IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Amit Jain
September 24, 2022

Outline

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

Executive Summary

- The project provides insights on the different feature that affect the rocket launching outcomes.
- Predictive model is developed to predict the outcome of the future launch based of various characteristic of the flight which includes launch site, booster version, payload mass, etc.
- Payload mass, orbit, launch site, booster version are key parameters that requires attention in the future launching
- Decision tree model predict the outcome with significantly good accuracy

Introduction

- The project provides insights on the different feature that affect the rocket launching outcomes.
- Predictive model is developed to predict the outcome of the future launch based of various characteristic of the flight which includes launch site, booster version, payload mass, etc.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - API calls and webscrping are used for data collection
- Perform data wrangling
  - Key features are used in the predictive analysis and classification features are converted into a numeric ones.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Models are build by tuning hyper parameters using GridSearchCV algorithm and then assessed for accuracy by dividing the dataset into training and testing sets.

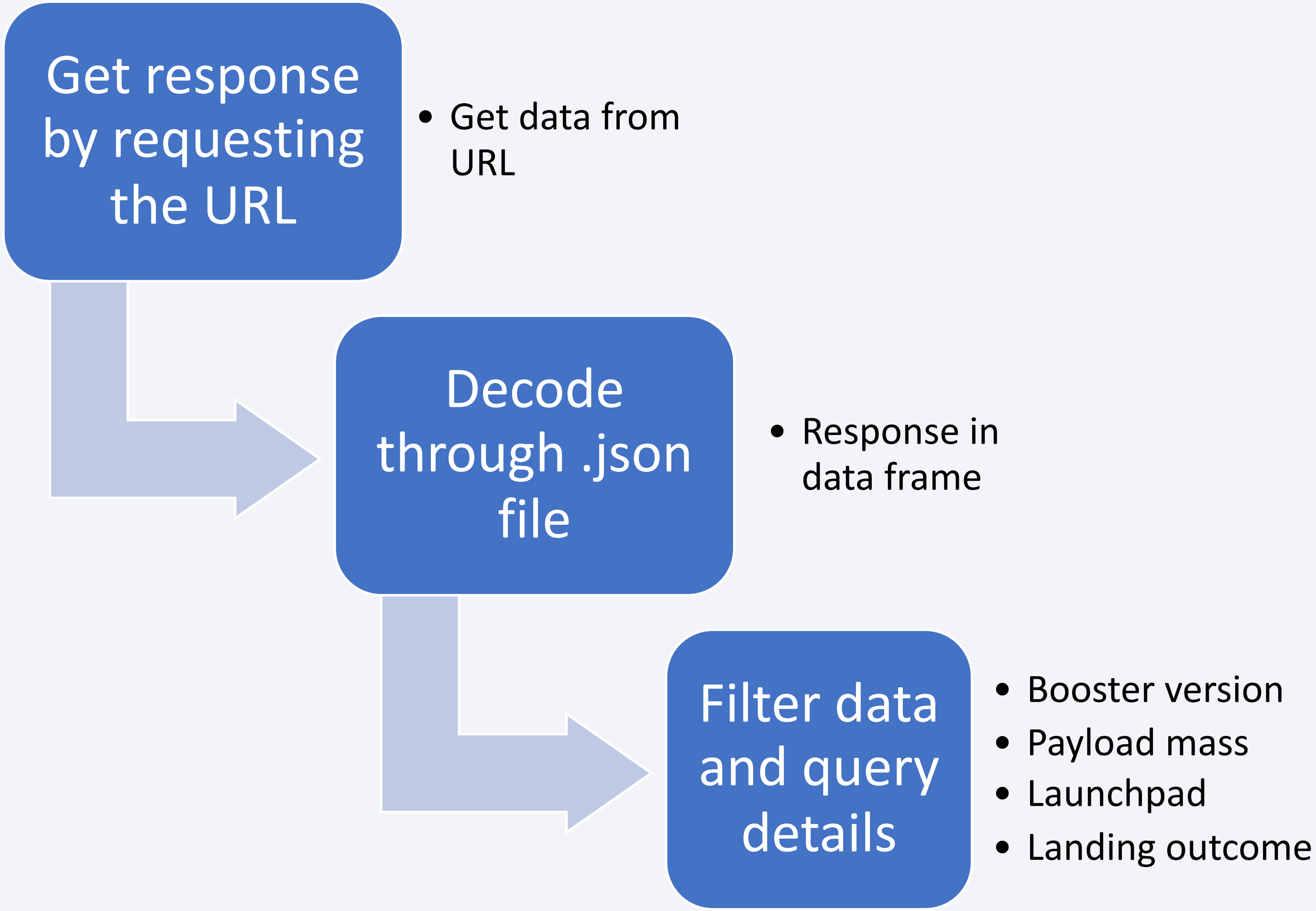
Data Collection

- Data sets were collected using two methods
  - Calling API
  - Web scraping
- API calls are by requesting response, modifying data using .json file and converting into a dataframe
- In web scraping, Wikipedia webpage is used to obtain information using ‘html’ parsing and then converting into a dataframe.

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



https://github.com/jain651/spacex_stage1_landing/blob/main/w1_collection_api.ipynb

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Get response with beautiful soup

- Get data from SpaceX wiki page

Parse the webpage using HTML parser

- Extract required table

Extract table rows into data frame

- Required data frame

https://github.com/jain651/spacex_stage1_landing/blob/main/w1_webscraping.ipynb

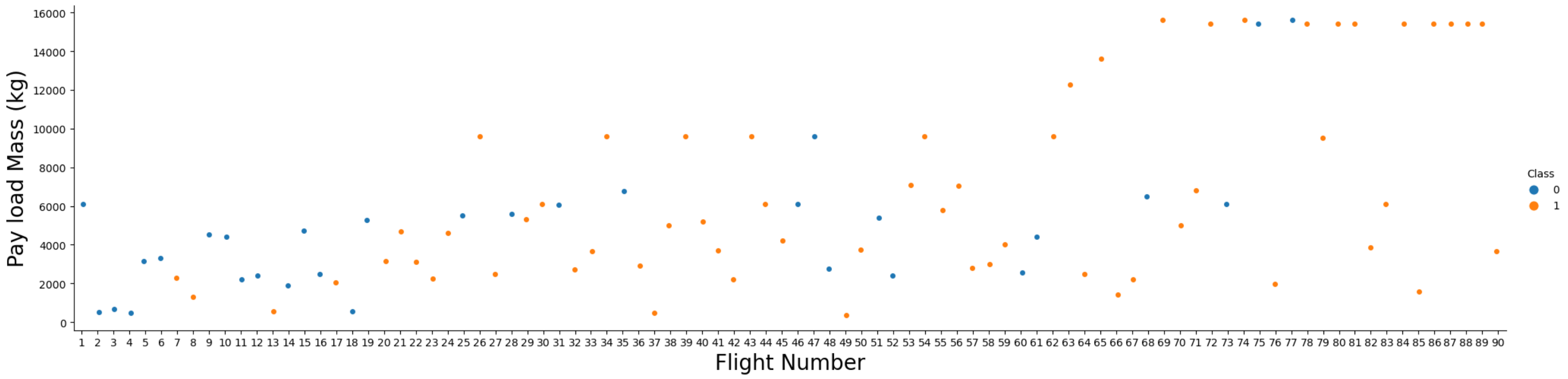
Data Wrangling

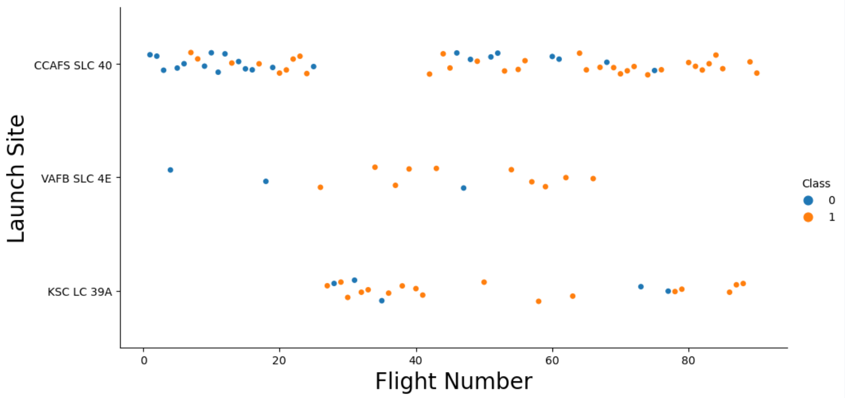
- At first the fraction of null values, data types were evaluated.
- Number of different orbits, launching sites, landing outcomes were extracted
- The landing outcomes were reduced binary class: ‘successful’ or ‘failure’

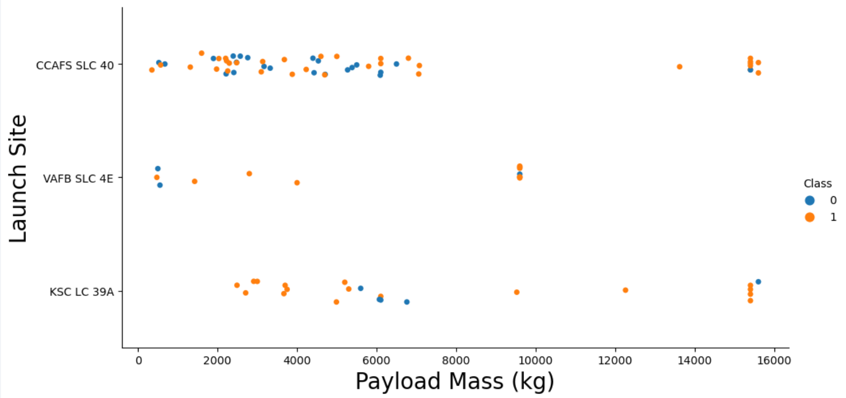
- GitHub URL: https://github.com/jain651/spacex_stage1_landing/blob/main/w1_wrangling.ipynb

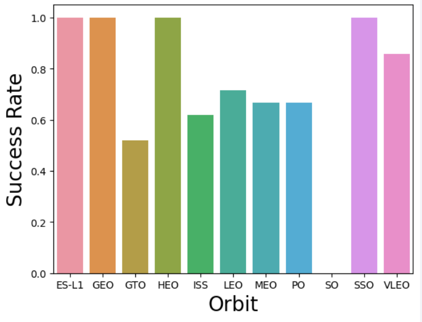
EDA with Data Visualization

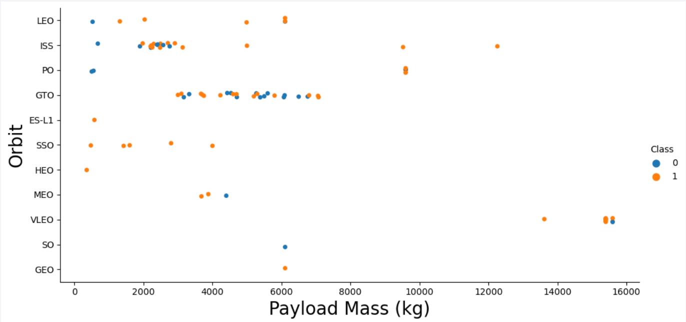
- Different features were plotted against each other to identify relationship between them
- One hot coding were done to replace categorical features into numerical
- GitHub URL: https://github.com/jain651/spacex_stage1_landing/blob/main/w2_eda_dataviz.ipynb

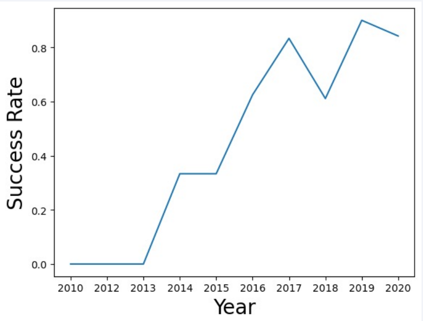










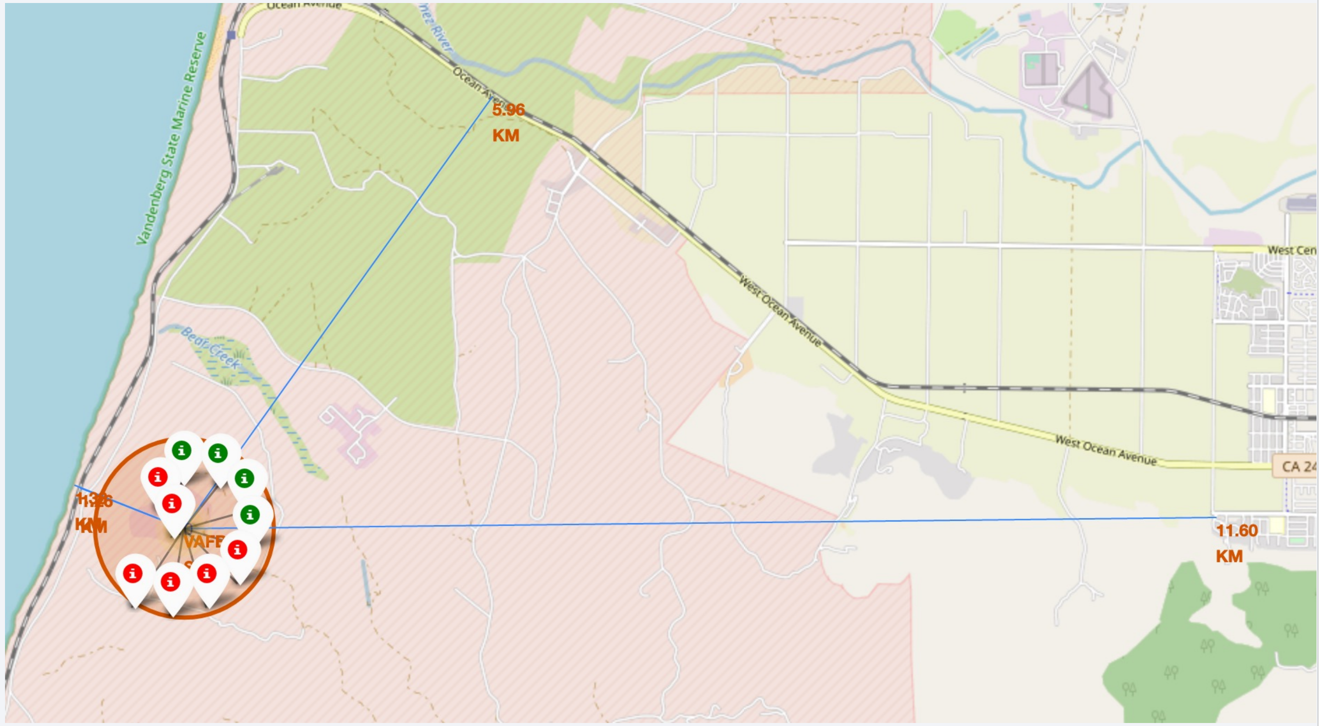


EDA with SQL

- Following SQL queries were performed to evaluate:
    1. Distinct launch sites
    2. Records where launch sites begin with '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 achieved
    6. List boosters having success in drone ship with payload between 4000 and 6000 kg
    7. Total number of successful and failure mission outcomes
    8. The booster versions that have carried the maximum payload mass
    9. Records for year 2015 displaying the month names, failure landing outcomes in drone ship, booster versions, and launch_site
    10. Rank the successful landing outcomes between 04-06-2010 and 20-03-2017
- GitHub URL: https://github.com/jain651/spacex_stage1_landing/blob/main/w2_eda_sqllite.ipynb

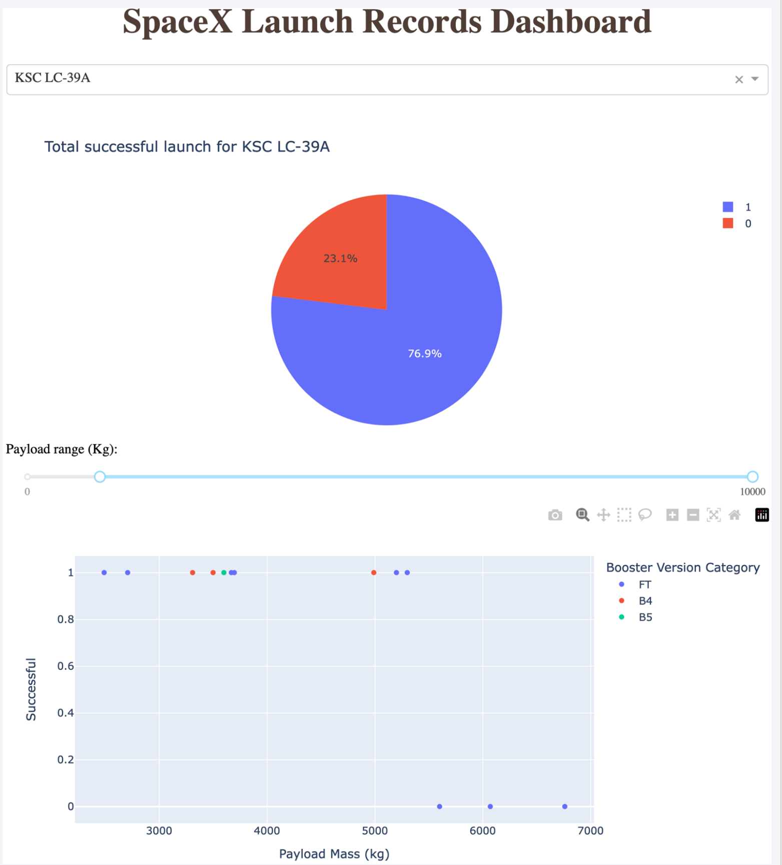
Build an Interactive Map with Folium

- Folium map is created to show launch sites their distances from nearest coastline, city, highway, and railway using following objects:
  - Markers
  - Circles
  - Lines
- These object quickly shows number of successful and failure attempts from each launching sites
- GitHub URL: https://github.com/jain651/spacex_stage1_landing/blob/main/w3_folium_viz.ipynb



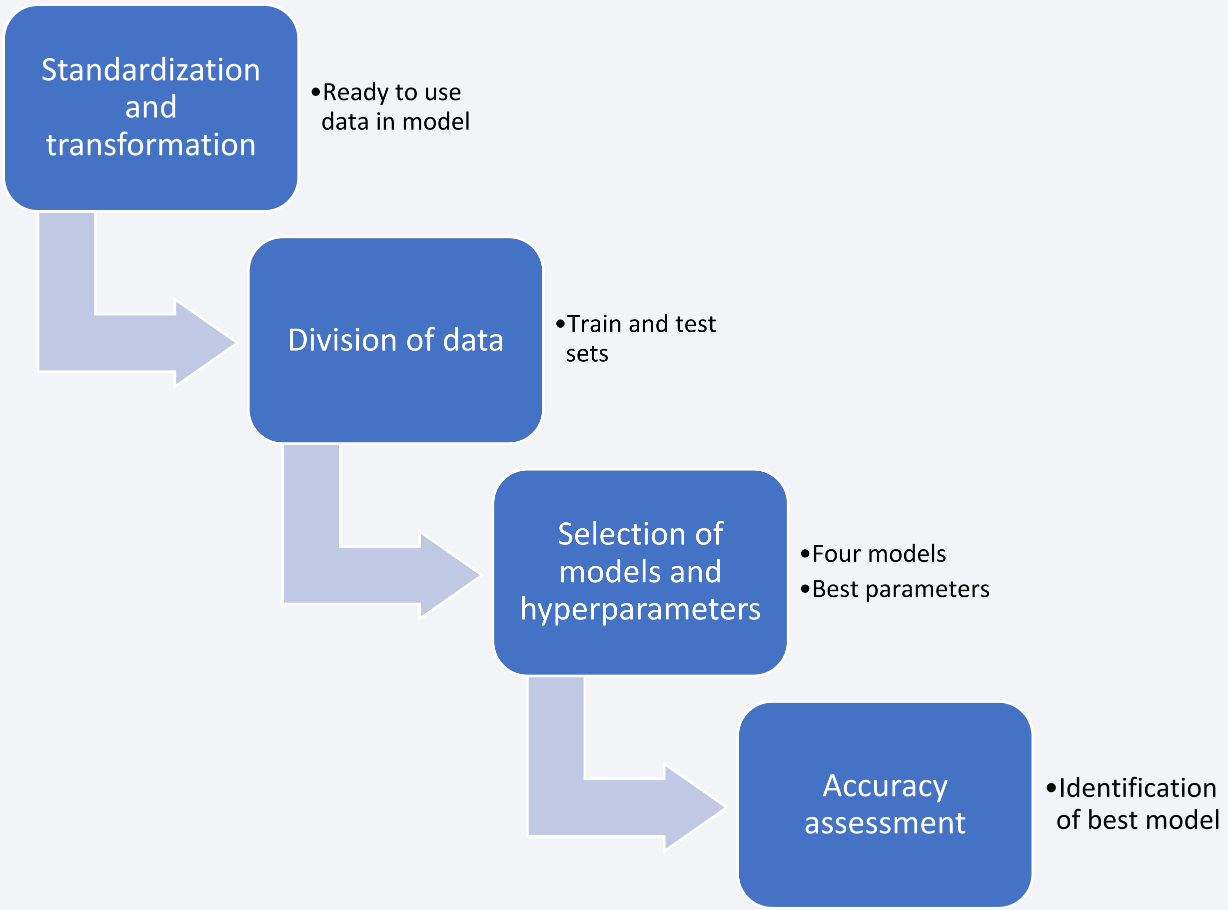
Build a Dashboard with Plotly Dash

- A glance on
  - Success rate for each launch site and all launch sites
  - Effect of payload mass on success and indication of the booster version category
- Dashboard is required to visually identify effectiveness of the launch site, payload mass and booster version on success.
- GitHub URL: https://github.com/jain651/spacex_stage1_landing/blob/main/w3_plotly_dash.py



Predictive Analysis (Classification)

- Following models were examined
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - k-Nearest Neighbor
- GridSearchCV was used to identify best hyperparameters for each model
- Decision tree was found to provide best prediction with 86% accuracy
- GitHub URL: https://github.com/jain651/spacex_stage1_landing/blob/main/w4_prediction.ipynb



Results

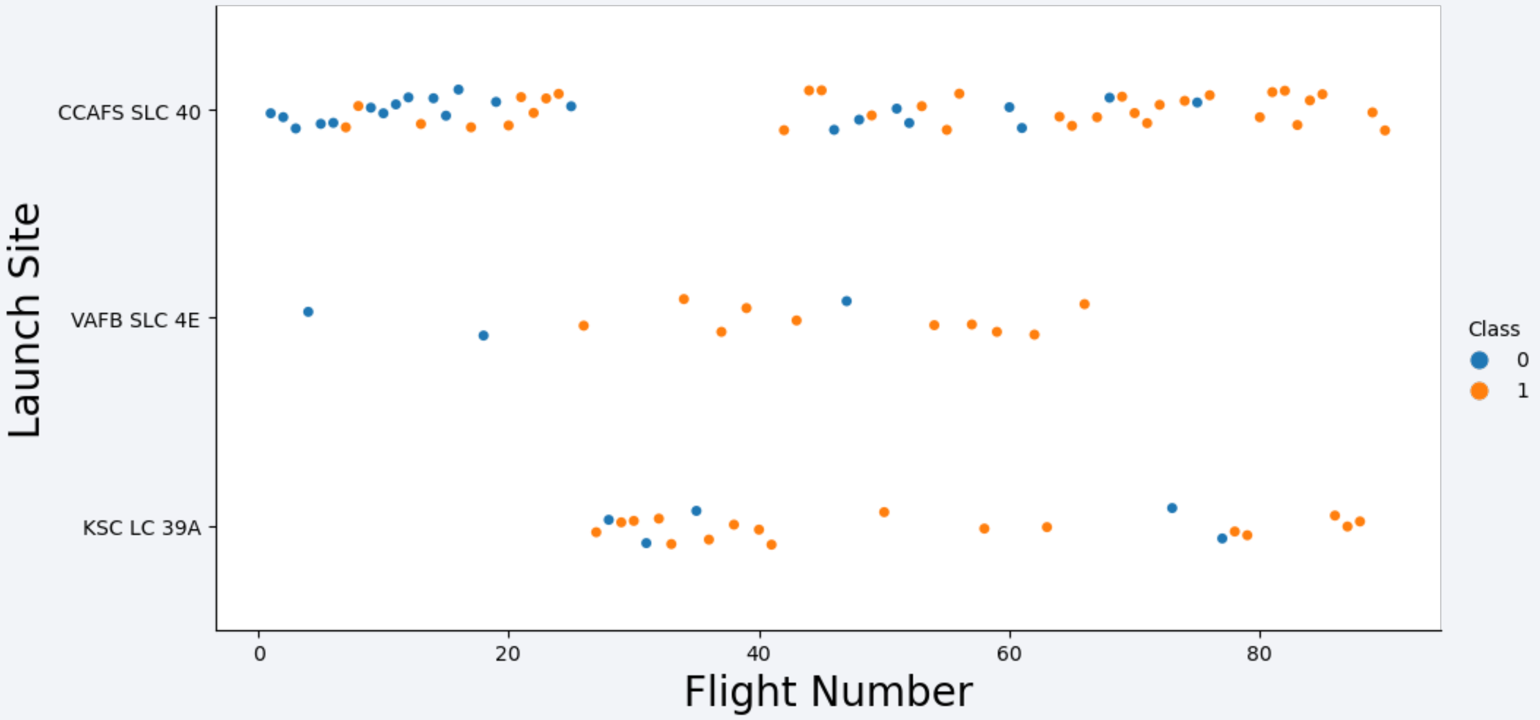
- Payload mass affect the accuracy of the results
- Success rate has significantly increased after 2014.
- Decision tree model provides best accuracy with 86%

Section 2

Insights drawn from EDA

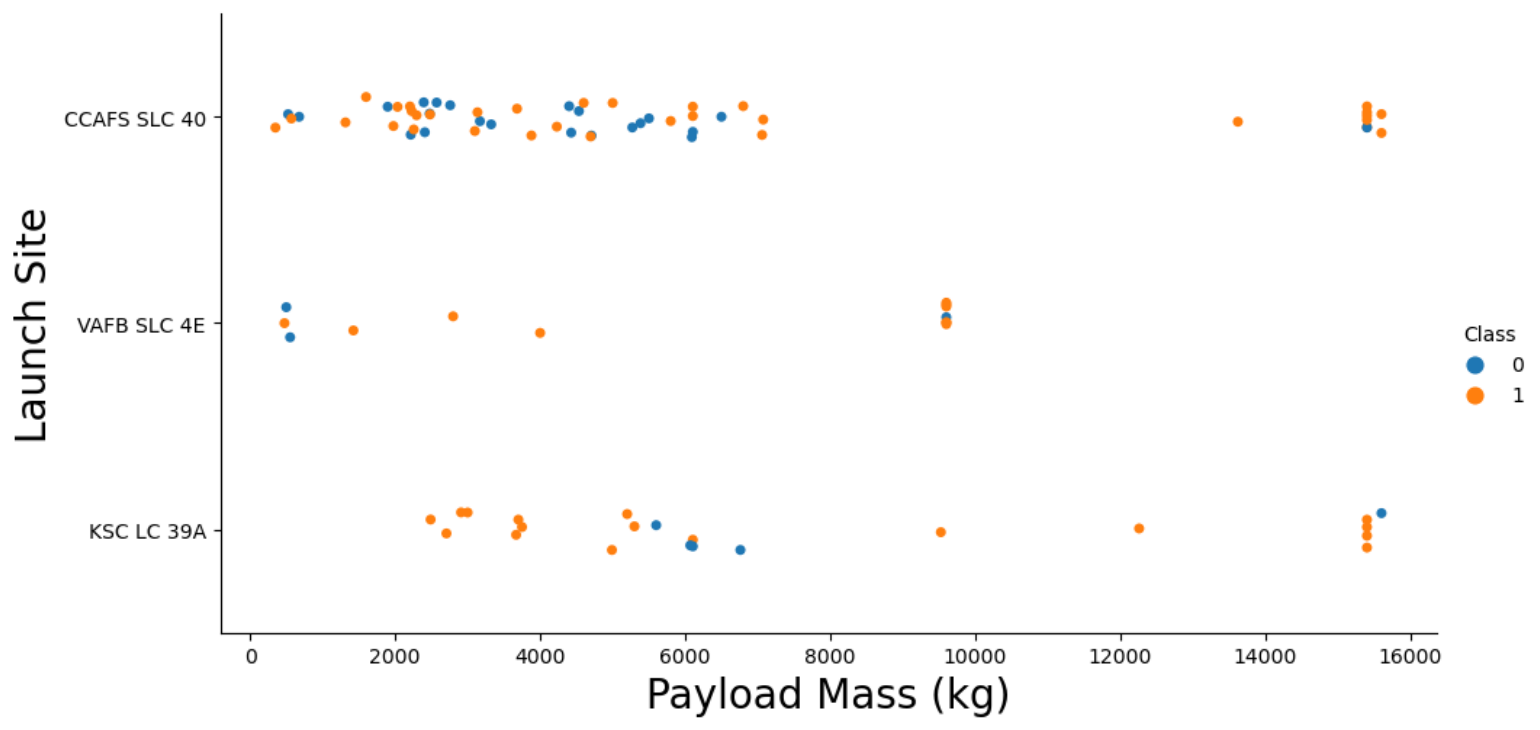
Flight Number vs. Launch Site

- Initial flights are mostly launched from CCAFS SLC 40.
- Most of the flights are launched from CCAFS SLC 40.
- KSC LC 39A has most successful launches.



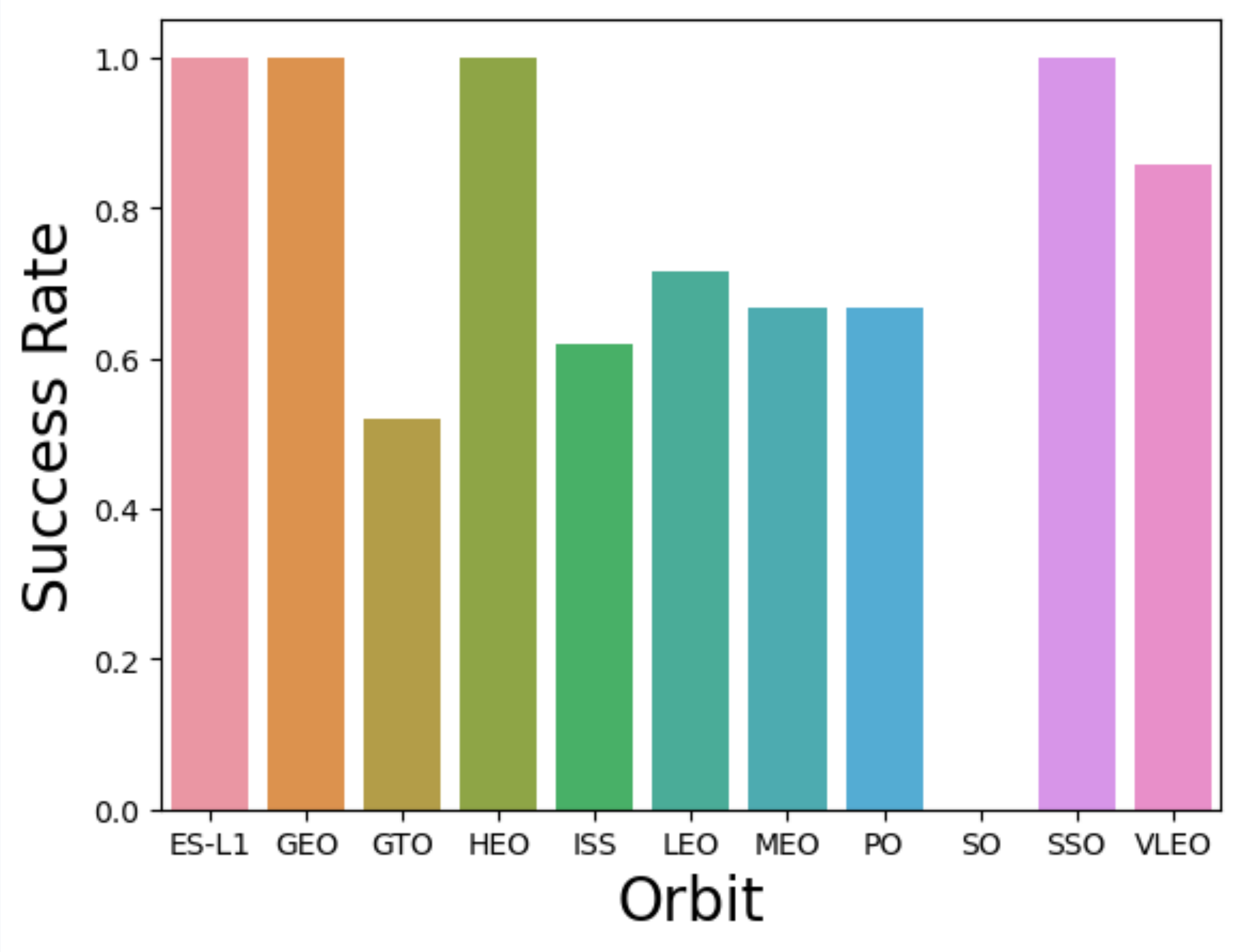
Payload vs. Launch Site

- Payload mass > 8000 kg has been most successful.
- Launch with heavy payload mass are done at either CCAFS SLC 40 or KSC LC 39A.



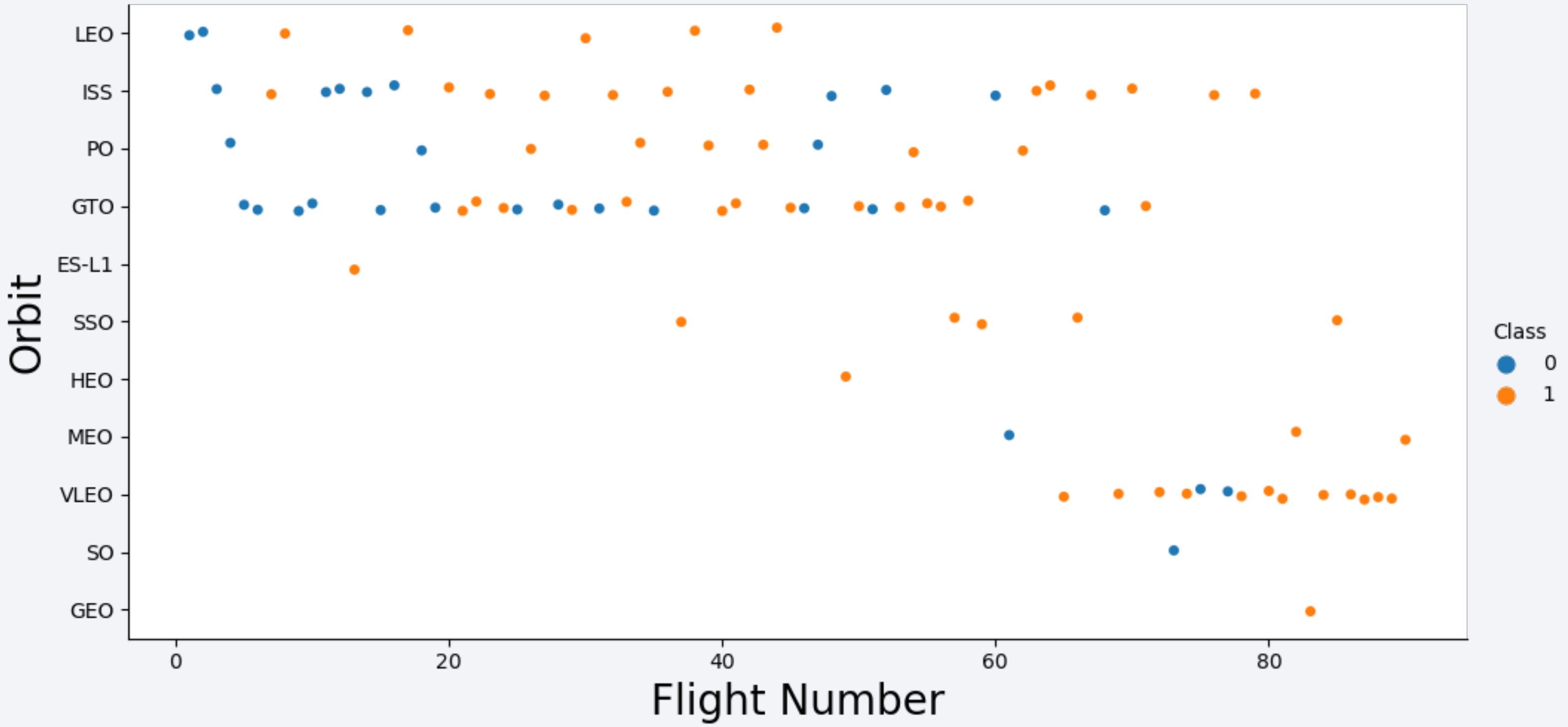
Success Rate vs. Orbit Type

- SO orbit has the least success rate followed by the GTO.
- ES-L1, GEO, HEO and SSO shows 100% success rate



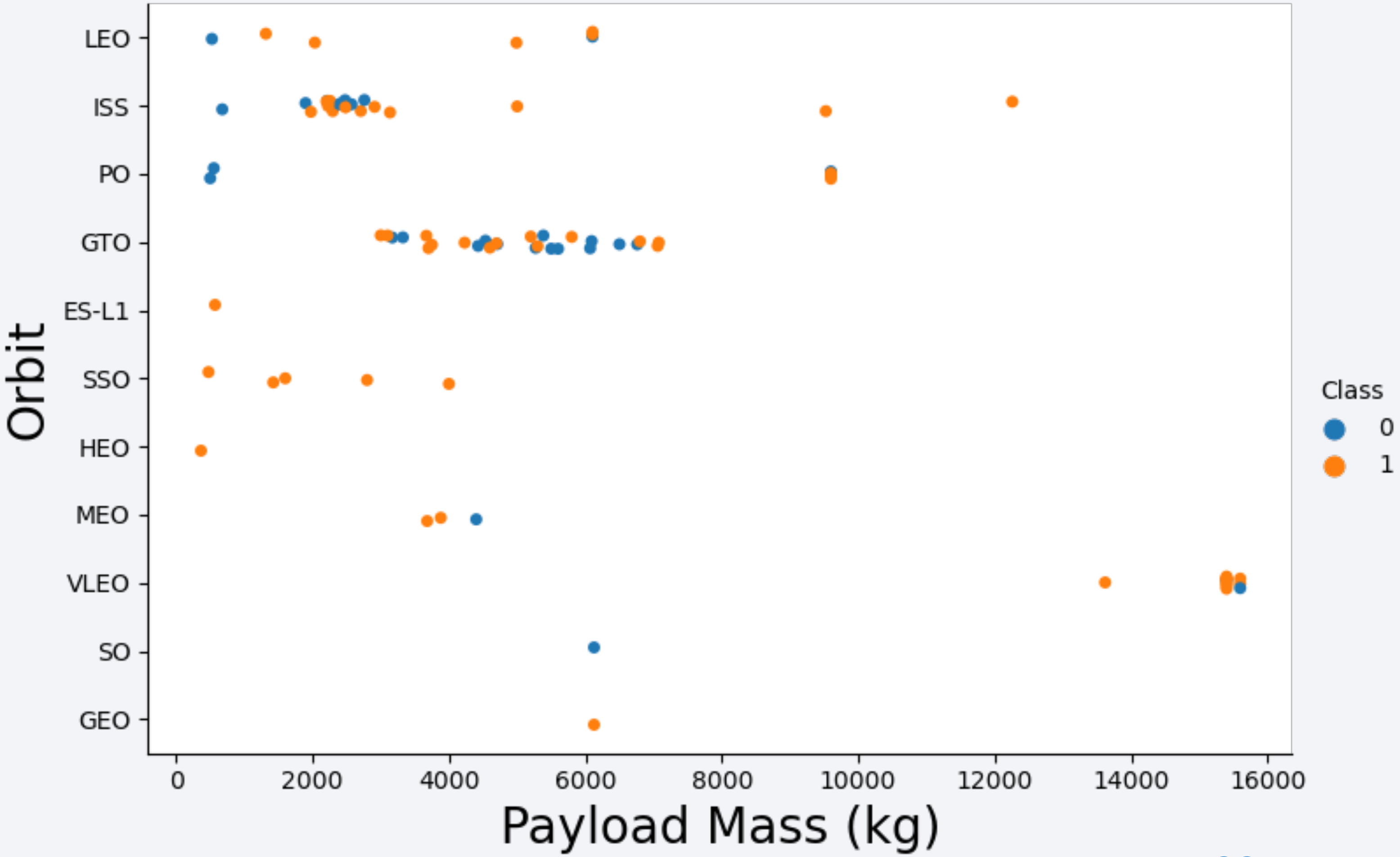
Flight Number vs. Orbit Type

- 100% success rate obtained in ES-L1, GEO, HEO and SSO is derived from very limited number of flights.
- Similarly, least success rate with SO is derived from only one flight.
- Recent flights are mostly targeted towards VLEO while initial flights were targeted towards LEO, ISS, GTO.



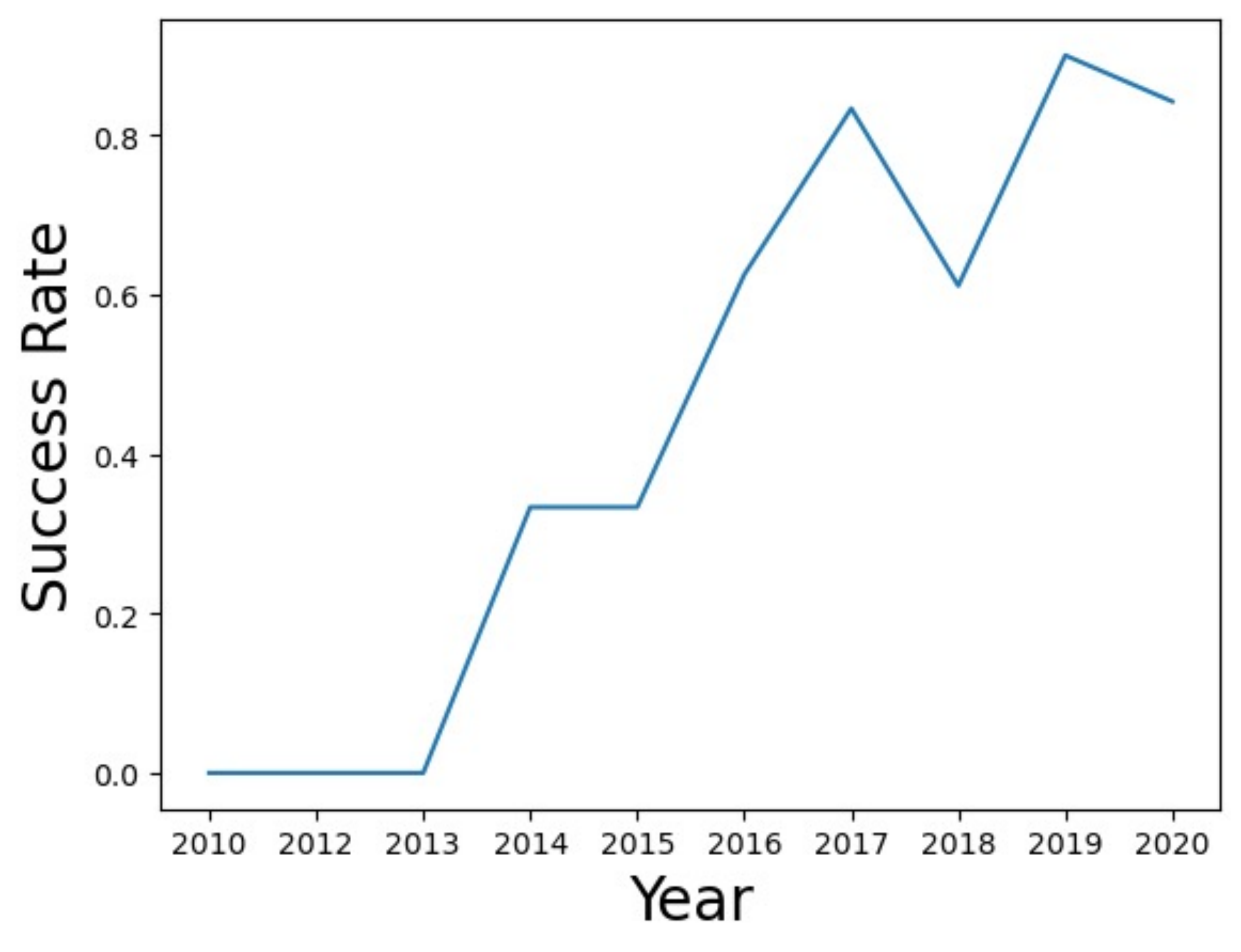
Payload vs. Orbit Type

- Heavy payload (>13000 kg) are sent to VLEO and are successful.
- Lighter payloads are mostly sent to SSO, GTO, ISS and LEO.



Launch Success Yearly Trend

- Success rate is significantly increasing since 2014.
- Success rate is about 80% in recent years



All Launch Site Names

- Unique launch site names are

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

- They are spatially located at mainly 3 geo-locations.

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
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- These are just 5 records, but other sites that begins with `CCA` can be obtained.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Total Payload by NASA (CRS)
45596

- Total payload mass by customer = NASA (CRS)

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Average Payload carried by booster version F9 v1.1
2928.4

- Average of the payload mass when booster_version = booster version F9 v1.1

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

First successful landing Outcome on ground pad
01-05-2017

- First successful landing for successful landing on ground pad

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

- Names of the boosters that have successfully landed on drone ship with payload mass in between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Mission_Outcome	Count(Mission_Outcome)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- Number of time the mission outcome occurs

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- List of boosters that have carried the maximum payload mass

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Month	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- When flight landing failed on the drone ship with booster version and launch site names.

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

Rank	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
1	08-04-2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2	06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
3	18-07-2016	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
4	14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
5	14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
6	19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
7	14-08-2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
8	07-09-2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990	LEO	U.S. Air Force	Success	Success (ground pad)
9	11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

- Ranking in the descending order of all flights between 2010-06-04 and 2017-03-20 with success on ground pad.

Section 3

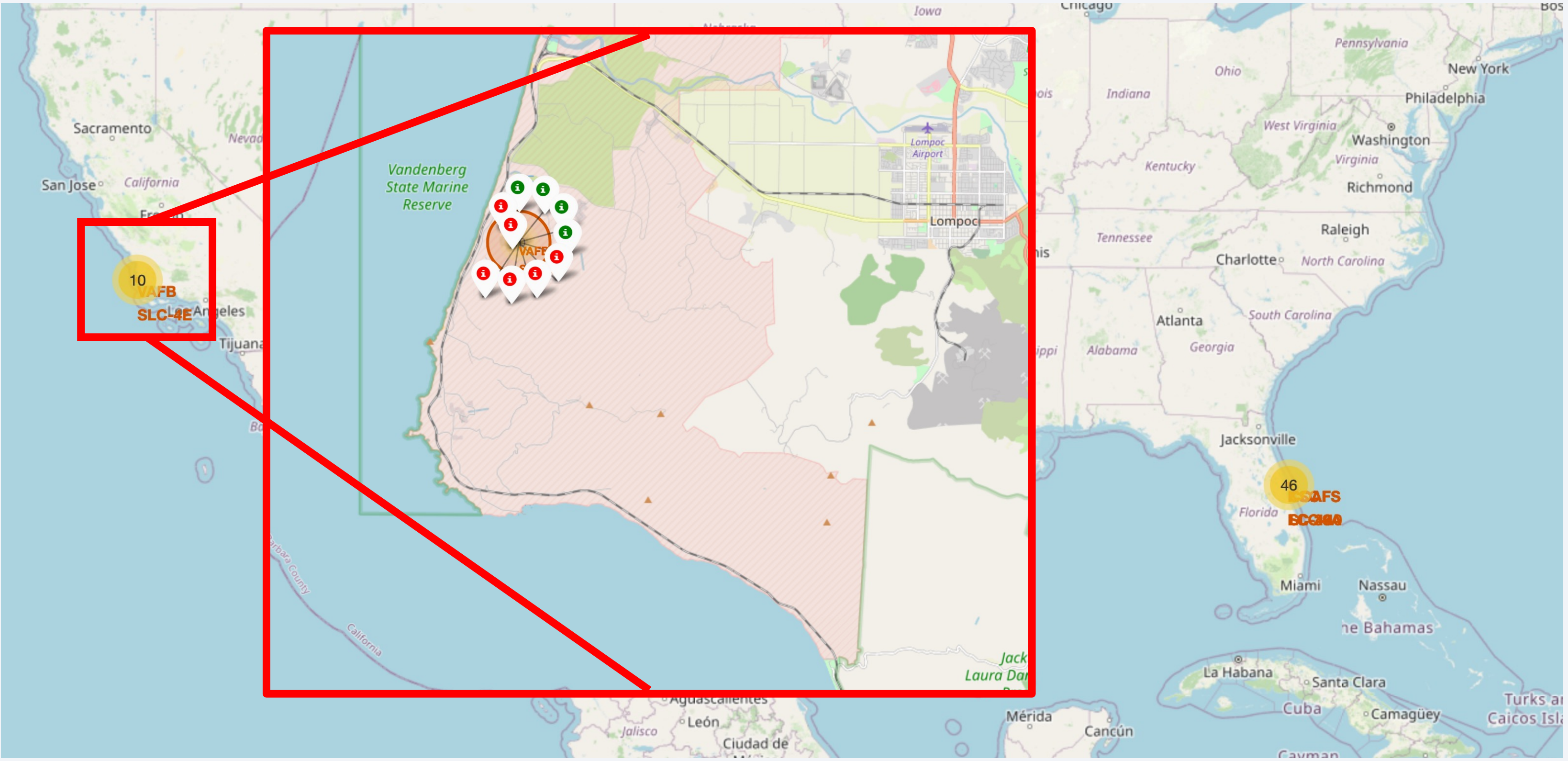
Launch Sites Proximities Analysis

Launching sites on map

- Launch sites are shown on map in red color texts.

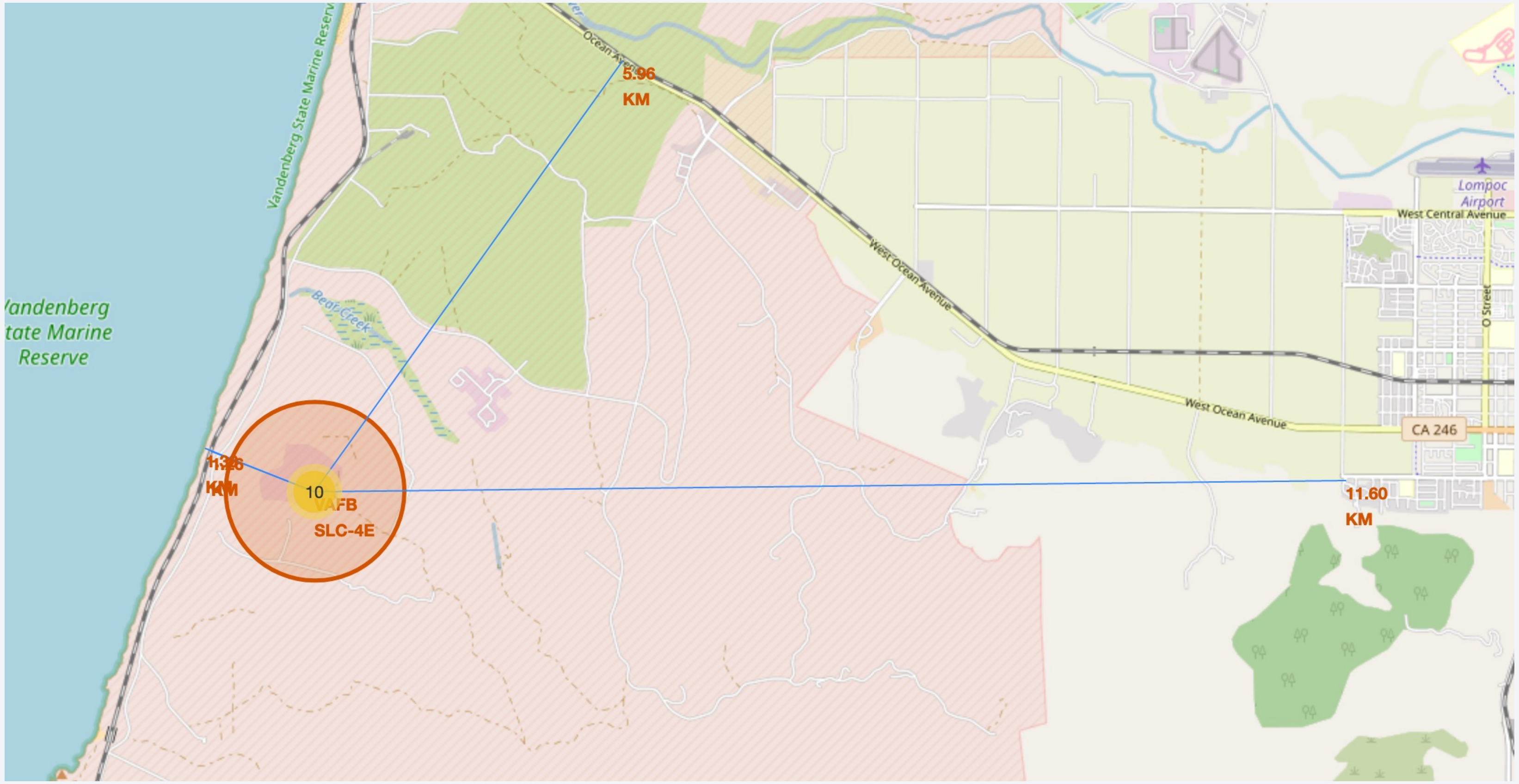
Launching sites with launch outcome markers

- Number of launch attempts are shown in yellow circle.
- Markers are shown in red and green color representing the failed and success outcome, respectively.



Launch site with distance to important elements

- Distance from the important elements are shown in the map with their location
- Blue line represent minimum distance from the important features.

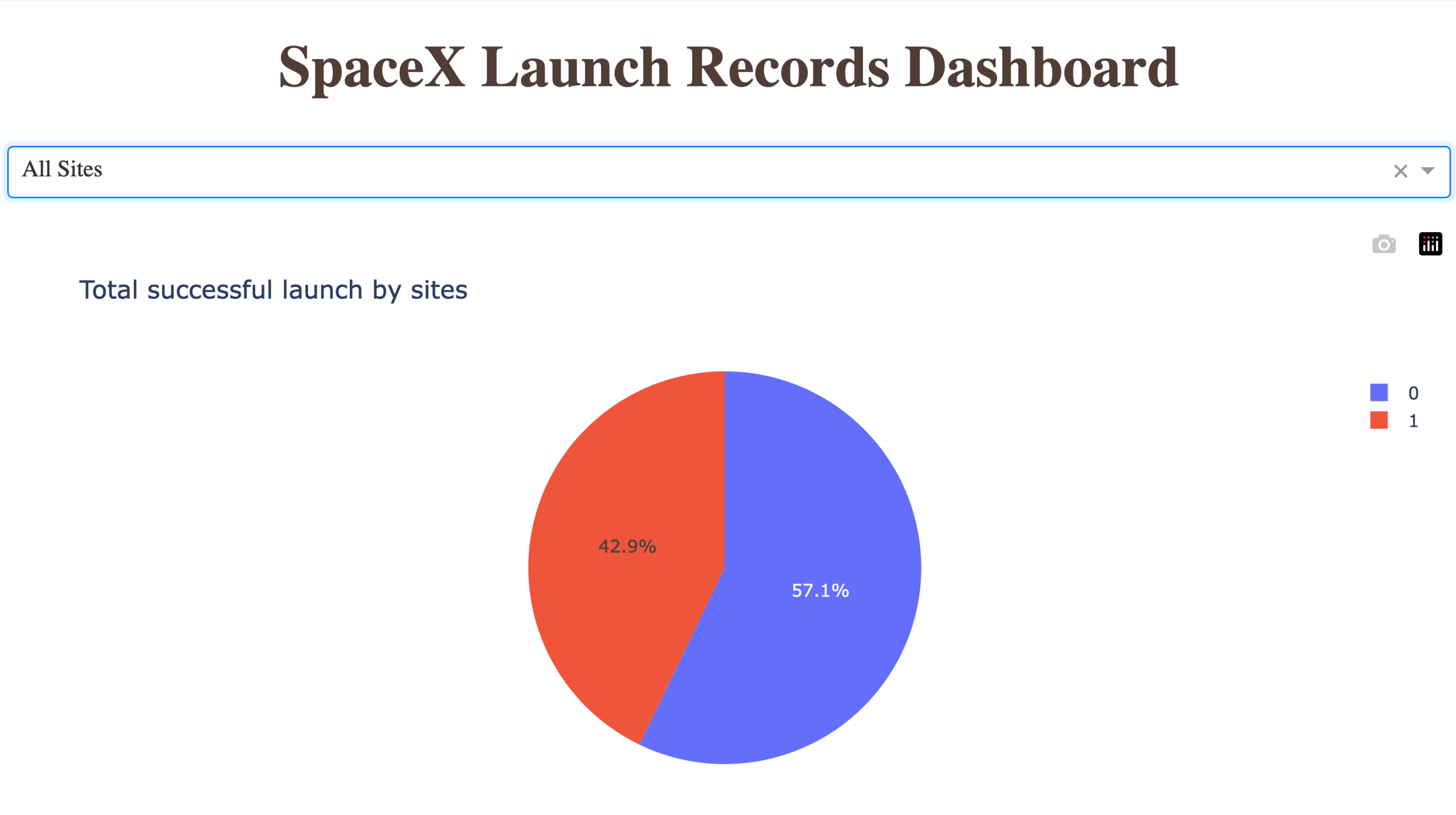


Section 4

Build a Dashboard with Plotly Dash

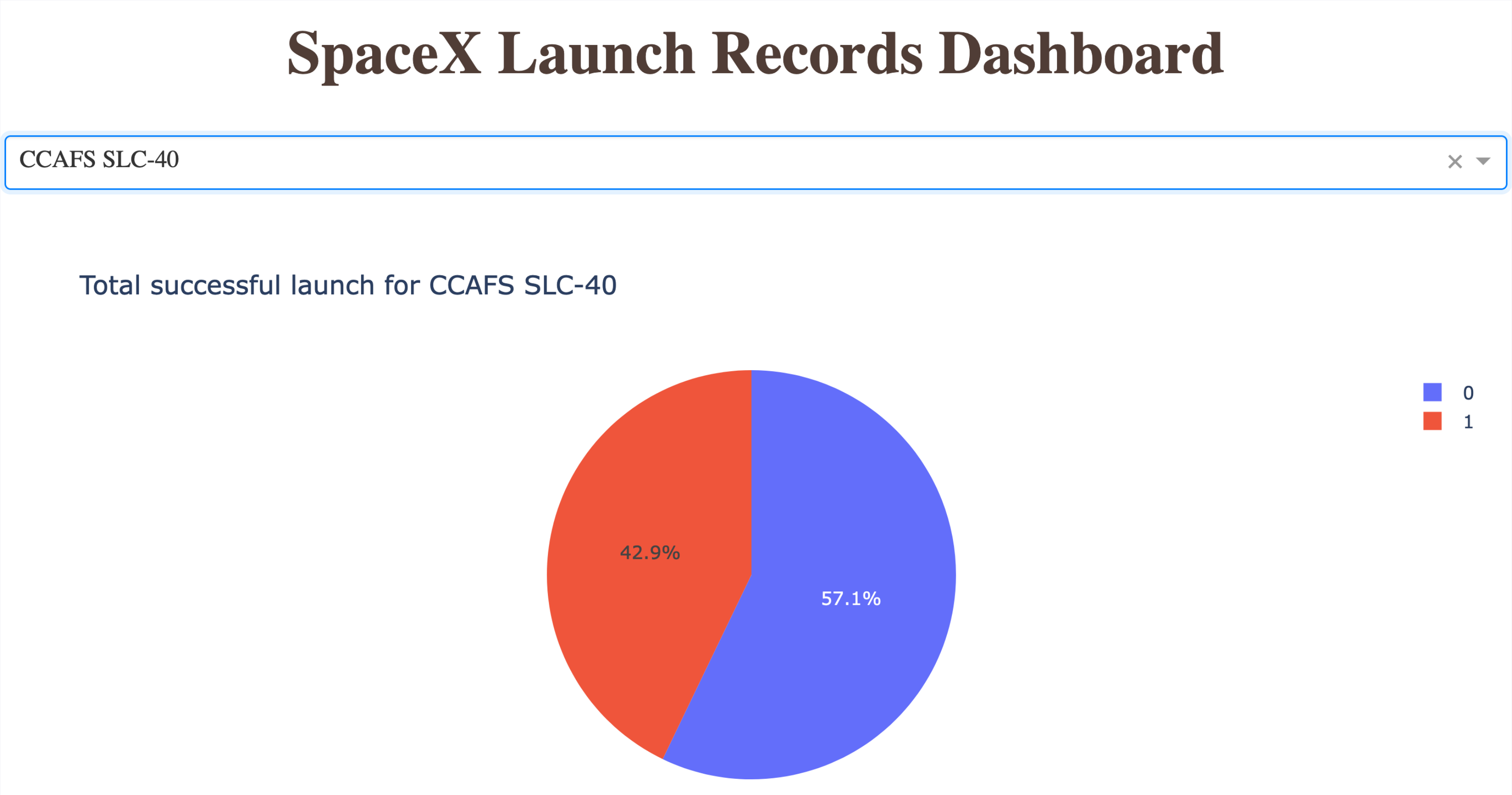
Successful launch by all sites

- Successful launch fraction is shown in red color in the pie chart.



Site with highest success rate

- Launch from CCAFS SLC-40 has the highest success rate of 42.9 %



Payload mass vs Booster version

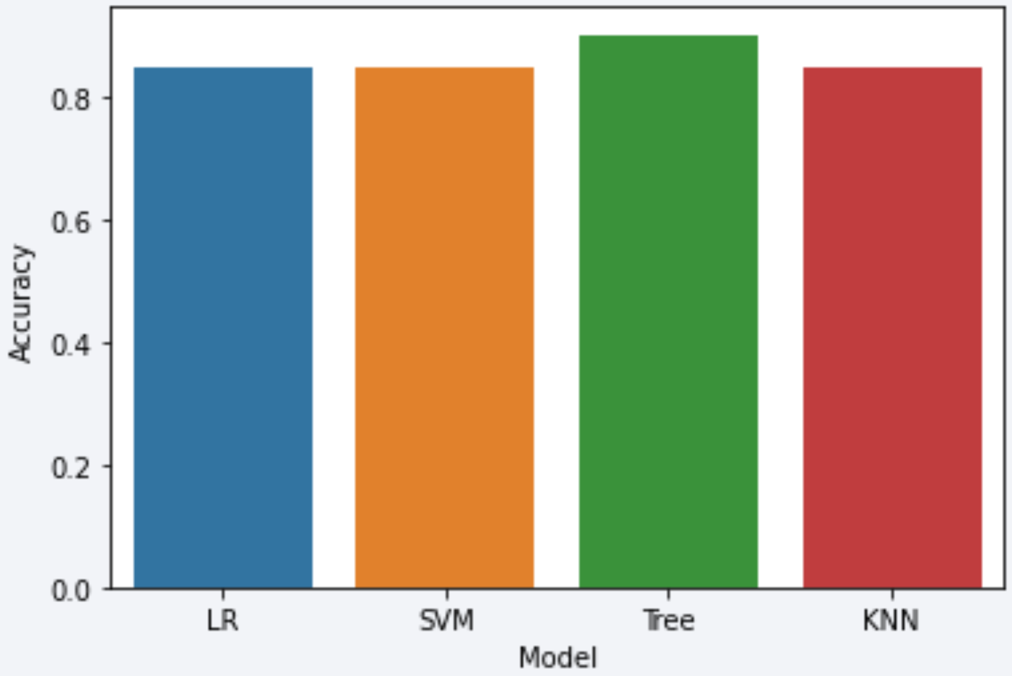
- For payload mass between 2000 and 7000 kg, Booster version with category 'FT' has highest success rate.

Section 5

Predictive Analysis (Classification)

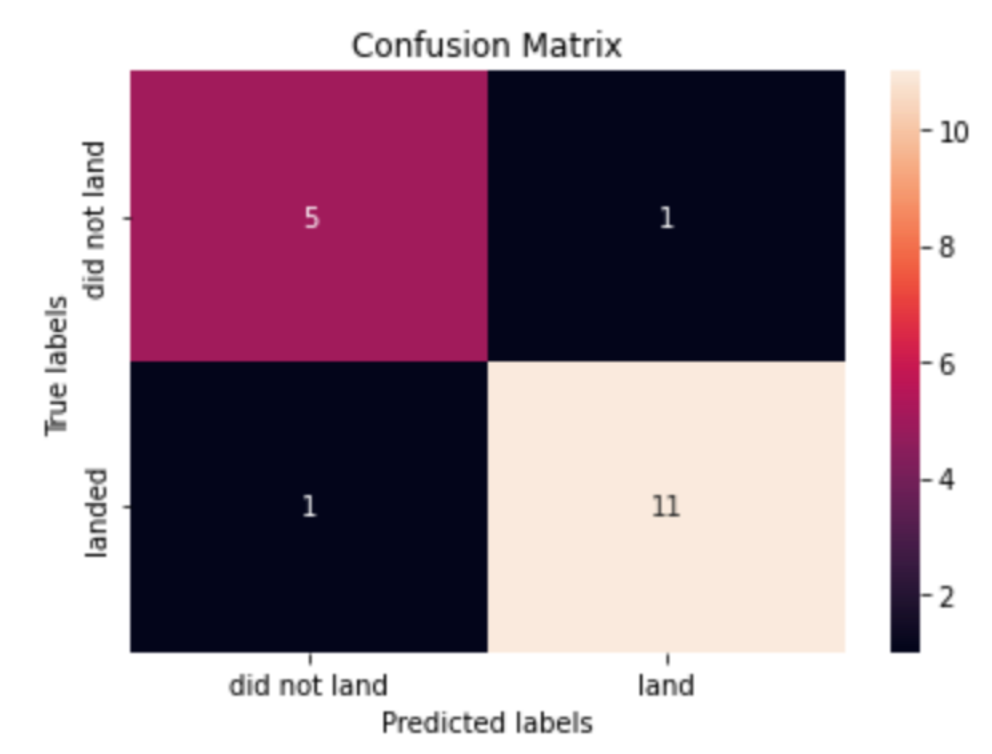
Classification Accuracy

- Decision Tree has the best prediction accuracy for this classification problem



Confusion Matrix

- Confusion matrix for the decision tree model.
- It shows that the number of cases for both false positive and false negative are very less as compared to true positive and true negative cases.



Conclusions

- Payload mass, orbit, launch site, booster version are key parameters that requires attention in the future launching
- Decision tree model predict the outcome with significantly good accuracy

Thank you!