Applied Data Science Capstone Project

Alif Putra Dewantara

31 December 2022

Outline

- EXECUTIVE SUMMARY
- INTRODUCTION
- METHODOLOGY
- RESULTS
- CONCLUSION
- APPENDIX

Executive Summary

- Summary of Methodologies
  - Data collection through API
  - Data collection with Web Scrapping
  - Exploratory Data Analysis (EDA) with SQL
  - Exploratory Data Analysis (EDA) with Data Visualization
  - Building a map visualization with Folium
  - Building a dashboard with Plotly Dash
  - Predictive Analysis (Classification)
- Summary of All Results
  - EDA results
  - Interactive analytics in screenshots
  - Predictive Analytics results from Machine Learning Lab

Introduction

- SpaceX is the most successful company in terms of commercial space travel. The company advertises its rocket launches, especially Falcon 9 as low as 62 million dollars when the other providers' cost is up to 165 million dollars. This cost is possible to make because SpaceX has revolutionary technologies in terms of its reusable rockets.As a data scientist at SpaceY, a startup company rivaling SpaceX, we need to create a machine learning pipeline to predict the landing outcome in the future. By that, we can also make more information in terms of bids against Space X.
- Question to be answered
  - How do the variables (payload mass, launch site, number of flight, orbits) affect the success of the first stage landing?
  - What is the rate of successful landings over the years?
  - What is the best algorithm that can be use for classification?

Methodology

Section 1

Data Collection

- We obtain SpaceX data from two sources:
    - Open Source SpaceX REST API
        - (FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude)
    - List of Falcon 9 and Falcon Heavy launches from Wikipedia (through web scrapping)
        - (Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time)

	Version, booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Success (drone ship)
h and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to ound-based astronomical observations.[15]								
),	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[17] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[18]	NASA (CTS)[19]	Success	No attempt
test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes, and splashed an 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[20] but that test article a ground test of SuperDraco engines on 20 April 2019.[21] The abort test used the capsule originally intended for the first crewed flight.[22] As expected, the booster y aerodynamic forces after the capsule aborted.[23] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its								
),	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Success (drone ship)
l and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the								
	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Failure (drone ship)
al and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching bit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[27] due to incorrect wind data.[28] This was the first time a flight failed to land.								
	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[30] (excl. Dragon mass)	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
hase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[31] Originally scheduled to launch on 2 March 2020, the pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[32] It was SpaceX's third flight 112 and the last launch of the cargo Dragon spacecraft.								
	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Failure (drone ship)
l launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May ds the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in However, the payload still reached the targeted orbit.[35] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual pped inside a sensor.[36]								
	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Success (drone ship)
l launch of Starlink satellites. The 84th flight of the Falcon 9 rocket, it surpassed Atlas V to become the most-flown operational US rocket.[38] Used fairings launched ugust 2019).[39]								
	F9 B5 B1058.1[41]	KSC, LC-39A	Crew Dragon Demo-2[42] (Crew Dragon C206.1 *Endeavour*)	12,530 kg (27,620 lb)[43]	LEO (ISS)	NASA (CCDev)	Success	Success (drone ship)
ital spaceflight from American soil since Space Shuttle STS-135 in July 2011, carrying NASA astronauts Bob Behnken and Doug Hurley to the International Space SpaceX live stream was peaked at 4.1 million viewers, while NASA estimated roughly 10 million people watched on various online platforms, and approximately gathered on Florida's space coast despite the risks of the COVID-19 pandemic.[44]								
	F9 B5 △ B1049.5	CCAFS, SLC-40	Starlink 7 v1.0 (60 satellites)	15,600 kg (34,400 lb)[14]	LEO	SpaceX	Success	Success (drone ship)
onal launch of Starlink satellites, occurred on the 10th anniversary of the first Falcon 9 flight. Included "VisorSat" satellite test that uses a sunshade to limit rst booster to successfully land five times, and first to land on Just Read The Instructions since it was moved to the East Coast.								
	F9 B5 △ B1059.3	CCAFS, SLC-40	Starlink 8 v1.0 (58 satellites),[48][49] SkySats-16, -17, -18	15,410 kg (33,970 lb)[47]	LEO	SpaceX Planet Labs	Success	Success (drone ship)
al launch of Starlink satellites, included the first rideshare in SpaceX's SmallSat Program, of three SkySat satellites.[50][49] One payload fairing half launched on fic 1 mission in December 2019. The other payload fairing half flew on Starlink 2 v1.0 in January 2020.[51] For the first time, SpaceX did not perform a static fire								

Data Collection through SpaceX API

- Get rocket launch data from SpaceX API
- Use .json() to decode response and convert it to dataframe using .json_normalize()
- Making data we already have into a dictionary
- Filtering the dataframe to only include the data we needed
- Performed data cleaning and filling the missing value by applying custom function
- Exporting the data to .csv

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

```
response = requests.get(spacex_url)
```

```
# Use json_normalize meethod to convert the json result
data = pd.json_normalize(response.json())
```

```
# Create a data from launch_dict
df = pd.DataFrame.from_dict(launch_dict)
```

```
data_falcon9 = df[df['BoosterVersion']!='Falcon 1']
```

Now that we have removed some values we should reset the FlgihtNumber column

```
data_falcon9.loc[:,'FlightNumber'] = list(range(1, data_falcon9.shape[0]+1))
data_falcon9
```

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

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Source:

[Applied-Data-Science-Capstone-IBM/1. jupyter-labs-spacex-data-collection-api.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]

Data Collection with Web Scrapping

```
# 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`

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text conte
soup = BeautifulSoup(response.text, 'html')
```

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

```
# Use soup.title attribute
soup.title
```

```
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
```

```
extracted_row = 0
#Extract each table
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowhea
   # get table row
    for rows in table.find_all("tr"):
        #check to see if first table heading is as number corresponding to launc
        if rows.th:
            if rows.th.string:
                flight_number=rows.th.string.strip()
                flag=flight_number.isdigit()
        else:
            flag=False
        #get table element
        row=rows.find_all('td')
        #if it is number save cells in a dictonary
        if flag:
```

```
df=pd.DataFrame(launch_dict)
```

- **Request**: Request the Falcon 9 Launch Wiki page from Wikipedia
- **Extract**: Extract all columns names from the HTML table header
- **Parse**: Parse the table and convert it to Pandas data frame

Source: [Applied-Data-Science-Capstone-IBM/2. jupyter-labs-webscraping.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]

Data Wrangling

Explore data

- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit and mission outcome per orbit

Create a landing outcome training label from 'Outcome' column

- True ASDS = drone ship landed succeeded
- None None = not attempted
- True RTLS = ground pad landed succeeded
- False ASDS = drone ship landing failed
- True Ocean = ocean landing succeded
- None ASDS = unable to be attempted due to launch failure
- False Ocean = ocean landing failed
- False RTLS = ground pad landing failed

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
Name: Outcome, dtype: int64
```

Source: [Applied-Data-Science-Capstone-IBM/3. labs-jupyter-spacex-Data-wrangling.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]

Exploratory Data Analysis (EDA) with SQL

The following SQL queries were performed:

• Names of the unique launch sites in the space mission;

• Top 5 launch sites whose name begin with the string '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 successful landing outcome in ground pad was achieved;

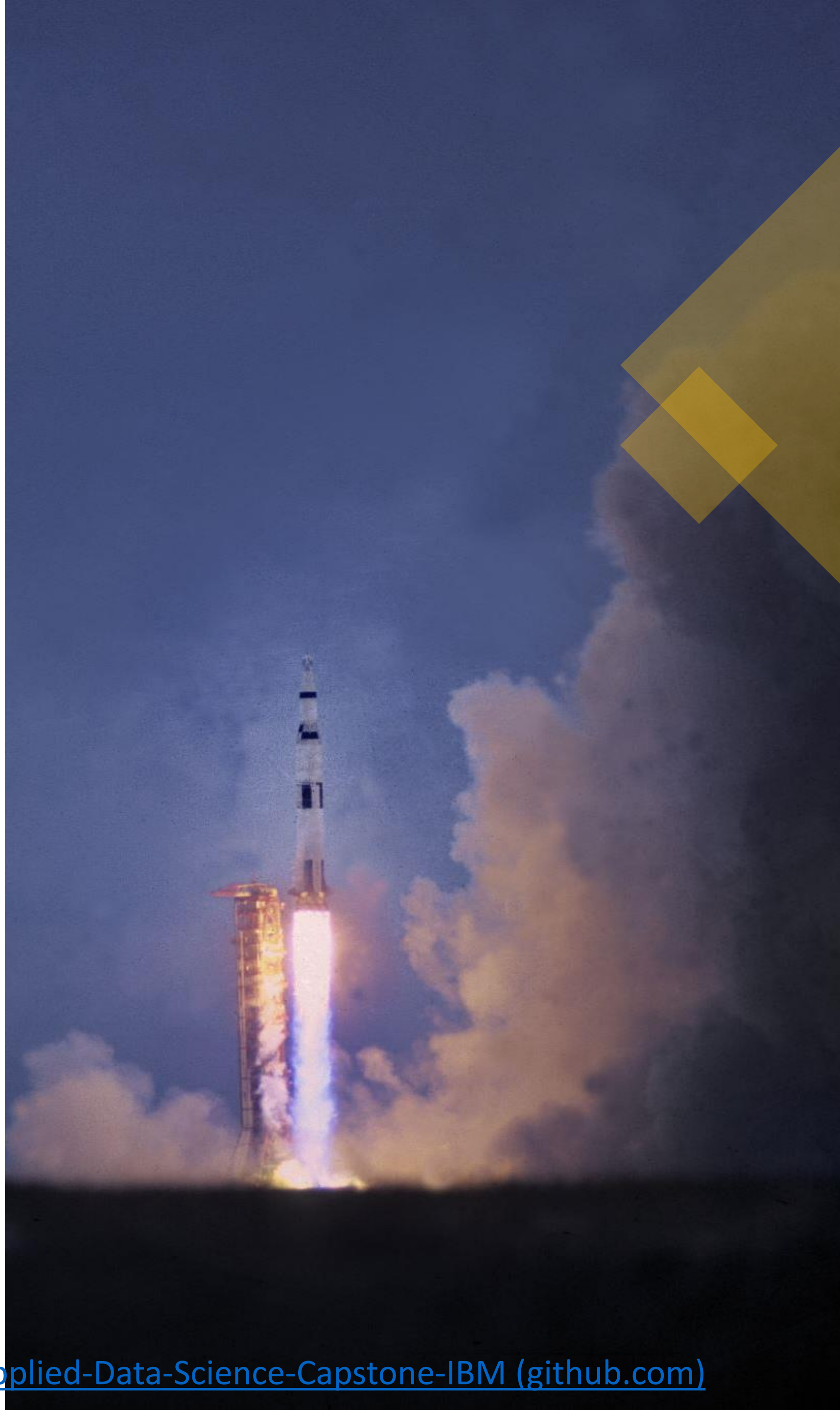
• Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg;

• Total number of successful and failure mission outcomes;

• Names of the booster versions which have carried the maximum payload mass;

• Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and

• 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.



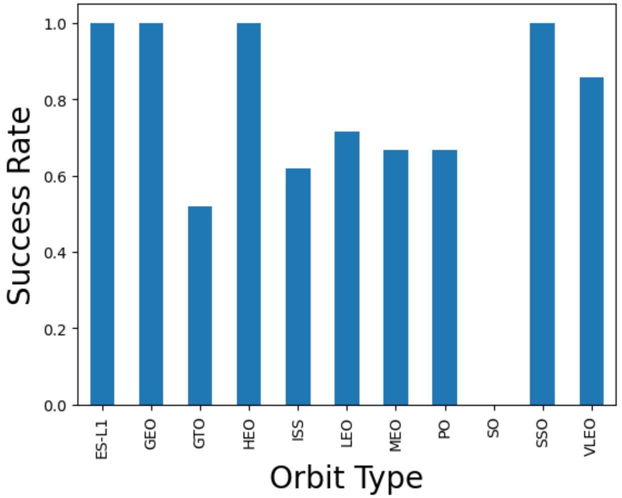
Source: [Applied-Data-Science-Capstone-IBM/4. jupyter-labs-eda-sql-coursera.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]

Exploratory Data Analysis (EDA) with Data Visualization

To explore data, scatterplots and barplots were used to visualize the relationship between pair of features:

- Payload Mass X Flight Number, Launch Site X Flight Number, Launch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit

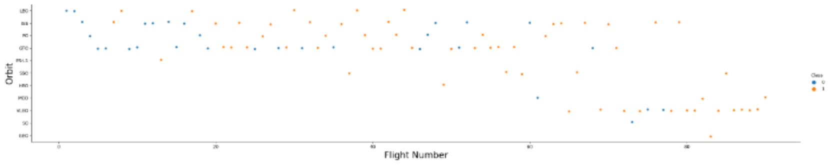
```python
# HINT use groupby method on Orbit column and get the mean of Class column
df_orbit = df.groupby('Orbit')['Class'].mean()
df_orbit.plot(kind='bar')
plt.xlabel('Orbit Type',fontsize=20)
plt.ylabel('Success Rate',fontsize=20)
plt.show()
```



TASK 4: Visualize the relationship between FlightNumber and Orbit type

For each orbit, we want to see if there is any relationship between FlightNumber and Orbit type.

```python
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the
sns.catplot(x='FlightNumber', y='Orbit', hue='Class', data=df, aspect = 5)
plt.xlabel('Flight Number', fontsize = 20)
plt.ylabel('Orbit', fontsize = 20)
plt.show()
```



Source: [Applied-Data-Science-Capstone-IBM/5. jupyter-labs-eda-dataviz.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]

```
site_map.add_child(circle)
```

Building a map visualization with Folium

Markers of all Launch Sites:

- Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location.
- Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.

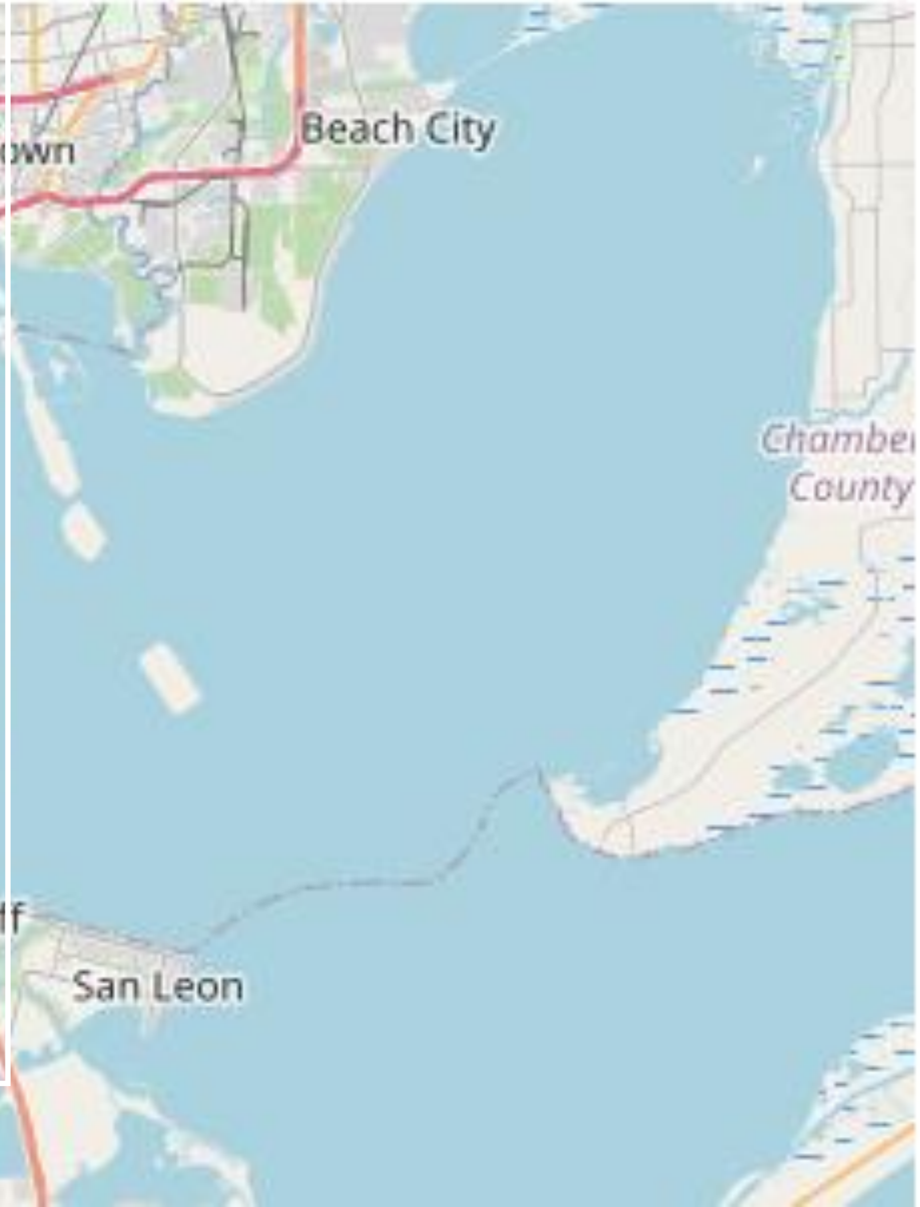
Colored Markers of the launch outcomes for each Launch Site:

- Added colored Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.

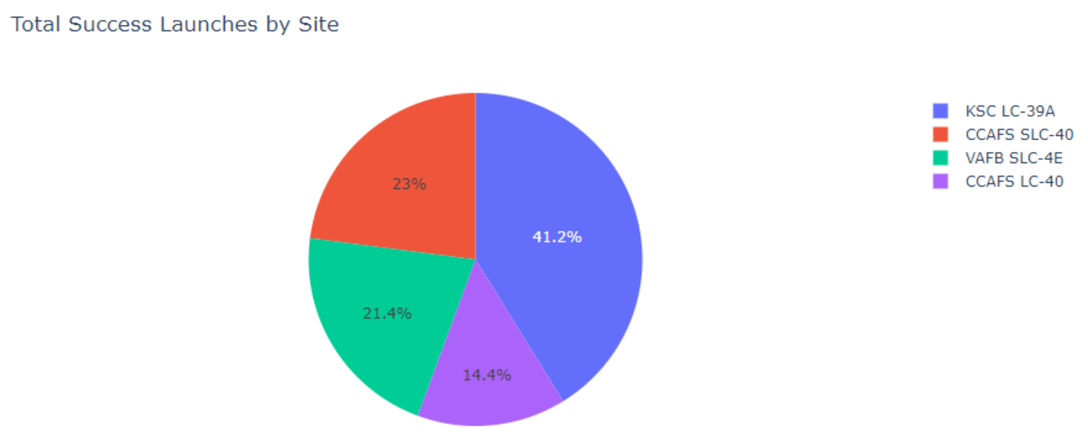
Distances between a Launch Site to its proximities:

- Added colored Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City

Source: [Applied-Data-Science-Capstone-IBM/6. lab_jupyter_launch_site_location.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]()



Building a Dashboard with Plotly Dash



- Launch Sites Dropdown List:

  Added a dropdown list to enable Launch Site selection.

- Pie Chart showing Success Launches (All Sites/Certain Site):

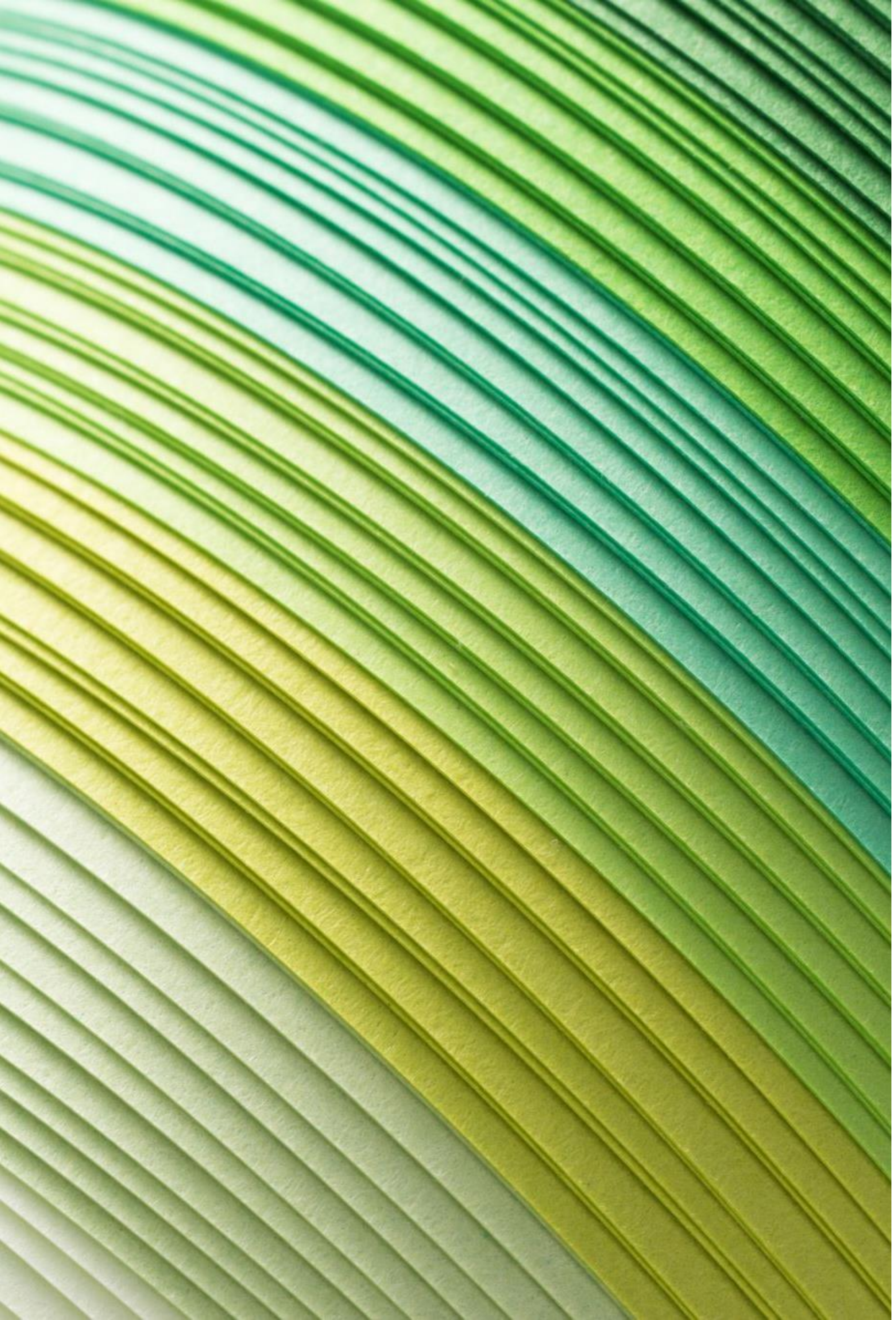
  Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected.

- Slider of Payload Mass Range:

  Added a slider to select Payload range.

- Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions:

  Added a scatter chart to show the correlation between Payload and Launch Success.



Predictive Analysis (Classification)

Creating a NumPy array from the column "Class" in data → Standardizing the data with StandardScaler, then fit and transform it → Splitting the data into training and testing sets with train_test_split function

→ Creating a GridSearchCV with cv = 10 to find the best parameters → Applying GridSearchCV on Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors models → Calculate the accuracy on the test data using the method .score() for all models

→ Examining the confusion matrix for all models → Finding the method performs best by examining the Jaccard_score and F1_score metrics

Source: [Applied-Data-Science-Capstone-IBM/8. SpaceX_Machine Learning Prediction_Part_5.ipynb at main · alifdewantaraa/Applied-Data-Science-Capstone-IBM (github.com)]

Results

The results will be categorized to main results:

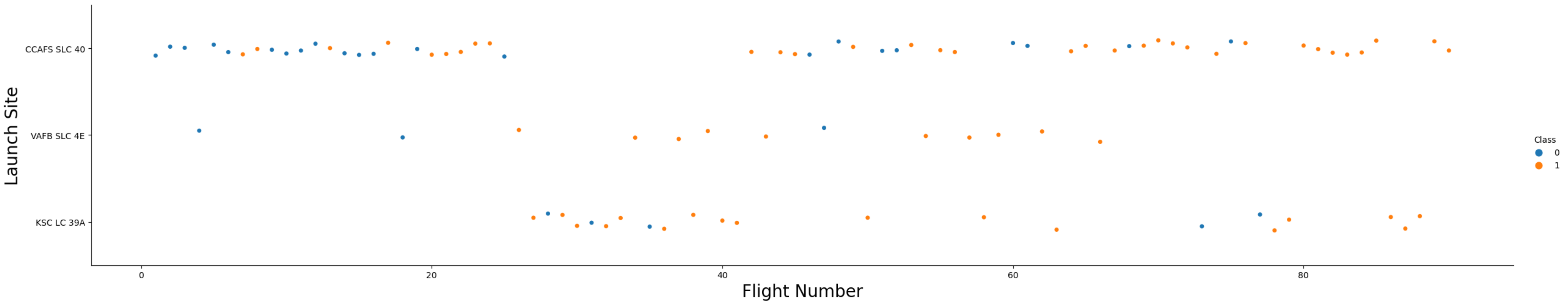
- Exploratory Data Analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Insights drawn from Exploratory Data Analysis

Section 2

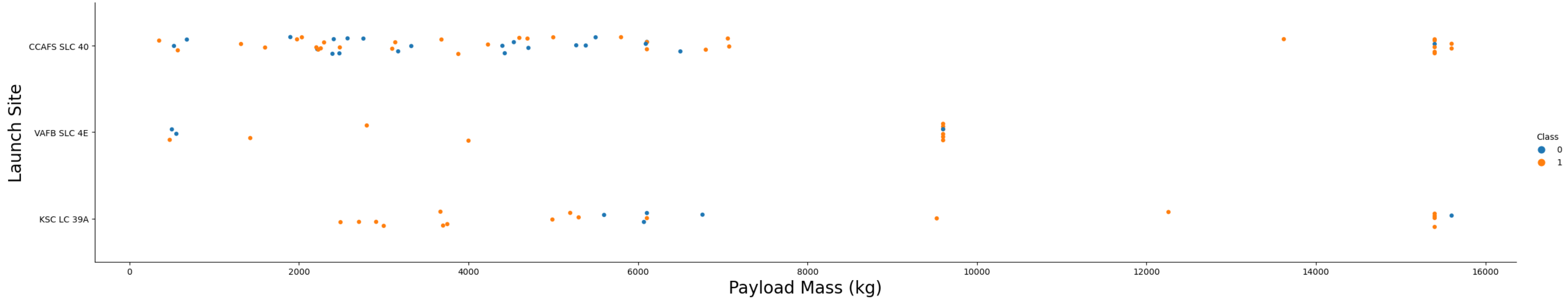
Flight Number vs. Launch Site

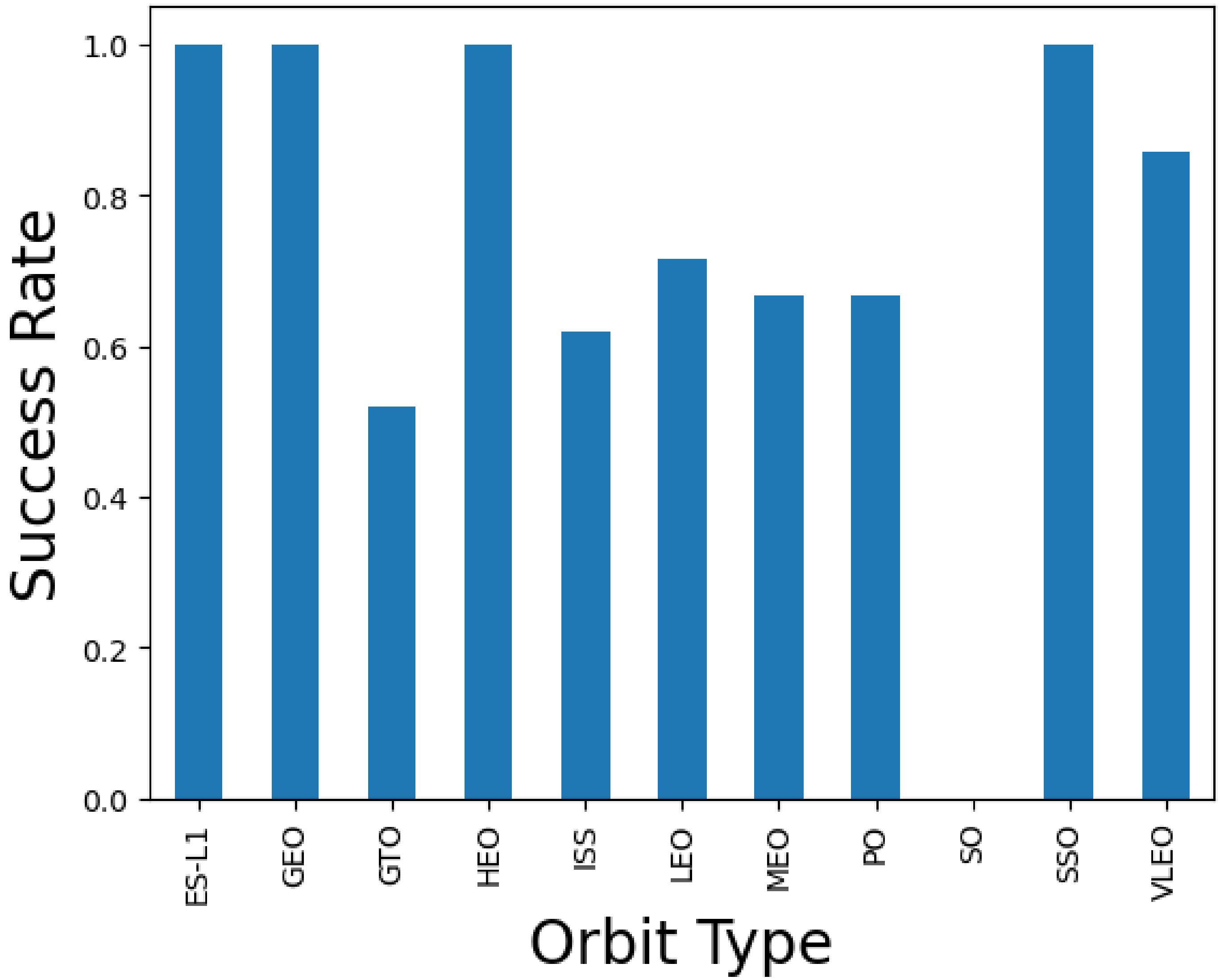
- This plot above explains that it is possible to verify that the best launch site nowadays is CCAF5 SLC 40
- The earliest flights all failed while the latest flight all succeded
- VAFB SLC 4E and KSC LC 39A have higher success rates
- It also explains that each new launch has higher rate of success



Payload vs. Launch Site

- For every launch site the higher the payload mass, the higher the success rate.
- Most of the launches with payload mass over 7000 kg were successful.
- KSC LC 39A has a 100% success rate for payload mass under 5500 kg too.





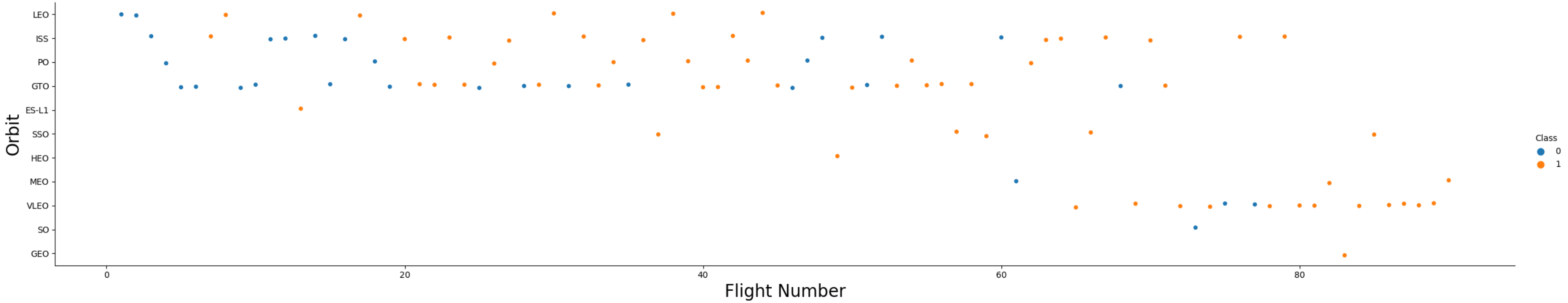
Success Rate vs. Orbit Type

- Orbits with 100% success rate:
  - ES-L1, GEO, HEO, SSO
- Orbits with 0% success rate:
  - SO
- Orbits with success rate between 50% and 85%:
  - GTO, ISS, LEO, MEO, P

However, deeper analysis show that some of this orbits has only 1 occurrence such as GEO, SO, HEO and ES-L1 which mean this data need more dataset to see pattern or trend before we draw any conclusion.

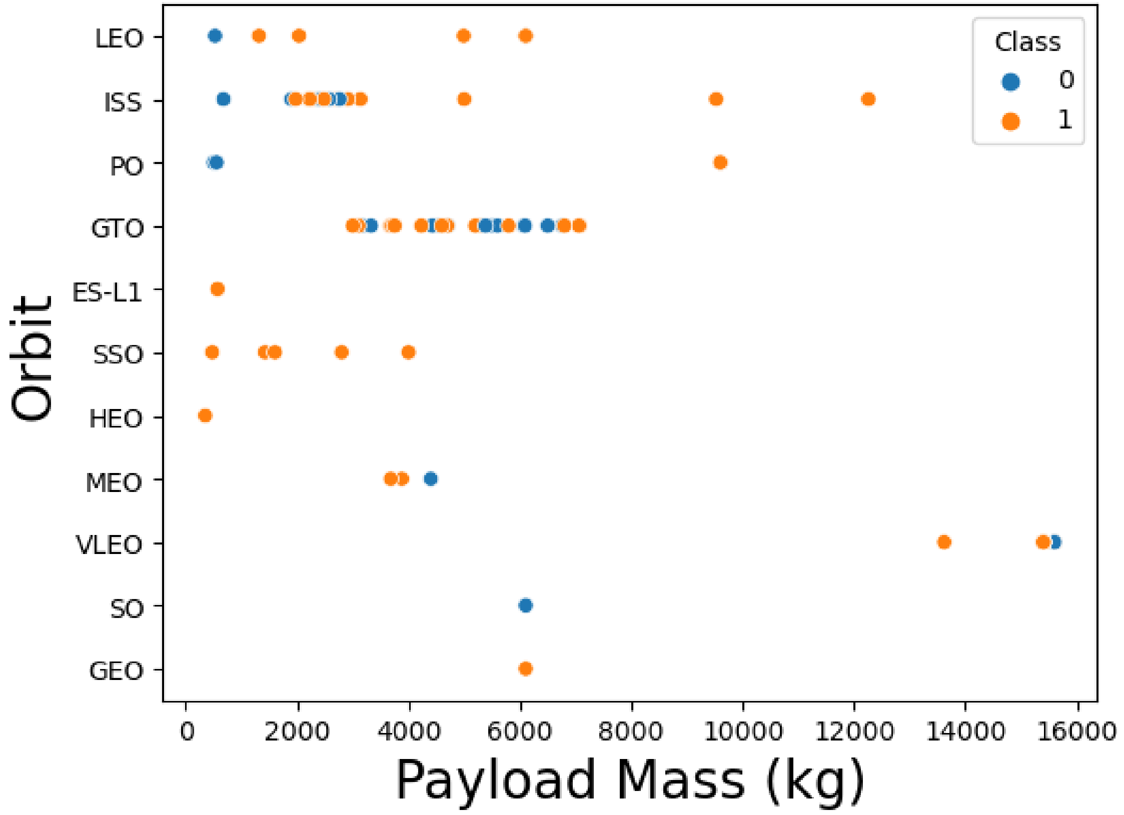
Flight Number vs. Orbit type

In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



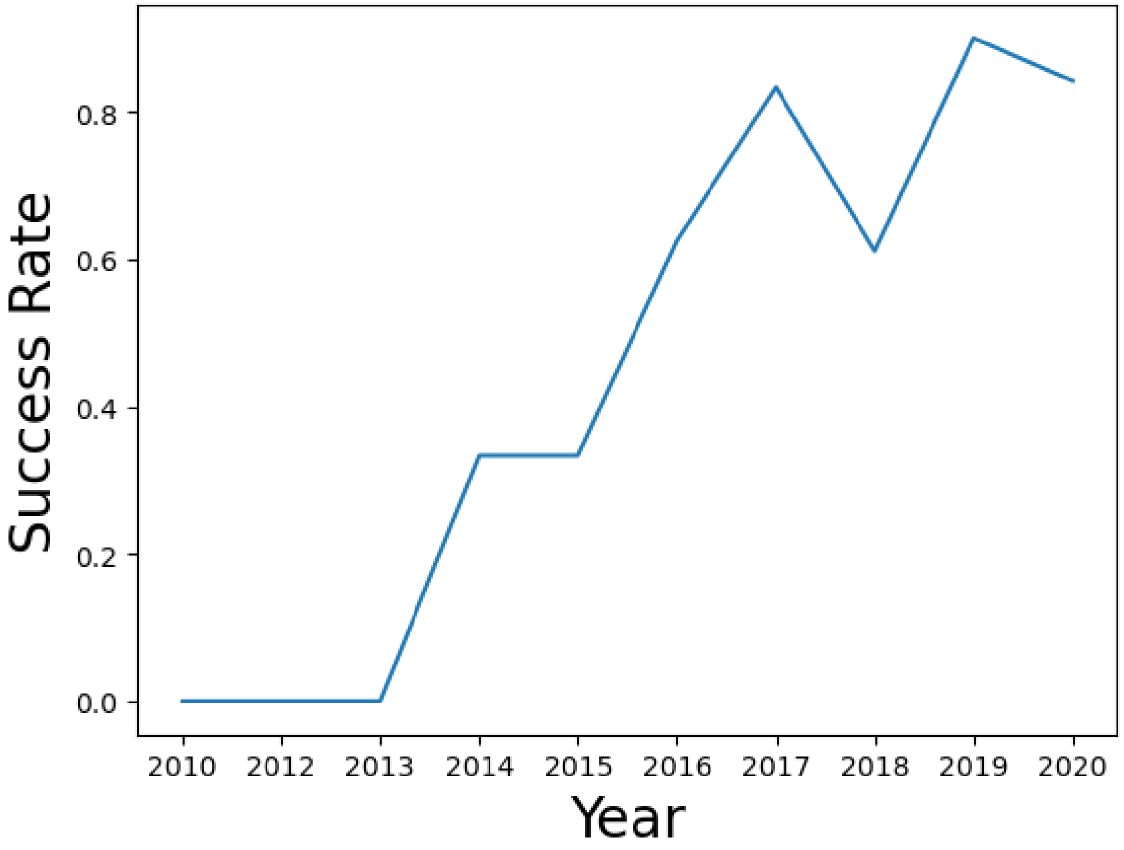
Payload Mass vs. Orbit Type

- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.



Launch Success Yearly Trend

This figures clearly depicted and increasing trend from the year 2013 until 2020. If this trend continue for the next year onward. The success rate will steadily increase until reaching 1/100% success rate.



All Launch Site Names

To show only unique launch sites from SpaceX data, we used keyword **DISTINCT**

```
%sql SELECT DISTINCT launch_site FROM SPACEX
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

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

Launch Site Names Begin with ‘CCA’

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

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

```
%sql SELECT * FROM SpaceX WHERE launch_site LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total payload mass

We calculated the total payload carried by boosters from NASA as **22007** using the query below

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

```
%sql SELECT SUM(payload_mass__kg_) AS total FROM SpaceX WHERE customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

total
22007

Average Payload Mass by F9 v1.1

We calculated the average payload mass carried by booster version F9 v1.1 as **3678**

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

```
%sql SELECT AVG(payload_mass__kg_) AS average FROM SpaceX WHERE booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomai
n.cloud:31864/bludb
Done.
```

average
3676

First Successful Ground Landing Date

- We use the *min()* function to find the result. We observed that the dates of the first successful landing outcome on drone ship pad was **05th July 2016**.

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

Hint:Use min function

```
%sql SELECT MIN(DATE) FROM SpaceX WHERE landing__outcome = 'Success (drone ship)'
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

1
2016-06-05

Successful drone ship landing with payload between 4000 and 6000

- We used the WHERE clause to filter for boosters which have successfully landed on drone ship and applied the AND condition to determine successful landing with payload mass 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 booster_version FROM SpaceX WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

booster_version
F9 FT B1022
F9 FT B1031.2

List the total number of successful and failure mission outcomes

```
%sql SELECT mission_outcome, COUNT(mission_outcome) FROM SpaceX GROUP BY mission_outcome
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomai
n.cloud:31864/bludb
Done.
```

mission_outcome	2
Success	44
Success (payload status unclear)	1

Total number of successful and failure mission outcomes

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

```
%sql SELECT booster_version FROM SpaceX WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SpaceX)
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

Boosters carried maximum payload

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

```
%sql SELECT landing__outcome, booster_version, launch_site FROM SpaceX WHERE landing__outcome = 'Failure (drone ship)' AND YEAR(DATE) = 2015
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40

2015 launch records

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(landing__outcome) AS count FROM SpaceX WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing__outcome
```

```
 * ibm_db_sa://hwz01292:***@21fecfd8-47b7-4937-840d-d791d0218660.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31864/bludb
Done.
```

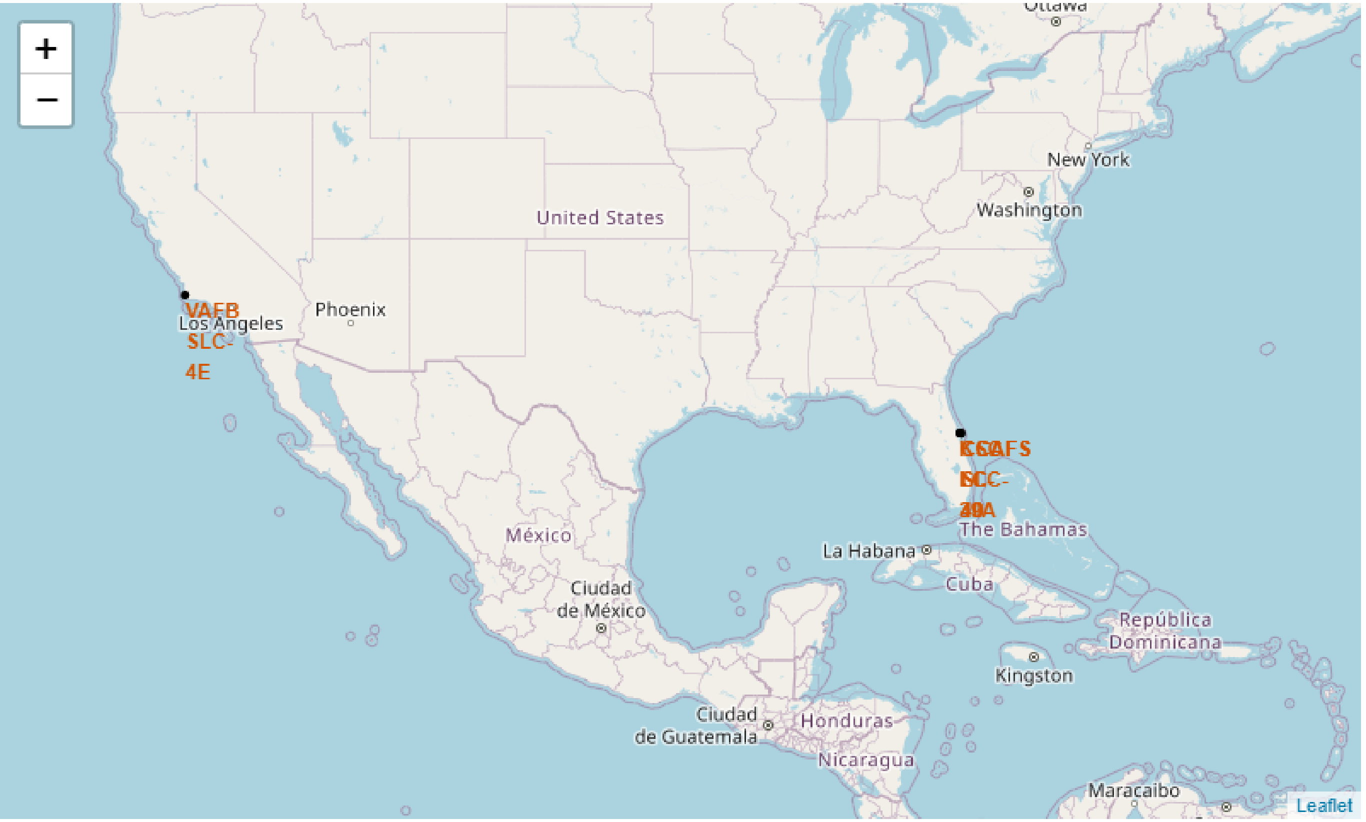
landing__outcome	COUNT
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

Rank success count between 2010-06-04 and 2017-03-20

Launch Sites Proximities Analysis

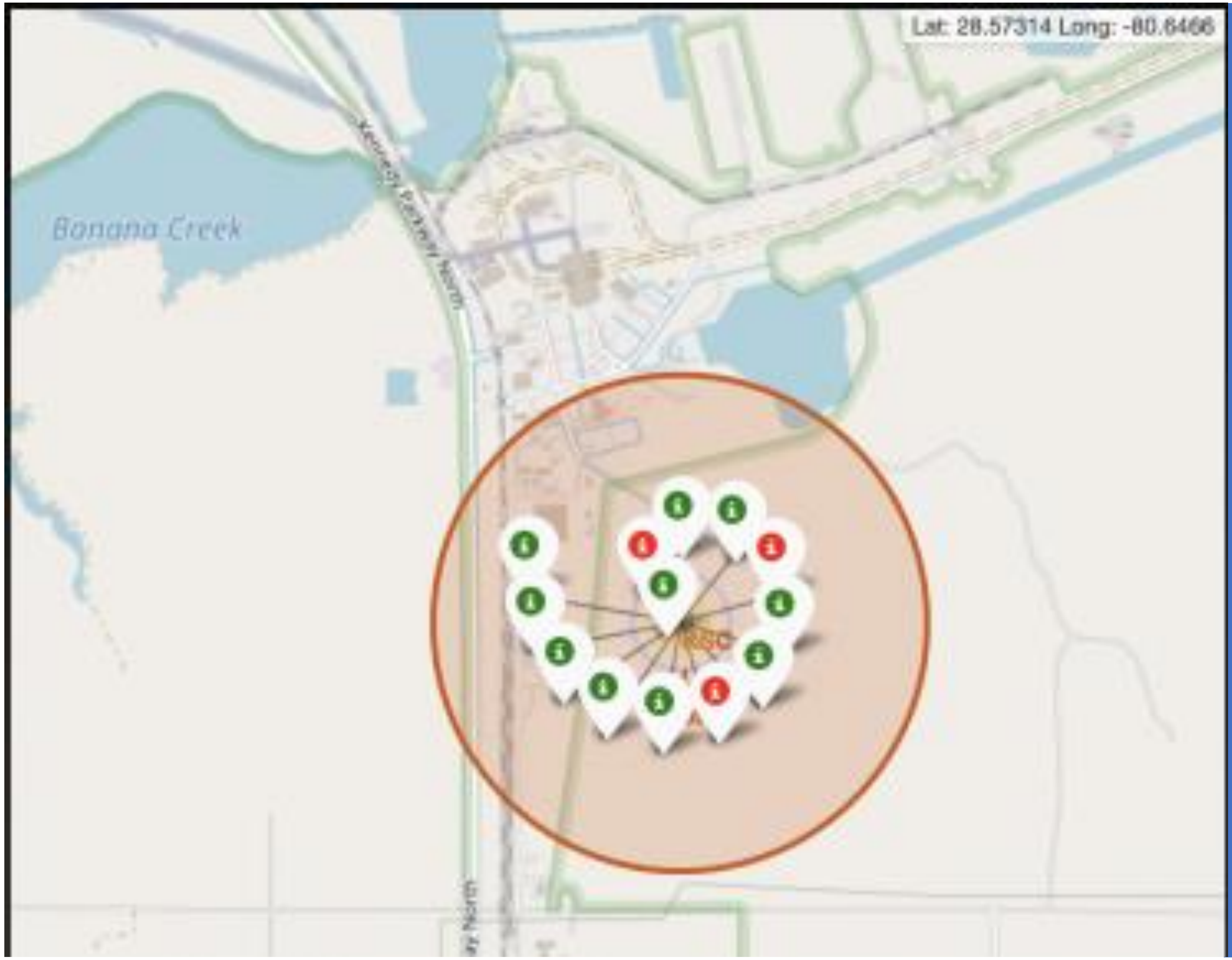
Section 3

```python
# Initialize the map
site_map = folium.Map(location=nasa_coordinate, zoom_start=4)
# For each launch site, add a Circle object based on its coordinate (Lat, Long) values. In addition, add Launch si
for index, row in launch_sites_df.iterrows():
    coordinate = [row['Lat'], row['Long']]
    folium.Circle(coordinate, radius=1000, color='#000000', fill=True).add_child(folium.Popup(row['Launch Site']))
    folium.map.Marker(coordinate, icon=DivIcon(icon_size=(20,20),icon_anchor=(0,0), html='<div style="font-size: 1
site_map
```



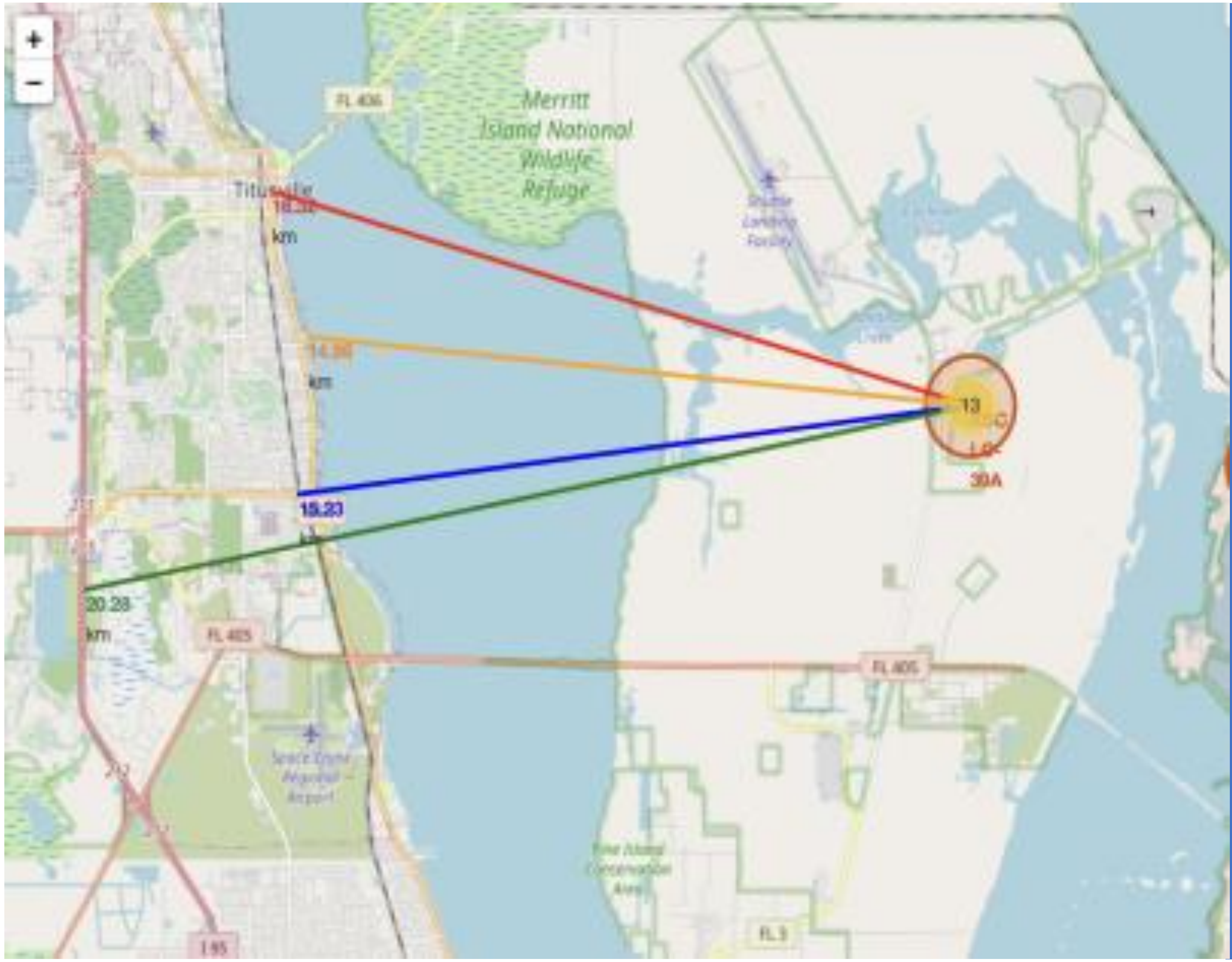
Location of all the Launch SItes

- Visualizing the launch sites on a map highlights the importance of launch site proximity to the coast and equator:



Colour-labeled launch records on the map

- From the colour-labeled markers we should be able to easily identify which launch sites have relatively high success rates.
  - Green Marker = Successful Launch
  - Red Marker = Failed Launch
- Launch Site KSC LC-39A has a very high Success Rate

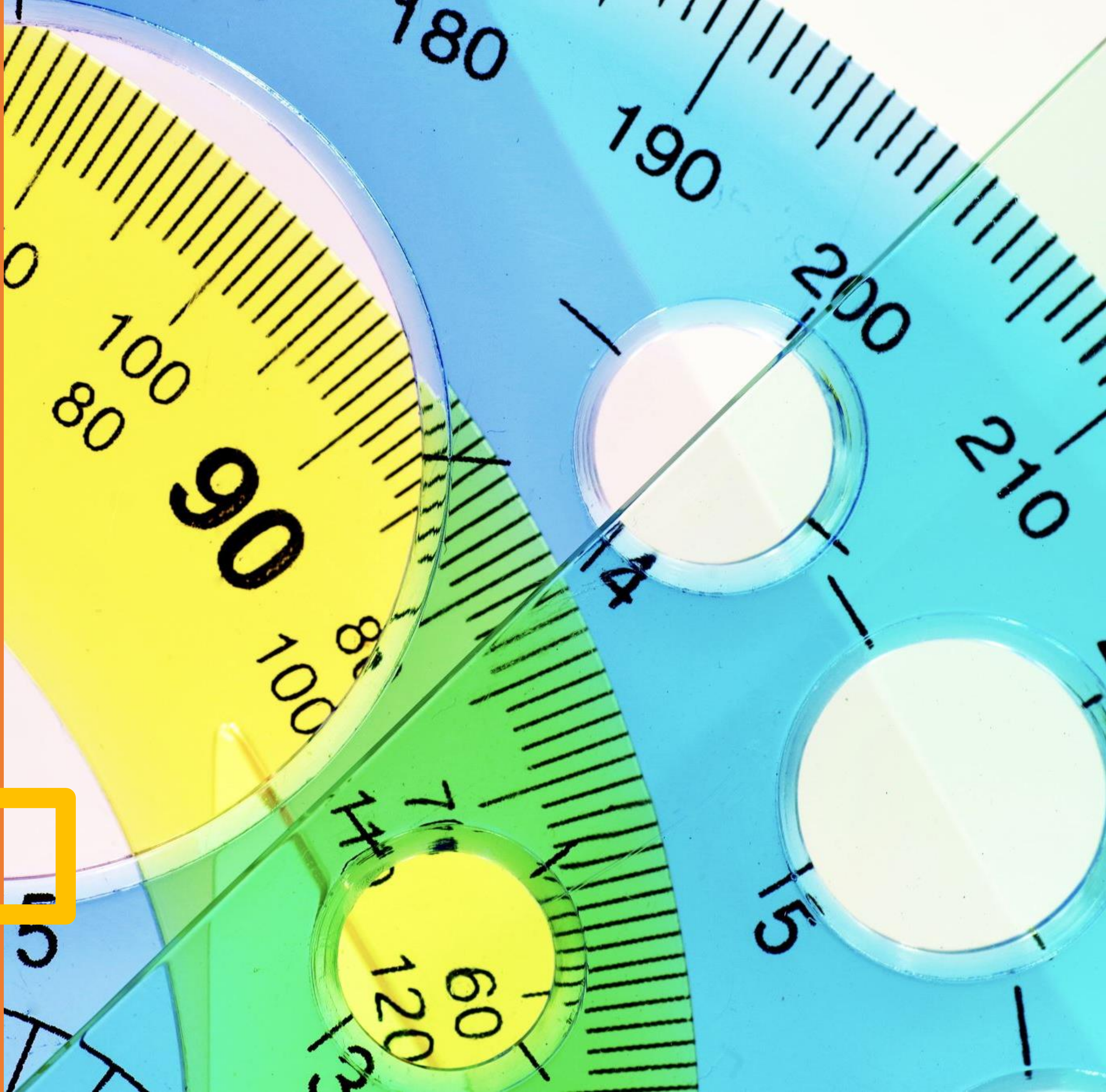


Distance from the launch site KSC LC-39A to its proximities

- From the visual analysis of the launch site KSC LC-39A we can clearly see that it is:
  - relatively close to railway (15.23 km)
  - relatively close to highway (20.28 km)
  - relatively close to coastline (14.99 km)
- Also, the launch site KSC LC-39A is relatively close to its closest city Titusville (16.32 km).
- Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas

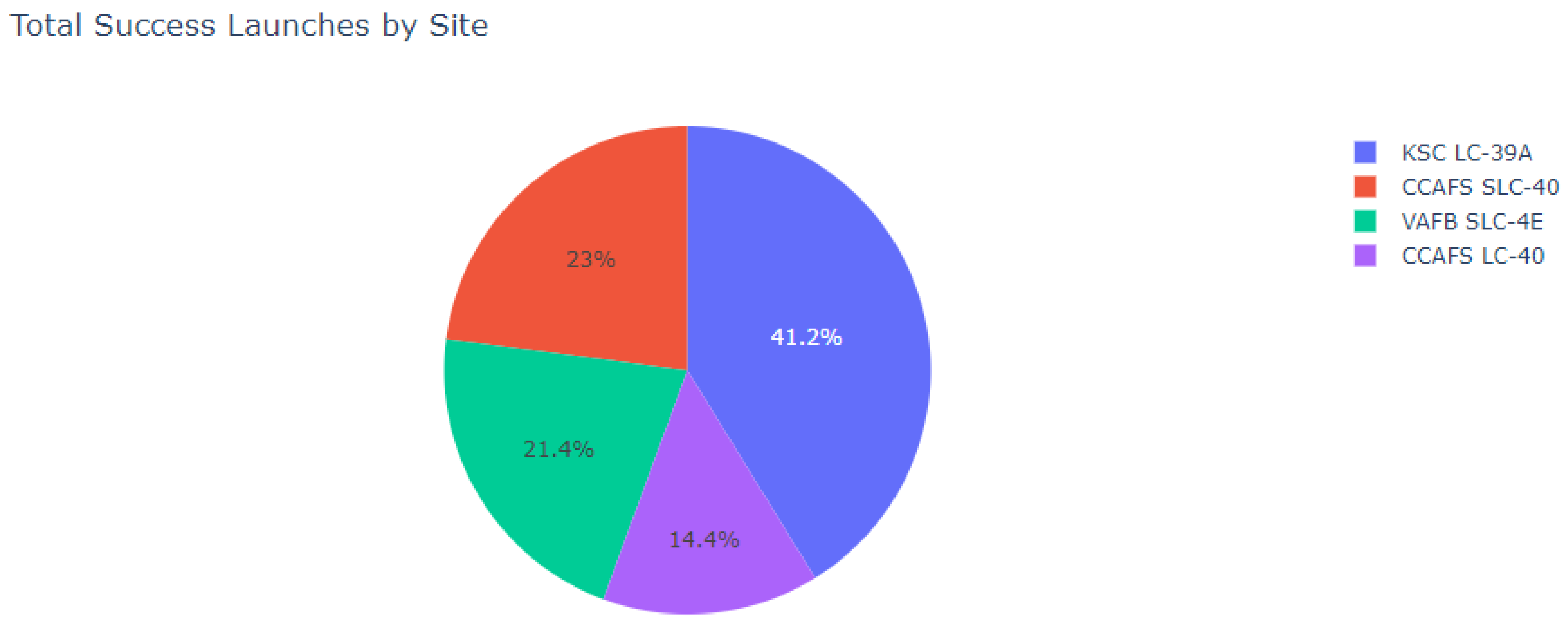
Build a pie chart with Plotly Dash

Section 4



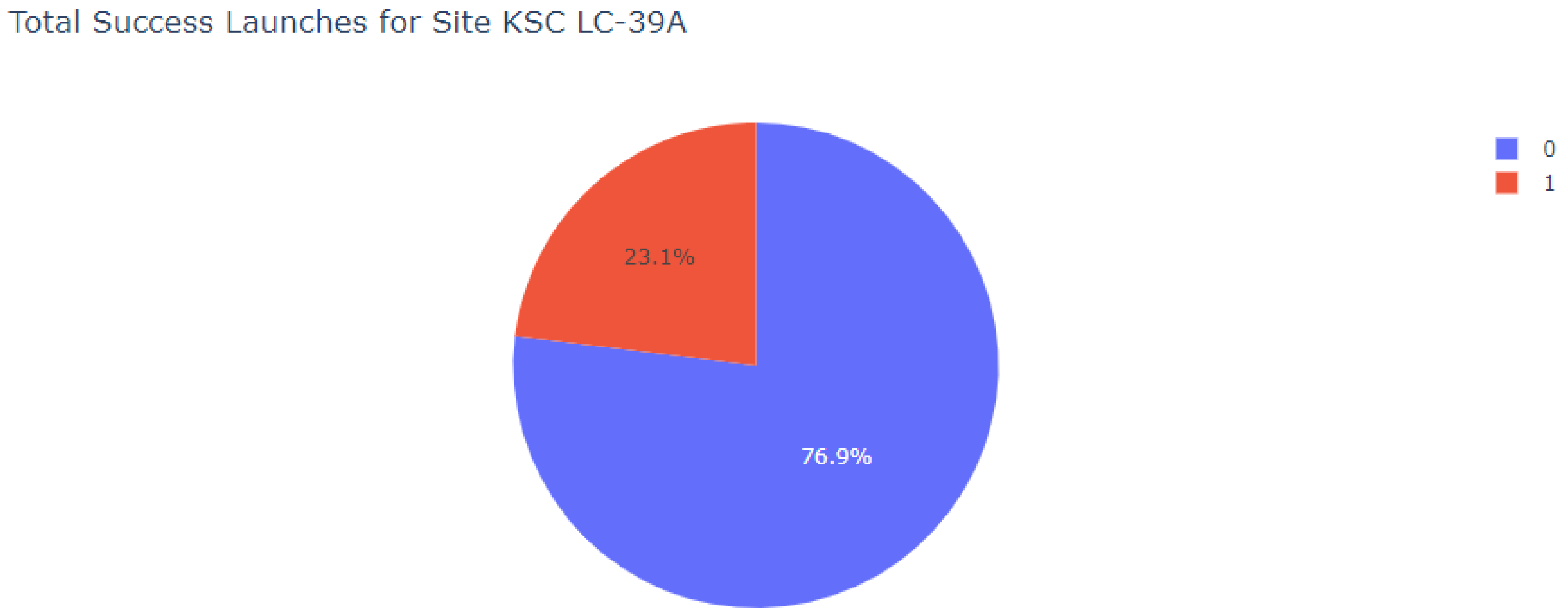
The success percentage by each sites

We can see that KSC LC-39A had the most ucessful launches from all the sites



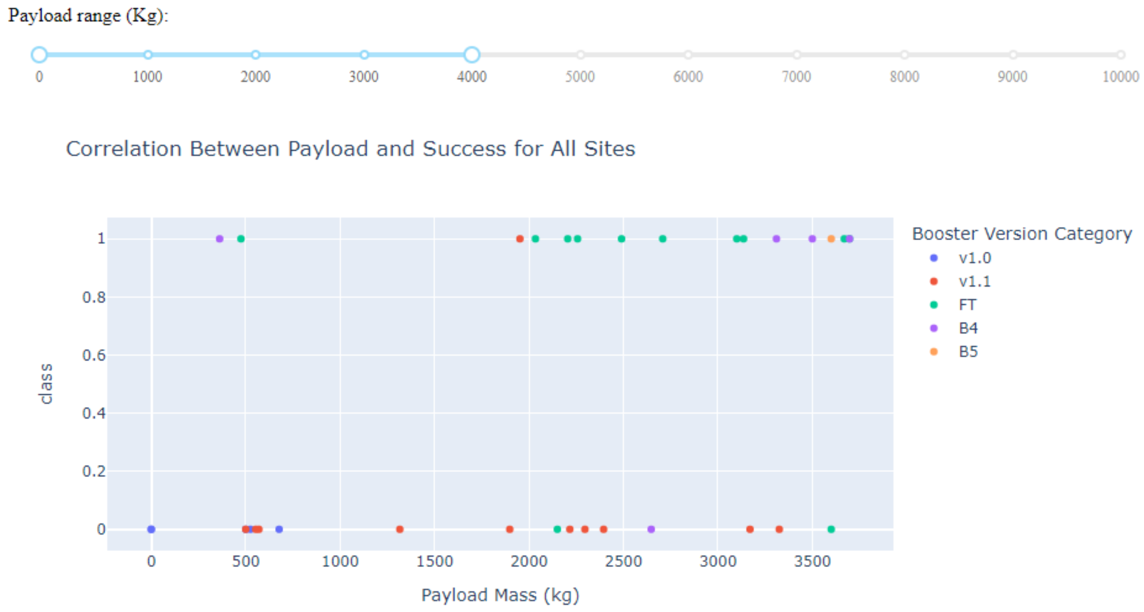
Launch site with highest launch success ratio

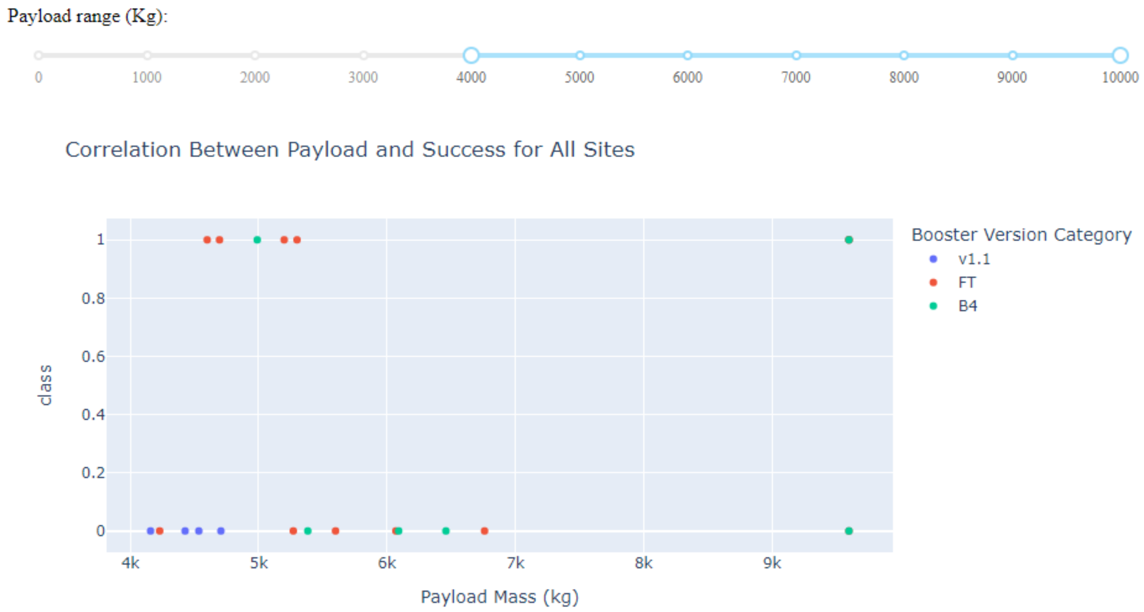
KSC LC-39A has the highest launch success rate (76.9%) with 10 successful and only 3 failed landings



Payload vs Launch Outcome Scatter Plot

We can see that all the success rate for low weighted payload is higher than heavy weighted payload





Predictive Analysis

Section 5

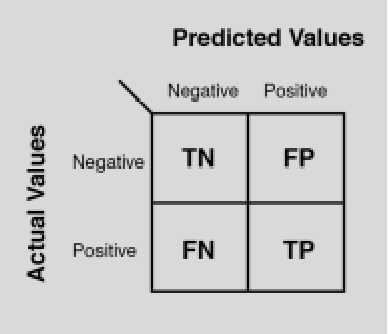
```python
algorithms = {'KNN':knn_cv.best_score_,'Tree':tree_cv.best_score_,'LogisticRegression':logreg_cv.best_score_}
bestalgorithm = max(algorithms, key=algorithms.get)
print('Best Algorithm is',bestalgorithm,'with a score of',algorithms[bestalgorithm])
if bestalgorithm == 'Tree':
    print('Best Params is :',tree_cv.best_params_)
if bestalgorithm == 'KNN':
    print('Best Params is :',knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best Params is :',logreg_cv.best_params_)
```

```
Best Algorithm is Tree with a score of 0.8892857142857145
Best Params is : {'criterion': 'entropy', 'max_depth': 6, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 
5, 'splitter': 'random'}
```

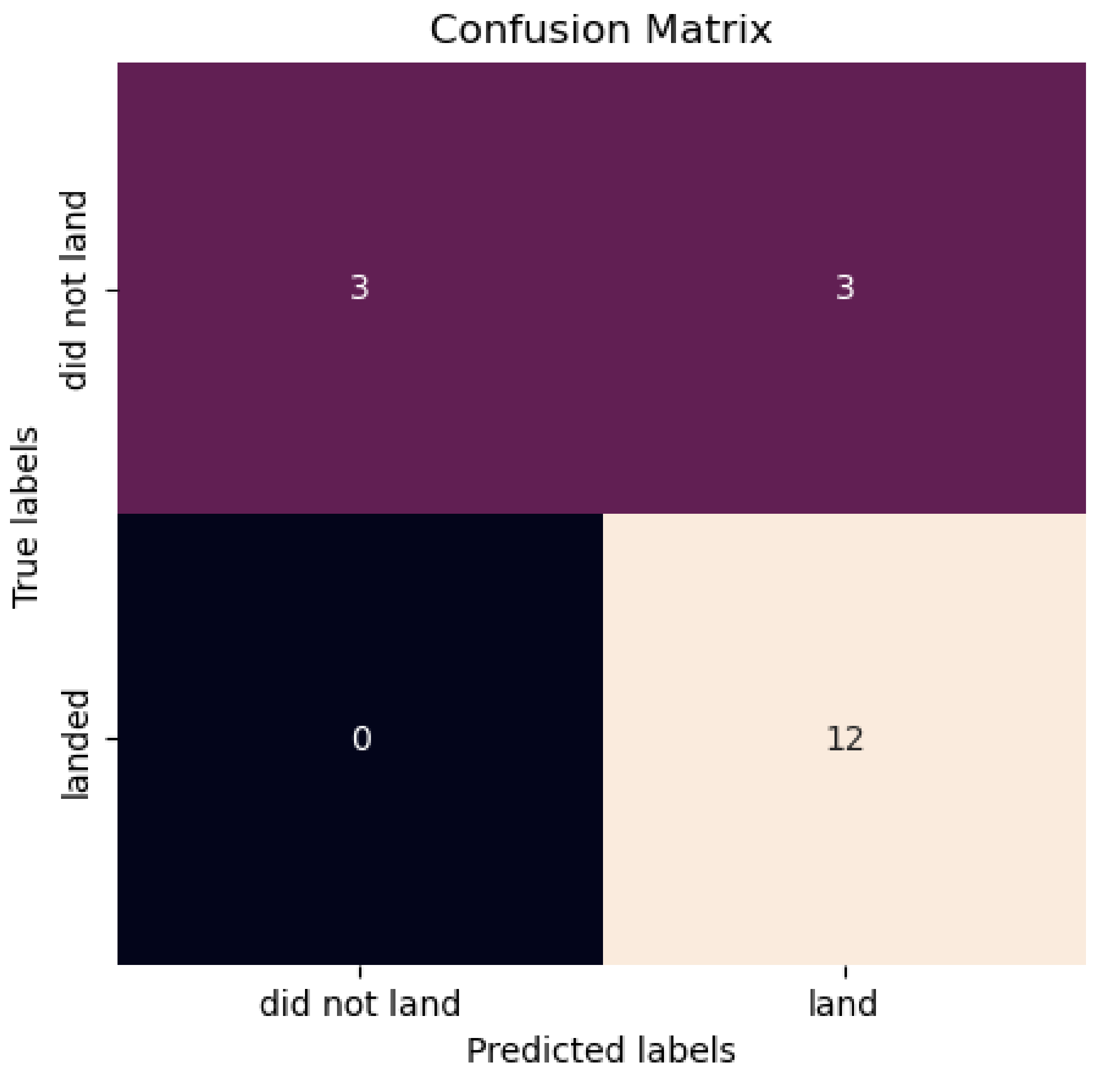
Classification Accuracy

- As we can see, by using the code as below: we could identify that the best algorithm to be the Tree Algorithm which have the highest classification accuracy.

Confusion Matrix



- Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives



Conclusion

- The Tree Classifier Algorithm is the best Machine Learning approach for this dataset.
- Launches with a low payload mass show better results than launches with a larger payload mass.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- The success rate of launches increases over the years.
- KSC LC-39A has the highest success rate of the launches from all the sites.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate

Thank You!

Thanks to:

[Instructors]

[Coursera]

[IBM]