Winning Space Race with Data Science

Aatis Sajid

25-09-2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- In this capstone project, we will predict if the SpaceX Falcon 9 first stage will land successfully
- If we can determine if the first stage will land, we can determine the cost of a launch.
- This will be achieved with the use of machine learning algorithms
- The methodolgy followed will include Data Collection, Data Wrangling and Preprocessing, Exploratory Data Analysis, Data Visualization and finally, Machine Learning Prediction.

INTRODUCTION

The main goal of this capstone project is to predict whether the Falcon 9 first stage will land successfully. SpaceX prides itself in being able to reuse the first stage of a rocket launch so much so that they advertise on their website that their rocket launches cost 62 million while other provides cost upward 165 million. Much of these savings are down to the first stage's reusability. If we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

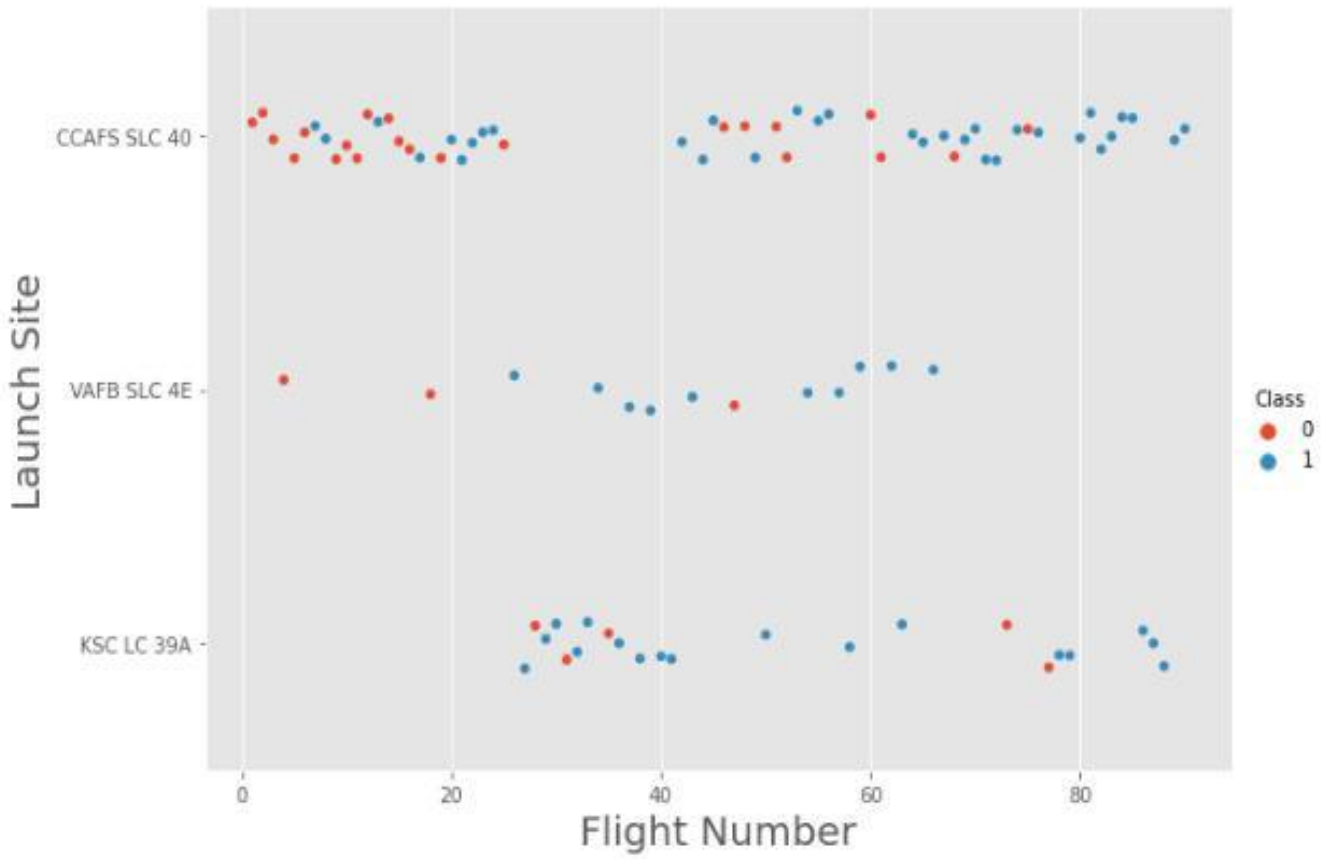
METHODOLOGY

- Data was collected through two methods:-
  - requesting data from the SpaceX API and web scraping launch data from a Wikipedia page.
  - Data wrangling was then performed to transform and clean the data using Python's pandas library.
- Exploratory Data Analysis (EDA) was performed using visualization tools such as Python's matplotlib and seaborn libraries.
- Python's interactive visualization packages were used to answer some analytical questions. Packages such as Folium and Plotly Dash.
- Four different machine learning classification models were used for the predictive analysis. Themodelsthat were used are logistic regression, support vector machines, k-nearest neighbour anddecisiontreeclassifier. Each model was trained, tuned and evaluated to find the best one.

RESULTS

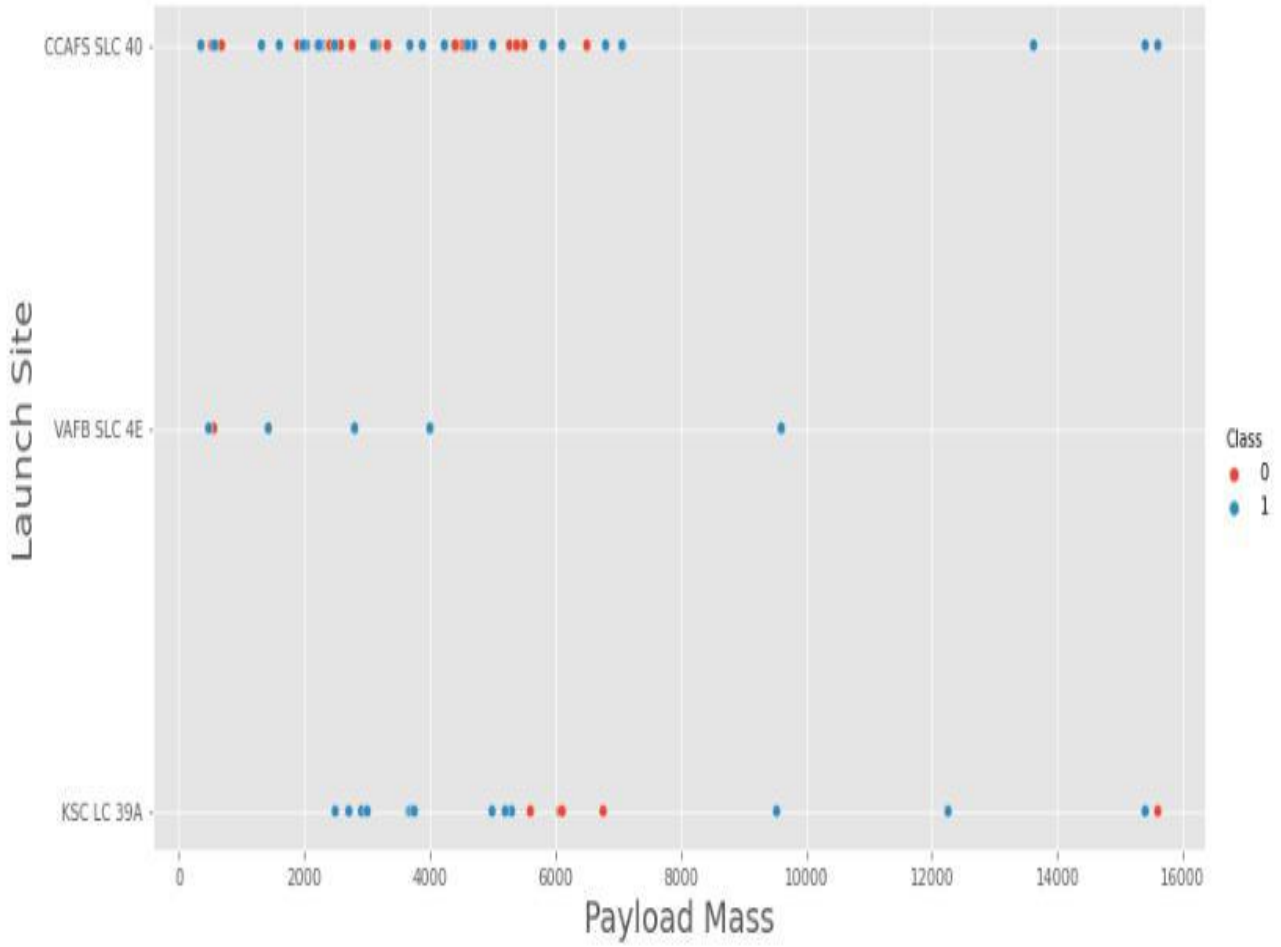
- The results of the exploratory data analysis revealed that the success rate of the Falcon9 landings was 66.66%.
- The predictive analysis results showed that the Decision Tree algorithm was the best classification method with an accuracy of 94%

Flight Number vs. Launch Site



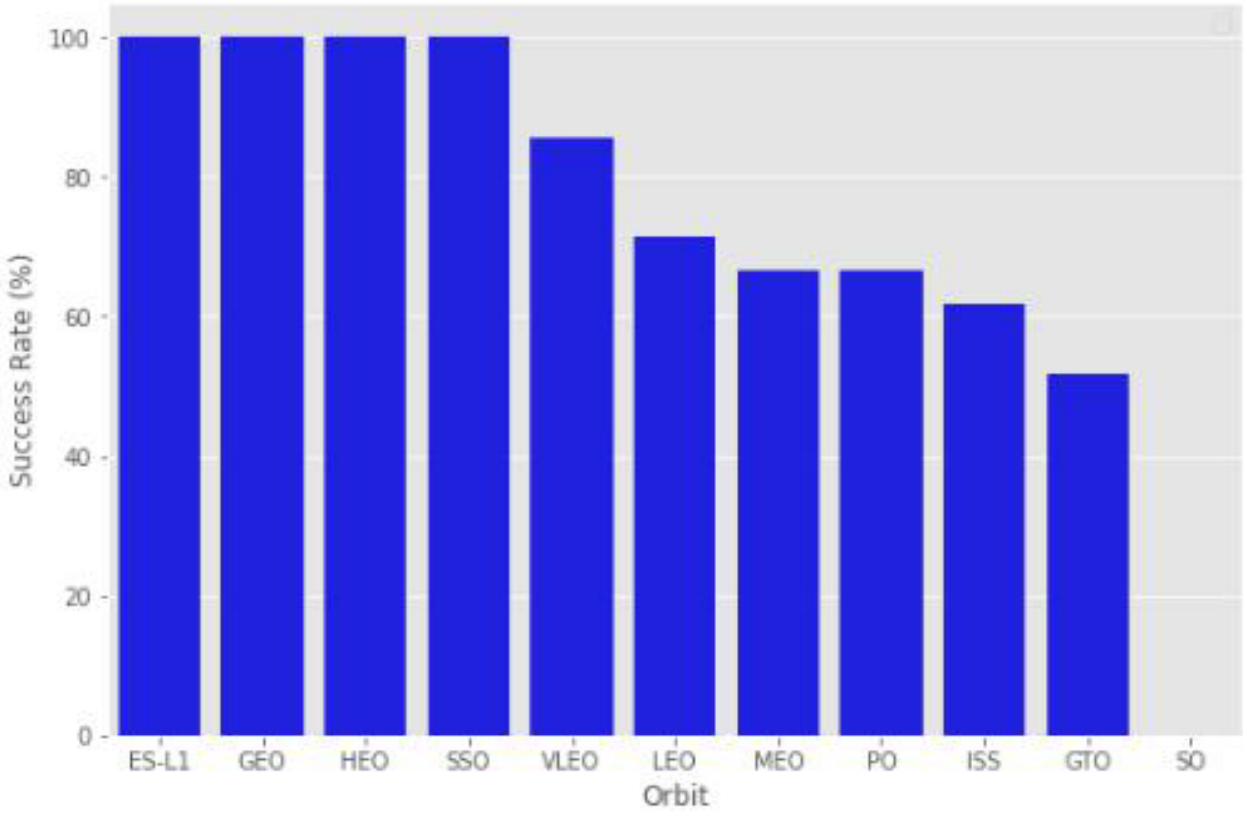
- This figure shows that the success rate increased as the number of flights increased.
- The blue dots represent the successful launches while the red dot represent unsuccessful launches.
- There seems to be an increase in successful flights after the 40th launch.

Payload vs. Launch Site



- The blue dots represent the successful launches while the red dots represent unsuccessful launches.
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass.
- There seems to be a weak correlation between Payload and Launch Site and therefore decisions cannot be made using this metric.

Success Rate vs. Orbit Type



- Orbits SSO, HEO, GEO, and ES-L1 have 100% success rates.
- SO orbit did not have any successful launches with a 0% success rate.

All Launch Site Names

- The DISTINCT clause was used to return only the unique rows from the launch_site column.
- The names of the launch sites are CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E .

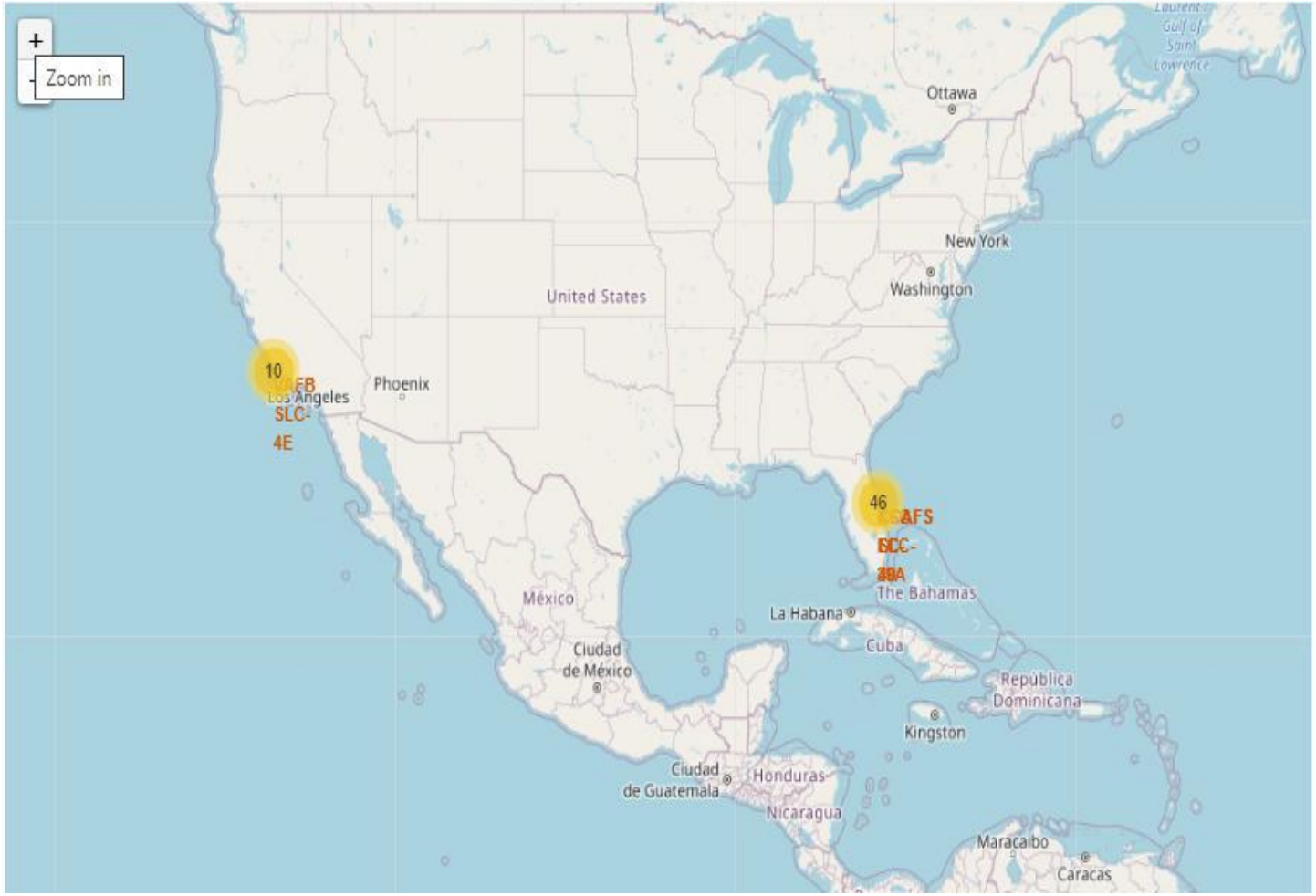
launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Total Payload Mass

- The SUM() function was used to the calculate the total payload carried by boosters from NASA from the total_payload_mass_kg column.

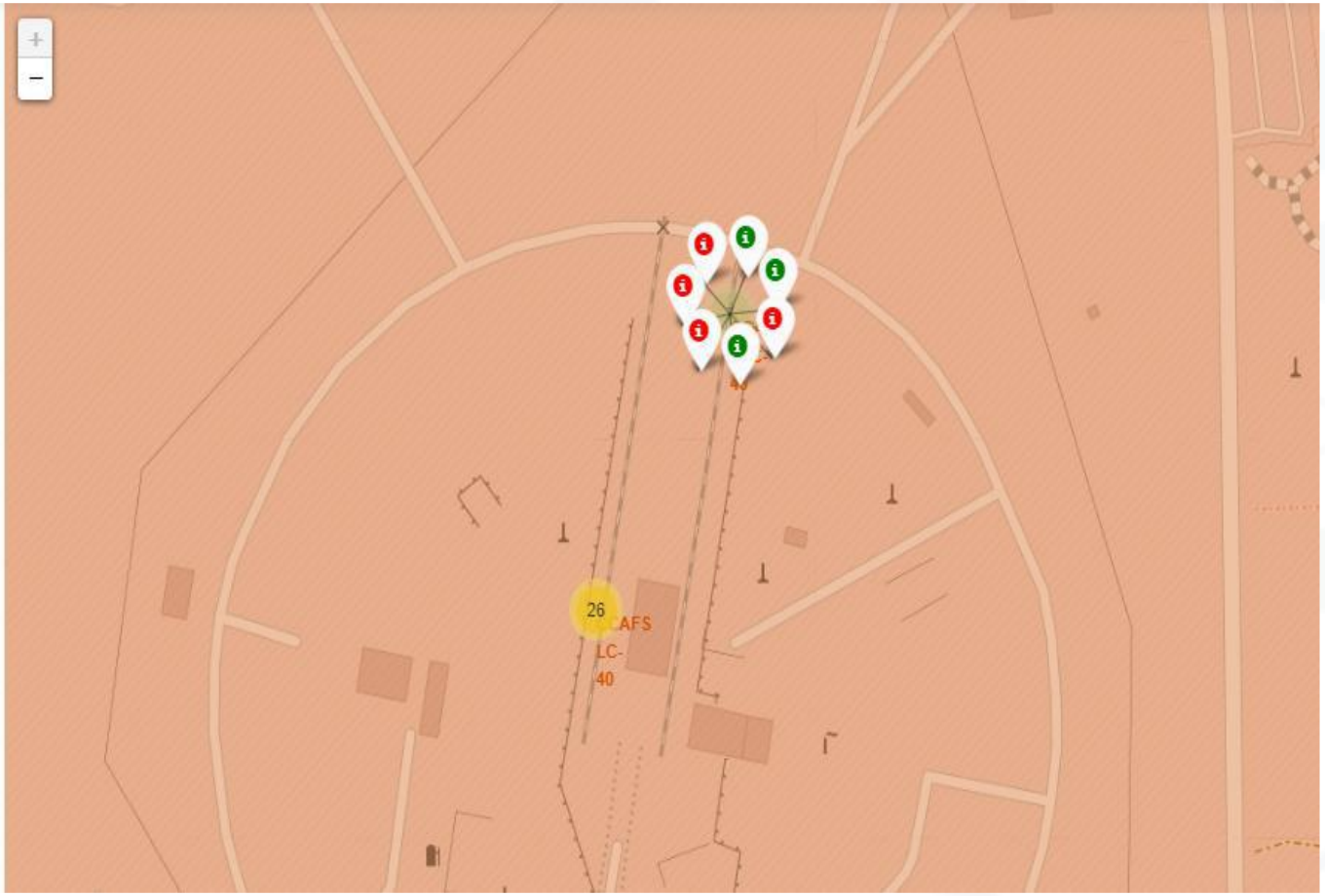
total_payload_mass_kg
45596

SpaceX Launch Sites Locations



- The yellow markers are indicators of where the locations of all the SpaceX launch sites are situated in the US.
- The launch sites have been strategically placed near the coast.

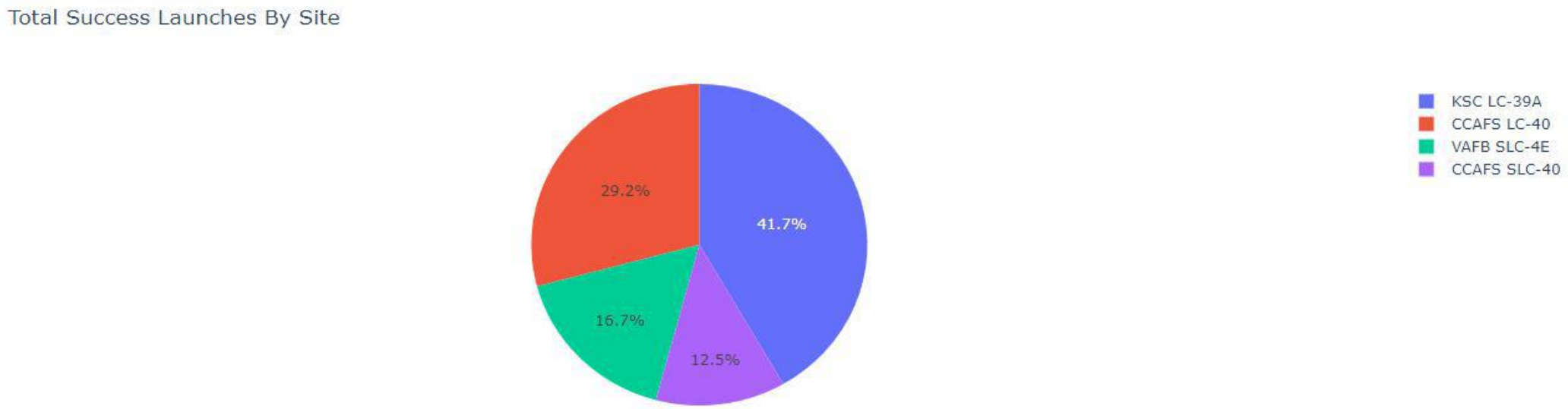
Success or Failure?



- When we zoom in on a launch site, we can click on the launch site which will display marker clusters of successful landings (green) or failed landing (red).

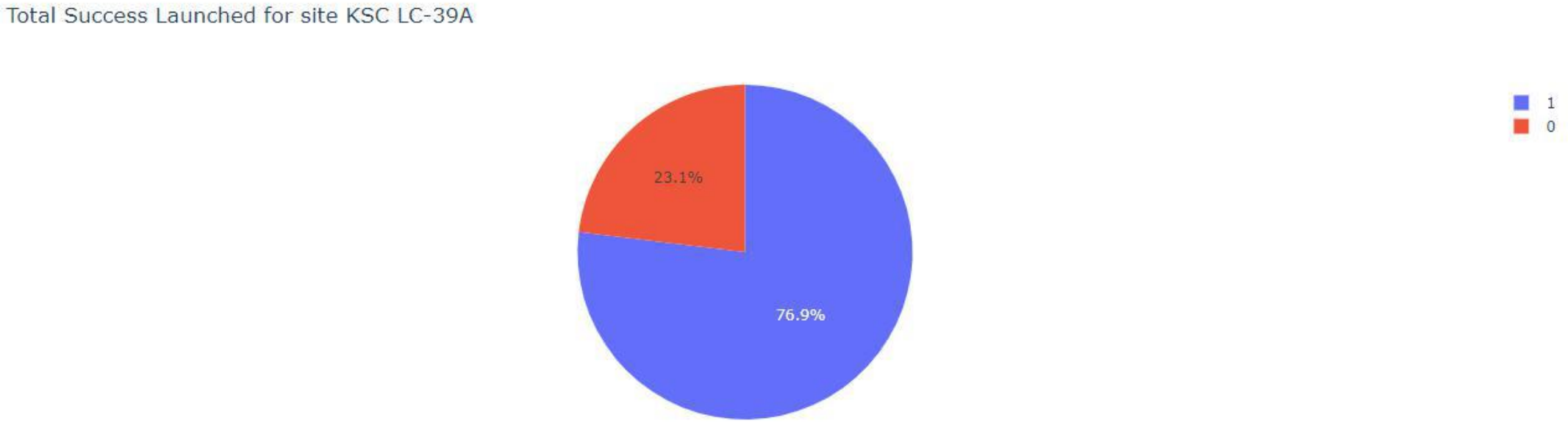
Total Successful Launches By Site

- The KSC LC-39A Launch site has the most successful launches with 10 in total.



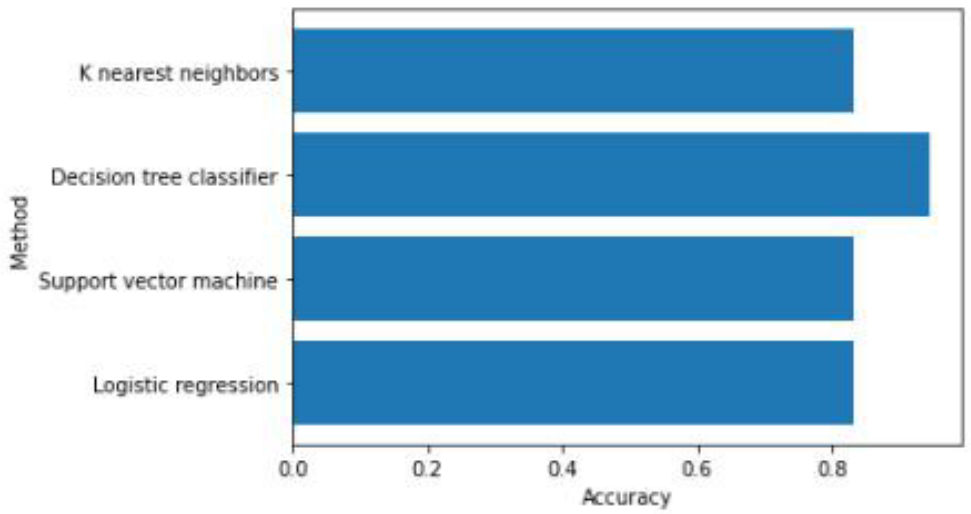
Launch Site With Highest Success Ratio

- The KSLC-39A has the highest success rate with 76.9%.



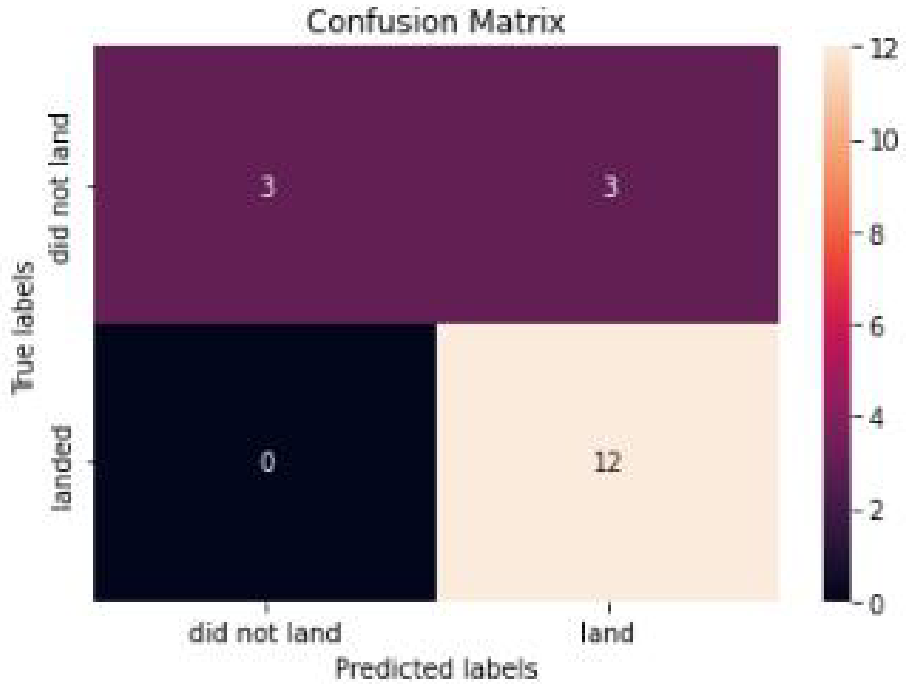
Classification Accuracy

- The **Decision Tree** classifier had the best accuracy at 94%.



	method	accuracy
0	Logistic regression	0.833333
1	Support vector machine	0.833333
2	Decision tree classifier	0.944444
3	K nearest neighbors	0.833333

Confusion Matrix



- The model predicted 12 successful landings when the True label was successful (True Positive) and 3 unsuccessful landings when the True label was failure (True Negative).
- The model also predicted 3 successful landings when the True label was unsuccessful landing (False Positive).
- The model generally predicted successful landings.

CONCLUSION

- The analysis showed that there is a positive correlation between number of flights and success rate as the success rate has improved over the years.
- There are certain orbits like SSO, HEO, GEO, and ES-L1 where launches were the most successful.
- Success rate can be linked to payload mass as the lighter payloads generally proved to be more successful than the heavier payloads.
- The best predictive model to use for this dataset is the Decision Tree Classifier as it had the highest accuracy with 94%.

APPENDIX

- Coursera Project Link: https://www.coursera.org/learn/applied-data-science-capstone/home/welcome
- GitHub Repository: https://github.com/aatissajid/IBM-Data-Science-Capstone-Project

Q & A

THANK YOU