Winning Space Race with Data Science

<Ijaz Ullah>
<2022-02-16>

Outline

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

Executive Summary

The following methodologies were used to analyze data

- Data collection, which involved using web scraping and the SpaceX API to gather information.
- Exploratory data analysis (EDA), which included tasks such as data wrangling, data visualization, and interactive visual analytics.
- Machine learning prediction, which helped to identify the most important features for predicting launch success.

Summary

Through these methodologies, valuable data was collected from public sources, and EDA was used to determine which features were most predictive of launch success. Additionally, the machine learning prediction model was able to identify the key characteristics that are important for optimizing launch opportunities, leveraging all of the 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/)

Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches), using web scraping technics.

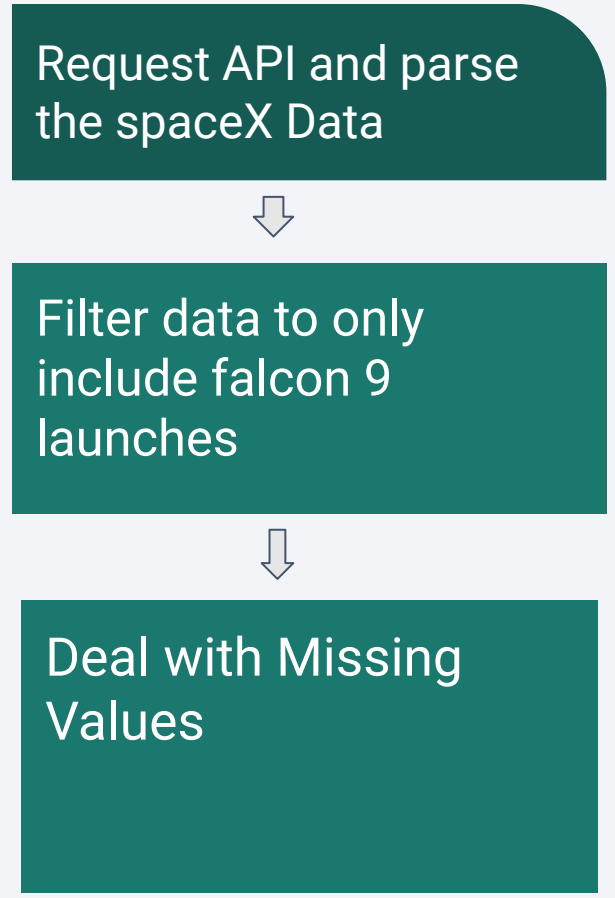
Data Collection – SpaceX API

- SpaceX offers a public API from

  where data can be obtained and

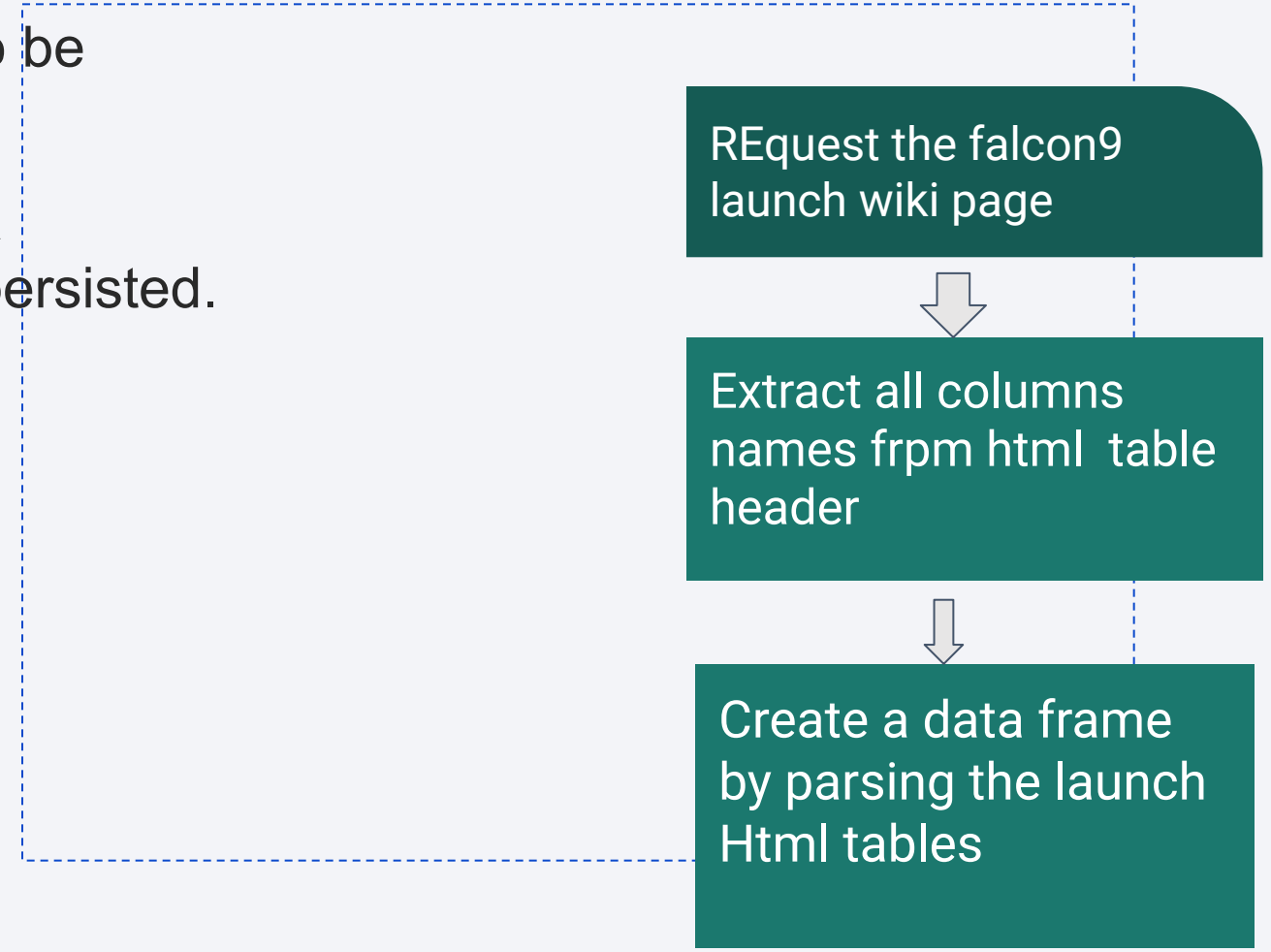
  then used;

- This API was used according to the flowchart beside and then data is persisted
- github https://github.com/ijazkhan101/Applied-Data-Science-Capstone/blob/main/Week%201/Data%20collection%20with%20Web%20Scraping/webscraping.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.

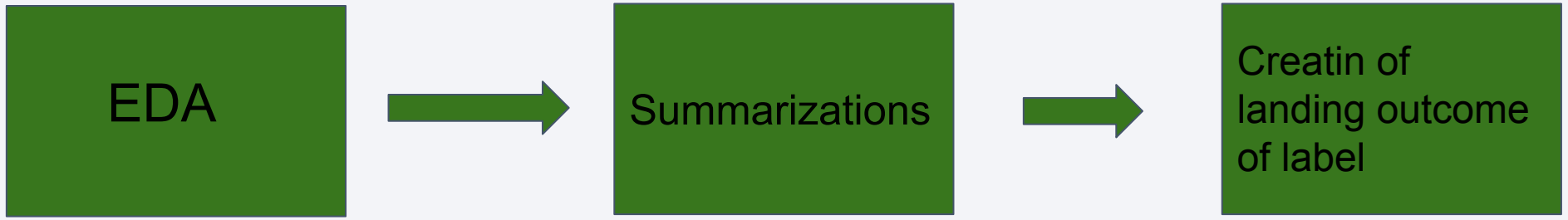


Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.

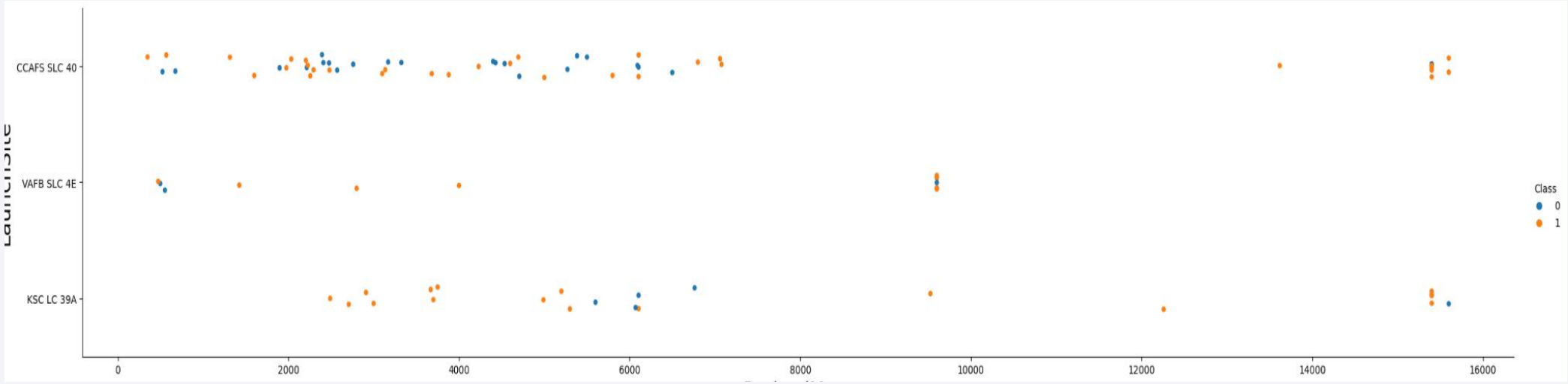
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.

- Finally, the landing outcome label was created from Outcome column.



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



code link
https://github.com/ijazkhan101/Applied-Data-Science-Capstone/blob/main/Week%202/Exploratory%20Data%20Analysis/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; 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.

code : github :
https://github.com/ijazkhan101/Applied-Data-Science-Capstone/blob/main/Week%202/Exploratory%20Data%20Analysis/Eda%20with%20sql.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 code : https://github.com/ijazkhan101/Applied-Data-Science-Capstone/blob/main/Week%203/interactive_dashboard.ipynb

Build a Dashboard with Plotly Dash

The following graphs and plots were used to visualize data

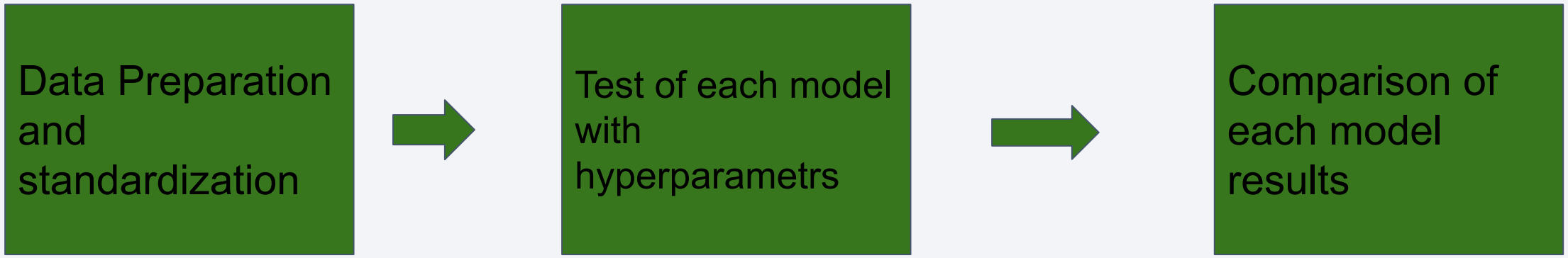
- Percentage of launches by site
- Payload range
- This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

code github :
https://github.com/ijazkhan101/Applied-Data-Science-Capstone/blob/main/Week%203/interactive_dashboard.ipynb

Predictive Analysis (Classification)

Four classification models were compared: logistic regression, support vector

machine, decision tree and k nearest neighbors.

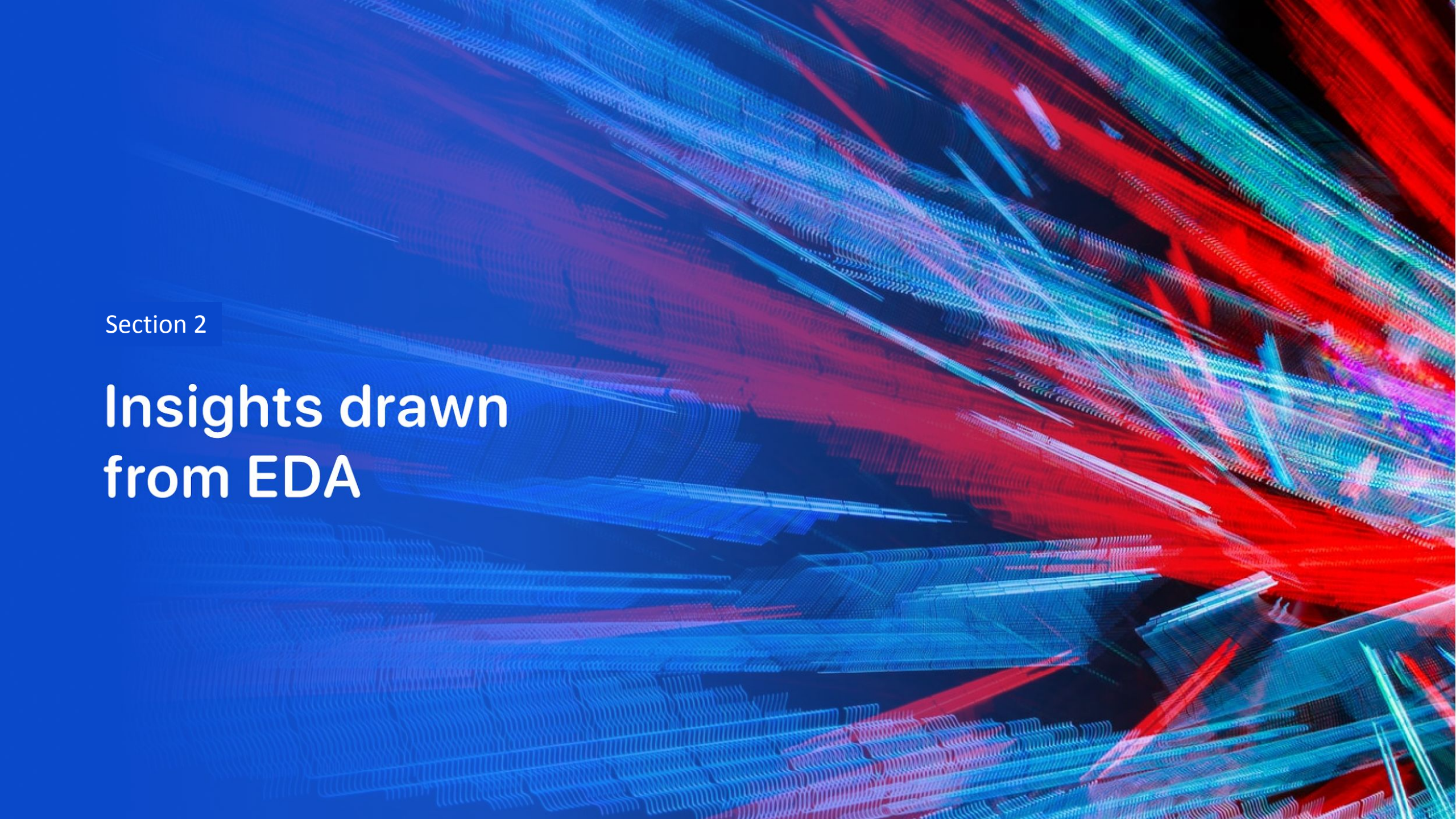


Results

- 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

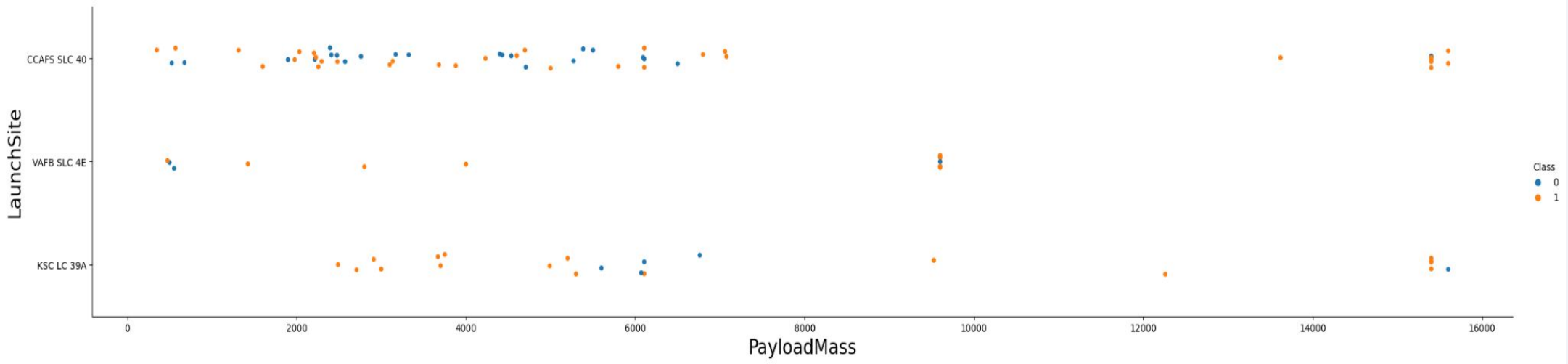
Section 2

Insights drawn from EDA



Flight Number vs. Launch Site

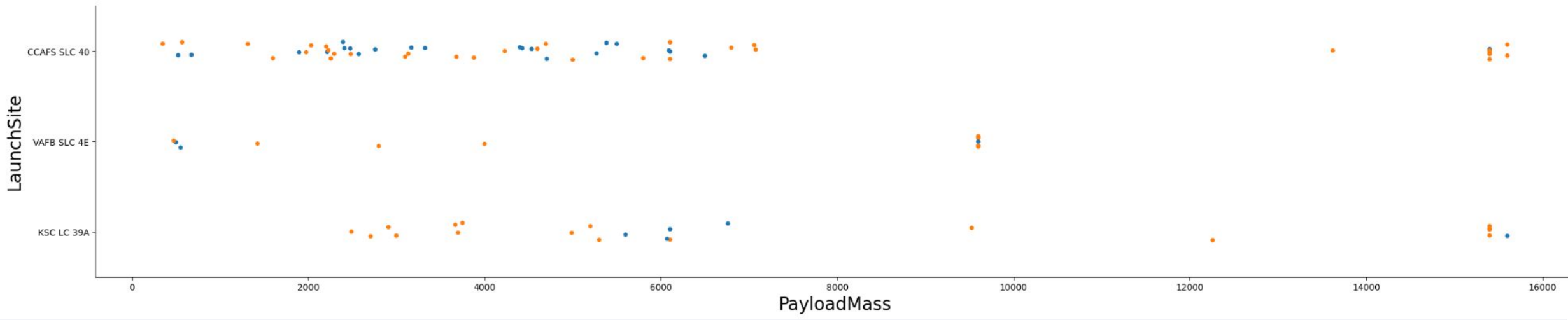
- According to the plot above, it’s possible to verify that the best launch site
- nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- • In second place VAFB SLC 4E and third place KSC LC 39A;
- • It’s also possible to see that the general success rate improved over time



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.



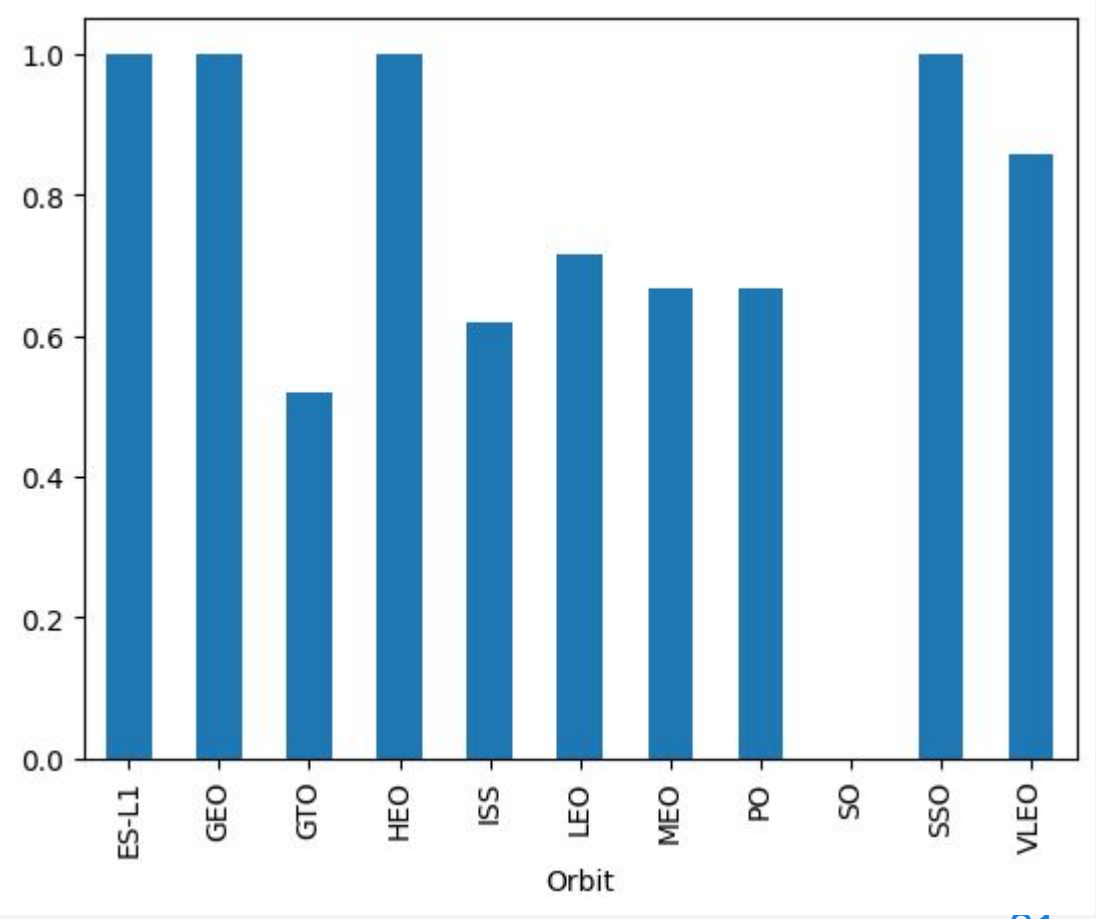
Success Rate vs. Orbit Type

The biggest success rates happens to orbits

- ES-L1;
- •GEO;
- HEO; and
- SSO.

Followed by:

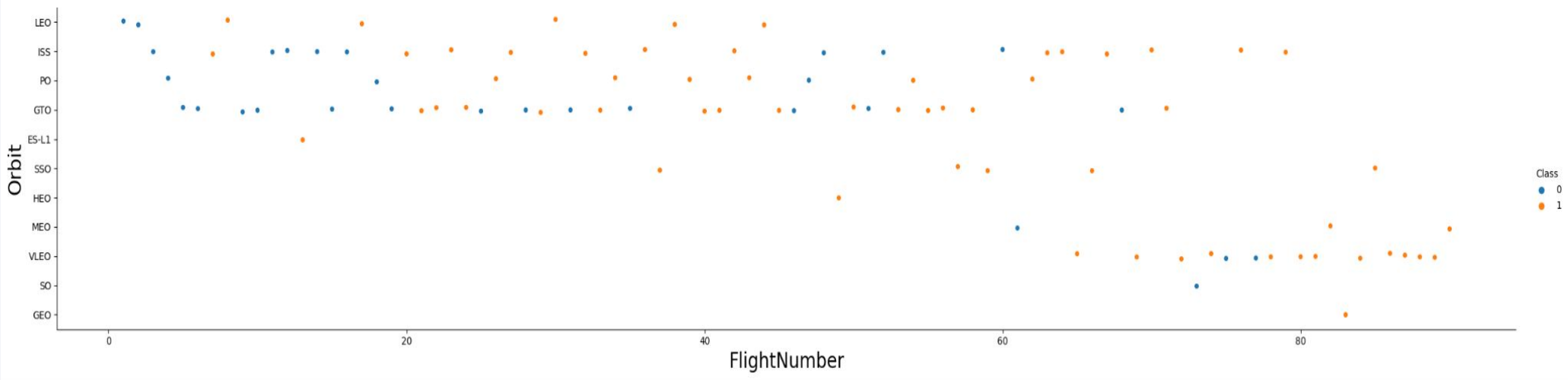
- VLEO (above 80%); and
- LFO (above 70%).



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 it frequency.

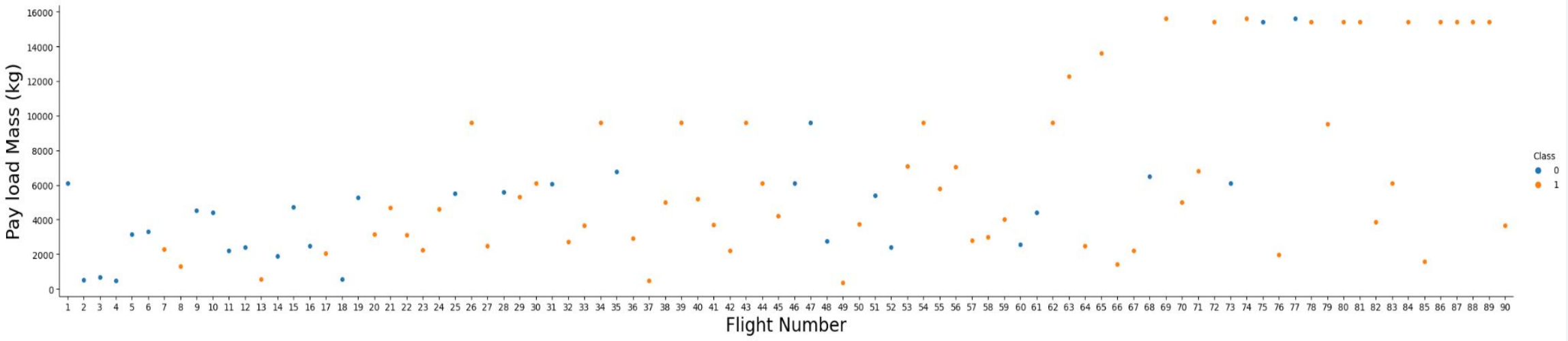


Payload vs. Orbit Type

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;

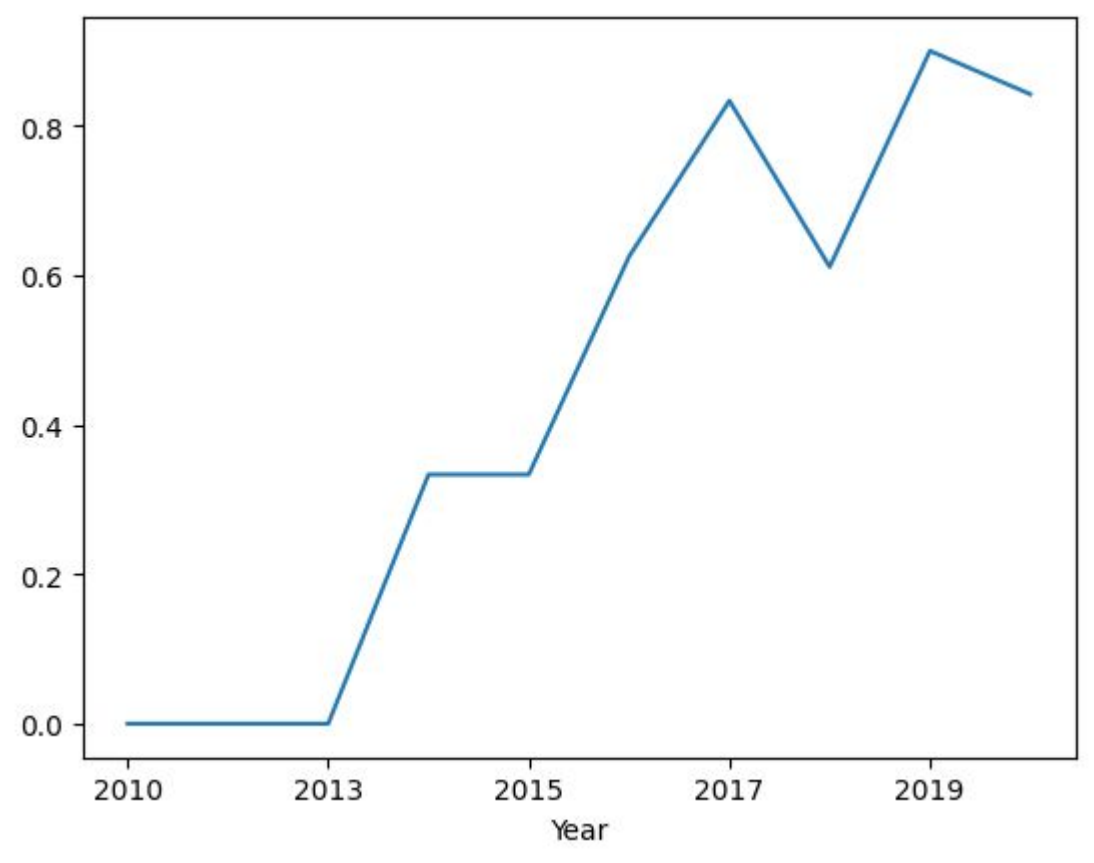
There are few launches to the orbits SO and GEO.



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

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

Launch Sites
CCAFS LC-40 CCAFS SLC-40 KSC LC-39A VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

5 records where launch sites begin with `CCA`:

Here we can see five samples of Cape Canaveral launches.

13]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Outcome	Flights	GridFins	Reused	Legs	Block	...	Serial_B1048	Serial_B1049	Serial_B1050
0	1	2010-06-04	Falcon 9	6104.959412	None None	1	False	False	False	1.0	...	0	0	0
1	2	2012-05-22	Falcon 9	525.000000	None None	1	False	False	False	1.0	...	0	0	0
2	3	2013-03-01	Falcon 9	677.000000	None None	1	False	False	False	1.0	...	0	0	0
3	4	2013-09-29	Falcon 9	500.000000	False Ocean	1	False	False	False	1.0	...	0	0	0
4	5	2013-12-03	Falcon 9	3170.000000	None None	1	False	False	False	1.0	...	0	0	0

5 rows × 86 columns

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) : 111.268

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 (KG) : 2.928

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 12/22/2015

Min Date : 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

- Number of successful and failure mission outcomes:
- Grouping mission outcomes and counting records for each group led us to the summary above.

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

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass
- 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

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

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

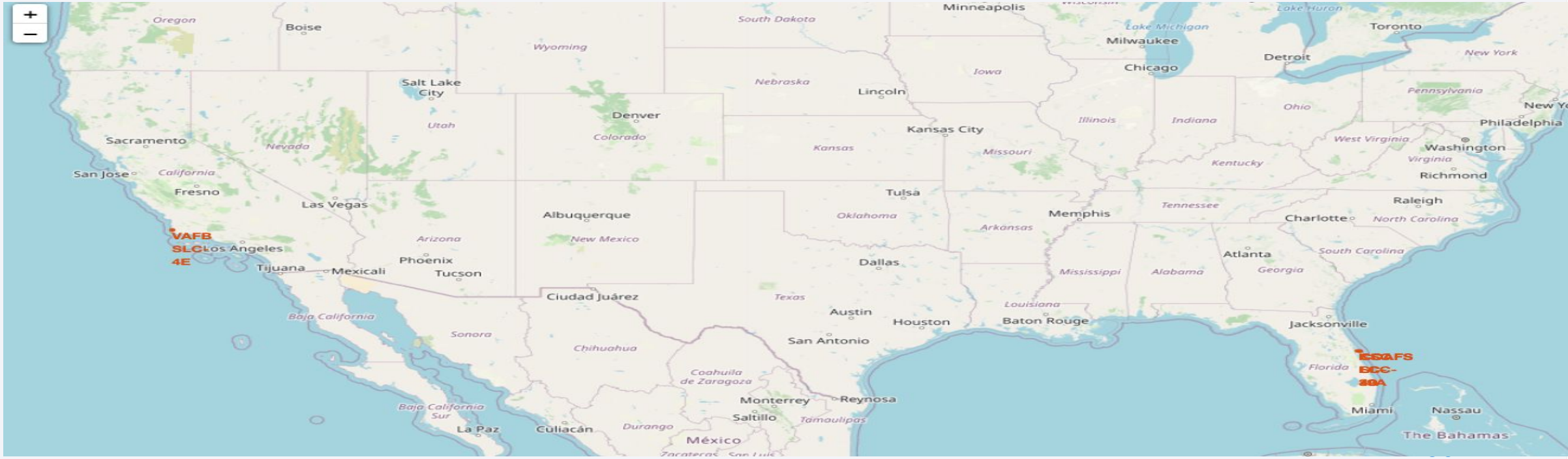
- Ranking of all landing outcomes between the date 2010-06-04 and 2017-03-20:
- This view of data alerts us that “No attempt” must be taken in account

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


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

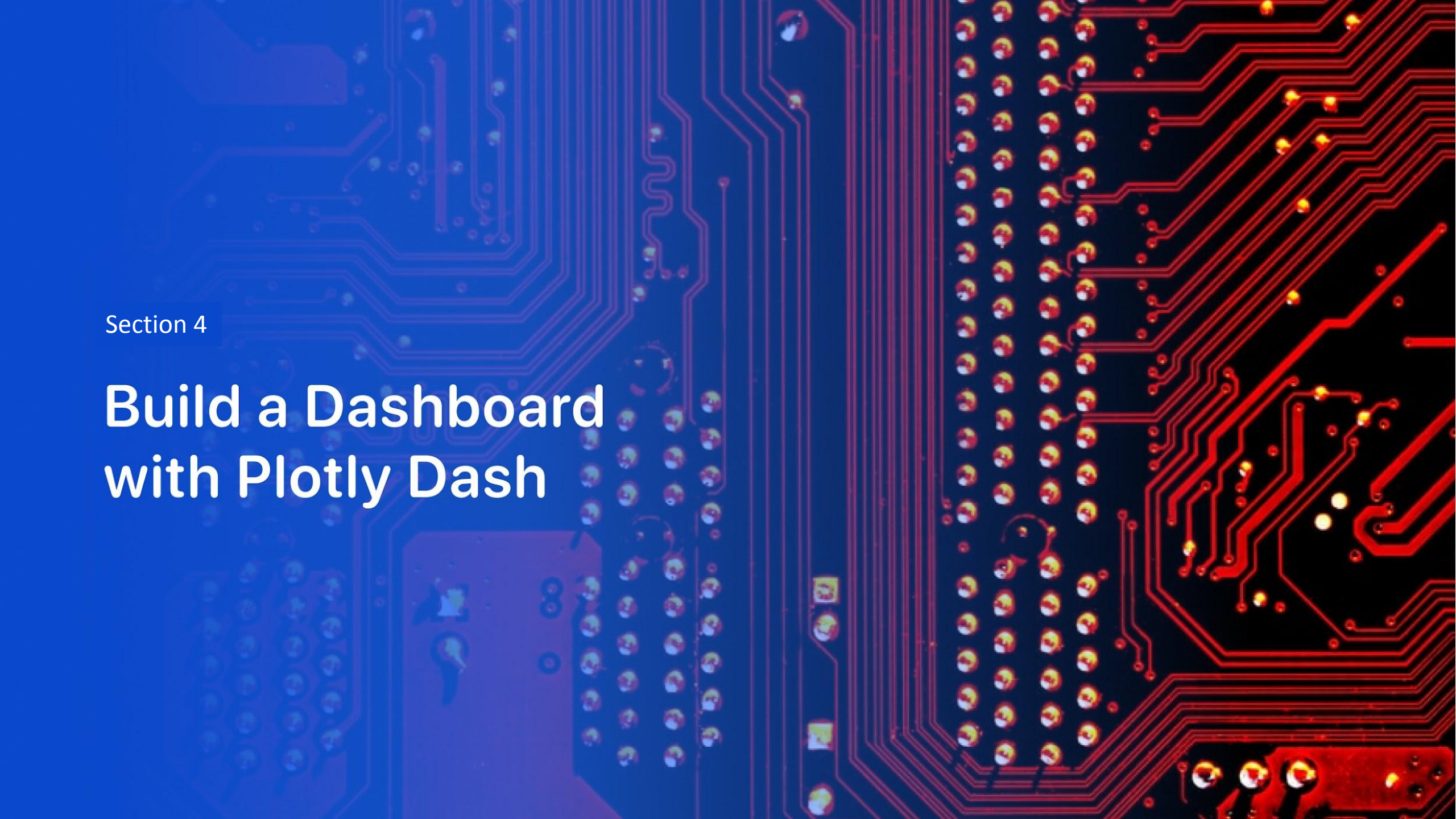
Example of KSC LC-39A launch site launch outcomes

- 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

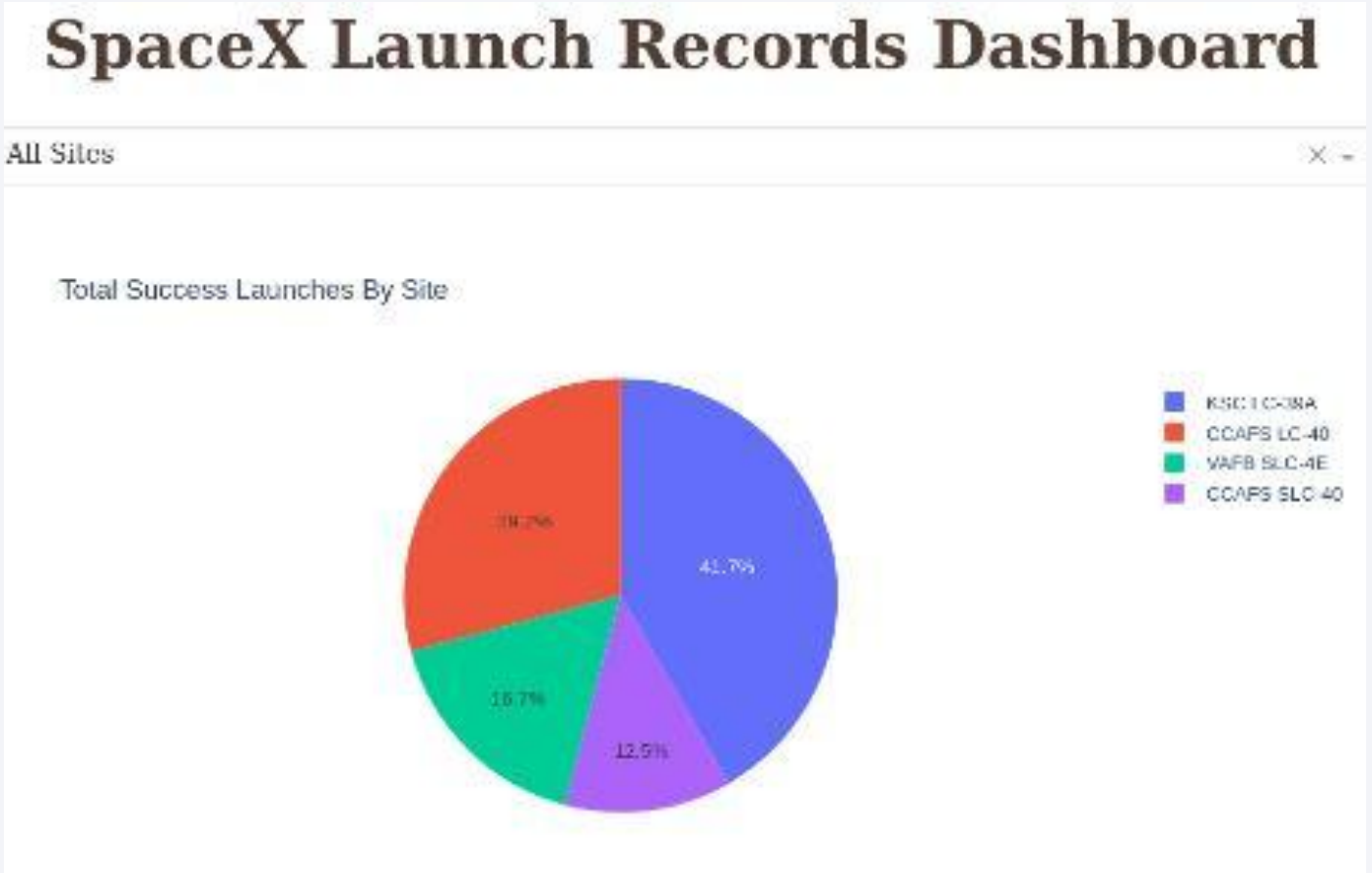
relatively far from inhabited areas.


Section 4
Build a Dashboard with Plotly Dash

Successful Launches by Site

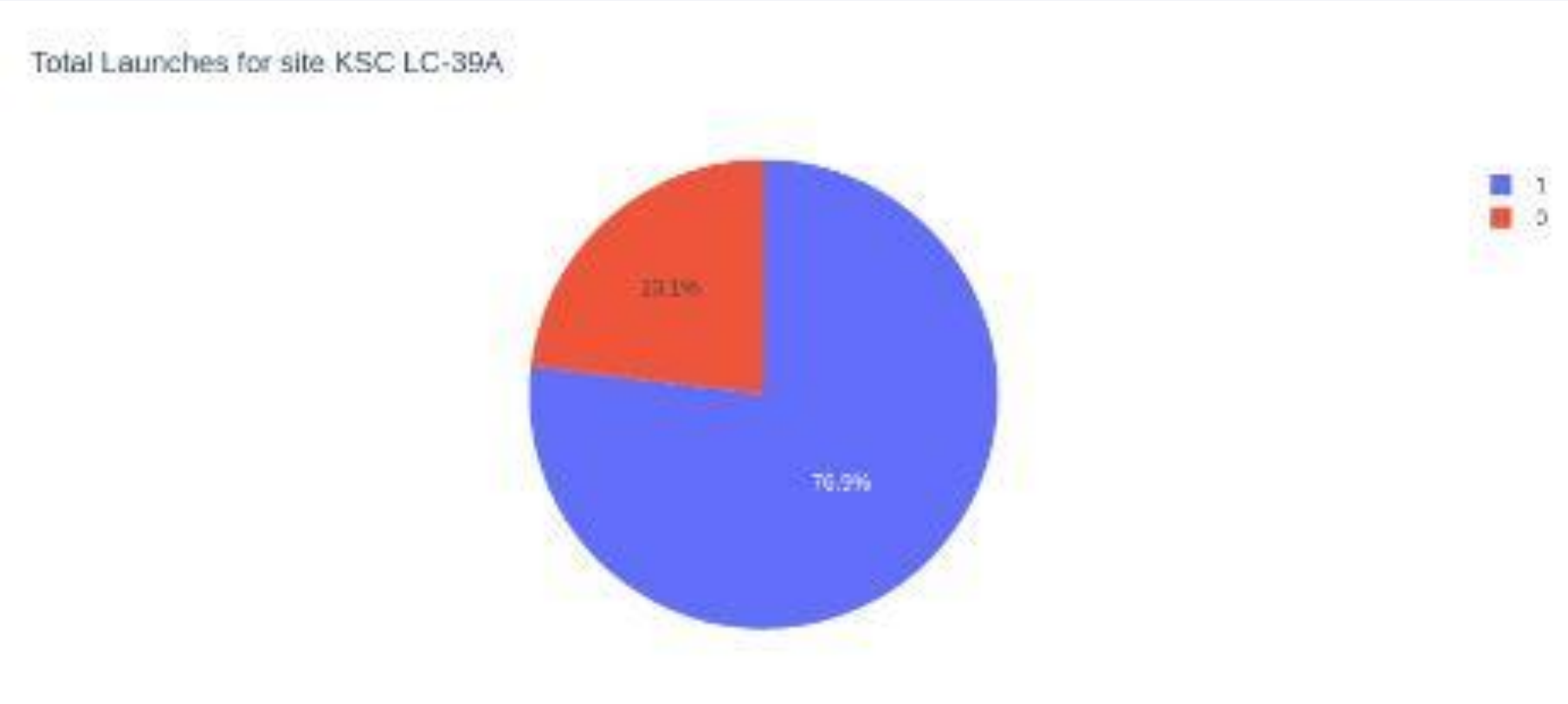
The place from where launches are done seems to be a very important factor of

success of missions.



Launch Success Ratio for KSC LC-39A

76.9% of launches are successful in this site.



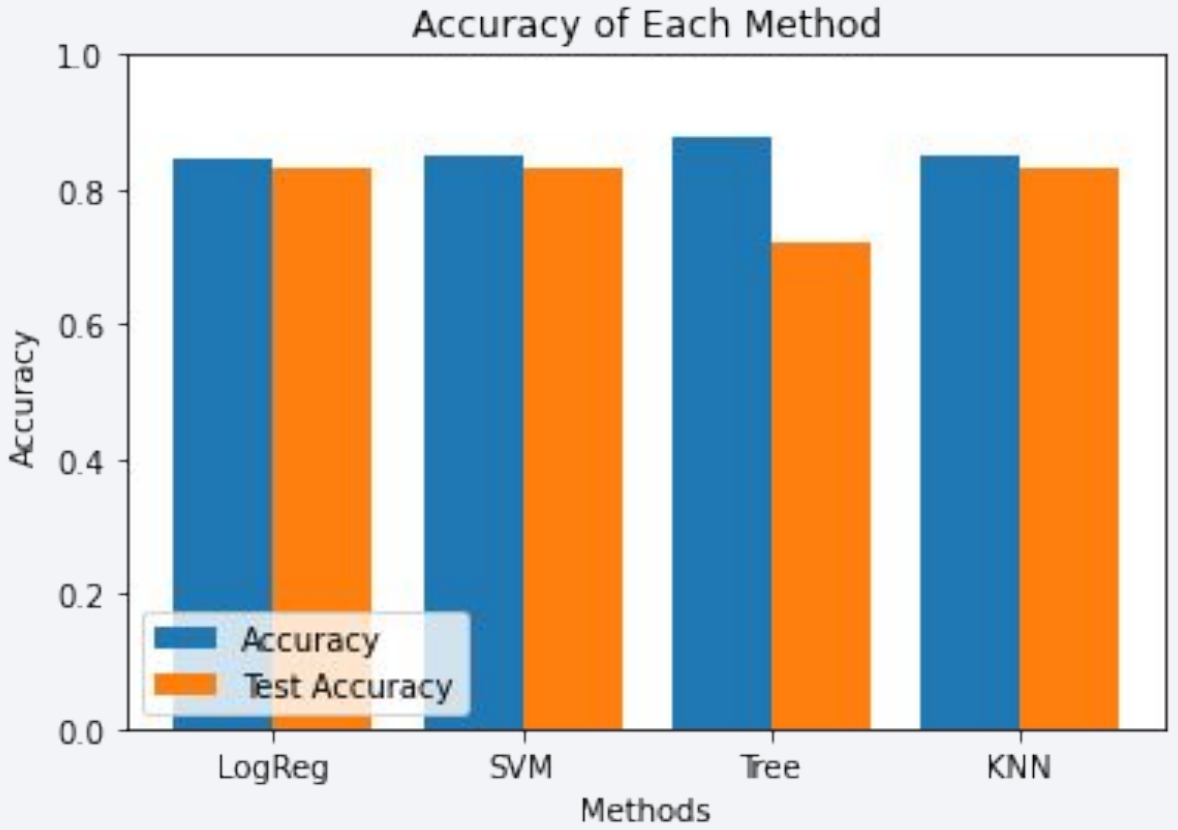
Section 5

Predictive Analysis (Classification)

Classification Accuracy

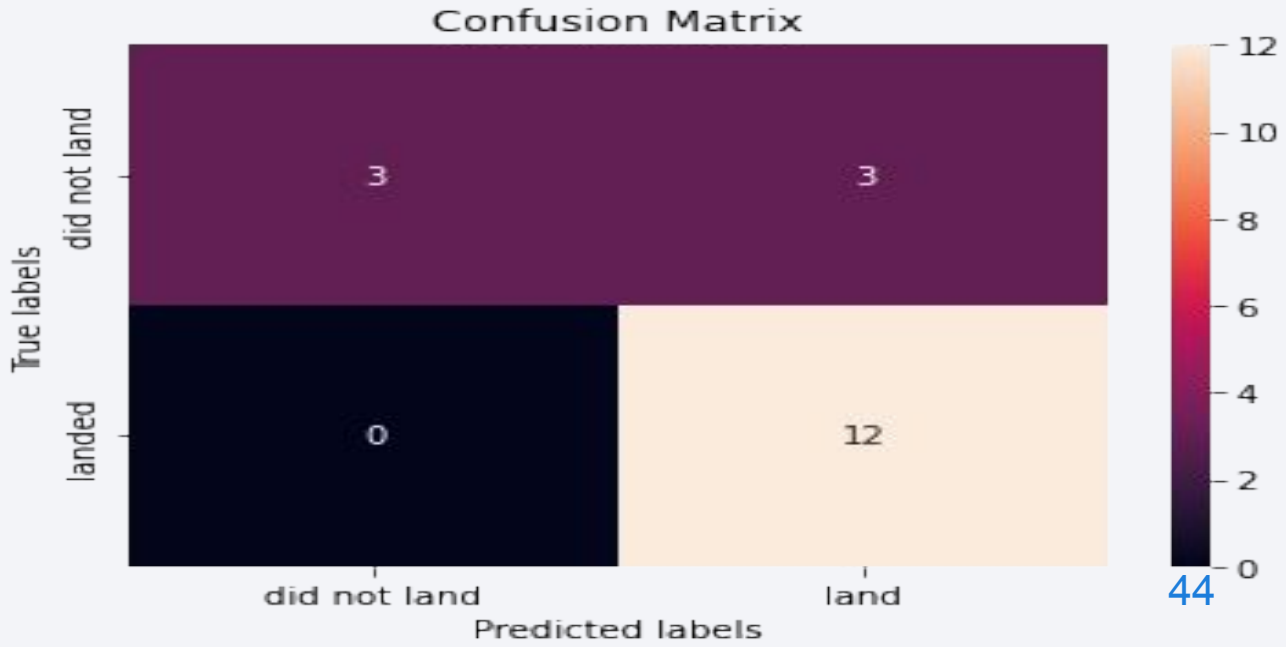
Four classification models were tested, and their accuracies are plotted beside;

The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%



Confusion Matrix 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 of processes and rockets;
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank you!