Rocket/Data Science

U M Khalid (27th May 2024)

Content

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Done by: U Mohamed Khalid

Github Link: https://github.com/villathi-villain/IBMDataScience

Executive Summary

We employed a comprehensive approach to analyze SpaceX launch data, involving:

- Data Collection: Web scraping and SpaceX API to gather valuable public data.
- Exploratory Data Analysis (EDA): Data wrangling, visualization, and interactive analytics to identify key patterns.
- Machine Learning Prediction: Developing models to pinpoint the most influential features for launch success.
- Results Synthesis: Summarizing key findings.

Our EDA identified crucial predictors of launch success, which informed our machine learning models. These models highlighted the most impactful characteristics, guiding strategies to enhance launch outcomes. This analysis showcases our ability to leverage data for strategic decision-making in space exploration.

Introduction

Objectives

Our objective is to assess the feasibility of the new company, Space Y, in competing with SpaceX.

The key questions we aim to answer include:

- Cost Estimation: Identifying the best methods to predict the total cost of launches by forecasting the successful landings of the first stage of rockets.
- Optimal Launch Sites: Determining the most advantageous locations for conducting launches.

Methodology

Methodology - Data Collection & Data Wrangling

Executive Summary

- Data Collection:
  - Data from SpaceX was obtained from 2 sources:
    - SpaceX API; https://api.spacexdata.com/v4/rockets/
    - Web Scraping
- Data Wrangling
  - Optimized Data
- EDA with visualisation and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis with classification models

Data Collection - SpaceX API

- SpaceX Falcon 9 first stage Landing Prediction
  - https://github.com/villathi-villain/IBMDataScience/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

- Request to the SpaceX API Established
- Cleaned the requested data

Predictive Analysis (Classification)

Logistic regression, support vector machine, decision tree and k nearest neighbours

https://github.com/villathi-villain/IBMDataScience/blob/main/7-SpaceX_Machine_Learning_Prediction_Part_5.ipynb

Insights from EDA

SQL Summary Slide

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

TOTAL_PAYLOAD
111268

AVG_PAYLOAD
2928.4

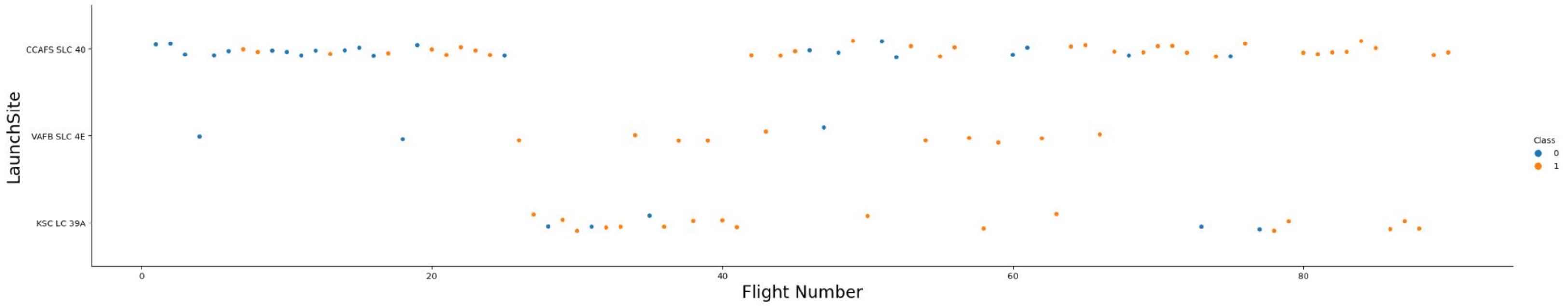
Mission_Outcome	QTY
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

FIRST_SUCCESS_GP
2015-12-22

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Landing_Outcome	QTY
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

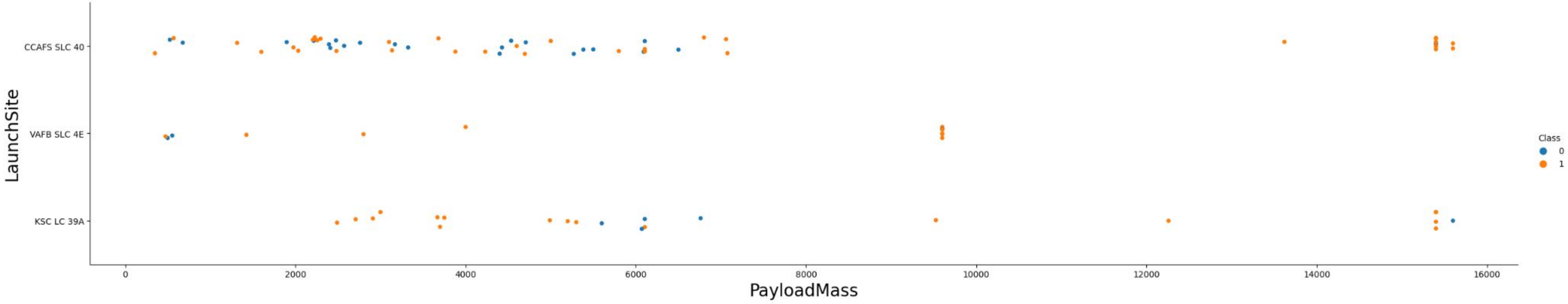
Flight Number V Launch Site



CCAF5 SLC 40 is probably the better launch site

- followed by VAFB SLS 4E and KSC LC 39A

Payload V Launch Site



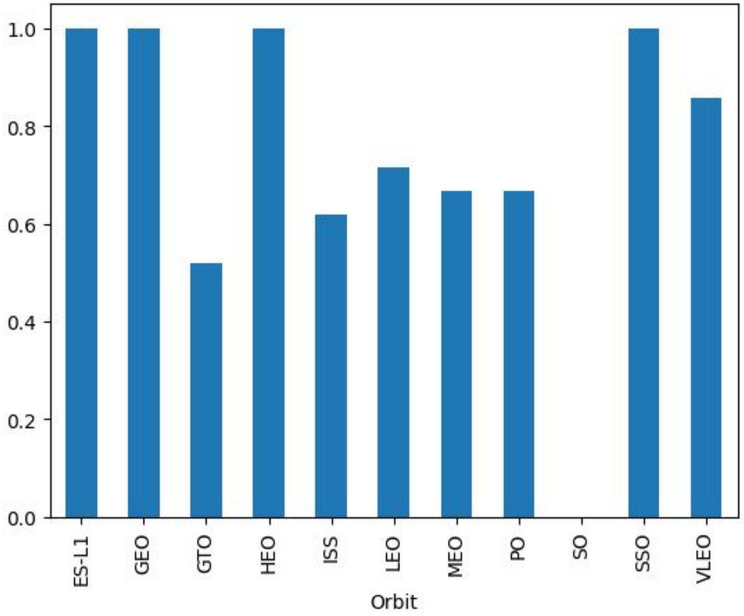
- Payloads more than 10,000 kg seems only possible with CCAF5 SLC 40 & KSC LC 39A
- CCAF5 SLC 40 has been the more frequent choice for payloads below 8,000 kg

The biggest success rates happens to orbits:

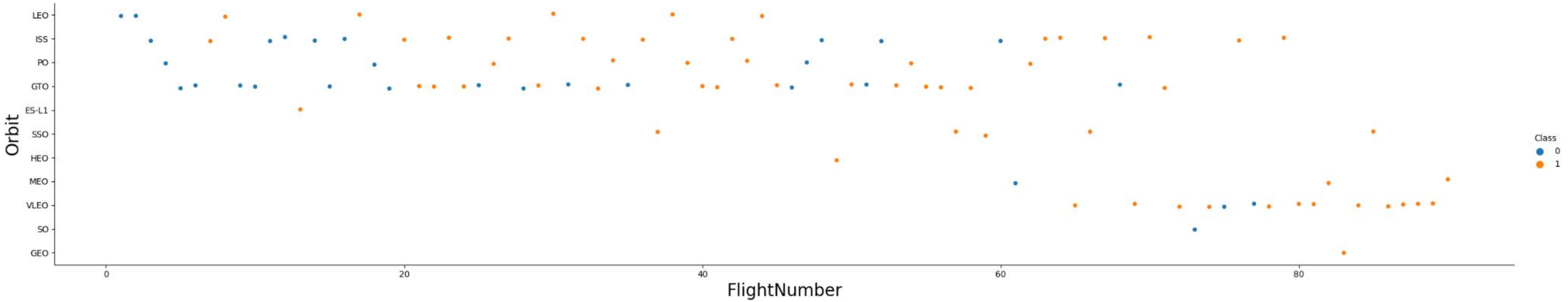
- ES-L1;
- GEO;
- HEO;
- SSO

Followed by:

- VLEO
- LFO



Flight Number V Orbit Type

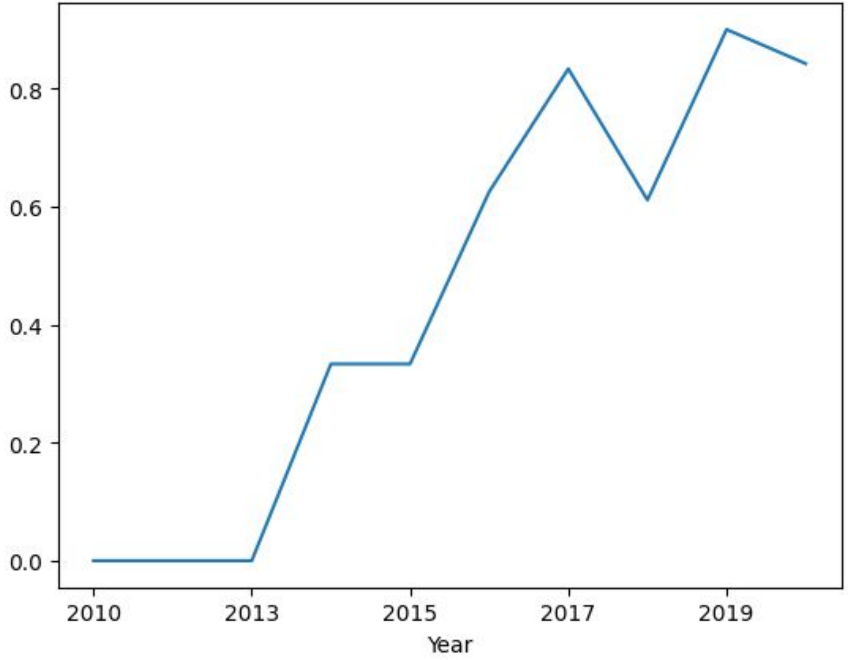


The success rate has been improving over time to all orbits;

VLEO orbit has a higher recent frequency

Launch Success Trend

you can observe that the success rate since 2013 kept increasing till 2020



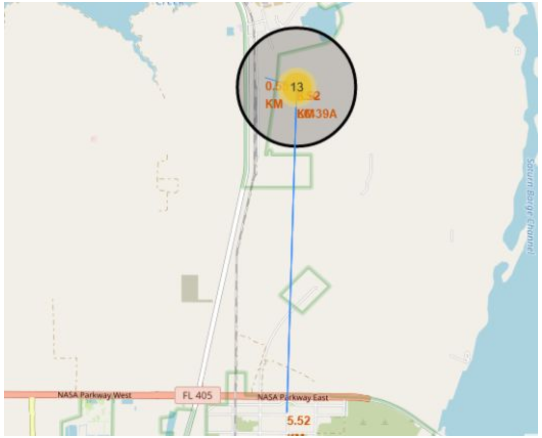
Launch Sites Proximities Analysis

All Launch Sites

The launches as you can see has always been in coastal areas

- This is for the concern of safety and convenience

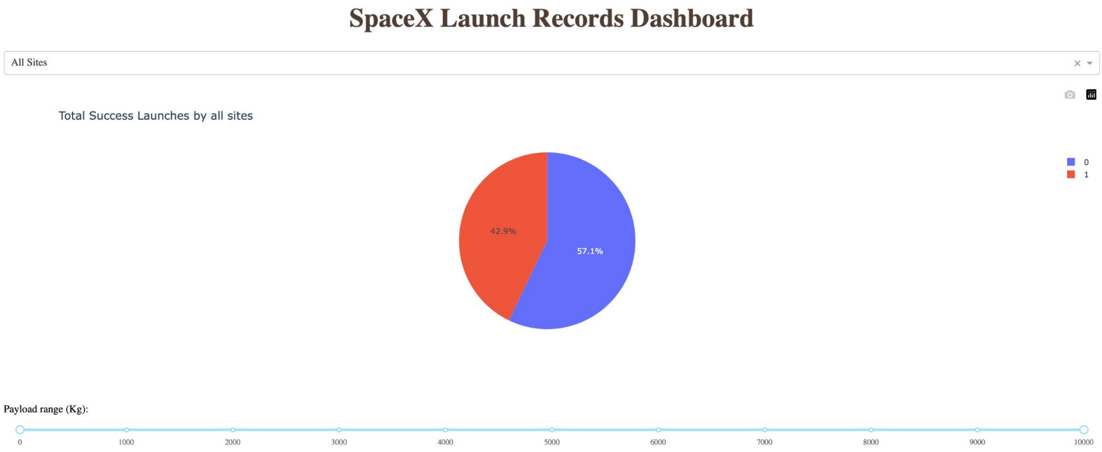
Launch Outcomes by Site



KSC LC-39A Good Logistics, near Railroad and transportation infrastructure and relatively far from inhabited areas.

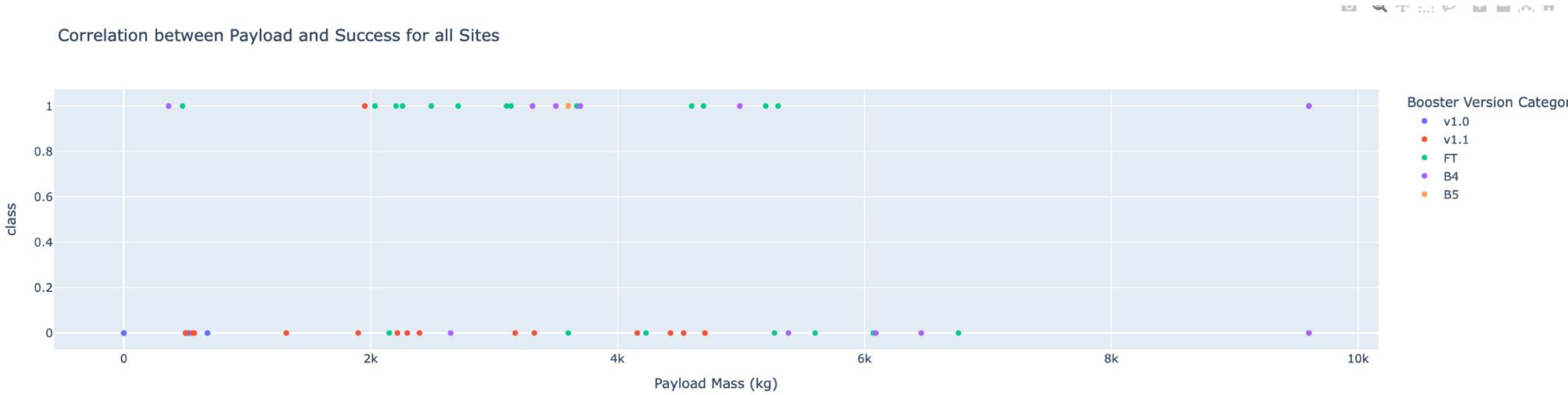
Dashboard with Plotly Dash

Dashboard with Plotly Dash



An interactive dashboard has been developed for stakeholder’s convenience

Dashboard with Plotly Dash



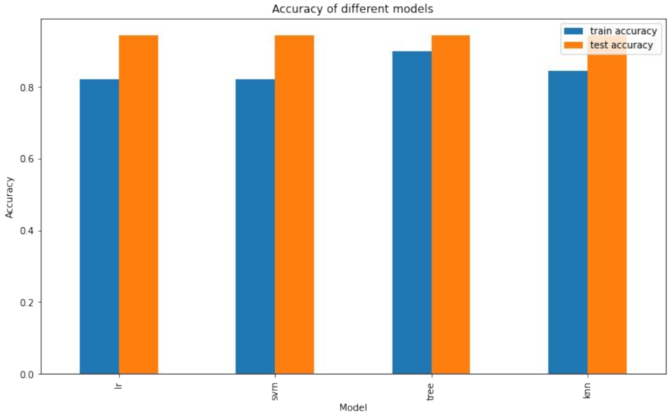
Correlation between Payload and Success for all sites is displayed above.

Predictive Analysis (Classification)

Classification Accuracy

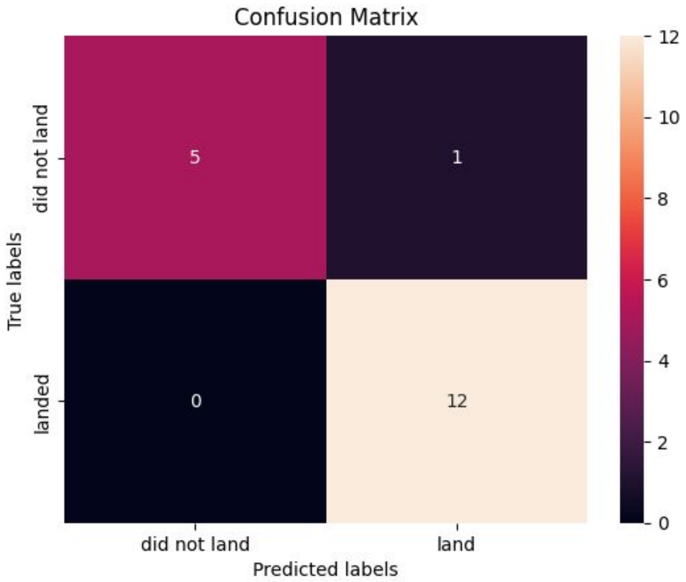
4 Classification models were tested and analysed

Decision Tree Classifier was the best model with over 87% accuracy.



Confusion Matrix (DTC)

The confusion matrix of Decision Tree Classifier **proves its accuracy** by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

Conclusions - TLDR;

Best Launch Site: KSC LC-39A;

Launches above 7,000 kg is less risky;

Successful landing outcomes improving over time

Decision Tree Classifier can be used to predict successful landings