IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gerardo Cuéllar Bonnard
18/12/2024

Outline

Executive Summary

- ***In this Project we will pull efforts towards an accurate prediction of the Space x Falcon 9 first stage landing , we will use several classification algorithyms to accomplish this.***

- *We will divide our project Into 4 main steps :*

1. Data collection, wrangling, and formatting
2. Exploratory data analysis
3. Interactive Data Visualization
4. Machine learning prediction

- Running the classification algorythims shows us that there are correlations between some feature of the rocket launching with the outcome of the launches particulary in success and failure

- ***We have concluded and recommend Decission three as the most efficient and secure way to predict a succesfull landing in the Falcon 9 First Stage.***

Introduction

- In the race of space, effictiveness is determined by speed and accuracy , **the main goal here is to enable the completion and safe return of the first stage of the launch rocket** , this will alow us to lower their operational costs and grant a margin within 62 million dollars , this put us in the same page as Space x **as a result we separate ourselfs from the rest of the competition and safe our company 103 million dollars .**
- Most unuccesfull landings tend to be paned , as a result of this Space x plan their ones in the ocean.
- **To answear the main question will the first stage of the rocket land succesfully/ ?** we have review the features corresponding payload mass / orbit type/ Launch site/ and many more

Section 1

Methodology

Methodology

- **Data collection methodology:**

  SpaceX API / Webscrapping

- **Perform data wrangling**

We process the data , to eliminate missing entries and we encoded categorical features using one - hot encoding / In the end we end up with 90 rows/ instances and 83 columns or features

- **Perform exploratory data analysis (EDA) using visualization and SQL**

Pandas and Numpy: We used Line Charts, Bar Charts , Scatter Plots

Sql/Ipython(Function)/Sqlite

Methodology

Perform interactive visual analytics using Folium and Plotly Dash

Matplot library and Seaborn

Folium

We used Markers, circles and lines ***to delimitate the launching areas*** an their proximity to railroads, highways and main highways within the folium map. ***A geographical approach was intended for the correct utilization of resources and logistcs.***

- ***Dash***

Interactive Scatter Plots / Pie Charts

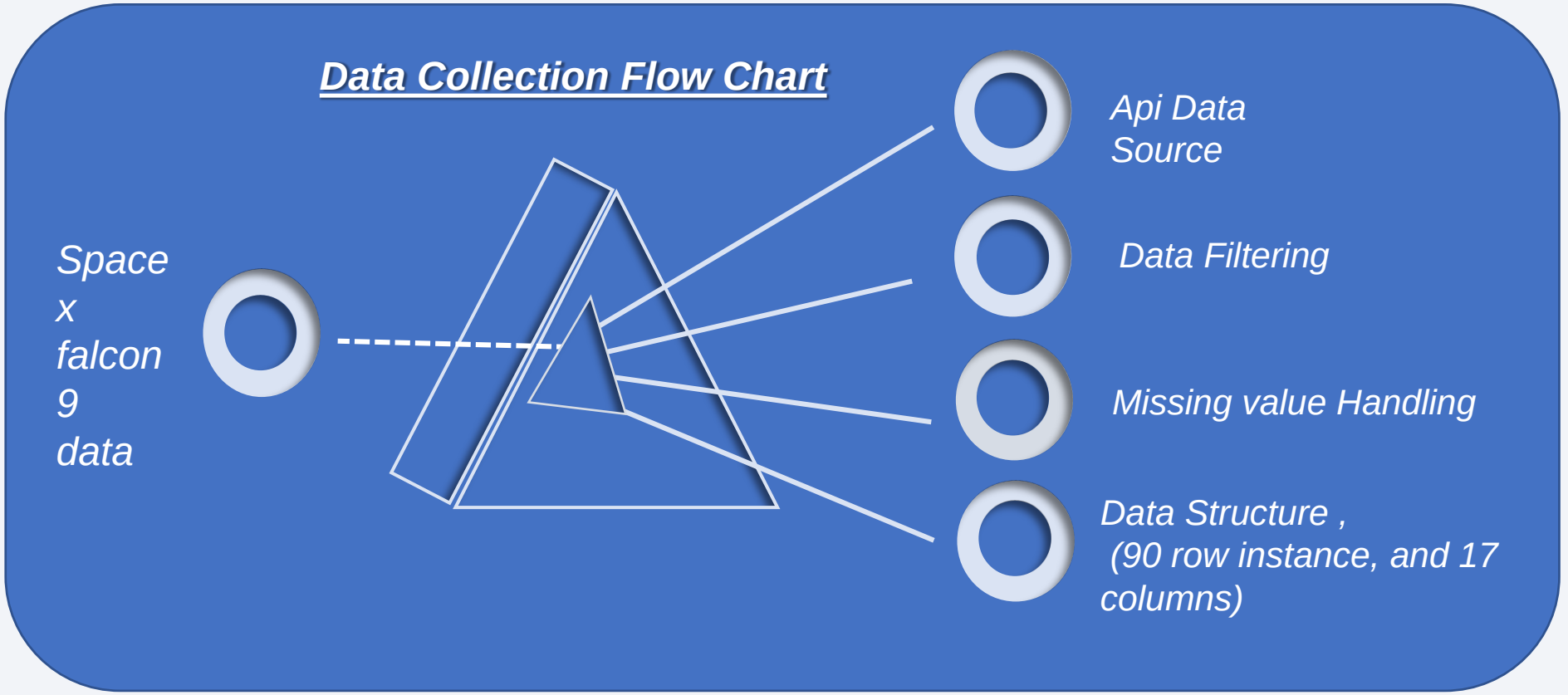
- **Perform predictive analysis using classification models**

How to build, tune, evaluate classification models

- Data Frames, Confusion Matrixs
- Logistic Regression /SVM***/ Decission Three***/ KNN

Data Collection – SpaceX API

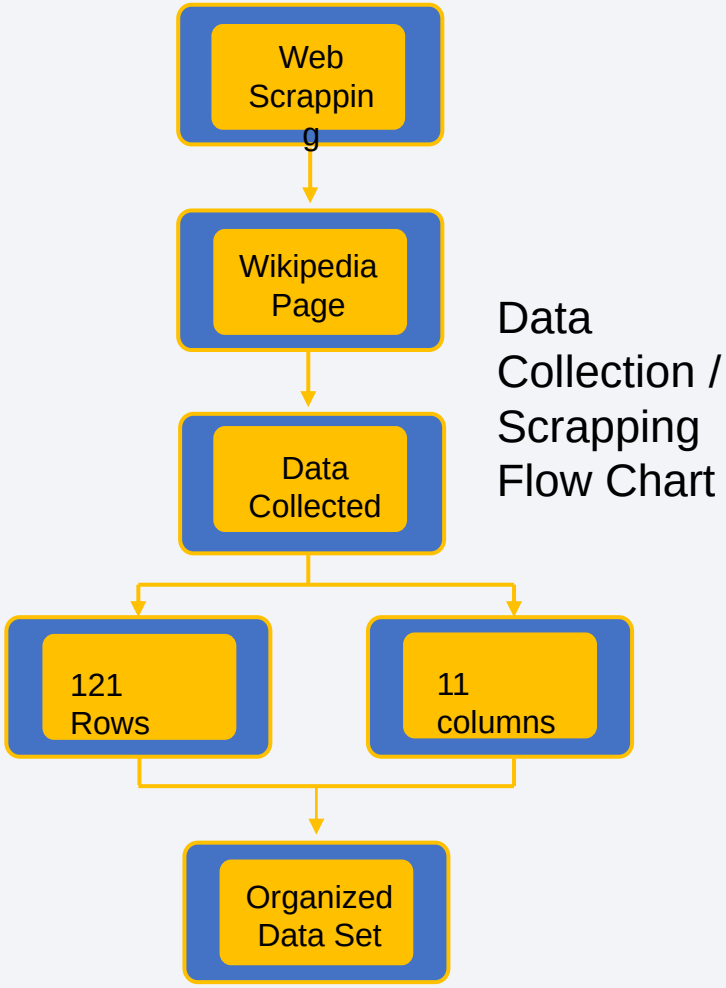
- The Api used is : https://api.spacexdata.com/v4/rockets/
- The api provides data about many types of rockets launches done by Space X, the data is therefore filter to include only Falcon 9 Launches
- Every missing value in the data is replaced .
- We end up with 90 rows or instances and 17 columns or features .



Data Collection Github permalink

Data Collection - Scraping

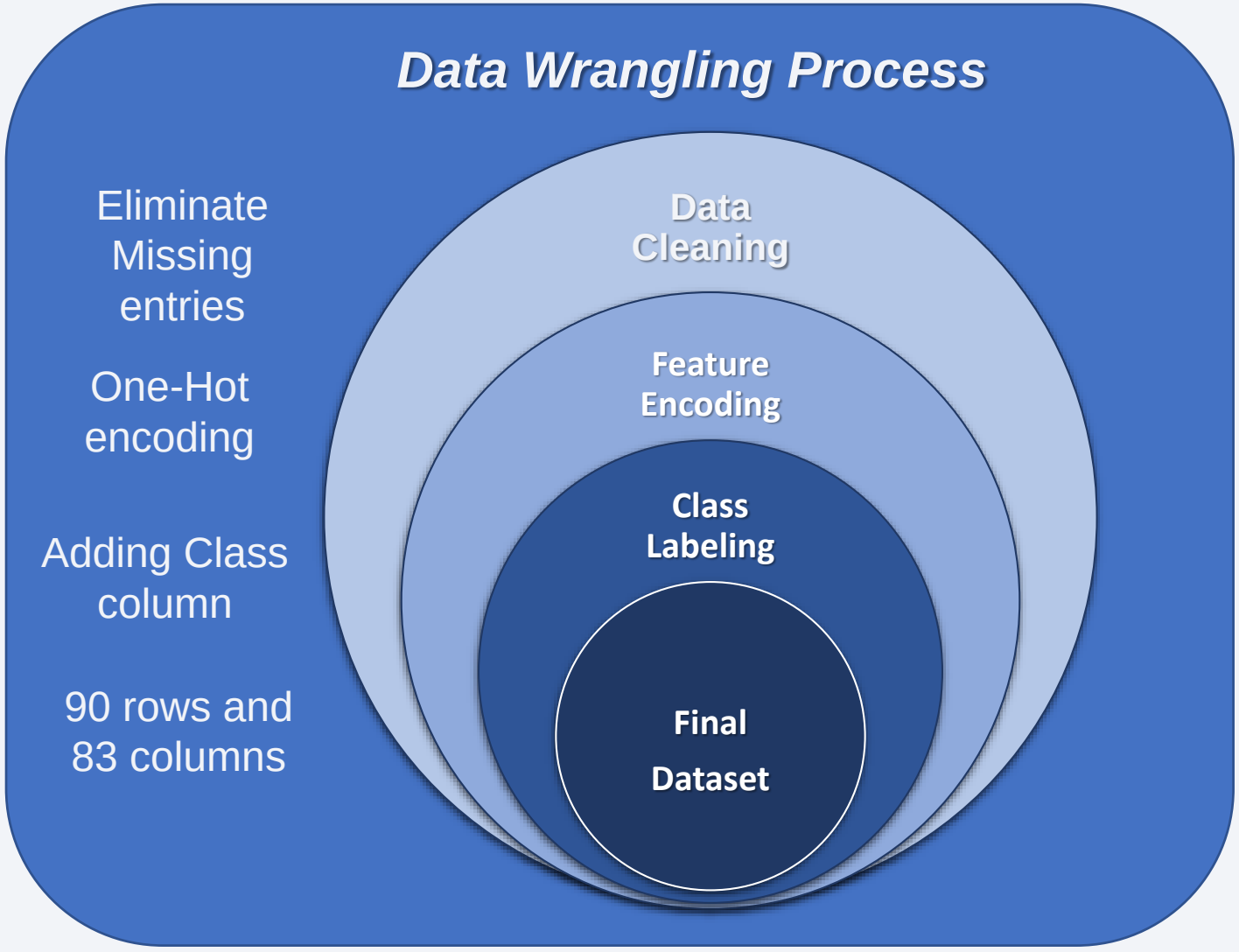
- Data is webscrapped from : https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- The website consist of exclusive falcon 9 Data
- We end up with 121 rows and 11 column features
- Git hub webScrapping Link



Data Collection / Scrapping Flow Chart

Data Wrangling

- We process the data , to eliminate missing entries and we encoded categorical features using one - hot encoding
- We then add an extra column called “Class” , this will help us determined if the launch is succesfull or not (1= succesfull landing , 0 = not succesfull landing)
- In the end we end up with 90 rows/ instances and 83 columns or features
- Data Wrangling Git Hub link



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

Libraries used for plotting :Pandas and Numpy, were used to delimit information about the data :

We used Scatter point charts , Bar Charts and line plots to :

1.- Determine the relationship between flight number and launch site ,

2.- To observe is there is any relationship between launch sites and their payload mass,

3.-To check if there is any relationship between success rate and orbit type ,

4.-to understand is there is any relationship between Flight number and orbit type.

5.- to reveal the relationship between Payload and orbit type and

6 .-to visualize the lunch success yearly trend

EDA with Visualization Github url

EDA with SQL

- **Unique launch sites in the space mission**
- **5 records where launch sites begin with the string 'CCA'**
- **Total payload mass carried by boosters launched by NASA (CRS)**
- **Average payload mass carried by booster version F9 v1.1**
- **Date when the first succesful landing outcome in ground pad was acheived.**
- **Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000**
- **Total number of successful and failure mission outcomes**
- **Names of the booster_versions which have carried the maximum payload mass.**
- **Records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.**
- **Landing outcomes count ranking (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.**
- **EDA With Sql Github link**

Build an Interactive Map with Folium

We used Markers, circles and lines to delimitate the launching areas an their proximity to railroads, highways and main highways, prominent water masses within the folium map***. A geographical approach was intended for the correct utilization of resources and logistcs.***

- Folium , github code

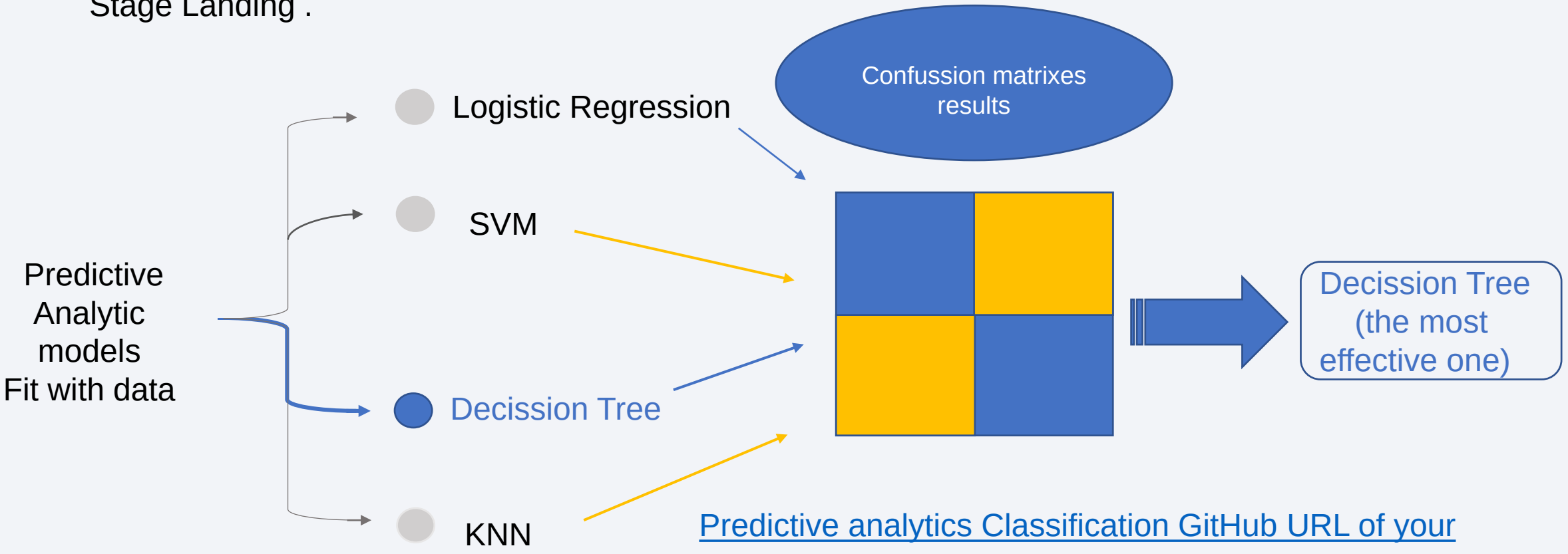
Build a Dashboard with Plotly Dash

We used, interactive Scatter Plots and pie charts , to determine the Launch Succes , the launch success ratio , to efectively understand the correlationship , between payload and Launch Succes.

Dash Github url Link

Predictive Analysis (Classification)

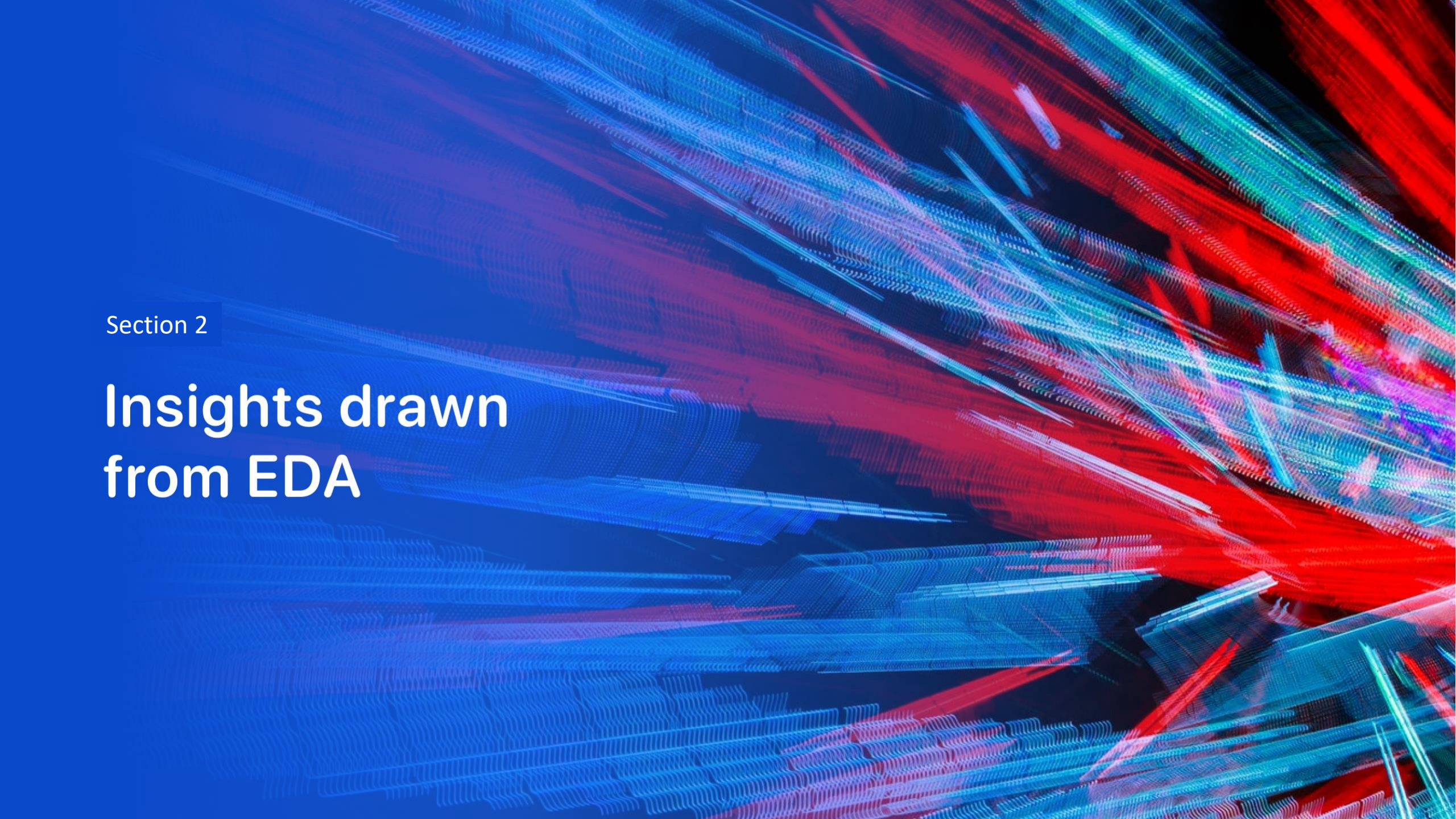
- We used and fit 4 *classification algorithyms* : Logistic Regression , SVM, **Decission Tree** and KNN into confussion Matrixes , to determine the most effective one, to accomplish the safe return of the Falcon 9 first Stage Landing .



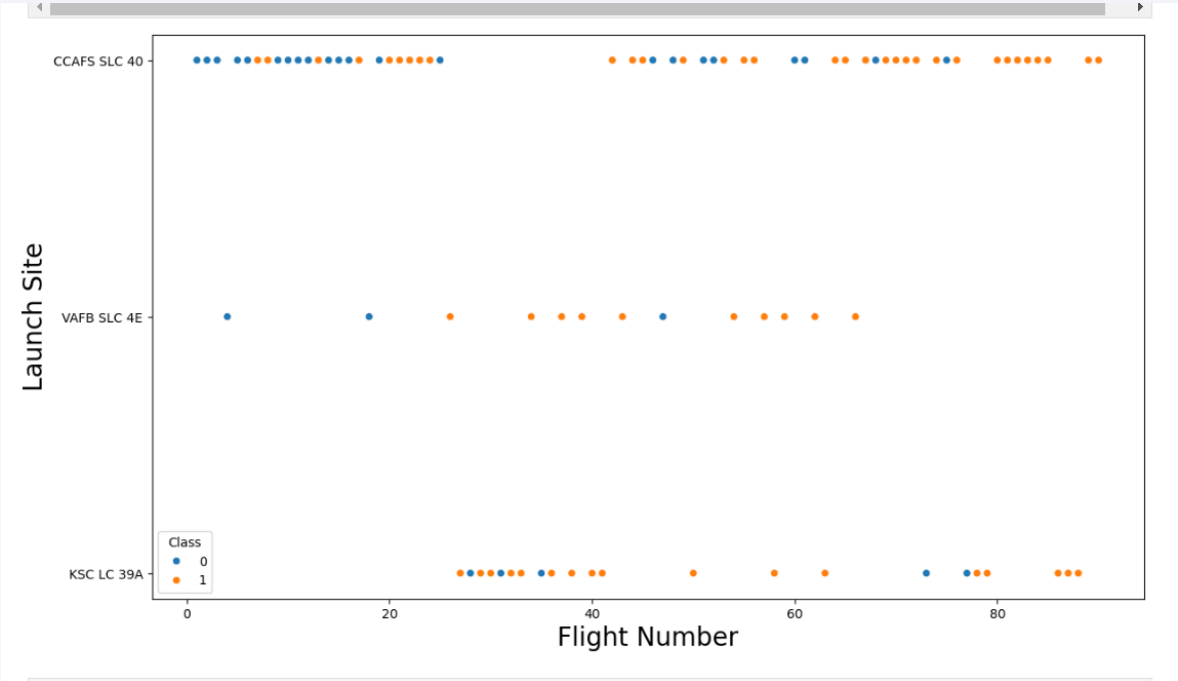
Predictive analytics Classification GitHub URL of your

Results

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

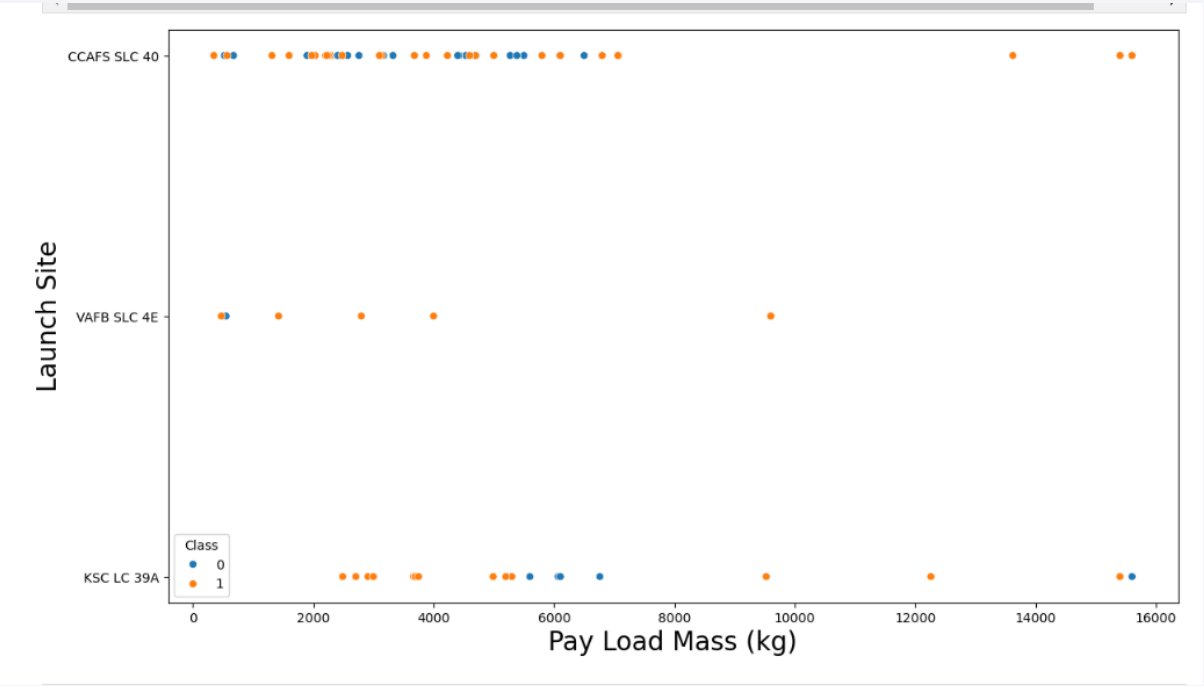

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



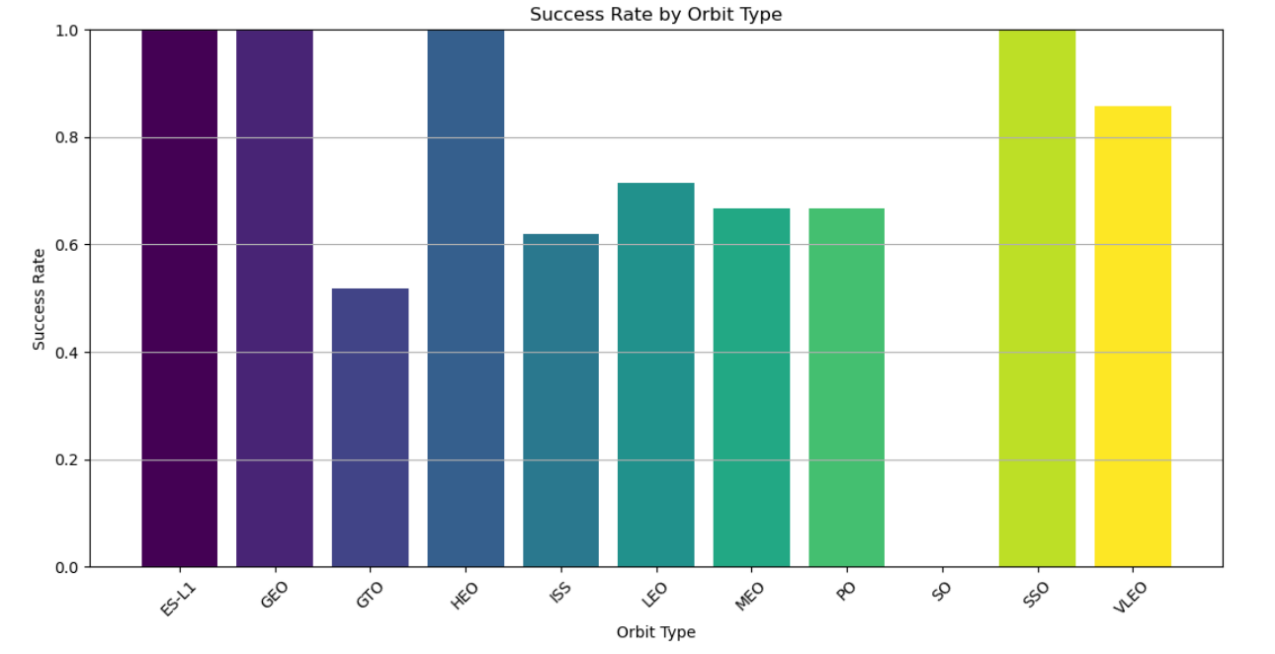
This scatter plot shows us the correct relationship between fight number in the x axis and Launch site in y axis. The class is represented in 0 and1 , 0 being an unsuccesfull launch and 1 representing a succesfull launch .

Payload vs. Launch Site

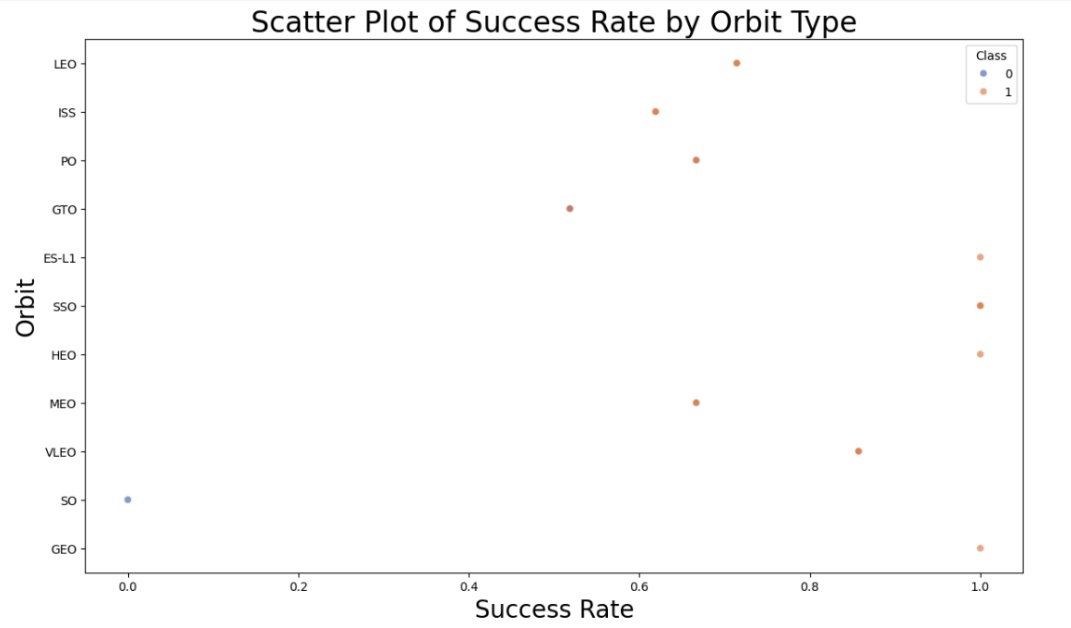


This scatter plot shows us the correct relationship between Payload mas in the x axis and Launch site in y axis. The class is represented in 0 and1 , 0 being an unsuccesfull launch and 1 representing a succesfull launch . Its interesting to denote that 8000 kg does not have a launch attempt .

Success Rate vs. Orbit Type Bar Chart

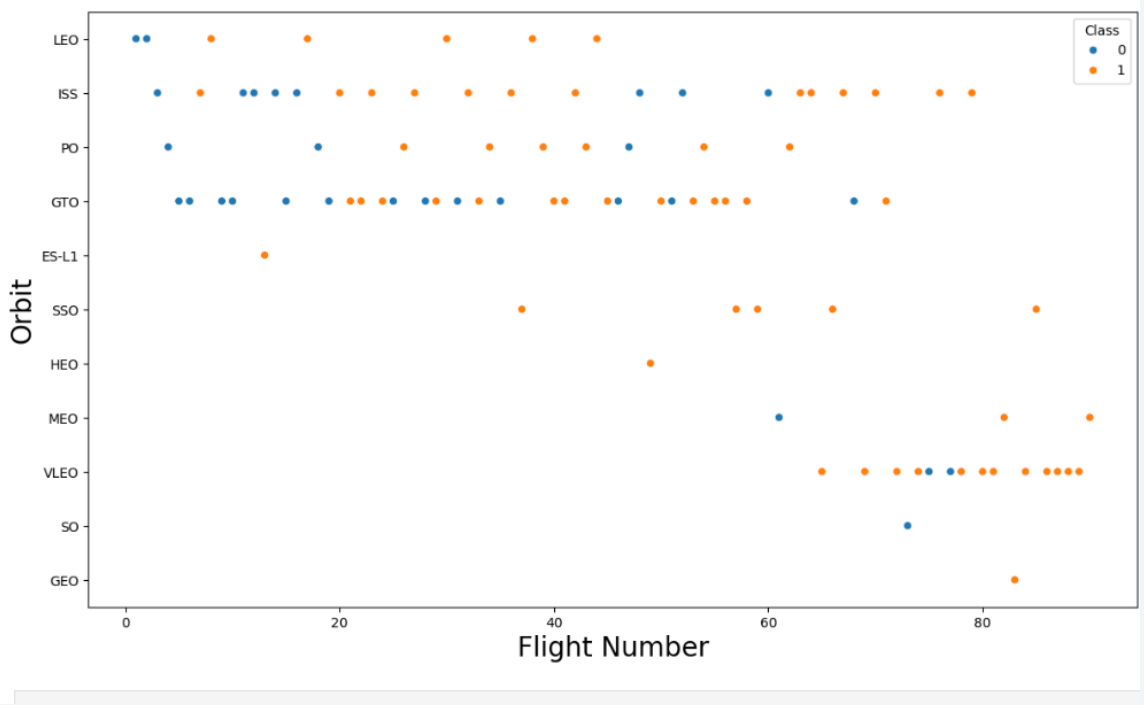


Success Rate vs. Orbit Type Scatter Plot



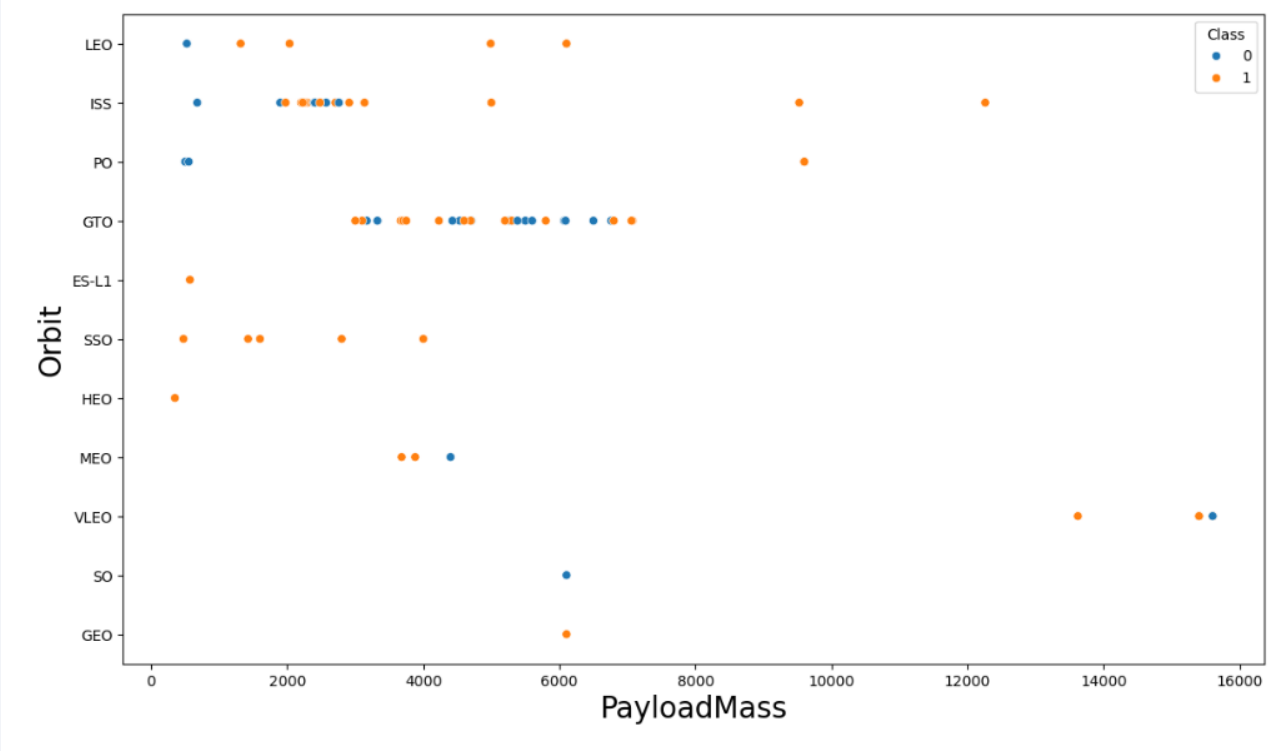
With this analysis we can understand insights of wich orbit types are more reliable in their performance x axis represents , success rate and y axis represents Orbit

Flight Number vs. Orbit Type



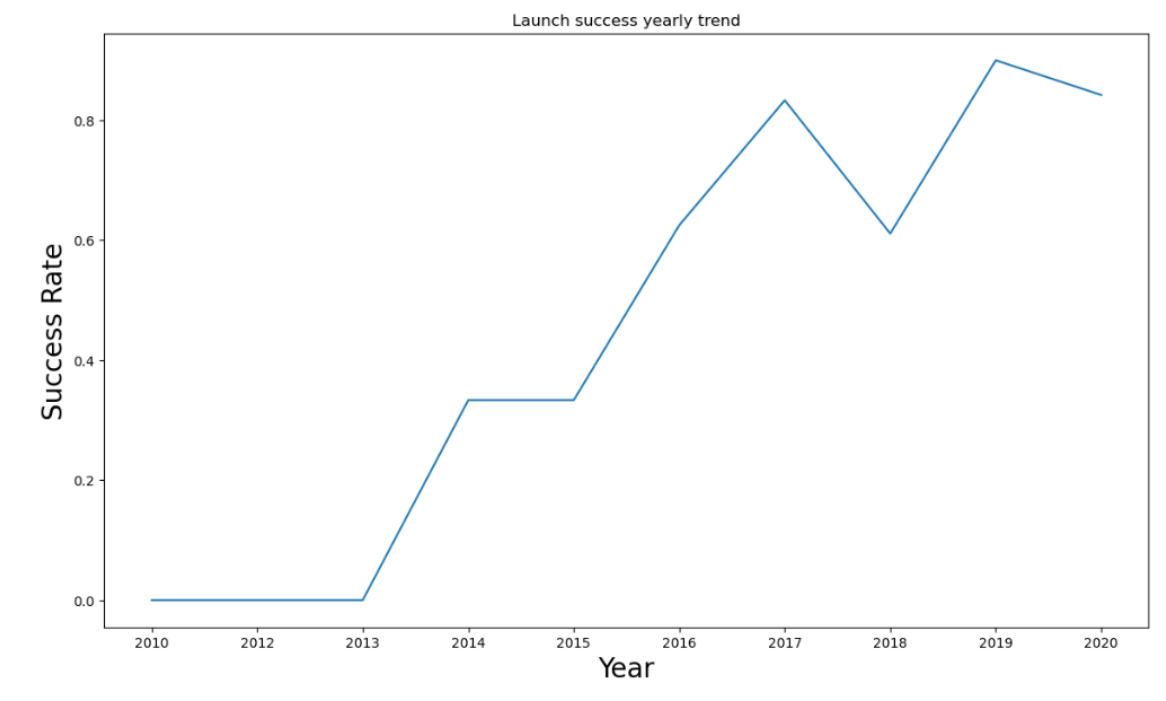
This scatter plot shows us the relationship between Flight number in the x axis and Orbit site in y axis. The class is represented in 0 and1 , 0 being an unsuccesfull launch and 1 representing a succesfull launch .

Payload vs. Orbit Type



This scatter plot shows us the correct relationship between Payload mas in the x axis and orbit in y axis. The class is represented in 0 and1 , 0 being an unsuccesfull launch and 1 representing a succesfull launch . Its interesting to denote that 8000 kg does not have a launch attempt .

Launch Success Yearly Trend



This Line cahrt , show us , the relationship , betweenYear in the x Axis and the success rate in the Y axes . Having and constant grow since 2010 with a considerable drop in 2018 , the highest relationship rate is found between 2018 and 2019 , with a minor drop in 2020 , probably due to Covid 19

All Launch Site Names

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

```
%sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;
```

This sql code Display the names of the unique launch sites in the space mission.

Launch Site Names that 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	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

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

This sql code displays 5 records where launch sites begin with 'CCA'

Total Payload Mass

Toltal_payload_mass_by_NASA_CRS
45596

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS Toltal_payload_mass_by_NASA_CRS FROM SPACEXTABLE WHERE CUSTOMER ='NASA (CRS)';
```

This Sql Code calculates the total payload by boosters From NaSA

Average Payload Mass by F9 v1.1

AveragePayloadMassByBoosterVersionF9v1_1
2928.4

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS
AveragePayloadMassByBoosterVersionF9v1_1 FROM SPACEXTBL WHERE
BOOSTER_VERSION = 'F9 v1.1';
```

This Sql code Calculates the average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

Date of first successful landing outcome in ground pad
2015-12-22

```
%sql SELECT MIN(DATE) AS 'Date of first successful landing
outcome in ground pad' FROM SPACEXTBL WHERE
LANDING_OUTCOME = 'Success (ground pad)';
```

This sql code finds and display the the dates of the first successful landing outcome on ground pad

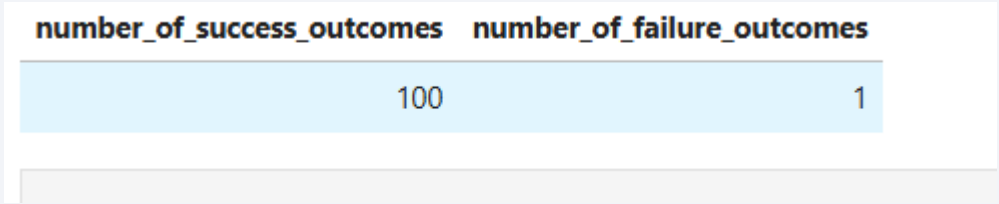
Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE LANDING_OUTCOME =
'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

This sql code lists the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes



number_of_success_outcomes	number_of_failure_outcomes
100	1

```
%sql SELECT number_of_success_outcomes, number_of_failure_outcomes
FROM (SELECT COUNT(*) AS number_of_success_outcomes FROM
SPACEXTBL WHERE MISSION_OUTCOME LIKE 'Success%') success_table,
(SELECT COUNT(*) number_of_failure_outcomes FROM SPACEXTBL WHERE
MISSION_OUTCOME LIKE 'Failure%') failure_table
```

This Sql code Calculates and displays the total number of successful and failure mission outcomes

Boosters Carried Maximum Payload

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

```
%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

This sql code lists the names of the booster which have carried the maximum payload mass

2015 Launch Records

4]:

Date	Booster_Version	Launch_Site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

```
%sql
SELECT
    CASE substr(DATE, 6, 2)
        WHEN '01' THEN 'January'
        WHEN '02' THEN 'February'
        WHEN '03' THEN 'March'
        WHEN '04' THEN 'April'
        WHEN '05' THEN 'May'
        WHEN '06' THEN 'June'
        WHEN '07' THEN 'July'
        WHEN '08' THEN 'August'
        WHEN '09' THEN 'September'
        WHEN '10' THEN 'October'
        WHEN '11' THEN 'November'
        WHEN '12' THEN 'December'
    END AS Month,
    LANDING_OUTCOME,
    BOOSTER_VERSION,
    LAUNCH_SITE
FROM SPACEXTBL
WHERE substr(DATE, 0, 5) = '2015'
AND LANDING_OUTCOME = 'Failure (drone ship)';
```

This Sql Lists the failed landing_outcomes in drone ship, their booster versions and launch site names for year 2015

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

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

```
%sql SELECT LANDING_OUTCOME,
COUNT(LANDING_OUTCOME) FROM
SPACEXTBL WHERE DATE BETWEEN '2010-
06-04' AND '2017-03-20' GROUP BY
LANDING_OUTCOME ORDER BY
COUNT(LANDING_OUTCOME);
```

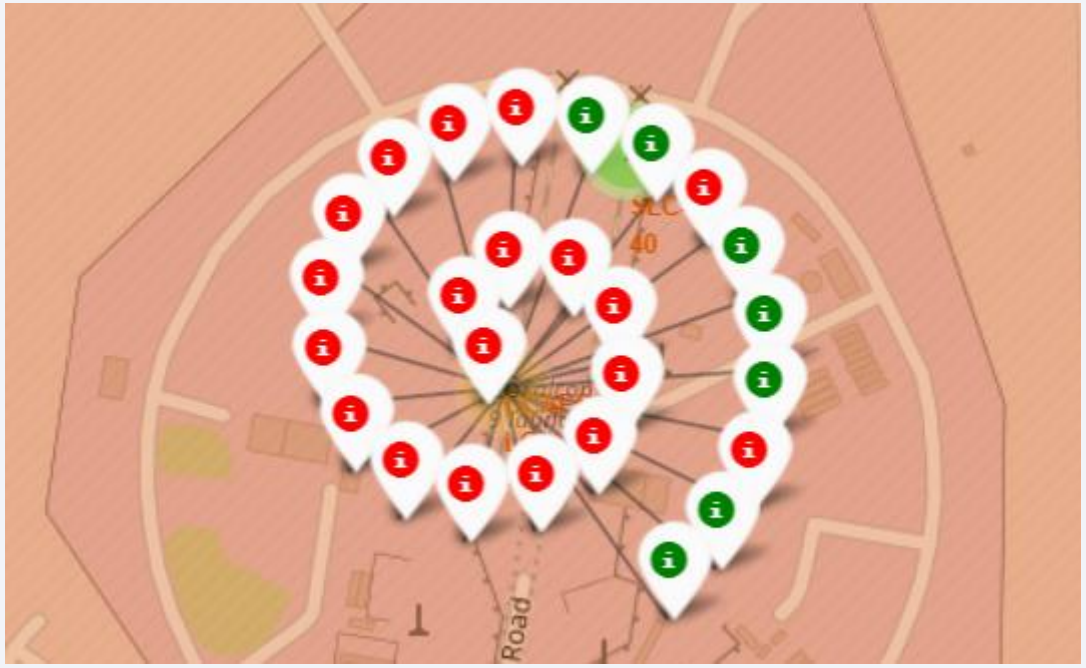
This sql Ranks and displays the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order


Section 3
Launch Sites
Proximities Analysis

Launch site delimitation

We have delimited the Launch sites,
Knowing these we have a proper understanding of the area and their surroundings

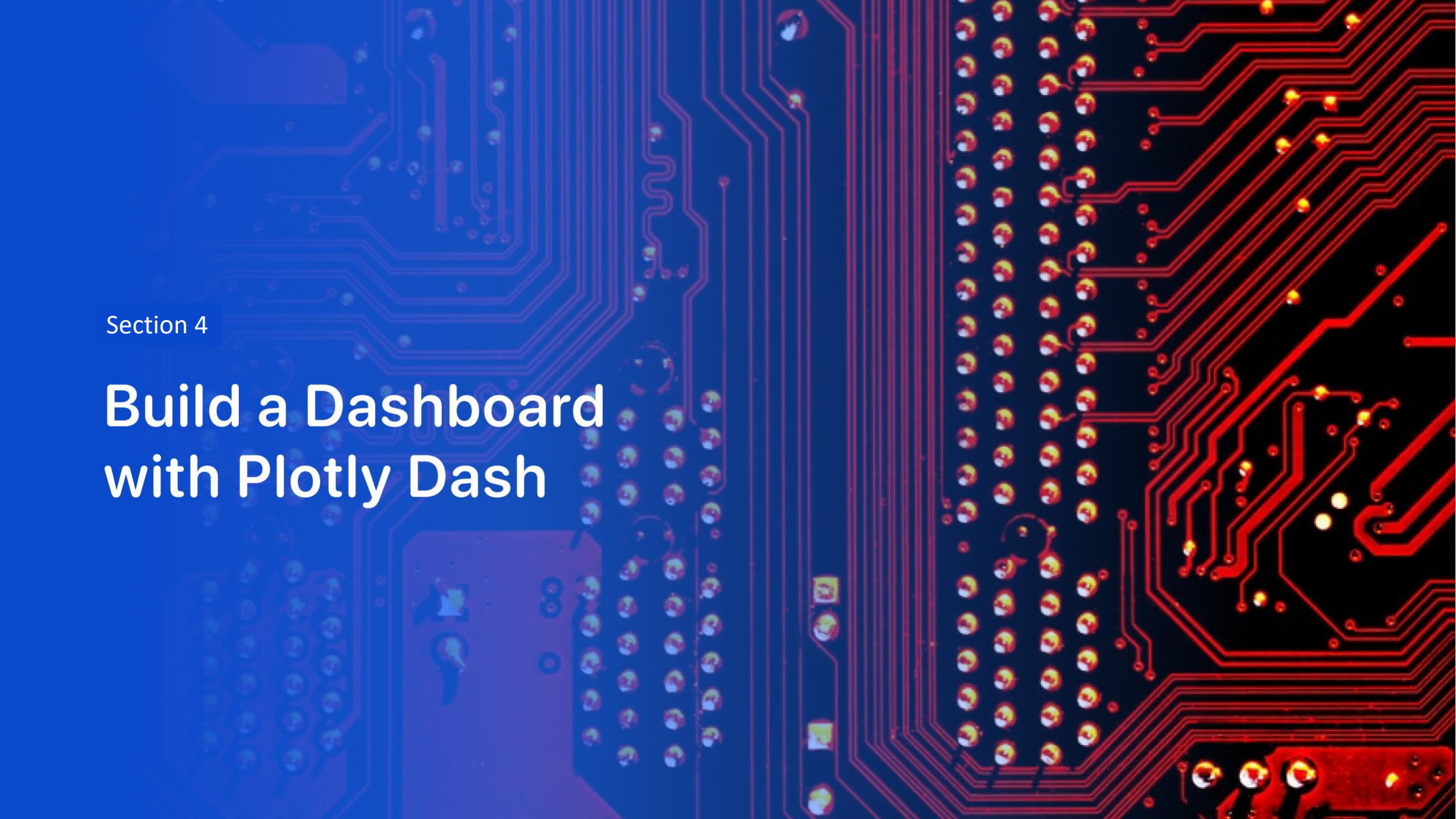
<Proper outcome and failure color launches>



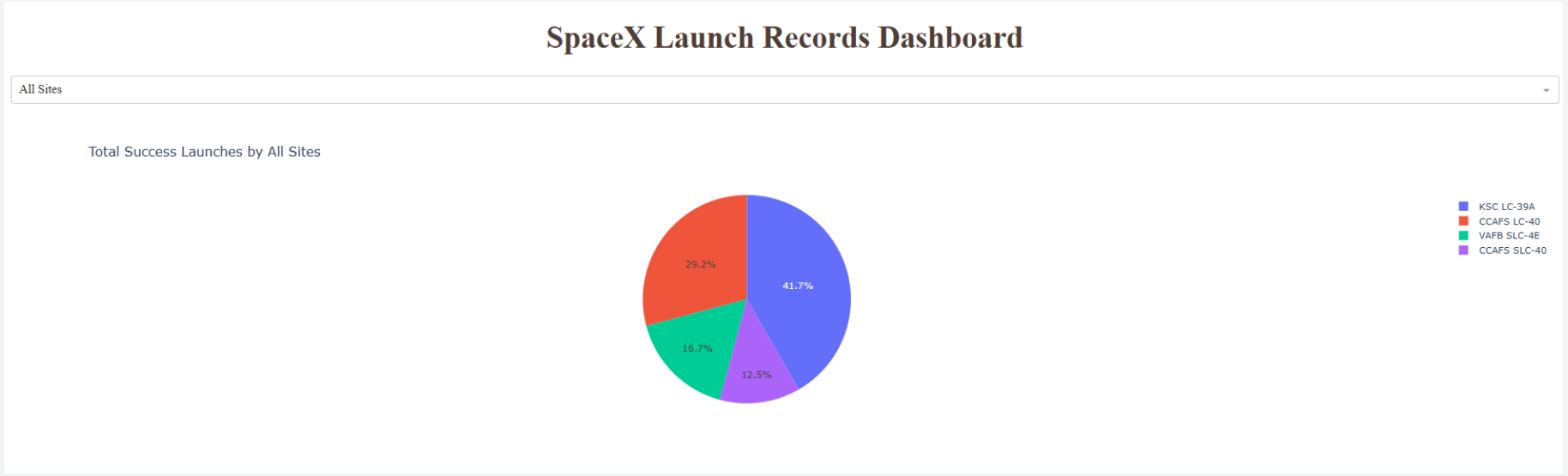
In here we can see an example of the Launch , represented by “Child markers”,
(Succesfull in green , unsucsesfull in red.) It os important because it gives us a visual representation of each launch

<Railway proximity example>

In here we can see two launch sites , wich are close to railroads . *A geographical understanding of the sorroundings allow us to make a proper analyses considering the correct utilization of resources and logistcs.*

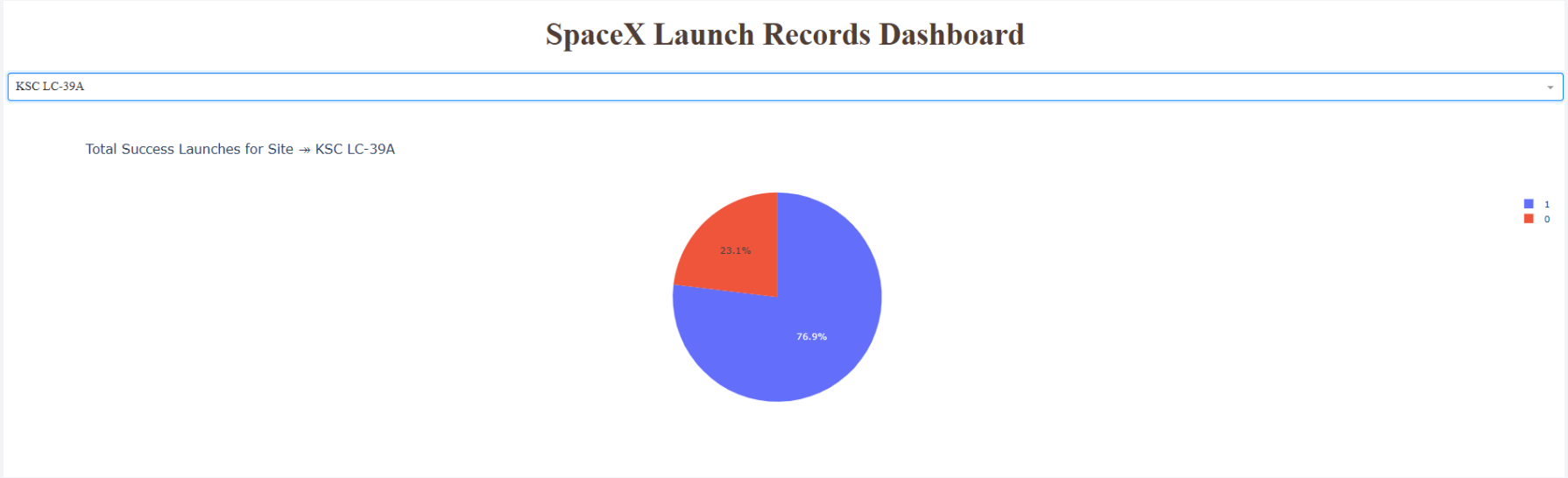

Section 4
Build a Dashboard with Plotly Dash

<Launch success Pie Chart>



- This here represents the Launch success Pie Chart for all sites, divided in 29.2% (CCAFS LC 40 launch site in color red) 41.7% (KSC LC-39A in color blue), 12.5%(CCAFS SLC-4E in color purple) , 16.7% (VAFB SLC-4E in color green)

<Highest launch success ratio>



- Here we can see that the highes launch success ratio , was found in KSC LC-39A . With an astonishing 76.9% succesfull rate

<Correlation between Payload and Launch Success >

In this scatter point plot we can see that between the range of 1000 kg to 10.000kg ‘ft’ booster version that the highesst rate. V1.1 and b5 booster version being the lowest

<Correlation between Payload and Launch Success >

In this scatter point plot ,we can see that between the range of 0 kg to 10.000kg ‘ft’ booster version has the highess success rate. ‘v1.1’ and ‘b5’ booster versions being the lowest .

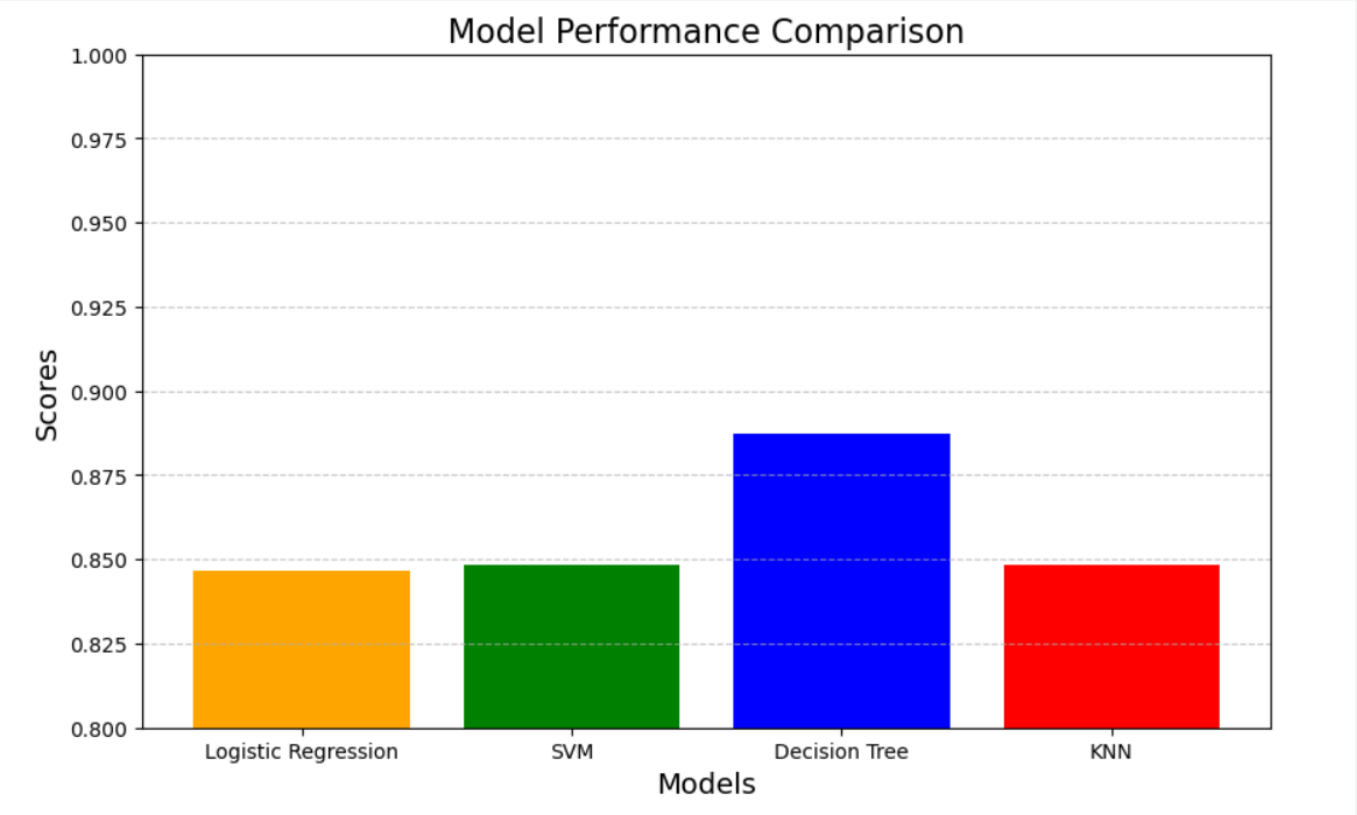
<Correlation between Payload and Launch Success >

In this scatter point plot ,we can see that between the range of 1000 kg to 5000kg ‘ft’ booster version has the highest success rate. ‘v1.1’ and ‘b5’ booster versions being the lowest .

Section 5

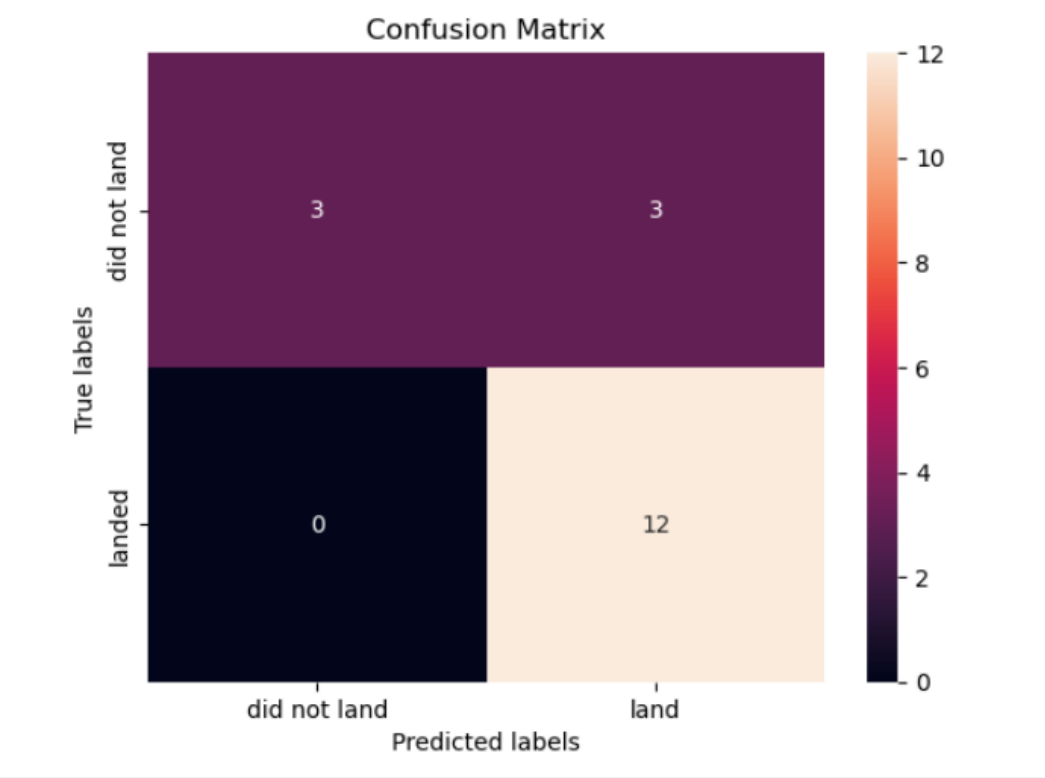
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

- Best confusion matrix performing: ***Decission Tree***



Model with an explanation

Decion tree predictive model , give us the higer success Rate performance in determine of wheter or not we have **completion and safe return of the first stage of the launch rocket.** This confussion Matrix is proof of that.

Conclusions

- In this project, we try to predict the landing outcome of a given falcon 9 launch in order to determine its cost and implement the same system to reduce operational costs, and *win the space race*
- We have consider each feature of the Falcon 9 Launch and determine that they individually and collectively have an affect in the mission. Therefore its *importance of the analys .*
- We employed several machine learning algorythims to learn patterns In data of past falcon 9 launches, this in order to produce predictive models that can produce the dessired outcome of a falcon 9 Launch.
- We have concluded that best predictive model by *Decission tree* perform the best among the 4 machine learning algorythims.

Appendix

Raw codes/Queries can be found in the determine links :

Data Collection Github permalink

Git hub webScrapping Link

Data Wrangling Git Hub link

EDA with Visualization Github url

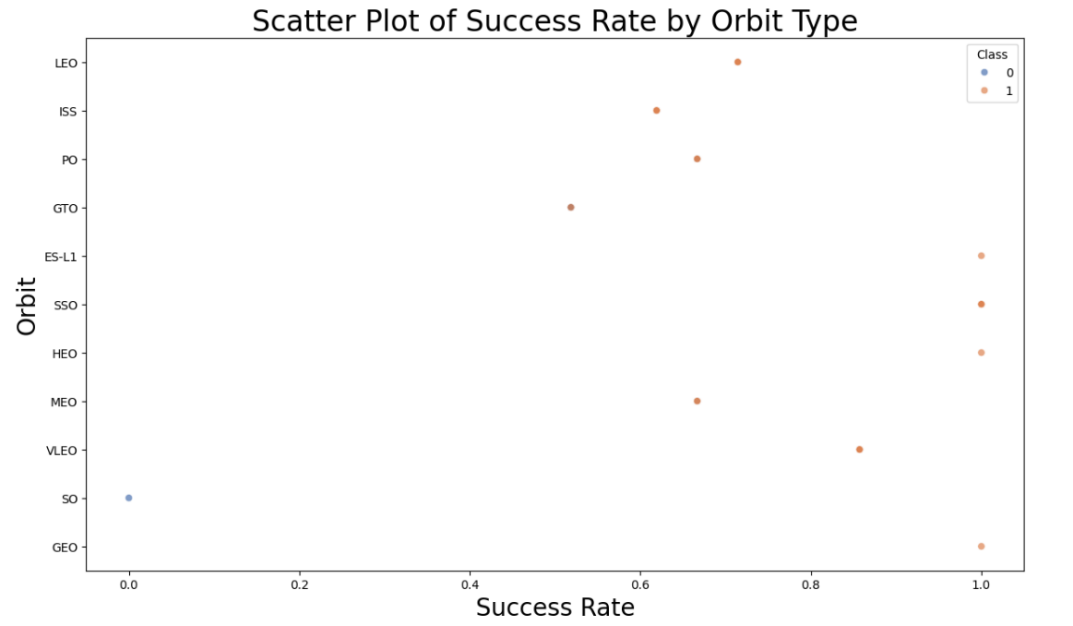
EDA With Sql Github link

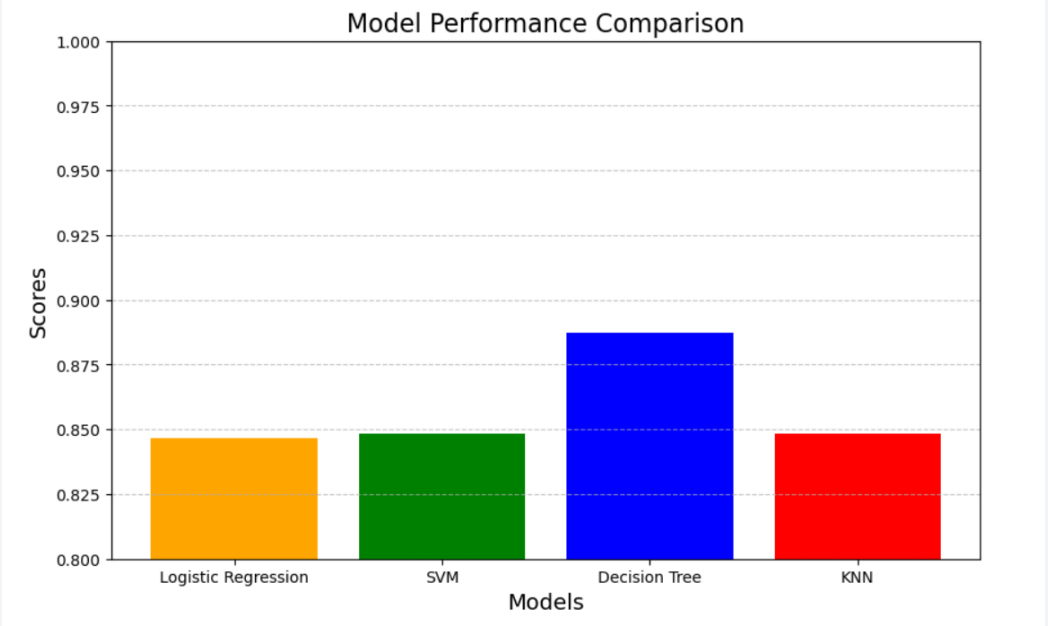
Folium , github code

Dash Github url Link

Predictive analytics Classification GitHub URL of your

Appendix / Additional Charts with code links

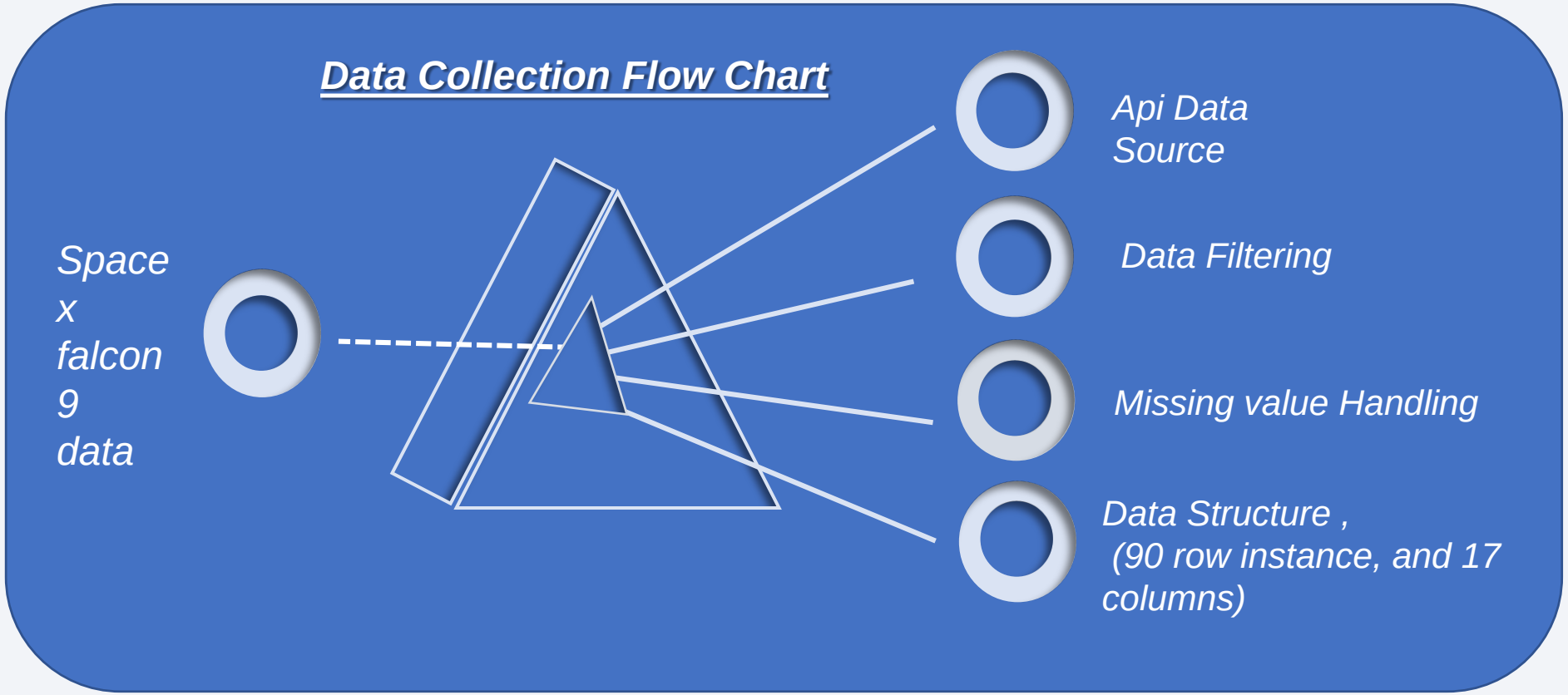




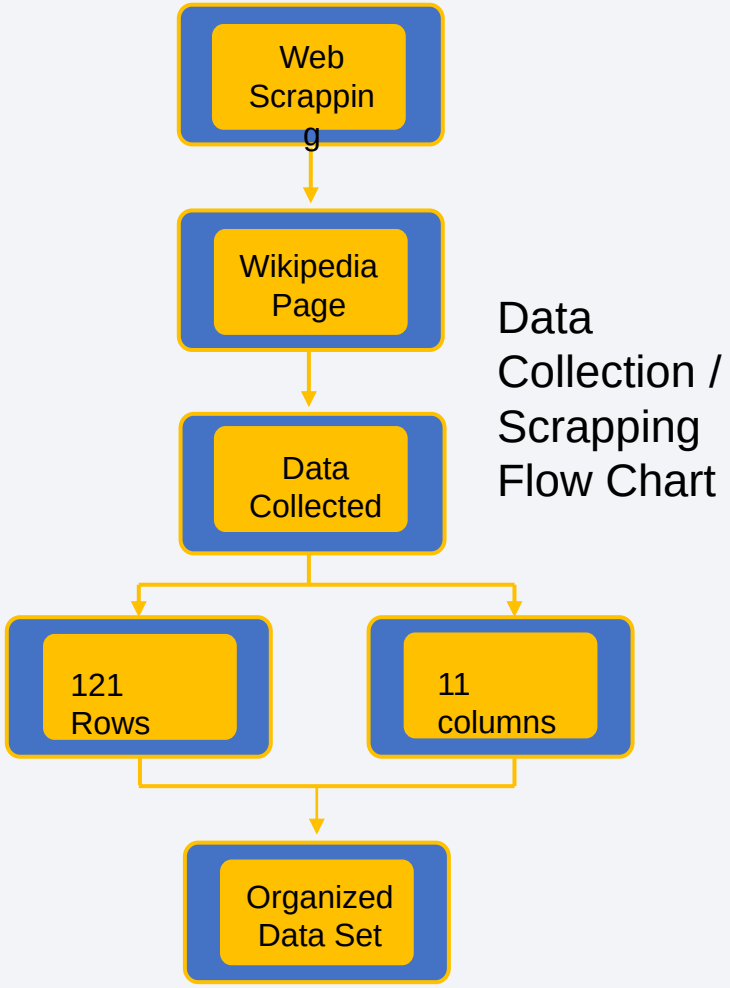
[Scatter Plot of Succes Rate vs Orbit type github url link]

[Model Performance Bar Chart, github Codes]

Appendix / Local (power point) Flow Charts

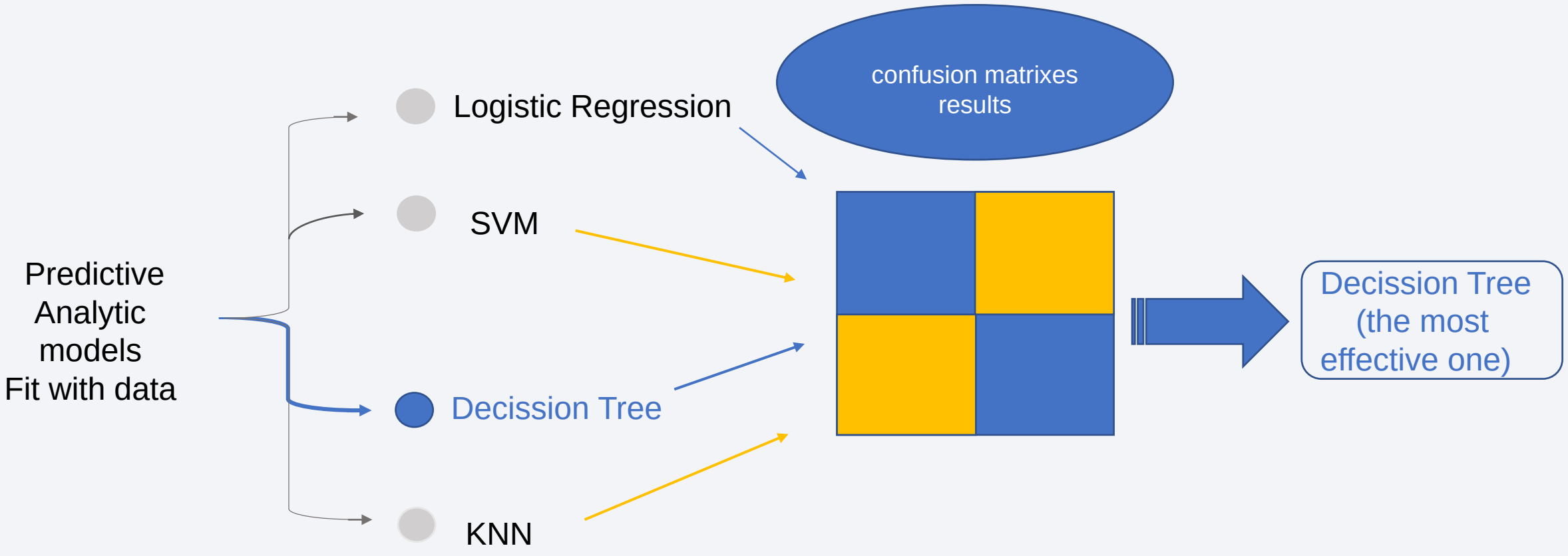


Appendix / Local (power point) Flow Charts



Data Collection / Scrapping Flow Chart

Appendix / Local (power point) Flow Charts



Thank you!