IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Niranjan Kumar
21/9/2021

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context
- Problems you want to find answers

Section 1

Methodology

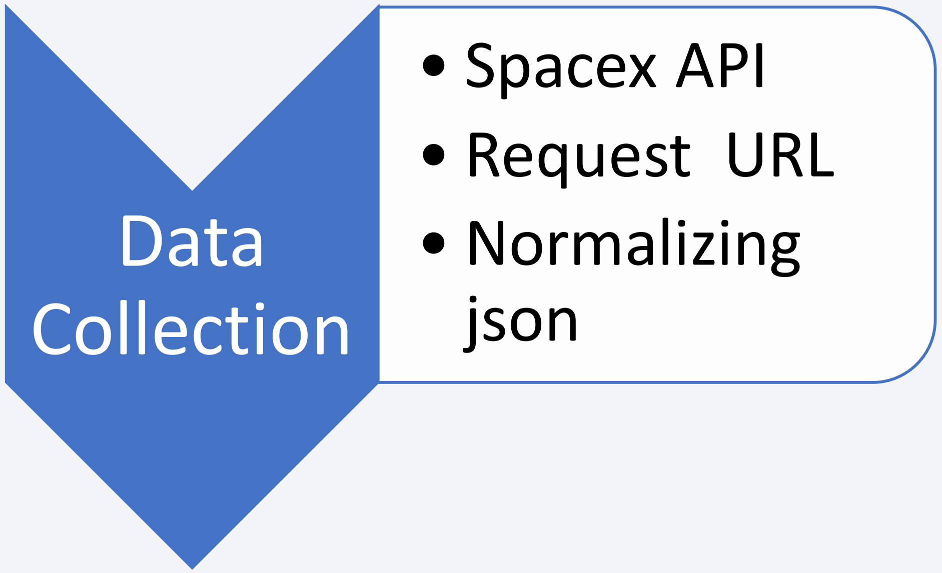
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected with the help of an API which helped us for analysis.
- Perform data wrangling
  - New ‘Class’ column has been created for better analysis.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - With the help of Python and SQL Explored the data
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Most of the Classification models are used; with the help of GridsearchCV, we tried to evaluate the best for each model.

Data Collection

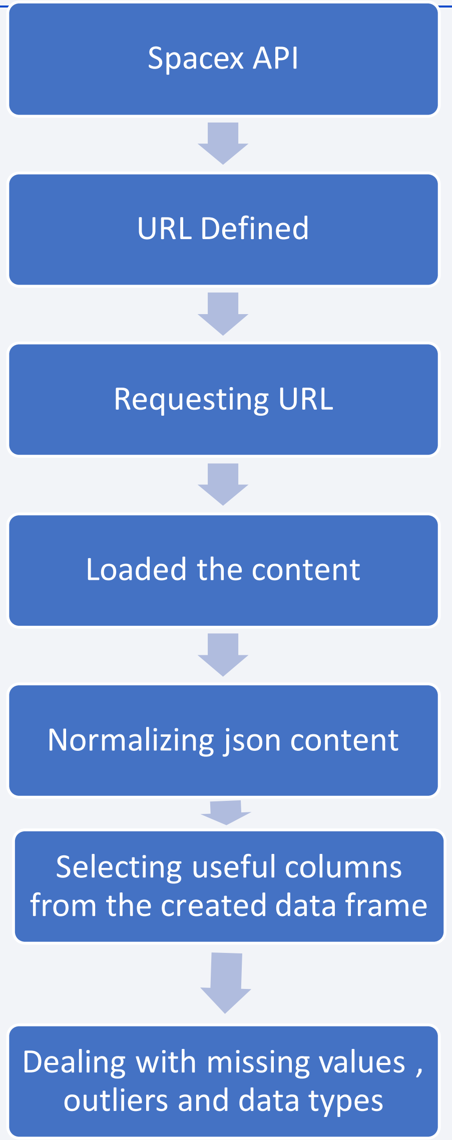
- Some amount of data is collected using API , while the rest of the data is collected by web scrapping from Wikipedia and then processed.



Data Collection – SpaceX API

1. Spacex API
2. Variable defined to ‘URL’
3. Created a request to the URL
4. Loaded the content
5. Normalized the json content
6. Selected the required columns and created a data frame for future process.
7. Dealing with missing values ,Outliers, different data types.
8. Finally creating extracting the cleaned data.

- [Notebook Link]



Data Collection - Scraping

1.Collecting [Website url]

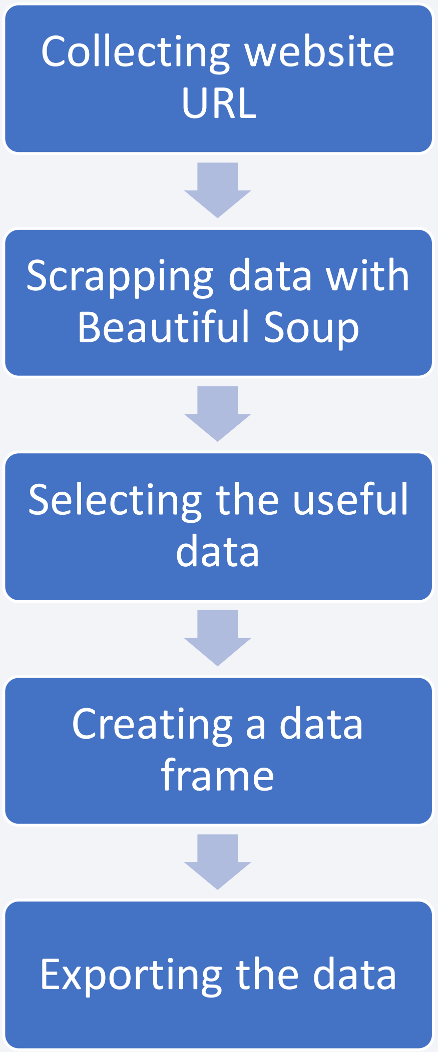
2. Scrapping with Beautiful Soup

3. Separating tables from rest

4.Seperating column names from table

5.Creating Data Frame.

- [Notebook]



Data Wrangling

- Data wrangling is the process of cleaning and unifying messy and complex data sets for easy access and analysis.
- I created new columns which are useful for the analysis.
- Notebook link

EDA with Data Visualization

1.scatter point chart:
 Find out how different variables would affect the launch outcome.
2.bar chart:
 Find which orbits have high success rate.
3.line chart:
 Visualize the launch success yearly trend

[Notebook Link]

EDA with SQL

SQL Queries Performed in the notebook

1.%sql select LAUNCH_SITE from SPACEXTBL group by LAUNCH_SITE

2.%sql select LAUNCH_SITE from SPACEXTBL where LAUNCH_SITE LIKE 'CCA%' limit 5

3.%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)’

4.%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1’

5.%sql select min(DATE) from SPACEXTBL where LANDING__OUTCOME = 'Success (ground pad)’

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

7.%sql select LANDING__OUTCOME from SPACEXTBL group by LANDING__OUTCOME

%sql select MISSION_OUTCOME, count(MISSION_OUTCOME) from SPACEXTBL group by MISSION_OUTCOME

8.%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)

9.%sql select DATE,BOOSTER_VERSION,LAUNCH_SITE from SPACEXTBL where LANDING__OUTCOME = 'Failure (drone ship)' and DATE LIKE '2015%’

10.%sql select LANDING__OUTCOME, count(LANDING__OUTCOME) as numbers from SPACEXTBL where DATE between '2010-06-04' and '2017-03-20' group by LANDING__OUTCOME order by numbers des

[Notebook Link]

Build an Interactive Map with Folium

1. **Object1:** folium Map object
   Key Code: site_map = folium.Map(location=nasa_coordinate, zoom_start=10)
   Explain: to add an initial map

2. **Object2:** circle object
   Key Code: circle=folium.Circle([lat, lng], radius=1000, color='#d35400', fill=True).add_child(folium.Popup(label))
   Explain : to create and add folium.Circle for each launch site on the site map

3. **Object3:** markerCluster object
   Key Code: site_map.add_child(marker_cluster)
   Explain : to simplify a map containing many markers having the same coordinate.

4. **Object4:** marker object
   Key Code: folium.Marker([lat, lng],icon=folium.Icon(color='white', icon_color=colors),
popup=label).add_to(marker_cluster)
   Explain : to mark the success/failed launches for each site on the ma

5. **Object5:** polyLine object
   Key Code: folium.PolyLine([(28.56341,-80.57678),(28.563197, -80.576820)],c

- We have explored this data to better understand , the successful projects and failed projects.

[Notebook Link]()

Predictive Analysis (Classification)

- First, we will split the data into 80% train data and 20% test data. There after we will define some parameters and use “GridSearchCv” to choose the best parameters and method for a given dataset.
- we will choose the best method and build a model.

Notebook Link

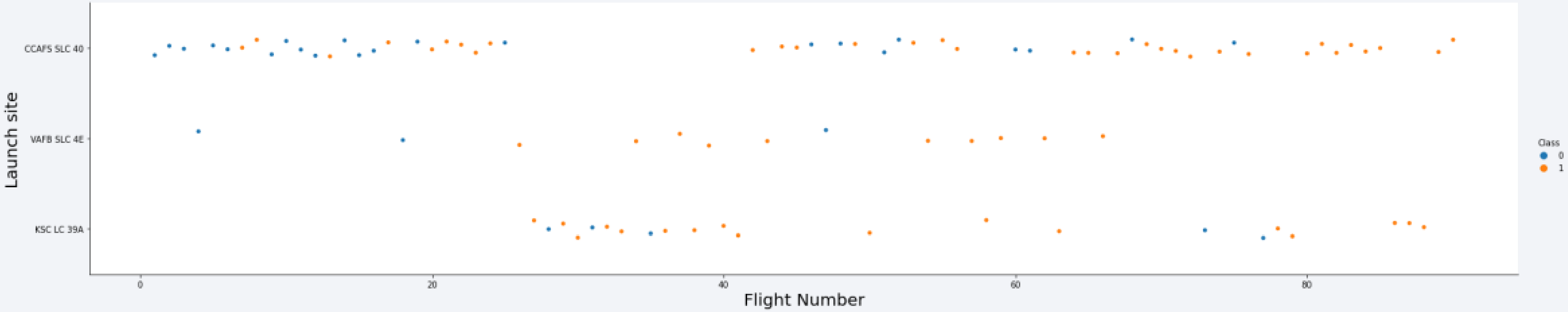
Results

- We found that data is normally distributed.
- Interactive analytics dashboards showed the successful launches vs failed launches.
- Predictive analysis shows us all the models are performing equally.

Section 2

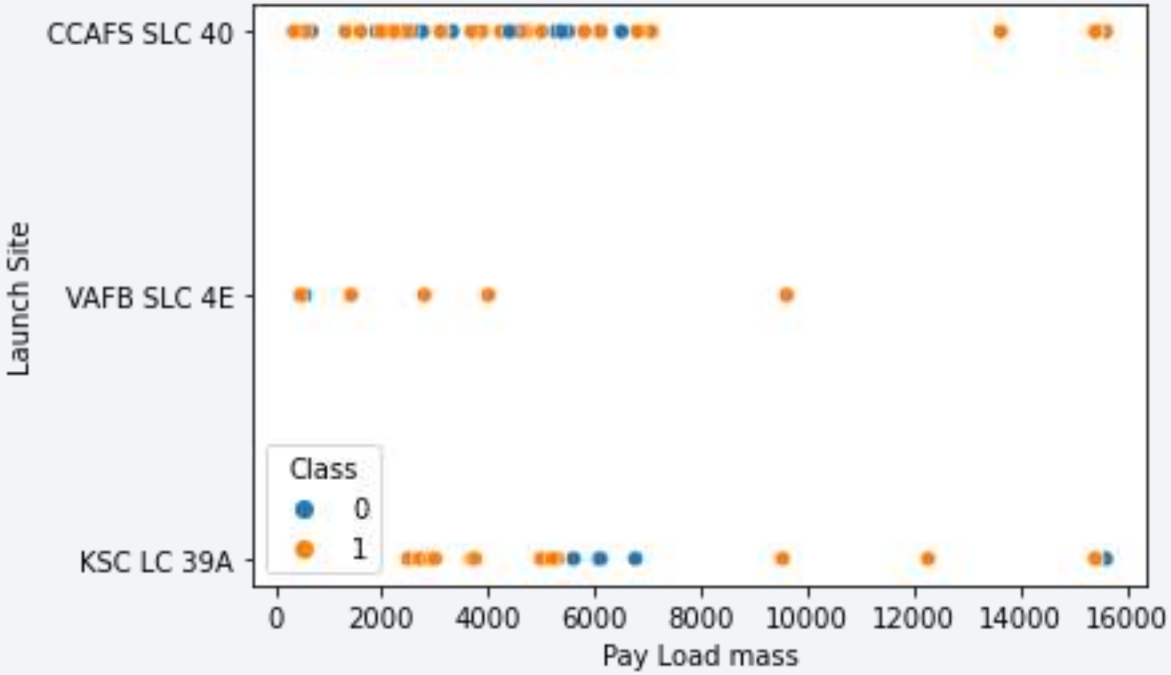
Insights drawn from EDA

Flight Number vs. Launch Site



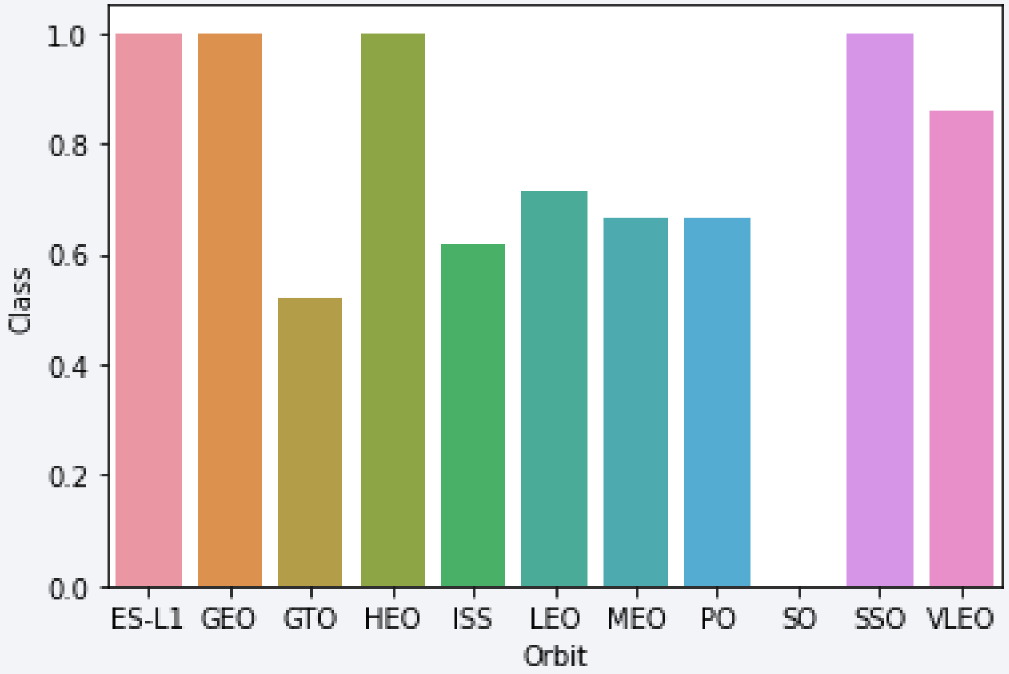
Here we can clearly see that most of the flights are launched from CCAFS SLC 40.

Payload vs. Launch Site

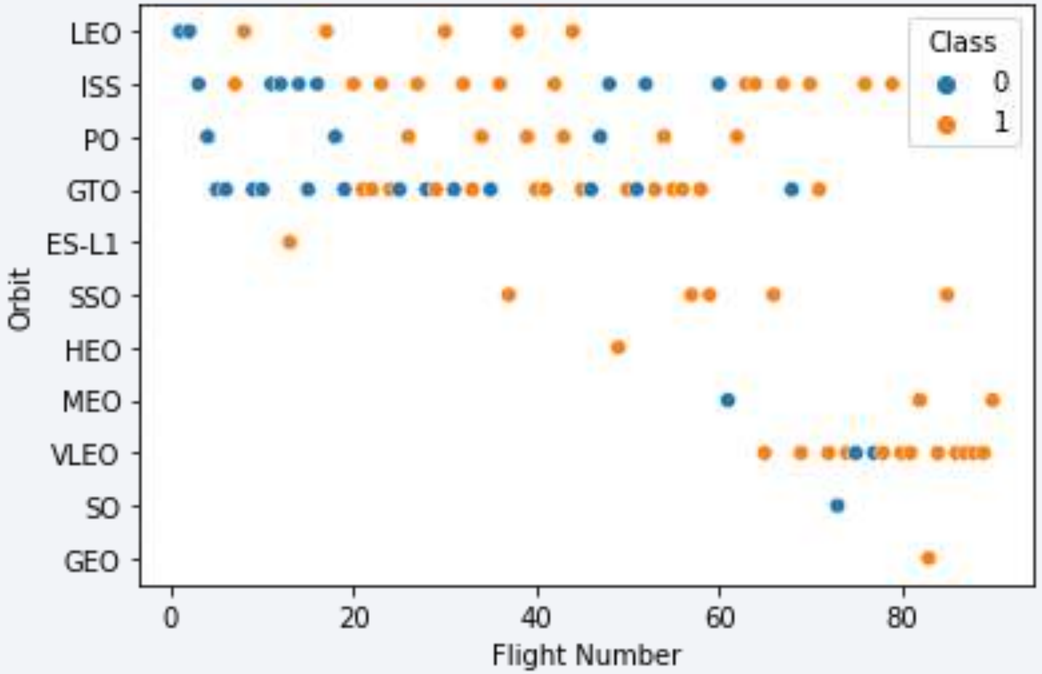


CCAFS SLC 40 is most used launch site and the rest of follows along

Success Rate vs. Orbit Type

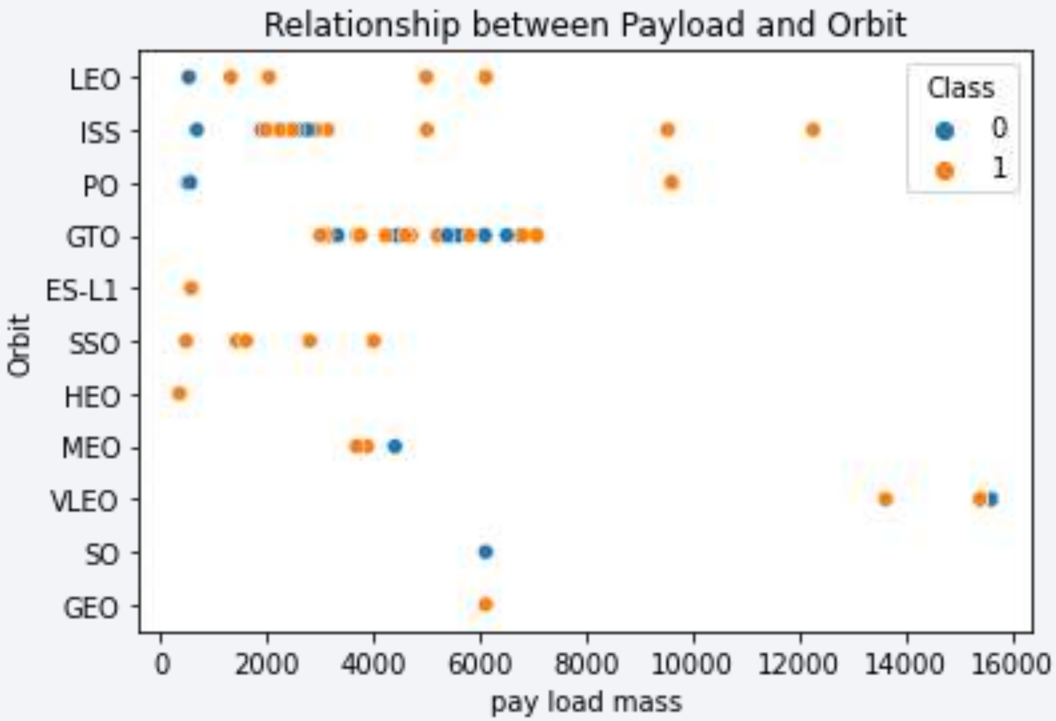


Flight Number vs. Orbit Type



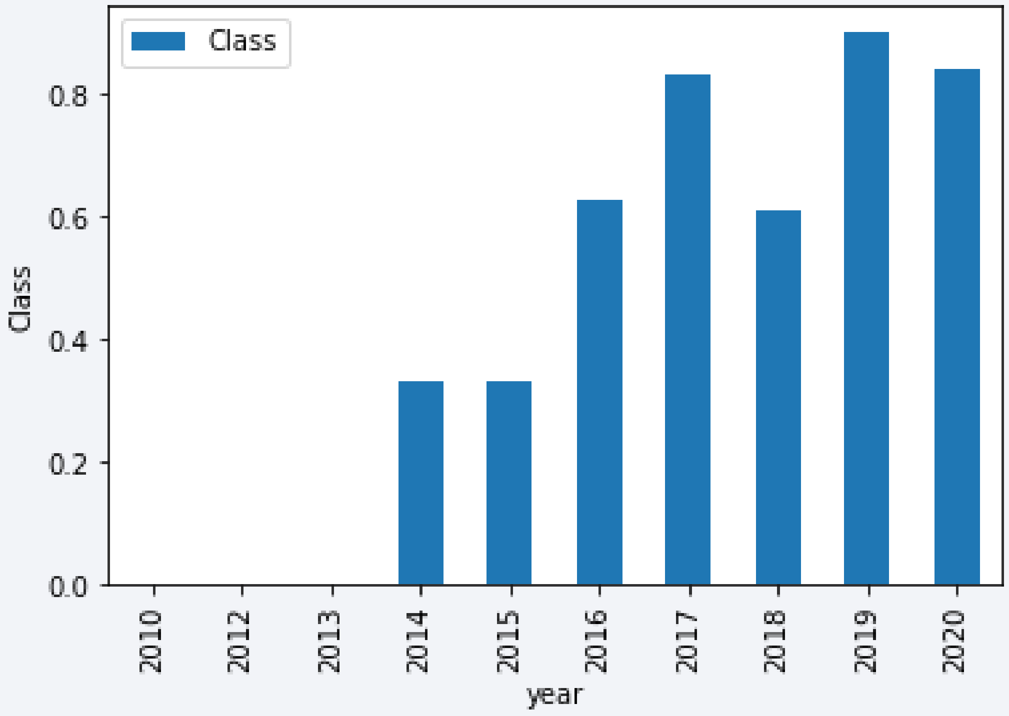
LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend



sucess rate since 2013 kept increasing till 2020

All Launch Site Names

In [4]:
```
%sql select DISTINCT launch_site from Spacextbl
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[4]:

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

Launch Site Names Begin with 'CCA'

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

In [5]:
```
%sql select * from Spacextbl where launch_site like 'CCA%' LIMIT 5;
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[5]:

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

Total Payload Mass

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

In [6]:
```
%sql select SUM(payload_mass__kg_) from Spacextbl where customer= 'NASA (CRS)';
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[6]:

1
45596

Average Payload Mass by F9 v1.1

```
In [7]: %sql select AVG(payload_mass__kg_) from Spacextbl where booster_version Like 'F9 v1.1%'

 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
Out[7]:
```

1
2534

First Successful Ground Landing Date

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In [8]:
```
%sql select DATE from Spacextbl WHERE mission_outcome = 'Success' ORDER BY DATE ASC Limit 1
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[8]:

DATE
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [9]: %sql select booster_version from Spacextbl where mission_outcome = 'Success' and 6000 <payload_mass__k
        g_ > 4000;

 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdom
ain.cloud:32731/bludb
Done.

Out[9]:
```

booster_version

Total Number of Successful and Failure Mission Outcomes

In [10]:
```
%sql select COUNT(mission_outcome) from Spacextbl where mission_outcome = 'Success';
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[10]:

1
99

In [11]:
```
%sql select COUNT(mission_outcome) from Spacextbl where mission_outcome like 'Fail%';
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[11]:

1
1

Boosters Carried Maximum Payload

```
In [13]: %sql select max(PAYLOAD_MASS__KG_) from Spacextbl

 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdom
ain.cloud:32731/bludb
Done.
```

Out[13]:

1
15600

2015 Launch Records

```
In [14]: %sql select DATE,BOOSTER_VERSION,LAUNCH_SITE from Spacextbl where LANDING__OUTCOME = 'Failure (drone s
hip)' and DATE LIKE '2015%'

 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdom
ain.cloud:32731/bludb
Done.
```

Out[14]:

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

In [15]:
```
%sql select LANDING__OUTCOME, count(LANDING__OUTCOME) as numbers from Spacextbl where DATE between '20
10-06-04' and '2017-03-20'  group by LANDING__OUTCOME order by numbers desc
```

```
 * ibm_db_sa://tsz62986:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdom
ain.cloud:32731/bludb
Done.
```

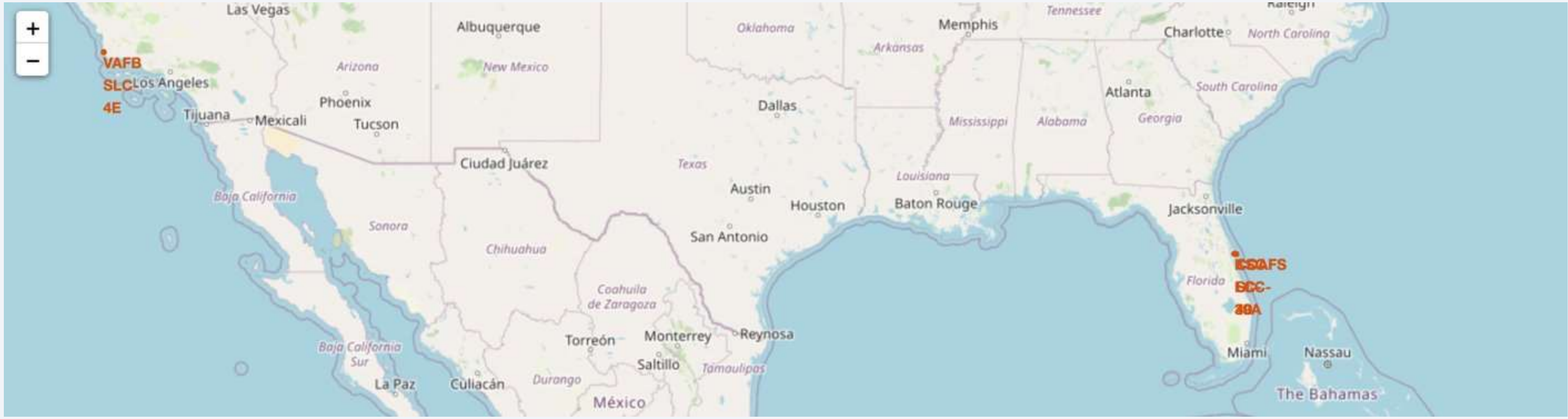
Out[15]:

landing__outcome	numbers
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

Section 4

Launch Sites Proximities Analysis

All Launch sites on map



The success/failed launches for each site on the map

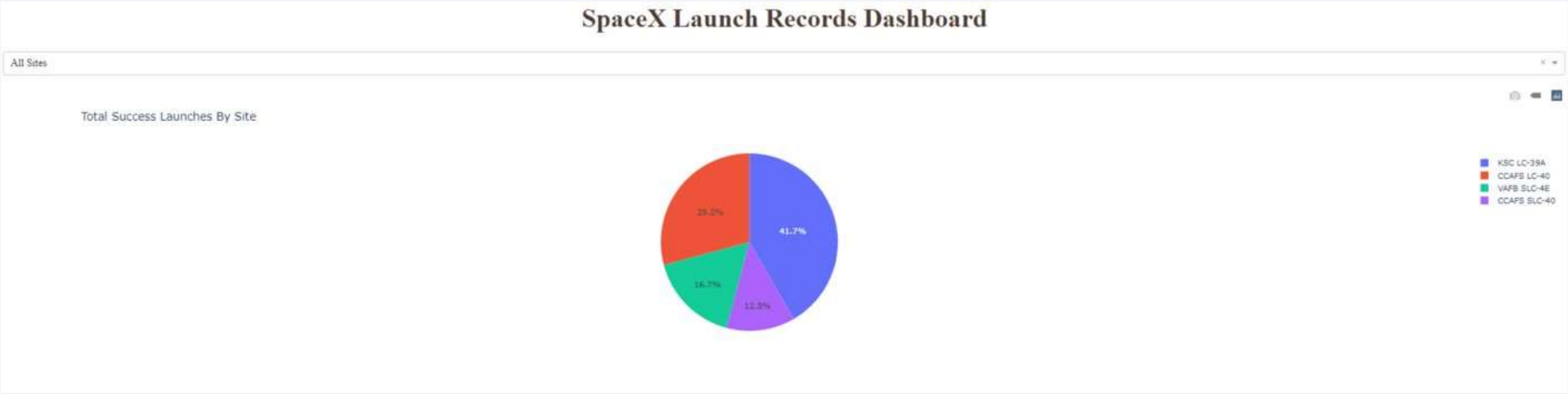
The distances between a launch site to its proximities

Section 5

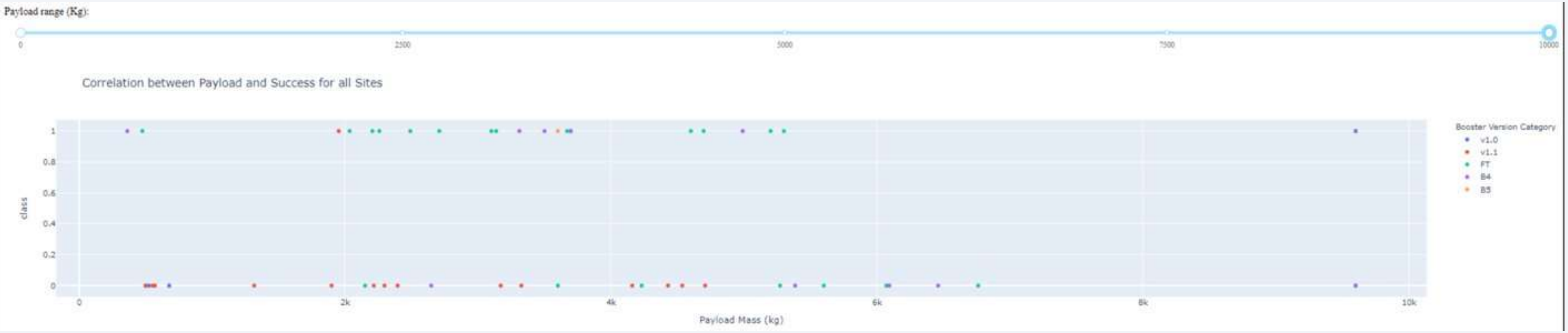
Build a Dashboard with Plotly Dash

Pie Chart-All Launch Site

- Explanations:
  - KSC LC-39A launch sites has the highest success rates
  - CCAFS SLC-40 launch sites has the lowest success rates



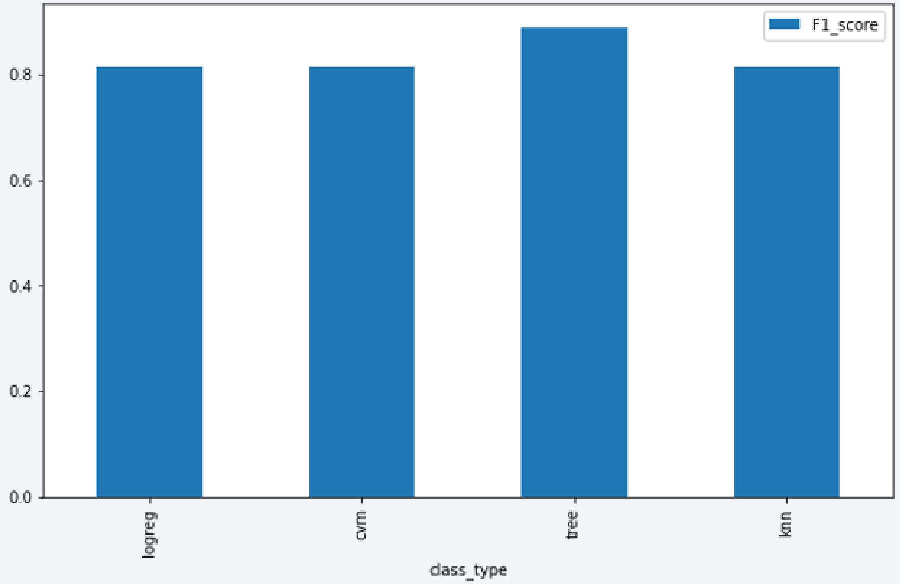
Scatter Plot- Payload vs. Launch Outcome



Section 6

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

Conclusions

- Orbit, Launch Site, Flight Number, Payload Mass, would affect the success rate
- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.
- The more massive the payload, the less likely the first stage will return
- Decision tree model is able to predict whether the Falcon 9 first stage will land successfully .

Thank you!