IBM Developer SKILLS NETWORK

Winning Space Race with Data Science

Atheer Alzaid
4-7-2023

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building interactive map with folium
- Building dashboard with Plotly dash
- Predictive analysis

Summary of Results

- We collected valuable data from public sources
- EDA help to identify the important features to predict success of launching
- We tested Machine learning models that are used for binary classification

Introduction

Project background and context

- SpaceX is the most successful company of the commercial space age, making space travel affordable. Much of the savings is because SpaceX can reuse the first stage. therefore, we want to determine if the first stage will land to determine the cost of a launch. Based on public information we gathered, we will determine this using prediction model

Problems you want to find answers

- How do variables such as number of flights, payload mass, launch site affect the success of the first stage landing?
- What is the best binary classification model to use in this case?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from SpaceX were collected from 2 sources:

1. Using SpaceX API to request rocket launch data
2. Web Scraping Falcon 9 launch data from Wikipedia

- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features

- Perform exploratory data analysis (EDA) using visualization and SQL

Methodology

Executive Summary

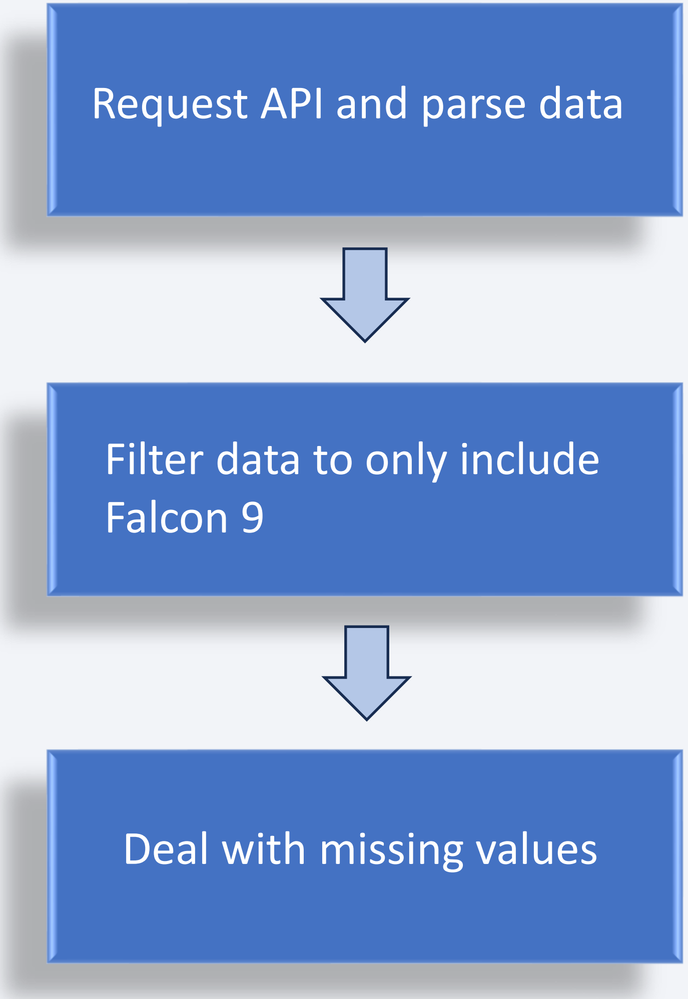
- Perform interactive visual analytics using Folium and Plotly dash
- Perform predictive analysis using classification models
  - We divided the collected data into training and testing sets and performed 4 different classification models, then evaluated the accuracy for each of the classification models

Data Collection

Data were collected using two techniques , first we used SpaceX API to collect Rocket launch data, and second was web Scpraping Falcon 9 launch data from Wikipedia

Data Collection – SpaceX API

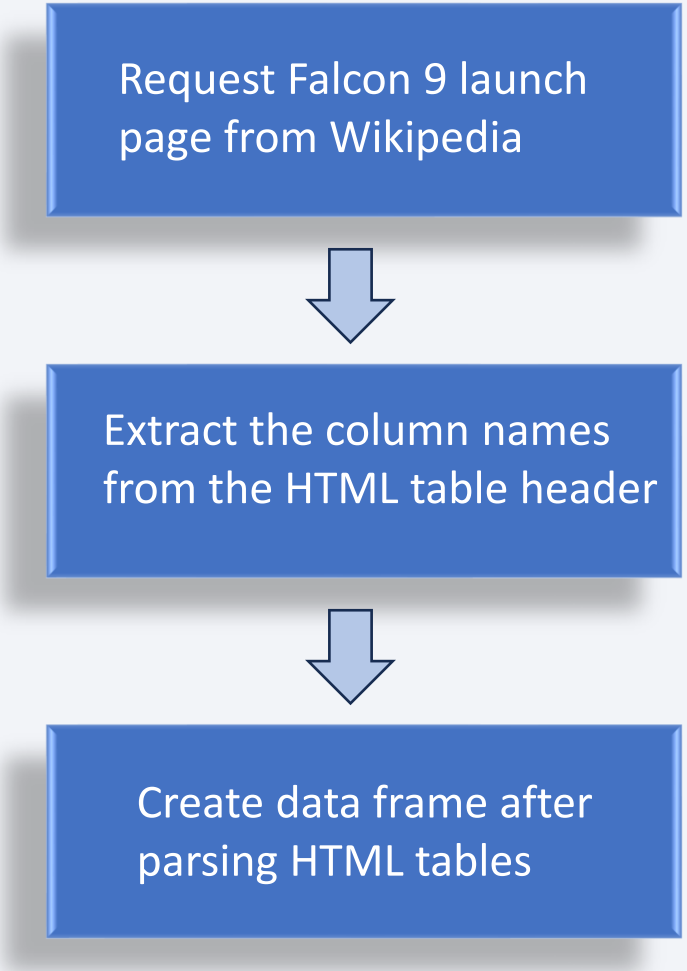
- SpaceX offers public API where we can obtain the data
- GitHub



Data Collection - Scraping

- Data from SpaceX can also be obtained from Wikipedia

- [GitHub]

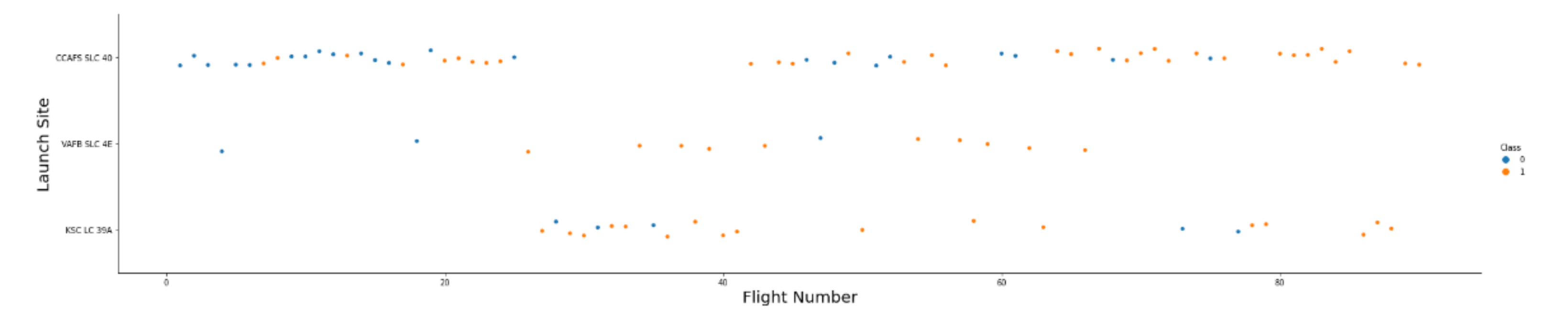


Data Wrangling

- Initially, we performed EDA on the dataset
- Then we calculated the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit
- Finally, the landing outcome label was created
- GitHub

EDA with Data Visualization

- To explore data, we used scatterplots and bar plots to visualize relationships between features: Payload mass X flight number, launch site X flight number, launch site X payload mass, orbit and flight number, payload and orbit



- GitHub

EDA with SQL

- The following SQL queries 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 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; and
  - 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.
- GitHub

Build an Interactive Map with Folium

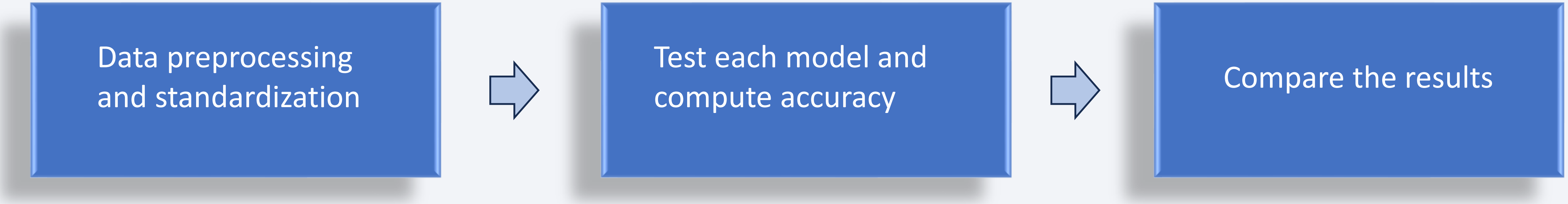
- Markers, circles, lines and marker clusters were used with Folium Maps
    - Markers indicate points like launch sites
    - Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
    - Marker clusters indicates groups of events in each coordinate, like launches in a launch site
    - Lines are used to indicate distances between two coordinates.
- GitHub

Build a Dashboard with Plotly Dash

- Launch Sites dropdown list
- Pie chart showing success of launch
- Slider of payload mass range
- Scatter chart of payload mass vs success rate for different booster versions
- GitHub

Predictive Analysis (Classification)

- 4 classification models were compared: logistic regression, SVM, decision tree and KNN



- GitHub

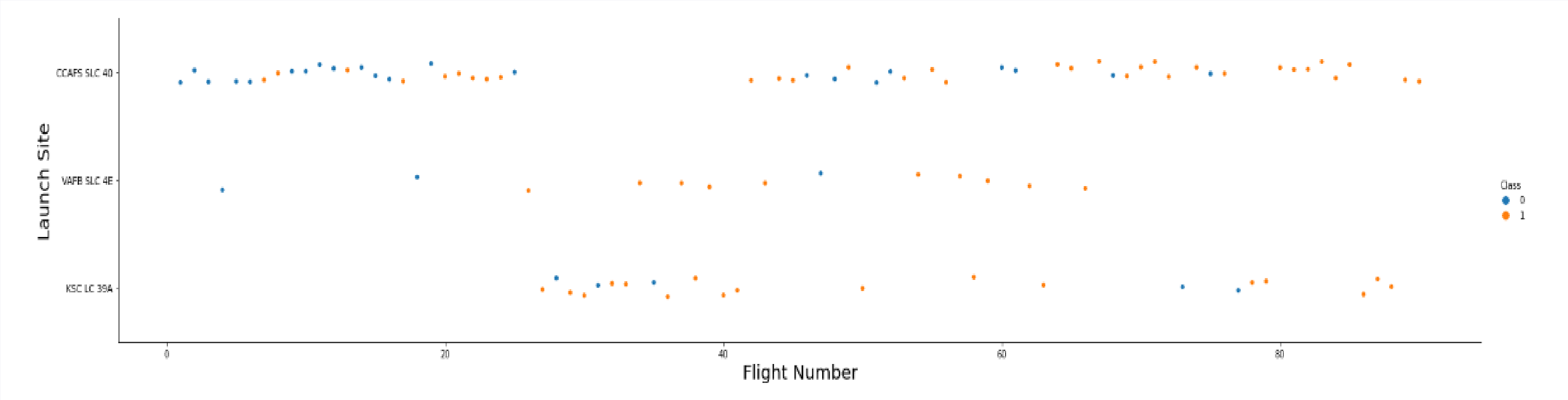
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

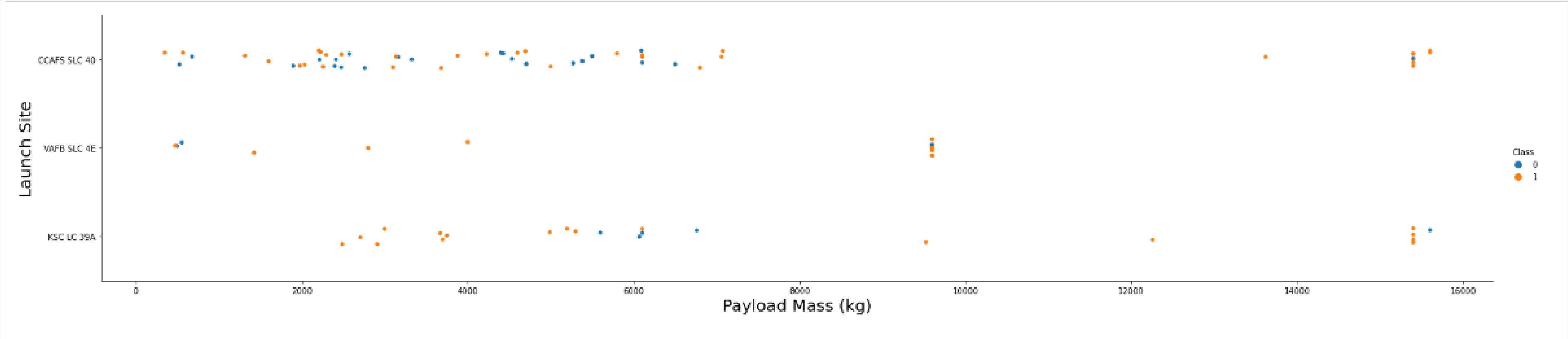
Insights drawn from EDA

Flight Number vs. Launch Site



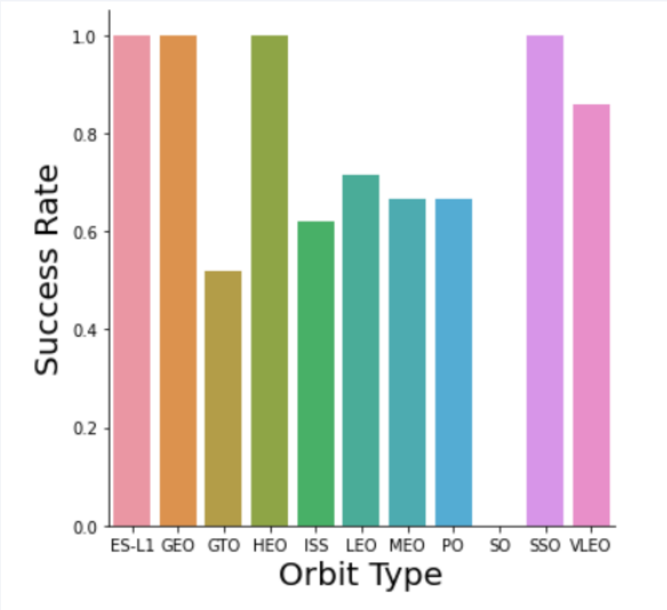
- According to the plot above, it's possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;

Payload vs. Launch Site



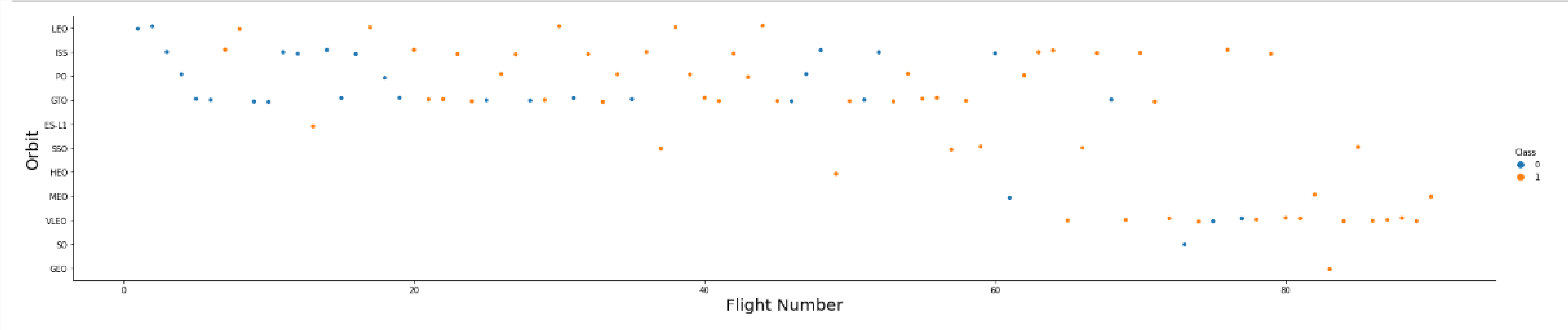
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Success Rate vs. Orbit Type



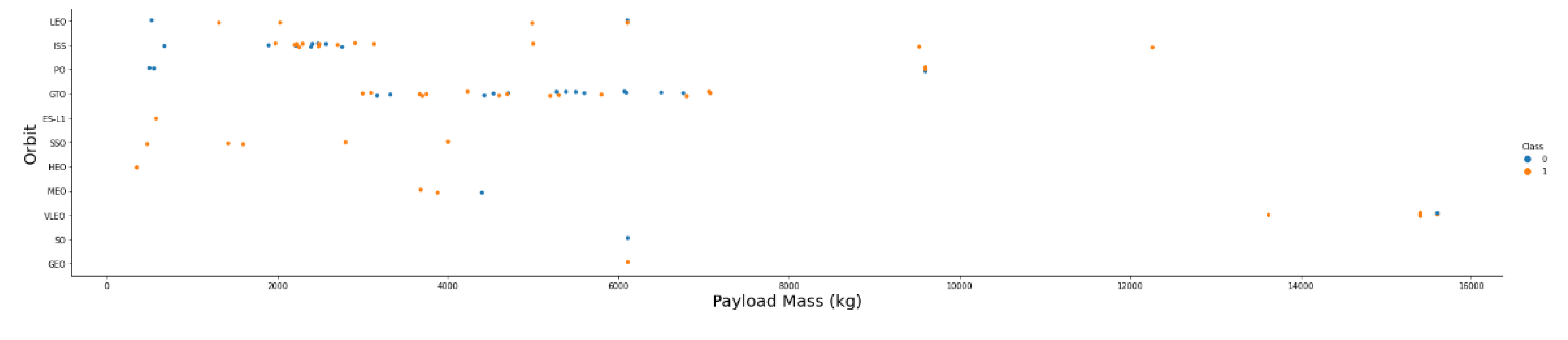
- The biggest success rates happens to orbits:
- ES-L1;
- GEO;
- HEO;
- SSO.

Flight Number vs. Orbit Type



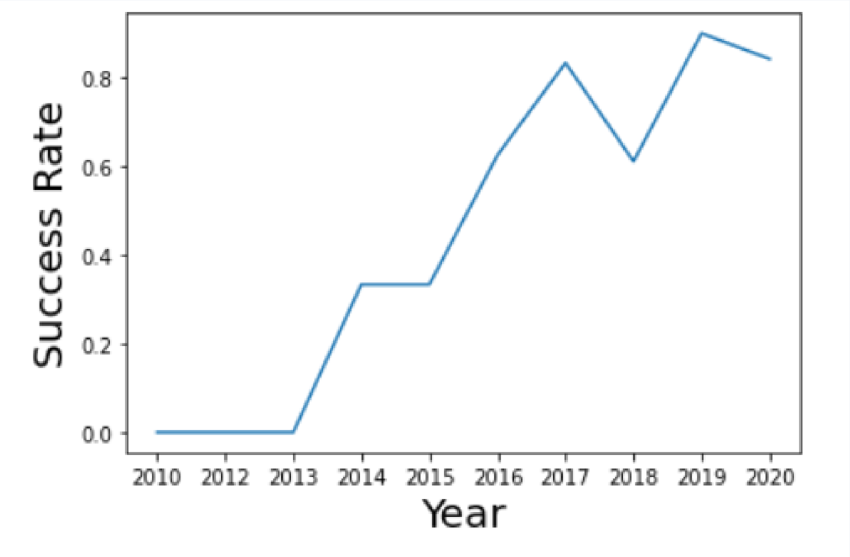
- Apparently, success rate improved over time to all orbits;

Payload vs. Orbit Type



- Apparently, there is no relation between payload and success rate to orbit GTO

Launch Success Yearly Trend



- Success rate started increasing in 2013 and kept until 2020

All Launch Site Names

Out[4]:

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'

Out[5]:

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

- Here we can see five samples of Cape Canaveral launches.

Total Payload Mass

Out[6]:

total_payload_mass
45596

- Total payload calculated above, by summing all payloads whose codes contain ‘CRS’, which corresponds to NASA.

Average Payload Mass by F9 v1.1

Out[7]:

average_payload_mass
2534

- Filtering data by the booster version above and calculating the average payload mass we obtained the value of 2,534 kg.

First Successful Ground Landing Date

Out[8]:

first_successful_landing
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

Out[9]:

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

- Selecting distinct booster versions according to the filters above, these 4 are the result.

Total Number of Successful and Failure Mission Outcomes

Out[10]:

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

- Grouping mission outcomes and counting records for each group led us to the summary above.

Boosters Carried Maximum Payload

Out[11]:

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- These are the boosters which have carried the maximum payload mass registered in the dataset.

2015 Launch Records

Out[12]:

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

Out[13]:

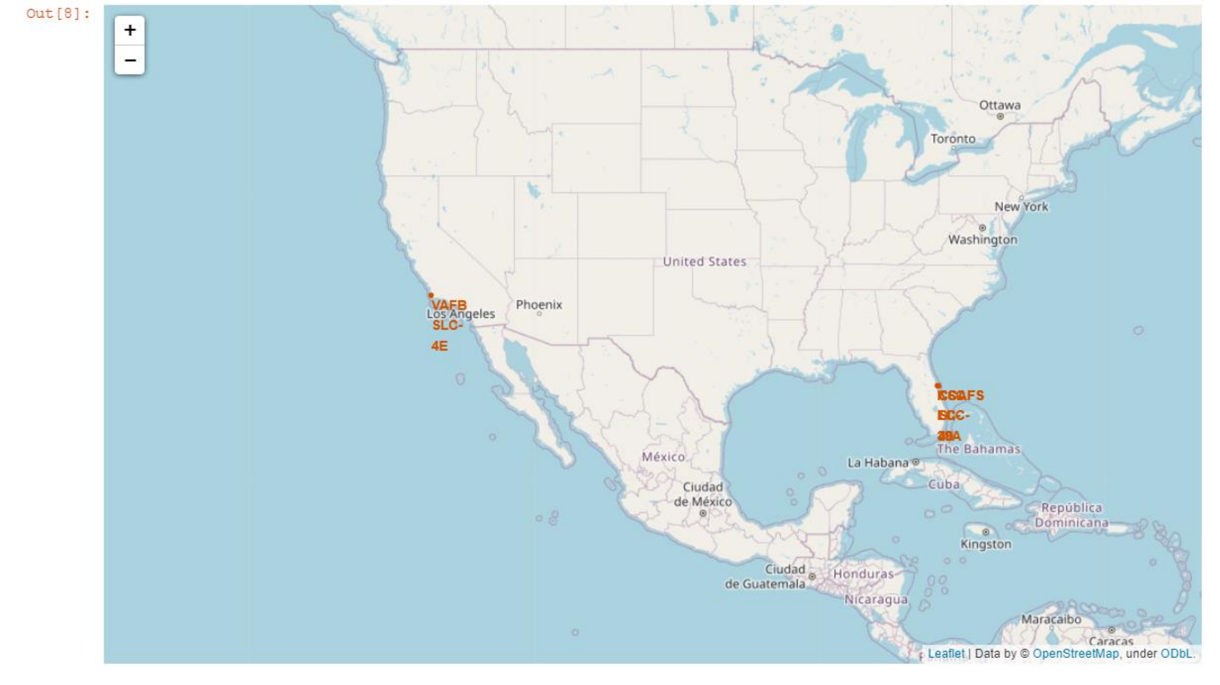
landing__outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

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

Section 3

Launch Sites Proximities Analysis

All launch sites



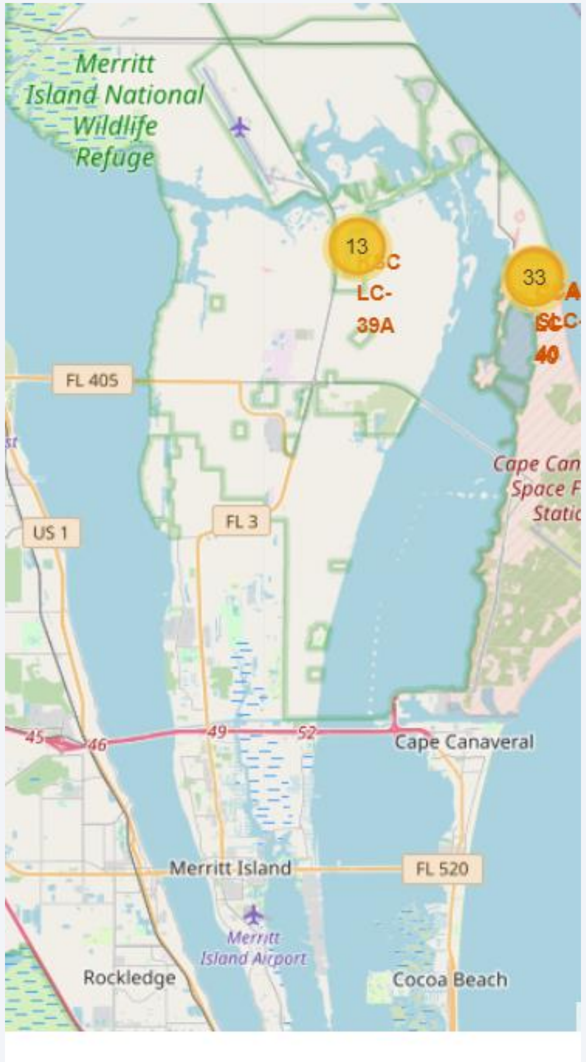
- Launch sites are near sea, probably by safety, but not too far from roads and railroads.

Launch outcomes by sites

- Green markers indicate successful and red ones indicate failure

Logistics and safety

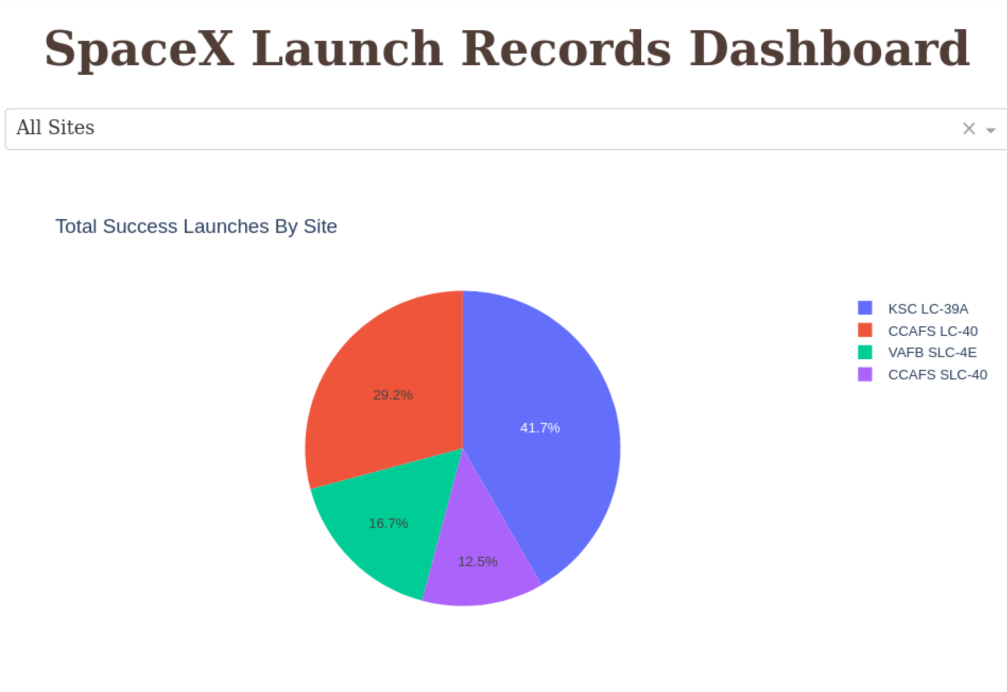
- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.



Section 4

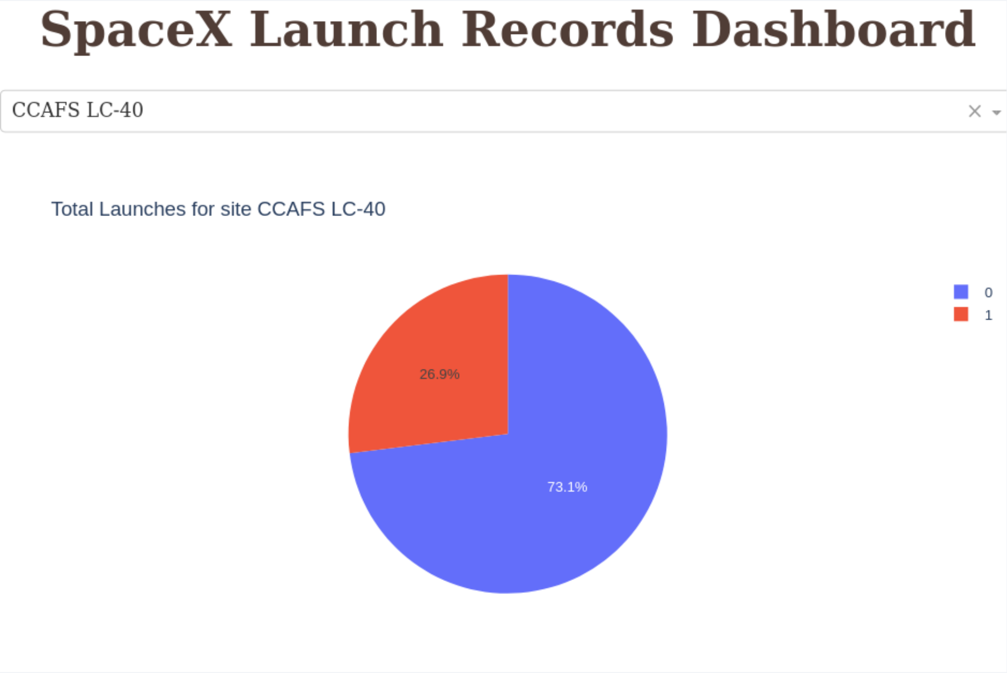
Build a Dashboard with Plotly Dash

Successful launches by site



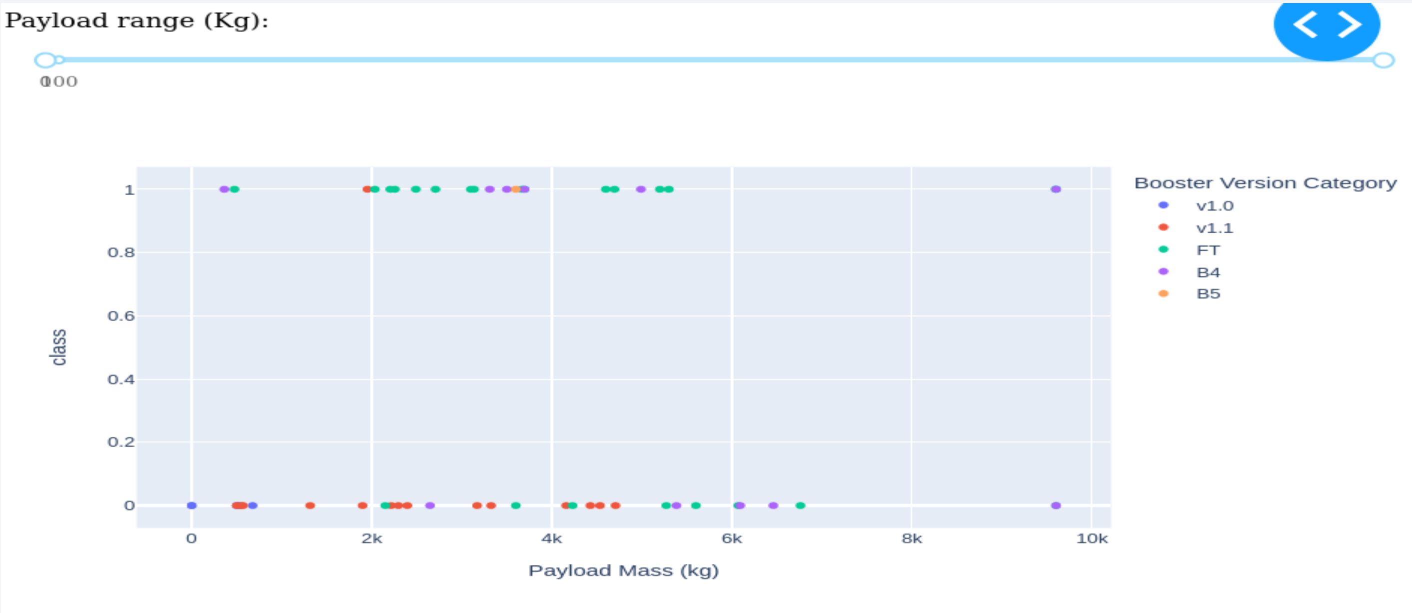
- The place from where launches are done seems to be a very important factor of success of missions.

<Launch success ratio for KSC LC-39A>



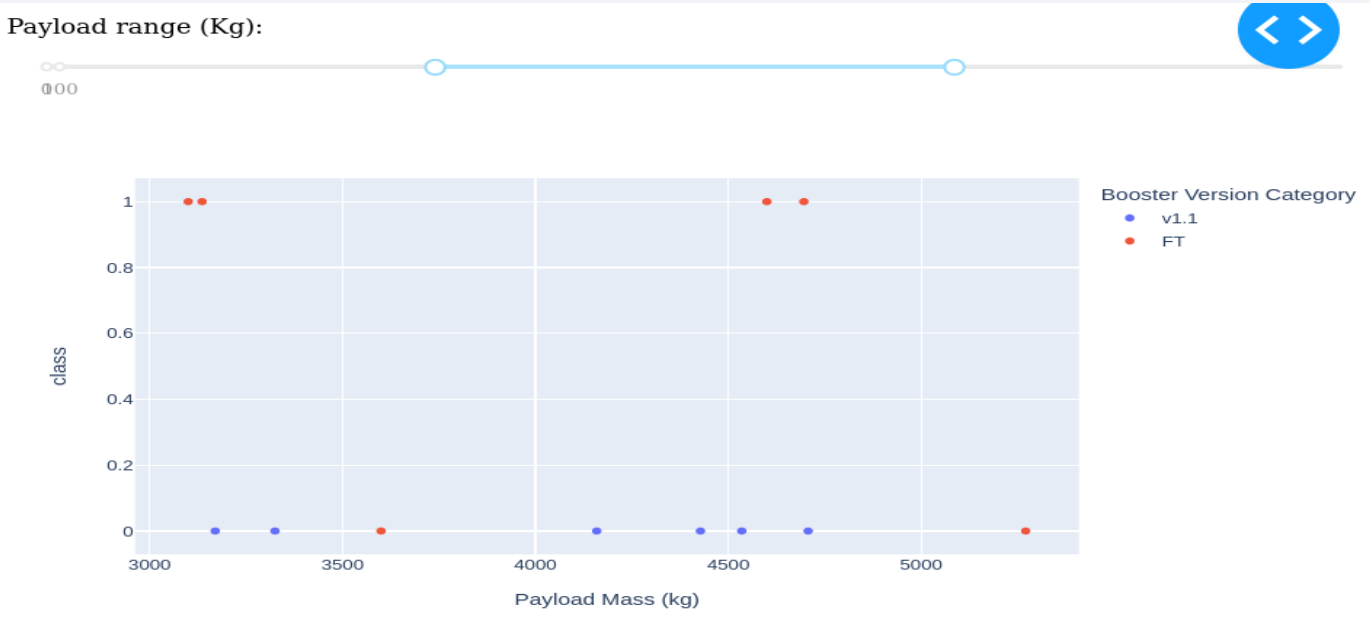
- 73.1% of launches are successful in this site.

Payload vs launch outcome



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

Payload vs launch outcome



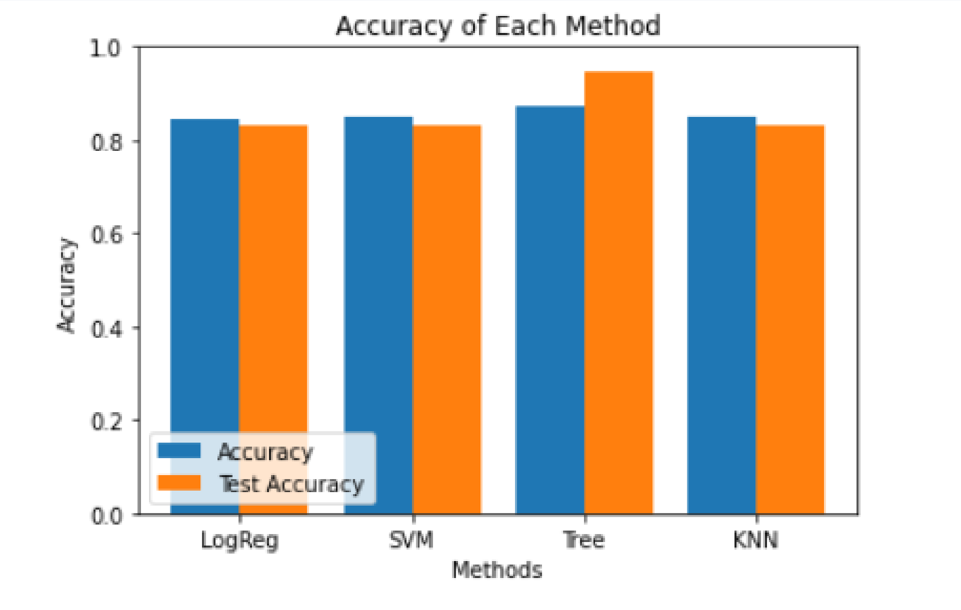
- Results of payloads between 3000kg and 5000kg

Section 5

Predictive Analysis (Classification)

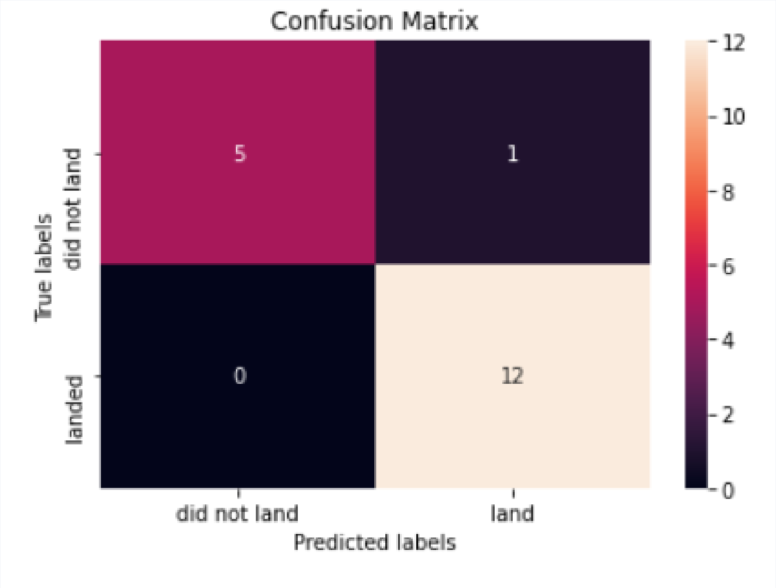
Classification Accuracy

- 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.
- 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

- Special thanks to:
    - Coursera
    - IBM

Thank you!