PROJECT SPACE Y

RAHUL STANLY KEECHERIL

07-04-2023

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- The following methodologies were used to analyze the data
  - Data Collection using web scraping and SPACEX API
  - Exploratory Data Analysis
  - Machine Learning
- Valuable data was collected using public data sources
- ML was used to find the best prediction model

INTRODUCTION

- The main purpose of this project is to understand how a new company SpaceY could compete against SpaceX
- Where to launch rockets from?
- How costly it could be?

METHODOLOGY

- Data Collection: Webscraping and SpaceX API
- Data Wrangling
- Perform EDA using SQL and Data visualization
- Apply Machine Learning
  - Data was normalized and divided into training and testing sets
  - Different models were used and their scores were assessed to decide the most accurate model

Data Collection and Data Wrangling

- Data was collected using web scraping and the public use API for SpaceX
- Data wrangling was done and labels were assigned to determine if the launch and landing were successful or not

EDA and interactive visual analytics methodology

- EDA(Exploratory Data Analysis) was done using SQL and data visualization tools like matplotlib and Folium
- SQL queries were done for the following,
  - Names of the unique launch sites in the space mission
  - Average payload mass carried by booster version F9 v1.1
  - The names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000
  - The total number of successful and failure mission outcomes
  - The names of the booster_versions which have carried the maximum payload mass.
  - 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,

Folium for Data Visualization

- Folium was used for creating maps with the locations of each launch site, the number of launches and to know whether it was a success or not.

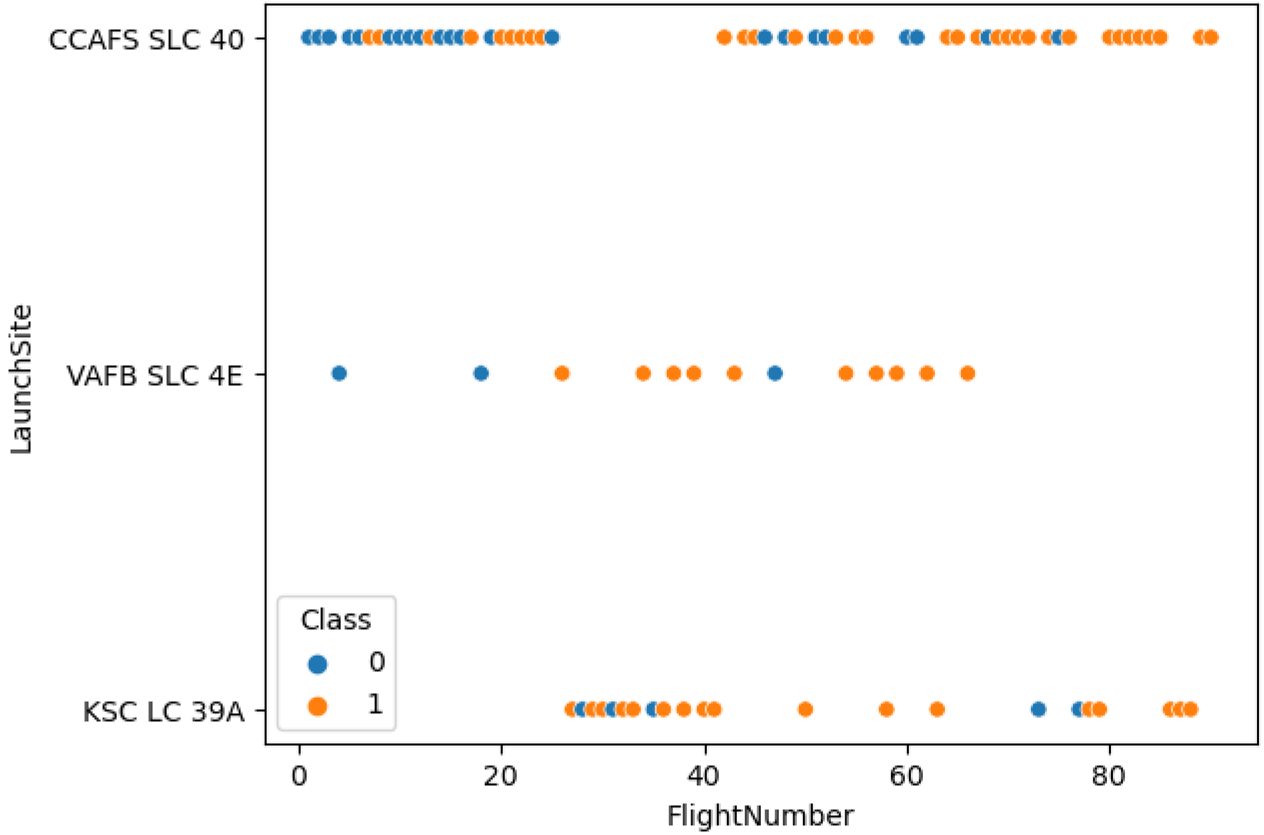
Predictive analysis

- 4 Classification models, namely logistic regression, support vector machines(SVM), decision tree and k nearest neighbors(KNN) were created and their accuracies were compared

Results

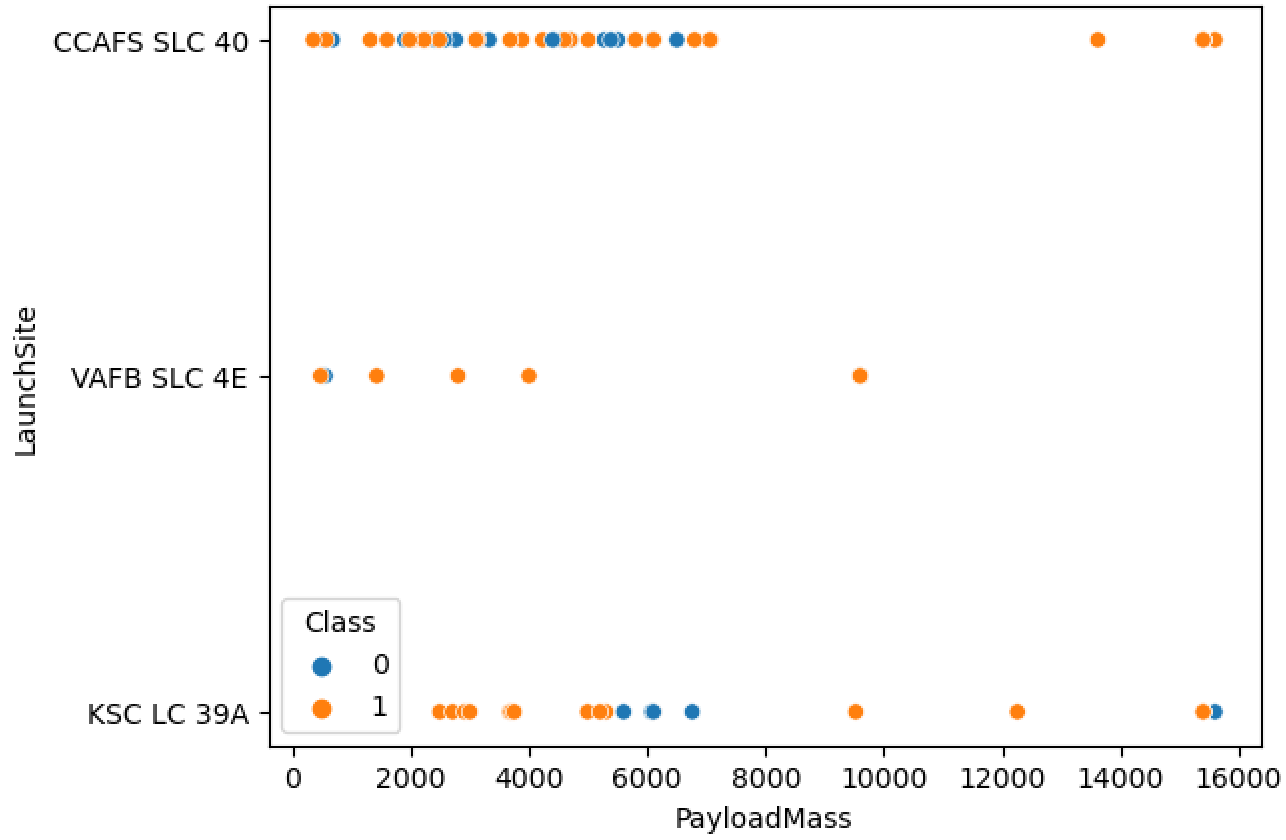
- SpaceX uses 4 different launch sites
- As the Flight number increases, the success rate also increases.
- Among different types of orbits, ES-L1, GEO, HEO, and SSO have the highest success rates.
- Success rates have reached as high as 0.87 during the year 2019.

Flight Number vs Launch Site



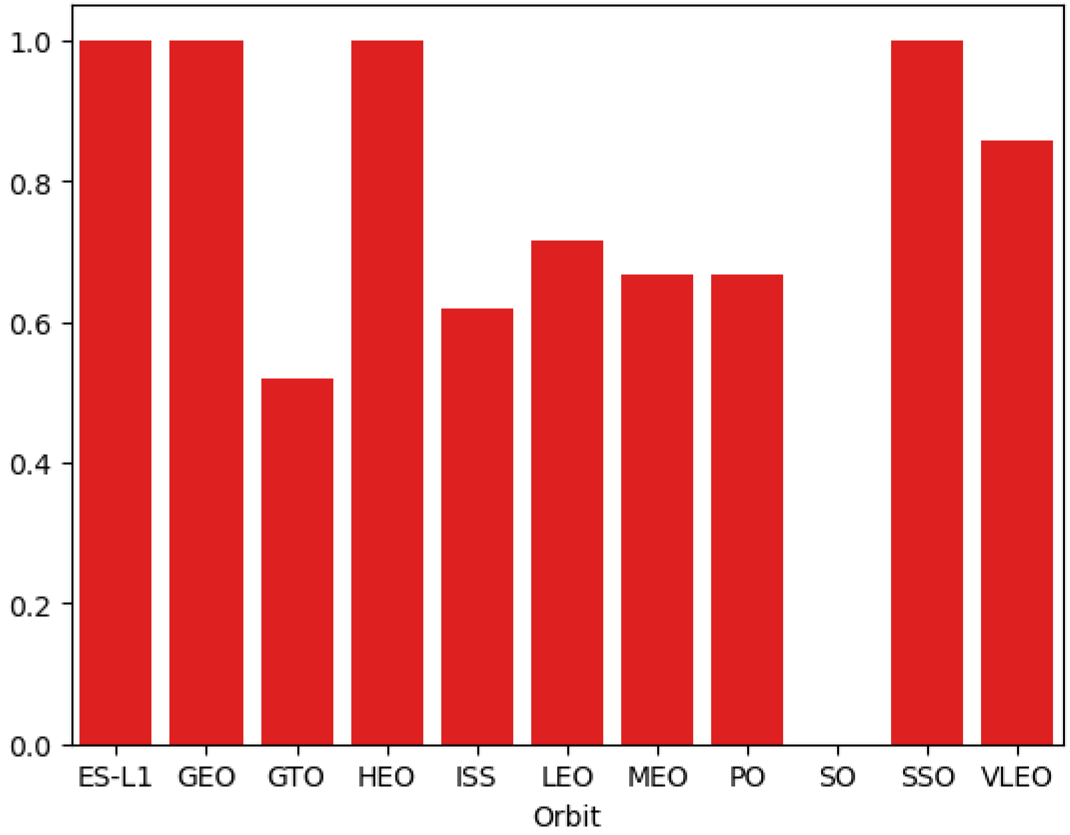
- We can see that the overall success rate of each site has improved over time.
- We can also see that the launch site CCAFS SLC 40 has had the most number of launches and that the success rate has improved after many launches.

Payload Mass vs Launch Site



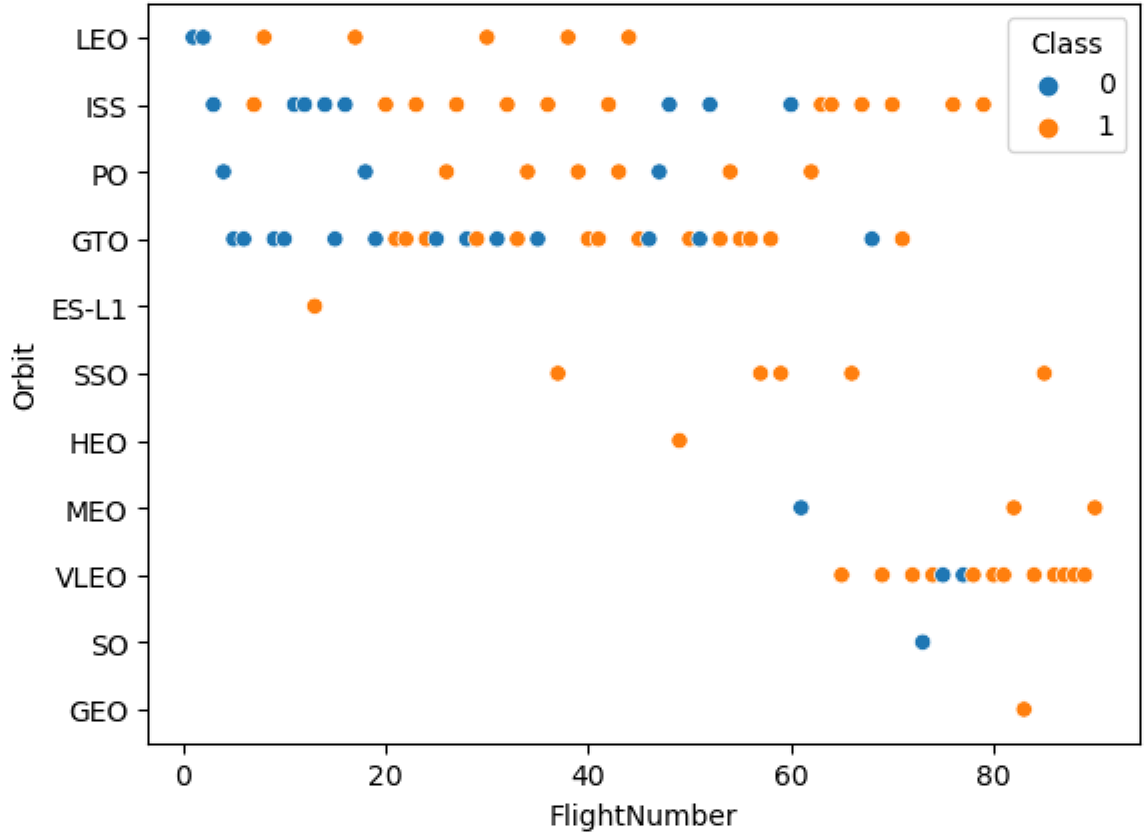
- The success rate for CCAFS SLC 40 is highest for huge payload masses compared to other launch sites

Orbit vs Success Rate



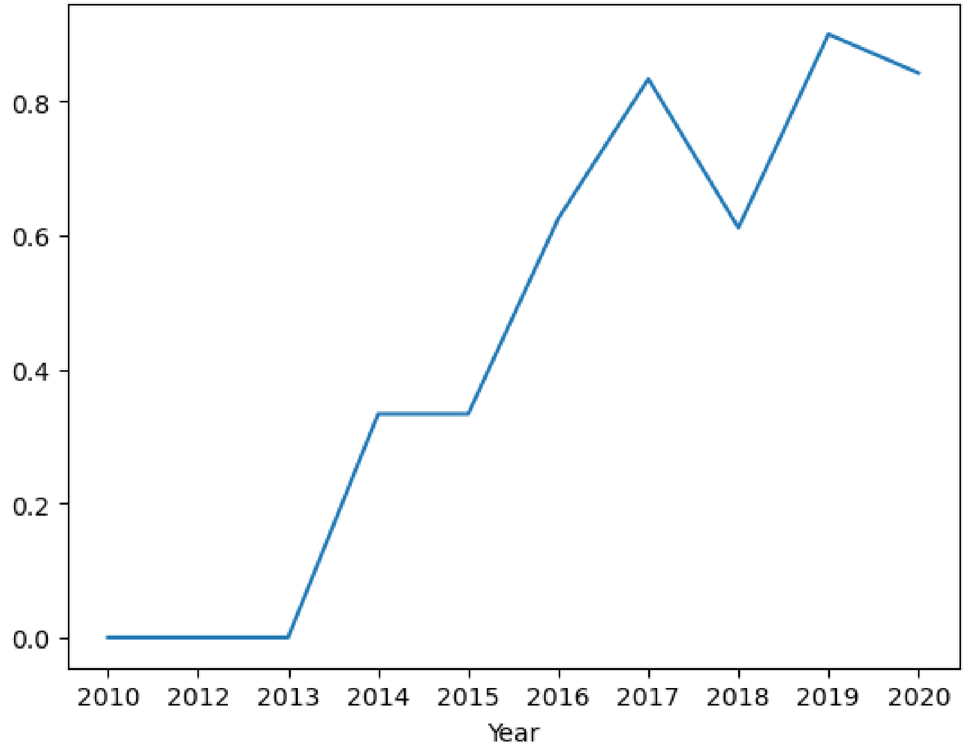
- Success rates are highest for ES-L1, GEO, HEO, SSO orbits.

Flight Number vs Orbit Type



- The number of flights is directly related to the success rate of LEO orbit and there seems to be no relation between GTO orbit and number of flights

Success rate vs Year



- Success Rate is directly related to year
- Peaked in 2019

Launch Site Names

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

- There are 4 different launch sites

Launch Sites beginning with CCA

DATE	time_utc	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- 5 Launch Sites with beginning with CCA and their information

Total payload mass for Boosters launched by NASA and average payload mass of booster version F9 v1.1

48213

- Shows the total payload mass of boosters launched by NASA

2534

- Shows the average payload mass of booster F9 v1.1

SQL queries

2016-06-05

- Date at which first successful landing outcome in ground pad was achieved

booster_version
F9 FT B1022
F9 FT B1031.2

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

Sql Queries contd.

mission_outcome	TOTAL NUMBER
Success	44
Success (payload status unclear)	1

- Total number of successful and failure mission outcomes

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1058.3
F9 B5 B1060.2

- Names of the booster versions which have carried the maximum payload mass

SQL Queries contd.

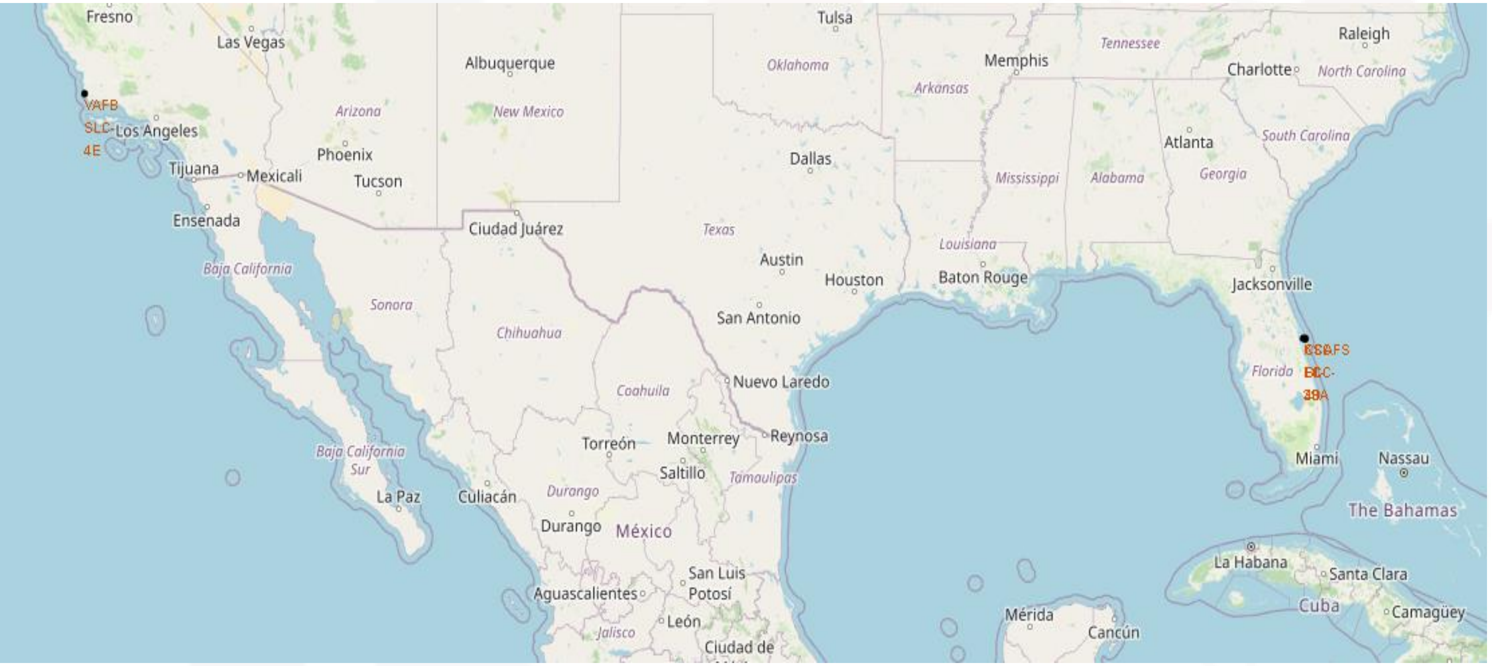
booster_version	landing_outcome	launch_site
F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40

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

landing_outcome	count_of_landing_outcomes
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

- The count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the dates 2010-06-04 and 2017-03-20, in descending order

Folium: Location of Launch Sites

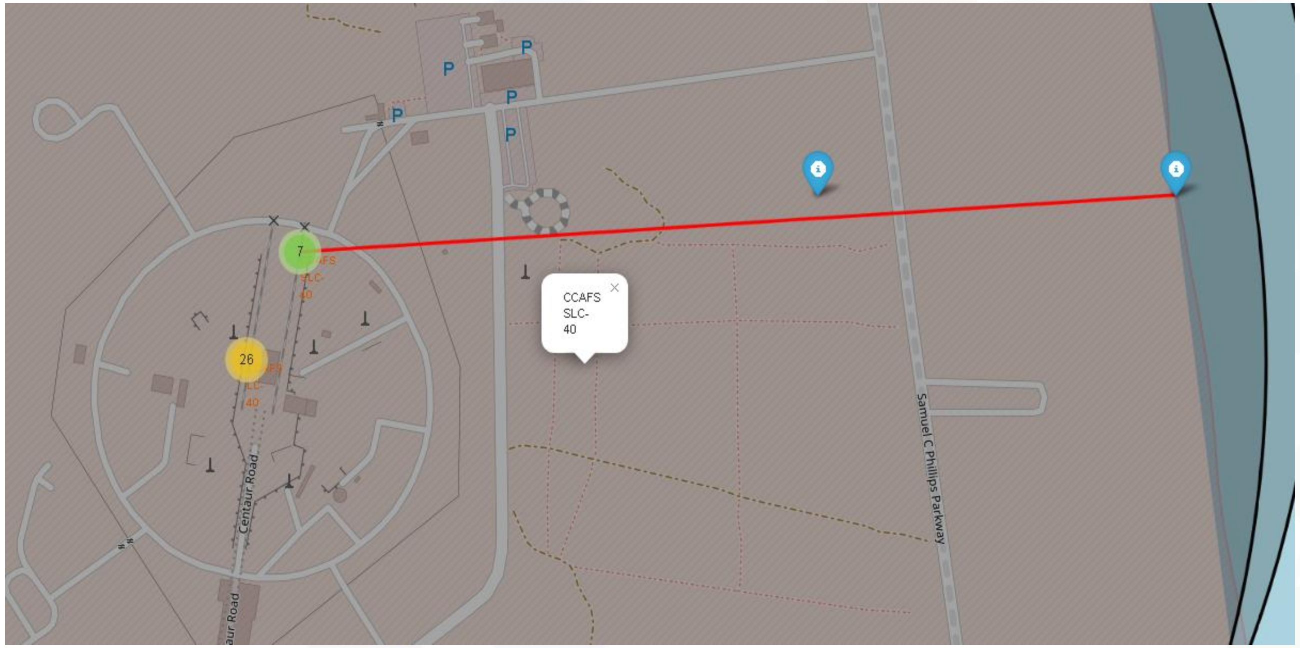


- Launch Sites are located close to the equator and coastal line

Folium: Landing Outcomes

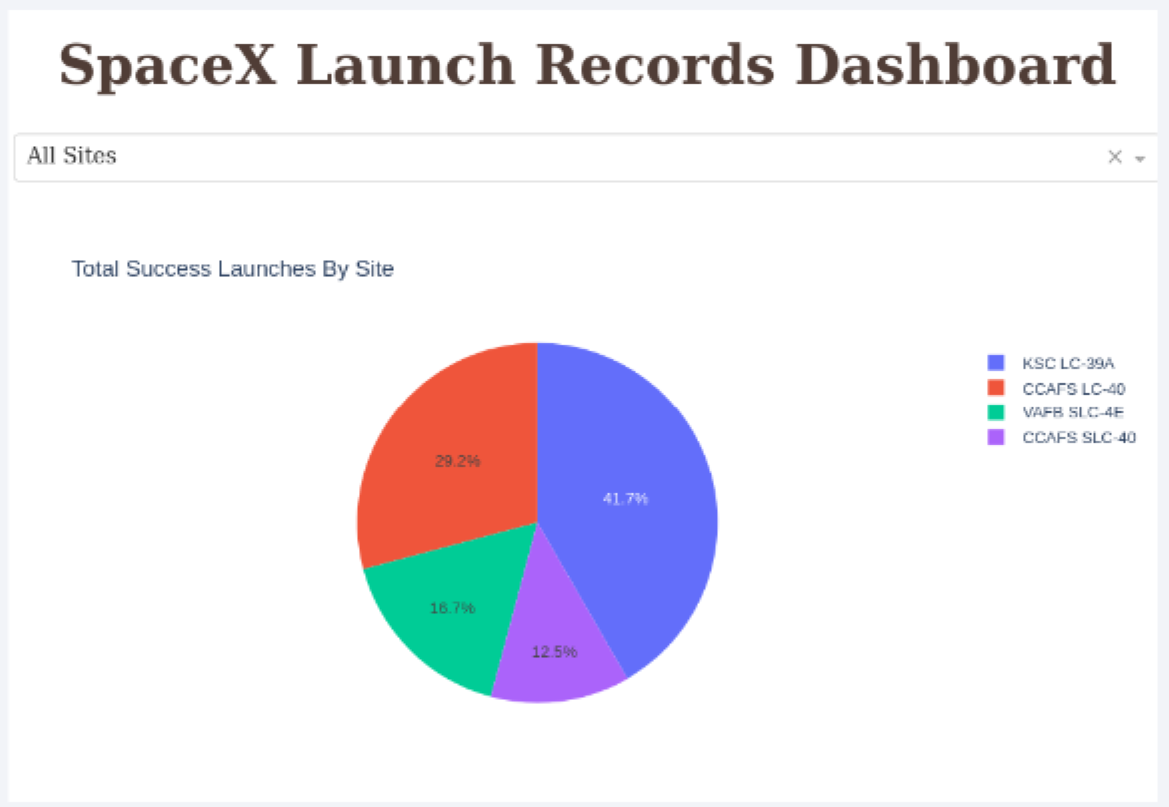
- Green markers indicate successful launches and red indicates failure.

Folium: Distance to Coast



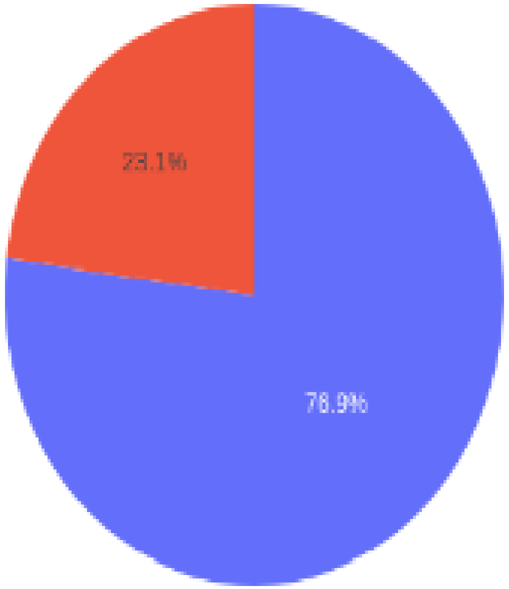
- The red line shows the distance between the launch site and the coast.

Dashboard Results



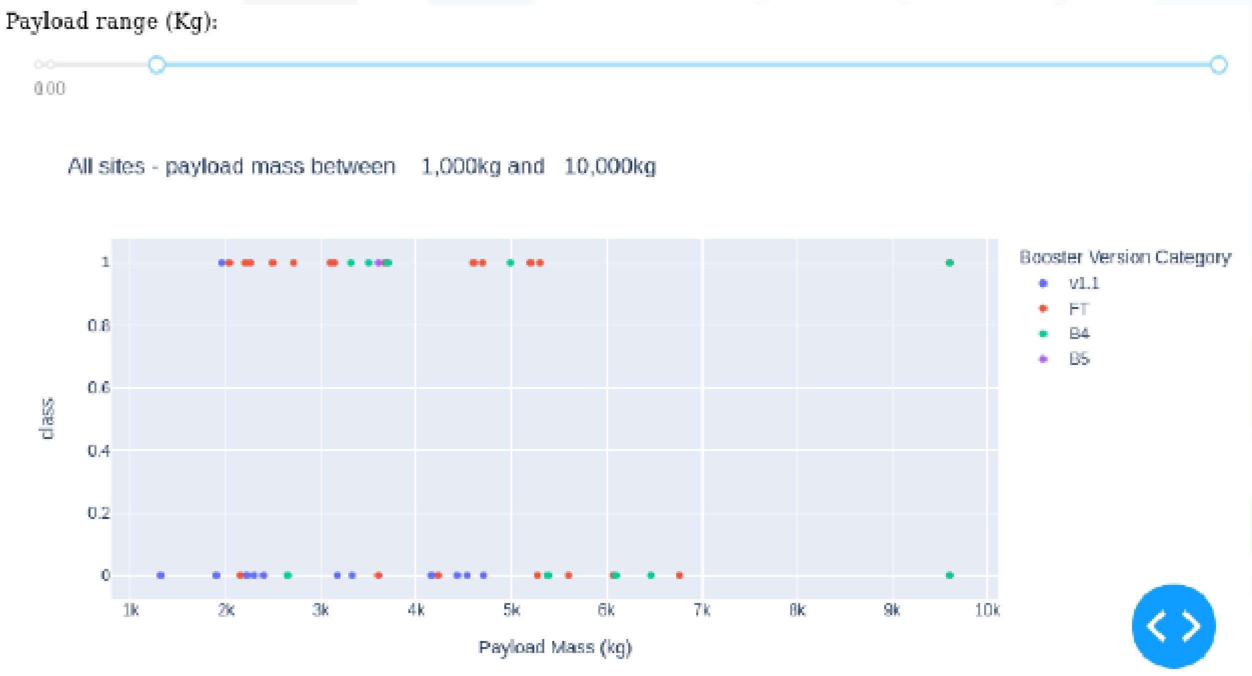
- The places from which the launching is done seems to be an important factor for launch outcome

Success rate for KSC LC-29A



- Has a high success rate of 76.9

Payload vs Launch outcome



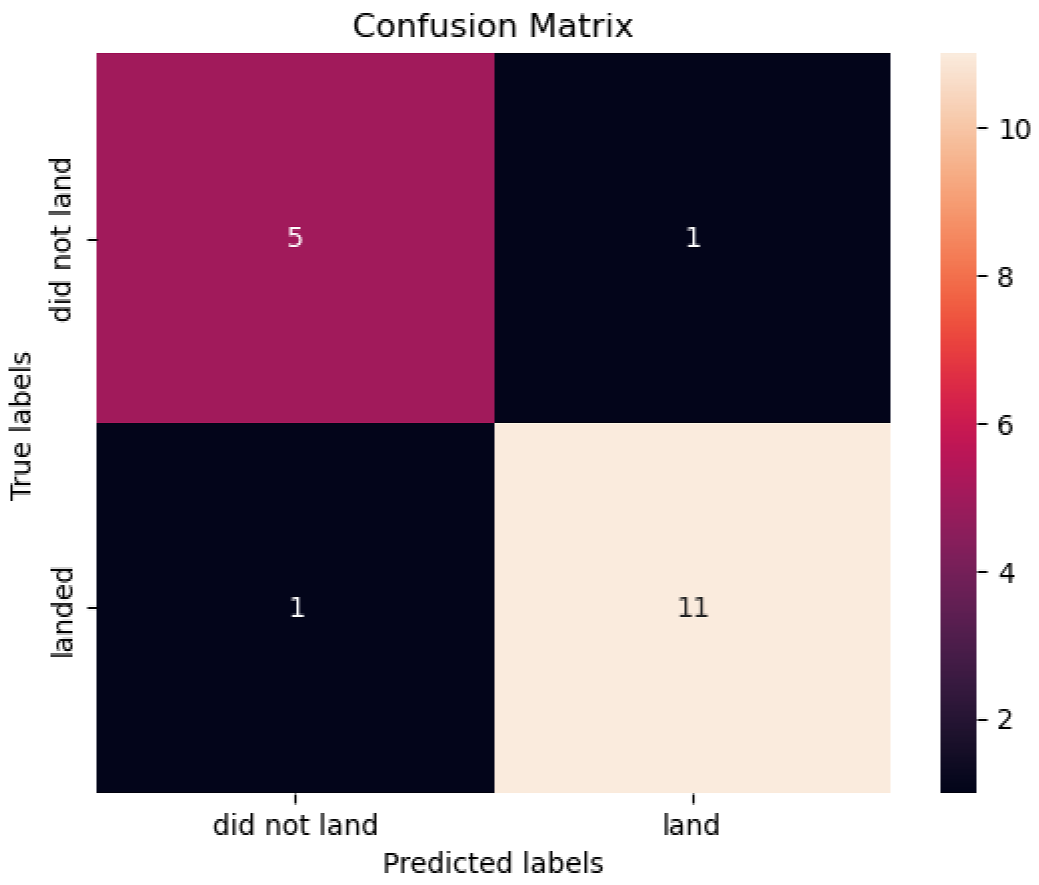
- Payloads under 6000kg and FT boosters are the most successful combination

Results: Machine Learning

```
Decision Tree accuracy 0.8888888888888888
SVM accuracy 0.8333333333333334
KNN accuracy 0.8333333333333334
Logistic Regression accuracy 0.8333333333333334
```

- Scores of different machine learning models were computed and it can be seen that decision trees have the best score of 89.9% accuracy

Confusion Matrix for Decision Tree



- Shows the confusion matrix to verify the accuracy of the model.

Conclusions

- Data was collected from many sources and analysed to provide meaningful insights
- Decision tree model can be used to predict landing outcomes more accurately
- Landing outcome success rate increases as number of flights increases