APPLIED DATA SCIENCE CAPSTONE PROJECT

- MANOJ V

22.04.2023

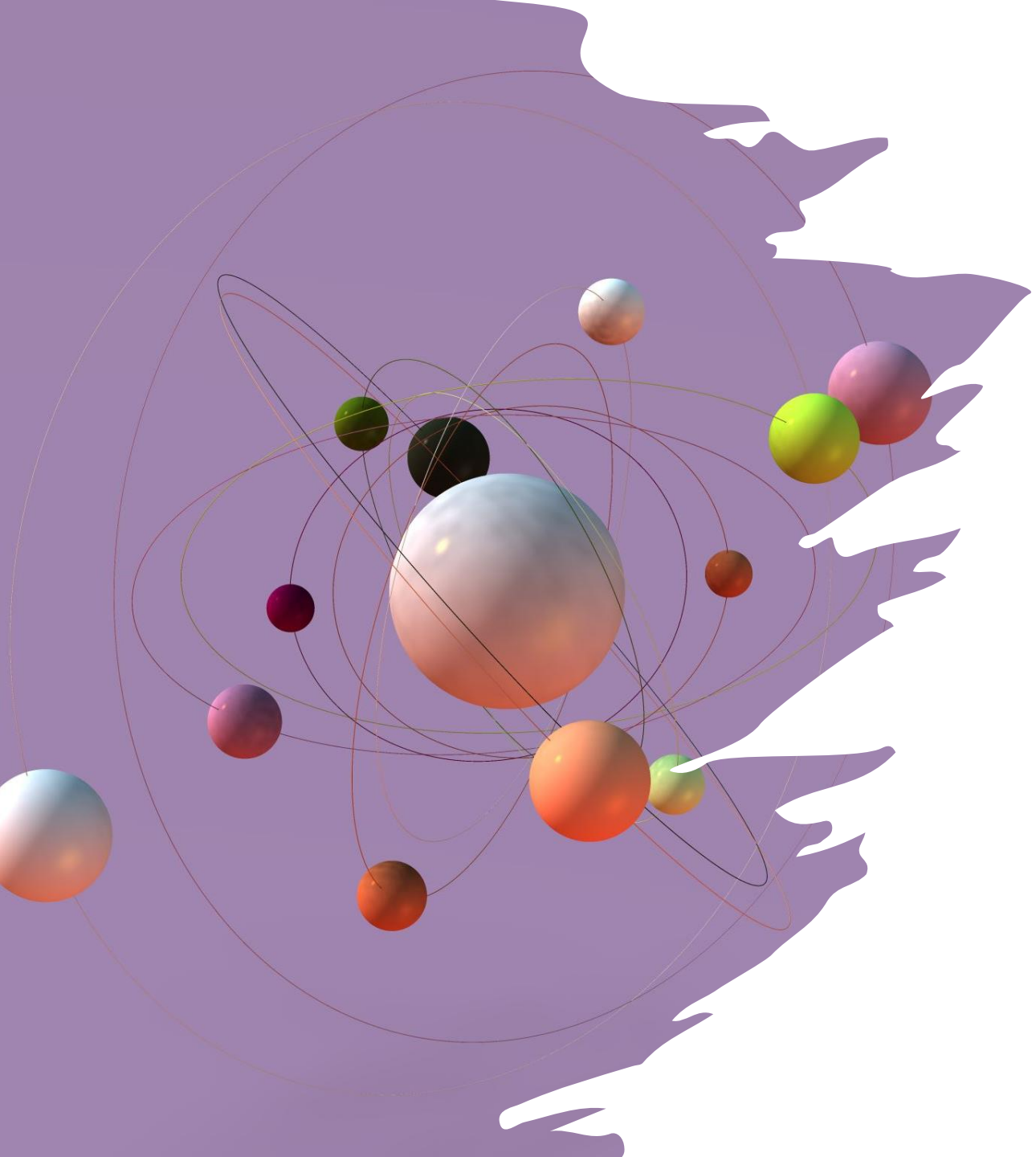
Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Following methodologies were used for data analysis:

- Data Collection using web scrapping and Data Wrangling.
- Exploratory Data Analysis (EDA) with SQL, Data Visualization and interactive dashboard analytics.
- Interactive dashboard results
- Machine Learning Predictions based on standardized data.



Introduction

Objective: To conduct a comprehensive analysis and assess the success rate of first stage landing of a novel company SpaceY in comparison to SpaceX.

Desirable Outcomes:

- The effects of features such as Payload Mass, Launch Site, Orbit on the success of first stage landing
- Presentation of success rate results over the years.
- A highly accurate prediction model for the success rate based on Machine Learning

Methodology

• Data Collection: Data was primarily collected from sources below

Sources:

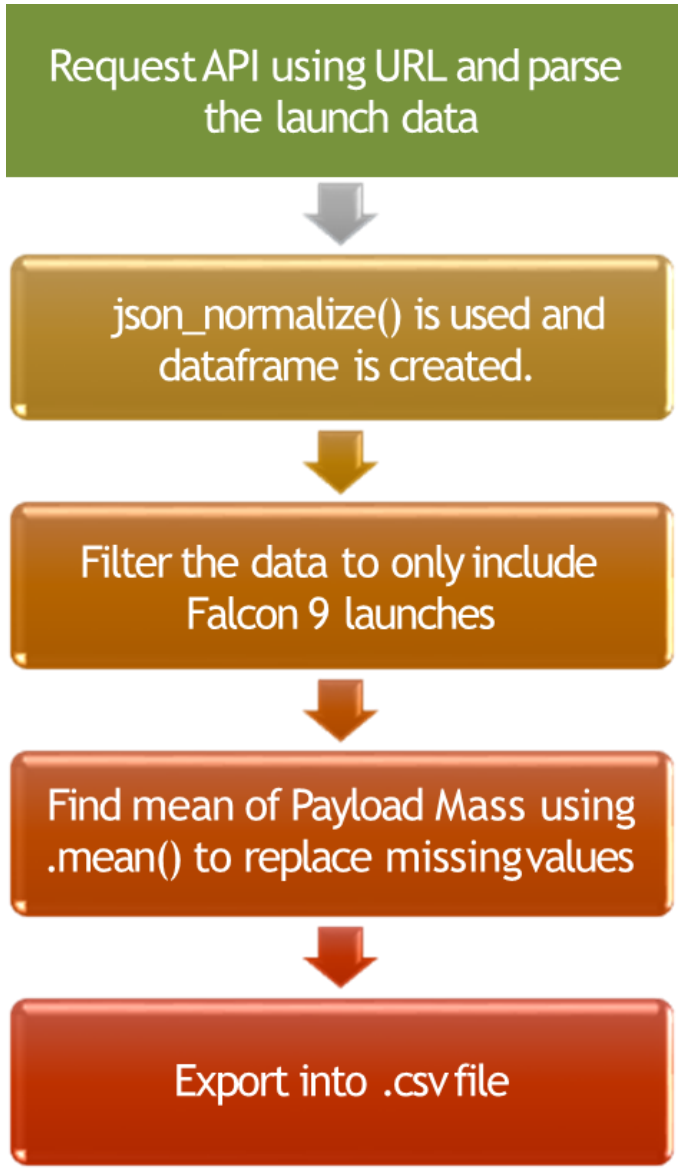
SpaceX API (https://api.spacexdata.com/v4/launches/past)

Web scrapping (Wiki Link)

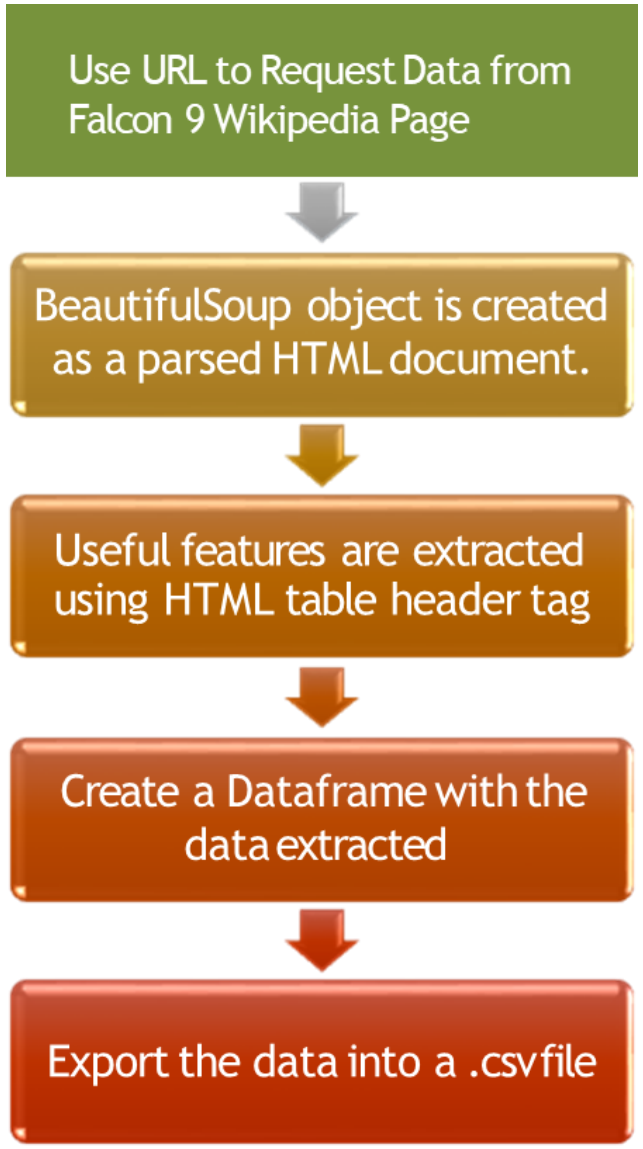
• Data Wrangling

Data was labelled based on the outcome of the launch, making it easier to work on in EDA

Data Collection - API



Data Collection – Web Scrapping

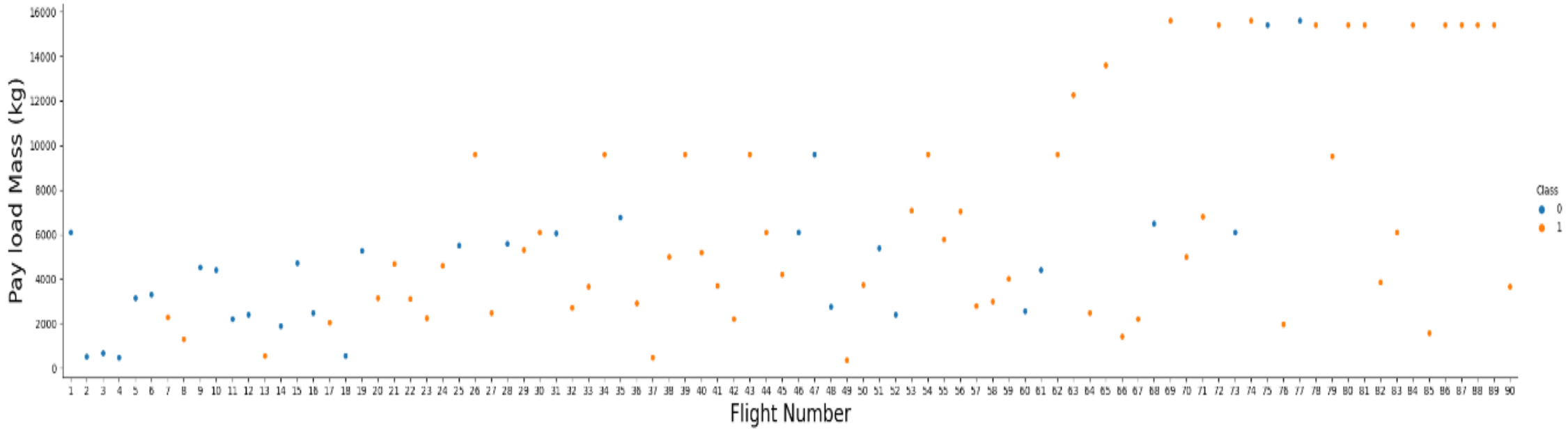


Exploratory Data Analysis (EDA) with SQL

- The following SQL queries were performed to extract valuable data:
- 1. Display the names of the unique launch sites in the space mission
- 2. Display 5 records where launch sites begin with the string 'CCA'
- 3. Display the total payload mass carried by boosters launched by NASA (CRS)
- 4. Display average payload mass carried by booster version F9 v1.1
- 5. List the date when the first successful landing outcome in ground pad was achieved.
- 6. List of names of boosters that succeeded and have payload mass between 4000-6000.
- 7. List the total number of successful and failure mission outcomes
- 8. List the names of booster versions that carried a max payload mass using Subquery.
- 9. List the records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
- 10. Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017 in descending order

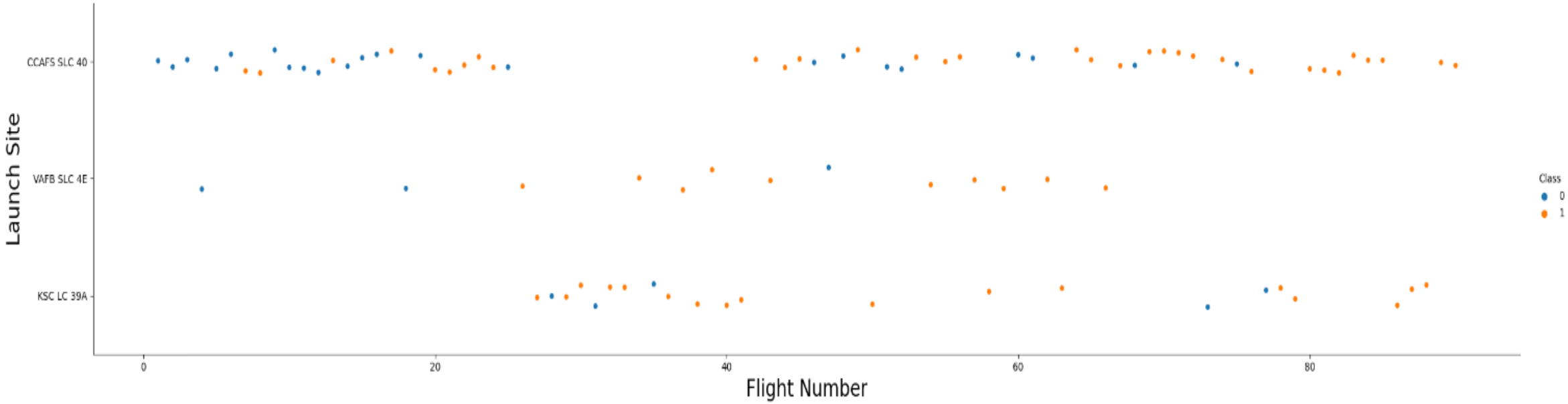
EDA with Data Visualization

- 1. Scatter Plot: Payload Mass vs Flight Number



EDA with Data Visualization

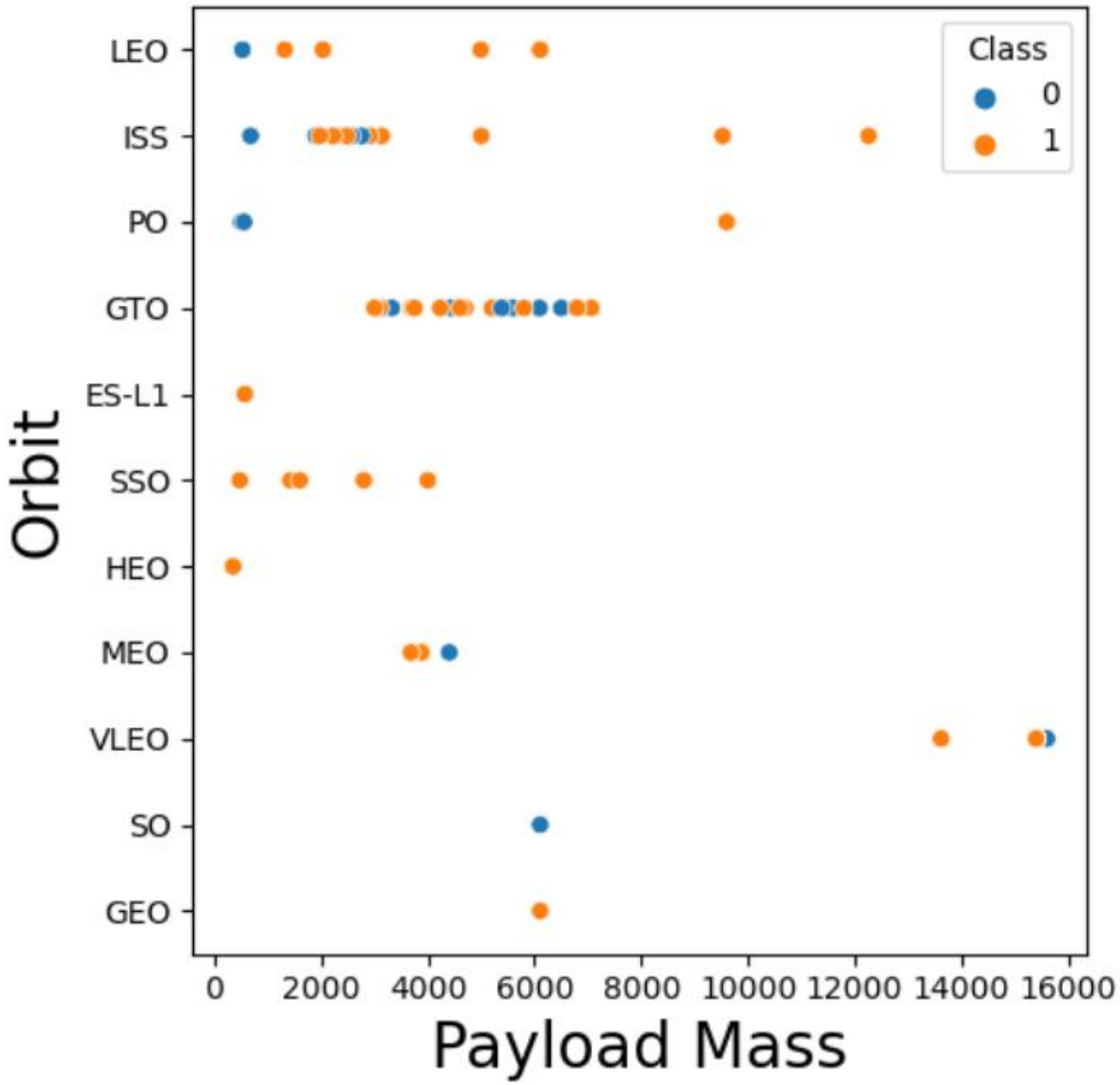
- 2. Scatter Plot: Launch Site vs Flight Number



EDA with Data Visualization

https://github.com/Rambo1806/IBM-Data-Science-Capstone/blob/main/EDA%20Visualization.ipynb

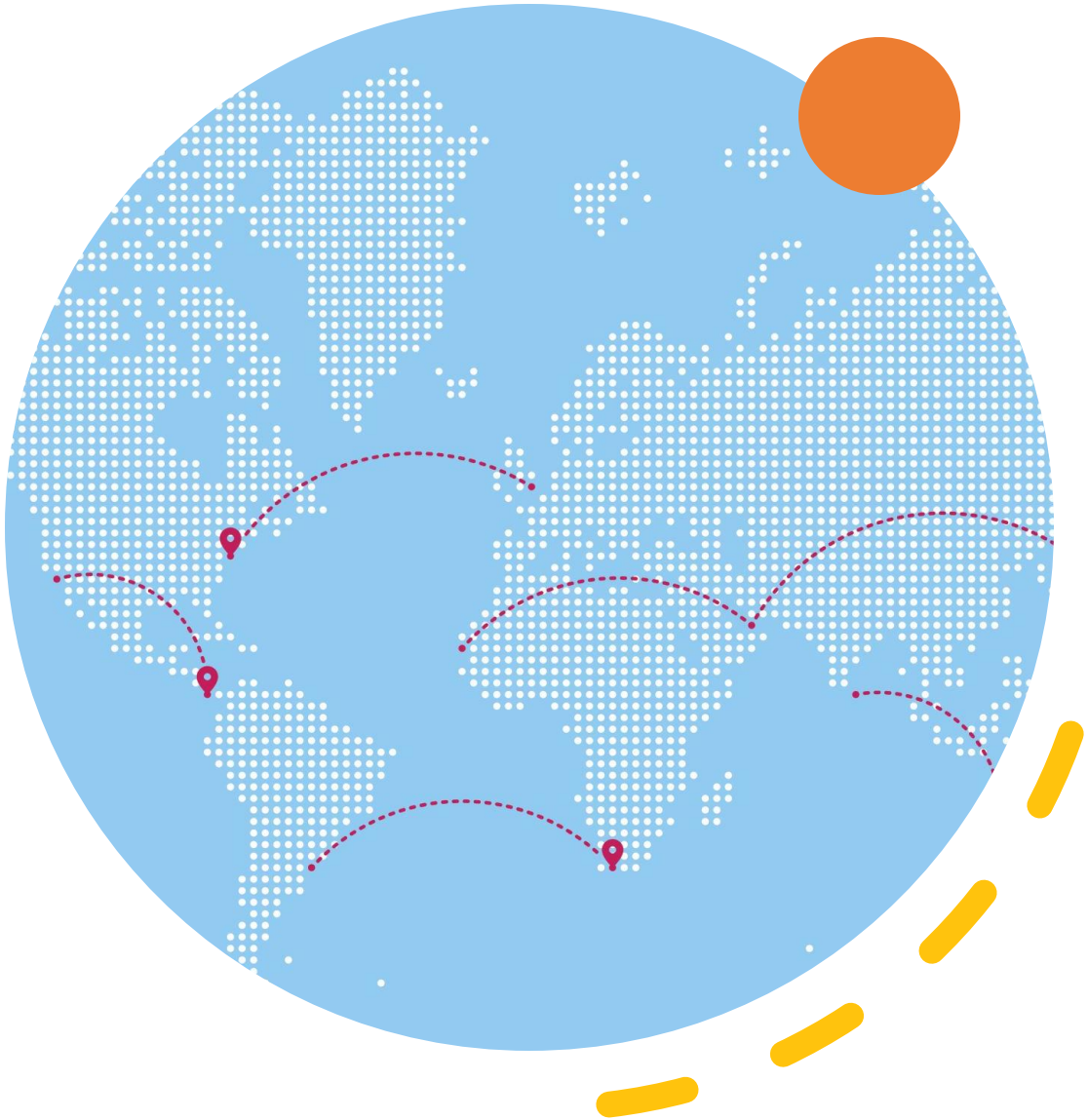
Visit Link for more clear Pictures and Details



Interactive Mapping with Folium

Folium enables interactive analysis enables the use of multiple leaflet maps and used in dashboarding.

- The following functions of folium were used along with Folium Library:
- 1. Markers: Used to mark coordinates of the data in a real-world map (e.g. launch sites)
- 2. Circles: Provides a circular highlight to the markers' specific location (SpaceX launch site)
- 3. Marker Clusters provide the option to mark as a group of occurrences in each coordinate. (e.g., multiple launches at a launch site).
- 4. Plotting line between two points provides distance between those coordinates.



Dashboarding with Plotly

- The following 2 plots were used to encapsulate and visualize data using an interactive dashboard:
- 1. Pie chart of Percentages of Launches by sites based on the choice of the dropdown menu.
- 2. A scatter plot to show relation between payload and Launch success. The payload range (Kg) is inputted using an interactive slider as shown in the source code.
- The key idea is to understand the relation between payload and launch success to pick the best launch site to have a first stage success.

Predictive Analysis (Classification)

The four classification models are:

1. Logisitic Regression
2. Support Vector Machines
3. Decision Tree
4. K nearest Neighbours

Results – Data Collection API

- Data collected from the SpaceX API that only contains information pertaining to Falcon 9 Booster Version

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False

Results – Data Collection Web scrapping

- Data collected from Wikipedia in reference to SpaceX and entered as a data frame using Pandas

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
237	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10	Success	9 May 2021	06:42
238	118	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	F9 B5B1058.8	Success	15 May 2021	22:56

Results – EDA with SQL

https://github.com/Rambo1806/IBM-Data-Science-Capstone/blob/main/SQL%20EDA.ipynb

Visit Link for further Details and Results

In [8]:
```
%sql select distinct(launch_site) from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

Out[8]:

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

Display average payload mass carried by booster version F9 v1.1

In [26]:
```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like '%F9 v1.1%';
```

```
 * sqlite:///my_data1.db
Done.
```

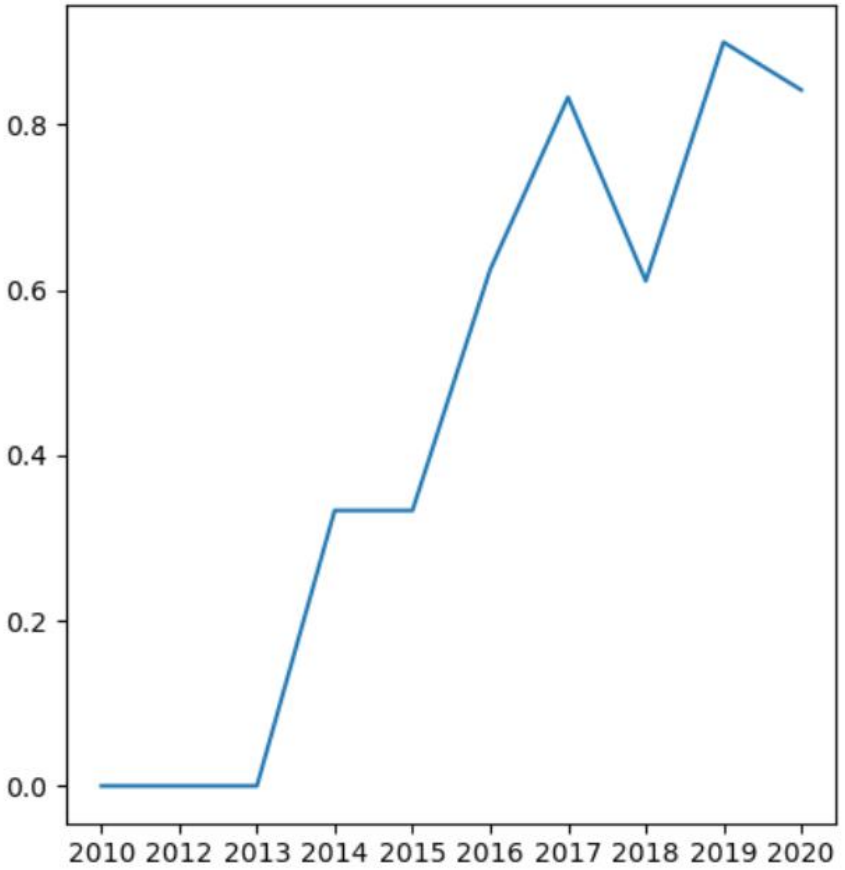
Out[26]:

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

Results – EDA with Visualization

https://github.com/Rambo1806/IBM-Data-Science-Capstone/blob/main/EDA%20Visualization.ipynb

Visit Link for further Details and Results



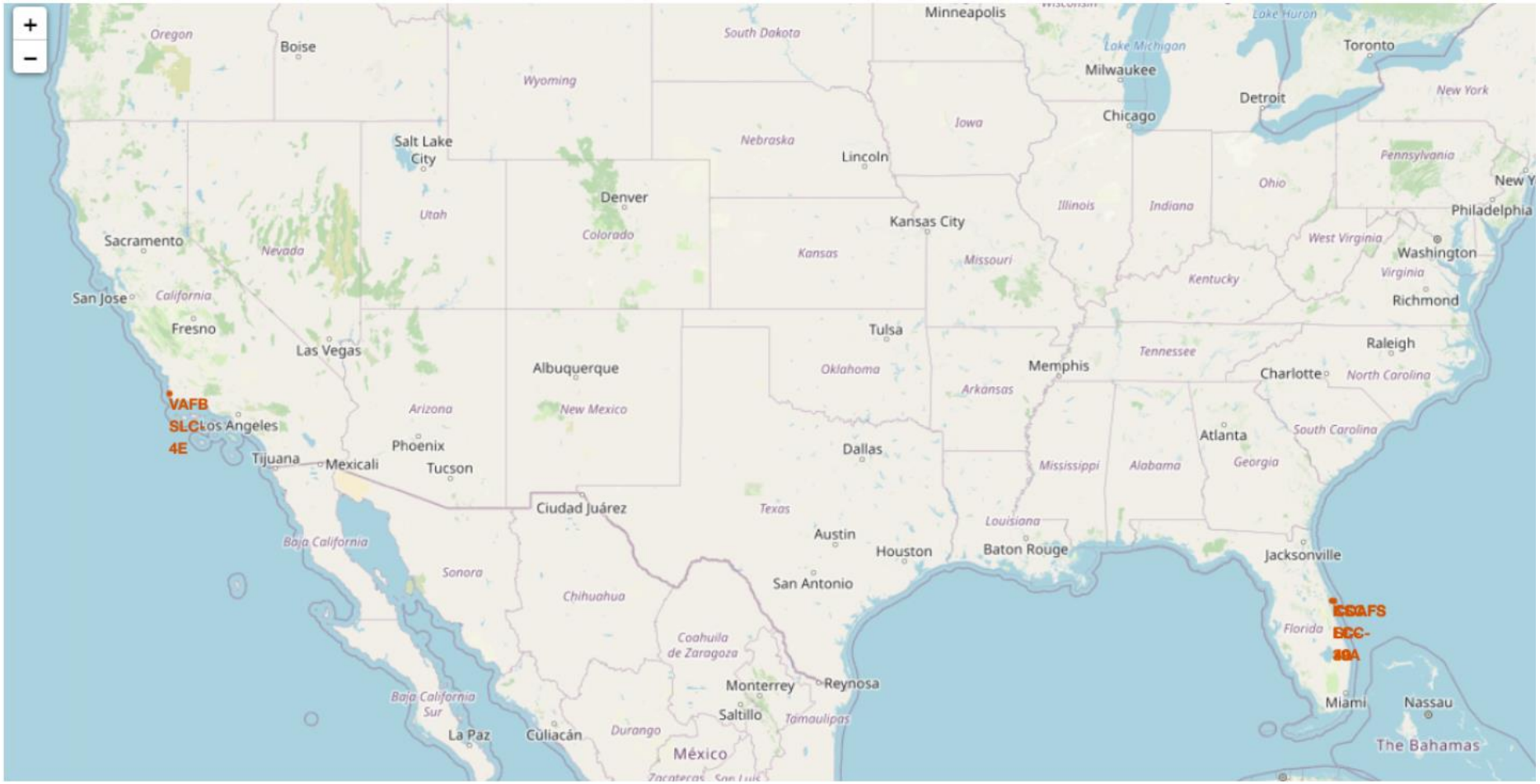
Success rate graph over the years

Results – Folium

https://github.com/Rambo1806/IBM-Data-Science-Capstone/blob/main/Folium%20Lab.ipynb

Visit Link for further Details and Results

Launch Sites on Map

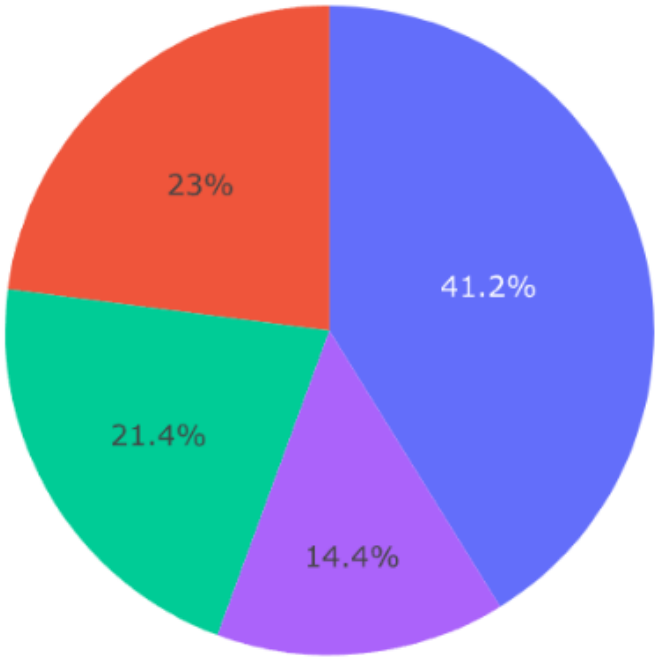


Results – Plotly Dashboard

https://github.com/Rambo1806/IBM-Data-Science-Capstone/blob/main/Plotly%20Application.pdf

Visit Link for further Details and Results

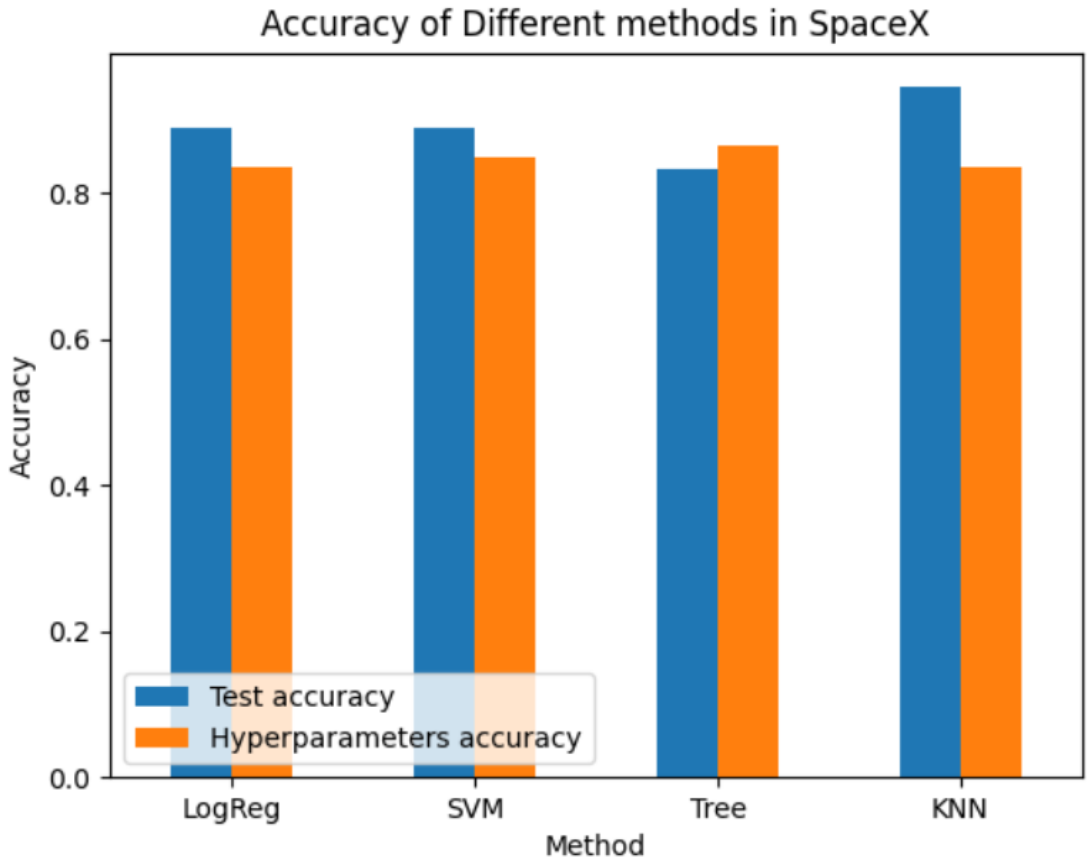
Total Success Launches by Site



Results – ML Predictive Analysis

https://github.com/Rambo1806/IBM-Data-Science-Capstone/blob/main/Machine%20Learning%20Prediction.ipynb

Visit Link for further Details and Results



Key Conclusions

Orbit Type "ES-L1", "GEO", "HEO" , "SSO", "VLEO" all have a success rate of 100% of landing the first stage rocket.

- The success rates have increased over the years with 2019 having the highest success rate of 98%.

- The launch site with most success is "KSC LC-39A"

- All the launch sites are situated closed to coastlines, to avoid rockets landing on land.

- The ML prediction models provides high accuracy to the datasets with Decision Trees methodology providing highest accuracy.

- The highest success rates come from payload mass range from 2000-5000 Kgs

Thank You

GitHub Repository:

https://github.com/Rambo1806/IBM-Data-Science-Capstone