IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Şafak Cankan ŞAHİN>
<02.12.2023>

Outline

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

Executive Summary

- **Summary of methodologies**
- Data Collection using web scraping and SpaceX API ;
- Exploratory Data Analysis(EDA), including data wrangling, data visualization and interactive visual analytics.
- Machine Learning Prediction.
- **Summary of all results**
- It was possible to collected valuable data from public sources.
- EDA allowed to identify which features are the best to predict success of launchings.
- Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.

Introduction

- **The objective is to evaluate the viability of the new company to compete with Space X.**
- **Desirable answers:**
- The best way to estimate the total cost for launches, by predicting succesful landing 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 SpaceX was obtained from 2 sources:
  - SpaceX API
  - WebScraping
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features.

Methodology

Executive Summary

- 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 that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

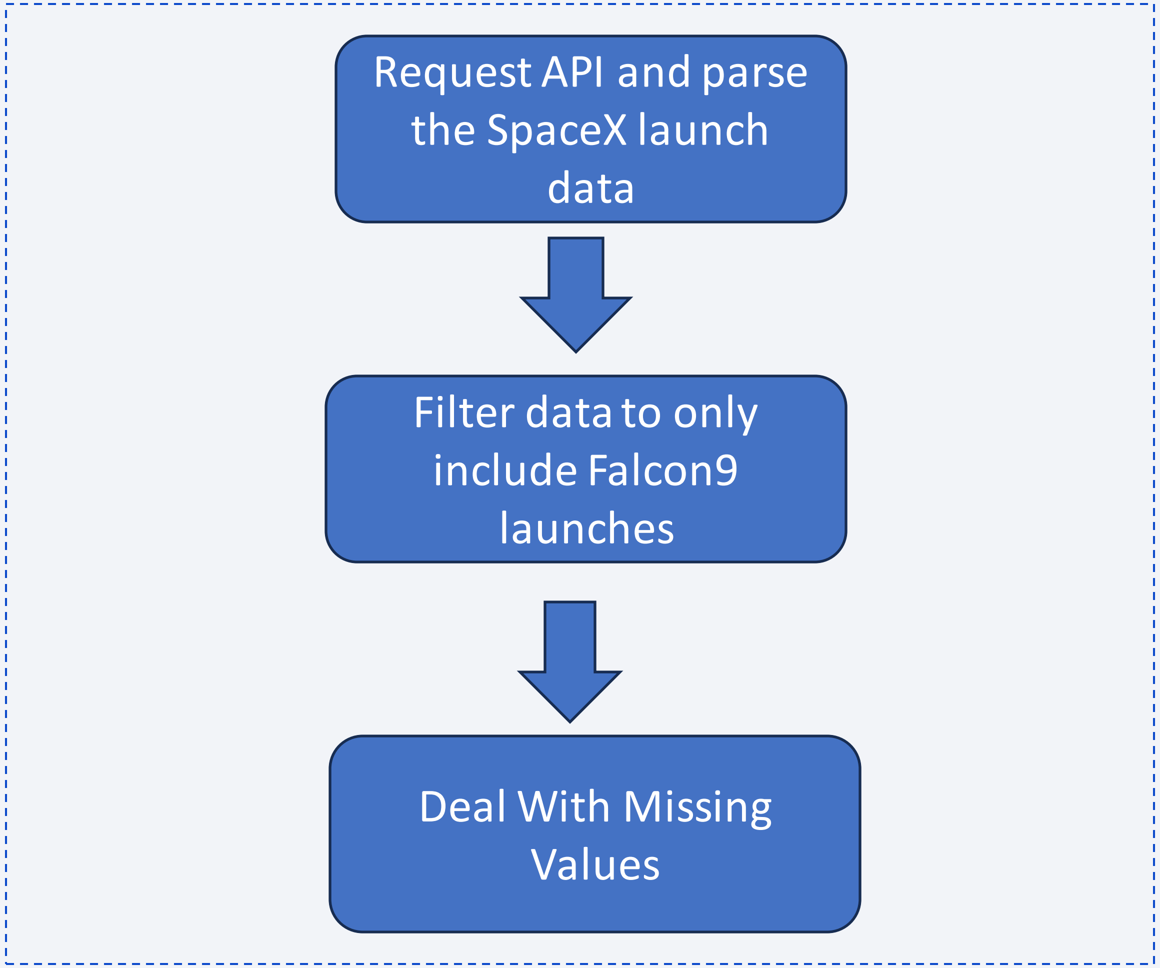
Data Collection

- Data sets were collected from SpaceX API and from Wikipedia using web scrapping technics.

Data Collection – SpaceX API

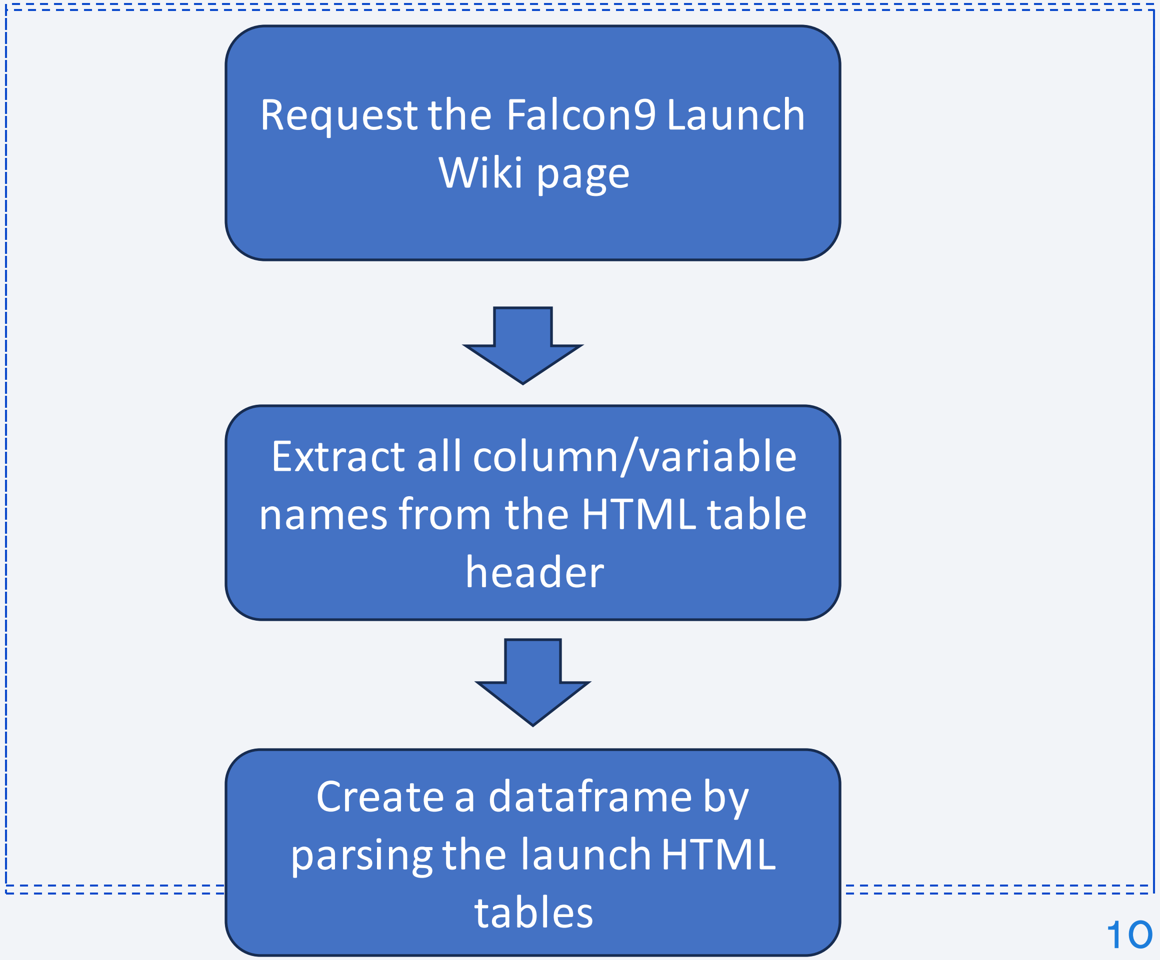
- SpaceX offers a public API from where data can be obtained and then used;
- This API was used according to the flowchart beside and then data is persisted.

- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia.
- Data are downloaded from Wikipedia according to the flowchart and then persisted.
- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb



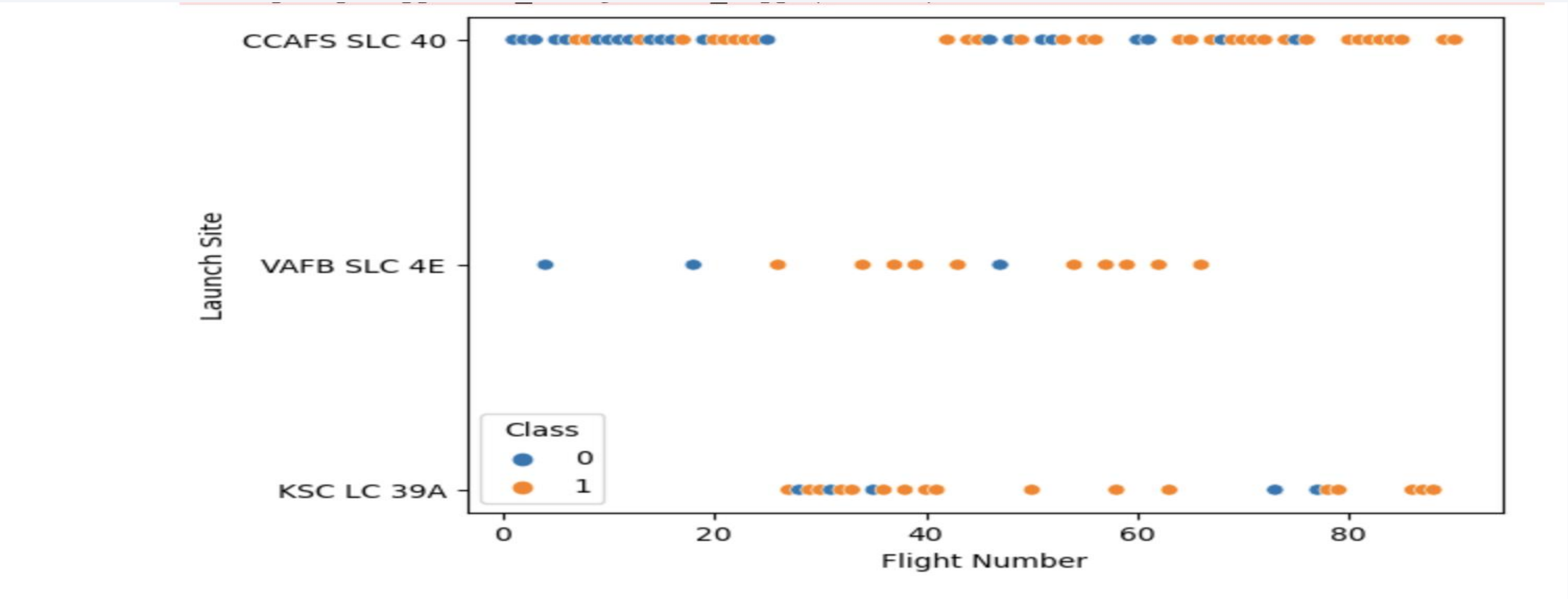
Data Wrangling

- 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.

- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:



- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- **The following SQL queries were performed:**

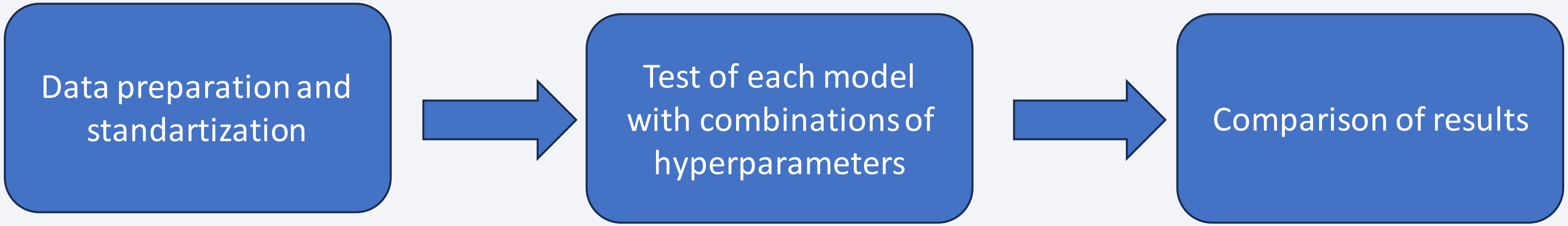
- Names of the unique launch sites in the space mission
- Top 5 records where launch sites begin with the string 'CCA'
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- Date when the first succesful landing outcome in ground pad was acheived.
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass. Use a subquery
- Failed landing in drone ship ,booster versions, launch site for the months in year 2015.
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.
- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.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 the best place to launch according to payloads.
- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/dash_interactivity.py

Predictive Analysis (Classification)

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



- Source code: https://github.com/Calaruin/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

- SpaceX uses 4 different launch sites.
- The first launches were done to SpaceX itself and NASA.
- The average payload of F9 v1.1 booster is 2,928 kg:
- The first success landing outcome happened in 2015 five year after the first launch.
- Many Falcon9 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.

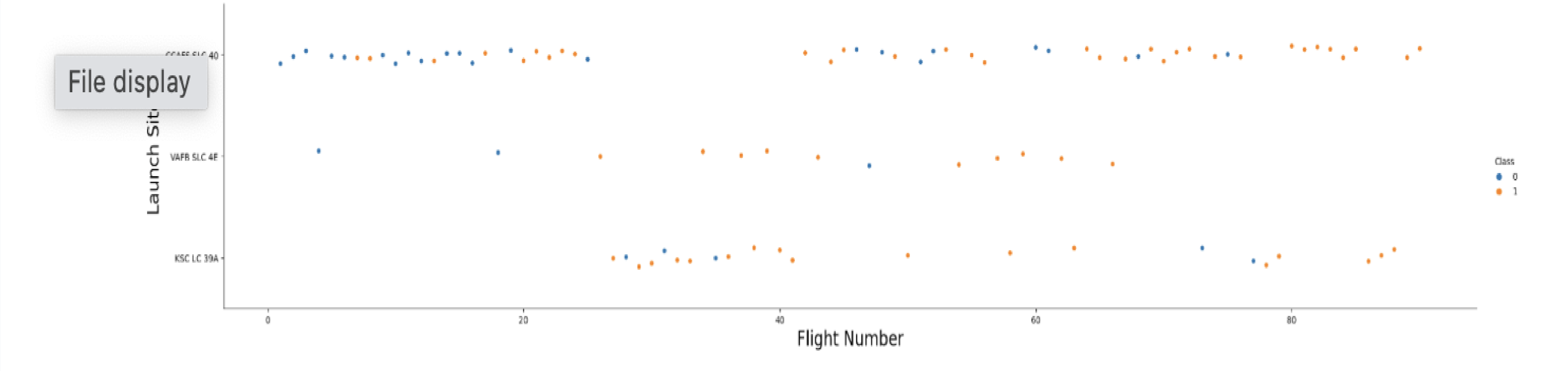
Results

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

Section 2

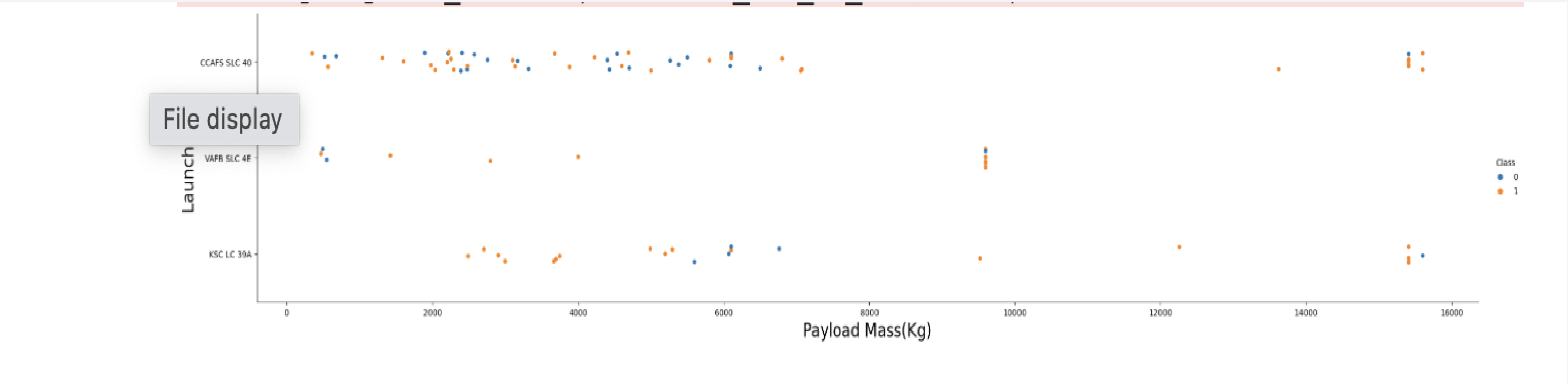
Insights drawn from EDA

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
- 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.

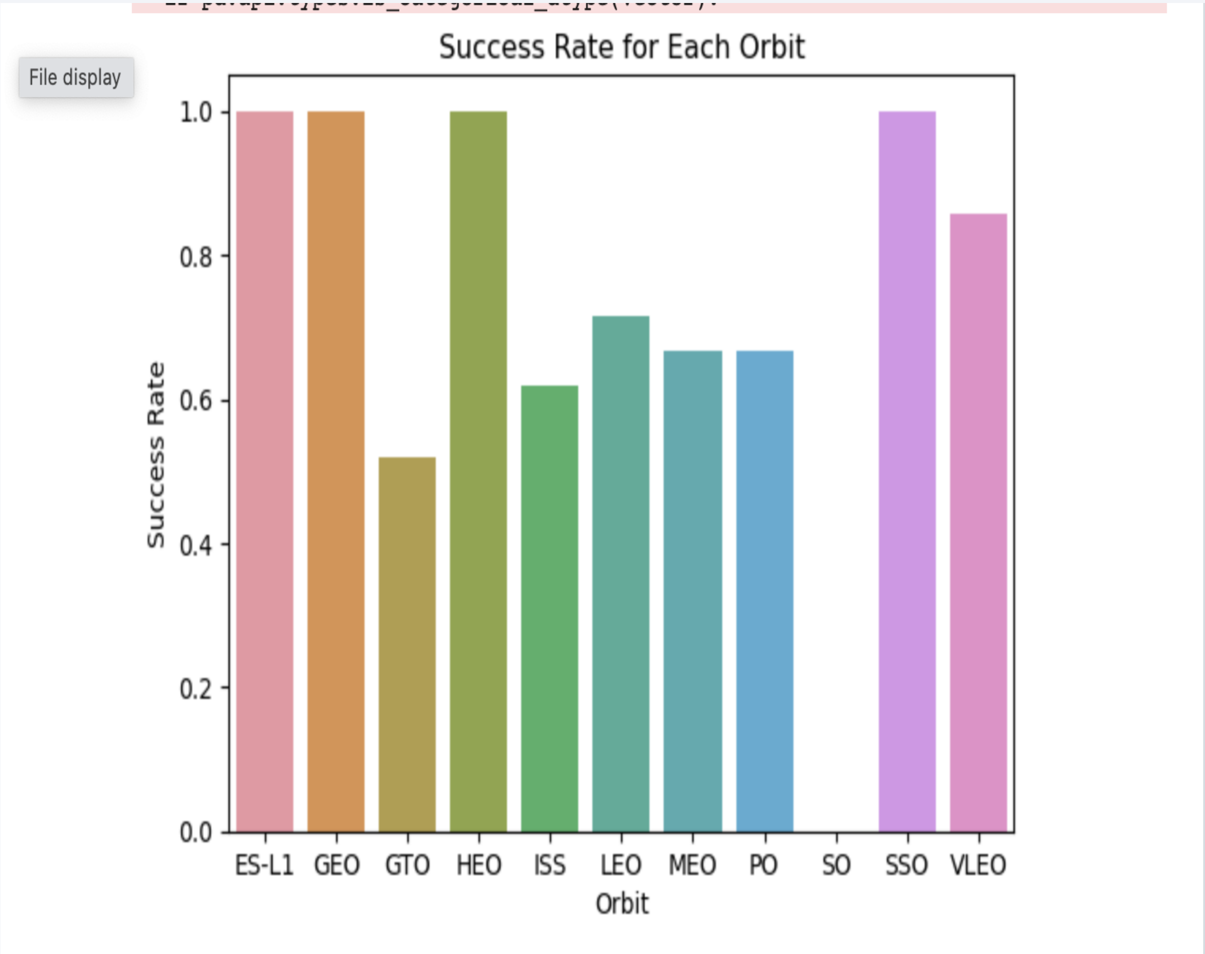
Payload vs. Launch Site



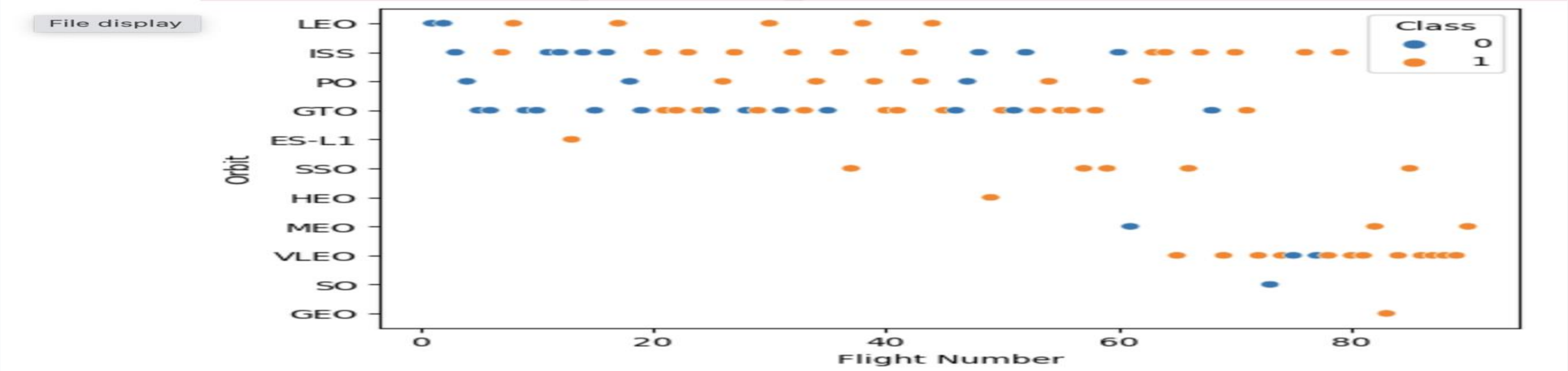
- Payloads over 9.000kg have excellent success rate.
- Payloads over 12.000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- **The biggest success rates happens to orbits:**
- ES-L1
- GEO
- HEO
- SSO
- **Followed by:**
- VLEO
- LFO

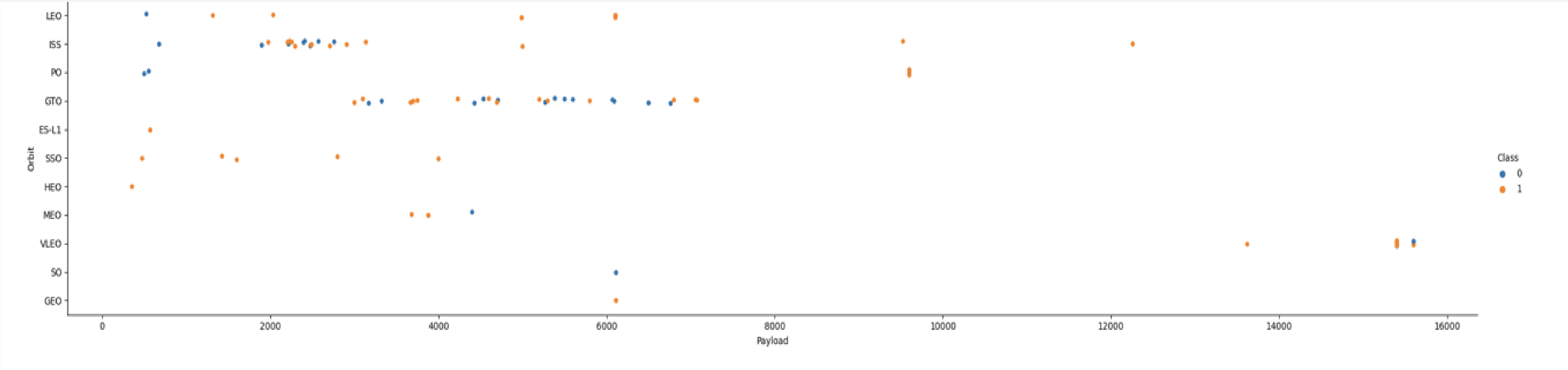


Flight Number vs. Orbit Type



- Apparently, success rate improved over time to all orbits.
- VLEO orbit seems a new business oppurtunity, due to recent increase of its frequency.

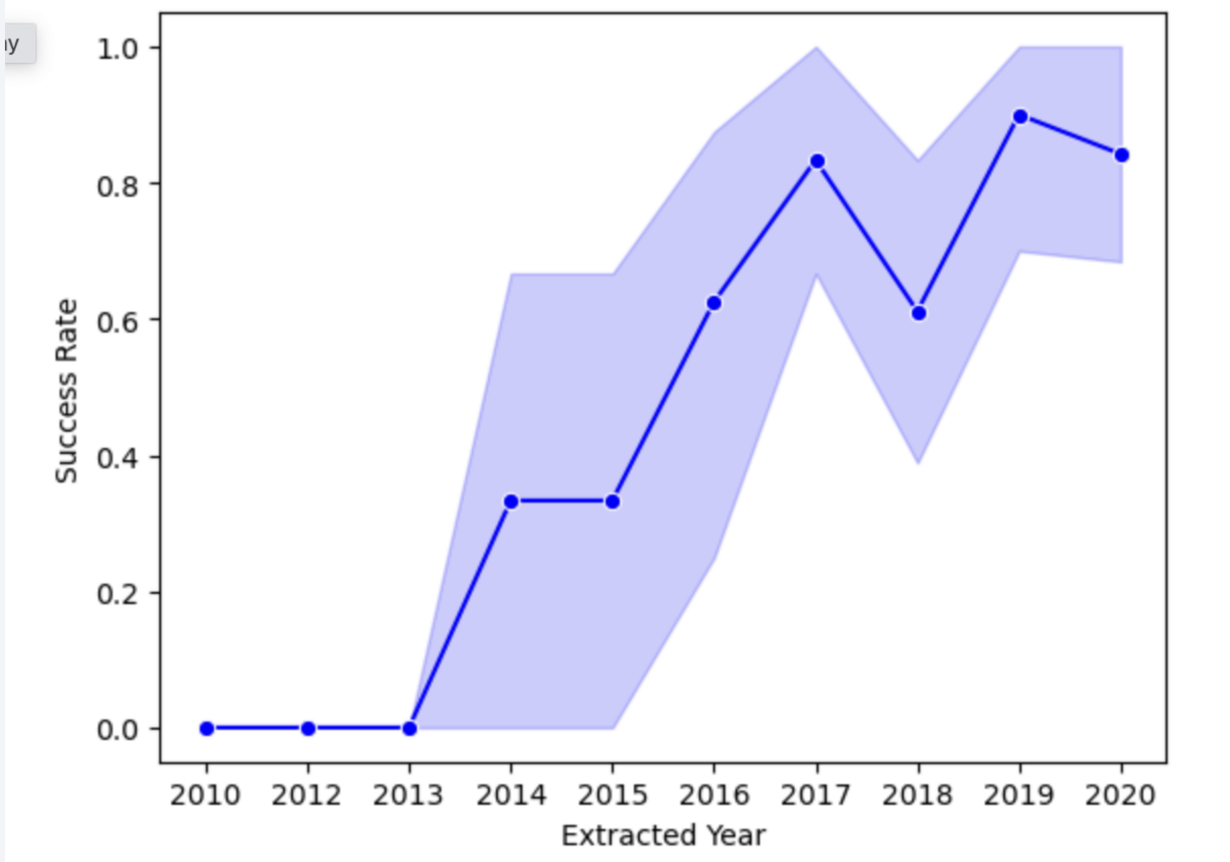
Payload vs. Orbit Type



- Apparently, there is no relation between payload and success rate to orbit GTO.
- ISS orbit has the widest range of payload and a good rate of success.
- There are few launches to the orbits SO and GEO.

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 adjust and improvement of technology.



All Launch Site Names

- According to data, there are four launch sites:

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

- They are obtained by selecting unique occurrences of "launch_site" values from the dataset.

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:

TOTAL_KG
45596

- Total payload calculated above, by summing all payloads whose codes contain "CRS" , which corresponds to NASA.

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1:

AVG_KG
2534.6666666666665

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2534,66 kg.

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

Date
2015-12-22

- By filtering data by successful landing outcome on ground pad and getting the minimum value for data it2s possible to identify the first occurrence, that happened on 22/12/2015.

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

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

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes.

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

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

- These are the boosters which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

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

Date	Booster_Version	Launch_Site	Landing_Outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- The list above has the only two occurrences.

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

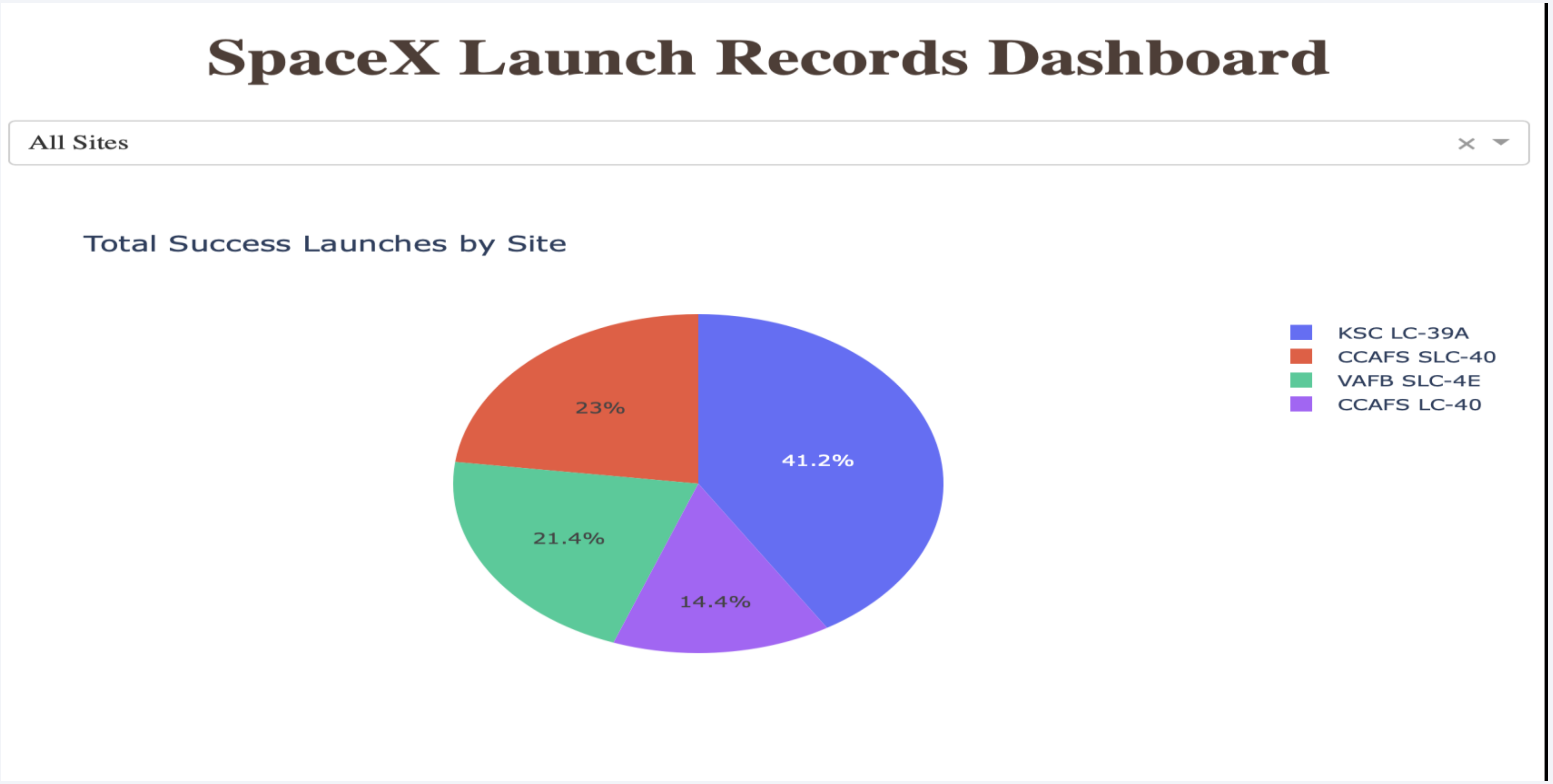
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Landing_Outcome	Landing_Count
Failure (drone ship)	5
Success (ground pad)	3

Section 4

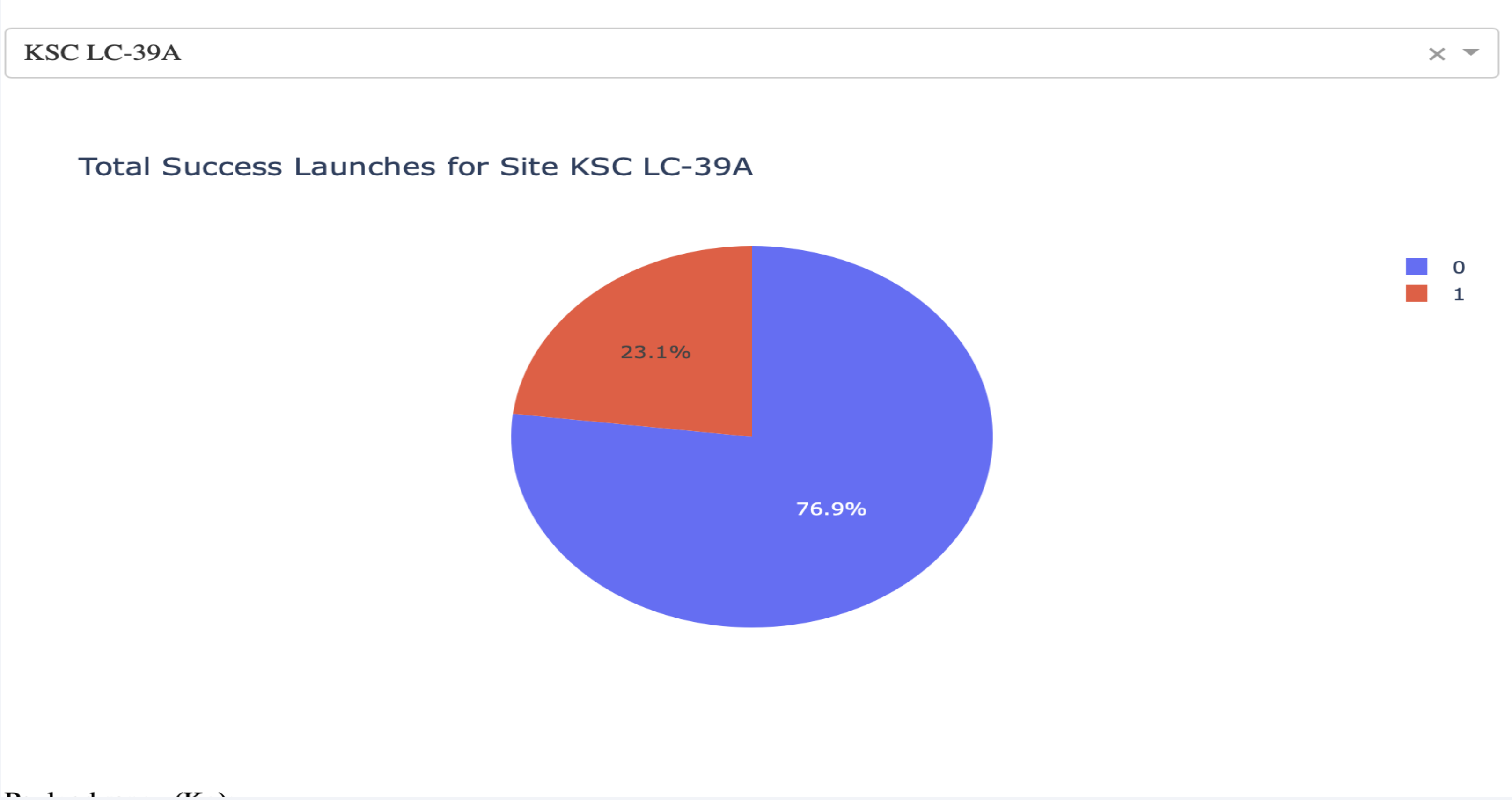
Build a Dashboard with Plotly Dash

Successful Launches by Site



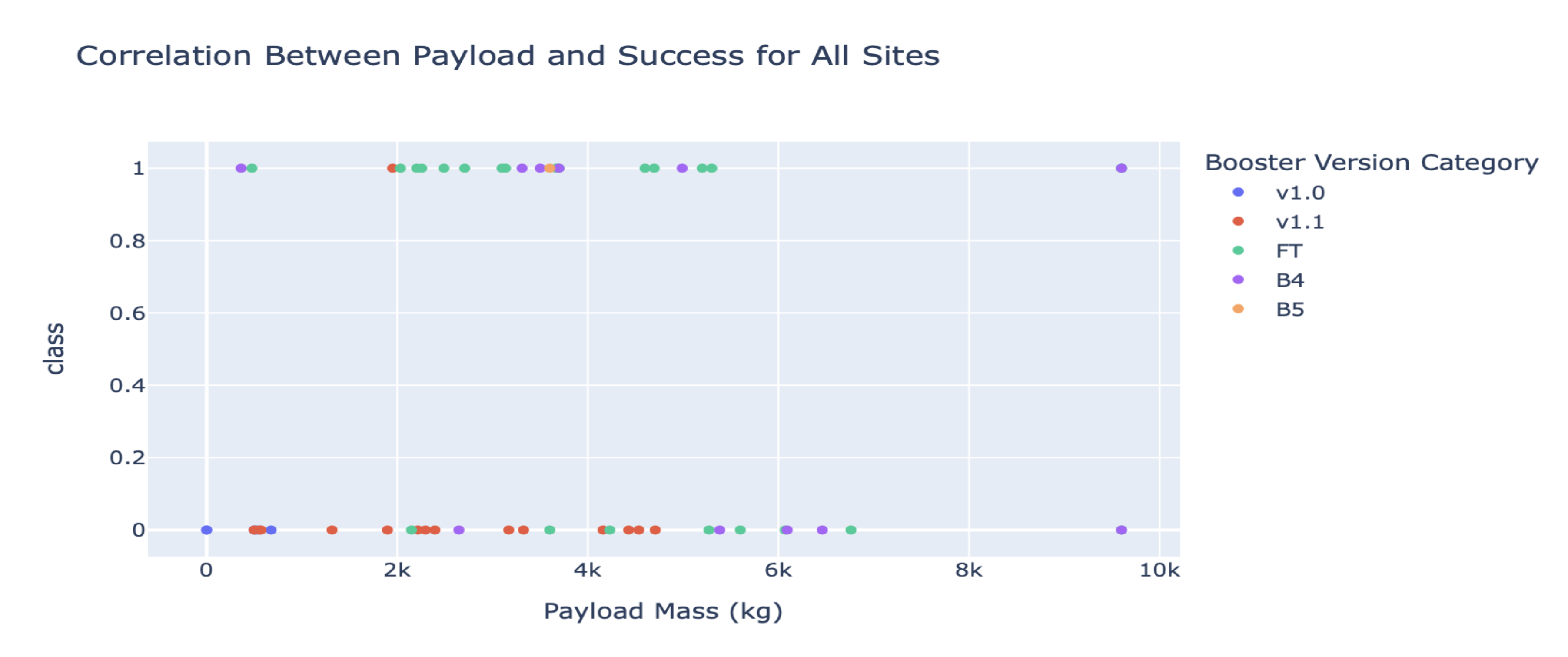
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



- %76.9 of launches are successful in this site.

Payload vs Launch Outcome



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

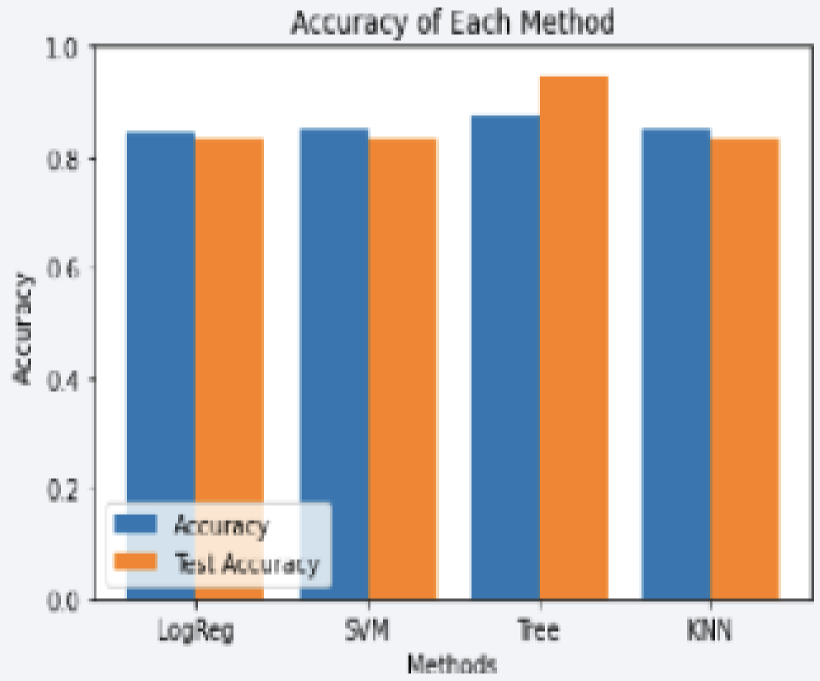
Section 5

Predictive Analysis (Classification)

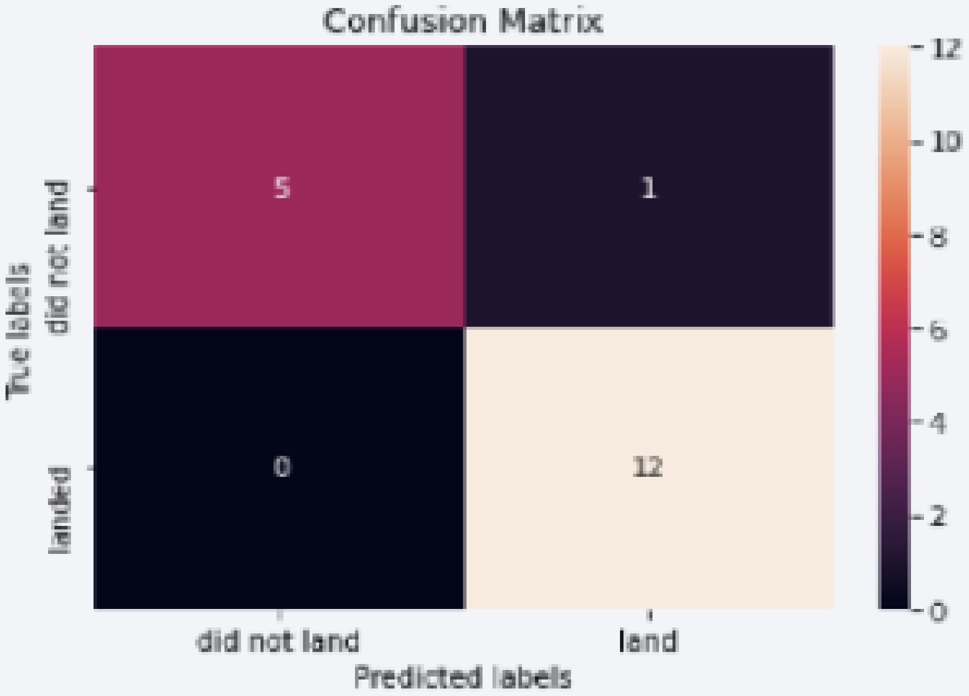
Classification Accuracy

- Four classification models were tested 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;

Thank you!