IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Federico Villa
04/06/2023

Outline

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

Executive Summary

- In the present work, were used different methodologies:
  - Data Collection using SpaceX API and web scraping
  - An Exploratory Data Analysis, with data visualization and interactive data visualization
  - Machine learning prediction to a bettere analysis
- Summary of all results:

  All the data were easy to obtain, due to their public status, and thanks to this feature we’ve been able to produce a research showing the best model to predict the better performance of rockets.

Introduction

- Is it possible for Space Y to compete with Space X?
- We need to predict the number of successful landings to better decide the cost of launches, and maximize the results of launches selecting the better location for make them.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data obtained from Space X API and from web scraping.
- Perform data wrangling
  - Data has been evaluated to create an interface for better understanding launch site positions and their features.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was collected and been used for testing various ML models.

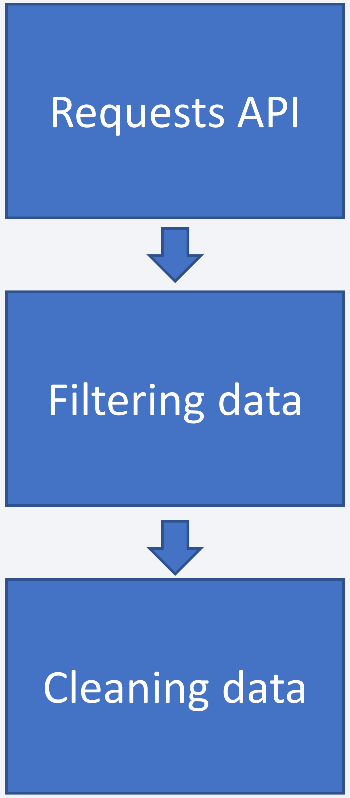
Data Collection

- Data were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/) and from a Wikipedia page, https://en.Wikipedia.or/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches.

Data Collection – SpaceX API

- SpaceX API is a public API that can be used to retrieve data from SpaceX databases.

- https://github.com/fevi98/Jup/blob/master/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- The second part of the data were obtained by Wikipedia via web scraping.

- https://github.com/fevi98/Jup/blob/master/jupyter-labs-webscraping.ipynb

Data Wrangling

- The next step has been the Exploratory Data Analysis on the dataset.

- In this phase, the database has been analyzed a first time.

- https://github.com/fevi98/Jup/blob/master/IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb

EDA with Data Visualization

- Data has been visualized using scatterplots and barplots.

- https://github.com/fevi98/Jup/blob/master/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- We've analyzed the database using SQL queries to properly select most relevant data.
- https://github.com/fevi98/Jup/blob/master/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- I've added objects to better understand the data on map.
- https://github.com/fevi98/Jup/blob/master/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- I created a dashboard to better read and navigate through data
- https://github.com/fevi98/Jup/blob/master/dash.py

Predictive Analysis (Classification)

- I’ve worked on various prediction model to find the better one for this case.

- https://github.com/fevi98/Jup/blob/master/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

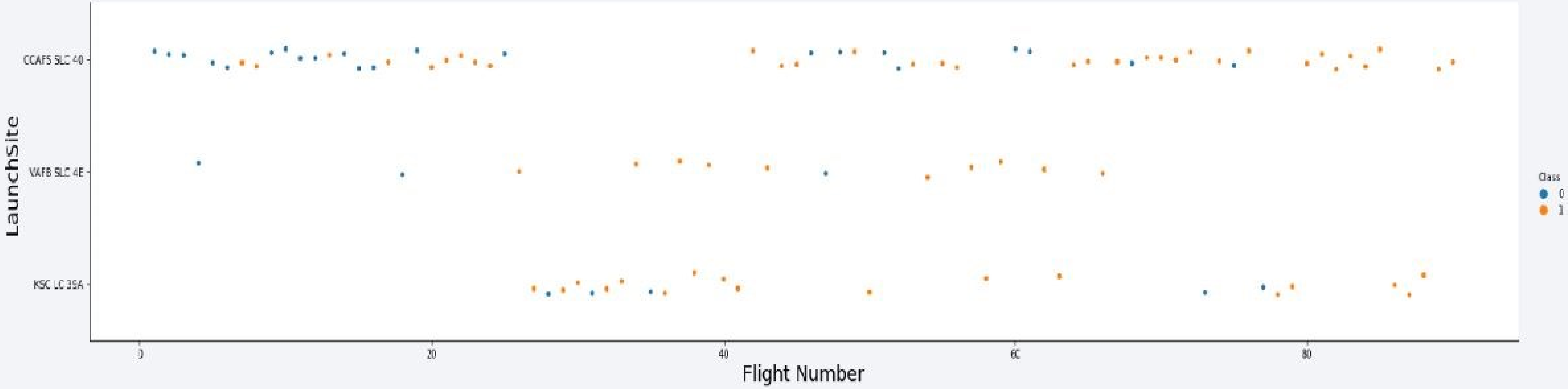
Results

- We’ve found that most of the launches were a success, and that most of them were tried from the US east coast. The results have become better every year.
- The predictive models analysis found out that a Decision Tree Classifier is the best model to predict the outcome of launches.

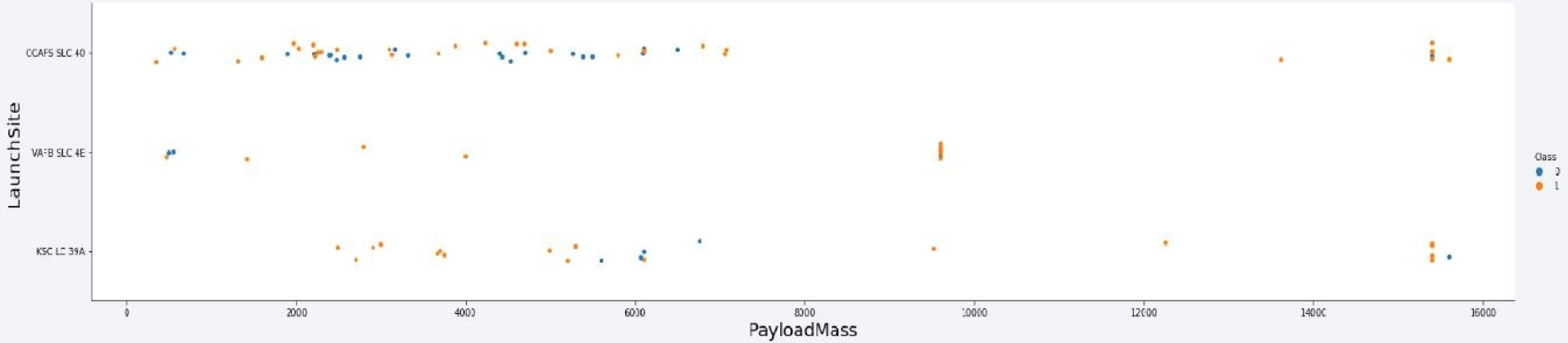
Section 2

Insights drawn from EDA

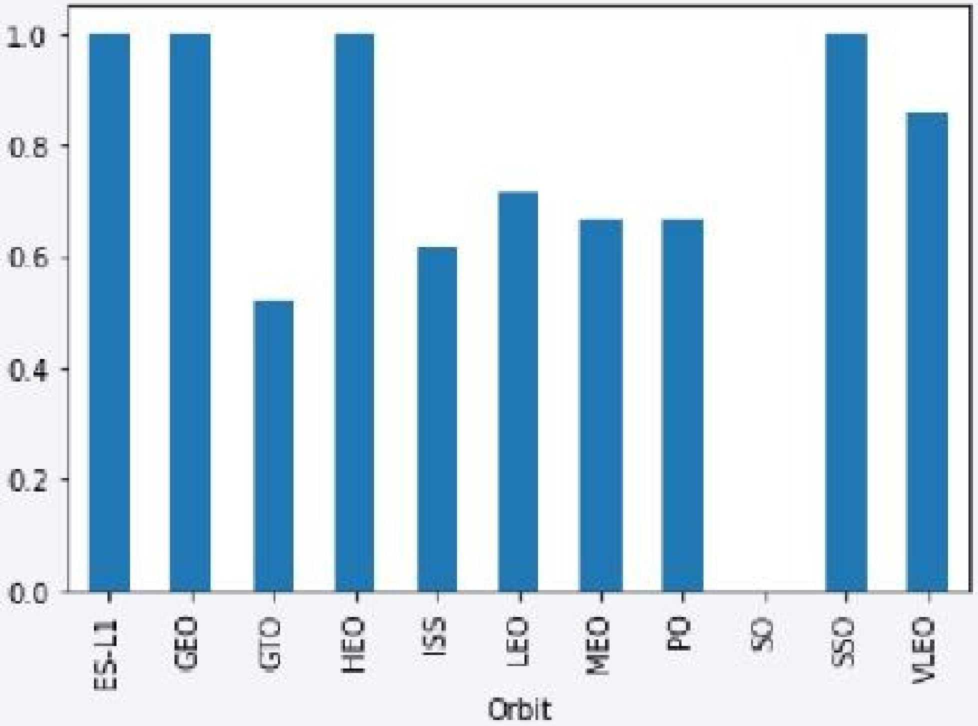
Flight Number vs. Launch Site



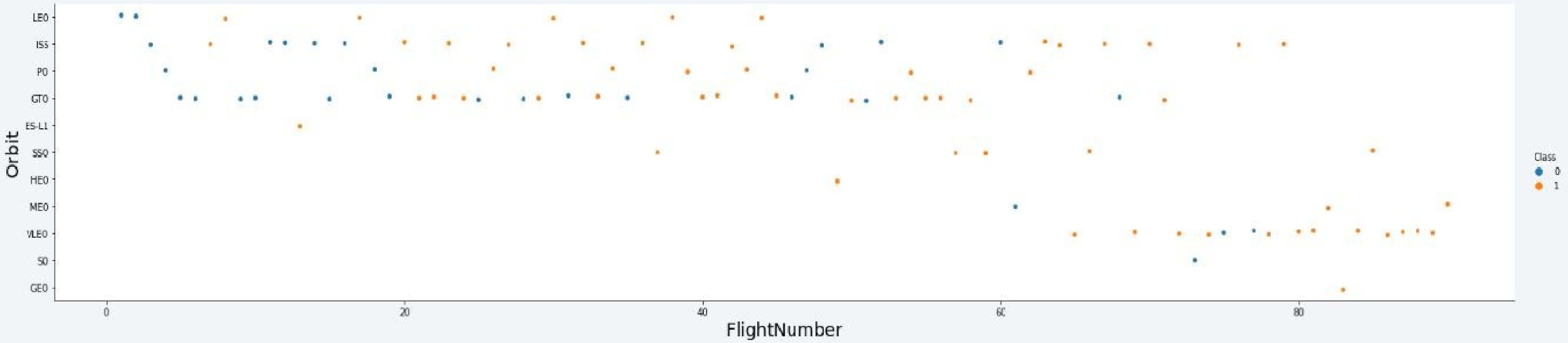
Payload vs. Launch Site



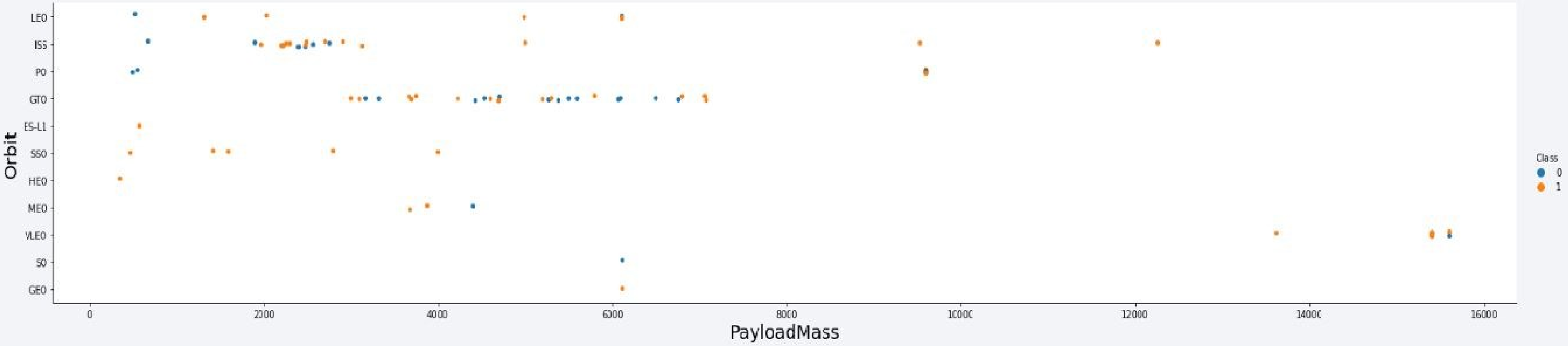
Success Rate vs. Orbit Type



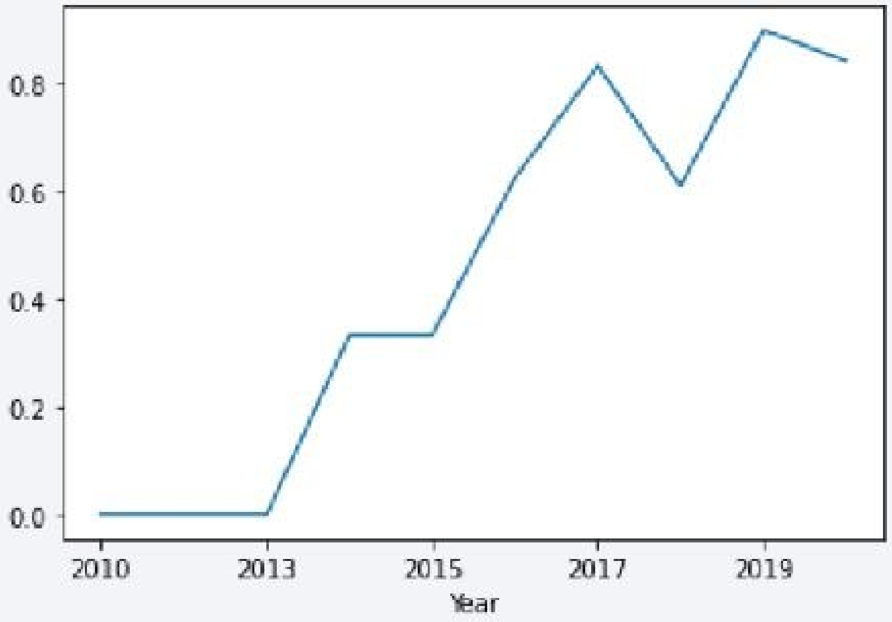
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

These are the launches site names found in the database:

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

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	**CCA**FS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	**CCA**FS 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	**CCA**FS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	**CCA**FS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	**CCA**FS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

This is the total payload of NASA launches:

Total Payload (kg)
111.268

Average Payload Mass by F9 v1.1

This is the average payload mass of F9 v1.1 launches:

Avg Payload (kg)
2.928

First Successful Ground Landing Date

Min Date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

Boosters Carried Maximum Payload

Booster Version
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

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

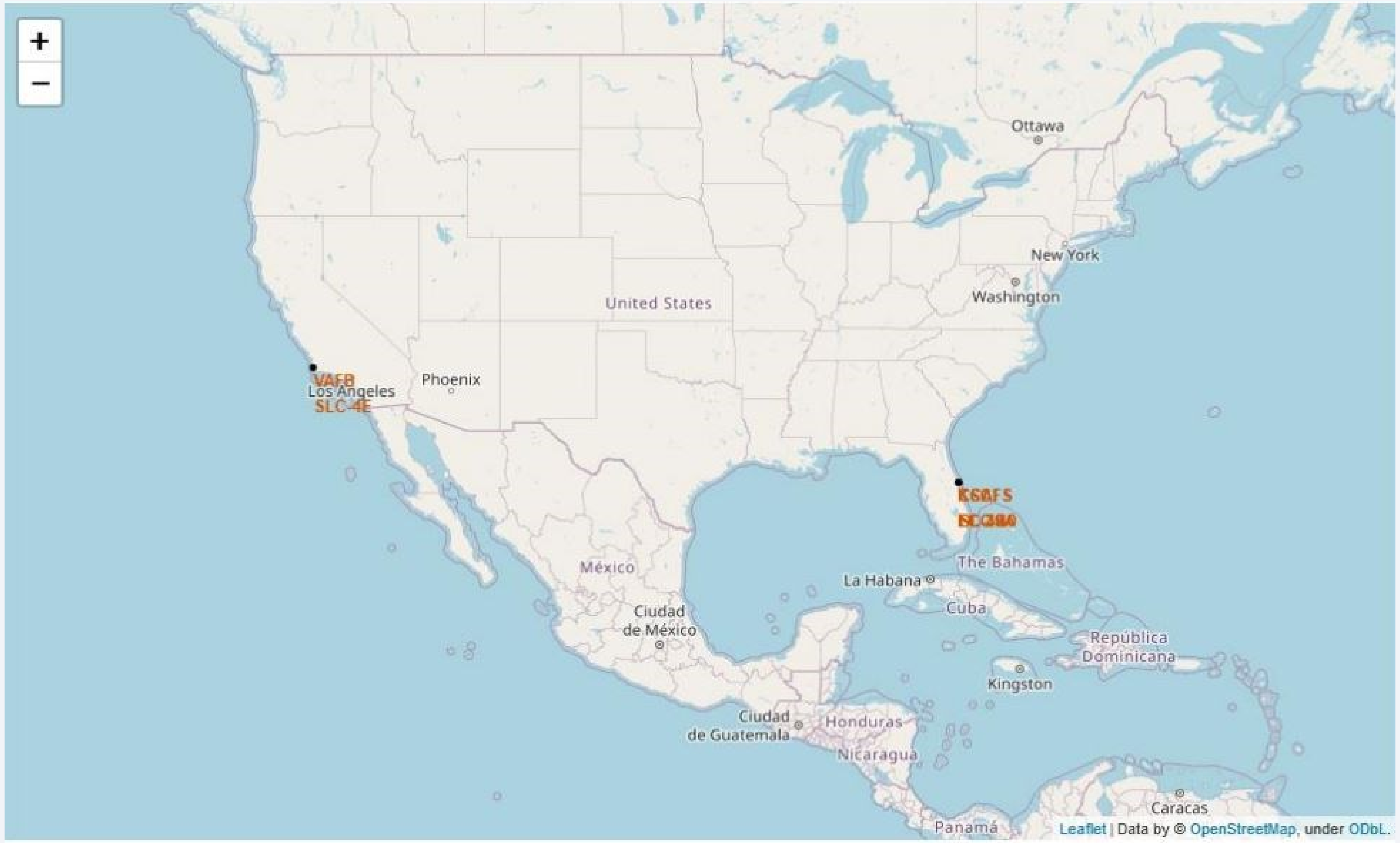
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

Landing Outcome	Occurrences
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

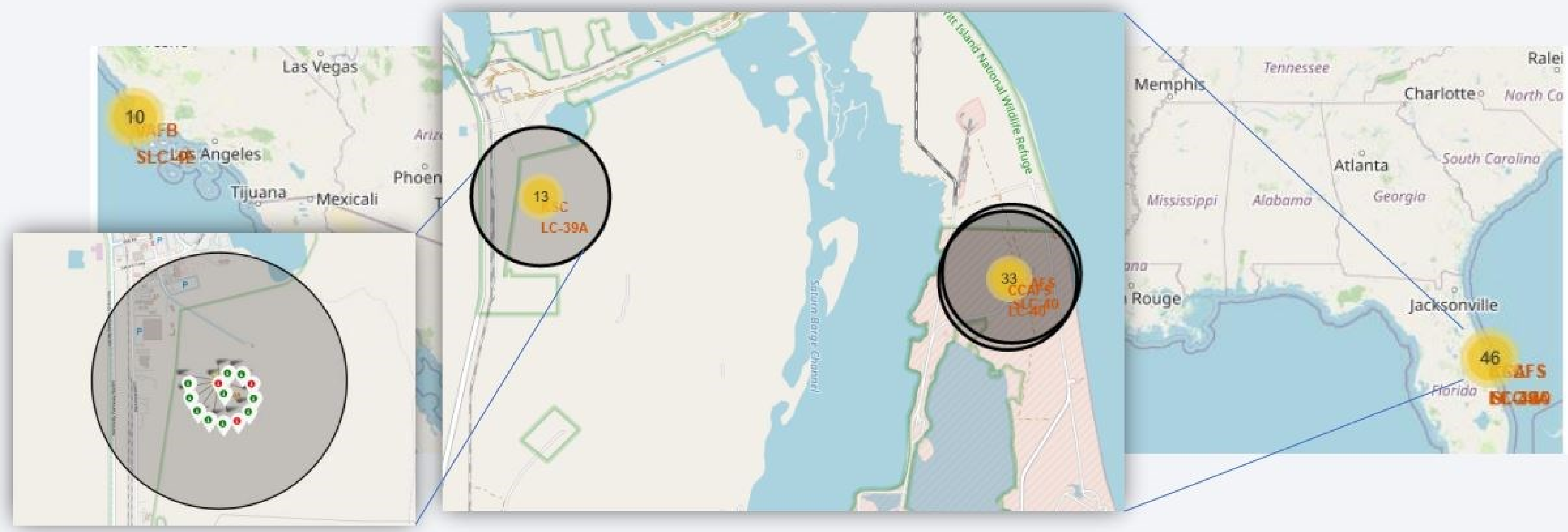
Launch Sites Proximities Analysis

Launch Sites



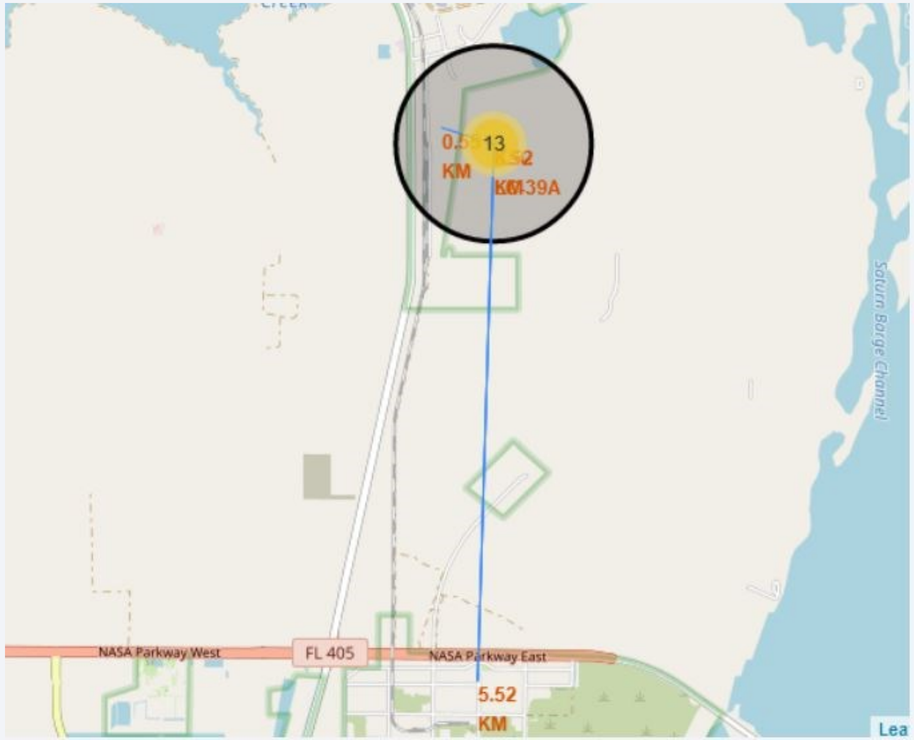
Launch Sites Results

Here we can see an example of Launch Site Result:



Launch Sites Surrounding Visualization

- Surroundings of launch site, to observe the distance from infrastructures:

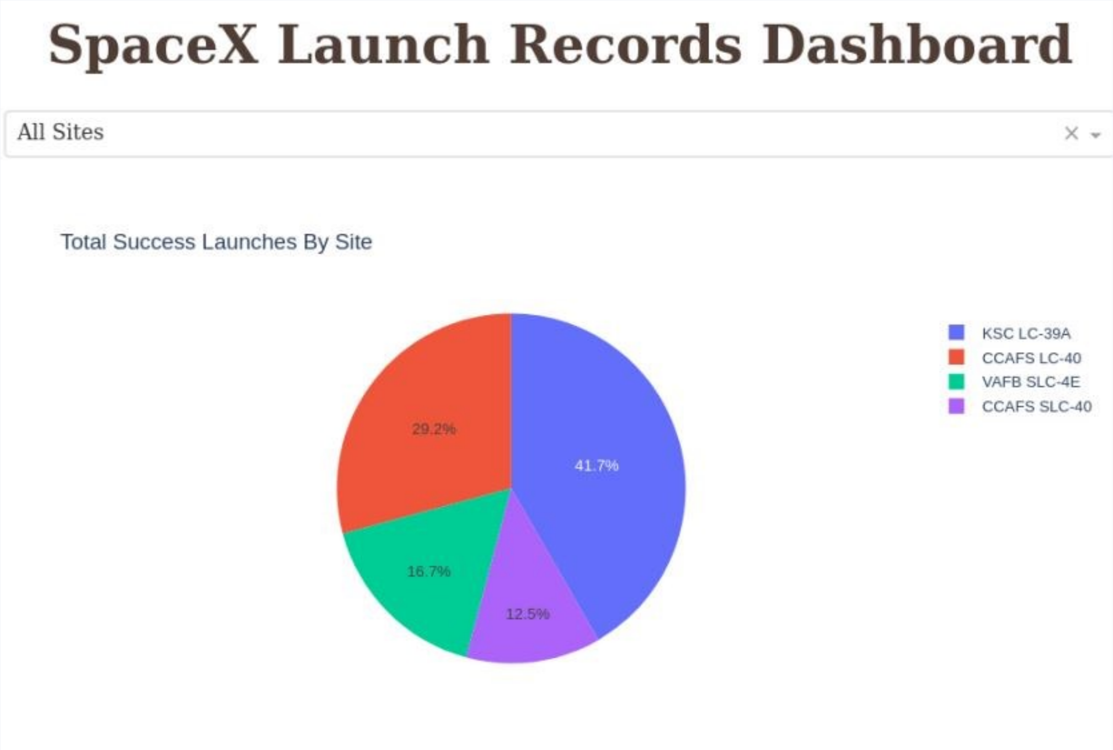


Section 4

Build a Dashboard with Plotly Dash

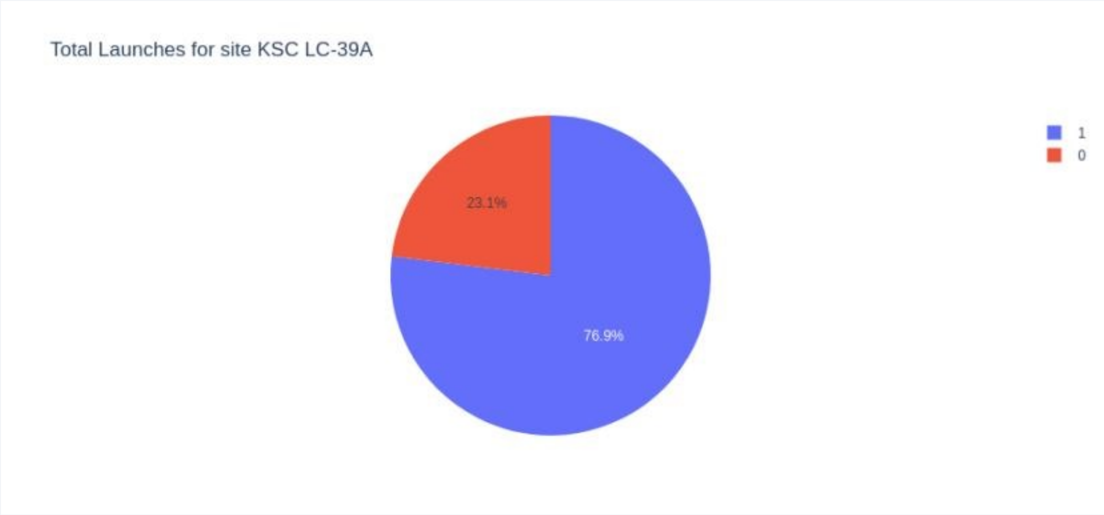
Dashboard View

- Here is the view of the dashboard:



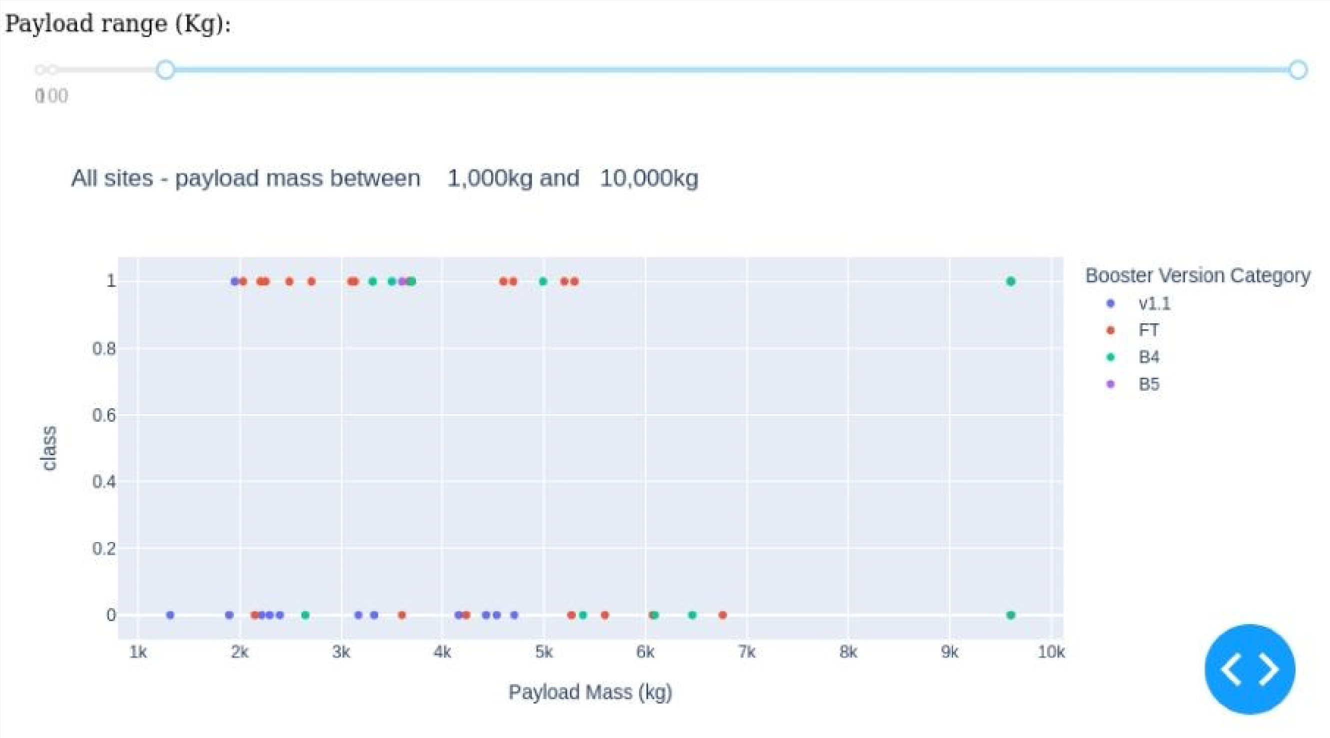
Best Launch Site

- Here we can view a piechart of the success rate of the best launch site, based on success percentage:



Payload range slider

- This is the view of the dashboard showing the success rate by the payload, with the possibility of using a slider:

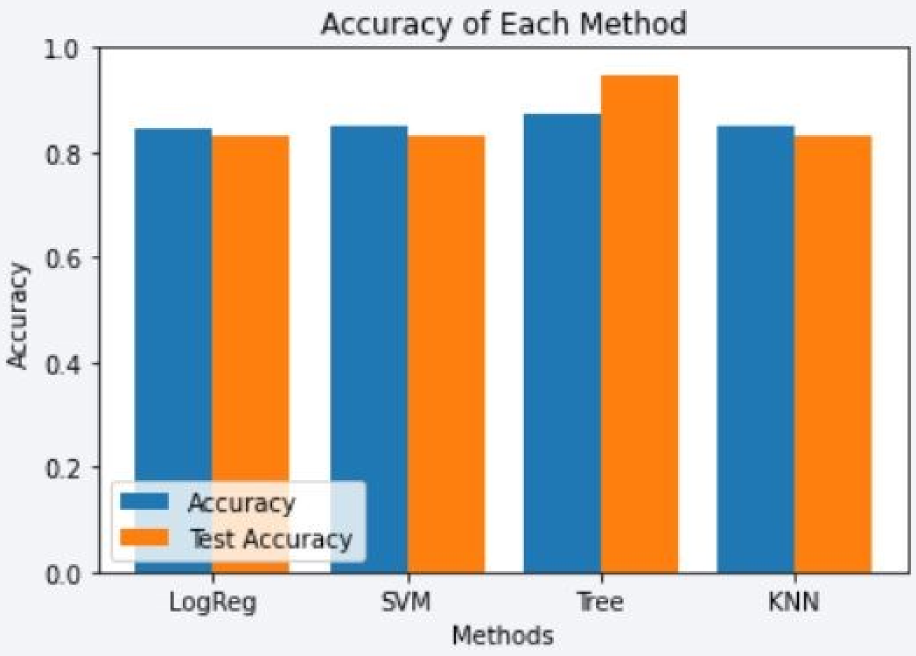


Section 5

Predictive Analysis (Classification)

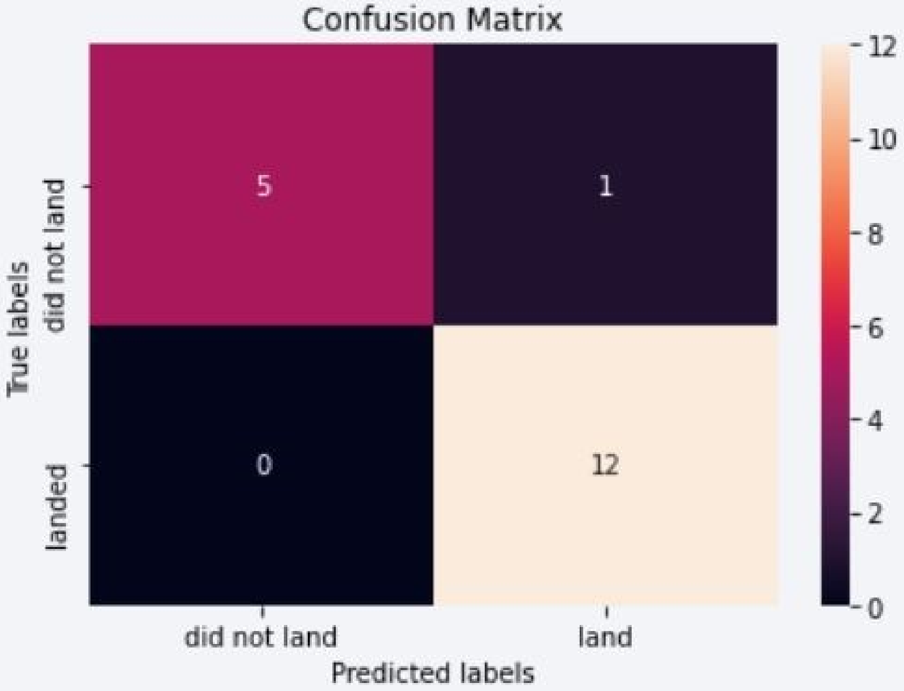
Classification Accuracy

We can now view the score of each method of prediction. The best one, as you can see, is the Decision Tree Classifier.



Confusion Matrix

This is the confusion matric of the Decisional Tree Classification, that shows how high is the model fitting the data evaluated.



Conclusions

- We’ve found out that the best launch site is KSC LC-39A.
- Launches over 7,000kg of payload are more safe.
- The safety of launches increase during time.
- Decision Tree Classifier can help us evaluating the chance of success for our launches.

Appendix

- Github doesn’t shows maps in a JupyterNotebook file.

Thank you!