IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rohan M
14/06/2023

Outline

Executive Summary

- Summary of methodologies
    - Data Collection - API
    - Data Collection – Web Scraping
    - Data Wrangling
    - EDA with SQL
    - EDA with Visualization
    - Visual Analytics with Folium
    - Machine Learning
- Summary of all results
    - EDA Result
    - Interactive Analytics
    - Predictive Analytics

Introduction

- Project background and context
  - SpaceX is revolutionizing the space age by making the launching of rockets as affordable as possible. They have been able to reduce the cost of a rocket launch to almost 1/3rd compared to other providers. This has been possible because SpaceX has been reusing the first stage rockets. Thus ,predicting if the stage 1 rocket will land will help other providers compete against SpaceX. The goal is to accurately predict if the first stage will land successfully.
- Problems you want to find answers
  - Identifying factors that will determine if the first stage will land successfully
  - Dependency on different factors that determine the success rate of a successful landing
  - Ideal conditions needed to ensure a successful landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected using 2 methods – Using SpaceX API and Web Scraping
- Perform data wrangling
  - The data was first filtered, then the missing values were removed from consideration, certain features were transformed from categorical data to numerical data to enable machine learning models to run on this data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Different classification models were built to predict the outcome .The models were tuned using GridSearch CV to tune the hyperparameters

Data Collection - API

- Describe how data sets were collected.
  - The data is gathered using SpaceX REST API
  - Using a get request we get the launch data
  - View the result calling the .json() method
  - Conversion from JSON to dataframe can be done using the json_normalize function
  - Export to csv file
- You need to present your data collection process use key phrases and flowcharts

Data Collection – Web Scraping

- Describe how data sets were collected.
  - Scraping the data from the HTML tables of Wikipedia using Beautiful Soup package
  - Parse the data from the HTML tables and convert into a Pandas data frame
  - Export to csv
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Using the Get request, gathered and parsed the SpaceX launch data
- A data frame was created to include the data of only Falcon 9
- Finally, we conducted some data wrangling by replacing missing values with mean

- https://github.com/rhnapril/IBM-capstone/blob/main/Week%201%20-%20API.ipynb

Now let's start requesting rocket launch data from SpaceX API with the following URL:

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

Data Collection - Scraping

- HTTP GET method was used to request Falcon 9 data from the HTML table
- Extracted relevant columns from the HTML table
- Data Frame was created with the data from the earlier task
- https://github.com/rhnapril/IBM-capstone/blob/main/Week%201%20-%20Web%20Scraping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```python
# use requests.get() method with the provided static_url
# assign the response to a object

r = requests.get(static_url)

htmlcontent = r.content
```

Create a `BeautifulSoup` object from the HTML `response`

```python
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content

soup = BeautifulSoup(htmlcontent,'html.parser')
```

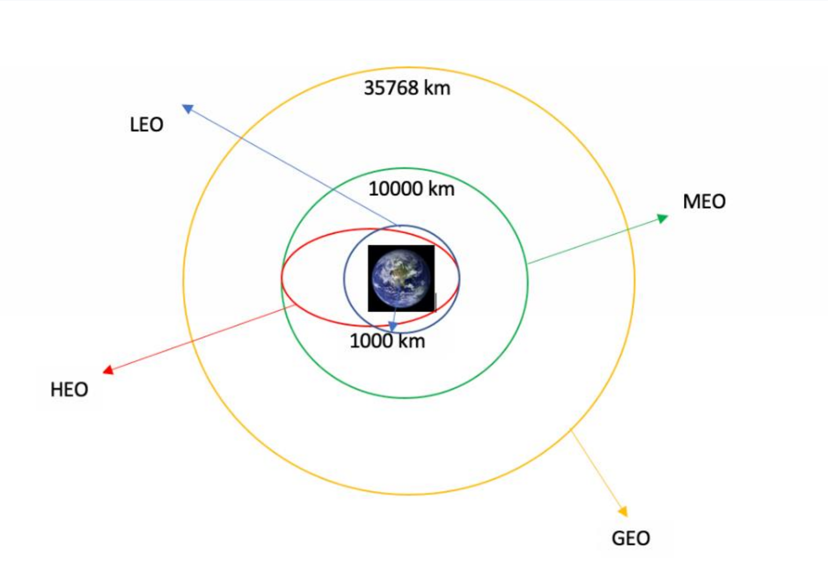
```python
column_names = []

extract_column_from_header = first_launch_table.find_all("th")

for i in extract_column_from_header:
    title = i.text
    column_names.append(title)

# Apply find_all() function with `th` element on first_launch_table
# Iterate each th element and apply the provided extract_column_from_header() to get a column name
# Append the Non-empty column name (`if name is not None and len(name) > 0`) into a list called column_names
```

Data Wrangling

- The number of launches from each site is calculated
- The number of occurrences of each orbit
- Calculated the number and occurrence of mission outcome per orbit type
- https://github.com/rhnapril/IBM-capstone/blob/main/Week%201%20-%20Data%20Wrangling.ipynb

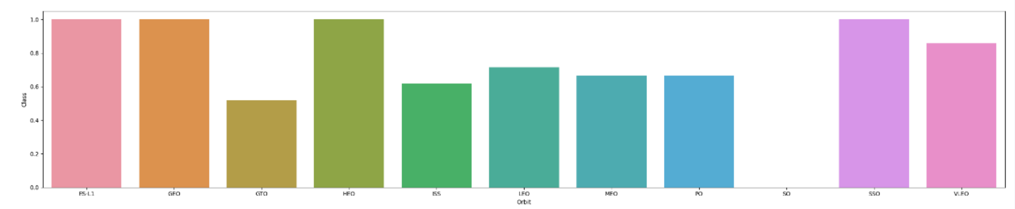
```python
# Apply value_counts on Orbit column
df.Orbit.value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
Name: Orbit, dtype: int64
```

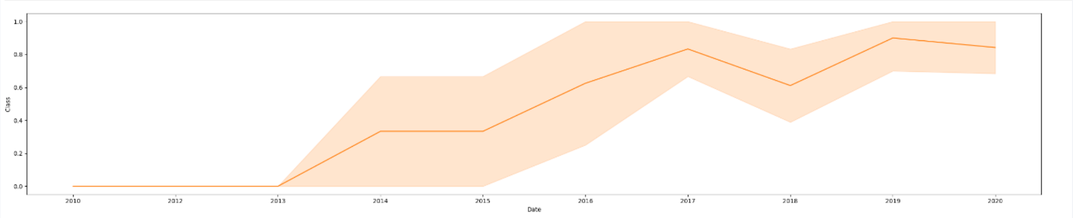


EDA with Data Visualization

- The success of the lunches improved drastically after 2013
- Find the relationship between flight number and launch Site, payload and launch site, , success rate of each orbit type
- https://github.com/rhnapril/IBM-capstone/blob/main/Week%202%20-%20EDA-%20Visualization.ipynb



Analyze the ploted bar chart try to find which orbits have high sucess rate.



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

EDA with SQL

- We applied EDA with SQL to get insight from the data. We wrote queries to find out for instance:
  - The names of unique launch sites
  - 5 records where the launch site begins with ‘CCA’
  - Total payload carried by NASA (CRS)
  - Total number of successful and failure mission outcomes
  - Names of the booster_version that carried maximum payload mass

- https://github.com/rhnapril/IBM-capstone/blob/main/Week%202%20-%20EDA%20-%20SQL.ipynb

Build an Interactive Map with Folium

- Identified all the sites used by SpaceX for launching the sites
- Identified the successful and failed launches from the launch sites
- Calculated the distances between the launch sites and proximities
- https://github.com/rhnapril/IBM-capstone/blob/main/Week%203%20-%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Plotted pie charts showing the total launches by a certain sites
- Plotted scatter graph showing the relationship with Outcome and Payload Mass (Kg) for the different booster version
- https://github.com/rhnapril/IBM-capstone/blob/main/Dash%20App.py

Predictive Analysis (Classification)

- Used different classification models like
  - Decision tree
  - KNN
  - SVM
  - Logistic Regression
- Built different machine learning models and tune different hyperparameters using GridSearchCV
- Calculated the accuracy scores for all the classification models
- https://github.com/rhnapril/IBM-capstone/blob/main/Week%204%20-%20Machine%20Learning.ipynb

Results

- EDA
  - Success of the launches have gone up since 2013
  - KSC LC -39 has the highest success rate
  - Orbits ES-L1, GEO, HEO and SSO have a cent per cent success rate
- Visual Analytics
  - All sites are located in the US
- Predicitive Analytics
  - Decision Tree is the most reliable classification model for the dataset

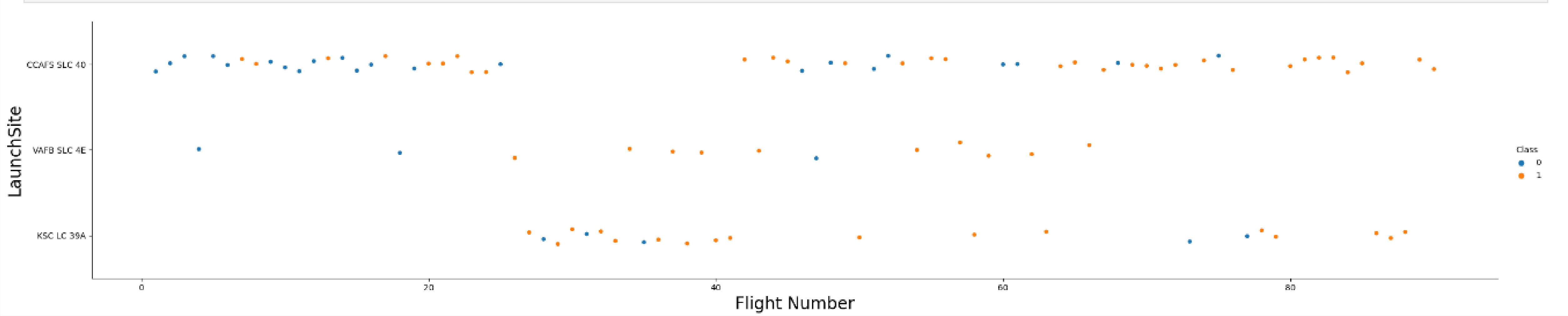
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

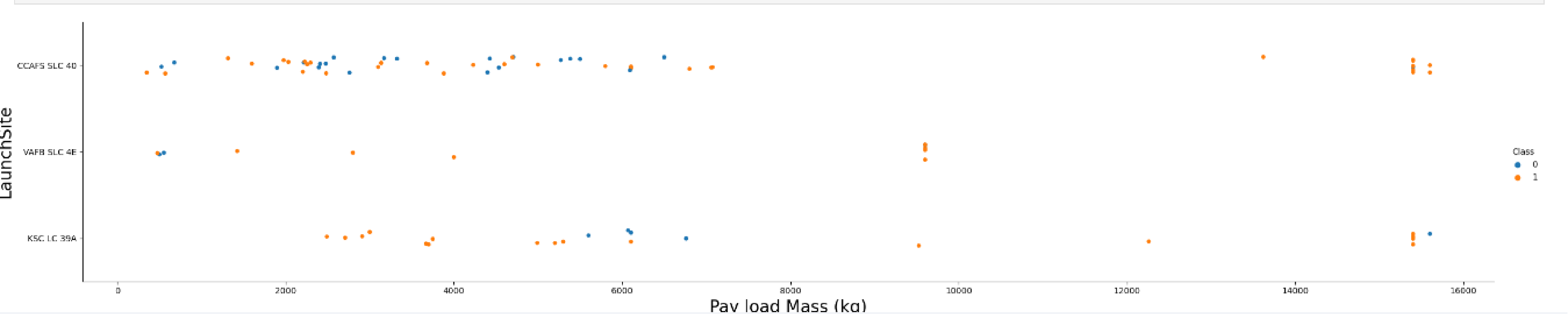
VAFB SLC 4E and KSC LC 39A have higher success rates

New launches have a higher success rate



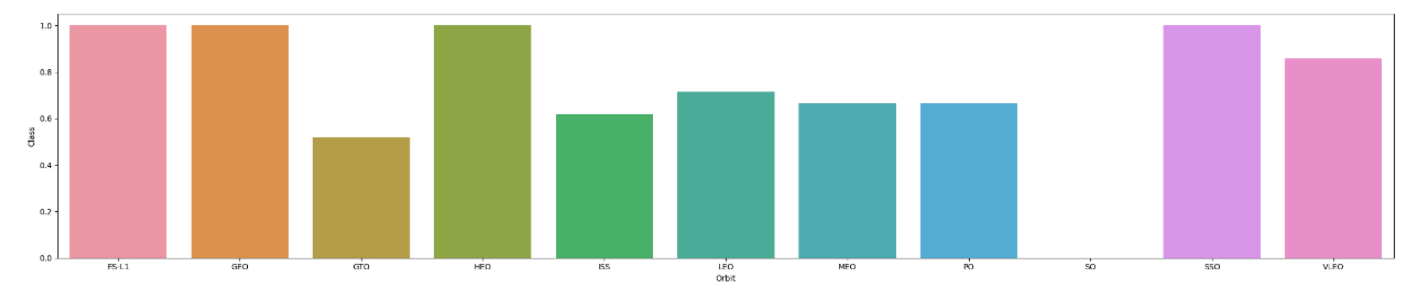
Payload vs. Launch Site

- KSC LC 39A has a 100% success rate for launches less than 5,500 kg
- Launces with a payload greater than 7,000 kg were successful



Success Rate vs. Orbit Type

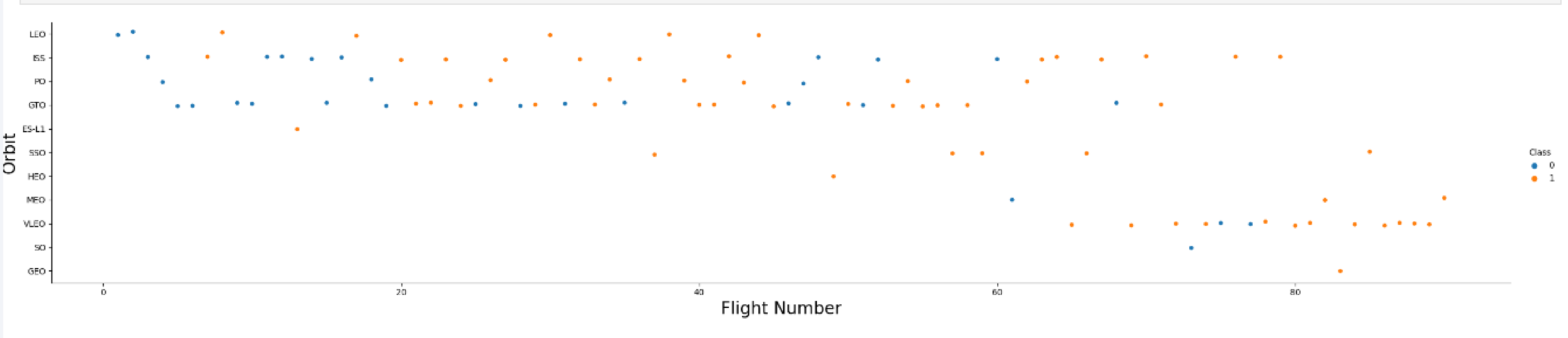
- Cent per cent success rate for ES-L1, GEO, HEO and SSO



Analyze the ploted bar chart try to find which orbits have high sucess rate.

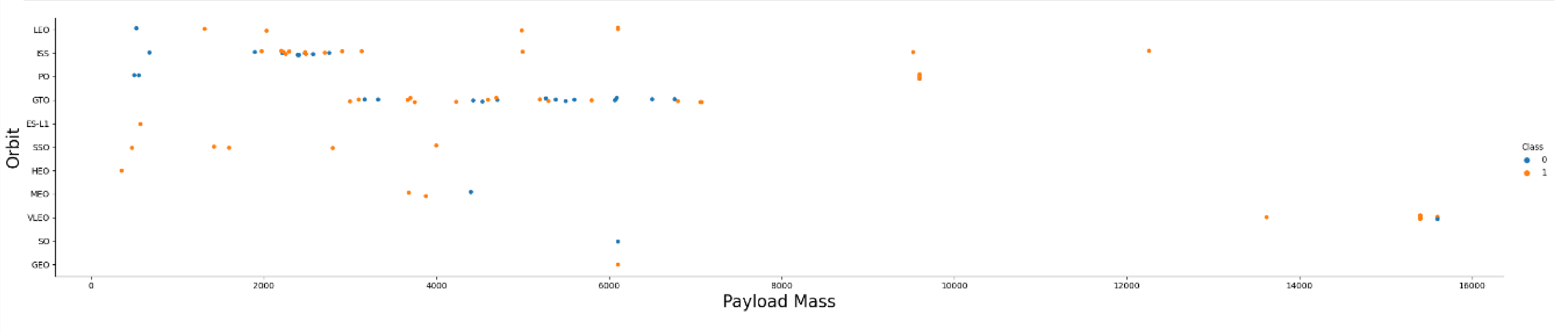
Flight Number vs. Orbit Type

- The plot below shows the Flight Number vs. Orbit type. We observe that in the LEO orbit, success is related to the number of flights whereas in the GTO orbit, there is no relationship between flight number and the orbit.



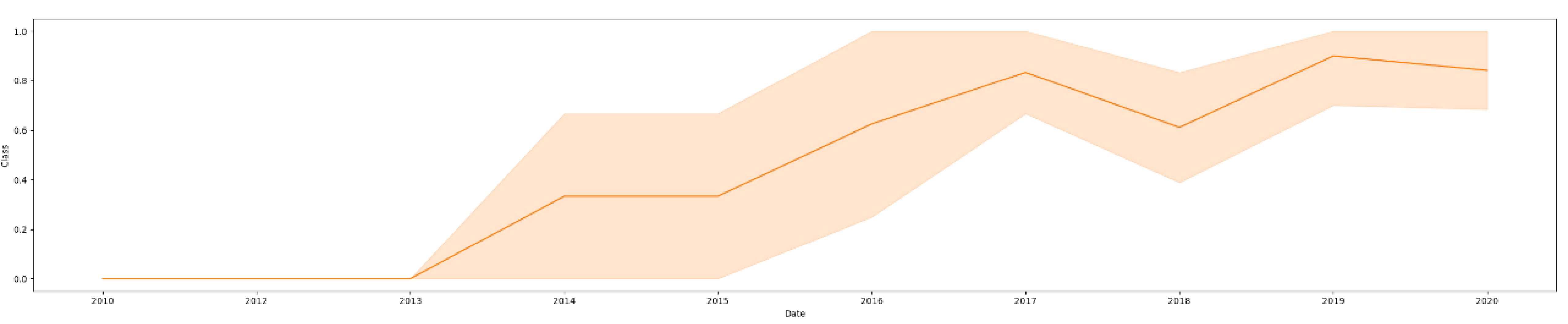
Payload vs. Orbit Type

- We can observe that with heavy payloads, the successful landing are more for PO, LEO and ISS orbits.



Launch Success Yearly Trend

- From the plot, we can observe that success rate since 2013 kept on increasing till 2020.



All Launch Site Names

We used the key word **DISTINCT** to show only unique launch sites from the SpaceX data.

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

```
%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

Launch Site Names Begin with 'CCA'

- We used the query above to display 5 records where launch sites begin with \`CCA\`

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

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

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

ate	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
010	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)
012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

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

```
%sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer like 'NASA (CRS)'
```

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

SUM(PAYLOAD_MASS__KG_)
45596.0

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like 'F9 v1.1'
```

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

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- We observed that the dates of the first successful landing outcome on ground pad was 22nd December 2015

[21]:

Date
22/12/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select * from SPACEXTBL where Landing_Outcome like 'Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
6/2016	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696.0	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
8/2016	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600.0	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
3/2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300.0	GTO	SES	Success	Success (drone ship)
1/2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200.0	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	count(Mission_Outcome)
Success	100

Mission_Outcome	count(Mission_Outcome)
Failure (in flight)	1

- We used wildcard like '%' to filter for **WHERE** MissionOutcome was a success or a failure.

Boosters Carried Maximum Payload

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

```
%sql select 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

```
%sql select substr(Date, 4, 2) as month, Booster_Version,Launch_Site from SPACEXTBL where Landing_Outcome like "Failure (drone ship)" and substr(Date,7,4)='2015'
```

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

month	Booster_Version	Launch_Site
10	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
22/12/2015	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034.0	LEO	Orbcomm	Success	Success (ground pad)
04/08/2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136.0	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
05/06/2016	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696.0	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
27/05/2016	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100.0	GTO	Thaicom	Success	Success (drone ship)
18/07/2016	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14/08/2016	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600.0	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14/01/2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600.0	Polar LEO	Iridium Communications	Success	Success (drone ship)
19/02/2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
30/03/2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300.0	GTO	SES	Success	Success (drone ship)
05/01/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300.0	LEO	NRO	Success	Success (ground pad)
06/03/2017	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
23/06/2017	19:10:00	F9 FT B1029.2	KSC LC-39A	BulgariaSat-1	3669.0	GTO	Bulsatcom	Success	Success (drone ship)
25/06/2017	20:25:00	F9 FT B1036.1	VAFB SLC-4E	Iridium NEXT 2	9600.0	LEO	Iridium Communications	Success	Success (drone ship)
14/08/2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
24/08/2017	18:51:00	F9 FT B1038.1	VAFB SLC-4E	Formosat-5	475.0	SSO	NSPO	Success	Success (drone ship)
09/07/2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OTV-5	4990.0	LEO	U.S. Air Force	Success	Success (ground pad)
10/09/2017	12:37:00	F9 B4 B1041.1	VAFB SLC-4E	Iridium NEXT 3	9600.0	Polar LEO	Iridium Communications	Success	Success (drone ship)
10/11/2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200.0	GTO	SES EchoStar	Success	Success (drone ship)
30/10/2017	19:34:00	F9 B4 B1042.1	KSC LC-39A	Koreasat 5A	3500.0	GTO	KT Corporation	Success	Success (drone ship)
15/12/2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205.0	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
01/08/2018	1:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000.0	LEO	Northrop Grumman	Success (payload status unclear)	Success (ground pad)
18/04/2018	22:51:00	F9 B4 B1045.1	CCAFS SLC-40	Transiting Exoplanet Survey Satellite (TESS)	362.0	HEO	NASA (LSP)	Success	Success (drone ship)
05/11/2018	20:14:00	F9 B5 B1046.1	KSC LC-39A	Bangabandhu-1	3600.0	GTO	Thales-Alenia/BTRC	Success	Success (drone ship)
22/07/2018	5:50:00	F9 B5B1047.1	CCAFS SLC-40	Telstar 19V	7075.0	GTO	Telesat	Success	Success
25/07/2018	11:39:00	F9 B5B1048.1	VAFB SLC-4E	Iridium NEXT-7	9600.0	Polar LEO	Iridium Communications	Success	Success
08/07/2018	5:18:00	F9 B5 B1046.2	CCAFS SLC-40	Merah Putih	5800.0	GTO	Telkom Indonesia	Success	Success
09/10/2018	4:45:00	F9 B5B1049.1	CCAFS SLC-40	Telstar 18V / Apstar-5C	7060.0	GTO	Telesat	Success	Success
10/08/2018	2:22:00	F9 B5 B1048.2	VAFB SLC-4E	SAOCOM 1A	3000.0	SSO	CONAE	Success	Success
15/11/2018	20:46:00	F9 B5 B1047.2	KSC LC-39A	Es hail 2	5300.0	GTO	Es hailSat	Success	Success
12/03/2018	18:34:05	F9 B5 B1046.3	VAFB SLC-4E	SSO-A	4000.0	SSO	Spaceflight Industries	Success	Success
01/11/2019	15:31:00	F9 B5 B1049.2	VAFB SLC-4E	Iridium NEXT-8	9600.0	Polar LEO	Iridium Communications	Success	Success
22/02/2019	1:45:00	F9 B5 B1048.3	CCAFS SLC-40	Nusantara Satu, Beresheet Moon lander, S5	4850.0	GTO	PSN, SpaceIL / IAI	Success	Success
03/02/2019	7:49:00	F9 B5B1051.1	KSC LC-39A	Crew Dragon Demo-1, SpaceX CRS-17	12055.0	LEO (ISS)	NASA (CCD)	Success	Success
05/04/2019	6:48:00	F9 B5B1056.1	CCAFS SLC-40	SpaceX CRS-17, Starlink v0.9	2495.0	LEO (ISS)	NASA (CRS)	Success	Success

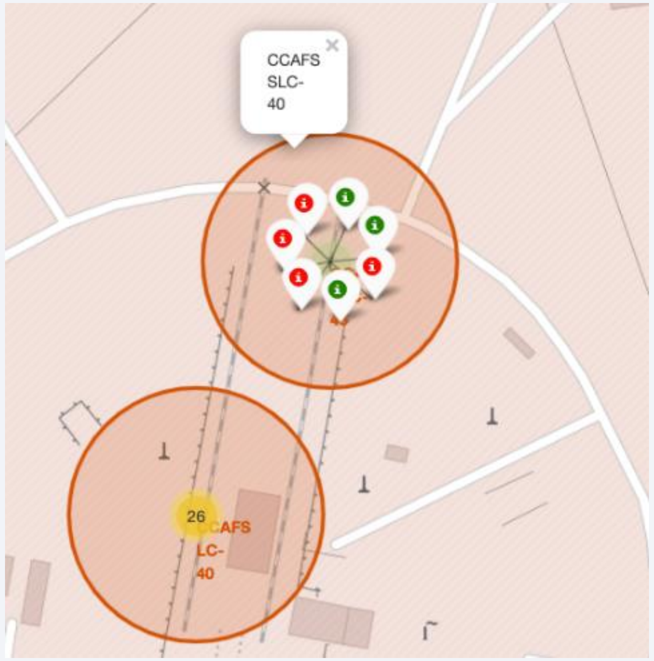
Section 3

Launch Sites Proximities Analysis

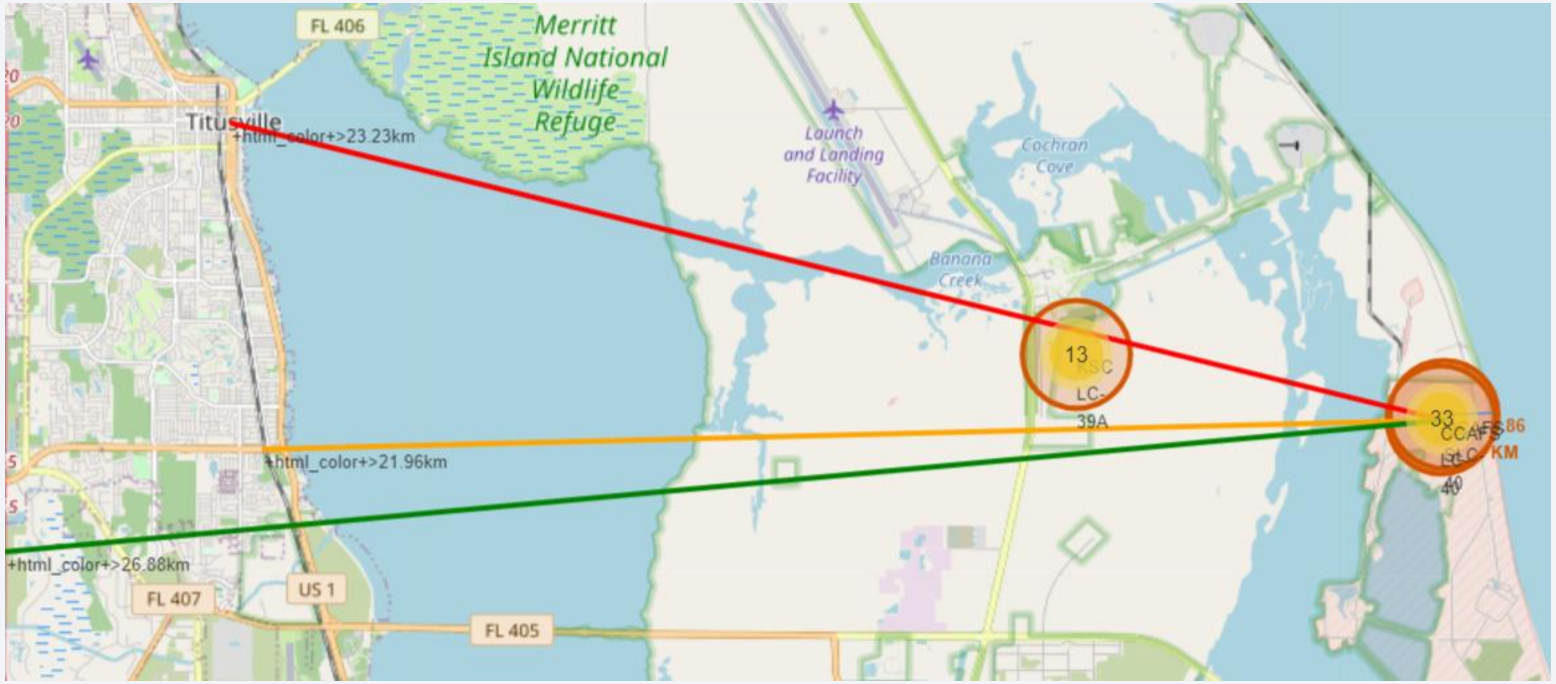
All launch sites

- All the launch sites are located in the east or the west coast of USA

Markers showing launch sites with color labels



Launch Site distance to landmarks

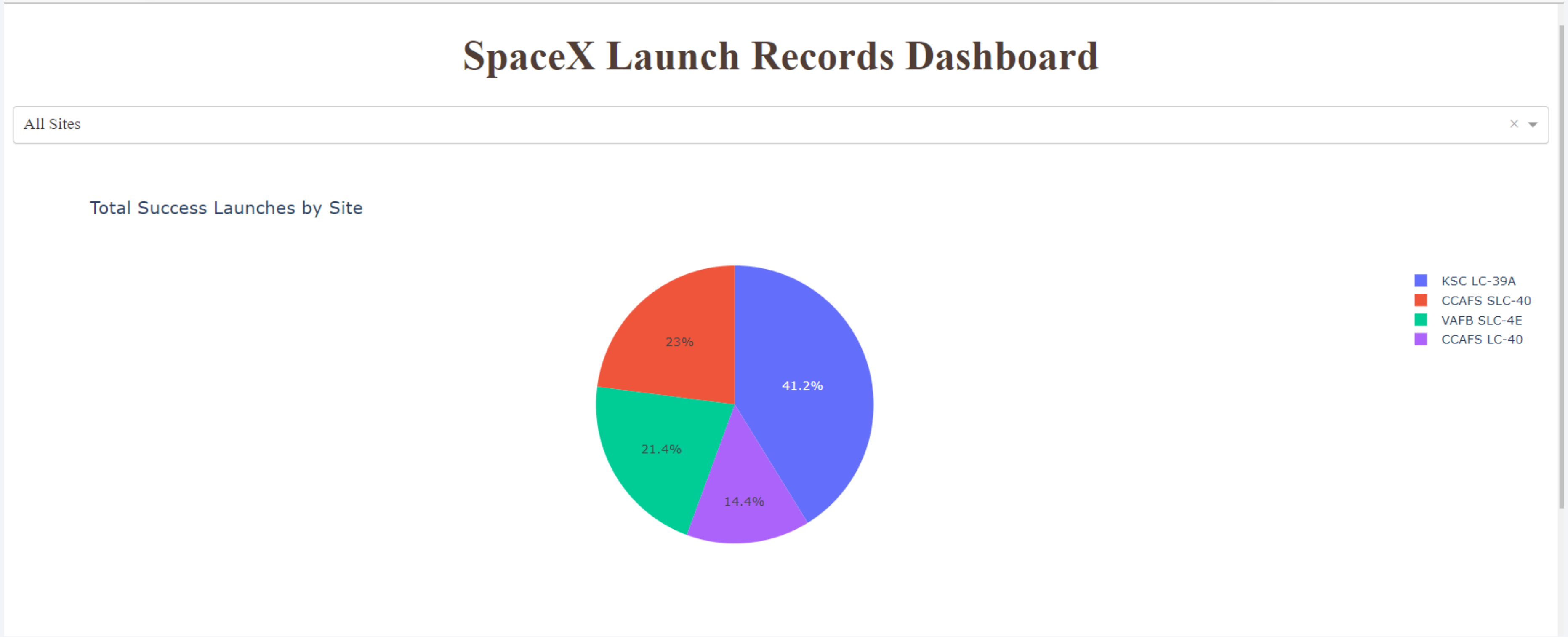


Section 4

Build a Dashboard with Plotly Dash

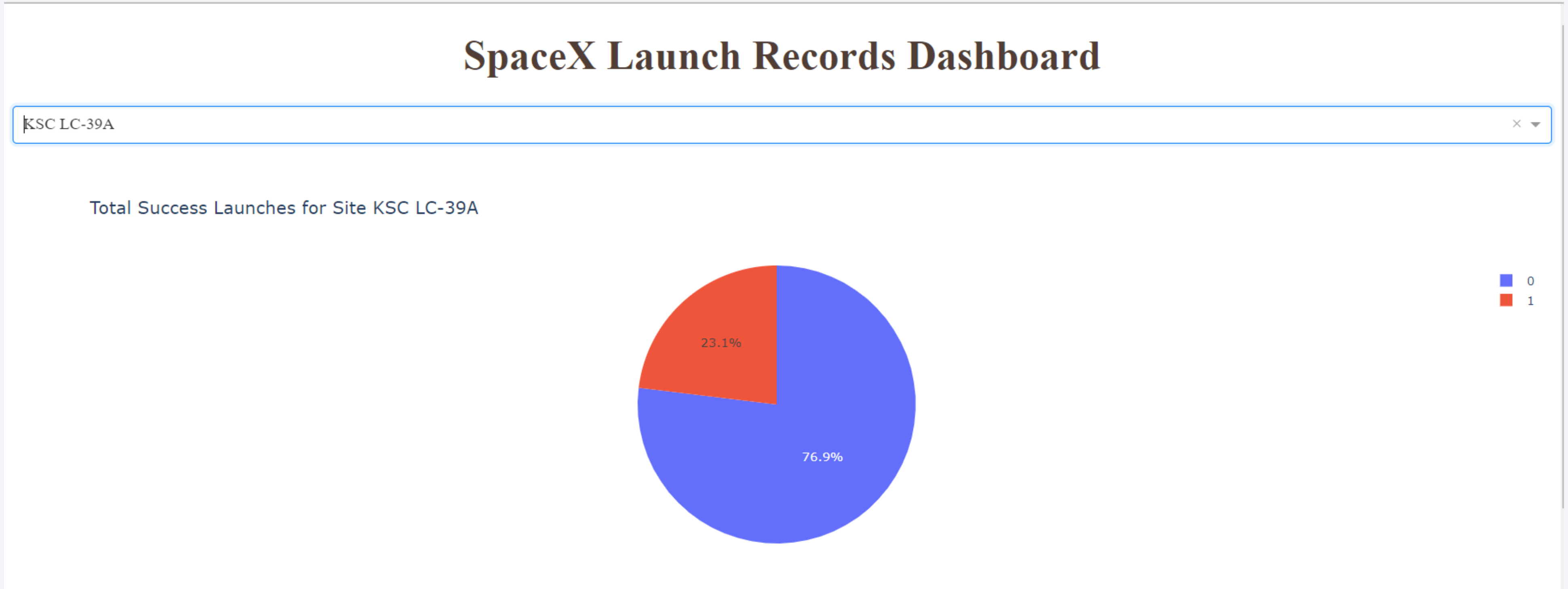
Pie chart showing the success percentage achieved by each launch site

- It is clear from the below pie char that the most successful launch site is KSC LC 39A



Pie chart showing the Launch site with the highest launch success ratio

- KSC LC 39A achieved a success rate of 76.9 %



Scatter plot of Payload vs Launch Outcome for all sites, with different payload selected in the range slider

Section 5

Predictive Analysis (Classification)

Classification Accuracy

```
print("tuned hpyerparameters :(best parameters) ",tree_cv.best_params_)
print("accuracy :",tree_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'random'}
accuracy : 0.8910714285714285
```

```
print("tuned hpyerparameters :(best parameters) ",knn_cv.best_params_)
print("accuracy :",knn_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
```

```
print("tuned hpyerparameters :(best parameters) ",svm_cv.best_params_)
print("accuracy :",svm_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```

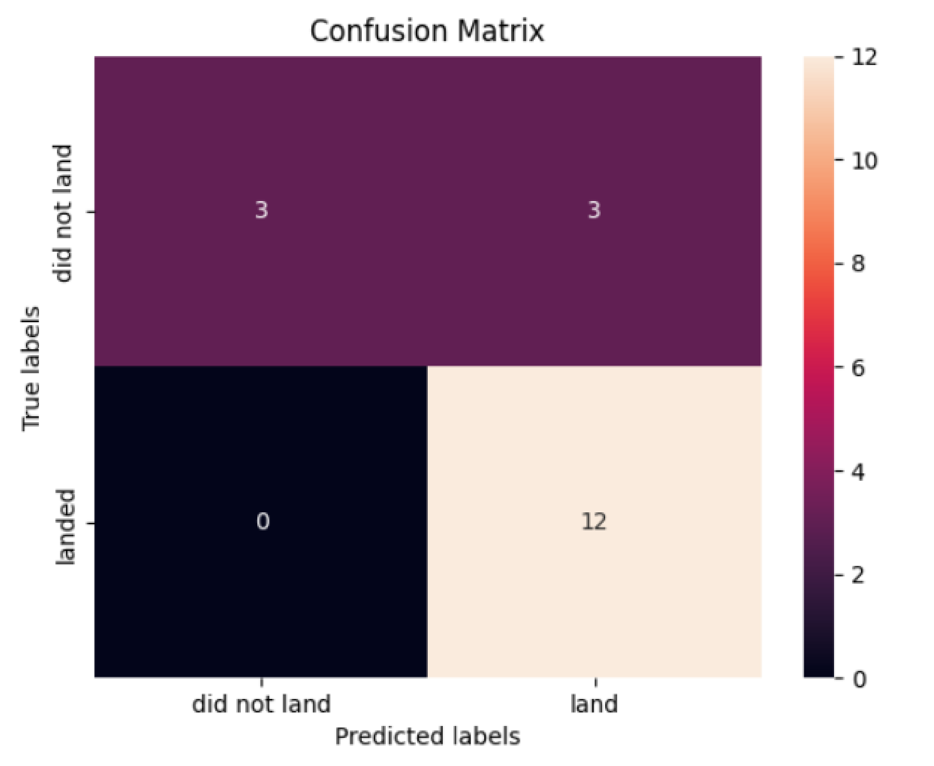
```
print("tuned hpyerparameters :(best parameters) ",logreg_cv.best_params_)
print("accuracy :",logreg_cv.best_score_)
```

```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

Confusion Matrix

- The Decision Tree Classifier is the most accurate prediction model

```
yhat_tree = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat_tree)
```



Conclusions

- Launch success rate has been on the increase since 2013
- KSC LC-39A had the most successful launches of any sites
- Decision Tree classifier is the most accurate prediction model

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!