DATA SCIENCE PORTFOLIO

Min Soe Thant

CONTENTS

CERTIFICATES

PROJECTS

CERTIFICATES

STATEMENT OF ACCOMPLISHMENT

#24,214,277

HAS BEEN AWARDED TO

Min Soe Thant

FOR SUCCESSFULLY COMPLETING

Introduction to Python

LENGTH

4 HOURS

COMPLETED ON

AUG 29, 2022

datacamp

Jonathan Cornelissen, CEO

10 Courses

What is Data Science?

Tools for Data Science

Data Science Methodology

Python for Data Science, AI & Development

Python Project for Data Science

Databases and SQL for Data Science with Python

Data Analysis with Python

Data Visualization with Python

Machine Learning with Python

Applied Data Science Capstone

May 6, 2022

MIN SOE THANT

has successfully completed the online, non-credit Professional Certificate

IBM Data Science

In this Professional Certificate, learners developed and honed hands-on skills in Data Science and Machine Learning. Learners started with an orientation of Data Science and its Methodology, became familiar and used a variety of data science tools, learned Python and SQL, performed Data Visualization and Analysis, and created Machine Learning models. In the process they completed several labs and assignments on the cloud including a Capstone Project at the end to apply and demonstrate their knowledge and skills.

Rav Ahuja
AI & Data Science
Program Director
IBM Skills Network

Verify this certificate at:
https://https://coursera.org/verify/professional-cert/VKHKFYM3EQQ8

Projects

- Project 1: Bank Customer Churn Prediction Project

https://github.com/MinSoeThant/Churn-Prediction-Project.git

- Project 2: IBM Data Science Capstone Project

IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Min Soe Thant>
<23042022>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Methodologies Used:

- Data collection using Web Scraping and SpaceX API
- Data Wrangling
- Exploratory Data Analysis (EDA) with Data Visualisation
- Predictive Analysis using Machine Learning

Summary of all results:

- EDA with data visualization helps to see analytics presented visually, so they can grasp difficult concepts or identify new patterns
- Predictive analysis using Machine Learning shows the best model to predict successful Stage 1 landings

Introduction

The objective is to evaluate whether Space Y can compete with SpaceX

Problems you want to find answers

- To determine the price of each launch
- To determine whether SpaceX will reuse the rockets
- To determine whether the Stage 1 of the rockets can land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from SpaceX was obtained from:
  - SpaceX API (https://api.spacexdata.com/v4/rockets/)
  - Webscraping (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and Falcon Heavy launches)
- Perform data wrangling
  - Collected data is improved by creating a landing outcome label based on outcome data after summarizing and analysing features

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 that was collected until this step were normalized, divided in training and test data sets and evaluating by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

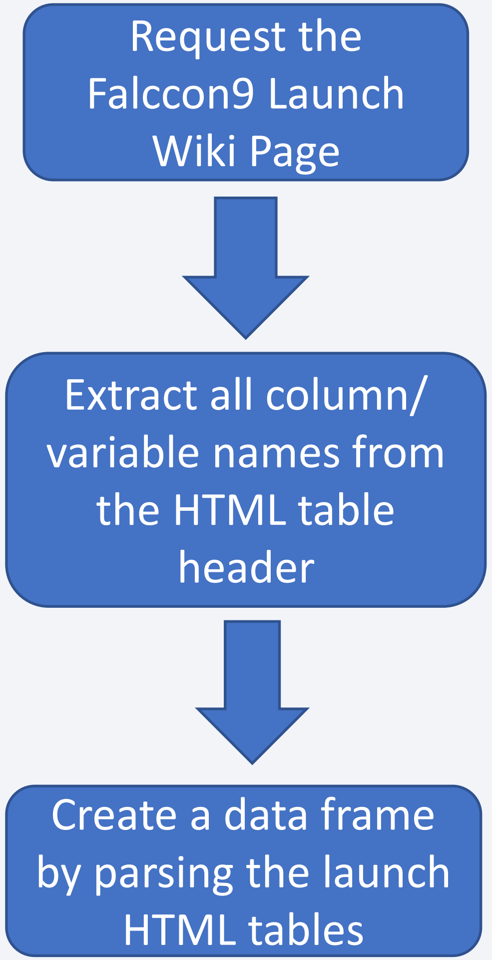
Data Collection

- Data collection process involved a combination of API requests from Space X public API and web scraping data from a table in Space X's Wikipedia entry.

- The next slide will show the flowchart of data collection from API and the following slide will show the flowchart of data collection from web scraping

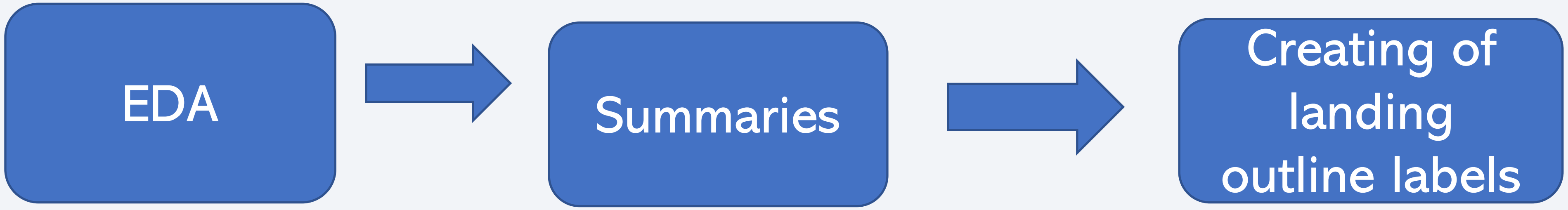
Data Collection – SpaceX API

Data Collection - Scraping



Data Wrangling

- Initially EDA was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- Finally, the landing outcome label was created from Outcome column.



EDA with Data Visualization

- To explore data, scatterplots and bar graphs were used to visualize the relationship between pair of features which are:
- Payload Mass and Flight Number,
- Launch Site and Flight Number,
- Launch Site and Payload Mass,
- Orbit and Flight Number,
- Payload and Orbit

EDA with SQL

- The following SQL queries were performed:
- Names of the unique 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;
- Average payload mass carried by booster version Falcon9;
- 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 6000kg;
- Total number of successful and failure mission outcomes

Build an Interactive Map with Folium

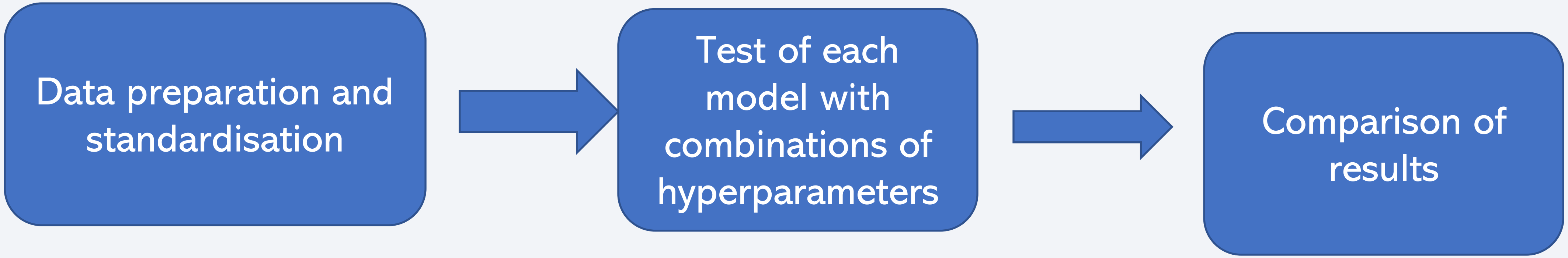
- Markers, circles, lines and marker clusters were used with Folium
- Markers indicate points like launch sites;
- Circles indicate highlighted areas specific coordinates, like NASA Johnson Space Centre;
- Marker clusters indicates groups of events in each coordinate, like launches in a launch and
-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

Build a Dashboard with Plotly Dash

- The following graphs and plots were used to visualize data
- Percentage of launces by site
- Payload range
- This combination allowed to quickly analyse the relation between payloads and launch sites, helping to identify where is best place to launch in accordance with payloads.

Predictive Analysis (Classification)

- Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbours

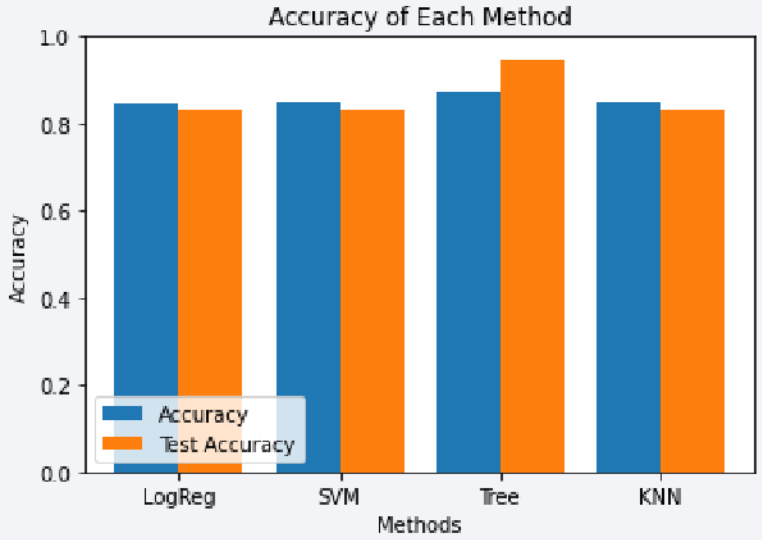


Results

- 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 launches happens at east cost launch sites.

Results

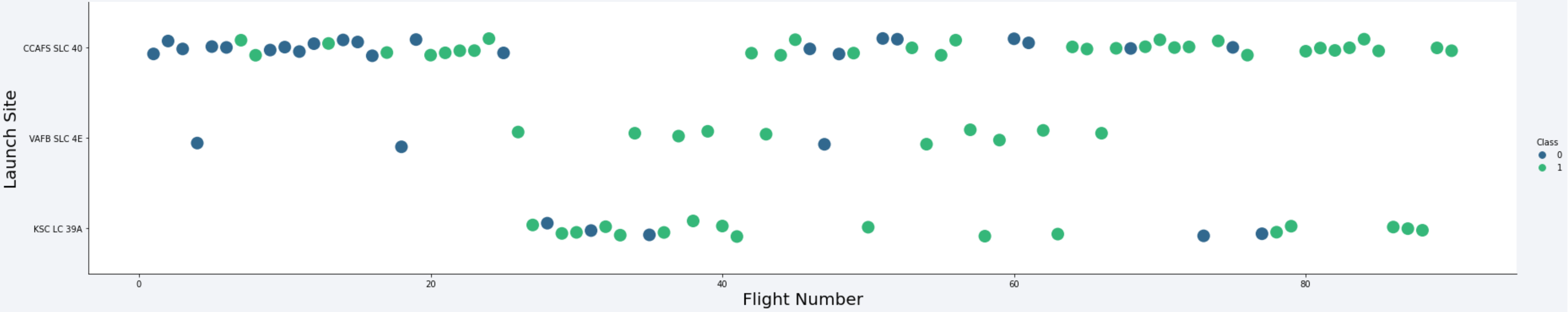
- Predictive Analysis showed that Decision Tree Classifier is the best model to predict successful landings, having accuracy over 87% and accuracy for test data over 94%.



Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



green indicates successful launch   blue indicates unsuccessful launch

The overall success rate increases over time.

Likely a big breakthrough around flight 20 which significantly increased success rate.

CCAFS appears to be the main launch site as it has the most volume.

Payload vs. Launch Site

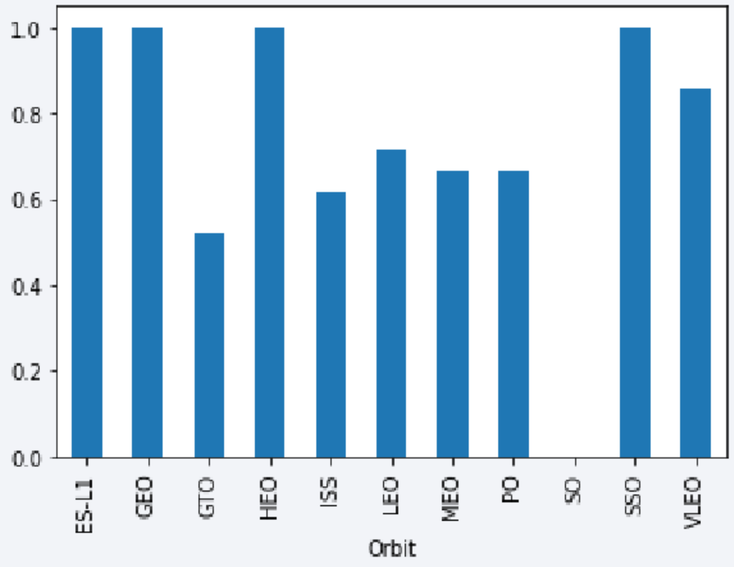
green indicates successful launch   blue indicates unsuccessful launch

Payload mass appears to fall mostly between 0-6000 kg.

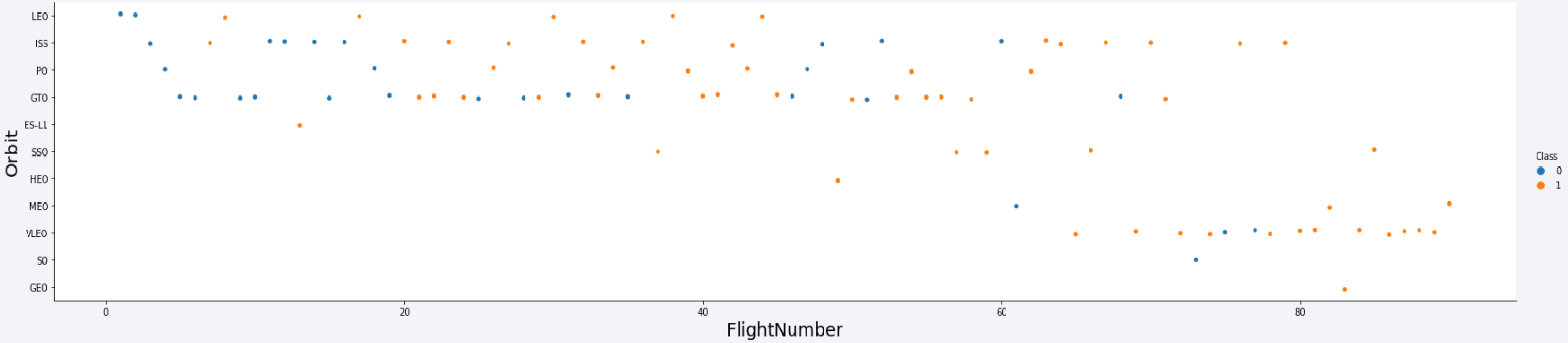
Different launch sites also seem to use different payload mass.

Success Rate vs. Orbit Type

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

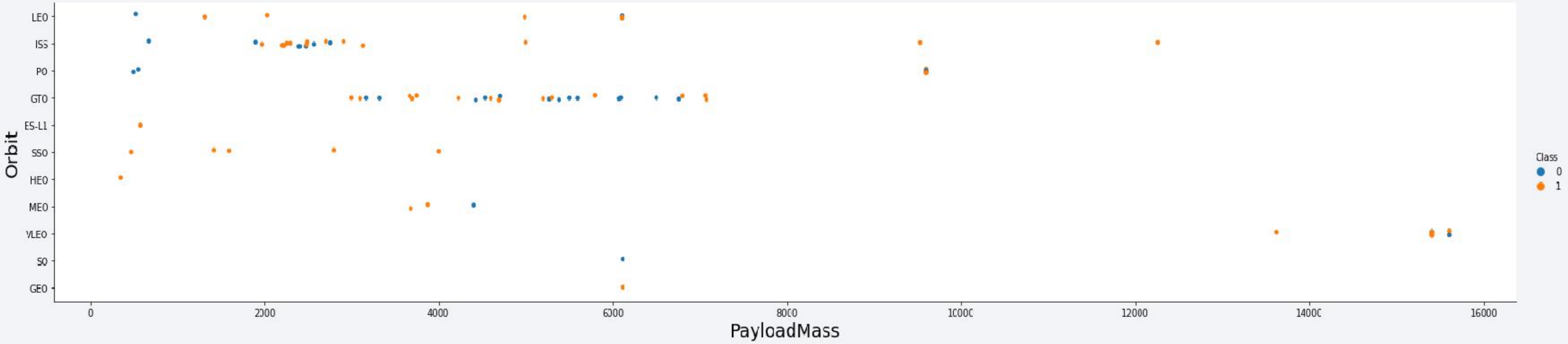


Flight Number vs. Orbit Type



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

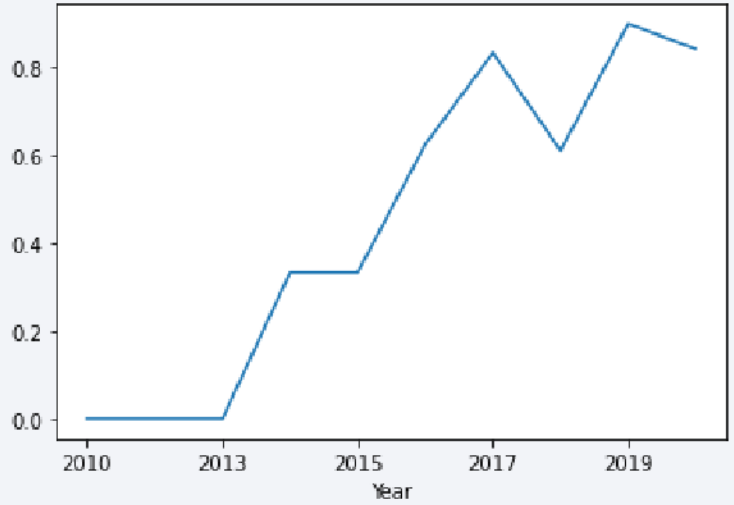
Payload vs. Orbit Type



- 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 carried by boosters from 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

- 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

- 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

- 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

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:

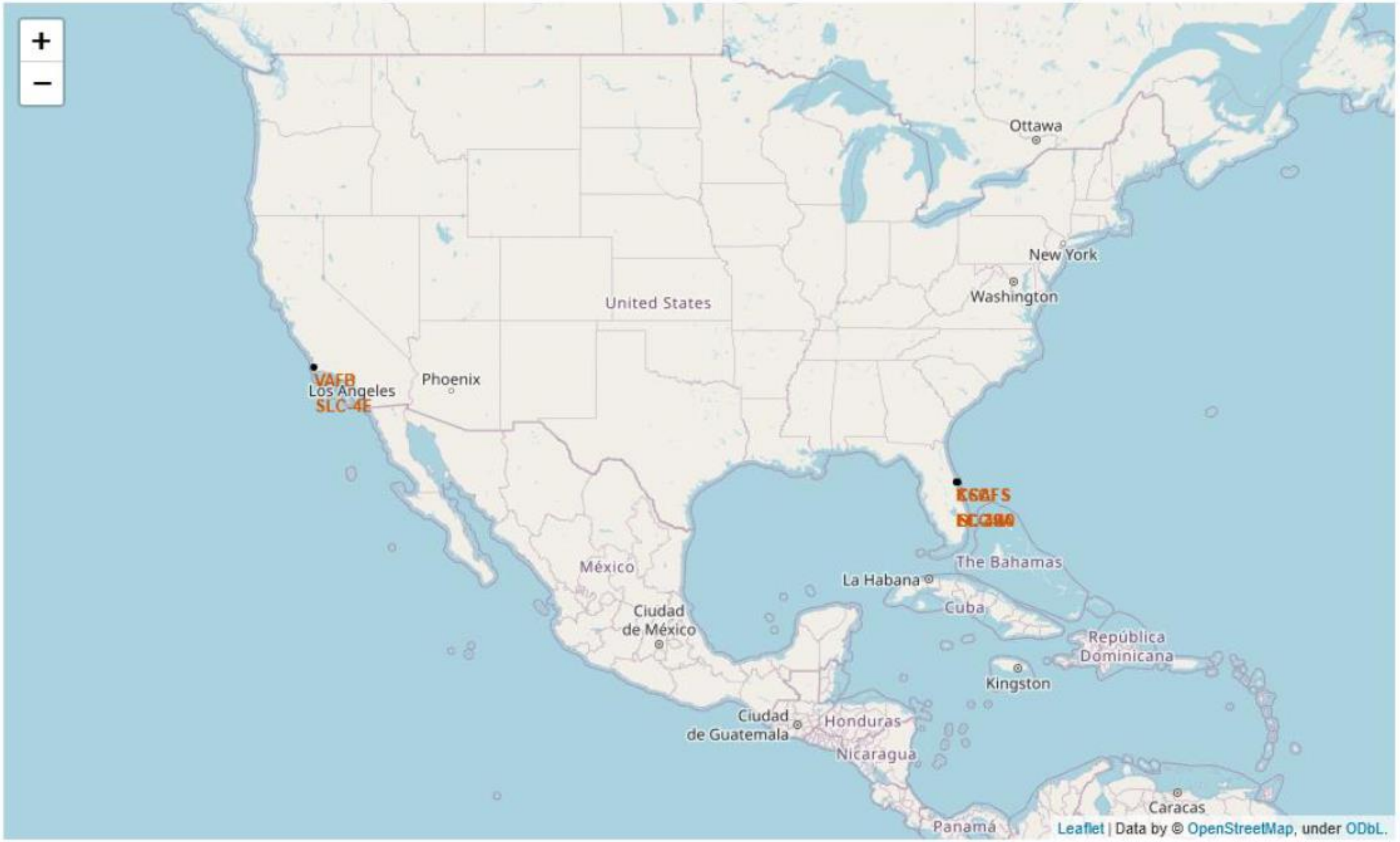
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

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

Section 3

Launch Sites Proximities Analysis

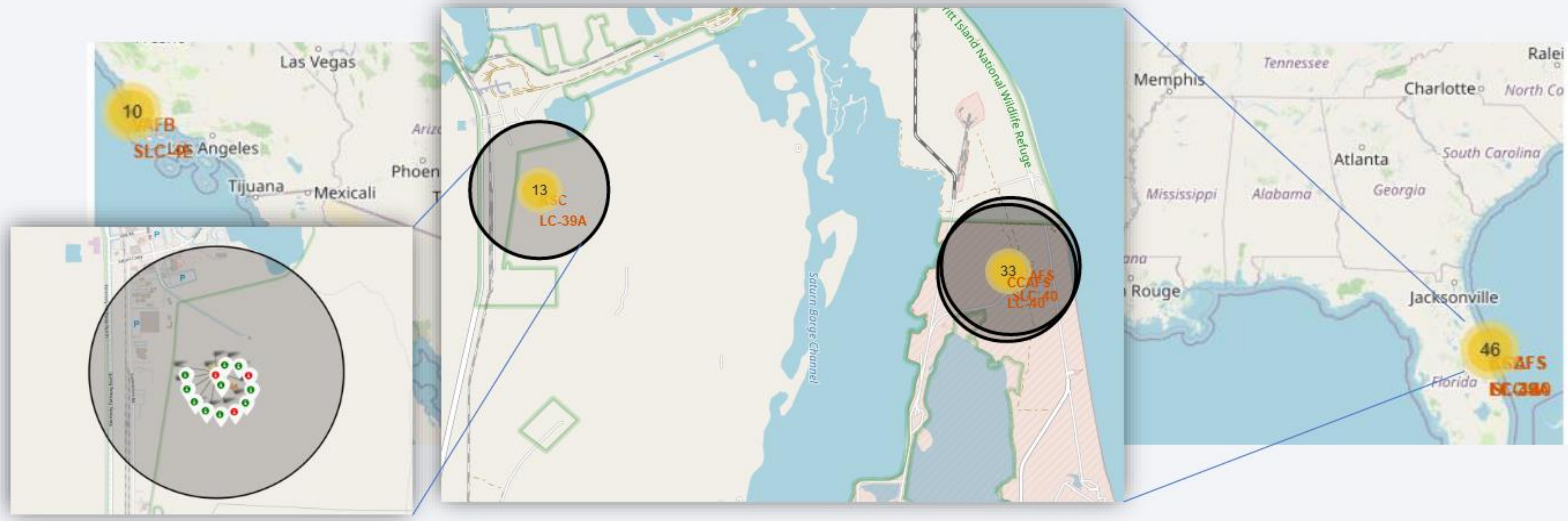
All launch sites



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

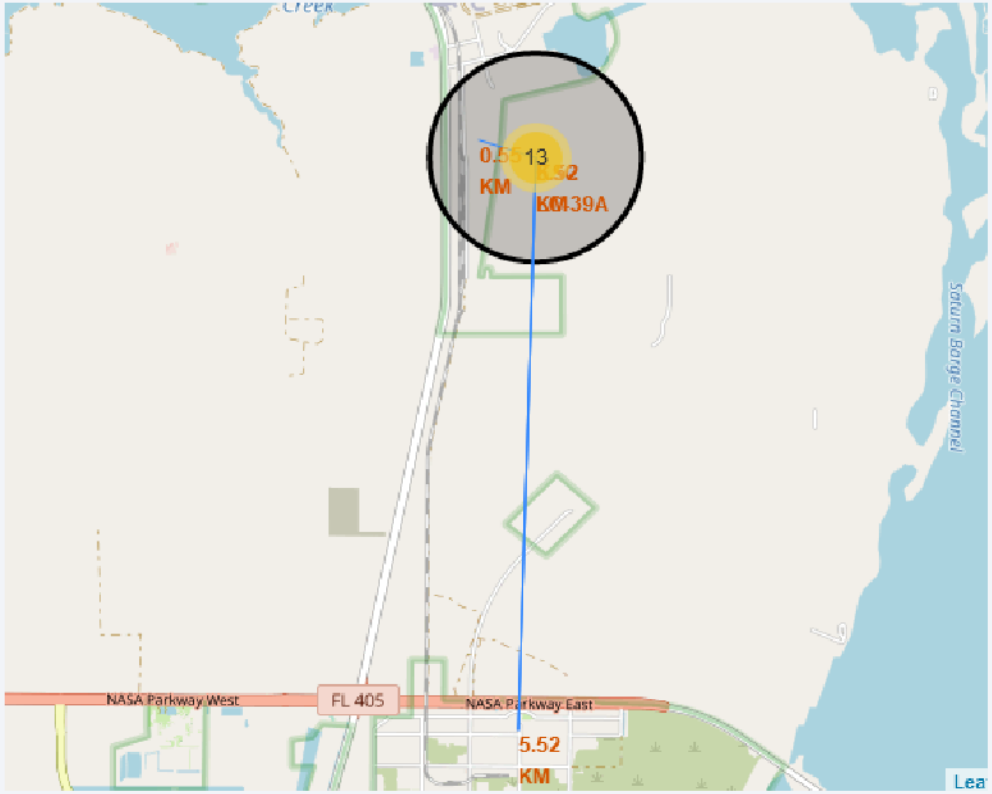
Launch Outcomes by Site

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



- Green markers indicate successful and red ones indicate failure.

Logistics and Safety

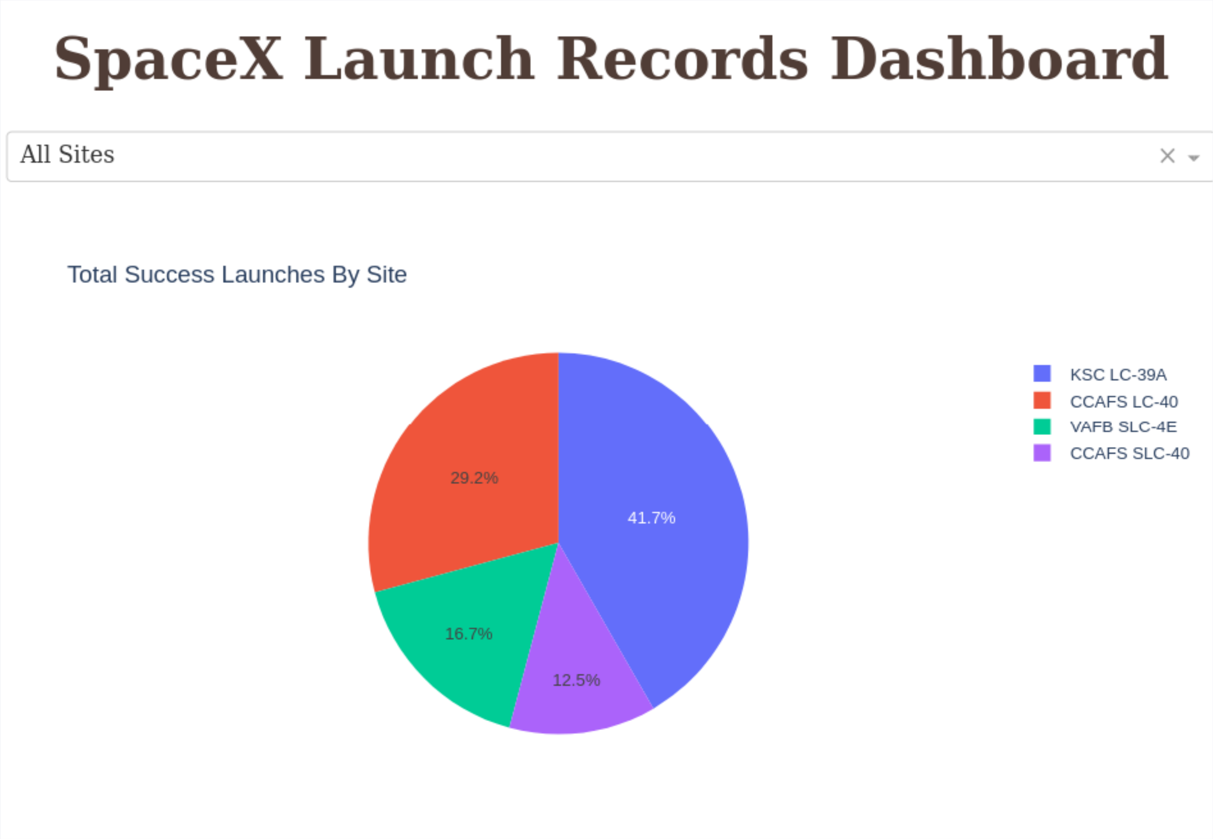


- Launch site KSC LC-39A has good logistics aspects, being near railroad and road and relatively far from inhabited areas.

Section 4

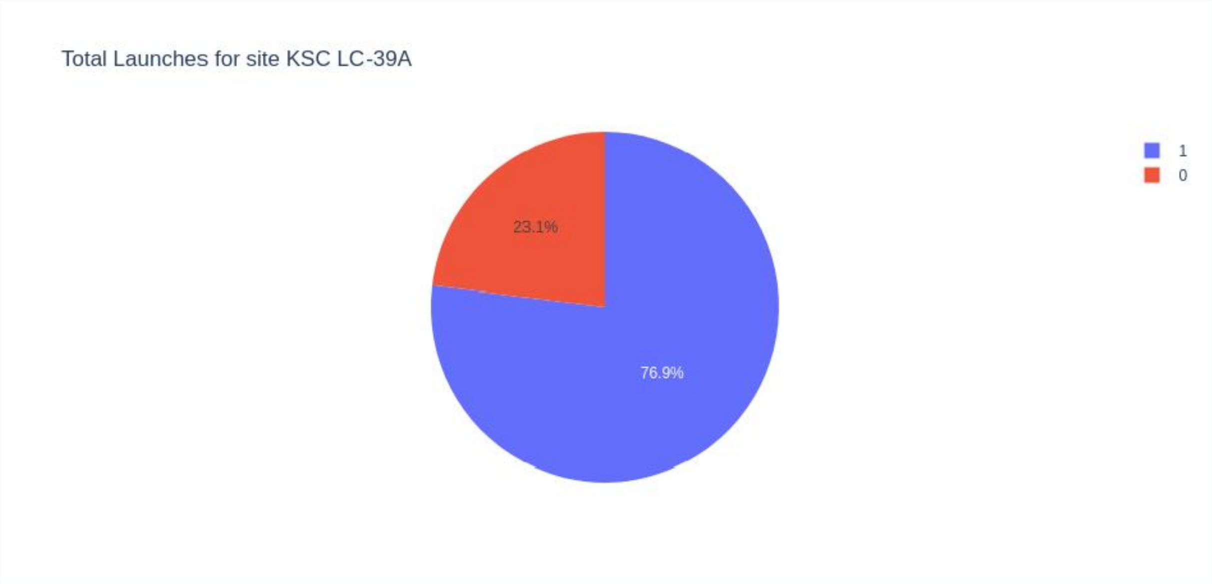
Build a Dashboard with Plotly Dash

Successful Launches by Site



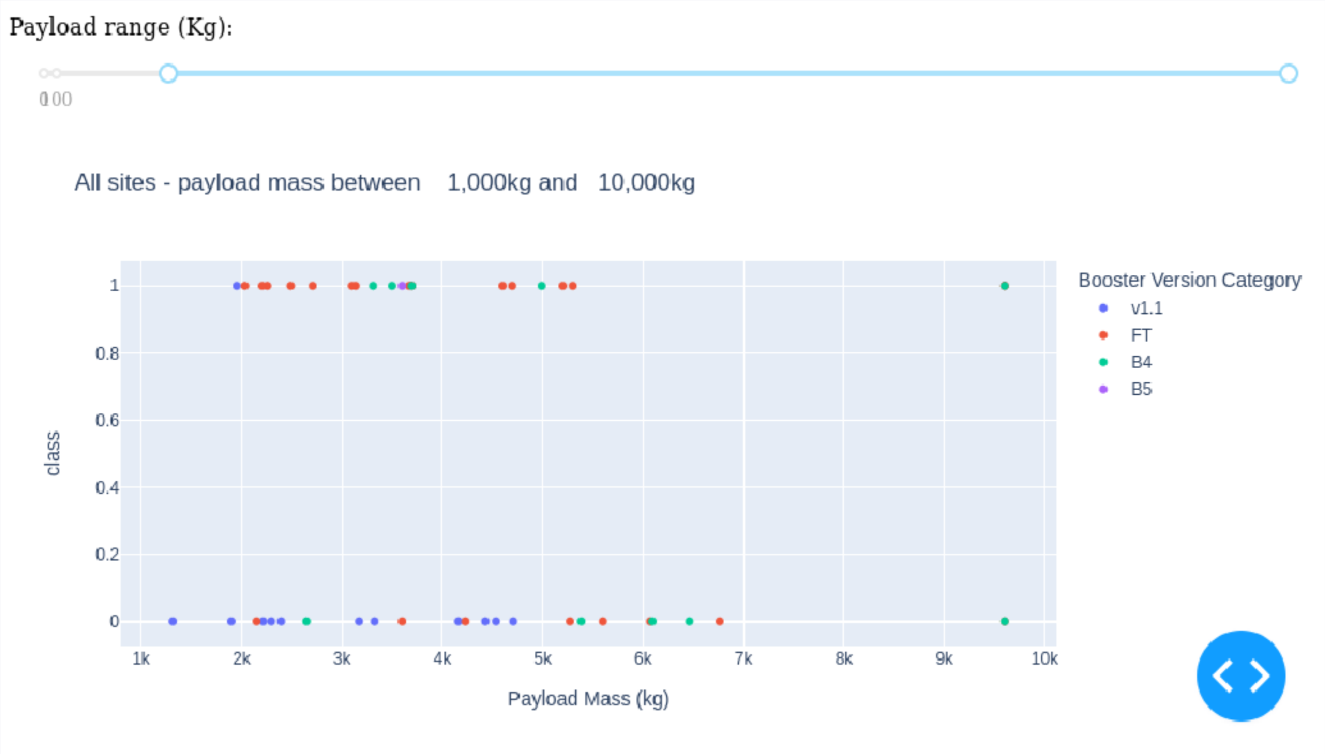
- The place from where launches are done seems to be a very important factor of success of missions.

Launch Success Ratio for KSC LC-39A



- 76.9% of launches are successful in this site.

Payload vs. Launch Outcome



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

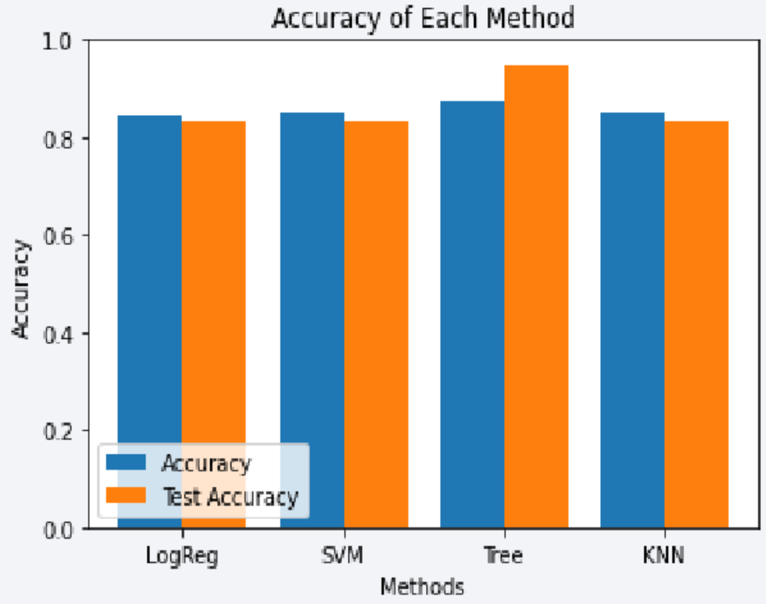
Section 5

Predictive Analysis (Classification)

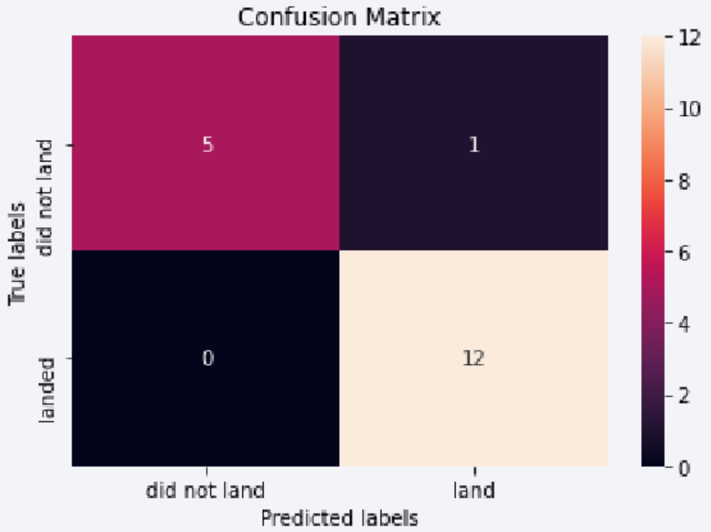
Classification Accuracy

- 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 of Decision Tree Classifier



- 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!