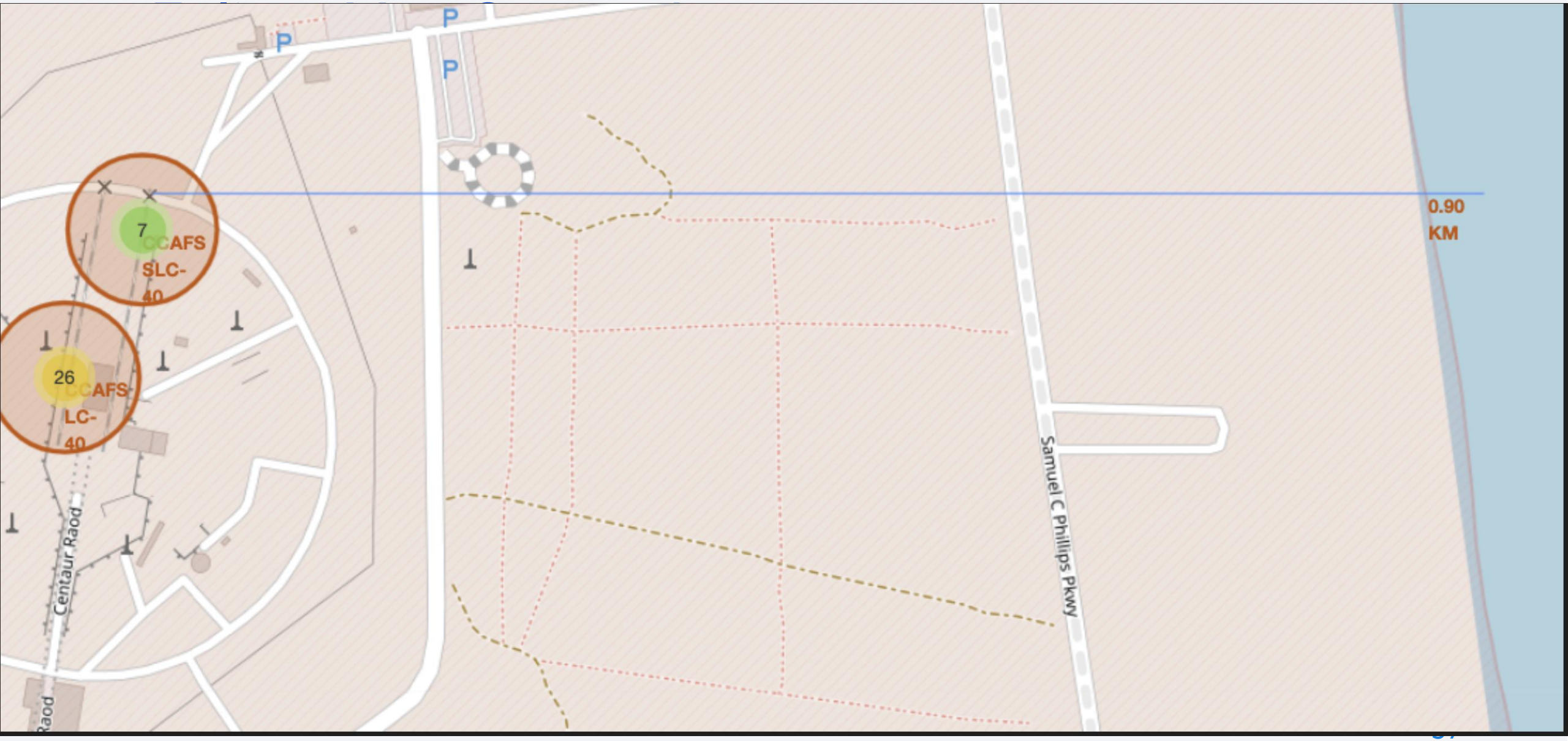
Landing_Outcome	count(*)
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

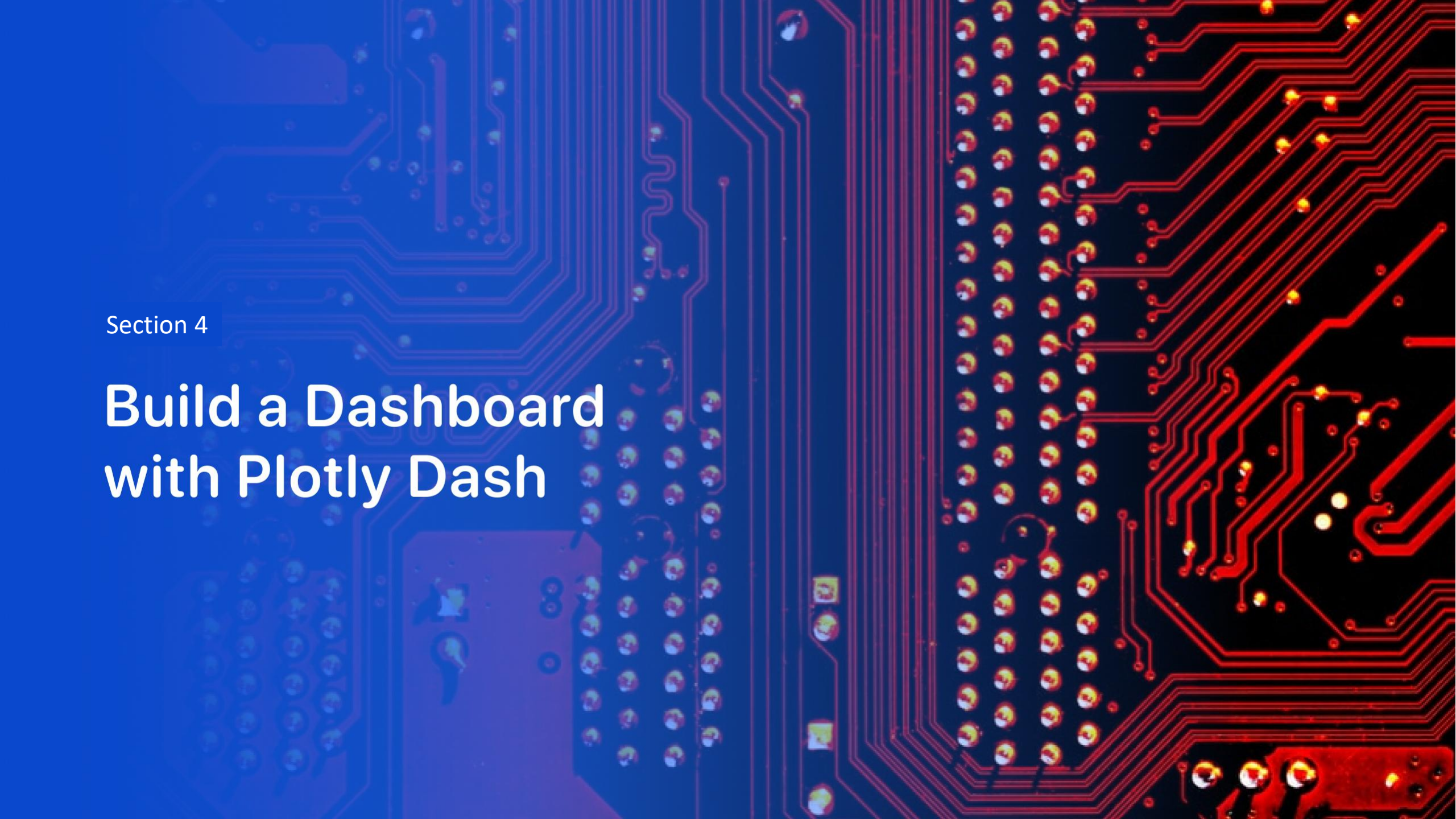
Section 3

Launch Sites Proximities Analysis


Minneapolis
Lake Huron
Toronto
Oregon
Boise
South Dakota
Lake Michigan
Milwaukee
Wyoming
Detroit
New York
Chicago
Iowa
Nebraska
Salt Lake City
Lincoln
Pennsylvania
Ohio
Denver
Utah
Illinois
Indiana
Philadelphia
Kansas City
Colorado
Sacramento
Nevada
Kansas
Missouri
West Virginia
Washington
Kentucky
Virginia
San Jose
California
Richmond
Fresno
Tulsa
Las Vegas
Tennessee
Raleigh
Albuquerque
Oklahoma
Memphis
Charlotte
North Carolina
Arkansas
VAFB
SLC
4E
Arizona
New Mexico
Los Angeles
Atlanta
South Carolina
Phoenix
Dallas
Tijuana
Mexicali
Tucson
Mississippi
Alabama
Georgia
Ciudad Juárez
Texas
Louisiana
Austin
Baton Rouge
Jacksonville
Baja California
Houston
Sonora
San Antonio
Chihuahua
Florida
Coahuila de Zaragoza
Monterrey
Reynosa
Miami
Nassau
Baja California Sur
Saltillo
Tamaulipas
La Paz
Culiacán
Durango
The Bahamas
México


7
AFS
SLC-
40
Leaflet | © OpenStreetMap contributors


P
P
P
7
CCAFS
SLC-
40
26
CCAFS
LC-
40
0.90
KM
Samuel C Phillips Pkwy
Centaur Raod

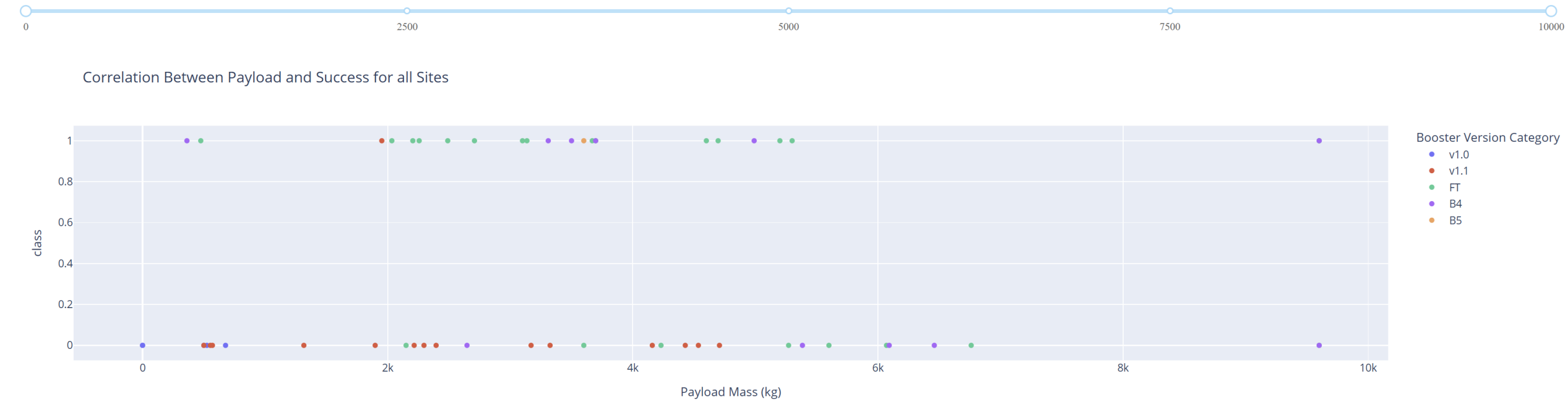

Section 4
Build a Dashboard with Plotly Dash

SpaceX Launch Records Dashboard

All Sites

Total Success Launches by Site

Payload range (Kg):



SpaceX Launch Records Dashboard

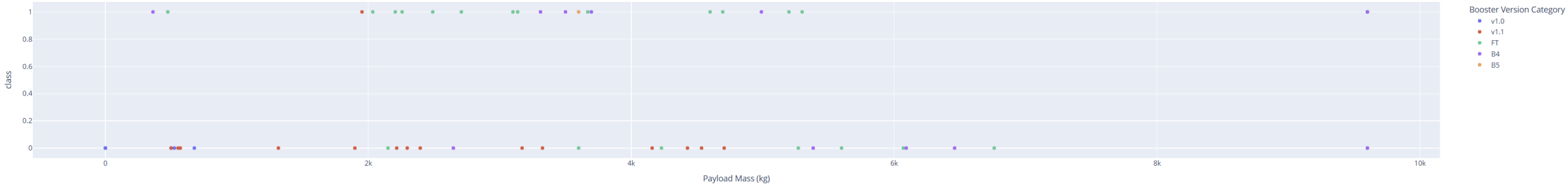
CCAFS LC-40

Total Success Launches by Site

Payload range (Kg):

0 2500 5000 7500 10000

Correlation Between Payload and Success for all Sites



Section 5

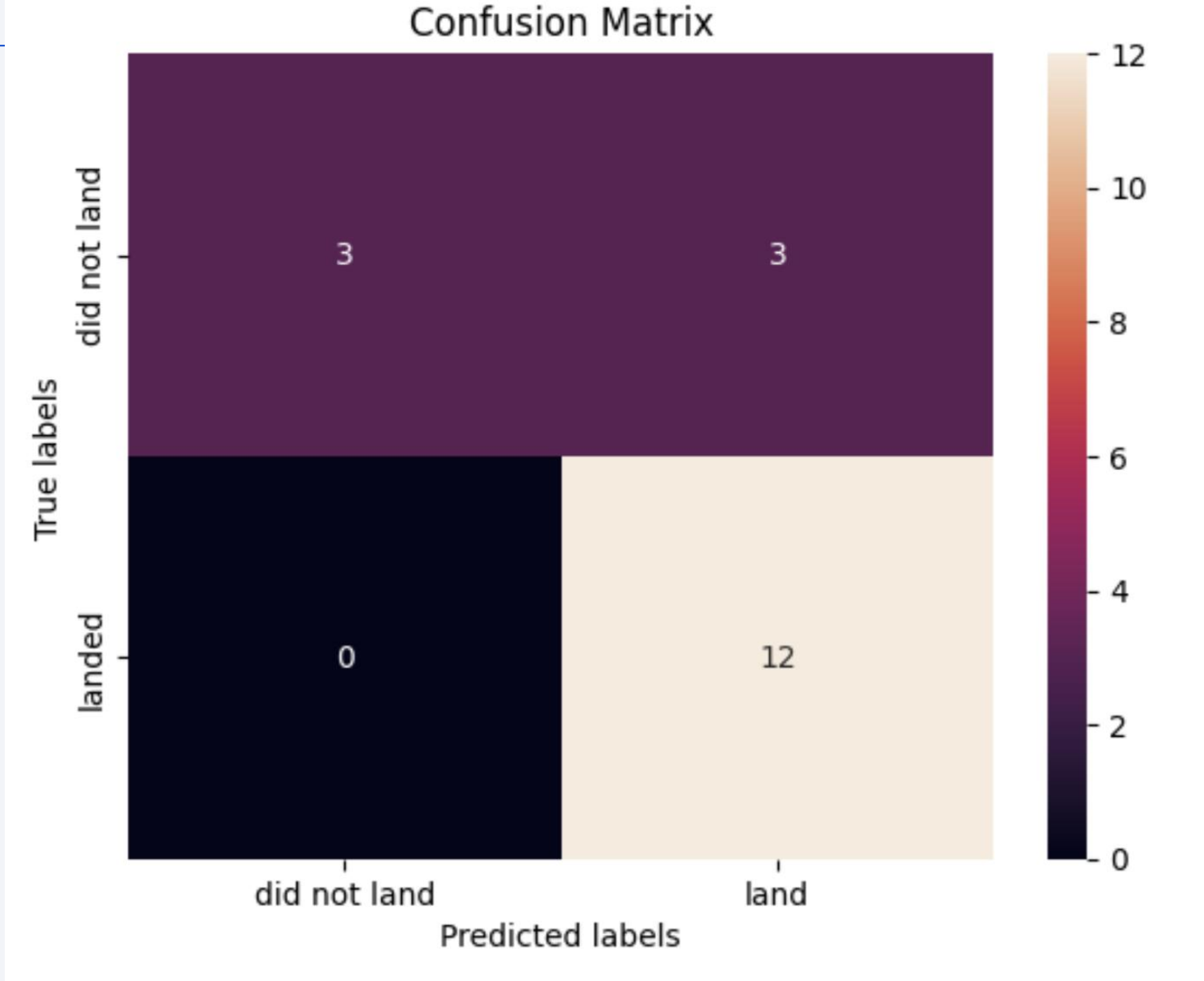
Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix



Conclusions

🛰️ **Project Highlights**

•Successfully developed a machine learning pipeline to **predict Falcon 9 first stage landings**

•Integrated **interactive visualizations and dashboards** for stakeholder exploration

•Conducted **deep exploratory analysis** to identify key features influencing landing success

🔍 **Key Findings**

•**Launch site, payload mass, booster version**, and **orbit type** are strong predictors of landing outcome

•**Landing success rate** has increased significantly since 2013, indicating improved tech and operational strategies

•**KSC LC-39A** emerged as the most reliable launch site for heavy payloads

•**LEO missions** are the most successful; **GTO missions** pose higher challenges for reusability

🤖 **Model Performance**

•Tested multiple ML models (Logistic Regression, SVM, Decision Tree, KNN)

•**Best performing model**: [Insert model name]

- Accuracy: 87%
- Evaluated with **Grid Search and Confusion Matrix**

🚀 **Next Steps**

•Incorporate **real-time launch data** for continuous model retraining

•Expand model to predict **landing method** (ASDS vs. RTLS)

•Apply insights to **Space Y** strategy for competitive launch planning

Appendix

```python
from math import sin, cos, sqrt, atan2, radians

def calculate_distance(lat1, lon1, lat2, lon2):
    # approximate radius of earth in km
    R = 6373.0

    lat1 = radians(lat1)
    lon1 = radians(lon1)
    lat2 = radians(lat2)
    lon2 = radians(lon2)

    dlon = lon2 - lon1
    dlat = lat2 - lat1

    a = sin(dlat / 2)**2 + cos(lat1) * cos(lat2) * sin(dlon / 2)**2
    c = 2 * atan2(sqrt(a), sqrt(1 - a))

    distance = R * c
    return distance
```

Python

Thank you!