IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rubén Pérez
2023-01-24

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection Using SpaceX API and web scrapping.
  - Exploratory Data Analysys (EDA)
  - Machine Learning exercise
- Summary of all results
  - Valuable data collection from public sources.
  - Identifying the best prediction features with EDA.
  - Choosing the best model to predict the target data using machine learning techniques.

Introduction

- Project background and context
  - Stage one, despegue
  - Stage two, driving payload to orbit
- Problems you want to find answers
  - The price of eah launch
  - Gather information about SpaceX and creating dashboards.
  - Determine if SpaceX will reuse the first stage.
  - Using machine learning and public information

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data for the Space X's project was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/launches/past)
    - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- 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

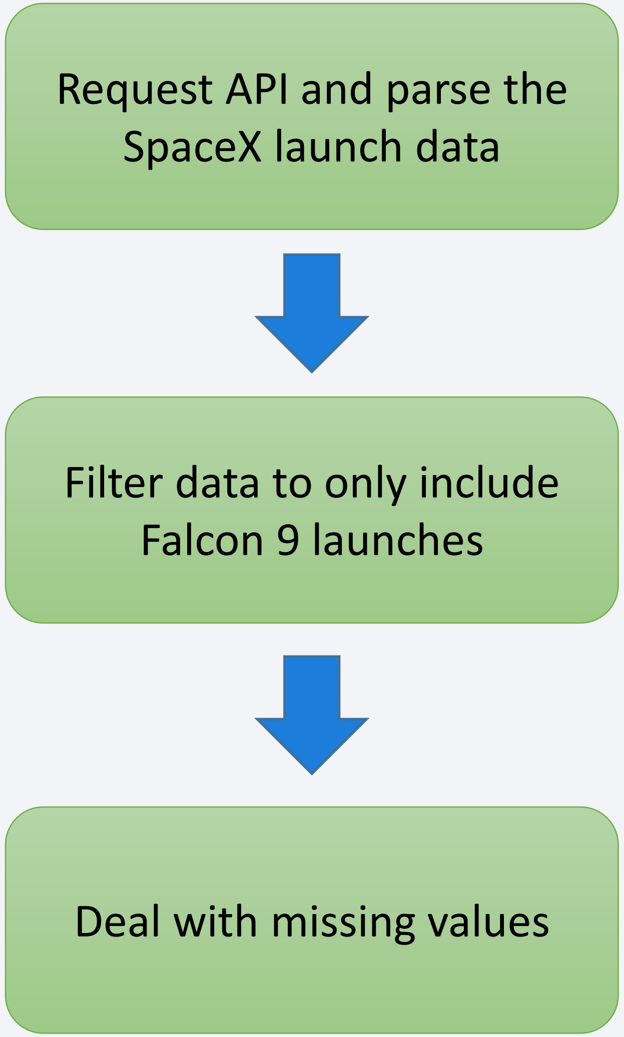
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

- First data set was collected directly using the Space X API requests (https://api.spacexdata.com/v4/rockets/), while the second data set was collected through WebScrapping techniques (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches).

Data Collection - SpaceX API

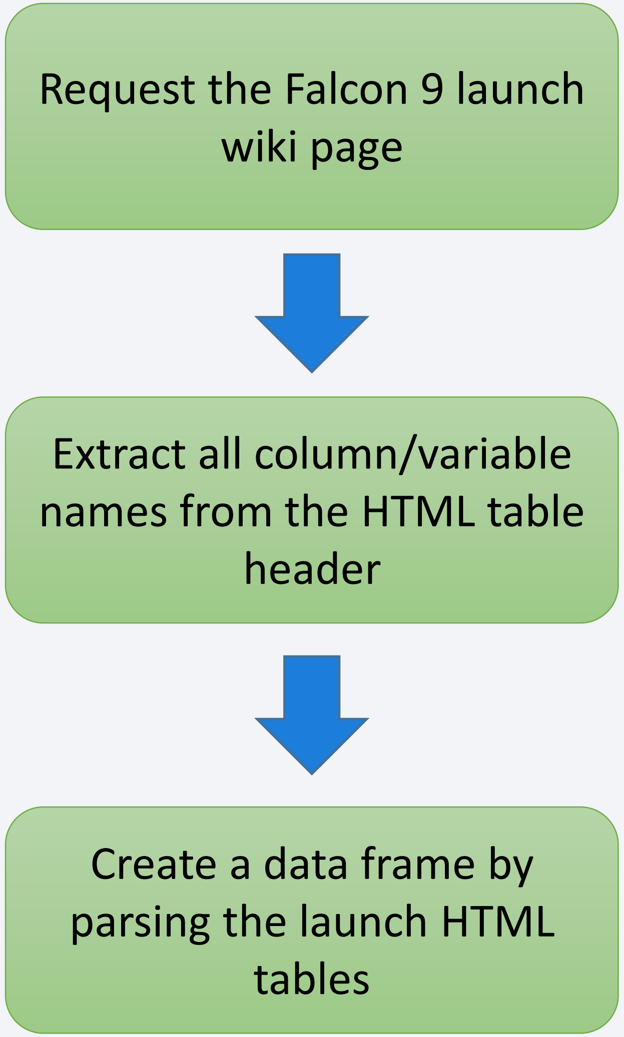
- Space X offers a public API from where data can be downloaded and esud
- This API was used according to the flowchart beside and then data is persisted.
- Source code: https://github.com/rperezpin/data_science_capstone_project/blob/44ea7d9dd4089d2db3e55a91017dfd51b38f605a/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

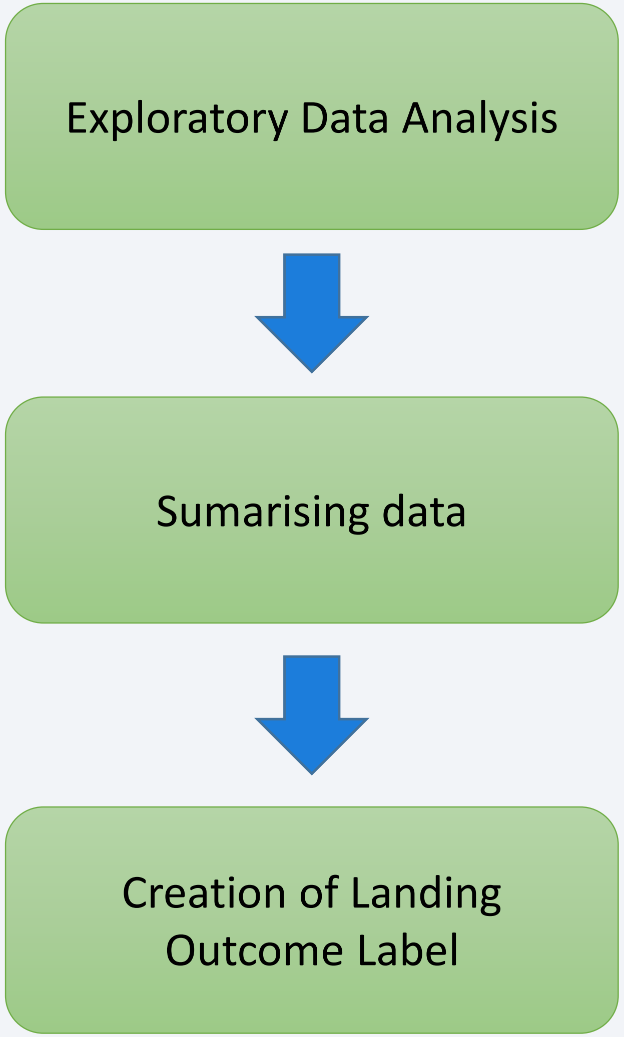
- Data from Wikipedia was directly obtained using python WebScraping

- Source code: https://github.com/rperezpin/data_science_capstone_project/blob/217eb0cda17d1163af5c4c35db1e8ea0d1eda83c/jupyter-labs-webscraping.ipynb



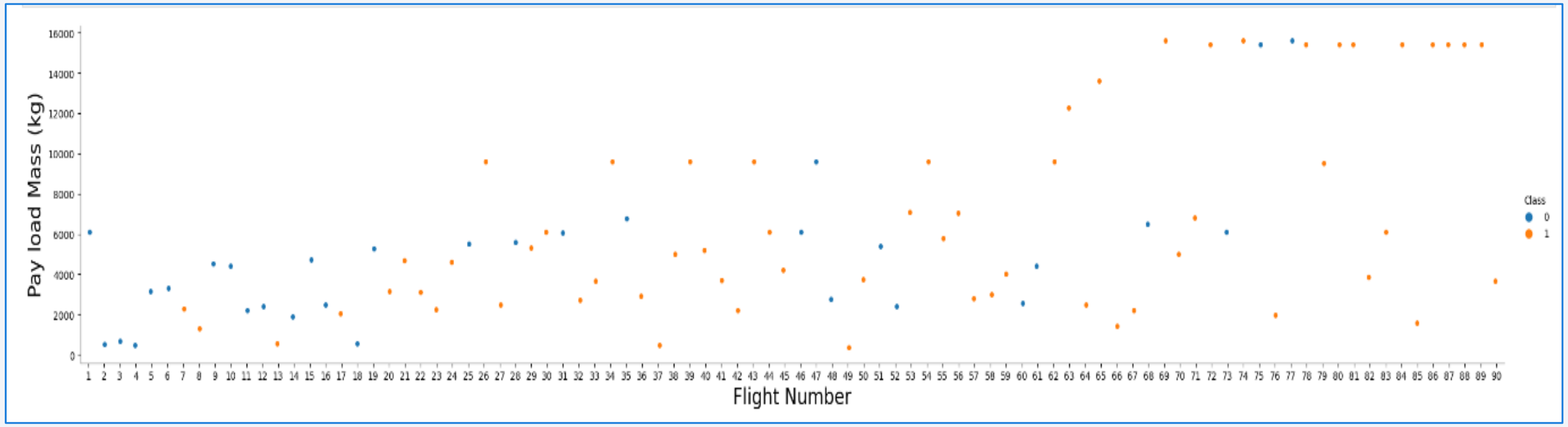
Data Wrangling

- It was performed a little exploratory data analysis and then summarized per site, occurrences of each orbit and occurrences of mission outcome per orbit type.
- Finally it was created the landing outcome label from the outcome column.



EDA with Data Visualization

- To visualize the data there have been used scatterplots and barplots due to the simply way to see the relationship between features.



- Source code: https://github.com/rperezpin/data_science_capstone_project/blob/b2c8cd656237f8398285a0135933f104b7a7b25f/jupyter-labs-eda-dataviz%20(1).ipynb

EDA with SQL

- The SQL queries that were performed are the following:
  - 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.
- Source code: https://github.com/rperezpin/data_science_capstone_project/blob/b2c8cd656237f8398285a0135933f104b7a7b25f/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

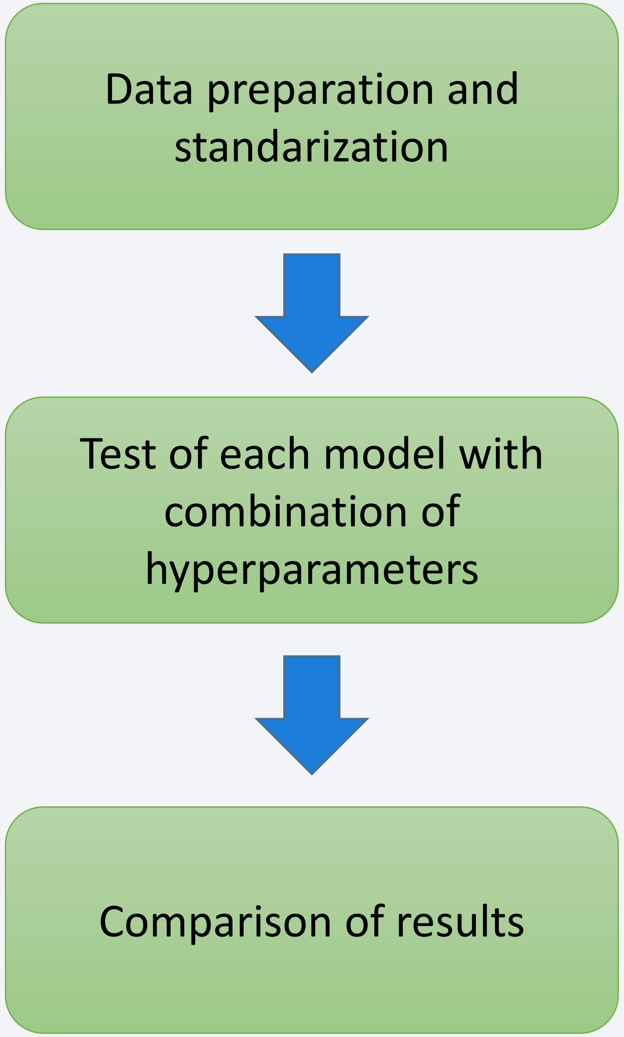
- Markers, circles, lines and marker clusters were used with Folium Maps.
  - Markers indicate points like launch sites.
  - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center.
  - Marker clusters indicates groups of events in each coordinate, like launches in a launch site; and.
  - Lines are used to indicate distances between two coordinates.
- Source code: https://github.com/rperezpin/data_science_capstone_project/blob/7da5ea7346664d2fbb2f2e2a203a4207930c03b6/lab_jupyter_launch_site_location%20(1).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.
- Source code: https://github.com/rperezpin/data_science_capstone_project/blob/7da5ea7346664d2fbb2f2e2a203a4207930c03b6/ploty_ashboard.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/rperezpin/data_science_capstone_project/blob/7da5ea7346664d2fbb2f2e2a203a4207930c03b6/SpaceX_Machine%20Learning%20Prediction_Part_5%20(1).ipynb

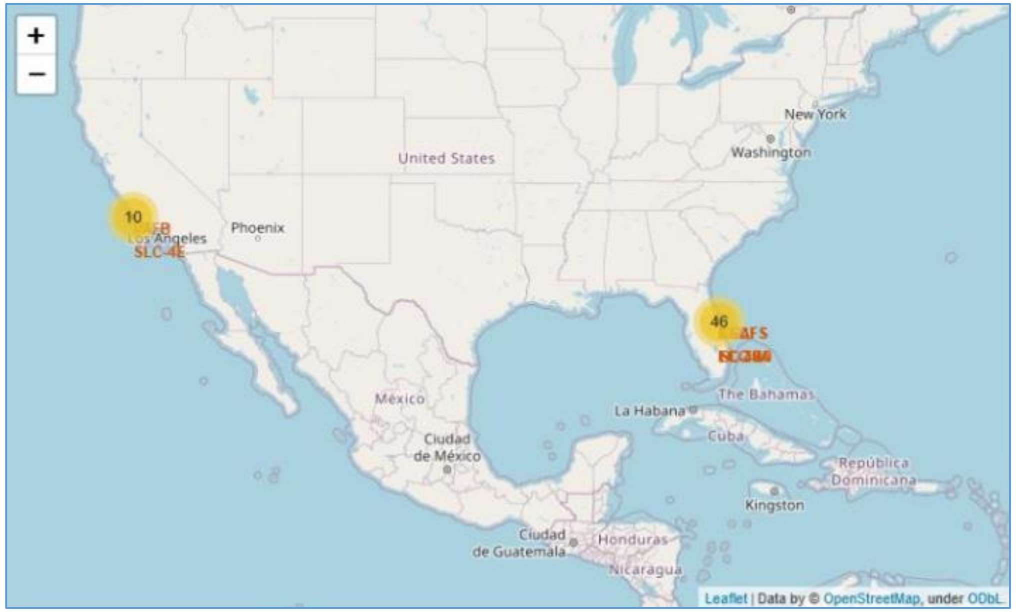


Results

- 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.

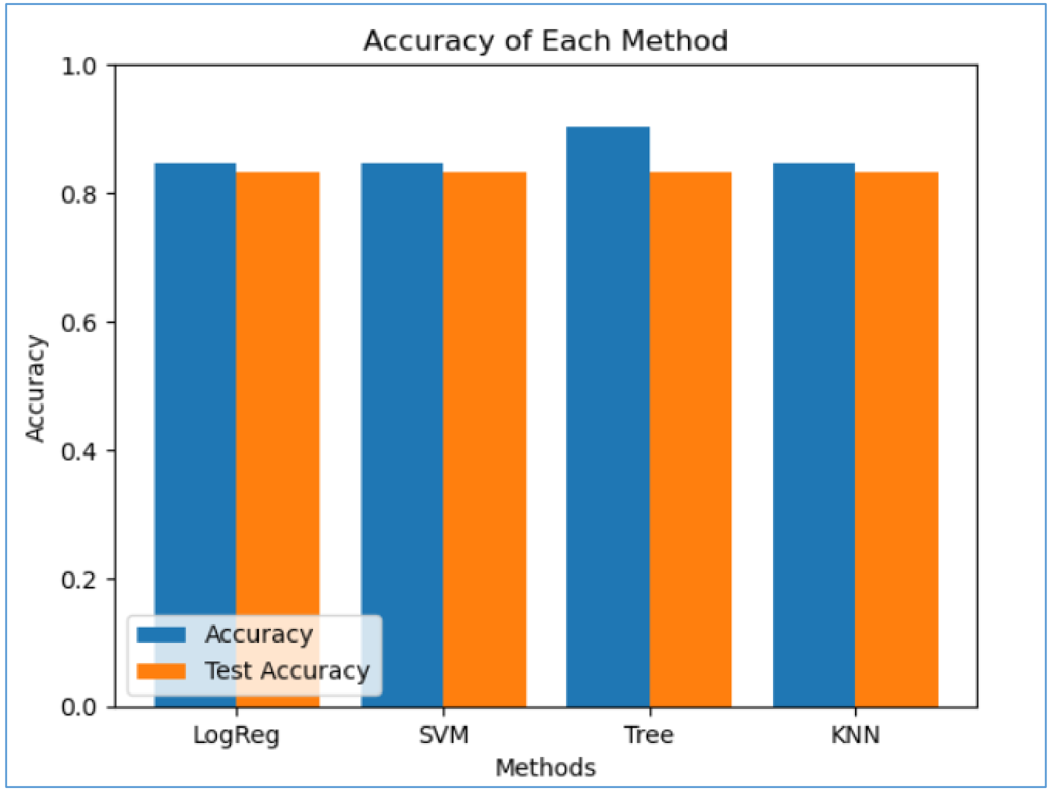
Results

- 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.



Results

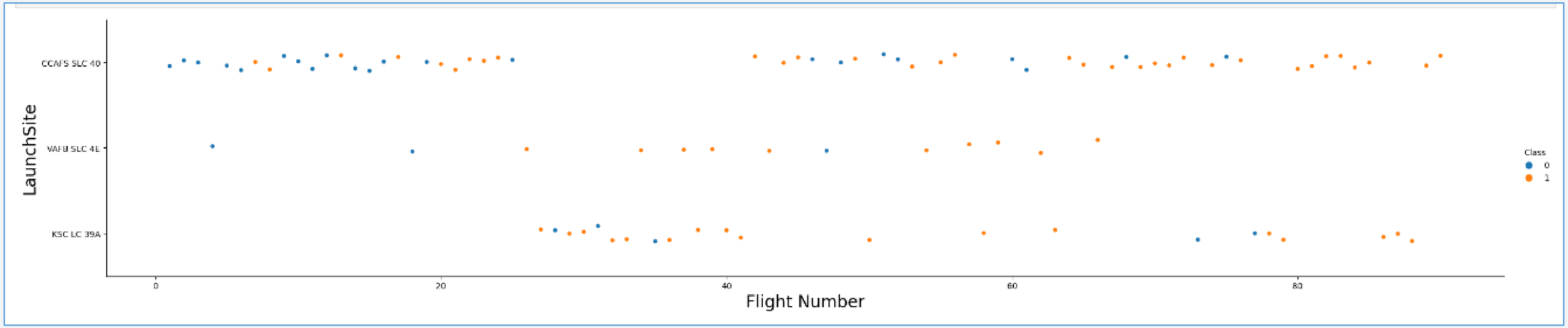
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 90% and accuracy for test data over 83%.



Section 2

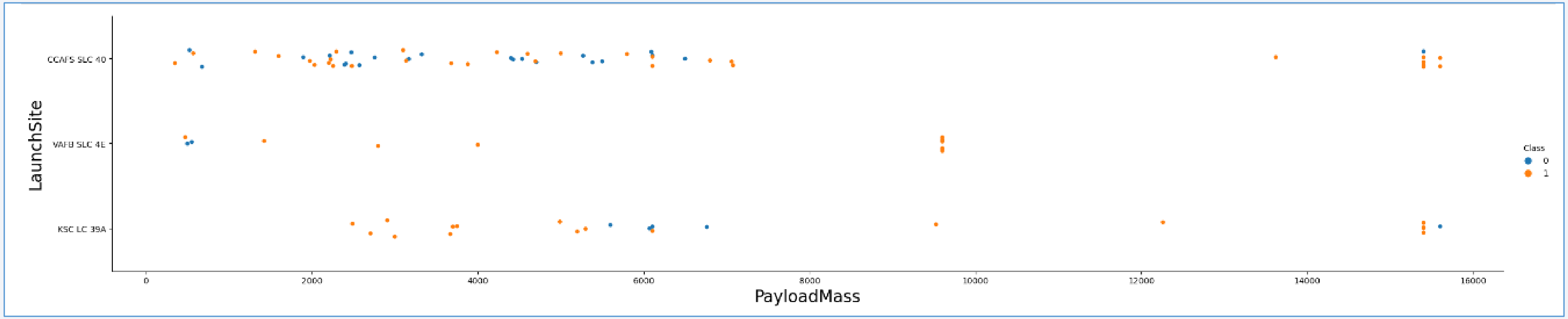
Insights drawn from EDA

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.

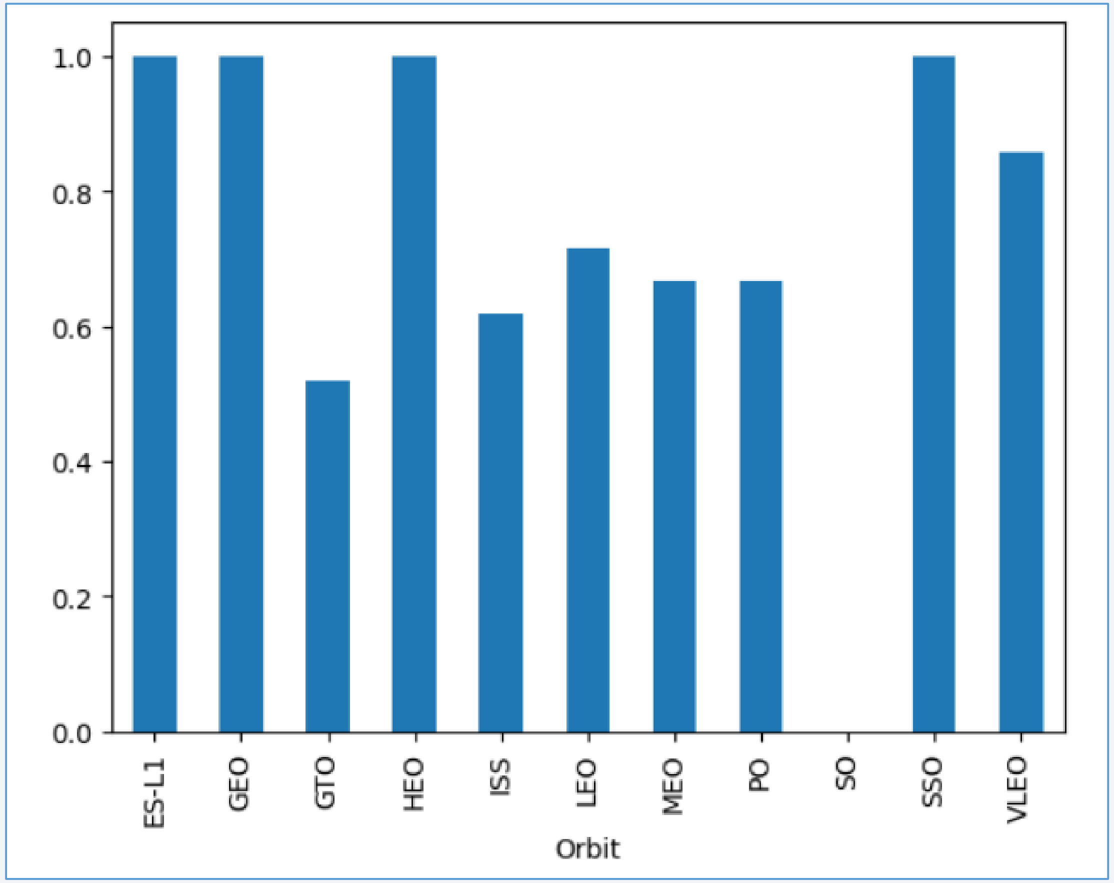
Payload vs. Launch Site



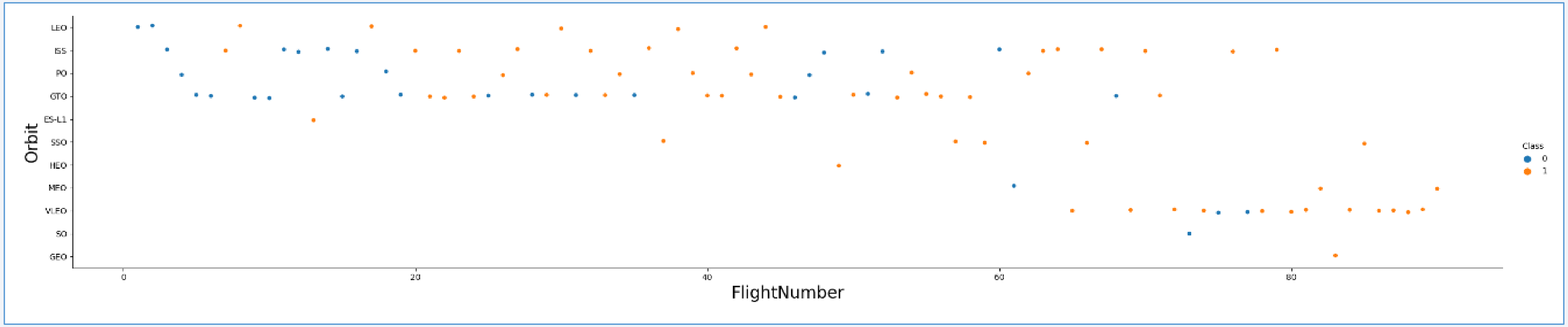
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate.
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites

Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
  - ES-L1.
  - GEO.
  - HEO.
  - SSO.
- Followed by:
  - VLEO (above 80%).
  - LFO (above 70%).

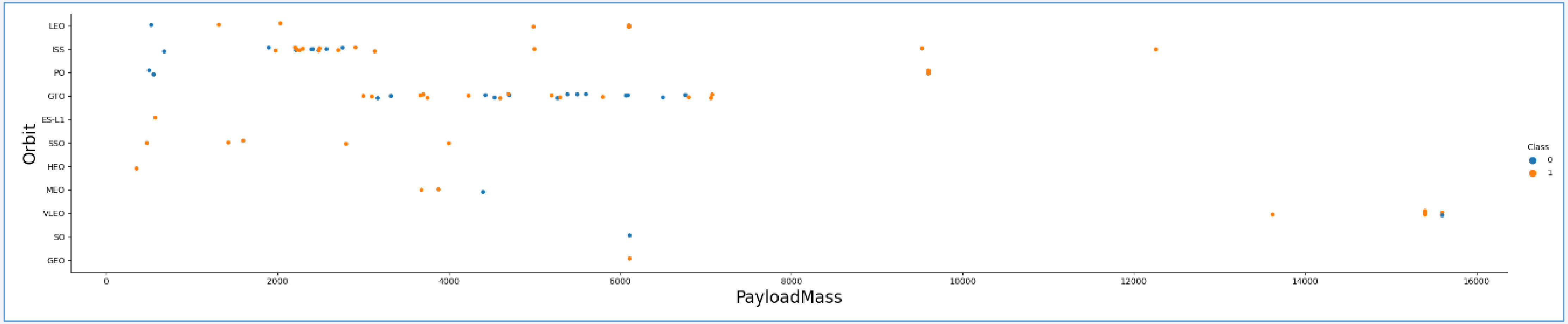


Flight Number vs. Orbit Type



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

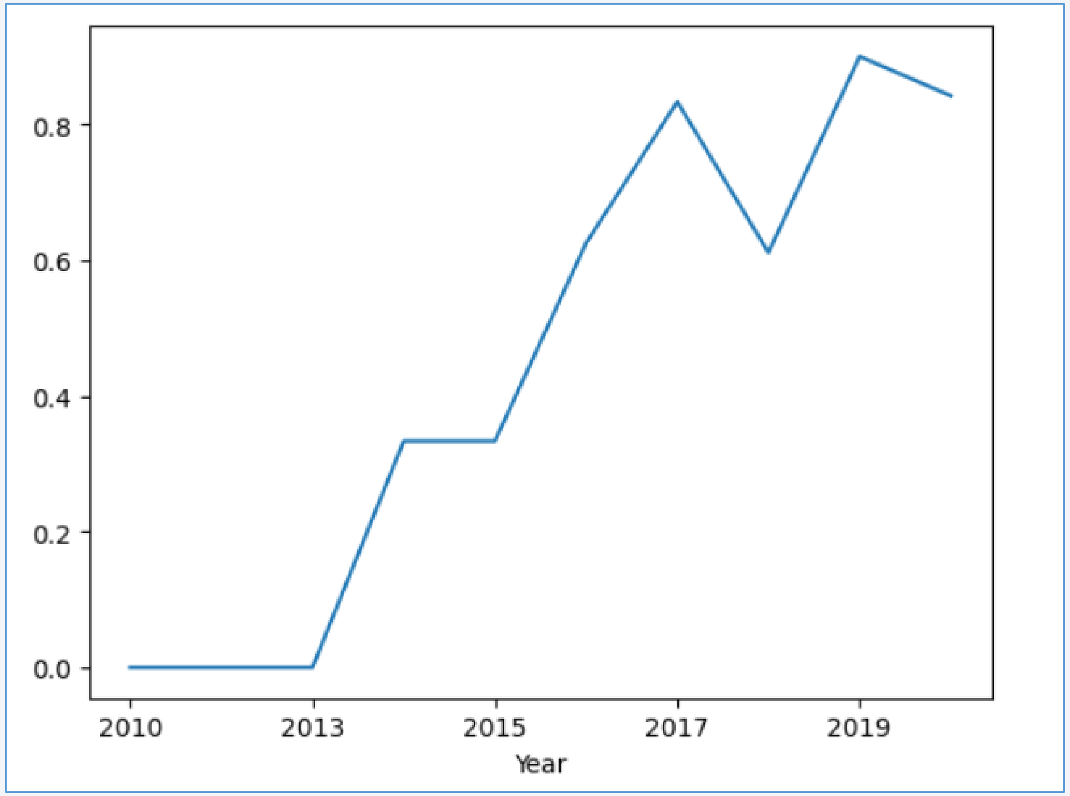
Payload vs. Orbit Type



- Shorently, 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 adjusts and improvement of technology.



All Launch Site Names

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

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 attempt

Total Payload Mass

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

total_payload
111268

Average Payload Mass by F9 v1.1

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,928 kg.

avg_payload
2928

First Successful Ground Landing Date

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

first_success_gp
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

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

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

Total Number of Successful and Failure Mission Outcomes

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

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

booster_version
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

- The list above has the only two occurrences.

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

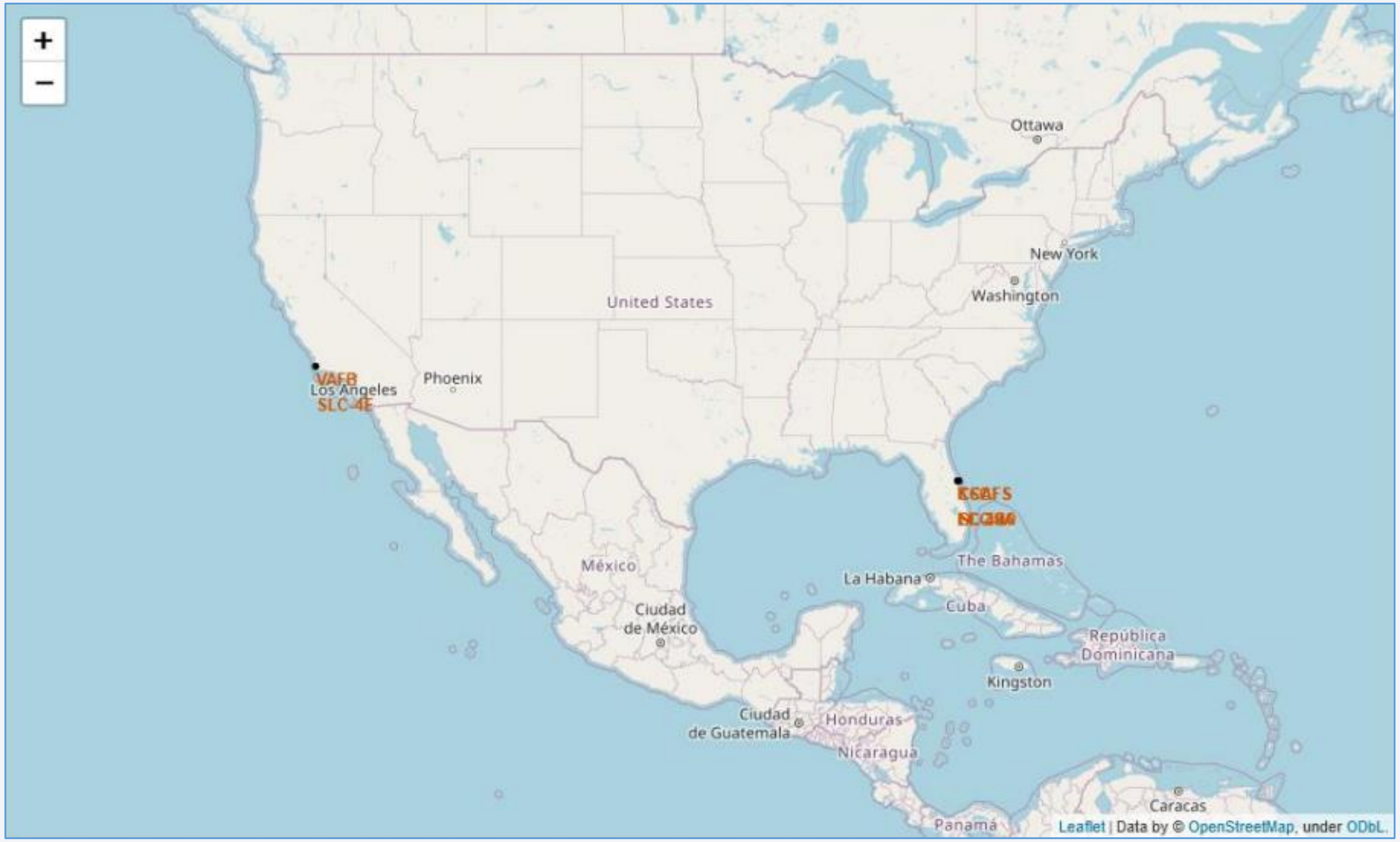
- Ranking of all landing outcomes between the date 2010-06-04 and 2017- 03-20:
- his view of data alerts us that “No attempt” must be taken in account.

landing_outcome	qty
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

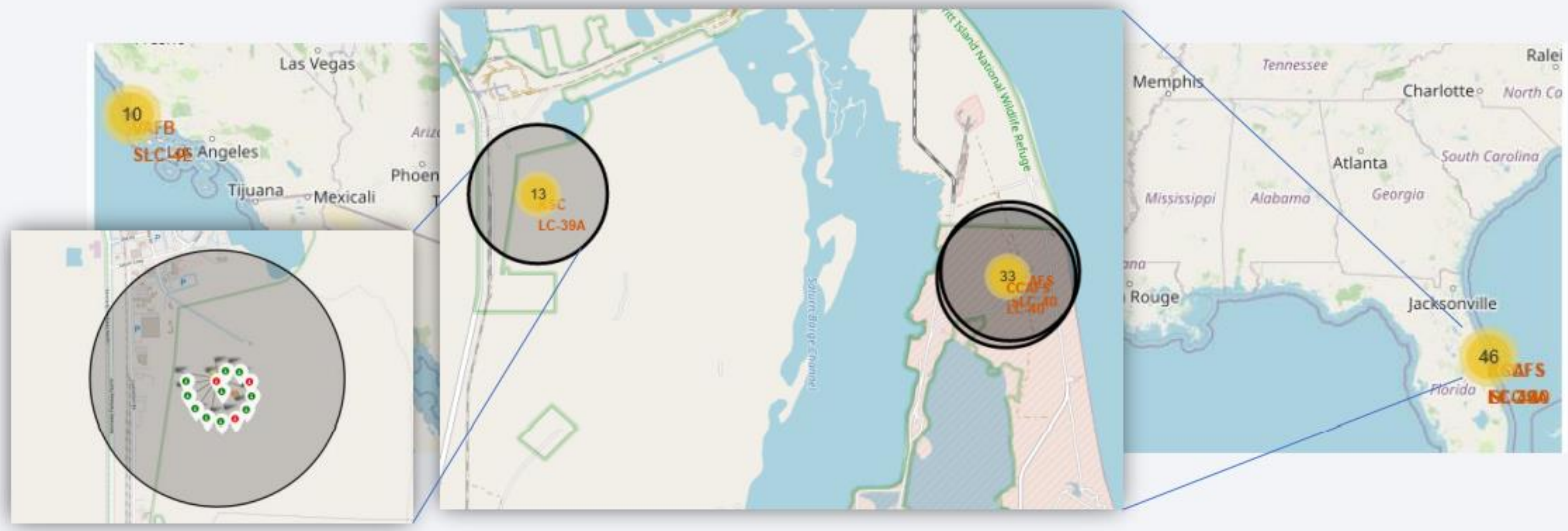
All launch sites



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

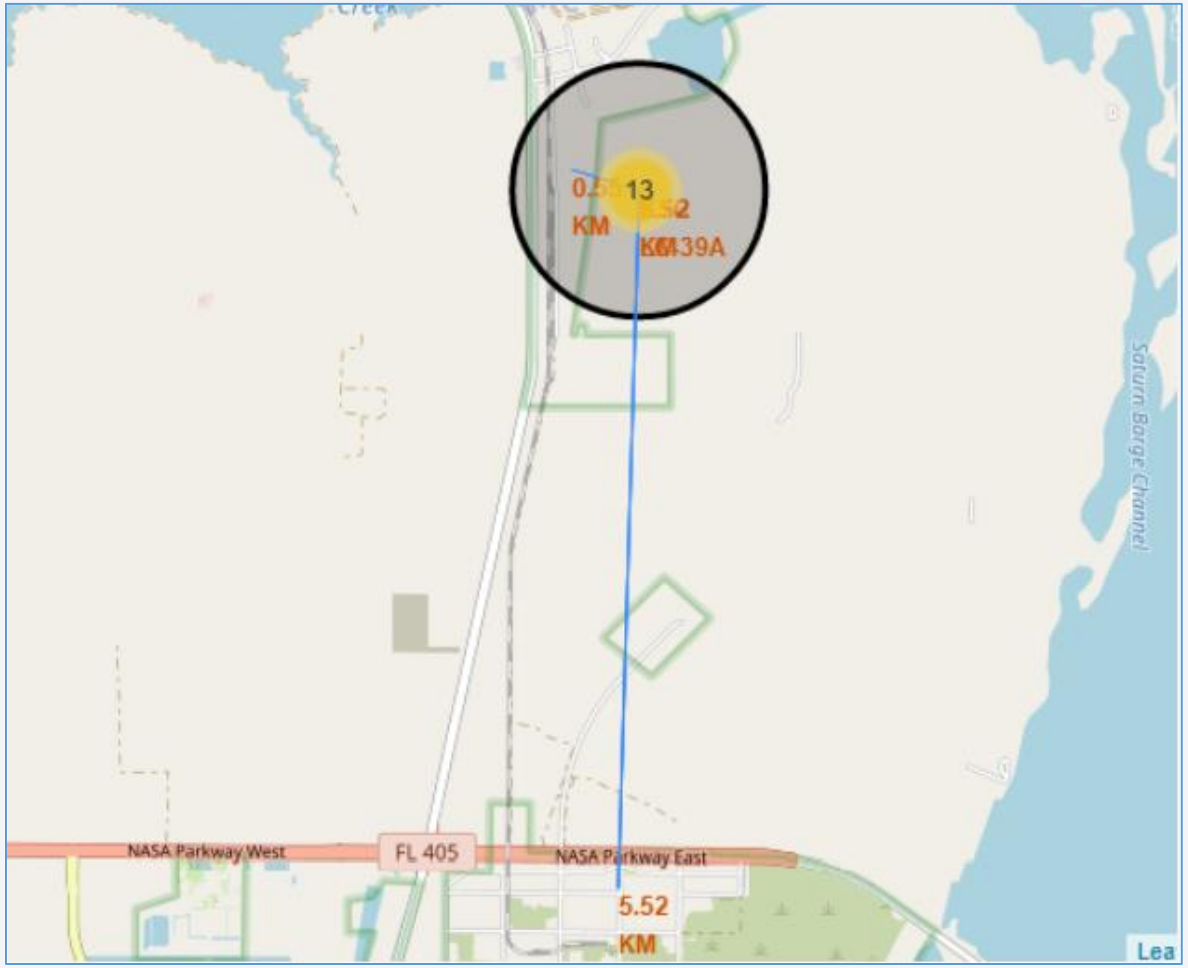
Launch Outcomes by site

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

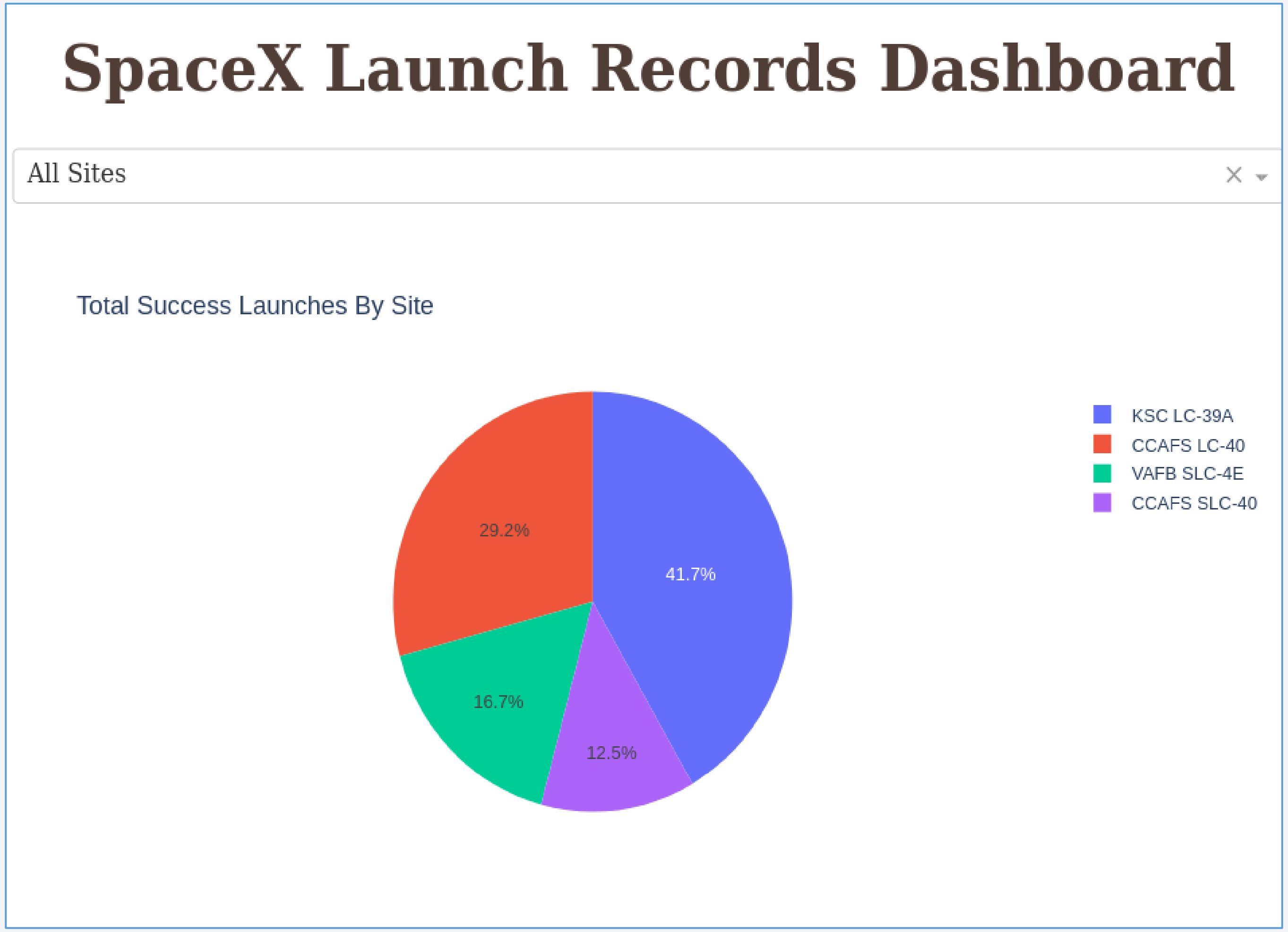


Section 4

Build a Dashboard with Plotly Dash

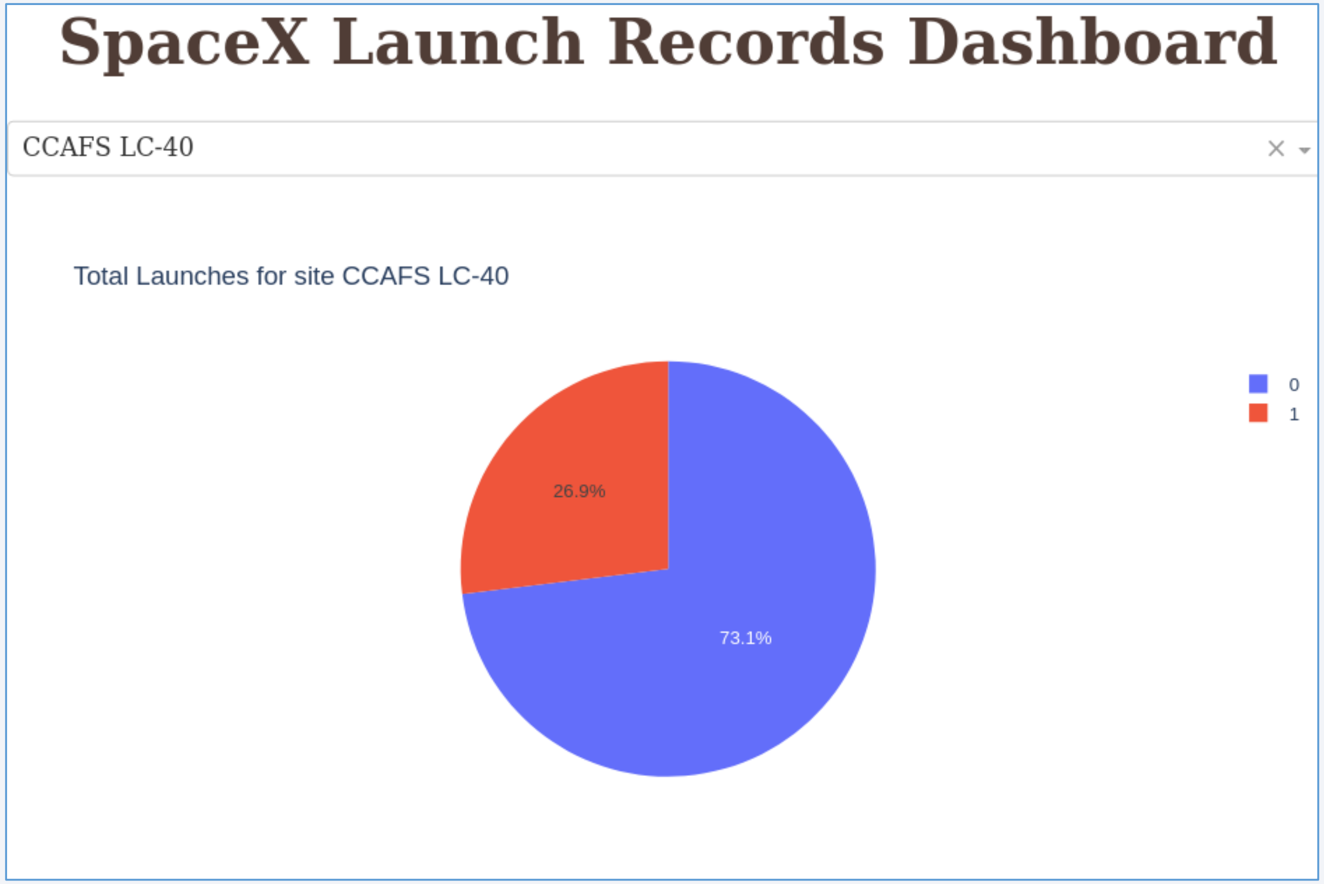
Successful Launches by Site

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



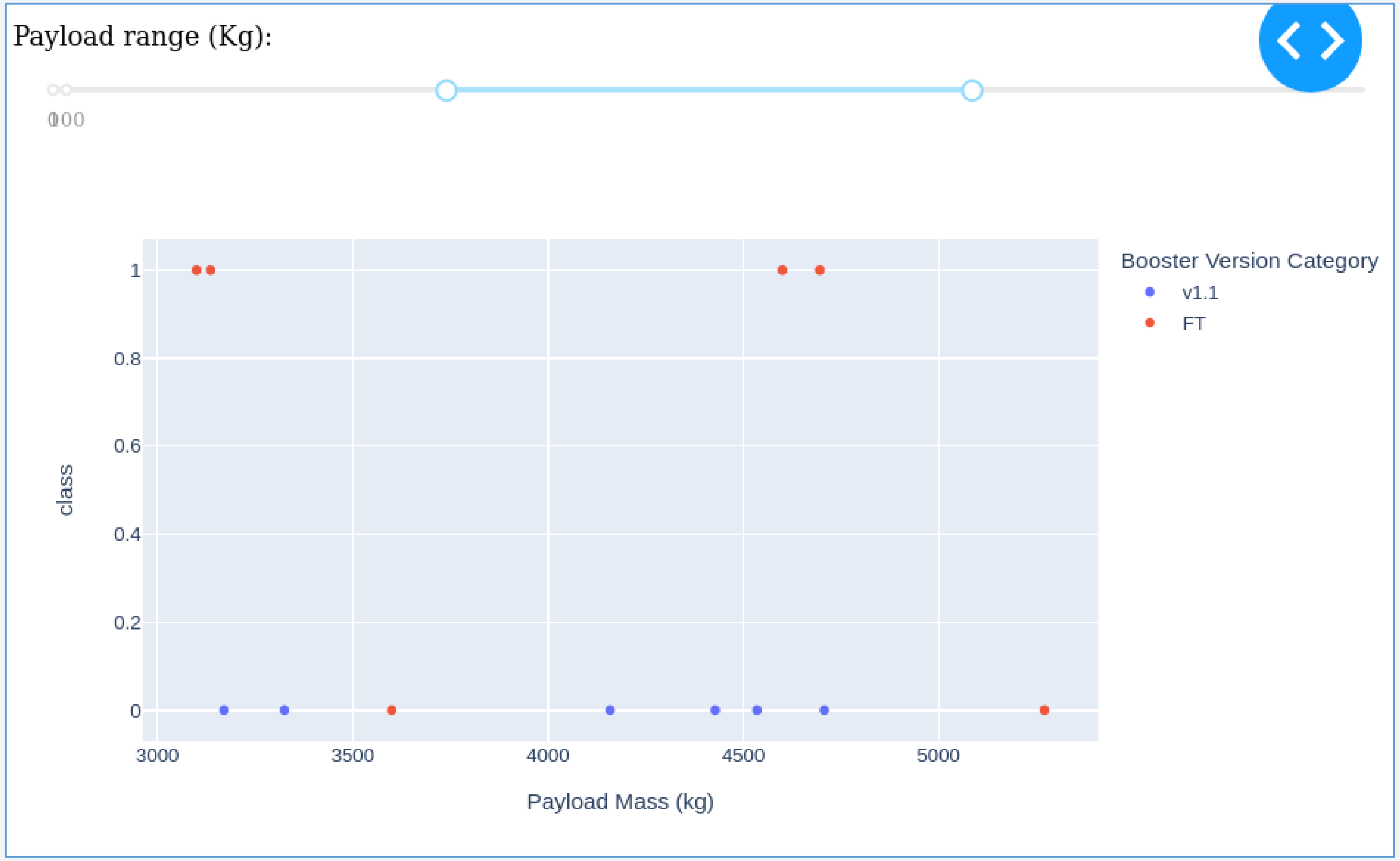
Launch Success Ratio for KSC LC-39A

- 76.9% of launches are successful in this site.



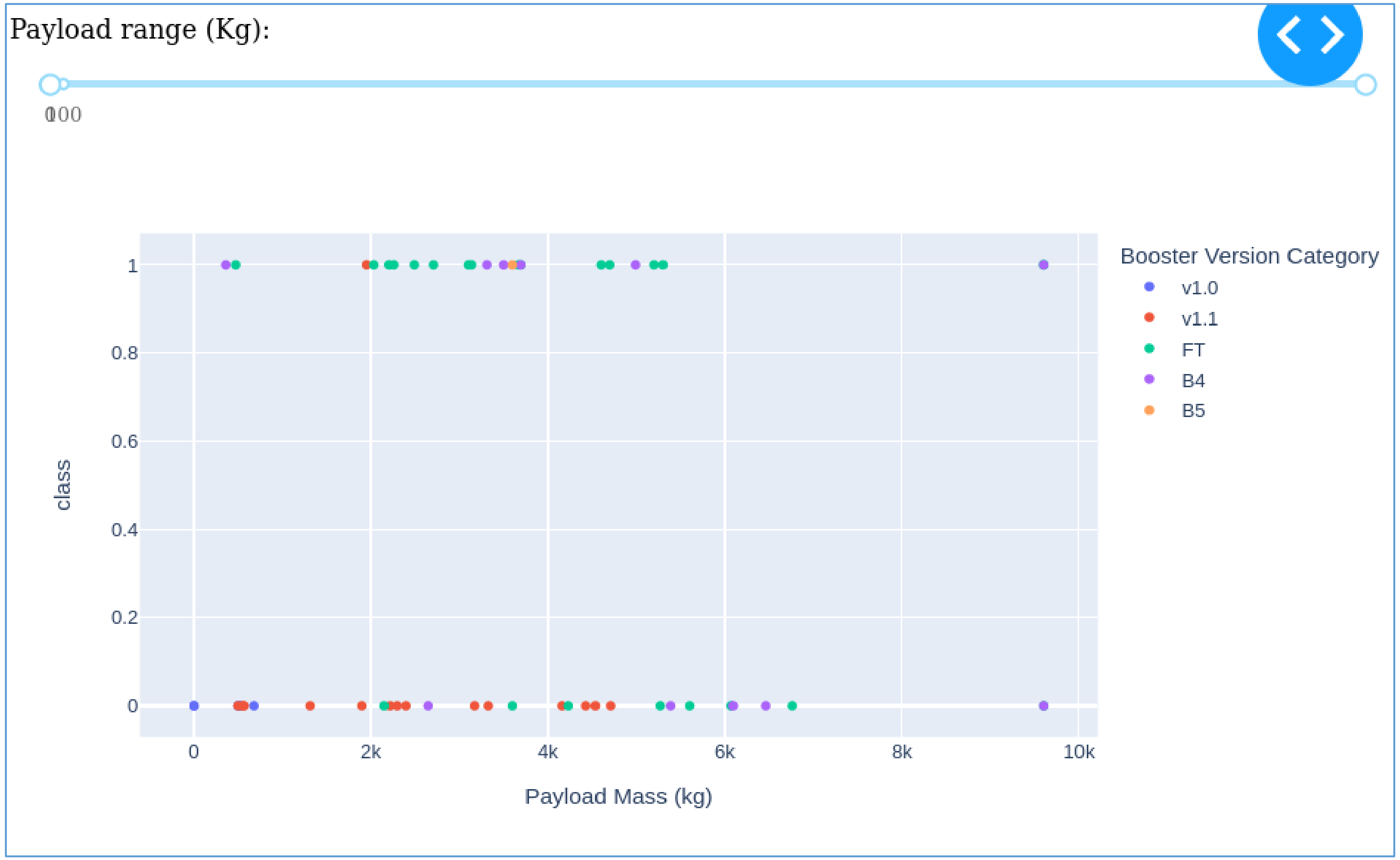
Payload vs. Launch Outcome

- Payloads under 6,000kg and FT boosters are the most successful combination.



Payload vs. Launch Outcome

- There's not enough data to estimate risk of launches over 7,000kg

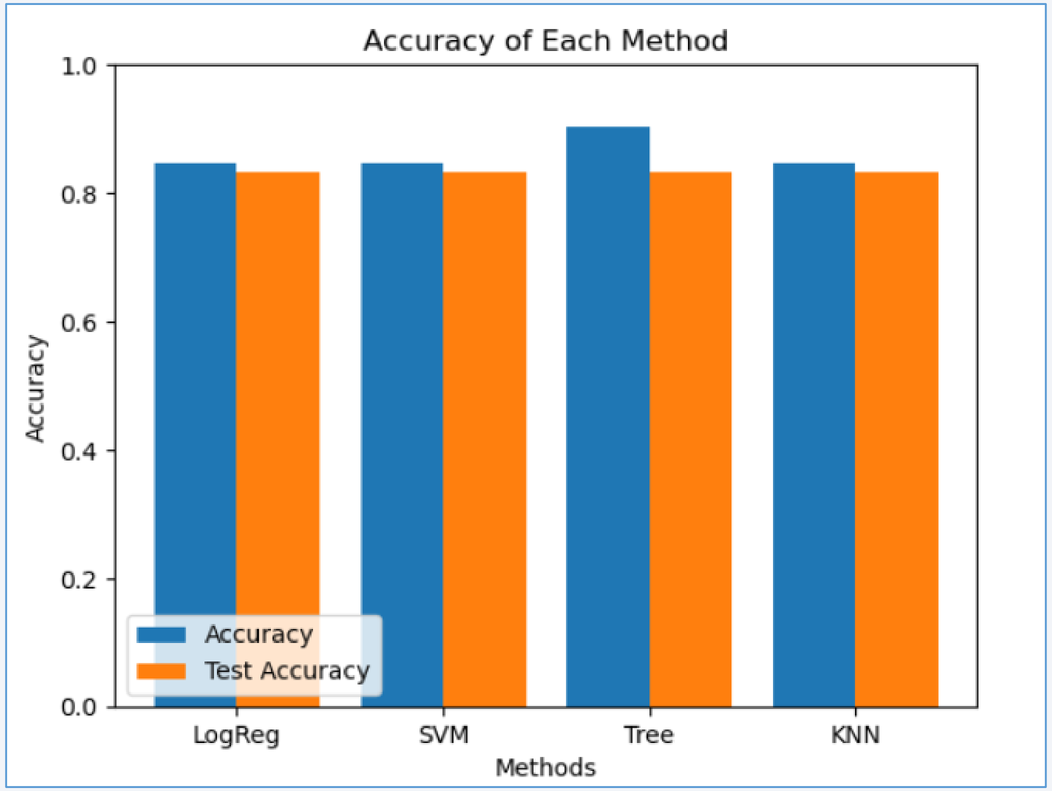


Section 5

Predictive Analysis (Classification)

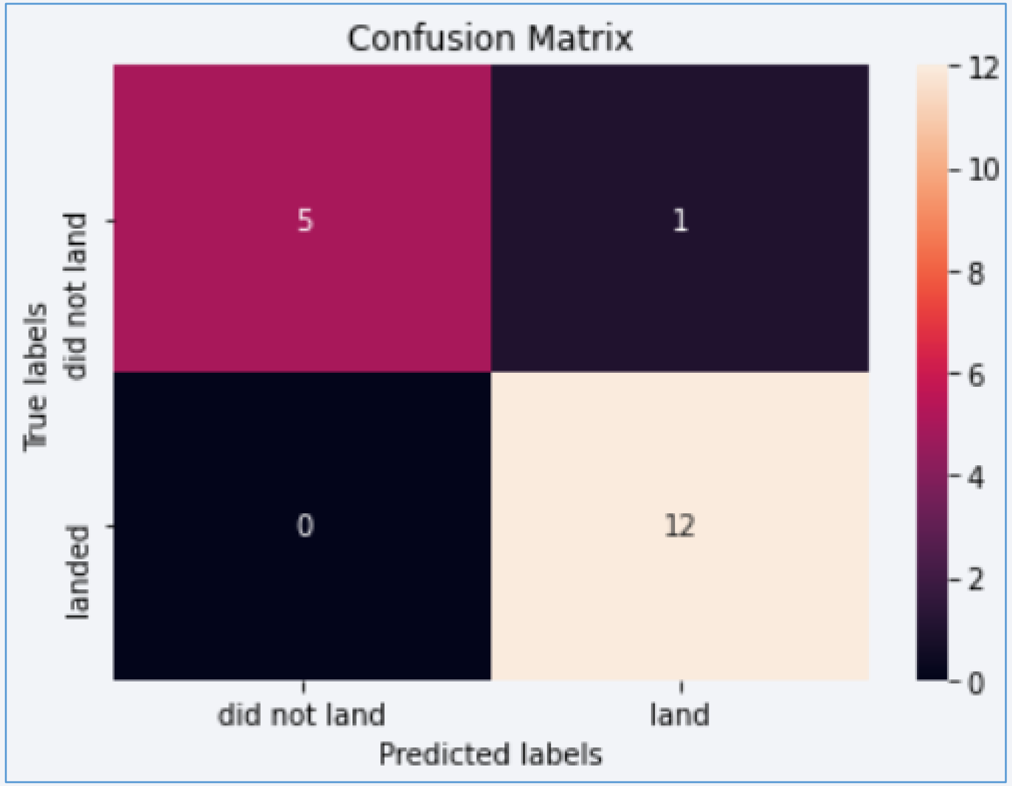
Classification Accuracy

- 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!