Capstone Presentation

based on the SpaceX Launch Data Analysis project

Executive Summary

Objective:
This project analyzes SpaceX launch data to determine the success rates of various launch sites and boosters, using Exploratory Data Analysis (EDA), SQL, visual analytics, and predictive modeling.

Outcome:

- Built interactive dashboards and visualizations using Plotly Dash
- Performed statistical analysis and classification models to predict launch success
- Delivered insights for improving mission planning and risk mitigation

Introduction

Problem Statement:
SpaceX aims to reduce spaceflight costs and improve reliability. Our task was to analyze historical launch data to:

- Understand trends
- Identify influential factors in launch success
- Build predictive models for future launches

Approach:
Followed the Data Science Methodology:

- Data Collection
- Data Wrangling
- Exploratory Analysis
- Modeling
- Visualization
- Deployment

Data Collection & Wrangling

Sources:

- SpaceX Launch Data (CSV from provided URL or API)
- Launch Site Coordinates (for Folium)
- Additional tables from Wikipedia using web scraping

Wrangling Tasks:

- Handled missing values
- Converted date columns to datetime
- Merged datasets
- Created new features like “Launch Success” (binary), “Payload Mass (kg)”, “Launch Site”

EDA & Interactive Visual Analytics Methodology

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data.

- Used **Pandas** for statistical summaries
- Created bar charts, pie charts, and scatter plots using **Plotly**
- Designed dropdowns and interactivity with **Dash**
- Used **Folium** to build interactive maps showing geographic distribution of launches

EDA with Visualization Results

Include these key visual results :

1. **Launch Success by Site** – Bar chart
2. **Payload Mass vs. Launch Success** – Scatter plot with regression
3. **Success Rate by Booster Version** – Grouped bar chart
4. **Time-Series of Launches** – Line chart
5. **Correlation Heatmap** – Show relationships (e.g., payload and success)
6. **Pie Chart** – Proportion of successful vs. failed launches

EDA with SQL Results

Show **SQL queries and results** from your SQLite or SQL notebook. Include:

1. Total number of launches
2. Launches per site
3. Success rate by launch site
4. Booster version with the highest success rate
5. Launch success by orbit type
6. Most frequent launch site
7. Total payloads launched per customer
8. Year-wise launch count
9. Min/max payload mass
10. Failed launches and reasons (if data available)

Interactive Map with Folium

Embed screenshot or GIF of the interactive Folium map:

- Map showing all launch sites
- Circle markers with popup info (e.g., site name, success rate)
- Highlighted successful sites in green, failures in red

Plotly Dash Dashboard Results

Show screenshots of the Dash app with these components:

- Dropdown to select launch site
- Range slider for payload mass
- Real-time update of charts
- Display of success rates based on selection

Predictive Analysis Methodology

Objective:
Predict the probability of a successful launch based on payload, booster type, and site.

Techniques Used:

- Feature encoding using `OneHotEncoder`
- Normalization of payload
- Classification models: Logistic Regression, SVM, Decision Tree
- Grid Search for hyperparameter tuning

Predictive Analysis Results

Display model comparison (accuracy, F1-score, confusion matrix)

Best performing model: e.g., **Logistic Regression with 92% accuracy**

Feature importance plot

Prediction examples:

Payload: 5000kg, Site: CCAFS LC-40 → **Prediction: Success**

Conclusion

- Launch site and payload are major factors in success
- CCAFS SLC-40 and KSC LC-39A have high success rates
- Heavier payloads decrease success likelihood
- Predictive model helps SpaceX plan better missions