IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Eric Fuentes Rico
2022-12-12

Outline

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

Executive Summary

- Summary of the applied methodology:
  - Data collection was made with WebScraping and SpaceX's API.
  - Exploratory Data Analysis (EDA) and Data Visualization were made with Python.
  - Predictions were made with ML algorithms.
- Summary of all results:
  - Open-source data from public sources was successfully gathered.
  - EDA allowed to identify which features are more correlated to the success rate.
  - ML algorithms found the best model to achieve the business' objectives.

Introduction

- **Project's objective:** to evaluate the viability of company Space Y competing with Space X.
- **Problems that need answers:**
  - Best places to launch missions
  - Best way to estimate the launches' net cost by predicting successful landings of the rockets' first stages.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX's data was collected from 2 sources:
    - Space X's 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

- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
    - Data collected until this step were normalized, divided into 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

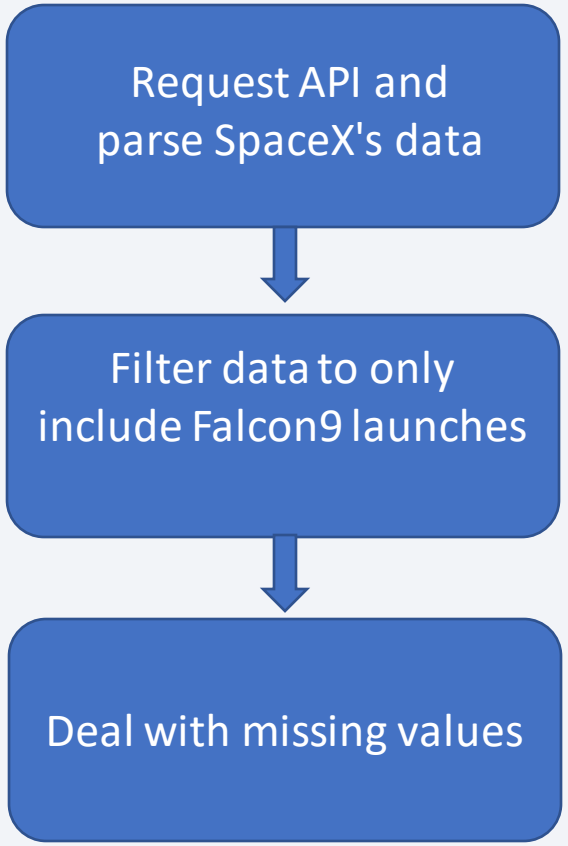
Datasets were collected from 2 sources:

- Space X's API (https://api.spacexdata.com/v4/rockets/)
- Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using webscraping technics.

Data Collection – SpaceX's API

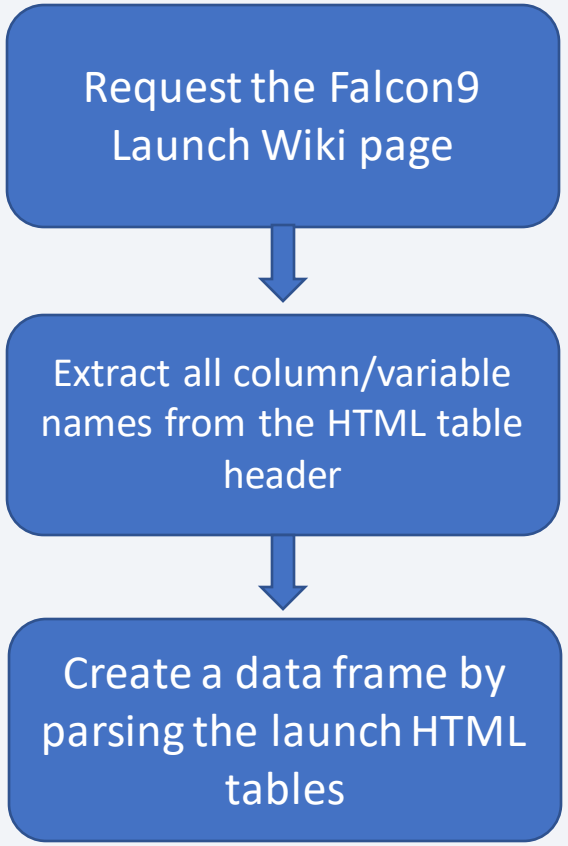
- SpaceX offers a public API that was used to gather the data.
- This API was used according to the flowchart on the right and then the data was cleaned.

- **Source code:** https://github.com/efuentesrico/IMB_Coursee/blob/main/1.SpaceX_Data_Collection_with_API.ipynb



Data Collection - Scraping

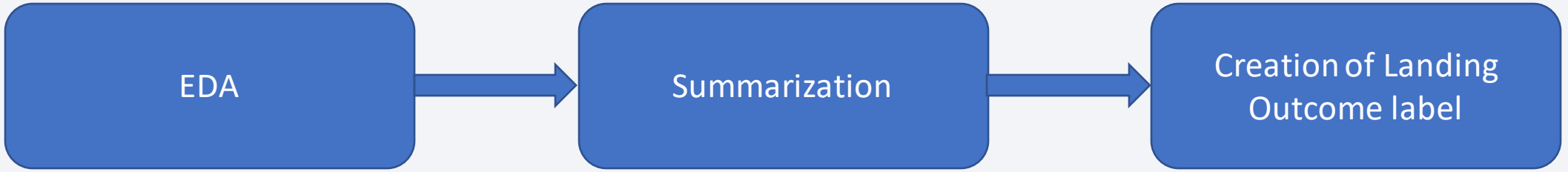
- Data from SpaceX launches were also obtained from Wikipedia
- Data were downloaded according to the flowchart and then cleaned.



- **Source code:** https://github.com/efuentesrico/IMB_Coursee/blob/main/2.Data_Collection_with_Webscraping.ipynb

Data Wrangling

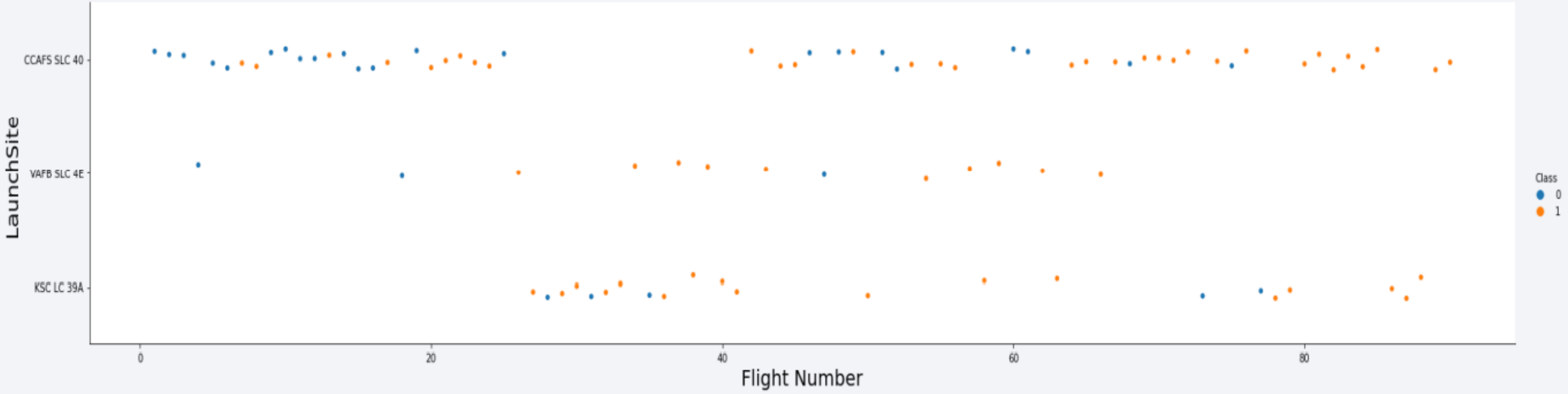
- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries of the launches per site, the occurrences of each orbit and the occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from the Outcome column.



- **Source code:** https://github.com/efuentesrico/IMB_Coursee/blob/main/3.Data_Wrangling.ipynb

EDA with Data Visualization

- To explore further explore the data, scatterplots and barplots were used to visualize the following relationships:
  - (Payload Mass, Flight Number), (Launch Site, Flight Number), (Launch Site, Payload Mass), (Orbit, Flight Number), (Payload, Orbit)



- **Source code:** https://github.com/efuentesrico/IMB_Coursee/blob/main/5.EDA_with_Data_Visualization.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 names 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/efuentesrico/IMB_Coursee/blob/main/4.EDA_with_SQL.ipynb

Build an Interactive Map with Folium

Markers, circles, lines and marker clusters were used with Folium Maps as follows:

- Markers indicate points like launch sites
- Circles indicate highlighted areas around specific coordinates, like NASA's Johnson Space Center
- Marker clusters indicate 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/efuentesrico/IMB_Coursee/blob/main/6.Interactive_Visual_Analytics_with_Folium.ipynb

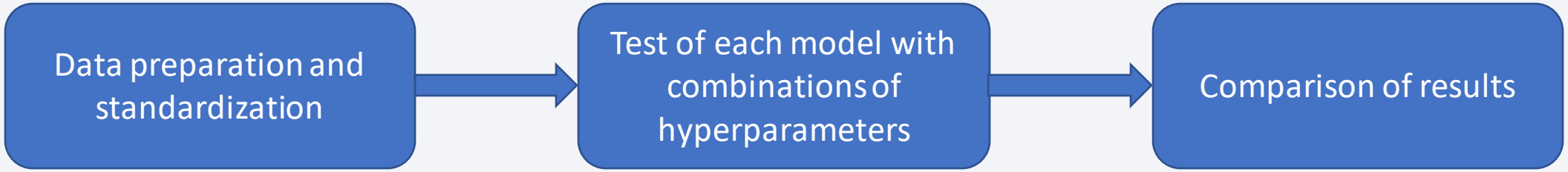
Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize the data:
  - Percentage of launches by site
  - Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify the best place to launch according to payloads.

- **Source code:** https://github.com/efuentesrico/IMB_Coursee/blob/main/SpaceX_Dash_App.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/efuentesrico/IMB_Coursee/blob/main/7.Machine_Learning_Prediction.ipynb

Results

Exploratory data analysis' results:

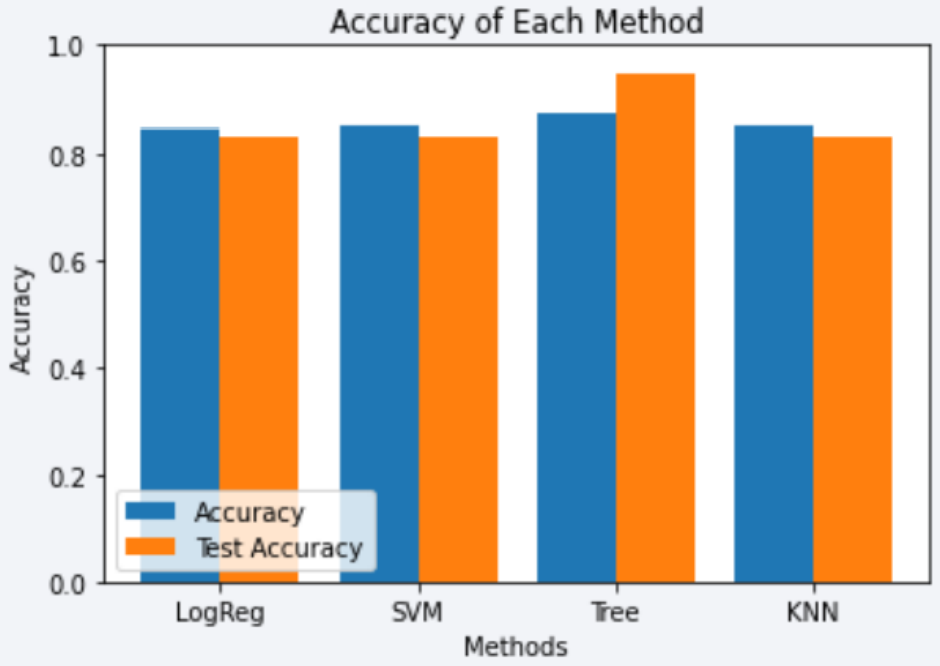
- Space X uses 4 different launch sites
- The first launches were done to Space X itself and NASA
- The average payload of F9 v1.1 booster is 2,928 kg
- The first success landing outcome happened in 2015, five years after the first launching
- Many Falcon 9 booster versions were successful at landing on 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 possible to identify that launch sites are usually located in safe places near the sea and have a good logistic infrastructure around.
- Most launches happen at east coast launch sites.

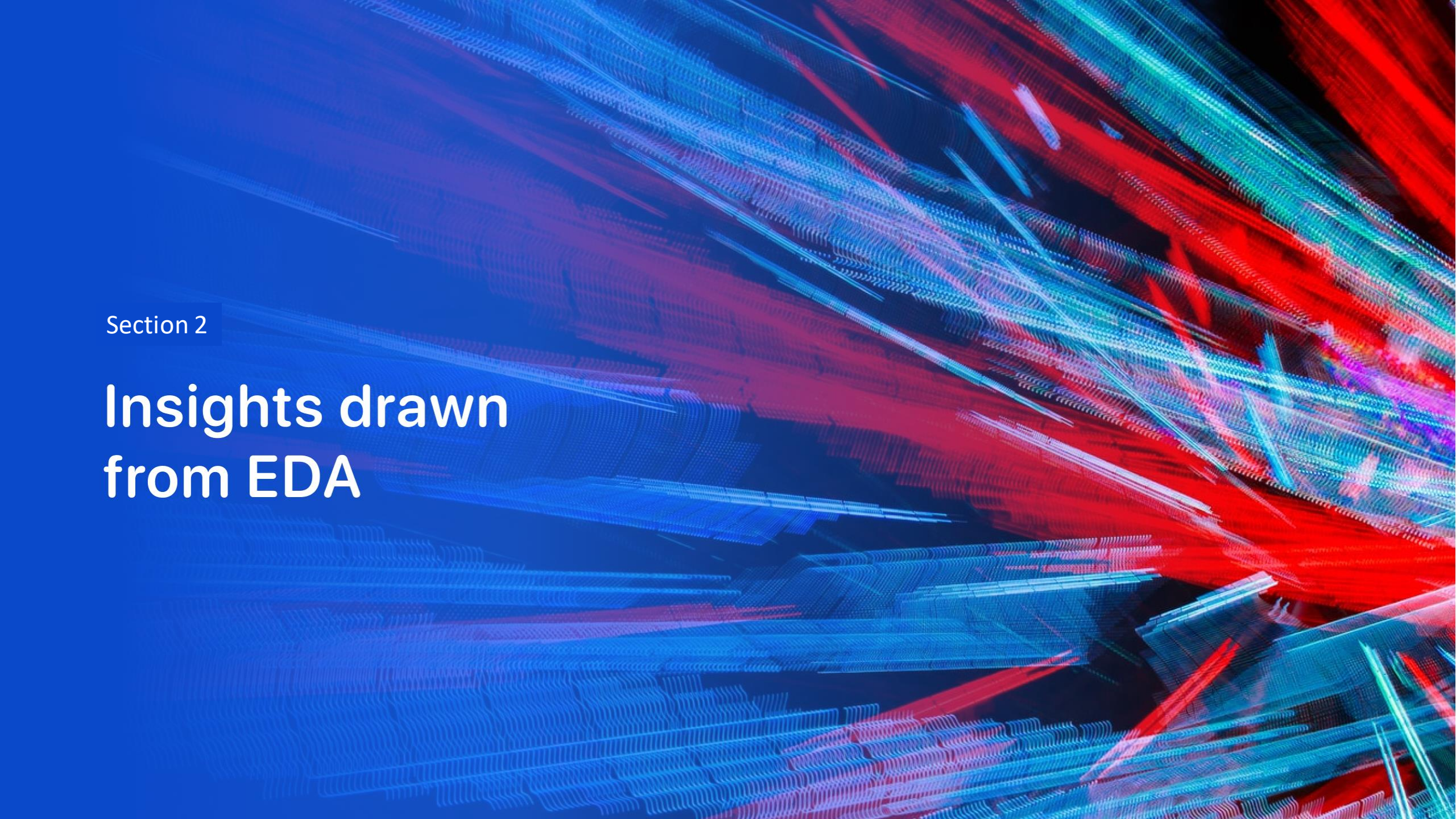
Results

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

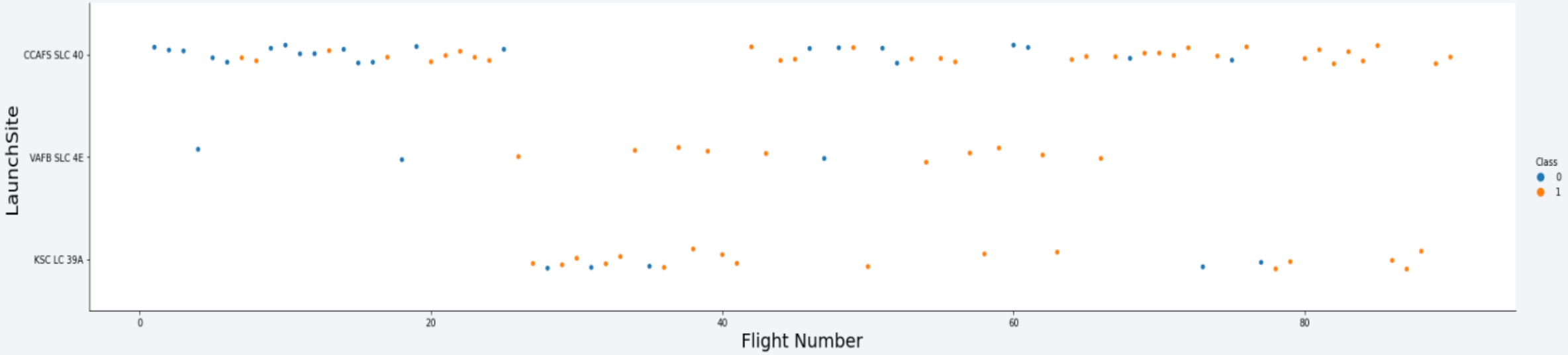


Section 2

Insights drawn from EDA

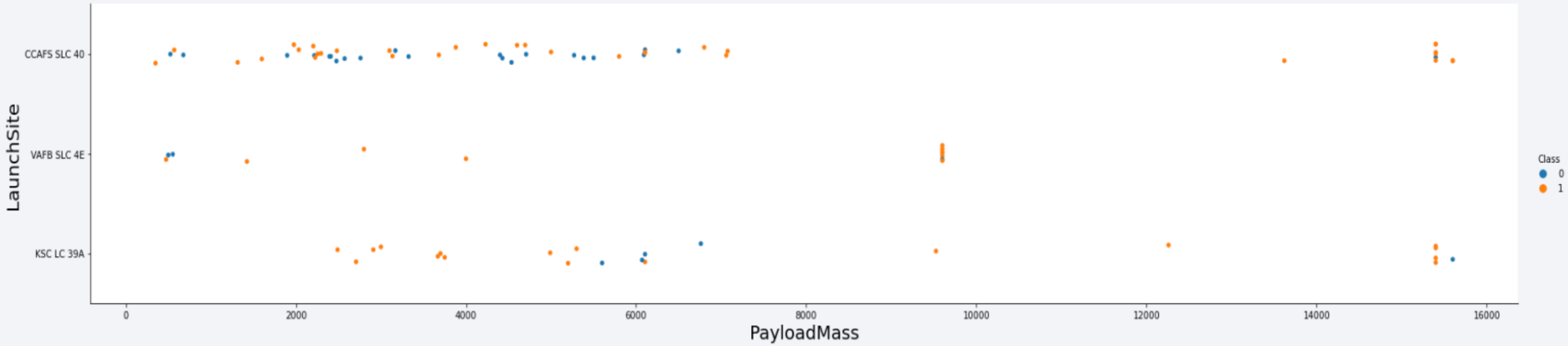


Flight Number vs. Launch Site



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

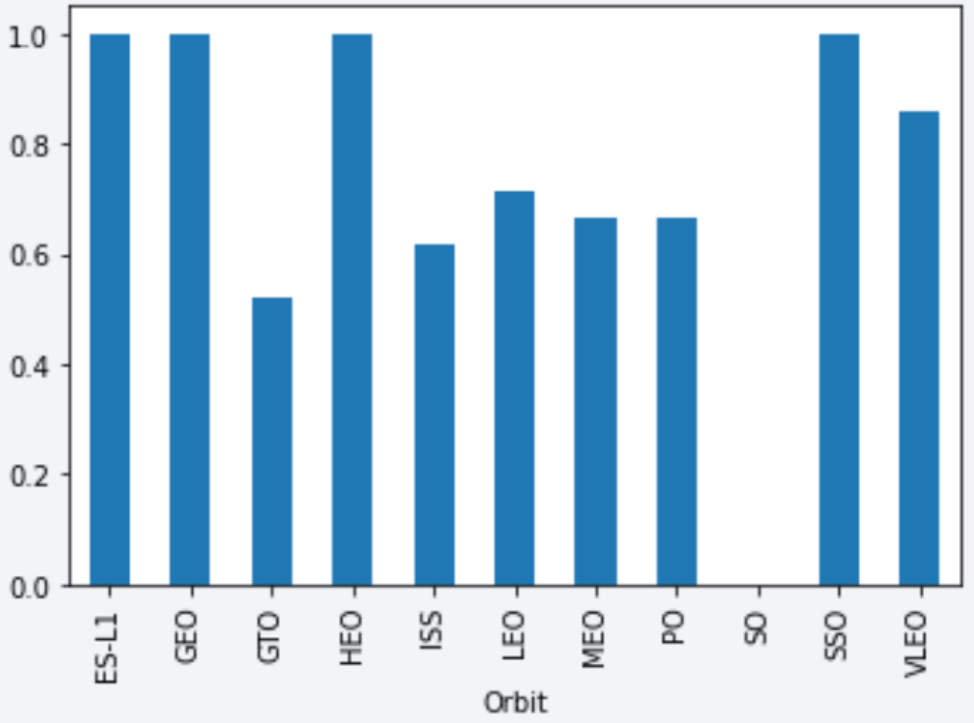
Payload vs. Launch Site



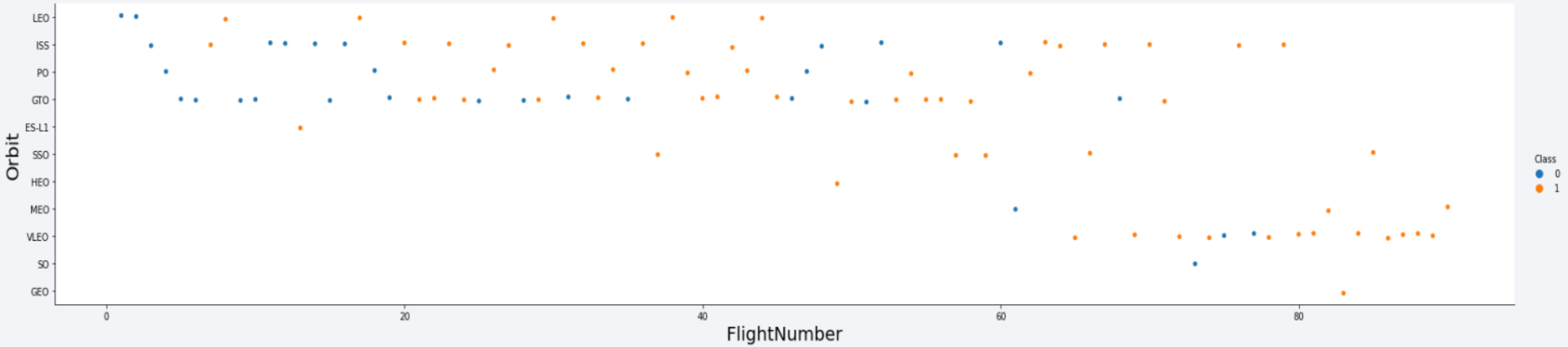
- Payloads over 9,000kg (about the weight of a school bus) have an excellent success rate
- Payloads over 12,000kg seem to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The orbits with the highest success rates are:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Followed by:
  - VLEO (above 80%)
  - LFO (above 70%)

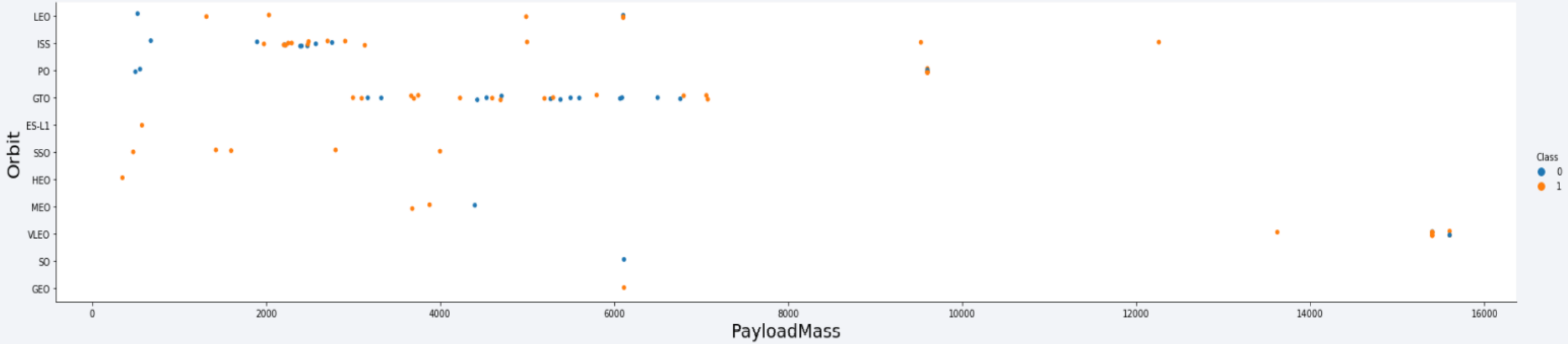


Flight Number vs. Orbit Type



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

Payload vs. Orbit Type

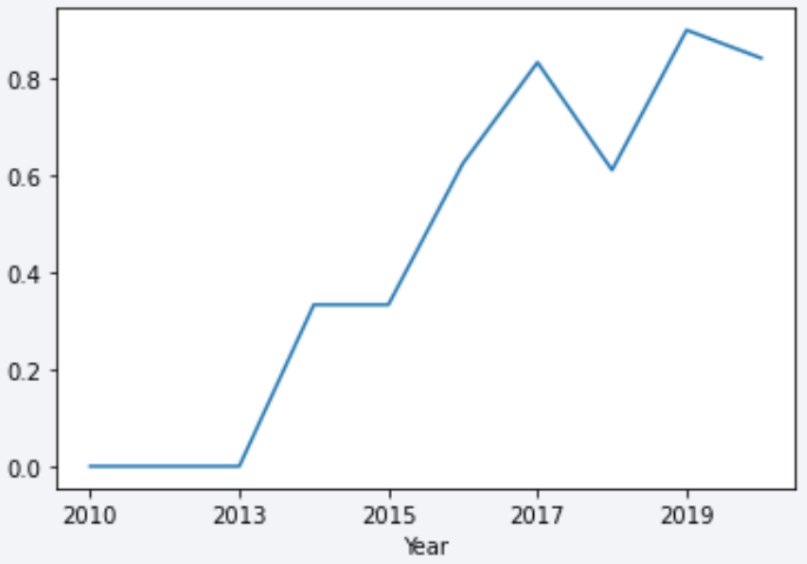


- Apparently, there is no relation between payload and success rate for orbit GTO
- ISS orbit has the widest range of payload and a good rate of success
- There are few launches to the orbits SO and GEO

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Success rate started increasing in 2013 and kept doing so until 2020
- It seems that the first three years were a period of adjustments and improvements in 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 gives 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 is possible to identify the first occurrence, which 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 results.

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 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 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

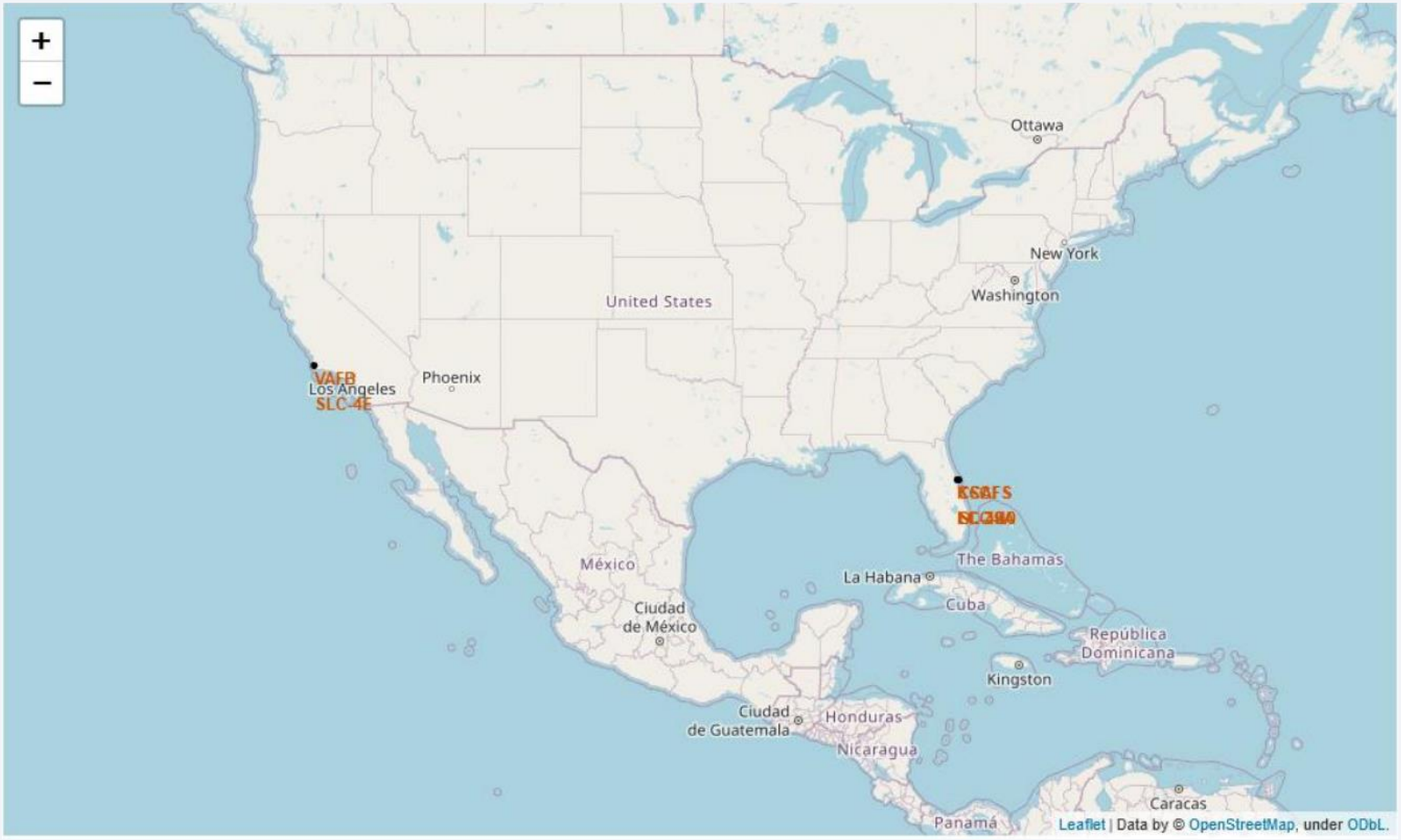
- Ranking of all landing outcomes between the date 2010-06-04 and 2017- 03-20:

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

- These data alerts us that “No attempt” must be taken in account, what means that the reason why the launching was not attempted must be further investigated.


Section 3
Launch Sites
Proximities Analysis

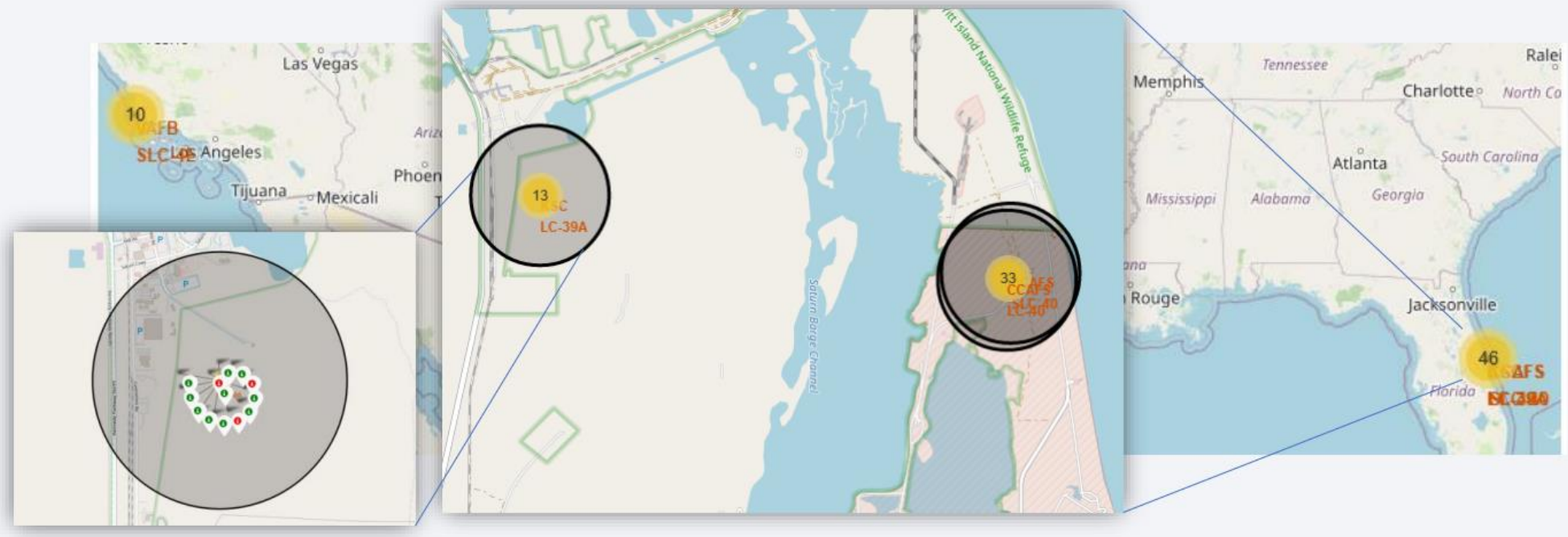
Location of all launch sites



- Launch sites are near sea (probably due to safety reasons) but not too far from roads and railroads.

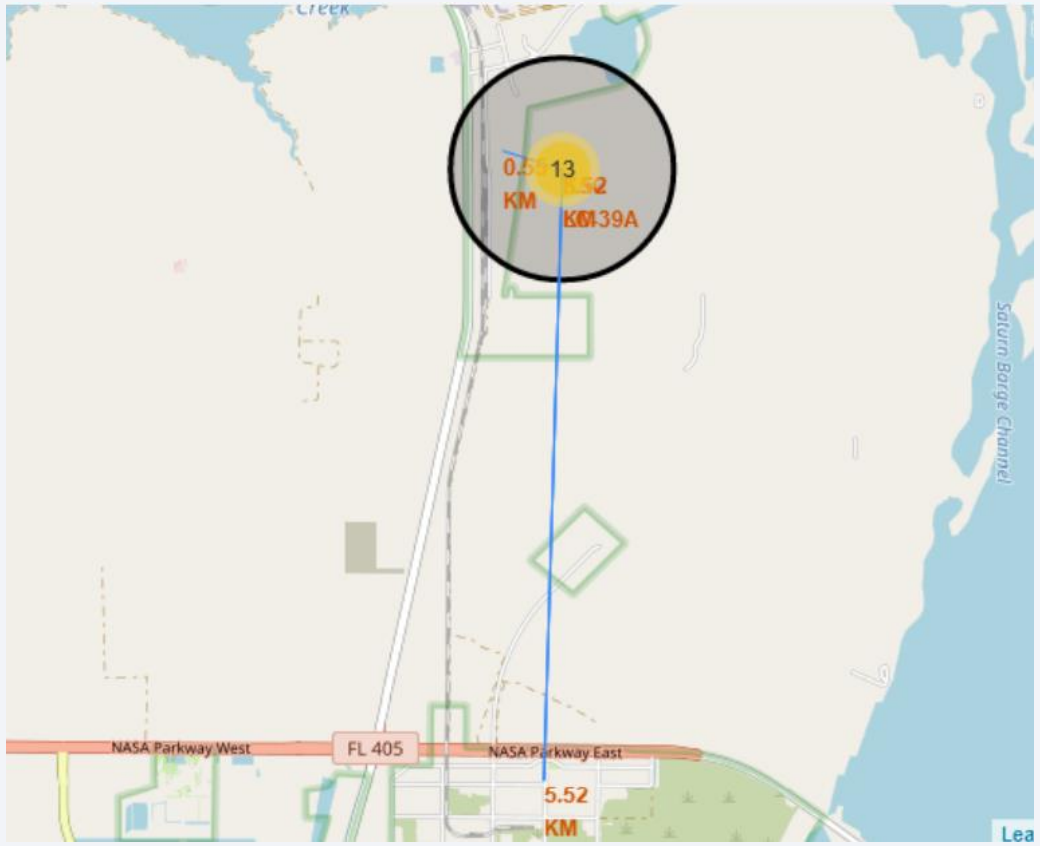
Launch Outcomes by Launch Site

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

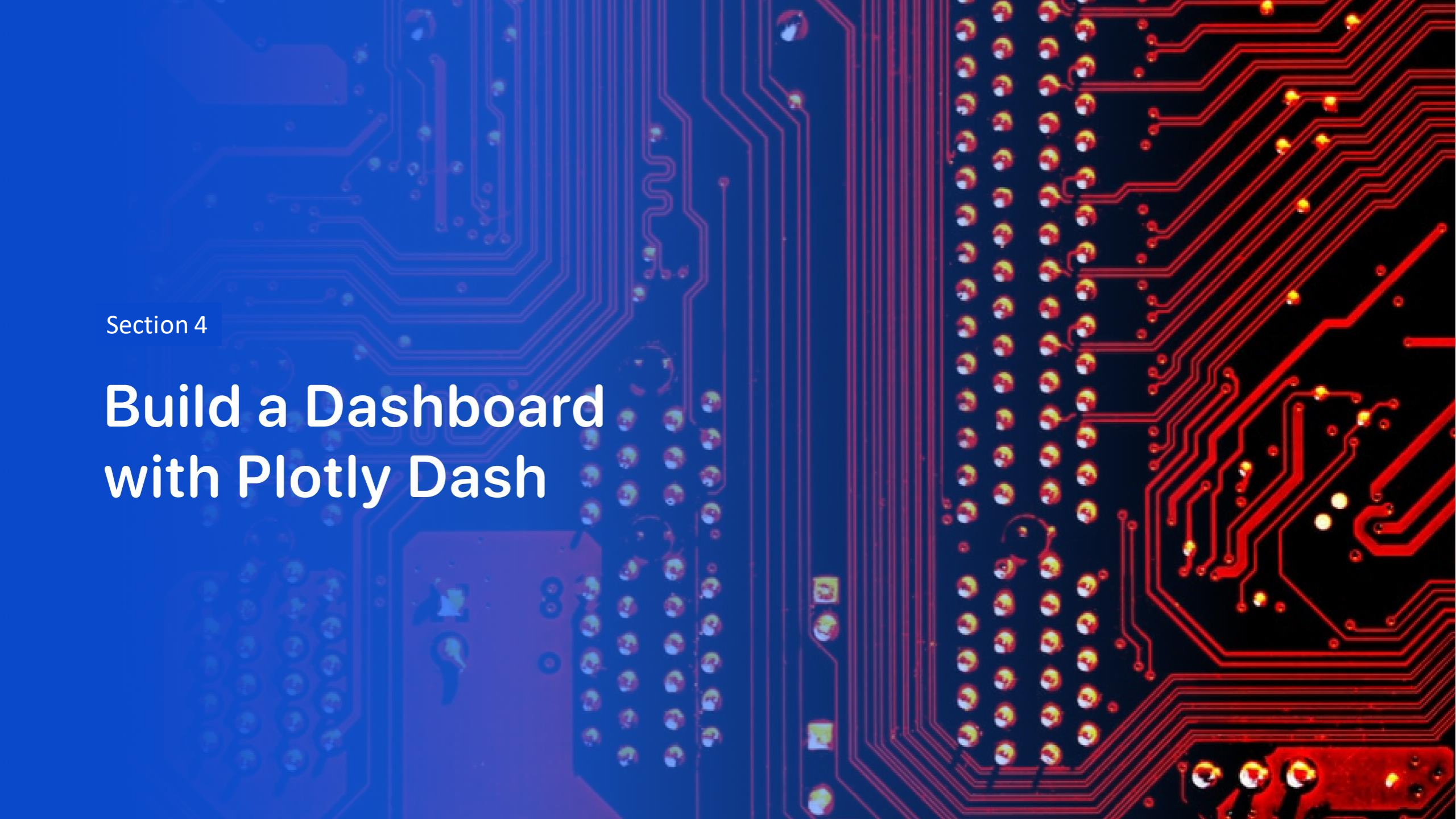


- Green markers indicate successful launches and red ones indicate failure.

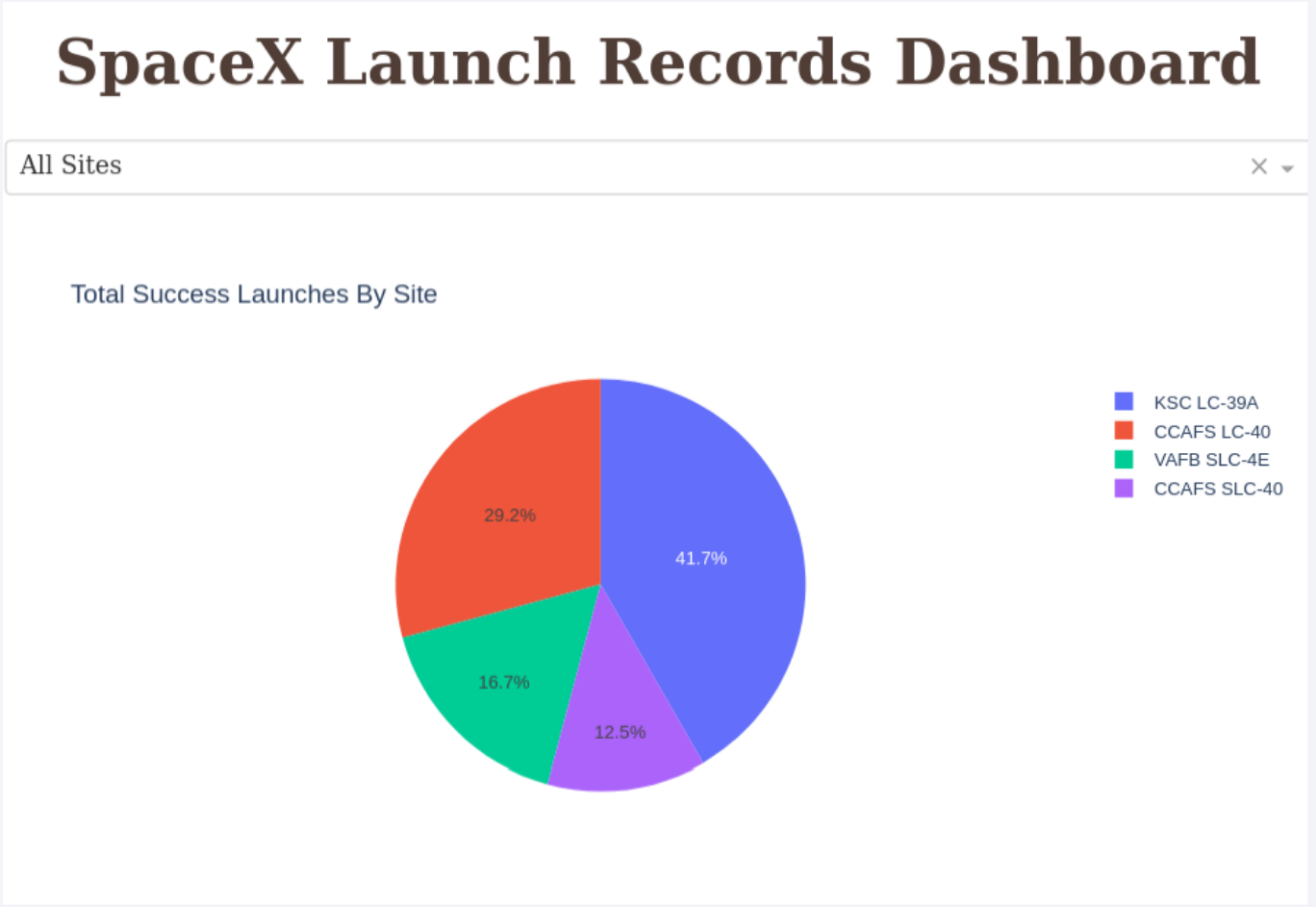
Logistics and Safety



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

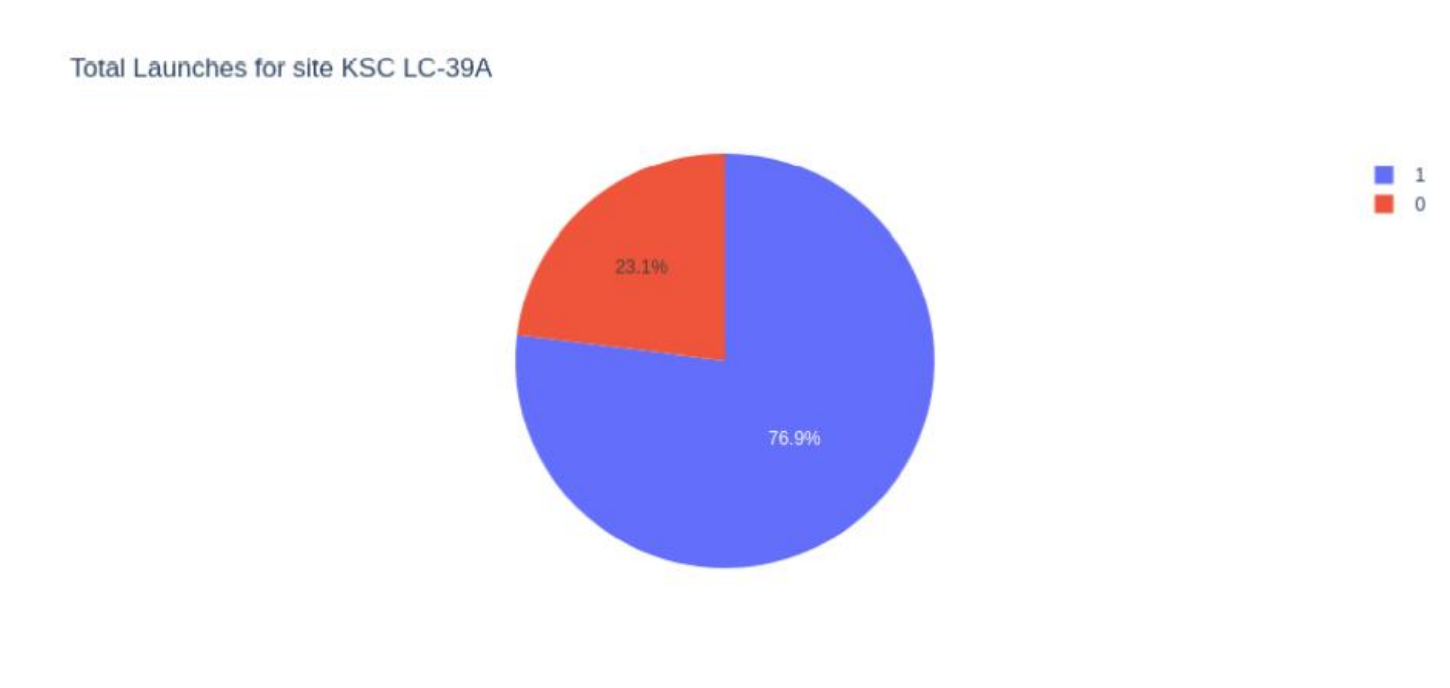

Section 4
Build a Dashboard with Plotly Dash

Successful launches by Launch Site



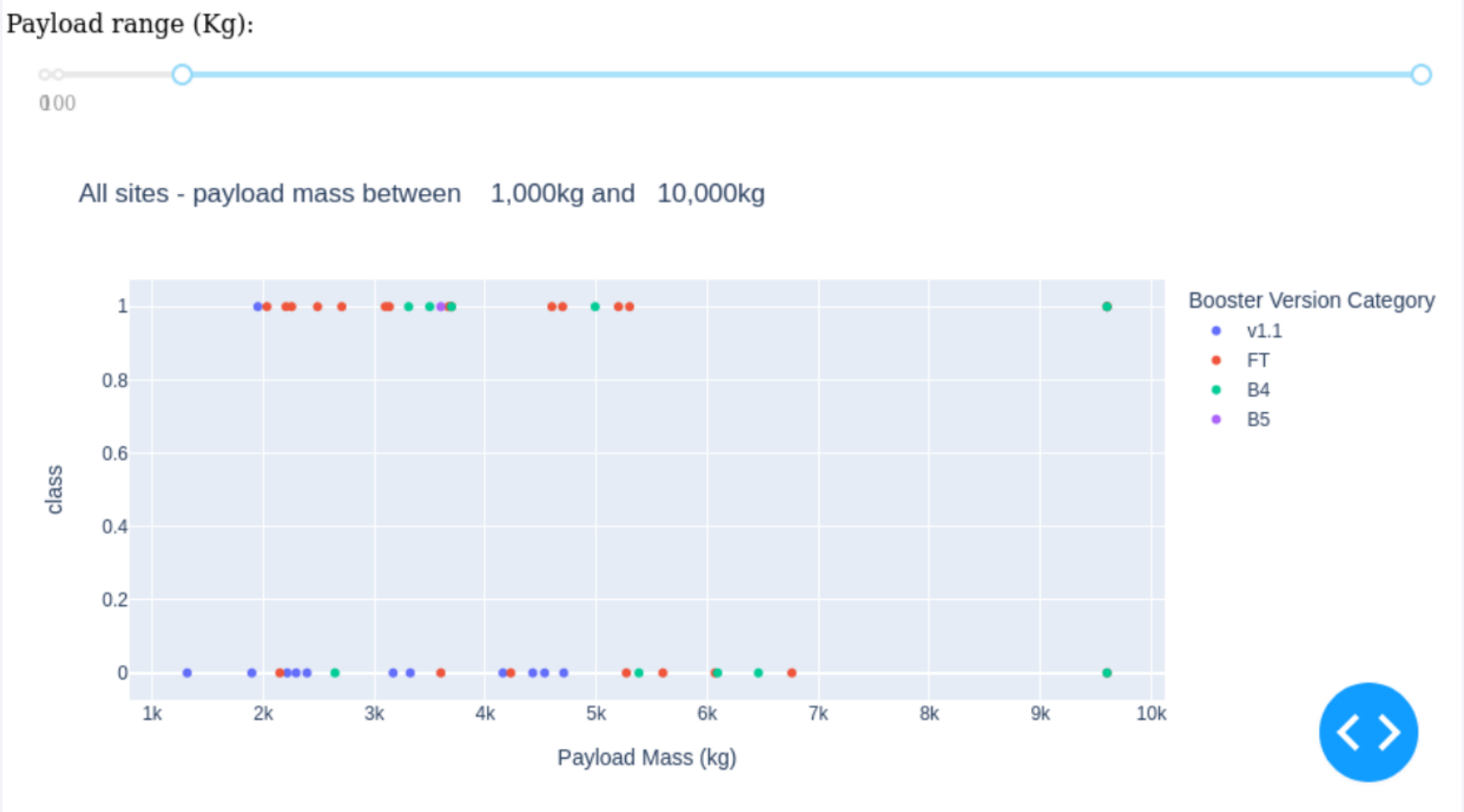
- The Launching Site seems to be a very important factor for the success of the mission.

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.

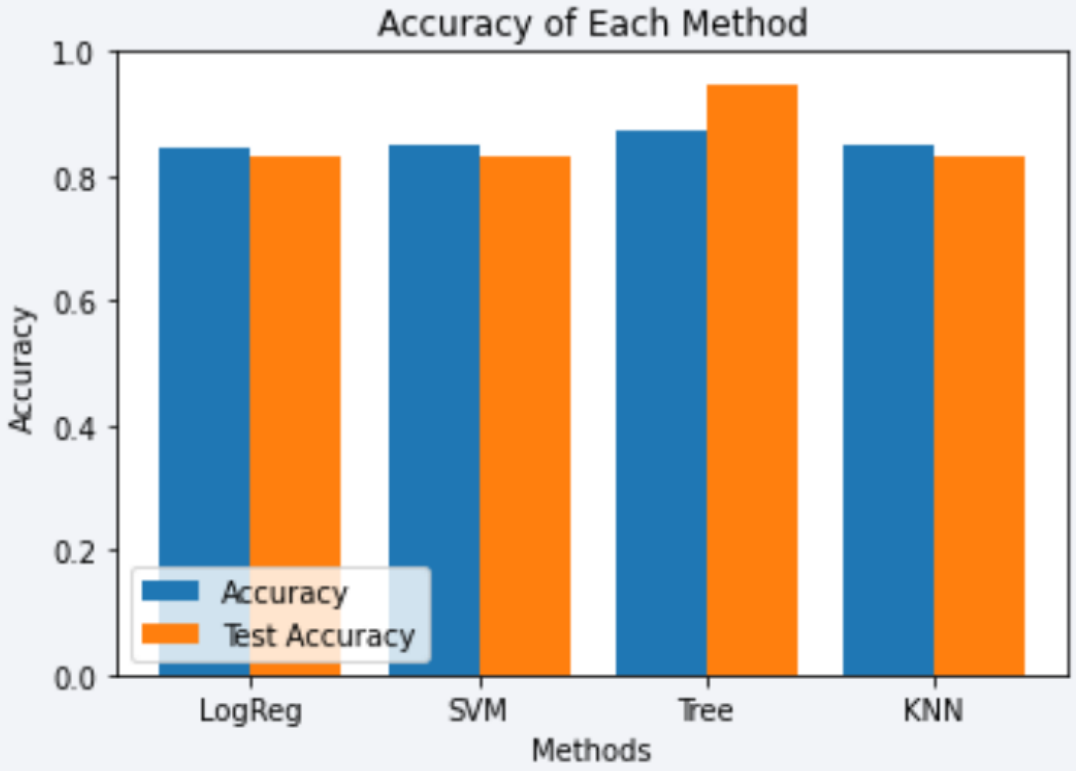
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four classification models were tested, and their accuracies are plotted on the right.

- 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 (12) and true negative(5) 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 the mission outcomes are successful, successful landing outcomes seem to improve over time, according to the evolution of processes and rockets
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

All the relevant jupyter notebooks with the Python code were already included in the previous slides, but the full list is going to be show below:

- **SpaceX API:** https://github.com/efuentesrico/IMB_Coursee/blob/main/1.SpaceX_Data_Collection_with_API.ipynb
- **Webscraping:** https://github.com/efuentesrico/IMB_Coursee/blob/main/2.Data_Collection_with_Webscraping.ipynb
- **Data wrangling:** https://github.com/efuentesrico/IMB_Coursee/blob/main/3.Data_Wrangling.ipynb
- **EDA with SQL:** https://github.com/efuentesrico/IMB_Coursee/blob/main/4.EDA_with_SQL.ipynb
- **EDA with Visualization:** https://github.com/efuentesrico/IMB_Coursee/blob/main/5.EDA_with_Data_Visualization.ipynb
- **Maps:** https://github.com/efuentesrico/IMB_Coursee/blob/main/6.Interactive_Visual_Analytics_with_Folium.ipynb
- **ML Prediction:** https://github.com/efuentesrico/IMB_Coursee/blob/main/7.Machine_Learning_Prediction.ipynb
- **Dashboard:** https://github.com/efuentesrico/IMB_Coursee/blob/main/SpaceX_Dash_App.py

Thank you!