Falcon 9 Landing Prediction for Cost-Effective Rocket Launches using Data Science

- Name: Chirayu Ahirrao

Outline

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

Executive summary

•In this project, we will predict whether the first stage of the Falcon 9 rocket will land successfully.

•The reuse of the first stage of the Falcon 9 rocket will reduce costs by 30%.

•The methodology followed will include data collection, data wrangling, and data preprocessing.

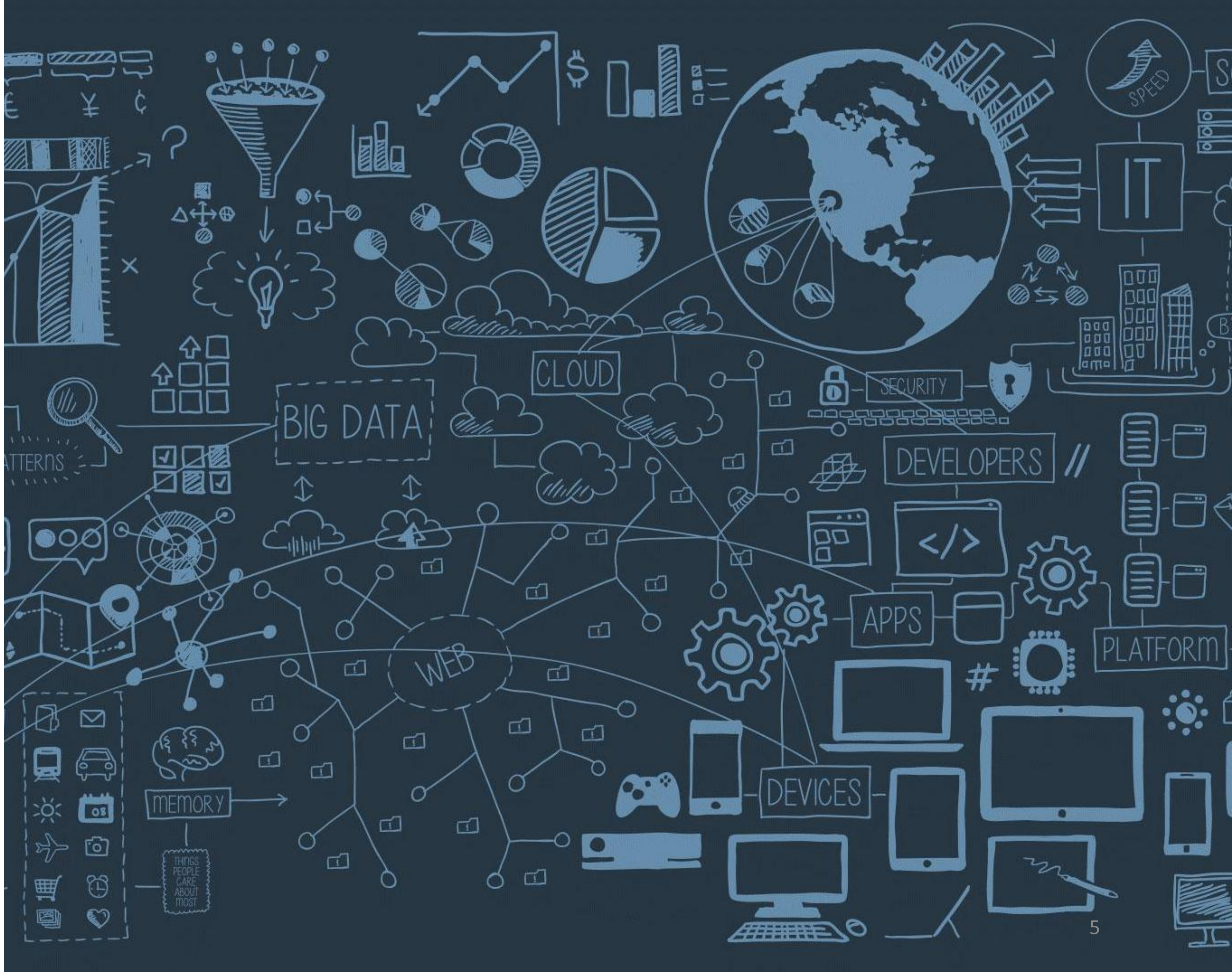
•After conducting the analysis, the outcomes indicate that some features are correlated with success and failure.

•Finally, on the basis of results we can conclude that the decision tree classifier method is the best fit for this analysis.

Introduction

- Falcon 9 is a partially reusable medium-lift launch vehicle.
- By reusing the first stage of the rocket, costs can be reduced by nearly 30%, allowing the company to stay competitive.
- The primary goal of this project is to predict the successful landing of the first stage of the rocket.
- In other words, we aim to answer the question: "Will the first stage of the rocket land successfully for a given set of features?"

Section 1: Methodology



Summary of methodology

Data have been collected by two methods: From SpaceX REST API and from web scraping of Wikipedia page. Data wrangling was used for transforming and cleaning data.

After cleaning data, exploratory data analysis (EDA) was performed using data visualization tools such as Python's Matplotlib and Seaborn libraries. SQL will be used for answering questions. Maps are drawn using Folium library for analyzing some other insights.

Following this comprehensive analysis, predictive analysis will be conducted using various machine learning models, including Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Decision Trees.

Data Collection: SpaceX rest API

1.request and parse the SpaceX launch data using get request.

2.Ceate a dataframe which contains normalized JSON response.

3. remove unwanted columns using auxilary functions.

4.Filter the data such that it has only falcon 9 launches.

5. handle missing values.

6. Export to CSV file

GitHub URL : [1.data-collection-api.ipynb]

Data collection: Web-scrapping

1.request the launch data from Wikipedia.

2.Extract all columns from the HTML table header.

3.create a Data frame by parsing the HTML tables.

4.Export to CSV file.

GitHub URL : [2.webscraping.ipynb]()

Data Wrangling

1.Count the number of launches on each site.

2.Count the number and occurrence of each orbit.

3.count the number and occurrence of mission outcome per orbit type

4.creating a landing outcome label from outcome column

5.Export to CSV.

GitHub URL : 3.data_wrangling.ipynb

EDA with Data Visualization

- Scatterplots represent the relationship between two variables, including:
    - Flight number vs. payload mass
    - Flight number vs. launch site
    - Payload mass vs. launch site
    - Flight number vs. orbit
    - Orbit vs. payload mass

- Bar Charts : Bar charts are utilized to compare different categorical or discrete variables, specifically using horizontal bar charts to compare success rates with different orbits.
- Line charts are employed to display trends, with a focus in this project on tracing the change in success rates over the years.

GitHub URL : [5.EDA with visualizatioj.ipynb]

EDA with SQL

- Display unique launch site names.
- Show 5 records with launch sites starting with 'CCA'.
- Calculate the total payload mass carried by NASA (CRS) boosters.
- Find the average payload mass carried by booster version F9 v1.1.
- Identify boosters with successful drone ship landings, payload mass between 4000 and 6000, and launch site names in 2015.
- List the total number of successful and failure mission outcomes.
- Determine the names of booster versions that carried the maximum payload mass.
- Highlight failed landing outcomes on drone ships, including their booster versions and launch site names in 2015.
- Rank the count of landing outcomes between the dates 2010-06-04 and 2017-03-20 in descending order.

GitHub URL : [4.eda-sql.ipynb]

Build an Interactive Map using Folium

We crafted objects and integrated them into a Folium map. Markers highlighted launch sites and their success/failure records, while Lines measured distances between launch sites and their nearby locations.

By analyzing these maps, we can answer following questions :

- Are launch sites in close proximity to railways? Yes
- Are launch sites in close proximity to highways? Yes
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes

GitHub URL : 6.interactive visual analytics.ipynb

Dashboard building using plotly dash

- **Features of the Dashboard** :

  1. Drop down Bar : in the Dashboard we can see the dropdown bar in which you can put the launch site which you want to analyze.
  2. Payload mass slider : slide has also been provided to change the range of payload mass for plotting scatter plot for success vs payload mass.

- **Contents of the Dashboard** :

  1. Pie Charts: in Dashboard we can analyze probability of success and failure for each Launch site using Pie charts.
  2. Scatter charts: scatter plots are drawn to show relation between success and payload mass.

GitHub URL : 9.spacex_dash_app.py

Predictive Analysis

1.create Class column and Standardize the data.

2. Make a split of data in test and train set.

3.Find best Hyperparameter for SVM, Decision Trees, KNN and Logistic Regression.

4. Evaluate the models using test data based on their accuracy scores and confusion matrix.

GitHub URL : 8.Machine_Learning_Prediction.ipynb

Results

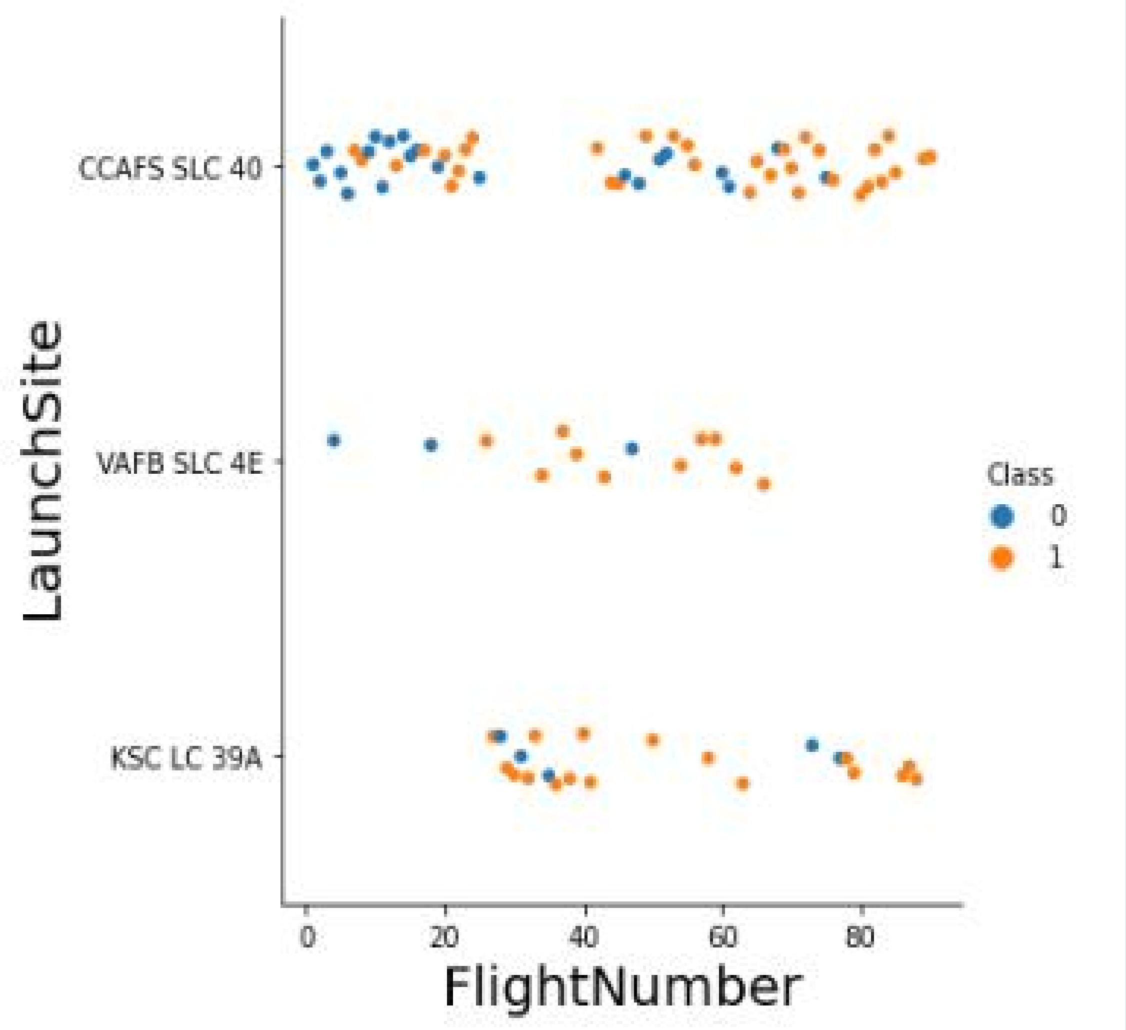
- From the results we get from Exploratory Data analysis we can say that success rate of Falcon 9 rocket landing is 66.66%

- After evaluating the models in predictive analysis, Decision tree model is the best fit with the accuracy of 94%

Section 2:

Insights Drawn from EDA

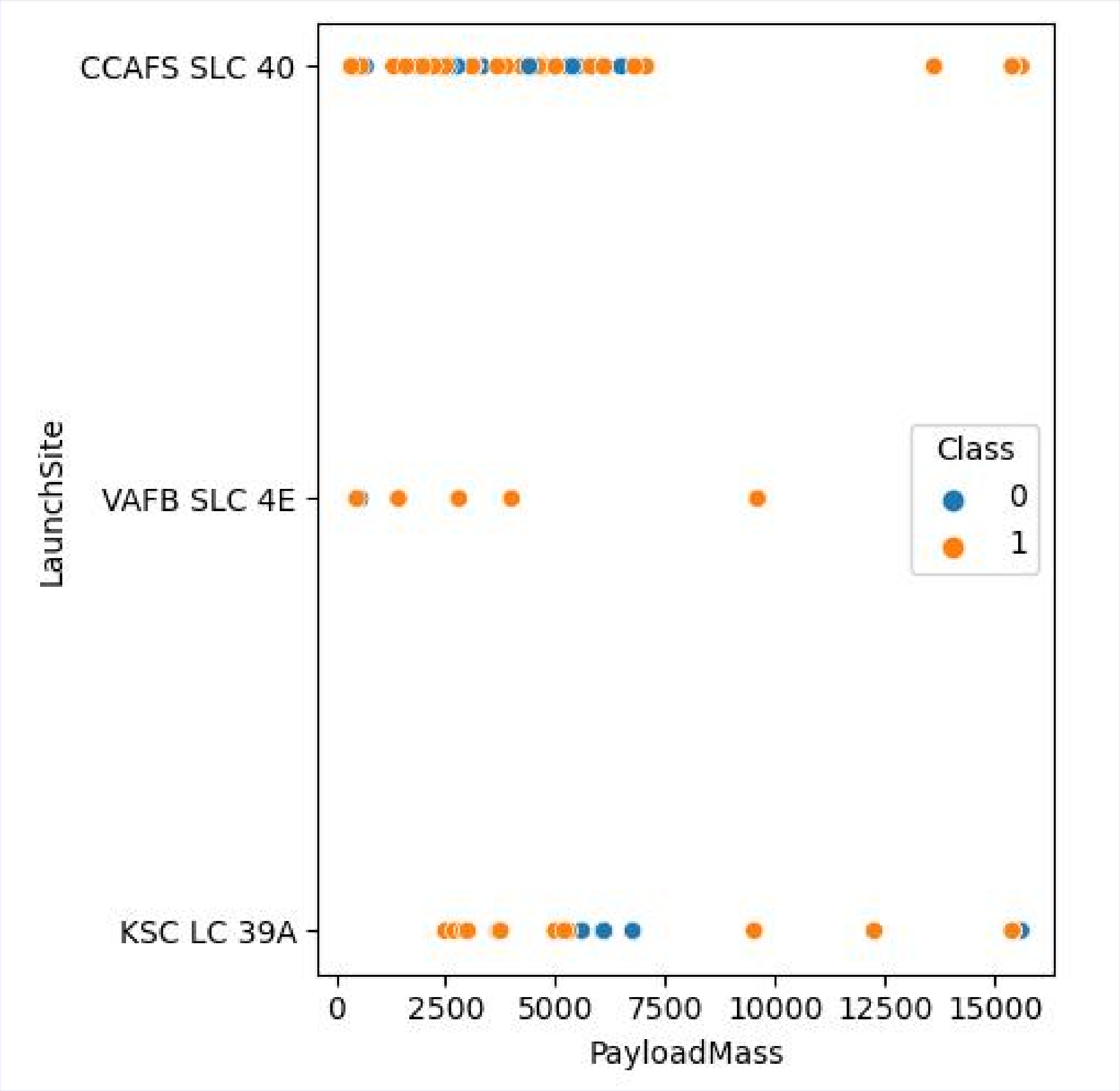
Flight number vs Launch Site



- In this figure we can see red and blue dots,
  Blue dots represents Failure,
  Red dots represents Success.

- In this plot we can see that probability of success is increased with increase in number of flights.

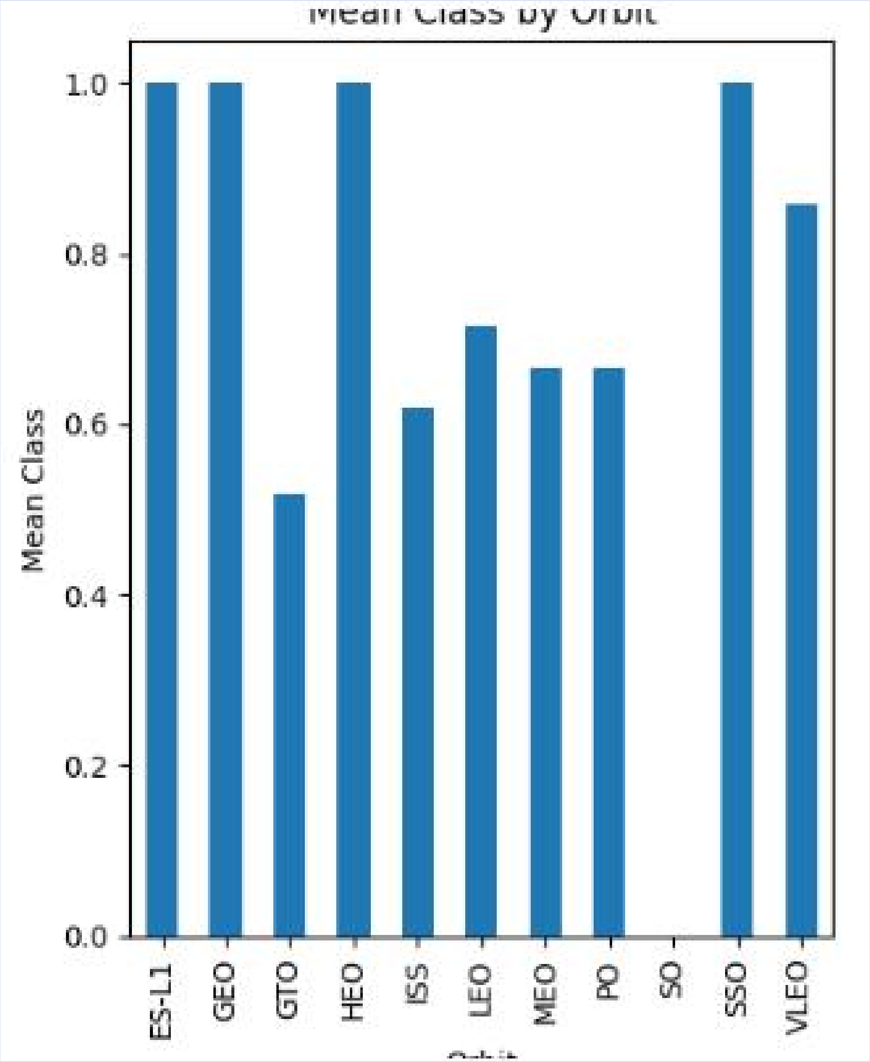
Payload vs Launch site



- In given plot we can see that, for VAFB-SLC launch site there are no heavy payload masses.

- There is very weak correlation between Payload and launch site

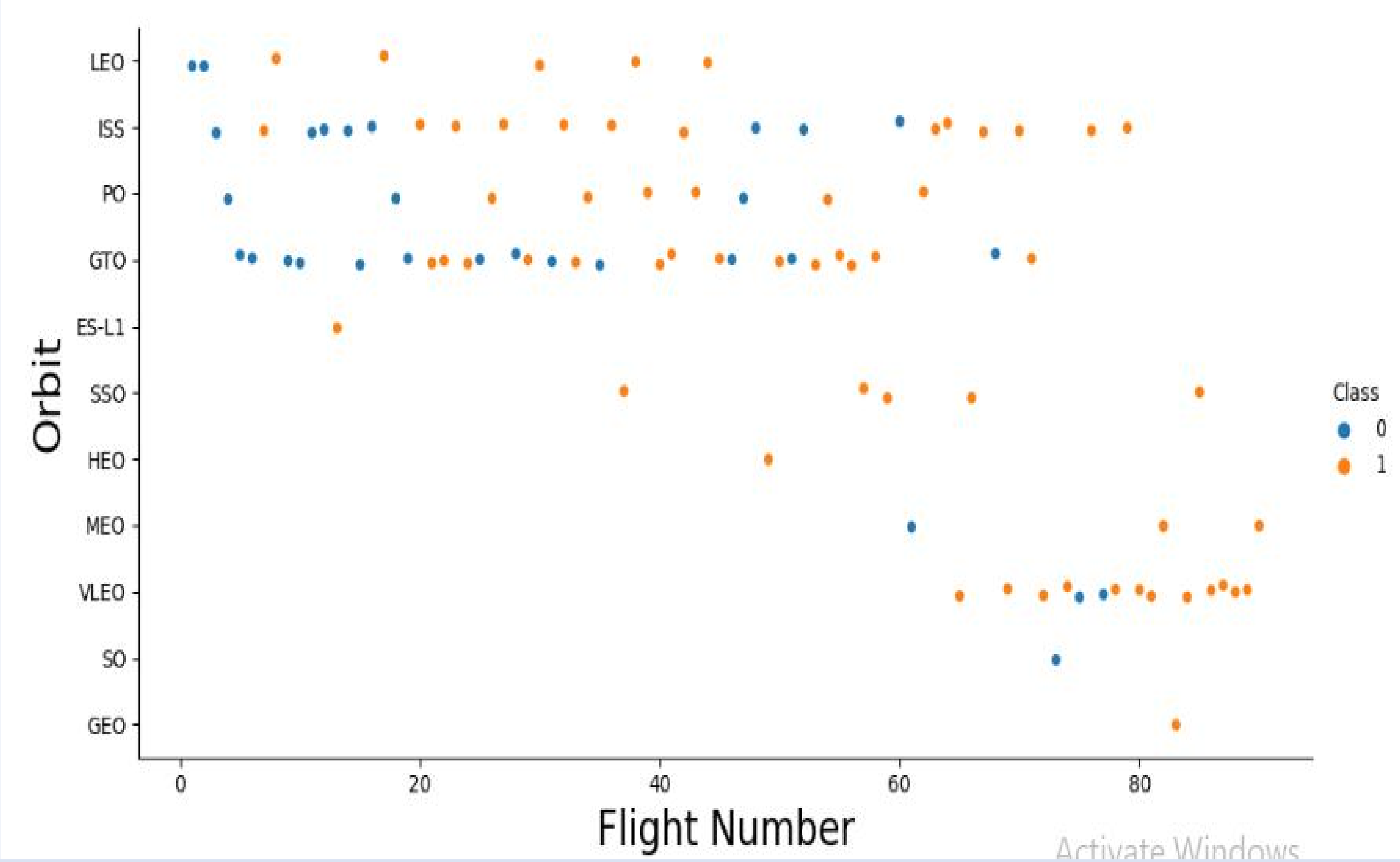
Success rate vs orbit type



- In this chart we can see that there are some orbits having success rate 100% which are, ES-L1, GEO, HEO, SSO.

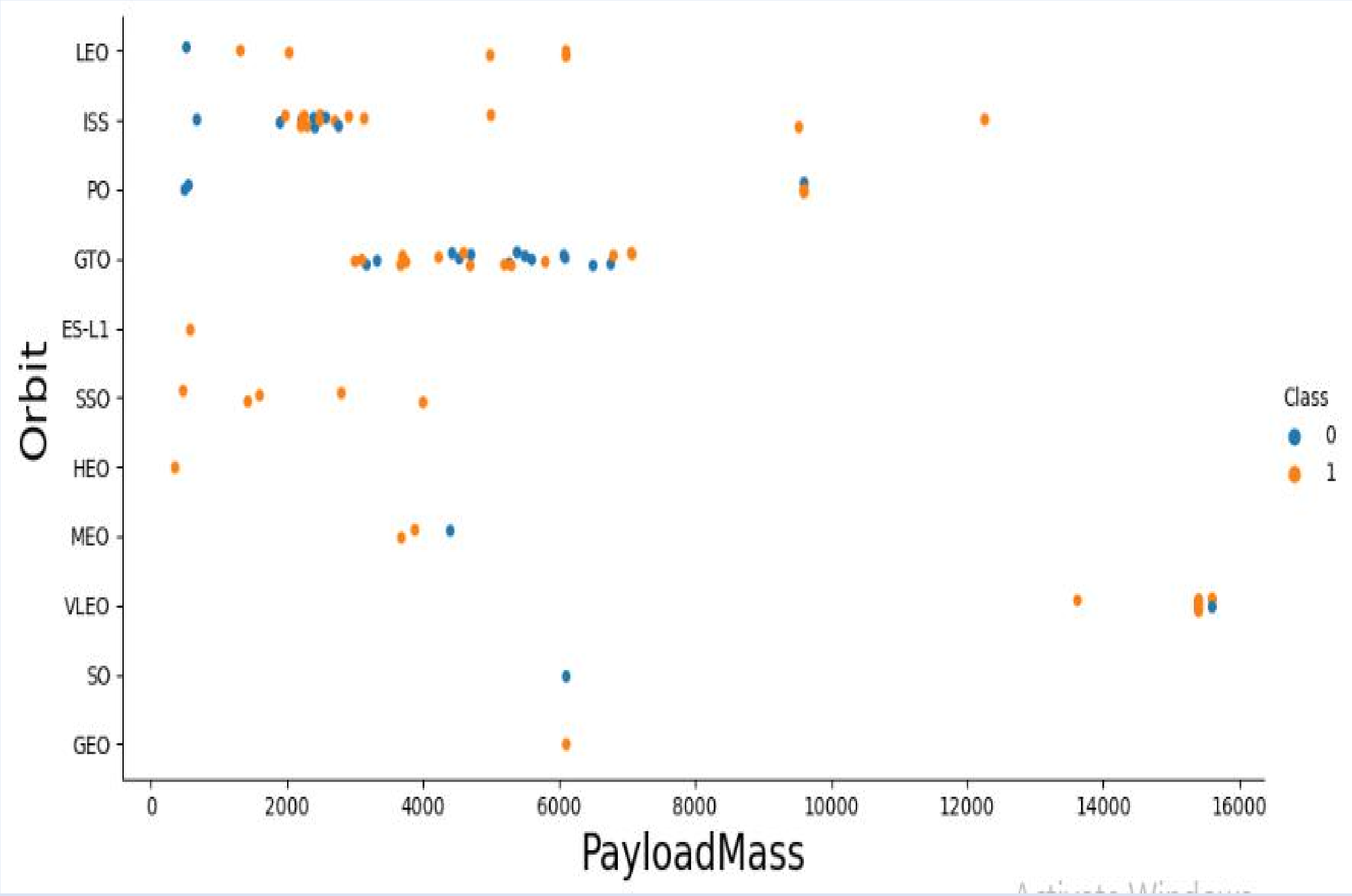
- Also we can see that orbit SO did not have successful launches.

Flight no. Vs orbit type



- In this chart we can see that the success rate of orbit LEO is positively correlated with Flight No.
- There seems to be no relationship between Flight no. And success rate for GTO orbit.
- Flight no. More than 40 have high success rate.

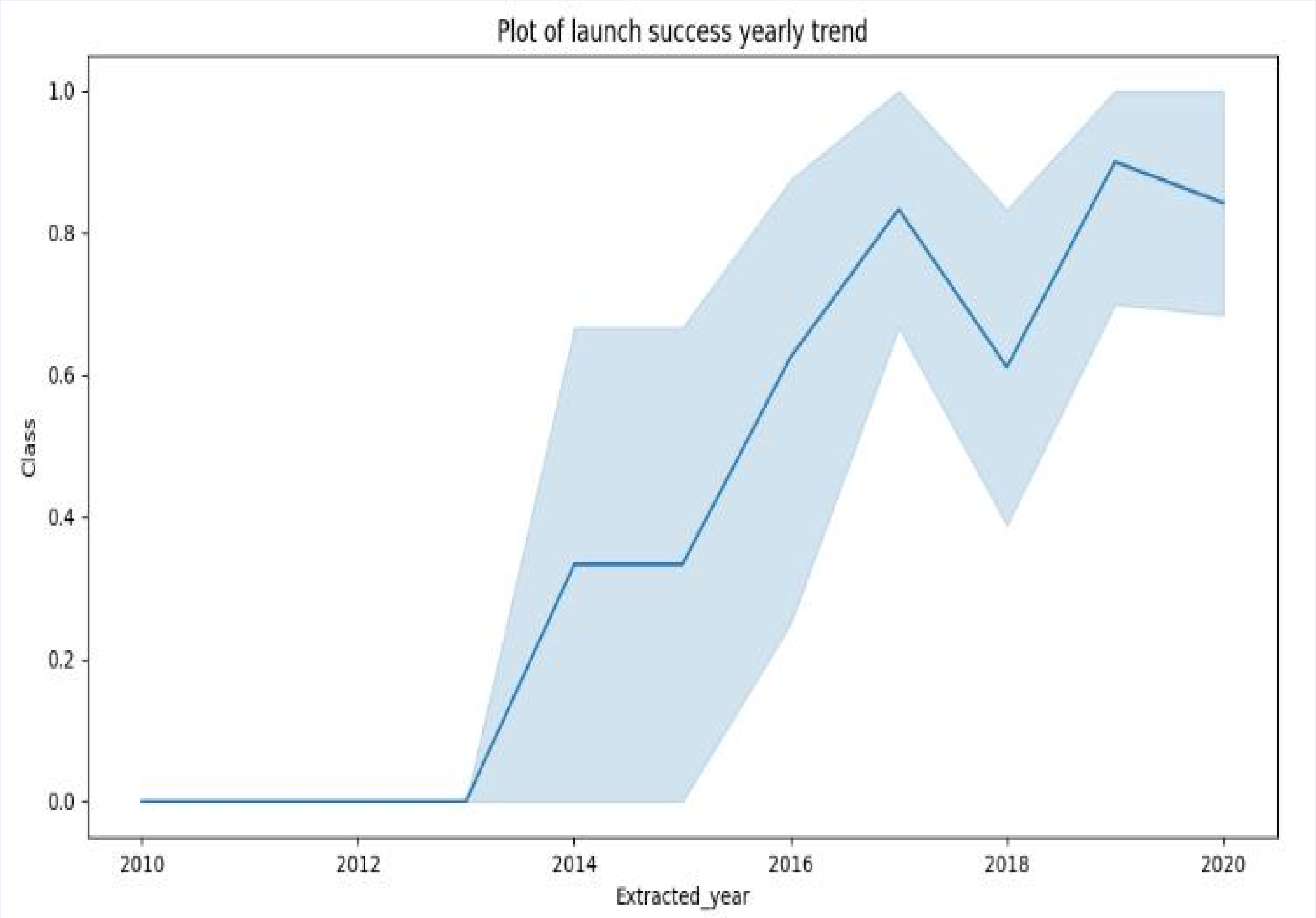
Payload vs orbit type



- In this figure we can see, for orbits LEO, ISS, SSO, PO success rate increases with increase in payload mass.

- There seems to be no correlation between orbit and payload
mass for orbit GTO.

Launch success yearly trend.



- In this trendline we can see that the launch success is trending up. There is a short dip in the year 2018.

Names of the sites.

- Using DISTINCT command in SQL we can print all unique sites from column *launch_site*.
- The names of launch sites are, CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E .

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

Launch site names starting with 'CCA'

- Using SQL command LIKE we can return the names of the launch sites which starts from 'CCA'
- These are the top 5 results,

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

- We can calculate total payload mass of boosters by NASA using SUM() function.

Total payload mass
45596

Average Payload mass by F9 V1.1

- The AVG() function was used to calculate average, and where clause was used to specify Booster version as V1.1.

average payload mass
2928.4

First successful Ground landing date

- The MIN(Date)function was used to find the date of successful landing on ground pad.

- The where clause was used to specify landing outcome as 'Success (ground pad)'.

MIN(Date)
2015-12-22

Successful Drone ship landing with payload between 4000 and 6000.

• The BETWEEN clause was used to retrieve only those results of payload mass greater than 4000 but less than 6000. The WHERE clause filtered the results to include only boosters which successfully landed on drone ship

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

Total Number of Successful and Failure Mission Outcomes

- To calculate total number of occurences of different mission outcomes COUNT() function was used with the help of GROUPBY clause on the mission_outcome column.
- There are total 101 launches out of which 99 mission outcomes are successful.

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

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

- The MAX() function was used in a subquery to retrieve a list of boosters which have carried the maximum payload mass

Launch records for the year 2015

- The Year(Date) function is used for retrieve only those rows containing year 2015 as a Launch date.

Landing_Outcome	Booster_Version	Launch_Site	substr(Date, 6, 2)
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	04

Ranking of the landing outcomes between 2010-06-04 and 2017-03-20.

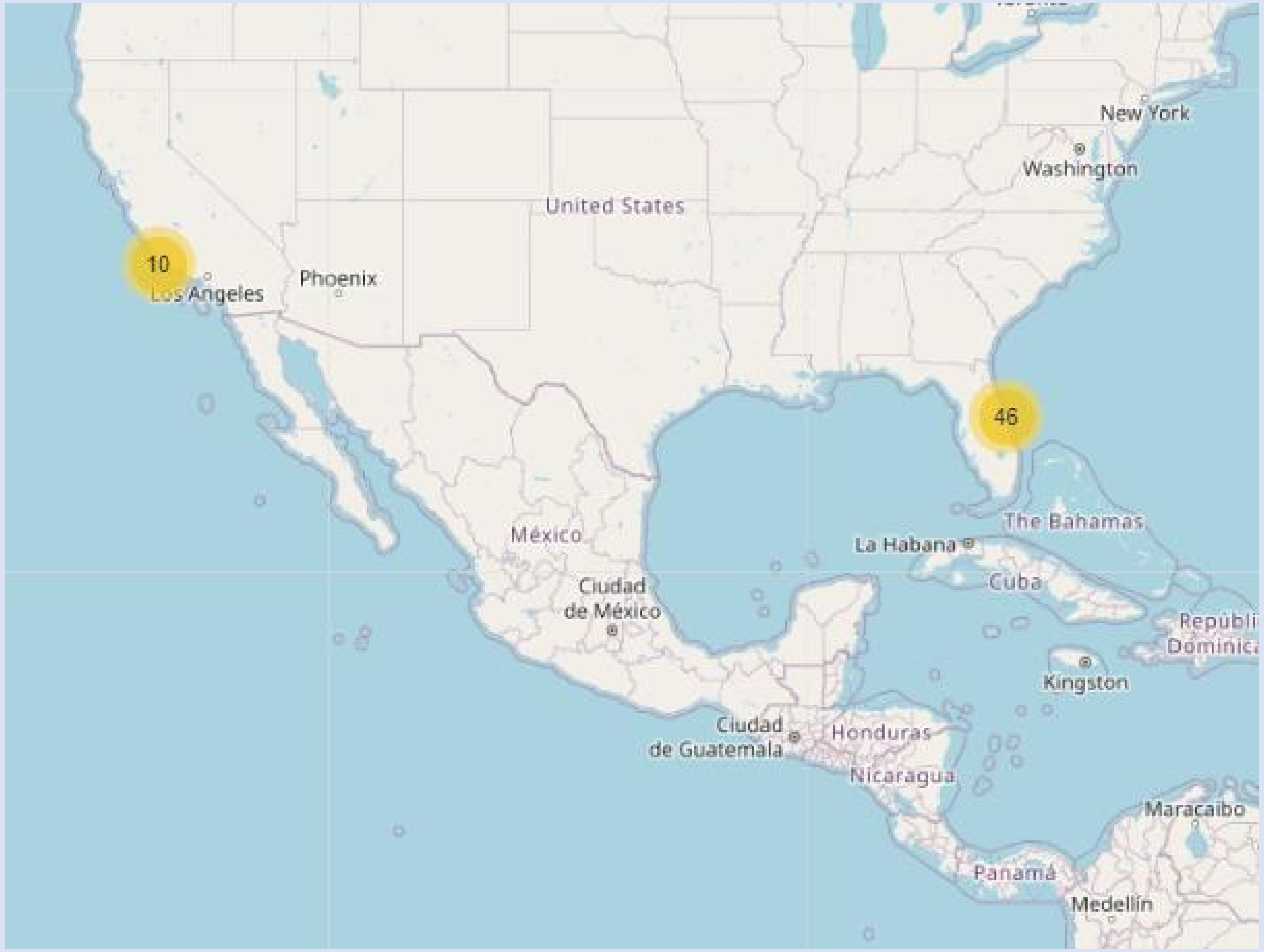
- COUNT() function was used to count the different *landing outcomes.* The WHERE and BETWEEN clauses filtered the results to only include results between 2010-06-04 and 2017-03-20. The GROUPBY clause ensure that the counts were grouped by their outcome. The ORDERBY and DESC clauses were used to sort the results by descending order.

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

Section 3:

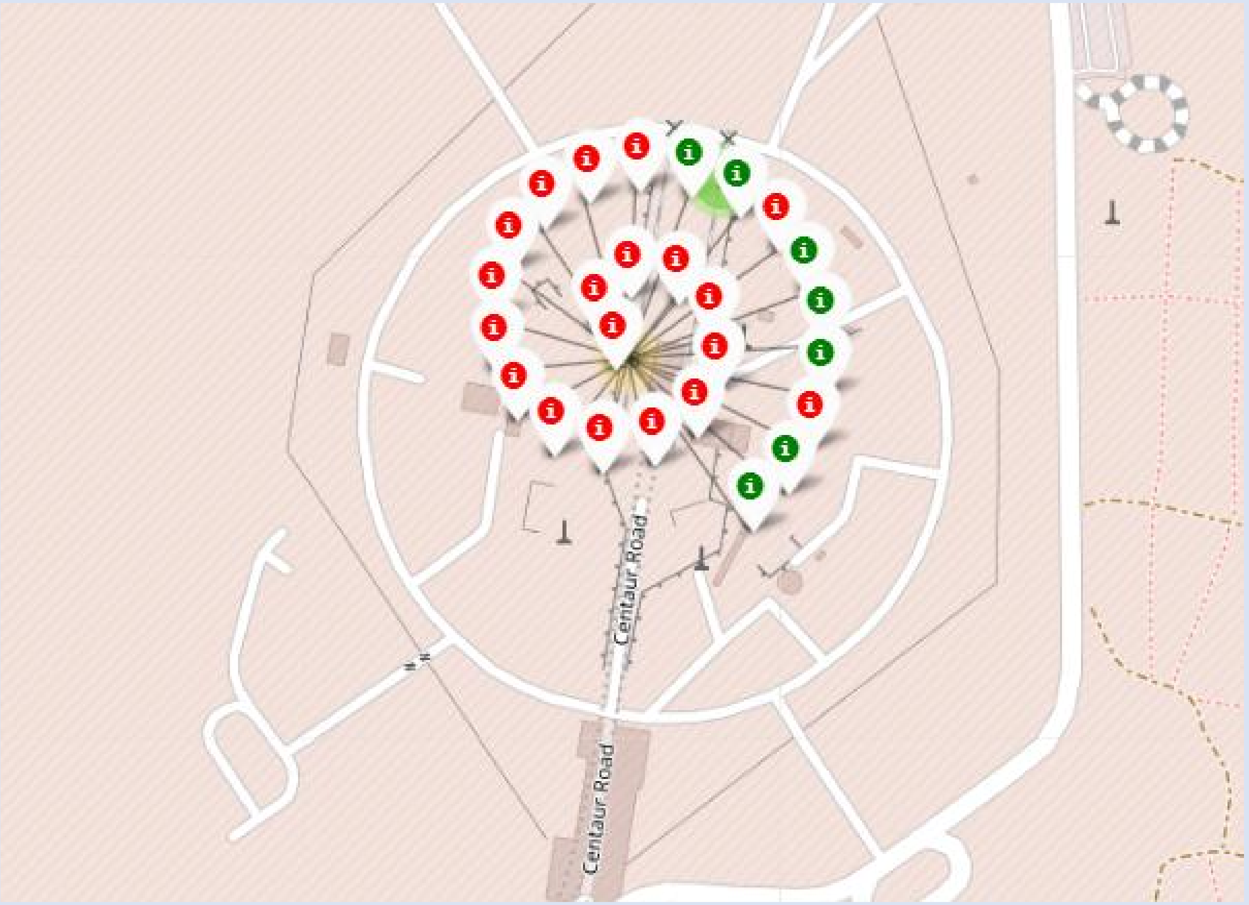
Launch Sites Proximities Analysis

Locations of Launch sites.



- The yellow markers on the map shows the Launch sites of the SpaceX in US.
- Both the Launch sites are placed near coast.
- The numbers shown indicates the number of Launches.

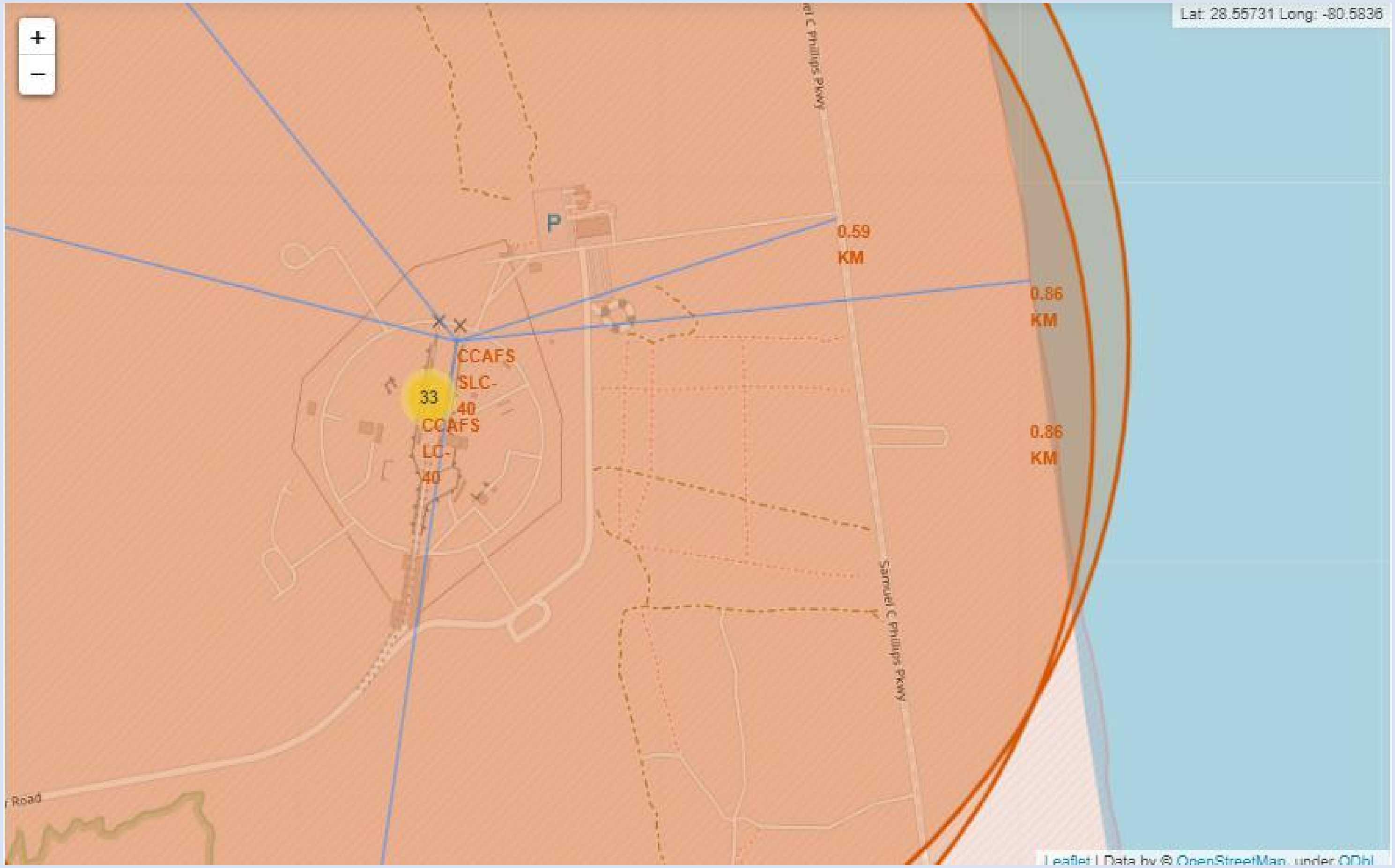
Zooming out the map



- Zooming out the map and clicking the launch site will show the marker clusters over there.

- This figure will indicate the number of successful (green) and unsuccessful landings (Red) on that particular site.

Launch Site Proximities

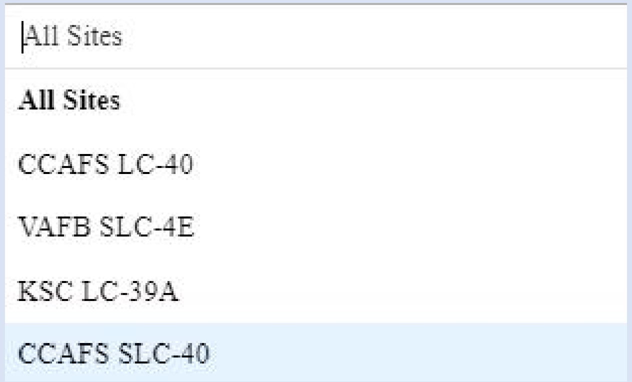


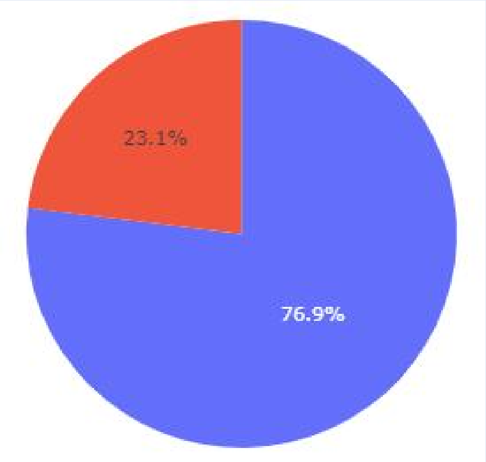
- The map shows that Launch site is near Coast and Highways for transportation of the equipment required.
- Also it shows that launch site maintains enough distance from city.

Section 4:

Building Dashboard

Overview of Dashboard



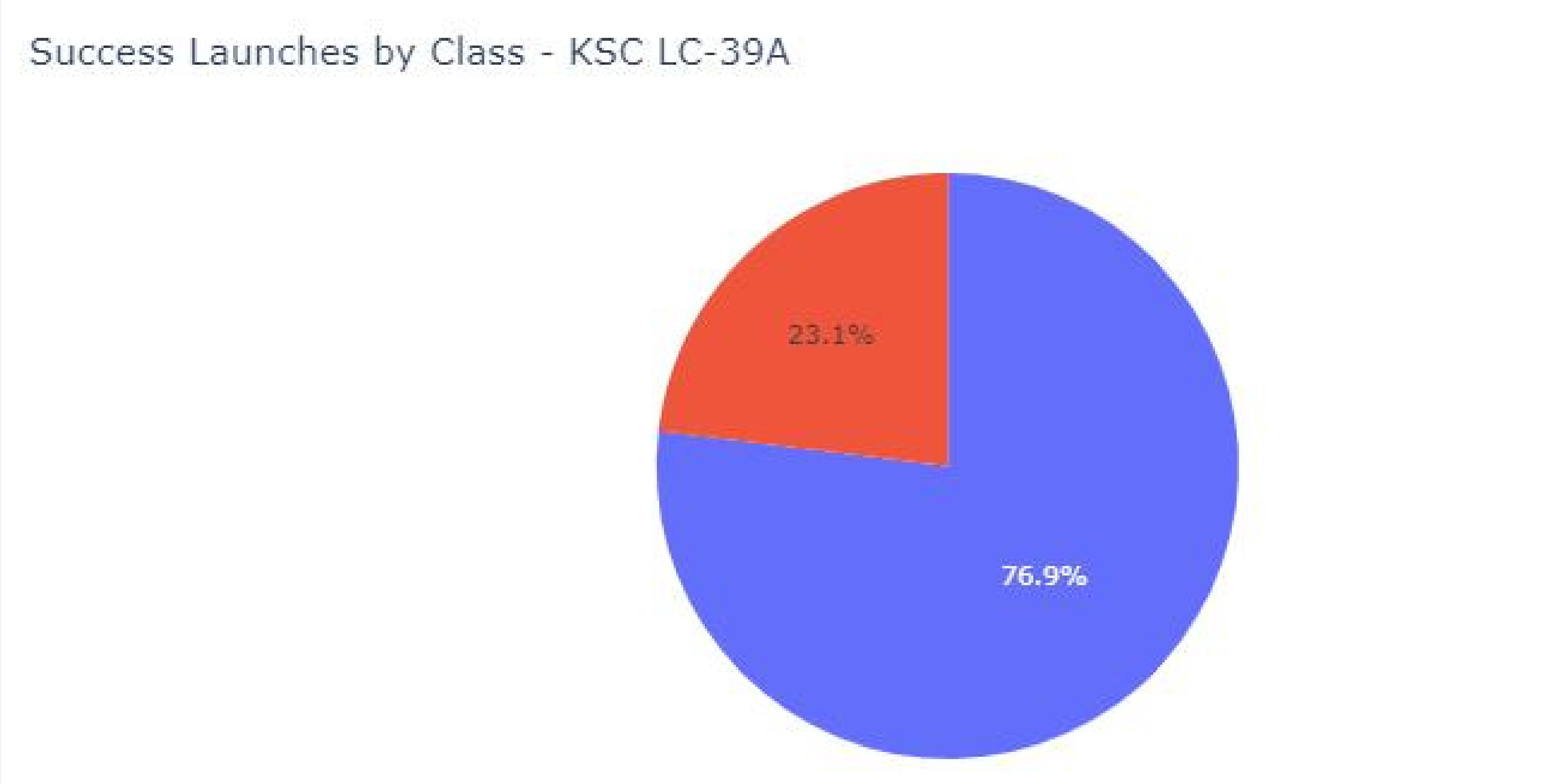


- Drop down bar: by clicking dropdown bar you can select the Launch site which you want to analyze.

- Pie chart and Scatter plot: Pie chart represents success of Launches in % while Scatter plot represents correlation between payload and Launch success.

Launch Site with highest success ratio

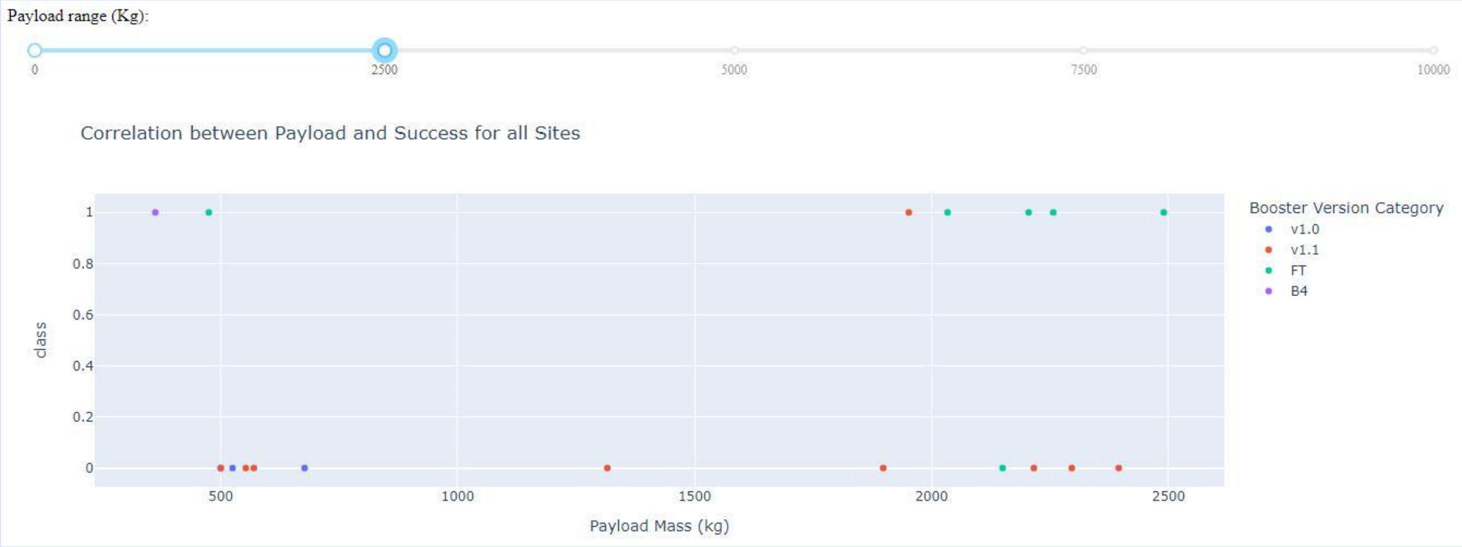
- The KSC LC-39A is the Launch Site with highest Success ratio of 76.9%



Payload Vs Launch outcome

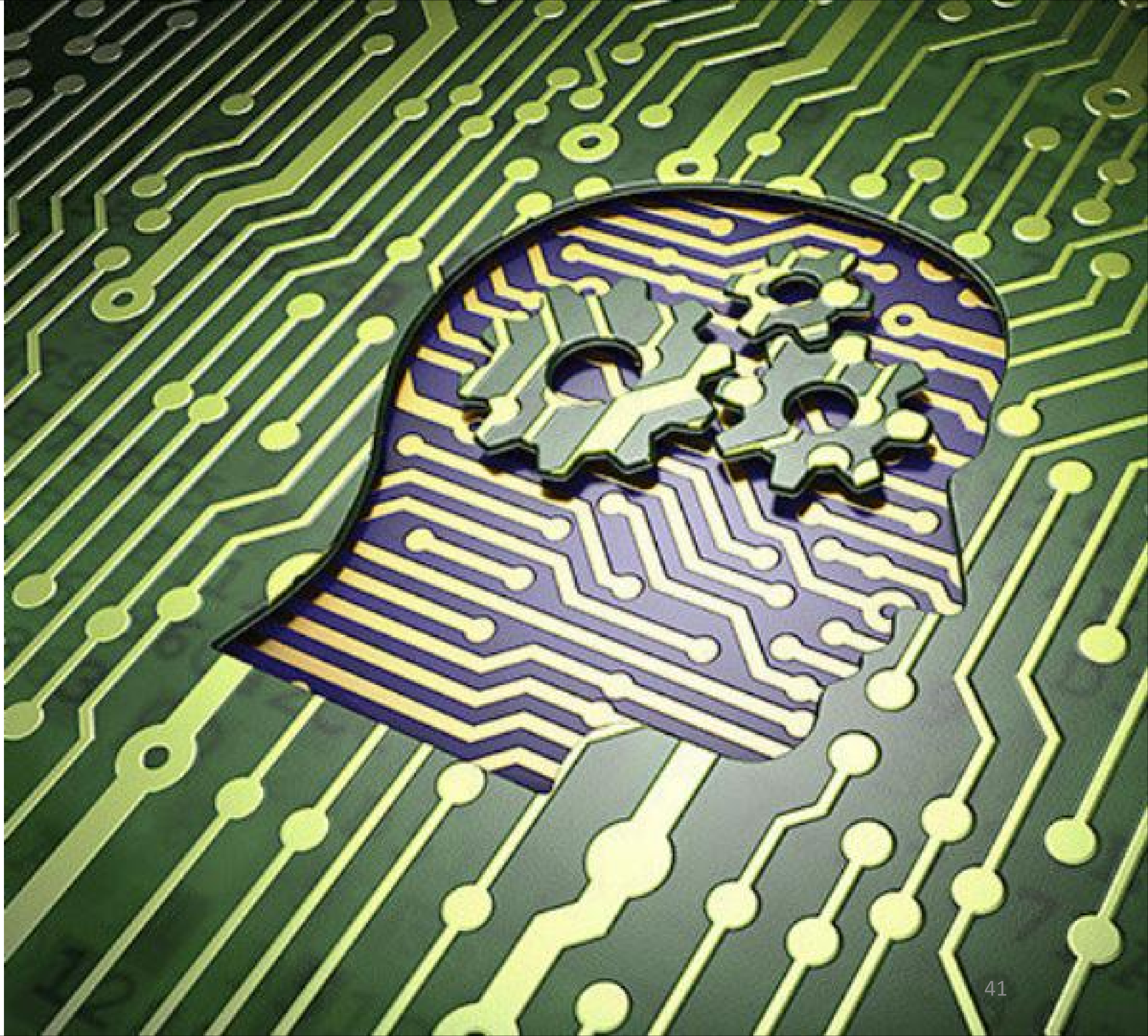
- The launch success rate for payloads 0-2500 kg is slightly lower than that of payloads 2500-5000 kg. There is in fact not much difference between the two.

- The booster version that has the largest success rate, in both weight ranges is the ***v1.1***



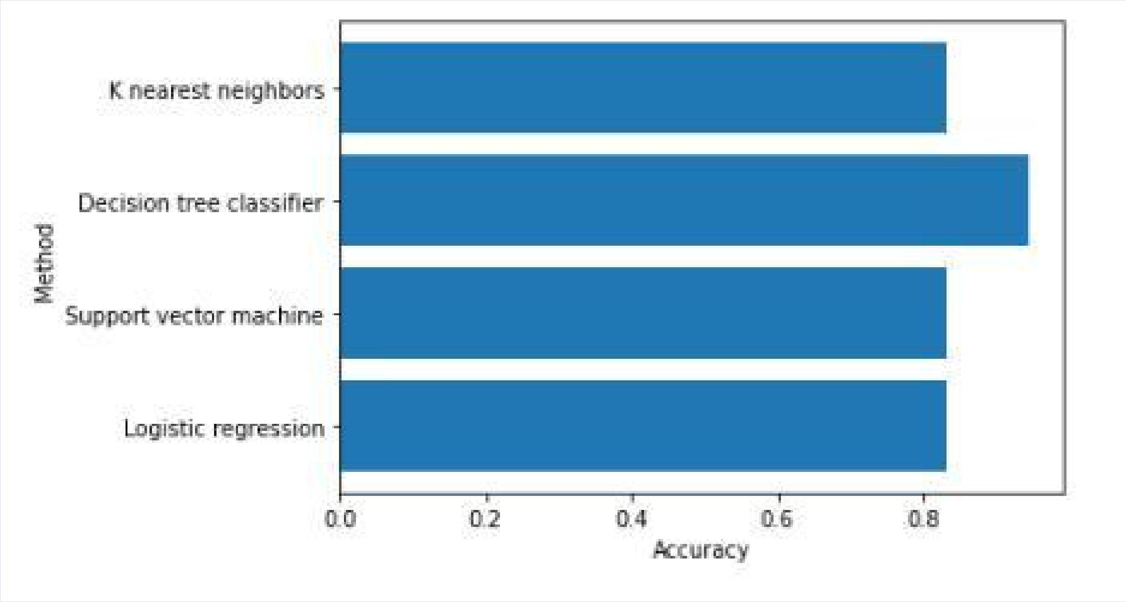
Section 5:

Predictive Analysis



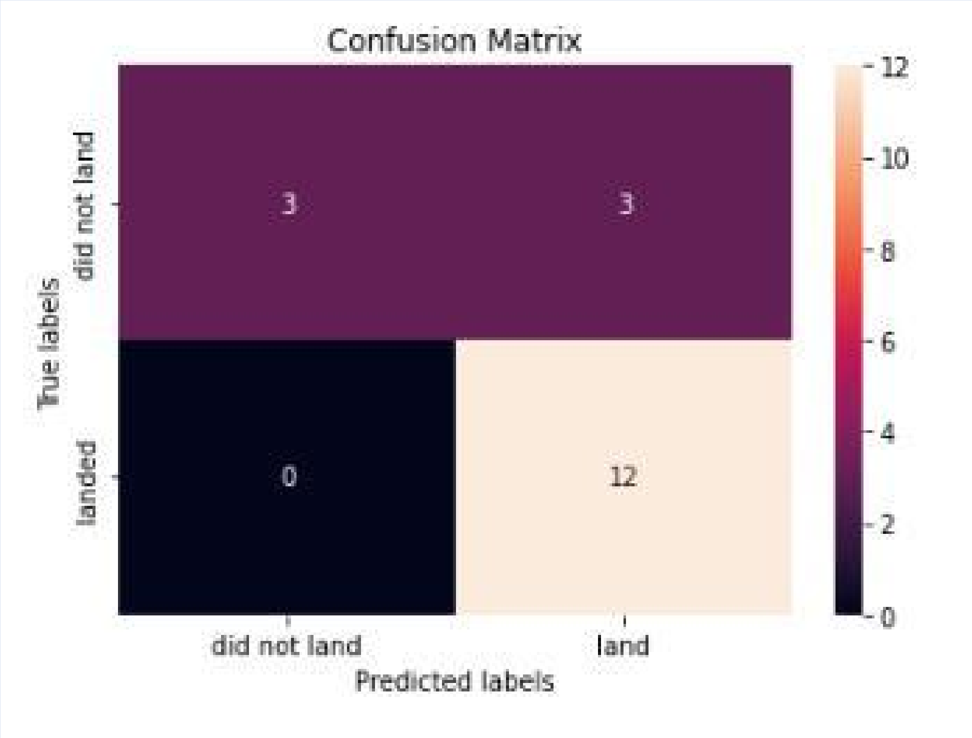
Classification Accuracy

- Out of all 4 classifier Decision Tree Classifier has highest accuracy of 94%.



	method	accuracy
0	Logistic regression	0.833333
1	Support vector machine	0.833333
2	Decision tree classifier	0.944444
3	K nearest neighbors	0.833333

Confusion Matrix



- The model predicted 12 successful landings when the True label was successful (True Positive) and 3 unsuccessful landings when the True label was failure (True Negative).

- The model also predicted 3 successful landings when the True label was unsuccessful landing (False Positive).

- The model generally predicted successful landings.

Conclusions

- We can see the uptrend of success rate over the years.
- The orbits like SSO, HEO, GEO and ES-L1 has most successful launches.
- Relation between success rate and Payload mass shows that the probability of successful landing increases if we reduce the payload.
- The maps show that the Launch sites are strategically placed near the Highways and coast for the transportation purpose, but also away from the cities for safety.
- Decision tree classifier has highest (94%) accuracy. Therefore it will be the best predictive model.

Appendix

GitHub Repository : https://github.com/Chirayu-spec/Redefining-the-Space-access-with-Data-Science.git

Thank you!