IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Presented by: Hong
Date: 15/3/2023

Outline

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

Executive Summary

- Methodologies used to analyze data:
  - Application of web scraping and SpaceX API in Data Collection
  - EDA:
    - Data Wrangling, Data Visualization and interactive visual analytic
  - Machine Learning
- Summary of all results
  - The existence of EDA enabled us to prove hypothesis and gain valuable insights for deciding the characteristics of rocket.
  - Machine Learning drive us the pathway to utilise certain independent variable for predictions.
  - Raw materials or data has to be purified before conducting analysis on data.

Introduction

- **My personal background:**
  - **Experienced Data Scientist** at Space X
  - Assigned to perform analysis on rocket characteristics.
- **SPACE Y**
  - Founded by Alien Mask
  - Competent against SPACE X
  - **Intention:** Outperform their rival by undermining the cost of launching a rocket in the first stage.
- This presentation will be covered the relationship between **a variety of related variable and the target which is the success rate of launching a rocket.**

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collected by requesting and parsing **SpaceX launch data** through **Space X API** and **HTTP** using Request library.
- Perform data wrangling
  - Removing **unnecessary column, convert descriptive text into binary** and **fill empty cells**.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Applying **a few machine learning models** tested with **diverse range of parameter** using **GridSearch** to **obtain the highest prediction score**.

Data Collection

- SpaceX's launching data is an open source on the web.

  1. With the aid of Request.get and BeautifulSoup, we were able to request the required information which is a collection of tables from SpaceX database and WIkipedia through HTTP as well as SpaceX API.

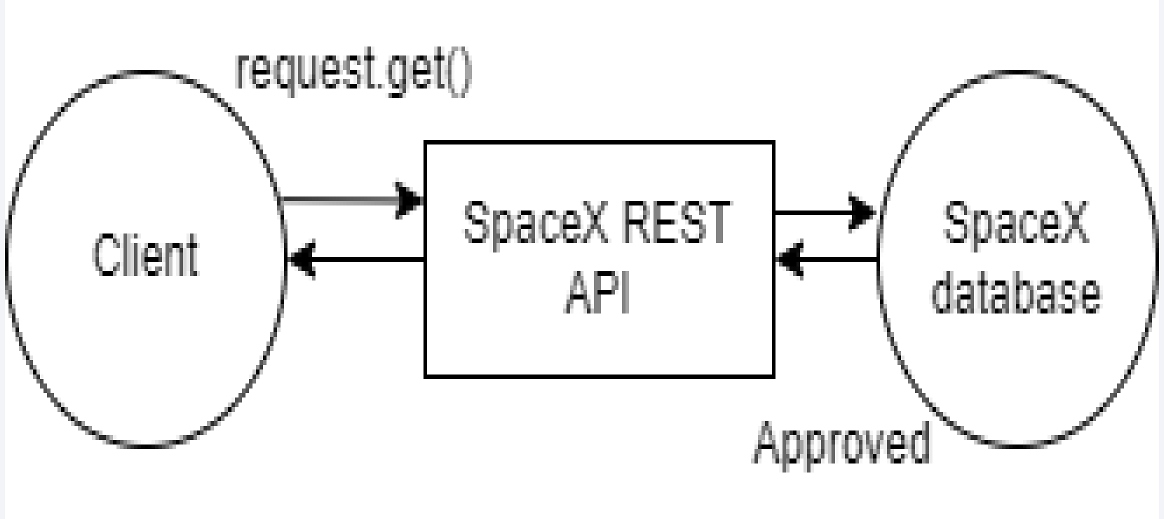
  2. Some data were originally classified as a Json file.

  3. Decoded with corresponding python function.

  4. Utilizing pandas to arrange the data into human readable content.

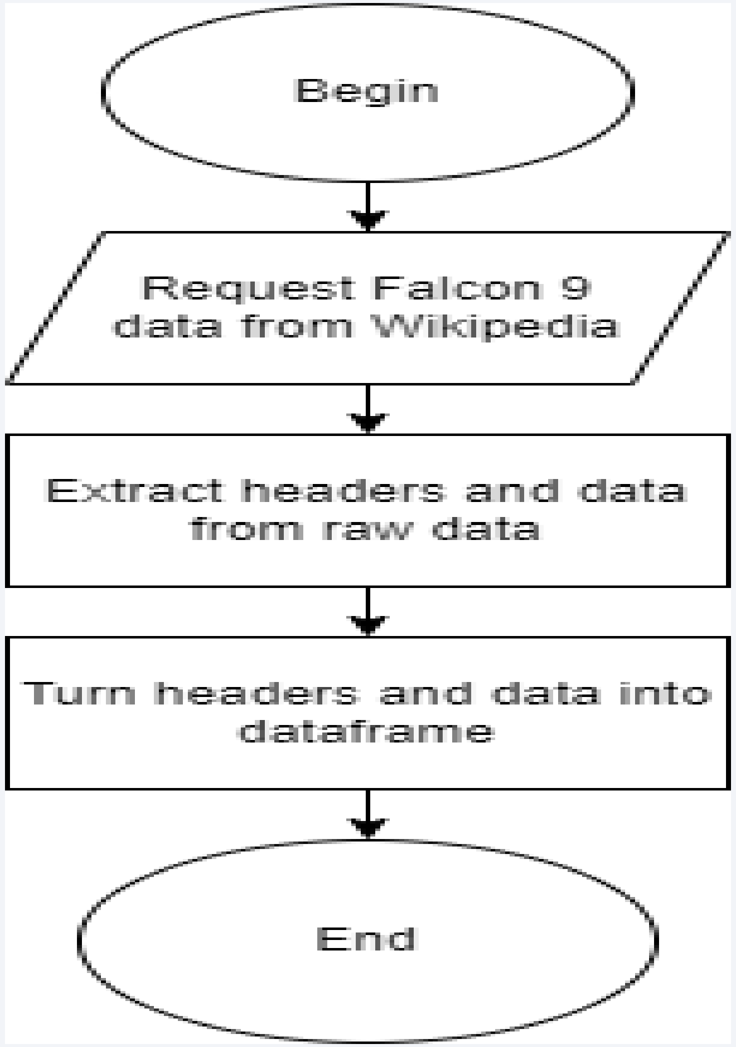
Data Collection – SpaceX API

- **We as a client** to the web send information request.
- Once it's approved, the data requested will be transferred via API.
- Link for reference:
  - [Module 1]()



Data Collection - Scraping

- **Request.get()** – request falcon data
- **BeautifulSoup()** – Parse and extract headers and data
- **Pd.Dataframe()** – Form a table filled with column names and information.
- Link for reference:
    - Module 2

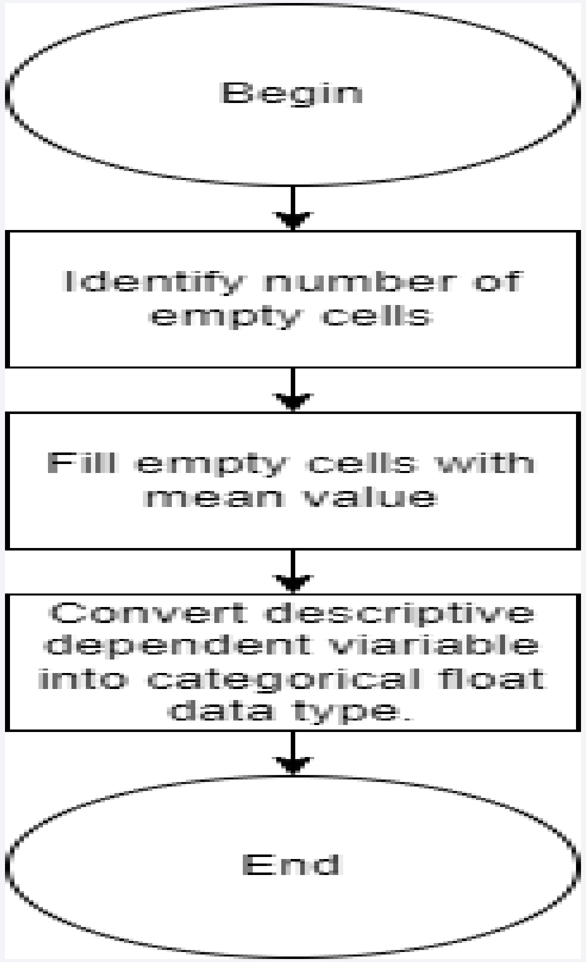


Data Wrangling

- Raw material collected consists of blank in certain cells due to unforeseen circumference.
- Dataframe provides us with the ability to process the data with mean numbers.
- The conversion of categorical target variable from text ease the analysis of data.

Link for reference:

- Module 3



EDA with Data Visualization

- There are few plot implemented for visualize purposes:
  - Categorial plot
  - Regression plot
  - Bar plot
  - Linegraph
- Link for reference:
  - IBM-Final-Project-SpaceY/IBM MODULE 5.ipynb at main · keemstarr123/IBM-Final-Project-SpaceY (github.com)

EDA with SQL

- SQL performed:
  - `select * from SPACEXTBL LIMIT 5`
  - `select distinct Launch_Site from SPACEXTBL`
  - `select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5;`
  - `SELECT SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer = 'NASA (CRS)';`
  - `select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version = 'F9 v1.1';`
  - `SELECT Date from SPACEXTBL where "Landing _Outcome" like 'Success (ground pad)' order by date desc limit 1`
  - `select Booster_Version from SPACEXTBL where Mission_Outcome = "Success" and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000;`
  - `select Mission_Outcome,COUNT(Mission_Outcome) from SPACEXTBL group by Mission_Outcome;`
  - `select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ =  (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)`
  - `select substr(Date, 4, 2), "Landing _Outcome", Booster_Version, Launch_Site from SPACEXTBL where "Landing _Outcome" like "Failure%" and substr(Date, 7, 4) = '2015';`
  - `select Date, "Landing _Outcome", COUNT("Landing _Outcome") from SPACEXTBL where Date between '04-05-2010' and '20-03-2017' group by "Landing _Outcome"  order by COUNT("Landing _Outcome") DESC;`
- Link for reference:
  - Module 4

Build an Interactive Map with Folium

- Map object that have been added:
  - Folium.marker – Mark the coordinate of rocket's launch site
  - MarkerCluster – Grouping a cluster of incident on map
  - Folium.PolyLine – Line on map linked to nearest city, highway and railway
- But why we added these stuff?
  - It's easier to generate valuable insights based on the interactive map.
  - Making comparison amongst different launch site which can be plotted in a map
  - Graph is not capable of performing the requested action related to geographical areas.
- Link for reference:
  - Module 8

Build a Dashboard with Plotly Dash

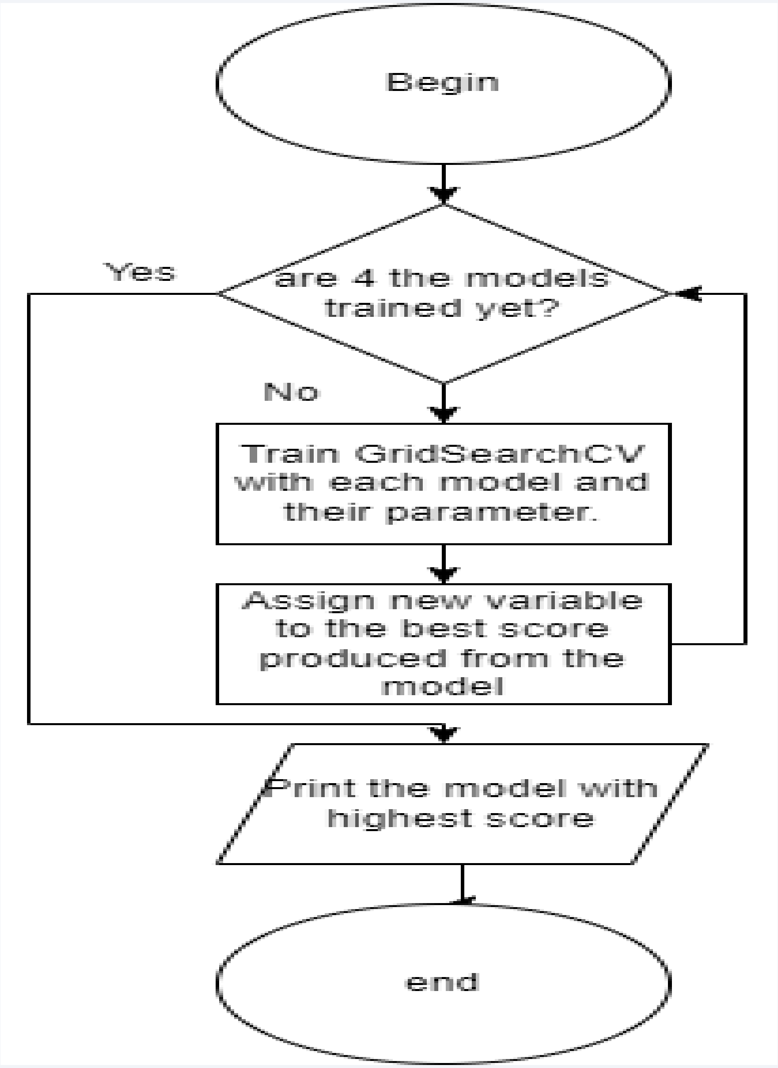
- There are several features which enable you to interact with to view plot/graph with the characteristic selected:
  - Markdown
  - SlideRanger
- Dashboard regarded as a interactive multimedia offers us the advantages to view the graphs with desired parameters that can be manipulated by the users.
- Link for reference:
  - Module 7

Predictive Analysis (Classification)

- 4 dissimilar machine learning are implemented in order to produce the best outcome:
    1. K-Nearest-Neighbors
    2. Tree Decision Classifier
    3. Support Vector Machine
    4. Logistic Regression

    For each model, a wide of parameter are assessed via GridSearchCV to evaluate the effectiveness of each parameter.

- Link for reference:
    - [Module 9]



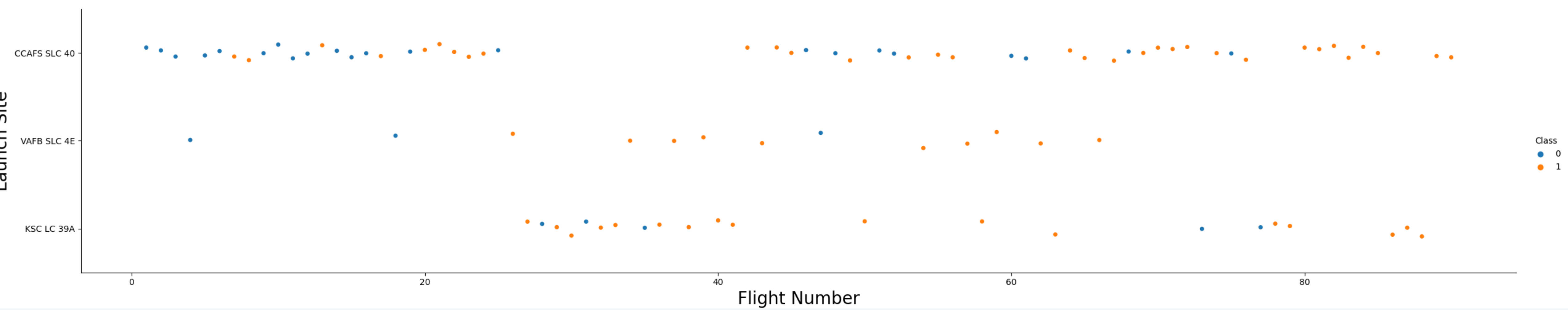
Results

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

Section 2

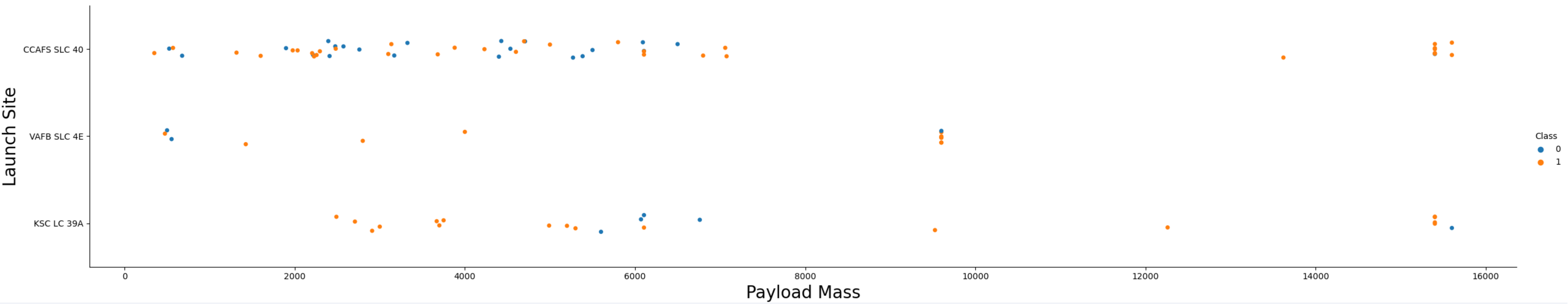
Insights drawn from EDA

Flight Number vs. Launch Site



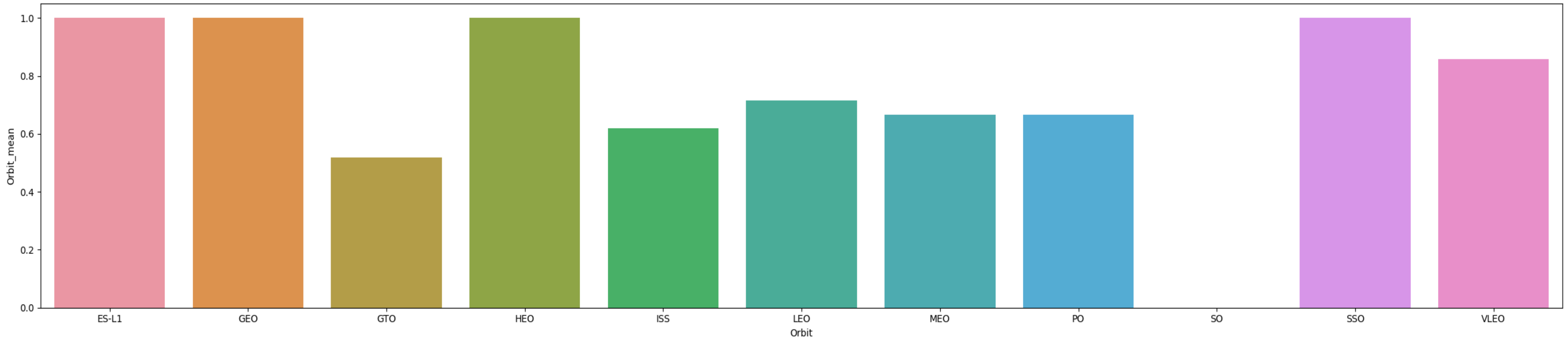
- KSC LC 39A has a lower rate of failure while CCAFS SLC 40 has the highest failure percentage.
- Thus, this indicates that KSC LC 39A is the best launch site among the three options

Payload vs. Launch Site



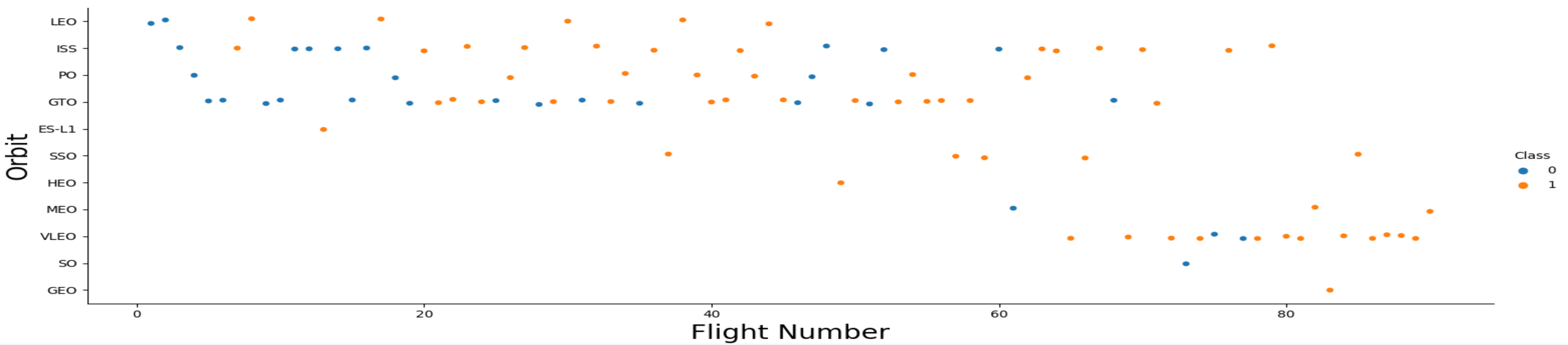
- Payload mass exceeding 8000kg have a higher possibility of succeeding.
- Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



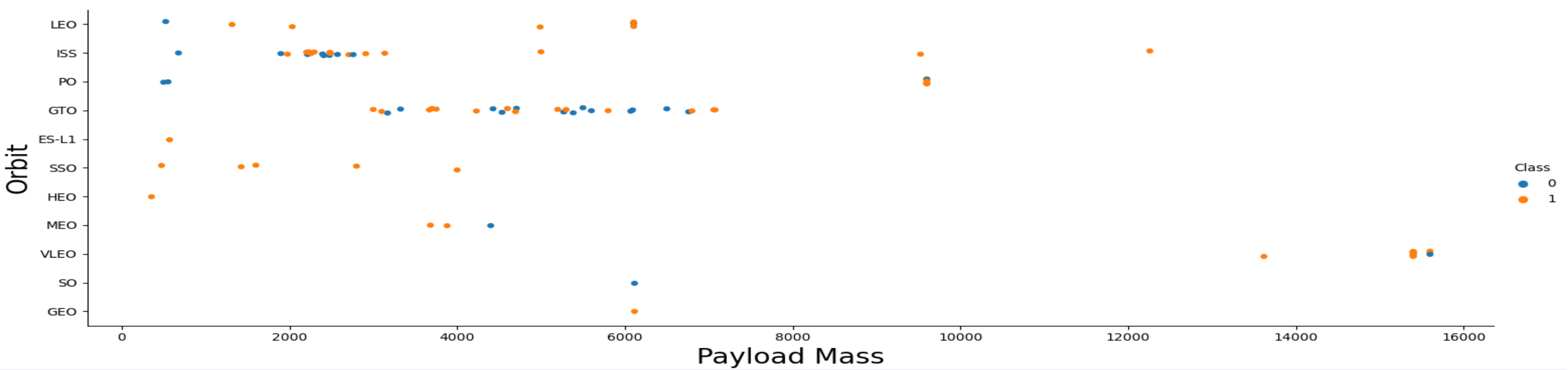
- Launching rocket to ES-L1, GEO, HEO and SSO has the highest rate of success.

Flight Number vs. Orbit Type



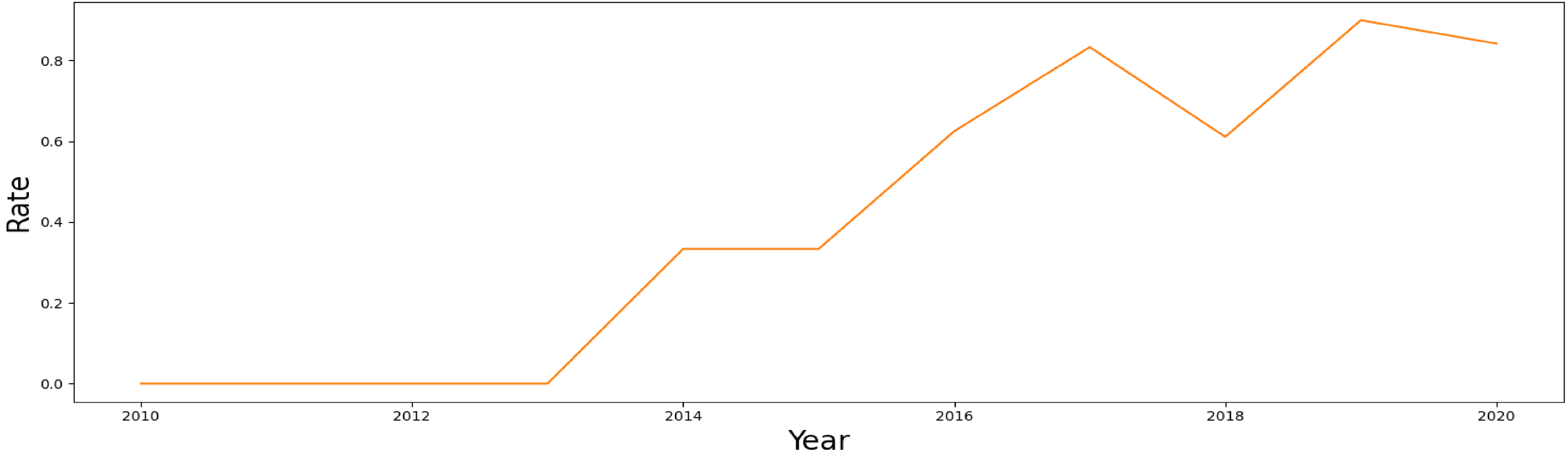
- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



- 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.

Launch Success Yearly Trend



- the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

Launch Site Names Begin with 'CCA'

- First 5 Record with Launch Site Name Begin with CCA

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- This is the total payload carried by boosters from NASA

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

Date
22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- The boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes

Mission_Outcome	COUNT(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- 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

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

substr(Date, 4, 2)	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- The ranking of count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Date	Landing _Outcome	COUNT("Landing _Outcome")
07-08-2018	Success	21
08-10-2012	No attempt	10
08-04-2016	Success (drone ship)	8
18-07-2016	Success (ground pad)	6
10-01-2015	Failure (drone ship)	4
05-12-2018	Failure	3
18-04-2014	Controlled (ocean)	3
04-06-2010	Failure (parachute)	2
06-08-2019	No attempt	1

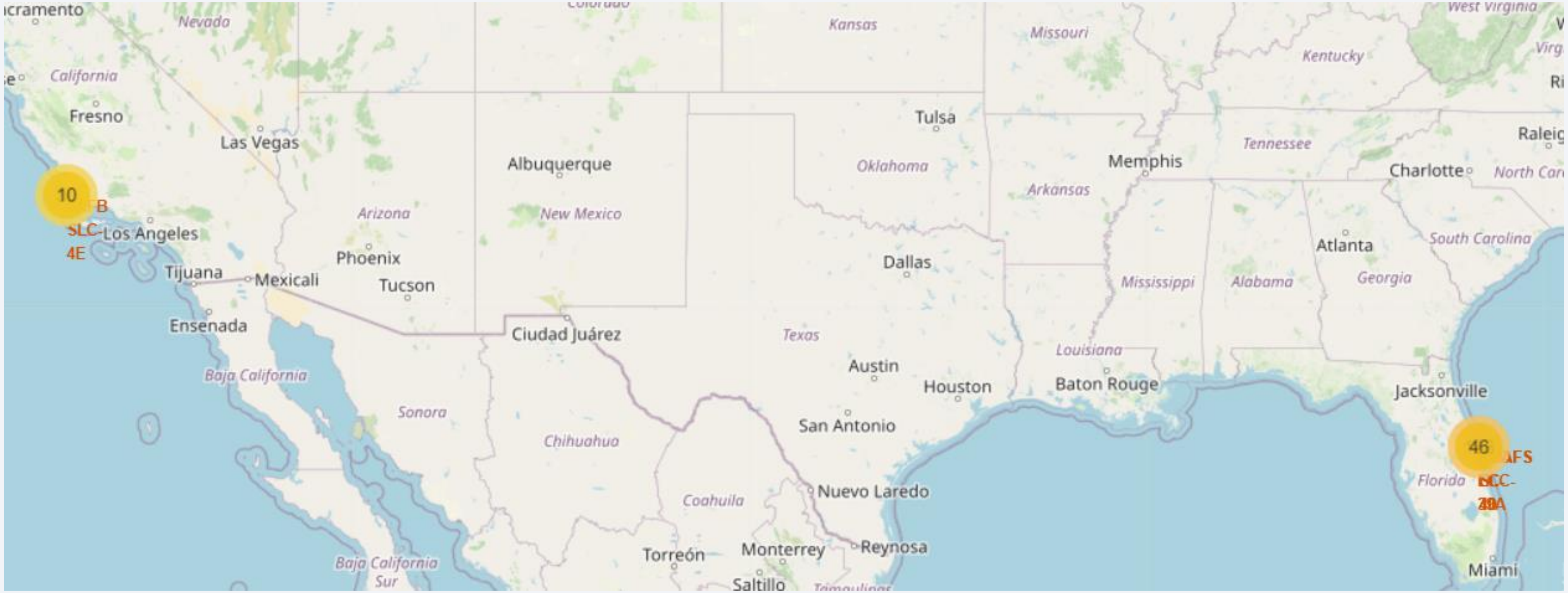
Section 3

Launch Sites Proximities Analysis

Launch Site Geographical Area

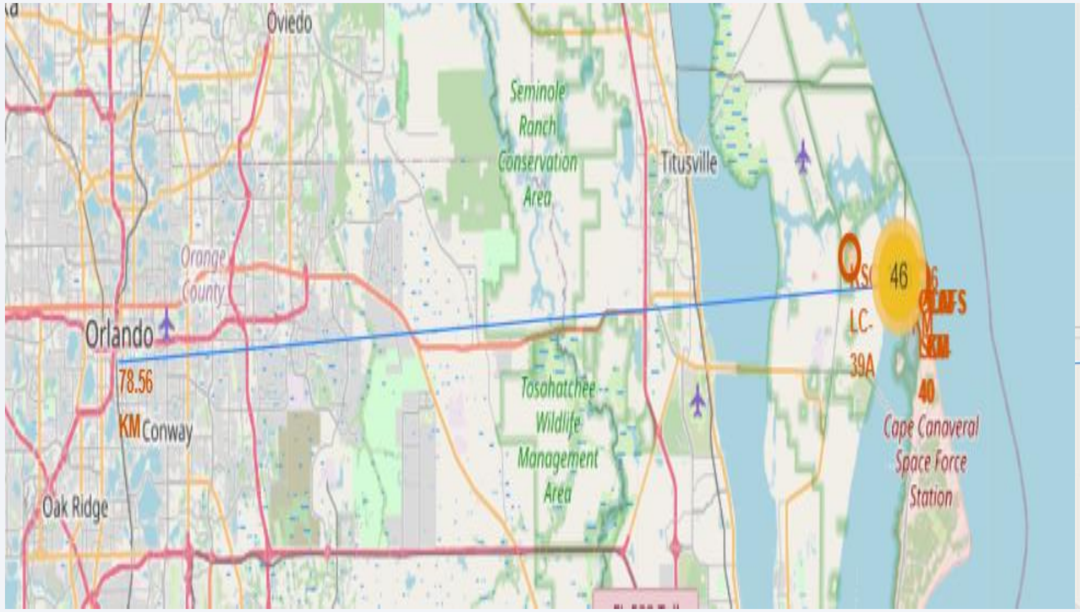
- As you may see on the map, the launch sites are marked with red dotes along with their name.
- Launch sites are chosen near to the coastline.

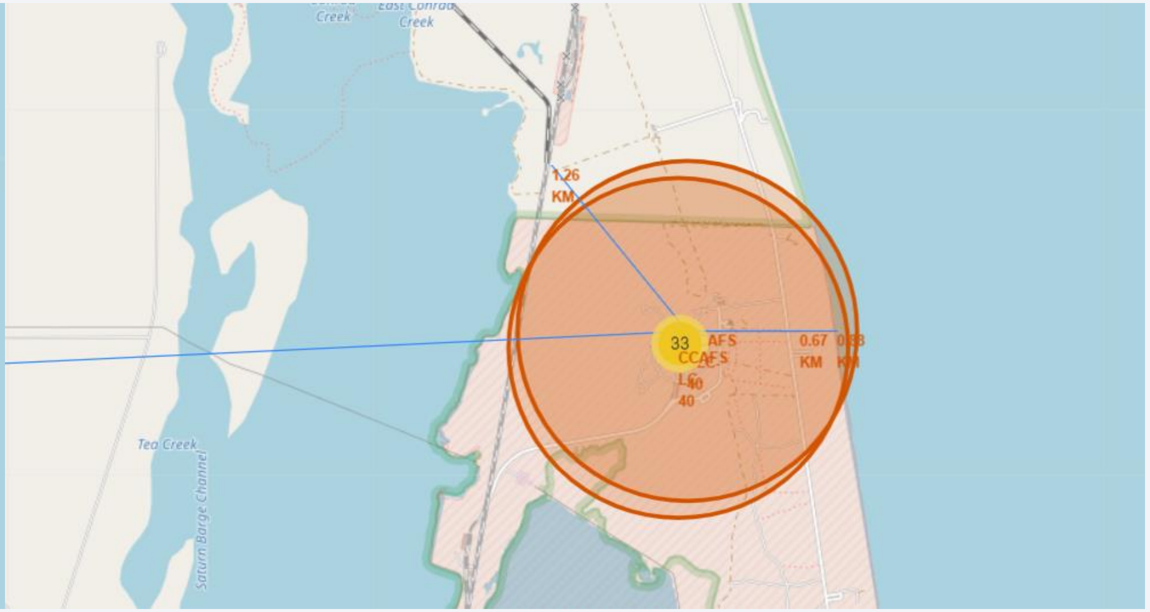
Rocket Launched Clusters



- The number in a circle demonstrate the amount of rocket launched in that area.

Distance between Launch site and infrastructures



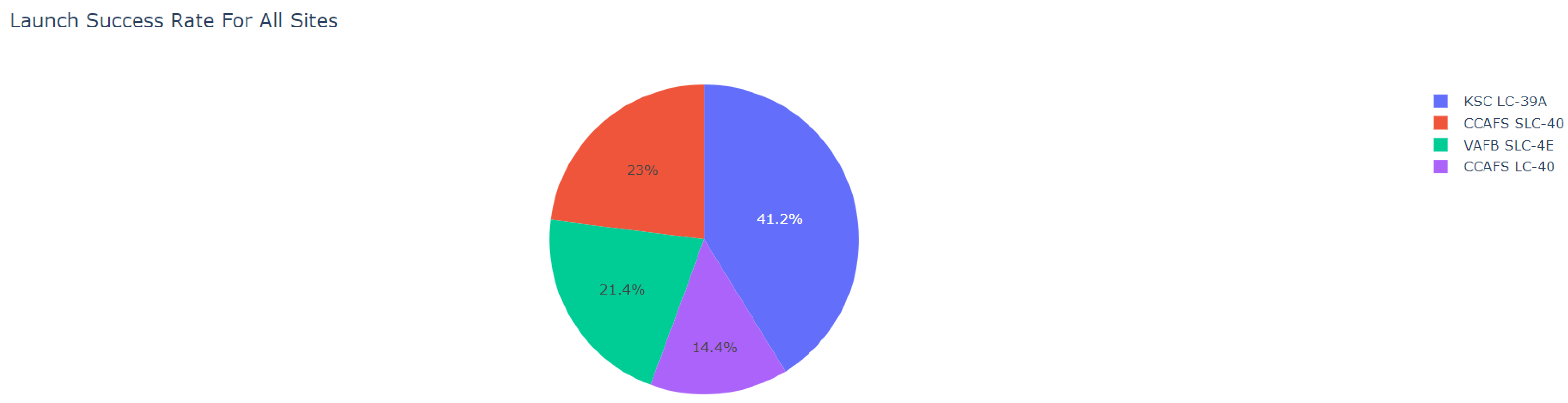


- Blue line on the map shows the distance between the launch site and the nearest city, railway and highway respectively.

Section 4

Build a Dashboard with Plotly Dash

Pie chart for launch success rate of all sites

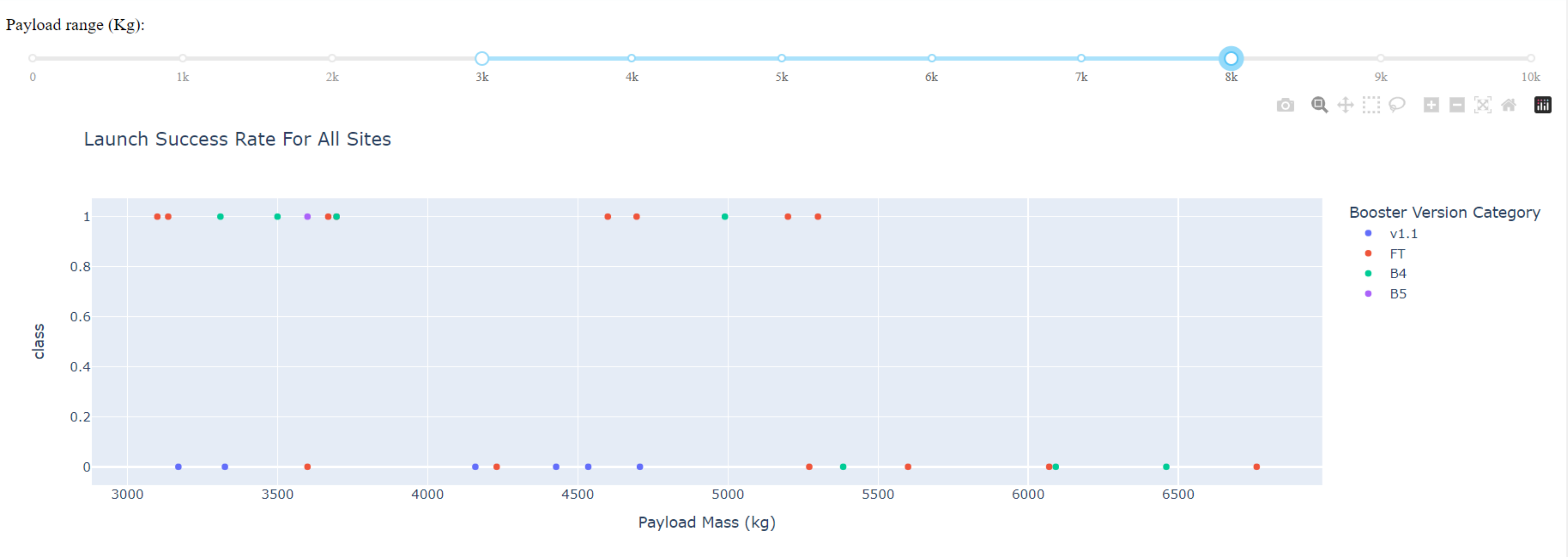


- Pie chart above shown the success rate of each rocket launching sites.
- Figure shown KSC LC-39A has the most success launches whereas CCAPS LC-40 has the least.

Highest Launching Success Rate

- Chart tells that KSC LC-39A has the highest success rate among all of the launch site for which it has a 76.9% chance to succeed.

Launch Success Rate with Payload Mass Ranging Between 3000 to 8000

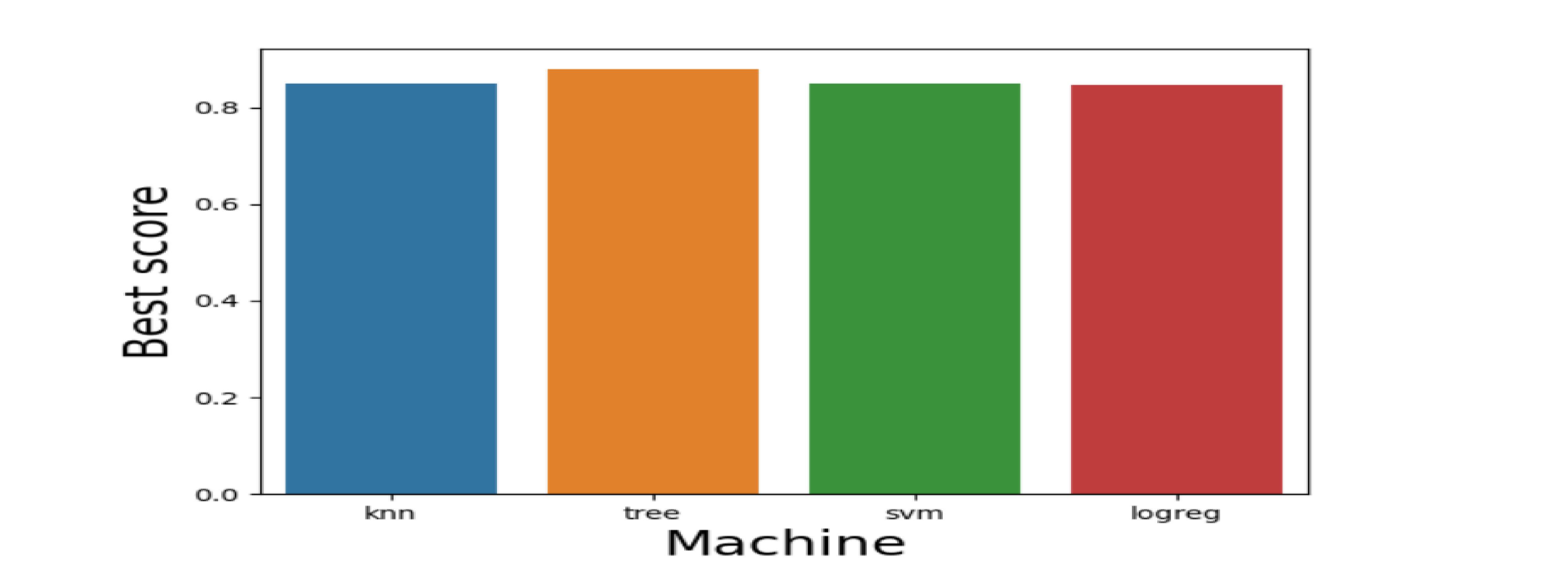


- Plot above shows the Launch Success Rate for all site with a payload mass between 3000 to 8000.
- Figure shows FT Booster version has the most success case.

Section 5

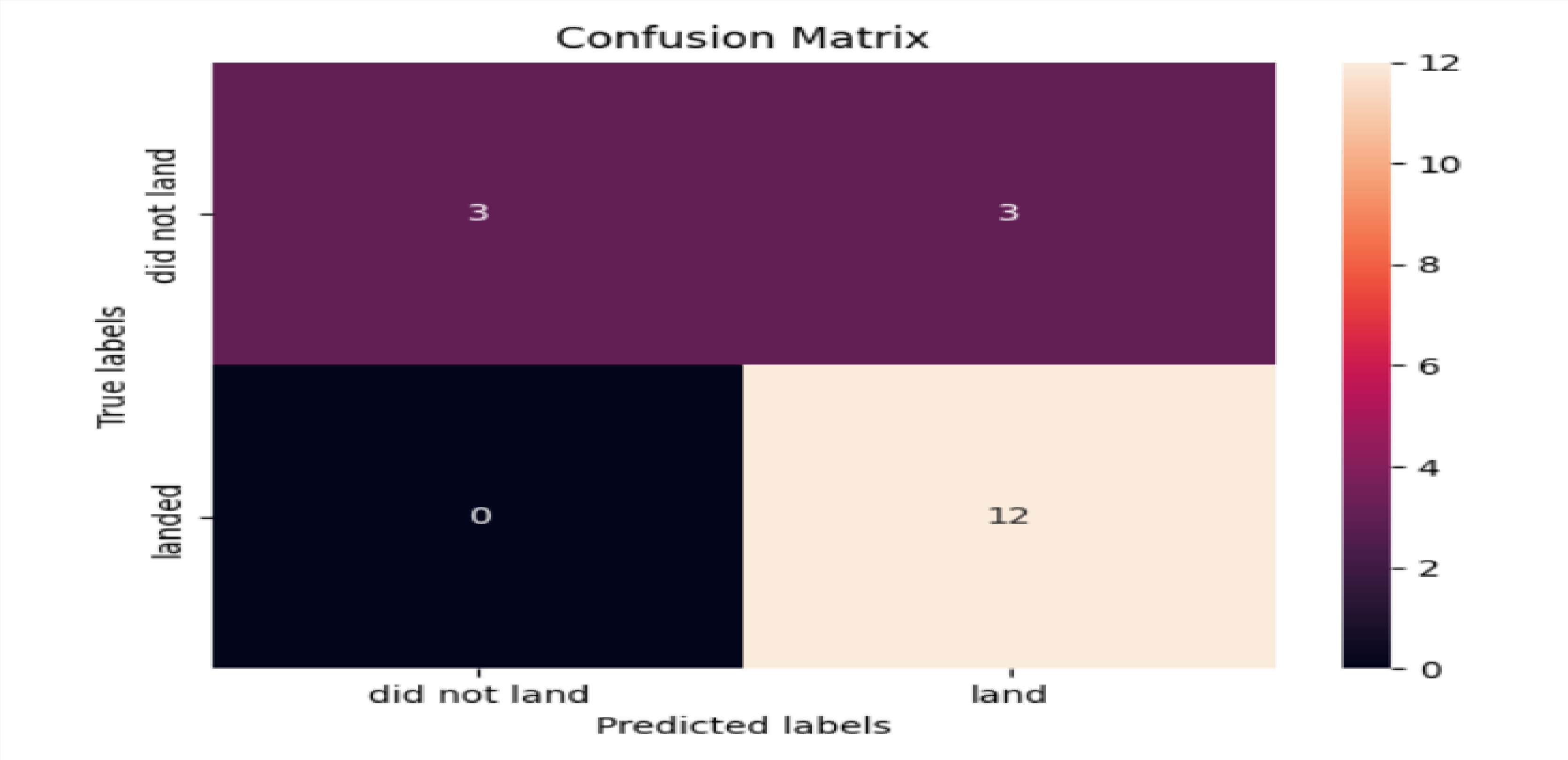
Predictive Analysis (Classification)

Classification Accuracy



- Bar chart above shows Tree Decision Model has the highest best score, 0.9 than the rest

Confusion Matrix



- Confusion matrix displayed portrays the major problem of the model which is false positive in which it predicted 3 did not land as land. Meanwhile, it anticipate all the correct output for landed.

Conclusions

- With the aim of increasing the possibility of succeed, KSC LC-39A is the best spot for rocket launching at this point.
- Furthermore, ES-L1, GEO, HEO and SSO are the recommended orbits to embark the launching of rocket
- Moreover, it is affirmative that payload with mass more than 8000kg can land or set off without crashing and hence we shall increase the mass of rocket until it exceeds 8000kg.
- Decision Tree Classifier is the best categorical model by far.

All the best for SPACE Y!

Thank you!