IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Saleh Binsuwaylih>
<2023/6/12>

Outline

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

Executive Summary

Metholedgy

- Collect using API and web scraping
- Data Wrangling
- EDA with Data Visualization
- EDA with SQL
- Build a Dashboard with Plotly Dash
- Predictive Analysis (Classification)

Result

- Utilizing multiple models, we trained the data and compared accuracy on both testing and training sets. This approach offers a concise yet thorough assessment of model performance and generalization capabilities.

Introduction

- This machine learning project is designed to predict the success of the first stage landing of a rocket. The primary objectives include identifying relationships between attributes and determining the most suitable model to fit the data, aiming for the highest accuracy.
- SpaceX has disrupted the space industry by offering Falcon 9 launches at just 62 million dollars, significantly undercutting competitors. Their groundbreaking idea of reusing the first stage through precise re-landing drives these cost savings. As a data scientist for a startup rivaling SpaceX, our project aims to predict first-stage landing outcomes using machine learning, crucial for bidding competitively. Challenges include identifying key landing factors, understanding variable relationships, and determining optimal conditions for success. • The objective for Space Y, a new player, is to assess its competitiveness against SpaceX. Our focus is on predicting successful first-stage landings to estimate total launch costs effectively. Additionally, the analysis aims to identify optimal launch locations, providing strategic insights for Space Y's market positioning and competition with SpaceX.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected from two source Api and Web Scraping. by the API from SpaceX Json, creating data frame from the response of the API. And the Web Scraping using beautifull soup.
  - Api (https://api.spacexdata.com/v4/rockets/)
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - The data procced using pandas and numpy library
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Train the data using Linear Logistic, KNN, SVM and decision tree to find the best model

Data Collection

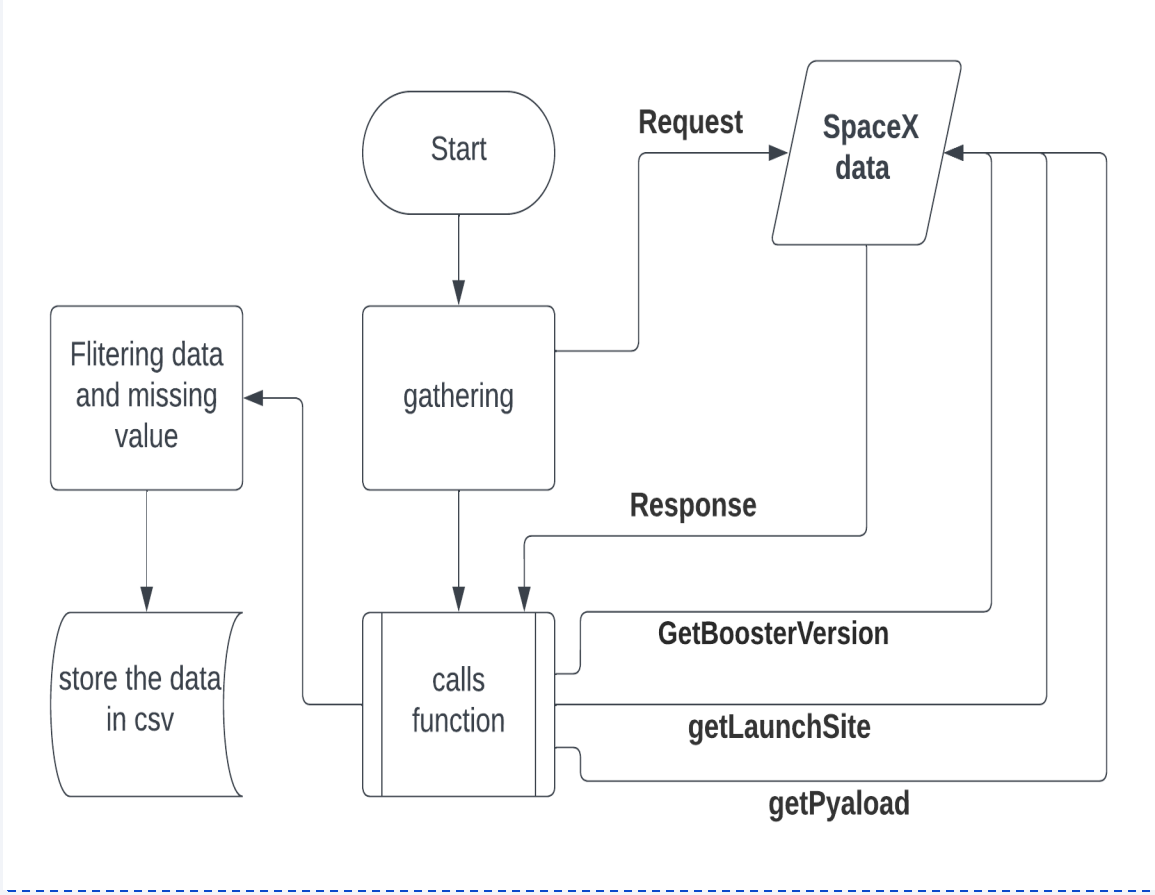
- The data was collected using the API as json and then parse in data frame it ,from SpaceX provider.
- Fetch the core of the rocket , payload and launch Site then store it in lists and make it data frame
- GetBoosterVersion() : taking data as parameter and fetch from API (name) using rocket column from data.
- getLaunchSite(): taking data as parameter and fetch from API (longitude, latitude, name) using launchpad column from data.
- getPayloadData(): taking data as parameter and fetch from API (mass_kg, orbit, name) using payloads column from data.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

Find more information in GitHub repository:

https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/spacex-data-collection-api-Part-1.ipynb



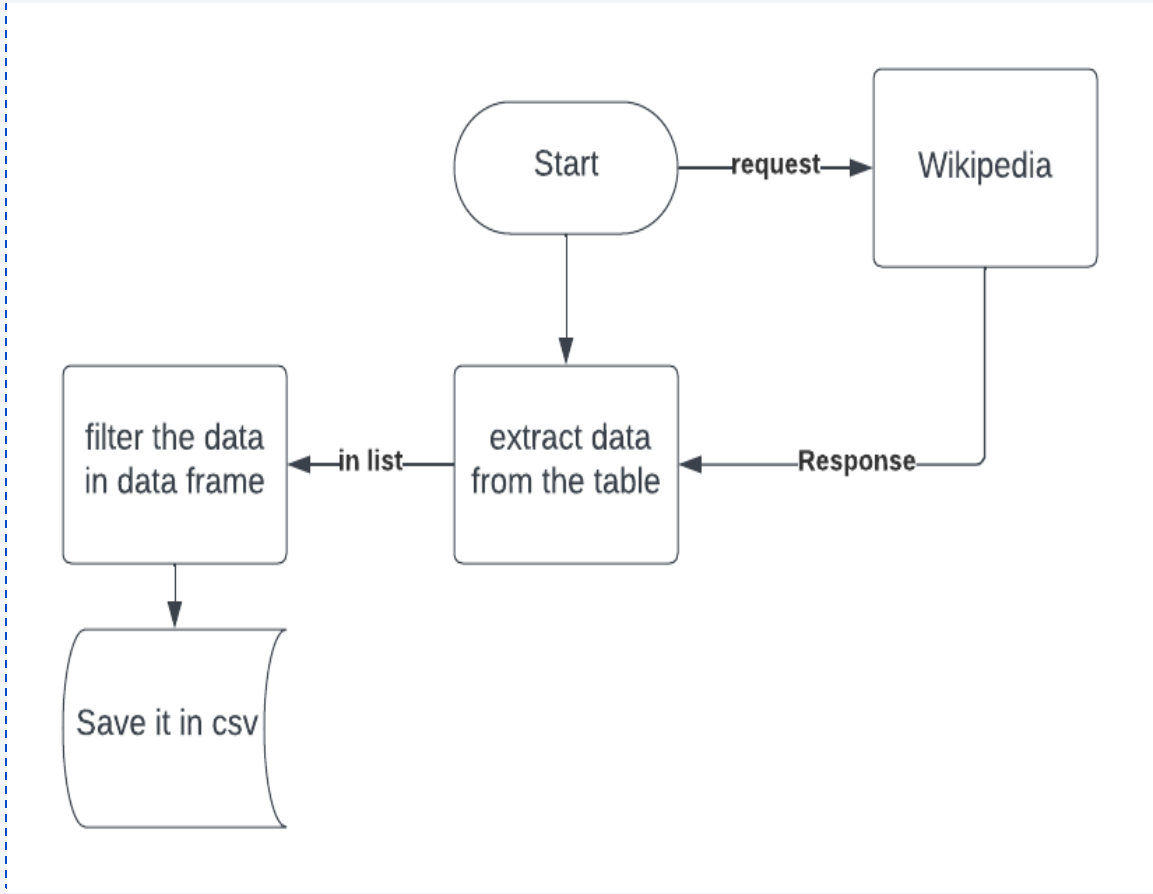
Data Collection

- The data was collected by the web scraping from Wikipedia and then parse in data frame it.
- Request the data from the table in Wikipedia using Beautifullsoup()
- Give the URL to the function and return the html. Find the table with find_all()
- And get all the cloumns to the list then make it in data frame then filter the data and save it in csv file

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Find more information in GitHub repository:
- https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/Webscraping.ipynb



Data Wrangling

- The data was cleaned after removing the null values
- Cheking from data types
- Count the number of launch site find that :
- CCSFS SLC has 40 site , 55, KSC LC 39A had 22 site and VAFB SLC 4E had 13 site
- Calculate the number and occurence of mission outcome of the orbits
- Crating class column which present if the land was successful assigned to 1, otherwise to 0 .
- Note that the success rate is 0.66.
- https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/labs-jupyter-spacex-Data%20wrangling-Part_2.ipynb

EDA with Data Visualization

- Use the scatter plot to discover if there any relation between the Payload Mass / LaunchSite / Orbit with FlightNumber.
- Apply the the function catplot() from the seaburn library
- plot to find relation between years and success rate.
- Lastly, convert the data using get_dummy() then one hot encoding for machine learning.
- Use the bar chart to check if there is relation between the success rate and orbit type

- https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/EDA-dataviz-Part_4.ipynb

EDA with SQL

- List the total number of successful and failure mission outcomes
- select Mission_Outcome, count(Mission_Outcome) as "Total of mission outcomes“ from 'SPACEXTABLE’ group by Mission_Outcome
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- select substr(Date,6,2) as month, substr(Date,0,5) as year, Booster_Version, Launch_Site from 'SPACEXTABLE' where Landing_Outcome = "Failure (drone ship)" and year = '2015'
- Display the names of the unique launch sites in the space mission
- select distinct Launch_Site from 'SPACEXTABLE'
- Display the total payload mass carried by boosters launched by NASA (CRS)

select Customer,sum(PAYLOAD_MASS__KG_) from 'SPACEXTABLE' where Customer like '%CRS%'

https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/EDA-SQL-Part_3.ipynb

Build an Interactive Map with Folium

- Creating map using Floium library with folium marker get the Latitud ,Longitude from the dataframe and put in location parmmeter in the function do display all the site with successful lanch site
- And show the distance between the coast and launch site

https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/launch_site_location-Part_5.ipynb

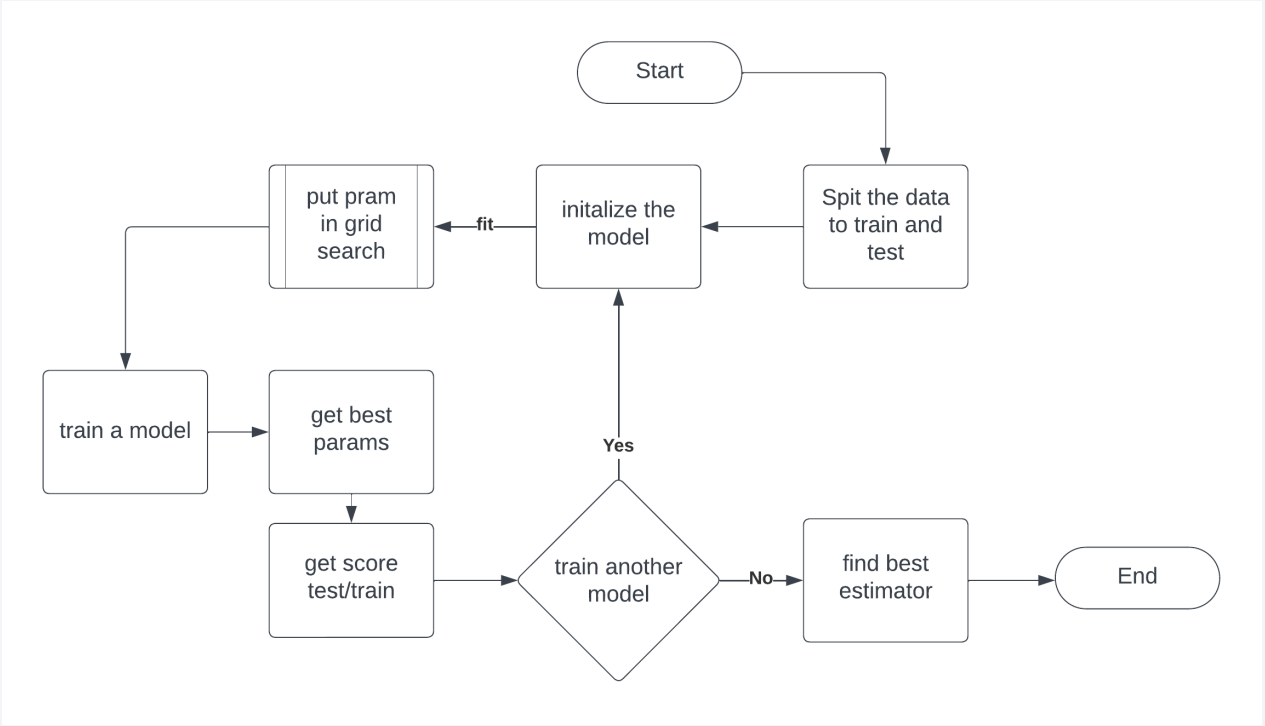
Build a Dashboard with Plotly Dash

- Use the pie chart and the scatter chart from plotly express to show the total success launches by the all sites.
- By grouping the data with class column and lancuh site we show the total success rate
- Also grouping the data with specific launch site to show the total of success rate in this site.

https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/dashboard.ipynb

Predictive Analysis (Classification)

- Using the sklearn library to use many machine learning model to find the best one to classification the data. for instance :
- Linear logistic regression, SVM, KNN, Decision tree
- Find the best parameter for each by using GridSearch.



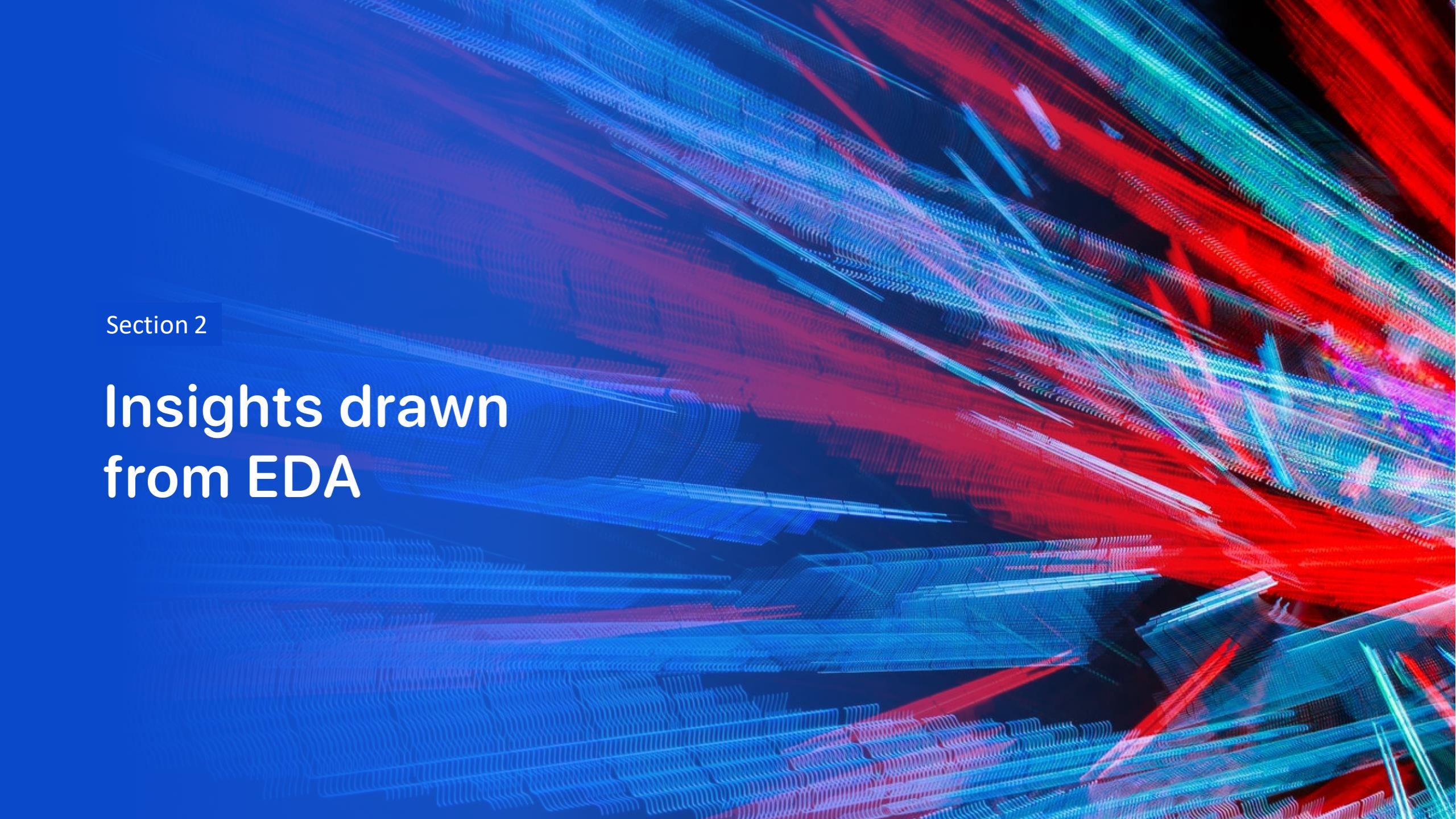
https://github.com/Sbinsuwaylih/SpaceX-Prediction-and-Analysis/blob/main/SpaceX_Machine_Learning_Prediction_Part_6.ipynb

Results

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

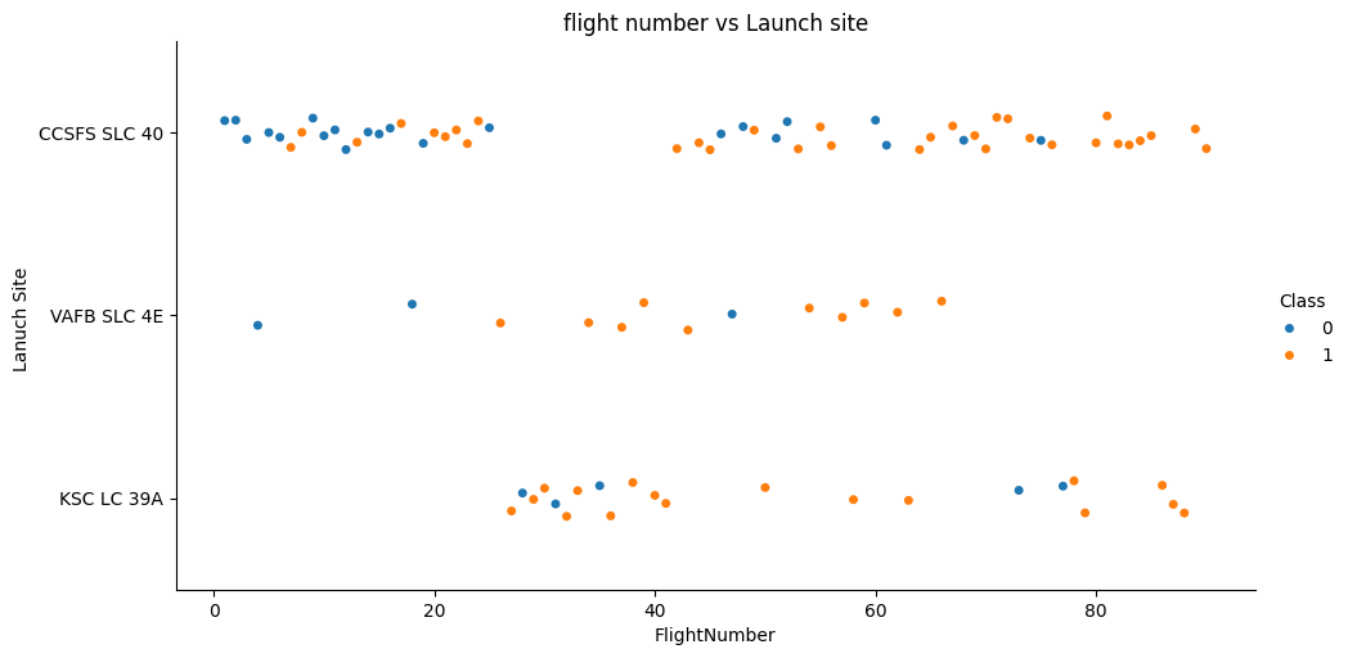
Section 2

Insights drawn from EDA



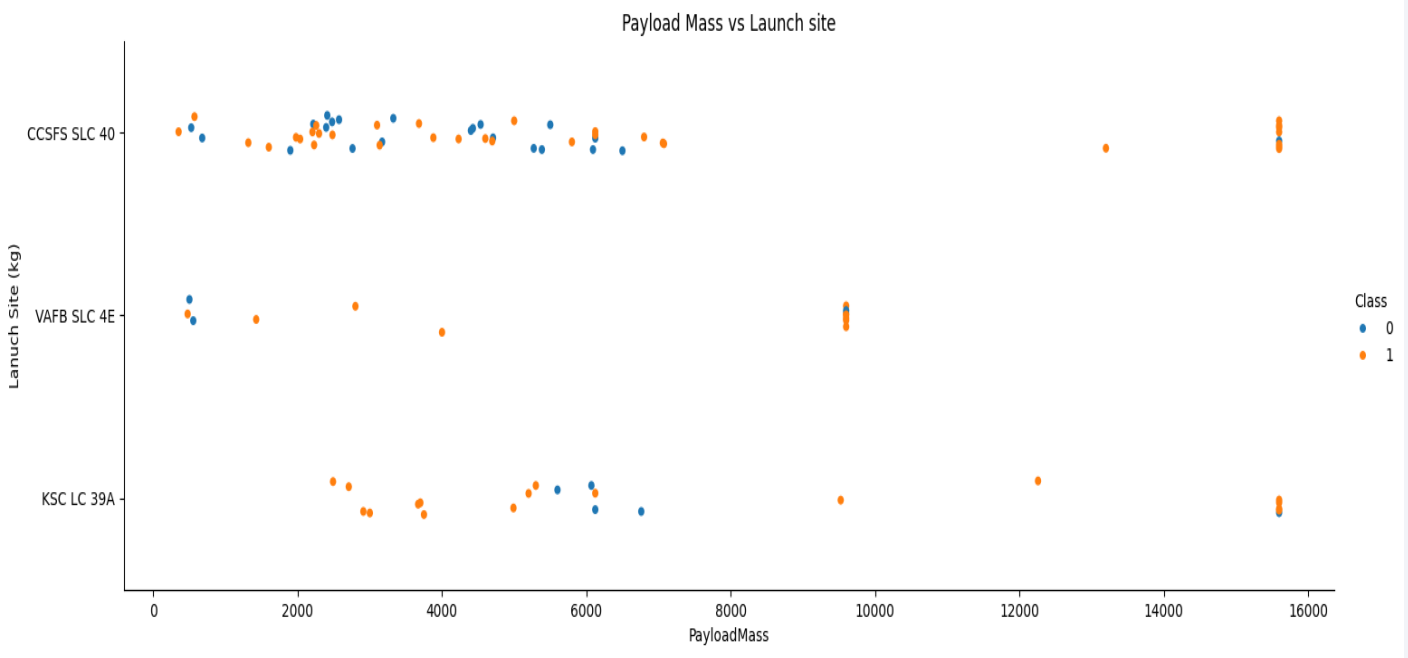
Flight Number vs. Launch Site

- This diagram represent the scatter chart between the flight number and launch site
- Observing that within the flight number increase the launch site also increase all launches site



Payload vs. Launch Site

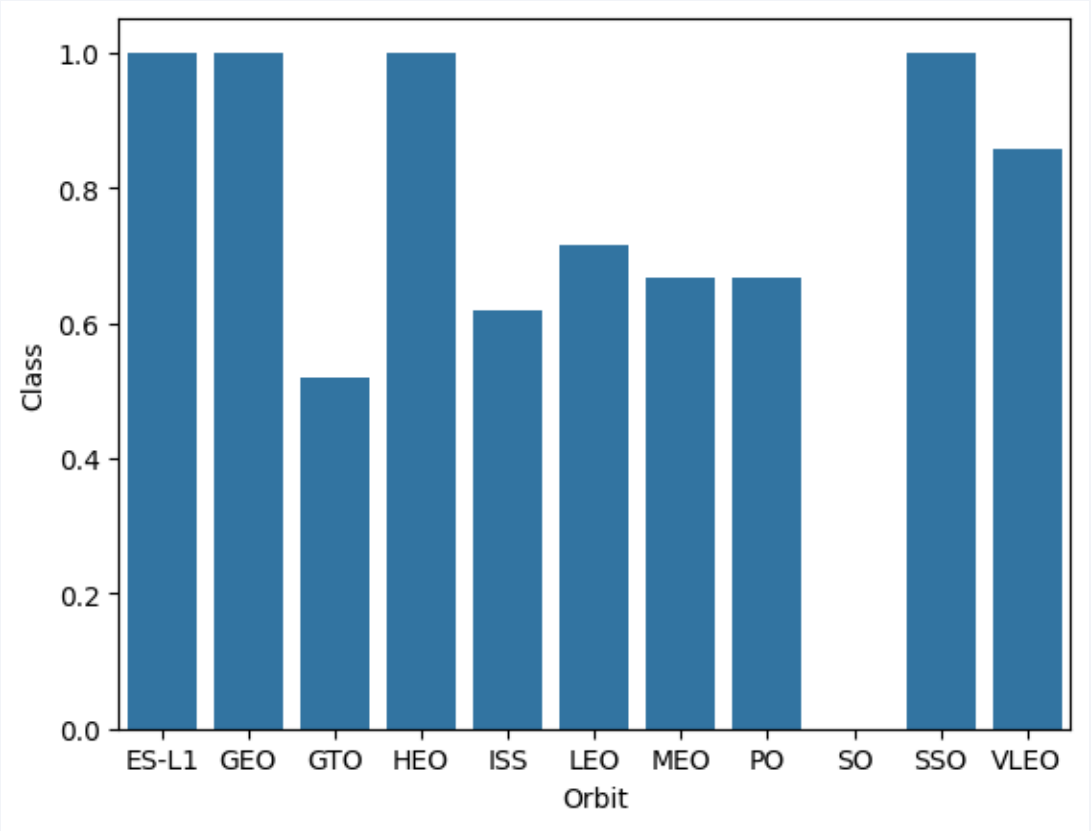
- This diagram represent the scatter chart between the Payload mass (kg)and launch site
- Observing that VAFB is not having paload mass more than 9600 unlike the others.



Success Rate vs. Orbit Type

- This diagram represents the bar chart depicting the highest success rates for all orbits.

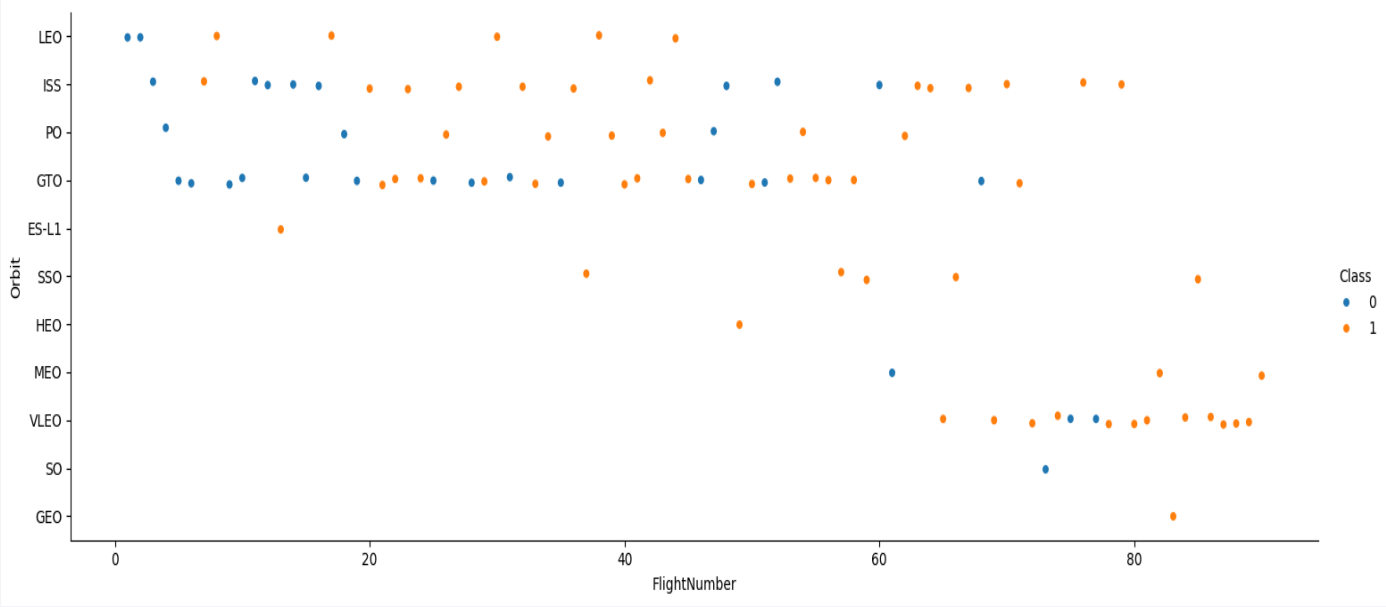
- SO has the lowest success rate, while ES, GEO, HEO, and SSO have the highest rates



Flight Number vs. Orbit Type

- This diagram represents the scatter chart depicting the relationship between the orbit type and flight number.

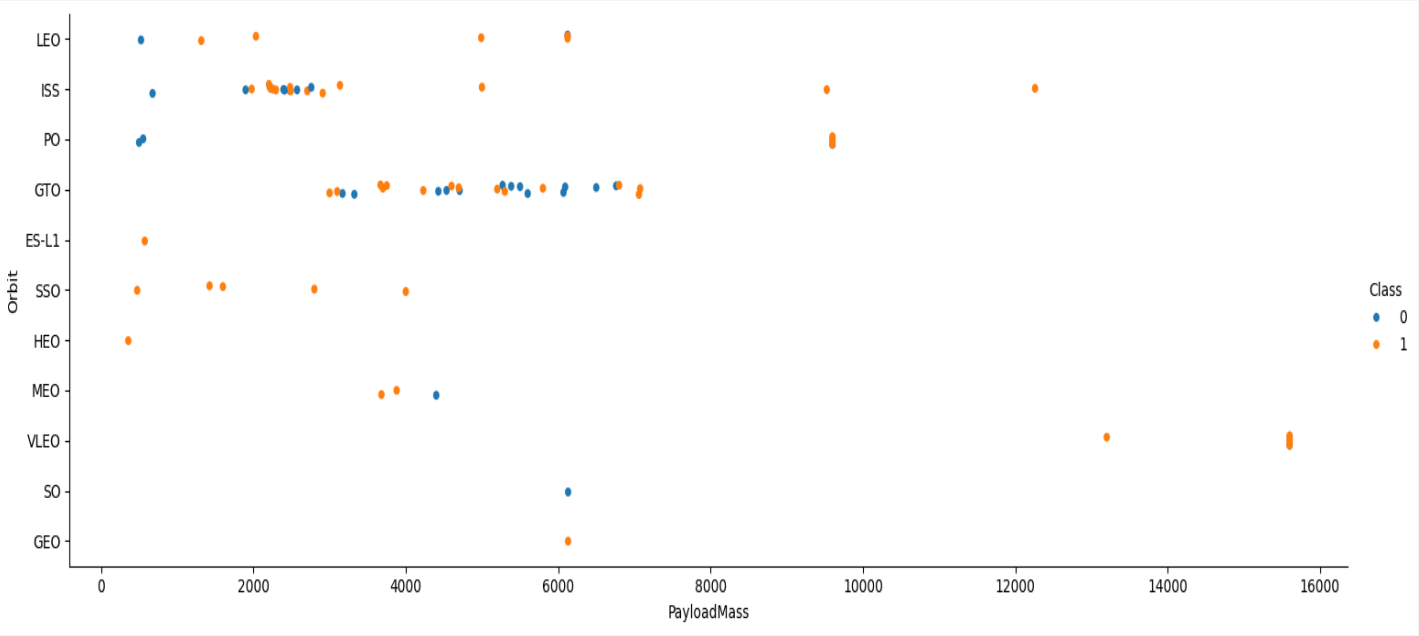
- In LEO orbit, there is a relation, while GTO exhibits an explicit correlation.



Payload vs. Orbit Type

- This diagram represents the scatter chart depicting the relationship between the orbit type and Payload mass.

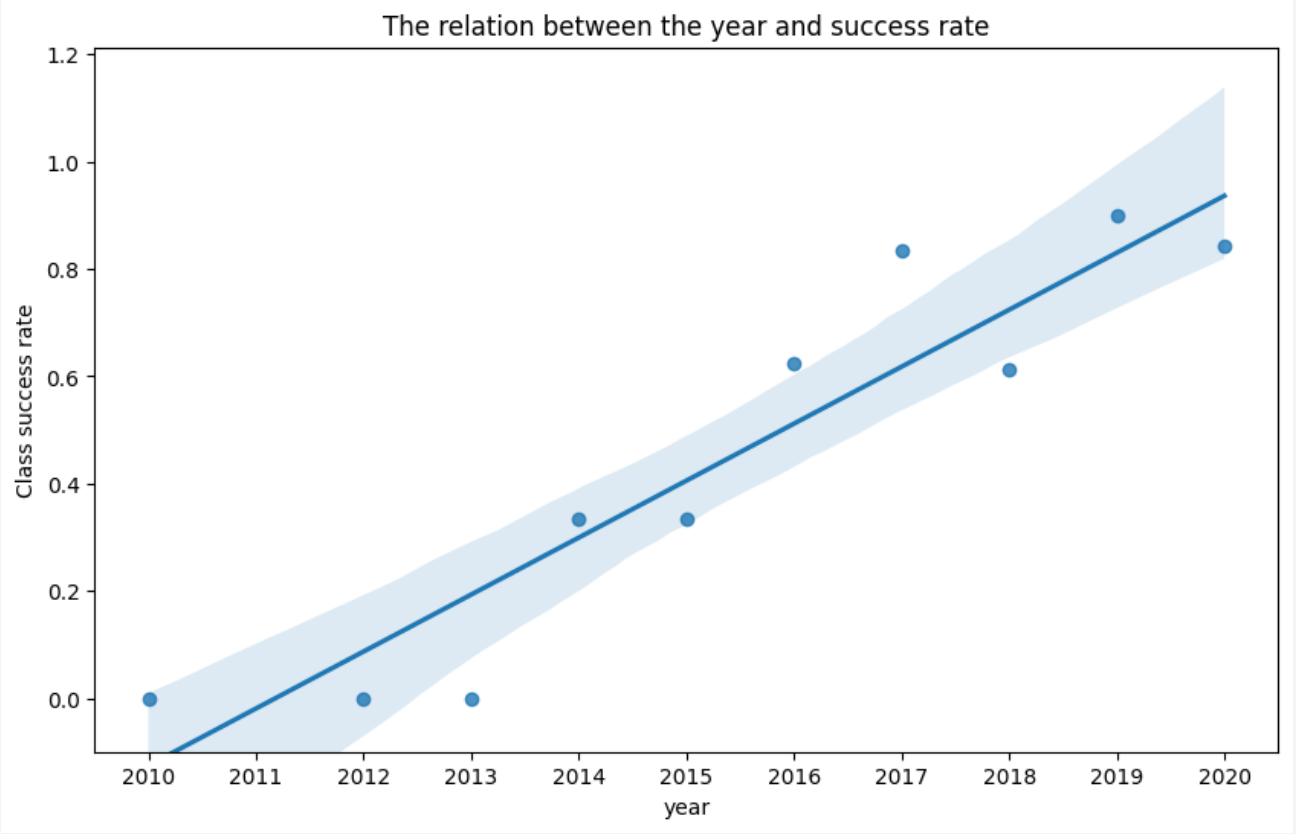
- VLEO,ISS has the largest success rate



Launch Success Yearly Trend

- This diagram represents the line chart depicting the relationship between the success rate with year.

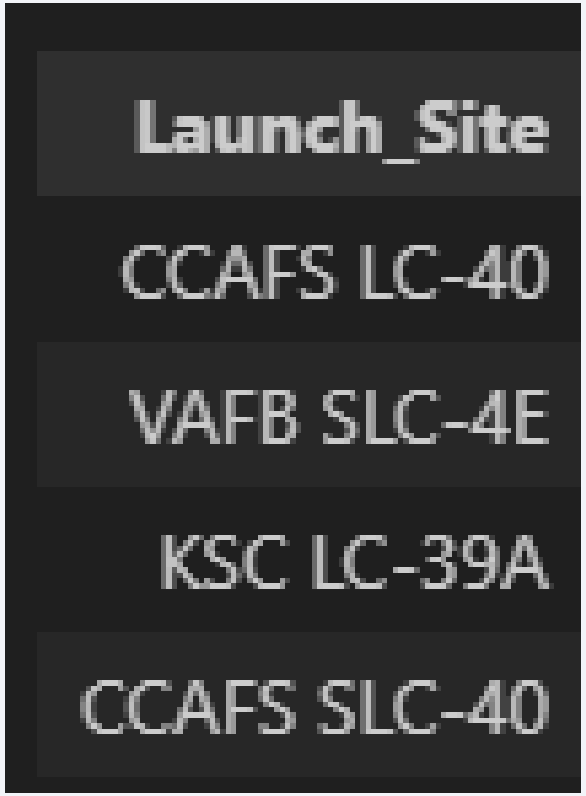
- The are positive correlation Starting from 2013 until 2020



All Launch Site Names

- This is list of all unique name in launch site from the SpaceX

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



Launch Site Names Begin with 'CCA'

- This is 5 records where launch sites begin with `CCA` from spaceX.
- Observing that mission outcome success and landing outcome id failure or not attempt and the orbit is LEO

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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
- We can see the total is 48213 kg from NASA (CRS)

Customer	sum(PAYLOAD_MASS__KG_)
NASA (CRS)	48213

Average Payload Mass by F9 v1.1

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

- The booster version F9 v1.1 B1003 the average for payload mass is 2534.67 kg.

Booster_Version	avg(PAYLOAD_MASS__KG_)
F9 v1.1 B1003	2534.6666666666665

First Successful Ground Landing Date

- This query shows the dates of the first successful landing outcome on ground pad
- Launch site in CCAF LC-40 with paload mass 2034 which is almost the average.

Date	Launch_Site	PAYLOAD_MASS__KG_	Landing_Outcome
2015-12-22	CCAFS LC-40	2034	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query shows list of names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Two of them are from customer SKY and two from SES

Booster_Version	Customer
F9 FT B1022	SKY Perfect JSAT Group
F9 FT B1026	SKY Perfect JSAT Group
F9 FT B1021.2	SES
F9 FT B1031.2	SES EchoStar

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes.
- Only one Failure in flight while there are 100 success .

Mission_Outcome	Total of mission outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

- Maximum payload mass is 15600 kg from booster F9 B5

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

2015 Launch Records

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

month	year	Booster_Version	Launch_Site
01	2015	F9 v1.1 B1012	CCAFS LC-40
04	2015	F9 v1.1 B1015	CCAFS LC-40

- We can see here in 2015 landing outcome wass failed from the same launch site.

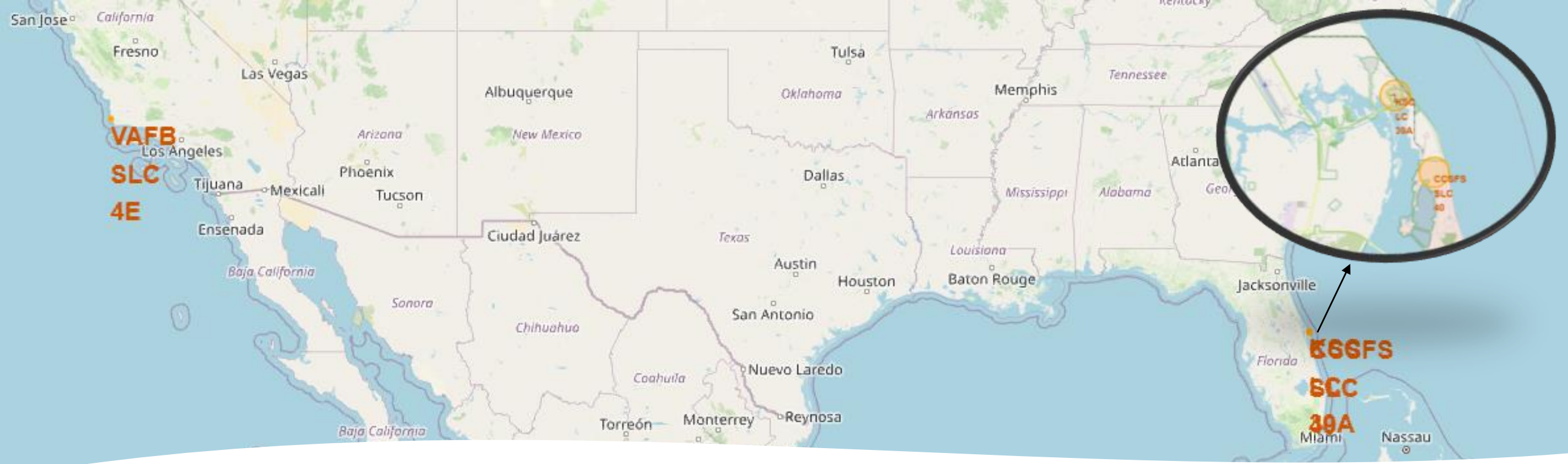
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Rank 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
- In 2012 there are a lot of no attempt landing outcome

Landing_Outcome	number_of_landing_outcome	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

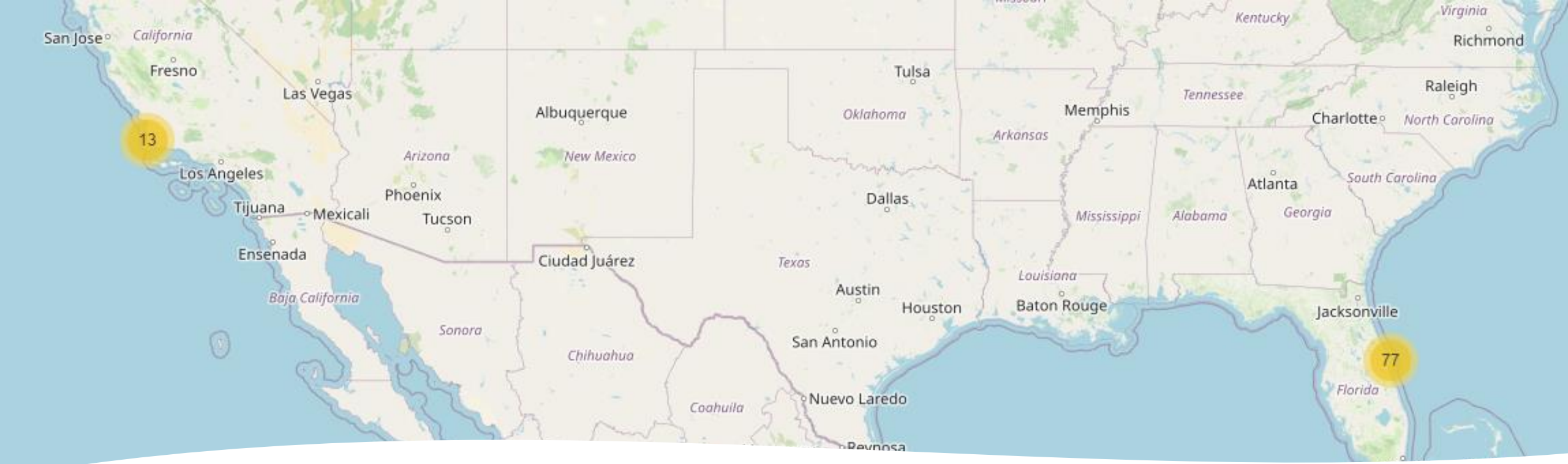
Section 3

Launch Sites Proximities Analysis



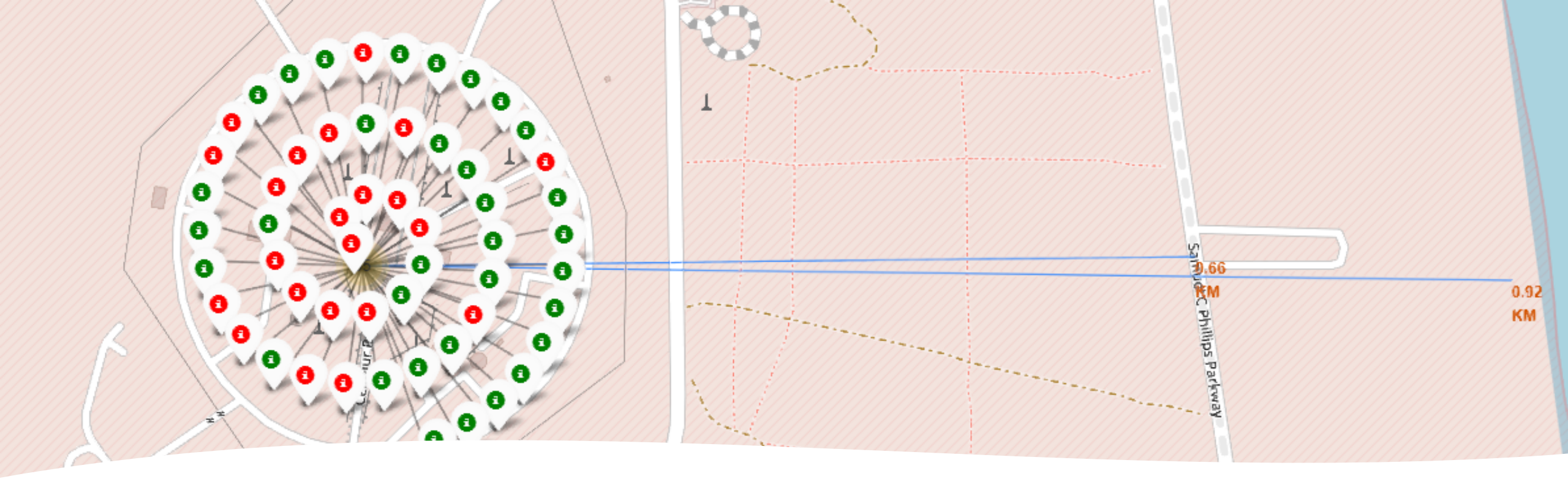
<Location of launch site in US>

- VAFB is located to the west of Los Angeles. In Florida, there are KSC and CCSFS.
- All launch sites are near the coast.



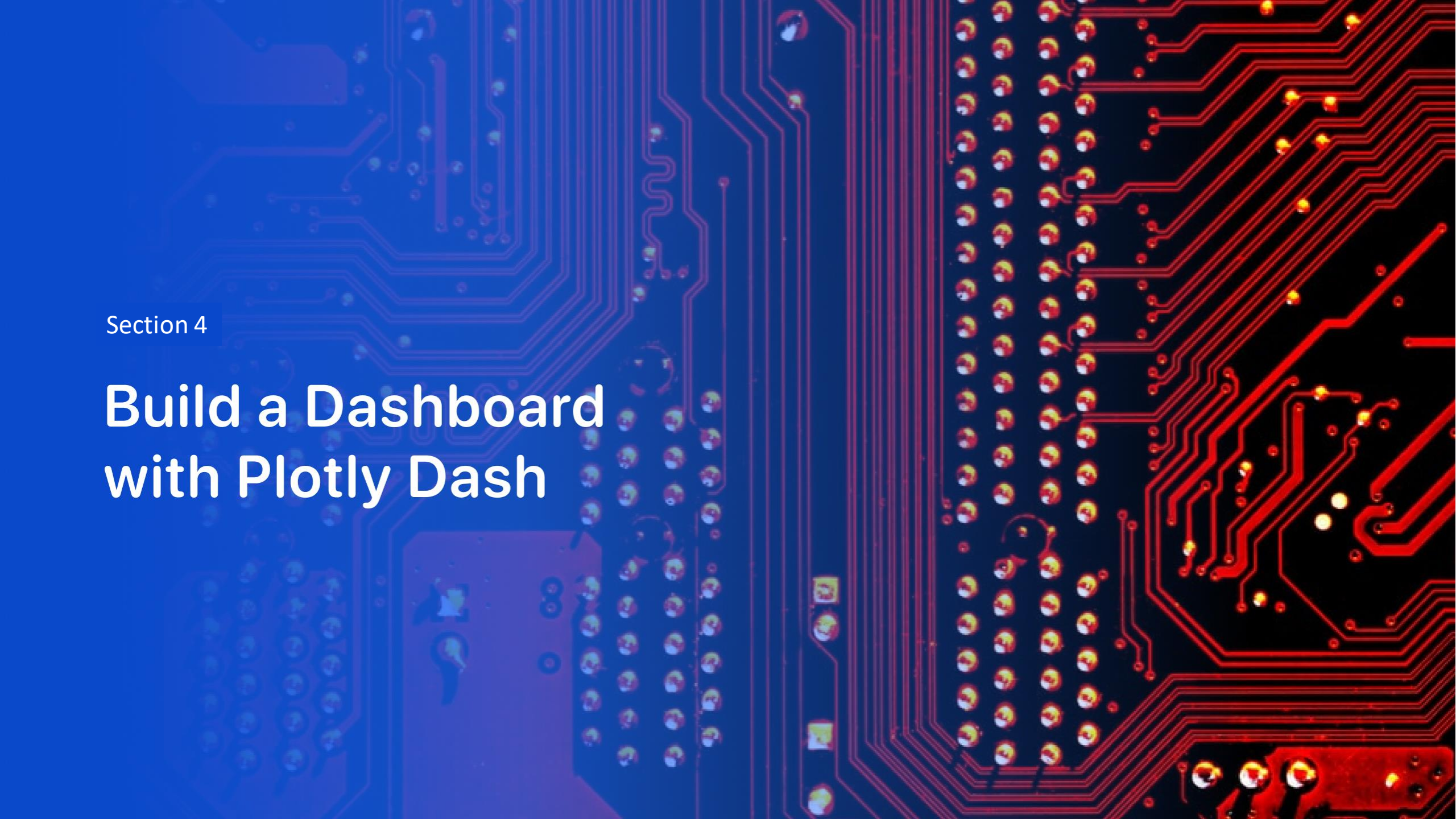
<Mark the launch site Location >

- Here create maker for each location and label the circle with number of launches.



< CCSFS lancuh site and distance from road>

- 1.Launch site Complex 40 is far from the railway by 0.6 km, whereas the other launch sites are near the railways. 2. All launch sites are close to the highways.
- 3. All launch sites are far from cities for safety

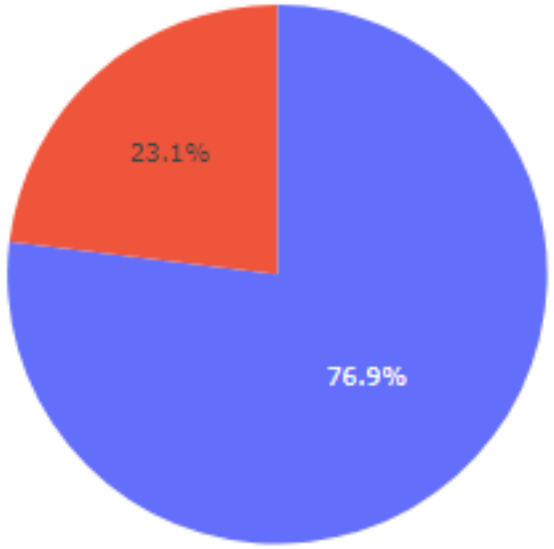

Section 4
Build a Dashboard with Plotly Dash

Total success Launches by Launch site

<Total launches>

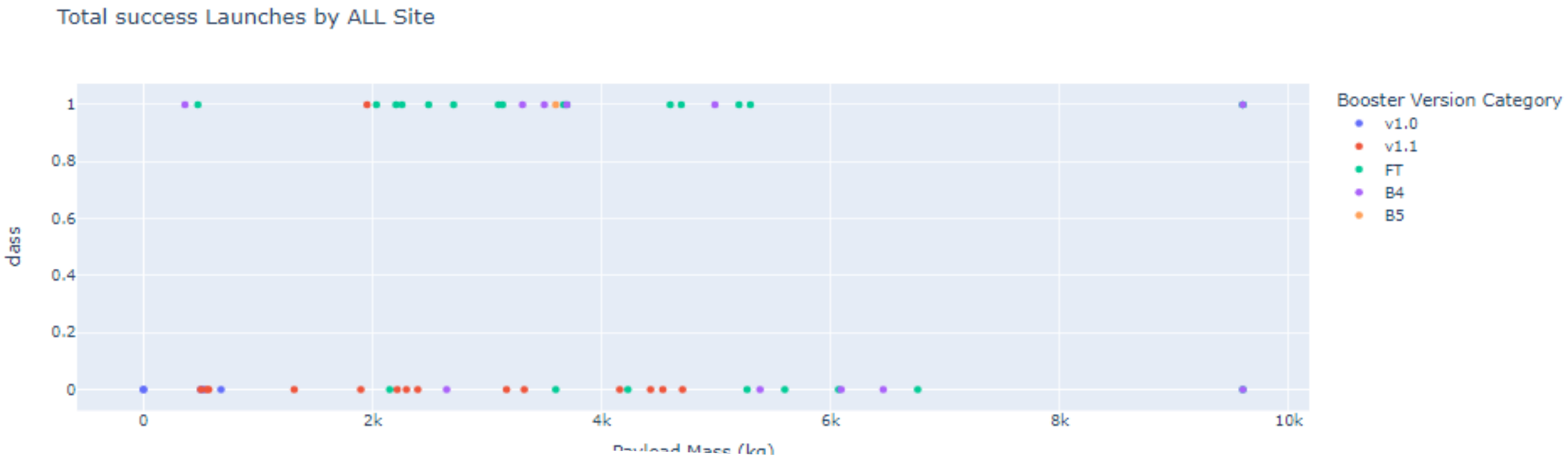
- KSC has the most success rate comparing with other launches with %41.7
- CCAFS has the lowest success rate with %12.5

Total success Launches by KSC LC-39A



<Total suceess launch by KSC>

- KSC has 76.9 success rate and 23.1 failure rate



<Total success launches by all site >

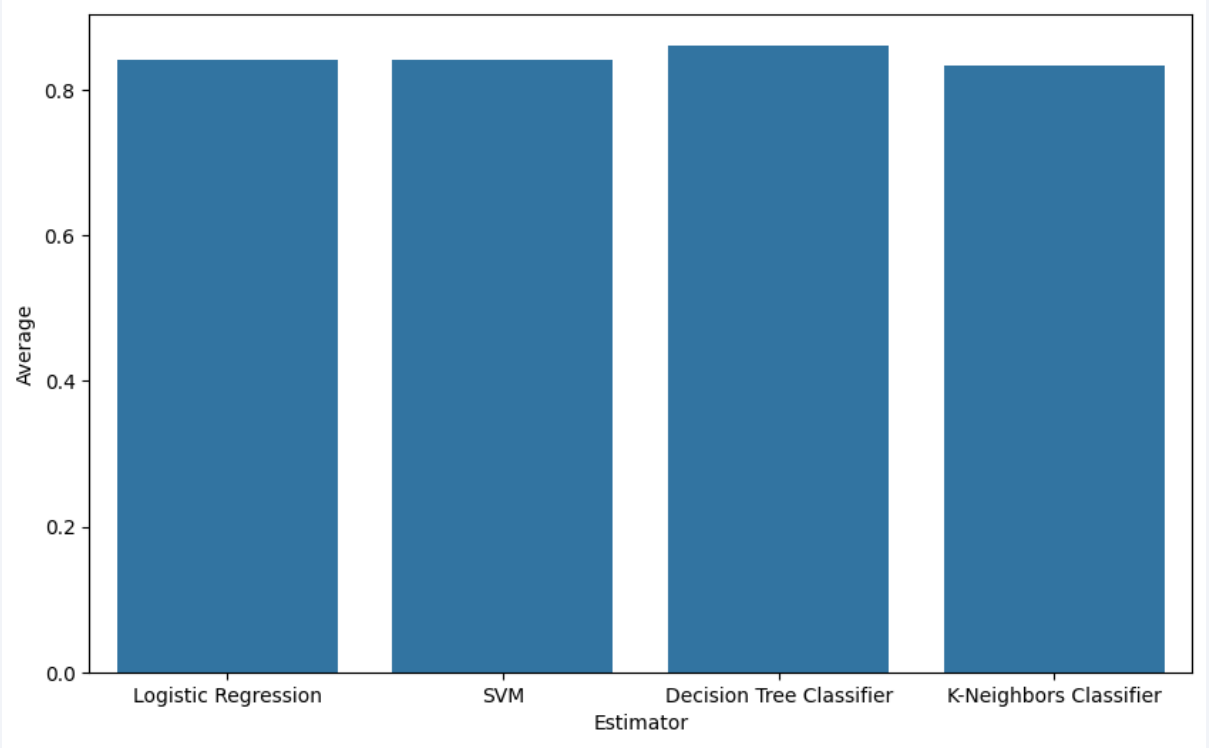
- 1. The launch site with the highest successful launch rate is **KSC LC-39A**, and it has the largest successful launch payload capacity at 38,463.65 kg
- 2. **VAFB SLC-4E** has the highest successful launch rate in the payload range(s) with a capacity of 96,000 kg. Its booster version category is **B4**. On the other hand, the lowest payload capacity is **at CCAFS SLC-40** with 362 kg.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

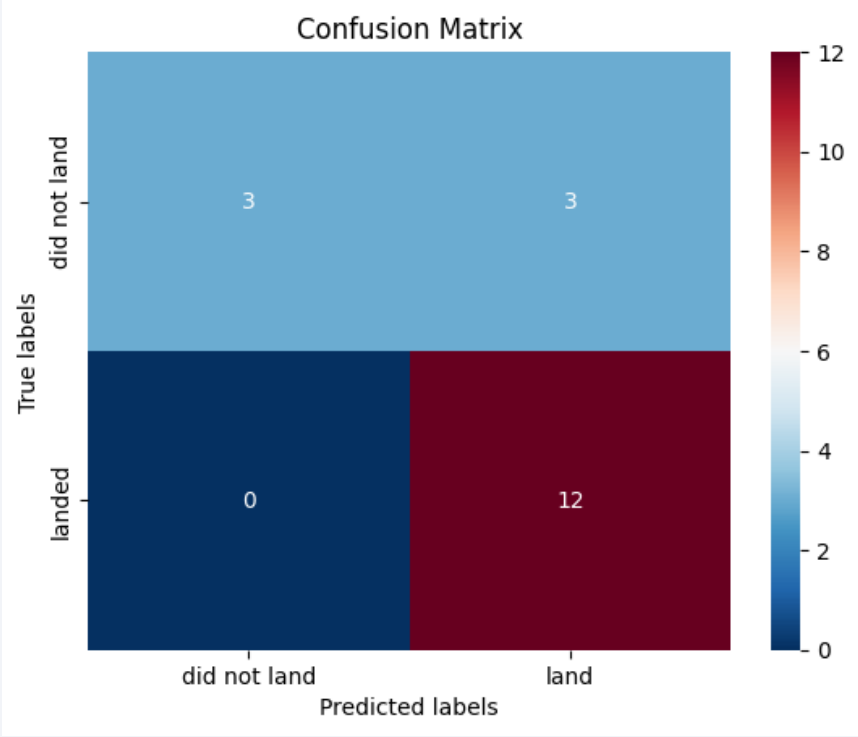
- The data trained using 4 machine learning models Logistic Regession, SVM, Decision Tree Classifier, KNN
- Random state in splitting is 2
- All model have accuracy around %80 but the highest one is Decision Tree Classifier with %86 accuracy.
- This bar chart are built with the average of test and train accuracy.



Confusion Matrix

- No misclassifying in true positive while in

True negative and false negative there are 3 misclassifying



Conclusions

- This project aims to predict whether the rocket's first stage will successfully land or not. The data was collected from the SpaceX API. The initial steps involved data cleaning and handling null values. Subsequently, a visualization of the columns with the target class was performed to identify any patterns in the data. The data was then processed to fit into machine learning models. Finally, various machine learning models were implemented to find the best one that provides the highest accuracy.
- Use the Decision tree and find %86 accuracy.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!