IBM Capstone

Ghanshyam Paunikar

Contents

Summary

Executive Summary

- **Overview:** This project analyzes SpaceX Falcon 9 launch data to predict first-stage landing success, leveraging cost-saving insights.
- **Approach:**
  - Collected data via SpaceX API and Wikipedia scraping.
  - Performed data cleaning, exploratory analysis, and visualization.
  - Built interactive maps and dashboards.
  - Applied machine learning for prediction.
- **Results:** Identified key factors influencing landing success and the best predictive model.
- **Demo:** Screenshots available at [Your-GitHub-Link-2] (e.g., link to visuals folder).

Introduction

- **Context:** SpaceX's Falcon 9 reduces launch costs to $62M (vs. $165M by competitors) by reusing its first stage.
- **Objective:** Predict if the first stage will land successfully using public data and machine learning.
- **Key Questions:**
  - How do payload mass, launch site, and orbit type impact landing success?
  - Has landing success improved over time?
  - Which algorithm excels at this binary classification?

Methodology

Data Collection:

- SpaceX REST API (e.g., FlightNumber, PayloadMass, Outcome).
- Web scraping from Wikipedia (e.g., Launch Site, Booster Landing).

Data Preparation:

- Filtered Falcon 9 launches, handled missing values, encoded outcomes (1 = success, 0 = failure).

Analysis Tools: Visualization, SQL, Folium maps, Plotly Dash dashboards, classification models.

Data Analysis & Visualizations

- **Exploratory Data Analysis (EDA):**
  - Scatter plots: Payload Mass vs. Flight Number, Launch Site vs. Success.
  - Bar charts: Success rate by orbit type.
  - Line charts: Yearly success trends.
- **Insights:**
  - Higher success at KSC LC-39A and VAFB SLC-4E.
  - Success rate rises with flight number and time.

Interactive Tools

- **Folium Map:**
  - Mapped launch sites with success/failure markers (green = success, red = failure).
  - Measured distances to nearby features (e.g., KSC LC-39A to coast: 14.99 km).
- **Plotly Dash Dashboard:**
  - Dropdown for launch site selection.
  - Pie chart of success rates.
  - Slider for payload range and scatter plot of payload vs. success.
- **Demo Links:** [Your-GitHub-Link-6] (Folium), [Your-GitHub-Link-7] (Dash app).

Predictive Modeling

Process:

- Standardized data and split into training/testing sets.
- Tested Logistic Regression, SVM, Decision Tree, and KNN with GridSearchCV (10-fold CV).

Results:

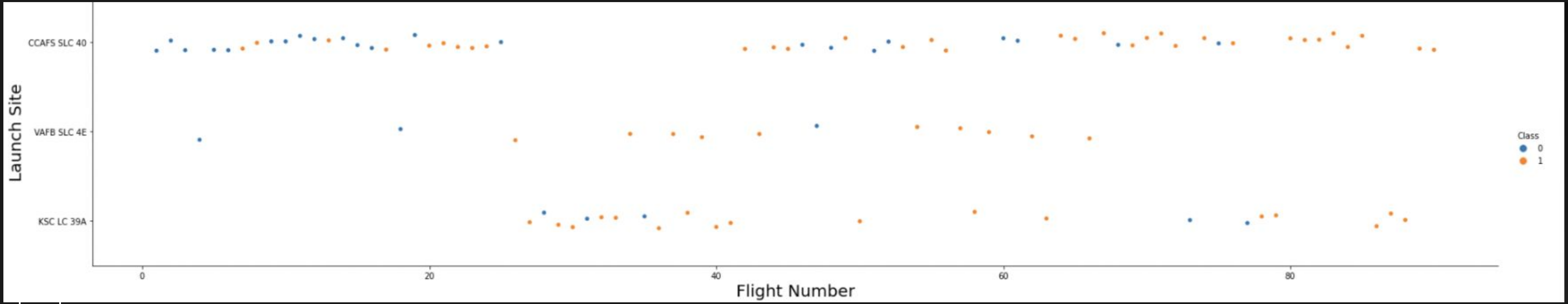
- Decision Tree outperformed with 91.11% accuracy on test set.
- Confusion matrix highlighted false positives as a challenge.

Model Code: [Your-GitHub-Link-8] (e.g., classification script).

Key Findings

- **EDA Insights:**
  - Launch sites near equator/coast (e.g., KSC LC-39A) have higher success.
  - Payloads < 9500 kg correlate with better outcomes.
  - Success rate increased from 2013 to 2020.
- **Model Outcome:** Decision Tree is optimal for predicting landing success.
-

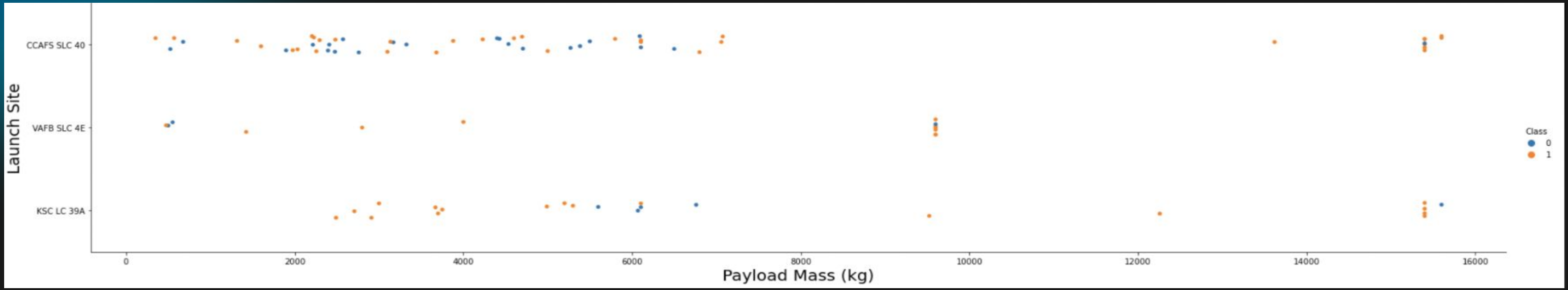
Flight Number vs. Launch Site



Explanation

Explanation:
- The earliest flights all failed while the latest flights all succeeded.
- The CCAFS SLC 40 launch site has about a half of all launches.
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- It can be assumed that each new launch has a higher rate of success.

Payload vs. Launch Site



Explanation:

• For every launch site the higher the payload mass, the higher the success rate.

• Most of the launches with payload mass over 7000 kg were successful.

• KSC LC 39A has a 100% success rate for payload mass under 5500 kg too.

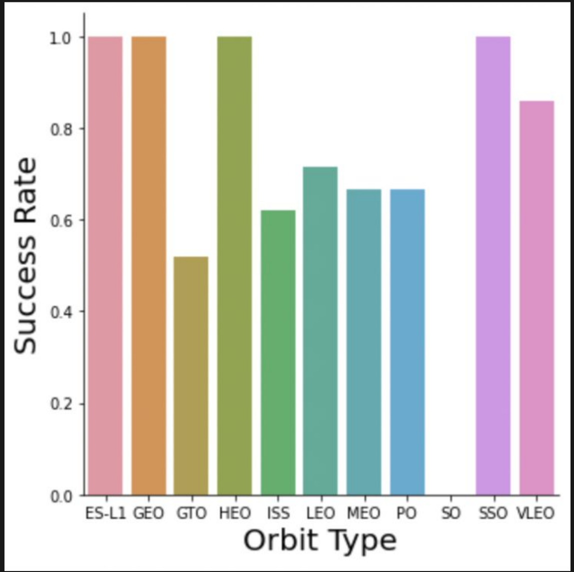
Success rate vs. Orbit type Explanation:

Explanation:

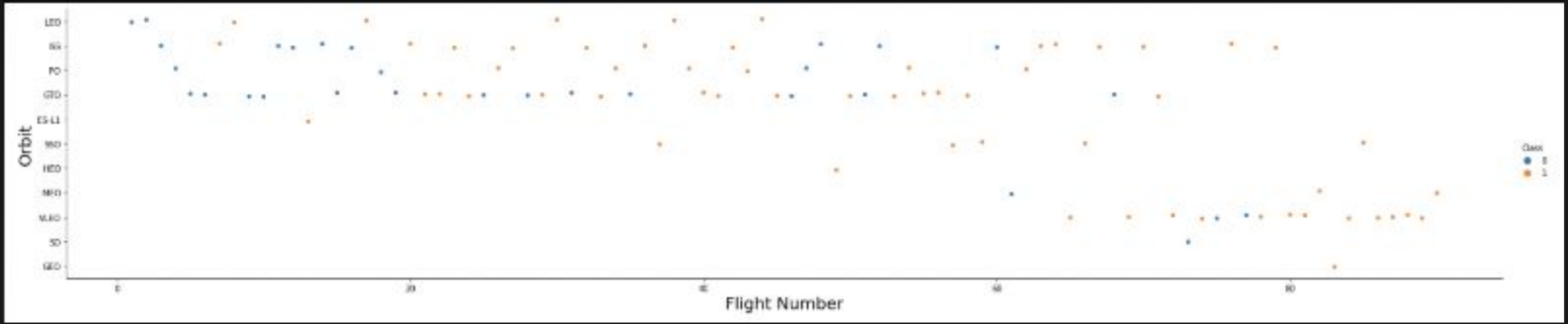
• Orbits with 100% success rate:
- ES-L1, GEO, HEO, SSO

• Orbits with 0% success rate:
- SO

• Orbits with success rate between 50% and 85%:
- GTO, ISS, LEO, MEO, PO



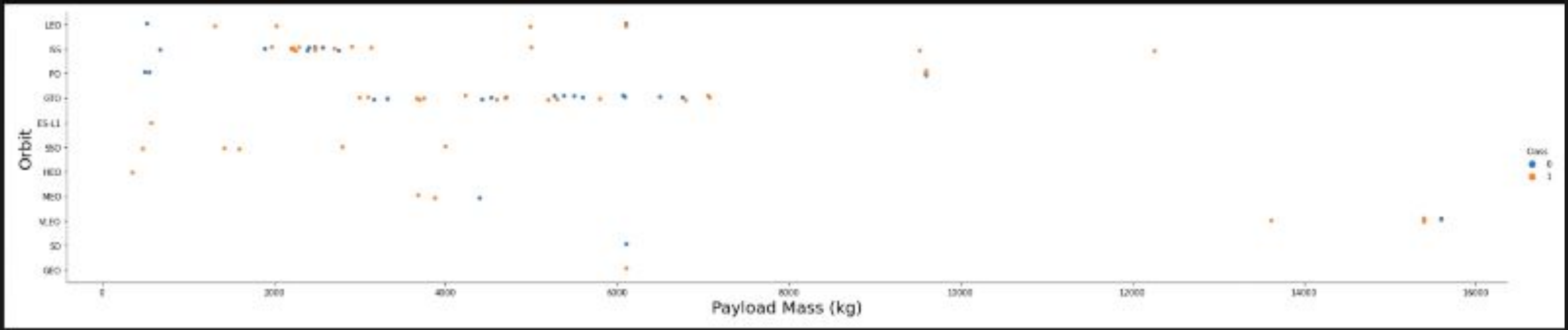
Flight Number vs. Orbit type



Explanation:

• In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

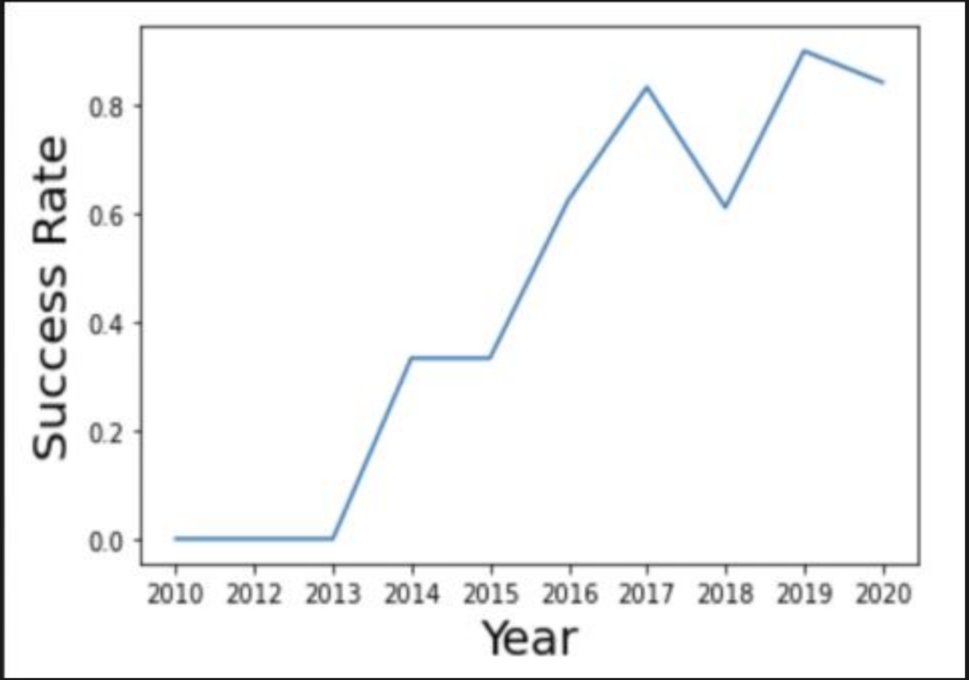
Payload Mass vs. Orbit type



Explanation:

- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch success yearly trend



Explanation:

• The success rate since 2013 kept increasing till 2020.

All launch site names

```
In [4]: %sql select distinct launch_site from SPACEXDATASET;

 * ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[4]:
```

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

Explanation:

• Displaying the names of the unique launch sites in the space mission.

Conclusion

Findings:

- Success tied to site location, payload <9500 kg, and time.
- Decision Tree is optimal (91.11% accuracy).

Appendix

Thanks: IBM instructors, SpaceX data providers.
Tools: Python, Pandas, Scikit-learn, Folium, Plotly Dash.