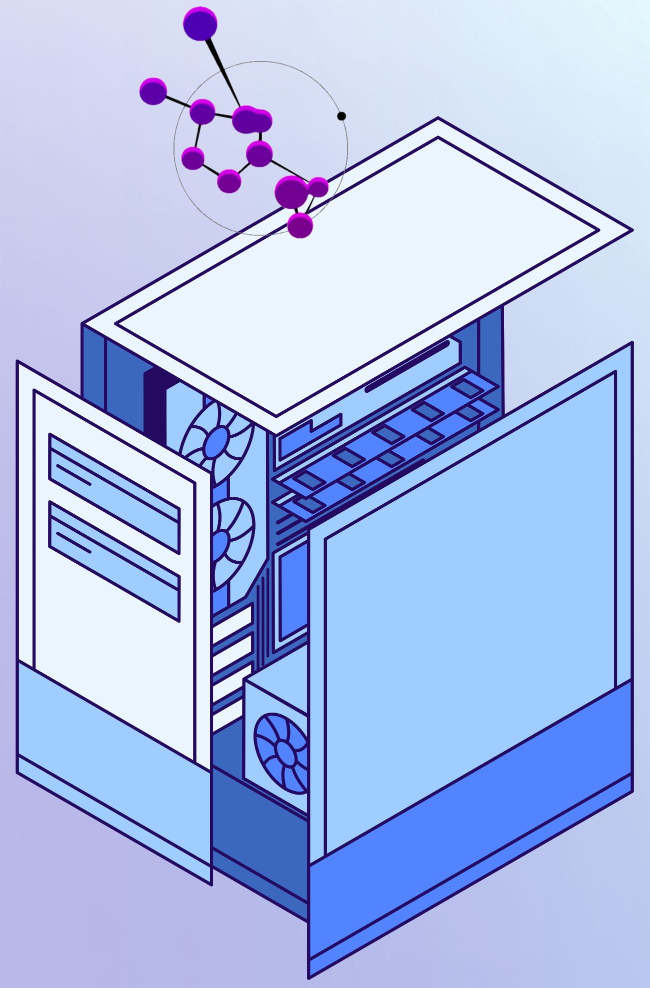
PREDICTIVE ANALYTICS FOR SPACEX LAUNCH SUCCESS USING MACHINE LEARNING

A DATA SCIENCE CAPSTONE PROJECT

Angela Paola Lozano Ochoa

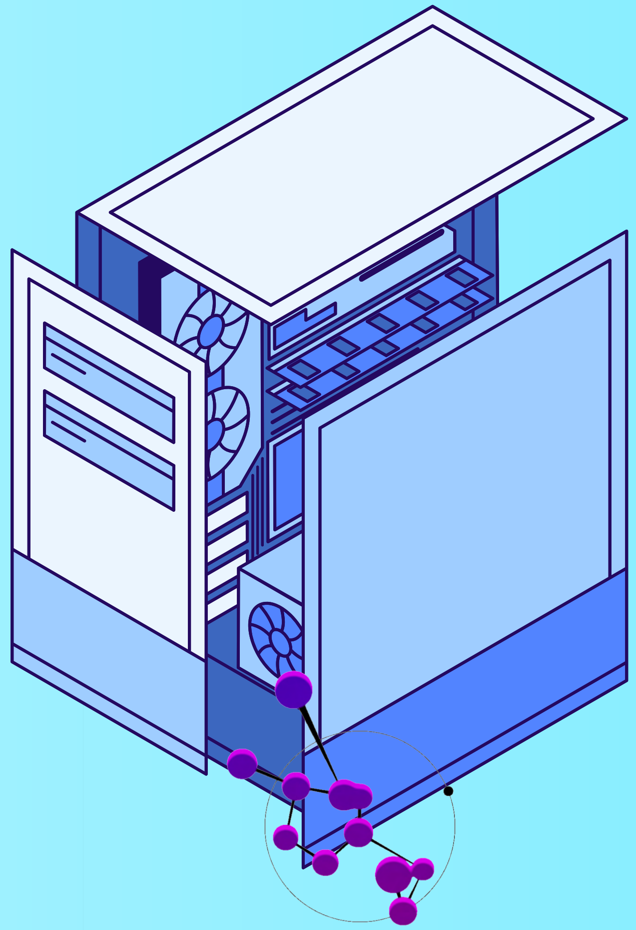
September 2024

EXECUTIVE SUMMARY

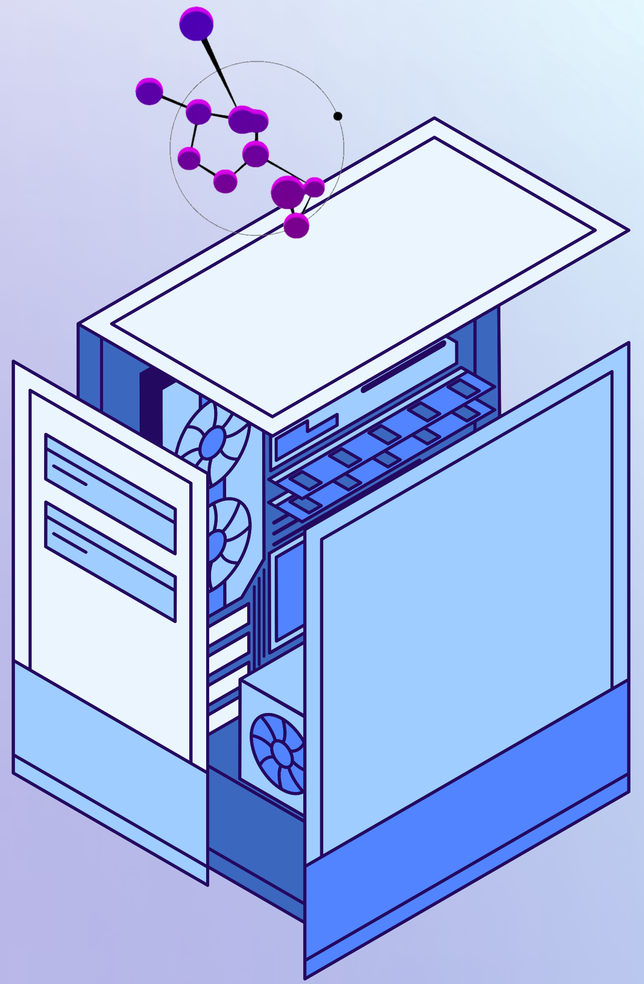


SEREINCAPSTONE Home About us Startegy Contact

This report presents the SpaceX Launch Success Prediction project, a Data Science Capstone developed for IBM's certification. The aim of this project was to predict the likelihood of a SpaceX launch being successful, based on historical data, using various data science techniques including web scraping, data wrangling, exploratory data analysis (EDA), and machine learning.

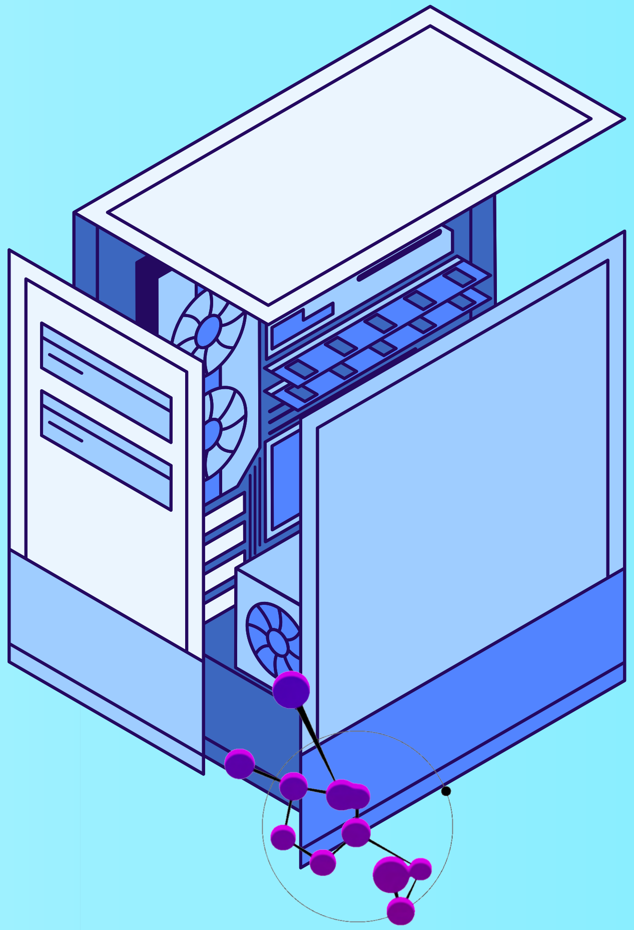


EXECUTIVE SUMMARY



SEREINCAPSTONE   Home   About us   Startegy   Contact

1. SpaceX Data Collection via API
2. Web Scraping for Supplementary Data
3. Data Wrangling and Preprocessing
4. Exploratory Data Analysis (EDA) with SQL
5. EDA and Data Visualization
6. Launch Site Location Analysis
7. Plotly Dash App Development
8. SpaceX Machine Learning Prediction



INTRODUCTION

SpaceX, a leader in private space exploration, has revolutionized the aerospace industry by developing reusable rocket technology. A Falcon 9 rocket launch costs $62 million, significantly lower than competitors, who charge upwards of $165 million per launch. This cost reduction is primarily due to SpaceX's ability to reuse the rocket’s first stage. Predicting whether the first stage will land successfully is crucial for optimizing costs and can provide insights for companies competing with SpaceX for rocket launches.

INTRODUCTION

Objective

The goal of this project is to create a comprehensive machine learning pipeline that predicts the successful landing of the Falcon 9 first stage. This project collects, processes, analyzes, and models SpaceX launch data to provide actionable insights on launch outcomes, particularly regarding cost optimization for reusability.

INTRODUCTION

Challenges

In this project, we aim to address several critical questions and challenges that are central to understanding and improving the success of SpaceX's Falcon 9 missions:

1. What factors influence the successful landing of the Falcon 9 first stage?
2. How do different launch sites affect the success rate of landings?
3. How does payload mass impact the landing success?
4. Can we predict landing outcomes using machine learning models?
5. What role does the type of booster play in landing success?
6. How can interactive data visualizations enhance understanding and decision-making?

IBM SPACEX

METHODOLOGY

Data Collection: Methods used to gather raw data from various sources, including APIs and web scraping.
Data Wrangling: Techniques for cleaning, transforming, and structuring the data to prepare it for analysis.
Exploratory Data Analysis (EDA): Procedures for investigating data patterns, relationships, and insights.
Feature Engineering: Processes for creating features that improve the performance of predictive models.
Model Building: Implementation of machine learning algorithms, including training, validation, and testing.
Visualization and Reporting: Use of interactive dashboards and visualizations to present findings and insights.

DATA COLLECTION

INTRODUCTION

Data collection is a crucial first step in any data science project. For our SpaceX project, it involved gathering reliable and structured data from two primary sources: SpaceX API and Web Scraping. By collecting this data, we can build a comprehensive dataset for further analysis, feature engineering, and predictive modeling.

IBM SPACEX

DATA COLLECTION

DATA COLLECTION WORKFLOW

1 

API/Web Scraping

2 Create DataFrame

3 Clean and Filter Data

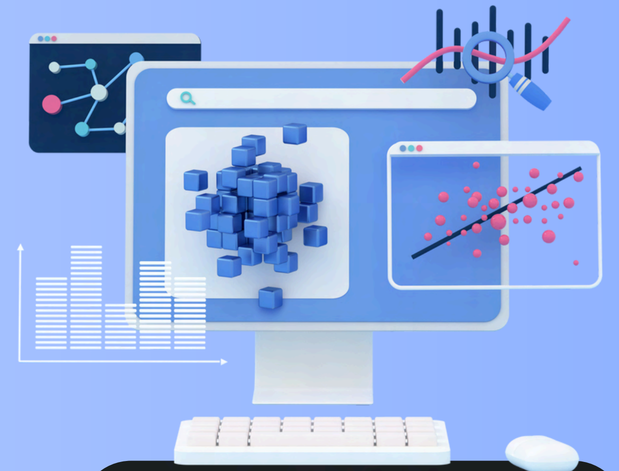
4 Export Data

EXPORT

DATA COLLECTION

SPACE X API

ANGELA PAOLA LOZANO OCHOA



1 API Integration

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
response.status_code
```

```
data = pd.json_normalize(response.json())
data.head()
```

2 Data Cleaning

Extracted relevant details such as Booster Version, Launch Site, Payload Data, and Core Data.

```
getBoosterVersion(data)
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```

Defined a dictionary to organize the data fields and convert it into a DataFrame for easier manipulation.

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
...
```

3 Data Filtering

Removed entries for Falcon 1 rockets, as they are not relevant to the Falcon 9 analysis.

Filled missing values in critical columns such as PayloadMass with the average value to ensure consistency.

```
data_falcon9 = data[data['BoosterVersion']
 != 'Falcon 1']
data_falcon9['PayloadMass'].fillna(value=mean_payloadMass,
 inplace=True)
```

4 Data Export

Exported the cleaned and filtered data to a CSV file


```
data_falcon9.to_csv
('dataset_part_1.csv',
index=False)
```

DATA COLLECTION

WEB SCRAPING

In this phase, we performed web scraping to collect historical launch records of Falcon 9 rockets from Wikipedia. The Wikipedia page titled "List of Falcon 9 and Falcon Heavy launches" was scraped using Python libraries to extract and structure the data.

1 Send GET Request

Fetch the HTML content from the Wikipedia page.

```
static_url = "https://en.wikipedia.org/w/
index.phptitle=List_of_Falcon_9_and_Falcon_
Heavy_launches&oldid=1027686922"
response = requests.get(static_url)
```

2 Parse HTML

BeautifulSoup

```
soup = BeautifulSoup(
response.content,
"html.parser"
)
```

3 Extract Data

Extract column names and table content.

```
html_tables = soup.find_all("table")
first_launch_table = html_tables[2]
```

4 Create DataFrame

Store the data in a pandas DataFrame.

```
launch_dict = dict.fromkeys(column_names)
df = pd.DataFrame({key: pd.Series(value) for
key, value in launch_dict.items()})
```

5 Save as CSV

Export the DataFrame to a CSV file for further use.


```
df.to_csv(
'spacex_web_scraped.csv',
index=False
)
```

IBM SPACEX

DATA WRANGLING

INTRODUCTION

Data wrangling, or data munging, refers to the process of transforming and preparing raw data into a usable format for analysis. In our SpaceX project, it was essential to clean, structure, and simplify the data collected from multiple sources (API and web scraping). This step ensures that the dataset is free of inconsistencies, missing values, and irrelevant information, making it suitable for feature extraction and machine learning.

IBM SPACEX

DATA WRANGLING

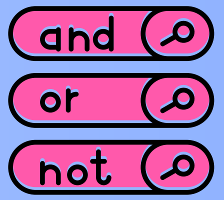
DATA WRANGLING WORKFLOW

1 Load Data

2 Create DataFrame

3 Clean the Data

4 Simplify Boolean Data



5 Export Data

DATA WRANGLING

Convert Mission outcomes to Training Labels

We converted the landing outcomes into binary labels:
1 indicates a successful landing.
0 indicates an unsuccessful landing.

1 Analyze Launch Sites

Calculate the number of launches on each site

```
df['LaunchSite'].value_counts()
LaunchSite
CCAFS SLC 40     55
KSC LC 39A       22
VAFB SLC 4E      13
Name: count, dtype: int64
```

2 Analyze Orbits

Calculate the number and occurrence of each orbit

```
df['Orbit'].value_counts()
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
```

3 Mission Outcome

Calculate the number and occurence of mission outcome of the orbits

```
landing_outcomes =
df['Outcome'].value_counts()
Outcome
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
```

4 Create Landing Outcome Labels

Create a landing outcome label from Outcome column

```
df['Class']=[0 if x in bad_outcomes
else 1 for x in df['Outcome']]
```

5 Export Data

```
df.to_csv(
"dataset_part_2.csv",
 index=False
)
```

IBM SPACEX

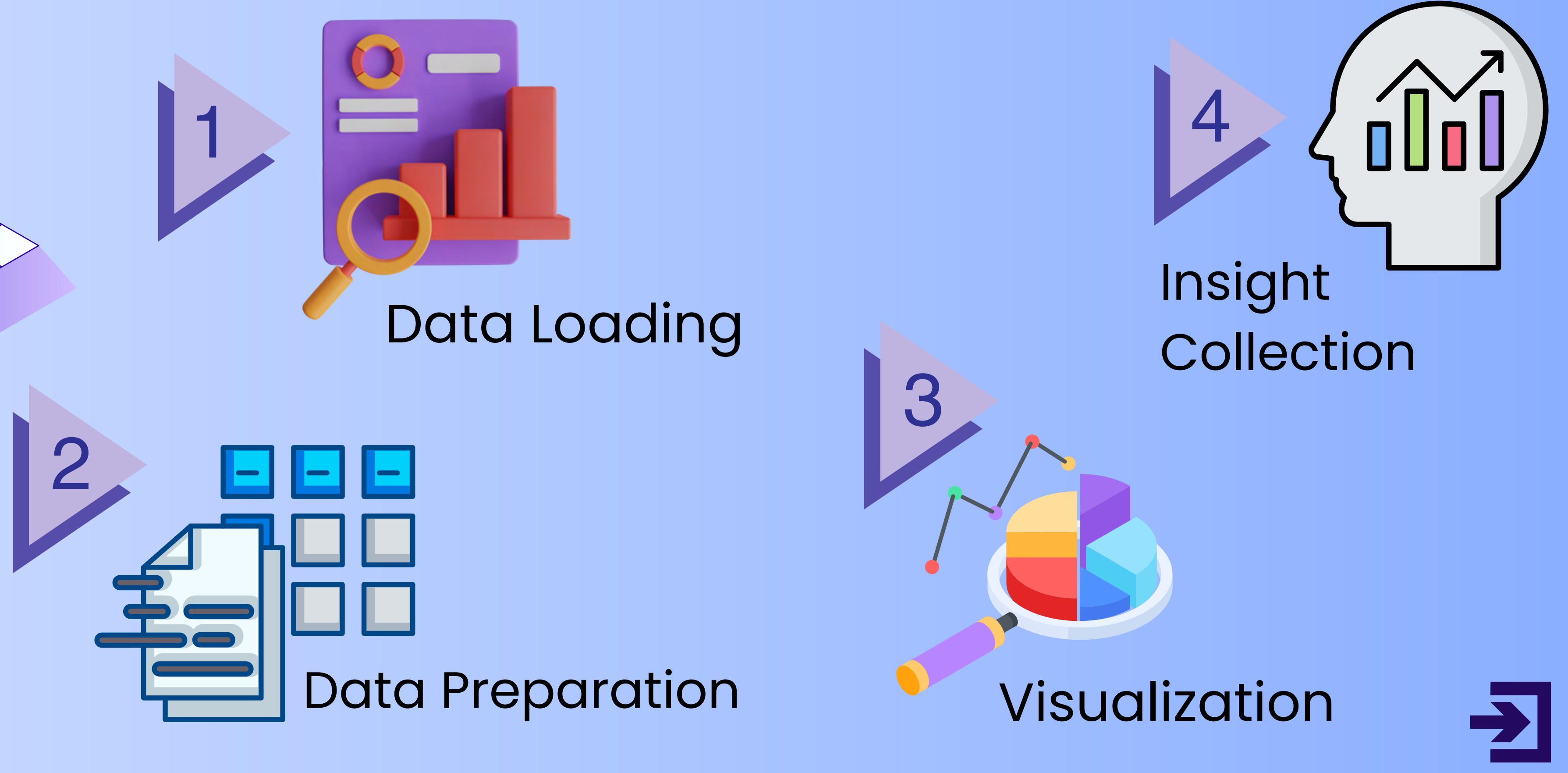
EDA

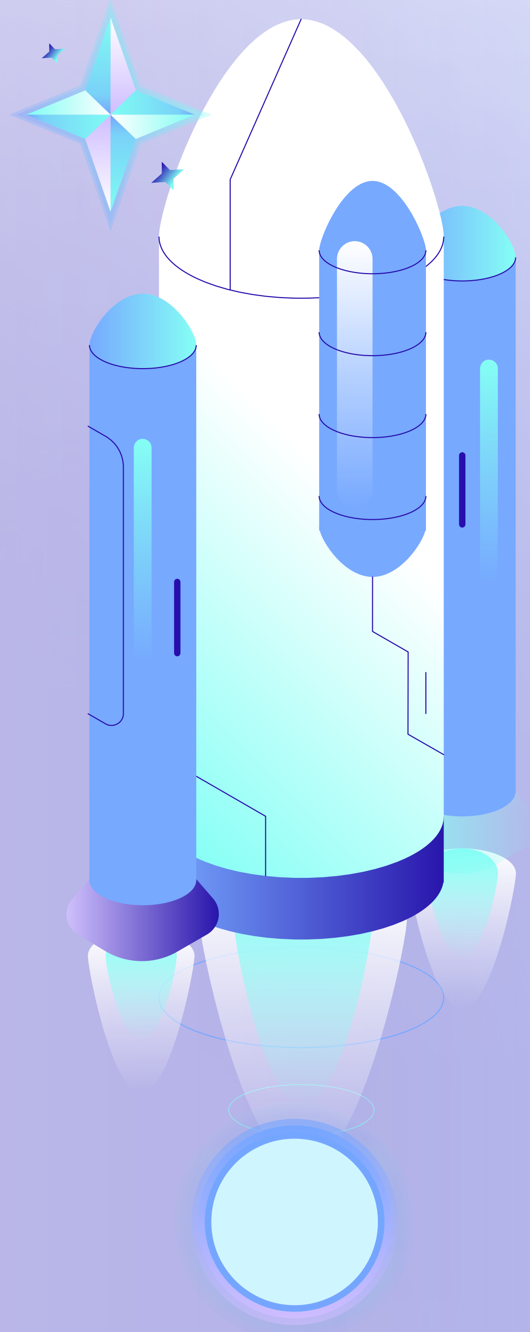
INTRODUCTION TO EDA

Exploratory Data Analysis (EDA) is a crucial phase in the data analysis process. It involves the systematic examination of data to uncover patterns, relationships, anomalies, and insights before applying more formal statistical or machine learning methods.

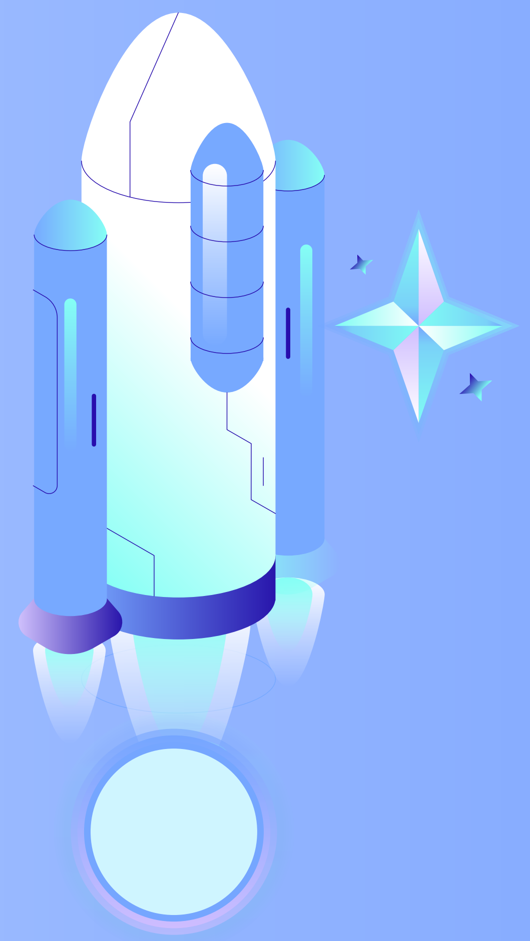
EDA

EDA WORKFLOW DIAGRAM





EDA WITH DATA VISUALIZATION



In this section, we used several types of visualizations to uncover insights and patterns from the SpaceX dataset. These graphs allowed us to explore relationships between key variables and better understand trends in launch success and performance.

Scatter Plots

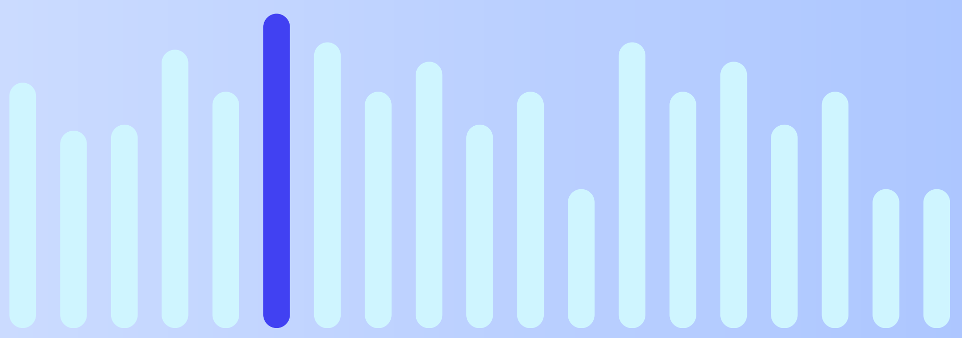
Scatter plots were used to investigate correlations between numeric variables and launch success.

1. *Payload vs. Flight Number*
2. *Flight Number vs. Launch Site*
3. *Payload vs. Launch Site*
4. *Flight Number vs. Orbit Type*
5. *Payload vs. Orbit Type*

Bar Charts

Bar charts helped us aggregate data to understand categorical relationships.

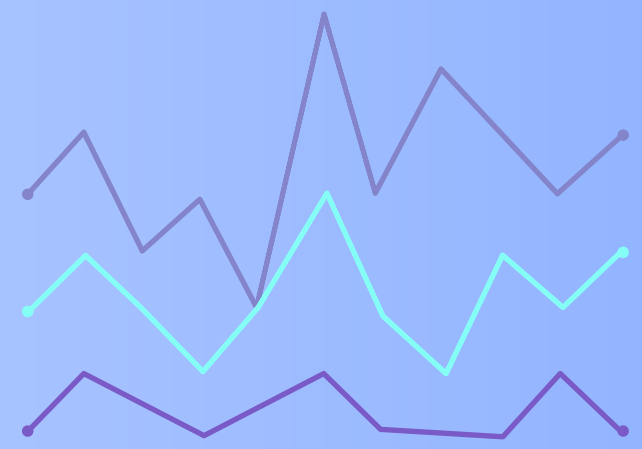
1. *Success Rate vs. Orbit Type*

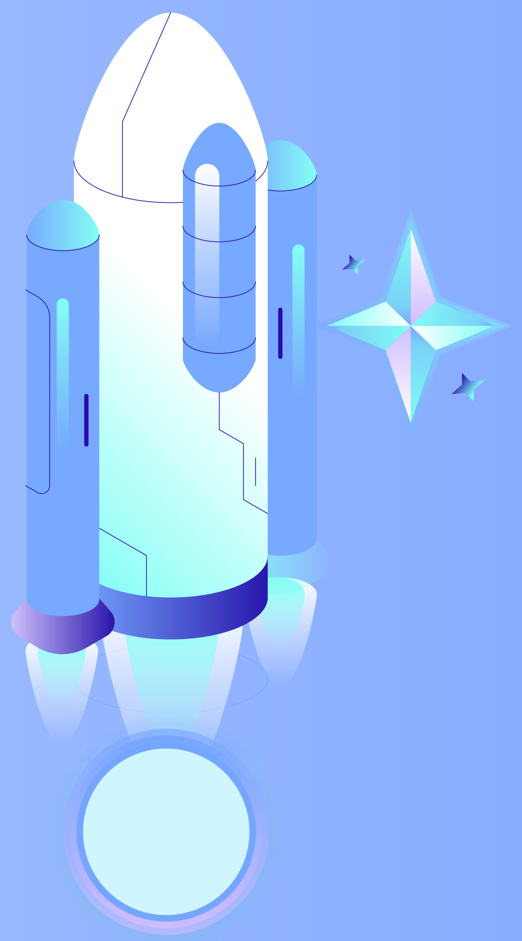


Line Charts

Line charts were used to analyze trends over time.

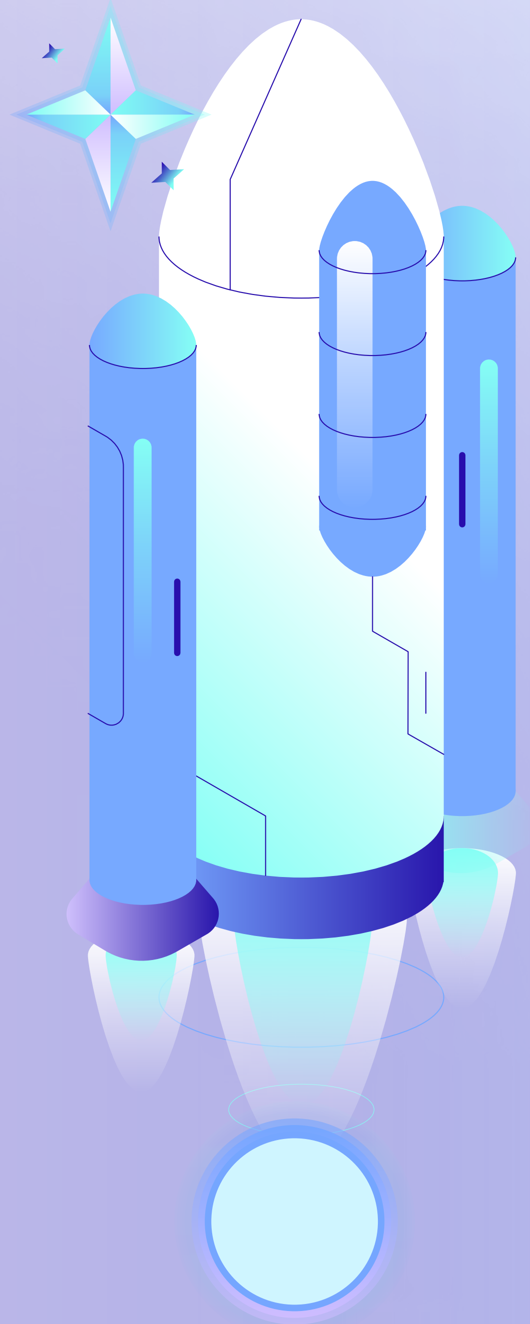
1. *Launch Success Yearly Trend*





EDA WITH SQL

SQL was utilized to perform complex queries on our SpaceX dataset, providing insights that were critical for understanding patterns and trends. By leveraging SQL, we were able to efficiently retrieve and analyze specific subsets of the data to answer key questions about launch sites, payloads, and mission outcomes.



Key Queries

1. *Display the names of the unique launch sites in the space mission.*
2. *Display 5 records where launch sites begin with the string 'CCA'.*
3. *Display the total payload mass carried by boosters launched by NASA (CRS).*
4. *Display the average payload mass carried by booster version F9 v1.1.*
5. *List the date when the first successful landing outcome in a ground pad was achieved.*
6. *List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.*
7. *List the total number of successful and failure mission outcomes.*
8. *List the names of the booster versions which have carried the maximum payload mass. Use a subquery.*
9. *List the records which will display the month names, failure landing outcomes in drone ship, booster versions, and launch site for the months in the year 2015.*
10. *Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order.*

INTERACTIVE MAP

Overview

In this section, we leverage Folium to create interactive visualizations that enhance our understanding of SpaceX launch sites. By visualizing launch sites and success rates on a map, we can gain insights into geographical factors affecting launch success. This interactive analysis helps identify patterns and relationships between launch sites and their performance.

IBM SPACEX

Folium

INTERACTIVE MAP

Objectives

1. Mark all launch sites on a map to visualize their geographical distribution.
2. Mark successful and failed launches for each site to identify patterns and performance metrics.
3. Calculate and visualize the distances between launch sites and their proximities to analyze spatial relationships.

Folium

MAP OBJECTS

Map Marker	folium.Marker()	Adds a marker to the map at a specified location.
Icon Marker	folium.Icon()	Customizes the appearance of markers with icons and colors.
Circle Marker	folium.Circle()	Creates a circle around a location to represent area of interest.
Polyline	folium.Polyline()	Draws lines between points to show trajectories or connections.
Marker Cluster Object	MarkerCluster()	Groups multiple markers into clusters to reduce map clutter.
AntPath	AntPath()	Draws animated paths to indicate movement or trajectories.

DASHBOARD

Overview

In this section, we create an interactive dashboard using Plotly Dash to visualize and analyze SpaceX launch data. The dashboard includes visualizations that help us understand launch success rates, payload impacts, and site-specific performance metrics.

DASHBOARD COMPONENTS

Pie Chart	Displays the distribution of successful and failed launches.
Scatter Chart	Illustrates the correlation between payload mass and launch success.

INTERACTIVE ELEMENTS

Dropdown	dcc.Dropdown()	Allows users to select a launch site to filter the data.
RangeSlider	dcc.RangeSlider()	Enables users to filter data based on payload mass range.

PREDICTIVE ANALYSIS

CLASSIFICATION

Overview

This section outlines the steps taken to build, train, evaluate, and select the best-performing classification model for predicting the success of Falcon 9 launches.

IBM SPACEX

PREDICTIVE ANALYSIS WORKFLOW

1

DATA PREPROCESSING

1. Create Target Variable
2. Standardize Features
3. Train-Test Split
4. Logistic Regression with GridSearchCV

2

MODEL TRAINING

1. Check Accuracy
2. Identify Best Hyperparameters
3. Confusion Matrix

3

BEST MODEL SELECTION

The model with the highest accuracy score is selected as the best-performing classification model.

4

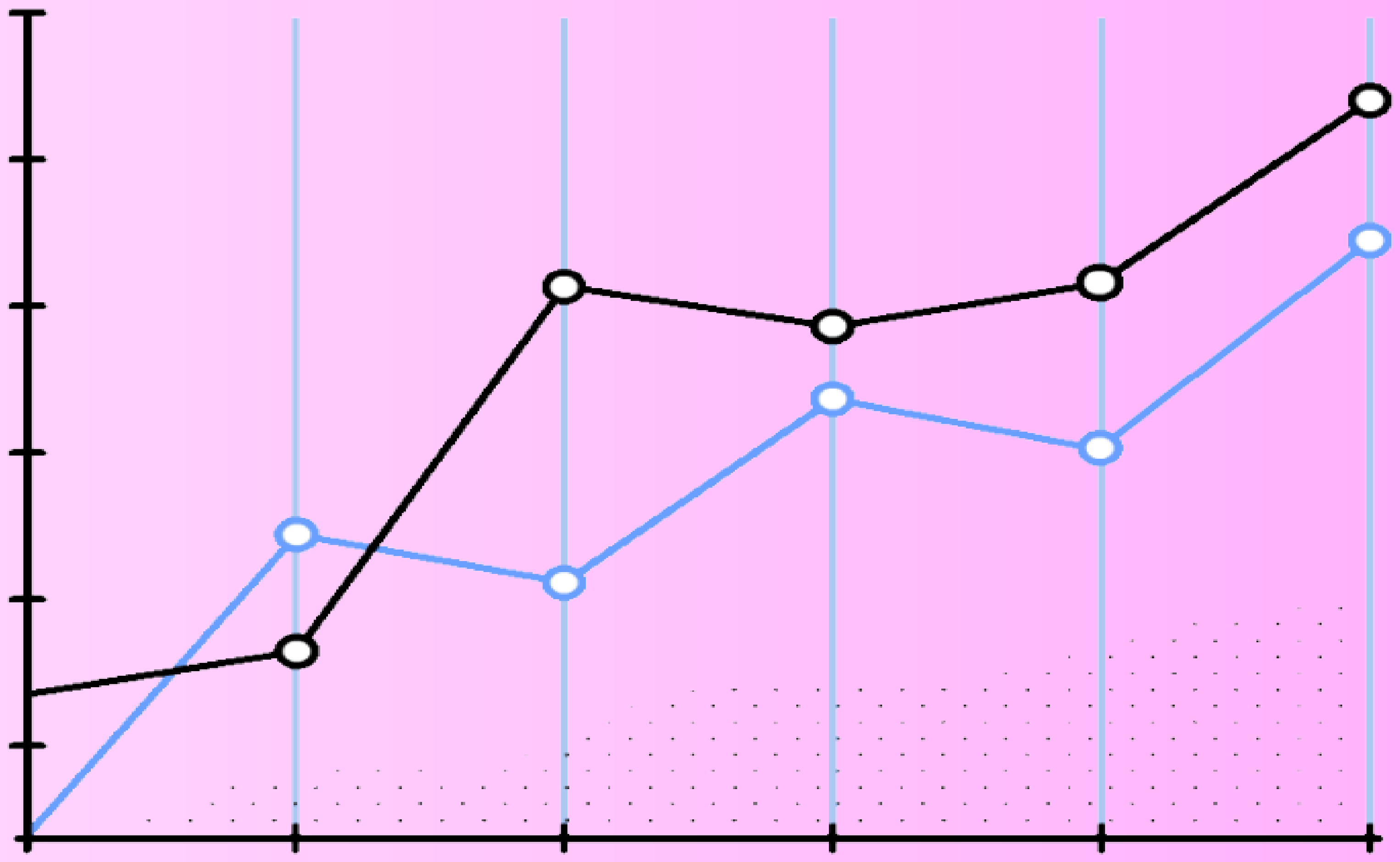
MODEL EVALUATION

Evaluate the final model on test data to validate its performance.

ANALYSIS AND MODEL OUTCOMES

RESULTS

In this section, we will present and review the key findings and outcomes of the analysis performed throughout the project. This includes results from our Exploratory Data Analysis (EDA) using visualizations, SQL queries, and the outcomes from the predictive analysis (classification models).



EDA WITH VISUALIZATION

Insights gained through scatter plots, bar charts, and line graphs to explore relationships between features such as payload mass, launch sites, and success rates.

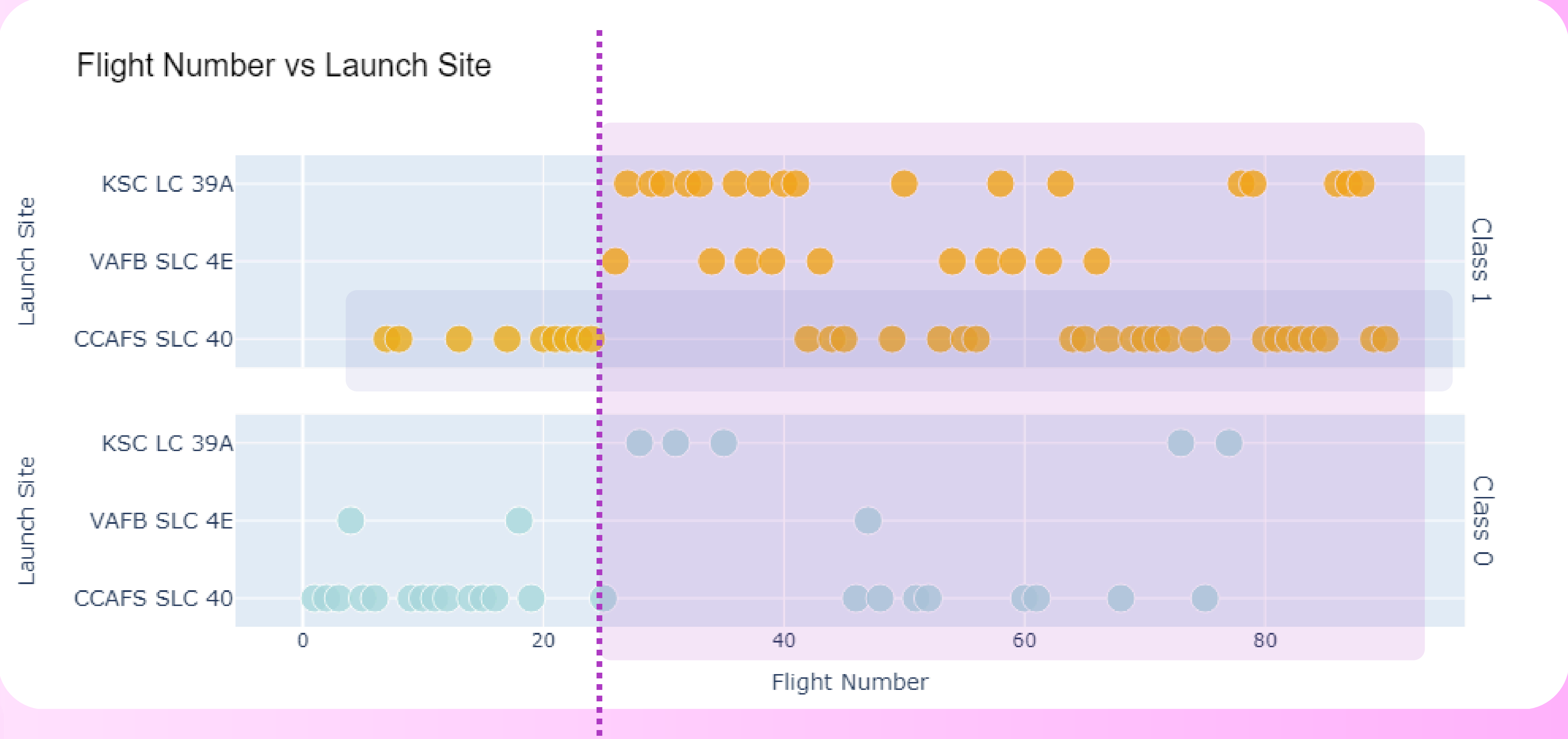
FLIGHT NUMBER BY LAUNCH SITE

EXPERT INSIGHTS

Class 1 (Successful Landings)

Success Increase Post-Flight 30
The trend suggests that from Flight Number 30 onwards, the majority of launches have resulted in successful landings. This can be attributed to SpaceX's refinement of both its Falcon 9 rocket systems and landing technologies over time.

High Success Rates at CCAFS SLC 40
The launch site at CCAFS SLC 40 demonstrates a high success rate for landings. The proximity to drone ships and favorable geographical conditions likely contribute to these results. The flat coastal environment offers fewer unpredictable atmospheric disruptions, making it optimal for re-entry trajectory calculations and fuel savings for reusability.



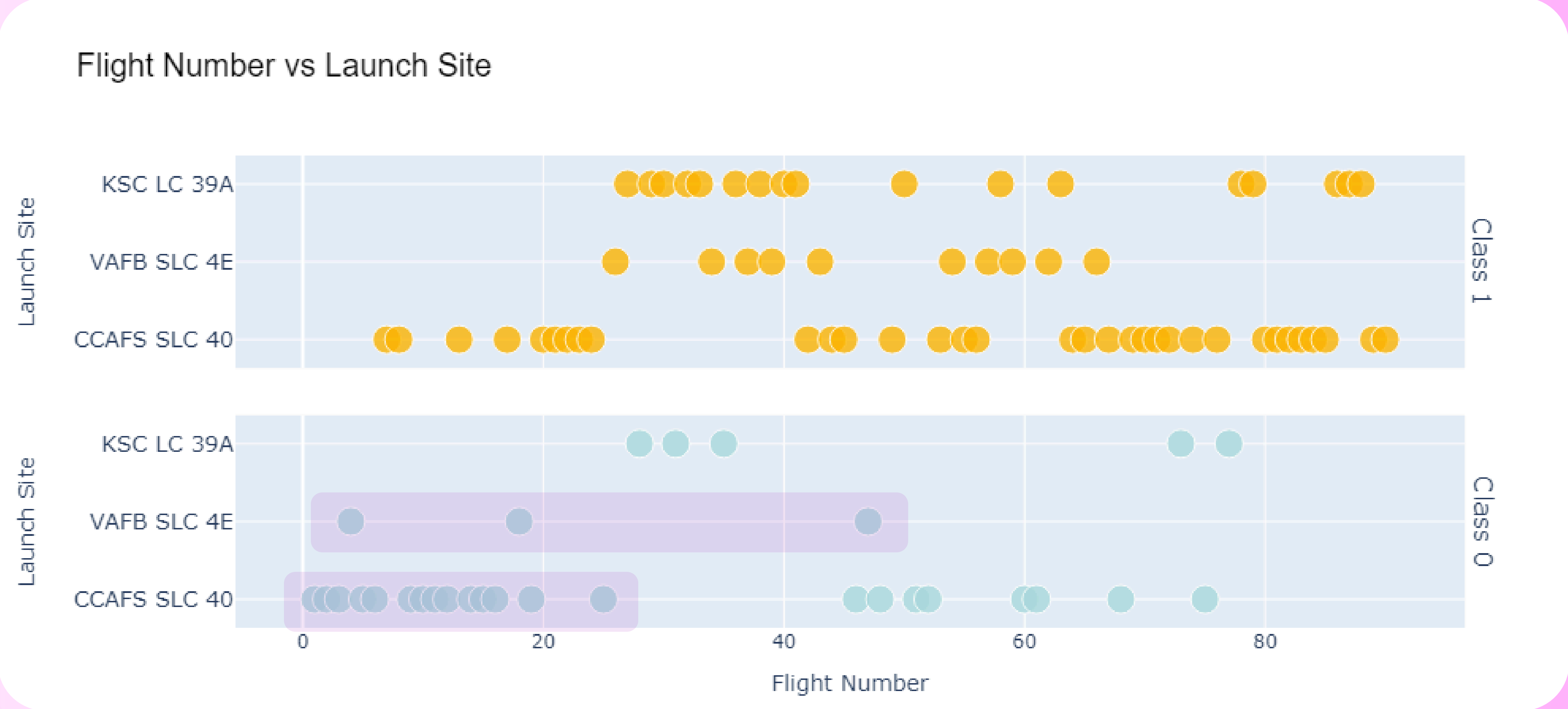
FLIGHT NUMBER BY LAUNCH SITE

EXPERT INSIGHTS

Class 0 (Failed Landings)

Early Launch Failures at CCAFS SLC 40
The concentration of failed landings in the early flight numbers, particularly at CCAFS SLC 40, suggests that initial missions were heavily experimental.

Fewer Failures at VAFB SLC 4E
Interestingly, VAFB SLC 4E has fewer recorded failures, indicating that while there are fewer launches, they are executed with more precision by the time they occur. The complexity of launching into polar orbits likely meant that only the more mature versions of Falcon 9 were used at this site, reducing the margin for error.

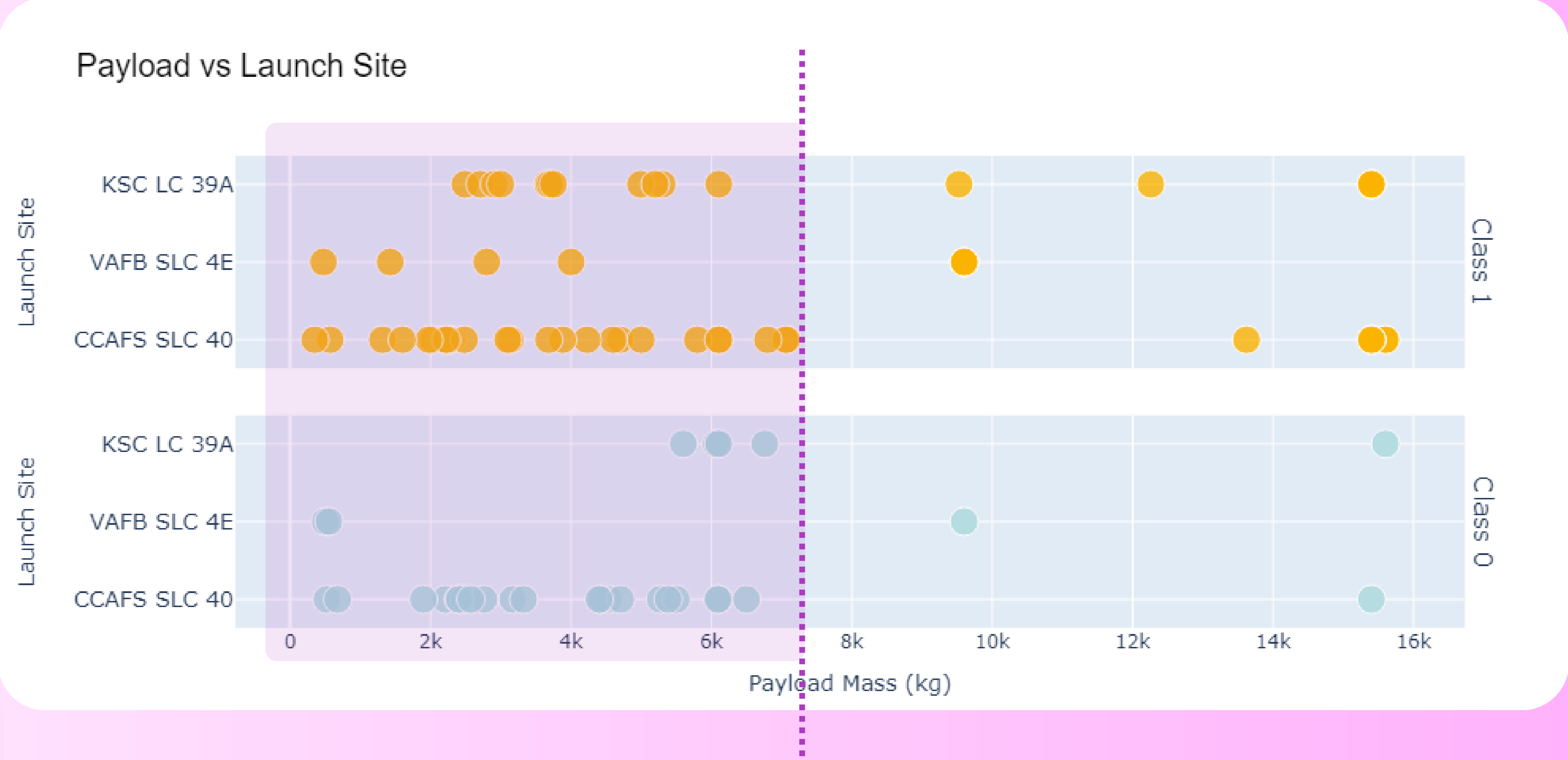


PAYLOAD MASS BY LAUNCH SITE

EXPERT INSIGHTS

Lower Payload, Higher Success
The data shows a trend where launches with lower payload masses (less than 8,000 kg) tend to have a higher rate of successful landings. The rocket's lower total mass leads to more flexibility during descent, allowing for finer control and corrections, which increases the likelihood of successful landings.

Inconsistent Pattern at higher Payloads
However, the success rates for higher payloads (above 10,000 kg) show a less consistent pattern, with successful landings still occurring at both KSC LC 39A and CCAFS SLC 40. This suggests that while heavier payloads introduce greater complexity in booster recovery.

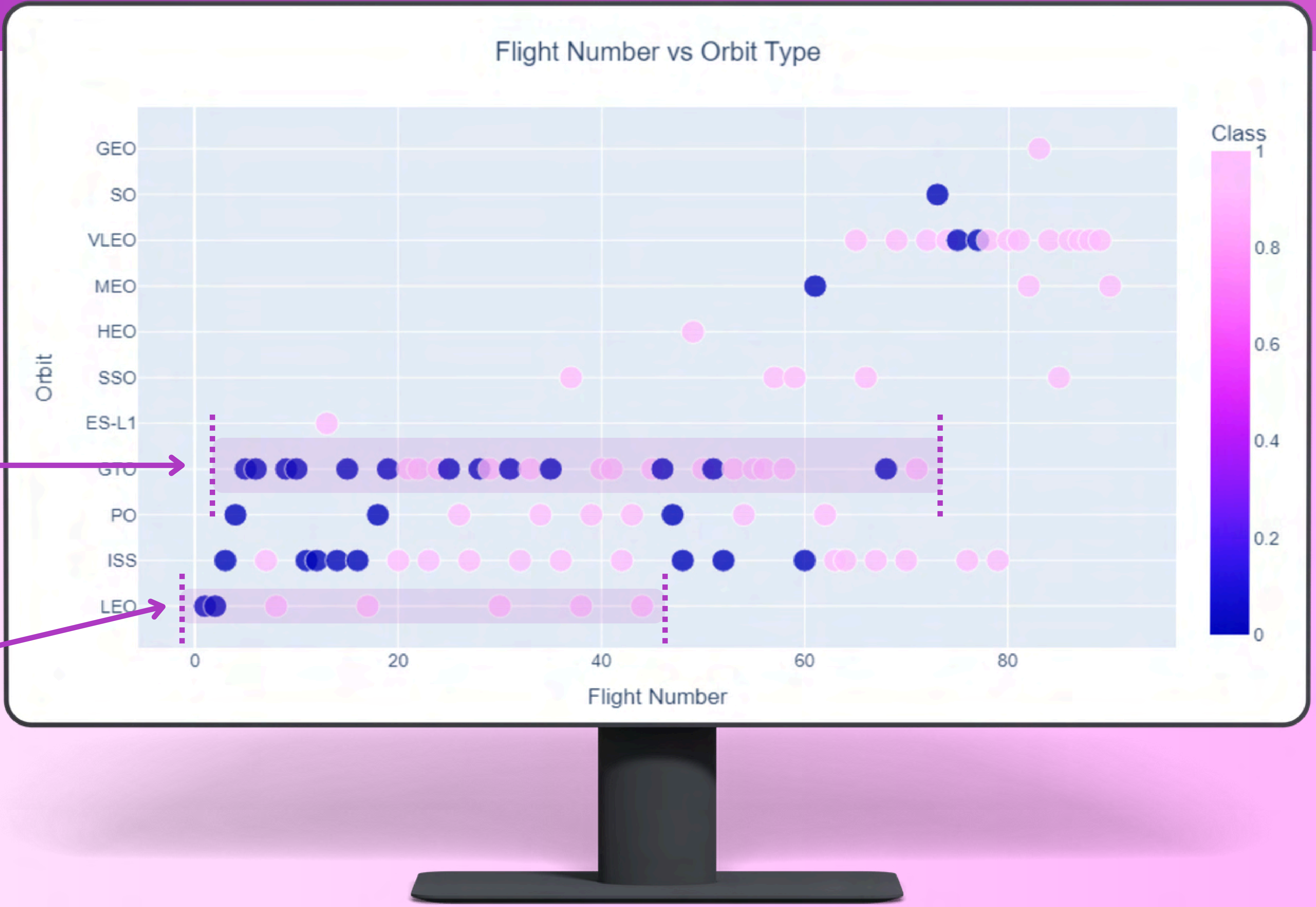


SUCCESS RATE OF EACH ORBIT TYPE

ADVANCED ANALYSIS

The orbits ES-L1, GEO, HEO, and SSO have achieved a 100% success rate, indicating SpaceX's mastery in handling missions targeting these specific orbits. These orbits, particularly GEO (Geostationary) and SSO (Sun-Synchronous Orbit), are critical for communication and earth observation satellites, respectively. The success in these orbits highlights SpaceX's advanced precision in orbital insertion and booster recovery during these high-stakes missions, where maintaining an exact altitude and position is crucial for the satellite’s operational lifetime.

FLIGHT NUMBER BY ORBIT TYPE



ADVANCED ANALYSIS

A clear upward trend is visible in **LEO**, where the number of **successful launches increases as the flight numbers progress**. This trend suggests that SpaceX has achieved greater reliability and refinement in its operations over time, particularly for LEO missions. These missions, which are typically less energy-intensive compared to higher orbits, benefit from the high frequency of launches such as Starlink deployments.

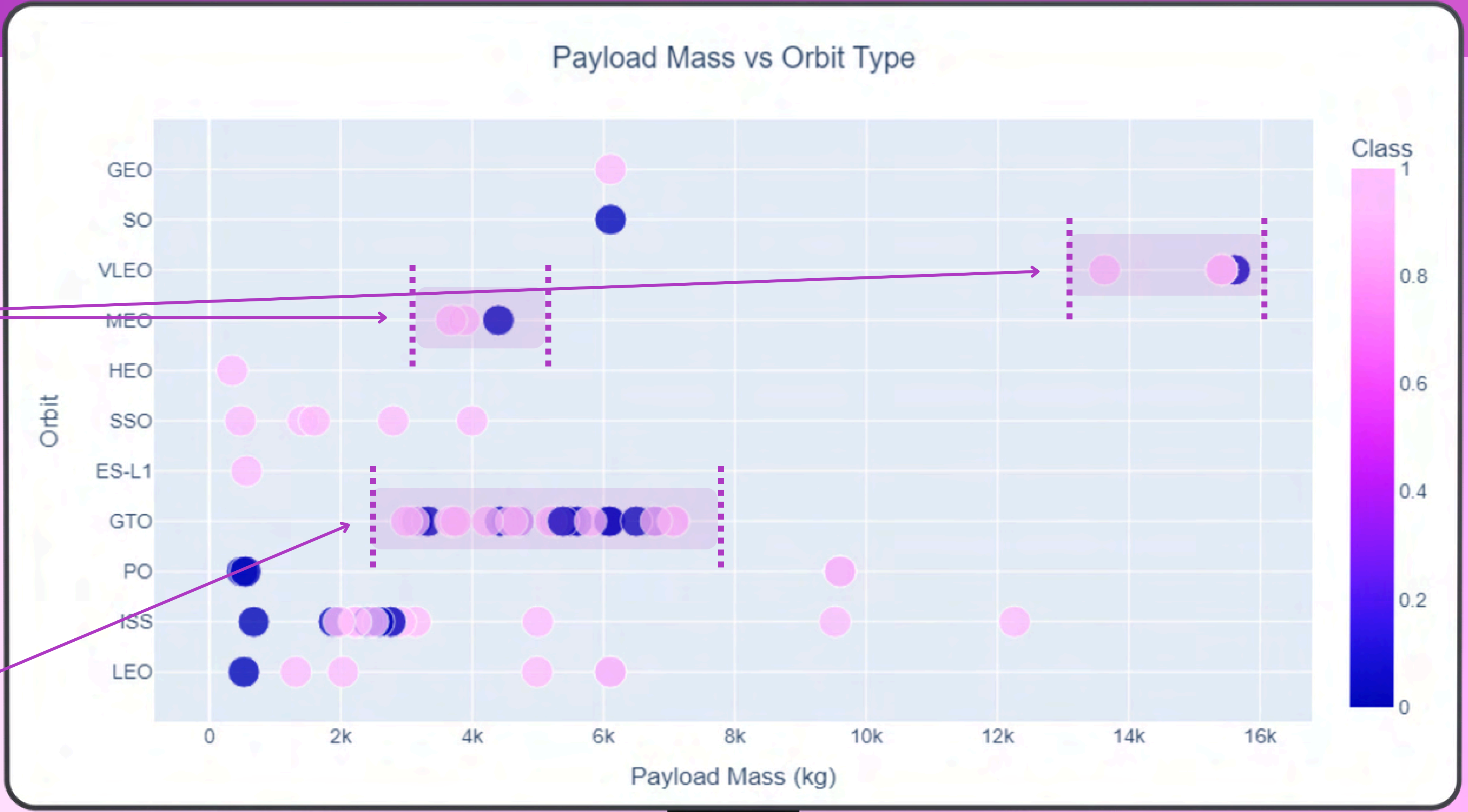
GTO presents a more erratic pattern, with no clear correlation between flight number and success rate. This could be due to the unique challenges posed by GTO missions, as these orbits require high-energy transfers to reach geostationary altitudes. GTO missions often involve heavier payloads and more complex upper-stage burns, leading to variable success outcomes.

PAYLOAD MASS BY ORBIT TYPE

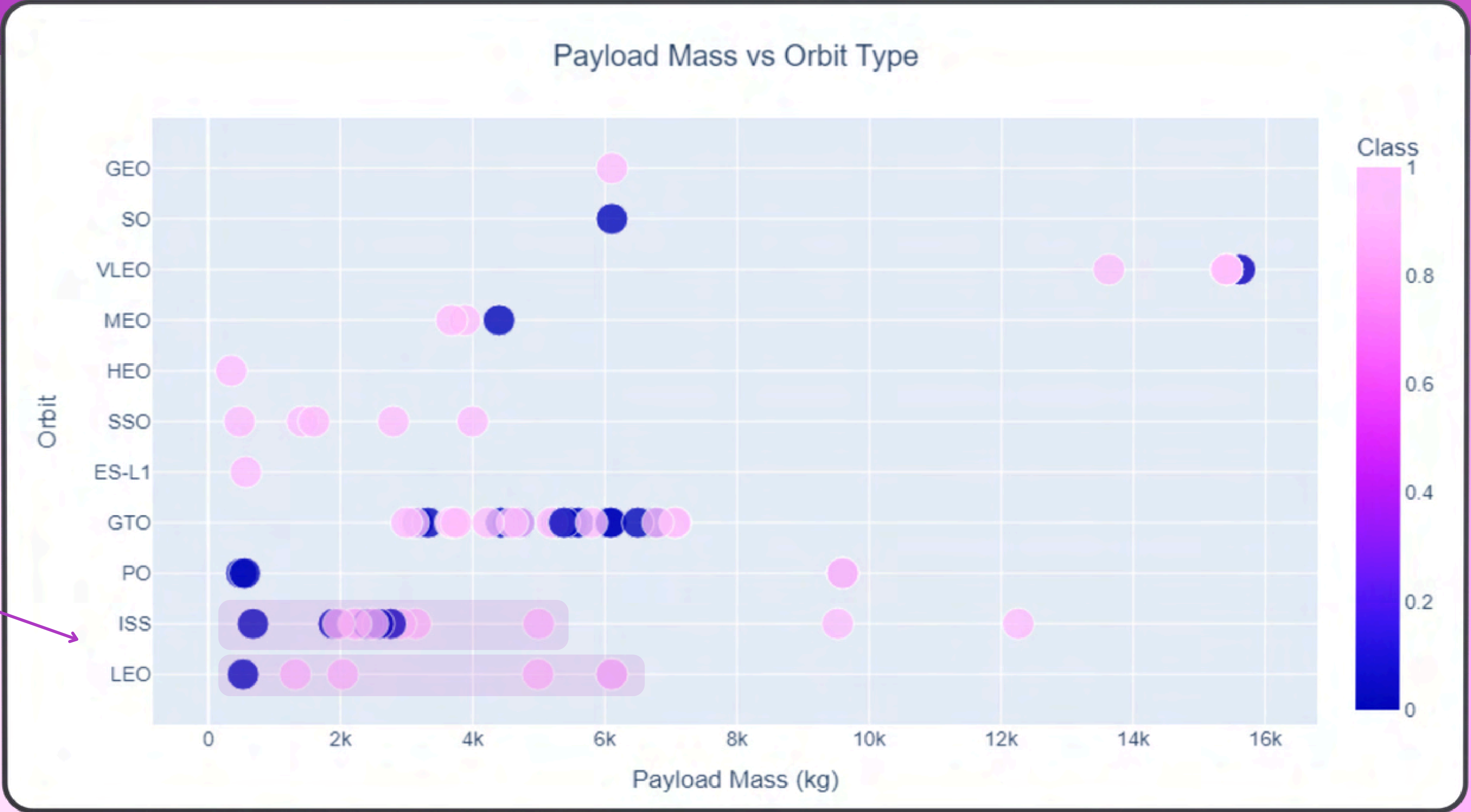
ADVANCED ANALYSIS I

For both **MEO** and **VLEO**, the data reveals a clear inverse relationship between payload mass and success rate. **As payload mass increases, the probability of mission success tends to decrease.** This correlation can be attributed to the heightened complexity and energy requirements associated with carrying heavier payloads to medium and very low orbits.

GTO continues to show **no clear pattern** with respect to payload mass and success rate. This orbit type, typically requiring high energy for the transfer, involves complex orbital dynamics that are less influenced by payload mass and more by the specific mission profile, trajectory adjustments, and upper-stage burns.



PAYLOAD MASS BY ORBIT TYPE



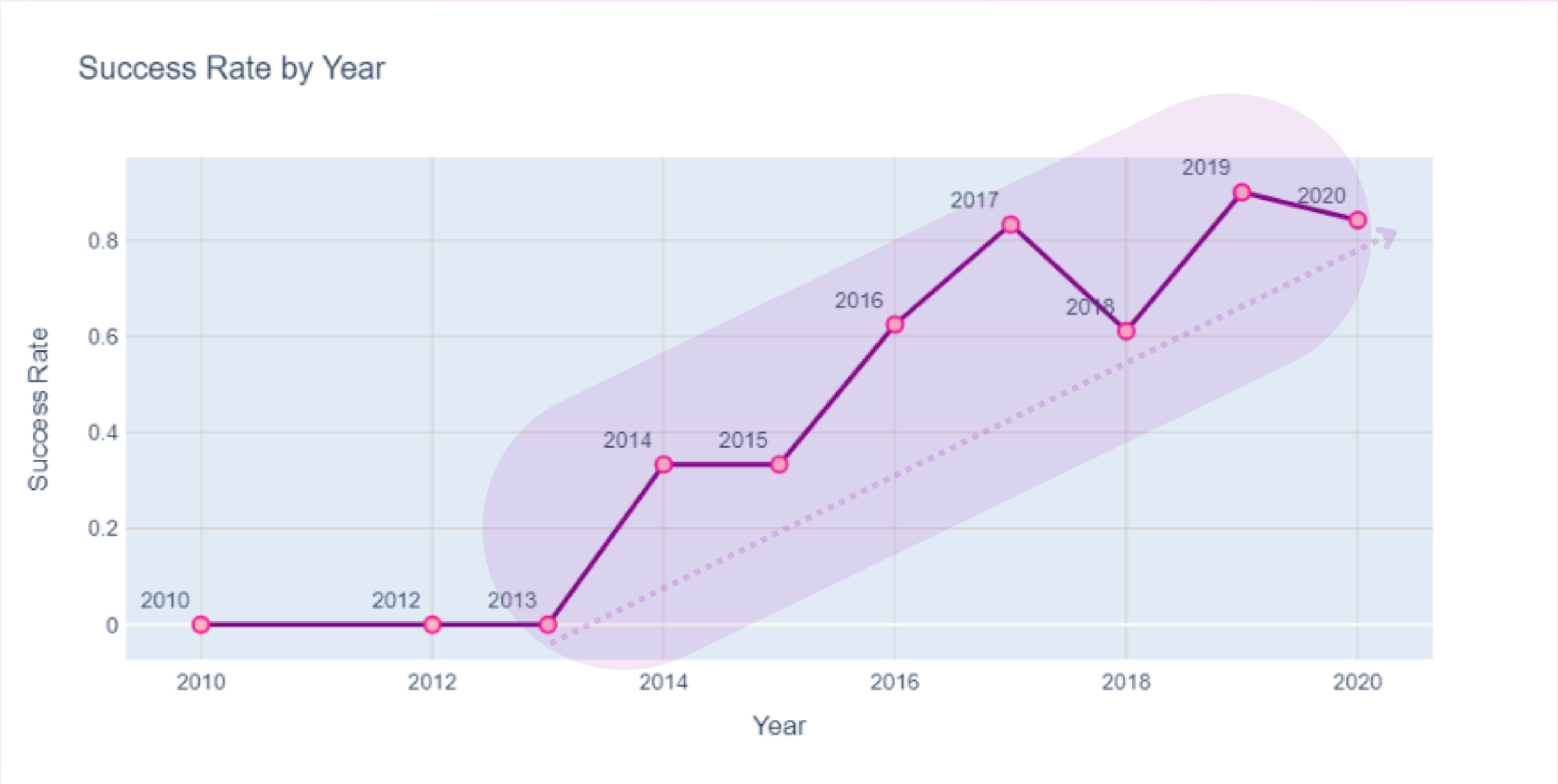
ADVANCED ANALYSIS II

Interestingly, both **LEO** and **ISS** missions show a positive correlation between payload mass and success rate. **As the payload mass increases, so does the probability of mission success.** This pattern likely reflects SpaceX's extensive experience with missions to these orbits, including routine cargo resupply missions to the ISS and frequent deployments of Starlink satellites to LEO.

LAUNCH SUCCESS YEARLY TREND

EXPERT ANALYSIS

From 2013 to 2020, there has been a steady and significant increase in SpaceX's launch success rate. This growth reflects the company's rapid technological advancements, system refinements, and operational optimizations over this period.



EDA WITH SQL

Key statistics derived from SQL queries, providing a deeper understanding of launch sites, payload trends, and mission outcomes.

UNIQUE LAUNCH SITES

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

Query

%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE

Description

This query is essential to identify all the distinct locations from which SpaceX missions were launched. By filtering out duplicates, it helps us focus on analyzing only the relevant and unique launch sites, allowing for a more accurate and streamlined exploration of site-specific performance and success rates.

LAUNCH SITES BEGIN WITH "CCA"

Query

%sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_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

LAUNCH SITES BEGIN WITH “CCA”

Description

We used the query SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5 to display records where the launch site name begins with "CCA." This query allows us to focus on launch sites associated with Cape Canaveral (identified by the "CCA" prefix).
By narrowing down the dataset to these specific sites, we can investigate patterns or trends that might be unique to launches from this region, providing deeper insights into their performance and outcomes. The LIMIT 5 ensures that we retrieve only a sample of the results for quick analysis.

TOTAL PAYLOAD MASS BY NASA (CRS)

Query

%sql SELECT SUM("PAYLOAD_MASS__KG_") as 'Total Payload Mass By NASA (CRS)' FROM SPACEXTABLE WHERE "Customer" = 'NASA (CRS)'

Total Payload Mass By NASA (CRS)
45596

Description

The query calculates the total payload mass transported by boosters for NASA's Commercial Resupply Services (CRS) missions. This is critical in understanding NASA's contribution to space station resupply efforts. By summing the payload mass, we gauge the scale of NASA's cargo deliveries, which is vital for evaluating mission success and logistical planning for the International Space Station (ISS) operations.

AVG. PAYLOAD MASS BY F9 V1.1

Query

%sql SELECT AVG("PAYLOAD_MASS__KG_") as 'Average Payload Mass by Booster Version F9 v1.1' FROM SPACEXTABLE where "Booster_Version" = 'F9 v1.1'

Average Payload Mass by Booster Version F9 v1.1
2928.4

Description

This query calculates the average payload mass carried by the Falcon 9 v1.1 booster version. Understanding the average mass it successfully delivers helps engineers and mission planners optimize future launches. This also allows for comparisons with other booster versions to evaluate efficiency and reliability across mission types.

FIRST SUCCESFUL GROUND LANDING DATE

Query

```
%sql SELECT MIN("Date") as 'First Succesful Landing Outcome in Ground Pad' FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)'
```

First Succesful Landing Outcome in Ground Pad
2015-12-22

Description

This query identifies the earliest recorded successful landing on a ground pad. This data is crucial because it marks a significant milestone in SpaceX's landing technology, transitioning from drone ships to stable ground landings. Tracking this date helps highlight the progress made in reusability, a core principle of SpaceX's business model for sustainable space exploration.

BOOSTERS WITH DRONE SHIP SUCCESS AND 4K-6K KG PAYLOAD

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

Query

%sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" BETWEEN 4000 AND 6000

Description

This query retrieves the names of the boosters that successfully landed on a drone ship while carrying payloads between 4000 kg and 6000 kg. This is a critical range in payload capacity, as it highlights boosters that are capable of medium-lift missions, which are often commercially viable and technologically challenging.

TOTAL COUNT OF SUCCESSFUL AND FAILED MISSION OUTCOMES

Successful	Failure_drone
100	1

Query

%sql SELECT COUNT(CASE WHEN MISSION_OUTCOME LIKE '%Success%' THEN 1 END) AS Successful, COUNT(CASE WHEN MISSION_OUTCOME like '%Failure%' THEN 1 END) AS Failure_drone FROM SPACEXTABLE;

TOTAL COUNT OF SUCCESSFUL AND FAILED MISSION OUTCOMES

Successful	Failure_drone
100	1

Description

This query is essential to understanding SpaceX's overall mission success rate. It provides a clear breakdown of successes versus failures, which allows us to assess the reliability of launches and mission outcomes. By analyzing both, we can draw technical conclusions about the continuous improvement of rocket technology, operational consistency, and risk mitigation strategies adopted by SpaceX over time.

BOOSTER VERSIONS WITH MAXIMUM PAYLOAD MASS

Booster Versions which carried the Max Payload Mass
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

Query

```
%sql SELECT DISTINCT "Booster_Version" as 'Booster Versions which carried the Max Payload Mass' FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE)
```

Description

This query utilizes a subquery to find the maximum payload mass and then retrieves the booster versions that carried it. It efficiently highlights the top-performing boosters based on payload capacity.

RANKED LANDING OUTCOMES (2010-2017)

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

Query

%sql SELECT SUBSTR(Date, 6, 2) AS Month, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTABLE WHERE SUBSTR(Date, 1, 4) = '2015' AND LANDING_OUTCOME = 'Failure (drone ship)';

RANKED LANDING OUTCOMES (2010-2017)

Month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

Description

This query helps to identify patterns in failed drone ship landings by highlighting which months in 2015 had more failures. Knowing the specific booster versions and launch sites involved in these failures provides technical insights into what configurations or conditions might have contributed to these outcomes.

BOOSTER VERSIONS WITH MAXIMUM PAYLOAD MASS

Landing_Outcome	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

Query

%sql SELECT LANDING_OUTCOME, COUNT(*) AS Outcome_Count FROM SPACEXTABLE WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING_OUTCOME ORDER BY Outcome_Count DESC;

Description

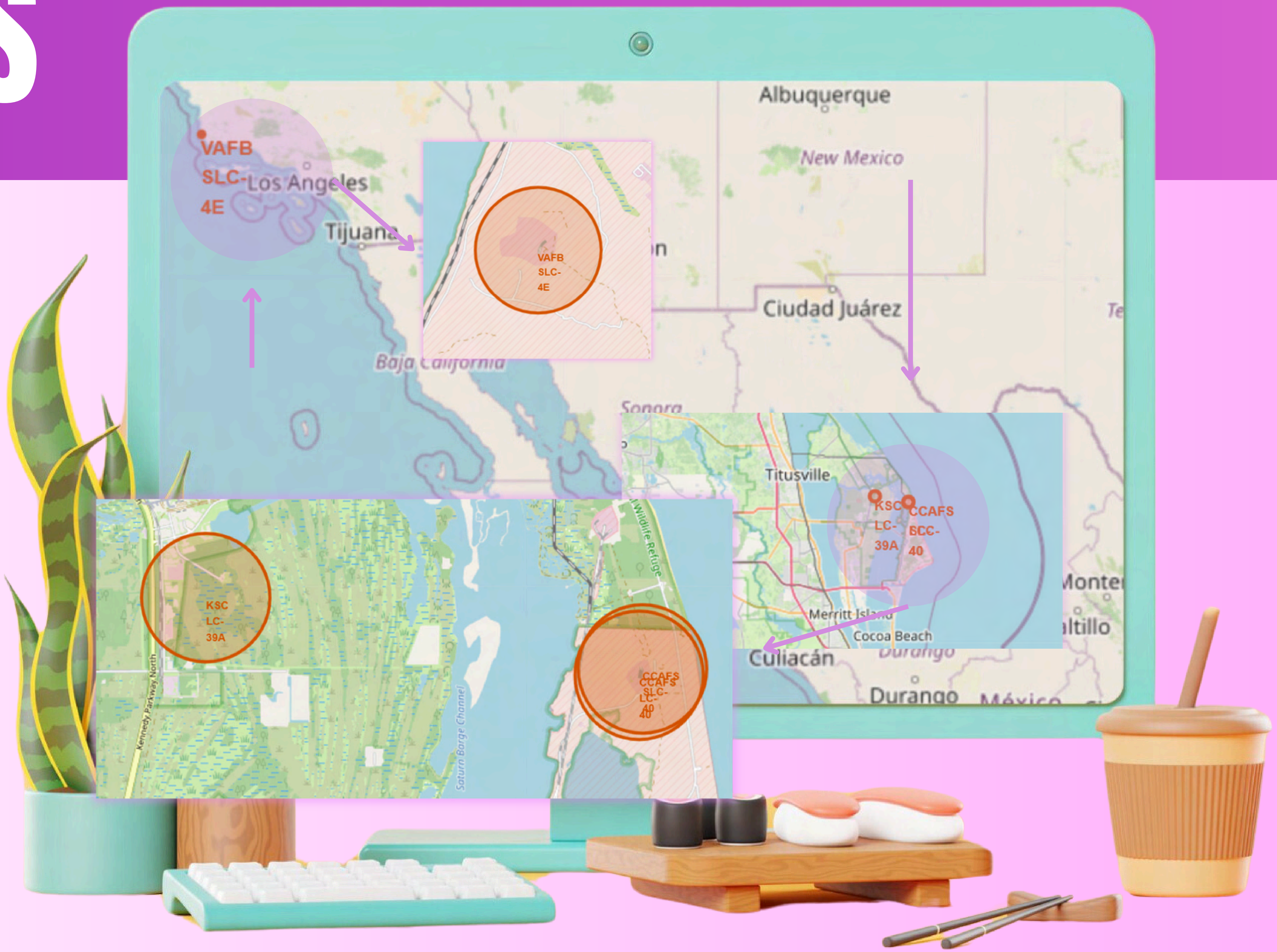
This query identifies the most frequent landing outcomes between 2010 and 2017 by grouping and counting them. It helps assess performance trends during a critical period for SpaceX landings.

INTERACTIVE MAP

Folium

Interactive maps created with Folium reveal geographical patterns and relationships between launch sites and their success rates.

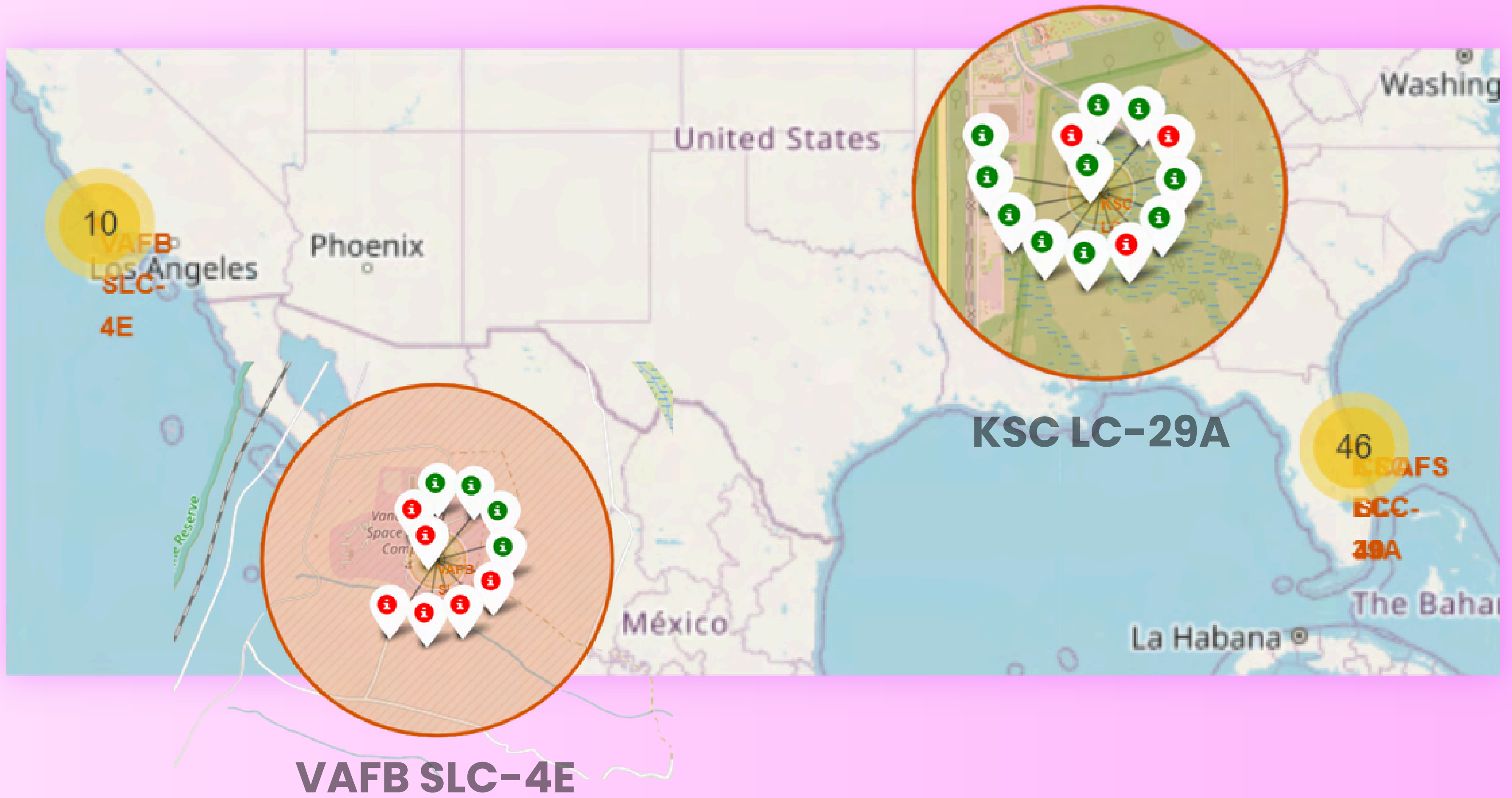
LAUNCH SITES



The analysis reveals that the majority of SpaceX launch sites are strategically located near the coastal regions of **Florida** and **California**. This geographical placement is crucial as it allows rockets to launch over open water, maximizing safety and efficiency. Additionally, the proximity to the equator in Florida's case provides an advantage for certain orbital launches, as it allows rockets to gain more velocity from Earth's rotation.

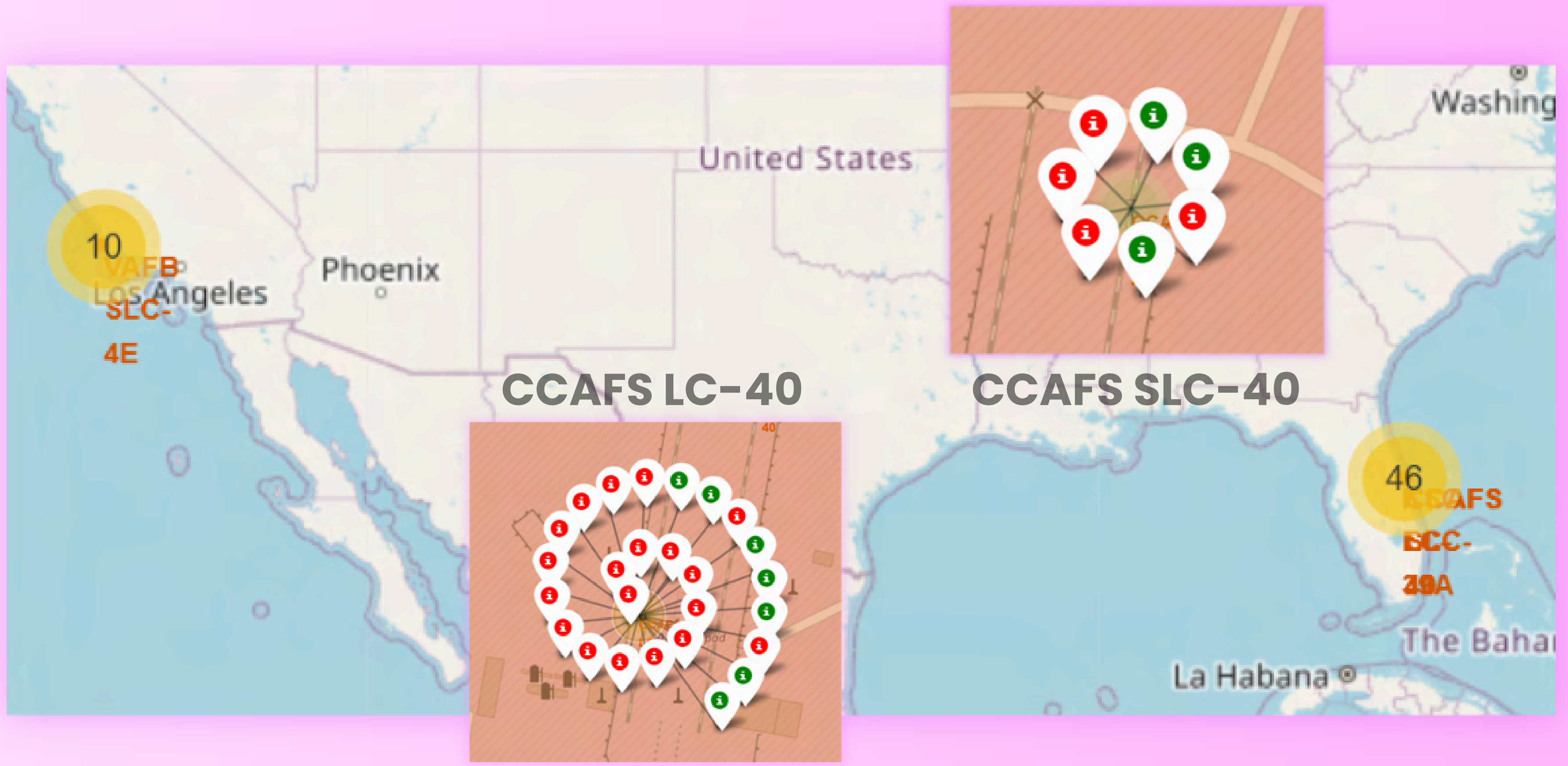
SUCCESS AND FAILED LAUNCHES FOR EACH SITE

For the second map, marking the success and failed launches for each site, the distribution of outcomes becomes visually clear through the color-coded markers. Successful launches (green) dominate the map, especially concentrated around **KSC LC39A**, highlighting it as the site with the highest success rate.



SUCCESS AND FAILED LAUNCHES FOR EACH SITE

For the second map, marking the success and failed launches for each site, the distribution of outcomes becomes visually clear through the color-coded markers. Successful launches (green) dominate the map, especially concentrated around **KSC LC39A**, highlighting it as the site with the highest success rate.

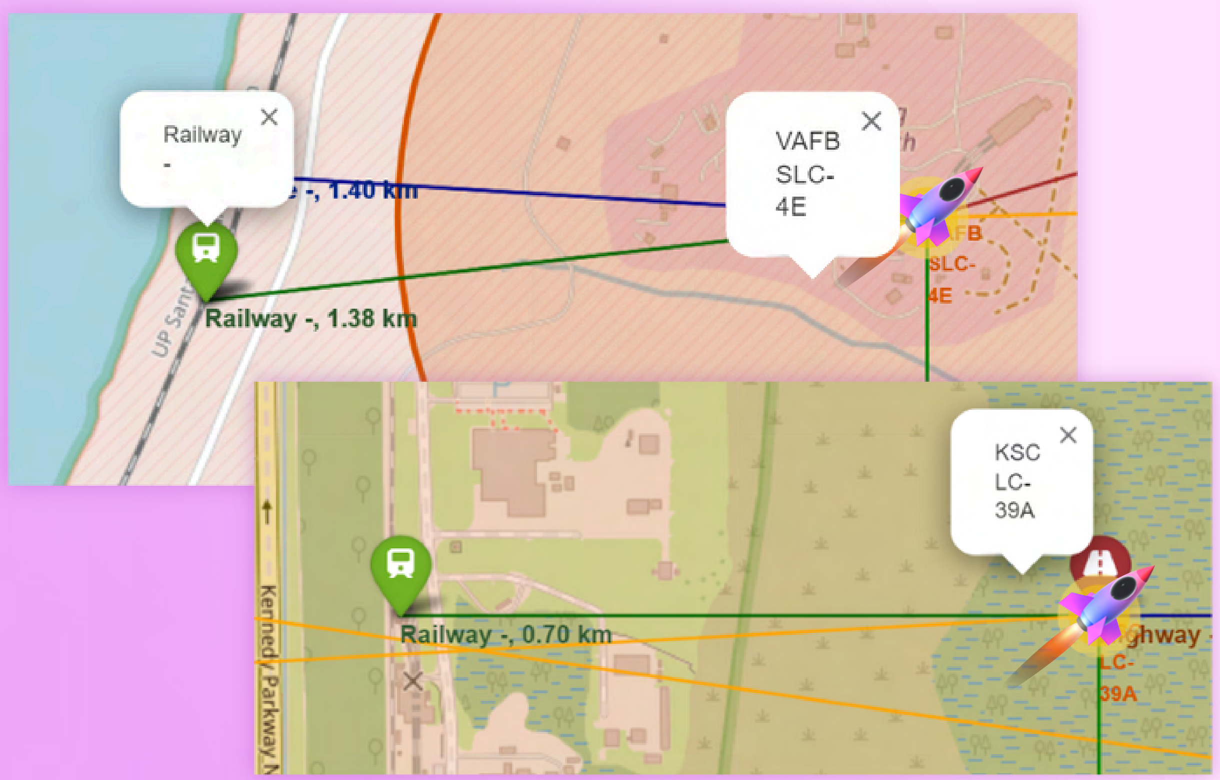


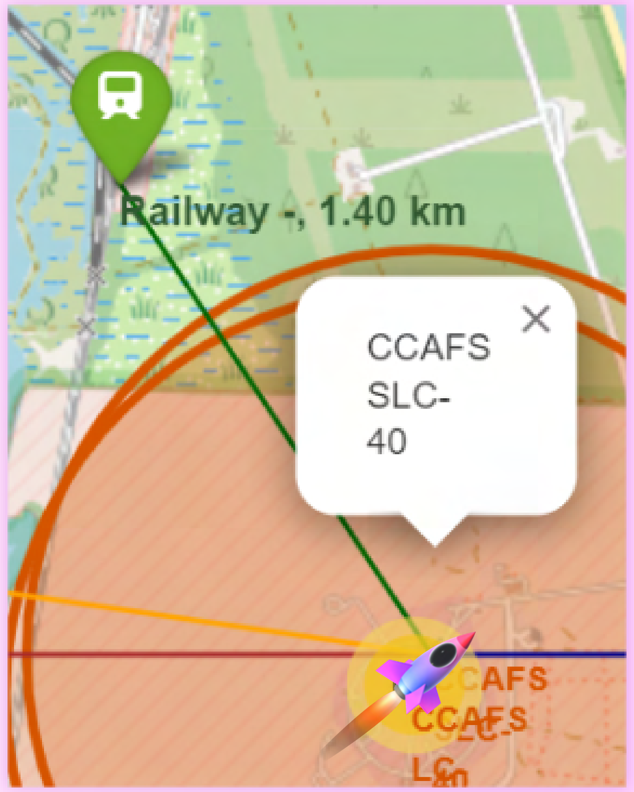
LAUNCH SITES DISTANCES

DISTANCE TO EQUATOR

The finding that all launch sites are over 3000 kilometers from the equator can be attributed to safety considerations, as sites are typically located in remote areas to minimize risks to populated regions.

LAUNCH SITES DISTANCES

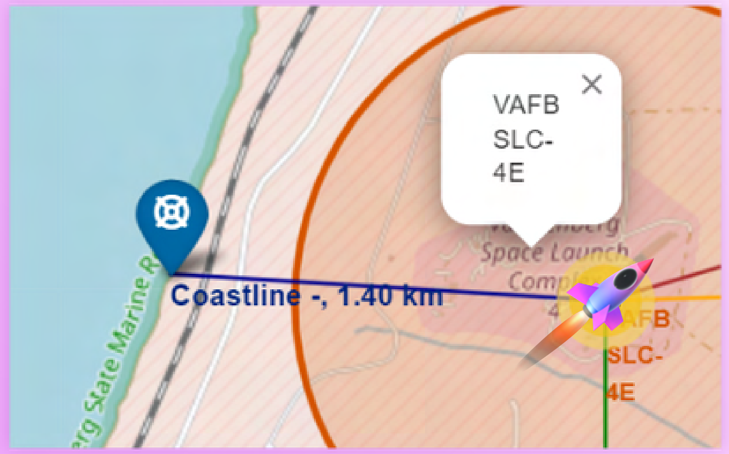


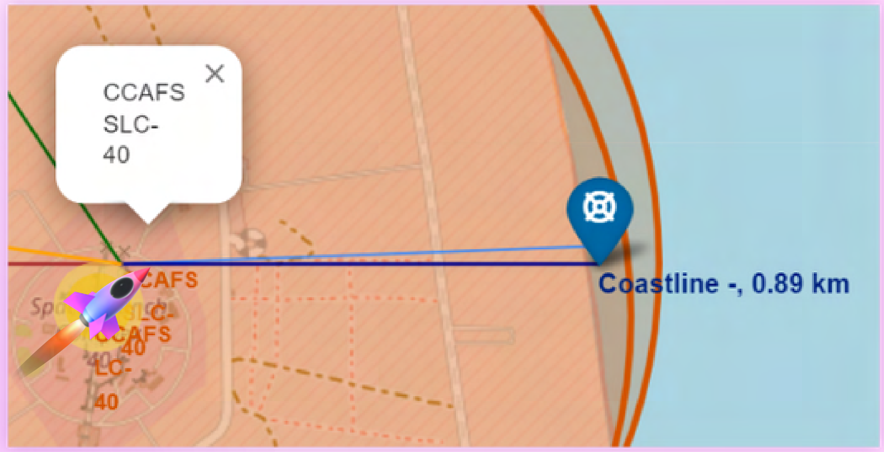


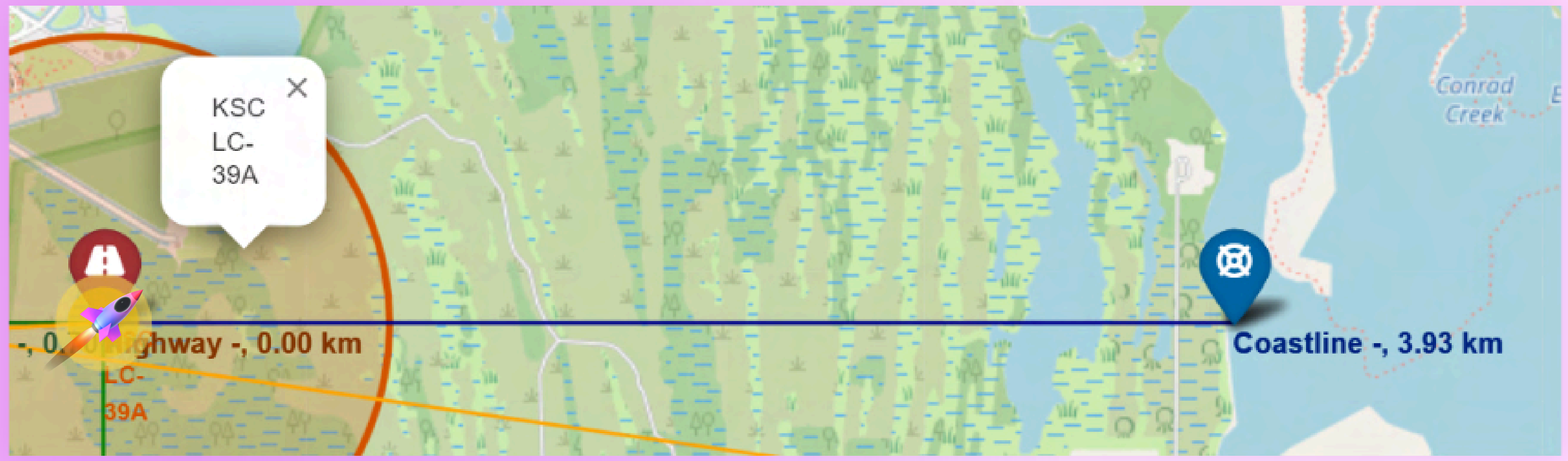
DISTANCE TO RAILWAYS

The distances to railways, ranging from **0.70 km to less than 2 km**, reflect relatively short distances. This could be due to the need for logistical support and efficient transport links, allowing for the quick movement of equipment and personnel to the launch sites.

LAUNCH SITES DISTANCES



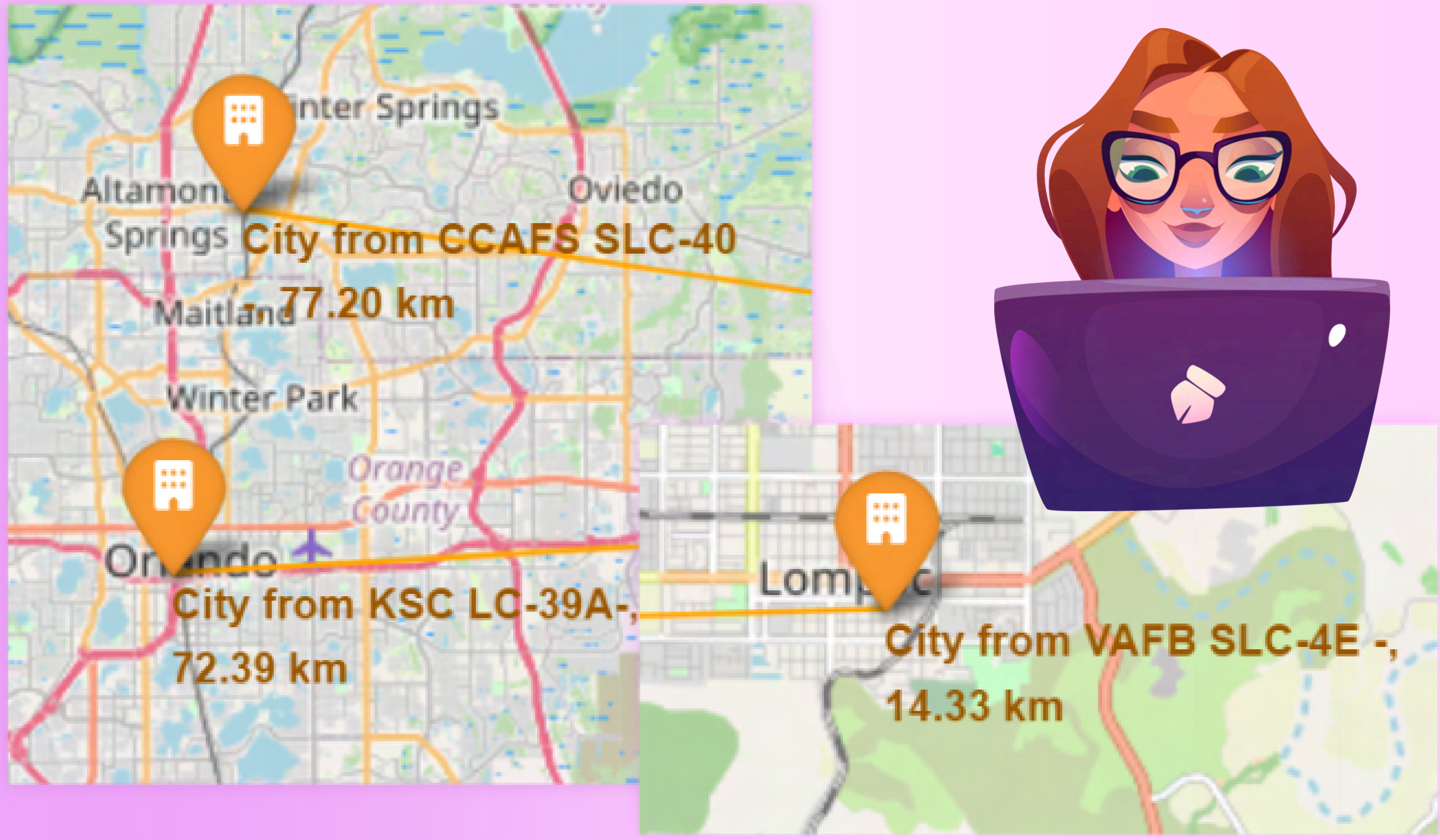




DISTANCE TO COASTLINES

The distances to coastlines, ranging **from 0.70 km to less than 4 km**, reflect relatively short distances. This proximity is likely due to safety protocols, as launching rockets near open water reduces the risk to populated areas in the event of a failure during launch or ascent.

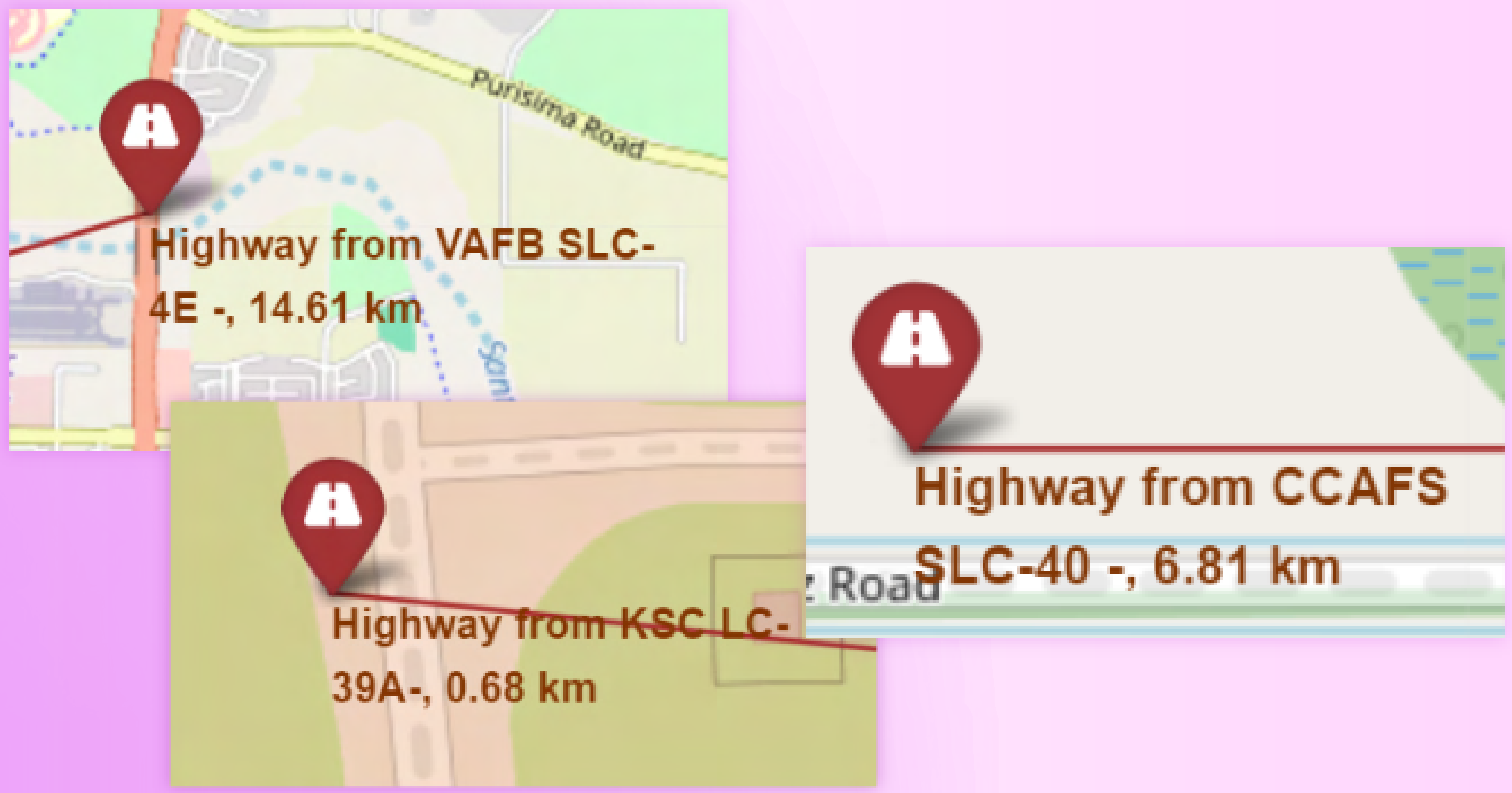
LAUNCH SITES DISTANCES



DISTANCE TO CITIES

The distances to cities are **over 14 km, with launch sites near Florida being more than 70 km away**. This greater distance is likely due to stringent safety regulations, ensuring that populated areas are far enough to minimize risk in case of launch anomalies or failures, particularly in high-traffic regions like Florida.

LAUNCH SITES DISTANCES



DISTANCE TO HIGHWAYS

The distances to highways range **from 0 to 15 km**. This variation is likely due to the need for convenient access to transportation infrastructure, allowing for the efficient movement of materials, equipment, and personnel, while also maintaining a safe buffer zone to minimize risks during launches.

PROXIMITY ANALYSIS OF SPACEX LAUNCH SITES

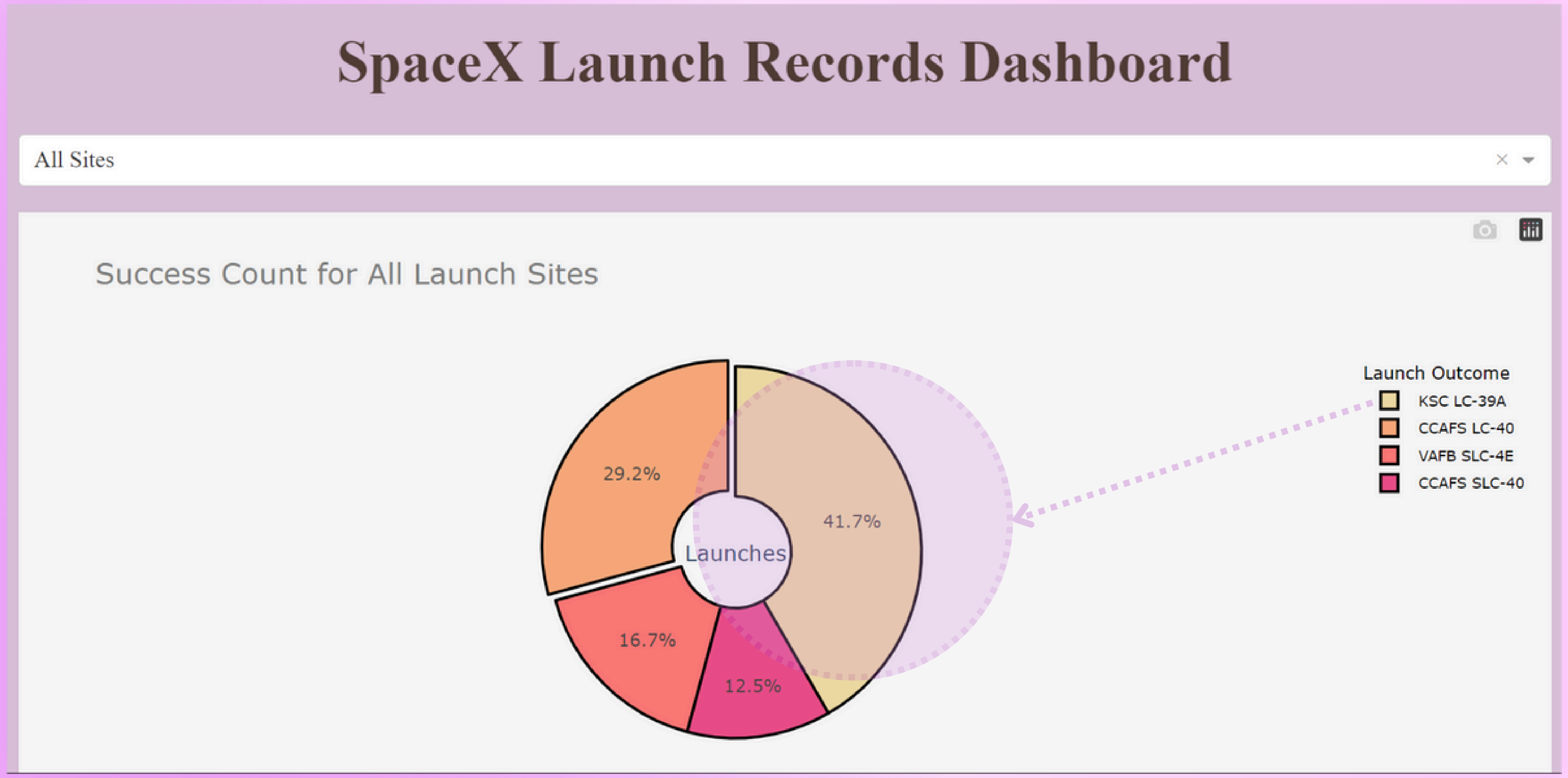
- **Are launch sites in close proximity to railways?** Yes, all launch sites are within 0.7 km to 2 km of railways, allowing for efficient transportation of equipment and materials.
- **Are launch sites in close proximity to highways?** Yes, distances to highways range from 0 to 15 km, providing quick access for logistical and operational needs.
- **Are launch sites in close proximity to coastlines?** Yes, launch sites are strategically placed 0.7 km to 4 km from coastlines, minimizing risks to populated areas.
- **Do launch sites keep a certain distance away from cities?** Yes, all sites are significantly distanced from cities, with a minimum of 14 km and up to 70 km, ensuring safety in case of emergencies.

INTERACTIVE DASHBOARDS

plotly

Interactive dashboards created to analyze and visualize SpaceX launch data, providing insights into the relationship between launch sites, payload mass, and mission outcomes.

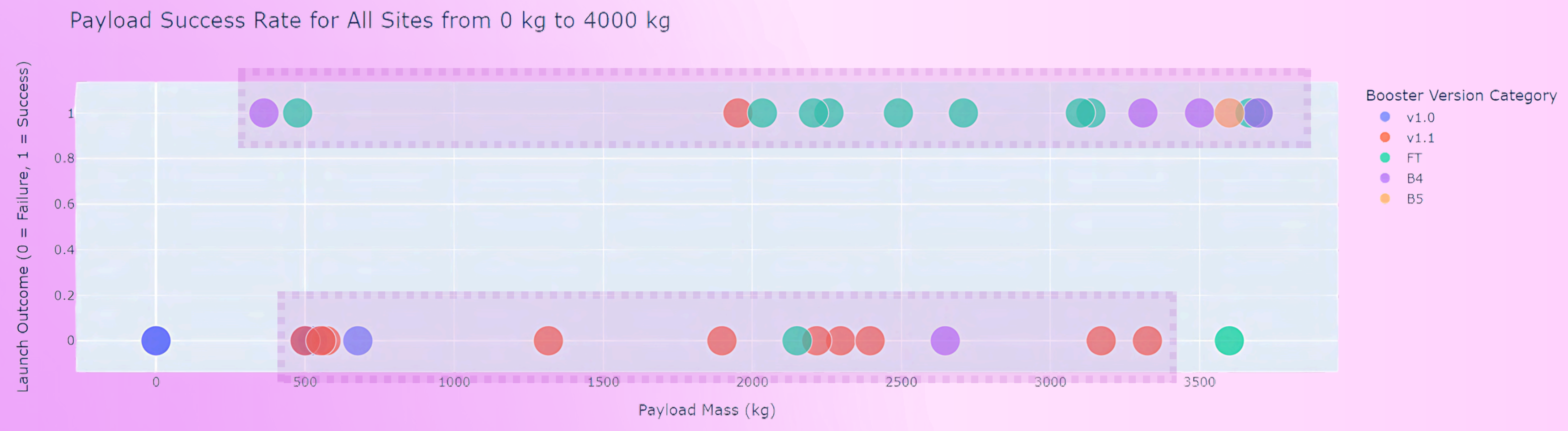
LAUNCH SUCCESS DISTRIBUTION

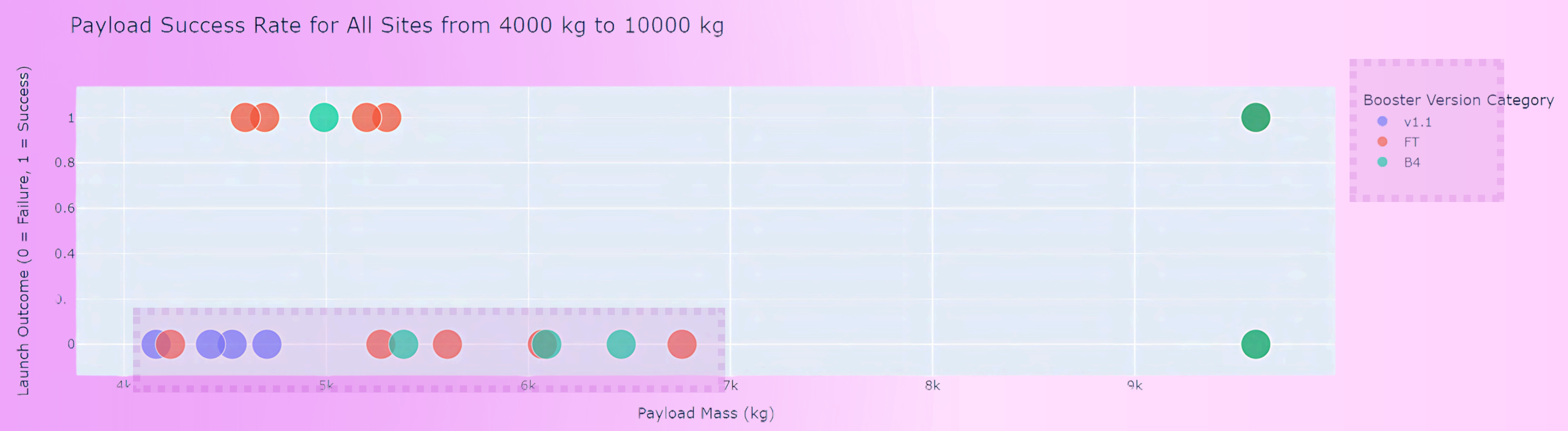


EXPERT ANALYSIS

From our pie chart, it is observed that **KSC LC-39A has the highest percentage** of successful launches. This may be due to it being one of SpaceX's most frequently used sites, equipped with advanced infrastructure that supports complex and high-frequency missions, thereby enhancing the likelihood of success. Additionally, its strategic location and favorable geographical conditions might contribute to these successful outcomes.

PAYLOAD MASS VS LAUNCH SUCCESS BY BOOSTER VERSION

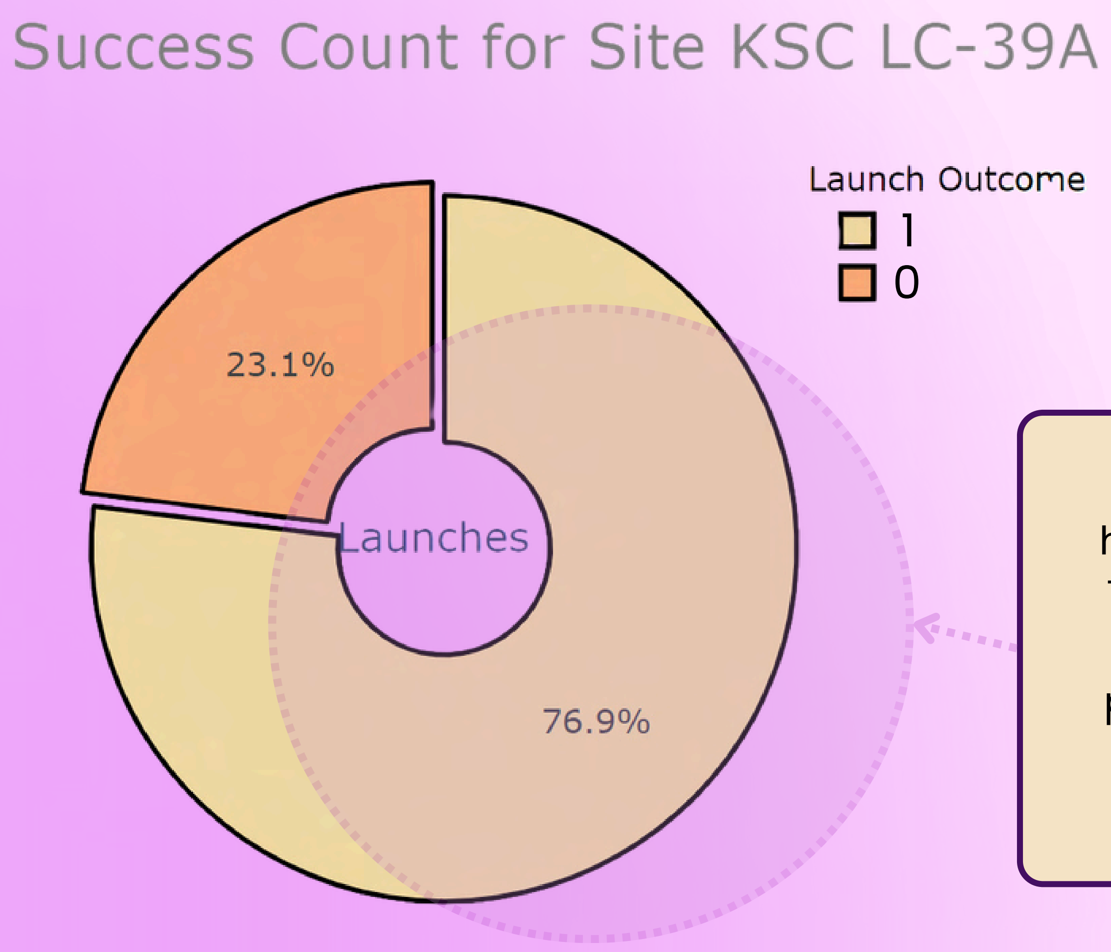




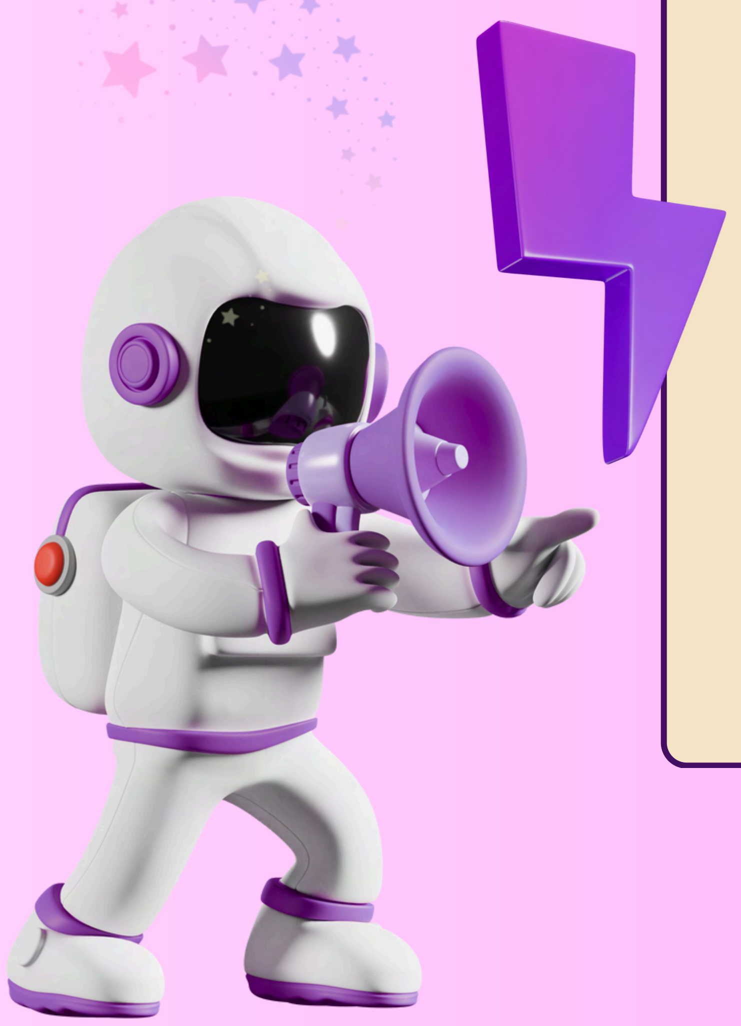
EXPERT ANALYSIS

The results demonstrate a clear **correlation between lower payload mass and higher success** rates in launches. The Falcon 9 Full Thrust (FT) booster version is observed to have the highest number of successful launches in the dataset. Additionally, it is important to note that there are no recorded successful launches for Falcon 9 v1.0 or Falcon 9 Block 5 with payloads exceeding 4000 kg.

TOP SUCCESS LAUNCH SITE



The results indicate that **KSC LC-39A** holds the highest success rate at 76.9%. This may be attributed to its advanced infrastructure, optimized logistics, and proximity to technical support facilities, ensuring minimal delays and a higher chance of successful launches.



RESULTS

Highest success launch site:
KSC LC-39A

Payload range with most success:
2000 kg to 6000 kg

Payload range with lowest success:
7000 kg to 10,000 kg

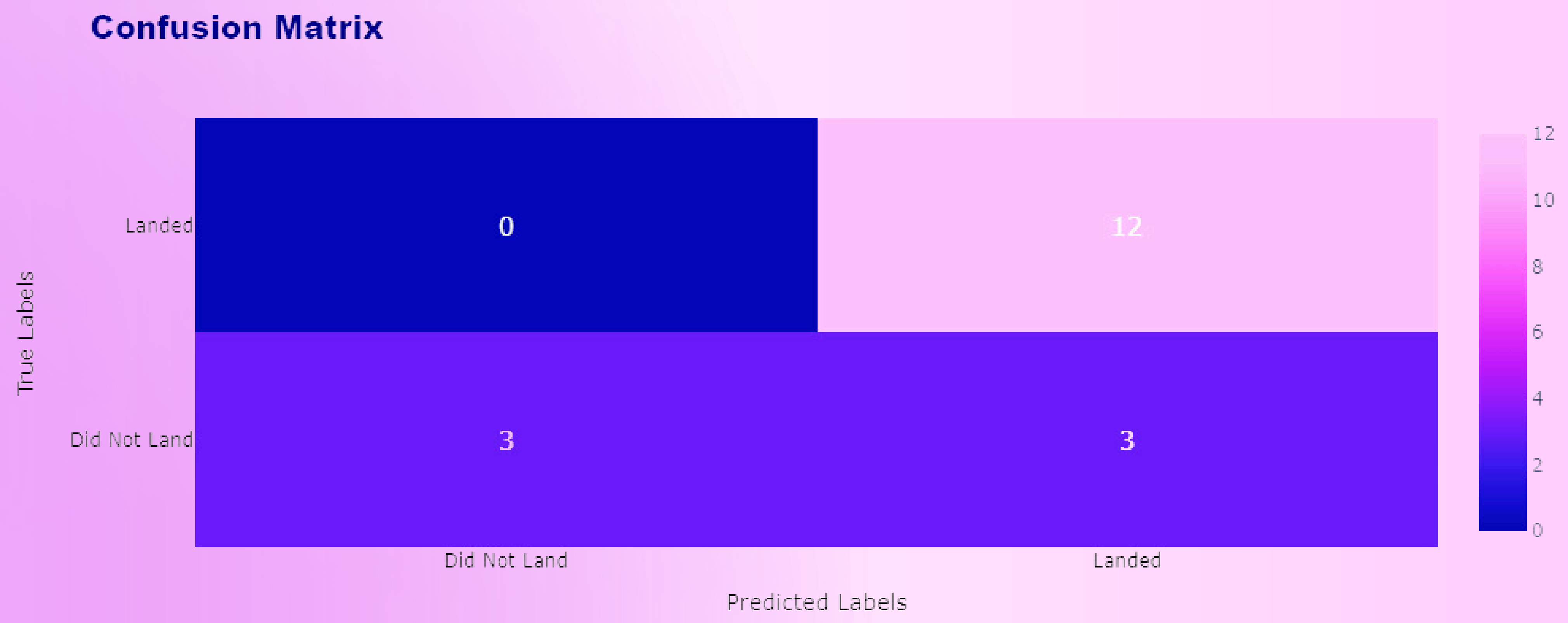
Booster version with most success:
Falcon 9 FT

ML PREDICTION FOR SPACEX LAUNCH OUTCOMES

An advanced predictive model was developed to forecast SpaceX mission outcomes. After hyperparameter tuning across Support Vector Machines (SVM), Decision Trees, and Logistic Regression, the model with the best predictive accuracy on test data was selected for final deployment.

LAUNCH SUCCESS DISTRIBUTION

Logistic Regresion

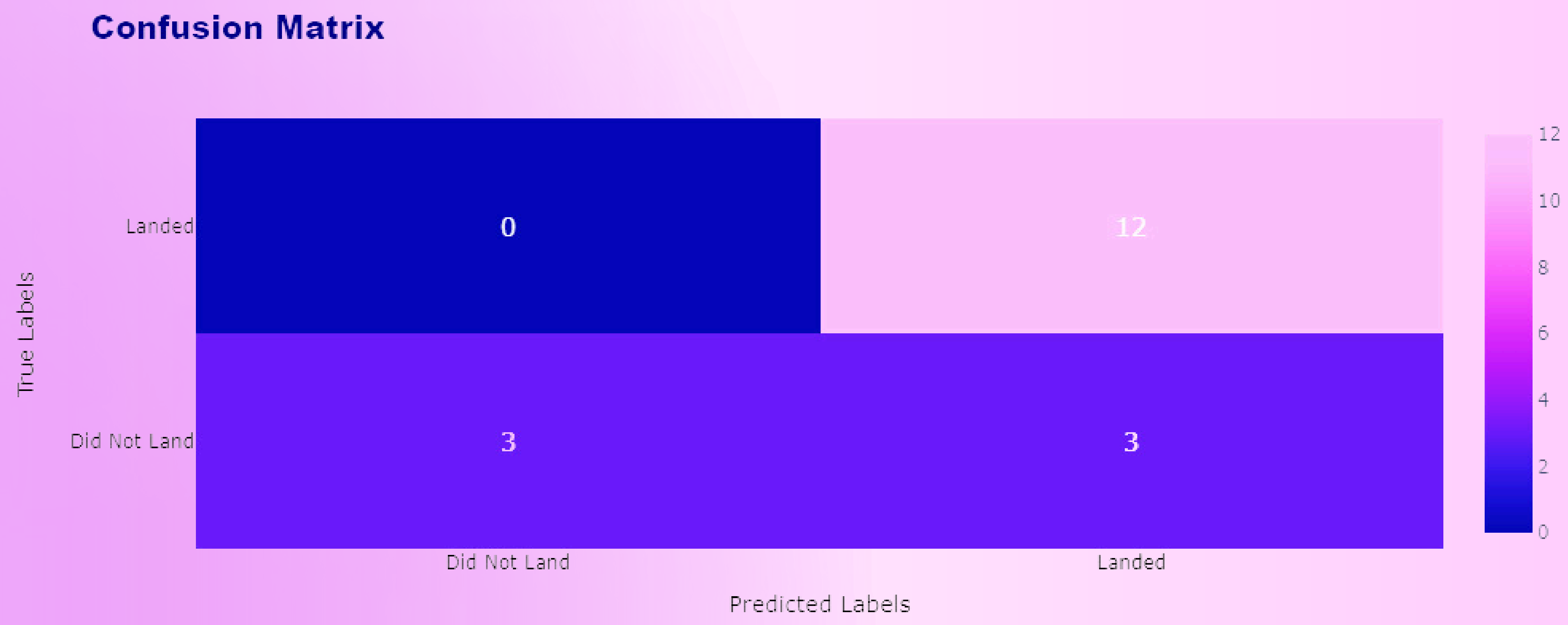


Tuned Hpyerparameters	Value
c	0.01
penalty	l2
solver	lbfgs

Best Score	0.846429

LAUNCH SUCCESS DISTRIBUTION

Support Vector Machine (SVM)

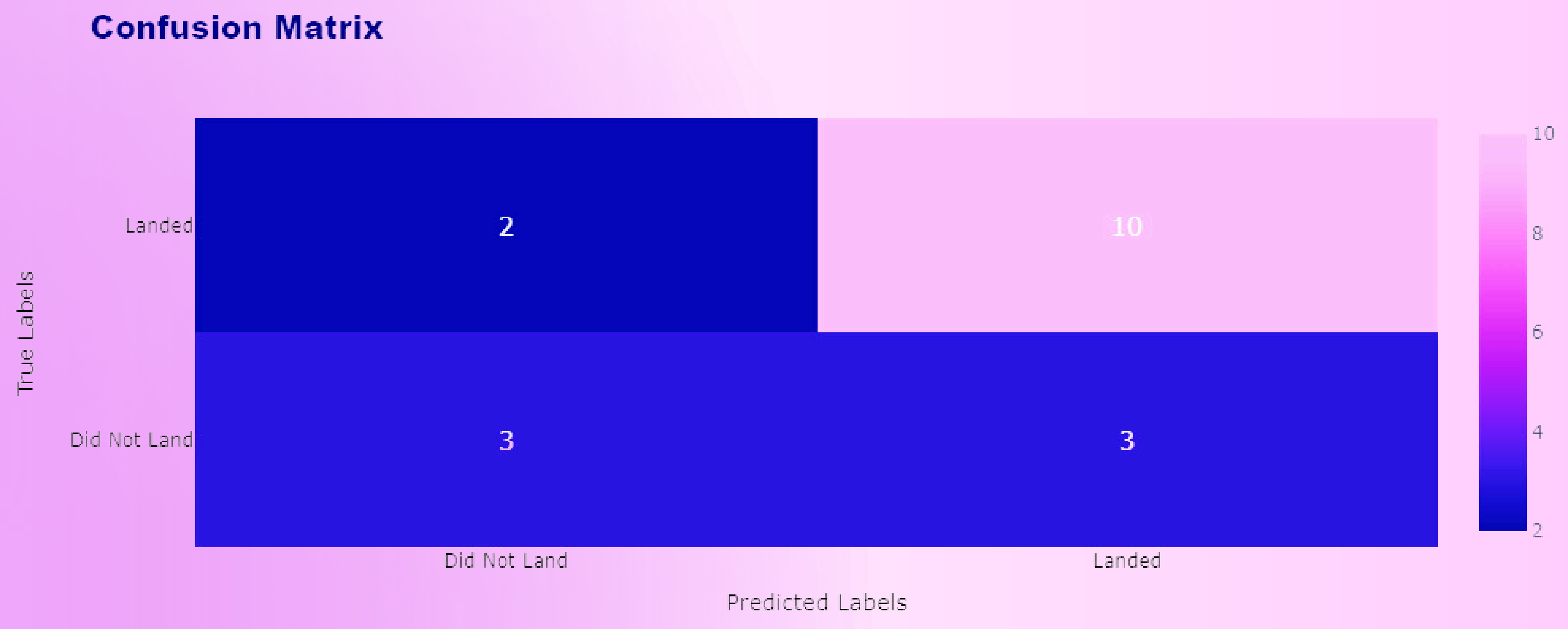


Tuned Hpyerparameters	Value
c	1.0
gamma	0.031623
kernel	sigmoid

Best Score	0.848214

LAUNCH SUCCESS DISTRIBUTION

Decision Tree

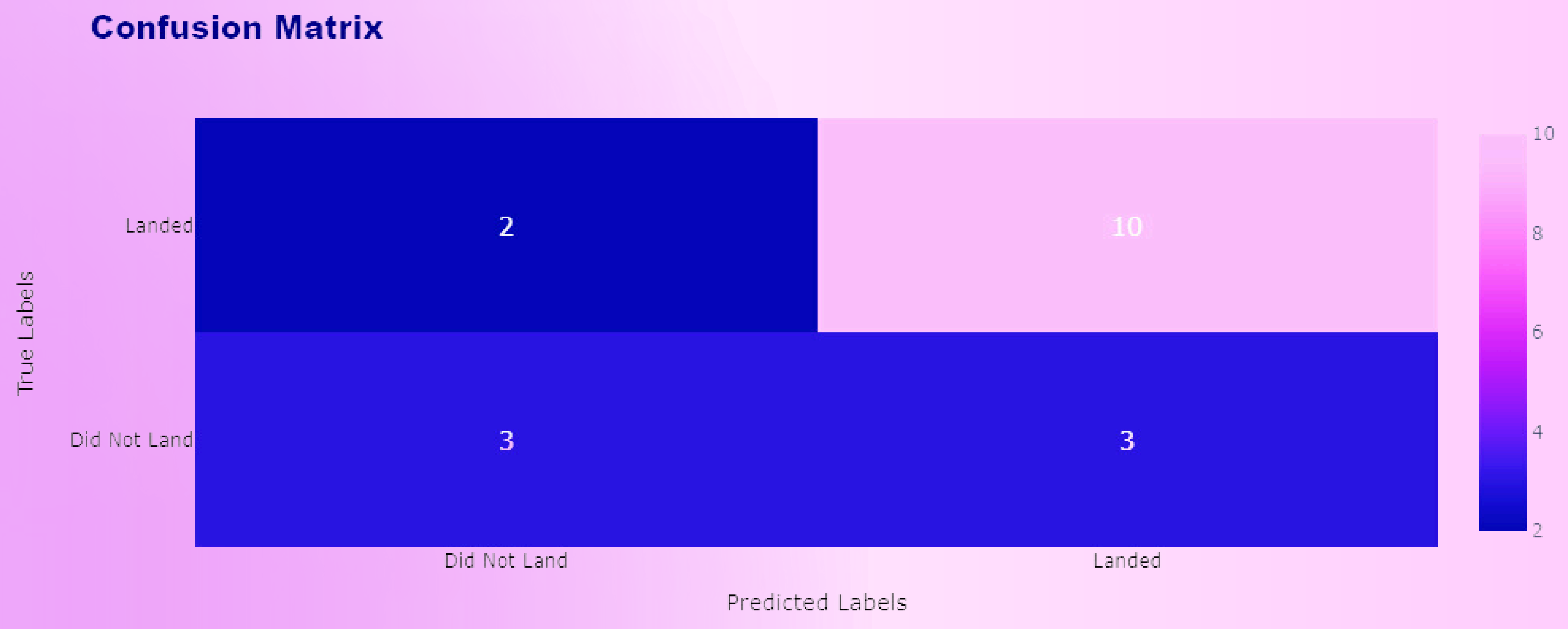


Tuned Hpyerparameters	Value
criterion	entropy
max_depth	6
max_features	sqrt
min_samples_leaf	1
min_samples_split	5
splitter	best

Best Score	0.876786

LAUNCH SUCCESS DISTRIBUTION

K Nearest Neighbors (KNN)



Tuned Hpyerparameters	Value
algorithm	auto
n_neighbors	10
p	1

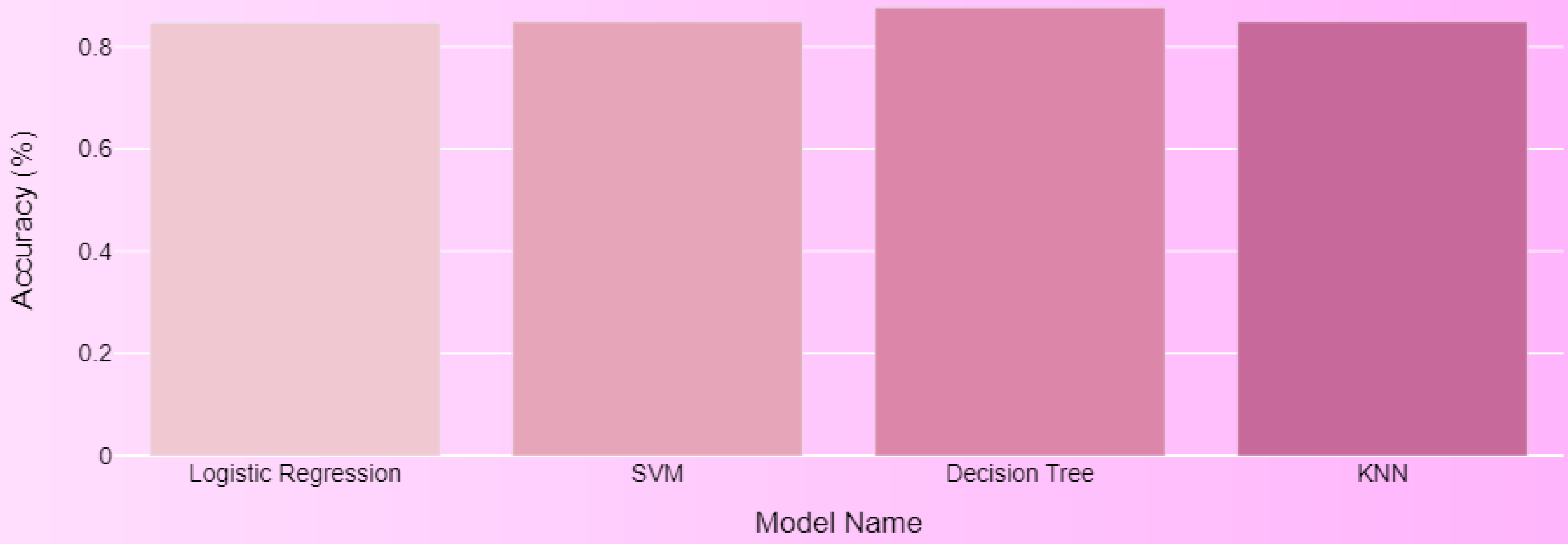
Best Score	0.848214

LAUNCH SUCCESS DISTRIBUTION

Model	Accuracy
Logistic Regression	0.846429
SVM	0.848214
Decision Tree	0.876786
KNN	0.848214

Best Model	Decision Tree
Score	0.876786

Model Accuracy Comparison

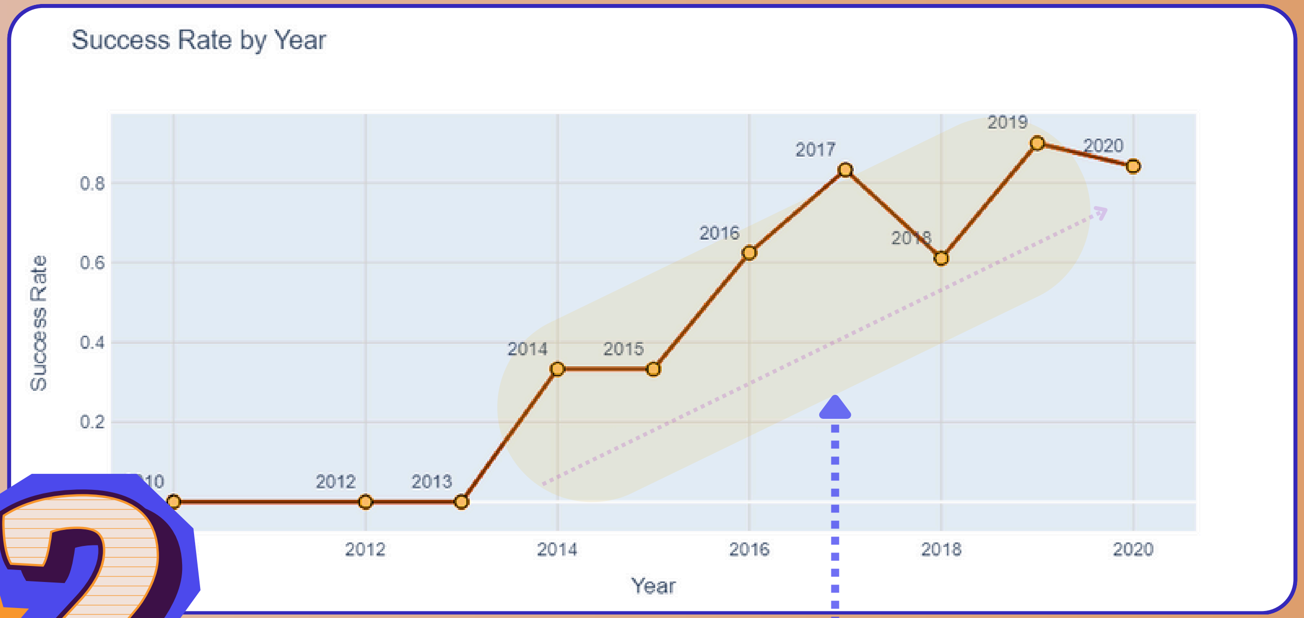


The best performing model is Decision Tree with an accuracy of 0.8767857142857143

CONCLUSION

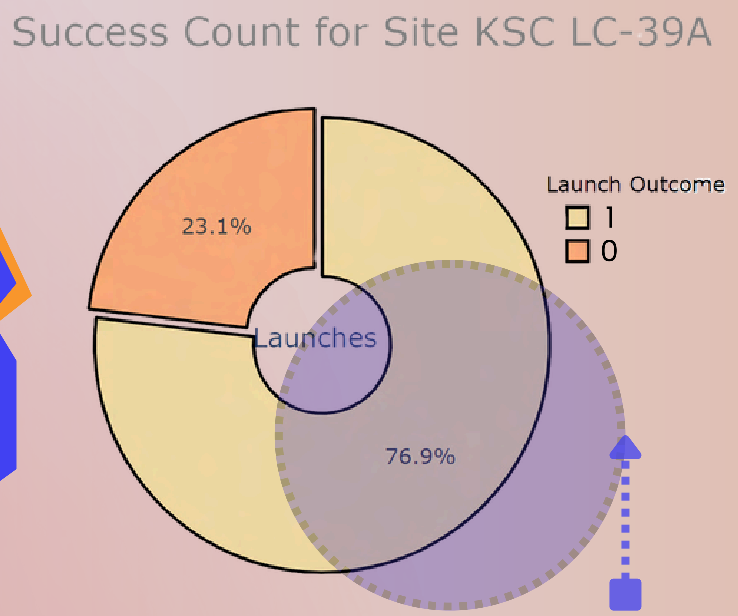
Comprehensive insights obtained through SQL queries, revealing critical patterns in launch site performance, payload capacities, and mission success rates.

Orbits ES-L1, GEO, HEO, and SSO exhibit the highest success rates.



From 2013 to 2020, SpaceX's launch success rate has steadily and significantly increased.

IBM SPACEX



3

KSC LC-39A stands out with the highest percentage of successful launches.

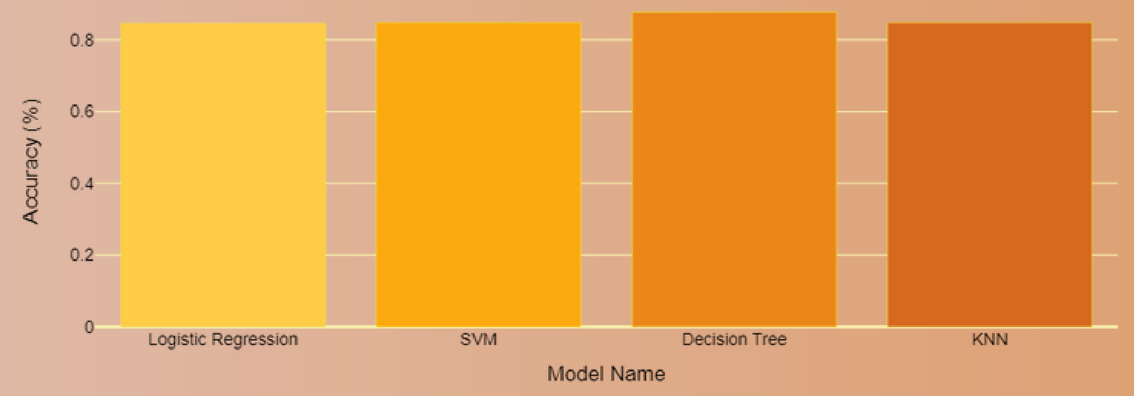
4

A clear correlation exists between lower payload mass and higher launch success rates.

5

The most successful booster version is Falcon 9 FT.

6



The Decision Tree model achieved the best performance.

IBM SPACEX

THANK YOU!