SpaceX Launch Analysis & Prediction

Thet Myat Noe Oo

CONTENTS OF THIS TEMPLATE

Executive Summary

Objectives:

- Retrieve and process SpaceX launch data from the API.
- Conduct EDA to identify patterns and relationships within the data.
- Visualize launch site success rates and other relevant metrics.
- Develop and evaluate machine learning models to predict successful rocket landings.
- Create an interactive dashboard for data exploration.

Key Findings:

- Launch Site Trends
- Payload Mass Impact
- Predictive Model Performance
- Folium Map Insights
- Dashboard insights

Introduction

SpaceX has revolutionized the aerospace industry with its reusable rocket technology. Understanding the factors that influence successful landings is crucial for cost reduction and mission success.

Objectives & Research Questions

1. What factors contribute to successful rocket landings?
2. How does payload mass affect launch outcomes?
3. Can we build a predictive model to forecast landing success?

The issues included:

- Identifying all factors that influence the landing outcome.
- The relationship between each variables and how it is affecting the outcome
- The optimal conditions required to enhance the likelihood of a successful landing.

Data Collection & Wrangling Methodology

Data Sources:

SpaceX API: Endpoints: /launches/, /payloads/, /cores/.
Python requests used for data retrieval.

```python
import requests
api_url = "https://api.spacexdata.com/v4/launches/"
response = requests.get(api_url)
data = response.json()
```

Data Cleaning: Missing Values, Data Types, and Duplicates

Key Processes: Used json_nomalize and custom functions to parse nested JSON
Addressed API rate limits
Ensured data consistency

Exploratory Data Analysis (EDA) Methodology

Objective:

- To identify patterns, relationships, and anomalies within the SpaceX launch dataset.
- To formulate hypotheses and guide subsequent predictive modeling.

Technique Employed:

- Descriptive Statistics: Calculated measures of mean, median, and quartiles for numerical features. Used pandas.describe() to summarize data distributions
- Visual Analysis: Histograms and box plots to examine the distribution of individual variables. Count plots for categorical variables. Scatter plots to explore relationships between numerical variables. Bar charts and grouped bar charts to compare categorical variables. Line plots to observe trends in launch outcomes and payload over time.
- SQL Queries: Utilized SQL to aggregate, filter, and summarize data, complementing visual analysis. Performed queries to count launch outcomes by site, aggregate payload mass, and explore launch data over time.

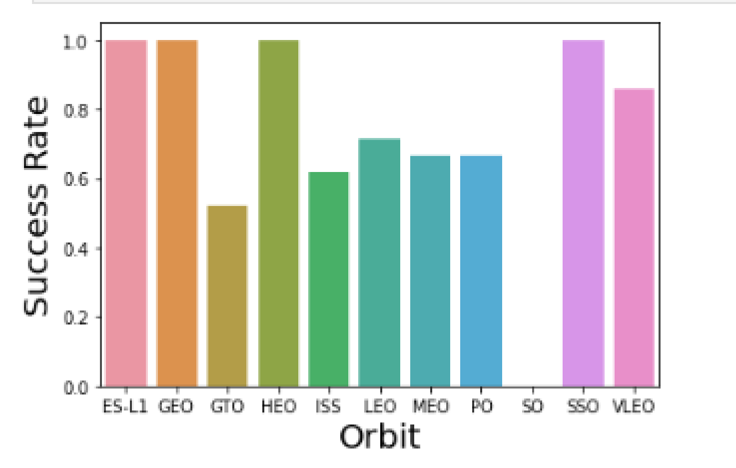
Tools & Libraries:

- **Pandas:** Data manipulation and statistical analysis.
- **Matplotlib:** Basic data visualization.
- **Seaborn:** Advanced statistical visualizations and enhanced aesthetics.
- **SQL:** Data querying and aggregation.

EDA with Visualization Results

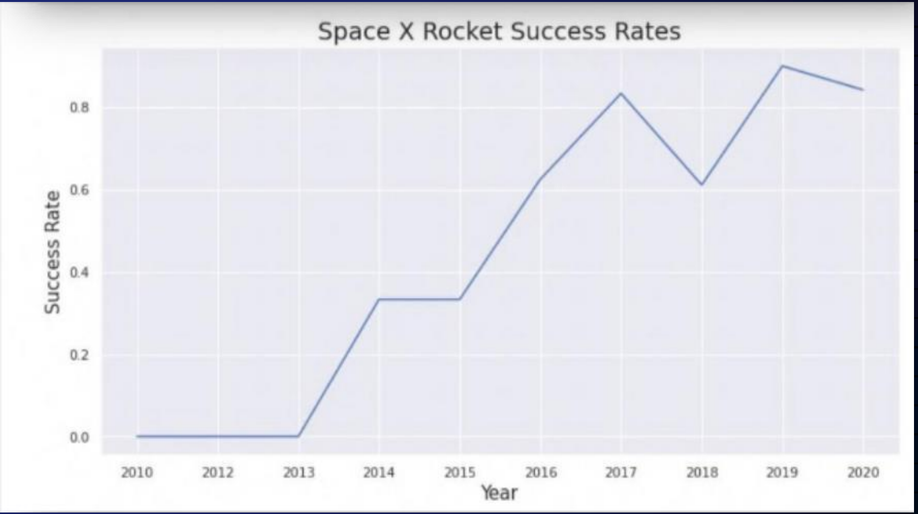
Bar graphs is easy to interpret the relationship between the attributes. This bar graph illustrates the launch outcome over number of launches. It compares unsuccessful launch and successful launch in which unsuccessful launches are half of the successful launch.

https://github.com/ThetMyat23/Data-Science-Final-Project/blob/main/EDA%20with%20Visualization.ipynb



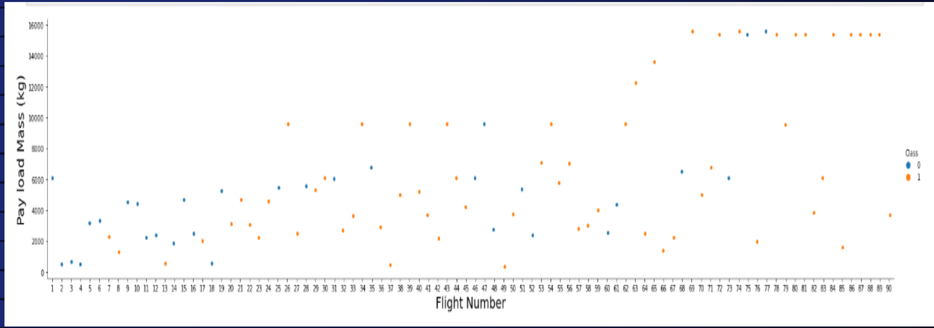
EDA with Visualization Results

This line graph shows Space X Rocket Success Rate. It shows a trend or pattern of the attribute over time which is used to see the launch success yearly trend.

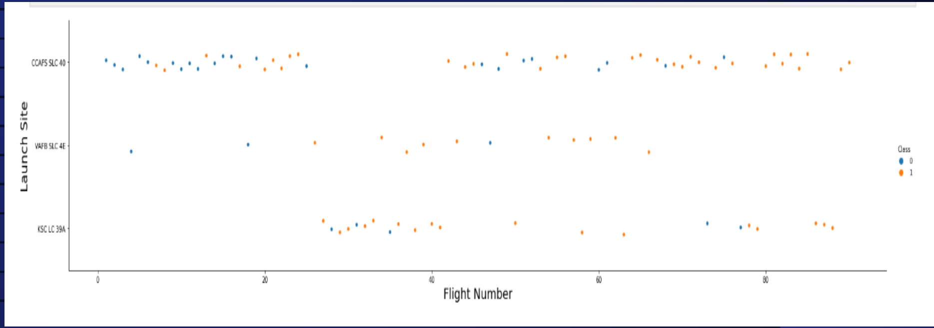


EDA with Visualization Results

Scatter Plot of Flight Number vs. Pay load Mass by Class

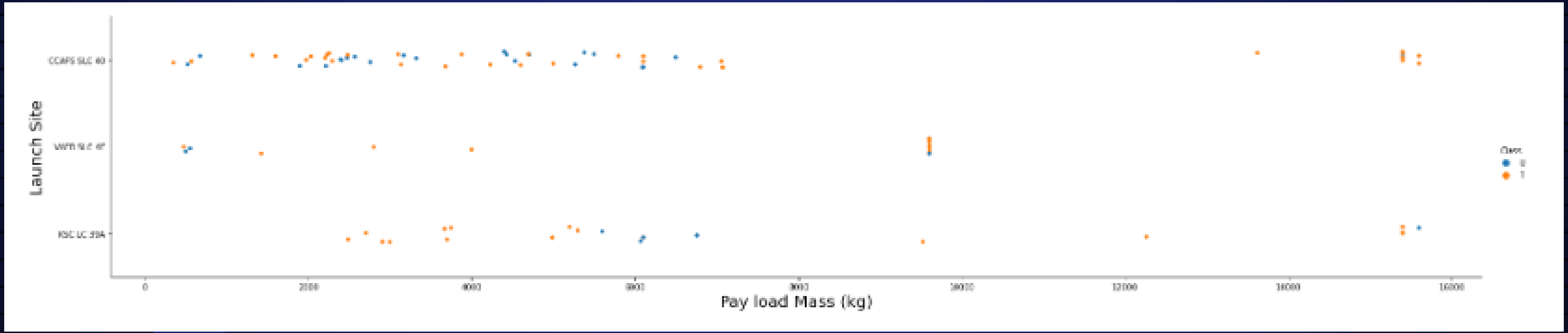


Scatter Plot of Flight Number vs. Launch Site by Class



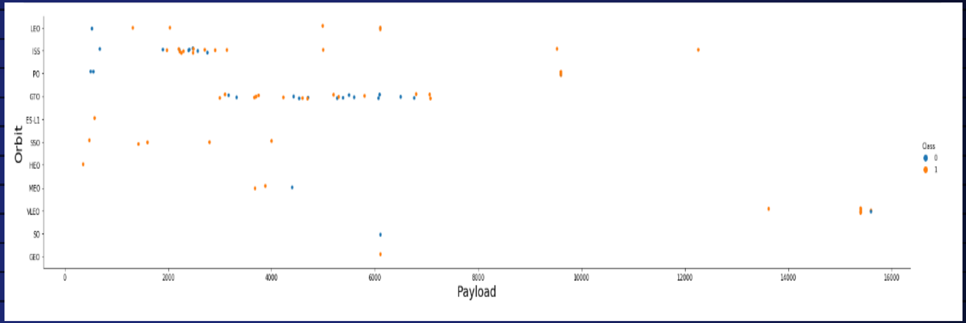
EDA with Visualization Results

Scatter Plot of Payload vs. Launch Site by Class

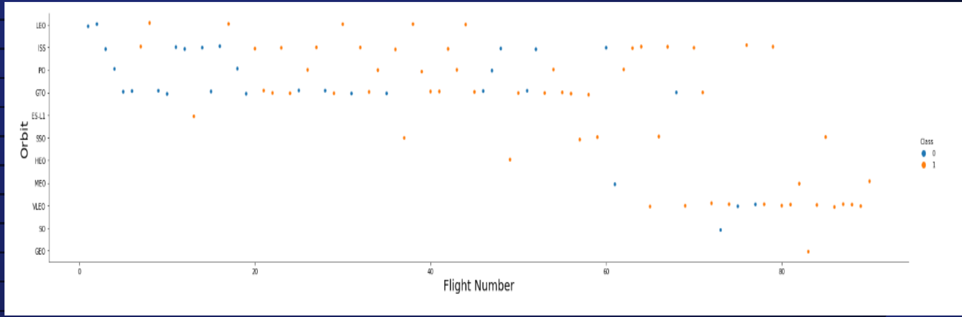


EDA with Visualization Results

Scatter plot of Payload vs. Orbit by Class



Scatter plot of Flight Number vs. Orbit by Class



EDA with SQL Results

Using SQL, we had performed many queries to get better understanding of the dataset.

- Display a simple table or bar chart showing the count of successful vs. failed launches.
- Display the number of successful launches for each launch site.
- Display the average payload mass for successful and failed launches.
- List the total number of successful and failure mission outcomes.

%sql select distinct Launch_Site from SPACEXTBL

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

EDA with SQL Results

%sql select * from SPACETBL where Launch_Site like 'CCA%' limit 5

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

EDA with SQL Results

%sql select PAYLOAD_MASS__KG_ from SPACEXTBL where Customer = 'NASA (CRS)'

PAYLOAD_MASS__KG_
500
677
2296
2216
2395
1898
1952
3136
2257
2490
2708
3310
2205
2647
2697
2500
2495
2268
1977
2972

%sql select avg(payload_mass__kg_) from SPACEXTBL WHERE booster_version = 'F9 v1.1'

avg(payload_mass__kg_)
2928.4

%sql select min(DATE) from SPACEXTBL WHERE landing_outcome = 'Success (ground pad)'

min(DATE)
2015-12-22

EDA with SQL Results

%sql select Booster_Version from SPACEXTBL where Landing_Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000

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

%sql select mission_outcome, count(mission_outcome) from SPACEXTBL GROUP BY mission_outcome

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

EDA with SQL Results

%sql select booster_version, payload_mass__kg_ from SPACEXTBL where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

%sql select Booster_Version, Launch_Site from SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' and Year(Date,7,4)=2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

EDA with SQL Results

%sql select count(landing_outcome), landing_outcome from SPACEXTBL \ where DATE between '2010-06-04' **and** '2017-03-20' group by landing_outcome \ order by count(landing_outcome) desc

https://github.com/ThetMyat23/Data-Science-Final-Project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

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

Interactive Map with Folium

Purpose:

- To geographically represent SpaceX launch sites and provide interactive access to their locations and details.
- To understand the spatial distribution of launch activities and potential correlations with success rates.

Map Features:

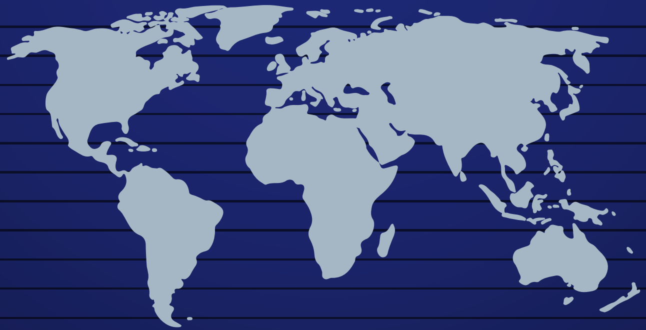
- **Base Map:**
  - Used [Specify base map style, e.g., "OpenStreetMap" or "Stamen Terrain"] for clear geographical context.
  - Centered the map to encompass all launch site locations.
- **Markers:**
  - Placed markers at the latitude and longitude coordinates of each launch site.
  - Used custom marker icons to differentiate sites.

Marker Clusters:

- Implemented MarkerCluster plugin to group nearby markers, improving map readability.
- This prevents overcrowding and allows for easier exploration of clustered sites.

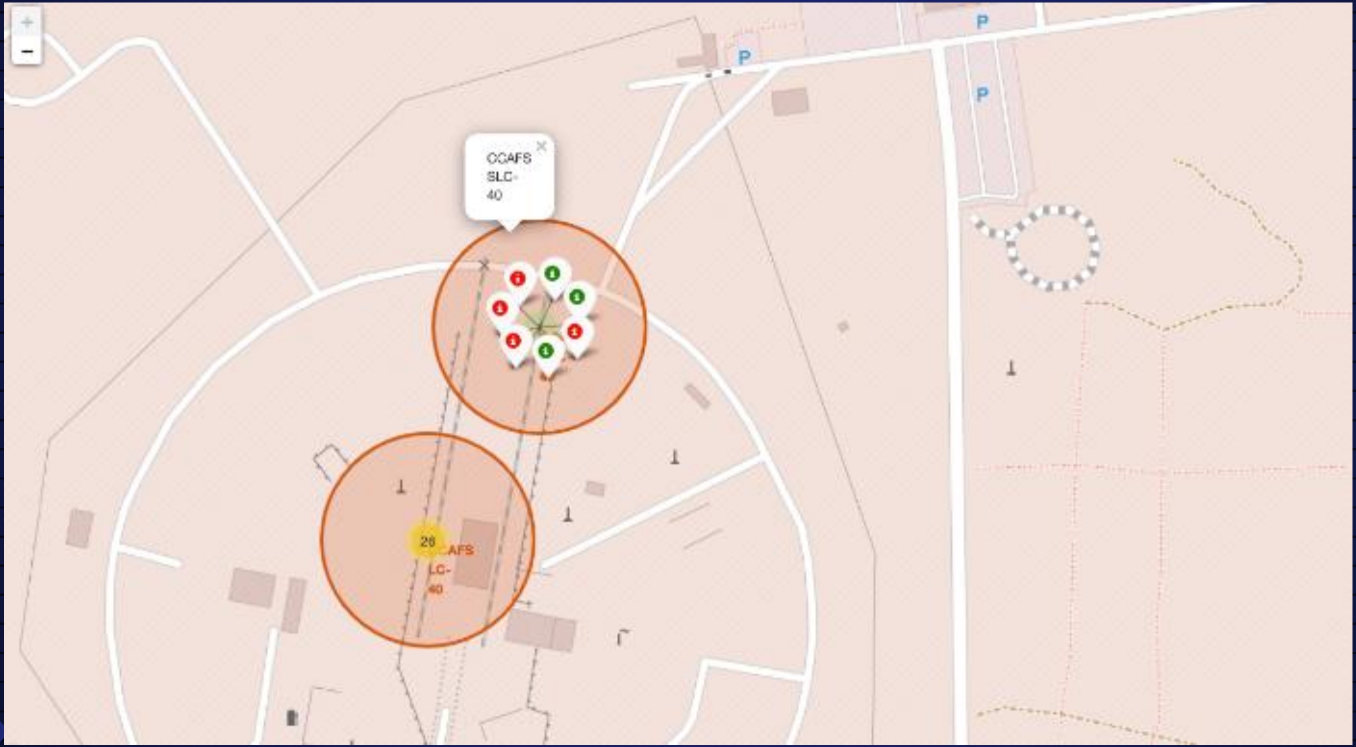
Insights:

- ❑ Describe the distribution of launch sites.
- ❑ Mention any observations about the proximity of launch sites
- ❑ Discuss the ease of accessing of information through the interactive map



Interactive Map with Folium

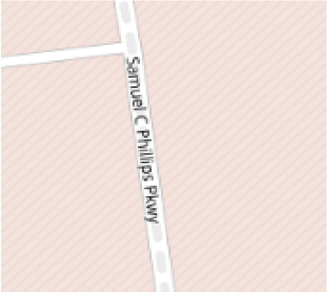
Interactive Map with Folium



Interactive Map with Folium

A railway map symbol may look like this:

A highway map symbol may look like this:



A city map symbol may look like this:

https://github.com/ThetMyat23/Data-Science-Final-Project/blob/main/lab_jupyter_launch_site_location.ipynb

Plotly Dash Dashboard Results

Purpose:

- To provide an interactive platform for exploring SpaceX launch data.
- To enable users to visualize trends, patterns, and relationships through dynamic visualizations.
- To facilitate data-driven decision-making through real-time interaction.

Dashboard Features:

- Dropdown Filters: Launch Site selection, Year selection, Outcome selection
- Interactive Visualizations:
  - Pie Charts
  - Scatter Plots
  - Line Plots

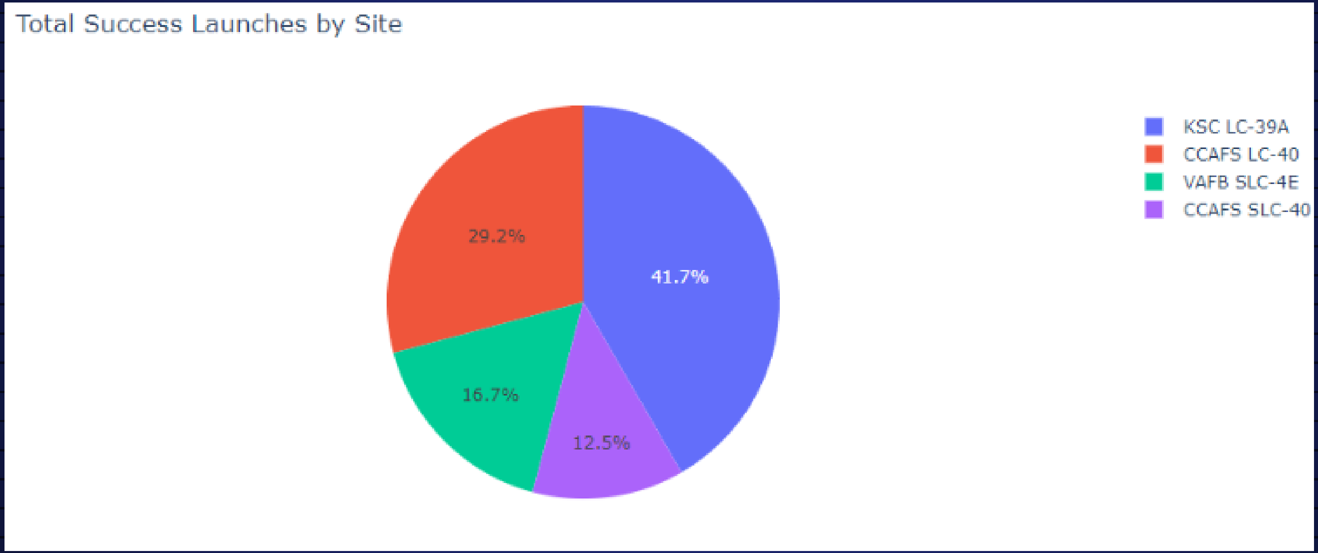
Key Insights & Functionality:

- **Comparative Analysis:** Users can easily compare launch success rates across different sites or time periods.
- **Trend Identification:** The dashboard facilitates the identification of trends and patterns in launch data.
- **Data Exploration:** Users can explore the relationships between different variables through interactive scatter plots.
- **Customizable Views:** Users can tailor the visualizations to their specific interests using the dropdown filters.

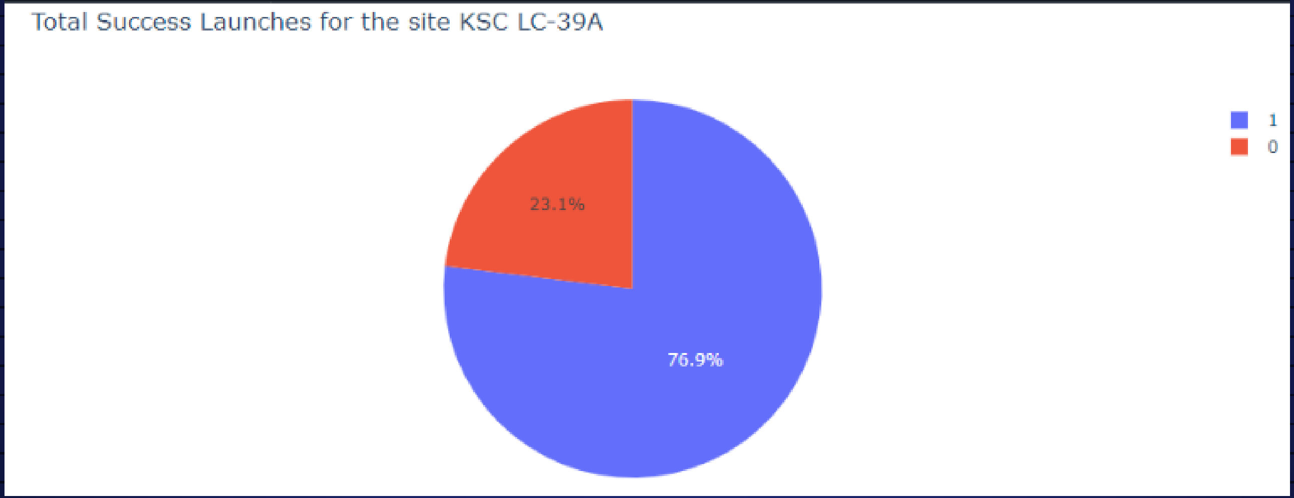
Benefits:

✓ Enhanced Data Exploration
✓ Improved Decision-Making
✓ Accessibility

Plotly Dash Dashboard Results



Plotly Dash Dashboard Results



Plotly Dash Dashboard Results

Predictive Analysis Methodology

Objective:

- To develop and evaluate classification models capable of predicting whether a SpaceX launch will result in a successful landing (Class 1) or failure (Class 0).

Building the Model:

- Load the dataset into NumPy and Pandas
- Transform the data and then split into training and test datasets
- Decide which type of ML to use
- Set the parameters and fit it to dataset

Evaluating the Model:

- Check the accuracy for each model
- Get tuned hyperparameters for each type of algorithms
- Plot the confusion matrix

Improving the Model:

- Use Feature Engineering and Algorithm Tuning
- Divided the dataset into training and testing sets using train_test_split

Find the Best Model:

- The model with the best accuracy score will be the best performing model

Performance Metrics:

- **Accuracy:** Overall correctness of predictions.
- **Precision:** Proportion of correctly predicted positive cases.
- **Recall:** Proportion of actual positive cases correctly predicted.
- **Confusion Matrix:** Visual representation of true positives, true negatives, false positives, and false negatives.
- **Classification Report:** Provides detailed metrics for each class.

Predictive Analysis Results

Key Findings:

- The Random Forest model achieved the highest accuracy, making it the best predictor for landing success.
- Payload mass and launch site were significant features influencing the landing outcome.
- The classification models effectively distinguish between successful and unsuccessful landings.

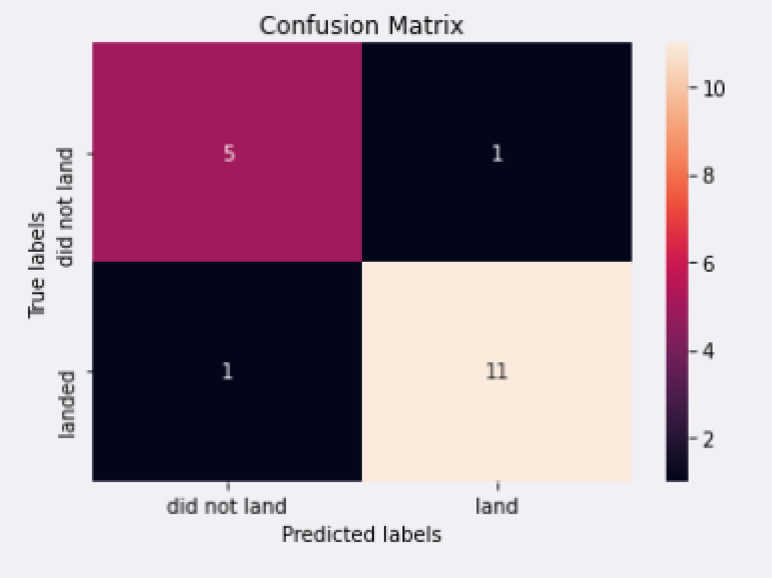
They can be categorized into many results:

- Data Collection and Wrangling Result
- Folium Map Results
- SQL Query Insights
- Exploratory Data Analysis Results

Predictive Analysis Results

Confusion Matrix of the best Model

It shows 11 true positive which is correctly predicted successful launches, 5 true negative results which is correctly predicted failed launches, 1 false positive which is incorrectly predicted successful launches, and 1 false negative which is incorrectly predicted failed launches.



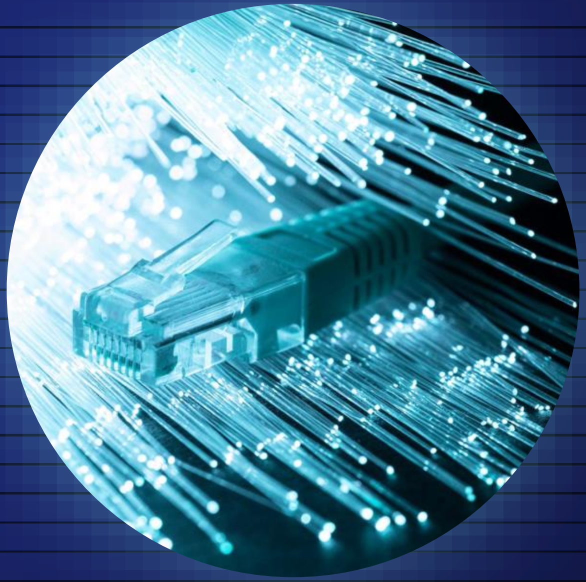
CONCLUSIONS

The low number of false negatives indicates that the model is effective at identifying successful launches.

Through EDA, we identified critical trends in launch site success rates and payload mass impact. Predictive models, particularly Random Forest, demonstrated strong accuracy in forecasting launch success. These insights can help SpaceX optimize operations, assess risks, and improve mission reliability.

This analysis showcases the power of data science in extracting valuable insights from complex aerospace data. The findings provide a strong foundation for future advancements in launch technology and operational efficiency.

Thank You



Credits to Slides go