IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tan Kai Yang
30 Dec 2024

Outline

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

Flow:

1. Request data from the url
2. Decode the response content with .json()
3. Turn the json into panda dataframe with .json_normalize()
4. Request information from SpaceX API with customized function
5. Create data frame with intended data and information
6. Filter the data frame to include only Falcon 9 launches
7. Replace missing value with the mean

Github:
https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/Data_Collection-API.ipynb

Data Collection - Scraping

Flow:

1. Request HTML data from the url
2. Create a BeautifulSoup object with the response
3. Locate the table from the BeautifulSoup
4. Turn the table into panda dataframe

Github:
https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/Data_Collection-Web_Scracping.ipynb

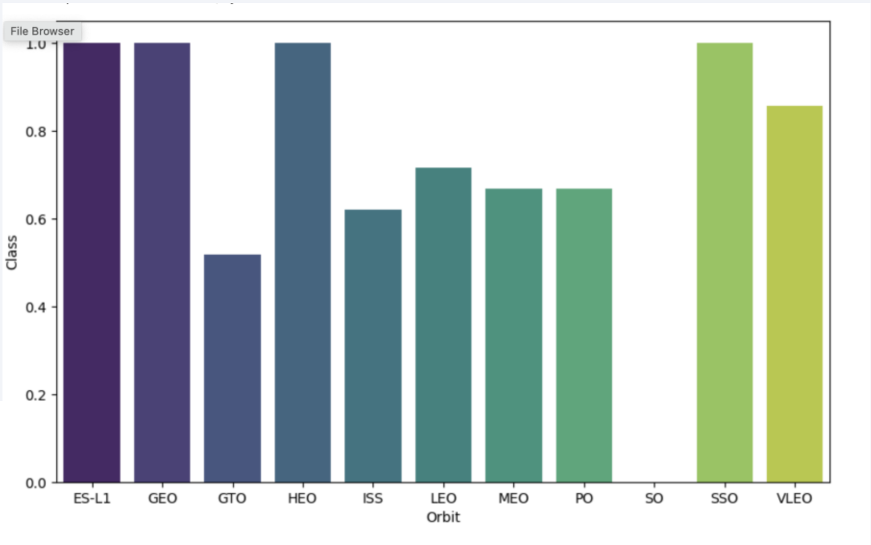
Data Wrangling

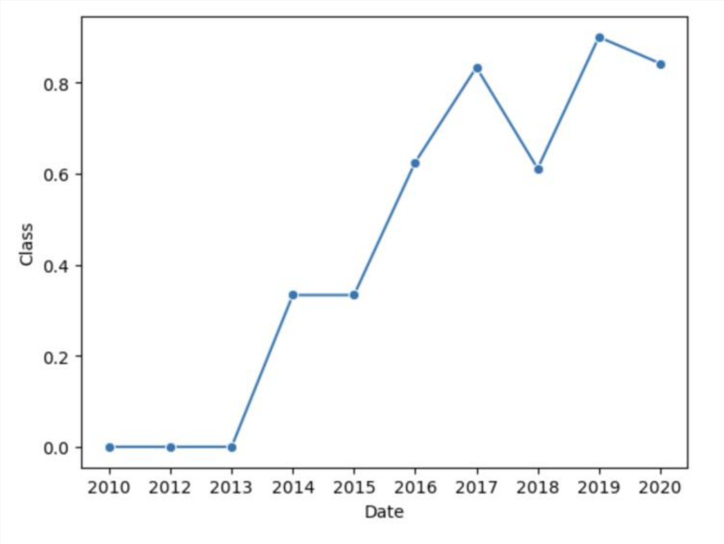
- Read data from csv file
- Check the existence of null
- Check the landing outcomes
- Assign 1 or 0 to good or bad outcomes and rename it as Class
- Check the mean of the Class to know the success rate

- https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/Data_wrangling.ipynb

EDA with Data Visualization

- Look at data of different orbits, launching_site, dates, payload mass. And analyze their relationship with success class.
- https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/EDA-Data_Visualization.ipynb





EDA with SQL

- Read SPACEX table from csv and turn it into sql
- Explore data:
  - Check the distinct launching site
  - Know the total payload of NASA
  - Display average payload mass carried by booster version F9 v1.1
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass.
  - List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- . https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/EDA-SQL.ipynb

Build an Interactive Map with Folium

- Added markers to the launching sites
- Mark them as red or green based on the success class
- Also clustered them for better visualization
- https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/Folium.ipynb

Build a Dashboard with Plotly Dash

- Showcasing success classes in pie chart and scatter chart
- The interactions let users choose the launching sites and payload mass range
- These interactions let users know the success rates of wanted launching sites and payload mass

Predictive Analysis (Classification)

- Transform the data with standard scaler
- Split data set into training and testing sets
- Used logistic regression, SVM, decision tree, and k-nearest neighbors for the gridsearch machine learning model
- Set different parameters for each of these models to perform gridsearch
- Output their confusion matrix and test with testing data set
- https://github.com/kaiyang5029/IBM-Data-Science-Assignment/blob/main/Capstone%20Project/Machine_Learning_Prediction.ipynb

Results

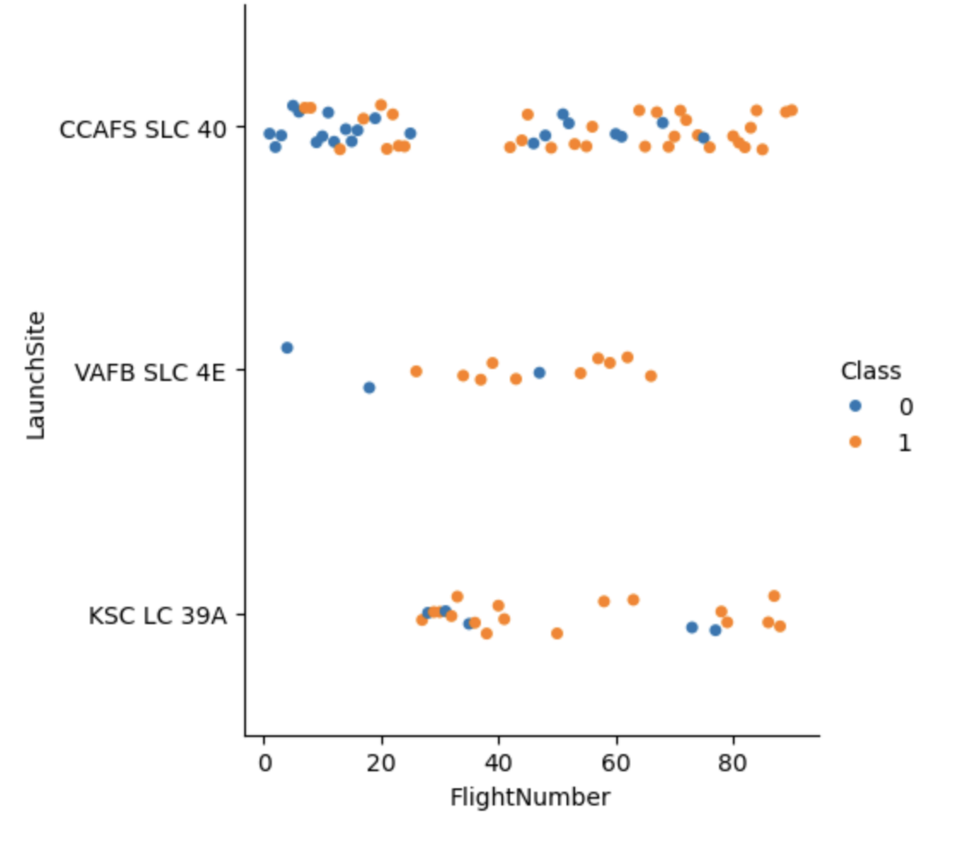
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

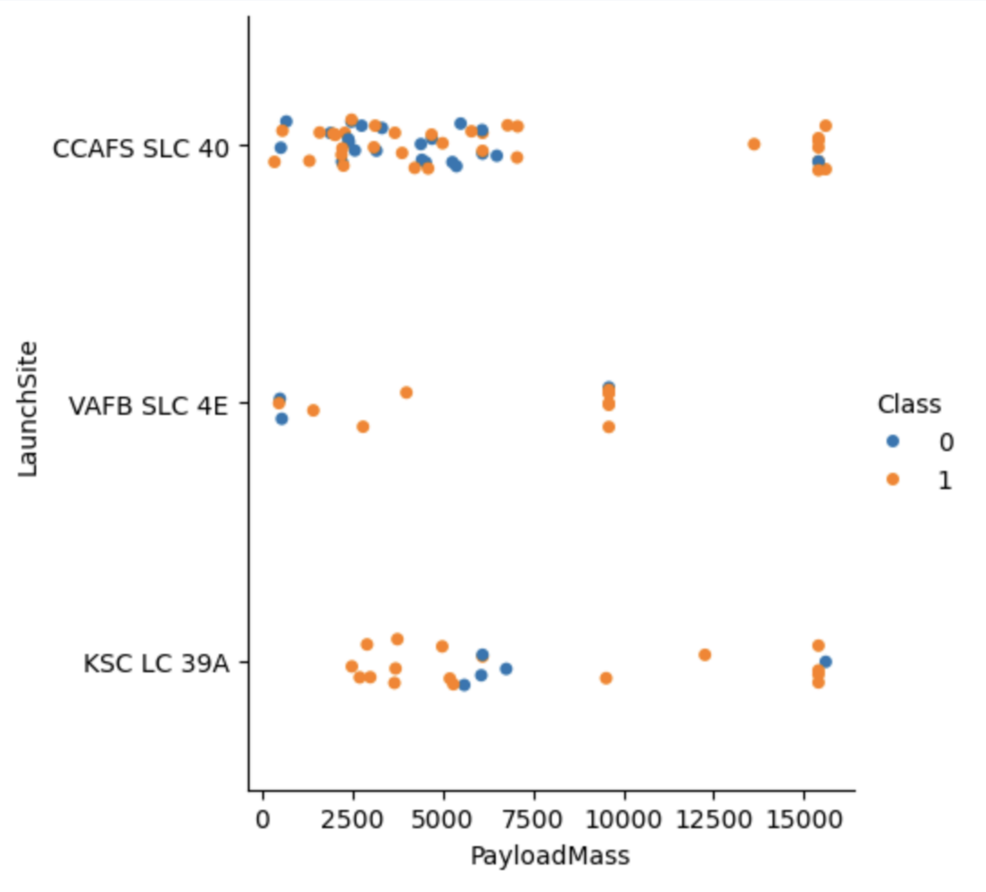
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site.
- The success rates are lower for the older numbers. After a certain flight number, they are mostly successes for each launching sites. The VAFB site reaches more successes sooner.



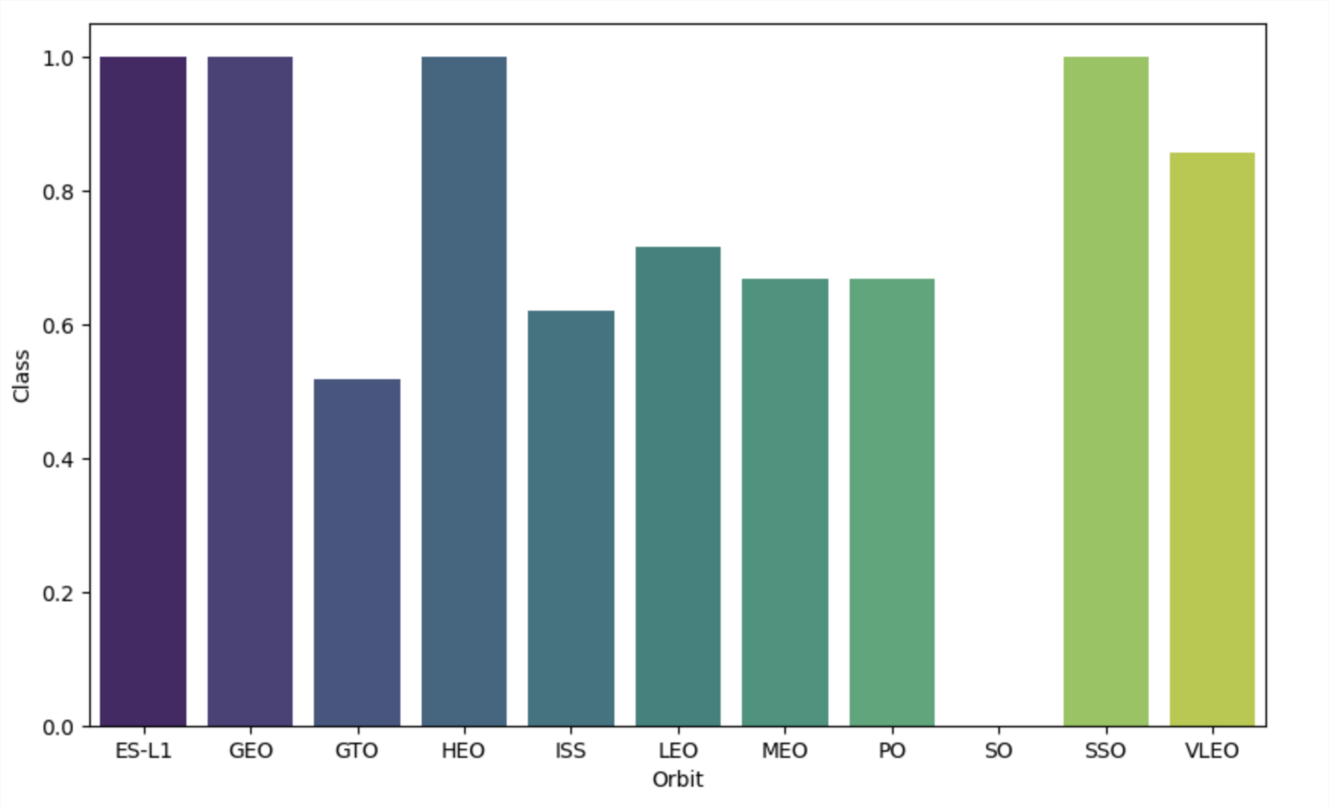
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site
- There are mixed successes for different payload mass. The VAFB site only takes small to middle range of payload mass.



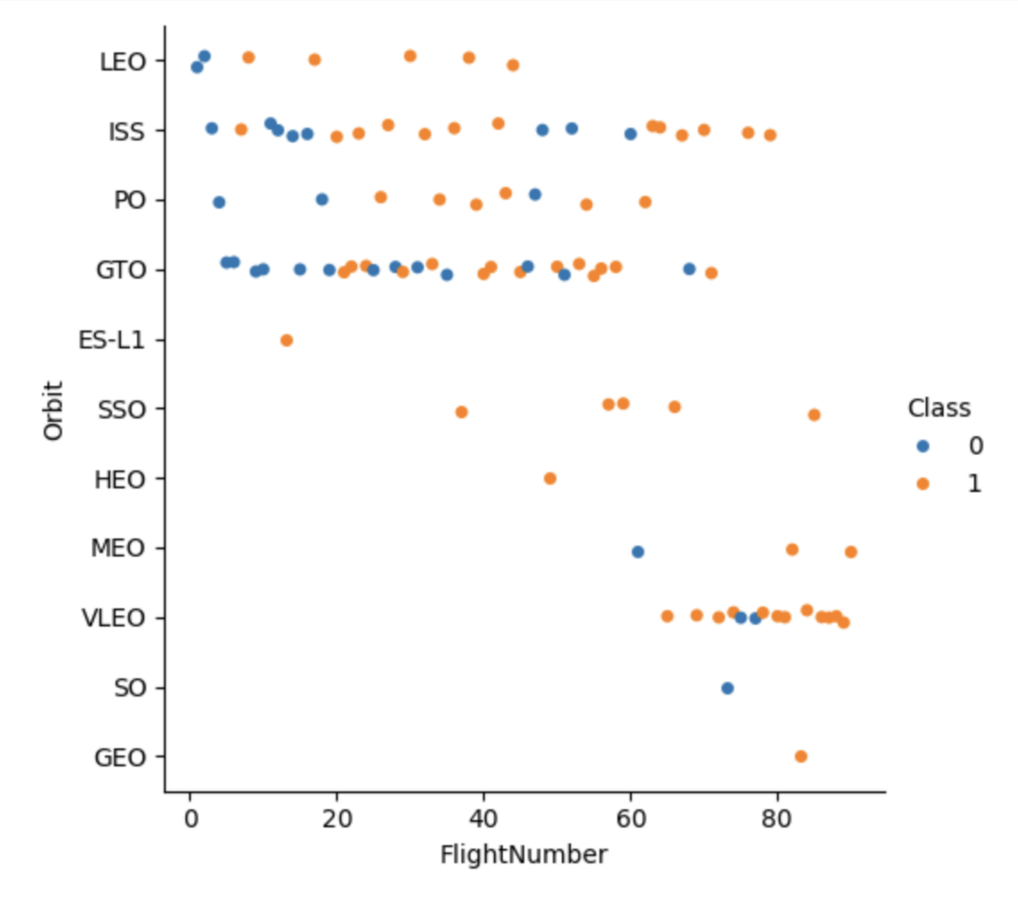
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type
- SO orbit has the lowest success rate, while ES-L1,GEO, HEO, and SSO have the highest success rate



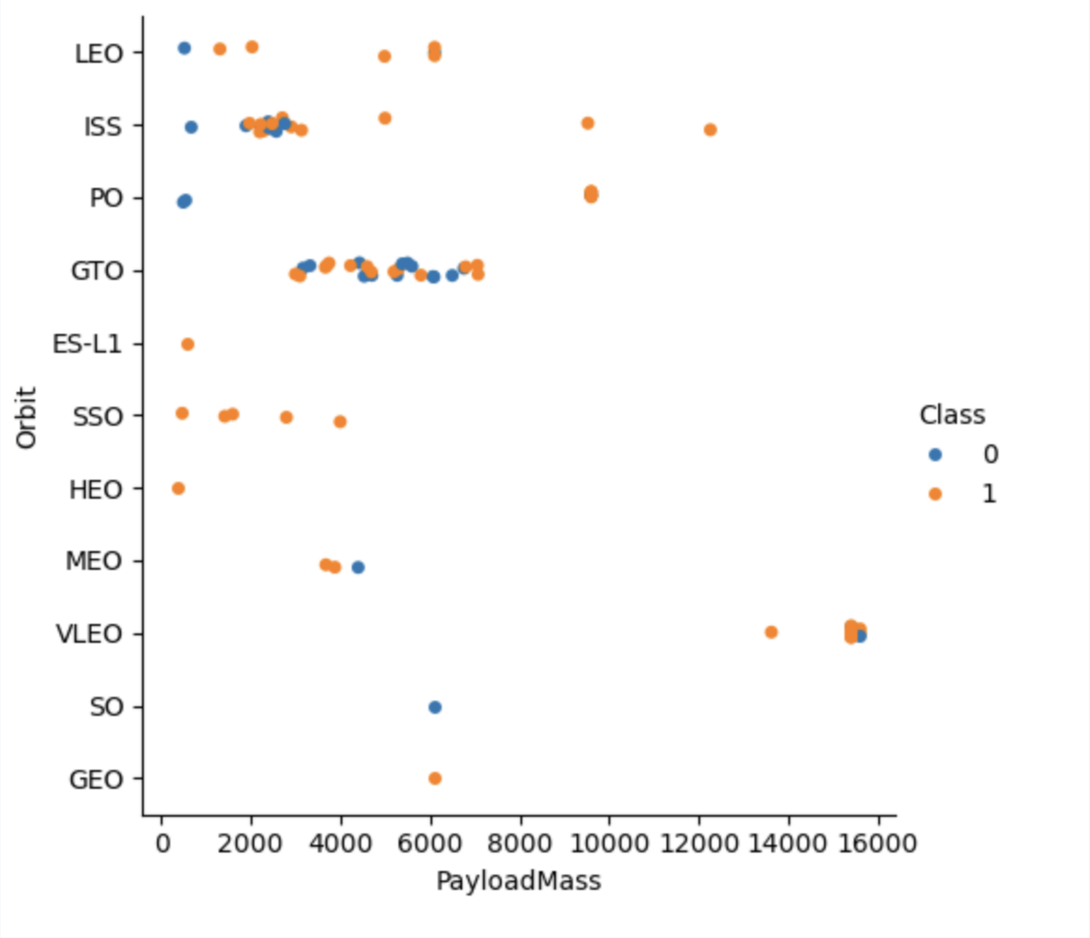
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type
- LEO orbit seems to be becoming more successful under increasing flight numbers, while GTO orbit has mixed results



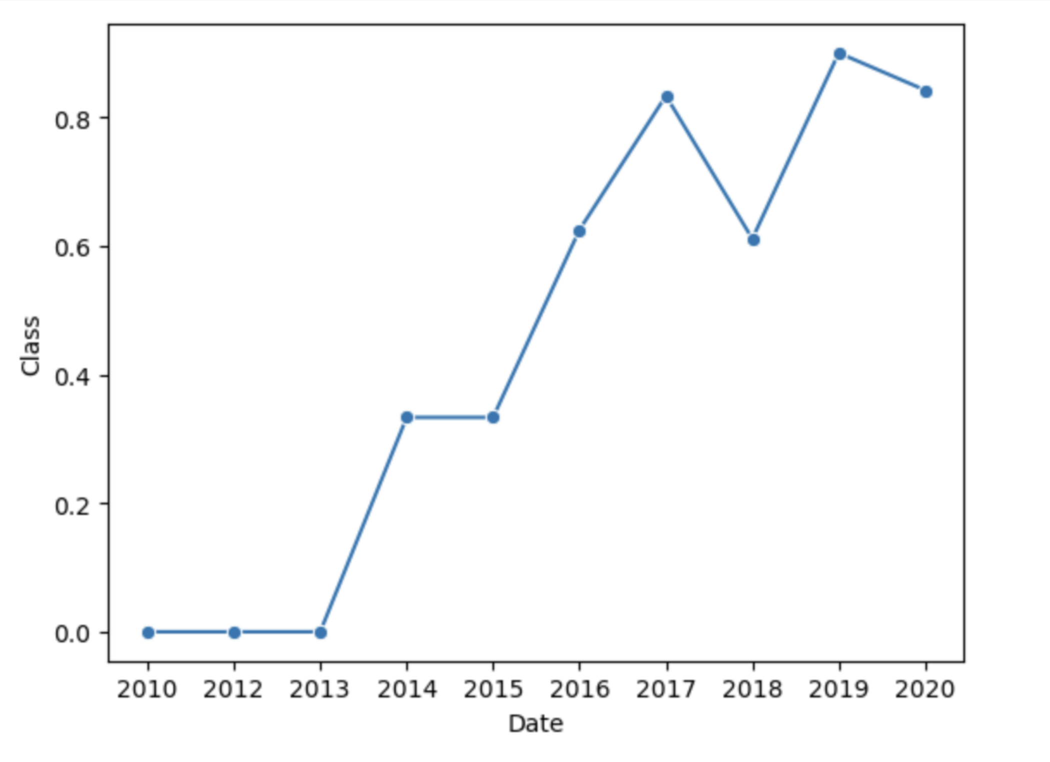
Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type
- For LEO, ISS, and PO, the success rates are higher at bigger mass, while GTO and VLEO give mixed results



Launch Success Yearly Trend

- Line chart of yearly average success rate
- Success rate is increasing over the years



All Launch Site Names

- There are 4 distinct launching sites, as shown at the right table.

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

Launch Site Names Begin with 'CCA'

- The first 5 records where launch sites begin with `CCA`
- The first five all gave successful outcomes

```
%sql SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' limit 5
```

```
* sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

- The total payload mass by NASA is 45596 kg

```
%sql SELECT sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where Customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 340.4kg

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version like 'F9 v1.0%'
```

```
 * sqlite:///my_data1.db
Done.
```

avg(PAYLOAD_MASS__KG_)
340.4

First Successful Ground Landing Date

- The first successful landing outcome on ground pad is on 2015 December 22

```
%sql select min(Date) from SPACEXTABLE where Landing_Outcome like 'Success%'
```

```
* sqlite:///my_data1.db
Done.
```

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are all F9 FT boosters

```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome like 'Success (drone ship)' and PAYLOAD_MASS__
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- A total of 61 successes, with 10 failures

Outcome_Category	Total
None	30
Failure	10
Success	61

Boosters Carried Maximum Payload

- Boosters which have carried the maximum payload mass are F9 B5 boosters

```
%%sql
SELECT Booster_Version, PAYLOAD_MASS__KG_
FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (
    SELECT MAX(PAYLOAD_MASS__KG_)
    FROM SPACEXTABLE
);
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- The failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015 are listed on the right.
- A total of 2 failure, all with booster version F9 v1.1, at launching site CCAFS LC-40

Month_Name	Booster_Version	Launch_Site	Landing_Outcome
January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- The count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order is on the right
- Without landing attempt has the most outcome, while drone ship ranks the second

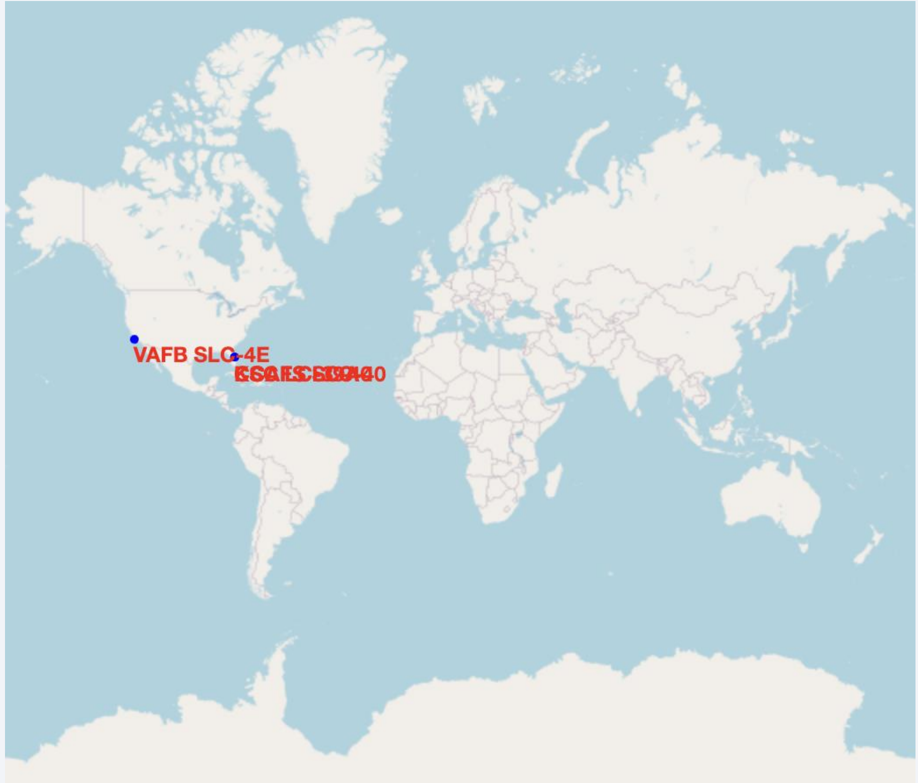
Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

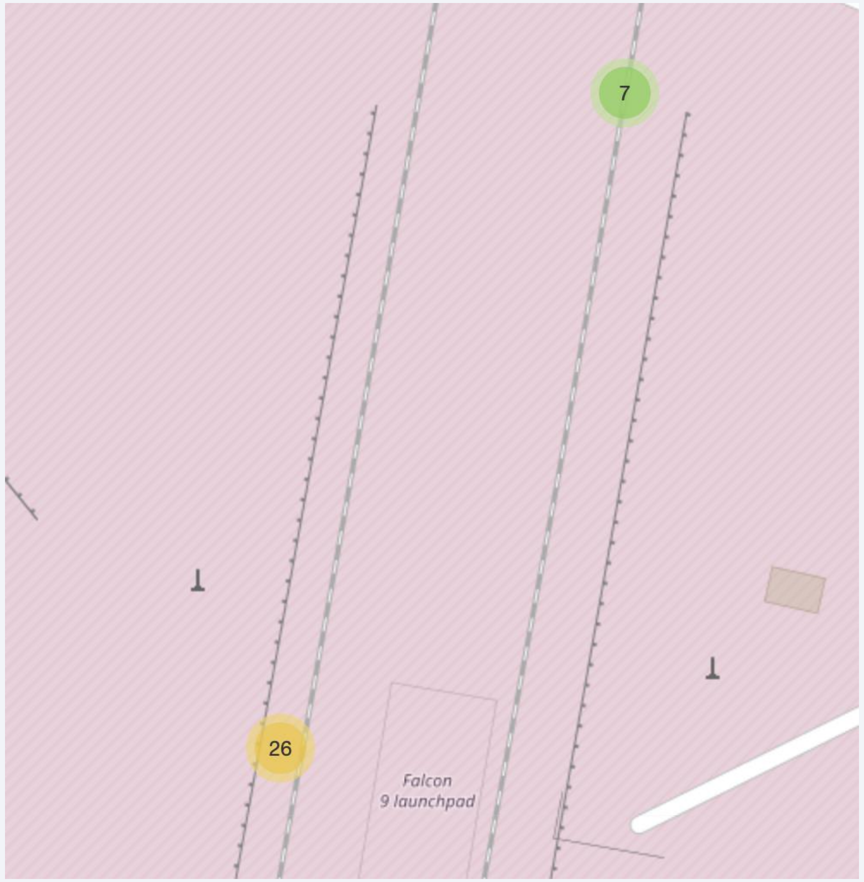
Launching Sites on Global Map

- One launching site at the west coast of the US, the rest on the east coast of the US.



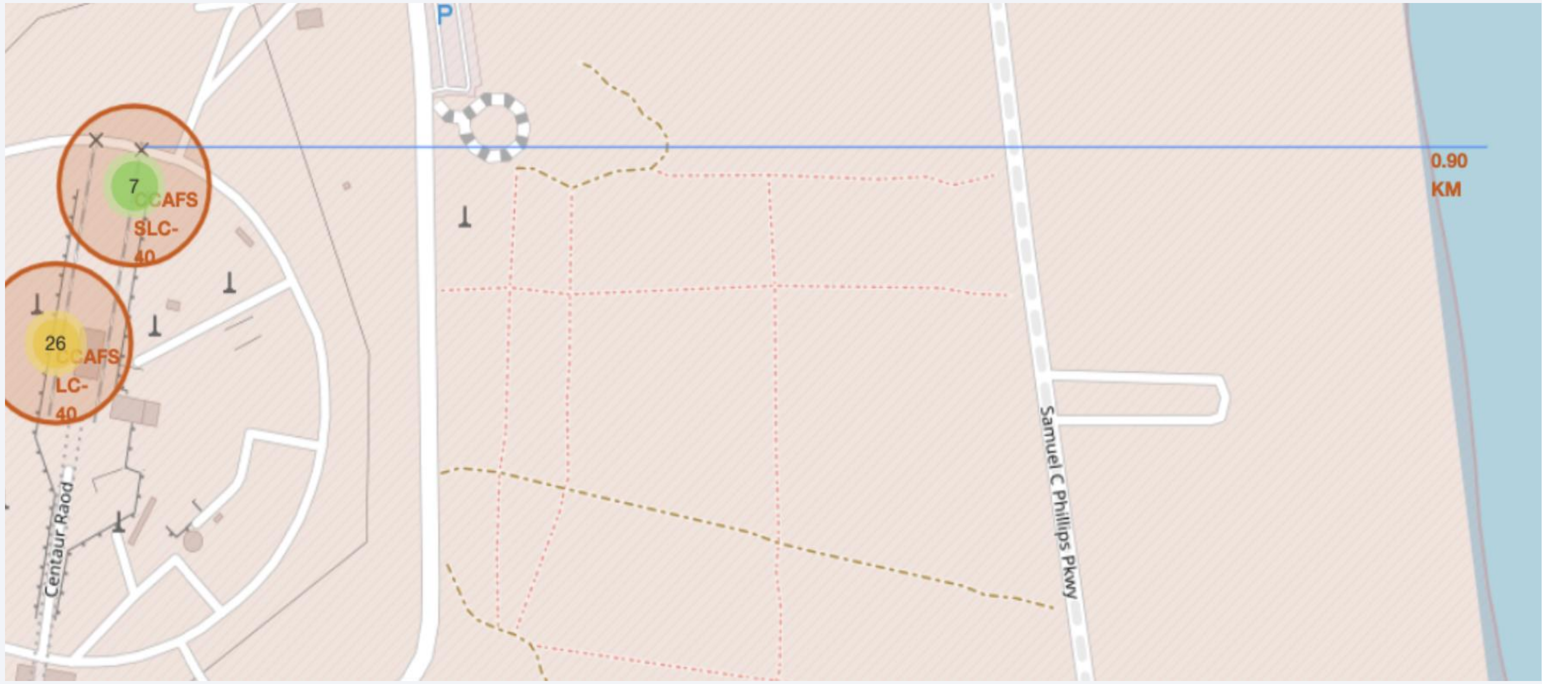
Launching Outcome on Map

- Green implies successful landing, while red implies unsuccessful landing.
- Yellow represents a mixed of both results.



Launch Sites to Proximities

- The CCAFS launch site is 0.9km away from coastline

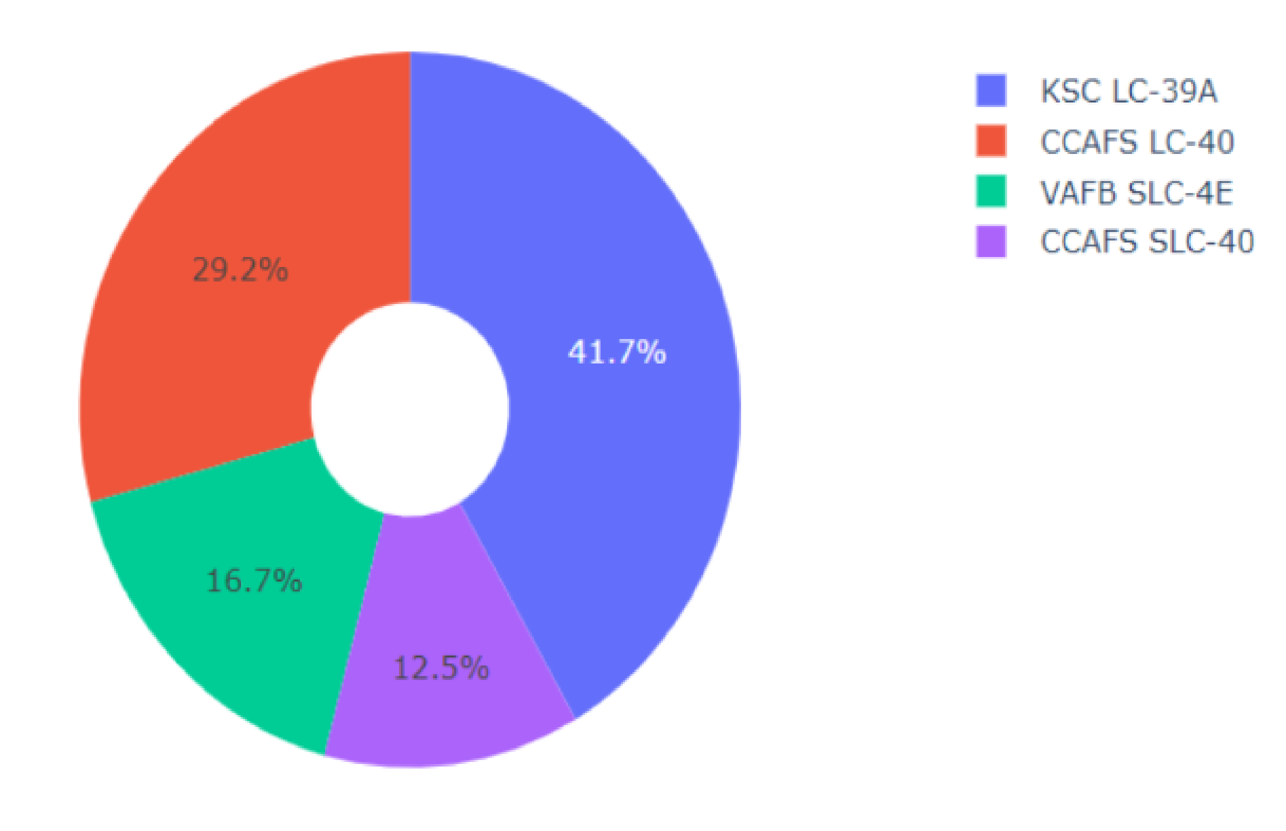


Section 4

Build a Dashboard with Plotly Dash

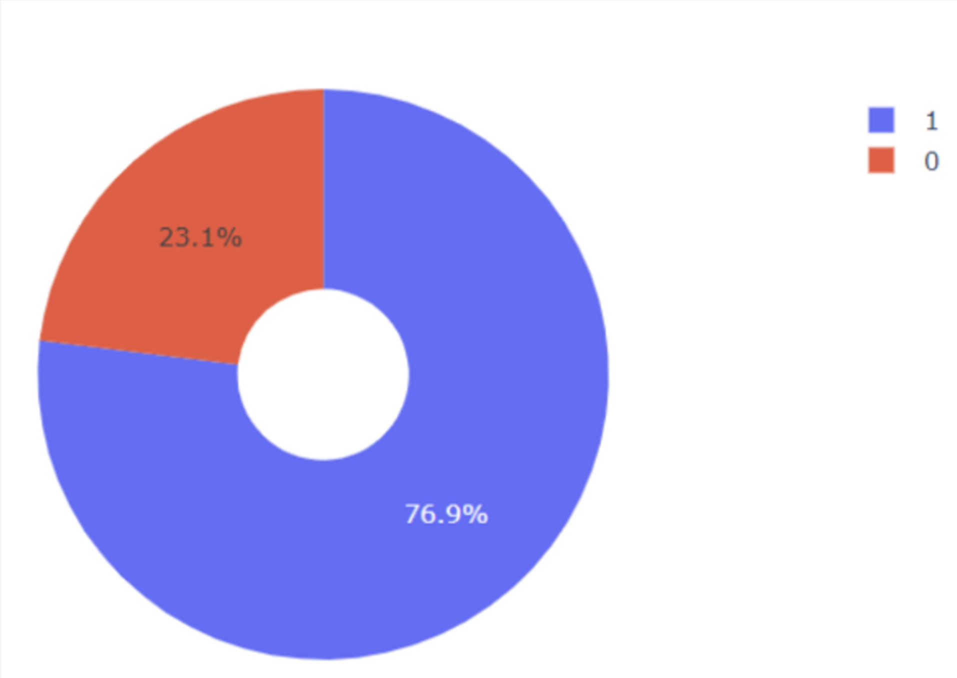
Launch Success at all Sites

- Pie chart on the right showed the success rate proportion of different launch sites
- KSC has the most success rates, while the VAFB has the lowest.



Success Rate of the most Successful Launch Site

- Pie chart on the right showed the success rate of the KSC launching site.
- It successfully landed 76.9% of the times.

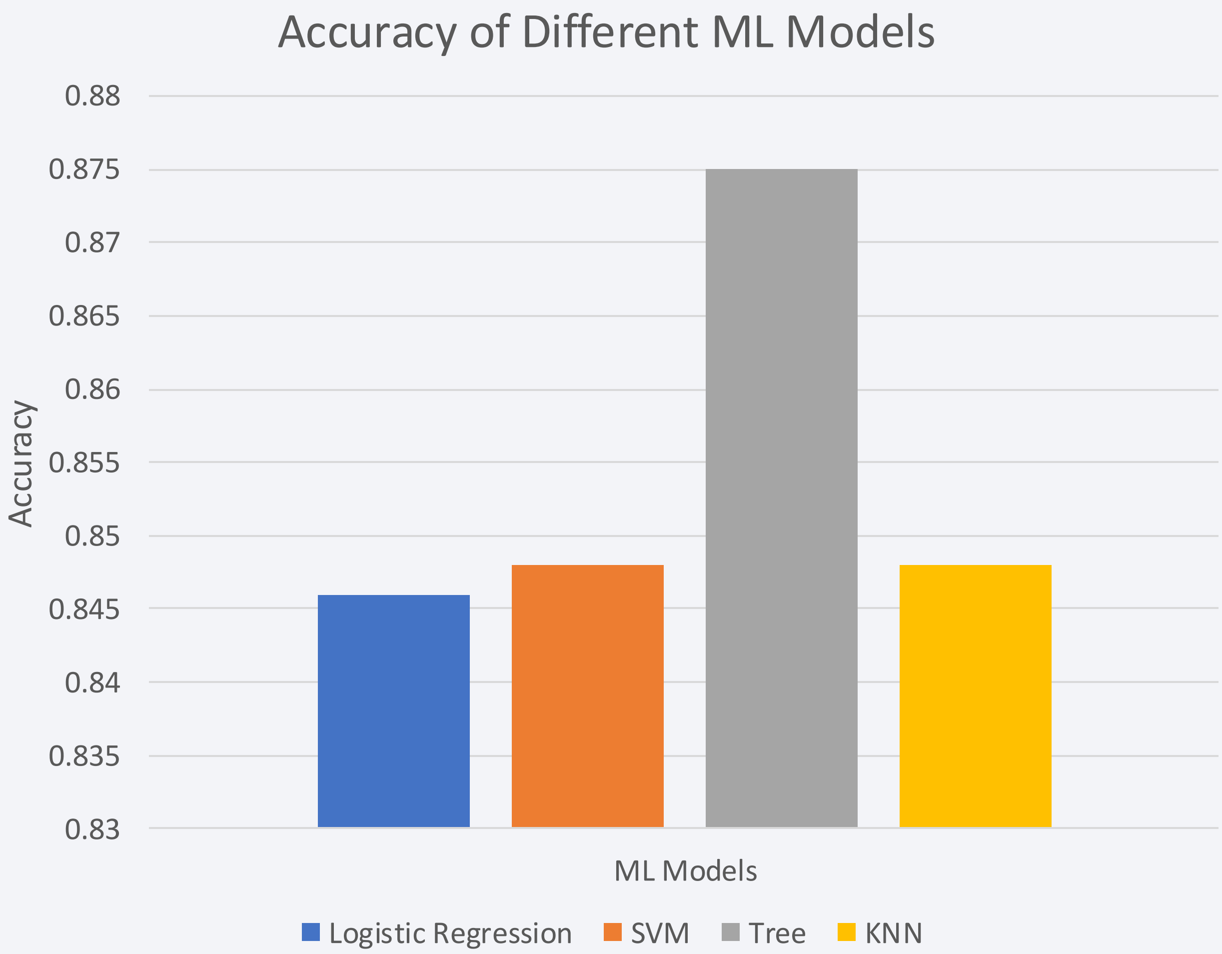


Section 5

Predictive Analysis (Classification)

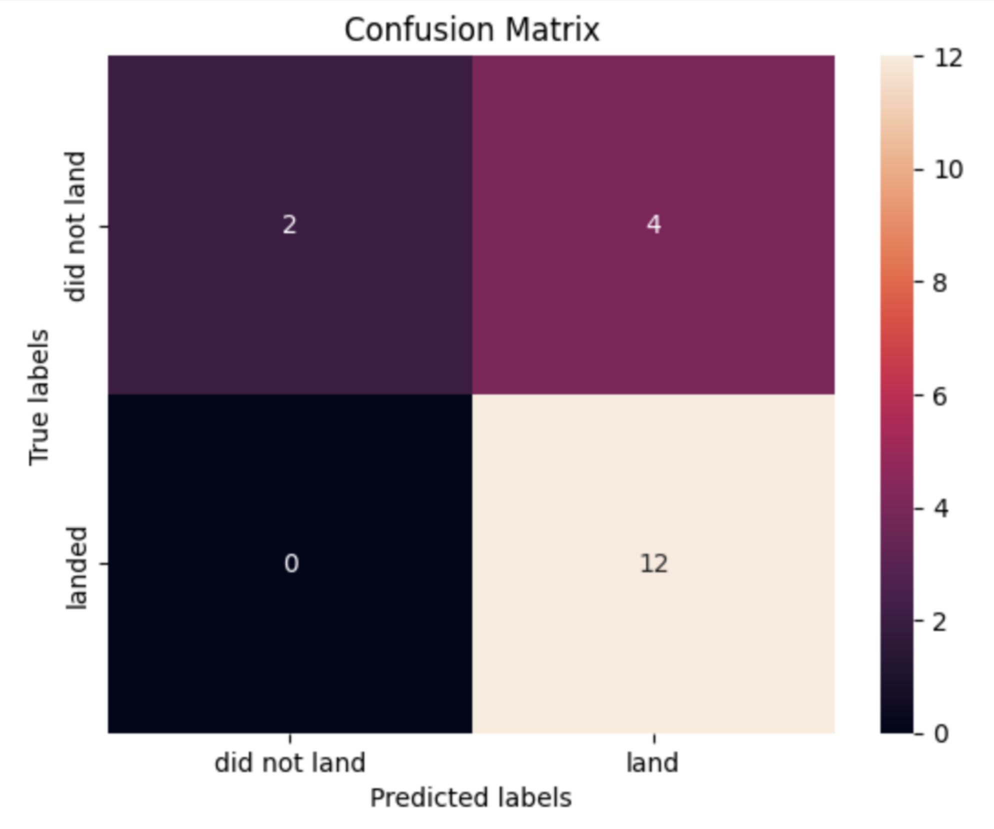
Classification Accuracy

- Decision Tree has the highest accuracy of 0.875



Confusion Matrix

- It is good at predicting landed, but not the unsuccessful landing



Conclusions

- The success rate is increasing under years
- SO orbit has the lowest success rate, while ES-L1,GEO, HEO, and SSO have the highest success rate
- The most successful launch site is KSC
- Decision tree is the best machine learning model for prediction, but it has issue on predicting the negative outcome

Thank you!