IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tasneem Makhlouf
08.11.2025

Outline

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

Executive Summary

- Summary of methodologies:

Collect SpaceX launch data, clean it, and prepare it for analysis. We explored the data using visualizations, created interactive maps and dashboards, and applied machine learning models to predict whether the Falcon 9 first stage would land successfully.

- Summary of all results:

The analysis showed that landing success is influenced by launch site, booster version, and payload mass. Among the machine learning models tested, the **Decision Tree Classifier** gave the best performance. The visualizations helped us clearly understand patterns in the data.

Introduction

- Project background and context:

SpaceX is working to reduce the cost of space travel by reusing rocket boosters. Predicting whether a booster will successfully land helps improve mission planning and reduce costs.

- Problems you want to find answers:

What factors affect the success of Falcon 9 booster landings?

Which launch sites have the highest landing success rates?

Can we accurately predict landing success using machine learning?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - How data was collected
- Perform data wrangling
  - How data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- The data was collected from publicly available SpaceX launch records and compiled into a structured dataset for analysis.
- **Practical Steps :**

  Data Source

  SpaceX API + Web Scraping.

  Data Extraction

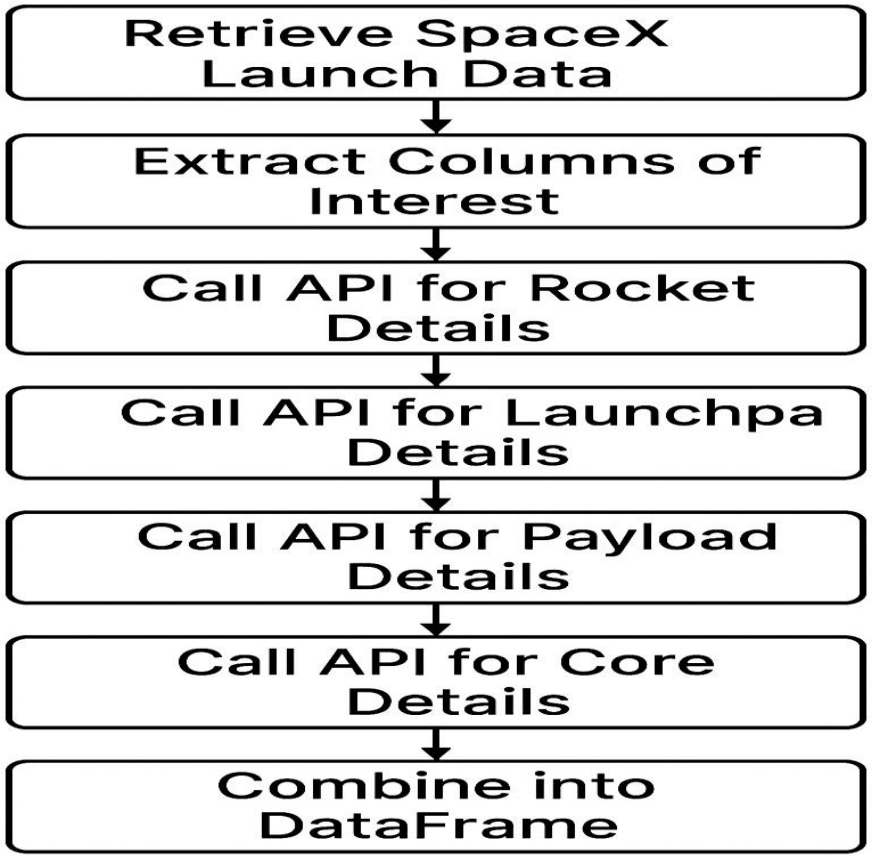
  Fetch data from API and scrape from web pages.

Data Collection – SpaceX API

Key Phrases:

- Retrieve Launch Data
- Extract Relevant Columns
- Filter Data
- API Calls for Details
- Combine Data
- Clean & Transform
- Final Dataset

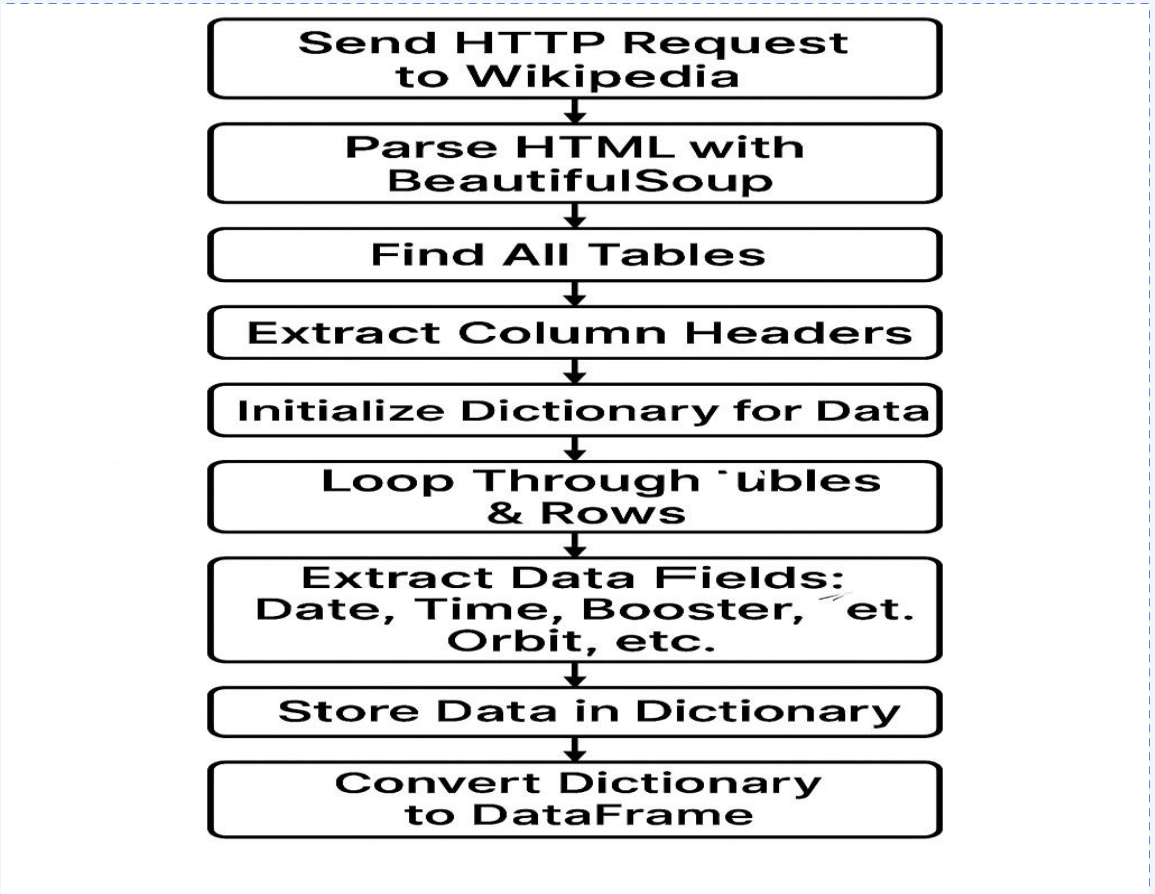
GitHub URL:[tasneemfaisal08/Data-Science-Capstone-Project]



Data Collection - Scraping

Key Phrases:

- Send HTTP Request
- Parse HTML Content
- Locate Tables
- Extract Column Headers
- Initialize Dictionary
- Loop Through Tables & Rows
- Extract Data Fields
- Store in Dictionary
- Convert to DataFrame



GitHub URL:[tasneemfaisal08/Data-Science-Capstone-Project]

Data Wrangling

Key Phrases for Data Wrangling Steps:

- **Load Dataset** :Read CSV file using pd.read_csv()
- **Inspect Data** :Use .head(), .dtypes, and .isnull().sum() to check structure and missing values
- **Handle Missing Values** :Identify and clean or impute missing data
- **Analyze Columns** :Apply value_counts() on LaunchSite, Orbit, and Outcome
- **Create Landing Class** :Define success/failure based on Outcome
- **Add New Feature** :Create Class column (binary: 1 for success, 0 for failure)
- **Check Statistics** :Compute mean success rate using .mean()

GitHub URL: tasneemfaisal08/Data-Science-Capstone-Project

EDA with Data Visualization

Charts Plotted and Why

- **Scatter Plots**
  - **Flight Number vs Launch Site**
    Why: To observe success/failure patterns across different launch sites.
  - **Flight Number vs Payload Mass**
    Why: To check if payload size affects landing success.
  - **Orbit vs Flight Number**
    Why: To analyze success rates for different orbit types.
  - **Orbit vs Payload Mass**
    Why: To see if orbit type correlates with payload size and success.
- **Bar Chart**
  - **Success Rate by Orbit**
    Why: To compare performance across orbit types.
- **Line Chart**
  - **Yearly Success Trend**
    Why: To visualize improvement in success rates over time.

GitHub URL: [tasneemfaisal08/Data-Science-Capstone-Project]

EDA with SQL

SQL EDA Summary

- **Get unique launch sites** → Identify all locations.
- **Count launches per site** → Compare site activity.
- **Filter by payload range** → Focus on medium payloads.
- **Retrieve successful landings** → Analyze success cases.
- **Group by booster version** → Check usage frequency.
- **Calculate success rate by orbit** → Compare orbit performance.

GitHub URL: tasneemfaisal08/Data-Science-Capstone-Project

Build an Interactive Map with Folium

Objects Added:

- **Map:** Base interactive map centered on launch sites.
- **Markers:** Show launch site locations and individual launches (green = success, red = failure).
- **Circles:** Highlight launch sites visually.
- **MarkerCluster:** Group multiple launches at the same site to reduce clutter.
- **MousePosition:** Display coordinates for exploration and distance calculation.
- **Distance Markers (DivIcon):** Show nearby points (coastline, highway, railway) and distances.
- **PolyLines:** Connect launch sites to nearby points to illustrate spatial relationships.

Purpose:

- Highlight launch sites and their outcomes.
- Show proximity to important features (coastline, cities, infrastructure).
- Reduce clutter and improve interactivity.
- Provide both visual and quantitative insights for analysis.

GitHub URL: tasneemfaisal08/Data-Science-Capstone-Project

Build a Dashboard with Plotly Dash

Plots & Graphs:

- **Pie Chart:** Shows total successful launches; updates per selected launch site.
- **Scatter Plot:** Payload vs. launch outcome, colored by Booster Version.

Interactions:

- **Dropdown:** Select “All Sites” or a specific launch site.
- **Range Slider:** Filter scatter plot by payload mass.

Purpose:

- Pie chart highlights top-performing sites.
- Scatter plot reveals correlation between payload and success.
- Interactions allow dynamic exploration and insights on site performance, payload ranges, and booster reliability.

GitHub URL: tasneemfaisal08/Data-Science-Capstone-Project

Predictive Analysis (Classification)

Model Development Process (Key Phrases):

- **Data Preparation:** Collected, cleaned, and standardized SpaceX launch data.
- **Feature Selection:** Chose relevant features (payload mass, launch site, booster version)
- **Train-Test Split:** Split data into training **(**80%**)** and test **(**20%**)** sets.
- **Model Selection:** Tested Logistic Regression, SVM, Decision Tree, KNN.
- **Hyperparameter Tuning:** Applied GridSearchCV with cross-validation (cv=10).
- **Model Evaluation:** Measured accuracy and plotted confusion matrices.
- **Best Model:** Decision Tree Classifier achieved the highest CV accuracy (87.5%).
- **Interpretation:** Evaluated predictions, analyzed misclassifications, and validated performance.

GitHub URL: tasneemfaisal08/Data-Science-Capstone-Project

Results

EDA Results :

- **Launch Success Trend:** Success rate increased over time, reaching ~90% by 2019.
- **Launch Sites:** Most launches from CCAFS LC-40, followed by KSC LC-39A and VAFB SLC-4E.
- **Payload Mass:** Wide range; heavier payloads did not strongly affect success.
- **Orbit Performance:** LEO, SSO, ES-L1 had high success rates; GTO had the lowest.
- **Booster Versions:** Falcon 9 dominated; Falcon 1 excluded from analysis.
- **Landing Outcomes:** Failures mostly in early missions and ocean landings.
- **Overall Success Rate:** Average success across all missions ~76%.

Predictive analysis results:

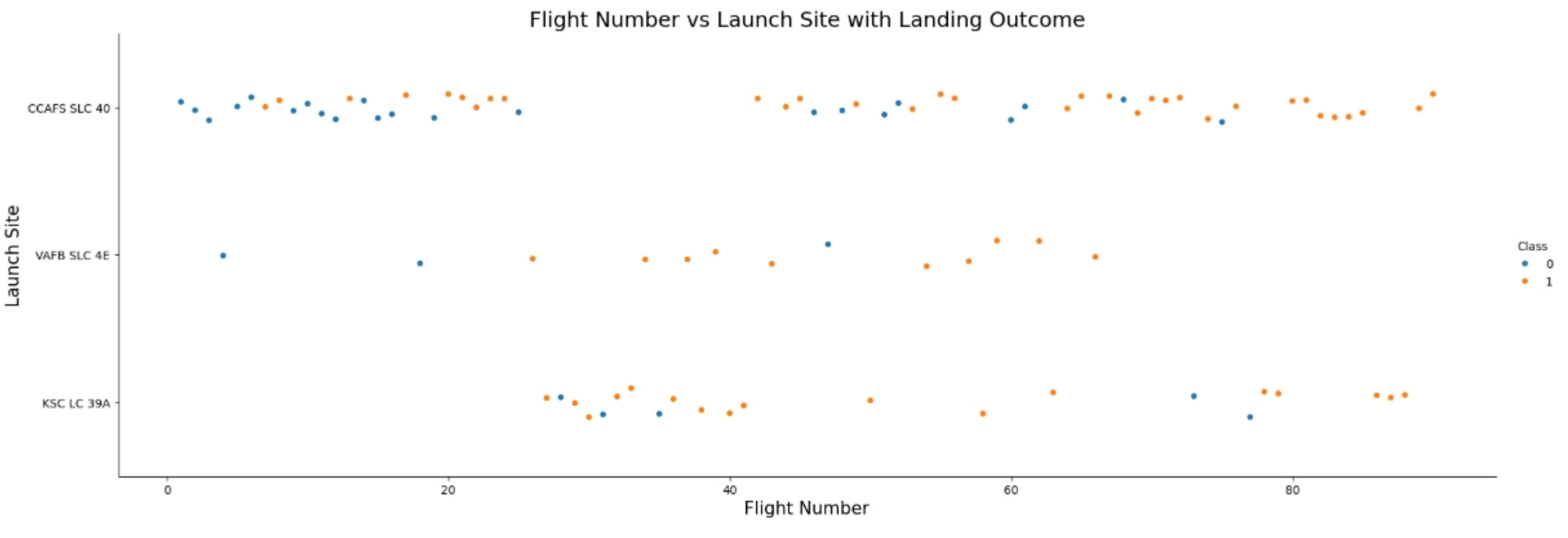
- Tested Logistic Regression, SVM, Decision Tree, KNN.
- Decision Tree had the highest CV accuracy **(**87.5%**)**.
- All models had similar test accuracy **(**~83**%)**.
- Confusion matrix shows most predictions correct, few misclassifications.
- Model can predict Falcon 9 booster landing success reliably.

Section 2

Insights drawn from EDA

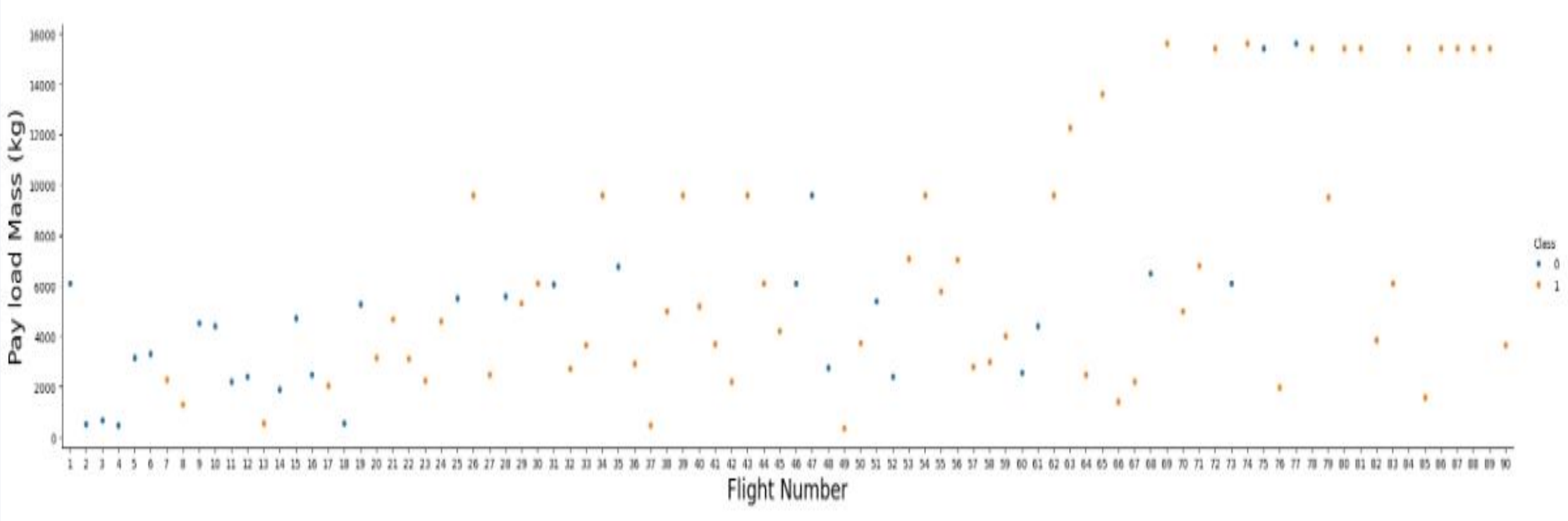
Flight Number vs. Launch Site

Each orange dot indicates a successful landing, and each blue dot indicates a failure, showing that success rates vary by launch site and flight number.



Payload vs. Launch Site

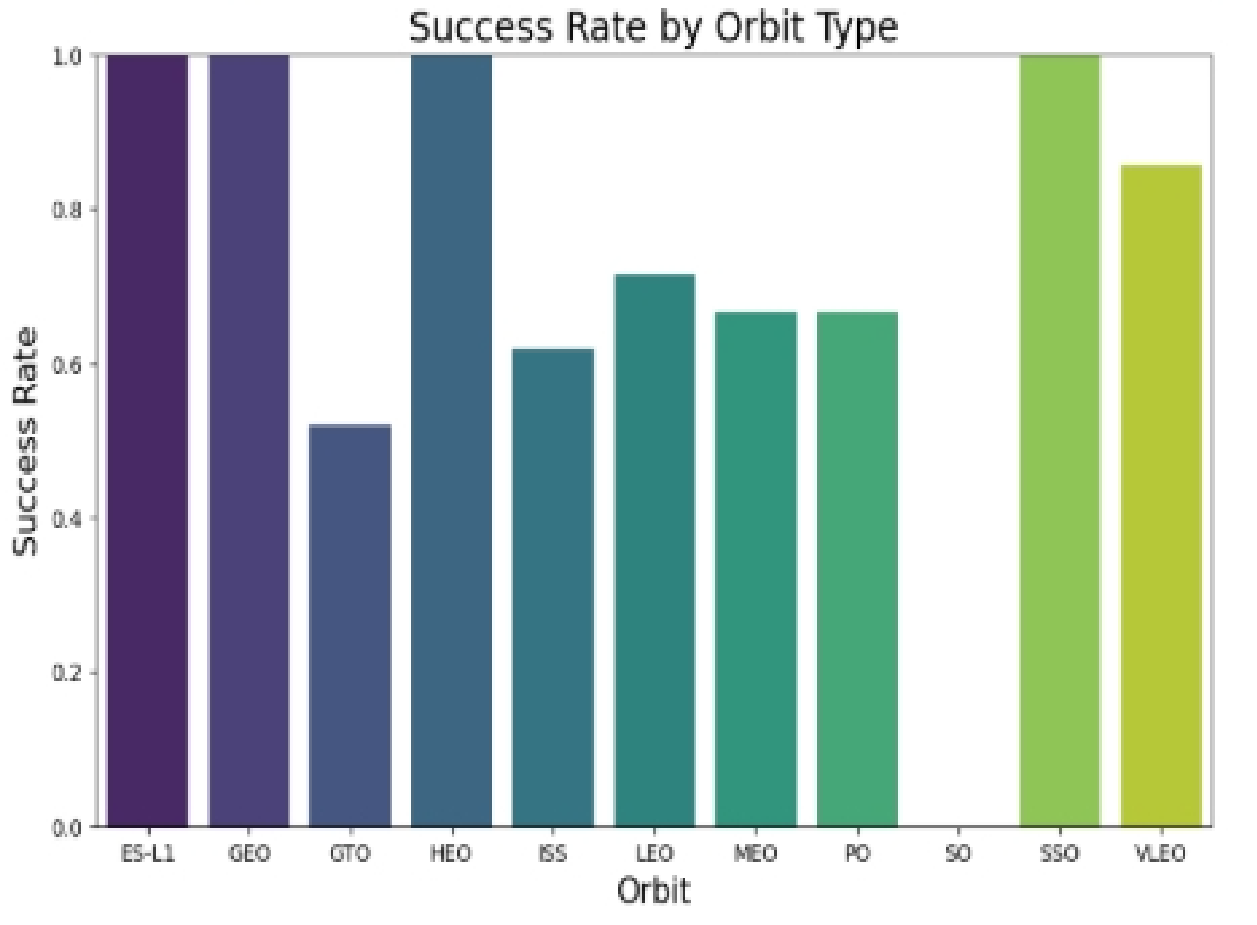
Each orange dot represents a successful landing and each blue dot a failure, illustrating how landing success varies with payload mass and flight number.



Success Rate vs. Orbit Type

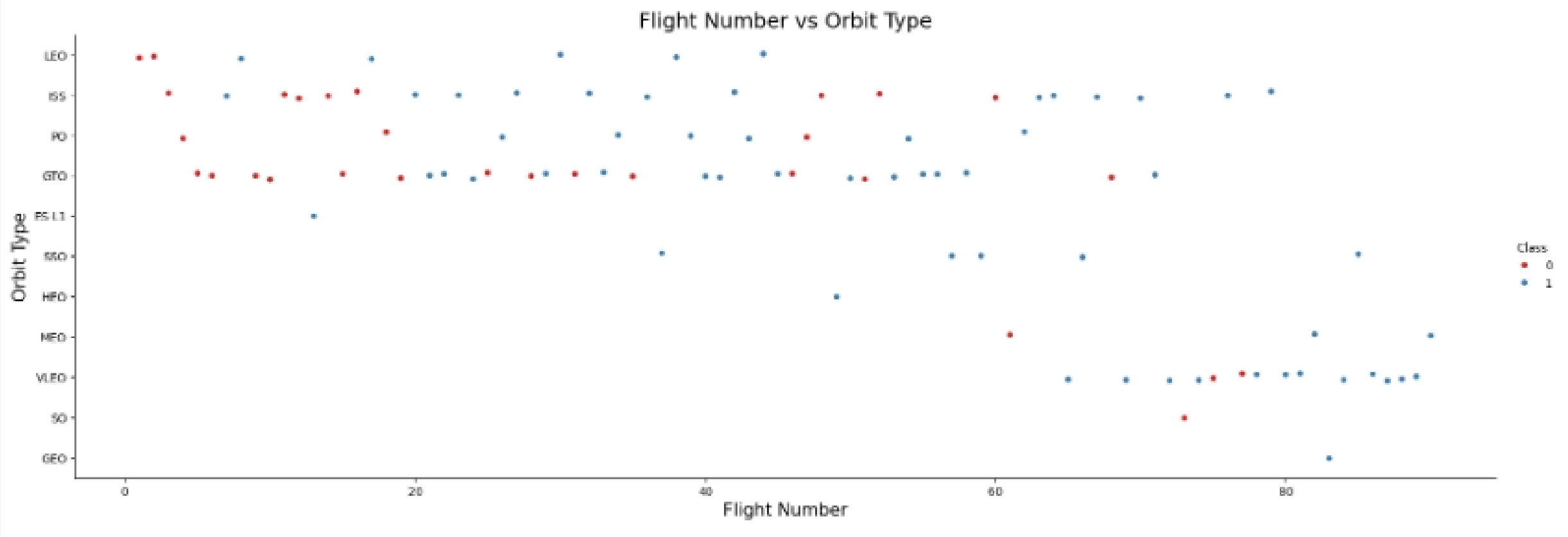
This bar chart shows **Success Rate vs Orbit type**. Each bar represents the success rate for a specific orbit:

- Orbits like **ES-L1, HEO, and SSO** have a success rate close to **1.0 (100%)**.
- Orbits like **GTO** have a much lower success rate (around 0.5).
- Others such as **LEO, MEO, PO** are in the mid-range (around 0.6–0.7).



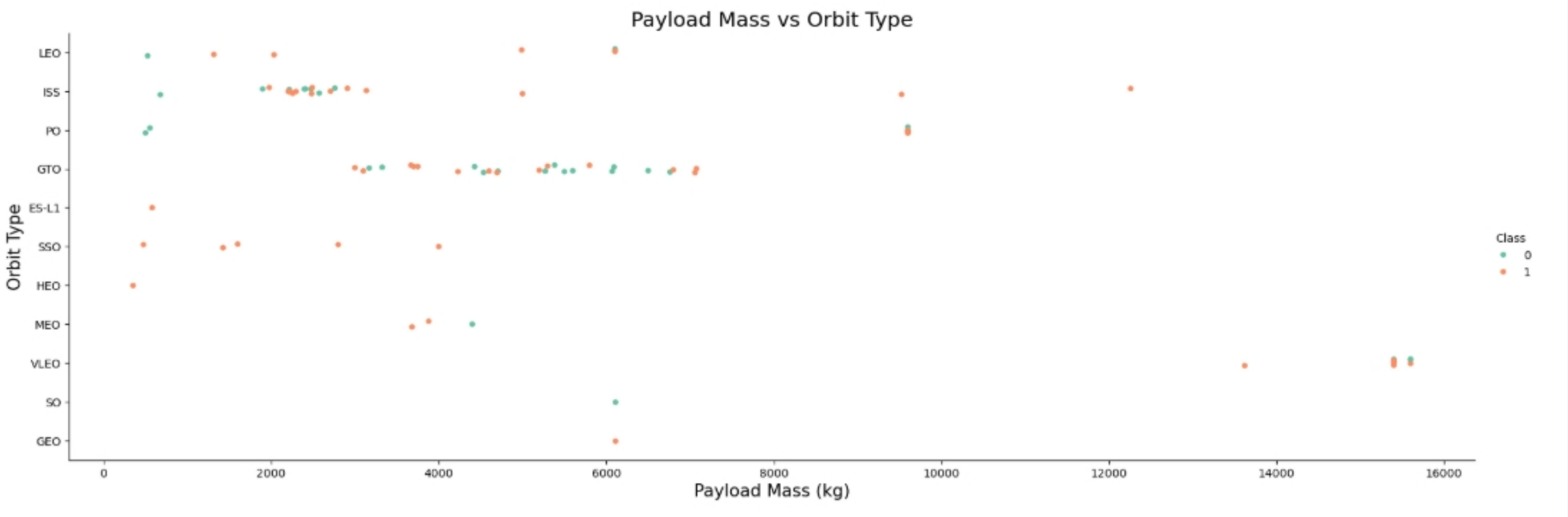
Flight Number vs. Orbit Type

Each blue dot represents a successful landing and each red dot a failure, showing how landing success varies across different orbit types and flight numbers.



Payload vs. Orbit Type

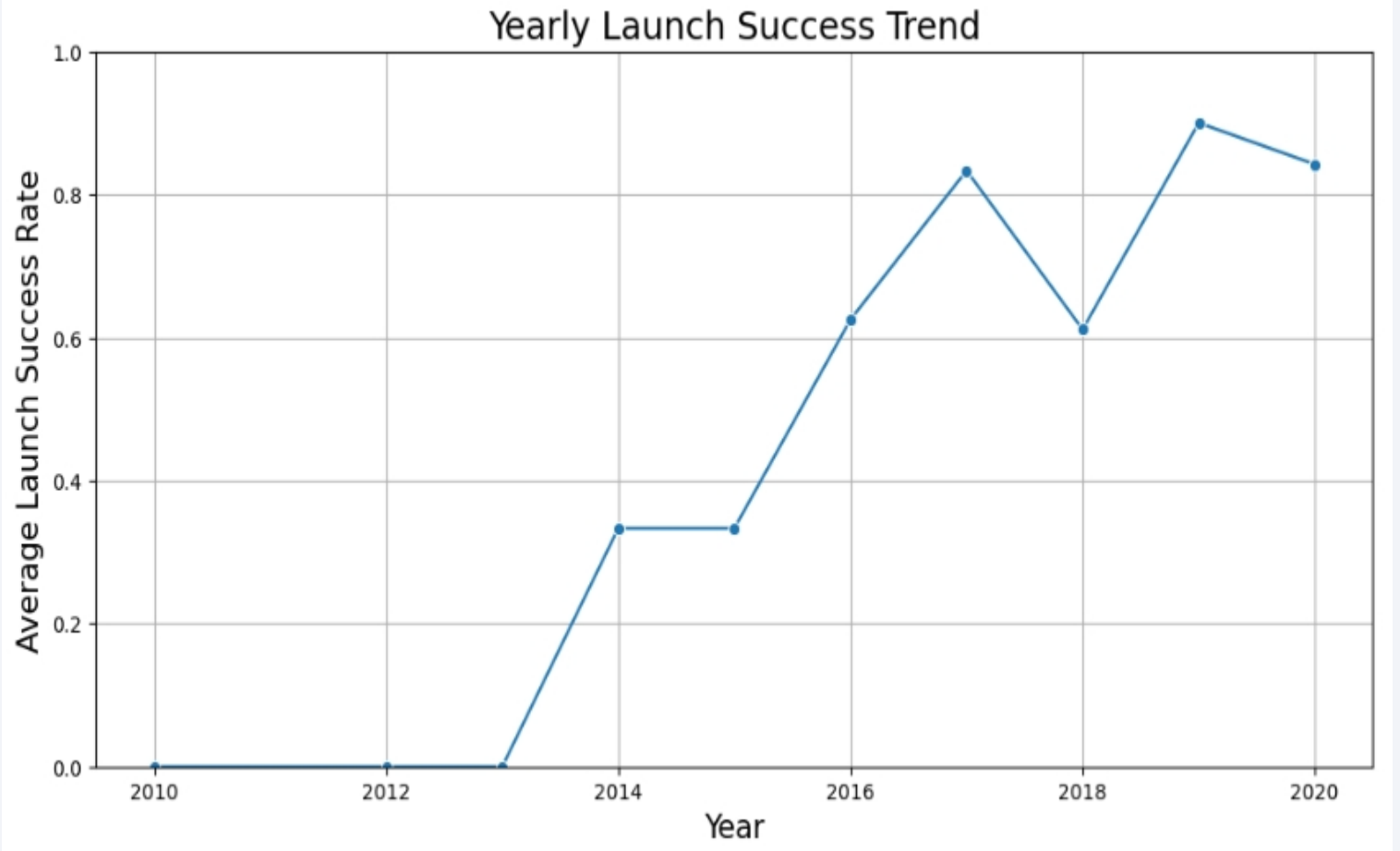
Each orange dot represents a successful landing and each green dot a failure, showing how landing success varies across different orbit types and payload masses.



Launch Success Yearly Trend

This line chart shows **Yearly Launch Success Trend** from 2010 to 2020:

- The **Average Launch Success Rate** starts near **0** in 2010–2013.
- It rises sharply around **2014**, then continues to increase, peaking near **0.9** in **2019**, with a slight dip in 2020.



All Launch Site Names

The query returns all distinct launch sites used in the space missions, which are CCAFS LC-40, VAFB SLC-4E, KSC LC-39A,and CCAFS SLC-40

Display the names of the unique launch sites in the space mission

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

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

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

Launch Site Names Begin with 'CCA'

Display 5 records where launch sites begin with the string 'CCA'

[16]:
```
%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
```

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

[16]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_O
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (pa
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 (pa
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No

Total Payload Mass

Display the total payload mass carried by boosters launched by NASA (CRS)

[17]:
```
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass
FROM SPACEXTABLE
WHERE Customer = 'NASA (CRS)';
```

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

[17]:

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

[18]:
```
%%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass
FROM SPACEXTABLE
WHERE Booster_Version LIKE 'F9 v1.1%';
```

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

[18]:

Average_Payload_Mass
2534.6666666666665

First Successful Ground Landing Date

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

[19]:
```
%%sql SELECT MIN(Date) AS First_Successful_Landing
FROM SPACEXTABLE
WHERE Landing_Outcome LIKE 'Success (ground pad)%';
```

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

[19]:

First_Successful_Landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

[20]:
```
%%sql SELECT Booster_Version
FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (drone ship)'
  AND PAYLOAD_MASS__KG_ > 4000
  AND PAYLOAD_MASS__KG_ < 6000;
```

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

[20]:

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

21]:
```
%%sql SELECT Mission_Outcome, COUNT(*) AS Total
FROM SPACEXTABLE
GROUP BY Mission_Outcome;
```

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

21]:

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

Boosters Carried Maximum Payload

List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function

[23]:
```
%%sql
SELECT Booster_Version, PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
)
```

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

[23]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600

2015 Launch Records

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

```
%%sql SELECT
    CASE substr(Date, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month,Landing_Outcome,Booster_Version,Launch_Site
FROM SPACEXTABLE
WHERE substr(Date, 0, 5) = '2015'
  AND Landing_Outcome LIKE '%Drone%'
  AND Landing_Outcome LIKE '%Failure%';
```

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

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

[25]:
```
%%sql SELECT
    "Landing_Outcome",
    COUNT(*) AS Count
FROM SPACEXTABLE
WHERE Date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing_Outcome"
ORDER BY Count DESC;
```

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

[25]:

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

Global Launch Sites Overview

Explanation:

- The map shows all SpaceX launch sites in the US.
- Each launch site is represented by a text Marker with the site name.

Findings:

- All launch sites are located near the coast, mainly on the east and west coasts of the US.
- Sites are safely away from major cities and railways, highlighting safety considerations.

Launch Outcomes by Site

Explanation:

- The map shows every launch with color-coded outcomes:
- Green = Successful launch
- Red = Failed launch

Findings:

- Some sites, like KSC LC-39A, have mostly successful launches.
- CCAFS SLC-40 shows some failures, with multiple launches clustered at the same location.
- This gives a quick visual understanding of the success rate per site.

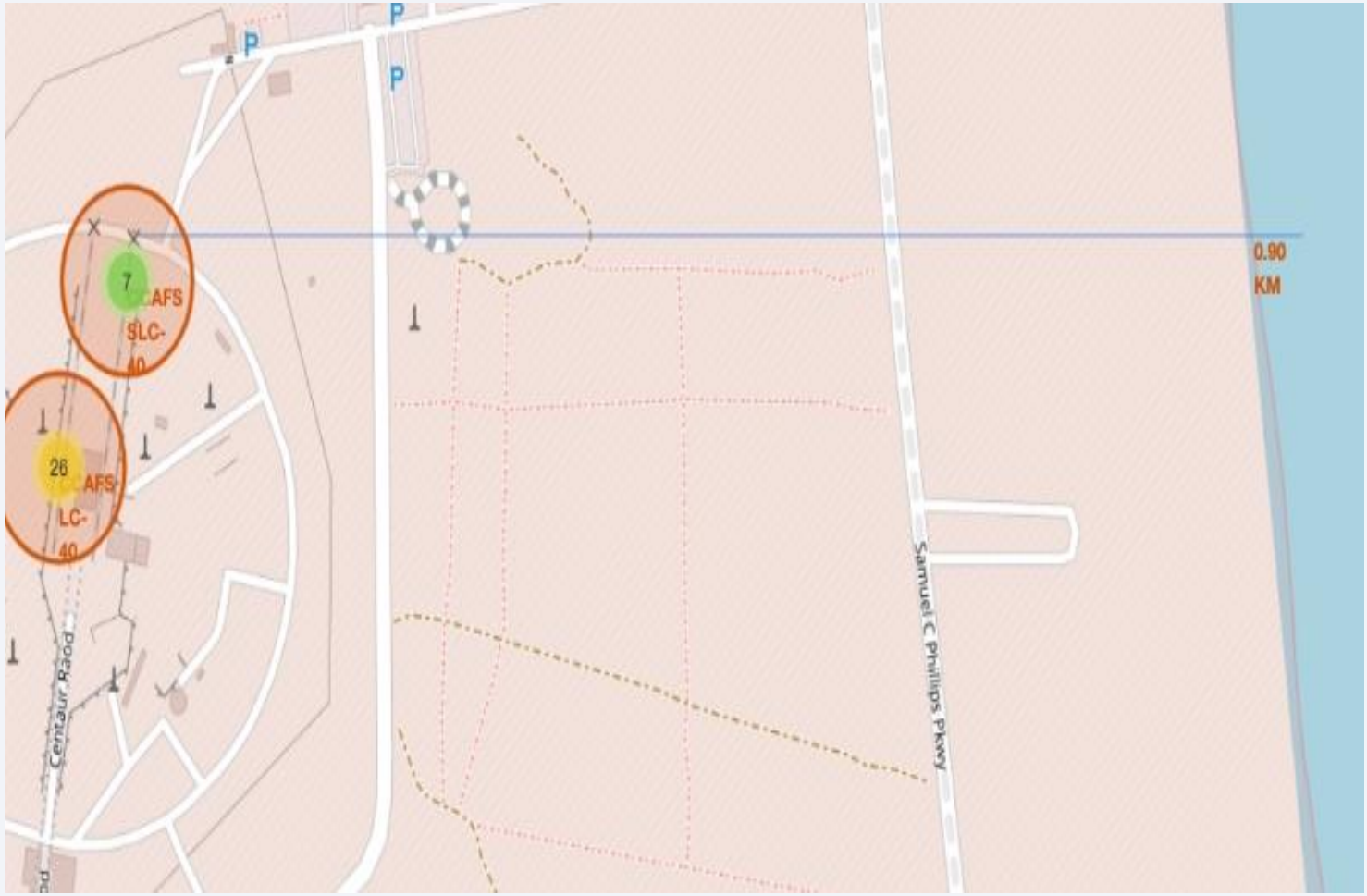
Launch Site Proximities and Distances

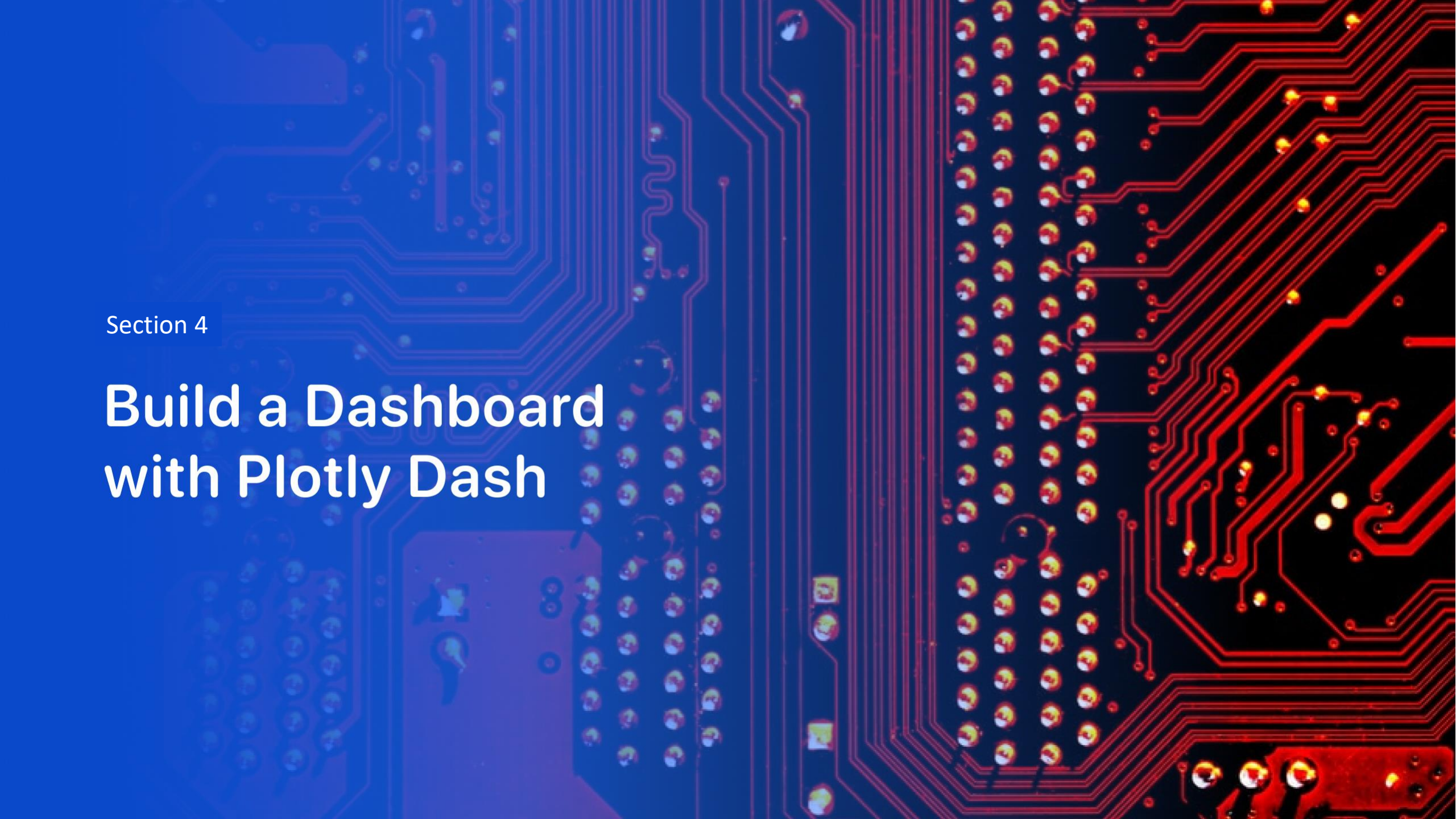
Explanation:
The map focuses on a selected launch site and shows:

- Nearby points of interest such as railway, highway, and coastline.
- PolyLines connecting the launch site to each nearby point.
- Markers displaying the distance (in km) to each point.

•Findings:

- Launch site is very close to the coastline to ensure a safe flight path over water.
- It is located away from railways and cities to minimize risks.
- The highway is close enough to facilitate logistics and transportation.



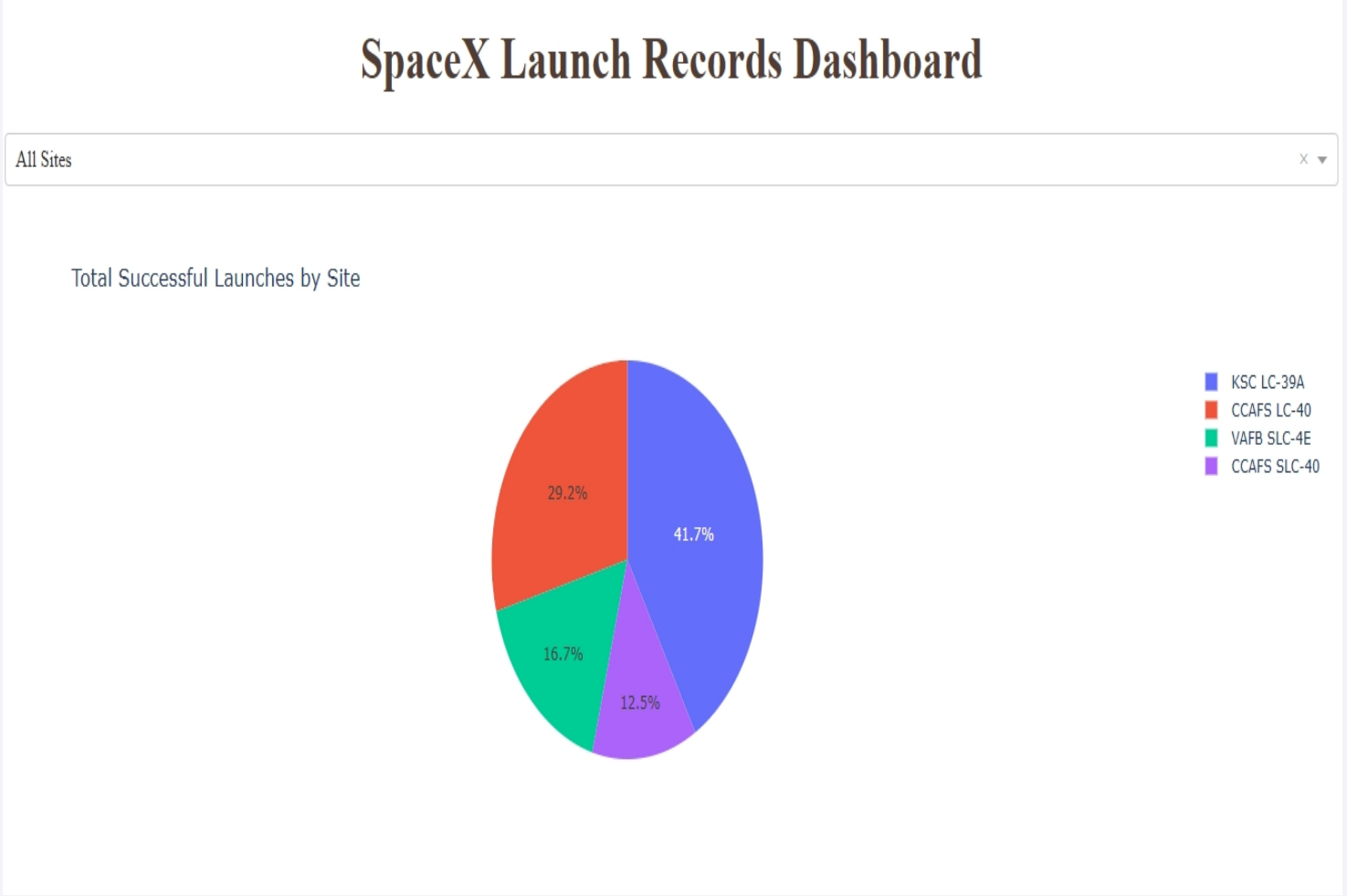

Section 4
Build a Dashboard with Plotly Dash

Total Successful Launches by Site

Key Observations:

- Shows which site has the highest number of successful launches.
- The distribution provides an overview of each site's performance relative to total launches.

Insight: KSC LC-39A has the largest slice in the chart, it indicates that this site has the highest number of successful launches.

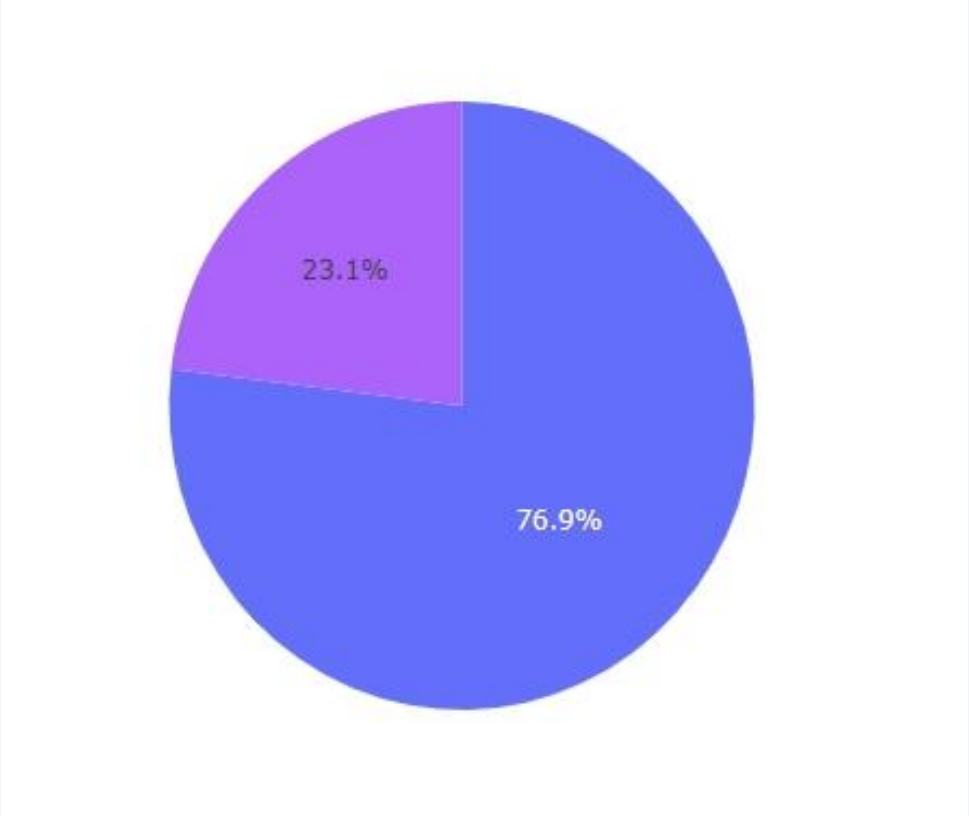


Success vs. Failure for Top Launch Site

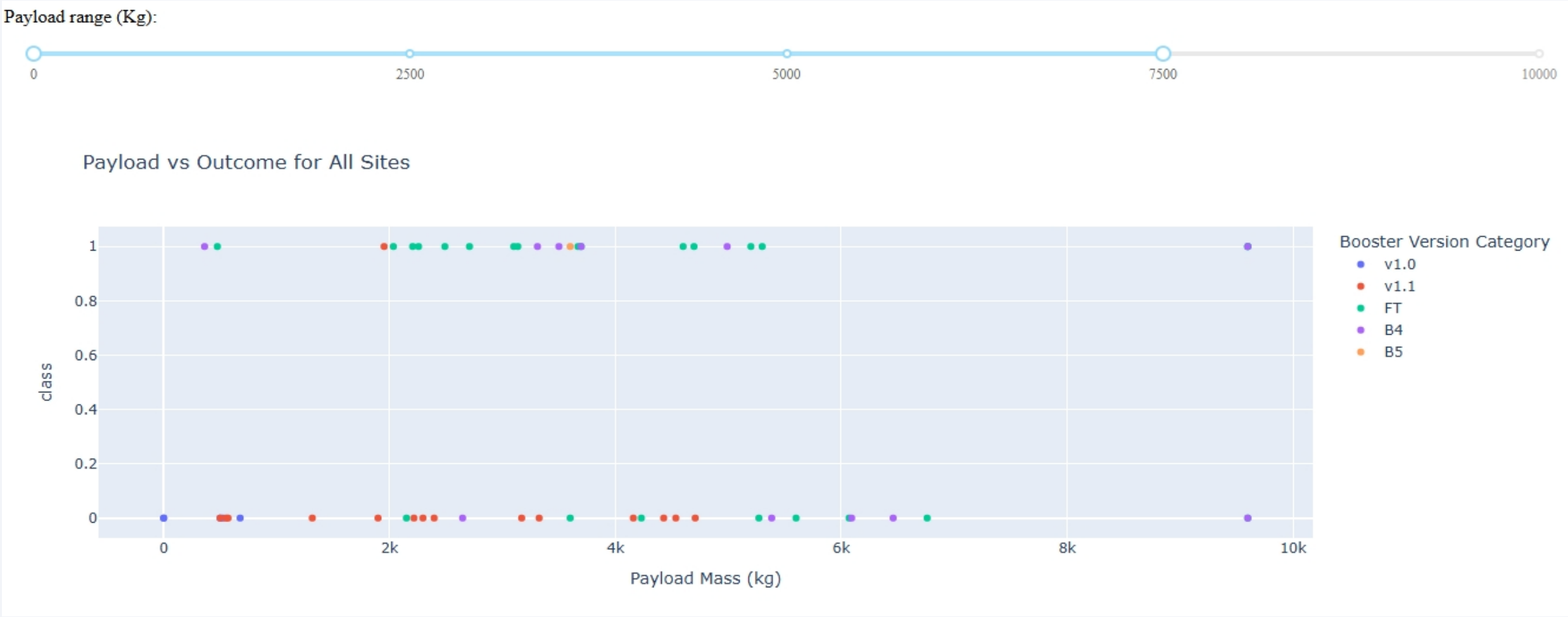
Key Observations:

- Compares the number of successful launches (76.9%) versus failed launches (23.1%).
- Helps evaluate the reliability of the site.

Insight: The majority of launches are successful, this indicates that the site is highly reliable.



Payload vs. Launch Outcome Scatter Plot



Payload vs. Launch Outcome Scatter Plot

Key Observations:

•Color indicates the Booster Version used for each launch.

•The Range Slider allows selecting different payload ranges to see patterns in success rates.

Insights:

•Certain payload ranges (lighter or medium weights) may have higher success rates.

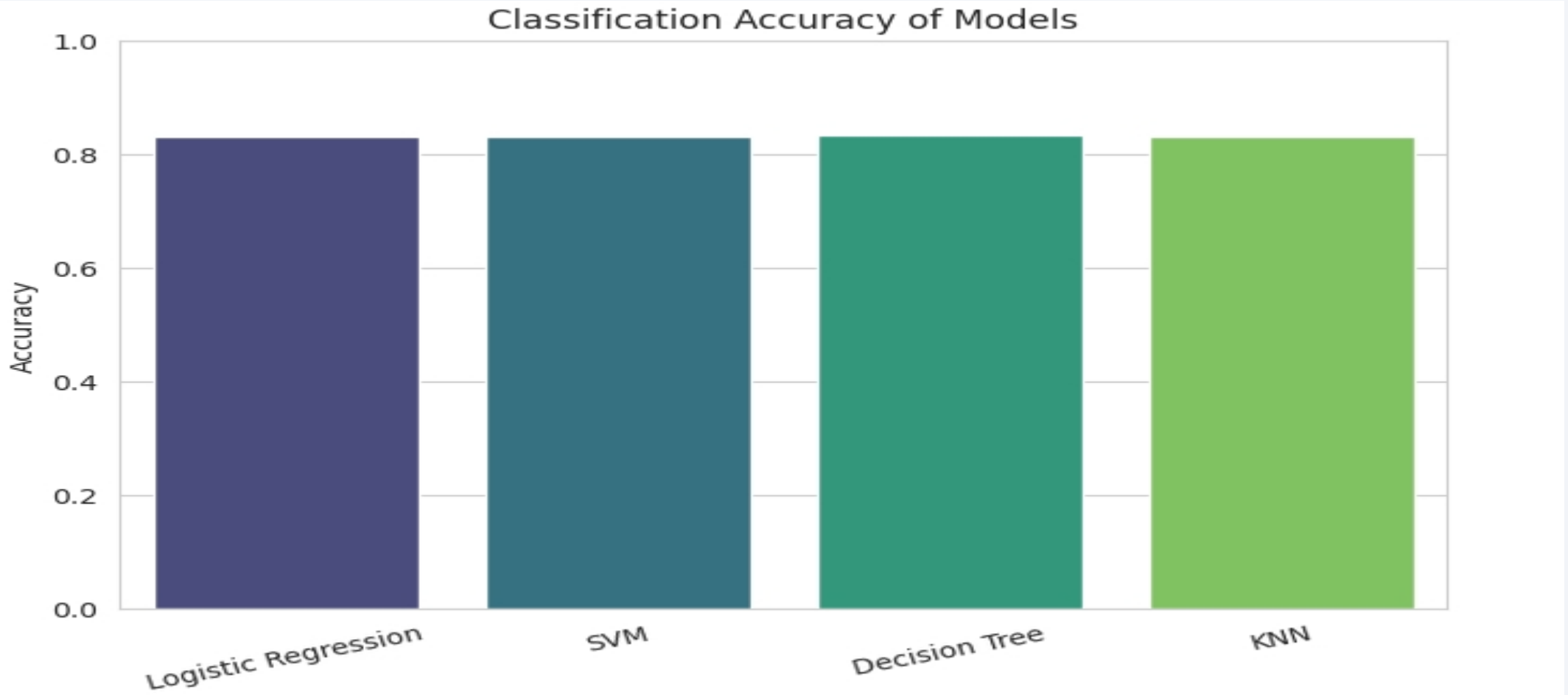
•Some Booster Versions (V1.1) may achieve higher success rates than others.

•This visualization shows how payload mass and booster type influence launch success.

Section 5

Predictive Analysis (Classification)

Classification Accuracy



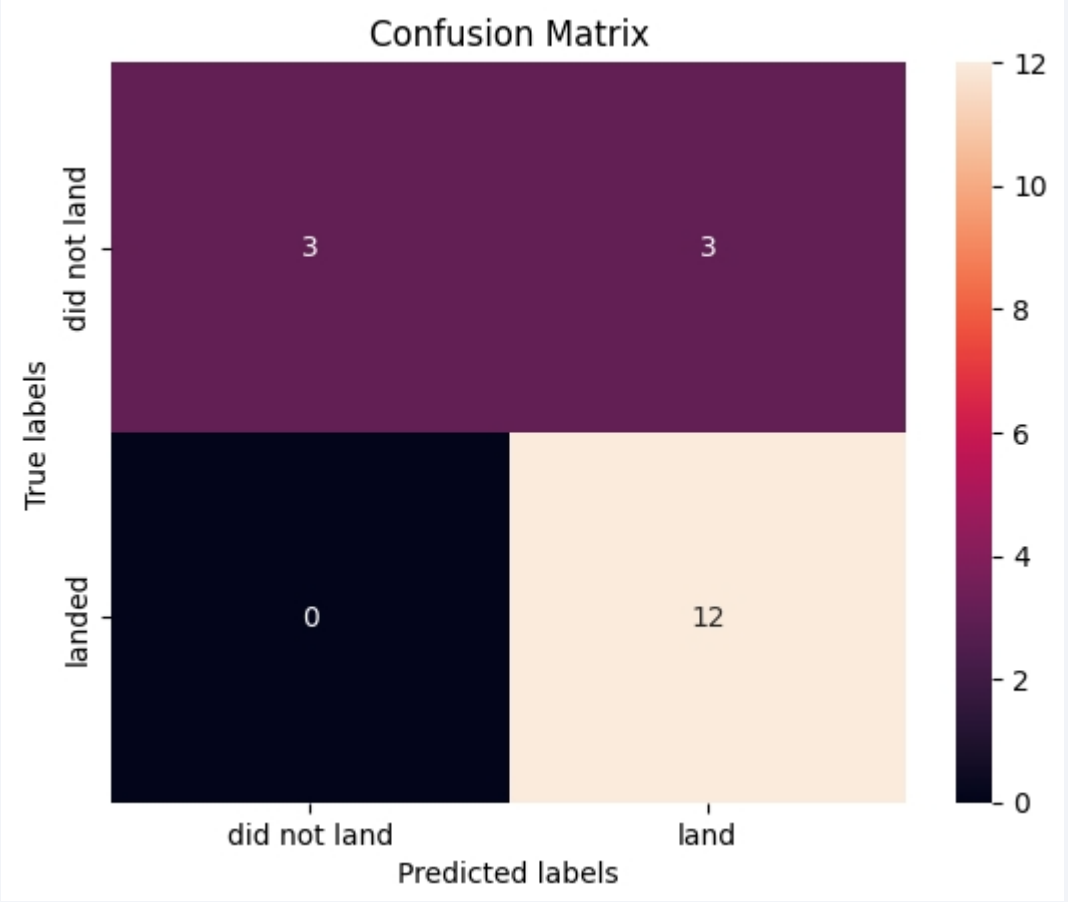

```
The model with the highest accuracy is Decision Tree with accuracy 0.83
```

Confusion Matrix

Examining the confusion matrix, we see that Decision Tree can distinguish between the different classes. We see that the problem is false positives.

Overview:

- True Postive - 12 (True label is landed, Predicted label is also landed)
- False Postive - 3 (True label is not landed, Predicted label is landed)



Conclusions

- SpaceX Falcon 9 booster landing success depends on **launch site, booster version, and payload mass**.
- Among the tested machine learning models, the **Decision Tree Classifier** had the **best performance** (CV accuracy 87.5%).
- All models performed similarly on **test data** (~83% accuracy).
- Confusion matrix analysis shows the model correctly predicts most landings and failures, with a few misclassifications.
- Interactive dashboards and visualizations make it easy to explore launch outcomes and identify patterns.
- Machine learning can **support mission planning** by predicting booster landing success, improving efficiency and reducing costs.

Appendix

GitHUB Project URL : tasneemfaisal08/Data-Science-Capstone-Project

Thank you!