SpaceX with Data Science

Theresa
31 Aug 2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Methodologies used for data analysis as below:
  - Data Collection using web scraping and SpaceX API
  - Exploratory Data Analysis (EDA) that include data wrangling, data visualization and interactive visual analytics.
  - Machine Learning Prediction.
- Summary of all results:
  - Collect valuable data from public sources is possible.
  - EDA able to identify the best features to predict success of launchings.
  - Machine Learning Prediction able to use the collected data and showed the best model of prediction which is important to drive the the best opportunity.

Introduction

- The objective is to evaluate the viability between Space Y and Space X.
- Desirable result:
  - The prefer mode to estimate the total cost for launches, by predicting successful and safer landings of the first stage of rockets.
  - Where is the best location to make launches.

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Date from Space X was obtained from 2 sources
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Collected data was enhanced by creating a landing outcome label based on outcome data after summarising and analysing features.
- Perform exploratory data analysis (EDA) using visualisation and SQL.

Methodology

Executive Summary

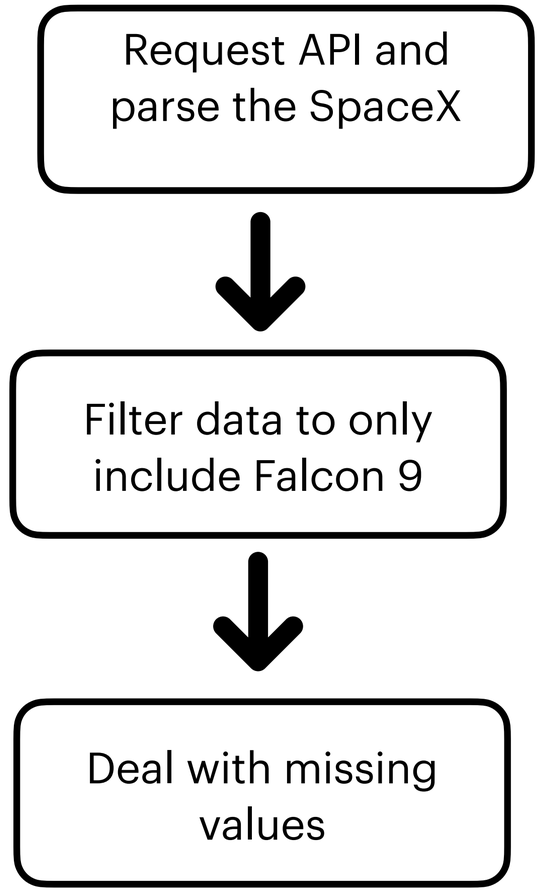
- Perform interactive visual analytic by using Plotly Dash and Folium.
- Perform predictive analysis by using classification models.
  - Data that collected were normalised, divided in training and test data sets and evaluated by 4 classification models. It being evaluated the accuracy of each model by using various combinations of parameters as well.

Data Collection

- Data sets were collected from SpaceX API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches) by using web scraping technics.

Data Collection - SpaceX API

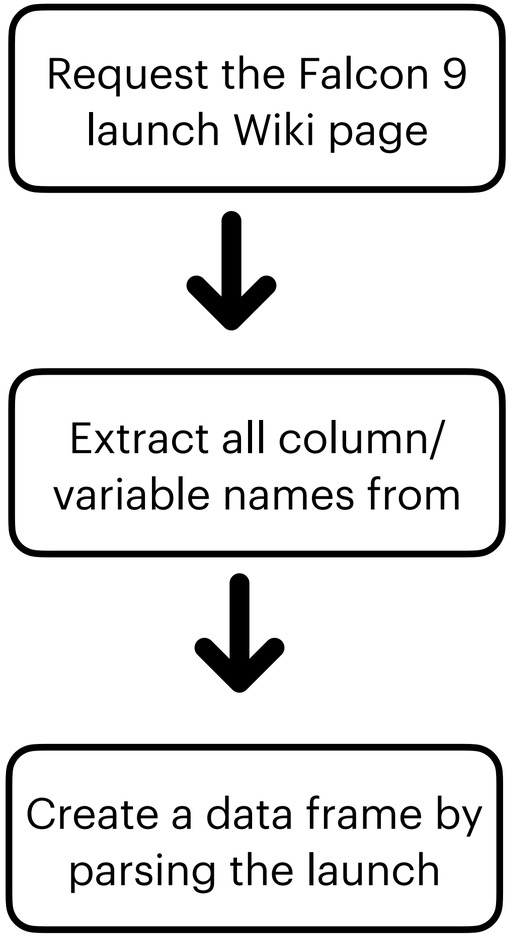
- Spacex provides a public API from where data can be obtained and then used.
- The flowchart beside and the data persist in the link below was used the API.



- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/Data%20Collection%20API.ipynb

Data Collection - Scraping

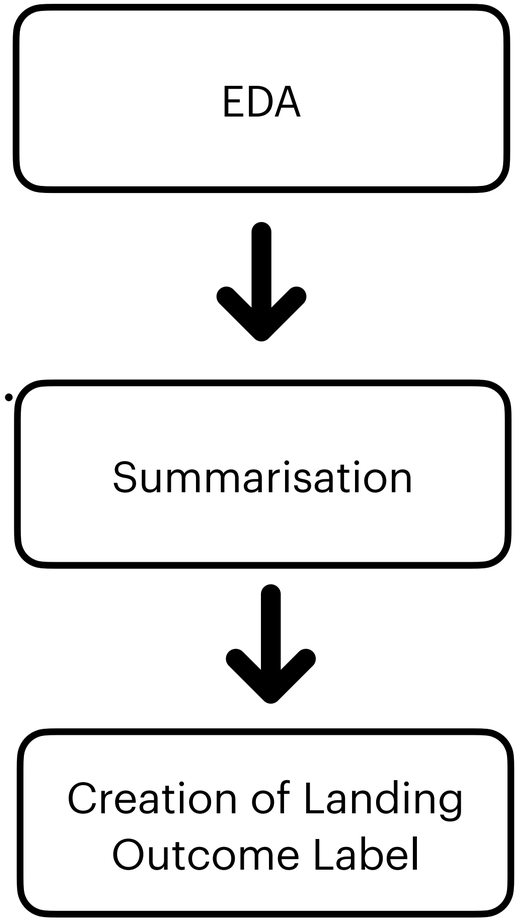
- Data from SpaceX launches can also be obtained from Wikipedia.
- Data downloaded from Wikipedia is according to the flowchart and persisted in the link below.



- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

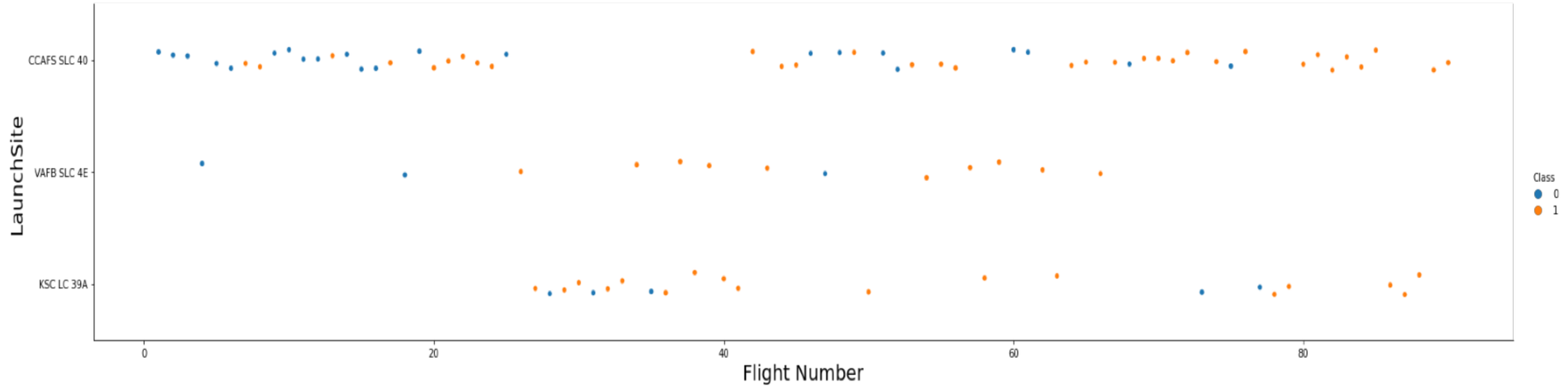
1. Exploratory Data Analysis (EDA) was performed on the dataset.
2. Summaries launches per site, happen of each orbit and occurrences of mission outcome per orbit type were calculated.
3. The landing outcome level was created from Outcome column.



- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/Data%20Wrangling.ipynb

EDA with Data Visualisation

- To explore data, scatterplots and bar plots were used to visualise the relationship between pair and 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/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- Following are SQL queries performed:
  - Names of the unique lunch site in the space mission
  - Top 5 launch sites whose name begin with the string ‘CCA’
  - Total payload mass carried by booster launched NASA (CRS)
  - Average payload mass carried by booster version F9 v 1.1
  - Date when the 1st successful landing outcome on ground pad was achieved
  - Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kgtotal 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 year 2015
  - Rank of the landing outcome count (such as Failure ‘drone ship’ or Success ‘ground pad’) between the date 4 Jun 2010 and 20 Mar 2017.
- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/EDA.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 indicate groups of events in each coordinate, like launches in a launch site
  - Lines are used to indicate distances between two coordinates

- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

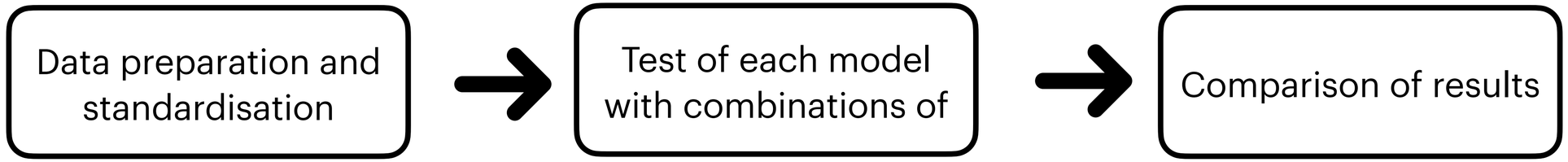
Build a Dashboard with Plotly Dash

- The following elements were used to visualise data
  - Percentage of launches by site
  - Payload range
- This combination allowed to analyse the relation between payloads and launch sites, helping to identify the best place to launch according to payloads.

- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/spacex_dash_app.py

Predictive Analysis (Classification)

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



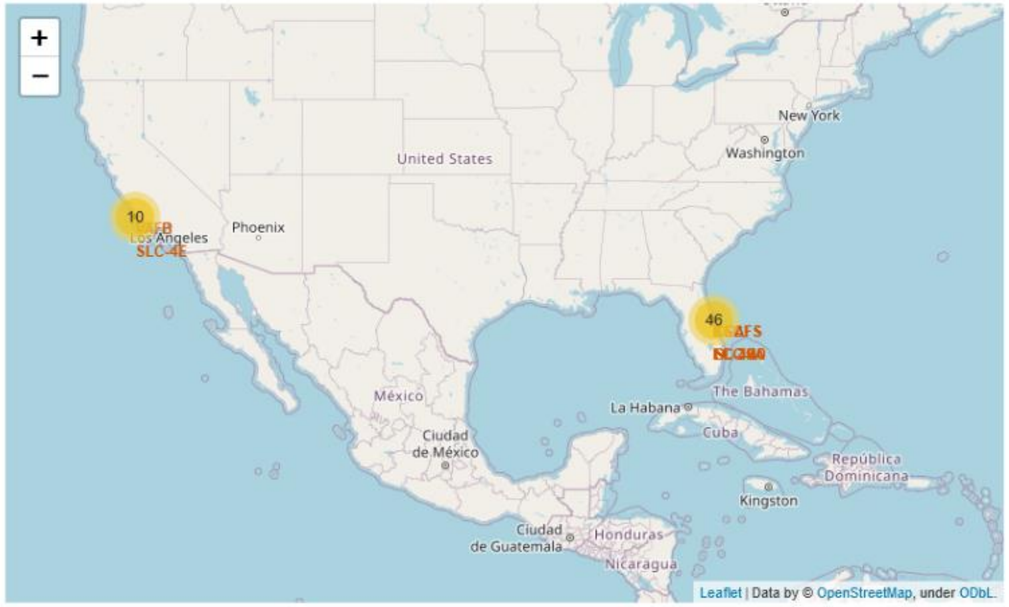
- Source code: https://github.com/Theresa20191/IBM-Applied-Data-Science-Capstone/blob/1eb0f60175343494fe69045e861b89346779d5bd/Machine%20Learning%20Prediction.ipynb

Results

- Exploratory data analysis results:
  - Space X uses 4 launch sites
  - The 1st launches were done to Space X itself and NASA
  - The average payload of F9 v 1.1 booster is 2,928 KG
  - The 1st success landing result happen in year 2015
  - Many Falcon 9 poster 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 this in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
  - The number of landing outcomes became as better as years passed.

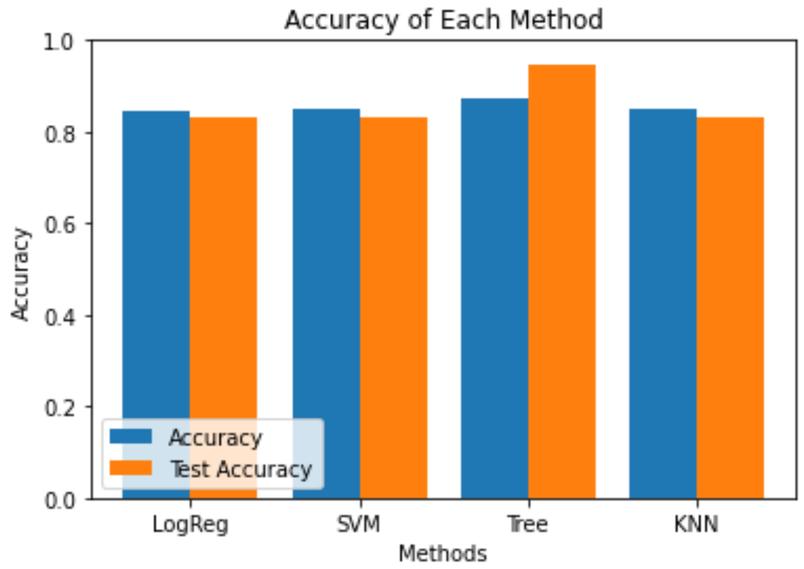
Results

- 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 launched happens at east cost launch sites.



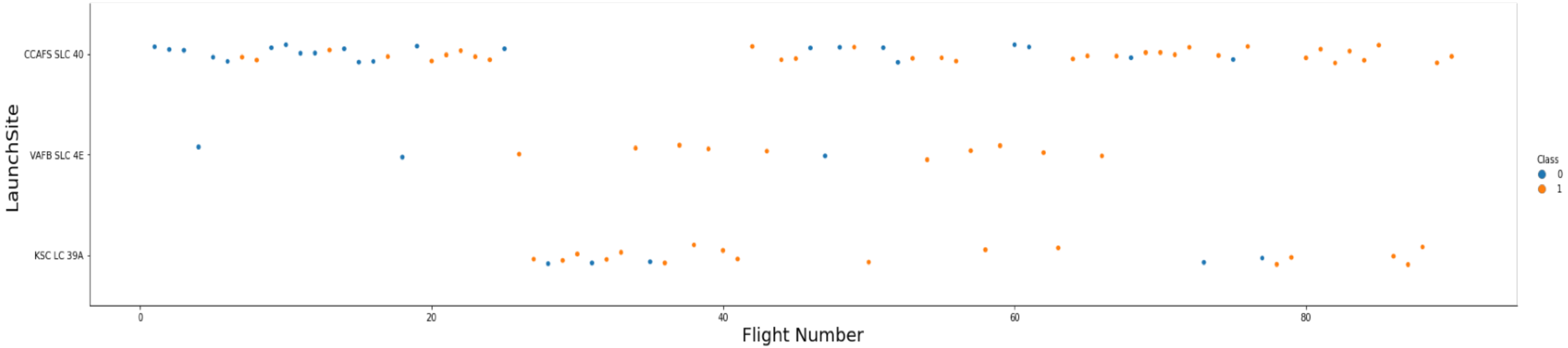
Results

- 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%.



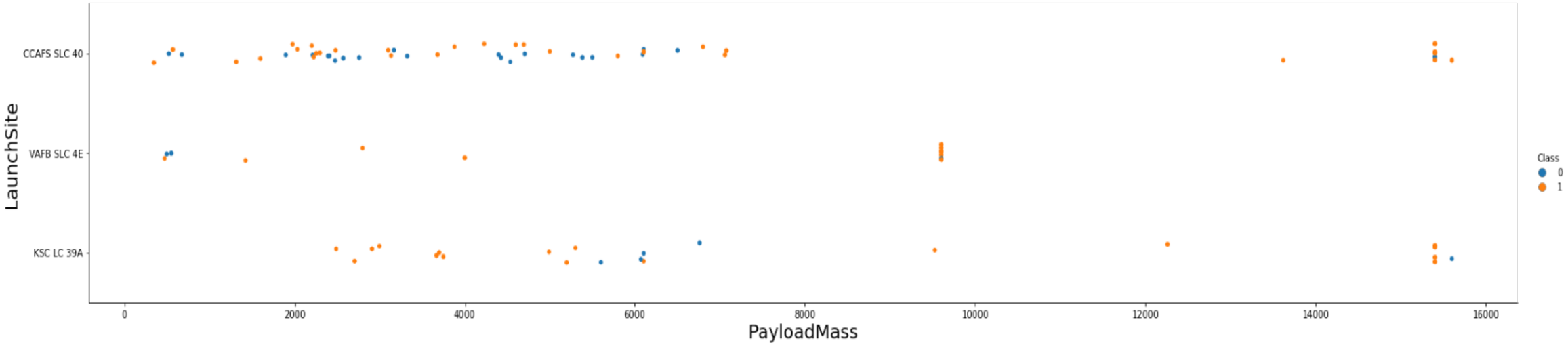
Insights drawn from EDA

Flight Number vs Launch Site



- Based on the plot above, it is to verify the best launch site is CCAF5 SLC 40, where is most of the launches were successfully
- The 2nd place is VAFB SLC 4E and the 3rd place is MSc lc 39a
- We can see that the general success rate has improved over time.

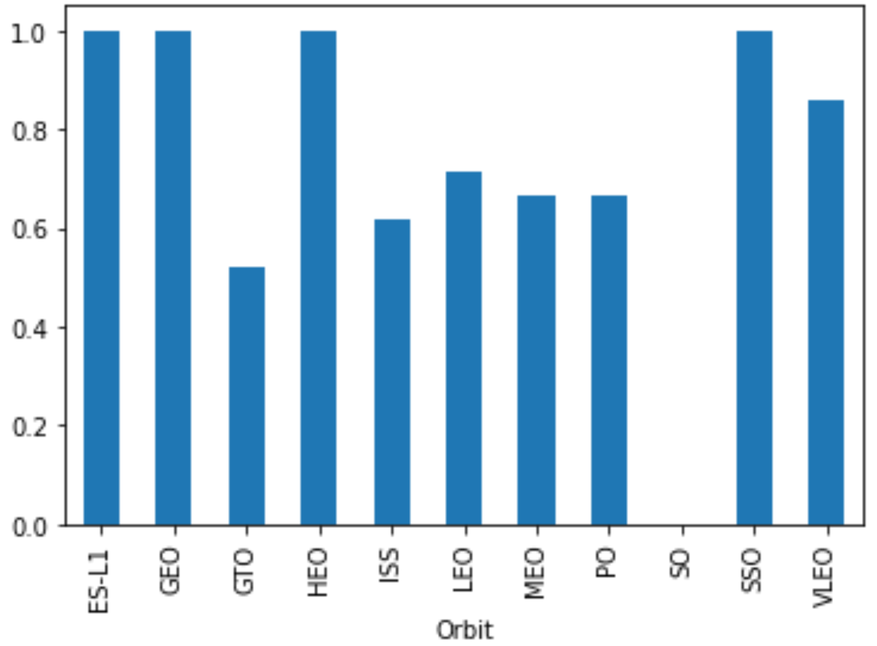
Payload vs Launch Site



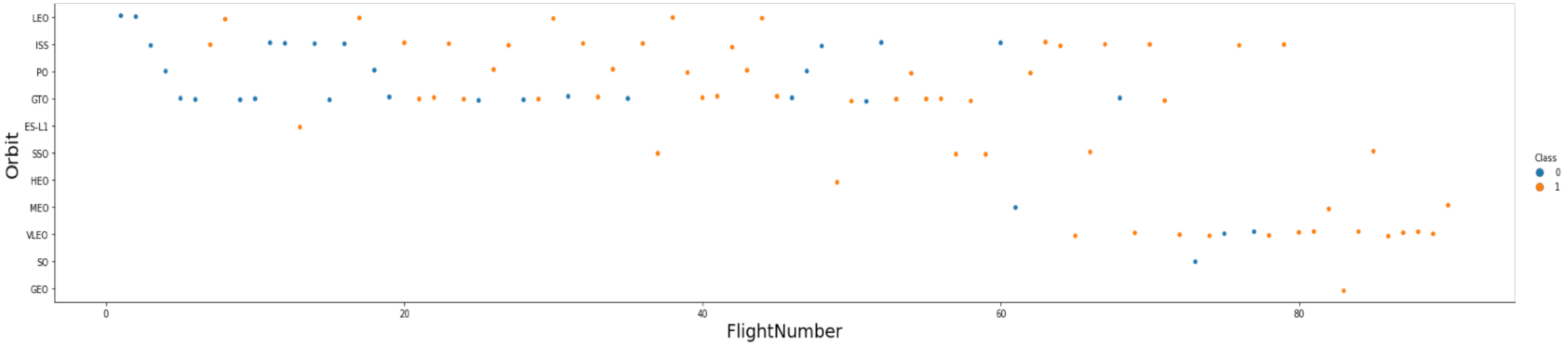
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate
- Payloads over 12,000kg seems only possible land on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs Orbit Type

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

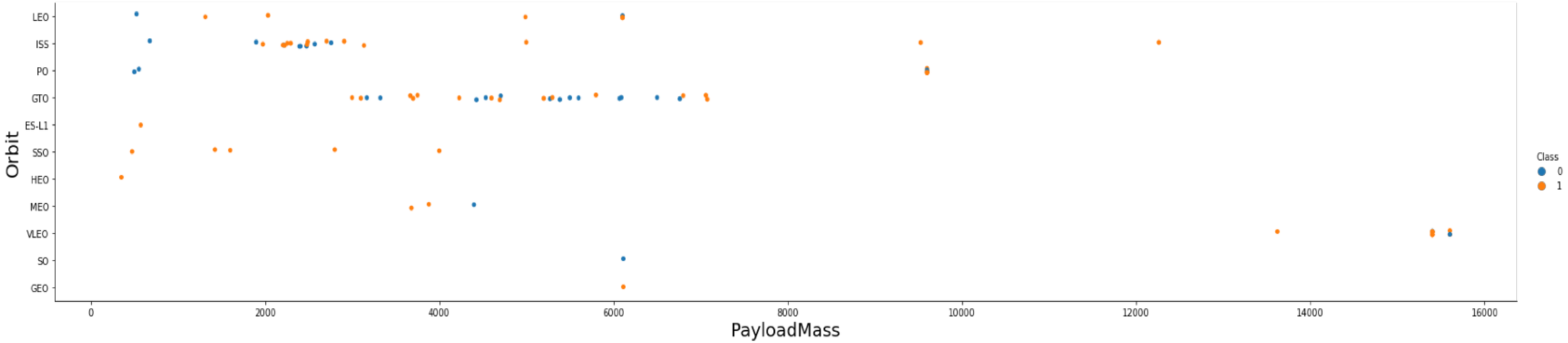


Flight Number vs Orbit Type



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

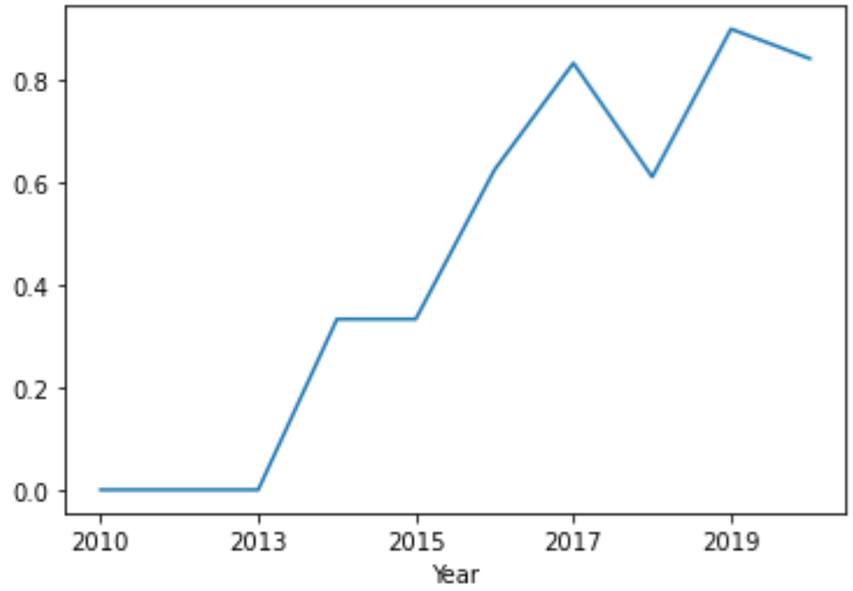
Payload vs Orbit Type



- Apparent, 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 1st 3 years were a period of adjust and improvement of technology.



All Launch Site Names

- According to data, there are 4 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	**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, 2 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 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 is by summing all payloads whose codes contain NASA ‘CRS’.

Average Payload Mass by F9 v 1.1

- Average payload mass carried by booster version F9 v1.1:

Avg Payload (kg)
2.928

- Filtering data by the booster version above and calculating 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 1st occurrence that happened on 22 Dec 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

- Select distinct booster versions according to the filters above, which is the result as above show.

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.

Booster 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 Sites
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- The list above has the only 2 occurrences.

Rank Landing Outcomes Between 4 Jun 2010 & 20 Mar 2017

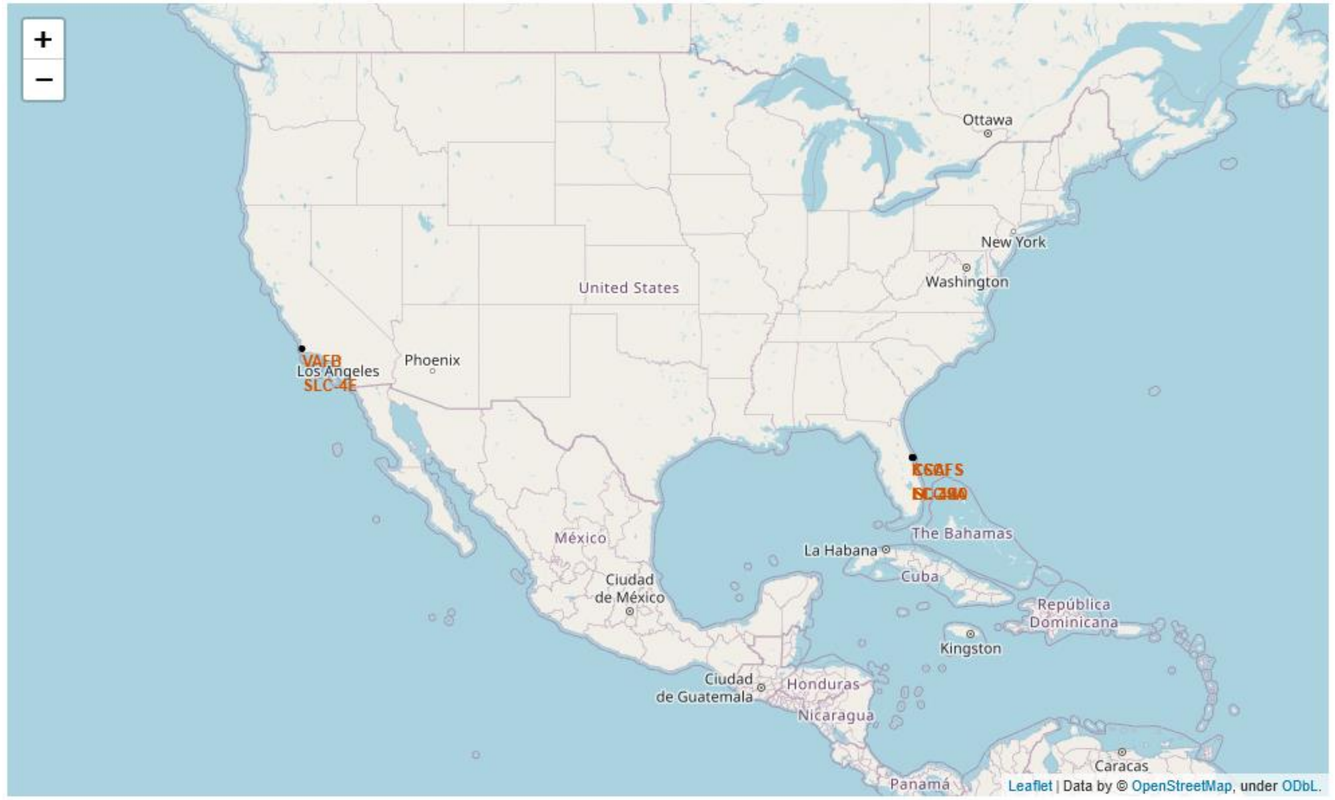
- Ranking of all landing outcomes between the date 4 Jun 2010 and 20 Mar 2017

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

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

Launch Sites Proximities Analysis

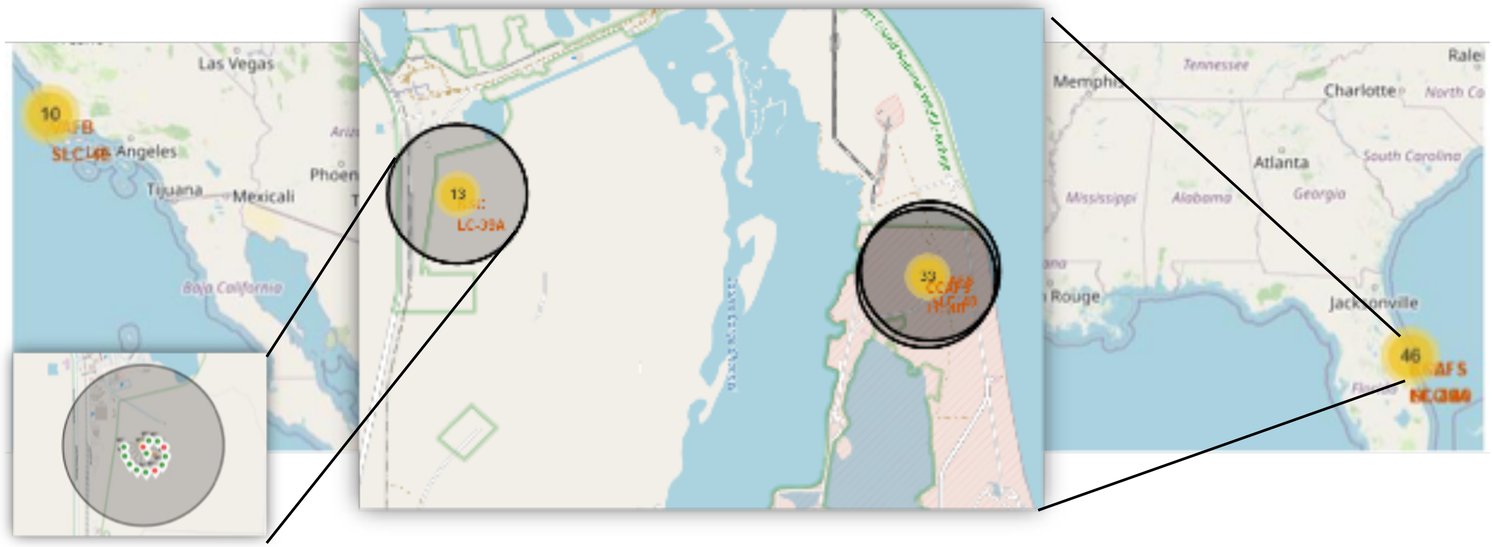
All launch sites



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

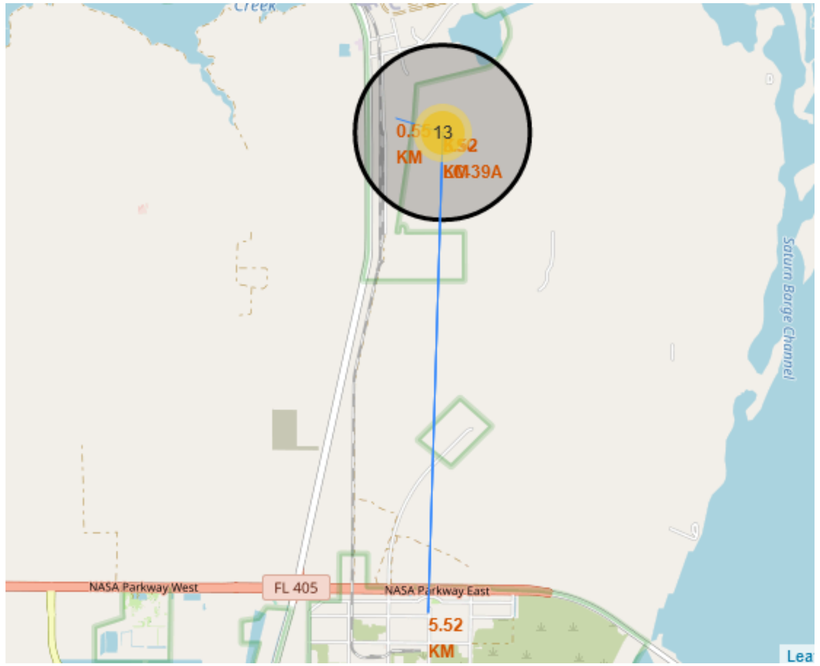
Launch Outcomes by Site

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



- Green markers indicate successful and red indicate as failure.

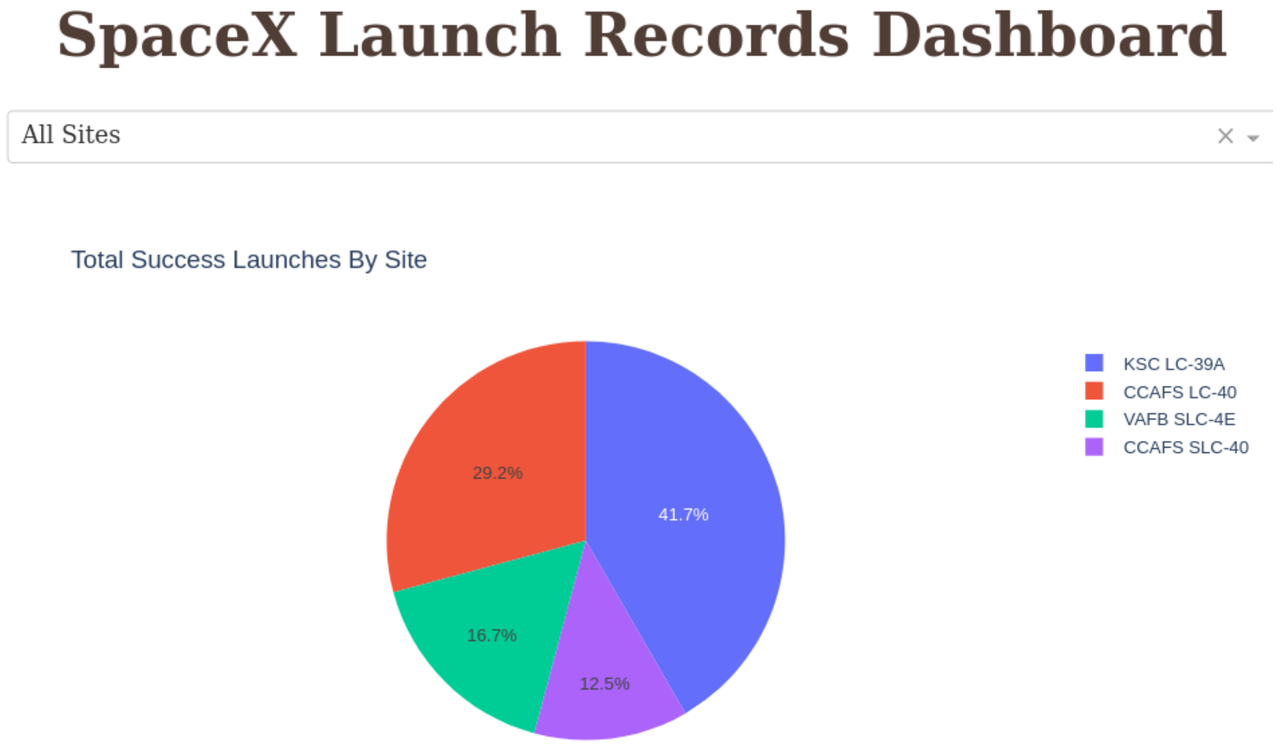
Logistics and Safety



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

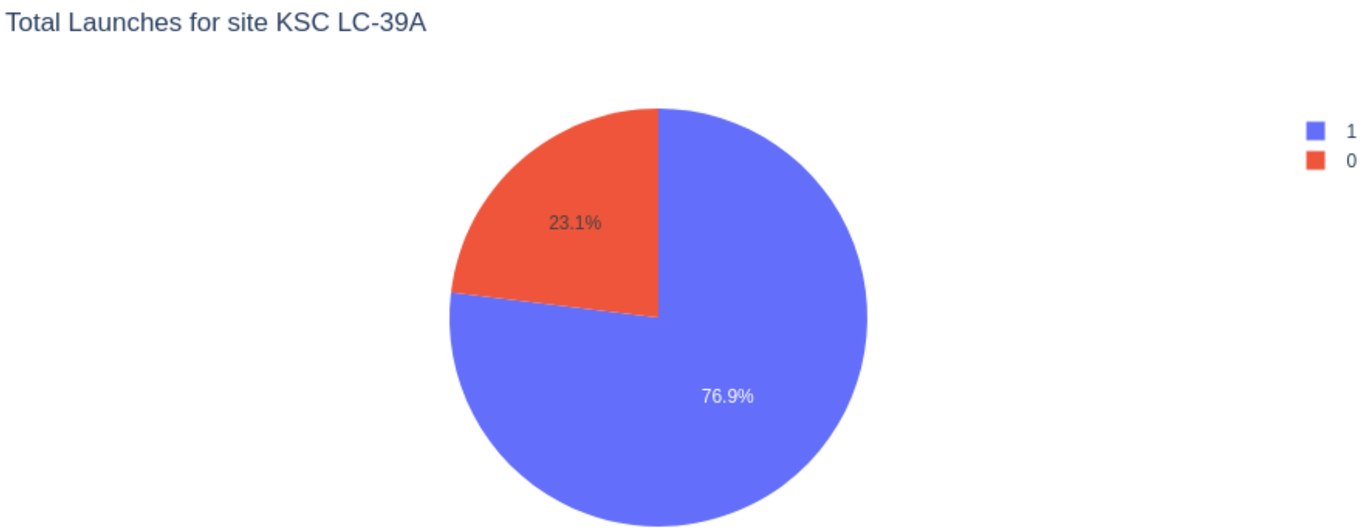
Build a Dashboard with Plotly Dash

Successful Launches by Site



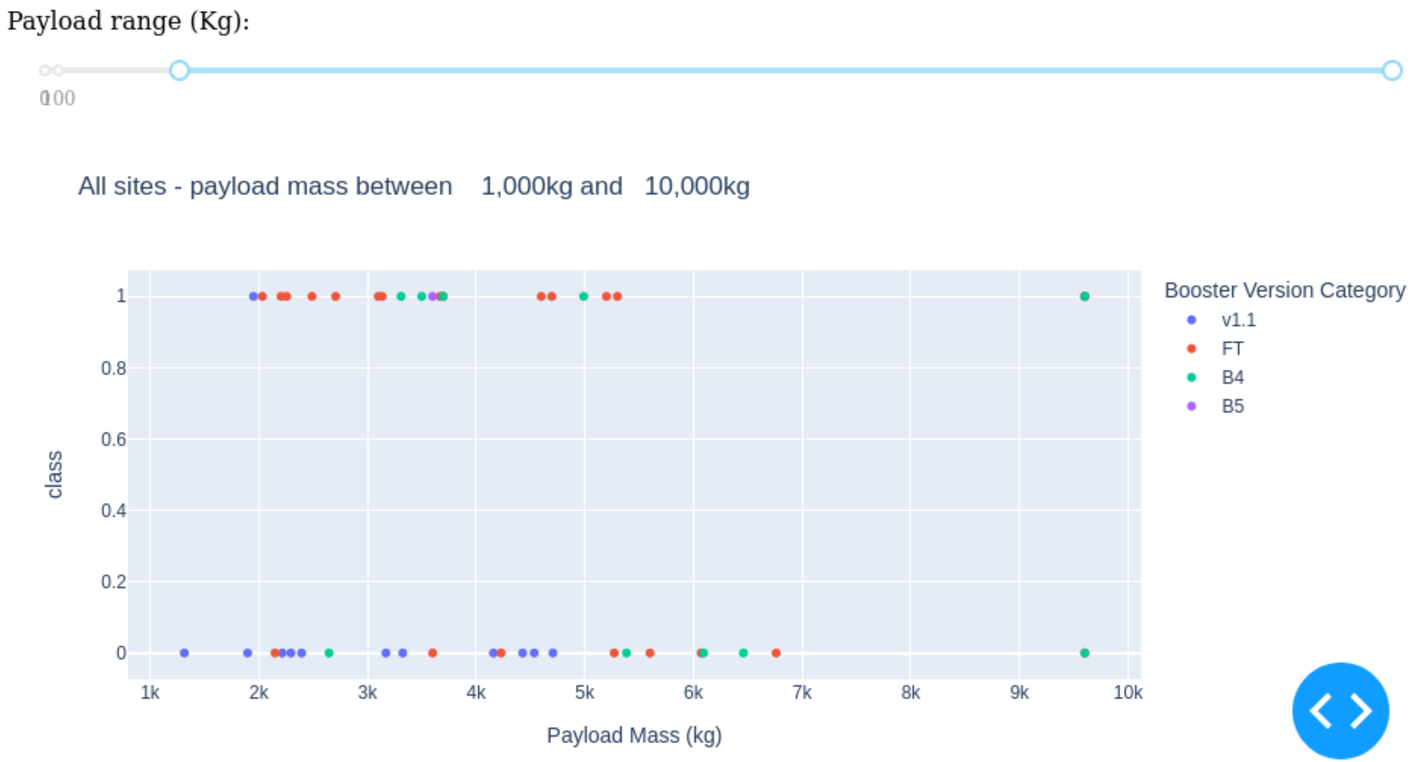
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



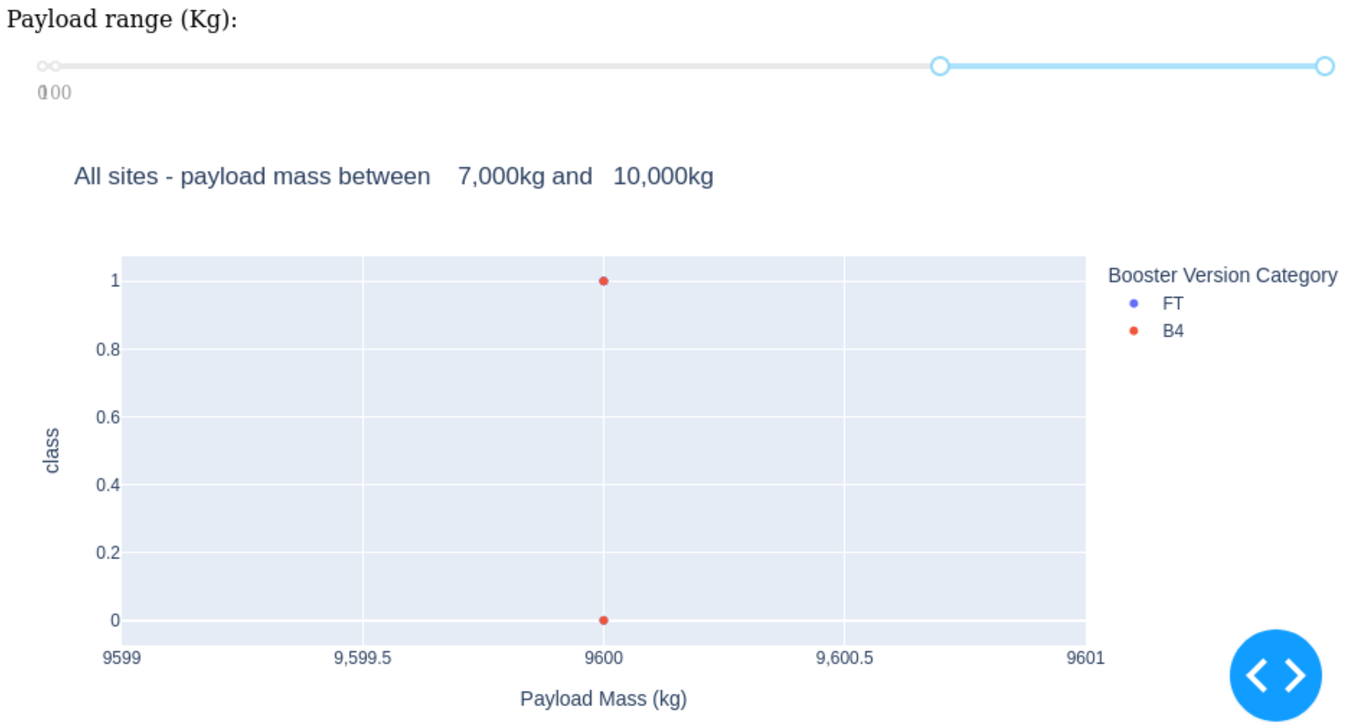
- 76.9% of launches are successful in this site.

Payload vs Launch Outcome



- Payloads under 6,oookg and FT boosters are the most successfully combination.

Payload vs Launch Outcome



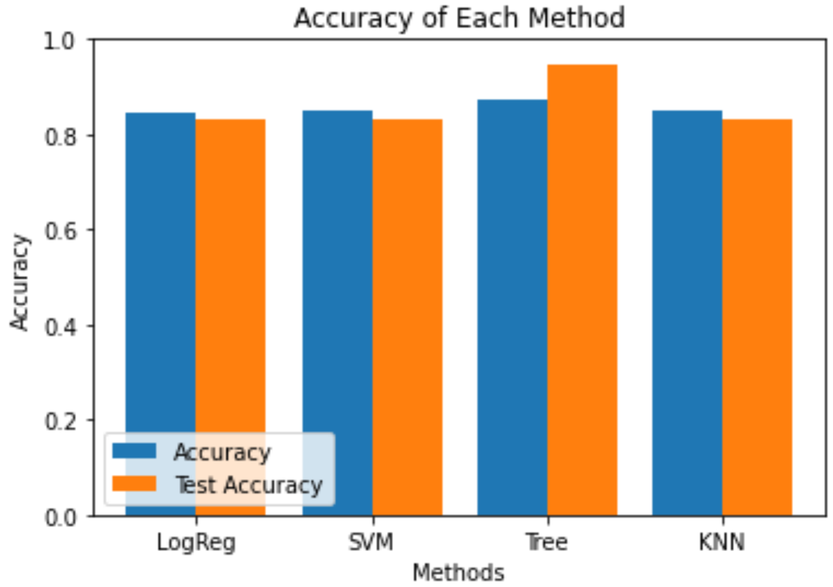
- There’s not enough data to estimate risk of launches over 7,000kg.

Predictive Analysis (Classification)

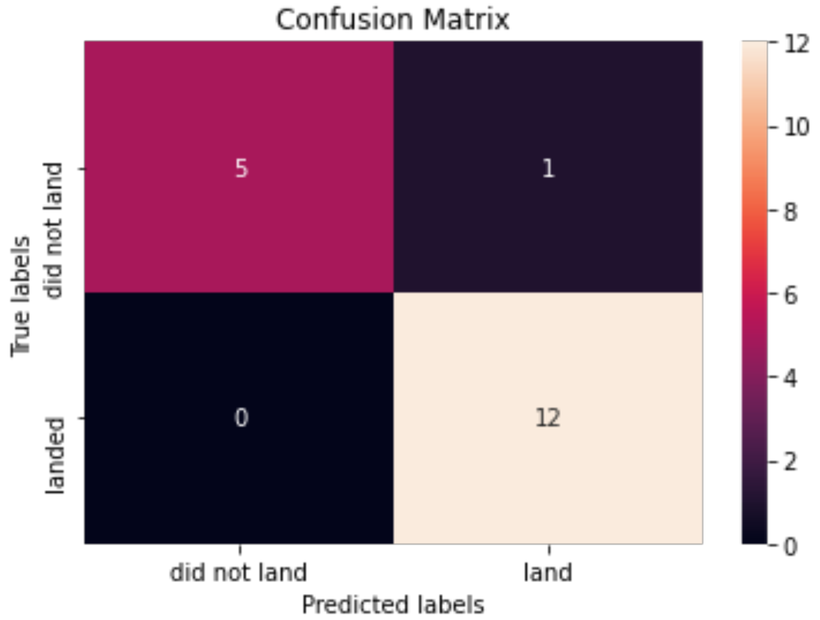
Classification Accuracy

- 4 classification models were tested, and their accuracies are plotted besides

- The model with th 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 +ve and true -ve compared to the false ones.

Conclusions

- Different data sources were analysed to refine conclusions along the process
- The best launch site is KSC LC-39A
- Launches above 7,000 kg are relatively less risky
- Successful landing outcomes seem to improve over time, according to the evolution of processes and rockets even though most of the mission result are successful
- Decision Tree Classifier is a good way to predict successful landings and increase profits.

Thank you!