IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Het Patel
20^{th} August, 2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection Using API
  - Data Collection Using Webscraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics and Dashboard
  - Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis Result
  - Interactive Visual Analytic Result
  - Predictive Analysis Result

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. The Goal of Project is To predict if the Falcon 9 first stage will land successfully or not.

- Problems you want to find answers
  - To predict if the Falcon 9 first stage will land successfully or not.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX API and web scraping.
- Perform data wrangling
  - Apply One hot encoding to prepare data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Build, tune, evaluate classification models like Logistic Regression, Decision Tree, etc.

Data Collection

- Data collected through SpaceX API.
- Also, Data was collected from Falcon 9 launch records Wikipedia webpage using BeautifulSoup.

Data Collection – SpaceX API

- Request and parse the SpaceX launch data using the GET request.

- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex-data-collection-api.ipynb

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
[6]: static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/dataset
```

We should see that the request was successfull with the 200 status response code

```
[7]: response = requests.get(static_json_url)
```

```
[8]: response.status_code
```

```
[8]: 200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
[9]: # Use json_normalize meethod to convert the json result into a dataframe
     data = pd.json_normalize(response.json())
```

Using the dataframe `data` print the first 5 rows

```
[10]: # Get the head of the dataframe
      data.head()
```

Data Collection - Scraping

- Request the Falcon9 Launch Wiki page from its URL and parse Data From HTML tables.

- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex_Data_collection_webscraping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

[4]:
```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url)
```

Create a `BeautifulSoup` object from the HTML `response`

[5]:
```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response.text, 'html.parser')
```

Print the page title to verify if the `BeautifulSoup` object was created properly

[6]:
```
# Use soup.title attribute
soup.title
```

[6]:
```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

TASK 2: Extract all column/variable names from the HTML table header

Next, we want to collect all relevant column names from the HTML table header

Data Wrangling

- How data were processed

  Calculate the number of launches on each site.

  Calculate the number and occurrence of each orbit.

  Calculate the number and occurrence of mission outcome of the orbits.

  Create a landing outcome label from Outcome column.

- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex-Data_wrangling.ipynb

EDA with Data Visualization

- Scatter Plot to Visualize the relationship between
  - Flight Number and Launch Site
  - Payload and Launch Site
  - Flight Number and Orbit type, etc.
- Bar Plot to Visualize the relationship between
  - Success rate of each orbit type
- Line Plot to Visualize launch success yearly trend
- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex_EDA_Visualization.ipynb

EDA with SQL

- Using SQL query we perform following task
  - Display the names of unique Launch site
  - Display Total Payload mass carried by boosters launched by Nasa
  - List the Date when first successful landing outcome in ground pad was achieved.
  - List the total number of successful and failure mission outcomes, etc.
- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex_EDA_with_SQL.ipynb

Build an Interactive Map with Folium

- Create and add Circle and Marker for each launch site on the site map
- Create and add MarkerCluster to Mark the success/failed launches for each site on the map
- Create and add line object to display the distances between a launch site to its proximities
- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex_Folium.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Build classification models like Logistic Regression, Decision Tree, SVM.
- Split data into train and test set
- Train models using train sets.
- Use GridsearchCV to find best parameters for models.
- Use Accuracy score to find best model
- GitHub URL: https://github.com/hetpatel37/DS_Capstone_Project/blob/main/SpaceX_Machine_Learning_Prediction.ipynb

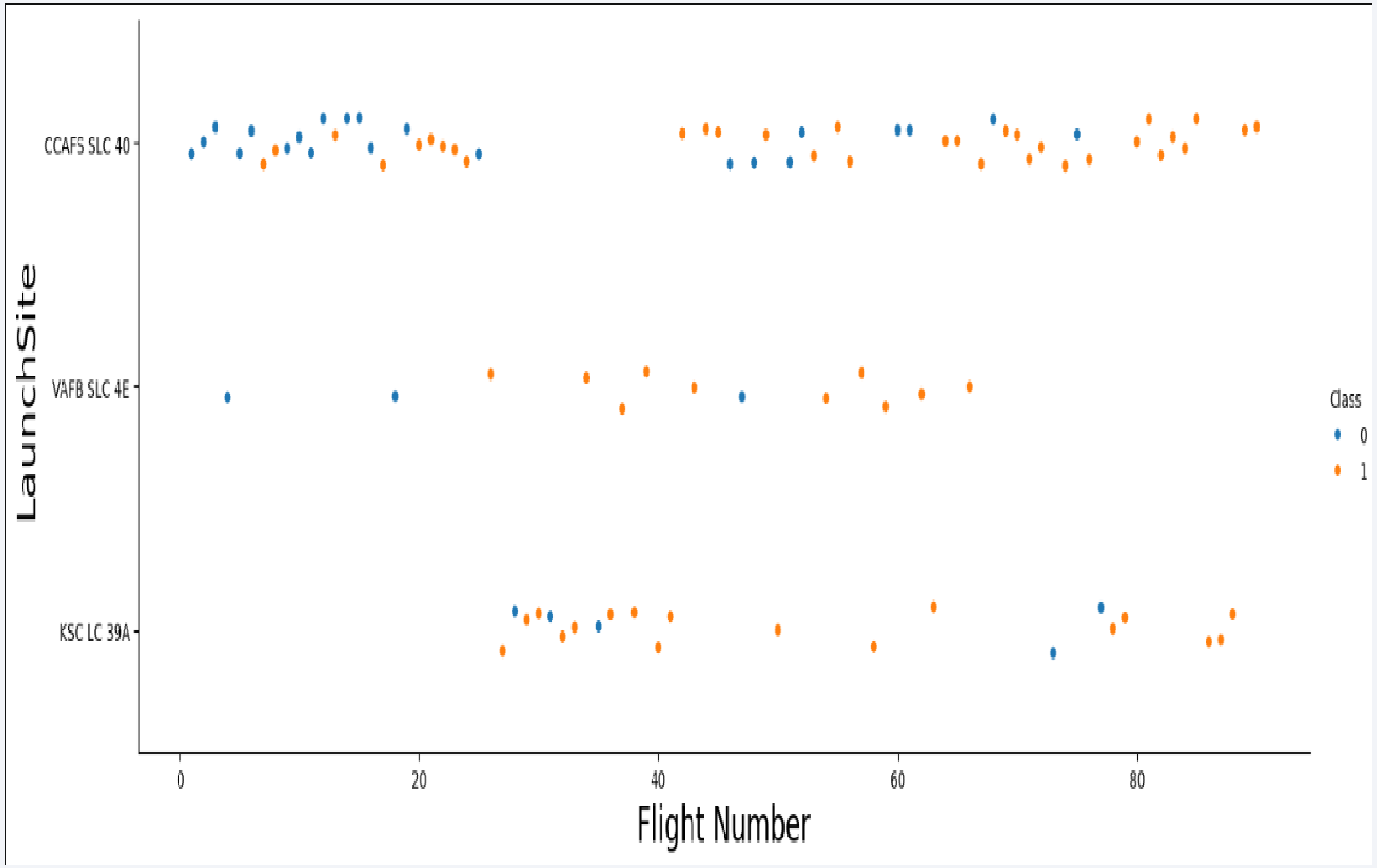
Results

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

Section 2

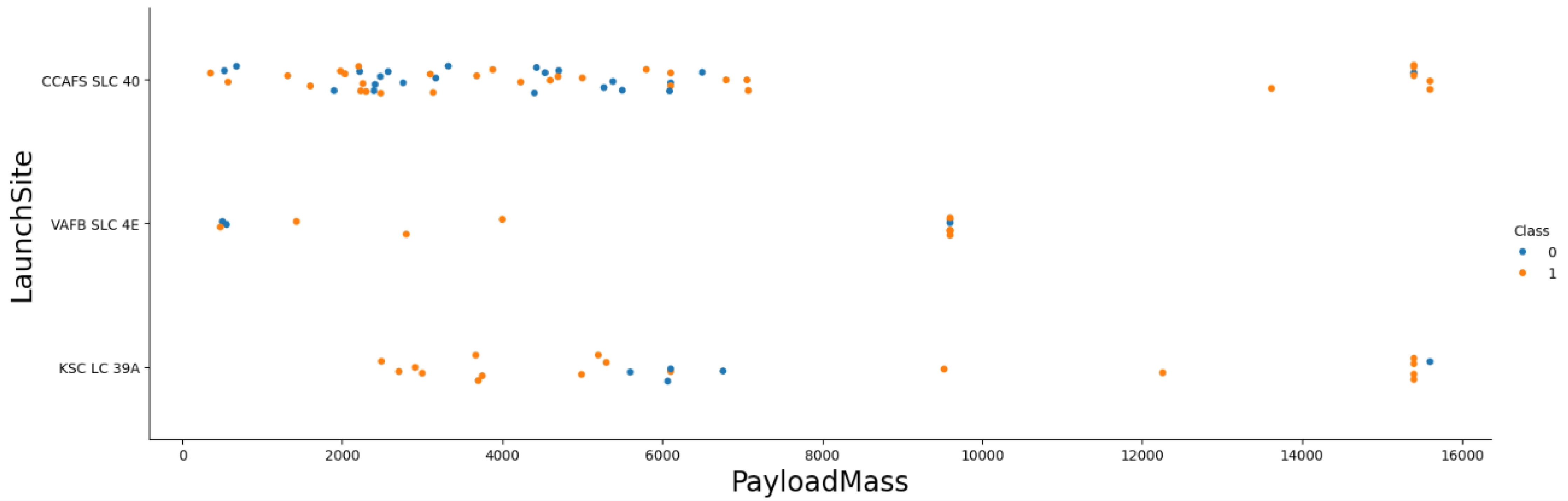
Insights drawn from EDA

Flight Number vs. Launch Site



- Explanations:
- Scatter Plot between Flight Number and Launch Site is shown in figure.
- From graph we observe that, As Flight Number increases, success rate increase.

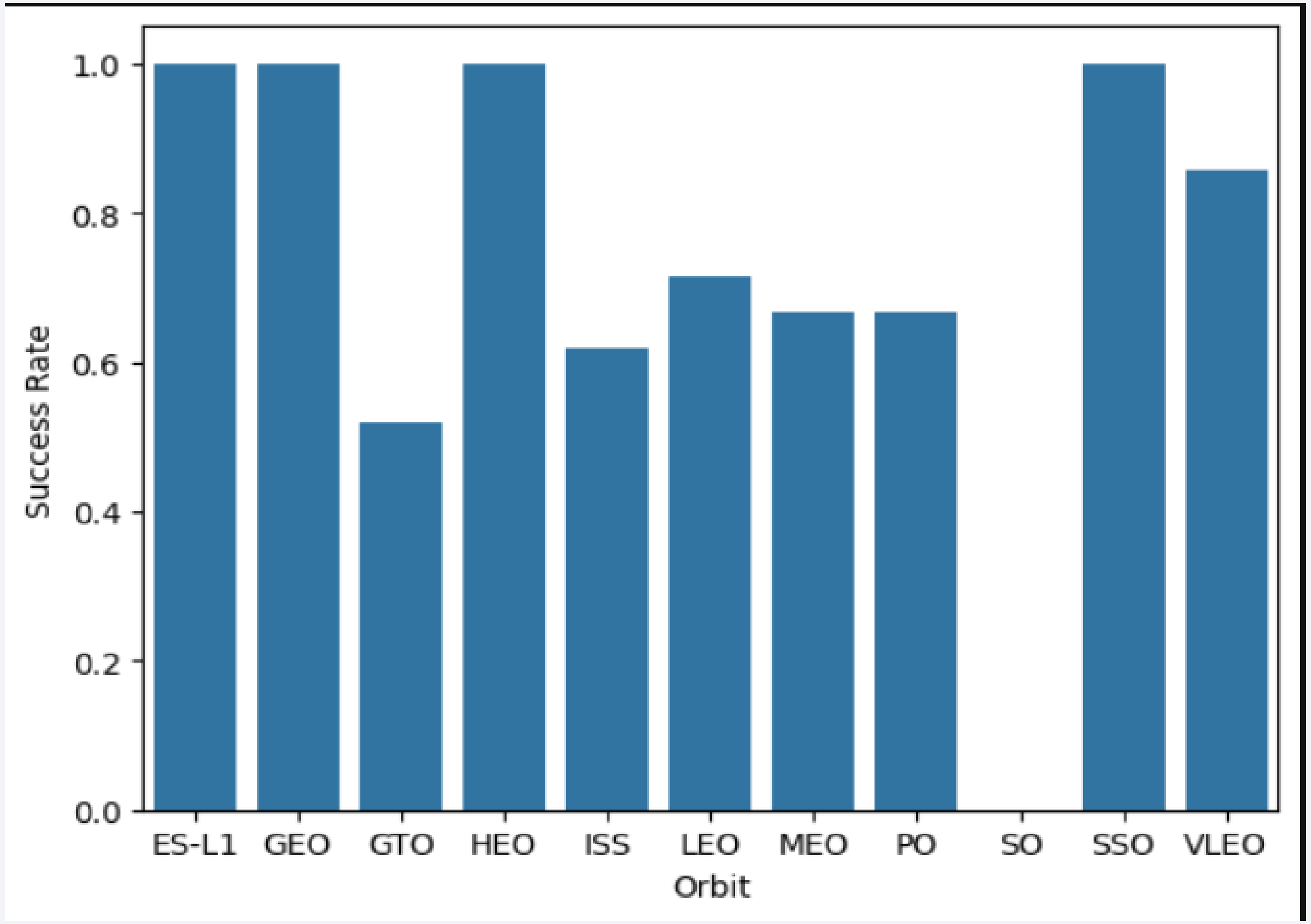
Payload vs. Launch Site



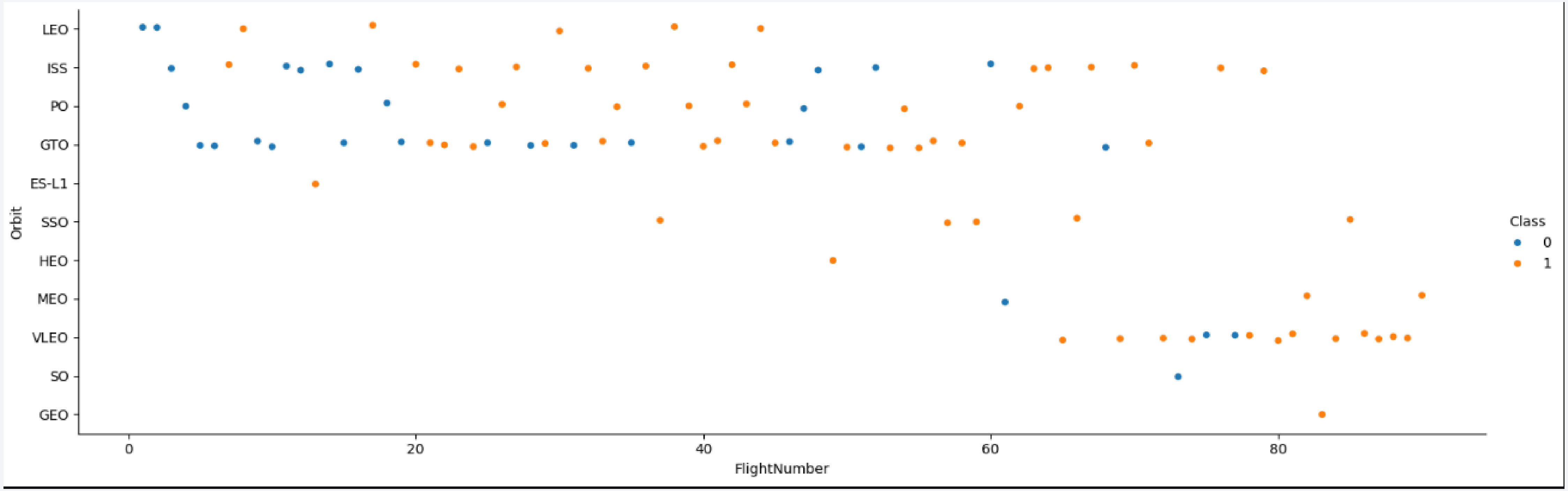
- For higher payload(>12000 kg), CCAFS SLC40 has higher success rate.
- For lower Payloadmass(<6000 kg), VAFB SLC4E and KSC LC 39A have higher success rate.

Success Rate vs. Orbit Type

- Orbits ES L1, GEO, HEO, SSO have higher success rate.

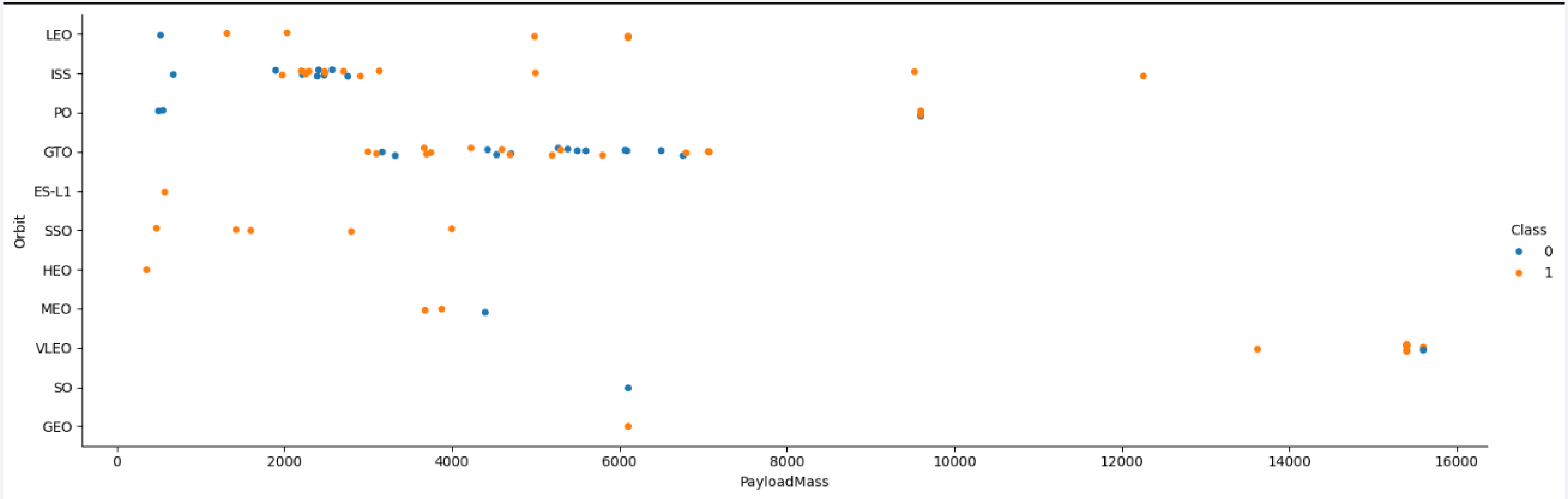


Flight Number vs. Orbit Type



- As Flight Number increases, success rate increases for all orbits except orbit GTO.

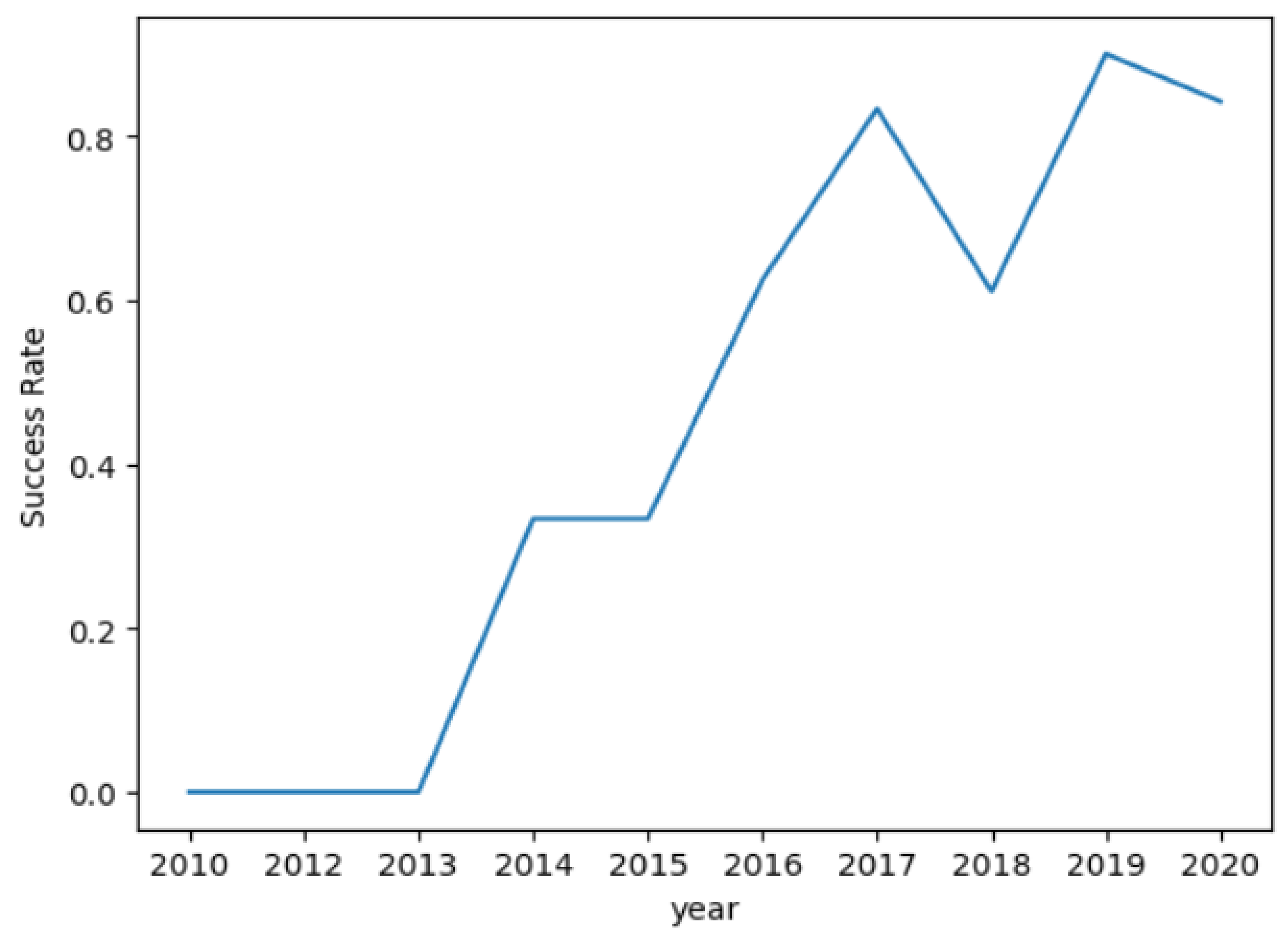
Payload vs. Orbit Type



- For higher PayloadMass, orbits LEO, PO and VLEO have higher successful landings.
- For lower PayloadMass, orbit SSO has higher successful landings.

Launch Success Yearly Trend

- We observe that Success rate increases over the years.



All Launch Site Names

- Displaying unique names of launch site using SQL query.

Task 1

Display the names of the unique launch sites in the space mission

```
[11]: %sql select DISTINCT "Launch_Site" from SPACEXTABLE
```

 * sqlite:///my_data1.db
Done.

[11]:

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

Launch Site Names Begin with 'CCA'

- Display Records where launch sites begin with `CCA` using SQL

Task 2

Display 5 records where launch sites begin with the string 'CCA'

[12]:
```
%sql select * from SPACEXTABLE where "Launch_Site" like 'CCA%' limit 5
```

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

[12]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
[13]: %sql select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Customer"='NASA (CRS)'
```

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

[13]:

sum(PAYLOAD_MASS__KG_)
45596

- As we show that the total payload carried by boosters from NASA is 45596 Kg.

Average Payload Mass by F9 v1.1

Task 4

Display average payload mass carried by booster version F9 v1.1

```
[14]: %sql select avg("PAYLOAD_MASS__KG_") from SPACEXTABLE where "Booster_Version" like 'F9 v1.1%'
 * sqlite:///my_data1.db
Done.
```

[14]:

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

- We show that the average payload mass carried by booster version F9 v1.1is approximately 2534.67 Kg.

First Successful Ground Landing Date

```
[16]: %sql select min("Date") from SPACEXTABLE where "Landing_Outcome" = 'Success (ground pad)'

 * sqlite:///my_data1.db
Done.

[16]: min(Date)
      2015-12-22
```

- The date of the first successful landing outcome on ground pad is 22nd December, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

- The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are shown in below picture.

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
[17]: %sql select Distinct("Booster_Version") from SPACEXTABLE where "Landing_Outcome" = 'Success (drone ship)' and "PAYLOAD_MASS__KG_" be
```

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

[17]:

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

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes are as follows

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

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass are shown in right picture.

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

2015 Launch Records

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

Date	substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
2015-01-10	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- 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, in descending order are as follows

Landing_Outcome	count(Landing_Outcome)
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

All launch sites global map markers

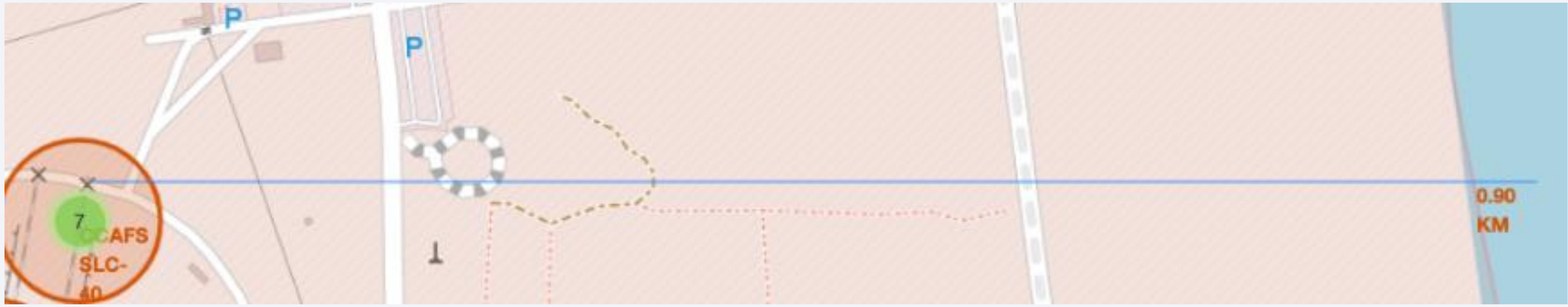
- Folium map shown in figure.
- All launch sites' location markers on a global map shown in red color.

Markers showing launch sites with color labels

- success/failed launches for each site on the map are labelled with Green/Red color as shown in figure.

Launch Site distance to landmarks

- Distance between Launch site CCAFS SLC40 and coastline shown in below map.

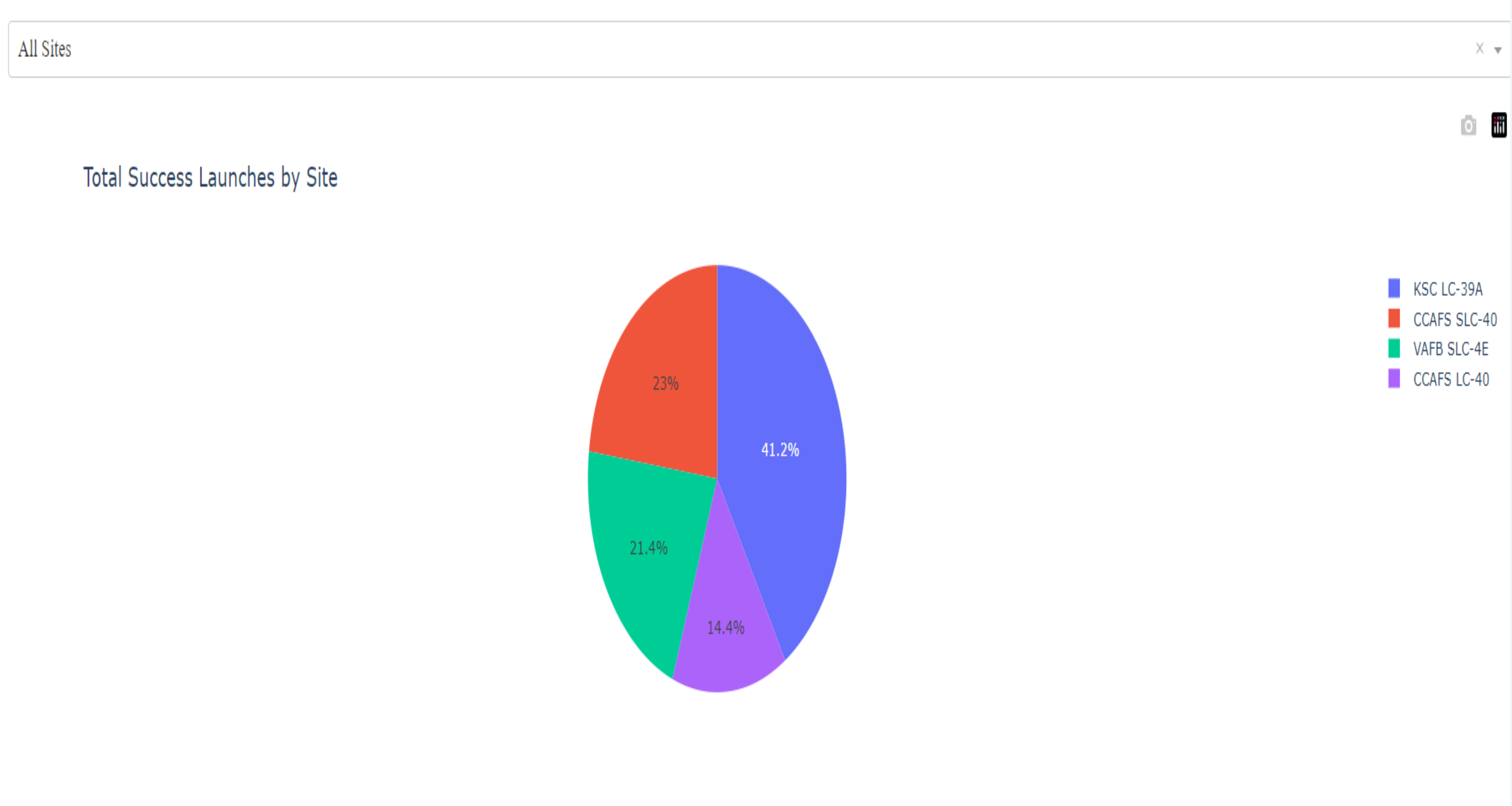


Section 4

Build a Dashboard with Plotly Dash

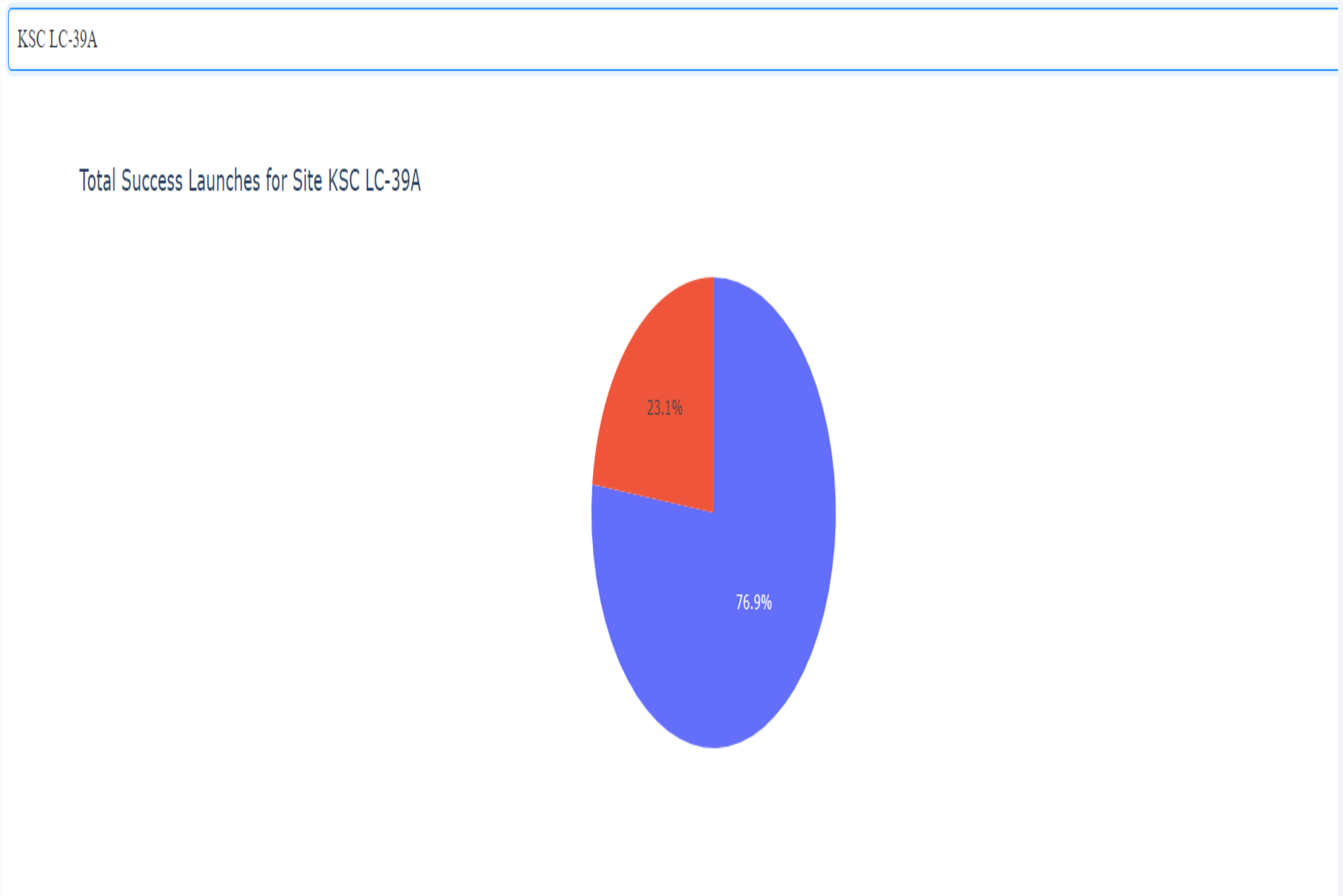
Pie chart showing the success percentage achieved by each launch site

- success percentage achieved by each launch site shown in figure
- Launch Site KSC LC39A has highest success rate.

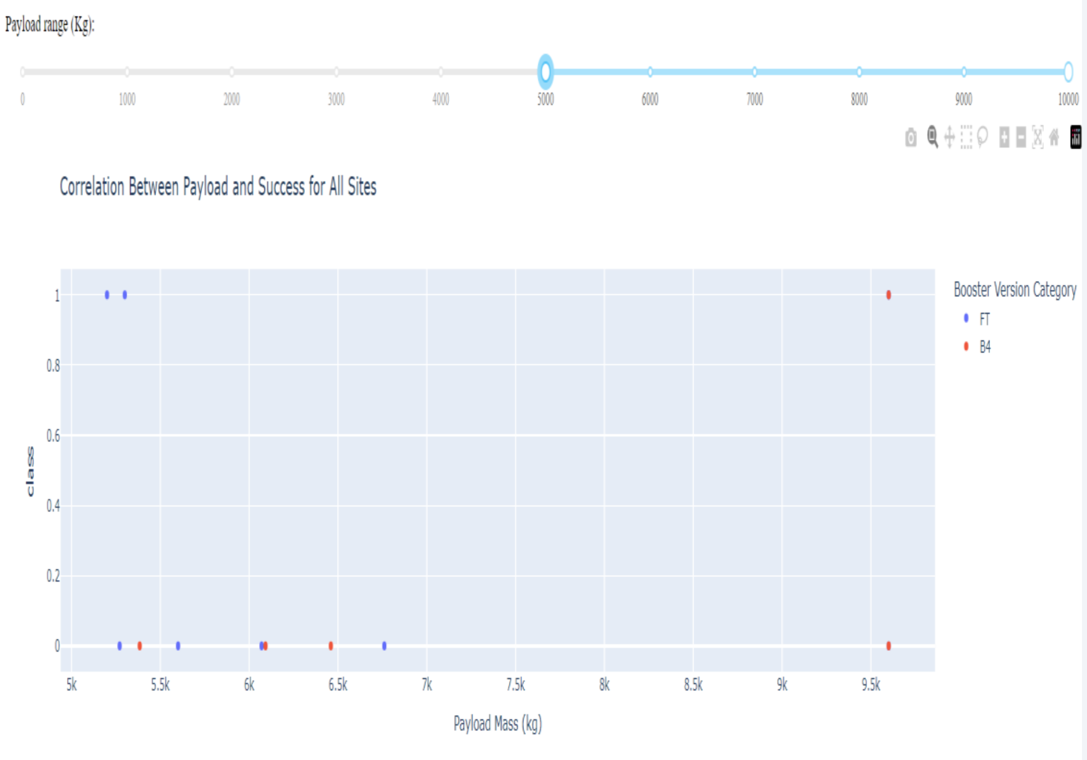


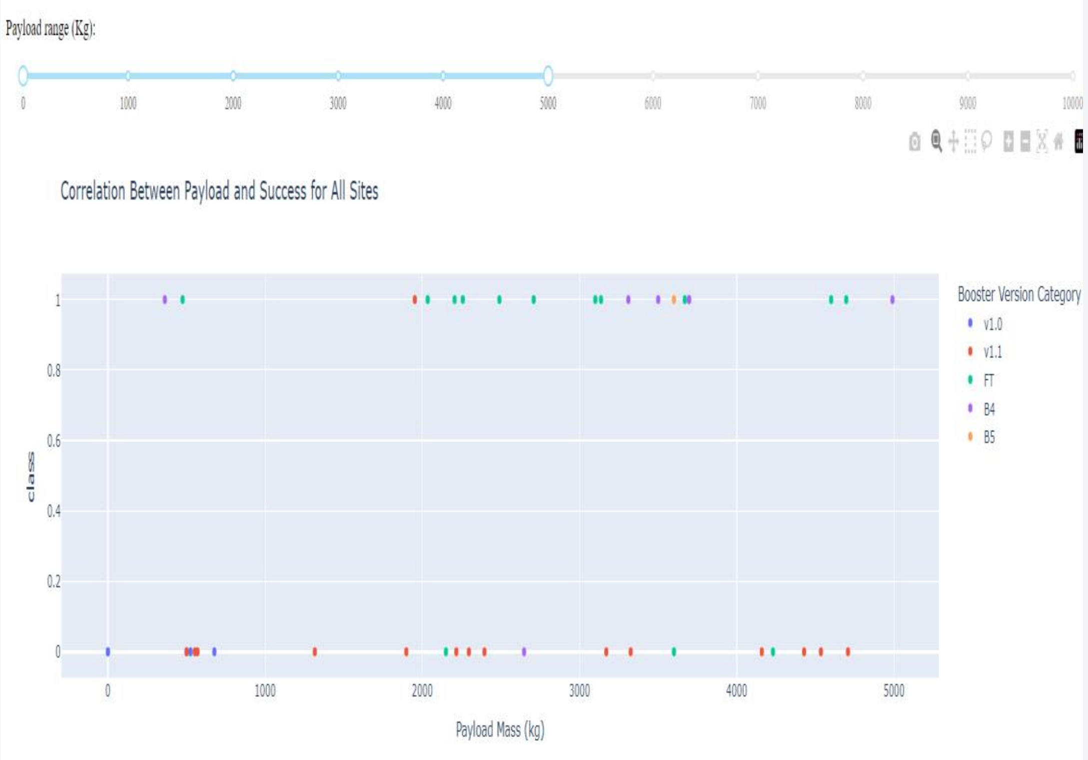
Pie chart showing the Launch site with the highest launch success ratio

- KSC LC39A has 76.9% success rate.



Scatter Plot of Payload vs Success





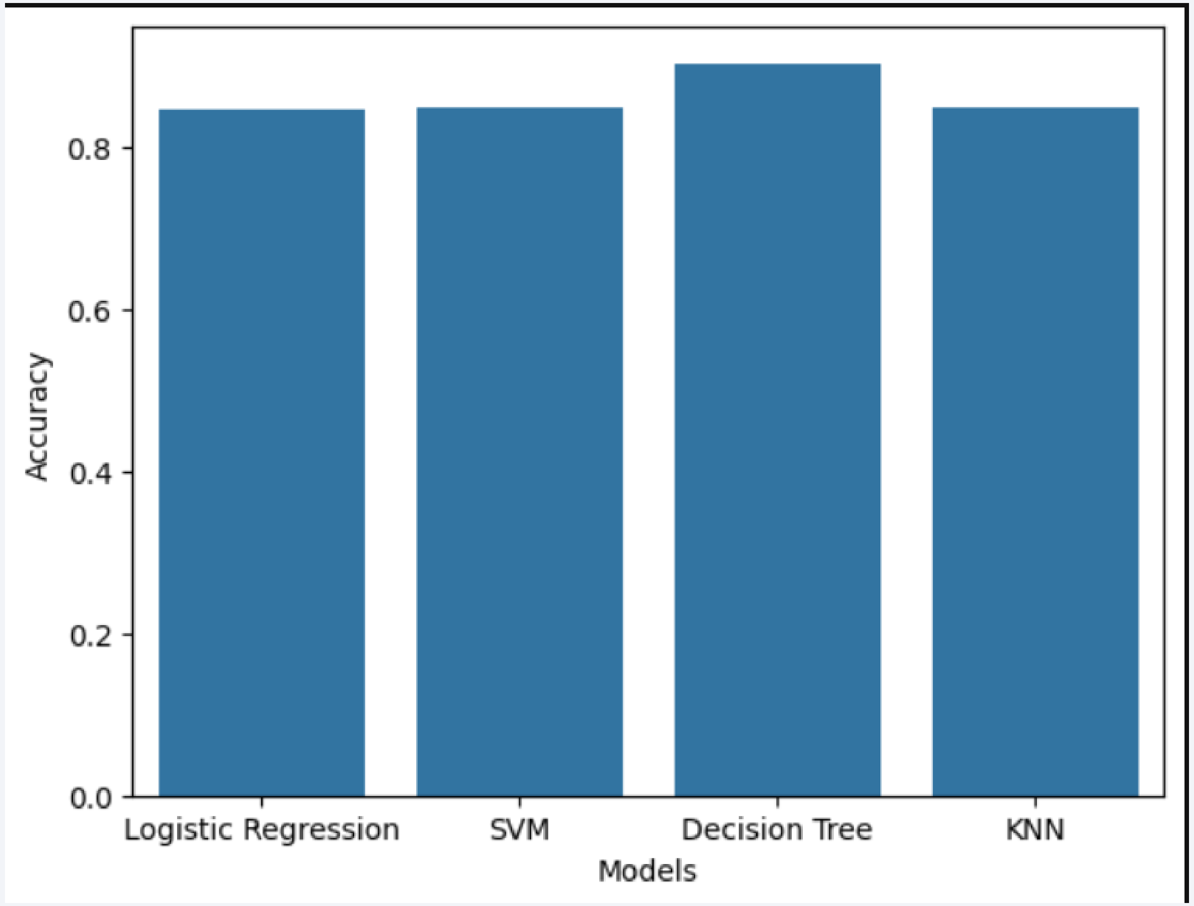
- Scatter Plot of Payload vs Success for different payload range

Section 5

Predictive Analysis (Classification)

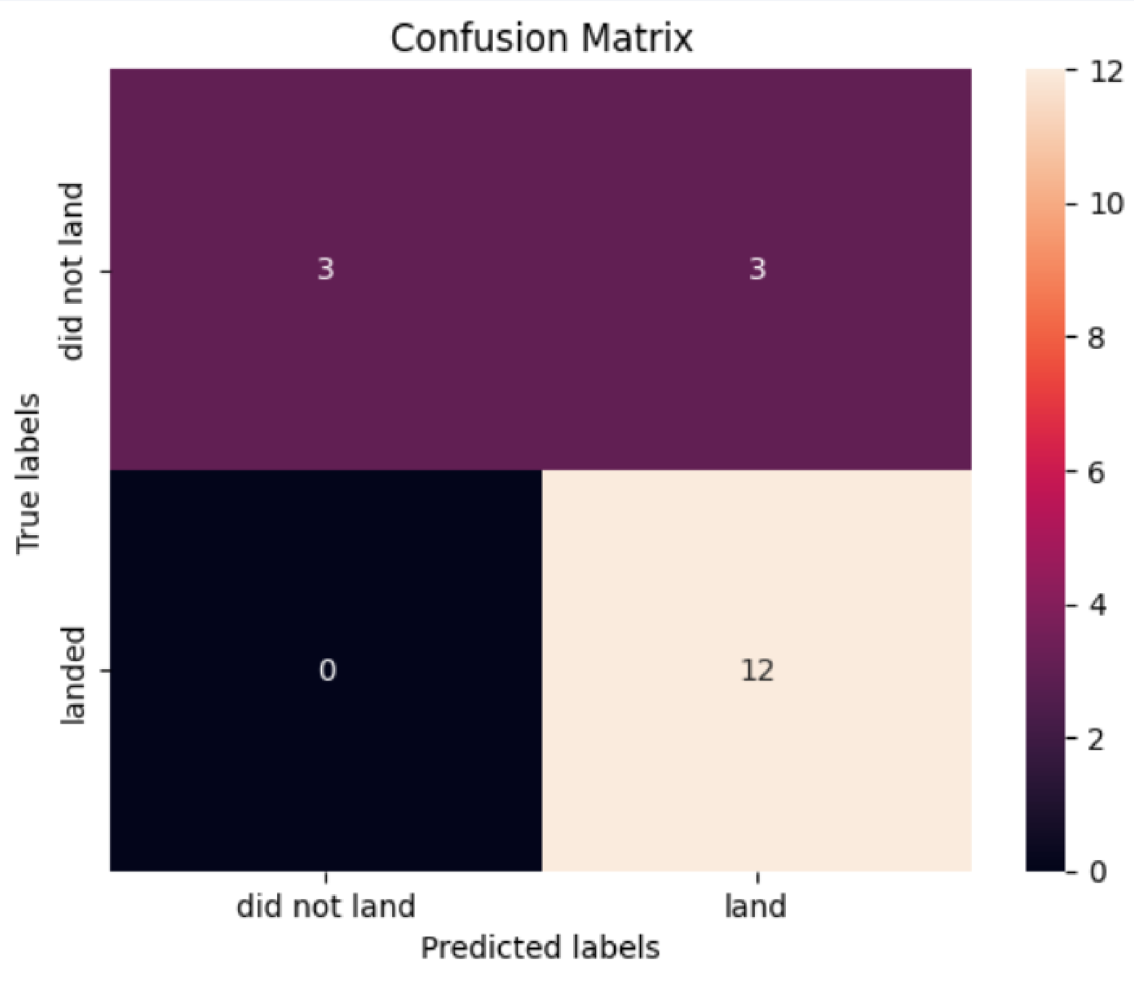
Classification Accuracy

- The built model accuracy for all built classification models, shown in a bar chart
- Decision Tree has highest accuracy.



Confusion Matrix

- Confusion matrix of the Decision Tree model shown in figure.
- Prediction for land outcome is 100% correct.
- Out of 6 did not land outcome model predict 3 as land outcome.



Conclusions

We can conclude that:

- As Flight Number increases, success rate increase at Launch Site.
- For higher payload(>12000 kg), CCAFS SLC40 has high success rate and for lower Payloadmass(<6000 kg), VAFB SLC4E and KSC LC 39A have high success rate.
- Orbits ES L1, GEO, HEO, SSO have higher success rate.
- Success rate start increasing after the year 2013.
- Decision Tree Classification is best model to predict landing outcome for this project.

Thank you!