,

Data Science Capstone Project Presentation

Niloufar Baba Ahmadi

02.09.2023

Overview

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Methods Employed:

- Data Collection: Web scraping and SpaceX API.
- Exploratory Data Analysis (EDA): Data wrangling, visualization, and interactive analytics.
- Machine Learning Predictions: Applied for insights.

Key Findings:

- Valuable data collected from diverse sources.
- EDA identified crucial launch success predictors.
- Machine Learning highlighted key success factors.

Introduction

Project Goal:

- Assess Space Y's potential to compete effectively with Space X.

Desired Inquiries Include:

- Determining the most accurate method for estimating the total launch cost by predicting successful first-stage rocket landings.
- Identifying the optimal launch location.

Methodology

Here is a refined summary of our data collection and analysis methodology:

Data Collection Methodology:

1. We sourced data from Space X through two primary channels:
   - Space X API (Accessed at: `https://api.spacexdata.com/v4/rockets/`)
   - Web Scraping (Referenced at: `https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches`)
2. After obtaining the data, we conducted data wrangling to prepare it for analysis.
3. We enhanced the collected data by introducing a landing outcome label based on the results of our feature summarization and analysis.

Exploratory Data Analysis (EDA):

- We performed an extensive exploratory data analysis (EDA), utilizing visualization techniques and SQL queries to gain valuable insights.

Interactive Visual Analytics:

- We employed interactive visual analytics tools such as Folium and Plotly Dash to facilitate a deeper understanding of the data.

Predictive Analysis:

1. Our analysis involved predictive modeling using classification algorithms.
2. We normalized the data up to this stage and divided it into training and test datasets.
3. We evaluated the data using four distinct classification models, assessing the accuracy of each model through various parameter combinations.

This methodology allowed us to comprehensively analyze the data from multiple sources, enhance its quality, and extract meaningful insights to support our evaluation of Space Y's competitiveness with Space X.

Results

Exploratory Data Analysis Findings:

- Space X operates from four different launch sites.
- Initial launches were conducted for Space X itself and NASA.
- The average payload for the F9 v1.1 booster is approximately 2,928 kg.
- The first successful landing occurred in 2015, five years after the first launch.
- Many Falcon 9 booster versions achieved successful landings on drone ships, typically with payloads surpassing the average.
- Nearly 100
- In 2015, two booster versions (F9 v1.1 B1012 and F9 v1.1 B1015) failed to land on drone ships.
- Over time, the success rate of landings improved.

By employing interactive analytics, we were able to identify that launch sites tend to be located in safe areas, often near bodies of water, and are well-equipped with strong logistic infrastructure. Furthermore, we observed that the majority of launches take place at launch sites on the east coast.

Predictive Analysis Results:

- In our predictive analysis, we determined that the Decision Tree Classifier emerged as the most effective model for predicting successful landings. It achieved an accuracy rate exceeding 87

Conclusion

- Summarize the project's key takeaways
- Reiterate the problem statement and its solution
- Mention any limitations and potential future work

Conclusion

Based on the insights drawn from the analysis:

- Currently, the most favorable launch site is CCAF5 SLC 40, boasting a high success rate in recent launches.
- The second and third positions are held by VAFB SLC 4E and KSC LC 39A, respectively.
- Overall, there is an observable improvement in the success rate of launches over time.

Further observations include:

- Launches carrying payloads exceeding 9,000 kg (approximately the weight of a school bus) tend to have an excellent success rate.
- Payloads surpassing 12,000 kg appear feasible primarily at CCAFS SLC 40 and KSC LC 39A launch sites.

- The highest success rates are associated with specific orbits, such as ES-L1, GEO, HEO, and SSO, followed by VLEO (with a rate above 80%) and LFO (with a rate above 70%).
- Over time, success rates have notably improved for all orbits, with VLEO emerging as a promising business opportunity due to its recent increase in frequency.
- Interestingly, there seems to be no significant correlation between payload weight and success rates for GTO orbits.
- The ISS orbit exhibits a wide range of payload options and maintains a good success rate.
- There are relatively fewer launches to SO and GEO orbits.
- The success rate began increasing in 2013 and continued until 2020, with the initial three years marked as a period of adjustments and technological improvements.

Throughout the data analysis process, multiple data sources were scrutinized, leading to refined conclusions, including:

- KSC LC-39A being identified as the top launch site.
- Launches involving payloads above 7,000 kg being less risky.
- An overall improvement in the success of landing outcomes over time, reflecting advancements in processes and rocket technology.
- The recommendation of using the Decision Tree Classifier for predicting successful landings to enhance profitability.

Thank You

- Thank you for your attention!