IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Douglas Grings
November 12th 2022

Outline

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

Executive Summary

- The following methodologies were used to analyze data:
  - 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 Space Y to compete with Space X.
- Desirable answers:
  - 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 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)
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing 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
    - 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.

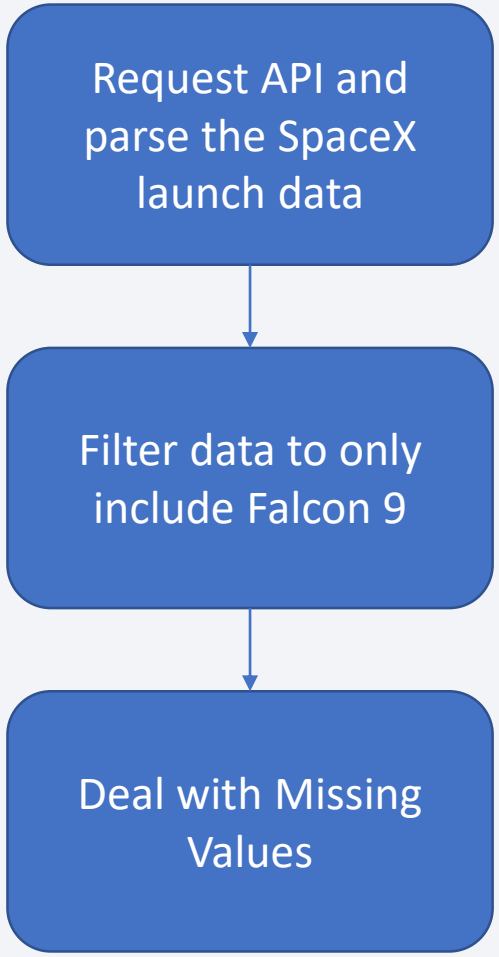
Data Collection

- Data sets were collected from 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

- 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.

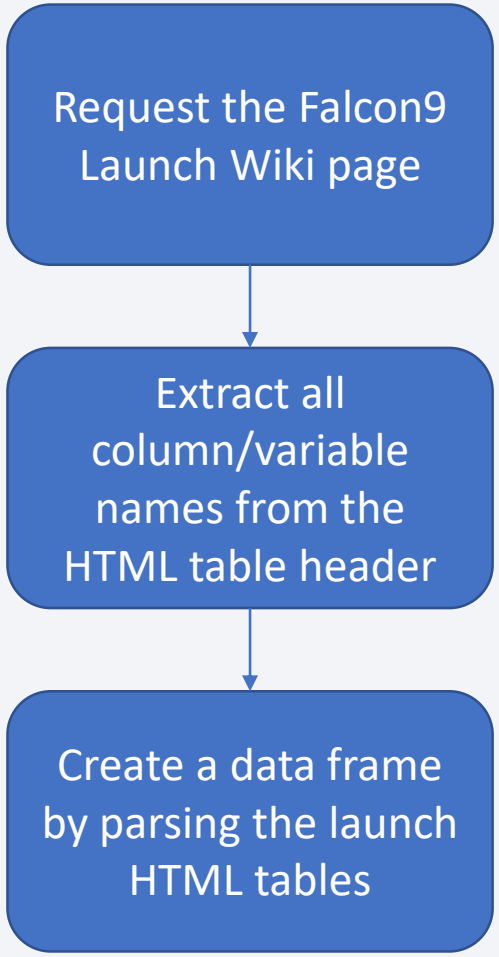
- GitHub URL:
- https://github.com/MscDouglas/aiup/blob/master/M%2010%20S1:%20Collectin%20the%20data.ipynb



Data Collection - Scraping

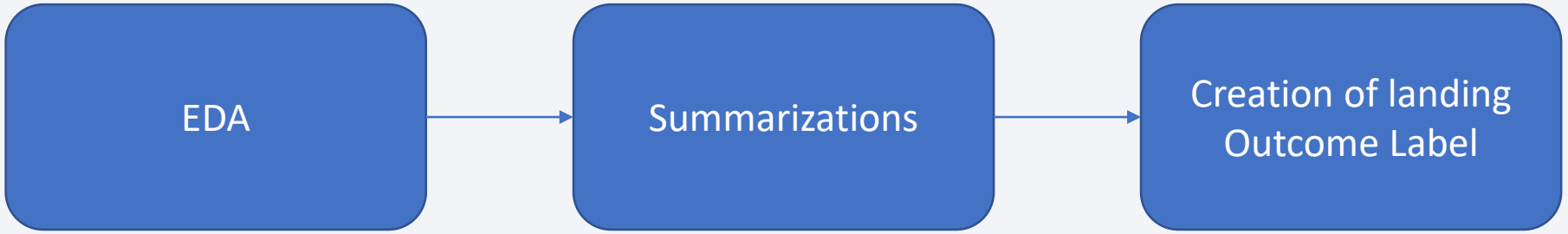
- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.
- GitHub URL:

https://github.com/MscDouglas/aiup/blob/master/Modulo%2010%20-%20Sem%201:%20Web%20scraping%20Falcon%209%20and%20Falcon%20Heavy%20Launch.ipynb



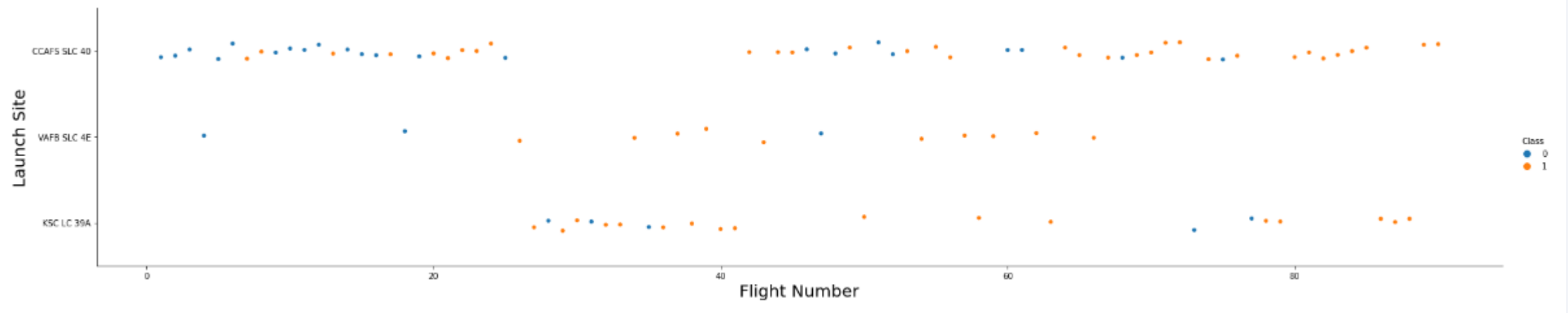
Data Wrangling

- 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.
- GitHub: https://github.com/MscDouglas/aiup/blob/master/M10%20S1.1:%20First%20Stage%20Landing%20Prediction.ipynb



EDA with Data Visualization

- To explore data, scatterplots and barplotswere 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



- GitHub: https://github.com/MscDouglas/aiup/blob/master/M10%20-%20S2.2:%20Stage%20Landing%20Prediction.ipynb

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.
- GitHub URL: https://github.com/MscDouglas/aiup/blob/master/M10%20-%20S2:%20SQL%20NOtebook%20for%20Peer%20Assignment.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.
- GitHub URL: https://github.com/MscDouglas/aiup/blob/master/M10%20-%20S3.1:%20Launch%20Sites%20Locations%20Analysis%20with%20Folium.ipynb

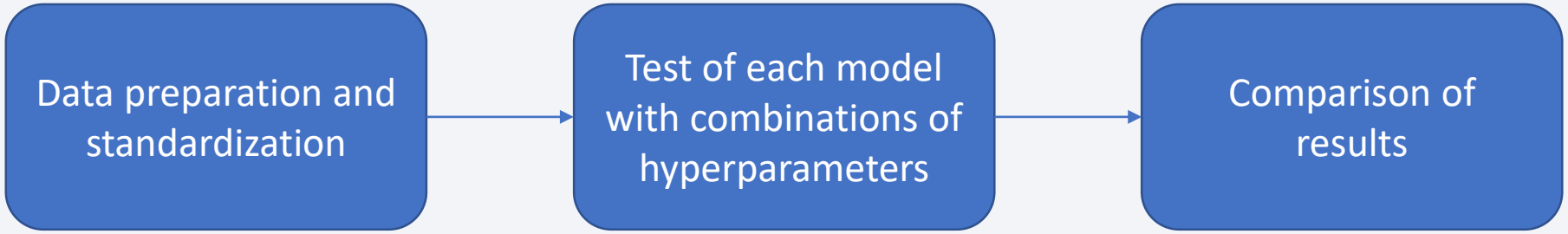
Build a Dashboard with Plotly Dash

- 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.

- GitHub URL: https://github.com/MscDouglas/aiup/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

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



- GitHub URL: https://github.com/MscDouglas/aiup/blob/master/M10%20-%20S4.1:%20Fisrt%20stage%20landing%20prediction.ipynb

Results

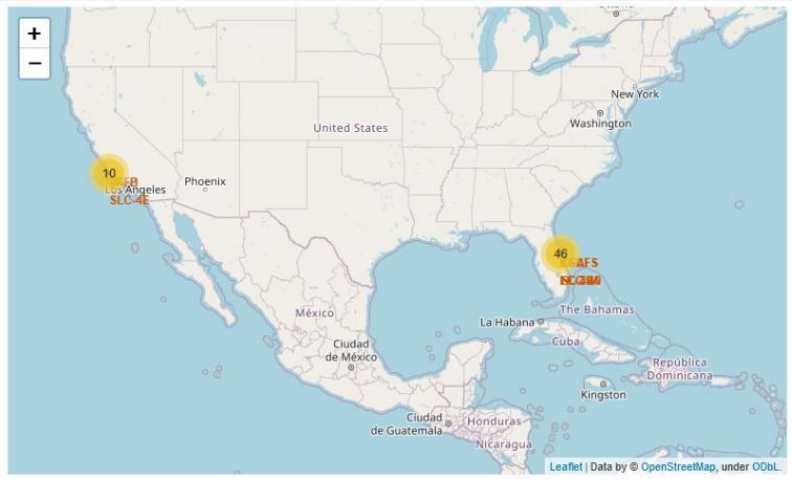
- Exploratory data analysis results:
  - Space X uses 4 different launch sites;
  - The first launches were done to Space X itself and NASA;
  - The average payload of F9 v1.1 booster is 2,928 kg;
  - The first success landing outcome happened in 2015 fiver year after the first launch;
  - Many Falcon 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.

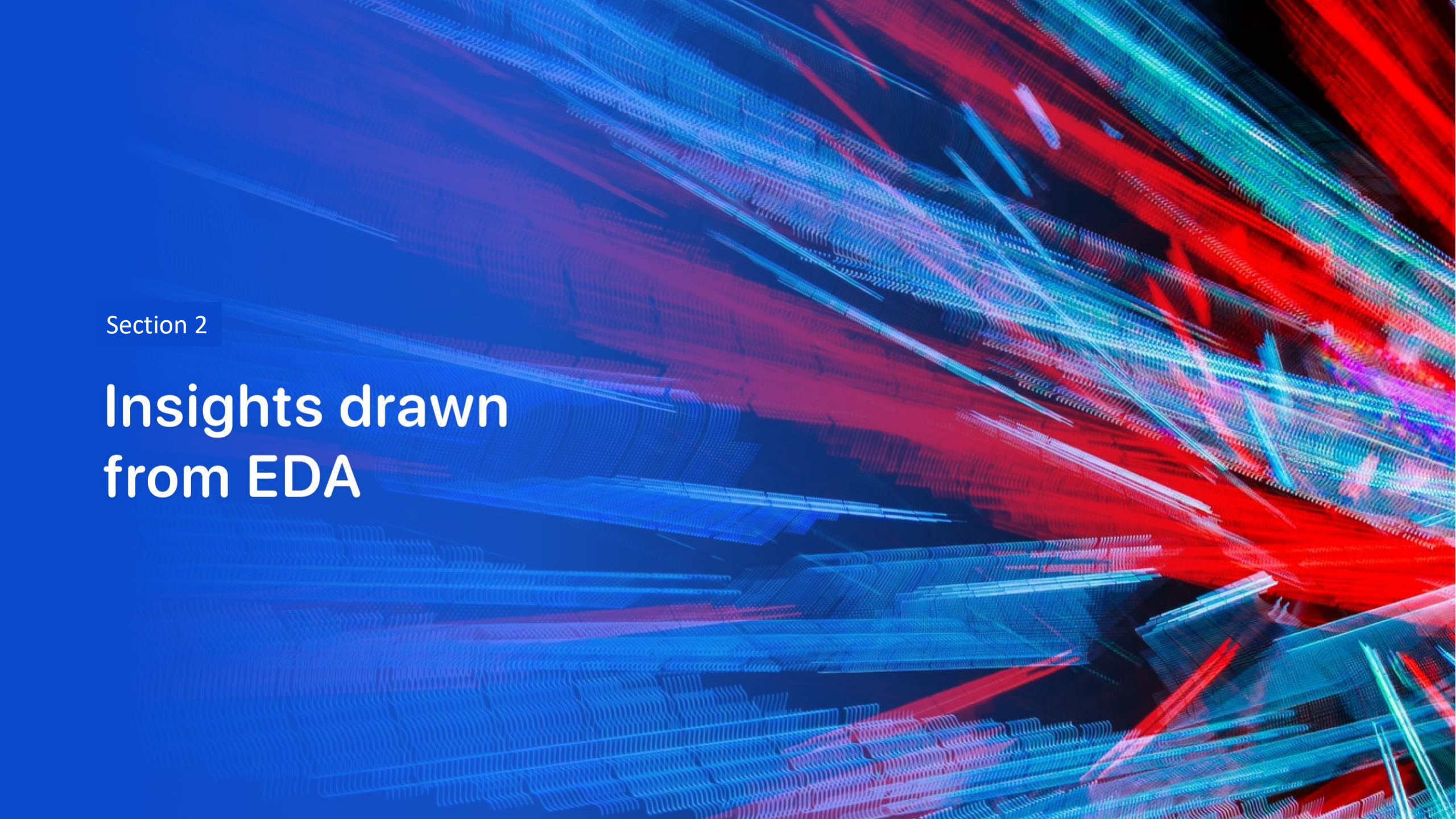
Results

- 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.

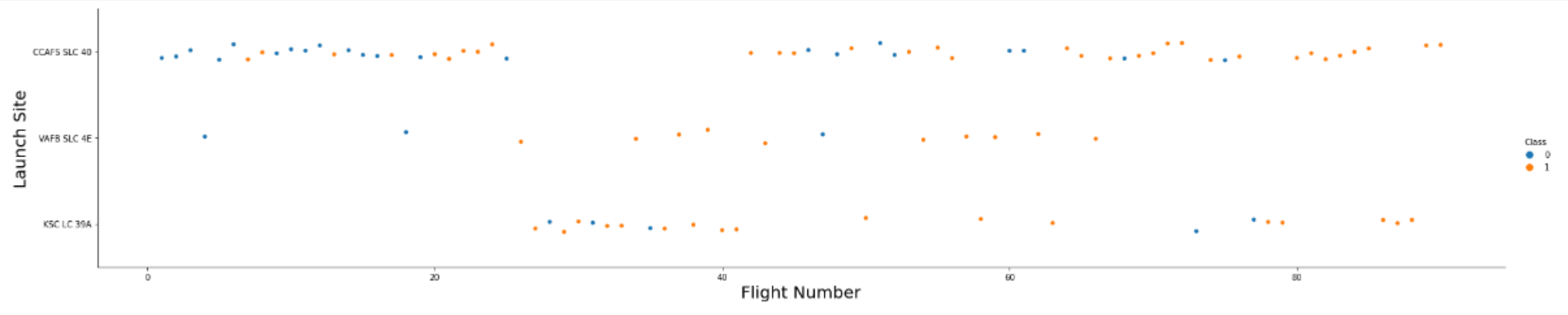
Results

- 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.



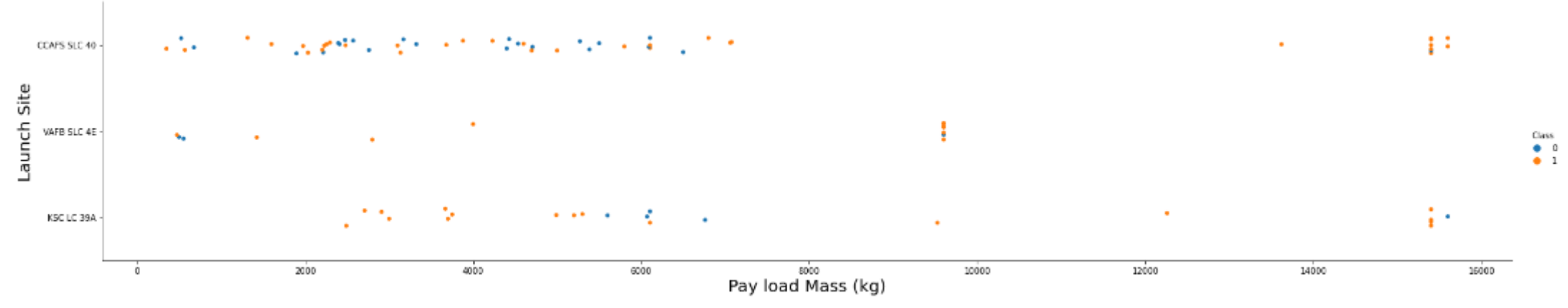

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



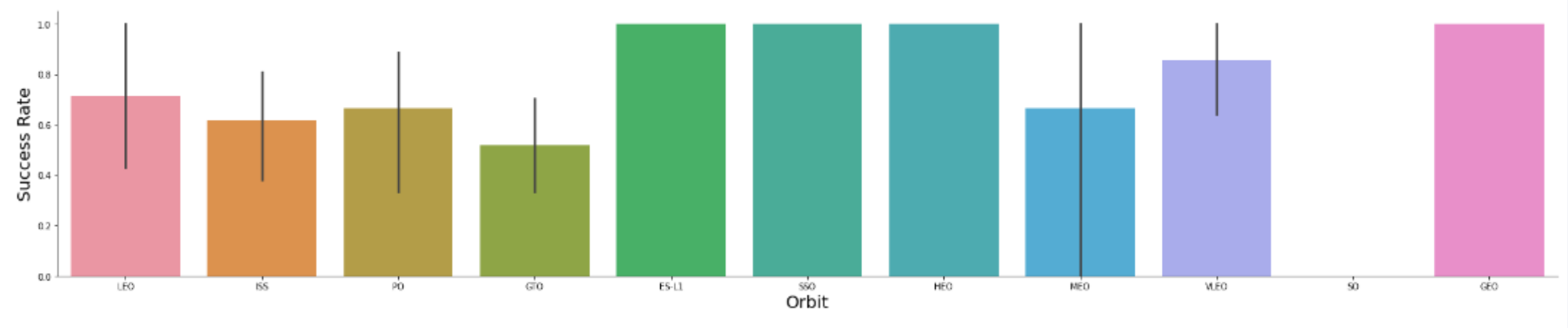
- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It's also possible to see that the general success rate improved over time.

Payload vs. Launch Site



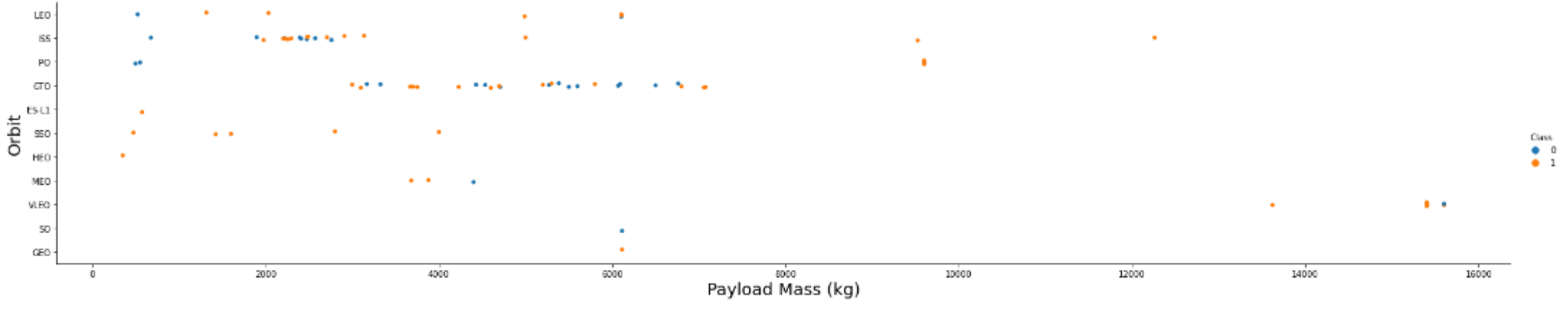
- Payloads over 8000 kg have a very high success rate, almost all between 9000 and 10000 kg launched from VAFB SLC 4E, the others are all newer launches;
- Smaller payloads have a lower success rate, generally because the earlier launches were all with these smaller payloads.

Success Rate vs. Orbit Type



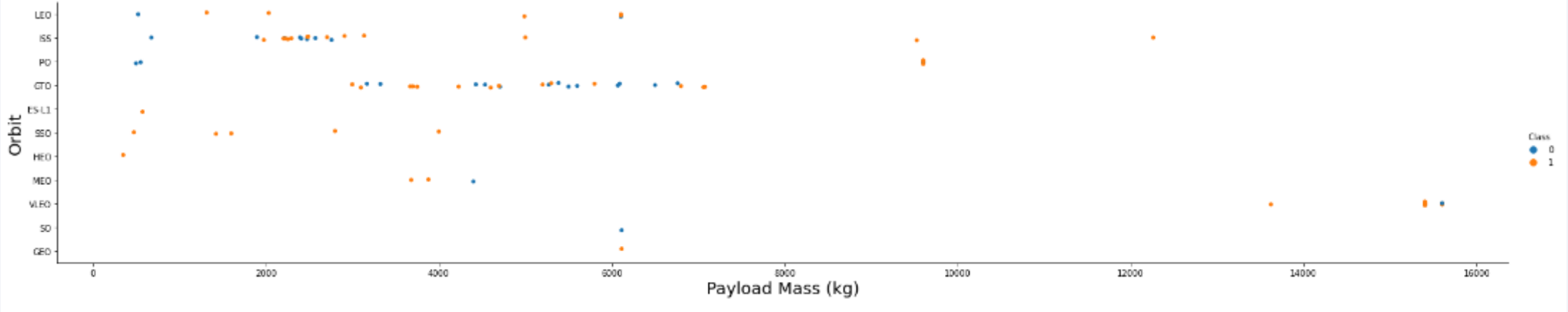
- All of the orbits have more than 50% success rate;
- The most successful orbits, with 100% success rate are: ES-L1, SSO, HEO and GEO;
- The only other orbit with over 80% success rate is VLEO.

Flight Number vs. Orbit Type



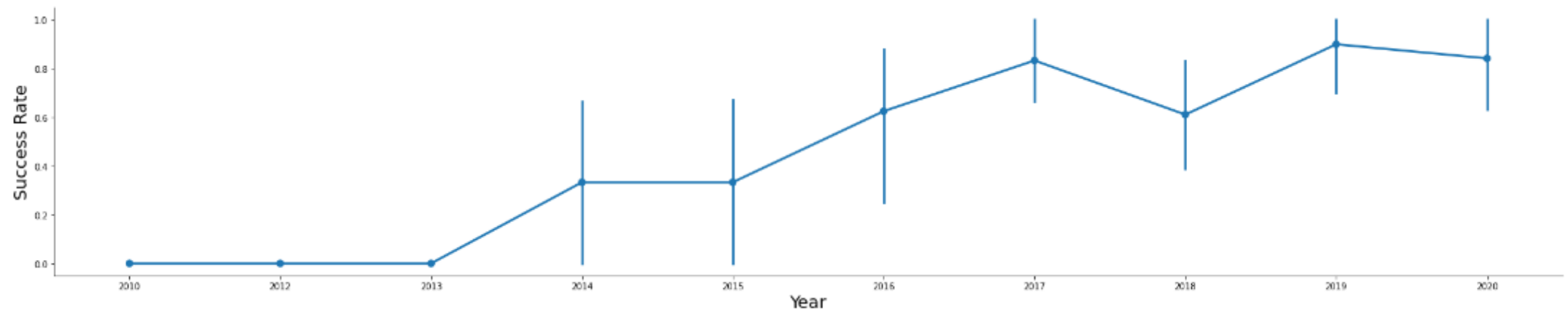
- LEO, ISS, PO andGTO wereusedthemostin earlylaunches, butare notbeingusedas frequentlyanymore;
- Fromaroundflight60 andonwarda preferencetotheVLEO orbitseemstobeforming.

Payload vs. Orbit Type



- Most orbits are pretty versatile when it comes to payload mass launches, having a pretty big range of launched masses;
- The very large payloads seem to prefer the VLEO orbit, but it might be just that there are so few of those launches still.

Launch Success Yearly Trend



- The success rate is trending up, but it looks to be slowing down since 2017;
- A dip in success rate in 2018 might be due to new orbits or technology being tested;
- From 2010 to 2013 are likely the experimental years where the concepts were being tested and validated.

All Launch Site Names

- The data we had showed four different launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC039A
  - VAFB SLC-4E

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- These were the distinct values on the “launch_sites” column of our database;

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	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

- We searched for 5 records from the “CCA” launch sites;
- We can see that all of them use the same orbit;
- All of them were successful missions, but none landed successfully;

Total Payload Mass

- By aggregating the payload masses from the customer “NASA (CRS)” we found that the boosters launched by them totaled 45596 kg;

- There are other “NASA” customers, but this analysis was made exclusively for the “(CRS)” variant.

Average Payload Mass by F9 v1.1

- By averaging the payload masses from boosters “F9 v1.1” we found that these boosters launched an average of 2928 kg;

- The same could be queried for different booster versions.

First Successful Ground Landing Date

- By querying the minimal date for record with value “Success (ground pad)” on the “landing__outcome” column, we found that the first successful ground pad landing was in December 22nd2015;

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- By creating a filter on the landing outcome andon the payload mass columns, we were able to retrieve which booster versions successfully landed on Drone Ship with payload between 4000 and 6000 kg:
  - F9 FT B1022;
  - F9 FT B1026;
  - F9 FT V1021.2;
  - F9 FT V1031.2.

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

Total Number of Successful and Failure Mission Outcomes

- By grouping the records by mission outcome, we can extract the total number of records for each different outcome:
  - Success: 99;
  - Success (payload status unclear): 1;
  - Failure (in flight): 1.

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- We can see that even if the landings are not always successful, the missions are over 99% successful.

Boosters Carried Maximum Payload

- By querying the maximum payload and subsequently the missions carrying this payload size, we can list all the booster version that have carried maximum payload:
  - 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

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

2015 Launch Records

- During the year 2015 only two missions had failure landing in drone ships, both were launched from the CCAFS LC-40 launch site, with the following booster versions:
  - F9 v1.1 B1012
  - F9 v1.1 B1015

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- Grouping by landing outcome, and filtering by date we can rank the most common landing outcomes within the period of June 6th2010 to March 3rd 2017:

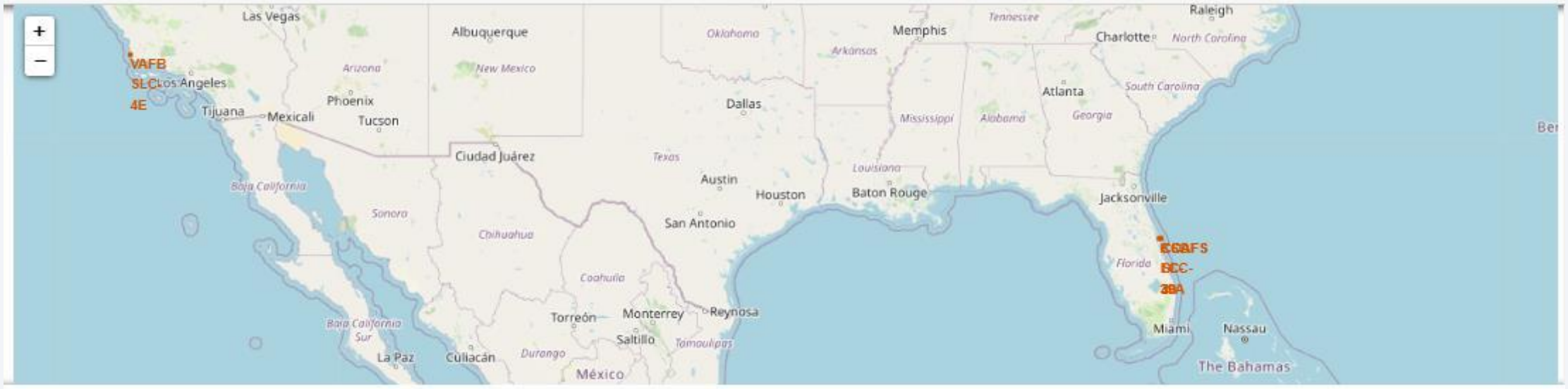
  - No attempt: 10 missions;
  - Failure (drone ship): 5;
  - Success (drone ship): 5;
  - Controlled (ocean): 3;
  - Success (ground pad): 3;
  - Uncontrolled (ocean): 2;
  - Failure (parachute): 1;
  - Precluded (drone ship): 1;

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

Section 3

Launch Sites Proximities Analysis

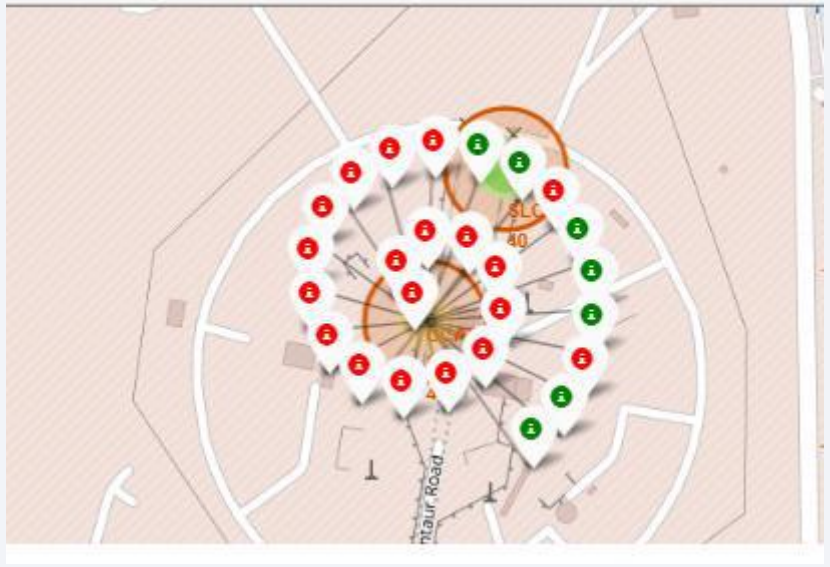
Location of Launch Sites



- Other than VAFBSLC-4E, which is located in California, all the other launch sites are located in Florida;
- In both cases they are located near the shore.

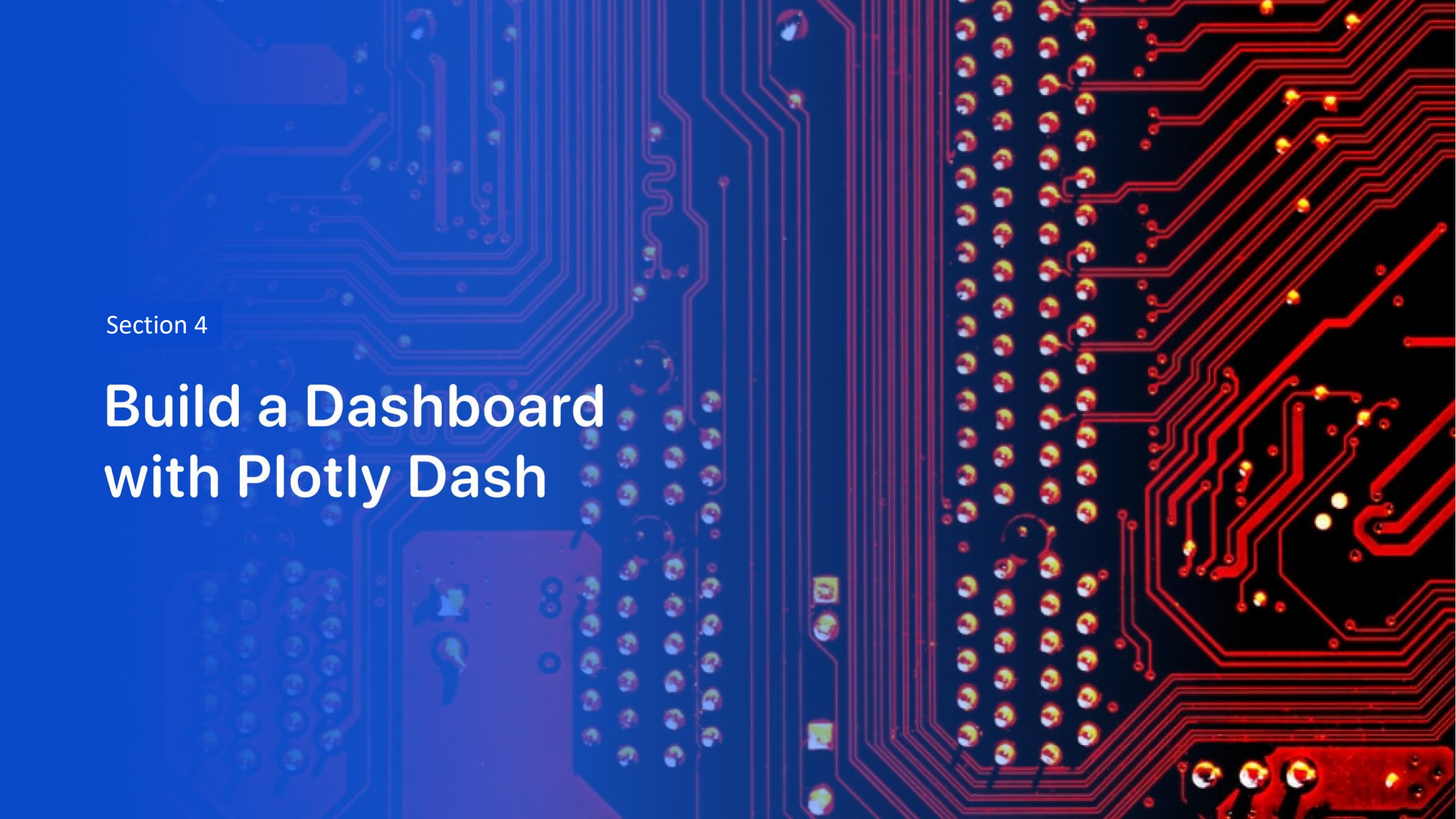
Landing Outcomes Marks

- On each launch site we can now see each launch marked, as well as the classification of success (green) and failure (red) for the landing outcome;
- The picture indicating the missions for the CCAFS LC-40 site.



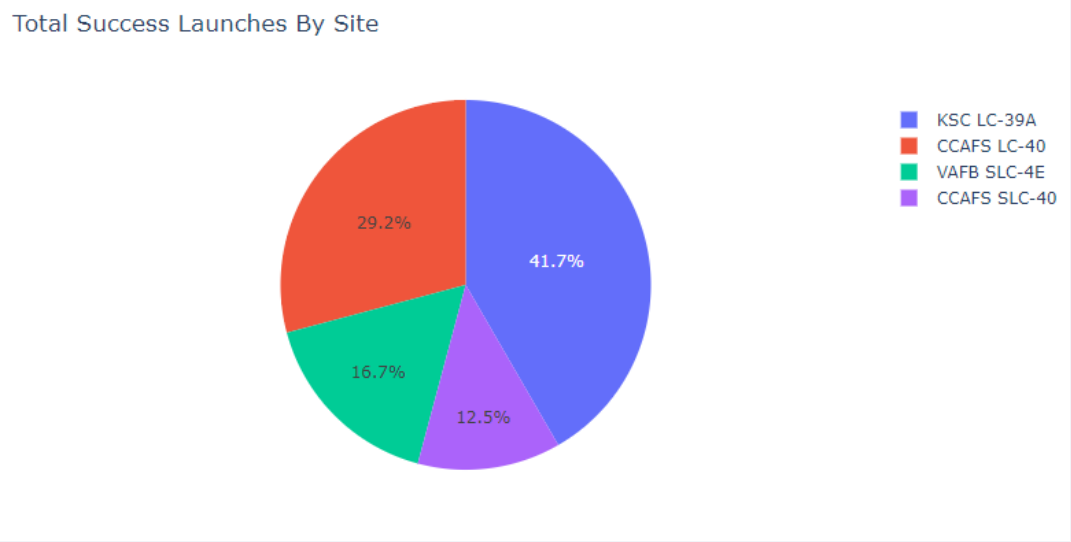
Launch Site Proximities

- Here we have exampled the distance of site CCAFS SLC-40 to the shore, we can see the labeled distance of 0.86 km;
- All other sites are distanced similarly to the shore, but much farther to inhabited areas;


Section 4
Build a Dashboard with Plotly Dash

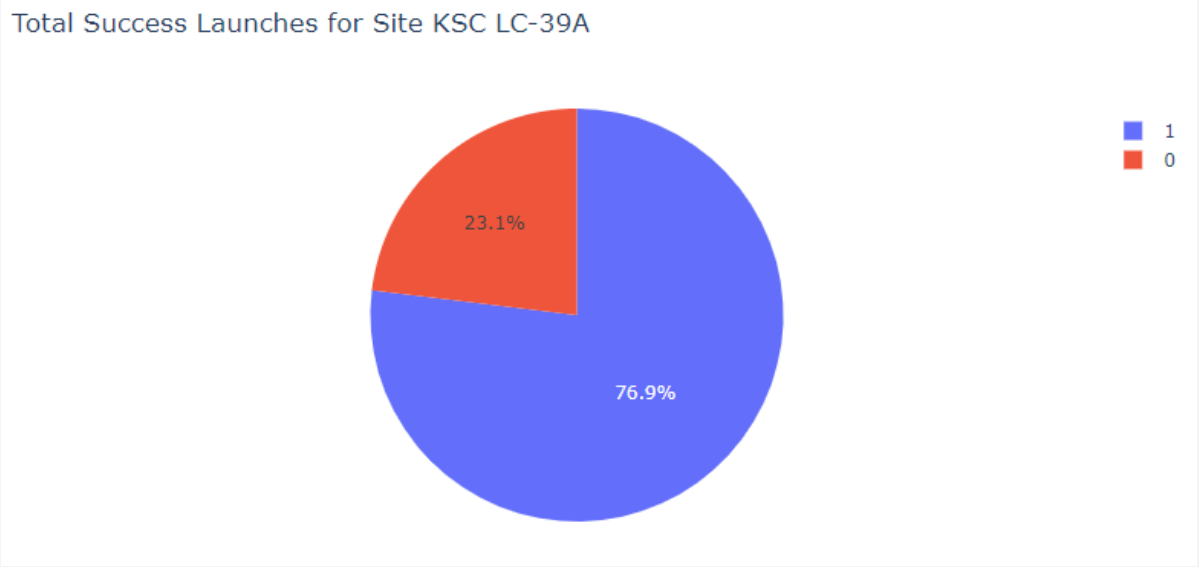
Total Successful Launches

- The following chart shows that the KSC LC-39A site is the one with the most successful launches;
- It is interesting that the CCAFS's have a huge difference in total successful launches.



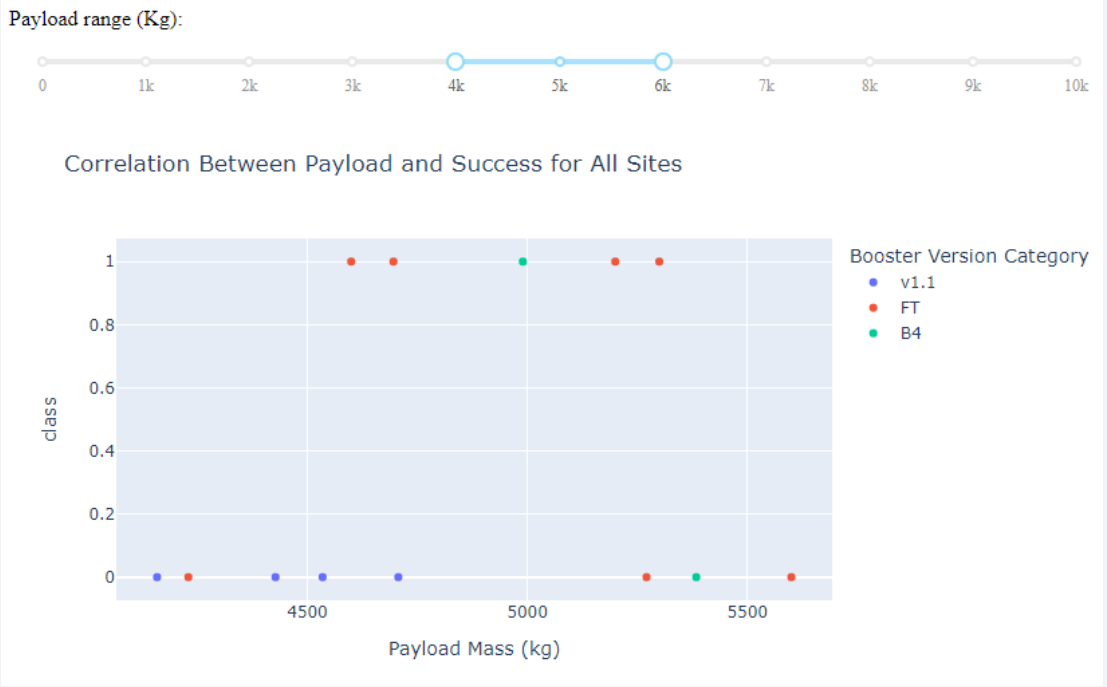
Success Rate for Site KSC LC-39A

- The following chart shows the success rate for the set KSC LC-39A, which is the one with the most successful launches;
- •We can see that it also have a high success rate of 76.9%.



Correlation Between Payload and Success Rate

- Within the range of 4000kg to 6000kg, the FT boosters were the most successful;
- Within the same range all the successes are placed between the 4500kg and 5500kg range.

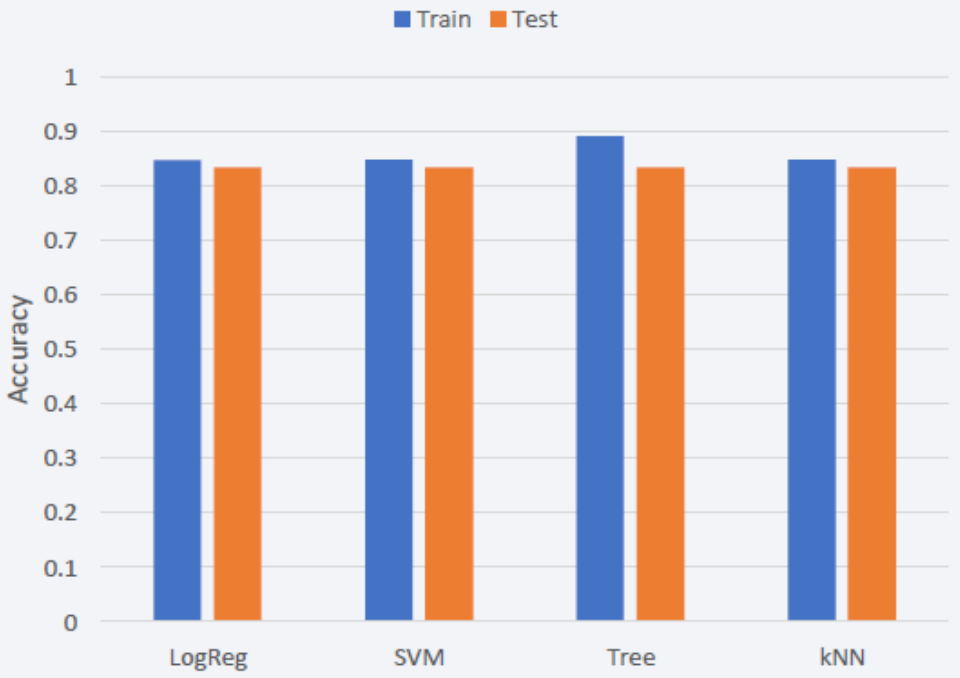


Section 5

Predictive Analysis (Classification)

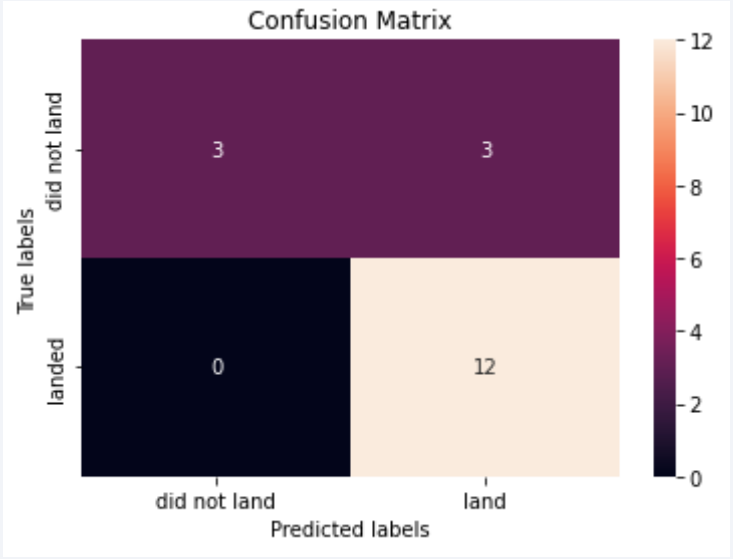
Classification Accuracy

- In the following chart are the accuracies of the four models trained;
- The model with the highest train accuracy was the decision tree, however all of the models performed the same on the test dataset.



Confusion Matrix

- In the following figure is the confusion matrix for the decision tree model;
- It shows a few false positives (3) and no false negatives, with a high accuracy for correct predictions (15).



Conclusions

- The success rate of the landings has been increasing over time, even with a dip in 2018;
- VLEO seems to be the new go-to orbit for launches, and has a high landing success rate;
- There are four launch sites currently, and they are all located near the shore and far from the general population;
- Even though the landings are not always successful the missions are still 99% successful;
- All models have resulted in a similar prediction for the test data, however, there is a low number of datapoints to test.

Thank you!