Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

Introduction

- Project background and context

SpaceX advertise the launch of Falcon 9 rockets on its website with a cost of 62 million dollars, while other providers cost up to 165 million dollars

Problems you want to find answers

We aim to discover the probability of successful launch and return of the Falcon 9 as well as make analysis of the launch and company development over the years

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The methods utilized are as follows ; EDA with Data Visualization ,EDA with SQL ,Build an Interactive Map with Folium Build ,a Dashboard with Plotly Dash
- Perform data wrangling
  - Describe how data was processed
- 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

- Describe how data sets were collected.

Data will be collected using a flowchart.

- You need to present your data collection process use key phrases and flowcharts

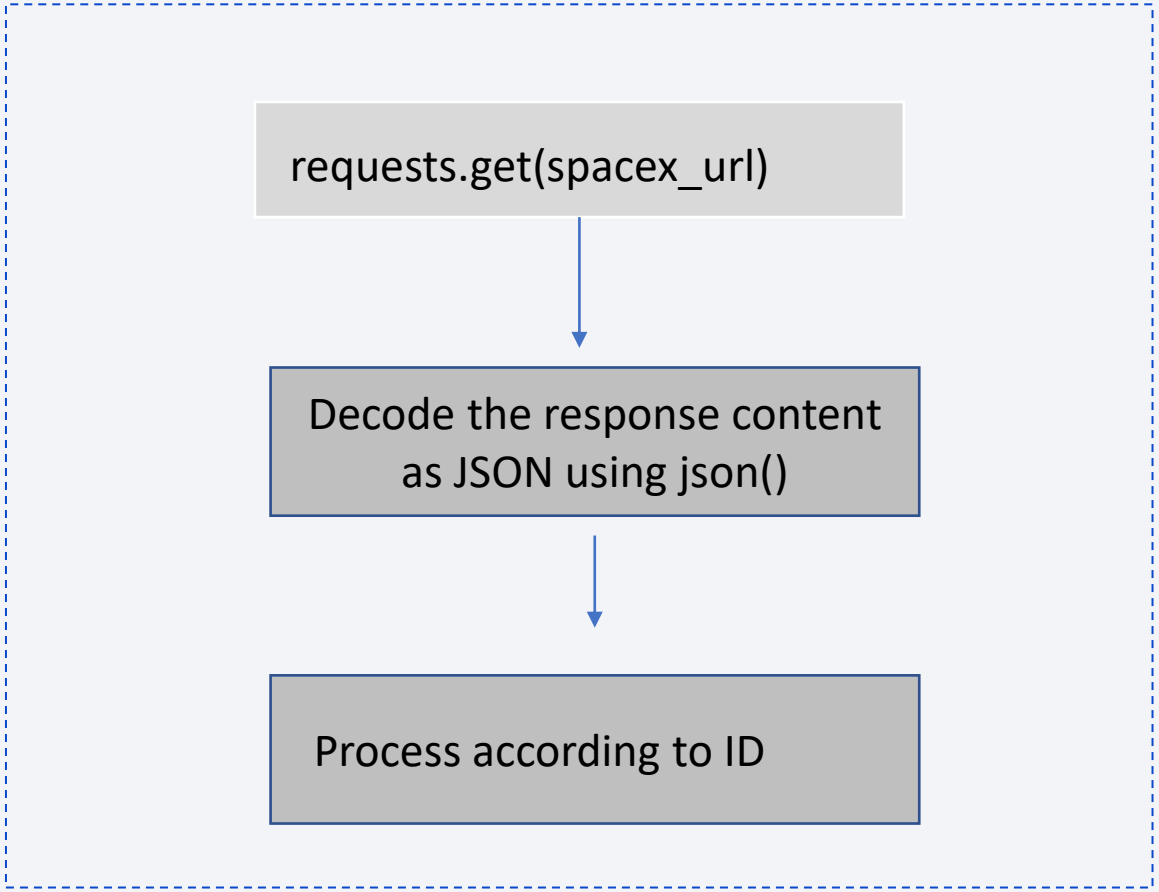
https://github.com/emeldam/Applied-Data-science-capstone/blob/1c0063b62575793dcc2029cc732a1ce17d5c225b/Lab%203%20:%20Data%20Wrangling%20.ipynb

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

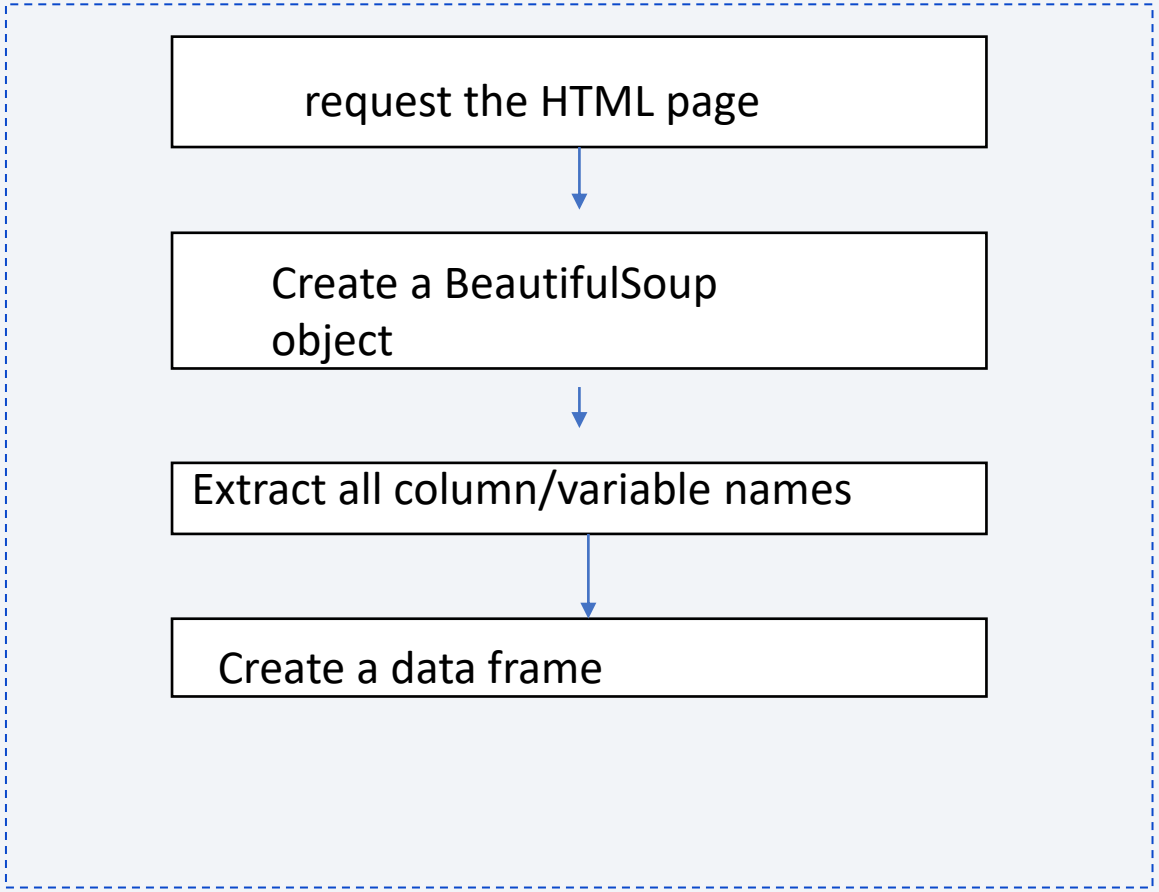
GitHub URL ：

https://github.com/emeldam/Applied-Data-science-capstone/blob/1c0063b62575793dcc2029cc732a1ce17d5c225b/Lab%201%20-%20Collecting%20Data



Data Collection - Scraping

GitHub URL :
https://github.com/emeldam/Applied-Data-science-capstone/blob/1c0063b62575793dcc2029cc732a1ce17d5c225b/Lab%202%20:%20Data%20Collection%20with%20Web%20Scrapping%20.ipynb

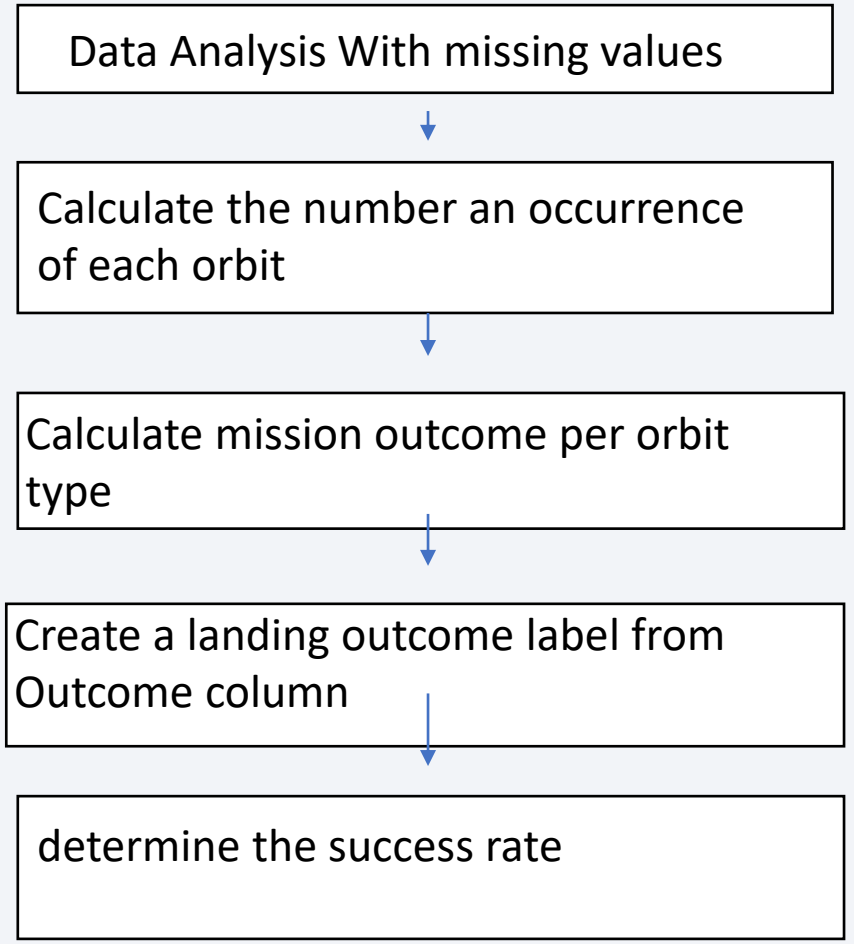


Data Wrangling

Describe how data were processed

Identify and calculate the percentage of the missing values in each attribute .And found that the landing pad column is missing. Hence, We execute the flow chart on the right to make up for the missing values ,data wrangling process using key phrases and flowcharts

- GitHub URL : https://github.com/emeldam/Applied-Data-science-capstone/blob/1c0063b62575793dcc2029cc732a1ce17d5c225b/Lab%203%20:%20Data%20Wrangling%20.ipynb
-



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

We make use of line chart and scatter plot as these are good way to show relationship

GitHub URL : https://github.com/emeldam/Applied-Data-science-capstone/blob/master/Lab%205%20:EDA%20with%20Data%20Visualization%20Lab

EDA with SQL

- GitHub URL : https://github.com/emeldam/Applied-Data-science-capstone/blob/1c0063b62575793dcc2029cc732a1ce17d5c225b/Lab%205%20:EDA%20with%20Data%20Visualization%20Lab

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard

We make use of line chart and scatter plot as these plots are a good way to show relationship

- GitHub URL:https://github.com/emeldam/Applied-Data-science-capstone/blob/master/Lab%205%20:EDA%20with%20Data%20Visualization%20Lab

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model

We create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

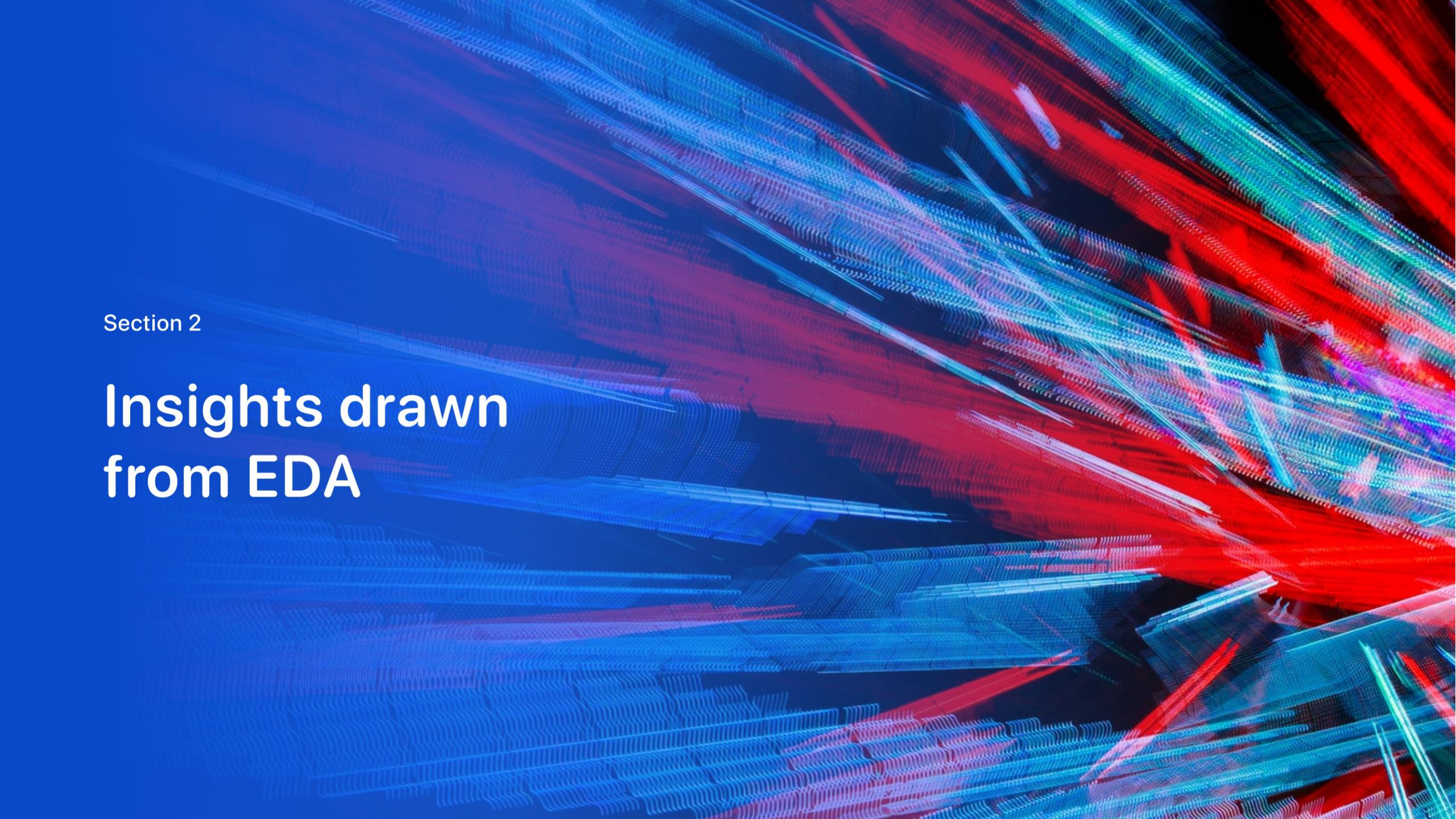
Add the GitHub URL:https://github.com/emeldam/Applied-Data-science-capstone/blob/1c0063b62575793dcc2029cc732a1ce17d5c225b/Lab%207%20:%20Machine%20Learning%20Prediction.ipynb

Results

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

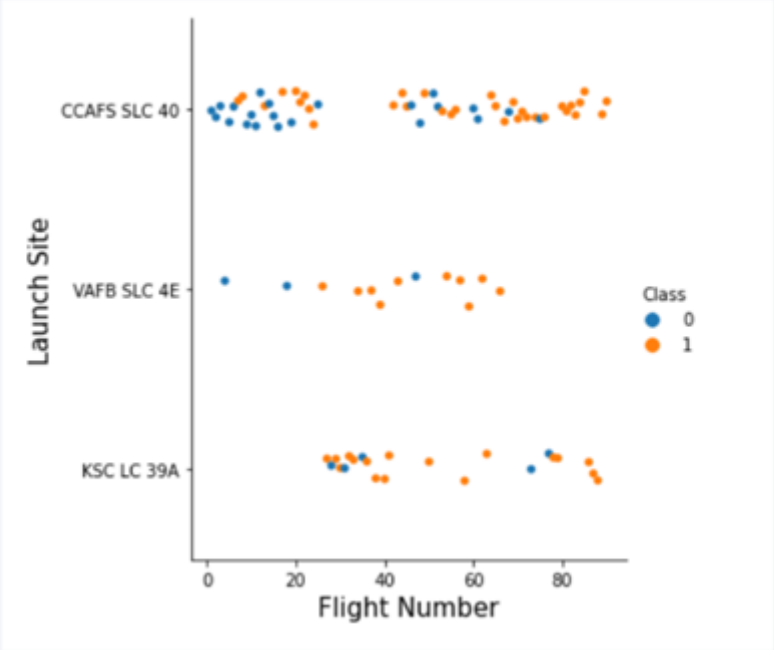
Section 2

Insights drawn from EDA



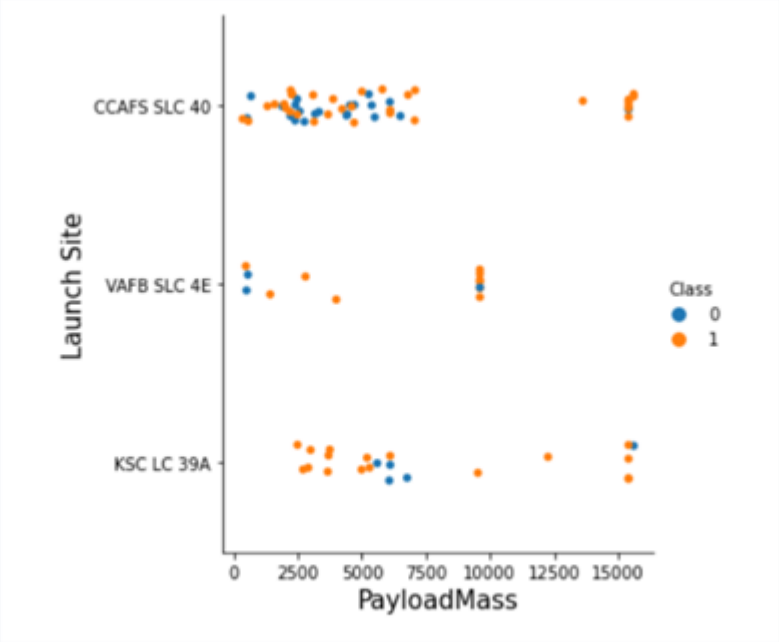
Flight Number vs. Launch Site

- The scatter plot of Flight Number vs. Launch Site
- Based on the diagram below , class represent the classification variable that represents the outcome of each launch .Given the value is 0 , the first stage does not land successfully ; while one means landed successfully



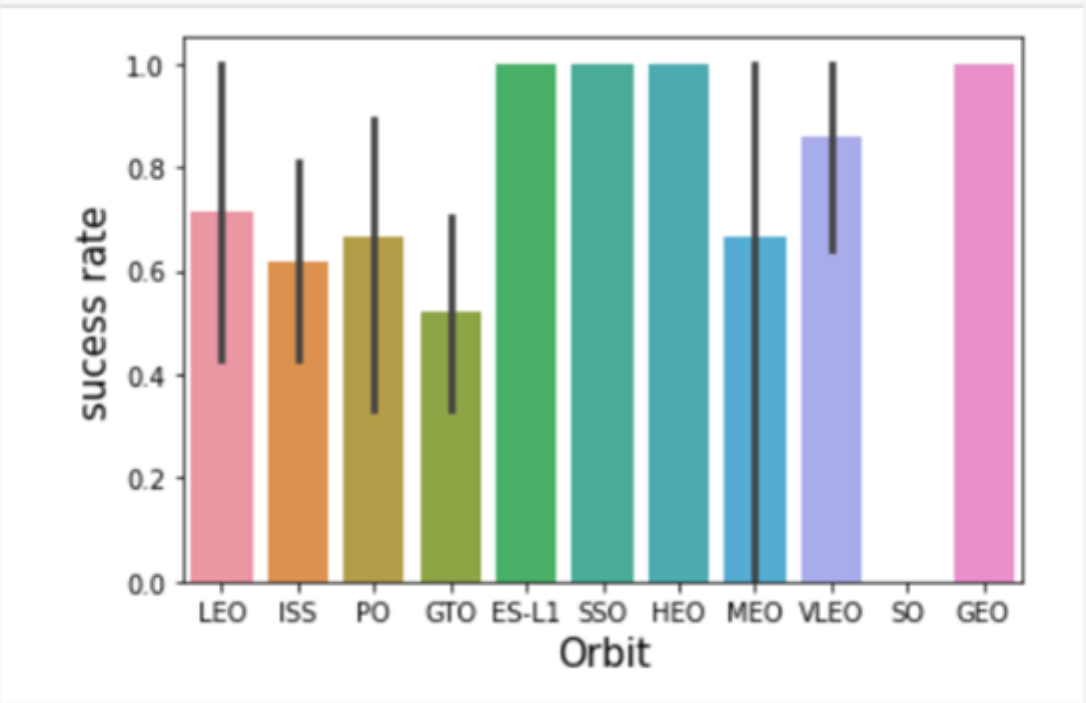
Payload vs. Launch Site

- The Scatter plot of Payload vs. Launch Site
- Below the diagram shows that in VAFB SLC sites ,the rocket launched a smaller load compared to the other two



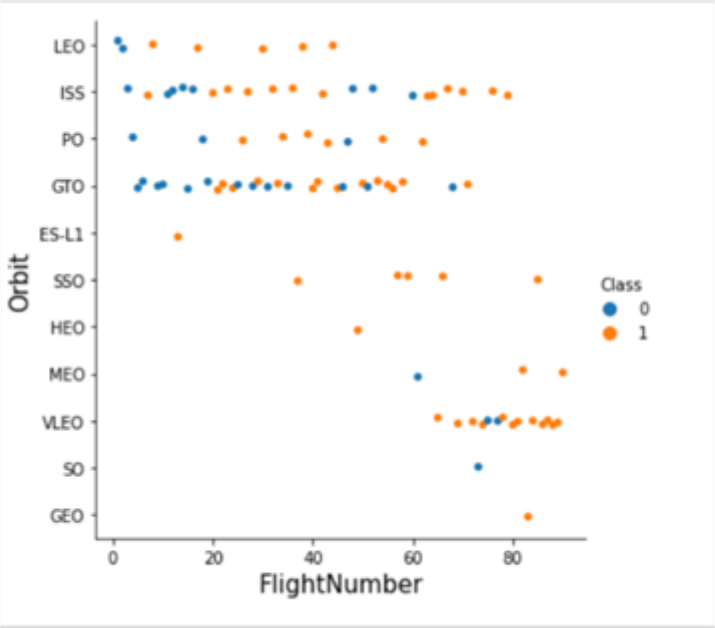
Success Rate vs. Orbit Type

- The bar chart for the success rate of each orbit type
- The results show that Po, ISS and MEO had the lowest success rate while ES L1 SSO HEO GEO had 100% success rate



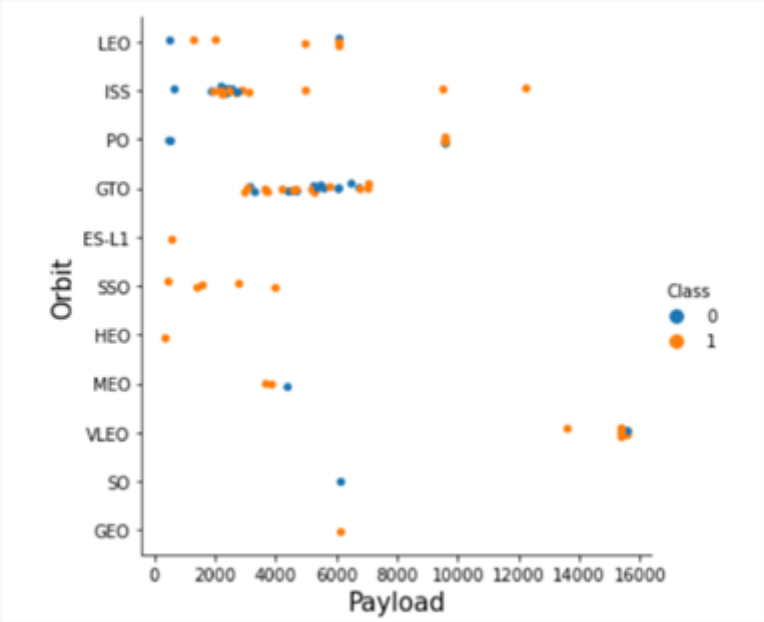
Flight Number vs. Orbit Type

- The scatter point of Flight number vs. Orbit type
- As seen below 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.



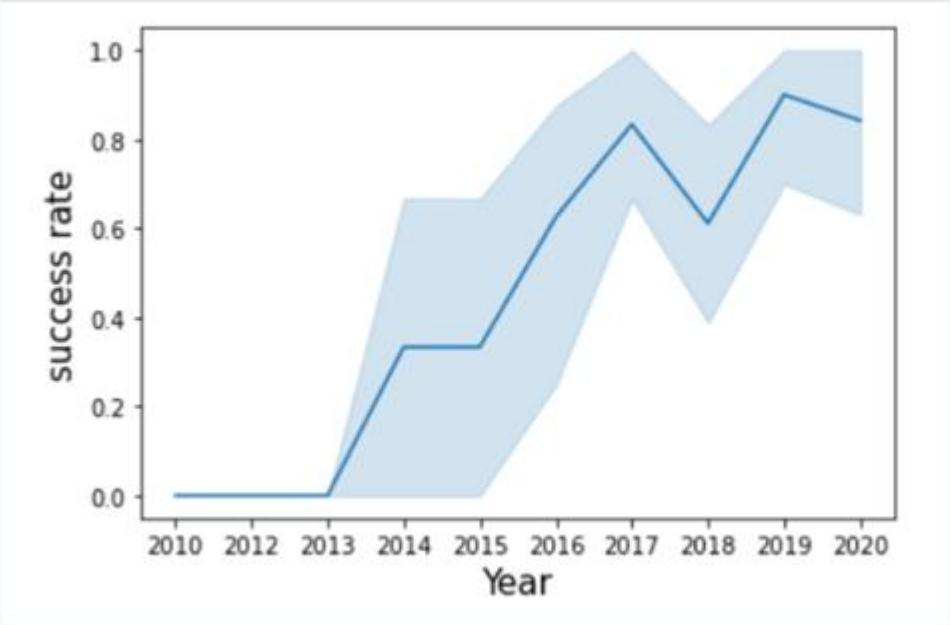
Payload vs. Orbit Type

- The scatter point of payload vs. orbit type
- On GTOs , the heavy payloads have a negative influence while on GTO and Polar LEO (ISS) obits they have a negative influence



Launch Success Yearly Trend

- The line chart of yearly average success rate
- As seen below they has been an increase since 2013 to 2020



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`
- To Output the below make use of this : Use sql : SELECT * FROM Spacex WHERE launch site LIKE 'CCA%' LIMIT 5

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

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- To Output the below make use of USE SQL : SELECET DISTINCT booster version FROM Spacext WHERE landing Outcome LIKE '%Success%' AND payload mass_kg_>4000 AND payload_mass_kg_<6000

booster_version
F9 B4 B1040.1
F9 B4 B1043.1
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5 B1058.2
F9 B5B1060.1
F9 B5B1062.1
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026
F9 FT B1032.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

2015 Launch Records

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

- Present your query result with a short explanation here

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

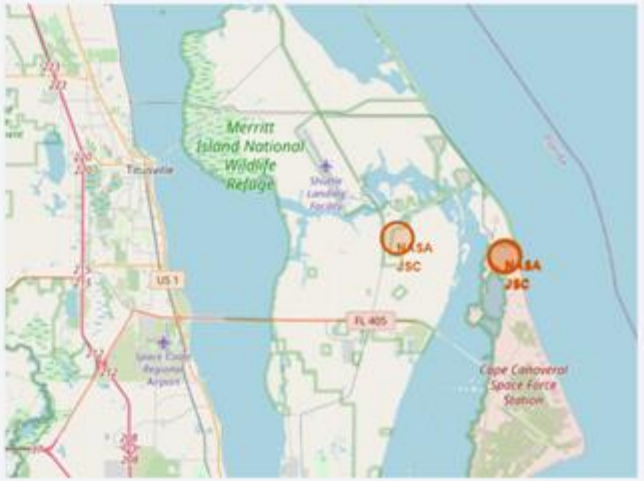
- Present your query result with a short explanation here

Section 4

Launch Sites Proximities Analysis

<ALL LAUNCE SITES ON MAP>

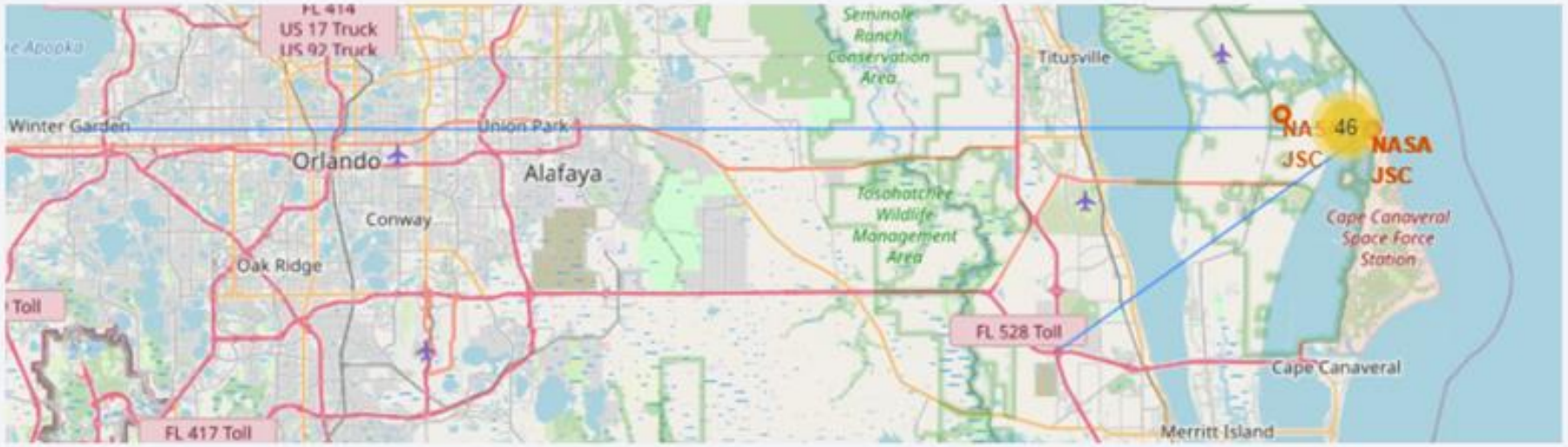
- They are 4 launch sites , with 3 closely located to each other



<Failed Launch Sites >

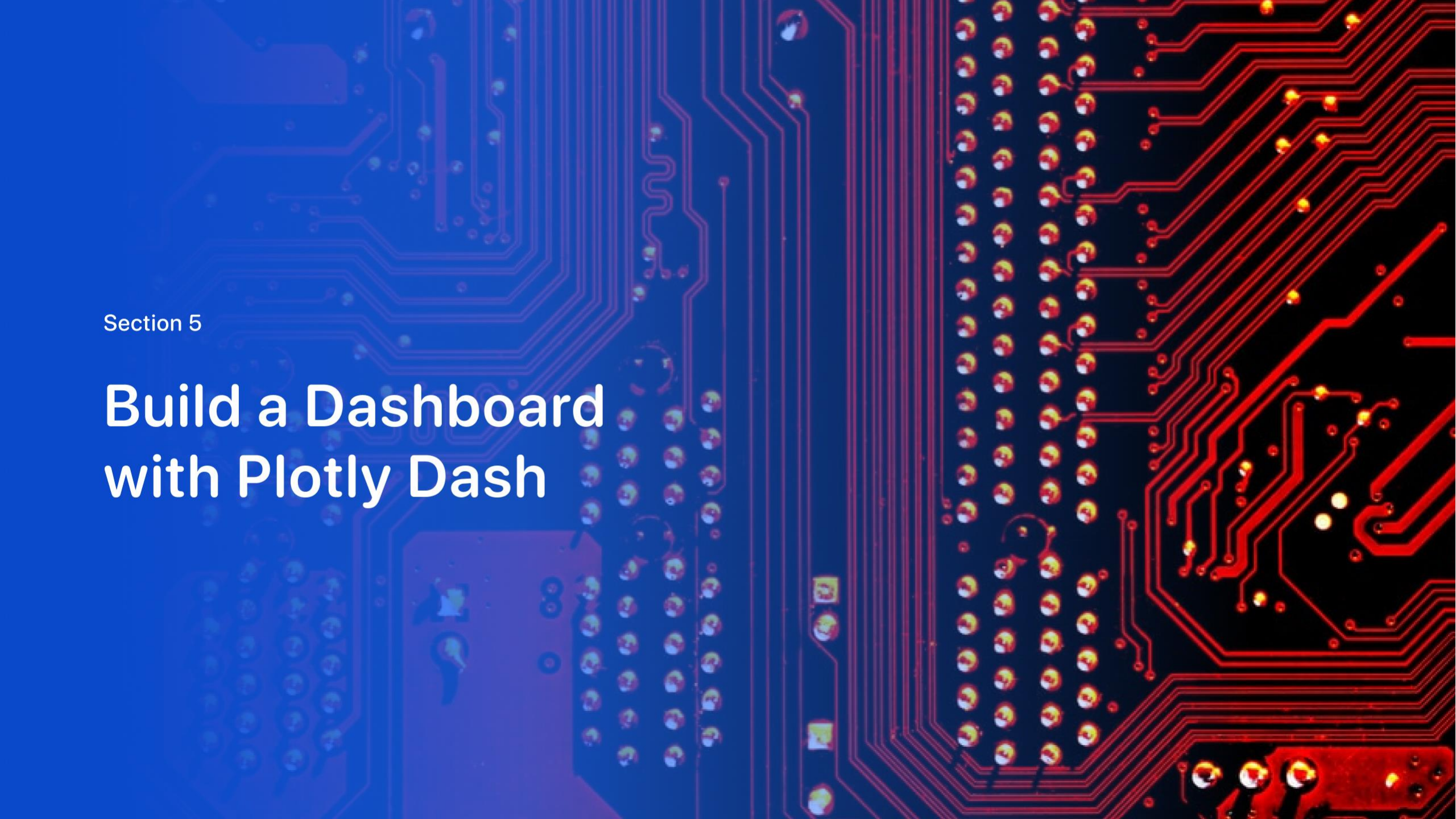
- As seen below , the distribution of each number.

<Calculation of the distance between launch sites to its proximities >



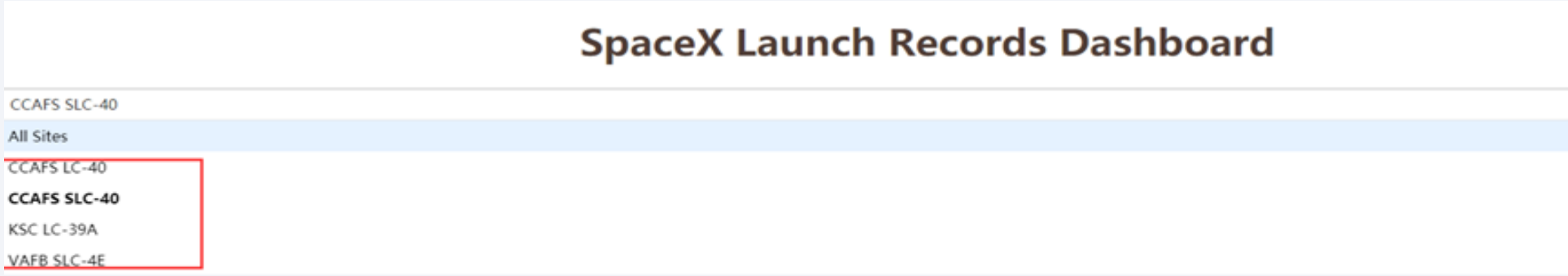
Section 5

Build a Dashboard with Plotly Dash

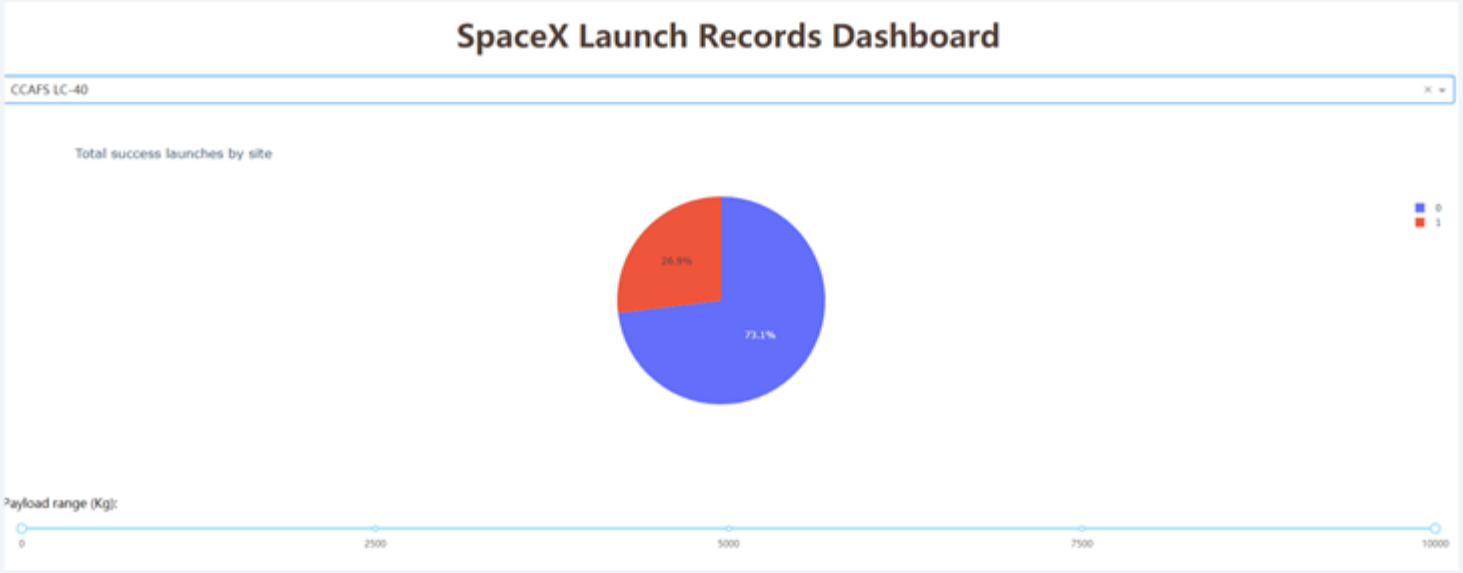


<Adding a Launch Site Drop-down Input Component>

With four different launch sites, we aim to find the one with the largest success count as well as select one specific site to check its detailed success rate (class=0 vs. class=1).

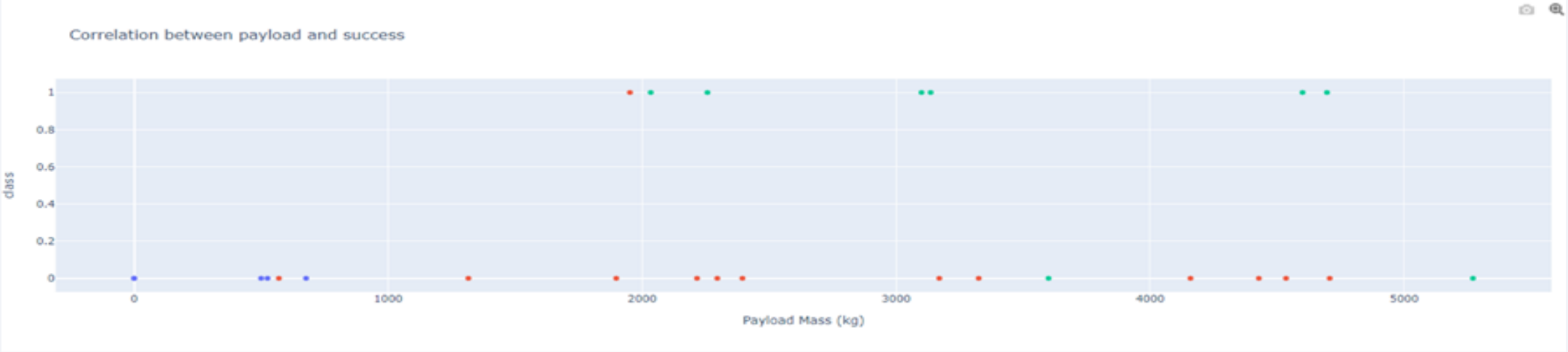


<Success pie chart on selected site dropdowns >



<Success-payload-scatter-chart scatter plot>

- Below is a scatter plot whereby x axis is the payload, and the y axis is the launch outcome (i.e., class column). Hence, we can visually observe how payload may be correlated with mission outcomes for selected site(s).

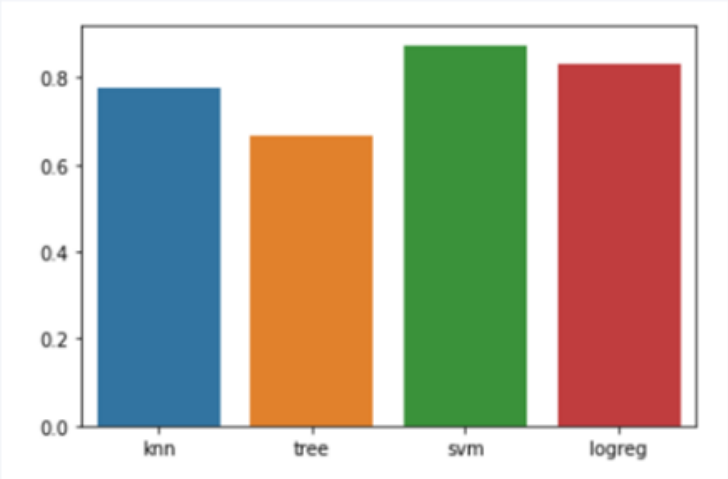


Section 6

Predictive Analysis (Classification)

Classification Accuracy

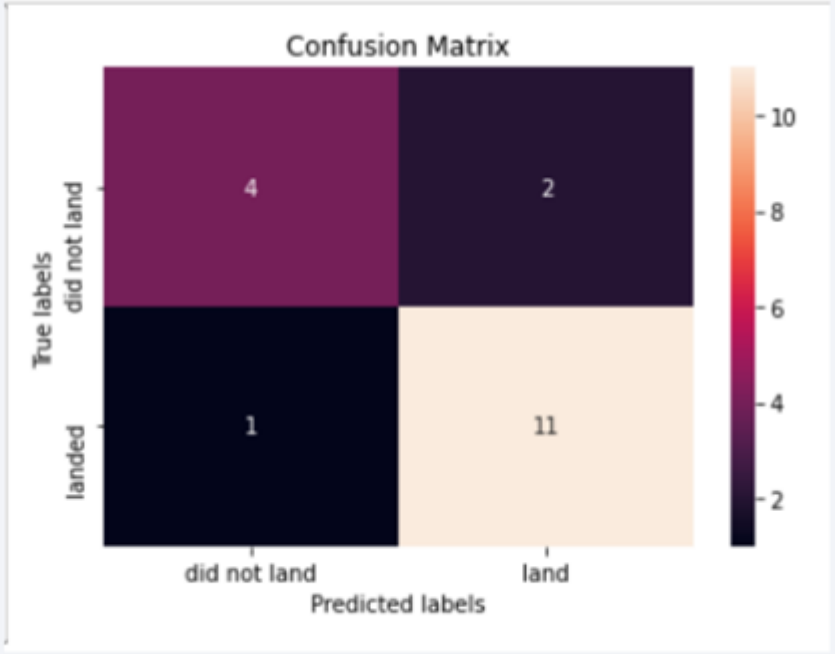
- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy



- The diagram shows the SVM model with the highest classification accuracy of 0.833%

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

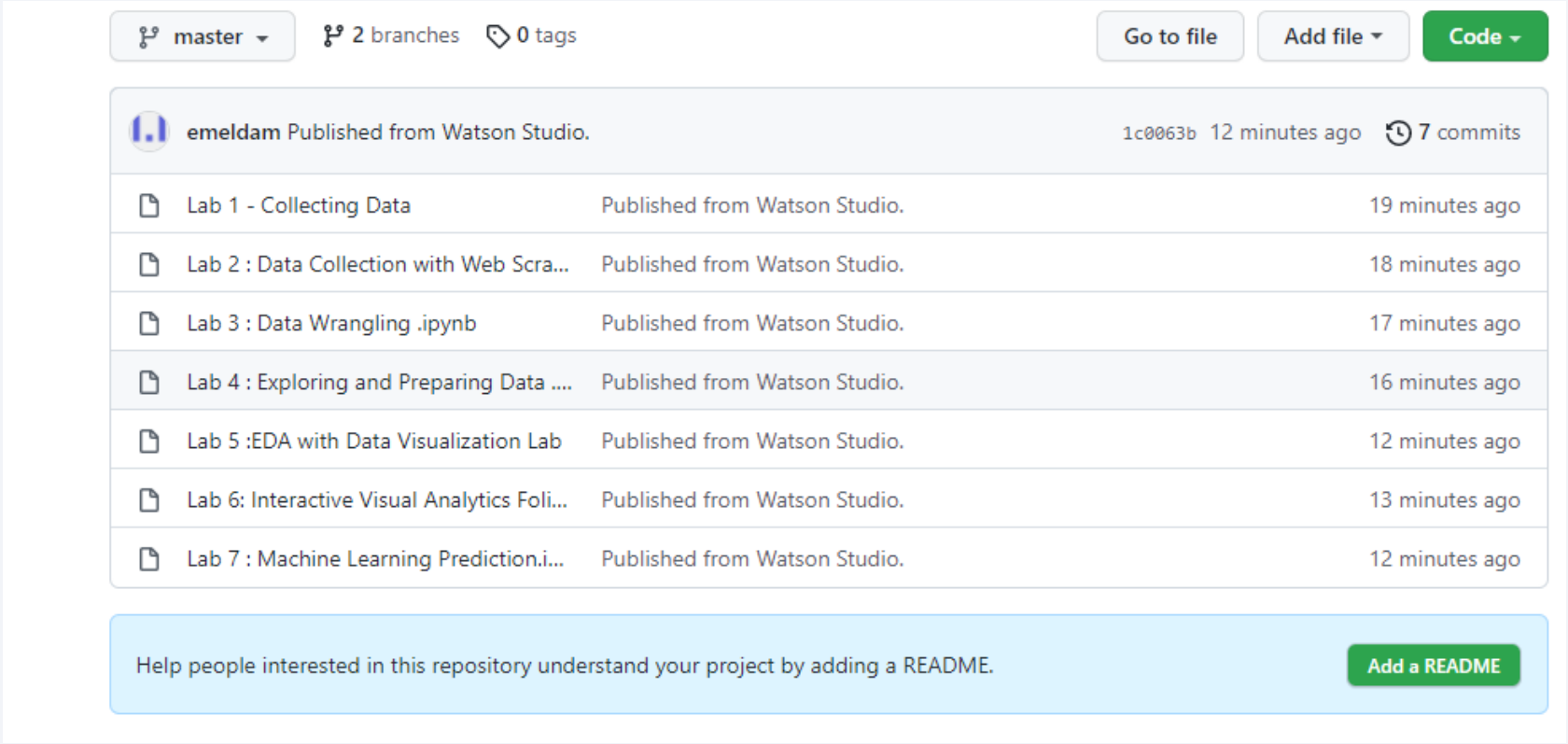


Conclusions

- The current best classifier for the current task is SVM
- It gives an idea of SPACEX's launch
- Better data analysis allows us to understand the content more intuitively

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!