IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Naphon Santisukwongchot>
<15/11/2024>

Executive Summary

In this project, we aimed to predict the success or failure of SpaceX rocket launches using machine learning techniques. The process involved several key steps: web scraping to collect launch data, data exploration and preprocessing, applying exploratory data analysis (EDA) techniques, creating visualizations, and implementing machine learning models to make predictions. Below is a summary of the methodologies used and the results obtained at each stage.

Summary of Methodologies

- Web Scraping & Data Collection
- Exploratory Data Analysis (EDA)
- Data Visualization
- Machine Learning Models

Summary of Accuracy

Logistic Regression: 83%

K-Nearest Neighbors (KNN): 83%

Support Vector Machine (SVM): 83%

Decision Tree : 61%

Introduction

SpaceX, a leading private aerospace manufacturer and space transportation company, has conducted numerous launches over the years, often with varying outcomes. Understanding the factors that contribute to the success or failure of these launches is crucial for improving their reliability and operational efficiency. With the increasing number of launches and the complexity of space missions, predicting the success of a launch based on historical data could provide valuable insights. In this project, we focus on predicting the success or failure of SpaceX rocket launches based on several key features of each mission. By using data from past launches, including payload details, launch sites, rocket versions, and other mission characteristics, we aim to build predictive models that can classify launches as successful or failed. The goal is to enhance understanding of the factors that influence launch outcomes and to create a tool that can help SpaceX and other space agencies optimize their planning and execution of future launches.

- What factors influence the success or failure of SpaceX rocket launches?
- Can we predict the success or failure of a launch based on historical data?
- Which machine learning models perform the best in predicting launch outcomes?
- What is the success rate of launches over time?
- How can we visualize launch data to uncover patterns or trends?
- Can we predict future launch outcomes based on past trends and mission characteristics?

Methodology

Data for this project was sourced from a publicly available dataset, containing detailed records of SpaceX's rocket launches. The dataset included information about launch sites, payloads, launch outcomes, rocket types, and other relevant mission features. Data was accessed in CSV format and loaded into a Pandas dataframe for processing.

- **Data Cleaning:**
  Missing values were handled by either replacing them with suitable imputed values (e.g., mean, median) or removing rows with missing values.
  Inconsistent data formatting was fixed, including date formats and categorical values like launch outcomes.
  Irrelevant columns and duplicates were removed to ensure the dataset was concise and focused on key features.
- **Feature Engineering:**
  New features were derived, such as splitting the date column into year, month, and day.
  Categorical features like LaunchSite, Orbit, and BoosterVersion were encoded using techniques like One-Hot Encoding to facilitate machine learning model training.
- **Data Type Conversion:**
  Data types were converted as necessary, for example, converting the date columns to datetime objects and numeric columns to float type to enable numerical analysis.

Data Collection

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

In [33]:
```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets
```

We should see that the request was successfull with the 200 status response code

In [39]:
```
response = requests.get(static_json_url)
data = response.json
data
```

In [47]:
```
# Use json_normalize meethod to convert the json result into a dataframe
import pandas as pd

if response.status_code == 200:
    # Parse the JSON content
    data = response.json()  # Call .json() to decode the response as JSON
else:
    print("Request failed with status code:", response.status_code)

# Now, let's use json_normalize to convert the JSON result into a DataFrame
df = pd.json_normalize(data)
```

Data Wrangling

In [73]:
```
data_falcon9.isnull().sum()
```

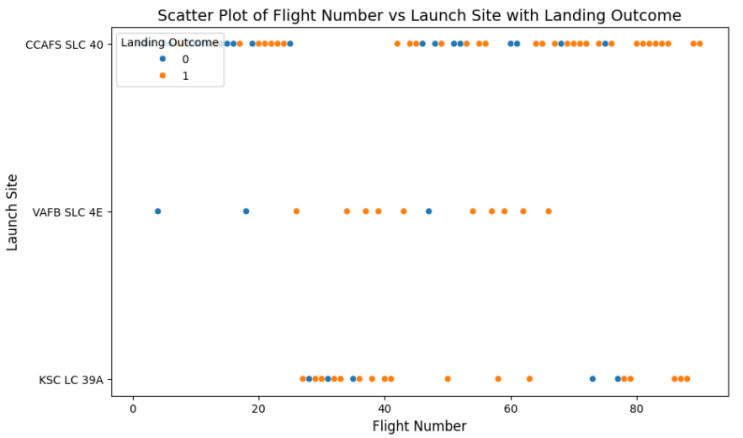
In [75]:
```
# Calculate the mean value of PayloadMass column
payload_mass_mean = data_falcon9['PayloadMass'].mean()
print(f"Mean of PayloadMass: {payload_mass_mean}")

# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'] = data_falcon9['PayloadMass'].replace(np.nan, payload_mass_mean)
print(data_falcon9['PayloadMass'].isna().sum())
```

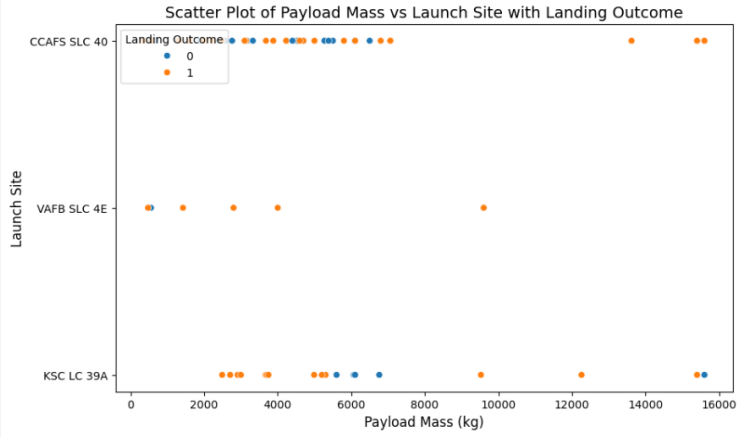
```
Mean of PayloadMass: 6123.547647058824
0
```

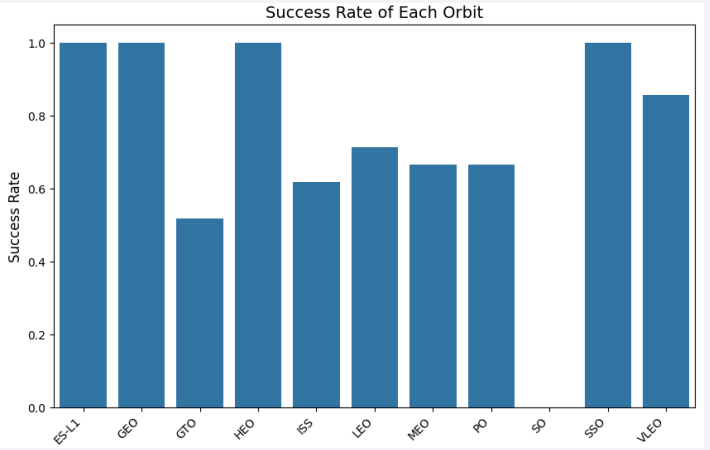
https://github.com/naphon1999/IBM_Data_Science/blob/main/Capstone_Project/jupyter-labs-webscraping.ipynb

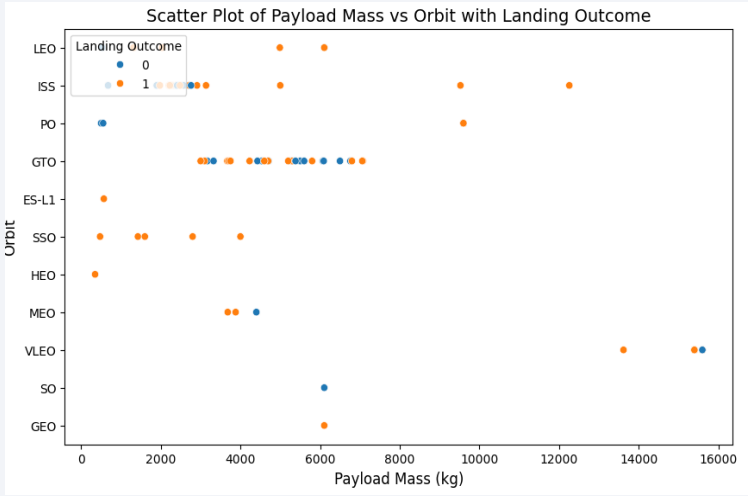
https://github.com/naphon1999/IBM_Data_Science/blob/main/Capstone_Project/jupyter-labs-spacex-data-collection-api.ipynb

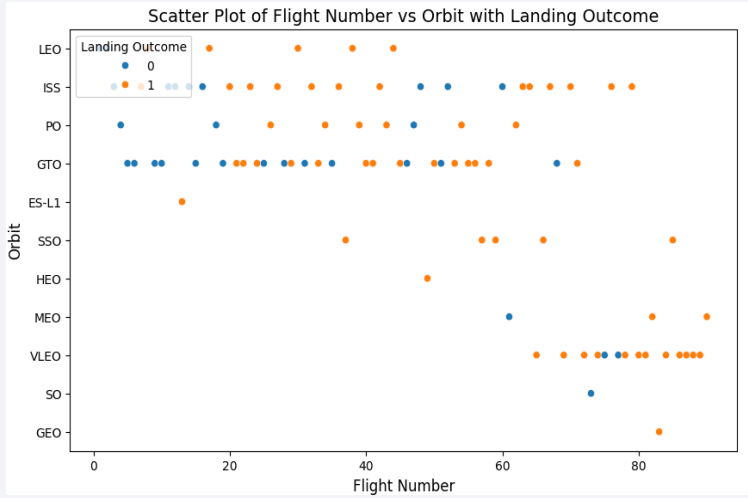
EDA with Data Visualization

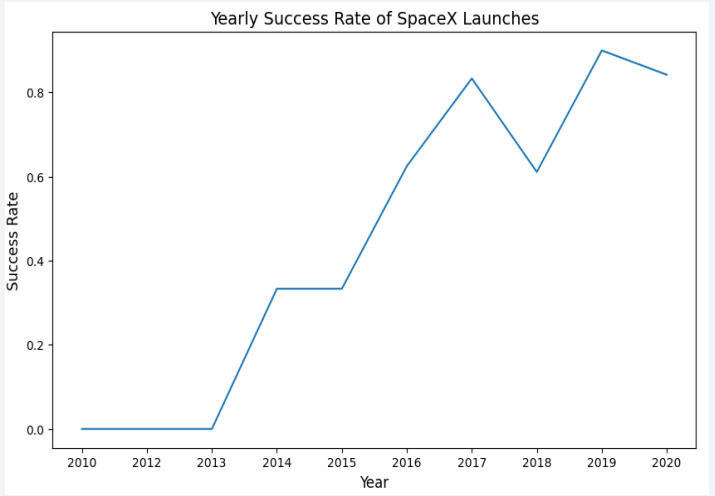












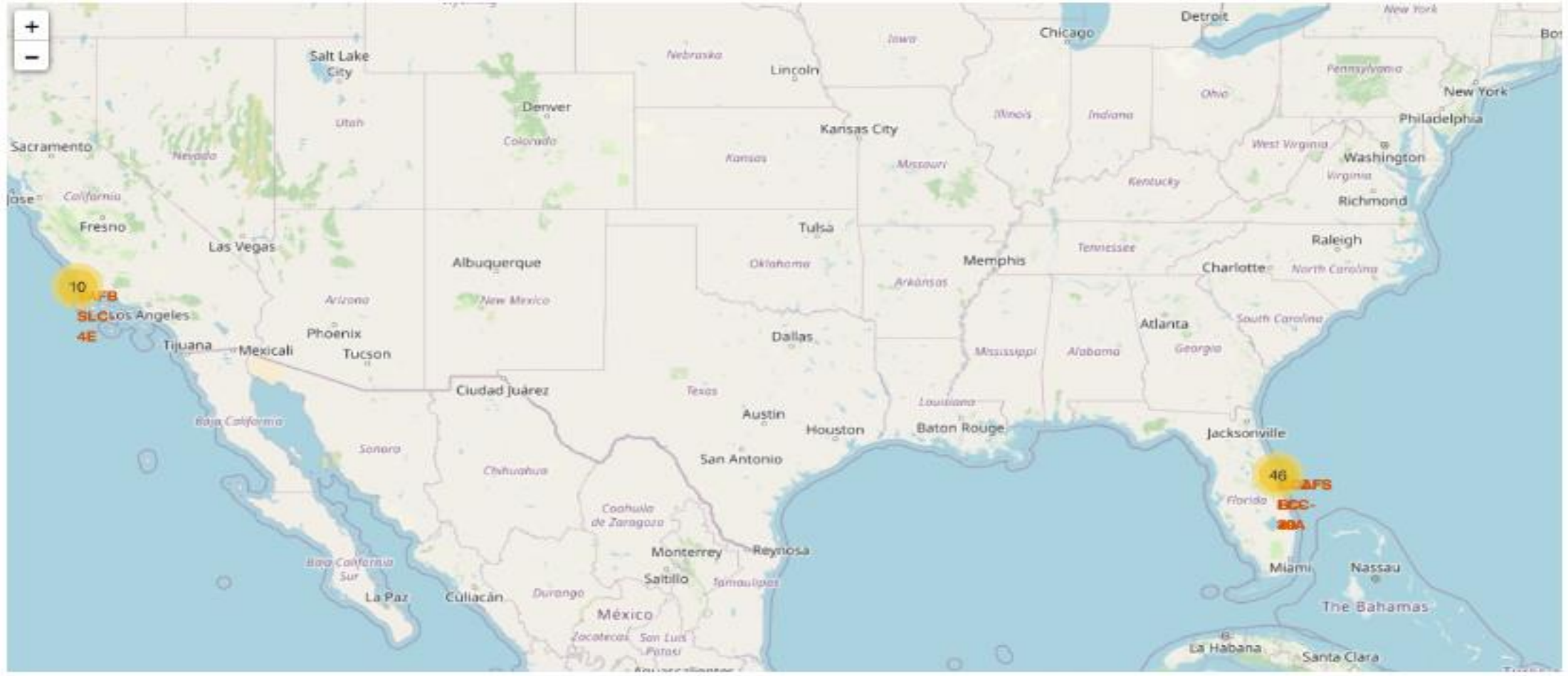
https://github.com/naphon1999/IBM_Data_Science/blob/main/Capstone_Project/edadataviz.ipynb

EDA with SQL

- Unique Launch Sites
- Launch Sites Starting with 'CCA'
- Total Payload Mass by NASA (CRS)
- Average Payload Mass by Booster Version
- First Successful Landing Outcome in Ground Pad
- Booster Versions with Payload Mass Between 4000 and 6000
- Count of Successful and Failed Mission Outcomes.
- Booster Versions with Maximum Payload Mass.
- Failure Landing Outcomes in Drone Ship for 2015.
- Landing Outcome Count Between 2010-06-04 and 2017-03-20.

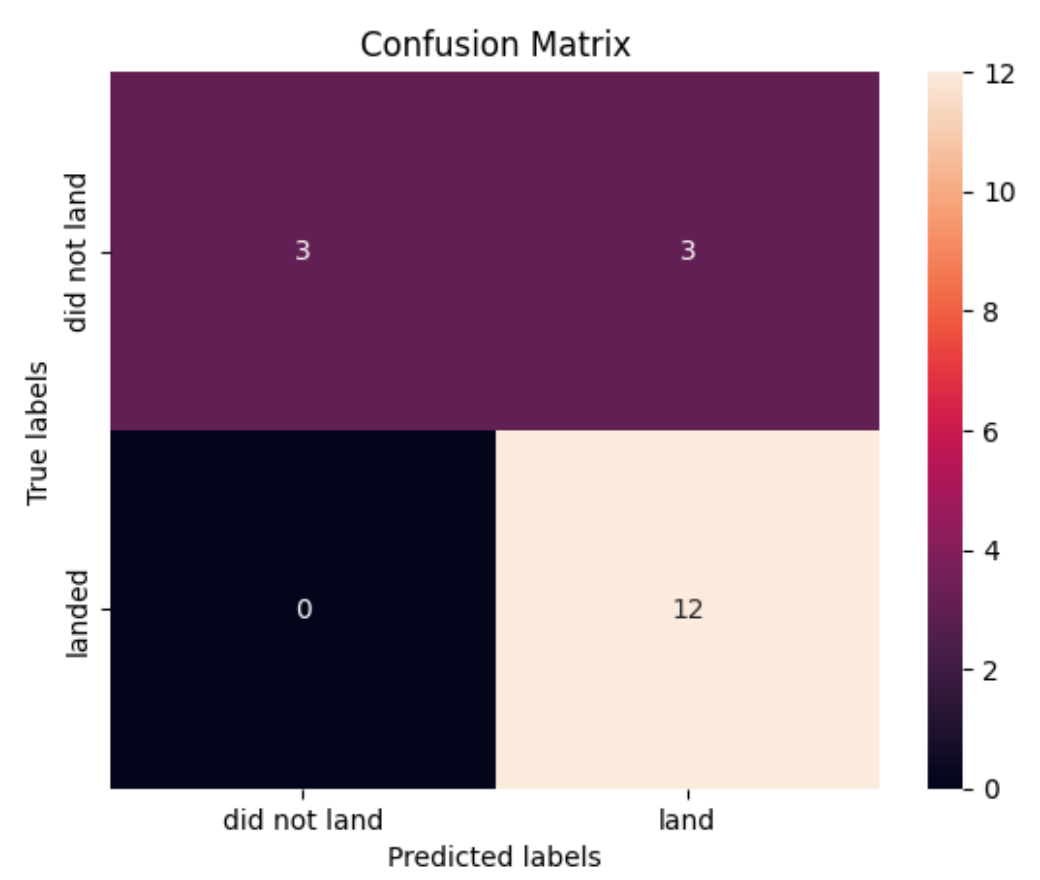
https://github.com/naphon1999/IBM_Data_Science/blob/main/Capstone_Project/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium



https://github.com/naphon1999/IBM_Data_Science/blob/main/Capstone_Project/lab_jupyter_launch_site_location.ipynb

Predictive Analysis (Classification)



https://github.com/naphon1999/IBM_Data_Science/blob/main/Capstone_Project/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Conclusions

In this project, we successfully analyzed SpaceX mission data through several key stages, employing a variety of methodologies to explore, visualize, and predict outcomes. Below are the main conclusions drawn from the process.

Overall, this project demonstrated the power of data science techniques in understanding complex datasets and predicting outcomes. The integration of SQL, Python libraries, and machine learning models provided valuable insights into SpaceX’s operations and helped identify factors influencing mission success rates.

Thank you!