IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Mark Petros
20/03/2024

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 such as Folium
- Machine Learning Prediction using classification techniques

Summary of all results

- Machine Learning Prediction showed the best classification technique to predict with accuracy, the number of true successful and unsuccessful landing and in which locations.
- This was possible and highly aided by Exploratory Data Analysis, which visually showed the expected success of launches/landings

Introduction

Project background and context

This project aims to predict whether SpaceX's Falcon 9 will successfully land in the first stage. It is crucial to determine this to help the board budget for the cost of each launch. Falcon 9 is a relatively cheaper rocket compared to other industry providers, therefore, there is still a marginal opportunity to leverage costs and launches for advertising/marketing purposes.

Problems you want to find answers

We aim to resolve whether the launches are successful using past launch data and metrics such as Orbit size, Long, Latitude, geographical location, to predict future successful launches and landings and be able to quantify what each launch will cost. This will help in the development phase helping SpaceX be a leader in rocket launches using the Falcon 9.

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - Data from Space X was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- **Performed data wrangling**
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features
- **Performed exploratory data analysis (EDA) using visualization and SQL**

Methodology

Executive Summary

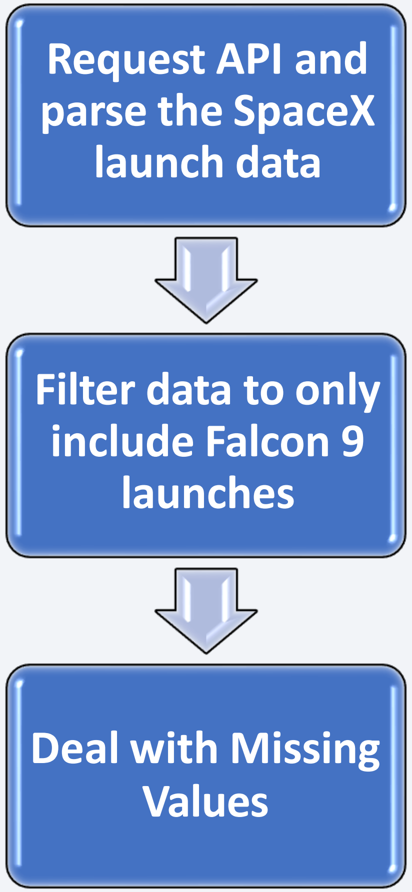
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification techniques**
  - Data that was collected and normalized, divided into training and test data sets, and evaluated by four different classification models, being the accuracy of each model evaluated using a combination of different parameters.

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 technique.

Data Collection - SpaceX API

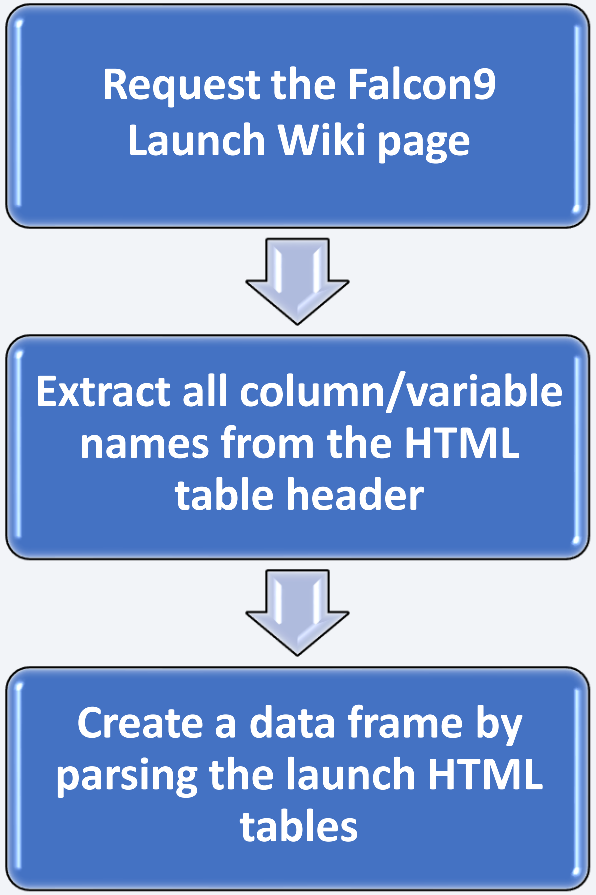
- SpaceX offers a public API where data can be obtained;



Source code:
https://github.com/Trosadventures/Applied-Data-Science-SpaceX-Project

Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- Data is downloaded from Wikipedia according to the flowchart and then parsed.



Source code:
https://github.com/Trosadventures/Applied-Data-Science-SpaceX-Project

Data Wrangling

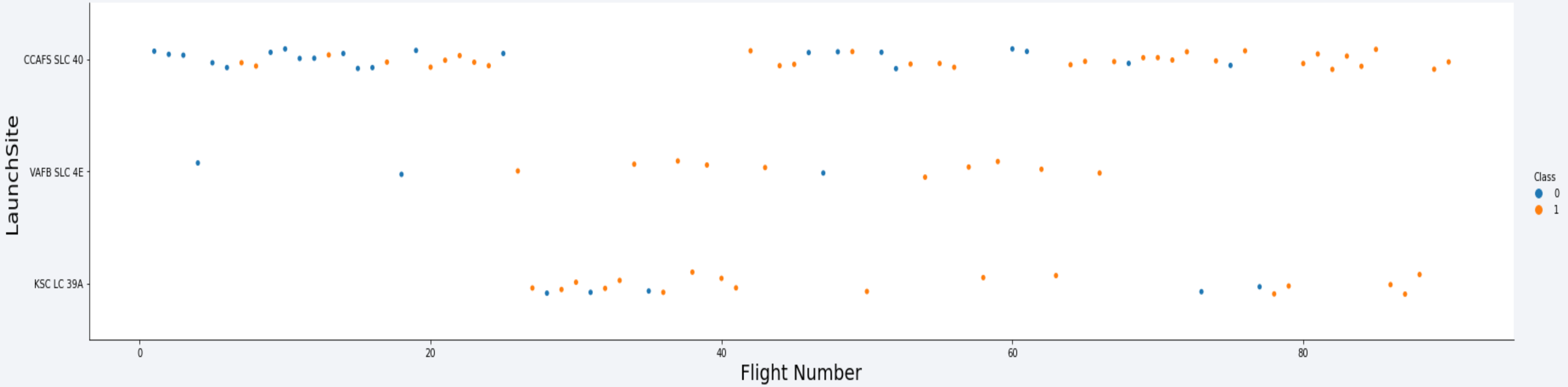
- Initially, some Exploratory Data Analysis (EDA) was performed on the dataset.
- Secondly, a summary of launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from the output.

Source code:
https://github.com/Trosadventures/Applied-Data-Science-SpaceX-Project

EDA with Data Visualization

- To explore the data, scatterplots and barplots were used to visualize the relationship between the following attributes:

Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



Source code:
https://github.com/Trosadventures/Applied-Data-Science-SpaceX-Project

EDA with SQL

- **The following SQL queries were 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; and
  - 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/Trosadventures/Applied-Data-Science-SpaceX-Project

Build an Interactive Map with Folium

- **Markers, circles, lines and marker clusters were used with Folium Maps**
  - Markers indicate points such as launch sites;
  - Circles indicate highlighted areas around specific coordinates, like the 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/Trosadventures/Applied-Data-Science-SpaceX-Project

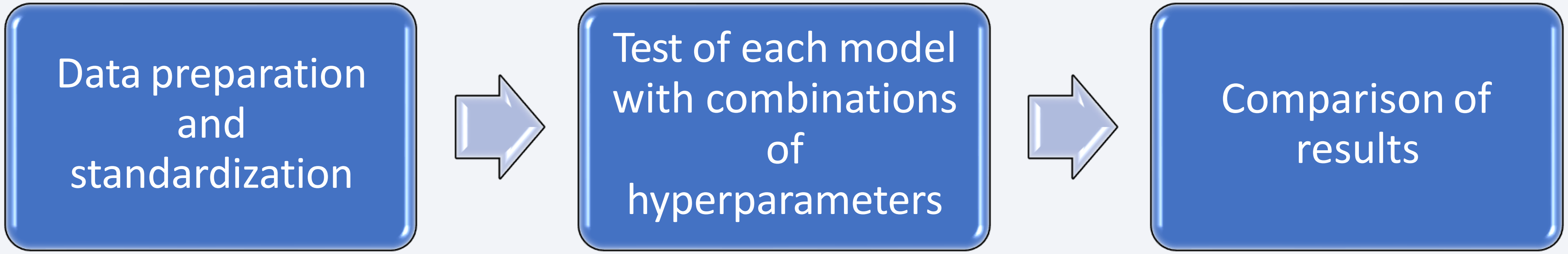
Build a Dashboard with Plotly Dash

- **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 as per the payloads.**

Source code:
https://github.com/Trosadventures/Applied-Data-Science-SpaceX-Project

Predictive Analysis (Classification)

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



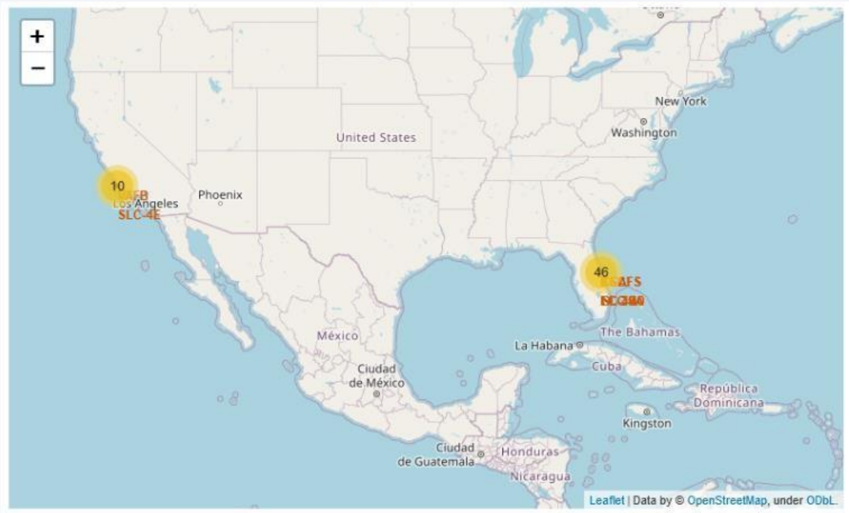
Source code:
https://github.com/Trosadventures/Applied-Data-Science-SpaceX-Project

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.

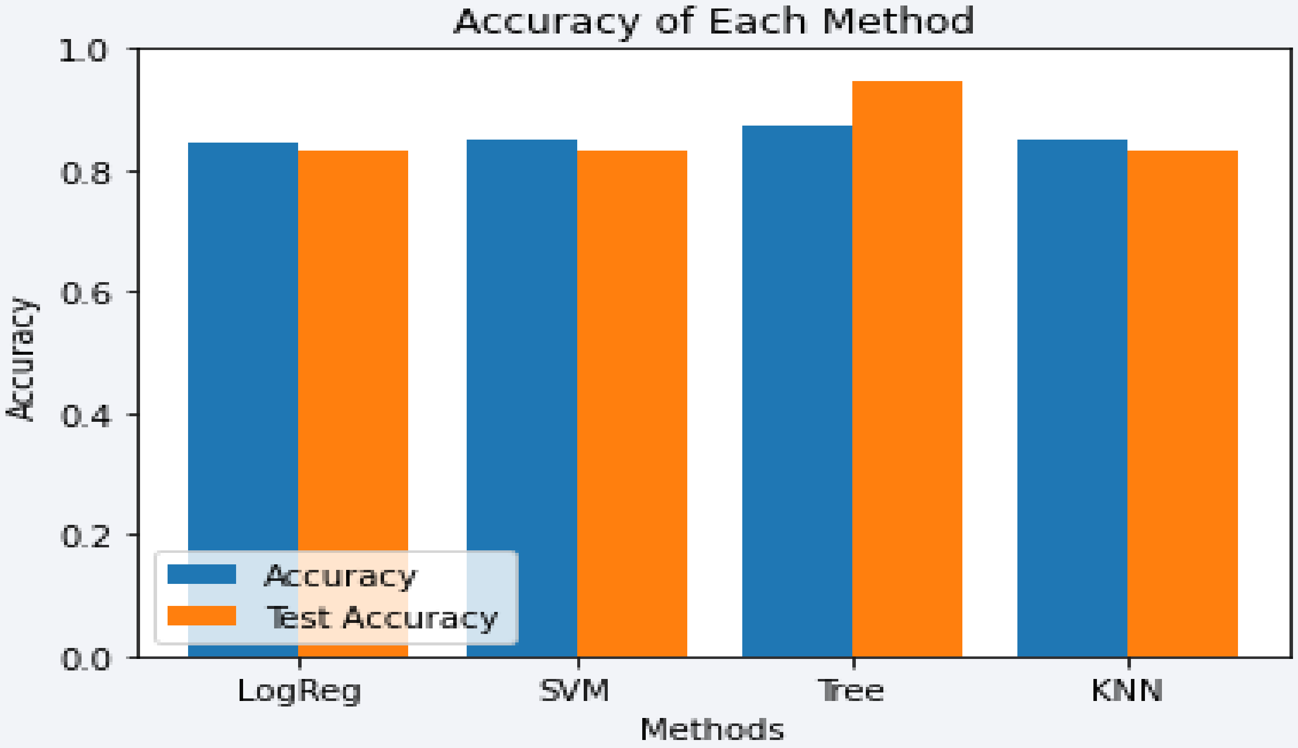
Results

- Using interactive analytics, it was clear to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around such as railways and proximity to highway.
- Most launches happens at east cost launch sites.



Results

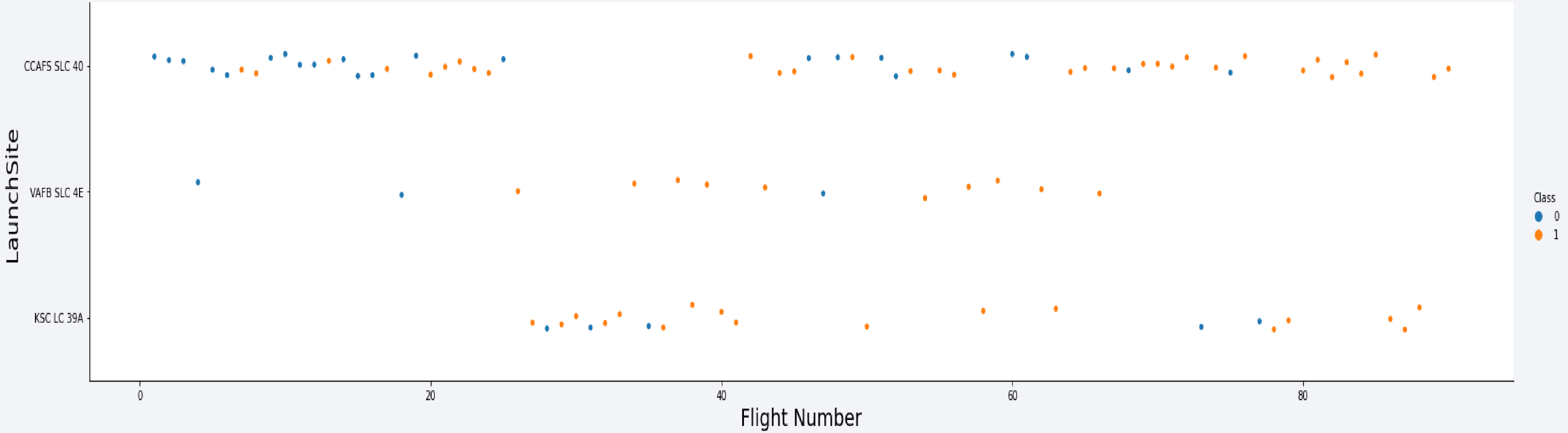
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.



Section 2

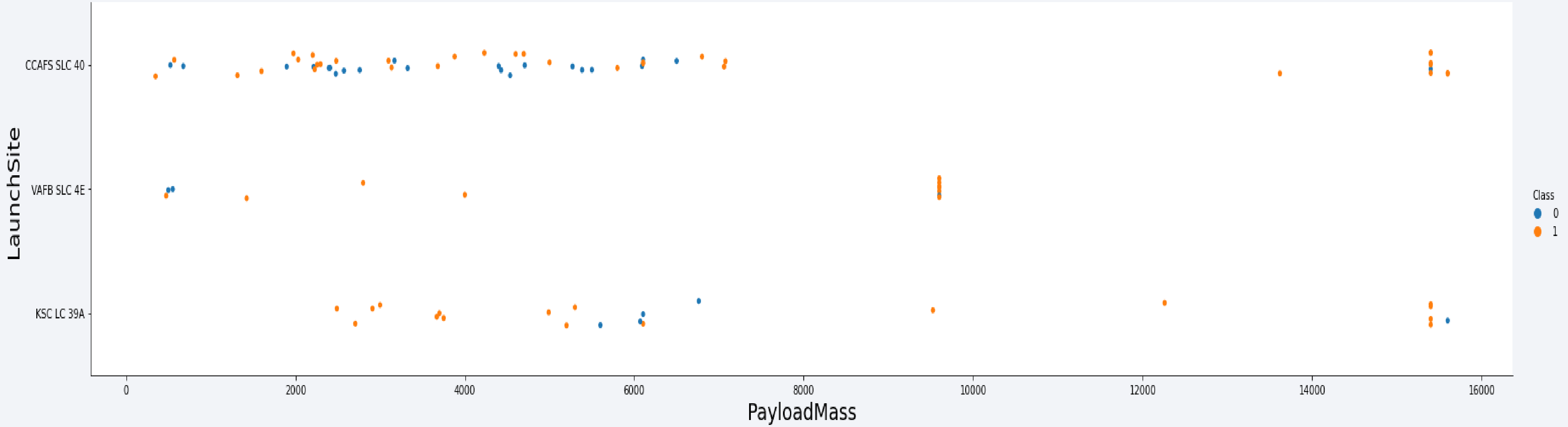
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot above, it's possible to verify that the best launch site is CCAF5 SLC 40, where most of recent launches were successful;
- Following closely are VAFB SLC 4E and KSCLC 39A;
- It's also possible to see that the general success rate improved over time.

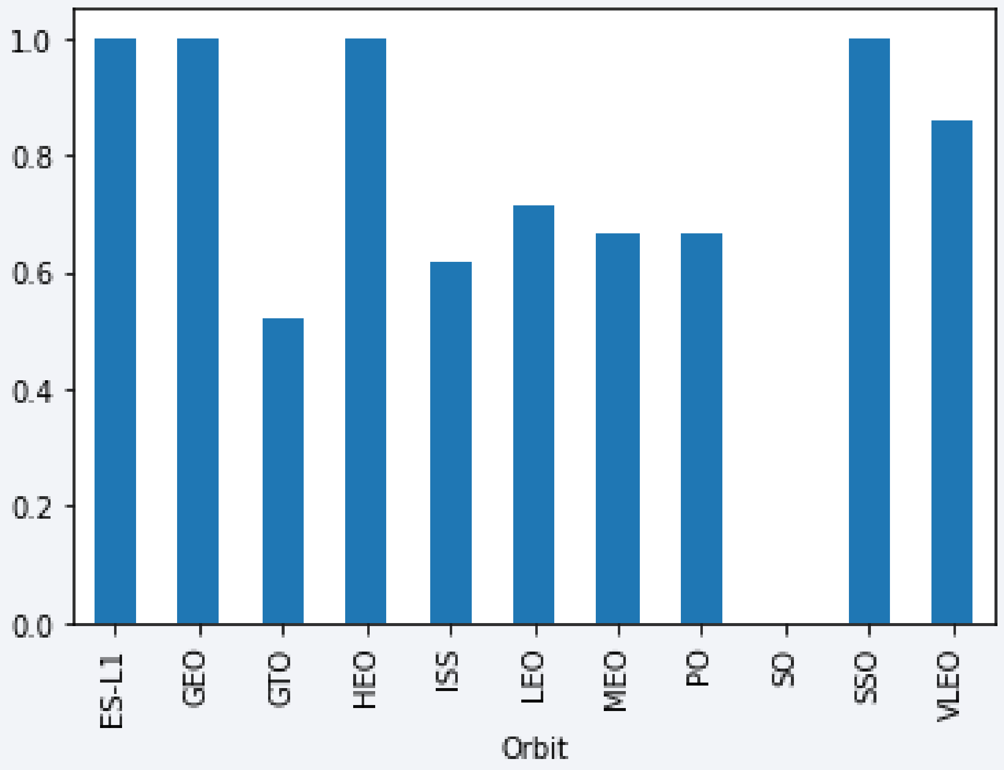
Payload vs. Launch Site



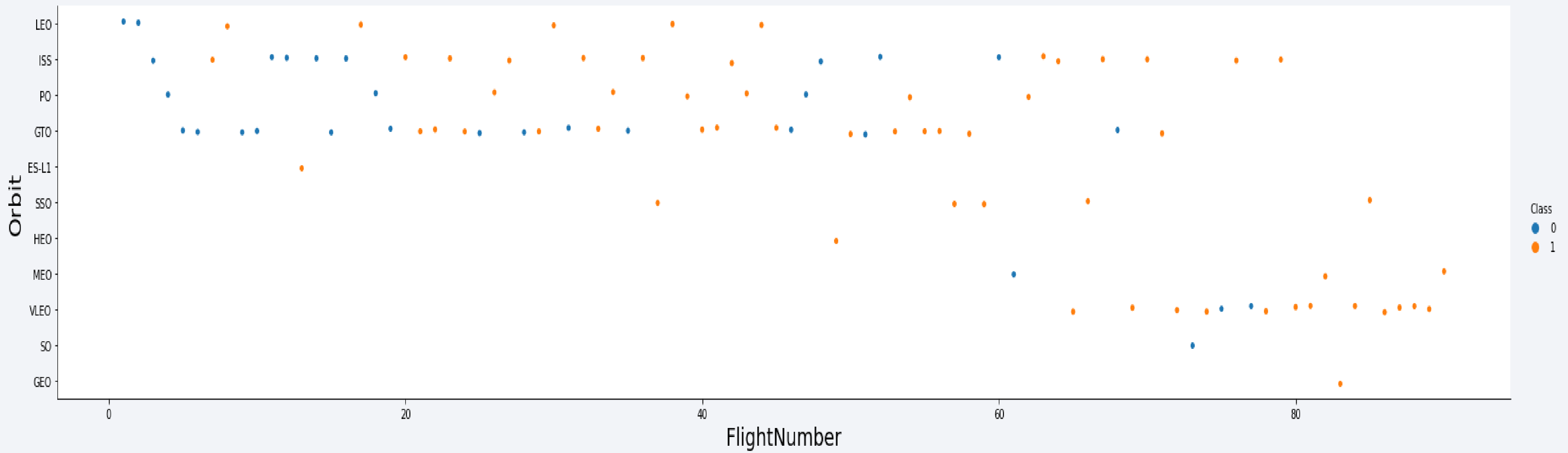
- Payloads over 9,000kg have excellent success rates;
- Payloads over 12,000kg seems to be more suited for CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

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

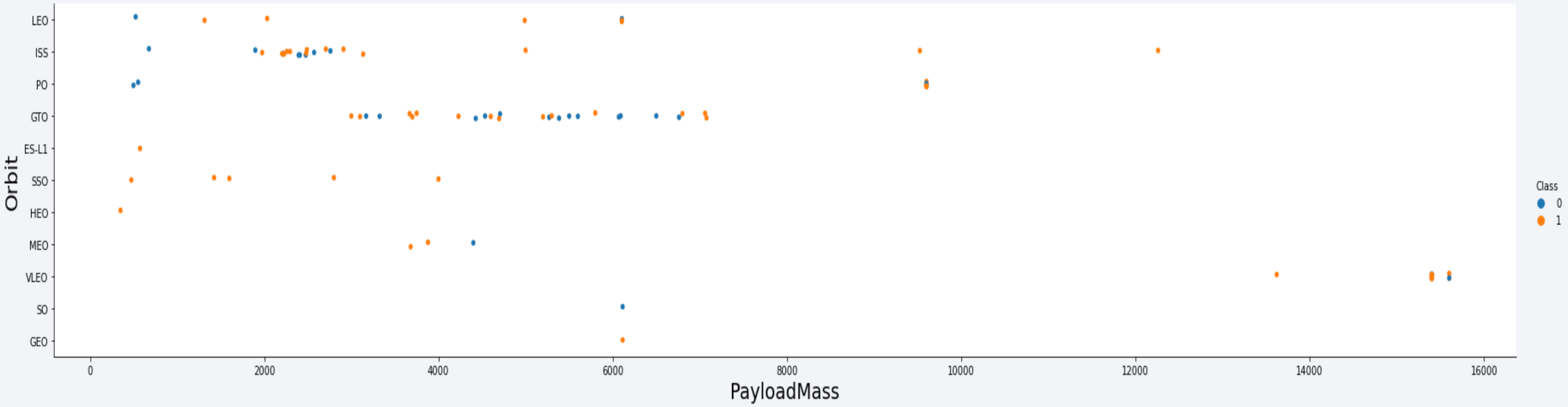


Flight Number vs. Orbit Type



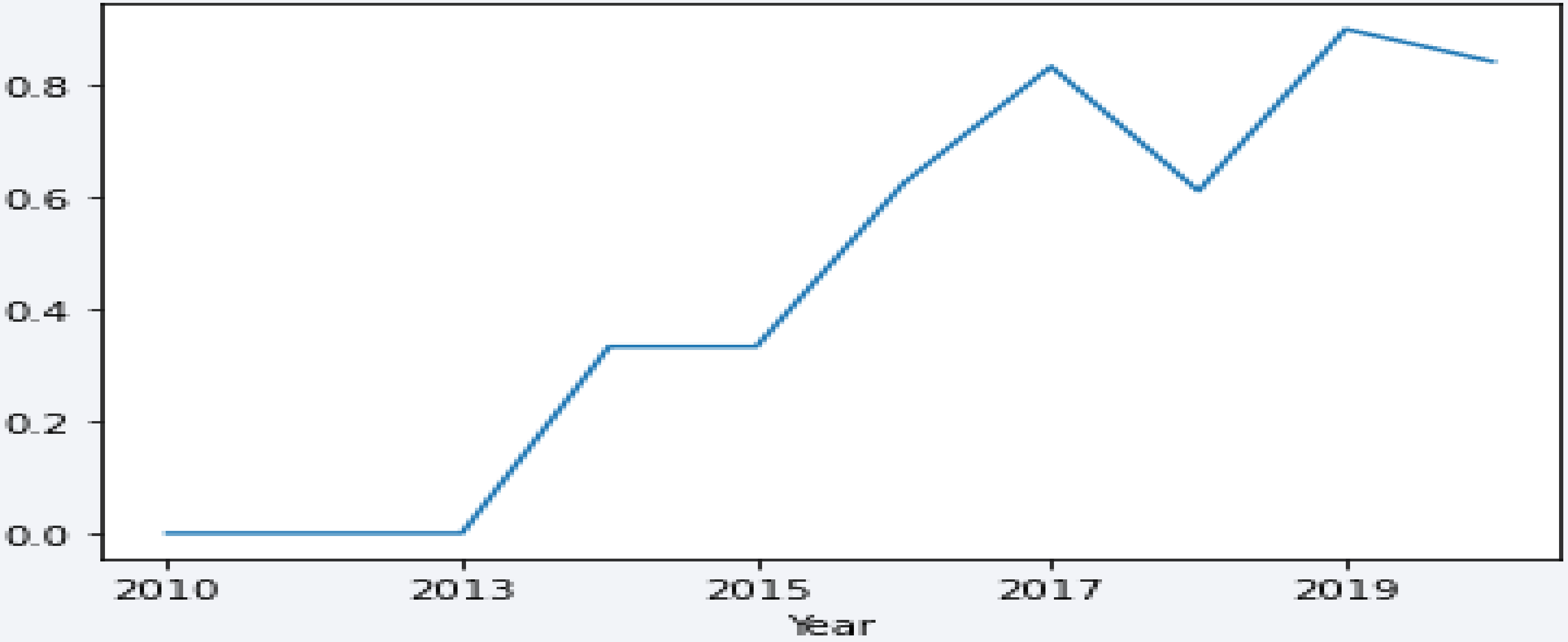
- Apparently, success rate improved over time on all orbits;
- VLEO orbit seems like a new business opportunity, due to the recent increase of its frequency.

Payload vs. Orbit Type



- Apparently, there is no correlation 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 gradually grew to 2020;
- It seems that the first three years were a period of trials, R&D 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	**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

- Here we can see five samples of Cape Canaveral launches.

Total Payload Mass

- Total payload carried by boosters from NASA:

Total Payload (kg)
111.268

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

Average Payload Mass by F9 v1.1

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

Avg Payload (kg)
2.928

- 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

- First successful landing outcome on ground pad:

Min Date
2015-12-22

- 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.

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

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

Total Number of Successful and Failure Mission Outcomes

- Number of successful and failure mission outcomes:

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

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

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass

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

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

- 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

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

- The list above has the only two occurrences.

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:

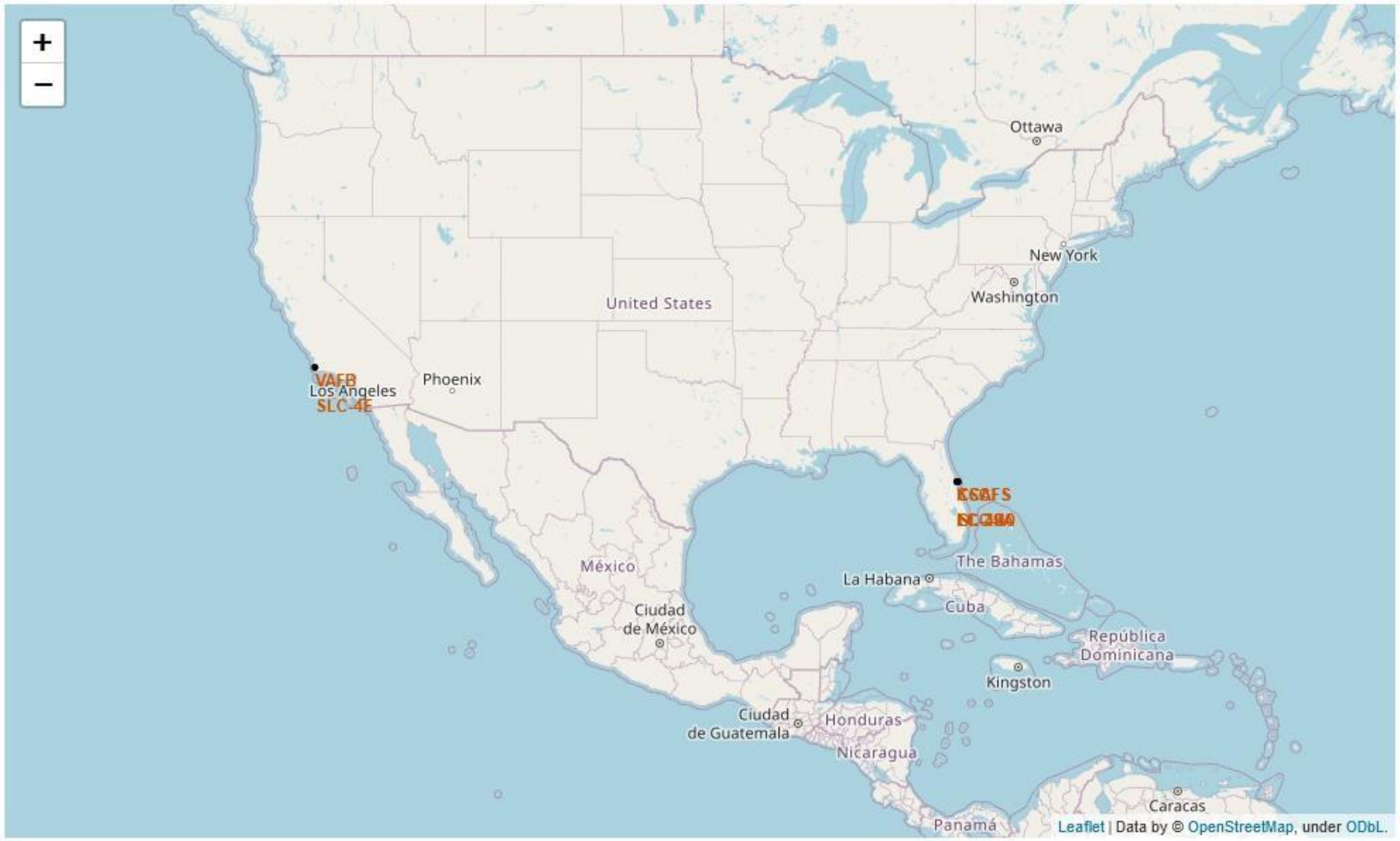
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This view of data alerts us that “No attempt” must be taken in account.

Section 4

Launch Sites Proximities Analysis

All launch sites



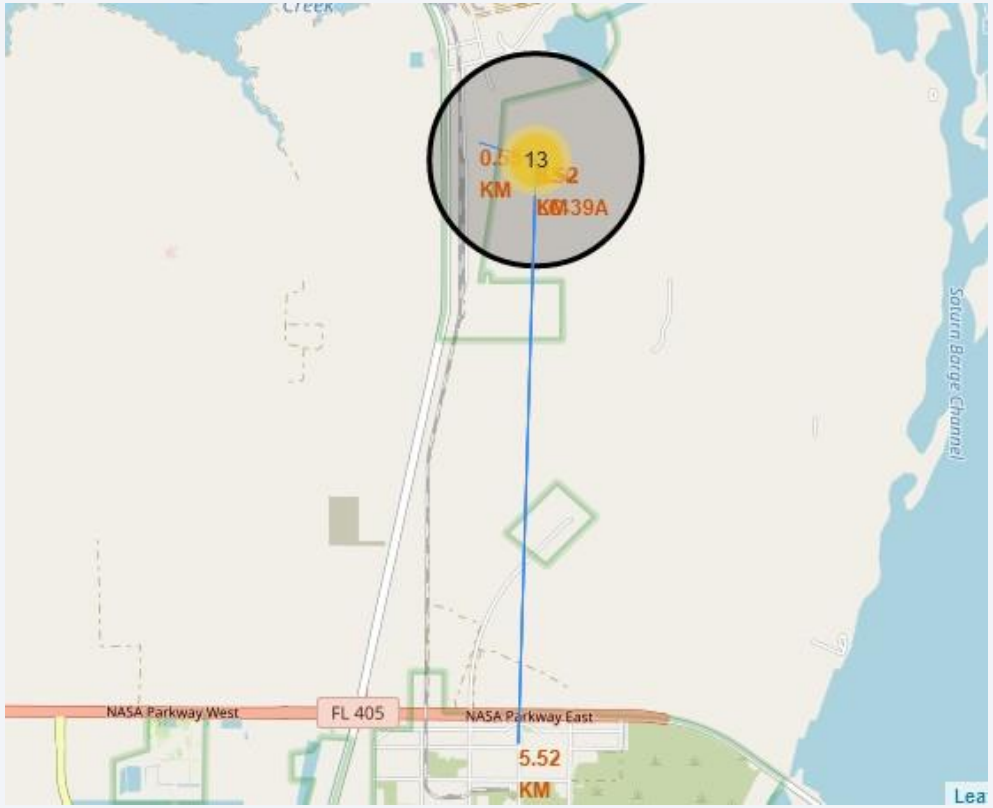
- Launch sites are near sea, probably for safety, but not too far from highways and railroads.

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

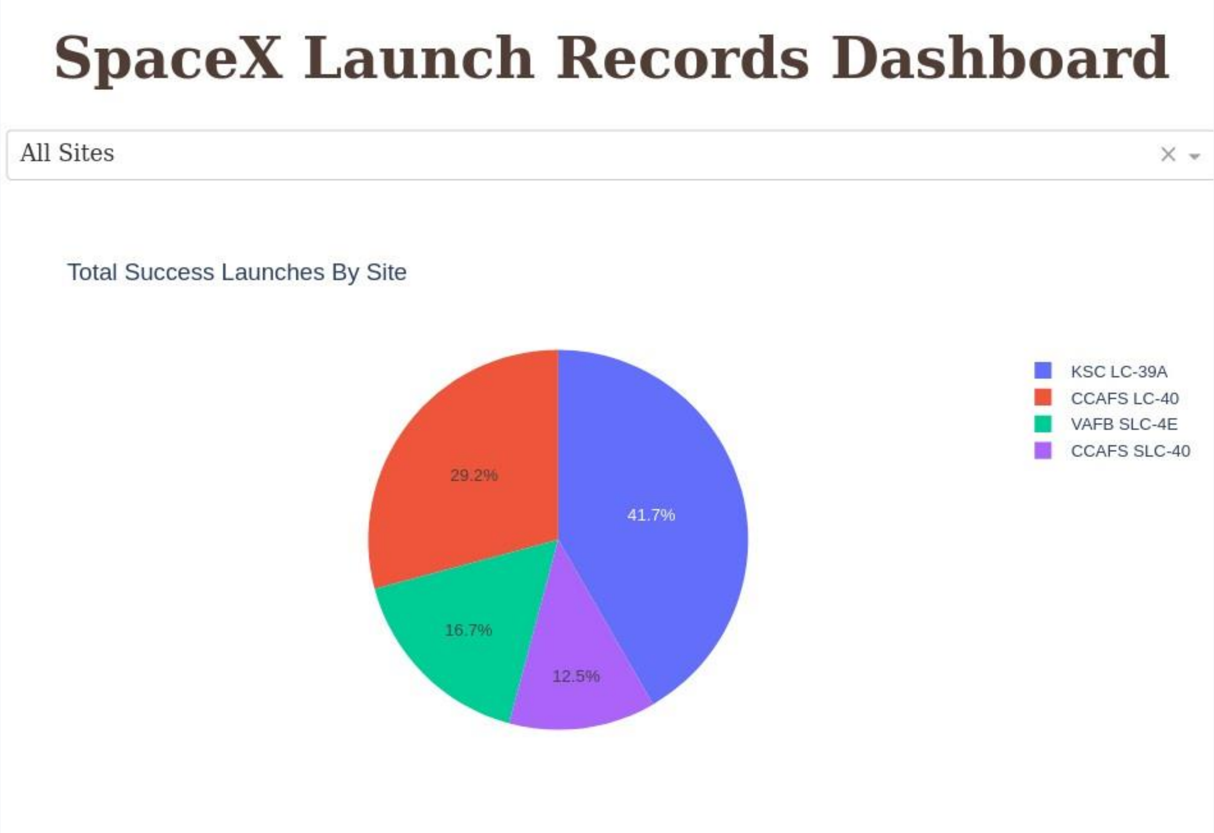


- Launch site KSCLC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 5

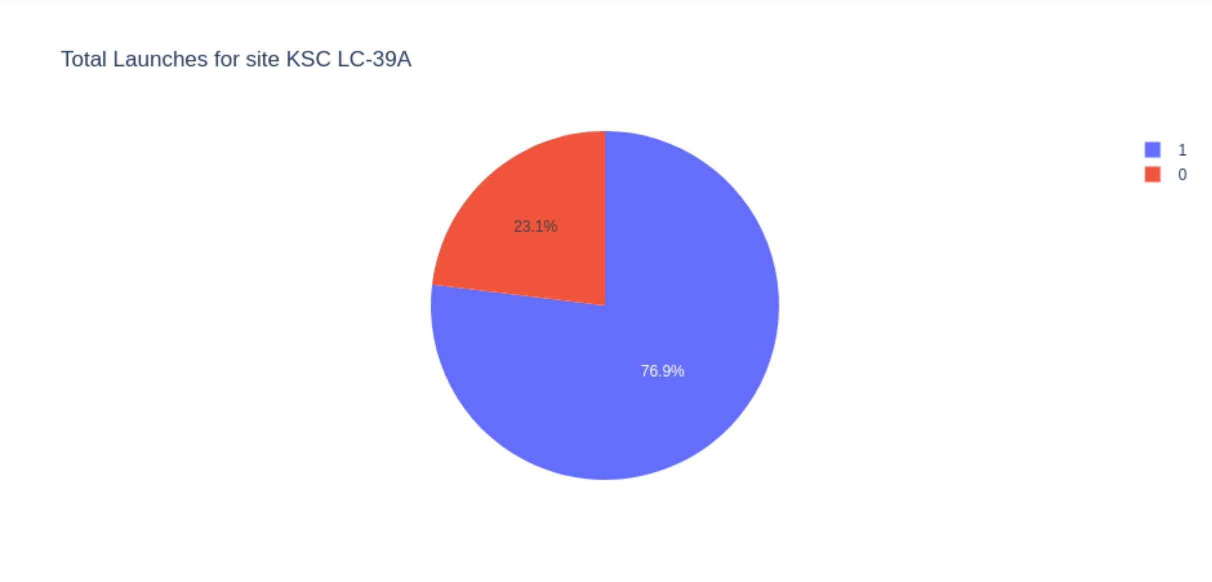
Build a Dashboard with Plotly Dash

Successful Launches by Site



- The place where launches are done seems to be a very important factor for Successful 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

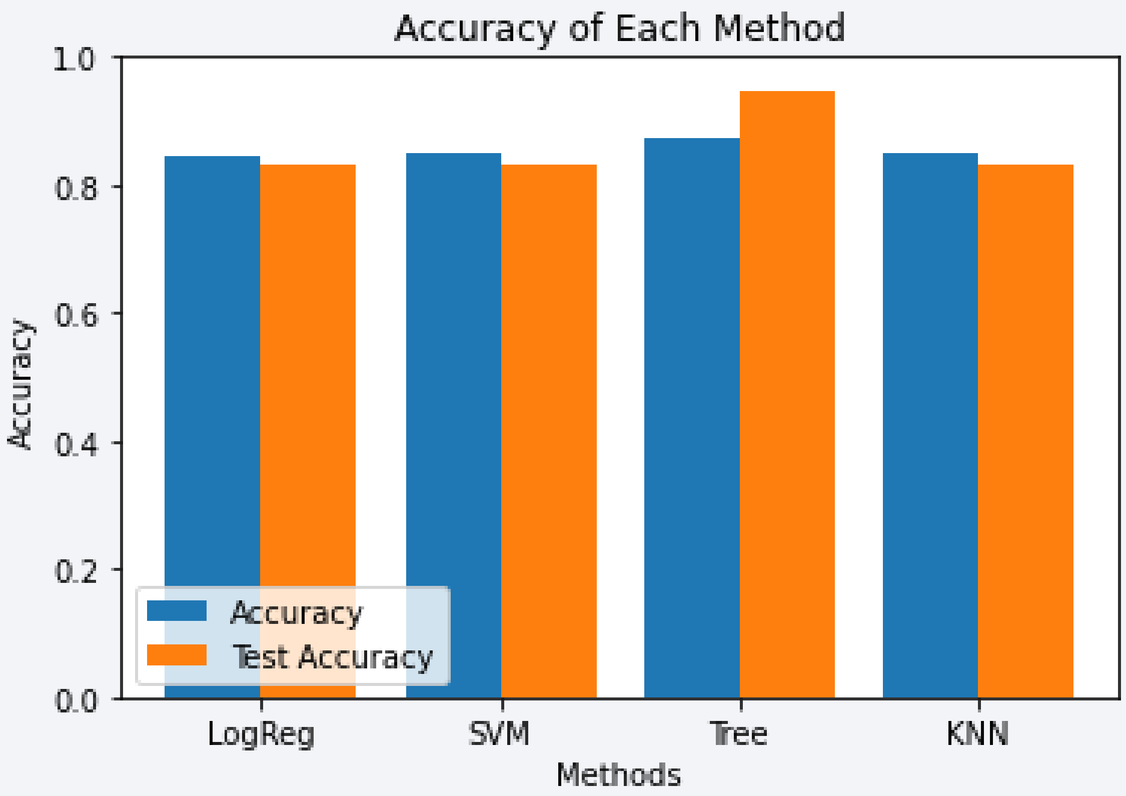
Section 6

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 of Decision Tree Classifier

- 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 and improvement 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!