IBM Data Science: Capstone Project

Agenda

EXECUTIVE SUMMARY

INTRODUCTION

METHODOLOGY

RESULTS

CONCLUSION

APPENDIX

Executive Summary

- Methodology:
  - Data collection
  - Data wrangling
  - EDA with data viz
  - EDA with SQL
  - Interactive map with Folium
  - Dashboard with Plotly Dash
  - Predictive Analytics
- Findings:
  - EDA, map, dashboard, and predictive analytics results

Introduction

SpaceX is the commercial leader in rocket development and space travel with reusable rockets leading to massive cost savings.

Research:

- How do operational variables such as payload mass affect successful landings?
- Has landing success increased over the years?
- What classification models can be used to predict landing success?

Methodology

Data Collection:

- SpaceX Rest API, web scraping

Data Wrangling:

- Filtering, missing values, encoding

EDA using data viz and SQL

Visual analytics using Folium and Plotly

Classification models

- Model building, training, testing, evaluation

Methodology

Data Collection:

- SpaceX Rest API, web scraping

Data Wrangling:

- Filtering, missing values, encoding

EDA using data viz and SQL

Visual analytics using Folium and Plotly

Classification models

- Model building, training, testing, evaluation

Data

SpaceX Rest API Data:

- FlightNumber, Date, BoosterVersion, PayloadMass, Orbiot, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Web Scraping from Wikipedia:

- Flight No., Launch Site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

EDA Data Viz

- Charts were plotted to determine relationships between variables
- Scatter plots show the linear relationships between variables and if present, can be used in ML models
- Bar charts were used to compare categories
- Line charts were used to depict trends historically

EDA SQL

SQL queries were performed on data to:

- Filter
- Analyze
- List
- Count
- Rank

Using SQL, it was possible to rank landing outcomes and count mission failures and successes over time.

Folium Map

Built an interactive map with Folium which included:

- Markers of all launch sites
- Colored markers for launch outcomes
- Distances between launch sites

Plotly Dashboard

Built a Plotly dashboard which included:

- Launch Sites Dropdown List
- Pie Chart showing operational successes
- Slider which depicted payloads
- Scatter charts of payload and success for different boosters

Classification

Modeling process:

- Standardization with StandardScaler
- Splitting into train and test using train_test_split
- Creating a GridSearchCV object where cv = 10
- LogReg, SVM, Decision Tree, KNN
- Accuracy using method.score()
- Confusion matrix
- Jaccard_score for optimal method

Results

Many failures in SpaceX's early history with many successes in recent times

VAFB SLC 4E and KSC LC 39A have highest success rates

The higher the payload mass, the higher the success rate

Launches over 7000 kg were mostly successful

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

Since 2013, major growth in success rate with approximately 80% success following 2020

GitHub Results

GitHub repository with all relevant files and analysis:

https://github.com/cr2230/ibmdatascience