IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected from @ sources
    - Spacex API
    - Web Scraping from wikipedia
- Perform data wrangling
  - Data was enriched by appending a landing outcome label, following the summarization and analysis of features.
- Perform exploratory data analysis (EDA) using visualization and SQL

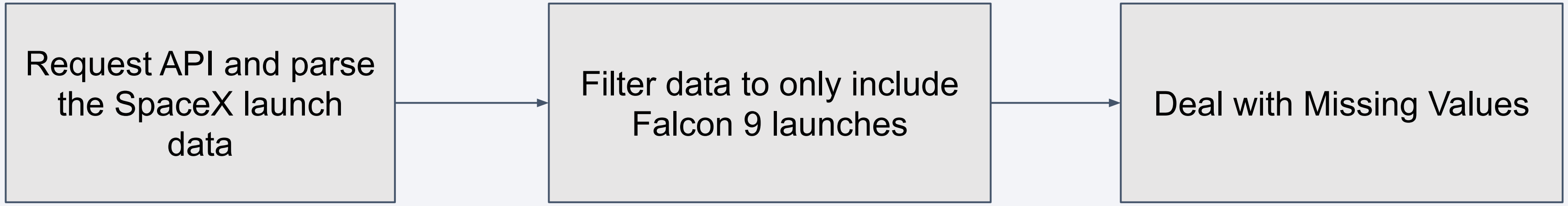
Methodology

Executive Summary

- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Collected data, up to this point, underwent normalization, split into training and test sets, and assessed using four classification models. Model accuracy was evaluated with varied parameter combinations.

Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.



Data Collection – SpaceX API

Data from SpaceX launches can also be obtained from Wikipedia; •

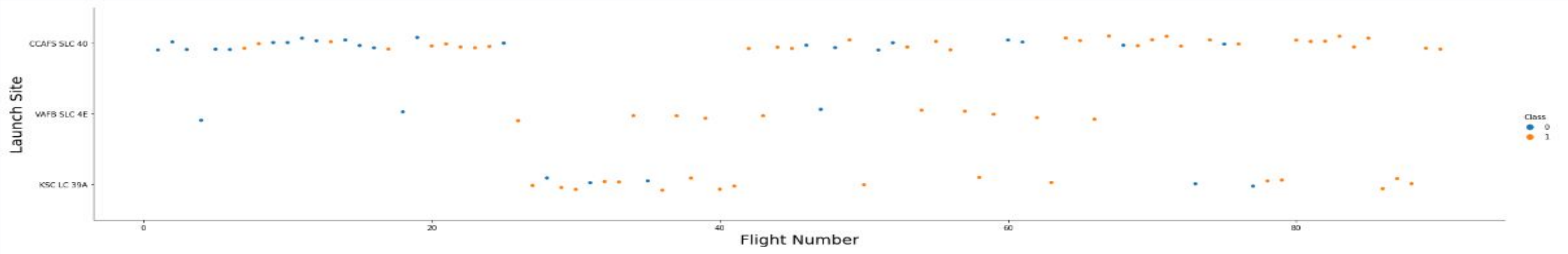
Data are downloaded from Wikipedia according to the flowchart and then persisted.

Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:
- Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



https://github.com/JADE-36/IBM-applied-data-science-capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- Executed SQL queries include:
- Identifying unique launch sites in the space mission;
- Listing the top 5 launch sites starting with 'CCA';
- Calculating total payload mass from NASA-launched boosters (CRS);
- Determining average payload mass from booster version F9 v1.1;
- Locating the date of the first successful landing on a ground pad;
- Retrieving boosters succeeding on a drone ship with payload mass between 4000 and 6000 kg;
- Summing up the total number of successful and failed mission outcomes;
- Listing booster versions carrying the maximum payload mass;
- Identifying failed landing outcomes on a drone ship, including booster versions and launch site names in 2015;
- Ranking the count of landing outcomes between June 4, 2010, and March 20, 2017.

https://github.com/JADE-36/IBM-applied-data-science-capstone/blob/main/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps • Markers indicate points like launch sites;
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and • Lines are used to indicate distances between two coordinates.

https://github.com/JADE-36/IBM-applied-data-science-capstone/blob/main/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data

  Percentage of launches by site

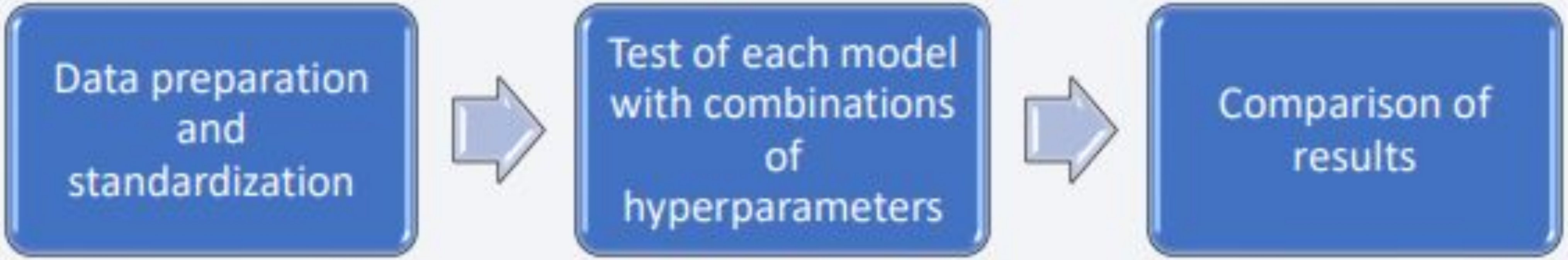
  Payload range

- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

https://github.com/JADE-36/IBM-applied-data-science-capstone
/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.



Results

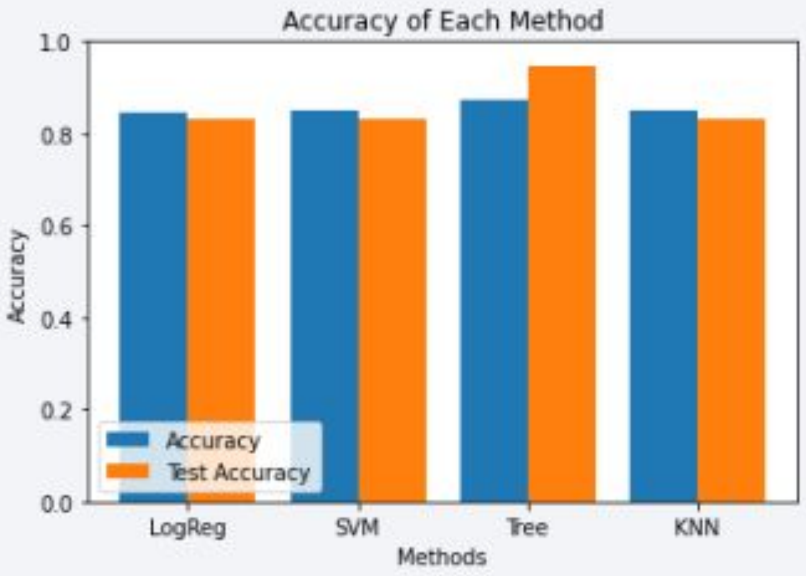
Exploratory data analysis findings:

- There are 4 distinct launch sites utilized by Space X;
- Initial launches targeted Space X itself and NASA;
- F9 v1.1 booster's average payload stands at 2,928 kg;
- The first successful landing occurred in 2015, five years post the inaugural launch;
- Numerous Falcon 9 booster versions achieved successful drone ship landings with payloads surpassing the average;
- Nearly 100% of mission outcomes resulted in success;
- Two booster versions, F9 v1.1 B1012 and F9 v1.1 B1015, faced landing failures on drone ships in 2015;
- The proficiency of landing outcomes improved over the years.

Results

- Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.

- Most launches happens at east cost launch sites.



Section 2

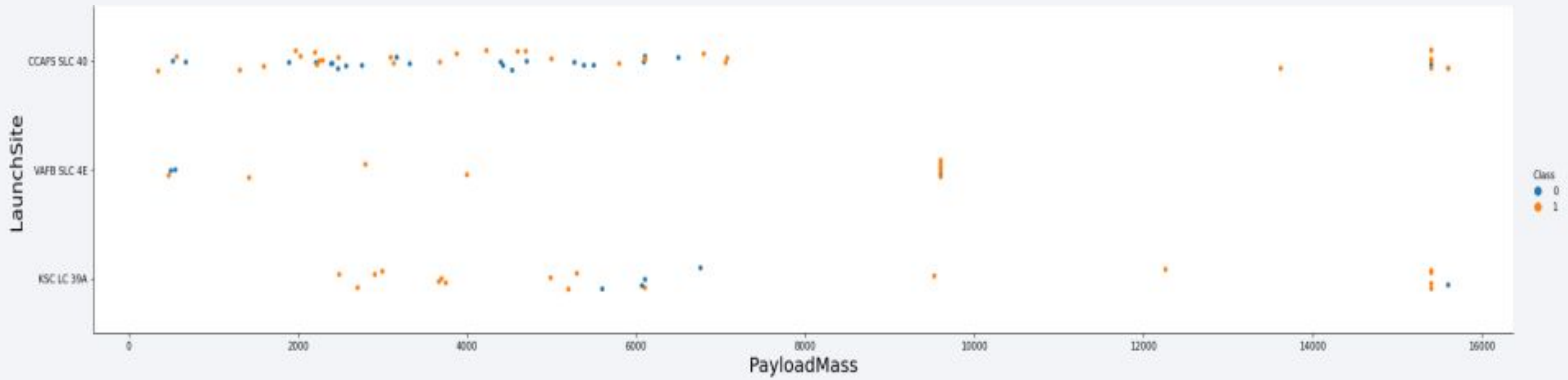
Insights drawn from EDA

Flight Number vs. Launch Site

According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;

- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.

Payload vs. Launch Site



• Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;

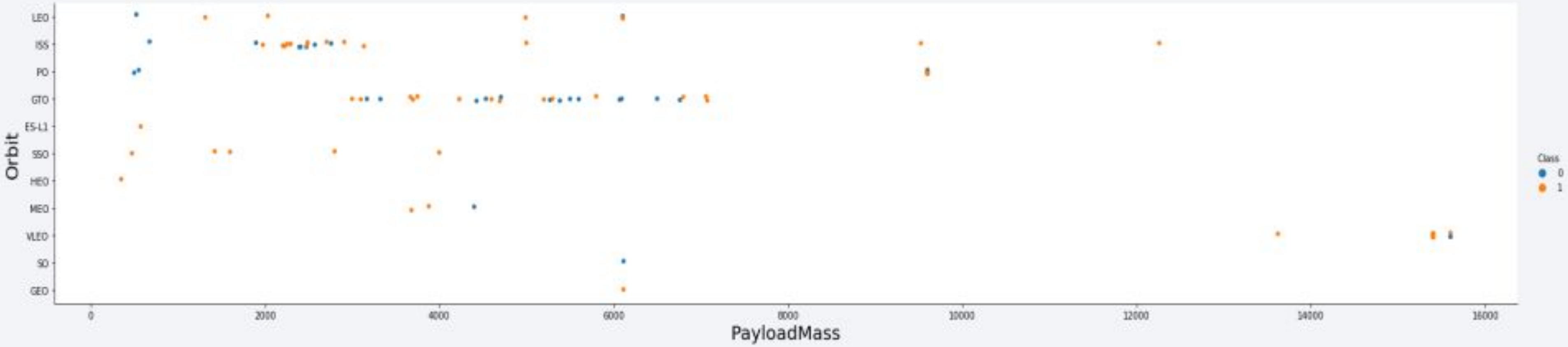
• Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- Orbits with the highest success rates include:
- ES-L1;
- GEO;
- HEO; and
- SSO.
- Subsequently:
- VLEO (above 80%); and
- LFO (above 70%).

Flight Number vs. Orbit Type

Payload vs. Orbit Type



Launch Success Yearly Trend

The success rate exhibited a continuous increase from 2013 to 2020, suggesting that the initial three years marked a period of adjustments and technological improvements.

All Launch Site Names

The dataset indicates four launch sites with the following names:

CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E
These names were derived by selecting unique occurrences of "launch_site" values from the dataset.

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

The total payload carried by boosters from NASA is 111,268 kg. This calculation involved summing all payloads with codes containing 'CRS', corresponding to NASA.

Average Payload Mass by F9 v1.1

Average payload mass carried by booster version F9 v1.1: • Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

First Successful Ground Landing Date

The first successful landing outcome on a ground pad occurred on December 22, 2015. This was identified by filtering the data for successful landings on ground pads and retrieving the minimum date value.

Successful Drone Ship Landing with Payload between 4000 and 6000

The boosters that successfully landed on a drone ship with a payload mass greater than 4000 but less than 6000 are as follows:

F9 FT B1021.2

F9 FT B1031.2

F9 FT B1022

F9 FT B1026

These distinct booster versions were selected based on the specified criteria.

Total Number of Successful and Failure Mission Outcomes

The number of successful and failure mission outcomes, based on the grouped and counted records, are as follows:

Success: 99 occurrences

Success (payload status unclear): 1 occurrence

Failure (in flight): 1 occurrence

Boosters Carried Maximum Payload

The boosters that have carried the maximum payload mass, as recorded in the dataset, include the following:

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

In 2015, the failed landing outcomes on a drone ship, along with their respective booster versions and launch site names, are as follows:

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

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

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

When considering all landing outcomes between June 4, 2010, and March 20, 2017, the occurrences and rankings are as follows:

No attempt: 10 occurrences

Failure (drone ship): 5 occurrences

Success (drone ship): 5 occurrences

Controlled (ocean): 3 occurrences

Success (ground pad): 3 occurrences

Failure (parachute): 2 occurrences

Uncontrolled (ocean): 2 occurrences

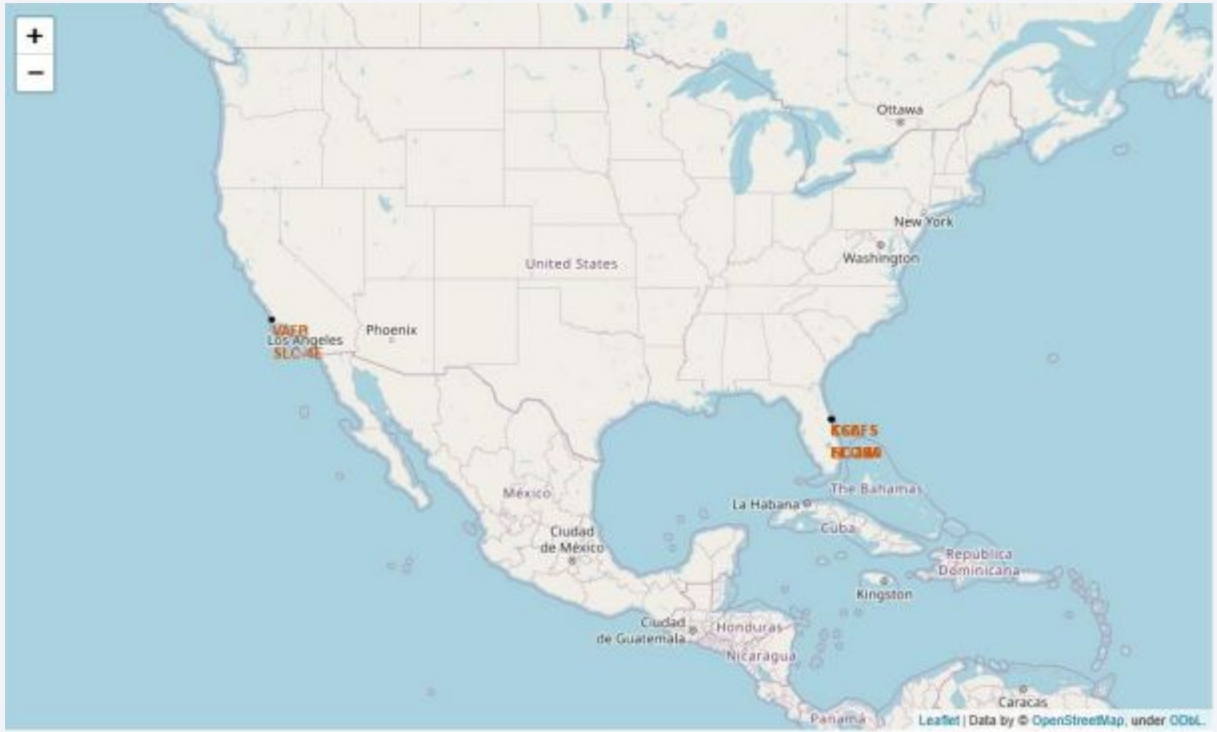
Precluded (drone ship): 1 occurrence

It's important to note that "No attempt" should be taken into account in the analysis.

Section 3

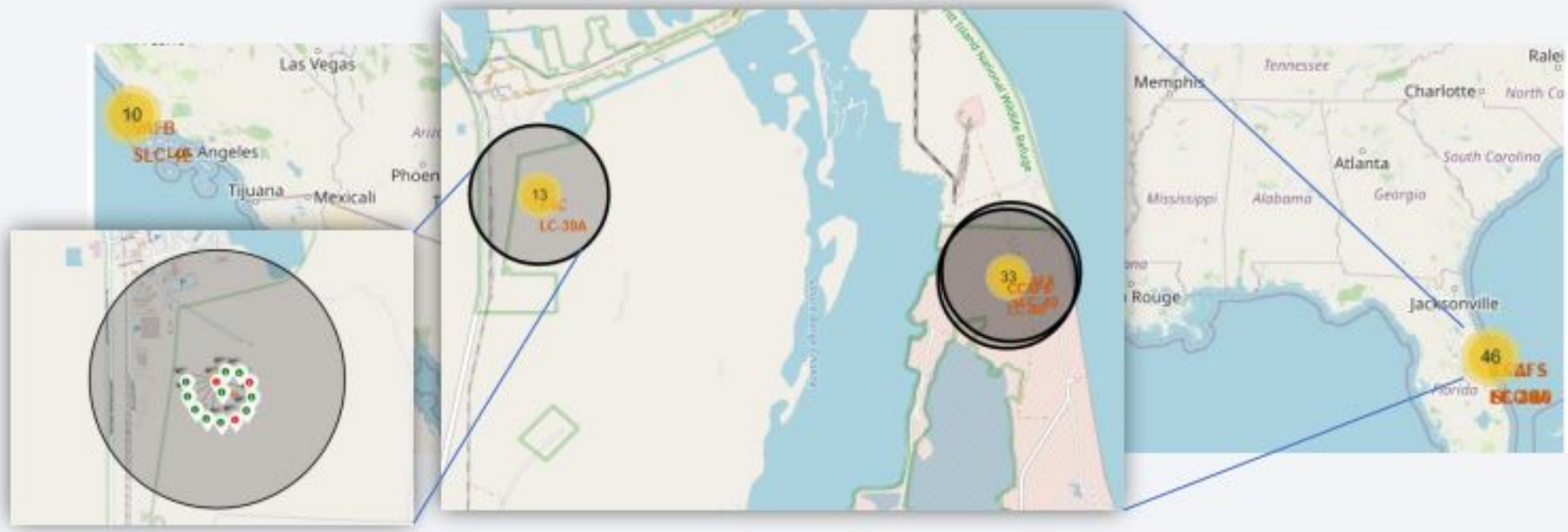
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



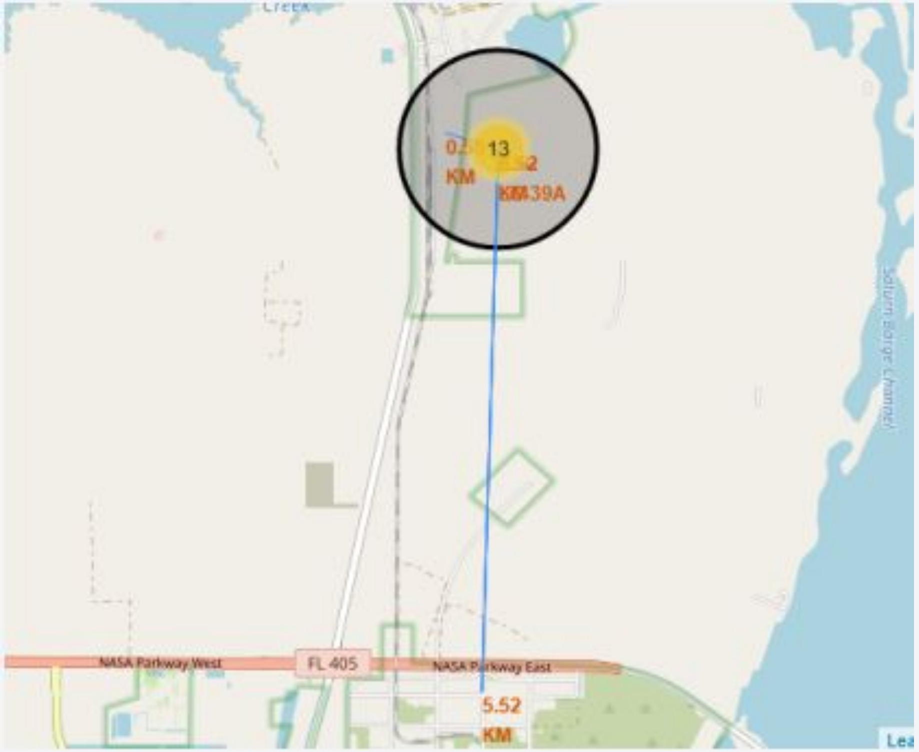
Launch sites are near sea, probably by safety, but not too far from roads and railroads.

Launch Outcomes by Site



Green markers indicate successful and red ones indicate failure.

Logistics and Safety



Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas

Section 4

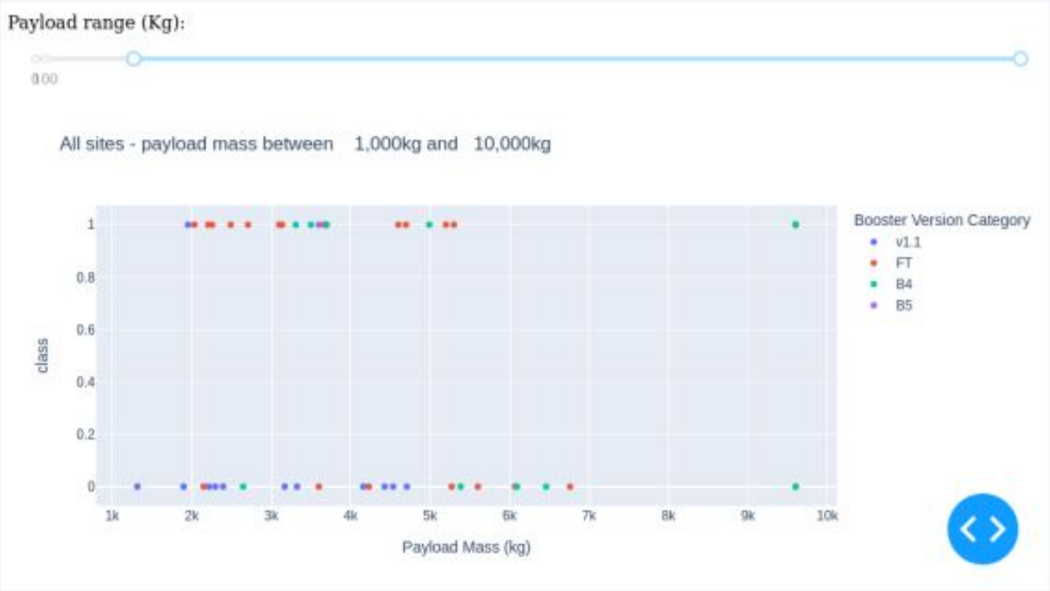
Build a Dashboard with Plotly Dash

Successful Launches by Site

Launch Success Ratio for KSC LC-39A

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

Payload vs. Launch Outcome



Section 5

Predictive Analysis (Classification)

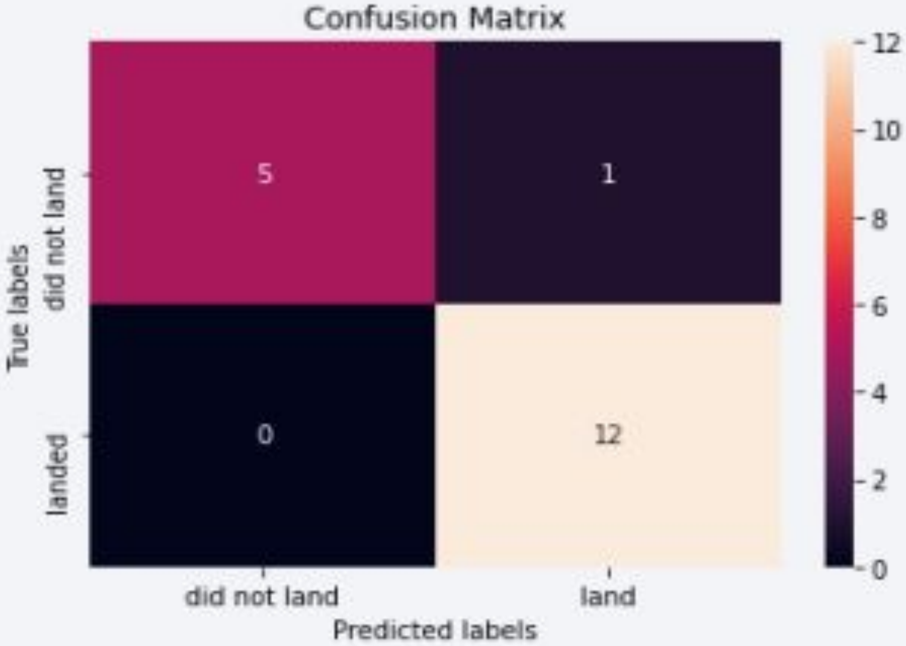
Classification Accuracy

Four classification models were tested, and their accuracies are plotted beside;

The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.

Confusion Matrix

Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- As an improvement for model tests, it's important to set a value to np.random.seed variable;
-
- • Folium didn't show maps on Github, so I took screenshots.

Thank you!