WHEN SKIES NOT THE LIMIT @ SPACE Y

IBM Developer
SKILLS NETWORK

Tijo John

12 April 2024

OUTLINE

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

EXECUTIVE SUMMARY

- **Summary of Methodologies**

  - Space Y Data Collection using Space Y API
  - Space Y Data Collection with Web Scraping
  - Space Y Data Wrangling
  - Space y Exploratory Data Analysis using SQL
  - Space Y EDA DataViz using Python – Pandas & Matplotlib
  - Space Y launch Site Analysis - Folium Interactive Visual Analytics & Plotly Dash
  - Space Y Machine Learning Landing Prediction

- **Summary of all Results**

  - EDA Results
  - Interactive Visual Analytics & Dashboards
  - Predictive Analysis - Classification

INTRODUCTION

Project Background and Context:

- Space Y advertises Falcon 9 rocket launches with a cost of 62 million dollars while other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Therefore 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.

Desired Results:

- We will predict if the Falcon 9 first stage will land successfully

SECTION 1:

METHODOLOGY

- Data Collection Methodology
- Perform Data Wrangling
- Perform Exploratory Data Analysis – Visualization & SQL
- Perform Interactive Visual Analytics – Folium & Plotly Dash
- Perform Predictive Analysis – Classification models

DATA COLLECTION – SPACE Y API

Data collected using Space Y API by making a GET request, then parsed the launch data and decoded the response content as a JSON result which was then converted into Pandas data frame.

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

DATA COLLECTION – WEB SCRAPING

Performed Web scraping to collect Falcon 9 records from Wikipedia using BeautifulSoup and Request and then created a data frame by parsing the launch HTML.

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

DATA WRANGLING

After obtaining and creating a Pandas data frame from the collected data, data was filtered using the BoosterVersion, then dealt with the missing date values in the LandingPad and PayLoadMass columns – where missing data values were replaced using mean value of column.

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

EDA WITH DATA VISUALIZATION

- Scatter plots: Scatter plots were used to represent the relationship between two variables. Different sets of features were compared such as *Flight Number vs. Launch Site, Payload vs. Launch Site, Flight Number vs. Orbit Type and Payload vs. Orbit Type.*

- Bar chart: Bar charts were used makes it easy to compare values between multiple groups at a glance. The x-axis represents a category and the y-axis represents a discrete value. Bar charts were used to compare the *Success Rate* for different *Orbit Types*

- Line chart: Line charts are useful for showing data trends over time. A line chart was used to show *Success Rate* over a certain number of *Years.*

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

EDA WITH SQL

A list of some of the SQL queries performed on the dataset is listed below:

- Displaying the names of the unique launch sites in the space mission --- Displaying 5 records where launch sites begin with the string 'CCA' --- Displaying the total payload mass carried by boosters launched by NASA (CRS) --- Displaying average payload mass carried by booster version F9 v1.1,

- Listing the date when the first successful landing outcome in ground pad was achieved --- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 --- Listing the total number of successful and failure mission outcomes --- Listing the names of the booster versions which have carried the maximum payload mass --- Listing the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 --- Ranking the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order.

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

BUILD AN INTERACTIVE MAP WITH FOLIUM

Here, we created a Folium map to mark all the launch sites and created map objects, circles, lines to mark the success or failure of launches of each site.

Created a launch set outcomes (failure = 0 or success = 1)

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

BUILD AN DASHBOARD WITH PLOTLY DASH

Built an interactive dashboard application with Plotly Dash by:

- Adding a Launch Site Drop down Input component
- Adding a callback function to render success pie chart based on selected site drop down
- Adding a Range slider to select Payload
- Adding a callback function to render success payload scatter chart plot

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

MACHINE LEARNING PREDICTION

- Summary of how the project was built, evaluated, improved and reached the best performing classification model.
- After loading the date as a Panda Data frame, set out to perform Exploratory Data Analysis and determine the training labels.
- In order to find the best ML model/method that would perform best using the test data between SVM, Classification Trees, K-nearest Neighbours and Logistic Regression.

Completed GITHUB URL notebook: https://github.com/TiJohnMath/SpaceX-Falcon-9-first-stage-Landing-Prediction/tree/main

RESULTS

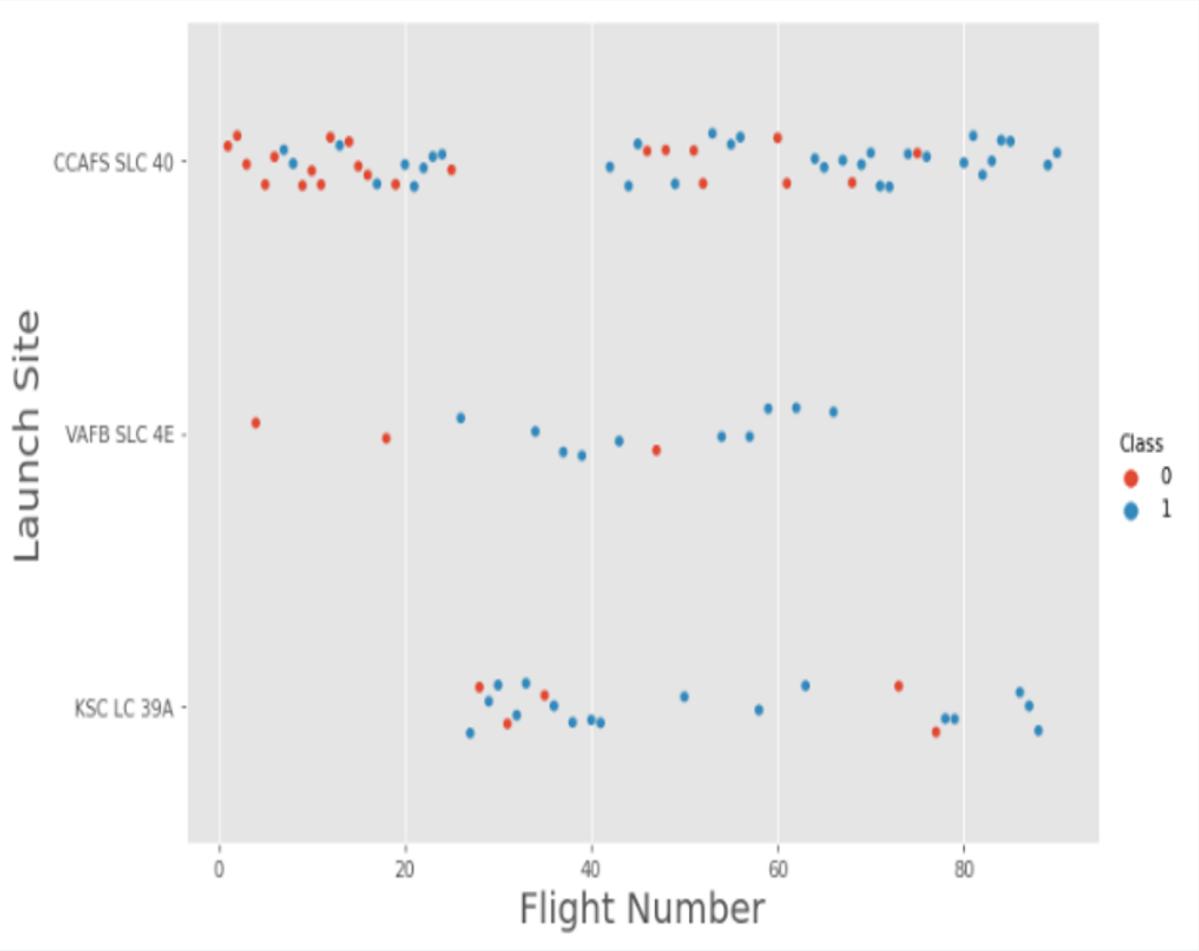
The results of the exploratory data analysis revealed that the success rate of the Falcon 9 landings was 66.66%.

The predictive analysis results showed that the Decision Tree algorithm was the best classification method with an accuracy of 94%

SECTION 2:

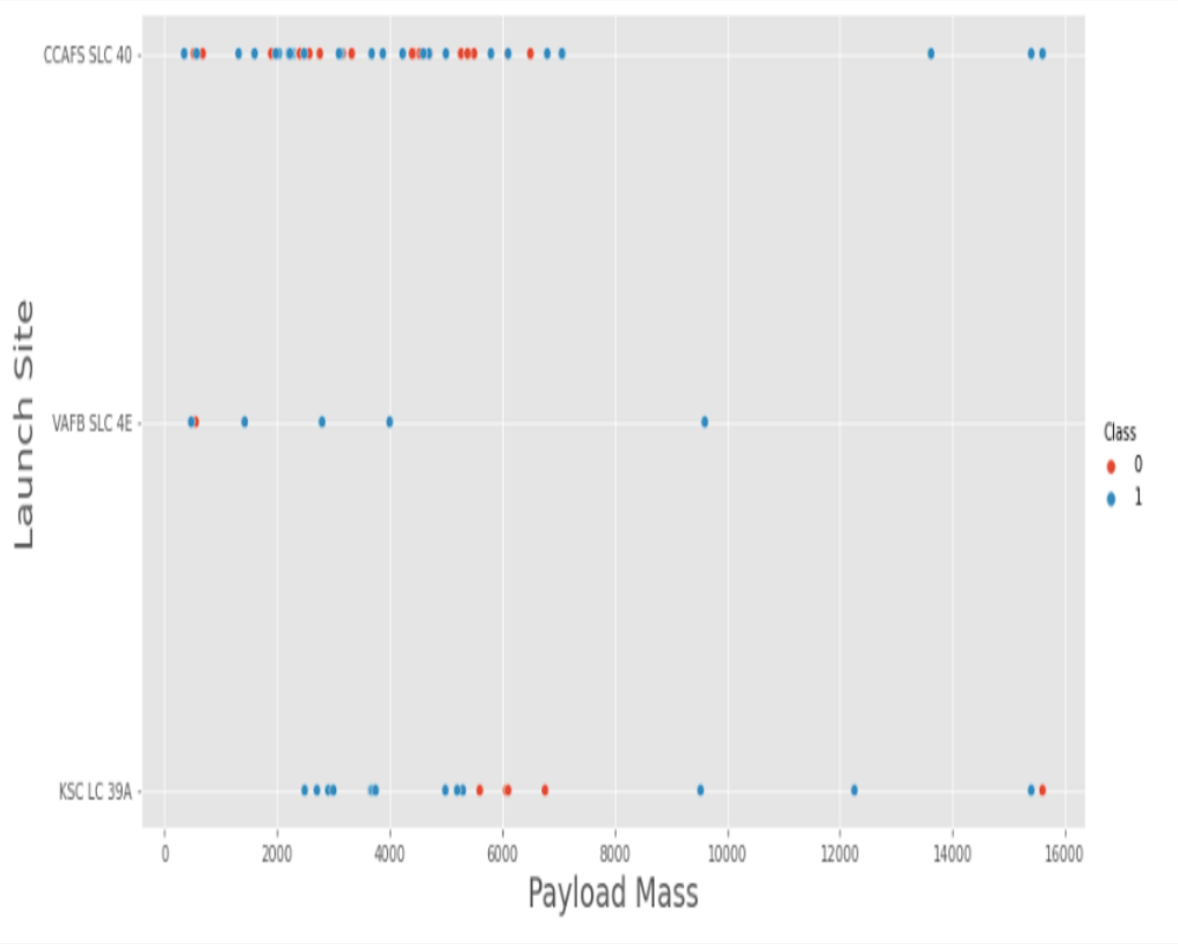
INSIGHTS DRAWN FROM EDA

Flight Number vs. Launch Site



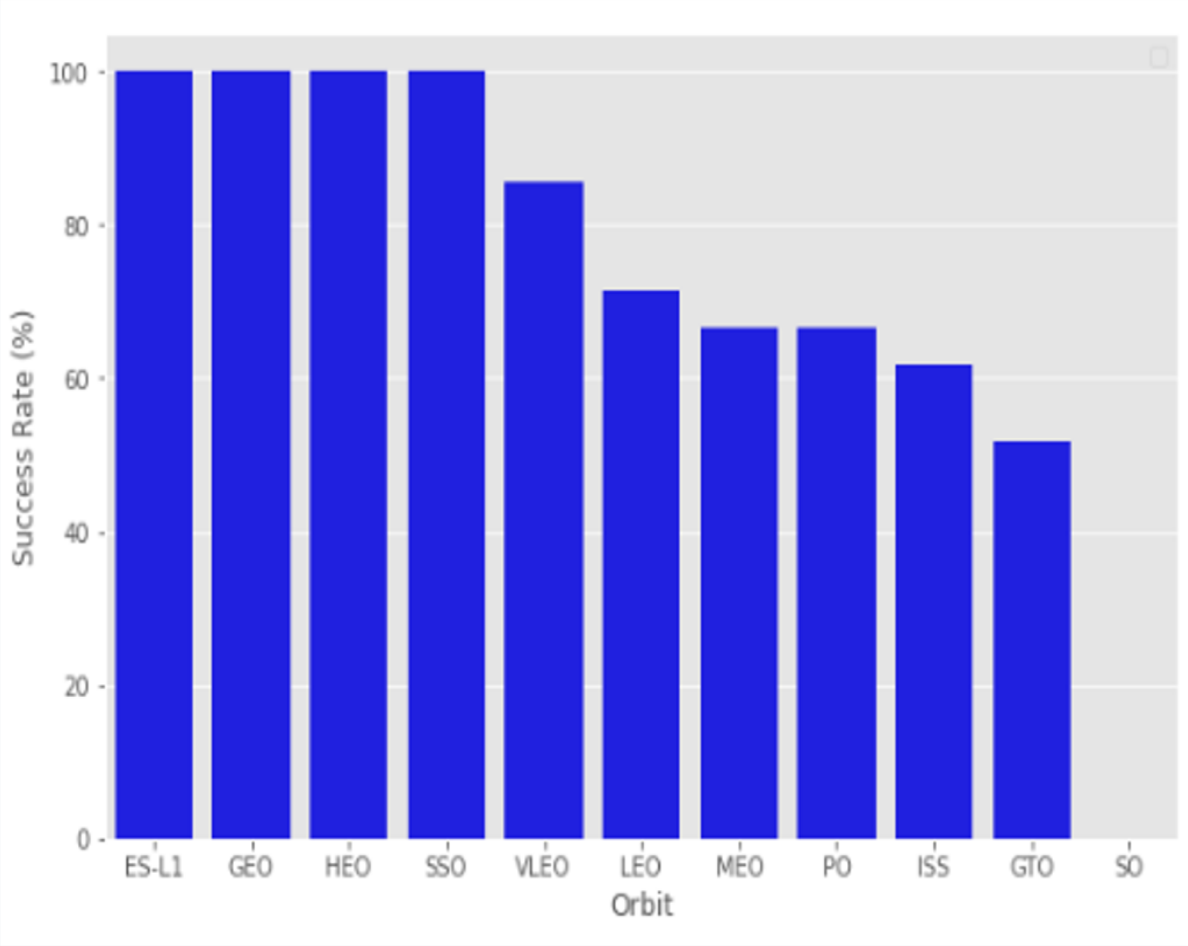
- This figure shows that the success rate increased as the number of flights increased.
- The blue dots represent the successful launches while the red dot represent unsuccessful luanches.
- There seems to be an increase in successful flights after the 40th launch.

Payload vs. Launch Site



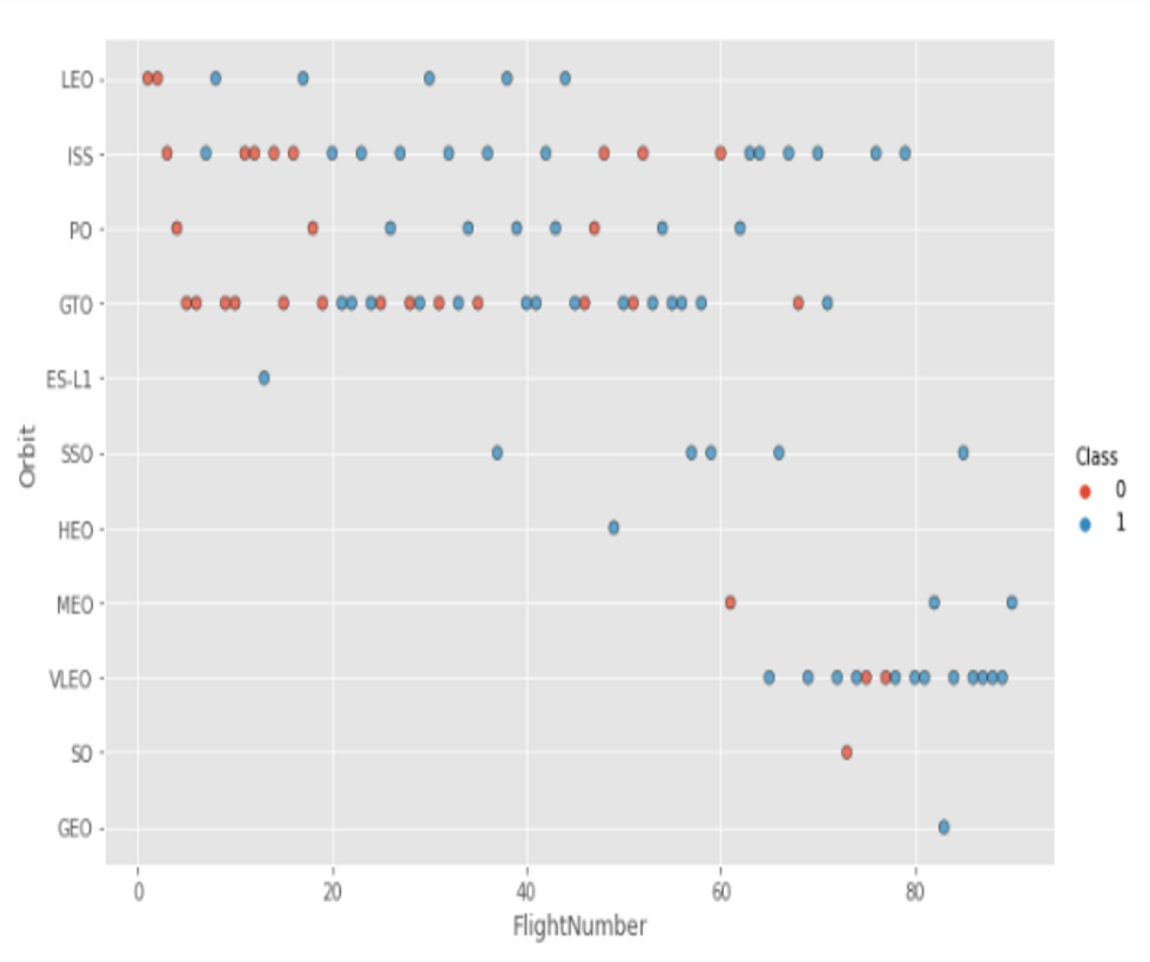
- The blue dots represent the successful launches while the red dots represent unsuccessful launches.
- For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass
- There seems to be a weak correlation between Payload and Launch Site and therefore decisions cannot be made using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

Success Rate vs. Orbit Type



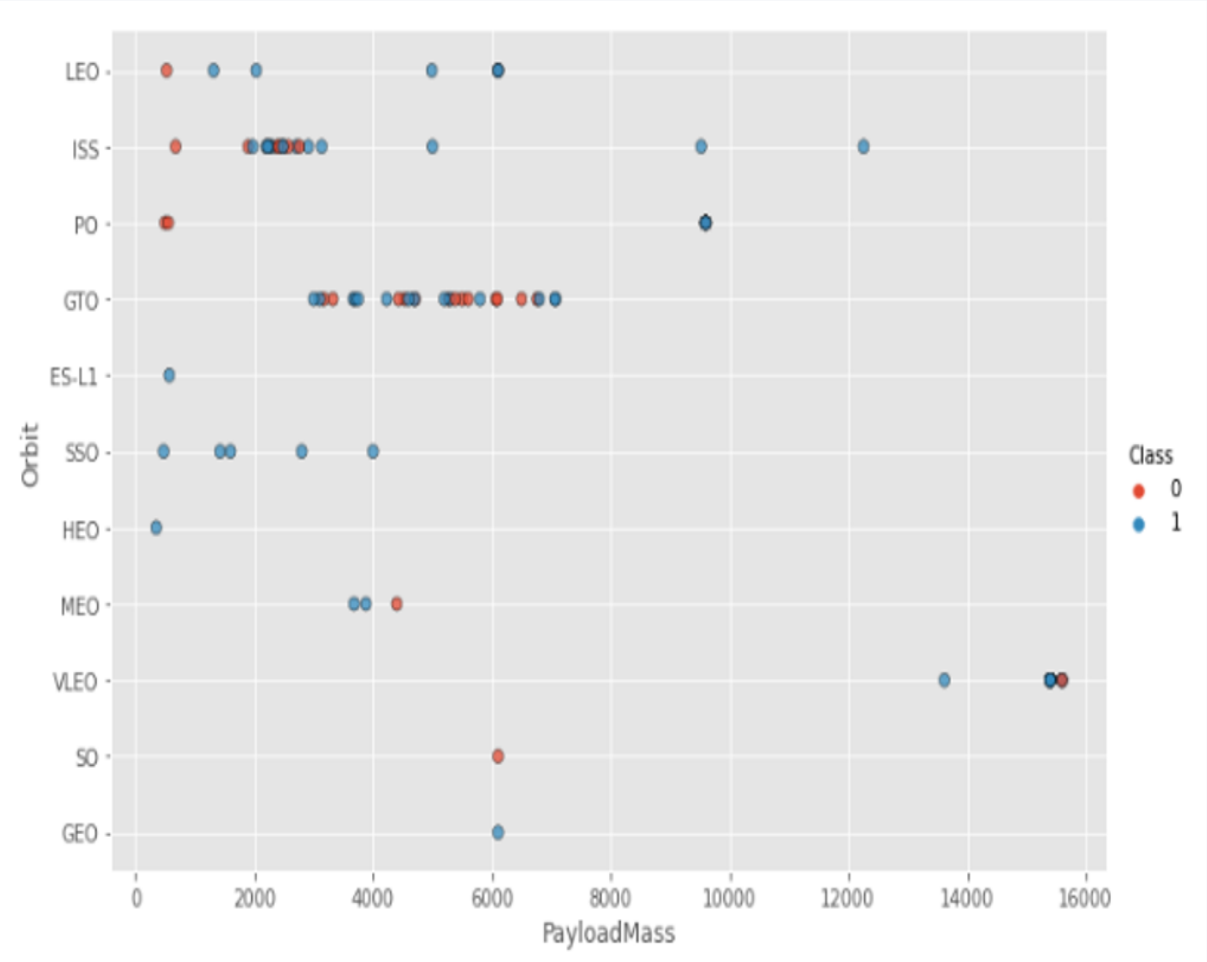
- Orbits SSO, HEO, GEO, and ES-L1 have 100% success rates.
- SO orbit did not have any successful launches with a 0% success rate.

Flight Number vs. Orbit Type



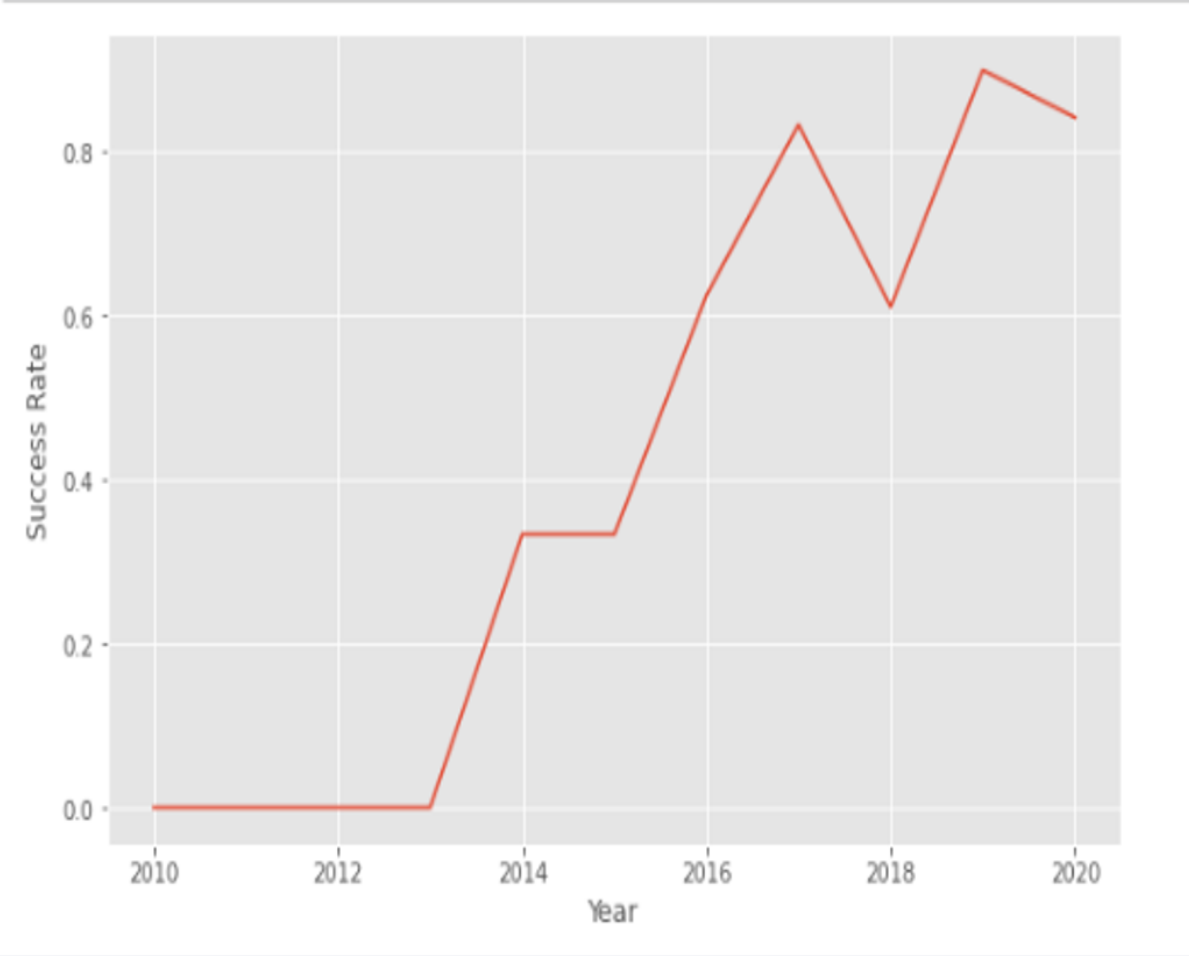
- In the LEO orbit, the success is positively correlated to the the number of flights.
- There seems to be no relationship between flight number in the GTO orbit.
- The SSO orbit has a 100% success rate however with fewer flights than the other orbits
- FLights numbers greater than 40 have a higher success rate than flight numbers between 0-40.

Payload vs. Orbit Type



- As the payloads get heavier, the success rate increases in the PO, SSO, LEO and ISS orbits.
- There seems to be no direct correlation between obrit type and payload mass for GTO orbit as both successful and failed launches are eually present

Launch Success Yearly Trend



- The general trend of the chart shows an increase in landing success rate as the years pass. There is however a dip in 2018 as well as in 2020.

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

Launch Site Names Begin with 'CCA'

- The LIMIT and LIKE clauses were used to display only the top five results where the *launch_site* name starts 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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Successful Drone Ship Landing with Payload between 4000 and 6000

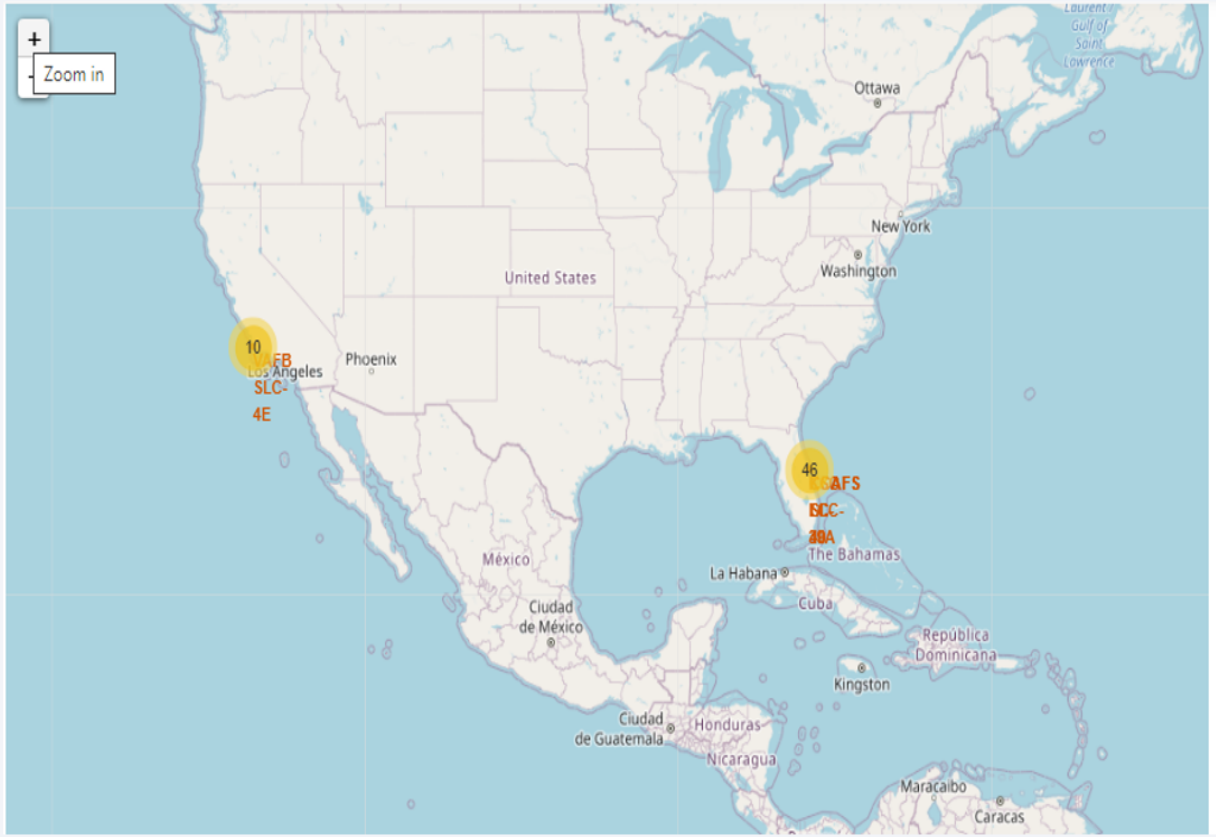
- The BETWEEN clause was used to retrieve only those results of payload mass greater than 4000 but less than 6000. The WHERE clause filtered the results to include only boosters which successfully landed on drone ship

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Section 3:

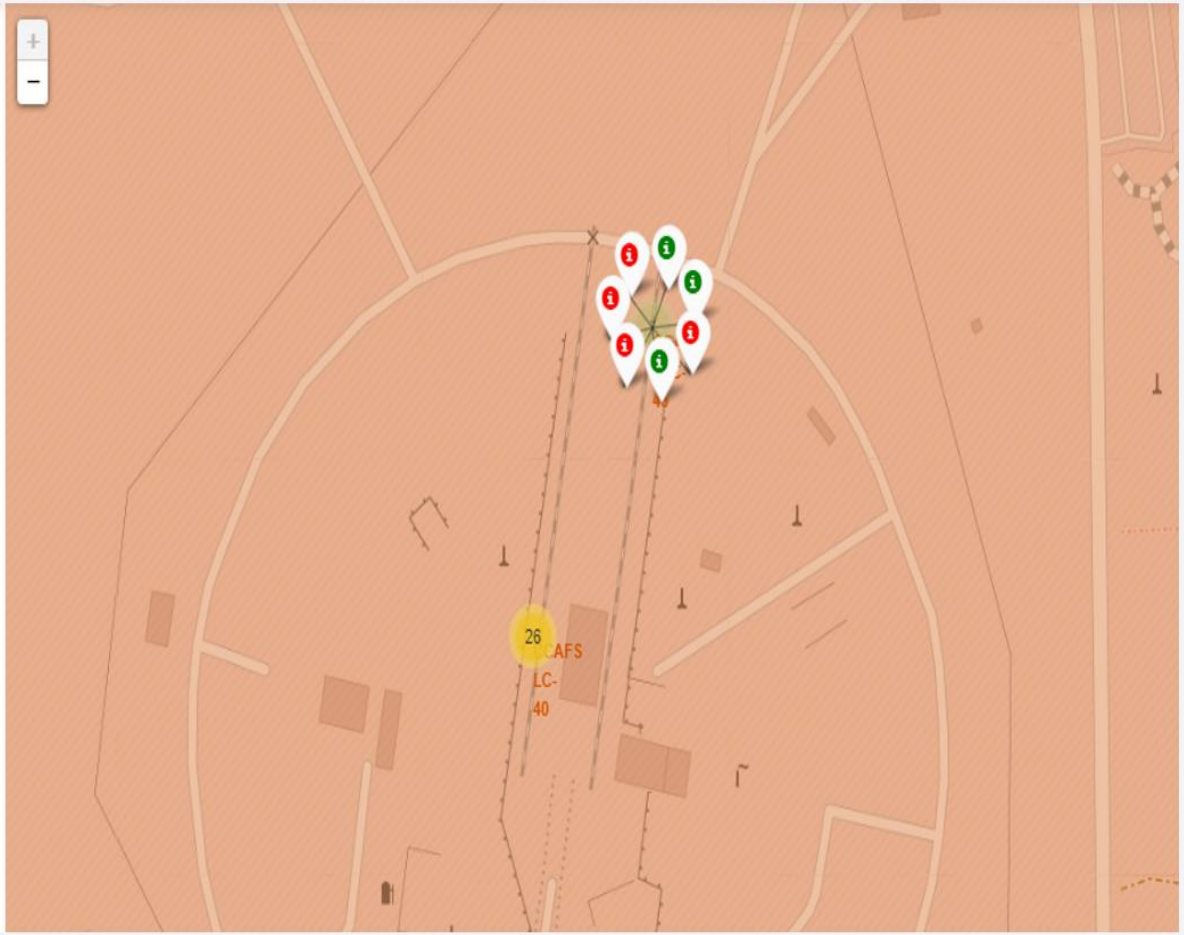
Launch Sites and Proximity Analysis

SpaceX Launch Sites Locations



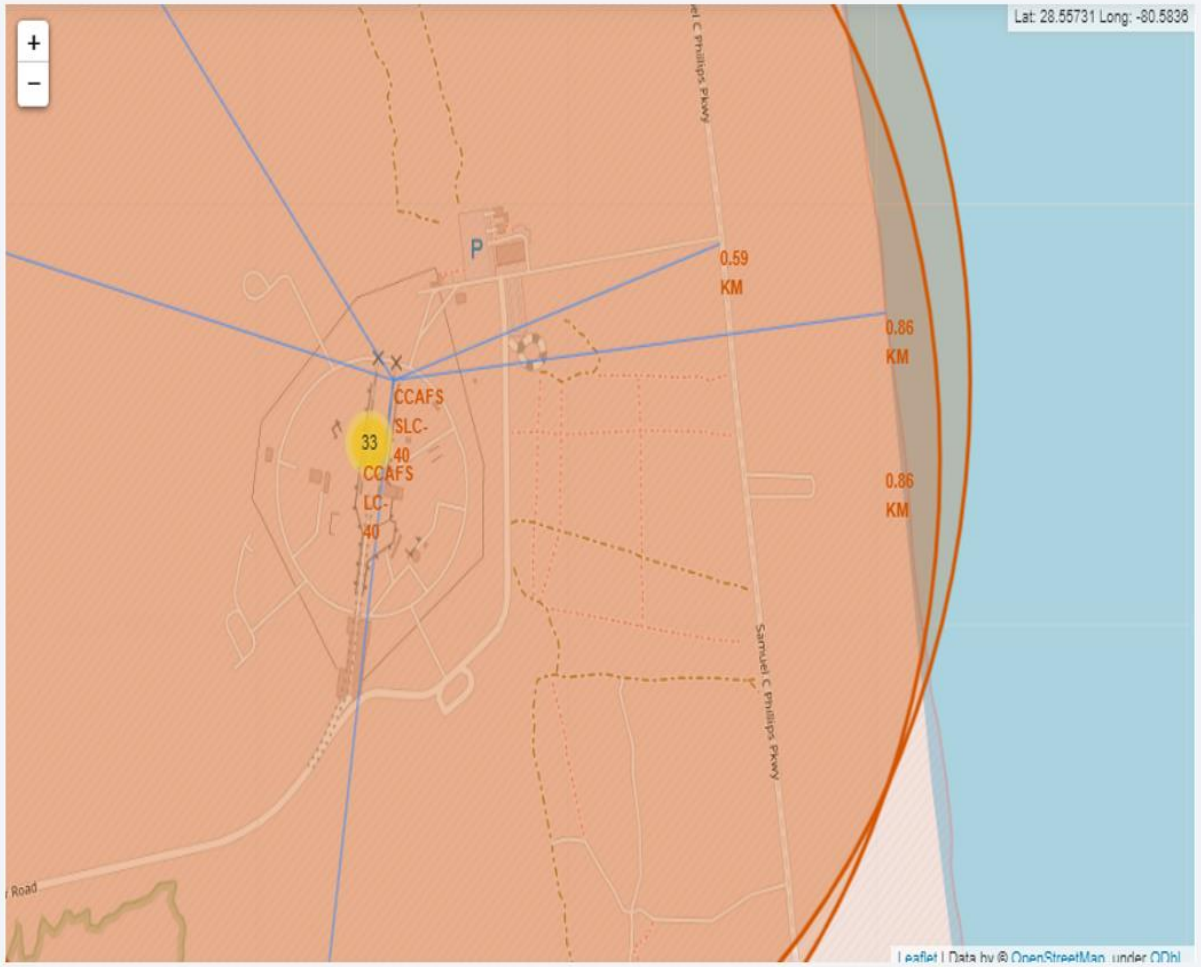
- The yellow markers are indicators of where the locations of all the SpaceX launch sites are situated in the US.
- The launch sites have been strategically placed near the coast

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).

Launch Site Proximities



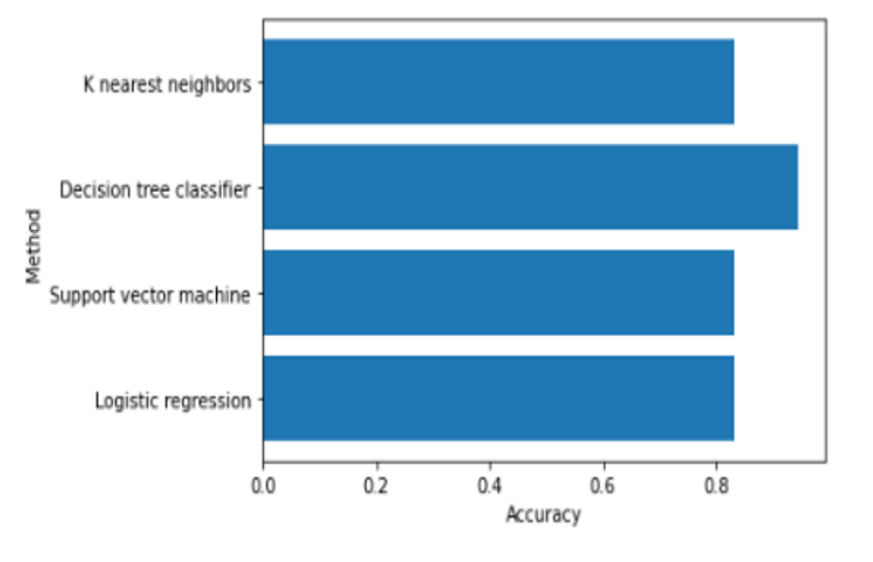
- The generated map shows that the selected launch site is close to a highway for transportation of personnel and equipment. The launch site is also close to the coastlines for launch failure testing.
- The launch sites also maintain a certain distance from the cities. (Can be viewed in notebook).

Section 4:

Build a Dashboard with Plotly Dash

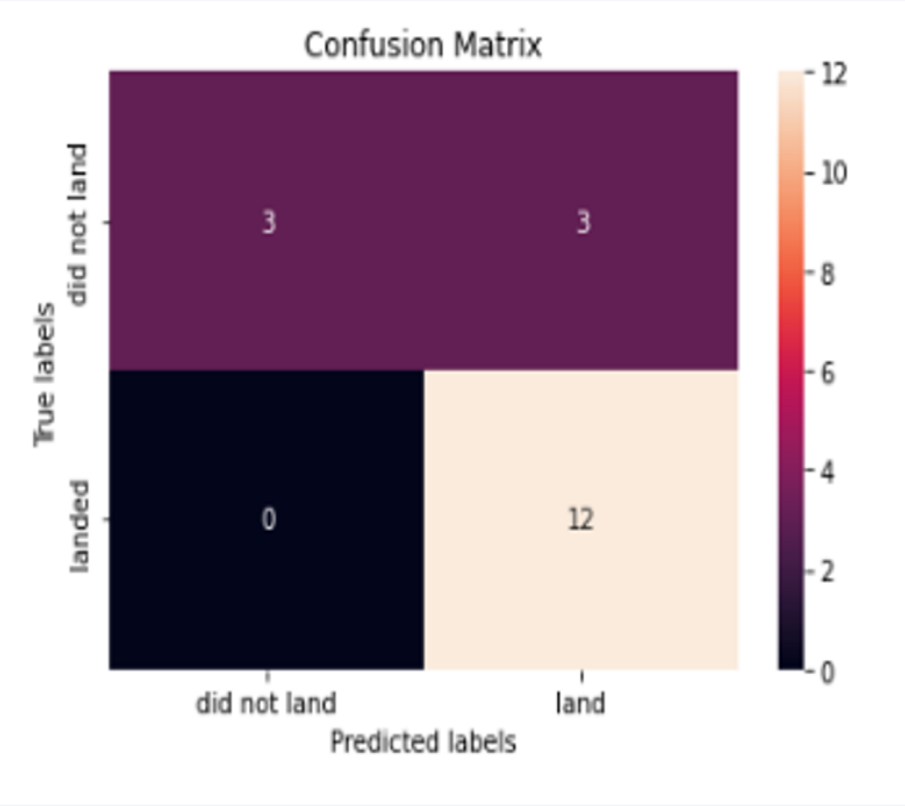
Classification Accuracy

- 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.

Conclusions

- 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 launch sites are strategically located near highways and railways for transportation of personel and cargo, but also far away from cities for safety.
- 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/withouttheh/IBM-Data-Science-Capstone-Project

Thank you !