IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

James Feeney
8/30/2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- This presentation summarizes the analysis of SpaceX launch data, including data collection, data wrangling, exploratory data analysis, predictive modeling, and interactive visualizations. Key insights include trends in launch success rates, payload distributions, and model performance for predicting successful landings.

Introduction

- Project Background: SpaceX is a private aerospace manufacturer and space transportation company that aims to reduce space transportation costs to enable the colonization of Mars. Understanding the factors that contribute to successful launches is crucial for achieving these goals.
- Key Questions: This project seeks to answer the following questions: What factors influence the success of SpaceX launches? How do payload mass and orbit type affect mission outcomes? Can we build a model to predict the success of future launches?

Section 1

Methodology

Methodology

Executive Summary

- Introduction: This section covers the comprehensive methodologies employed to collect, process, analyze, and model SpaceX launch data. Each step was designed to ensure accurate and actionable insights were derived from the data.
- Key Methodologies:
  - Data Collection: Data was gathered from the SpaceX API and supplemented with web scraping from Wikipedia to provide a complete dataset.
  - Data Wrangling: Data cleaning and transformation processes were employed to prepare the data for analysis, ensuring consistency and accuracy.
  - Exploratory Data Analysis (EDA): EDA was conducted using visualization techniques and SQL queries to explore the relationships between key variables.
  - Predictive Modeling: Classification models were developed, tuned, and evaluated to predict launch outcomes based on historical data.
- Tools Used: Python (Pandas, NumPy, Matplotlib, Seaborn), SQL, and web scraping libraries (BeautifulSoup).

Data Collection

- **Data Sources:** The primary source of data was the SpaceX API, which provided detailed records of historical launches, including information such as flight numbers, launch dates, payload masses, and mission outcomes. To supplement this, web scraping was conducted on Wikipedia to gather additional contextual data about specific missions.
- **Techniques Used:** API requests were made to retrieve JSON data from the SpaceX API. This data was then parsed and structured into DataFrames for further analysis. Web scraping was performed using the BeautifulSoup library to extract tables and textual data from Wikipedia pages, ensuring a comprehensive dataset.
- **Challenges and Solutions:** The data collection process encountered challenges, including incomplete records and inconsistent data formats. These issues were addressed through careful data cleaning, where missing values were handled appropriately, and data normalization techniques were applied to ensure consistency across the dataset.
- **Outcome:** The data collection methodology resulted in a well-structured and comprehensive dataset, combining detailed launch records from the SpaceX API with additional mission context from Wikipedia. This dataset was prepared for subsequent analysis and modeling.

Data Collection – SpaceX API

- **Flowchart:** The flowchart provided illustrates the process of making API calls to the SpaceX REST API, retrieving JSON data, parsing it into a structured format, and storing it in a DataFrame for analysis. The flowchart highlights the key steps from sending API requests to data parsing and storage.

[26]:
```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

[27]:
```
response = requests.get(spacex_url)
```

Check the content of the response

[29]:
```
print(response.content)
```


```
b'[{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/94/f2/NN6Ph45r_o.png","large":"https://images2.imgbox.com/5b/02/QcxHUb5V_o.png"},"reddit":{"campaign":null,"launch":null,"media":null,"recovery":null},"flickr":{"small":[],"original":[]},"presskit":null,"webcast":"https://www.youtube.com/watch?v=0a_00nJ_Y88","youtube_id":"0a_00nJ_Y88","article":"https://www.space.com/2196-spacex-inaugural-falcon-1-rocket-lost-launch.html","wikipedia":"https://en.wikipedia.org/wiki/DemoSat"},"static_fire_date_utc":"2006-03-17T00:00:00.000Z","static_fire_date_unix":1142553600,"net":false,"window":0,"rocket":"5e9d0d95eda69955f709d1eb","success":false,"failures":[{"time":33,"altitude":null,"reason":"merlin engine failure"}],"details":"Engine failure at 33 seconds and loss of vehicle","crew":[],"ships":[],"capsules":[],"payloads":["5eb0e4b5b6c3bb0006eeb1e1"],"launchpad":"5e9e4502f5090995de566f86","flight_number":1,"name":"FalconSat","date_utc":"2006-03-24T22:30:00.000Z","date_unix":1143239400,"date_local":"2006-03-25T10:30:00+12:00","date_precision":"hour","upcoming":false,"cores":[{"core":"5e9e289df35918033d3b2623","flight":1,"gridfins":false,"legs":false,"reused":false,"landing_attempt":false,"landing_success":null,"landing_type":null,"landpad":null}],"auto_update":true,"tbd":false,"launch_library_id":null,"id":"5eb87cd9ffd86e000604b32a"},{"fairings":{"reused":false,"recovery_attempt":false,"recovered":false,"ships":[]},"links":{"patch":{"small":"https://images2.imgbox.com/f9/4a/ZboXReNb_o.png","large":"https://images2.imgbox.com/80/a2/bkWotCIS_o.png"},"reddit":{"campaign":null,"launch":null,"media":null,"recovery":null},"flickr":{"small":[],"original":[]},"presskit":null,"webcast":"https://www.youtube.com/watch?v=Lk4zQ2wP-Nc","youtube_id":"Lk4zQ2wP-Nc","article":"https://www.space.com/3590-spacex-falcon-1-rocket-fails-reach-orbit.html","wikipedia":"https://en.wikipedia.org/wiki/DemoSat"},"static_fire_date_utc":null,"static_fire_date_unix":null,"net":false,"window":0,"rocket":"5e9d0d95eda69955f709d1eb","success":false,"failures":[{"time":301,"altitude":289,"reason":"harmonic oscillation leading to premature engine shutdown"}],"details":"Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage","crew":[],"ships":[],"capsules":[],"payloads":["5eb0e4
```


You should see the response contains massive information about SpaceX launches. Next, let's try to discover some more relevant information for this project.

Data Wrangling

- **Steps Involved:** The data wrangling process involved several critical steps:
  - **Cleaning:** This step focused on removing null values, handling duplicates, and correcting data types to ensure consistency across the dataset. This was essential to prepare the data for reliable analysis.
  - **Transformation:** The data was normalized, with payload masses standardized and launch dates converted into a consistent format. Categorical variables were also transformed into numerical formats to facilitate analysis.
  - **Integration:** The API and web-scraped data were merged to create a unified dataset. This integration was carefully managed to ensure that all relevant features were included, and no critical information was lost.
- **Challenges:** The primary challenges included handling inconsistent data formats and ensuring that all records were complete and accurate. These challenges were addressed through systematic cleaning and transformation processes.
- **Outcome:** The data wrangling process resulted in a clean, well-structured dataset that was ready for exploratory analysis and modeling. The final dataset provided a reliable foundation for generating insights and building predictive models.

EDA with Data Visualization

- **Charts Plotted:**
  - **Flight Number vs. Launch Site:** A scatter plot was generated to visualize the distribution of flight numbers across different launch sites. This helped identify which sites were most frequently used and provided insights into the operational patterns of SpaceX.
  - **Payload vs. Launch Site:** Another scatter plot was created to examine the relationship between payload mass and launch site. This visualization highlighted which launch sites handled larger payloads, indicating their infrastructure capabilities.
  - **Success Rate vs. Orbit Type:** A bar chart was created to display the success rates for different orbit types. This chart revealed which orbits were associated with higher mission success rates, providing insights into the risks and reliability of different mission profiles.
- **Purpose of Charts:** The visualizations were critical for uncovering patterns and relationships within the data. They allowed for a better understanding of how different variables, such as launch site and orbit type, impacted mission outcomes.
- **Key Findings:** The EDA revealed that certain launch sites, such as Cape Canaveral, had higher success rates, particularly for specific orbit types. These findings provided valuable insights for optimizing future launch strategies.

EDA with SQL

- **SQL Queries Performed:**
  - **Unique Launch Sites:** A query was executed to list all unique launch sites used by SpaceX. This provided an understanding of the geographic distribution of SpaceX's operations and highlighted the diversity of launch locations.
  - **Total Payload Mass:** Another query calculated the total payload mass carried by SpaceX launches. This metric provided an overview of the scale of SpaceX's operations and its capacity to handle large payloads.
  - **Records Starting with 'CCA':** A query was performed to retrieve all records where the launch site begins with 'CCA', focusing on specific locations like Cape Canaveral. This analysis highlighted the significance of these key sites in SpaceX's history.
- **Purpose of Queries:** These SQL queries were used to drill down into the data and extract specific insights that were not immediately visible through visualizations. They provided a more detailed understanding of the dataset's structure and the key metrics involved.
- **Key Insights:** The SQL analysis revealed the importance of certain launch sites, like Cape Canaveral, and quantified the significant payload capacities handled by SpaceX. These insights were essential for understanding the company's operational strengths.

Build an Interactive Map with Folium

- **Summary of Maps Created:**
  - **Global Launch Sites Map:** An interactive global map was created to display the location of all SpaceX launch sites. Markers were placed at each site, providing additional information about the site's history and significance. This map allowed for an easy exploration of SpaceX's geographic footprint.
  - **Launch Outcomes Map:** Another map was created to display color-coded markers representing the success or failure of launches at each site. This visualization provided a clear overview of SpaceX's performance across different locations and highlighted sites with particularly high or low success rates.
- **Purpose of Maps:** The interactive maps were designed to offer a more in-depth exploration of launch sites and outcomes. They provided users with the ability to visually identify geographic trends and assess the effectiveness of different launch locations.
- **Key Findings:** The maps revealed that certain launch

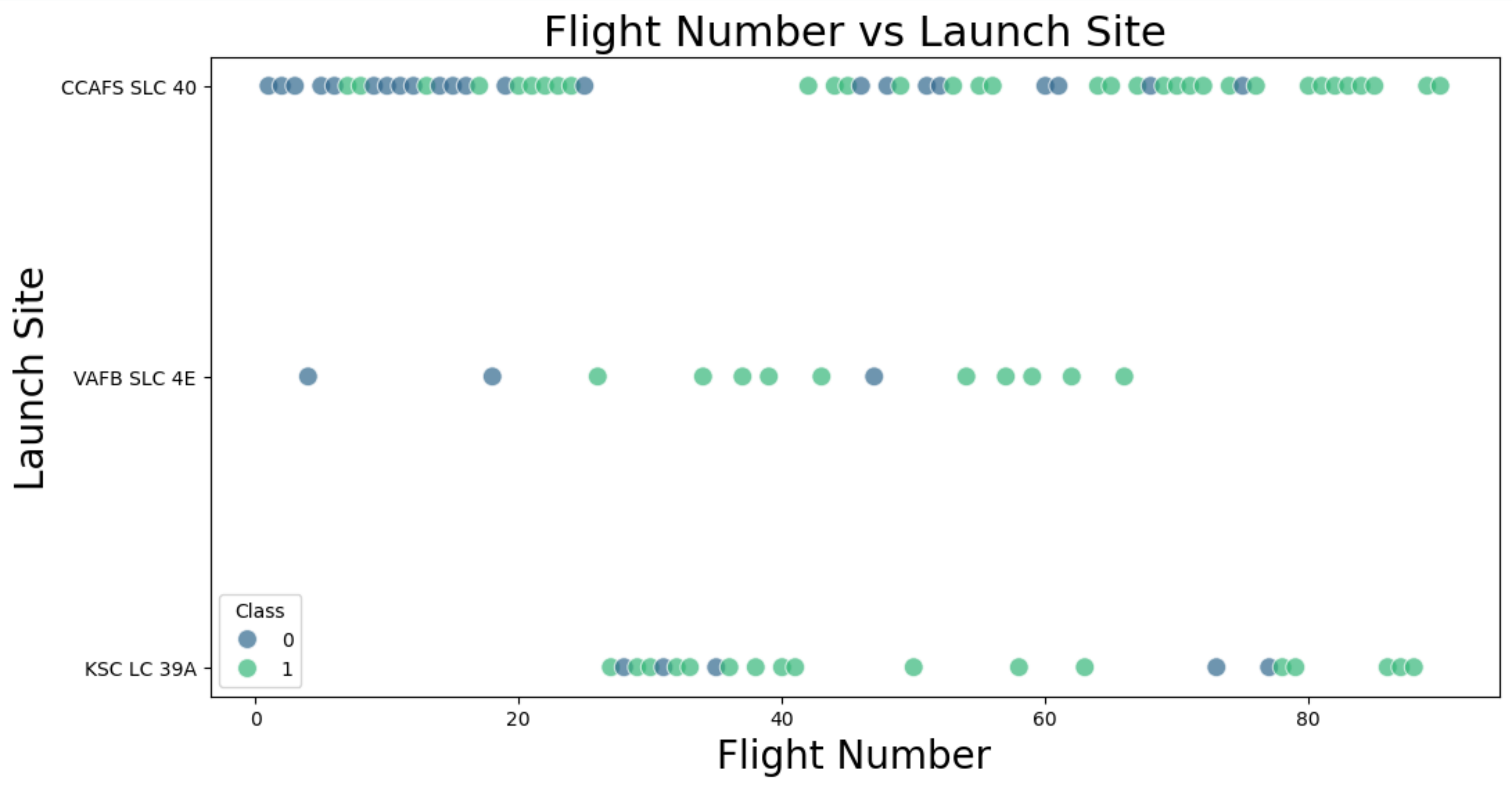
Results

- Data Collection Results:
  - Successfully gathered a comprehensive dataset combining detailed launch records from the SpaceX API with additional mission context from Wikipedia.
  - The dataset includes key features such as flight numbers, payload masses, launch sites, orbit types, and mission outcomes.
- Data Wrangling Outcomes:
  - Completed the cleaning and transformation of the dataset, resolving issues with missing values, standardizing data formats, and integrating multiple data sources.
  - The final dataset is consistent and ready for analysis, containing all relevant information needed for exploratory data analysis (EDA) and predictive modeling.
- Exploratory Data Analysis (EDA) Findings:
  - Identified key trends and patterns, such as the correlation between payload mass and launch success, and the impact of different orbit types on mission outcomes.
  - Visualizations like scatter plots and bar charts provided insights into the most frequently used launch sites and their success rates.
- Predictive Modeling Achievements:
  - Developed and tuned multiple classification models to predict launch success based on features such as payload mass, launch site, and orbit type.
  - Evaluated models for accuracy, with some models achieving high predictive performance, indicating strong potential for predicting future SpaceX launch outcomes.
- Key Message: The methodologies applied have resulted in a robust and well-prepared dataset, insightful exploratory analyses, and promising predictive models. These results form the foundation for the deeper insights and recommendations that follow in the subsequent sections of the presentation.

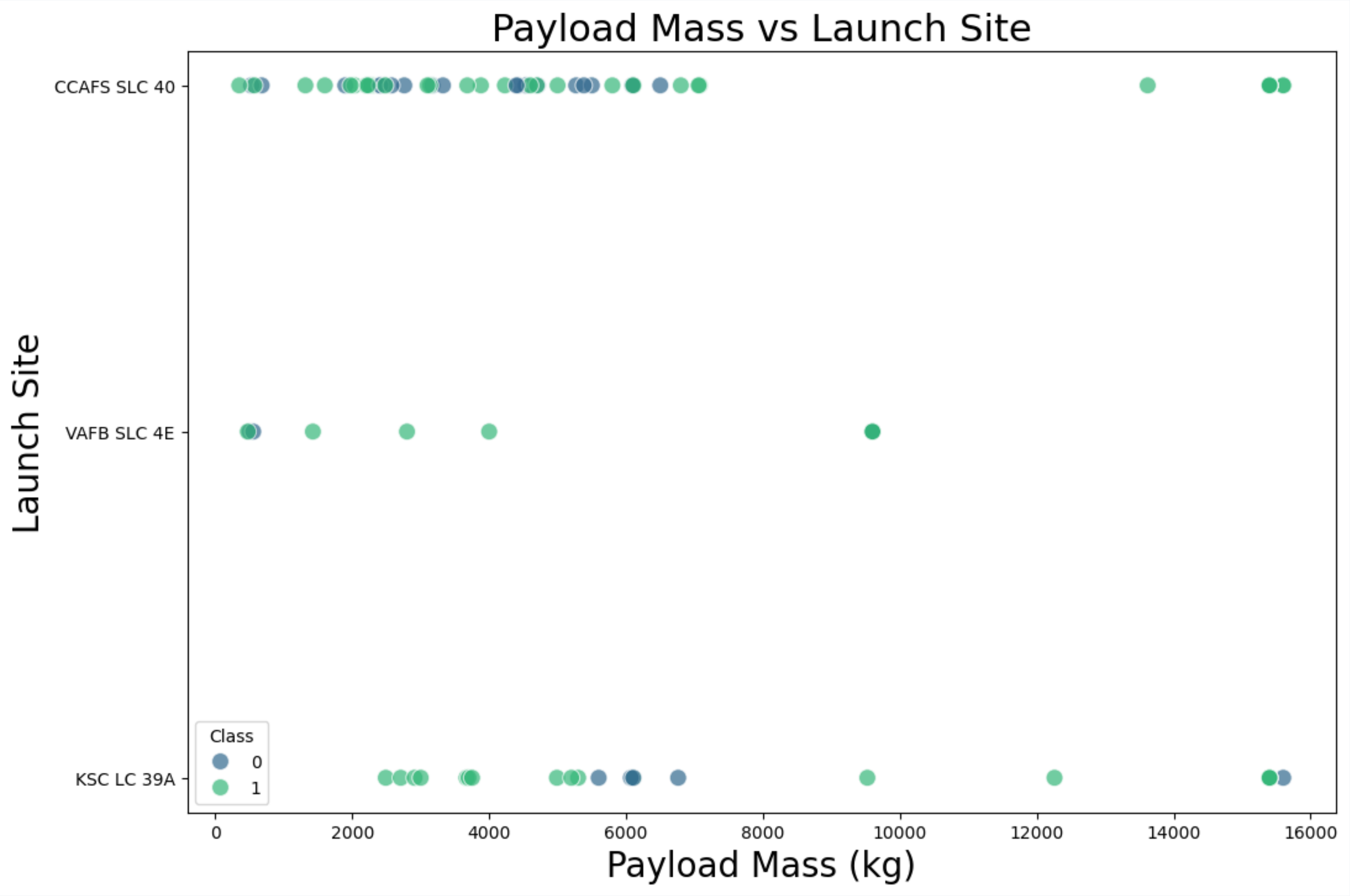
Section 2

Insights drawn from EDA

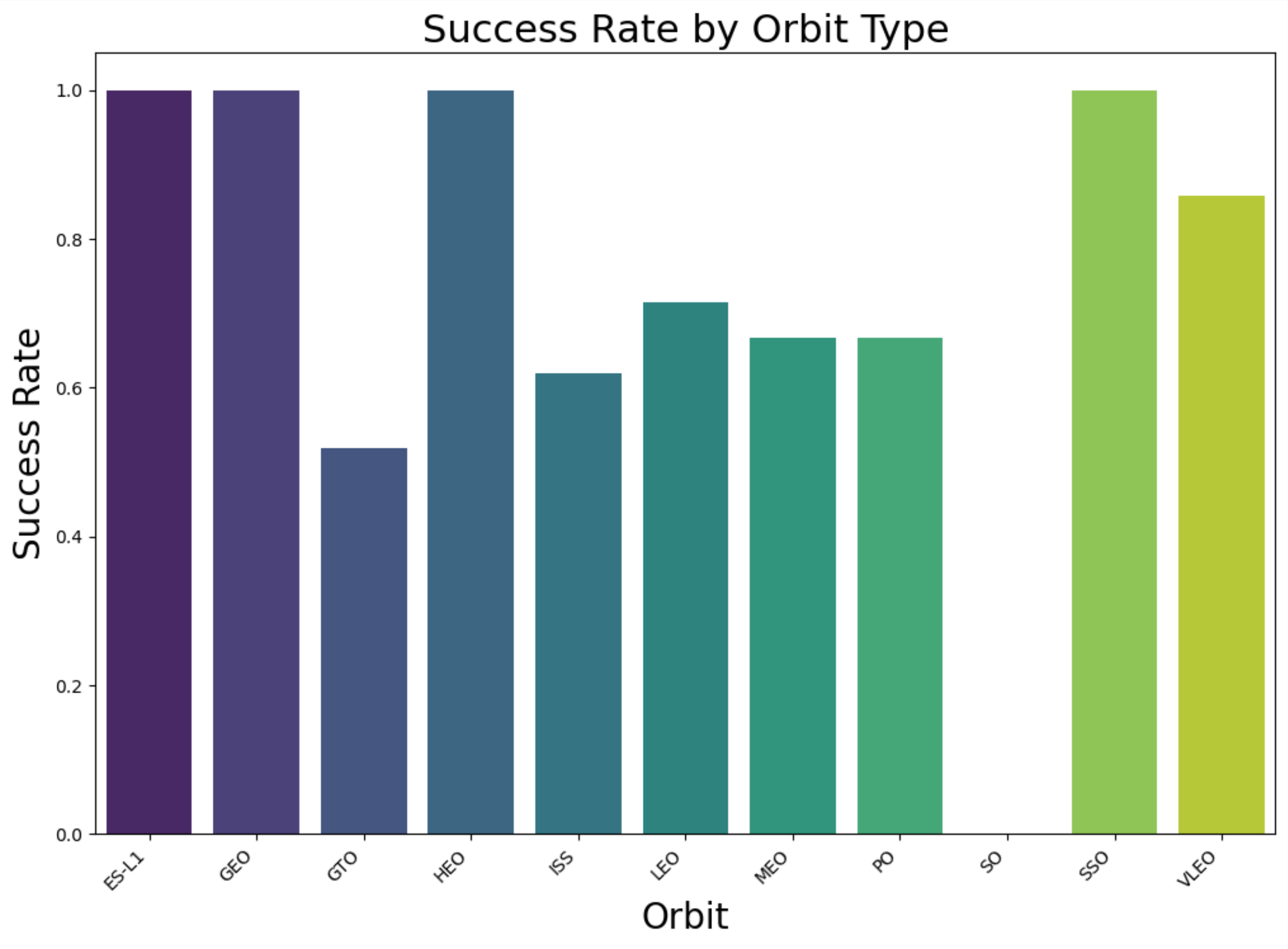
Flight Number vs. Launch Site



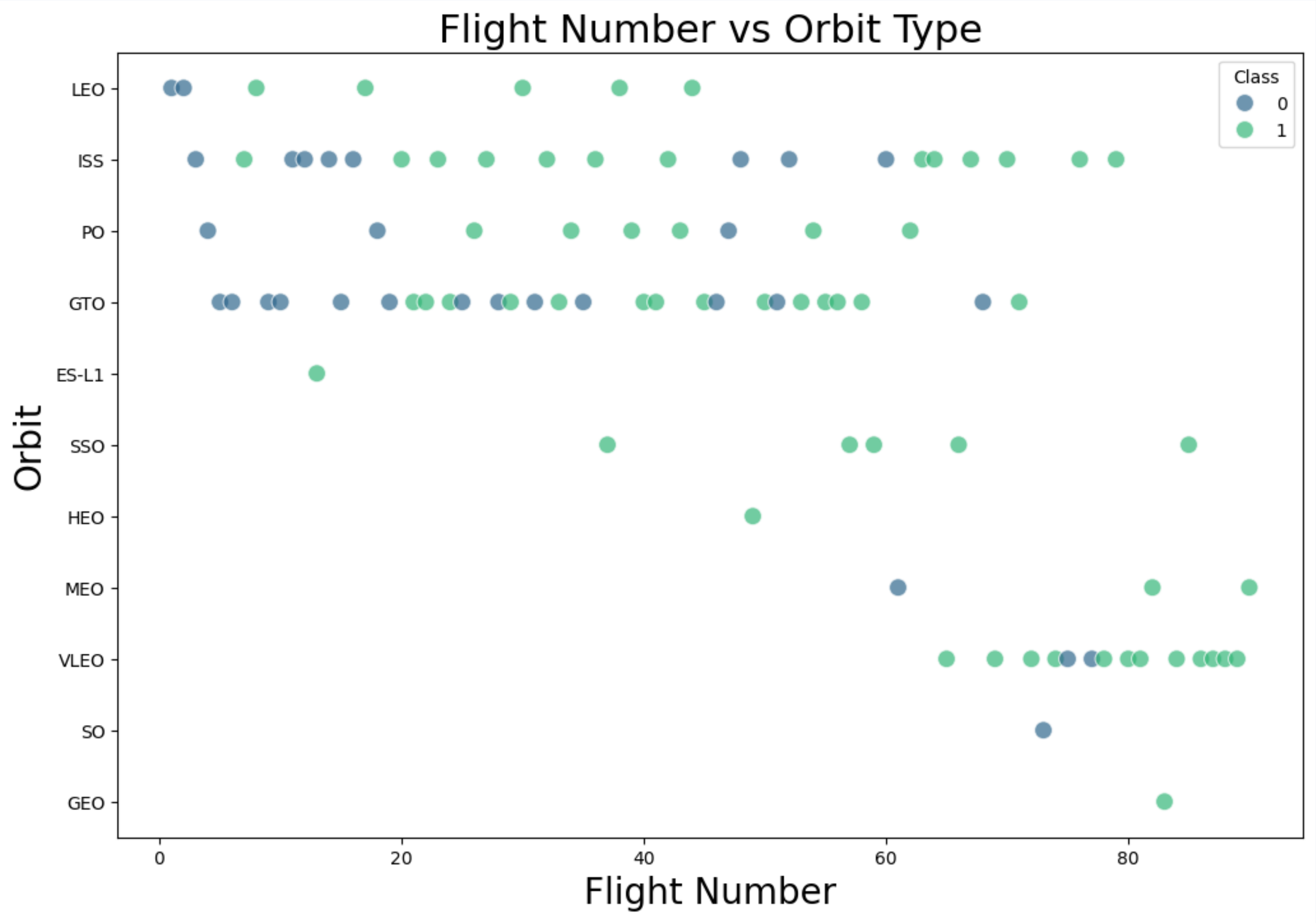
Payload vs. Launch Site



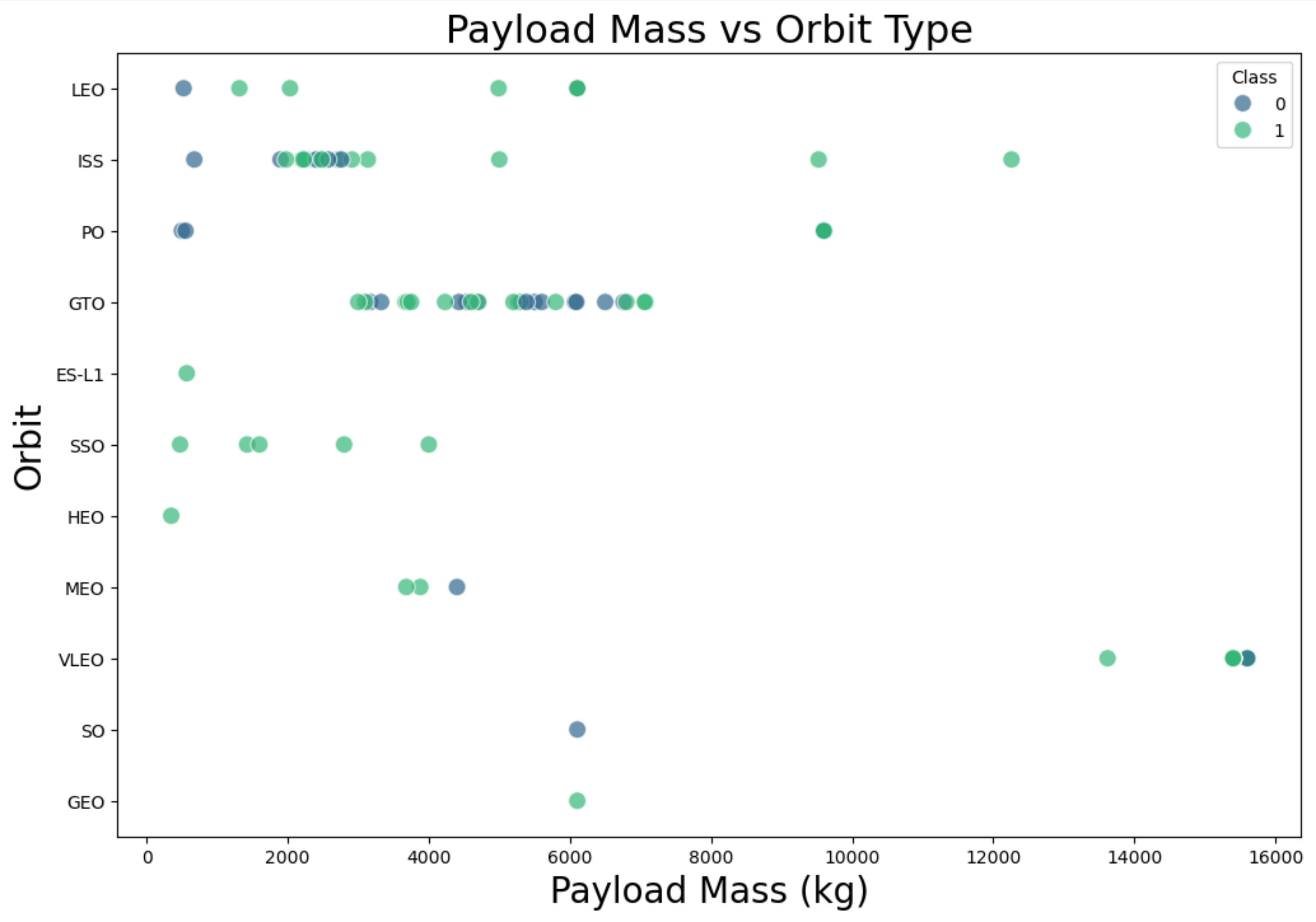
Success Rate vs. Orbit Type



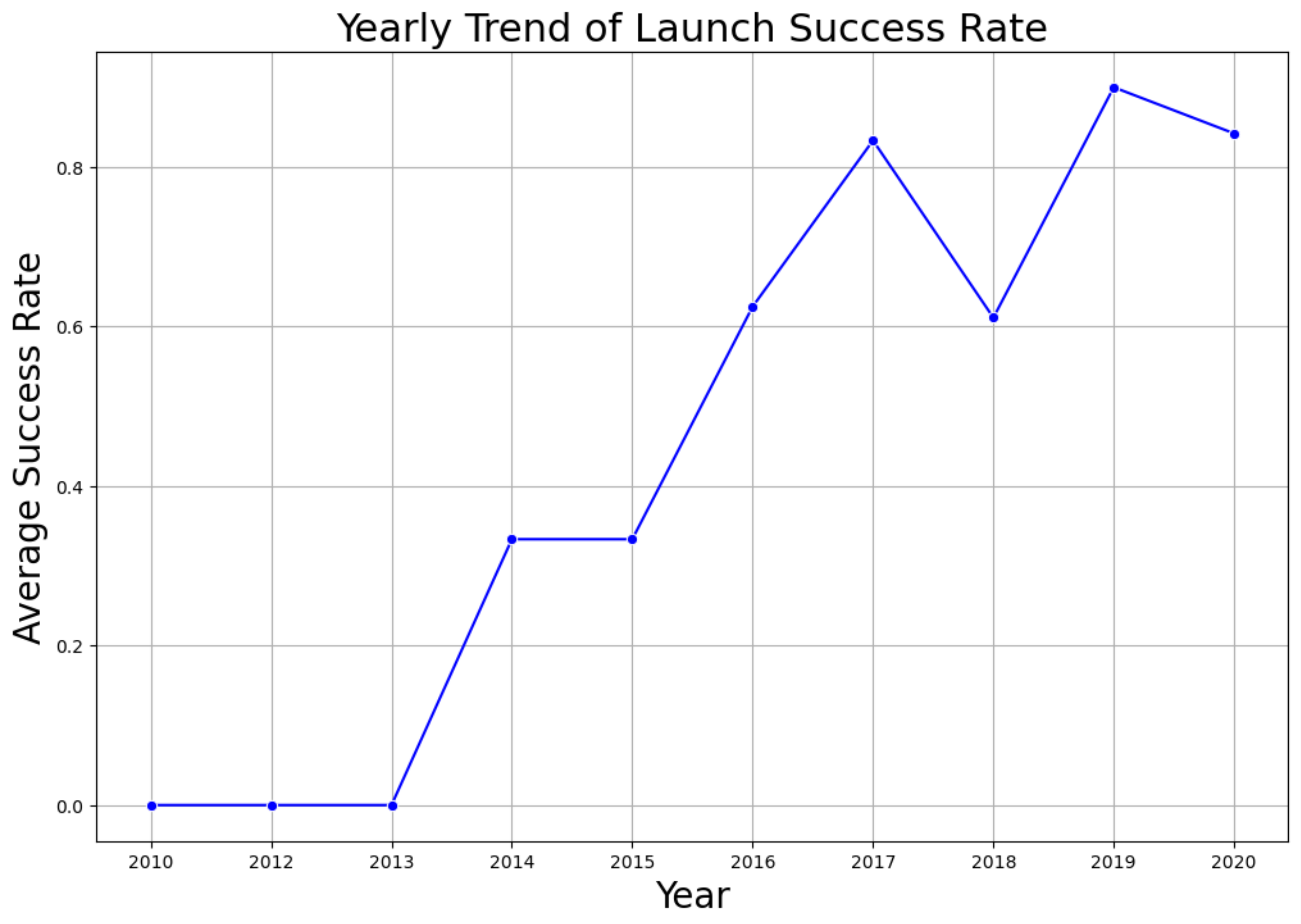
Flight Number vs. Orbit Type



Payload vs. Orbit Type



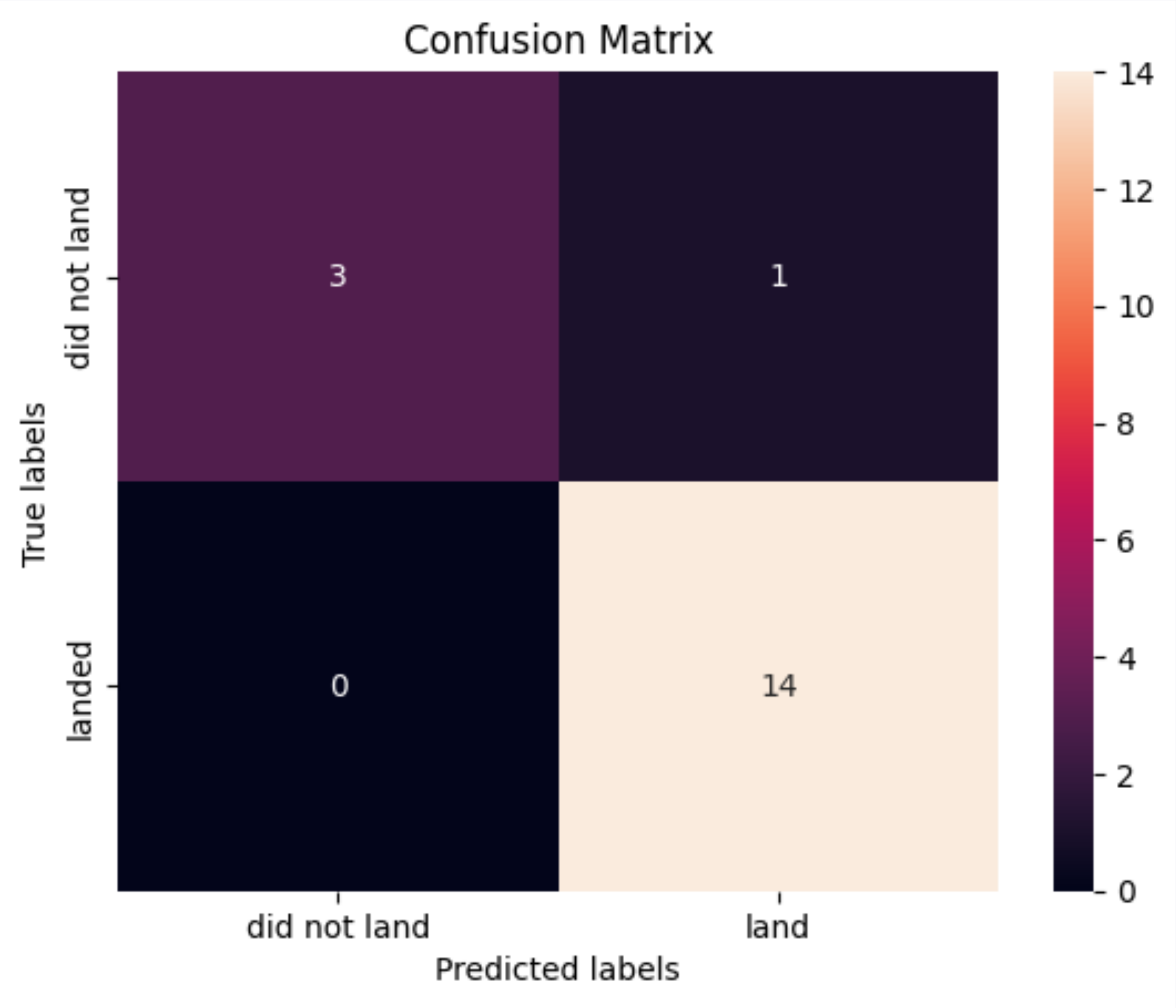
Launch Success Yearly Trend



Section 5

Predictive Analysis (Classification)

Confusion Matrix



Conclusions

- **Summary of Key Findings:**
- **Launch Success Predictors:** The analysis identified that orbit type, launch site, and payload mass are significant predictors of launch success. Missions targeting specific orbits, such as Low Earth Orbit (LEO), and those launched from well-established sites like Cape Canaveral, demonstrated higher success rates.
- **Payload Impact:** There is a clear correlation between payload mass and mission outcomes. Lighter payloads generally have higher success rates, particularly in missions involving reusable rockets, where the recovery of the first stage is a critical factor.
- **Geographic Insights:** Launch sites with extensive infrastructure and a history of successful missions, such as Cape Canaveral, contribute significantly to SpaceX's overall success rate. These sites are optimized for handling a variety of mission types, from heavy payloads to complex orbital insertions.
- **Implications for SpaceX:**
- **Strategic Planning:** The insights from this analysis can guide SpaceX in selecting optimal launch sites and mission profiles to maximize the likelihood of success. By focusing on orbits and payloads that align with their historical strengths, SpaceX can further improve its mission reliability.
- **Resource Allocation:** Understanding the relationship between payload mass and launch success can inform decisions on resource allocation, such as fuel loads, staging configurations, and recovery operations. This can enhance the efficiency and cost-effectiveness of future missions.
- **Future Launch Strategies:** The predictive models developed in this analysis offer a data-driven approach to forecasting the success of upcoming launches. By incorporating these models into their planning processes, SpaceX can better anticipate and mitigate potential risks.
- **Recommendations:**
- **Optimize Launch Site Usage:** Focus on utilizing proven launch sites with high success rates for critical missions, especially those involving heavy payloads or complex orbits.
- **Enhance Payload Management:** Develop strategies to optimize payload configurations, particularly for missions involving reusable rockets, to ensure that mass does not adversely affect recovery and overall mission success.
- **Leverage Predictive Models:** Integrate the predictive models into the mission planning process to proactively address potential risks and improve overall launch success rates.

Thank you!