IBM Data Science Project
* SPACE EX*

Frederick Duff

Table of Contents

I. Executive Summary

- Launching a Falcon 9 rocket costs an estimated $67 million, with the booster making up 60% of the cost. However, since the booster is recoverable, the cost can be reduced significantly if it's reused for subsequent launches. The upper stage, fairing, and launch costs make up the remaining 40%. Due to the need for the booster to retain enough fuel for landing, there's a payload reduction, but this can be compensated for by recovering the booster multiple times. To estimate the cost of a launch, it's essential to predict whether a booster will be recovered.
- In summary, the Falcon 9's relatively low cost is due in large part to SpaceX's ability to recover and reuse the booster. By doing so, SpaceX can significantly reduce the overall cost of launching a rocket. Additionally, the ability to predict whether the booster will be reusable for future launches is crucial in creating a proforma and estimating the cost of a launch. Overall, understanding the costs associated with launching rockets is crucial in developing and supporting new rocket companies that can compete with industry leaders like SpaceX.
- Overall, this project aims to provide insight into the cost-effectiveness of rocket launches and support the development of new rocket companies like Space Y that can compete with industry leaders like SpaceX.

Introductory Slide

- Fundamentally if we can determine a predictability to the ability to recover stages of the Falcoln 9 , then we can develop a competitive business model . Further if we are able to determine factors which enhance recovery then we may be able to alter the curve and develop a superior model for competitiveness and efficiency.

Data collection and data wrangling

- Imported data from the SpaceX API-

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

- Data wrangling- checked for null (missingness) we used the mean value to substitute

```
# Calculate the mean value of PayloadMass column
mean_payload = data_falcon9['PayloadMass'].mean()
# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'].replace(np.nan, mean_payload, inplace=True)
```

Exploratory Data Analysis

- 1. Determined the number of launches from each site:

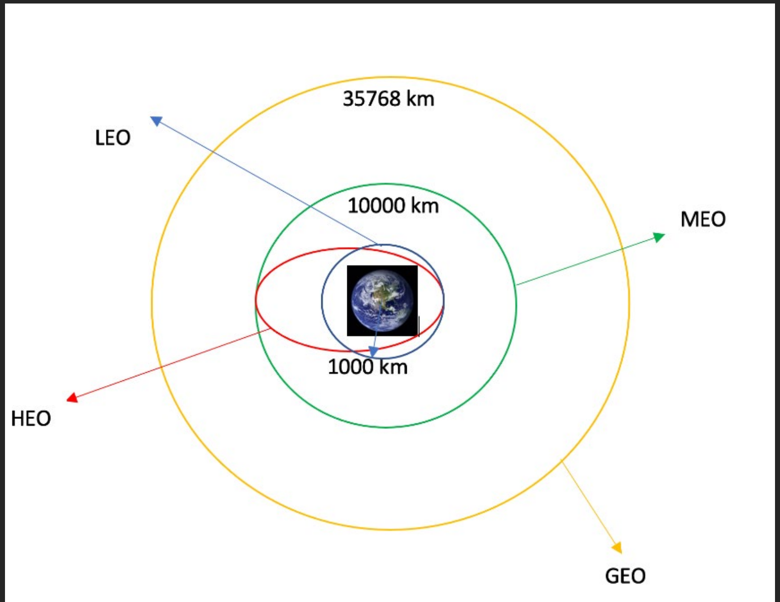
  CCAFS SLC 40 (Cape Canaveral) – 55

  KSC LC 39a (Kennedy Space Center complex 39a) – 22

  VAFB SLC 4E (Vandenberg Air Force Base) - 13

Exploratory Data Analysis – Orbit types

- **LEO**: Low Earth orbit (LEO)is an Earth-centred orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth),[1] or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25.[2] Most of the manmade objects in outer space are in LEO [1].
- **VLEO**: Very Low Earth Orbits (VLEO) can be defined as the orbits with a mean altitude below 450 km. Operating in these orbits can provide a number of benefits to Earth observation spacecraft as the spacecraft operates closer to the observation[2].
- **GTO** A geosynchronous orbit is a high Earth orbit that allows satellites to match Earth's rotation. Located at 22,236 miles (35,786 kilometers) above Earth's equator, this position is a valuable spot for monitoring weather, communications and surveillance. Because the satellite orbits at the same speed that the Earth is turning, the satellite seems to stay in place over a single longitude, though it may drift north to south," NASA wrote on its Earth Observatory website [3] .
- **SSO (or SO)**: It is a Sun-synchronous orbit also called a heliosynchronous orbit is a nearly polar orbit around a planet, in which the satellite passes over any given point of the planet's surface at the same local mean solar time [4] .
- **ES-L1** :At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies. L1 is one such point between the sun and the earth [5] .
- **HEO** A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth [6].
- **ISS** A modular space station (habitable artificial satellite) in low Earth orbit. It is a multinational collaborative project between five participating space agencies: NASA (United States), Roscosmos (Russia), JAXA (Japan), ESA (Europe), and CSA (Canada) [7]
- **MEO** Geocentric orbits ranging in altitude from 2,000 km (1,200 mi) to just below geosynchronous orbit at 35,786 kilometers (22,236 mi). Also known as an intermediate circular orbit. These are "most commonly at 20,200 kilometers (12,600 mi), or 20,650 kilometers (12,830 mi), with an orbital period of 12 hours [8]
- **HEO** Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi) [9]
- **GEO** It is a circular geosynchronous orbit 35,786 kilometres (22,236 miles) above Earth's equator and following the direction of Earth's rotation [10]
- **PO** It is one type of satellites in which a satellite passes above or nearly above both poles of the body being orbited (usually a planet such as the Earth [11]



Exploratory Data Analysis (cont)

- 2 – Calculate the number and occurrence of each orbit
- - geosynchronous orbit led the number of types of orbit with a total count of 27

```
In [5]:
# Apply value_counts on Orbit column
orbit_count = df['Orbit'].value_counts()
print(orbit_count)
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit, dtype: int64
```

Exploratory Data Analysis –(cont)

- 3. Number of occurrences of mission outcomes of the orbits

True	ASDS	41
None	None	19
True	RTLS	14
False	ASDS	6
True	Ocean	5
False	Ocean	2
None	ASDS	2
False	RTLS	1

- *True Ocean – successful landing in a region in the ocean- 5*
- *False Ocean- unsuccessful landing in a region in the ocean- 2*
- *True RTLS- mission outcome successfully landed on a ground pad- 14*
- *False RTLS- unsuccessful landing on the ground pad- 1*
- *True ASDS – mission outcome was successful landing on a drone ship- 41 observations*
- *False ASDS – mission outcome was unsuccessful landing on a drone ship- 6*

Exploratory Data Analysis (cont)

66% of landings were successful , meaning we have a lot of room for development and improvement

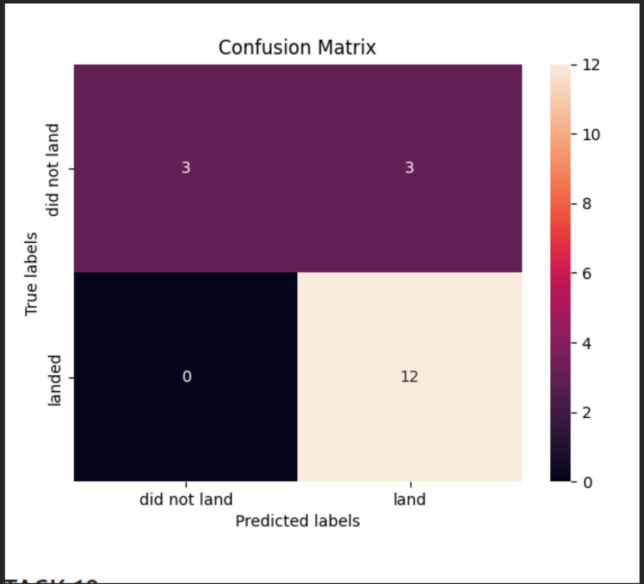
```
df["Class"].mean()
```

```
0.6666666666666666
```

Predictive Analytics

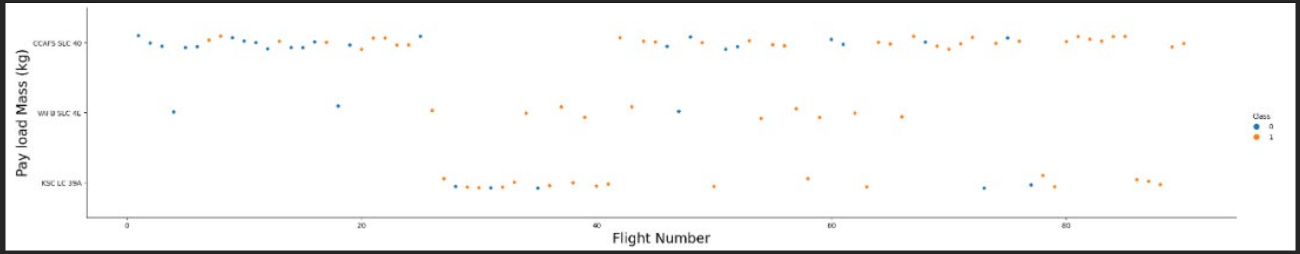
Decision Tree performed best with a F1 score of .94 and similar accuracy from the other models. Going forward In production we would want to pay close attention since Decision trees are notorious for overfitting

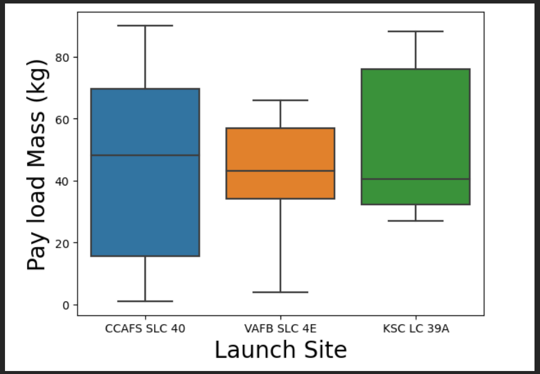
```
DecisionTreeClassifier
              precision    recall  f1-score   support

           0       1.00      0.83      0.91         6
           1       0.92      1.00      0.96        12

    accuracy                           0.94        18
   macro avg       0.96      0.92      0.93        18
weighted avg       0.95      0.94      0.94        18

KNeighborsClassifier
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

LogisticRegression
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18

SVC
              precision    recall  f1-score   support

           0       1.00      0.50      0.67         6
           1       0.80      1.00      0.89        12

    accuracy                           0.83        18
   macro avg       0.90      0.75      0.78        18
weighted avg       0.87      0.83      0.81        18
```



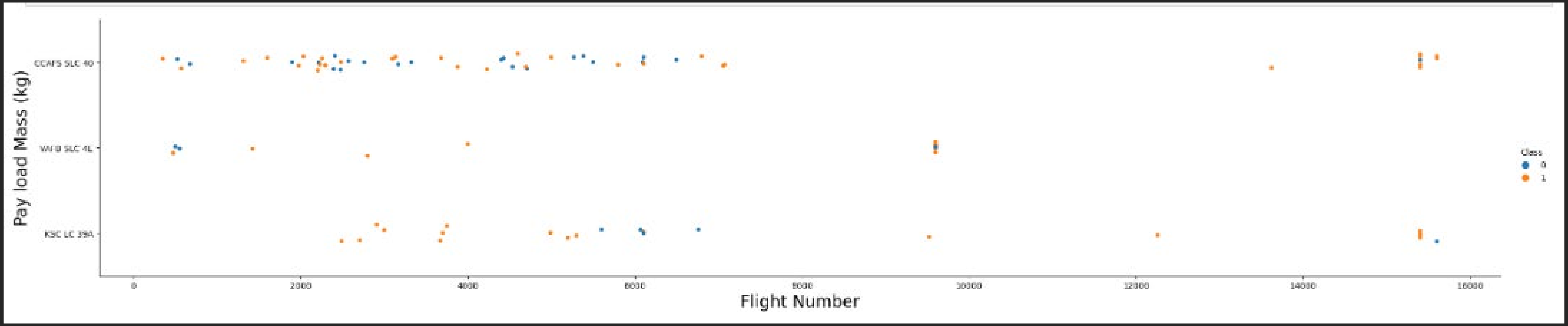
EDA with visualization results

CCAFS had a large range of payloads and on average were higher masses than the other launch sites

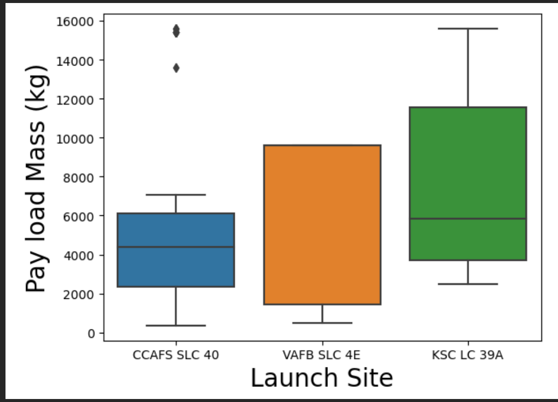




EDA with visualization results- cont

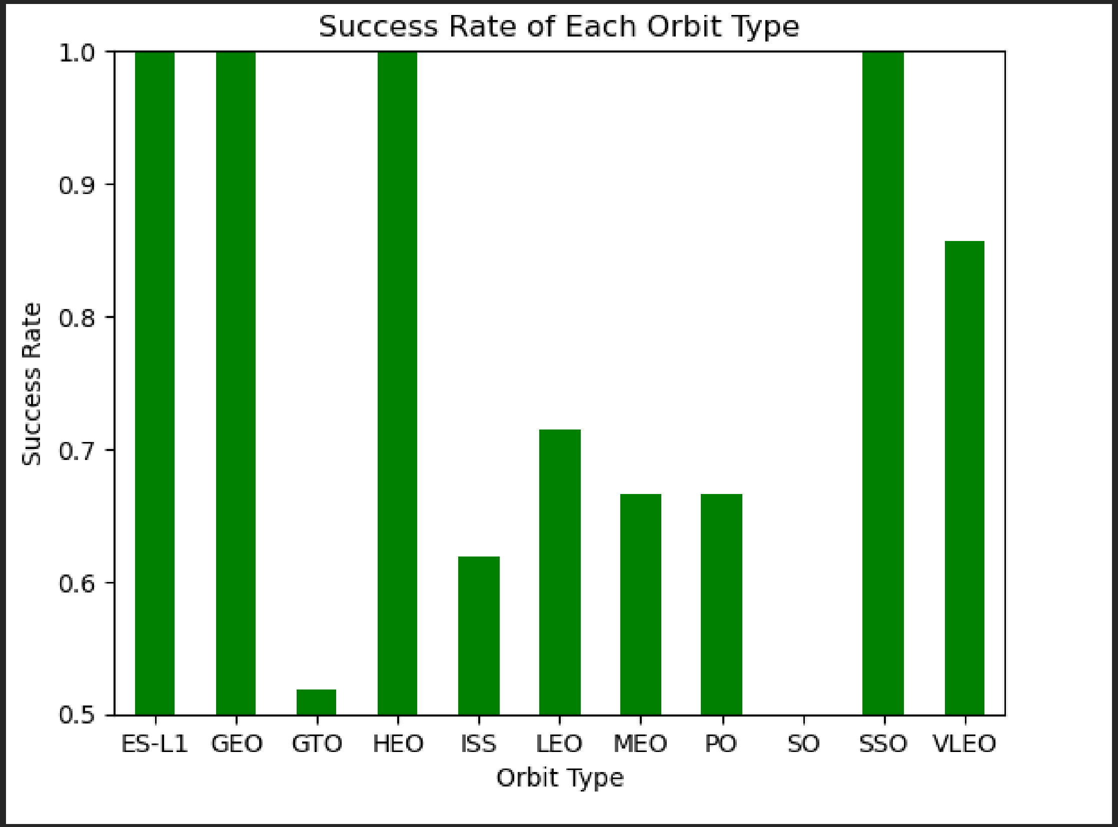


We can see that the payload mass was higher for KSC



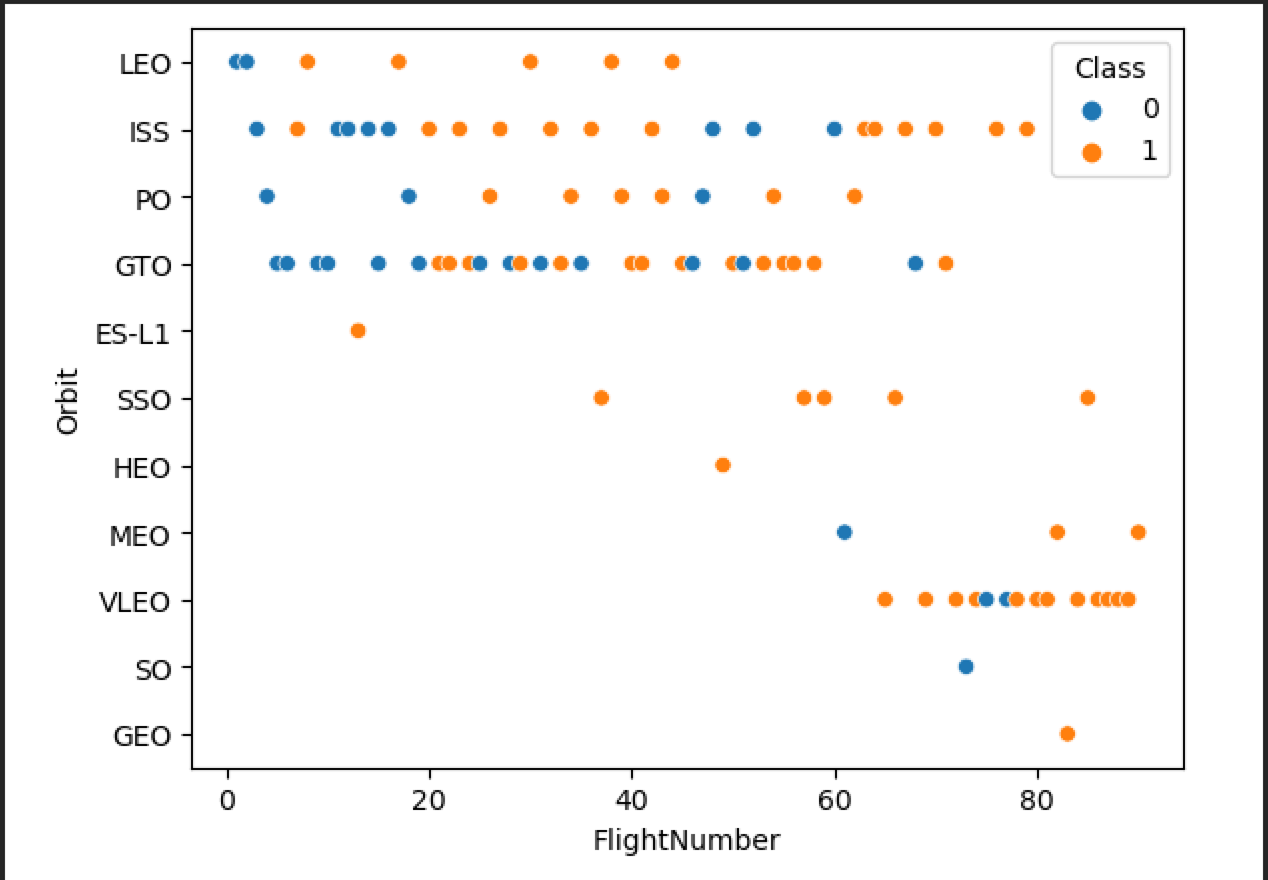
EDA with visualization results- cont

Types of orbits that were more successful
ES-L1, GEO , HEO and SSO



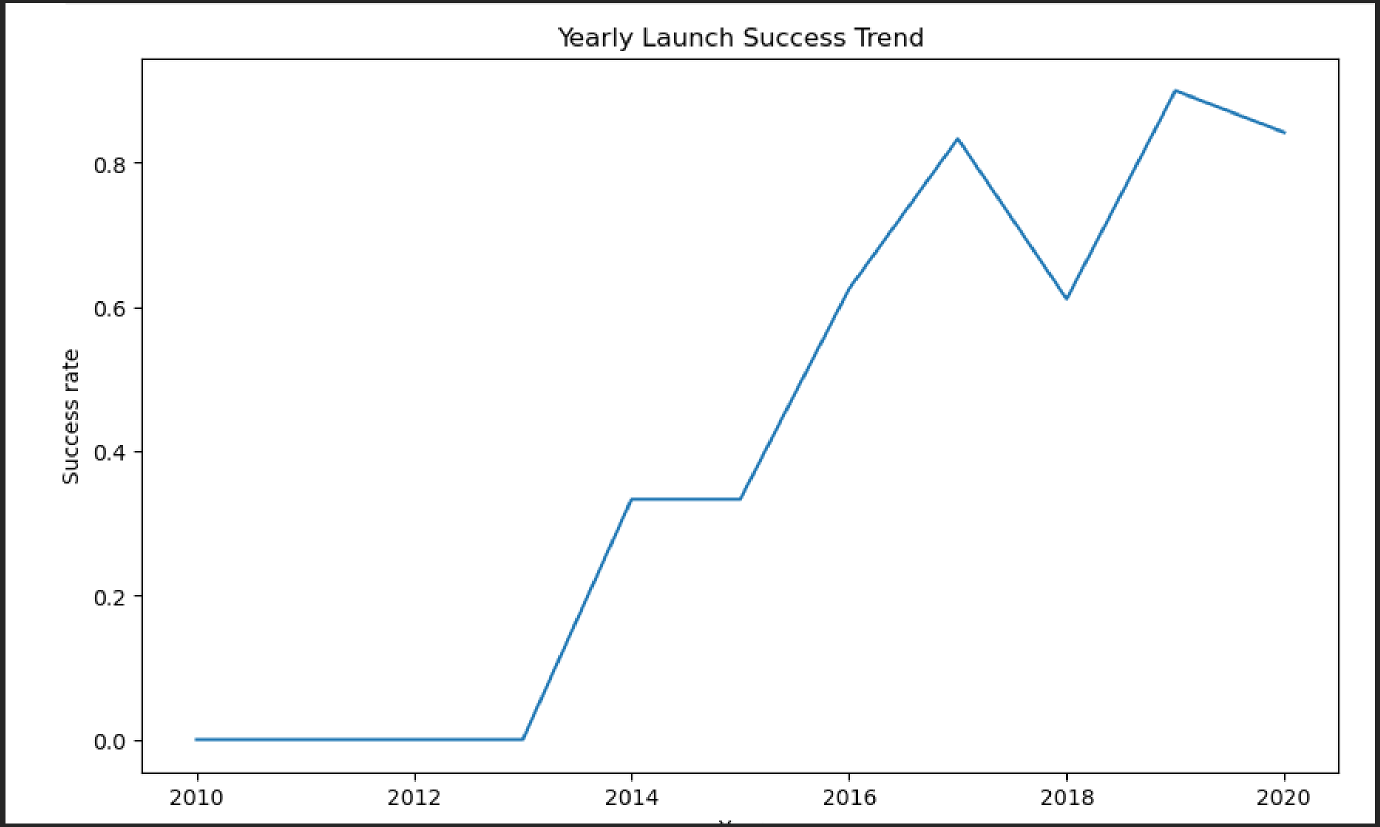
EDA with visualization results- cont

- Does not seem to be a relationship with the flight number and the orbit type



EDA with visualization

Clearly as development and testing occur the success has trended upward with the exception of 2018.



EDA with EDA- SQL

- Displayed the names of each unique launch site

Task 1

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

In [7]:

```
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTBL;
```

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

Out[7]:

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

EDA with EDA- SQL

Task 2

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

In [8]:
```
%sql SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA' LIMIT 5;
```

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

Out[8]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome

Task 3

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

In [9]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) as Total_Payload_Mass FROM SPACEXTBL WHERE Customer LIKE 'NASA (CRS)%';
```

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

Out[9]:

Total_Payload_Mass
48213

Task 4

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

In [10]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) as "Average Payload Mass (kg)" FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1';
```

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

Out[10]:

Average Payload Mass (kg)

EDA with EDA- SQL

Task 5

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

Hint:Use min function

In [11]:
```
%%sql SELECT MIN(Date) AS "Date of First Successful Landing on Ground Pad"
FROM SPACEXTBL
WHERE "Landing _Outcome" LIKE '%Success% (ground pad)';
```

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

Out[11]:

Date of First Successful Landing on Ground Pad
01-05-2017

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 [12]:
```
%%sql SELECT DISTINCT Booster_Version AS "Booster Name"
FROM SPACEXTBL
WHERE "Landing _Outcome" = "Success (drone ship)"
AND "Payload_mass__kg_" > 4000
AND "Payload_mass__kg_" < 6000;
```

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

Out[12]:

Booster Name
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Task 7

List the total number of successful and failure mission outcomes

In [13]:
```
%sql SELECT "Mission_Outcome", COUNT(*) as "Total Count" FROM SPACEXTBL GROUP BY "Mission_Outcome";
```

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

EDA with EDA- SQL

Task 8

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

In [14]:
```
%sql SELECT booster_version FROM SPACEXTBL WHERE payload_mass=(SELECT MAX(payload_mass) FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) no such column: payload_mass
[SQL: SELECT booster_version FROM SPACEXTBL WHERE payload_mass=(SELECT MAX(payload_mass) FROM SPACEXTBL);]
(Background on this error at: http://sqlalche.me/e/e3q8)
```

Task 9

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

In [15]:
```
%sql SELECT substr(Date,4,2) as Month, "Landing _Outcome", booster_version, launch_site FROM SPACEXTBL WHERE substr(Date,7,4)='2015' AND "Landing _Out
```

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

Out[15]:

Month	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

Task 10

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

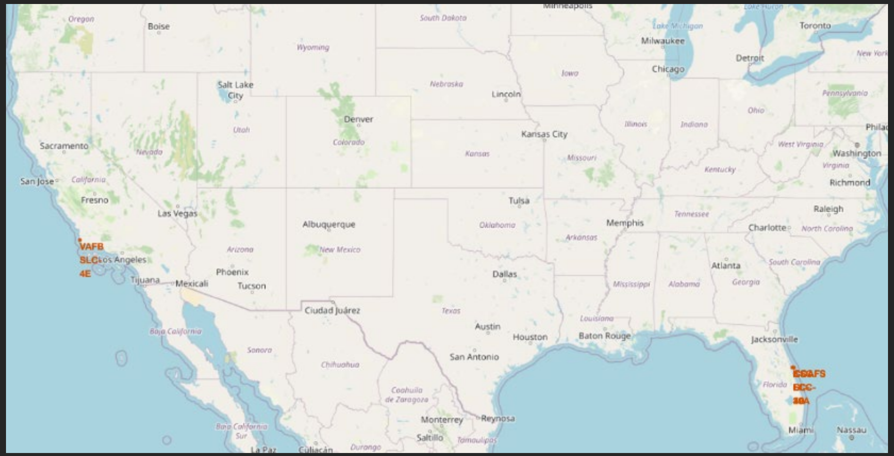
In [16]:
```
%sql SELECT "Landing _Outcome", COUNT(*) as "Total Count" FROM SPACEXTBL WHERE Date BETWEEN '04-06-2010' AND '20-03-2017' AND "Landing _Outcome" LIKE
```

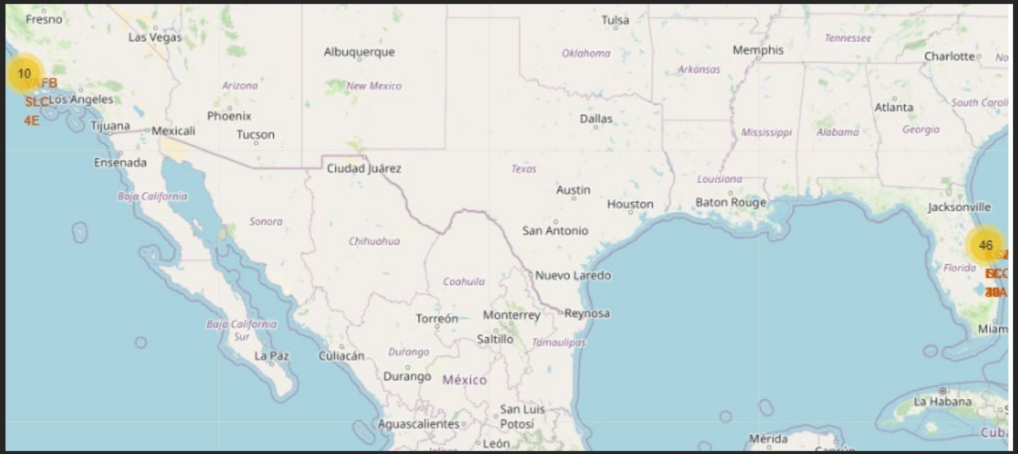
```
 * sqlite:///my_data1.db
Done.
```

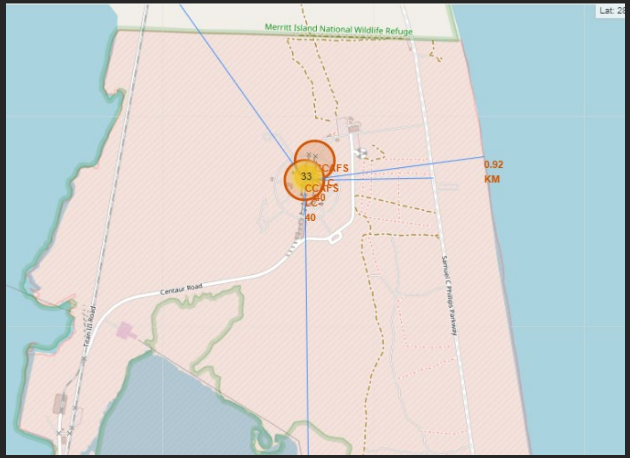
Out[16]:

Landing _Outcome	Total Count
Success	20
Success (drone ship)	8
Success (ground pad)	6

Interactive map with folium results







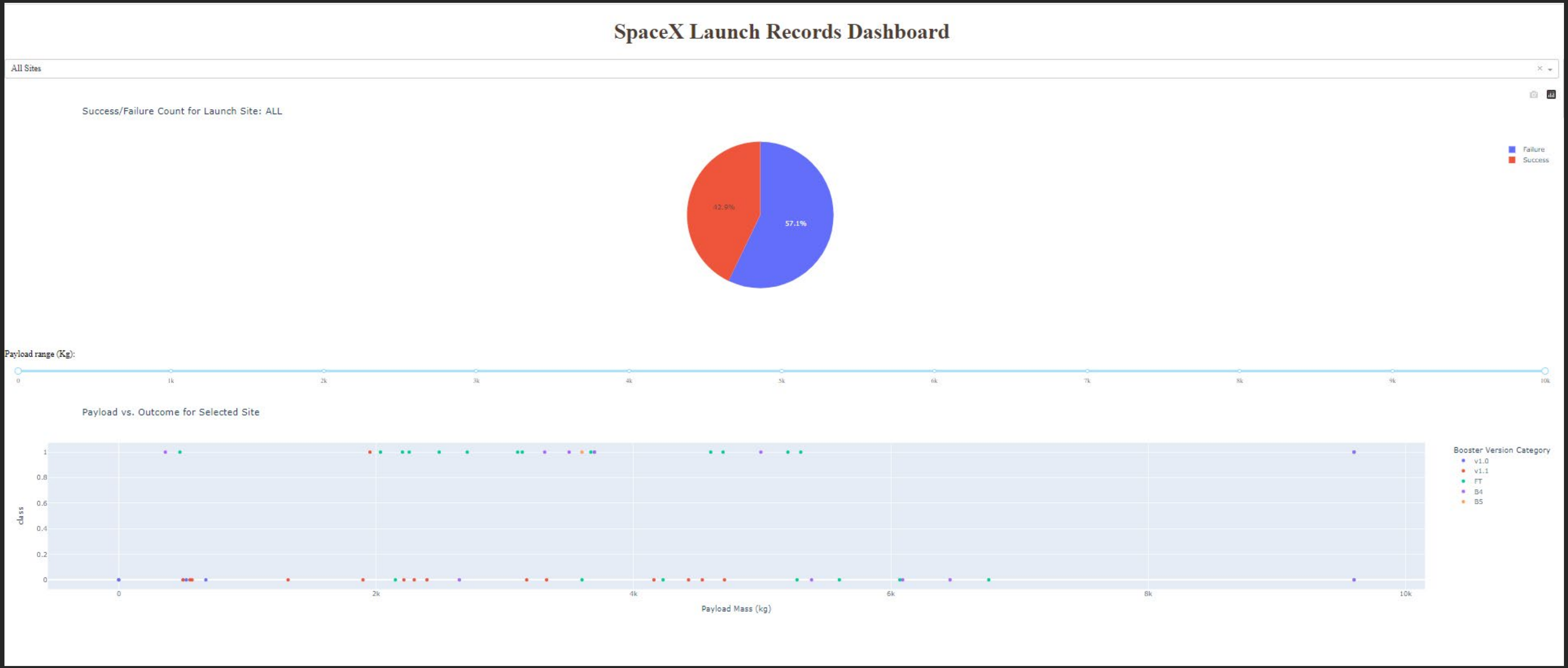
- Are launch sites in close proximity to railways? yes they
- Are launch sites in close proximity to highways?
- Are launch sites in close proximity to coastline?
- Do launch sites keep certain distance away from cities?

In [27]:

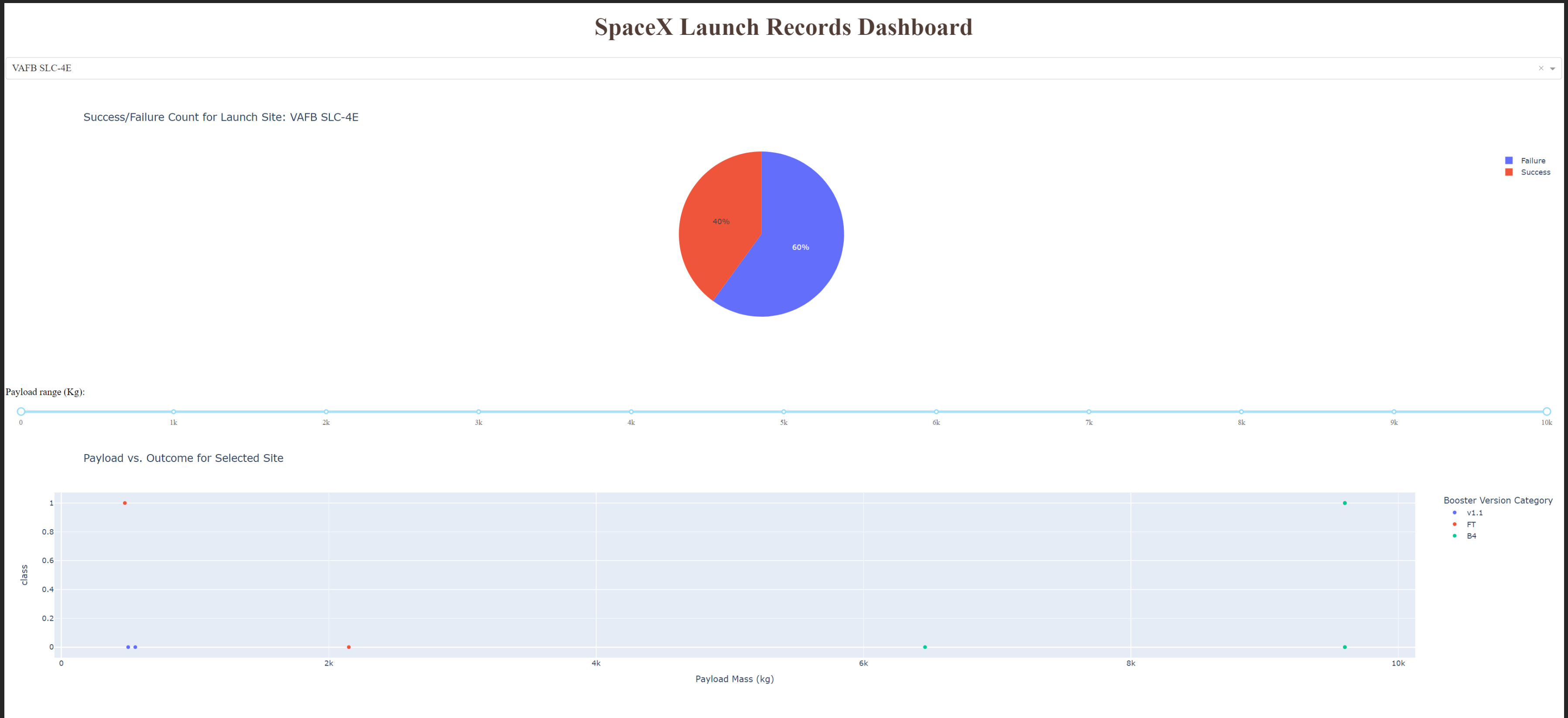
```
print(f'The distance to the coastline from the launch pad is {distance_coastline} km.')
print(f'The distance to the railway from the launch pad is {distance_railway} km.')
print(f'The distance to the highway from the launch pad is {distance_highway} km.')
print(f'The distance to the city of Melbourne from the launch pad is {distance_Melbourne} km.')
```

```
The distance to the coastline from the launch pad is 0.9229554298579413 km.
The distance to the railway from the launch pad is 1.3389196853620255 km.
The distance to the highway from the launch pad is 0.6571726261163688 km.
The distance to the city of Melbourne from the launch pad is 51.42599073416263 km.
```

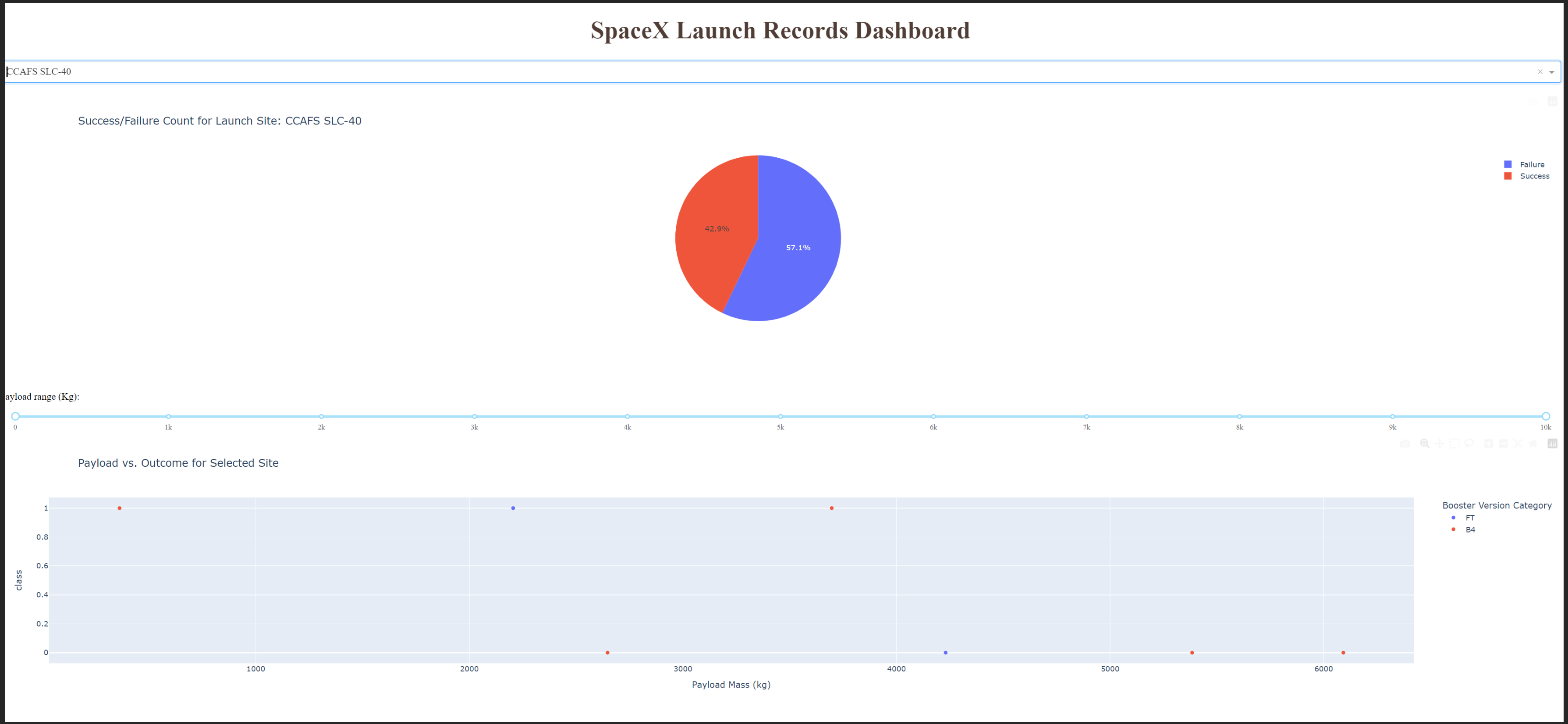
Plotly Dash Results



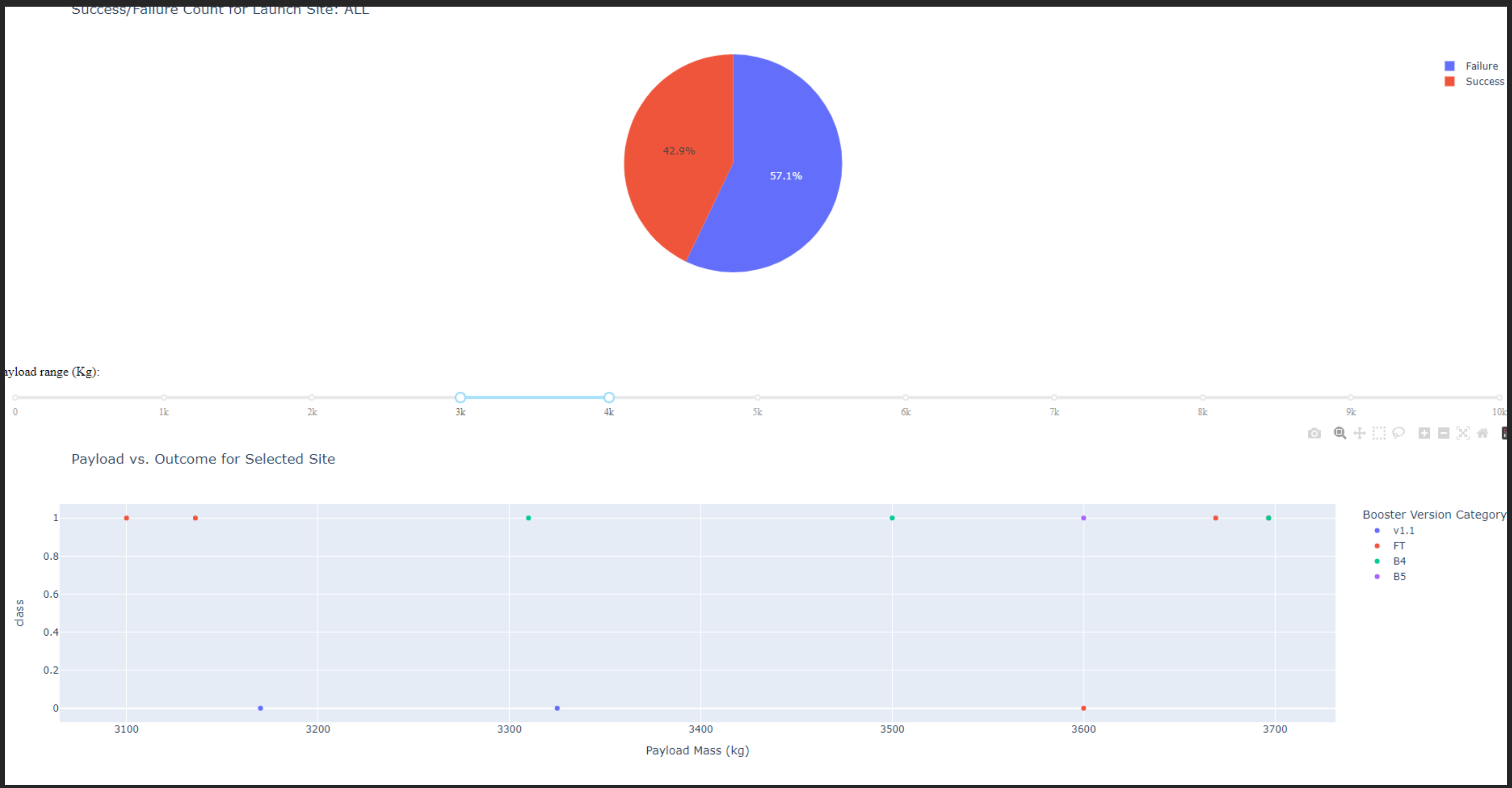
Plotly Dash Results- largest successful launches



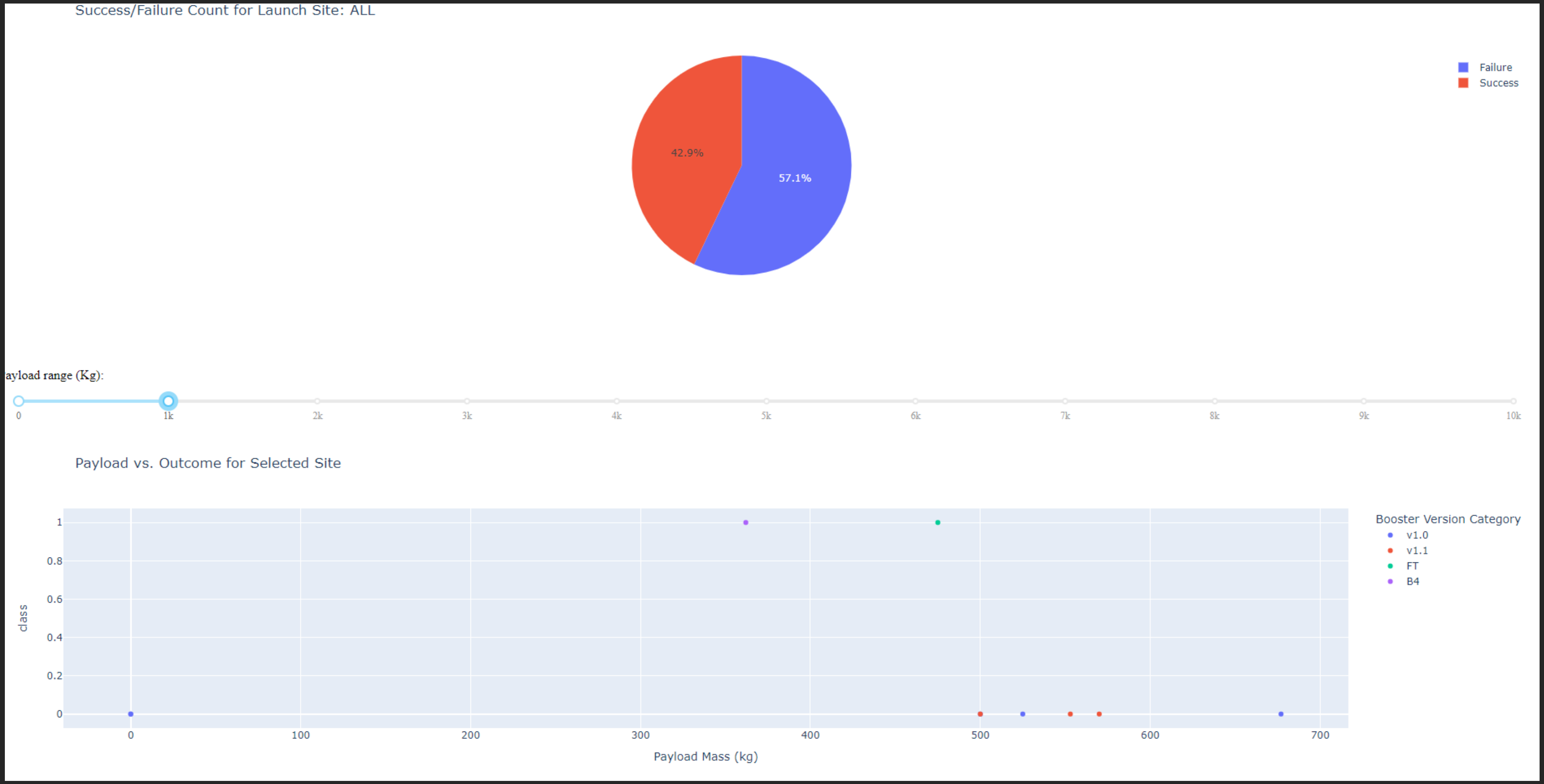
Plotly Dash Results – highest launch success rate



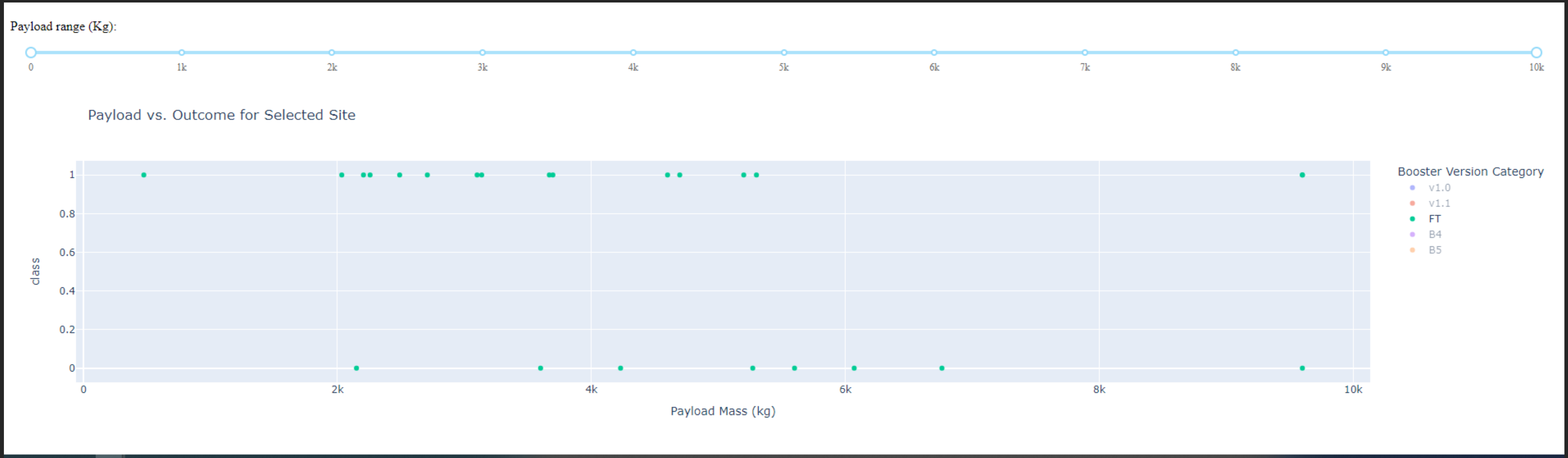
Plotly Dash Results- payload success 3k-4k



Plotly Dash Results- payload with lowest success rates

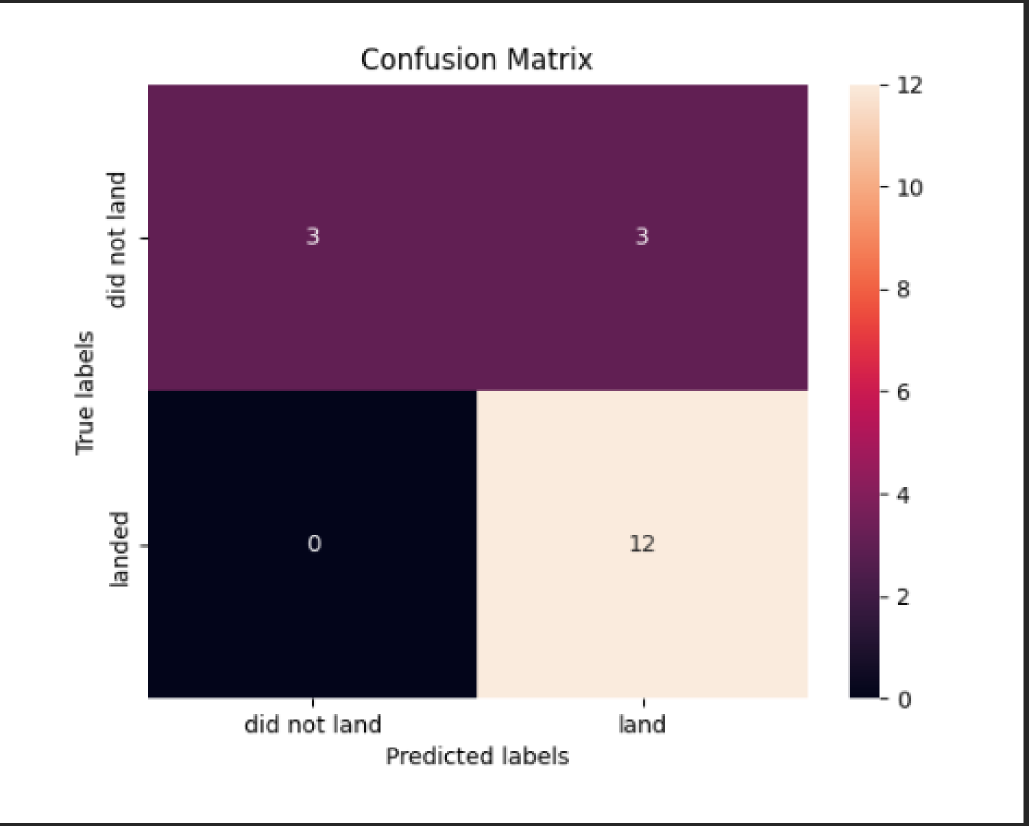


Plotly Dash Results- highest success rated booster- FT



Predicted analysis

We are able to predict 66% of the time
 a successful launch retrieval.
 We are able to determine
 similarly 33% of the time a unsuccessful recovery



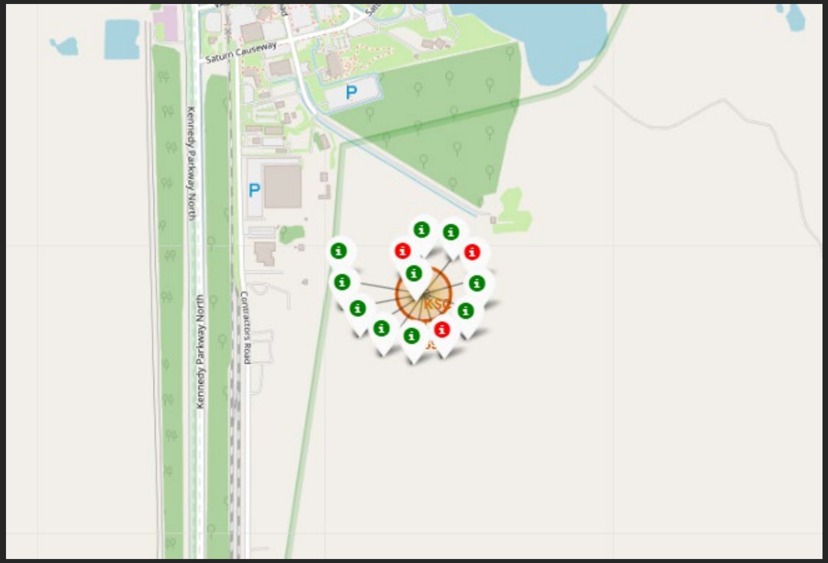
Conclusion

KSC LC 39a has the highest success rate from the launch sites available. Further insight needs to be placed on why This is the case

Try to keep payloads between 3k and 4k as much as possible to have a increased booster return rate. Avoid the high load and low loads as much as possible.

Use drone ships as often as possible (ASDS)

Have boosters poised for geosynchronous orbits to improve return potential.



Innovative insights

- There needs to be a full evaluation on the dates which the flights occurred to insure that they were not in the early stage of development. Insight from failed launches early on are outliers which were needed to learn to successfully execute the return effort.
- Also , I am skeptical about the launch sites – further analysis needs to be taken on the conditions for which launch site is determined. I would further like to see weather, humidity, wind considered.
- As additional tests are run, the modeling would need to be run again to insure that overfitting was not occurring.

The END.

- Thank you.

- Frederick Duff