IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Lucas Santos
2022-10-14

Outline

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

Executive Summary

Collected data via SpaceX API:

- Made manipulation to repair the masks being manipulated;
- Rocket filtering;
- Dealing with nulls;

Exploratory Data Analysis:

- Data understanding and feature engineering;

Machine learning prediction:

- Perform ML models to predict response;
- Show the best model

Introduction

Goal:

- The objective is to evaluate the viability of the new company Space Y to compete with Space X.

Questions to answer:

- The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets;

- Where is the best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - Data from Space X was obtained from:
    - Space X API (https://api.spacexdata.com/v4/rockets/);

- **Perform data wrangling:**
  - After understanding the data, a feature engineering was made to prepare data to the ML model;

- **Perform exploratory data analysis (EDA) using visualization and SQL:**
  - Get insights about data

- **Perform interactive visual analytics using Folium and Plotly Dash;**

- **Use classification models to predict response**

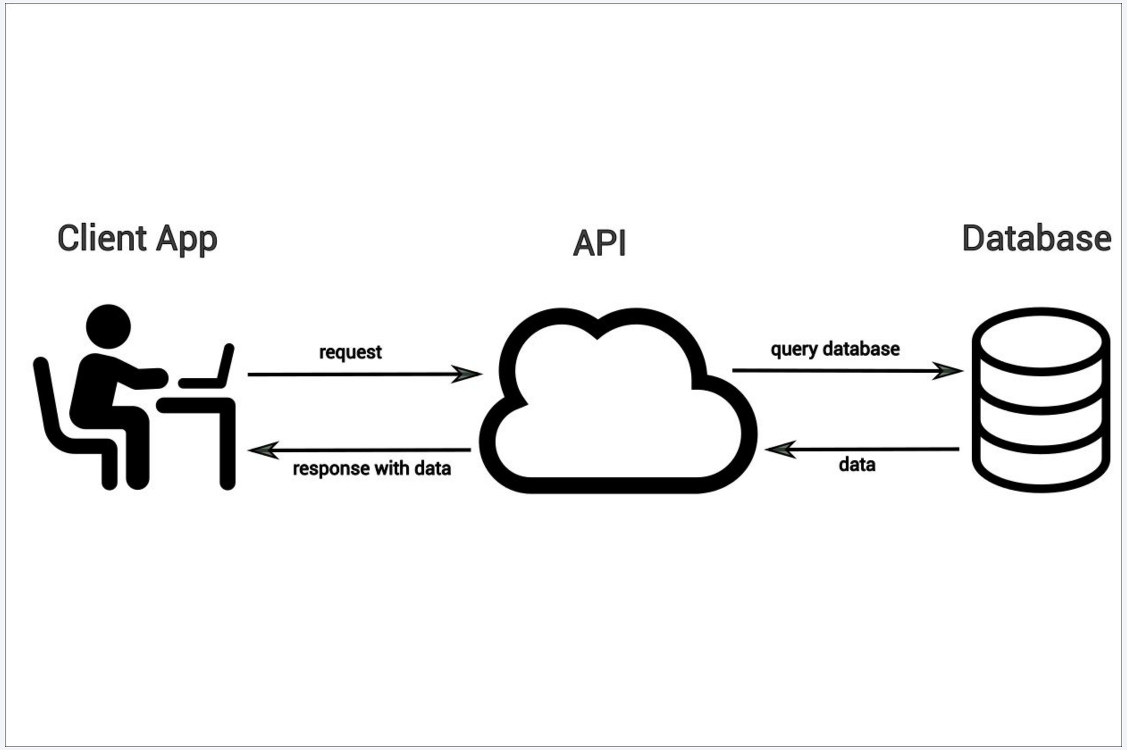
Data Collection

- Data sets were collected from Space X API

Data Collection – SpaceX API

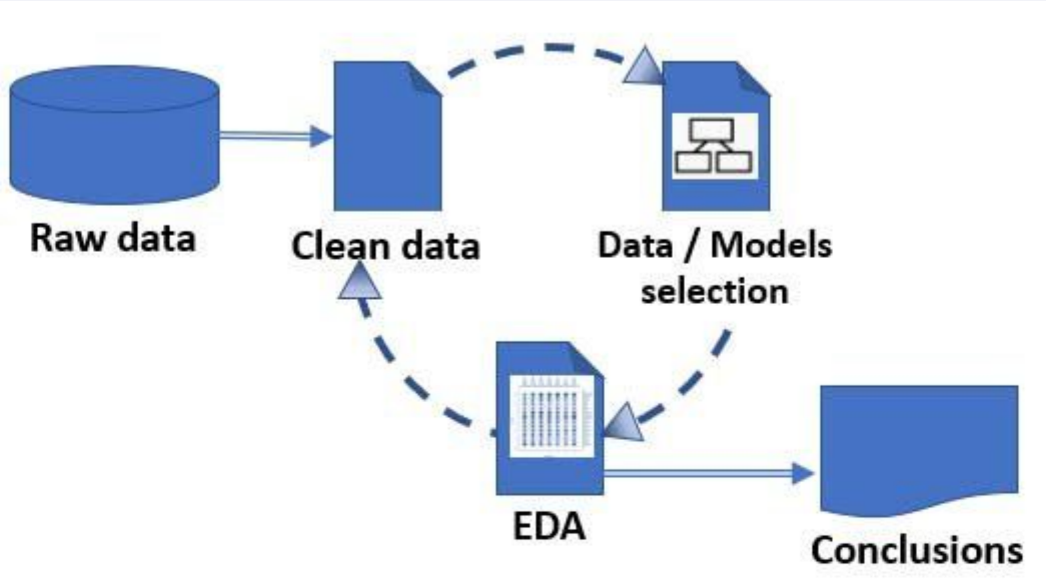
- SpaceX offers a public API where you can download the data;

- File to collect the data:
    - Data Collection API.ipynb



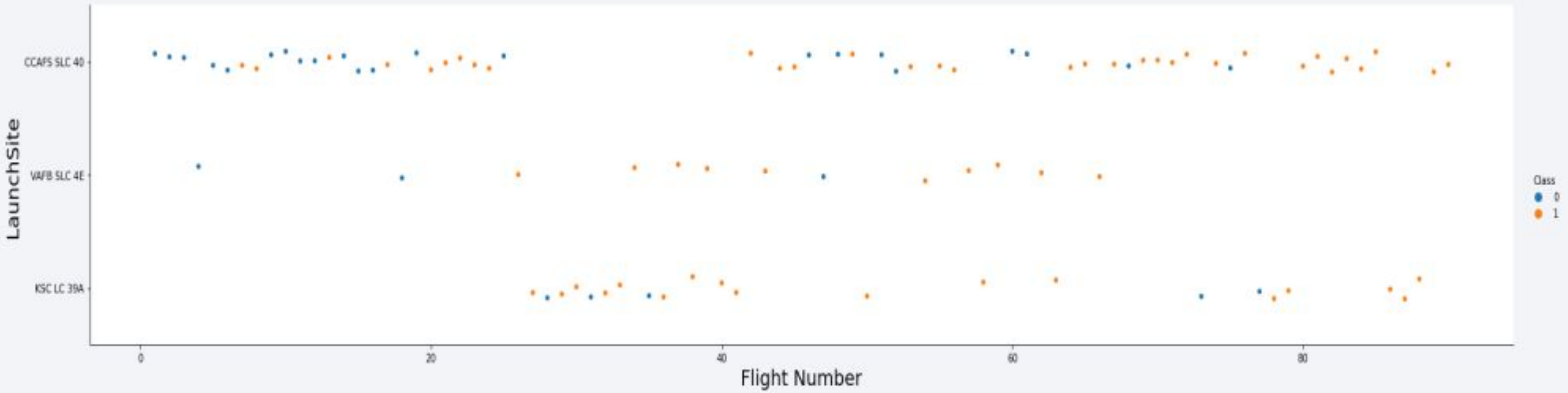
Data Wrangling

- 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



EDA with SQL

- **SQL queries performed:**
  - Names of the unique launch sites in the space mission;
  - Top 5 launch sites whose name begin with the string 'CCA';
  - Total payload mass carried by boosters launched by NASA (CRS);
  - Average payload mass carried by booster version F9 v1.1;
  - Date when the first successful landing outcome in ground pad was achieved;
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;
  - Total number of successful and failure mission outcomes;
  - Names of the booster versions which have carried the maximum payload mass;
  - Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015;
  - Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

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;
  - Lines are used to indicate distances between two coordinates.

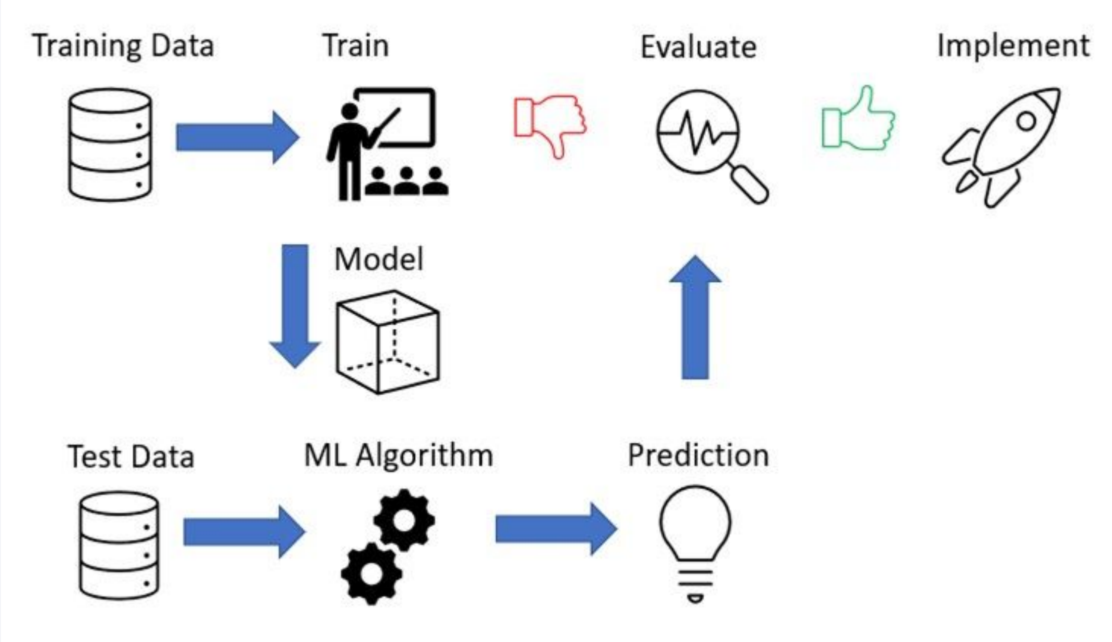
Build a Dashboard with Plotly Dash

- **The following graphs and plots were used to visualize data:**
  - Percentage of launches by site
  - Payload range

- **It helped to get insights to identify the best place to launch**

Predictive Analysis (Classification)

- **Four classification models were compared:**
  - Logistic Regression,
  - Support Vector Machine,
  - Decision Tree
  - K Nearest Neighbors (KNN).

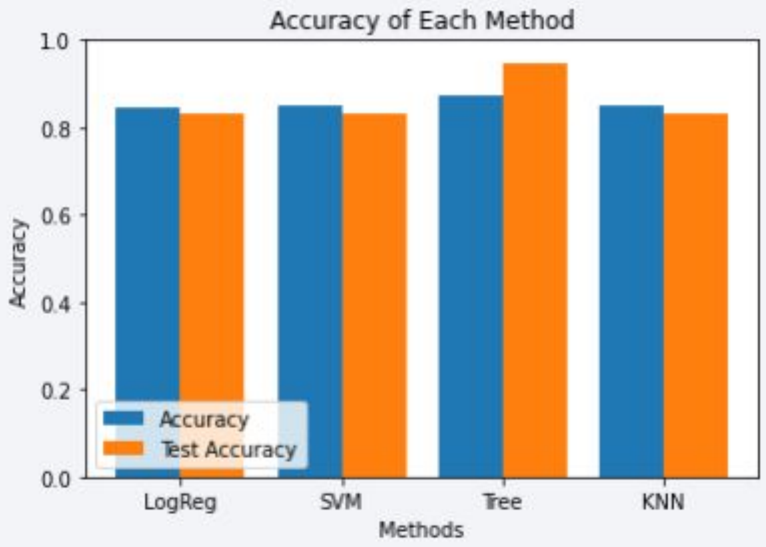


Results

- **Exploratory data analysis results:**
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
  - Almost 100% of mission outcomes were successful;
  - Two booster versions failed at landing in drone ships in 2015:
    - F9 v1.1 B1012 and F9 v1.1 B1015;
  - The number of landing outcomes became as better as years passed.

Results

- **Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.**



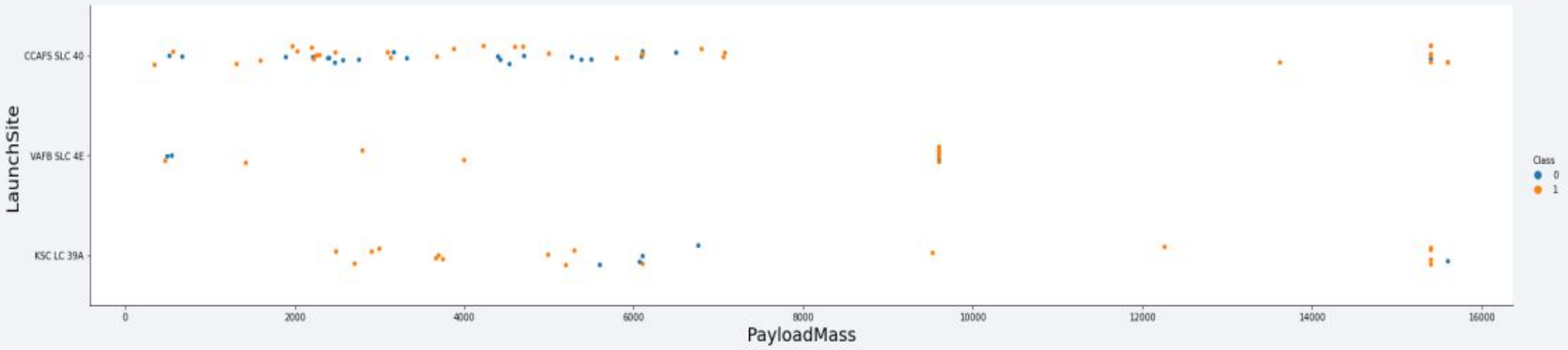
Section 2

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;

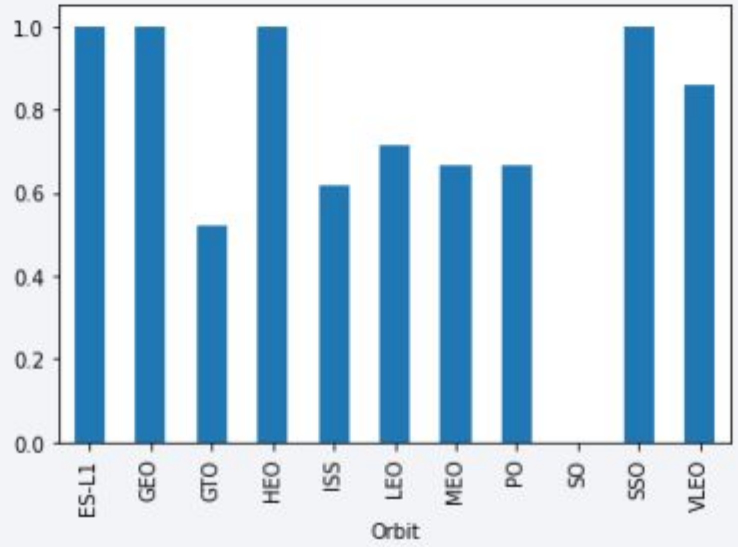
Payload vs. Launch Site



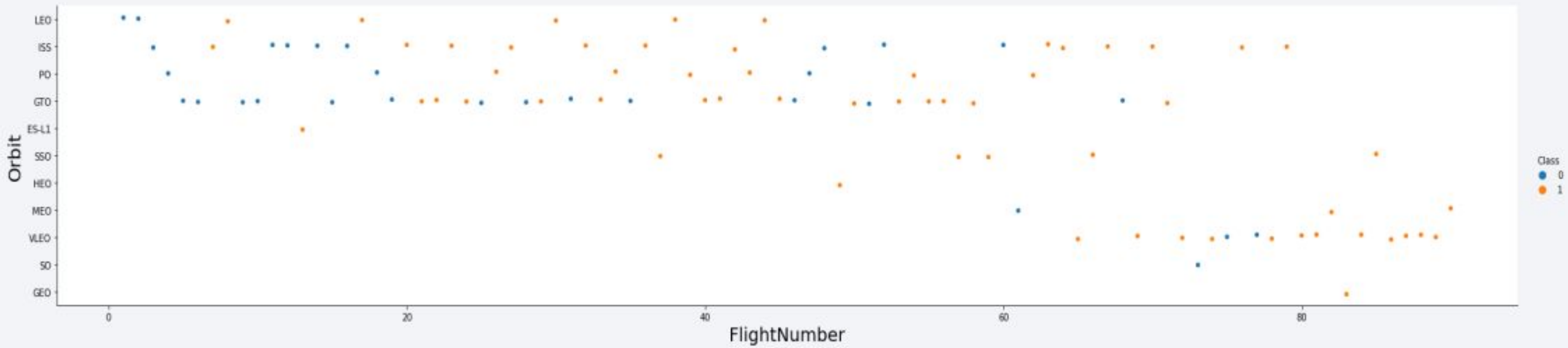
- According to the plot above, Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;

Success Rate vs. Orbit Type

- **How to see, the biggest success rates happens to orbits:**
- ES-L1;
- GEO;
- HEO;
- SSO.

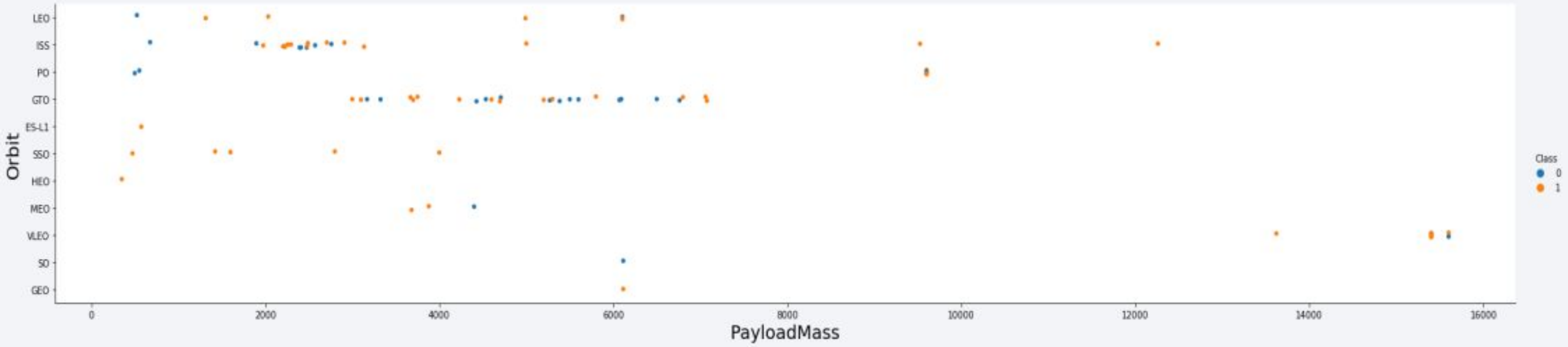


Flight Number vs. Orbit Type



- Apparently, success rate improved over time to all orbits;

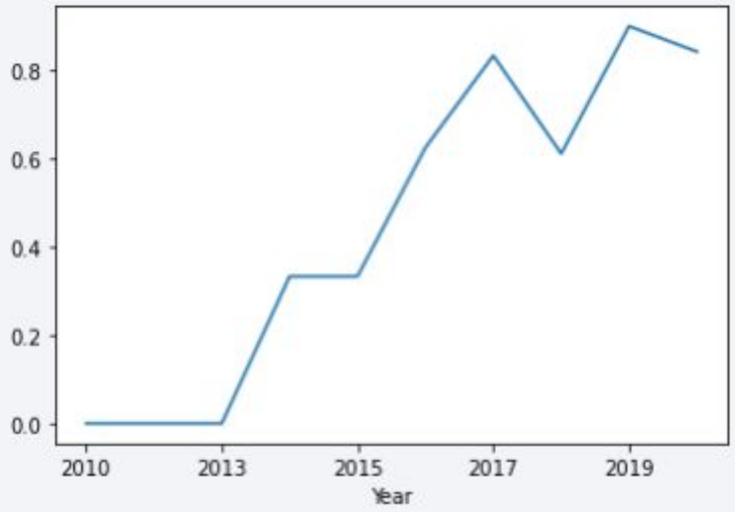
Payload vs. Orbit Type



- Apparently, there is no relation between payload and success rate

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept until 2020:
    - It seems that the first three years were a period of adjusts and improvement of technology.



All Launch Site Names

- There are four launch sites, obtained by “launch site” information in the data:

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites 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 attemp

Total Payload Mass

- Total payload carried by boosters from NASA:

  111268 Kg

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1:

    2928 Kg

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

  2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

  - **F9 FT B1021.2**
  - **F9 FT B1031.2**
  - **F9 FT B1022**
  - **F9 FT B1026**

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes:

  - **Success = 99**
  - **Success (payload status unclear) = 1**
  - **Failure (in flight) = 1**

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass:
  - **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

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015:

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

- Ranking of all landing outcomes between the date 2010-06-04 and 2017- 03-20:

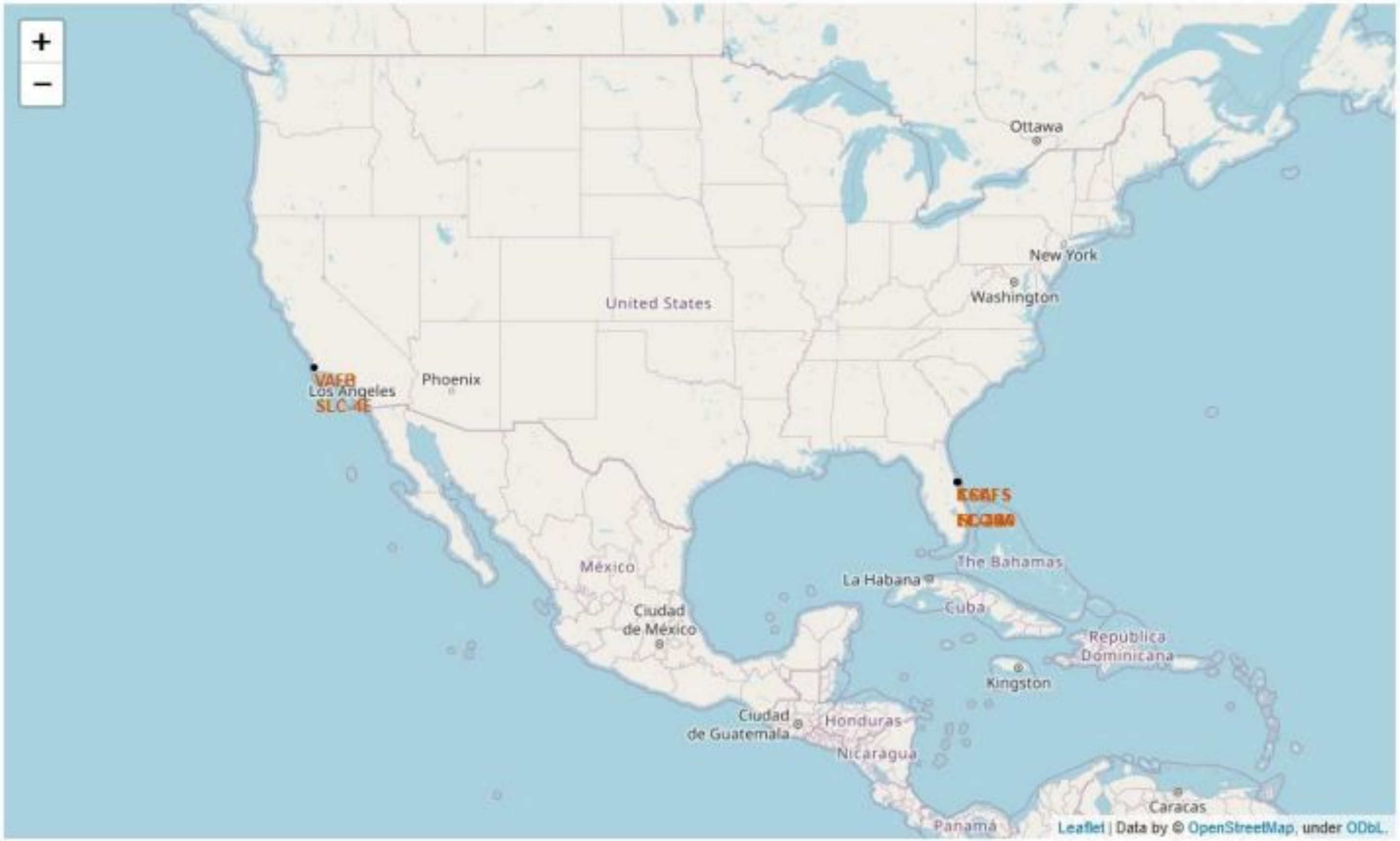
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

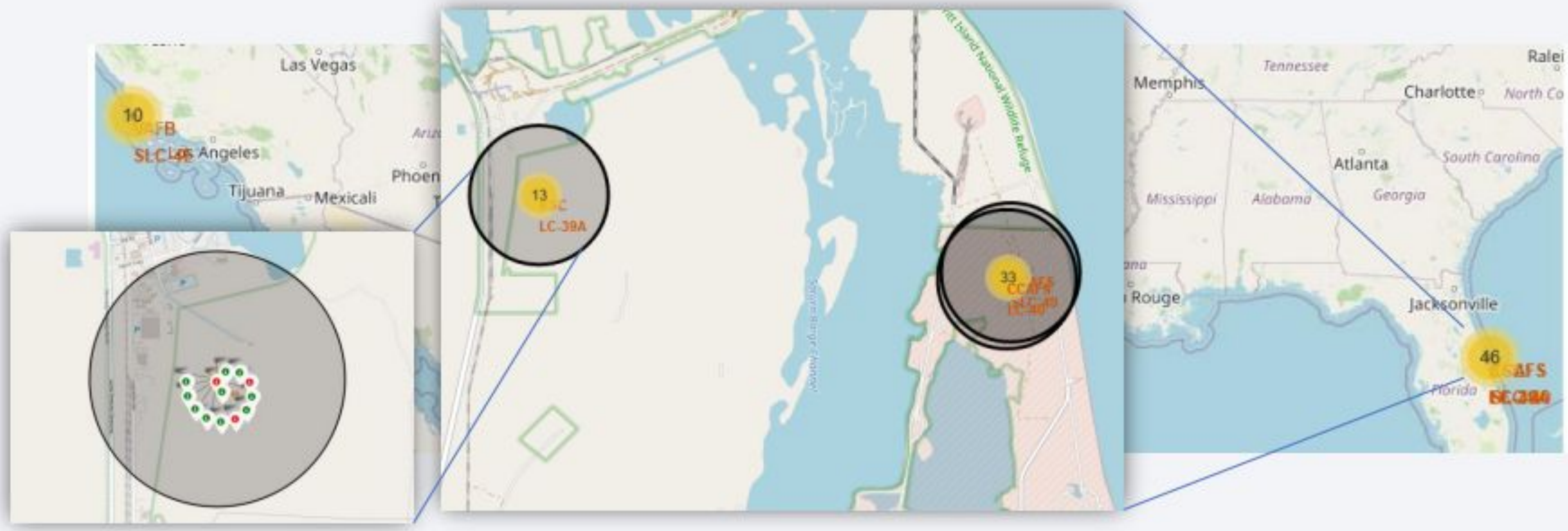
All launch sites

- Closer to the sea, more safer



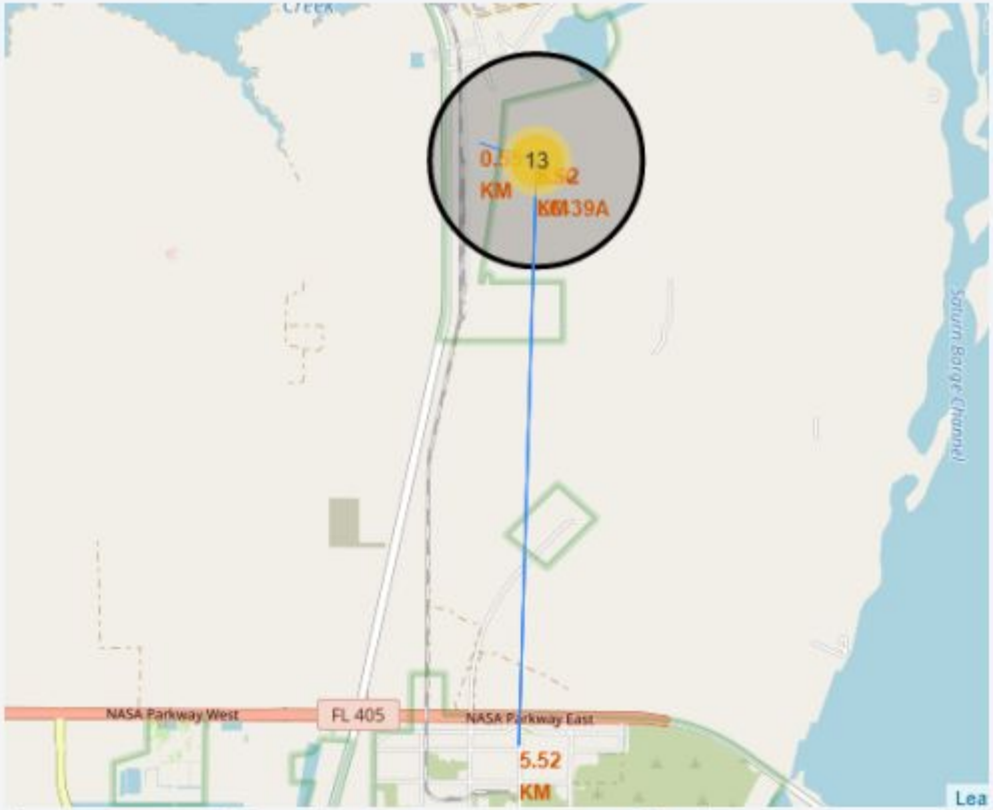
Launch Outcomes by site

- Example of KSC LC-39A launch site launch outcomes:



Logistics

- Close to the sea and empty areas, but also with access road close by

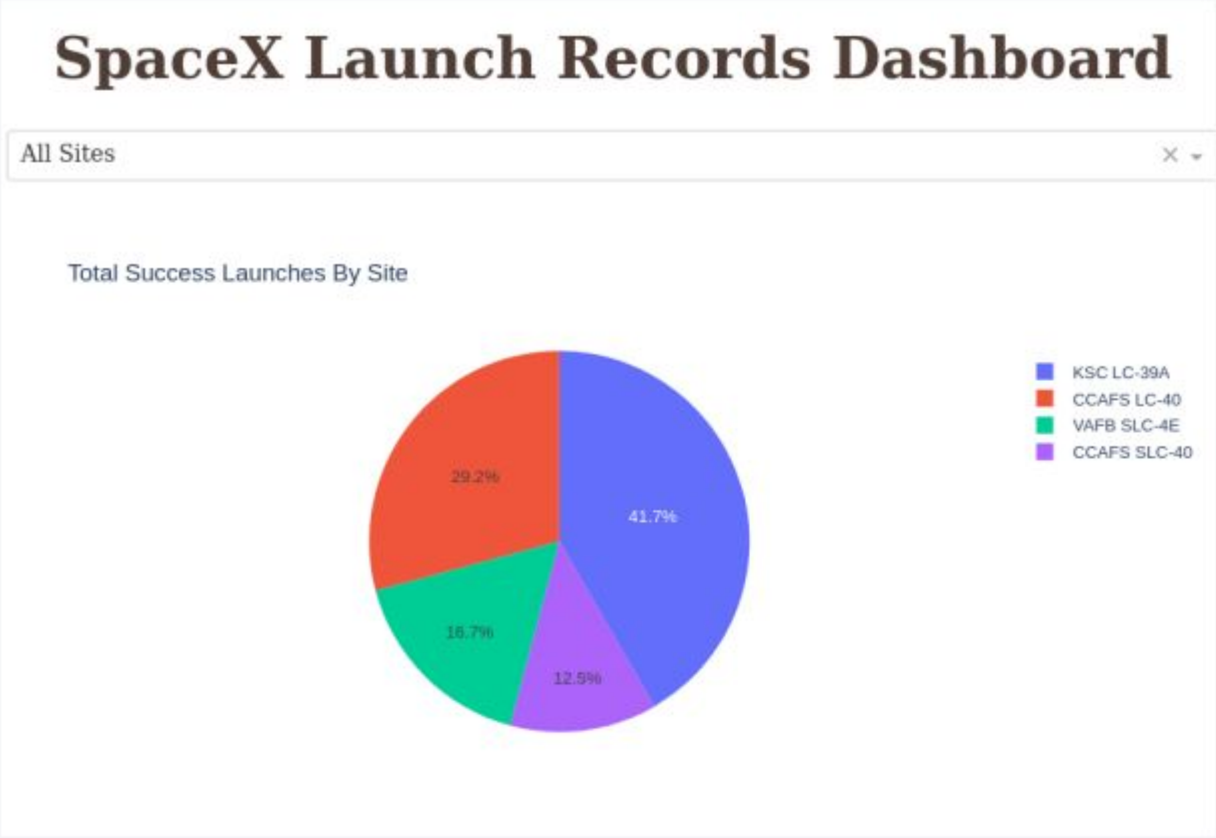


Section 4

Build a Dashboard with Plotly Dash

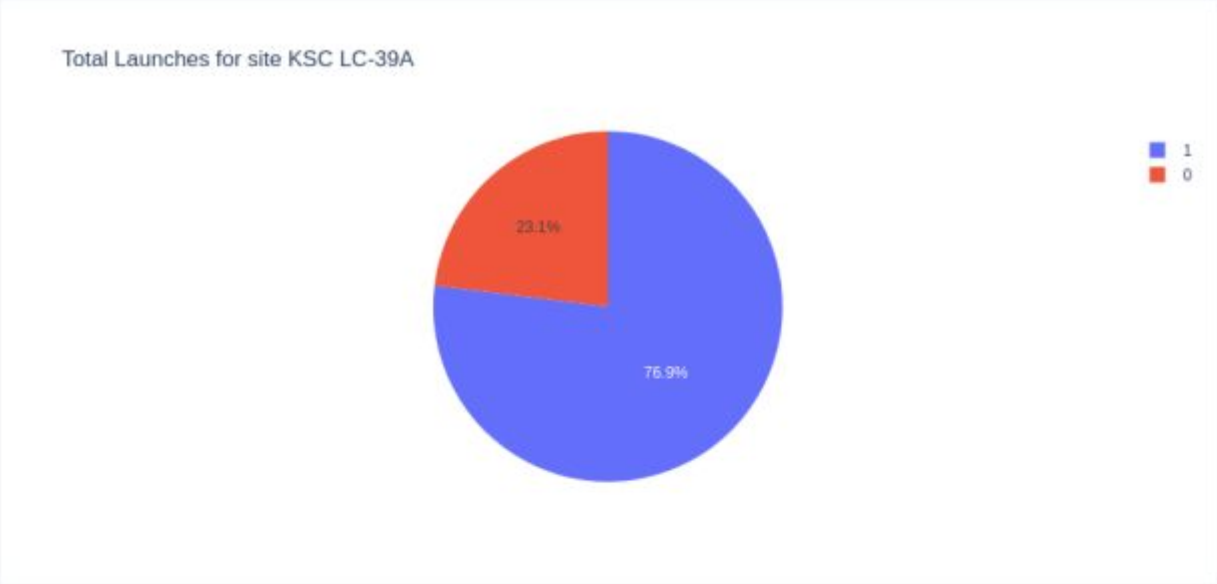
Sucessful launches by site

- The place from where launches are done seems to be a very important factor of success of missions.



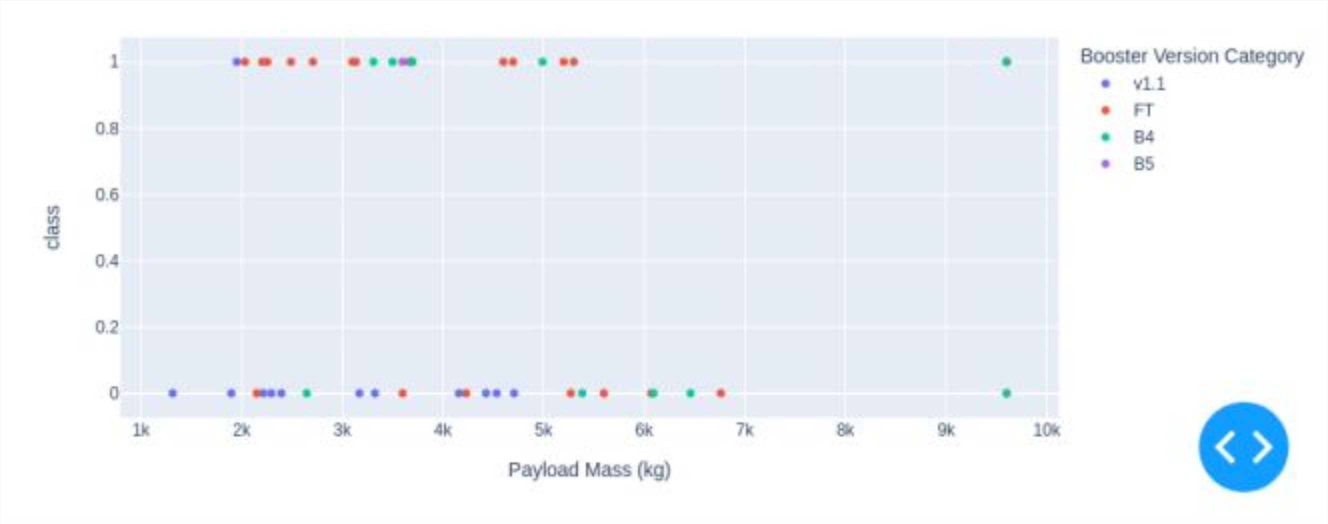
Launch Success Ratio for KSC LC-39A

- Most successes are in this site



Payload vs Launch Outcome

- Payloads under 6.000Kg and FT boosters are the most successful combination.



Section 5

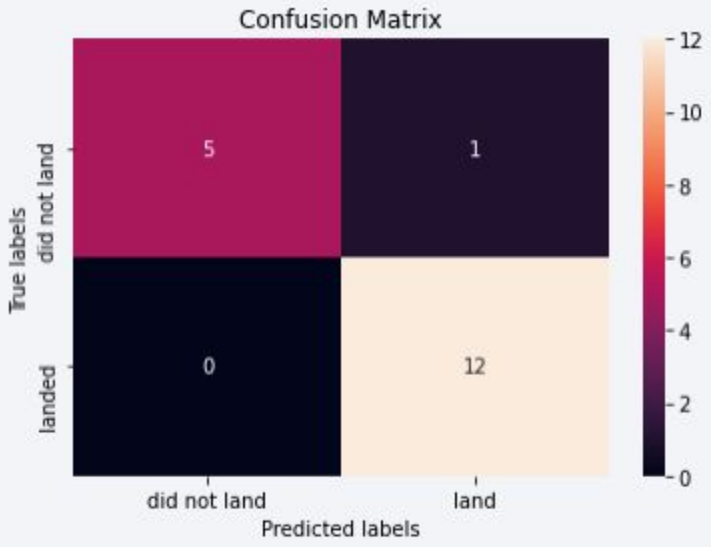
Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested:
    - KNN
    - Decisions Tree
    - SVM
    - Logistic regression
- 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:



Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7.000 Kg 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.

Thank you!