IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hasan ÜNLÜ
06/22/2023

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- In this project, data is collected by using both SpaceX REST API and Wikipedia Web page. Then, collected data is wrangled and the landing outcome determined into ‘Class’ attributte as 0 and 1 class which are successful landing for 1 and failed landing for 0. At next step, Explonatory Data Analysis is applied to get insights from the data set with both visualization methods and SQL queries. There are several correlations between attributes such as payload mass and landing success/fail as well as several other relationships. After these EDA results, data set is used to build an interactive dashboard and interactive maps. In the dashboard, it can be seen that success/fail landing rates on the pie chart for launch sites in the data set and also relationship between success situations of the booster versions on the each and every launch site depending on the payload mass that is carried by the boosters. Also, in the interactive map it can be seen that locations of the launch sites and succeded/failed landing attempts on the every launch location. As a result of interactive map, a launch site needs to be near to the coast, railway, and highways and needs to be far to the settlement areas. After data driven insights, 4 machine learning algorithms are used to obtain the best performed predictions. Those ML algorithms are LogisticRegression, SupportVectorMachine, DecisionTree and KNearestNeighbors classification algorithms with the GridSearch method. In order to use these ML algorithms data set is split into train and test data sets all models are trained with the train data sets and evaluated with the test data sets. In the evaluation stage, R squared method and F1 score evaluation methods are used. As a conclusion of the evaluation step, Decision Tree Classification method was the best performed method for this project with 88.89% success percentage.

Introduction

- In this project, it is aimed that find an answer to a question: a new designed or need to be designed rocket will be able to land properly?
- In order to answer this question some insights from the data set, that comes from the SpaceX company, need to be obtained as the first stage. These insights are about relationships between payload mass, orbit, booster version, launch sites and their locations on the map.
- After obtaining necessary relationships from the data set, 4 Machine Learning algorithms are used to have answer the question at the beginning. Used Machine Learning algorithms are LogisticRegression, SupportVectorMachine, DecisionTree and KNearestNeighbors. All of the ML models are trained by using the train data set that is obtained from the data set by splitting it into train and test data sets. After training the models, test data sets are used to get predictions from the models. All models' predictions are evaluated by using r_squared and f1 score fuctions. As a result, Decision Tree Classification Model is the best performed method with the with 88.89% success percentage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

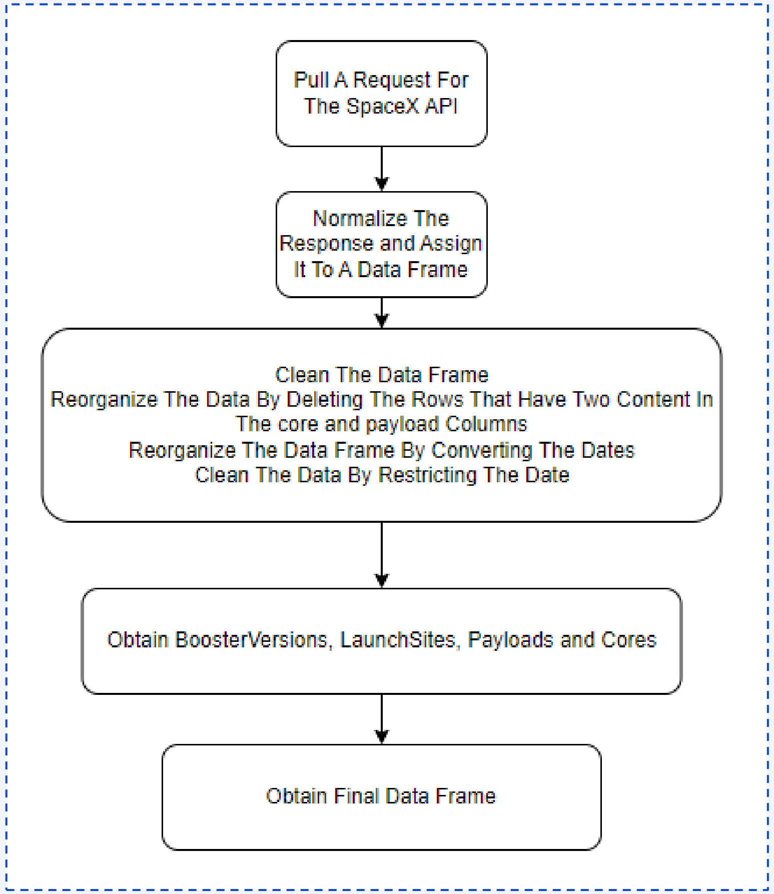
Data Collection

- There are two methods that are used to collect the data, SpaceX API and Web Scrapping.
- In the SpaceX API method, data is requested through the API and json formatted response is normalized in order to read into the '*data*' data frame. Then, there are several additional data that needs to be cleaned and reorganized. Also, BoosterVersion, LaunchSite, PayloadData and CoreData data are obtained through different sections of the SpaceX API.
- In Web Scrapping method, Wikipedia web page is used to obtain data. At first, a request has pulled and json formatted response is parsed by using BeautifulSoup object. Then, tables and their data contents are determined and found with BeautifulSoup library methods. At last step, a list that containes the data in the tables has prepared to get a data frame. Then, this list converted to the '*df*' data frame.

Data Collection – SpaceX API

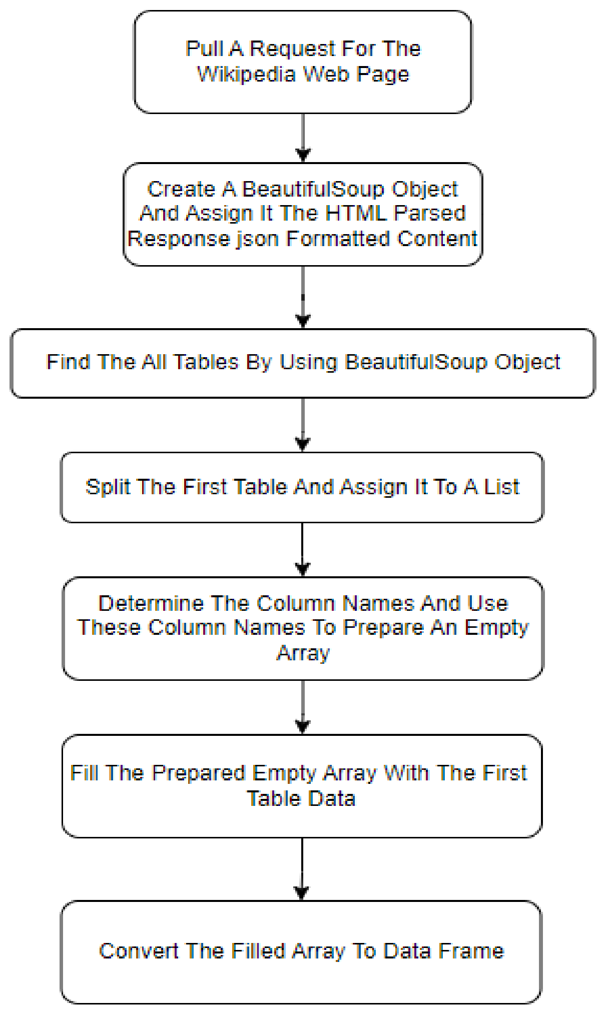
- Basic used codes are as follows,
- resp = requests.get(static_json_url)
- norm_resp = pd.json_normalize(resp.json())
- data = pd.DataFrame(norm_resp)

- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_1_Lab_1_Data_Collection_SpaceX_API_REST.ipynb



Data Collection - Scraping

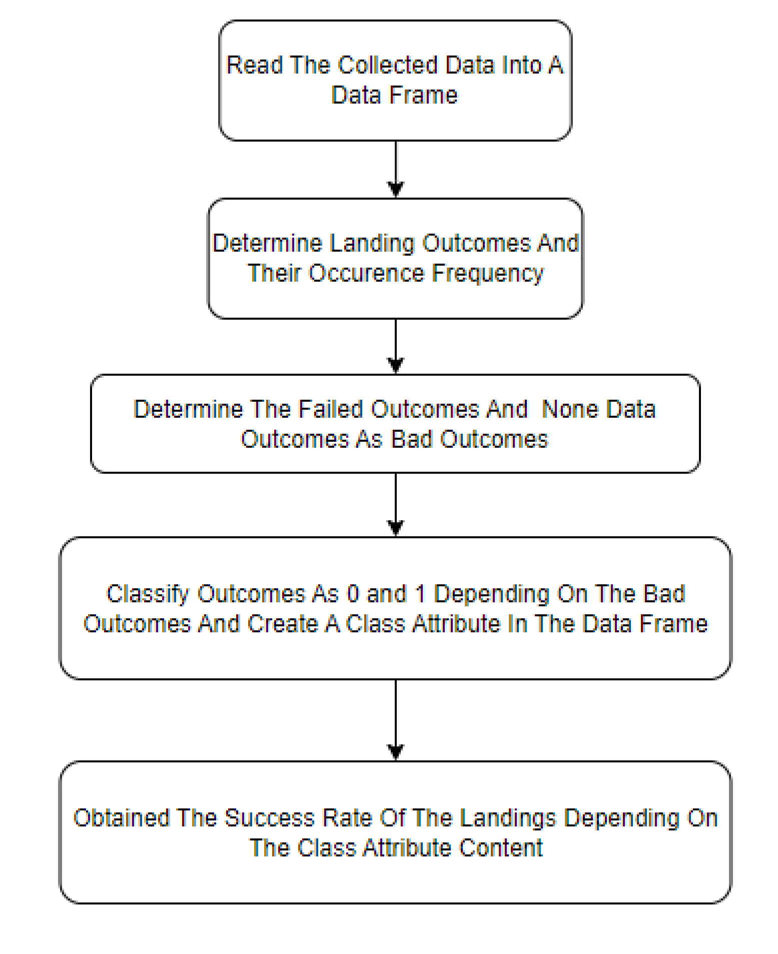
- Basic used codes are as follows,
  - response = requests.get(static_url)
  - soup = BeautifulSoup(response.content, 'html.parser')
- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_1_Lab_1_Data_Collection_Web_Scrapping.ipynb



Data Wrangling

- The collected data in the Data Collection stage is analyzed to get a class that shows the landing is failed or succeded. In order to get class attribute, bad landing outcomes are determined to get failed landing and none data outcomes. Then, these failed landing outcomes are marked as 0 in the ‘Class’ column and others are marked as 1.

- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_1_Lab_2_Data_Wrangling.ipynb



EDA with Data Visualization

- In this stage of the project, Scatter Plot, Line Plot, Categorical Plot(catplot()) and Bar Plot is used.
- Scatter Plot is used in order to see succeded/failed landing distribution of the data points of two attributes.
- Line Plot is used to identify succeded/failed landing trends depending on the years.
- Categorical Plot is used classify the succeded/failed landings depending on the Launch Sites .
- Bar Plot is used to visualize relationships between orbit categories and their yielding success rates of landings.

- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_2_Lab_1_EDA_Data_Visualization.ipynb

EDA with SQL

- select distinct("Launch_Site") from SPACEXTBL query is used to obtain unique values.
- select * from SPACEXTBL where "Launch_Site" like '%CCA%' limit 5 query is used to get 5 value of Launch Sites that are begin with the 'CCA'.
- select sum("PAYLOAD_MASS__KG_") from SPACEXTBL where "Customer" = 'NASA (CRS) ' query is used to obtain total payload mass carried by boosters launched by NASA (CRS).
- select avg("PAYLOAD_MASS__KG_") from SPACEXTBL where "Booster_Version" = 'F9 v1.1' query is used to obtain average payload mass carried by booster version F9 v1.1
- select "Date" from SPACEXTBL where "Landing_Outcome" = 'Success (ground pad) ' query is used to obtain date when the first successful landing outcome in ground pad was acheived.
- select "Booster_Version" from SPACEXTBL where "Landing_Outcome" = 'Success (drone ship)' and "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_" < 6000 query is used to obtain names of the boosters which have success in drone ship and have payload mass greater than 4000 kg but less than 6000 kg.
- select "Mission_Outcome", count("Mission_Outcome") from SPACEXTBL group by "Mission_Outcome" query is used to obtain total number of successful and failure mission outcomes.
- select distinct("Booster_Version") from SPACEXTBL where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPACEXTBL) query is used to obtain names of the *booster_versions* which have carried the maximum payload mass.

EDA with SQL

```
%sql select \
"Date", \
substr("January  February March    April    May      June     July     August   SeptemberOctober  November December ", substr("Date", 4, 2)*9+1, 9) as Month_Names, \
"Booster_Version", \
"Landing_Outcome", \
"Launch_Site" \
from SPACEXTBL \
where \
"Landing_Outcome" = (select "Landing_Outcome" from SPACEXTBL where "Landing_Outcome" like 'Failure (drone ship)') \
and \
"Launch_Site" = (select "Launch_Site" from SPACEXTBL where substr("Date",7,4) = '2015')
```

- Above query is used to 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.

```
%sql select "Landing_Outcome", count("Landing_Outcome") as rank from (select "Landing_Outcome" from SPACEXTBL where "Date">'04/06/2010' and "Date"<'20/03/2017') \
group by "Landing_Outcome" \
order by count("Landing_Outcome") desc
```

- Above query is used to rank the count of successful *landing_outcomes* between the date 04-06-2010 and 20-03-2017 in descending order.

- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_2_Lab_2_EDA_SQL.ipynb

Build an Interactive Map with Folium

- In this stage of the project, markers, circles, lines, marker_cluster and mouse_position are used to build an interactive map.
- Markers are added for showing the names of the launh sites or points.
- Circles are added to make the launch sites noticable on the map.
- Lines are added show distance line between two data points on the map.
- Marker Cluster is added to increase readibility of the data point that are on the same launch site(so in the same coordinates) in aspect of succes/fail. Success category is represented as green i label and fail category is represented as red i label.
- Mouse Position is added to get the coordinates on the map by using mouse position.

- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_3_Lab_1_Interactive_Visual_Analytics_with_Folium.ipynb

Build a Dashboard with Plotly Dash

- In the dashboard, Pie Chart, Scatter Chart, Dropdown Menu and Range Slider properties are added.
- Dropdown Menu is added to select launch sites that are shown in the Pie Plot.
- Pie Plot shows the successful and failed landings in the each launch sites.
- The Scatter Plot is used to show distribution of successful/failed landings depending on the PayloadMass and inputs that come from the Range Slider and Dropdown Menu are used to shape the Scatter Plot.

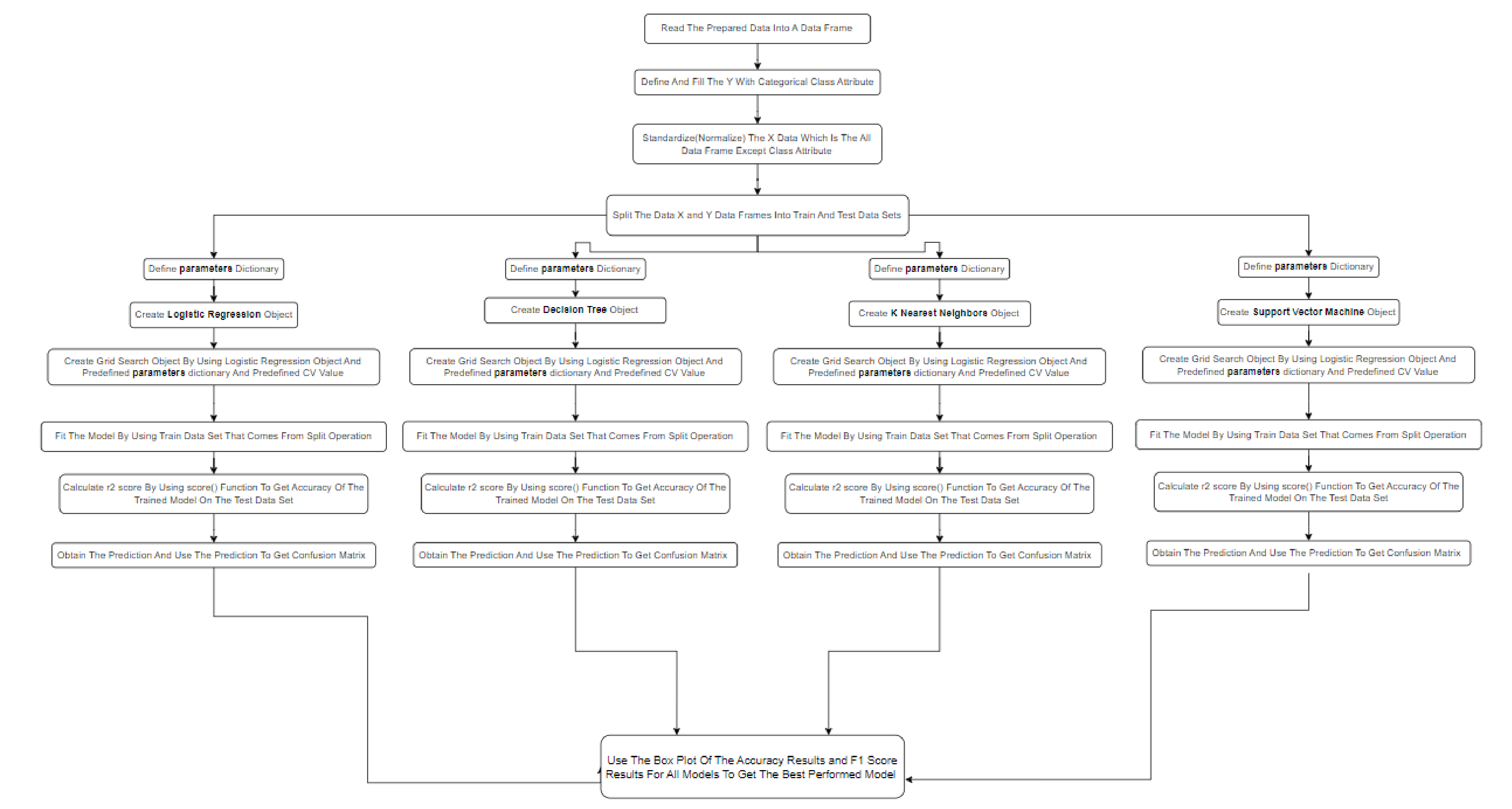
- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_3_Lab_2_Interactive_Dashboard_With_Plotly_Dash.py

Predictive Analysis (Classification)

- In this stage, data split into train and test data sets and they are used in the LogisticRegression, SupportVectorMachine, DecisionTree and KNearestNeighbors models with the *GridSearch* method. Roughly, all these models are applied to the data sets in same way.
- At first, model object is defined and these model object is pass as parameter to GridSearch method as well as the predefined parameters which are ‘dictionary’ and determined ‘cv’ value. Then, model is fitted(trained) with the train data sets.
- At the last step, predicted value of the test data set is obtained and these prediction are used for obtaining the r2 scores(accuracies), confusion matrix and F1 Scores. Accuracies and F1 Scores are used find best performing classification model.
- In aspect of accuracies and F1 Scores three of all four models have same values and DecisionTree Model has higher accuracy and F1 score than other models. Therefore, best performed model is Decision Tree Model.

- https://github.com/HasanUnlu09/IBM_Applied_Data_Science_Capstone/blob/main/Module_4_Lab_1_Predictive_Analysis_Classification.ipynb

Predictive Analysis (Classification)

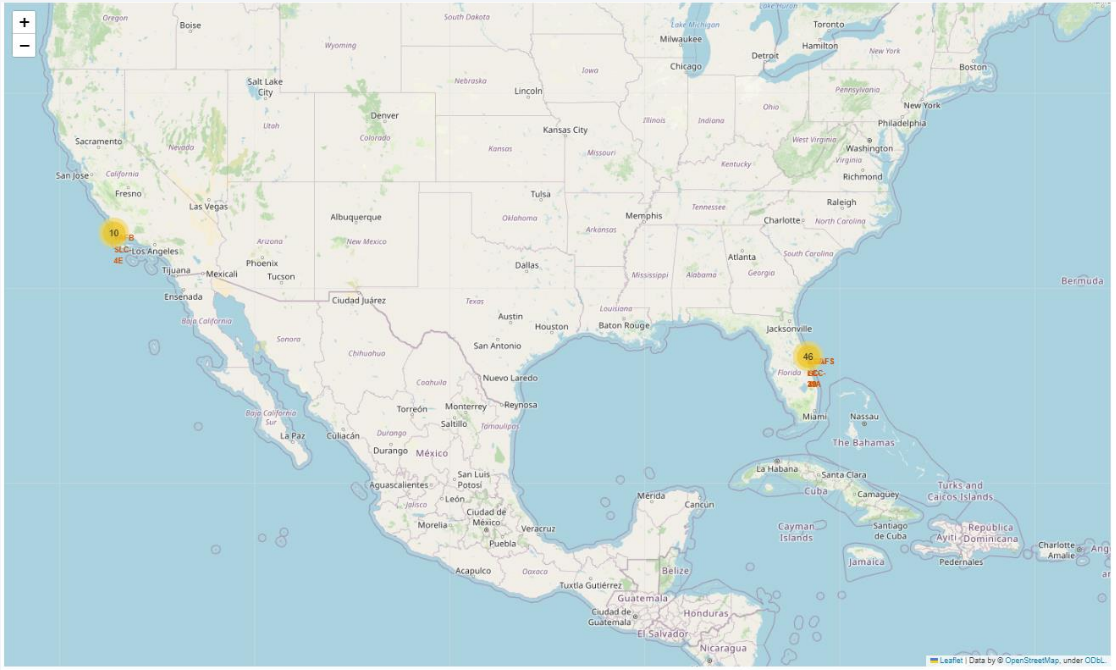


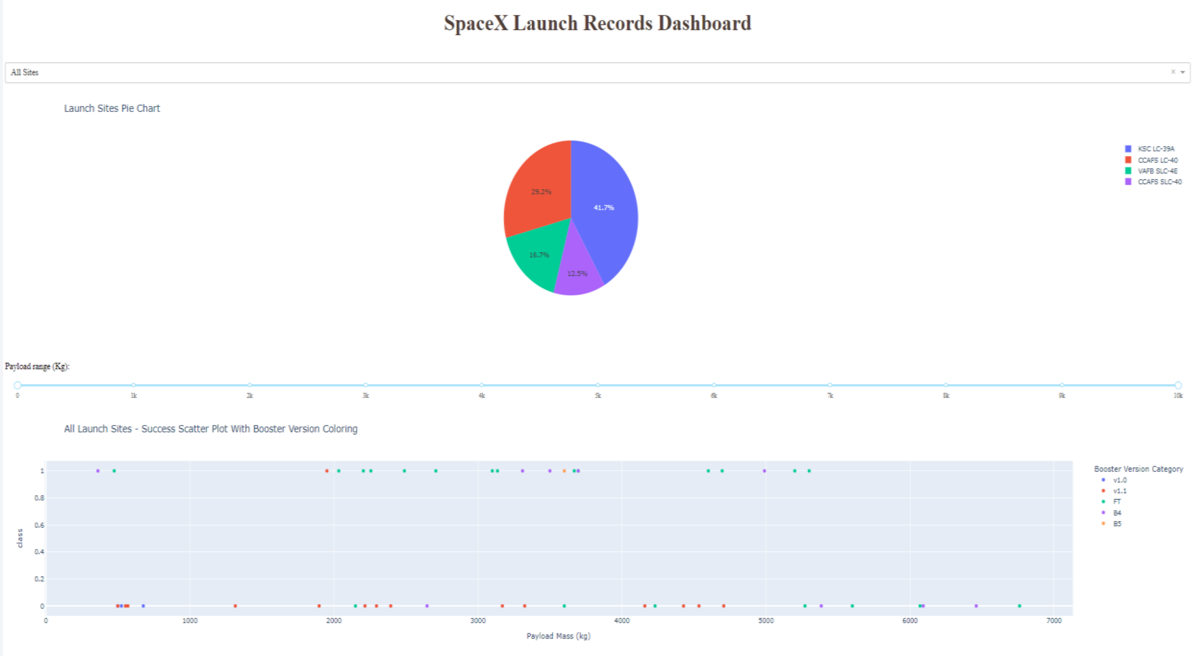
Results(EDA Analysis)

- In the Exploratory Data Analysis step, correlation between Flight Number - Launch Site, Payload - Launch Site, Success Rate - Orbit, Flight Number - Orbit, Payload Mass - Orbit, Launch Success - Year has been investigated.
- Firstly, succeded landings have been increasing since 2014.
- There is a relationship between Payload Mass and Orbit types because some payload amounts fails for some orbits and some payload amounts are more successful at some orbits. For example, around 1000 kg ISS, LEO and PO orbits have no successful launches.
- There is a relationship between Payload Mass and Launch Sites because some launch sites has more successful for some payload mass. For example, VAFB SLC 4E launch site has no failed launch in the range of around 2000 kg and 4000 kg payload mass.
- Also, there is four orbit types that have better success rates, ES-L1, GEO, HEO and SSO.
- Total payload mass that is carried by boosters from NASA is 45596 kg.
- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 can be listed as F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2.
- Also, some payload masses are more successful at some launch sites. Furthermore, there are 4 launch sites named as CCAFS LC-40, CCAFS SLC-40, VAFB SLC-4E and KSC LC-39A.
- Just to keep clear the slide, more detailed results can be found in the following slide pages.

Results(Interactive Analytics)

• A Demo For Interactive Folium Map





• A Demo For Interactive Plotly Dashboard

Results(Predictive Analysis)

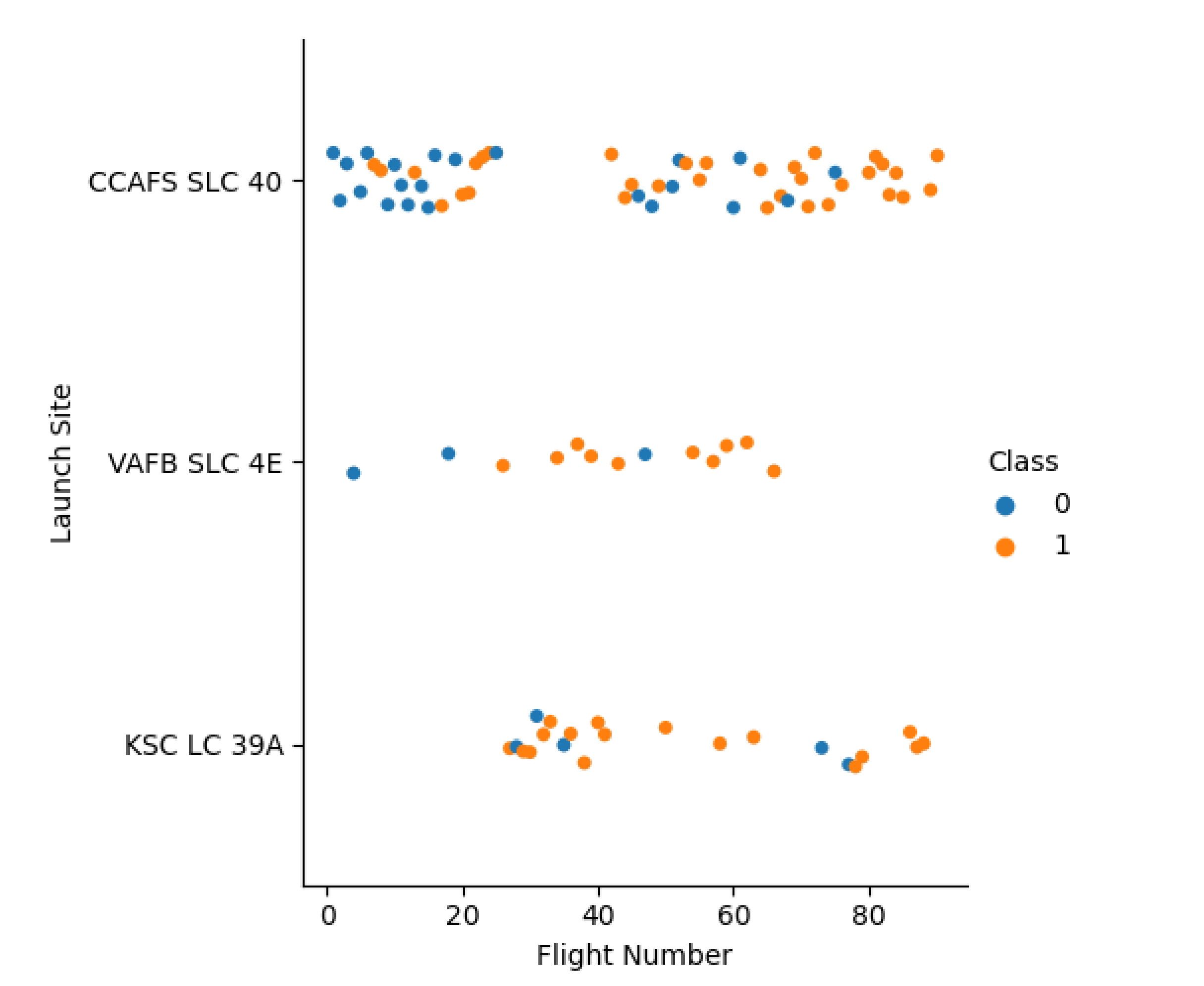
- There are four methods that are applied to the data set as LogisticRegression, SupportVectorMachine, DecisionTree and KNearestNeighbors by using GridSearch method.
- Results of these models are evaluated as accuracy and F1 Score and conclusion from these evaluation step is the best performed model is Decision Tree Model.

Section 2

Insights drawn from EDA

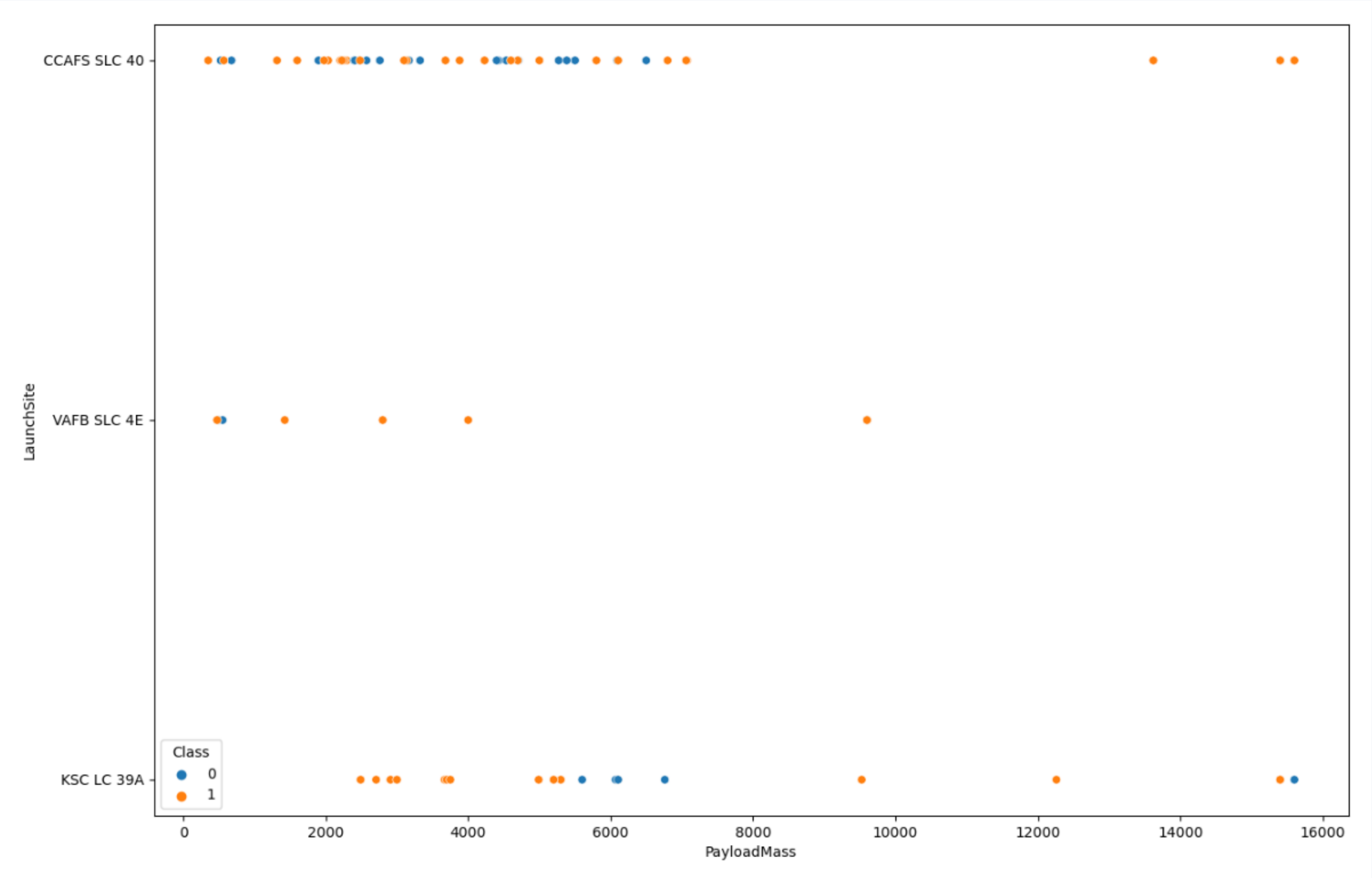
Flight Number vs. Launch Site

- According to categorical graph, success/fail landing distribution varies on three launch sites depending on the flight number.
- For the CCAFS SLC 40 launch site, it can be said that landings above 40 flight number are ended more successfully than below 40 flight number .
- For the VAFB SLC 4E launch site, despite there is no succeded landing below 20 flight number, almost perfect sequence of successful landings is obtained above 20 flight number.
- For the KSC LC 39A launch site, except between 40 and approximately 70 flight numbers, there is almost homogeneously distributed succeded/failed landing data points. Which leads to an ambiguity for future landing predictions.



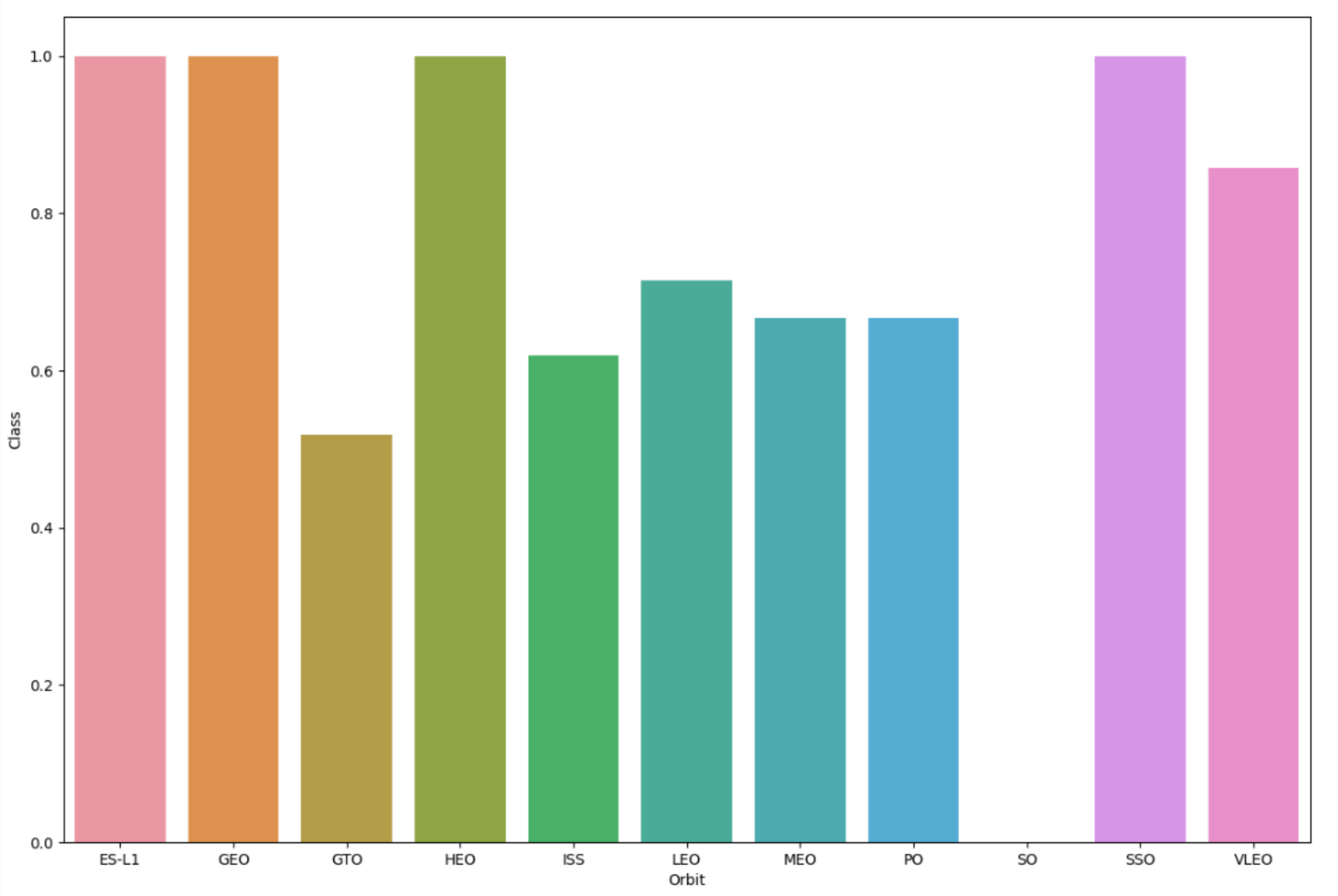
Payload vs. Launch Site

- According to graph, succeded/failed landings distribution varies on three launch sites depending on the payload mass.
- For the CCAFS SLC 40 launch site, it can be said that landings above 12000 kg are perfect. Although, there is a homogoneus succeded/failed landings distribution for other payload mass. Which means that this launch site is more appropriate for landing for high payload massses.
- For the VAFB SLC 4E launch site, it can be said that, almost all landings are ended with success. Therefore, probably, rockets that carry approximately above 1000 kg to 10000 kg payload mass will end up with succeded landings.
- For the KSC LC 39A launch site, it can be said that, landings except around 6000 kg and around 16000 kg will end up with success.



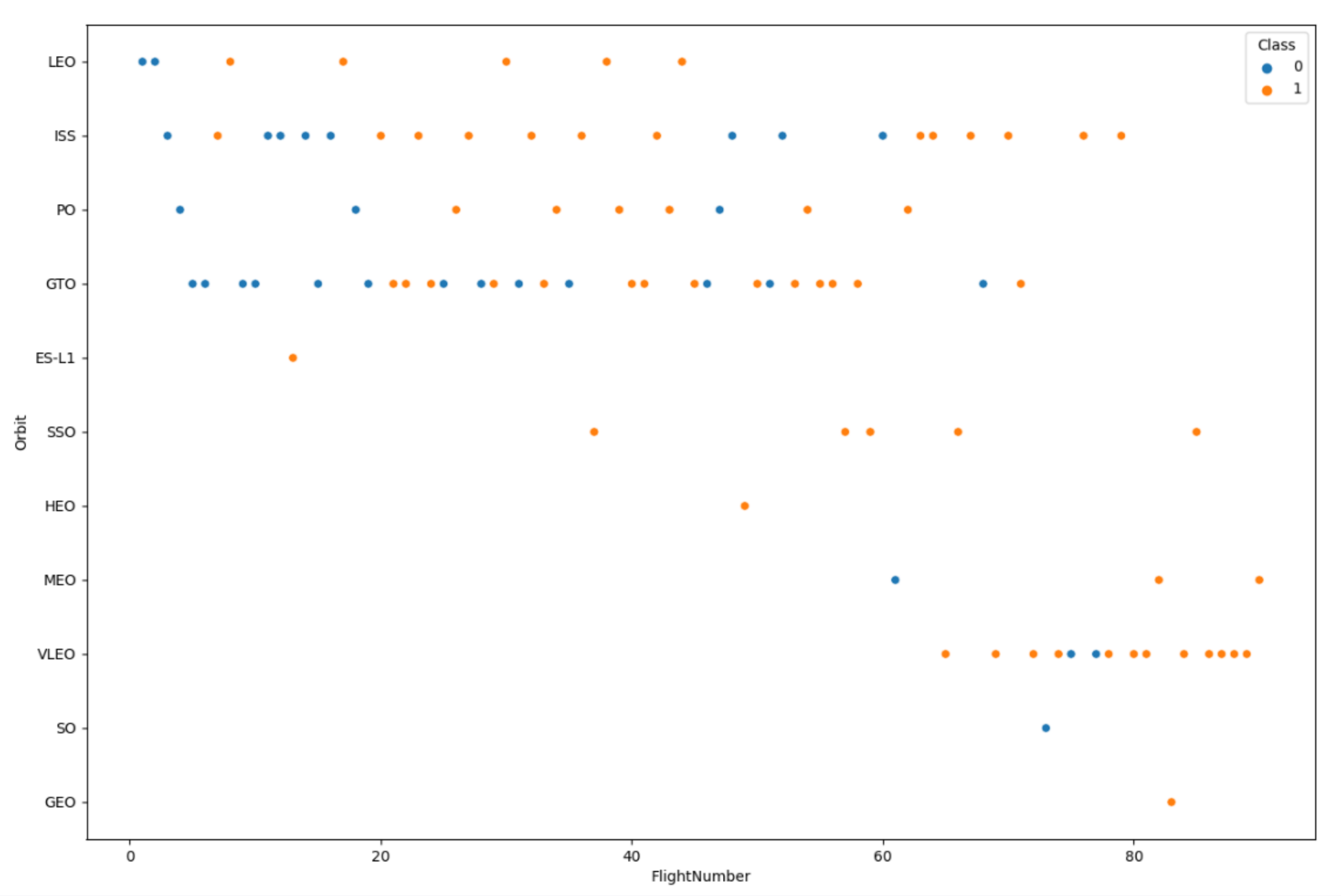
Success Rate vs. Orbit Type

- In the bar chart, successful rates for each orbit has shown.
- According to the graph, ES-L1, GEO, HEO and SSO orbits have perfect success rates despite GTO orbit which has approximately failed one of the two times. Also, there is no success data point for the SO orbit.
- As conclusion, this chart can give insights about the success possibility for a new launch for an orbit.



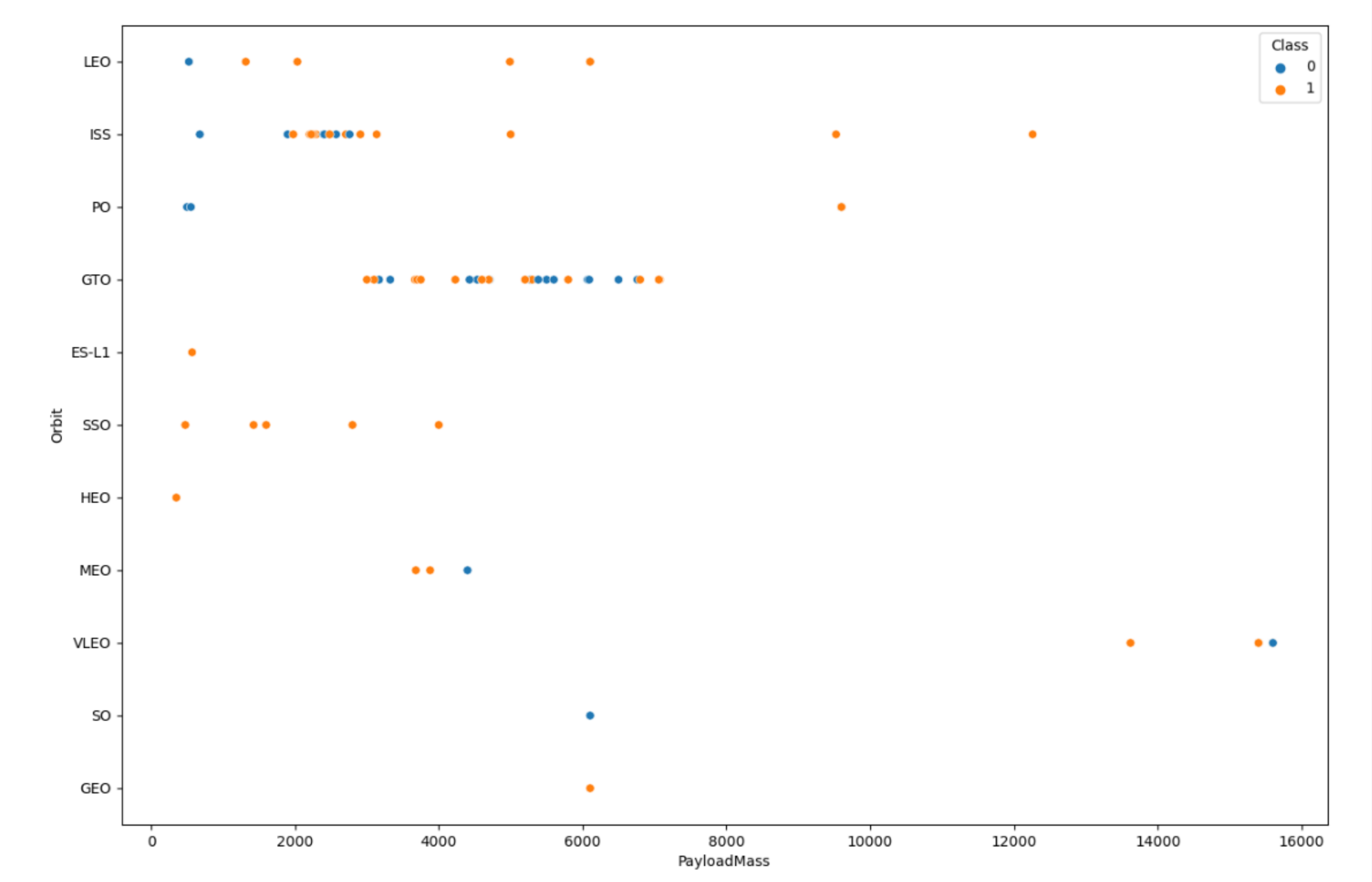
Flight Number vs. Orbit Type

- In this scatter plot, succesfull/failed landings per orbits are shown.
- ES-L1, SSO, HEO, GEO and approximately MEO orbits have perfect landings results for each flight number. Which can be said that probably landings for these orbits will end up with success.
- LEO and VLEO orbits have also high success rate which means new landings for these orbits will possibly be the successful landings.
- ISS, PO and GTO orbits have varying distribution of success/fail launch. So, a new landings can not be interpreted as it will be successful or failed.



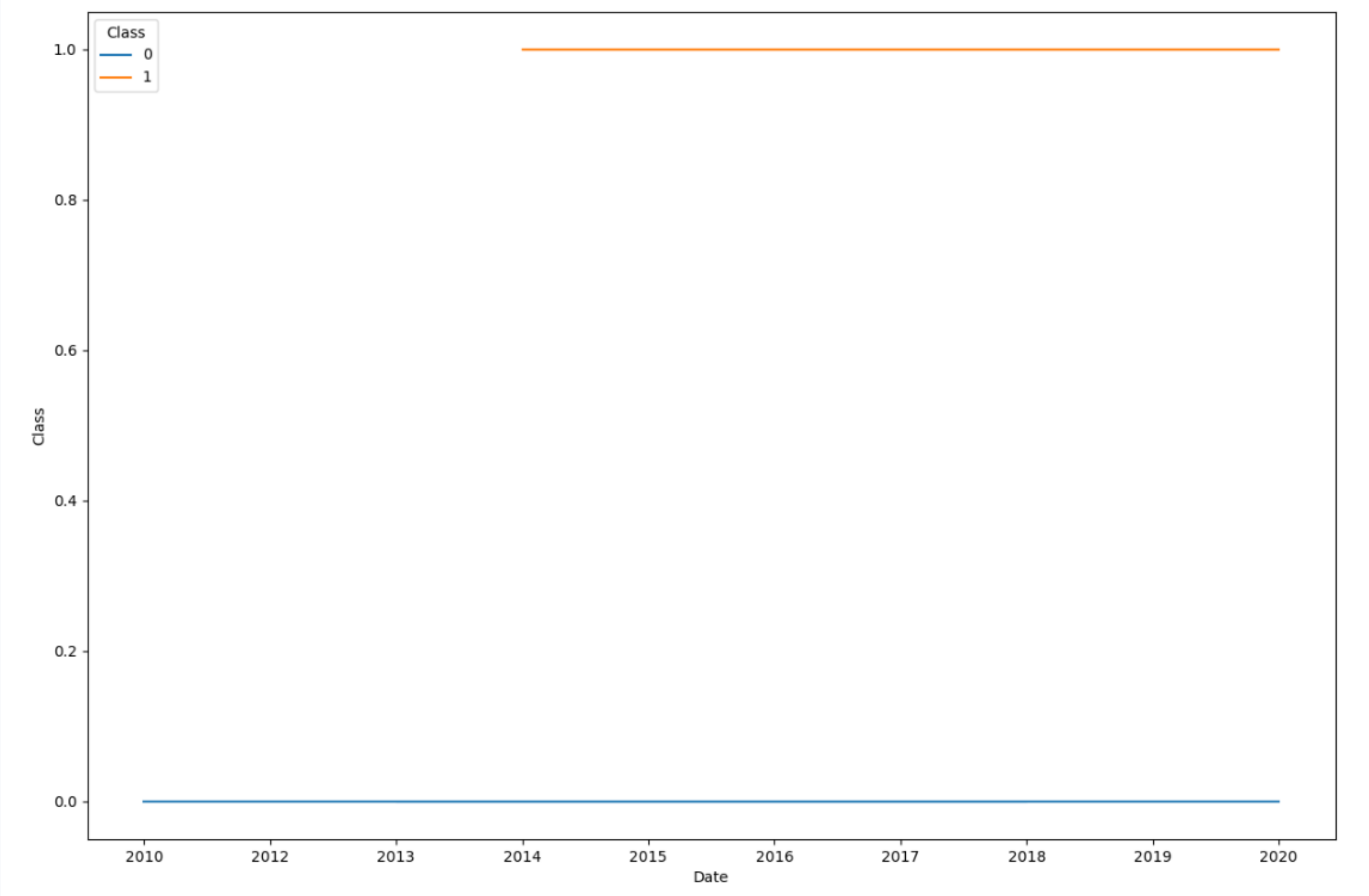
Payload vs. Orbit Type

- In this scatter plot, succesfull/failed landings that depends on the payload mass are shown per orbits.
- It can be said that, payload masses up to 4000 kg is more suitable for ES-L1, SSO, HEO and MEO orbits.
- Above 4000 kg payload masses, LEO, ISS, PO, VLEO and GEO orbits are more suitable.
- In the GTO orbit it cannot be said that there is a payload mass range that is more suitable. There is approximately homogeneus success/fail landing distribution for ranging payload masses.
- Also, ISS orbit is not suitable for approximately 2000kg - 4000kg range.



Launch Success Yearly Trend

- In this line plot, successes and fails of landings are shown depending on the years.
- From the chart, it can be said that successful landings are in trend from approximately 2014.



All Launch Site Names

- According the query result, there are 4 different launch sites in the dataset and also there are none launch sites.

```
%sql select distinct("Launch_Site") from SPACEXTBL
```

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

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL where "Launch_Site" like '%CCA%'limit 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_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

- First five launch site results that begins with the ‘CCA’ is ‘CCAFS LC-40’ as it is indicated with the red rectangle in the above image.

Total Payload Mass

- Total payload mass that is carried by boosters from NASA is 45596 kg as it is shown in the below image.

```
%sql select sum("PAYLOAD_MASS__KG_") from SPACEXTBL where "Customer" = 'NASA (CRS)'
```

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

sum("PAYLOAD_MASS__KG_")
45596.0

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928.4 kg as it is shown in the below image.

```
%sql select avg("PAYLOAD_MASS__KG_") from SPACEXTBL where "Booster_Version" = 'F9 v1.1'
```

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

avg("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad is shown in the right side image.
- It was started from 2015 and it continued upto 2018.

```
%sql select "Date" from SPACEXTBL where "Landing_Outcome" = 'Success (ground pad)'
```

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

Date
22/12/2015
18/07/2016
19/02/2017
05/01/2017
06/03/2017
14/08/2017
09/07/2017
15/12/2017
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 are shown in the below image.

```
%sql select "Booster_Version" from SPACEXTBL where "Landing_Outcome" = 'Success (drone ship)' and "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_" < 6000
```

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

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

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes are shown in the below image.
- There are only 1 failure out of 101 attempt.

```
%sql select "Mission_Outcome", count("Mission_Outcome") from SPACEXTBL group by "Mission_Outcome"
```

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

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

Boosters Carried Maximum Payload

- The names of the booster which have carried the maximum payload mass is shown in the right side image.
- There are 12 different booster version which have carried maximum payload mass.

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

```
%sql select distinct("Booster_Version") from SPACEXTBL where "PAYLOAD_MASS__KG_" = (select max("PAYLOAD_MASS__KG_") from SPACEXTBL)
```

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

2015 Launch Records

- The failed Landing Outcomes in drone ship, their booster versions, and launch site names for in year 2015 is listed as shown in the below image.

```
%sql select \
"Date", \
substr("January  February March    April     May       June      July      August    SeptemberOctober  November December ", substr("Date", 4, 2)*9+1, 9) as Month_Names, \
"Booster_Version", \
"Landing_Outcome", \
"Launch_Site" \
from SPACEXTBL \
where \
"Landing_Outcome" = (select "Landing_Outcome" from SPACEXTBL where "Landing_Outcome" like 'Failure (drone ship)') \
and \
"Launch_Site" = (select "Launch_Site" from SPACEXTBL where substr("Date",7,4) = '2015')
```

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

Date	Month_Names	Booster_Version	Landing_Outcome	Launch_Site
01/10/2015	November	F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40
14/04/2015	May	F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40
03/04/2016	May	F9 FT B1020	Failure (drone ship)	CCAFS LC-40
15/06/2016	July	F9 FT B1024	Failure (drone ship)	CCAFS LC-40

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

- The ranked list 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 are as shown in the following image.

```
%sql select "Landing_Outcome", count("Landing_Outcome") as rank from (select "Landing_Outcome" from SPACEXTBL where "Date">'04/06/2010' and "Date"<'20/03/2017') \
group by "Landing_Outcome" \
order by count("Landing_Outcome") desc
```

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

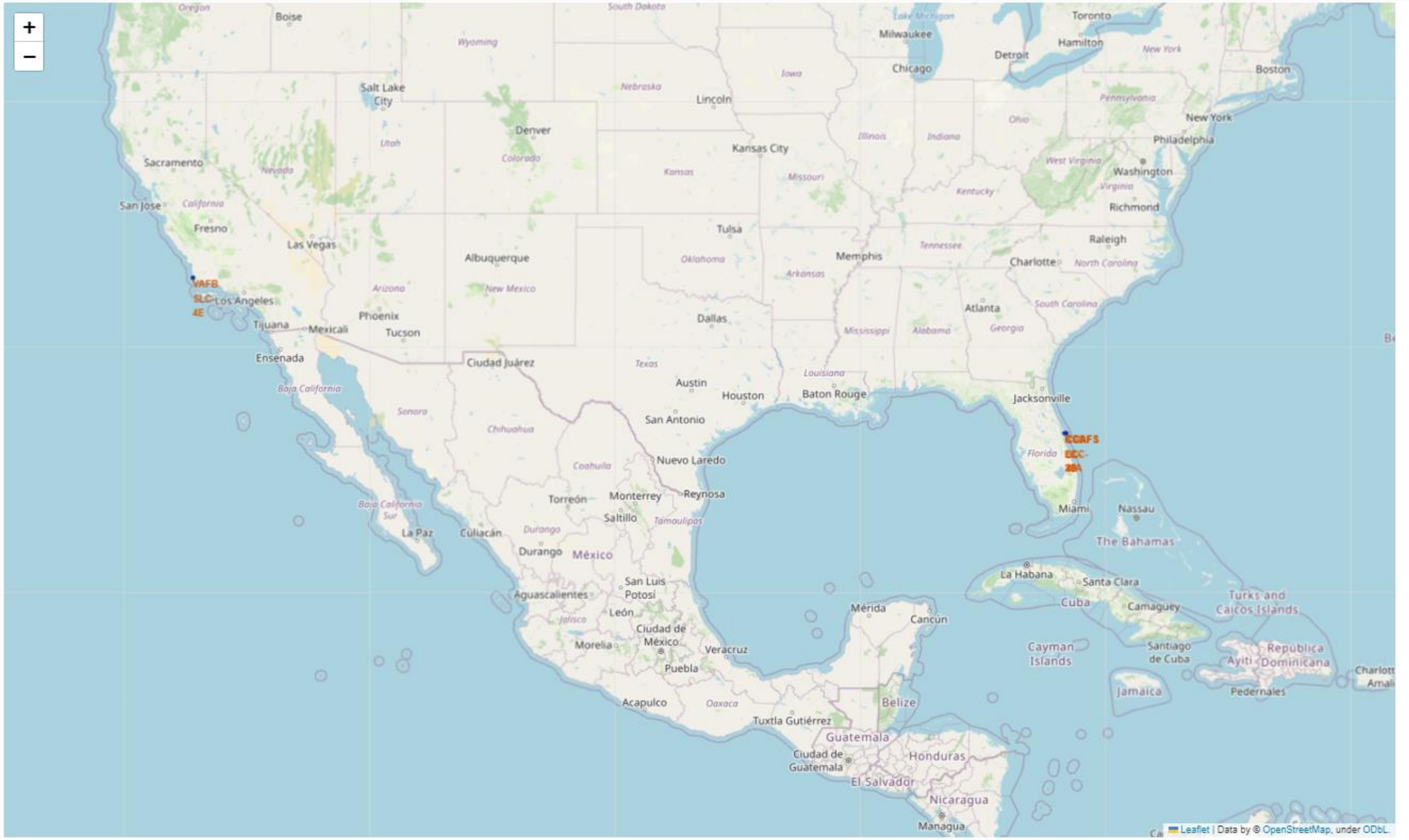
Landing_Outcome	rank
Success	20
No attempt	9
Success (drone ship)	8
Success (ground pad)	7
Failure (drone ship)	3
Failure	3
Failure (parachute)	2
Controlled (ocean)	2
No attempt	1

Section 3

Launch Sites Proximities Analysis

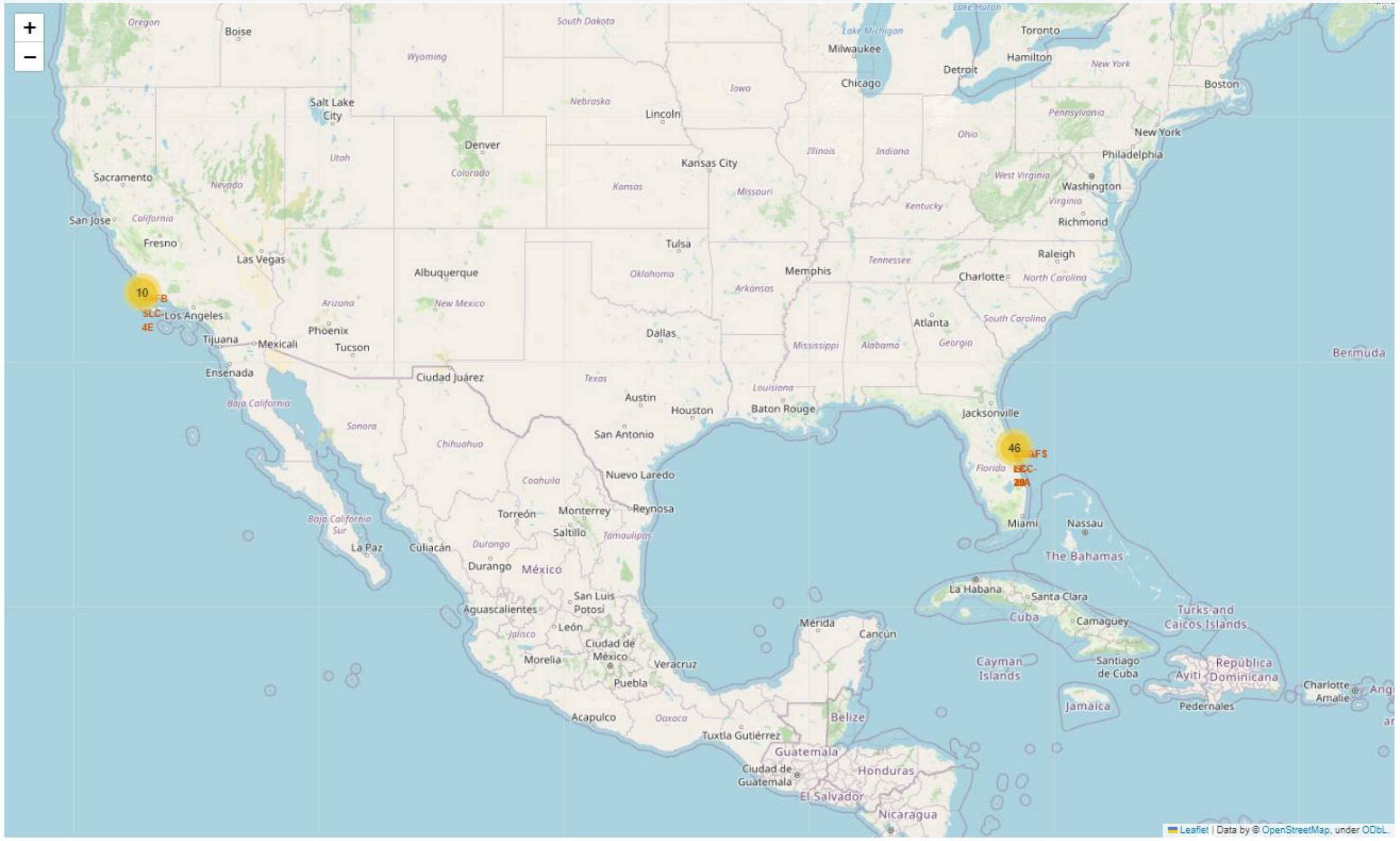
Launch Sites On Map

- According to launch sites on the map, there should be a necessity to be close the coast. Since, all launch sites are nearby the sea.



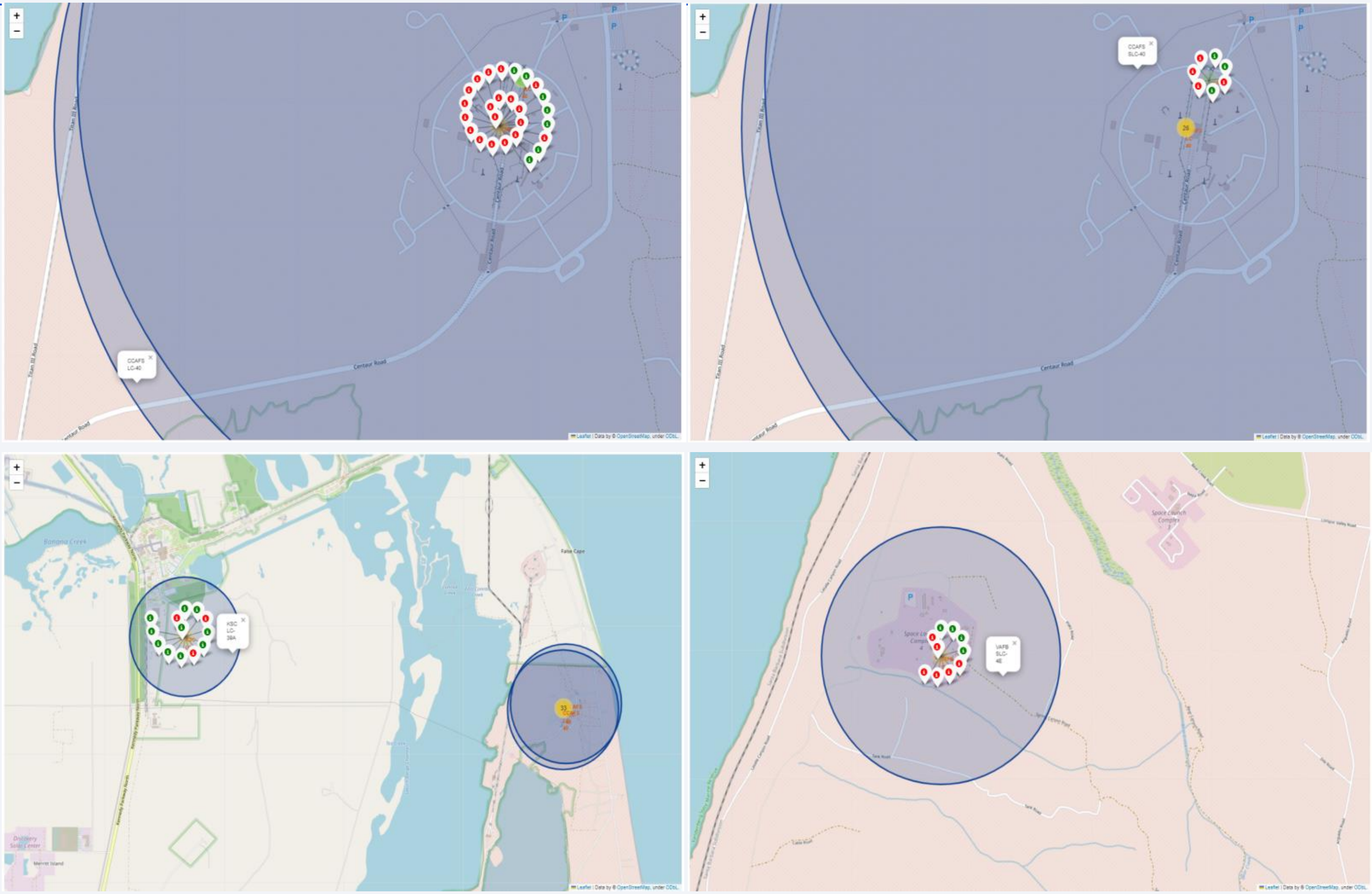
Success/Failed Launches For Each Site On The Map

- Map is shown as before zooming in the yellow dot areas.
- Image for the success/fail map is in the next slide.



Success/Failed Launches For Each Site On The Map

- There are 4 launch sites as CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E.
- Launch site at the top left image is CCAFS LC-40 and it has 26 launches. Only 7 of 26 landings are successful in this site.
- Launch site at the top right image is CCAFS SLC-40 and it has 7 landings and 3 of 7 landings are successful in this site.
- Launch site at the bottom left image is KSC LC-39A and it has 13 landings and 10 of 13 landings are successful in this site.
- Launch site at the bottom right image is VAFB SLC-4E and it has 10 landings and 4 of 10 landings are successful in this site.

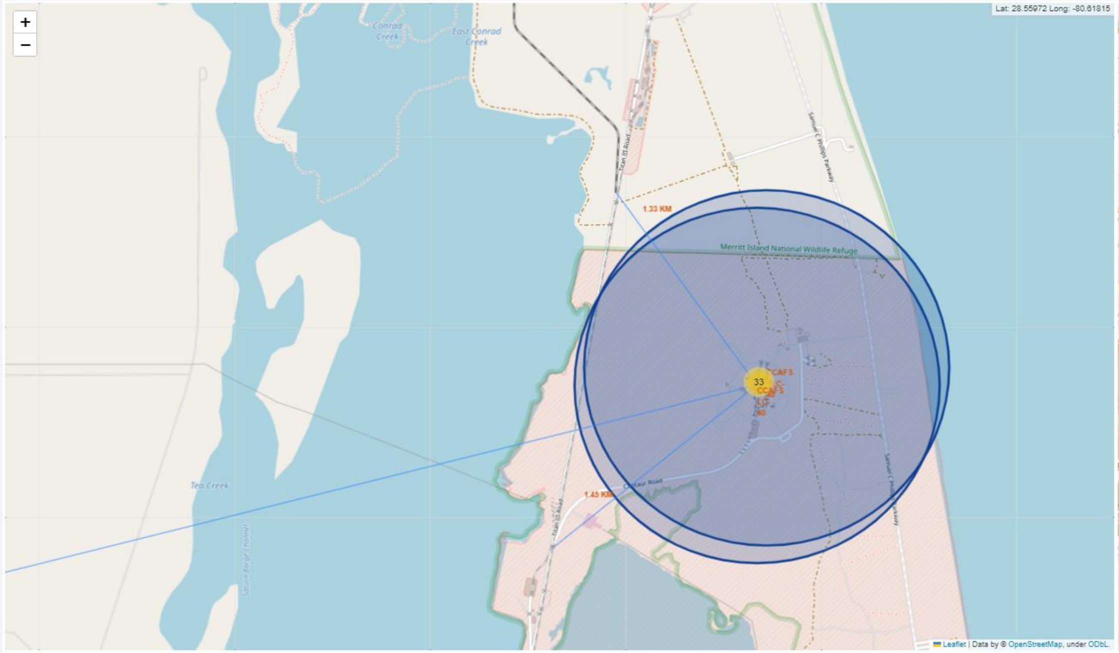


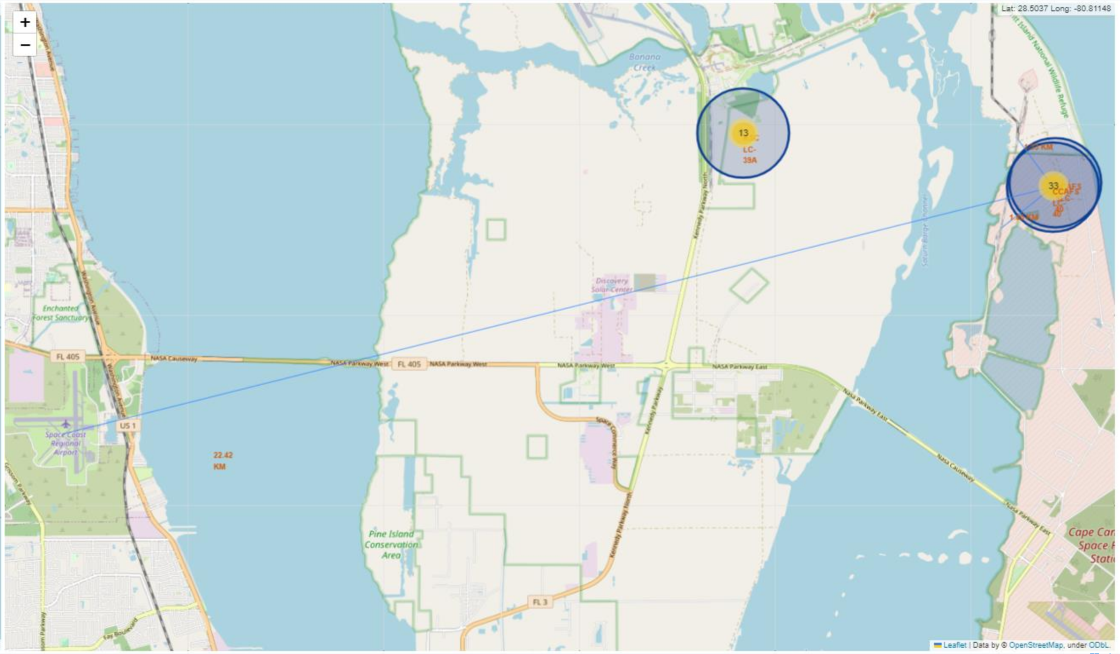
- According to above findings, it can be said that landings in the CCAFS SLC-40 site are more successful than CCAFS LC-40 launch site despite they are too close to each other.
- KSC LC-39A launch site is the best performed launch site despite closeness of this site to CCAFS LC-40 and CCAFS SLC-40 sites.
- VAFB SLC-4E launch site is in the other side of the country, yet it is too close to the coast as all others.

Distances Between A Launch Site To Its Proximities On Map

- As it can be seen from the left side image, the closest railway distance is 1.33 km and closest highway distance is 1.43 km. Also, the closest airport is 22.42 km away as it can be seen from the right side of the image.

- From this aspect, transportation way locations are selected to be relatively close and nearby assembly area locations such as airports, cities, etc. are selected as far as it can be. Which makes sense in order to protect living beings and structures in case of a disaster such as booster explosion.

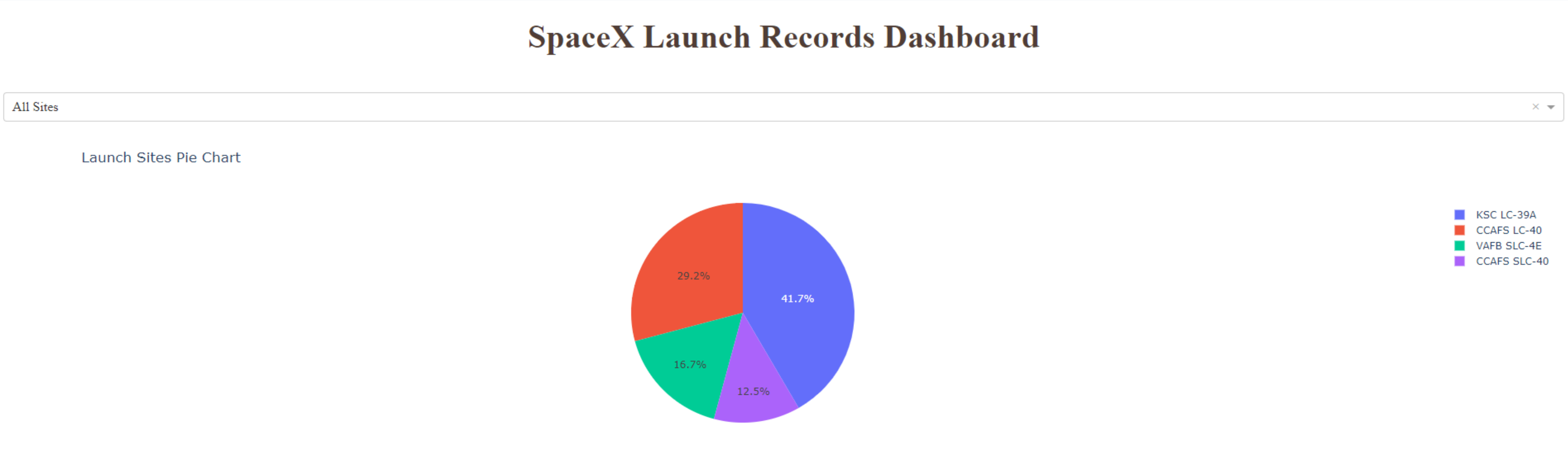




Section 4

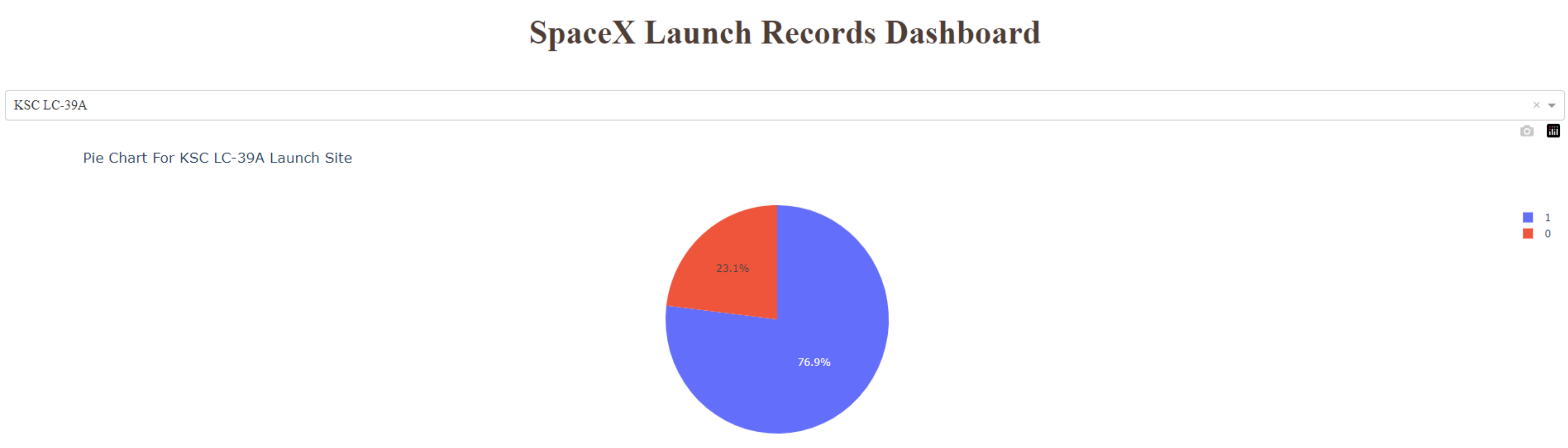
Build a Dashboard with Plotly Dash

Launch Success Count For All Sites



- According to the pie chart, if a new launch needs to be made, then KSC LC-39A is the most success promising option among other sites.
- !!! Just to remind, launches' success depends on several varibles such as orbits, payload mass etc. Therefore, jumping to a conclusion that KSC LC-39A launch site provide a successful landing will not be a correct approach.

Launch Site With Highest Launch Success Ratio



- In the pie chart, launc site(KSC LC-39A) that has the highest landings success ratio is shown.
- Blue colored area represents the successful landings on the KSC LC-39A site and red colored area represents the failed landings.

Payload vs. Launch Outcome For All Sites With Different Payload

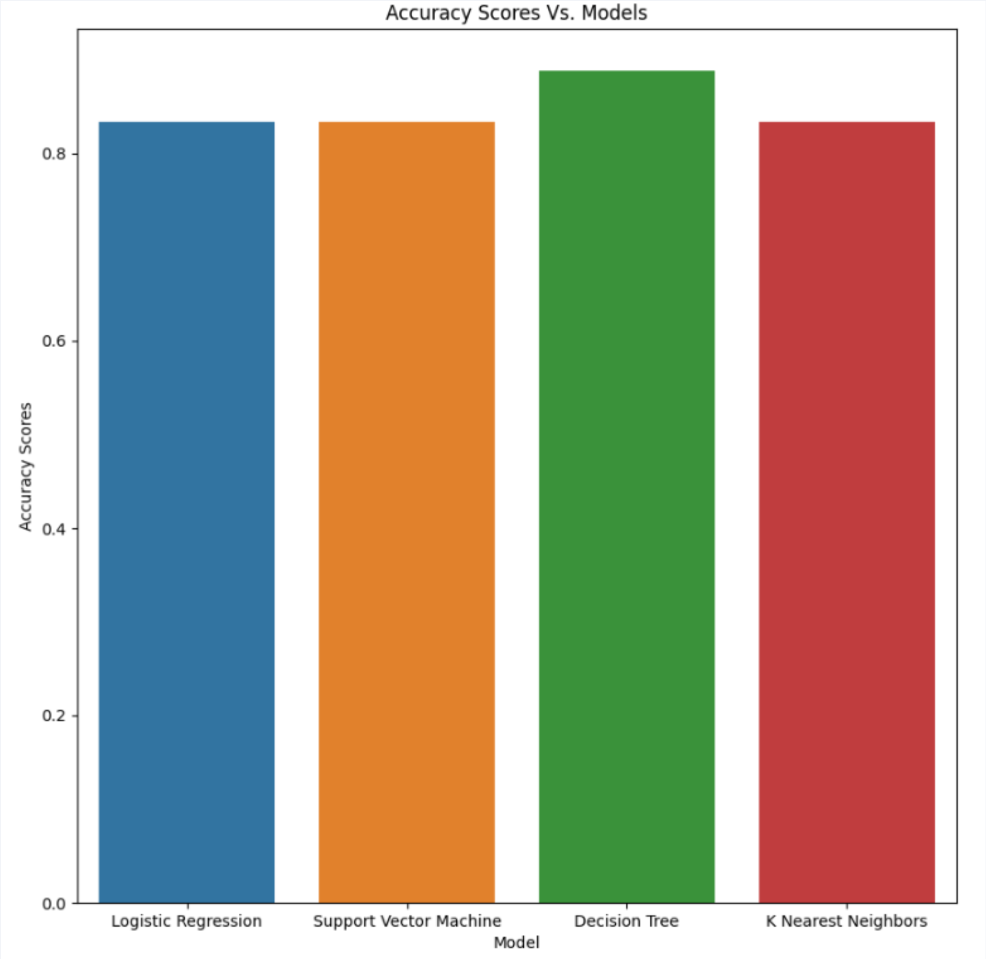
- Dashboard in the above image is about All Launch Sites and scattered succeded/failed landing distribution of these sites depending on payload mass by coloring booster versions.
- In the dashboard payload range is limited between 1000 kg and 7000 kg for all launch sites. Therefore, it can be said that the best performed booster verison is the B5 version. Since, it has only one landing in this range and it is succesful. However, other booster version that have more samples should give more insight. Therefore, actual good performance between all booster versions in this range is FT version. Since, 12 launces out of 19 landings ended up with success which makes the perfomance percenteage is 63.16%.

Section 5

Predictive Analysis (Classification)

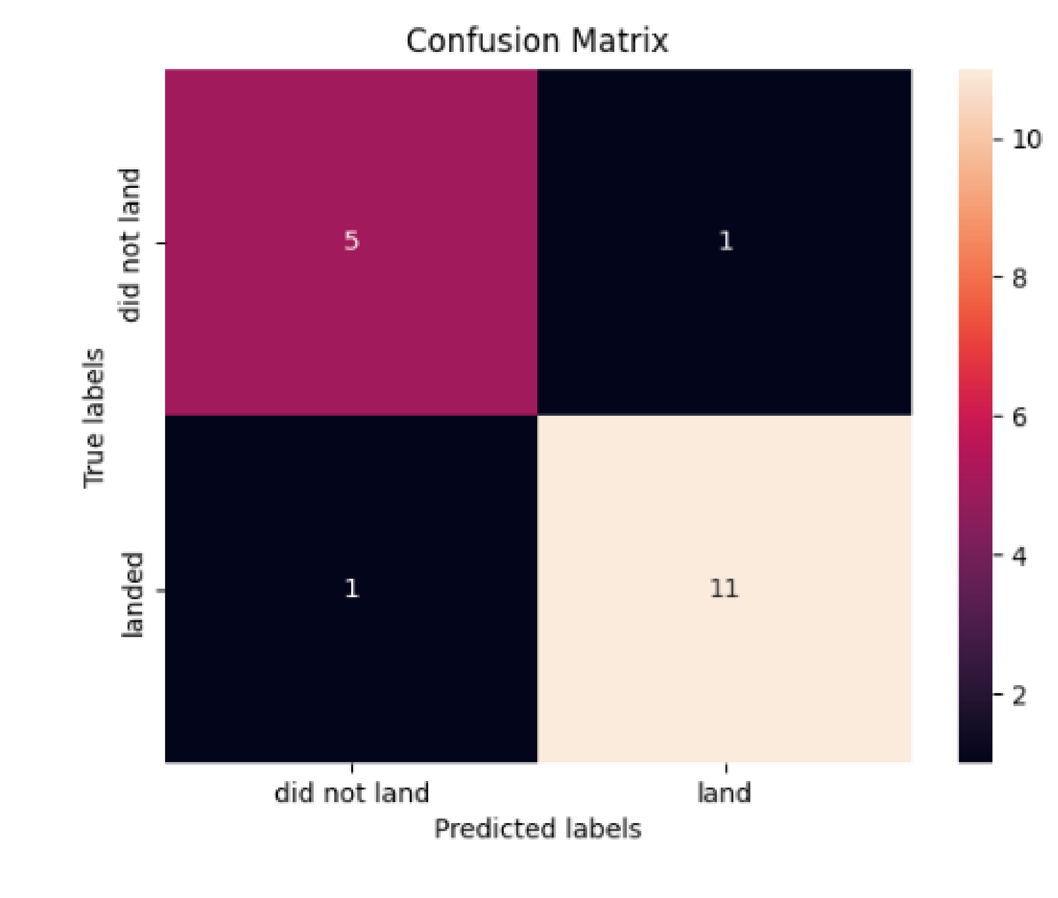
Classification Accuracy

- Right side bar chart is about accuracy scores of the four tried models.
- In the chart, models are in the x axis and the yielding accuracy results in the y axis.
- As it can be understood from the bar plot, the best performed classification model is Decision Tree.



Confusion Matrix

- Confusion Matrix of the Decision Tree Model, which performs best, is shown.
- In the matrix, there are 6 'did not land' labeled data point that is shown in the first row of the matrix. Only one of these 6 data points couldn't be predicted correctly by the Decision Tree Model.
- There is also 'landed' labeled data points which represents the successful landings in the data set and it is shown in the second row of the confusion matrix. Again only one of 12 data points couldn't be predicted by the Decision Tree Model.
- From the results it can be said that, only 2 of data points couldn't be predicted correctly by the Decision Tree Model out of 18 data points. It results as 88.89% success percentage.



Conclusions

- There is a strong relationship between of payload mass and launch sites in aspect of success/fail landings.
- Also, there is a strong relationship between of payload mass and orbit types in aspect of success/fail landings.
- Launch sites are consciously build on the locations that are near to the coasts, railways, and highways and far to the publicly dense areas.
- Booster Versions can effect the success/fail rate of the landing.
- The best performed Machine Learning Classification algorithm is the Decision Tree Classification Model.

Thank you!