Space Race with Data Science

Ana García París

23/08/2024

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- For the capstone project, I predict if SpaceX Falcon 9 first stage is going to land successfully by applying different machine learning algorithms
- This project includes

  Data collection, wrangling

  Exploratory data analysis

  Data visualizations

  ML predictions
- The graphs provide some reason to believe that specific features of rocket launches are correlated with whether they succeed or fail
- The decision tree is potentially the best algorithm

INTRODUCTION

- For the capstone project, I predict if SpaceX Falcon 9 first stage is going to land successfully by applying different machine learning algorithms. SpaceX is capable of saving about 100 million dollars because they are able to reuse the first stage. Moreover, if I can decipher whether a rocket's first stage will land, I can help determine the cost of the launch. This would benefit rivaling companies of SpaceX.
- I'm looking to determine for a specific set of features in a rocket launch like the mass, orbit type, launch site, and others, will the first stage land?

Part 1

Methodology

METHODOLOGY

Executive Summary:

- Data collection methodology:

  SpaceX API

  WebScraping

- Perform data wrangling:

  Based on outcome data

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models:

  Train and set tests -> classification algorithms -> make conclusion

Data Collection – SpaceX API

- SpaceX API o API provides different types of SpaceX rocket launches o Missing data replaced by mean o 90x17 of data

- https://github.com/agprays/Capstone-agprays/blob/main/1_jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Webscraping

https://github.com/agprays/Capstone-agprays/blob/main/2_jupyter-labs-webscraping.ipynb

Data Wrangling

- EDA -> summarizations -> Outcome Label

https://github.com/agprays/Capstone-agprays/blob/main/3_labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Use scatterplots and boxplots to show relationships

https://github.com/agprays/Capstone-agprays/blob/main/5_edadataviz.ipynb

EDA with SQL

- SQL queries o Launch site names

  Total payload mass by NASA

  Average payload mass by booster

https://github.com/agprays/Capstone-agprays/blob/main/4_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Used markers, circles, lines and marker clusters
- Shows successes, fails, distances, and locations

https://github.com/agprays/Capstone-agprays/blob/main/6_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Lauches by site percentages
- Payload range
- Helps show the relationship between payload and launch site

Predictive Analysis (Classification)

- Classification Model

  Logistic regression

  Support Vector Machine

  Decision Tree

  K-Nearest Neighbors

- Data prep -> combinations -> conclusion

https://github.com/agprays/Capstone-agprays/blob/main/8_SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

RESULTS

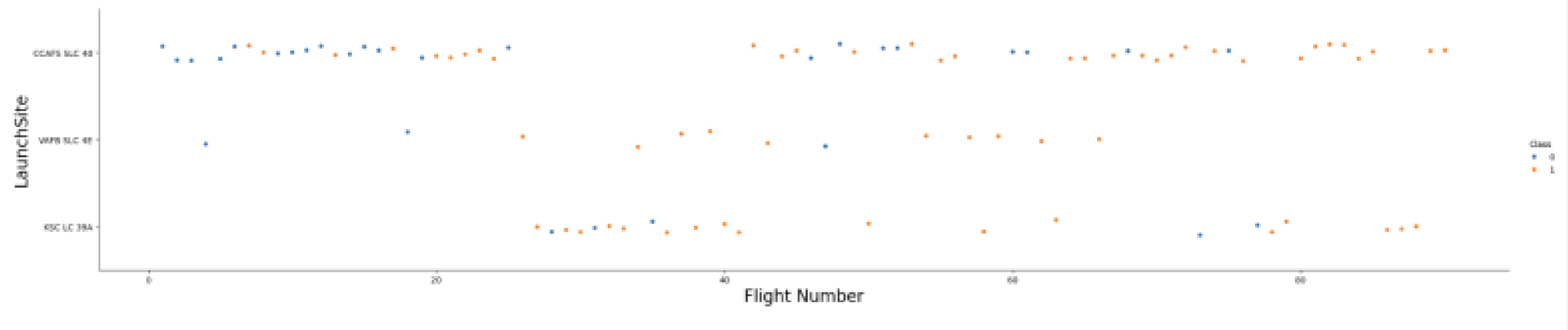
- Four launch sites
- Two booster versions
- Improvement over years

Part 2

Insights drawn fron EDA

Flight Number vs. Launch Site

Flight Number vs. Launch Site o success over time

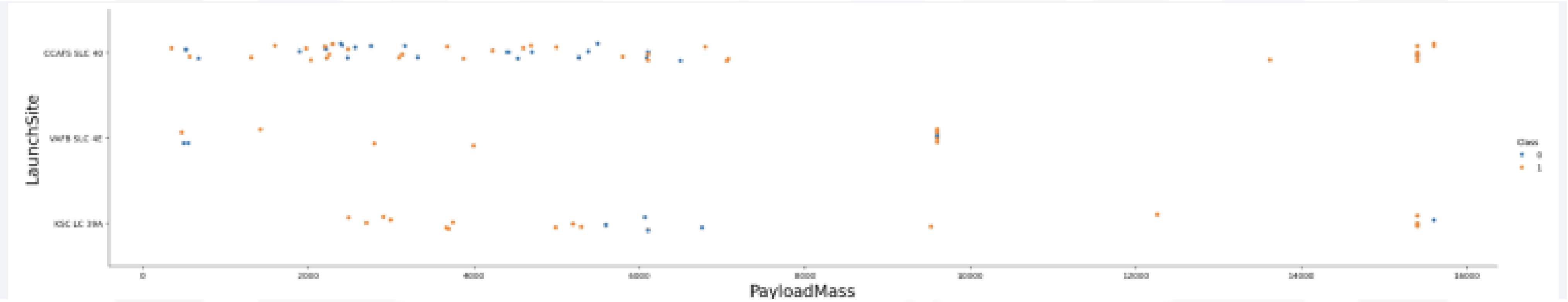


IBM Developer

SKILLS NETWORK

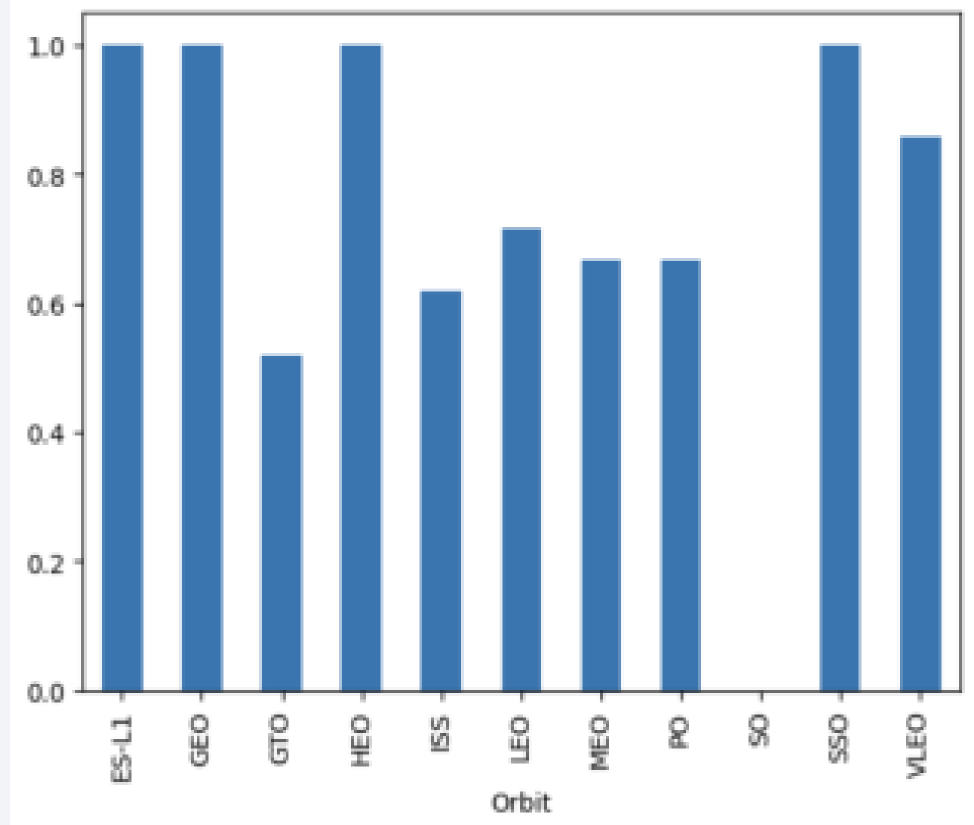
Payload vs. Launch Site

Payload vs. Launch Site o Payloads > 9,000kg have good success rate



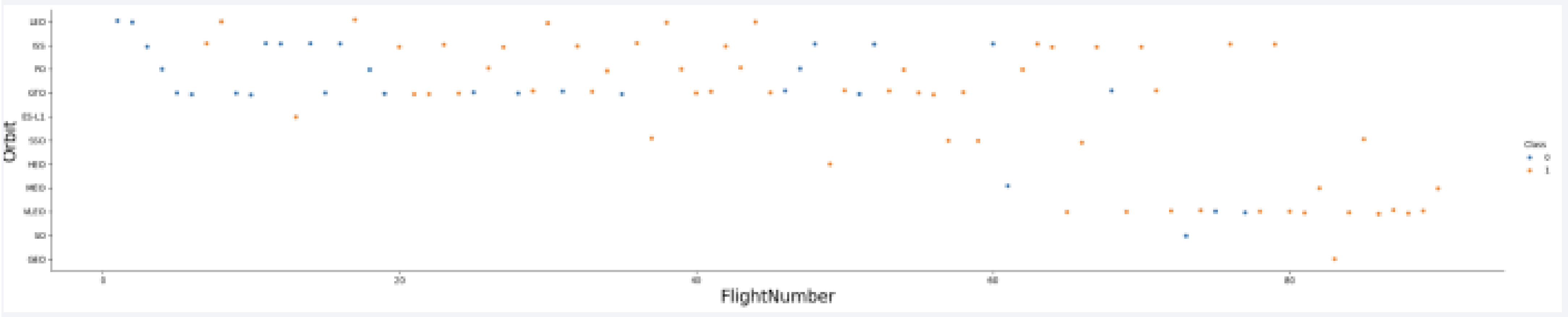
Success Rate vs. Orbit Type

ES-L1, GEO, HEO, SSO are all tied at the top



Flight Number vs. Orbit Type

Flight number vs. Orbit type o Success over time

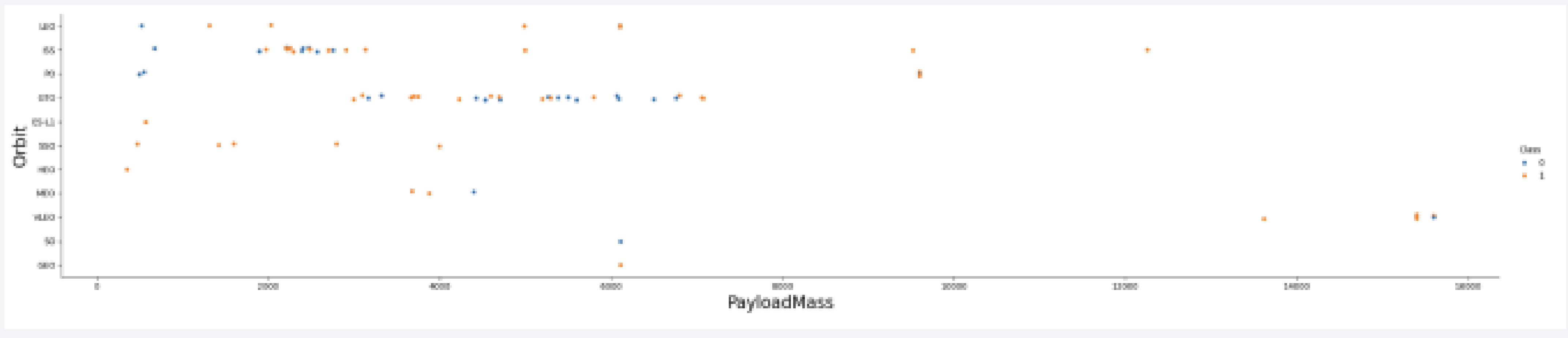


IBM Developer

SKILLS NETWORK

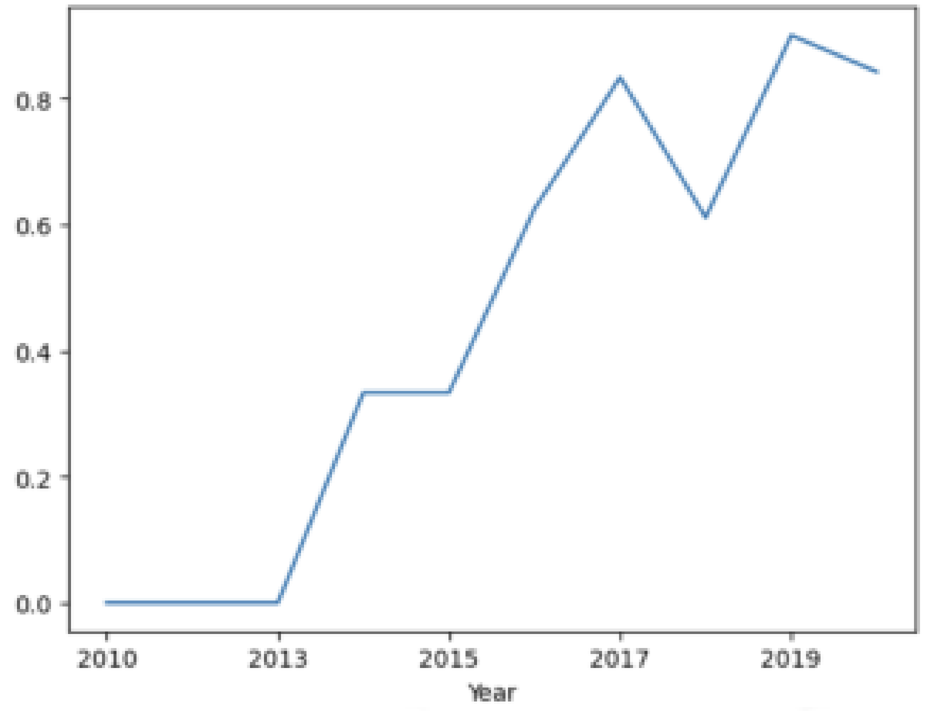
Payload vs. Orbit Type

payload vs. orbit type o No relation



Launch Success Yearly Tren

No improvement at 2010- 2013 but that increases a lot



All Launch Site Names

Four Launch sites

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

Launch Site Names Begin with 'CCA'

5 of these records

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcon
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachut
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachut
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attem
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attem
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attem

Total Payload Mass

the total payload carried by boosters from NASA

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

first successful landing outcome on ground pad

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
sql SELECT MIN(DATE) AS FIRST_SUCCESS_GP FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)';
```

Successful Drone Ship Landing with Payload between 4000 and 6000

boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

```
sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING__OUTCOME =
```

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes

Mission_Outcome	QTY
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

Booster which have carried the maximum payload mass

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

2015 Launch Records

The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

```
sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND DATE_PART(
```

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

Task 10

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.

```
sql SELECT LANDING__OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LA
```

```
* sqlite:///my_data1.db
```

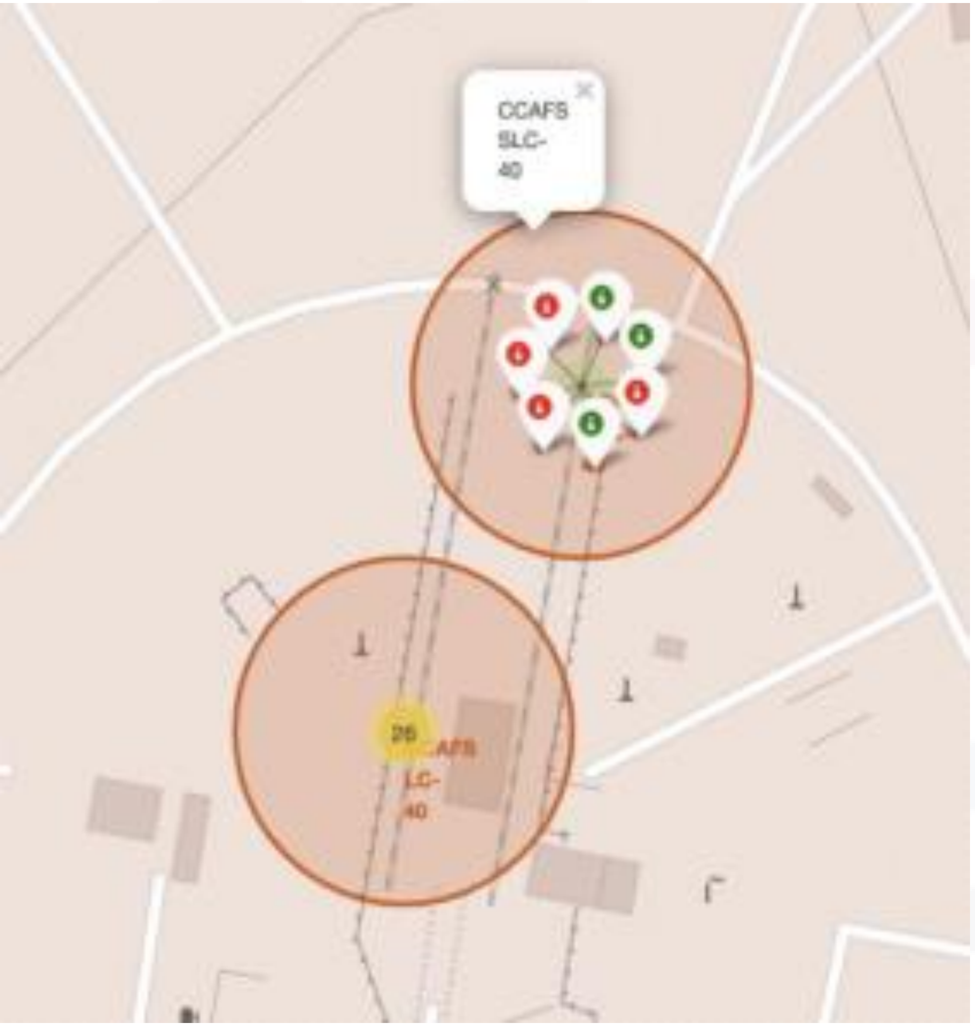
Part 3

Launch Sites Proximities Analysis

Launch Sites

Sites are on the coasts

Launch Successes and Fails



Distance from Site to Proximity

Part 4

Build a Dashboard with Plotly Dash

DASHBOARD TAB 1

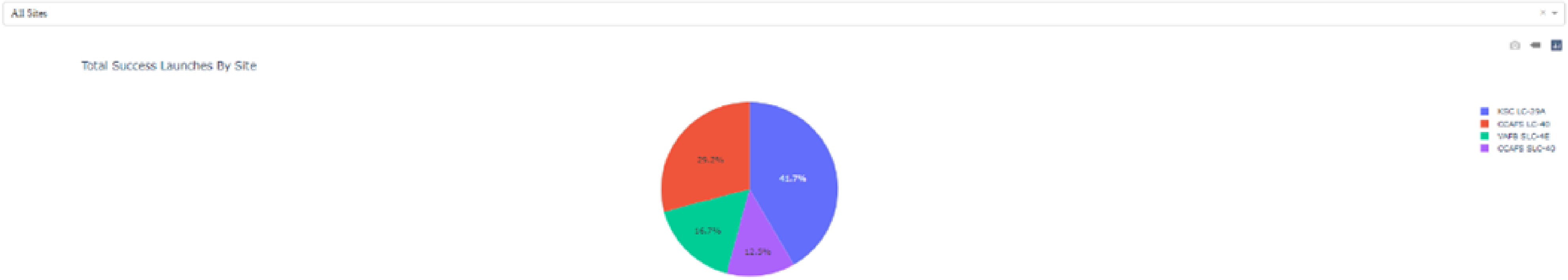
SpaceX Launch Records Dashboard

All Sites

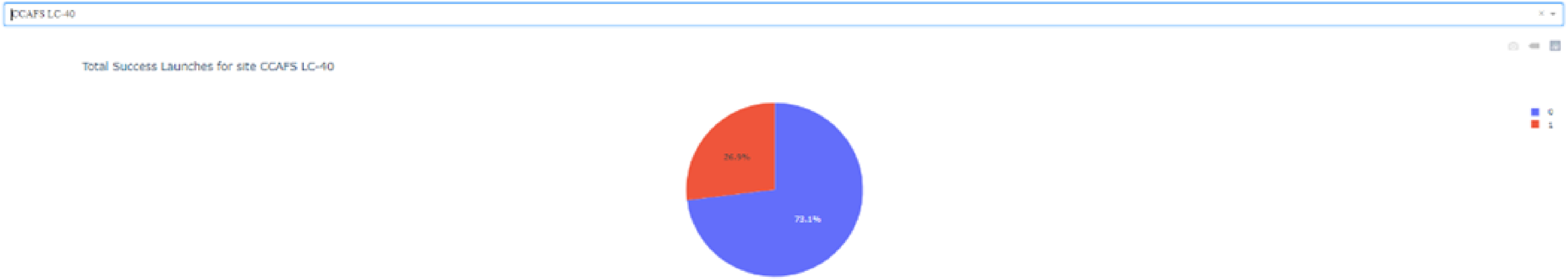
- All Sites
- CCAFS LC-40
- VAFB SLC-4E
- KSC LC-39A
- CCAFS SLC-40

DASHBOARD TAB 2

- Pie chart for all sites are selected



- Pie chart for is selected

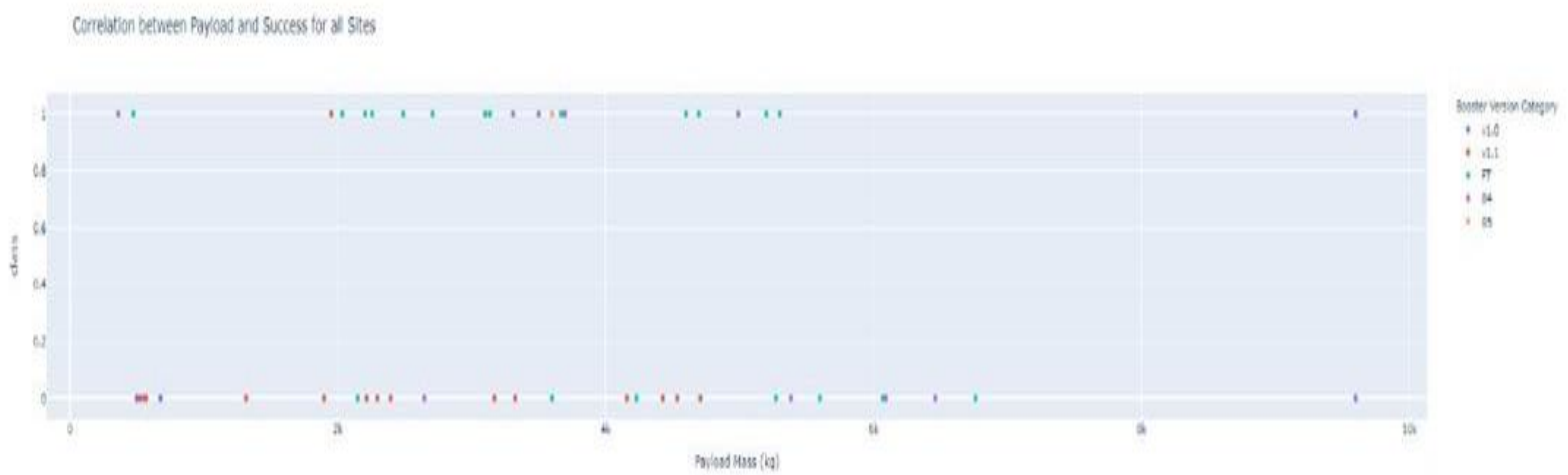


IBM Developer

SKILLS NETWORK

DASHBOARD TAB 3

DASHBOARD TAB 4



Part 5

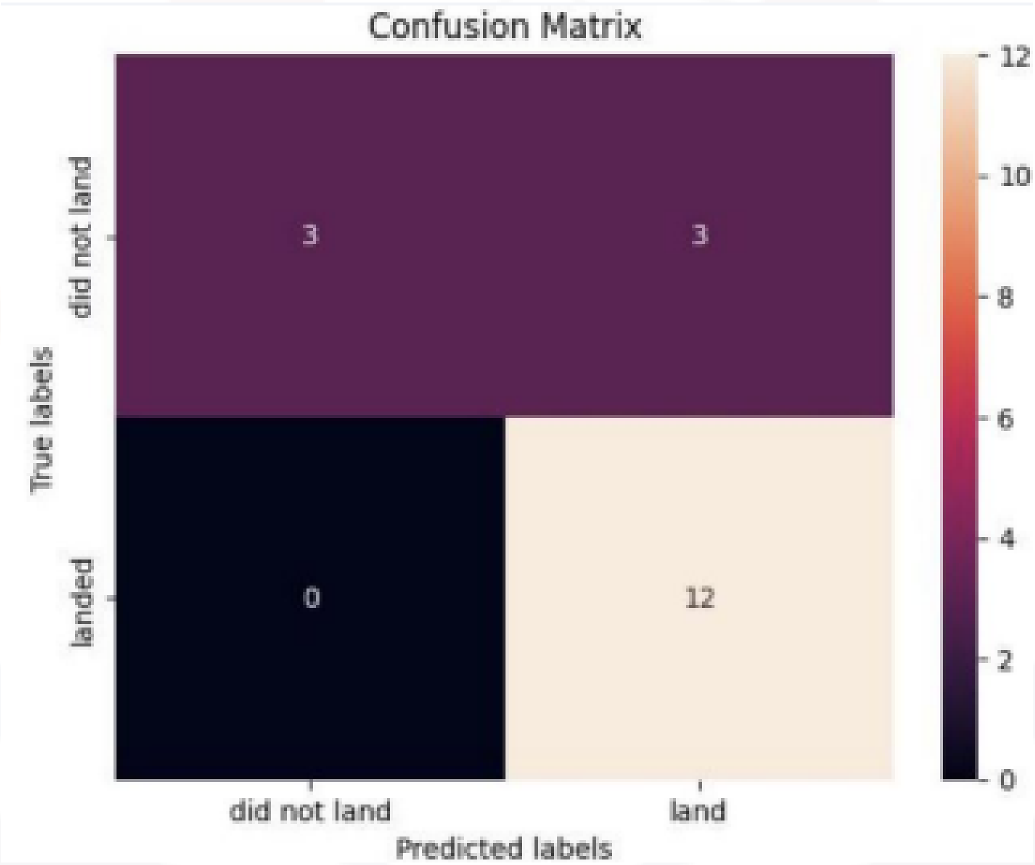
Predictive Analysis(Classification)

Classification Accuracy

They were all equal

Logistic_Reg	0.8333
SVM	0.8333
Decision Tree	0.8333
KNN	0.8333

Confusion Matrix



CONCLUSION

- Data was from SpaceX API and Wikipedia
- KSC LC-39A was the strongest site
- Launches under 7,000kg has more danger
- Over time, the data improved
- Sites are located near coasts; ensure less things in proximity
- More tests and algorithms are recommended for more accurate conclusions

APPENDIX

Numpy

Pandas

Matplotlib

Bye, thank you.