capstone project- SpaceX launch analysis

Abir Al Ajmi

24th of October 2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Summary of Methodology
  - Data collection using SpaceX REST API
  - Data collection with web scrapping
  - Data wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis using Pandas and Matplotlib
  - Interactive visual analytics and dashboard
  - Machine learning prediction
- Summary of the Results
  - Exploratory Data Analysis results
  - Visual Dashboard
  - Predictive results

INTRODUCTION

- **Project Background**

Falcon 9 is a reusable, two-stage rocket designed and manufactured by SpaceX for the reliable and safe transport of people and payloads into Earth orbit and beyond. Falcon 9 is the world's first orbital class reusable rocket. Reusability allows SpaceX to reuse the most expensive parts of the rocket, which in turn drives down the cost of space access.

- **Questions To Solve?**
  - Will the first stage of falcon 9 land successfully?
  - Which factors impact the successful landing?

METHODOLOGY

- **Data Collection**
  - **SpaceX REST API**
  - We used get request to collect the data, then decoded it to a json file.
  - We used json_normalize() function to perform a panda dataframe from the json file
  - We filled the null values where was necessary using the mean of the data

  - **Web Scrapping from Wikipedia**
  - We used Python BeautifulSoup package to web scape some HTML tables
  - Then, parsed the data from those table and convert them into pandas dataframe

METHODOLOGY

- **Data Wrangling**

  - We calculated the number of launches on each site
  - We calculated the number and occurrence of each orbit
  - We calculated the number and occurrence of mission outcome per orbit type
  - We created a landing outcome label from Outcome column

METHODOLOGY

- **Exploratory Data Analysis with SQL**

  SQL queries performed:

  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

METHODOLOGY

- **Exploratory Data Analysis using Pandas and Matplotlib**
  - Visualize the relationship between Flight Number and Launch Site
  - Visualize the relationship between Payload and Launch Site
  - Visualize the relationship between success rate of each orbit type
  - Visualize the relationship between Payload and Orbit type
  - Visualize the launch success yearly trend

METHODOLOGY

- **Interactive visual analytics (Launch Sites Locations Analysis with Folium)**
  - Marked all launch sites on a map
  - Marked the success/failed launches for each site on the map
  - Calculated the distances between a launch site to its proximities

METHODOLOGY

- **Machine learning prediction (Classification)**
  - We preprocessed the data by standardize it
  - We splited the data into train and test sample with the test size to be 20%
  - We tested four types of classification: K-Nearest point, Decision Tree, Logistic Regression, Support Vector Machine

RESULTS (Exploratory Data Analysis results)

We found the different launch site using DISTINCT.

```
In [5]: %sql select distinct (LAUNCH_SITE) from SPACEXDATASET

 * ibm_db_sa://jbd99186:***@c77d6f2-5da9-48a9-81f8-86b
s.appdomain.cloud:31198/bludb
Done.
```

Out[5]:

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

RESULTS (Exploratory Data Analysis results)

We calculated the total payload carried by boosters from NASA as 45596.

```
In [11]: %sql select SUM(PAYLOAD_MASS__KG_) from SPACEXDATASET where CUSTOMER = 'NASA (CRS)'

 * ibm_db_sa://jbd99186:***@c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg
s.appdomain.cloud:31198/bludb
Done.

Out[11]:
1
45596
```

RESULTS (Exploratory Data Analysis results)

We found the date for first successful landing to be 22/12/2015.

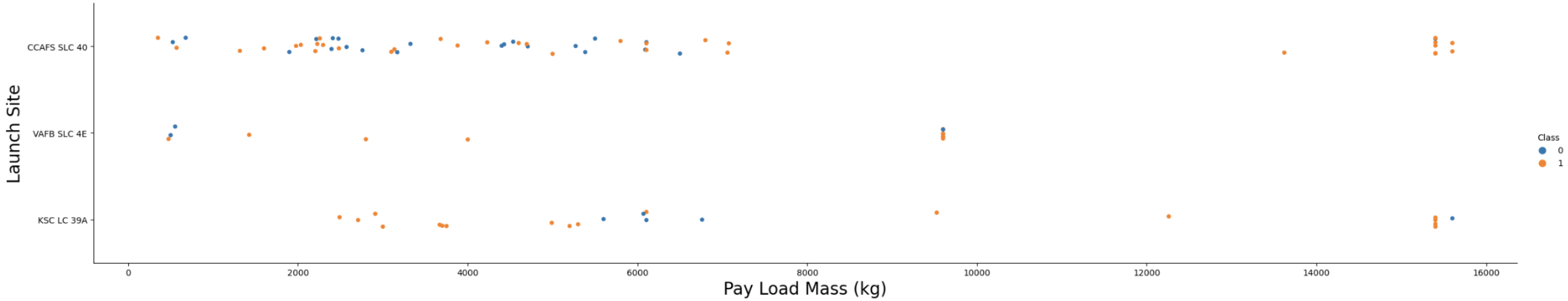
```
In [13]: %sql select MIN(DATE) from SPACEXDATASET where LANDING__OUTCOME like 'Success%'

 * ibm_db_sa://jbd99186:***@c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od
s.appdomain.cloud:31198/bludb
Done.

Out[13]:
        1
2015-12-22
```

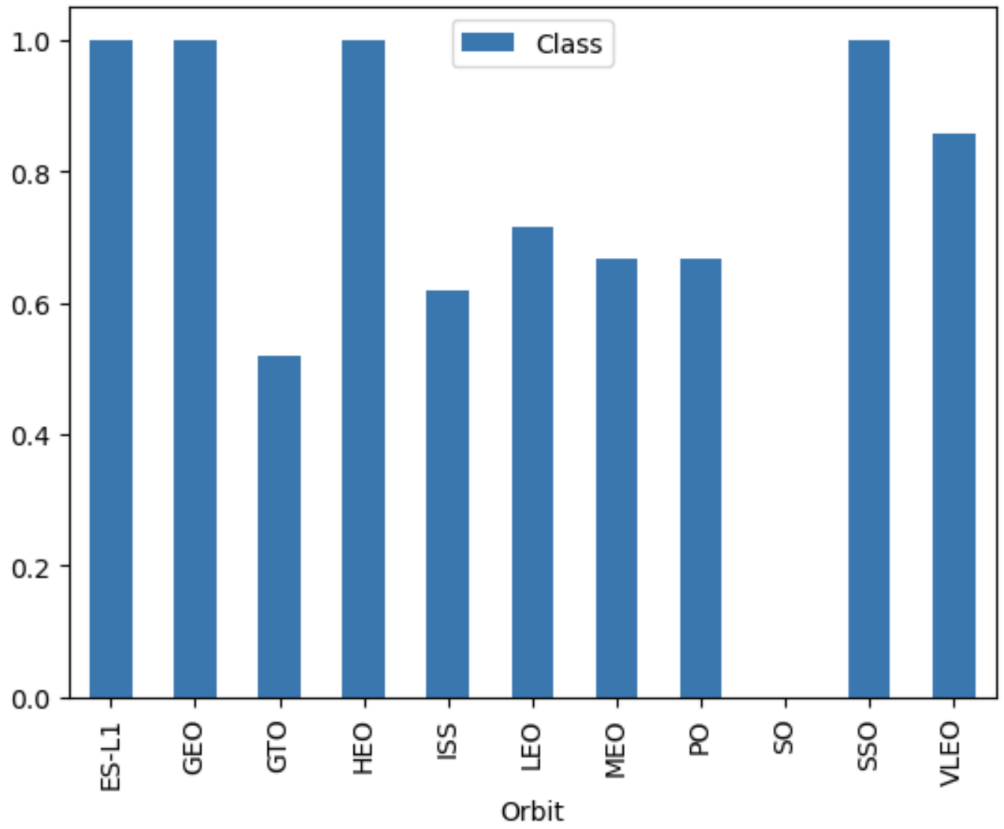
RESULTS (Exploratory Data Analysis results)

Observing Payload Vs. Launch Site scatter point chart, we found for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).



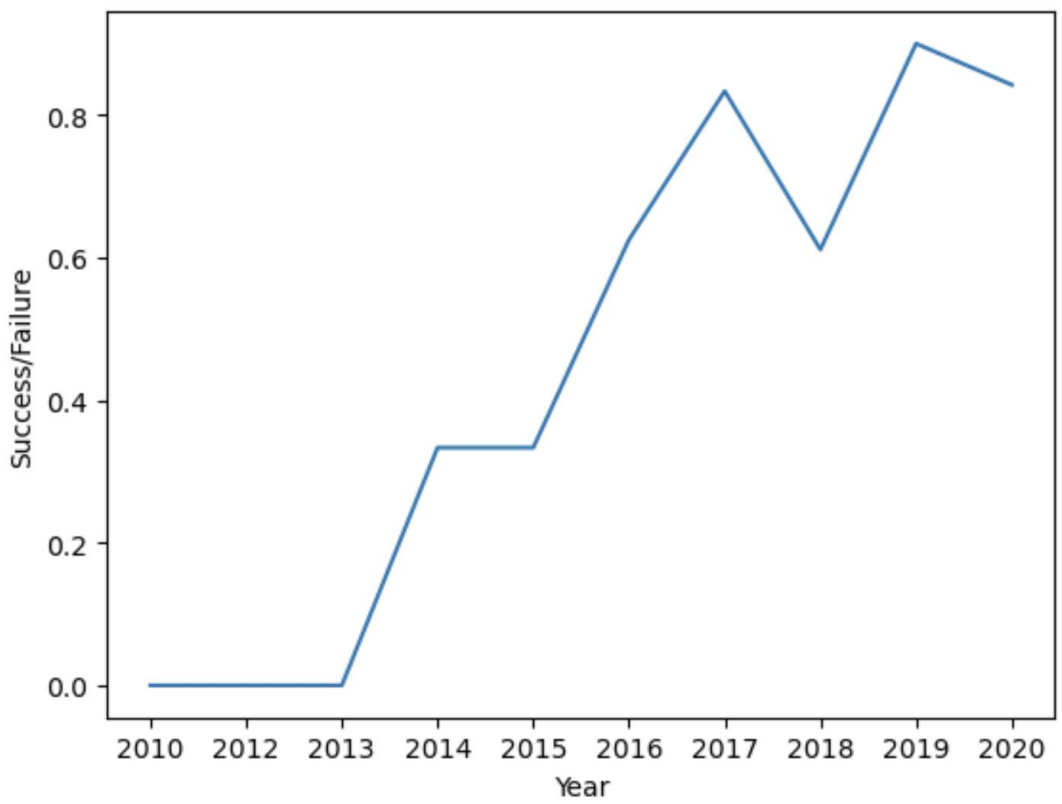
RESULTS (Exploratory Data Analysis results)

From the plot, we can see that ES-L1, GEO, HEO, SSO had the most success rate.



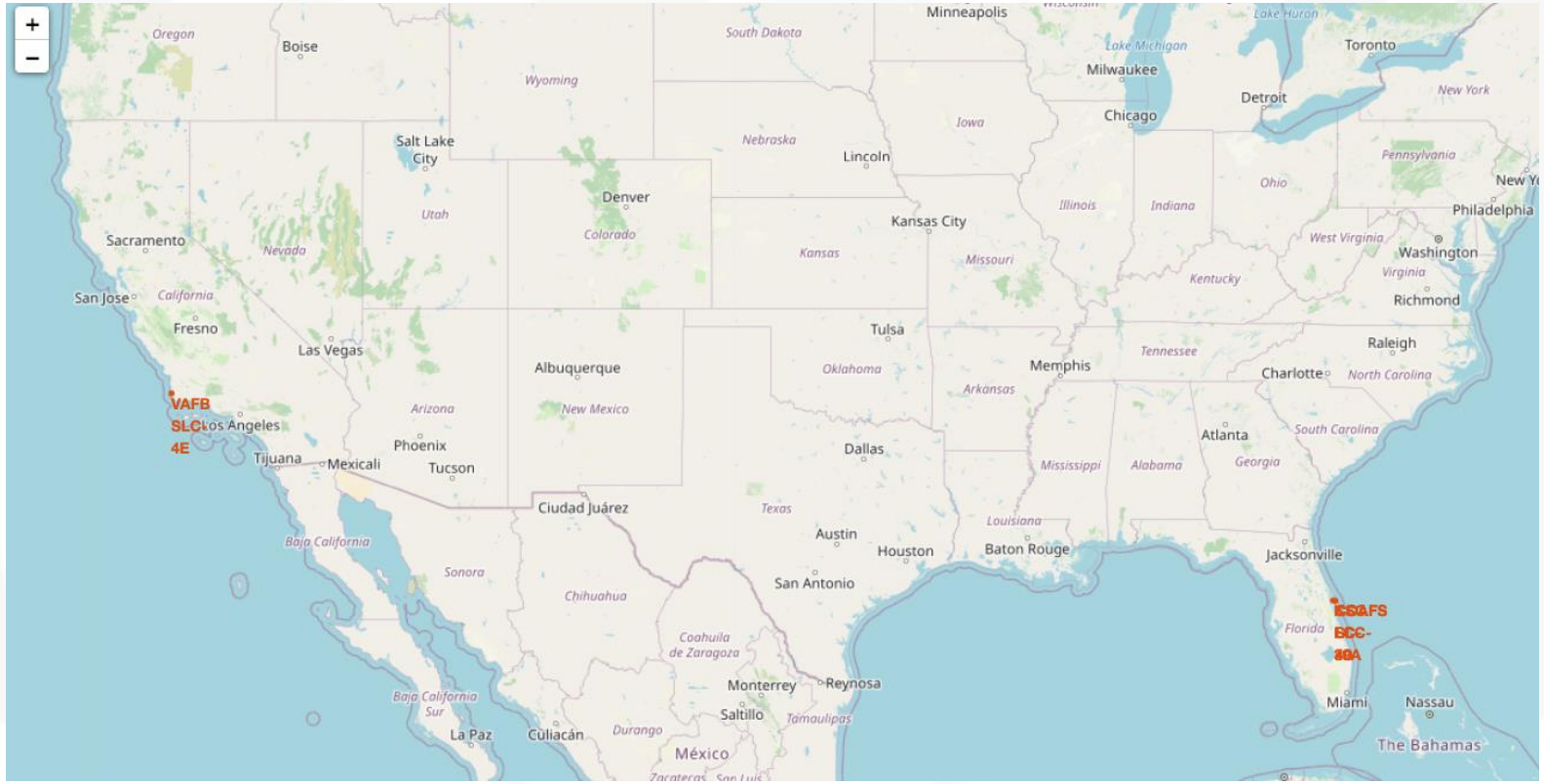
RESULTS (Exploratory Data Analysis results)

From the plot, we can see the success rate start increasing in 2013.



RESULTS (Interactive visual analytics using Folium)

Launch sites across the US (Florida and California).

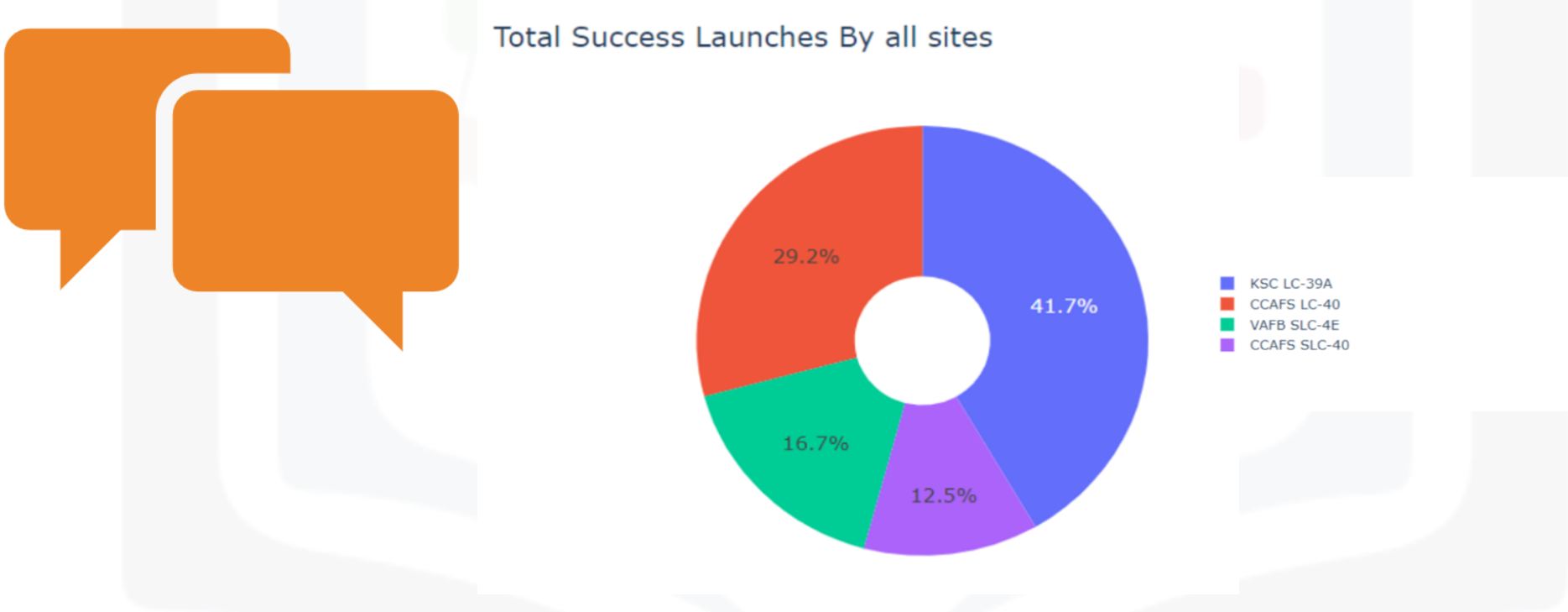


RESULTS (Interactive visual analytics using Folium)

Launch sites with colors where green indicates high success rate.

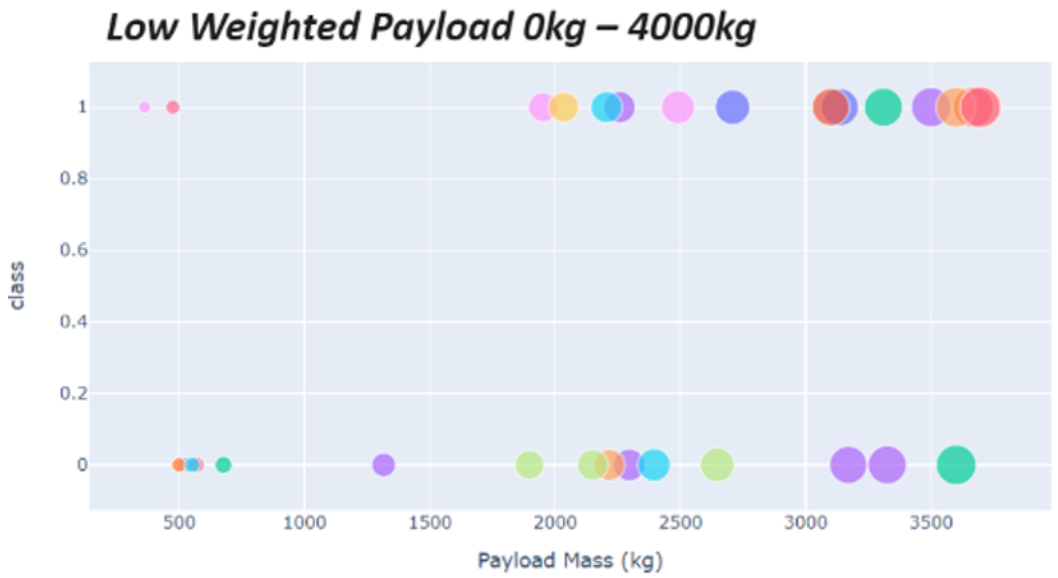
DASHBOARD

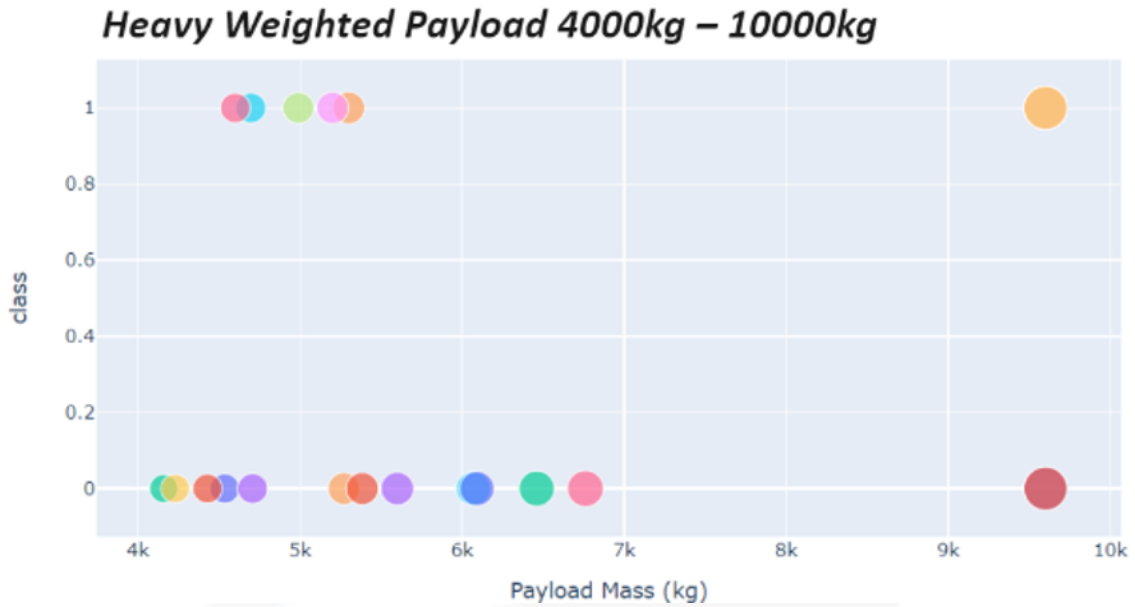
Launch site KSC LC-39A has the most successful landing.



DASHBOARD

Success rates for low weighted payloads is higher than the heavy weighted payloads

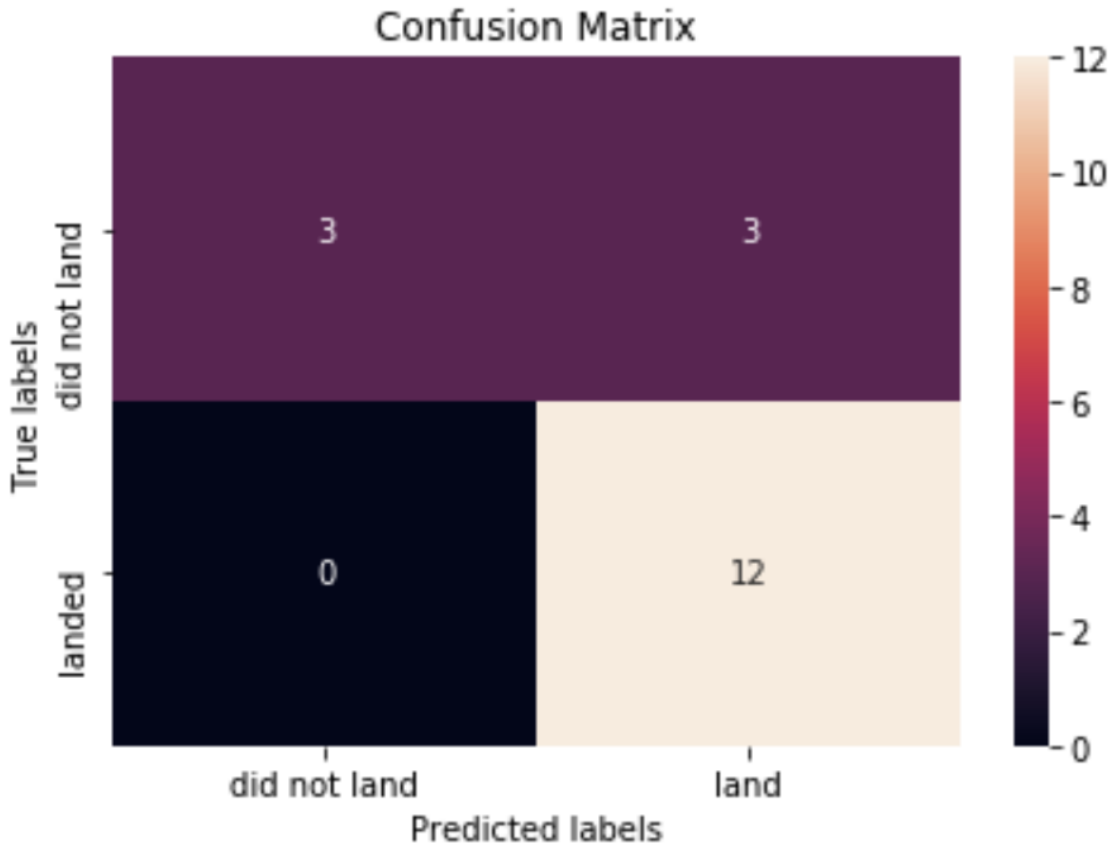




RESULTS (Predictive Classifier)

The decision tree had the highest classificion accuracy.

we see that decision tree can distinguish between the different classes. We see that the major problem is false positives.



CONCLUSION

- Orbits ES-L1, GEO, HEO, SSO had the most success rate.
- •Launch success rate is directly proportional with time since 2013
- •KSC LC-39A had the most successful launches of any sites.

•The Decision tree classifier is the best machine learning algorithm for this task.

APPENDIX

- Include any relevant additional charts, or tables that you may have created during the analysis phase.