IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Valentyn Lazepka
18/02/2023

Outline

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

Executive Summary

General aim of our study was to define the best conditions which provide successful launch and landing of Falcon 9 space carrier.

For our study we've used traditional, well-known and time-tested approaches such as:

- Data Collection through API
- Data Collection with Web Scrapping
- Data Wrangling
- Data Analysis with SQL and Data Visualization
- Visual Analytics
- Machine Learning Prediction

As a result of our study we've defined the best ML approach to create a model and created the model to determine the price of each launch. Also we've determined key factors that influence on successful of each launch. According to our research the most precise result can be received using Decision Tree model.

Introduction

We are living in the time when commercial space launches has became a reality. Among other companies SpaceX stands out with it’s technology of reusing the first stage of the rocket. Because:

- not each launch is successful
- The company use different launch pads and landing places

there is a problem of determination of the best conditions which increase a probability of a first stage accident-free landing. It will give a possibility to forecast a cost of each launch.

During our research we tried to find the answers on the following problems:

- to determine the price of each launch
- to determine if SpaceX will reuse the first stage

Section 1

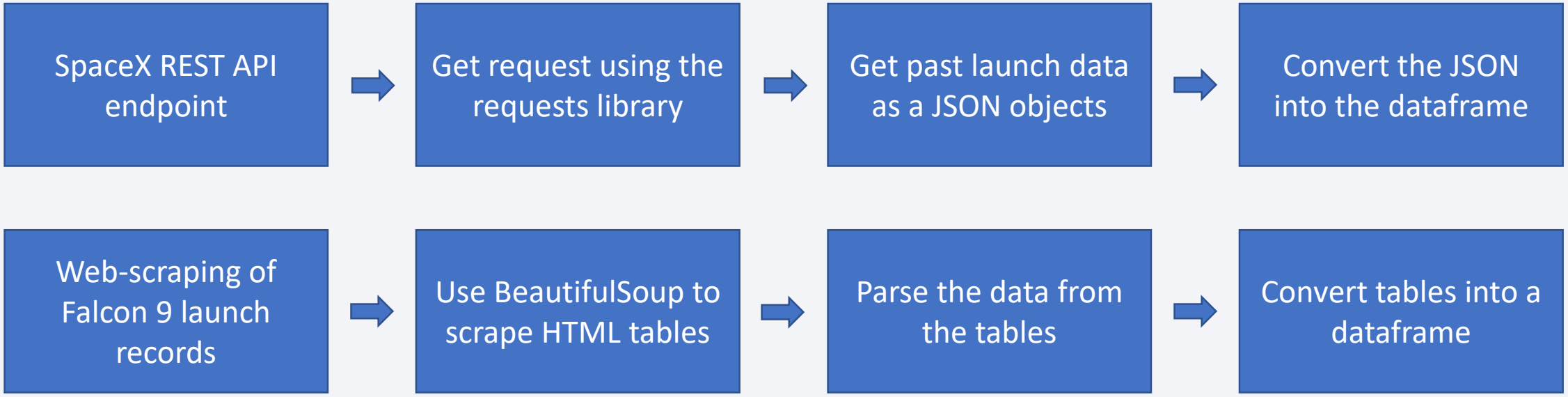
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - WEB-scraping from WIKI pages
  - SpaceX Rest API
- Perform data wrangling
  - Data were collected in JSON and HTML formats, converted into Pandas dataframe for further visualization and analysis
- 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’ve built different classification models for predictive analysis and select the most accurate

Data Collection



Data Collection – SpaceX API

Collect the data:

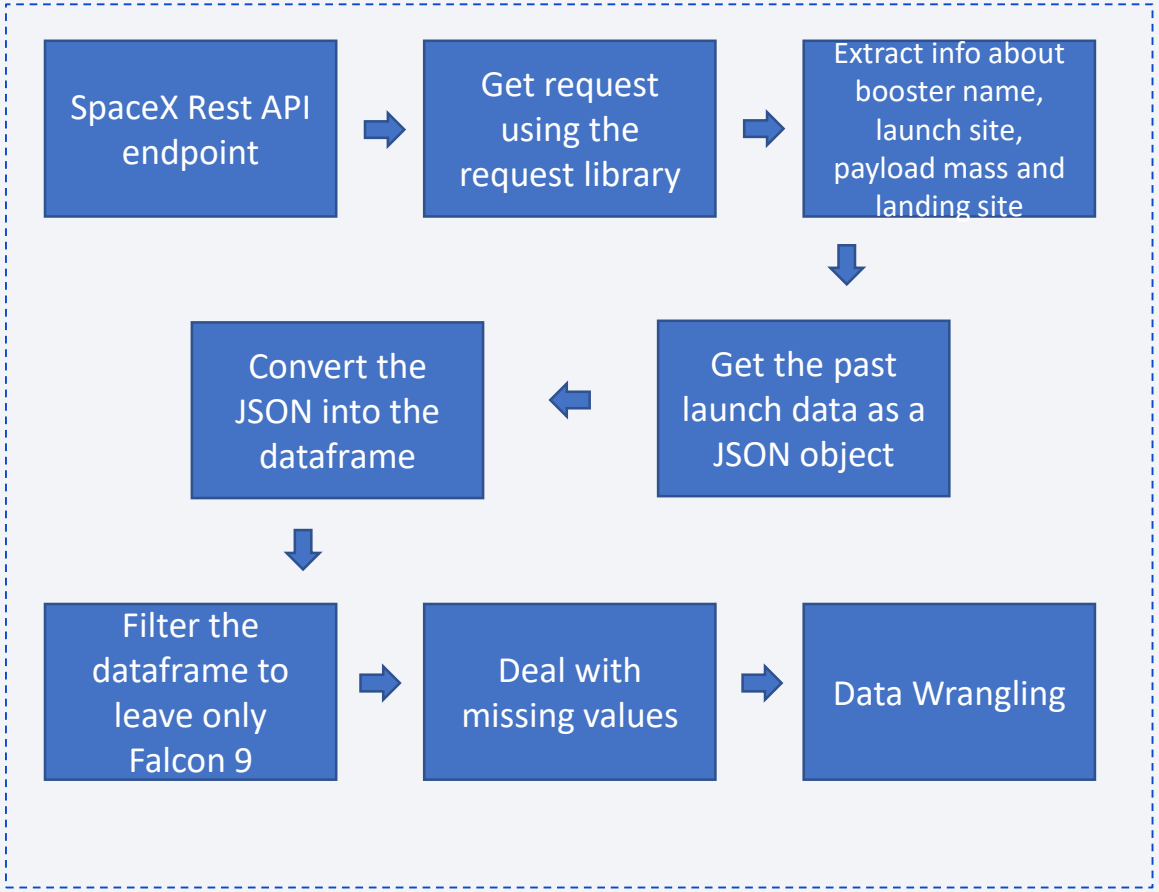
```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url
```

Make sure the requested JSON results are correct.

We decoded the response content as a Json using .json() and turn it into a Pandas dataframe
using .json_normalize()

Create and filter the Pandas dataframe to only include Falcon 9 data

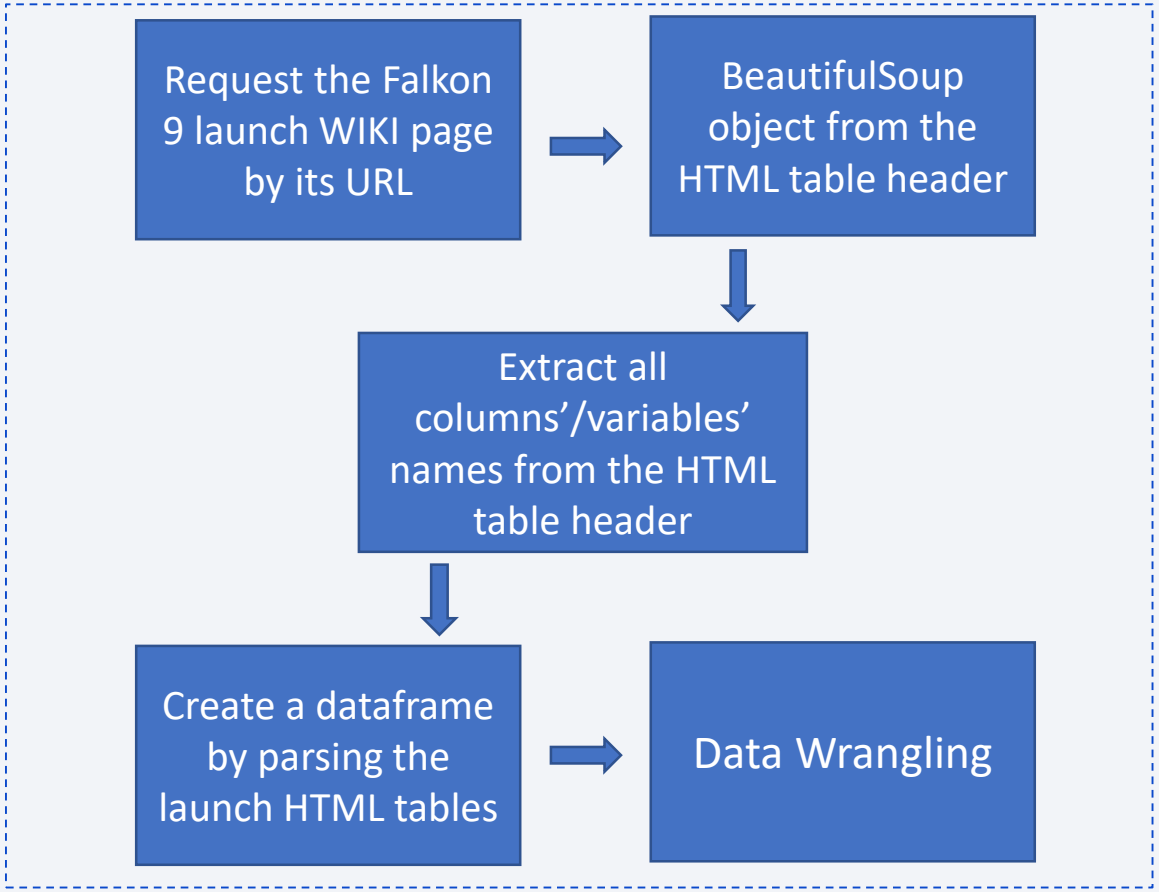
https://github.com/VALENTYN1974/rocket-launch/blob/main/Data%20Collection%20API.ipynb



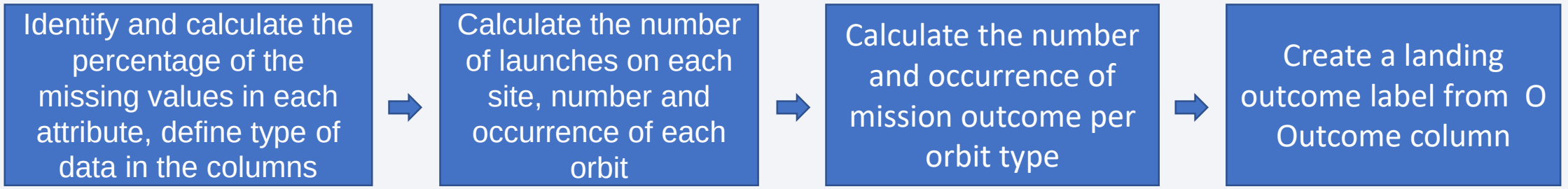
Data Collection - Scraping

- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame

https://github.com/VALENTYN1974/rocket-launch/blob/main/Data%20Collection%20with%20Web%20Scraping%20(1).ipynb



Data Wrangling



Landing outcome label represents the outcome of each launch. One means that the booster landed successfully. Zero means that the booster has crashed.

https://github.com/VALENTYN1974/rocket-launch/blob/main/EDA.ipynb

EDA with Data Visualization

Summary of charts were plotted:

- Catplot to visualize the relationship between Flight Number and Payload
- Catplot to visualize the relationship between Flight Number and Launch Site
- Catplot to visualize the relationship between Payload and Launch Site
- Bar Cart to visualize the relationship between success rate of each Orbit type
- Catplot to visualize the relationship between Flight Number and Orbit type
- Catplot to visualize the relationship between Payload and Orbit type
- Line Chart to visualize the yearly trend of successful launches

https://github.com/VALENTYN1974/rocket-launch/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL (Part 1)

Display the names of the unique launch sites in the space mission

%sql select distinct LAUNCH_SITE from SPACEXTBL

Display 5 records where launch sites begin with the string 'CCA‘

%sql select * from SPACEXTBL where launch_site like "CCA%" limit 5

Display the total payload mass carried by boosters launched by NASA (CRS)

%sql select sum(payload_mass__kg_) from SPACEXTBL where Customer like 'NASA%'

Display average payload mass carried by booster version F9 v1.1

%sql select avg(payload_mass__kg_) from SPACEXTBL where booster_version like 'F9 v1.1%'

List the date when the first successful landing outcome in ground pad was achieved

%sql select min(date) from SPACEXTBL where "Landing _Outcome" like "Success (ground pad)“

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

%sql select "Booster_Version", "Landing _Outcome" from SPACEXTBL where "Landing _Outcome" like "Success (drone ship)" and payload_mass__kg_>4000 and payload_mass__kg_<6000

EDA with SQL (Part 2)

List the total number of successful and failure mission outcomes

%sql select count(*) from SPACEXTBL where “Mission _Outcome" like "%Success%" or “Mission _Outcome" like "%Fail%"

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

%sql select "Booster_Version" from SPACEXTBL where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL)

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

%sql select "Booster_Version", "Launch_Site", "Landing _Outcome", from SPACEXTBL where "Landing _Outcome" like "Failure (drone ship)" and substr(Date,7,4)=‘2015’

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

%sql select "Landing _Outcome", count("Landing _Outcome") from SPACEXTBL where date between '04-06-2010' and '20-03-2017' and "Landing _Outcome" like “%Success%’

https://github.com/VALENTYN1974/rocket-launch/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb

Build an Interactive Map with Folium

Map objects that was created and added to a folium map:

- Folium Circle and folium Marker to add a highlighted circle area with a text label on a specific coordinate for each launch site on the site map
- MarkerCluster object to simplify a map containing many markers having the same coordinate
- MousePosition on the map to get coordinate for a mouse over a point on the map
- MousePolyLine object to draw a line between a launch site to its closest city, railway and highway

https://github.com/VALENTYN1974/rocket-launch/blob/main/module_3_lab_jupyter_launch_site_location.jupyterl.ipynb

Build a Dashboard with Plotly Dash

Summary of plots/graphs and interactions we've added to a dashboard to perform interactive visual analytics of SpaceX launch data in real-time/

This dashboard application contains input components such as a dropdown list and a range slider to interact with a pie chart and a scatter point chart.

- A launch site drop-down input component. There are four different launch sites and drop-down menu let us select different launch sites
- A callback function to render **success-pie-chart** based on the selected site drop-down. The general idea of this callback function is to get the selected launch site from a site-dropdown and render a pie-chart visualizing launch success counts
- A range Slider to Select Payload. The slider is to be able to easily select different payload range and see if we can identify some visual patterns
- A callback function to render **the success-payload-scatter-chart** scatter plot. To visually observe how payload may be correlated with mission outcomes for selected sites.

https://github.com/VALENTYN1974/rocket-launch/blob/main/spacex_dash_app.py

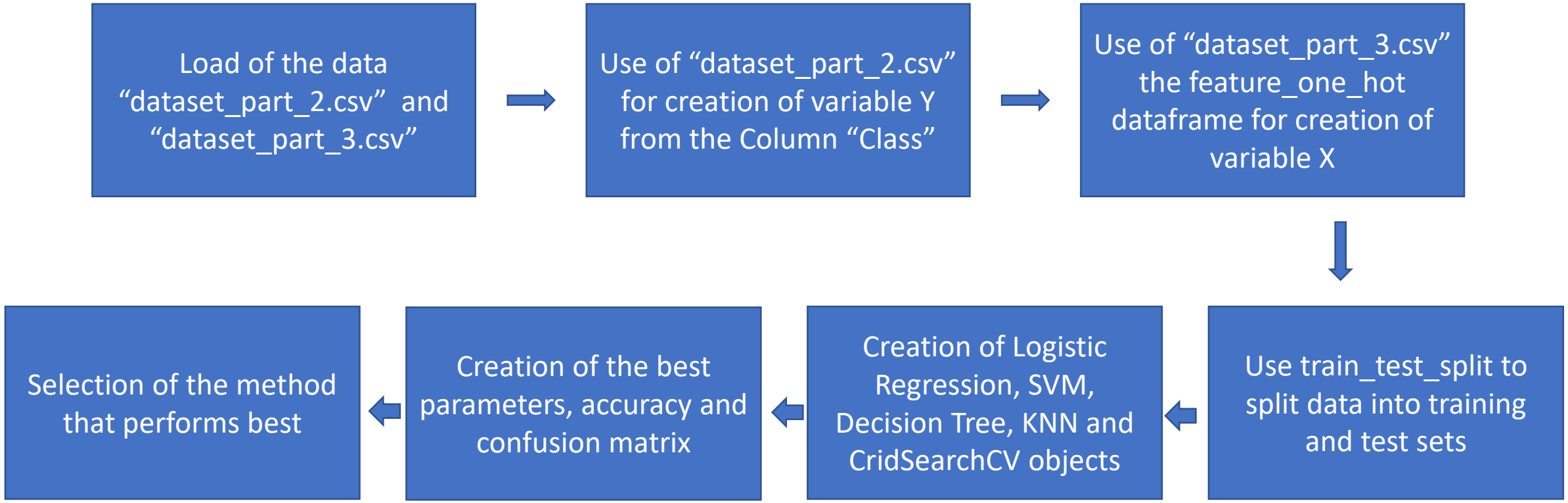
Predictive Analysis (Classification) (Part 1)

Summary of building, evaluation, improvement and founding the best performing classification model:

- Create a NumPy array from the column Class in data, make sure the output is a Pandas series
- Standardize the data
- split the data into training and test data
- Fit the object
- Search for the best hyperparameters for logistic regression, SVM, decision tree and KNN classifiers
- Search for the method that performs best using test data

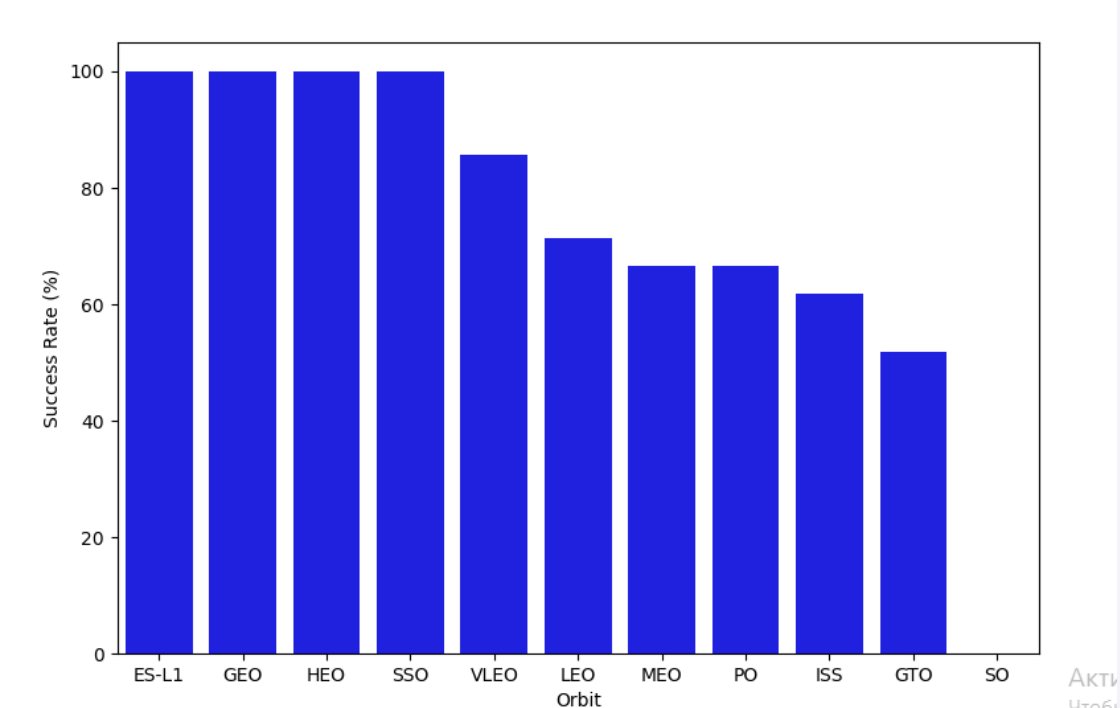
https://github.com/VALENTYN1974/rocket-launch/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Ma.ipynb

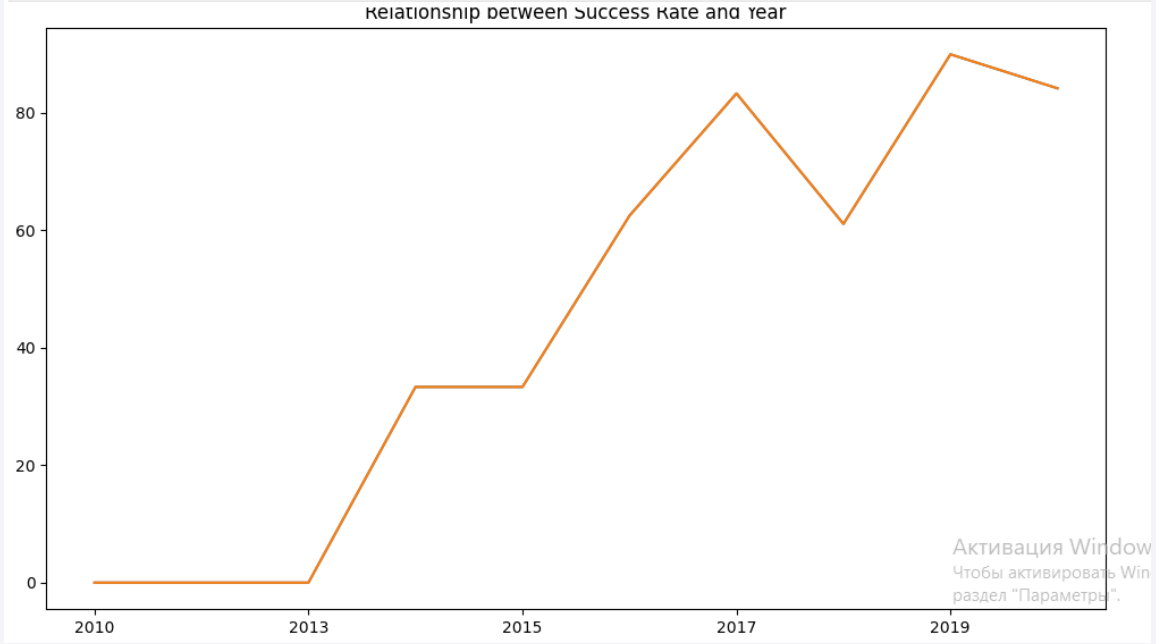
Predictive Analysis (Classification) (Part 2)

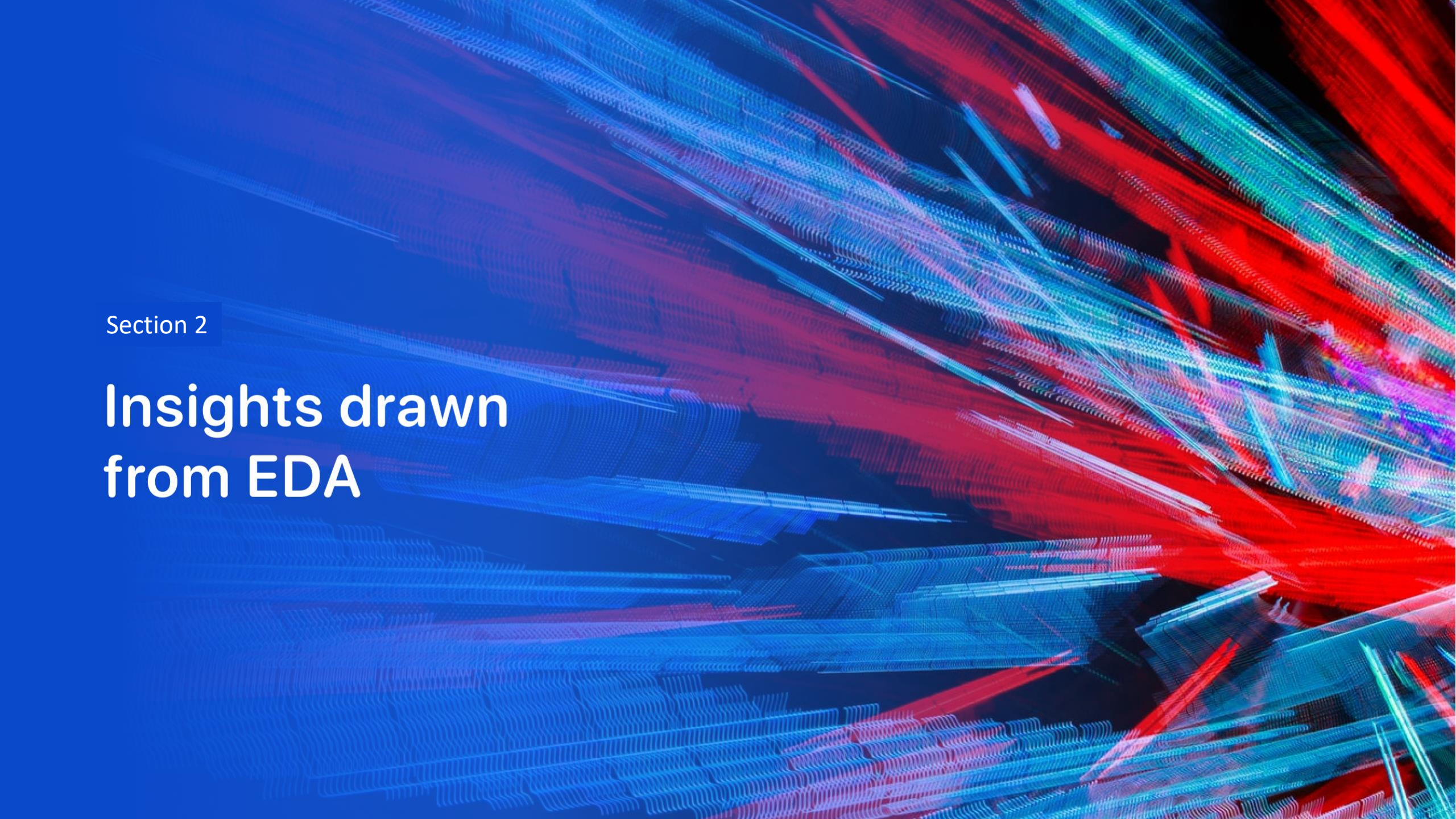


Results

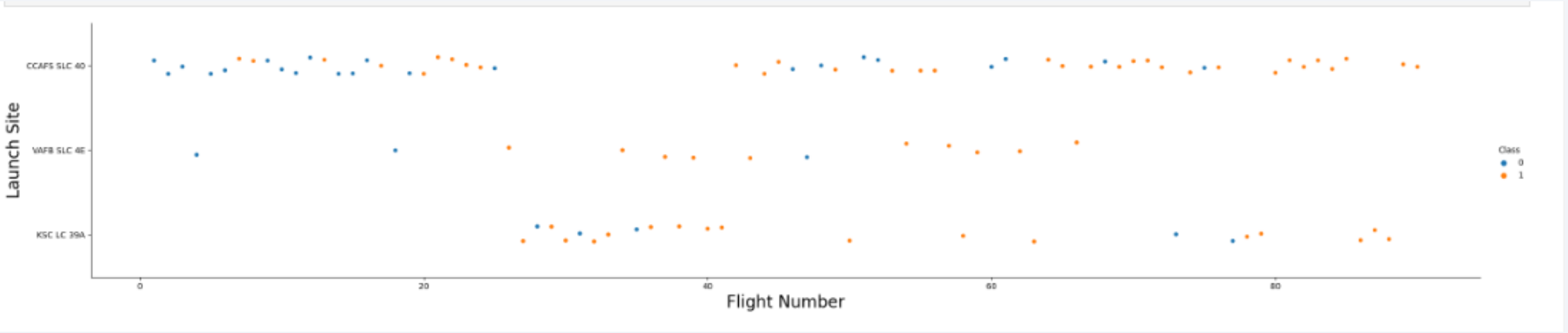
- There were four of the unique launch sites in the space mission, the booster_versions which have carried the maximum payload mass is F9 B5, there were more successful landings on drone ships
- success rate since 2013 kept increasing till 2020; the best success rates have orbits ES-L1, GEO, HEO, SSO;
- Best Algorithm is Decision Tree with a score of 0.8875






Section 2
Insights drawn from EDA

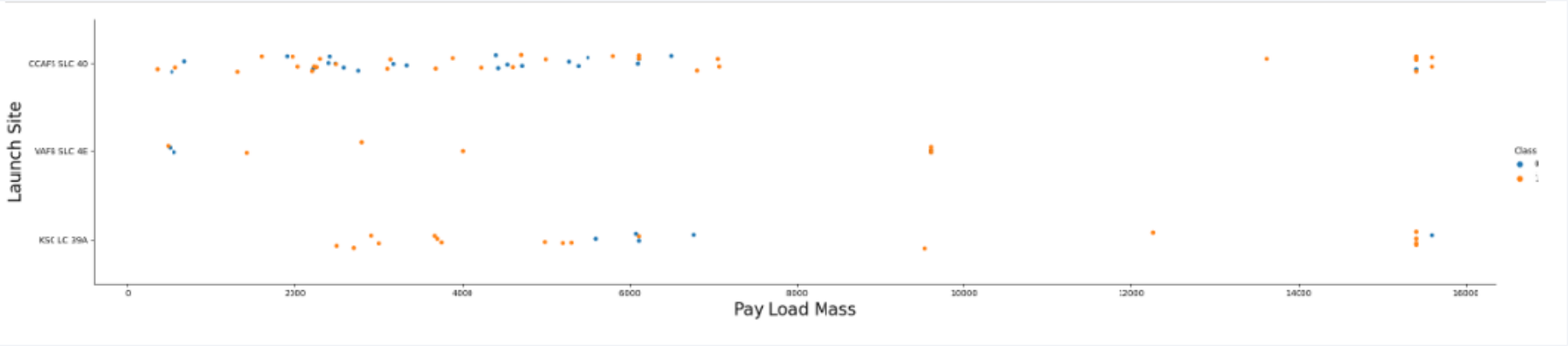
Flight Number vs. Launch Site



With time the success rate has increased for every launch site, especially for CCAFS SLC 40, where are concentrated the majority of the launches.

VAFB SLC 4E and KSC LC 39A has a higher success rate but represents only one third of the total launches.

Payload vs. Launch Site



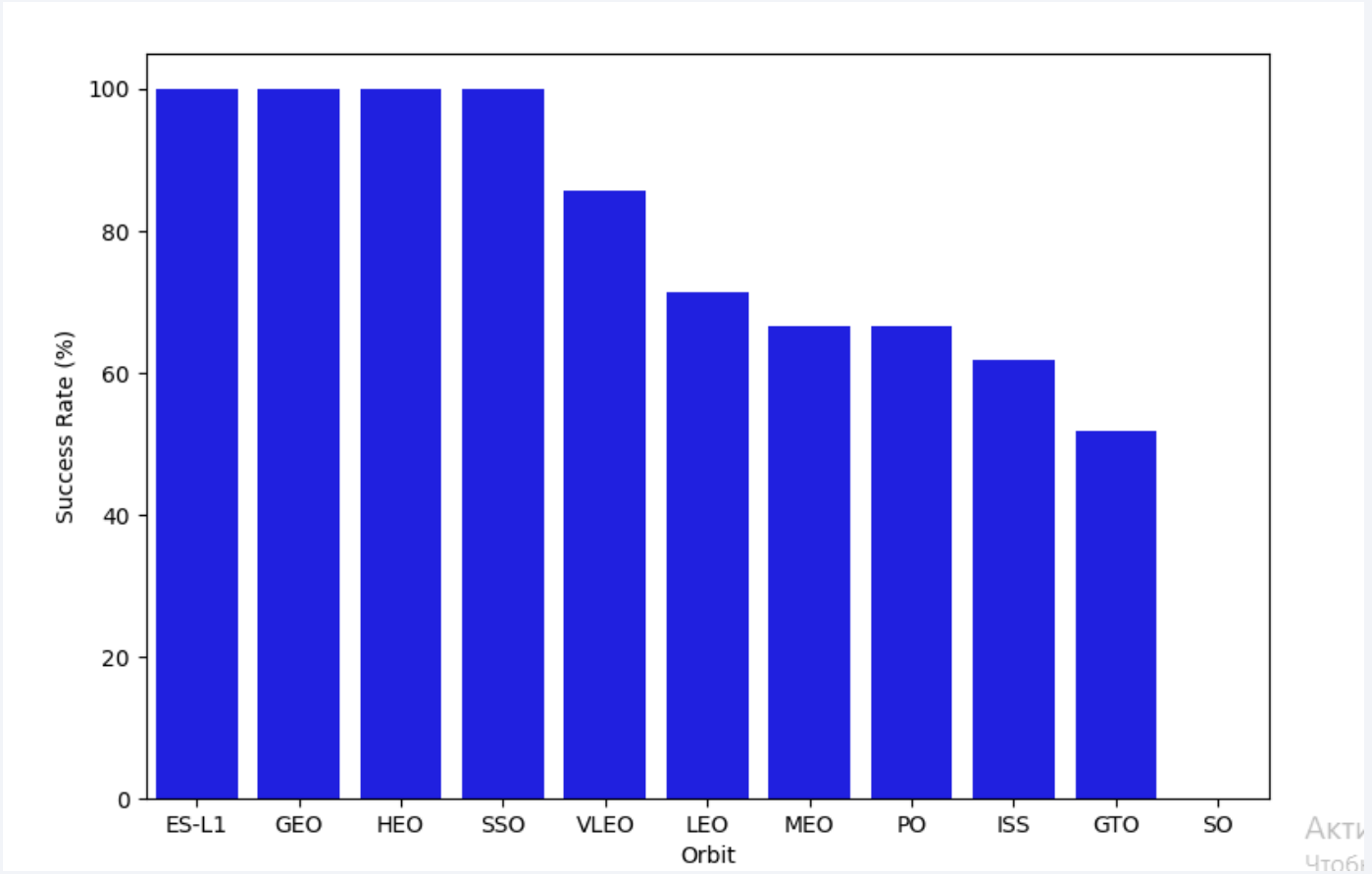
In VAFB-SLC launch site there are no rockets launched for heavy payloadmass (more than 10000 kg)

In KSC LC launch site there are no rockets launched for lower payloadmass (less than 2500 kg)

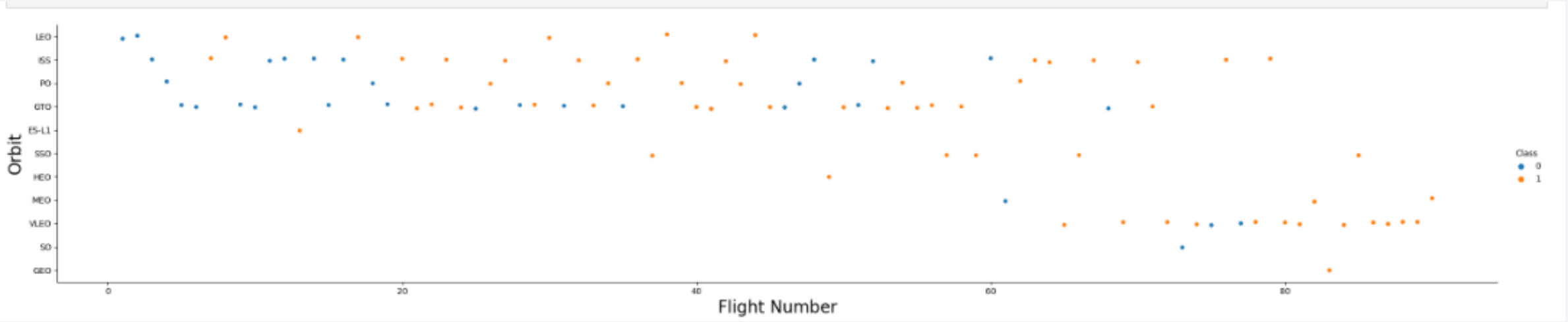
CCAFS SLC has launched rockets less than 7500 kg and more than 13000kg payloadmass but not in between

Success Rate vs. Orbit Type

The best success rates have orbits ES-L1, GEO, HEO, SSO but we don’t know a number of launches from that sites



Flight Number vs. Orbit Type

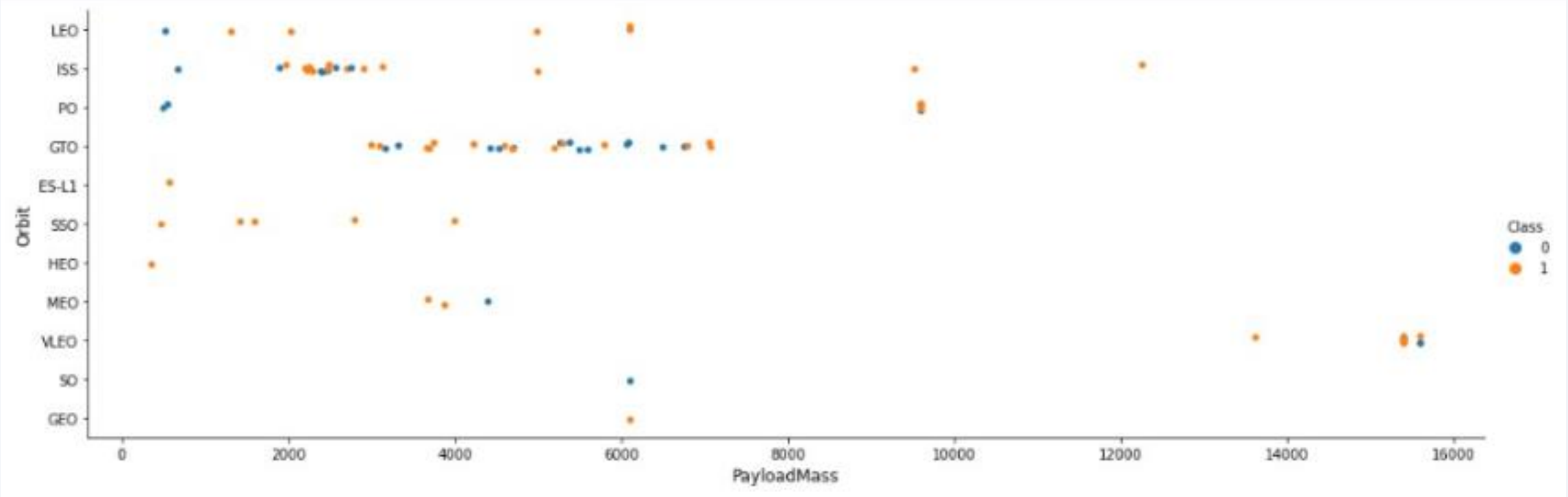


It is not surprised that there are more failures at the beginning of the series of launches but after the first 40 launches the ratio improves by reducing the 50%of unsuccessful landings.

GTO and ISS orbits has the higher concentration of launches with the lowest ratio of successful landings

The orbits with higher successful rate has one or just a few number of launches

Payload vs. Orbit Type



Most number of launches has less then 8000 kg of payload

GTO orbit doesn’t have visible difference between number of successful and failed launches

Heavy payloads are visibly more successful then light

Launch Success Yearly Trend

Success 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

There are four unique launch sites

For request code see slides 12-13

Launch Site Names 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

As we see there are no success landings for launches from this site (among these 5). And all of them have light payload mass

For request code see slides 12-13

Total Payload Mass

sum(payload_mass__kg_)
99980

For request code see slides 12-13

Average Payload Mass by F9 v1.1

avg(payload_mass__kg_)
2534.6666666666665

Average payload mass by F9 v1.1 is higher then average payload mass of all launches (99980/71=1408,169 kg)

For request code see slides 12-13

First Successful Ground Landing Date

min(date)
01-05-2017

SpaceX needed 1 and a half year to achieve first successful ground landing

For request code see slides 12-13

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster_Version	Landing _Outcome
F9 FT B1022	Success (drone ship)
F9 FT B1026	Success (drone ship)
F9 FT B1021.2	Success (drone ship)
F9 FT B1031.2	Success (drone ship)

There are only two booster versions which were used for such payload mass and had successful landing on drone ship

For request code see slides 12-13

Total Number of Successful and Failure Mission Outcomes

count(*)
101

For request code see slides 12-13

Boosters Carried Maximum Payload

Maximum Payload was on B5 version of Falcon 9

For request code see slides 12-13

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

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

For request code see slides 12-13

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

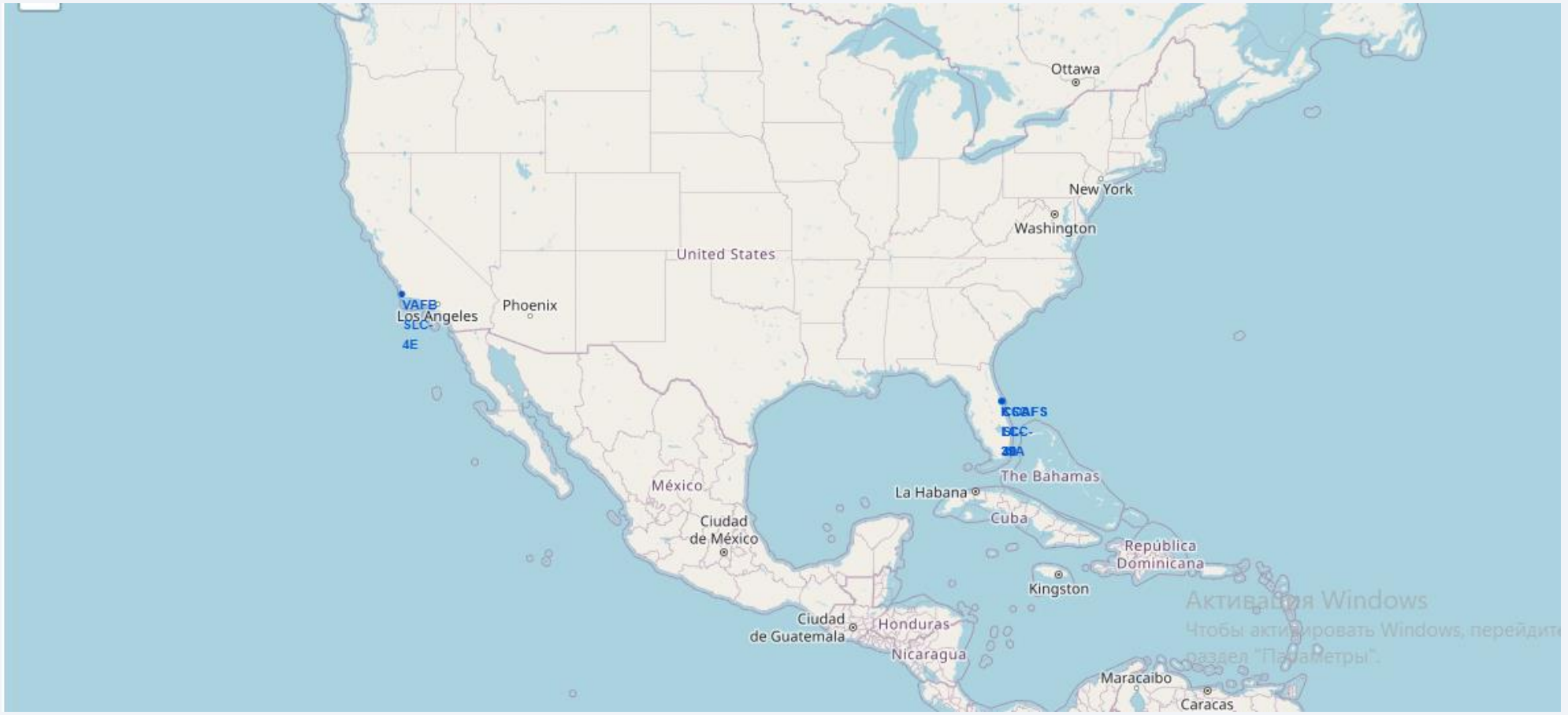
Landing _Outcome	count("Landing _Outcome")
Success	20
Success (drone ship)	8
Success (ground pad)	6

For request code see slides 12-13

Section 3

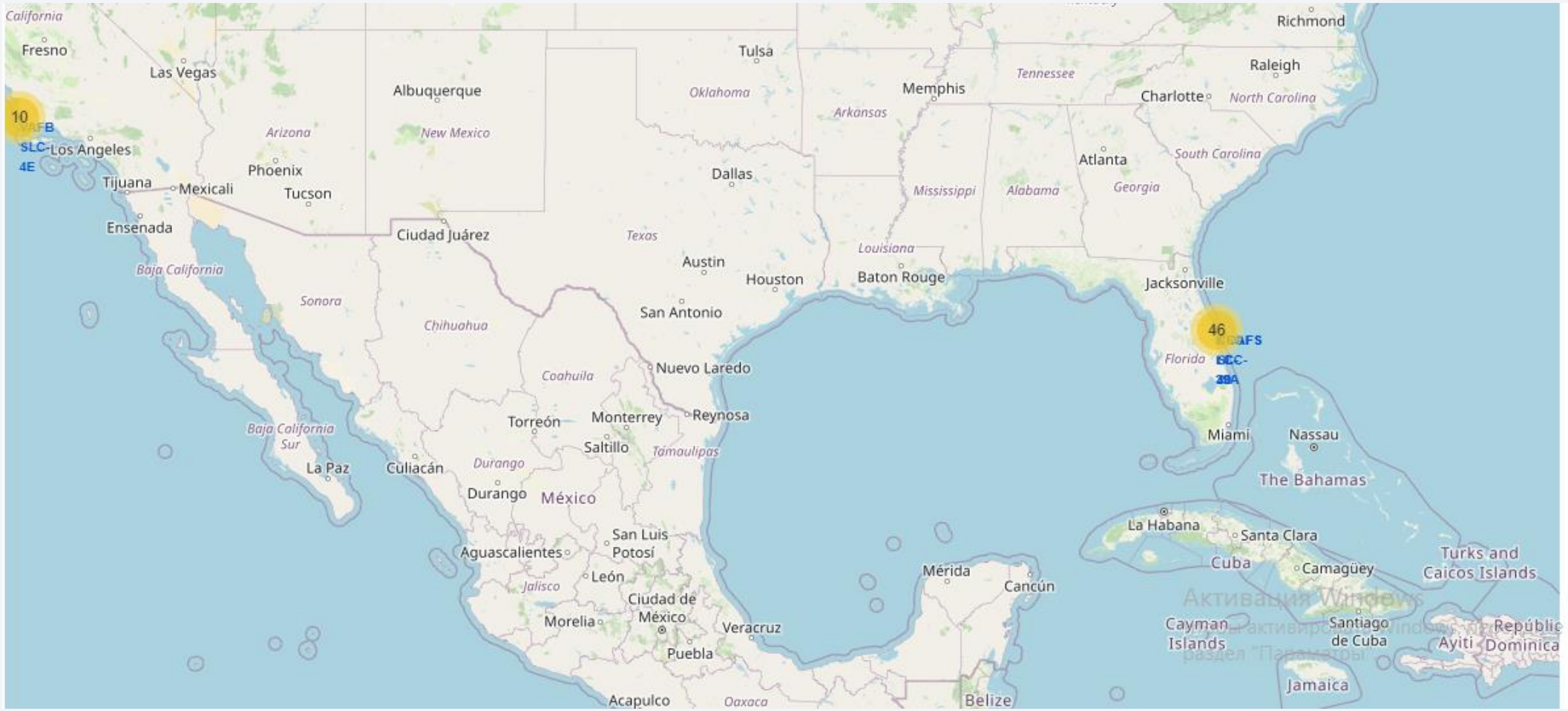
Launch Sites Proximities Analysis

Map with marked launch sites



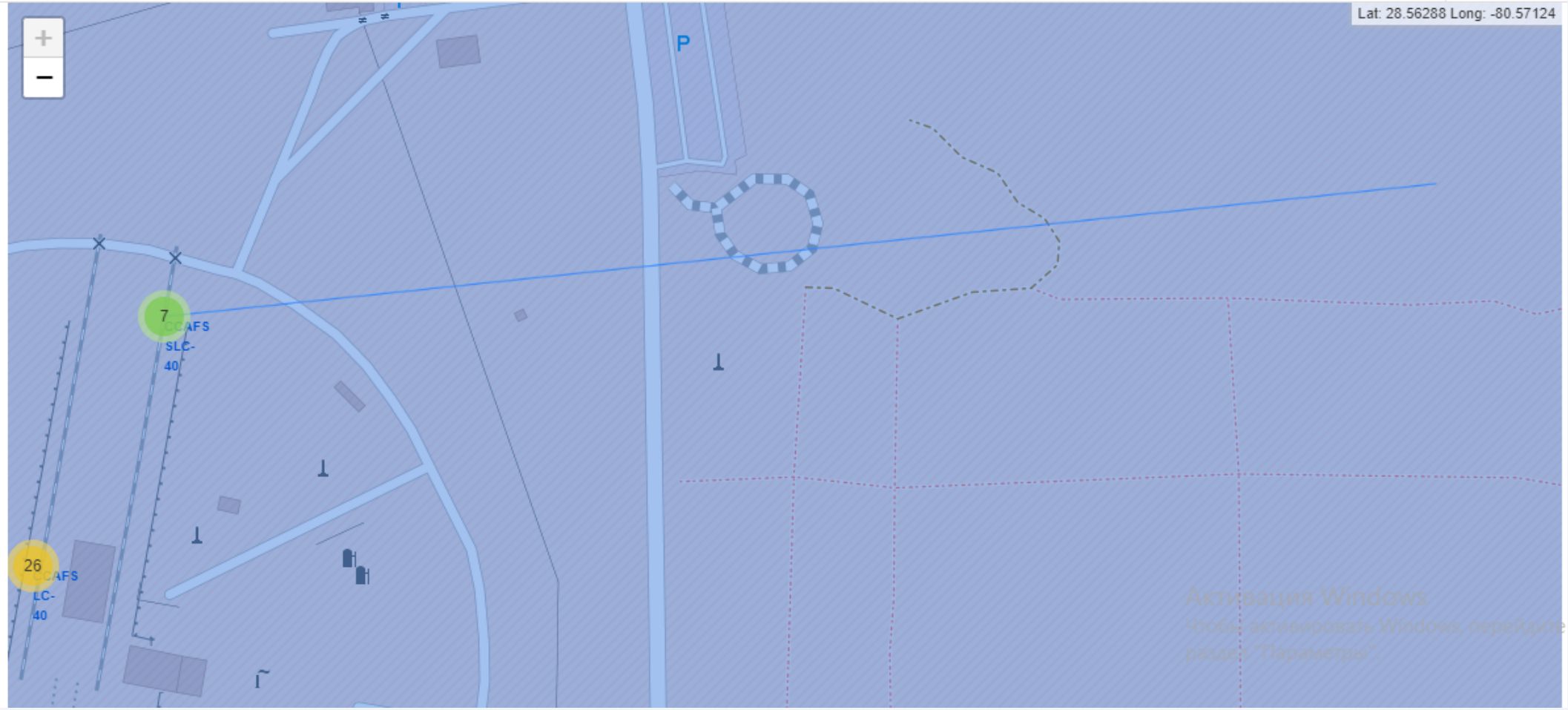
All launch sites are near ocean shores

Launch sites with number of launches

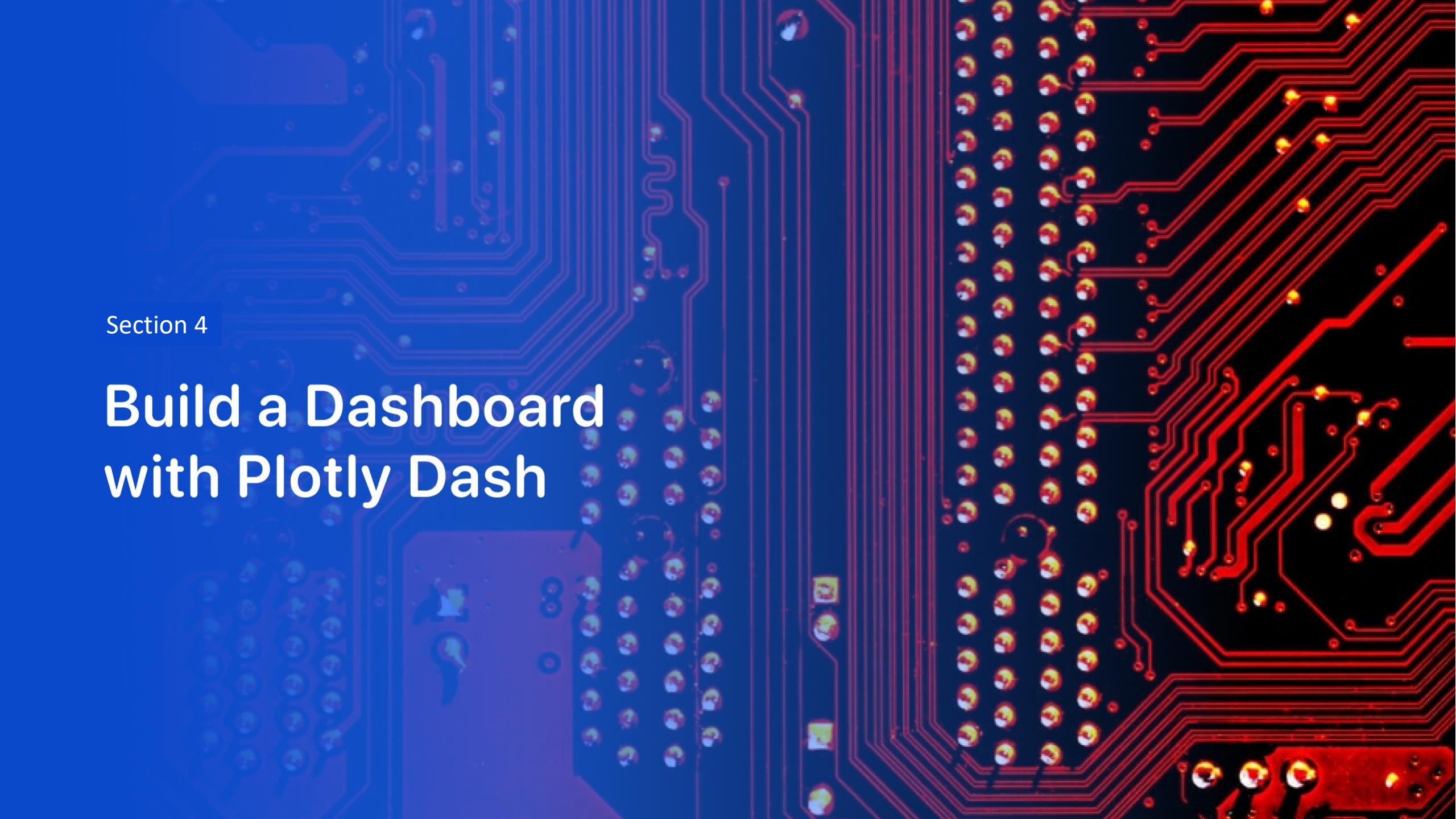


More launches were done from the Atlantic ocean shore

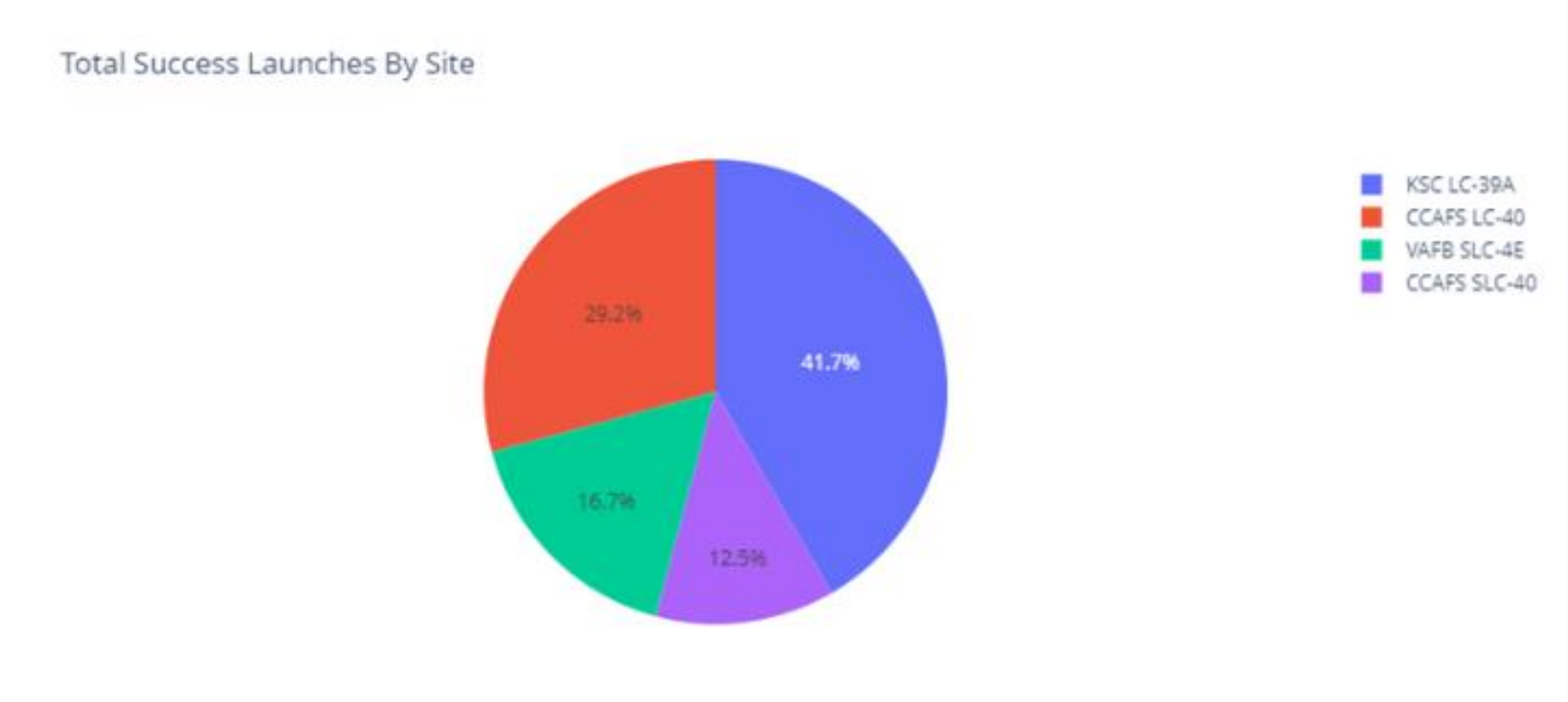
Distance to selected point



- Launches done in uninhabited area

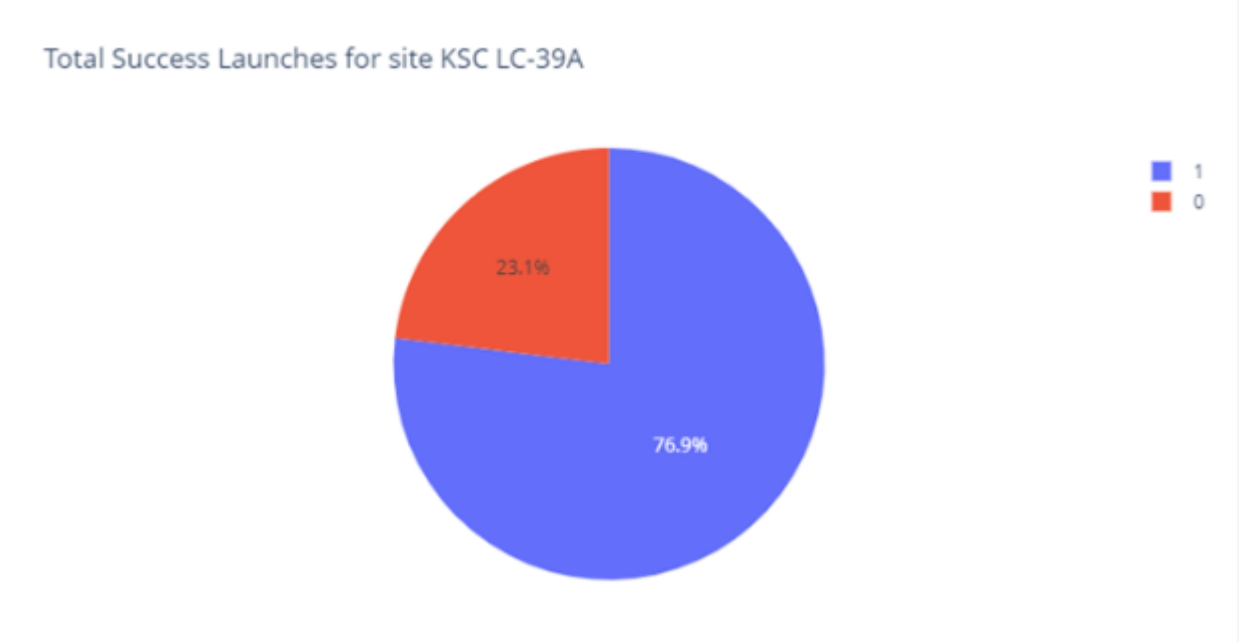

Section 4
Build a Dashboard with Plotly Dash

Total succesful launches for all sites



KSC LC-39A is a site with the most launches

Launch site with highest success ratio



- The most successful launch site has 76.9% of successful launches. Also because it is a site with the most number of launches it can be recommended as a preferable place for launches saving the same payload mass characteristics and booster models

Payload success range

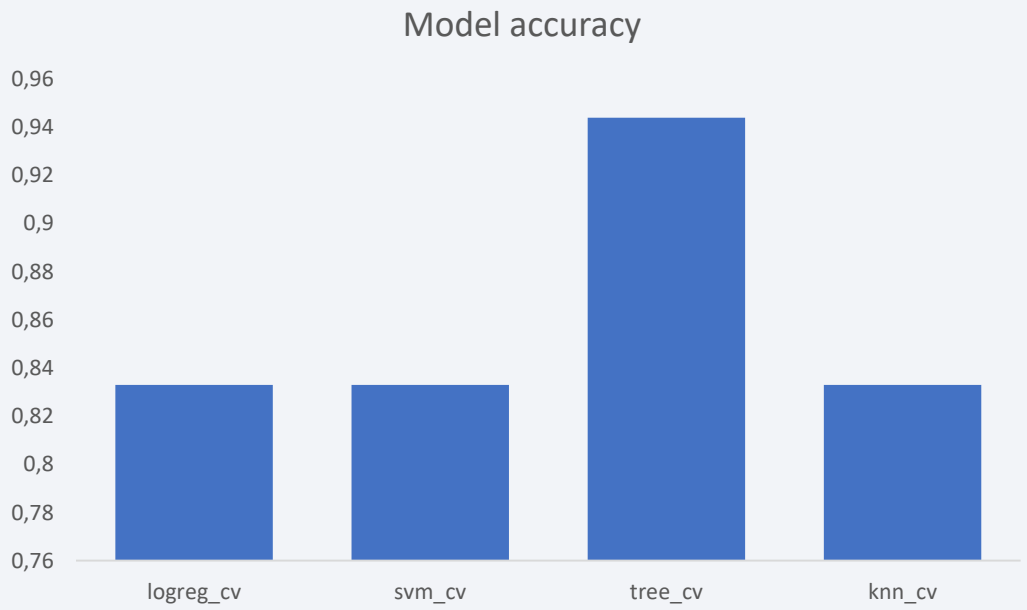
- The largest success rate has FT booster version

Section 5

Predictive Analysis (Classification)

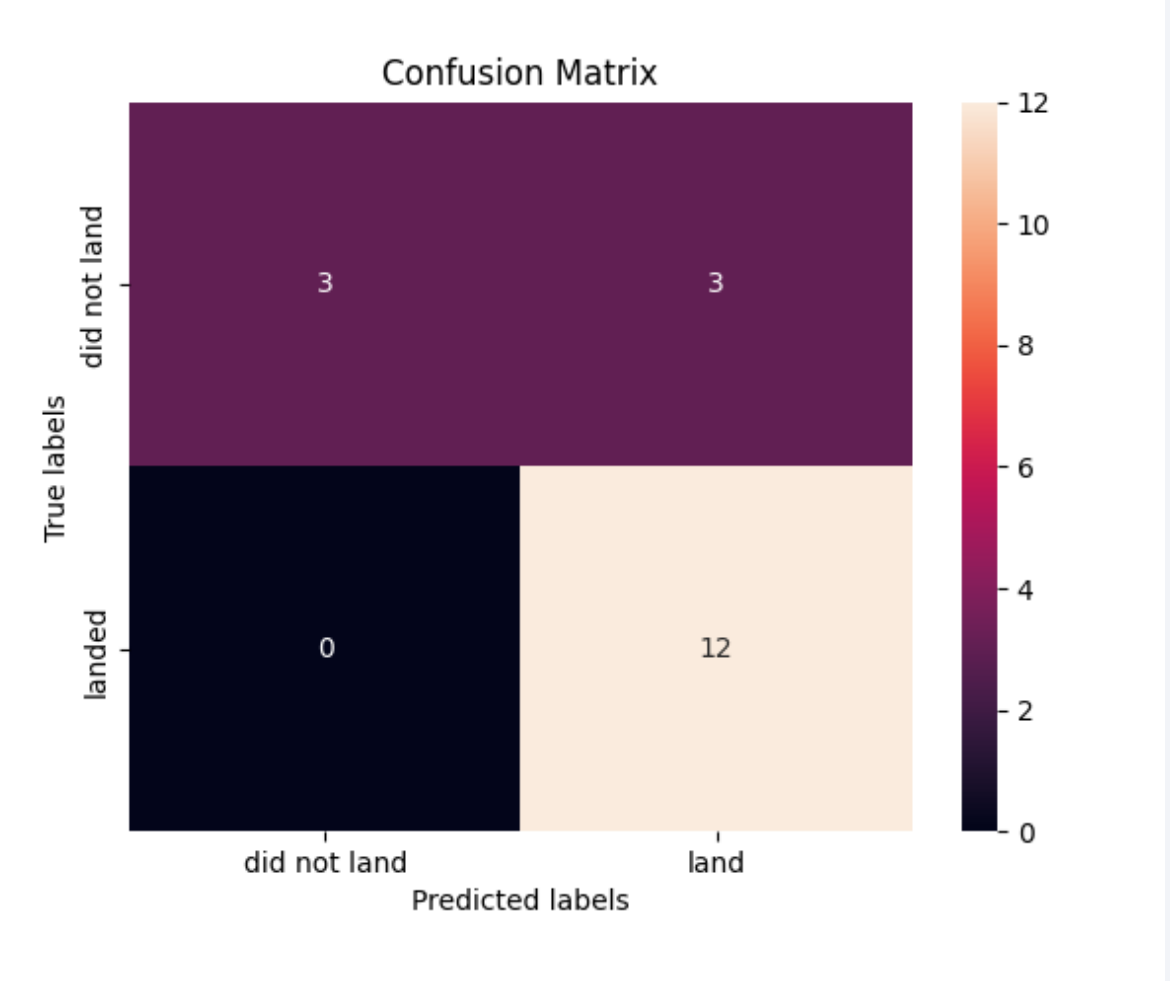
Classification Accuracy

Decision Tree Classification model has the highest model accuracy



Confusion Matrix

- This is a confusion matrix of Decision Tree classification model
- As you can see the model demonstrate 15 correct results among 18 in the test dataset
- It demonstrates mistakes with classification of unsuccessful launches predicting them as successful
- Actually the confusion matrix is the same for all models



Conclusions

- As the result of our research the have Decision Tree classification model that demonstrate the best results in prediction of launch result
- We defined the most efficient launch site for Falcon 9 launches
- It was expected that share of successful launches will grow from year to year
- Also we define best success rates orbits: ES-L1, GEO, HEO, SSO;

Appendix

You can acquainted with all materials here:
https://github.com/VALENTYN1974/rocket-launch

Thank you!