IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aditya Nema

30th July 2023

Outline

Executive Summary

Summary of methodologies
- Data Collection
- Data Wrangling
- EDA with data visualization
- EDA with SQL
- Build an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive Analysis

Summary of all results
- Exploratory data analysis results
- Interactive analysis.
- Predictive analysis results

Introduction

- In this analysis we will focus on the study of the first stage of the SpaceX Falcon9 Rocket in order to obtain conclusions that allow us to make cost projections as well as obtain insight on the implications in the area beyond the economic ones.
- We will work on finding the following solutions:
  - How to predict if the rocket will land successfully.
  - What parameters can determine the success rate of a landing successful.
  - Project and determine the costs of future launches.

Section 1

Methodology

Methodology

- Data collection methodology:
  - Data was collecting from past SpaceX missions. SpaceX API
  - Web Scrapping from Wikipedia
- Perform data wrangling
  - Calculated the number of launches on each site
  - Calculated number and occurrences of each orbit
  - Calculated the number and occurrence of mission outcome per orbit type
  - Created a landing outcome label from Outcome column
- 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 sets were collected from previous SpaceX mission and Wikipedia pages and below processes were obtained to Filter, clean and Transform the data to prepare for Modeling

SpaceX API

We make a get request to the SpaceX API. We also perform some basic data wrangling and formatting.

Web Scraping

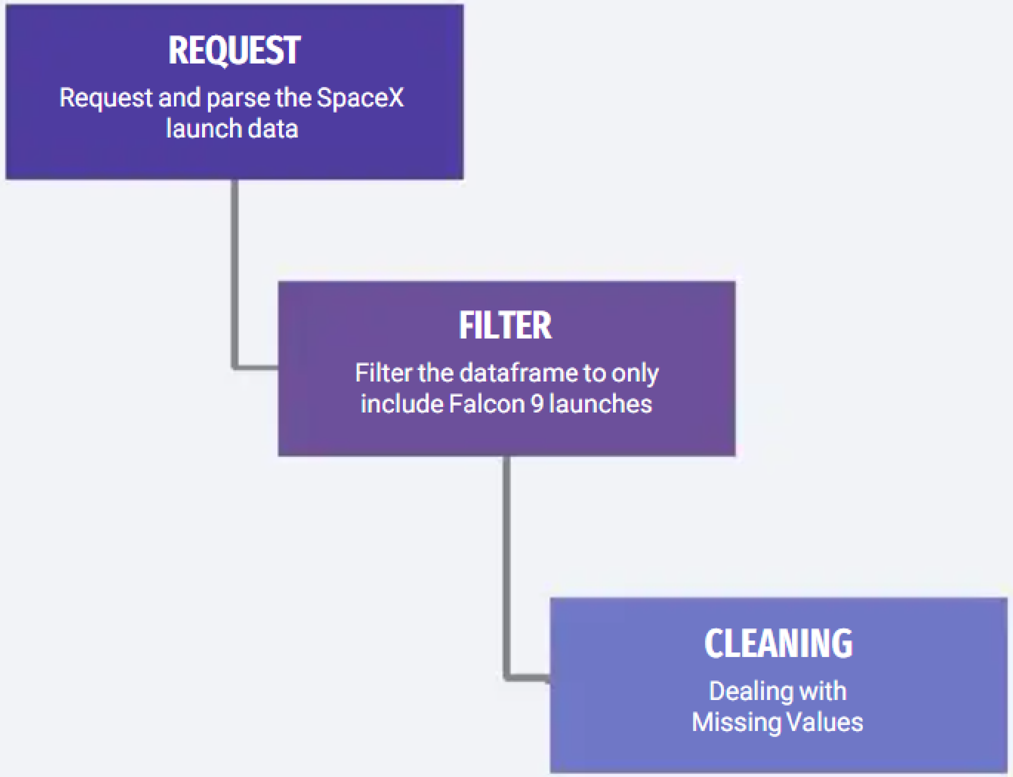
We performed web scraping to collect historical Falcon 9 launch records from a Wikipedia.

Data Wrangling

Through a preliminary exploratory analysis identifying the transformations that are required in the data set to prepare them.

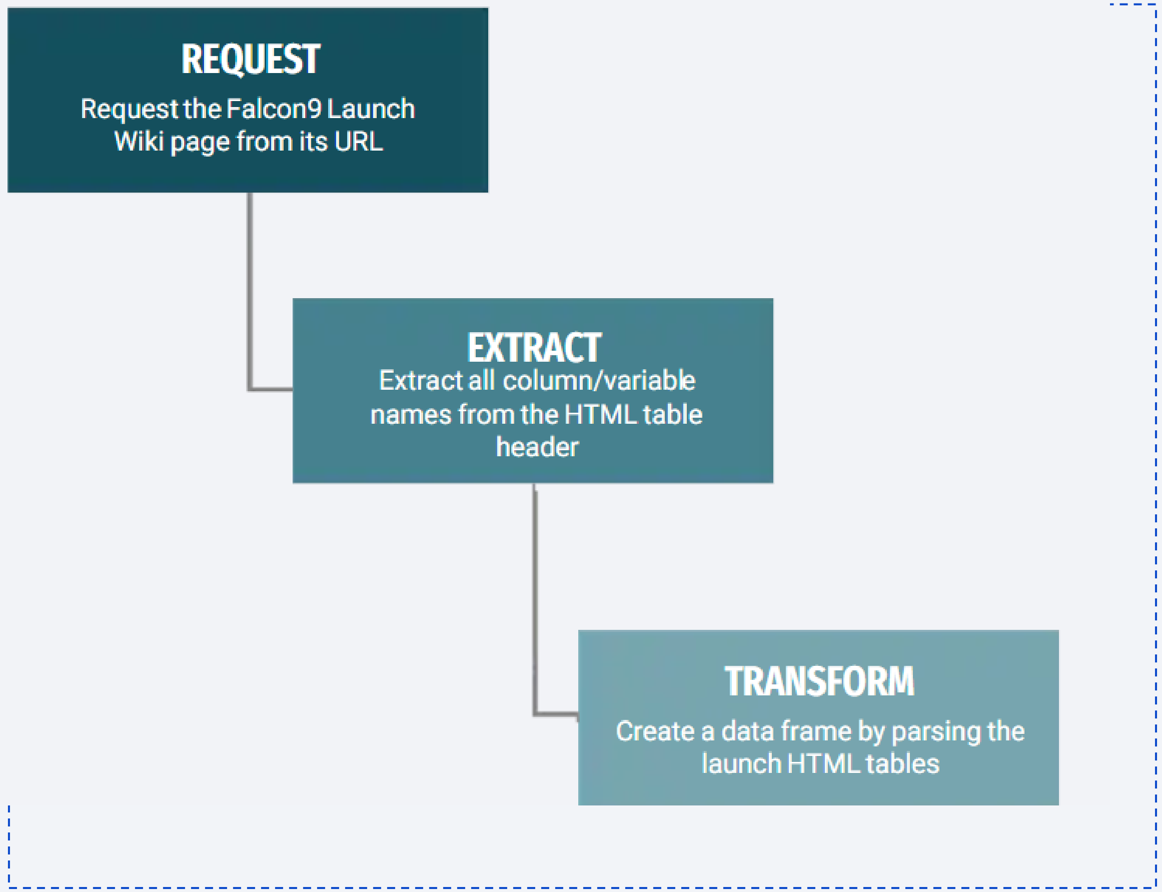
Data Collection – SpaceX API

- We make a get request to the SpaceX API. We also perform some basic data wrangling and formatting.
- It can be seen in detail in the following- https://github.com/adityanema91/data-science-capstone-coursera/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

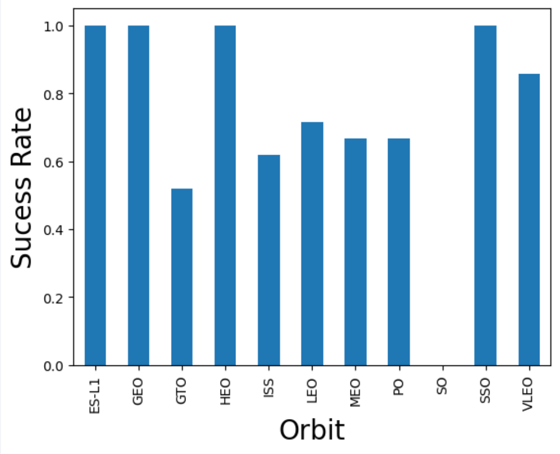
- We performed web scraping to collect historical Falcon 9 launch records from a Wikipedia page titled “List of Falcon 9 and Falcon Heavy launches”.
- It can be seen in detail in the following- https://github.com/adityanema91/data-science-capstone-coursera/blob/main/jupyter-labs-webscraping.ipynb



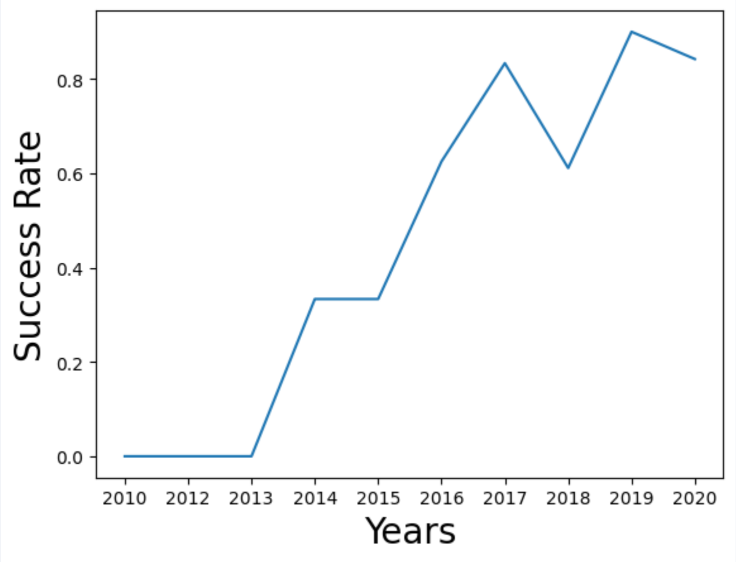
Data Wrangling

- Through a preliminary exploratory analysis identifying the transformations that are required in the data set to prepare them
- We will process the landing data into valid tags for training the predictive models later.
- Training tags with "1" will mean the rocket landed successfully and "0" means it was unsuccessful.
- See in detail in GitHub link

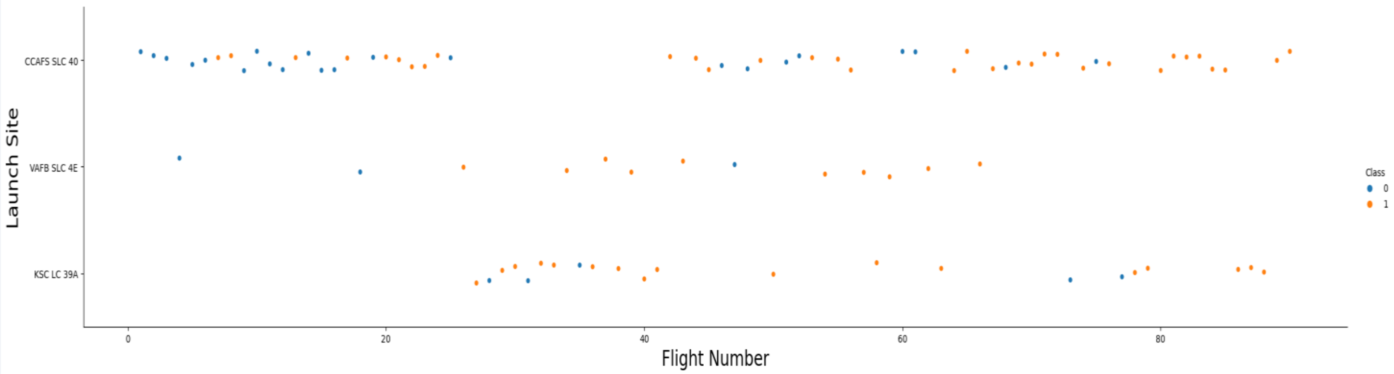
EDA with Data Visualization



- Exploratory Data Analysis to visualize the relationship between:
  - Flight Number and Launch Site.
  - Payload and Launch Site.
  - Success rate of each orbit type.
  - Flight Number and Orbit type.
  - Payload and Orbit type.
  - Visualize the launch success yearly



- See in details in [link]



EDA with SQL

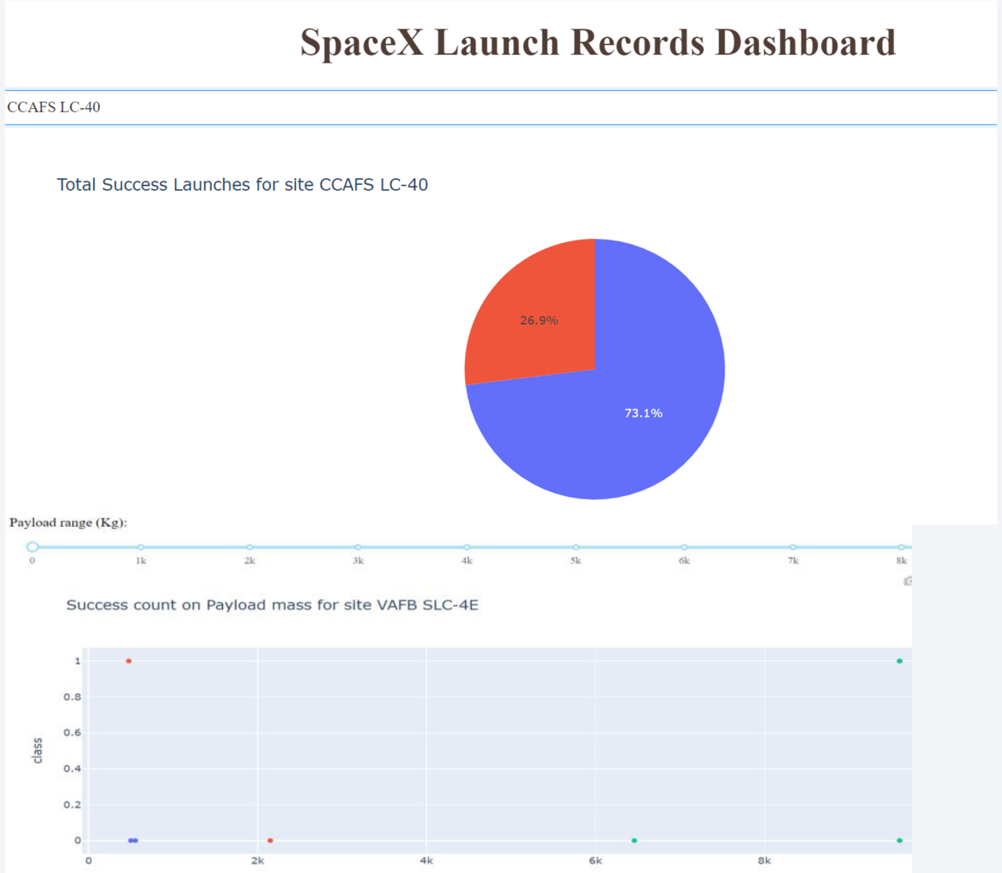
- SQL Queries performed-
  - Names of the unique launch sites in the space mission
  - Top 5 launch sites whose name begin with the string 'CCA’
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful 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
  - 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
- See details in [link]

Build an Interactive Map with Folium

- Markers, circles, lines and marker clusters were used with Folium Maps.
- Indications of each element:
  - Markers indicate points like launch sites
  - Circles indicate highlighted areas around specific coordinates like NASA Johnson Space Center
  - Marker clusters indicates groups of events in each coordinate like launches in a launch site
  - Lines are used to indicate distances between two coordinates.
- See details in [link]

Build a Dashboard with Plotly Dash

Dashboard



[LINK FOR CODE]

- **Elements**
  - Dropdown list for the launch site.
  - Range Slider for selecting the payload mass.
  - PieChart: for showing the success rate of each launch site, or showing the number of successful landing outcomes.
  - Scatterplot: Show success/failure by payload and booster version.

- **Findings:**
- Which site has the largest successful launches? KSC LC-39A.
- Which site has the highest launch success rate? KSC LC-39A (success rate76.9%).
- Which payload range(s) has the highest launch success rate? 2000-4000.
- Which payload range(s) has the lowest launch success rate? 6000-8000.
- Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest.
- launch success rate? B5 (only one successful start), apart from that FT (15successes, 8 failures)

Predictive Analysis (Classification)

- We create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs
  - Perform exploratory Data Analysis and determine Training Labels
  - Create a column for the class
  - Standardize the data
  - Split into training data and test data
- Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
- Find the method performs best using test data.
- GitHub URL- link

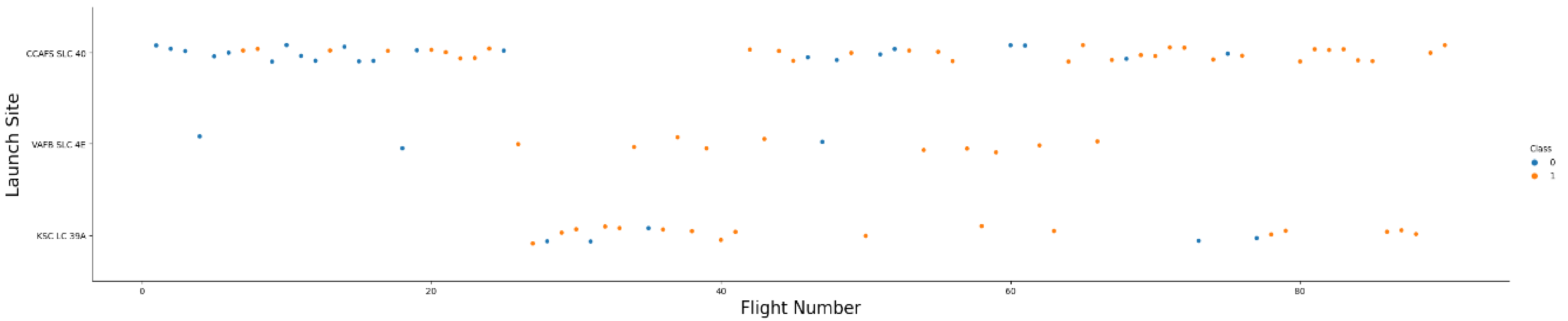
Results

- Exploratory data analysis results
  - Launch success rate increases over time
  - Higher success rate for higher orbits
- Interactive analytics demo in screenshots
  - Higher success rate for higher payload mass
  - Low success rate for booster versions v1.0, v1.1, high success rate for FT, B4, B5
  - Higher success rate for Kennedy Space center and recent starts at CapeCanaveral
- Predictive analysis results
  - Best prediction results with KNN and Support Vector Machine

Section 2

Insights drawn from EDA

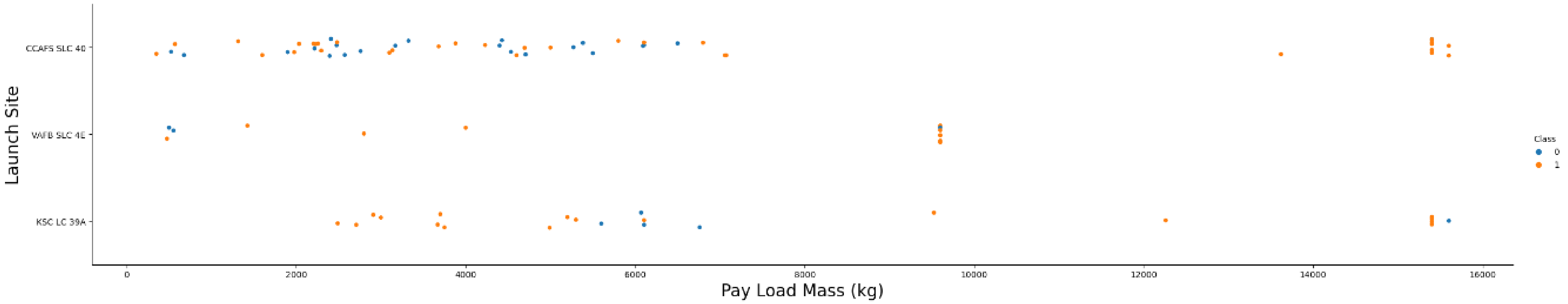
Flight Number vs. Launch Site



- Explanations

We can see that the CCAFS LC-40 launch site has more attempts than KSC LC-39A and VAFB SLC 4E.We cansee that the CCAFS LC-40 launch site has more attempts than KSC LC-39A and VAFB SLC 4E.

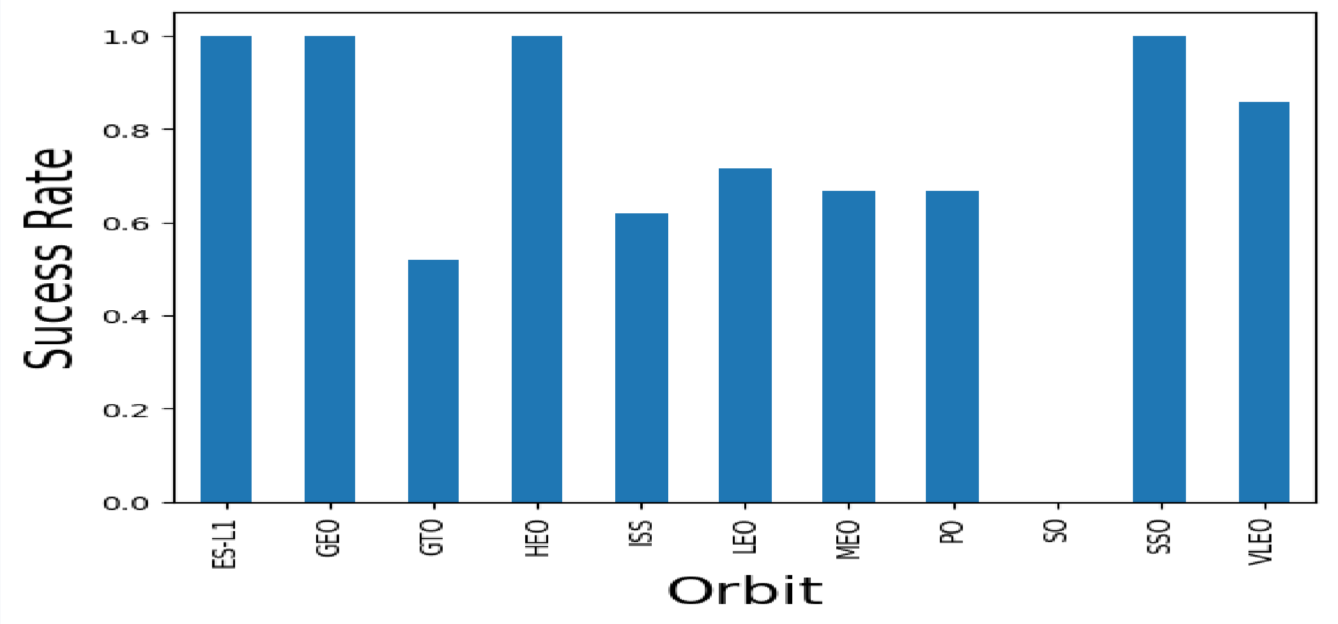
Payload vs. Launch Site



- Explanation

Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there areno rockets launched for heavypayload mass(greater than 10000).

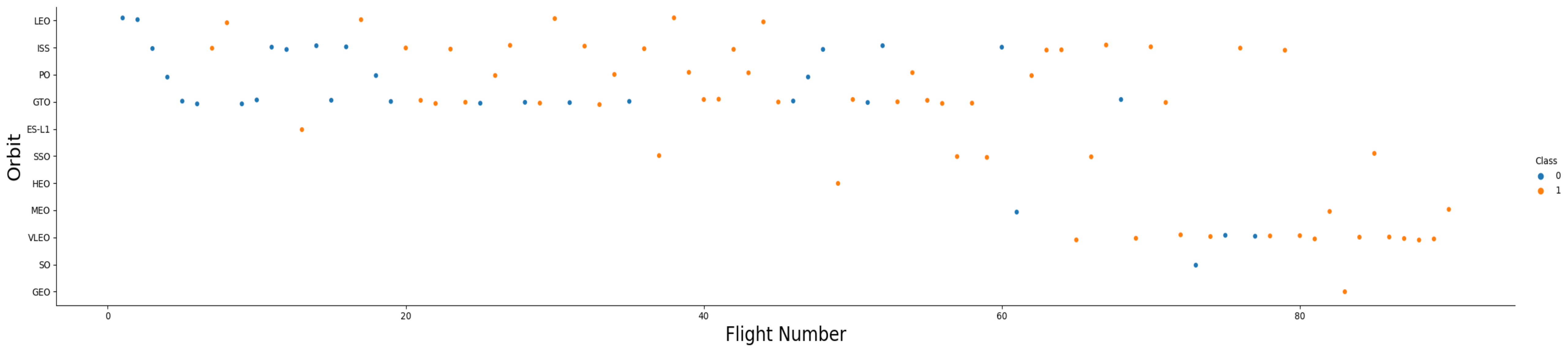
Success Rate vs. Orbit Type



Explanations

- Low Earth Orbits-GTO; ISS; LEO; MEO; PO; VLEO
- High Earth Orbits- ES-L1; GEO; HEO;SSO

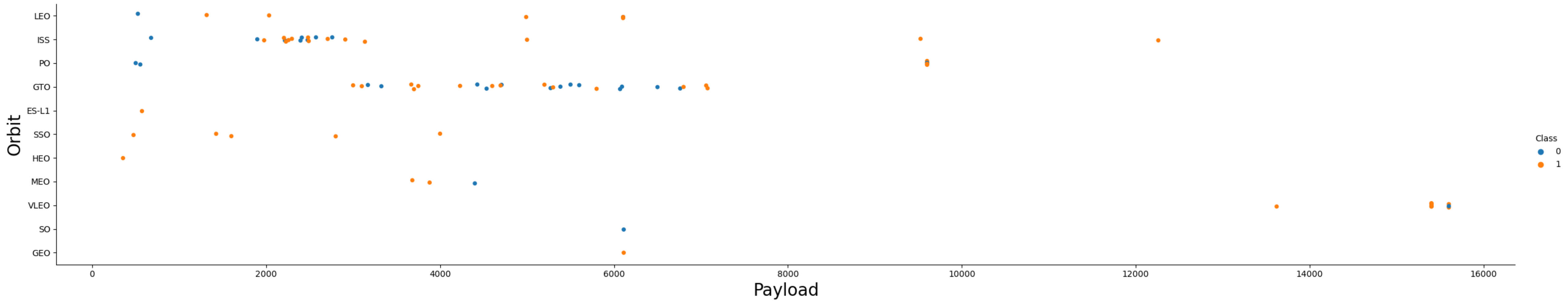
Flight Number vs. Orbit Type



Explanations

- You should see that 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.

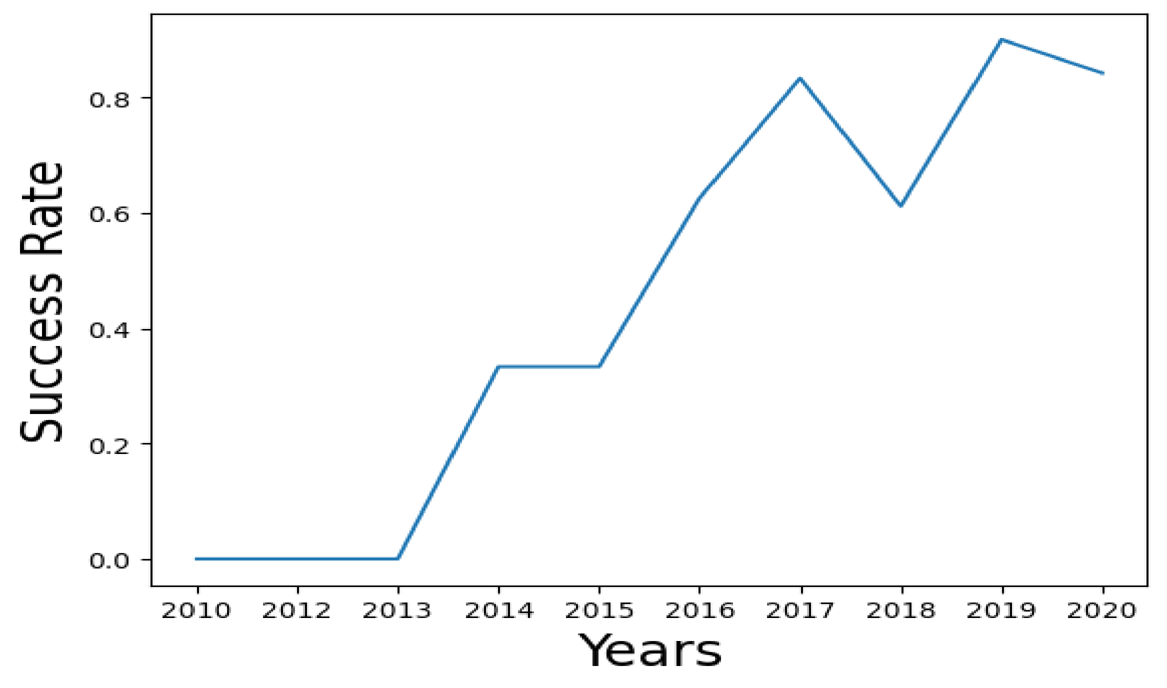
Payload vs. Orbit Type



Explanations

- With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here

Launch Success Yearly Trend



Explanations

- you can observe that the success rate since 2013 kept increasing till 2020

All Launch Site Names

- Unique Launch Sites-
  - CCAFS LC-40
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40
  - None
- Query

```
results = %sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL;
print(results)
```

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with the string 'CCA'CCAFS LC-40

```
----------------+-----------+------------------------------------+--------------------+------------------------+
|    Date     | Time (UTC) | Booster_Version | Launch_Site |                          Payload                          | P
YLOAD_MASS__KG_ |   Orbit   |                Customer               |   Mission_Outcome   |    Landing_Outcome      |
+------------+----------+------------------+-------------+-------------------------------------------------------+--
----------------+-----------+------------------------------------+--------------------+------------------------+
| 06/04/2010 |  18:45:00   |  F9 v1.0  B0003 | CCAFS LC-40 |           Dragon Spacecraft Qualification Unit          |
0.0          |    LEO    |                 SpaceX                |       Success       |  Failure (parachute)   |
| 12/08/2010 |  15:43:00   |  F9 v1.0  B0004 | CCAFS LC-40 | Dragon demo flight C1, two CubeSats, barrel of Brouere cheese |
0.0          | LEO (ISS) |            NASA (COTS) NRO            |       Success       |  Failure (parachute)   |
| 22/05/2012 |  7:44:00    |  F9 v1.0  B0005 | CCAFS LC-40 |                  Dragon demo flight C2                  |
525.0          | LEO (ISS) |               NASA (COTS)              |       Success       |       No attempt       |
| 10/08/2012 |  0:35:00    |  F9 v1.0  B0006 | CCAFS LC-40 |                       SpaceX CRS-1                       |
500.0          | LEO (ISS) |                NASA (CRS)              |       Success       |       No attempt       |
| 03/01/2013 |  15:10:00   |  F9 v1.0  B0007 | CCAFS LC-40 |                       SpaceX CRS-2                       |
677.0          | LEO (ISS) |                NASA (CRS)              |       Success       |       No attempt       |
```

- Query

```
results = %sql SELECT * FROM SPACEXTBL where Launch_Site LIKE 'CCA%'LIMIT 20;
print(results)
```

Total Payload Mass

- Total payload carried by boosters from NASA-
  - 45596.0 Kg

- Query

```
In [15]: results = %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Customer='NASA (CRS)';
         print(results)

 * sqlite:///my_data1.db
Done.
+-----------------------+
| SUM(PAYLOAD_MASS__KG_) |
+-----------------------+
|        45596.0        |
+-----------------------+
```

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
  - 340

- Query
- %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXDATASET WHERE Booster_Version LIKE 'F9 v1.0%';

First Successful Ground Landing Date

- Dates of the first successful landing outcome on ground pad
  - 01/08/2018

- Query

```
In [17]: results = %sql SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome='Success (ground pad)';
         print(results)

 * sqlite:///my_data1.db
Done.
+------------+
| MIN(Date)  |
+------------+
| 01/08/2018 |
+------------+
```

Successful Drone Ship Landing with Payload between 4000 and 6000

- names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
In [18]: results = %sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
         print(results)

 * sqlite:///my_data1.db
Done.
+-----------------+
| Booster_Version |
+-----------------+
|     F9 v1.1     |
|  F9 v1.1 B1011  |
|  F9 v1.1 B1014  |
|  F9 v1.1 B1016  |
|   F9 FT B1020   |
|   F9 FT B1022   |
|   F9 FT B1026   |
|   F9 FT B1030   |
|  F9 FT  B1021.2 |
|  F9 FT B1032.1  |
|  F9 B4 B1040.1  |
|  F9 FT  B1031.2 |
|  F9 B4 B1043.1  |
|  F9 FT  B1032.2 |
|  F9 B4  B1040.2 |
|  F9 B5 B1046.2  |
|  F9 B5 B1047.2  |
|  F9 B5 B1046.3  |
|    F9 B5B1054   |
|  F9 B5 B1048.3  |
|  F9 B5 B1051.2  |
|   F9 B5B1060.1  |
|  F9 B5 B1058.2  |
|   F9 B5B1062.1  |
+-----------------+
```

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

```
+----------------------------------+-----------------------+
|         Mission_Outcome          | COUNT(Mission_Outcome) |
+----------------------------------+-----------------------+
|               None               |           0           |
|       Failure (in flight)        |           1           |
|             Success              |           98          |
|             Success              |           1           |
| Success (payload status unclear) |           1           |
+----------------------------------+-----------------------+
```

- Query

```
In [23]:
results = %sql SELECT Mission_Outcome,COUNT(Mission_Outcome) FROM SPACEXTBL GROUP BY Mission_Outcome;
print(results)
```

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

```
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
```

- Query

```
In [24]:
results = %sql SELECT Booster_Version FROM SPACEXTBL where PAYLOAD_MASS__KG_=(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
print(results)
```

2015 Launch Records

- Failed landing outcomes in 2015

```
+-------+----------------------+---------------+
| month |    Landing_Outcome   |  Launch_Site  |
+-------+----------------------+---------------+
|   10  | Failure (drone ship) | F9 v1.1 B1012 |
|   04  | Failure (drone ship) | F9 v1.1 B1015 |
+-------+----------------------+---------------+
```

- Present your query result with a short explanation here

```
In [25]:
results = %sql SELECT substr(Date, 4, 2) month, Landing_Outcome,Booster_Version Launch_Site FROM SPACEXTBL where Landing_Out
print(results)
```

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

LANDING__OUTCOME	TOTAL_NUMBER
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3

- Query

- %%sqlSELECT LANDING__OUTCOME, COUNT(LANDING__OUTCOME) AS TOTAL_NUMBER FROM SPACEXDATASETWHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOMEORDER BY TOTAL_NUMBERDESCRank Landing Outcomes Between 2010-06-04 and 2017-03-20

Section 3

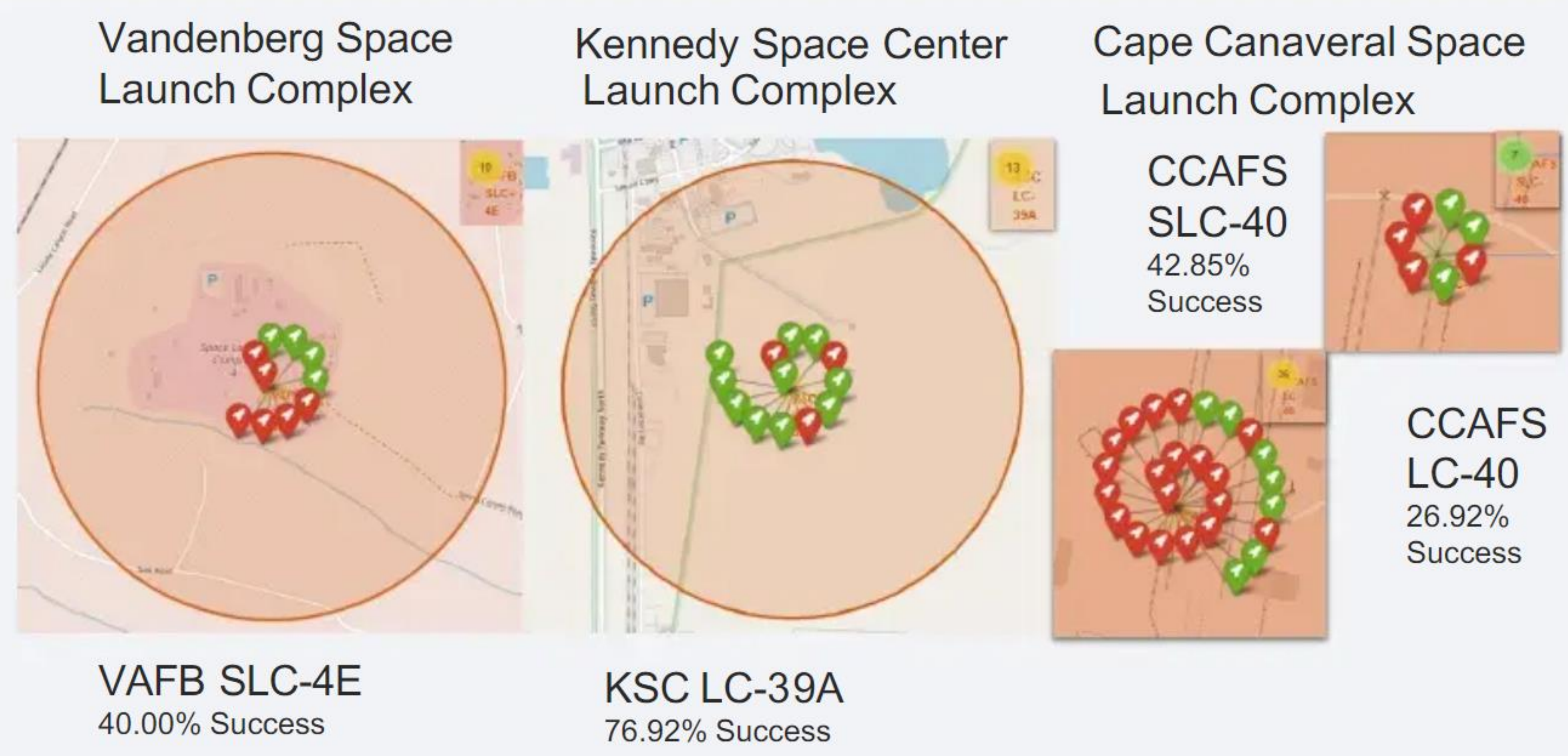
Launch Sites Proximities Analysis

Folium Map- Launch Sites

- Launch sites are at the East and West coast, near the southernmost U.S.mainland area, ;which is Florida and; California

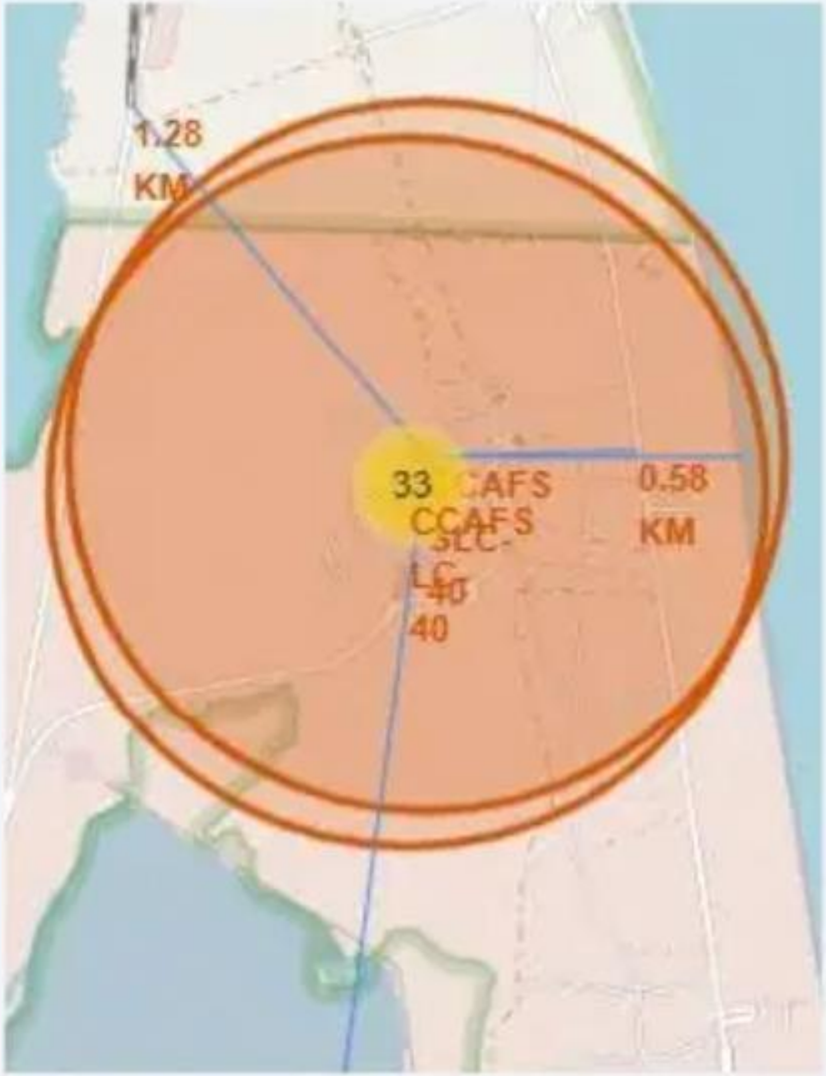
CCAFS [Cape Canaveral Space Launch Complex](#)
KSC [Kennedy Space Center Launch Complex](#)
VAFB [Vandenberg Space Launch Complex](#)

Folium Map- Stage 1 Landing success



Folium Map- Logistics

- Launch site KSC LC-39A has good logistics aspects, being near railroad and road
- Also it is far from inhabited areas



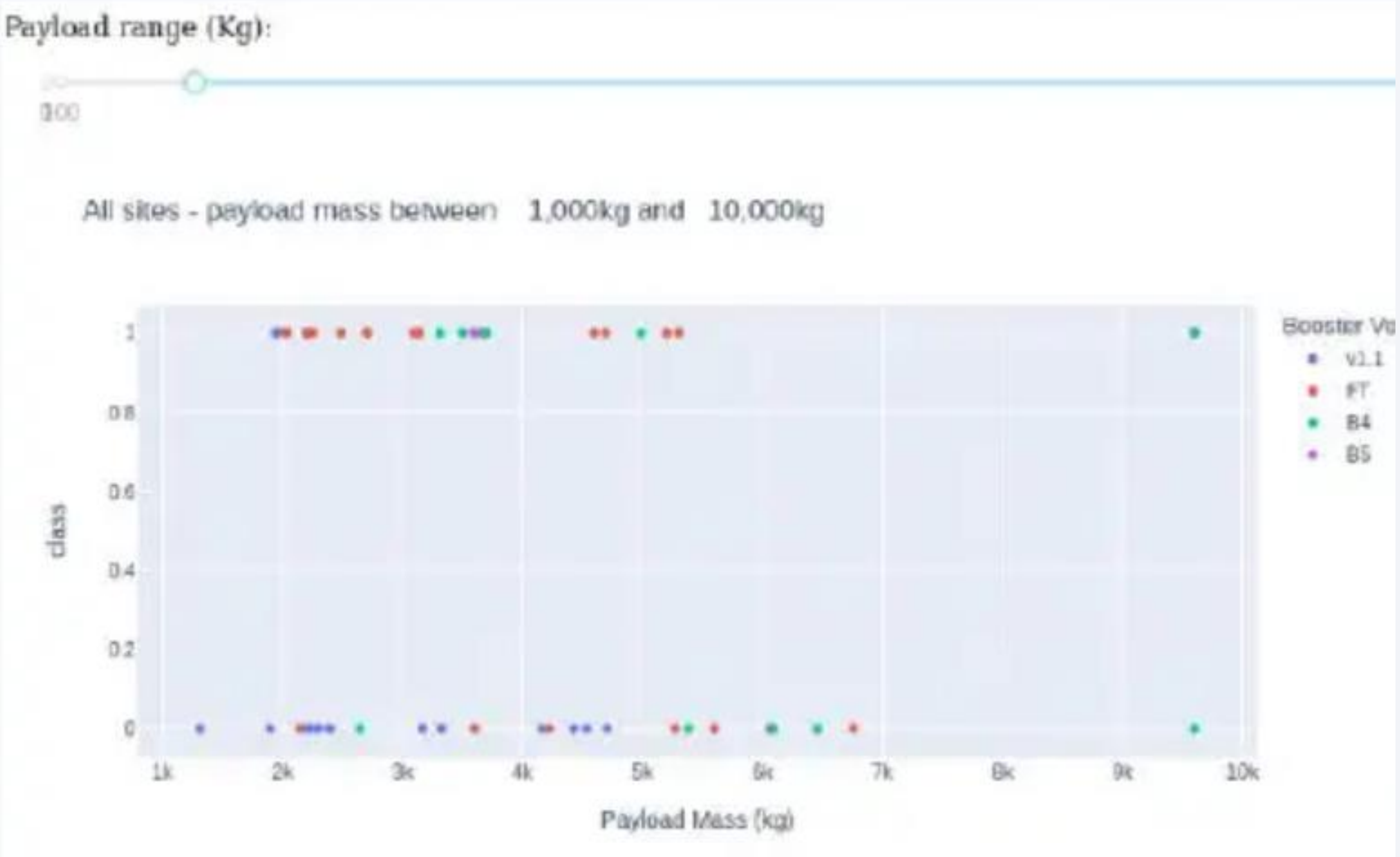
Section 4

Build a Dashboard with Plotly Dash

Dashboard- Success for all sites

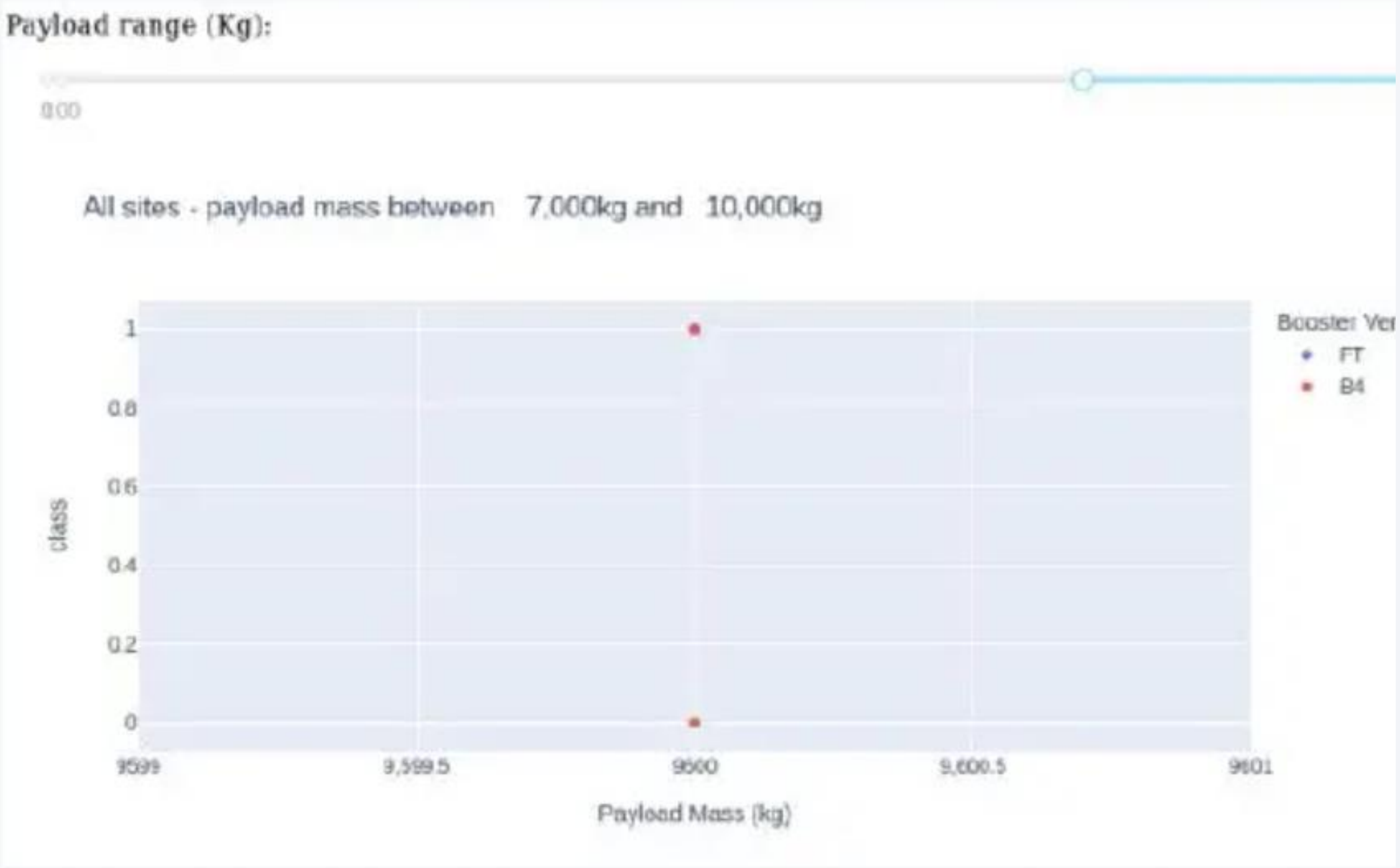
- Kennedy Space Center (KSCLC-39A) hast the most successful stage-1 landings
- Vandenberg Air Force Base(VAFB SLC-4E) has the least number of successful stage-1landings

Dashboard- Payload vs launch outcome



- Payloads under 6,000kgand FT boosters are the most successful combination.

Payload vs launch Outcome

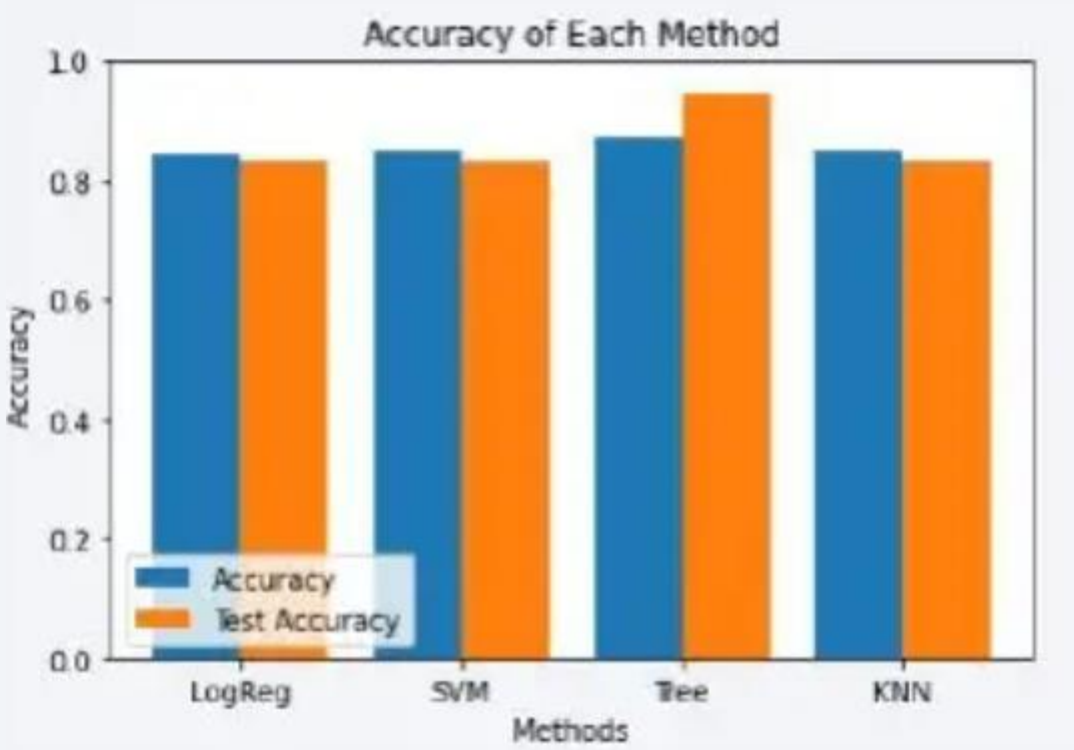


- There’s not enough data to estimate risk of launches over7,000kg

Section 5

Predictive Analysis (Classification)

Classification Accuracy



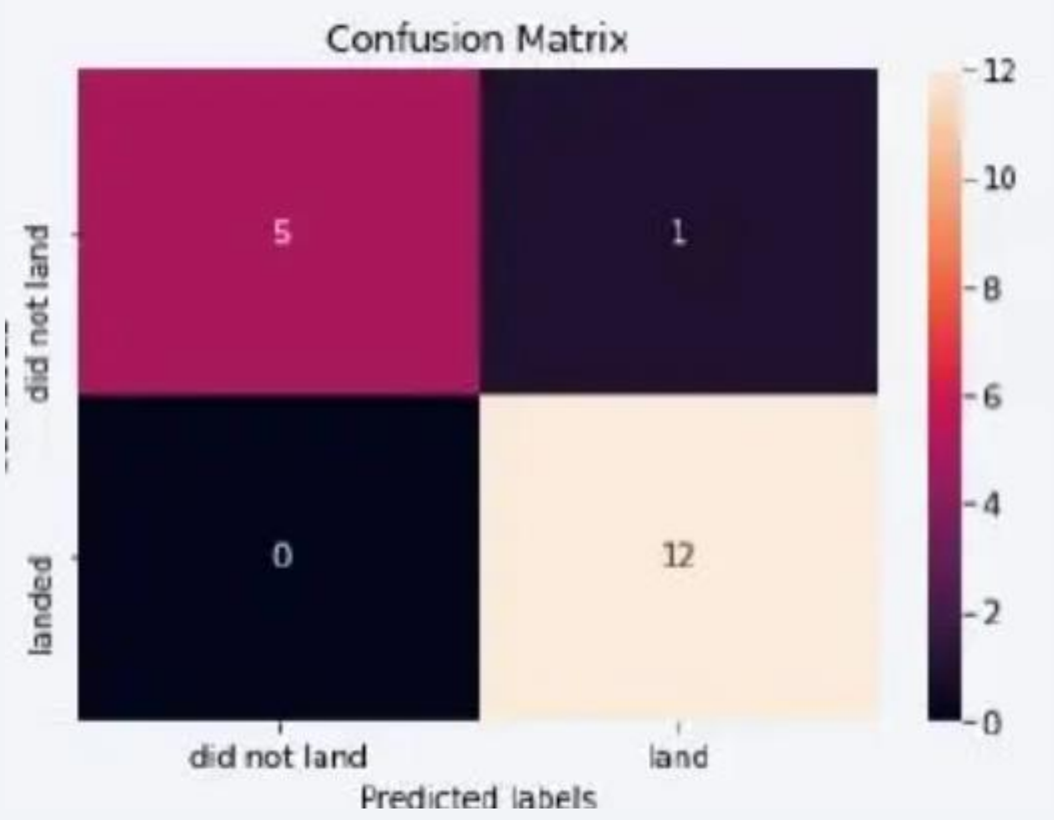
- Four classification models were tested and their accuracies are plotted beside

- The model with the highestclassification accuracy is Decision TreeClassifier, which has accuracies overthan 87%

Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.

True Positives	12
True Negatives	5
False Positives	1
False Negatives	0



Conclusions

- The best launch site is KSC LC-39A
- Launches above 7,000kg are less risky
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets
- None of the models had false negatives
- All models had at least one false positive
- Prediction with Logistic Regression is quite accurate•
- Support Vector Machine also provide a good result for predicting the landing outcome

Appendix

- You can see references and details of my project on this link

Thank you!