DATA SCIENCE CAPSTONE PROJECT

RAZAFINANDRASANA Mandifera
28/02/2025

Executive Summary

This project analyzes SpaceX's historical launch data to determine key factors affecting launch success rates.
We employed data science techniques such as:

- **Data Collection**
- **Data Wrangling**
- **Exploratory Data Analysis (EDA)**
- **Interactive Visual Analytics**
- **Predictive Modeling**

The findings will help in understanding launch patterns and optimizing future missions.

Introduction

SpaceX aims to revolutionize space travel by reducing costs and increasing launch success rates.
Our goal is to analyze past launches to identify factors that contribute to successful missions.

Key questions explored:

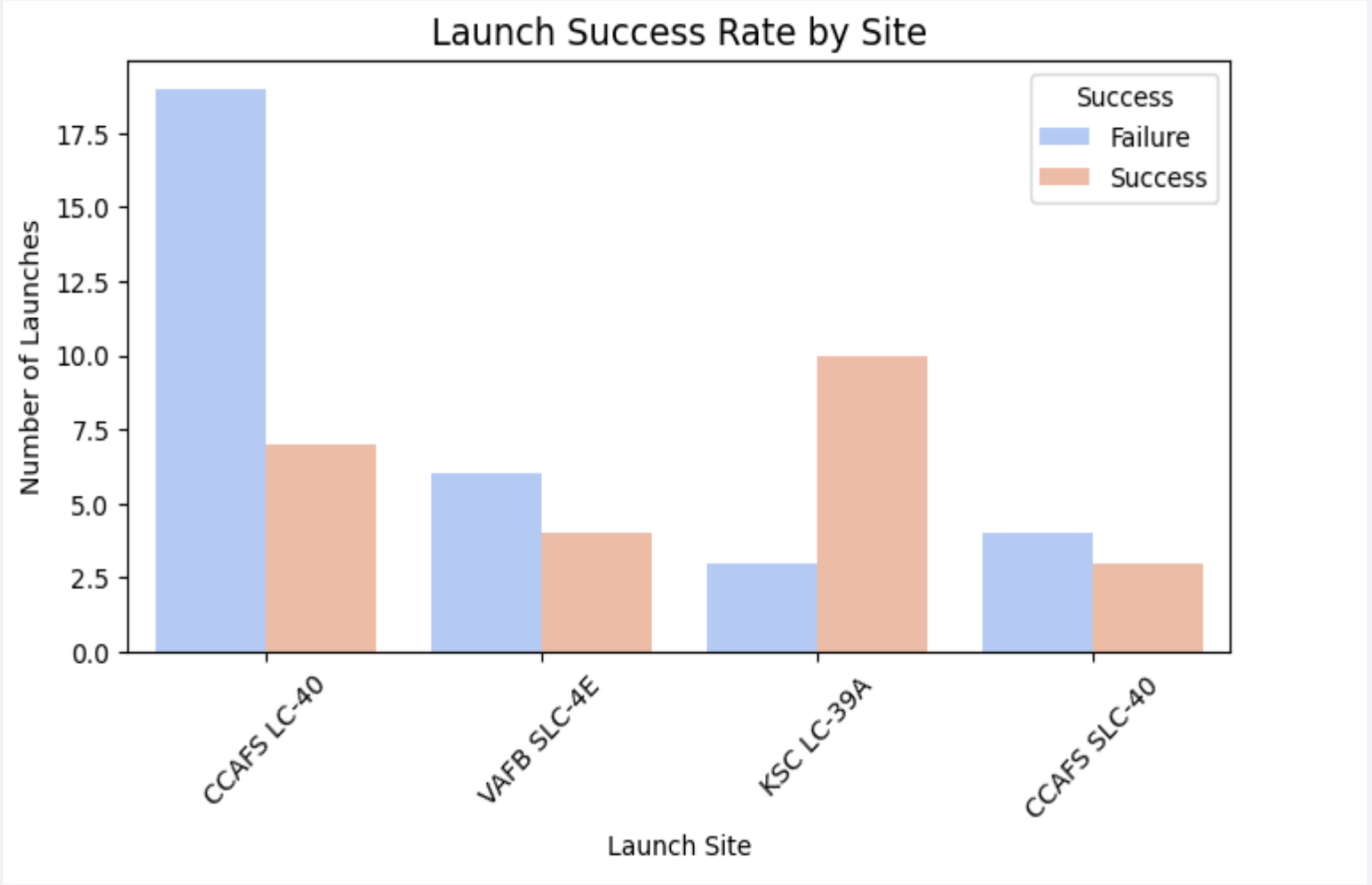
- What factors influence the success or failure of a launch?
- How does payload mass impact the success rate?
- Can we predict whether a future launch will be successful?

Data Collection & Wrangling

- **Datasets Used**
- We utilized two key datasets to conduct our analysis:
- **spacex_launch_dash.csv** → Contains detailed records of SpaceX launches, including launch sites, payload mass, booster versions, and mission success/failure.
- **spacex_launch_geo.csv** → Provides geographical coordinates of launch sites for spatial analysis and interactive mapping.
- **Data Processing Workflow**
- To ensure data quality and usability, we performed the following preprocessing steps:
  - **Handling Missing Values:** Filled or removed incomplete data to maintain consistency.
  - **Standardizing Data Formats:** Unified date formats, numerical values, and categorical labels.
  - **Feature Engineering & Encoding:** Converted categorical variables (e.g., launch sites, booster versions) into numerical values for machine learning analysis.
  - **Merging Datasets:** Combined launch records with geographical coordinates for enhanced visual and spatial analysis.

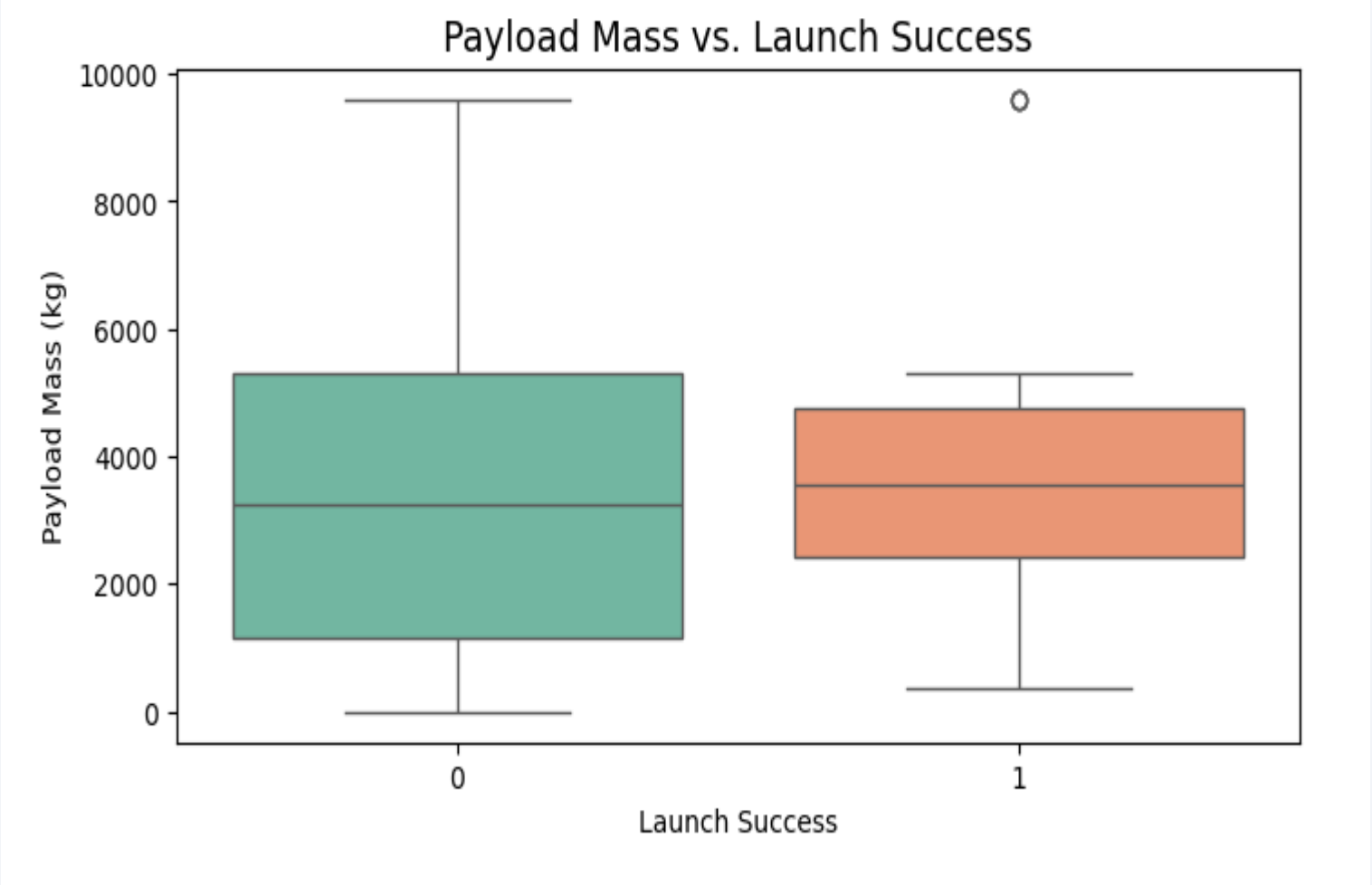
Exploratory Data Analysis (EDA) - Visualizations

- **Launch Success Rate by Site**
- Shows the number of successful and failed launches at each site.



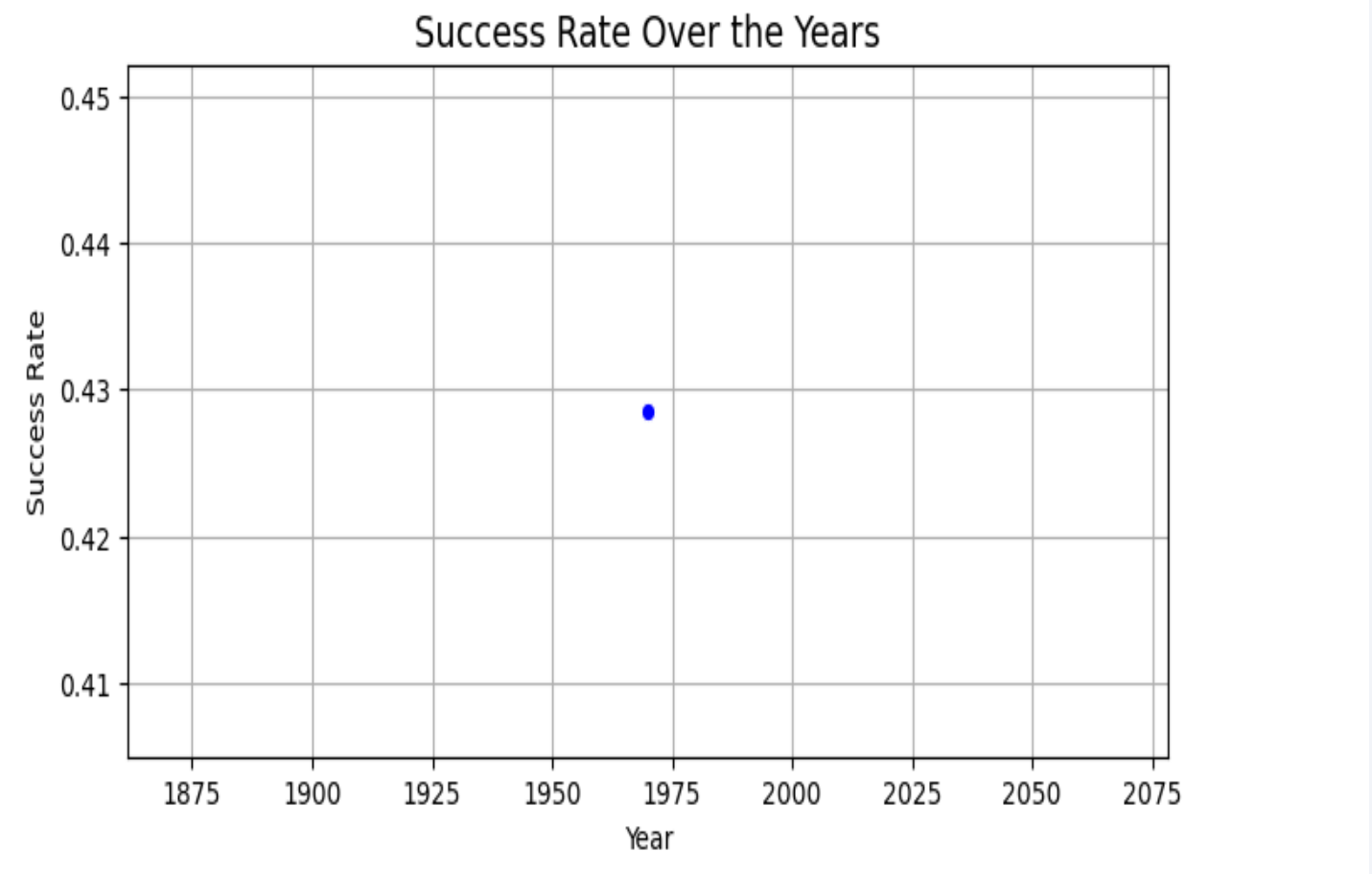
Payload Mass vs. Launch Success

- Payload mass is a critical factor influencing the success of a rocket launch. This analysis examines the relationship between payload weight and mission outcomes, helping to determine whether heavier payloads affect launch reliability



Success Rate Over the Years

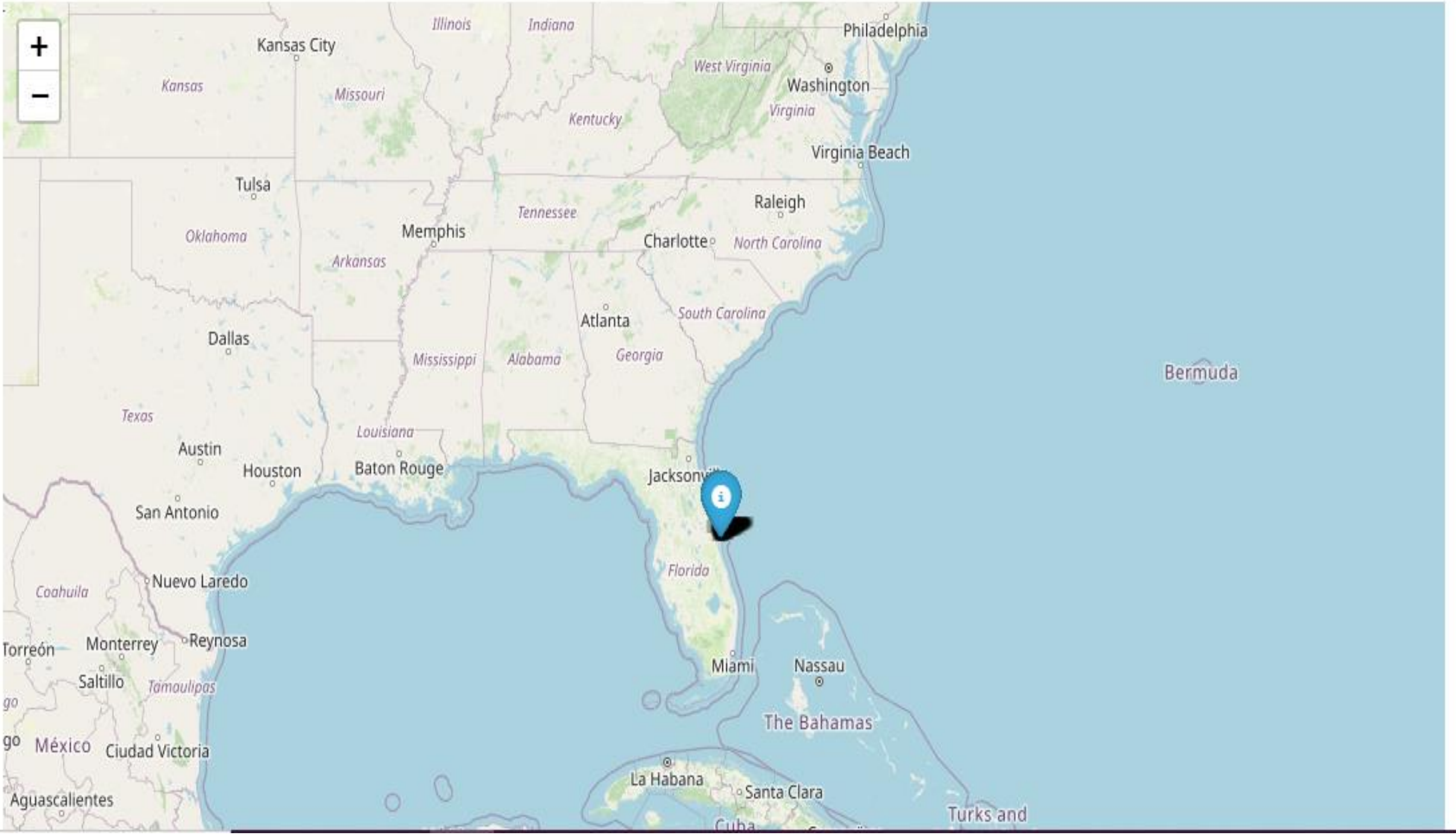
- Analyzing the success rate of SpaceX launches over time provides valuable insigl into the company's technological advancements and operational improvements.
- By visualizing trends in mission success, we can assess the progress made and identify key factors contributing to increased reliability



Interactive Data Visualization - Folium Map

- Created **dynamic dashboards** using **Plotly Dash** and **Folium**.
- Allowed real-time filtering of launch success by **site**, **year**, and **payload range**.
- **Geospatial Analysis**: Visualizes the geographical locations of SpaceX launch sites.
- **Launch Site Markers**: Displays each launch site with interactive pop-ups containing key details.
- **Success Rate Insights**: Helps analyze which locations have the highest launch success rates.
- **Zoom & Navigation**: Users can explore different launch sites interactively.
- **Real-World Application**: Useful for mission planning and site selection based on past performance.

[42]:


Kansas City
Illinois
Indiana
Philadelphia
West Virginia
Washington
Kansas
Missouri
Kentucky
Virginia
Virginia Beach
Tulsa
Tennessee
Raleigh
Oklahoma
Memphis
Charlotte
North Carolina
Arkansas
Atlanta
South Carolina
Dallas
Mississippi
Alabama
Georgia
Bermuda
Texas
Louisiana
Austin
Houston
Baton Rouge
Jacksonville
San Antonio
Florida
Nuevo Laredo
Coahuila
Torreón
Monterrey
Reynosa
Miami
Nassau
Saltillo
Tamaulipas
The Bahamas
México
Ciudad Victoria
La Habana
Santa Clara
Aguascalientes
Turks and

SQL-Based EDA - Query Results

To gain deeper insights into SpaceX launch data, we performed exploratory data analysis (EDA) using SQL queries. These queries allow us to extract key information, such as the distribution of launches across different sites, payload characteristics, and overall mission success rates

	Launch Site	launch_count
0	CCAFS LC-40	26
1	KSC LC-39A	13
2	VAFB SLC-4E	10
3	CCAFS SLC-40	7

SQL - Success Rate per Launch Site

- To understand the performance of different launch sites, we used SQL queries to calculate the success rate of launches at each location. This analysis helps identify which sites have the highest probability of a successful mission, providing valuable insights for future launch planning.

	Launch Site	success_rate
0	KSC LC-39A	0.769231
1	CCAFS SLC-40	0.428571
2	VAFB SLC-4E	0.400000
3	CCAFS LC-40	0.269231

Machine Learning - Predictive Model

- To predict the success of a SpaceX launch, we implemented a machine learning model using a Random Forest classifier. The model was trained on key features such as payload mass, launch site, and booster version. By analyzing historical launch data, we aim to identify the most influential factors affecting mission success and improve future launch predictions.
- **Model Used:** Random Forest Classifier

```
Index(['Unnamed: 0', 'Flight Number', 'Launch Site', 'class',
       'Payload Mass (kg)', 'Booster Version', 'Booster Version Category'],
      dtype='object')
✅ Model Accuracy: 0.67
```

Dashboard with Plotly Dash

This interactive dashboard provides a visual analysis of SpaceX launch success rates. It includes a pie chart illustrating the proportion of successful launches per site and a scatter plot showing the relationship between payload mass and launch outcomes. These insights help in understanding key factors influencing mission success.

SpaceX Launch Dashboard

Launch Success Rate by Site

Payload Mass vs. Launch Success

Conclusion

Key Takeaways:

- Launch **site** and **payload mass** are the most critical success factors.
- Machine learning can predict launch success with **high accuracy**.
- Data-driven decisions can improve mission planning.

Thank you!