Winning Space Race with Data Science

Vinicius Castillo de Lima
22/05/2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Introduction

- In this project I will take the disponible data of Space X to analysis what make the launches have a success landing
- I will make some analysis to understand better the data set, like some charts, statistics and a folium map
- Some charts I will try to create on a dashboard to anyone could make their analysis
- And I will try to find the best model to predict the landing success

Section 1

Methodology

Methodology

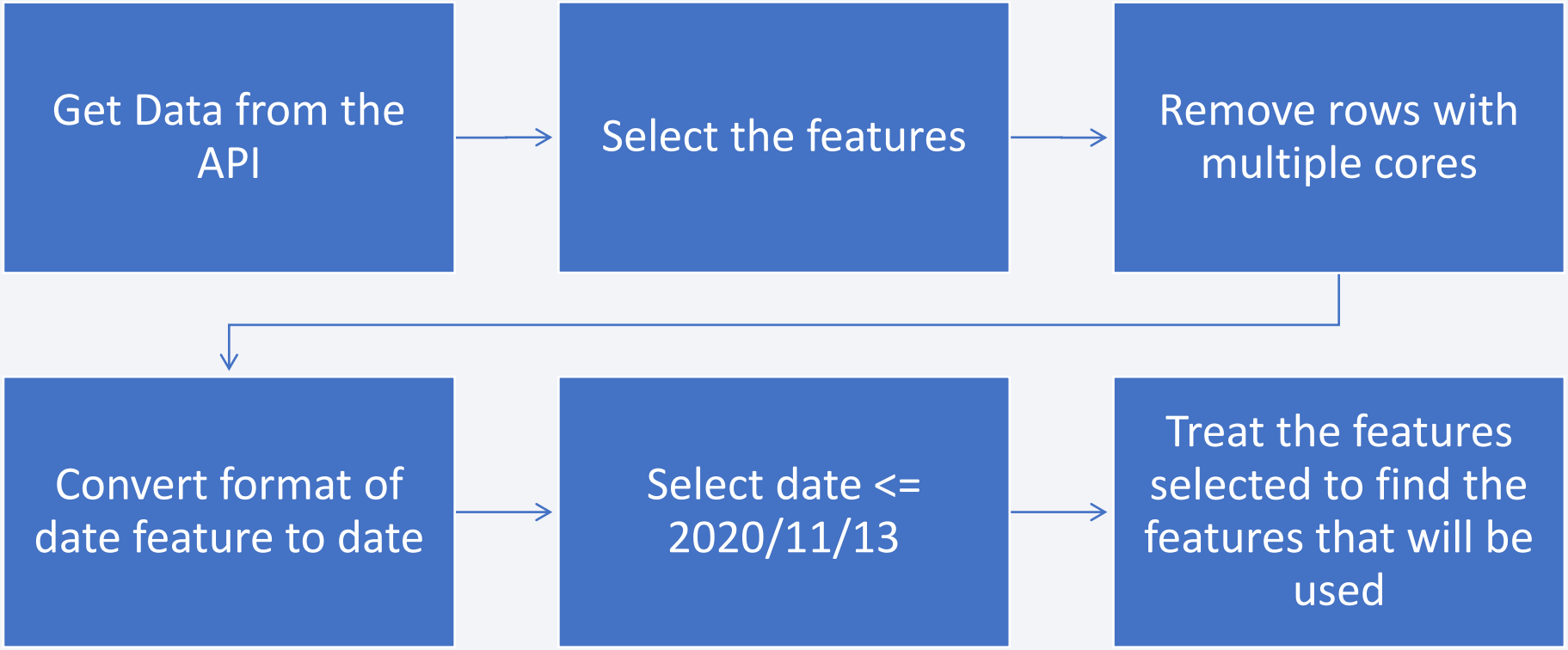
Executive Summary

- Data collection methodology:
  - Use Space X API and web scrapping
- Perform data wrangling
  - Create a Class label to identify the booster successfully landed or not.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Try 4 models and find the best parameters of each one with Grid Search CV.
  - Choose the better model using accuracy and confusion matrix

Data Collection

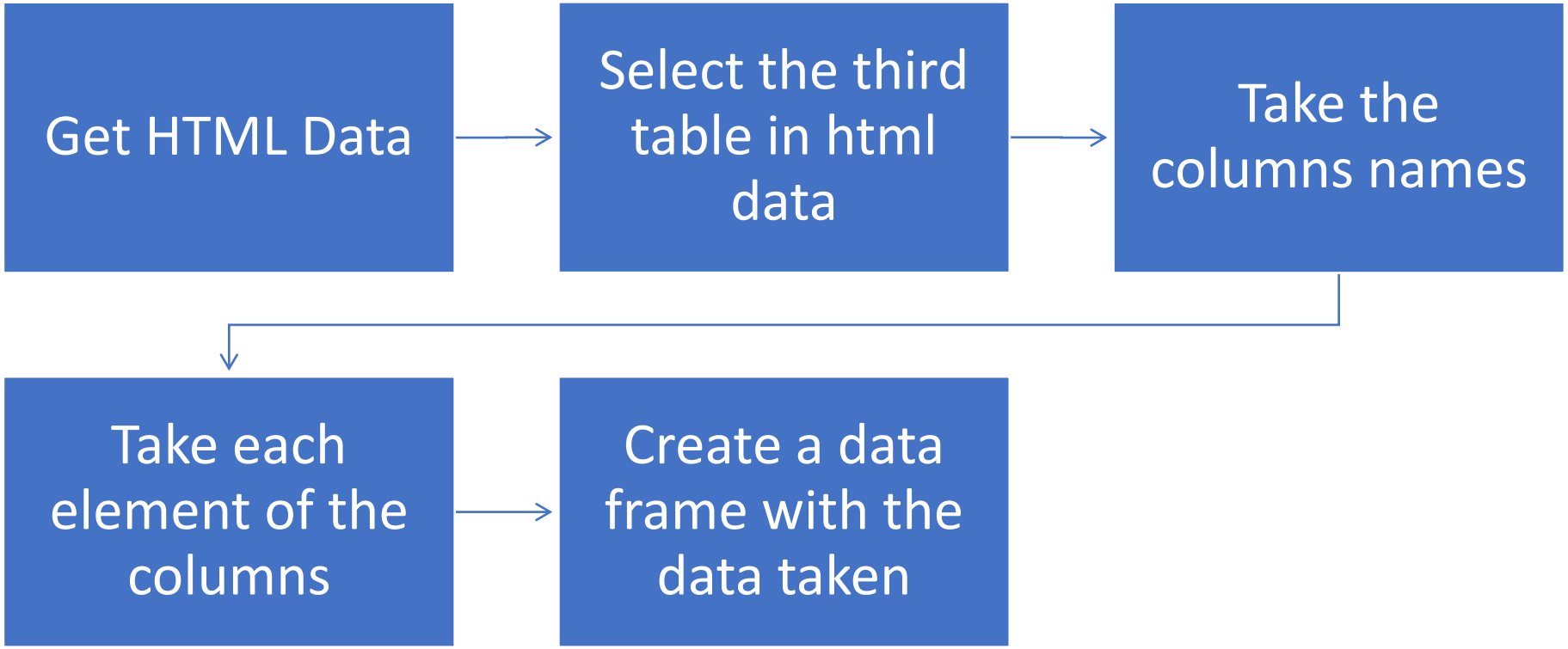
- Collect data from SpaceX API link: https://api.spacexdata.com/v4/launches/past (a static response could be access here)
- Use requests python library to capture the data from the API and make some treatments to make the dataset.
- More details on the next slide
- Other font of data is Wikipedia List of Falcon 9 and Falcon Heavy launches
- Use requests and beautifulsoup4 python library to capture the data from the Wikipedia and make some treatments to make the dataset.
- More details ahead on the Data Collection – Scraping slide

Data Collection – SpaceX API



- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/Data Collection API.ipynb

Data Collection – Scraping



- Link GitHub: [Data_Science_IBM_Certificate_Capstone_Project/Data Collection with Web Scraping.ipynb]

Data Wrangling

- Become to explorer the data to understood some features and find the better way do create a Training Labels
- Create a Training Labels with 1 means the booster successfully landed
  0 means it was unsuccessful

- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/Data wrangling.ipynb

EDA with Data Visualization

- Scatterplots to see some correlations:
  - Flight Number vs. Payload Mass colored by success or not
  - Flight Number vs. Launch Site colored by success or not
  - Payload mass vs. Launch Site colored by success or not
  - Flight Number vs. Orbit colored by success or not
  - Payload mass vs. Orbit colored by success or not
- Bar chart to see the success rate of each orbit type
- Line plot to observe the annual evolution of the average success rate
- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/EDA with Visualization.ipynb

EDA with SQL

- View the launch site options
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- First successful landing outcome in ground pad
- List of the boosters which have success in drone ship
- Total number of successful and failure mission outcomes
- List of the booster which have carried the maximum payload mass
- Booster versions and launch site that failed outcome in drone ship in year 2015
- Count of landing outcomes between the date 2010-06-04 and 2017-03-20

- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/EDA with SQL.ipynb

Build an Interactive Map with Folium

- Create a folium map with circles and marks on the launch sites
- Create green marks for launch was successful on landing and red marks for launch was failed
- Create line and marks to show the distance between the launch site of Florida to:
  - Closest coastline
  - Closest railway
  - Closest highway
  - Closest cities

- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/Interactive Visual Analytics with Folium.ipynb

Build a Dashboard with Plotly Dash

- Created a dashboard with a pie chart with the success launches for each launch site
- Included a scatter chart with the correlation between the payload mass vs. success or fail on landing and colored by the booster version
- Included a possible to select the Launch Site to filter both plots
- Included a possible to choose the payload mass range to filter the scatter plot

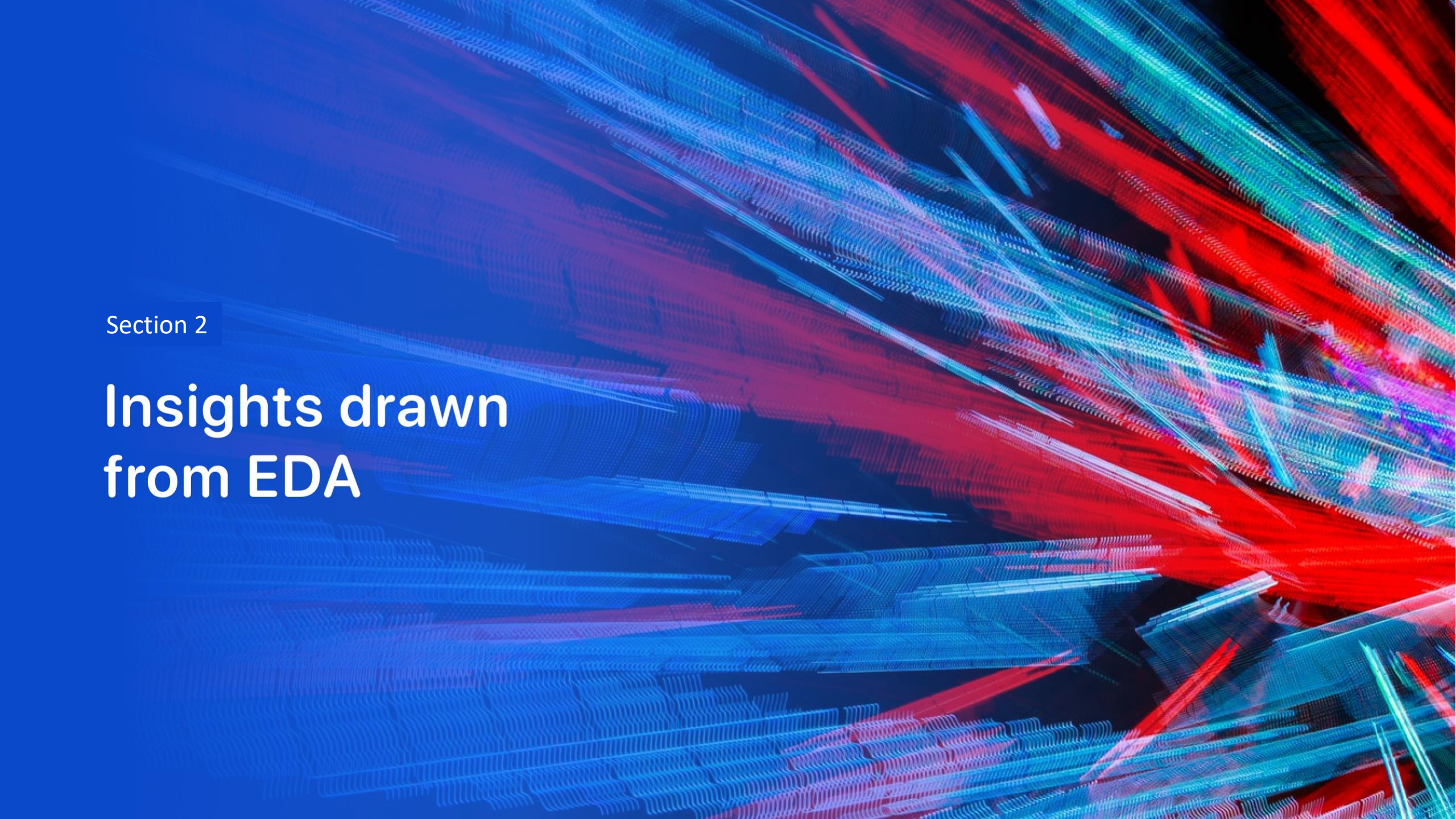
- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/Dashboard with Plotly Dash.py

Predictive Analysis (Classification)

- We take the data with the categorical features transformed by one hot encoding, this transformation was made on the EDA with Data Visualization
- Separate the features data (X) and the target (Y)
- Apply standard scaler on the features
- Separate the data between train and test sets
- Find the best parameters of each model (logistic regression, support vector machine, decision tree and k nearest neighbors) with Grid Search CV
- Use accuracy and confusion matrix to find the best model
- Link GitHub: Data_Science_IBM_Certificate_Capstone_Project/Machine Learning Prediction.ipynb

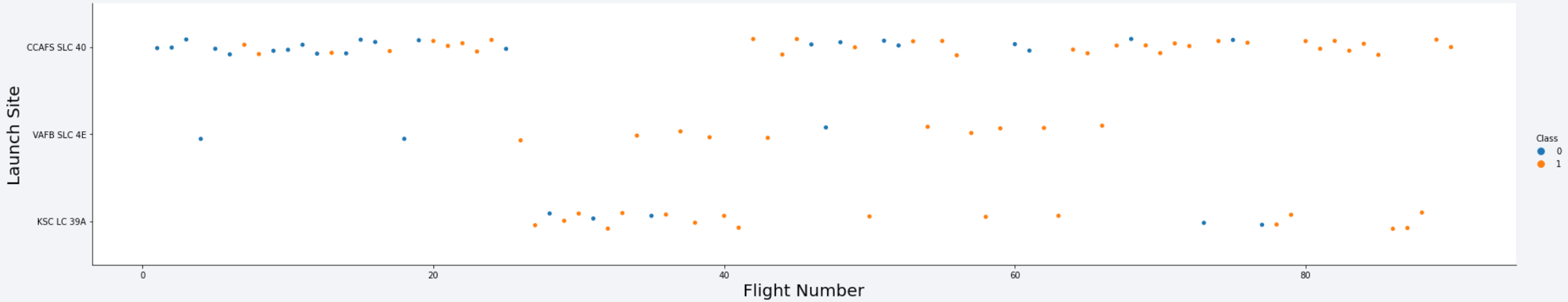
Results

- Exploratory data analysis results (section 2 and 3)
- Interactive analytics demo in screenshots (section 4)
- Predictive analysis results (section 5)


Section 2
Insights drawn from EDA

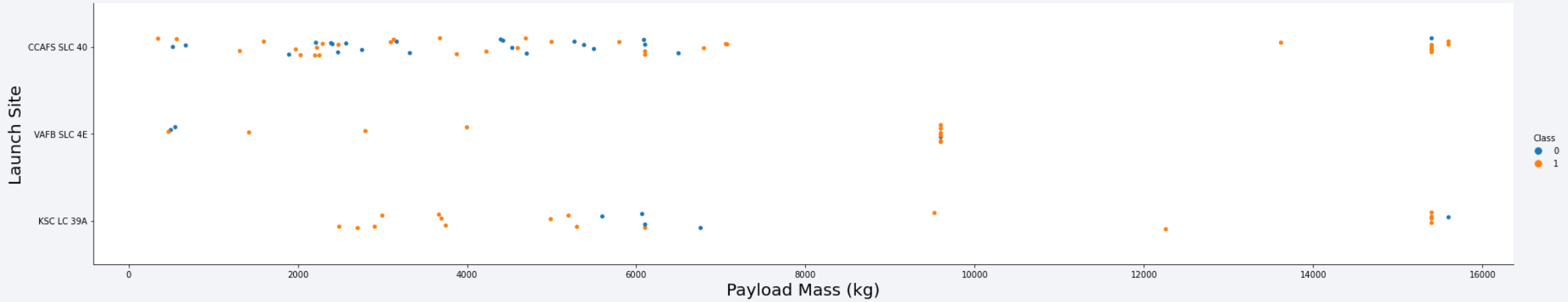
Flight Number vs. Launch Site

- This chart show us that CCAFS LC-40 success ratio is lower than the others



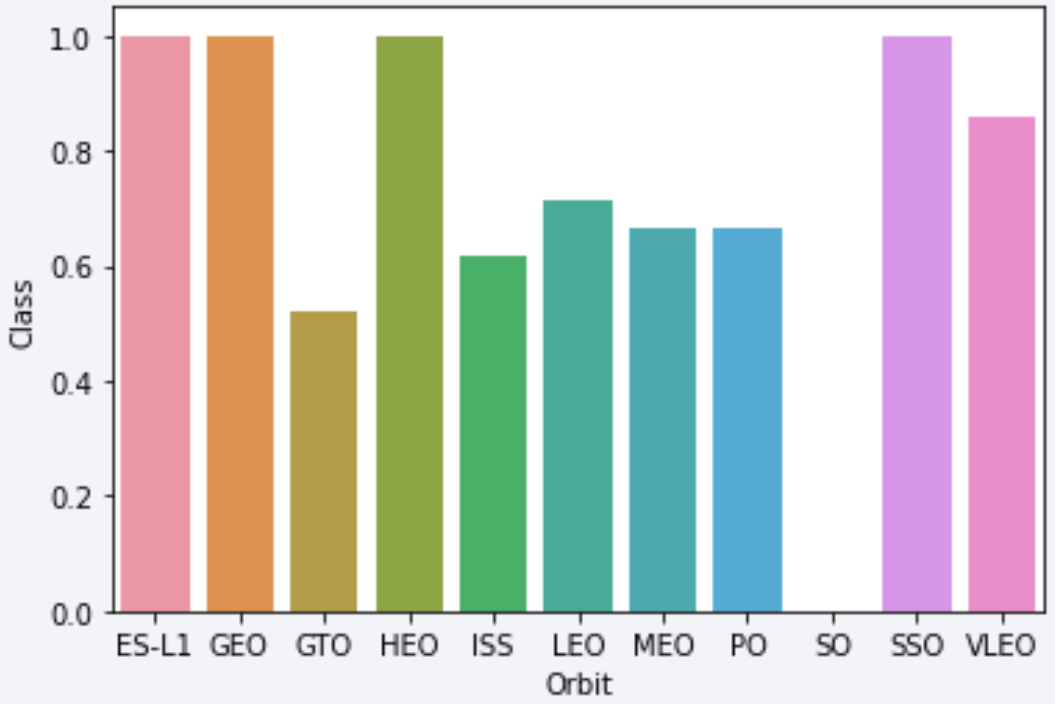
Payload vs. Launch Site

- Here we can see the reduction of the success when the payload is great than 6000 kg
- Another observation is the max payload mass launch from VAFB-SLC site is lower than 10.000 kg



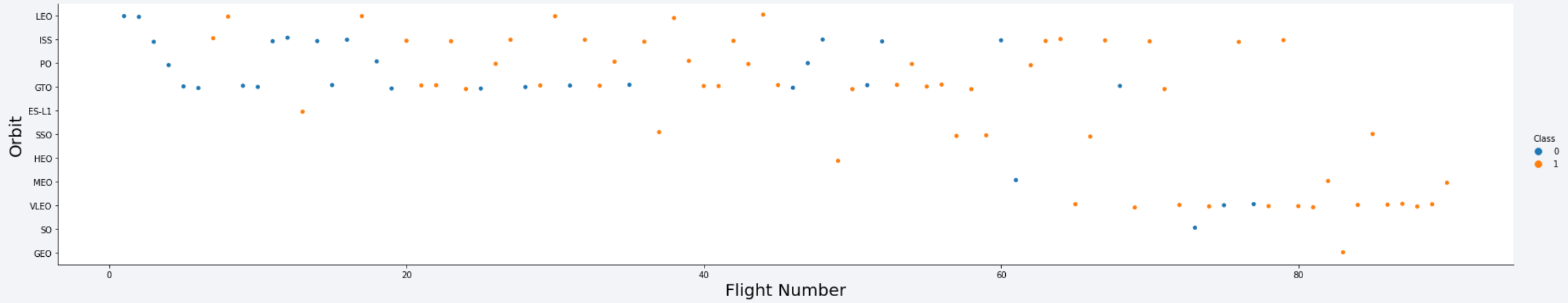
Success Rate vs. Orbit Type

- Here we can see the orbits ES-L1, GEO, HEO and SSO had 100% of success ratio while the orbit SO had 100% of failure



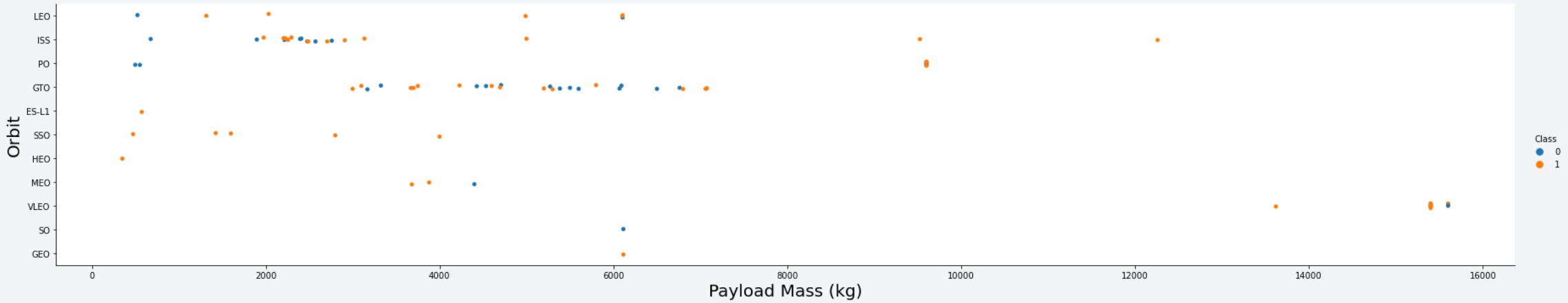
Flight Number vs. Orbit Type

- We could see that in the LEO orbit the Success appears related to the number of flights
- Others such as GTO don't appears have any correlation



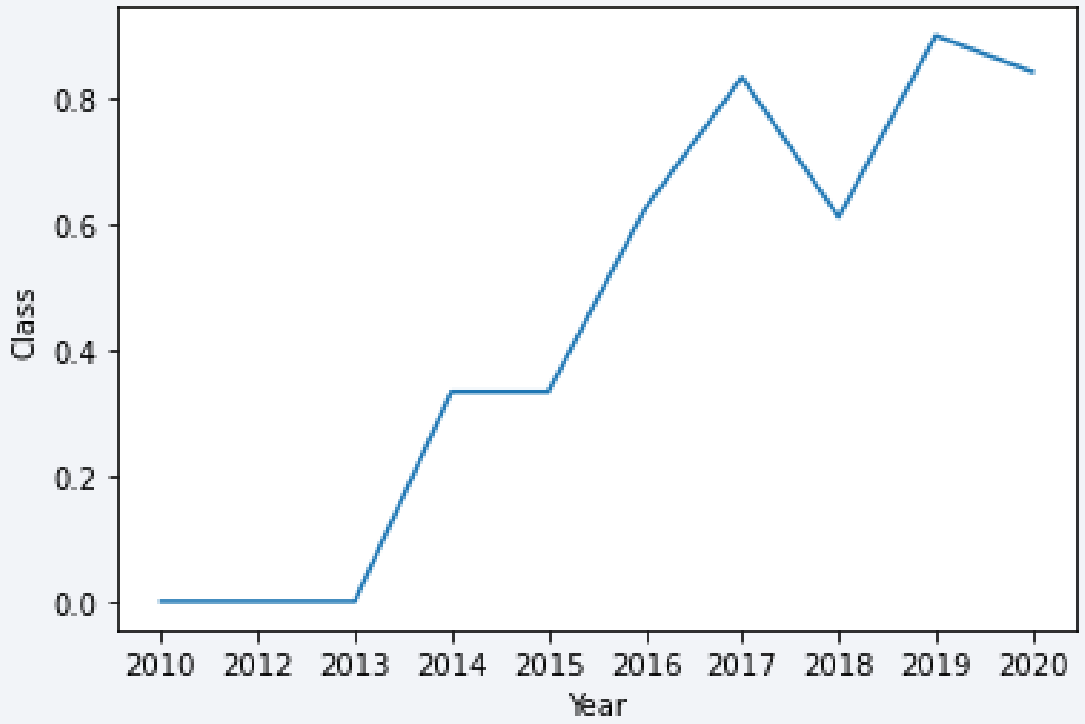
Payload vs. Orbit Type

- Like on the chart before, the correlation is easily observed on orbit LEO and is difficult to see on the GTO.



Launch Success Yearly Trend

- We can see the success rate increases with the years



All Launch Site Names

- We discovered that have only four launch sites, and they are:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- When we look to the five first records where launch sites begin with `CCA`, no one had success on landing, but the mission was successful

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

- The total payload mass carried by boosters launched by NASA(CRS) was 45596 kg.

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 was 2534 kg

First Successful Ground Landing Date

- The first successful landing outcome in ground pad was achieved in 2015/12/22

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 was:
  - F9 FT B1022
  - F9 FT B1026
  - F9 FT B1021.2
  - F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- When we see the number of successful and failure mission outcomes, we see only one failure on 101 launches.

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

Boosters Carried Maximum Payload

- The booster which have carried the maximum payload mass was:
  - F9 B5 B1048.4
  - F9 B5 B1048.5
  - F9 B5 B1049.4
  - F9 B5 B1049.5
  - F9 B5 B1049.7
  - F9 B5 B1051.3
  - F9 B5 B1051.4
  - F9 B5 B1051.6
  - F9 B5 B1056.4
  - F9 B5 B1058.3
  - F9 B5 B1060.2
  - F9 B5 B1060.3

2015 Launch Records

- The booster versions and launch site names of the failed landing outcomes in drone ship, for in year 2015 was:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

- 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 was:

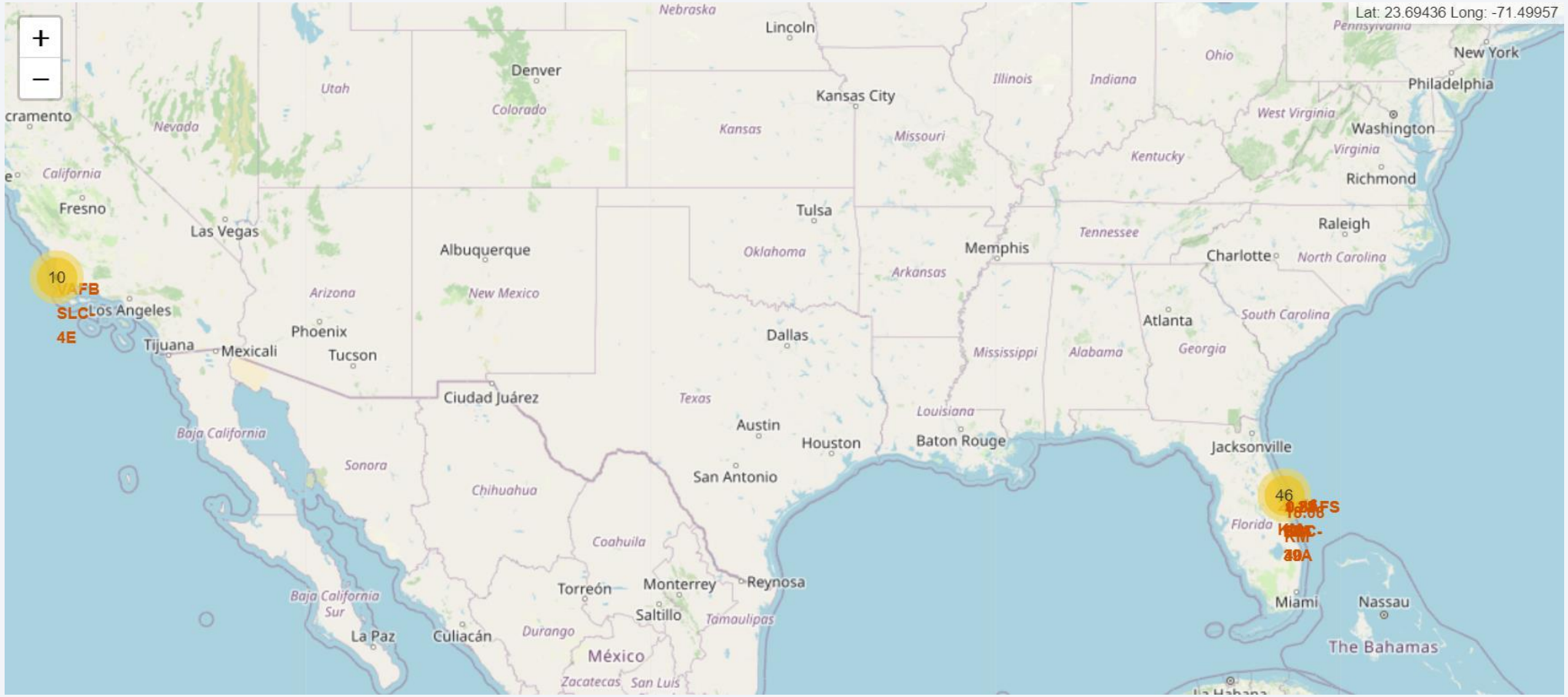
landing_outcome	number
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

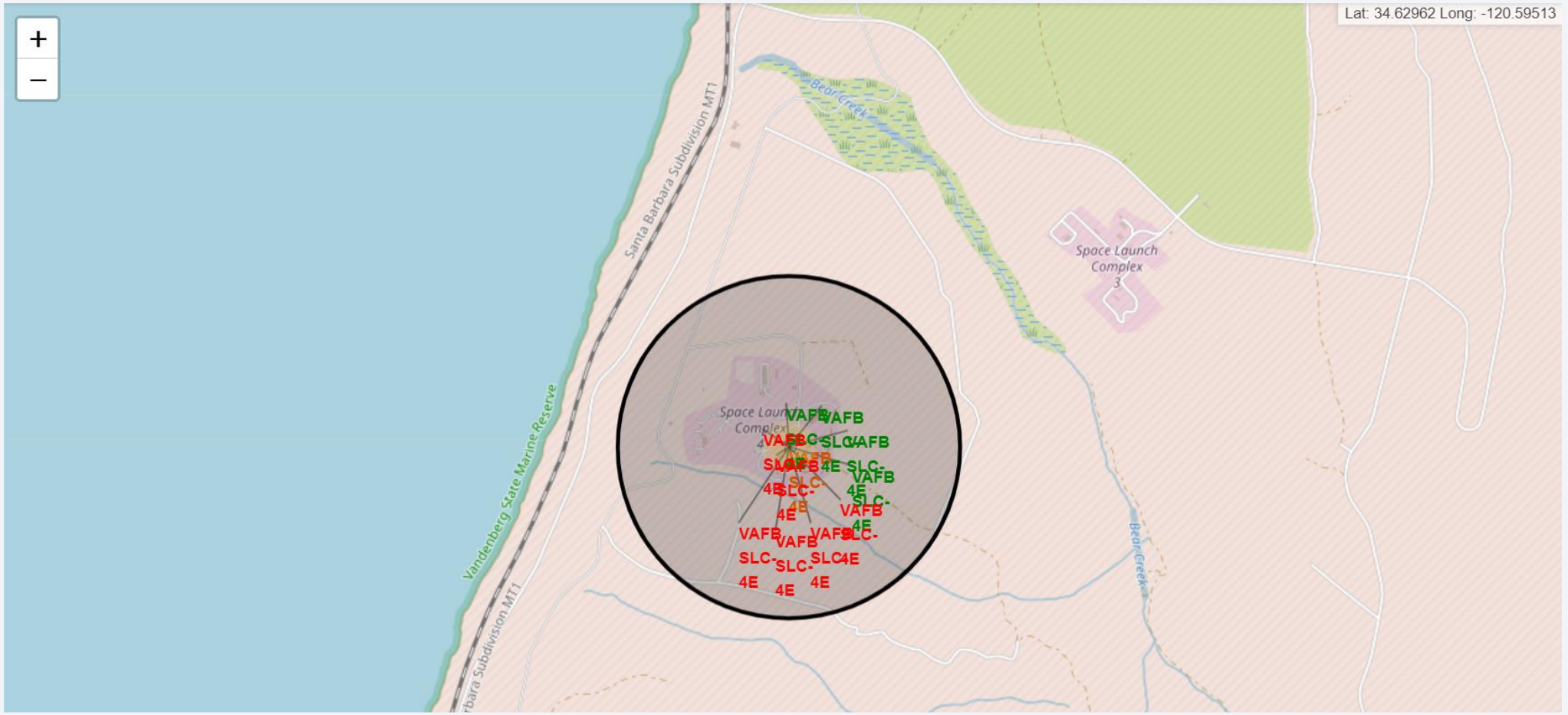
Map of the launch sites

- In this map appears all launch sites with the number o launches happened on them



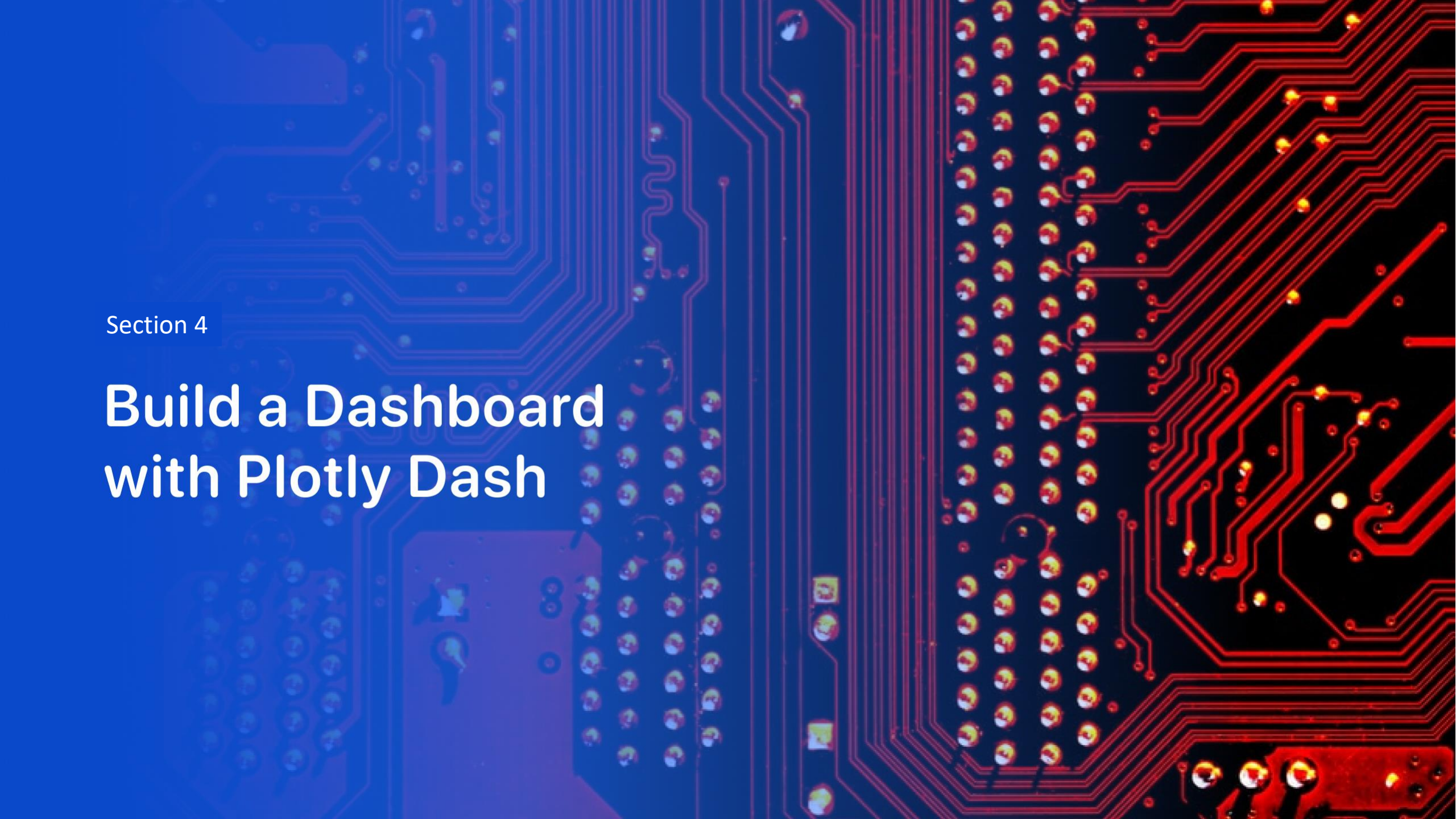
Color mark for launches

- Here are the marked red the launches was failure and green the launches was success



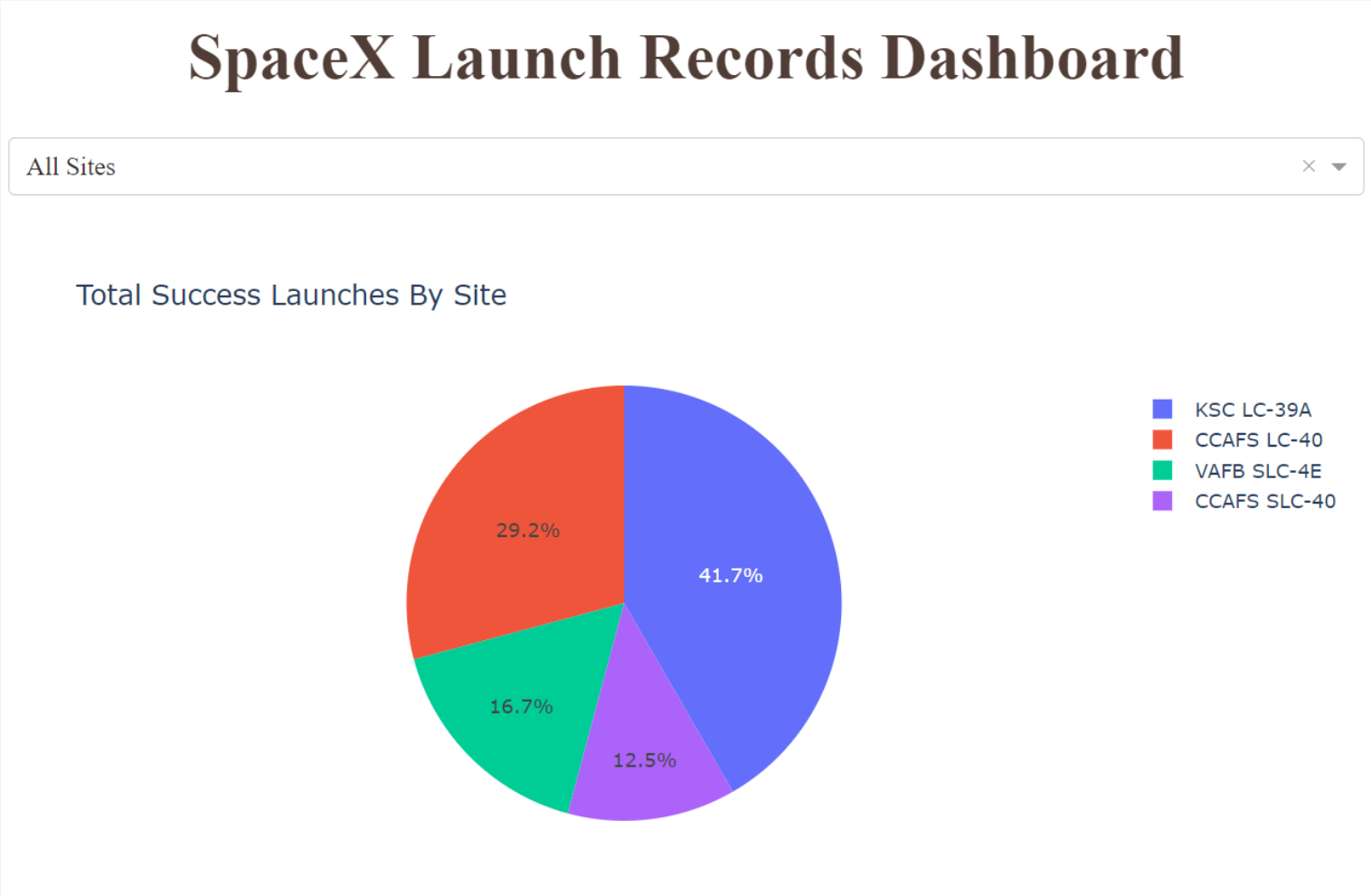
Distances from the launch site

- This have lines and the distance from some points, like closest coastline and highway


Section 4
Build a Dashboard with Plotly Dash

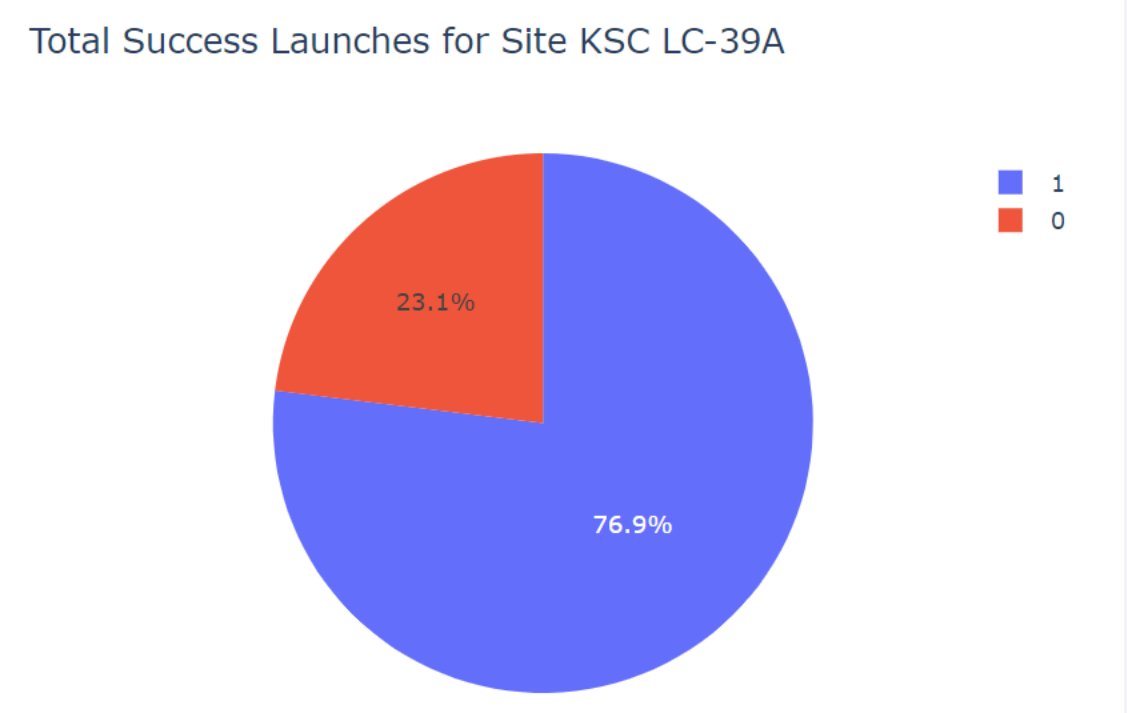
Launch sites success on dashboard

- This first chart on the dashboard show the total success launches of each launch site
- Is possible to see the KSC LC launch site have more launches with success than the others



The most success ratio

- This is the chart of the launch site with the most success ratio
- It is the same launch site that have the most success launches observed before

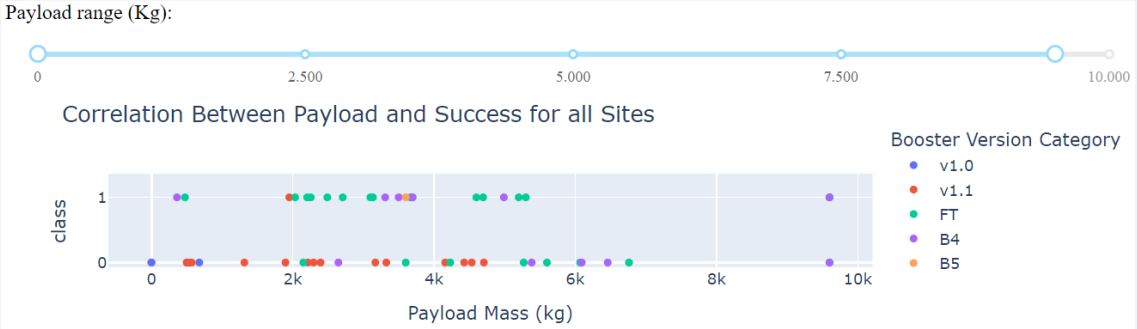


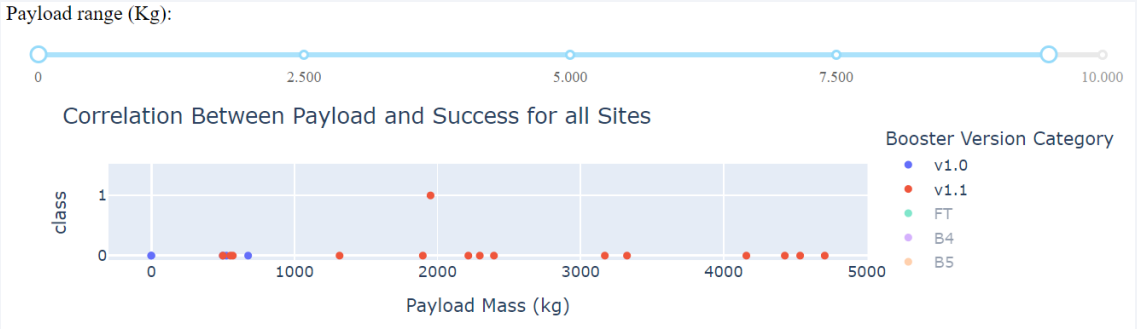
Payload and Booster Version vs. Success

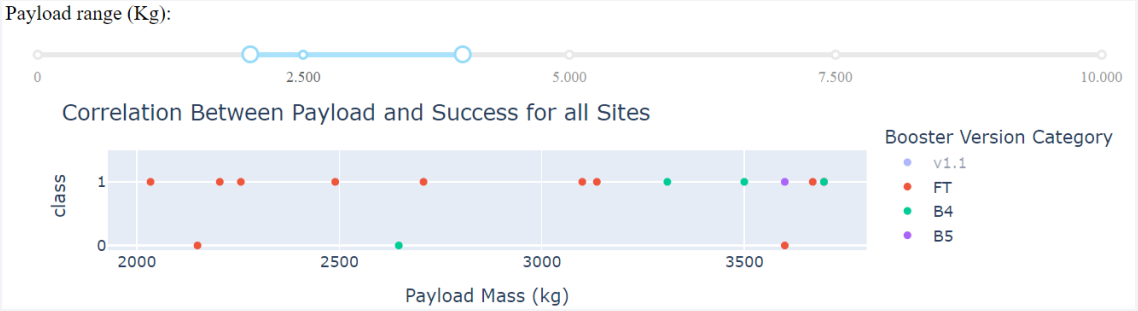
- View all sites information together is possible to see some concentration of colors in fail (class = 0) range

- When is selected the booster versions v1.0 and v1.1 almost all points is in failure range

- Other think is the range (2000 until 4000 kg) of the payload have the most success on landing





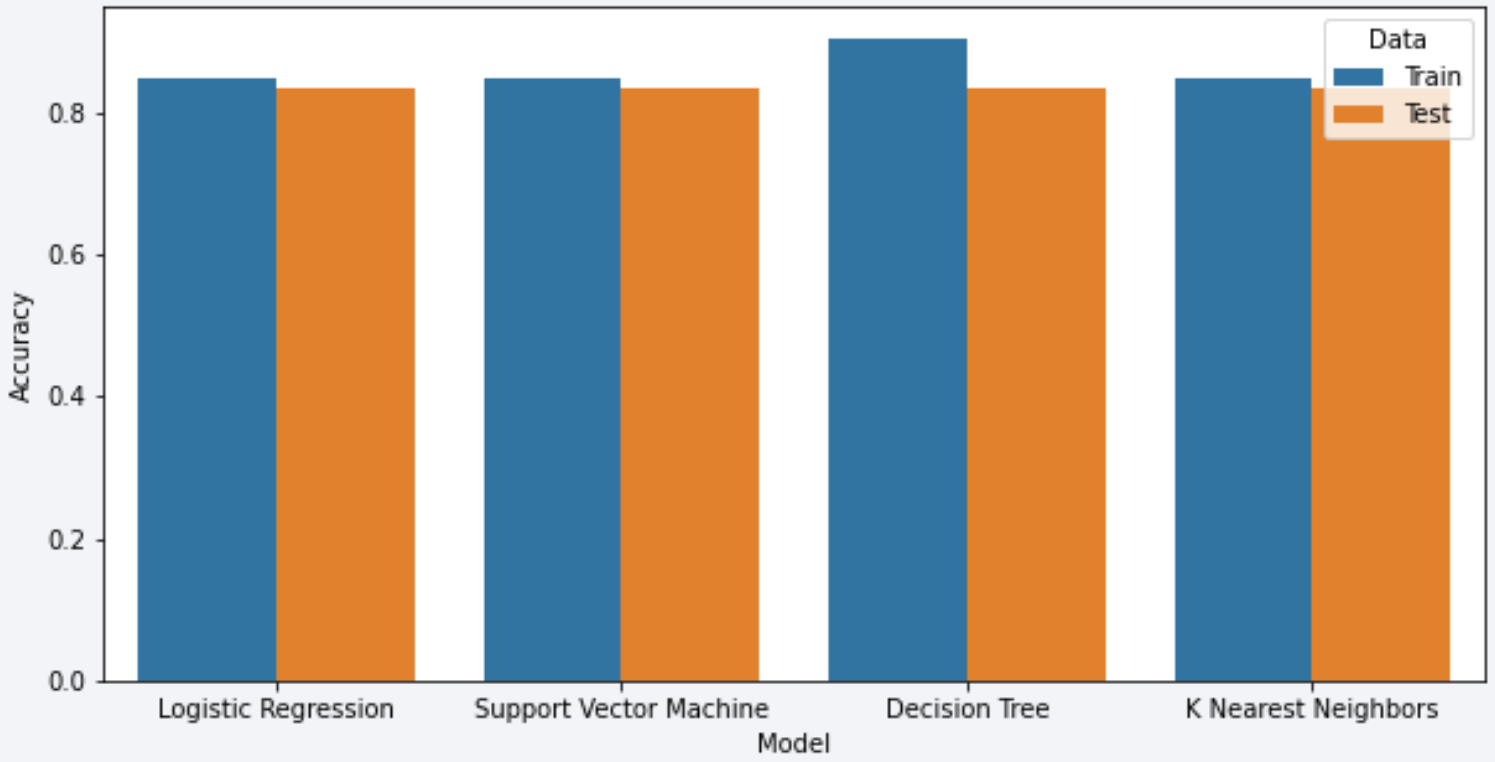


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- All models had the same accuracy and confusion matrix with the test data
- The best model is the Decision Tree, because in the train set had the best accuracy – 90% and, as said before, the same result of the others model on the test data



Confusion Matrix

- The model could distinguish between the different class
- The problem is the false positive

Conclusions

- We find a good model with accuracy of 83,33%
- We can observe some interesting correlations such as orbit, payload mass and the success
- We could calculate the distance between the launch site and some important locals and see that have security distance to important sights like cities
- We create a dashboard where anyone could take their analysis

Thank you!