IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sujit Kumar Pradhan
10-08-2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using SpaceX API and Web scraping
  - Exploratory Data Analysis using SQL, Pandas and Matplotlib
  - Data Visualization using Seaborn and Folium
  - Interactive Visual analytics and Dashboard using Plotly
  - Predictive Analysis using machine Learning Algorithms
- Summary of all results
  - Creating both Visual and Data driven analytics using the above methodologies to find out the answers to the Problem Statement

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.
  - Determining whether the first stage will land and further determination of the cost of the same and using the above information to showcase SpaceX’s success and cost effectiveness against other companies’ bid.
- Problems you want to find answers
  - Whether the Falcon 9 first stage will land successfully ?
  - Whether we can determine the cost of launch ?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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

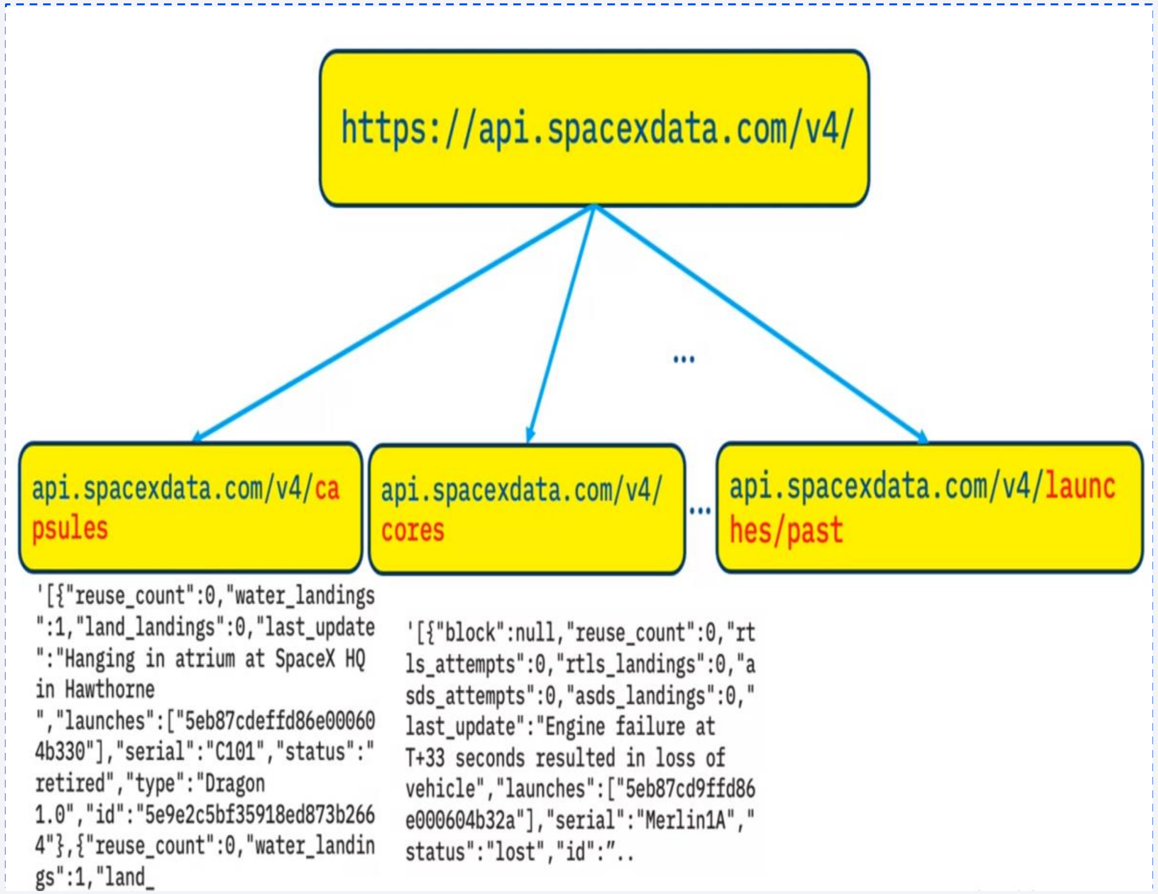
Data Collection

- Data was first collected from SpaceX API through a get request to it done by defining a series of functions that would extract information using identification numbers in the launch data and then requesting launch data from the URL.

- Finally to make the requested JSON results more consistent, they were decoded using .json() and turned into a Pandas dataframe using .json_normalize().

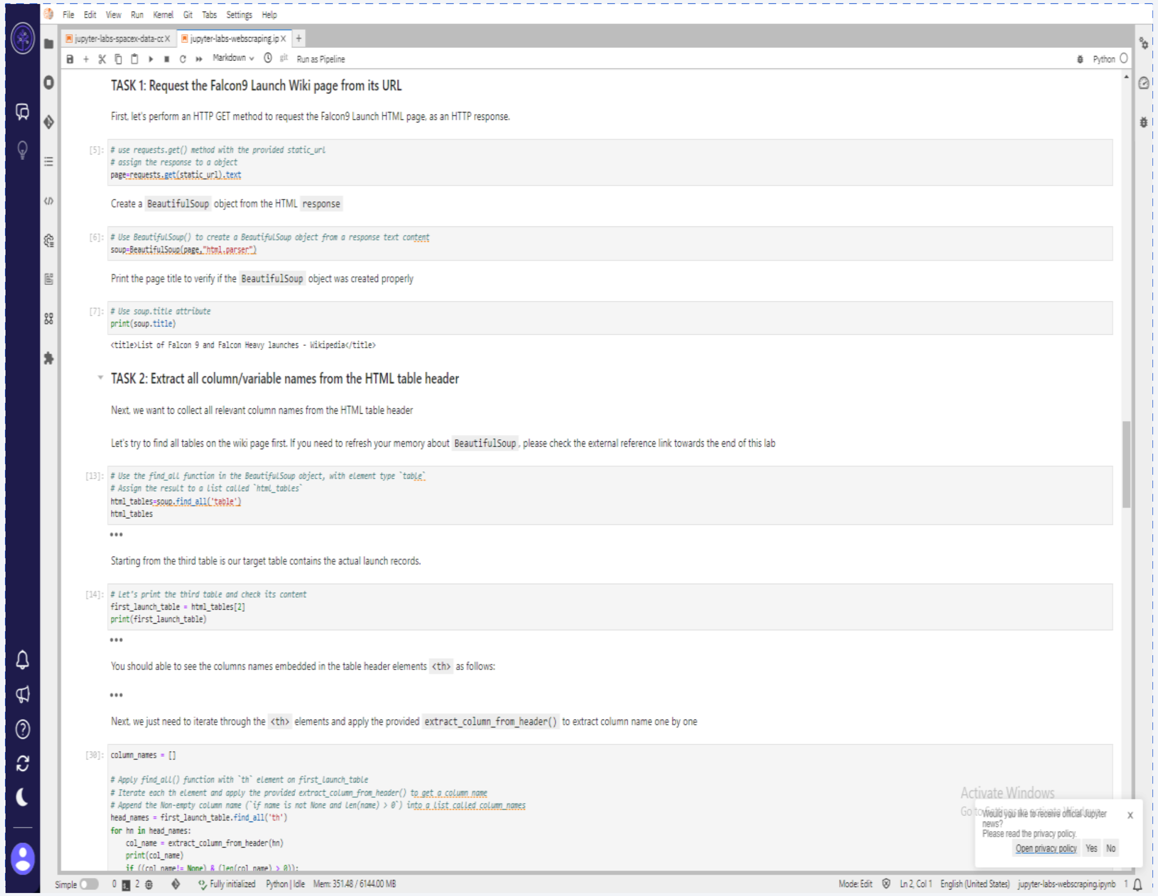
Data Collection – SpaceX API

- SpaceX launch data gathered from SpaceX REST API which provides us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcomes
- URL :- https://api.spacexdata.com/v4/launches/past



Data Collection - Scraping

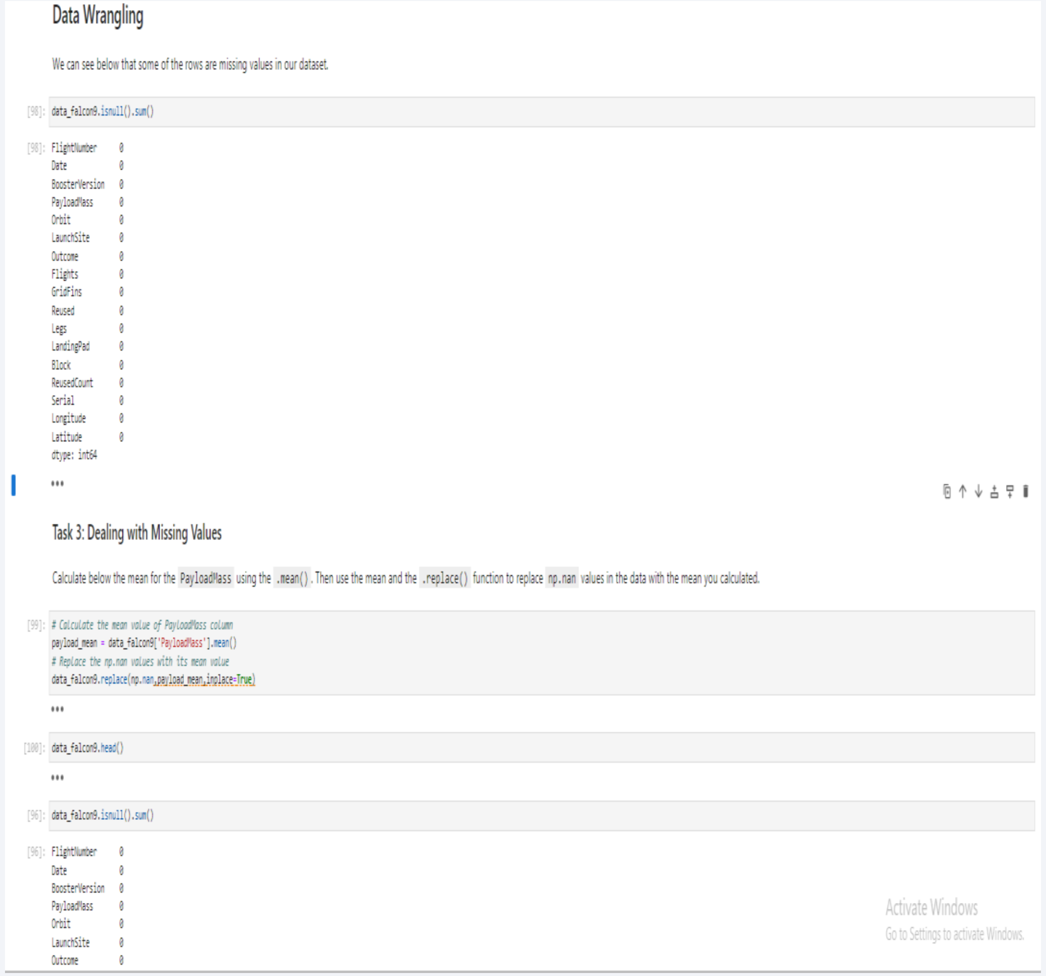
- Falcon 9 launch records are scraped with the help of `BeautifulSoup`object to Extract a Falcon 9 launch records HTML table from Wikipedia and the table is parsed and converted to a Pandas data frame
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

- Once the dataframe is created, data is filtered only to keep the Falcon 9 launches, missing data values in respective columns dealt with.
- Once the missing values are dealt with, a final check is done through isnulll().sum() in order to ensure that we have no missing values in our dataset.

Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose



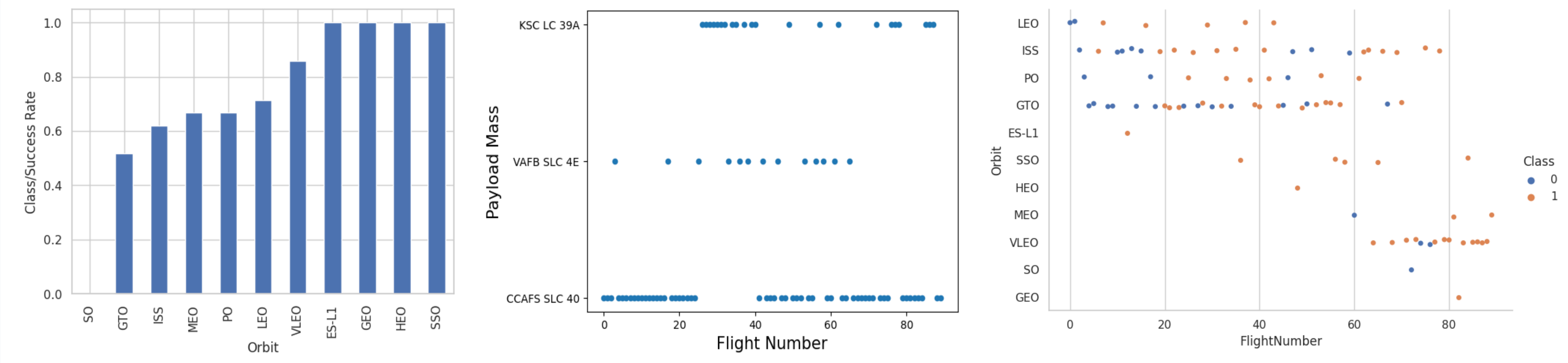
EDA with Data Visualization

- The following charts were plotted for EDA with visualization :-

1. **Scatterplots** to visualize the relationship between ‘Flight number’ and ‘Launch Site’, between ‘Payload’ and ‘Launch Site’, between ‘Flight Number’ and ‘Orbit’ and so on

2. **Bar Charts** to visualize the relationship between success rate of each orbit type.

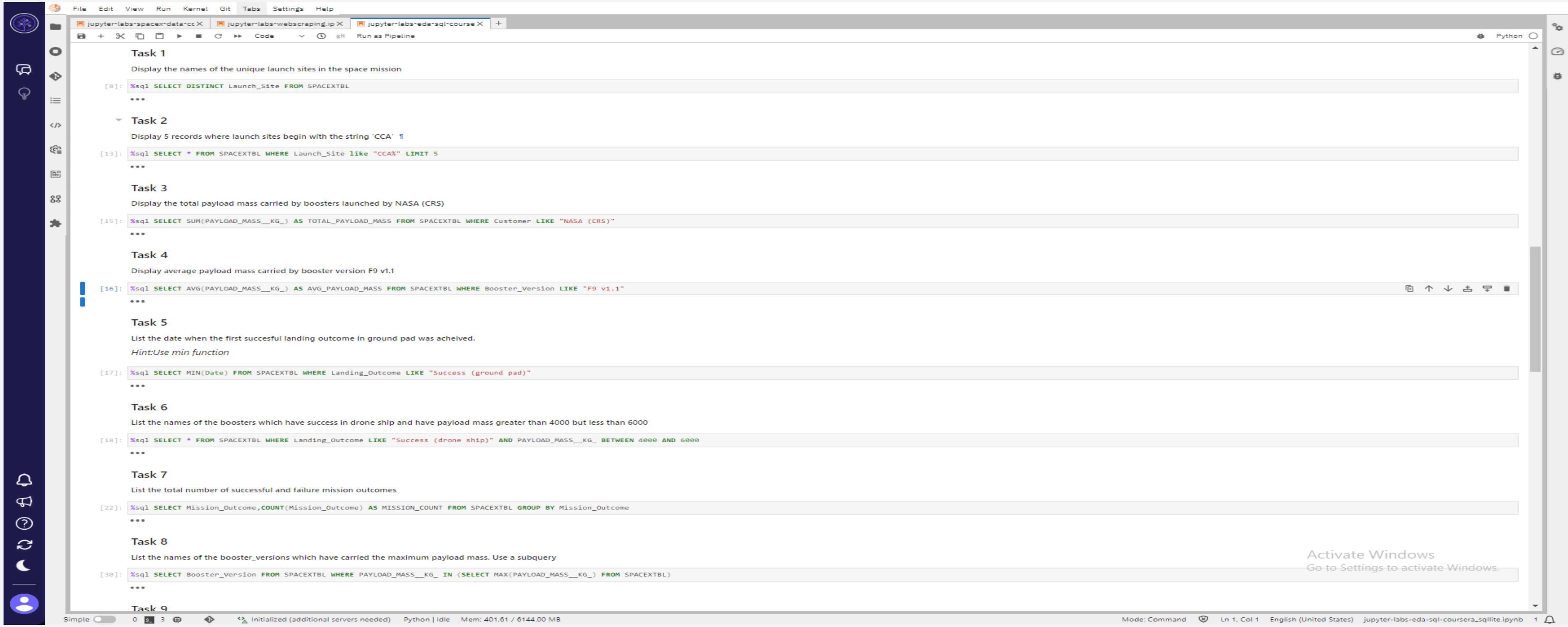
3. **Line plots** to visualize the yearly success trend of launches



Add the GitHub URL of your completed EDA with data visualization notebook, as an

EDA with SQL

- The following queries were used for EDA :



- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

- More interactive visual analytics were done using `Folium` to perform the following tasks :
  1. Mark all launch sites on a map
  2. Mark the success/failed launches for each site on the map
  3. Calculate the distances between a launch site to its proximities

- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- Built an interactive dashboard application by:
    - Adding a Launch Site Drop-down Input Component
    - Adding a callback function to render success-pie-chart based on selected site dropdown
    - Adding a Range Slider to Select Payload
    - Adding a callback function to render the success-payload-scatter-chart scatter plot

- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

- Exploratory Data Analysis was performed and Training Labels were determined to :
  - create a column for the class
  - Standardize the data
  - Split into training data and test data
  - Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
  - Find the method performs best using test dataYou need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

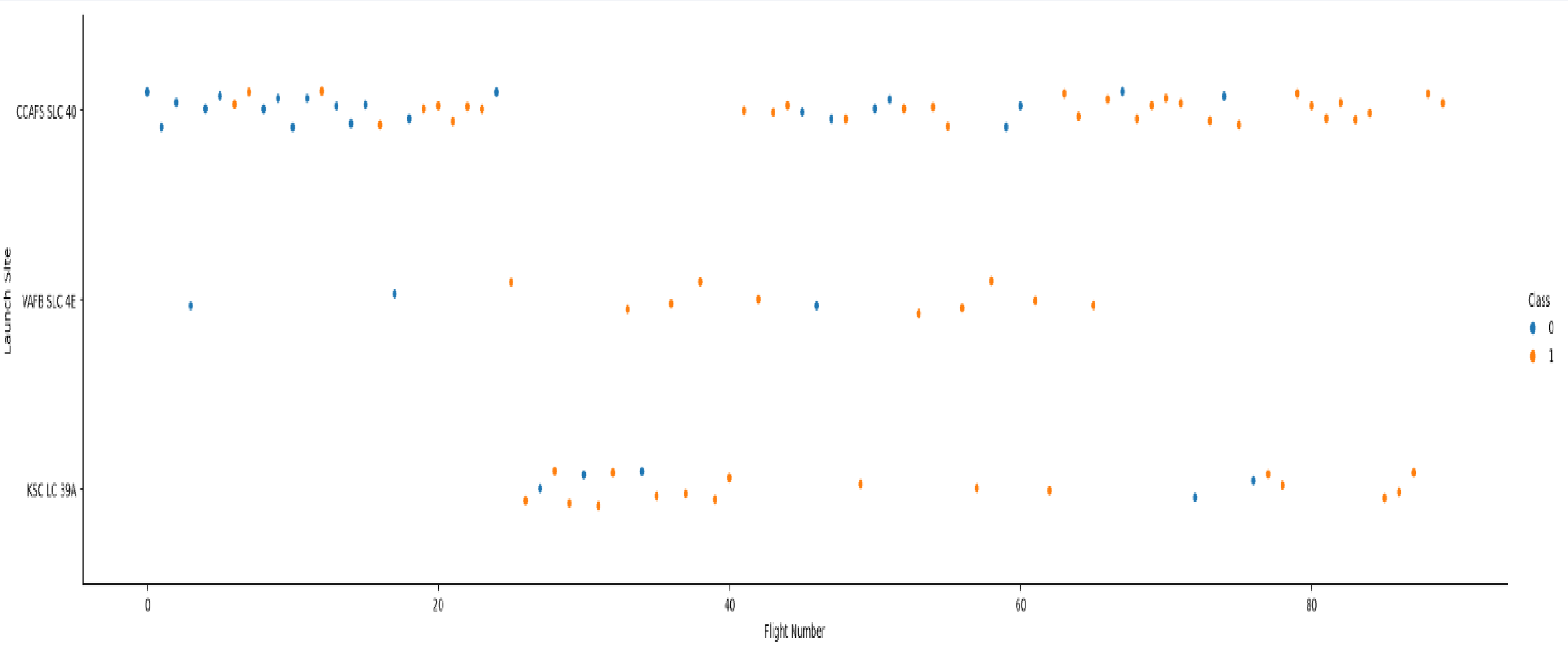
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

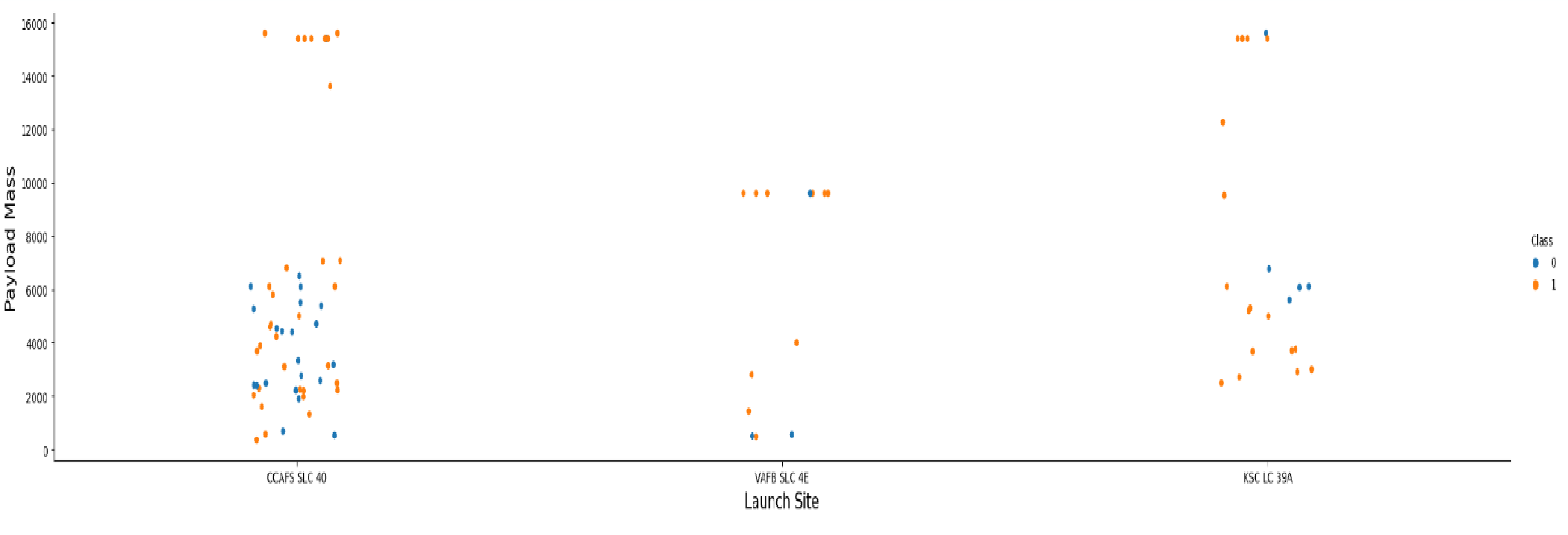
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

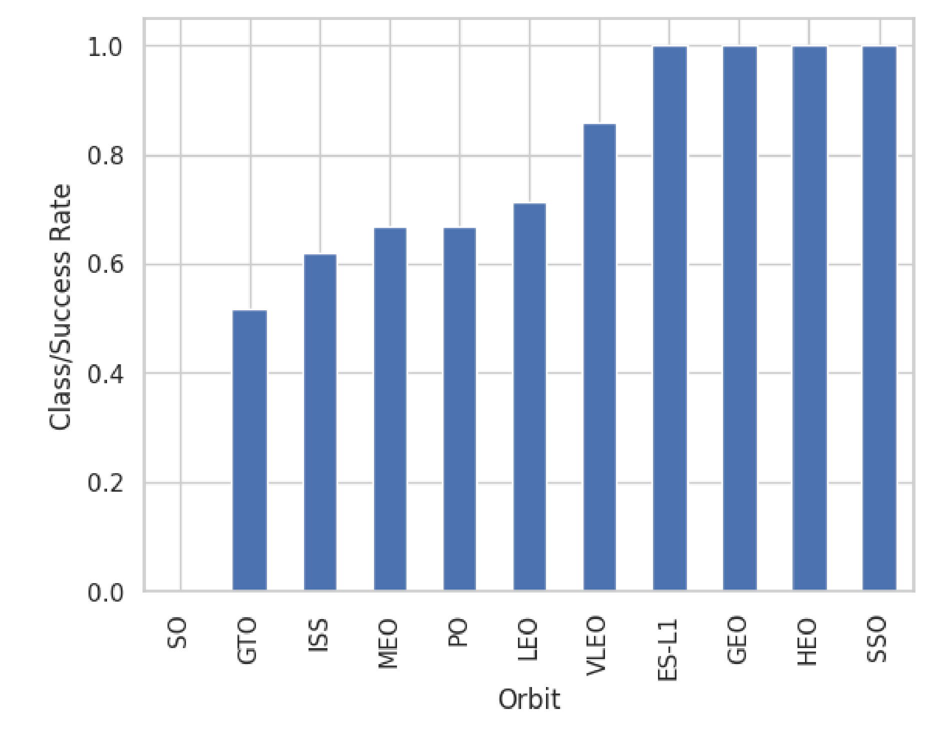


Payload vs. Launch Site

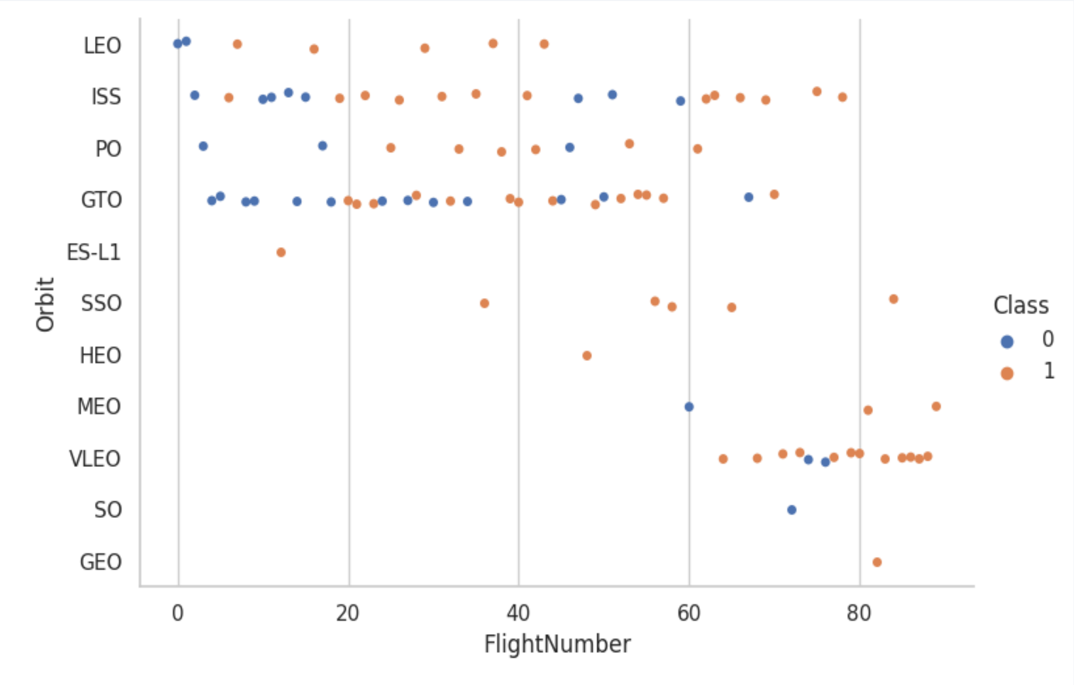


- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

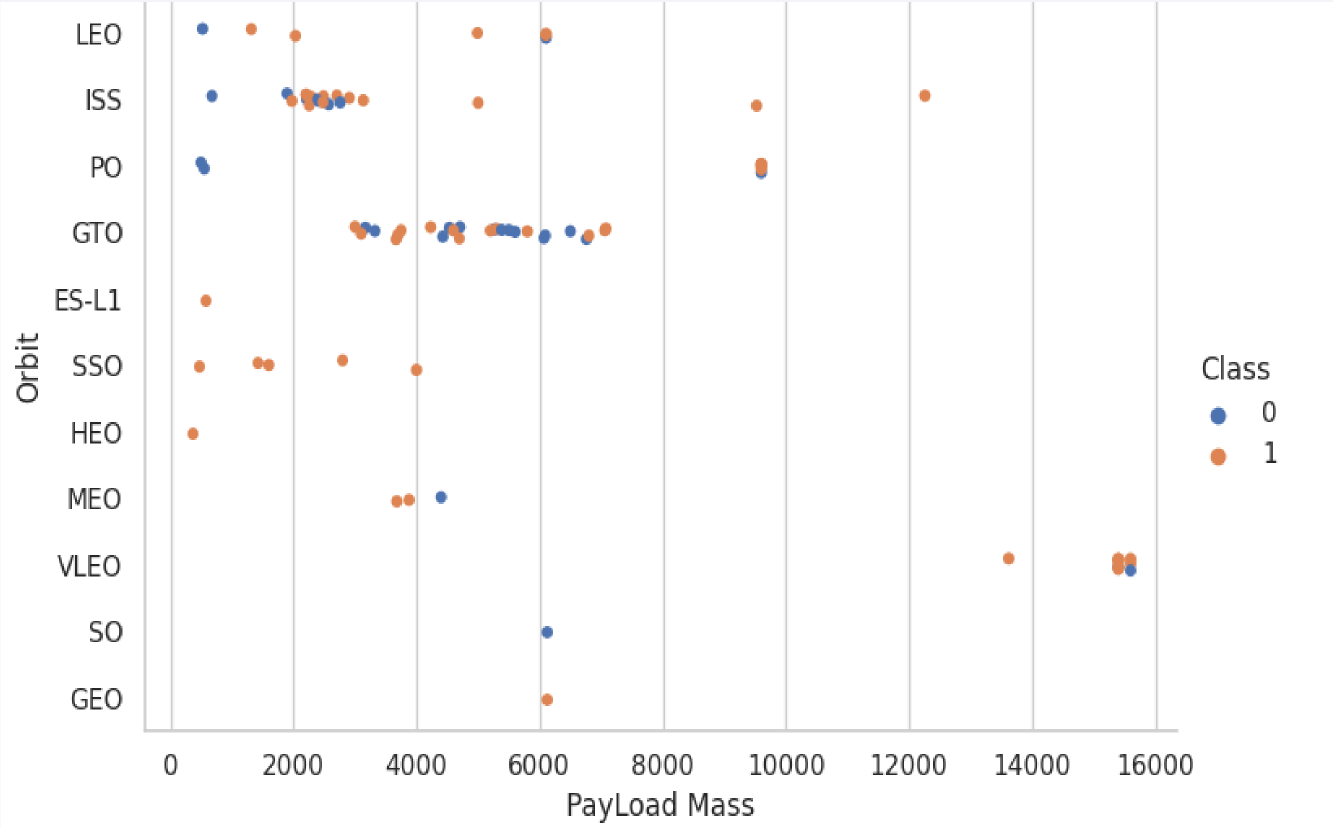


Flight Number vs. Orbit Type



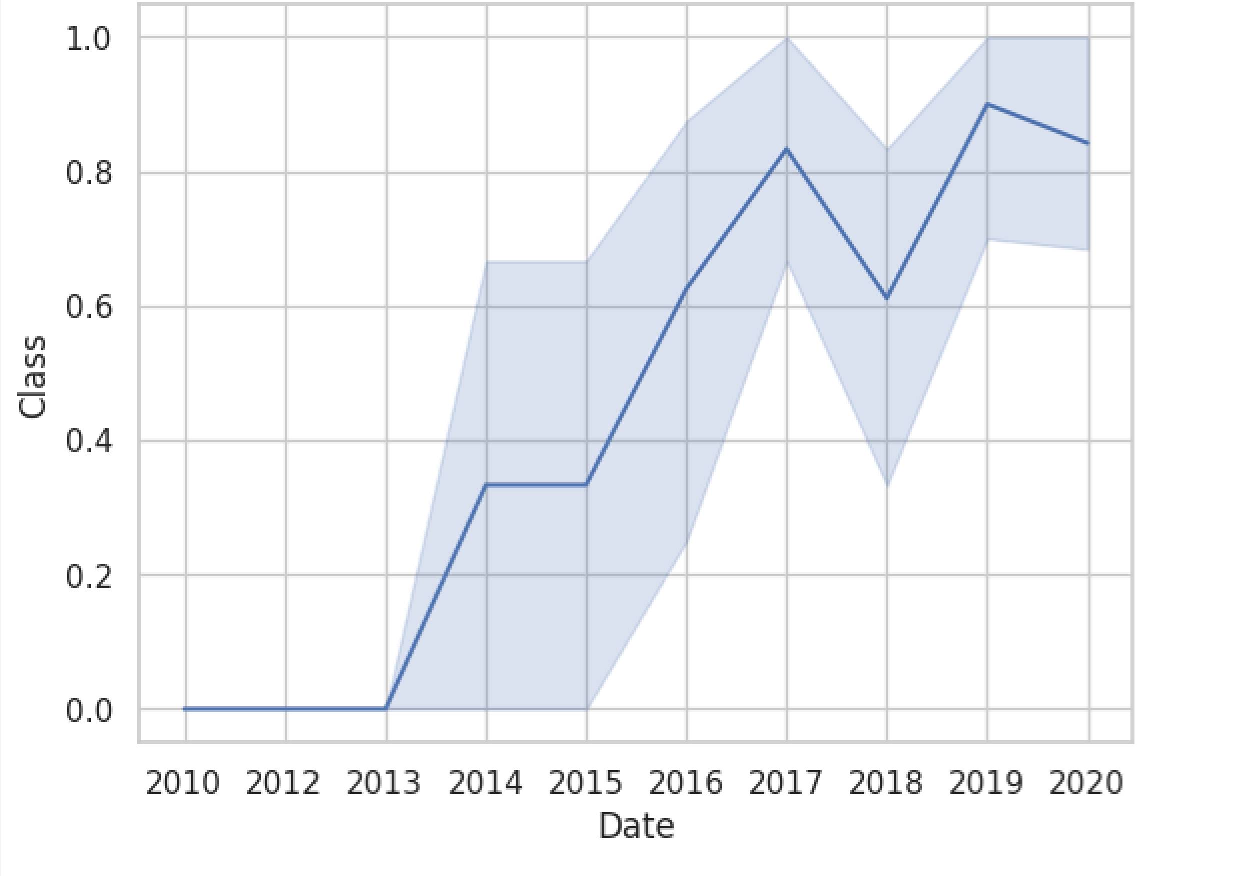
- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

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



- the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

```
[8]: %sql SELECT DISTINCT Launch_Site FROM SPACEXTBL
```

 * sqlite:///my_data1.db
Done.

[8]:

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

Launch Site Names Begin with 'CCA'

```
[13]: %sql SELECT * FROM SPACEXTBL WHERE Launch_Site like "CCA%" LIMIT 5

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

[13]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE Customer LIKE "NASA (CRS)"
```

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

TOTAL_PAYLOAD_MASS
45596

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD_MASS FROM SPACEXTBL WHERE Booster_Version LIKE "F9 v1.1"
```

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

AVG_PAYLOAD_MASS
2928.4

First Successful Ground Landing Date

```
%sql SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome LIKE "Success (ground pad)"
```

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

MIN(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 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
2016-06-05	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-11-10	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT Mission_Outcome,COUNT(Mission_Outcome) AS MISSION_COUNT FROM SPACEXTBL GROUP BY Mission_Outcome
```

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

Mission_Outcome	MISSION_COUNT
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

```
%sql SELECT Booster_Version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ IN (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,Landing_Outcome,Booster_Version,Launch_Site FROM SPACEXTBL  WHERE Landing_Outcome LIKE "Failure (drone ship)" AND Date LIKE '%2015%'
```

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

MONTH	Landing_Outcome	Booster_Version	Launch_Site
5-	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
5-	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

```
%sql SELECT COUNT(Landing_Outcome),Date FROM SPACEXTBL WHERE Landing_Outcome IN (SELECT DISTINCT(Landing_Outcome) FROM SPACEXTBL WHERE Landing_Outcome="Failure (drone ship)" OR Landing_Outcome="Success (ground pad)") AND Date BETWE
```

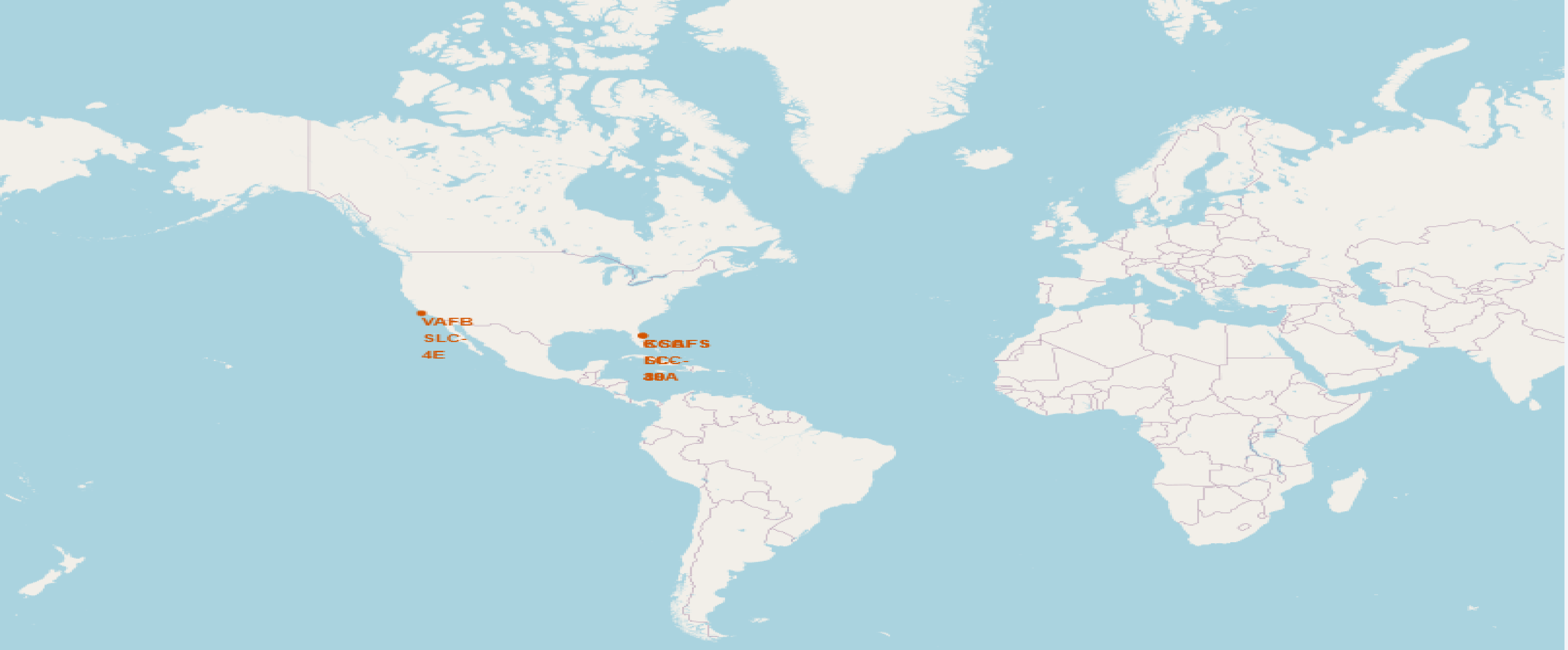
```
 * sqlite:///my_data1.db
Done.
```

COUNT(Landing_Outcome)	Date
10	2015-10-01

Section 3

Launch Sites Proximities Analysis

Mark all launch sites on a map

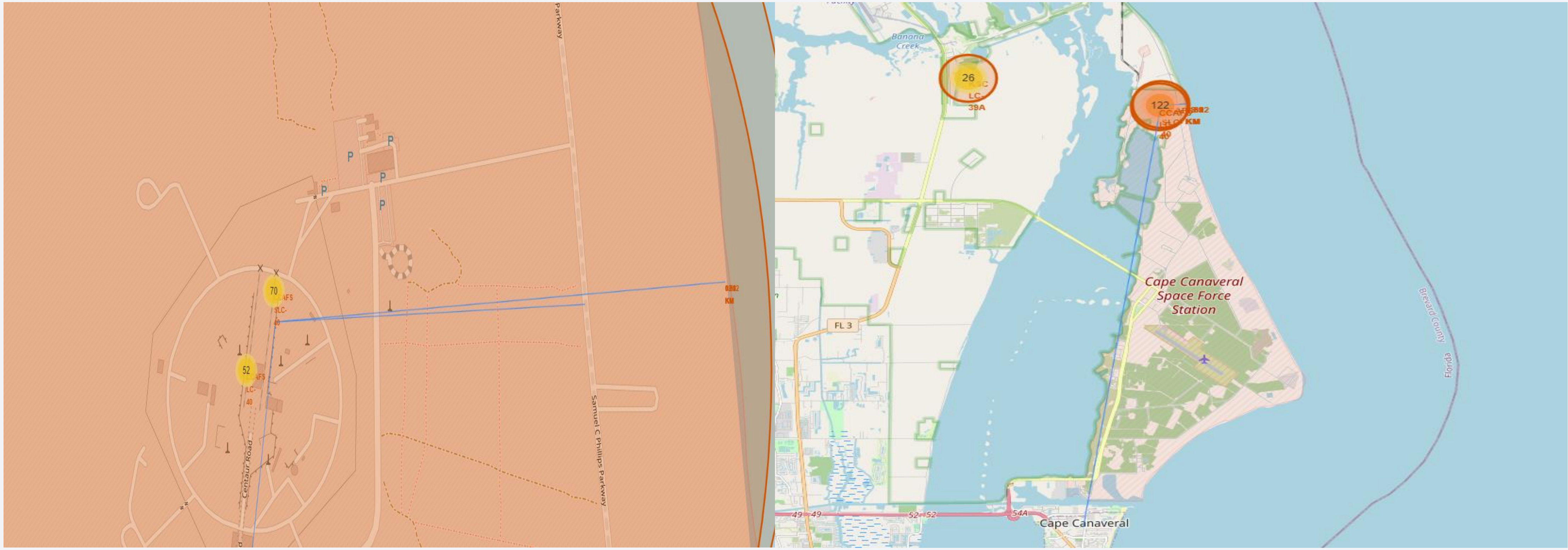


- All launch sites are in proximity of the Equator and in close proximity to coasts as well.

Mark the success/failed launches for each site on the map

In the Eastern coast (Florida) Launch site KSC LC-39A has relatively high success rates compared to CCAFS SLC-40 & CCAFS LC-40

Calculate the distances between a launch site to its proximities

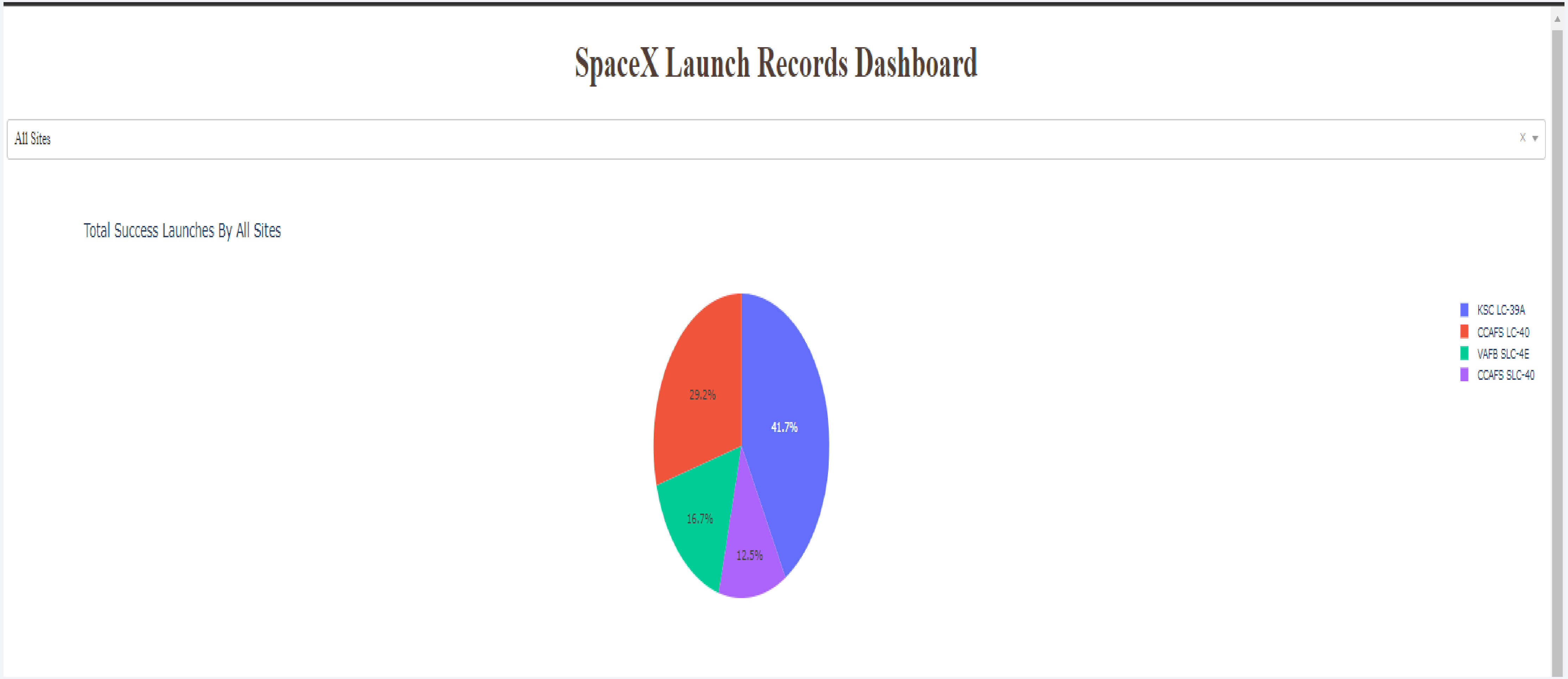


CCAFS SLC-40 is 0.1 km,0.59 km and 18.22 km from the nearest rail line, highway and city respectively

Section 4

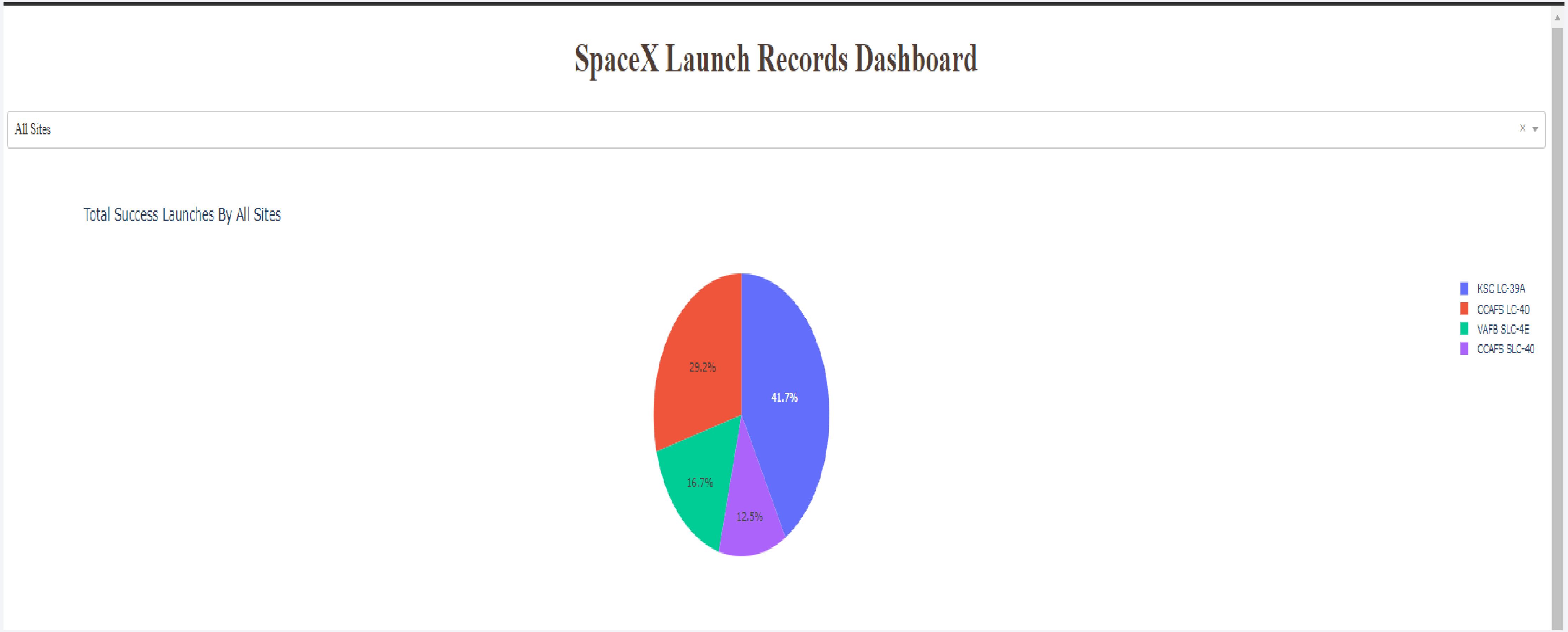
Build a Dashboard with Plotly Dash

Pie chart for the launch site with 2nd highest launch success ratio



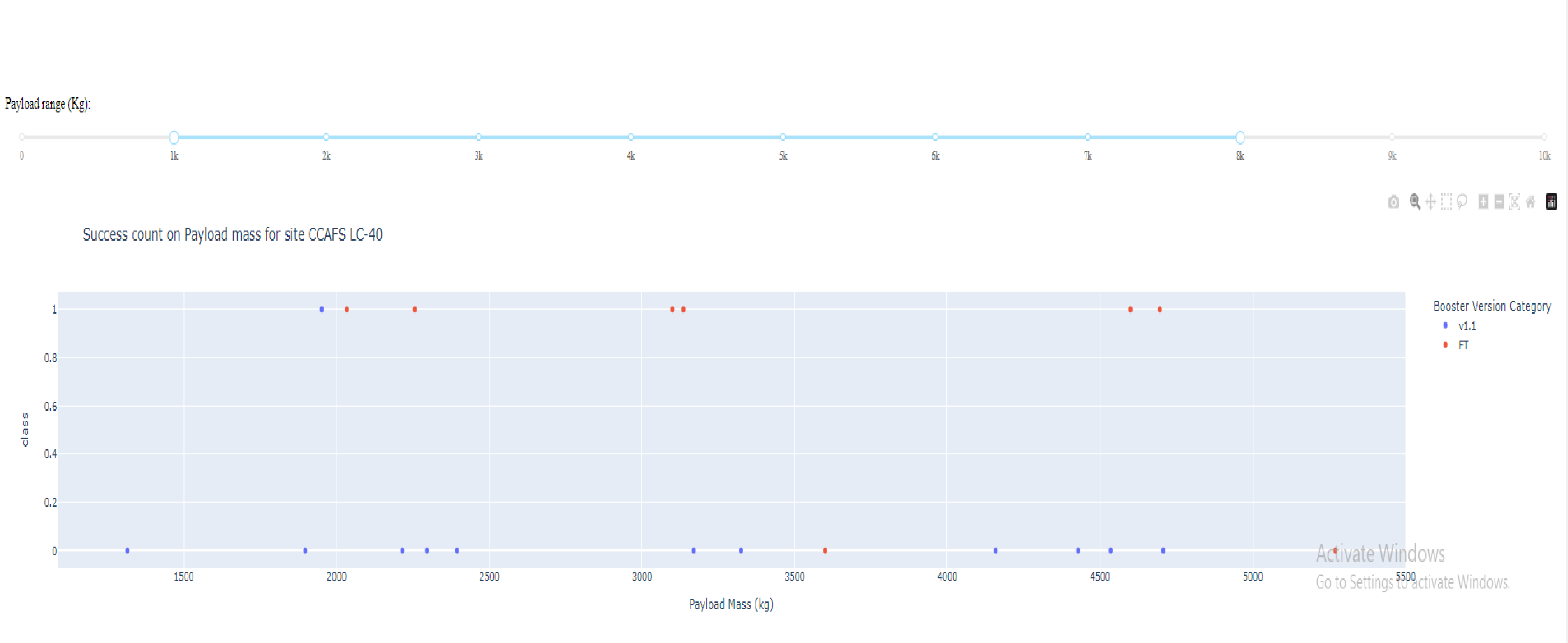
- Launch site KSC LC-39A has the highest launch success rate at 42% followed by CCAFS LC-40 at 29%, VAFB SLC-4E at 17% and lastly launch site CCAFS SLC-40 with a success rate of 13%

Pie chart for the launch site with 2nd highest launch success ratio



- Launch site CCAFS LC-40 had the 2nd highest success ratio of 73% success against 27% failed launches

Payload vs. Launch Outcome scatter plot for all sites

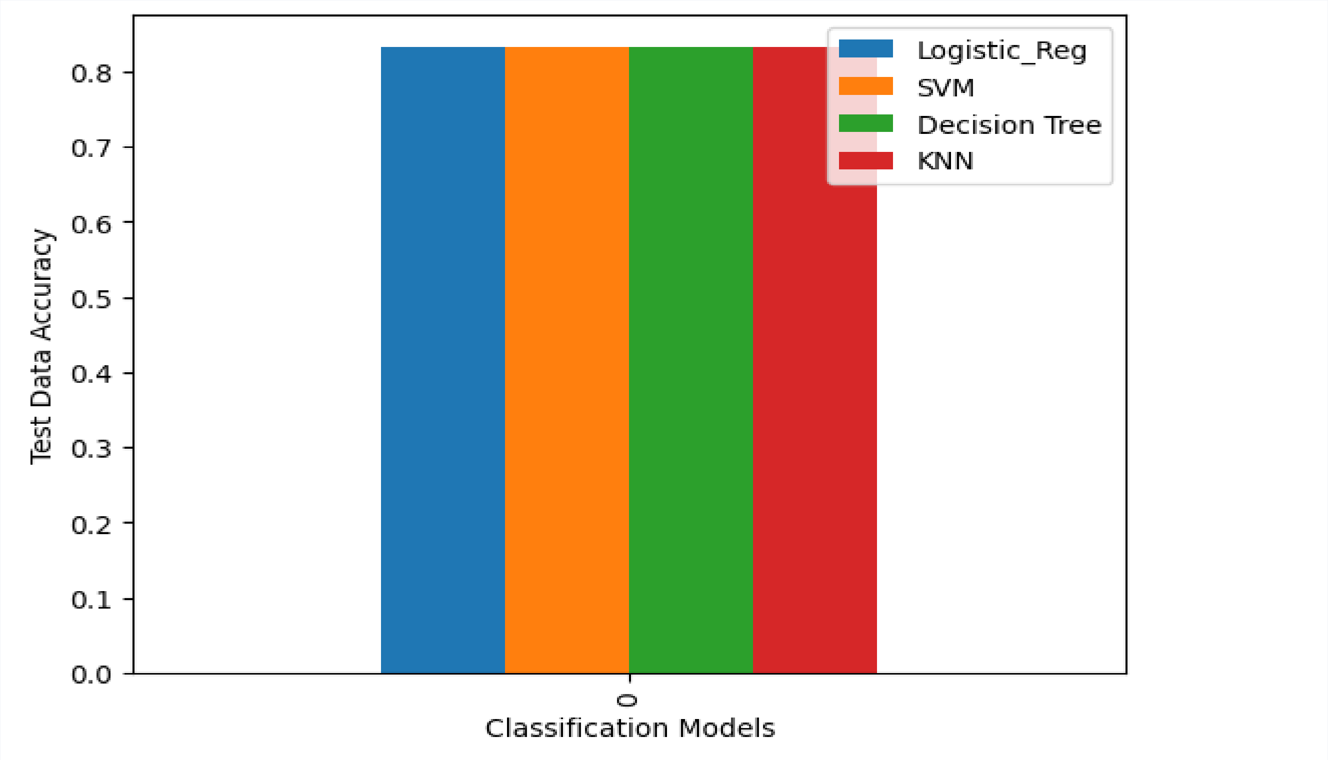


- For Launch site CCAFS LC-40 the booster version FT has the largest success rate from a payload mass of >2000kg

Section 5

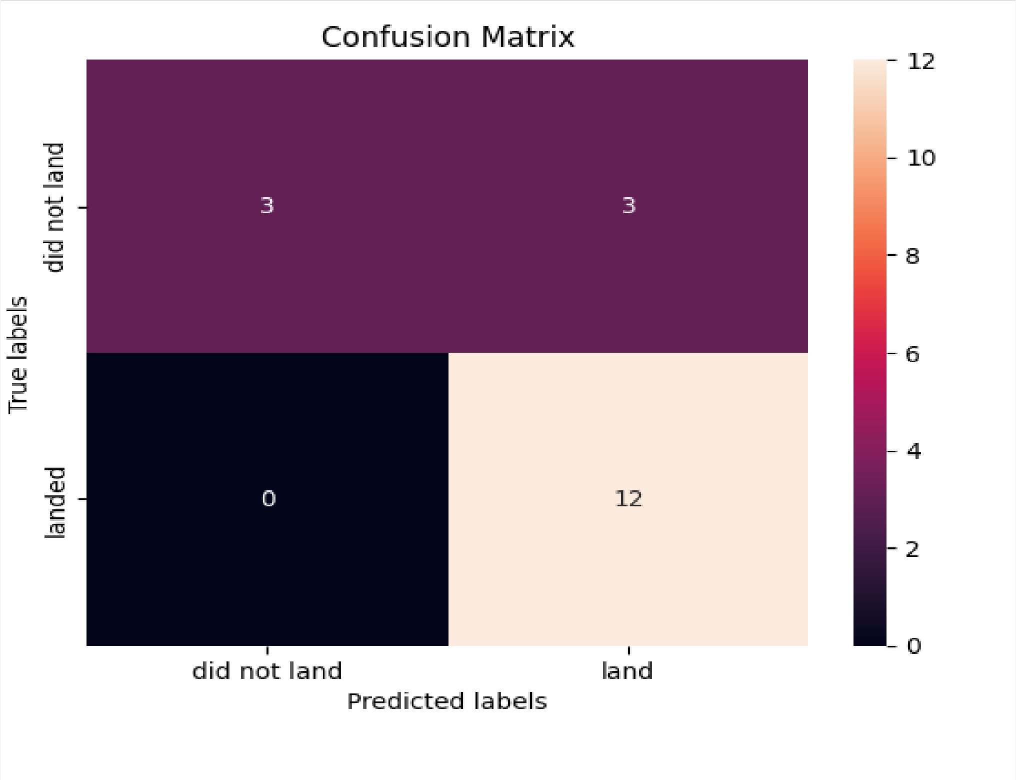
Predictive Analysis (Classification)

Classification Accuracy



- All the classification models have the same test accuracy.

Confusion Matrix



- All the 4 classification model had the same confusion matrixes and were able equally distinguish between the different classes. The major problem is false positives for all the models.

Conclusions

- Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

• We can deduce that, as the flight number increases in each of the 3 launch sites, so does the success rate. For instance, the success rate for the VAFB SLC 4E launch site is 100% after the Flight number 50. Both KSC LC 39A and CCAFS SLC 40 have a 100% success rates after 80th flight

• In Payload Vs. Launch Site scatter point chart ,the VAFB-SLC launch site has no rockets launched for heavy payload mass(greater than 10000).

• Orbits ES-L1, GEO, HEO & SSO have the highest success rates at 100%, with SO orbit having the lowest success rate at ~50%. Orbit SO has 0% success rate.

• 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 Conclusions 57

• 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

• The success rate since 2013 kept increasing till 2020.

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!