IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection and wrangling
  - EDA with data visualization and SQL
  - Interactive map using Folium
  - Dashboard with Plotly Dash
  - Machine learning modeling
- Summary of all results
  - Insight drawn from EDA
  - Dashboard and map and graphs for visualization
  - Predictive analysis(classification)

Introduction

Project background and context

- Space X is a company built with the goal of space travel.The success of SpaceX's launch and landing is an important issue that determines the success of the company.
- With changes in technology and software, the success rate of Falcon 9 so far is examined through various data.

Problems that need to be answered

- Find out which variables influenced for Falcon 9 landing success
- Predict successful rate analysis with classification modeling
- Weather the success rate has increased according to various changes over the year change

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology**
  - *Data was obtained from 2 sources*

    Space X API

    Web scraping : Falcon 9 launch dat from wikipedia

- **Perform data wrangling**
  - *Data were transformed and create a landing outcome label.*

- **Perform exploratory data analysis (EDA) using visualization and SQL**

- **Perform interactive visual analytics using Folium and Plotly Dash**

- **Perform predictive analysis using classification models**
  - *Collected data were divided into train and and test data and predict binary result using 4 classification Machine learning model.(Logistic Regression, SVC, DecisionTree and KNeighbors)*

Data Collection

- Data set is collected these 2 data sources

Space X API (“https://api.spacexdata.com/v4/rockets/“)

Request and parse the SpaceX launch data → Filter the dataframe to only include `Falcon 9` launches → Dealing with Missing Values → Export the file to CSV ‘dataset_part_1.csv’

Web scraping Falcon 9 launch data from wikipedia

Request the Falcon9 Launch Wiki page from its URL → Extract all column/variable names With beautiful soup → Create a data frame by parsing the launch HTML tables → Export the file to CSV ‘spacex_web_scraped.csv'

Data Collection – SpaceX API

1.Request the SpaceX launch data using the GET request

2. Turn the responses file into a data frame using `.json_normalize()`

3. Create a Pandas data frame from the dictionary launch_dict with new column names

4. Filter the data frame to only include `Falcon 9` launches

5. Convert to CSV file “data_falcon9.to_csv(‘dataset_part_1.csv', index=False)”

https://github.com/delicate99/SpaceX-capston/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

1. Request the Falcon9 Launch Wikipedia web page from its URL

2. Extract all column/variable names from the HTML table header with Beautifulsoup object

3. Creating launch_dict and fill up it with launch records extracted from table rows.

4. Converting the dictionary to DataFrame

5. Make a cdv file “spacex_web_scraped.csv”

https://github.com/delicate99/SpaceX-capston/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

1. Data loading and import libraries

2. Identify and calculate the percentage of the missing values

3. Calculate the number of launches on each site

4.Calculate the number and occurrence of each ‘Orbit’ column

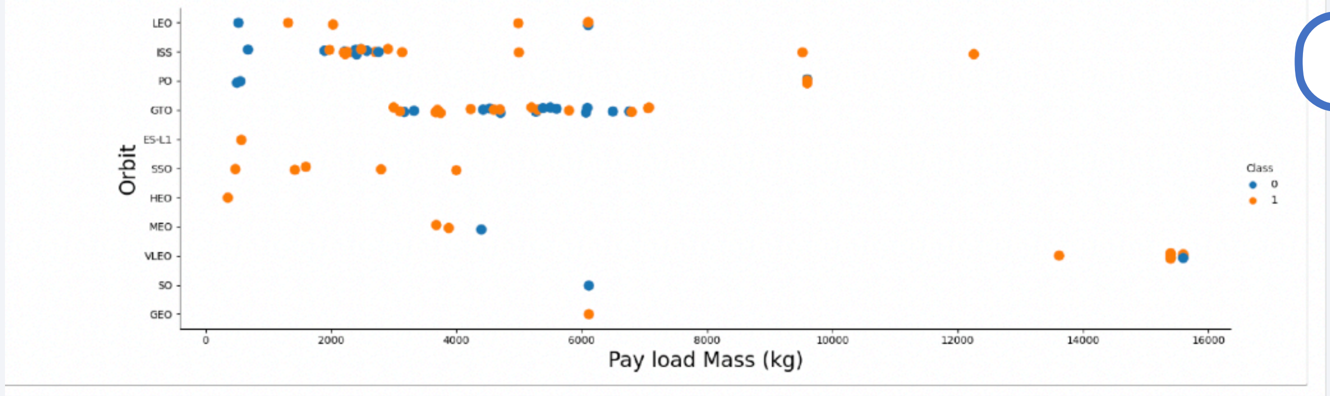
5.Calculate the number and occurrence of mission outcome per orbit type

6. Create a landing outcome label from Outcome column (Success :1, Failure :0)

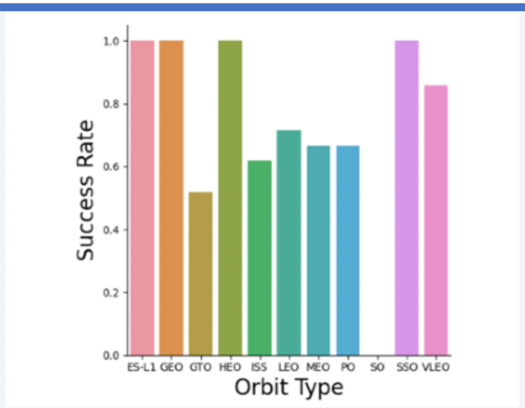
https://github.com/delicate99/SpaceX-capston/blob/main/jupyter-labs-spacex-data_wrangling_jupyterlite.ipynb

EDA with Data Visualization

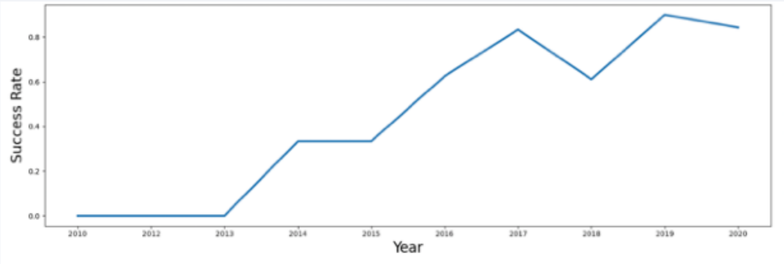
1. Scatter plot (for analysis for launching outcome with the variables.
'Flight Number', 'Launch site' ,'PayloadMass', 'Orbit type' relationship for outcome)



2. Bar plot for success outcome depends on orbit type



3. Line plot for launch success trend by year



Use "seaborn" Library

https://github.com/delicate99/SpaceX-capston/blob/main/jupyter-labs-EDA-Visualization.ipynb

EDA with SQL

SQL queries.

- 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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

https://github.com/delicate99/SpaceX-capston/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Map objects such as markers, circles, lines, etc. I created and added to a folium map
  - Colored mark for the launch –site for discriminate for success and failure
  - Cluster object used for the clusters launch sire like NASA. When you zoom in you can see the individual launch site.
  - folium.Circle and folium.Marker for each launch site on the site map.
  - Lines were used for the distance of two coordinates (Latitude and Longitude)

https://github.com/delicate99/SpaceX-capston/blob/main/jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

1. Launch Site Dropdown List
   - all and 4 different launch site for the choice

2. Pie chart showing success rate for each site
   - all site shows the relative success rate compare 4 launch site
   - Individual site shows the its own success and failure rate

3. Range slider was used for the selection of payload mass range.
   - ranges are 0 -10000kg
   - In the selected ranges you can see the success and failure distribution in all or each launch site

4. Scatter chart for the success and failure distribution
   - shows the outcome according to the booster version categories in the selected payload mass ranges .

https://github.com/delicate99/SpaceX-capston/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

4 classification models are used for prediction
(Logistic Regression, Support Vector Machine, Decision tree classifier, K nearest Neighbor classifier)

1. Data collection and normalization.

2. Data were divided for train and test set

3. Classification models established.
Use GridsearchCV for testing various parameters.
And find best parameter for each models.

4. Calculate accuracy use .score()
Draw confusion matrix.

https://github.com/delicate99/SpaceX-capston/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- **Exploratory data analysis results**
  - ***Graphs and SQL for the analysis the data***

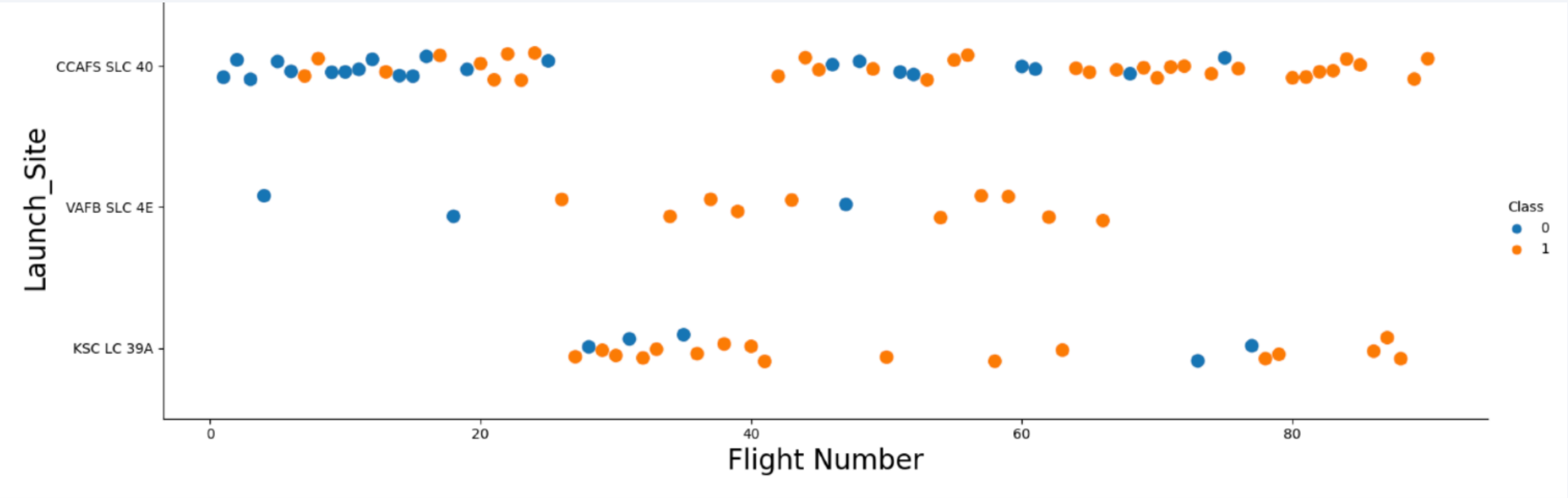
- **Interactive analytics demo in screenshots**
  - **Dashboard graphs for easy intuition**

- **Predictive analysis results**
  - **Predictive accuracy confusion matrix**

Section 2

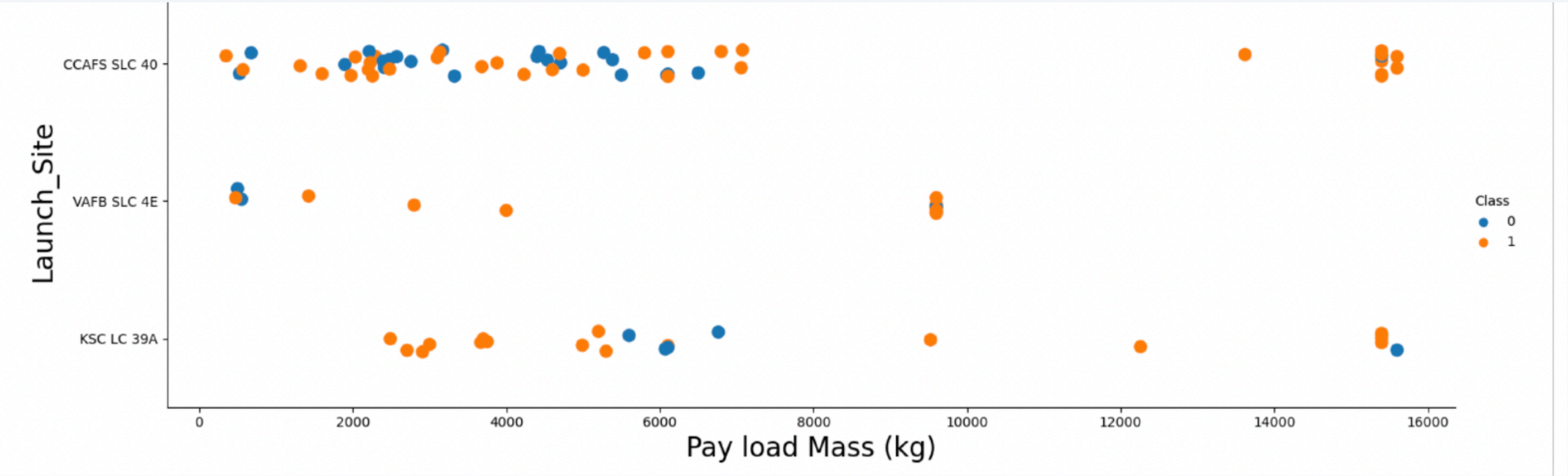
Insights drawn from EDA

Flight Number vs. Launch Site



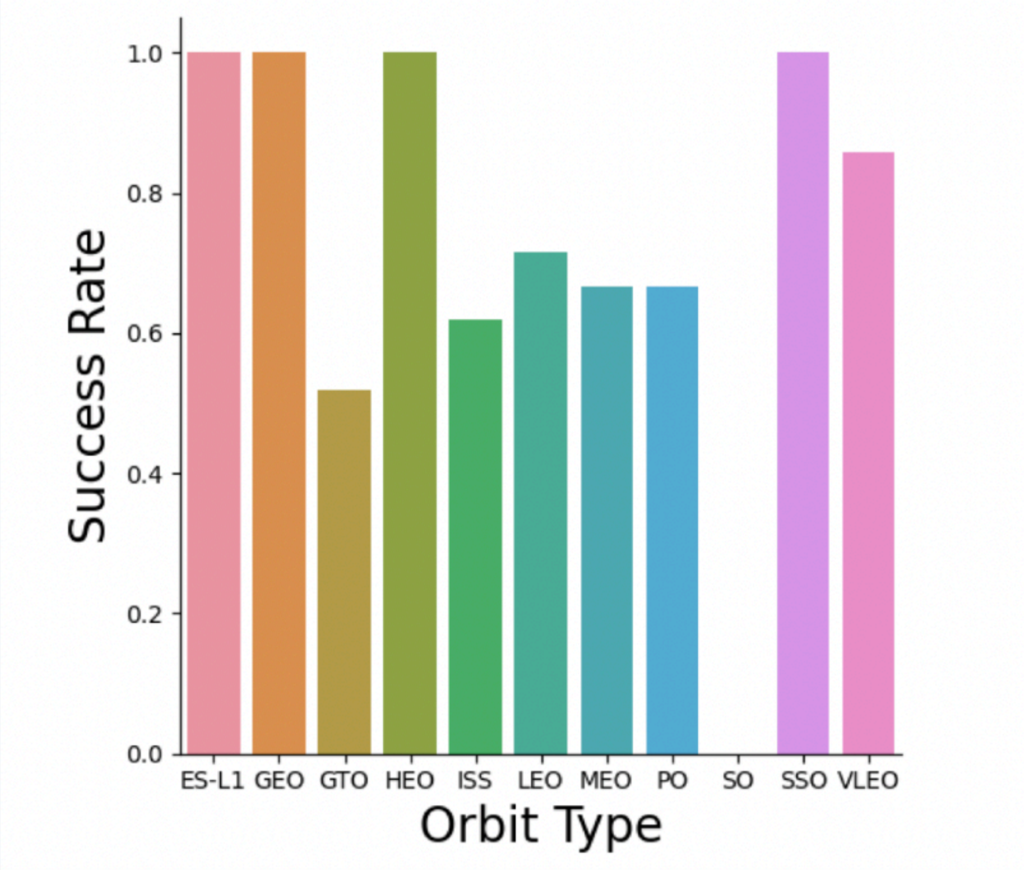
- KSC LC 39A launch -site is the most successful.
- In case of CCAFS SC 40 site, the success rate increases as the flight number increases.
- Many trials in the CCAFS SC 40 launch site

Payload vs. Launch Site



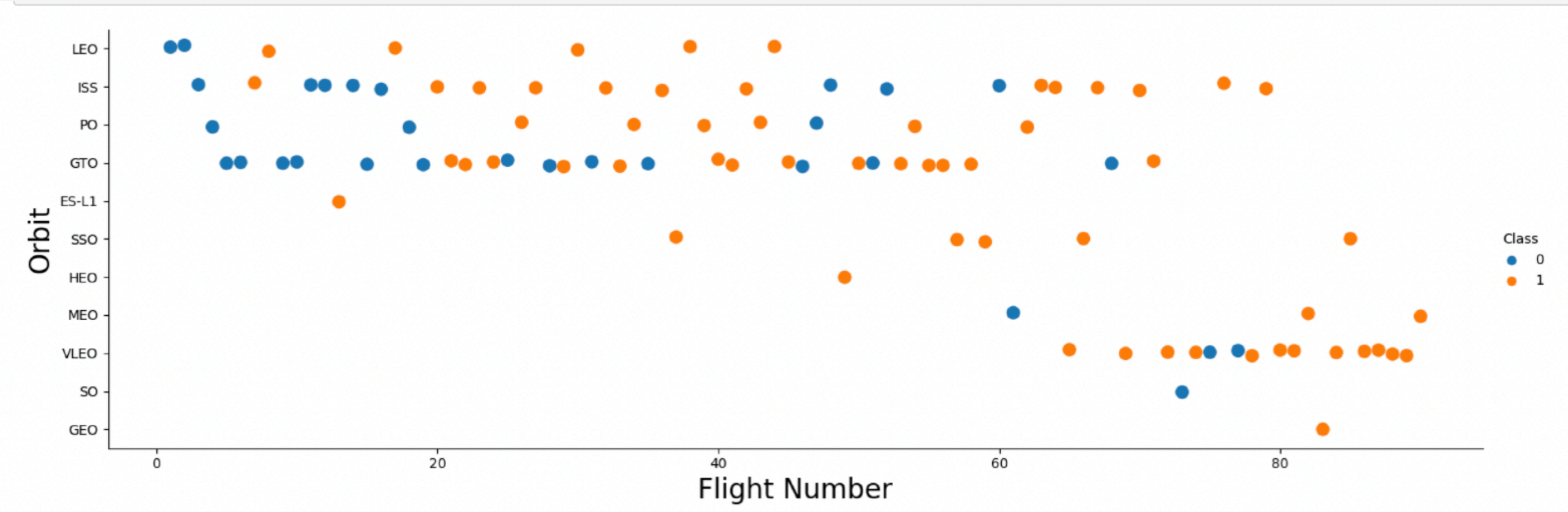
- Pay load mass more than 7000kg shows 100% successful case IN THE CCADS SLC 40 launch site.
- In the VAFB SLC 4E site shows 100% successful rate over the 2000kg payload mass.
- In the KSC LC 39A site shows that 100% success less that 5000kg payload mass.

Success Rate vs. Orbit Type



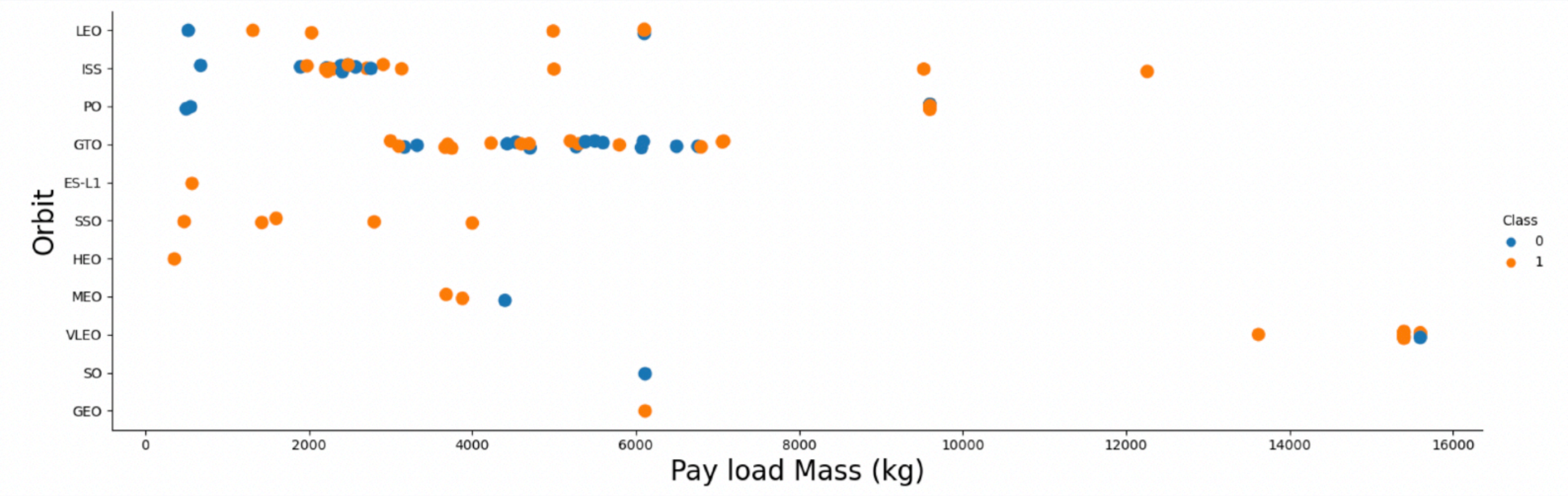
- Orbit type of 'ES-L1', 'GEO', 'HEO', ' SSO' showed 100 % success rate
- There is no success at the orbit type 'SO'
- Other orbit types showed more that 50% success rate

Flight Number vs. Orbit Type



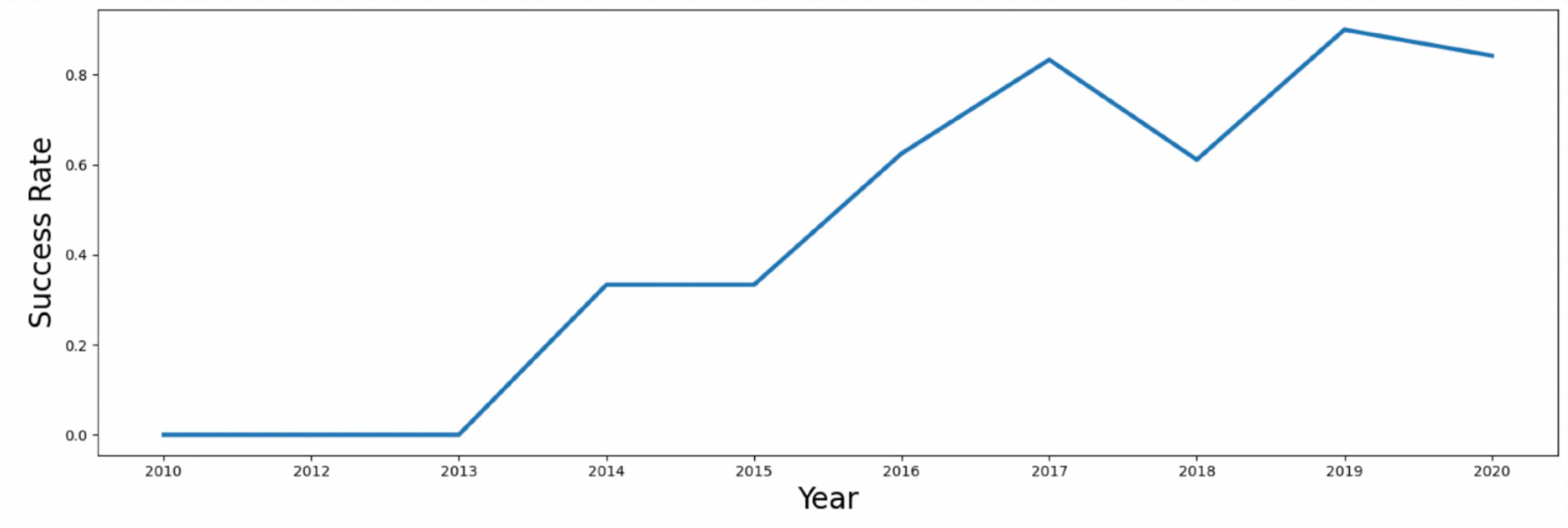
- Orbit type ‘VLEO” shows that there are some relationship between flight number and success outcome.

Payload vs. Orbit Type



- In the Orbit type ES-L1, SSO, HEO, MEO 100% success rate with the less than 4000kg pay load mass.
- More than 12000kg Pay load Mass was attempted in the VLEO orbit and successful rate is high.

Launch Success Yearly Trend



- Since 2013, the success rate has increased rapidly, showing a success rate of over 80% from 2017, slightly decreased in 2018, but recovered by 2020

All Launch Site Names (SQL Queries)

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

→ *“SELECT distinct Launch_Site FROM SPACEXTBL;”*

	Launch_Site
0	CCAFS LC-40
1	VAFB SLC-4E
2	KSC LC-39A
3	CCAFS SLC-40

- **There are 4 district launch sites**

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

→ *"SELECT * from SPACEXTBL where Launch_Site like 'CCA%' limit 5;"*

	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
0	2010-04-06 00:00:00	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2010-08-12 00:00:00	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	2012-05-22 00:00:00	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	2012-08-10 00:00:00	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	2013-01-03 00:00:00	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- **First launch date was 2010 Apr 6th and shows all columns and 5 rows**
- **Use Limit function for showing for 5**

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

→ *"SELECT sum(PAYLOAD_MASS__KG_) FROM SPACEXTBL where CUSTOMER = 'NASA (CRS)';"*

	sum(PAYLOAD_MASS__KG_)
0	45596

- **Apply SUM on the "PAYLOAD_MASS_KG_" column**

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

→ *"SELECT avg(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE Booster_Version = 'F9 v1.1';"*

	avg(PAYLOAD_MASS__KG_)
0	2928.4

- **Apply AVG function on the "PAYLOAD_MASS_KG_" column**

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

→ *'SELECT min(Date) from SPACEXTBL WHERE "Landing _Outcome" = "Success (ground pad)”;’*

	min(Date)
0	2015-12-22 00:00:00

- **Used MIN for search first landing day.**

Successful Drone Ship Landing with Payload between 4000 and 6000

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

→ *'SELECT "Booster_Version" FROM SPACEXTBL WHERE "Landing _Outcome" ="Success (drone ship)" AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;'*

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

- **Use AND condition for all matching finding**

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

→ *'SELECT "Mission_Outcome",count("Mission_Outcome") from SPACEXTBL GROUP BY "Mission_Outcome";'*

	Mission_Outcome	count("Mission_Outcome")
0	Failure (in flight)	1
1	Success	98
2	Success	1
3	Success (payload status unclear)	1

- **Used count and Group by for discriminate the distinct outcome**

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

*'SELECT BOOSTER_VERSION from SPACEXTBL *

where PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_) from SPACEXTBL);'

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	F9 B5 B1049.7

- Used subquery for search maximum payload

2015 Launch Records

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

→ *'SELECT substr(Date, 6, 2) AS MONTH, substr(Date, 1,4) AS YEAR,"Booster_Version", "Launch_Site" FROM SPACEXTBL WHERE "Landing _Outcome" = "Failure (drone ship)" and substr(Date,1,4) = "2015";'*

	MONTH	YEAR	Booster_Version	Launch_Site
0	10	2015	F9 v1.1 B1012	CCAFS LC-40
1	04	2015	F9 v1.1 B1015	CCAFS LC-40

- Make substructure for extract only month or year

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

'SELECT "Landing _Outcome", count(*) as Count FROM SPACEXTBL

WHERE date between "2010-06-04" and "2017-03-20" GROUP BY "Landing _Outcome" Order BY Count desc;')

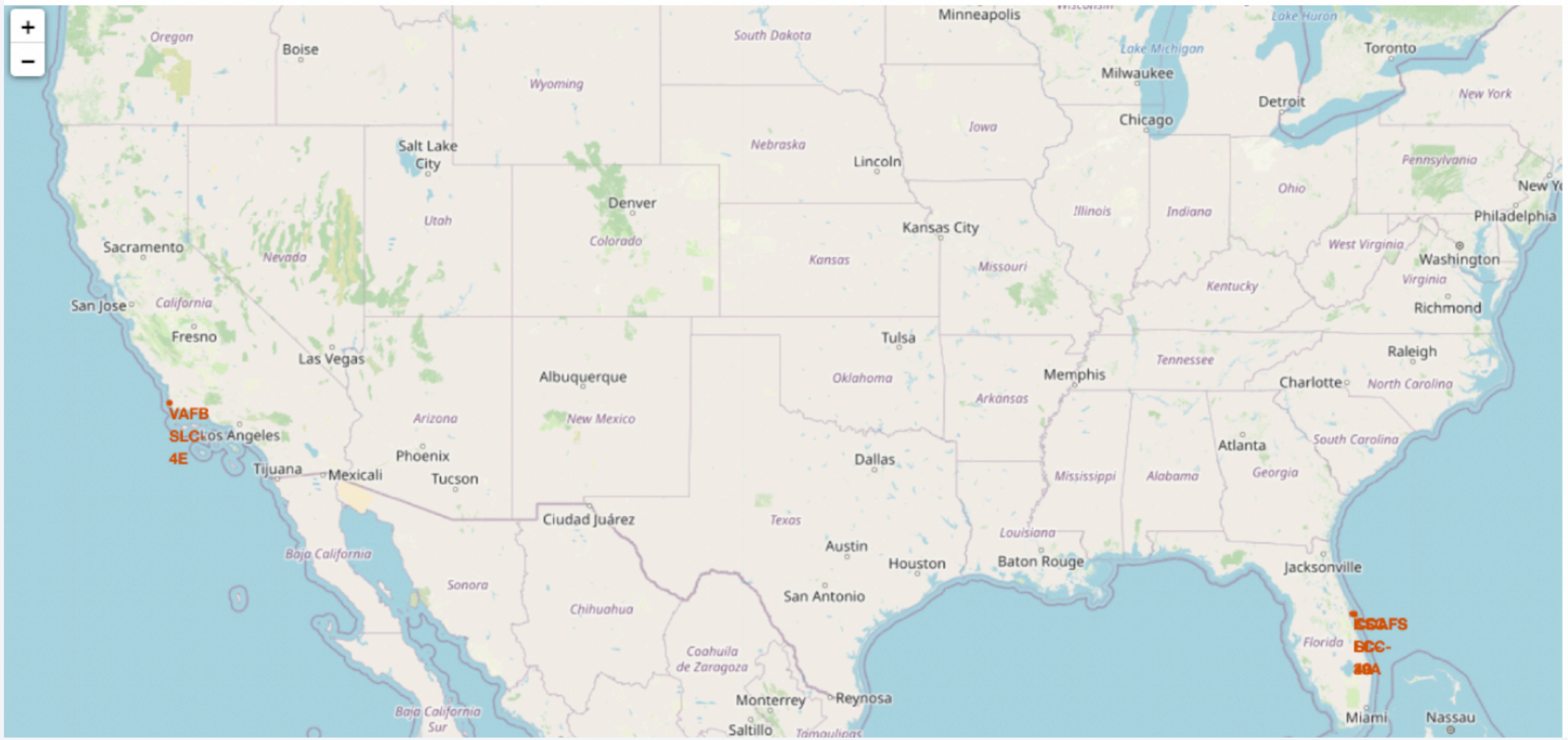
	Landing _Outcome	Count
0	No attempt	10
1	Success (ground pad)	5
2	Success (drone ship)	5
3	Failure (drone ship)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

- **Used for Group by and descending order for sorting**

Section 3

Launch Sites Proximities Analysis

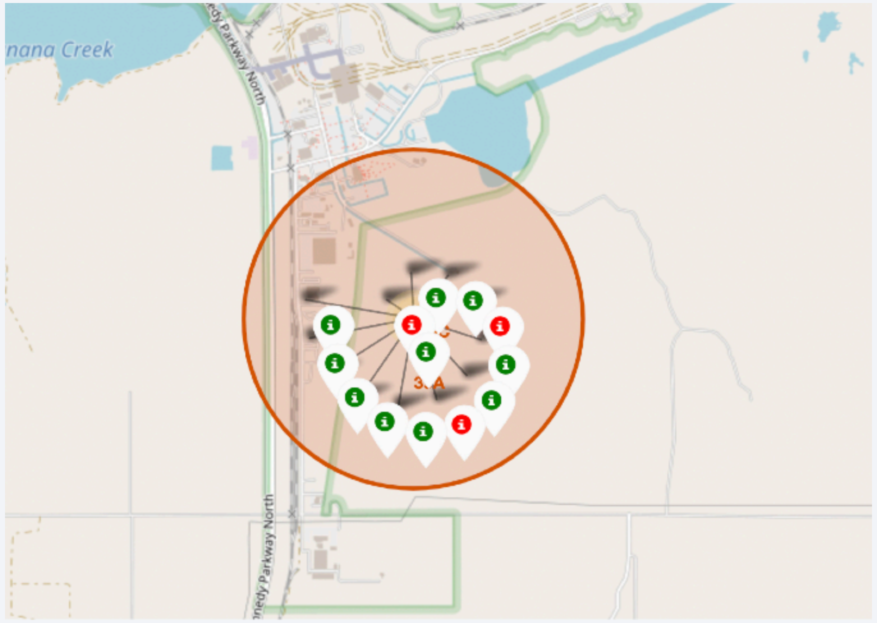
All launch sites' location marker on a global map



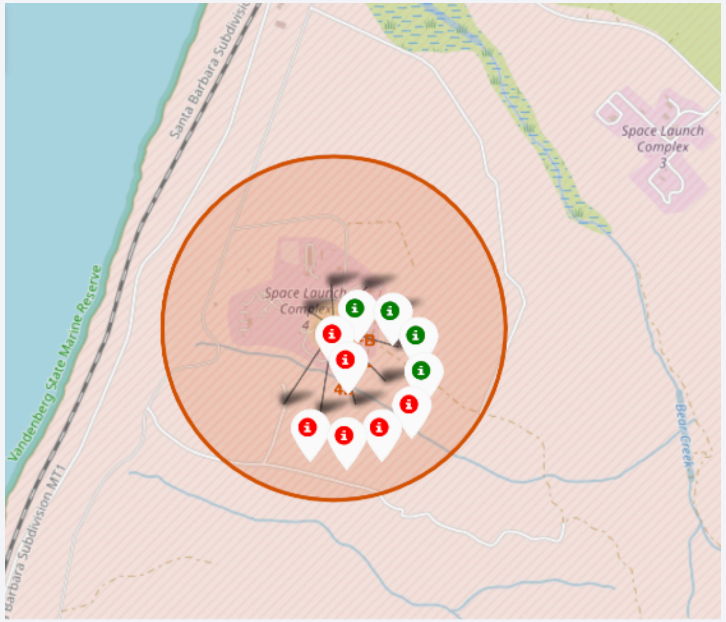
- We can see the launch site neat east coast and west coast.

Color labeled launch outcomes on the map

KSC-LC-39A site neat east coast

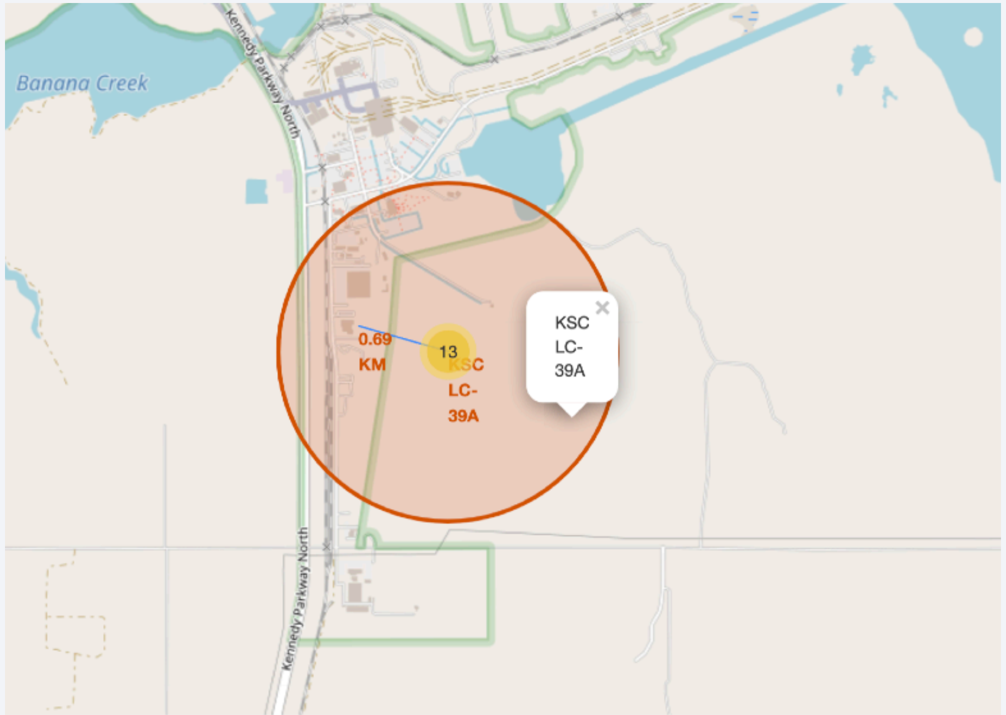


VAFBSLC-4E site near west coast



- KSC_LC _39A showed high success rate.
- On VAFBSLC -4E site had 40 % of success rate among 10 attempts.

The distance between launch site and to its proximities

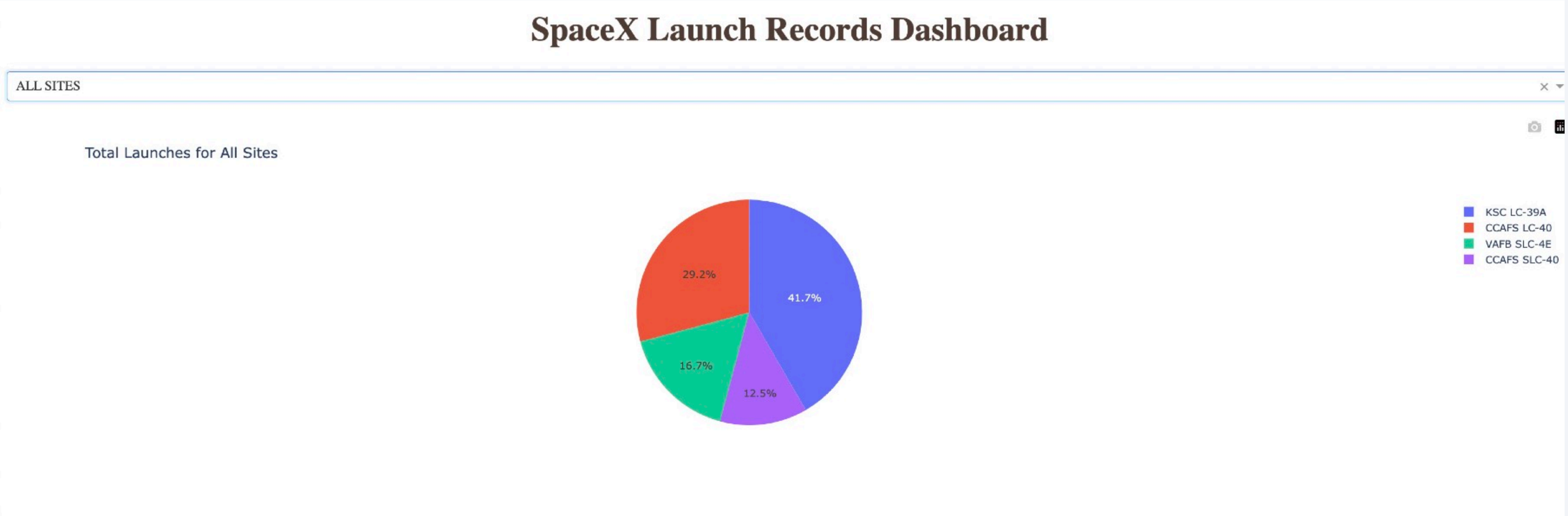


- KSC-LC-4E site and to nearest rail road distance : 0.69km (It’s near)

Section 4

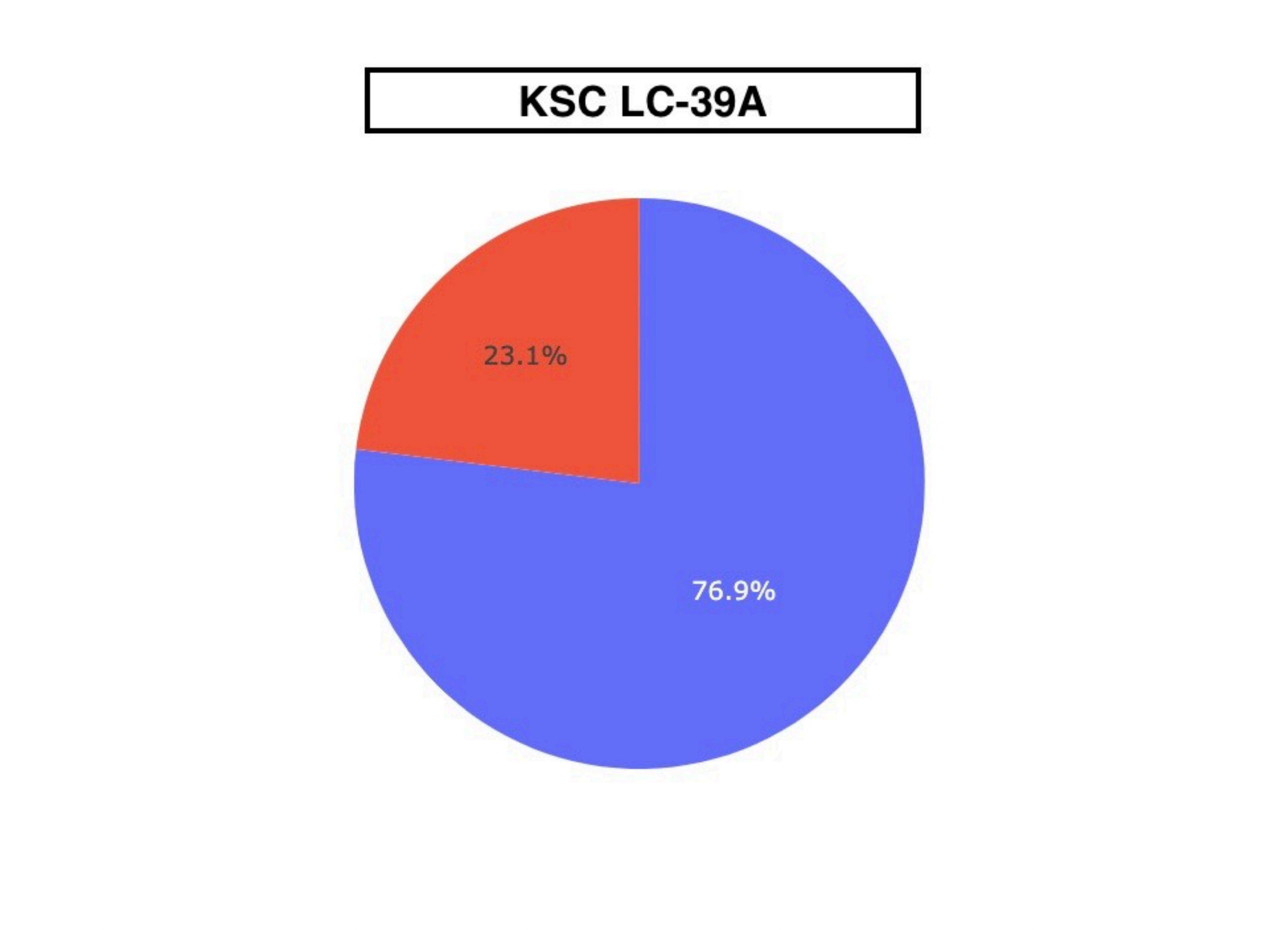
Build a Dashboard with Plotly Dash

Total Success Percentage by All Launch_sites(4 Sites)



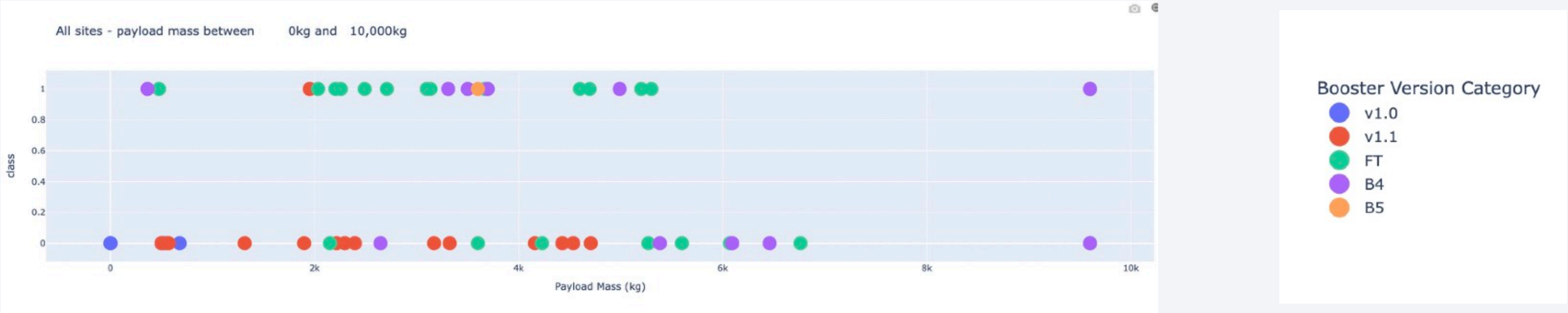
- ‘KSC LC 39A ‘ is most successful launch site(41.7%)

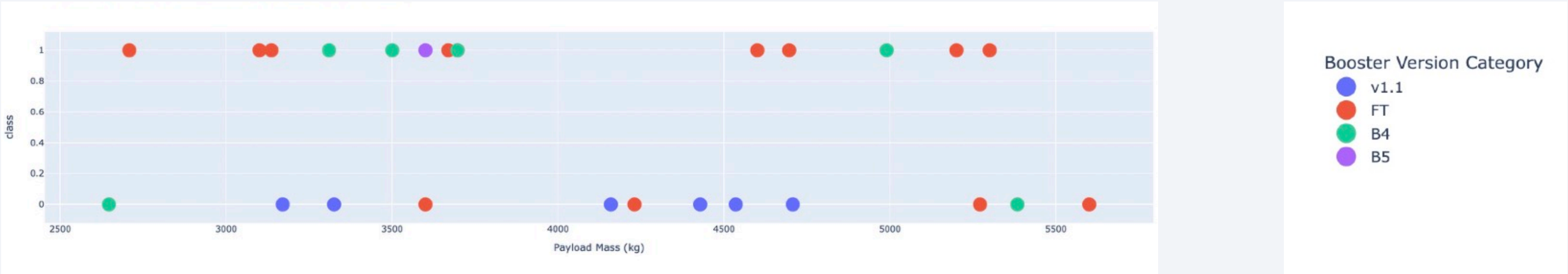
The most highest success rate site



- In the KSC LC-39A site shows that 76.9% success rate and 23.1% failure rate.

Relationship Payload and launch success in all sites



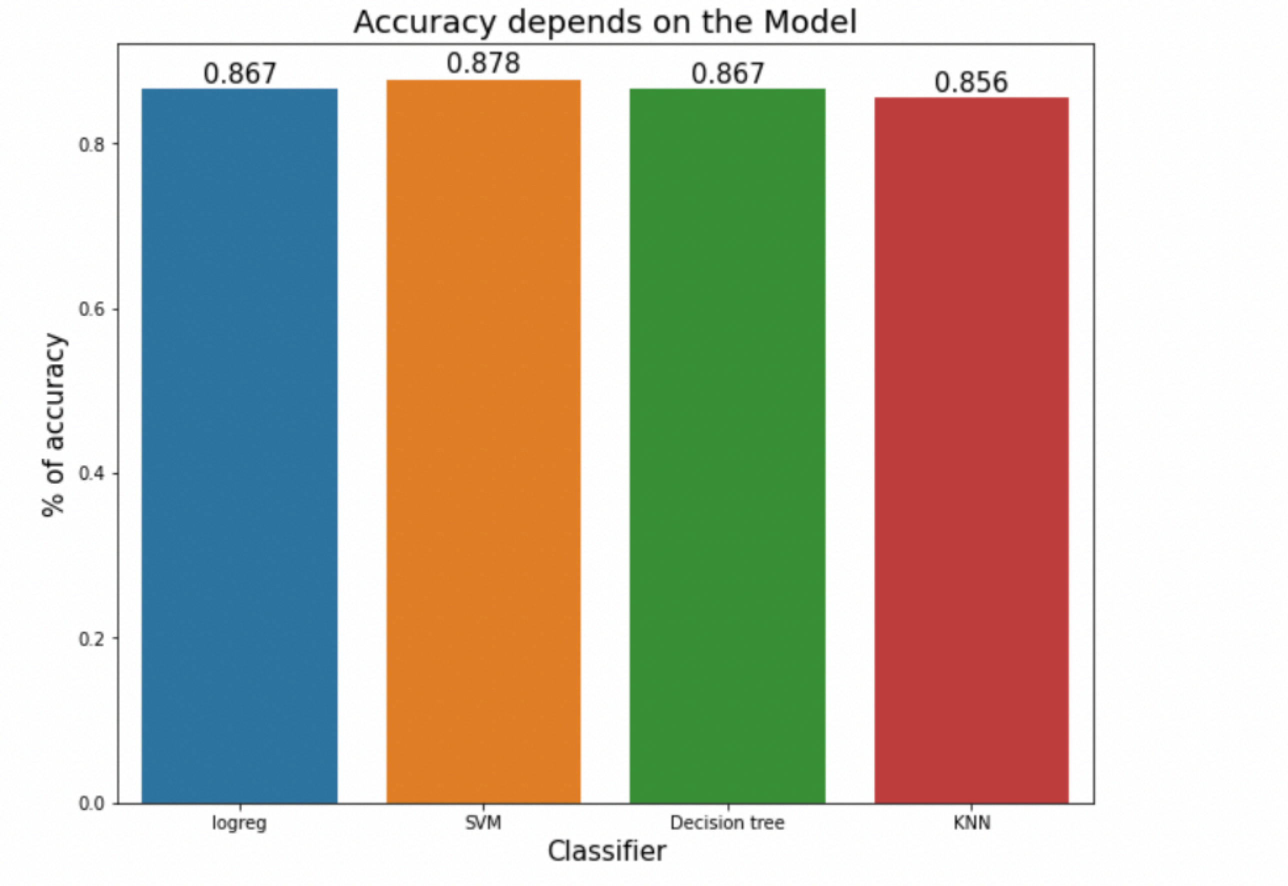


- Payload Mass between 2000 - 6000 kg has high successful late and booster version ‘FT’ was most affected.
- Regardless payload mass, v1.1 booster showed very low success rate.

Section 5

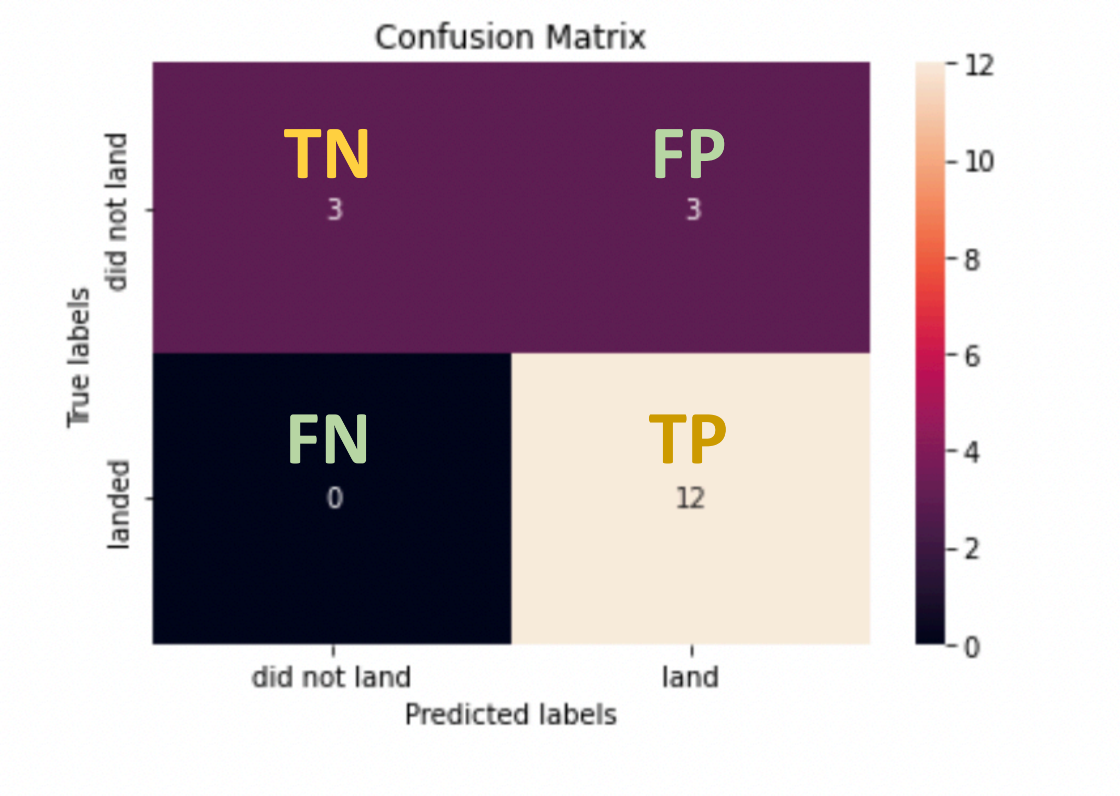
Predictive Analysis (Classification)

Classification Accuracy



- Accuracy is almost similar in 4 models
- SVM is the highest accuracy model

Confusion Matrix



- TN : True Negative
- TP : True Positive

- FP : False Positive
- FN : False Negative

- Sensitivity = TP /(TP+FN) =1.0
- Precision = TP / (TP+FP) = 0.8

Conclusions

- KSC-LC-39A site is the most successful site.
- Success rate has been increased with each passing year
- Pay loss mass is affected by orbit type and booster version
- SVC(support vector machine) is the good model for prediction of success.
- We can see the specific launch site on the global map and can calculate the distance between its proximities.

Appendix

- You can find all relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that I have created in my github.

https://github.com/delicate99/SpaceX-capston#spacex-capston

Thank you!