DATA SCIENCE FINAL CAPSTONE PROJECT

NAME-SAFDER SHAKIL

DATE-29TH OF OCTOBER,2022

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- The data were collected and different methodologies were used to derive meaningful conclusions.
- The following methodologies were used to analyze data :
  - Data Collection using web scraping and SpaceX API.
  - Exploratory Data Analysis (EDA), including data wrangling,data visualization and interactive visual analytics.
  - Machine Learning Prediction.
- Summary of results

  It was possible to collect data from public sources.

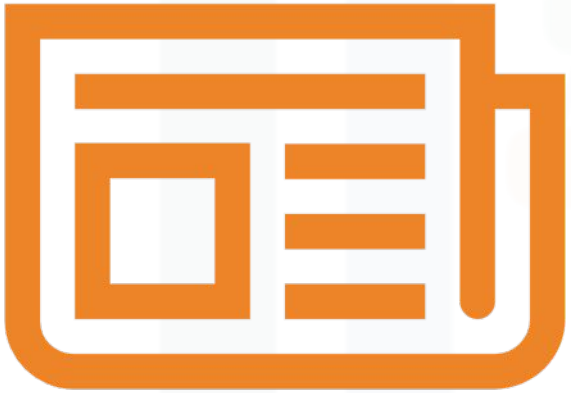
  EDA allowed to identify which features are the best to predict success of launchings.

  Machine Learning Predictions showed the best model to predict which characteristics are important to drive this opportunity by the best way ,using all collected data.

INTRODUCTION

- The objective is to evaluate the viability of the new company Space Y to compete with Space X.
- Desirable answers:
  - The best way to estimate the total cost for launches, by predicting successful landing of the first stage of rockets.
  - Where is the best place to make launches.

METHODOLOGY



- Data collection methodology:
  Data from space X was obtained from 2 sources . (1) Space X API (2) Web scraping from wikipeida .
- Perform Data Wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and plotly dash
- Perfrom Predictive analysis using classification models.

RESULTS

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.

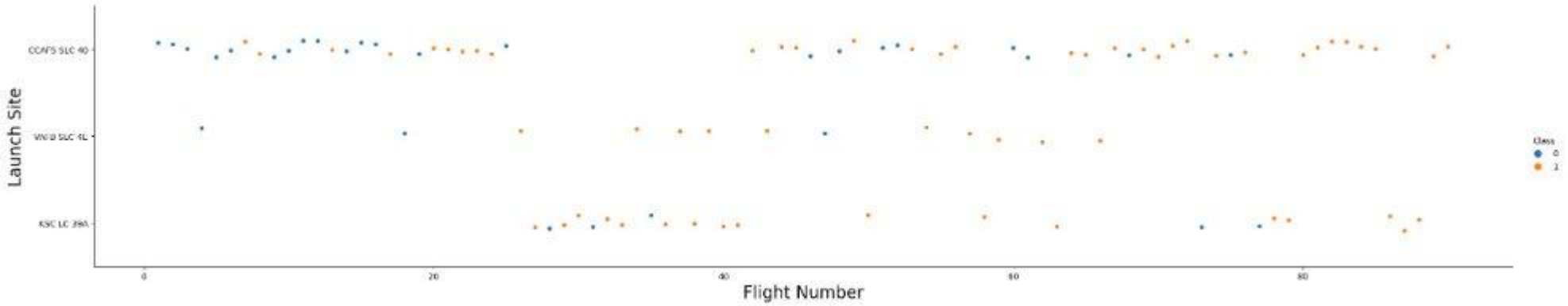
Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around

- Most launches happens at east cost launch site.

- 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 Drawn from EDA

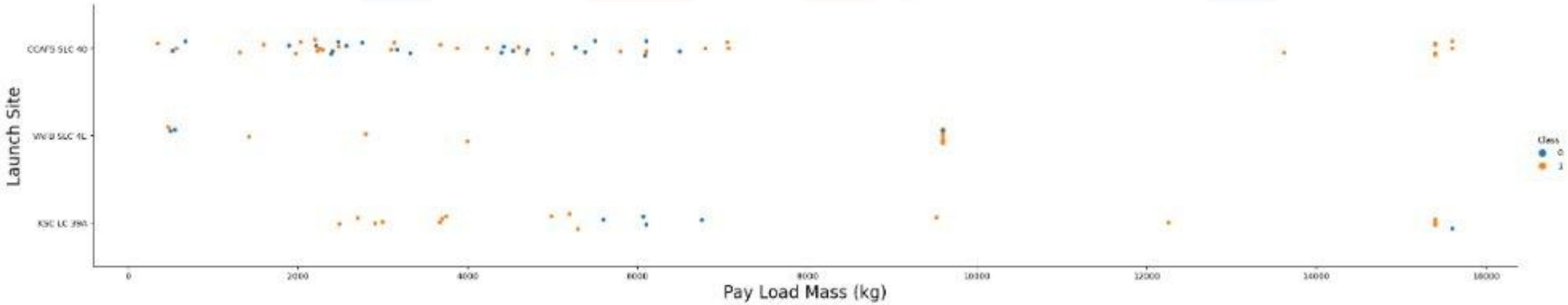


According to the plot above, it’s possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful.

In second place VAFB SLC 4E and third place KSC LC 39A.

It’s also possible to see that the general success rate improved over time.

Payload VS launch Site



Payloads over 9,000kg (about the weight of a medium size truck) have excellent success rate.

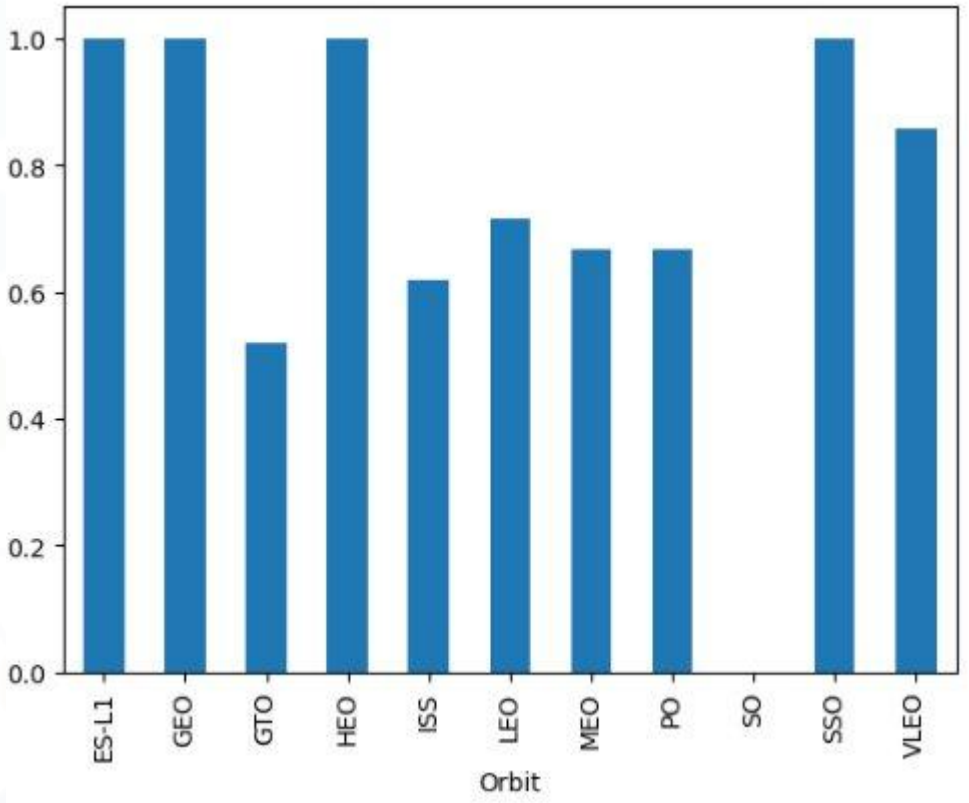
Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success rate VS Orbit type

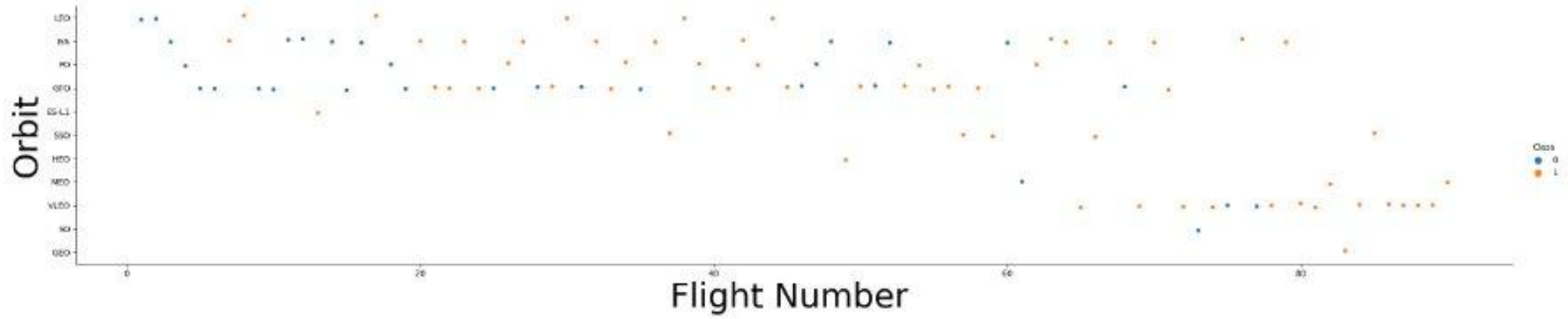
- The biggest success rates happens to orbits:
  - ES-L1;
  - GEO;
  - HEO; and
  - SSO.

Followed by:

- VLEO (above 80%); and
- LFO (above 70%)



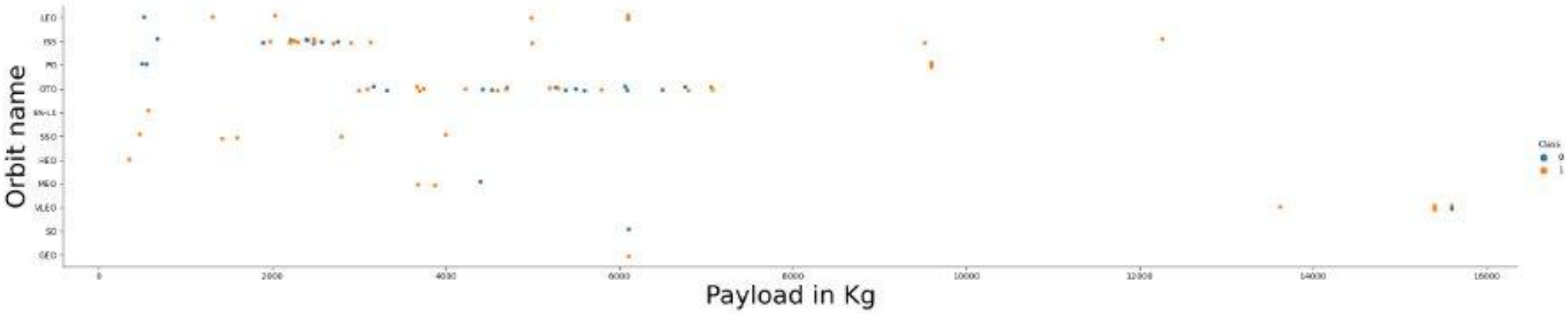
Flight Number VS Orbit Type



1. VELO orbit due to increase in its frequency seems like a good business opportunity.

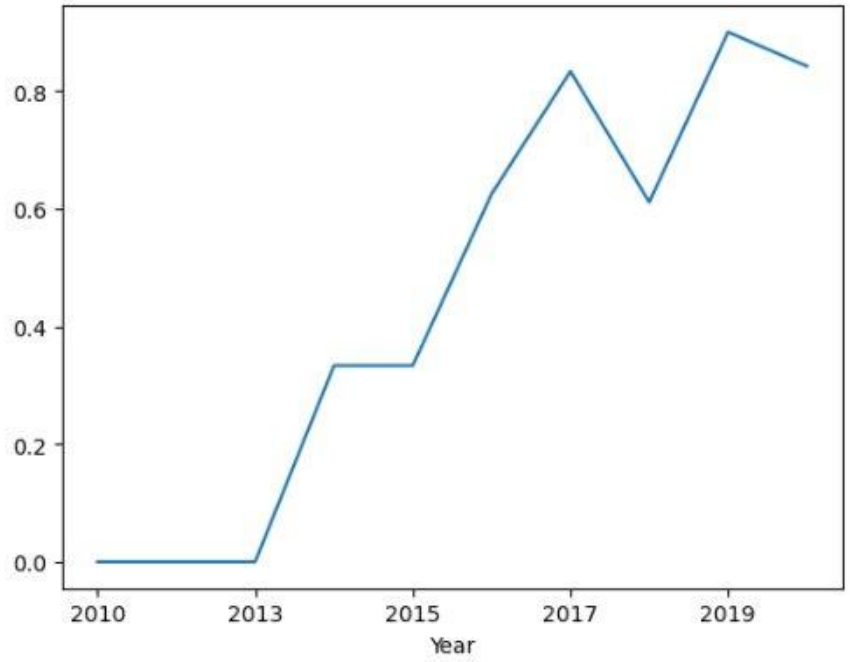
2. Success rate of all the orbits have improved over time.

Payload VS Orbit Type



- Apparently, there is no relation between payload and success rate to orbit GTO;
- ISS orbit has the widest range of payload and a good rate of success;
- There are few launches to the orbits SO and GEO.

Yearly Success Launch Trends



- The success increased exponentially after 2015.
- The first three years were stagnant.

All launch sites names

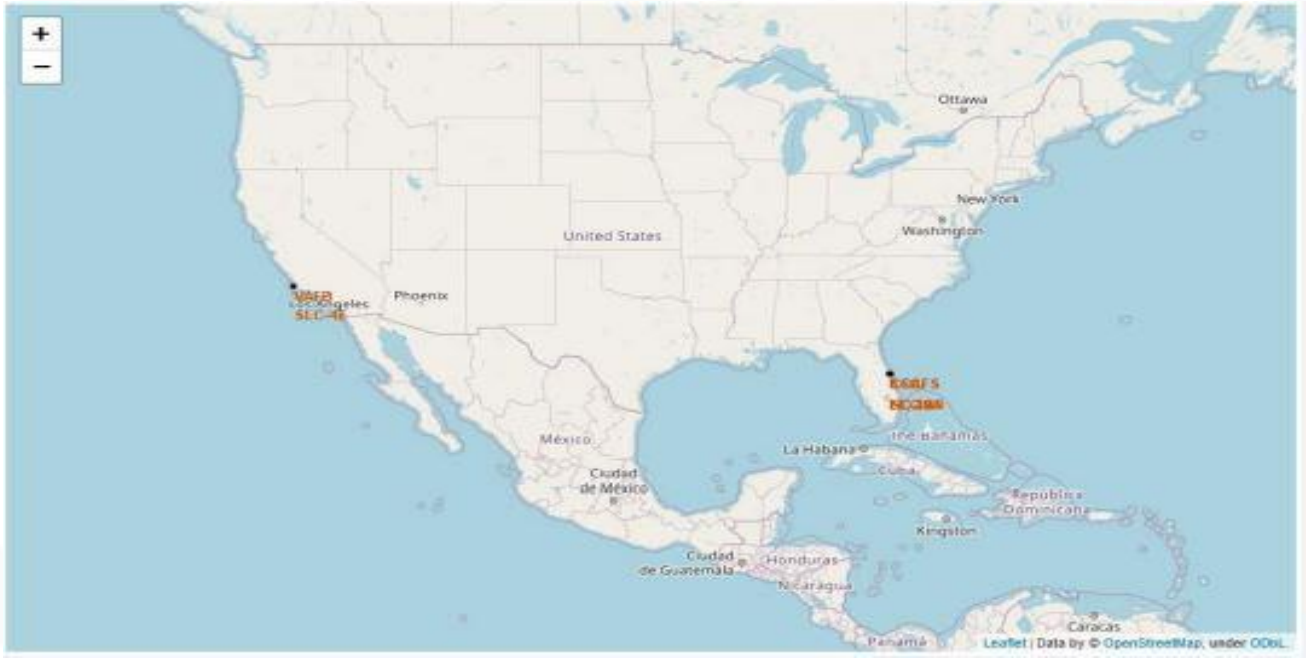
	Launch Site	Lat	Long
0	CCAFS LC-40	28.562302	-80.577356
1	CCAFS SLC-40	28.563197	-80.576820
2	KSC LC-39A	28.573255	-80.646895
3	VAFB SLC-4E	34.632834	-120.610746

There are four launch sites.

They are selected by unique occurrence of launch site values from data set.

Lauch Site approximate analysis

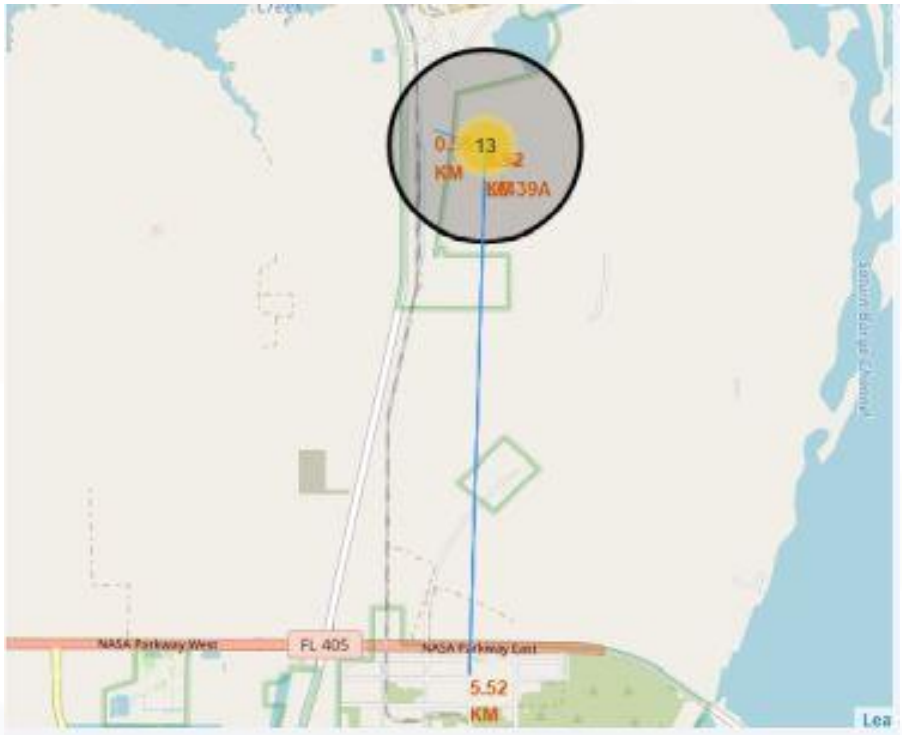
ALL launch sites- launch sites are near sea because of safety measures.



Launch Outcome by sites

Green ones are the successful ones and red ones are failed ones.

Logistics and safety



Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas

EDA WITH SQL

- The following SQL queries were performed: • Names of the unique launch sites in the space mission;
- Top 5 launch sites whose name begin with the string 'CCA';
- Total payload mass carried by boosters launched by NASA (CRS);
- Average payload mass carried by booster version F9 v1.1;
- Date when the first successful landing outcome in ground pad was achieved;
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and
- Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

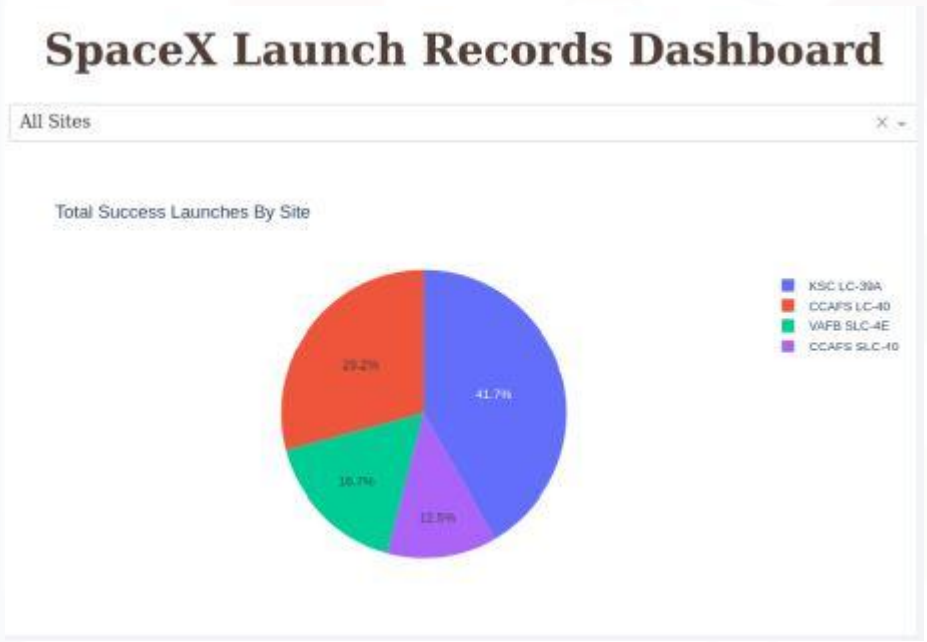
github link-
https://github.com/safder777/Capstone-IBM-Data-Science/blob/2ef9722c30cf67f38a10784730dfffad69f2991a/jupyter-labs-eda-sql-coursera_sqllite.ipynb

DASHBOARD

Github link for Dasboard :

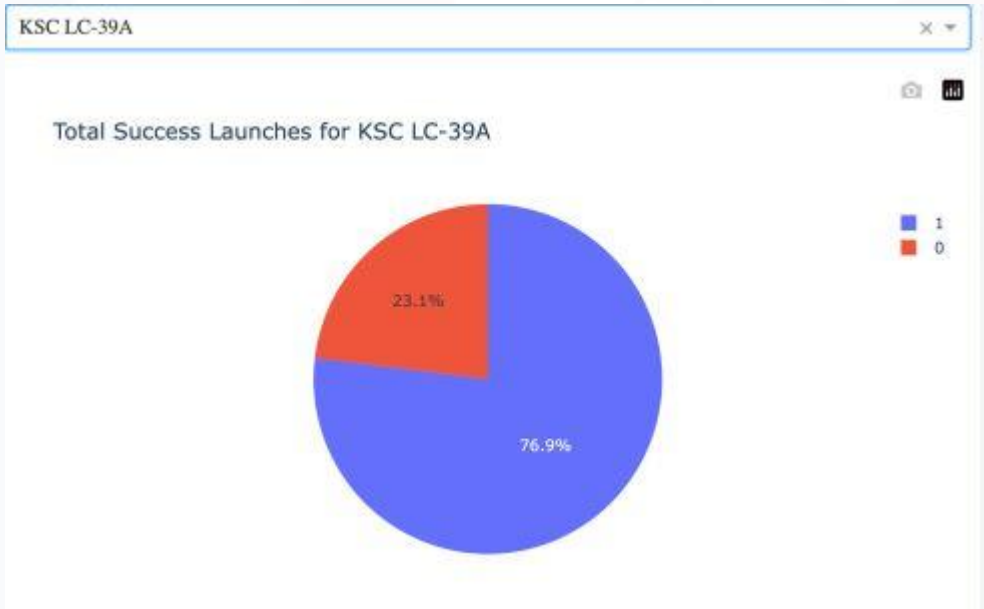
https://github.com/safder777/Capstone-IBM-Data-Science/blob/1fe9c9599c1d890068757b68b75ff49f36647aa8/spacex_dash_app.py

DASHBOARD TAB 1|Total Success Launches for All Sites



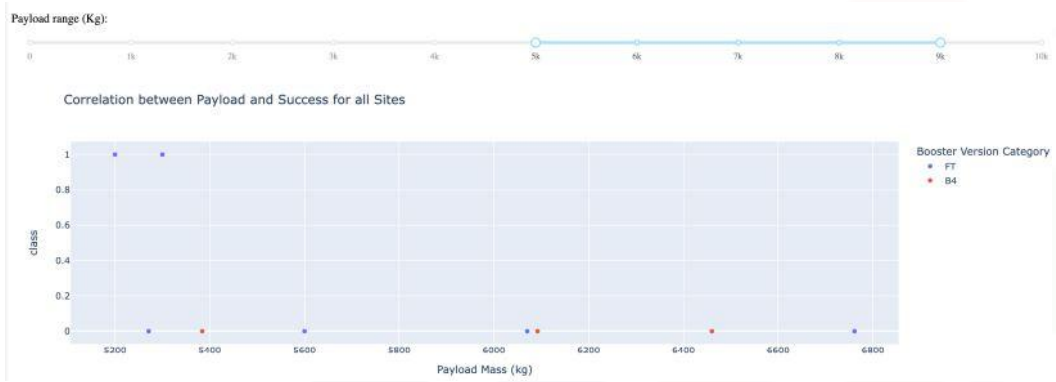
The launch site plays an important role in success of the missions.

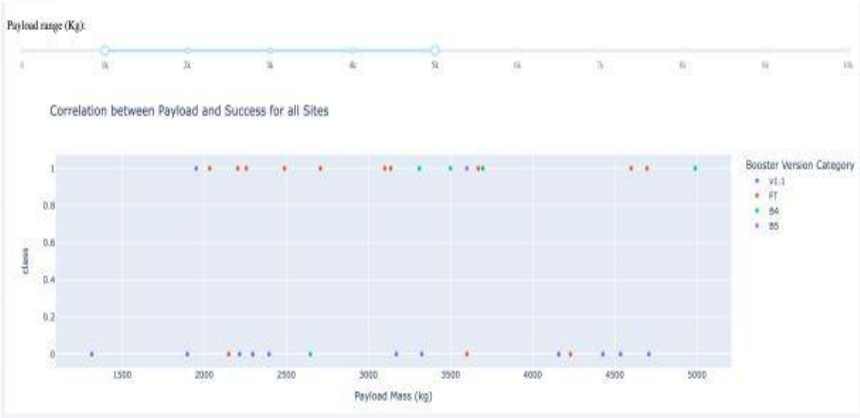
DASHBOARD TAB 2 | Success Ratio for KSC LC-39A

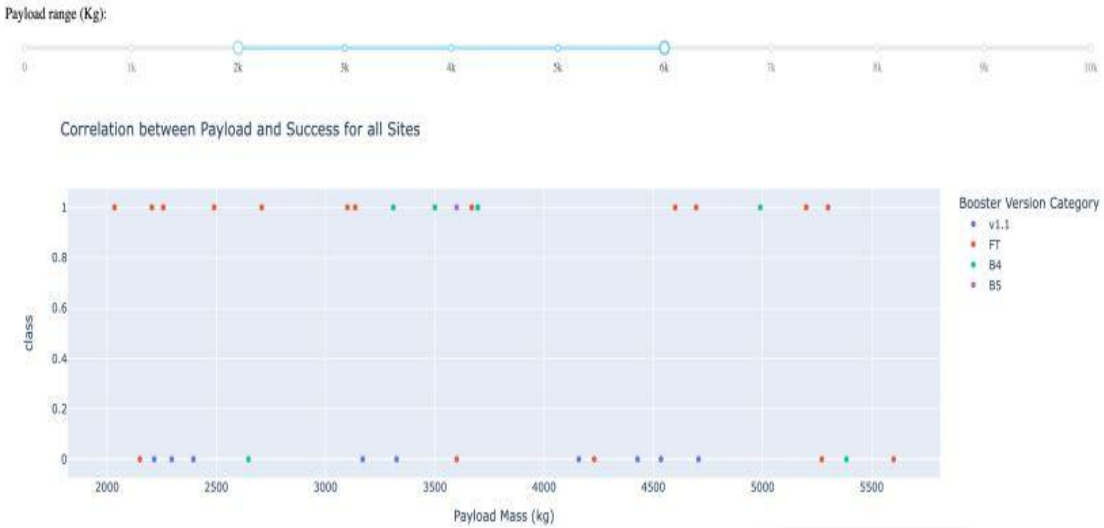


The launch site with highest launch success ratio is KSC LC-39A. It has a success rate of 76.9%

DASHBOARD TAB 3 | Correlation between Payload and success





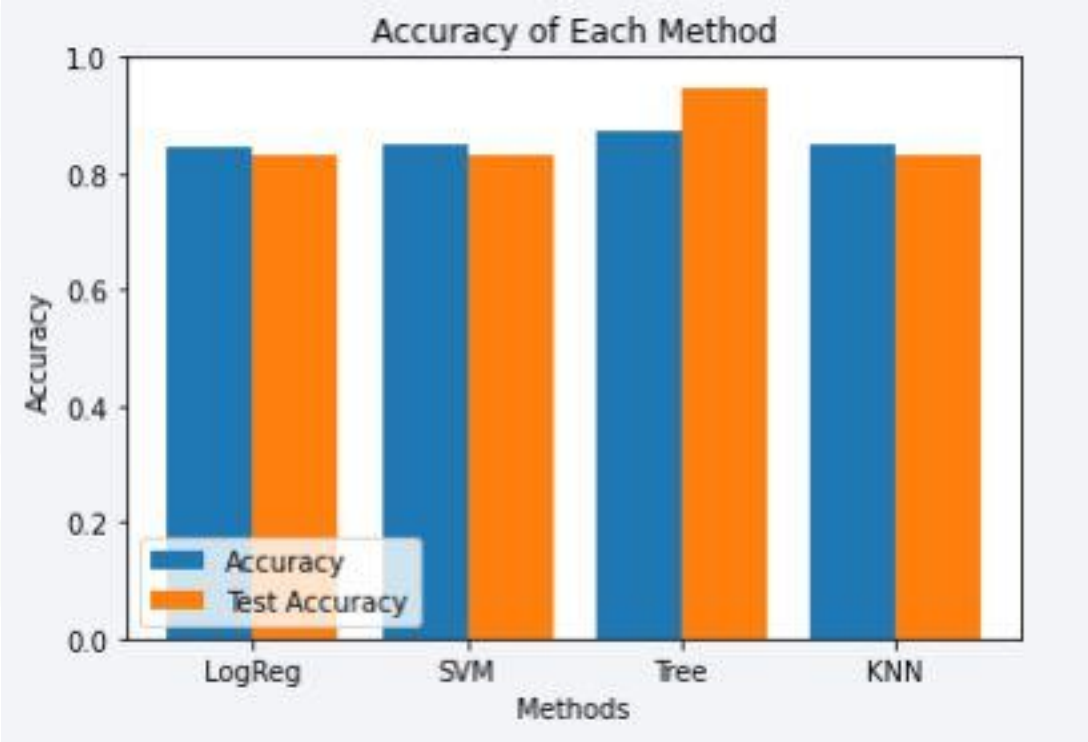


- Booster version of FTT has the largest success rate.
- Pyload version between 3000-4000 has the largest success rate.

Predictive Analysis (Classification)

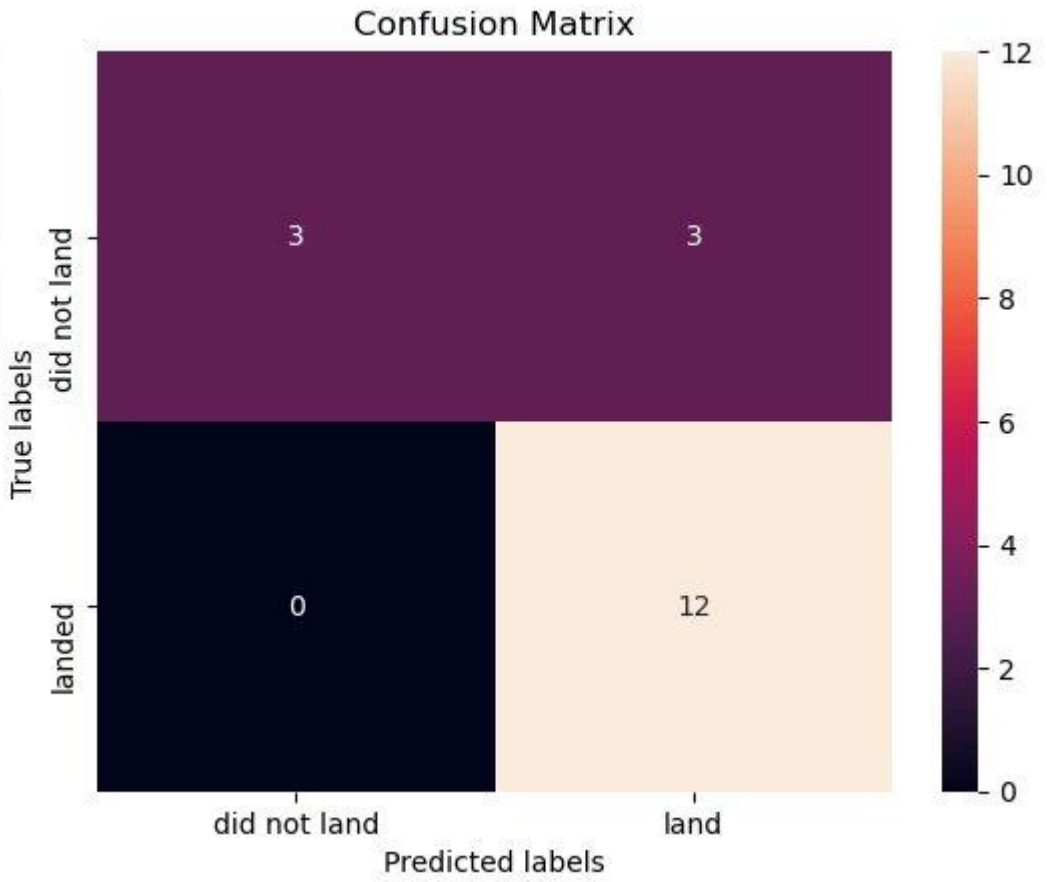
- Four classification models were tested, and their accuracies are plotted beside;

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



Confusion Matrix of Decision Tree Classifier

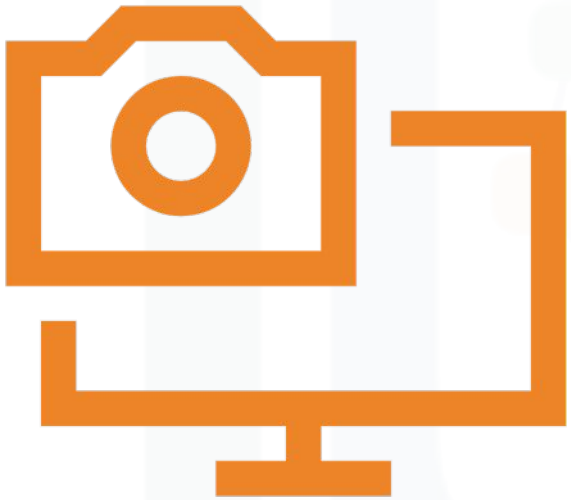
The confusion matrix shows that the number of true positives is much more than the false positives and negatives.



CONCLUSION

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

APPENDIX



- The github of link of the files is given below.

https://github.com/safder777/Capstone-IBM-Data-Science