SpaceX Falcon 9 Launch Prediction

ANDRES
FELIPE ORTIZ
ZAPATA
10/15/2024
<Date>

Outline

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

Executive Summary

- API and webscraping are used to collect data, SQL is used to explore the data on questions such as uniques, dates for first launch. Visualisation is used to make comparisons on factors such as success and payload. Map and dashboard are used for better visualization. Finally prediction is made using classification methods and the best one is explored.
- Some relations can be found between success rate and factors such as payload and locations. Prediction is made using classification method.

Introduction

- This project is to predict if the Falcon 9 first stage will land successfully. We will explore what factors will affect the success rate and which method is the best for prediction.

Section 1

Methodology

Executive Summary

- Data collection methodology
- Perform data wrangling
- 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 Collection

Data was collected using API and web scraping

```
import requests
import pandas as pd
import numpy as np
import datetime
pd.set_option('display.max_columns', None)
pd.set_option('display.max_colwidth', None)

spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)

response.status_code
json = response.json()
data = pd.json_normalize(json)
data.head(5)
```

1. Request using SpaceX API
2. Decode the response using .json
3. Convert into a pandas dataframe using json.normalize()
4. Check status code (success = 200)
5. Preview the first 5 rows using .head()

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and
data = requests.get(static_url).text
soup = BeautifulSoup(data, "html.parser")
```

```
html_tables = []
html_tables = soup.find_all('table')
first_launch_table = html_tables[2]
first_launch_table

column_names = []
for row in first_launch_table.find_all("th"):
    name = extract_column_from_header(row)
    if name != None and len(name) > 0:
        column_names.append(name)
```

1. Use BeautifulSoup to scrape falcon9 data from Wikipedia page
2. Extracting the headers
3. Create a dictionary, parse the content and append to the dictionary
4. Convert the dictionary into datafram

Data Wrangling

Data cleansing

```python
data = data[data['cores'].map(len)==1]
data = data[data['payloads'].map(len)==1]

data['cores'] = data['cores'].map(lambda x : x[0])
data['payloads'] = data['payloads'].map(lambda x : x[0])

data['date'] = pd.to_datetime(data['date_utc']).dt.date

# Using the date we will restrict the dates of the launches
data = data[data['date'] <= datetime.date(2020, 11, 13)]
```

1. Remove rows from ‘cores’ and ‘payload’ columns
2. Extract values from list
3. Converting date into datatime datatype

```python
data_falcon9.isnull().sum()
mean = data_falcon9.PayloadMass.mean()
data_falcon9.PayloadMass.replace(np.nan, mean, inplace = True)
# Replace the np.nan values with its mean value
data_falcon9.info()
```

1. Check for missing data
2. Replacing missing value with mean of the column

[Machine-Learning/DataCollection.ipynb at main · Samantha688/Machine-Learning (github.com)]()

EDA with Data Visualization

- Scatter plot, bar chart and line chart are used. One hot encoding is used for feature engineering
- Machine-Learning/EDA with visualisation.ipynb at main · Samantha688/Machine-Learning (github.com)

EDA with SQL

Queries made using SQL:

- Unique launch sites
- Records of launch sites which have names with 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 succesful landing outcome in ground pad was achieved
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Total number of successful and failure mission outcomes
- Names of the booster versions which have carried the maximum payload mass.
- Records which will display the month, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015
- Rank of the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Machine-Learning/SQL EDA.ipynb at main · Samantha688/Machine-Learning (github.com)

Build an Interactive Map with Folium

- To visualize the launch sites and see if there is any connections between the locations and success rates
- Initiate the map, add 4 circles and markers for the 4 launch sites, add marker cluster for the success or failure icons and add a line from the launch site to coastline and distance calculated
- Machine-Learning/Map.ipynb at main · Samantha688/Machine-Learning (github.com)

Build a Dashboard with Plotly Dash

- A drop down is created to select either all sites or selected sites. A pie chart is used to visualise proportion of launch outcomes for each site.A scatter plot is used to visualise launch outcome and payload with payload range as a slider
- Machine-Learning/Dashboard.ipynb at main · Samantha688/Machine-Learning (github.com)

Predictive Analysis (Classification)

- Create numpy array from class column from dataset 2
- Transform dataset 3 using Standard Scaler
- Perform train_test_split to the two datasets
- Create objects and fit the datasets using different classification models – Logistic regression, SVM, decision tree and KNN
- Find the best parameters, calculate the accuracy on test data and check confusion matrix

• Machine-Learning/PredictiveAnalysis.ipynb at main · Samantha688/Machine-Learning (github.com)

Results

- Success rate can be dependent on orbit and for each orbit, the flight number and payload can affect the success rate for the specific orbit. But in general success rate increases over time.
- KSC LC-39A has the highest success rate, likely due to its location farther from coastline
- Decision trees provide the best accuracy in prediction while support vector matrix provides the best result in confusion matrix

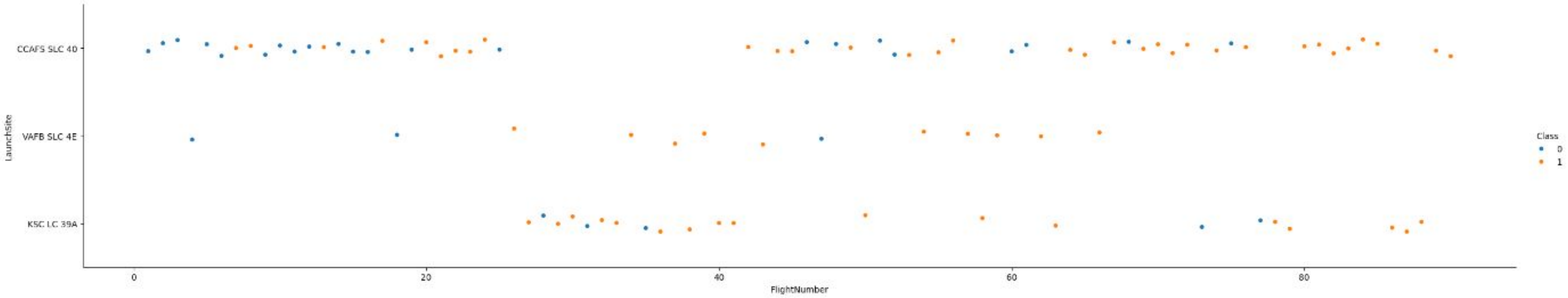
Section 2

Flight Number vs. Launch Site

- Flight number increases, success increases

```
[7]: sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
```

```
[7]: <seaborn.axisgrid.FacetGrid at 0x2416cbb8590>
```

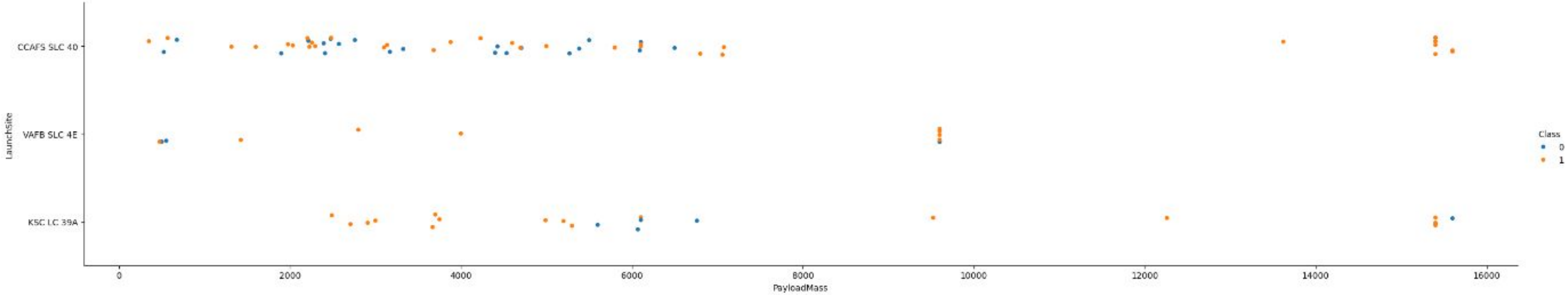


Payload vs. Launch Site

- Less launches for higher payload

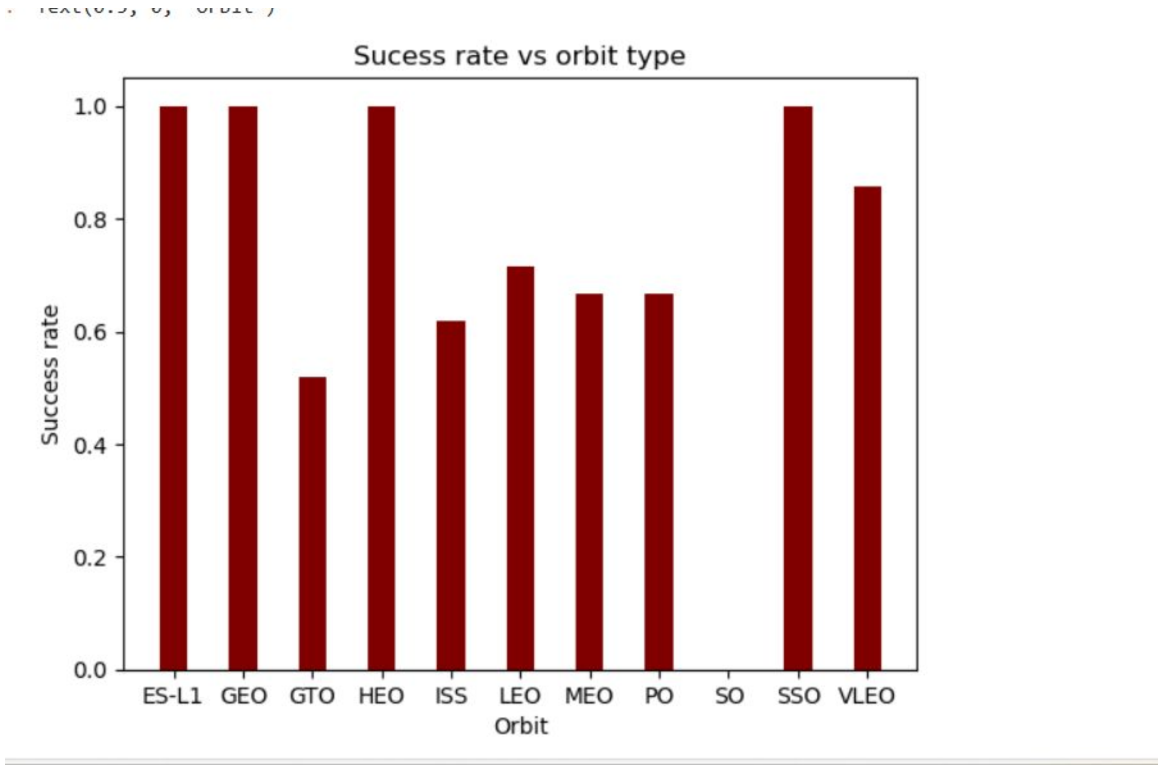
```
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 5)
```

<seaborn.axisgrid.FacetGrid at 0x2416d23eb10>



Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO have the highest success rate

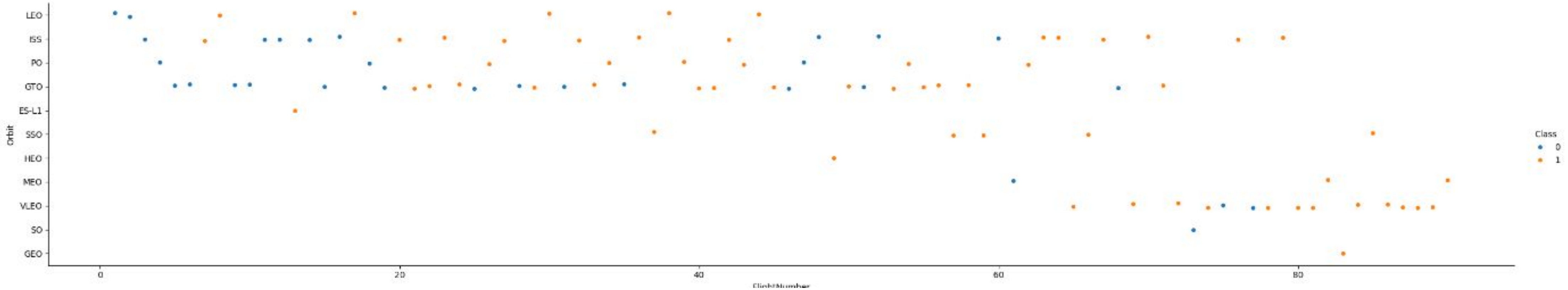


Flight Number vs. Orbit Type

- For some orbits flight number increases, success increases while for some there is no relation

```
sns.catplot(y="Orbit", x="FlightNumber", hue="Class", data=df, aspect = 5)
```

```
<seaborn.axisgrid.FacetGrid at 0x2416d3a26f0>
```

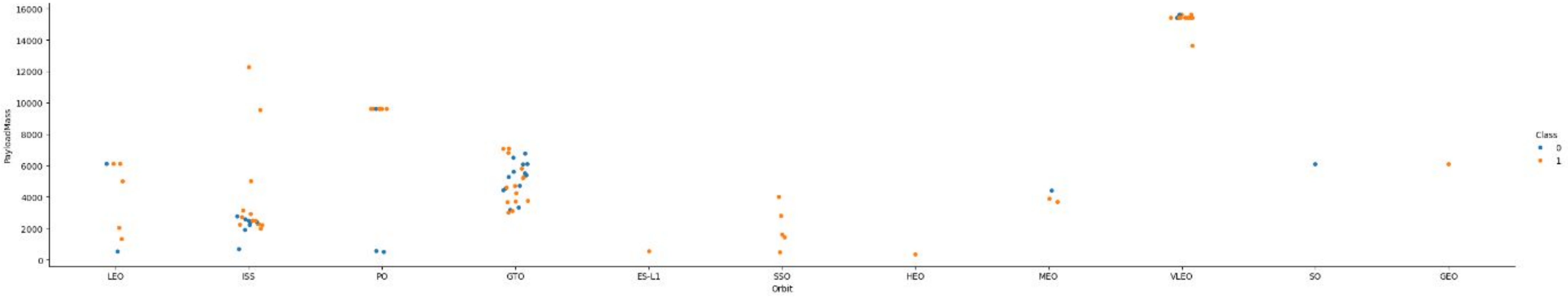


Payload vs. Orbit Type

- Payload increases success increases for some orbit but not all

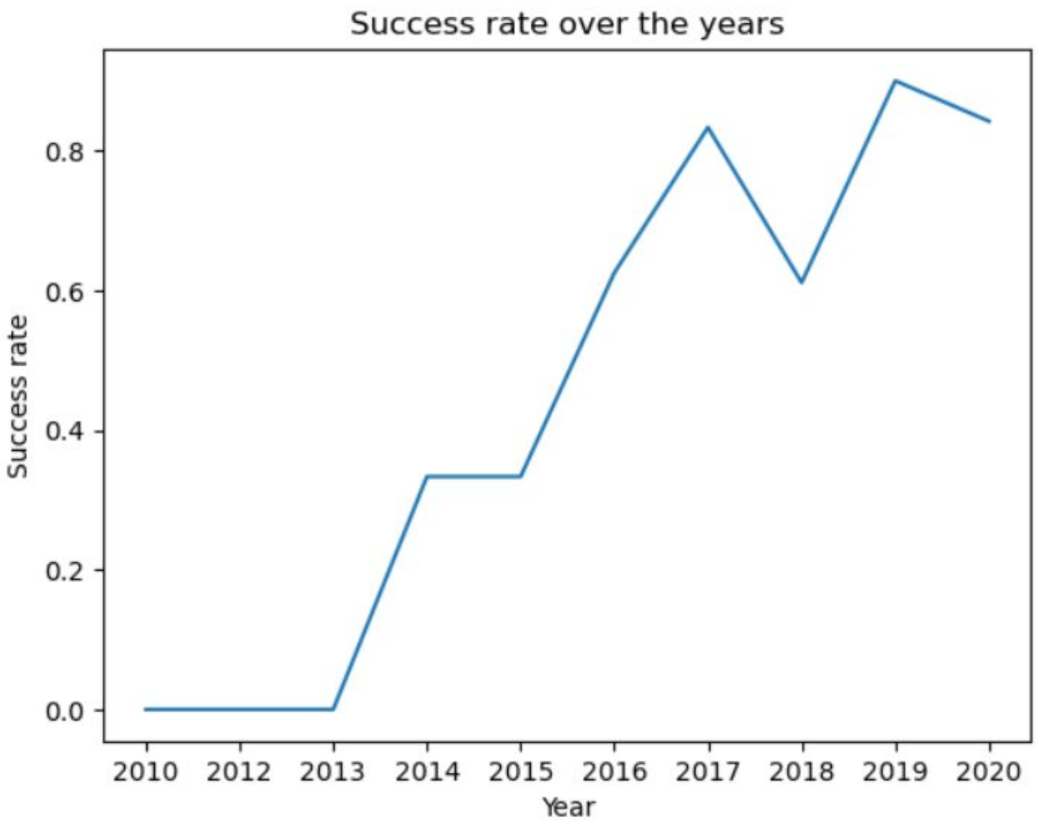
```
sns.catplot(y="PayloadMass", x="Orbit", hue="Class", data=df, aspect = 5)
```

<seaborn.axisgrid.FacetGrid at 0x2416d5515b0>



Launch Success Yearly Trend

- Success rate increases over the year



All Launch Site Names

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

```
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE
```

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

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

Launch Site Names Begin with 'CCA'

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

```
%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	Missio
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	
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	
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	

Total Payload Mass

Task 3

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

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS Total_Payload_Mass FROM SPACEXTABLE WHERE Customer = "NASA (CRS)"
```

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

Total_Payload_Mass
45596

Average Payload Mass by F9 v1.1

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS Average_Payload_Mass FROM SPACEXTABLE WHERE Booster_Version like "F9 v1.1%"
```

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

Average_Payload_Mass
2534.6666666666665

First Successful Ground Landing Date

```
%sql SELECT Min(Date) AS First_date FROM SPACEXTABLE WHERE Landing_Outcome = "Success (ground pad)"
```

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

First_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql SELECT Payload FROM SPACEXTABLE WHERE Landing_Outcome = "Success (drone ship)" AND (PAYLOAD_MASS__KG_) > 400 AND (PA
```

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

Payload
SpaceX CRS-8
JCSAT-14
Thaicom 8
JCSAT-16
SES-10
BulgariaSat-1
Formosat-5
SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql SELECT COUNT(*) FROM SPACEXTABLE WHERE Landing_Outcome = "Success%" OR Landing_Outcome like "Failure%"
```

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

COUNT(*)
10

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
]: %sql SELECT Booster_Version FROM SPACEXTABLE WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTABLE)
```

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

]:

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

```
%sql SELECT substr(Date,6,2), Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTABLE WHERE substr(Date,0,5) = "20
```

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

substr(Date,6,2)	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

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.

```
%sql SELECT Landing_Outcome, COUNT(*) AS Number FROM SPACEXTABLE WHERE Date >= "2010-06-04" AND Date <= "2017-03-20" GROUI
```

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

Landing_Outcome	Number
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

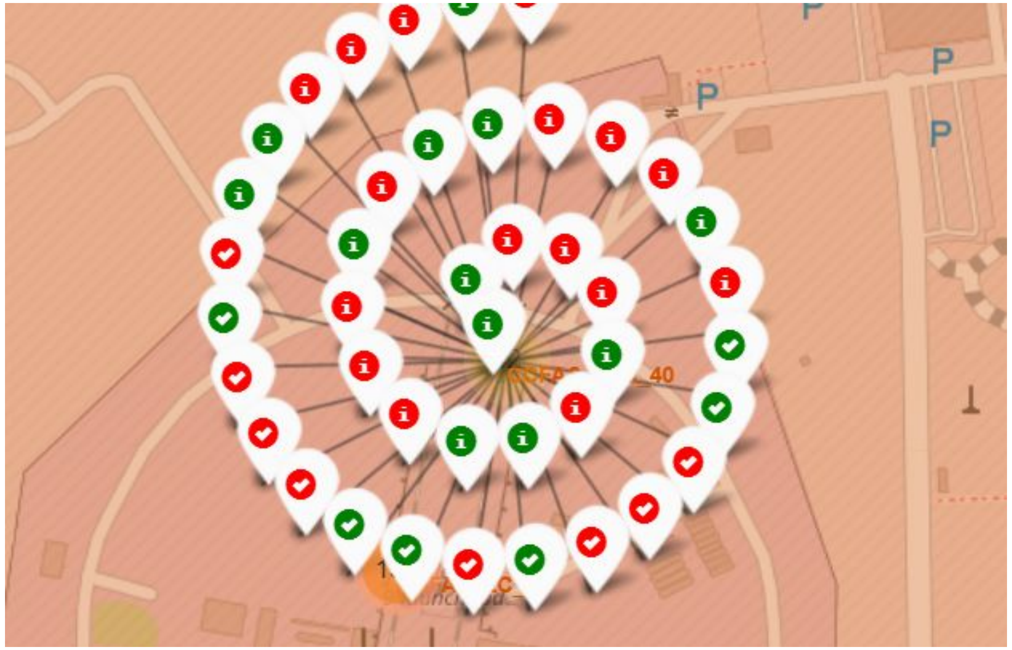
Section 3

Launch sites on a map

- To visualize the launch site and investigate any relation with success rate
- It is noted that all launch sites are close to coastline except KSC LC-39A

Launch outcomes on sites

- It is noted that the sites closer to coastline has lower success rates

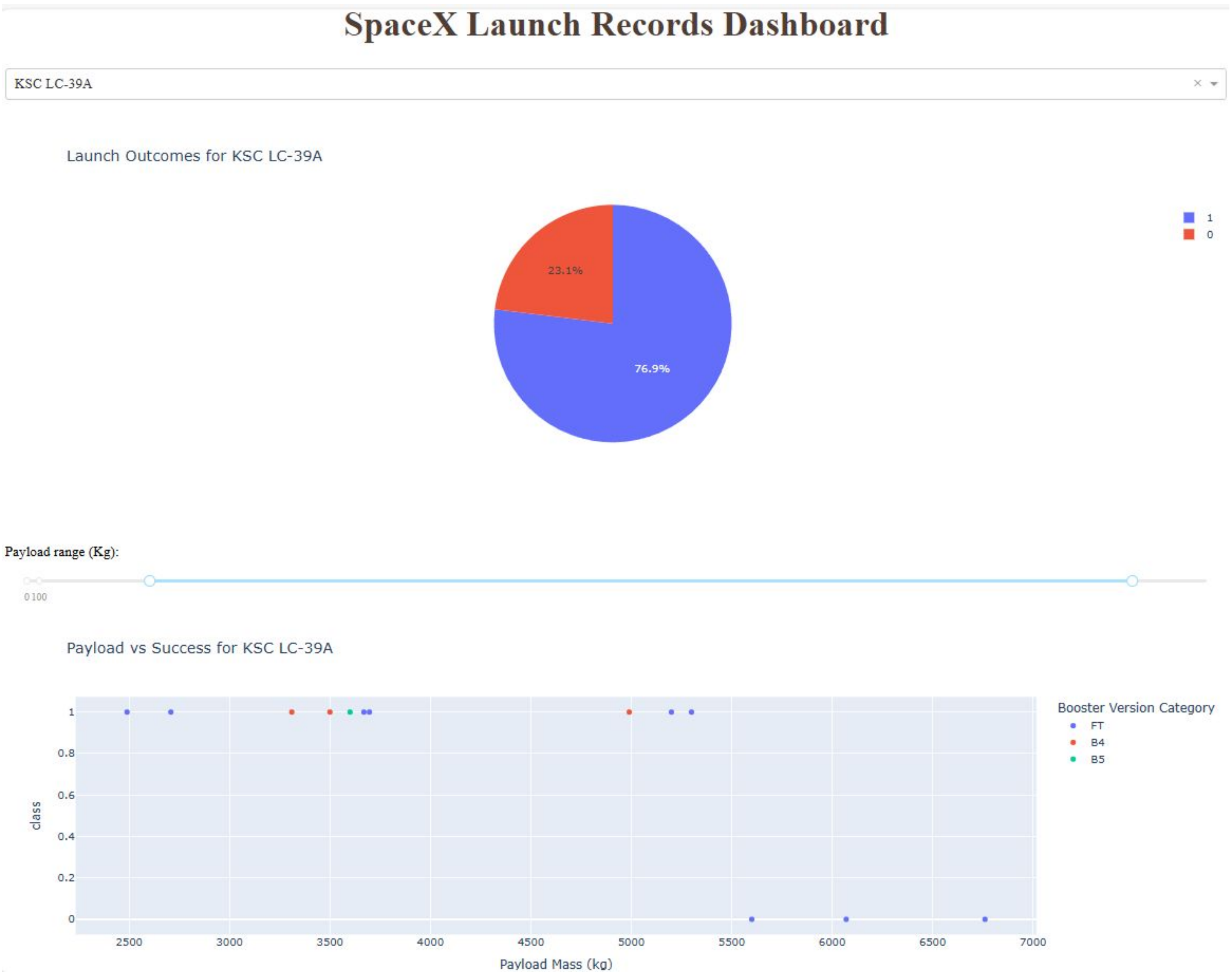


Distance between coastline and sites

- We can calculate the distance between coastline and sites and mark the line on the map

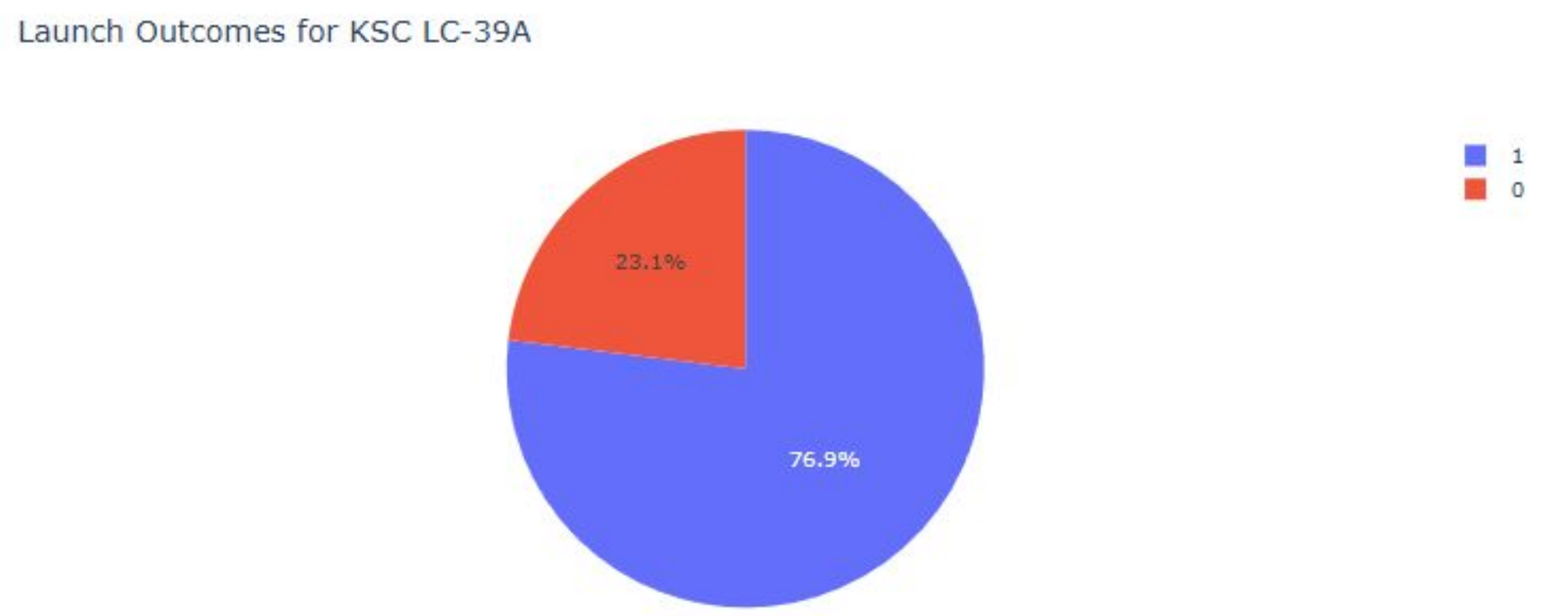
Section 4

Dashboard with pie chart and scatter plot



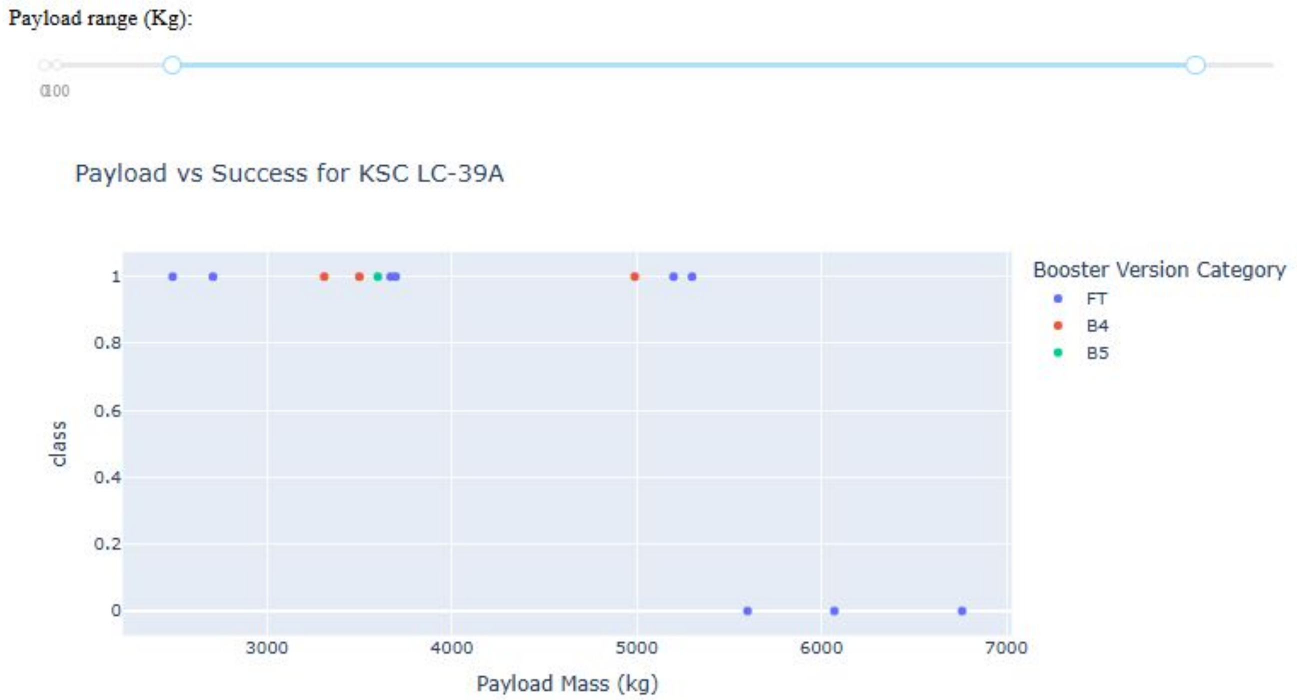
Pie chart of sites vs launch outcomes

- KSC LC-39A has the highest success rate with 76.9%



Scatter plot of payload vs success rate

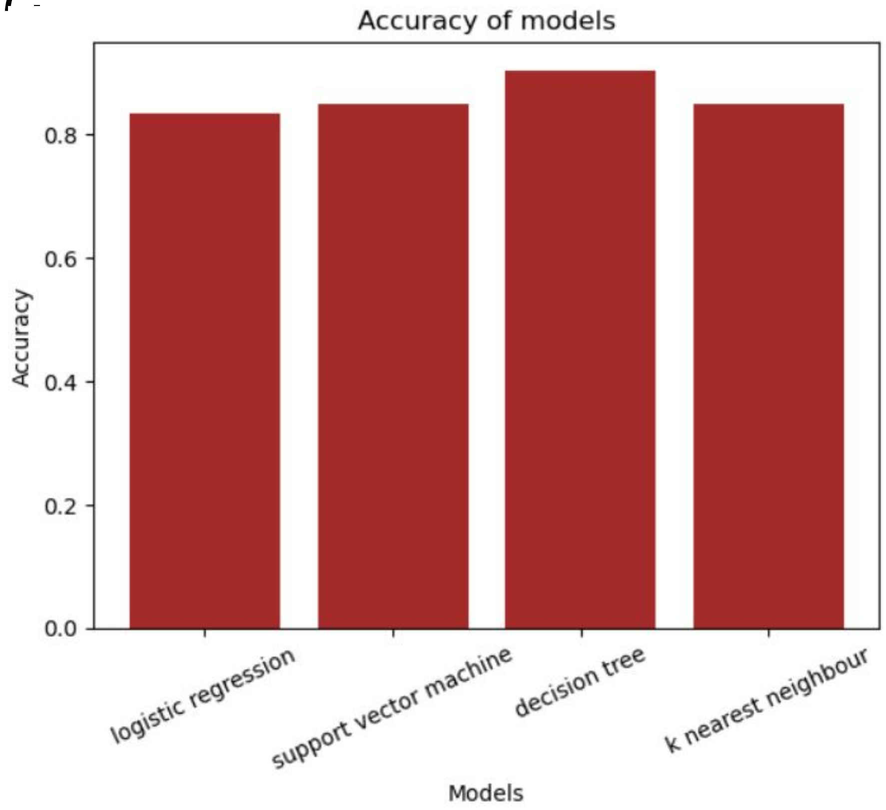
- It is noted that success rate is higher with lower payload



Section 5

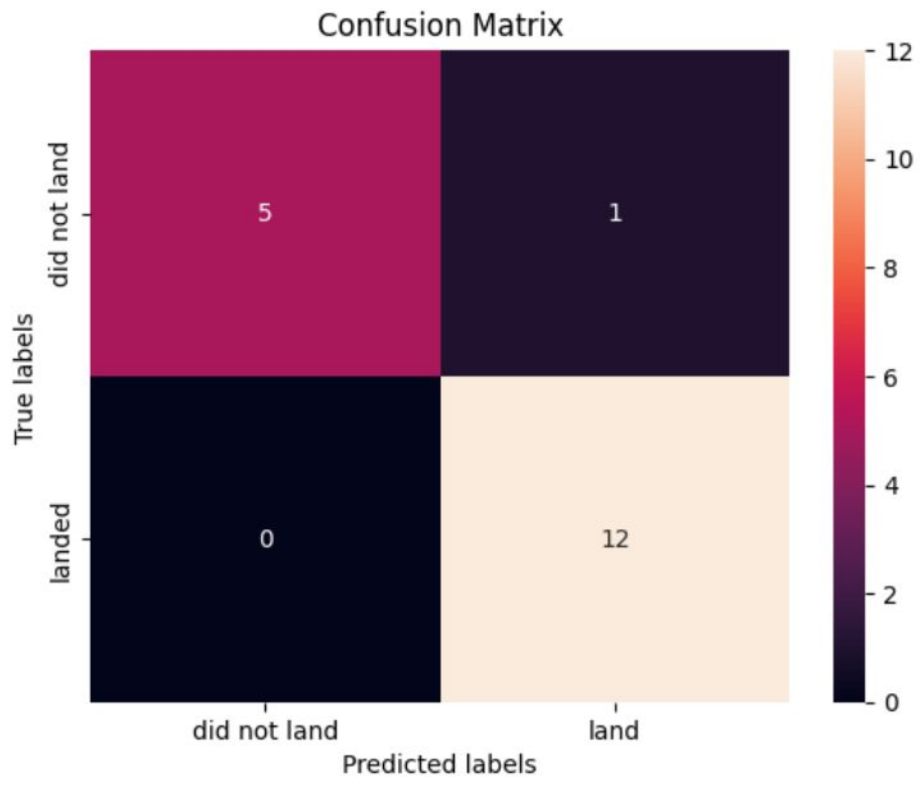
Classification Accuracy

- Decision Tree has the highest accuracy on the test data – which a number of 0.9035714285714287



Confusion Matrix

- Support vector machine has the best performance with 12 true positives and 5 false negat



Conclusions

- Success rate is dependent on orbit, flight number and payload which can be orbit-specific
- Success rate can be related to the location of the launch sites, whether it is close to coastline or not
- Decision tree and support vector matrix are considered better in prediction among the 4 classification methods
- Success rate has improved over time

Appendix

- Visit the GitHub URL for codes