IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Predicting Falcon 9 First Stage Landing Success

Presented By: Hammad Qaiser

Date: August 20, 2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Our data-driven analysis successfully predicted the landing success of the Falcon 9 rocket's first stage. By gathering data from the SpaceX API and web scraping, we identified key factors that influence a successful landing, including **payload mass**, **orbit type**, and **launch site**. Our exploratory analysis with visualizations and SQL queries confirmed these findings, revealing a clear upward trend in SpaceX's success rate and highlighting the most reliable launch locations. The **Logistic Regression Classifier** proved to be the most accurate predictive model, providing a reliable tool for our startup to make more informed and competitive bids. These insights offer a significant strategic advantage, allowing us to target market segments with a higher probability of success and thereby increase our competitiveness against SpaceX.

Introduction

Our startup is competing with **SpaceX** for rocket launch contracts. To make more informed and competitive bids, we need to predict the success of the first stage landing. A successful landing means a reusable rocket, which significantly lowers costs.

=

The central question is: **Will the first stage of the Falcon 9 rocket land successfully?** We aim to use data-driven insights and predictive models to answer this question.

Section 1

Methodology

Methodology

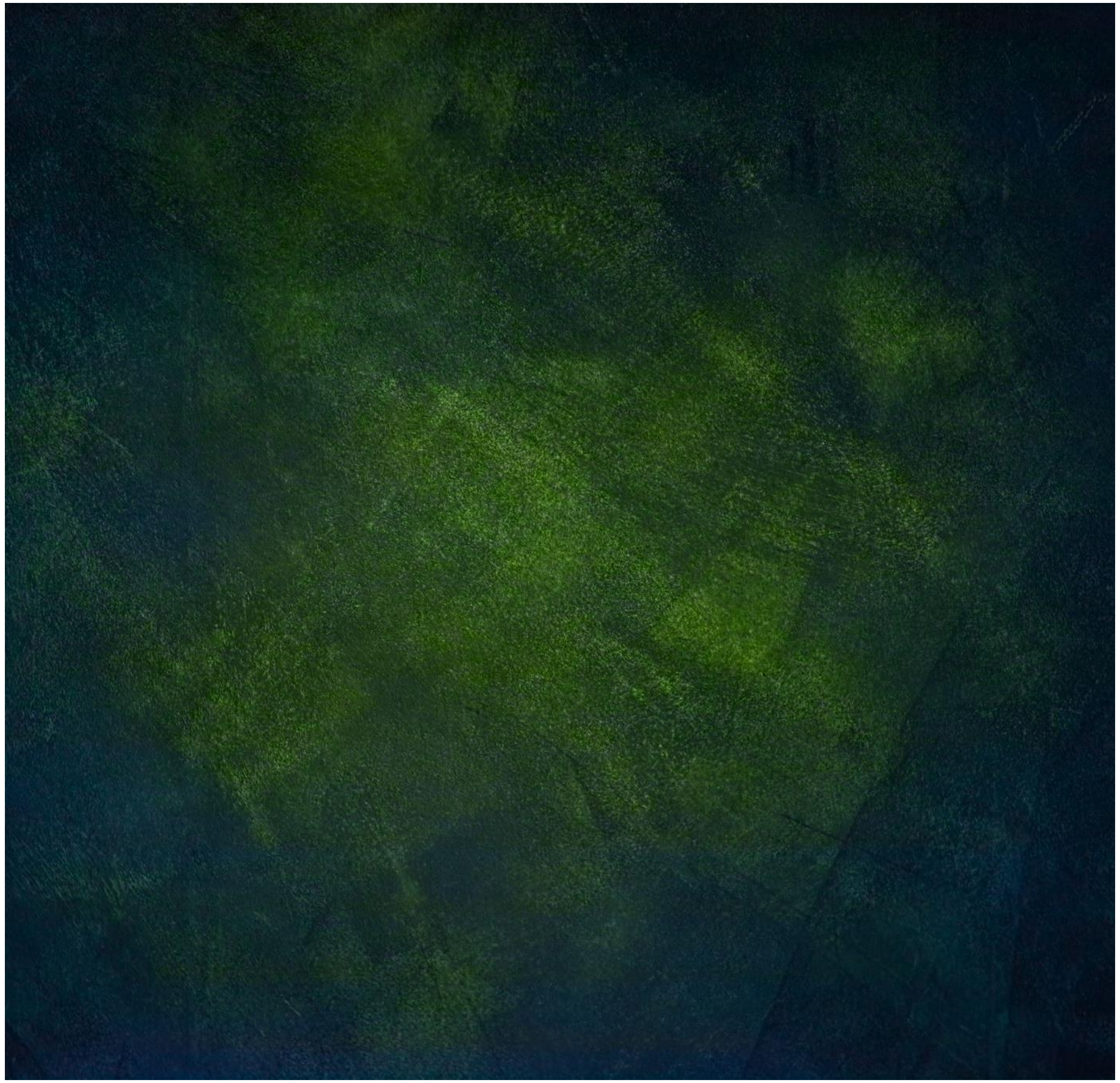
Data Collection: Gathered data from the SpaceX REST API and performed web scraping to get a complete dataset.

Data Wrangling: Cleaned and processed the data to handle missing values and prepare it for analysis.

Exploratory Data Analysis (EDA): Used visualization and SQL queries to explore the data and uncover patterns.

Interactive Visual Analytics: Built an interactive map with Folium and a dashboard with Plotly Dash for deeper analysis.

Predictive Analysis: Developed and evaluated classification models to predict landing success.

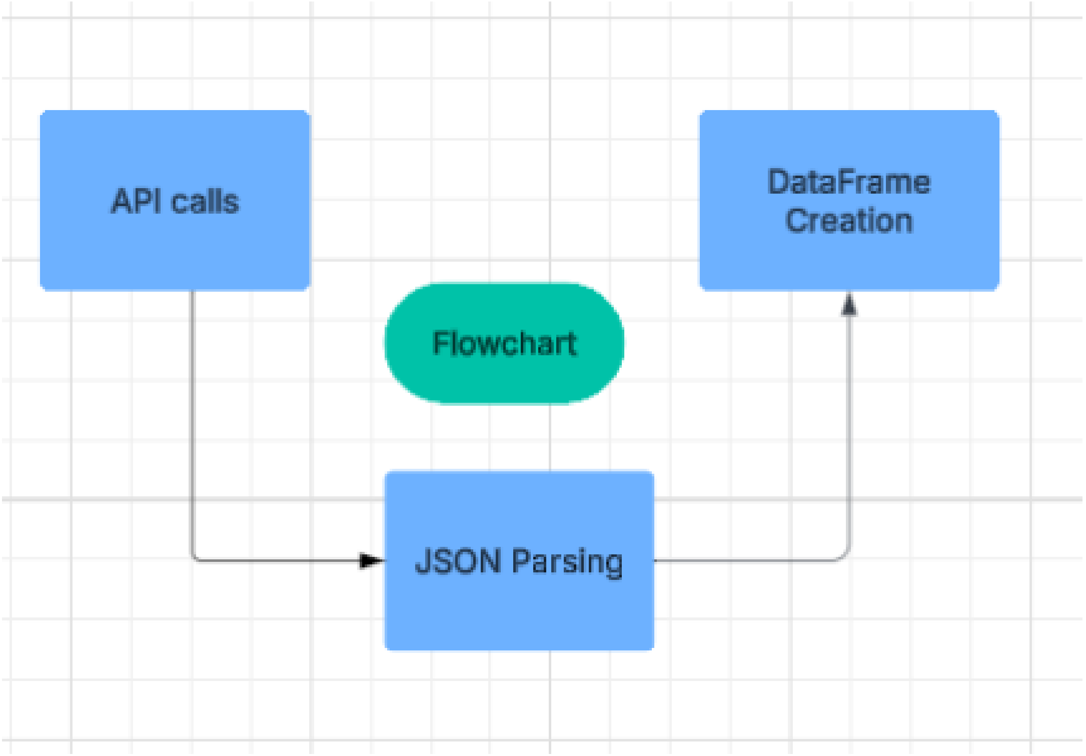


Data Collection

- **Process**: We used a combination of the SpaceX REST API and web scraping to collect our dataset. This hybrid approach allowed us to gather both structured launch data and unstructured historical data from a Wikipedia table

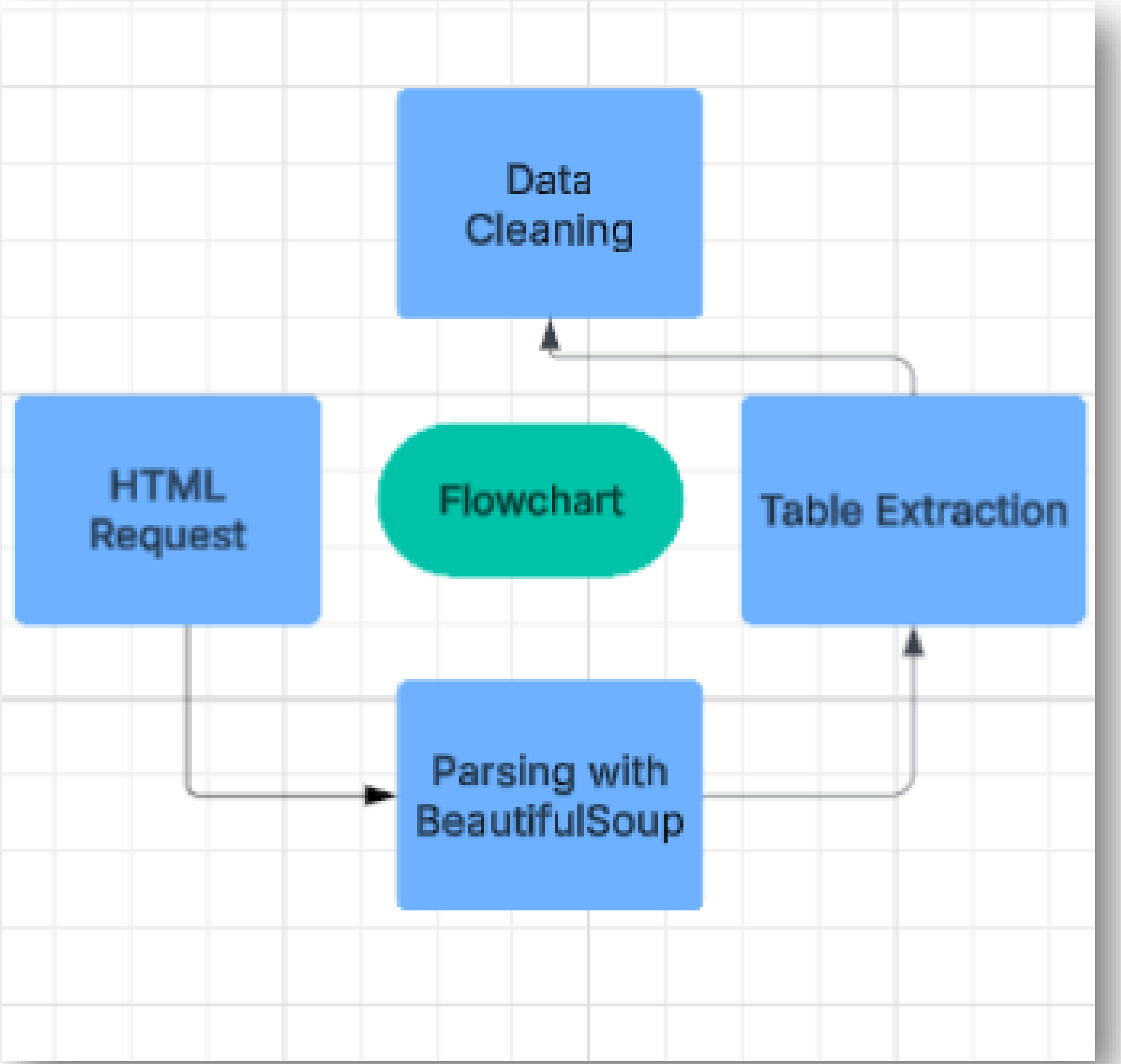
Data Collection – SpaceX API

- Used Python's requests library to make REST API calls to the SpaceX API.
- Parsed the JSON response to extract key launch information like flight number, date, and launch site.
- Transformed the data into a Pandas DataFrame for initial processing.
- **GitHub URL:** [View the full Jupyter Notebook on GitHub]



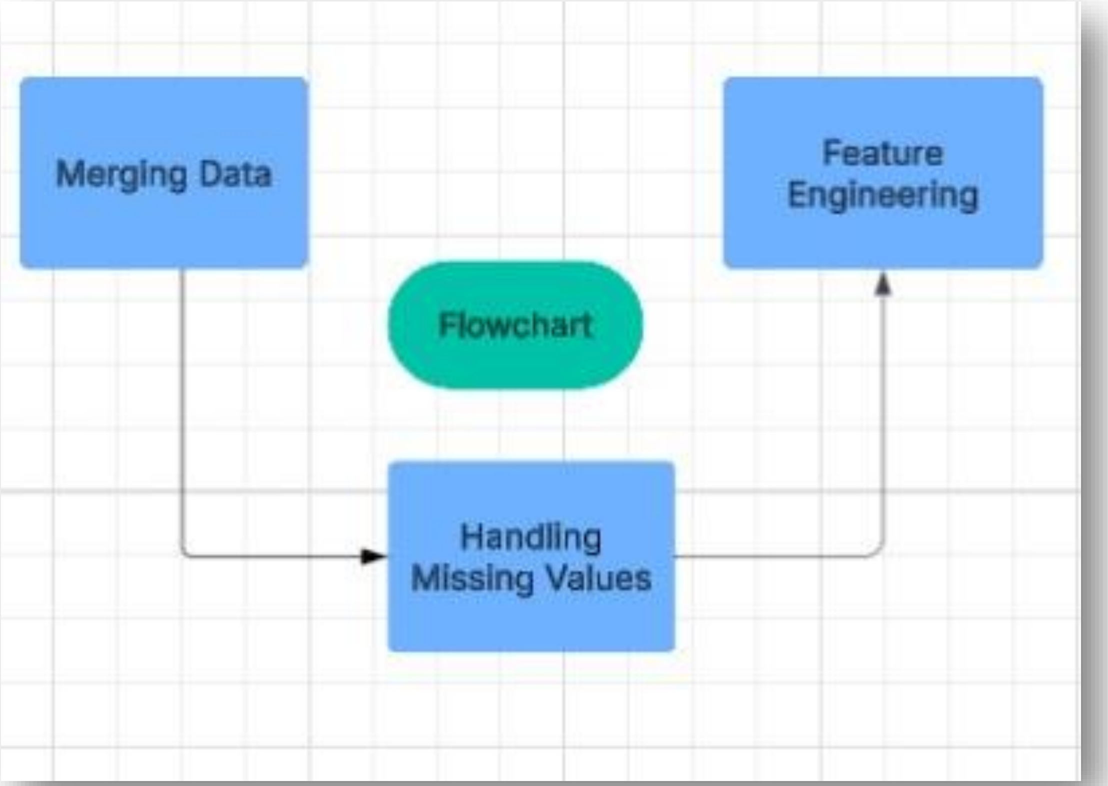
Data Collection - Scraping

- Utilized the BeautifulSoup library to parse the HTML content of the SpaceX launch Wikipedia table.
- Scraped the table to obtain detailed landing outcomes that were not available in the API data.
- Cleaned and processed the scraped data before merging it with the API data
- **GitHub URL:** View the full Jupyter Notebook on GitHub



Data Wrangling

- **Merging Data**: Combined the data from the API and web scraping notebooks into a single DataFrame.
- **Handling Missing Values**: Imputed or dropped missing values to ensure data quality.
- **Feature Engineering**: Created new features, such as categorical variables from the Orbit and LaunchSite columns
- **GitHub URL:** [View the full Jupyter Notebook on GitHub]



EDA with SQL

Communicate Python with SQL

Performed several SQL queries on cleaned dataset to get specific, aggregate insights.

Counted total successful and failure outcomes.

Calculated average payload mass for specific booster versions.

Found the first successful ground landing date.

Ranked landing outcomes by count over a specific time range.

GitHub URL: View the full Jupyter Notebook on GitHub

EDA with Data Visualization

Used **various plots to visualize relationships between variables and the landing outcome.**

- **Scatter Plots**: To show the relationship between continuous variables.
- **Bar Charts**: To compare success rates across different categories.
- **Line Charts**: To visualize trends over time.

GitHub URL: View the full Jupyter Notebook on GitHub

Build an Interactive Map with Folium

- **Summary:** We created an interactive map to visualize launch site locations and outcomes geographically.
- **Markers**: Placed markers for each launch site.
- **Color-coding**: Used different colors to show the outcome of each launch (success/failure).
- **Proximity Analysis**: Calculated and displayed distances to nearby infrastructure to assess logistical advantages.

- **GitHub URL:** View the full Jupyter Notebook on GitHub

Build a Dashboard with Plotly Dash

Summary: We built a dynamic dashboard to allow for interactive exploration of the data.

Pie Chart: Visualized the success count for all launch sites.

Scatter Plot: Enabled stakeholders to filter launch outcomes by PayloadMass using a range slider.

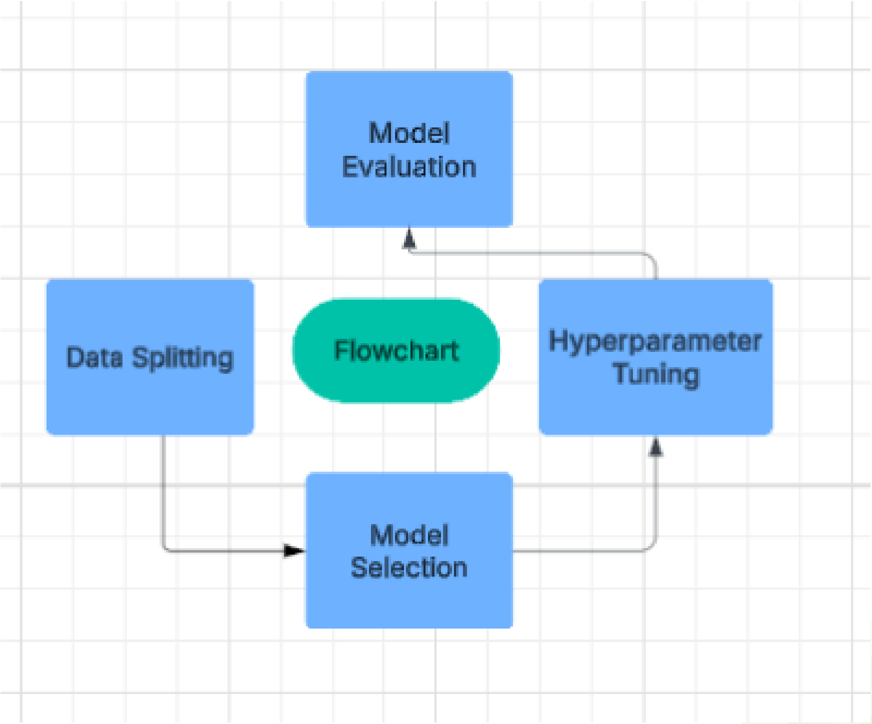
GitHub URL: View the full Python code on GitHub

Predictive Analysis (Classification)

Summary: Built, tuned, and evaluated multiple classification models to find the best predictor for the Falcon 9 landing outcome.

- **Data Splitting**: Splitting the cleaned data into training and testing sets.
- **Model Selection**: Developed four distinct machine learning models
- **Hyperparameter Tuning**: Used GridSearchCV.
- **Evaluation:** Based on Accuracy.

GitHub URL: View the full Jupyter Notebook on GitHub

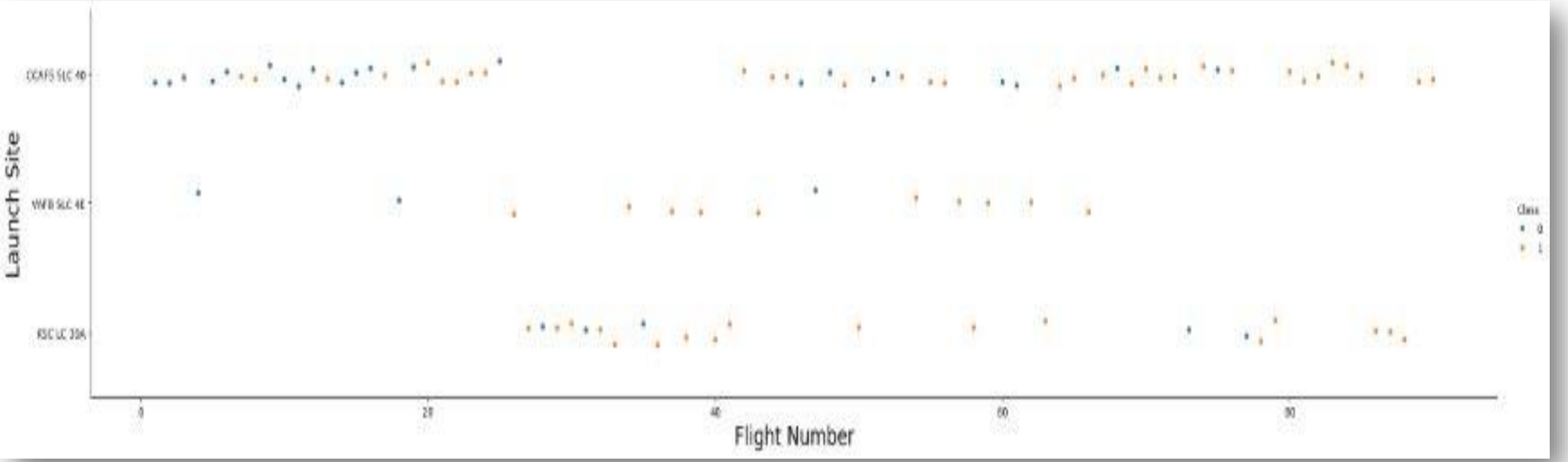


Section 2

Insights drawn from EDA

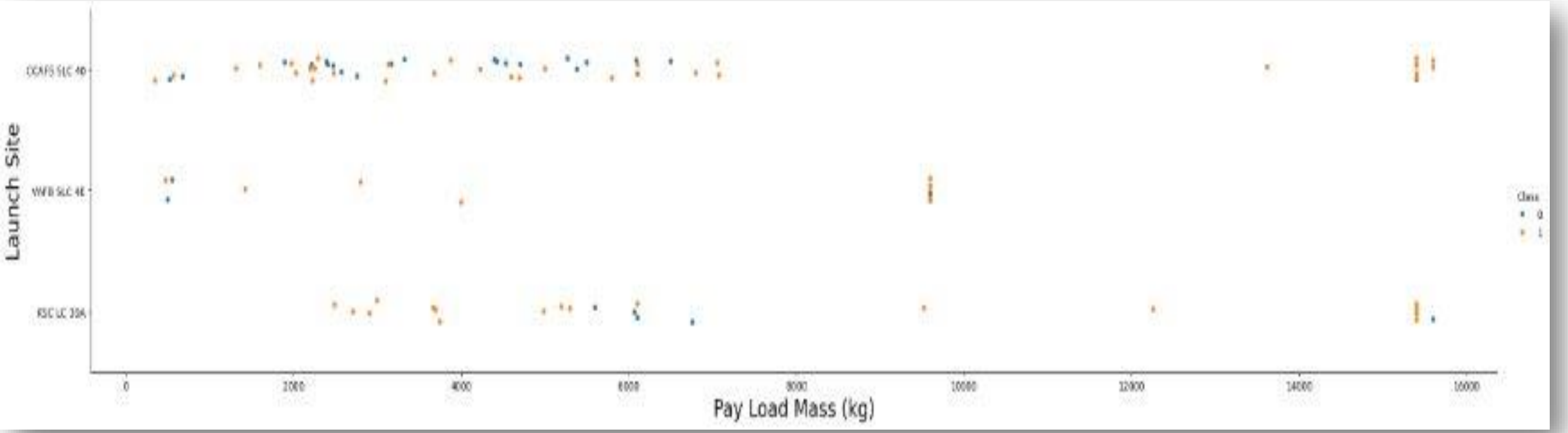
Flight Number vs. Launch Site

- **Explanation**: The scatter plot shows the launch outcomes at different sites over time. We can observe that certain launch sites, like VAFB SLC 4E, have a consistently high success rate.
- **Scatter Plot**



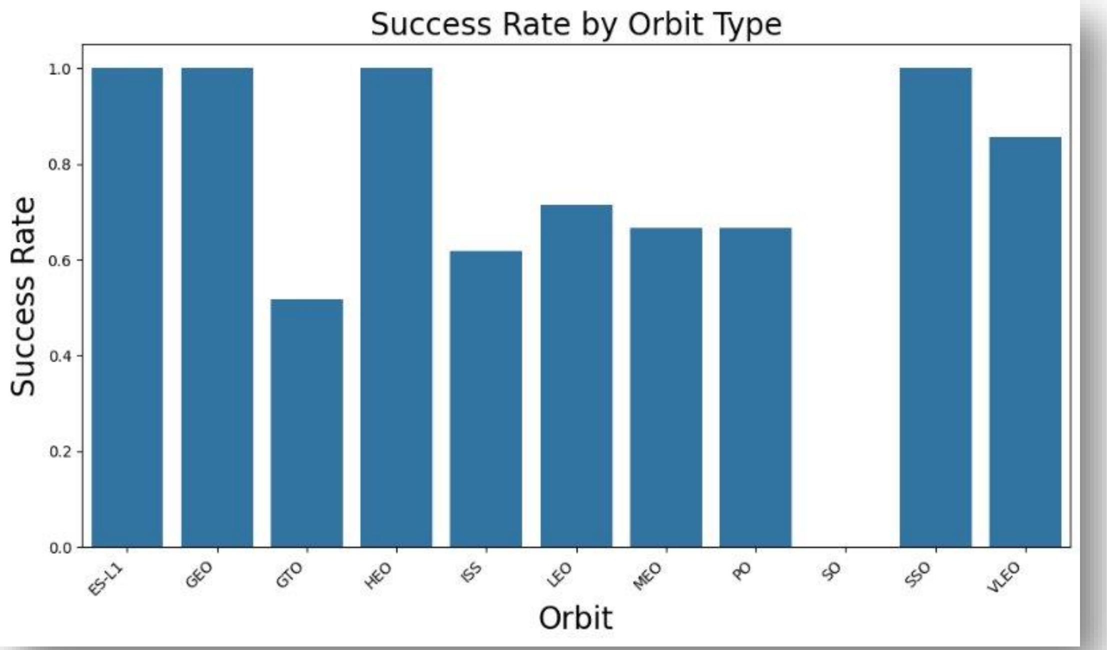
Payload vs. Launch Site

- **Explanation**: This plot helps visualize how PayloadMass affects landing outcome at each site. It shows that higher payload masses tend to be associated with successful landings, especially at the CCATS SLC 40 site.
- **Scatter Plot**



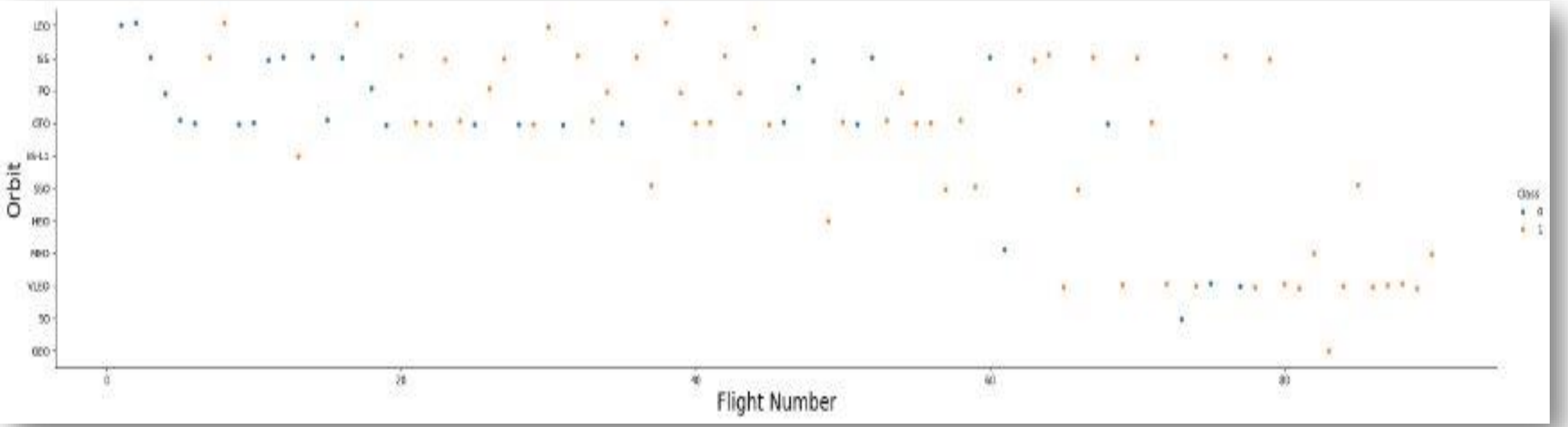
Success Rate vs. Orbit Type

- **Explanation**: This chart is critical for strategic decision-making. It reveals that some orbits, such as SSO and ES-L1, have a 100% success rate, while others, like GTO, have a lower success rate.
- **Bar Chart**



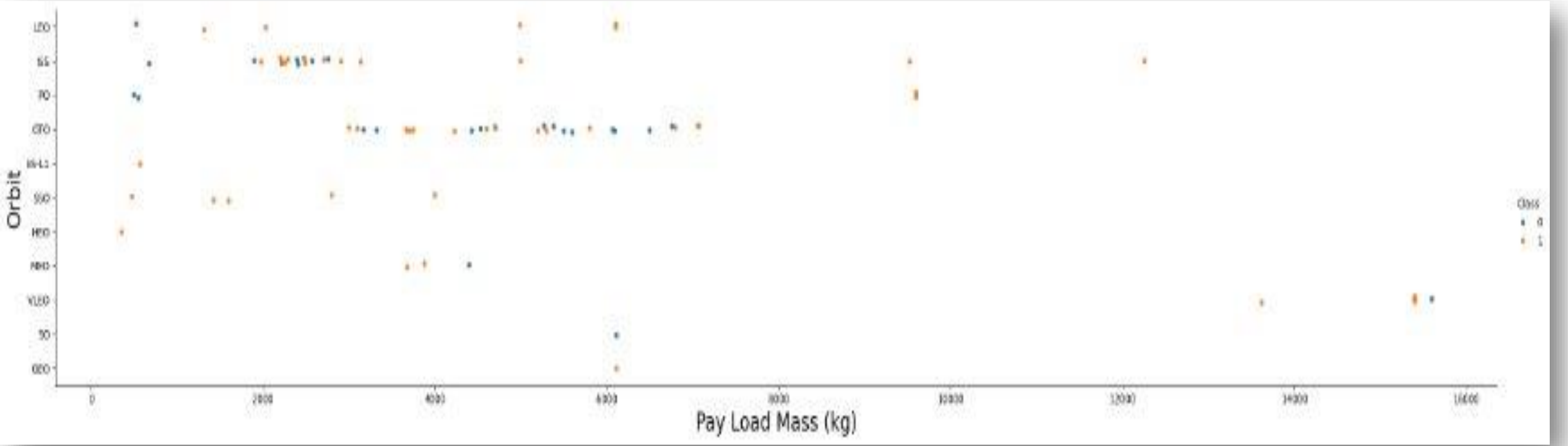
Flight Number vs. Orbit Type

- **Explanation**: This plot shows the evolution of launch outcomes for different orbits over time. It indicates a significant increase in successful landings for certain orbits as SpaceX’s technology matured.
- **Scatter Plot**



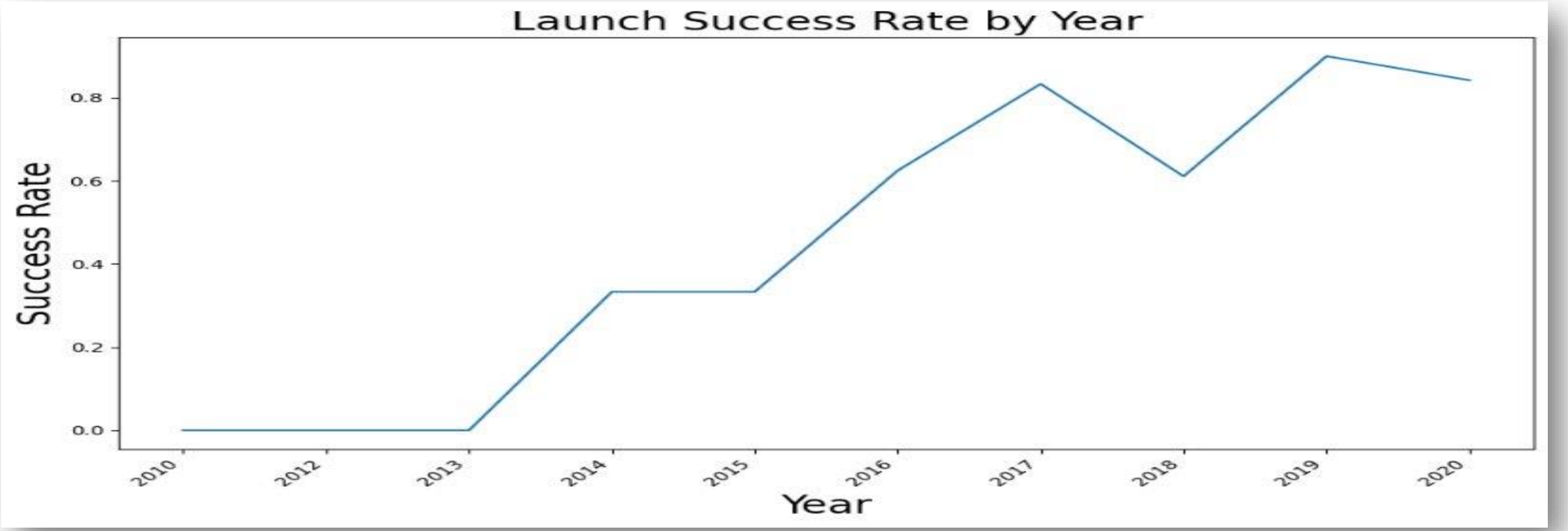
Payload vs. Orbit Type

- **Explanation**: This plot illustrates that a successful landing is possible across a wide range of payload masses for various orbits.
- **Scatter Plot**



Launch Success Yearly Trend

- **Explanation**: The line chart clearly shows a positive trend in the yearly average success rate, indicating a continuous improvement in SpaceX's landing technology.
- **Line Chart**



All Launch Site Names

- **Explanation**: This query identified all the unique launch sites used by SpaceX, providing a complete list of operational locations.

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

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

Total Payload Mass

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

TotalPayloadMass
45596

- **Explanation**: This result quantifies the total payload mass launched for NASA, highlighting a significant market for our competitors.

Average Payload Mass by F9 v1.1

- **Explanation**: This result provides insight into the typical payload for a specific, older booster version, which helps in understanding the evolution of booster capabilities.

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

AveragePayloadMass
2928.4

Launch Site Names Begin with 'CCA'

- **Explanation**: This query specifically identified the Cape Canaveral Space Launch Complex launch site, a key location for our analysis.

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

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

First Successful Ground Landing Date

- **Explanation**: This query pinpointed the date of a key technological milestone for SpaceX.

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

FirstSuccessfulGroundPadLanding
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- **Explanation**: This result lists the booster versions that have successfully performed drone ship landings with a medium range payload mass, which is a common payload for commercial satellites.

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

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

Total Number of Successful and Failure Mission Outcomes

- **Explanation**: This provides ahigh-level overview of SpaceX’s overall track record, showing the total number of successful vs failed landings.

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

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

Boosters Carried Maximum Payload

- **Explanation**: This query identifies the most powerful booster versions capable of carrying the heaviest payloads.

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

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

- **Explanation**: This result highlights the early challenges SpaceX faced with drone ship landings, an important point for risk assessment.

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

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

- **Explanation**: This query provides a ranking of landing outcomes during a specific, critical period of SpaceX's development.

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

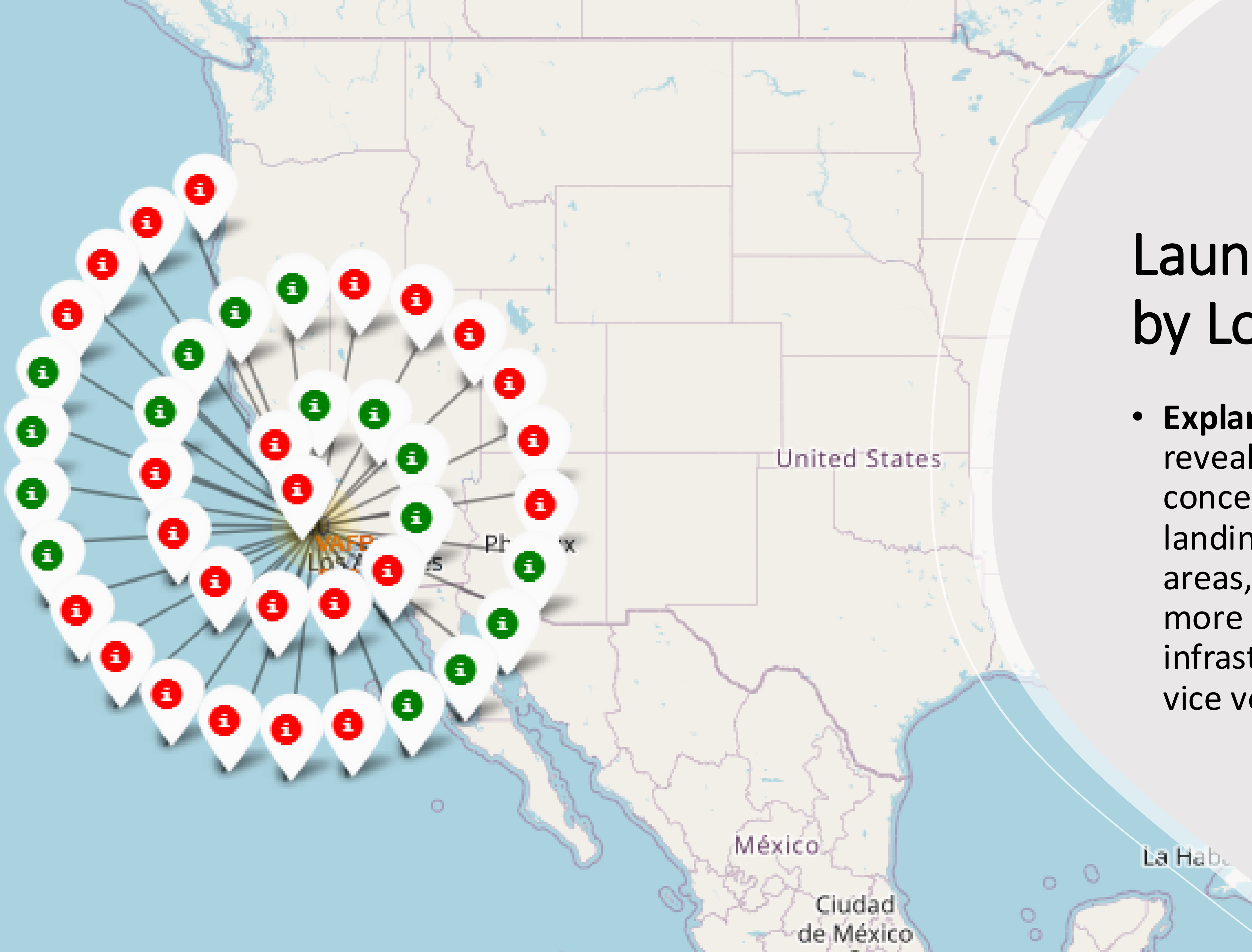
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

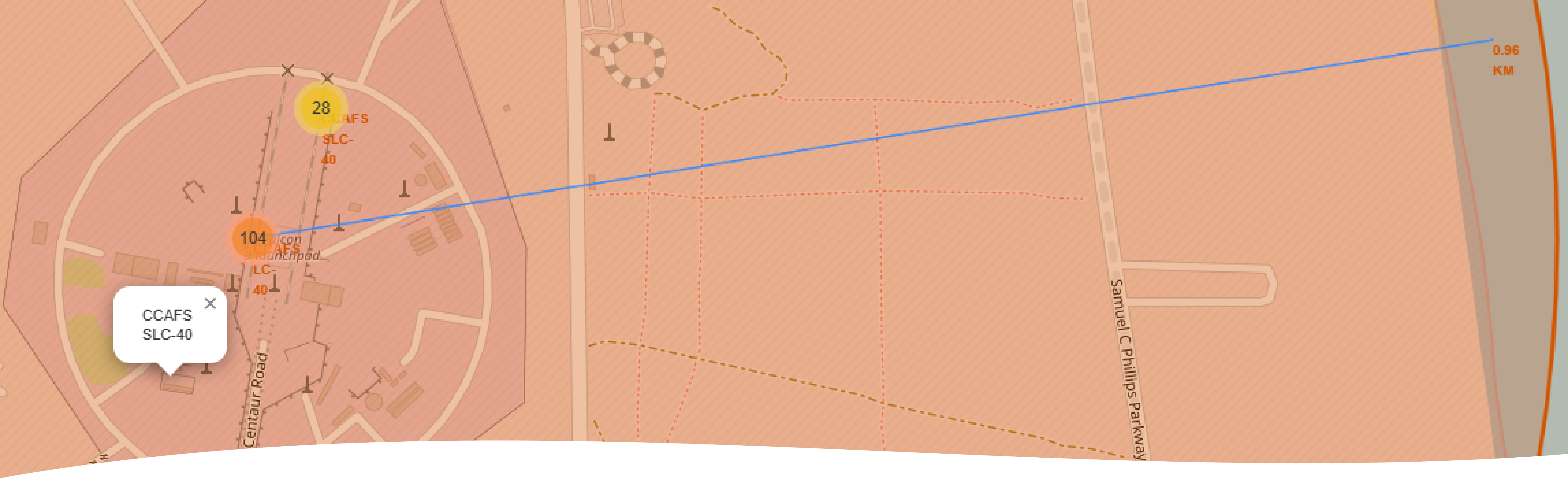
Launch Sites on a Global Map

- **Explanation**: This map provides a geographical perspective on SpaceX's launch operations, showing the location of all launch sites



Launch Outcomes by Location

- **Explanation**: The color-coding reveals a visual trend: a higher concentration of successful landings in certain geographic areas, which could be due to more favorable weather or infrastructure conditions and vice versa.



Launch Sites and their Proximities

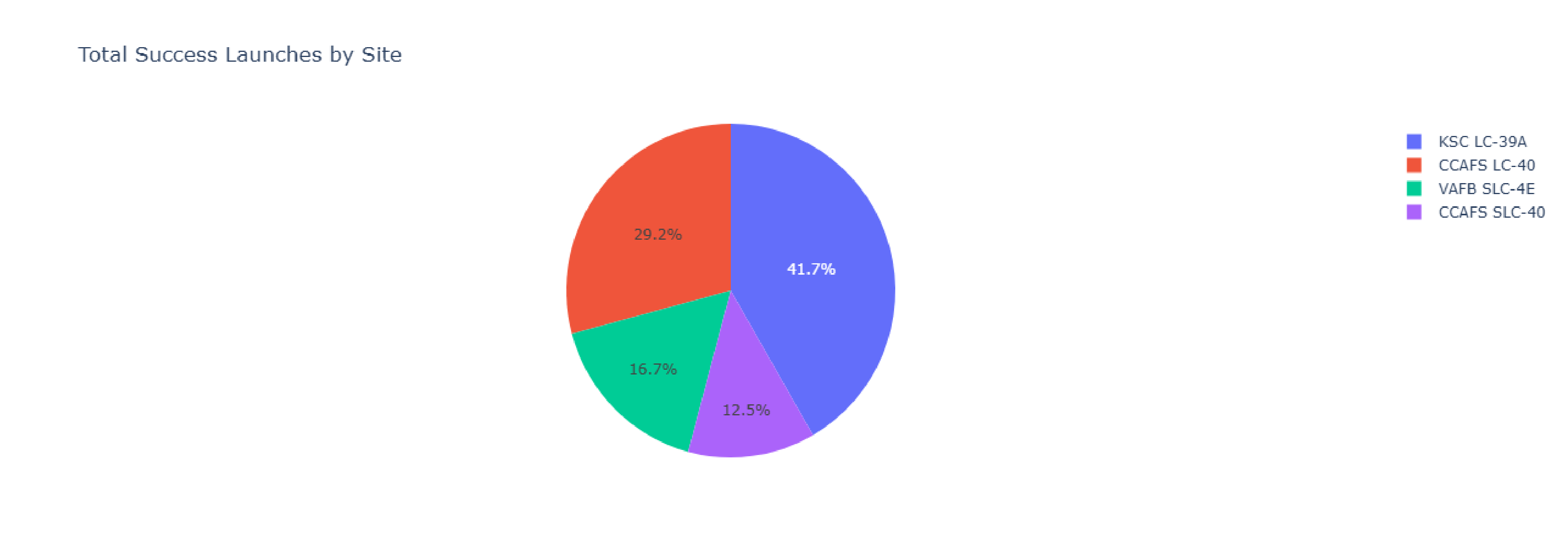
- **Explanation**: This analysis shows the logistical advantages of each launch site by measuring its proximity to key transport routes, which is crucial for moving rocket components.

Section 4

Build a Dashboard with Plotly Dash

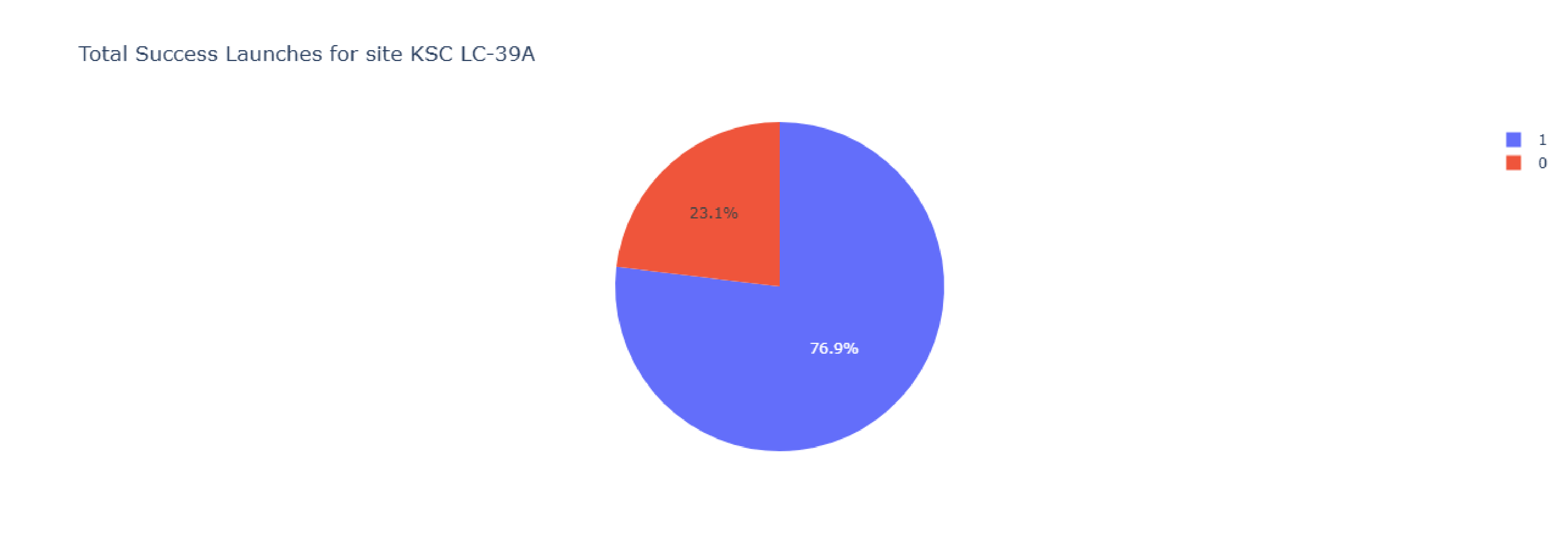
Launch Success Rates by Site

- **Explanation**: This pie chart offers a quick, at-a-glance view of the overall success and failure rates across all launch sites, providing a clear summary of SpaceX's performance.



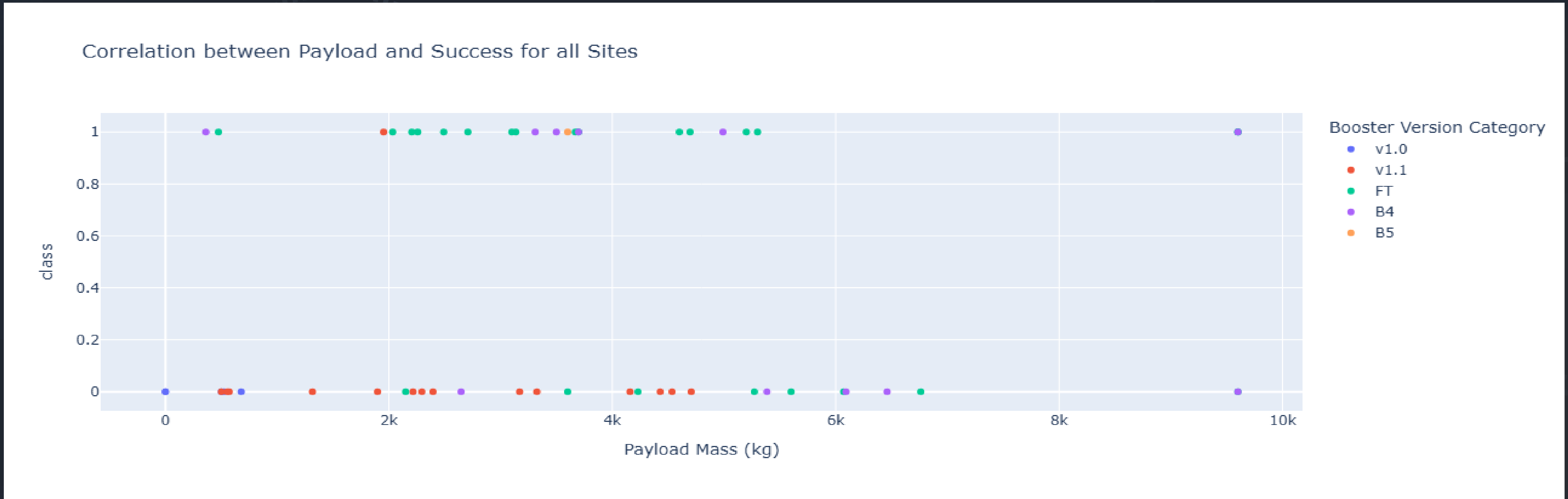
Highest Launch Success Ratio

- **Explanation**: By drilling down, we identified the launch site with the highest success ratio, a key finding that informs our understanding of the most reliable launch locations



Payload vs. Launch Outcome

- **Explanation**: This interactive plot demonstrates that a specific payload range (e.g., 4000-6000 kg) and certain booster versions have the largest success rate. This insight is extremely valuable for targeting profitable market segments.

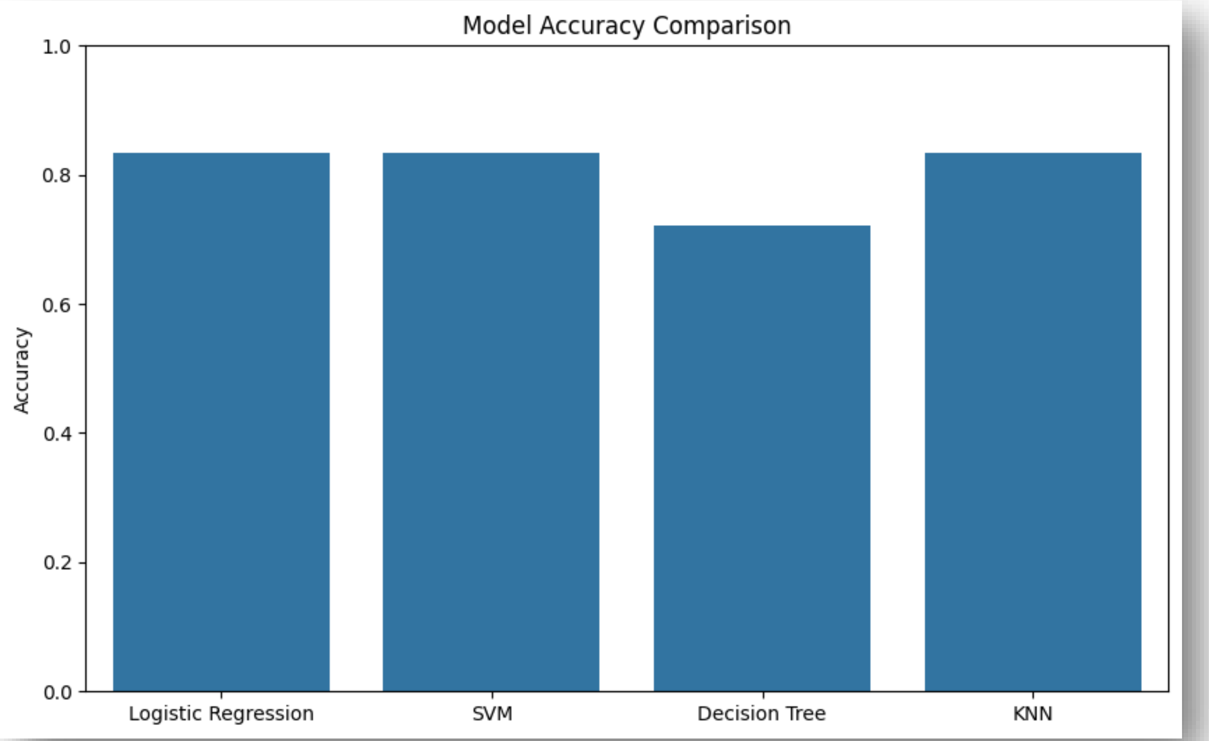


Section 5

Predictive Analysis (Classification)

Classification Accuracy

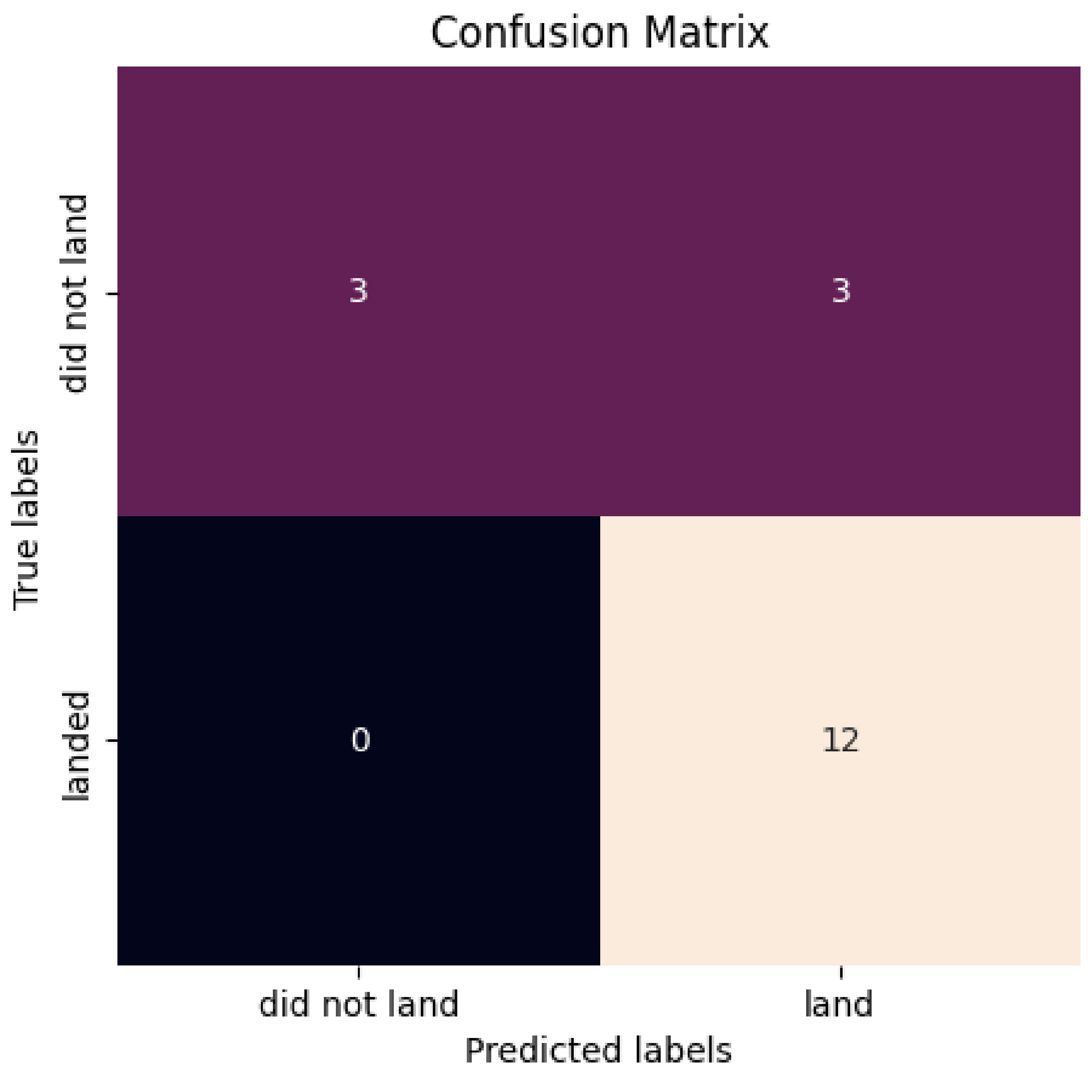
- **Explanation**: The bar chart shows that the **Logistic Regression Classifier** achieved the highest accuracy on test data. This model is our best predictor of a successful landing.



Confusion Matrix of Best Performing Model

Explanation: Total Samples: **18**

- **True Positives**: The model correctly predicted successful landings.
- **True Negatives**: The model correctly predicted failed landings.
- **False Positives**: The model incorrectly predicted a failed landing as successful.
- **False Negatives**: The model incorrectly predicted a successful landing as failed.
- The matrix shows that our model has a low rate of misclassification, making it a reliable tool for predicting landing success.



Results

Summary: Our comprehensive analysis yielded a clear and actionable path for our startup.

Exploratory Data Analysis (EDA) revealed that Orbit type, PayloadMass, and LaunchSite are the most influential factors in determining a successful landing. We observed a consistent **upward trend** in SpaceX's landing success over time.

The **interactive visualizations** provided powerful insights. The Folium map highlighted the logistical advantages of different launch sites, while the Plotly Dash dashboard demonstrated that specific booster versions have a significantly higher success rate.

The **predictive analysis** confirmed our findings. The "Logistic Regression Classifier" emerged as the best-performing model with a high accuracy on test data, providing a reliable tool to predict the landing outcome.

Conclusions

Data-Driven Insights: The analysis shows a clear upward trend in SpaceX's landing success rate, driven by improvements in Orbit type, PayloadMass, and LaunchSite.

Actionable Recommendations: Our startup can leverage these findings to make more informed bids against SpaceX. By focusing on launch sites and orbits with higher success rates, we can maximize our chances of a successful and cost-effective mission.

Predictive Power: Our Logistic Regression Classifier provides a robust and accurate prediction of landing success, which is a significant advantage in the competitive space launch market.

Thank you!