IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Amjad ALKadi
JUL 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:**
  - Estimating the best total cost of launching missiles and working to anticipate the success of the first stage of launching the missiles
  - Trying to find the best place to launch the missile

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/)
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- **Perform data wrangling**
  - I tried to enrich the data collected by making a special label for the result of the drop based on the data taken in the results after summarizing and analyzing all the advantages (columns) in order to reach the best results and the best decisions
  - **Perform exploratory data analysis (EDA) using visualization and SQL**

Methodology

Executive Summary

- **Perform exploratory data analysis (EDA) using visualization and SQL**
- **Perform interactive visual analyzes with Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - The data was collected in order to improve the model and divided it into training and testing the two data sets and working on evaluating this model through other additional models, with evaluating the accuracy of each model using some parameters

Data Collection

- **Data sets were collected from**

- Space X API

  (https://api.spacexdata.com/v4/rockets)

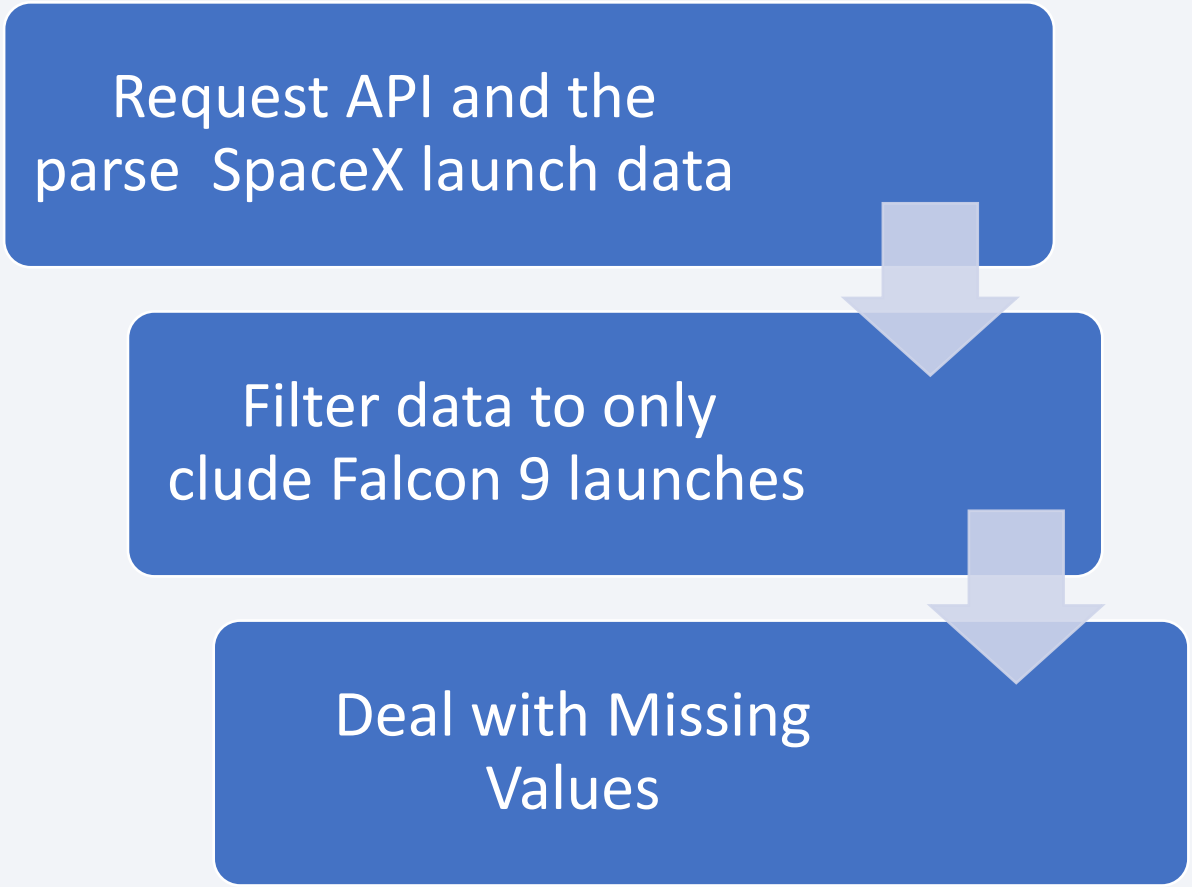
- Wikipedia

  (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.
- Source code:

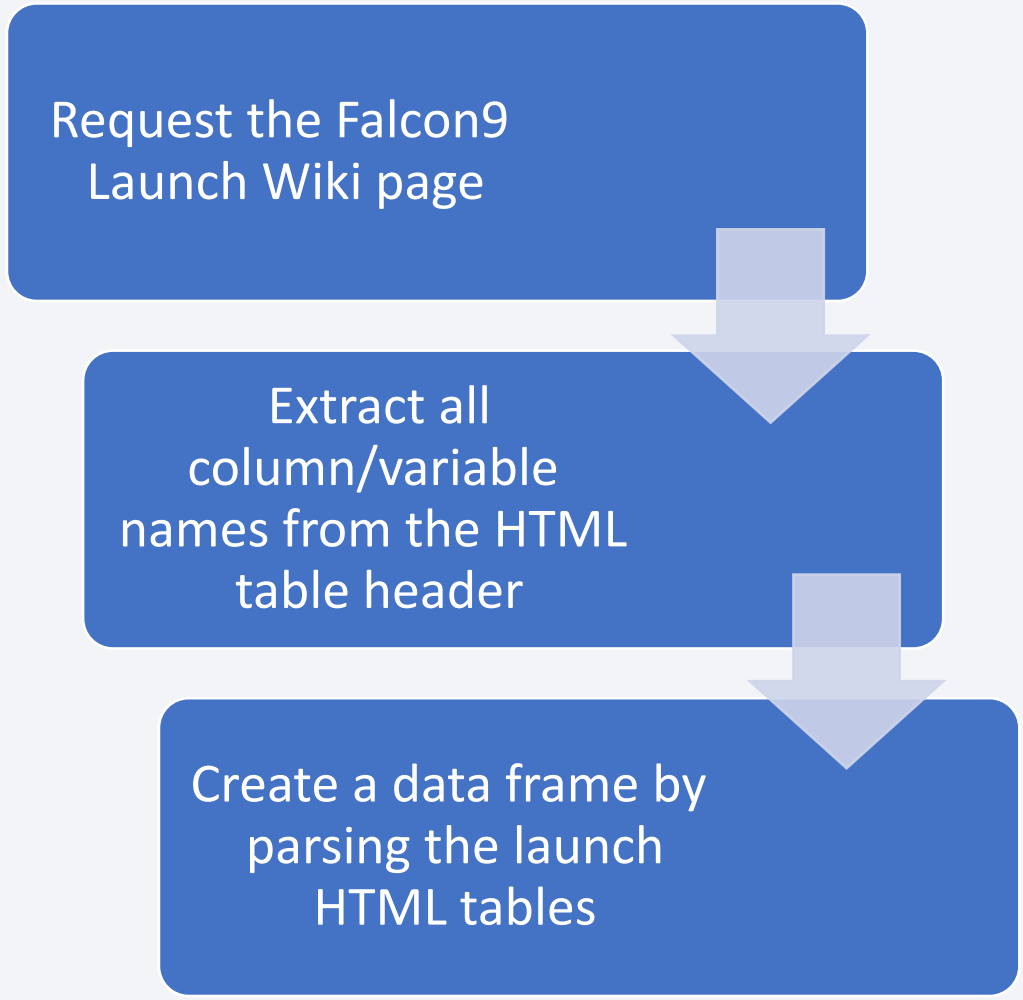
https://github.com/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

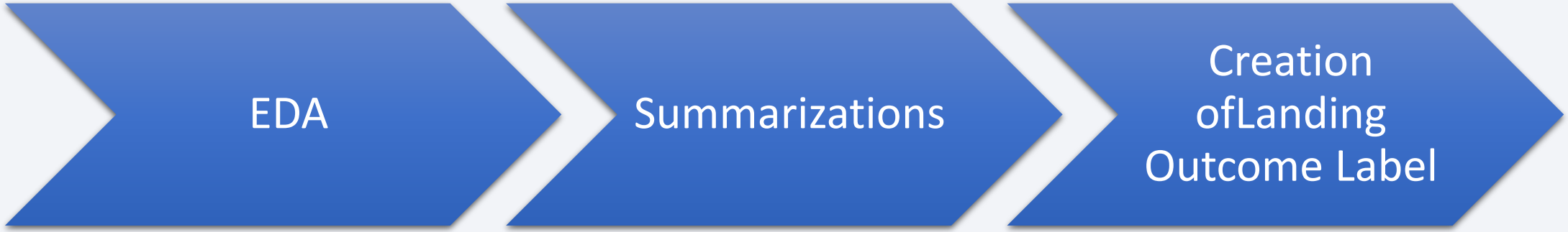
- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.
- Source code :

https://github.com/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/jupyter-labs-webscraping.ipynb



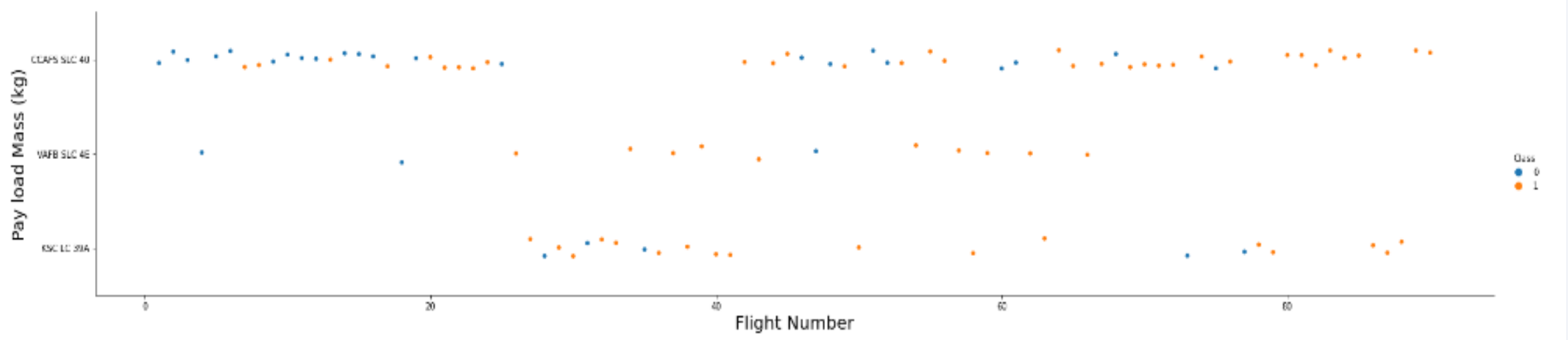
Data Wrangling

- 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/AmjadALKadi/Applied_Data_Science_Capstone_proj/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:
- 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/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/jupyter-labs-eda-dataviz.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
  - 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/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

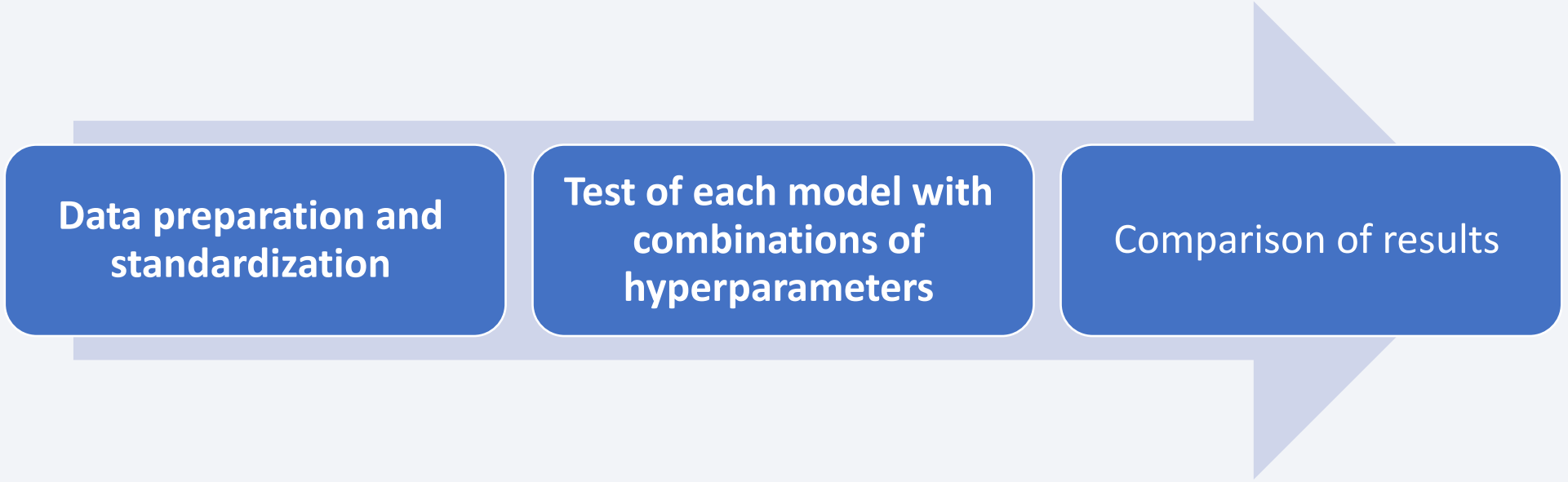
- Markers, circles, lines and marker clusters were used with Folium Maps
- Markers indicate points like launch sites
- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site
- Lines are used to indicate distances between two coordinates.
- Source Code : https://github.com/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/lab_jupyter_launch_site_location.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/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/SpaceX%20Plotly.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/AmjadALKadi/Applied_Data_Science_Capstone_proj/blob/main/Machine%20Learning%20Prediction.ipynb



Results

- **Exploratory data analysis results**
  - There are different launchers for Space X 4
  - The first launches of Space X were conducted by NASA
  - The average payload of the boosted missile F9 v1.1 is approximately 2928 kg
  - Remove the result of a successful landing was in 2015 5 years after the first attempt
  - Falcon 9 boosted versions succeeded in landing a drone in a ship with a higher payload
  - 100% of the results were very important in determining the launch,
  - There is an interface failure encountered two enhanced versions when landing in the case of drone ships in 2015, these versions are: F9 v1.1 B1012 and F9 v1.1 B1015
  - noticeable improvement in the results of the landing without any problems with the experience with the progression of time due to the experience

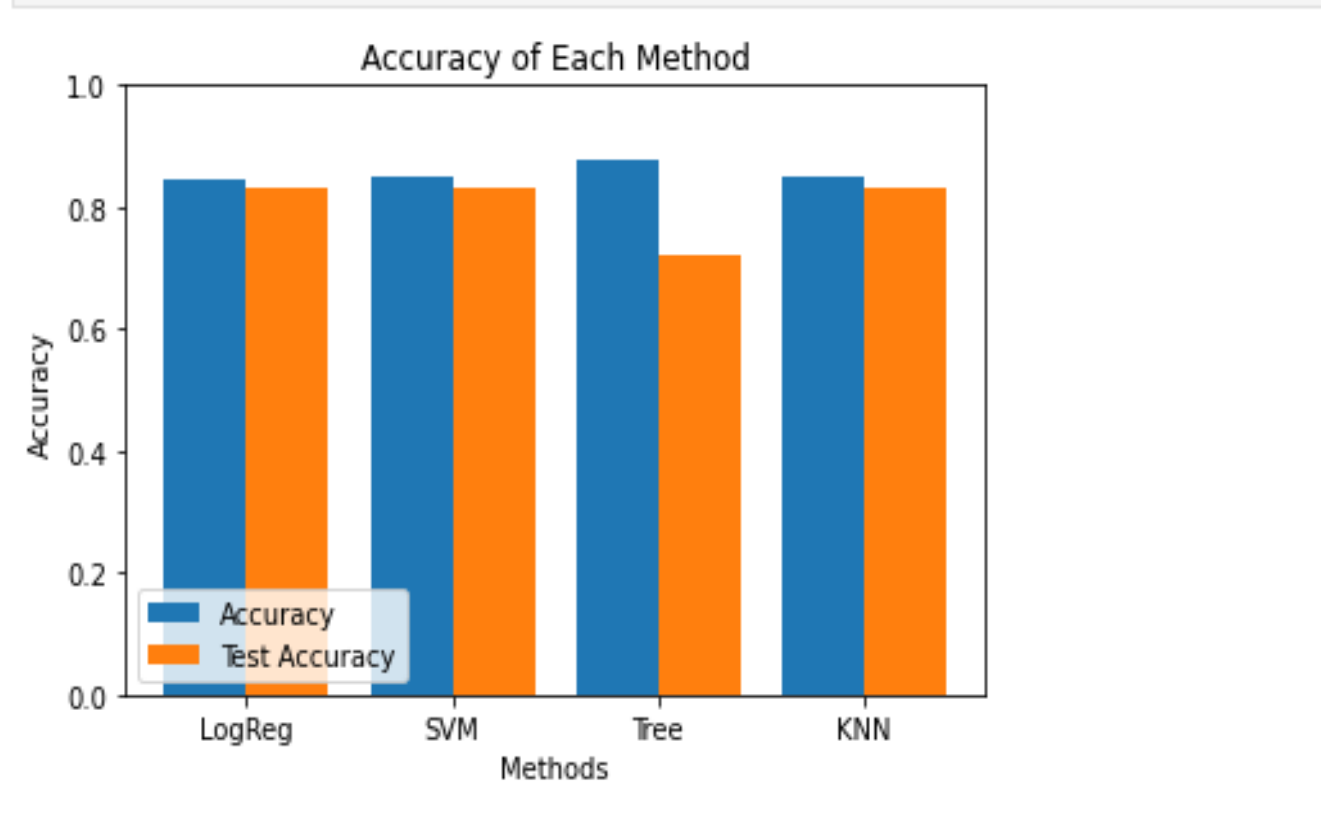
Results

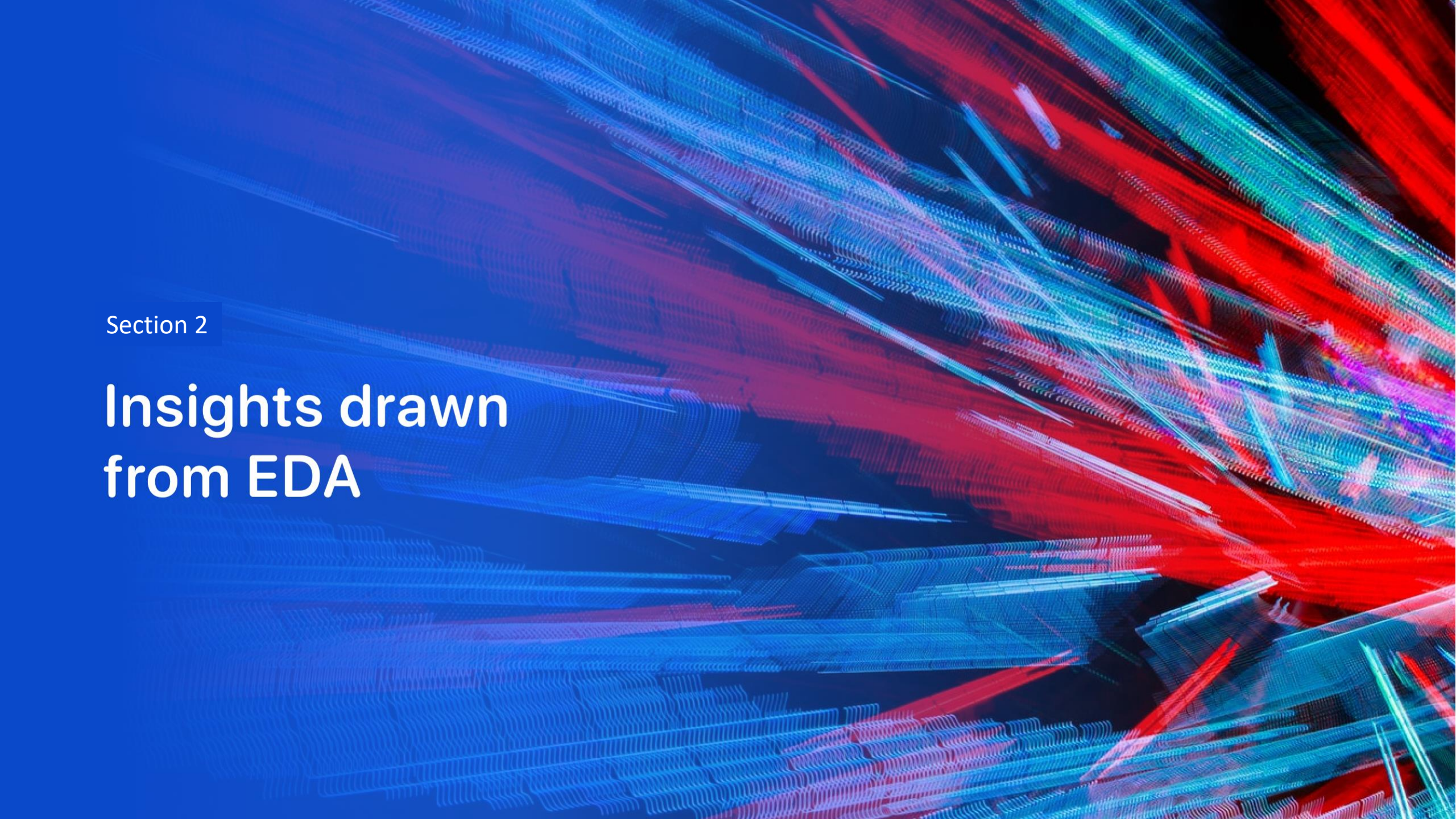
Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.

Most launches happens at east cost launch sites.

Results

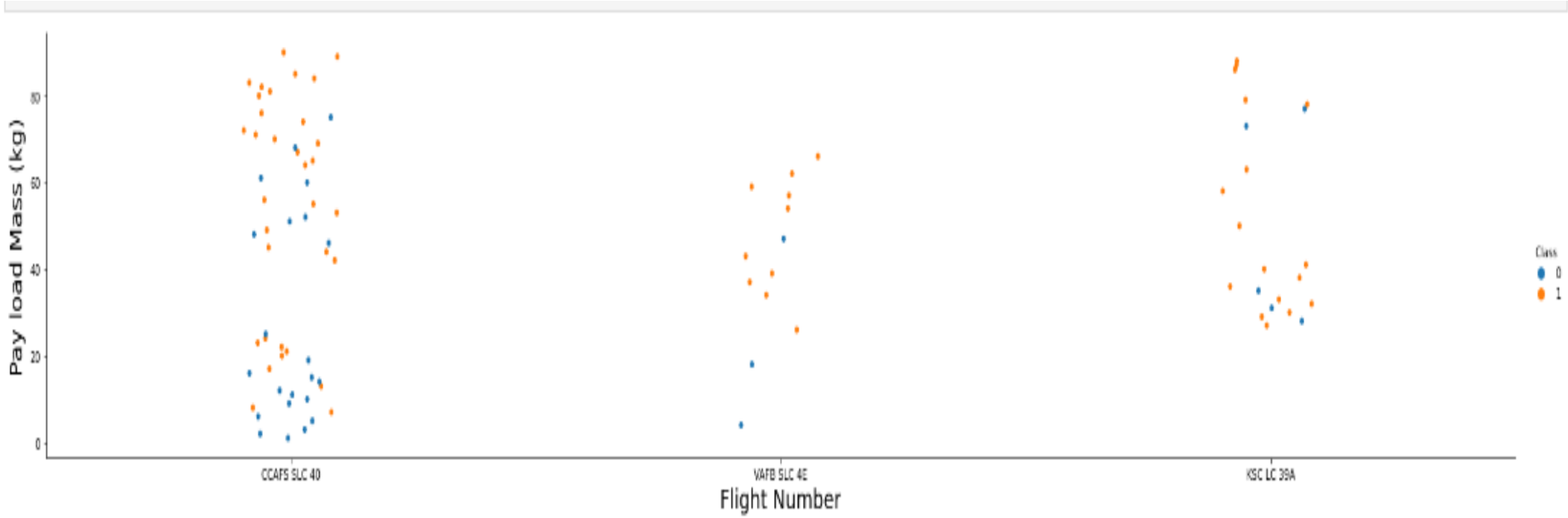
Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.




Section 2
Insights drawn from EDA

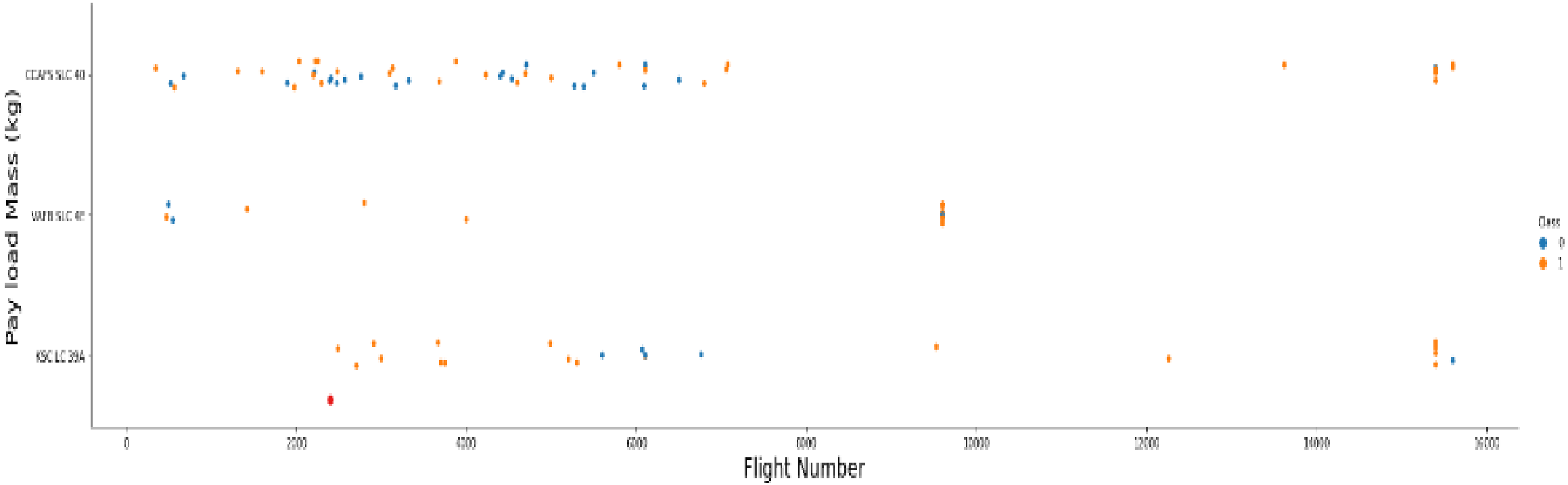
Flight Number vs. Launch Site

- 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



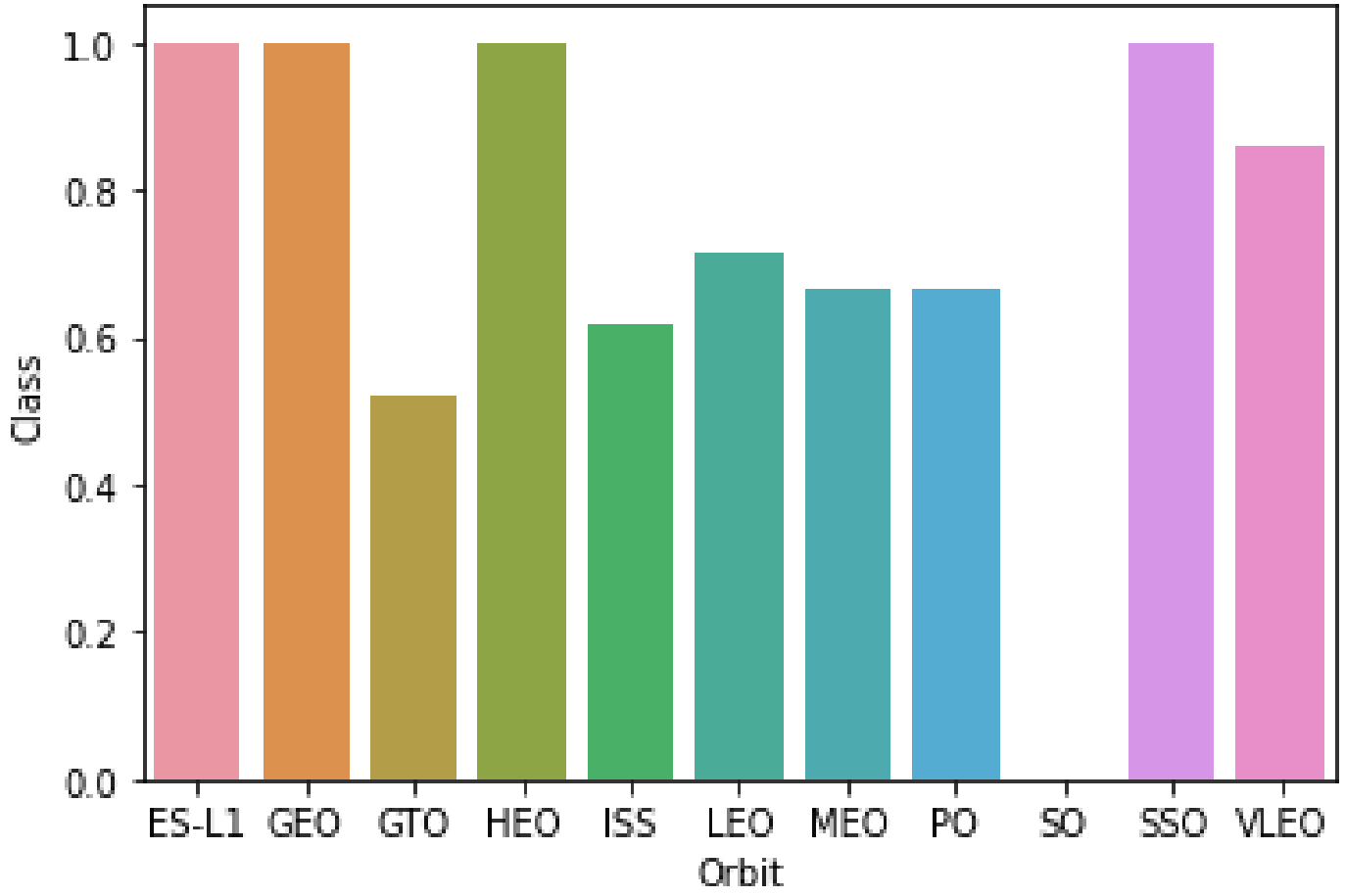
Payload vs. Launch Site

- Payloads over 9,000kg (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 L 39A launch sites.



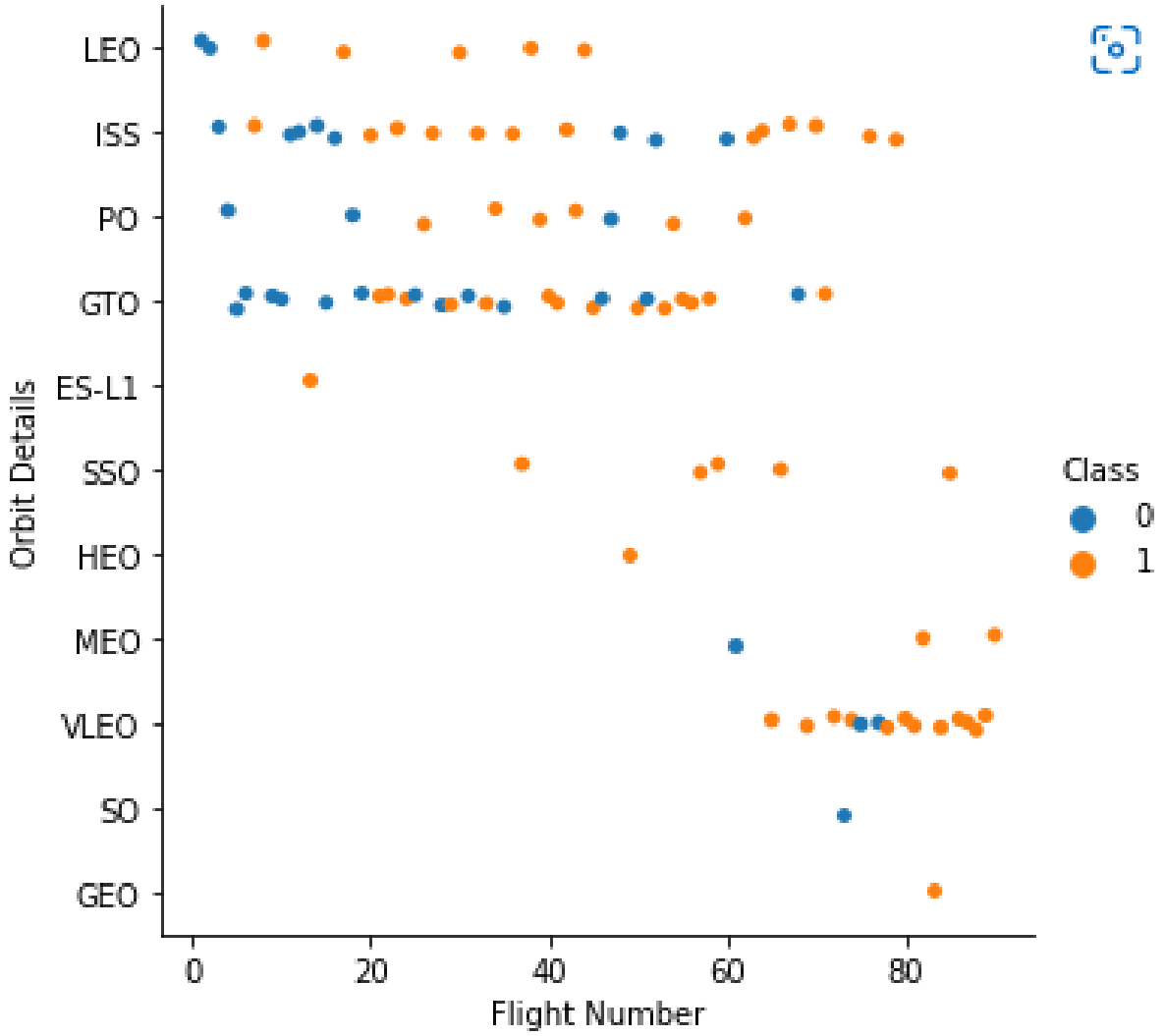
Success Rate vs. Orbit Type

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



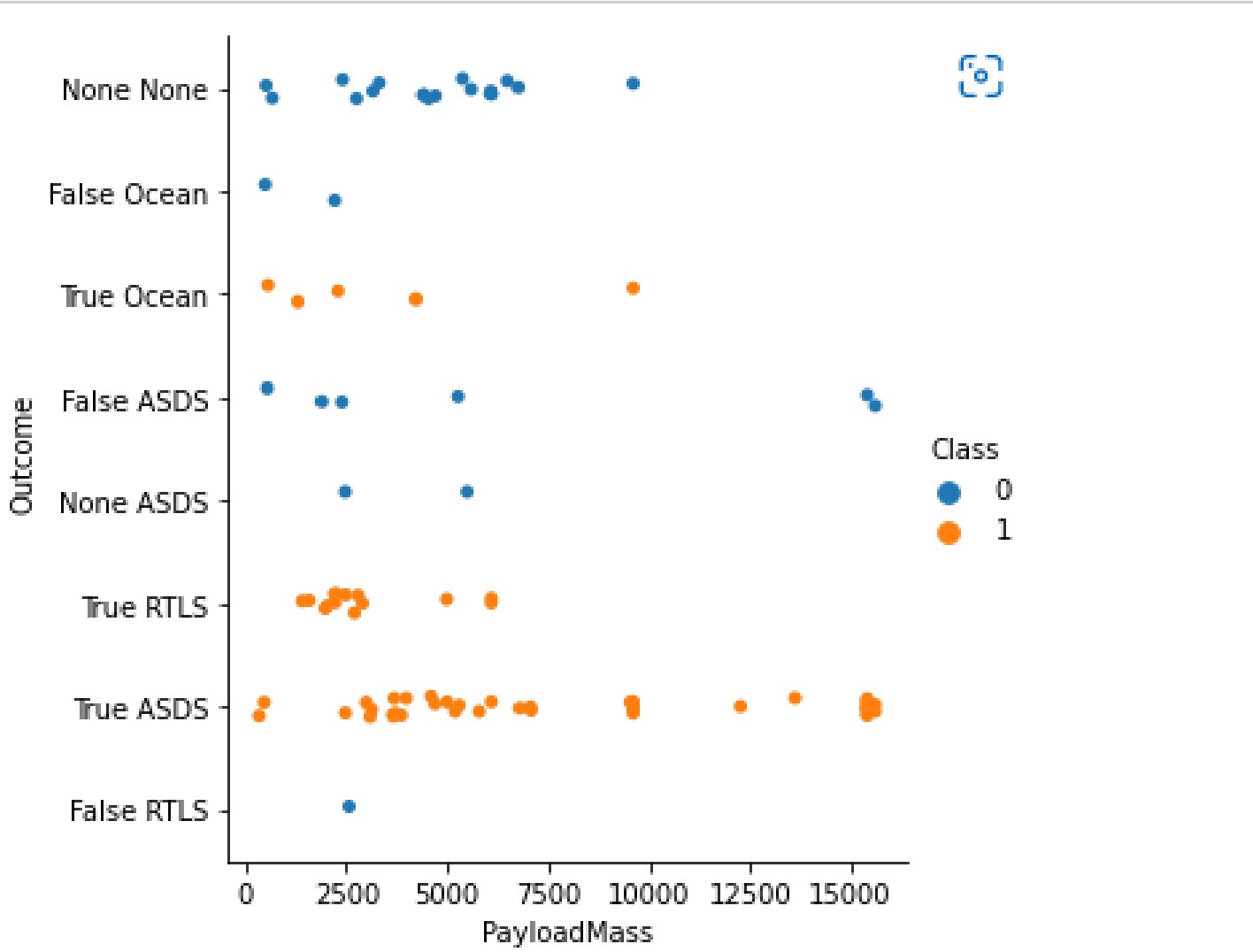
Flight Number vs. Orbit Type

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



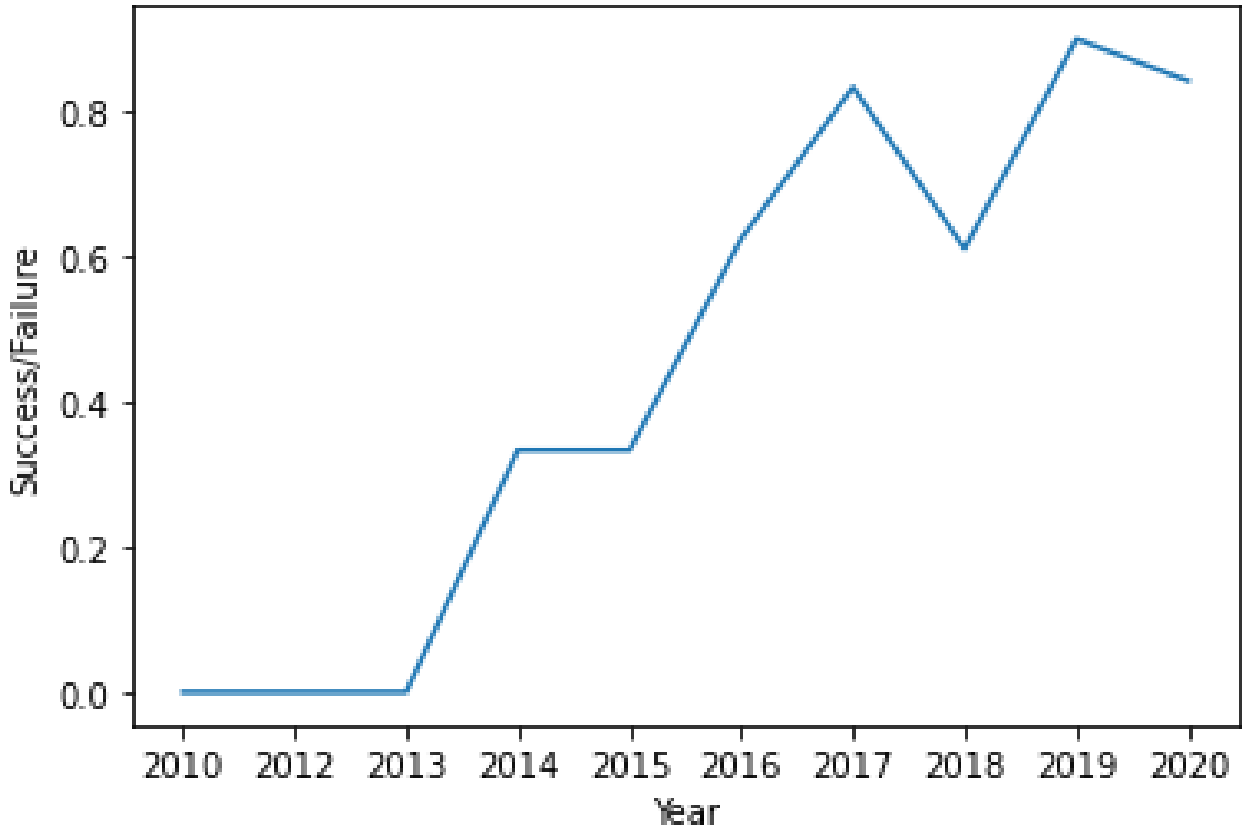
Payload vs. Orbit Type

- 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 adjusts and improvement of technology.



All Launch Site Names

- According to data, there are four launch sites:

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

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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 carried by boosters from NASA:
- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

TOTAL_PAYLOAD_kg
111268

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version 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_MASS__KG_)
2928.4

First Successful Ground Landing Date

- First successful landing outcome on ground pad:
- 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 01/07/2015.

Min_data
01-05-2017

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

Total Number of Successful and Failure Mission Outcomes

- 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	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Boosters which have carried the 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 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

2015 Launch Records

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

Landing_Outcome	QTY
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Controlled (ocean)	3
Failure	3
Failure (parachute)	2
No attempt	1

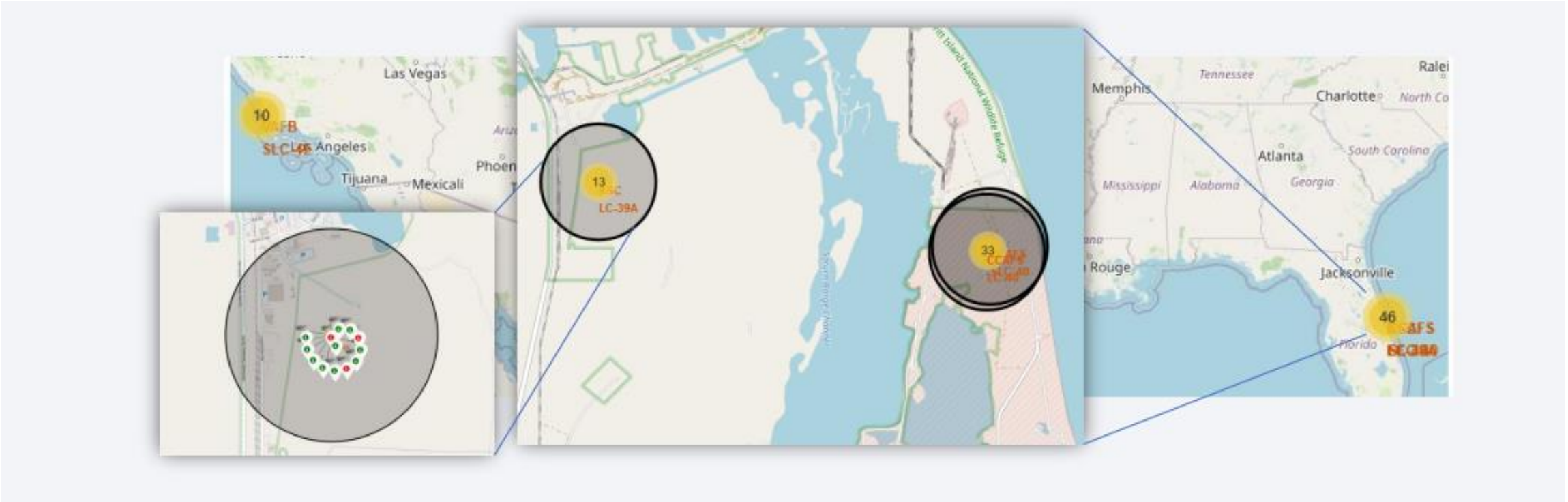

Section 3
Launch Sites
Proximities Analysis

All launch sites

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

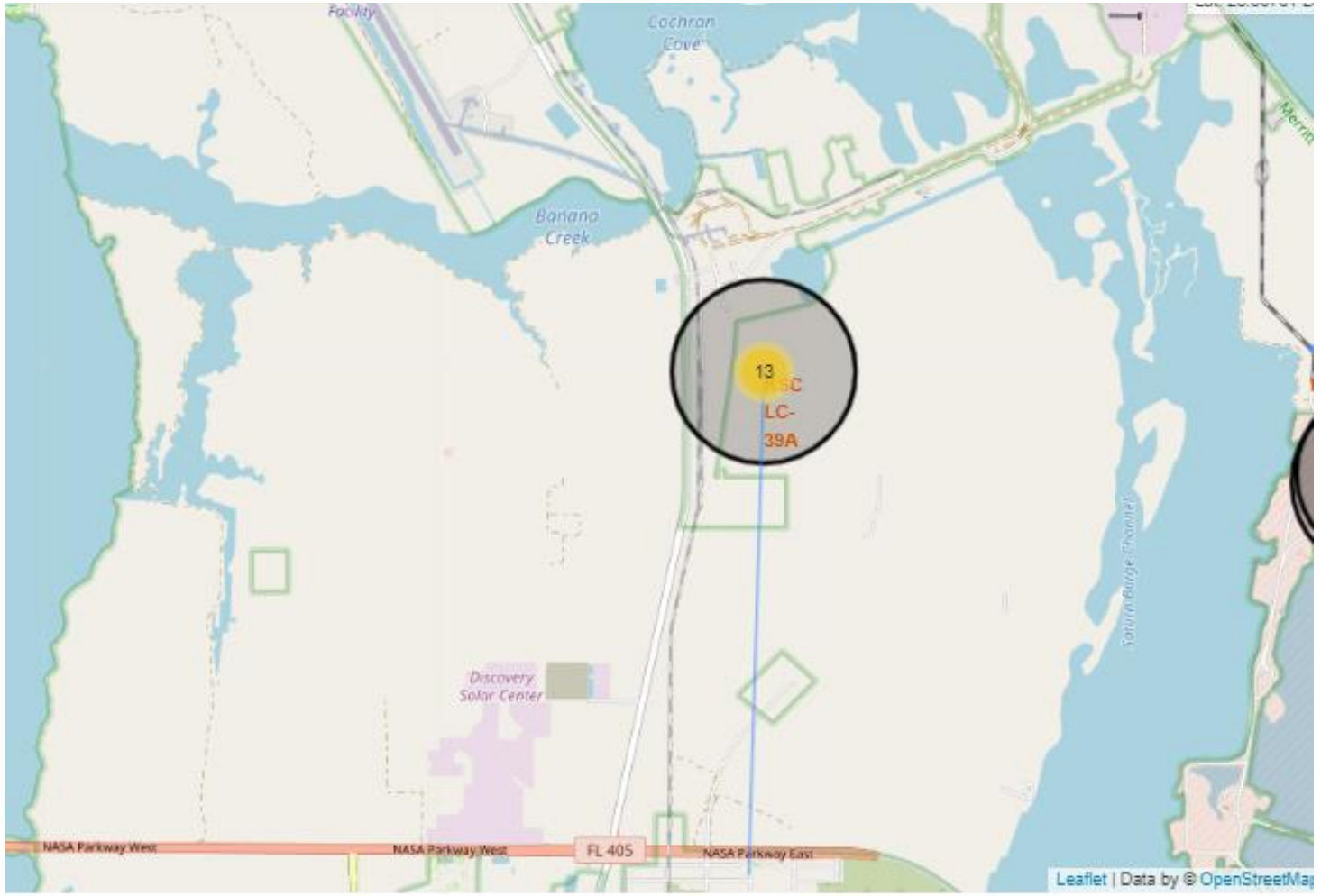
Site's Launch Outcomes

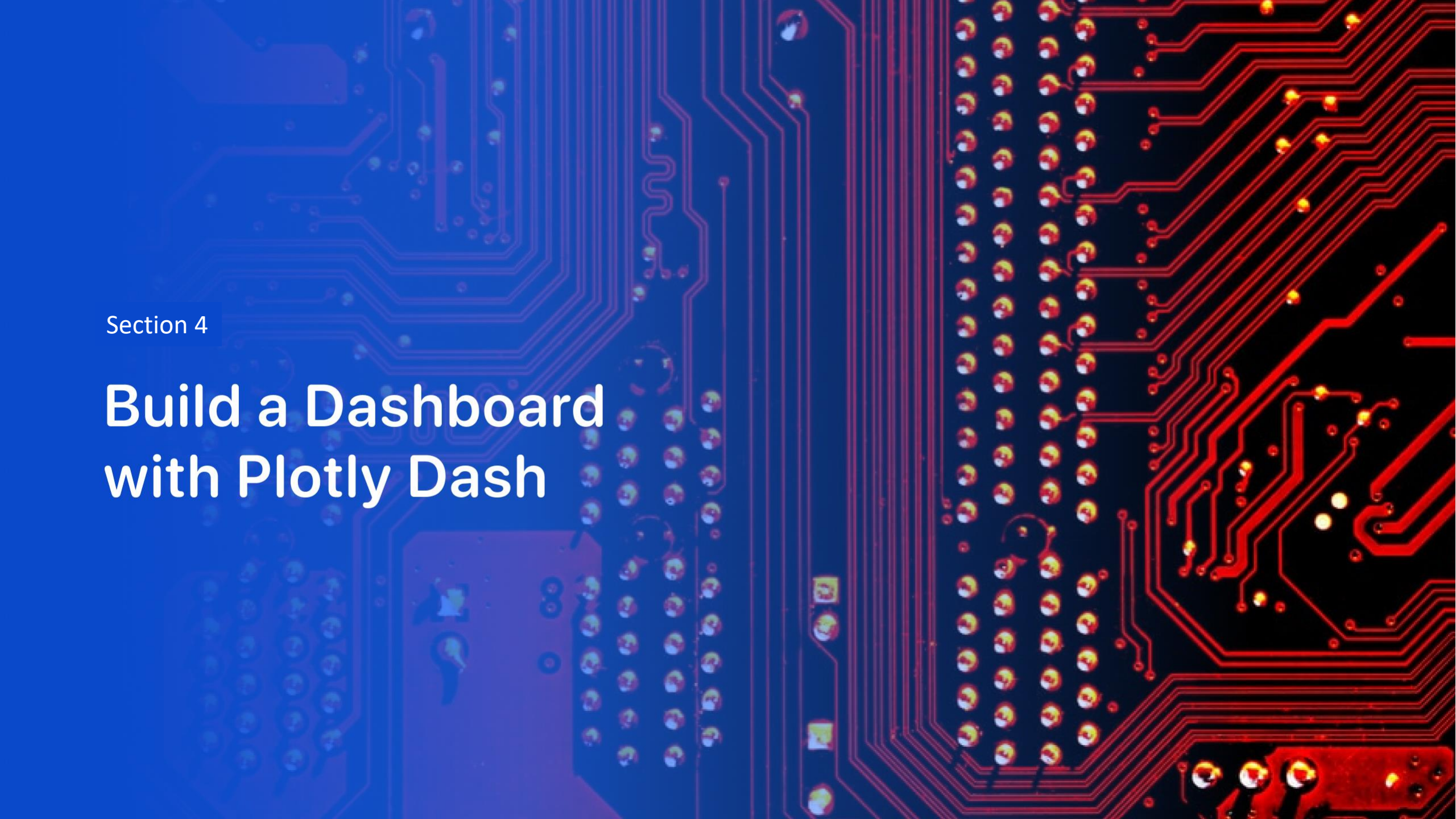
- Example of KSC LC-39A launch site launch outcomes
- Green markers indicate successful and red ones indicate failure



Logistics and Safety

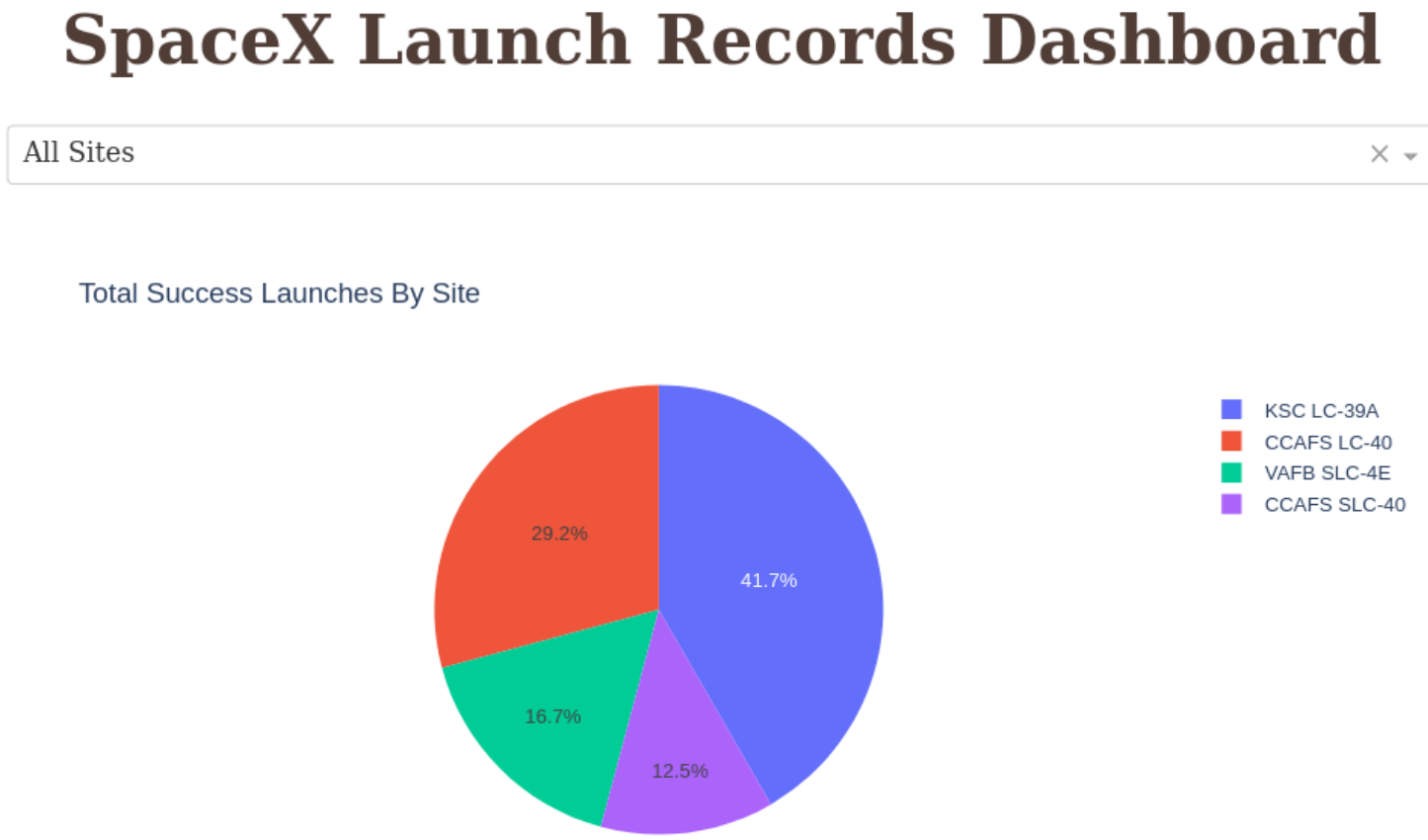
- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.




Section 4
Build a Dashboard with Plotly Dash

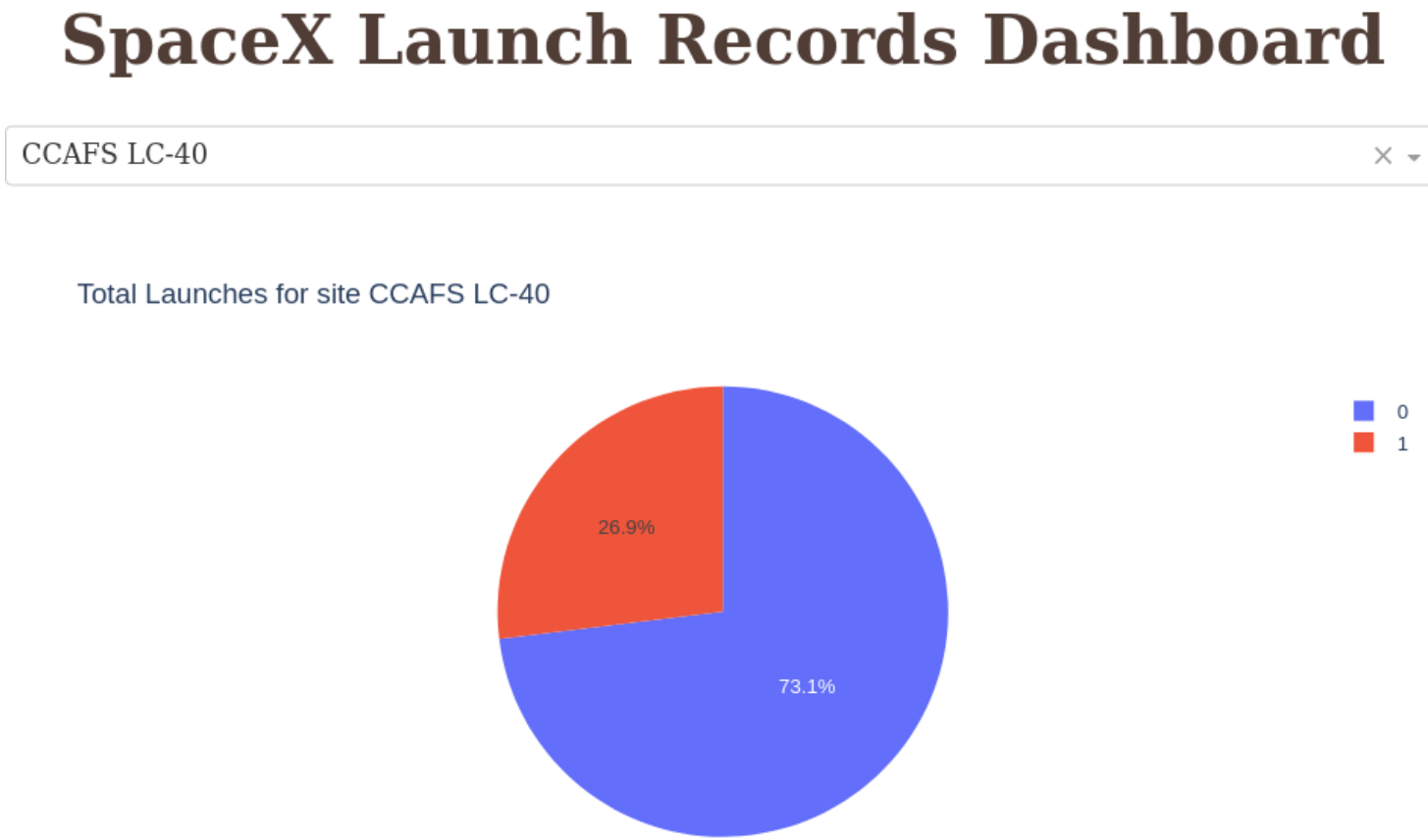
Successful Launches by Site

- 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

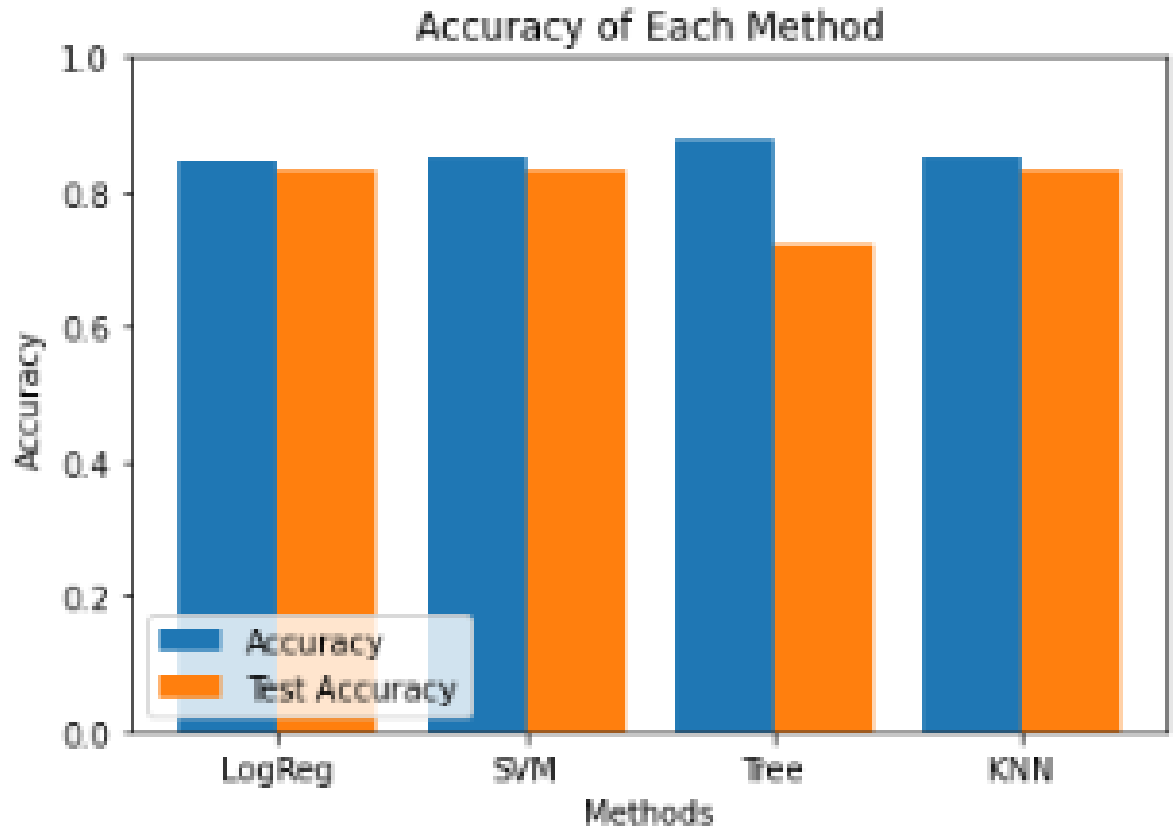
- There’s not enough data to estimate risk of launches over 7,000kg

Section 5

Predictive Analysis (Classification)

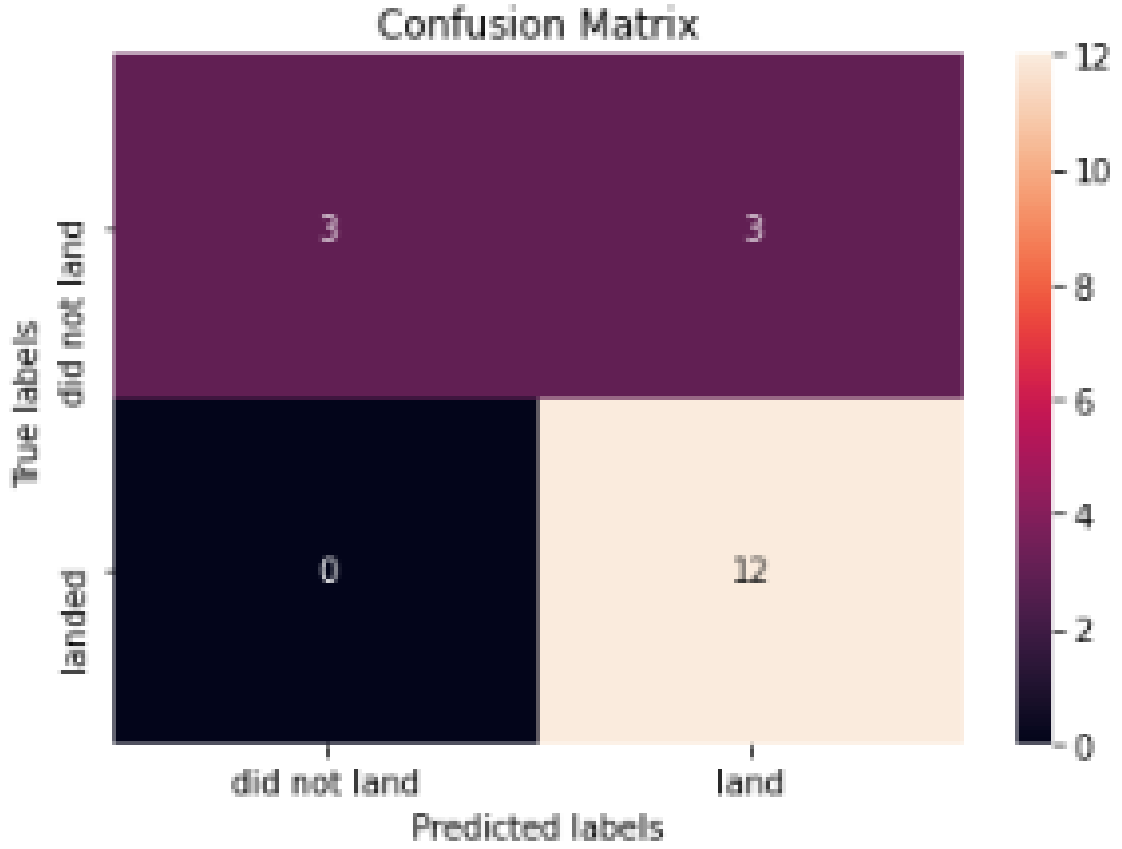
Classification Accuracy

- 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

- **Work was done on several data from several sources and as a result of the analysis and revision of the results during data processing As a result of the analyzes after revision we recommend:**
  - The best platform and launch site was the KSC LC-39A
  - Launches with a payload of more than 7000 kg are successful launches and no steps were recorded
  - There are important and successful results and conclusions. Despite this, the good landing and the security in the beginning was weak, but with the passage of time and gaining experience, the landing began to improve as a result of the development of experiences, the development of the tools and means used, the development in missile technologies and the management of launch operations
  - Based on the conclusions, we can benefit from the classification, especially the classification matrix, in order to predict successful landing operations, which contributes to increasing profits, of course, after evaluating the matrix model, verifying its validity and updating it.

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!