IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

EL AOUDI REDOUANE
12/02/2023

Outline

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

Executive Summary

- Methodologies:
  - RESTful API and web scraping are used to get necessary data
  - Exploratory data analysis like Data wrangling ,data visualization and plotly dash interactive visualization are made
  - Machine learning is used to make prediction
- Summary of all results:
  - We are able to extract the necessary features to make predictions
  - Machine learning prediction, allowed to make a best prediction of the Falcon 9 first-stage landings success.

Introduction

- The cost of Falcon 9 rocket launches of SpaceX is 62 million dollars, while other competitors cost upward of 165 million dollars each, because SpaceX can reuse the first stage.
- Our goal is to determine if the first stage will land successfully, and so, we can determine the cost of a launch. This information can be used by SpaceX's competitors for a rocket launch,
- We will predict the landing success of Falcon 9 first stage of SpaceX

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected from SpaceX Api and webscrapping from wikipedia
- Perform data wrangling
  - We converted outcomes into training labels with 1 means the booster successfully landed and 0 means it was unsuccessfully landed
- 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 standardized the data, split it into training data and test data, search best hyperparameter for SVM, Classification Trees and Logistic Regression and choose the best model

Data Collection

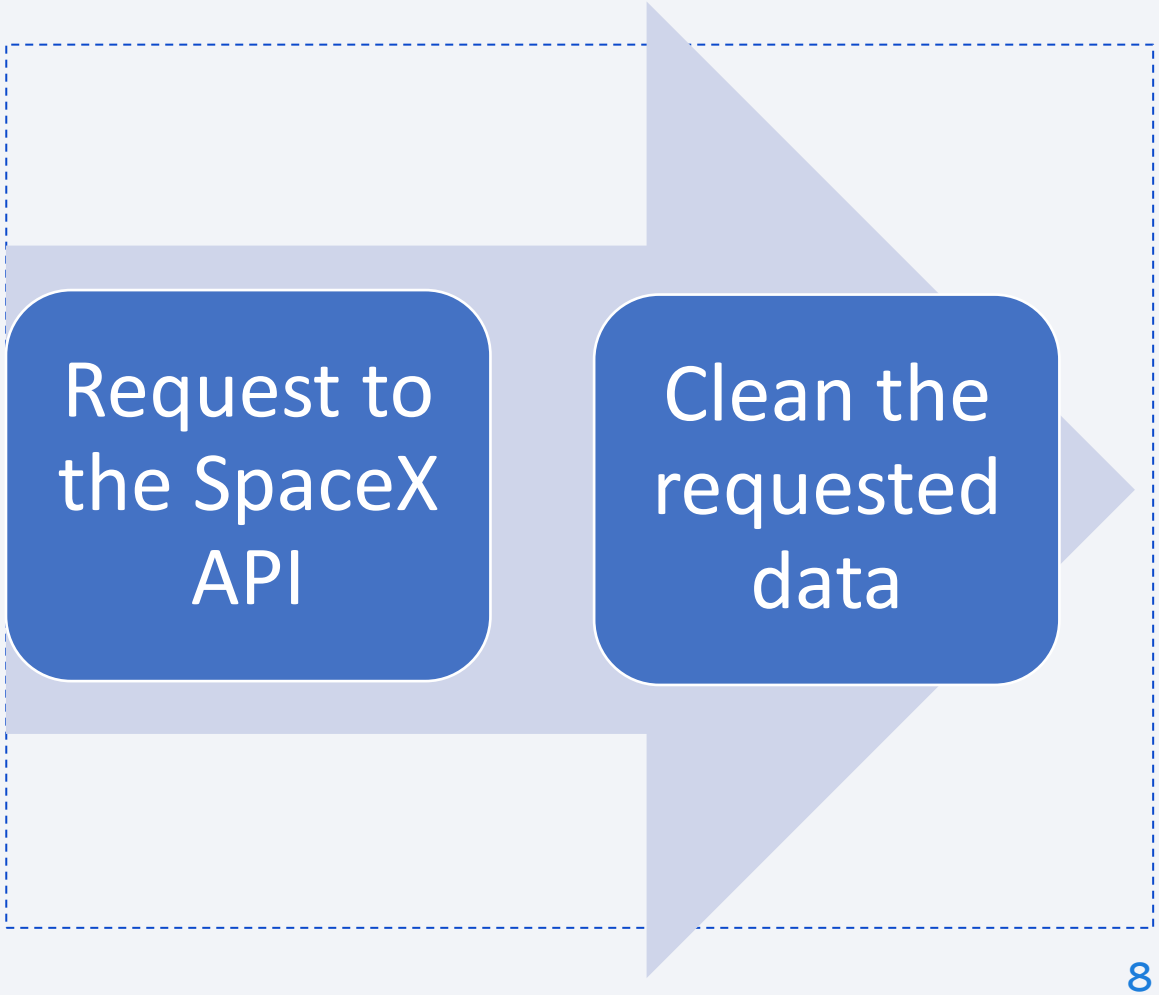
- Data sets were collected by using RESTful API from SpaceX website and using webscaping from wikipedia website

Data Collection – SpaceX API

- SpaceX API was used to collect data and some manipulations applied to cleaning data (see flowchart)

- Source code:

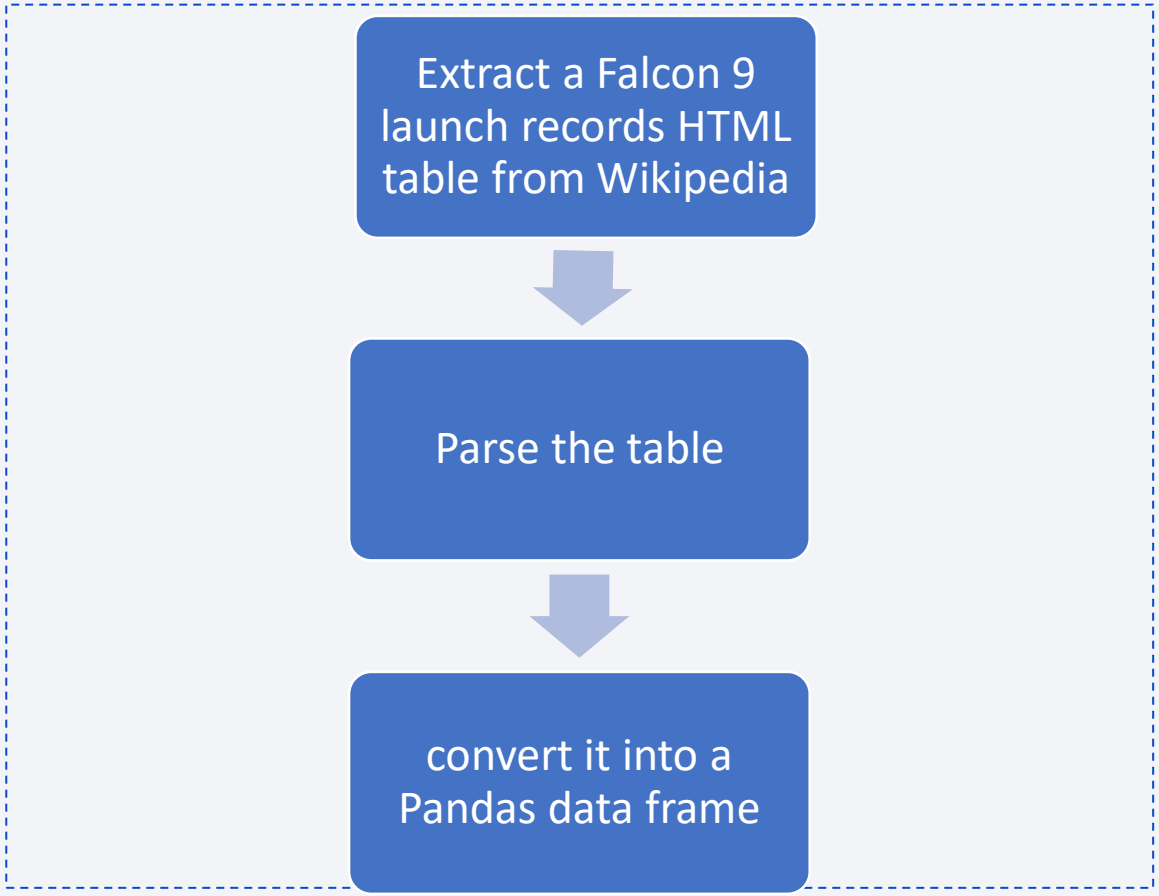
https://github.com/relaoudi/DataScienceCapstone/blob/b0895f1649e28bd86bc3e0cdde2a31a9e1aff2fd/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

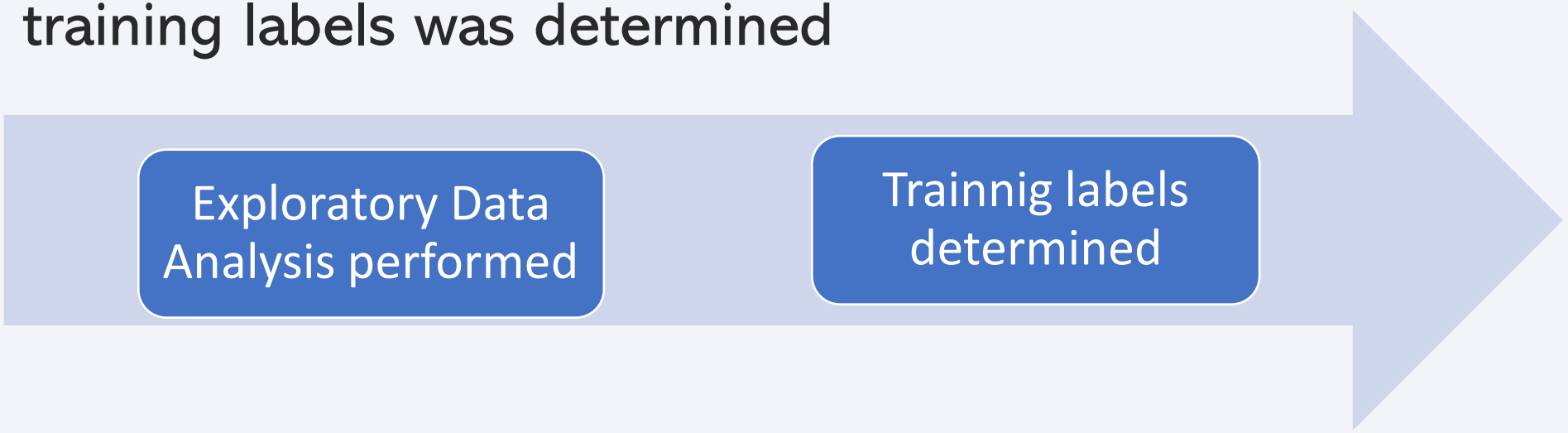
- We have performed a web scraping to collect Falcon 9 historical launch records from a Wikipedia
- Source Code:

https://github.com/relaoudi/DataScienceCapstone/blob/b0895f1649e28bd86bc3e0cdde2a31a9e1aff2fd/jupyter-labs-webscraping.ipynb



Data Wrangling

- Exploratory Data Analysis has been performed and training labels was determined

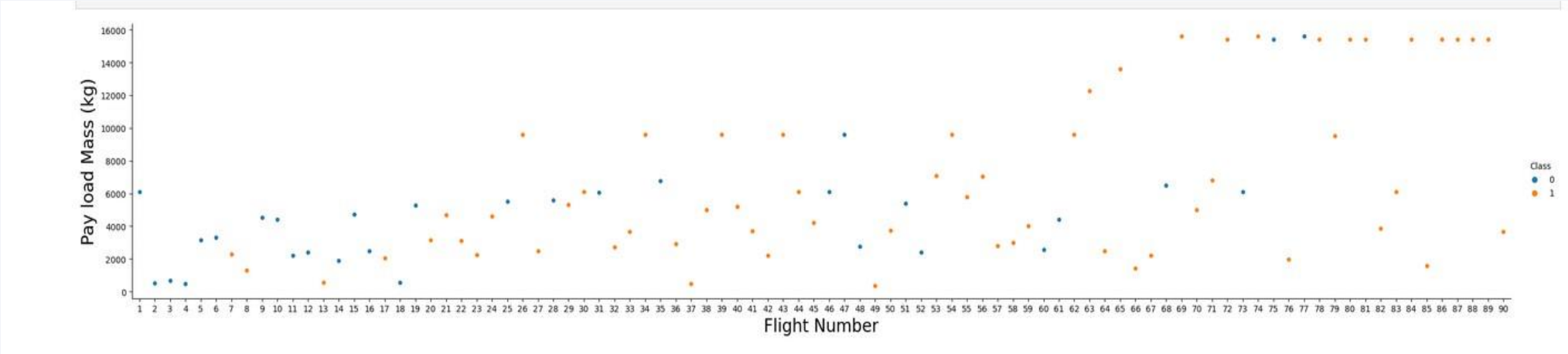


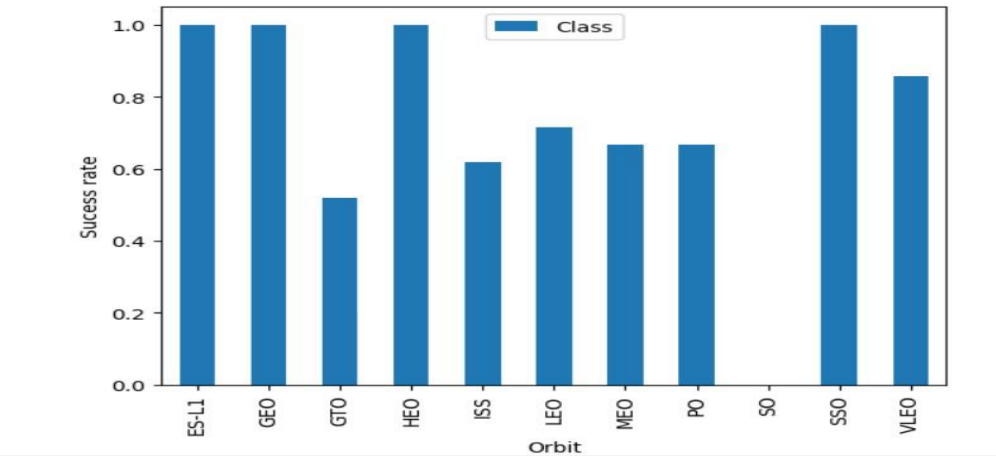
Code source:

https://github.com/relaoudi/DataScienceCapstone/blob/b0895f1649e28bd86bc3e0cdde2a31a9e1aff2fd/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Exploratory Data Analysis and Feature Engineering are used to determine the relationship between features and some interesting graphs was created like this:





For more details see the following file:
https://github.com/relaoudi/DataScienceCapstone/blob/b0895f1649e28bd86bc3e0cdde2a31a9e1aff2fd/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

the SQL queries has performed like following:

- ❑**Display the names of the unique launch sites in the space mission**
- ❑**Display the total payload mass carried by boosters launched by NASA (CRS)**
- ❑**Display average payload mass carried by booster version F9 v1.1**
- ❑**List the date when the first succesful landing outcome in ground pad was acheived.**
- ❑**And more -see the following file:**

https://github.com/relaoudi/DataScienceCapstone/blob/b0895f1649e28bd86bc3e0cdde2a31a9e1aff2fd/jupyter-labs-eda-sql-coursera_sqllite%20.ipynb

Build an Interactive Map with Folium

We created a map objects such as markers, circles, lines, and markers clusters in a folium map:

- Markers define a sites on map like launch sites
- Circles define areas around a specific point
- Lines define a distance between points
- Markers clusters define a set of events in specific point
- Explain why you added those objects

Code source:

https://github.com/relaoudi/DataScienceCapstone/blob/11e1b913ca2a07ae9c2e63c371cc69018f18cfa6/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite%20.ipynb

Build a Dashboard with Plotly Dash

Our dashboard consists of:

- Pie-chart to show a percentage of success and fail for evey site launching
- Scatter plot betwen payload mass and success for every site launching chosen in pie-chart related to booster version category

It allows to determine for each launch site the best booster for a given load

See file:

https://github.com/relaoudi/DataScienceCapstone/blob/80d119fa8fd87e25c3f471ca1eed4712964a23ad/spacex_dash_app.ipynb

Predictive Analysis (Classification)

After features normalization and data splitting into training data and test data, four classification models was studied and compared, these are the following models:

1. Logistic Regression
2. Svm
3. DecisionTree
4. Kneighbors

See file: https://github.com/relaoudi/DataScienceCapstone/blob/80d119fa8fd87e25c3f471ca1eed4712964a23ad/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

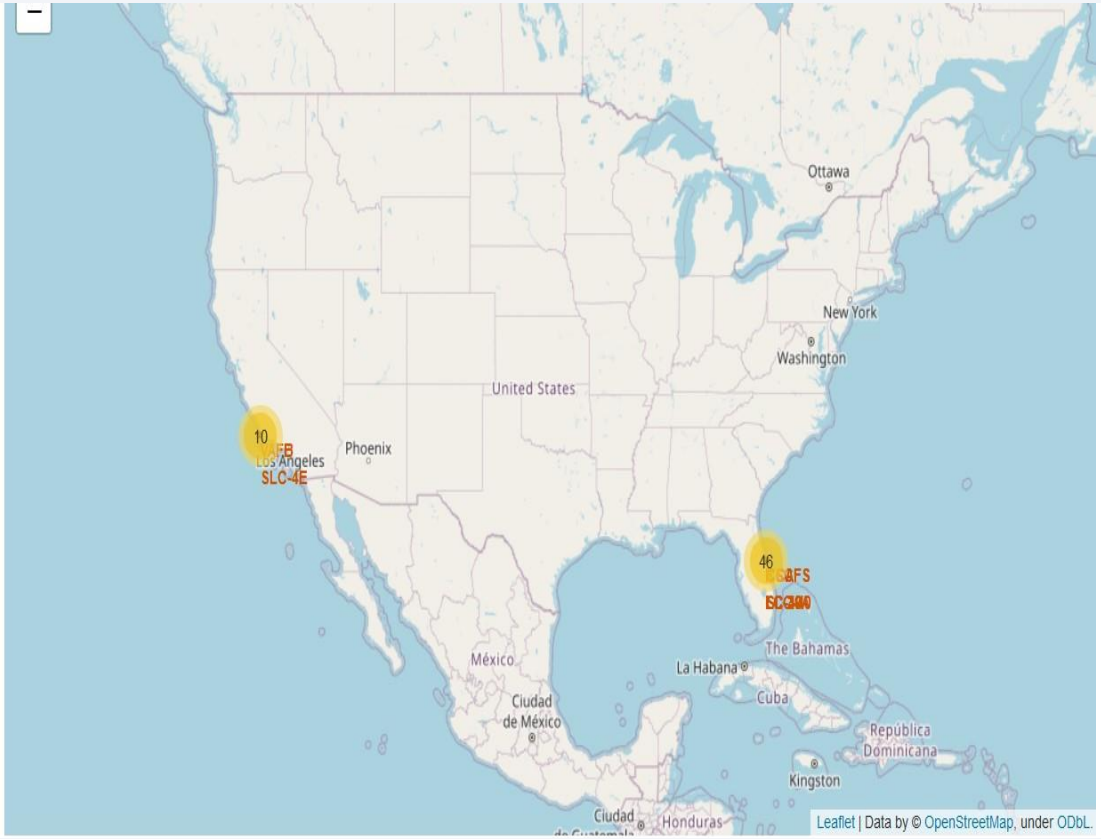
Results

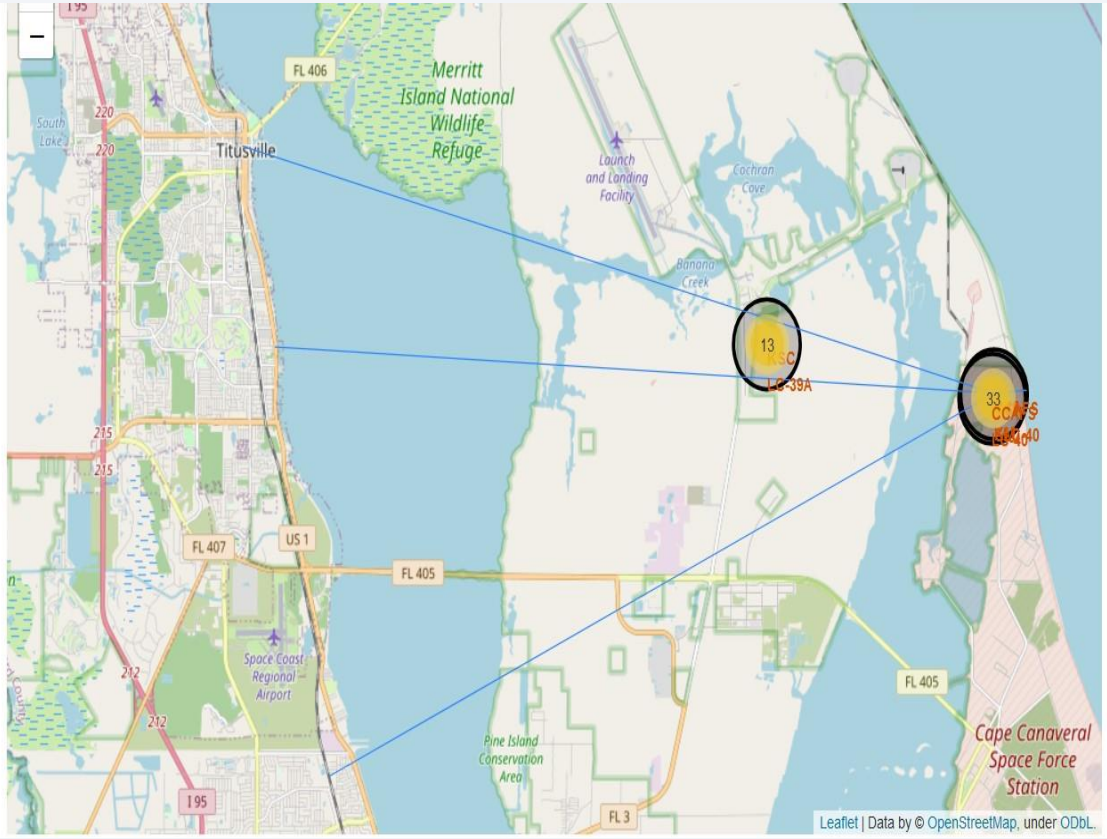
Exploratory data analysis results:

- We see that different launch sites have different success rates
- Orbits with high sucess rate are ES-L1,GEO ,HEO and SSO
- In LEO orbit the Success appears related to the number of flights;
- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- The sucess rate since 2013 kept increasing till 2020
- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg
- the first succesful landing outcome in ground pad was achieved in 2017

Results

the interactive analysis made it possible to locate the launchsites at the level of the coast far from the cities and the presence of an important infrastructure (train, highway ,ect,)

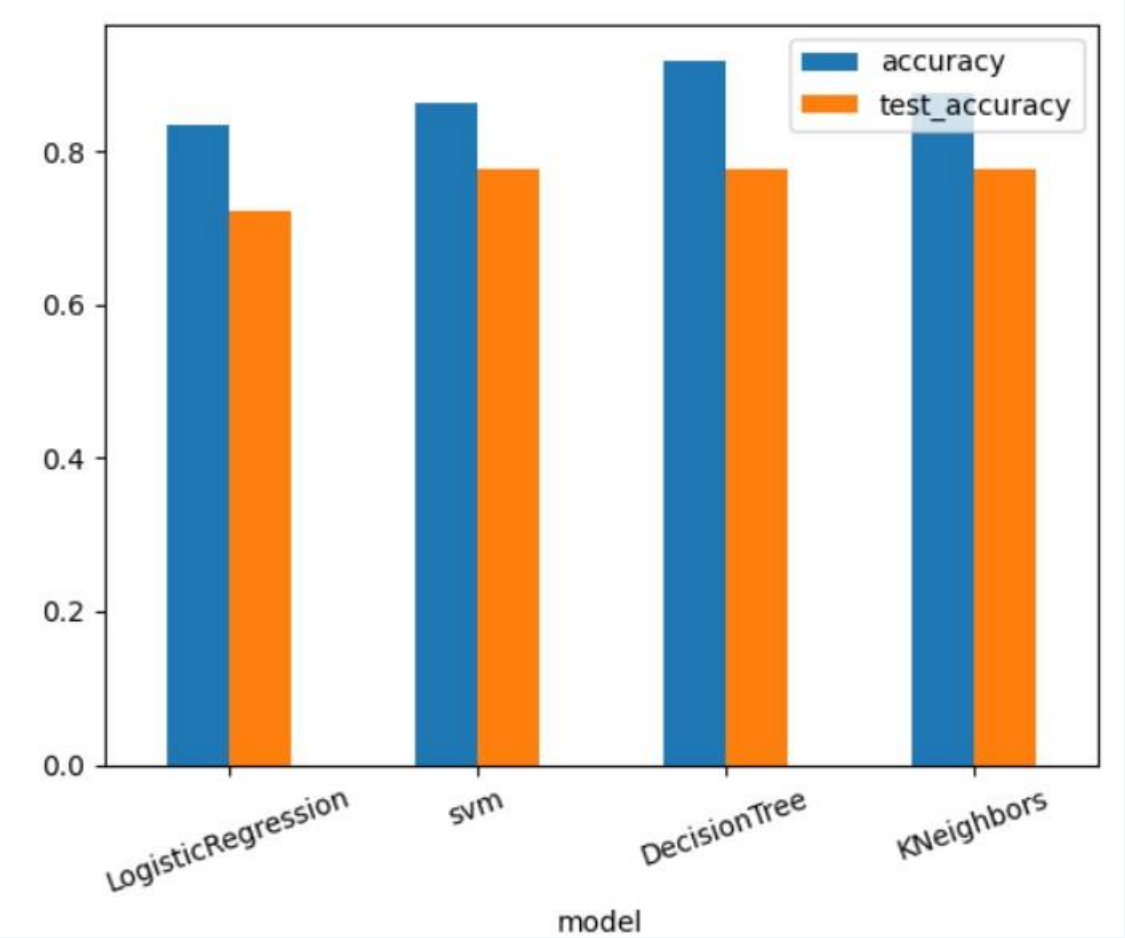


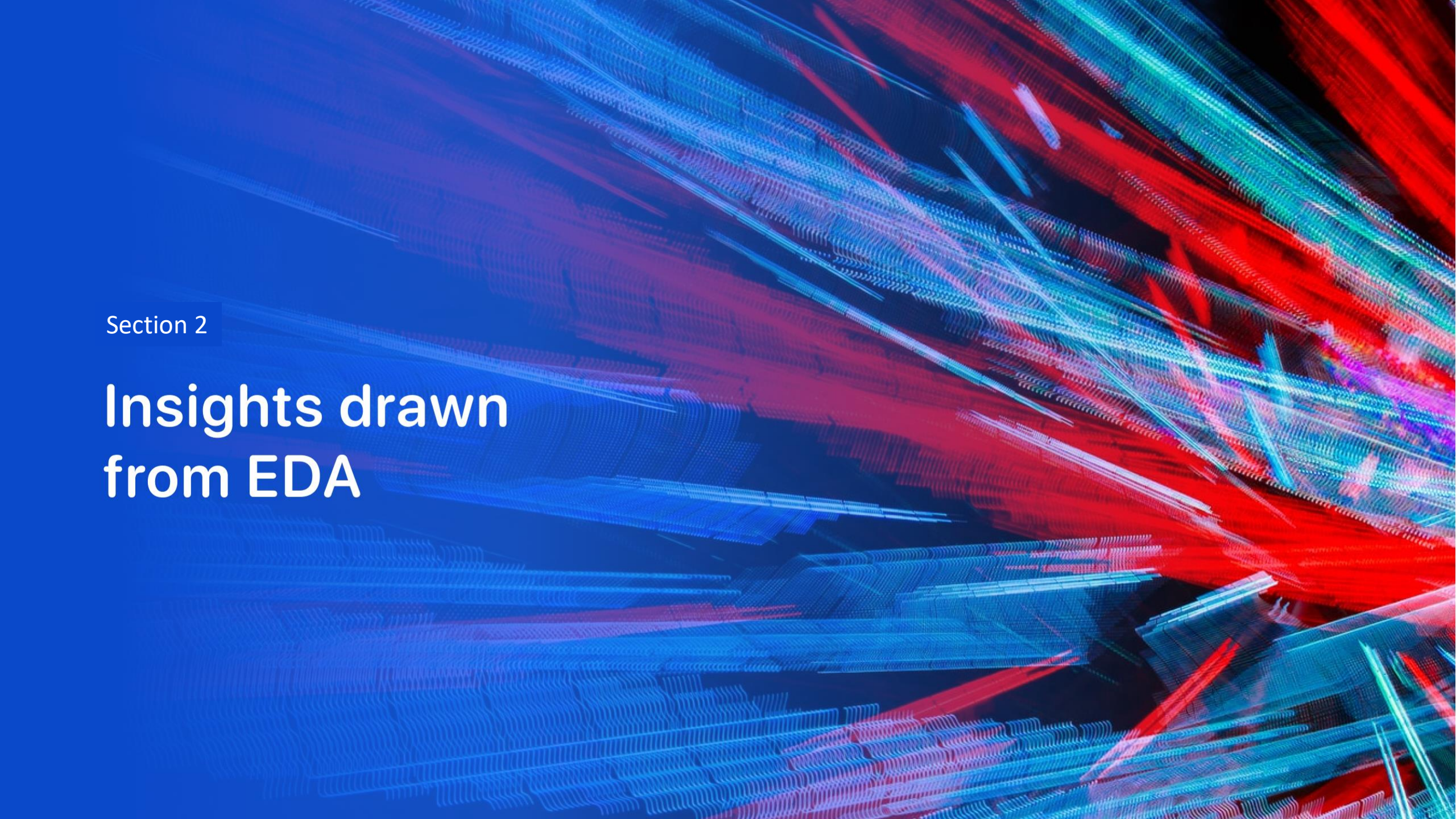


Results

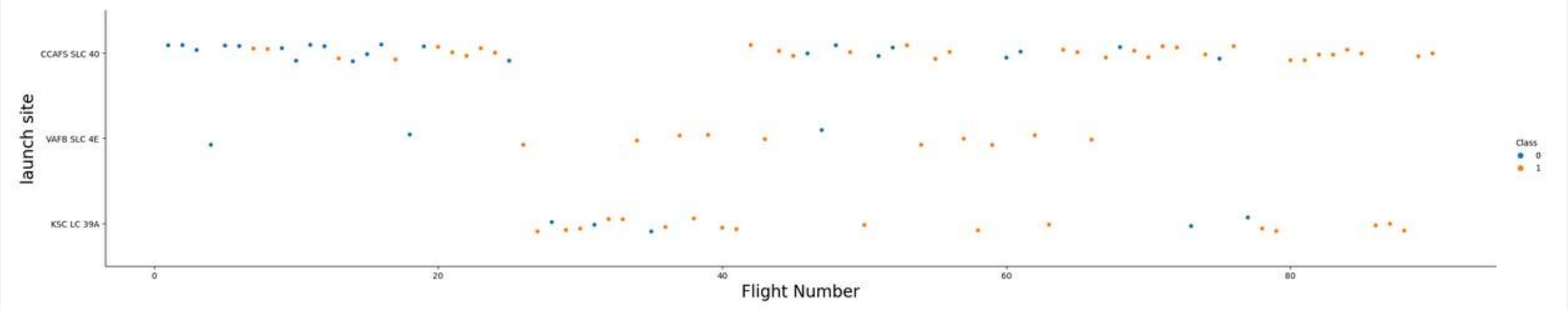
- Predictive analysis has shown that Decision Tree is the best model with high score

model	accuracy	test_accuracy
LogisticRegression	0.835	0.722
svm	0.862	0.777
DecisionTree	0.919	0.777
KNeighbors	0.876	0.777



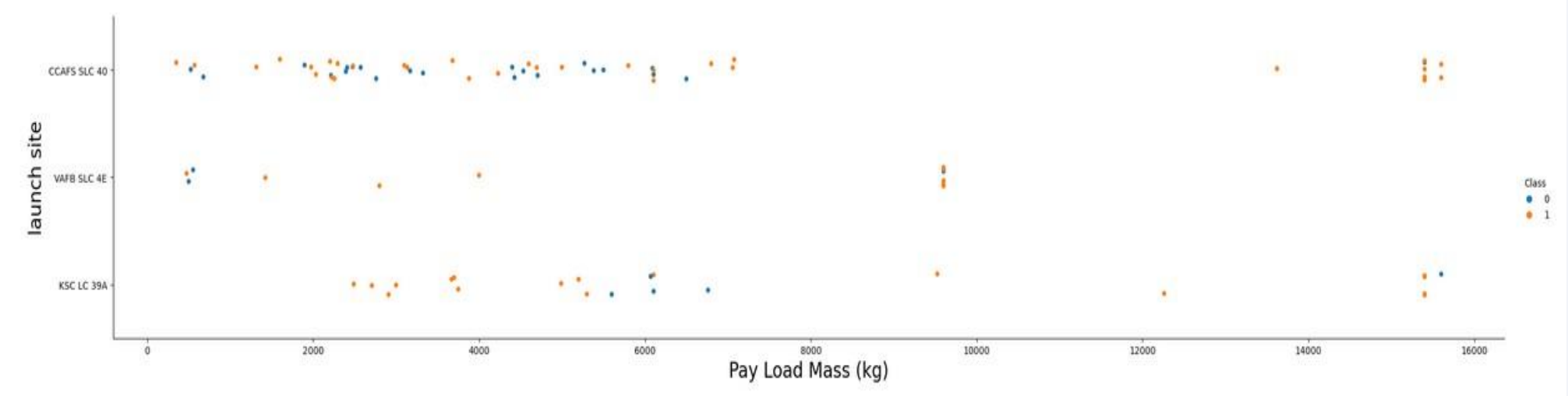

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



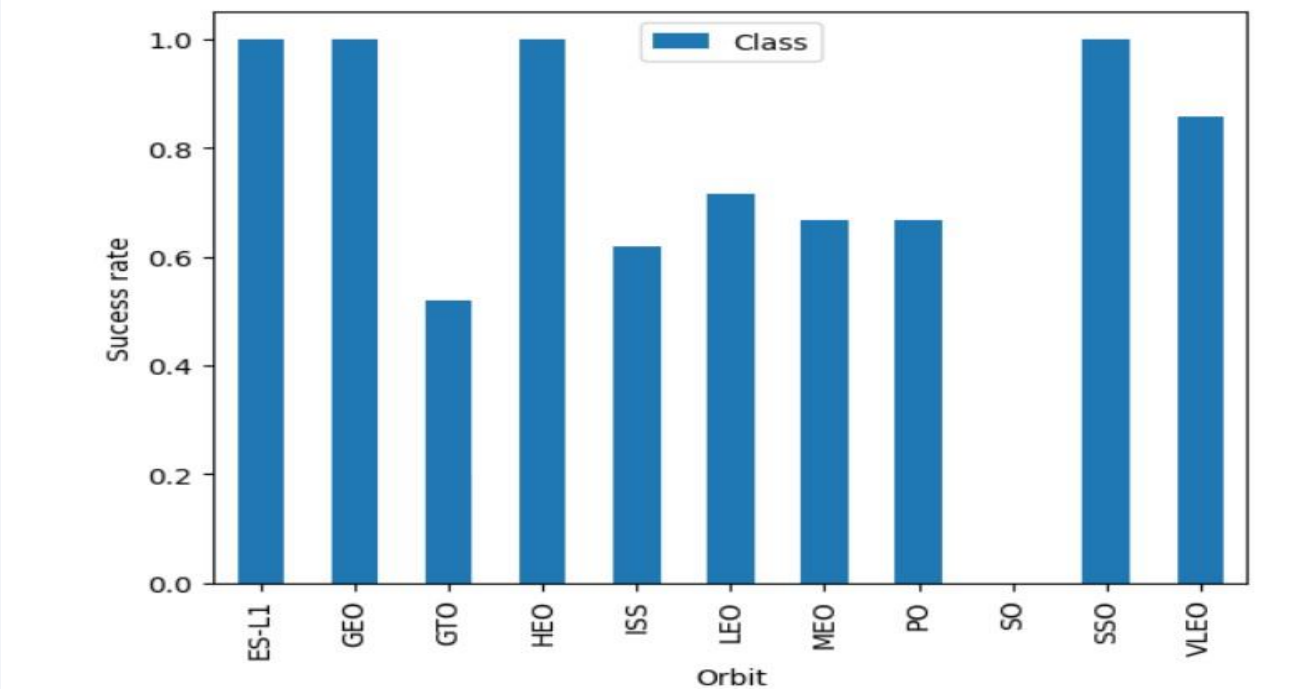
- The success improved with flight number increase
- The maximum number of launches was in site CCAF5 SLC 40 where most of recent launches were successful

Payload vs. Launch Site



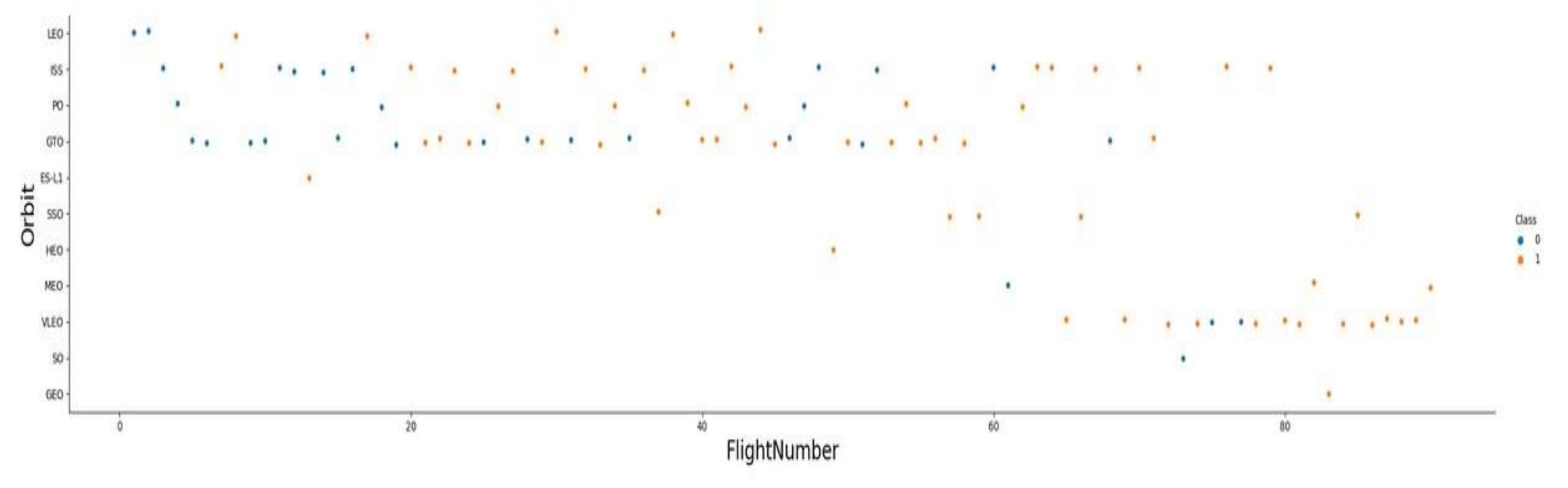
- Payload less than 6000 kg has more success in site KSC LC 39A
- Payload greater than 14000 has almost 100% of success in other launch sites than KSC LC 39A

Success Rate vs. Orbit Type



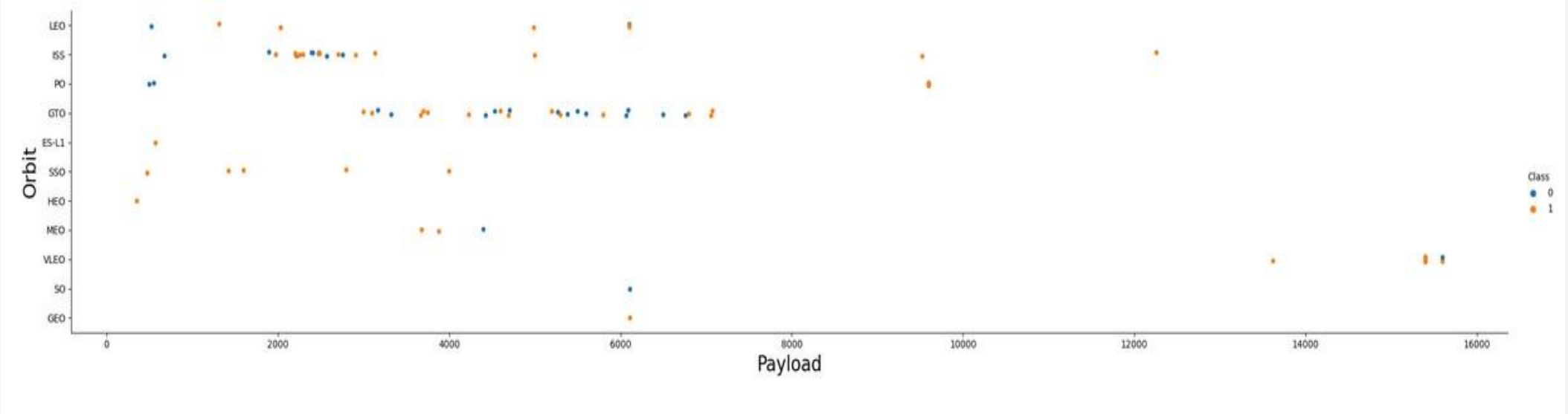
- The high success rate(100%) is in orbits ES-L1,GEO, HEO and SSO

Flight Number vs. Orbit Type



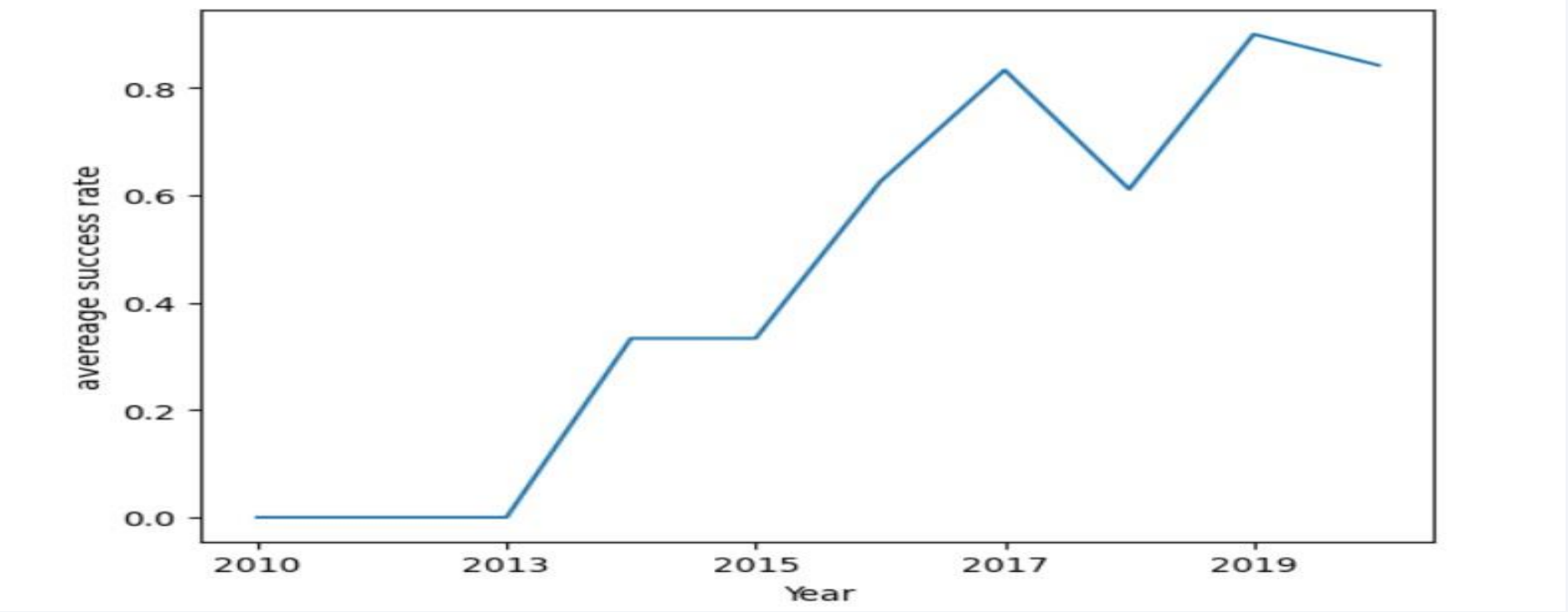
- In the LEO orbit, the Success appears related to the number of flights; but, there seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type



- With heavy payloads the successful landing are more for Polar,LEO and ISS

Launch Success Yearly Trend



- The average success rate kept increasing till 2020 since 2013

All Launch Site Names

- **the names of the unique launch sites in the space mission are:**

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

- **There are a four unique site**

Launch Site Names Begin with 'CCA'

- The 5 records where launch sites begin with `CCA` are:

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

- The total payload carried by boosters from NASA is:

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

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

avg(PAYLOAD_MASS__KG_)
2928.4

- It's like car 'mass

First Successful Ground Landing Date

- **The date when the first successful landing outcome in ground pad was achieved is**:

min(Date)
01-05-2017

- It's a great day

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is:

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

- Failure is rare

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass:

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

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are:

month	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

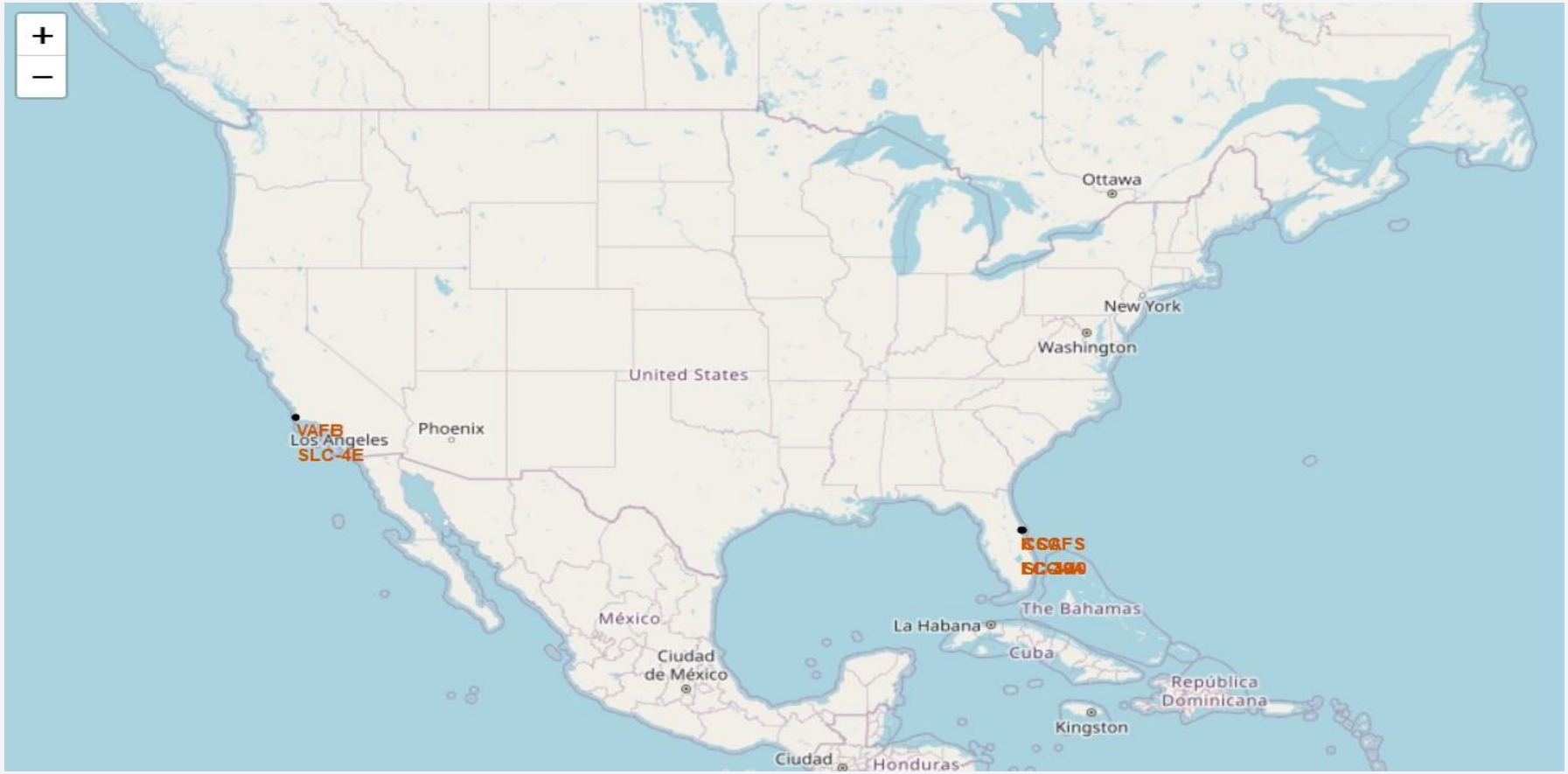
- **The rank of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order is:**

Landing _Outcome	count(*)
Success	20
Success (drone ship)	8
Success (ground pad)	6

- **There is a 34 success**

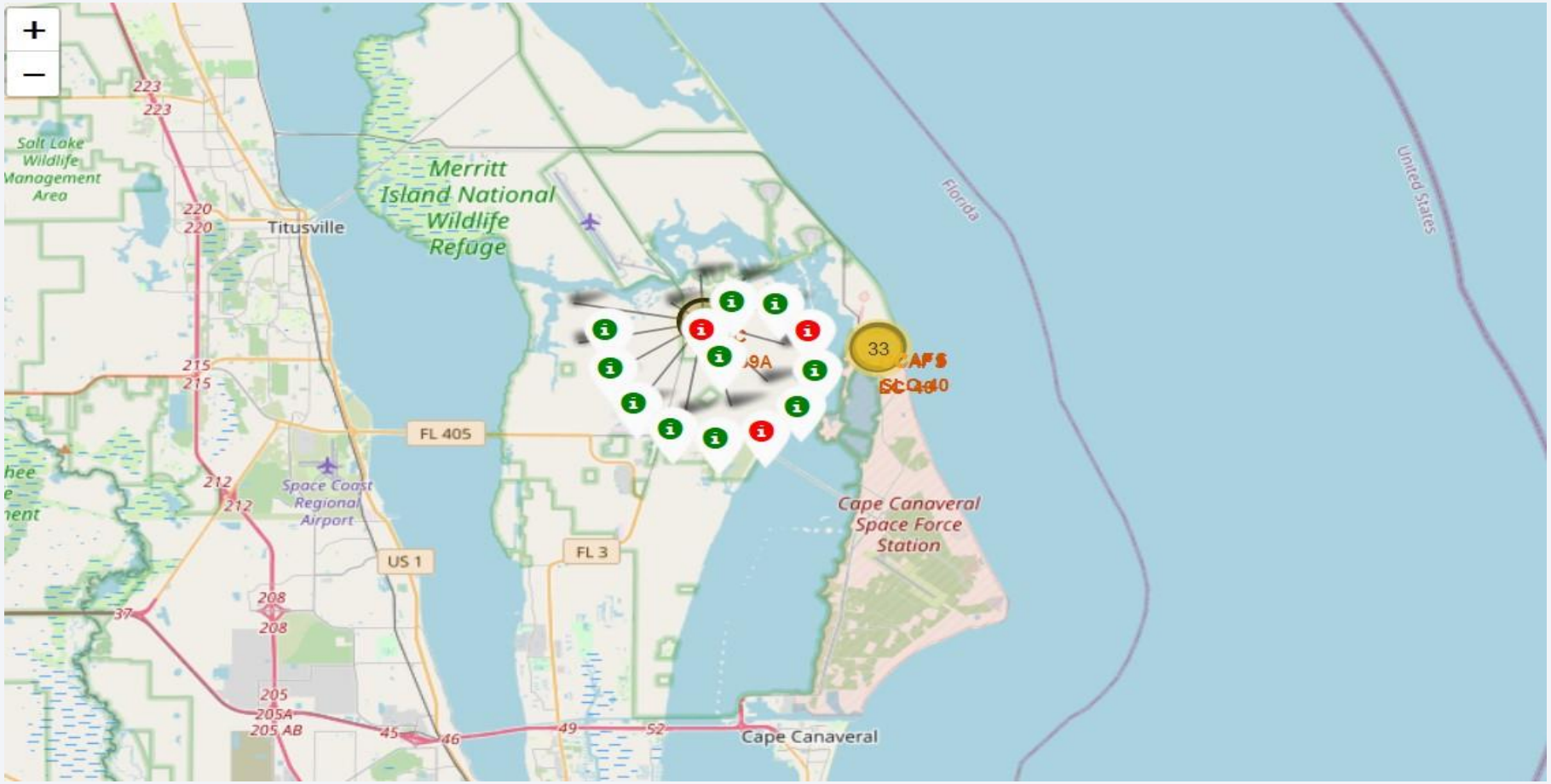

Section 3
Launch Sites
Proximities Analysis

launch sites on the site map



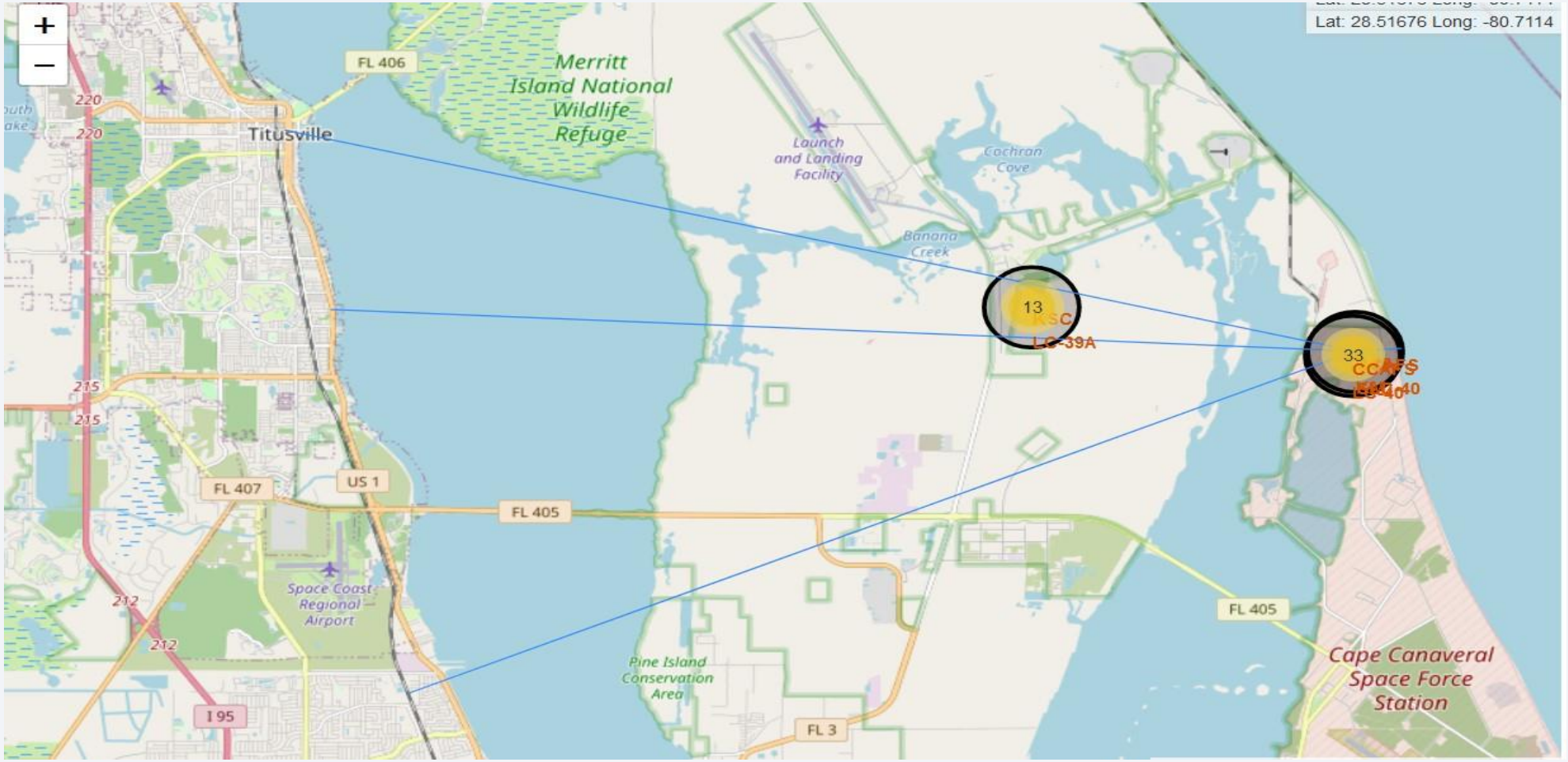
the launch sites are located near the coast and far from the cities, probably for safety

launch result

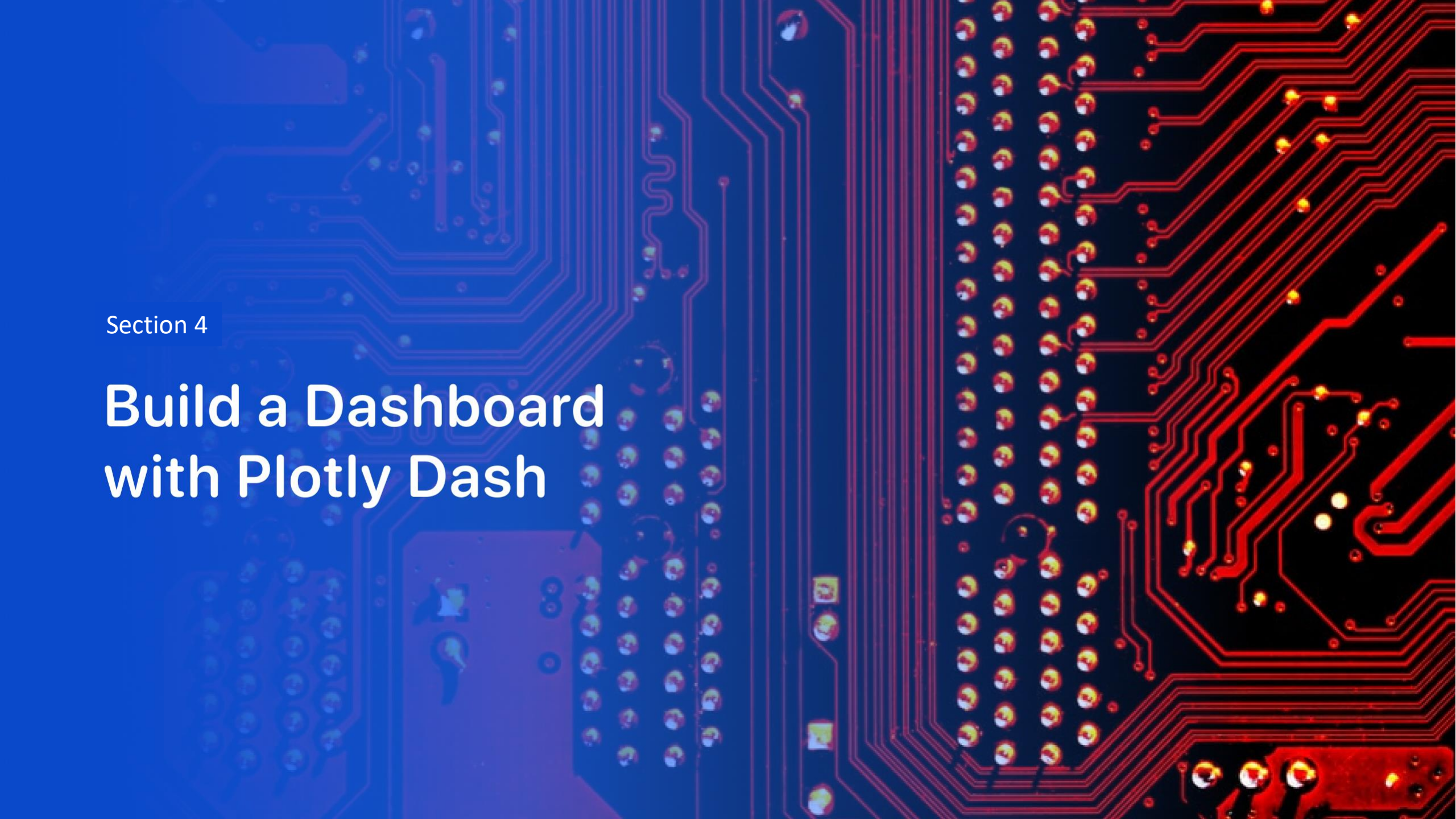


- For example in the launch site KSC LC-39A,there is a 13 launches and green marker indicate success and red one indicate the failure

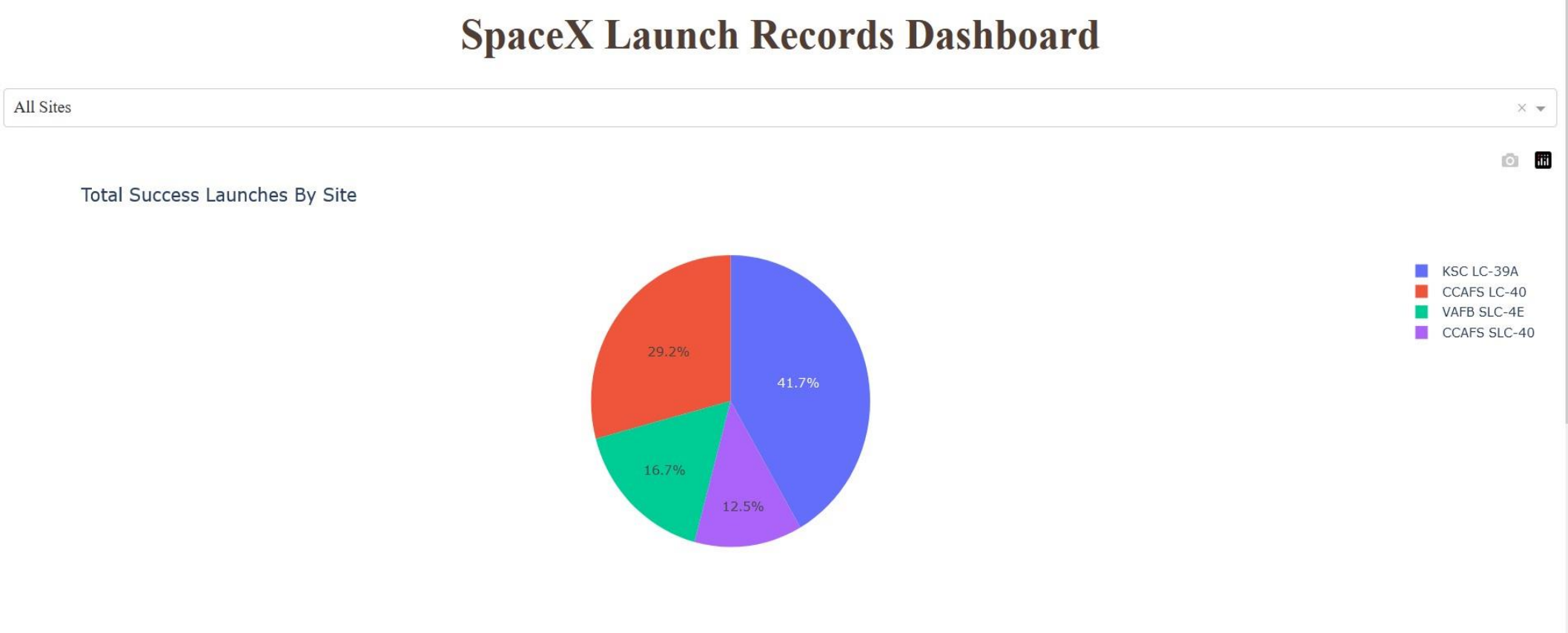
Infrastructure



- There is an important infrastructure around launch sites(road ,railway, bridges, ect,,,)

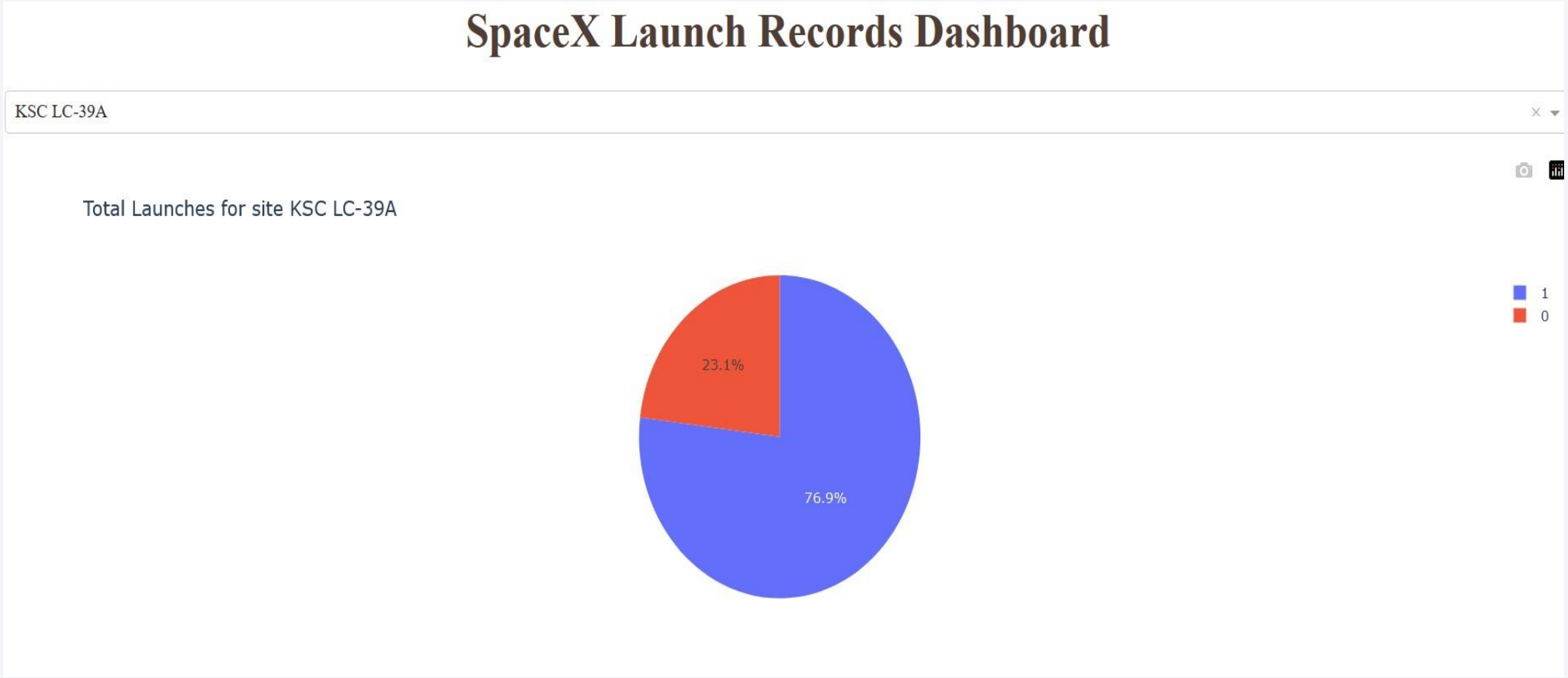

Section 4
Build a Dashboard with Plotly Dash

Total success launches by site



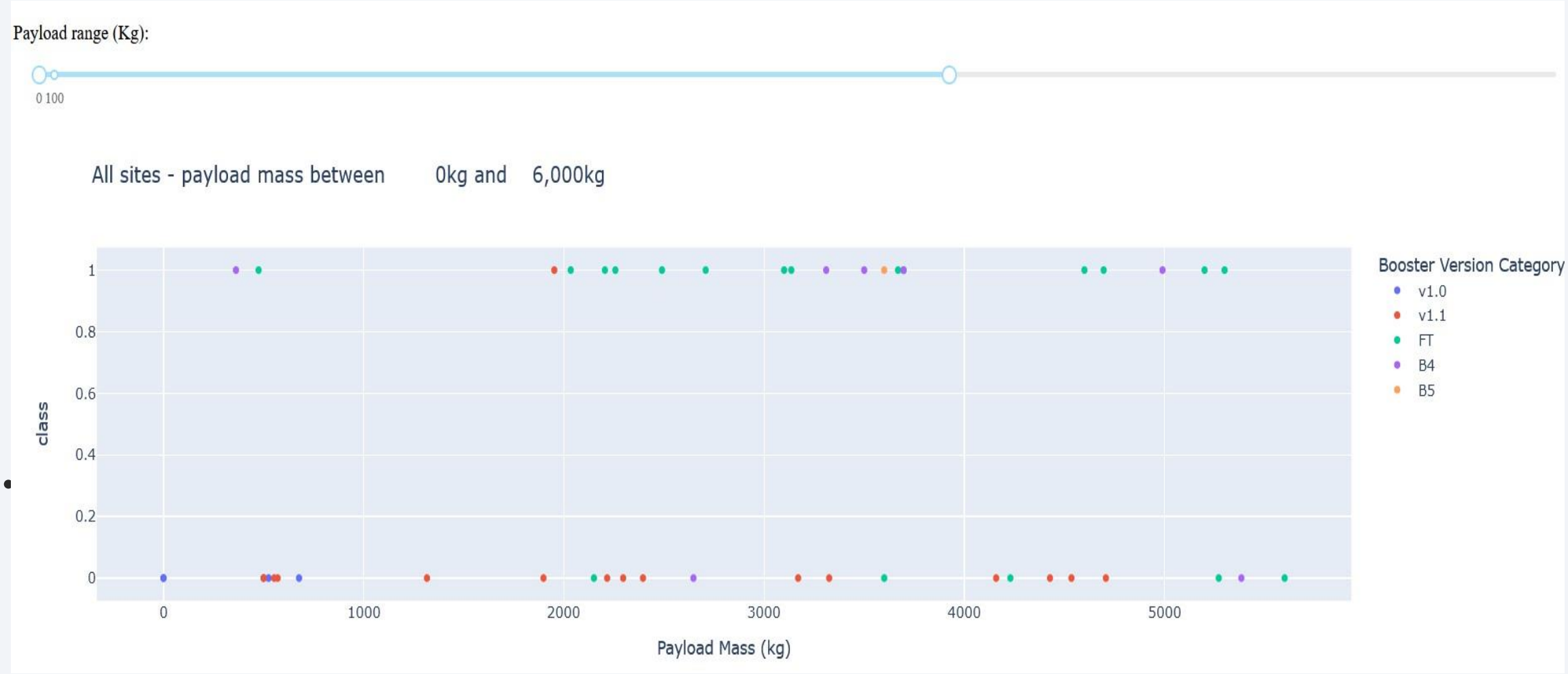
- The best launch site is KSC LC-39A with 41,7% of total launches site success

Best launch site



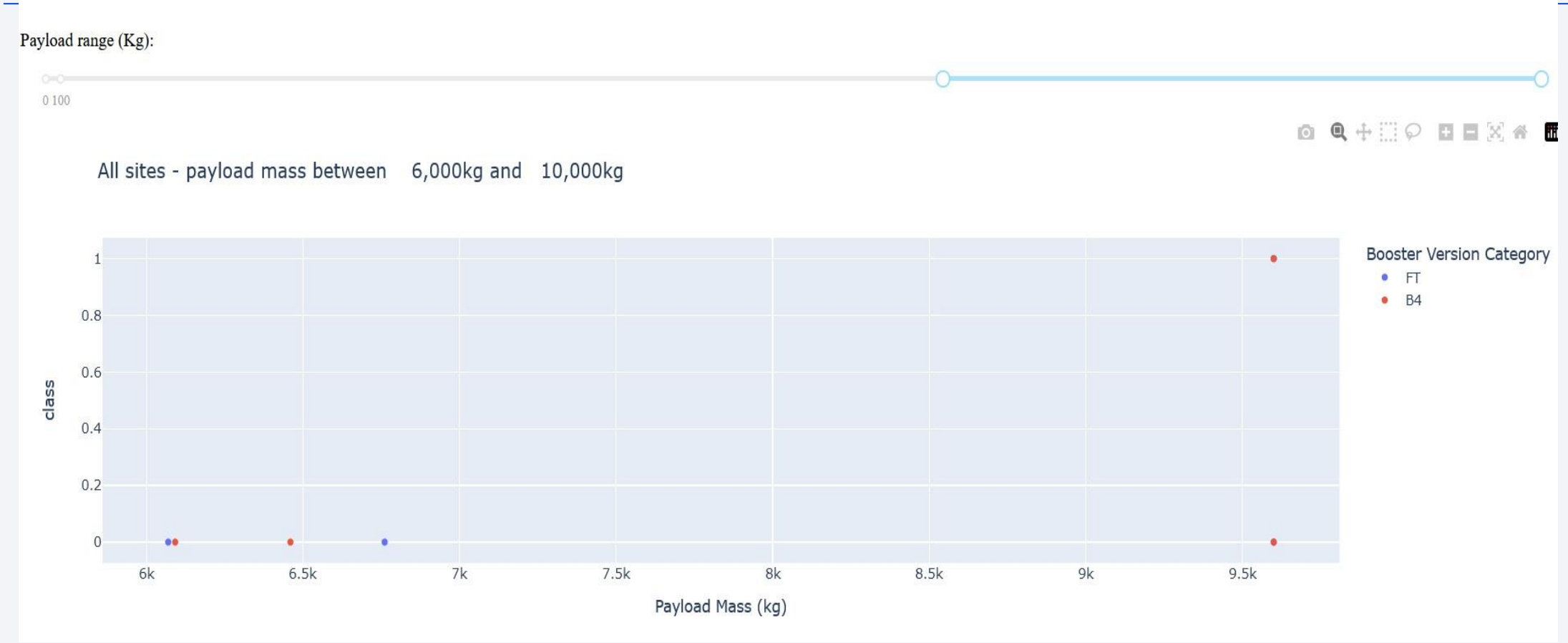
- The highest launch success ratio is in KSC LC-39A site,76,9 % of launches are successful in this site

Payload vs. Launch Outcome



- The FT booster version have the largest success rate on playload between 0 and 6000kg

Payload vs. Launch Outcome

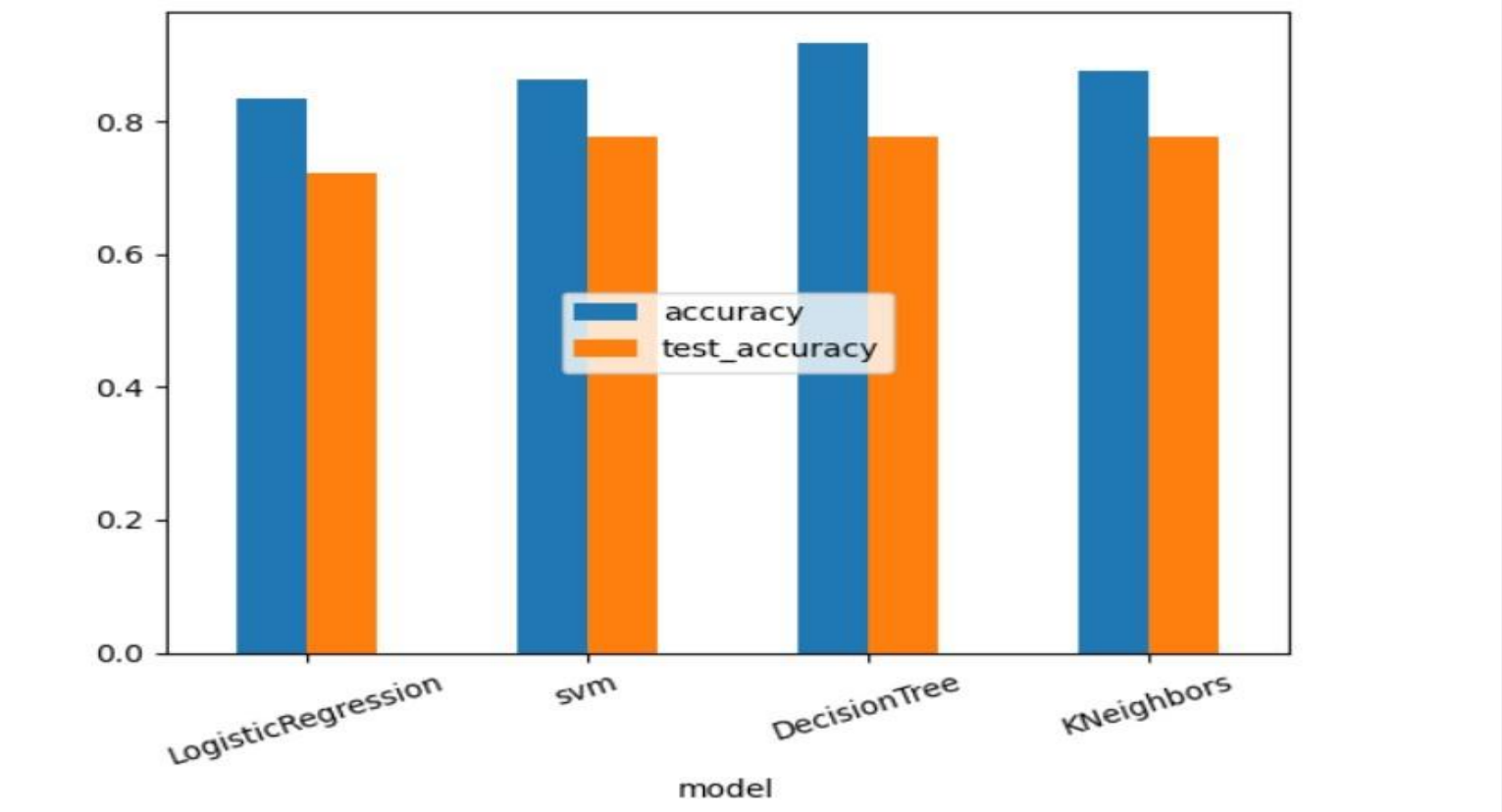


- The B4 booster version have the largest success rate on playload between 6000 and 9000kg,
- There is a few launches in this category of payload

Section 5

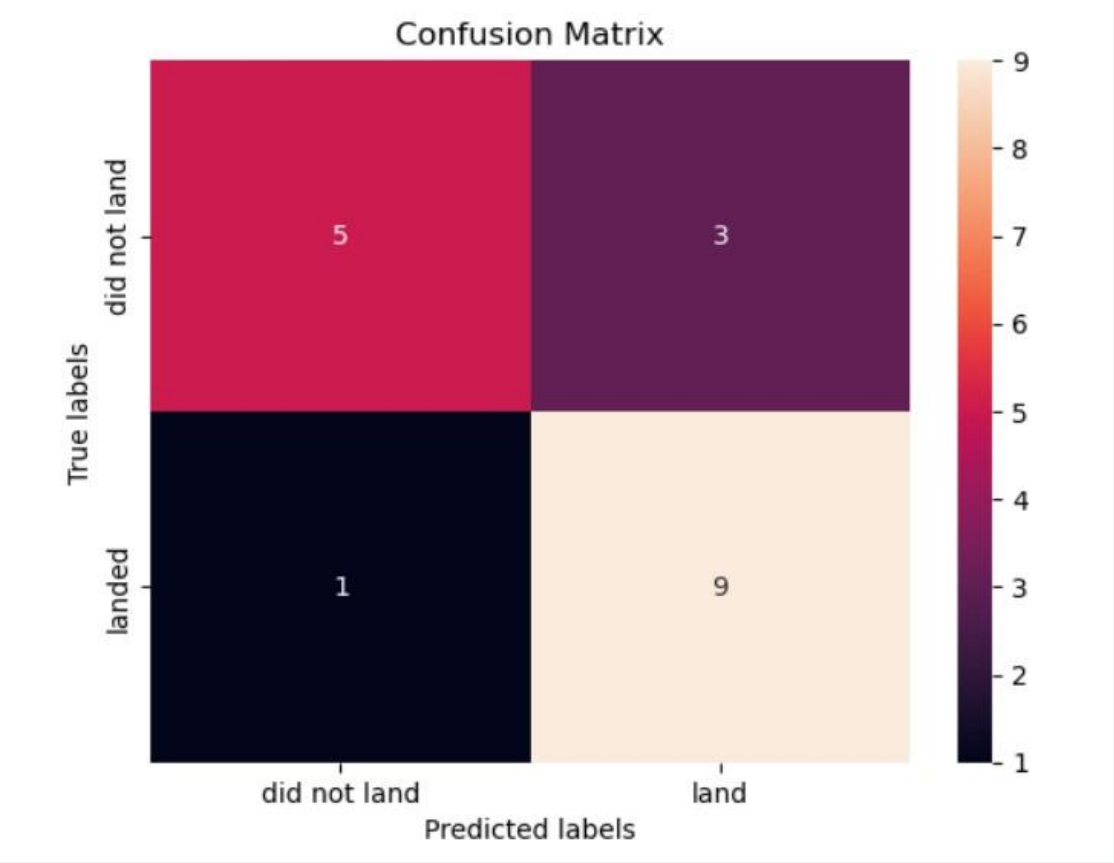
Predictive Analysis (Classification)

Classification Accuracy



- The model with the highest classification accuracy is Decision Tree

Confusion Matrix



- The false positive is weak compared to true positive and the false negative is also very low compared to true negative so the accuracy is very high

Conclusions

- The highest launch success ratio is in KSC LC-39A launch site.76,9 % of launches are successful in this site
- The high success rate(100%) is in orbits ES-L1,GEO, HEO and SSO
- The FT booster version have the largest success rate on playload between 0 and 6000kg
- The B4 booster version have the largest success rate on playload between 6000 and 9000kg
- The success improved with flight number increase and the average success rate kept increasing till 2020 since 2013
- The decision tree is the best model to predict Space X Falcon 9 First Stage Landing

Appendix

- *NOTA BENE: FOLIUM did’nt show maps in github*

Thank you!