

Winning Space Race with Data Science

André Rodrigues Vicente
05/02/2023



Outline

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

Executive Summary

- Summary of methodologies:
 1. Data was collected with Webscrapping and SpaceX API;
 2. Then, Exploratory Data Analysis (EDA) was conveyed, including data wrangling, data visualization and interactive visual analytics;
 3. Finally, Machine Learning Prediction.
- Summary of all results:
 1. It was possible to collect valuable data from public sources;
 2. EDA allowed to identify which features are the best to use in success of launchings' predictions;
 3. Through Machine Learning Prediction it was built the best model to predict which characteristics are important to drive this opportunity the best way, using all collected data.

Introduction

- Project background and context:
 - The main goal of the project was to evaluate the ability of a new company (SpaceY) to compete with SpaceX
- Problems you want to find answers:
 - The best model to estimate the total cost for launches
 - The best location to make launches.

Section 1

Methodology

Methodology

Executive Summary

1. Data collection methodology:

- Data from SpaceX was obtained via:
 - SpaceX API: <https://api.spacexdata.com/v4/rockets/>
 - WebScraping:
https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

2. Perform data wrangling

- It was created a landing outcome label based on outcome data after summarizing and analyzing features

3. Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

4. Perform interactive visual analytics using Folium and Plotly Dash
5. Perform predictive analysis using classification models
 - Data was preprocessed, then divided in training and testing data sets. An evaluation of 4 different models was done and the accuracy of each model using certain parameters was the criteria to choose the best one.

Data Collection

- Data sets were collected through:
 - SpaceX API: <https://api.spacexdata.com/v4/rockets/>
 - WebScraping:
https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- SpaceX offers a public API, where data can be extracted, as seen on the flowchart

Request API and parse the SpaceX launch data



Filter data to include just Falcon 9 launches



Deal with missing values

- Source: [Applied-Capstone-Project/Data Collection API.ipynb at main · andre vicente97/Applied-Capstone-Project \(github.com\)](https://github.com/andre vicente97/Applied-Capstone-Project)

Data Collection - Scraping

- Data also can be obtained via WebScraping from Wikipedia

Request Falcon9
Launch Wiki page



Extract all
column/variable
names from the
HTML table header

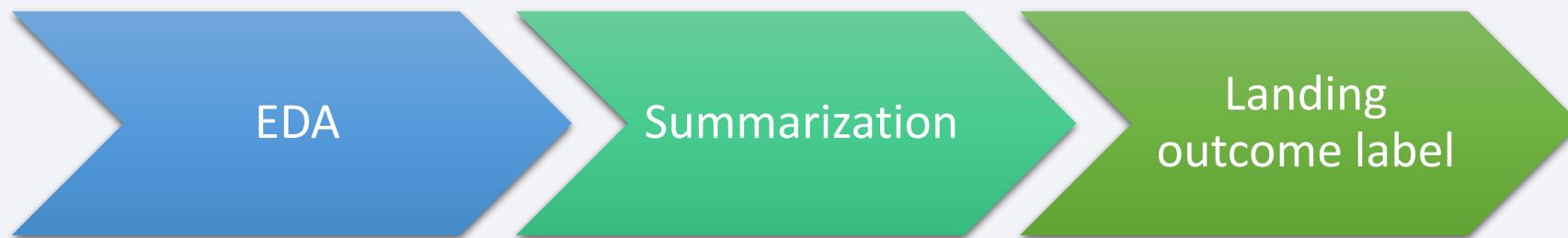


Create a dataframe
by parsing the
launch HTML table

- Source: [Applied-Capstone-Project/Web scrapng.ipynb](https://github.com/andrevicente97/Applied-Capstone-Project/blob/main/Applied-Capstone-Project/Web%20scraping.ipynb) at main · andrevicente97/Applied-Capstone-Project (github.com)

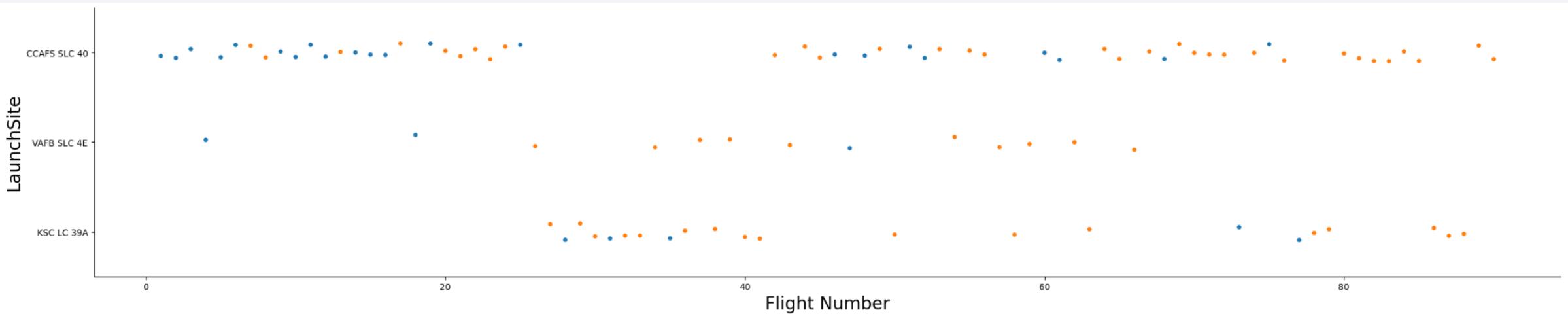
Data Wrangling

1. Exploratory Data Analysis (EDA);
2. Summarization of launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type;
3. Creation of Landing outcome label.



EDA with Data Visualization

- Scatterplots and barplots were used to visualize the data and the relationships between the features:
 - Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



EDA with SQL

- The SQL queries performed were:
 - 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:** <https://github.com/andre-vicente97/Applied-Capstone-Project/blob/main/EDA%20with%20SQL%20lab.ipynb>

Build an Interactive Map with Folium

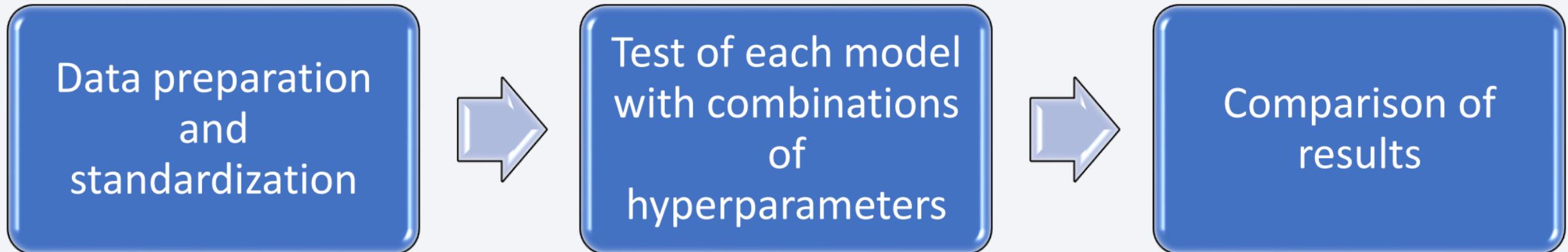
- Markers, circles, lines and marker clusters were created and added to a folium map
 - Markers to indicate launch sites;
 - Circles to indicate highlighted areas around specific coordinates;
 - Marker clusters to indicate groups of events in each coordinate;
 - Lines to indicate distances between 2 coordinates.

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

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



Results

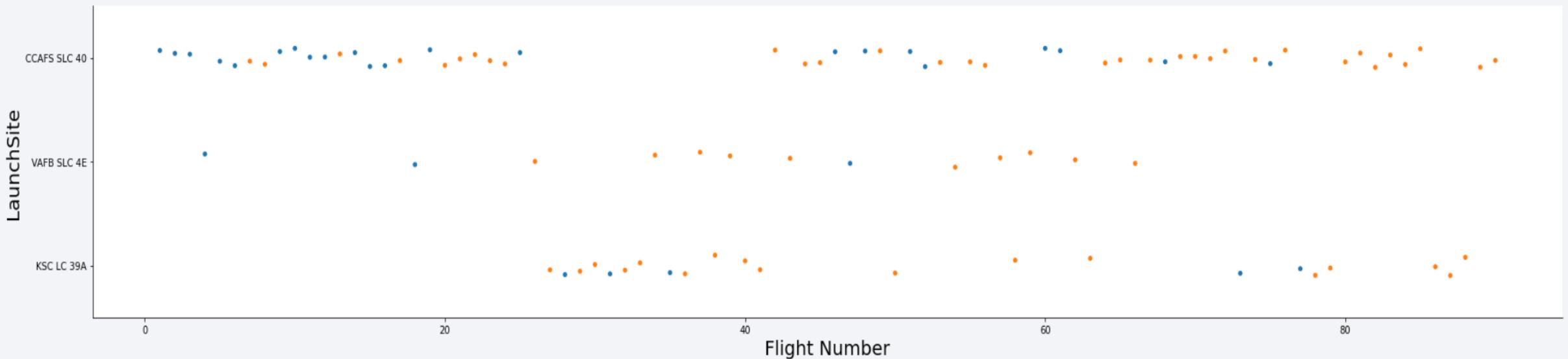
- Exploratory data analysis results:
 - Space X uses 4 different launch sites;
 - The first launches were done to Space X itself and NASA;
 - The average payload of F9 v1.1 booster is 2,928 kg;
 - The first success landing outcome happened in 2015 five years after the first launch;
 - Many Falcon 9 booster 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 better as years passed.

The background of the slide features a complex, abstract digital visualization. It consists of numerous thin, glowing lines that create a sense of depth and motion. The lines are primarily blue and red, with some green and purple highlights. They form a grid-like structure that curves and twists across the frame, resembling a three-dimensional space or a network of data points. The overall effect is futuristic and dynamic.

Section 2

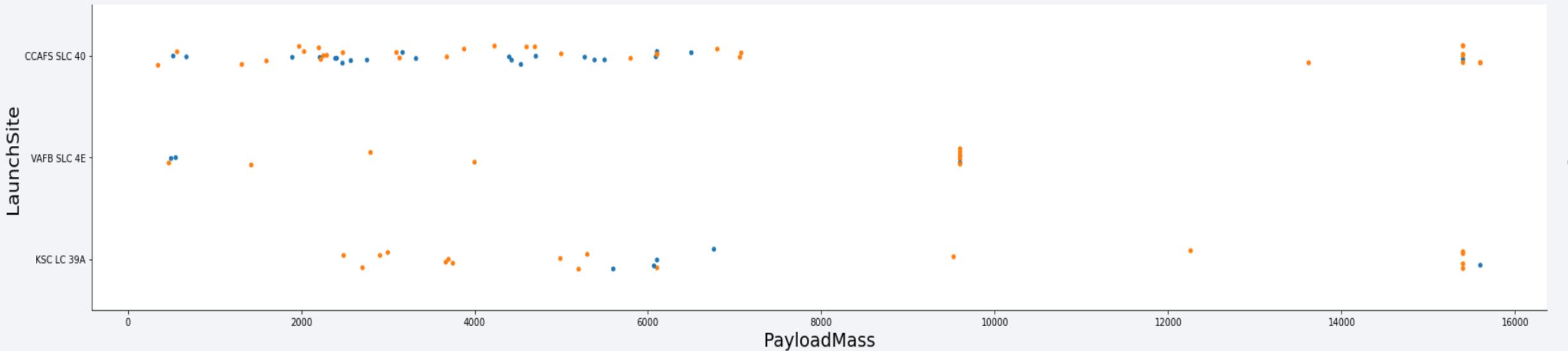
Insights drawn from EDA

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.

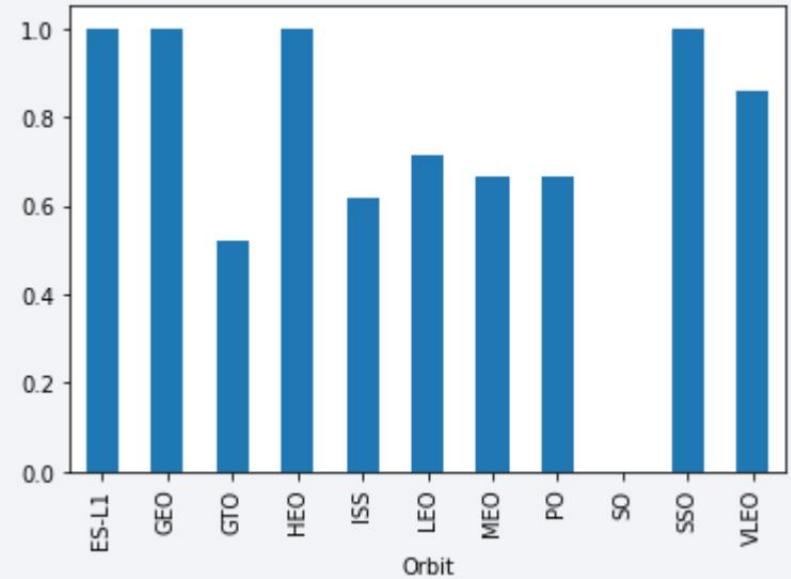
Payload vs. Launch Site



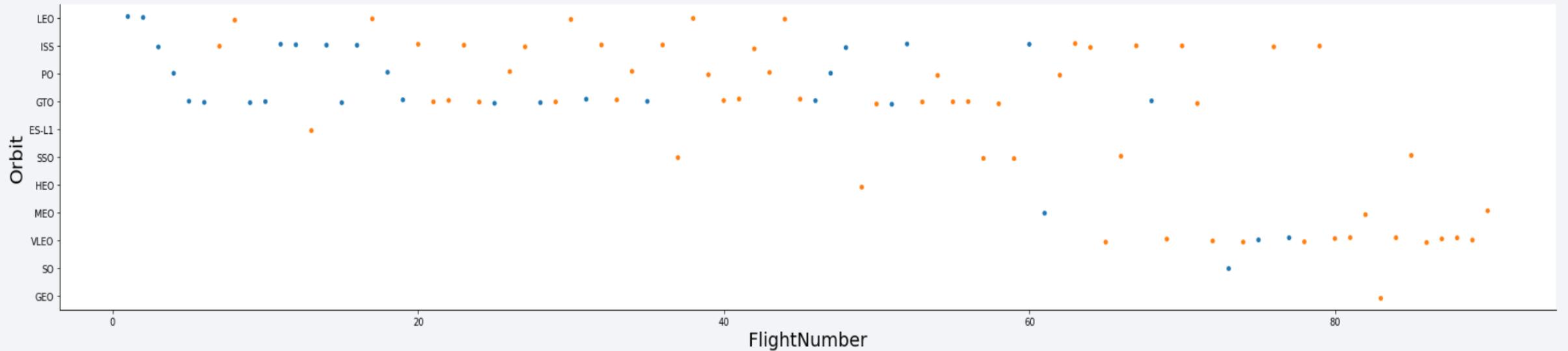
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

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%).

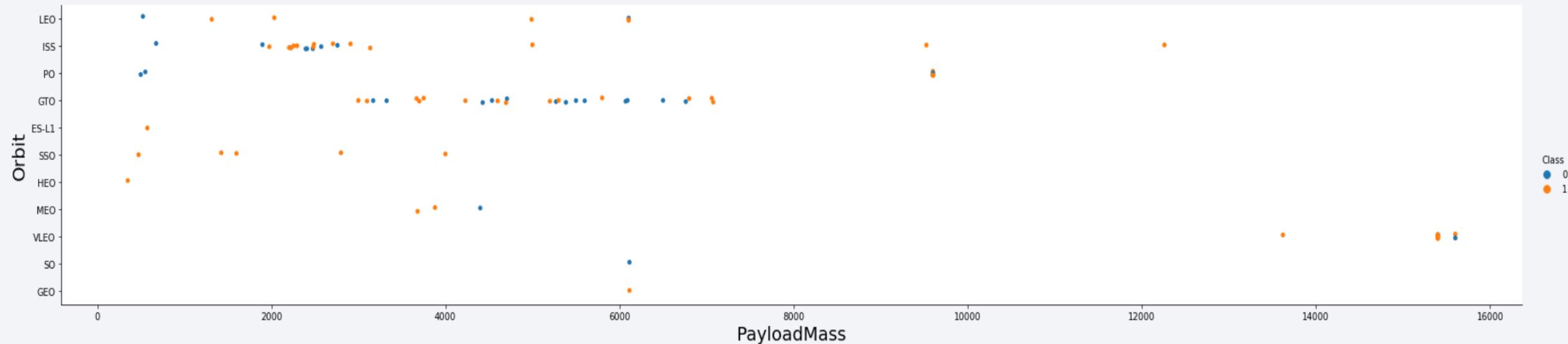


Flight Number vs. Orbit Type



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

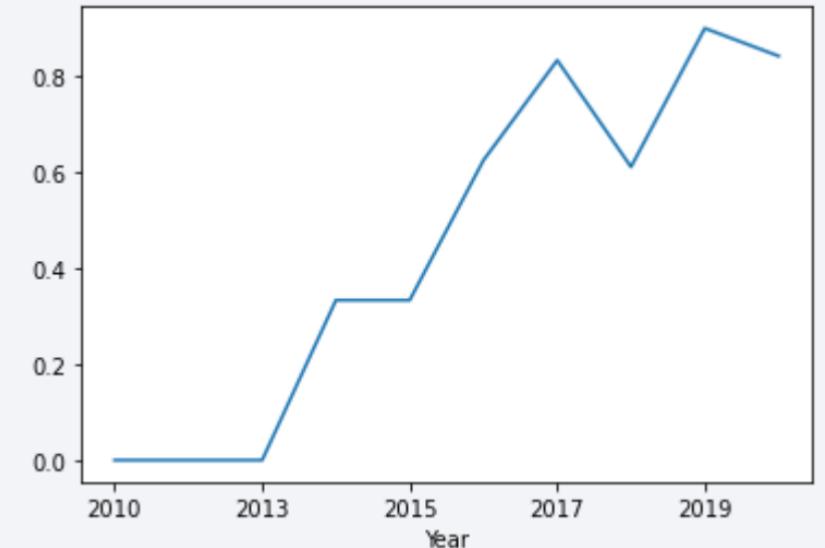
Payload vs. Orbit Type



- 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
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	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 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 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

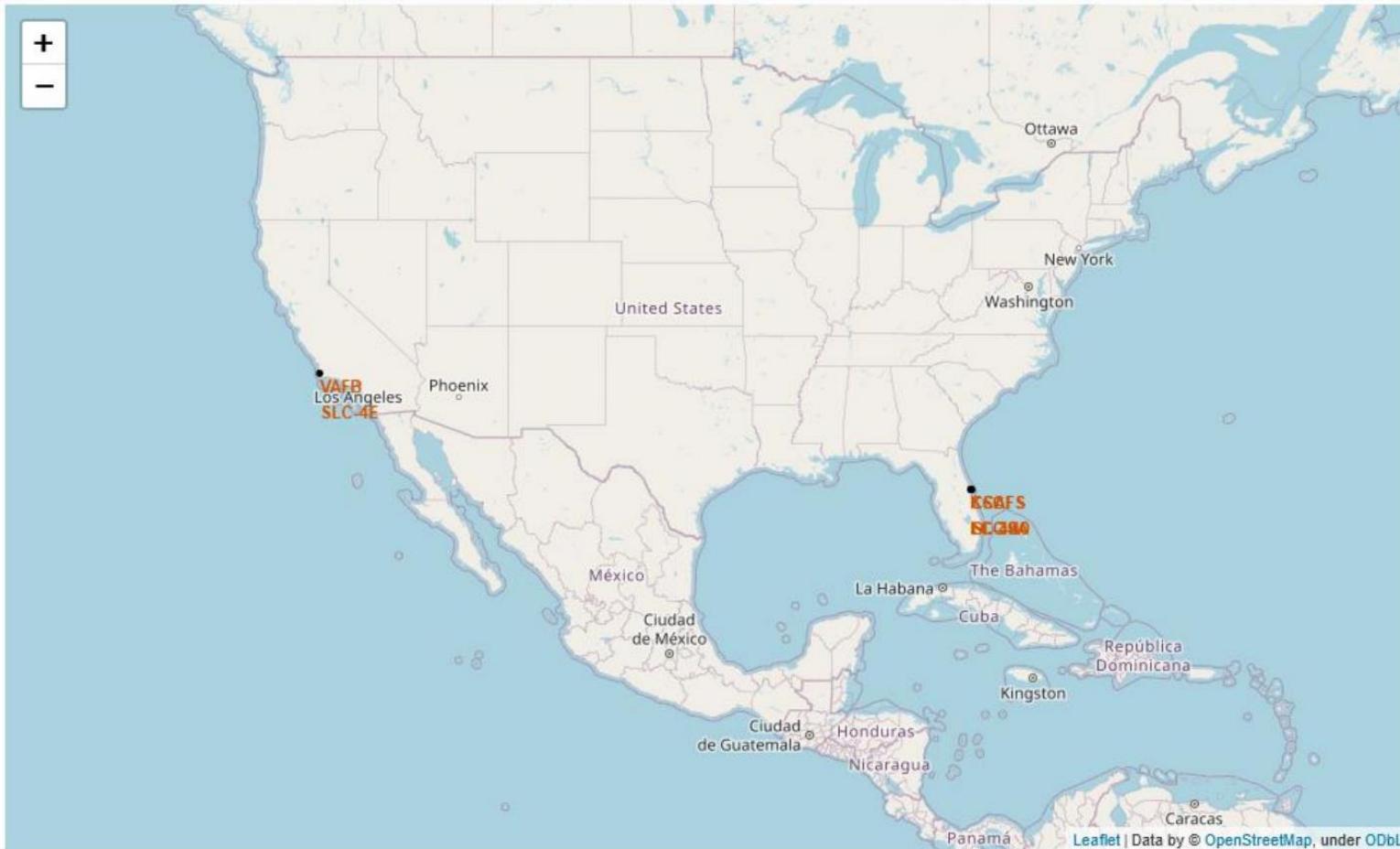
- This view of data alerts us that “No attempt” must be taken in account.

The background of the slide is a photograph taken from space at night. It shows the curvature of the Earth's horizon against a dark blue sky. Numerous glowing yellow and white points represent city lights, concentrated in coastal and urban areas. In the upper right quadrant, there are bright green and yellow bands of light, likely the Aurora Borealis or Australis. The overall atmosphere is dark and mysterious.

Section 3

Launch Sites Proximities Analysis

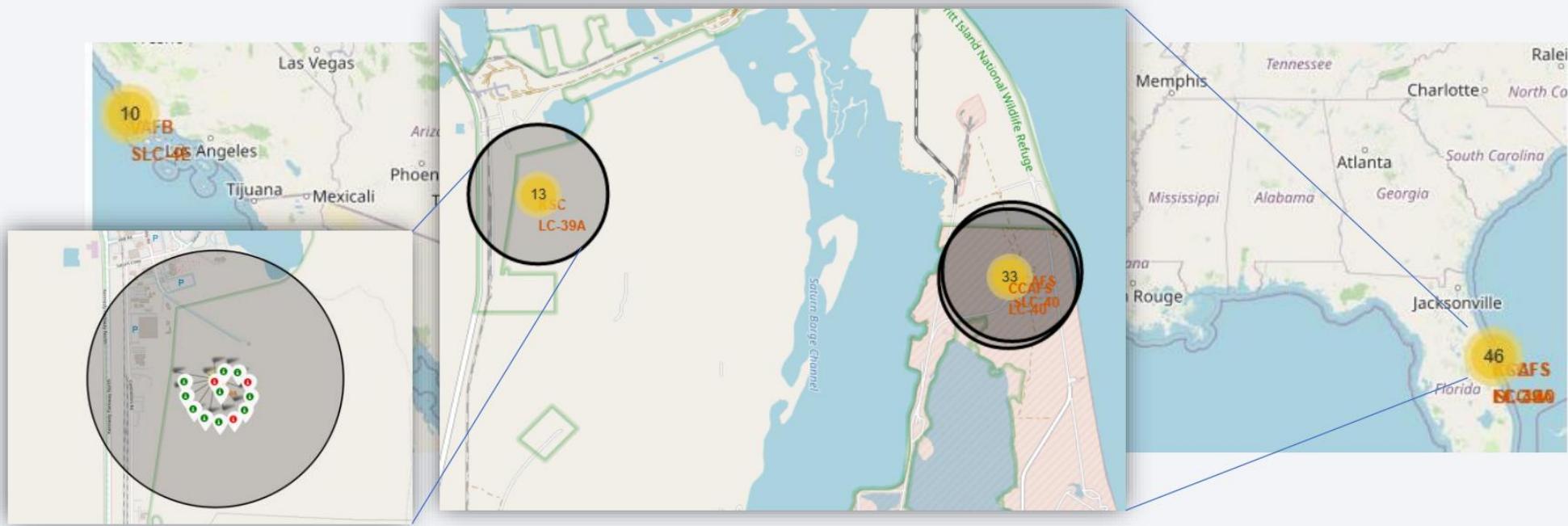
All launch sites



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

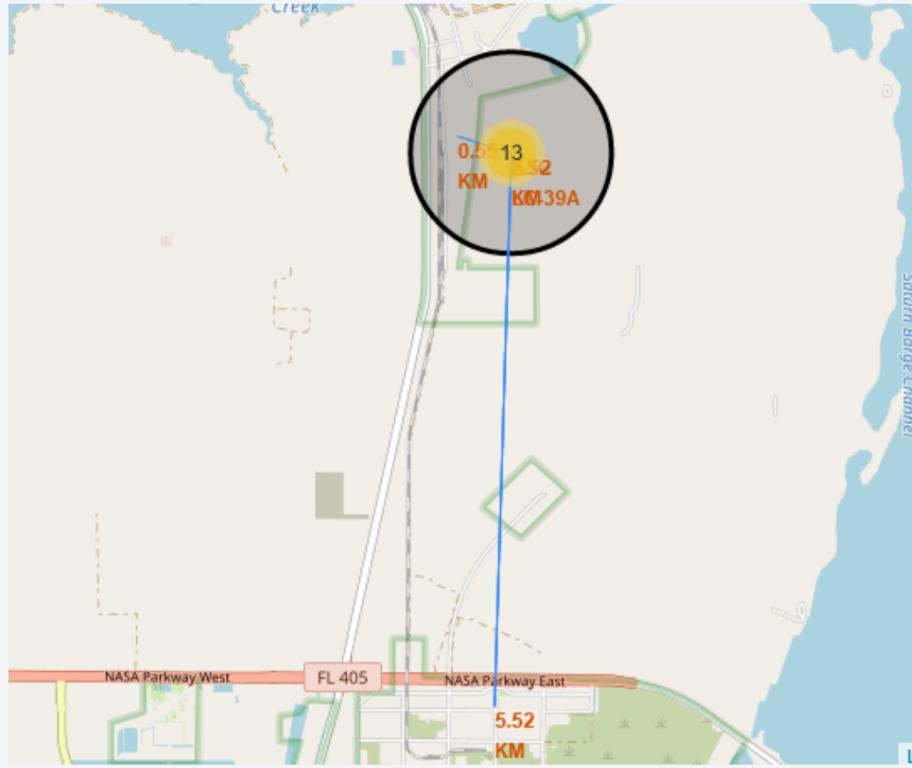
Launch outcomes by site

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



- Green markers indicate successful and red ones indicate failure.

Logistics and safety



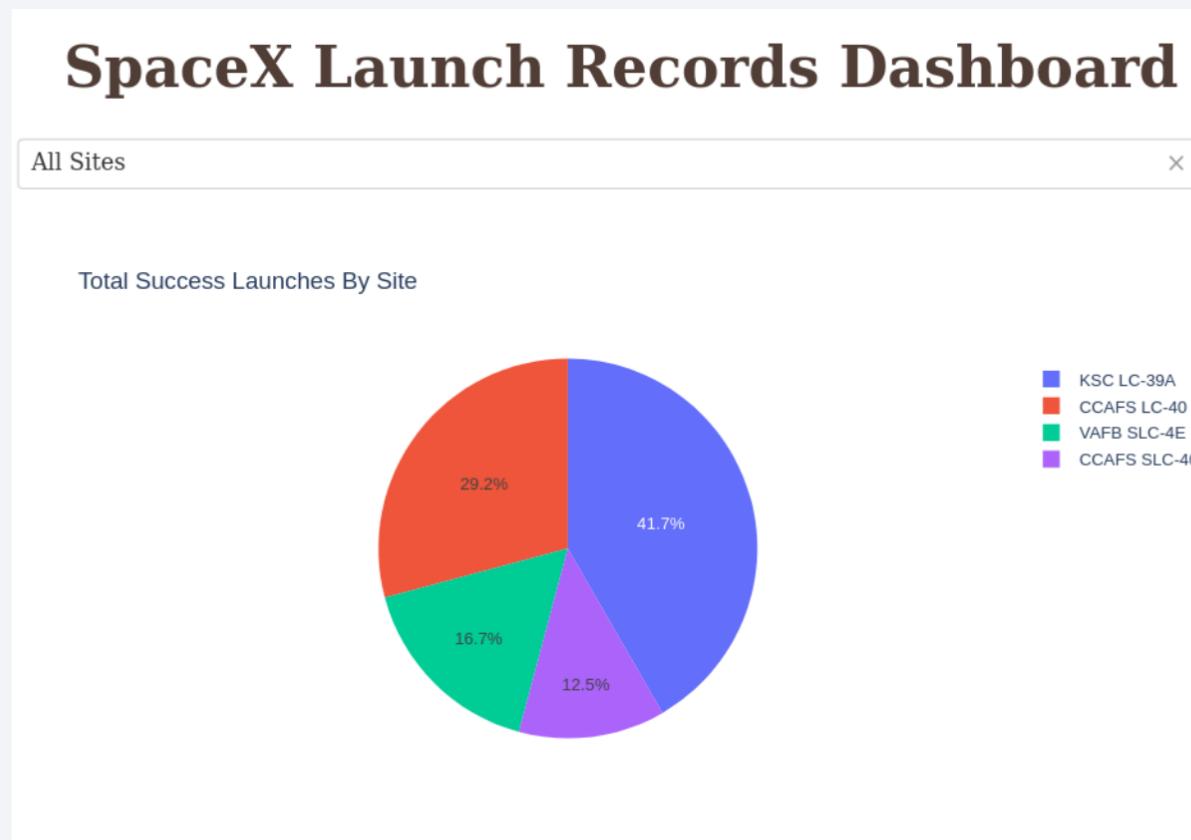
- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

The background of the slide features a close-up photograph of a printed circuit board (PCB). The left side of the image has a blue color overlay, while the right side has a red color overlay. The PCB itself is dark grey or black, with numerous red and blue printed circuit lines (traces) connecting various components. Components visible include a large integrated circuit chip on the left, several surface-mount resistors, capacitors, and other small electronic parts. A few yellow circular components, likely SMD capacitors, are also scattered across the board.

Section 4

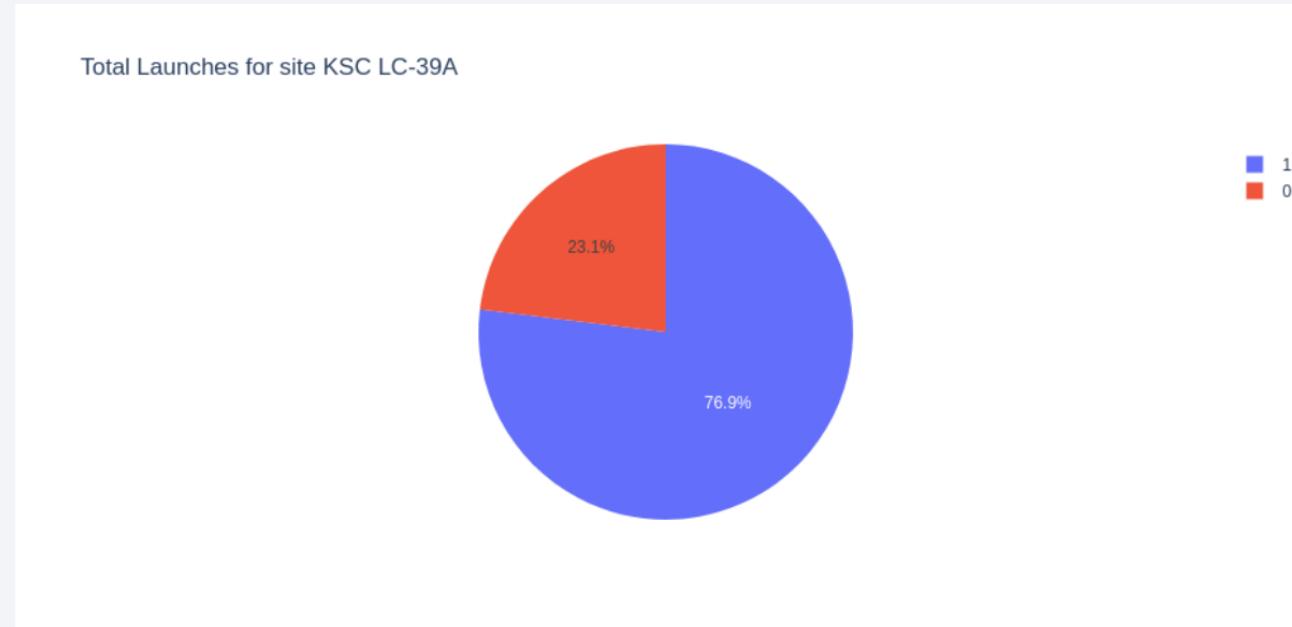
Build a Dashboard with Plotly Dash

Successful Launches by Site



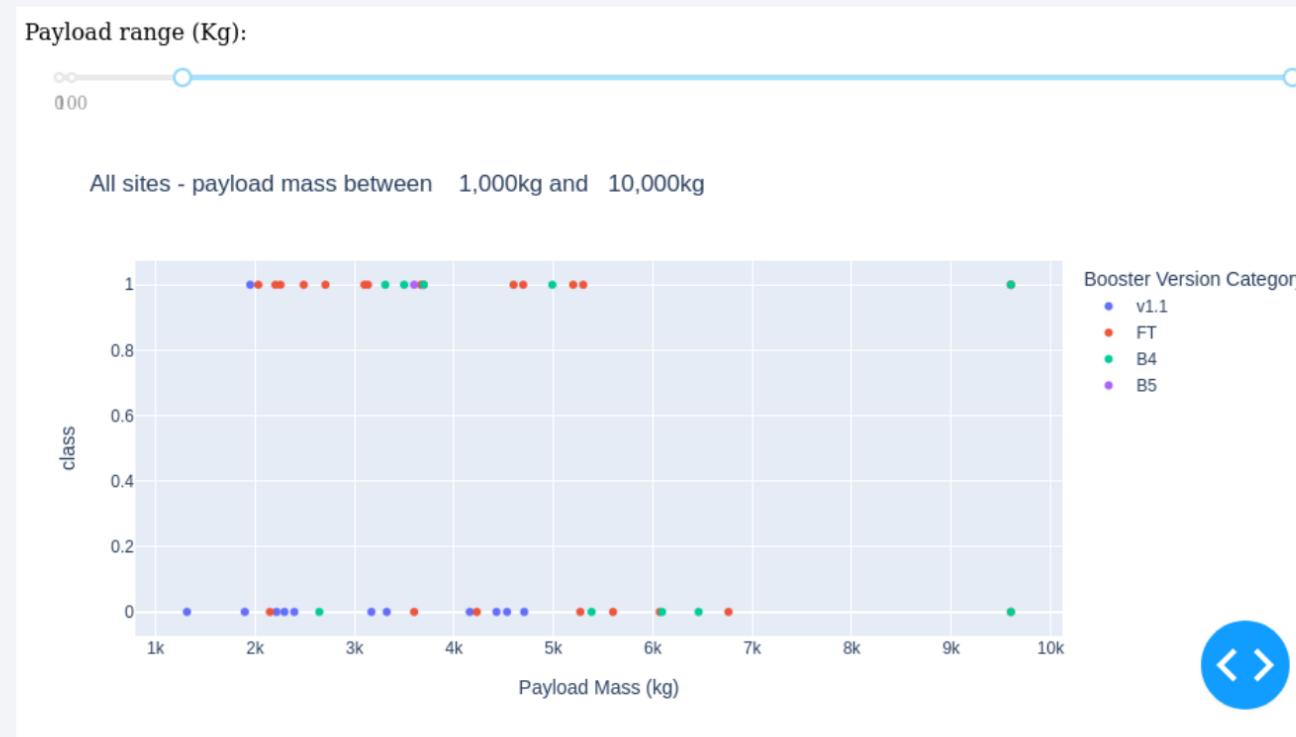
- The place from where launches are done seems to be a very important factor of success of missions.

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.

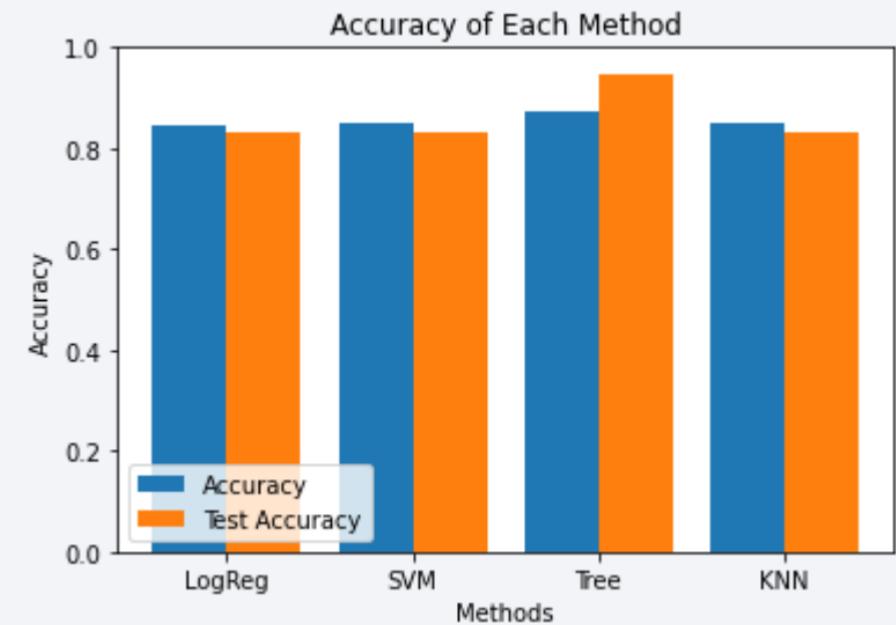
The background of the slide features a dynamic, abstract design. It consists of several curved, overlapping bands of color. A prominent band on the left is a bright blue, while another on the right is a warm yellow. These colors transition into lighter, more diffused tones towards the edges of the frame. The overall effect is one of motion and depth.

Section 5

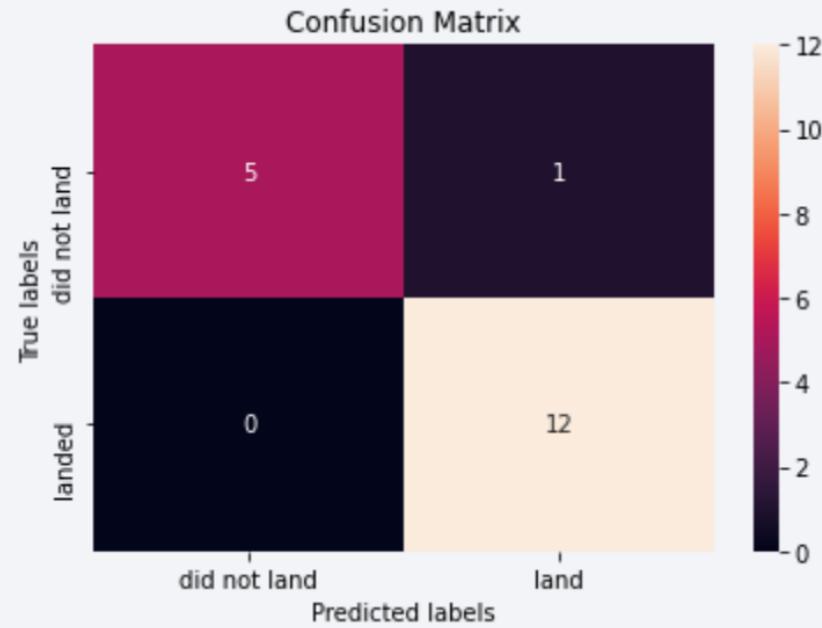
Predictive Analysis (Classification)

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

- 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 appears to be the best model 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!

