IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alexej Tjurenkov
13.03.2025

Outline

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

Executive Summary

- Summary of methodologies

Data was collected through SpaceX API and web scraping. Data wrangling was performed to clean and prepare data for analysis. EDA was conducted using visualizations and SQL queries. Interactive analytics were developed using Folium and Plotly Dash. Predictive analysis was done through classification models to predict launch success.

- Summary of all results

The analysis identified key factors affecting launch success, including payload mass, launch site, and orbit type. The best-performing classification model was selected after tuning and evaluation. The interactive dashboard provided insights into trends and success rates across various parameters.

Introduction

- Project background and context

The SpaceX dataset contains historical information on rocket launches. Predicting launch success is critical for improving space mission planning and optimizing resources.

- Problems you want to find answers

Can we predict the success or failure of a SpaceX launch based on historical data? Which features most impact the success rate of a launch?

Section 1

Methodology

Methodology

Executive Summary

- Data Collection Methodology:

Data was collected via API calls to the SpaceX REST API and web scraping for supplementary data. The collected data included launch details such as launch site, payload, success, and orbit type.

- Data Wrangling:

Raw data was cleaned by handling missing values, converting categorical variables into numerical representations, and ensuring consistent formatting for analysis.

- Exploratory Data Analysis (EDA):

Various charts were plotted (scatter plots, bar charts, line charts) to understand relationships between different features and launch success. SQL queries helped uncover trends like success rates per orbit type.

- Interactive Visual Analytics:

Folium maps were used to visualize launch site locations and outcomes. Plotly Dash was used to build interactive dashboards for exploring the data and trends.

- Predictive Analysis Using Classification Models:

- Classification models (Logistic Regression, SVM, etc.) were trained to predict launch success. Models were evaluated based on accuracy and tuned using hyperparameter optimization techniques.

Data Collection

- Describe how data sets were collected.

Data was gathered through SpaceX API and web scraping. The SpaceX API was queried for launch-related data, while additional information was scraped from websites.

Data Collection – SpaceX API

- The SpaceX API provided data on rocket launches, including details like launch site, payload, orbit type, and success/failure. API calls were made for each required endpoint to fetch this data

- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

Start with API request → Fetch data → Store in a structured format (CSV/JSON).

Data Collection - Scraping

- **Web Scraping:** Web scraping was used to gather supplementary data not available through the API. Data was extracted from websites like SpaceX's official pages or other launch databases.
- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

Web scraping tool → Extract data → Clean and store data.

Data Wrangling

- **Data Processing:**

The raw data was cleaned by filling missing values, removing duplicates, and standardizing formats. Feature engineering was performed to create new columns that aid in predictive modeling.

- **Flowchart:** Raw data → Handle missing values → Feature engineering → Clean data.

- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

EDA with Data Visualization

- **Charts Plotted:**

Various visualizations were created:Scatter plot: Flight number vs. Launch siteBar chart: Success rate by orbit typeLine chart: Yearly trend of launch success These charts helped understand patterns and relationships in the data.

- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

EDA with SQL

SQL Queries:

- Unique launch site names.
- Success rate grouped by orbit type.
- Total number of successful and failed launches.
- Failed landing outcomes for 2015. Each query was executed to extract valuable insights into the dataset.

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- **Map Objects Created**:
  Markers were added for each launch site, color-coded by launch success or failure. Circles were used to represent payload mass, and lines were drawn to show the proximity of launch sites to roads or coastline. These elements were added to provide clear geographic and success-related insights

- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

Build a Dashboard with Plotly Dash

Dashboard Features:
The dashboard included interactive plots like: Success rate pie chart per launch site.

- Payload vs. launch outcome scatter plot. Interactions like filtering and range sliders were added to allow users to explore data dynamically.
- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

Predictive Analysis (Classification)

- **Model Development**:
  Multiple classification models were built, including Logistic Regression and Support Vector Machines (SVM). Models were evaluated using accuracy scores and confusion matrices. Hyperparameters were tuned to improve model performance.
- **Flowchart**: Data split → Model training → Hyperparameter tuning → Evaluation.

- https://github.com/AlexejGMC/Python-Poject-for-Data-Science.git

Results

- **Exploratory Data Analysis Results:**

EDA revealed that payload mass and orbit type significantly influence launch success. A higher payload and specific orbit types were associated with greater success rates.

- **Interactive Analytics Demo:**

Screenshots of the Folium maps and Plotly dashboards were included to demonstrate how users can interact with the data.

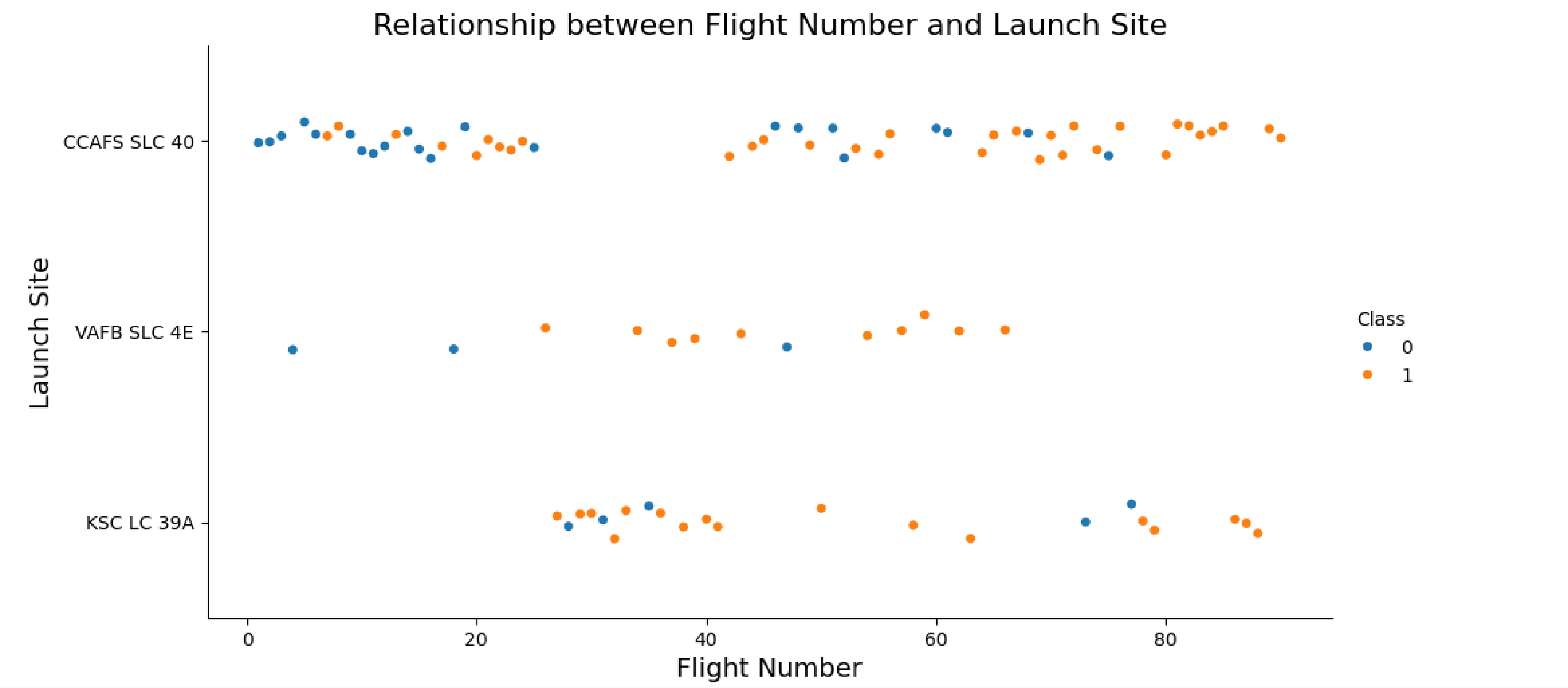
- **Predictive Analysis Results:**

The best classification model (e.g., Logistic Regression) showed high accuracy in predicting launch success. The confusion matrix helped identify the model's strengths and weaknesses.

Section 2

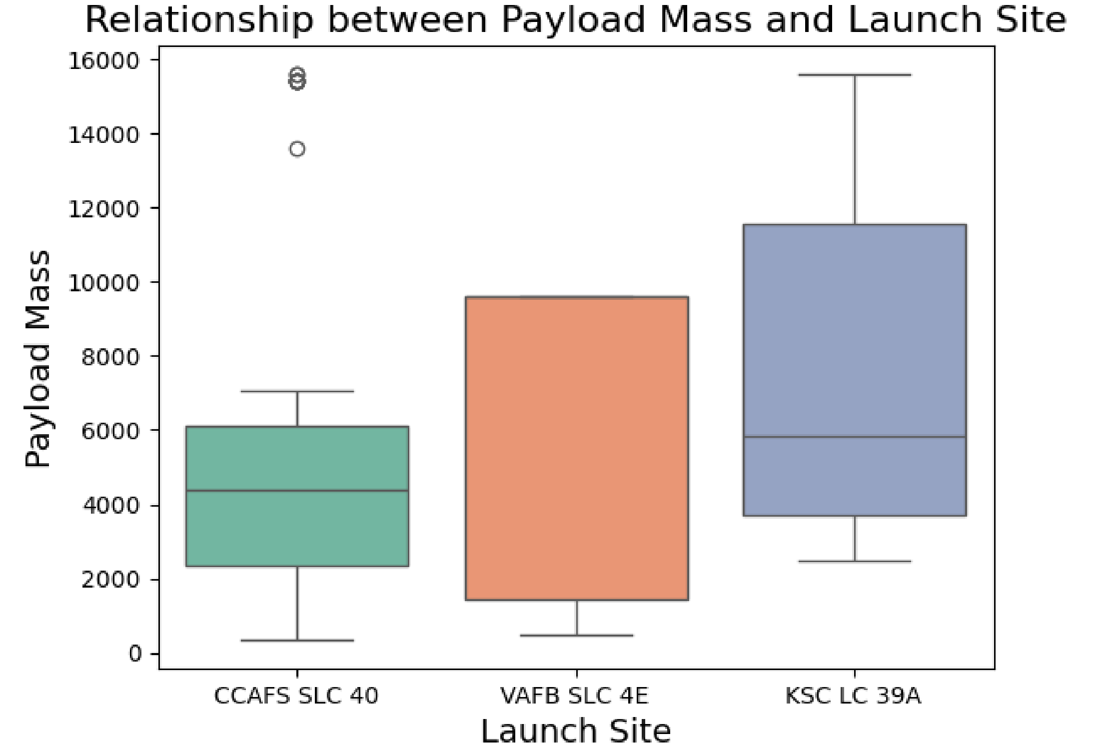
Insights drawn from EDA

Flight Number vs. Launch Site



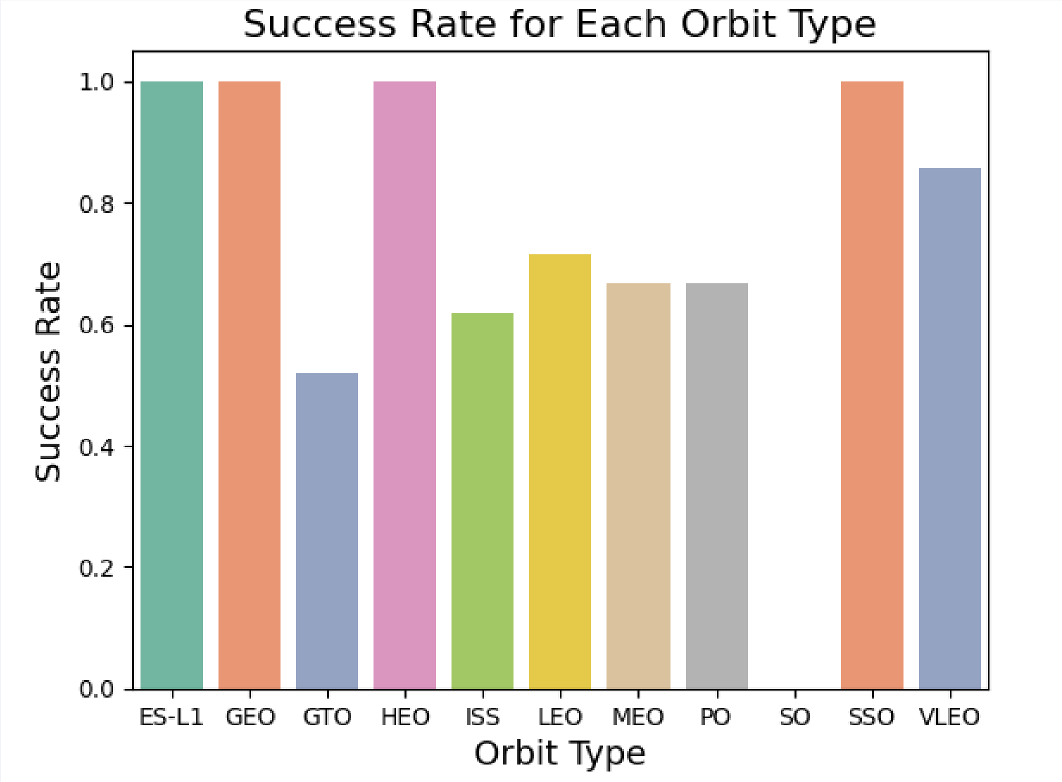
This plot showed the relationship between payload mass and launch site. It helped identify any significant correlation between payload size and site-specific launch success.

Payload vs. Launch Site



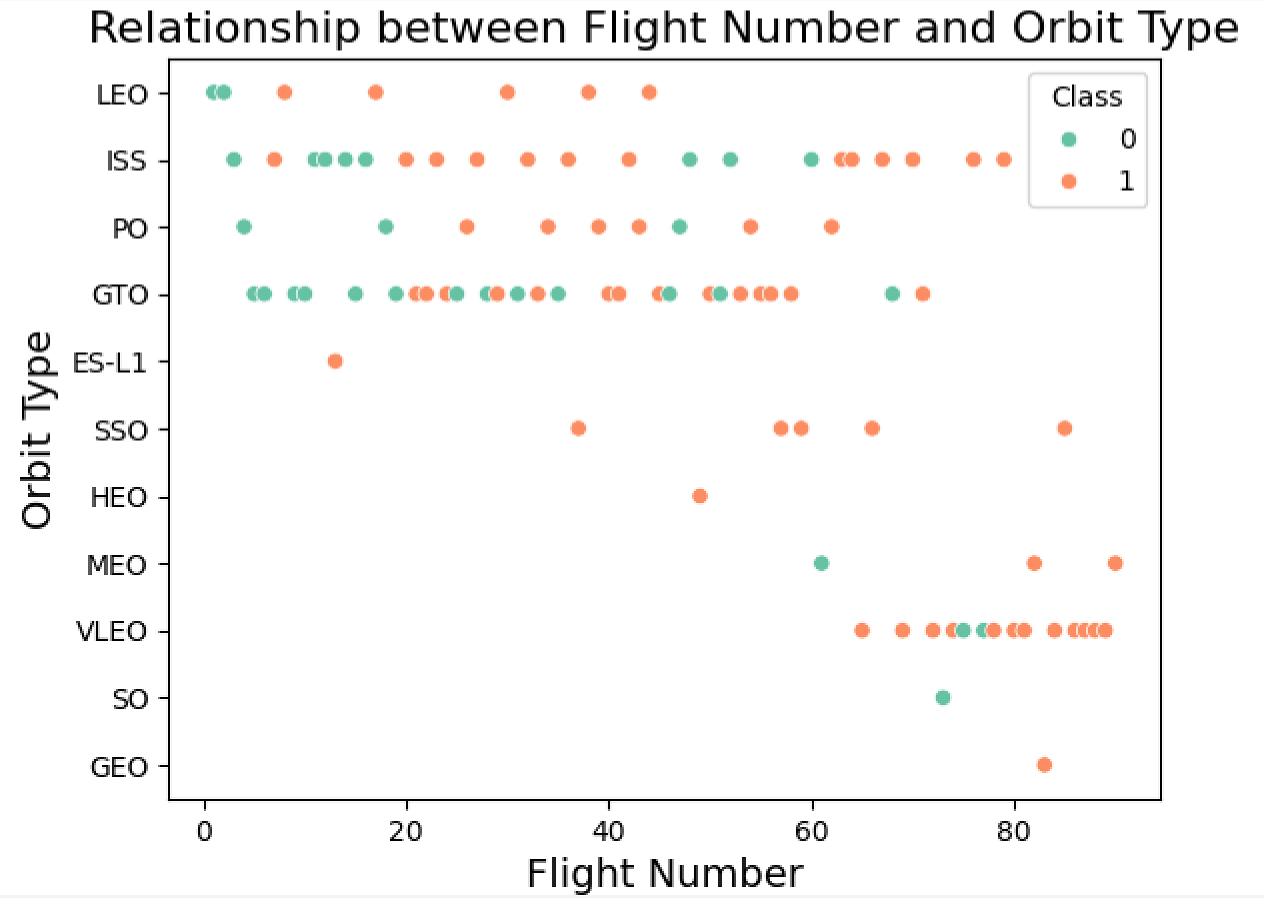
This plot showed the relationship between payload mass and launch site. It helped identify any significant correlation between payload size and site-specific launch success.

Success Rate vs. Orbit Type



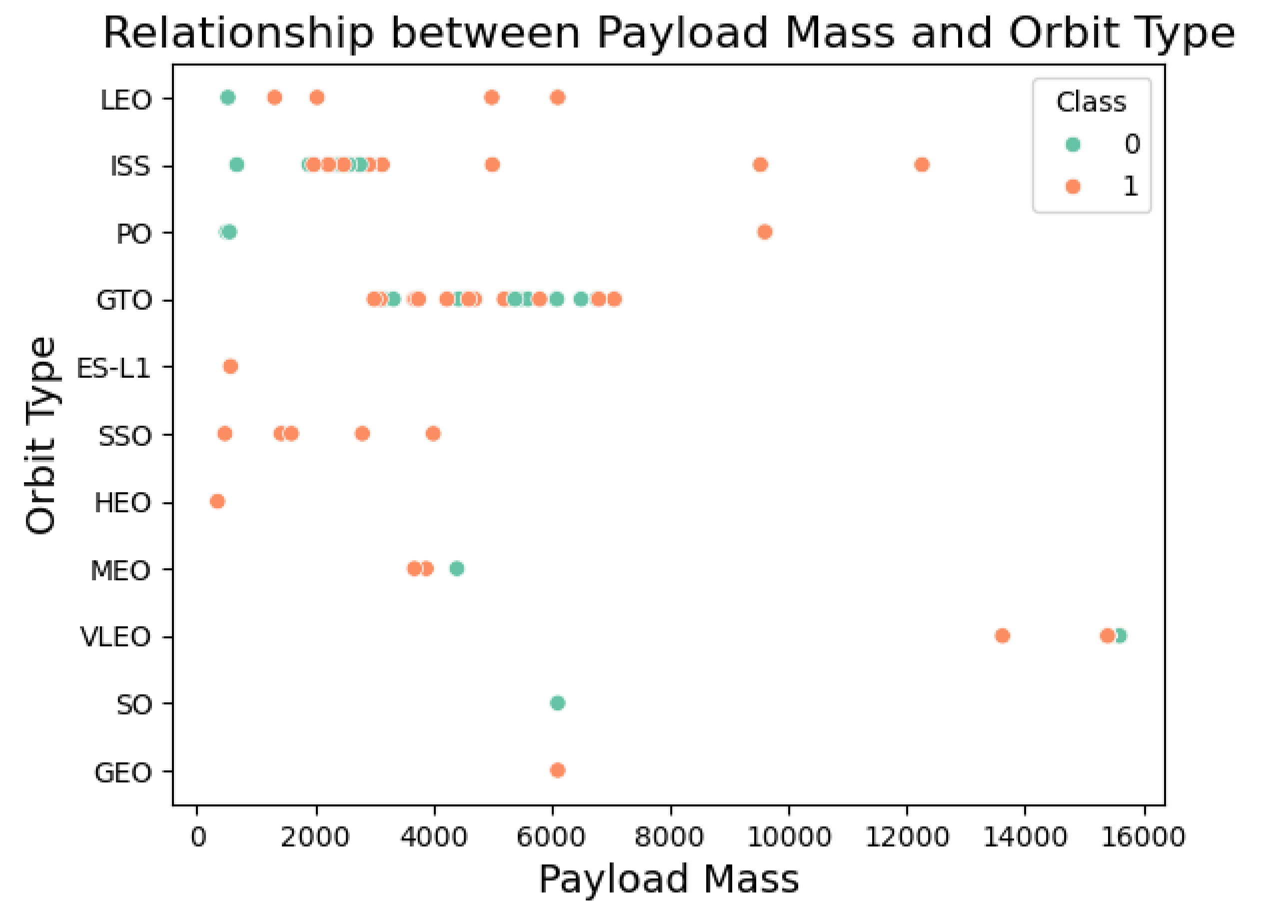
A bar chart was plotted to show the success rate for each orbit type. This chart highlighted which orbits had the highest success rates, guiding SpaceX's future mission planning.

Flight Number vs. Orbit Type



This plot illustrated how the number of flights varied by orbit type, helping identify trends or patterns in SpaceX's mission types over time.

Payload vs. Orbit Type



This plot visualized the relationship between payload and orbit type, revealing if certain payload sizes were associated with specific orbit types.

All Launch Site Names

- Find the names of the unique launch sites

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

A query was executed to list all unique launch site names, providing an overview of all the locations SpaceX operates from.

Launch Site Names Begin with 'CCA'

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Five records of launch sites beginning with "CCA" were retrieved, providing insight into how SpaceX uses specific sites.

Total Payload Mass

4556

The total payload carried by NASA boosters was calculated, helping analyze the capacity and scale of SpaceX's missions.

Average Payload Mass by F9 v1.1

the average payload mass carried by booster version F9 v1.1 is 2928.4

The average payload mass for booster version F9 v1.1 was calculated, revealing payload capabilities specific to this booster type.

First Successful Ground Landing Date

- first successful landing outcome on ground pad was 2015-12-22

- Dates of the first successful ground landings were identified, marking milestones in SpaceX's technological progress.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Boosters that successfully landed on a drone ship with payload mass between 4000 and 6000 kg were listed, providing insights into landing success with specific payload sizes.

Total Number of Successful and Failure Mission Outcomes

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

The total number of successful and failed mission outcomes was calculated, providing an overall success rate for SpaceX missions.

Boosters Carried Maximum Payload

Booster Versions with 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

The boosters carrying the maximum payload were identified, showcasing the largest payload capacity of SpaceX's rockets.

2015 Launch Records

Year	Month Names	Landing Outcomes in drone ship	Booster_Version	Launch_Site
2015	October	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015	April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Failed landing outcomes for drone ships in 2015 were listed, revealing any patterns in failed landings for that year.

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

Landing_Outcome	Total Number Ranked
Success (ground pad)	5
Failure (drone ship)	5

Landing outcomes between 2010-06-04 and 2017-03-20 were ranked by frequency, showing the distribution of successful and failed landings over time.

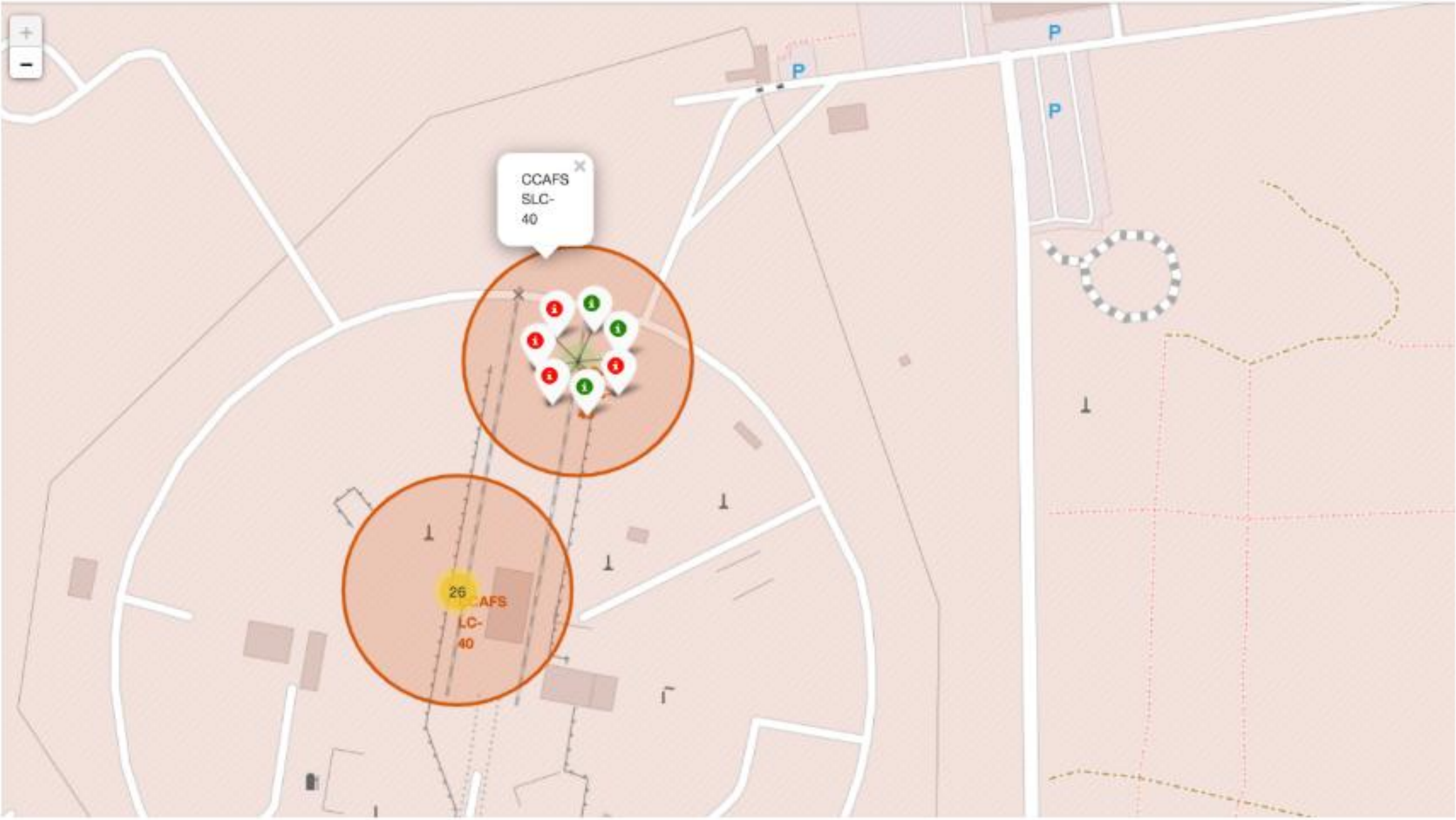
Section 3

Launch Sites Proximities Analysis

Map Visualizations

Screenshots of interactive maps created with Folium, showing launch site locations, outcomes, and proximity to infrastructure (railways, highways, etc.).

Map Visualizations



Section 4

Build a Dashboard with Plotly Dash

Dashboard Screenshots

Screenshots showing interactive dashboards with visualizations like launch success pie charts, payload vs. outcome scatter plots, and success rate analysis.

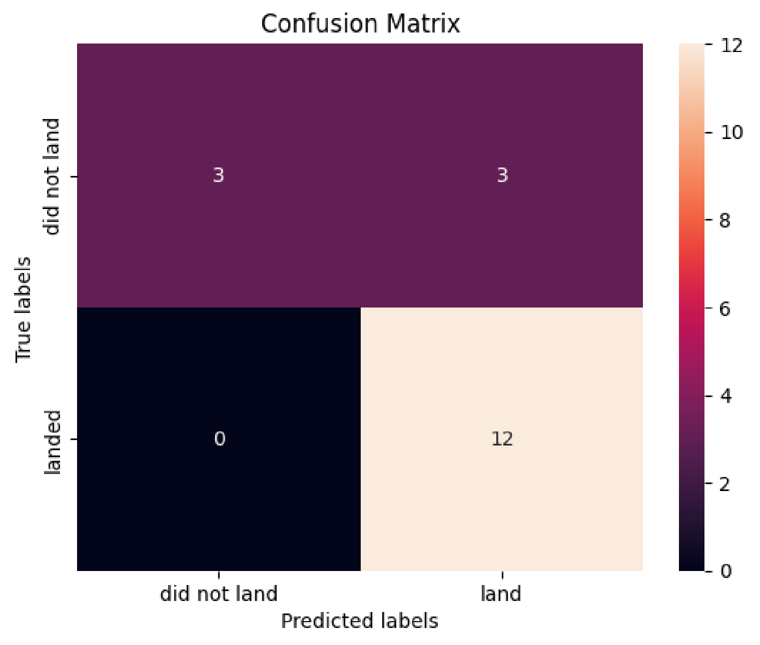
Section 5

Predictive Analysis (Classification)

Classification Accuracy

A bar chart was used to compare the accuracy of different classification models. The best-performing model was identified and chosen for further evaluation.

Confusion Matrix



The confusion matrix of the best model was displayed to show true positives, false positives, true negatives, and false negatives.

Conclusions

- **Key Findings**: Payload and orbit type are critical factors in launch success.
- The best classification model was Logistic Regression, with high accuracy.
- SpaceX's success rate has improved significantly over the years.
- Interactive analytics provided valuable insights for mission planning.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!