



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Muhammad Taimoor Khan
04/12/2023



Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
 - SpaceX data was collected from two different sources:
 - The SpaceX API (<https://api.spacexdata.com/v4/rockets/>)
 - Web scraping of the Wikipedia page on Falcon 9 and Falcon Heavy launches (https://en.wikipedia.org/wiki/List_of_Falcon/9_and_Falcon_Heavy_launches).
- Perform data wrangling
 - The collected data was enhanced by generating a landing outcome label, which was derived from summarized and analyzed outcome data features.
- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

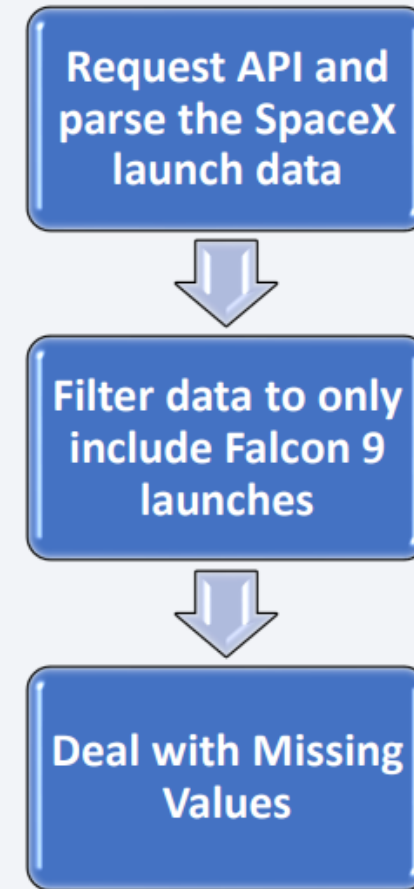
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - The data collected up to this point underwent normalization, was split into training and test sets, and was assessed using four distinct classification models. The accuracy of each model was evaluated using various parameter combinations.

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.

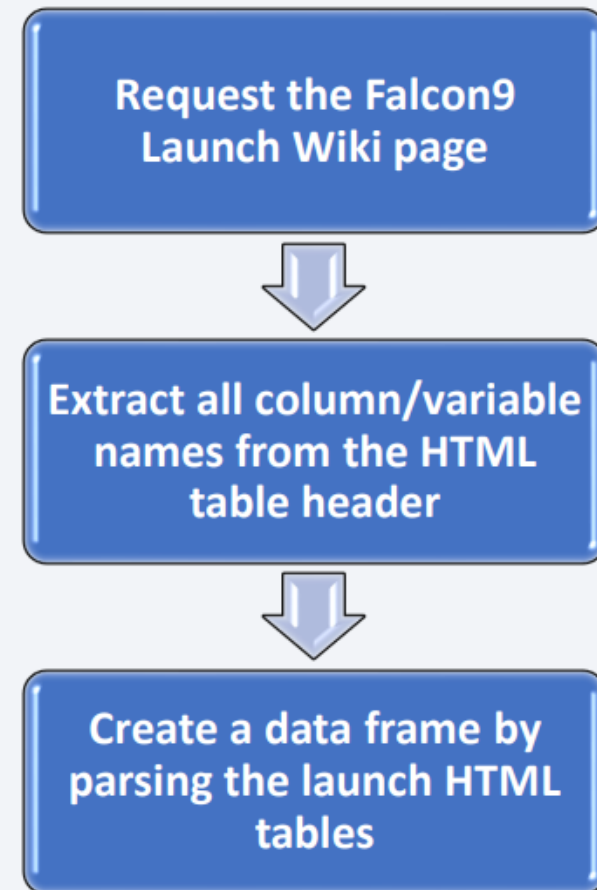
Data Collection – SpaceX API

- 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/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/data%20collection/spacex_data_collection_api.ipynb



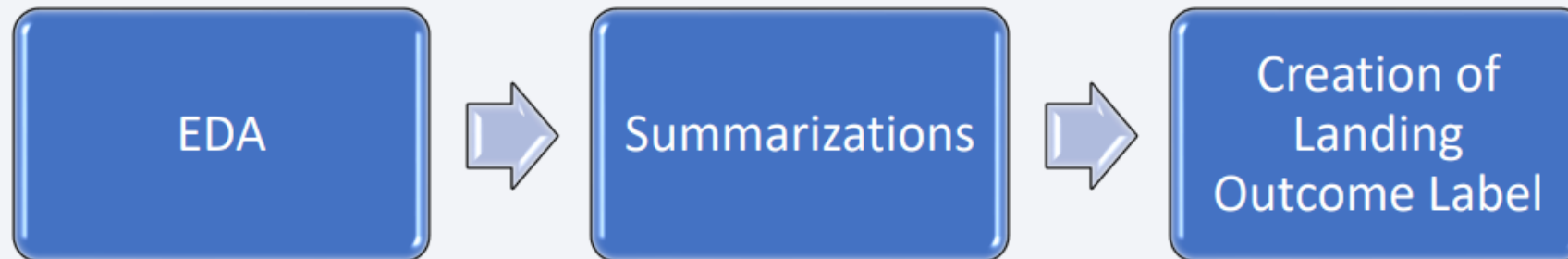
Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- Data is downloaded from Wikipedia according to the flowchart and then persisted.
- Source Code:
https://github.com/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/data%20collection/spacex_data_collection_webscrapping.ipynb



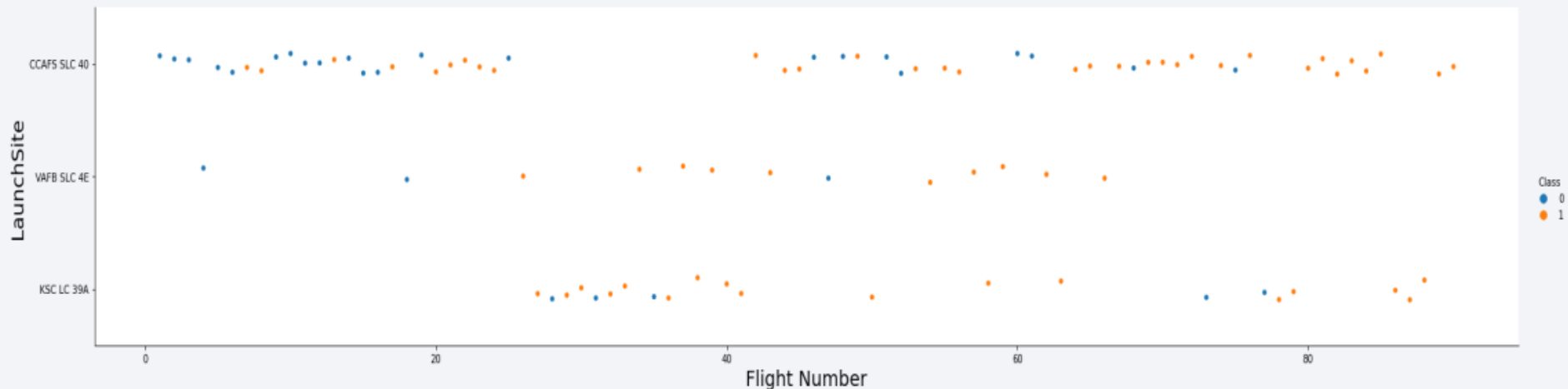
Data Wrangling

- 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/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/data%20wrangling/spacex_data_wrangling.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/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/EDA%20with%20Pandas/spacex_eda_pandas.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 the 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/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/EDA%20with%20SQL/spacex_exploratory_analysis.ipynb

Build an Interactive Map with Folium

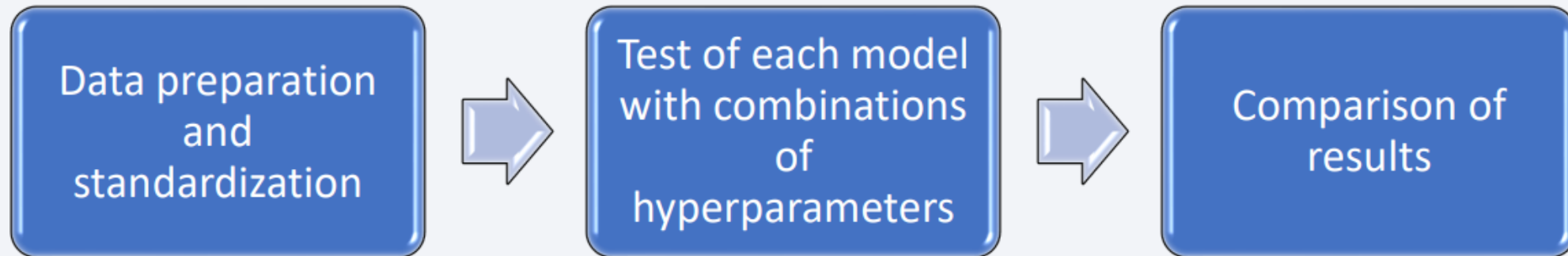
- Markers, circles, lines and marker clusters were used with Folium Maps.
 - Markers indicate points like launch sites
 - Circles indicate highlighted areas around specific co-ordinates, like NASA Johnson Space Center.
 - Marker Clusters indicate groups of events in each co-ordinate, like launches in a launch site.
 - Lines are used to pin point the distance between two co-ordinates.
- Source Code: https://github.com/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/Interactive%20Analytics%20Dashboards/launchsite_location_analysis.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 where is best place to launch according to payloads.
- Source Code: https://github.com/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/Interactive%20Analytics%20Dashboards/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/MTaimoorK/Applied-Data-Science-Capstone-IBM/blob/main/predictive%20analysis/spacex_ml_prediction.ipynb

Results

- Exploratory data analysis results:
 - SpaceX uses 4 different launch sites.
 - The first launches were done to SpaceX itself & NASA.
 - The average payload of F9 v1.1 booster is 2,928 KG.
 - The first success landing outcome happened in 2015 five 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 at drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015.
 - The number of landing outcomes became better as years passed by.

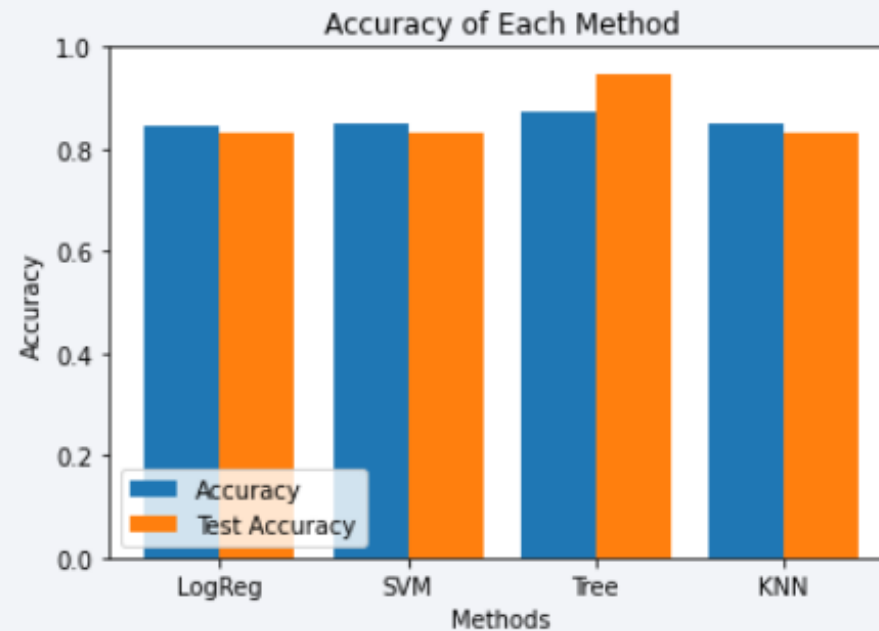
Results

- 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%

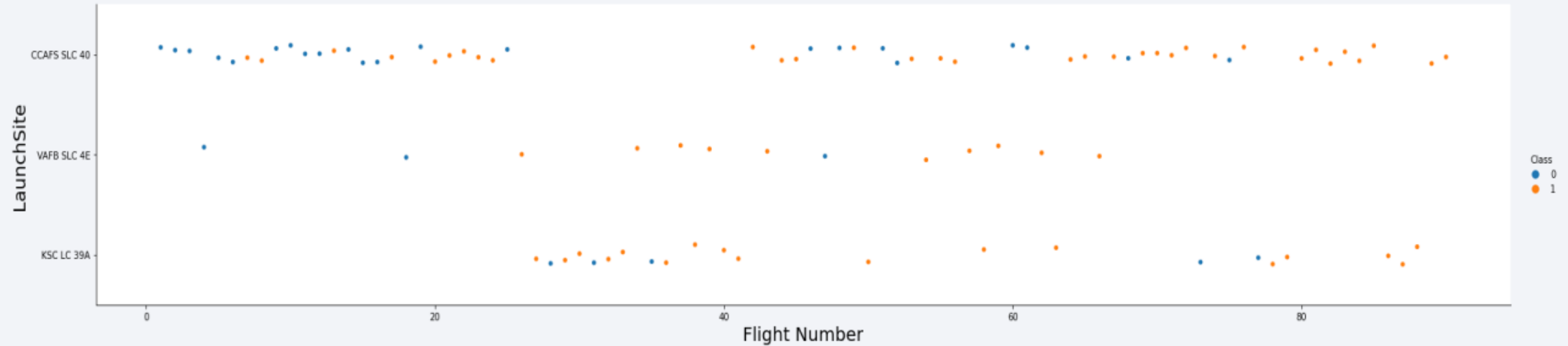


The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

Section 2

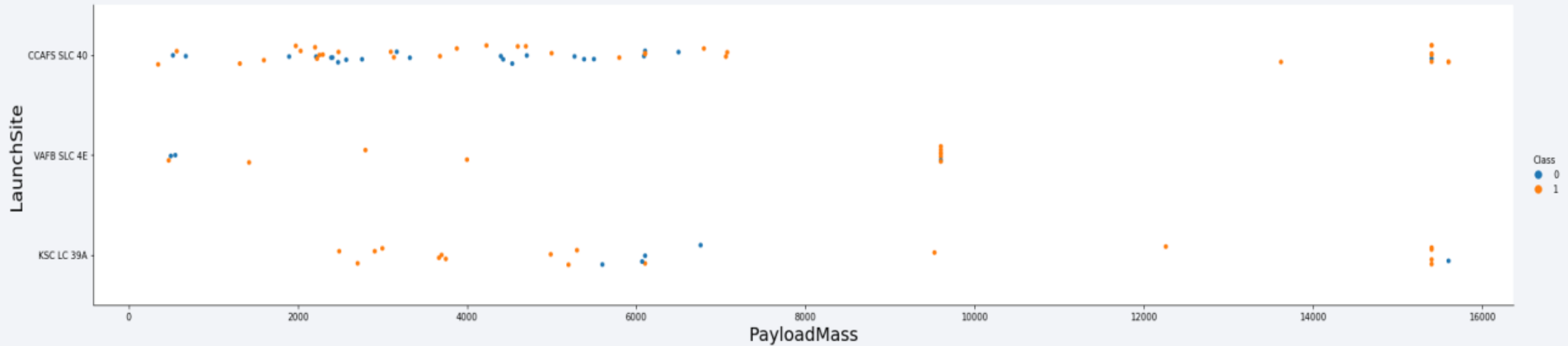
Insights drawn from EDA

Flight Number vs. Launch Site



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

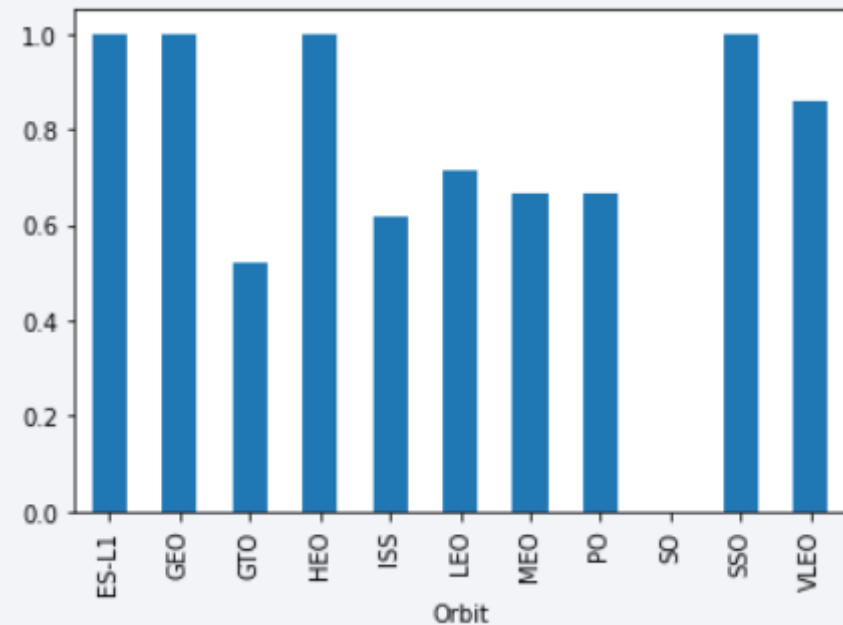
Payload vs. Launch Site



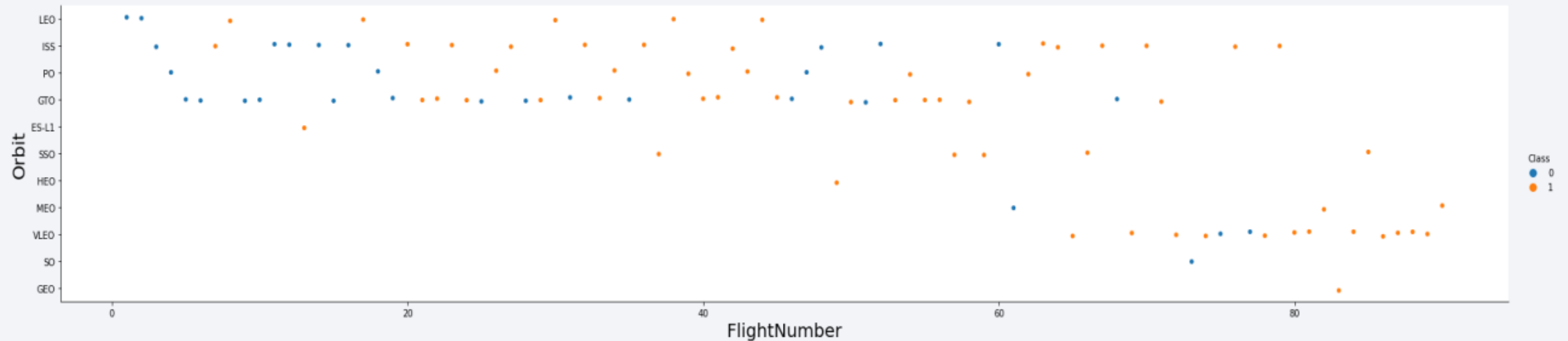
- Payload over 9,000 KGs have excellent success rate.
- Payloads over 12,000 KGs seems to be successful only on CCAFS SLC 40 and KSC LC 39A

Success Rate vs. Orbit Type

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

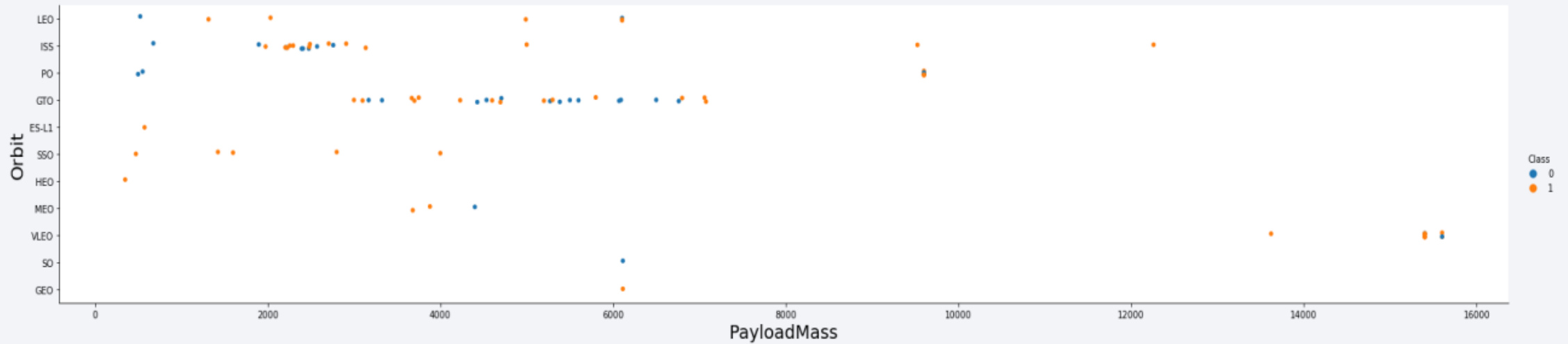


Flight Number vs. Orbit Type



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

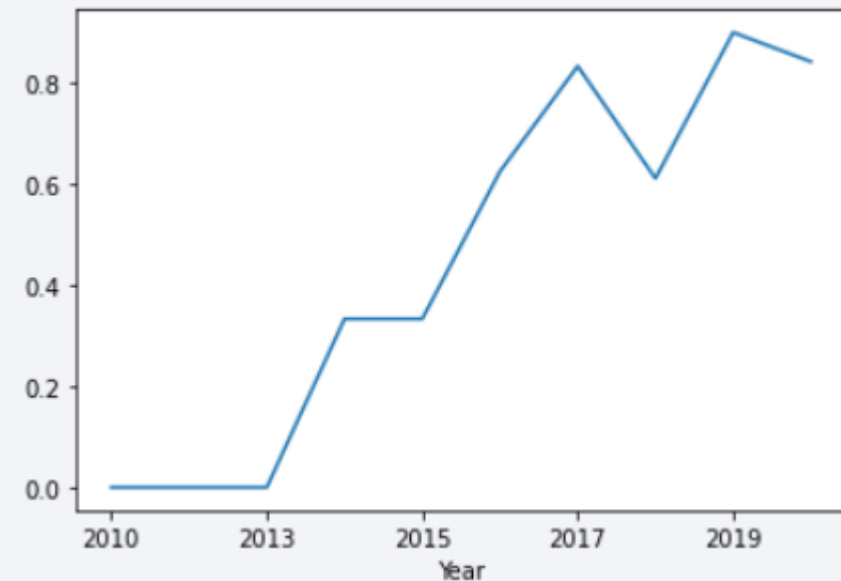
Payload vs. Orbit Type



- Seemingly, there is no relation between payload and success rate to orbit GTO.
- ISS orbit have the widest range of payload and a good success rate.
- There are fewer launches to the orbit SO and GEO

Launch Success Yearly Trend

- Success rate started increasing in 2013 and kept rising until 2020.
- It seems that the first three years were a period of adjusts and improvement of technology



All Launch Site Names

- According to the dataset, there are four launch sites:

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

- They are retrieved by selecting unique occurrence of “LaunchSites” values from the dataset.

Launch Site Names Begin with 'CCA'

- 5 records where the 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 attempt

- Here we can see the 5 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 the 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 the data by the booster version above and calculating the average payload mass we obtained the value of 2,928 KGs

First Successful Ground Landing Date

- First successful landing outcome on ground pad came on:

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 (...)	Booster Version
F9 B5 B1048.4	F9 B5 B1051.4
F9 B5 B1048.5	F9 B5 B1051.6
F9 B5 B1049.4	F9 B5 B1056.4
F9 B5 B1049.5	F9 B5 B1058.3
F9 B5 B1049.7	F9 B5 B1060.2
F9 B5 B1051.3	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

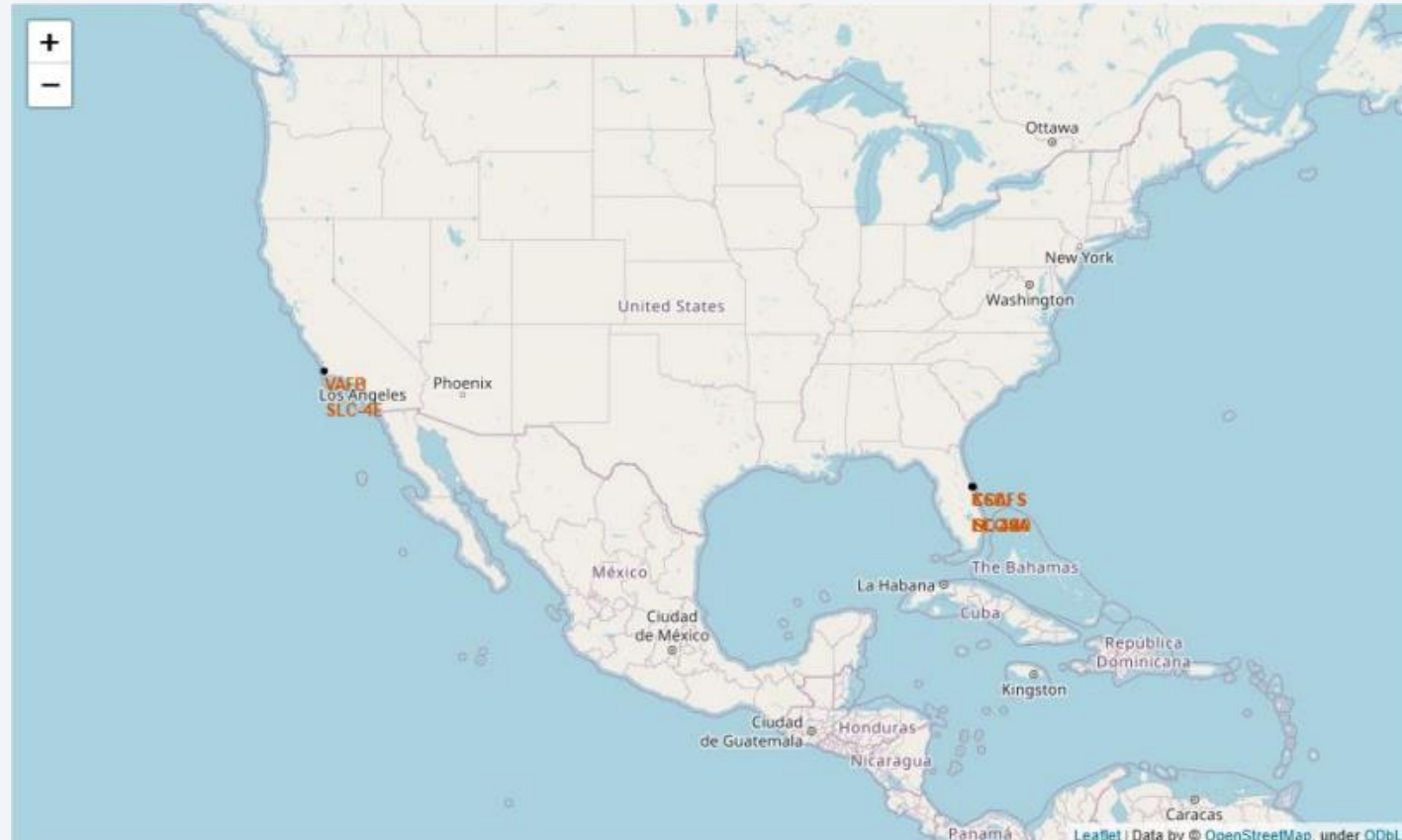
- This view of data alerts us that “No attempt” must be taken in account.

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

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.

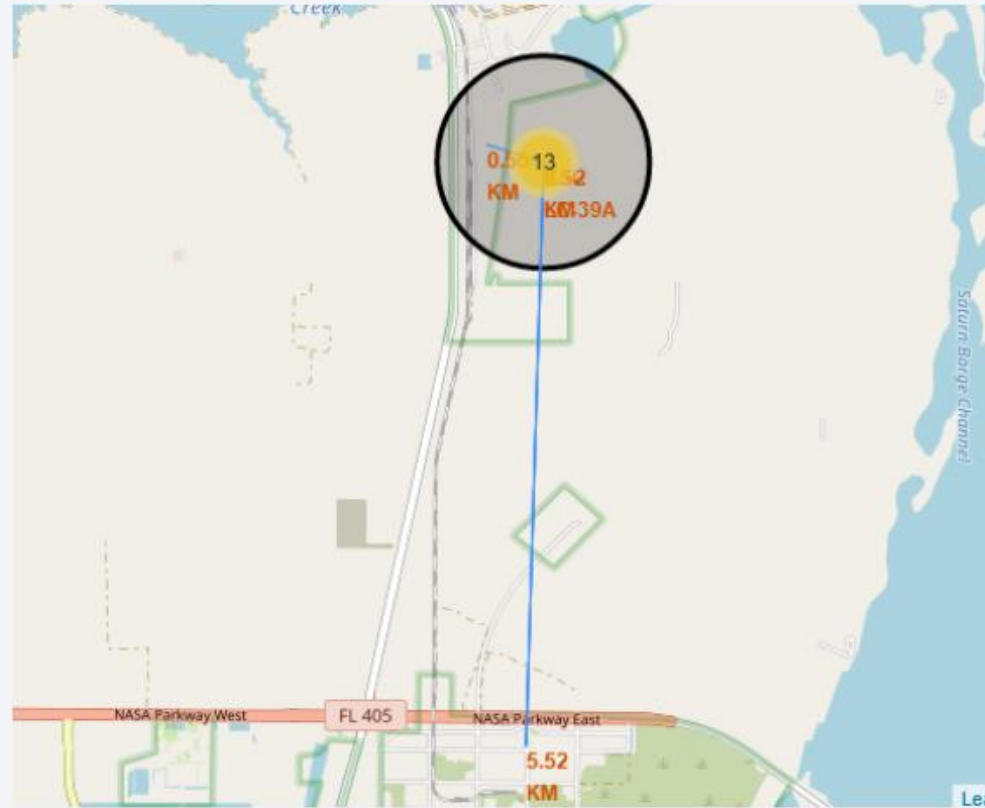
Launch Outcomes By Sites

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



- Green markers indicate successful and red ones indicate failure.

Logistics & Safety



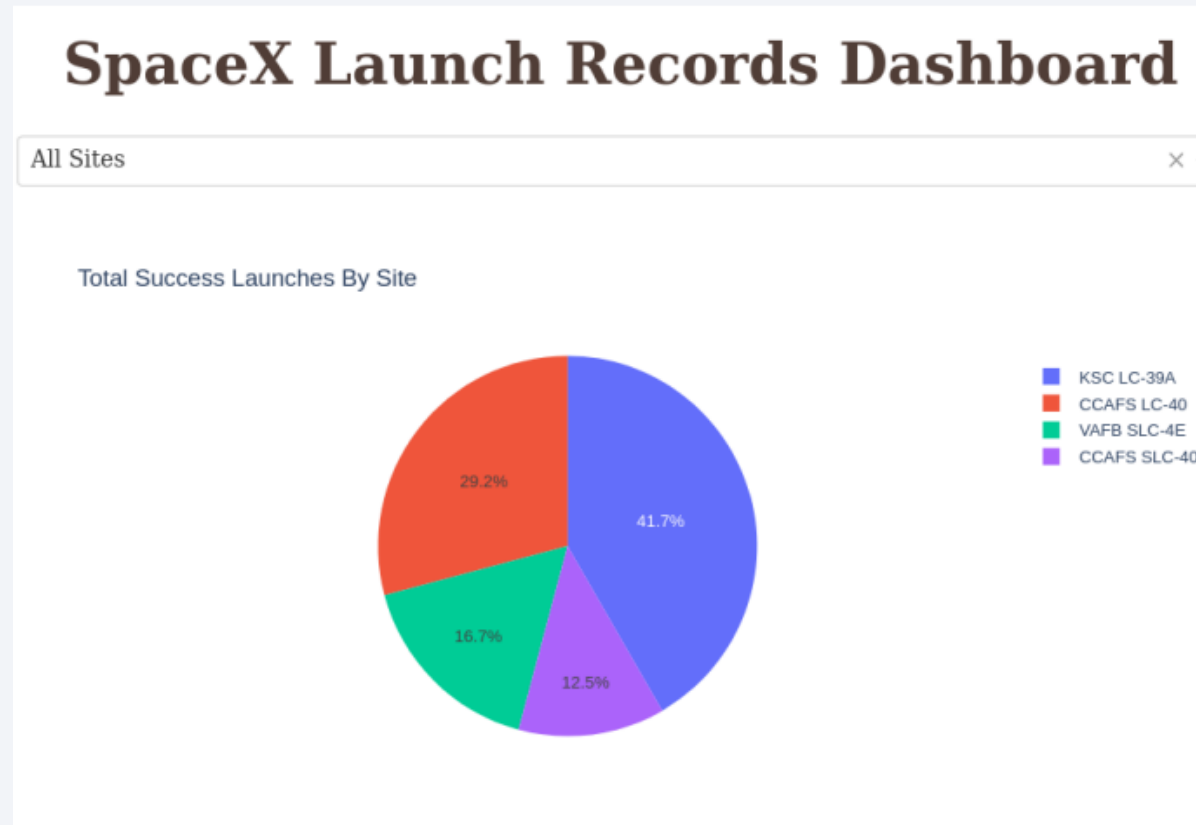
- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas



Section 4

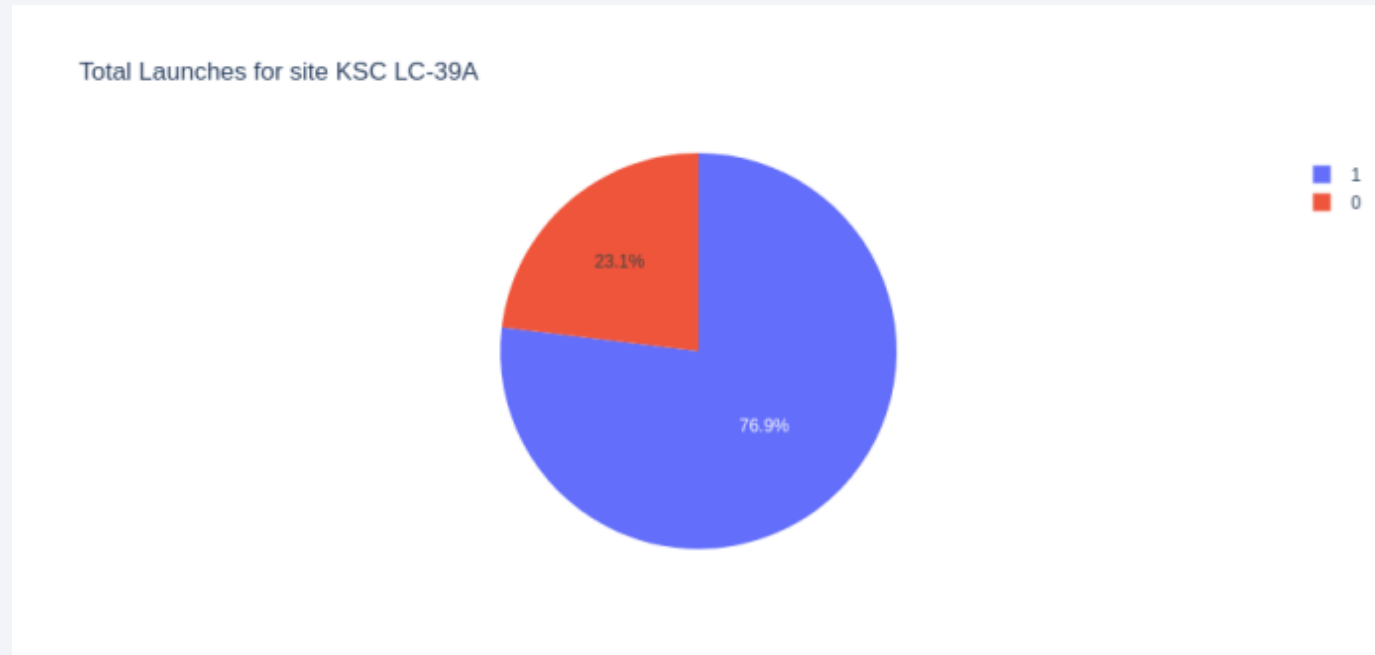
Build a Dashboard with Plotly Dash

Successful Launches by Site



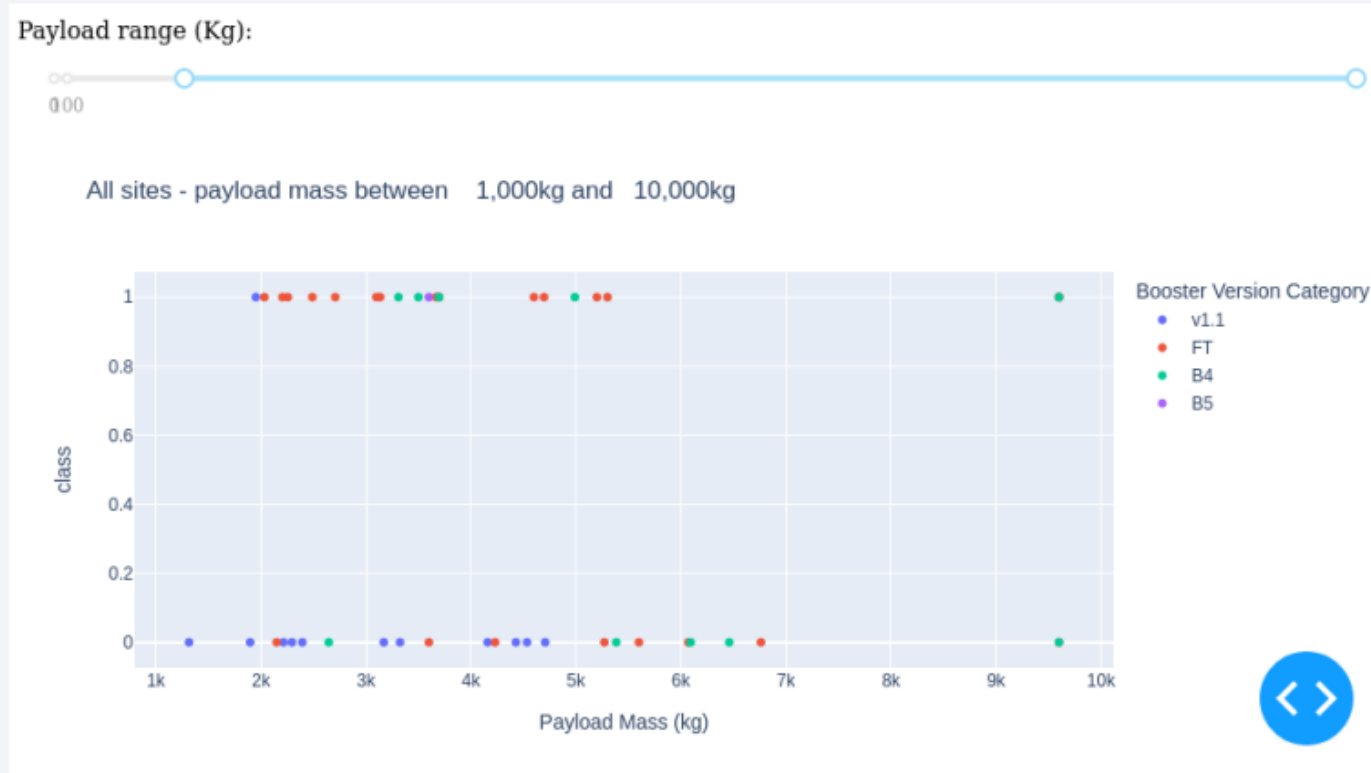
- The place from where launches are done seems to be a very important factor of success of missions.

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



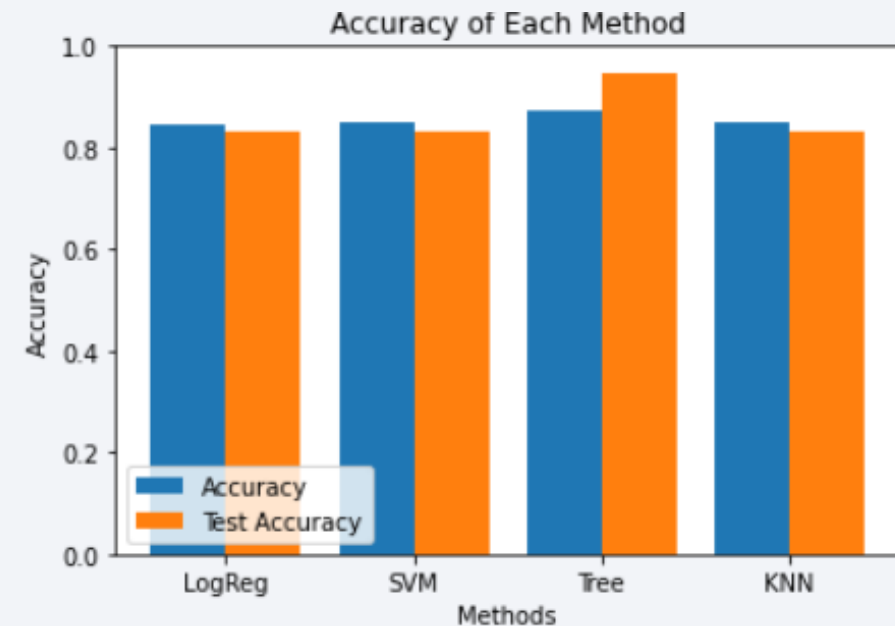
- There's not enough data to estimate risk of launches over 7,000kg

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

- 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

Appendix

- As an improvement for model tests, it's important to set a value to `np.random.seed` variable.

Thank you!

