Winning Space Race with Data Science

Muhammad Shafwan Faturrahman
5 October 2021

Outline

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

Executive Summary

- Summary of methodologies

  Data gathered from the website of SpaceX and Wikipedia. Select the data of Falcon 9 only. Then fix the data to handle the missing values. Use SQL and visualization library to found out some information of its launches. Make a dashboard by using plotly dash. The last step is predict the success rate of rocket launches by using machine learning.

- Summary of all results

  Based on the research to its data, there is no best method to predict the success rate of rocket launch. All of the machine learning method that have been used has the same accuracy. The accuracy of its prediction is 0.8334.

Introduction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars. Other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. If we can determine that the first stage will land, we can determine the cost of a launch.
- Problems you want to find answers
  - Find out the launches development by launches trend for each year.
  - Find the best sites to launch a rocket.
  - Make the best model of machine learning that can predict whether the rocket will successfully launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

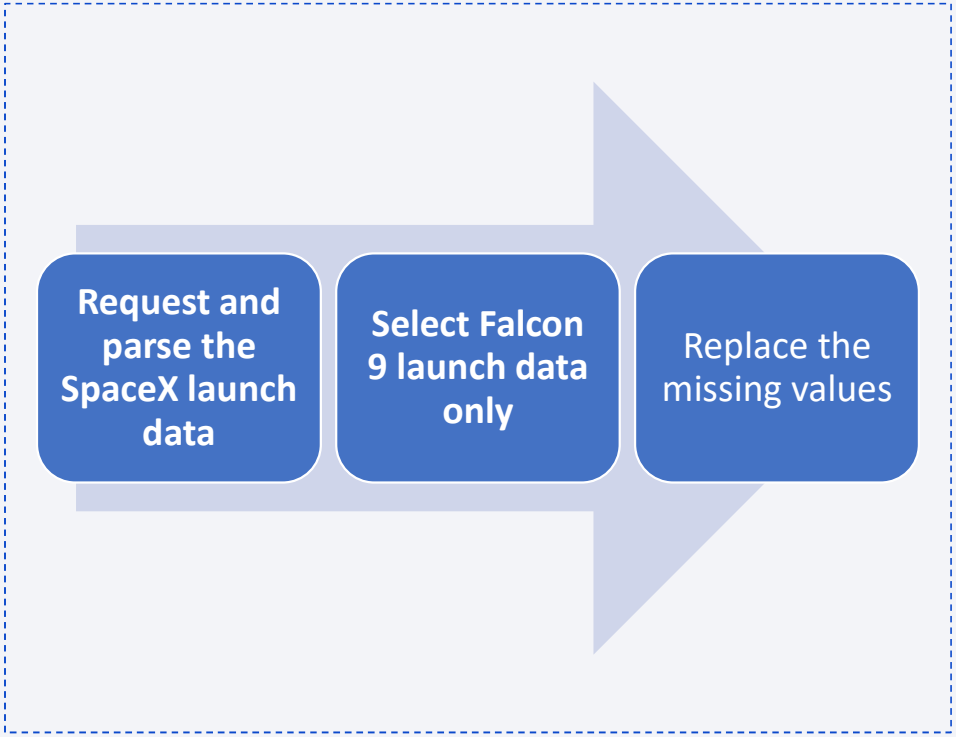
Data Collection

Datasets collected from 2 sources, the API of SpaceX official website and Wikipedia.

Data Collection – SpaceX API

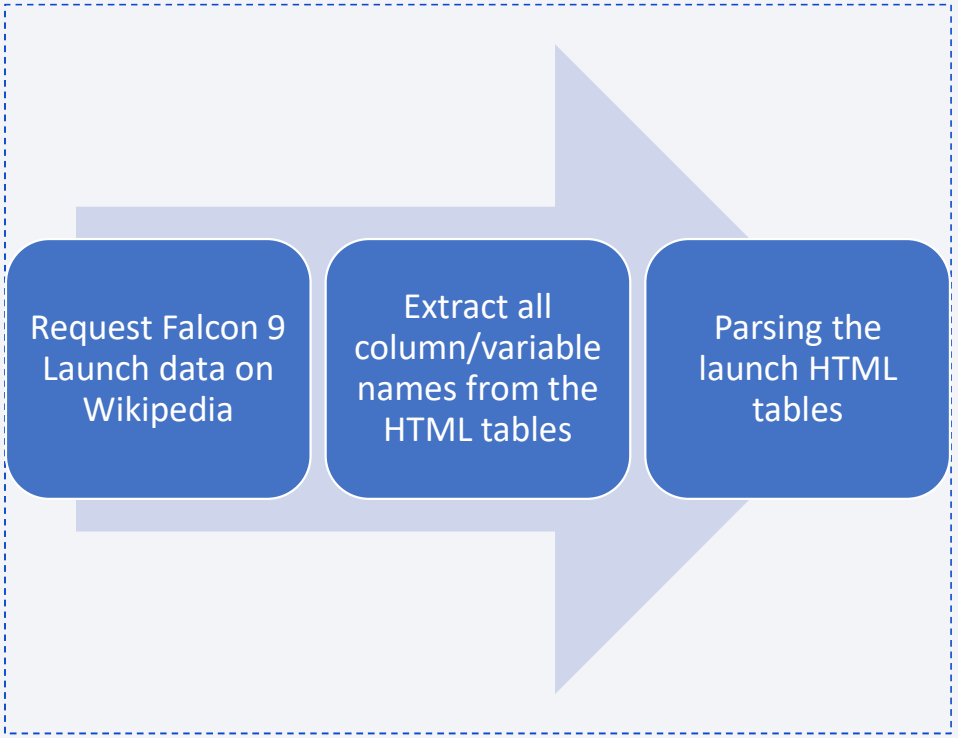
- Data was collected by using get request. Then filter the data to only include Falcon 9 Launches. Fix the missing values by replace it with the mean of the data.

- Data collection process: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/jupyter-labs-spacex-data-collection-api.ipynb



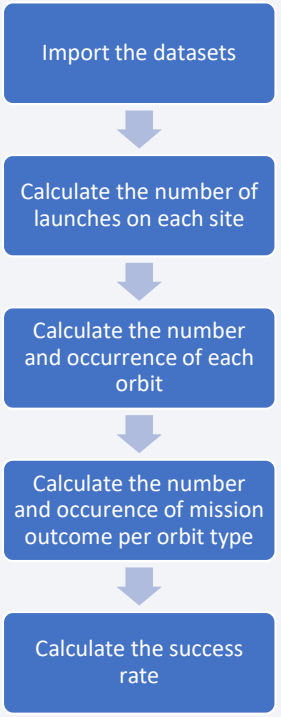
Data Collection - Scraping

- Data was collected by using get request on Wikipedia. Then the column/variable from the HTML tables. Get the data by parsing the launch HTML tables.
- Data collection process: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/jupyter-labs-webscraping.ipynb



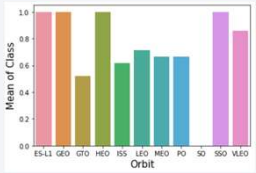
Data Wrangling

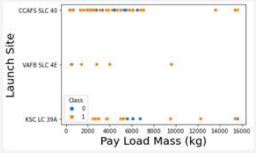
- First, import the datasets using pandas. Then doing the process on the flowchart to get the results. The forth process is used to get the new column class that will be used to calculate the success rate. The success rate of the launches is 0.667.
- Data wrangling process: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/labs-jupyter-spacex-Data%20wrangling.ipynb

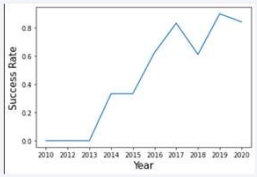


EDA with Data Visualization

- In this process, I used three kind of chart to visualize the datasets. In general, scatter point chart is used to show the relationship between two features. Bar chart is used to show some value like count, mean or average for each categorical data. Line chart is used to show the trend of some feature in time series.







- Scatter Point Chart: Flight Number and Launch Site, Payload and Launch Site, FlightNumber and Orbit type, Payload and Orbit type
- Bar Chart: Success rate of each orbit type
- Line Chart: Success Launch yearly trend

- Exploratory data analysis process: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- In this process, I used SQL queries to analyze the data. Here is the point I search:
  - Display the names of the unique launch sites in the space mission.
  - Display 5 records where launch sites begin with the string 'CCA'.
  - Display the total payload mass carried by boosters launched by NASA (CRS).
  - Display average payload mass carried by booster version F9 v1.1.
  - List the date when the first successful landing outcome in ground pad was achieved.
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
  - List the 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.
  - List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
  - 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.
- Exploratory data analysis process: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/EDA%20with%20SQL%20lab.ipynb

Build an Interactive Map with Folium

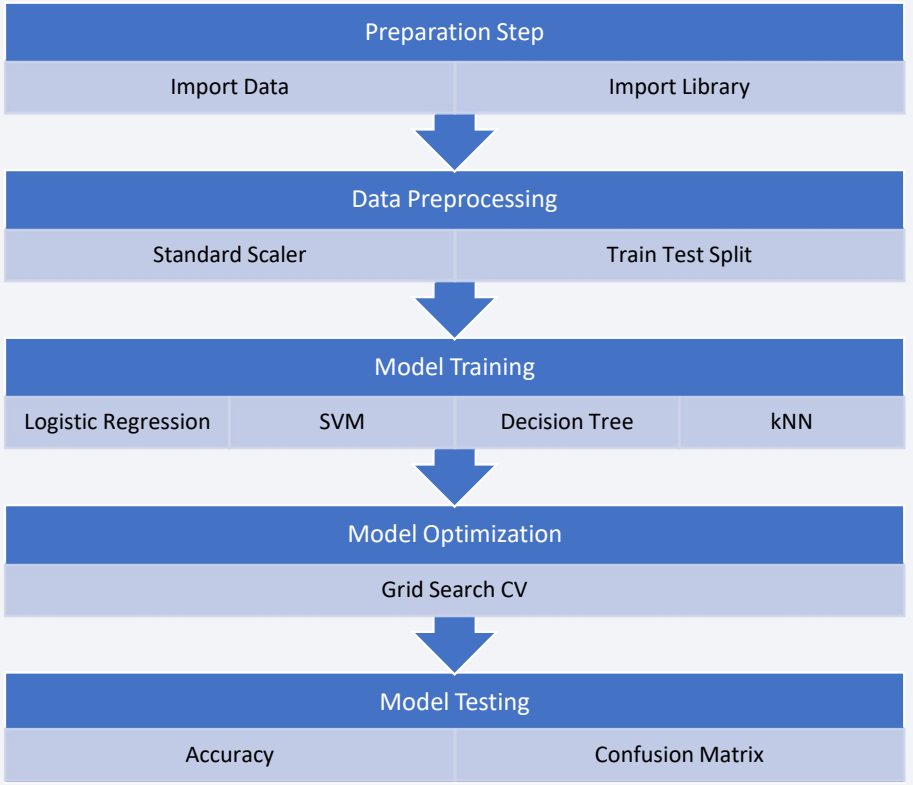
- The goal of this process are Mark all launch sites on a map, mark the success/failed launches for each site on the map and determine the distances between a launch site to its proximities.
  - Markers was used to point a specific location on maps.
  - Circles was used to round the selected location on maps.
  - Polyline was used to make a line between two selected point on maps.
- Interactive map with folium: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- In the dashboards, I present two plots about the percentage of success launch using pie chart and the success launch based on the payload mass using scatter plot.
  - Pie-Chart: The plot has two function. The first one is to point out the percentage of number success launch for all sites. And the second is to point out the percentage of success launch for specific sites. This two mode can be control by using the dropdown feature on the dashboards.
  - Scatter-Plot: This plot used to point out the success launch based on its payload mass. The color of its plot indicate the different type of booster version category. The payload mass range of its plot can be control by using the slider on the dashboards.
- Dashboards code: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/spacex_dash_app.py

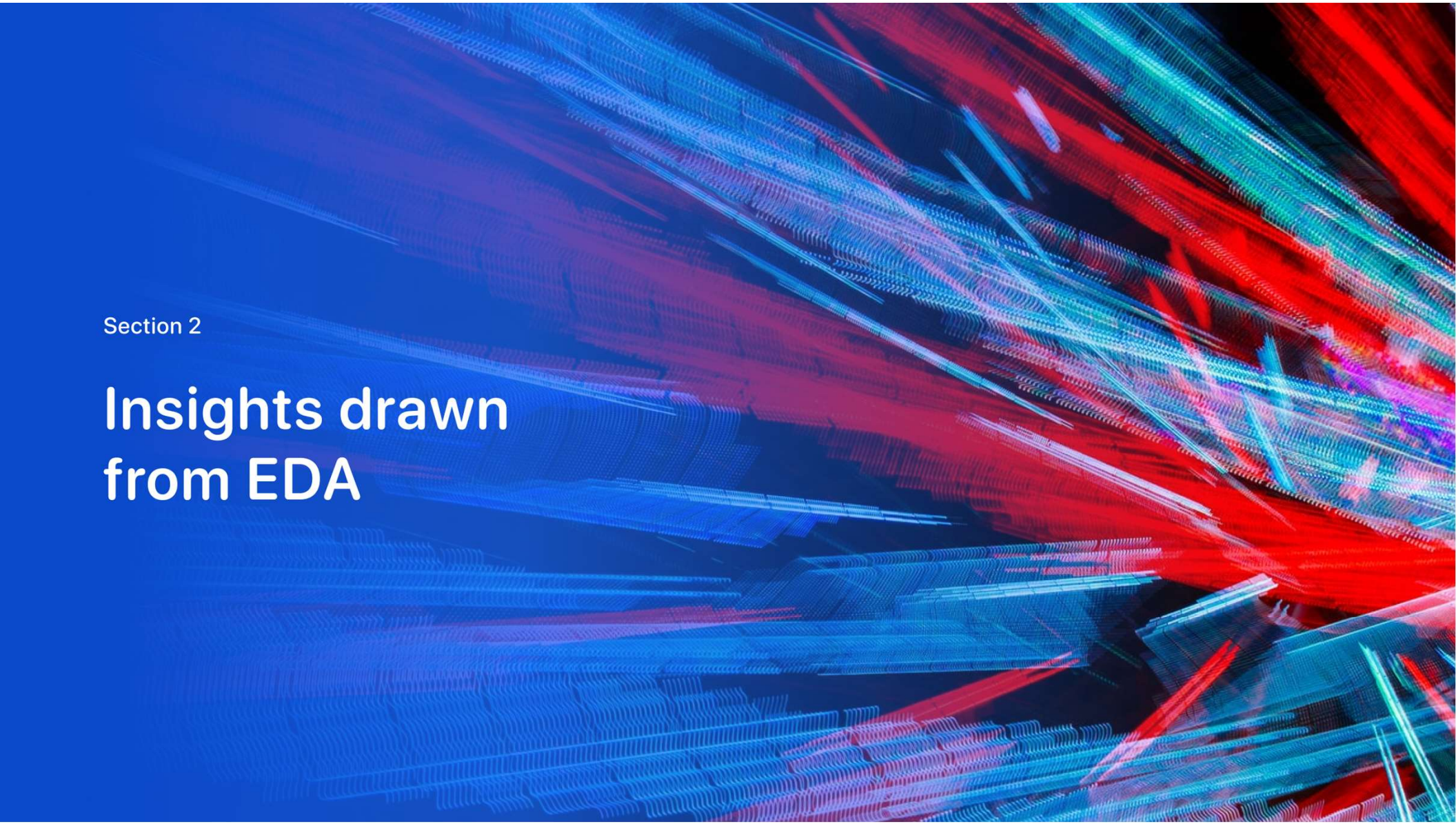
Predictive Analysis (Classification)

- Data used for this analysis is SpaceX data of Falcon 9 only. First the data is processed by standard scaler so it has optimize range of data. Then data was split for train and test. Then train the data with some machine learning algorithm as shown in right side. Grid Search CV used to optimize the model so it can give the best hyperparameter of the model. Finally test the model to get the accuracy and confusion matrix of the results.
- Predictive analysis process: https://github.com/msf-coursera/Coursera_Capstone/blob/main/Final%20Project/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb



Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

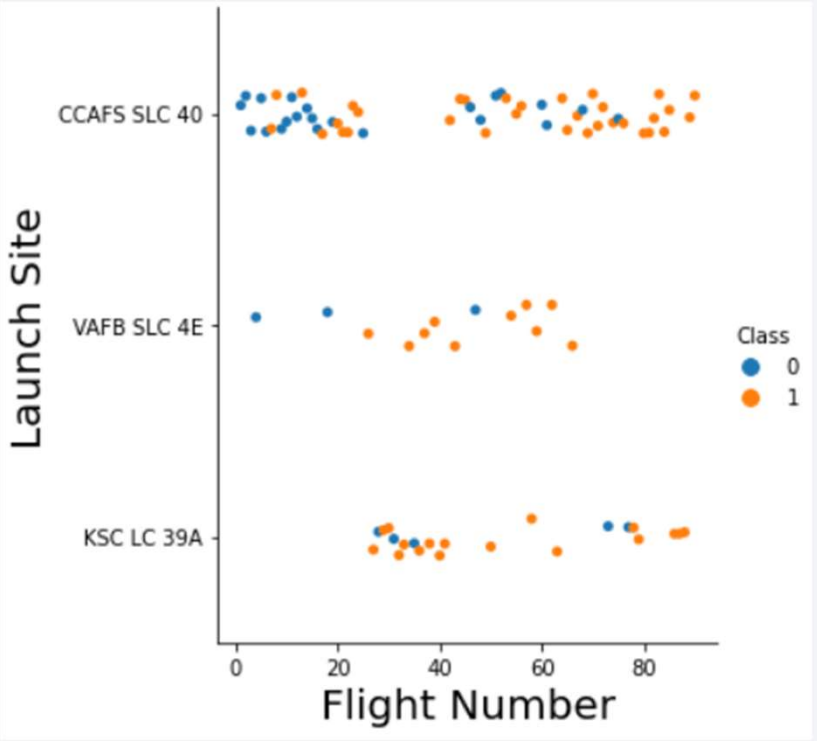


Section 2

Insights drawn from EDA

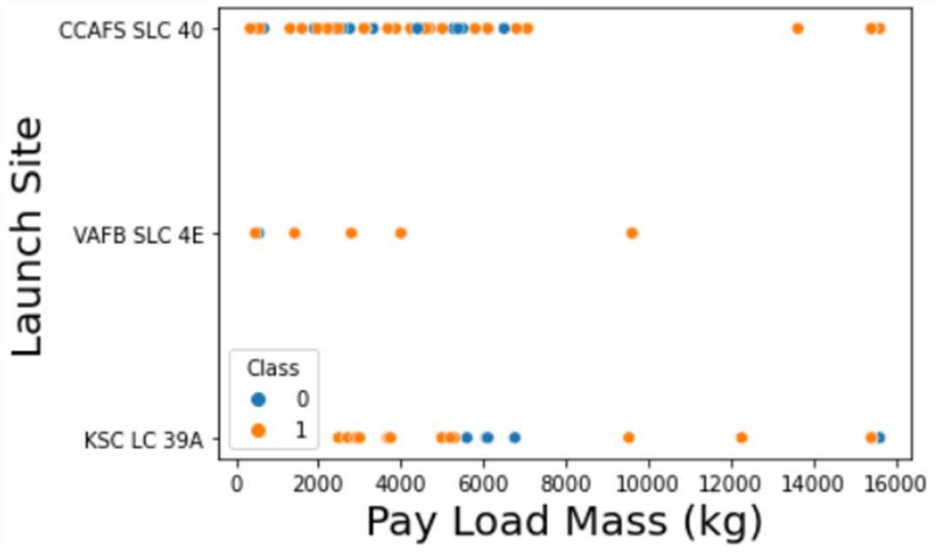
Flight Number vs. Launch Site

Based on the figure, we can see that KSC LC 39A have the best percentage of success rate.



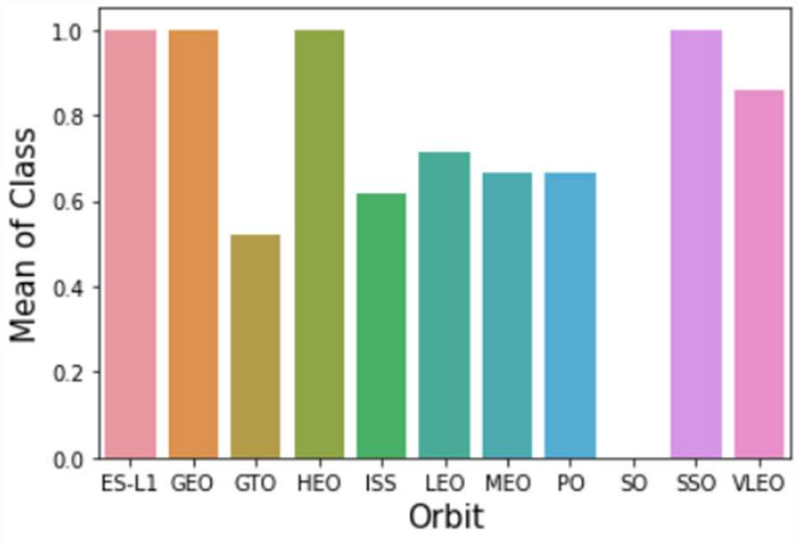
Payload vs. Launch Site

- Most of its launches that have payload mass over 8000kg is successful.
- Most of its launches have the payload mass range from 0-6000kg.



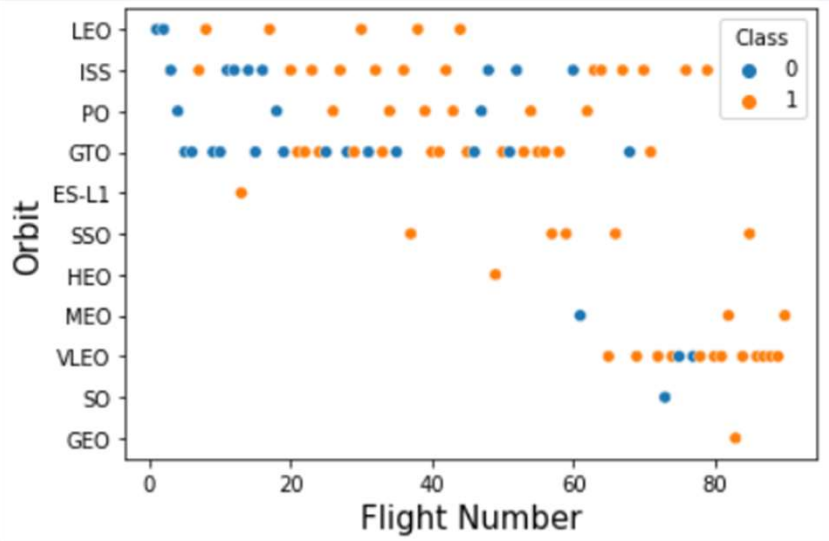
Success Rate vs. Orbit Type

The most successful orbit to launch a rocket are ES-L1, GEO, HEO and SSO



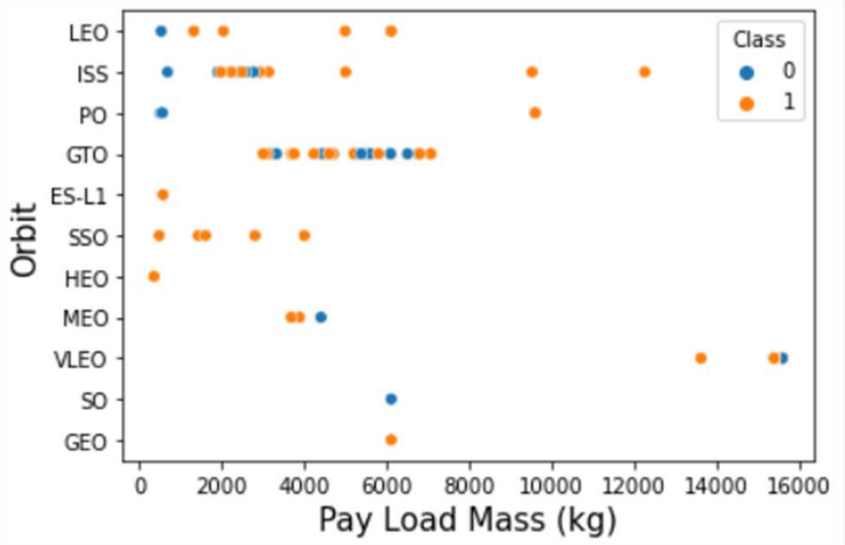
Flight Number vs. Orbit Type

Based on the figure, we did not launch the rocket to LEO anymore and it is been changed to VLEO and any other orbit. The one that has been used from then until now is ISS.



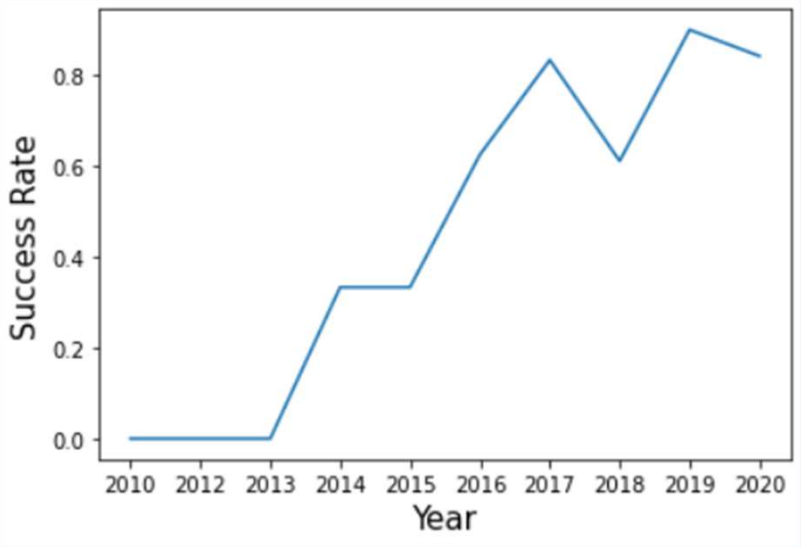
Payload vs. Orbit Type

- There are just a few of rocket which launch with high payload mass.
- ISS is the only orbit which can handle rocket launch on variety payload mass.



Launch Success Yearly Trend

Launch success trend has been increased since 2013. In spite of its instability development, overall it launches better gradually.



All Launch Site Names

```
%%sql
SELECT DISTINCT launch_site
FROM spacextbl
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

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

There are 4 different sites that used to launch a rocket.

Launch Site Names Begin with 'CCA'

```
%%sql
SELECT DISTINCT launch_site
FROM spacextbl
WHERE launch_site LIKE 'CCA%'
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

launch_site
CCAFS LC-40
CCAFS SLC-40

There are only 2 site names which start with ‘CCA’.

Total Payload Mass

```
%%sql
SELECT SUM(payload_mass__kg_)
FROM spacextbl
WHERE customer LIKE 'NASA (CRS)'
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

1
45596

The total payload mass by NASA (CRS) is 45596 kg.

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(payload_mass__kg_)
FROM spacextbl
WHERE booster_version LIKE 'F9 v1.1'
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

1
2928.400000

Average payload mass which used the booster version of F9 v1.1 is 2928.4 kg.

First Successful Ground Landing Date

```
%%sql
SELECT MIN(date)
FROM spacextbl
WHERE landing__outcome LIKE 'Success (ground pad)'
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

1
2015-12-22

The first successful ground landing happen on december 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT customer
FROM spacextbl
WHERE landing__outcome LIKE 'Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

customer
SKY Perfect JSAT Group
SKY Perfect JSAT Group
SES
SES EchoStar

There are 4 times and 3 different customer which has successful drone ship landing with the payload mass between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

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

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

There are 100 success mission and 1 failed mission.

Boosters Carried Maximum Payload

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

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
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

There are 12 booster version carried on the maximum payload mass.

2015 Launch Records

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

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

There are 2 records of failure drone ship landed on 2015.

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

```
%%sql
SELECT landing__outcome, COUNT(*)
FROM spacextbl
WHERE date BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome
ORDER BY COUNT(*) DESC
```

```
 * ibm_db_sa://hnv97156:***@dashdb-txn-sbox-yp-dal09-08.services.dal.bluemix.net:50000/BLUDB
Done.
```

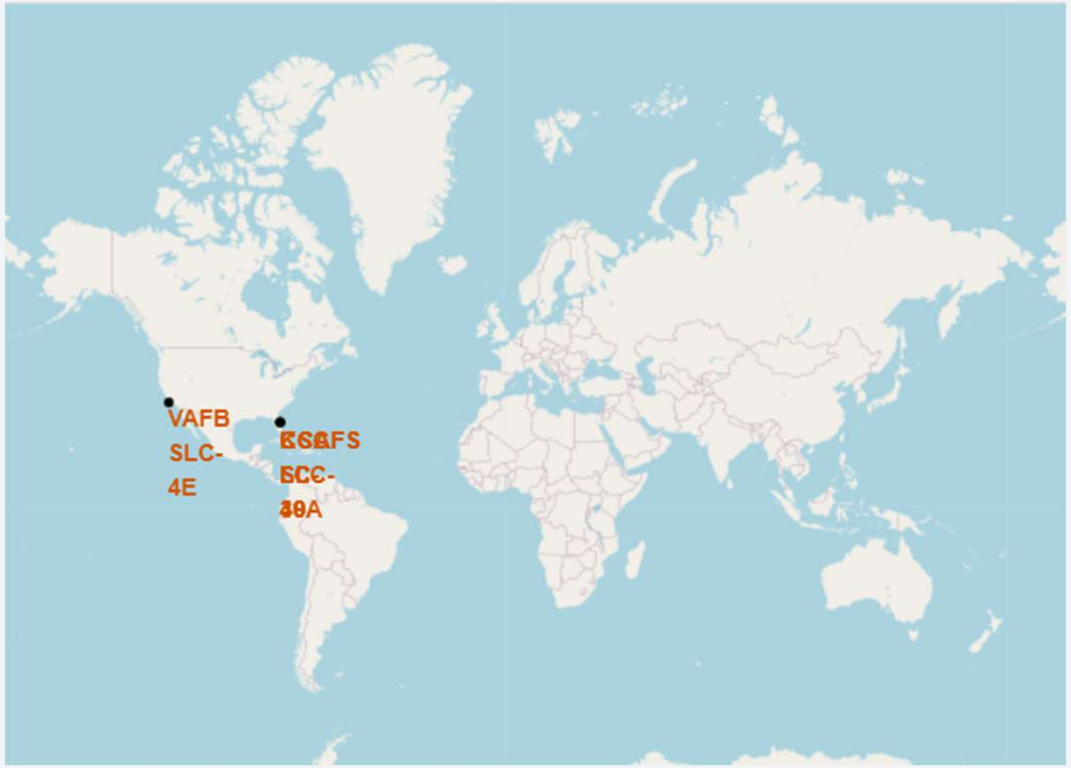
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

There are so many type of landing outcome that happen between 2010-06-04 and 2017-03-20.


Section 4
Launch Sites
Proximities Analysis

All Launch Sites on Global Map

- All launch sites are stated only in USA.
- There are four places that used for launch the falcon 9.
- The sites are:
  - VAFB SLC-4E
  - KSC LC-39A
  - CCAFS SLC-40
  - CCAFS LC-40



Success/Failed Launches for Each Sites

- VAFB SLC-4E (top-left)
  - Success/Failed : 4/6
- KSC LC-39A (top-right)
  - Success/Failed : 10/3
- CCAFS SLC-40 (bottom-left)
  - Success/Failed : 3/4
- CCAFS LC-40 (bottom-right)
  - Success/Failed : 7/19

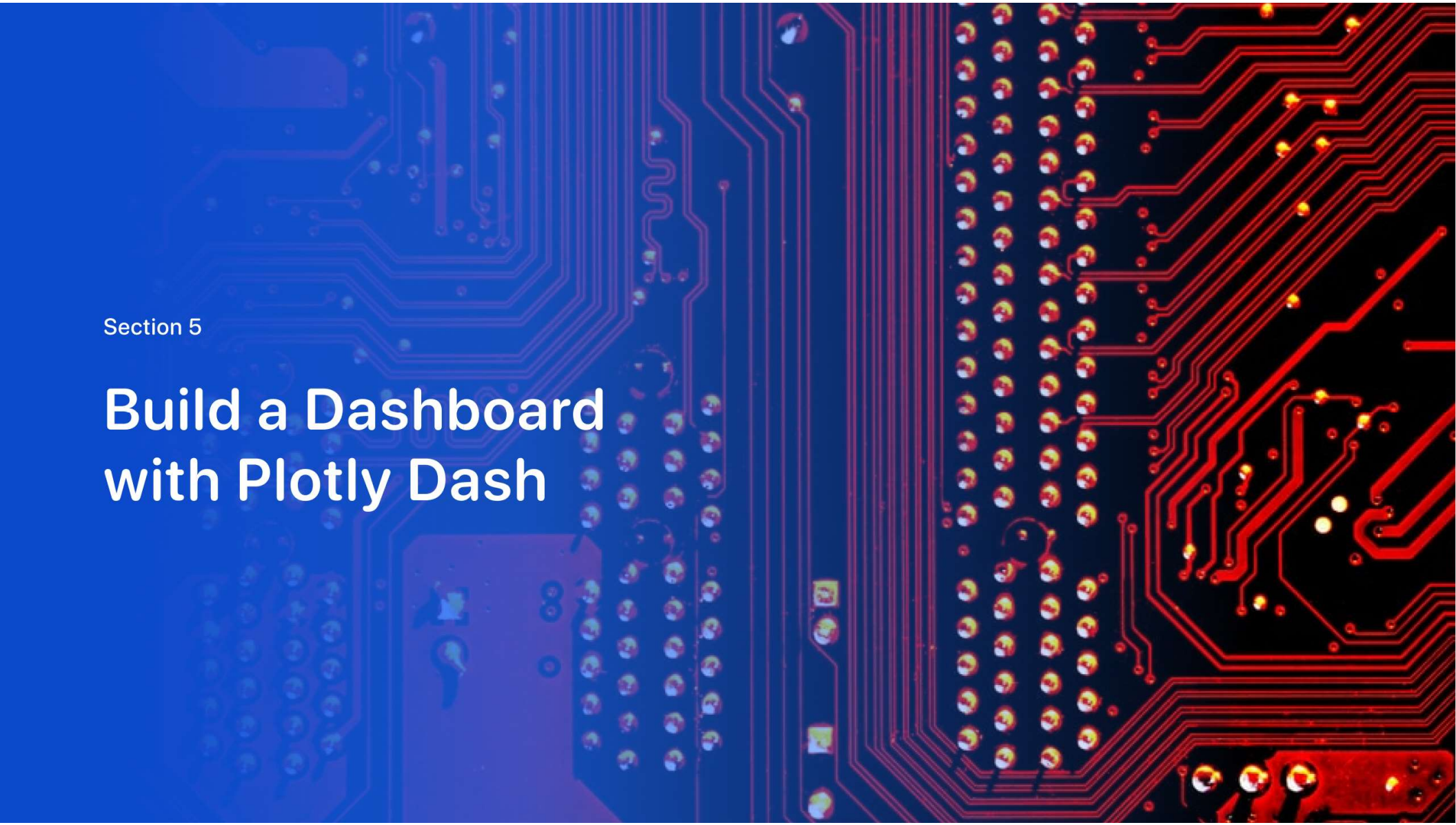
Different Distance of Launch Sites to its Proximities

- Highway (top-left)
  - Coordinates [lat/long]:

    [28.573254570378147, -80.64689528960382]
  - Distance: 5.444 km
- City (top-right)
  - Coordinates [lat/long]:

    [28.56230196799018, -80.57735647504778]
  - Distance: 17.401 km
- Coastline (bottom)
  - Coordinates [lat/long]:

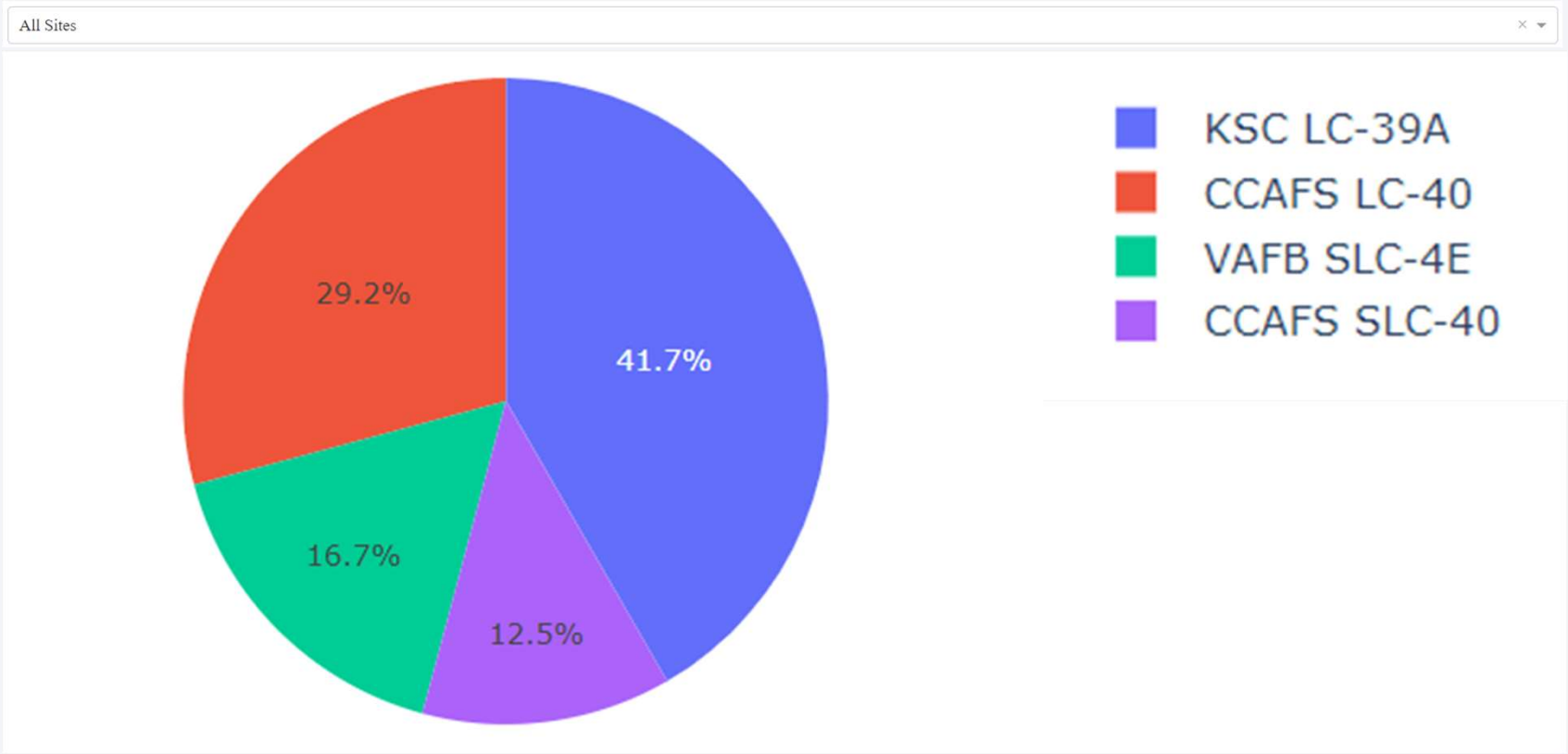
    [28.563197177407144, -80.57682003124195]
  - Distance: 0.858 km



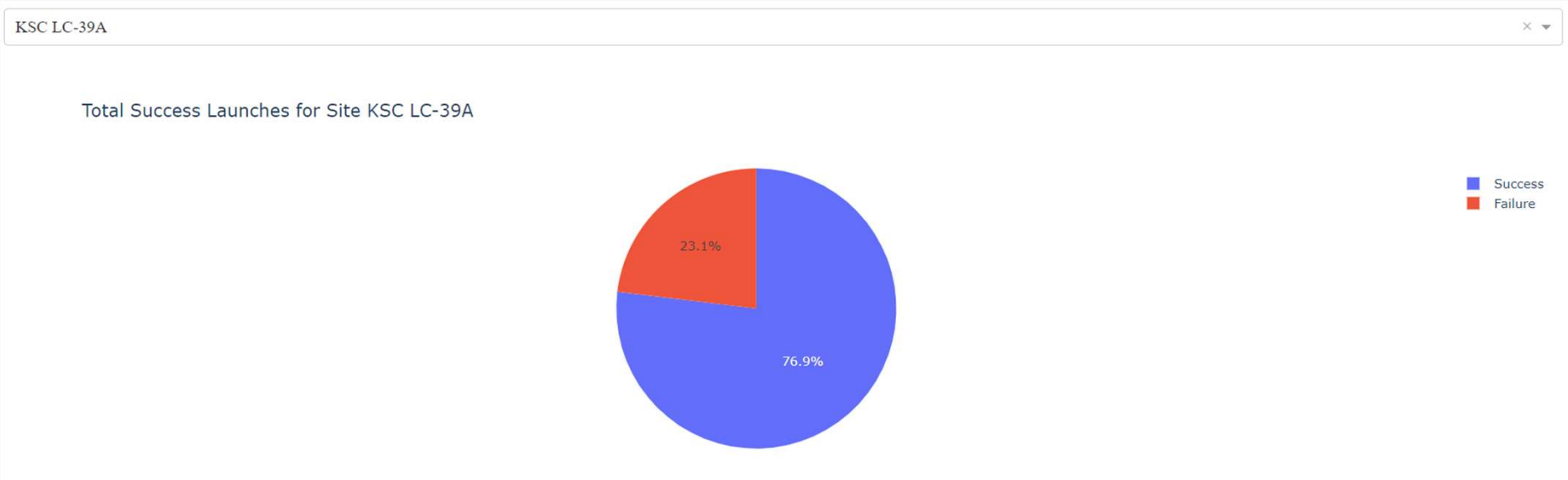
Section 5

Build a Dashboard with Plotly Dash

Dashboard of Total Success Launches for All Sites



Dashboard of Launch Site with Highest Launch Success Ratio



- Other success launch percentage for each sites
  - CCAFS LC-40 (73.1%)
  - VAFB SLC-4E (60%)
  - CCAFS SLC-40 (57.1%)

Dashboard of Payload Success Rate

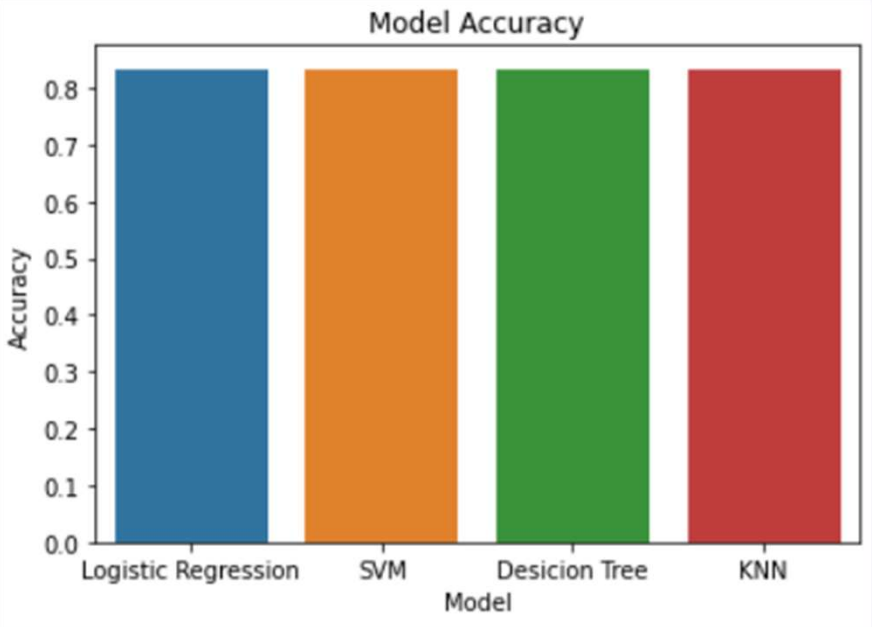
- This two plot shows success rate based on payload mass for each launches.
- First plot shows launch when the payload mass is between 0-5000kg, while the second plot is between 5000-10000kg.
- Most of the success launches when the payload is between 0-5000kg.
- Booster version by the most of the success launches is FT.

Section 6

Predictive Analysis (Classification)

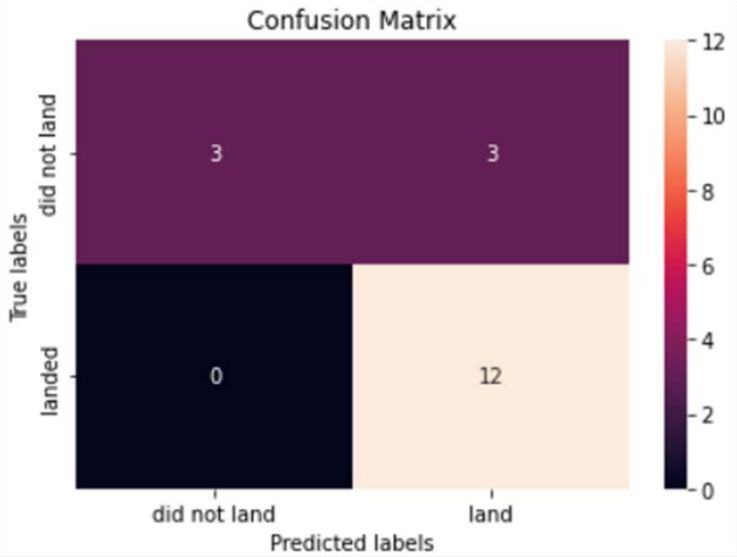
Classification Accuracy

- All of the model has the same accuracy of 0.834. Based on this data, there is no best model or worst model because the model has the same accuracy percentage.



Confusion Matrix

- Since the model has the same accuracy percentage, the confusion matrix of all the model has the same value.
  - True Positive: 12
  - False Negative: 0
  - True Negative: 3
  - False Positive: 3
- Accuracy = (12+3)/(12+3+3+0) = 0.8334



Conclusions

- Rocket launch development is getting better each year.
- The best sites to launch a rocket is KSC LC-39A
- There is no best model to predict the rocket launch. All of the model gives the same accuracy of 0.8334.

Appendix

- Github repository:

  https://github.com/msf-coursera/Coursera_Capstone/tree/main/Final%20Project

Thank you!