IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jayavarshini
09-01-2022

Outline

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

Executive Summary

- Summary of methodologies:
  - Data collected from SPACEX REST API and by web scraping wiki pages and data is prepared for further analysis and visualization purpose
  - Exploratory data analysis is done using visualization, SQL queries and plotly dashboard.
  - Data is converted into numerical for machine learning purpose and a model is created to predict the landing outcome.
- Summary of all results:
  - With visualization techniques we understand the effect of payload mass, orbit, flight numbers and their relationship to the successful landing.
  - With dashboard visualization we understand the role of payload mass and booster version category in the landing outcomes.

Introduction

- The project involves collecting the data, preparing it for analysis and finding the effect of every feature on the outcome of landing.
- Problems that needs to be answered:
  - We need to predict whether the first stage of a new launch will successfully land or not with the data we obtained.
  - We also need to know what are the factors that most affects the success and failure of the first stage landing.

Section 1

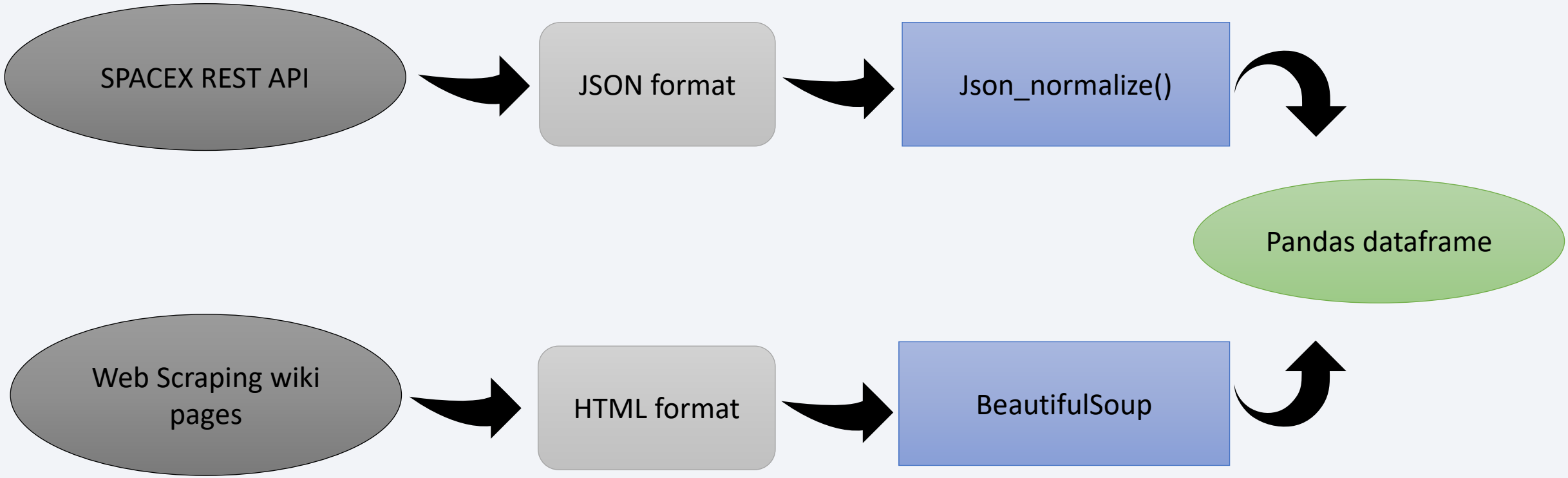
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data is gathered from SPACEX REST API (in JSON format) and through web scraping wiki pages (as HTML tables).
- Perform data wrangling
  - Filtered for Falcon 9 launches, replaced NULL values (with mean in payload mass) and converted categorical columns into numerical values.
- 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 data, train-test-split to train model and perform grid search (to find best hyperparameters), test the models and output the confusion matrix.

Data Collection

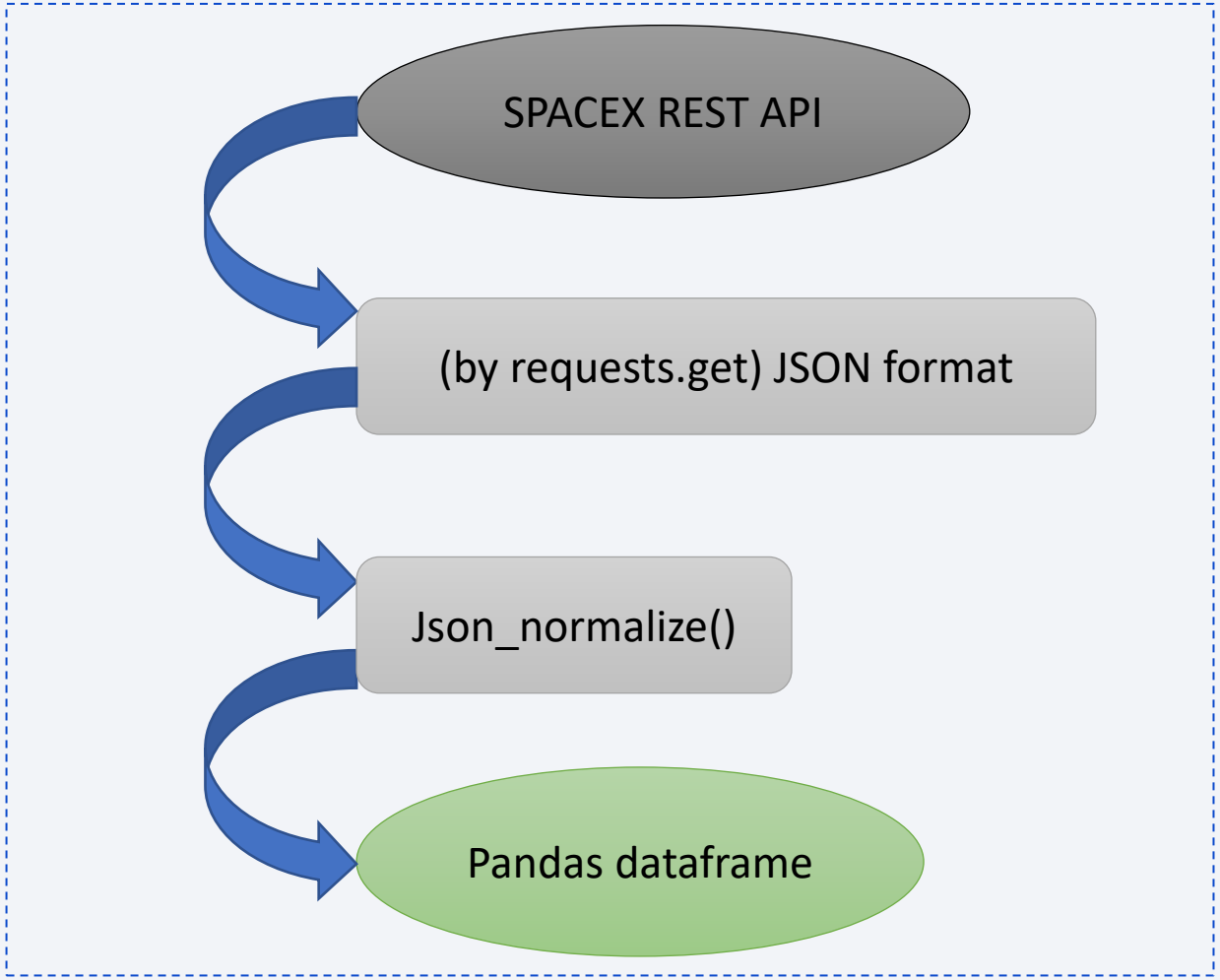


Data Collection – SpaceX API

GitHub URL :

datacollectionAPI

(for external reference and peer-review purpose)

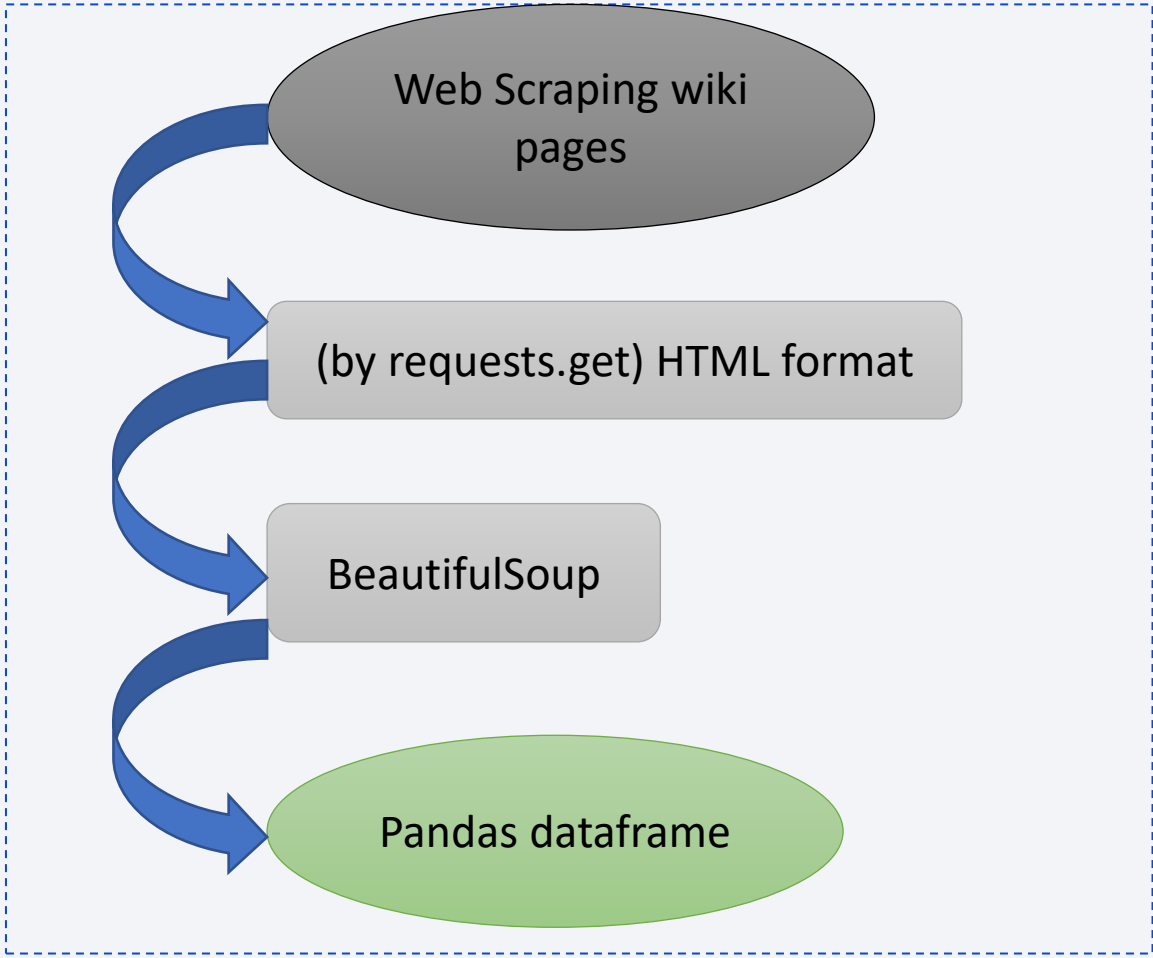


Data Collection - Scraping

GitHub URL :

DataCollectionWebScraping

(for external reference and peer-review purpose)

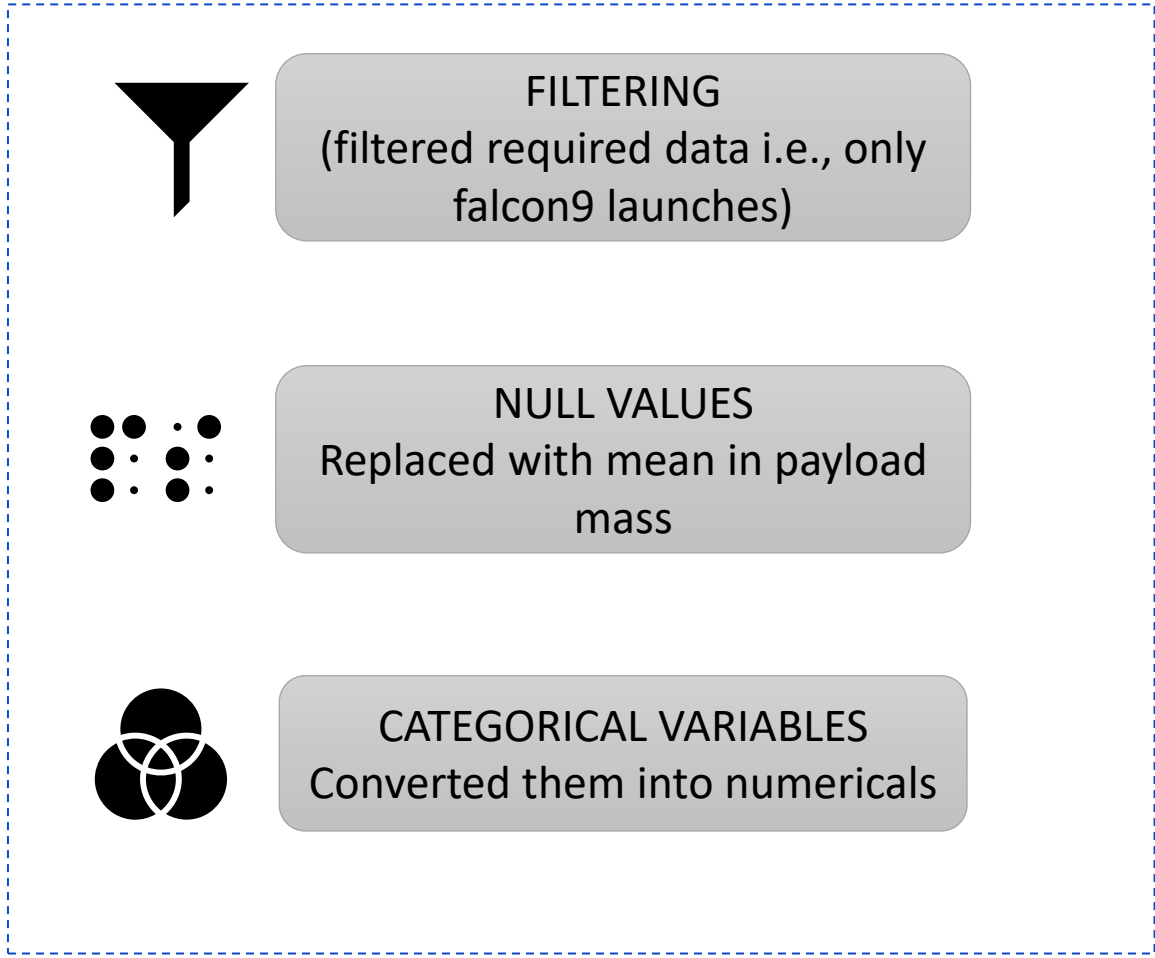


Data Wrangling

GitHub URL :

dataWrangling

(for external reference and peer-review purpose)



EDA with Data Visualization

- Line plot between year and success rate.
- Bar chart between orbit and success rate.
- Scatter plots where used to find relationship between:
    - Payload mass and launch site
    - Payload mass and flight numbers
    - Payload mass and orbit
    - Launch site and flight numbers
    - Launch site and payload mass
    - Orbit and flight number
- GitHub URL : EDAwithdatavisualization (for external reference and peer-review purpose)

EDA with SQL

- SQL queries were done to list launch sites, outcome counts, booster versions of different outcomes and first date of successful landing.
- Also to calculate total payload mass, average payload mass, maximum payload mass (along with their booster versions)
- Also to rank the count of different outcomes.

- GitHub URL : EDAwithSQL (for external reference and peer-review purpose)

Build an Interactive Map with Folium

- Map objects such as markers, circles, lines, marker cluster, mouse position and distance marker were created and added to a folium map
  - Circle-to mark a location along with popup
  - markers-to mark name of the location
  - marker cluster-to mark outcomes at the site
  - mouse position-to display latitude and longitude points on the map
  - distance marker-to mark distance
  - Lines-to draw lines connecting sites to closest proximities.

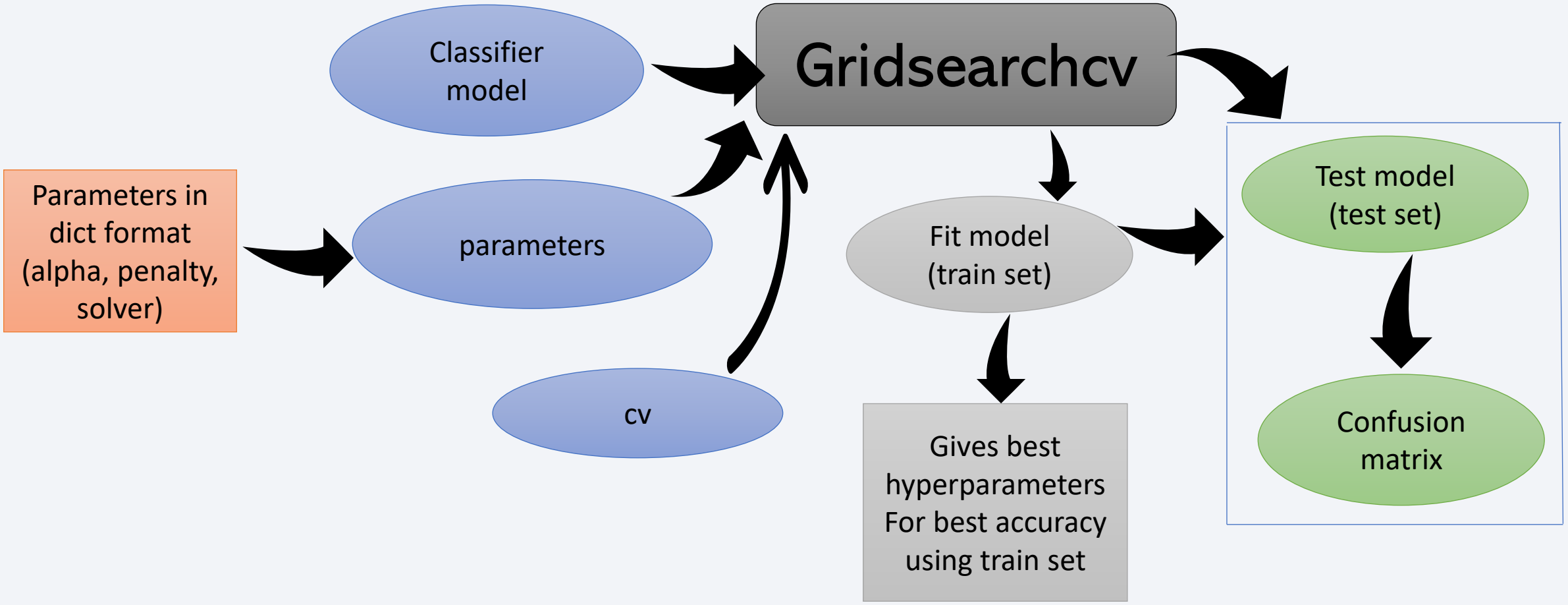
- GitHub URL : FoliumMap (for external reference and peer-review purpose)

Build a Dashboard with Plotly Dash

- Graph and chart used:
  - scatter plot between payload mass and success rate with different booster version categories
  - Pie chart for displaying the success rates for each launch site and for all sites
- Interactions:
  - Dropdown: to choose the site for which the success rate information is needed
  - Range slider: to set the range of payload mass

- GitHub URL : dashboard (for external reference and peer-review purpose)

Predictive Analysis (Classification)



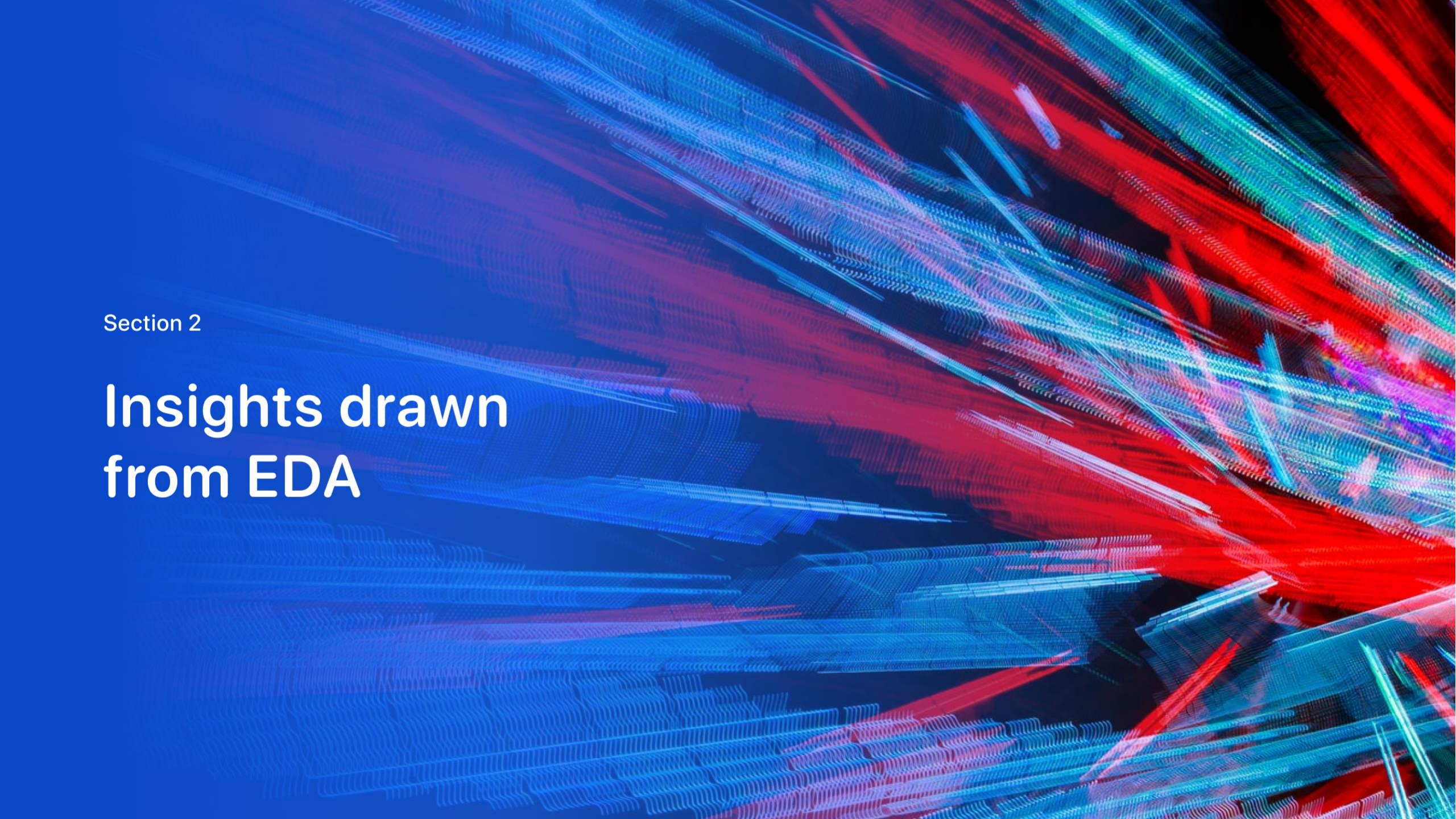
GitHub URL : machinelearning (for external reference and peer-review purpose)

Results

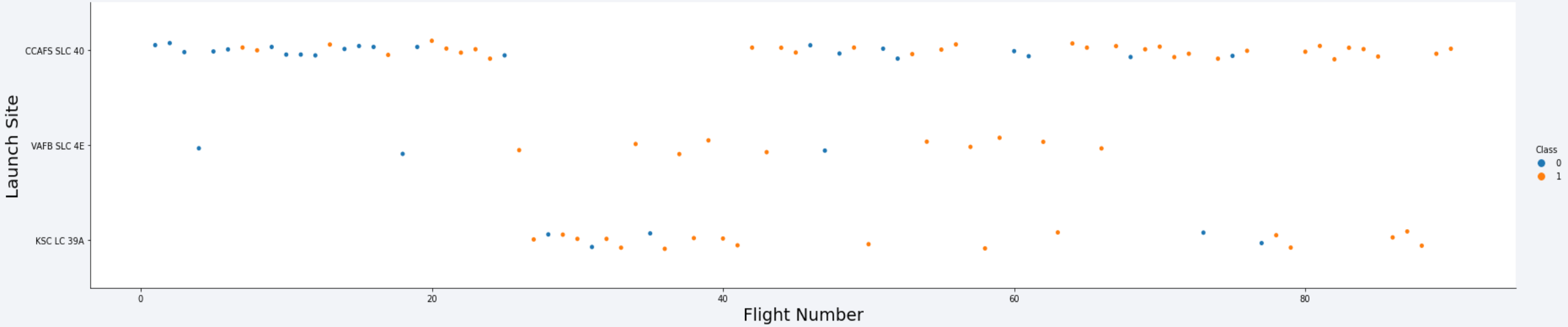
- From exploratory data analysis it is clear that payload mass has an impact in the landing outcome, along with the orbit and booster versions.
- Folium map displays the proximities of launch sites to nearby coastlines, highway and railways.
- Predictive analysis helps to predict the outcome with 83.3% of accuracy.

Section 2

Insights drawn from EDA

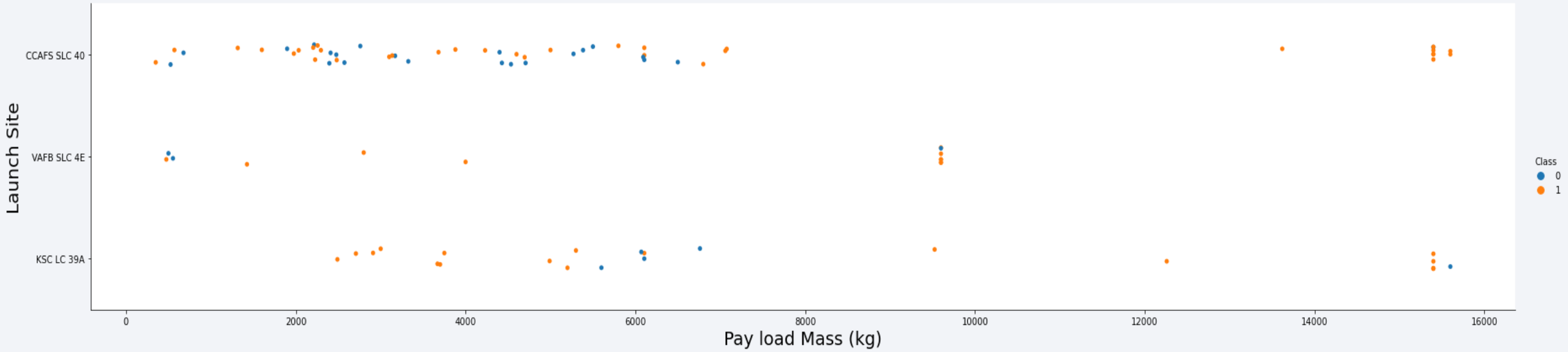


Flight Number vs. Launch Site



- In CCAFS SLC-40 site success appears related to the flight numbers.
- In VAFB SLC-4E site there are very less launches.

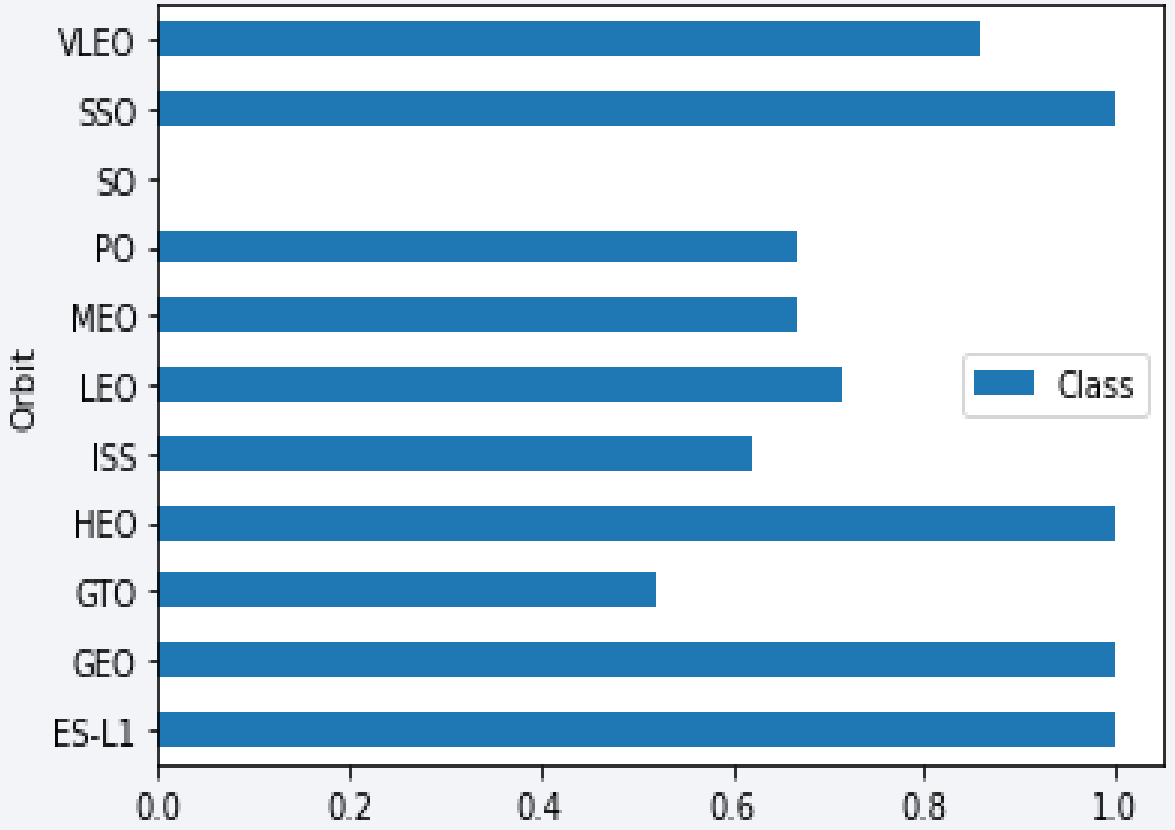
Payload vs. Launch Site



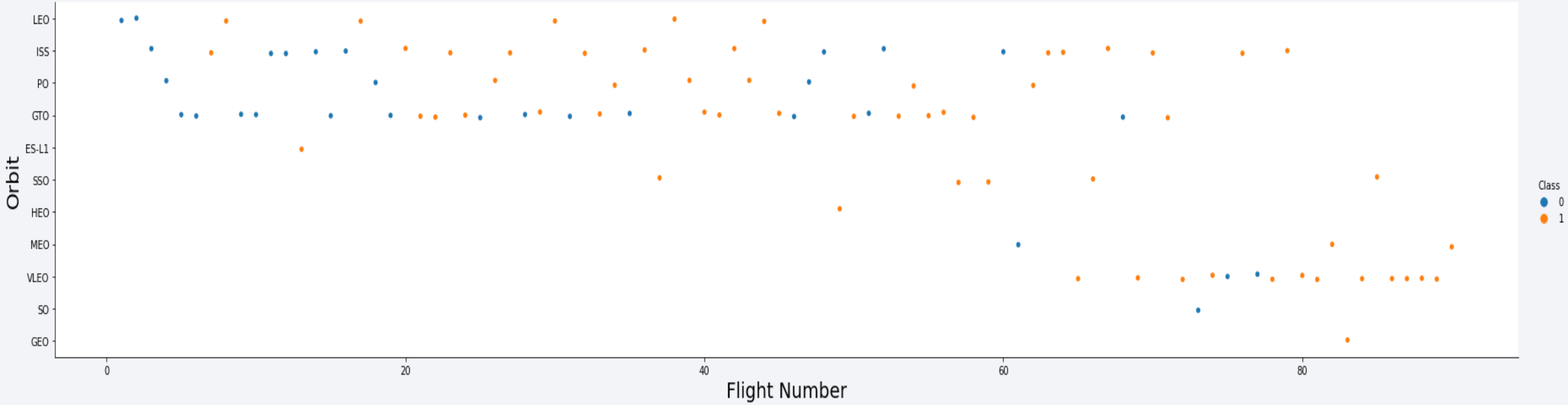
- VAFB-SLC launch site there are no launches for payload mass >10000kg.
- For payload mass > 8000kg, it is a success in CCAFS SLC-40 site.

Success Rate vs. Orbit Type

- In ES-L1, GEO, HEO, SSO orbits it is always a success.
- GTO orbit has the lowest success rate

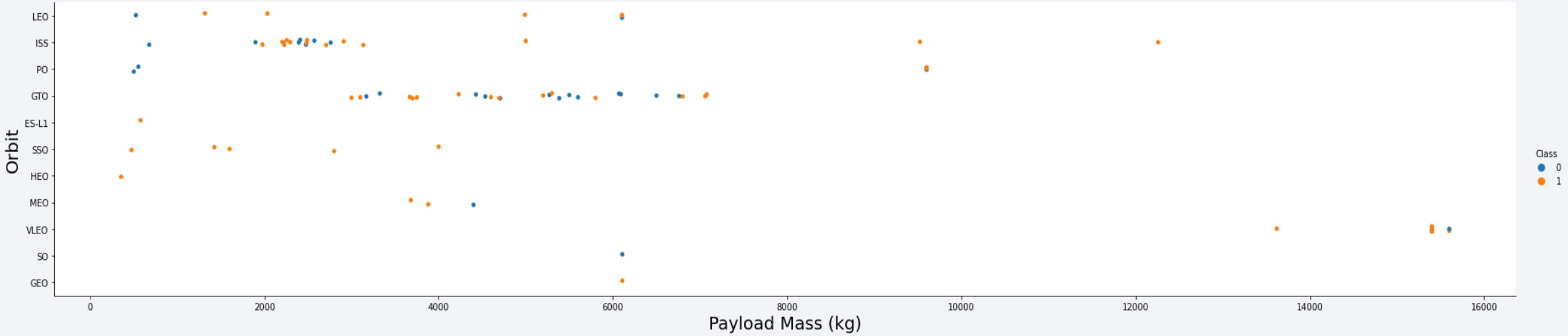


Flight Number vs. Orbit Type



- In the LEO orbit the Success appears related to the number of flights.
- GTO orbit has no relationship with the flight numbers.

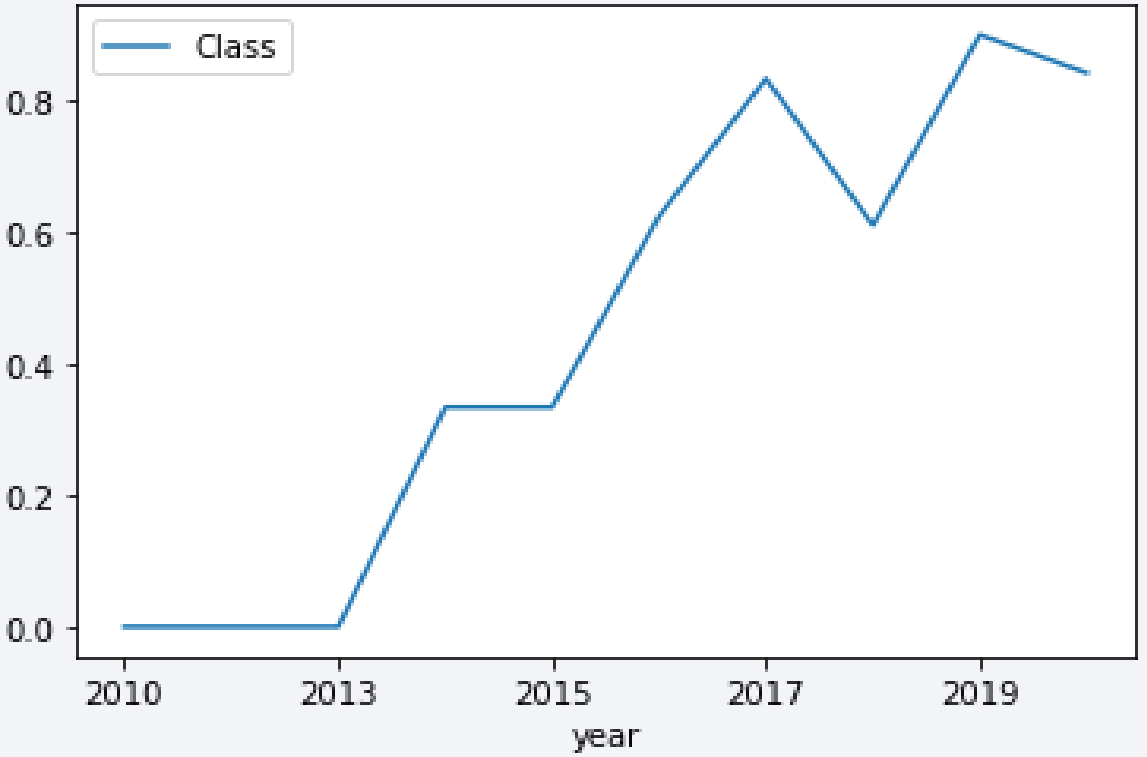
Payload vs. Orbit Type



- The success rate is more with heavy payloads in PO, LEO and ISS and with lower payloads in SSO, HEO, ES-L1.
- GTO has no relationship with payload mass.

Launch Success Yearly Trend

- The success rate since 2013 kept on increasing till 2020.



All Launch Site Names

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

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

Out[11]:

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

- Where CCAFS LC-40 and CCAFS SLC-40 are in close proximity compared to others.

Launch Site Names Begin with 'CCA'

```
%%sql select * from SPACEXDATASET
where LAUNCH_SITE like 'CCA%'
limit 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

Data record of launch sites beginning with CCA (CCAFS SLC-40 and CCAFS LC-40)

Total Payload Mass

45596 kg

```
In [13]: %%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET
         where customer = 'NASA (CRS)'

Out[13]: total_payload_mass
         45596
```

- Total payload mass carried by boosters launched by NASA (CRS) is 45596 kg.

Average Payload Mass by F9 v1.1

2928 kg

```
In [14]: %%sql select avg(payload_mass__kg_) as avg_payload_mass from SPACEXDATASET
         where booster_version = 'F9 v1.1'
```

Out[14]:

avg_payload_mass
2928

- The average payload mass carried by booster version F9 v1.1 is 2928 kg.

First Successful Ground Landing Date

2015-12-22

```
In [15]: %%sql select min(DATE) as first_successful_landing from SPACEXDATASET
         where landing__outcome = 'Success (ground pad)'

Out[15]: first_successful_landing
         2015-12-22
```

- The date of the first successful landing outcome on ground pad is 2015-12-22.

Successful Drone Ship Landing with Payload between 4000 and 6000

Booster versions
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

```
In [16]: %%sql select booster_version,landing__outcome from SPACEXDATASET
         where (landing__outcome = 'Success (drone ship)') and (payload_mass__kg_ between 4000 and 6000)
```

Out[16]:

booster_version	landing__outcome
F9 FT B1022	Success (drone ship)
F9 FT B1026	Success (drone ship)
F9 FT B1021.2	Success (drone ship)
F9 FT B1031.2	Success (drone ship)

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

Total Number of Successful and Failure Mission Outcomes

Mission outcome	count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

```
In [17]: %%sql select mission_outcome,count(*) as count from SPACEXDATASET
         group by mission_outcome
```

Out[17]:

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

- The total number of successful and failure mission outcomes are listed.

Boosters Carried Maximum Payload

Booster versions	
F9 B5 B1048.4	F9 B5 B1049.5
F9 B5 B1049.4	F9 B5 B1060.2
F9 B5 B1051.3	F9 B5 B1058.3
F9 B5 B1056.4	F9 B5 B1051.6
F9 B5 B1048.5	F9 B5 B1060.3
F9 B5 B1051.4	F9 B5 B1049.7

```
%%sql select booster_version,payload_mass__kg_ from SPACEXDATASET
where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET)
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

- The names of the booster(12) which have carried the maximum payload mass(15600 kg) are listed.

2015 Launch Records

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)

```
%%sql select DATE,booster_version,launch_site,landing__outcome from SPACEXDATASET
where (landing__outcome = 'Failure (drone ship)') and (DATE like '2015%')
```

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)

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015 are listed.

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

RANK	COUNT	Landing outcome
1	10	No attempt
2	5	Failure (drone ship)
2	5	Success (drone ship)
3	3	Controlled (ocean)
3	3	Success (ground pad)
4	2	Failure (parachute)
4	2	Uncontrolled (ocean)
5	1	Precluded (drone ship)

In [20]:
```
%%sql select count(landing__outcome) as count,landing__outcome from SPACEXDATASET
where DATE>='2010-06-04' and DATE<='2017-03-20'
group by landing__outcome
order by count desc
```

Out[20]:

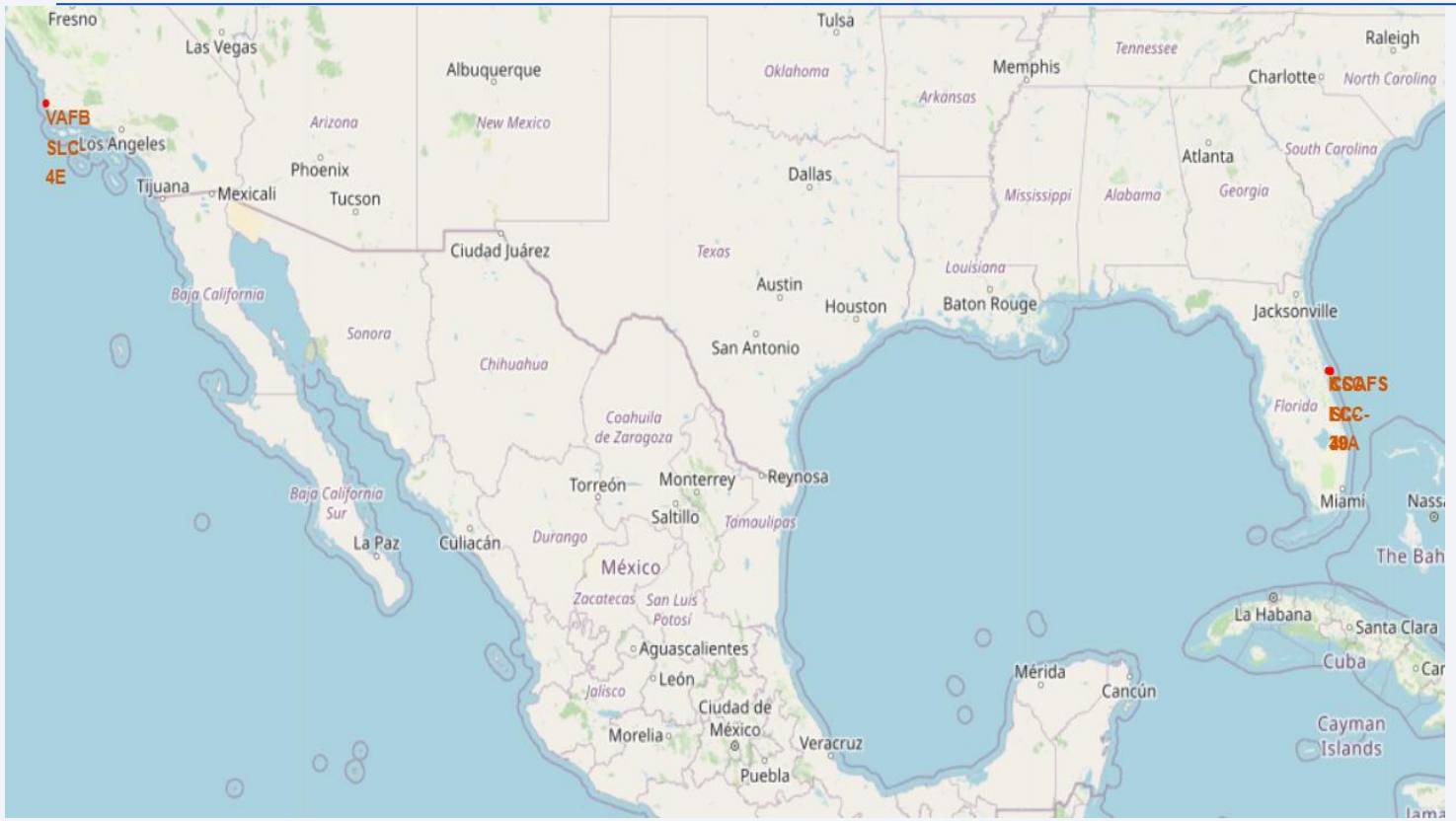
COUNT	landing__outcome
10	No attempt
5	Failure (drone ship)
5	Success (drone ship)
3	Controlled (ocean)
3	Success (ground pad)
2	Failure (parachute)
2	Uncontrolled (ocean)
1	Precluded (drone ship)

- The count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending(ranked) order are shown.

Section 4

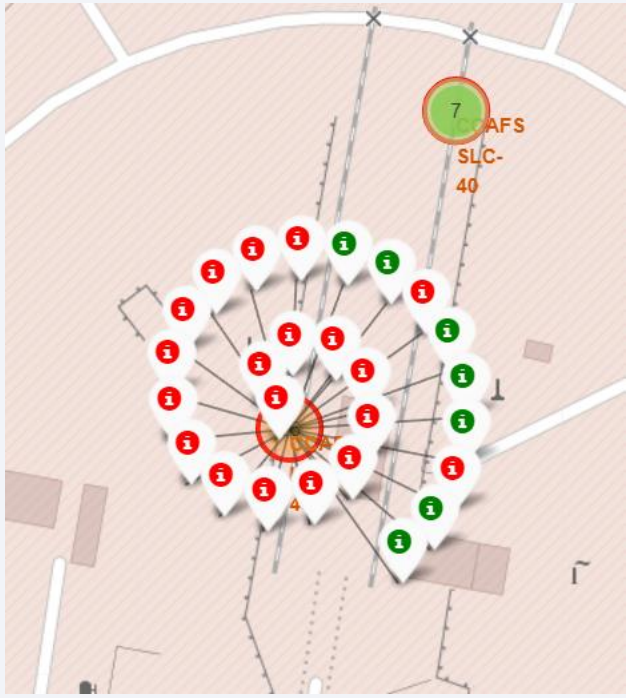
Launch Sites Proximities Analysis

Location of launch sites



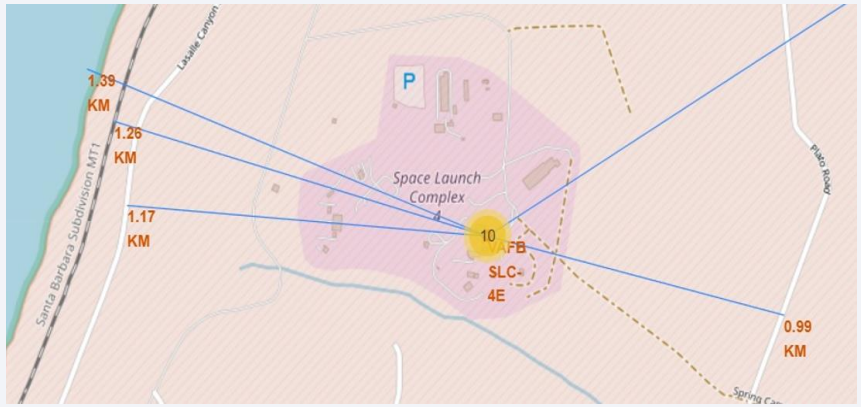
VAFB SLC-4E is located on the western part and all the other 3 on the eastern part, their zoomed in location is shown, where it is clear that CCAFS SLC-40 and CCAFS LC-40 lies in closer proximity.

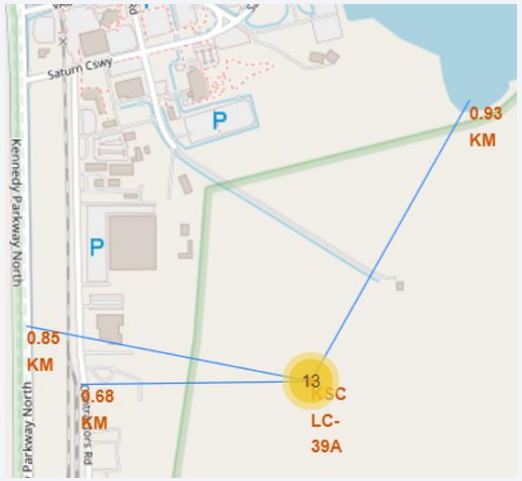
Success/failed launches for each site

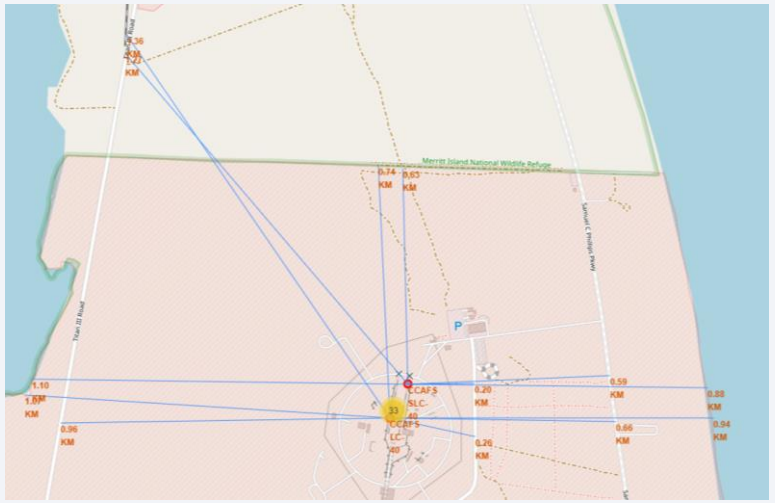


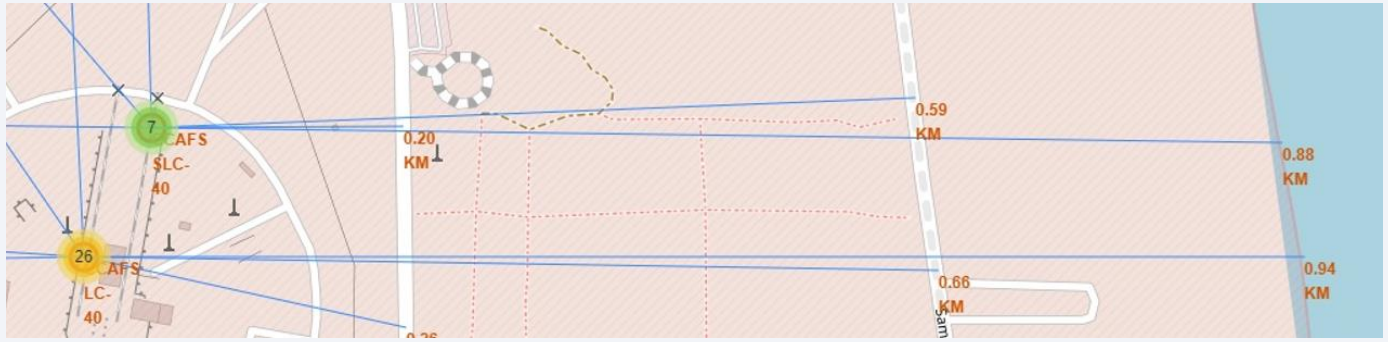
For each site the number of success(green) and failure(red) marks are marked to identify the success rate of each site on the map.

Close proximities to launch sites





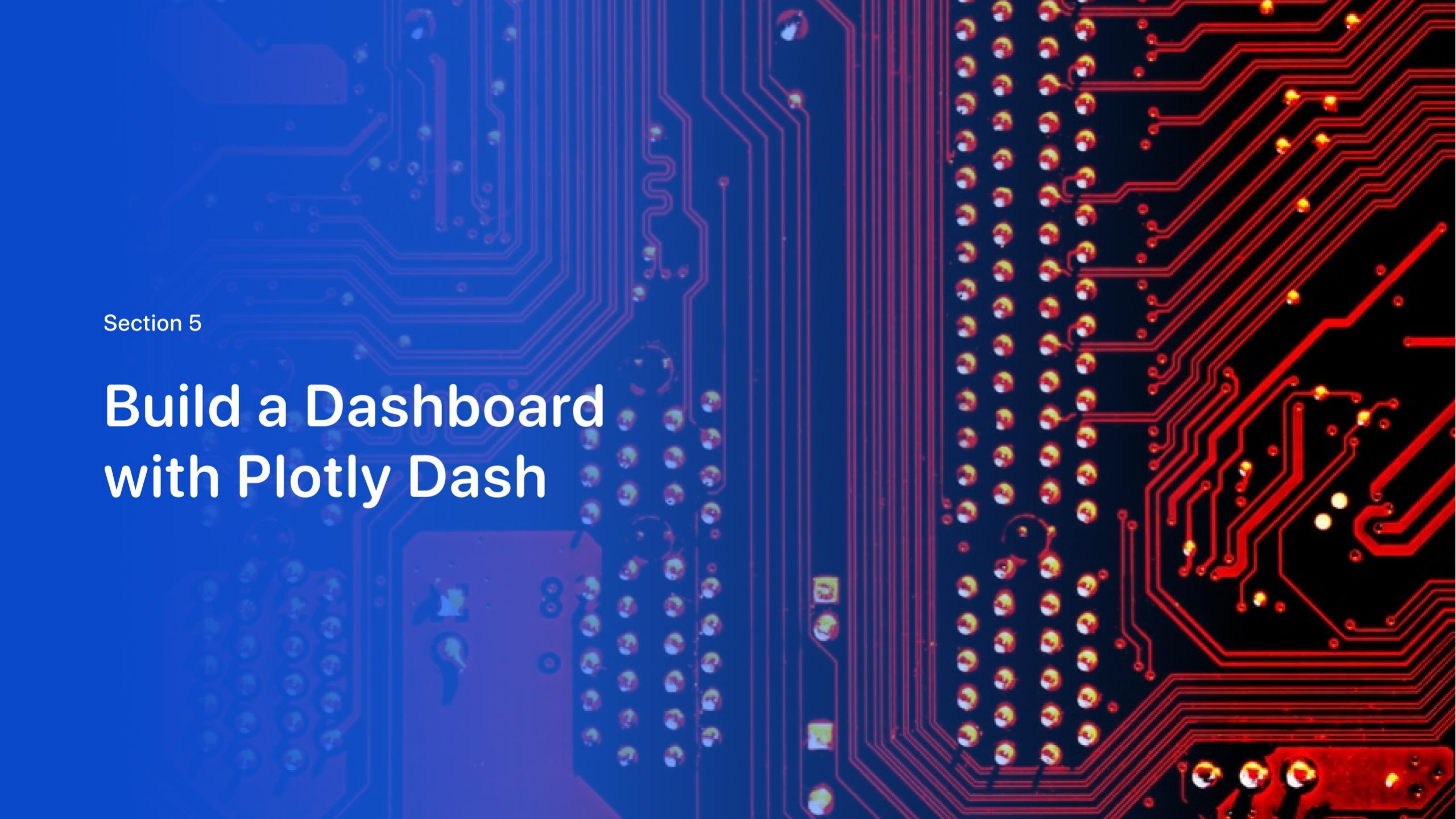




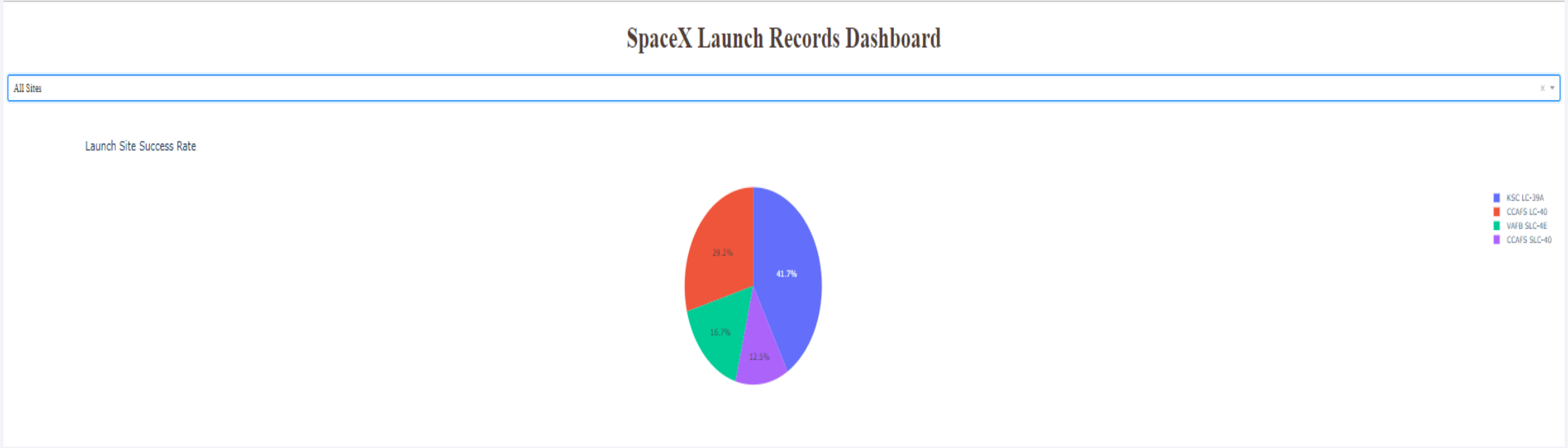
- All sites are located near coastline (around 1km) and no site is near any city
- One or two Railway roads and highway roads are seen around the sites(within 1km).

Section 5

Build a Dashboard with Plotly Dash

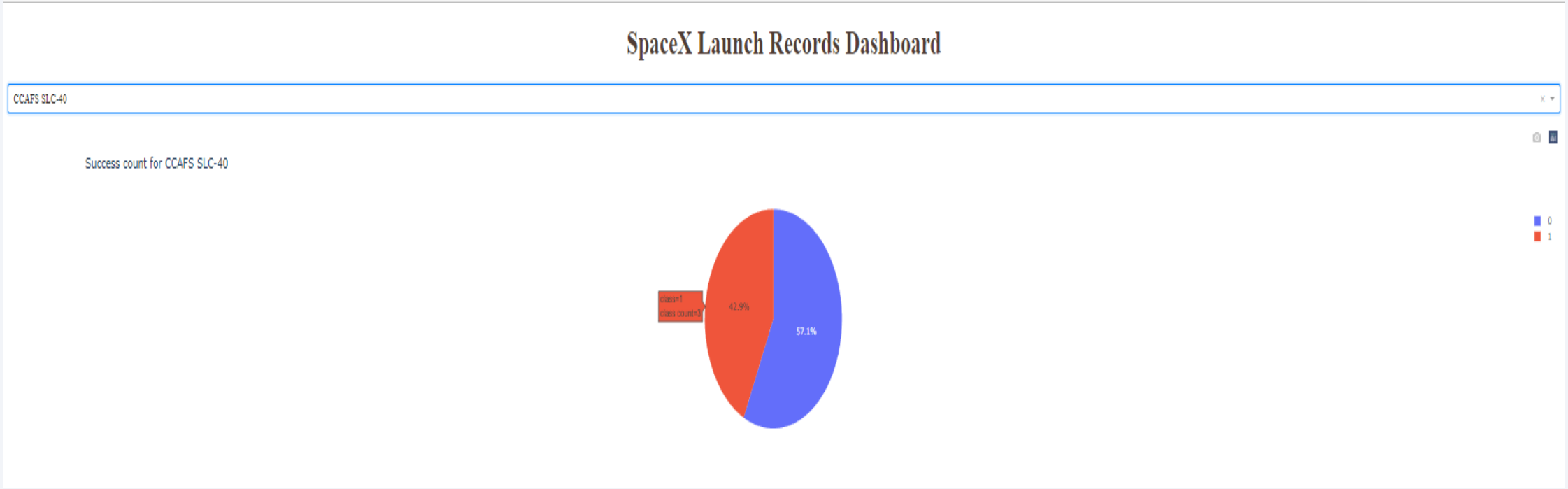


launch success count of all sites



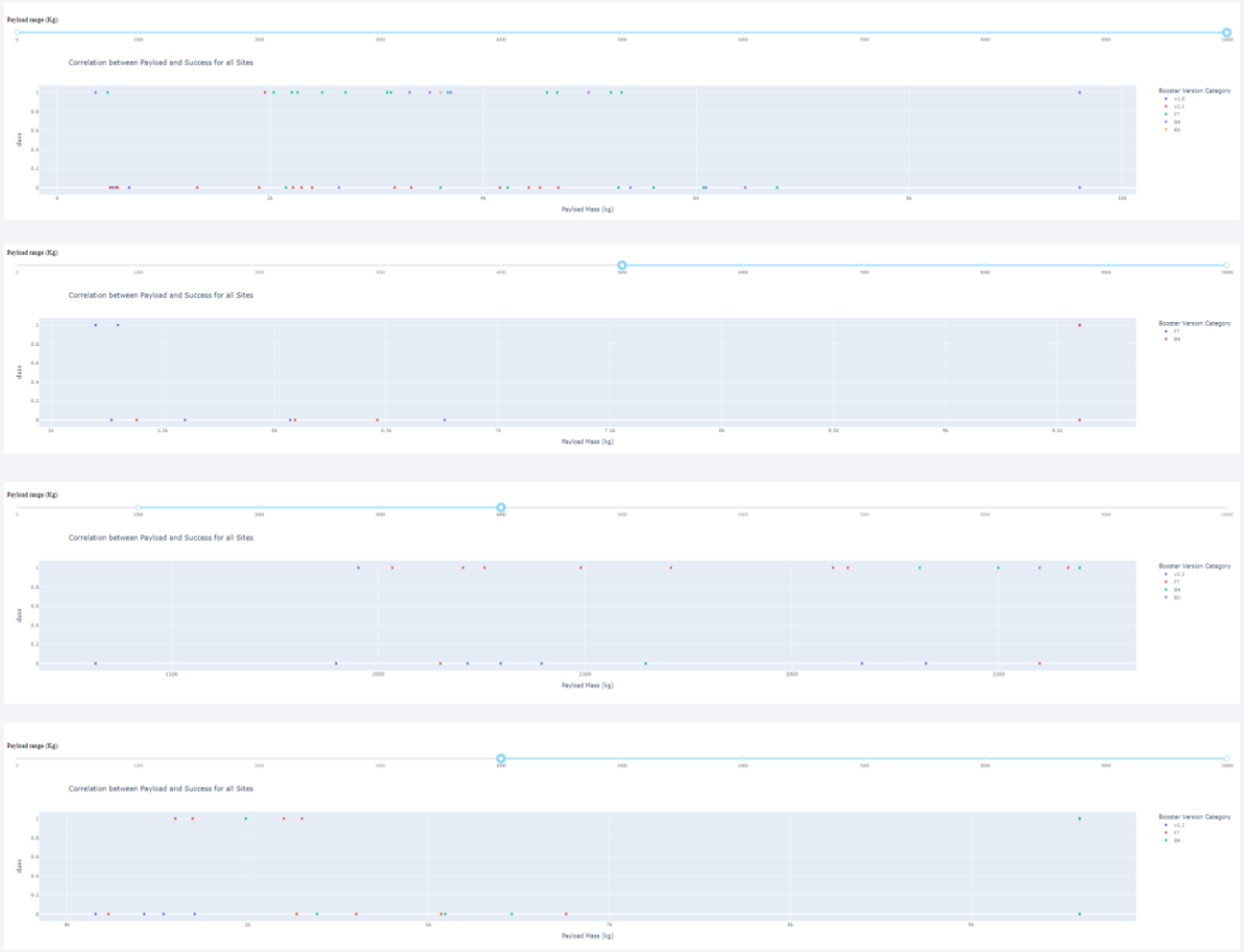
The KSC LC-39A has the most success count (10), whereas CCAFS SLC-40 has the lowest success count (3).

Launch site with highest success rate



the launch site with highest launch success ratio is CCAFS SLC-40 with success rate 42.9% with count=3(success ratio is 3/7)

Correlation between payload mass and success



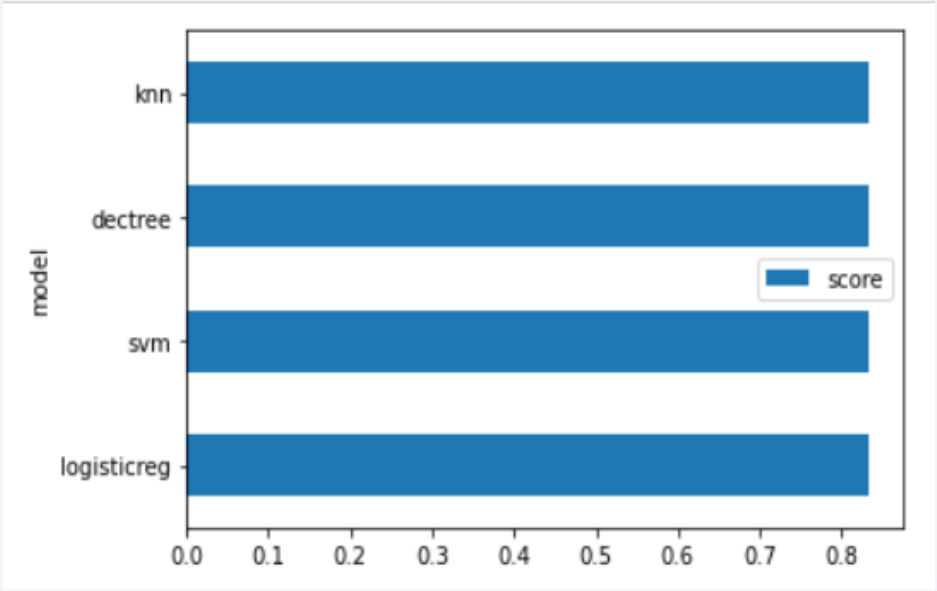
- For Heavy payload mass (5000 to 10,000 kg) the success counts are very low, when compared to low mass (1000 to 4000kg).
- Only v1.1, FT, B4 where used for above 4000kg payload mass.
- More success when mass between 1000 to 4000kg by FT booster version category.

Section 6

Predictive Analysis (Classification)

Classification Accuracy

- All four models has the same accuracy score(83.33%)
- The four models are
  - ✓ Logistic regression
  - ✓ Decision tree
  - ✓ Support vector machine
  - ✓ K- nearest neighbor



Confusion Matrix

- 12 landed outcomes are correctly predicted and 3 failed outcomes are correctly predicted.
- 3 more failed outcomes are falsely predicted.
- major problem is false positives.

Conclusions

- The success rate of spacex launch is proportional to time in years, hence the probability of success is even more in the upcoming years
- We can see that KSC LC-39A has the most successful launches but success ratio is highest in CCAFS SLC-40
- Orbit GEO, HEO, SSO, ES-L1 has best success rates
- Payload mass, booster version and orbit combined has a great effect on the outcome of landing

Appendix

- Python notebooks:
    - github repo
- SQL data:
    - Spacex DataSet
- Web scrapping:
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- SPACEX REST API:
    - Spacex rest api

Thank you!