Falcon9 Launch Cost Prediction

Xiaoyu Chen

2022-07-12

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Falcon 9 rocket can **re-use the first stage**(and hence reduce cost)
- Predict first stage success rate
  - Acquire historical data (Data Wangling)
  - Data Analysation (EDA)
  - Train ML models for further prediction

INTRODUCTION

- Falcon 9 claims a cost per launch 62 million USD
- Competitors: 165 million USD
- An accurate prediction cost per launch for Falcon9 based on historical data
- What we took into consideration:
  - Launch sites
  - Payload Mass
  - Orbitals
  - Year

METHODOLOGY

- Collect data
  - SpaceX API
  - Wikipedia
  - SQL Database
- Exploratory Data Analysis (EDA)
  - Determine what would be the label to train our models
- EDA prediction
  - Visualise relationship between independent variables*
- Train machine learning models
  - Logistic regression
  - Support vector machines
  - KNN
  - Decision tree

Data Source

A few distinct data sources are used in this project

Data Wrangling: Wikipedia Data

EDA: Wikipedia Data

SQL: uploaded data from a csv file*

Folium Map – study on Launch site: additional data provided in csv*

Dash: additional data

Model training: Wikipedia Data

RESULTS – Data Wrangling

Launches at each sites:

CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

Landing Success Rate:

0.67

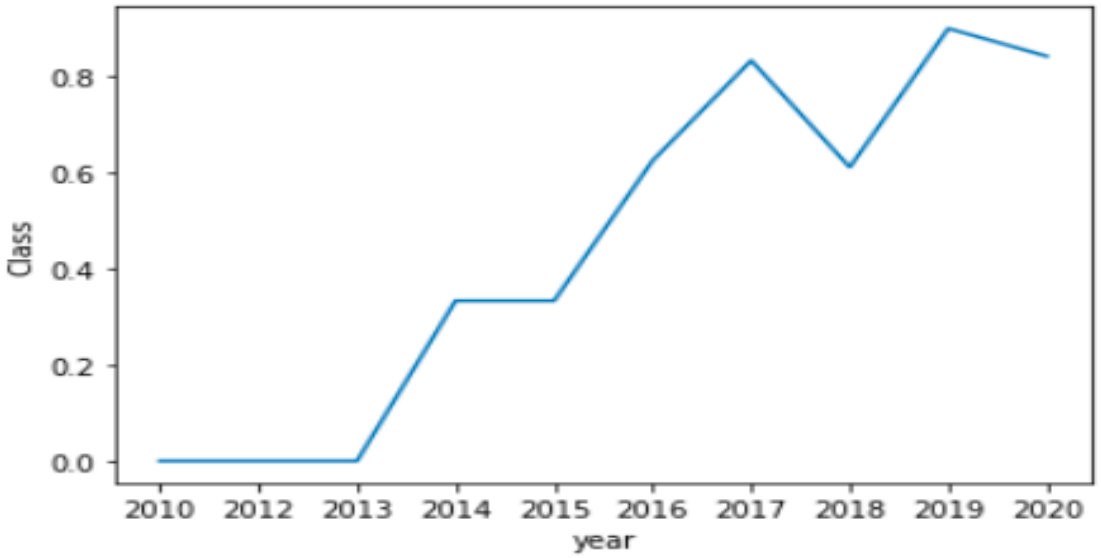
Launches per Orbital:

GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
ES-L1	1
HEO	1
SO	1
GEO	1

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/Data%20wrangling.ipynb

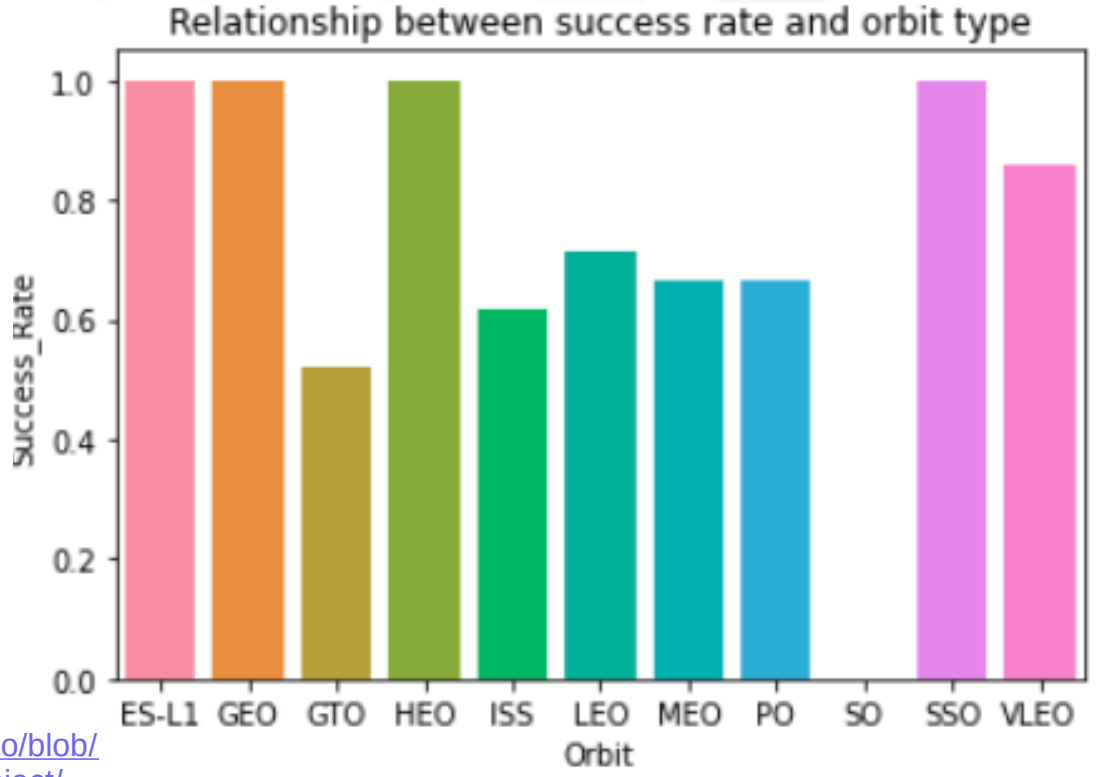
RESULTS – EDA

We see correlation for Launch Site, Payload Mass, Orbital and Year



Success Rate vs.Year

Success Rate per Orbital:



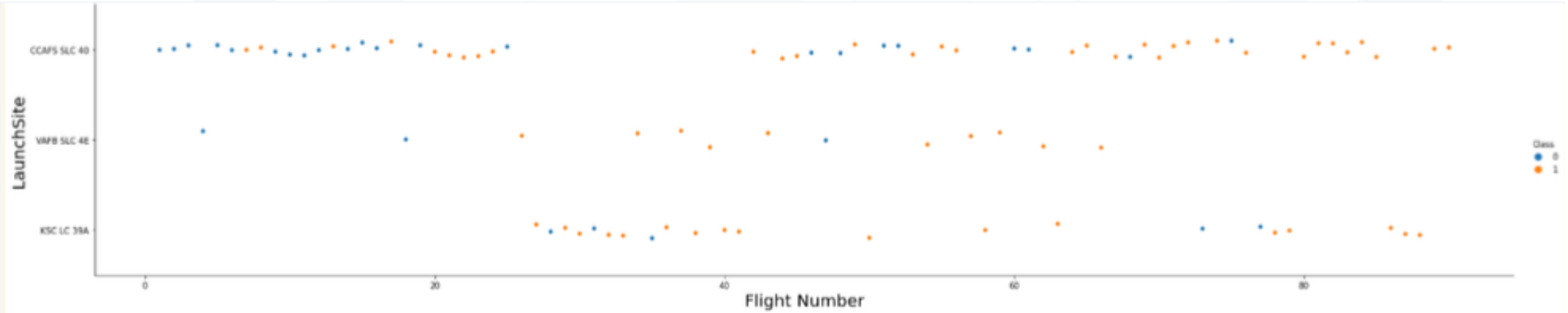
https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/EDA_visualisation.ipynb

IBM Developer

SKILLS NETWORK

EDA additional Info

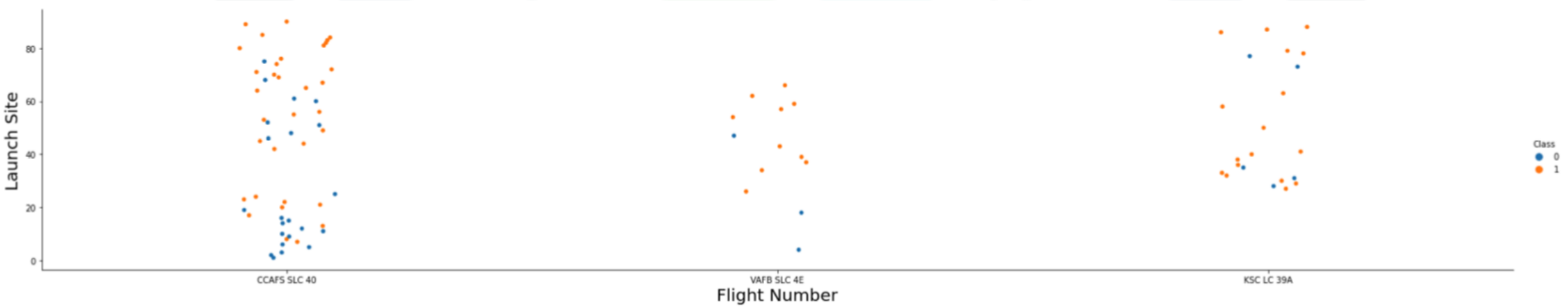
Flight Number vs. Launch Site



https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/EDA_visualisation.ipynb

EDA additional Info

PayloadMass vs. Launch Site

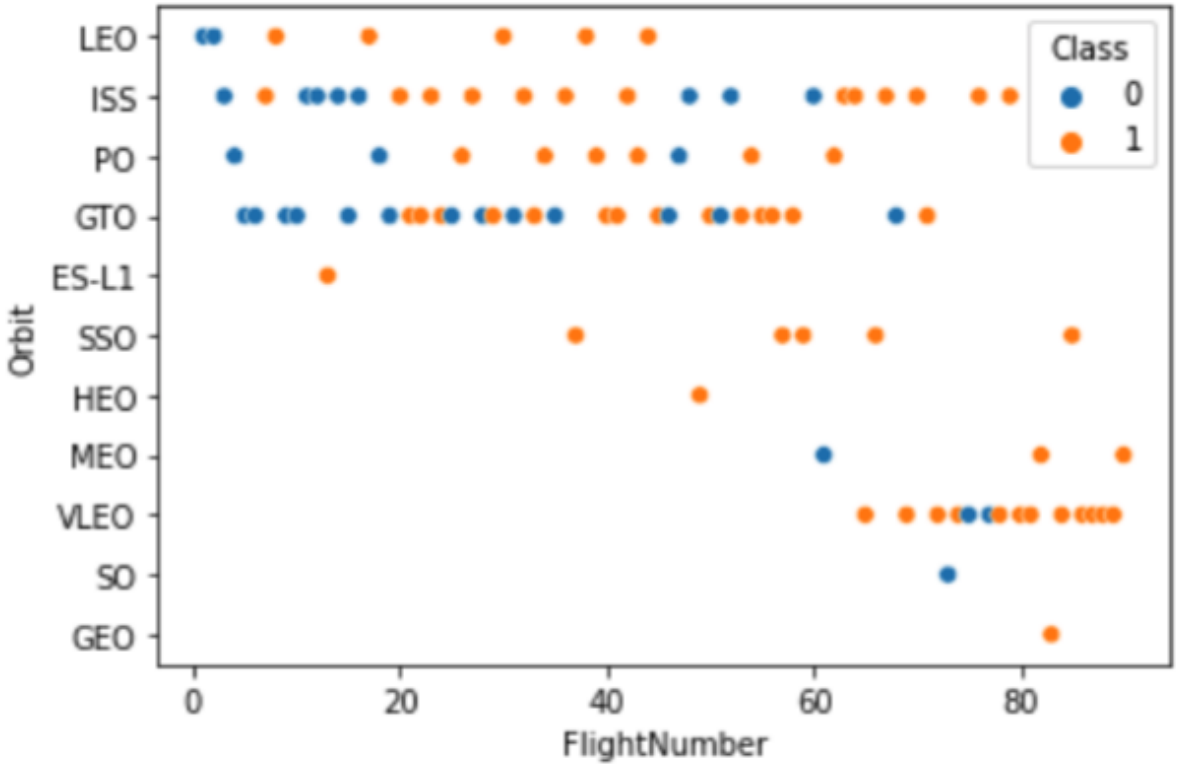


https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/EDA_visualisation.ipynb

IBM Developer
SKILLS NETWORK

EDA additional Info

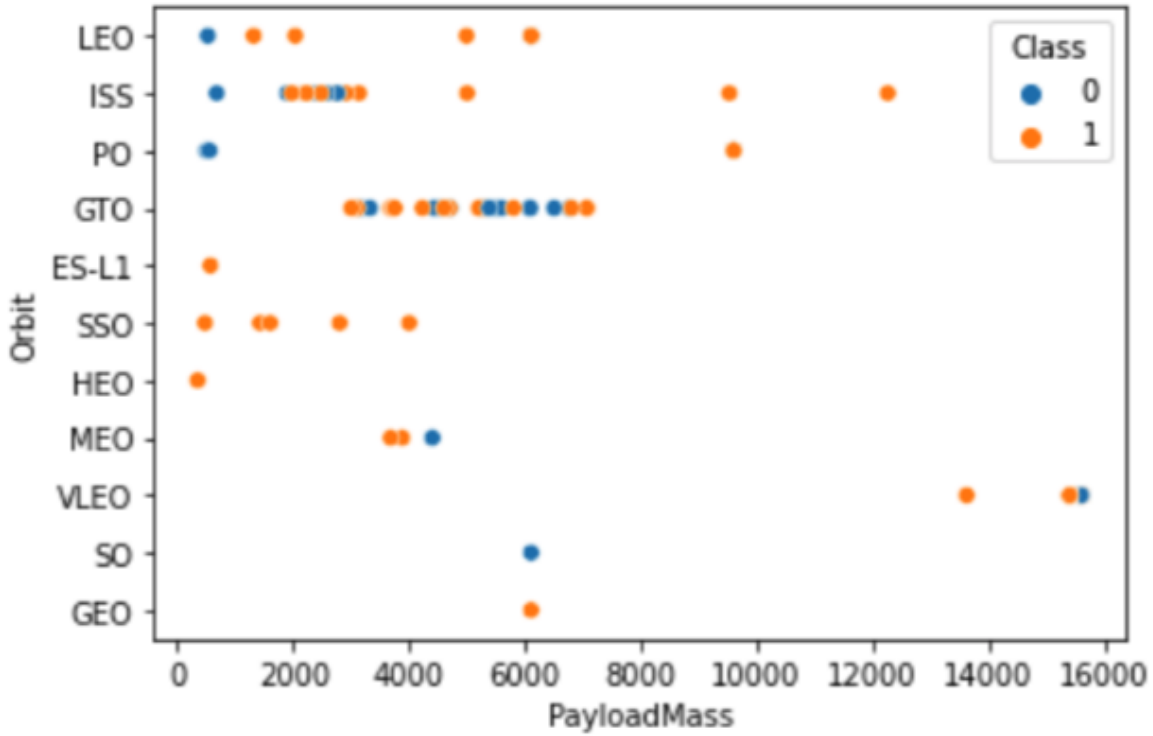
Flight Number vs. Orbital Type



https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/EDA_visualisation.ipynb

EDA additional Info

Payload vs. Orbital Type



https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/EDA_visualisation.ipynb

SQL Database Exercise

Additional data in a SQL database was also analysed

Mission Success Rate: **97%** (98/101)

Landing Success Rate: **65%** (66/101)

Distinct Launch Site:

CCAFS SLC-40	34
CCAFS LC-40	26
KSC LC-39A	25
VAFB SLC-4E	16

* Launch Site info from wikipedia

CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

2. Launch Site name begins with ‘CCA’

```
Select DISTINCT * from SPACEX where LAUNCH_SITE like ‘CCA%’
```

```
CCAFS SLC_40
CCAFS SL-40
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

3. Calculate the total payload carried by booster from NASA

```
Select sum(PAYLOAD_MASS__KG_) from SPACEX where CUSTOMER = ‘ NASA%’
```

```
45596
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

4. Calculate the average payload mass carried by booster F9 v1.1

```
Select AVG(PAYLOAD_MASS__KG_) from SPACEX where BOOSTER_VERSION = 'F9 v1.1'
```

```
2928
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

5. First successful ground landing date: find the date when the first successful landing outcome in ground pad

```
Select min(DATE) from SPACEX where LANDING__OUTCOME = 'Success%'
```

```
2015-12-22
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

6. List the names of boosters which have success in drone ship and have payload greater than 4000 but less than 6000

```
Select distinct BOOSTER_VERSION  from SPACEX where PAYLOAD__MASS_KG_
between 4000 and 6000
```

```
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

7. Calculate the total number of successful and failure mission outcomes

```
Select MISSION_OUTCOME, count(*) from SPACEX group by MISSION_OUTCOME
```

```
Failure (in flight)                 1
Success                             99
Success(payload status unclear) 1
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

8. List the names of the booster which have carried the maximum payload mass

```
Select distinct BOSTER_VERSION from SPACEX where PAYLOAD_MASS__KG_ in
(select max(PAYLOAD_MASS__KG_) from SPACEX)
```

F9 B5 B1048.4, F9 B5 B1049.4, F9 B5 B1051.3, F9 B5 B1056.4, F9 B5 B1048.5
F9 B5 B1051.4, F9 B5 B1049.5, F9 B5 B1060.2 F9 B5 B1058.3 F9 B5 B1051.6
F9 B5 B1060.3, F9 B5 B1049.7

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

9. List the launch records for months in 2015

```
Select *  from SPACEX where YEAR(DATE) = 2015
```

https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

SQL- Additional Info

10. Rank the count of successful landings between 2010-06-04 and 2017-03-20

```
Select count(LANDING_OUTCOME)  from SPACEX where LANDING_OUTCOME = upper('%SUCCESS%') And (DATE between '04-06-2010' and '20-03-2017)
```

34

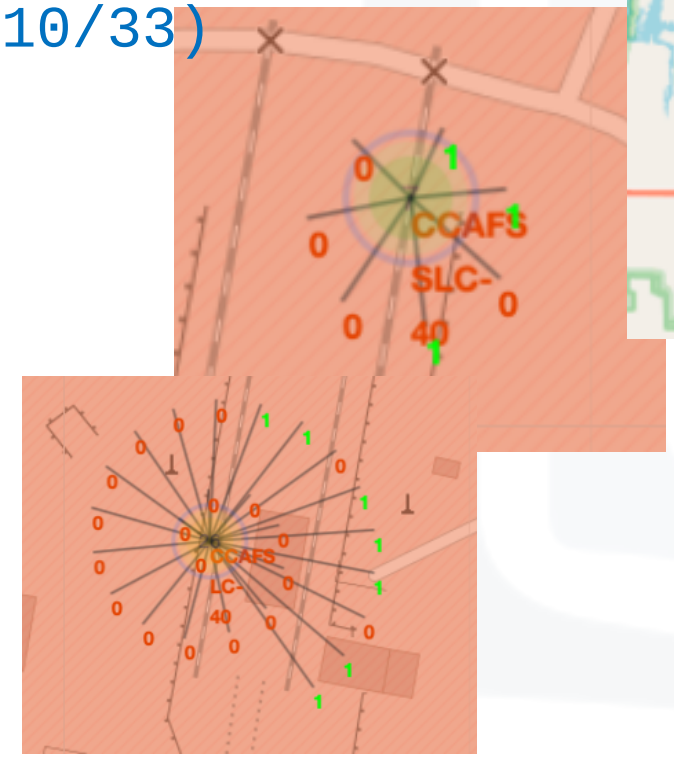
https://github.com/xyll313/Data_Science_Portfolio/blob/ibm_course/Data_Science_Course/capstone-project/SQL_exercise.ipynb

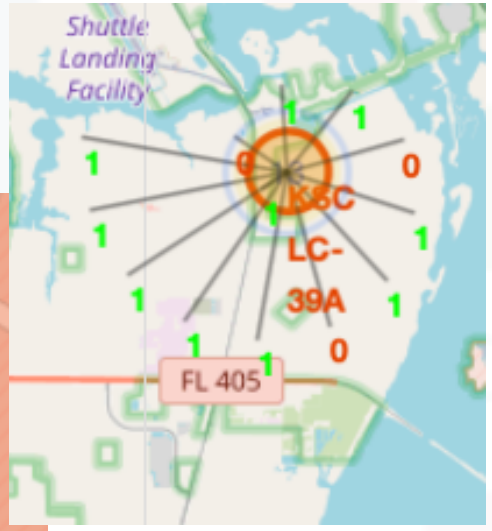
Launch-site Folium Exercise

All launch sites are close to the coasts

Folium Map – Success Rate

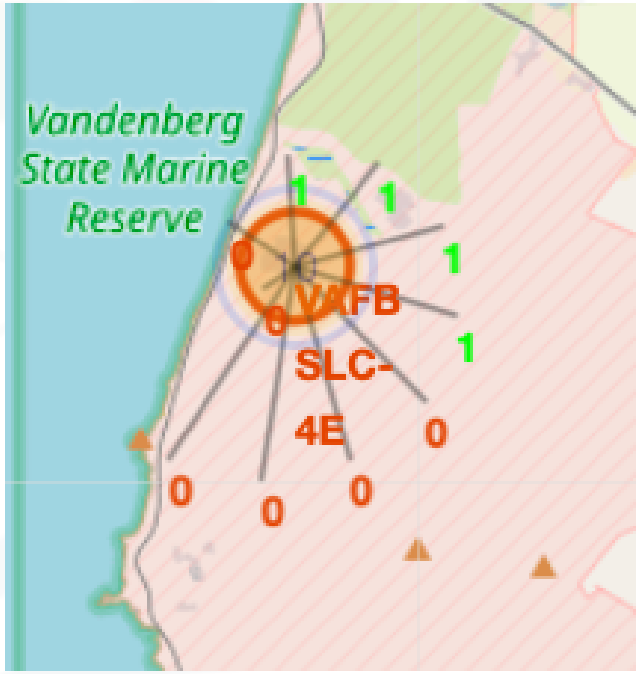
CCAFS SLC-40 **30%** (10/33)



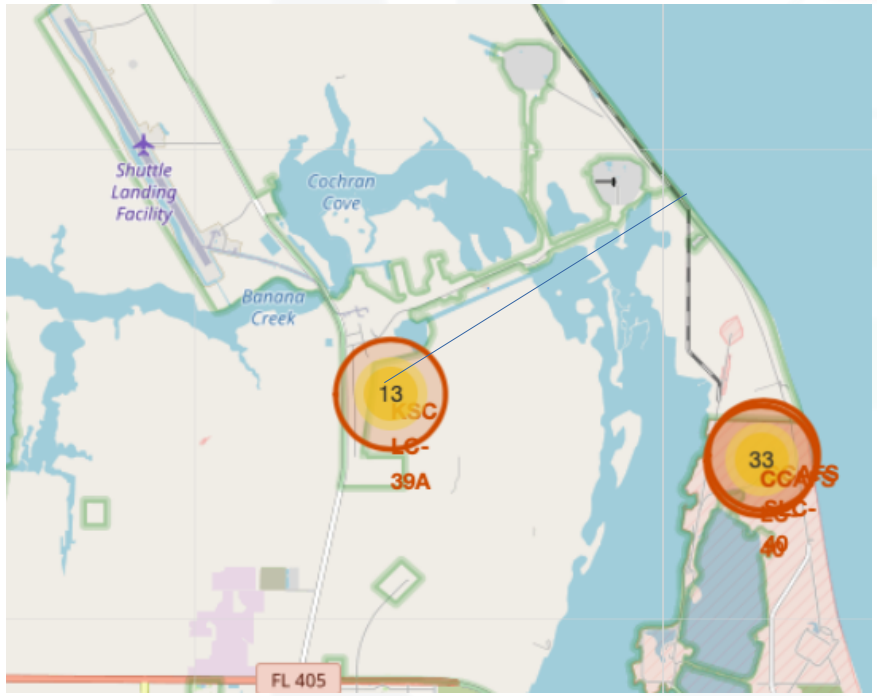


CCAFS SLC-39A **77%** (10/13)

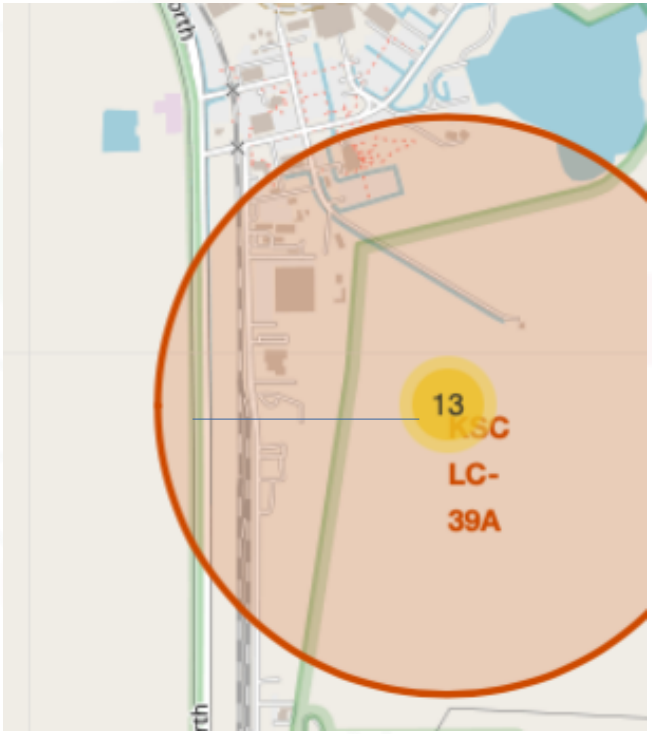
VAFB SLC-4E **40%** (4/10)



Folium Map – CCAFS SLC-39A

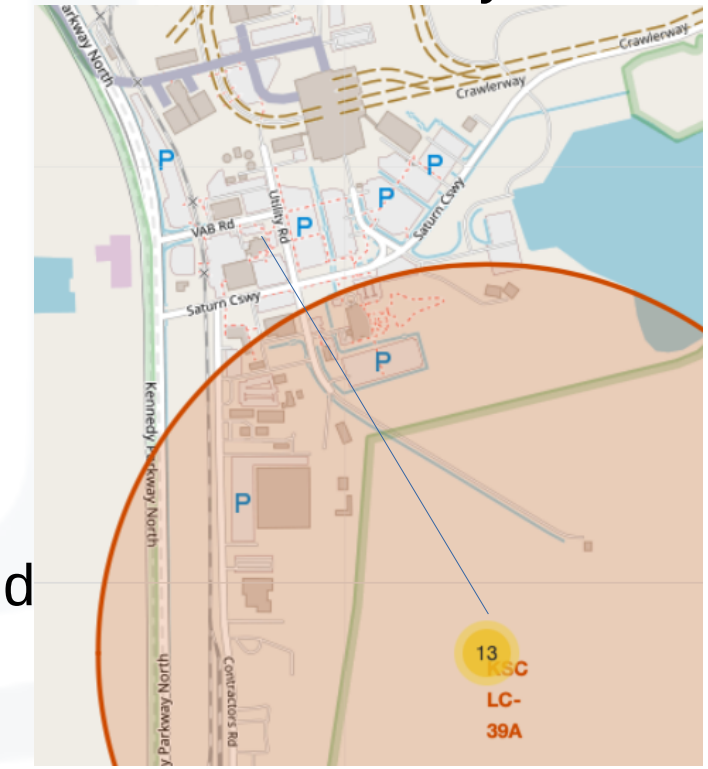


Close to coastline



Close to railway/road

Close to city



RESULTS – Summary

Launch Site info:

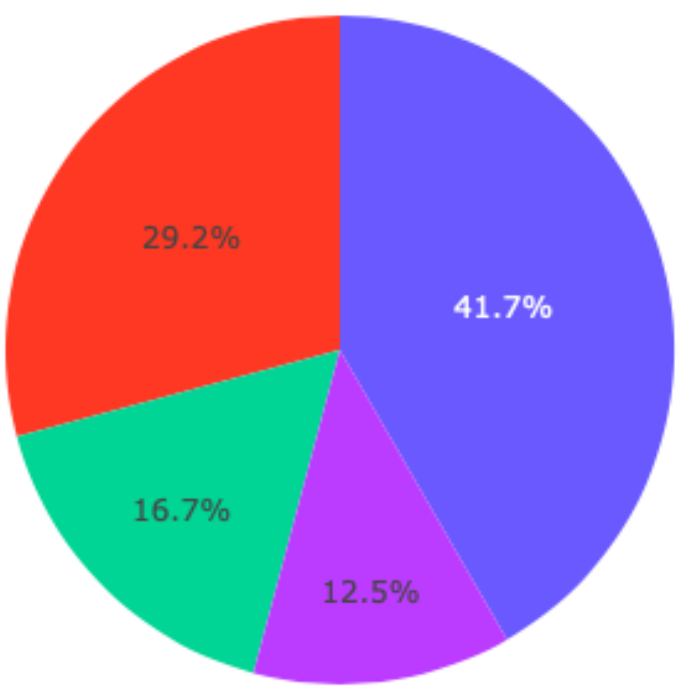
CCAFS	SLC-40	89
CCAFS	LC-40	26
KSC	LC-39A	47
VAFB	SLC-4E	29

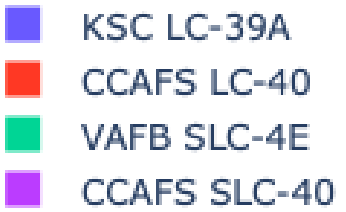
Overall Landing Success Rate: **68%** (132/192)

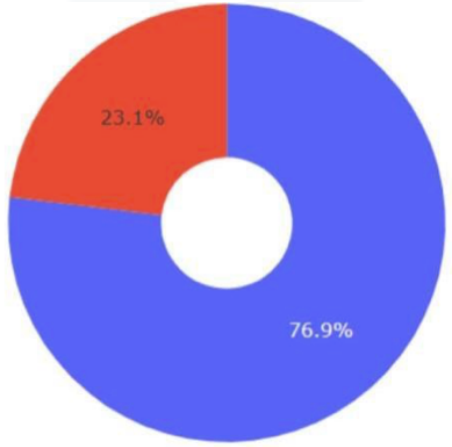
Independent Variables used for model training:
Launch Site, PayloadMass, Orbital, Year

DASHBOARD

Success Rate vs. Launch sites KSC LC-39A has the most launches



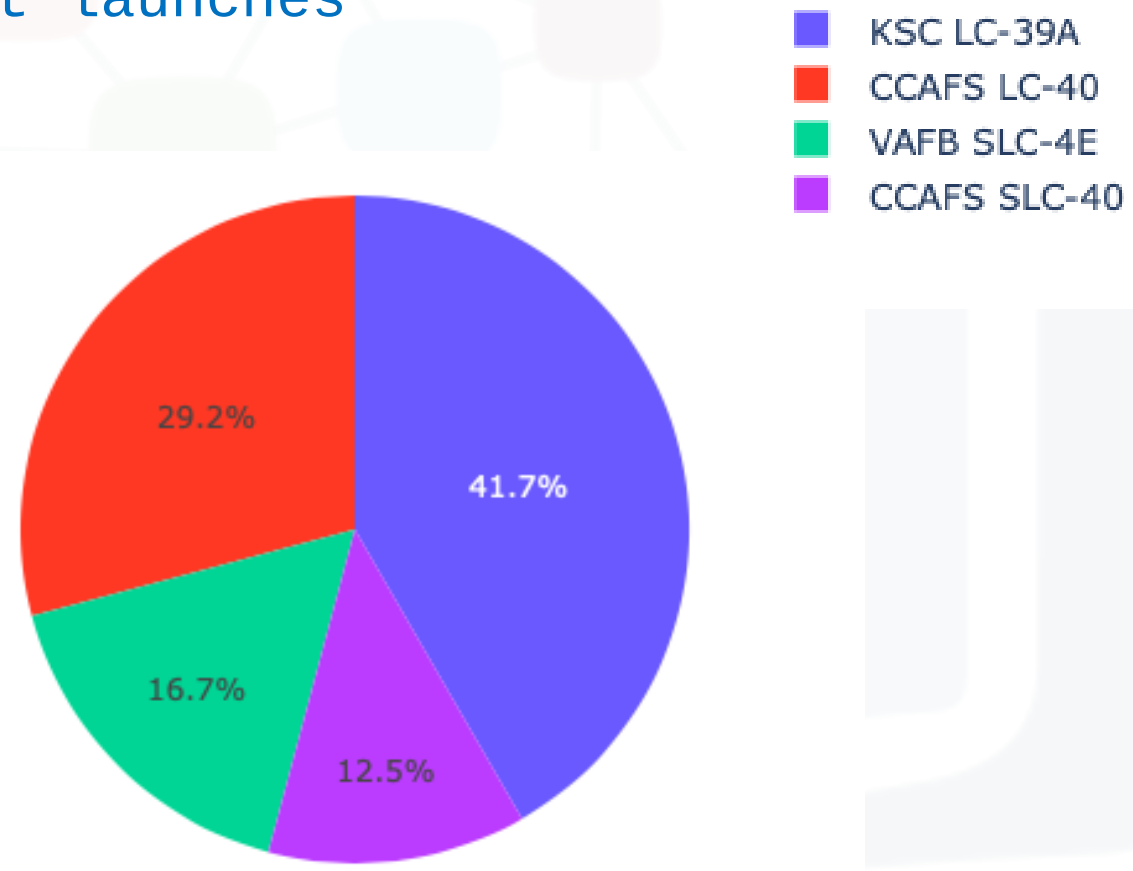




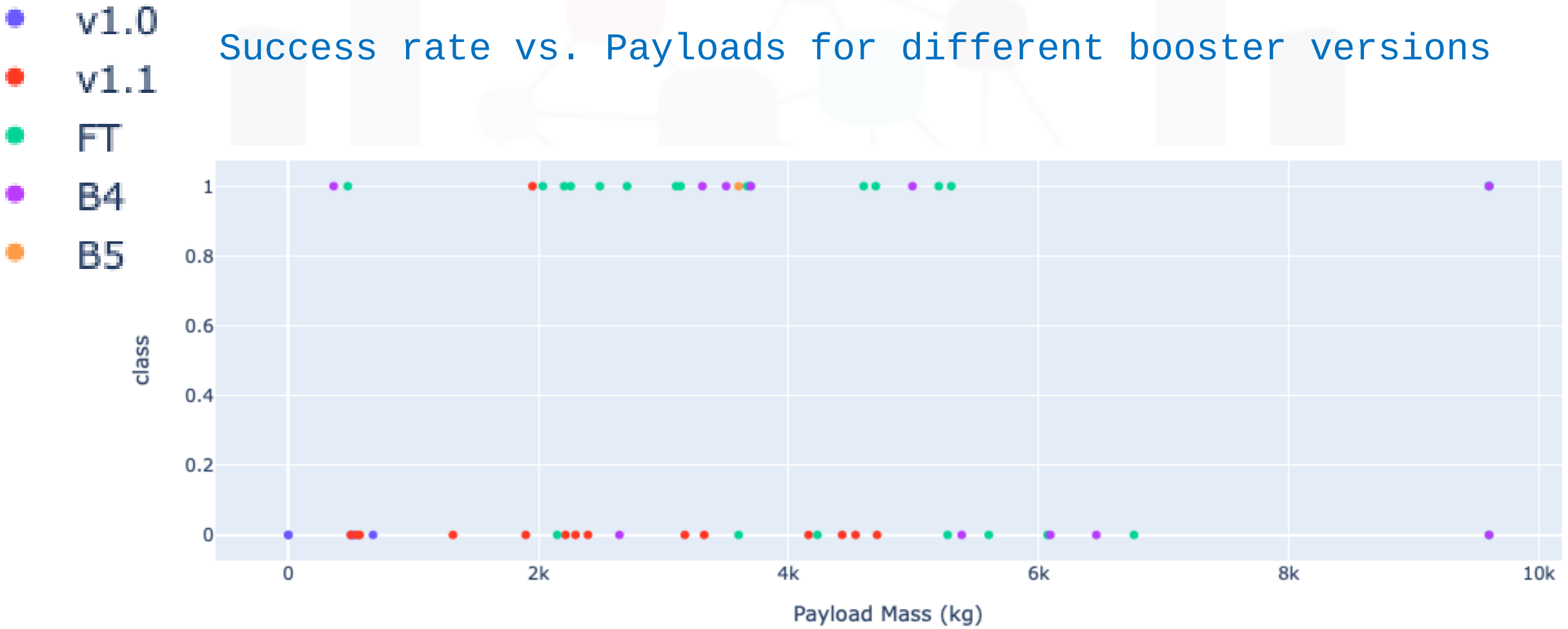
With success rate at 76.9%

DASHBOARD

KSC LC-39A has the most launches

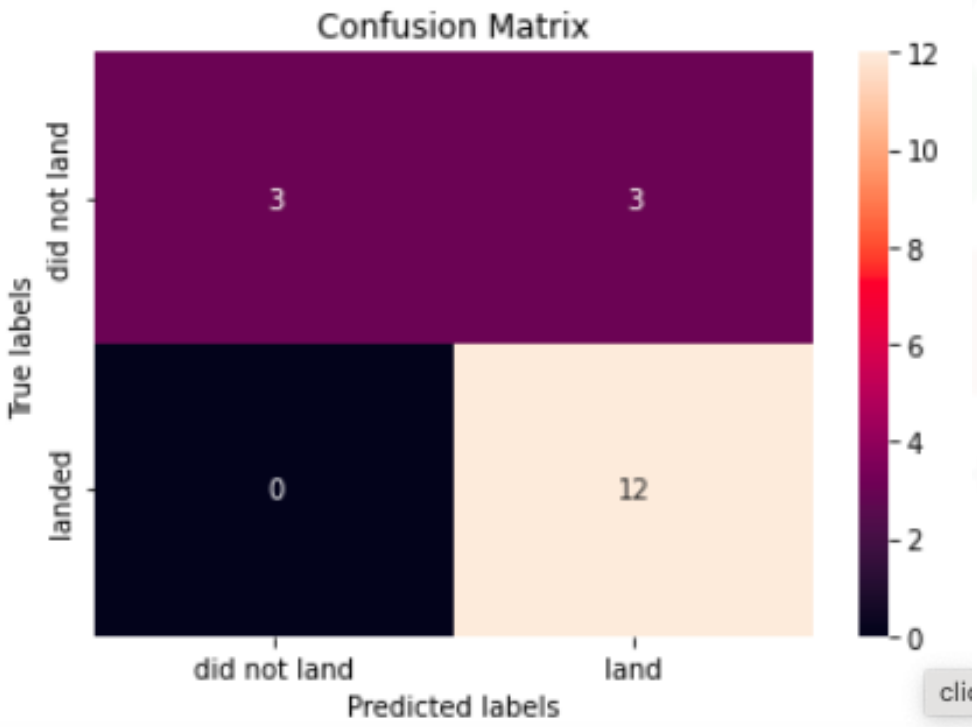


DASHBOARD


v1.0
v1.1
FT
B4
B5
Success rate vs. Payloads for different booster versions
class
1
0.8
0.6
0.4
0.2
0
0
2k
4k
6k
8k
10k
Payload Mass (kg)

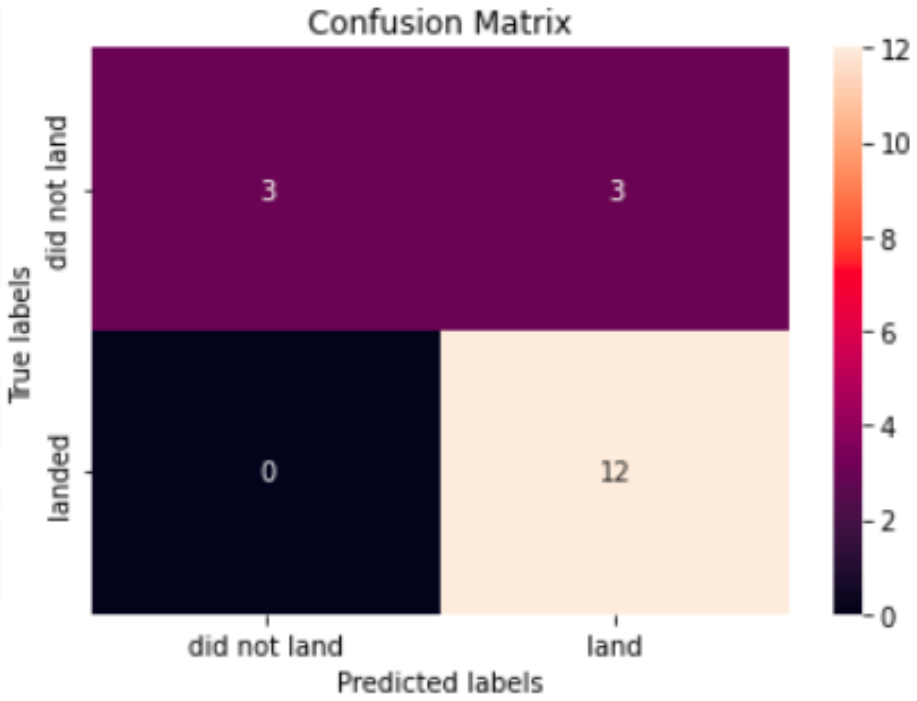
Machine Learning Models

Logistic Regression:



Accurancy Socre: 0.83

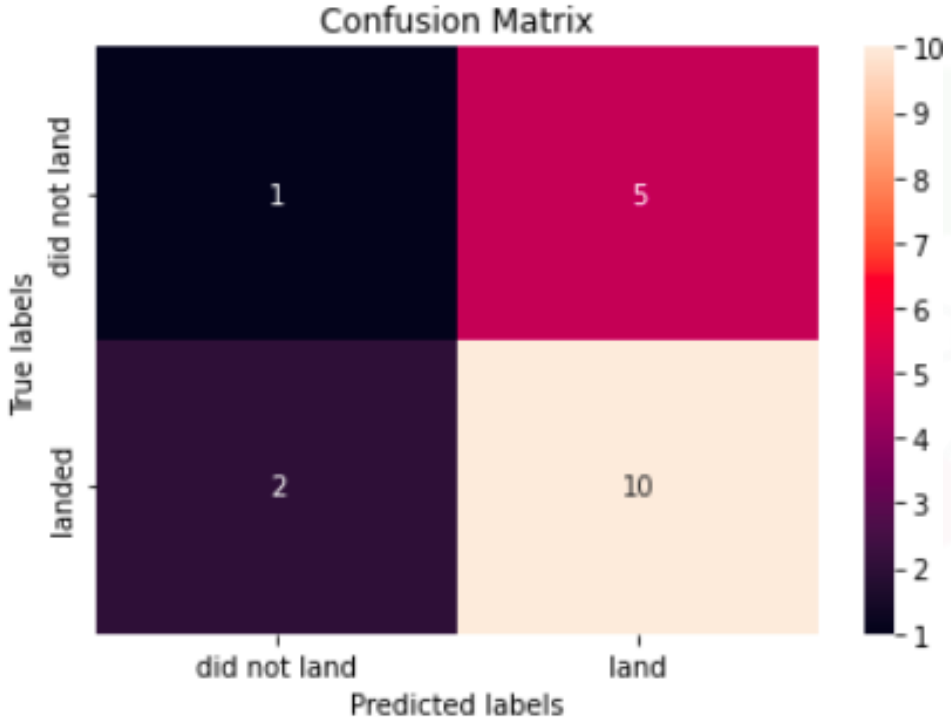
Support Vector Machine:



Accurancy Socre: 0.83

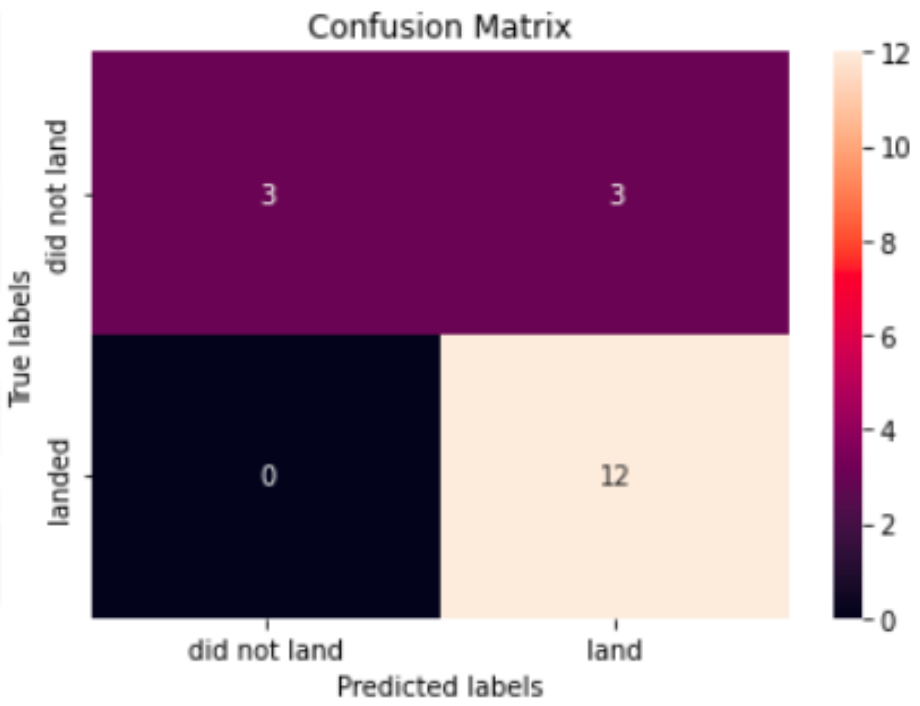
Machine Learning Models

Decision Tree:



Accurancy Socre: 0.61

KNN:



Accurancy Socre: 0.83

CONCLUSION

- Overall landing success Rate at **68%** based on historical Data
- CCAFS SLC-39 has the highest success rate
- Model training reveils that **Logistic Regression, Support Vector Machine and KNN** all have the same level of accuracy at 83%

Analysation & Further Work

- Launch Sites, PayLoadMass,Orbitals and Year all have impacts on success rate. Detailed analysation on Launch Site is required as the other factors can’t be easily modified.
- See if it’s the facilities at one Launch Site improved success rate, or if it’s because it only launches easier missions.
- Train the models again by using different selection of training and testing data to see if we obtain the same results.