IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies

This Data science project was about finding the key states behind a successful landing the part 1 rocket fuel for a relaunch.

Data was collected using wiki tables and SpaceX APIs, and ELT was performed using pandas. SOL with IBM DB2 for Exploratory Data Analysis and finally a model was generated

- Summary of all results

Results show the various details and highlight key circumstances that make the launch successful. Different sites had different success rates which depended upon its proximity to landmarks.

All these details were taken into account when building the model for predicting the success rate given various inputs.

Introduction

- Project background and context

The following Data Science Project was conducted for a company named SpaceY a competitor to SpaceX.

Using Various information source we analyze the SpaceX profit parameters and success rates.

- Problems you want to find answers

What makes SpaceX launch sites have less success rates compared to others?

Can we predict the outcome for a launch, given sufficient inputs?

What can increase the chances to recreate a perfect landing(sites, landmarks, orbit etc.)?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX Falcon APIs, missing info and data validation was performed by web scraping on wiki tables. Extracted Information was further use to get precise data using various other SpaceX APIs
- Perform data wrangling
  - We used the powerful yet simple programming language Python, and The Pandas Library to perform Data Wrangling and ELT(Extract Load Transform)

Methodology

Executive Summary

- Performed exploratory data analysis (EDA) using visualization and SQL from IBM DB2 and SQL Alchemy
- Performed interactive visual analytics using Folium and Plotly Dash and build a dash Application
- Perform predictive analysis using classification models
    - Models were tuned using cross validation and GridSearch to find the optimum parameters, F1 and r2_scores were the scoring metrics

Data Collection

Data was collected using the API provided : https://api.spacexdata.com

Also web scraped using the wiki:-

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetwork-Channel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2021-01-01

The Wiki data was in the table, web-scraped using Requests and Beautiful soup

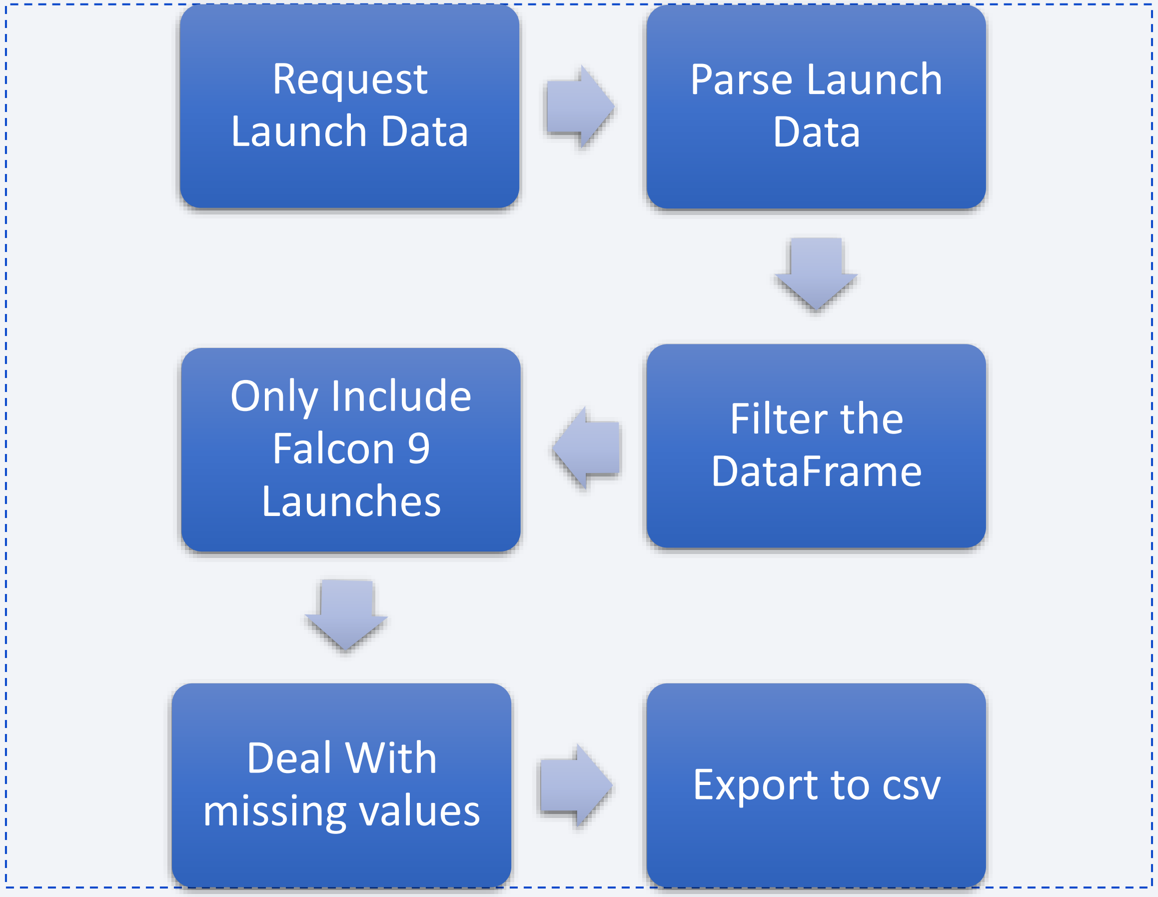
The Table contains List of Falcon 9 and Falcon Heavy Launches.

Data Collection – SpaceX API

- SpaceX Rest calls, call the SpaceX API to perform Data Collections

- GitHub:-

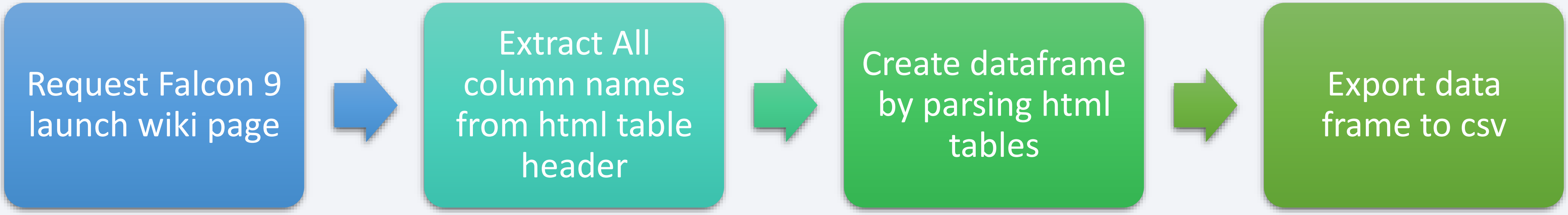
https://github.com/Aryan-909/SpaceX_Analysis_project/blob/master/Data%20Collection.ipynb



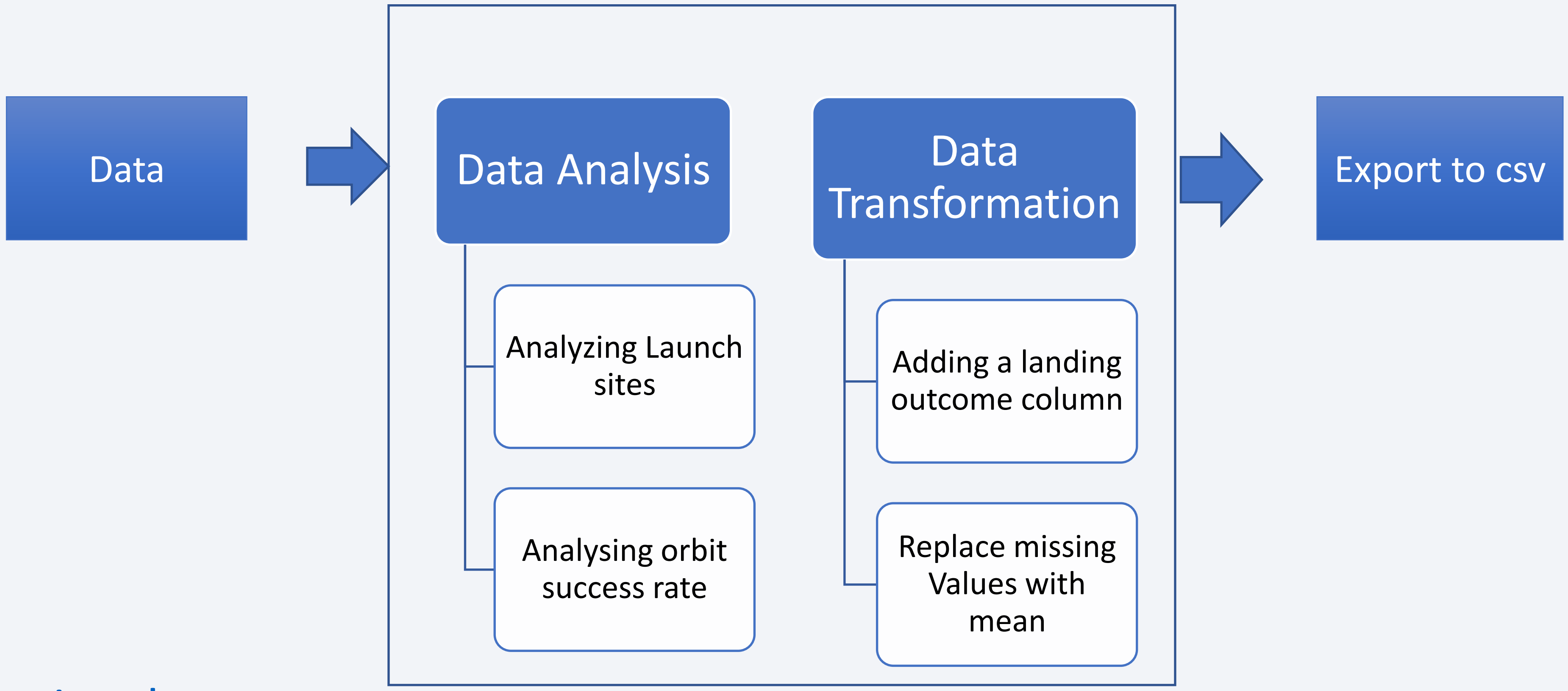
Data Collection - Scraping

Web Scraped Falcon9 and Falcon Heavy data using Request module and Beautiful soup.

GitHubURL: Data Collection using web scraping



Data Wrangling



GitHubURL:
DataWrangling.ipynb

EDA with Data Visualization

- Visualized the relationship between Launch sites and Flight no. using scatter plot.
- Scatter plot to visualize payload mass and launch site.
- Bar chart to show compare the success rate between each orbit.
- Visualized the relationship between Flight no., Orbit type and landing outcome.
- Scatter plot visualize payload mass and orbit type
- Visualized the yearly success trend.
- GitHubURL:[EDA_using_plotly_and_Pandas]

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- Display average payload mass carried by booster version F9 v1.1

EDA with SQL

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the *booster versions* which have carried the maximum payload mass
- List the failed *landing outcomes* in drone ship, for in year 2015
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20
- GitHubURL: EDA_using_sql

Build an Interactive Map with Folium

Foluim Objects used:-

- Markers – to mark locations of launch sites

- MarkerCluster – to generate landing outcome markers for individual launch sites

- Polyline – to check the distances of various landmarks such as coast, railroads, highways and cities.

GitHubURL: Visual_Analytics.ipynb

Build a Dashboard with Plotly Dash

Plots and DashBoards add to Web Application:

- Pie chart to visualize various landing outcomes or success rates for all sites
- Pie chart to similarly visualize for an individual landing site

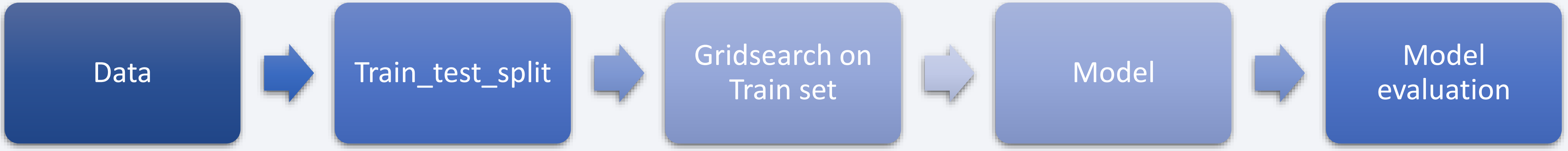
- Scatter plot to visualize the success rate and Payload mass for all sites or individual sites given a range of payload mass values. This range is accepted using the Range slider dash core component object

GitHubURL: Dash_app

Predictive Analysis (Classification)

We built 4 classification models using K-nearest-neighbours, Logistic Regression, Support Vector Machines and Decision tree Classifiers

The Model was improved using GridSearch and crossvalidation hyperparameter tuning with accuracy scores and confusion matrices to determine model accuracy



- GitHub URL: [Predictive_Analysis.ipynb]

Results

- Through the insights drawn from EDA we perform feature engineering on the data, using one-hot-encoding on all orbits and landing sites

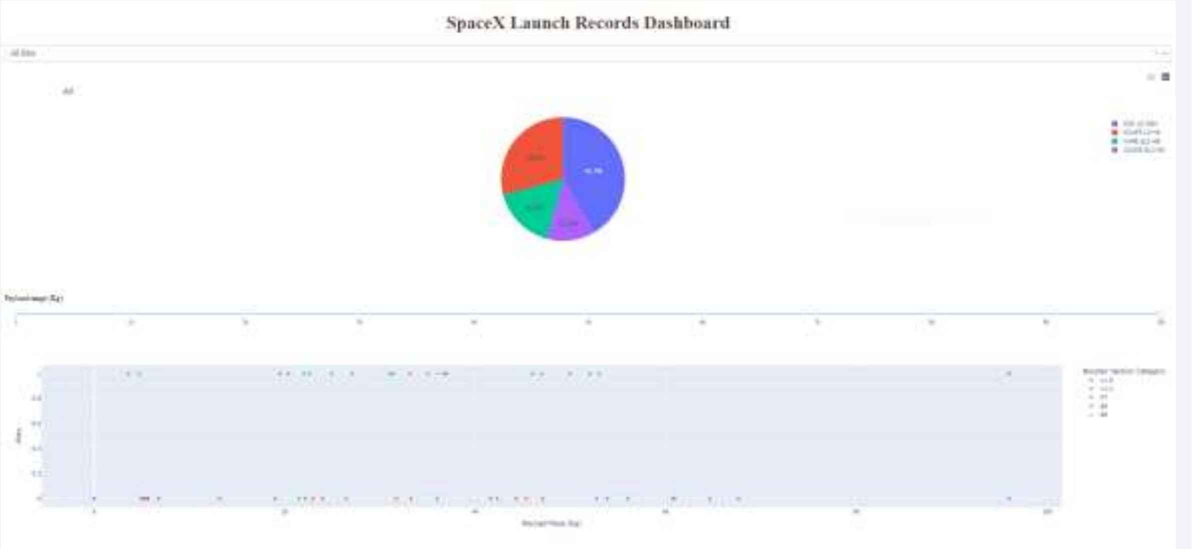
- In interactive analysis we saw the success rate of KLC-LC-39A was the highest.

- Also we discovered that some landing sites displayed a higher success ratio by having less payload mass making landing site success rate somewhat biased

Model Evaluation

- Decision tree showed the highest accuracy.

Which wasn’t surprising due to the uneven Distribution of data and a lot of categorical features

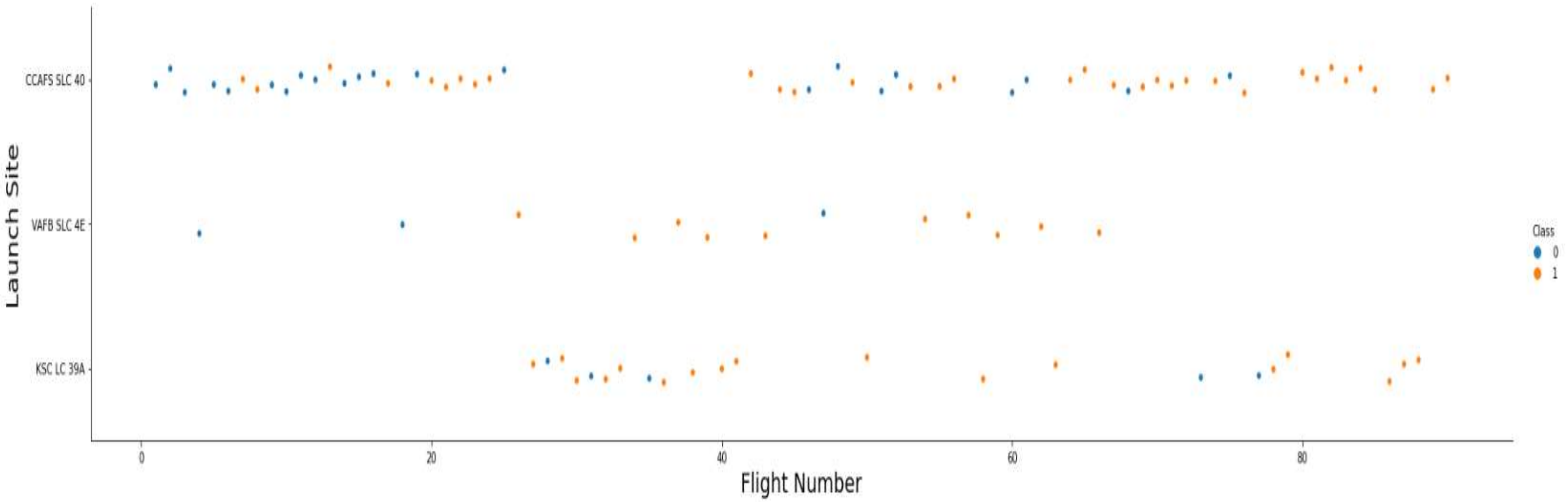


Section 2

Insights drawn from EDA

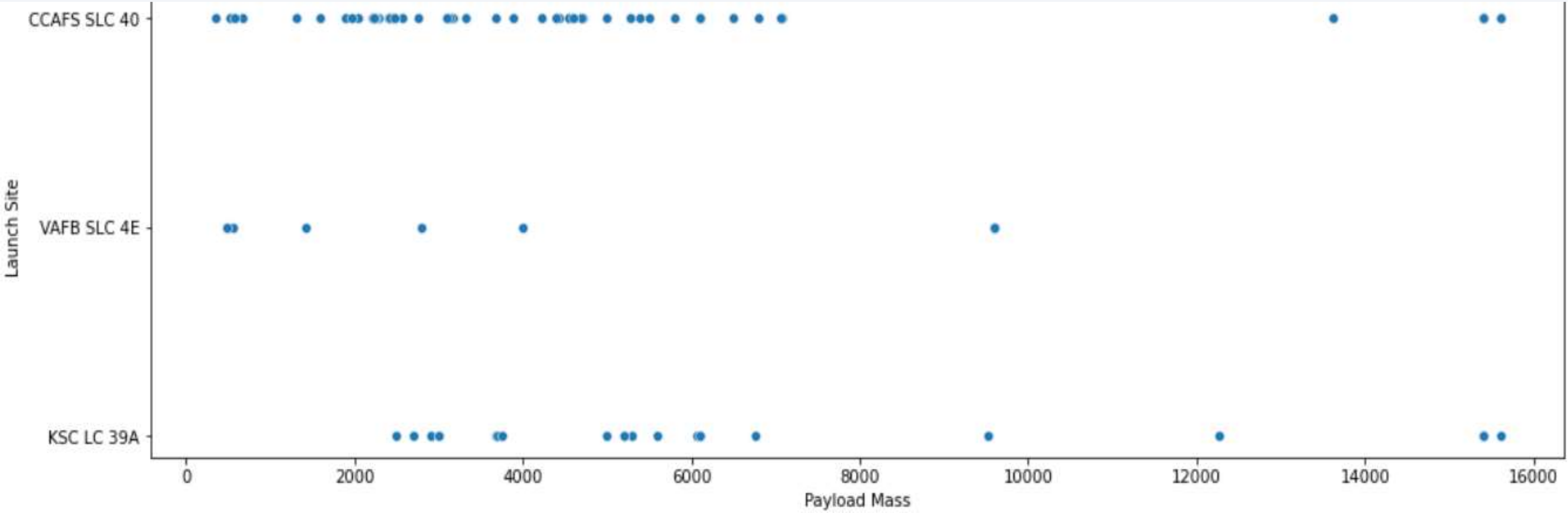
Flight Number vs. Launch Site

- We observe a trend that higher flight no. leads to higher chance of sucess



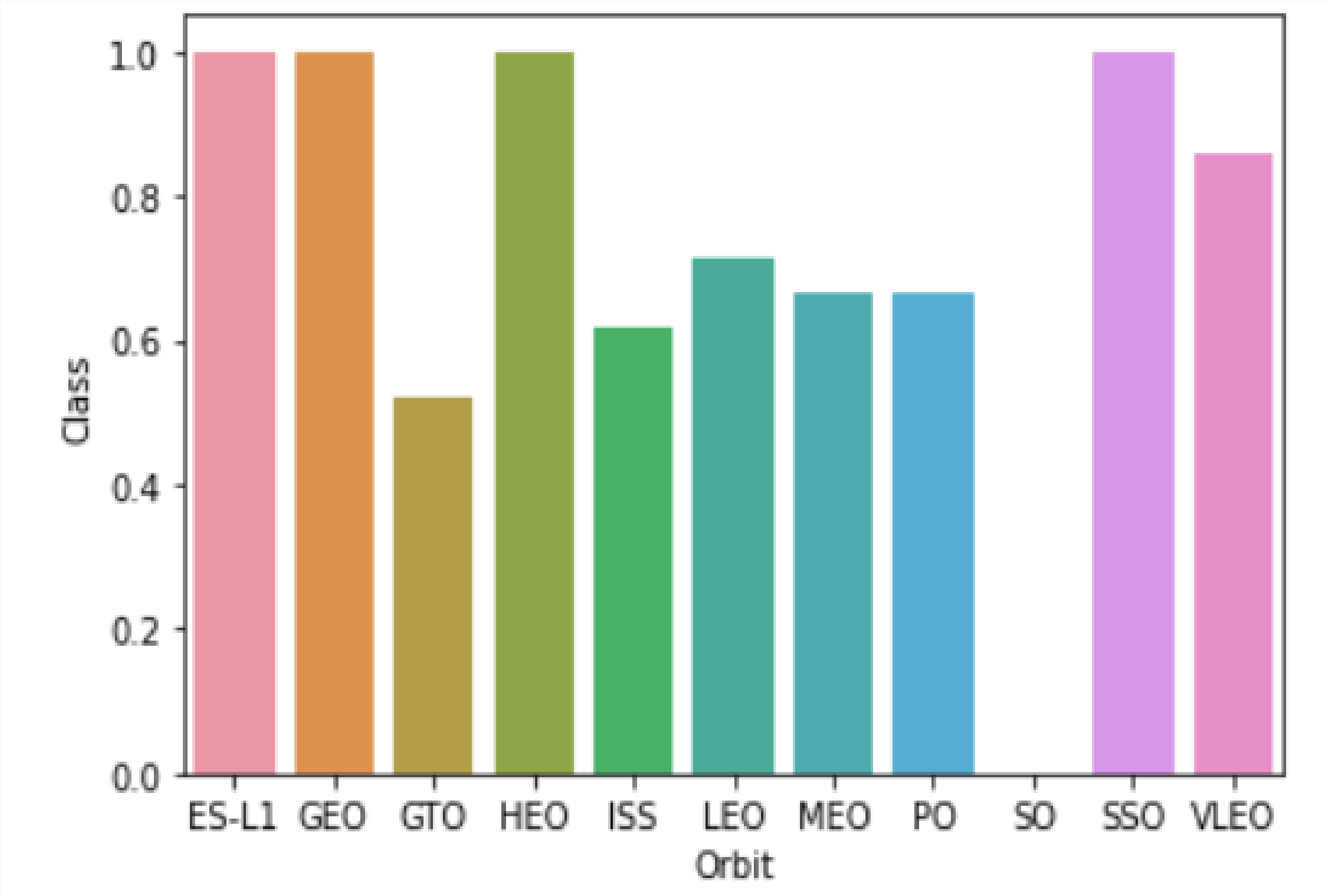
Payload vs. Launch Site

- We observe that payload mass of VAFB SLC 4E never exceeds 10k



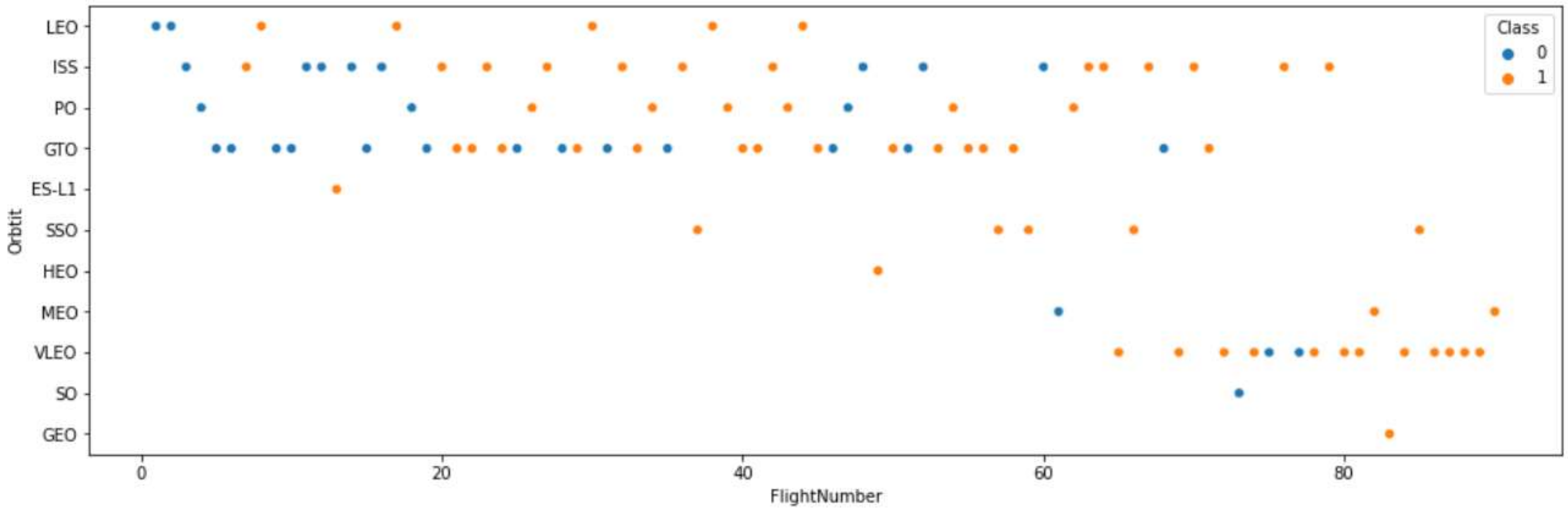
Success Rate vs. Orbit Type

- SO orbit type has the least success rate out of all
- ES-L1 and GEO have the highest success rates



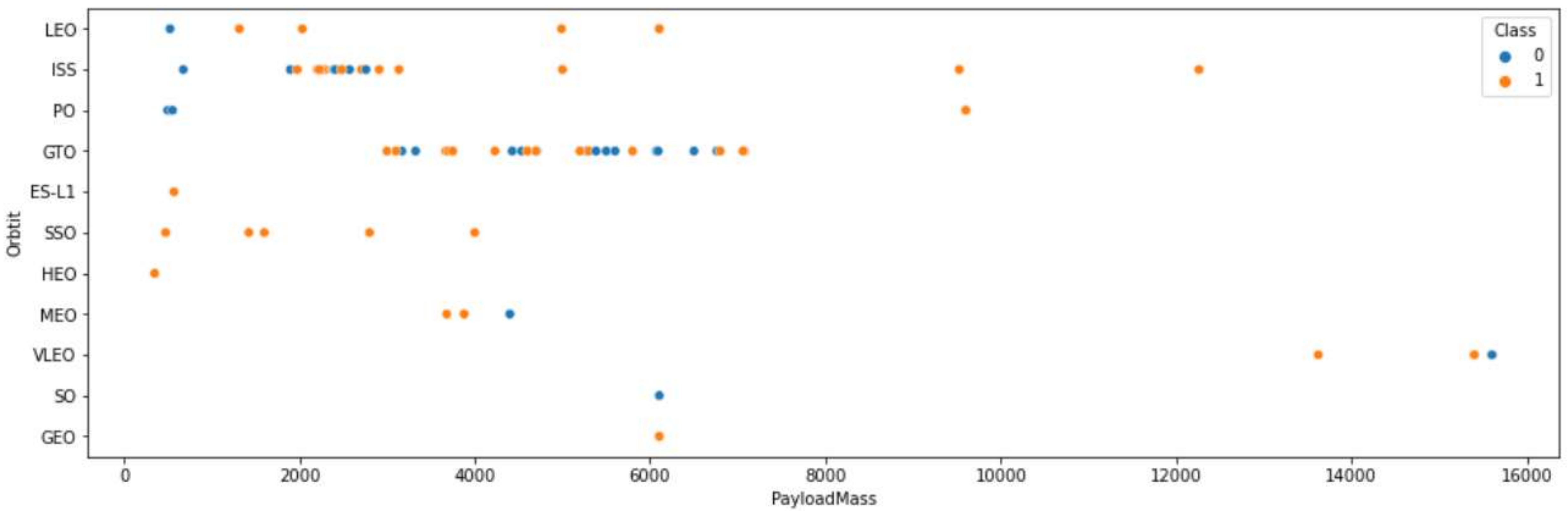
Flight Number vs. Orbit Type

- The success rate of ES-L1 shown in the previous slide is biased as there is only one successful flight recorded



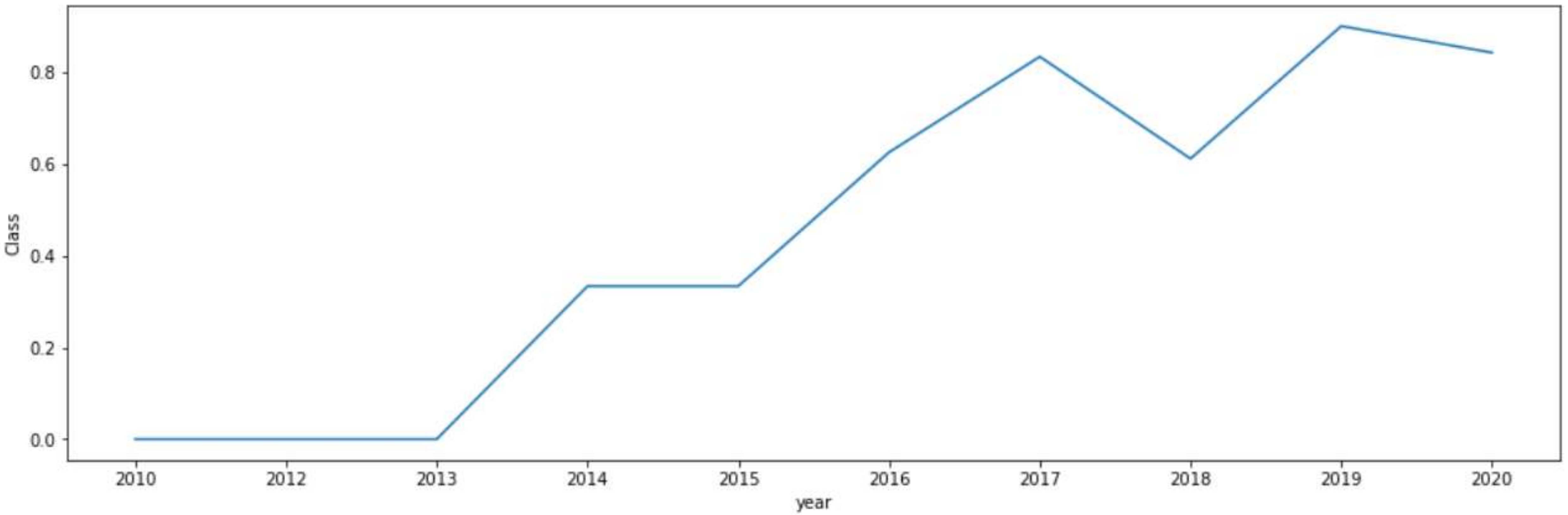
Payload vs. Orbit Type

- GEO success rate relies on payload mass of 6000k which is inconsistent making it again a bias option
- This makes VLEO, ISS the best pick when it comes to orbit type and Mass



Launch Success Yearly Trend

- The trend shows an increase in success rate from 2013.



All Launch Site Names

- List of all the Launch site names from a query result

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites 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	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

- Calculate the total payload carried by boosters from NASA

Query Result
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Query Result
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Query Result

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- Query Result for the given data extraction statement

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

- Most missions have a successful outcome (note this is not landing outcome)

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

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

2015 Launch Records

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

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

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

- We can see that drone ships were the most used landing platform with a 50%success rate

Landing _Outcome	Landing_Outcome_Count
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

Section 3

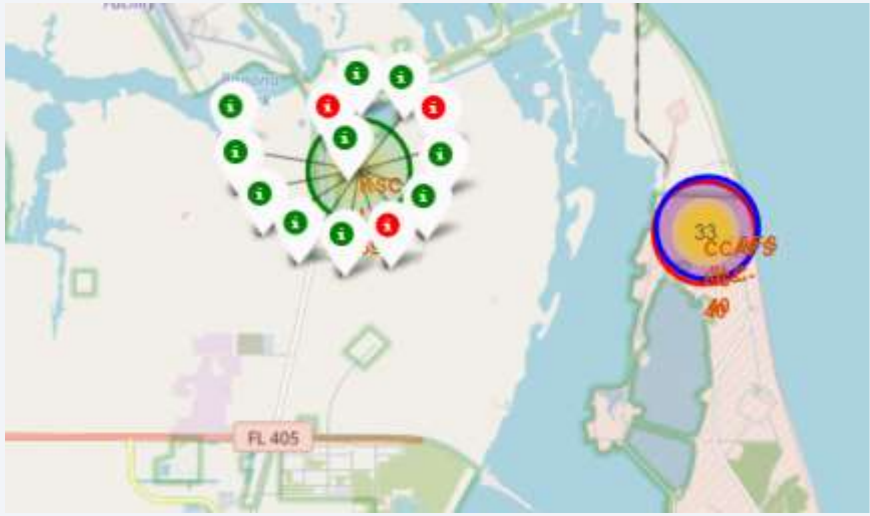
Launch Sites Proximities Analysis

Launch Sites

Launch sites are usually near coasts and nearer to the equator

Launch site Landing Outcomes

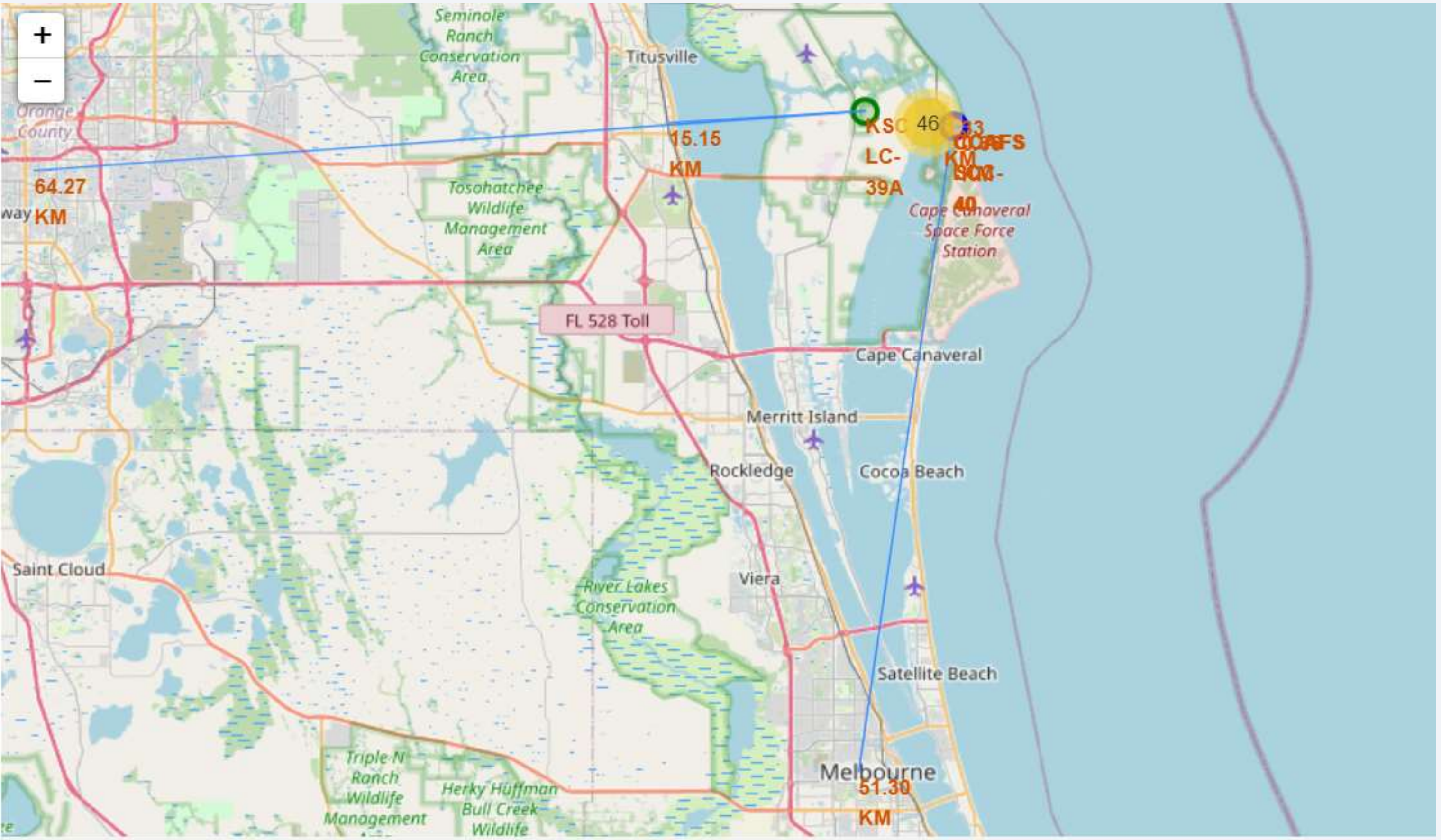
Launch site Landing Outcomes

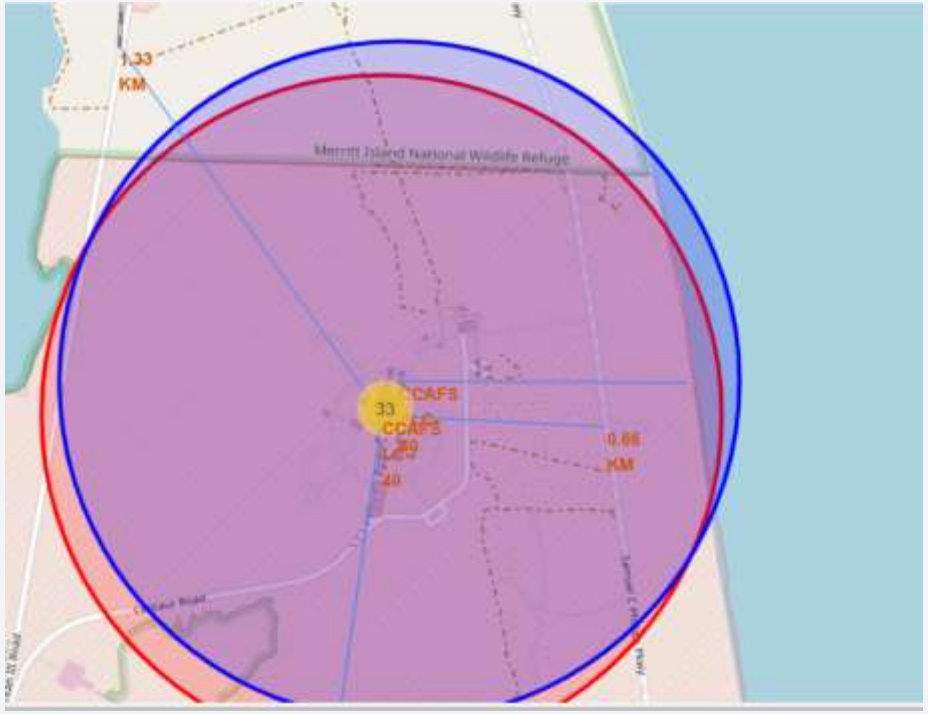


Launch site success outcomes, figure 1 shows the cluster for KSC LC-39A which is highest of all the landing sites

Fig 2 shows the landing sites which are in very close proximity

Launch Site close proximity land marks



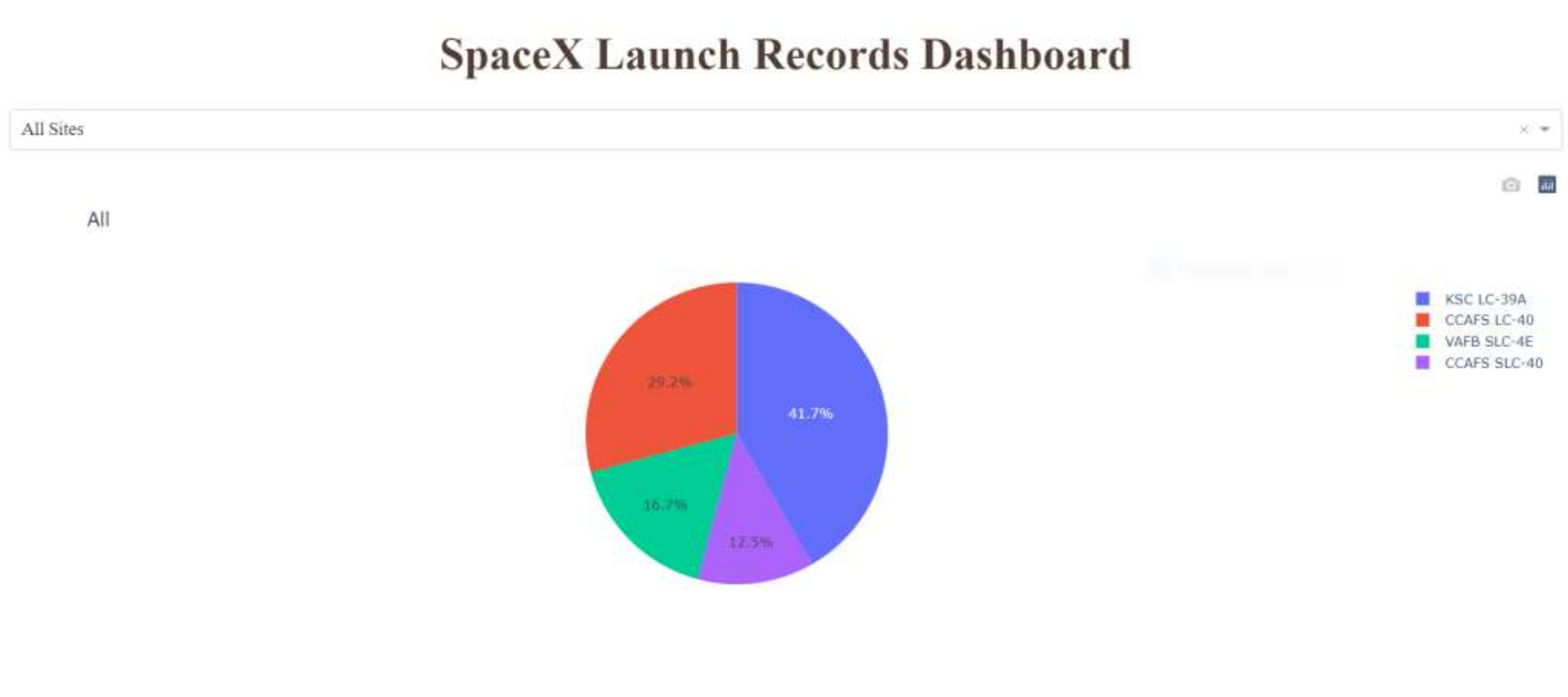


Lunch sites are significantly further from cities but closer to coast. No trend was observed for highways or railroads

Section 4

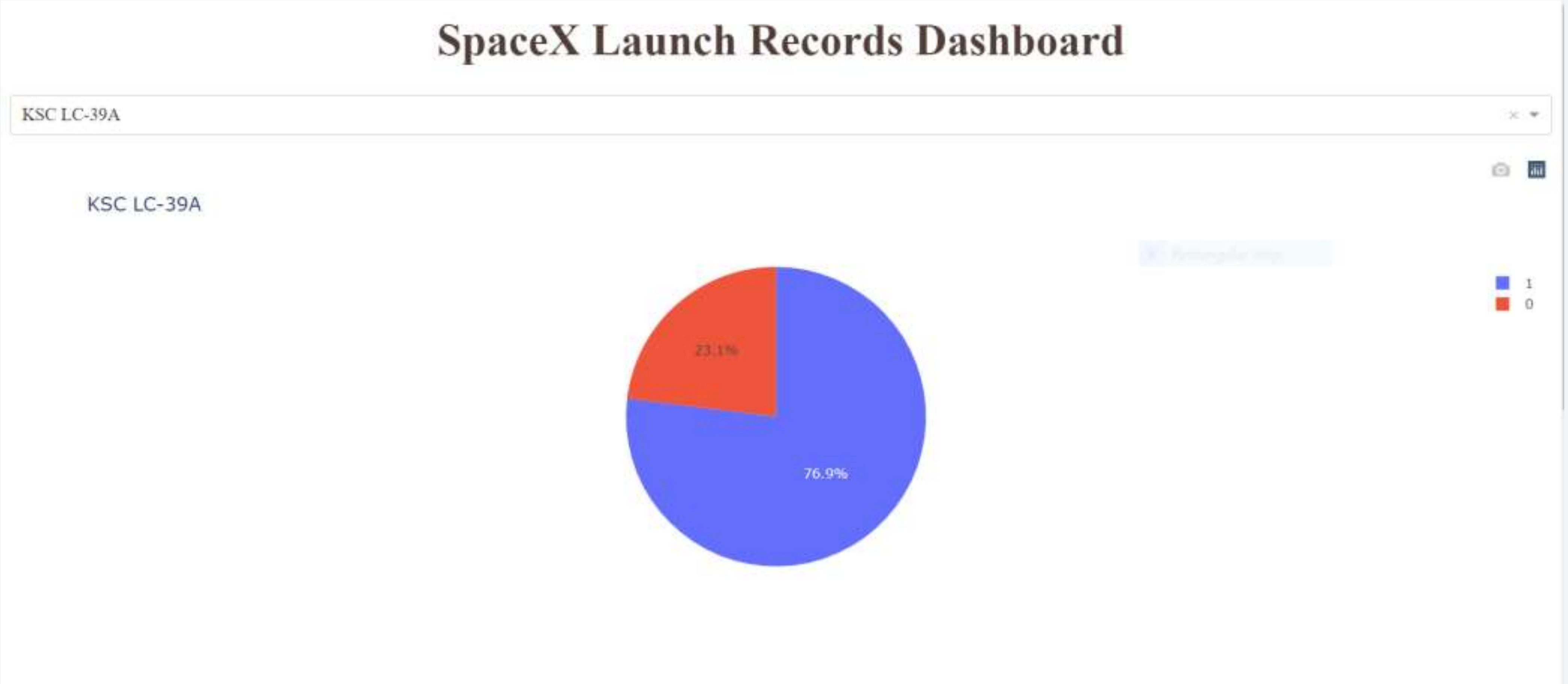
Build a Dashboard with Plotly Dash

SpaceX Launch Site Pie Chart

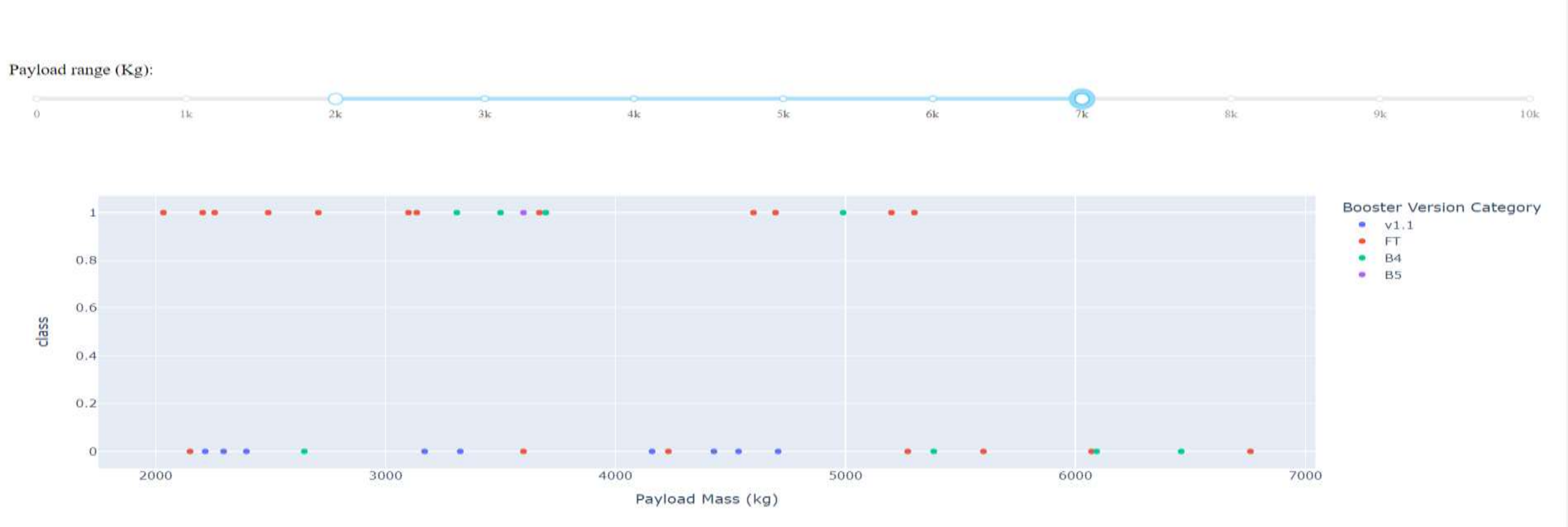


We can observe that the launch site with the highest success rate was KSC LC-39A

KSC LC-39A Landing Outcome pie chart



<Dashboard Screenshot 3>



- Booster version FT was able to show a high success rate between payload mass 2k and 4k
- Success rates seem to plummet beyond payload mass of 6k
- We can conclude that booster version 1.1 performed the worst even during optimal payload mass range of 2k-4k kg

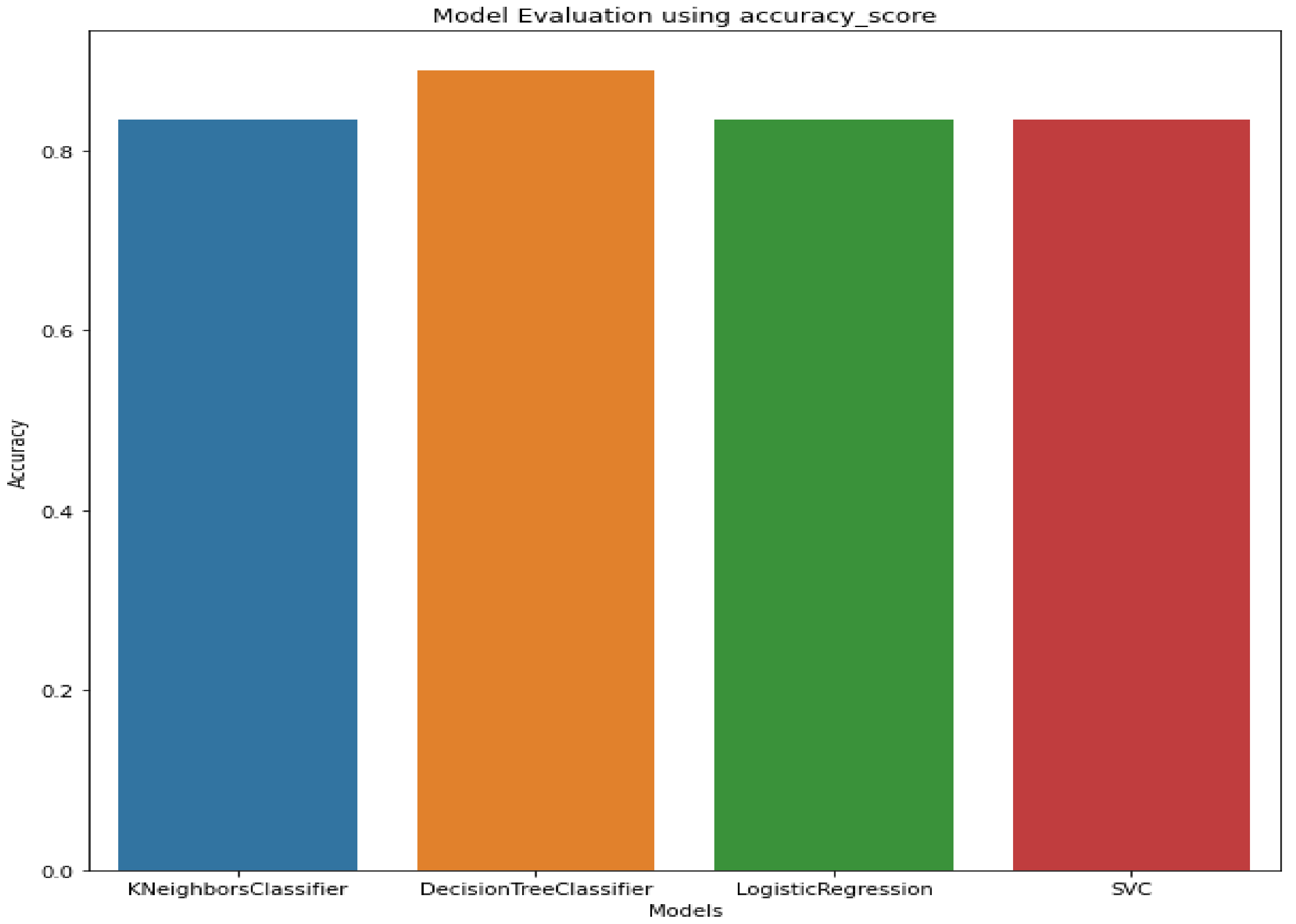
Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
bar_model_chart(mod,Y_test, X_test, accuracy_score)
```

```
[0.8333333333333334, 0.8888888888888888, 0.8333333333333334, 0.8333333333333334]
```

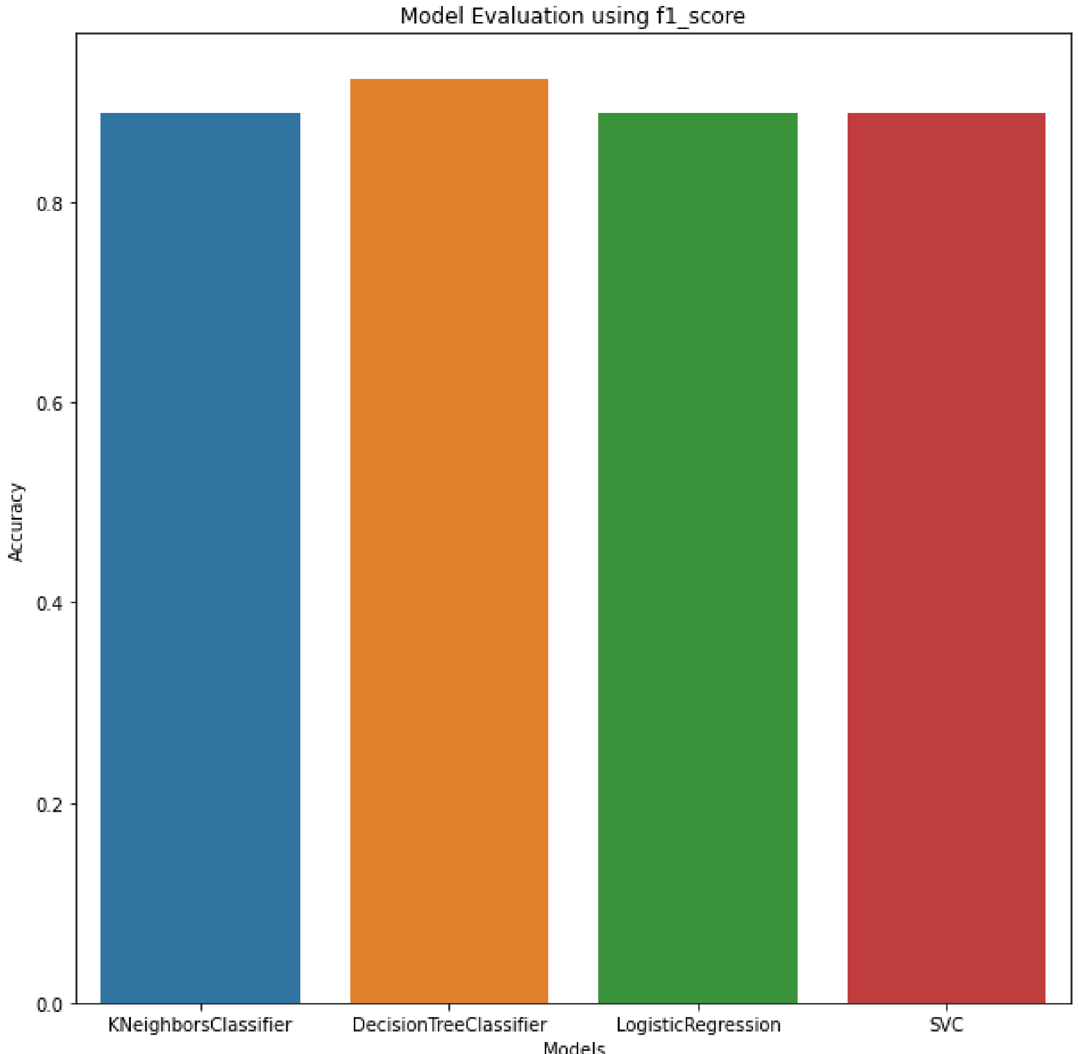


Model metric using Accuracy Score

Classification Accuracy

```
bar_model_chart(mod,Y_test, X_test, f1_score)
```

```
[0.888888888888889, 0.923076923076923, 0.888888888888889, 0.888888888888889]
```

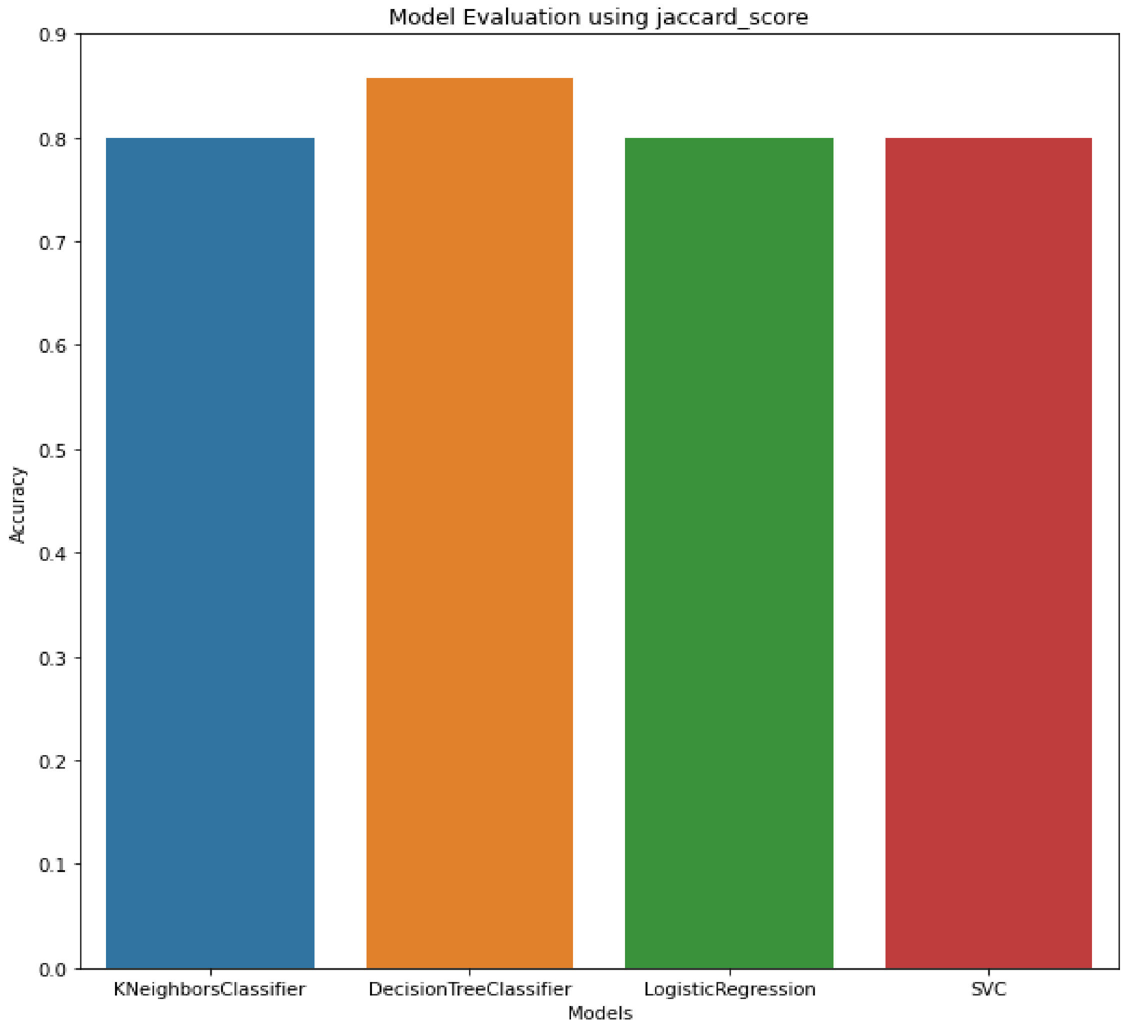


Model Evaluation using F1 score

Classification Accuracy

```
bar_model_chart(mod,Y_test, X_test, jaccard_score)
```

```
[0.8, 0.8571428571428571, 0.8, 0.8]
```



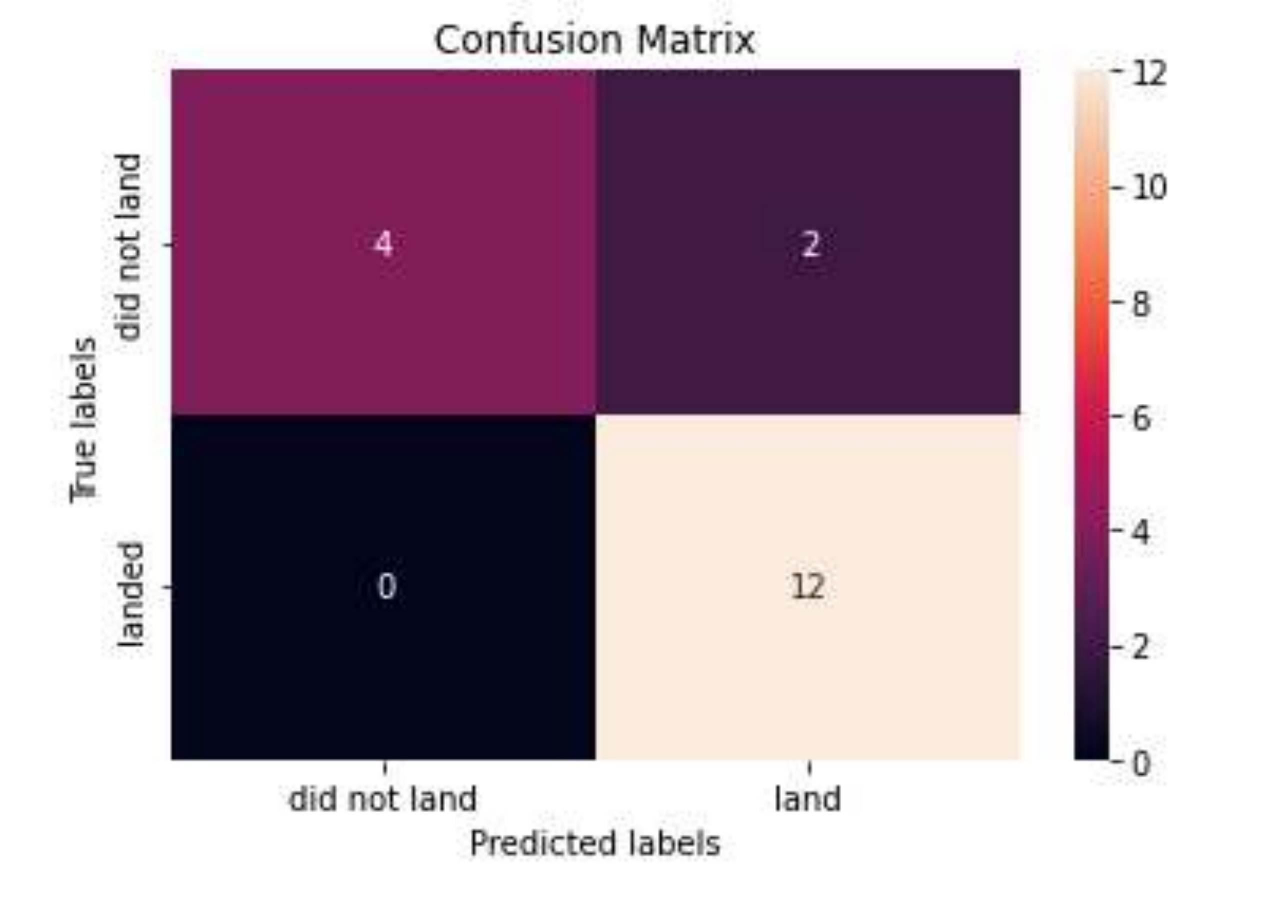
Model Evaluation using Jaccard Score

Classification Accuracy

Looking at the previous Three bar Graphs we can safely conclude that ***Decision tree Classifier*** is the best model for our project

Confusion Matrix

- Confusion Matrix of Decision Tree Classifier
- This shows that our model is great at identifying True Negatives(our rocket fuel landed).



Conclusions

- For the Highest success landing rate we need to have payload mass below 6k on better performing sites
- KSC LC-39 was the best performing site.
- Decision Tree Classifier managed to perform significantly better, as there were relatively more no. of categorical encoded columns
- Best performing Booster version was FT
- It was easier to predict the successful landing outcome by virtue of Decision Trees and a large amount of constraints.
- Best performing Orbit type was VELO

Appendix

Data sets used during the course of project

One_hot_encoding_Features :

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/api/dataset_part_3.csv

Base Features:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/api/dataset_part_2.csv

Visual_analytics_dataset -> dash_app/spacex_launch_dash.csv

Thank you!