SpaceX Falcon9 Landing Rate Prediction with Data Science

ViswaShanthi Bonala

07-JULY-2023

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- SpaceX advertises Falcon 9 rocket launches on its website which reduces the cost as it can reuse the first stage.
- Here, we determine if the first stage will land successfully or not by
  - SpaceX launch data collection from SpaceX REST API.
  - Data Wrangling
  - Exploratory Data Analysis using SQL and Visualization
  - Interactive Visual Analytics and Dashboard
  - Predictive Analysis using
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree Classifiers
    - K-Nearest Neighbors

INTRODUCTION

- SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars.
- To reduce the cost SpaceX wants to reuse the first stage.
- Here, we want to determine if the first stage will land successfully and the best place to make the Launch.
- This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

METHODOLOGY

- **Data Collection**:

  Data for SpaceX was collected from

  - Booster Names: https://api.spacexdata.com/v4/rockets/
  - The mass of the payload and the orbit: https://api.spacexdata.com/v4/launchpads/
  - The launch site being used, the longitude, and the latitude: https://api.spacexdata.com/v4/payloads/
  - The outcome of the landing, the type of the landing, number of flights with that core: https://api.spacexdata.com/v4/cores/
  - Rocket Launch Data: https://api.spacexdata.com/v4/launches/past

METHODOLOGY

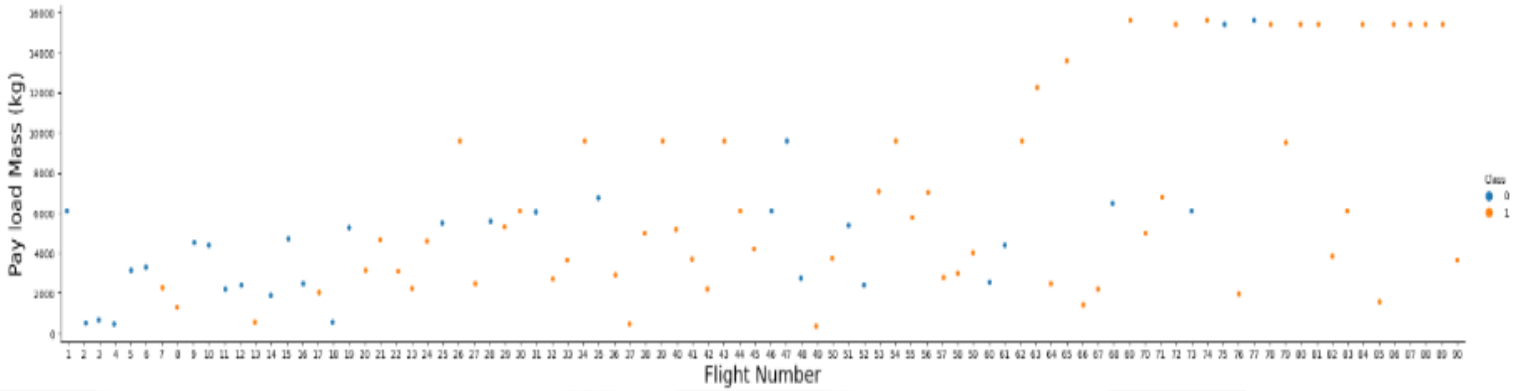
- **Data Wrangling**:
  - Handle the missing or null data
  - Create Landing Outcome column called Class where 0 means bad landing and 1 means successful landing.
- **Exploratory Data Analysis using SQL**
  - Understand the Spacex DataSet
  - Load the dataset into the corresponding table in a Db2 database
  - Check the data patterns using SQL Queries.

METHODOLOGY

- **Exploratory Data Analysis using Visualization**

Here we used, Scatter and bar plots to visualize the relationship between pair of features.

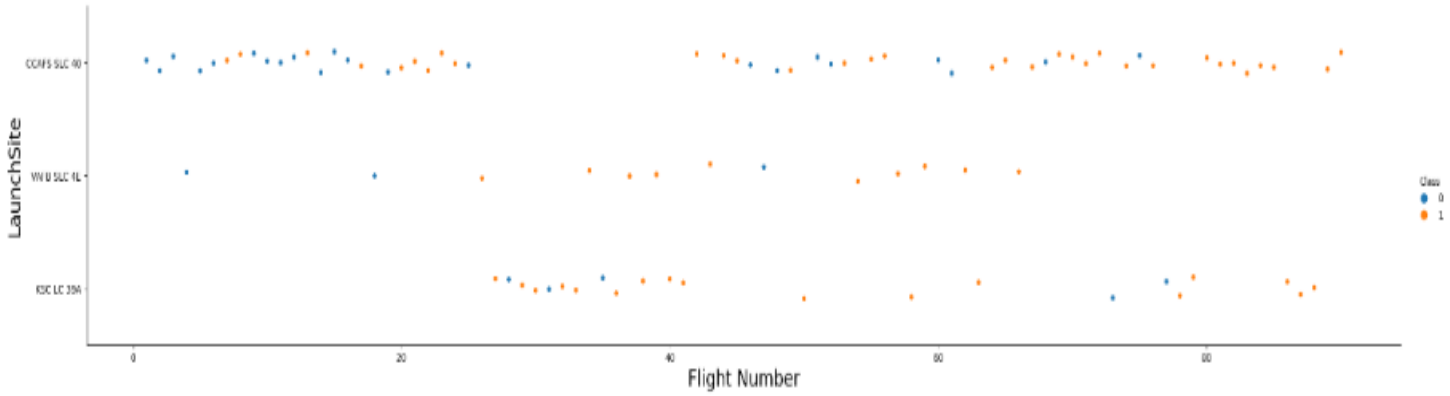
Relationship between Flight number and PayLoadMass



We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

METHODOLOGY

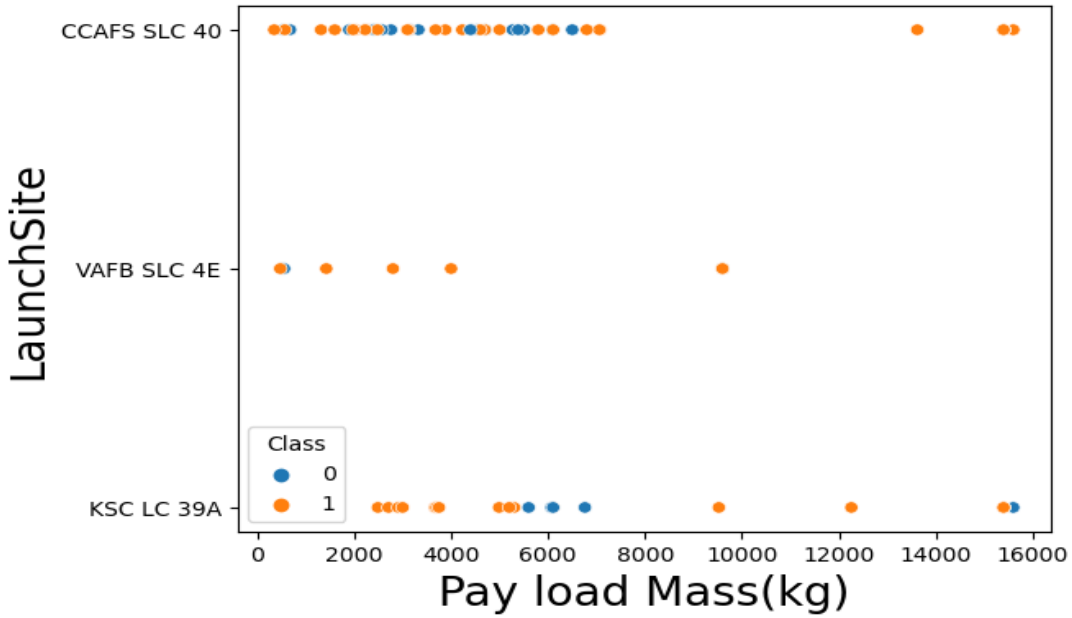
Relationship between Flight number and LaunchSite.



We see that as the flight number increases, the first stage is more likely to land successfully. It seems CCAFS LC-40 has more bad launches compare to KSC LC-39A and VAFB SLC 4E.

METHODOLOGY

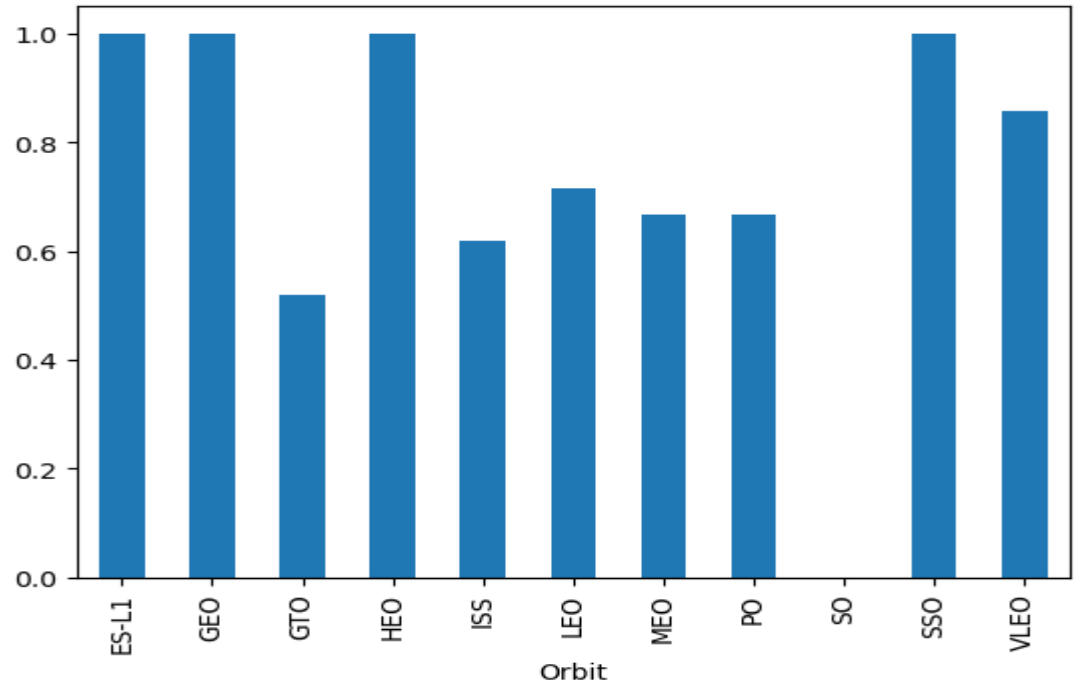
Relationship between PayloadMass and LaunchSite.



The VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).

METHODOLOGY

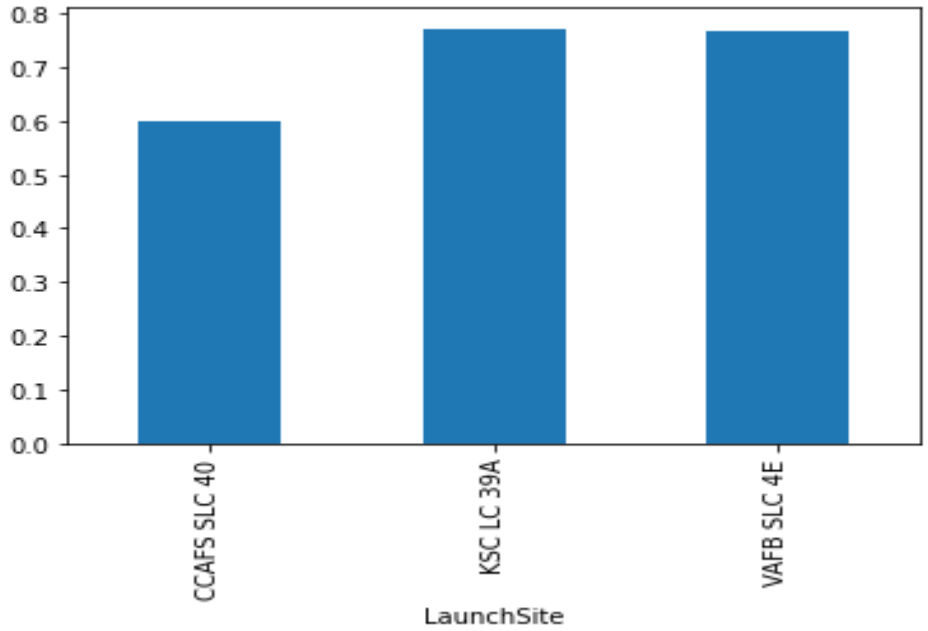
success rate of each orbit type.



ES-L1,GEO,HEO and SSO has highest success rate. if we observe closely, we could see SO does not have any success rate.

METHODOLOGY

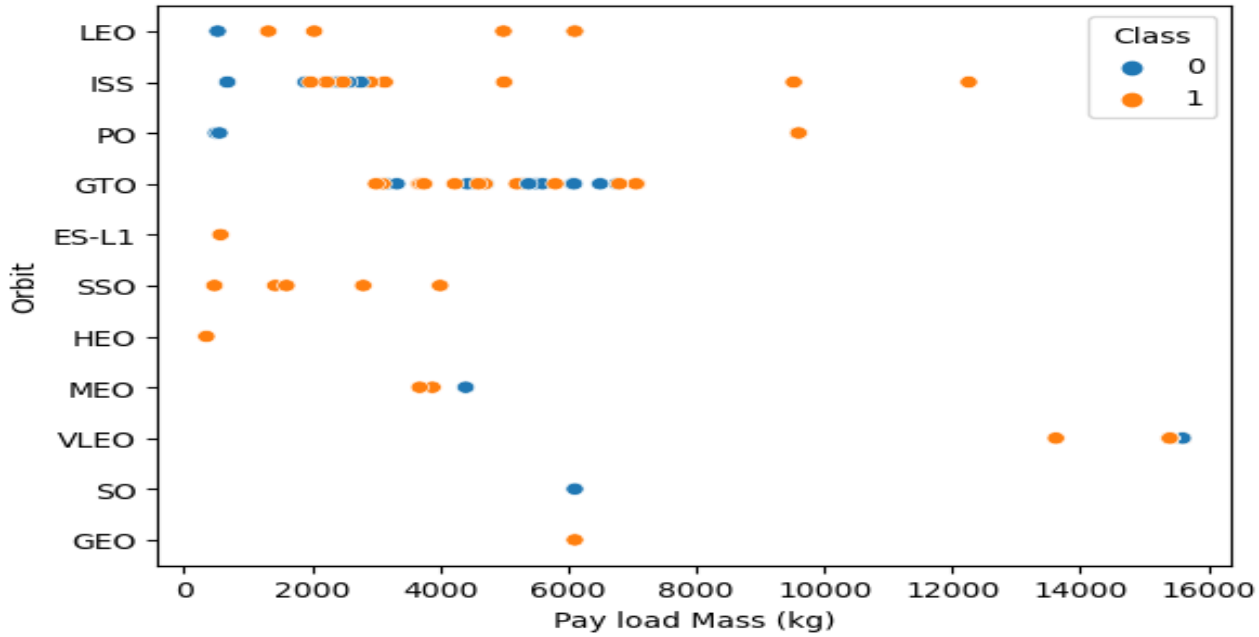
success rate of each Launching site.



KSC LC 39A has highest success rate.

METHODOLOGY

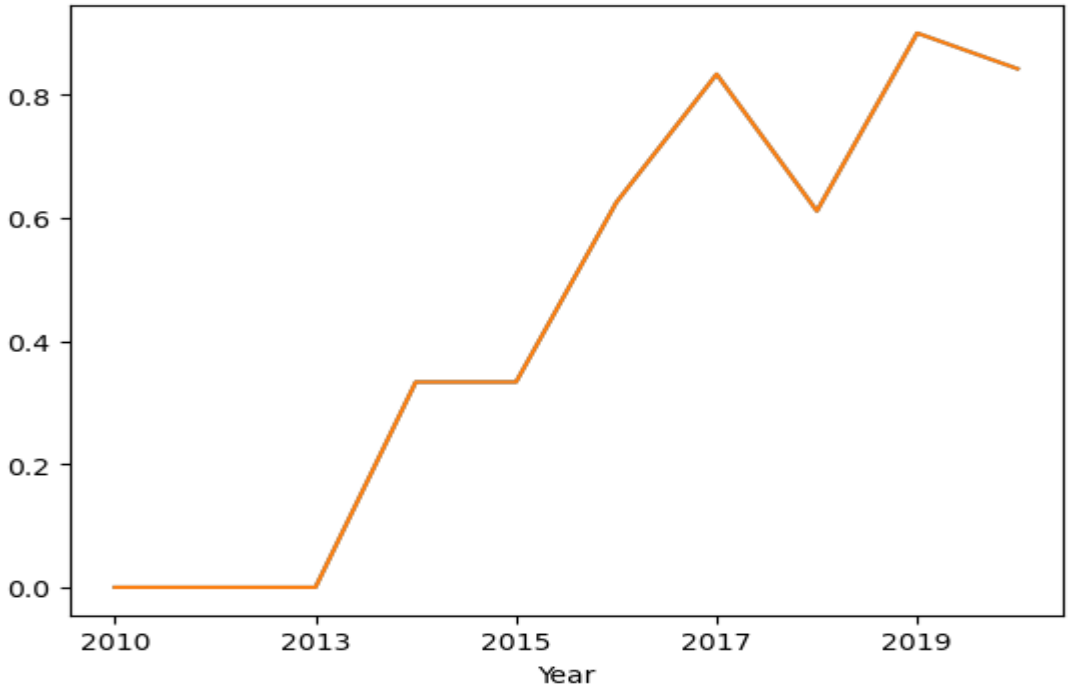
Relationship between Payload and Orbit type



With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS. However, for GTO we cannot distinguish this well as both positive landing rate and negative landing.

METHODOLOGY

Launch success yearly trend

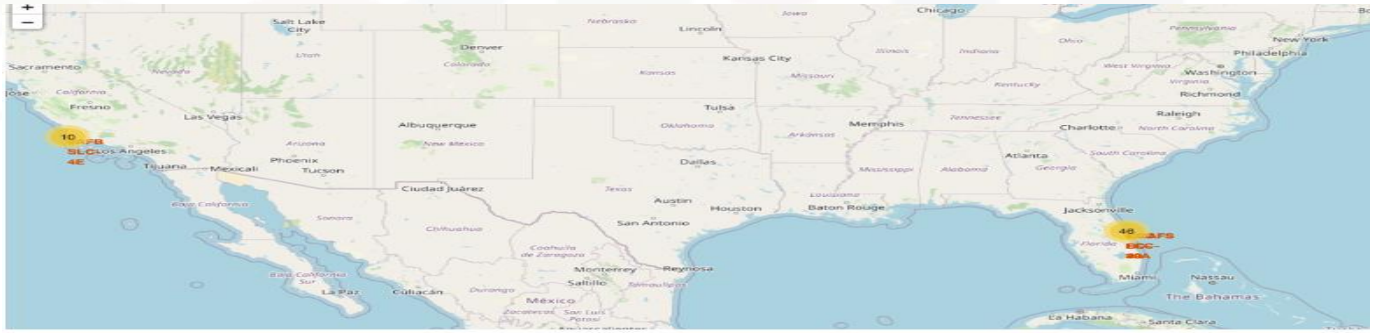


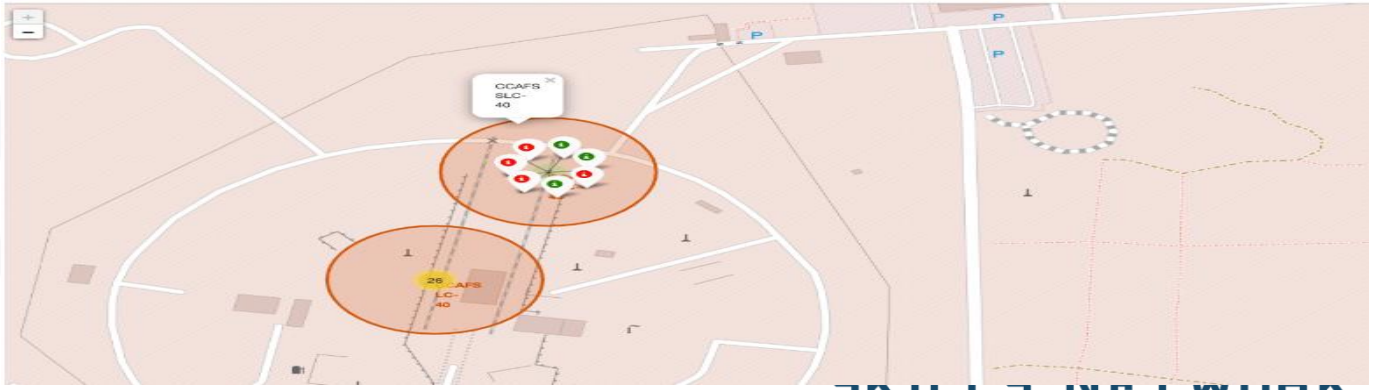
Observation is that the success rate since 2013 kept increasing till 2020.

METHODOLOGY

- **Interactive Visual Analytics and Dashboard**

In the Interactive Visualization, we have marked all the launching stations and success and failure launches of each site.





IBM Developer

SKILLS NETWORK

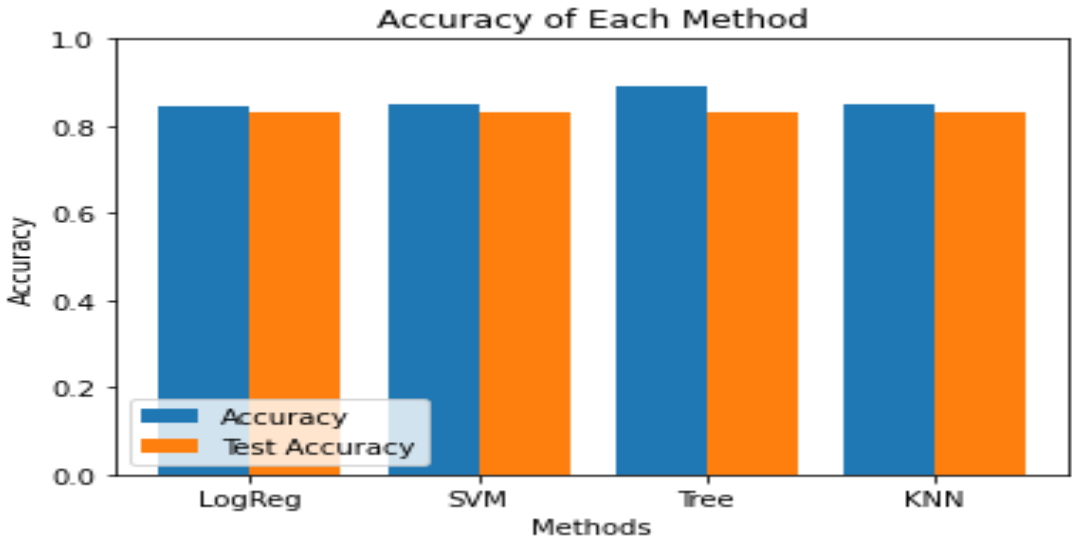
METHODOLOGY

- **Predictive Analysis**

  Predictive Analysis using

  - Logistic Regression
  - Support Vector Machine
  - Decision Tree Classifiers
  - K-Nearest Neighbors

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.88929	0.83333
KNN	0.84821	0.83333



RESULTS

- From Predictive Analysis, we could see that all the test scores are same, but Decision Tree accuracy is better that rest of the models.
- ES-L1,GEO,HEO and SSO Orbits have highest success rate.
- KSC LC 39A Launching Site has highest success rate
- The success rate since 2013 kept increasing till 2020.

CONCLUSION

Based on the data from different sources on SpaceX Launch data, we can conclude that

- The best Launch site is KSC LC 39A.
- Launches above 7000 kgs are less risky.
- As the FlightNumber (indicating the continuous launch attempts.) increases, the first stage is more likely to land successfully.
- ES-L1,GEO,HEO and SSO has highest success rate.
- Success rate since 2013 kept increasing
- Decision Tree Classifier can be used to predict successful landings and to increase profits.

APPENDIX

Folium charts for Interactive Visualization didn’t appear in git-hub page.

THANK YOU