IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Matija Kos
2022-07-28

Outline

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

Executive Summary

- Methodologies used to analyse data:
  - Data Collection
    - SpaceX API & WebScraping
  - Exploratory Data Analysis
    - Data wrangling, visualization and interactive visual analytics
  - Machine Learning Prediction
- Summary of all results:
  - Data collected from public sources
  - EDA – determined best features to predict success of launching
  - ML – best model to predict important characteristics

Introduction

- OBJECTIVE:
  - Evaluate the viability of the SpaceY to compete with SpaceX.

- DESIRABLE OUTCOME:
  - Best way to estimate the total cost of launches, by predicting successful landing of the first stage.
  - Best place to make launches.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - 2 sources: SpaceX API

    WebScraping
- Perform data wrangling.
  - Creating a landing outcome label based on summarizing and analyzing features.
- Perform Exploratory Data Analysis (EDA) using visualization and SQL.
- Perform interactive visual analytics using Folium and Plotly Dash.

Methodology

Executive Summary

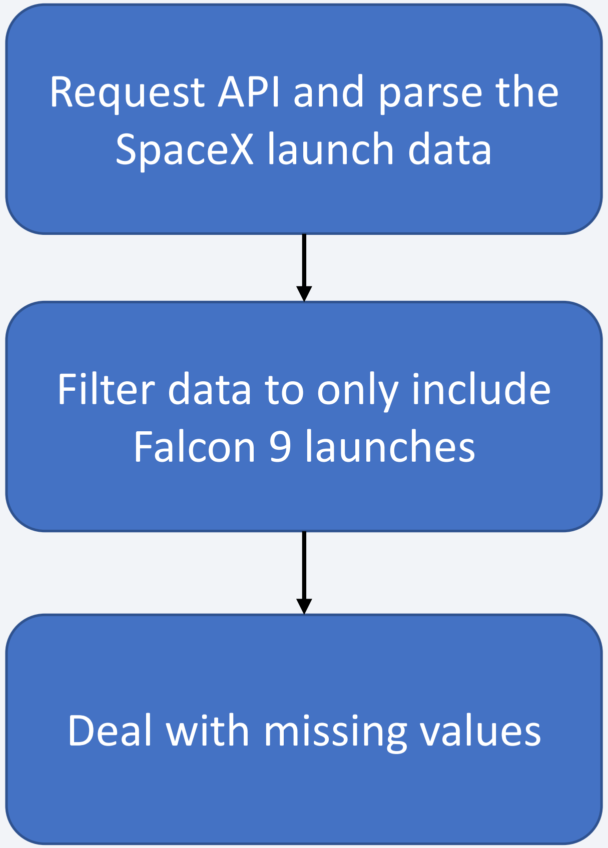
- Perform predictive analysis using classification models
  - Collected data was normalized, divided in training and testing data sets and evaluated by four different classification models. Accuracy of each model was evaluated using different combinations of parameters.

Data Collection

- Data sets were collected from:
  - SpaceX API (https://api.spacexdata.com/v4/rockets/)
  - Wikipedia using web scraping (https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922)

Data Collection - SpaceX API

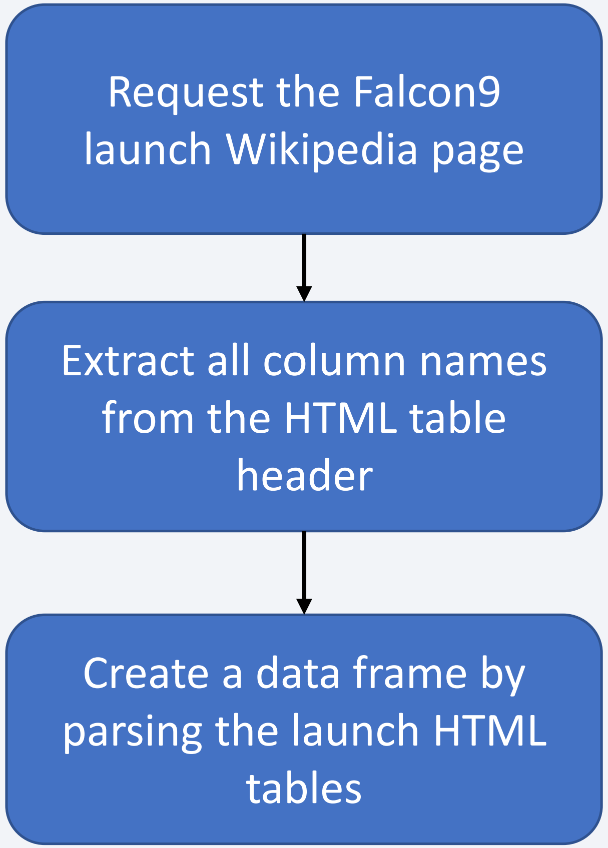
- SpaceX offers an API from where it is possible to obtain and use the dana.



Source:
https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/data%20collection%20with%20api.ipynb

Data Collection - Scraping

- Data from SpaceX launches can be obtained from Wikipedia.



Source:
https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/data%20collection%20with%20web%20scraping.ipynb

Data Wrangling

- Initial Exploratory Data Analysis was performed on the dataset.
- Calculated summaries of:
    - Launches per site
    - Occurrences of each orbit
    - Mission outcome per orbit
- Landing outcome label was created.

Source:
https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/data%20wrangling.ipynb

EDA with Data Visualization

- Scatterplots and barplots were used in order to visualize the relationship between features:

FEATURE 1	FEATURE 2
Payload Mass	Flight Number
Launch Site	Flight Number
Launch Site	Payload Mass
Orbit	Flight Number
Payload	Orbit

Source:
https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/eda%20with%20visualization.ipynb

EDA with SQL

- SQL queries that 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 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 landing and have payload mass between 4000 and 6000 kg
  - Total number of successful and failure mission outcomes
  - Names of the booster version which have carried the maximum payload mass
  - Failed landing outcome in drone ship, their booster version, and launch site names in year 2015
  - Count of landing outcomes between dates 04-06-2010 and 20-03-2017

Source: https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/eda%20with%20sql.ipynb

Build an Interactive Map with Folium

- Map objects used: markers, circles, lines, marker clusters
  - Markers for indicating points (e.g. launch sites).
  - Circles for indicating areas around specific coordinates (e.g. NASA Johnson Space Center).
  - Marker clusters for indicating groups of events in coordinate (e.g. launches in a launch site).
  - Lines for indicating distances between two coordinates.

Source: https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/interactive%20visual%20analytics%20with%20folium.ipynb

Pictures: https://github.com/Kos360/Applied-Data-Science-Capstone/tree/main/pics%20for%20interactive%20visual%20analytics%20with%20folium

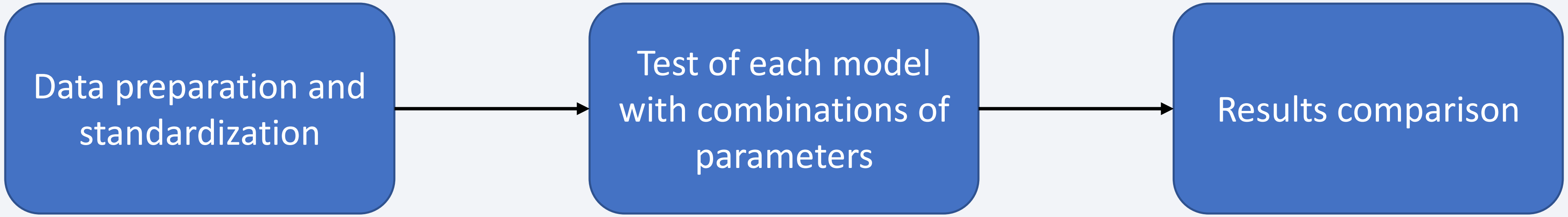
Build a Dashboard with Plotly Dash

- Graphs and plots used to visualize dana:
  - Percentage of launches by site
  - Payload range
- Those plots and interactions were used to analyze the relation between payloads and launch sites in order to help identify where is the best place to launch according to payloads.

Source: https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/dash%20app.py

Predictive Analysis (Classification)

- Compared classification models:
  - Logistic regression
  - Support vector machine (SVN)
  - Decision tree
  - K nearest neighbors (kNN)



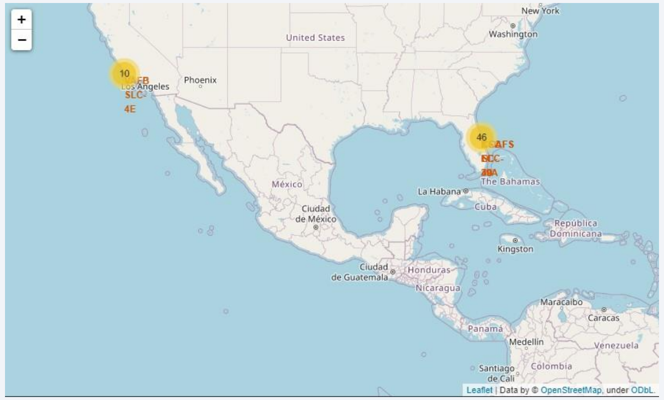
Source: https://github.com/Kos360/Applied-Data-Science-Capstone/blob/main/machine%20learning%20prediction.ipynb

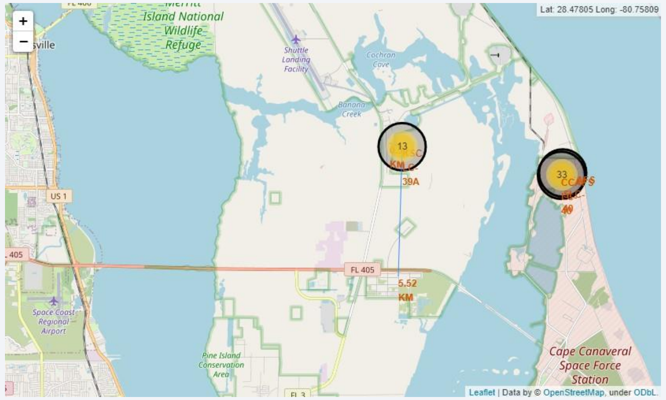
Results

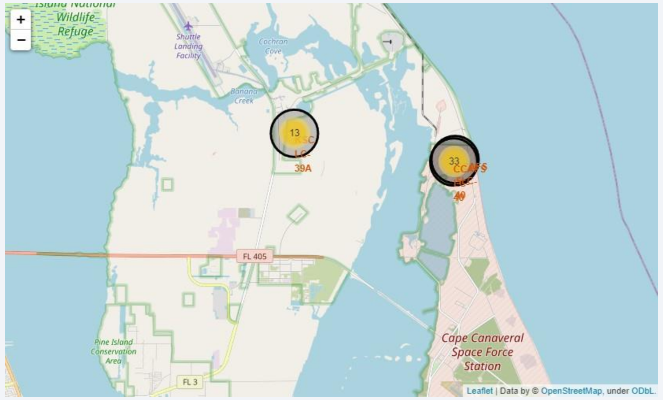
- Exploratory data analysis results:
    - 4 different launch sites are used by SpaceX,
    - The average payload of F9 v1.1 booster is 2,928 kg,
    - The first successful landing outcome happened in 2015,
    - Many Falcon 9 booster versions were successful at landing on drone ships having payload above the average,
    - ~100% mission outcomes were success,
    - Two booster versions failed at landing on drone ships in 2015,
    - The number of landing outcomes became better as years passed.

Results

- Interactive analytics demo in screenshots:
  - Launch sites are in safety places (near sea) and have a good logistic infrastructure.

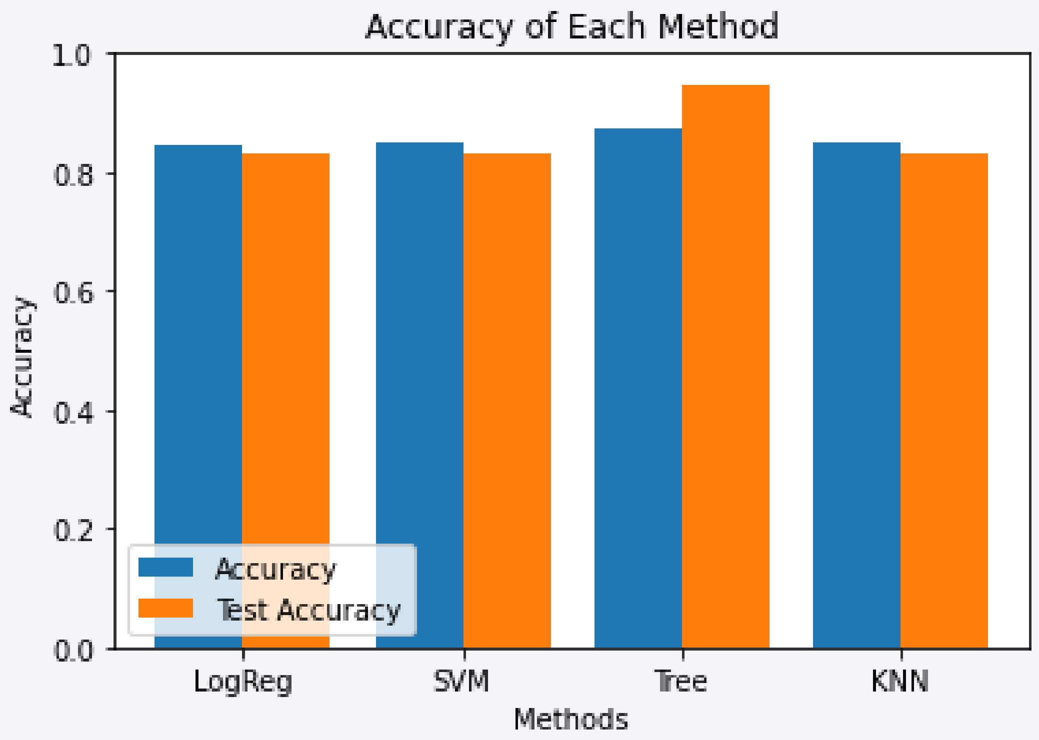






Results

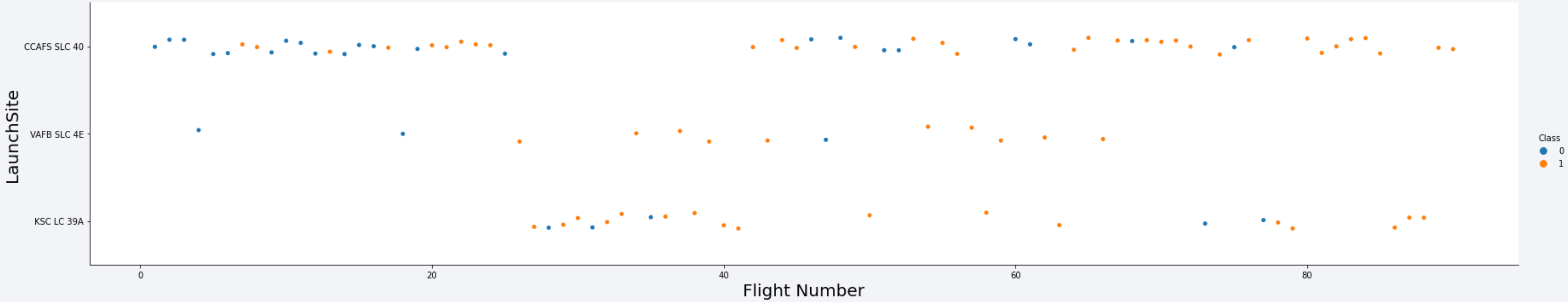
- Predictive analysis results:
  - Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landing with accuracy of 87% and test data accuracy of 94%.



Section 2

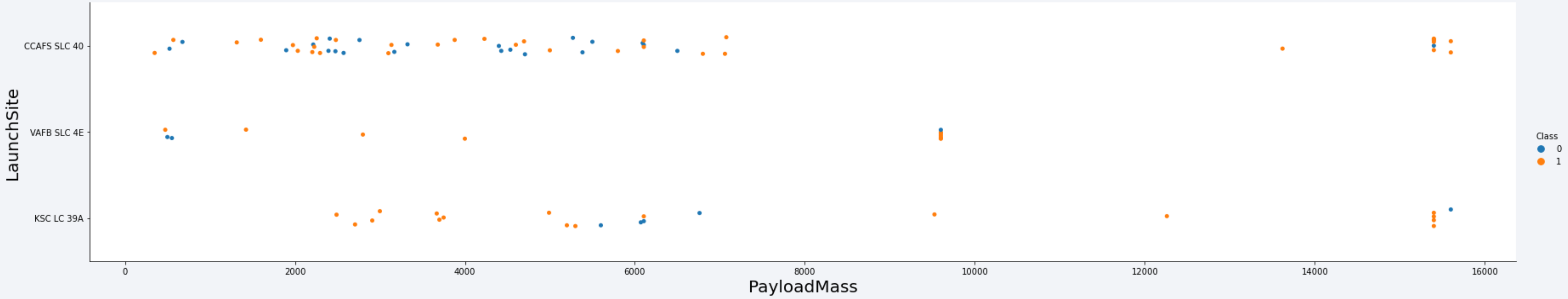
Insights drawn from EDA

Flight Number vs. Launch Site



- According to the plot, it is possible to deduct that the best launch site is CCAF5 SLC 40.
- VAFB SLC 4E is in second place and KSC LC 39A in third.
- General success rate improves over time.

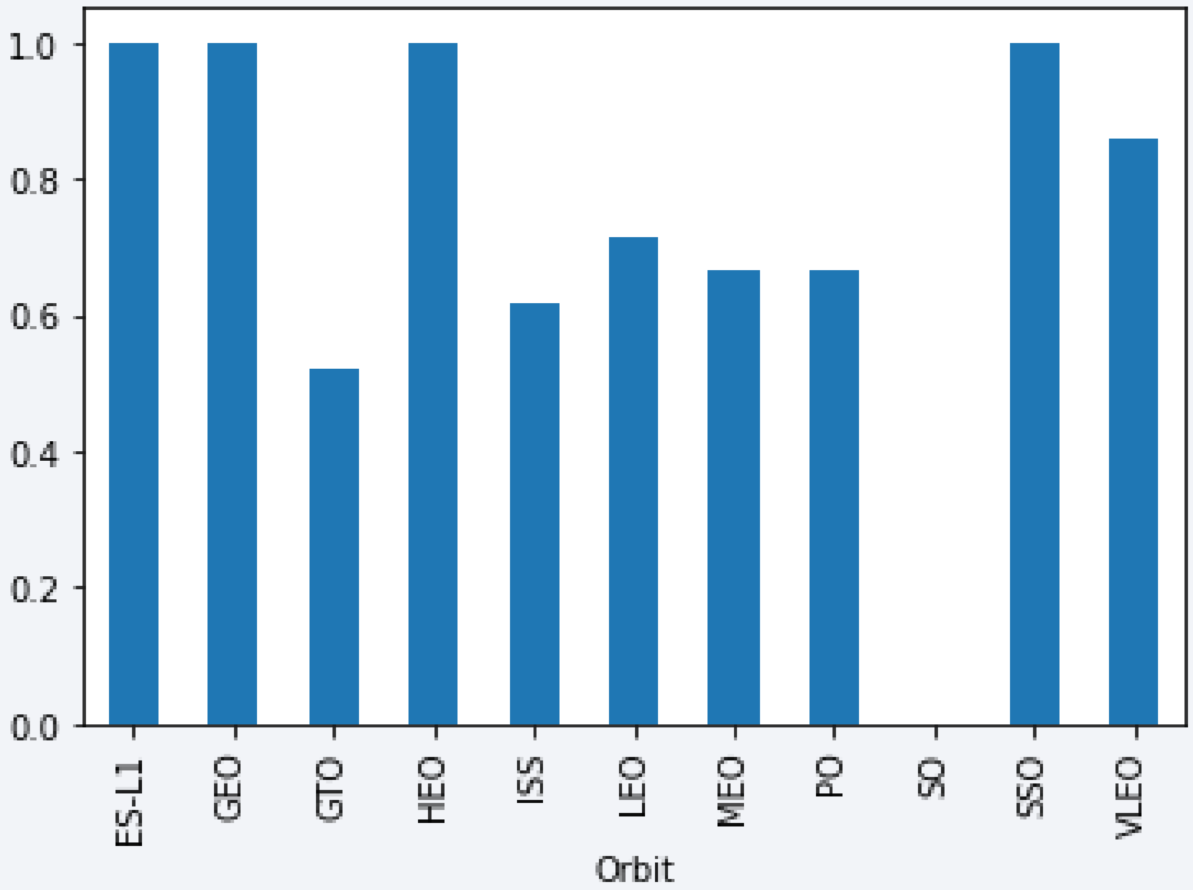
Payload vs. Launch Site



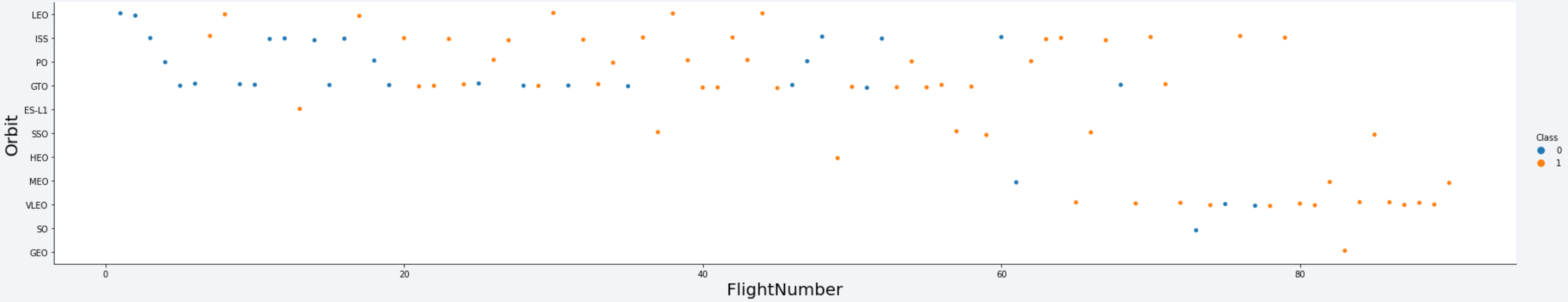
- Payloads over 9,000 kg have excellent success rate.
- Payloads over 12,000 kg are possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type

- The best success rates:
  - Orbit ES-L1
  - Orbit GEO
  - Orbit HEO
  - Orbit SSO

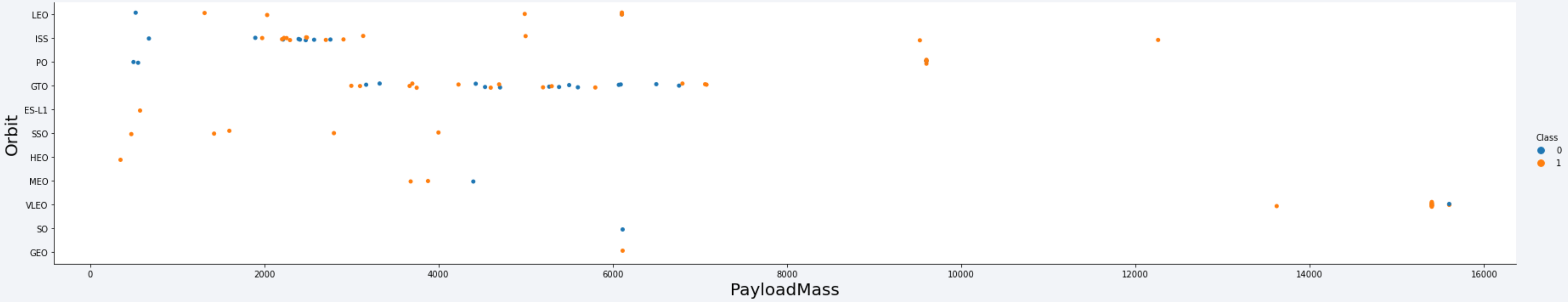


Flight Number vs. Orbit Type



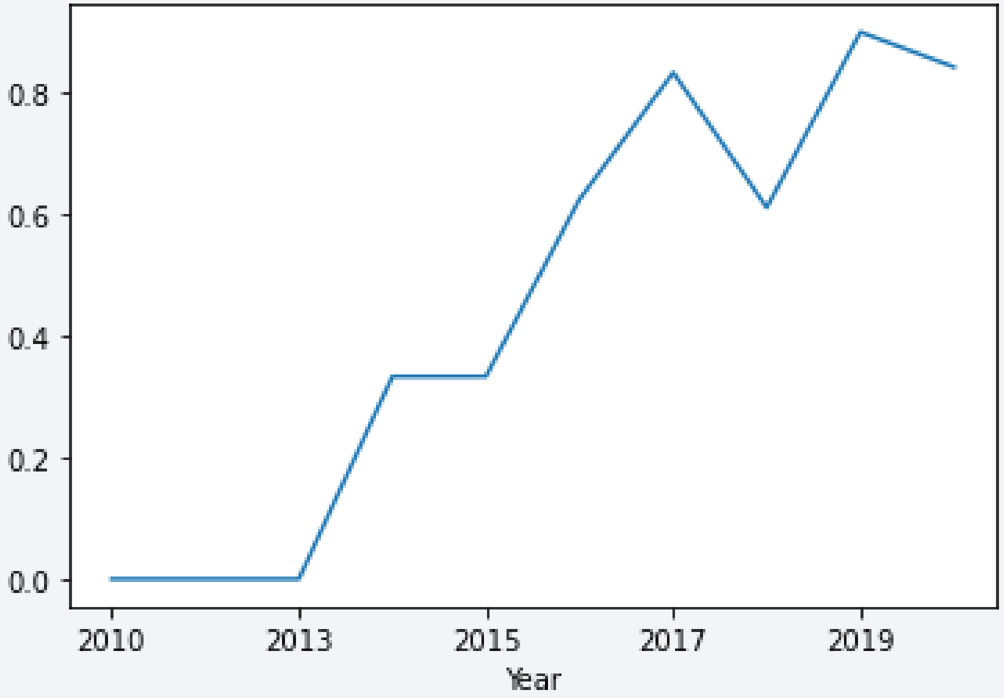
- Success rate improves over time on all orbits.
- VLEO orbit is possibly a opportunity, due to recent increase of its frequency.

Payload vs. Orbit Type



- There is no relation between payload and success rate to orbit GTO.
- ISS orbit has the widest range of payload and good rate of success.

Launch Success Yearly Trend



- Success rate started to increase in 2013 and kept rising until 2020.
- This could be the cause of adjusting and improving the technology in the first three years.

All Launch Site Names

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

LAUNCH SITE
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- They have been obtained by selecting unique occurrences of “launch_site” values.

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with “CCA”:

DATE	TIME	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 LS-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Total payload carried by boosters from NASA:

TOTAL PAYLOAD (kg)
111.268

- It is calculated by summing all payloads whose codes 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

- Average payload mass was calculated from filtered data by the booster version.

First Successful Ground Landing Date

- First successful landing outcome on ground pad:

MIN DATE
2015-12-22

- First successful landing outcome on ground pad was derived from filtering data by successful landing outcome on ground pad and then finding the minimum value.

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

- Select distinct booster version with filters:
  - Successful landing on drone ship
  - Payload mass in interval [4000, 6000]

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

- Groped mission outcome and counted records for each group.

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
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:

BOOSTER VERSION	LAUNCH SITE
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 04-06-2010 and 20-03-2017

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

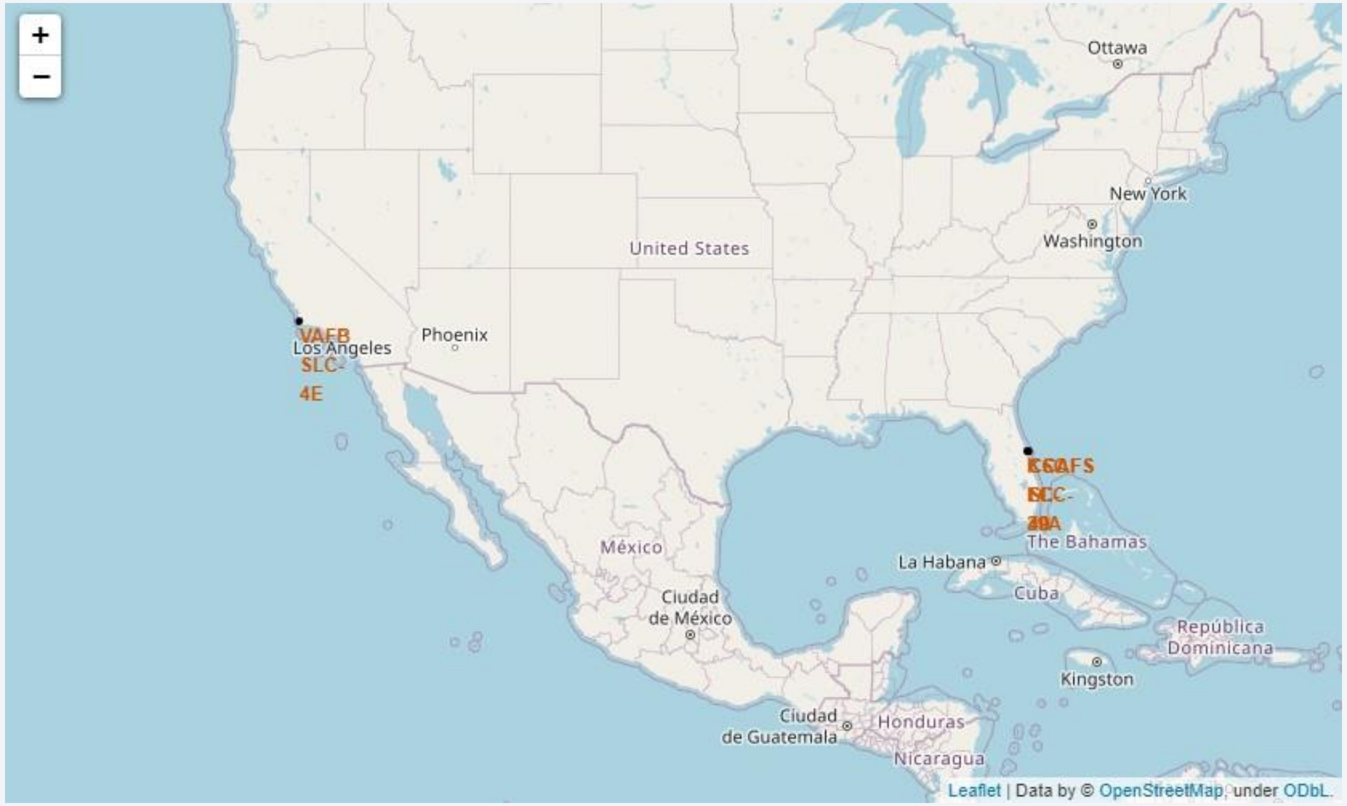
LANDING OUTCOME	OCCURRENCES
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- *"No attempt"* must be taken in consideration

Section 3

Launch Sites Proximities Analysis

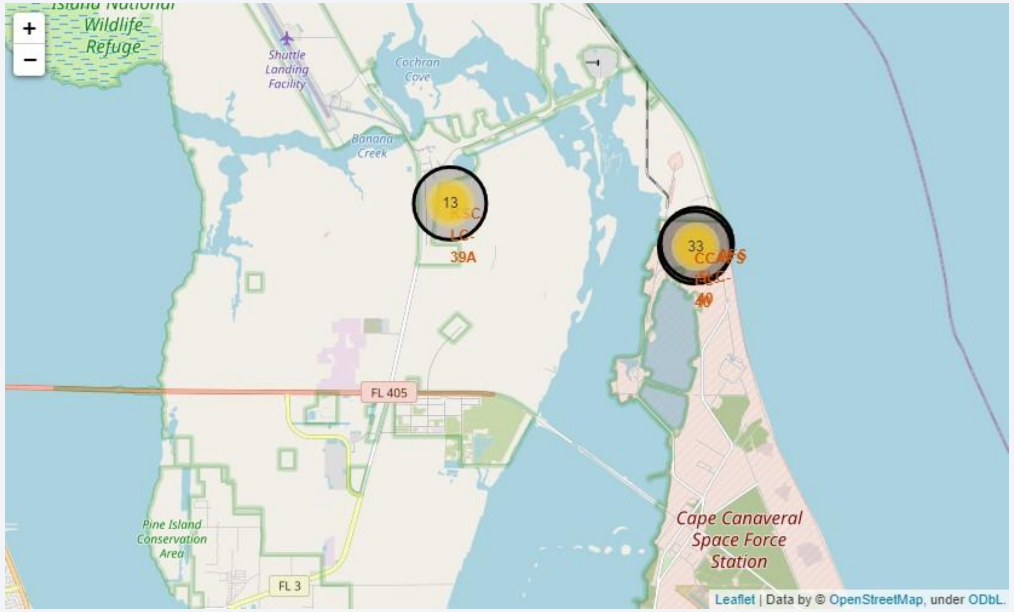
All launch sites



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

Launch Outcome by Site

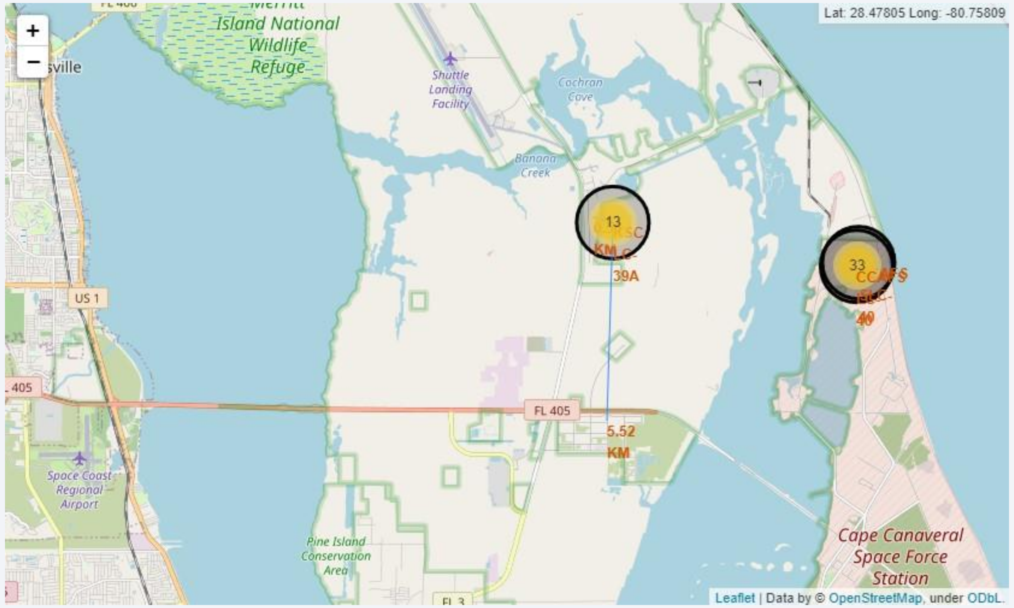
- Example of KSC LC-39A launch site outcomes:



- Green markers- success
- Red markers- failure

Logistics and Safety

- Screenshot of a selected launch site to its proximities:

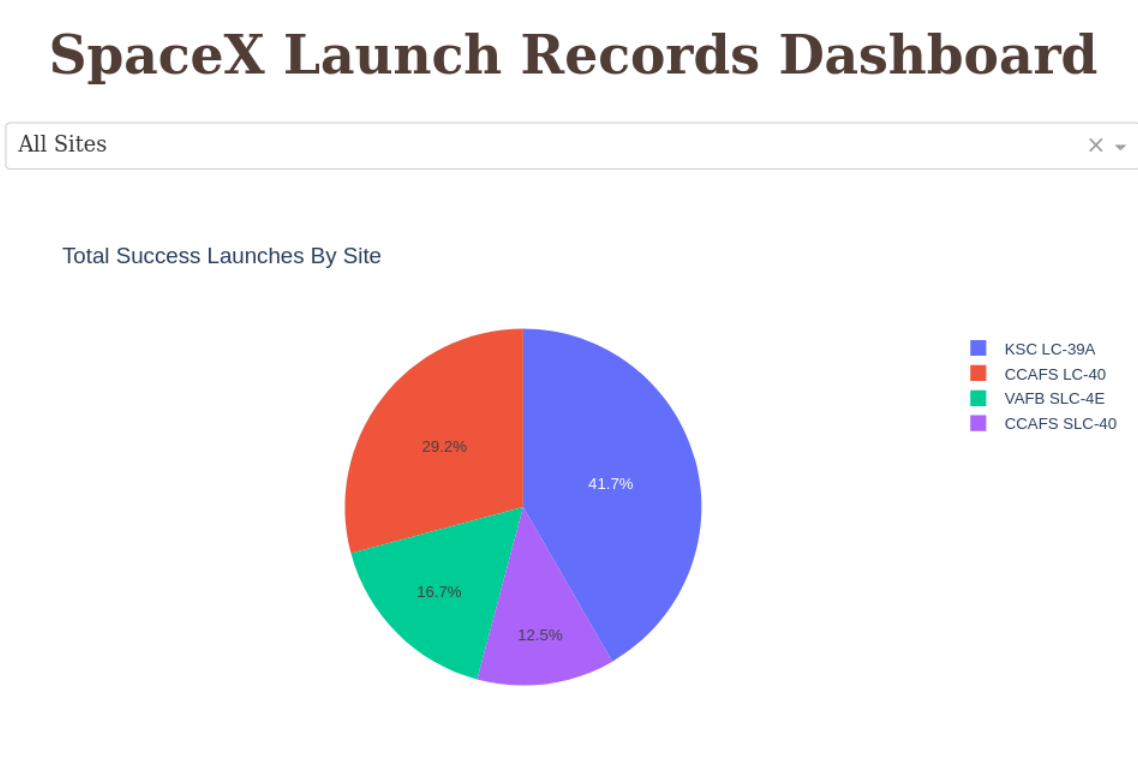


- Launch sites KSC LC-39A is logistically well connected and it`s far from populated areas.

Section 4

Build a Dashboard with Plotly Dash

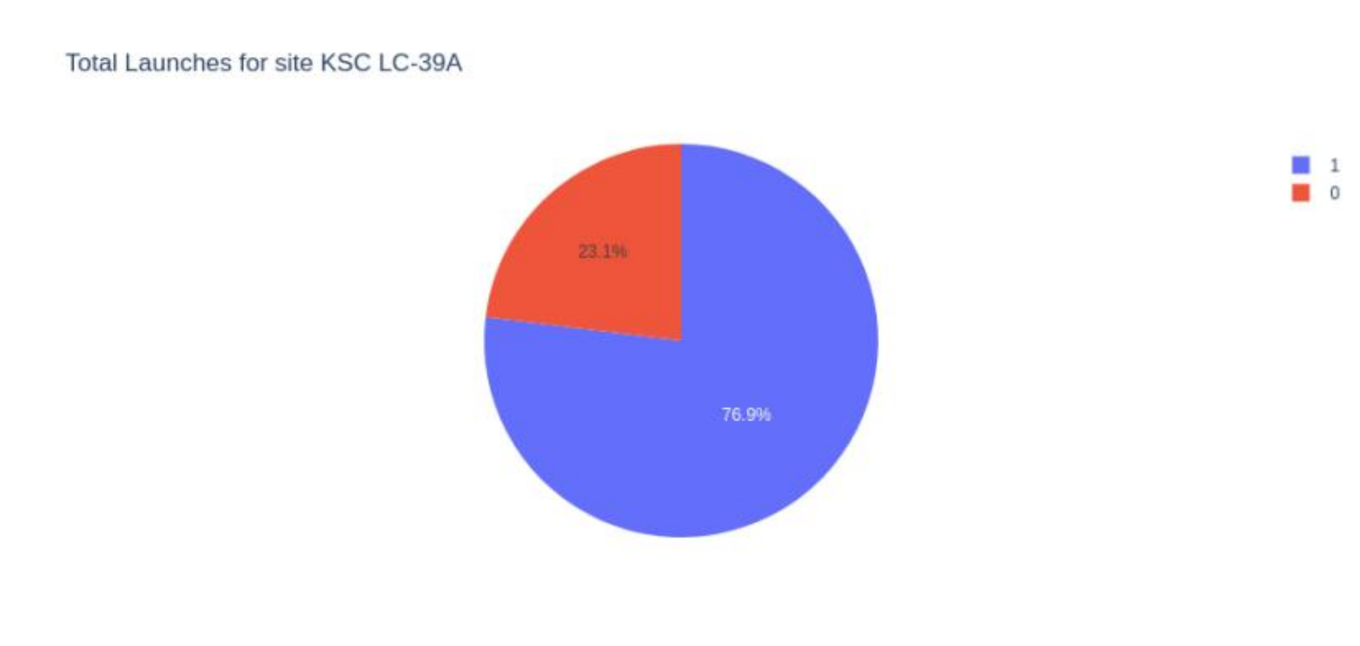
Successful Launches by Site



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

Launch Success Ratio for KSC LC-39A

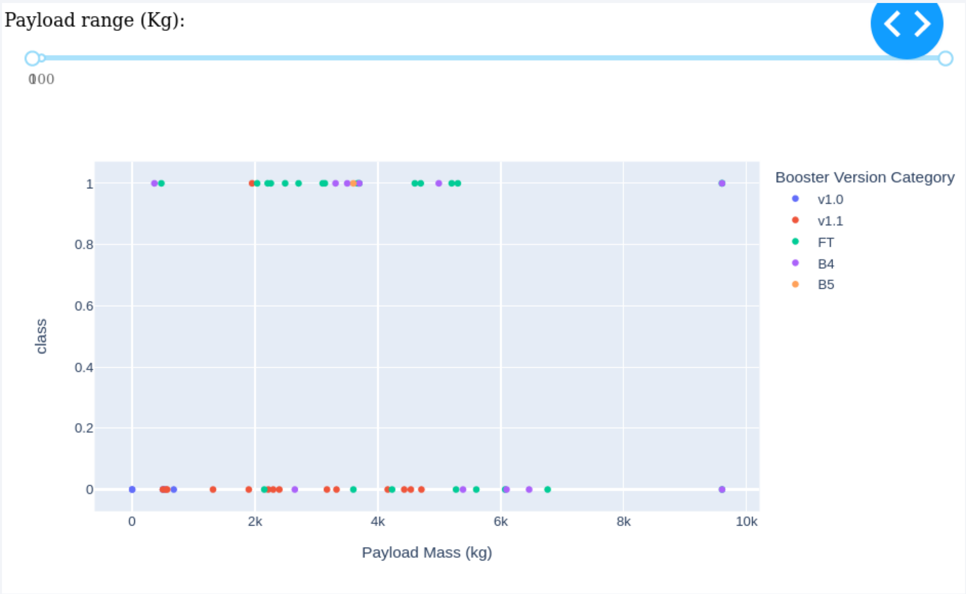
- Screenshot of the piechart for the launch site with highest launch success ratio:



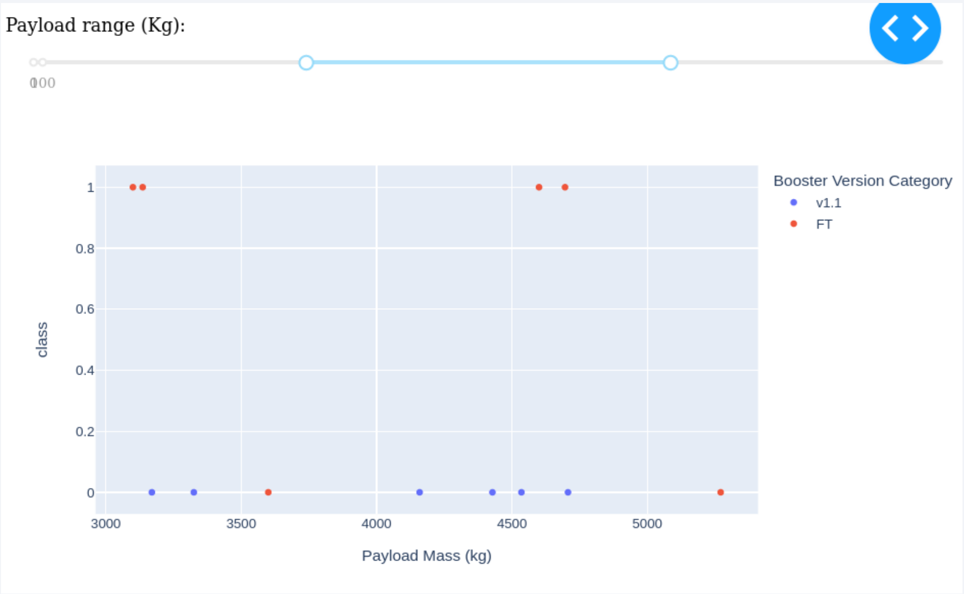
- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome

- Screenshots of Payload vs. Launch Outcome, with different payload selected:



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



There`s not enough data to estimate the risk of launches over 7,000 kg.

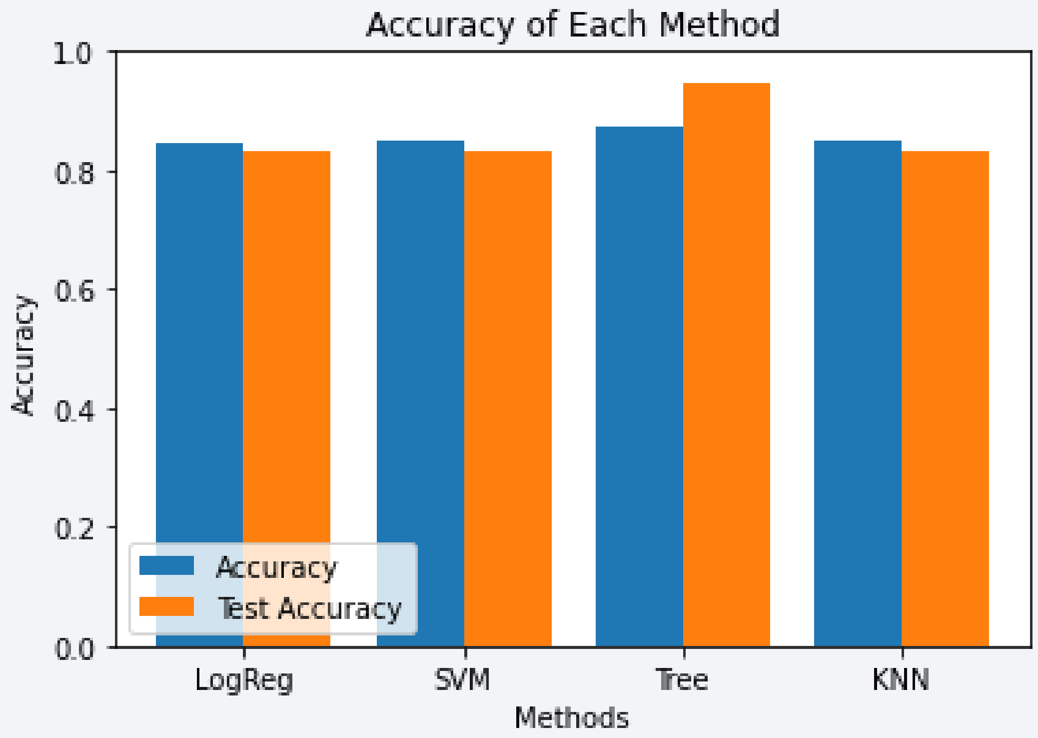
Section 5

Predictive Analysis (Classification)

Classification Accuracy

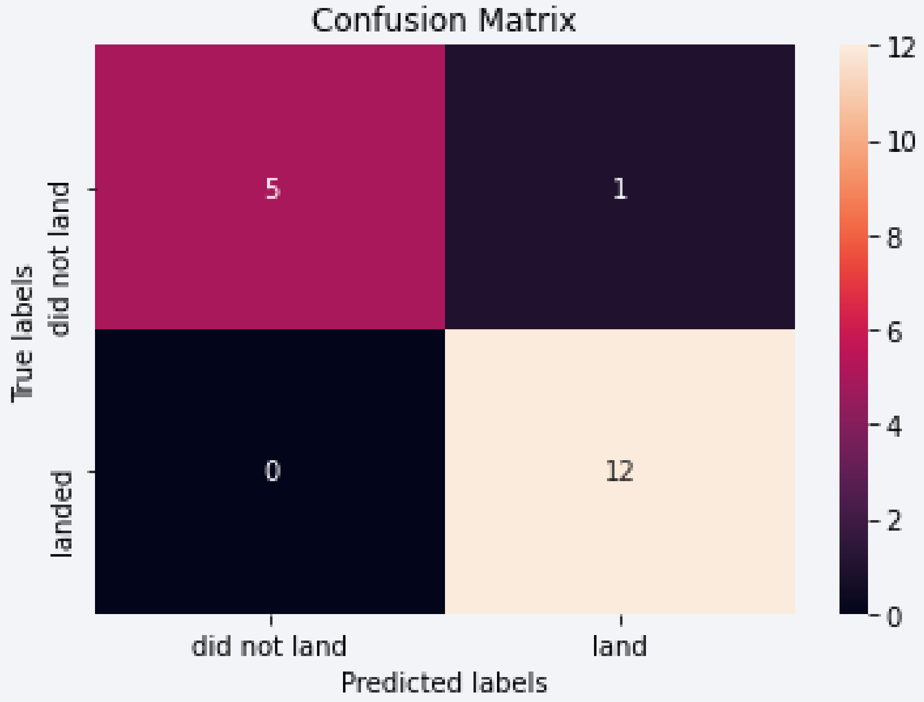
- Four classification models were tested and their accuracies plotted.

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



Confusion Matrix of Decision Tree Classifier

- Confusion matrix of Decision Tree Classifier:



Conclusions

- Different data sources were analyzed.
- The best launch site is KSC LC-39A.
- Although most of mission outcomes are successful, landing outcomes seem to improve over time.
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Thank you!