SpaceX Launch Prediction – Capstone Project

Jitesh Bhola

July 11,2023

Introduction

- In this course, the successful landing of Falcon 9 – first stage was predicted.
- SpaceX advertises the launch of Falcon 9 with price much less than other competitors due to the ability to reuse the first stage.
- Therefore, prediction of the first stage landing can help to determine the entire launch cost.
- This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

METHODOLOGY

- Data collection and Data Wrangling :
  - Falcon 9 Launch data will be collected from Wiki pages using web scraping.
  - Another way of gathering data is SpaceX REST API.
  - Raw data from table would be transformed to clean data.
  - Functions such as Booster, Launchpad, payload, and core will be used to deal with null values etc. to get actual values.

METHODOLOGY

- Data Visualization:
  - Use Catplot to visualize the data by choosing different groups of data
  - For example, compare between payload mass and flight number to see if successful
  - launch is dependent on the payload mass

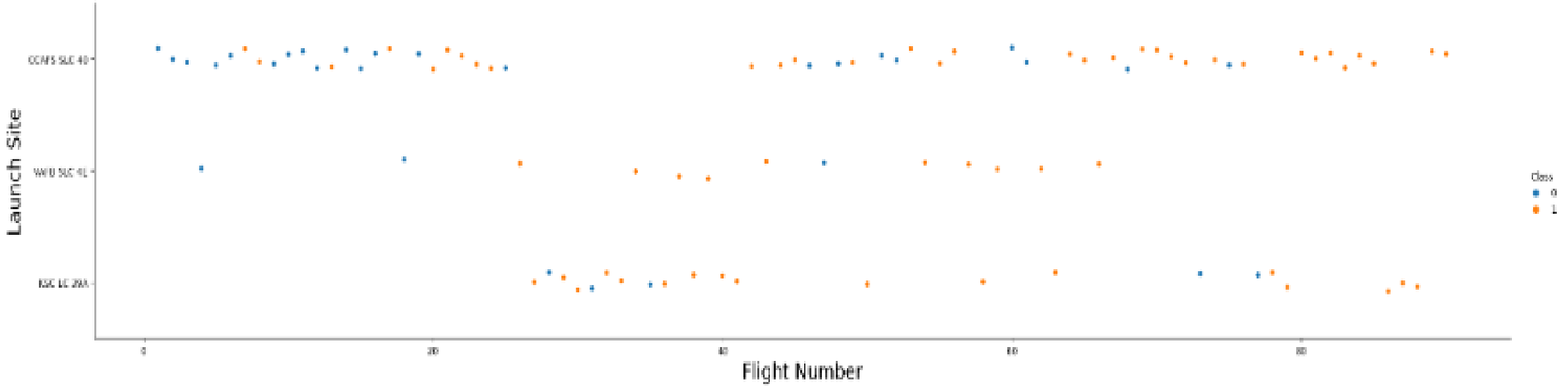
METHODOLOGY

- Predictive Analysis
  - Using machine learning methods to predict the launch
  - First need to standardize the dataset
  - Split into train and test sets
  - Use different methods to train the datasets and compare the accuracy

INSIGHTS FROM EDA

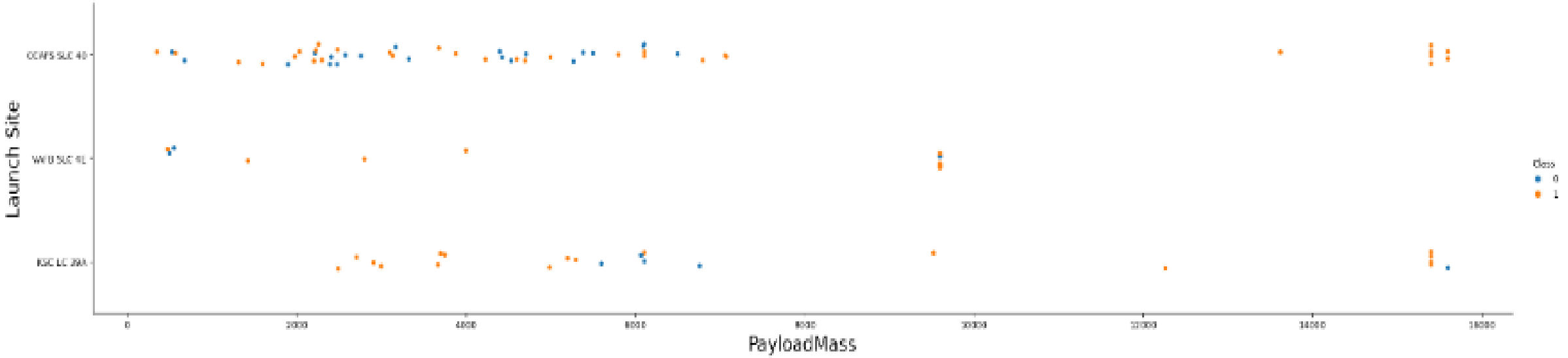
FLIGHT NUMBER VS. LAUNCH SITE

[6]:
```
### 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()
```



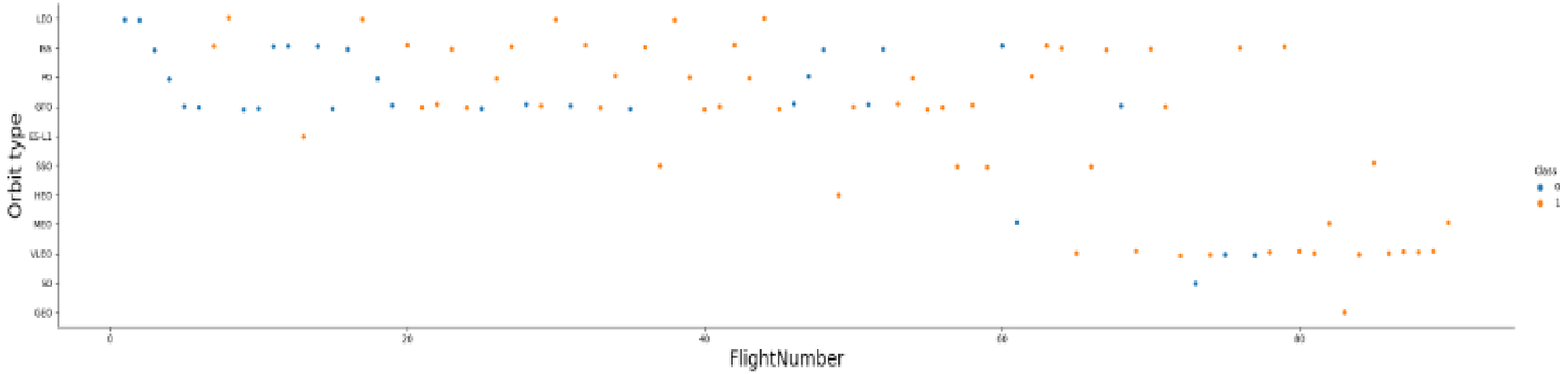
PAYLOAD VS. LAUNCH SITE

[7]:
```
### TASK 2: Visualize the relationship between Payload and Launch Site
sns.catplot(y="LaunchSite", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("PayloadMass",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



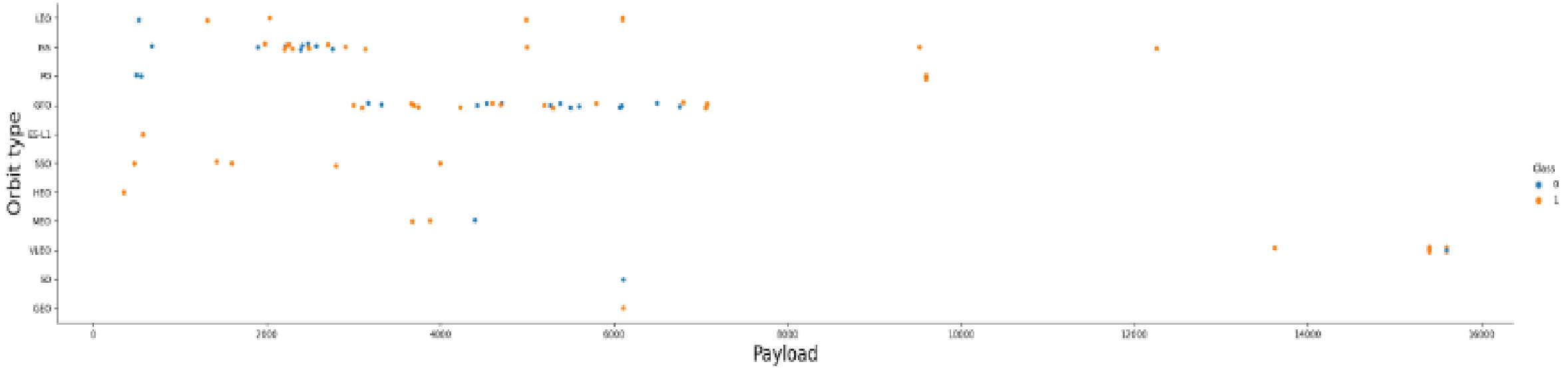
FLIGHT NUMBER VS. ORBIT TYPE

[8]:
```
### TASK  3: Visualize the relationship between success rate of each orbit type
sns.catplot(x='FlightNumber', y = 'Orbit', hue = 'Class', data = df, aspect = 5)
plt.xlabel("FlightNumber",fontsize=20)
plt.ylabel("Orbit type",fontsize=20)
plt.show()
```

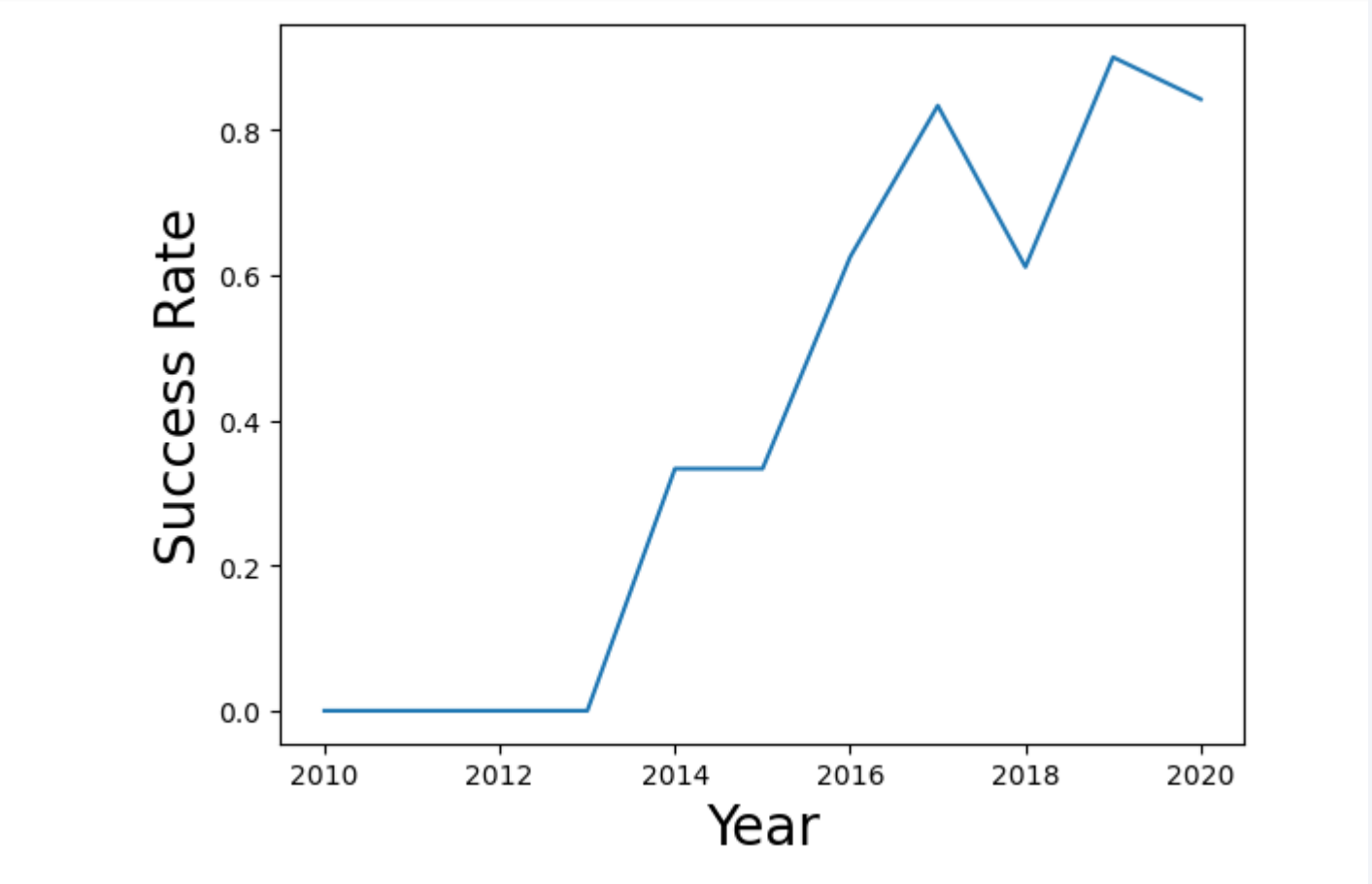


PAYLOAD VS. ORBIT TYPE

[9]:
```
### TASK  5: Visualize the relationship between Payload and Orbit type
sns.catplot(x='PayloadMass', y = 'Orbit', hue = 'Class', data = df, aspect = 5)
plt.xlabel("Payload",fontsize=20)
plt.ylabel("Orbit type",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 [19]: %sql select distinct(Launch_Site) from SPACEXTBL
```

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

Out[19]:

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'

```
In [23]: %sql select Launch_Site from SPACEXTBL where Launch_Site like "CCA%" limit 5
```

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

Out[23]:

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

TOTAL PAYLOAD MASS

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

In [31]:
```
%sql select sum(PAYLOAD_MASS__KG_) as "PAYLOAD_NASA(CRS)" from SPACEXTBL where Customer="NASA (CRS)"
```

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

Out[31]:

PAYLOAD_NASA(CRS)
45596.0

AVERAGE PAYLOAD MASS BY F9 V1.1

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

In [33]:
```
%sql select avg(PAYLOAD_MASS__KG_) as "PAYLOAD_F9" from SPACEXTBL where Booster_Version="F9 v1.1"
```

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

Out[33]:

PAYLOAD_F9
2928.4

FIRST SUCCESSFUL GROUND LANDING DATE

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

Hint:Use min function

In [41]:
```
%sql select min(Date) from SPACEXTBL where Landing_Outcome="Success (ground pad)"
```

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

Out[41]:

min(Date)
01/08/2018

SUCCESSFUL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

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

In [47]:
```
%sql select Booster_Version from SPACEXTBL where Landing_Outcome="Success (drone ship)" and PAYLOAD_MASS__KG_ between 4
```

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

Out[47]:

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

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

[34]:
```
%sql SELECT Count(Mission_Outcome), Mission_Outcome FROM SPACEXTBL group by Mission_Outcome
```

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

[34]:

Count(Mission_Outcome)	Mission_Outcome
0	None
1	Failure (in flight)
98	Success
1	Success
1	Success (payload status unclear)

BOOSTERS CARRIED MAXIMUM PAYLOAD

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

[44]:
```
%sql select Booster_Version,PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_)\
from SPACEXTBL)
```

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

[44]:

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600.0
F9 B5 B1049.4	15600.0
F9 B5 B1051.3	15600.0
F9 B5 B1056.4	15600.0
F9 B5 B1048.5	15600.0
F9 B5 B1051.4	15600.0
F9 B5 B1049.5	15600.0
F9 B5 B1060.2	15600.0
F9 B5 B1058.3	15600.0
F9 B5 B1051.6	15600.0
F9 B5 B1060.3	15600.0
F9 B5 B1049.7	15600.0

2015 LAUNCH RECORDS

[45]:
```
%sql select substr(Date, 4, 2) as Month,Landing_Outcome,Booster_Version,Launch_Site from SPACEXTBL\
where substr(Date,7,4)='2015' and Landing_Outcome =="Failure (drone ship)"
```

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

[45]:

Month	Landing_Outcome	Booster_Version	Launch_Site
10	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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

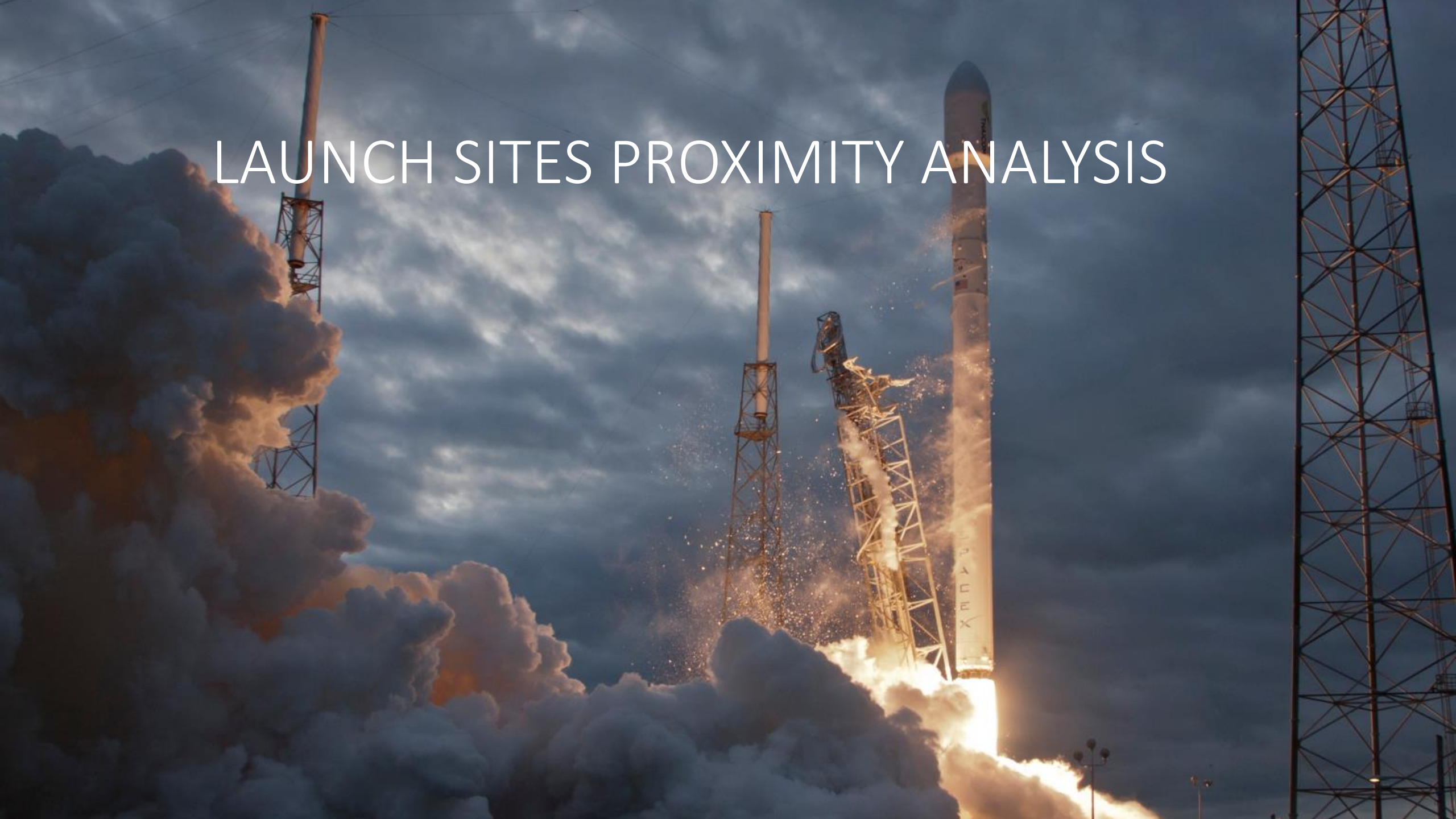
In [89]:
```
%sql SELECT COUNT("Landing _Outcome") FROM SPACEXTBL WHERE ("Landing _Outcome" LIKE 'Success%') GROUP BY "Landing _Outcome
```

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

Out[89]:

COUNT("Landing _Outcome")
38
14
9

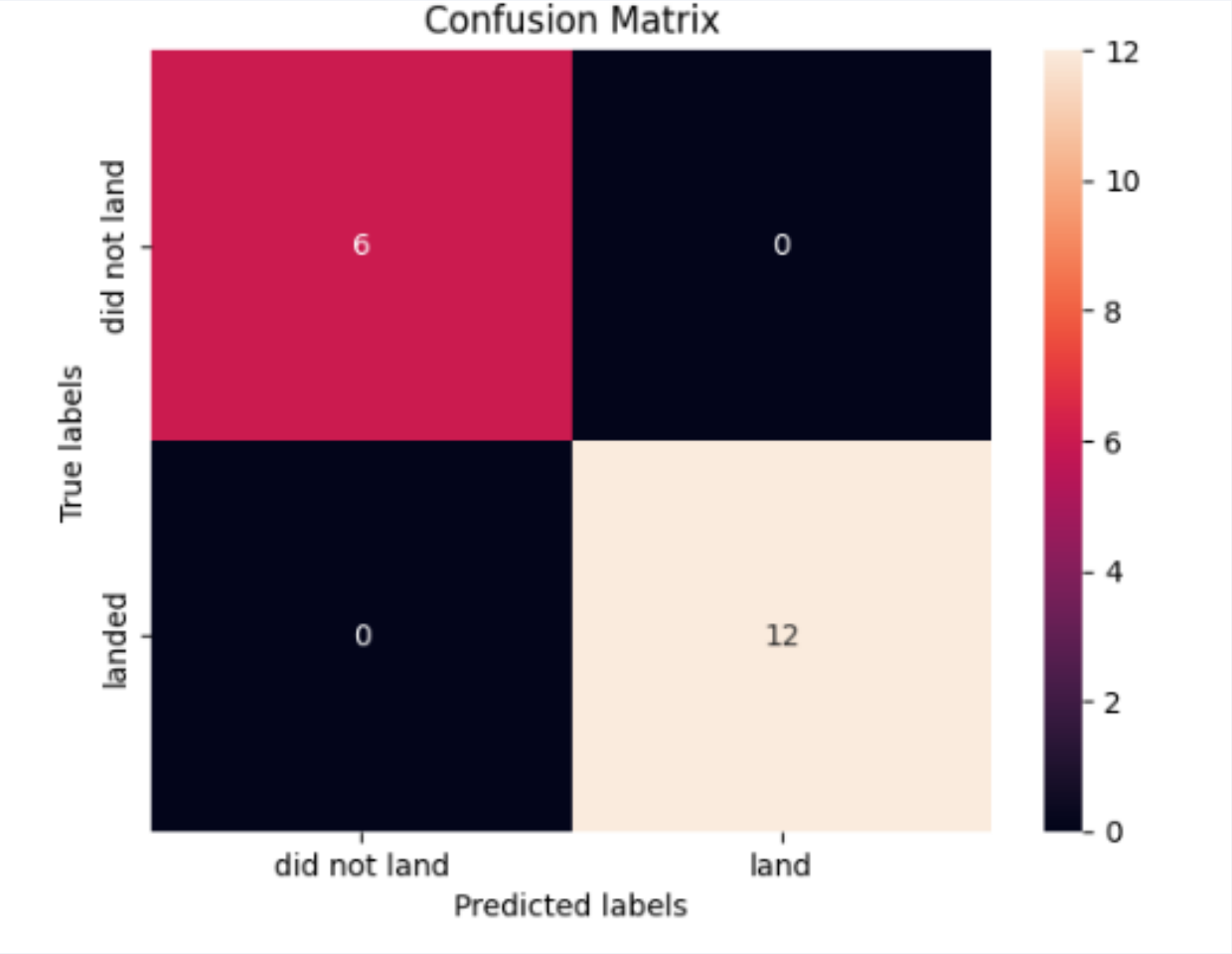
LAUNCH SITES PROXIMITY ANALYSIS



FOLIUM MAP WITH MARKERS OF LANDING SITES

PREDICTIVE ANALYSIS CLASSIFICATION

CONFUSION MATRIX



CONCLUSIONS

- Accuracy for Logistics Regression method: 0.8333333333333334
- Accuracy for Support Vector Machine method: 0.8333333333333334
- Accuracy for Decision tree method: 0.7777777777777778
- Accuracy for K nearest neighbors method: 0.8333333333333334


THANKYOU