IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jaydeep Jitendra
26/10/2021

Outline

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

Executive Summary

- ## Summary of methodologies

1. Data Collection
2. Data Wrangling
3. EDA with data visualization
4. EDA with SQL
5. Building an interactive map with Folium
6. Building a Dashboard with Plotly Dash
7. Predictive Analysis

- ## Summary of all results

1. Exploratary data analysis results
2. Interactive analytics demo in screenshots
3. Predictive analysis results

Introduction

- In this project we will try to predict if Falcon 9 first stage will land successfully or not by using previous data and different regression analysis.
- We will discuss through the problem and how to find solutions:

1. How to predict if the rocket will land successfully
2. For successful landing what efforts SpaceX will have to make to achieve best results.
3. How different parameters such as rocket parameters, location and etc can impact on determining the success rate of a successful landing.

Section 1

Methodology

Methodology

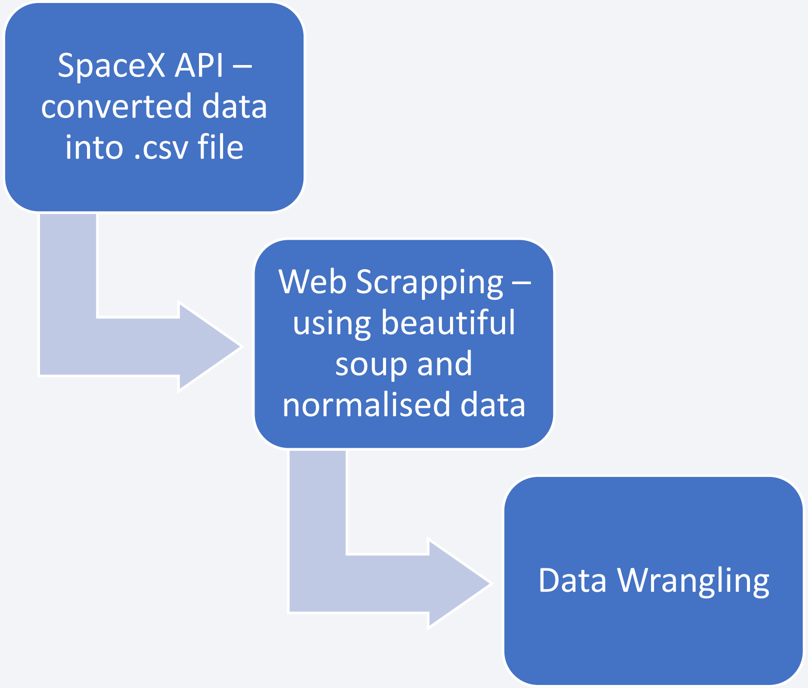
Executive Summary

- Data collection methodology:
  - Data was collecting from past SpaceX missions. From below URL

    https://api.spacexdata.com/v4/launches/past
  - Web scrapping from Wikipedia
- Perform data wrangling
  - Calculated the number of launches on each site
  - Calculated number and occurrences of each orbit
  - Calculated the number and occurrence of mission outcome per orbit type
  - Created a landing outcome label from Outcome column
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

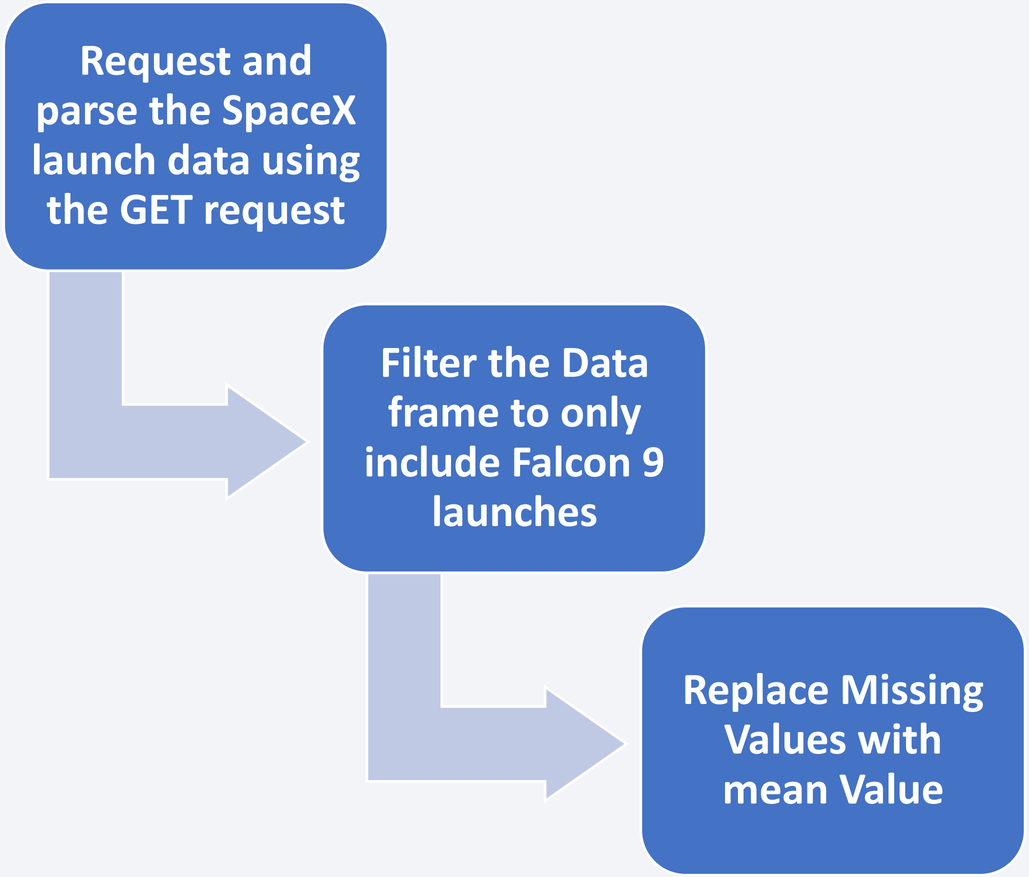
Data Collection

- Data sets were collected from previous SpaceX mission and Wikipedia pages and below processes were obtained to clean the data.



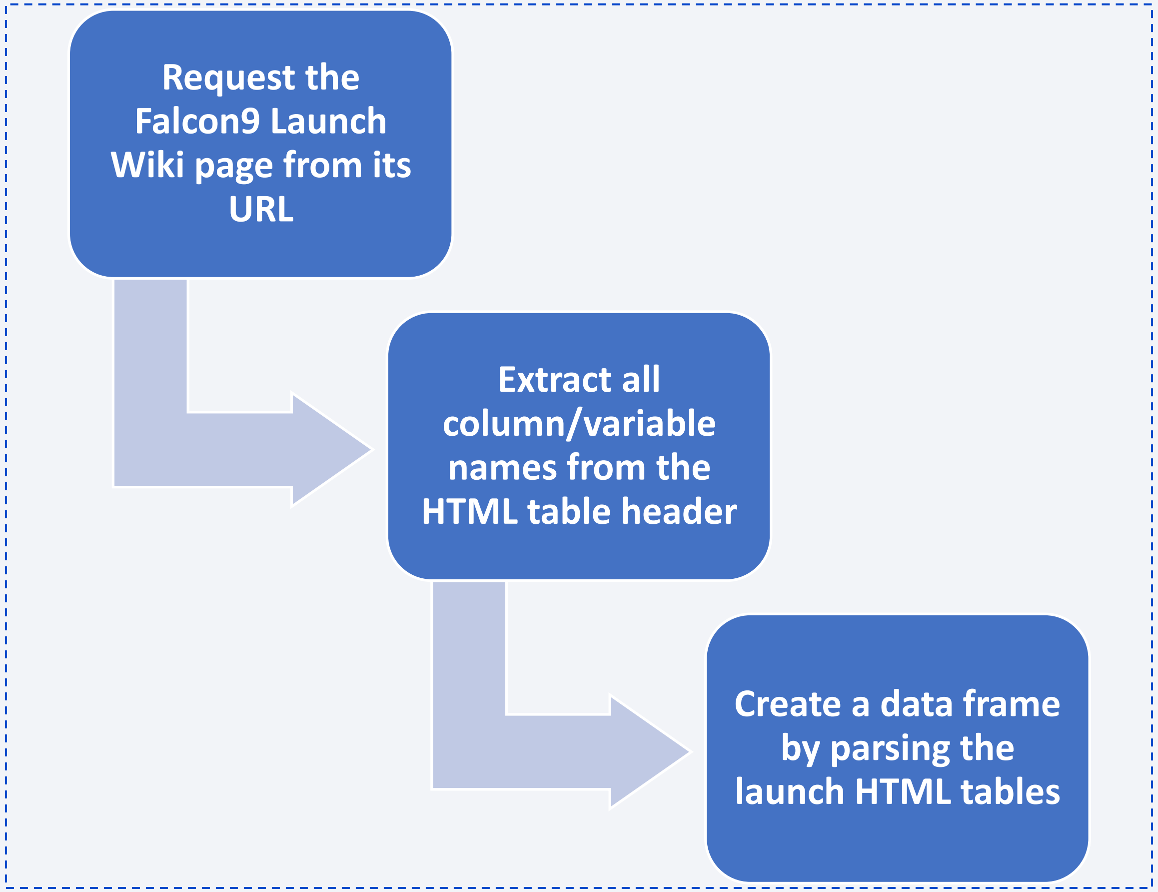
Data Collection – SpaceX API

- https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/jupyter-labs-spacex-data-collection-api.ipynb



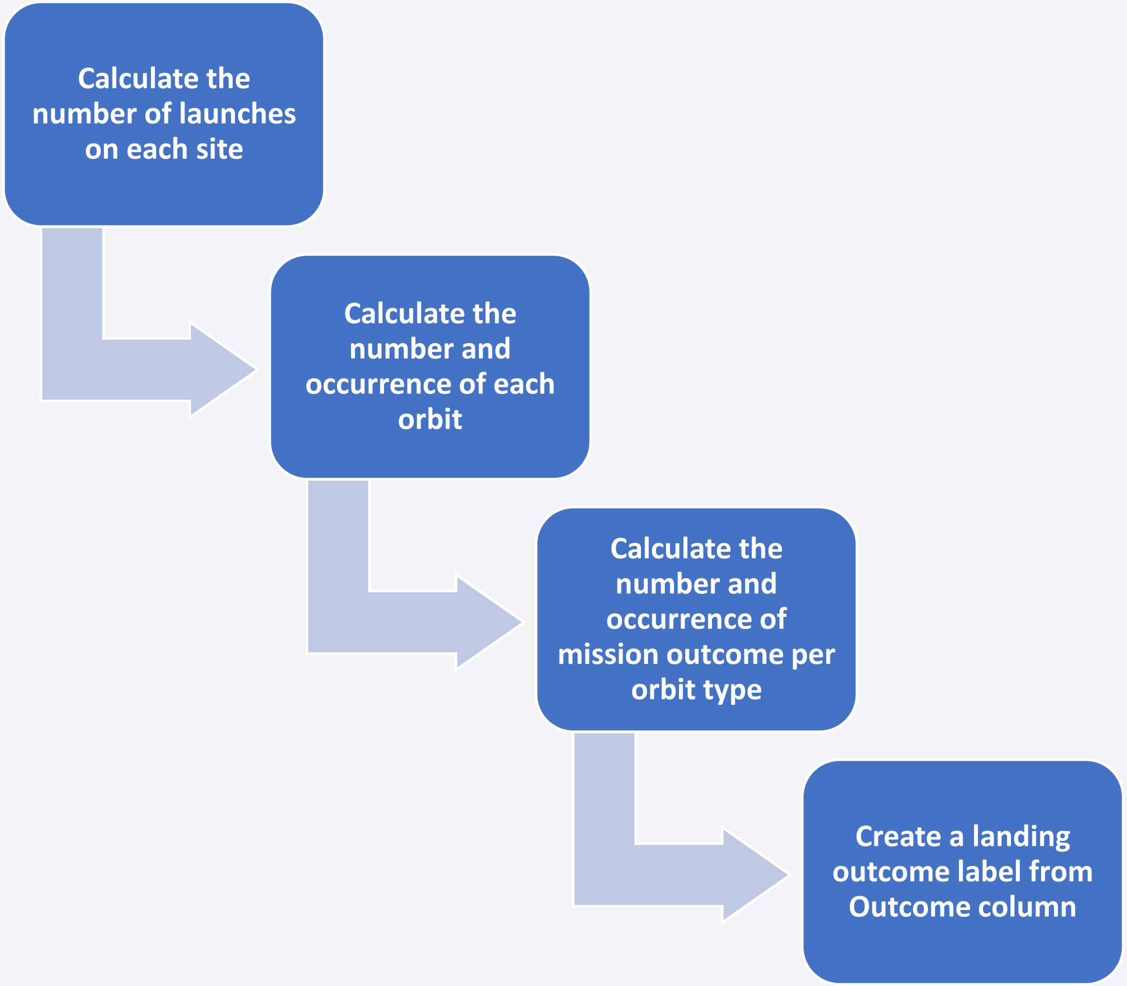
Data Collection - Scraping

- https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/jupyter-labs-webscraping.ipynb



Data Wrangling

https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/labs-jupyter-spacex-Data%20wrangling.ipynb



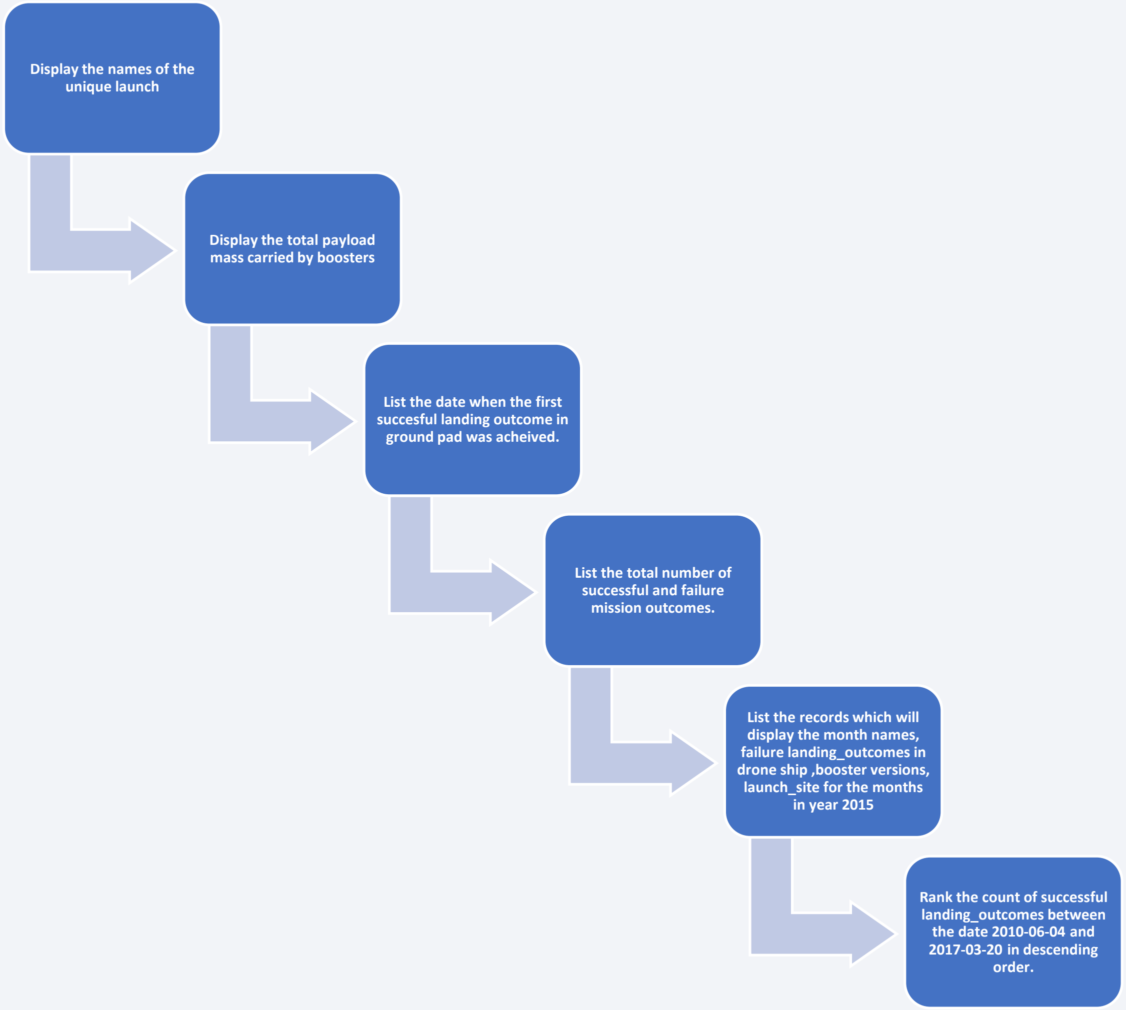
EDA with Data Visualization

- Plotted below charts to understand the data.
  1. Flight Number Vs Payload variables, how it would affect the launch outcome.
  2. Flight Number vs Launch Site
  3. Payload vs Launch Site
  4. Success rate vs Orbit type
  5. Flight Number vs Orbit type
  6. Payload vs Orbit type
  7. launch success yearly trend

https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

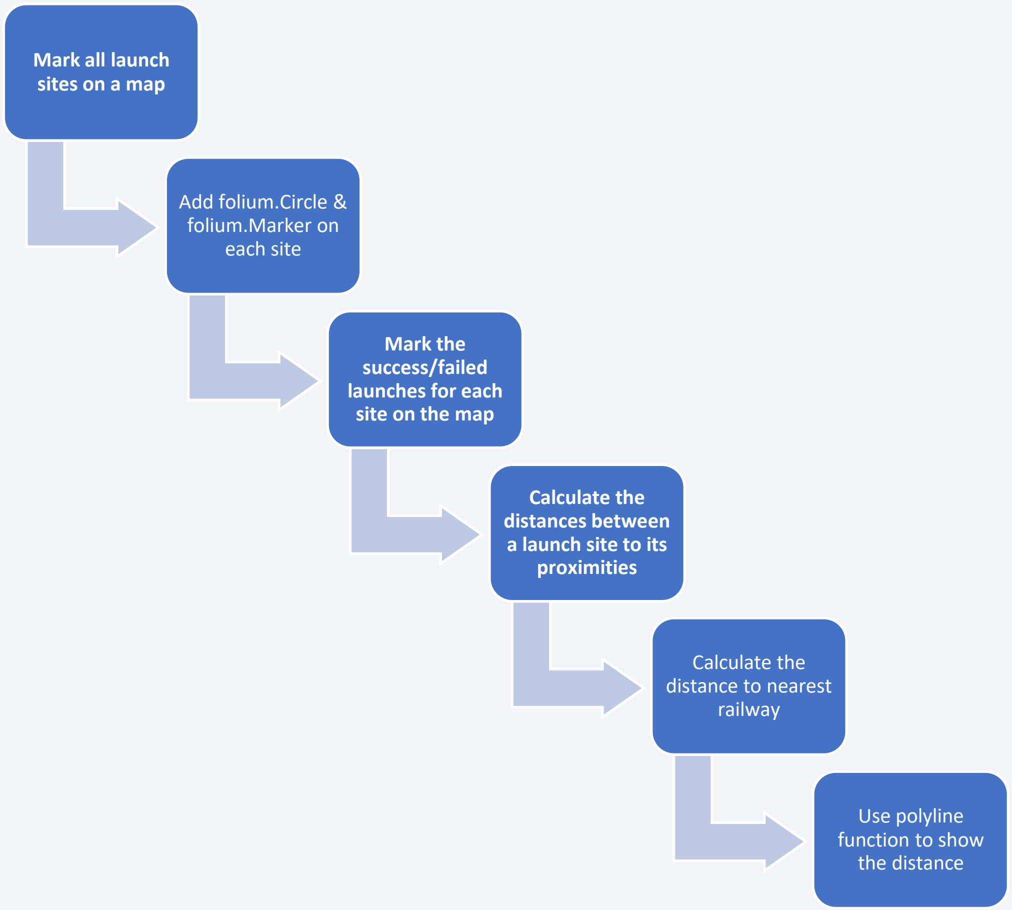
https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/jupyter-labs-eda-sql-coursera.ipynb



Build an Interactive Map with Folium

- The launch success rate may depend on many factors such as payload mass, orbit type, and so on. It may also depend on the location and proximities of a launch site, i.e., the initial position of rocket trajectories. Finding an optimal location for building a launch site certainly involves many factors.

https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/lab_jupyter_launch_site_location%20with%20Folium.ipynb



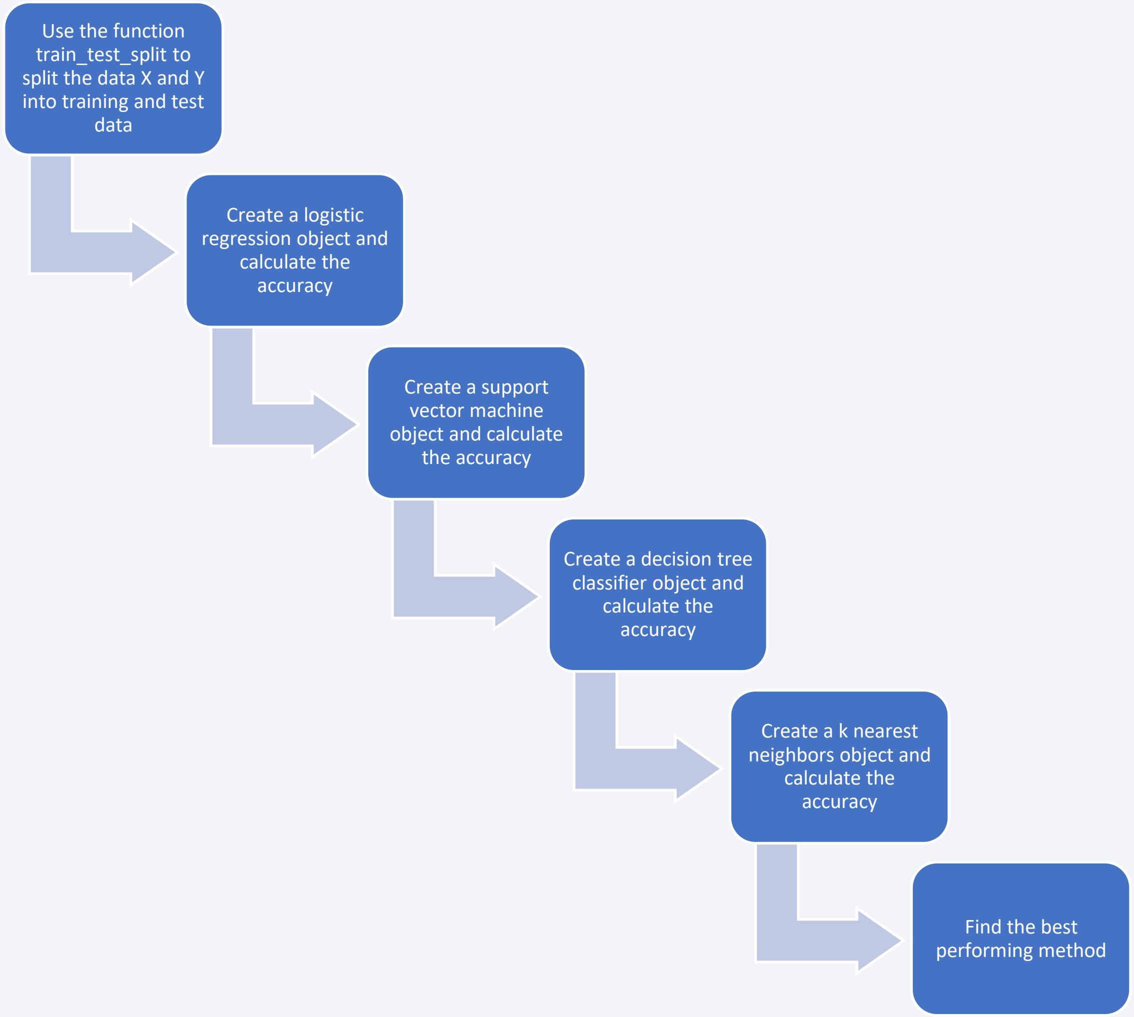
Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Split the data for training and testing, and performed regression analysis methods such as SVM, Classification trees and Logistic Regression to find the best hyperparameters and accuracy of the model.

https://github.com/jaydeep-j/SpaceX_Capstone_Project/blob/master/SpaceX%20Machine%20Learning%20Predictive%20Analysis.ipynb



Results

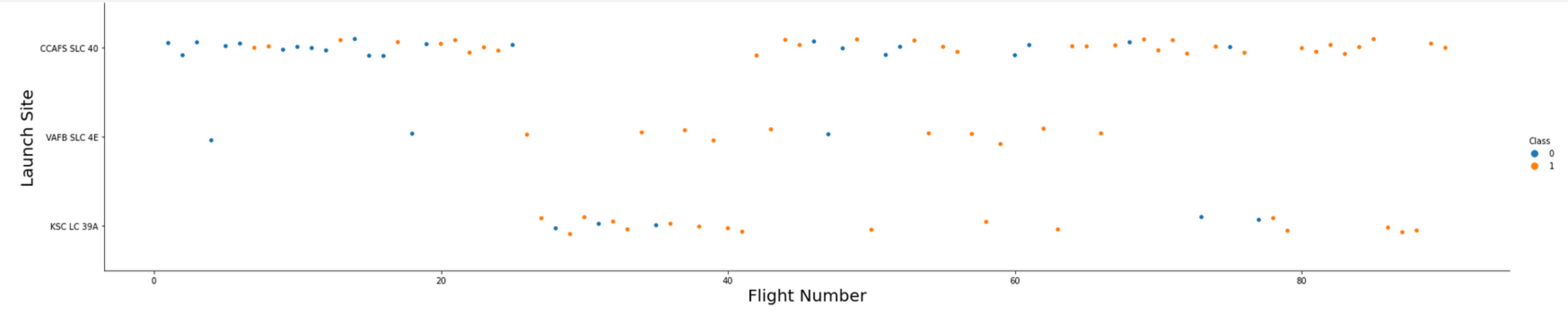
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

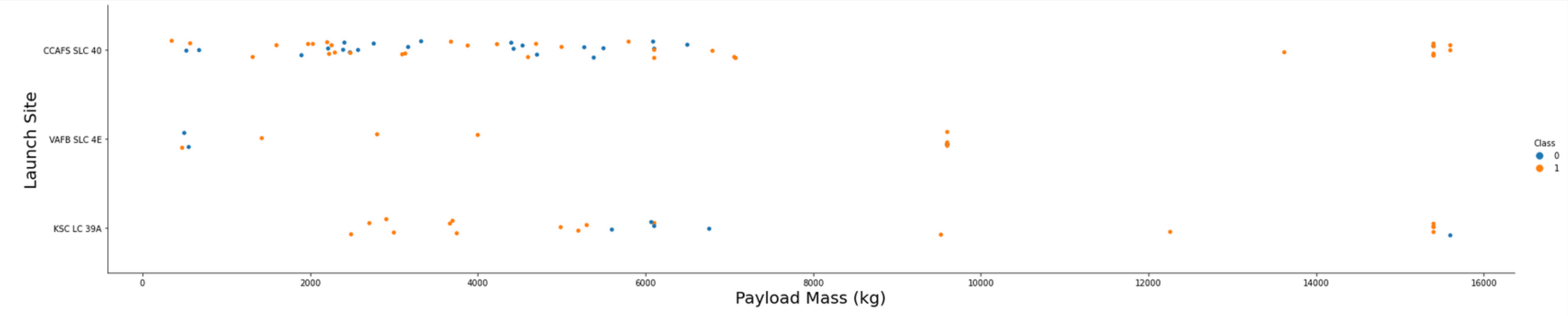
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site
- Explanation: The more amount of flights at a launch site the greater the success rate at a launch site.



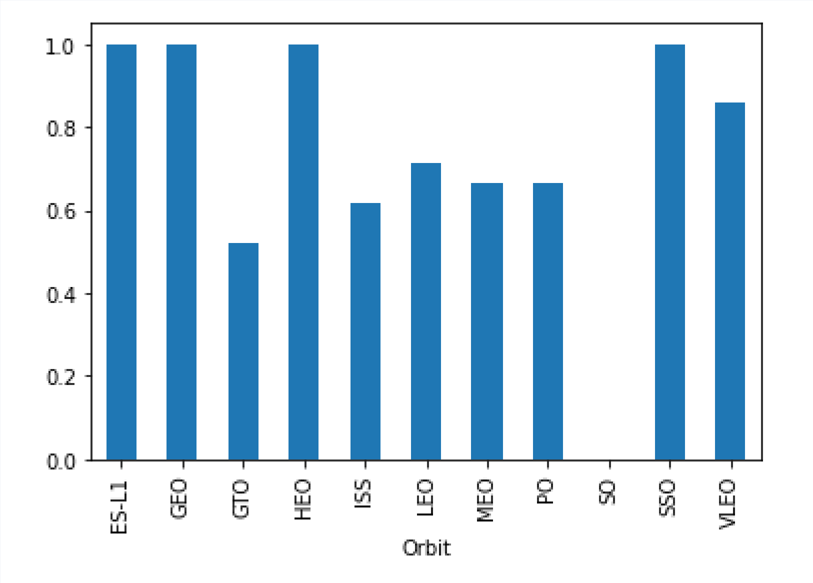
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- Explanations: The greater the payload mass for Launch Site CCAFS SLC 40 the higher the success rate for the Rocket. There is not quite a clear pattern to be found using this visualization to make a decision if the Launch Site is dependant on Pay Load Mass for a success launch.



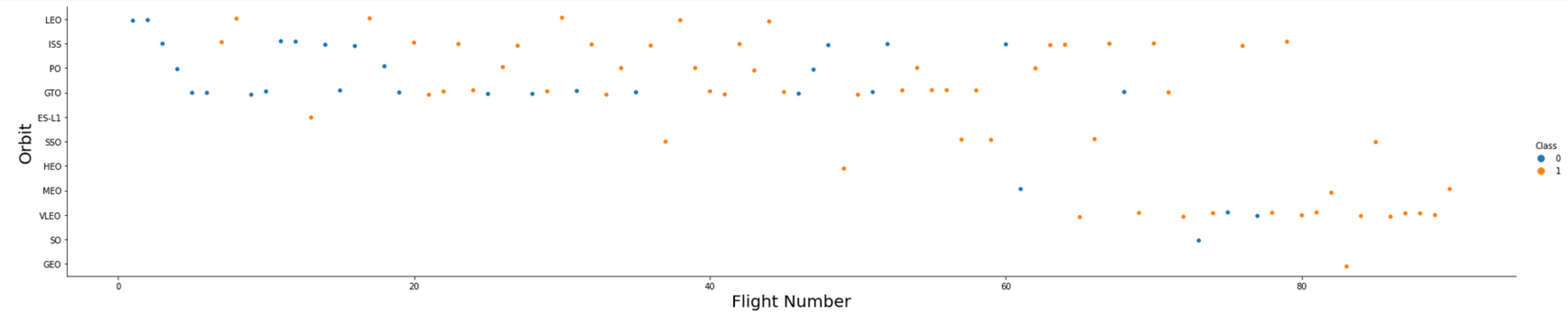
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type
- Explanations: Orbit GEO,HEO,SSO,ES-L1 has the best Success Rate



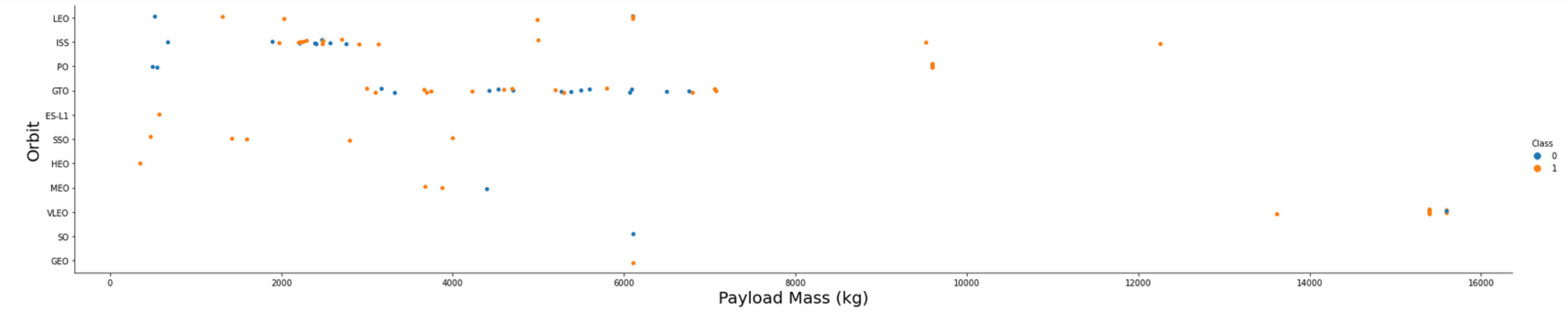
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type
- Explanations : You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



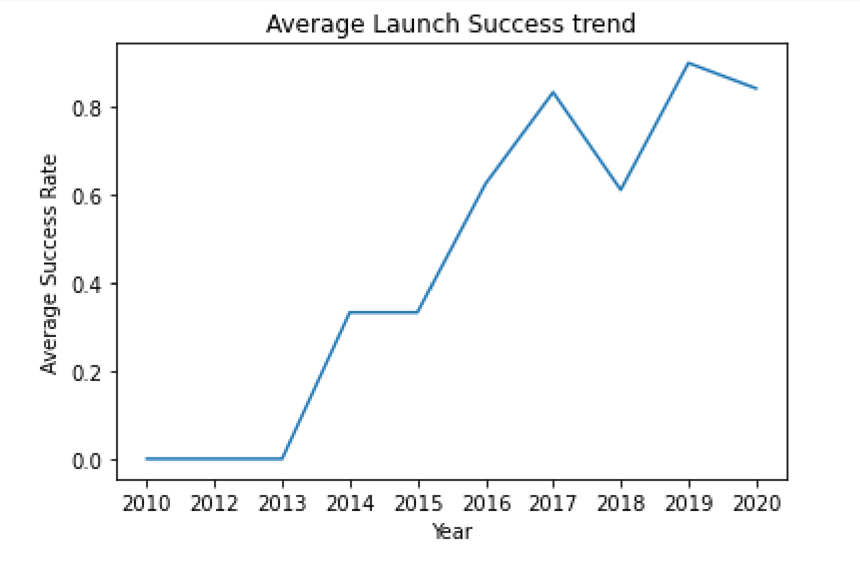
Payload vs. Orbit Type

- Scatter point of payload vs. orbit type
- Explanations: You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



Launch Success Yearly Trend

- Line chart of yearly average success rate.
- Explanations : It can be observed that the success rate kept increasing since 2013 to 2020.



All Launch Site Names

- Names of the unique launch sites:
    1. Cape Canaveral Space Launch Complex 40 VAFB SLC 4E
    2. Vandenberg Air Force Base Space Launch Complex 4E (SLC-4E)
    3. Kennedy Space Center Launch Complex 39A KSC LC 39A
- Query explanation: Using the word DISTINCT in the query means that it will only show Unique values in the Launch_Site column from tblSpaceX

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin 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

Total Payload Mass

- Total payload carried by boosters from NASA
- Explanation : Using the function SUM summates the total in the column PAYLOAD_MASS_KG_, The WHERE clause filters the dataset to only perform calculations on Customer NASA (CRS)

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
- Explanation : Using the function AVG works out the average in the column PAYLOAD_MASS_KG_

average_payload_mass
2928

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad
- Explanation : Using the function MIN works out the minimum date in the column Date

first_succesful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Explanation : Selecting only Booster_Version The WHERE clause filters the dataset to Landing_Outcome = Success (drone ship) The AND clause specifies additional filter conditions Payload_MASS_KG_ > 4000 AND Payload_MASS_KG_ < 6000

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass
- Explanation : Using the word DISTINCT in the query means that it will only show Unique values in the Booster_Version column from tblSpaceX GROUP BY puts the list in order set to a certain condition.DESC means it's arranging the dataset into descending order

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600
F9 B5 B1049.6	15440
F9 B5 B1059.3	15410
F9 B5 B1051.5	14932

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

MONTH	landing__outcome	booster_version	launch_site
1	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
4	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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, in descending order
- Explanation : Function COUNT counts records in column WHERE filters data LIKE (wildcard), AND (conditions), AND (conditions)

landing__outcome	2
Success (ground pad)	3
Success (drone ship)	5

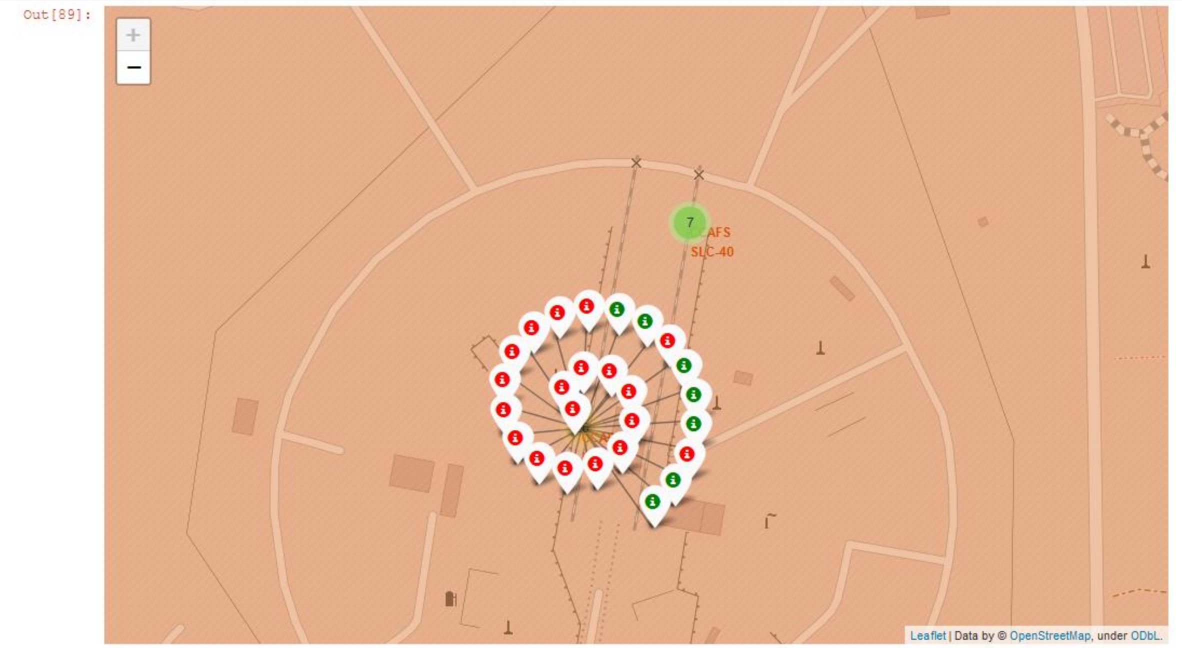
Section 4

Launch Sites Proximities Analysis

All Launch Sites on Map

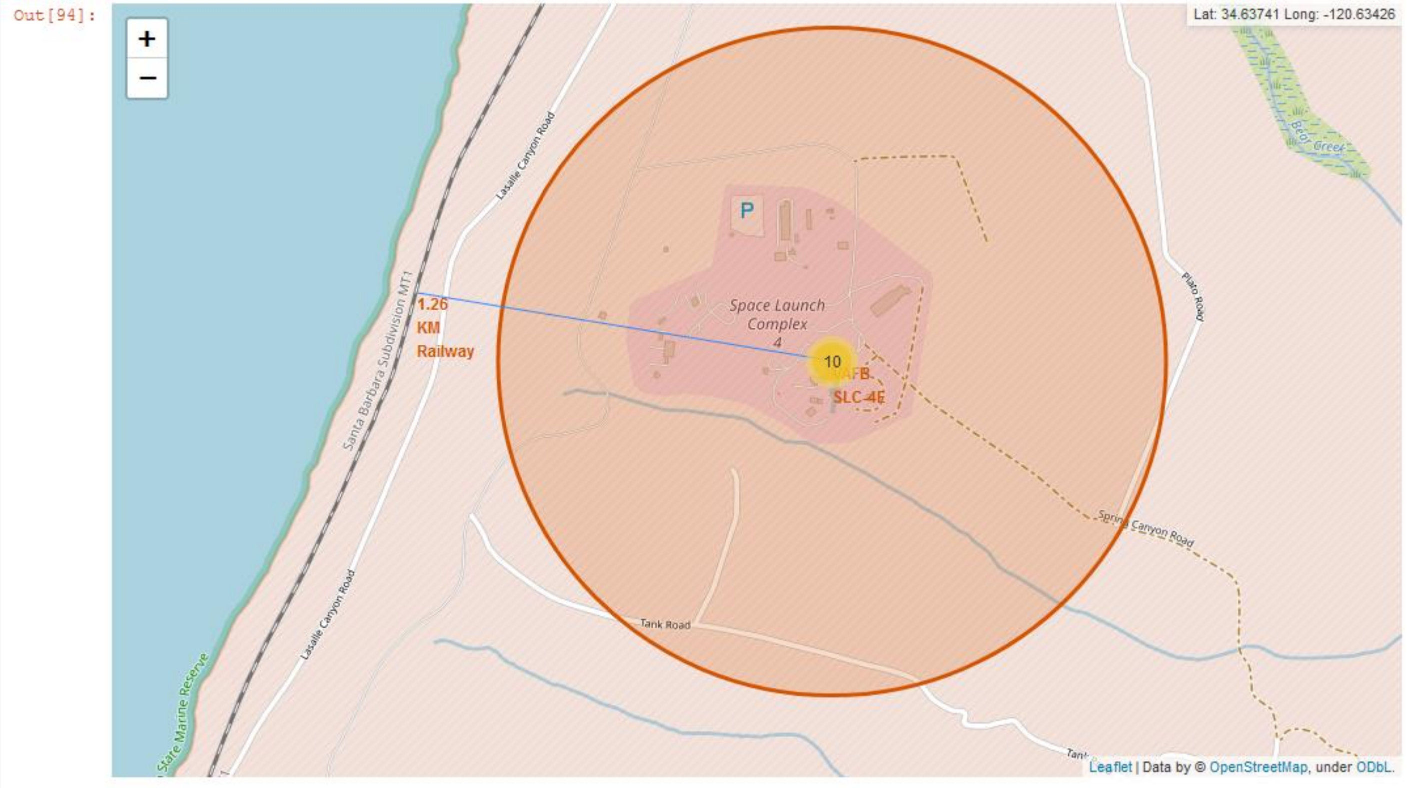
- We can see that the SpaceX launch sites are in the United States of America coasts. Florida and California.

Success/failed launches for each site on the map



- Green Marker shows successfulll launches and Red marker shows failures.

Distance from the Coastline

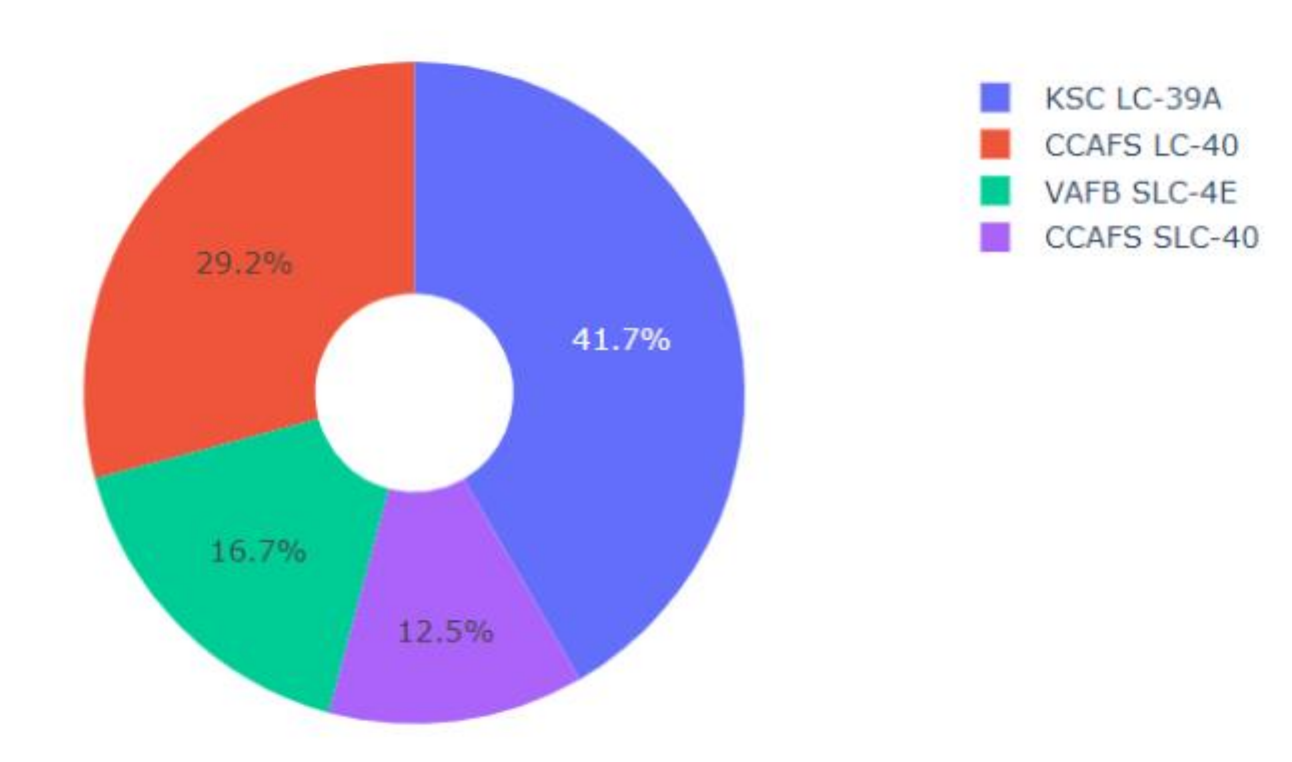


- We can see that the launch site is 1.26KM away from the coastline.

Section 5

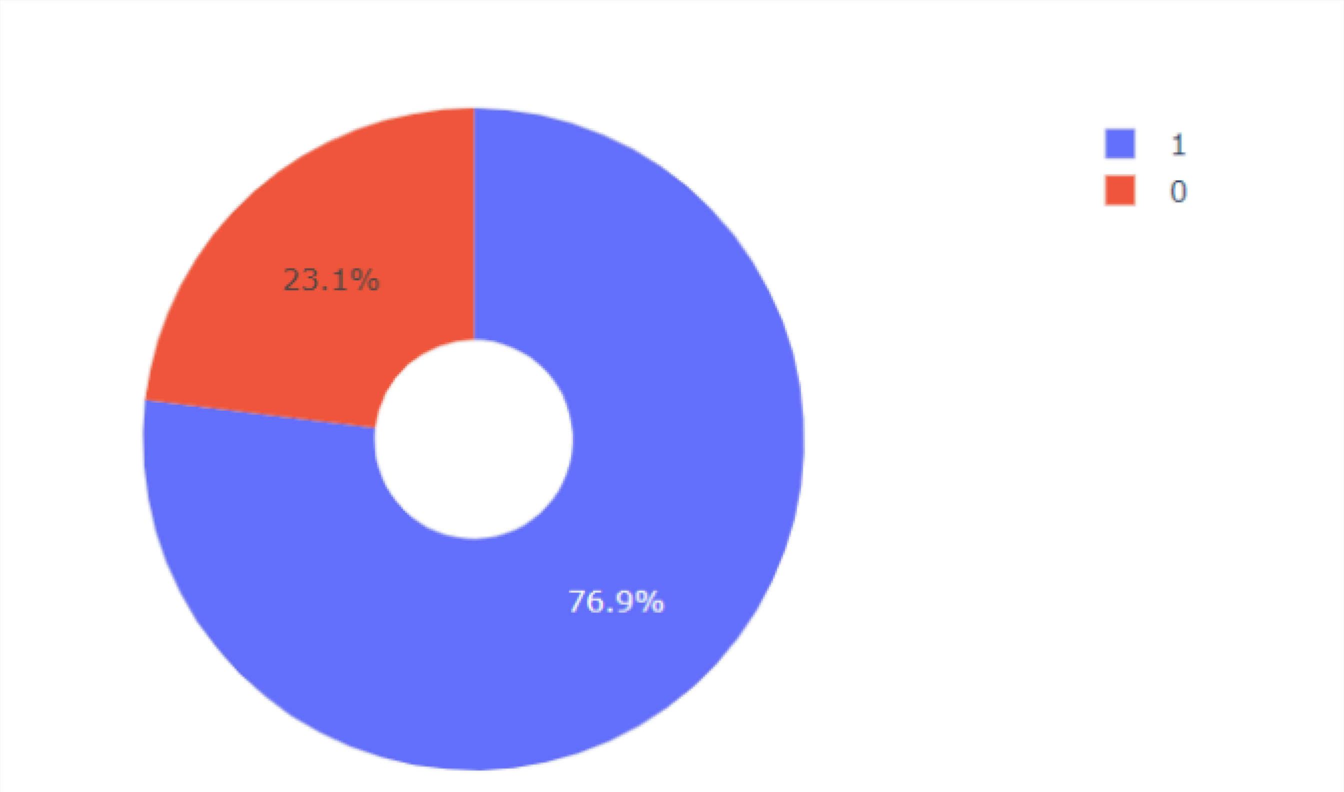
Build a Dashboard with Plotly Dash

Pie chart showing the success percentage achieved by each launch site



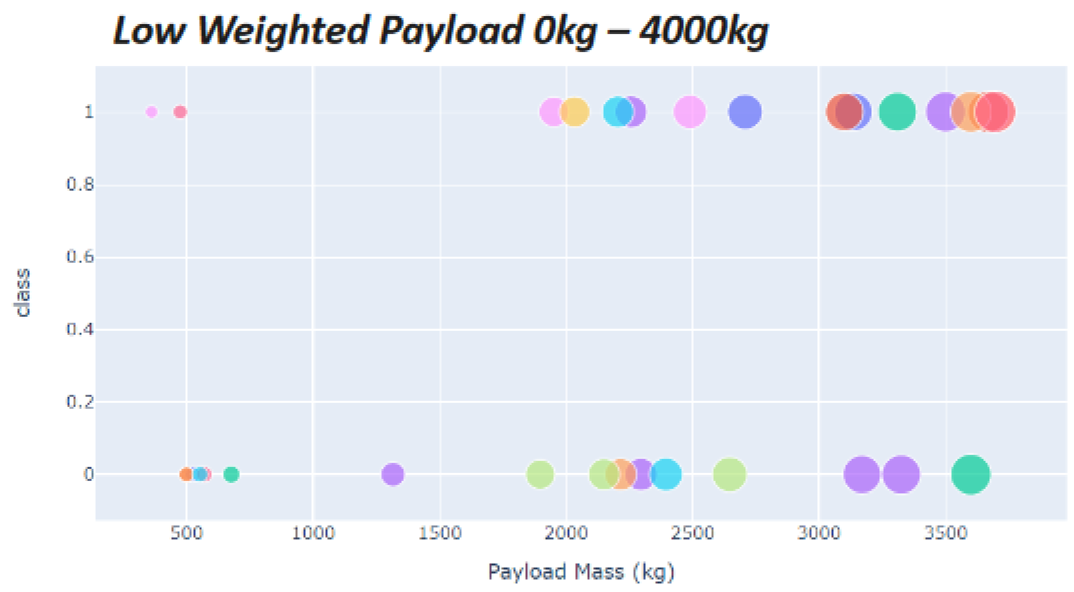
- We can see that KSC LC-39A had the most successful launches from all the sites

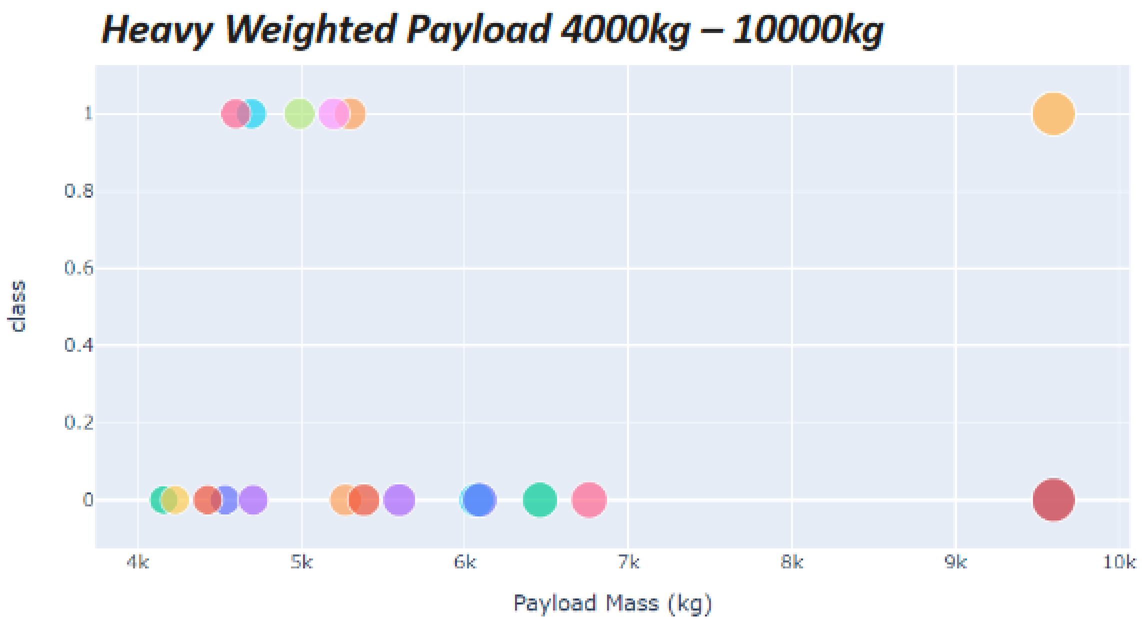
Pie chart for the launch site with highest launch success ratio



- KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider



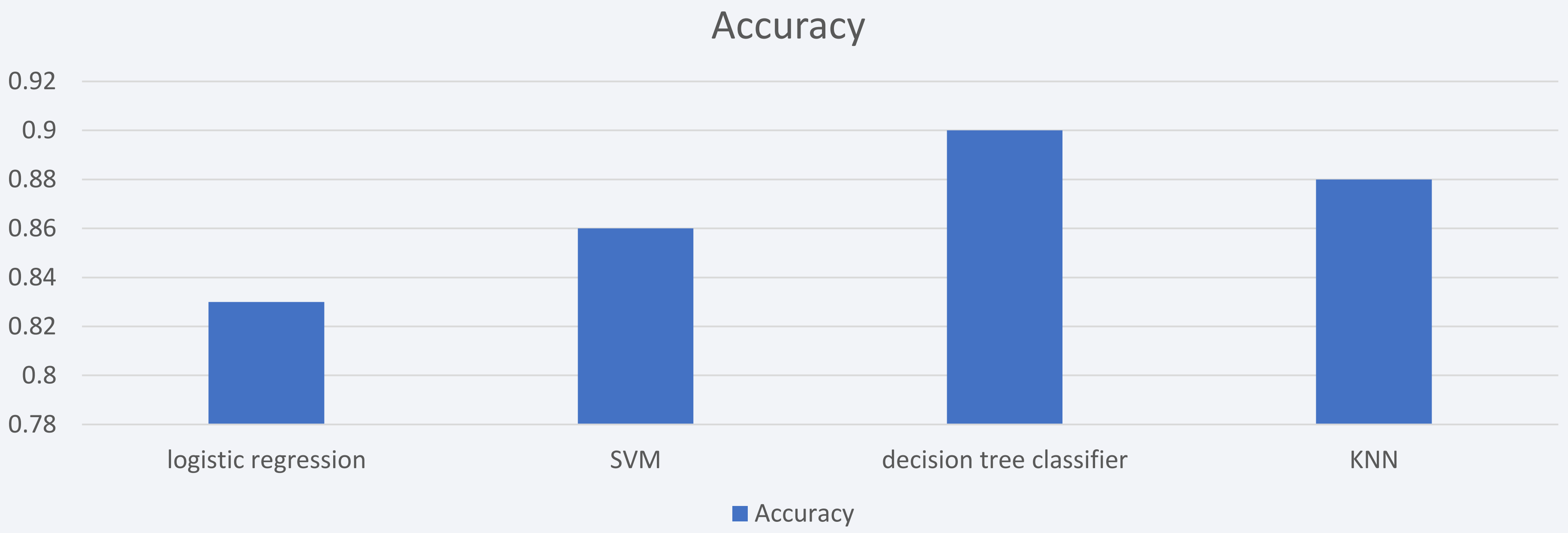


- We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Section 6

Predictive Analysis (Classification)

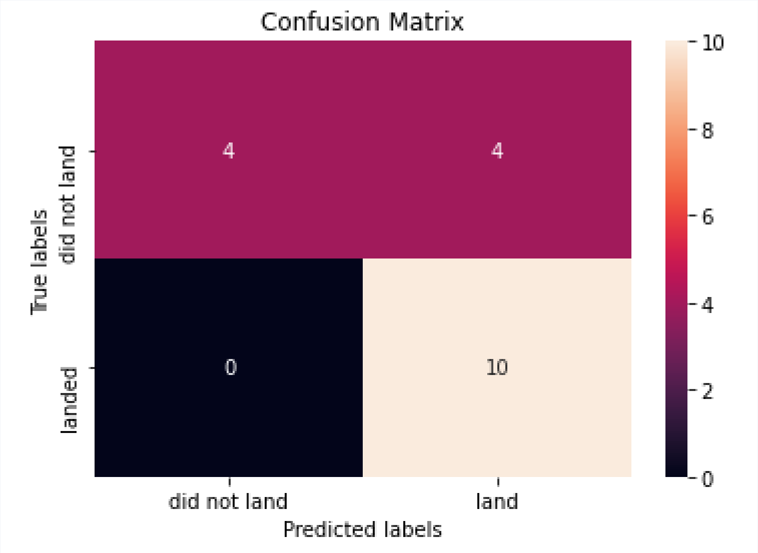
Classification Accuracy



- Decision tree classifier has the best accuracy.

Confusion Matrix

- Confusion matrix of the best performing model with an explanation



Conclusions

- The Tree Classifier Algorithm is the best for Machine Learning for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- We can see that KSC LC-39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES-L1 has the best Success Rate

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!