Winning The Space Race with Data Science

James Allen

27/07/2024

IBM

IBM SkillsBuild

OUTLINE

Executive Summary

Introduction

Methodology

Results

Discussion

Conclusion

Appendix

EXECUTIVE SUMMARY

- **Summary of Methodologies:**
- Data collection using webscraping & rest API queries
- Data wrangling to classify launches based on success and failure. Transform data into standardised numeric form
- Exploratory data analysis using SQL & visualisation packages for python.
- Interactive Plotly Dash web app to visualise payload and successful launch data at each launch site
- Exploring launch sites using interactive folium maps
- Predictive analysis for classification of rocket landing success
- **Summary of all results:**
- Exploratory data analysis results
- Predictive analysis results

INTRODUCTION

SpaceX advertises Falcon 9 rocket launches on its website at a cost of 62 million dollars per launch. Other rocket launching companies cost upwards of 165 million dollars per launch. SpaceX saving are due to the nature of their launches because a successful landing means they can reuse a lot of material. Therefore if we can determine if a launch will land we can then determine the cost of the launch. This information can be used by an alternate company who wants to bid against SpaceX as a rocket launching company. It is advantageous to be able to predict whether the Falcon 9 will land successfully for new missions.

To make valuable predictions we must solve the following:

- What factors of a mission influence Falcon 9 launch success?
- What conditions must be met by SpaceX to ensure the highest probability of success for a given mission?

METHODOLOGY

Data Collection Method:

- Requested past launch data from SpaceX REST API
- Webscraped tabular data on SpaceX launches

Data Wrangling Method:

- Dropped data on non Falcon 9 launches
- Used one hot encoding to transform categorical variables into factors
- Transformed factors into integers
- Replaced missing numerical data for payload mass with the sample mean
- Classified data as 1 for successful landing and 0 for failed landing

- **Explorotory Data Analysis Method:**

- Used scatter plots and bar graphs to visualise relationships between variables
- Used SQL queries to understand the data collected

Methodology Cont.

Interactive Visual Analytics using Folium and Plotly Dash method:

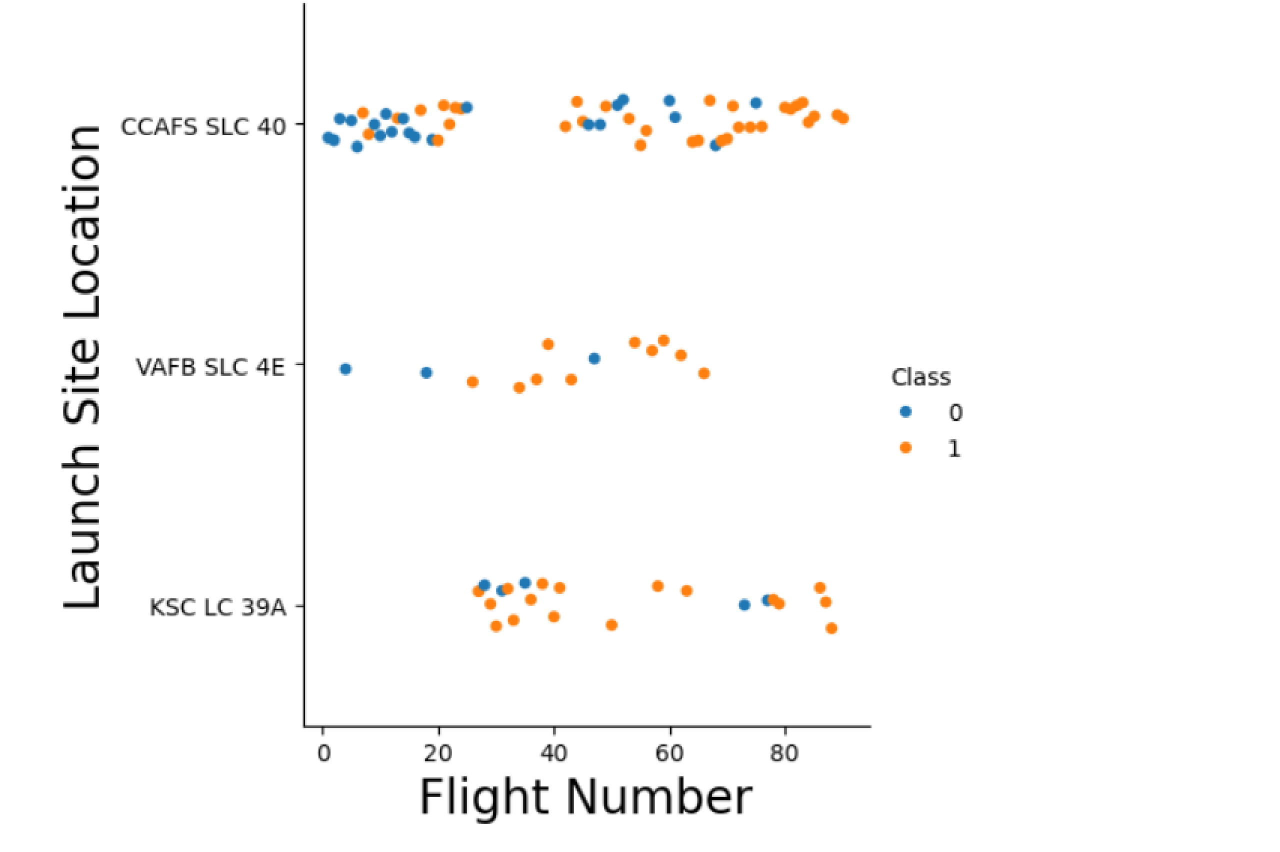
- Created an interactive plotly dash web app to visualise payload and success launch data.
- Exploring launch sites using interactive folium maps.

Predictive Analysis using Classification Models:

- Built and tuned multiple classification models to predict landing success
- Used grid search and cross validation to find the best model parameters for each model tested(Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest Neighbour)

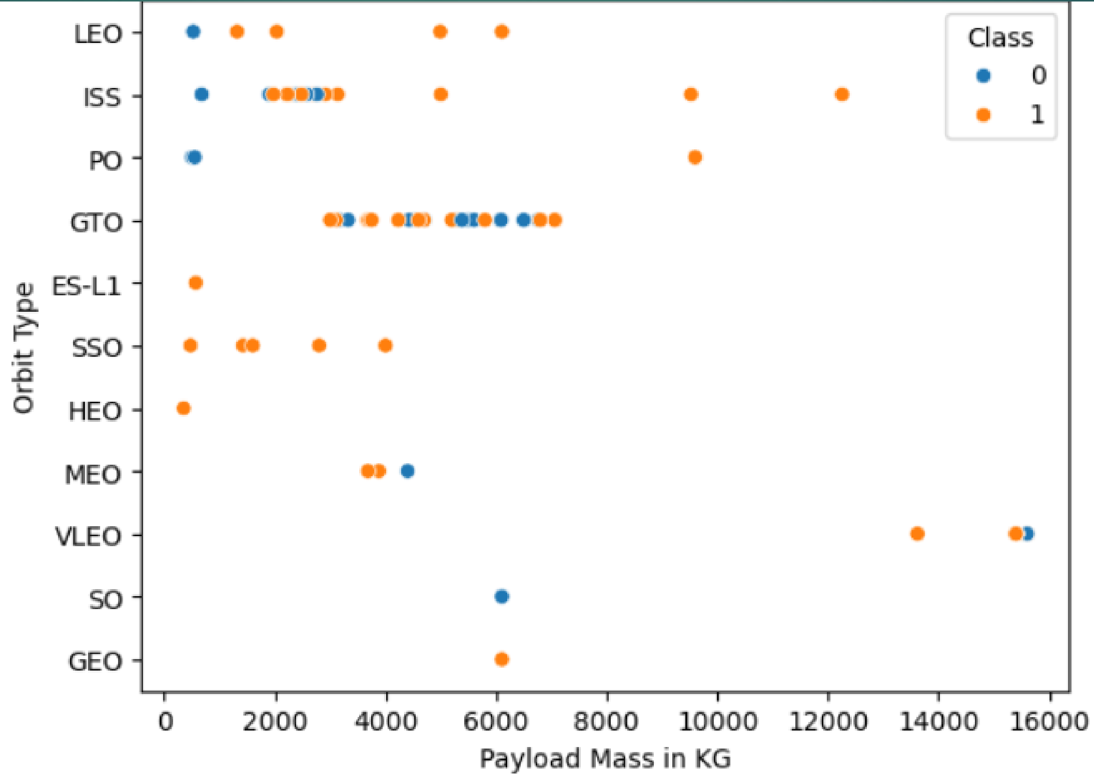
EDA with Visualization Results

This is a scatter plot of successful (1) and unsuccessful (0) landings with the site locations. It is visible during the earlier flight number stages there were a higher count of unsuccessful landings but as time went on the landing success rate is exponentially increasing.



EDA with Visualization Results (2)

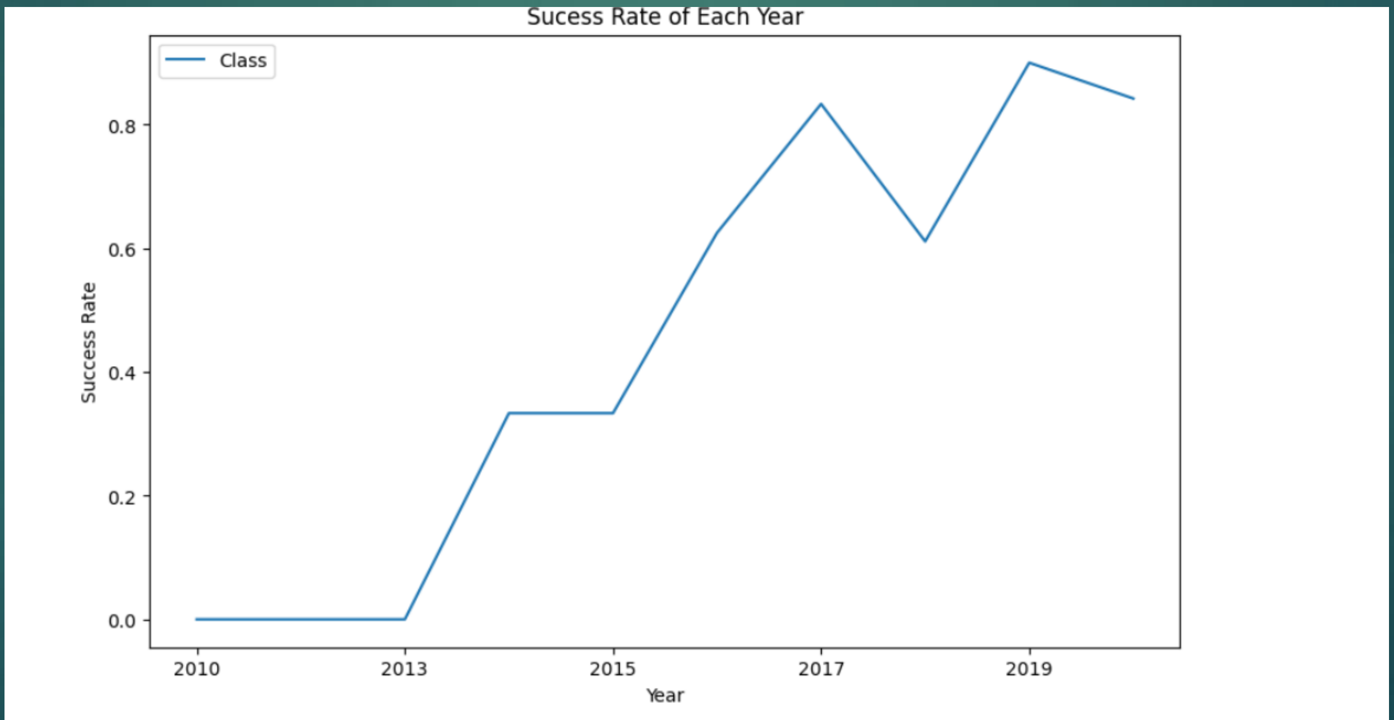
In this scatter plot it shows that the higher the payload mass the greater successful landings for LEO, ISS and PO orbit type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

EDA with Visualization Results (3)

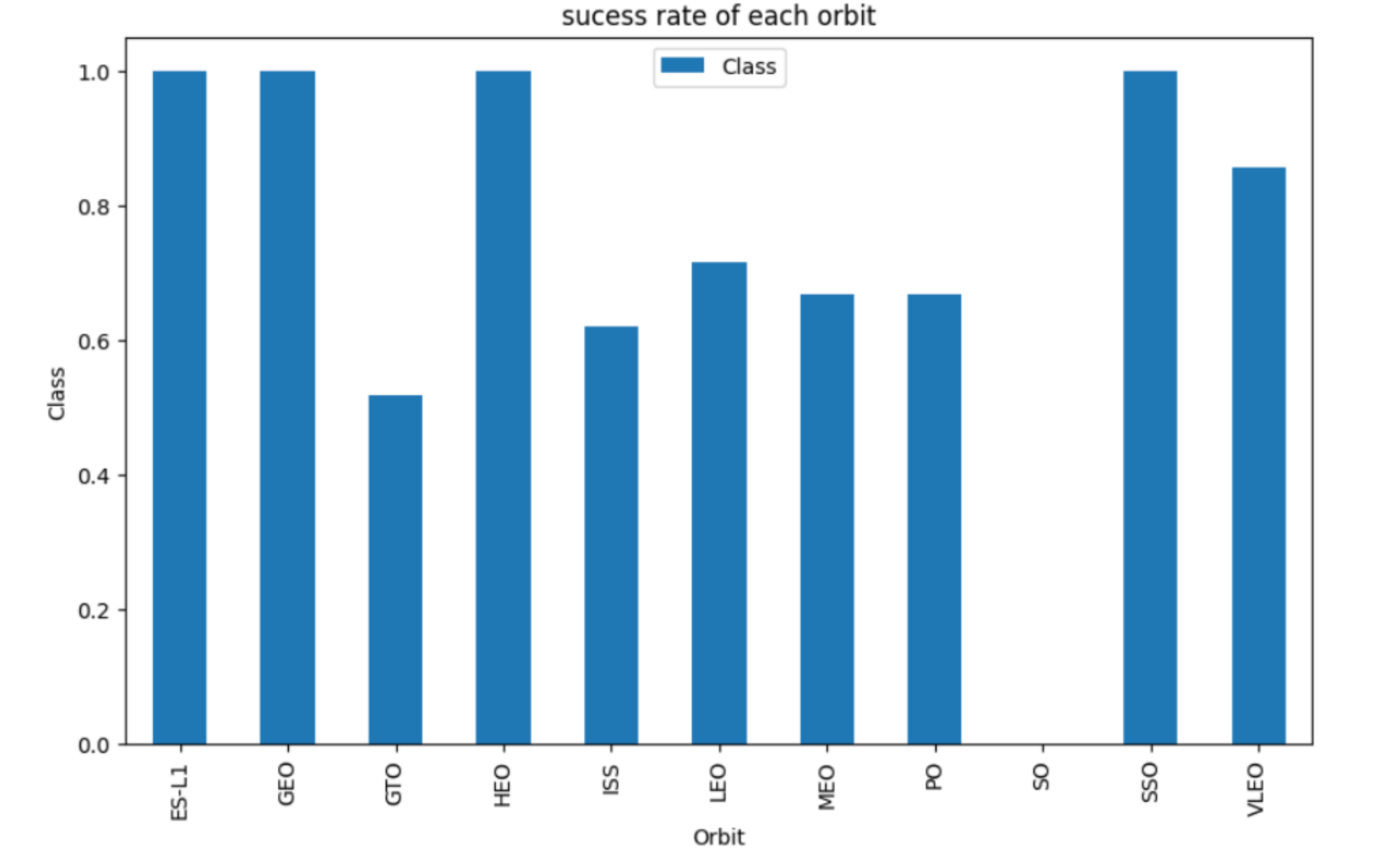
In this line graph it is showing a trend that since 2013 there has been a trend of a higher success rate per year. Except for 2018 when it dipped.



you can observe that the sucess rate since 2013 kept increasing till 2020

EDA with Visualization Results (4)

Here is the success rate of each orbit type. We can see the most successful are ES-L1, GEO and SSO



Analyze the plotted bar chart to identify which orbits have the highest success rates.

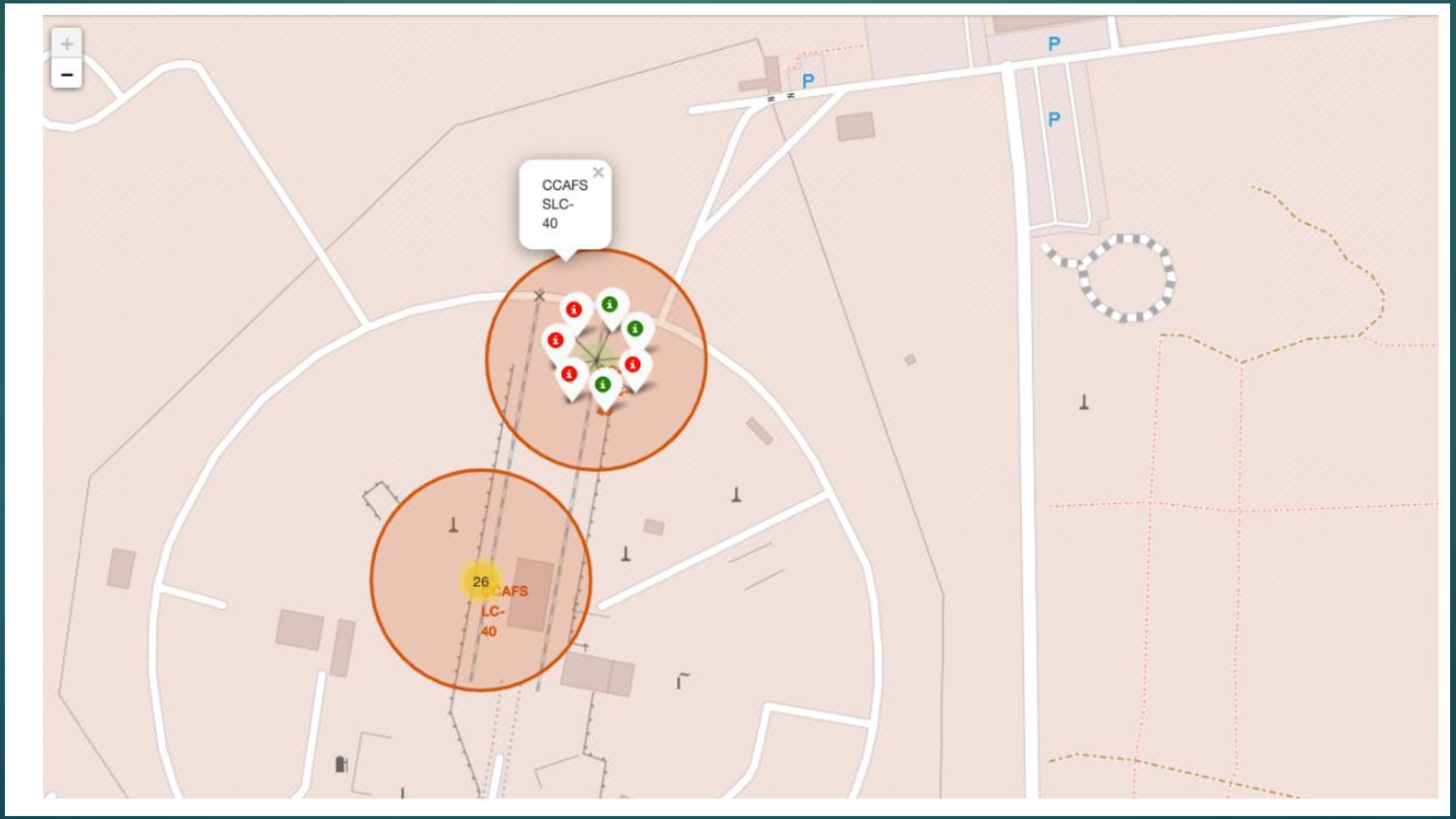
EDA with SQL results

The results here show the count or number of landing outcomes that has occurred throughout SpaceX missions.

count(landing_outcome)	Landing_Outcome
10	No attempt
5	Success (drone ship)
5	Failure (drone ship)
3	Success (ground pad)
3	Controlled (ocean)
2	Uncontrolled (ocean)
2	Failure (parachute)
1	Precluded (drone ship)

Interactive map with Folium results

Here is a map showing information clusters where you can hover to get a report on successful (green) and unsuccessful(red) markers

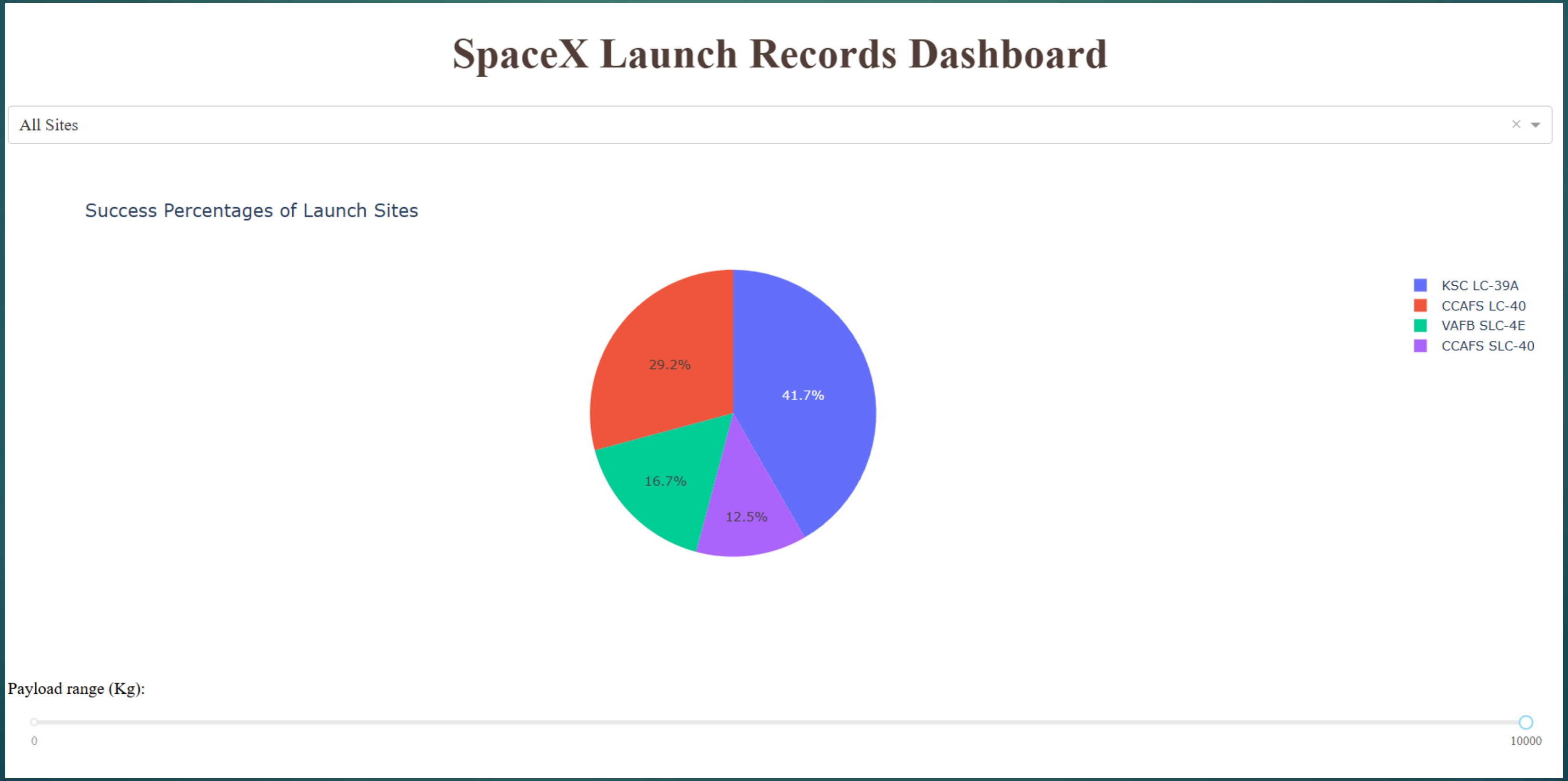


Interactive map with Folium results (2)

Here we have the distance between the coastline and the launchsit with a direct path which equates to 0.9 KM.

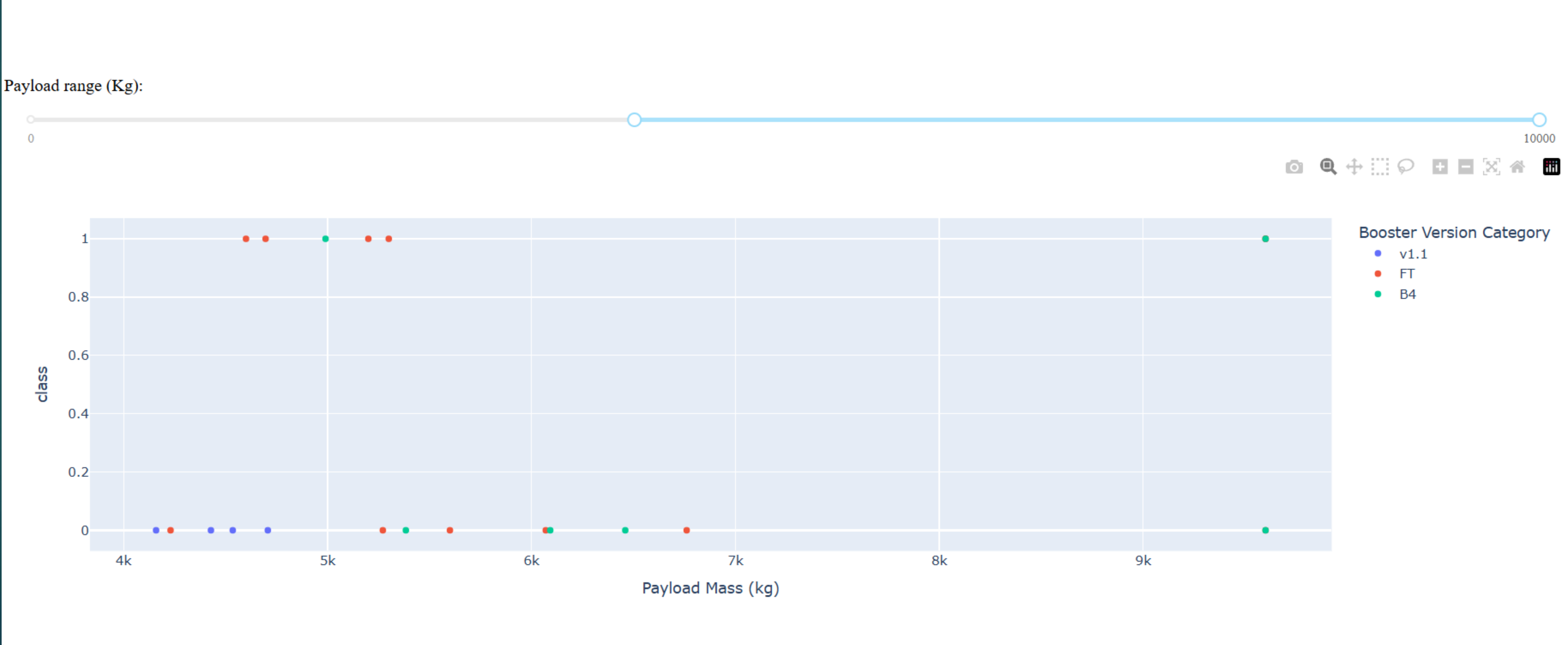
Plotly Dash dashboard results

This interactive dashboard is showing that KSC LC 39-A launchsite which is represented in blue has the highest percentage rate of successful launches with a figure of 41.7%



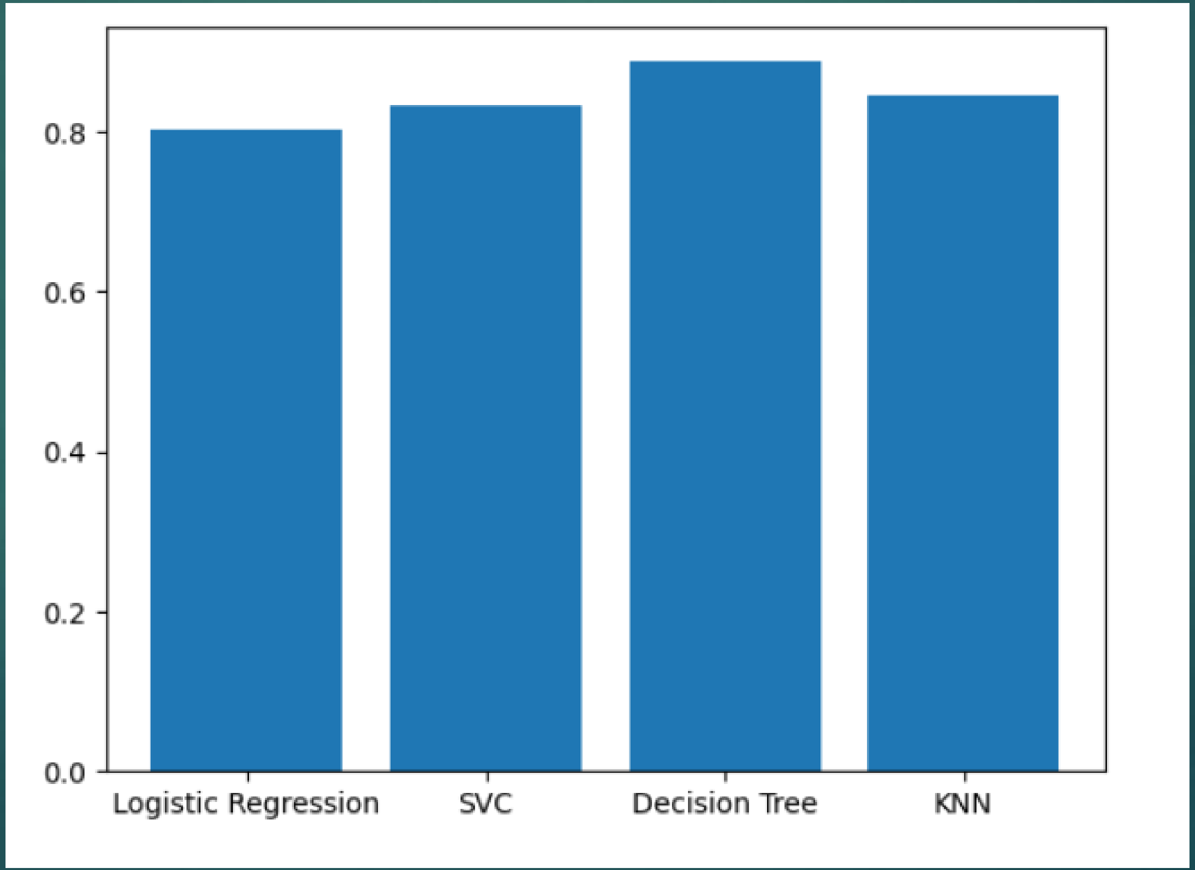
Plotly Dash dashboard results (2)

This is an interactive dashboard with a payload slider that shows different booster versions and how much payload mass. The class on the y axis shows if the dot is in class 1 it was successful and if it was in class 0 it is unsuccessful.We can gather that FT booster version had the most successful.



predictive analysis (classification) results

I used a bar chart to plot the best model for predictive analysis. The model with the best accuracy score was the Decision Tree with a value of 87 %. They are all grouped at are similar level showing that they all are relevant fits for the data.



DISCUSSION

I would like to point out that there was a dip in the line graph that showed the success rate of launches over each year. The dip was in 2018 so further investigation will have to occur for that data point.

Why did it happen in 2018?
Are there circumstances that affected it and could we find answers so it doesn't happen again?

These questions will need to be answered to gain further insights.

OVERALL FINDINGS & IMPLICATIONS

Findings

- Finding 1: The best location to launch is KSC LC 39-A.
- We can see the most successful orbit types are ES-L1, GEO and SSO.
- There is a correlation with high payload mass and successful launches.

Implications

- Implication 1: We can further get more data around this location to find out why it is successful and launch more rockets with similar conditions to this site
- Implication 2: We can launch more rockets to these orbit types to get more data on what is holding back other orbit types from being as successful. Are there barriers that need to be considered that haven't been?
- Implication 3: Increasing payload mass will result in a higher chance of successfully landing the rocket.

CONCLUSION

From what we have gathered there are strong correlations that need to be considered to factor if a landing is going to be successful or not.

One of these is the payload mass correlation. Launching more rockets with a higher payload in mass could be considered as sound logic to increase the chances of a successful landing.

Location, what is the climate of our successful launches. If anything shows from our data it is that location KSC LC 39-A is a great bet for a successful landing. Replicating the conditions that this location offers is desirable

Conducting more research into the events of 2018 is a key takeaway from this aswell. The trend in the line graph changed with a dip and then continued in the same positive trajectory.

What we have here are answers to many questions we did not know and we have also found a question that we did not know existed. All in the name of winning the space race with Data Science.

APPENDIX

- GitHub Repositories:
- All completed notebooks and python files can be found in my GitHub repository:
- J-R-Allen/IBM-Data-Science-Professional-Certificate-Capstone: IBM Capstone Assignment (github.com)