IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rasikha Swaminathan
21-Feb-2022

Outline

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

Executive Summary

Summary of methodologies

- Data collection
- Data wrangling
- EDA with data visualization
- EDA with SQL
- Building an interactive map with Folium
- Building a Dashboard with Plotly Dash
- Predictive analysis (Classification)

Summary of all results

- EDA results
- Interactive analytics
- Predictive analysis

Introduction

- Project background and context

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. 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 for a rocket launch.

- Problems you want to find answers

- Will the first stage of Falcon 9 land successfully or not ?

- Which factor influences the successful landing of the Rocket?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is collected through the SpaceX REST API and Webscraping from related Wiki pages
- Performed data wrangling
  - One Hot Encoding is performed and the irrelevant columns are removed
- Performed exploratory data analysis (EDA) using visualization and SQL
  - Dataset is first read into the Pandas dataframe and the summary of the data is observed
  - Scatter plots and Bar graphs are used to visualize the relationship between the columns
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models

Data Collection

- SpaceX launch data is gathered using the REST API
- This gives the information about the rocket launches , payload delivered, rocket used ,launch specifications, landing specifications and launch outcome
- Our task is to use this data and predict for the successful launch
- Data is collected from this URL https://api.spacexdata.com
- Launch data is also obtained by webscraping Wiki pages with the beautiful soup objects

Data Collection – SpaceX API

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

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

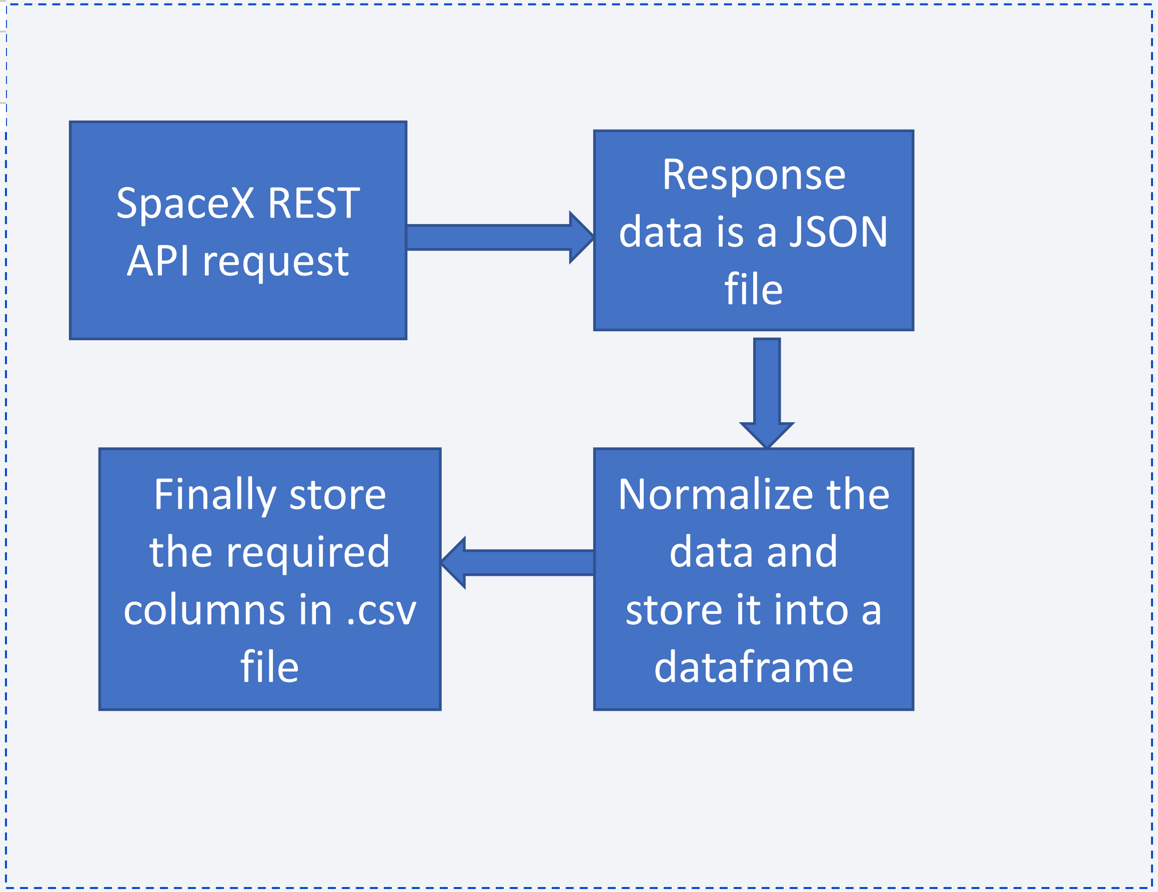
```python
data= pd.DataFrame(launch_dict)
```

Show the summary of the dataframe

```python
# Show the head of the dataframe
data.describe()
```

```python
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9 = data[data.BoosterVersion == 'Falcon 9']
data_falcon9
```

```python
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```



GitHub URL:
https://github.com/RasikhaV/testrepo/blob/master/Webscraping.ipynb

Data Collection - Scraping

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

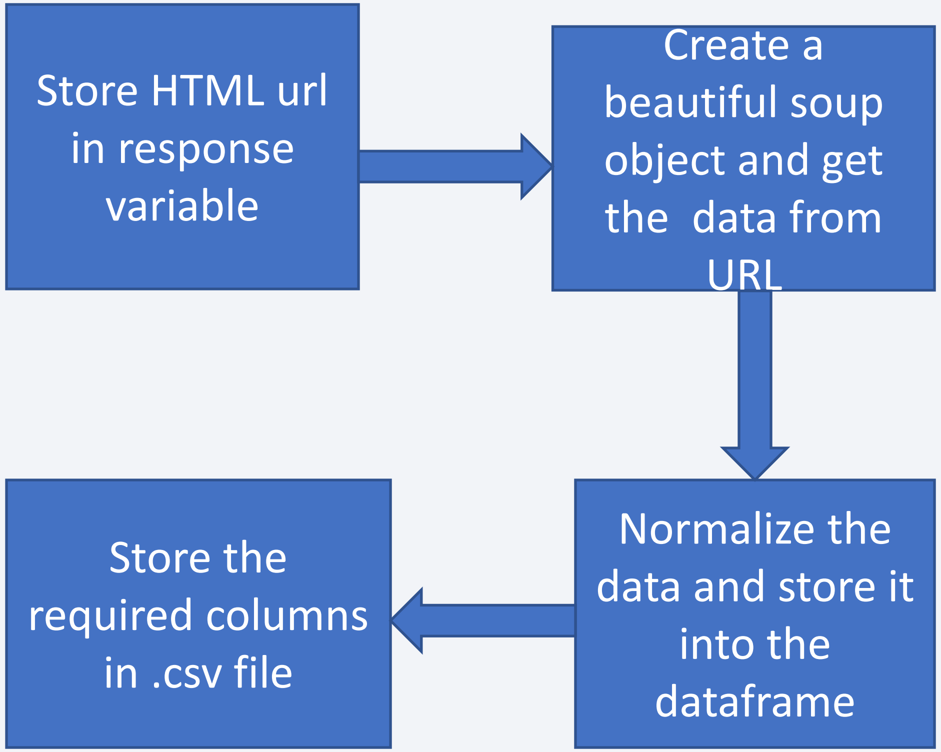
```
# use requests.get() method with the provided static_url
# assign the response to a object
response = requests.get(static_url).text
```

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

```
# Use BeautifulSoup() to create a BeautifulSoup object f
soup = BeautifulSoup(response , 'html.parser')
```

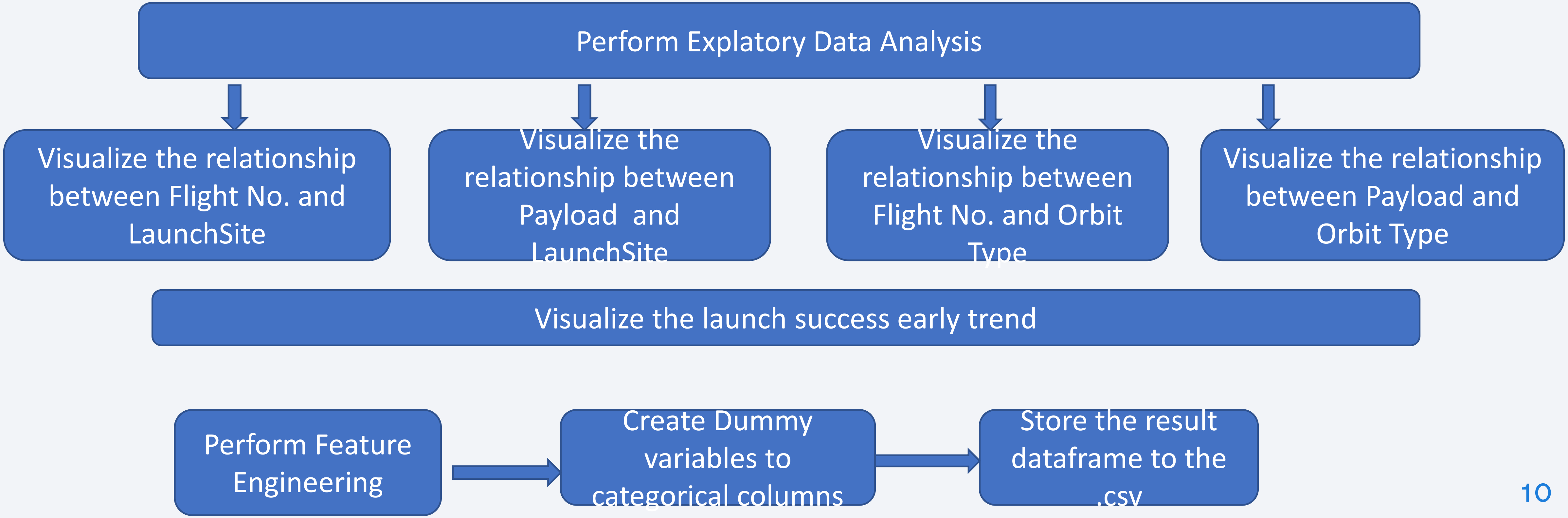
```
headings = []
for key,values in dict(launch_dict).items():
    if key not in headings:
        headings.append(key)
    if values is None:
        del launch_dict[key]

def pad_dict_list(dict_list, padel):
    lmax = 0
    for lname in dict_list.keys():
        lmax = max(lmax, len(dict_list[lname]))
    for lname in dict_list.keys():
        ll = len(dict_list[lname])
        if  ll < lmax:
            dict_list[lname] += [padel] * (lmax - ll)
    return dict_list

pad_dict_list(launch_dict,0)
df = pd.DataFrame.from_dict(launch_dict)
df.head()
```

GitHub URL :
https://github.com/RasikhaV/testrepo/blob/master/Webscraping1.ipynb



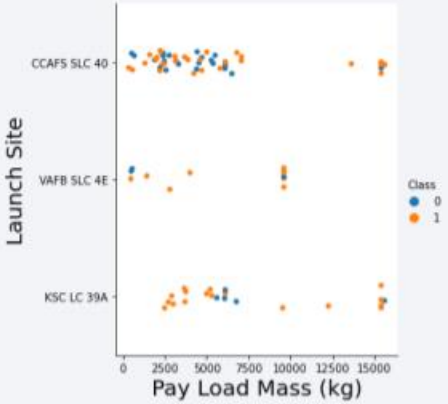
Data Wrangling

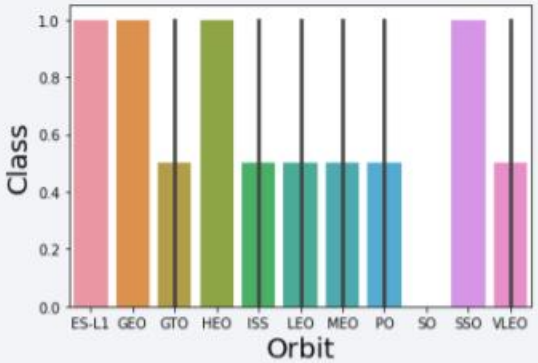
- Data wrangling is done by performing one hot encoding and removing the irrelevant columns

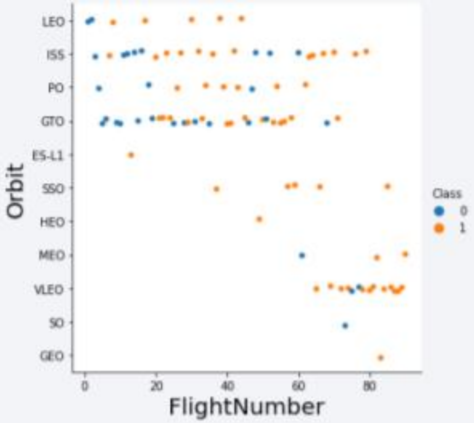


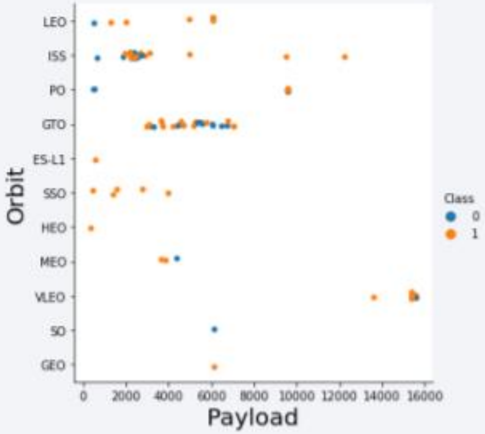
GitHub Url :

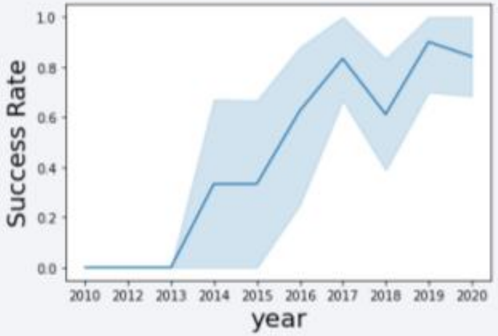
EDA with Data Visualization











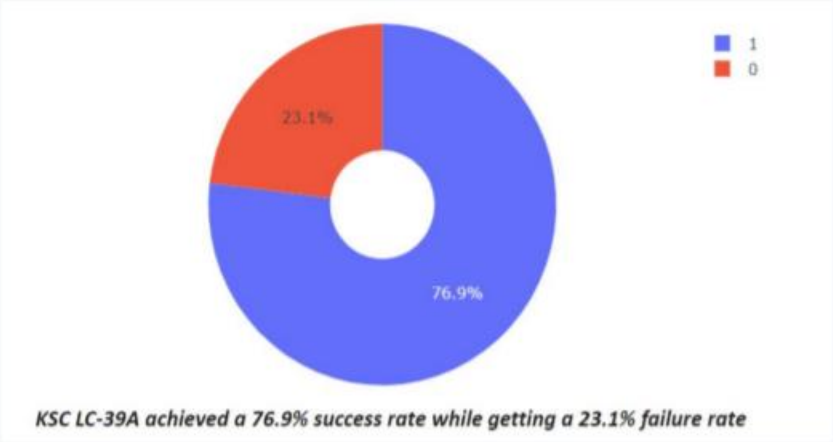
EDA with SQL

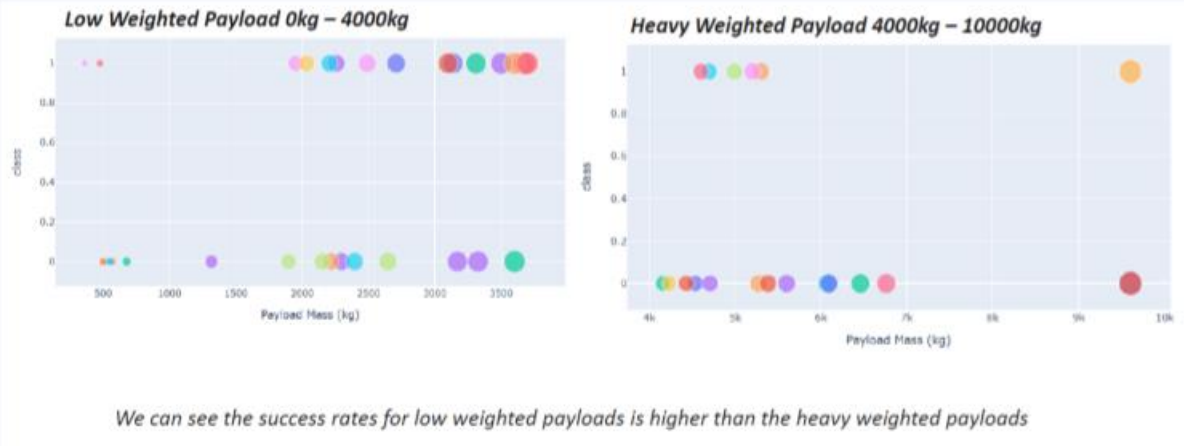
- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string ‘CCA’
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date where the successful landing outcome in drone ship was achieved.
- Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass.
- Listing the records which will display every column where landing outcomes is success for the months in year 2015
- Ranking the count of successful landing outcomes between the date 2010 06 04 and 2017 03 20 in descending order

Build an Interactive Map with Folium

GitHub Url :
https://github.com/RasikhaV/testrepo/blob/master/Interactive%20visual%20analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash





Predictive Analysis (Classification)

- **Developing a Model :**
  - Load our dataset into a pandas dataframe and transform it
  - Split our data into testing and training sets
  - Use a specific Machine learning Algorithm
  - Set our parameters and algorithm type to GridSearchCV object and train our datasets

- **Model Evaluation:**
  - Check for the accuracy in each model
  - Get the hyperparameters tuned
  - Plot confusion Matrix
- **Improving the Model:**
  - Perform feature engineering
  - Algorithm tuning

- **Finding best performing Model:**

  - Check for the higher accuracy obtained and that model would be the best performing model

  GitHub Url :
  - https://github.com/RasikhaV/testrepo/blob/master/Machine%20Learning%20Prediction.ipynb

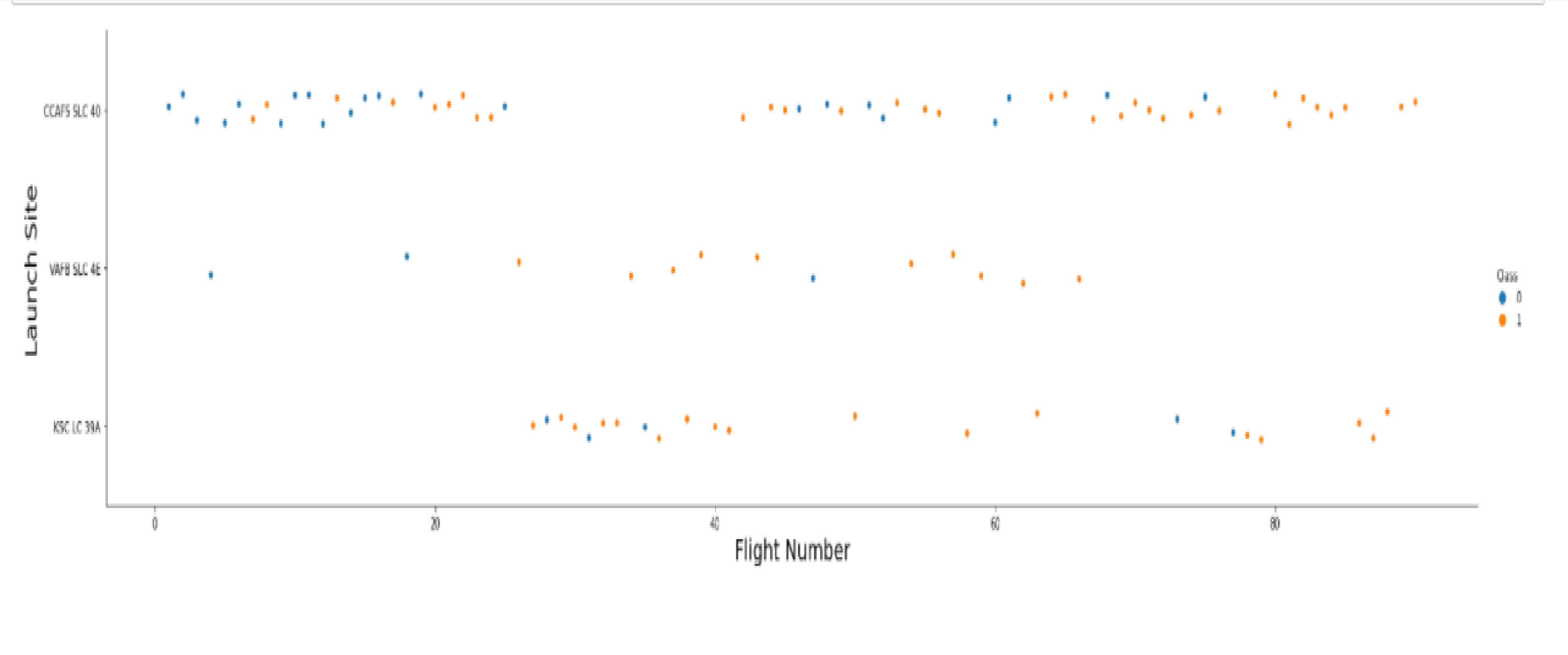
Results

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Section 2

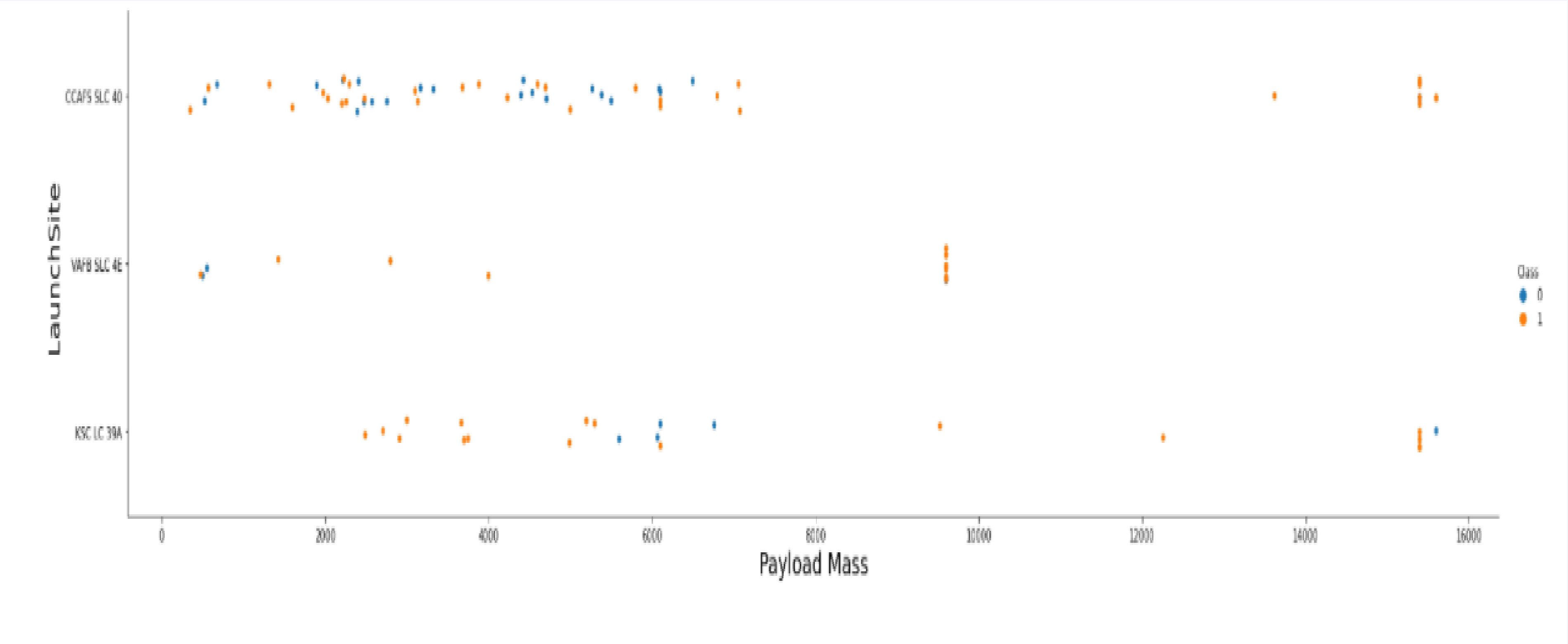
Insights drawn from EDA

Flight Number vs. Launch Site



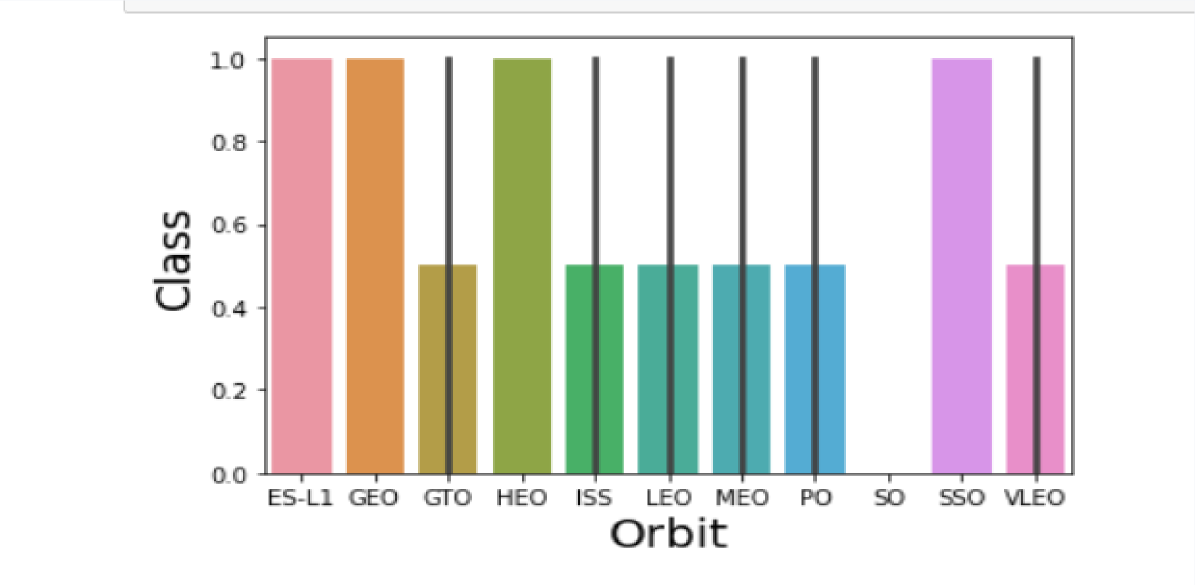
As the number of flight increases, the success rate at the launch site also increases

Payload vs. Launch Site



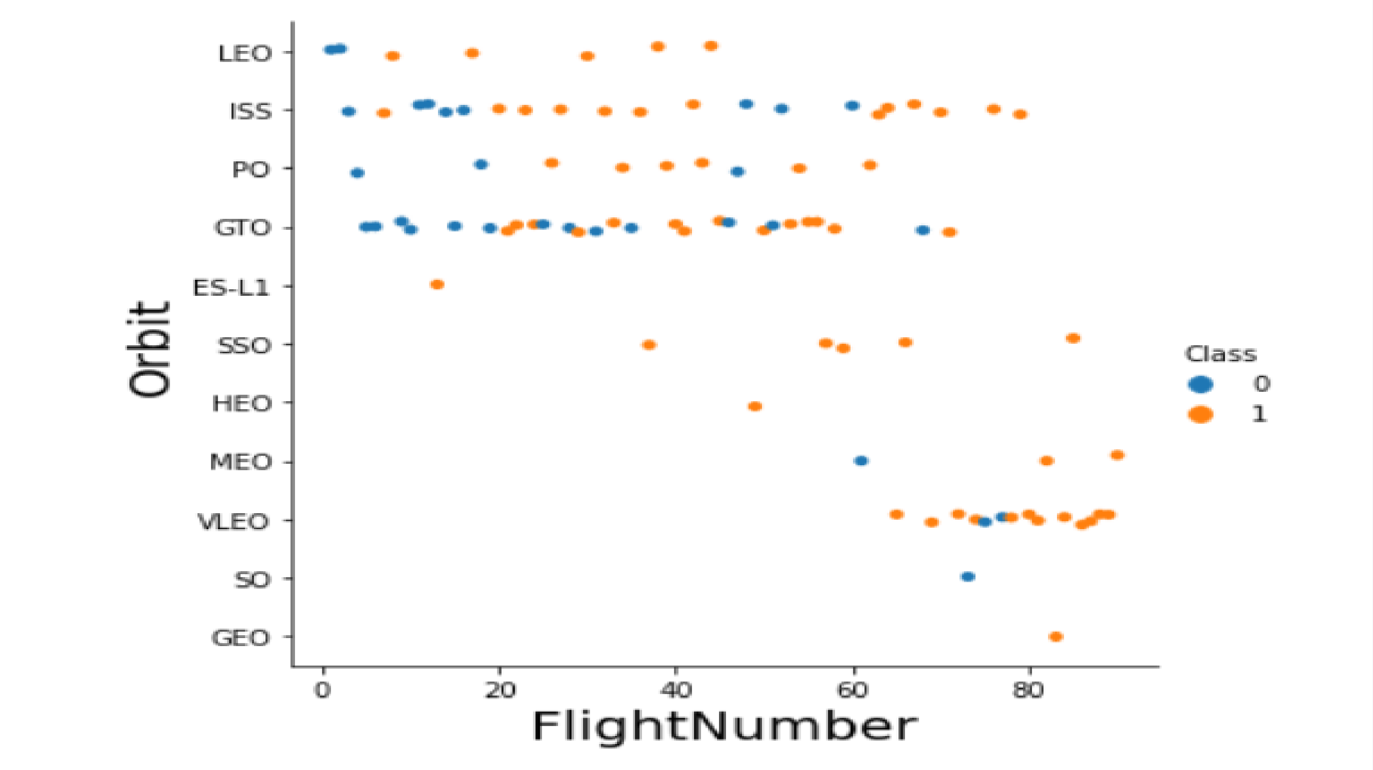
We can find that ,at the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



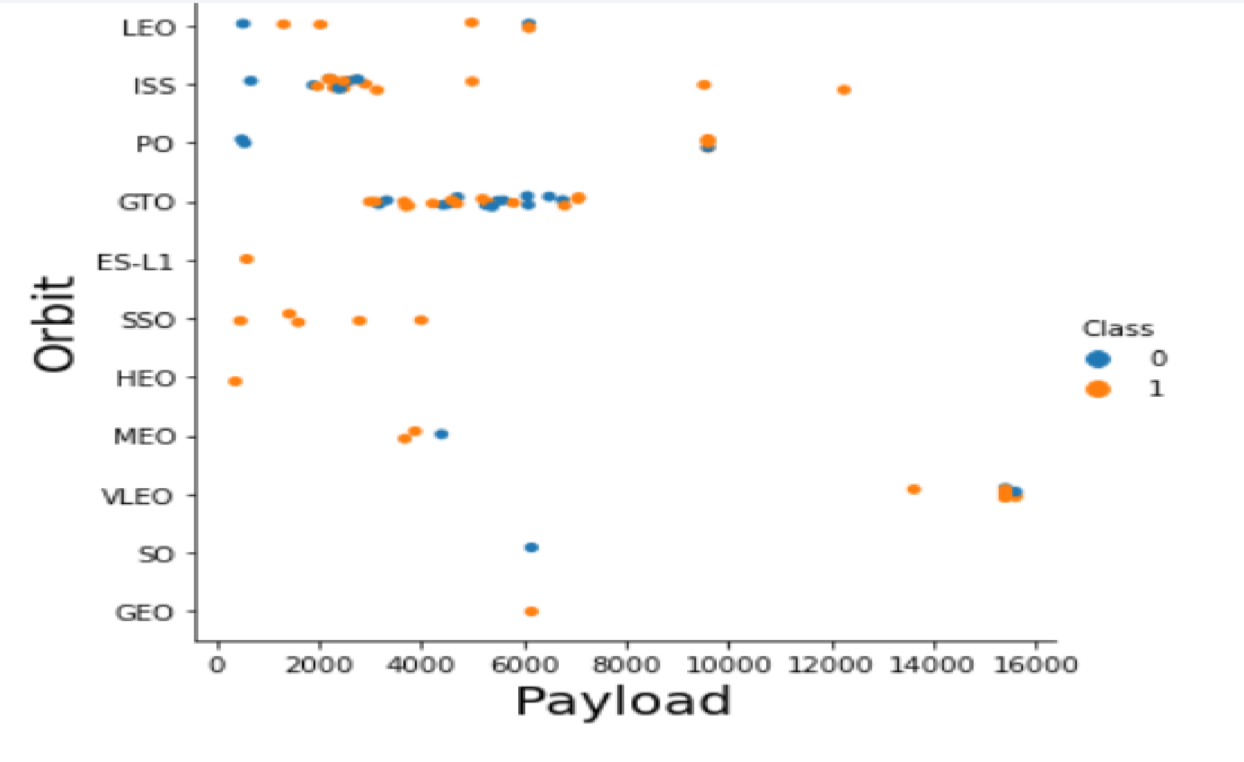
In the above bar plot , the success rate of the orbit types GEO, ES-L1 , SSO are at its best

Flight Number vs. Orbit Type



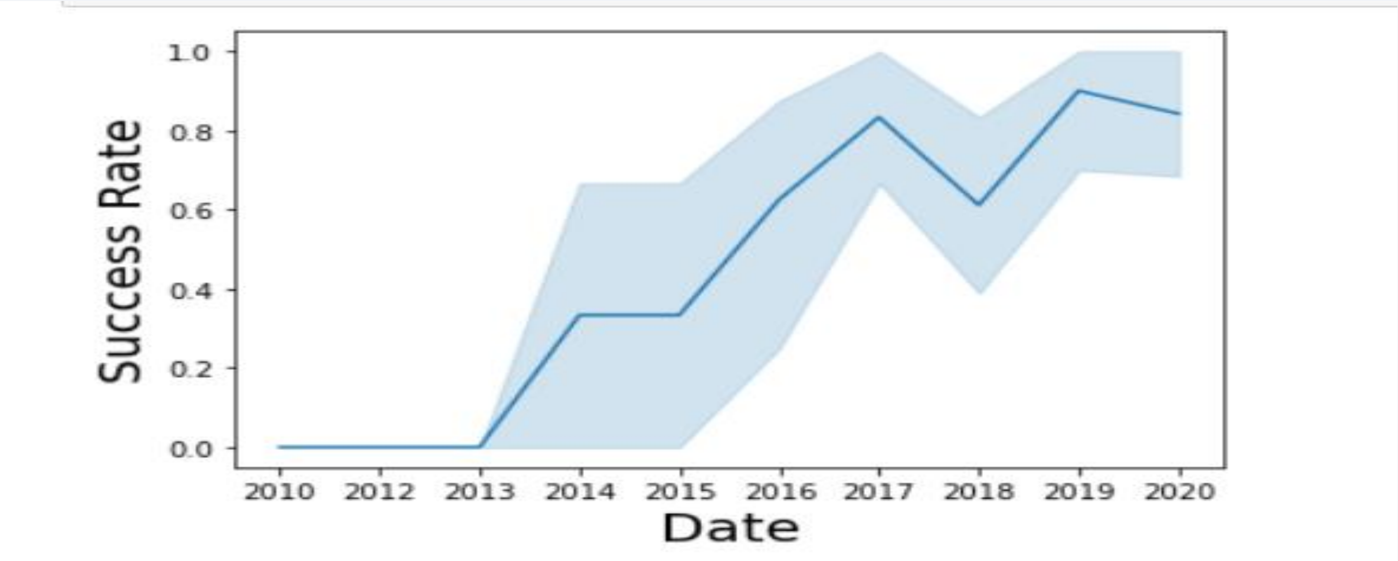
We could see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



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.

Launch Success Yearly Trend



We can observe that the success rate since 2013 kept increasing till 2020 with a slight drop in 2018

All Launch Site Names

```
In [11]: %sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

Out[11]:

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

Above listed are the unique launch sites.

Launch Site Names Begin with 'CCA'

```
In [12]: %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5
```

Out[12]:

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

Above result is for the query which gives 5records for the launch site name that starts with ‘CCA‘

Total Payload Mass

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

```
In [14]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

Out[14]:

1
45596

- 45596 is the total payload mass that are carried by the boosters launched by NASA

Average Payload Mass by F9 v1.1

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

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

```
In [15]: %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'
```

Out[15]:

1
2928

First Successful Ground Landing Date

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)'
```

]:

1
22-12-2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

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

Total Number of Successful and Failure Mission Outcomes

- total number of successful and failure mission outcomes

```
In [52]: %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
```

Out[52]:

1
100

Boosters Carried Maximum Payload

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

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

In [53]:
```
%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

Out[53]:

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

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

```
%sql select launch_site,booster_version from SPACEXTBL where  Landing__Outcome like 'Failure%' and (DATE between '2015-01-01' and '2015-12-31') order by date desc
```

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

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 * from SPACEXTBL where Landing_Outcome = ('Failure (drone ship)' or 'Success (drone ship)') and (DATE between '2010-06-04' and '2017-03-20') order by date
```

DATE	Time (UTC)	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	Landing _Outcome
10-01-2015	09:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
14-04-2015	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
17-01-2016	18:42:00	F9 v1.1 B1017	VAFB SLC-4E	Jason-3	553	LEO	NASA (LSP) NOAA CNES	Success	Failure (drone ship)
04-03-2016	23:35:00	F9 FT B1020	CCAFS LC-40	SES-9	5271	GTO	SES	Success	Failure (drone ship)
15-06-2016	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success	Failure (drone ship)

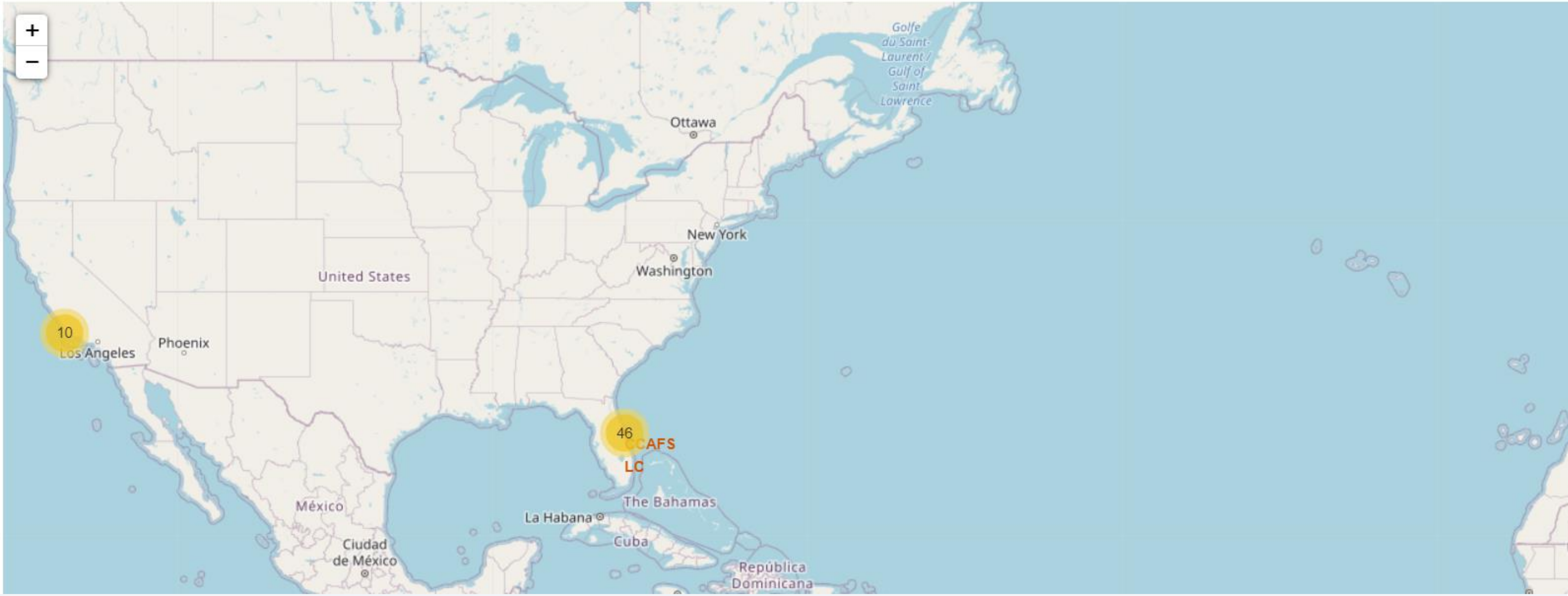
Section 3

Launch Sites Proximities Analysis

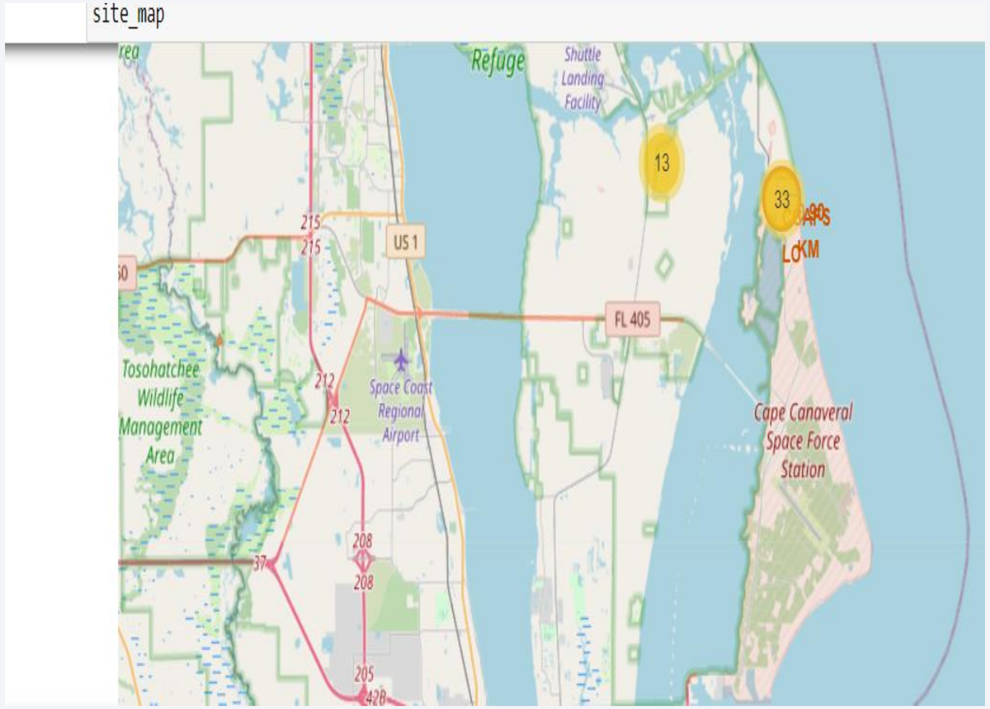
All launch sites marked on a map

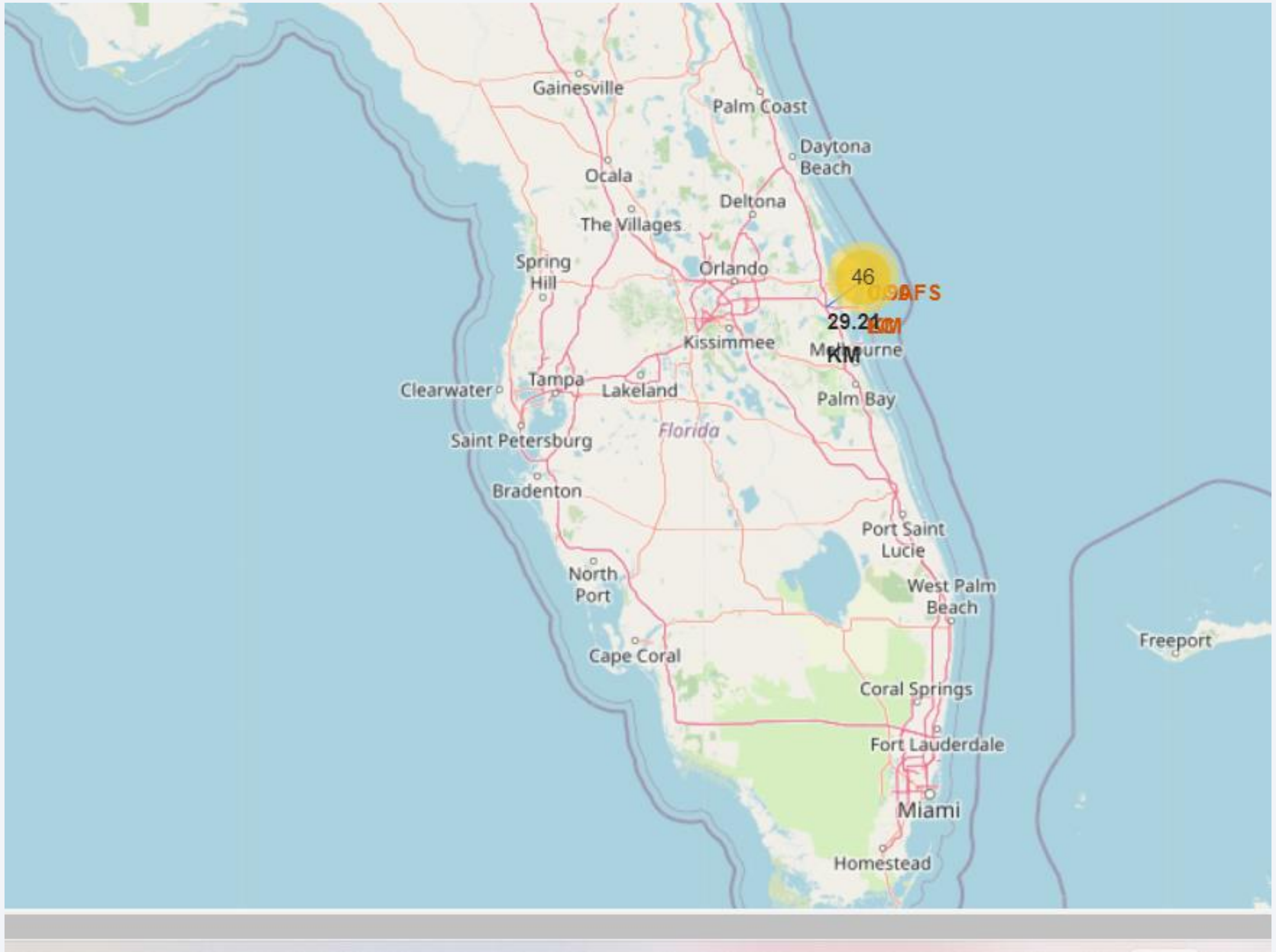
Success/failed launches marked on the map

Out[13]:



Distances between a launch site to its proximities

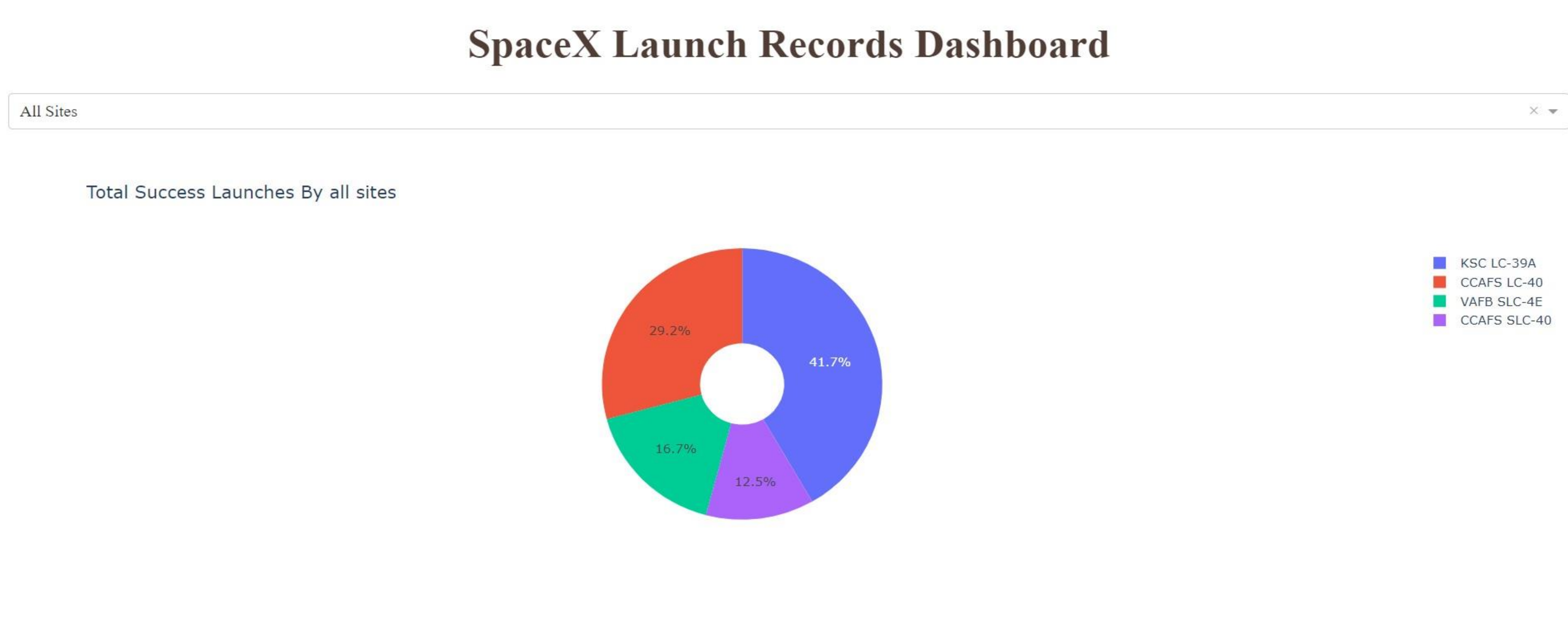




Section 4

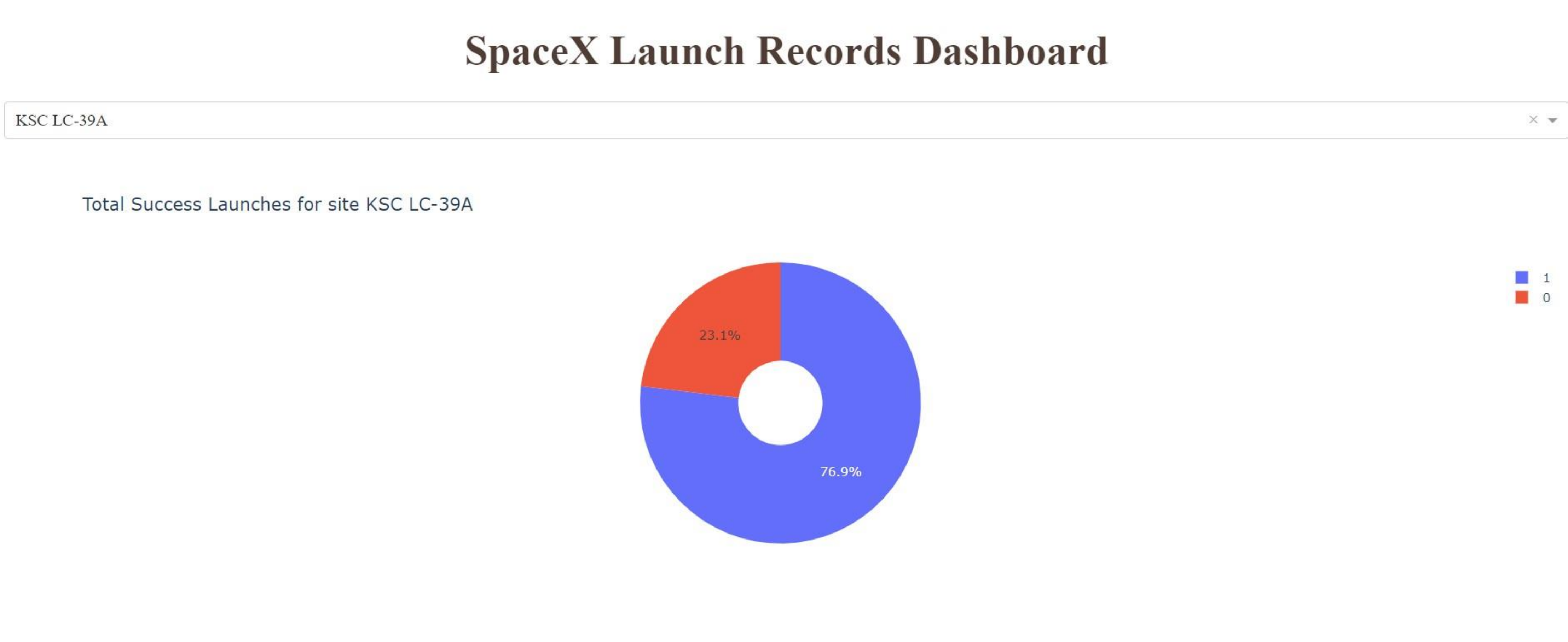
Build a Dashboard with Plotly Dash

Launch Success Count –All Sites



KSC LC -39A had the most successful launches from all sites

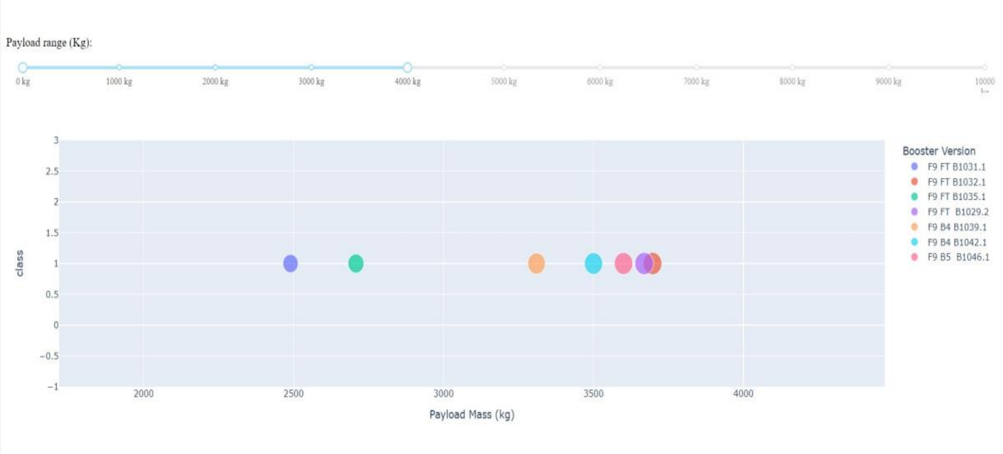
Highest Launch Success Ratio Site



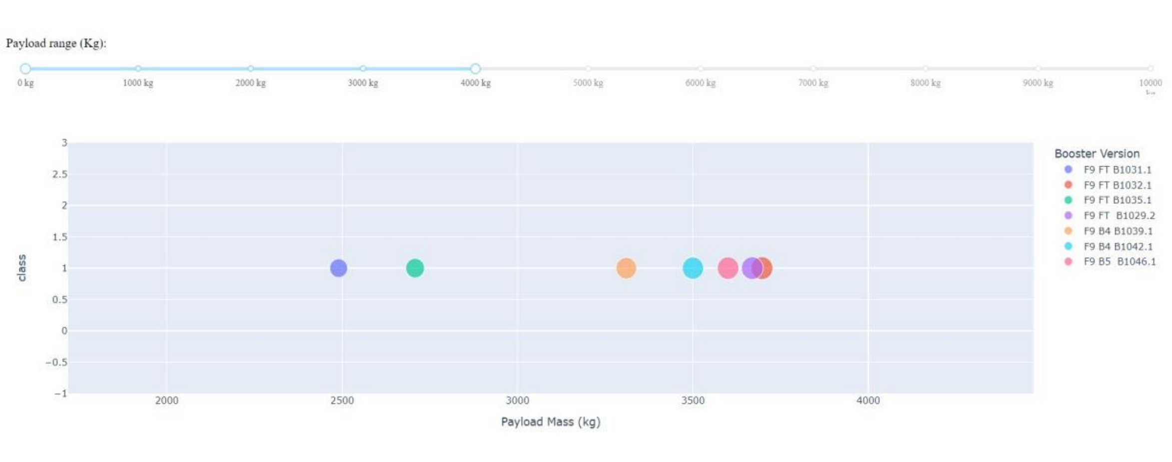
KSC LC-39A has the highest success rate of 76.9% and failure rate of 23.1%

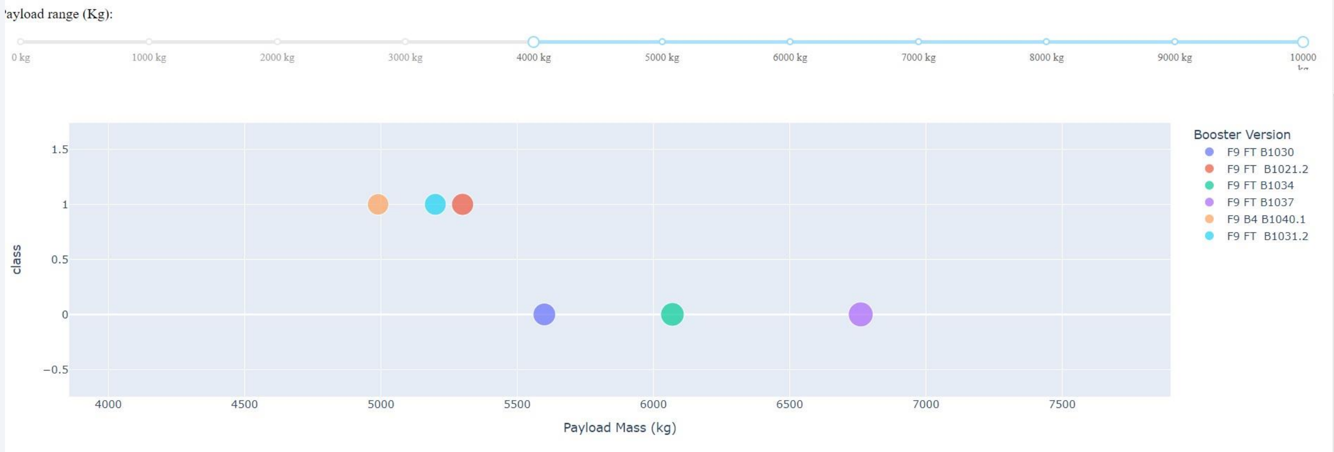
Payload vs. Launch Outcome Scatter Plot

Full Payload Range Slider (0-10K)



Payload vs. Launch Outcome (0-4k slider)





Payload vs.Launch Outcome (4k – 10 k slider)

Success rate is higher in 0 – 4K slider range

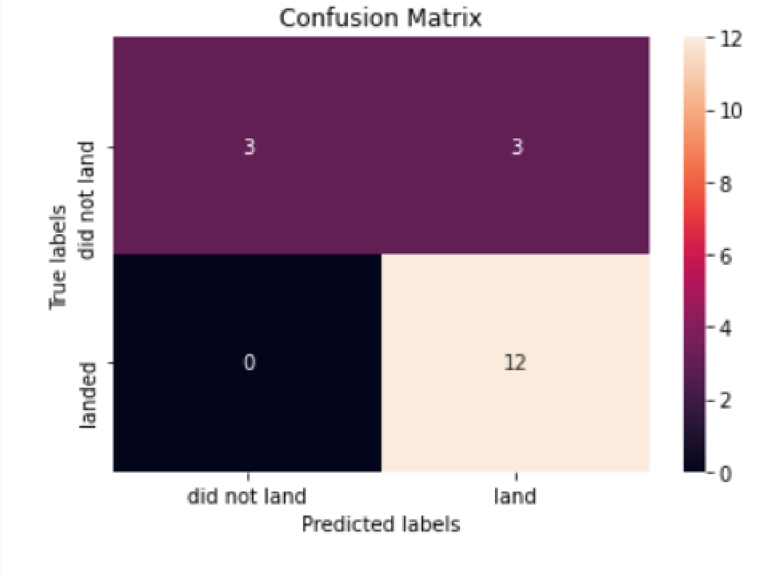
Section 5

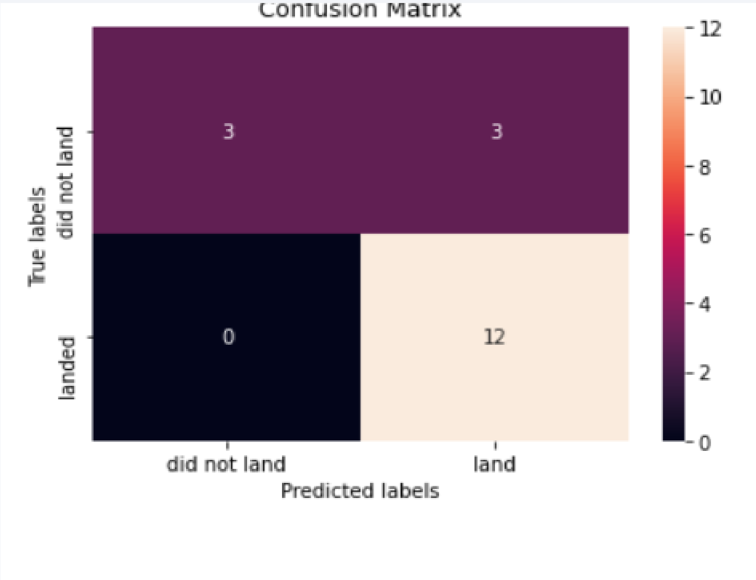
Predictive Analysis (Classification)

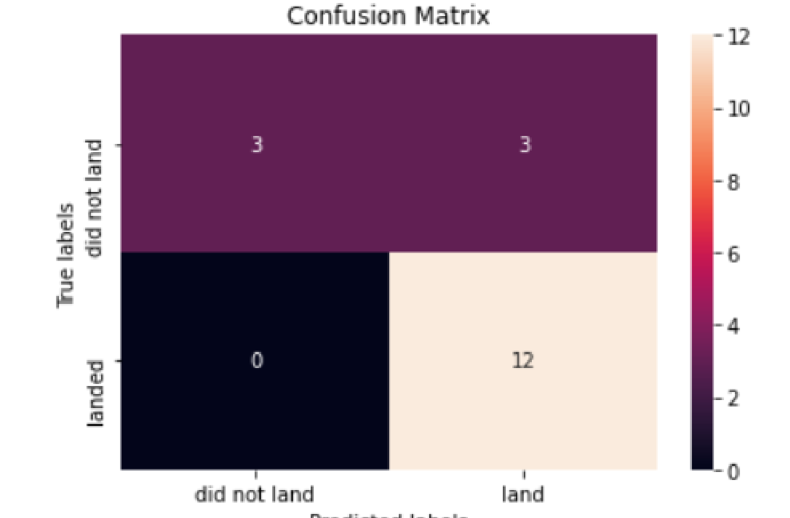
Classification Accuracy

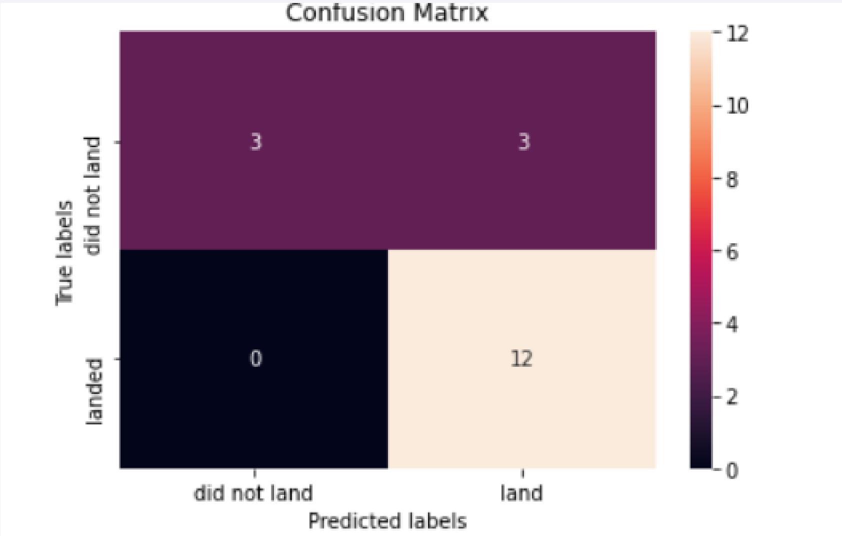
- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

Confusion Matrix









Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Thank you!