IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Walar Min Khant
07-06-2024

Outline

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

Executive Summary

Methodology summary in the case study

1. Data Collection Using API
2. Data collection by Web scraping
3. Data wrangling
4. Data analysis with SQL
5. Data analysis with visualizations
6. Interactive visual analytics with Folium
7. Machine learning prediction

Introduction

- This is a capstone project as a part of the IBM python Data Science course.
- In this case, i will be conducting a data analytics project on the launch data gathered from the space x
- This data encompasses every data regarding the rockets that have been launched by SpaceX
- Data visualizations are created based on these data
- This data will be used to construct the data model to predict the outcome of future space expeditions
- Instead of using Engineering knowledge to predict the success of rocket launches, i will be using the power of data analytics.

Section 1

Methodology

Methodology

Executive Summary

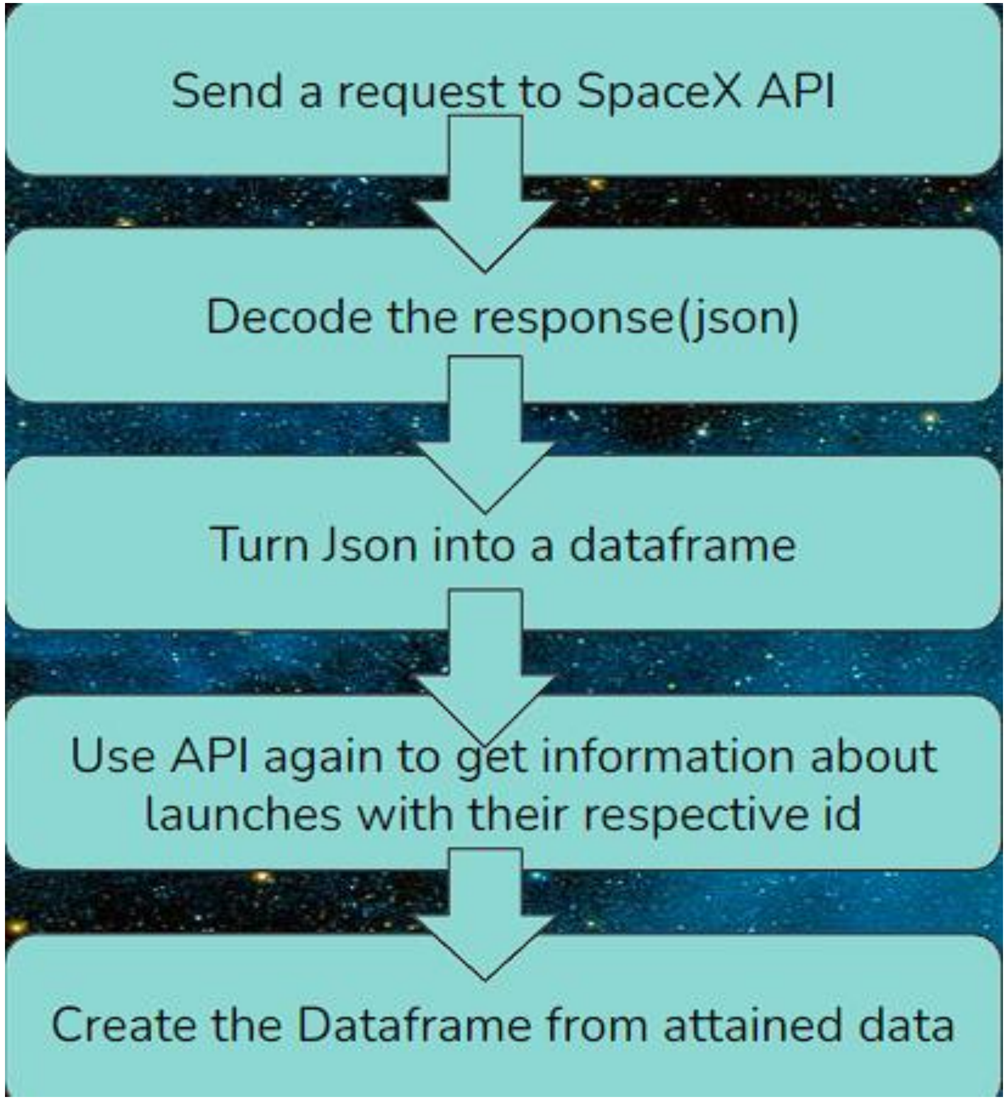
- Data collection methodology:
  - Get requests from SpaceX API and web scraping from Wikipedia
- Perform data wrangling
  - Clean and format the data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create a Machine learning model based on multiple parameters

Data Collection

- Acquire the necessary data from the specified link
- Turn that data into easily accessible data frame
- Remove unnecessary data or fill in the necessary data

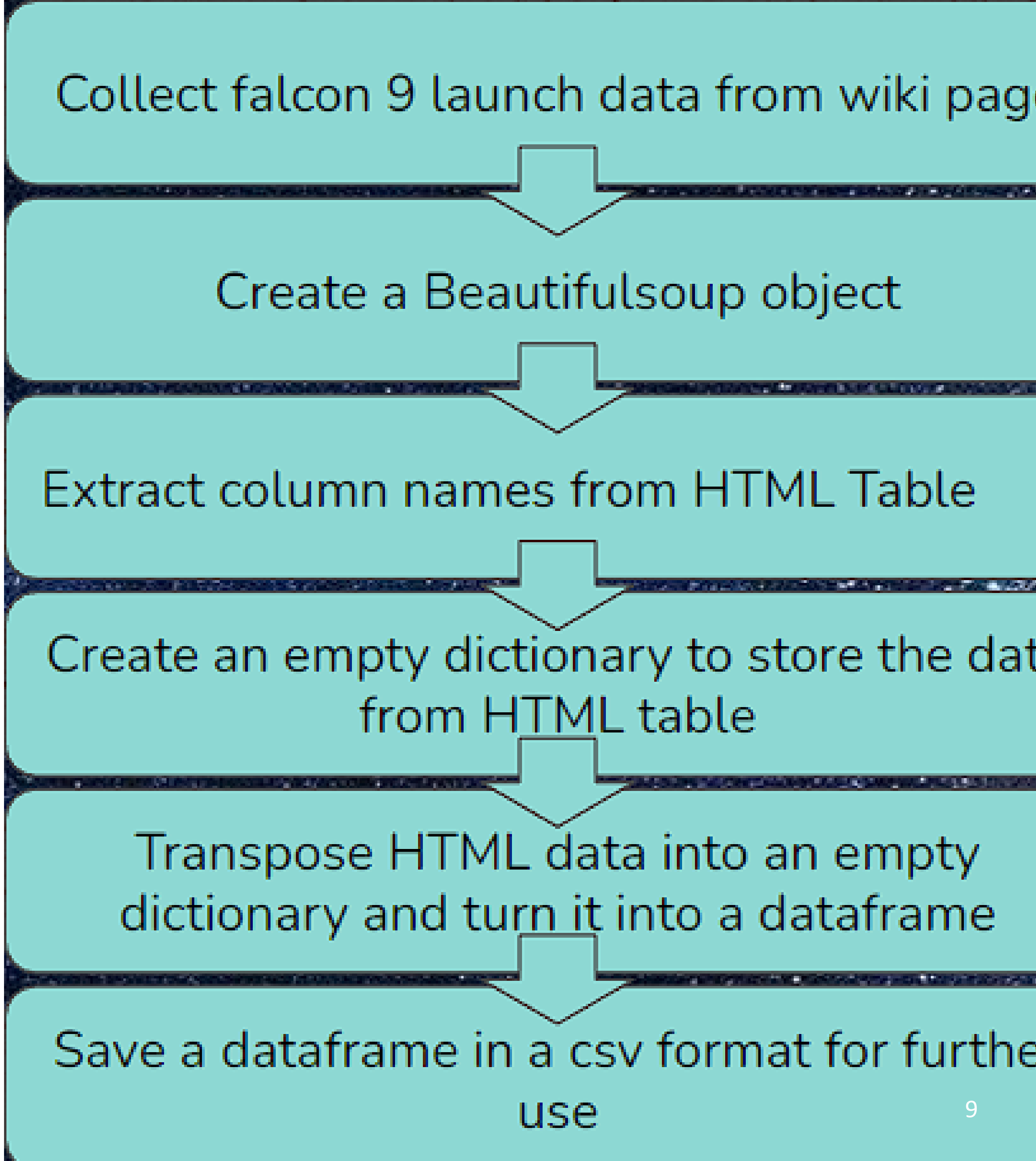
Data Collection – SpaceX API

- The process of collecting data with API is summarized on the flow chart at right side. For the details of the entire process, you can check the link below .
-
- https://github.com/walar2/capston-ibm-python/blob/a0f2a399cf6061c6fd5bc84436995f8be7dbb92a/jupyter-labs-spacex-data-collection-api.ipynb



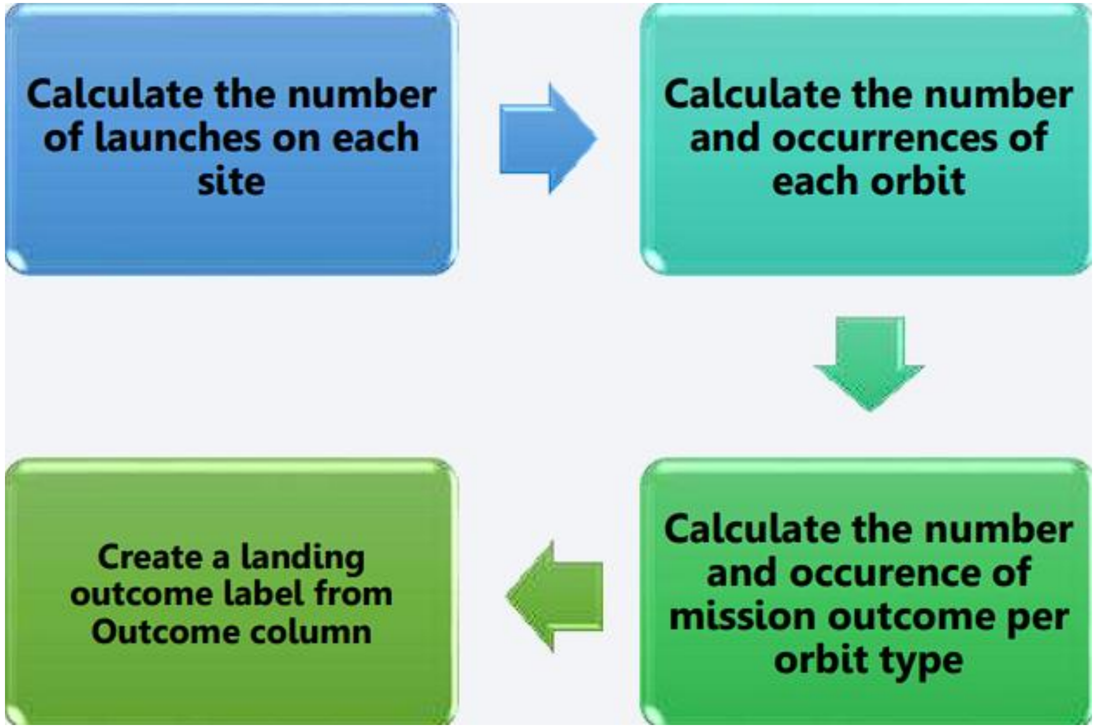
Data Collection - Scraping

- Same as API method, data collection by web scraping is summarized by a flow chart on the right. You can access to the full process in detail by following the link below.
- https://github.com/walar2/capston-ibm-python/blob/a0f2a399cf6061c6fd5bc84436995f8be7dbb92a/jupyter-labs-webscraping.ipynb



Data Wrangling

- Data extracted were cleaned and formatted and cleaned for further processing.
- https://github.com/walar2/capston-ibm-python/blob/e2b150e6d23db9d24e8d8fde9e98e0a720db64f5/labs-jupyter-spacex-Data%20wrangling.ipynb



EDA with Data Visualization

- Scatter plot were used to compare Flight number vs payload mass, flight number vs launch sites, Payload and launch sites, Flight number and orbit type , Payload and orbit type
- Bar chart is used to display the success rate of each orbit type
- Line chart is used to display the success rate of launches in specific time frame
- https://github.com/walar2/capston-ibm-python/blob/e2b150e6d23db9d24e8d8fde9e98e0a720db64f5/edadataviz.ipynb

EDA with SQL

Using the Sequel language to find

- The name of launch sites in the space mission
- 5 records where the launch sites begin with CCA(string)
- The total payload mass carried by boosters launched by NASA
- The average payload mass carried by Booster version F9 v1.1
- The date when the first successful landing outcome in ground pad was achieved
- The names of the boosters which have success in drone ship and have payload mass greater than 4000 kg and less than 6000 kg
- The total number of successful and failure mission outcomes
- The names of booster versions that have carried the maximum payload mass(used subquery)
- The failed landing outcomes in droning ship,their booster versions and launch site names for the year 2015
- Ranking the count of landing outcomes between 2010-06-04 and 2017-03-20

https://github.com/walar2/capston-ibm-python/blob/d7c4a1cd3e89339a9bc628fa48fd57fb93f525b6/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

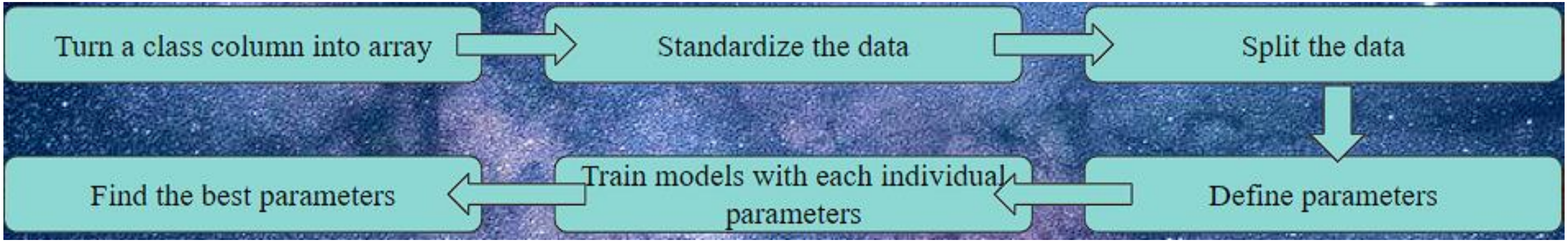
- In this section, i used folium library to display
- Launch site locations
- Successful/failed launches of each site on the map
- The distance between a launch site to its proximities
- Circles were used to highlight launch sites
- Dotted lines were used to represent the distance between launch site and its proximities
- https://github.com/walar2/capston-ibm-python/blob/e2b150e6d23db9d24e8d8fde9e98e0a720db64f5/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie charts and scatter plots are added.
- Afore mentioned charts are added to display the launch success records of spaceX
- Successful launch mean 1 in class values and failure means zero
- https://github.com/walar2/capston-ibm-python/blob/e2b150e6d23db9d24e8d8fde9e98e0a720db64f5/spacex_dash_app.py

Predictive Analysis (Classification)

- A machine learning model is built using Scikit-learn library using multiple parameters and built around the success rate of each launch sites
- https://github.com/walar2/capston-ibm-python/blob/e2b150e6d23db9d24e8d8fde9e98e0a720db64f5/spacex_dash_app.py



Results

- Exploratory data analysis has shown that most of the launch sites locates within 1-2km from the coastline and the landing outcomes success rate have been gradually increasing over the years at a steady rate since 2015
- Site also have access to the major roadways
- The machine learning had resulted in models that can predict the outcomes with 83.33% accuracy

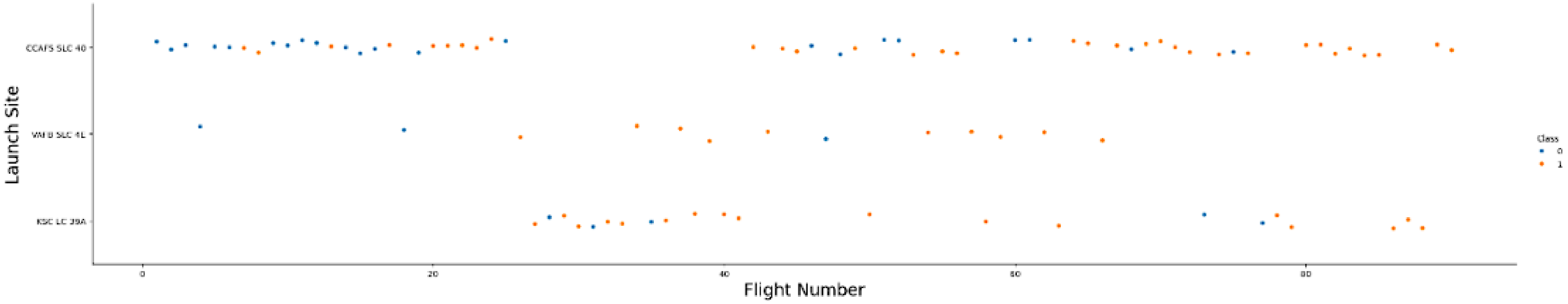
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- a scatter plot of Flight Number vs. Launch Site

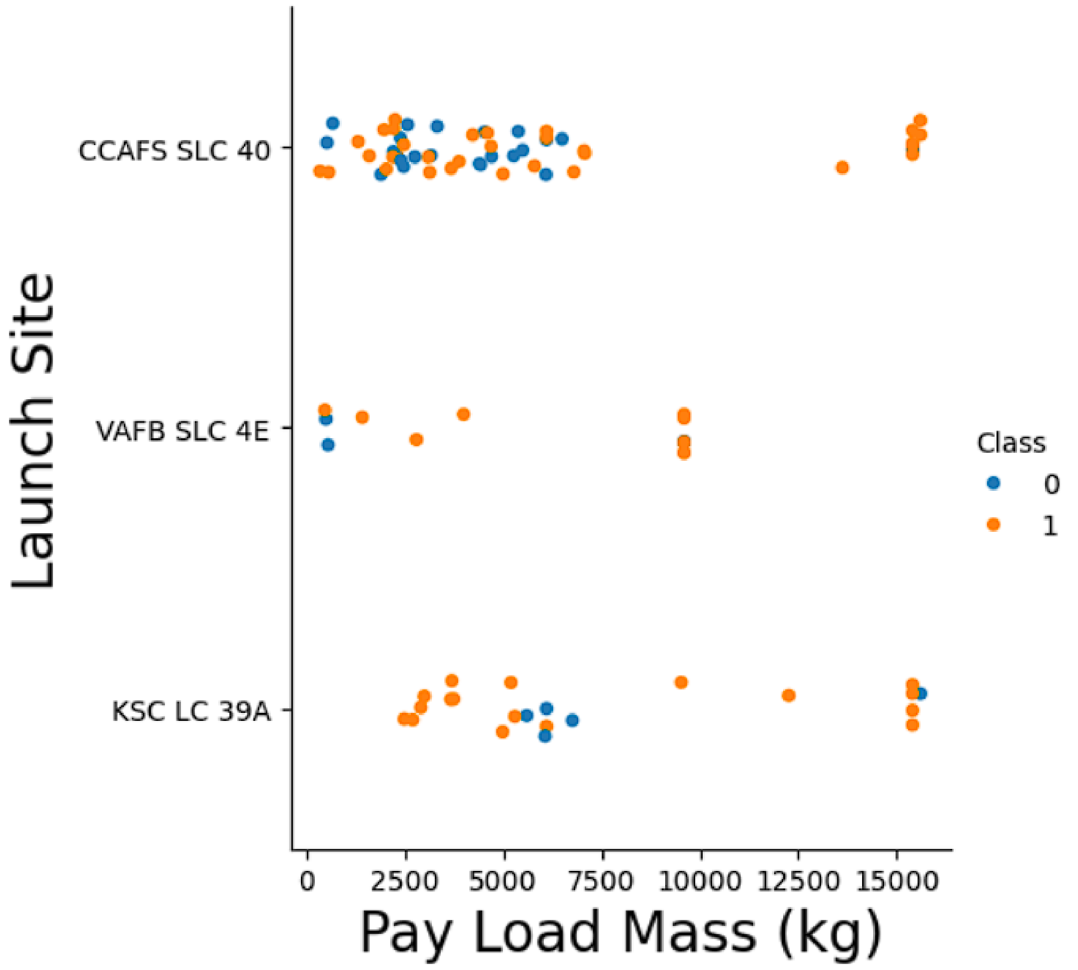
```
### TASK 1: Visualize the relationship between Flight Number and Launch Site
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



Payload vs. Launch Site

- a scatter plot of Payload vs. Launch Site

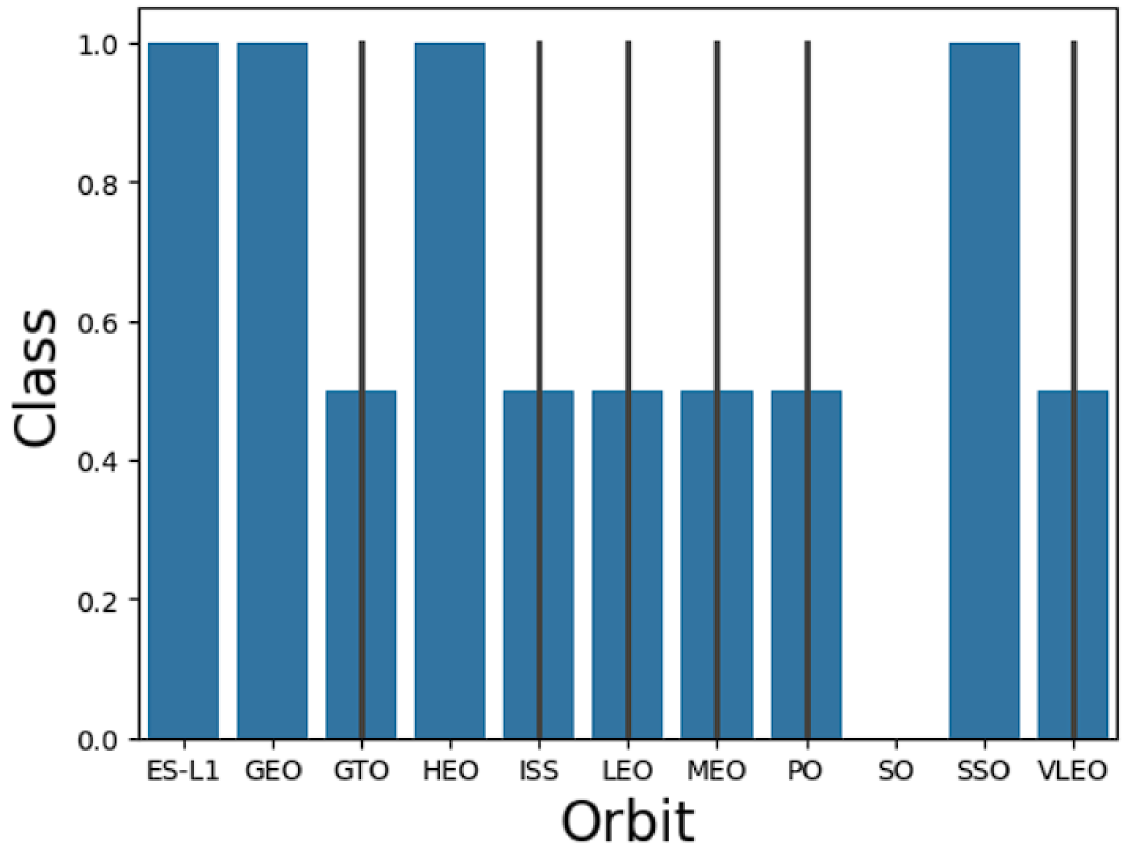
```python
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df)
plt.xlabel("Pay Load Mass (kg)",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



Success Rate vs. Orbit Type

- a bar chart for the success rate of each orbit type

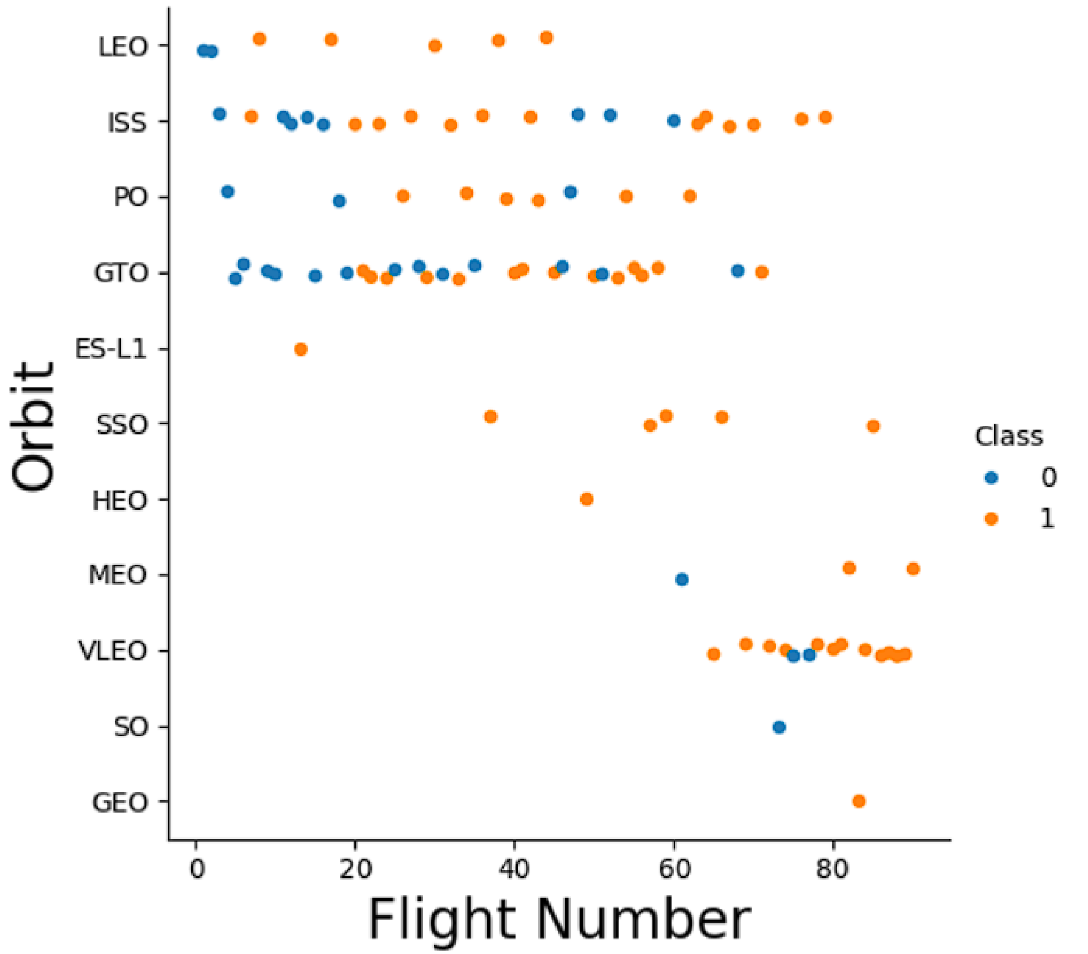
```
# Create a bar plot with custom colors
sns.barplot(y="Class", x="Orbit", data=ob)

# Set the labels
plt.xlabel("Orbit", fontsize=20)
plt.ylabel("Class", fontsize=20)

# Show the plot
plt.show()
```



Flight Number vs. Orbit Type

- a scatter point of Flight number vs. Orbit type

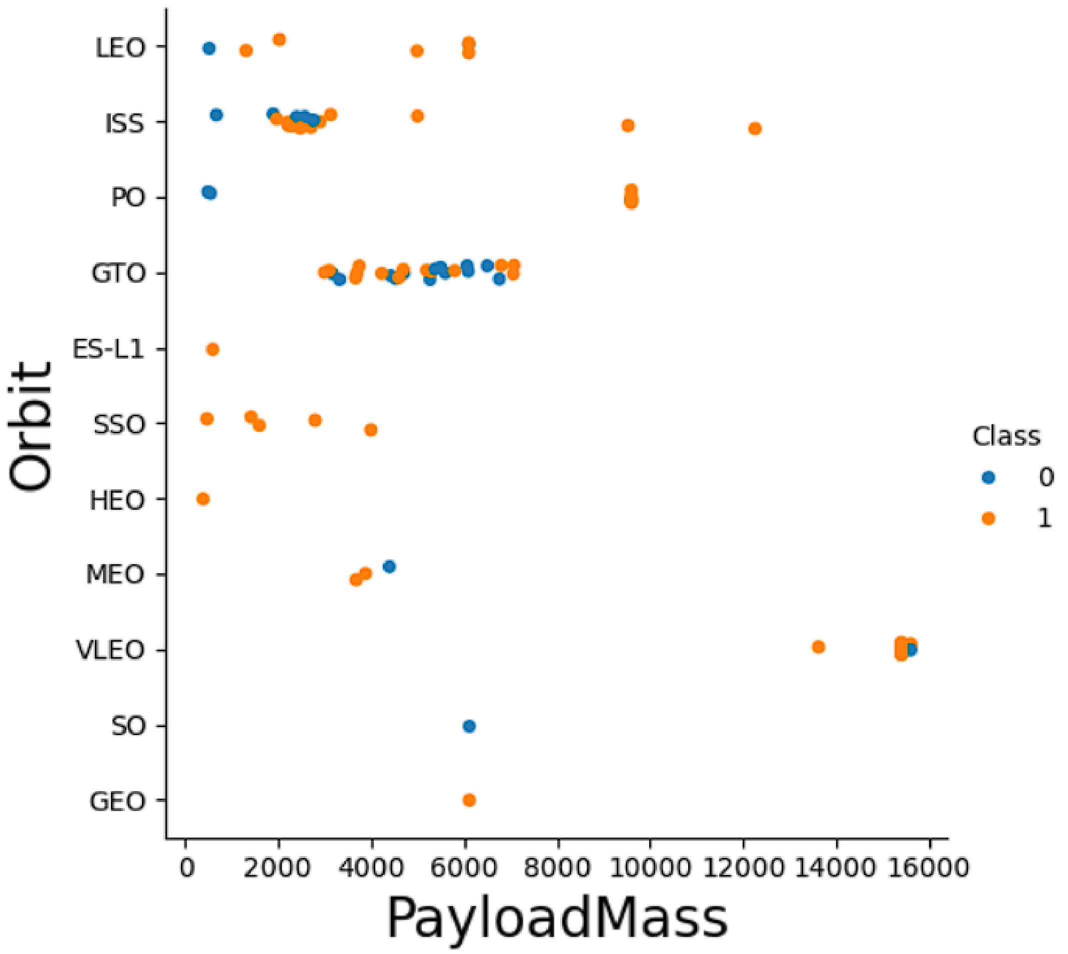
```
sns.catplot(x="FlightNumber",y="Orbit",hue="Class",data=df)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



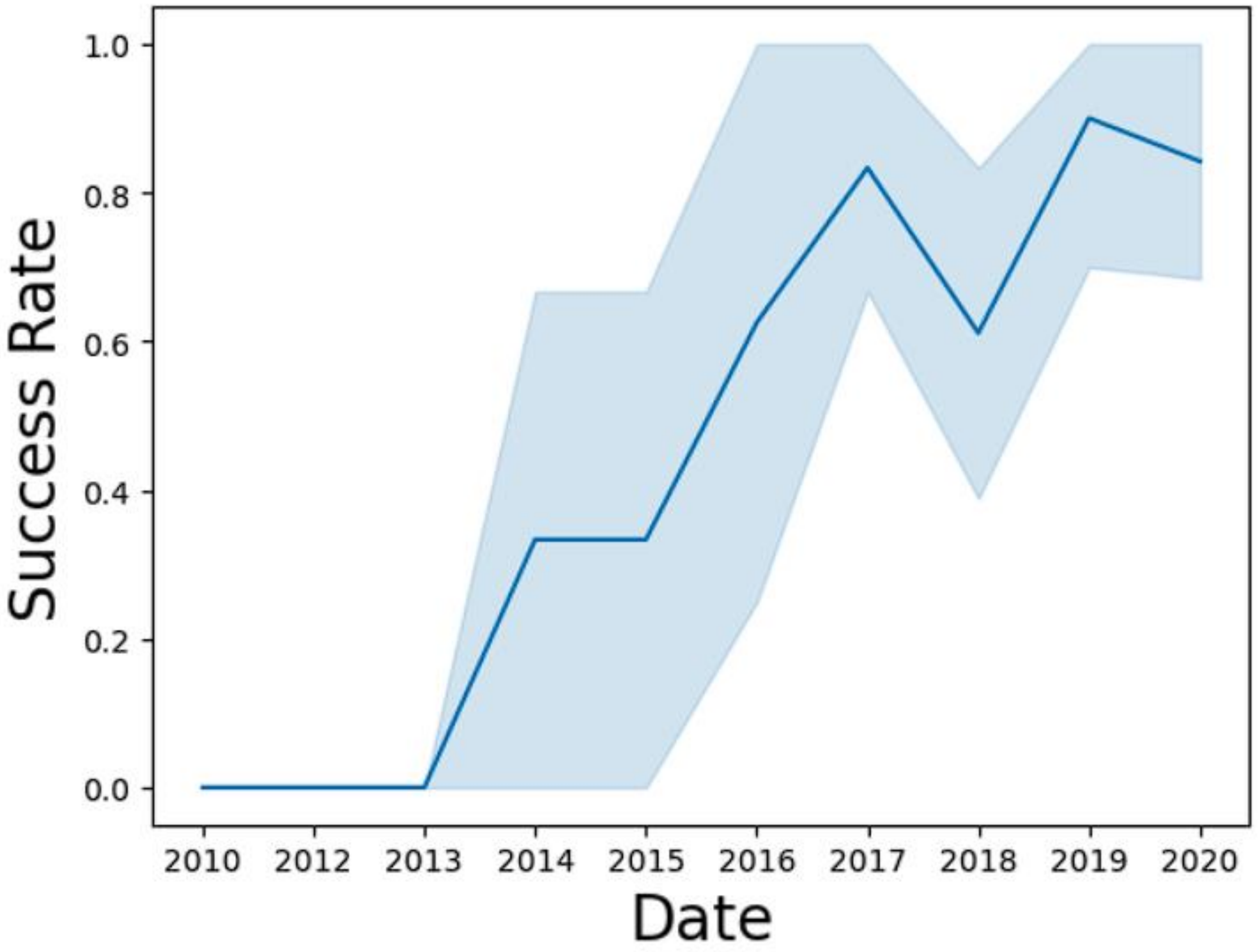
Payload vs. Orbit Type

- a scatter point of payload vs. orbit type

```
sns.catplot(x="PayloadMass",y="Orbit",hue="Class",data=df)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



Launch Success Yearly Trend



All Launch Site Names

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

In [9]:
```
%sql Select DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

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

Out[9]:

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

Launch Site Names Begin with 'CCA'

Task 2

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

```
%sql SELECT * FROM SPACEXTBL 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_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

These are 5 records where launch sites begin with the letters 'CCA'. As we can see, there are other organizations besides Space X that were testing their rockets.

Total Payload Mass

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

In [23]:

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA(CRS)';
```

```
 * ibm_db_sa://gfd86828:***@3883e7e4-18f5-4afe-be8c-fa31c41761d2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31498/bludb
Done.
```

Out[23]:

1

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version='F9 v1.1';
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql SELECT min(Date) FROM SPACEXTBL Where Landing_Outcome='Success (ground pad)';
```

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

min(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

In [14]:
```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE 4000<PAYLOAD_MASS__KG_ <6000 AND Landing_Outcome='Success (drone ship)';
```

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

Out[14]:

Booster_Version
F9 FT B1021.1
F9 FT B1022
F9 FT B1023.1
F9 FT B1026
F9 FT B1029.1
F9 FT B1021.2
F9 FT B1029.2
F9 FT B1036.1
F9 FT B1038.1
F9 B4 B1041.1
F9 FT B1031.2
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [15]:
```
%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight
```

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

Out[15]:

count(MISSION_OUTCOME)
99

Boosters Carried Maximum Payload

Task 8

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

```
%sql SELECT Booster_Version from SPACEXTBL WHERE PAYLOAD_MASS__KG_=(SELECT max(PAYLOAD_MASS__KG_)From SPACEXTBL);
```

```
 * 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

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

Out[37]:

month_name	Mission_Outcome	Booster_Version	Launch_Site
January	Success	F9 v1.1 B1012	CCAFS LC-40
February	Success	F9 v1.1 B1013	CCAFS LC-40
March	Success	F9 v1.1 B1014	CCAFS LC-40
April	Success	F9 v1.1 B1015	CCAFS LC-40
April	Success	F9 v1.1 B1016	CCAFS LC-40
June	Failure (in flight)	F9 v1.1 B1018	CCAFS LC-40
December	Success	F9 FT B1019	CCAFS LC-40

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

Present your query result with a short explanation here

```
%%sql
select *
from SPACEXTBL
where
   (Landing_Outcome = 'Success (ground pad)' OR 'Failure (drone Ship)')
    AND Date BETWEEN '2010-06-04' AND '2017-03-20' order by Date desc
```

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

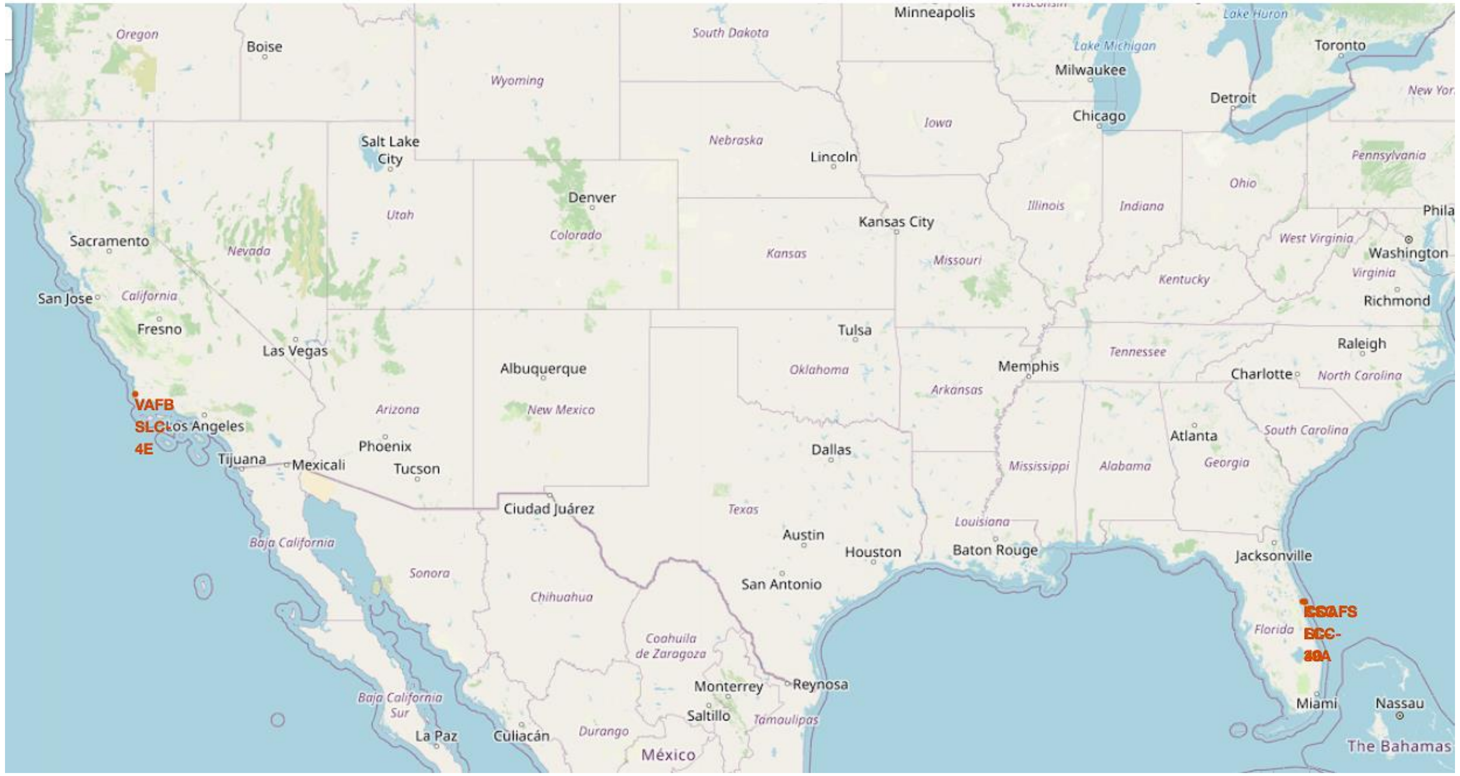
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 3

Launch Sites Proximities Analysis

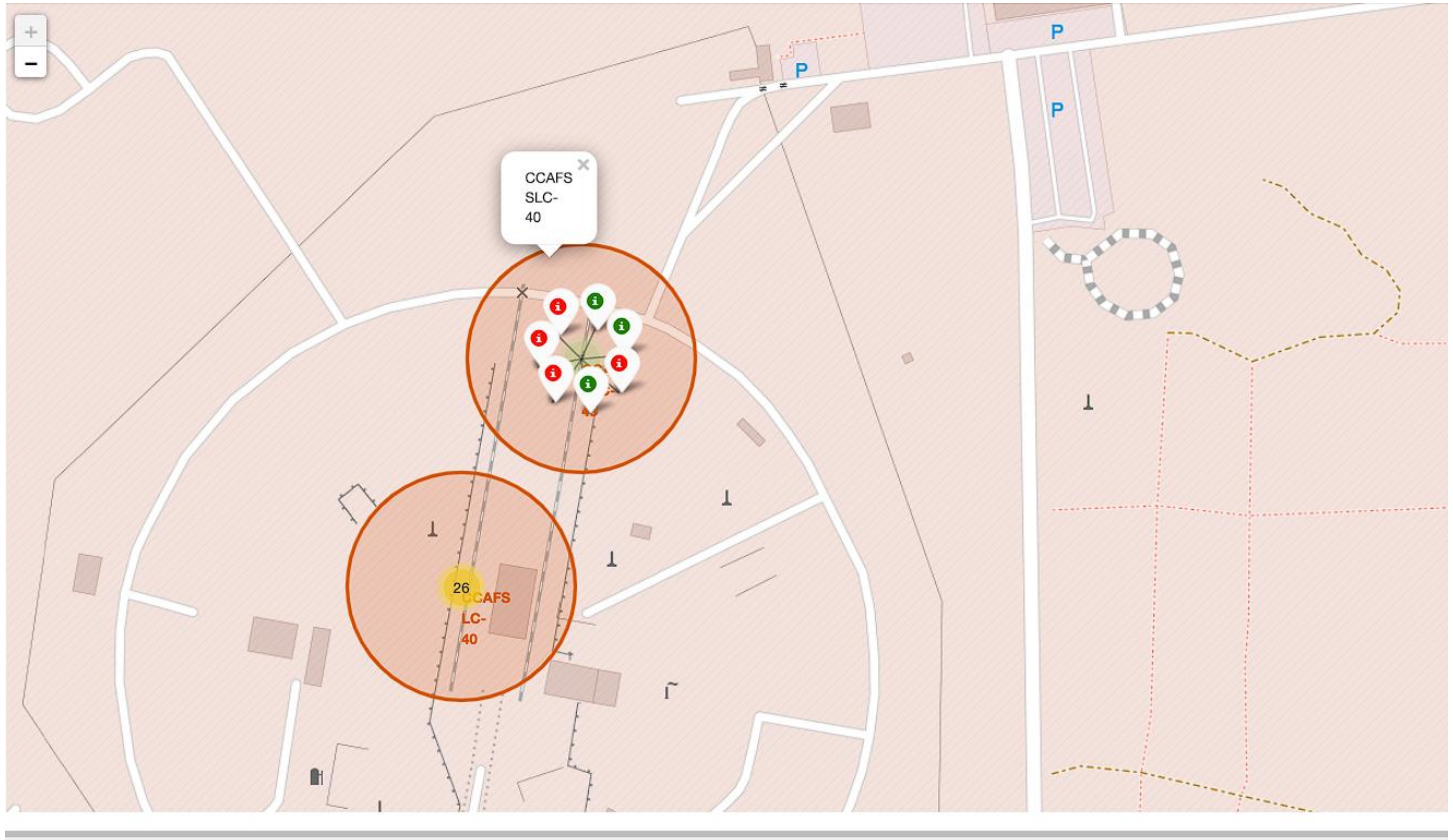
Launch site locations

- Launch sites are located with the couple kms from the coastlines



Success rate of launches

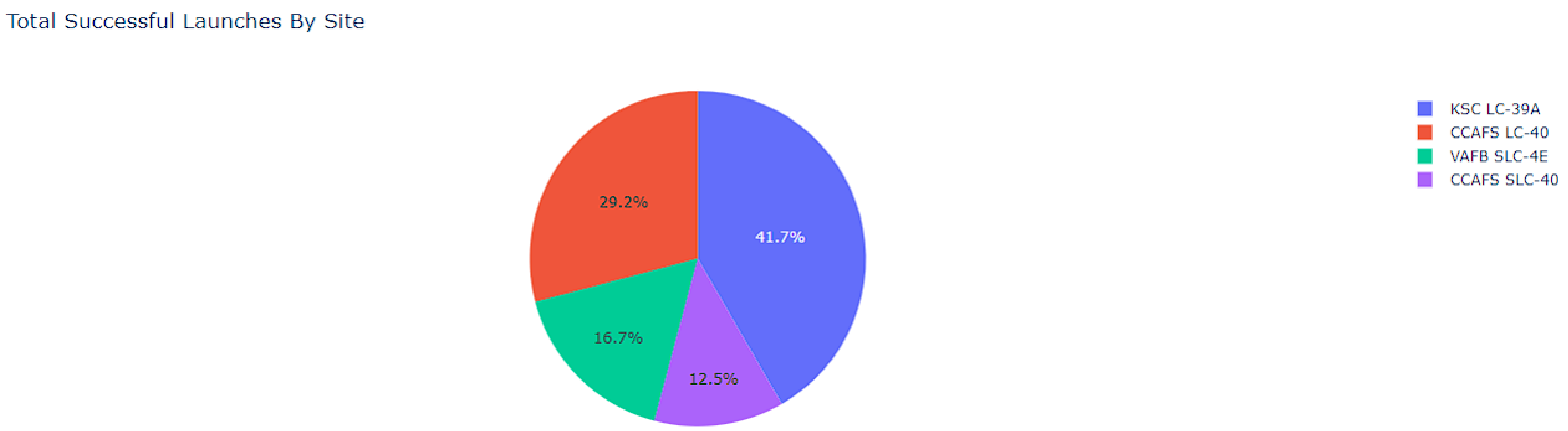
- Green represents successful launches, red means failures



The distance between launch site to its approximates

Section 4

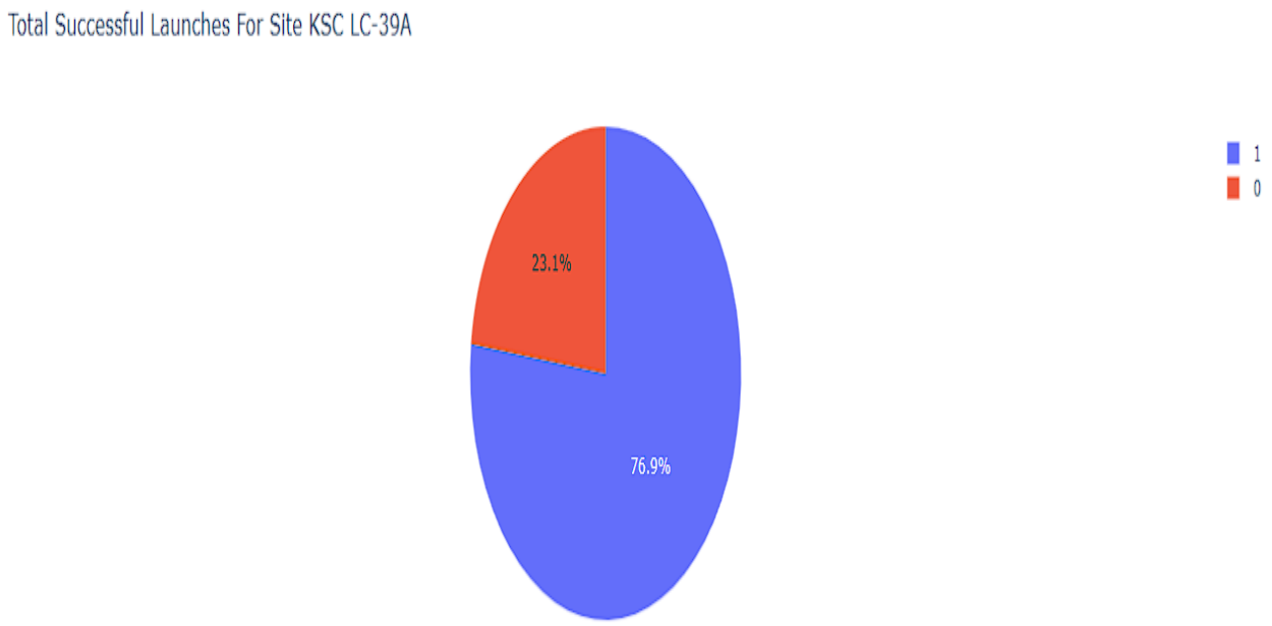
Build a Dashboard with Plotly Dash



Successful launches by site

- KSC LC-39A makes up majority of successful launches

Total launches breakdown of a specific site

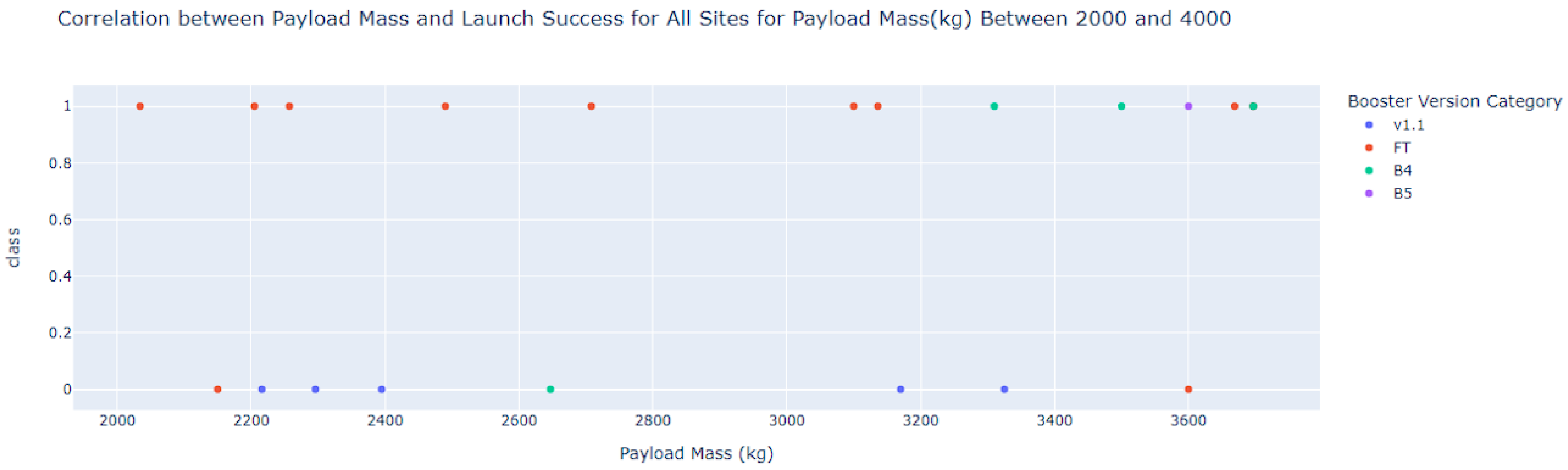


1- mean success

0- mean failure

Correlation between Payload mass and launch success for sites payload mass between 2000 and 4000

- Payload mass range between 2000 and 4000 has the highest successrate



Section 5

Predictive Analysis (Classification)

Classification Accuracy

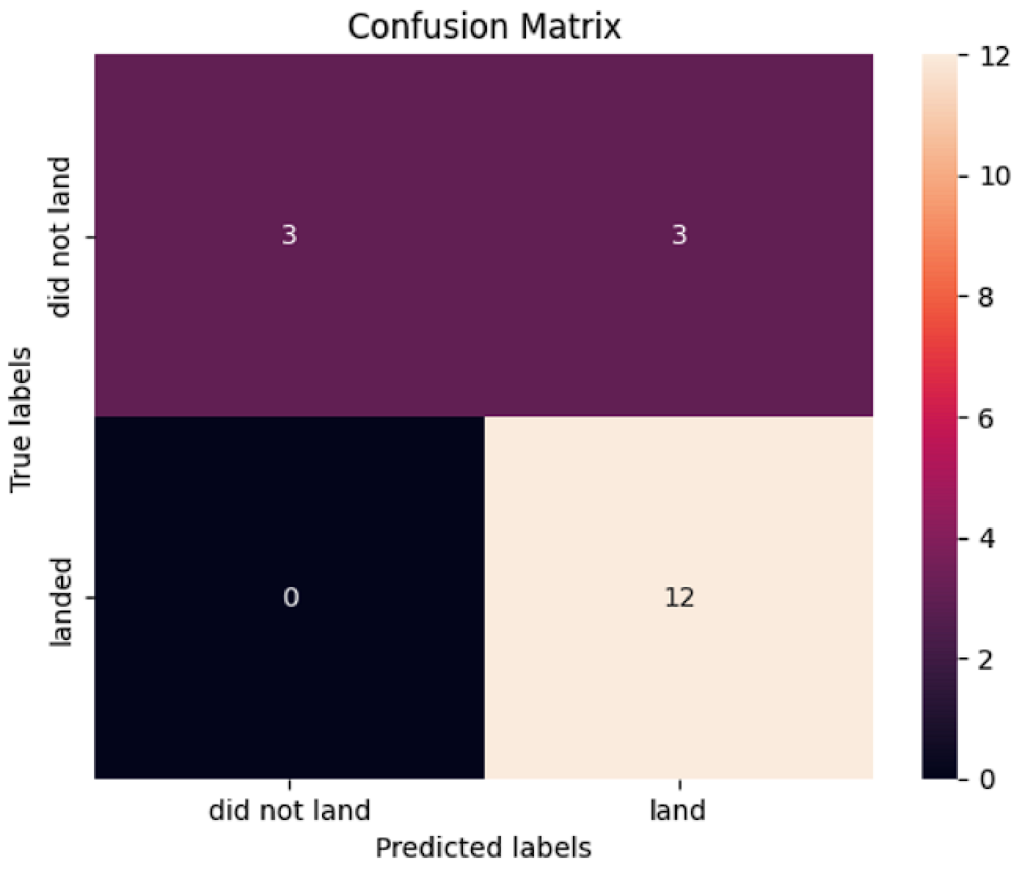
- All the methods have the same accuracy predictability hence doesn't make much difference which method is used

```
Accuracy=[svm_cv_score, logreg_score, knn_cv_scc
Accuracy=[i*100 for i in Accuracy]
methods=['Support Vector Machine', 'Logistic Reg
model={'ML method':methods,'Accuracy Score%':Acc
df=pd.DataFrame(model)
df
```

	ML method	Accuracy Score%
0	Support Vector Machine	83.333333
1	Logistic Regression	83.333333
2	K Nearest Neighbour	83.333333
3	Decision Tree	83.333333

Confusion Matrix

confusion matrix of any of the model is acceptable since they have same outcomes



Conclusions

- CCAFS SLC 40 is the best launch site for the payload between 12500 kg and 15000kg with KSC LC39A coming in second
- VAFB SLC4E is the best launch site for payload between 7500kg and 10000kg
- Orbit types ESL-1,GEO,HEO and SSO have the 100 percent success rate
- KSC LC 39-A makes up the majority of the successful launches
- No matter the type of machine learning model that we use, we can predict the outcome 83.33% accurately

Thank you!