IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Kunguma Gangai

08/06/2022

Outline

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

Executive Summary

Summary of methodologies

- Collect the data from a wikipedia page using SPACEX API and Web Scraping using BeautifulSoup.
- Data wrangling
  - Clean the dataset
  - Handle the missing values
  - Format the dataset
  - Export to CSV file
- Data visualization
  - Interactive Dashboard for the dataset using Dash
  - Plot launch sites on Global map using Folium
- Predictive Analysis using classification methodologies

Summary of All Results

- Decision tree model is best performing classification model with the accuracy of 87.5%

Introduction

- Project background and context
    - In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advert ises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX f or a rocket launch. In this lab, you will collect and make sure the data is in the correct format from an API.

- Problems you want to find answers
    - Determine what factors has influence on the land outcome of SpaceX rockets.
    - Determine a ML classification that will be able to predict a outcome landed rocket.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data's are collected from wikipedia page using SPACEX API and Web Scraping
- Perform data wrangling
  - Data's are first cleaned, handle missing values, Format the data and export to CSV file
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Standardize the data, Split the data, Fit the data, model the data and find accuracy value.

Data Collection

Datasets are collected from the URL - https://api.spacexdata.com/v4/launches/past (rocket launch data) by using SpaceX API (REST API)

↓

Make a get request to the SpaceX API to extract information using identification numbers in the rocket launch data.

↓

The response contains massive information about SpaceX launches.

↓

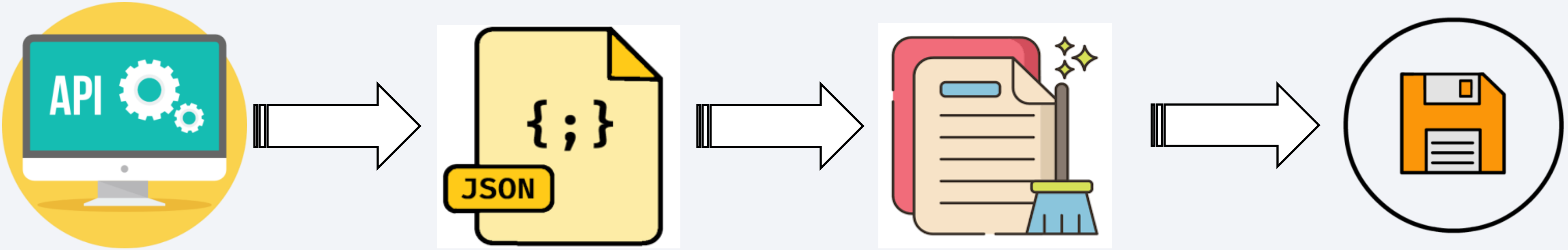
Collect and make sure the data is in the correct format

↓

Clean the requested data and try to discover some more relevant information for this project

Data Collection – SpaceX API

[GitHub - Link]



Using SpaceX API, Request rocket launch data

Decode & normalize response data as json and save it

Clean , Explore and format the data

Save the relevant data and export to csv file.

Data Collection - Scraping

GitHub - Link

Perform web scraping to collect Falcon 9 historical launch records from a Wikipedia page titled https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

`List of Falcon 9 and Falcon Heavy launches`

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]
78	7 January 2020, 02:19:21[492]	F9 B5 ♺ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a tes			
79	19 January 2020, 15:30[494]	F9 B5 ♺ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, rea site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that t crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[5			
80	29 January 2020, 14:07[501]	F9 B5 ♺ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. On			
81	17 February 2020, 15:05[503]	F9 B5 ♺ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 2 booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight prov			
82	7 March 2020, 04:50[506]	F9 B5 ♺ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 ♺)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloa decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successfu			
83	18 March 2020, 12:16[510]	F9 B5 ♺ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and t shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. Ho caused by residual cleaning fluid trapped inside a sensor.[513]			
84	22 April 2020, 19:30[514]	F9 B5 ♺ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)

Web scrap Falcon 9 launch records HTML tables from Wikipedia using Beautiful Soup

Parse the table and convert it into a Pandas Dataframe and export it to a CSV file

Data Wrangling

GitHub - Link

Objective – Perform Exploratory Data Analysis (EDA) to find patterns and determine the label for training supervised model

Load SpaceX dataset from last section

```
df=pd.read_csv("https://cf-courses-data.s3.us.cl
df.head(10)
```

Identify and calculate percentage of missing values in each attribute

```
In [3]: df.isnull().sum()/df.count()

Out[3]: FlightNumber      0.000
        Date              0.000
        BoosterVersion    0.000
        PayloadMass       0.000
        Orbit             0.000
        LaunchSite        0.000
        Outcome           0.000
        Flights           0.000
```

Identify numerical and categorical columns

```
In [4]: df.dtypes

Out[4]: FlightNumber        int
        Date                obje
        BoosterVersion      obje
        PayloadMass         float
```

Calculate the number of launches on each site

```
# Apply value_counts() on col
df['LaunchSite'].value_counts

5]: CCAFS SLC 40     55
    KSC LC 39A       22
    VAFB SLC 4E      13
    Name: LaunchSite, dtype: int6
```

Create a landing outcome label from the number of occurrence of each orbit and mission outcome per orbit type

```
]: GTO
   ISS
   VLEO
   PO
   LEO
   SSO
   MEO
```

```
]: True ASDS
   None None
   True RTLS
   False ASDS
   True Ocean
   None ASDS
```

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class = []
for i in range(len(df)):
    if df['Outcome'][i] in bad_outco
        landing_class.append(0)
    else:
        landing_class.append(1)
```

EDA with Data Visualization

GitHub - Link

Objective – Perform Exploratory data analysis and Feature engineering using Pandas and Matplotlib

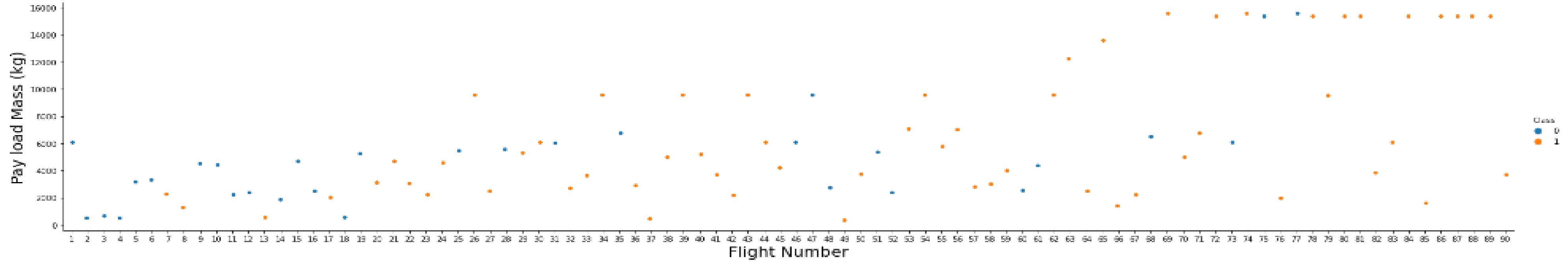
Scatter Plot

Using scatter plot to visualize how different variables would affect the launch outcome

Flight Number Vs Payload Mass

We can plot out the `FlightNumber` vs. `PayloadMass` and overlay the outcome of the launch. We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.

```
sns.catplot(y="PayloadMass", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```



We see that different launch sites have different success rates. `CCAFS LC-40`, has a success rate of 60 %, while `KSC LC-39A` and `VAFB SLC 4E` has a success rate of 77%.

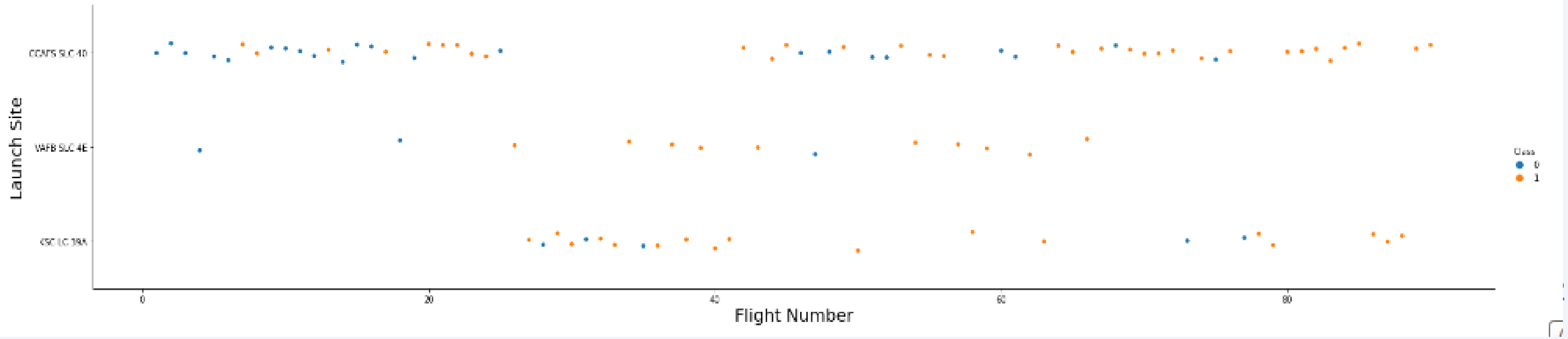
EDA with Data Visualization

[GitHub - Link]

Flight Number Vs Launch Sites

Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`,set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

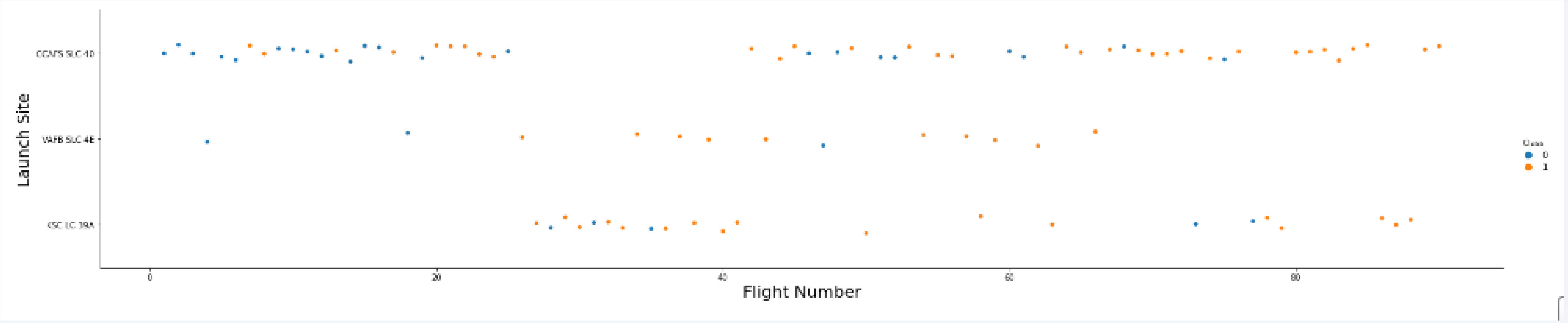
```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()

# We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and
```



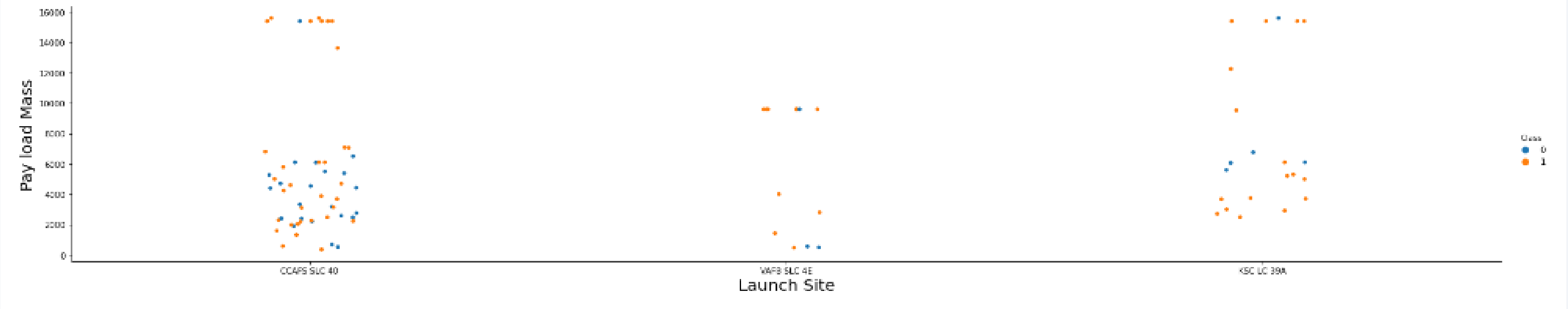
EDA with Data Visualization

[GitHub - Link]

Flight Number Vs Launch Sites



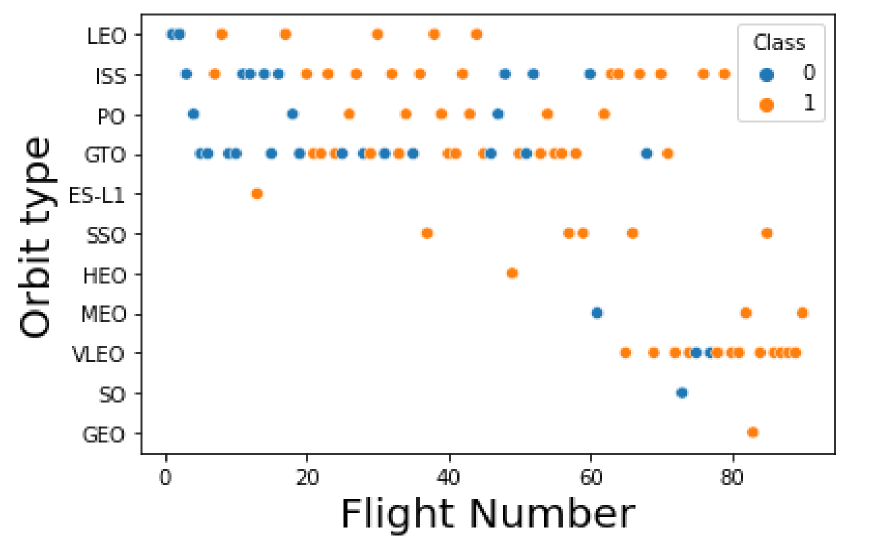
Payload Vs Launch Sites



EDA with Data Visualization

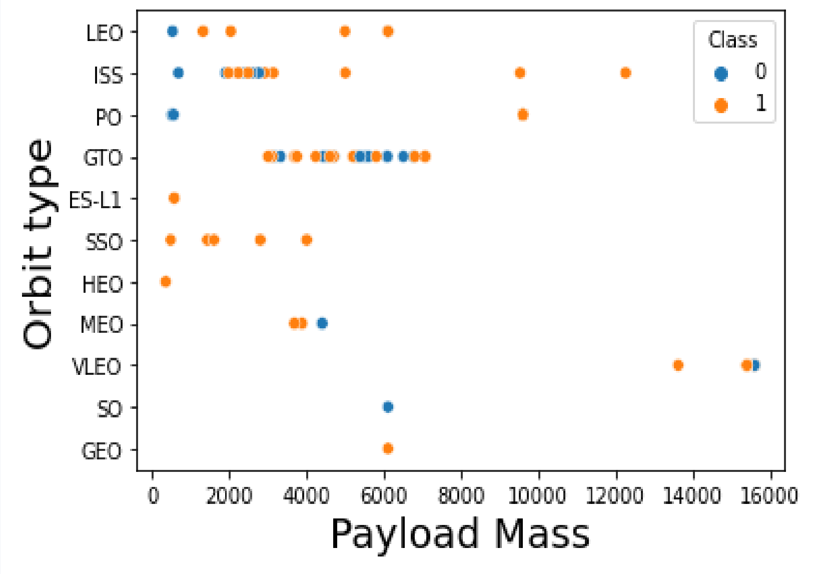
GitHub - Link

Flight Number Vs Orbit type



We can see, LEO orbit success appears related to the number of flights; On the other hand GTO orbit have no relationship with Flight number

Payload Mass Vs Orbit type



Polar, LEO and ISS have successful landing for heavy payload; but for GTO, couldn't distinguish success and failure landing

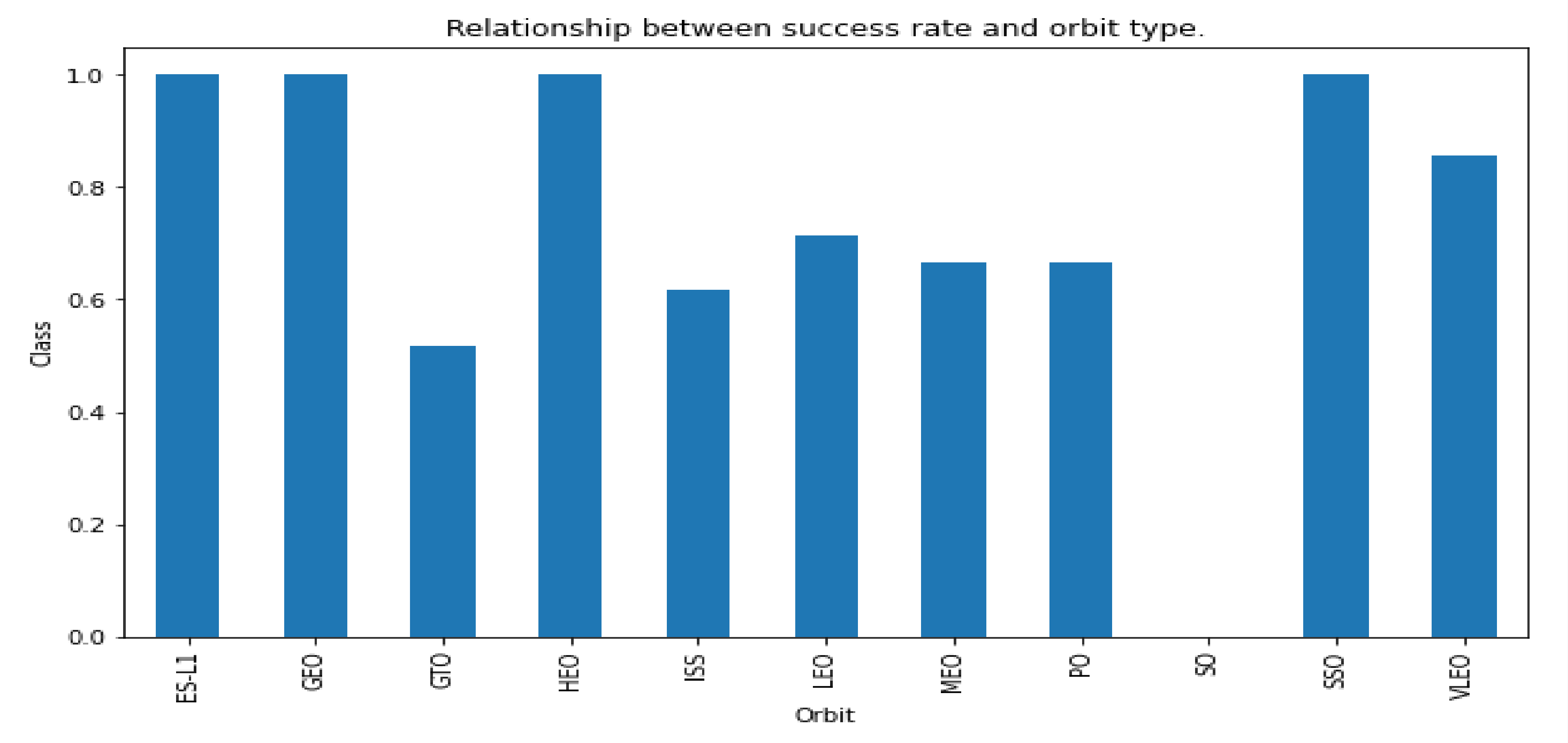
EDA with Data Visualization

[GitHub - Link]

Bar Chart

Using bar chart to visualize success rate of each orbit type

Success Rate Vs Orbit Type



EDA with SQL

[GitHub - Link]

Queries:

1. Display the names of the unique launch sites in the space mission
2. Display 5 records where launch sites begin with the string 'CCA'
3. Display the total payload mass carried by boosters launched by NASA (CRS)
4. Display average payload mass carried by booster version F9 v1
5. List the date when the first successful landing outcome in ground pad was acheived.
6. List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
7. List the total number of successful and failure mission outcomes
8. List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
9. List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
10. 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

1.
```sql
%sql SELECT DISTINCT launch_site FROM SPACEXTBL
```

2.
```sql
%%sql
SELECT *
FROM SPACEXTBL
WHERE launch_site LIKE 'CCA%'
LIMIT 5
```

3.
```sql
%%sql
SELECT SUM(PAYLOAD_MASS__KG_) AS "TOTAL PAYLOAD MASS (Kg)"
FROM SPACEXTBL
WHERE CUSTOMER LIKE 'NASA (CRS)'
```

4.
```sql
%%sql
SELECT AVG(PAYLOAD_MASS__KG_) AS "AVERAGE PAYLOAD MASS (Kg)"
FROM SPACEXTBL
WHERE BOOSTER_VERSION LIKE 'F9 v1.1'
```

5.
```sql
%%sql
SELECT MIN(DATE) AS "First Date - Successful Landing"
FROM SPACEXTBL
WHERE LANDING__OUTCOME LIKE 'Success (ground pad)'
```

6.
```sql
%%sql
SELECT
    BOOSTER_VERSION AS "Booster name"
FROM
    SPACEXTBL
WHERE
    LANDING__OUTCOME LIKE 'Success (drone ship)'
    AND (PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000)
```

7.
```sql
%%sql
SELECT
    DISTINCT (LANDING__OUTCOME),
    COUNT(LANDING__OUTCOME)
FROM
    SPACEXTBL
GROUP BY LANDING__OUTCOME
```

8.
```sql
%%sql
SELECT
    DISTINCT (BOOSTER_VERSION)
FROM
    SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL)
```

9.
```sql
%%sql
SELECT
    LANDING__OUTCOME,
    BOOSTER_VERSION,
    LAUNCH_SITE
FROM
    SPACEXTBL
WHERE
    DATE LIKE '2016%' AND LANDING__OUTCOME LIKE 'Failure (drone ship)'
```

10.
```sql
%%sql
SELECT
    LANDING__OUTCOME,
    COUNT(LANDING__OUTCOME)
FROM
    SPACEXTBL
WHERE
    DATE BETWEEN '2010-06-04' and '2017-03-20'
GROUP BY
    LANDING__OUTCOME
ORDER BY COUNT(LANDING__OUTCOME) DESC
```

Build an Interactive Map with Folium

GitHub - Link

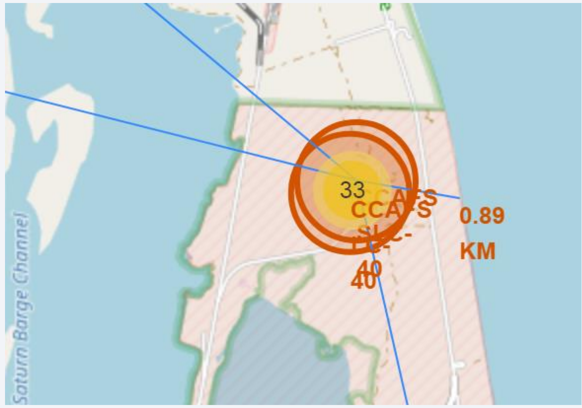
Folium Circles

Folium Circles are used to mark all launch sites

Folium Markers

Folium Markers are used to show the success and failure mission in each launch site

Folium Lines



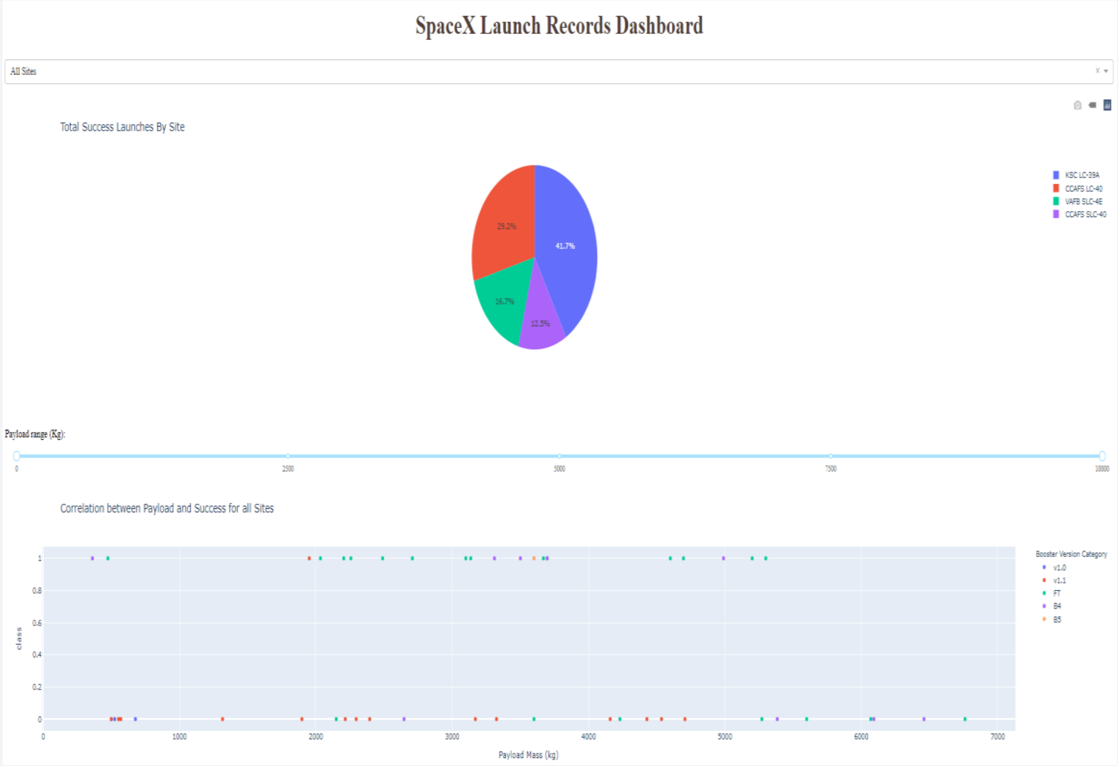
Folium lines are used to calculate the distance between launch sites and city, railway and highways

Build a Dashboard with Plotly Dash

GitHub - Link

Plots/Graphs

- Pie Chart - Used to show total success launches by all sites and specific launch site
- Scatter plot – Used to show correlation between payload and success in all sites and specific sites



Interaction

- Dropdown – Used to filter different launch sites
- Range Slider – Used to select the payload mass weight range of missions

Predictive Analysis (Classification)

GitHub - Link

Build the Model	Evaluate the Model	Find the best performing Model
Load and Convert the data into numpy arrays to do the math operations	Create the GridSearchCV object for all the classification model	Find the tuned hyperparameters using best_params_
Standardize the dataset	Fit all the model objects by passing training datasets	Find the accuracy on the validation data using best_score_
Split the data into training and testing dataset	Output the GridSearchCV object and analyze the values	Calculate the accuracy on the test data using score method
Create the classification objects of KNN, SVM, Decision tree and Logistic regression		Using yhat value plot the confusion matrix and examine it.

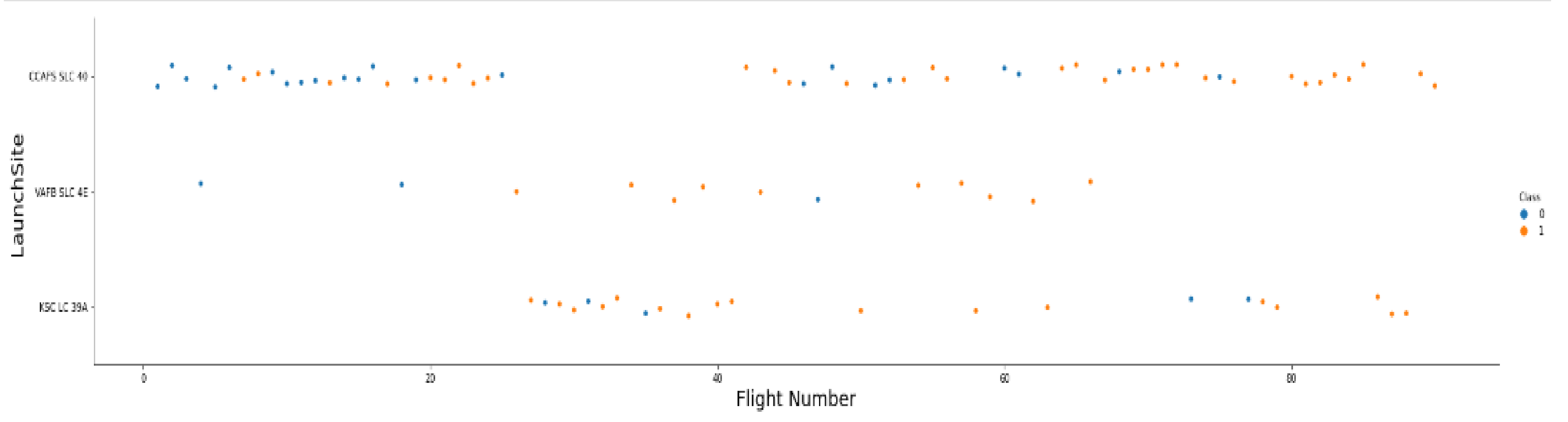
Results

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

Section 2

Insights drawn from EDA

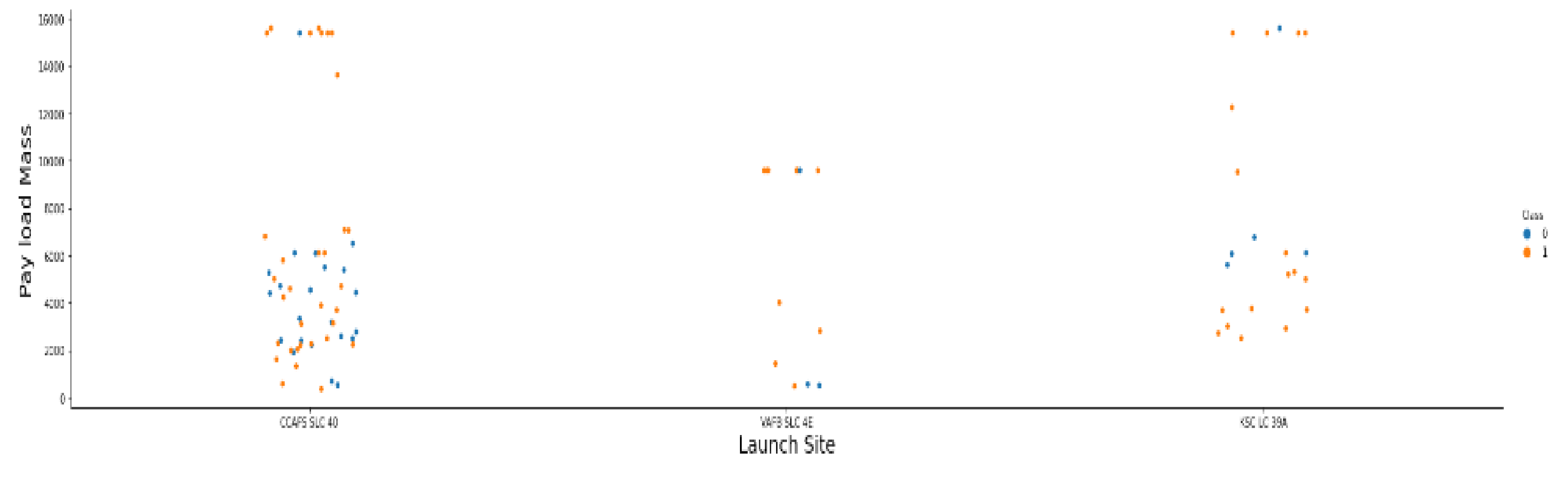
Flight Number vs. Launch Site



Insights / Explanation

We see that different launch sites have different success rates.
CCAFS LC-40, has a success rate of 60 %,
KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

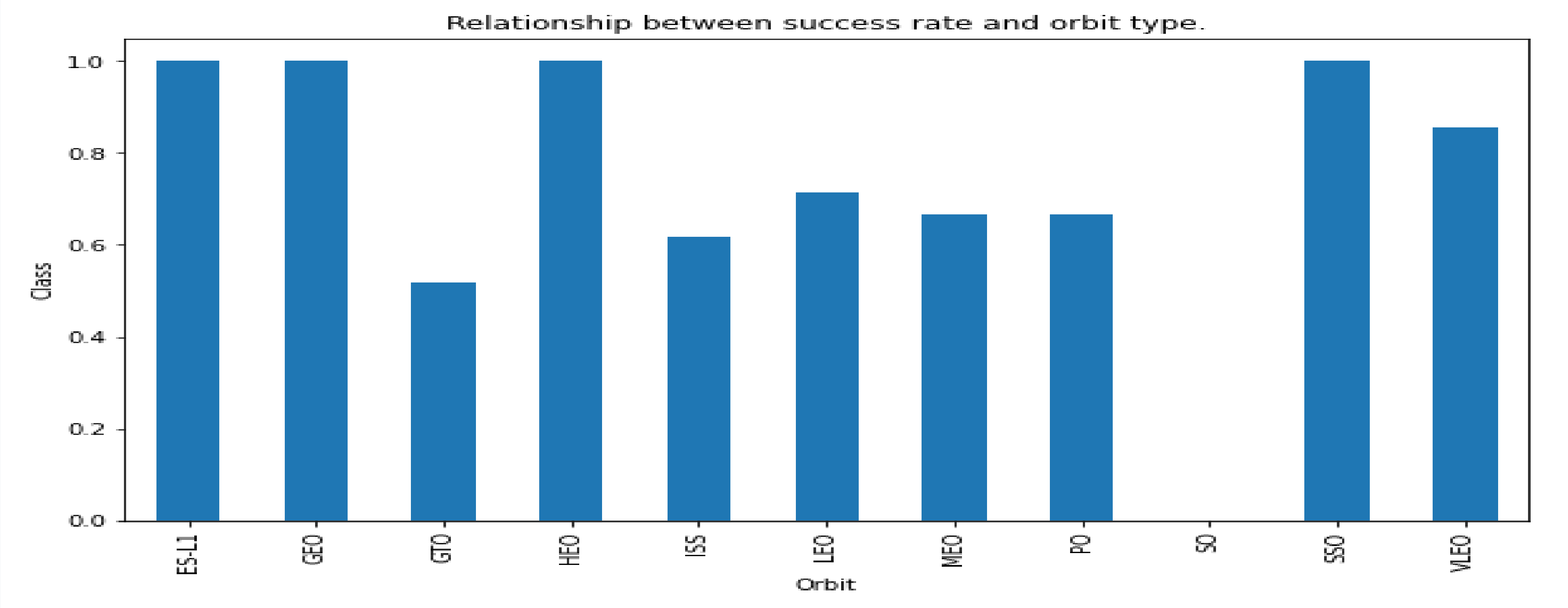
Payload vs. Launch Site



Insights/ Explanation

Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).
We can also see that success rate of launch site with heavy payload mass are higher when compared to launch sites with low payload mass

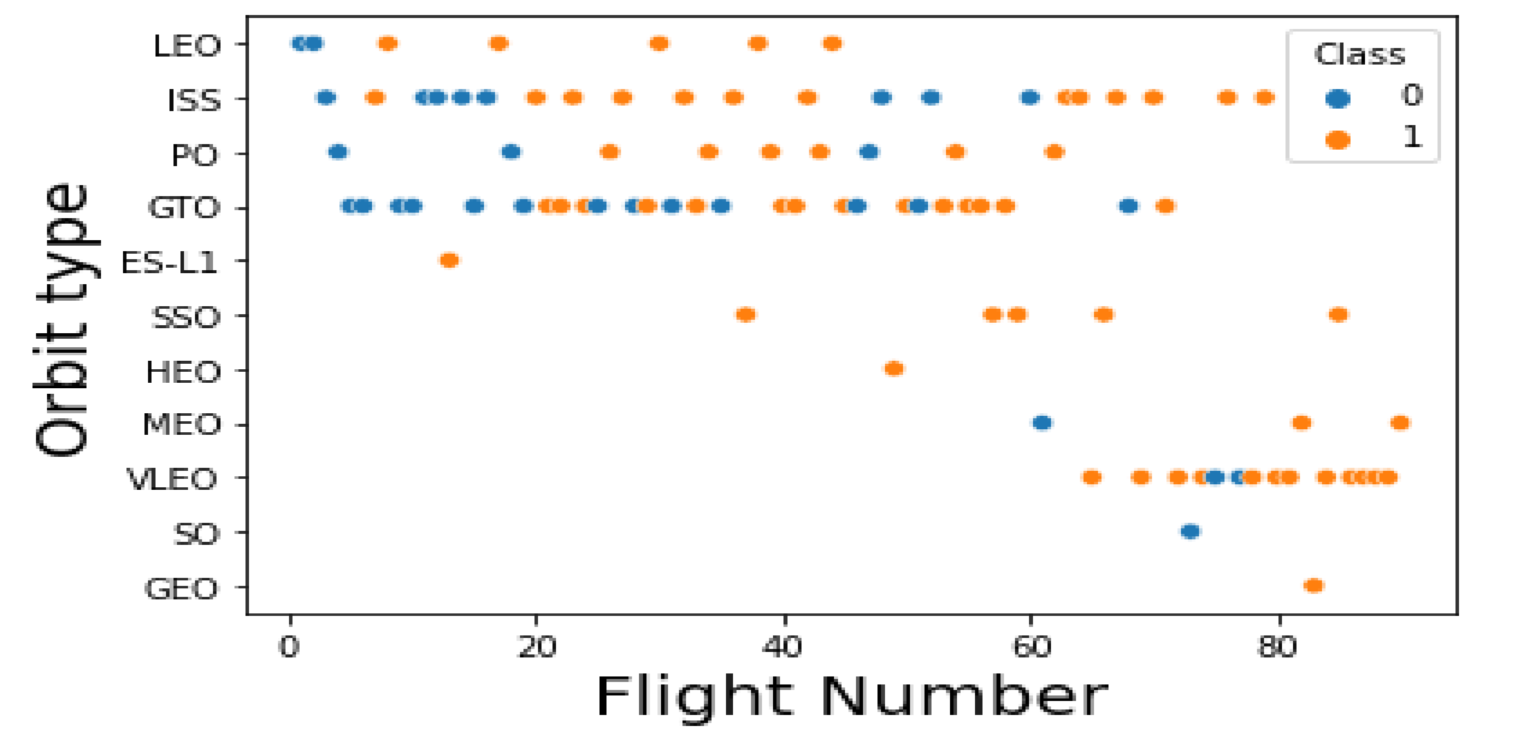
Success Rate vs. Orbit Type



Insights/ Explanation

We can observer that ES-L1, GEO, HEO, and SSO orbit have higher success rate when compared to other orbits. So from the chart we can conclude that the orbit type have significant impact on success rate of a mission launch.

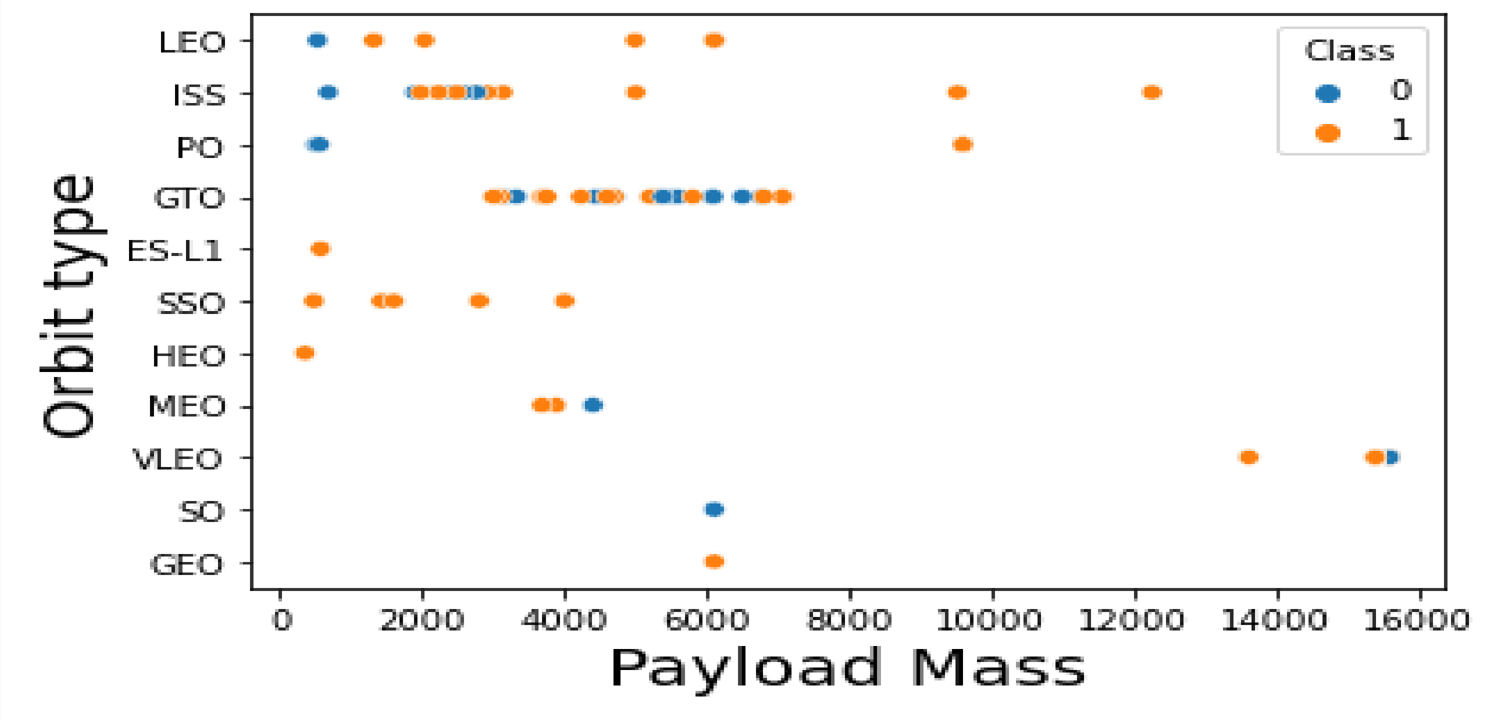
Flight Number vs. Orbit Type



Insights/ Explanation

We can 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.
So we cannot take flight number to predict the success rate of mission launch.

Payload vs. Orbit Type

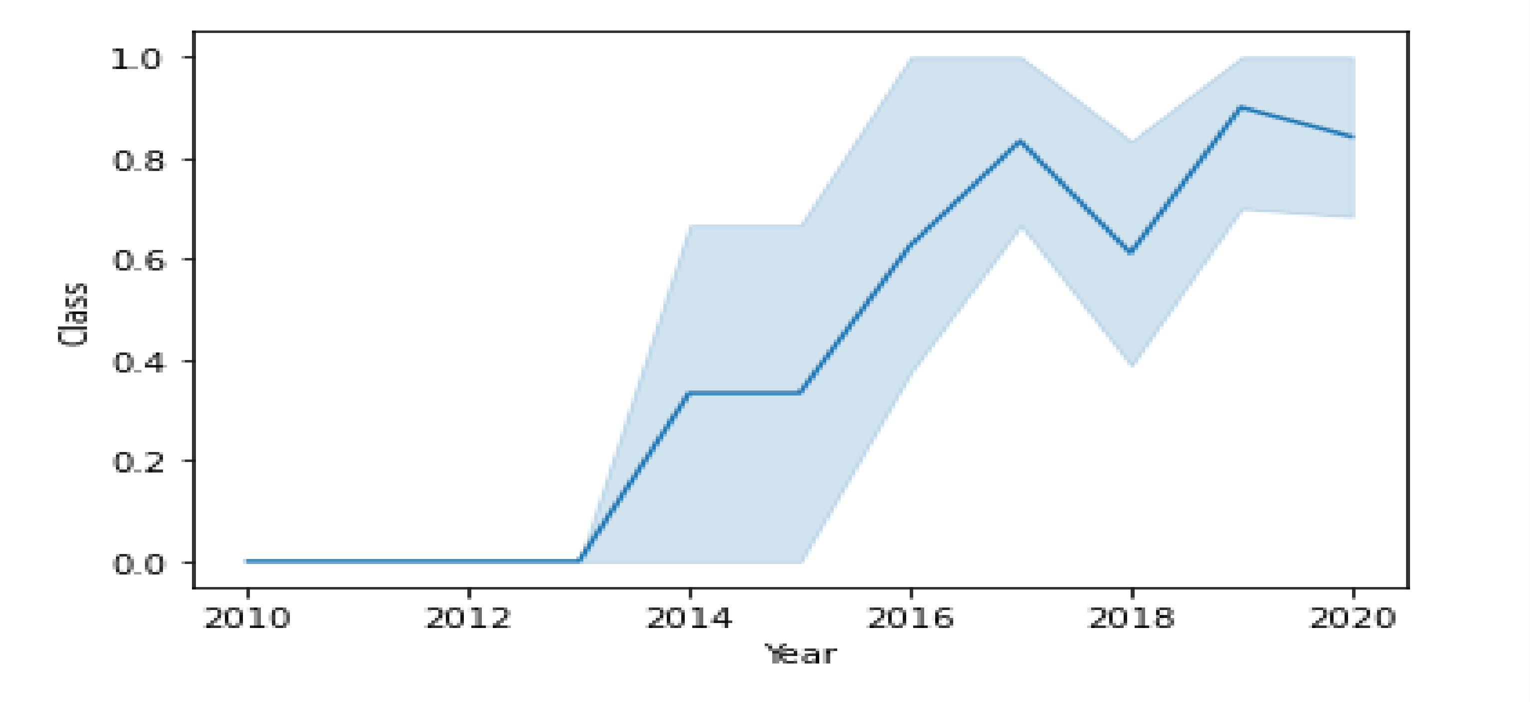


Insights/ Explanation

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.

So we can conclude that payload mass have no significant impact on orbit type success rate.

Launch Success Yearly Trend



Insights/ Explanation

The sucess rate since 2013 kept increasing till 2020
we can observe an increase of success rate by every year, this true positive rate maybe related to advanced technologies and prolonged studies and research on the rocket science field.

All Launch Site Names

- Find the names of the unique launch sites

```
In [76]: %%sql
select DISTINCT(launch_site) from SPACEXTBL

 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[76]:

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

Explanation

'SELECT' - used to get/read data from database
'DISTINCT' - used to get unique values in a mentioned column
SPACEXTBL – Database where we stored the rocket launch data.

Launch Site Names Begin with 'CCA'

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

In [4]:
```
%%sql
select * from SPACEXTBL where launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[4]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Explanation

Sometimes, we may not know/remember the name of a launch site, to get the full name of a launch site we can use 'LIKE' operator at the end of the query to get the name with the given string value.
In this case we would like to know the launch site name that starts with 'CCA'.
Above is the result of 5 records of launch sites name begins with 'CCA'.

Total Payload Mass

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

In [5]:
```
%%sql
select sum(payload_mass__kg_) as total_payload from SPACEXTBL where customer='NASA (CRS)'
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[5]:

total_payload
45596

Explanation

SUM() - Is a sql function used to calculate the total value of a given column
Here, we calculated the total payload mass carried by the boosters launched by NASA(CRS).
From the result we can see that total payload mass carried by NASA(CRS) is 45596 kg.

Average Payload Mass by F9 v1.1

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

In [6]:
```
%%sql
select avg(payload_mass__kg_) as average_payload_mass from SPACEXTBL where booster_version='F9 v1.1'
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[6]:

average_payload_mass
2928

Explanation

AVG() - Is a sql function used to calculate the average or mean values of a given column value
Here, we calculated the average value of payload mass of booster version F9 v1.1.
From result we see that average value of booster version F9 v1.1 payload mass is 2928

First Successful Ground Landing Date

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

```
In [21]: %%sql
select MIN(DATE) as first_success_groundpad from SPACEXTBL where landing__outcome='Success (ground pad)'

 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.

Out[21]:
```

first_success_groundpad
2015-12-22

Explanation

MIN() - Is a sql function used to calculate the minimum value of a given column
Here, we found the First successful landing date of a ground pad outcome. For Date type min() function will output the past date.
From result we see 2015-12-22 is the first successful landing outcome on ground pad.

Successful Drone Ship Landing with Payload 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

In [30]:
```
%%sql
select booster_version from SPACEXTBL where landing__outcome='Success (drone ship)' and payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[30]:

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

Explanation

BETWEEN (value1) AND (value2) - is a sql keyword used to get value range between value 1 and value2
Here we got the names of the booster version which are success in drone ship and have payload mass greater than 4000 and less that 6000.
Result shows the names of the booster version of different launch sites.

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

In [40]:
```
%%sql
select mission_outcome, count(mission_outcome) as total from SPACEXTBL group by mission_outcome
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[40]:

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

Explanation

COUNT() - Is a sql function used to calculate the total number of records in the given column.
GROUP BY – Is used to group the values in a column
Here, we calculated the total number of successful and failure mission from all launch sites
From result, we see success rate is very high (99 success mission outcome) and very less failure (1 failure in flight) and (1 success payload status unclear).

Boosters Carried Maximum Payload

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

In [46]:
```
%%sql
select booster_version from SPACEXTBL where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[46]:

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

Explanation

SUBQUERY – are nothing but nested query or query inside a query
Here, we found the names of the booster version which have carried the maximum payload mass
Form result, we can see that many booster have carried heavy payload mass.

2015 Launch Records

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

In [52]:
```
%%sql
select DATE, booster_version, launch_site, landing__outcome
from SPACEXTBL where landing__outcome = 'Failure (drone ship)' and DATE like '2015%'
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[52]:

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Explanation

To get the failed landing outcome in drone ship we mention column landing_outcome = 'Failure (drone ship)' in WHERE clause and to get only the outcomes that happened in the year 2015 we have used LIKE operator in DATE column.
From result we can see, there are 2 failure drone ship in the year 2015.

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

```
%%sql
select landing__outcome, count(landing__outcome) as rank from SPACEXTBL where DATE between '2010-06-04' and '2017-03-20'
group by landing__outcome
order by rank desc
```

```
 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

landing__outcome	RANK
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

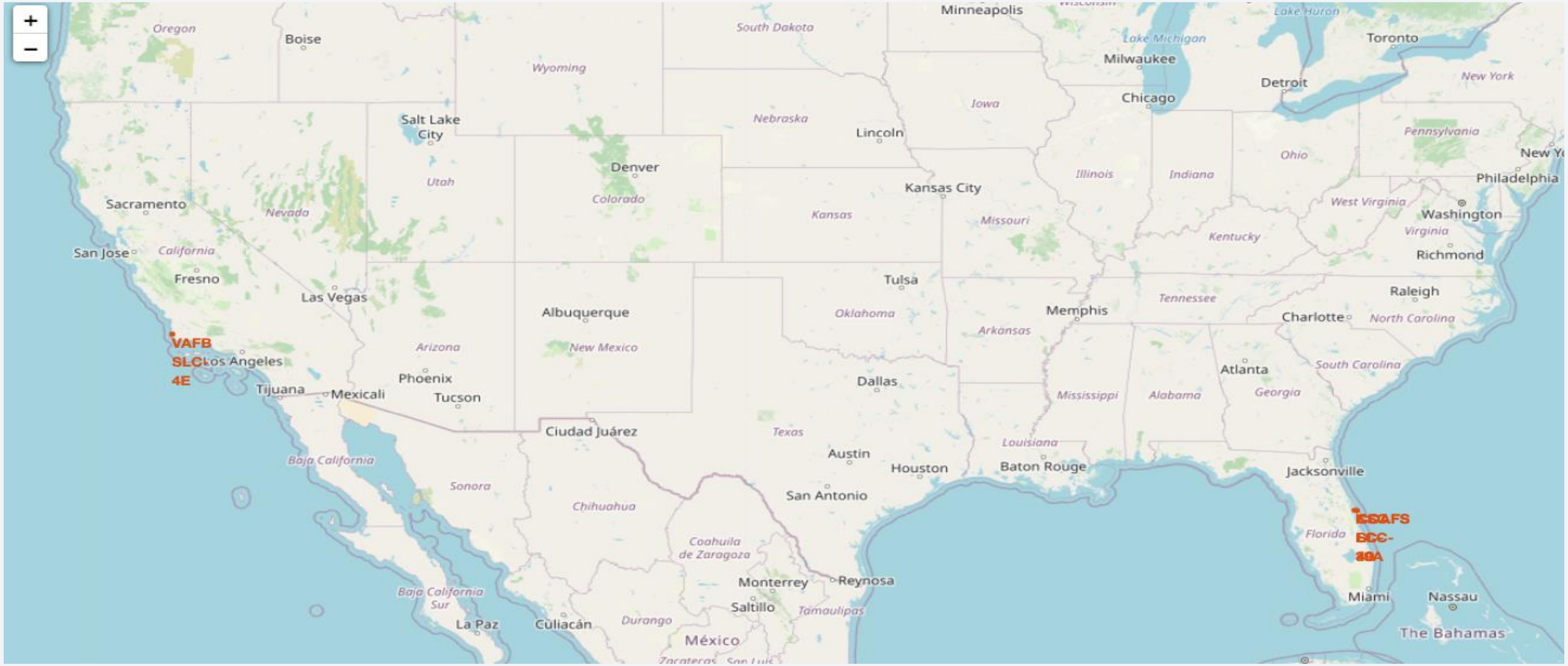
Explanation

ORDER BY – used to order the rows in the table by the values of the mentioned column.
DESC – used to order the value of a column in descending order
Here, we ranked the count of landing outcomes (such as success , failure) between the date 2010-06-04 and 2017-03-20.
From result, we can see that No attempt ranks 10 and precluded (drone ship) ranked 1.

Section 3

Launch Sites Proximities Analysis

All Launch Sites location markers on a Global Map

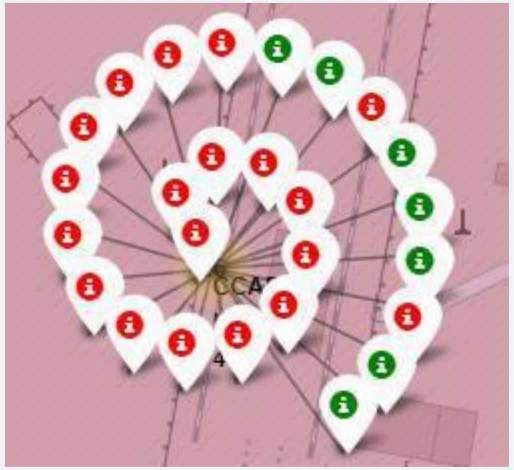


Insights/ Explanation

Above is a Folium Global Map, Here we marked all the locations of the launch sites using folium circle marker. The orange color spot with labels are the location of the launch sites.

Success/Failure Launches for each site on the map

CCAFS LC-40



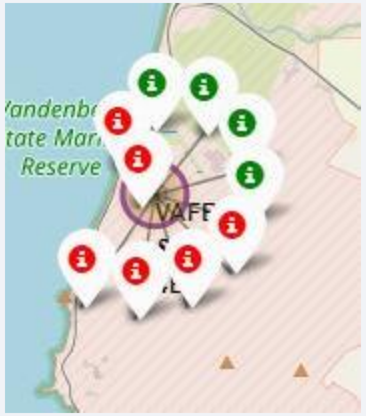
CCAFS SLC-40

Successful mission

Failed mission

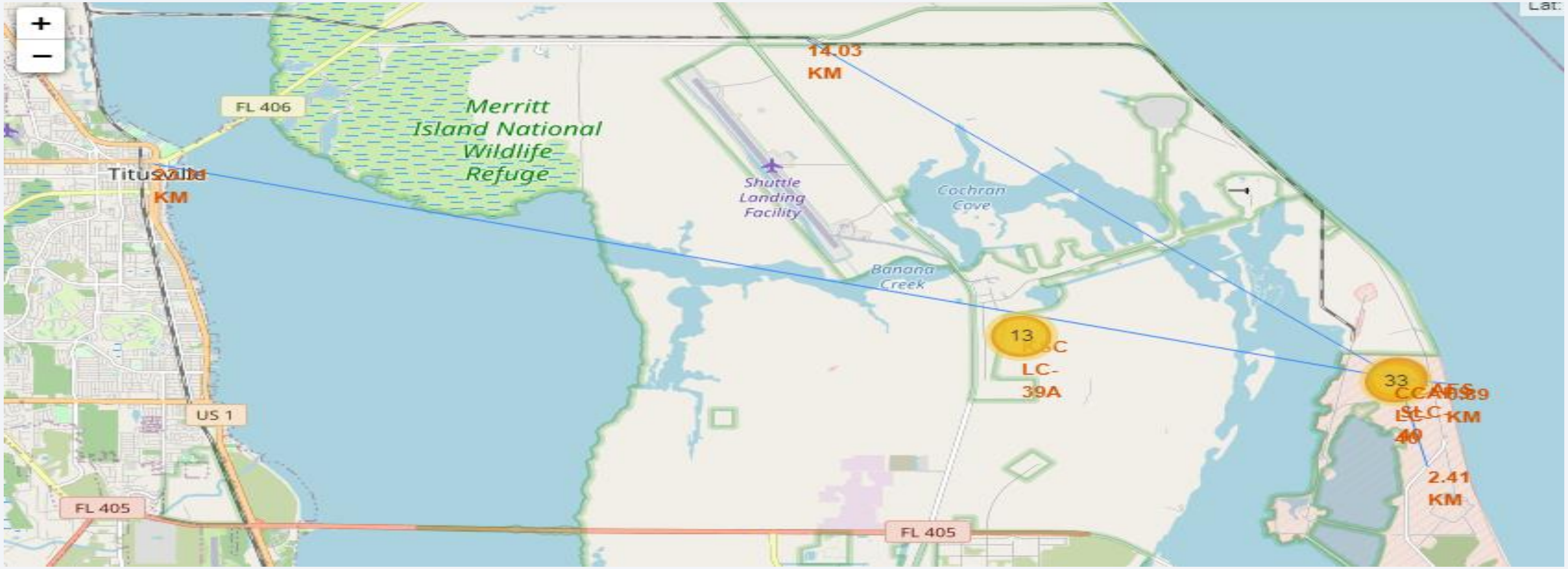
KSC LC-39 A

VAFB SLC-4E



- We have 4 launch sites, from the markers on map we can easily find the success and failure rate of a launch site.
- Here, CCAFS LC-40 have more red markers which means the launch site have more failure rate.
- KSC LC-39 A launch site have more green markers which means more success rate.
- CCAFS SLC-40 and VAFB SLC-4E have equal success and failure mission

Distance from Launch site to City, Railway, Highway and coastline



Insights/ Explanation

The map show the distance between the launch site and its closest coastline, city, railway and highway

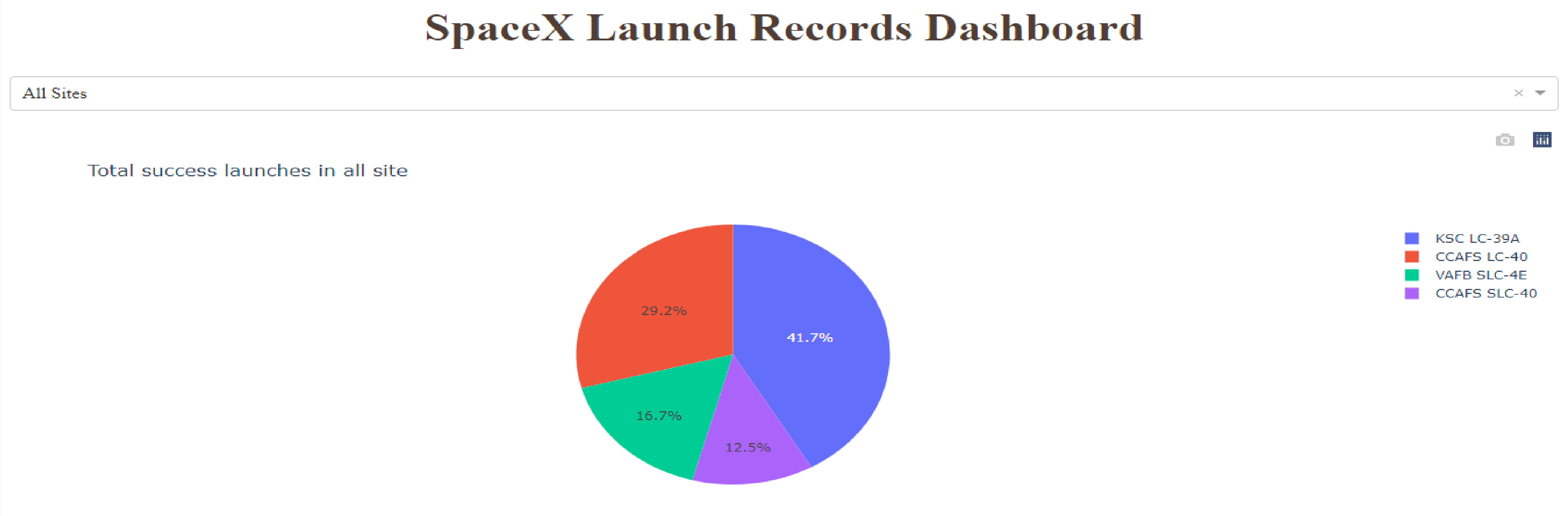
The yellow mark are the launch site locations,

The blue lines are the distance from launch site and coastline, city, railway and highway.

Section 4

Build a Dashboard with Plotly Dash

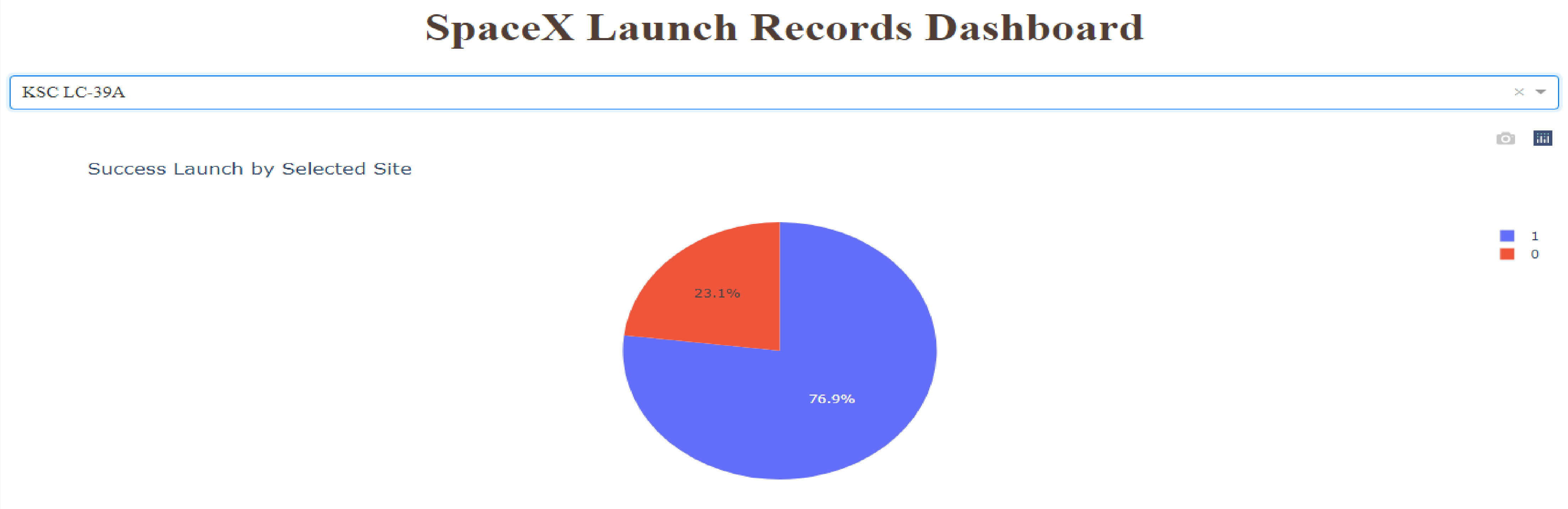
Pie Chart – Launch success count of all sites



Insights/ Explanation

- The Pie chart shows the total success launches in all the sites. The different colors are used to differentiate the launch sites and for easy understanding.
- From the pie chart we can see and easily understand that KSC LC 39A has 41.7 % success rate and marks the highest success rate.
- CCAFS SLC-40 with 12.5% marks the lowest success rate compared to all launch sites.

Launch site with highest success ratio

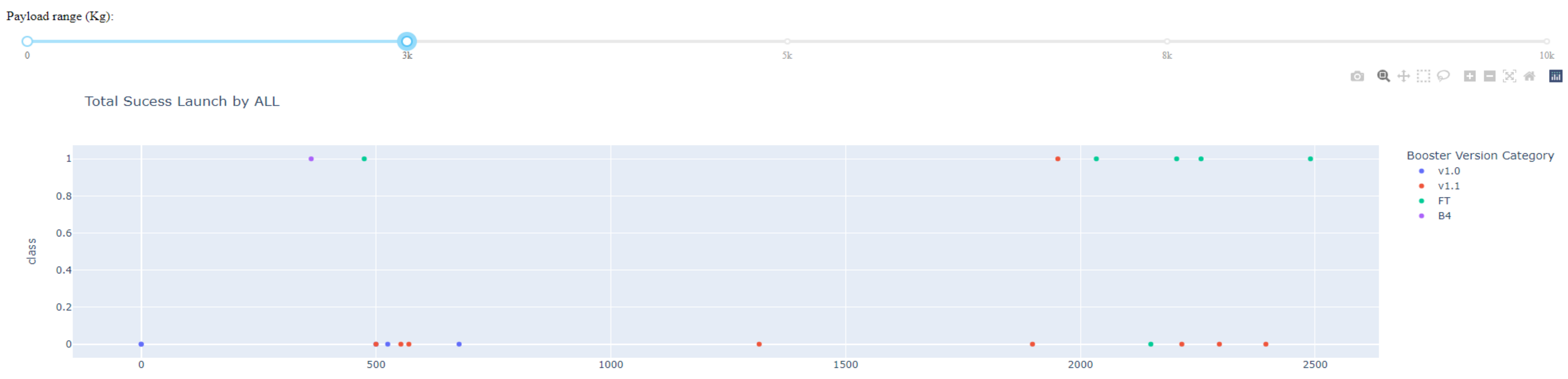


Insights/ Explanation

- Above pie chart shows the highest success ratio of a launch site.
- From the previous pie chart we found that KSC LC-39A marks the highest success ratio with 41.7%.
- In this pie chart we see the success ratio of KSC LC-39A alone.
- The blue color shows the success launches and red color shows the failed launches
- Pie chart shows KSC LC-39A have 76.9% success ratio and 23.1% failure ratio.

Range Slider - Payload vs. Launch Outcome scatter plot for all sites

Payload Mass from 0 to 3000 kg

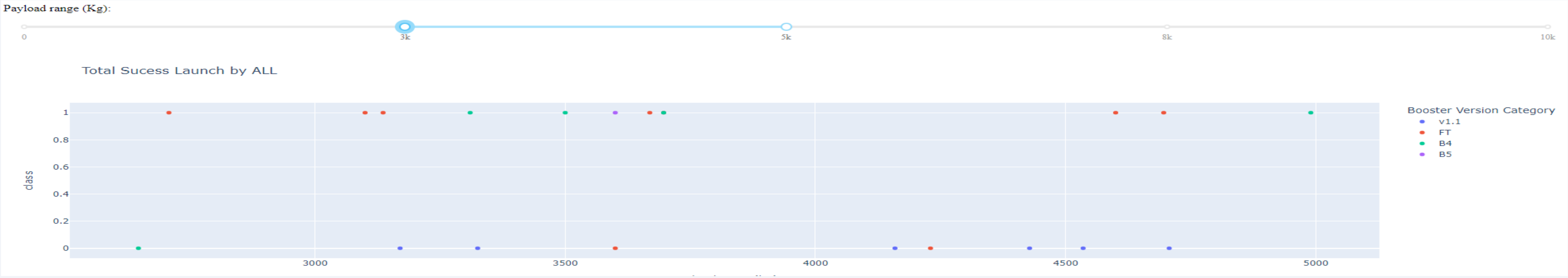


Insights/ Explanation

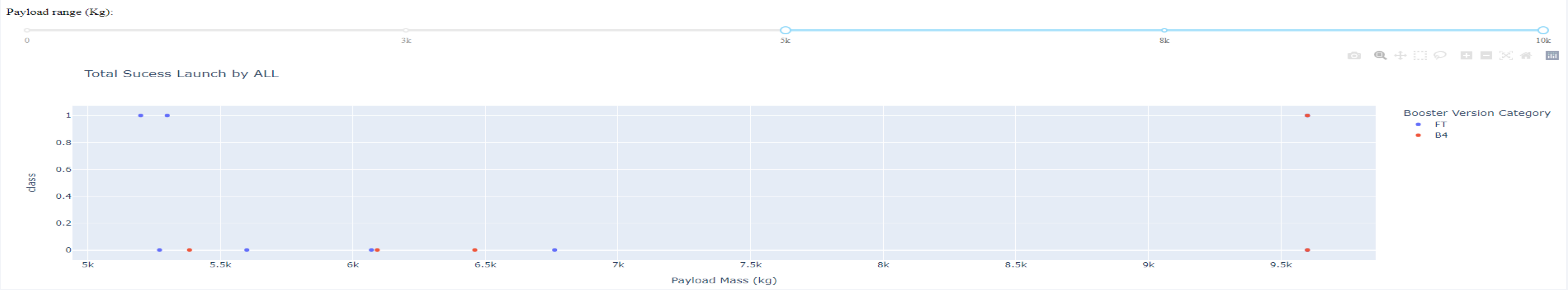
- Range slider is used to select values of different range.
- In the above scatter plot, we have used range slider to show the launch outcome of all sites for different payload mass.
- Screenshots shows the payload range for 0 to 3000 kg then 3000 to 5000 kg and last one shows 5000 to 10000 kg of payload mass values.
- From the scatter plot we can observe that higher the payload mass higher the success rate of a launch site.

Range Slider - Payload vs. Launch Outcome scatter plot for all sites

Payload Mass from 3000 kg to 5000 kg



Payload Mass from 5000 kg to 10,000 kg

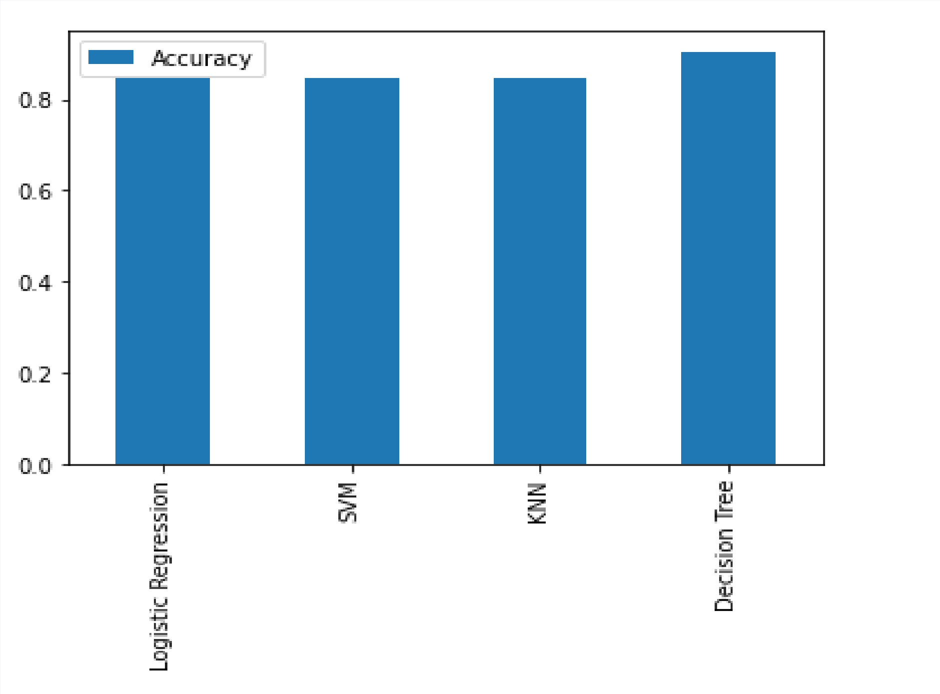


Section 5

Predictive Analysis (Classification)

Classification Accuracy

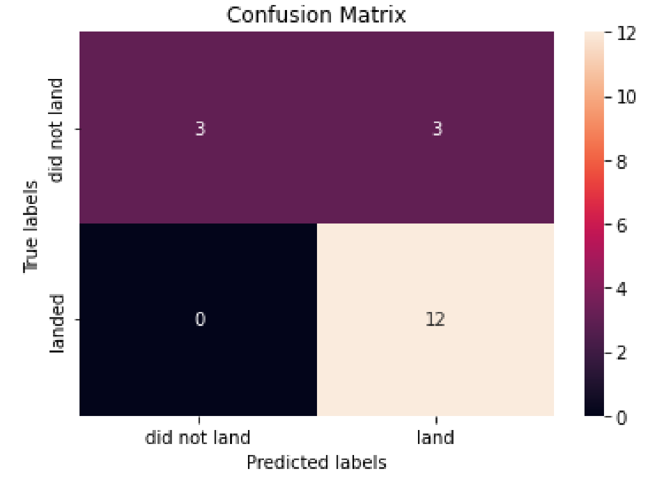
- The bar chart shows the built model accuracy for all the classification models (Logistic regression, SVM, KNN and Decision Tree).
- From the chart we can observe Decision tree bar is higher compared to other bars.
- The classification accuracy of each model are Logistic regression, SVM, KNN have 0.84 accuracy and Decision tree have 0.90 accuracy
- So we can conclude that 'DECISION TREE MODEL' have high accuracy value.



	Accuracy
Logistic Regression	0.846429
SVM	0.848214
KNN	0.848214
Decision Tree	0.903571

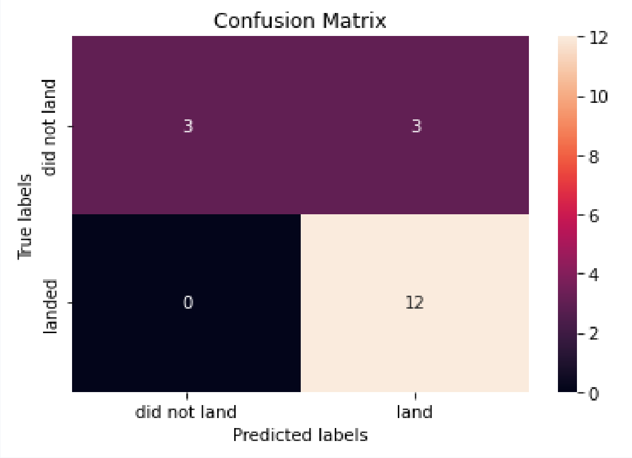
Confusion Matrix

Decision Tree

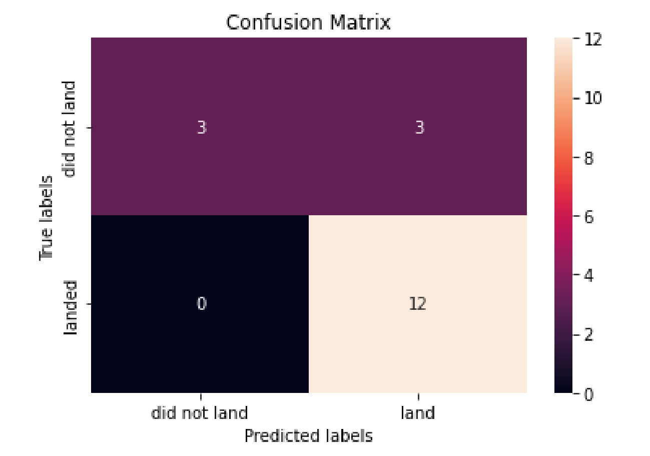


- From the confusion matrix we can see that DECISION TREE performs best compared to Logistic regression, SVM and KNN.
- Hence, DECISION TREE CLASSIFICATION MODEL is the BEST PERFORMING model.

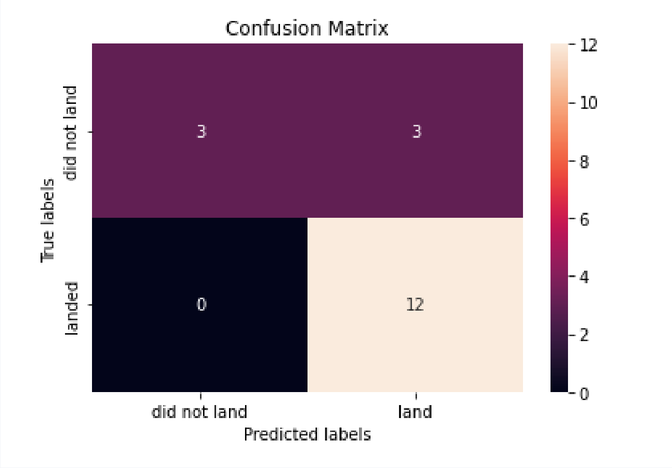
Logistic Regression



KNN



SVM



Conclusions

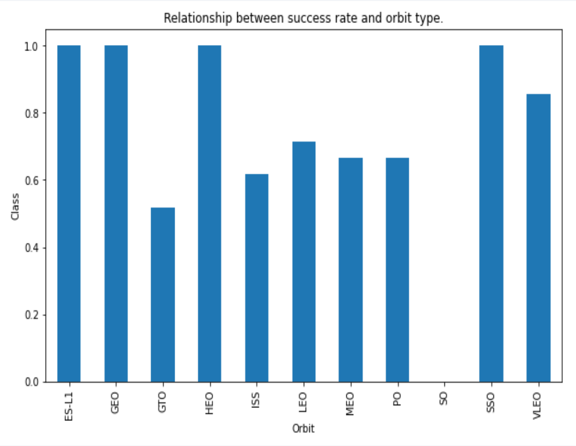
- Decision Tree classification model is the best model for the SPACEX dataset
- The Orbit type have significant impact on mission success rates. ES-L1 , GEO, HEO, SSO has the highest success rates.
- From 2013, The mission success rates have increased, with the help of advanced technologies and research in rocket science
- Payload Mass have significant impact on mission success rate.
- KSC LC-39A has the most successful launches, but increasing their payload mass have negative impact, that's a great point of improvement to get more successful mission delivering greater payloads

Appendix

Python code

```
spacex_dash_app.py X
spacex_dash_app.py > ...
# Create an app layout
app.layout = html.Div(children=[html.H1('SpaceX Launch Records Dashboard',
                                        style={'textAlign': 'center', 'color': '#503D36',
                                               'font-size': 40}),
                                # TASK 1: Add a dropdown list to enable Launch Site selection
                                # The default select value is for ALL sites
                                # dcc.Dropdown(id='site-dropdown',...)
                                dcc.Dropdown(id='site-dropdown',
                                             options=[  {'label': 'All Sites', 'value': 'ALL'},
                                                        {'label': 'CCAFS LC-40', 'value': 'CCAFS LC-40'},
                                                        {'label': 'CCAFS SLC-40', 'value': 'CCAFS SLC-40'},
                                                        {'label': 'KSC LC-39A', 'value': 'KSC LC-39A'},
                                                        {'label': 'VAFB SLC-4E', 'value': 'VAFB SLC-4E'}
                                                     ],
                                             value='ALL',
                                             placeholder="Select a Launch Site here",
```

Bar Chart



Notebook Output

```
# Apply value_counts() on col
df['LaunchSite'].value_counts
```

```
5]: CCAFS SLC 40    55
    KSC LC 39A      22
    VAFB SLC 4E     13
    Name: LaunchSite, dtype: int6
```

SQL Query

```
In [21]: %%sql
select MIN(DATE) as first_success_groundpad from SPACEXTBL where landing__outcome='Success (ground pad)'

 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[21]:

first_success_groundpad
2015-12-22

Dataset

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

```
In [4]: %%sql
select * from SPACEXTBL where launch_site like 'CCA%' limit 5

 * ibm_db_sa://pmj24486:***@2f3279a5-73d1-4859-88f0-a6c3e6b4b907.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30756/BLUDB
Done.
```

Out[4]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Thank you!