SPACE-X FALCON-9 - ANALYSIS

by :- Sandeep Vashisth

11-July-2023

EXECUTIVE SUMMARY

SUMMARY OF METHODLOGY:

The analysis/research tries to identify the factors for a successful rocket landing. Below methodology has been used for this analysis:

1. **Collected** data using API and Web Scrapping techniques.
2. **Wrangled** data to create outcome variables.(success/fail)
3. **Explored** data using visualisation techniques.
4. **Analyzed** the data using SQL and calculated the statistics.
5. Explored the launch sites and proximity to geographical ranges using Folium.
6. **Visualized** the most successful launch sites and payload ranges.
7. **Build models** to predict the landing outcome using Logistic Regression, SVM and KNN.

RESULTS:

- Overtime the launch success has improved.
- Most launch sites are near the equator and close to the coast.
- All models performed similar, however tree model slightly outperformed.

TABLE OF CONTENTS

INTRODUCTION

- The SpaceX is a spacecraft manufacturer, launcher and satellite communication company.
- SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- We'll be exploring how payload mass ,orbits , number of flights affect the landings and the rate of successful landings overtime.

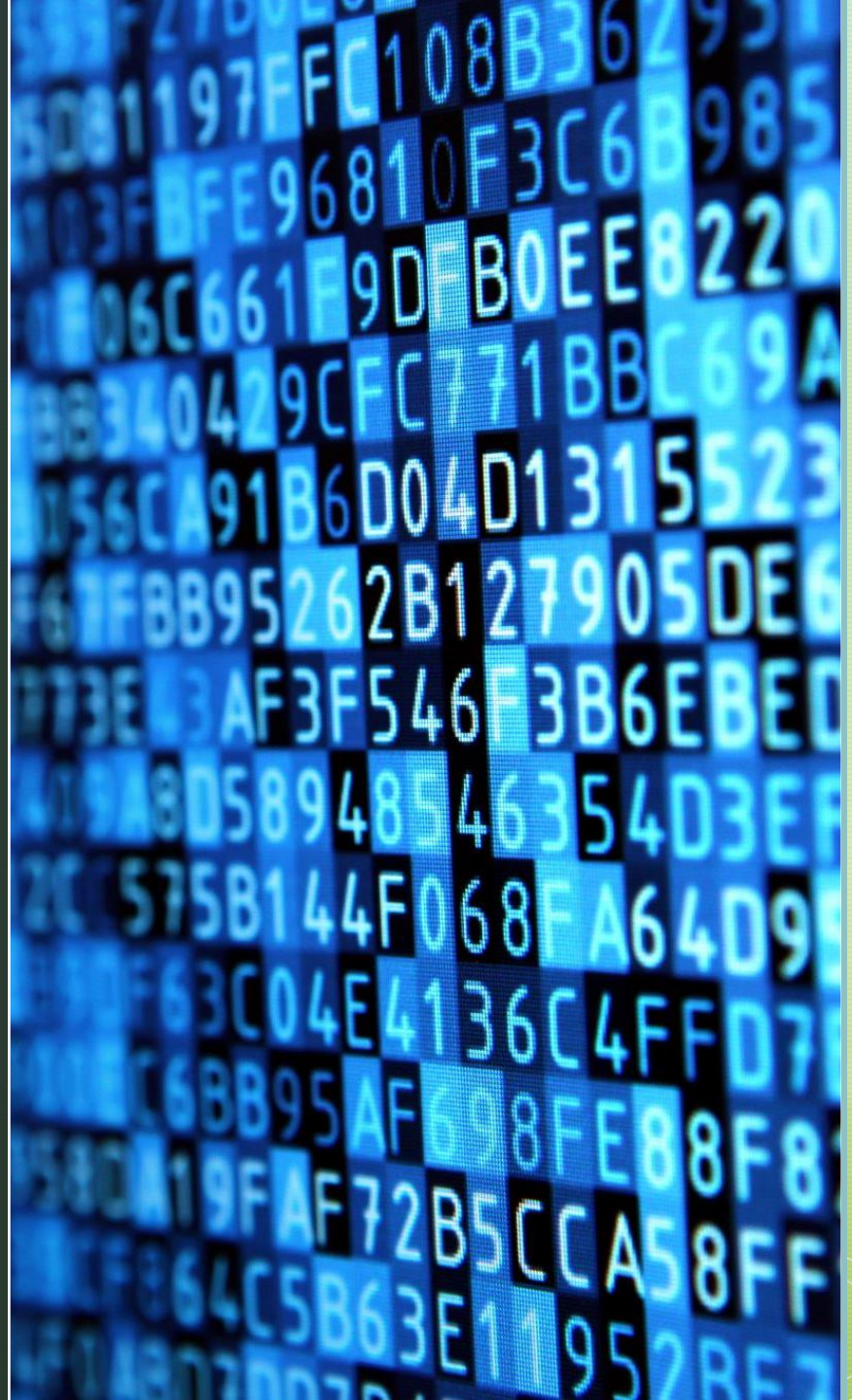
METHODOLOGY

Data Collection (API)

Request	Decode	Use	Combine	Filter	Replace	Export
Request data to SpaceX API and clean the requested data.	Decode the response content as a Json using .json() and tirn it into a pandas dataframe using .json_normalize() .	Use the API again to get information about the launches using the IDs given for each launch. Specifically, we will be using columns rocket, payloads, launchpad, and cores.	Combine the columns into a dictionary.	Filter data to only contain Falcon 9 launches.	Replace the missing values with the mean of the payload mass.	Export data to csv file.

Data Collection (Web Scrapping)

- Request data from Wikipedia(Falcon9 launch records).
- Create a BeautifulSoup object from the HTML response.
- Extract columns from HTML header table header.
- Create dataframe by parsing launch HTML tables.
- Export data to csv file.



Data Wrangling

- Import libraries(numpy , pandas) for EDA.
- Calculate :

  - No. of launches on each site.

  - No. and occurrence of each orbit.

  - no. and occurrence of mission outcome per orbit type.
- Create binary landing outcome label(fail = 0, success=1).
- Export to csv file.

- Landing Outcomes:

  True Ocean means the mission outcome was successfully landed to a specific region of the ocean .

  False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean.

  True RTLS means the mission outcome was successfully landed to a ground pad.

  False RTLS means the mission outcome was unsuccessfully landed to a ground pad.

  True ASDS means the mission outcome was successfully landed to a drone ship.

  False ASDS means the mission outcome was unsuccessfully landed to a drone ship.

  None ASDS and None None these represent a failure to land.

EDA with VISUALIZATION

Creating charts:

- Flight number VS Payload mass
- Flight number VS Launch site
- Payload mass VS Launch site
- Payload mass VS Orbit type

Analysis:

View relationship using scatter plots and show comparisons using bar plots among discrete categories.

EDA WITH SQL

1. Load the SQL extension and establish a connection with the database.
2. Analysis:
   - Unique launch sites in the space mission.
   - Launch sites that begins with ‘CCA’.
   - Average payload mass carried by booster version F9 v1.1
   - Total payload mass carried by boosters launched by NASA-CRS.
   - Date of first successful landing outcome in groudpad.
   - Number of successful and failed mission outcomes.
   - Booster versions having maximum payload mass.
   - Names of the boosters having success in drone ship and have payload mass greater than 4000 but less than 6000.
   - Month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015.
   - Count of landing outcome between the date 2010-06-04 and 2017-03-20.

MAP WITH FOLIUM

Markers indicating Launch Sites:

- Added a blue circle at NASA Johnson Space Center coordinate with a popup label showing its name using latitude and longitude.
- Added red circles to all the launch sites with a popup label showing their name.

Added coloured markers of green(successful) and red(fail), to be easily identify which launch sites have relatively high success rates.

Calculate and add coloured lines between a launch site and its proximities.

INTERACTIVE DASHBOARD WITH PLOTLY DASH

- Created an app layout with heading ‘SpaceX Launch Records Dashboard’.
- Added a dropdown list to enable all launch sites.
- Added a pie chart to show the total successful launches count for all sites,
- If a specific launch site was selected, show the Success vs. Failed counts for the site.
- Added a slider to select payload range.
- Added a scatter chart to show the correlation between payload and launch success.
- Add the callback functions and run the app.

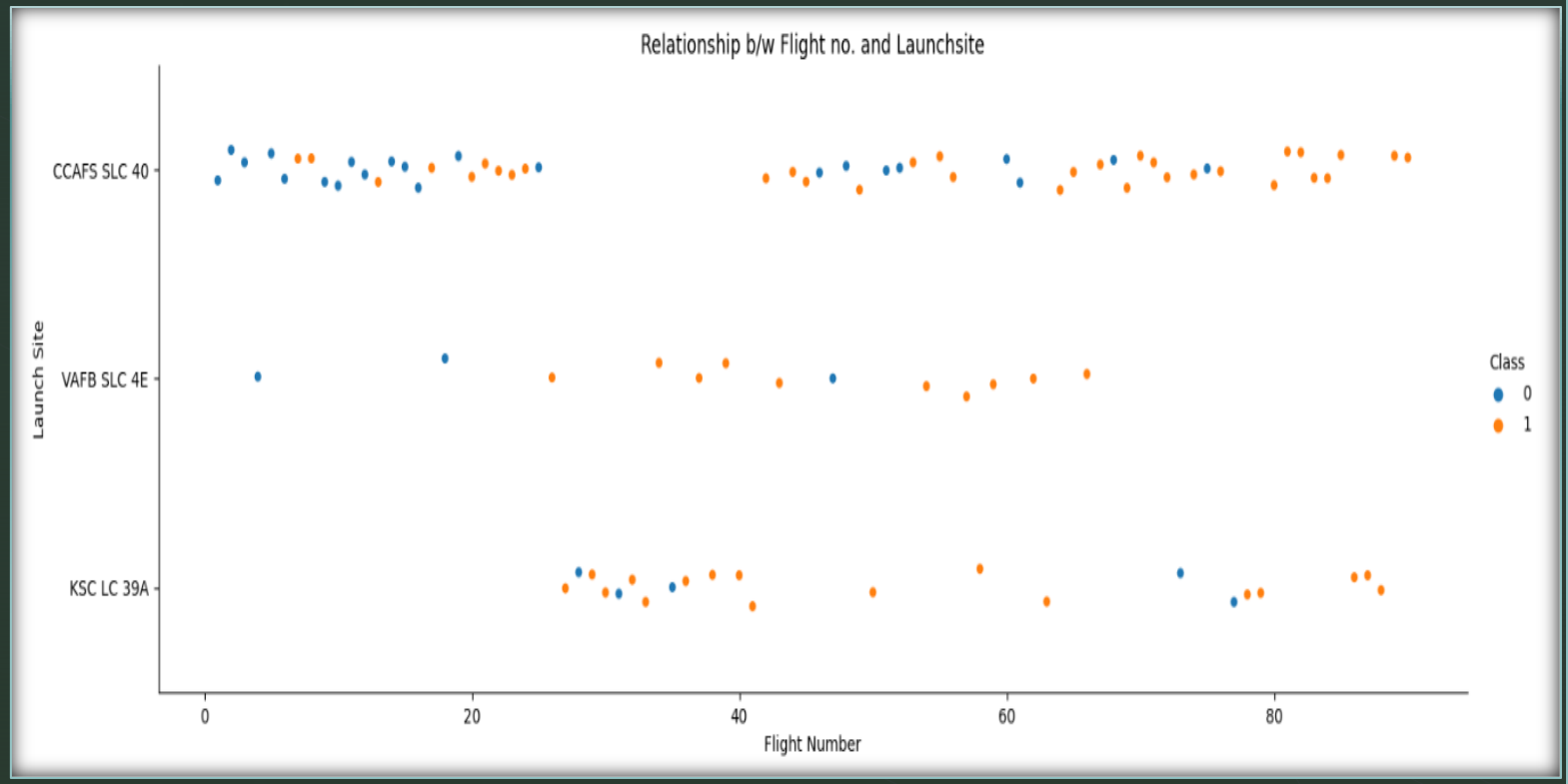
PREDICTIVE ANALYSTICS

Load	Load the data and create Numpy array from column Class.
Standardize	Standardize the data using StandardScalar and transform the data.
Split	Split the data using train_test_split.
Create	Create a GridSearchCV object with cv = 10 for parameter optimization and apply on different algorithms : • Logistic Regression, SVC,KNN, Decision tree.
Calculate	Calculate accuracy on the test data.
Plot	Plot the confusion matrix for all models.
Identify	Identify the best model.

RESULTS

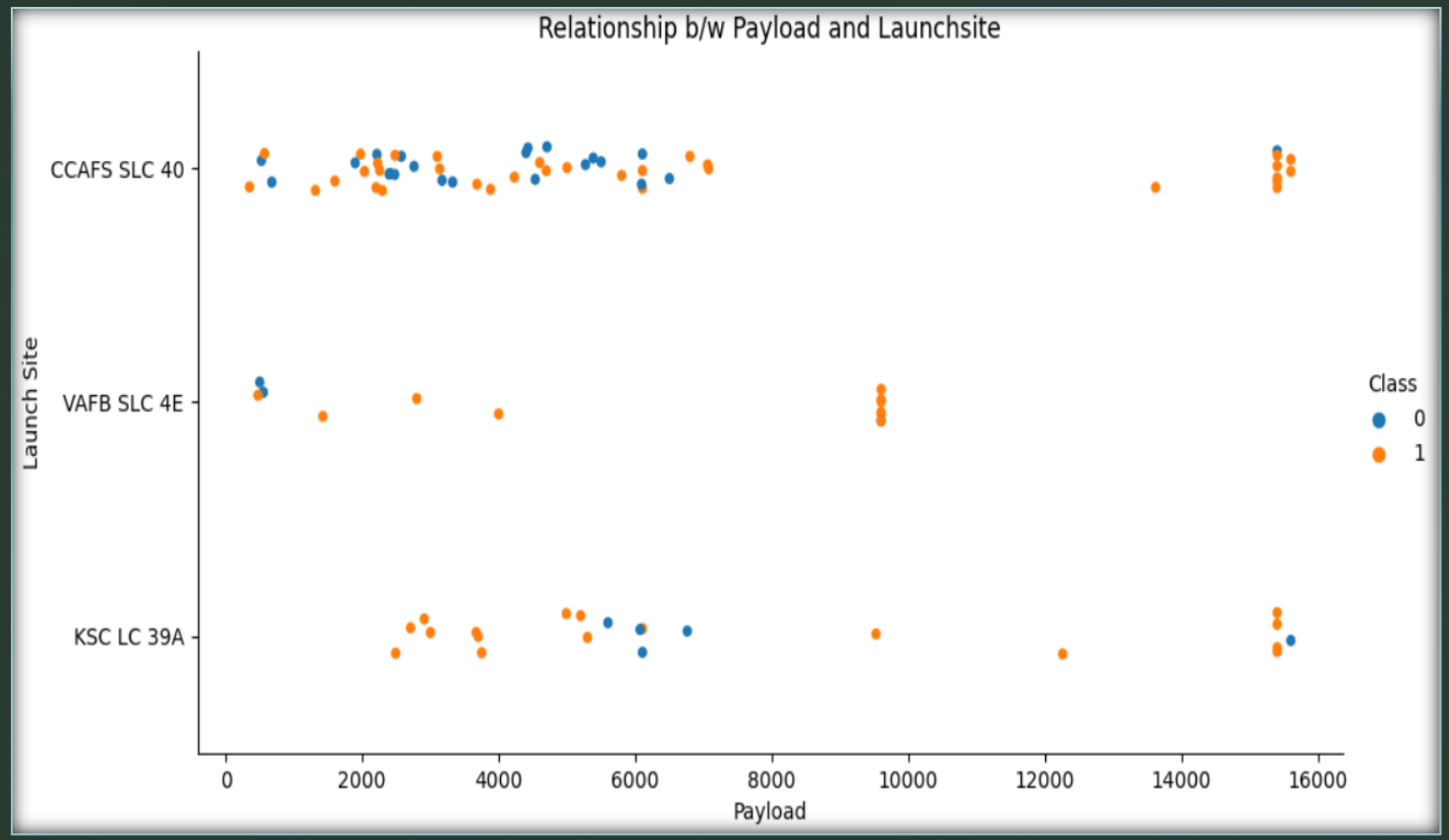
FLIGHT NUMBER VS LAUNCH SITE

- Majority of the launches are done by CCAFS SLC-40.
- VAFB SLC 4E and KSC LC 39A has a higher success rate.
- As the flight number is increasing, so as the success rates.
- Blue = ‘fail’, Orange = ‘Success’



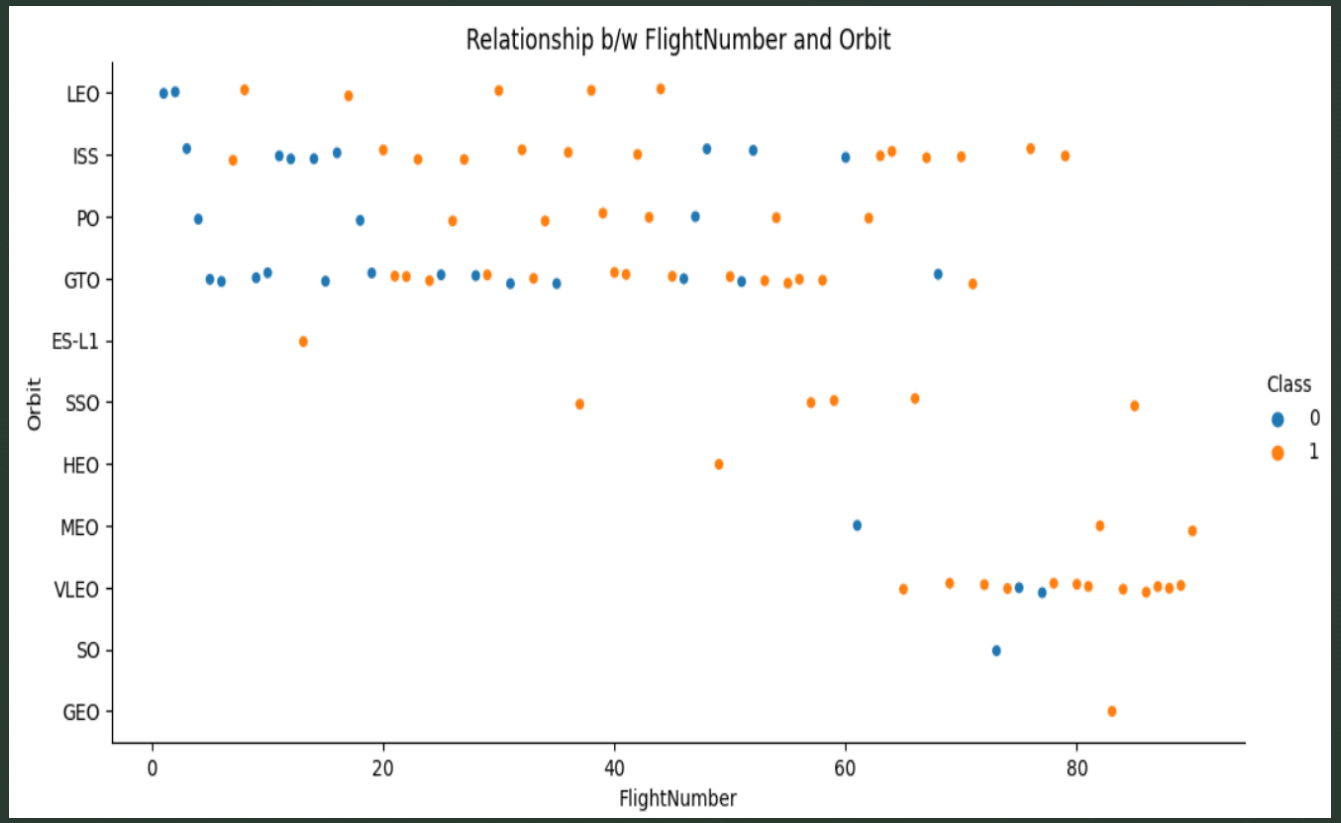
PAYLOAD VS LAUNCH SITE

- Higher the payload mass, higher is the success rates.
- KSC LC 39A has 100% success rate having payload up to 5000kg.
- VAFB SLC 4E hasn't launched anything having payload >10000kg.
- Launches having payload greater than approx. 7500kg were successful.



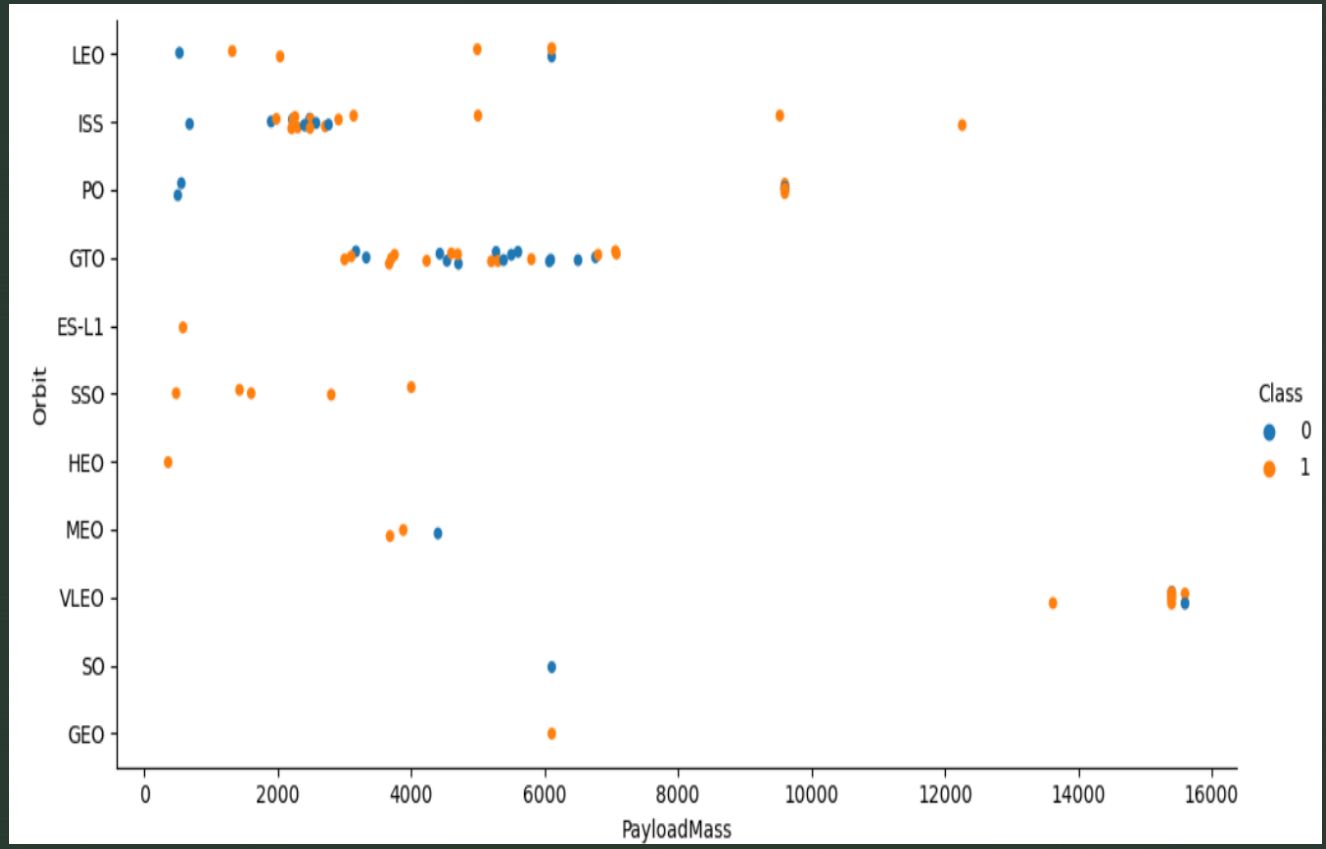
FLIGHT NUMBER AND ORBIT

- LEO orbit the Success appears related to the number of flights
- There seems to be no relationship between flight number when in GTO orbit.



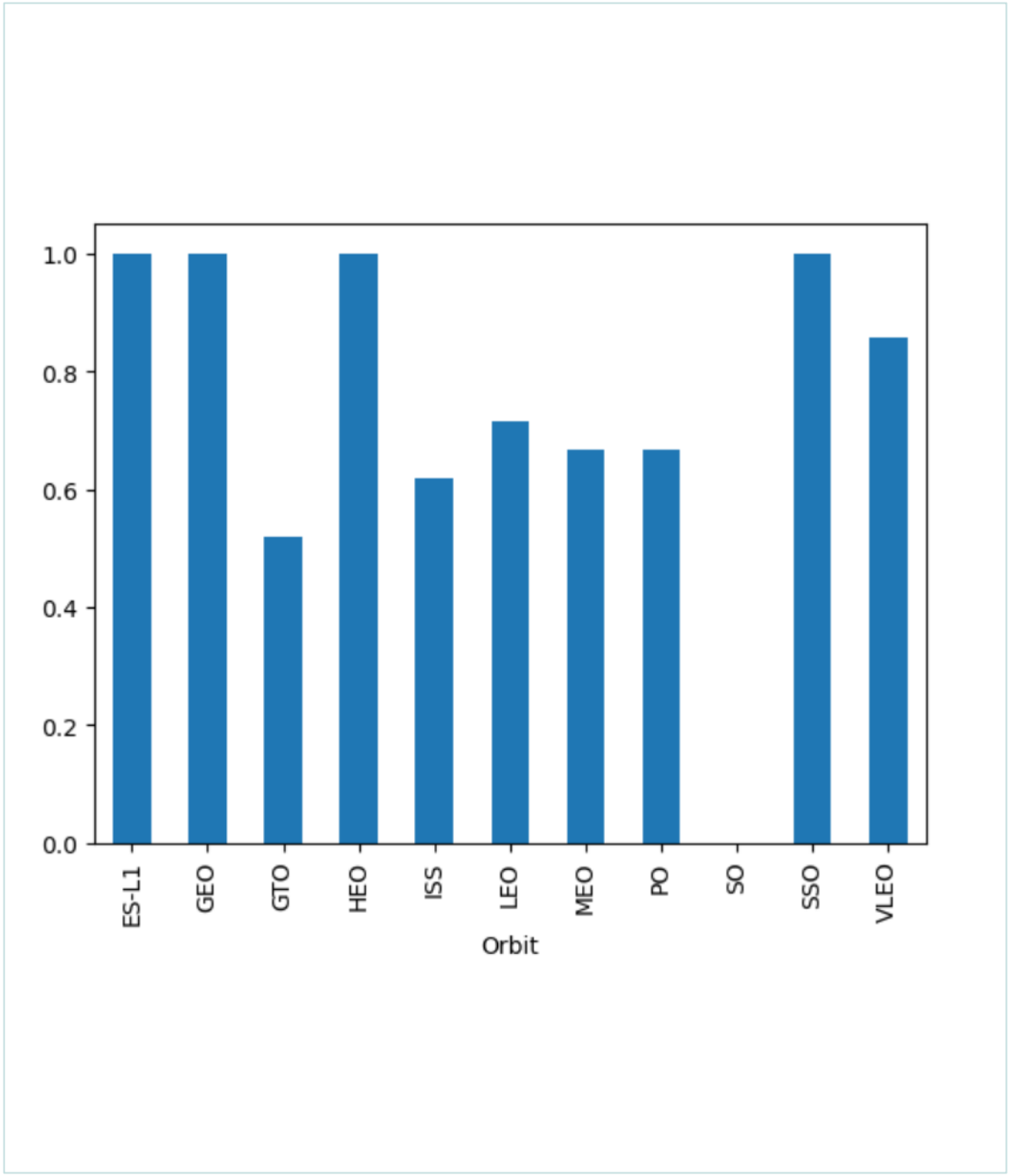
PAYLOAD VS ORBIT

- With heavy payloads the successful landing or positive landing rate are more for Polar , LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both mixed.



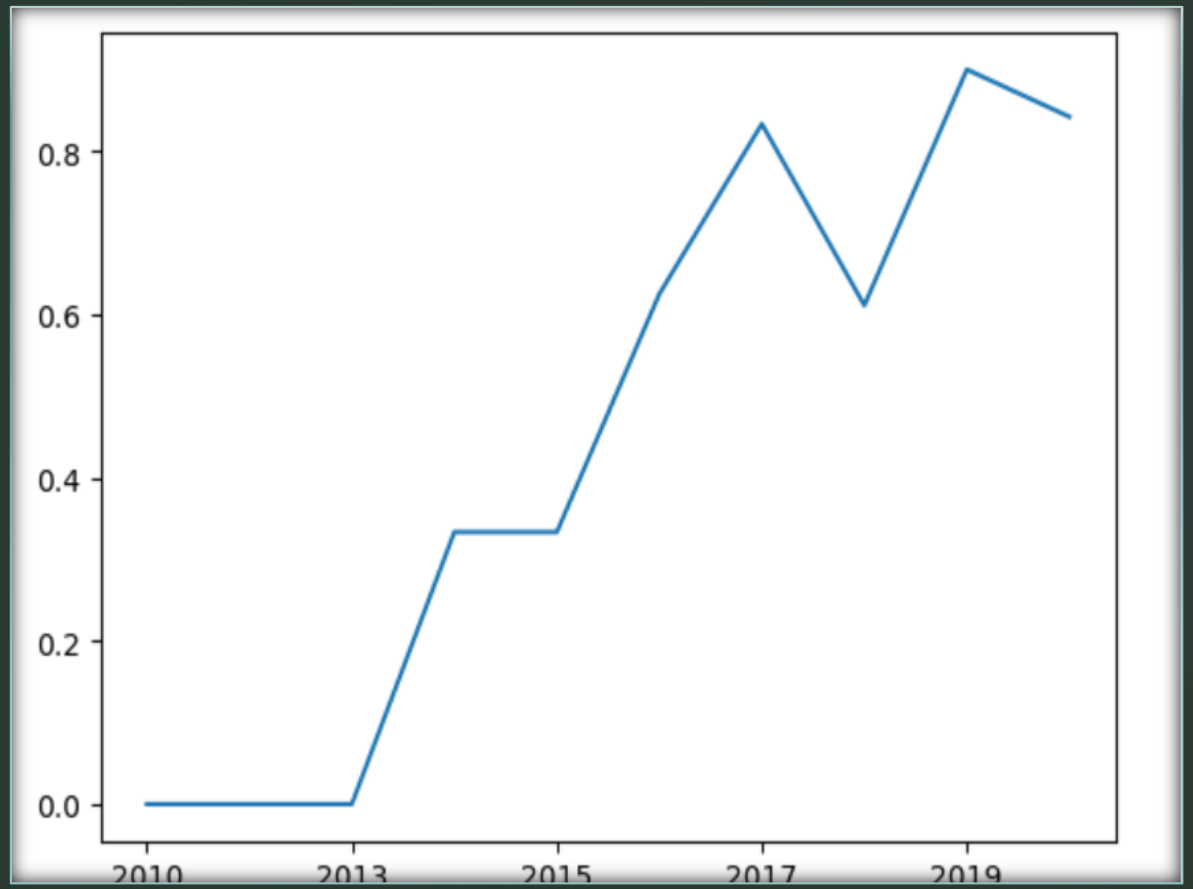
SUCCESS RATE BY ORBIT

- 100% success rate of ES-L1, GEO, HEO and SSO.
- Zero success rate for SO site.
- Other sites have moderate success rate.



LAUNCH SUCCESS OVER YEARS

- Success rate since 2013 has kept increasing till 2020.
- However , there is a little downfall in the success rate around 2018.



SQL EDA RESULTS

DIFFERENT LAUNCH SITES:

1. CCAFS LC-40
2. VAFB SLC-4E
3. KSC LC-39A
4. CCAFS SLC-40

Out[19]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

5 RECORDS WHERE LAUNCH SITES BEGIN WITH THE STRING 'CCA'

Out[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (paracl
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (paracl
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No atte
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No atte
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No atte

Out[9]:

SUM(PAYLOAD_MASS__KG_)
45596.0

Out[10]:

AVG(PAYLOAD_MASS__KG_)
2928.4

PAYLOAD MASS

Total payload mass carried by boosters launch by NASA (CRS).

Average payload mass carried by booster version F9 v1.1

SUCCESSFUL LANDING WITH PAYLOAD B/W 4000 AND 6000KG

4]:

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

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

Out[15]:

Mission_Outcome	COUNT(*)
None	898
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- 99 Success
- 1 failure (in flight)
- 1 Success (payload status unclear)

Booster Versions with Max. Payload Mass

- We noticed that the maximum load was carried mostly by booster version starting with F9 B5.

Out[61]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600.0
F9 B5 B1049.4	15600.0
F9 B5 B1051.3	15600.0
F9 B5 B1056.4	15600.0
F9 B5 B1048.5	15600.0
F9 B5 B1051.4	15600.0
F9 B5 B1049.5	15600.0
F9 B5 B1060.2	15600.0
F9 B5 B1058.3	15600.0
F9 B5 B1051.6	15600.0
F9 B5 B1060.3	15600.0
F9 B5 B1049.7	15600.0
F9 B5 B1049.6	15440.0
F9 B5 B1059.3	15410.0
F9 B5 B1051.5	14932.0
F9 B5 B1049.3	13620.0

List of Failure Landing Outcomes Records in 2015

Done.

Out[16]:

month_names	Booster_Version	Launch_Site	Landing_Outcome
10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- There were two failure landing outcomes in the year 2015 in the month of April and October.
- And both the outcomes are from CCAFS LC-40 launch site.

LANDING OUTCOMES B/W 2010-06-04 AND 2017-03-20

Out[19]:

Landing_Outcome	No. of landing outcomes
Success (ground pad)	7
Success (drone ship)	8
Success	20
No attempt	1
No attempt	10
Failure (parachute)	2
Failure (drone ship)	3
Failure	3
Controlled (ocean)	2

Done.

Out[13]:

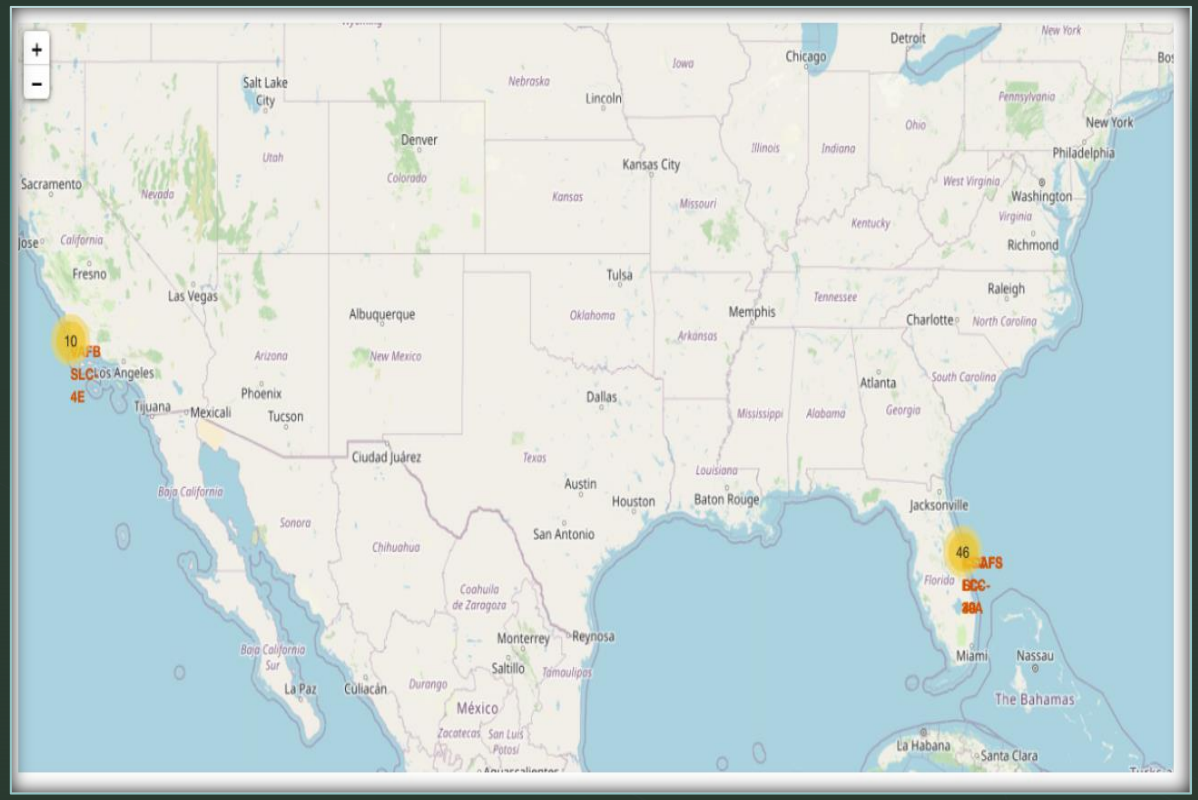
MIN(DATE)
01/08/2018

FIRST SUCCESSFUL LANDING ON GROUND PAD

LAUNCH SITES ANALYSIS WITH FOLIUM

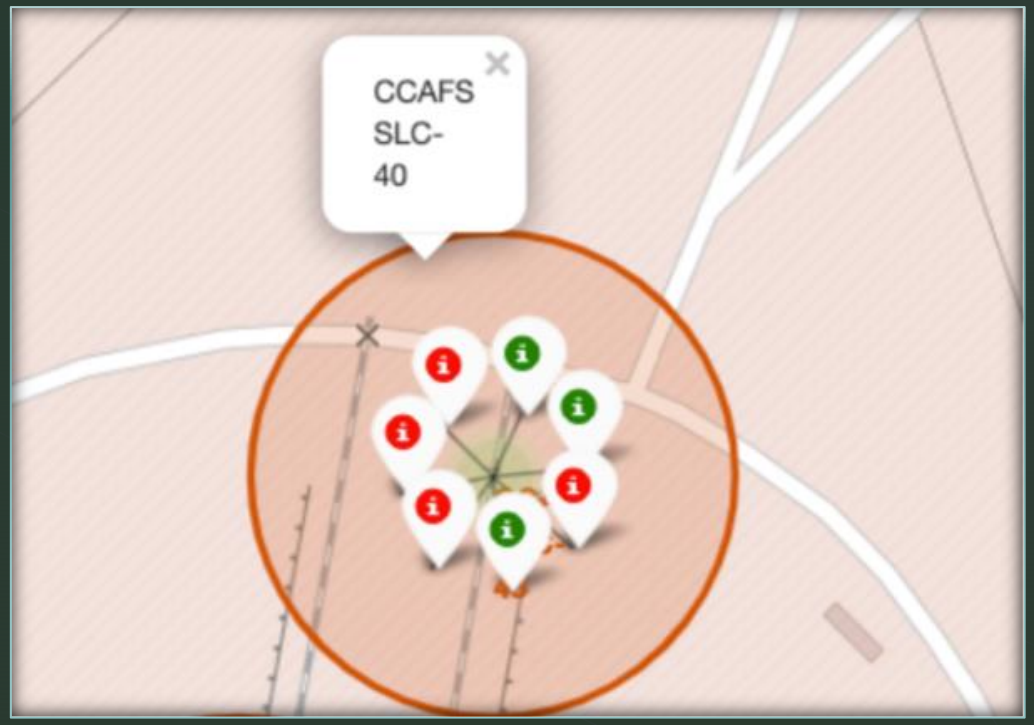
LAUNCH SITES LOCATION

- All the launch sites are near coastal area, which is good and not far from roads and railway network as well.
- Near to the equator, easier to launch to equatorial orbit.



LAUNCH OUTCOMES

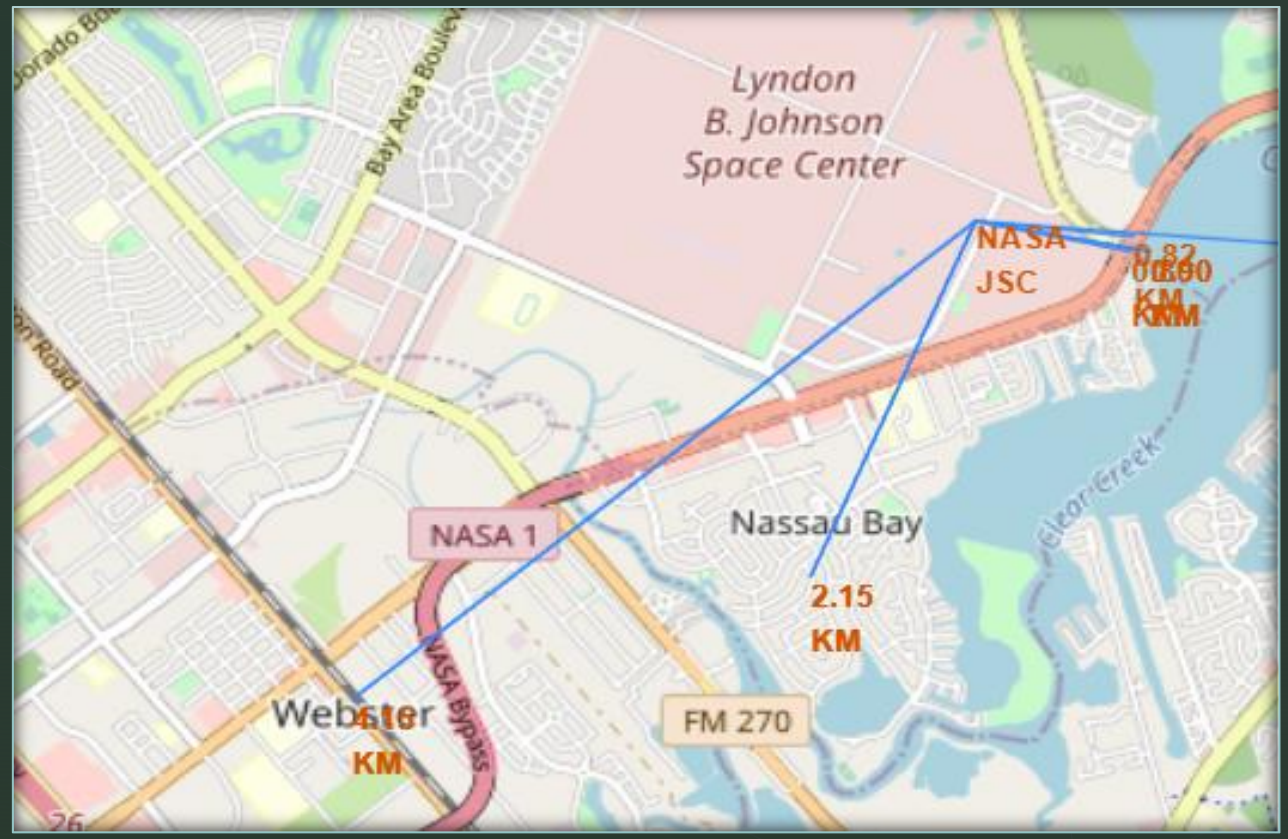
- GREEN pop-ups are the successful launches while the RED ones are the unsuccessful launches.
- Launch site CCAFS SLC-40 has below average successful outcomes.



DISTANCE TO PROXIMITIES

Distance b/w the launch site and the proximities:

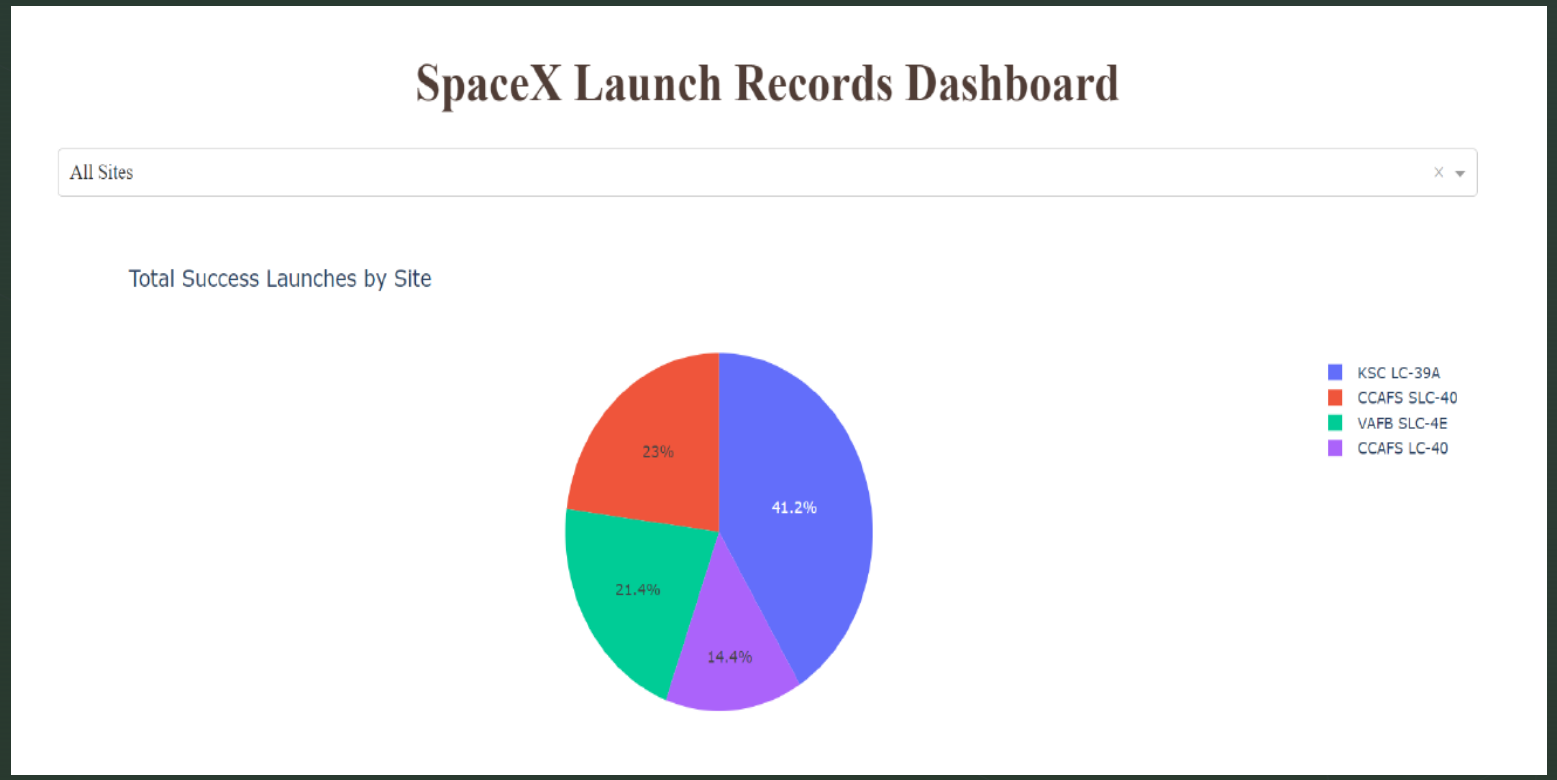
- Coast : 0.82 KM
- Railway : 4.15 KM
- Highway : 8.15 KM
- City : 2.15 KM



INTERACTIVE DASHBOARD WITH PLOTLY DASH

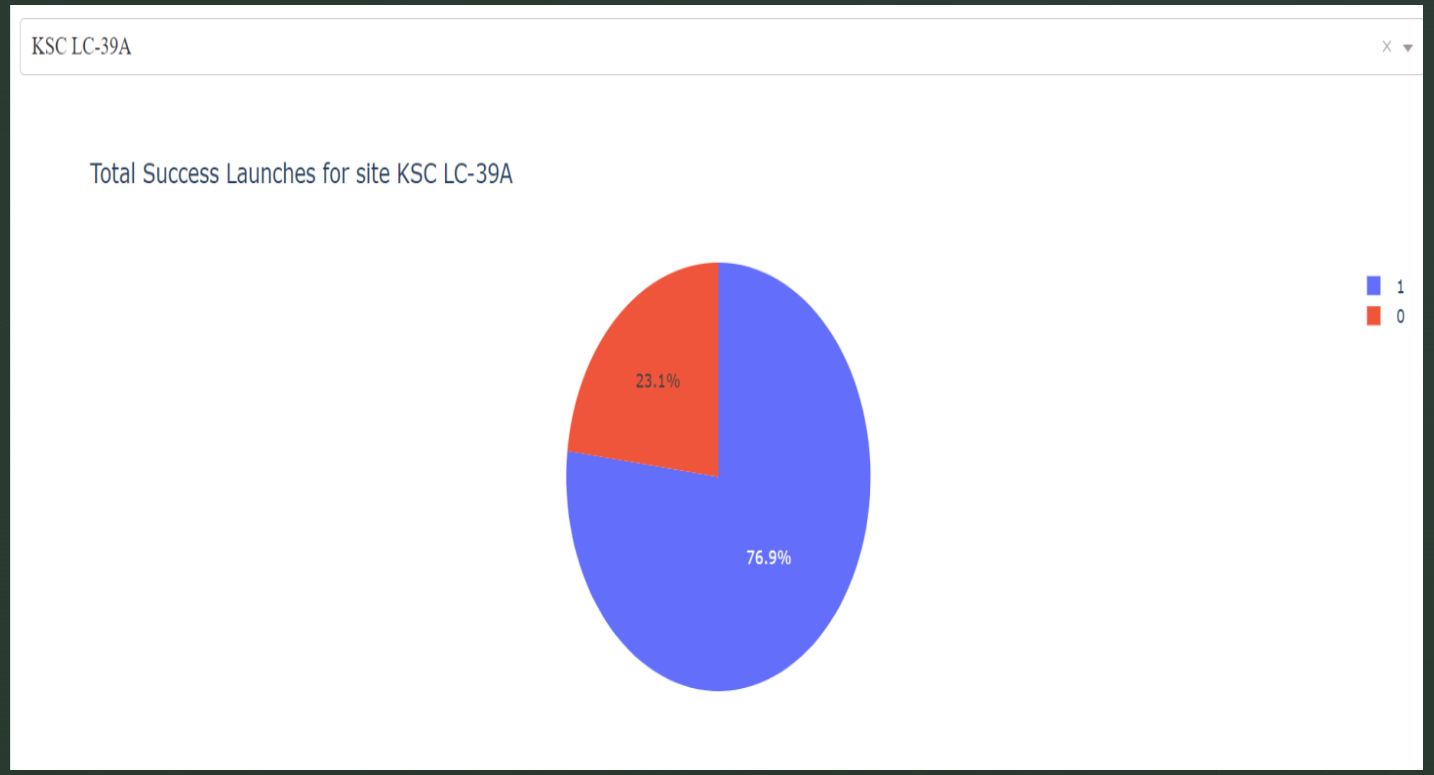
LAUNCH SUCCESS BY SITE

- We can notice that the most successful launch site is KSC LC-39A with 41.2%.



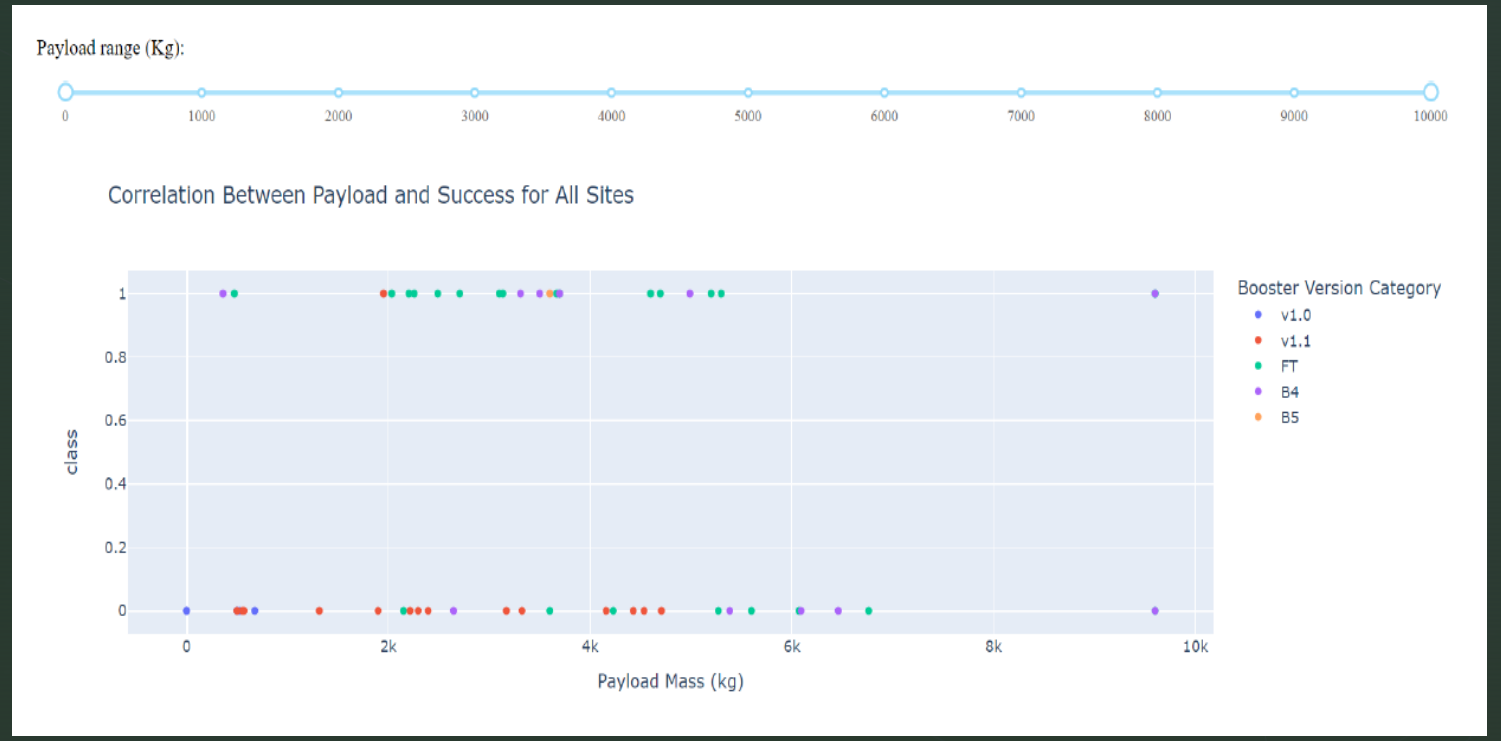
LAUNCH SUCCESS OF KSC LC 39A

- KSC LC -39A has the highest success rate with almost 77%.

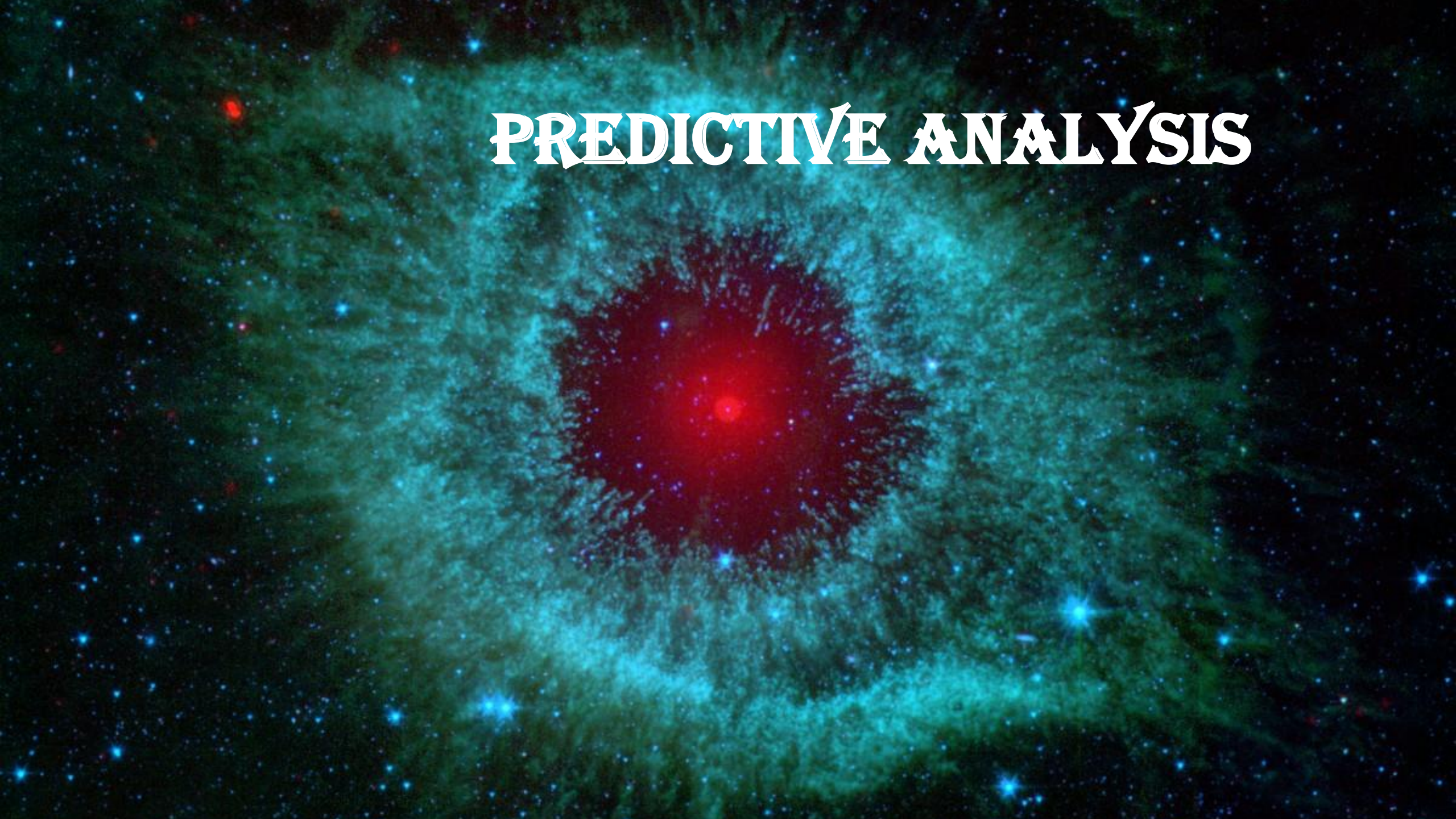


Payload Mass and Success with Scatter Chart and Slider

- We observed that payloads b/w 2000 kg and 6000 kg has the highest success rate.



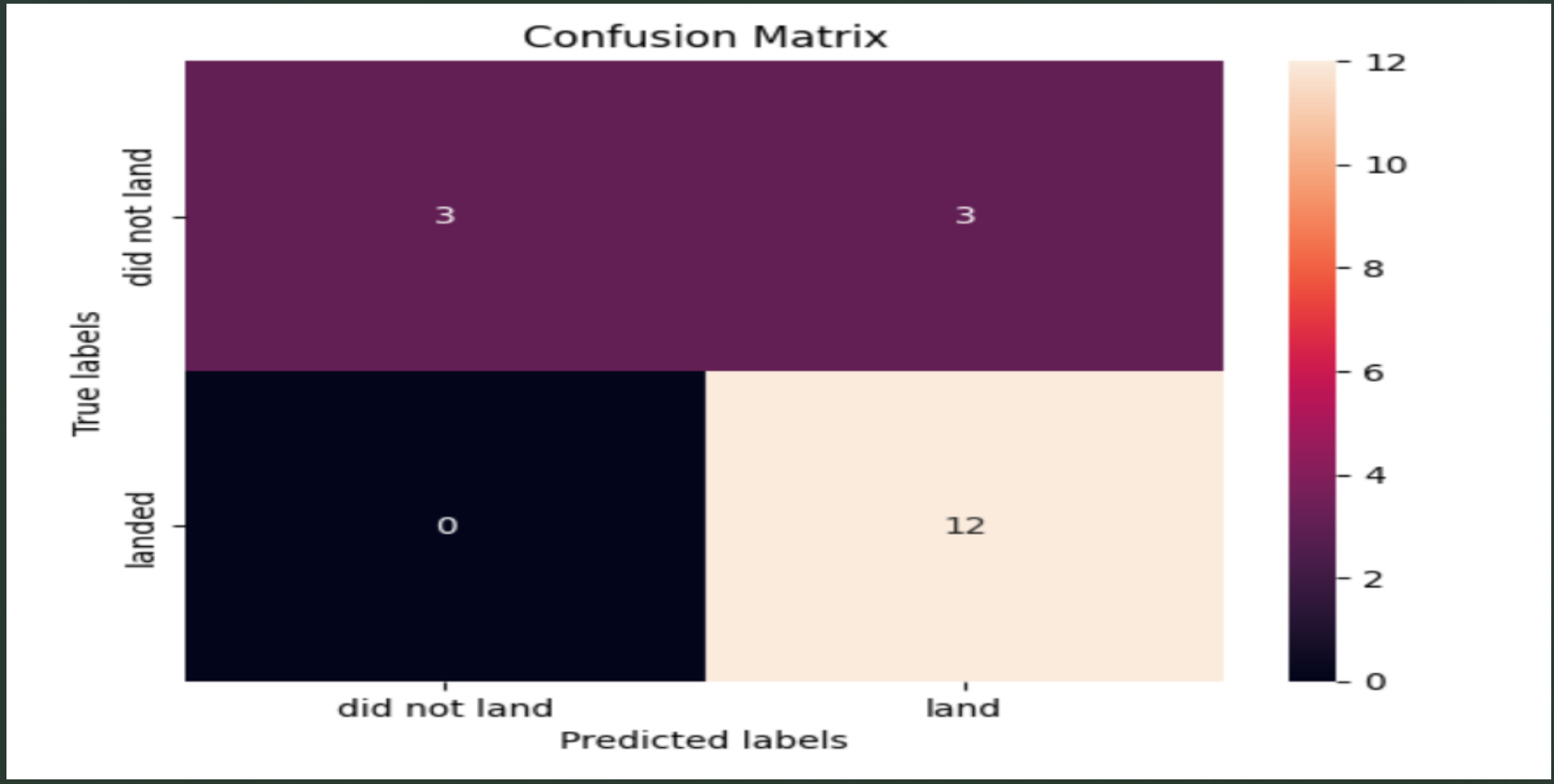
PREDICTIVE ANALYSIS



```
Best model is DecisionTree with a score of 0.9
Best params is : {'criterion': 'gini', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2,
'splitter': 'random'}
```

BEST PREDICTIVE MODEL

CONFUSION MATRIX OF DECISION TREE MODEL



OUTCOMES FROM PREDICTIVE ANALYSIS

1. The best model among SVC , Decision tree model, KNN, Logistic Regression is the **Decision Tree Model** with accuracy of **0.9**
2. Confusion matrix of Decision tree proves its accuracy as the true positives and true negatives are in big numbers

CONCLUSION

- The best launch site is KSC LC-39A.
- Launches above 7,000kg are more successful.
- Overtime the successful landings have improved, stating that in future unsuccessful outcomes can be minimized.
- For the prediction, Decision Tree happens to be more suitable to predict the landings more accurately and provides benefit to the company.
- All the launch sites are near coast lines.
- ES-L1, GEO, HEO, and SSO have a 100% success rate
- KSC LC-39A has the highest success rate.

GITHUB REPOSITORY WHICH INCLUDES ALL THE COMPLETED NOTEBOOKS AND FILES

- https://github.com/svashisth-git/testrepo/tree/main

- **Repository name : testrepo**


United States
Atlantis