IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Eduard Fadieiev
28-10-2022

Outline

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

Executive Summary

- **Methodologies used for analysis:**
  - Data were collected using web scraping and through the SpaceX API
  - For EDA (Exploratory Data Analysis) were used: data wrangling, data visualization, interactive visual analytics
  - And MLP (Machine Learning Prediction) was used on top of this

- **Summary of all results:**

  - EDA helped with the identification which features were the best for successful launches

  - MLP, with all collected data, helped to predict the most important characteristics

Introduction

- The main objective is to evaluate the possibility of the new SpaceNow company to compete SpaceX
- Problems you want to find answers:
    - Estimate total costs
    - Place for the most successful launches
    - Place for the most successful landings of the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Two sources were used for gathering the data:
    - Web Scraping

      https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
    - SpaceX API

      https://api.spacexdata.com/v4/rockets/
- Perform data wrangling
  - Collected data was enriched by summarizing and analyzing data after landing outcomes

Methodology

Executive Summary

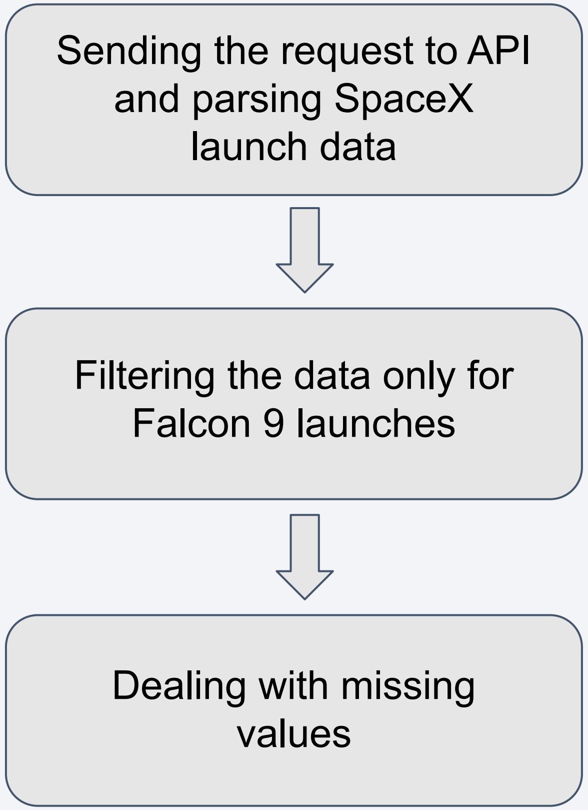
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash

  Perform predictive analysis using classification models:

  - Data were collected, normalized, divided into data sets for training and testing. After this evaluated by four classification models with different combinations of parameters

Data Collection

- Data sets were collected using SpaceX API and from Wikipedia

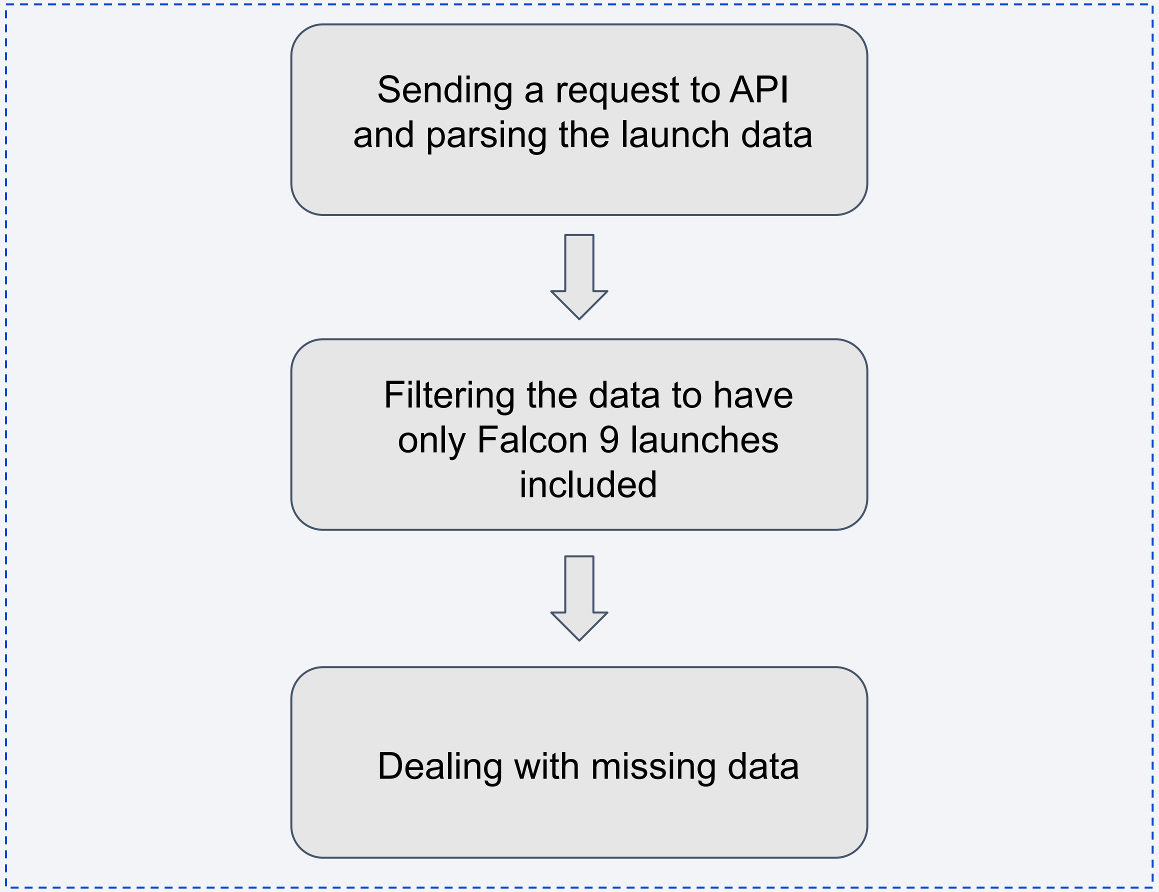


Data Collection – SpaceX API

- SpaceX offers a public API through which the data can be collected
- API was used accordingly to the flowchart

Source code:

https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

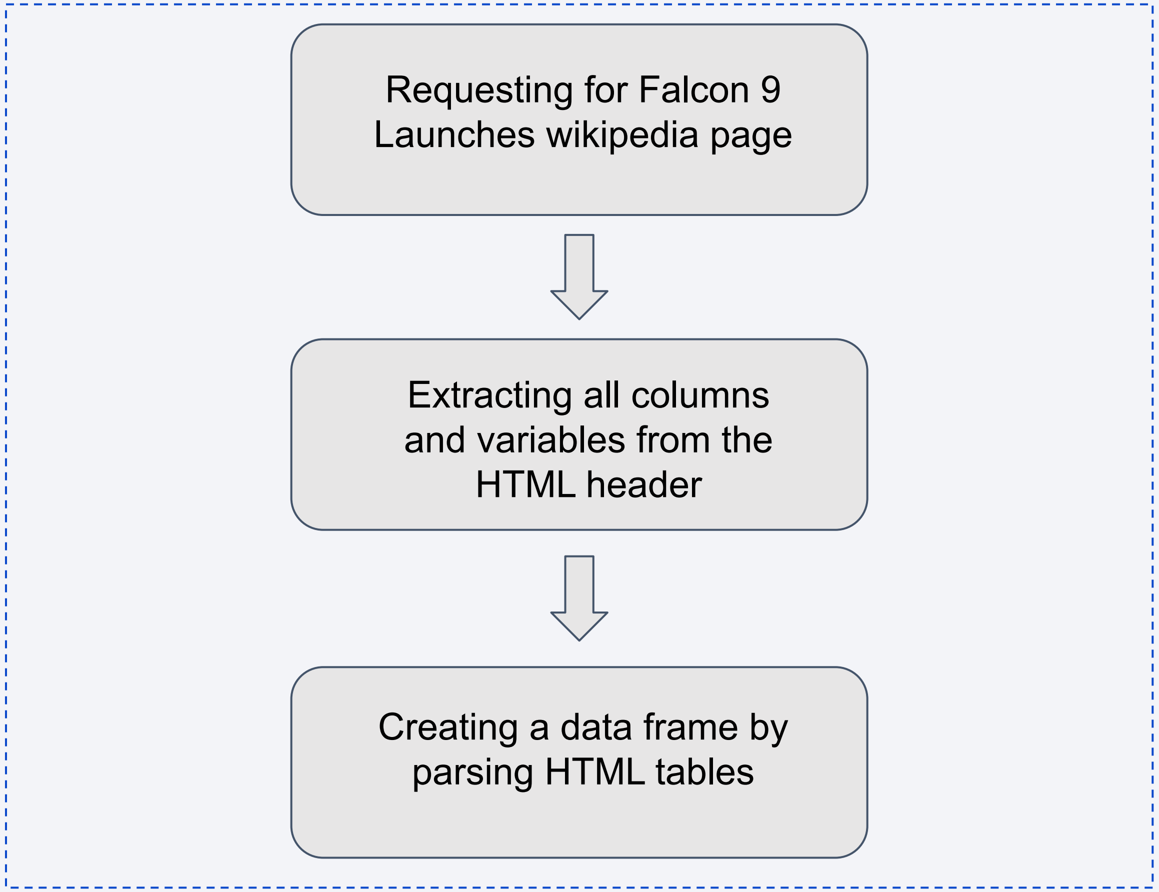


Data Collection - Scraping

- Data of Falcon 9 launches can be also found and obtained from Wikipedia
- Hence, data were downloaded accordingly to the flowchart

Source code:

https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

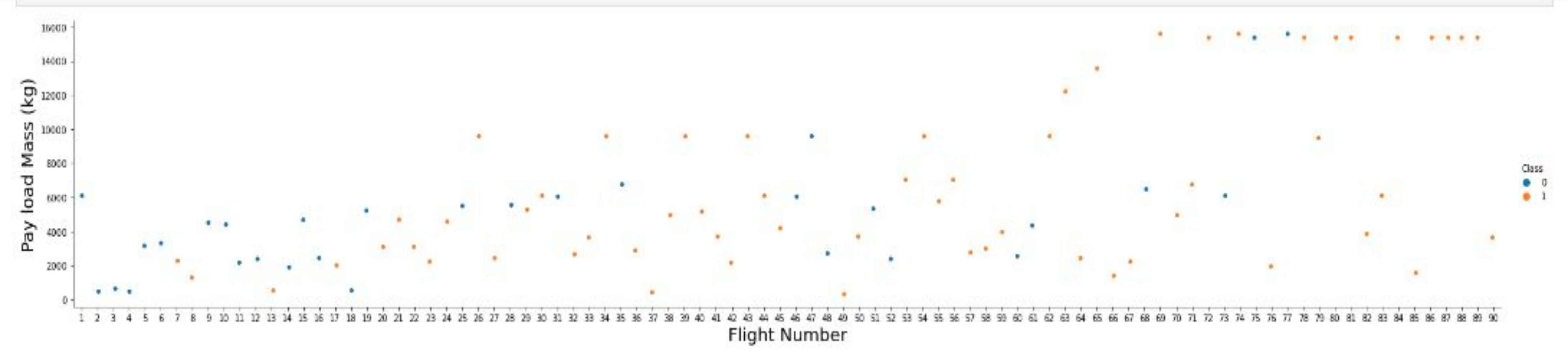
- Firstly EDA was performed on the dataset
- After this summaries on launches per site, occurrences of each orbit and occurrences of mission outcome per orbit were made
- Lastly, the Landing outcome per orbit type was made

Source code:

https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scaterplots and barplots were used for for the visualization of relationships



Source code:

https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- SQL queries performed:
  - Names of the unique launch sites
  - Top 5 launch sites beginning with 'CCA'
  - Total payload mass carried by NASA (CRS) boosters
  - Average payload mass carried by Falcon 9
  - Date of the first successful landing on ground
  - Names of the boosters with payload mass between 4000 and 6000 kg
  - Total number of successful and failure missions
  - Names of the booster versions with the maximum payload mass
  - Failed landing outcomes in drone ship, launch site names and booster versionsin 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/eddyfadeev/applied-data-science-capstone/blob/main/EDA%20with%20Data%20Visualization.ipynb

Build an Interactive Map with Folium

- Markers are indicating launch sites
- Lines are indicating distances between two coordinates
- Circles are highlighting the areas with specific coordinates
- Marker clusters are standing for groups of elements

Source code:

https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Graphs and plots used:
  - Payload range
  - Launches by site in percents
- It allowed to analyze the relation between payloads and sites quickly and helped to identify the best place for launch with according payload

Source code:

https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/dash_app.py

Predictive Analysis (Classification)

- Regression, SVM, Decision tree and k-nearest neighbors were compared

Data preparation → Data standardization → Test of each model → Comparison of results

Source code:

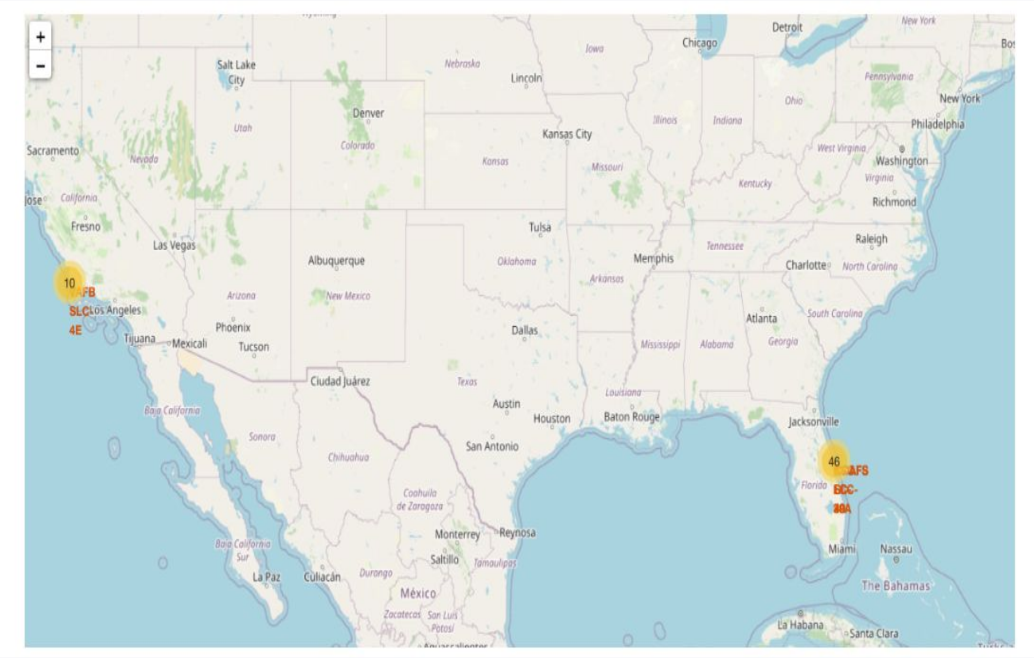
https://github.com/eddyfadeev/applied-data-science-capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Exploratory data analysis results:
  - SpaceX uses 4 different launch sites
  - The first launches were done by SpaceX and NASA
  - Avg. payload of F9 is 2,928 kg
  - The first success launch is dated 2015

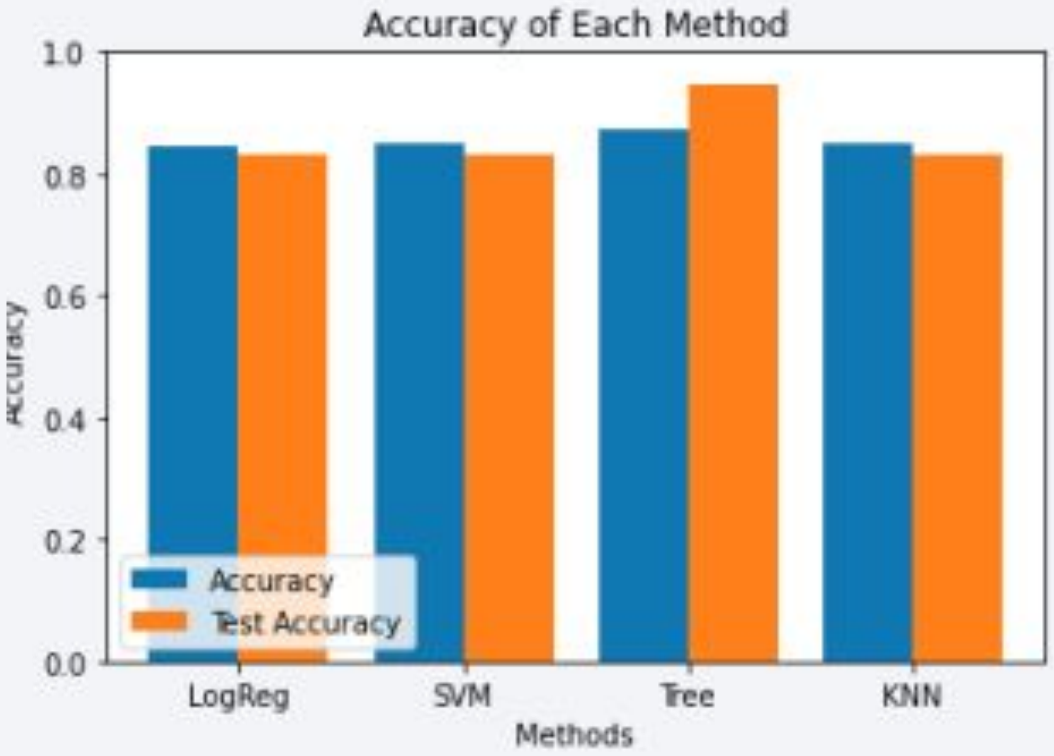
Results

- Most launches happened on the east coast sites



Results

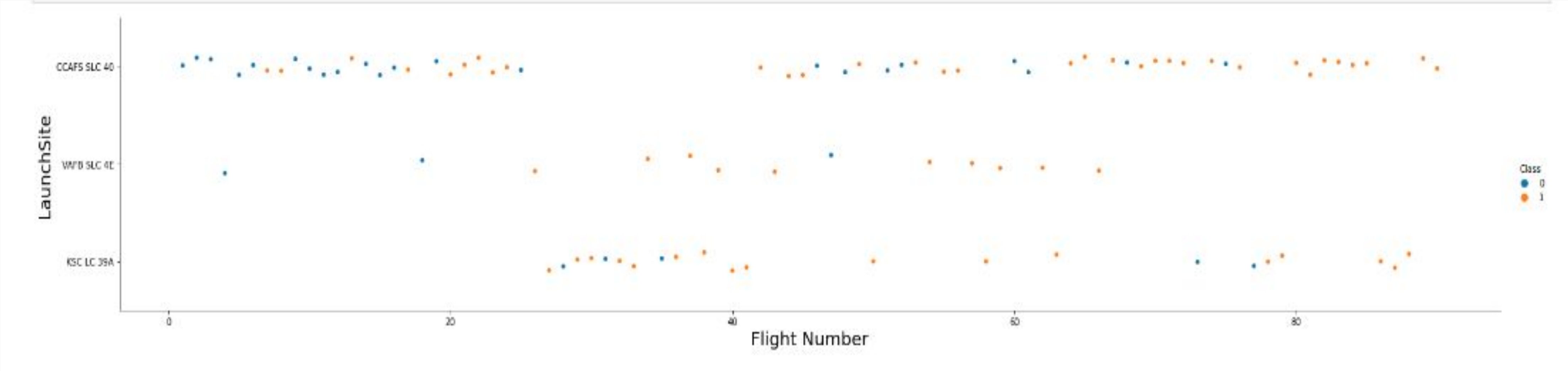
- Predictive analysis results



Section 2

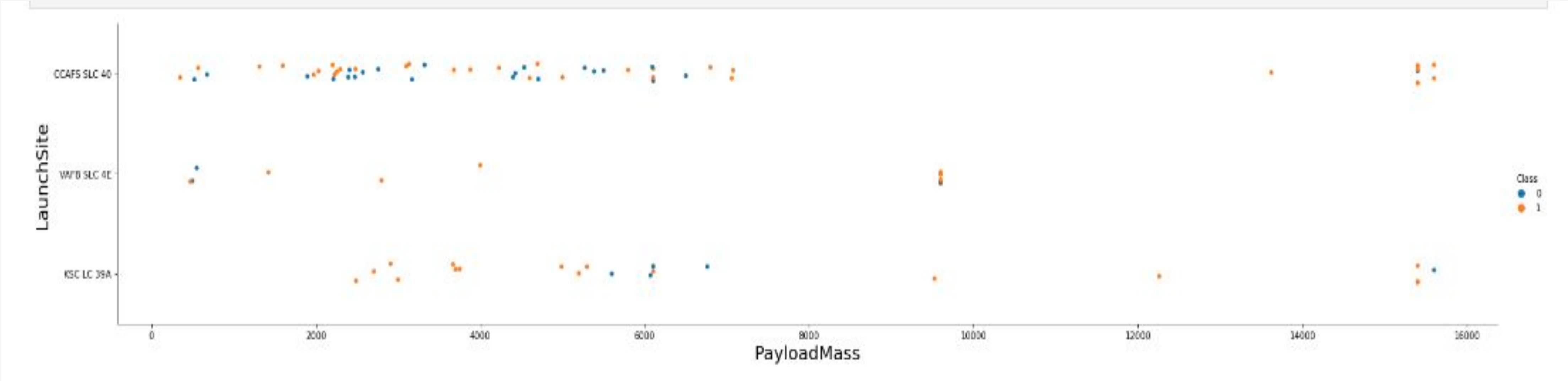
Insights drawn from EDA

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

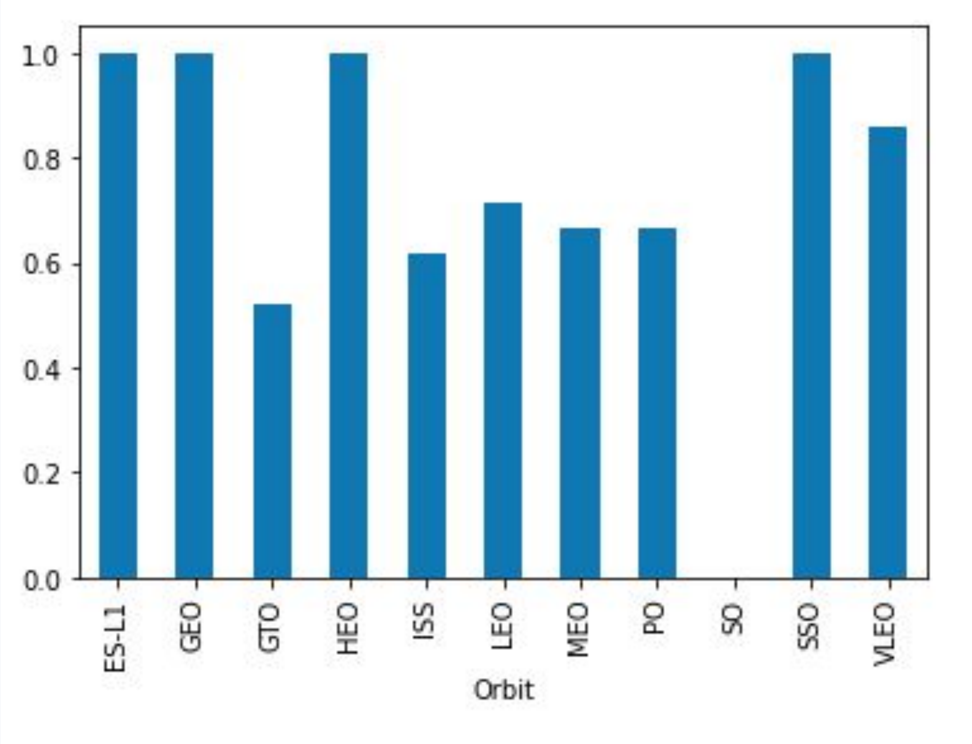


- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate; • Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

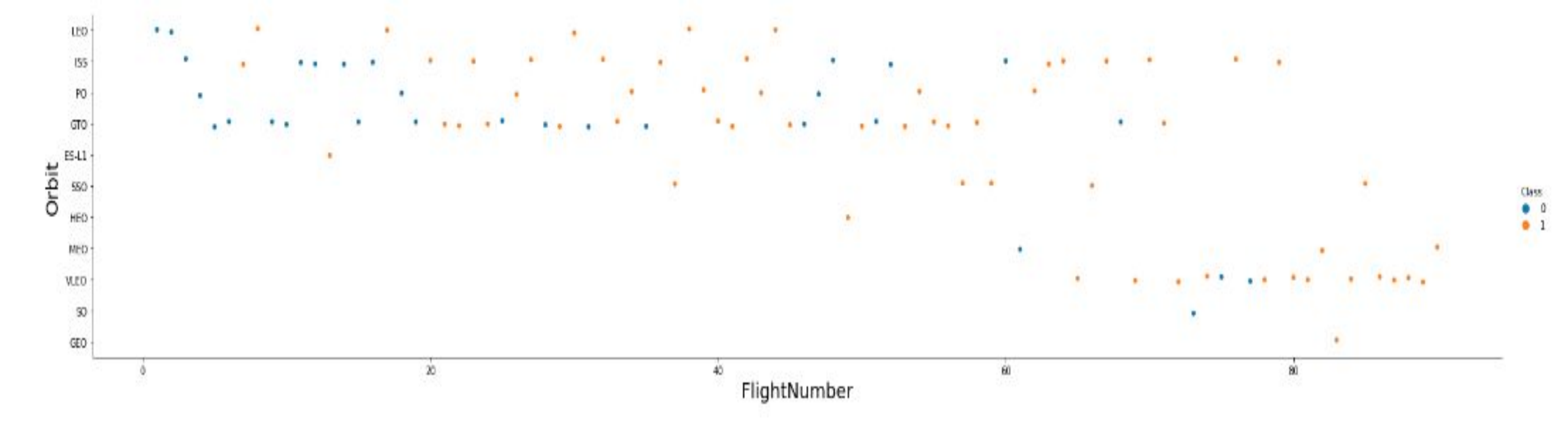
Success Rate vs. Orbit Type

- The biggest success rates happens to orbits:
    - ES-L1
    - GEO
    - HEO
    - SSO

- Followed by:
    - VLEO (above 80%)
    - LFO (above 70%).

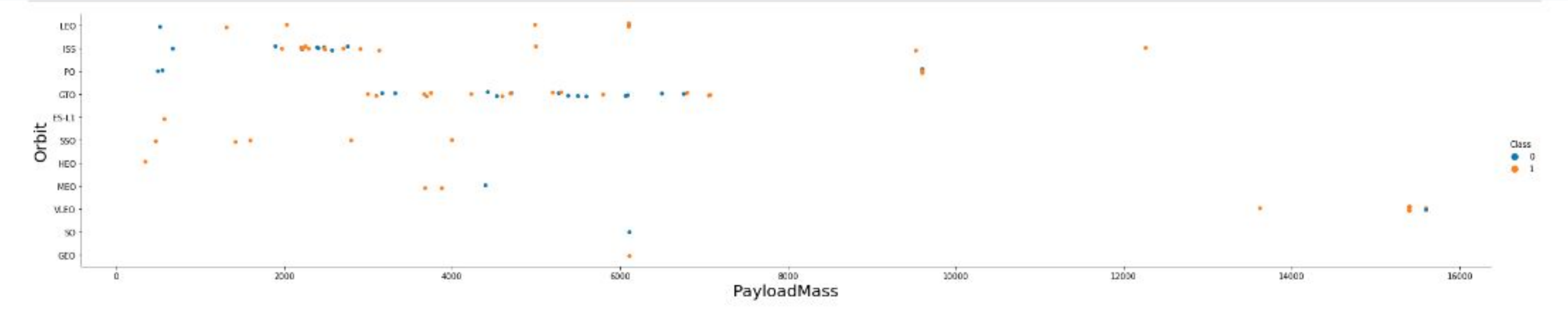


Flight Number vs. Orbit Type



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

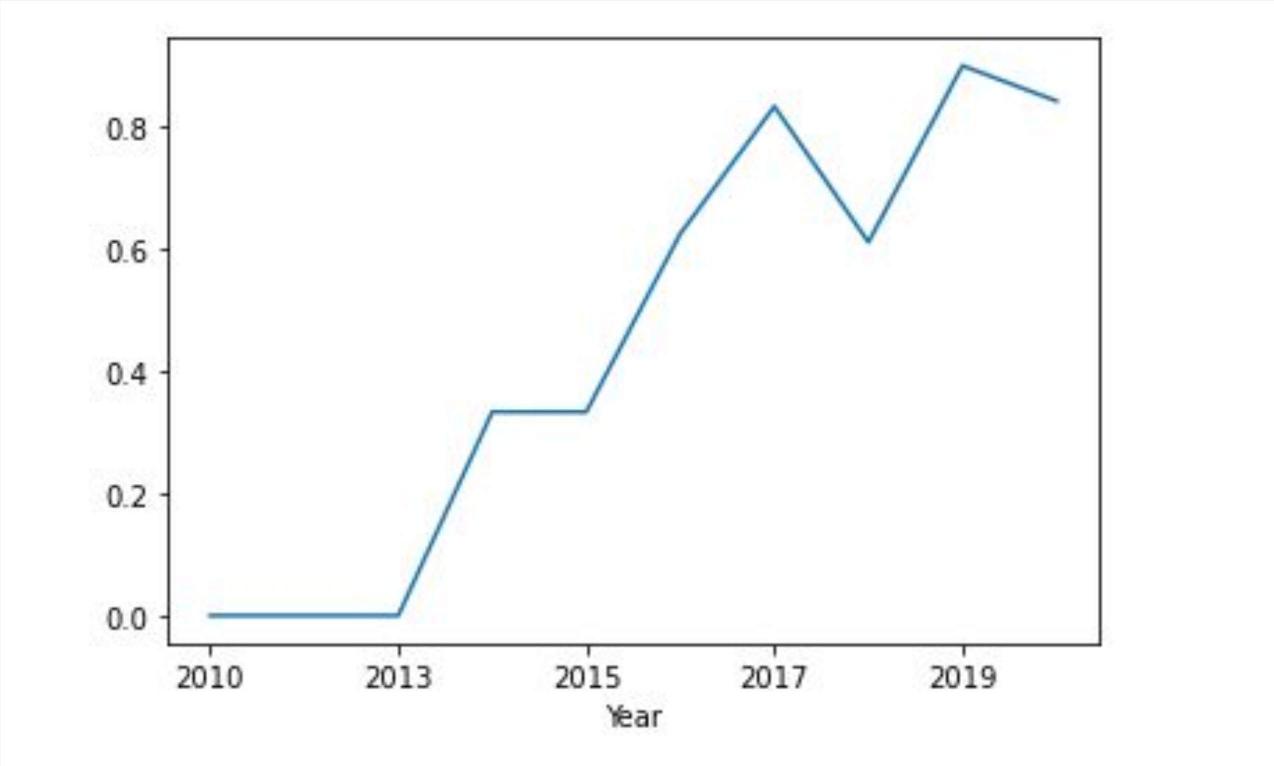
Payload vs. Orbit Type



- 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 by 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

- The 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

- The 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

- Booster 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

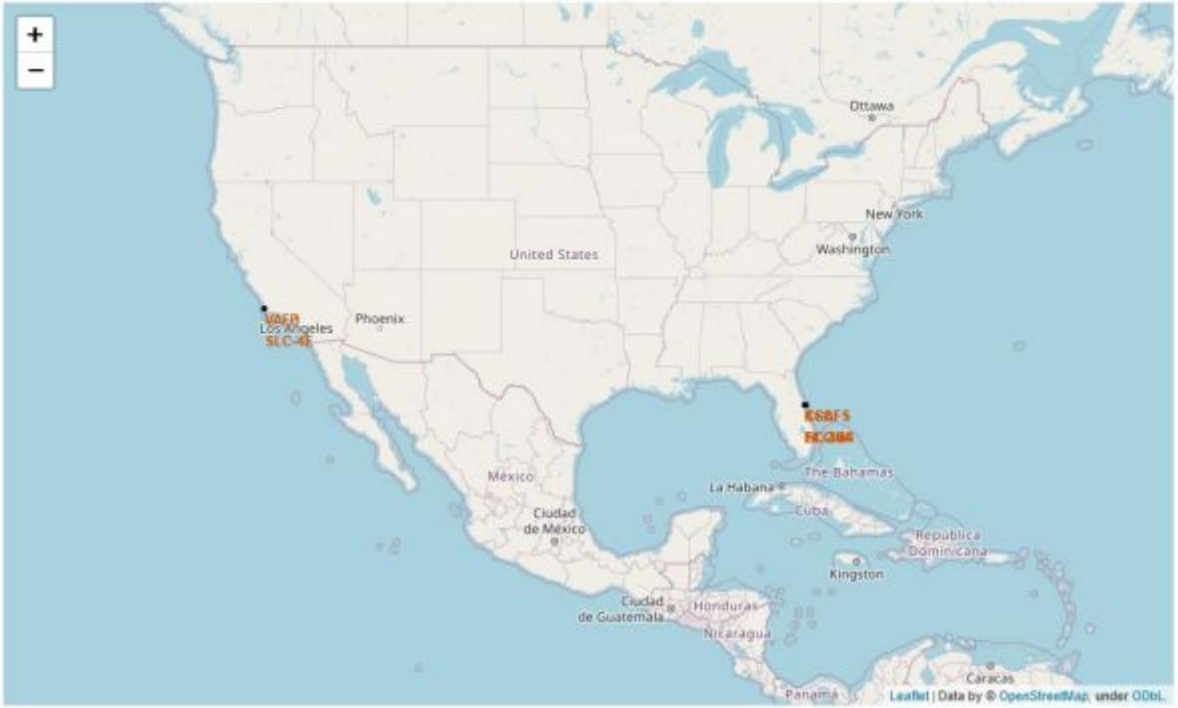
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

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

Section 3

Launch Sites Proximities Analysis

All launch sites

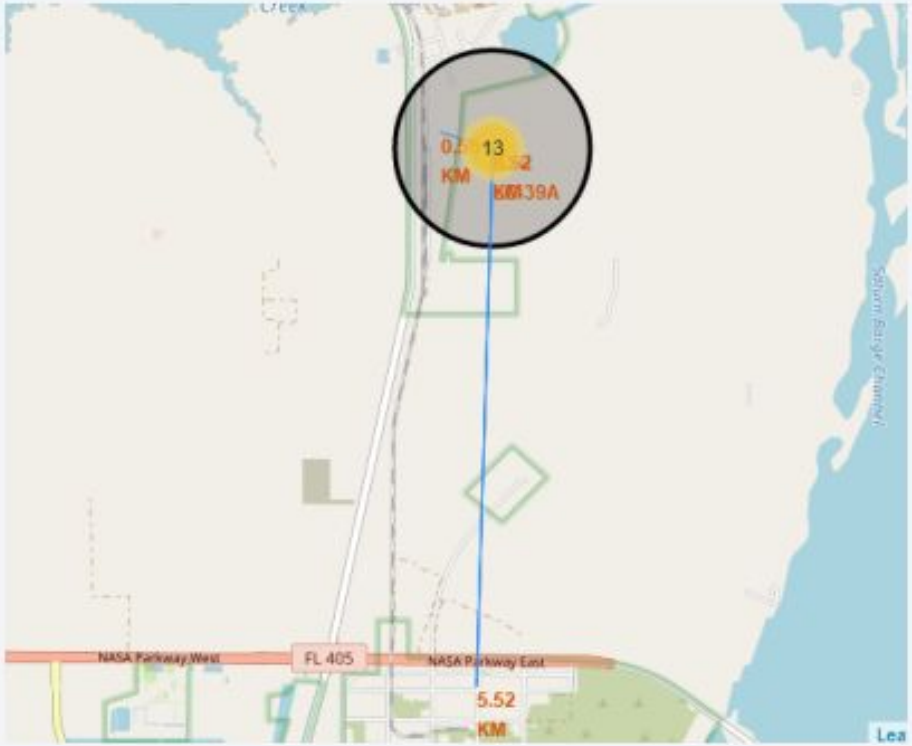


- Launch sites are near sea, probably by safety, but not too far from roads and railroads

Launch outcomes

- Green markers indicate successful and red ones indicate failure.

Logistics and Safety

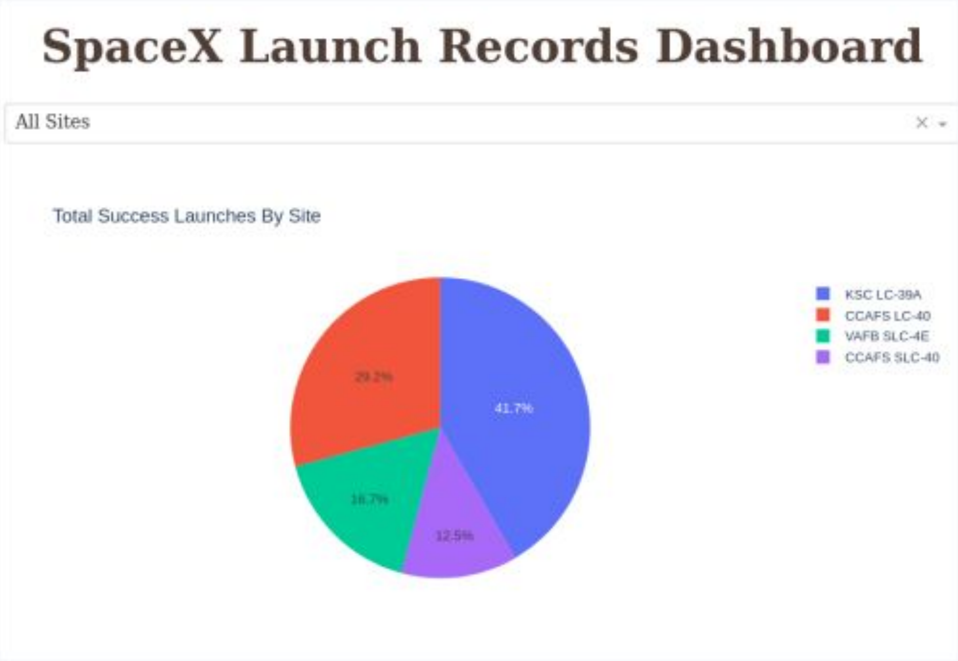


- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

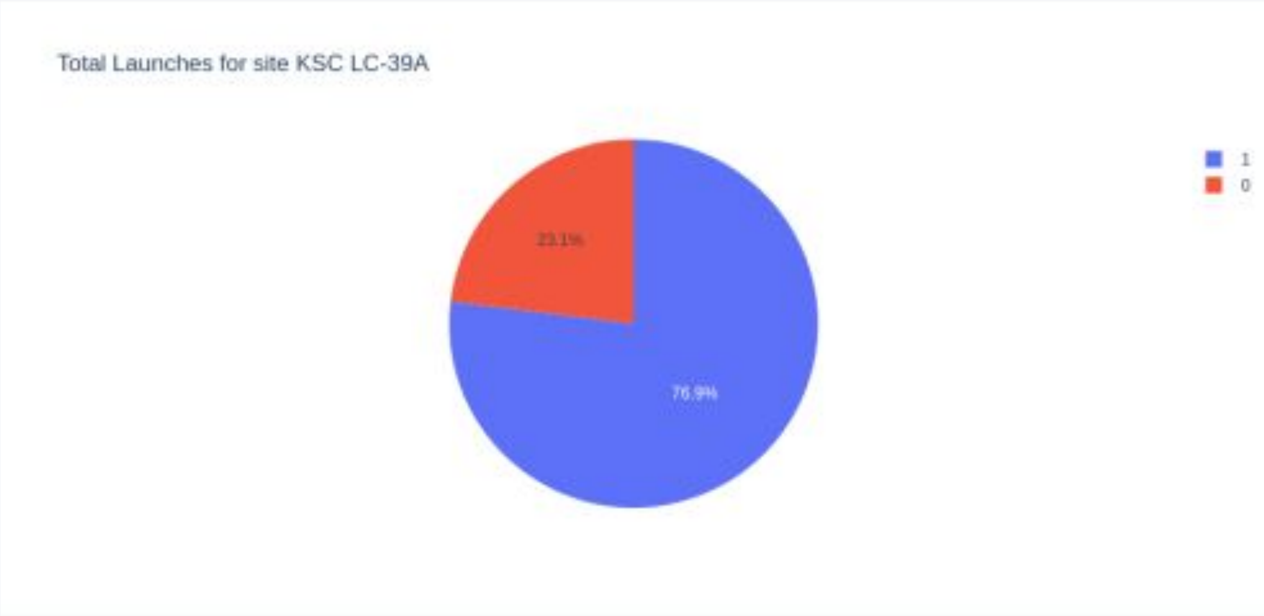
Build a Dashboard with Plotly Dash

Successful Launches by Site



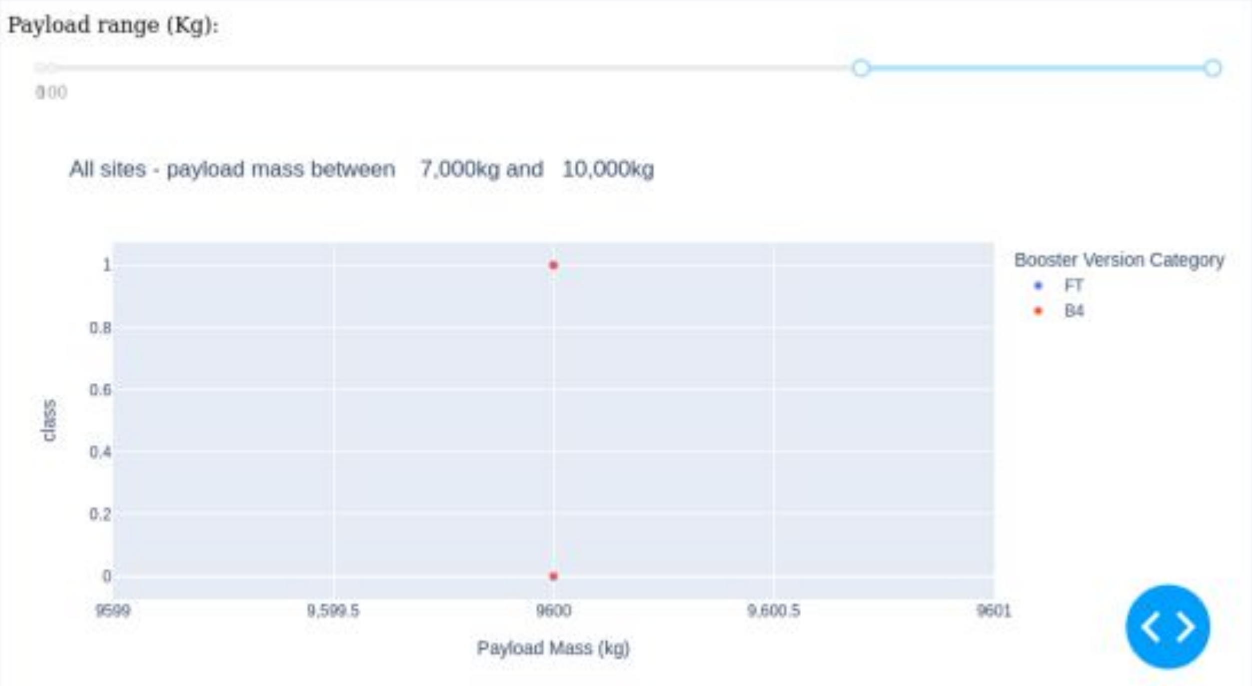
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



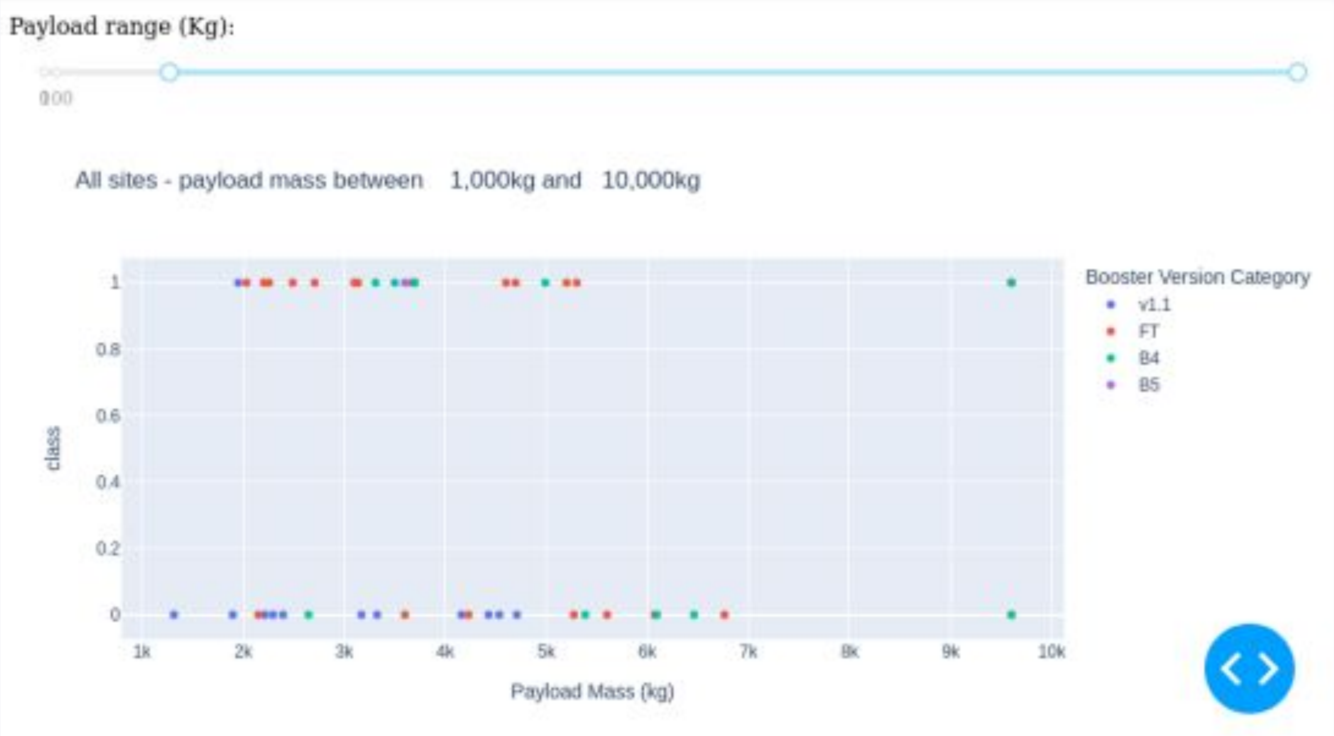
• 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



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

Payload vs. Launch Outcome



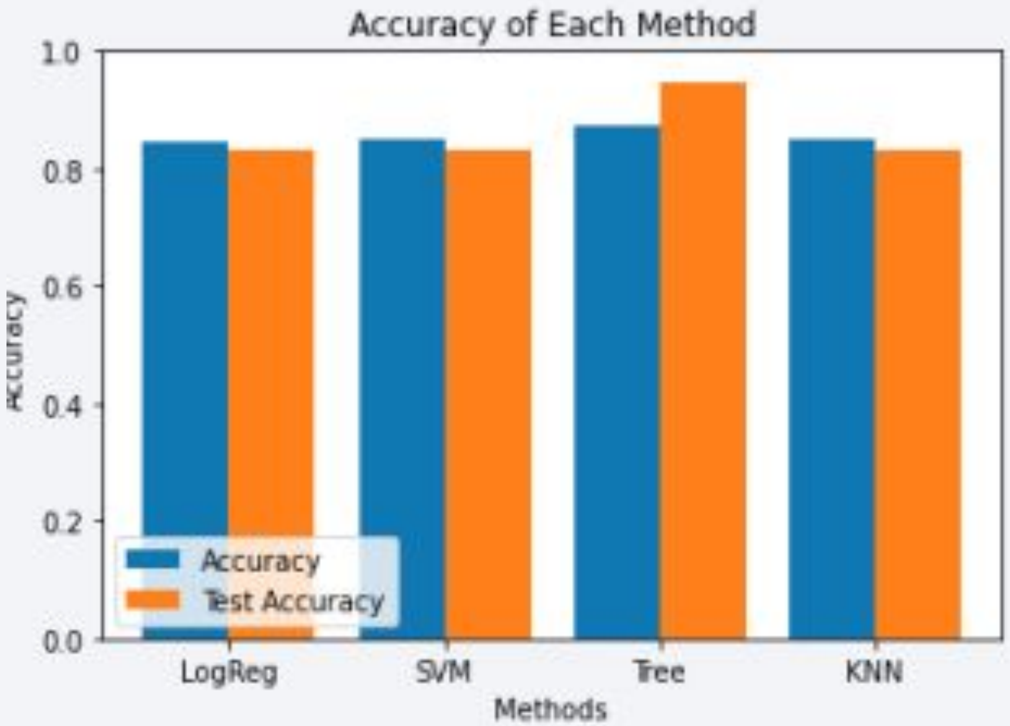
- Payloads under 6,000kg and FT boosters are the most successful combination.

Section 5

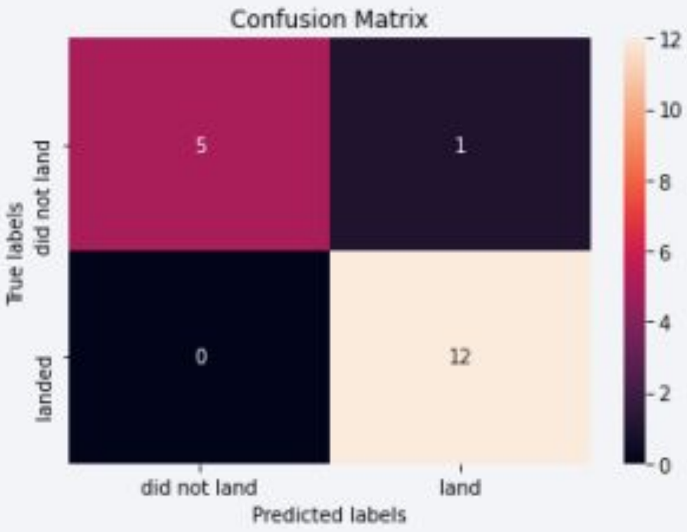
Predictive Analysis (Classification)

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.

Thank you!