IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

TAYYABA RIZ

25-08-2023

Table of Contents

Executive Summary

- The data analysis involved gathering data through webscraping and the SpaceX API, conducting Exploratory Data Analysis (EDA), encompassing tasks like data preprocessing,visualization, and interactive visual analytics and employing Machine Learning for predictive analysis.
- The outcomes of this analysis are the acquisition of meaningful data from publicly available sources was successfully achieved, the EDA facilitated the identification of prime predictors for launch success and Machine Learning Prediction unveiled the optimal model for forecasting launch outcome based on specificattributes.

Introduction

- The analysis of data encompassed various methodologies, including data collection through web scraping and the SpaceX API, followed by Exploratory Data Analysis (EDA) involving tasks such as data wrangling, visualization, and interactive visual analytics.
- Subsequently, Machine Learning Prediction was applied to the dataset. The results were summarized, highlighting the successful acquisition of valuable information from public sources, the EDA's role in identifying key predictive features for launch success, and the Machine Learning Prediction model's effectiveness in forecasting launch outcomes based on specific characteristics.

Section 1

Methodology

Methodology

• Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/ rockets) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_ Falcon_ Heavy_launches), using web scraping technics.

• First Data wrangling operations were performed on the dataset then exploratory data analysis was conducted using SQL and Data Visualization. Interactive Dashboard was created using Folium and predictive analysis was done using classification models.

Data Collection

Initiate an API request to retrieve SpaceX launch data and parse the obtained information. Subsequently, apply filtering to isolate entries related to Falcon 9 launches, and address any instances of missing data.

• Initiate a request for the Falcon 9 Launch Wiki page, extract the names of all columns or variables from the header of the HTML table, and generate a data frame by parsing the HTML tables related to the launches.

Data Wrangling

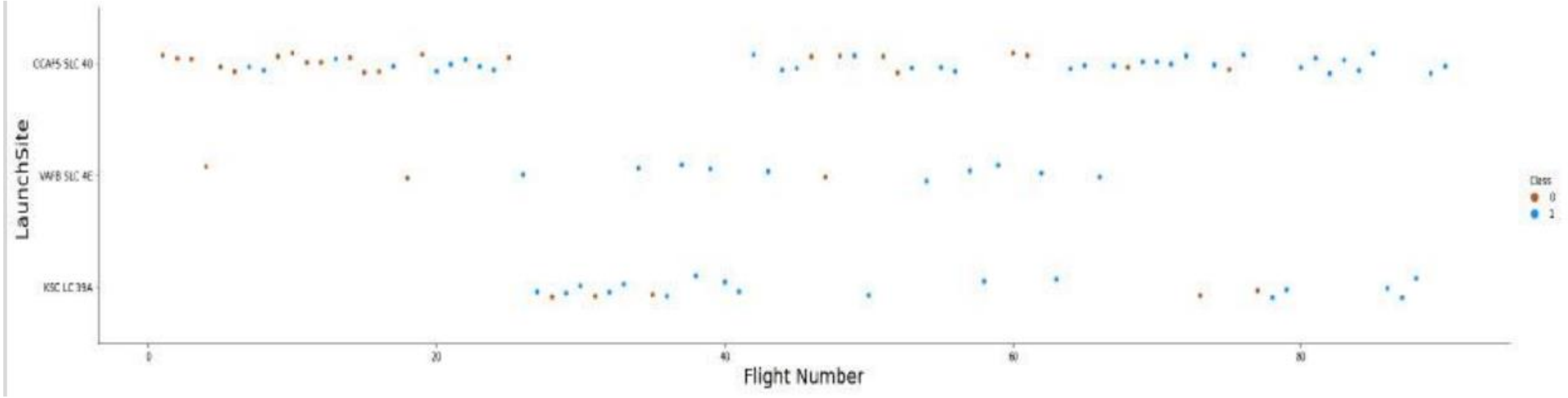
The initial phase involved conducting Exploratory Data Analysis (EDA) on the dataset.

• Subsequently, calculations were performed to determine launch counts per site, frequency of each orbit type, and occurrences of mission outcomes for each orbit category.

• Finally, the landing outcome label was derived from the data in the Outcome column.

Exploratory Data Analysis

For data exploration, scatterplots and barplots were employed to visually represent associations between pairs of features, including Payload Mass vs. Flight Number, Launch Site vs. Flight Number, Launch Site vs. Payload Mass, Orbit vs. Flight Number, and Payload vs. Orbit.



Exploratory Data Analysis with SQL

Executed SQL queries:

- Extracting names of distinct launch sites
in space missions.
- Identifying the top 5 launch sites with names starting with 'CCA'.
- Calculating the total payload mass transported by NASA (CRS) boosters.
- Determining the average payload mass conveyed by booster version F9 v1.1.
- Finding the date of the initial successful
landing outcome on a ground pad.
- Listing the names of boosters that achieved success on drone ships while carrying payload masses between 4000 and 6000 kg.
- Computing the overall count of successful and failed mission outcomes.
- Identifying booster versions that carried the highest payload masses.
- Enumerating failed landing outcomes on drone ships in 2015, along with corresponding booster versions and launch site names.
- Ranking the count of landing outcomes (e.g., Failure (drone ship) or Success (ground pad)) between June 4, 2010, and March 20, 2017.

Interactive Map with Folium

- Folium Maps were enhanced through the incorporation of markers, circles, lines, and marker clusters, each serving distinct purposes.

- Markers were used to denote significant points like launch sites, circles highlighted specific areas around particular coordinates, such as the NASA Johnson Space Center, marker clusters provided a visual grouping of events at individual coordinates like launches within a launch site, and lines were employed to illustrate distances between pairs of coordinates.

Dashboard with Plotly

- Various graphs and plots were employed to visualize the data effectively, including the percentage distribution of launches by site and the payload range.

- This combination of visualizations facilitated a swift analysis of the connection between payloads and launch sites, aiding in the identification of optimal launch locations based on payload considerations.

Predictive Analysis

Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.

- Each model was iterated through steps that included preparation and standardization of data, evaluation of each model using various hyperparameter combinations, and subsequent comparison of outcomes.

Results and Insights

Exploratory Analysis Result

Exploratory data analysis results:

• Space X uses 4 different launch sites;

• The first launches were done to Space X itself and NASA;

• The average payload of F9 v1.1 booster is 2,928 kg;

• The first success landing outcome happened in 2015 fiver year after the first launch;

• Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;

• Almost 100% of mission outcomes were successful;

• Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;

• The number of landing outcomes became as better as years passed.

Dashboard with Plotly

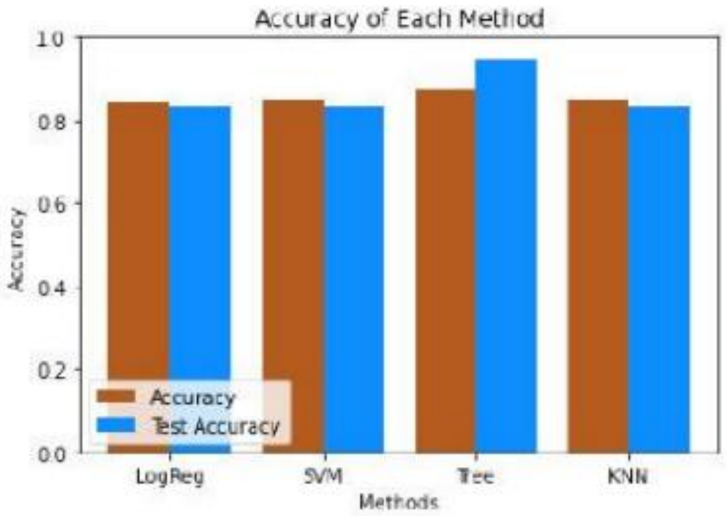
- Various graphs and plots were employed to visualize the data effectively, including the percentage distribution of launches by site and the payload range.

- This combination of visualizations facilitated a swift analysis of the connection between payloads and launch sites, aiding in the identification of optimal launch locations based on payload considerations.

Interactive and Predictive Analysis Results

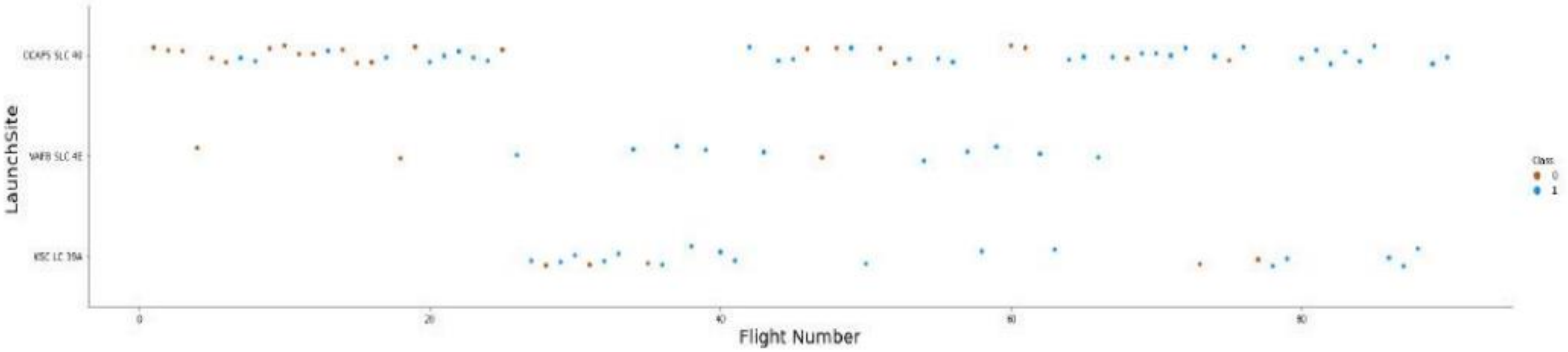
- Interactive analytics revealed that launch sites are typically located in secure areas, often near bodies of water such as the sea, and are strategically positioned with strong logistical infrastructure in their vicinity. Additionally, a predominant majority of launches take place at launch sites situated along the east coast.

- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.

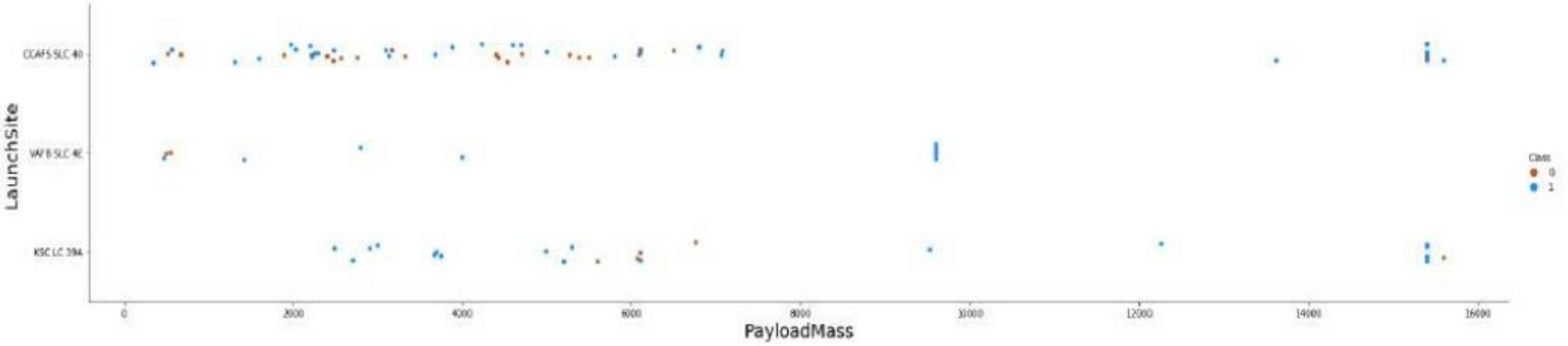


Insights

CCAF5 SLC 40 is the best launch site

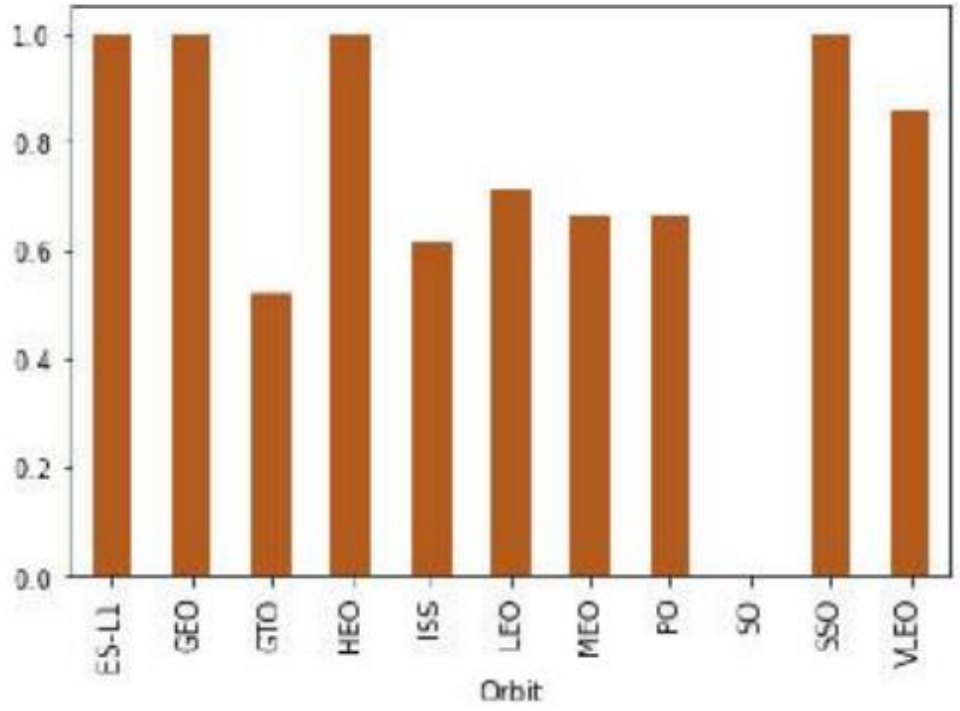


Payloads exceeding 12,000 kg appear to be feasible exclusively at the CCAFS SLC 40 and KSC LC 39A launch sites.

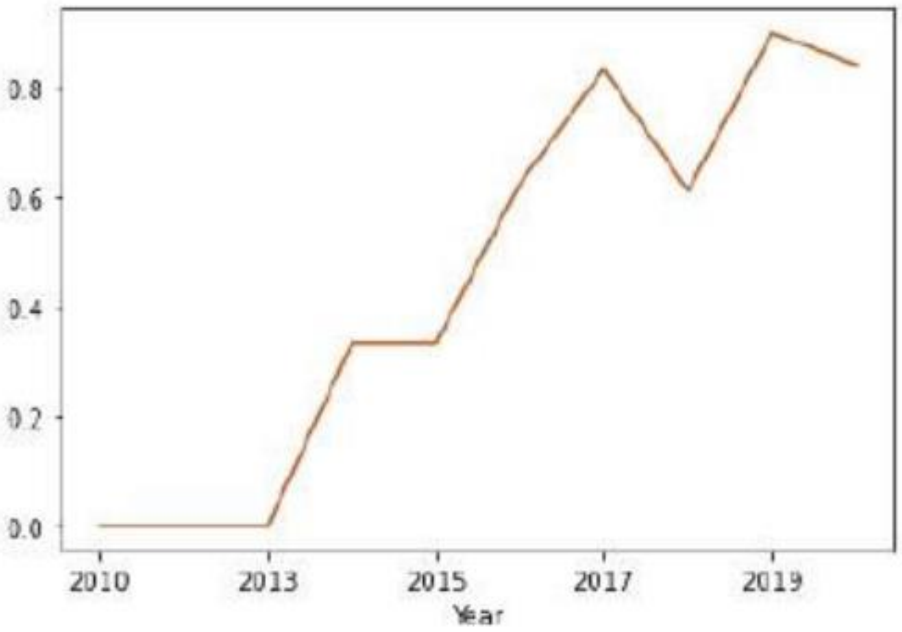


Insights

ES L1, GEO, HEO, SSO have highest success rate



It seems that the first three years were periods of adjustments

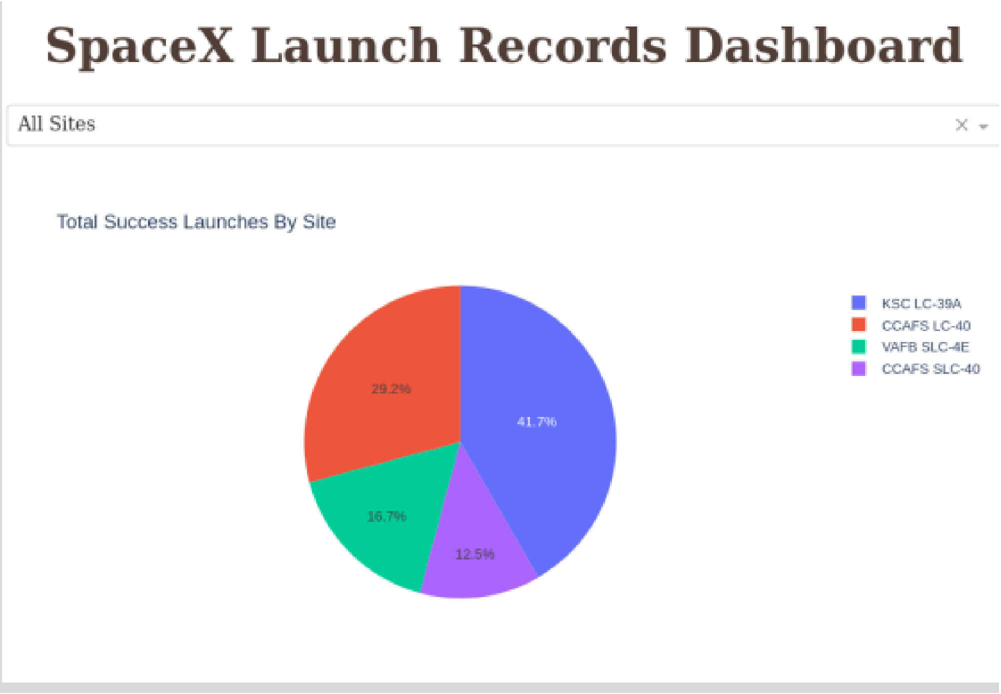


Insights

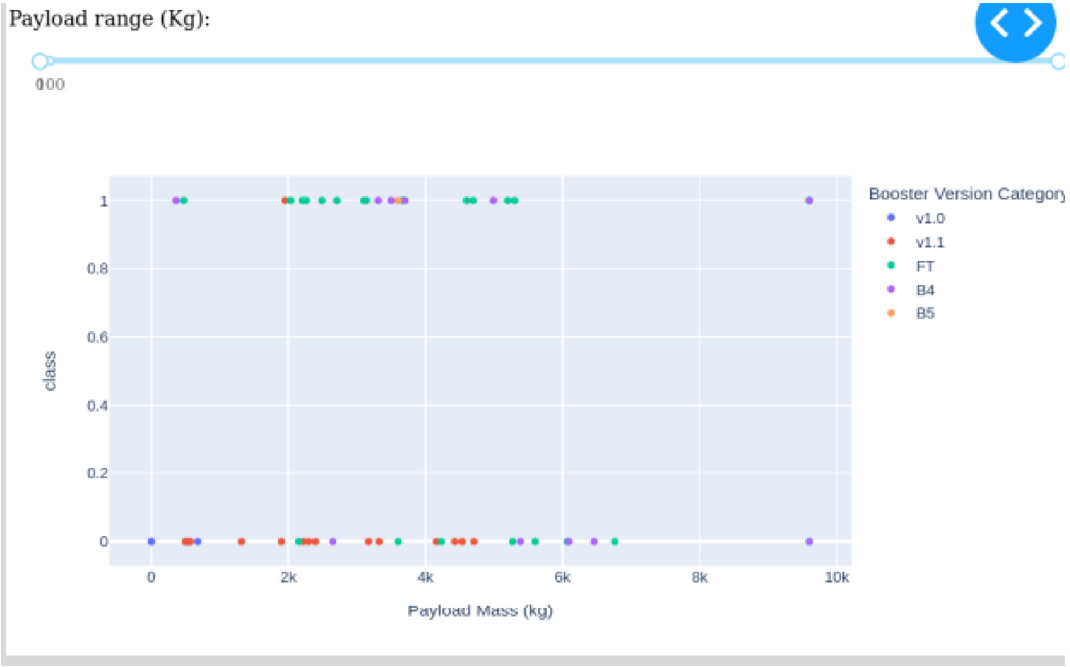
During the analysis it was determined:

- The total pay load mass is 111.268 kg.
- The average pay load mass is 2.928 kg.
- The first landing outcome on ground pad is dated to be on12/22/2015.
- Boosters that have achieved successful landings on drone ships and carried payload masses exceeding 4000 but below 6000 are F9 FT B1021.2, F9 FT B1031.2, F9 FT B1022 and F9 FT B1026.
- There were 99 successes mission outcomes and 1 failure mission outcome.irst landing outcome on ground pad is dated to be on12/22/2015.

Dashboards

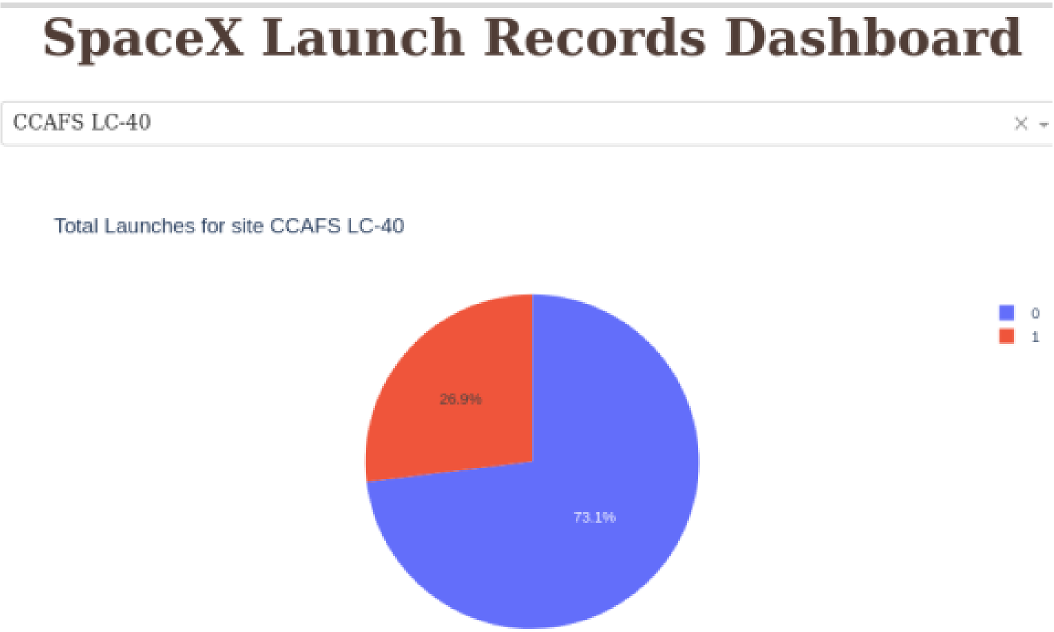


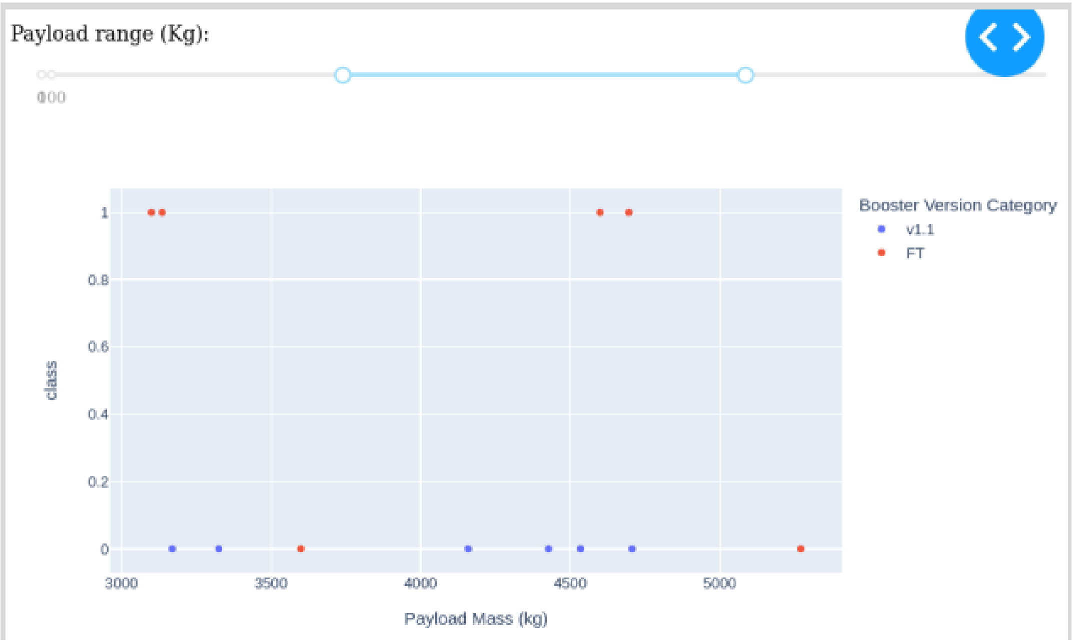
SpaceX Launch Records Dashboard
All Sites
Total Success Launches By Site
29.2%
41.7%
16.7%
12.5%
KSC LC-39A
CCAFS LC-40
VAFB SLC-4E
CCAFS SLC-40



Payload range (Kg):
class
Payload Mass (kg)
0
2k
4k
6k
8k
10k
0
0.2
0.4
0.6
0.8
1
Booster Version Category
v1.0
v1.1
FT
B4
B5

Dashboards

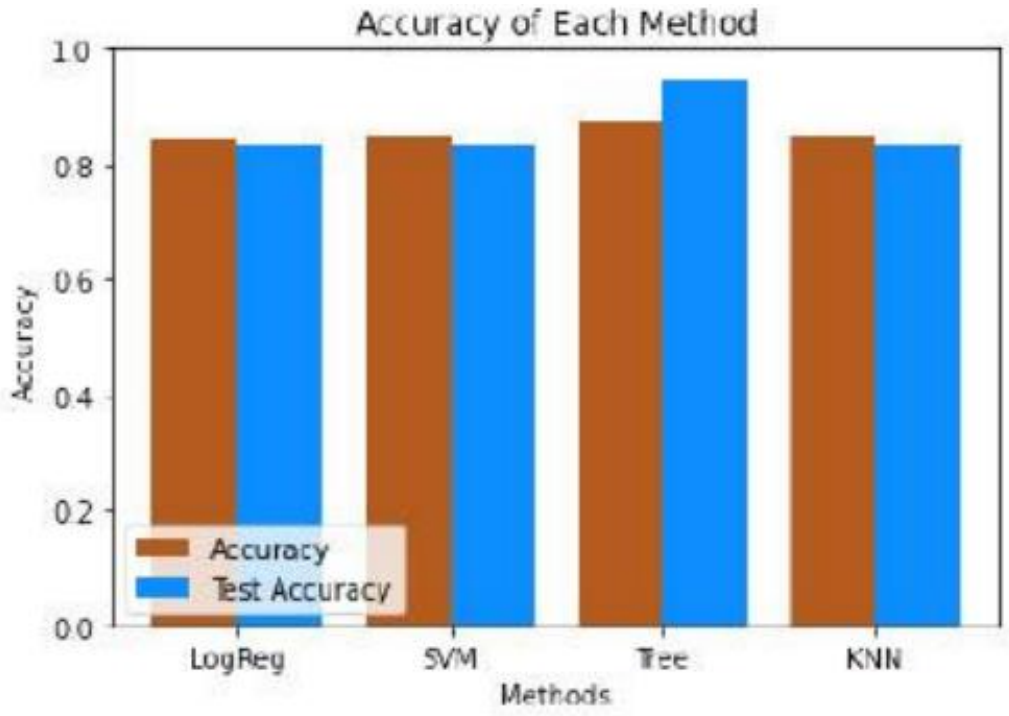

SpaceX Launch Records Dashboard
CCAFS LC-40
Total Launches for site CCAFS LC-40
0
1
26.9%
73.1%


Payload range (Kg):
Booster Version Category
v1.1
FT
class
1
0.8
0.6
0.4
0.2
0
3000
3500
4000
4500
5000
Payload Mass (kg)

Predictive Analysis

The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.

Confusion matrix demonstrates notable figures for true positives and true negatives in contrast to the occurrences of false outcomes.



Conclusions

The analysis involved the examination of various data sources,

leading to progressively refined conclusions:

- The optimal launch site is KSC LC 39A.
- Launches with payloads exceeding 7,000 kg exhibit reduced risk.
- While a majority of mission outcomes are successful, successful landing results appear to enhance gradually over time in alignment with process and rocket advancements.
- The Decision Tree Classifier is a valuable tool for forecasting successful landings, potentially contributing to increased profits.

Appendix

Check the source files using this link:

https://github.com/tabzyriaz/Coursera.git

Thank you!