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
  - ✓ Data Collection
  - ✓ Data Wrangling
  - ✓ Data exploration and analysis
  - ✓ Predictive modeling
- ✓ Summary of all results
  - ✓ The data was pre-processed and analysed, using statistical and visual tools
  - ✓ Classification models were developed
  - ✓ Presenting the results

Introduction

- Project background and context

Falcon 9 was developed by SpaceX with the main purpose of reusing its parts. This will drastically reduce the costs of space exploration while making it more reliable for people and payloads.

- Problems you want to find answers

The goal is to build a model that can help us as a company to determine what aids in getting a succesfull landing.

Section 1

Methodology

Methodology

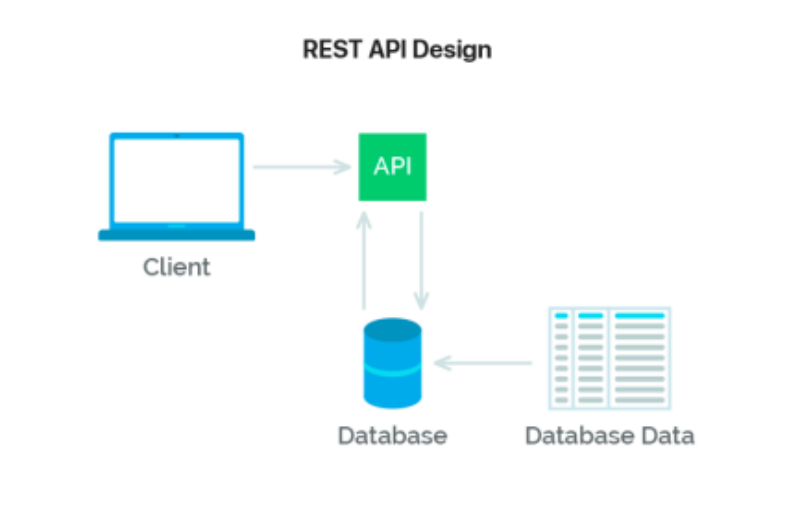
Executive Summary

- Data collection methodology:
  - the data was collected with the help of SpaceX API and Wikipedia scrapping
- Perform data wrangling
  - Feature extraction and transforming using python libraries
  - Create new data to improve models accuracy
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Few classification models were build and tested to choose the best one

Data Collection

Using the SpaceX API, you can access data without using any credentials.

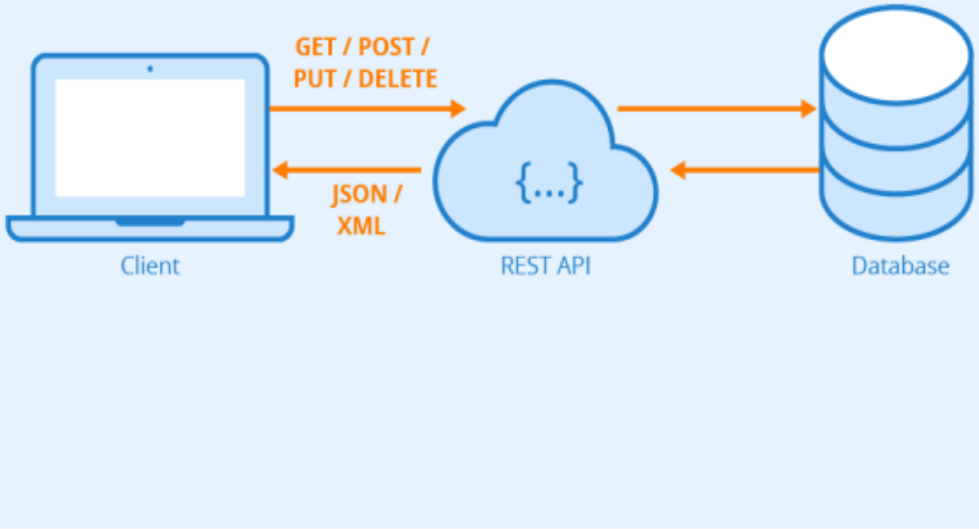
Also, to get more useful data, web scrapping was performed on the Wikipedia page



Data Collection – SpaceX API

- The result of the API is a JSON file that was transformed into a Data Frame

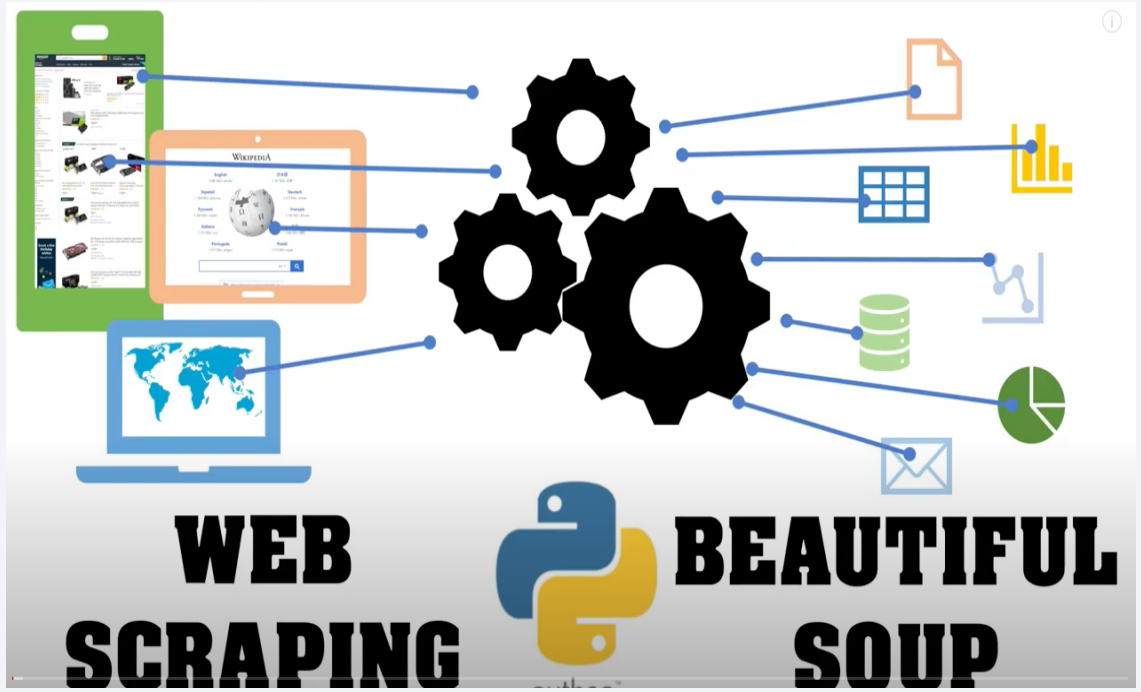
- The code used cand be found here:
- https://github.com/AndreiToma89/coursera-capstone-project/blob/master/capstone.ipynb



Data Collection - Scraping

- using the beautifulsoup python library, additional data was collected

- The code for this operation can be found here:
- https://github.com/AndreiTom a89/coursera-capstone-project/blob/master/Data%20Collection%20with%20Web%20Scraping%20.ipynb



Data Wrangling

- The collected data was preprocessed. The methods involved were replacing missing values, feature extraction and addition of new columns

- https://github.com/AndreiToma89/coursera-capstone-project/blob/master/Data%20Wrangling.ipynb

EDA with Data Visualization

- A number of charts were created to highline the correlation between data. The visual representation of the data was more initiative in explain few findings as well as finding important features

https://github.com/AndreiToma89/coursera-capstone-project/blob/master/Eda%20with%20Visualisation.ipynb

EDA with SQL

- Using SQL we could dive more into the data and we could extract more valuable information.

- GitHub URL: https://github.com/AndreiToma89/coursera-capstone-project/blob/master/Eda%20with%20SQL.ipynb

Build an Interactive Map with Folium

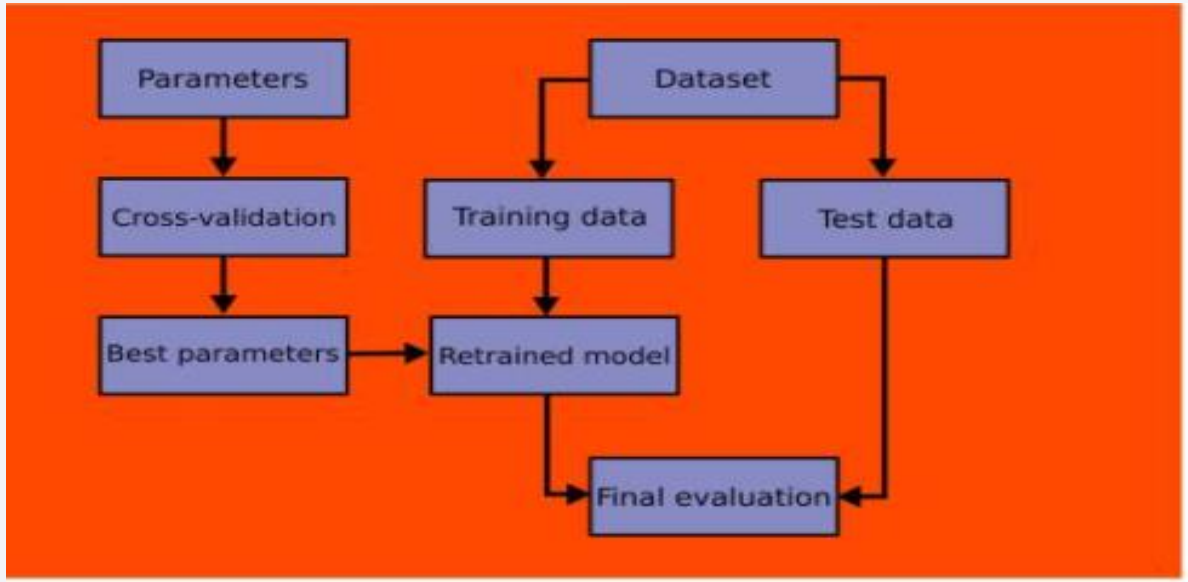
- An interactive map with Folium was developed to help understanding the data with graphs and maps. The map shows circular marks for each site with color coding depending on the outcome.
- Adding the markers and marking points on the map, help in visualize the data better in the geo contex.
- GitHub URL: https://github.com/AndreiToma89/coursera-capstone-project/blob/master/Folium.ipynb

Build a Dashboard with Plotly Dash

- Adding an interactive dashboard was the next step in exploring the data better. Each user can get their desired visualizations, in order to extract valuable information and insights.
- GitHub URL: https://github.com/AndreiToma89/coursera-capstone-project/blob/master/Dashboard.ipynb

Predictive Analysis (Classification)

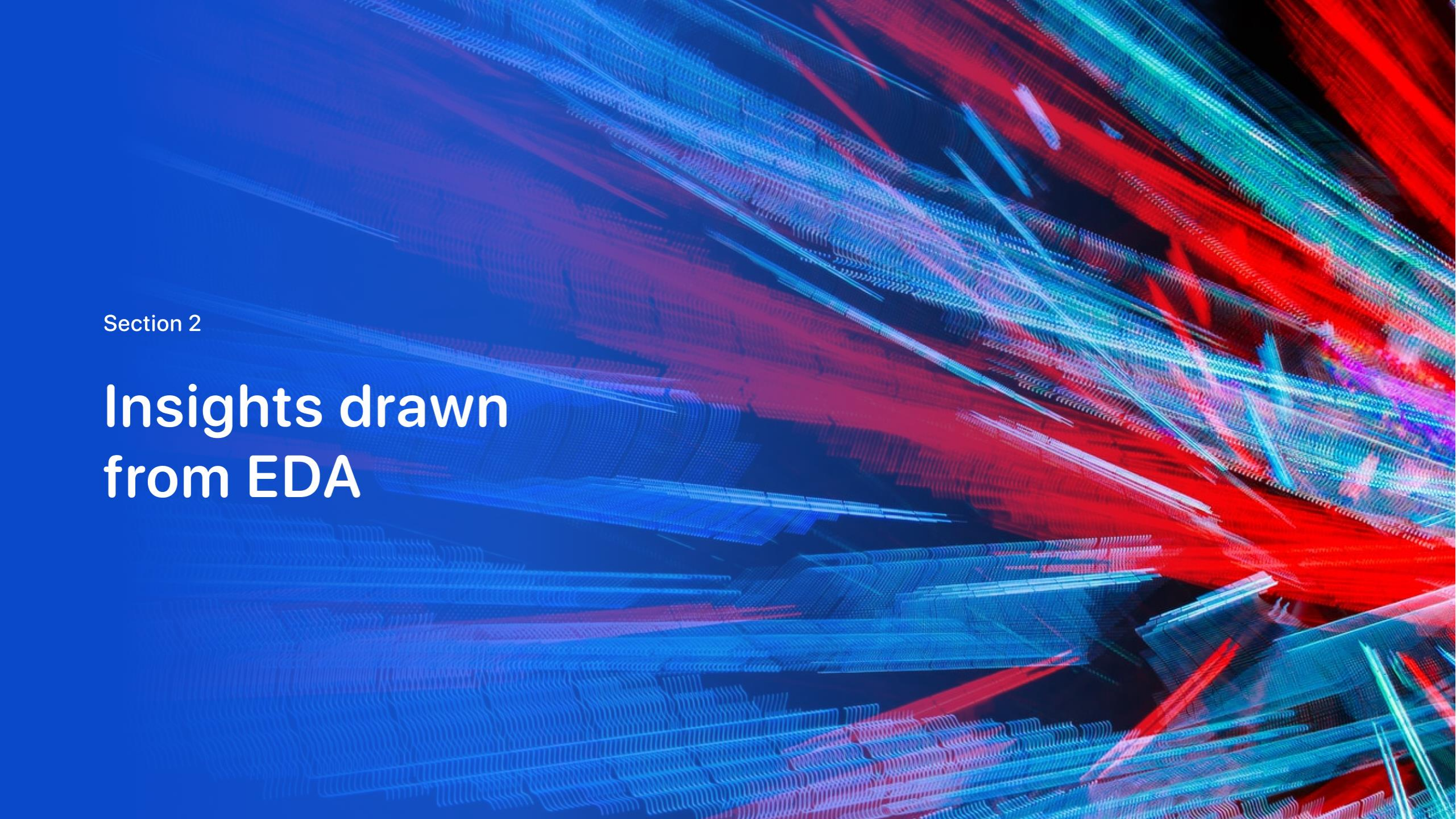
- In order to get the best result, several classification models were created, and, combined with the power of a Grid Search, we could find the best nodel for our data.



- GitHub URL: https://github.com/AndreiToma89/coursera-capstone-project/blob/master/ML%20prediction.ipynb

Results

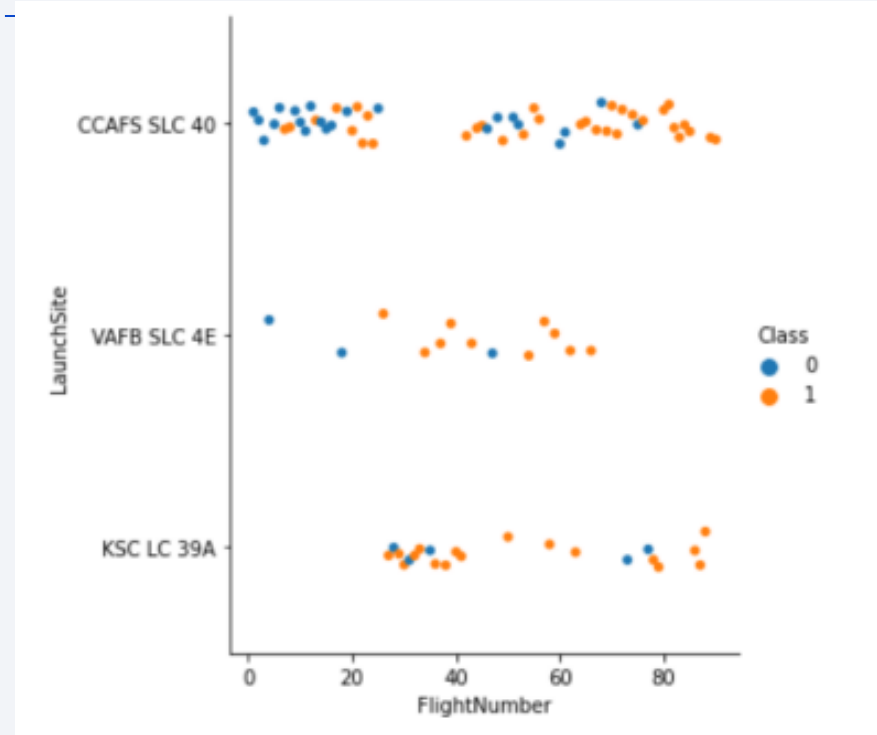
- EDA showed that orbit, launch site, payload and the number of flight were significantly correlated with the outcome.
- All the models had almost the same accuracy, having the highest accuracy of the models at 83%



Section 2

Insights drawn from EDA

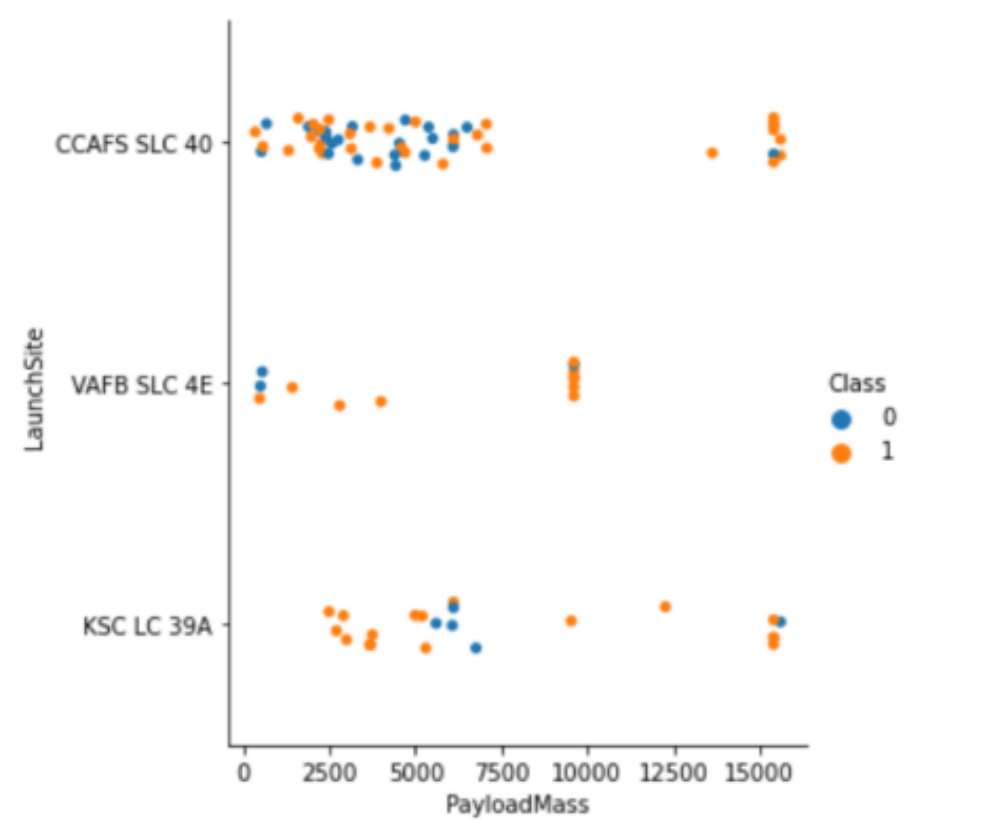
Flight Number vs. Launch Site



There is a string correlation between the number of flights and the success rate

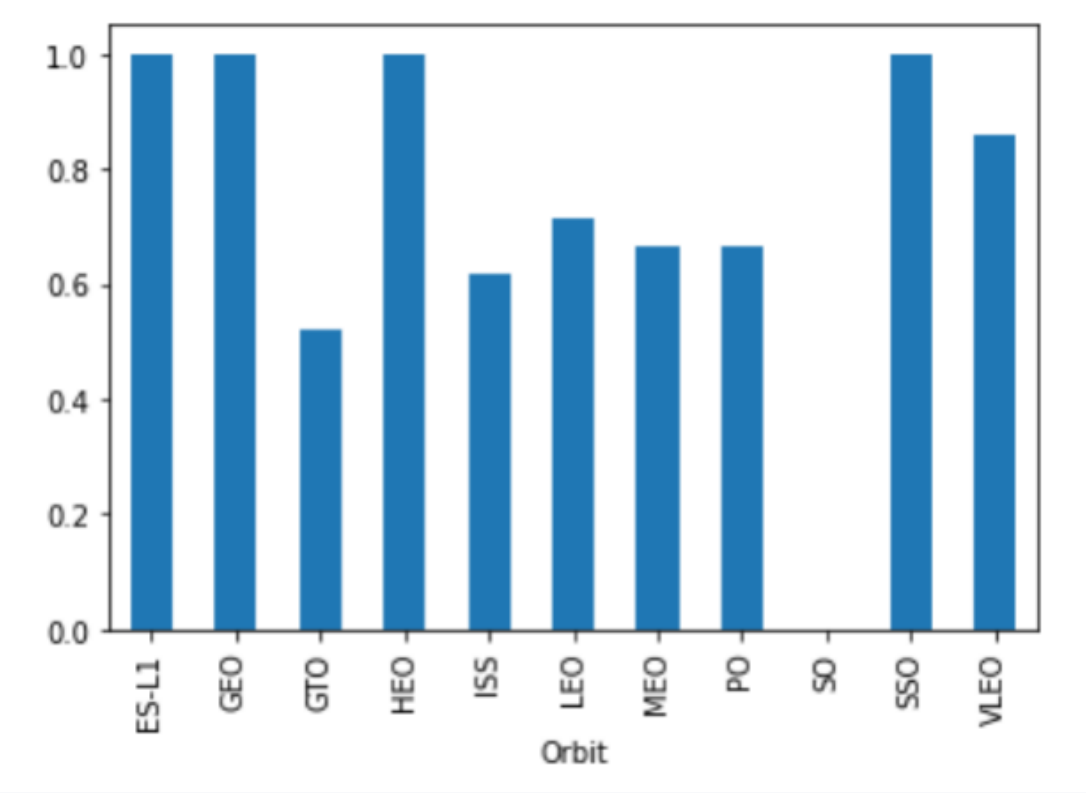
We have no connection between launch location and the success rate

Payload vs. Launch Site



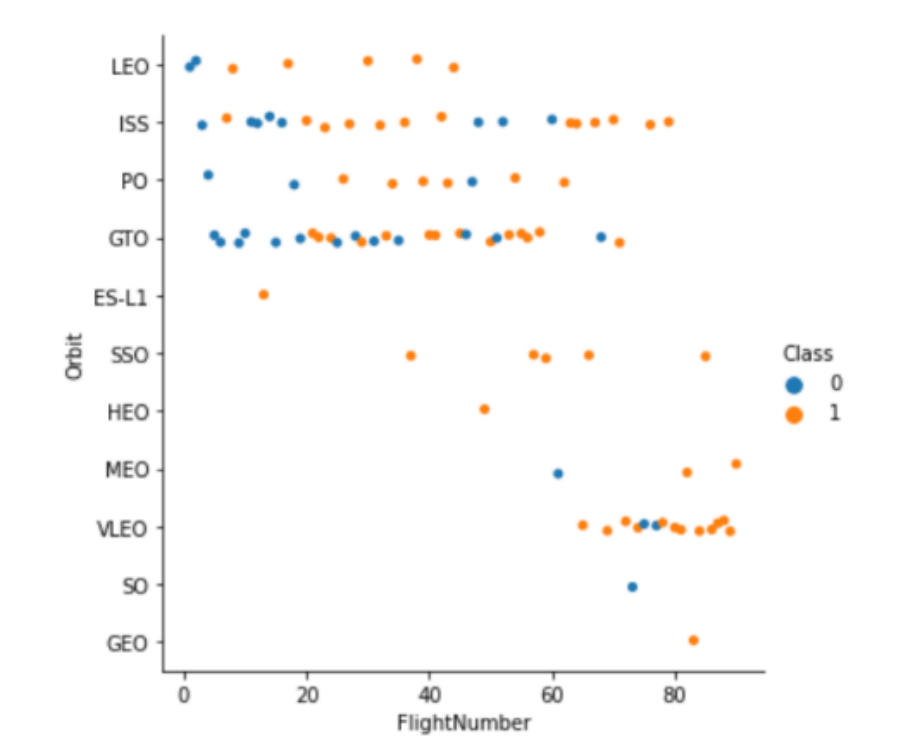
- There's a slight correlation between payload and outcome

Success Rate vs. Orbit Type



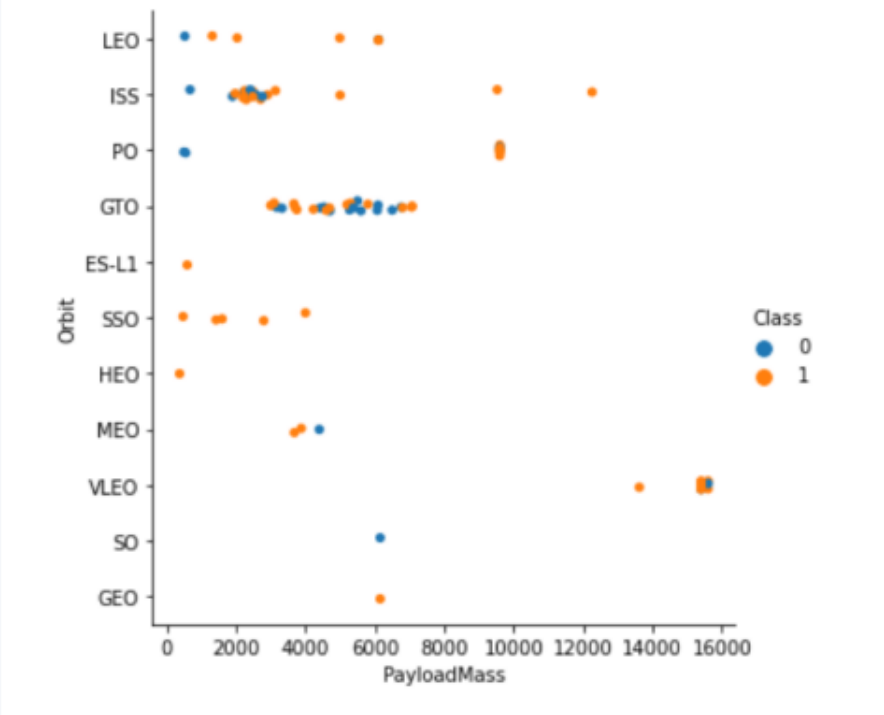
- The graph illustrates the success rate by orbit. We can observe that few of them have a high success rate.

Flight Number vs. Orbit Type



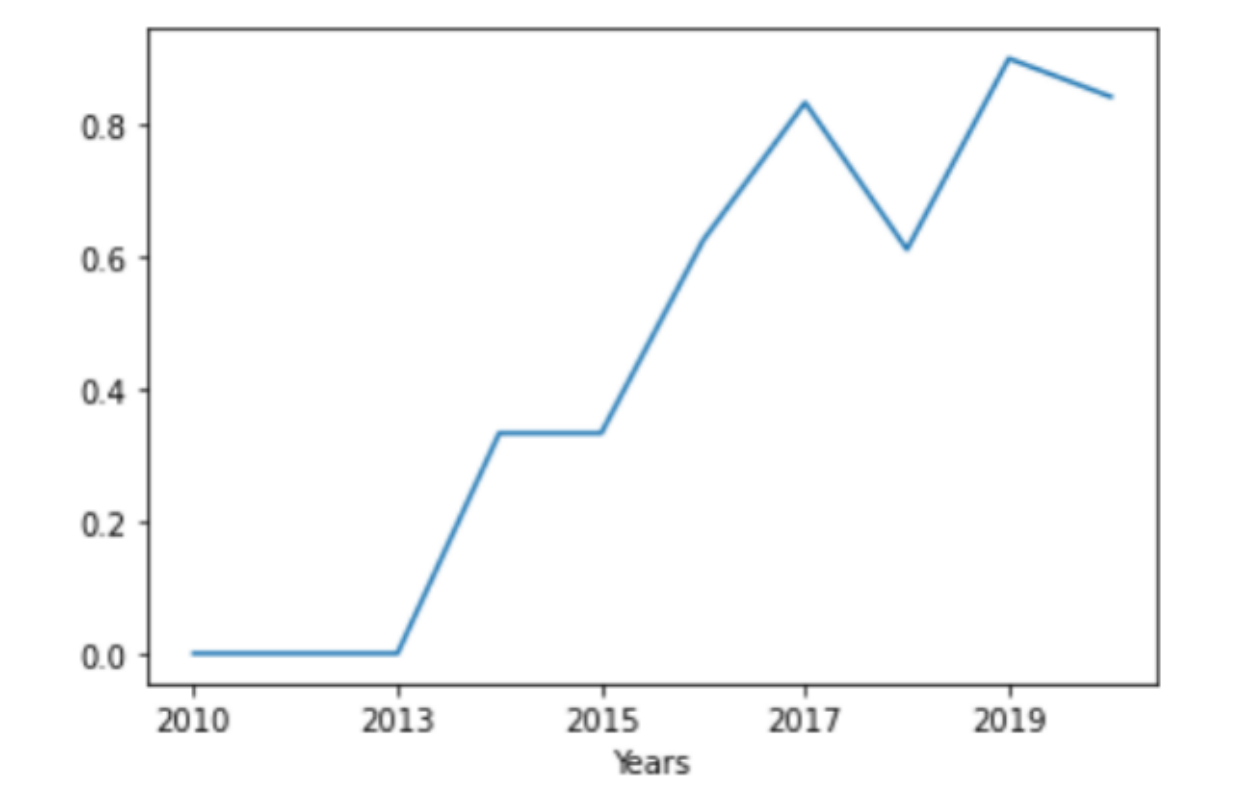
- This graph again shows increased success rate with increased flight number. The VLEO has good success rate partly due to increased flight number

Payload vs. Orbit Type



- The plot shows the success rate per PayloadMass and orbits. There's an invers correlation between pay load and success rate.

Launch Success Yearly Trend



- As expected, the success rate is higher as the time goes.

All Launch Site Names

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

These are the four launch sites

Launch Site Names 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

Above we have the top 5 rows of the launch site "CCA"

Total Payload Mass

- The total payload carried for NASA was 45596 kg. This might be an important figure which can be asked while estimating cost and profitability

Average Payload Mass by F9 v1.1

- The average payload mass carried by F9 v1.1: 2534 kg. The new version has improve payload carrying capacity.

First Successful Ground Landing Date

- The date of the first successful landing was: 2015-12-22.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

In this list we have the boosters with a payload between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

The result of each mission outcome is presented below.

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

Boosters Carried Maximum Payload

booster_version
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

These are the names of the boosters carried with maximum payload mass

2015 Launch Records

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

- This is how 2015 was for us in terms of landing outcomes, boosters and launch sites.

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

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

- The total count of landing outcomes in for dates between 2010-06-04 and 2017-03-20

Section 4

Launch Sites Proximities Analysis

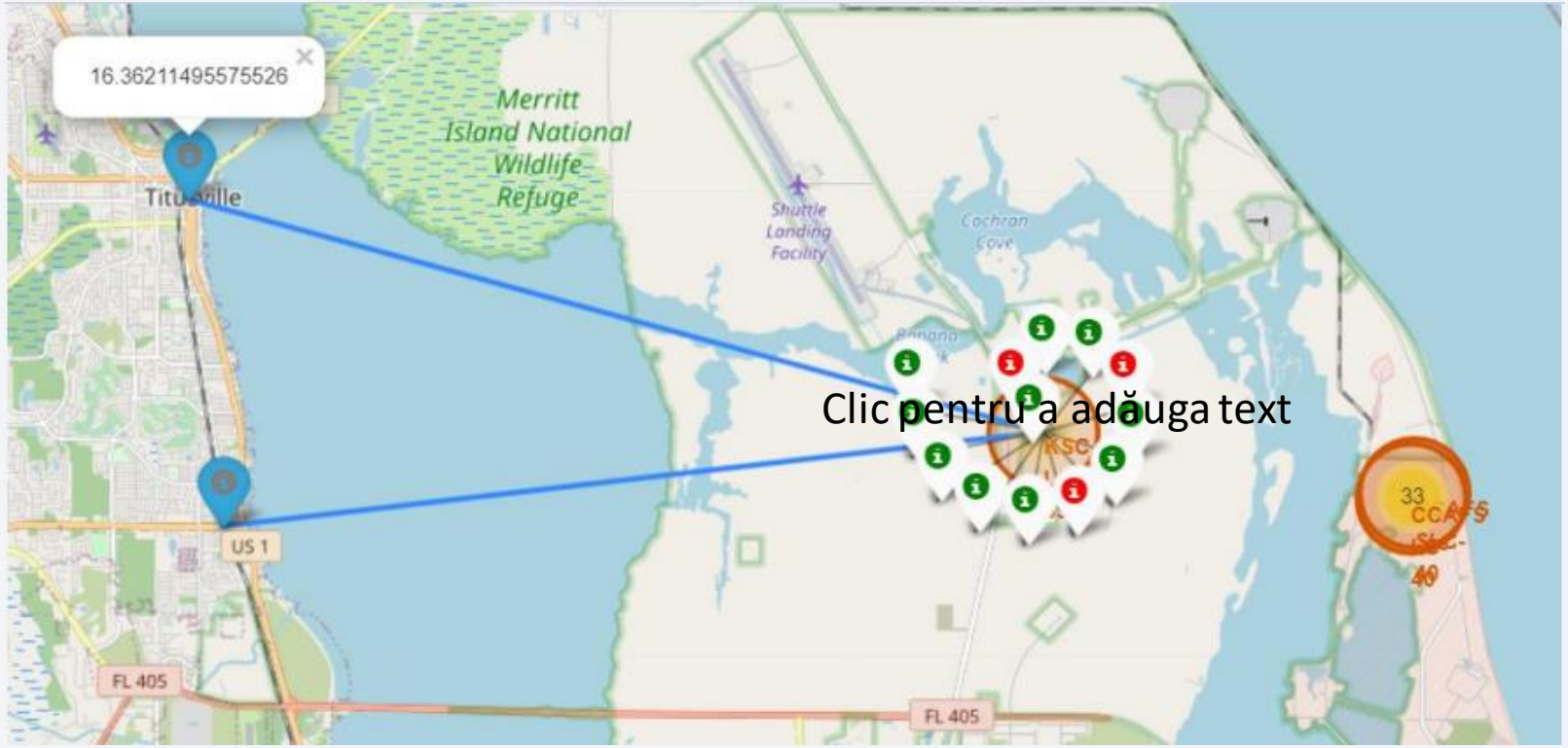
<Folium Map Screenshot 1>

Here we have displayed all our launch sites

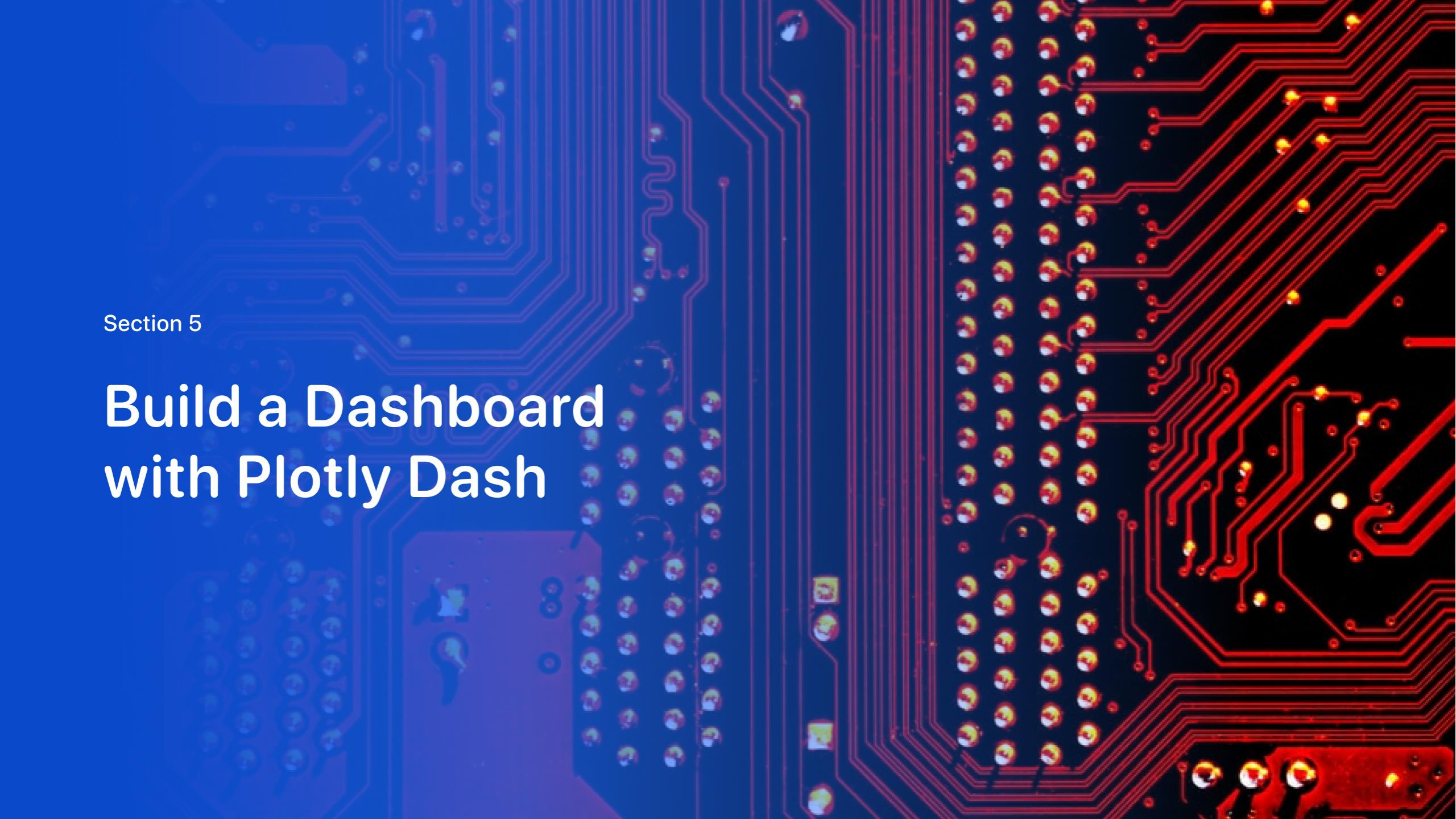
<Folium Map Screenshot 2>

- As we can see, 46 launches were made in the east while only 10 in west

<Folium Map Screenshot 3>



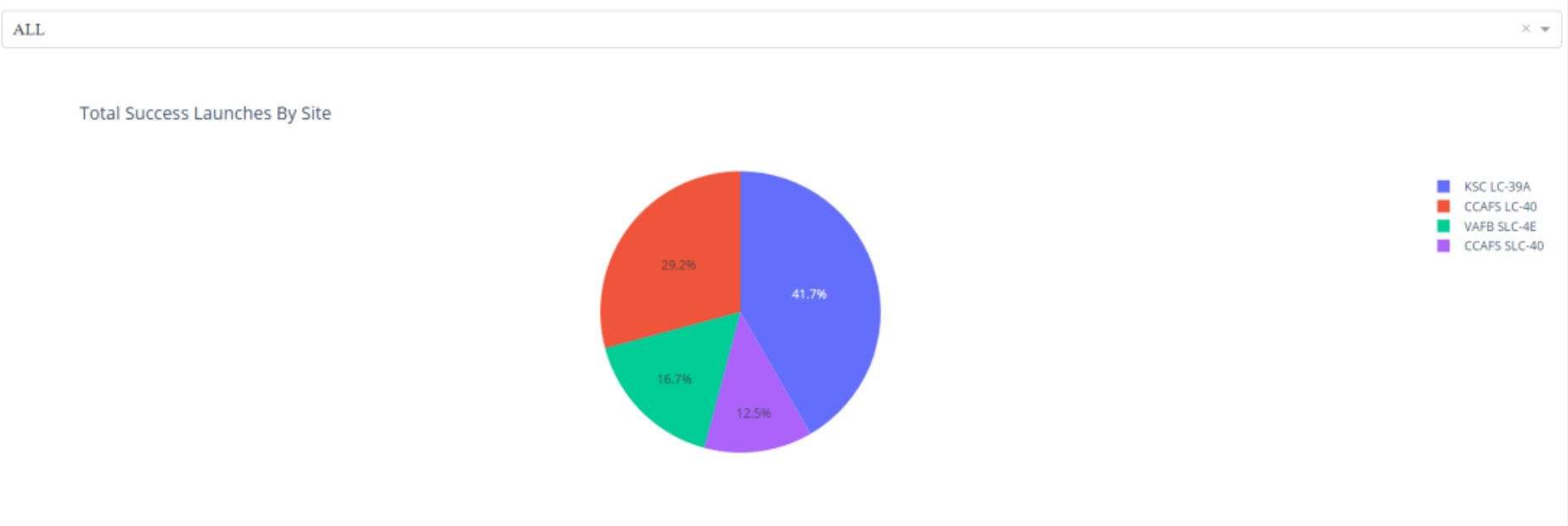
Above you can see which sites hada successful landing (in green) and which not(in red)



Section 5

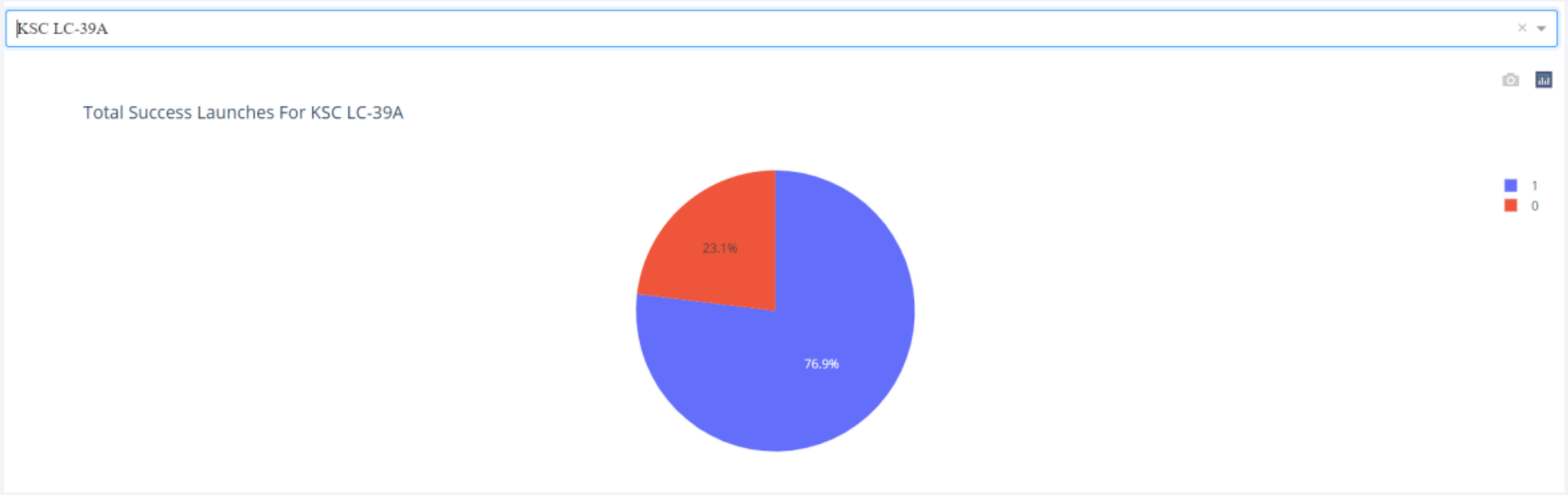
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



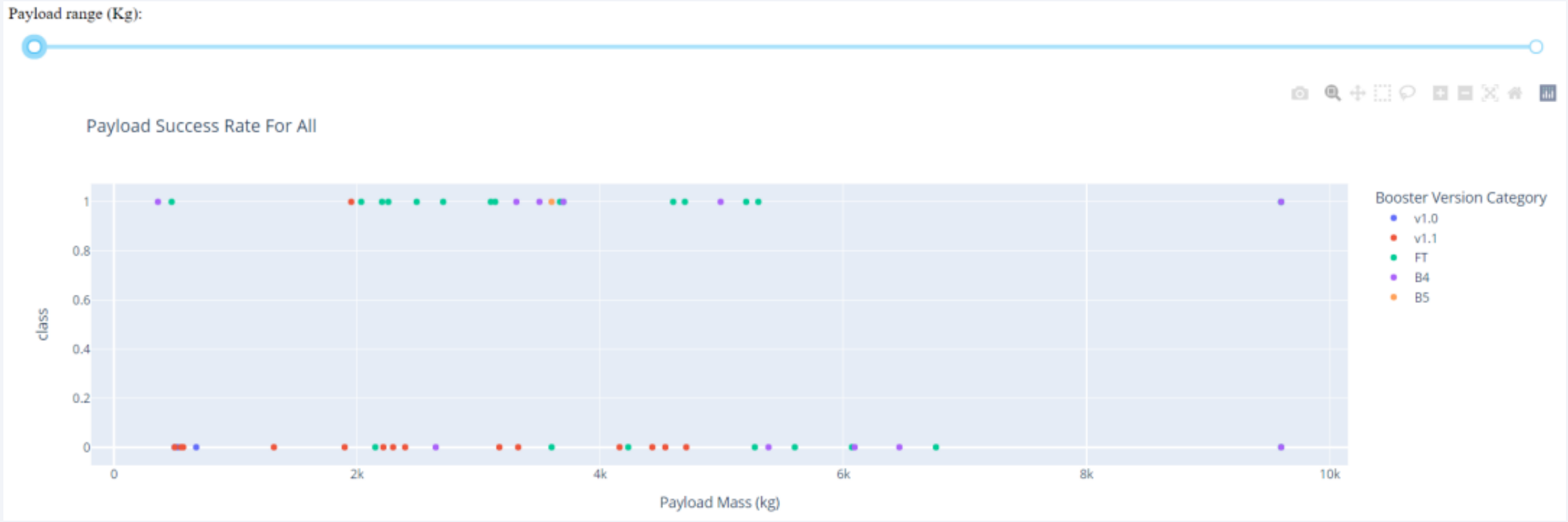
In the above piechart we have the succesful rate per each launch site

<Dashboard Screenshot 2>



- The above piechart displays the success rate only for one launch site

<Dashboard Screenshot 3>

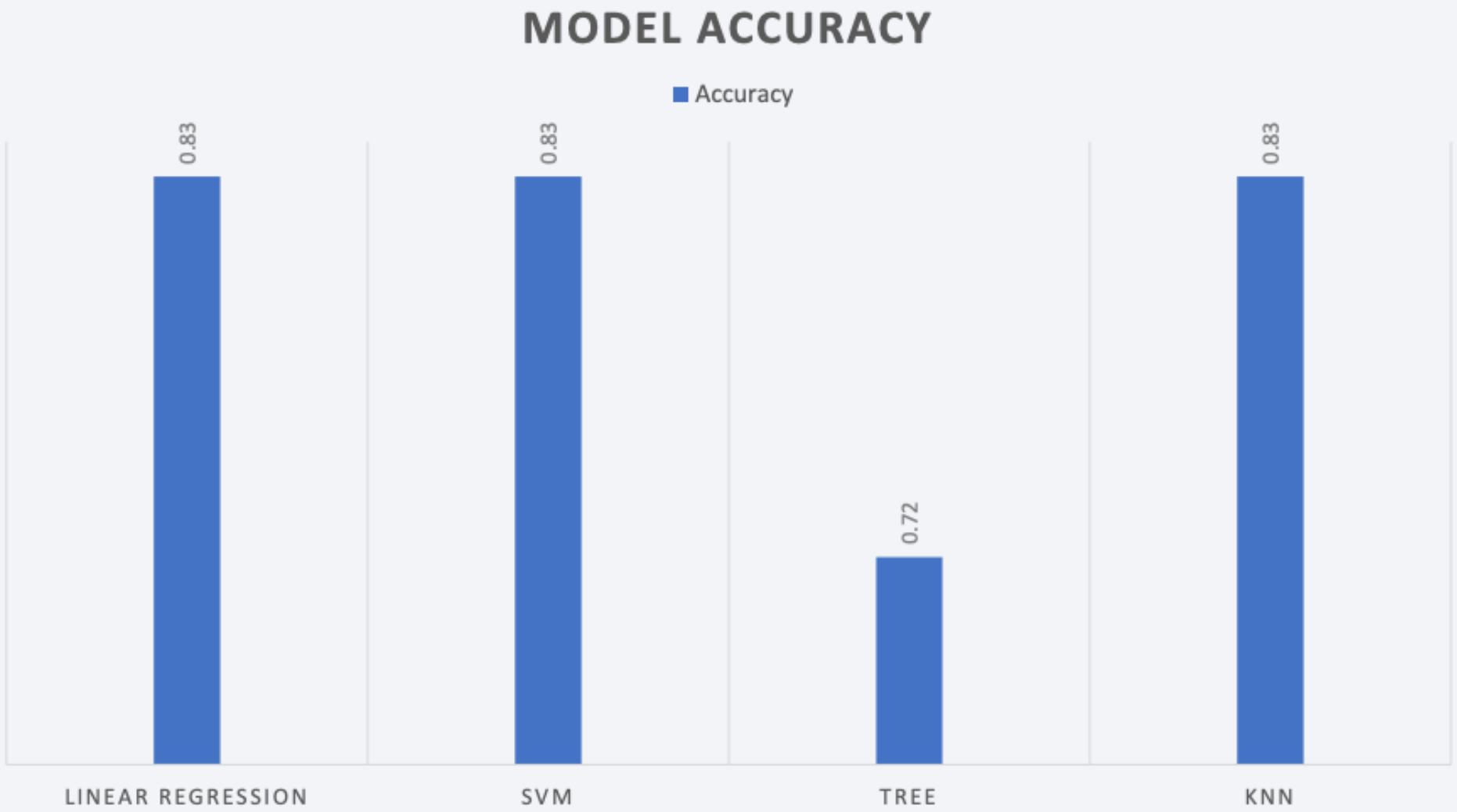


Payload vs success rate plot

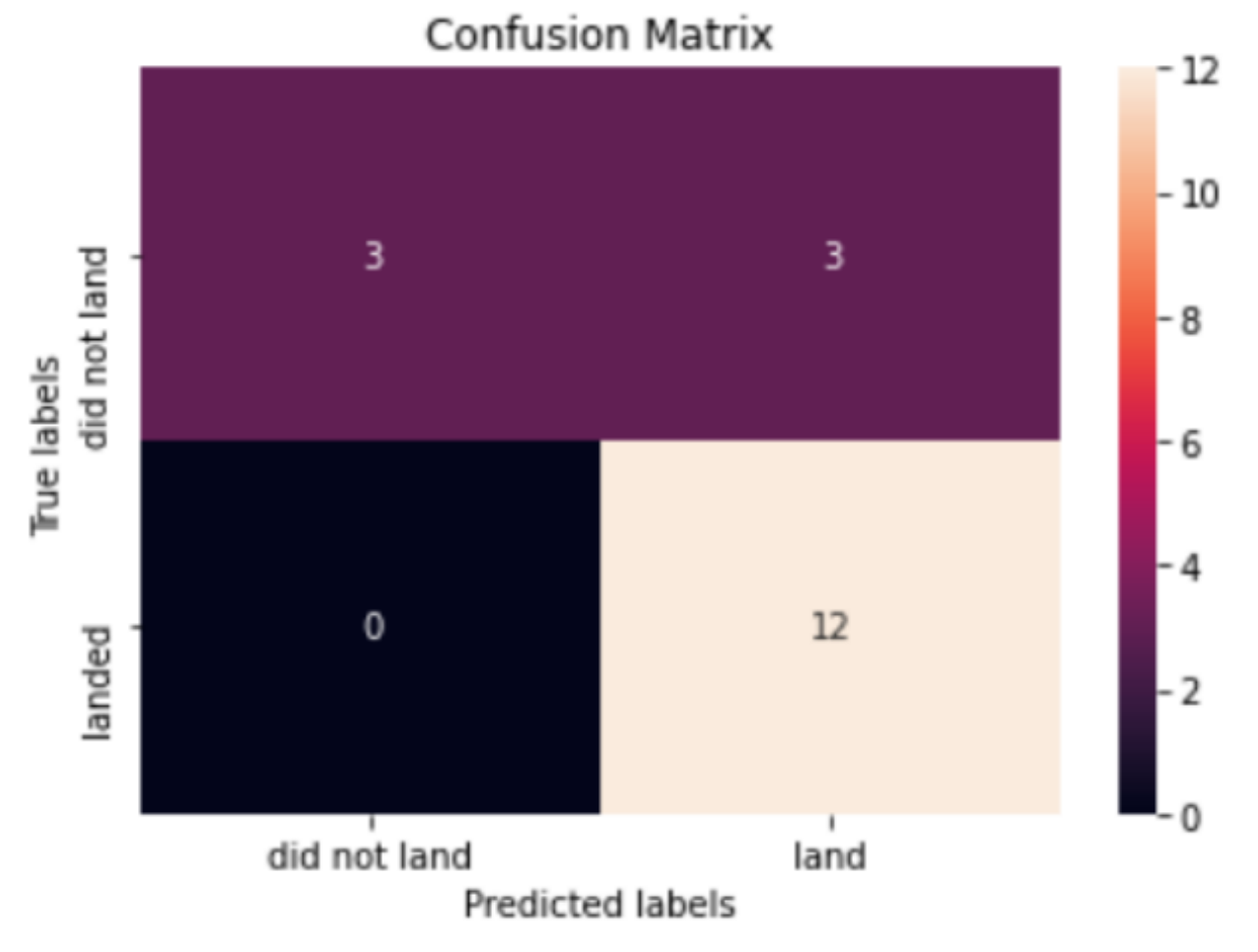
Section 6

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



- Confusion matrix showing true positive, false positive, true negative and false negative in numbers

Conclusions

- With predictive modelling technique, we predicted success rate of SpaceX launch outcome with 83% accuracy.
- The insights obtained were

1. lower payload increased the success rate. However, payload is also linked to profitability. We have to come up with a optimal payload for acceptable success rate (assuming that rest of the parameters are constant). Another way is to retest the hypothesis with new technology and invasion

2. effect of orbits on success rate

3. launch site related features. Are there any room for improvement?

Appendix

- All the code and notebooks are available on :
- https://github.com/AndreiToma89/coursera-capstone-project

Thank you!