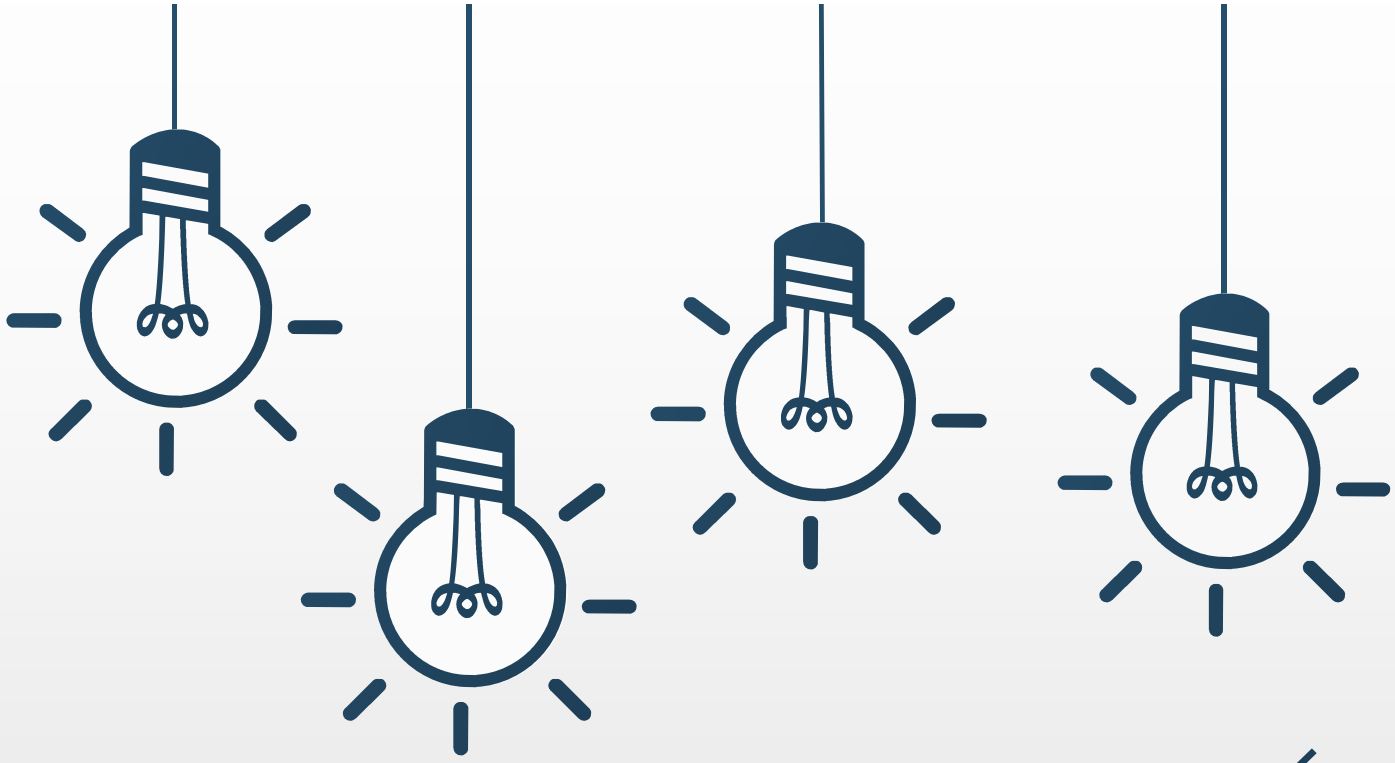
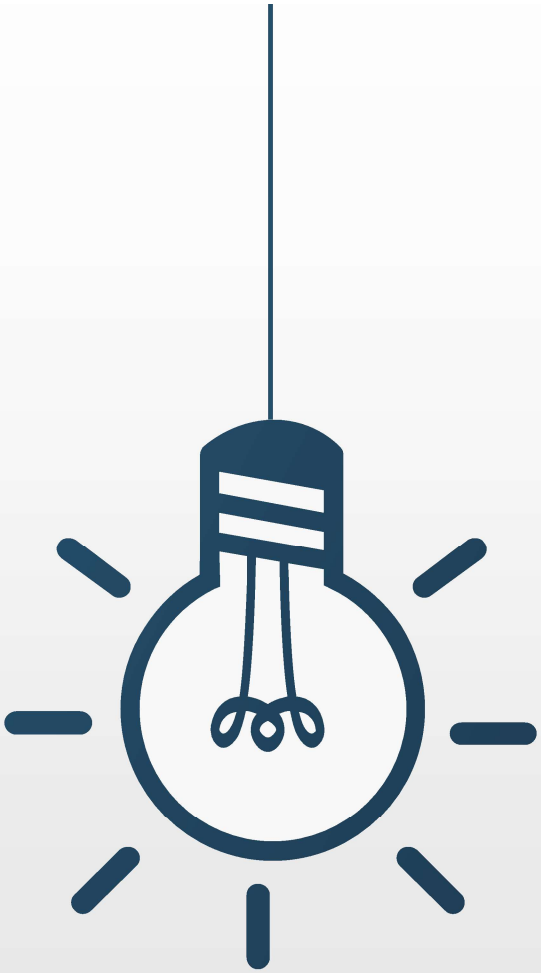
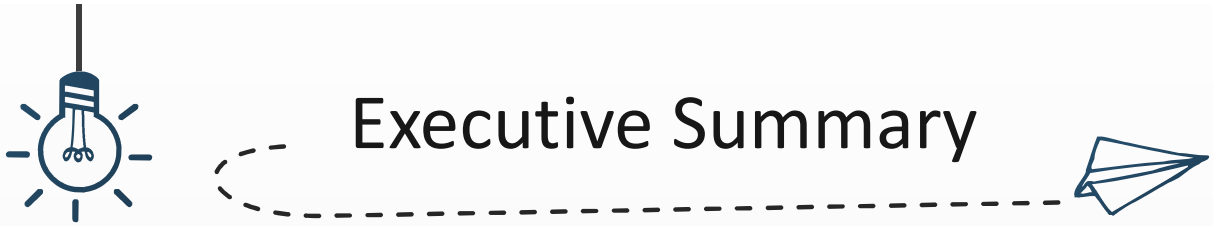


Winning space race with Data Science

REPORTER
VY TANG



Executive Summary



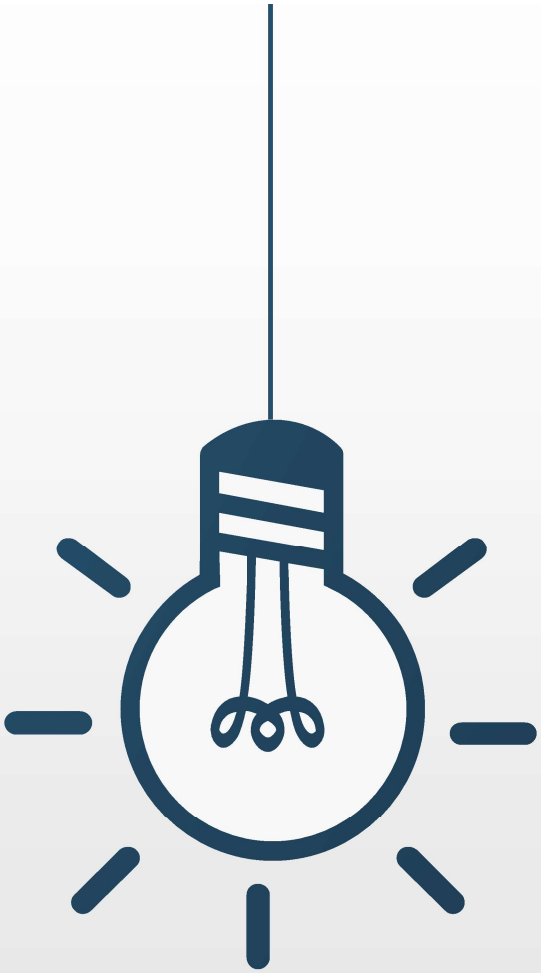
Executive Summary

The following methodologies were used to analyze data:

- Firstly, Data Collection using web scraping and SpaceX API;
- Secondly, Exploratory Data Analysis (EDA) is implemented,which include data wrangling, data visualization and interactive, visual analytics;
- Finally, Bulding Machine Learning Prediction.

Summary of all results

- It was possible to collected valuable data from public sources;
- EDA allowed to identify which features are the best to predict success of launchings;
- Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data.



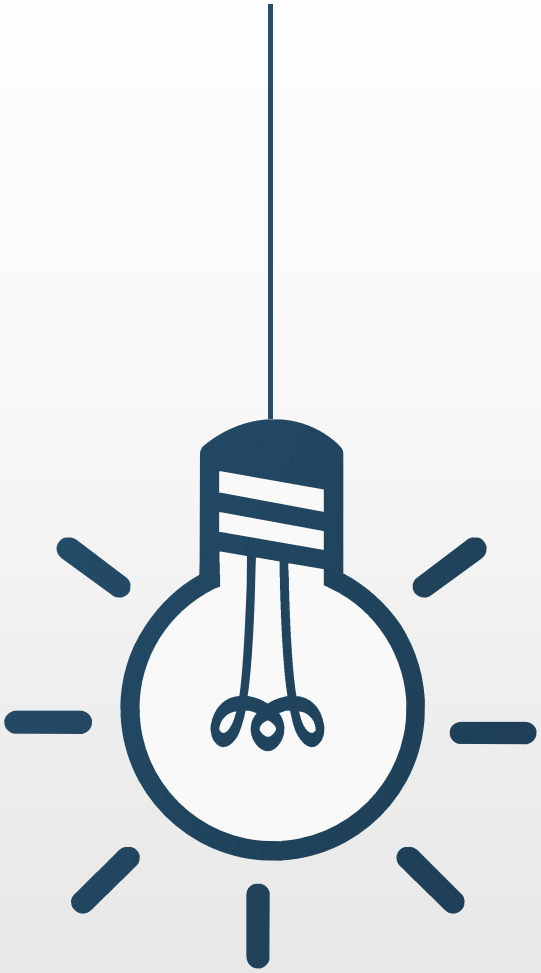
Introduction

Introduction

- The objective is to evaluate the viability of the new company Space Y to compete with the Space X company
- Desirable answers:

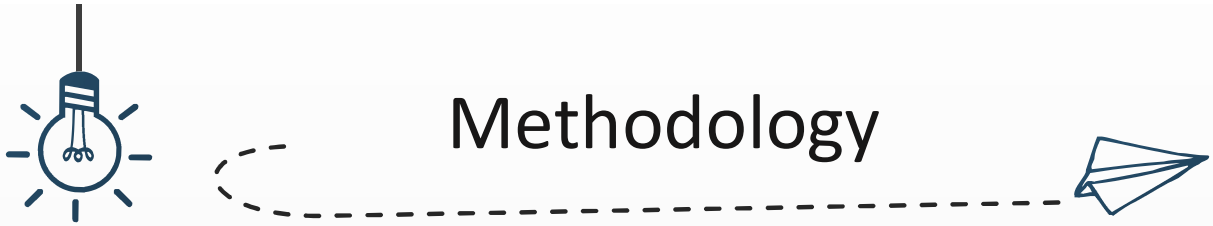
  Find out :

  1. what is the best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets and
  2. Where is the best place to make launches.



Section 1

Methodologies



Methodology

Executive Summary

- **Data collection methodology:**
  - Collect data from the resources the lab provided
- **Perform data wrangling**
  - After analyzing and summarizing the features, the collected data was enhanced by producing a landing outcome label based on the outcome data.
- **Perform exploratory data analysis (EDA) using visualization and SQL**
  - Collect insights from visualization and SQL query
- **Perform interactive visual analytics using Folium and Plotly Dash**
- **Perform predictive analysis using classification models**
  - Normalize and divide data in training and test sets.
  - Build the model and evaluate the score (accuracy)

Data Collection

Through the lab, Data was obtained from 2 sources:

- Space X API (https://api.spacexdata.com/v4/rockets/)
- WebScraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Data Collection

1. Request API and parse the SpaceX launch data

2. Filter data to only include Falcon 9 launches

3. Deal with Missing Values

Reference from my code:
https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/Data%20Collection%20API.ipynb

Data Collection- Web Scraping

1. Request the Falcon 9 Wiki Page
2. Extract data from HTML header
3. Create data frame by parsing the HTML data table

Reference from my code:
https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

1. EDA was performed on the dataset

2. Summarize and calculat launch per site, occurrences of each orbit and the occurrence of mission outcome per orbit

3. Create landing outcome label

Reference from my code:
https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/Data%20Wrangling.ipynb

Exploratory data analysis using visualization and SQL

1. Data Visualization: Plots

Scatterplots were used to visualize the relationship between all pair of features:

- Flight number x launchsite
- Flight number x orbit
- Payloadmass x orbit
- Launchsite x payloadmass

Reference from my code:https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/EDA%20Data%20Visualization.ipynbv

Exploratory data analysis using visualization and SQL

2. EDA: SQL

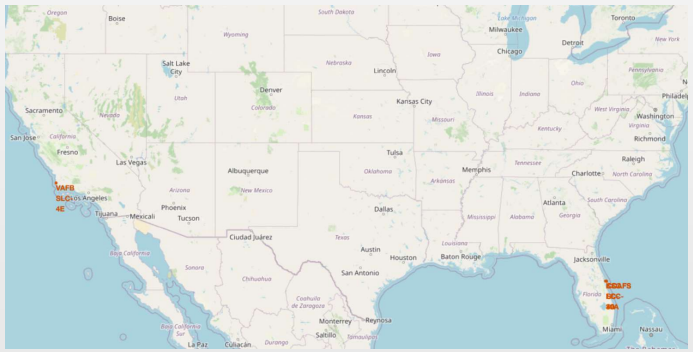
SQL queries were performed:

1. Names of the unique launch sites in the space mission
2. Top 5 launch sites whose name begin with the string 'CCA’
3. Total payload mass carried by boosters launched by NASA (CRS)
4. Average payload mass carried by booster version F9 v1.1
5. Date when the first successful landing outcome in ground pad was acheived.
6. Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg
7. Total number of successful and failure mission outcomes
8. Names of the booster versions which have carried the maximum payload mass
9. The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. The rank of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20

Reference from my code:https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/EDA%20with%20sql.ipynb

Visual Analytics with Folium

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Mark the distances between a launch site to its proximities

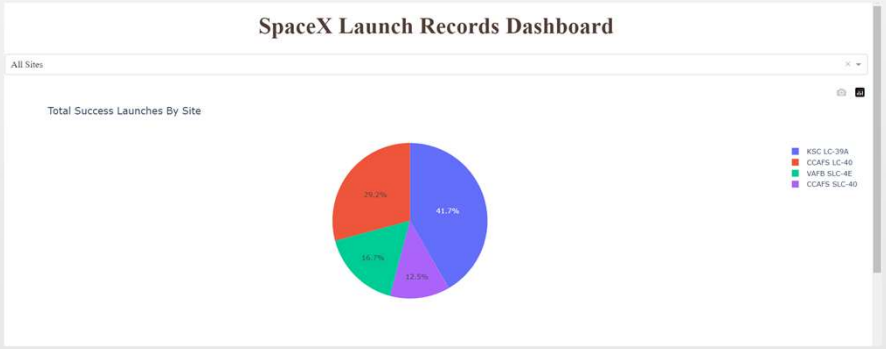


Reference from my code:https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Dashboard with Plotly Dash

Graphs and plots were used to visualize the following:

- Percentage of launch by all site and particular site
- Adjustable Payload range(Kg)



Reference from my code:https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/spacex_dash_app.py

Predictive Analysis

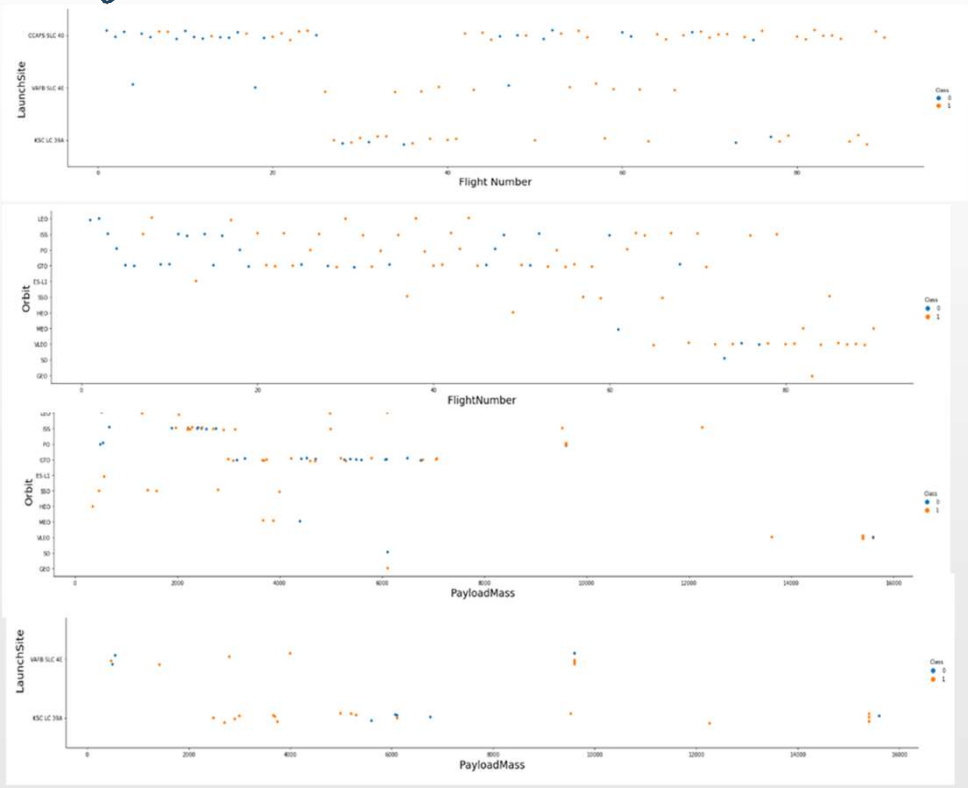
1. Normalize and prepare divide data in training and test sets.
2. Build the model (logistic regression, support vector machine, decision tree and K-nearest neighbors)
3. Evaluate the score (accuracy) and compare results

Reference from my code:https://github.com/davidzeng28/Applied-Data-Science-Capstone-Final-Project/blob/master/Machine%20Learning%20Prediction.ipynb

Results

1. EDA result

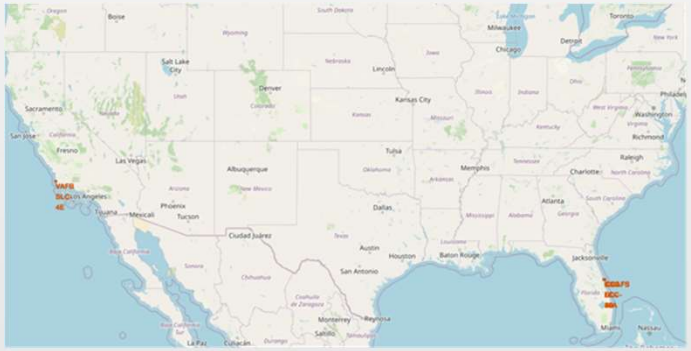
- Space X uses 4 different launch sites in total
- The average payload of F9 v1.1 booster is 2,928 kg;
- The first success landing outcome happened in 2015 fiver year after the first launch;
- Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average;
- Almost 100% of mission outcomes were successful;
- Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015;
- The number of landing outcomes became as better as years passed.



Results

2. Interactive analysis

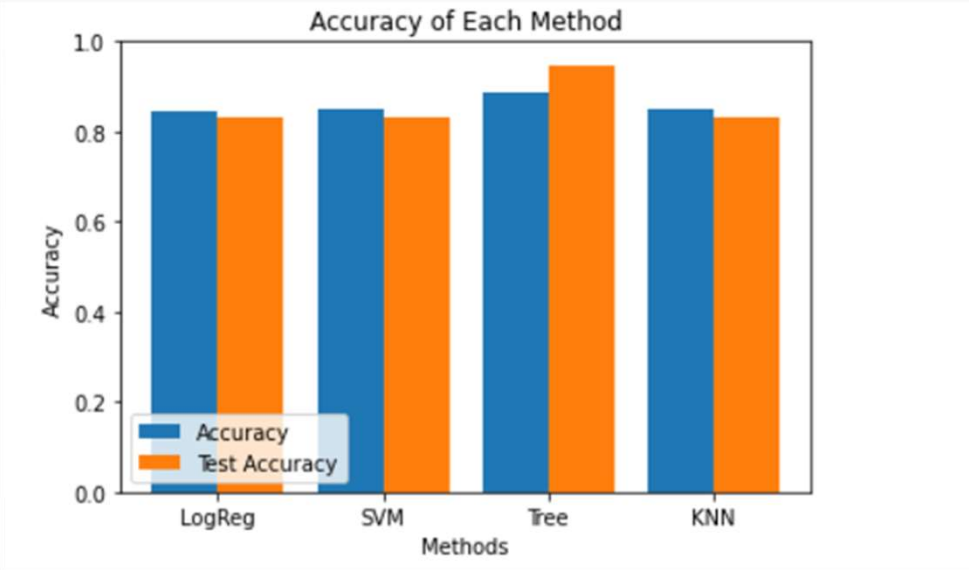
- Identify launch sites be in suitable places
- Most launches happens at east cost launch sites

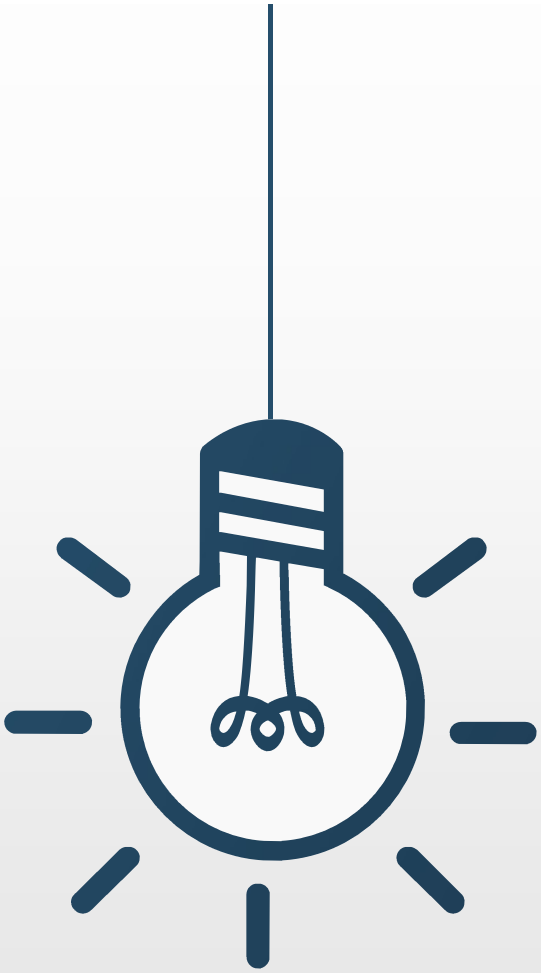


Results

3. Model results

- Decision Tree is the best model to predict



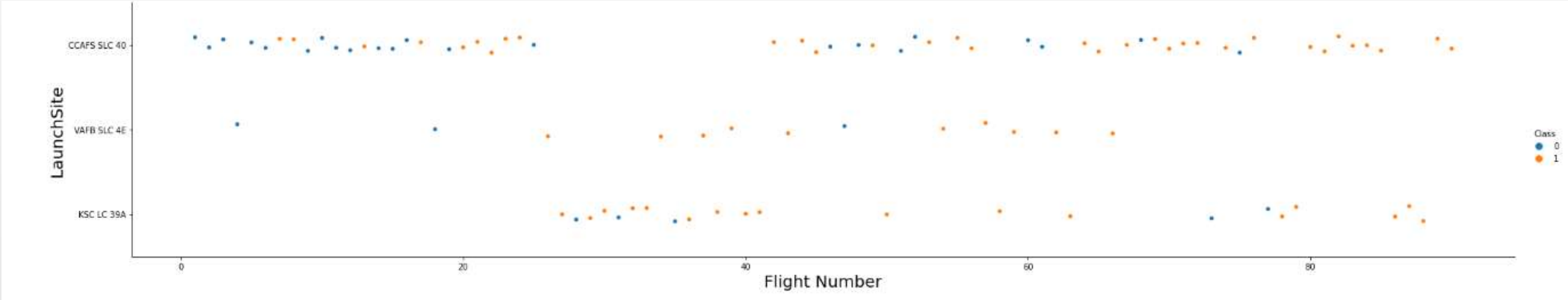


Section 2

INSIGHTS DRAWN FROM E D A

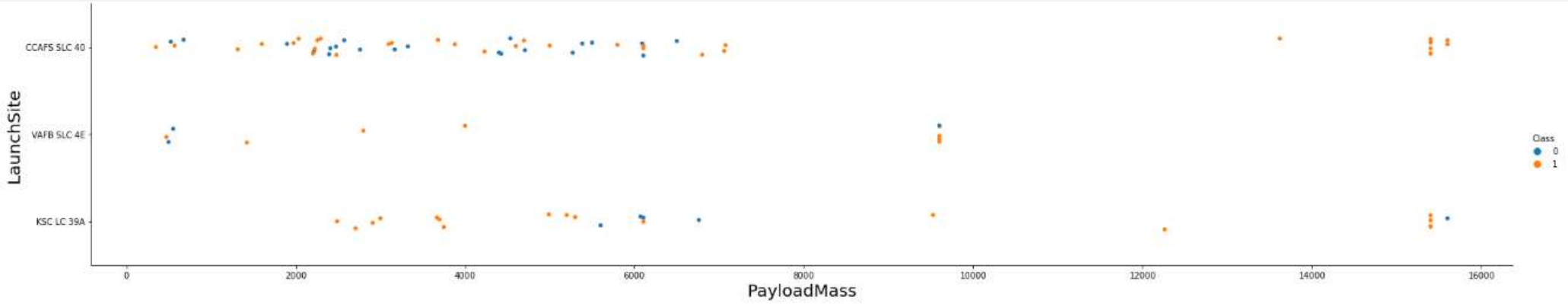
Flight number x Launch Site

- Best launch site is CCAF5 SLC 40, VAFB SLC 4E in second, and KSC LC 39A is the final one
- Success rate have improved over time



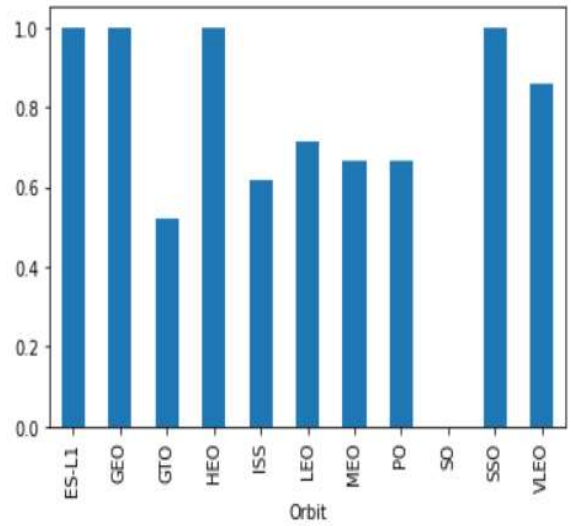
PayloadMass x LaunchSite

- Most launches with payload mass under 10,000 kg are from any launch site, but heavier ones happens mainly at CCAFS SLC 40 and KSC LC 39A



Success rate of each orbit type

Out[6]:

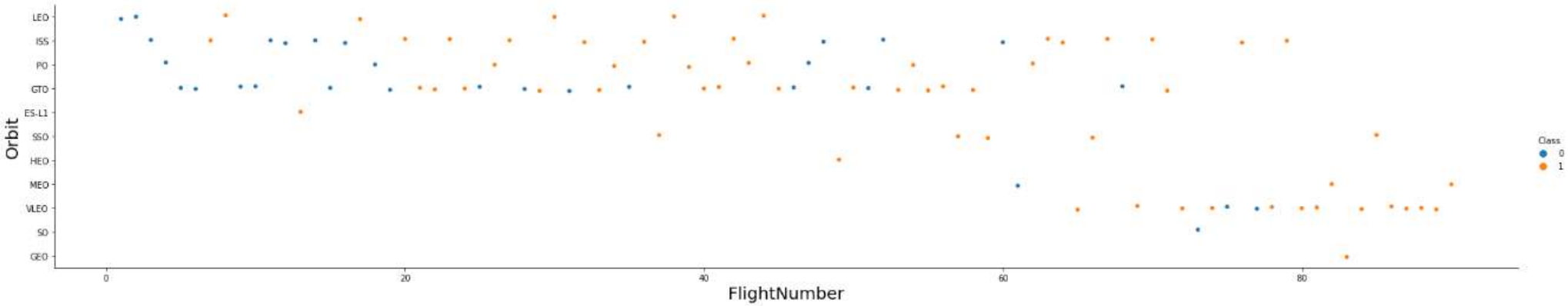


There are four biggest success rate of orbit type:

- ES-L1
- GEO
- HEO
- SSO

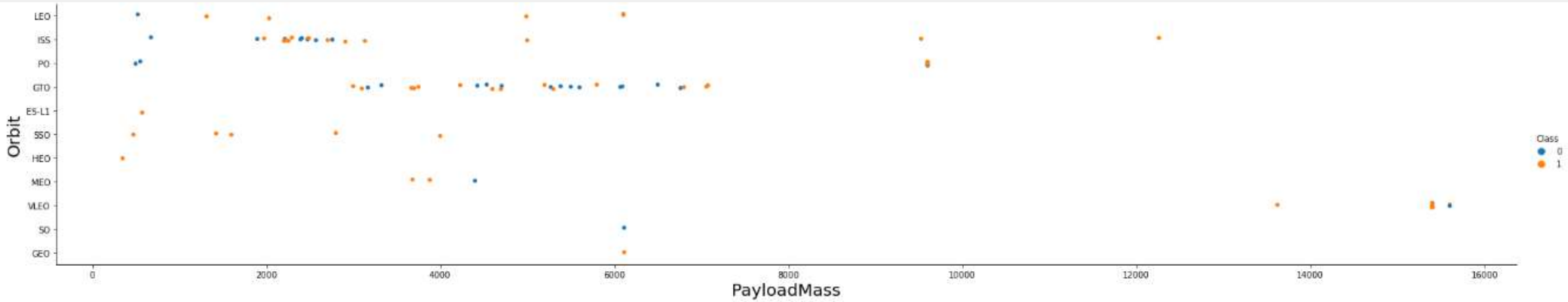
Orbit x FlightNumber

- LEO orbit the Success appears related to the number of flights
- GTO orbit and Flight number has no relationship
- VLEO have the really high frequency



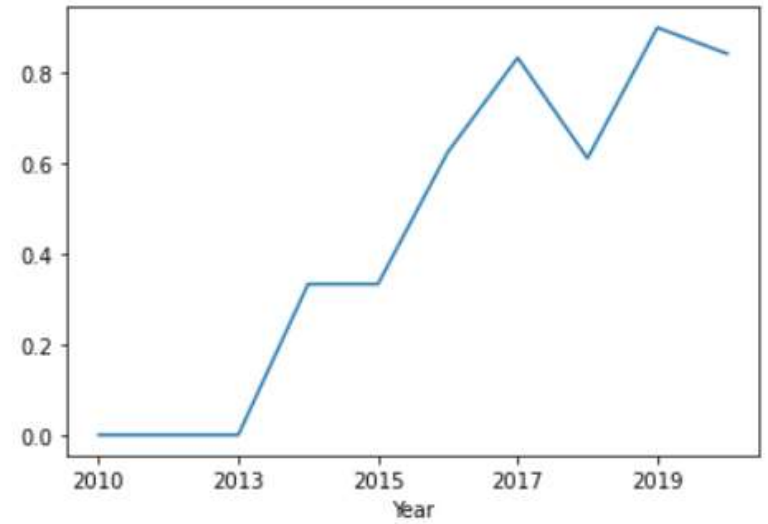
Orbit x PayloadMass

- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.
- SO and GEO have lowest launches
- ISS have the widest range of payload



Launch success yearly trend

Out[10]:



- The rate started increasing from 2013, but face a drop in 2018 and 2020
- It seems no success launches in the first three year due to the backward technology
- It reaches it peak in 2019

All Launch Site Names

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- According to data and using sql to extract, there are four launch sites in total

All Launch Site Names

- 5 example records which launch sites begin with ‘CCA’

Out[11]:

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 attempt

Total Payload Mass

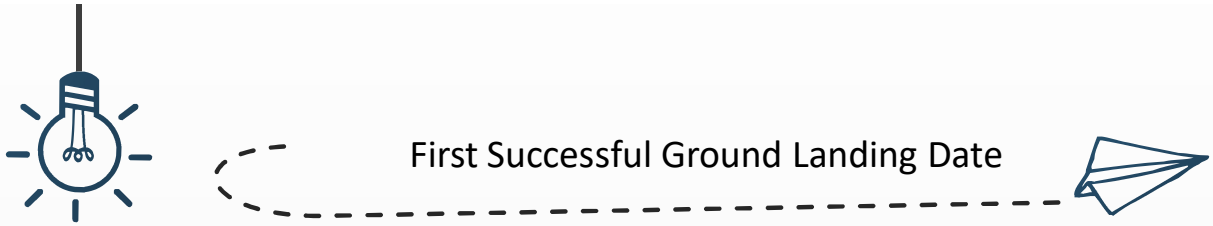
total_payload
111268

- Total payload mass carried by boosters launched by NASA (summing all Payload contain “CRS”)

Average Payload Mass by F9 v1.1

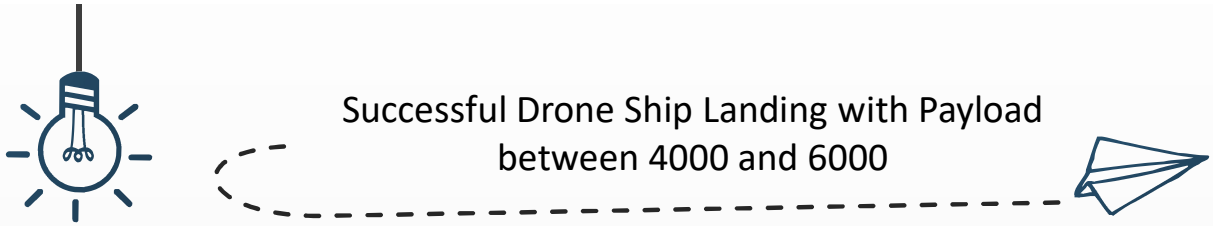
avg_payload
2928

- Show the data in the filter of booster version above and calculate the average payload mass, which is 2928 kg



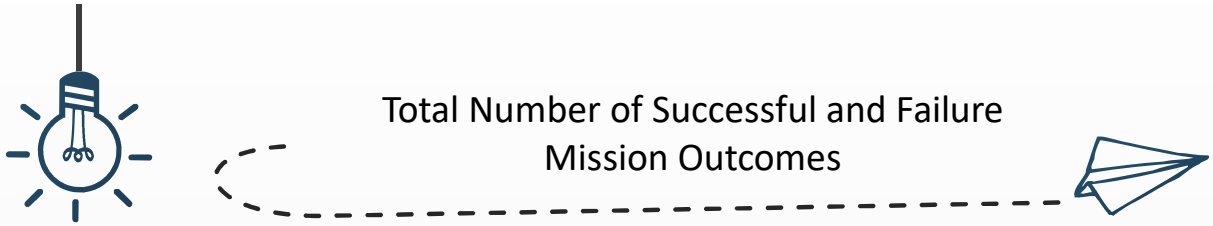
first_success_gp
2015-12-22

- Filter on ground pad and get the min value of the date, which is the first successful date “2015-12-22”



booster_version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

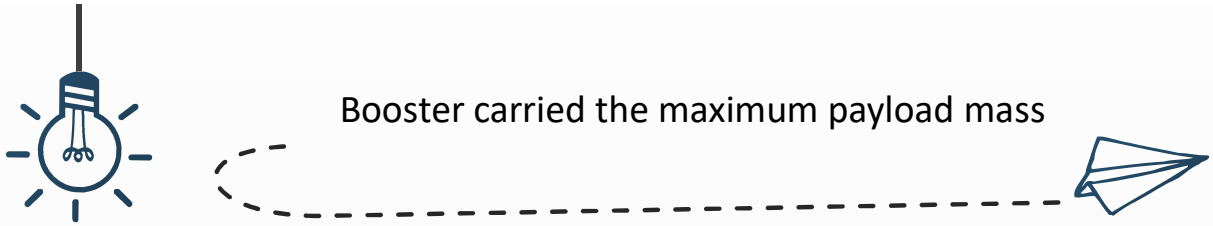
- 4 results are retrieve with the filtering Payload 4000- 6000



Total Number of Successful and Failure Mission Outcomes

mission_outcome	qty
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

- The table show count records of each group and grouping failure mission outcomes



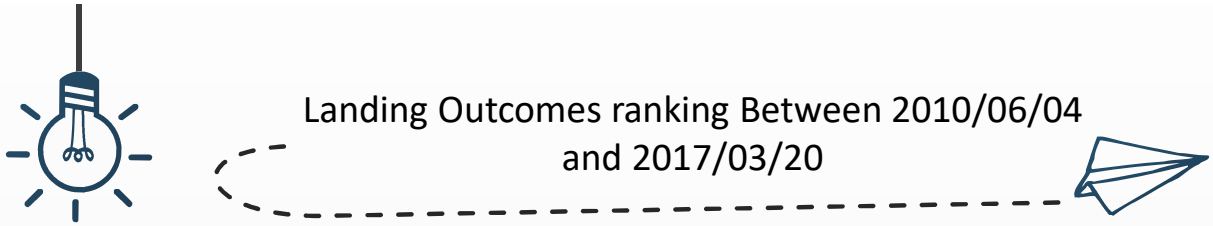
Booster carried the maximum payload mass

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

- The table show the boosters which have carried the maximum payload mass In the dataset

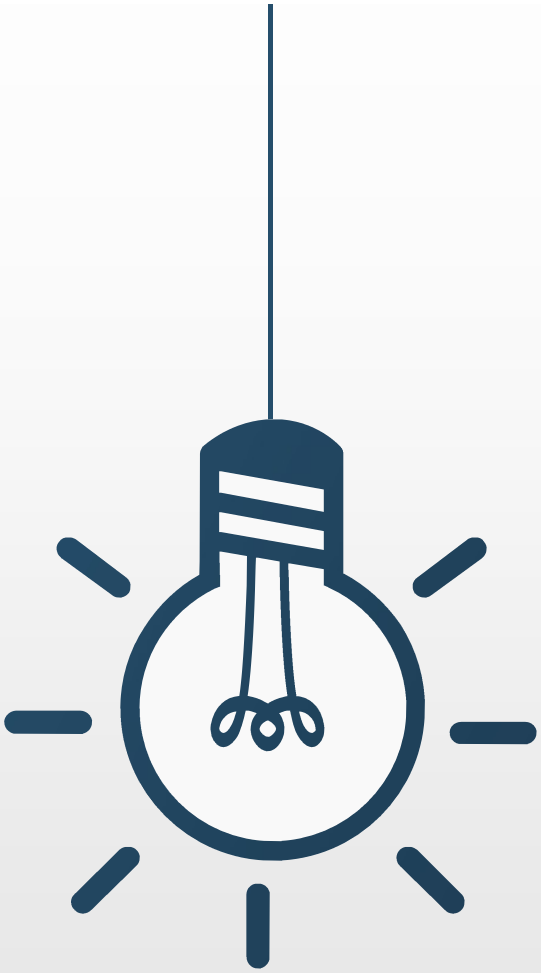
- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 include 2 occurences

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40



- The no attempt outcome is the biggest and need to take in account

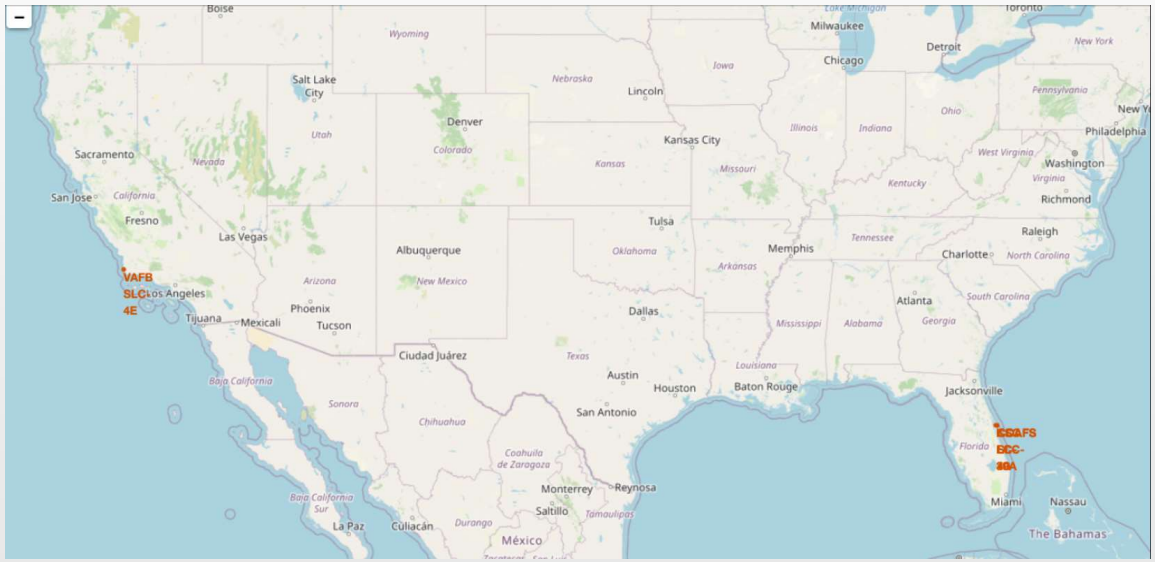
landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1



Section 3

Launch Sites Proximities ANALYSIS

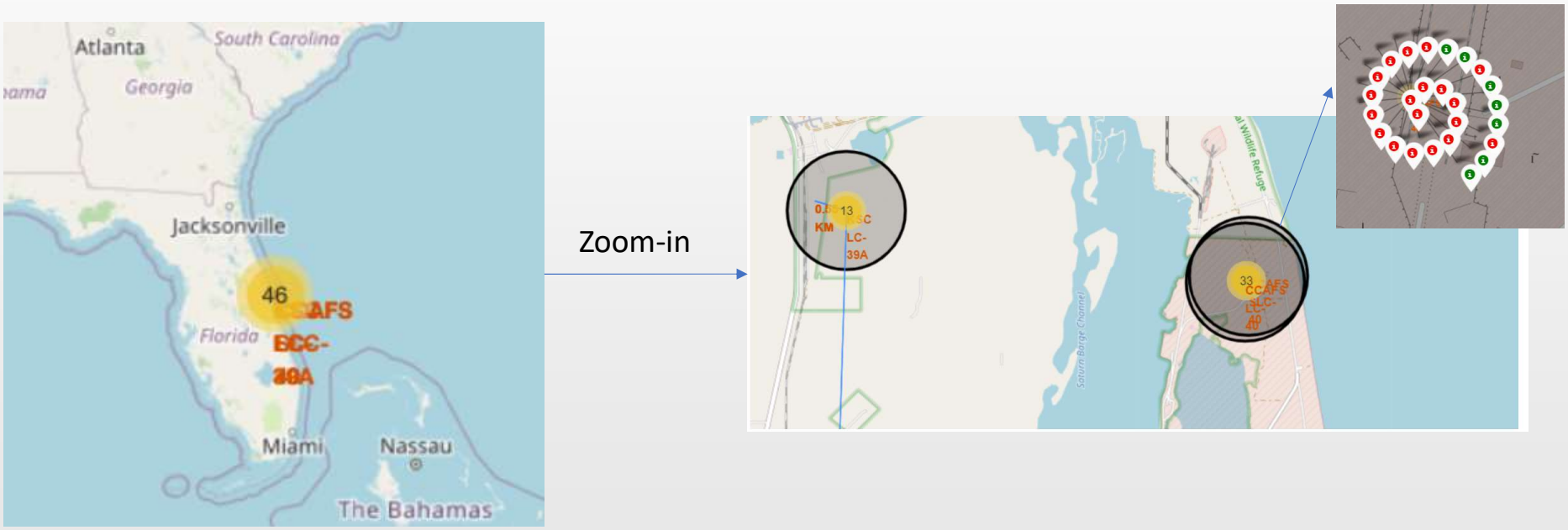
All launch sites



- 2 launch sites are located near areas that near the sea

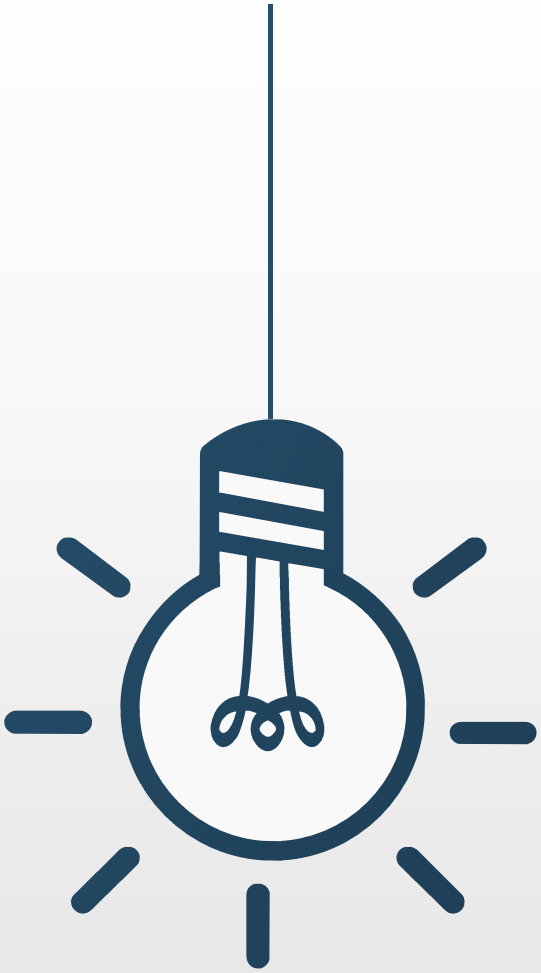
Launch Outcomes by Site

Example: CCAFS LC-40 of launch site launch outcomes when zoom-in zoom-out



Logistic road

For example: This picture show that this launch site has a really good logistic
Near road and railroad and far from inhabited areas

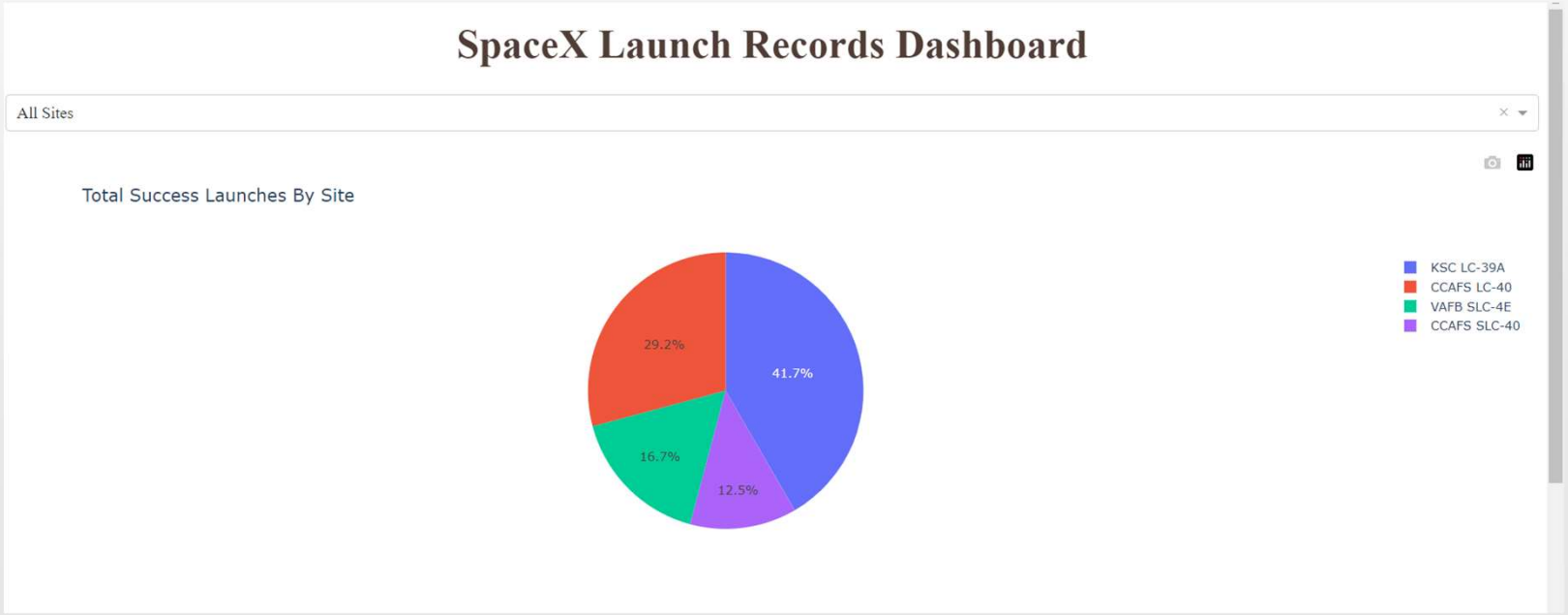


Section 4

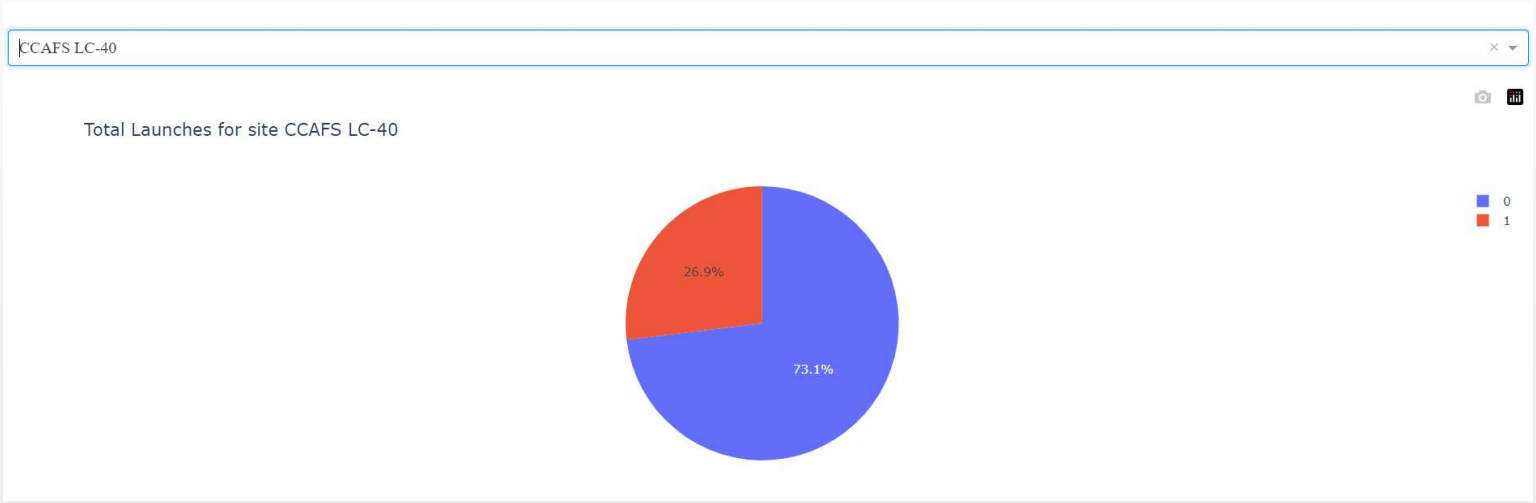
Dashboard with Plotly Dash

Successful Launches by Site

- The sites is a key factor of success of the mission



Example for launch site



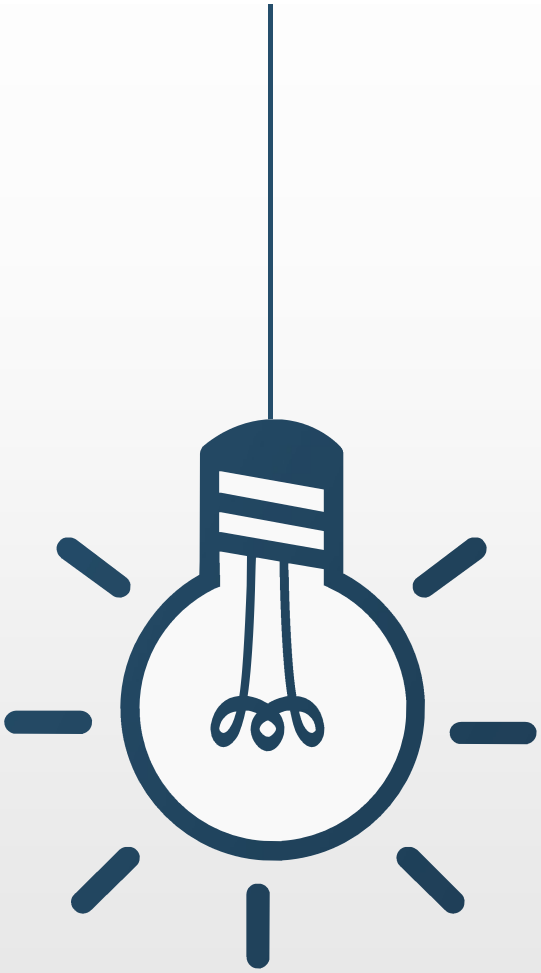
- CCAFS LC-40 is the chosen example
- 73.1 success rate in this site

Payload vs. Launch Outcome

- Payload from 1000-6000 and the “FT” booster version together have the highest rate of success

Payload vs. Launch Outcome

- Only 2 result from over 7000 kg ; hence, we don't have enough data to conclude or give any insight

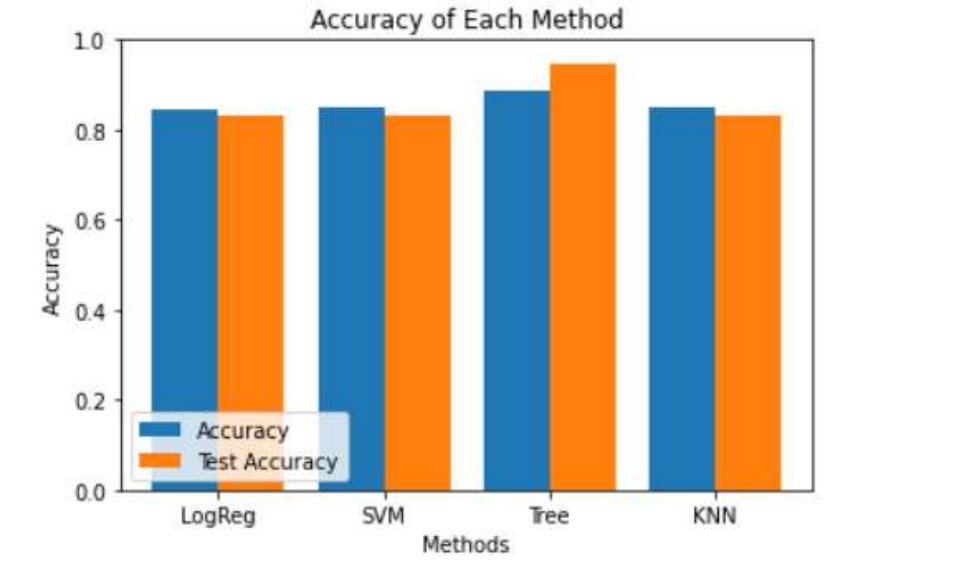


Section 5

Predictive Analysis (Classfication)

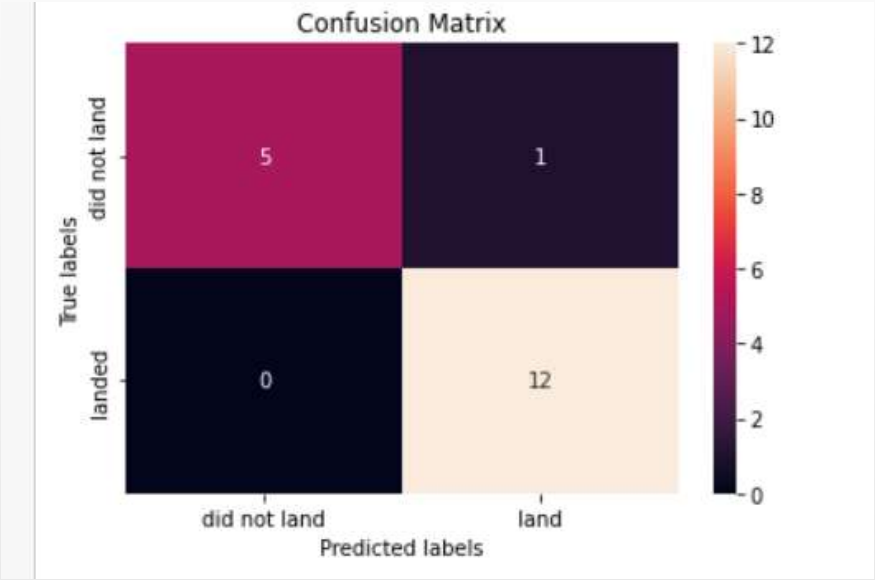
Classification accuracy

- Logistic regression, support vector machine, decision tree and K-nearest neighbors are used
- The highest accuracy rate is decision tree classfier, which is more than 88%



Confusion Matrix of Decision Tree Classifier

- The big numbers of true positive and true negative compared to the false ones show that the confusion matrix of decision tree has the highest accuracy

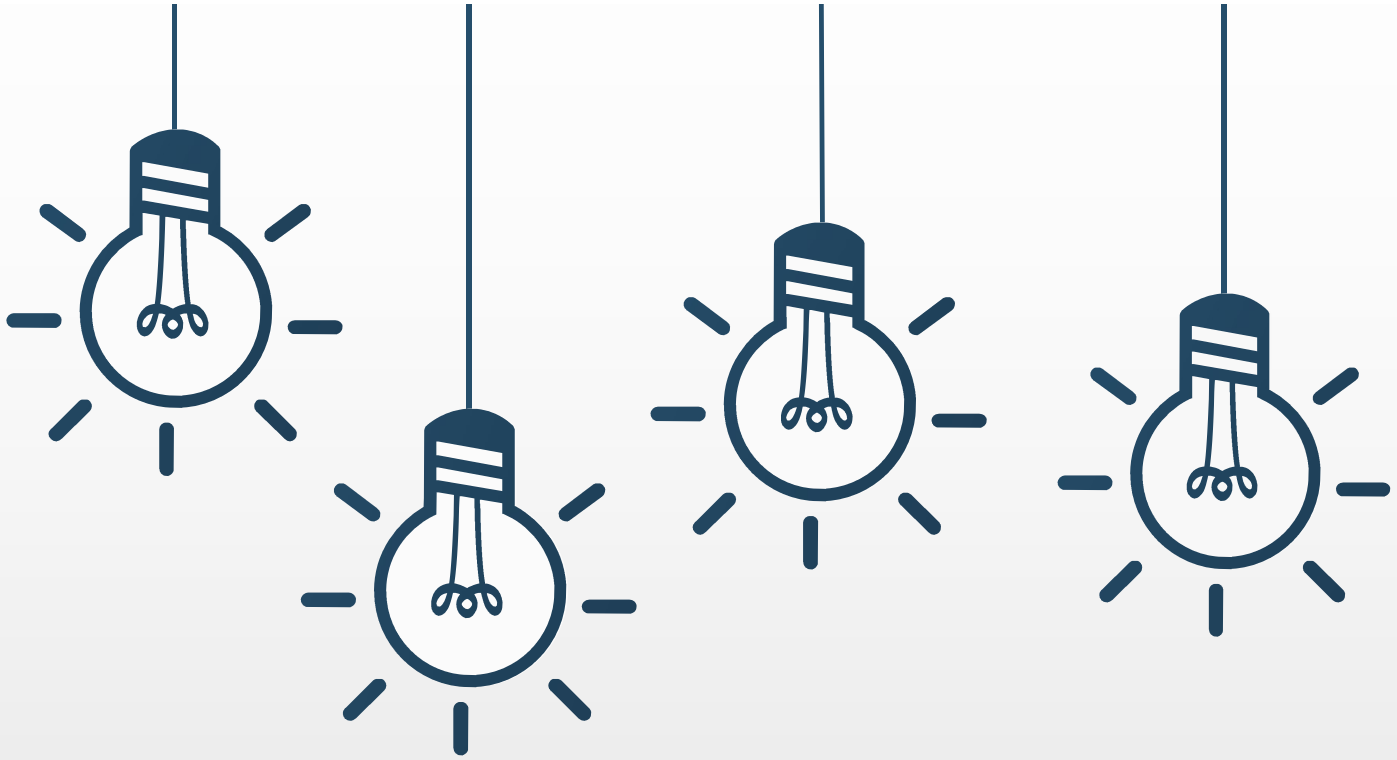


Conclusion

- Success landing outcomes improve overtime
- The best launch site will be KSC LC-39A
- Launches above do not have enough data to conclude, hence it’s risky
- Decision tree classifier is the best model to predict success landings rate and increase profit

Appendix

- Because Folium doesn’t show map on Github, so I attached image in Jupyter Notebook file, please download to see the full version



THANKS FOR LISTENING