'''FALCON9 LANDING OUTCOME USING DATA SCIENCE'''

Naman Kumar

1st July 2022

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Summary of Methodologies & Tools:

- Data Collection:
  - Using APIs
  - Using Web Scrapping
- Data Wrangling & Pre-processing
- Exploratory Data Analysis:
  - Using SQL
  - Using Data Visualization
- Interactive Visual Analytics:
  - Using Folium
  - Dashboard using Dash plotly
- Machine Learning Prediction:
  - Predictive Analysis(Classification)

Summary of Results:

- Exploratory Data Analysis
- Interactive Analytics
- Predictive Analysis

INTRODUCTION

- **Project Context and Background**
  - 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.

- **Problems to answer**
  - Predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully.
  - What factors and conditions influence the probability of the rocket landing successfully?

METHODOLOGY

Our adopted methodology goes in the following manner:

- We took a business approach and first understood the problem at hand
- Data was collected using SpaceX API and web scraping
- Perform data wrangling and other pre-processing techniques to data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- We try to keep every step iterative and flexible to ensure feedback and improvement in model tuning.
- Find the best hyper parameters and serve the final model with them

DATA COLLECTION USING SpaceX API

- Using get from requests function, we collected the data using SpaceX API.

Now let's start requesting rocket launch data from SpaceX API with the following URL:

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

- We converted the collected response data to a normalized .json file.

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

- Then we stored the data in DataFrame.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None

- [Jupyter Notebook]()

DATA COLLECTION USING Web Scrapping

- We used BeautifulSoup function from bs4 module to web scrap HTML page.

```
from bs4 import BeautifulSoup
```

- We parsed the table and converted it into a pandas dataframe using soup constructor

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.content, "html.parser")
```

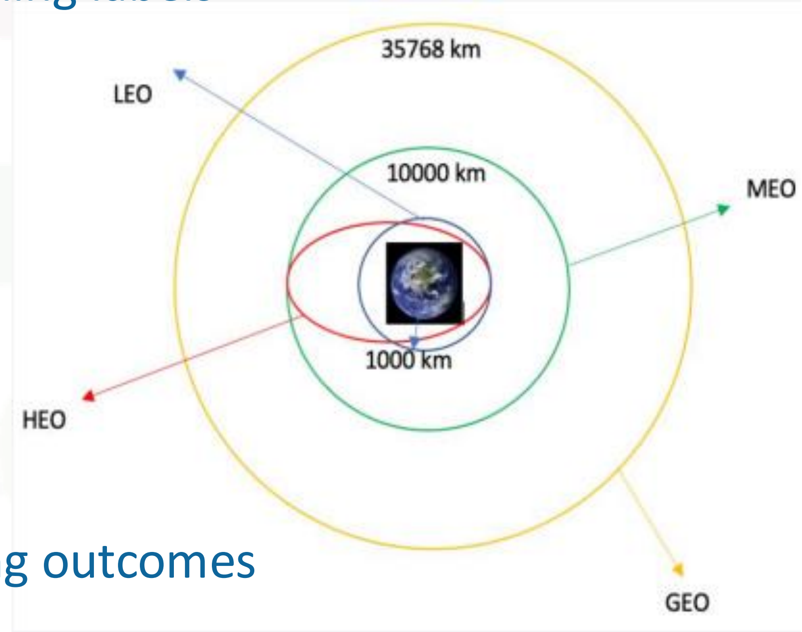
- Then we stored the data in DataFrame.

```
df=pd.DataFrame(launch_dict)
df
```

- [Jupyter Notebook]()

DATA WRANGLING

- We performed Exploratory Data Analysis to determine training labels
  - Pandas
  - Numpy
  - Matplotlib



- We found number of launches per site

CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

- We also created to create a landing outcome class for landing outcomes

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
```

- Jupyter Notebook

EDA WITH SQL

- Using db2 and python connectivity we performed sql queries on our database remotly

```
!pip install sqlalchemy==1.3.9
!pip install ibm_db_sa
!pip install ipython-sql
```

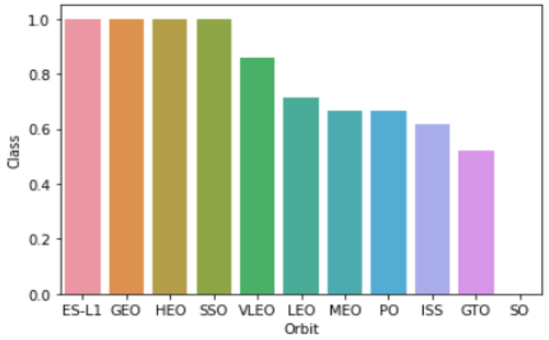
- Through these queries we used EDA to understand the data better and find new insights and relations

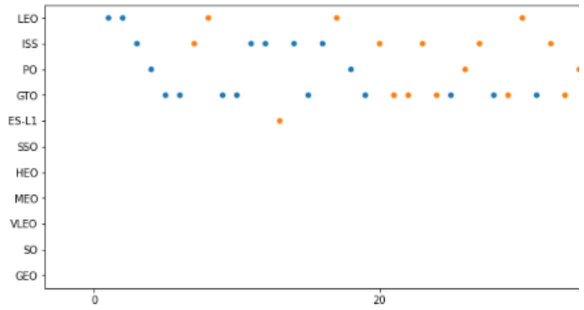
  - The names of unique launch sites in the space mission.
  - The total payload mass carried by boosters launched by NASA (CRS)
  - The average payload mass carried by booster version F9 v1.1
  - The total number of successful and failure mission outcomes
  - The failed landing outcomes in drone ship, their booster version and launch site names.

- [Jupyter Notebook]()

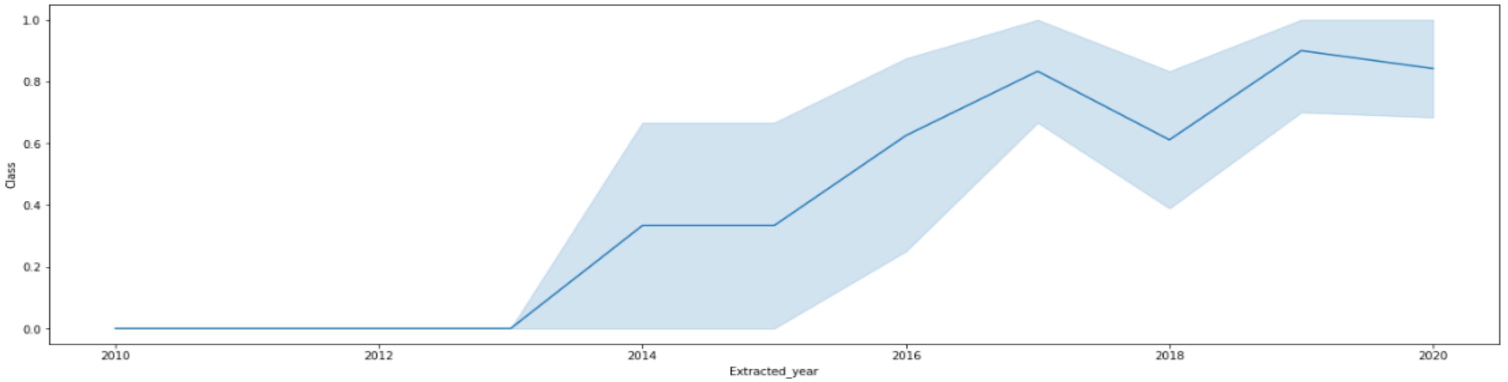
EDA WITH DATA VISUALIZATION

- Here, we Perform exploratory Data Analysis and Feature Engineering using Pandas and Matplotlib
  - Exploratory Data Analysis
  - Preparing Data Feature Engineering





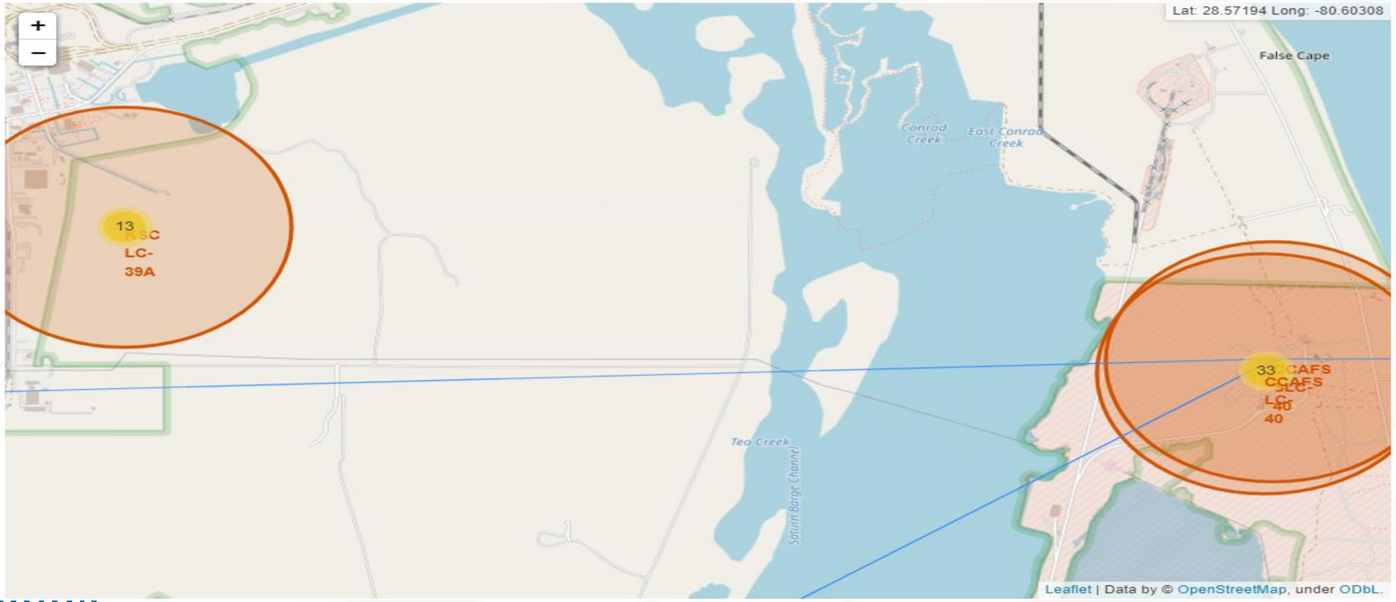
- Through visualizing many different feature of our data, we could unravel new relations among various attribute of our data to develop a yearly success data trend.



- [Jupyter Notebook]

INTERACTIVE ANAYLITCS USING FOLIUM

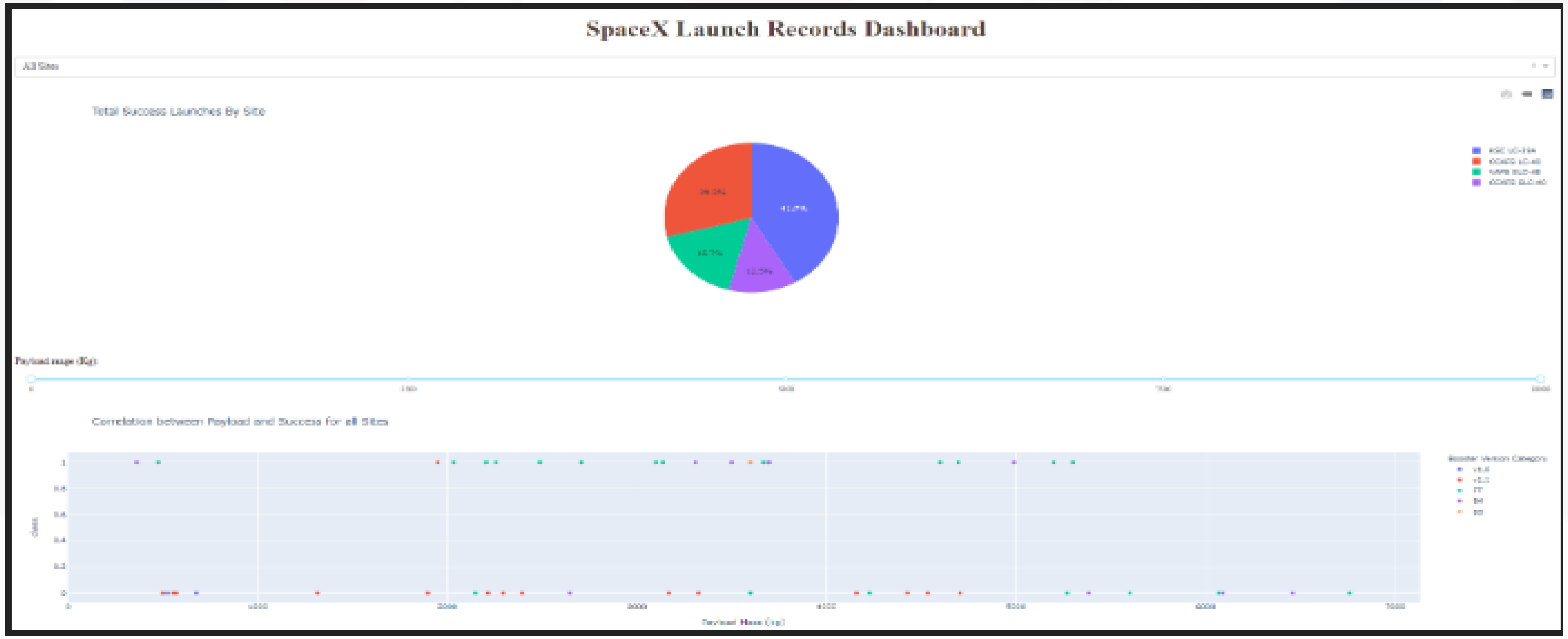
- Using the Geo Visualization library Folium, we created an interactive map that displays
  - Mark all launch sites
  - The success/failed launches for each site
  - Calculate the distances between a launch site to its proximities



- Jupyter Notebook

INTERACTIVE DASHBOARD USING PLOTLY DASH

- Using Dash from Plotly, we built a reactive web based Dashboard
- We plotted pie charts showing the total launches by a certain sites
- We plotted scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version.
- We even built a input slider for payload mass(Kg)



- .py file

IBM Developer

SKILLS NETWORK

RESULTS AND INSIGHTS

- Found from EDA and Analytics
  - SQL
  - Data Visualization
  - Geo Visualization Folium
  - Dashboard

RESULTS & INSIGHTS FROM EDA & ANALYTICS

Here, we Ranked 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

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

RESULTS & INSIGHTS FROM EDA & ANALYTICS

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

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

RESULTS & INSIGHTS FROM EDA & ANALYTICS

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

List of the names of the booster_versions which have carried the maximum payload mass, using a sub query

RESULTS & INSIGHTS FROM EDA & ANALYTICS

Listing the total number of successful and failure mission outcomes

```
Total number of successful mission outcomes +------+
|  1   |
+------+
| 100 |
+------+
Total number of failure mission outcomes +---+
| 1 |
+---+
| 1 |
+---+
```

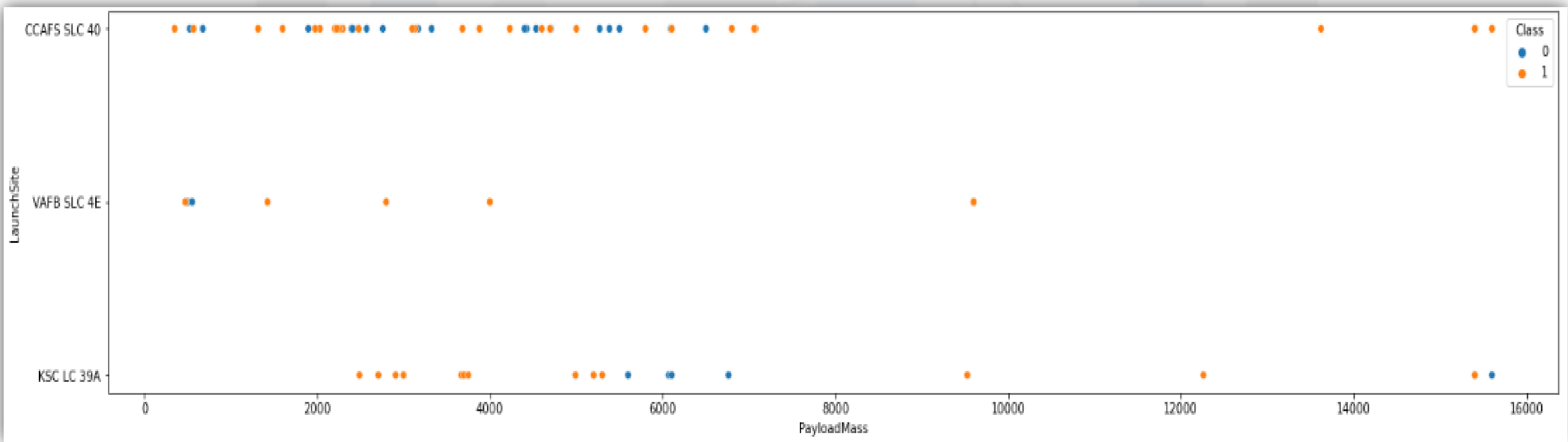
RESULTS & INSIGHTS FROM EDA & ANALYTICS

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

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

RESULTS & INSIGHTS FROM EDA & ANALYTICS

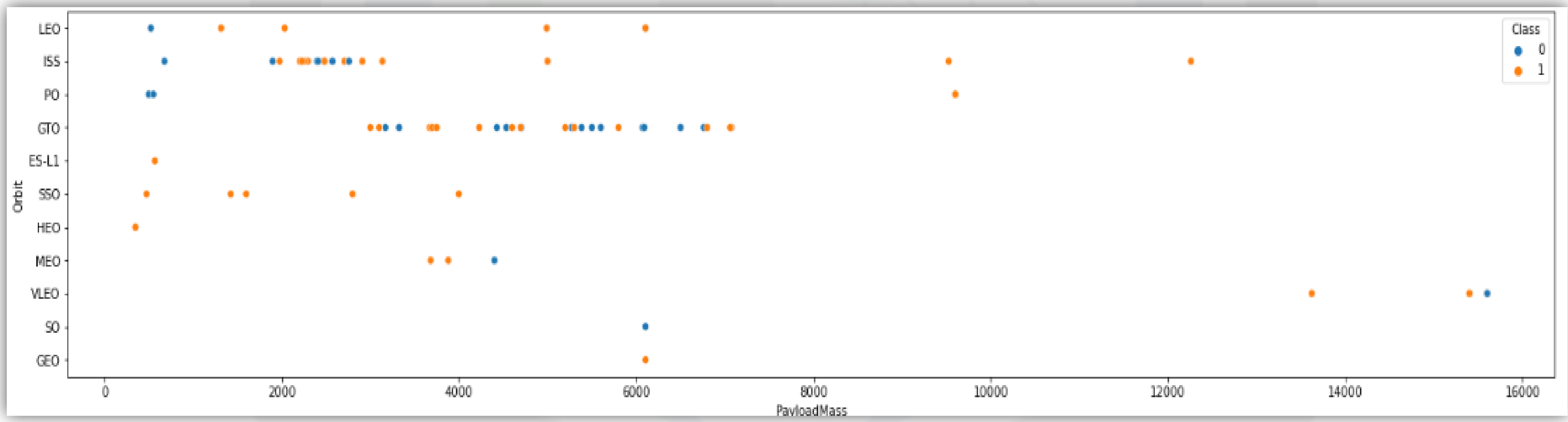
In Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).



RESULTS & INSIGHTS FROM EDA & ANALYTICS

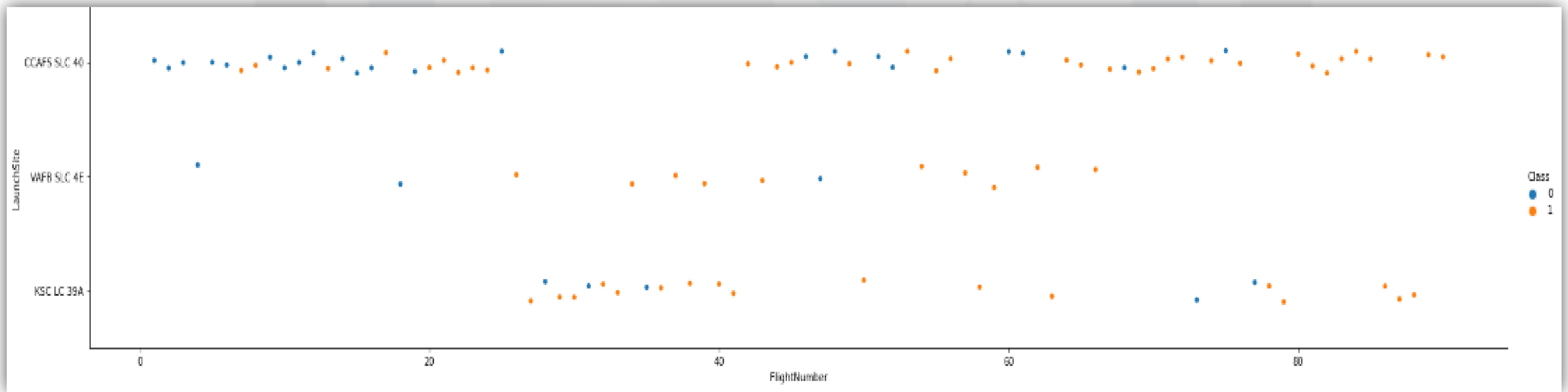
With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



RESULTS & INSIGHTS FROM EDA & ANALYTICS

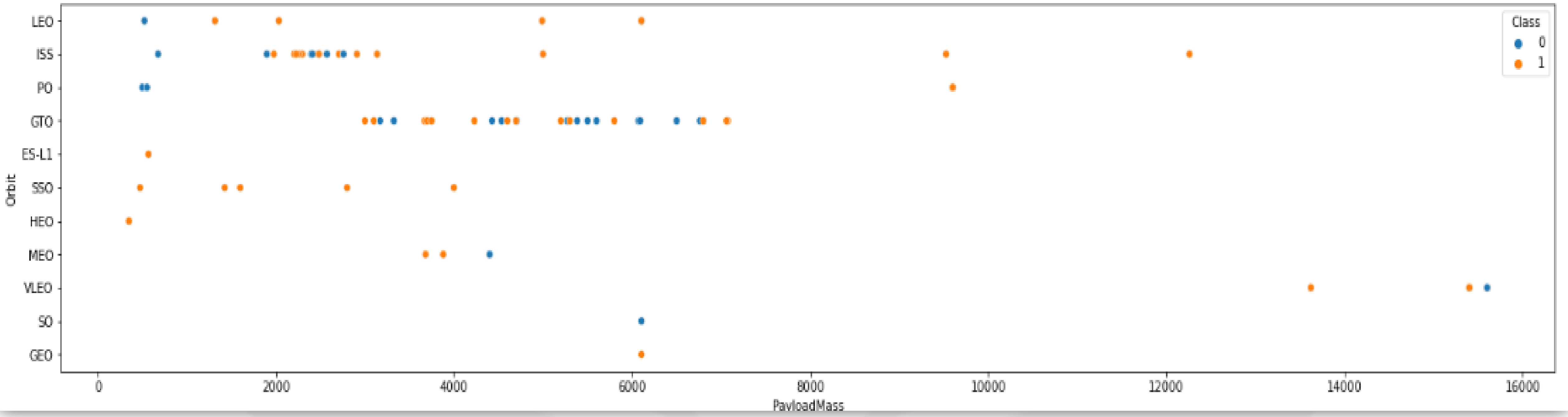
From the below plot, we can see that larger the flight amount at a launch site, the greater the success rate at a launch site.



RESULTS & INSIGHTS FROM EDA & ANALYTICS

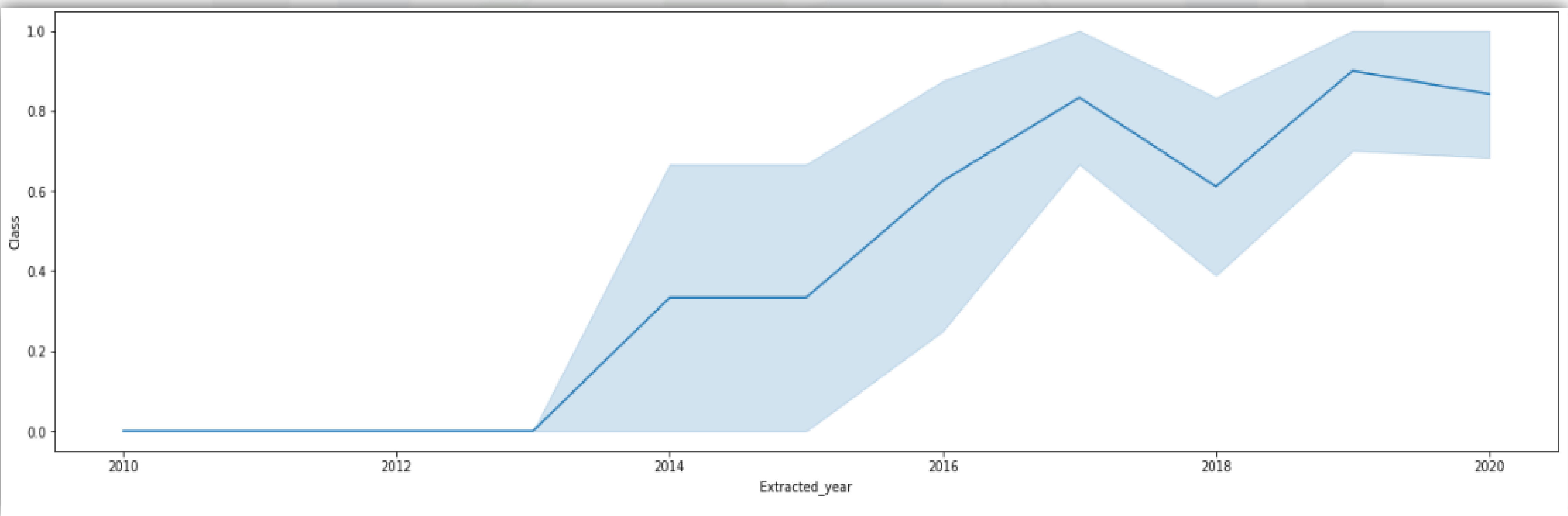
With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



RESULTS & INSIGHTS FROM EDA & ANALYTICS

Below line chart clearly indicates, The Number of flights kept on increasing from 2013 onwards.

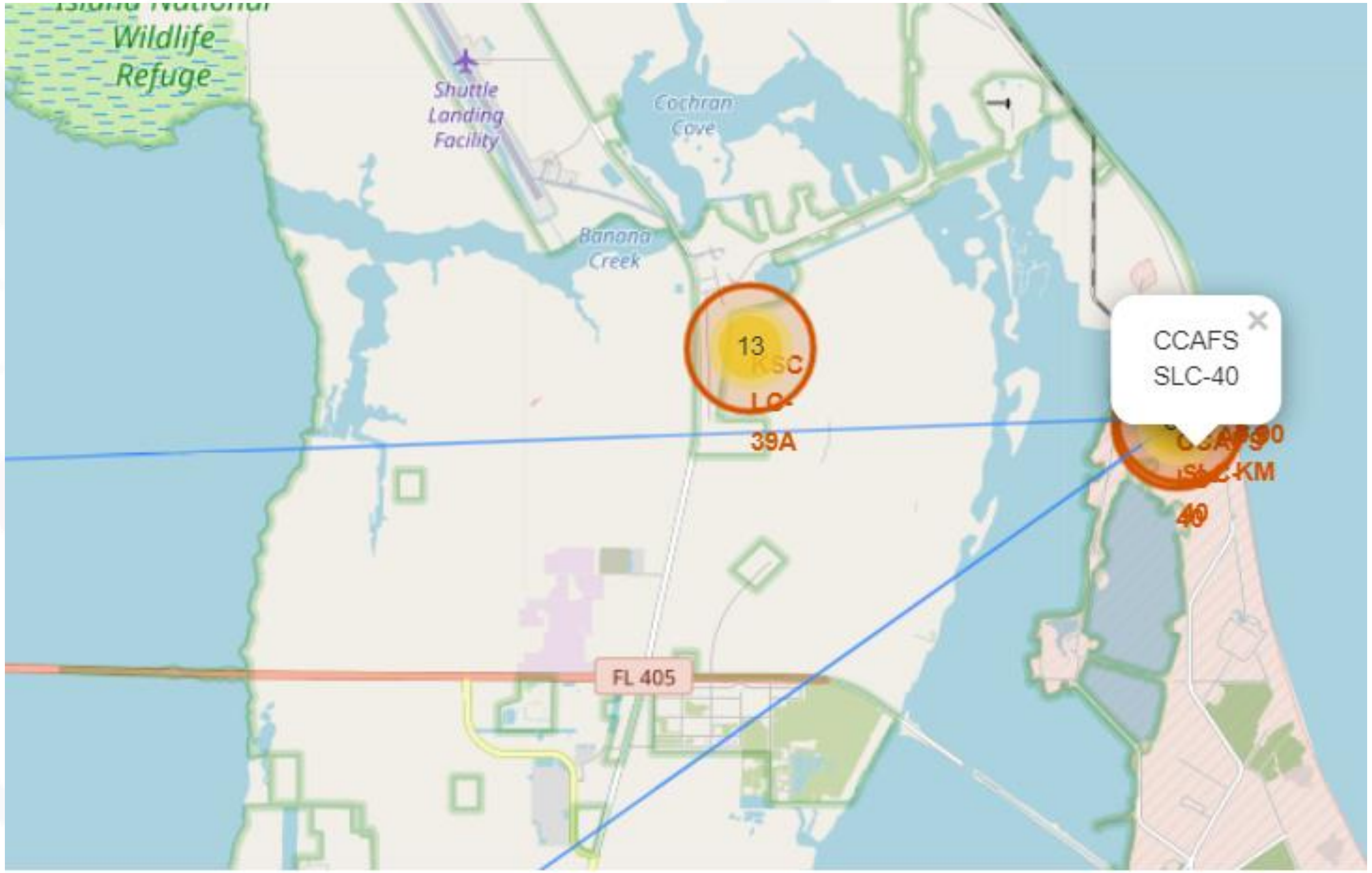


RESULTS & INSIGHTS FROM EDA & ANALYTICS

SpaceX launch sites are in USA coastline. California & Florida

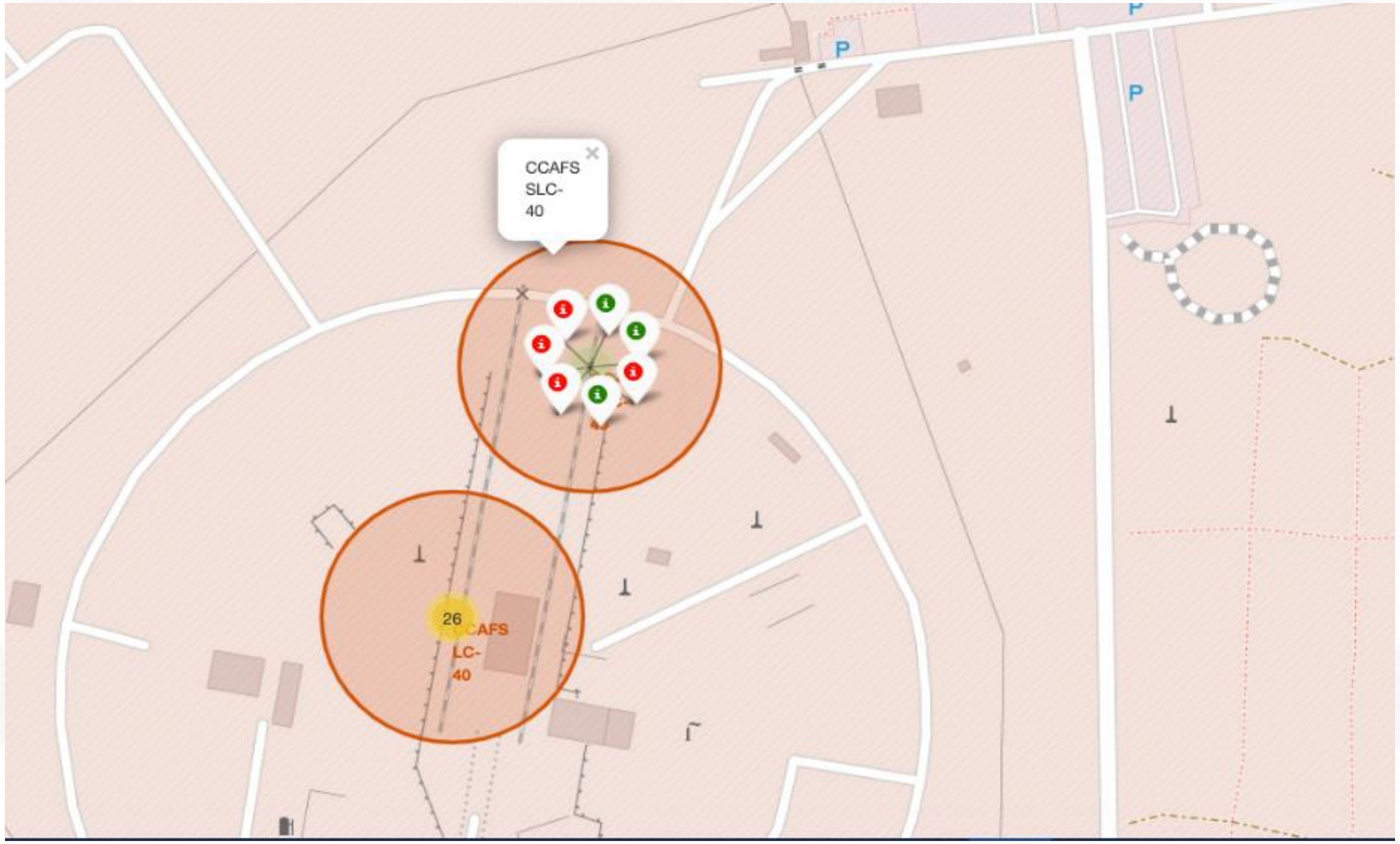
RESULTS & INSIGHTS FROM EDA & ANALYTICS

- Are launch sites in close proximity to railways? No
- Are launch sites in close proximity to highways? No
- Are launch sites in close proximity to coastline? Yes
- Do launch sites keep certain distance away from cities? Yes



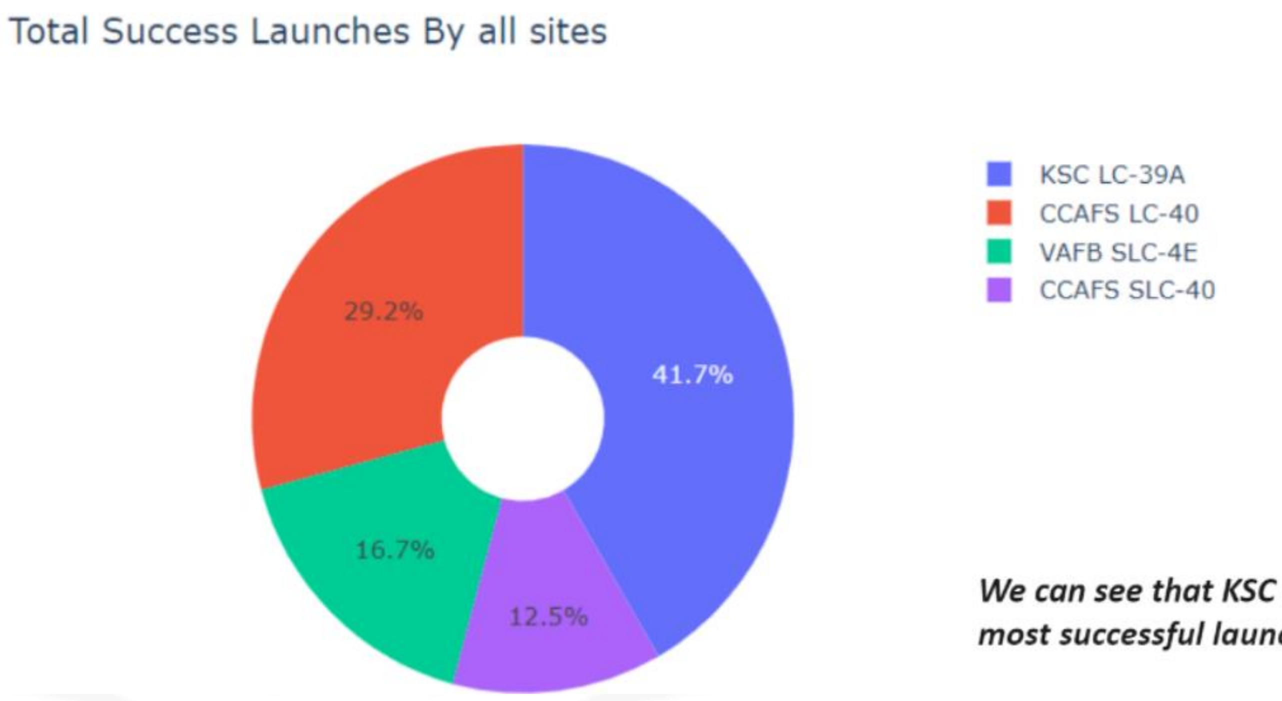
RESULTS & INSIGHTS FROM EDA & ANALYTICS

From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates.



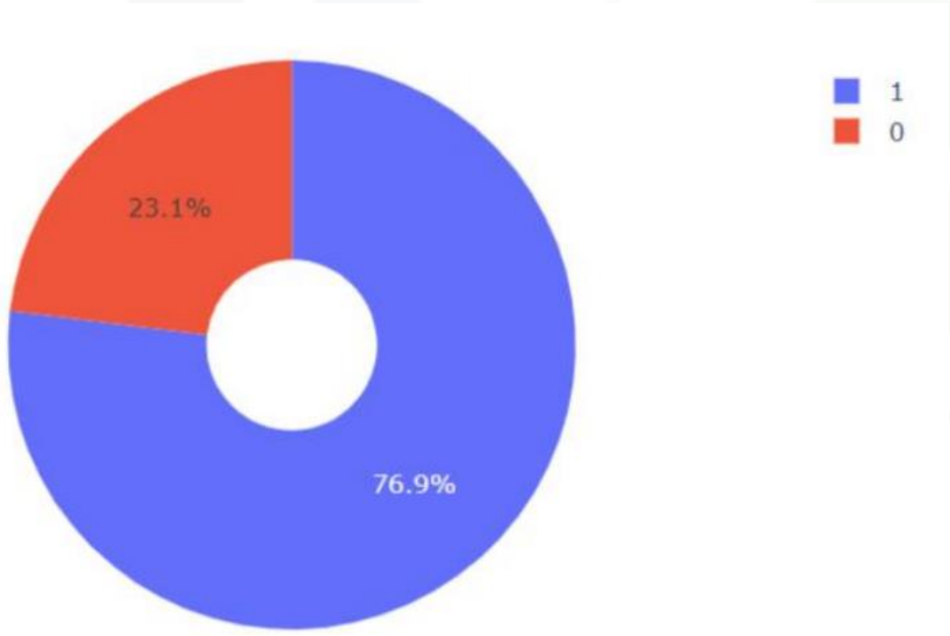
RESULTS & INSIGHTS FROM EDA & ANALYTICS

Pie chart showing the success percentage achieved by each launch site



We can see that KSC LC-39A had the most successful launches from all the sites

RESULTS & INSIGHTS FROM EDA & ANALYTICS



Pie chart showing the Launch site with the highest launch success ratio

PREDICTIVE ANALYSIS(CLASSIFICATION)

Here, we perform exploratory Data Analysis and determine Training Labels

- create a column for the class
- Standardize the data
- Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data

IBM Developer

[Jupyter Notebook]

SKILLS NETWORK

PREDICTIVE ANALYSIS(CLASSIFICATION)

After comparing all the model's accuracy, Decision Tree comes out with the best accuracy.

```
models = {'KNeighbors':knn_cv.best_score_,
              'DecisionTree':tree_cv.best_score_,
              'LogisticRegression':logreg_cv.best_score_,
              'SupportVector': svm_cv.best_score_}

bestalgo = max(models, key=models.get)
print(f'Model with Best performance: {bestalgorithm} with a score of {models[bestalgorithm]}')
```

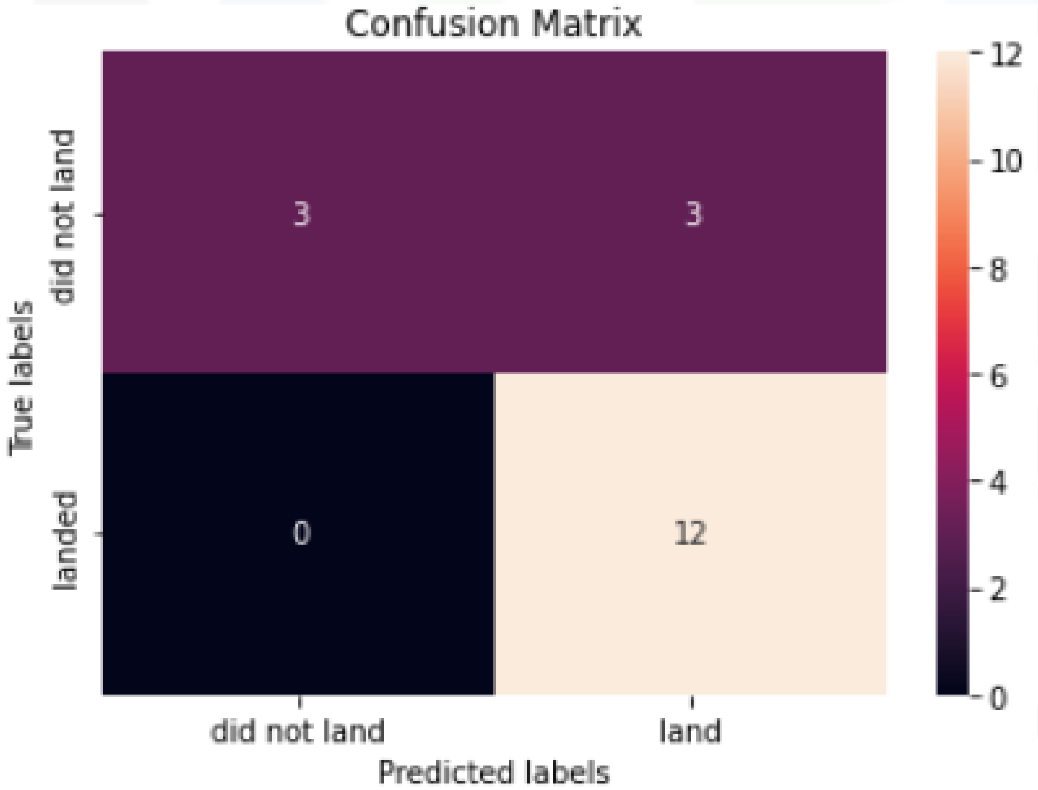
```
Model with Best performance: DecisionTree with a score of 0.875
```

Parameters:

```
tuned hpyerparameters :(best parameters)  {'criterion': 'gini', 'max_depth': 4, 'max_features': 'auto', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'random'}
accuracy : 0.875
```

PREDICTIVE ANALYSIS(CLASSIFICATION)

Using Confusion Matrix as our evaluation metric we see That,



There's a noticeable portion with

False Positives, meaning

The model is predicting

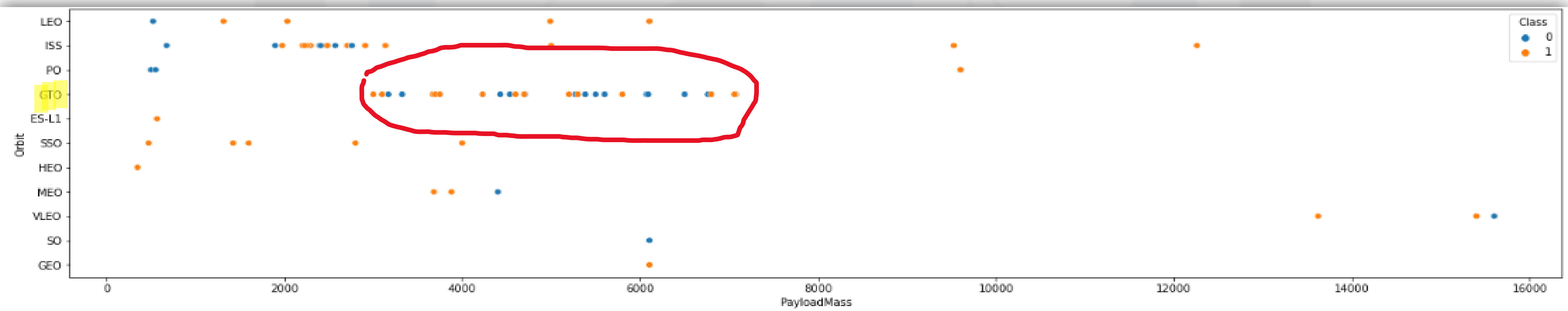
Successful landing even

When landing was

Unsuccessful.

INNOVATIVE INSIGHT

In payload vs orbit Scatter plot we observe that, Although for Polar, LEO and ISS, with heavy payloads the successful landing or positive landing rate are more.



However there is no such trend in GTO, as both positive landing rate(successful mission) and negative landing(unsuccessful mission) are both there. Making this particular Orbit absurd for our analysis and may even act as an outlier in the process.

'''CONCLUSION'''

After rigorous analysis, model training and hyper tuning, we come to the conclusion that:

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.