IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Devin M>
<11/24/24>

Outline

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

Executive Summary

- In this project, python and SQL were utilized in junction to determine if spacex will reuse the first stage of their rocket. This Is important to competitors because of the investment required to design and launch a rocket.
- With 83% accuracy we can predict when SpaceX will reuse the first stage of a rocket.

Introduction

- SpaceX is an aerospace company with a focus in space travel. One of the first steps to accomplishing long term space missions is to make them as efficient as possible. One of the way to do so is to recover the first stage of the rocket. This stage is where most of the propulsion systems are located.
- Problem A: Statistically predict when SpaceX will reuse stage one of the rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected through open-source websites VIA web scraping utilizing Pythons request library
- Perform data wrangling
  - The data was molded into a cohesive form by utilizing SQL and Python
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

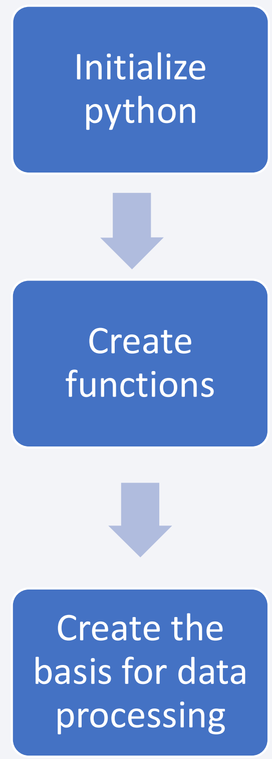
- The data was collected by using multiple python libraries. The data was sourced from reliable public sources.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Data collection begins by understanding where the data comes from and initializing python. Once the data is scraped or acquired then the formatting of the data begins.
- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/jupyter-labs-spacex-data-collection-api-v2.ipynb

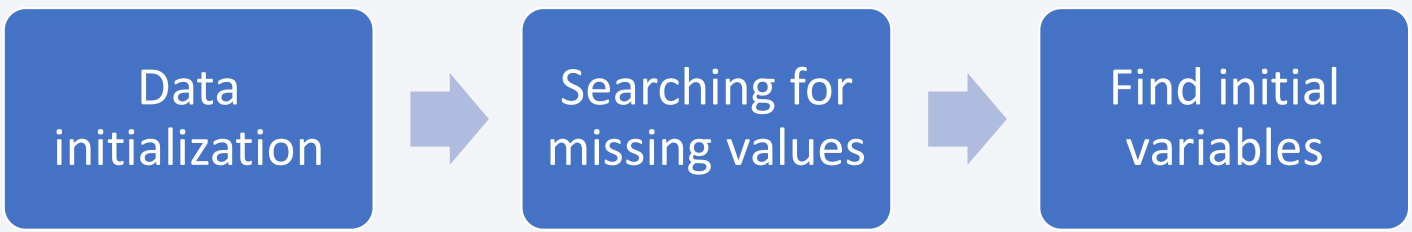
Data Collection - Scraping

- Web scraping begins by finding a source of useable data from an open source/public data set. After the source is found, the scraping takes place by literally extracting the data. Once the data is collected, formatting is applied to make the data more pliable.
- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Once the data is in the ecosystem, data wrangling takes place. This is the act of formatting the data in a way that facilitates exploratory data analysis(EDA)
- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/labs-jupyter-spacex-Data wrangling-v2.ipynb



EDA with Data Visualization

- Scatter plots, Bar plots and Line plots were all used in the EDA process. The scatter plots determine the overall trends of the data with multiple variabels. The line plots do a similar action but by using to mode of the time series. Bar plots were used to plot categorical values against a single factor.
- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/jupyter-labs-eda-dataviz-v2.ipynb

EDA with SQL

- Determine unique names of launch sites
- Finding the total and average payload mass carried at NASA
- Finding the first successful landing date
- Finding best boosters, payload and dates

- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Using Folium, map markers were placed in the locations of launches. Lines were also made to determine the distances from each launch as well as from the nearest water mass.
- I added these objects to show a clear correlation between the geographic location and the landing percentage
- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

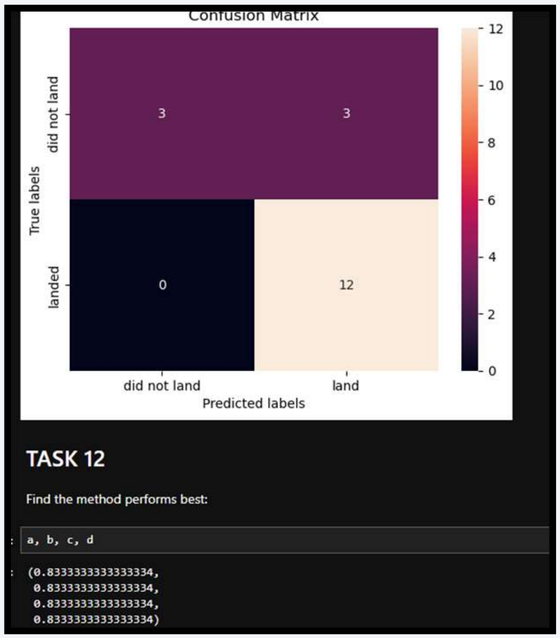
- A pie chart and scatter plot was used to convey the data. The plots displayed were able to showcase payloads for varying boosters and for all the launch sites
- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

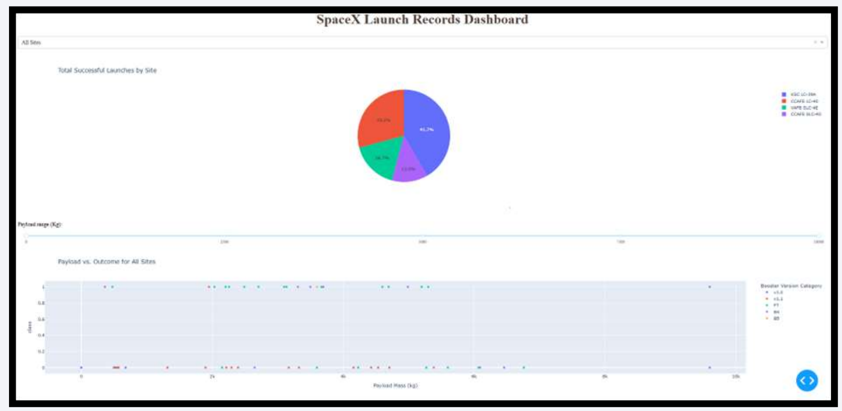
- In summary, first the splitting of data into training and testing was done first to initialize the data and normalize it. After this parameter optimization was done on multiple models to test out what model best fit the dataset.

- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/blob/main/SpaceX-Machine-Learning-Prediction-Part-5-v1.ipynb

Results



Confusion Matrix
True labels
did not land
landed
3
3
0
12
did not land
land
Predicted labels
12
10
8
6
4
2
0
TASK 12
Find the method performs best:
a, b, c, d
(0.8333333333333334,
0.8333333333333334,
0.8333333333333334,
0.8333333333333334)



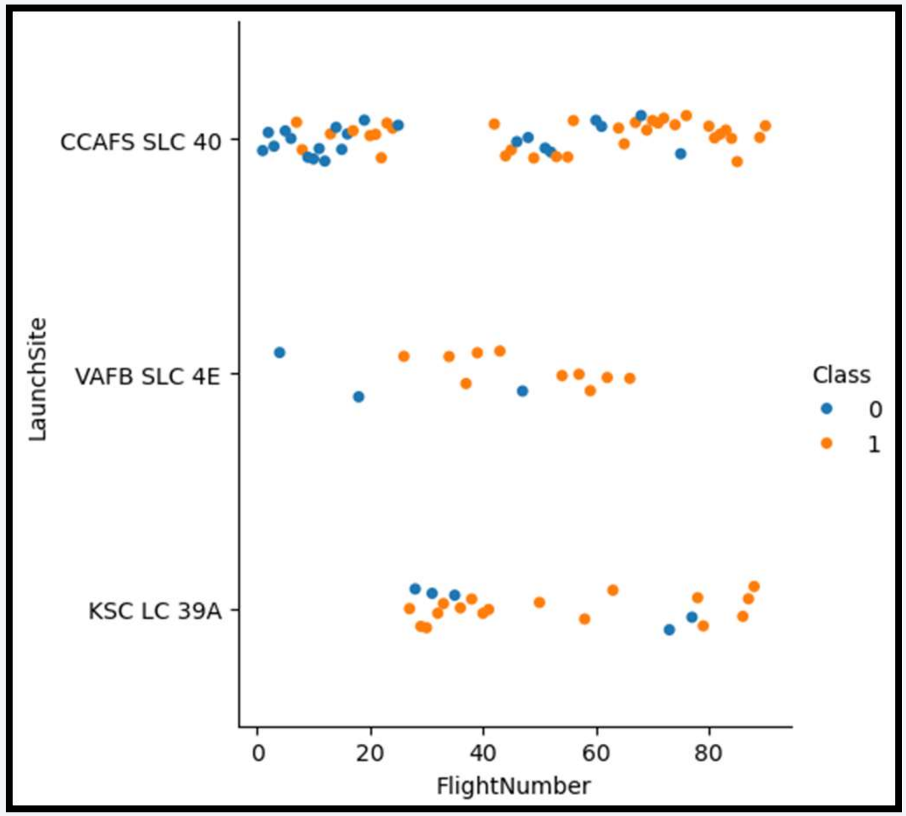
SpaceX Launch Records Dashboard

Section 2

Insights drawn from EDA

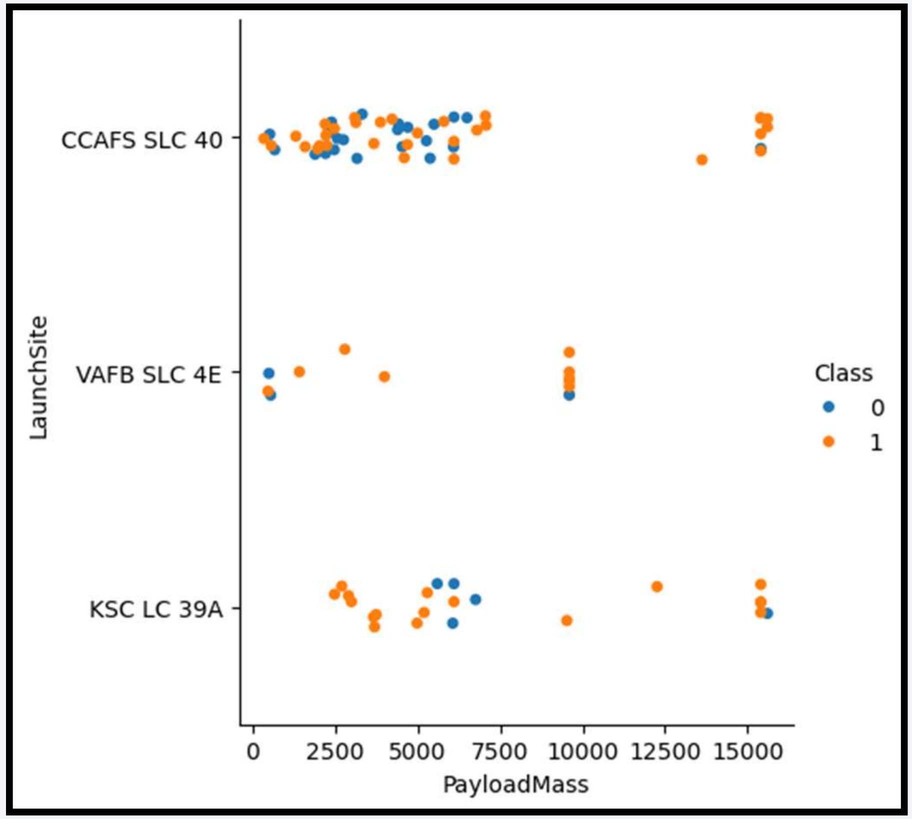
Flight Number vs. Launch Site

- The figure displays which site yields the most successful and unsuccessful landings. The dots are keyed to the classes of 0 and 1.



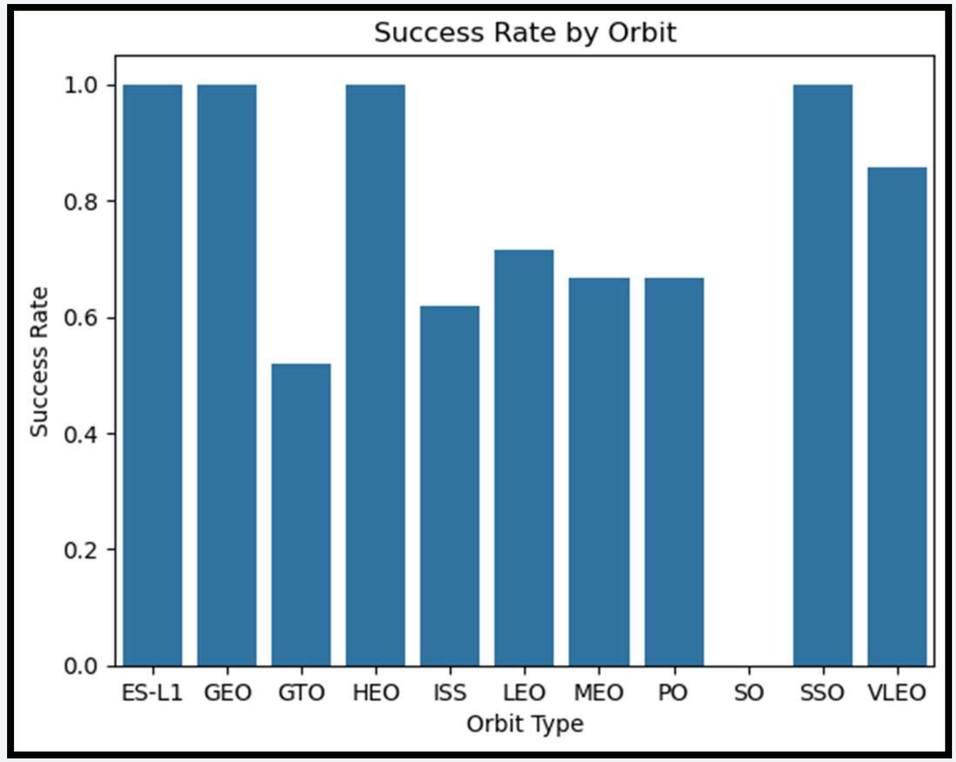
Payload vs. Launch Site

- The figure displays the same Y-axis however the X-axis is the payload mass. Displays the correlation between mass, launch site and class



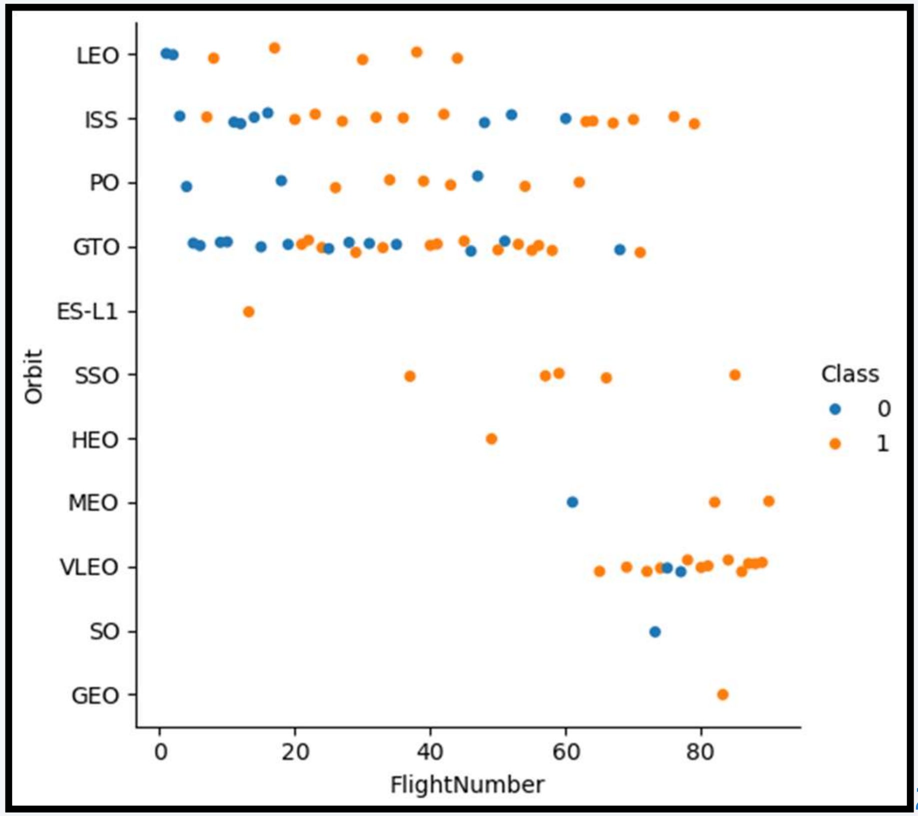
Success Rate vs. Orbit Type

- The bar chart displays the correlation between orbit type and success rate. With "SO" being zero.



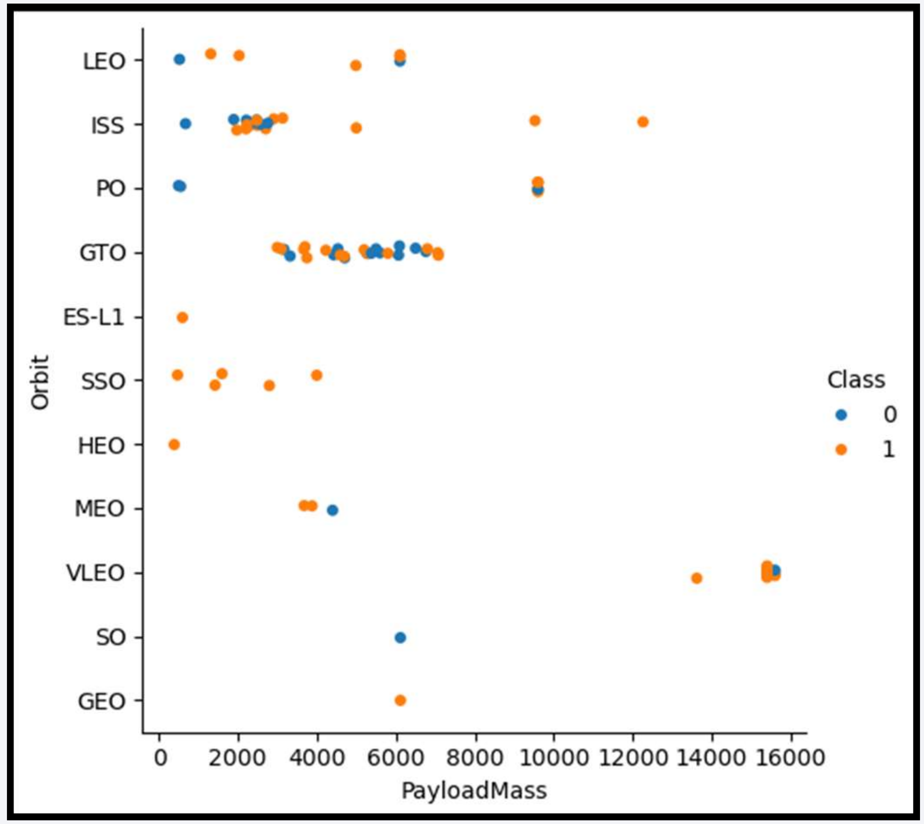
Flight Number vs. Orbit Type

- The figure displays the trend for flight number and orbit type. Showing the favored type of orbit and newer trials. While keying the values to the class of the launch.



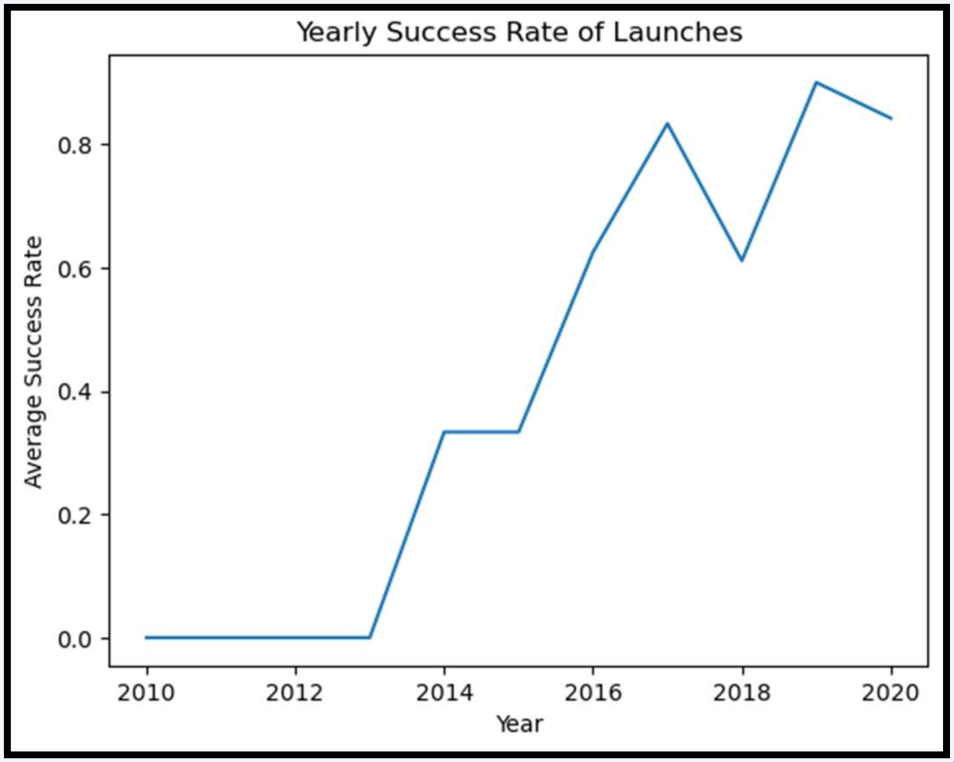
Payload vs. Orbit Type

- The scatter plot defines what the optimal payload mass is for a given orbit type and denotes this by keying the intersections to the launch class.



Launch Success Yearly Trend

- The figure displays the trend over time of successful launches.



All Launch Site Names

- %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
- By using the above query, launch site names can be retrieved.

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

Launch Site Names Begin with 'CCA'

- %sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
- The above query results in the table below. The limit 5 only shows 5 items, however there are many more.

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

Total Payload Mass

- %sql SELECT SUM("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Customer" LIKE 'NASA (CRS)’;
- The above query finds the total payload mass.

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

- %sql SELECT AVG("PAYLOAD_MASS__KG_") AS Total_Payload_Mass FROM SPACEXTABLE WHERE "Booster_Version" LIKE 'F9 v1.1';
- The above query yields the total payload mass and averages it.

Total_Payload_Mass
2928.4

First Successful Ground Landing Date

- %sql SELECT MIN("Date") AS First_Success_Date FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)’;
- The above query results in the first date of a successful landing

First_Success_Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
- The above query shows the booster versions for specified payload masses

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

Total Number of Successful and Failure Mission Outcomes

- %sql SELECT "Mission_Outcome", COUNT(*) AS Count FROM SPACEXTABLE GROUP BY "Mission_Outcome";
- The above query depicts the failures and successes

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

Boosters Carried Maximum Payload

- %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
- The query above retrieves the boosters that can take max payload mass

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

2015 Launch Records

- %sql SELECT substr("Date", 6, 2) AS Month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE substr("Date", 0, 5) = '2015' AND "Landing_Outcome" = 'Failure (drone ship)';
- The above query displays the booster version outcome

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- %sql SELECT "Landing_Outcome", COUNT(*) AS Count FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY Count DESC;
- Makes a table of descending order landing outcomes

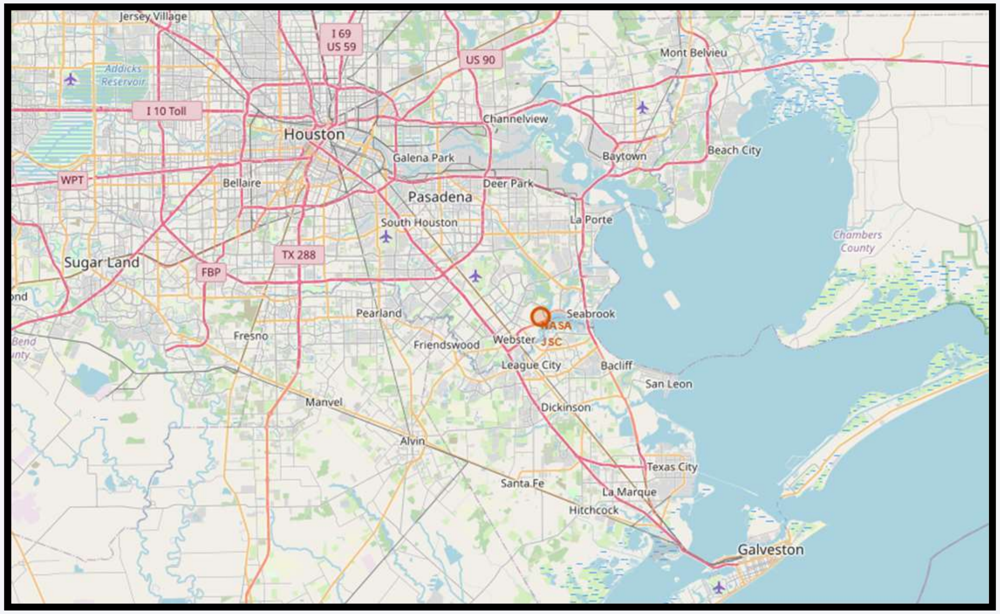
Landing_Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

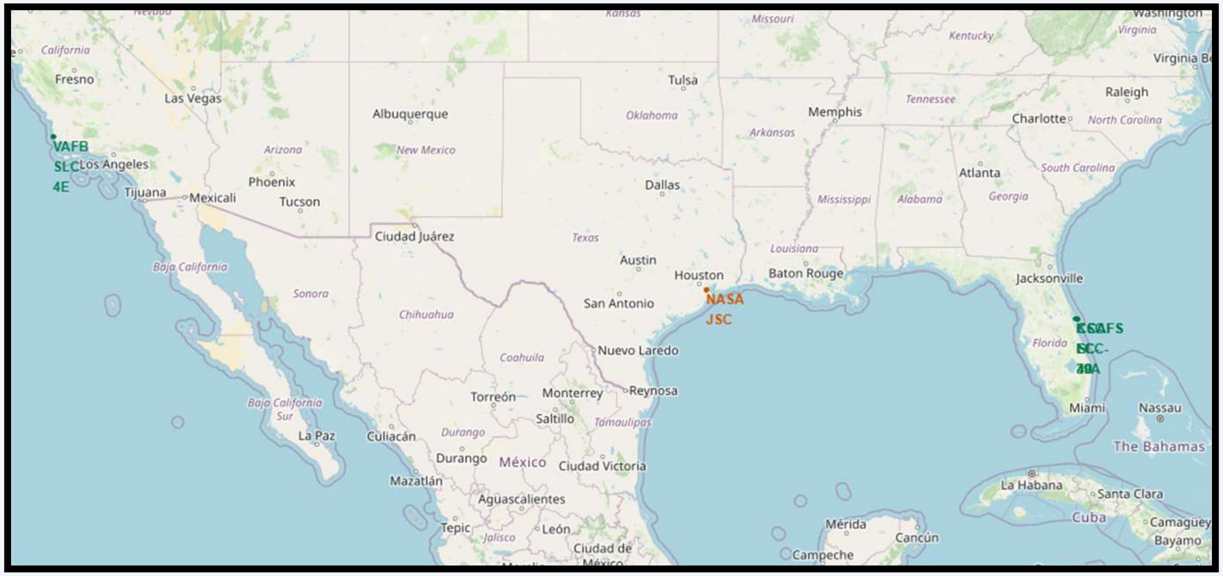
Nasa location map

- The map displays where NASA is with respect to Houston Texas. The orange circle denotes the exact location.



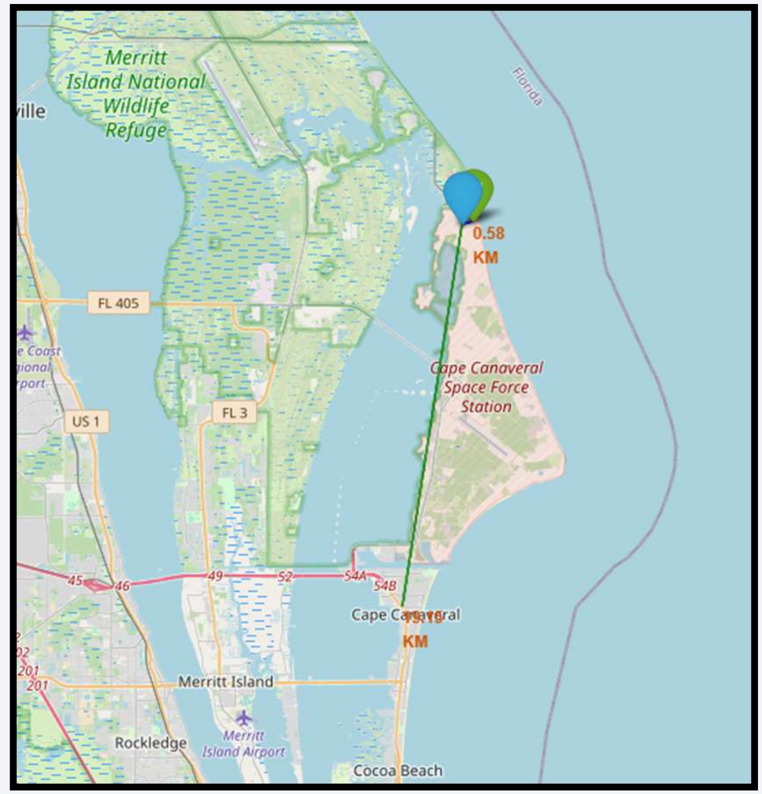
Launch location map

- The map highlights the launch locations in the united states.



Launch site distances

- The map displays the distances for each launch site in the cape Canaveral area

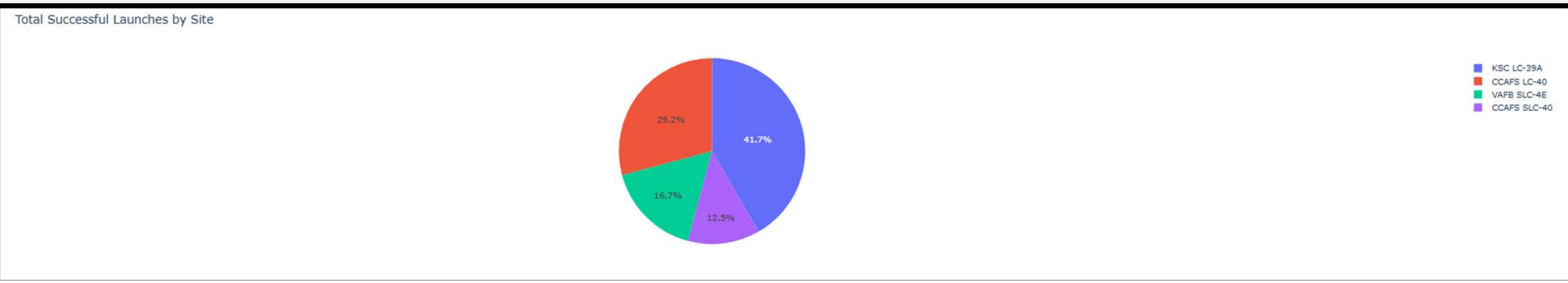


Section 4

Build a Dashboard with Plotly Dash

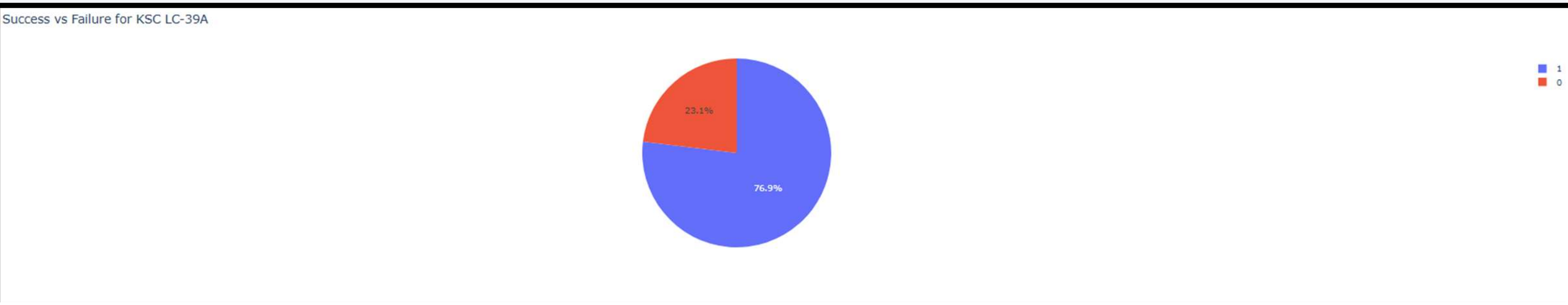
Total successful launches by site

- As shown in the figure below, KSC LC has the highest success rate out of all the locations



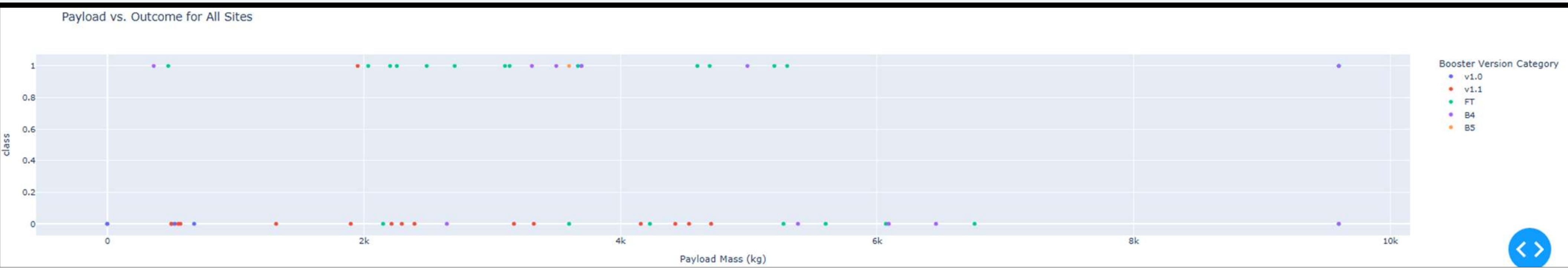
Failure Vs Success for KSC-LC-39A

- As indicated below, this location has the highest amount of successful landings



Payload Vs. Outcome

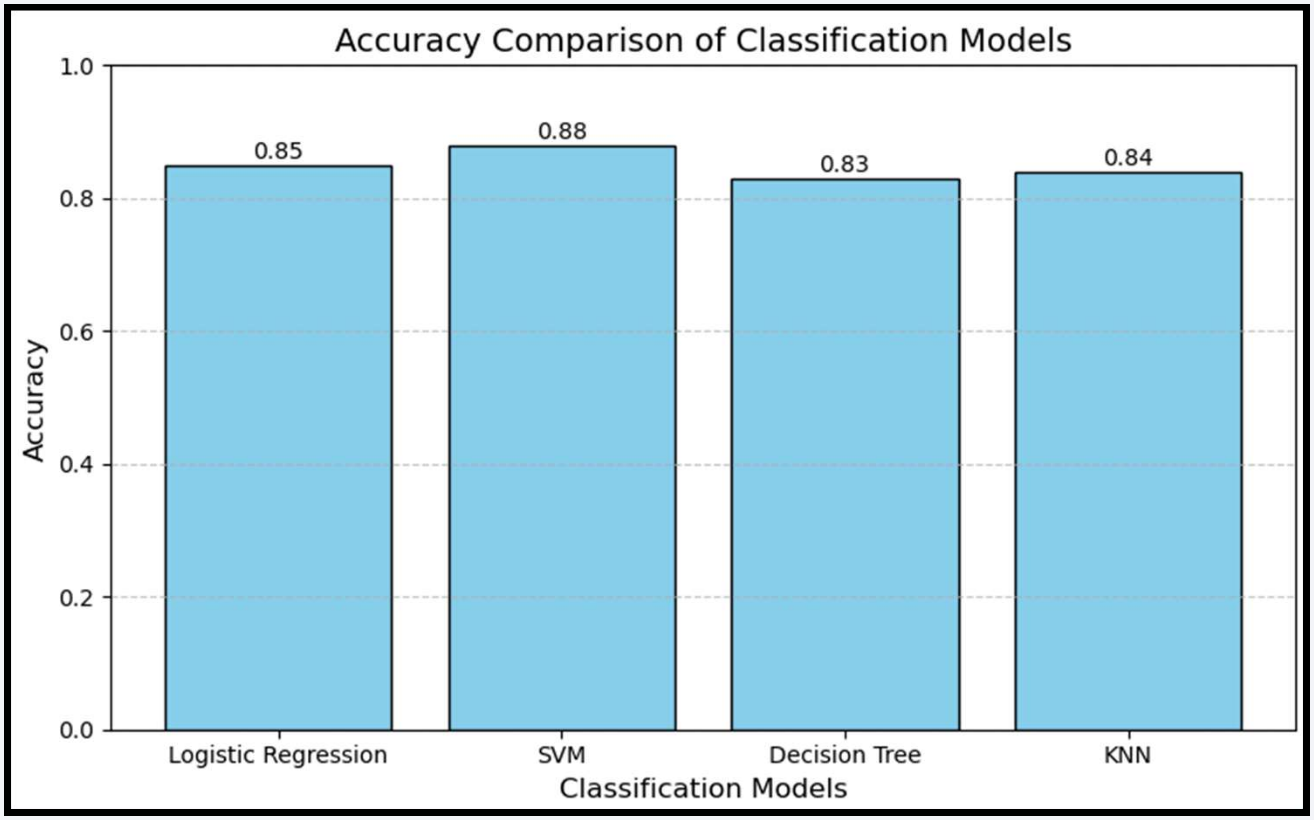
- The figure below illustrates the relationship between payload and landing outcome



Section 5

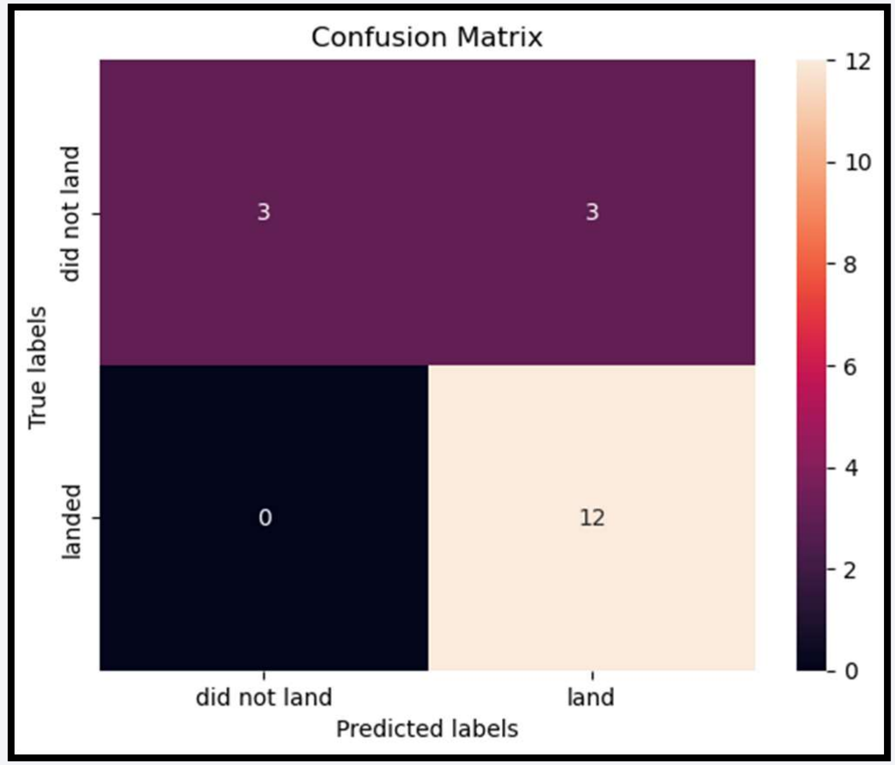
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- In the confusion matrix shown, there are 12 true positives and 3 false negatives.
- This indicates a high amount of accuracy



Conclusions

- Time and place and orbits effect the landing of rockets
- KSC-LC-39A has the best landing ratio
- The best model to predict stage reusage is SVM

Appendix

- All snippets are included on my GitHub repository

- https://github.com/Devinm345/IBM-Data-Science-Professional-Certificate-Capstone/tree/main

Thank you!