IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Aleem Noorani
7DEC21

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

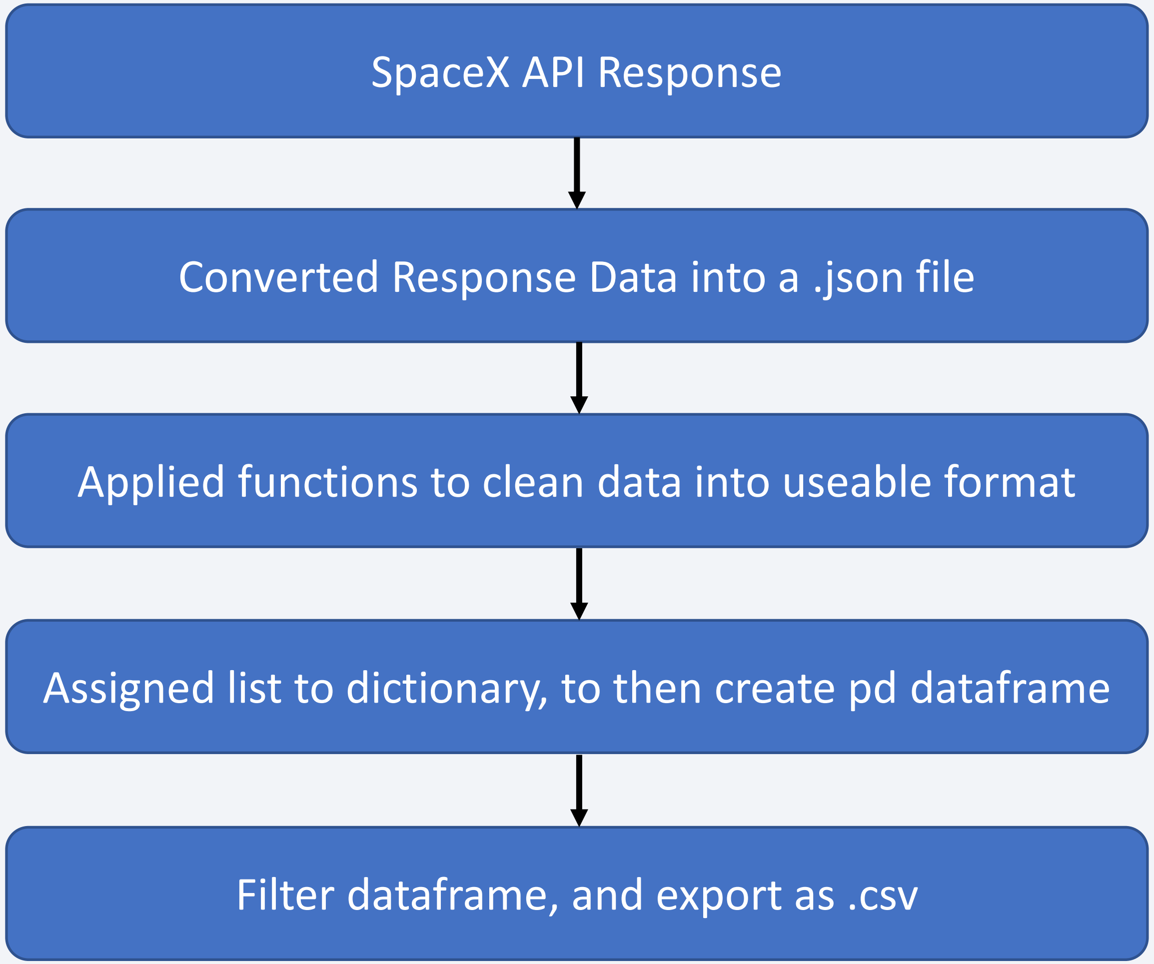
- Data collection methodology:
    - SpaceX Rest API
    - Wikipedia Web Scrapping
- Perform data wrangling
    - Processed with Python code
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

Data Collection

- Data from SpaceX Rest API.
    - Data published directly from SpaceX
- Webscrapping of Wikipedia
    - Used as data to test SpaceX data against

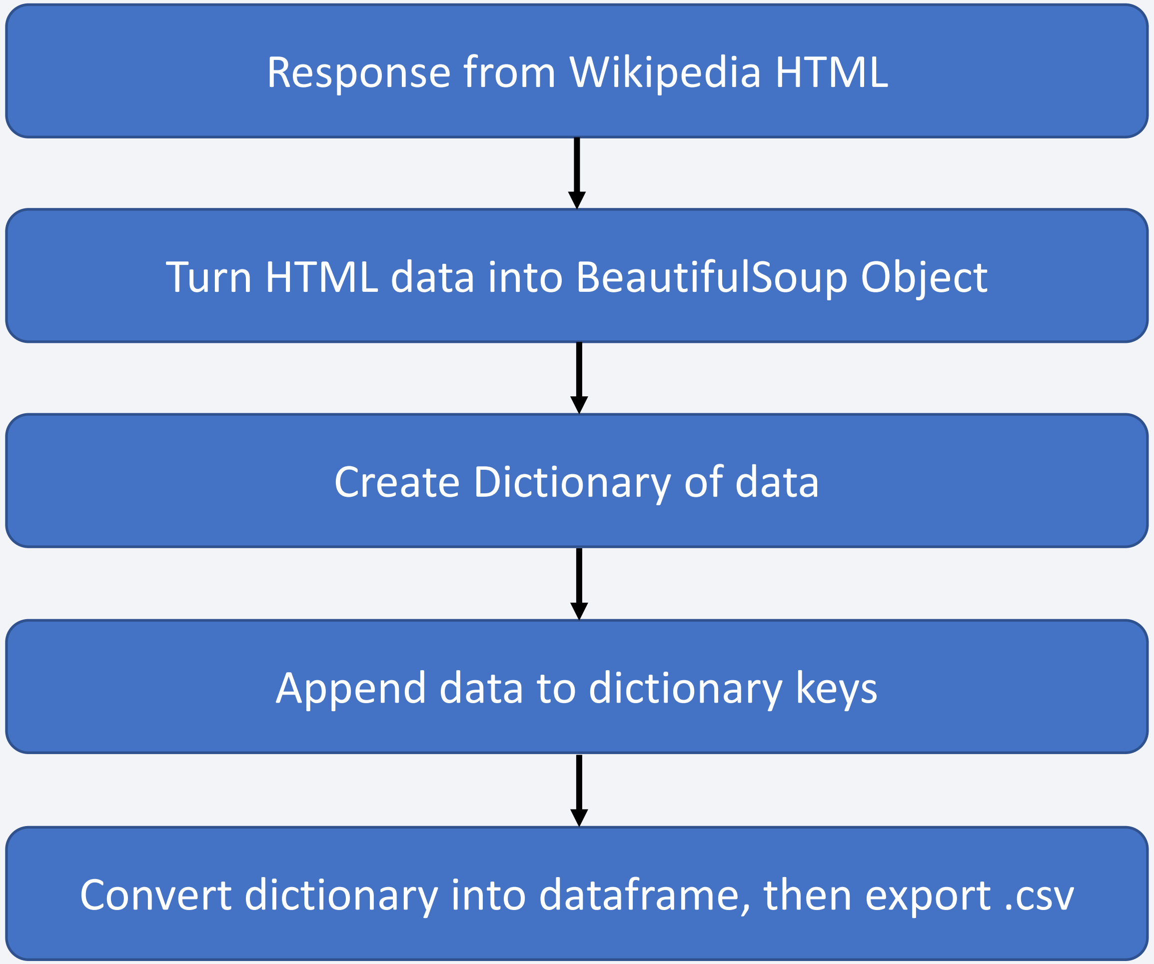
Data Collection – SpaceX API

- SpaceX API data was converted into useable filetype .csv
- GitHub URL of Files



Data Collection - Scraping

- BeautifulSoup used to collect data from Wikipedia
- [Github URL of File]()



Data Wrangling

- Data Wrangling was simple process

- Perform EDA on datasets
- Calculate the number of launch sites
- Calculate the number of orbits
- Calculate Mission outcomes
- Analyze success rate for landings
- Export to .csv

EDA with Data Visualization

- We created multiple charts
  - Flight vs Mass, Flight vs Launch Site, Mass vs Launch Site, Flight vs Orbit, Mass vs Orbit, Payload vs Flight, Payload vs Orbit
- We created SQL queries
  - Identify unique launch sites, average payload mass, total launches by NASA, names of boosters who succeeded, annualized success metrics
- We also created an interactive data map using Folium
  - Check out the Github link to utilize the maps

GITHUB URL of Python EDA

GITHUB URL of SQL EDA

GITHUB URL of Folium EDA

Build a Dashboard with Plotly Dash

- We also create a Dashboard with Plotly Dash
  - The interactions allow for anyone to see the relationship between range, payload mass, and launch sites.
  - Easy to use and see for people who aren't used to data analysis

GITHUB URL of Plotly Dash

Predictive Analysis (Classification)

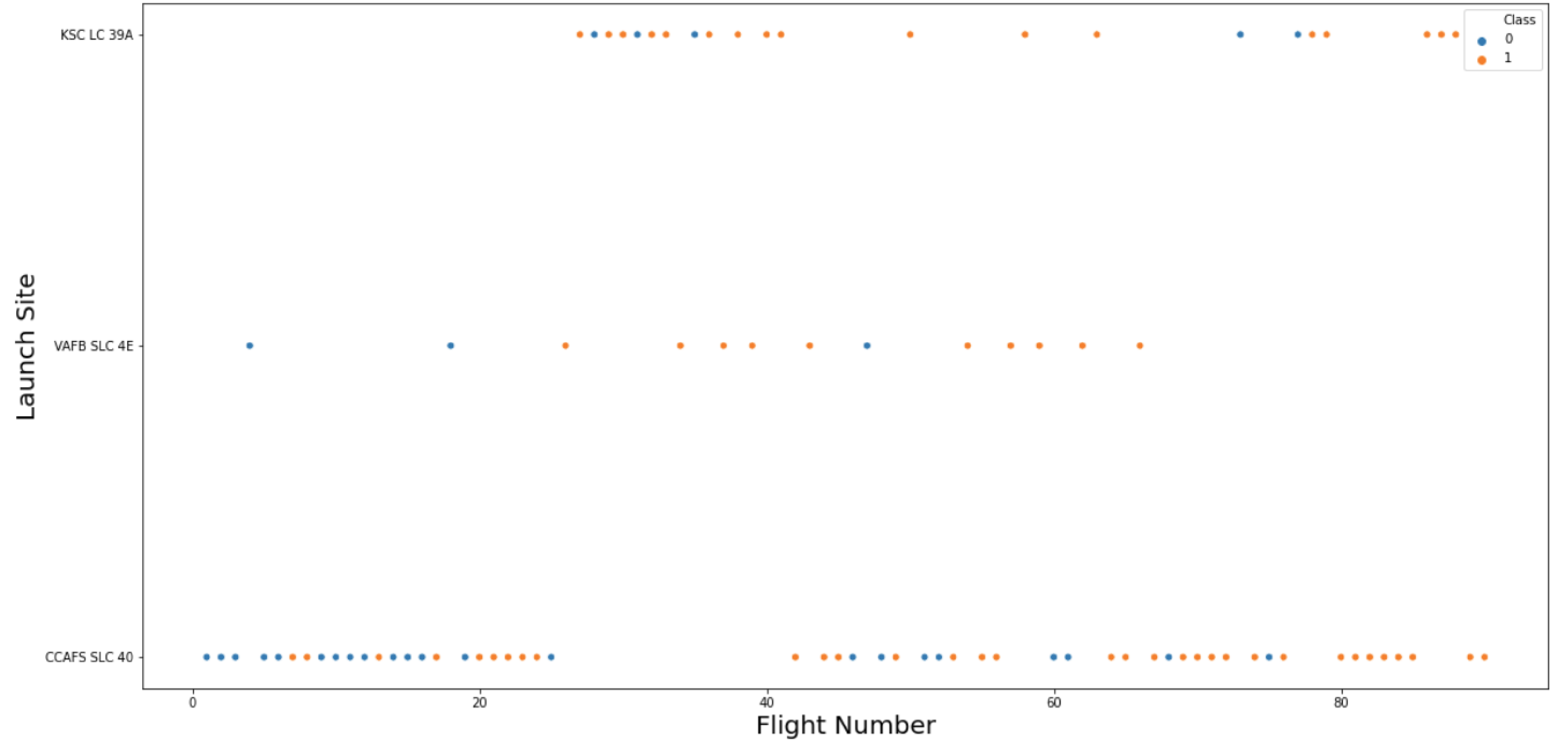
- We utilized Machine Learning predictive analysis to identify best outcomes and models to see whether a launch will be successful or not
- Models used
  - NumPy Array
  - GridSearchCV

GITHUB URL of Plotly Dash

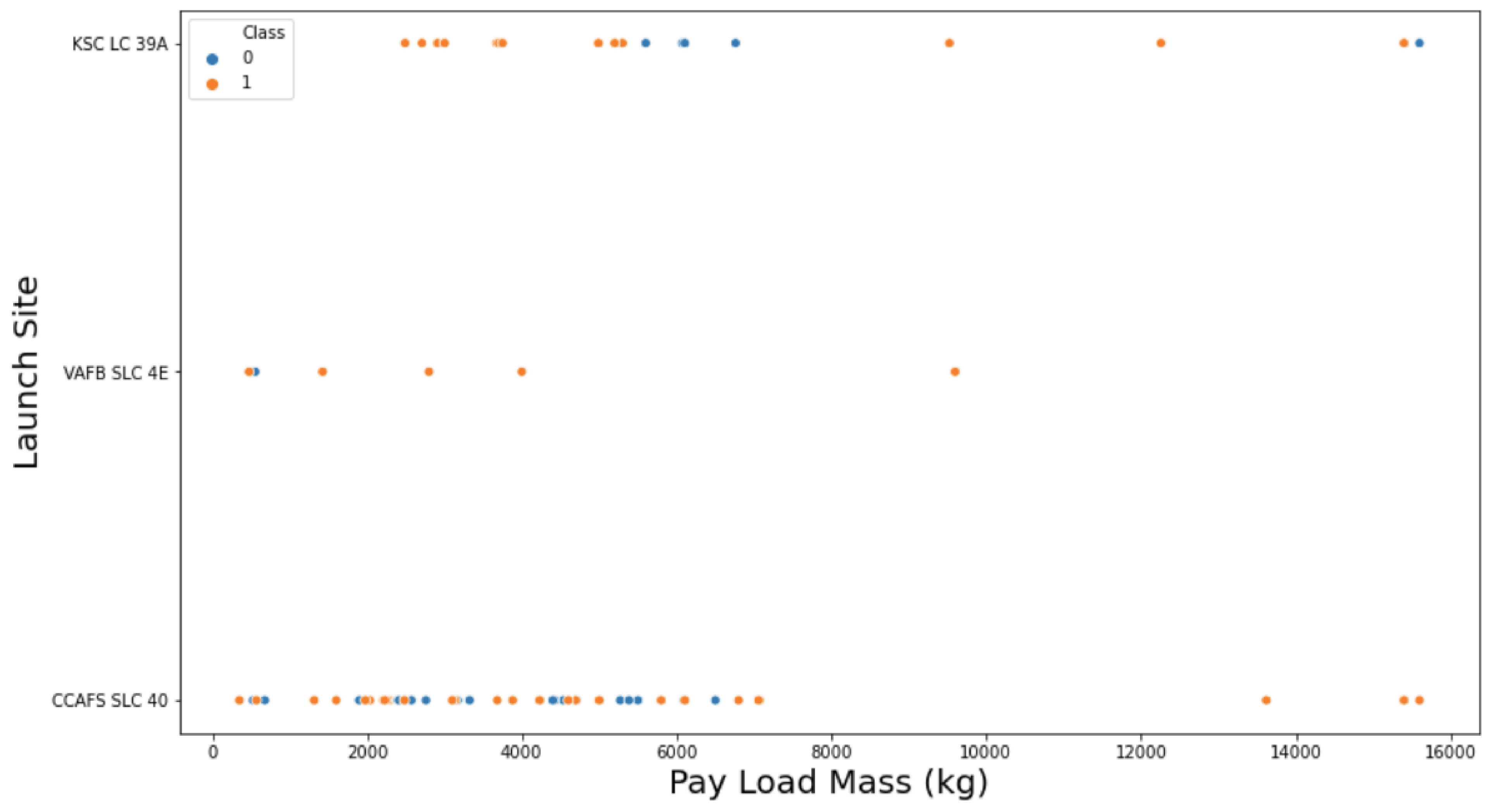
Section 2

Insights drawn from EDA

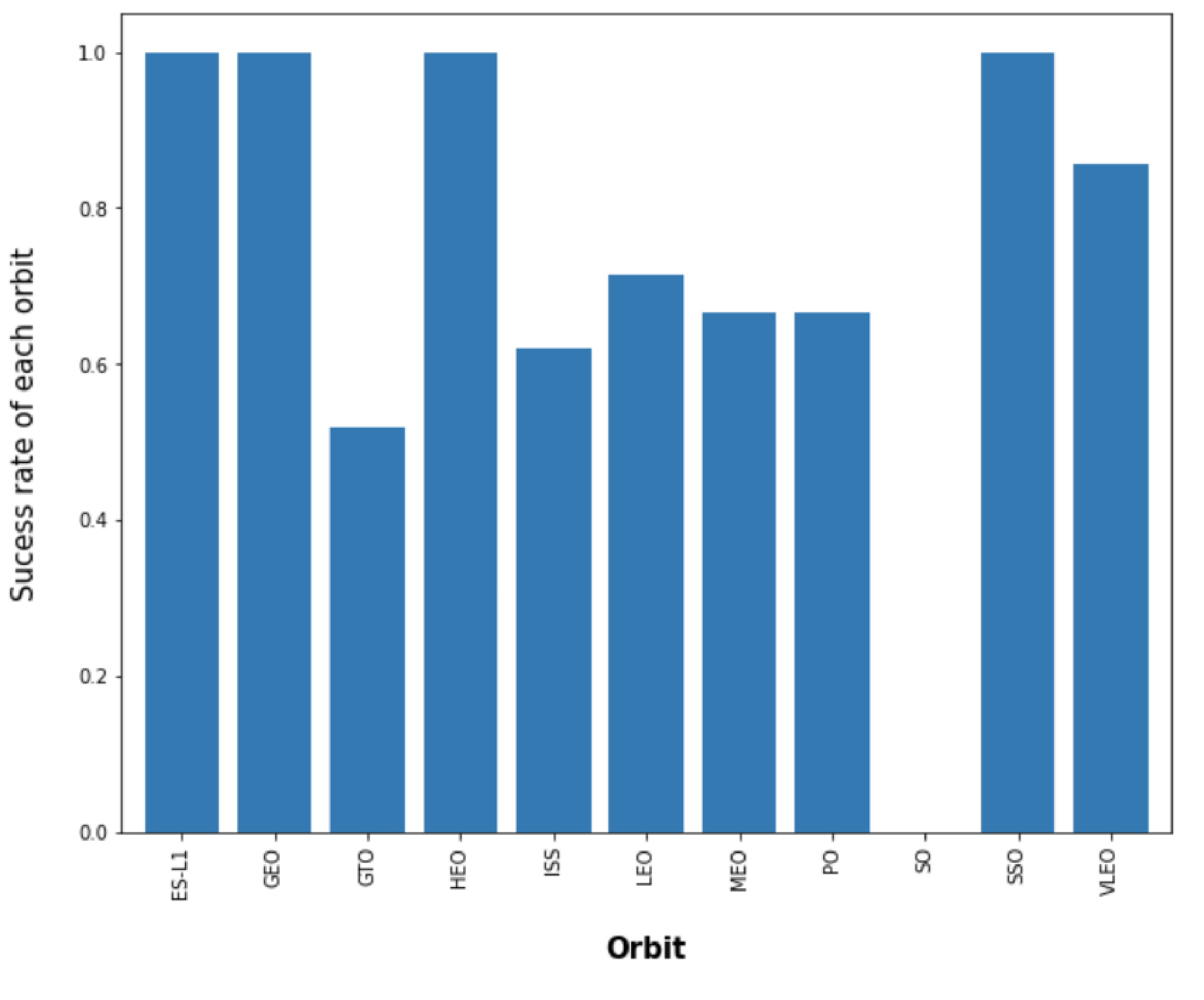
Flight Number vs. Launch Site



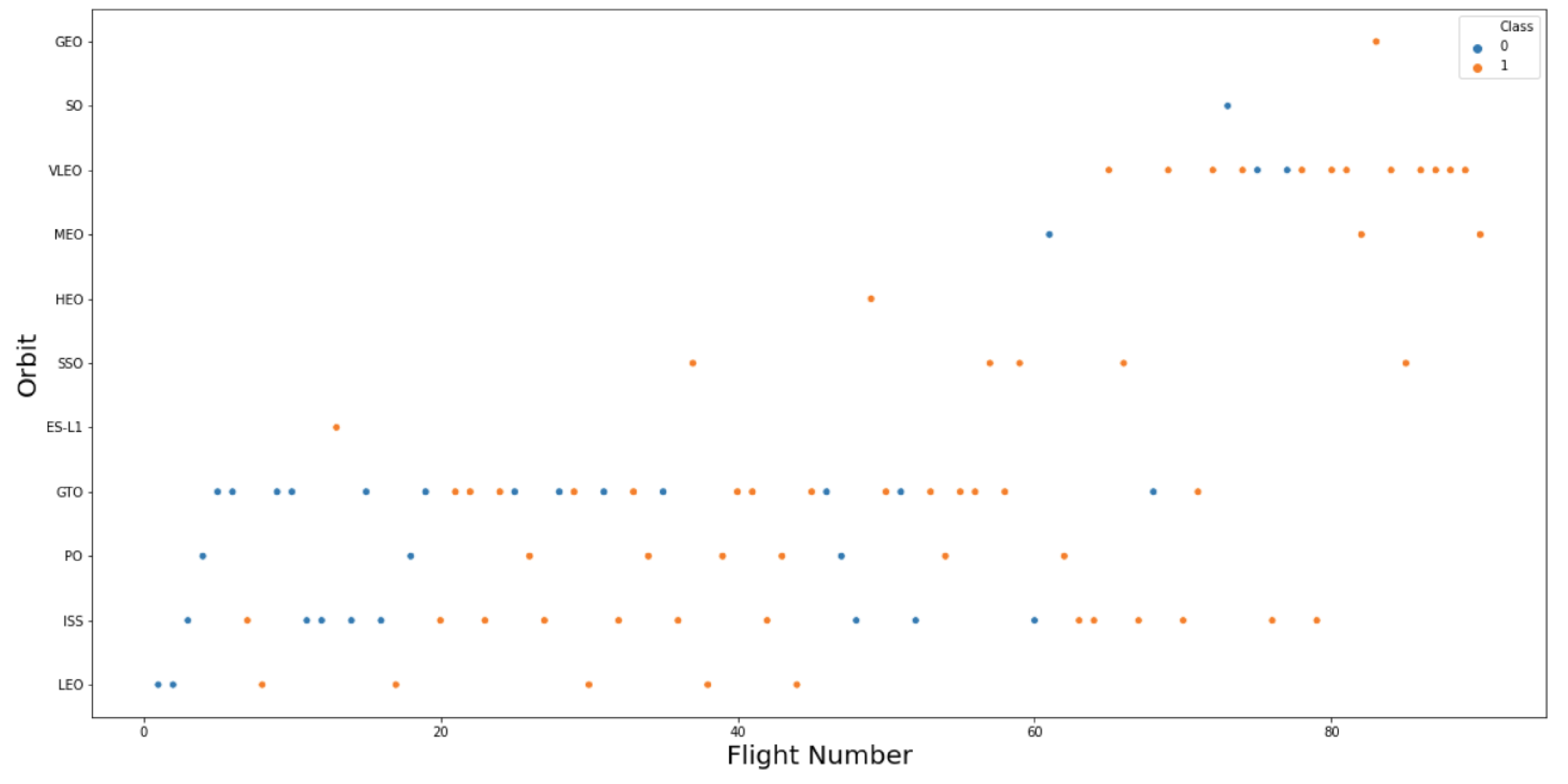
Payload vs. Launch Site



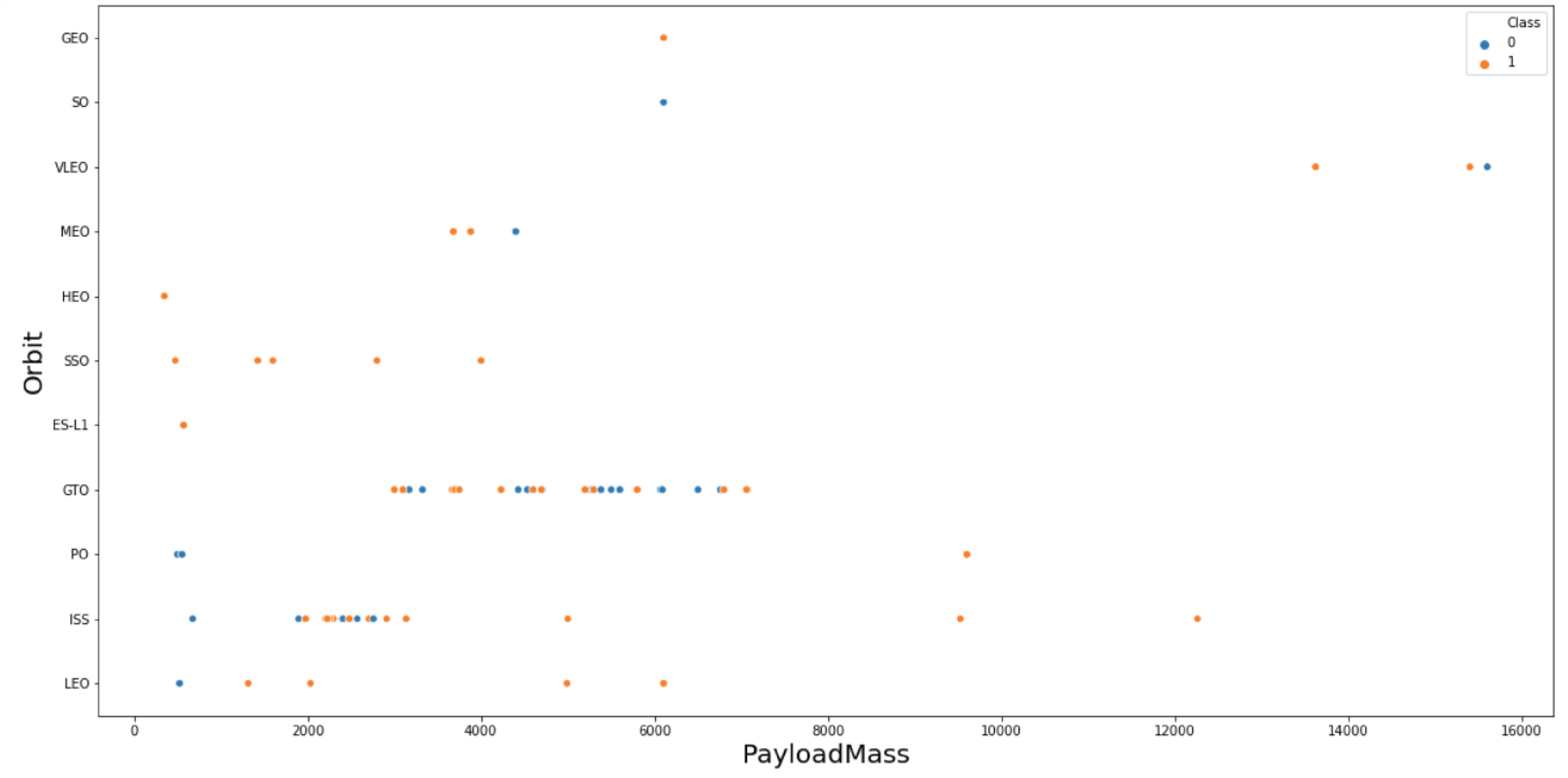
Success Rate vs. Orbit Type



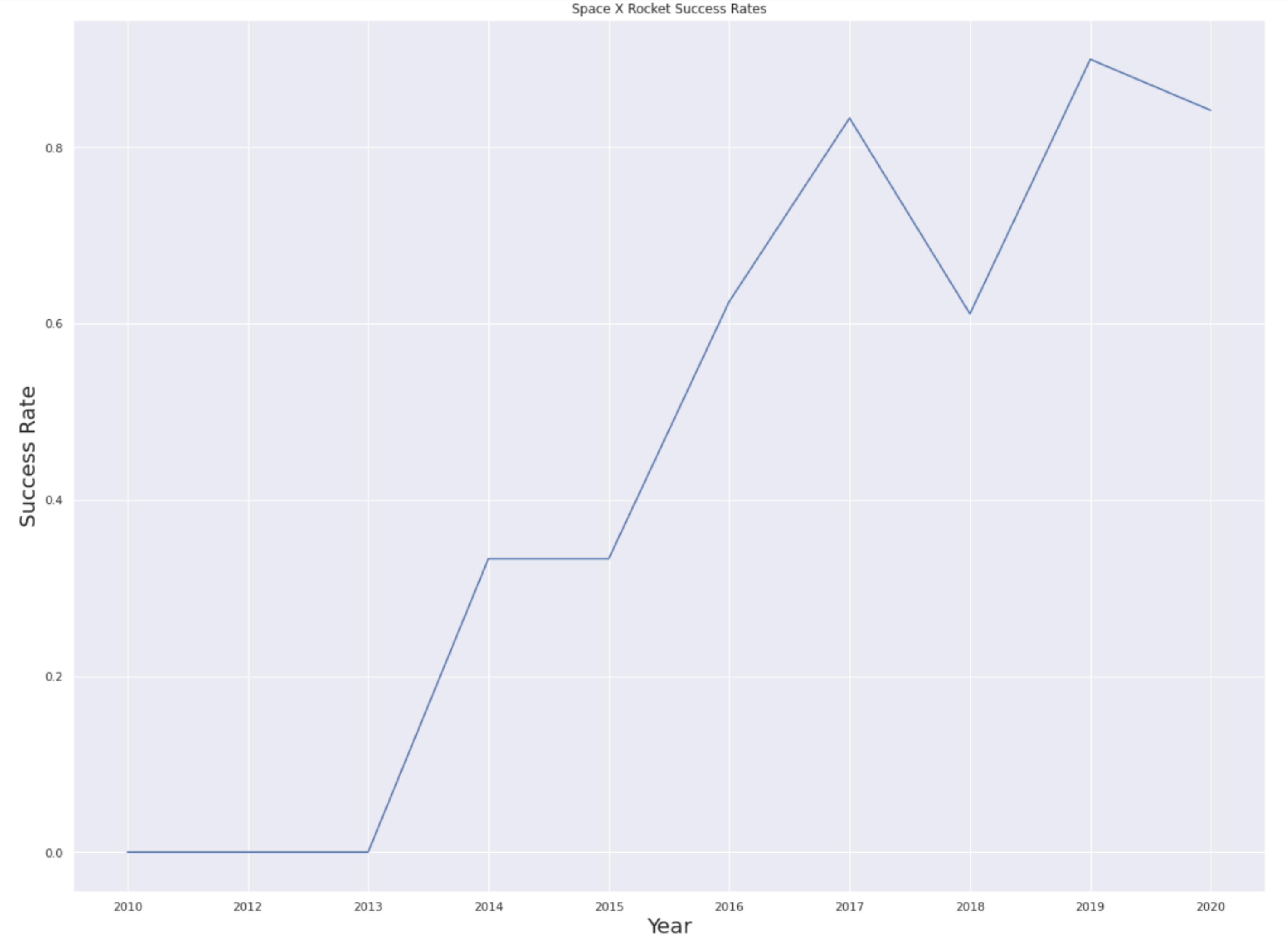
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

- Names of the unique launch sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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

- Calculate the total payload carried by boosters from NASA

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.
```

1
45596

Average Payload Mass by F9 v1.1

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

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

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

```
 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.
```

1
2928

First Successful Ground Landing Date

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

```
In [16]: %sql SELECT DATE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)' ORDER BY DATE ASC LIMIT 1

 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.

Out[16]:
```

DATE
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
In [9]: %sql Select  BOOSTER_VERSION, PAYLOAD_MASS__KG_ as PAYLOAD FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_

 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.
```

Out[9]:

booster_version	payload
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

List the total number of successful and failure mission outcomes

```
In [10]: %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'

 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.

Out[10]:
```

1
100

Boosters Carried Maximum Payload

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

Task 8

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

In [11]:
```
%sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.
```

Out[11]:

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

Task 9

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

In [12]:
```
%sql Select DATE, BOOSTER_VERSION, LANDING__OUTCOME, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND YEAR(DATE)=
```

```
 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.
```

Out[12]:

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

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

Task 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

In [13]:
```
%sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc
```

```
 * ibm_db_sa://rmb71214:***@1bbf73c5-d84a-4bb0-85b9-ab1a4348f4a4.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:32286/BLUDB
Done.
```

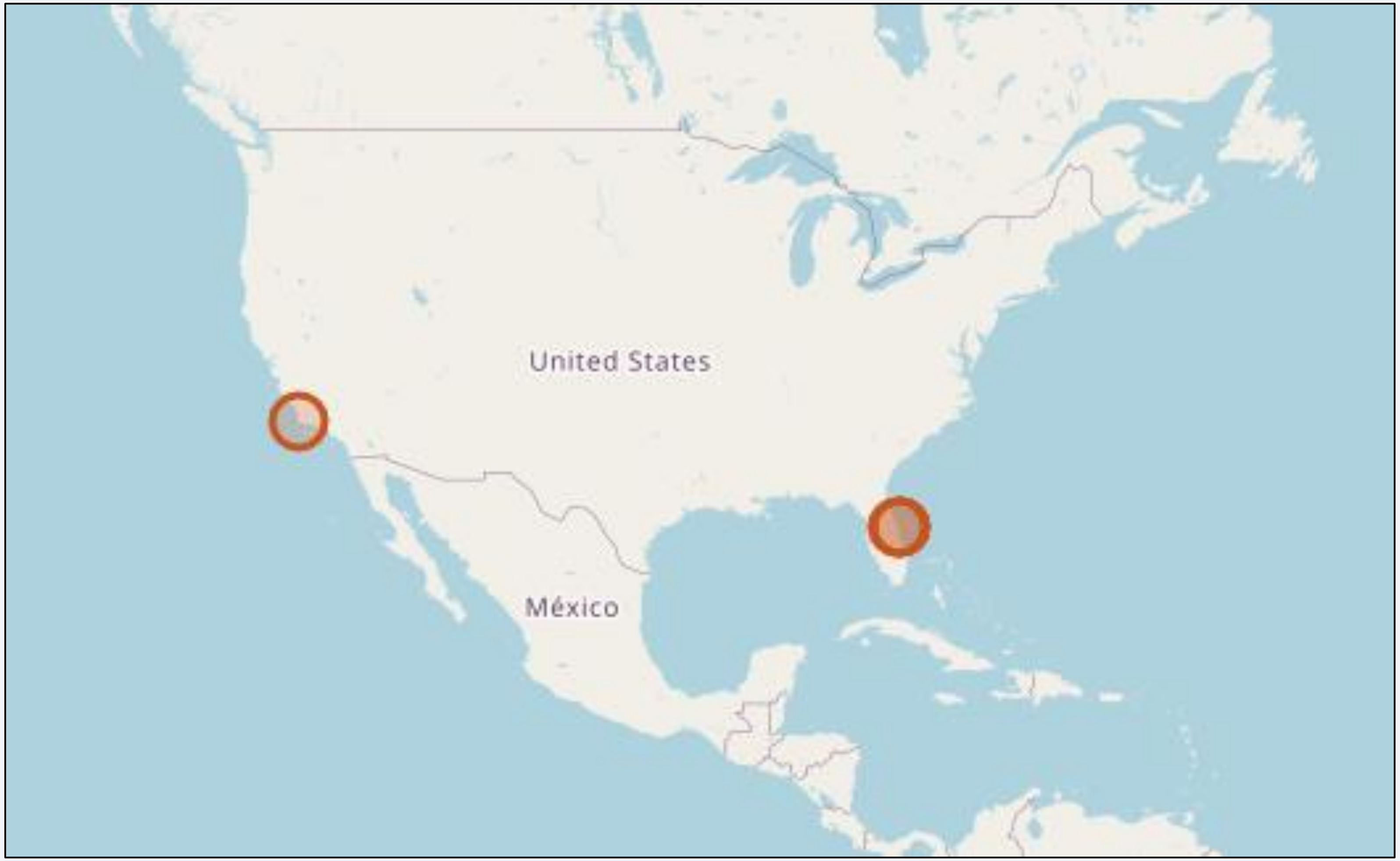
Out[13]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

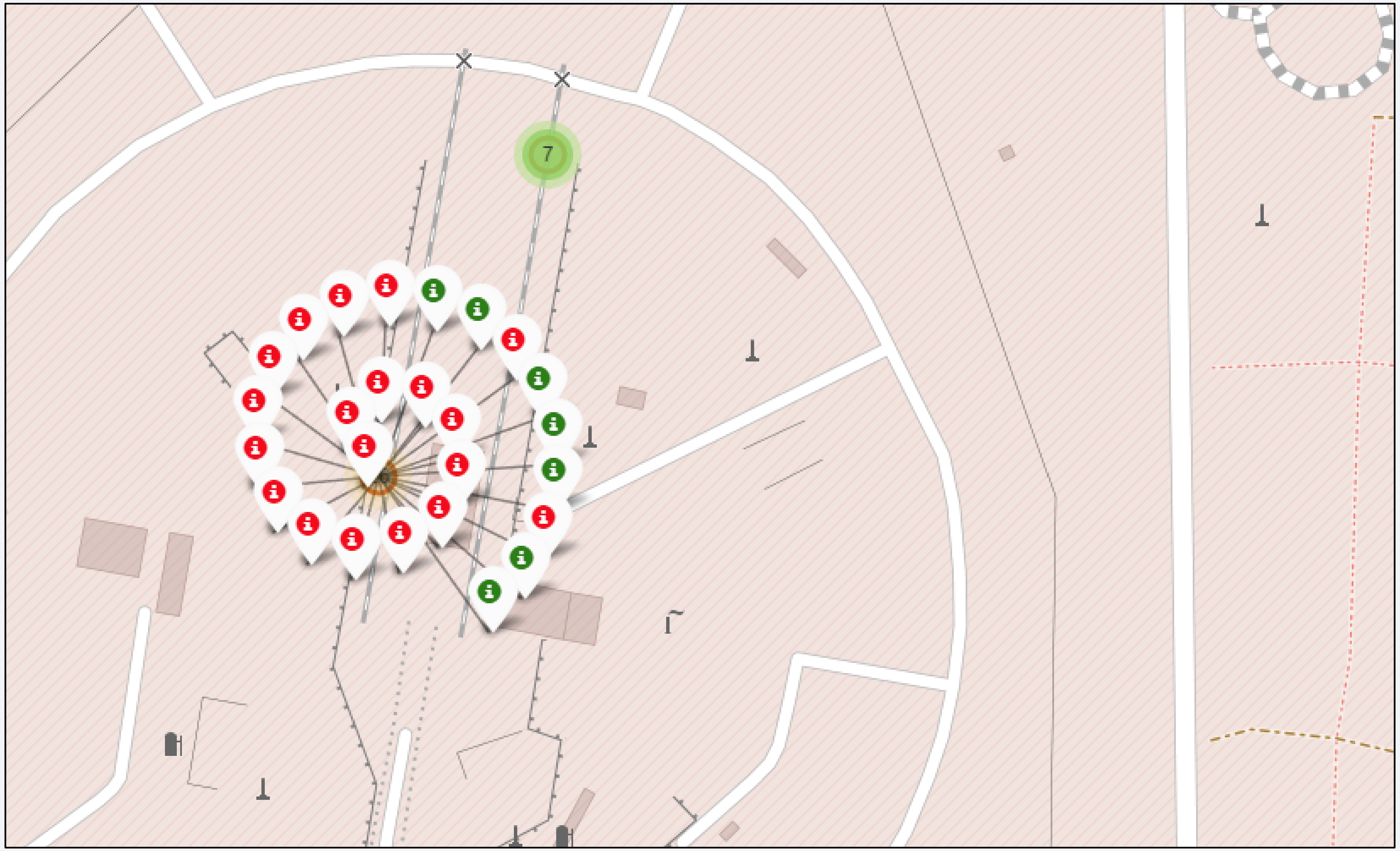
Section 4

Launch Sites Proximities Analysis

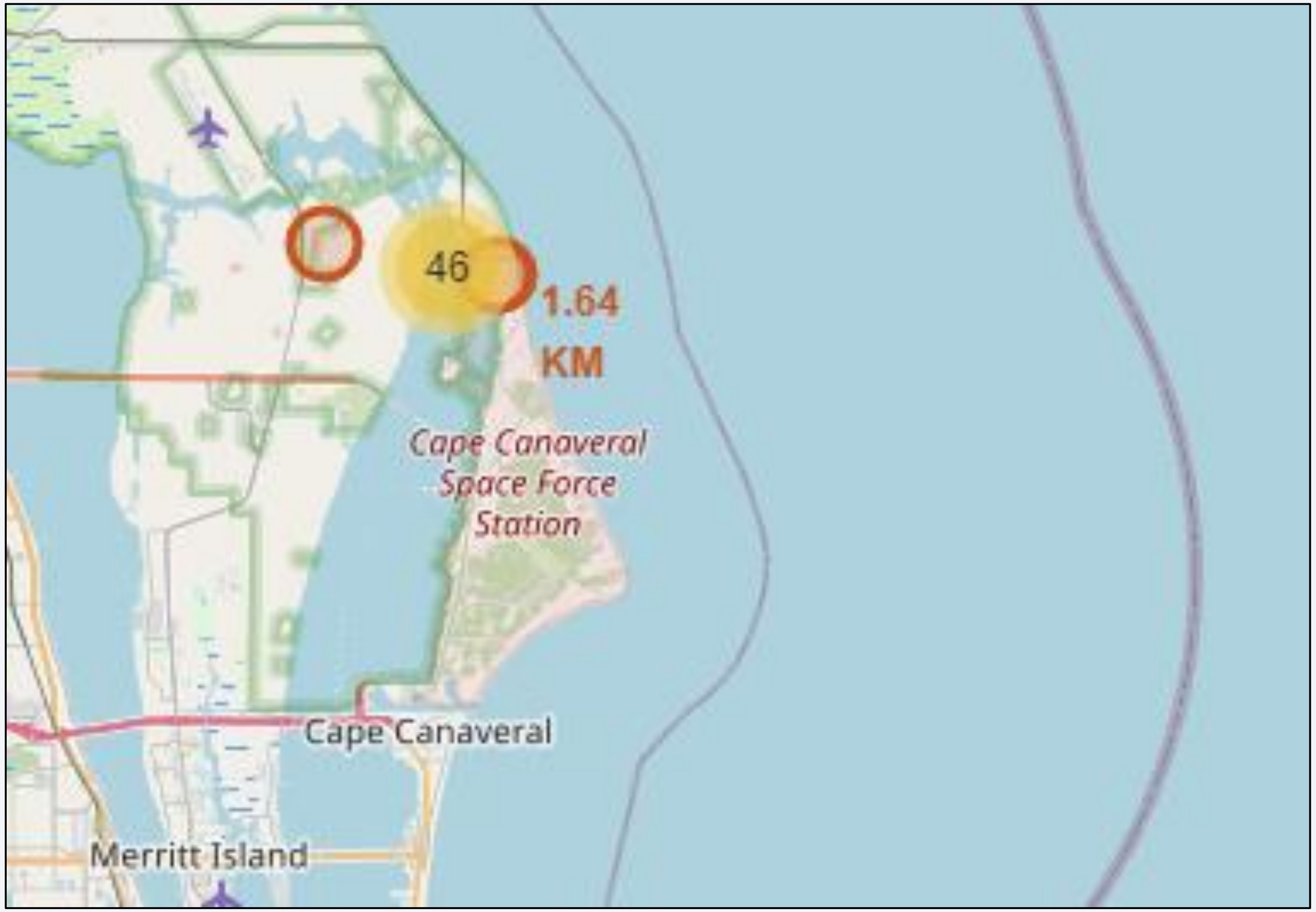
CONUS Launch Site Locations



CCAFS Launch Outcomes



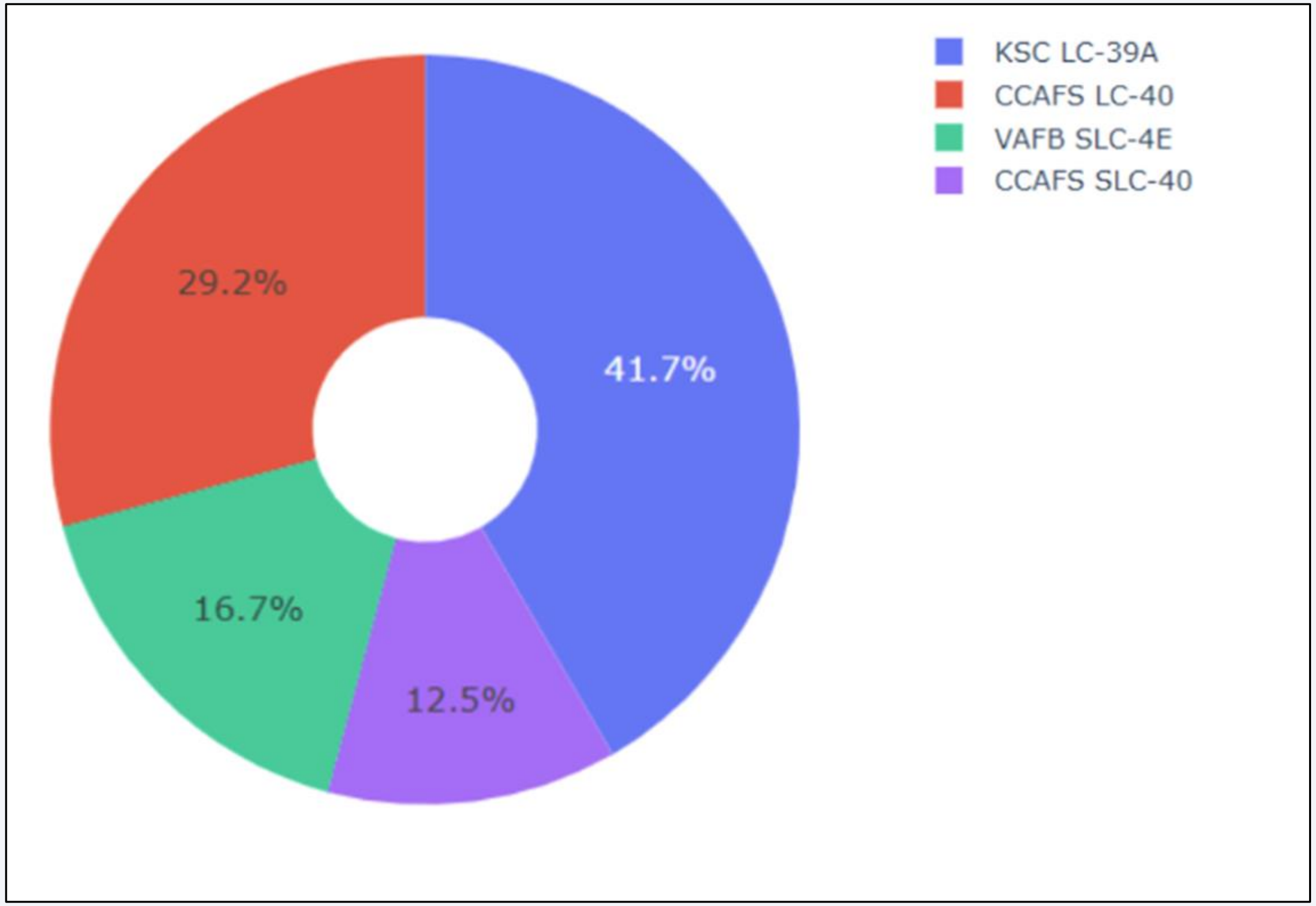
CCAFS launch site distance to coast



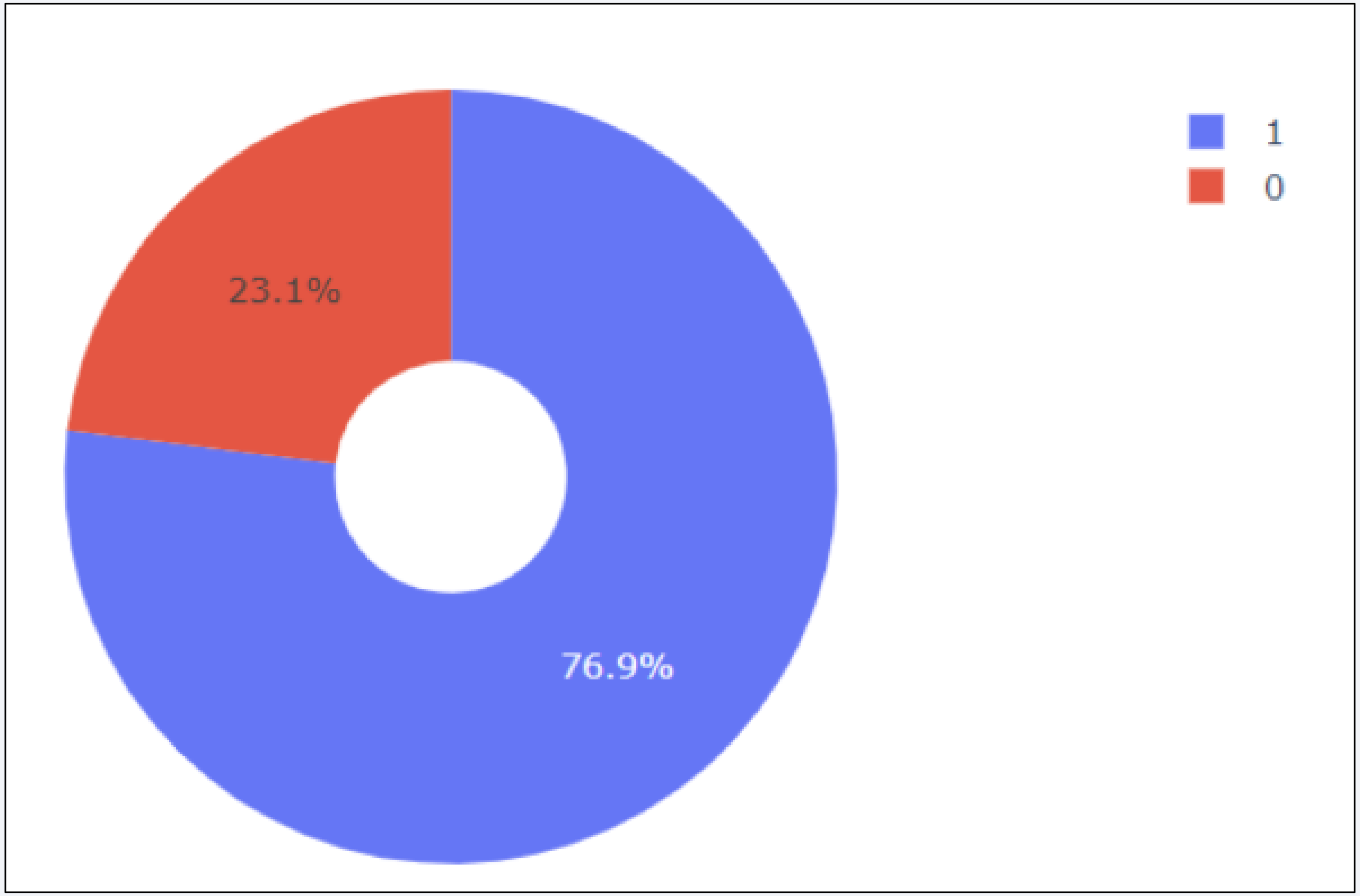
Section 5

Build a Dashboard with Plotly Dash

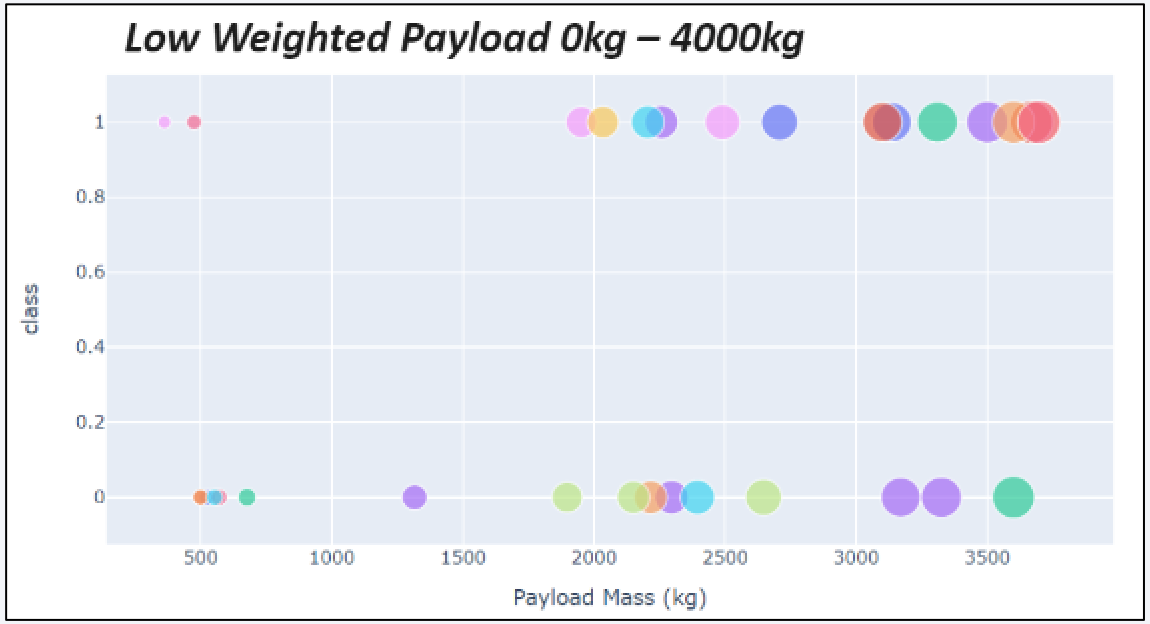
Launches By Sites

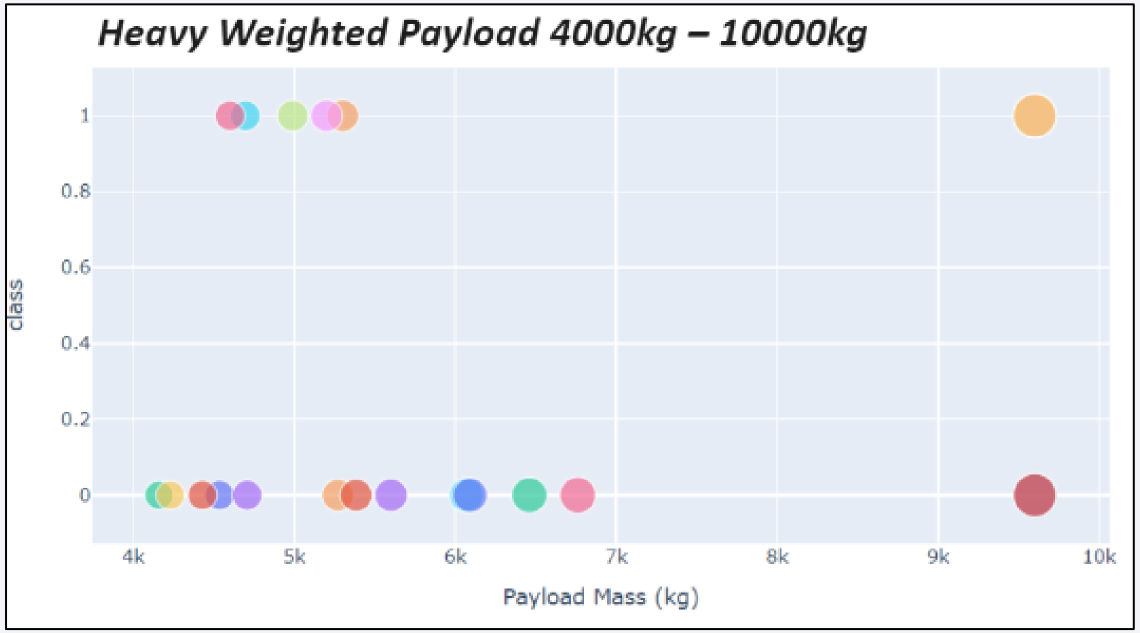


Most Successful Launch Site (KSC LC-39A)



Payload Success Rates based on Weight



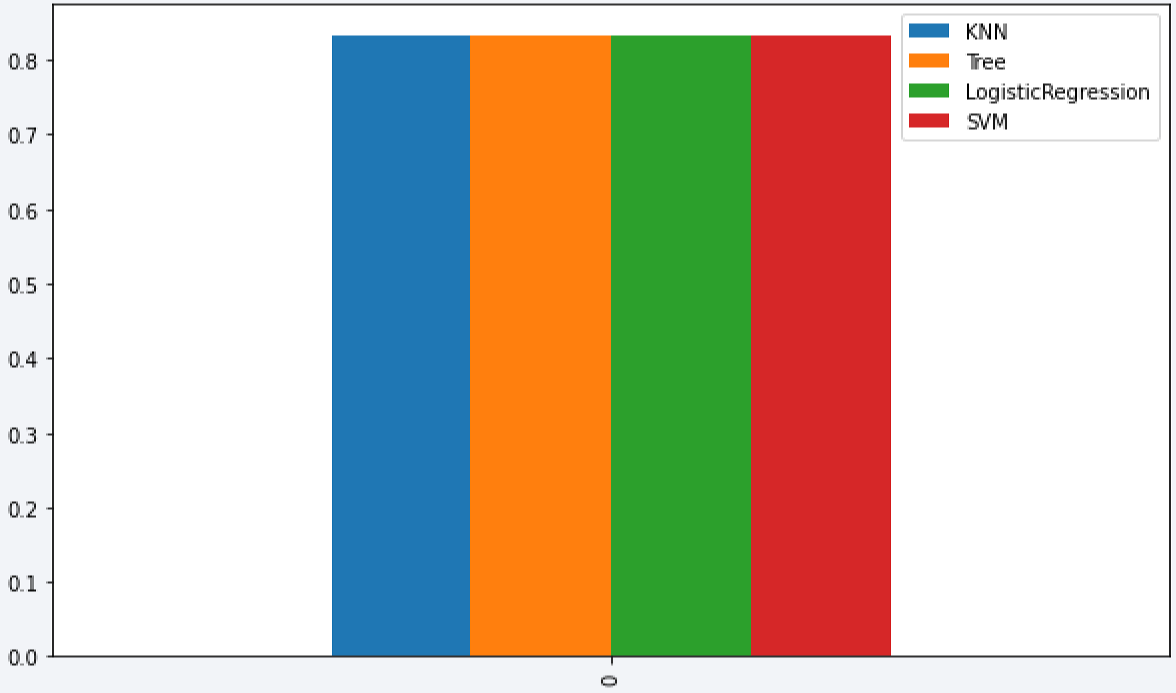


Section 6

Predictive Analysis (Classification)

Classification Accuracy

- Best Algorithm is Tree, but barely
- They are all so close that it doesn't really matter



Confusion Matrix

- Confusion matrix showing the Tree which allows us to distinguish between type 1 and type 2 errors

Conclusions

- Tree was the best algorithm for prediction
- Low weight is better than heavy weight payload
- Launches are getting more successful with time
- GEO, HEO, SSO, and ES-L1 are the best orbits
- CCAFS is the best launch site

Thank you!