Winning Space Race with Data Science

Indumathi Pandiyan
9/30/2022

Outline

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

Executive Summary

- Summary of methodologies

Data was collected from the SpaceX API and scraped from the SpaceX Wikipedia webpage. The data was analyzed and cleaned. Missing values were corrected where necessary. After gathering insights with graphs and summary statistics from the data, features were selected. With the selected features, predict models were trained and hyperparameters were fine-tuned with grid search. Models were tested with test data and each model was scored.

- Summary of all results

While cleaning the data, there were 5 values for payload mass that were missing. The mean value was used to replace the missing values. From graphing, a relationship between reuse of the first stage rocket and several features were identified. The models tested all had similar accuracy scores of 83%. Since all of the tested models performed similarly, the Logistic Regression was selected for the added feature of providing the probability of being able to reuse first stage rockets.

Introduction

Project background and context

Rocket launches can have some sky-high price tags, upwards of $165 million each. Reusing the first stage rockets can be bring that cost a little more down to earth at about $62 million. Being able to predict when a rocket can be reused can lead to understanding the conditions in which a rocket can be reused.

- Problems you want to find answers

We want to be able to predict when a first stage rocket can be reused. With a model that can predict this, we can determine the cost of a launch for customers.

Section 1

Methodology

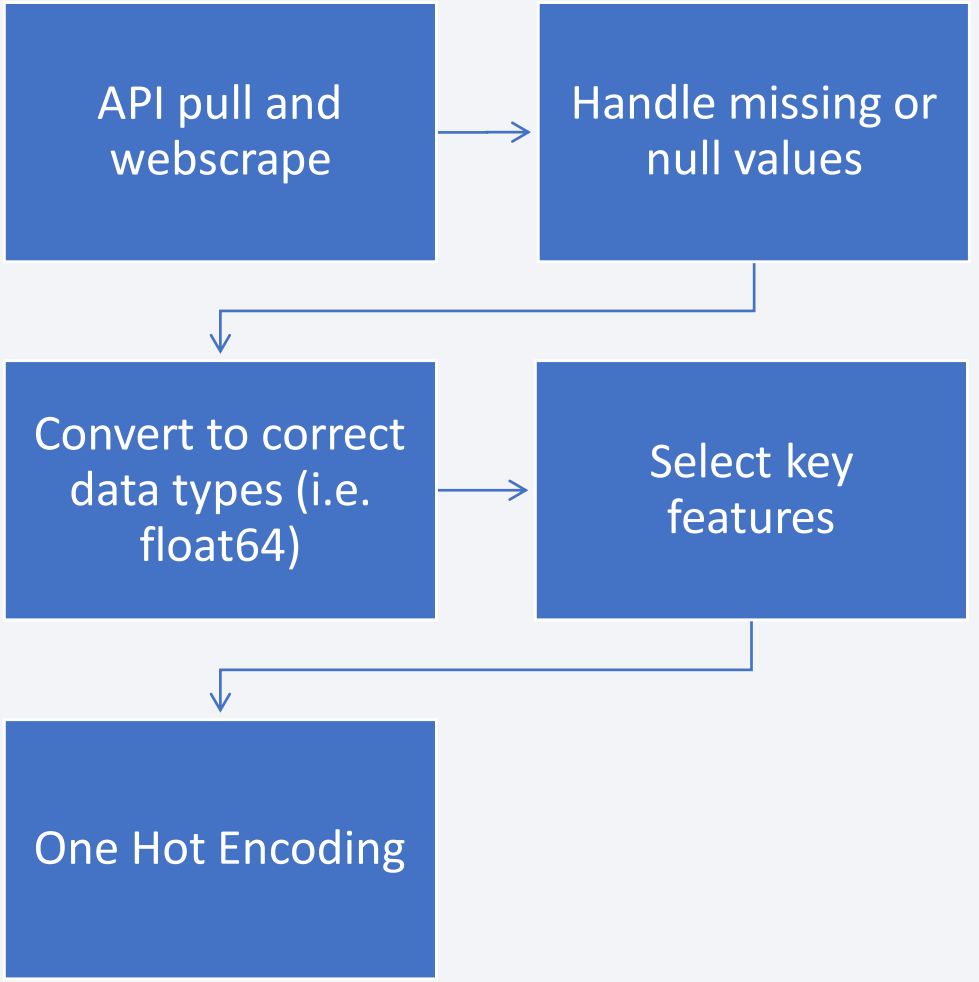
Methodology

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

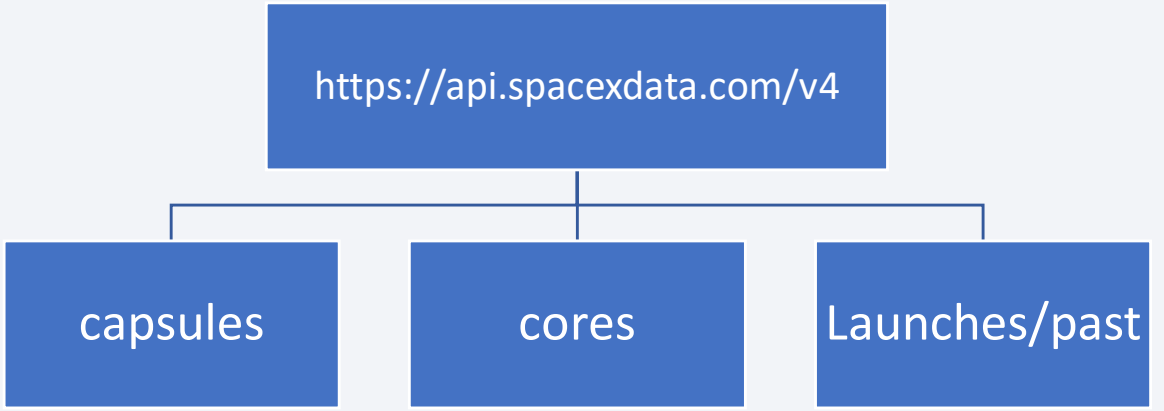
Data was pulled from the SpaceX API using the requests library. Additional information was obtained by scraping the SpaceX Wikipedia webpage using the BeautifulSoup library.



Data Collection - SpaceX API

Spacex Data Collection API notebook

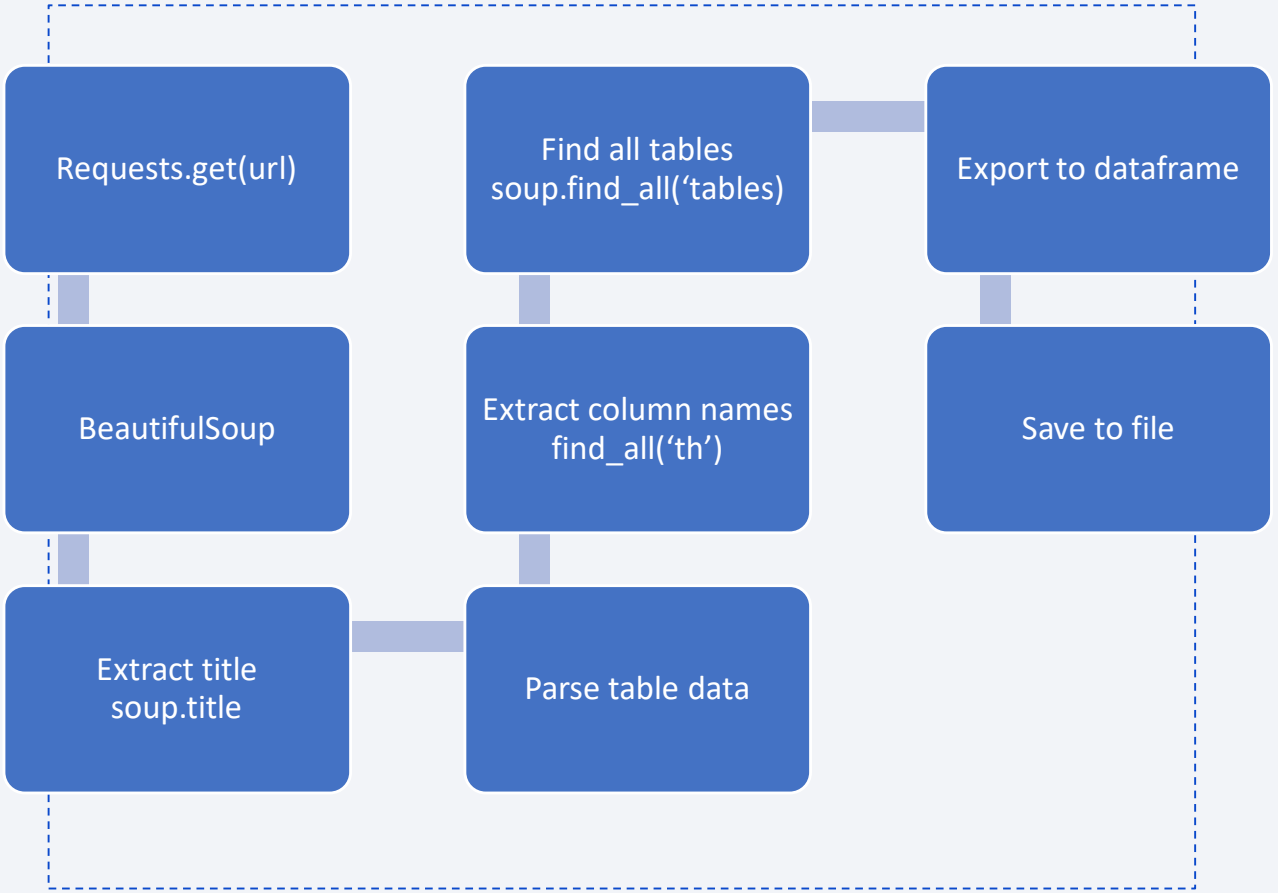
- Requests library was used for the API calls



Data Collection - Scraping

Data Collection with Webscraping notebook

Pulled data with requests library and used the BeautifulSoup library to extract data.

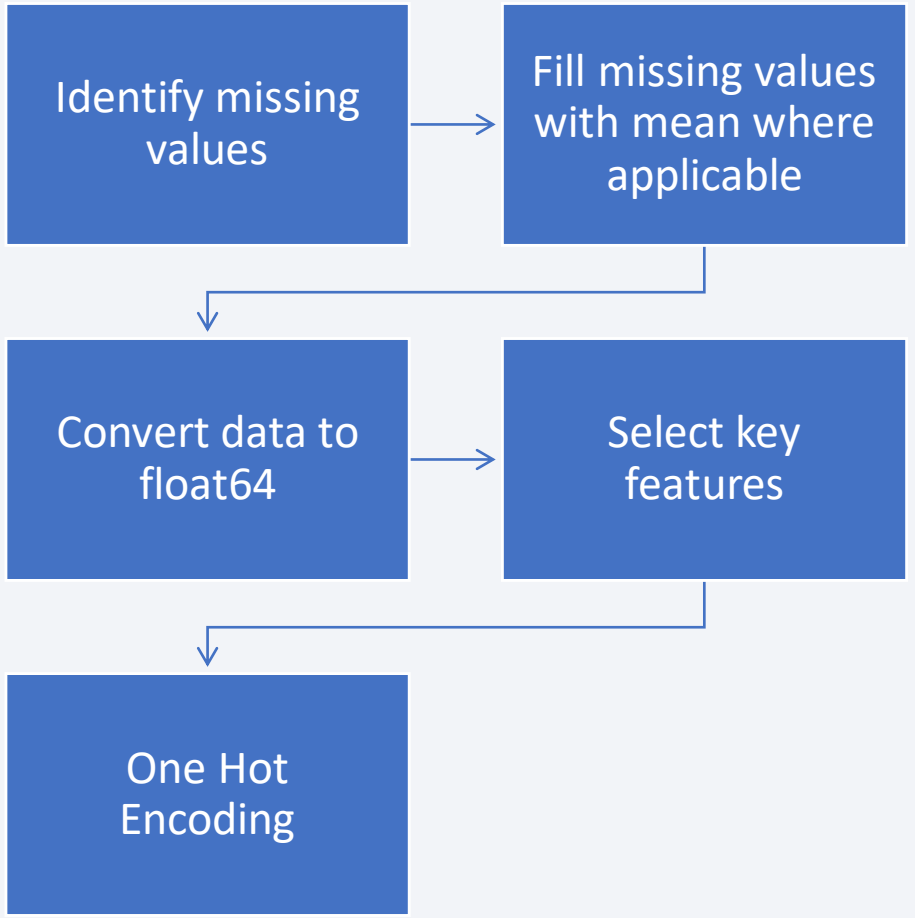


Data Wrangling

Describe how data were processed

You need to present your data wrangling process using key phrases and flowcharts

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



EDA with Data Visualization

- Flight number vs Payload with hue for Success
- Flight number vs Launch Site with hue for Success
- Payload Mass vs Launch Site with hue for Success
- Orbit vs Success Rate
- Flight number vs Orbit
- Payload vs Orbit
- Year vs Success Rate
- Data Visualization Notebook

EDA with SQL

- Get the names of the launch sites
- Display 5 records of launches from site names containing ‘CCA’
- Display total payload from launches by NASA (CRS)
- Display average payload for booster F9 v1.1
- Display the first date of a successful landing on a ground pad
- Display names of boosters with success on drone ship and payload mass between 4,000 and 6,000 kg
- Display total number of mission successes and failures
- Display names of boosters that carried the max payload
- Display records for the year 2015
- SQL Data Analysis Notebook

Build an Interactive Map with Folium

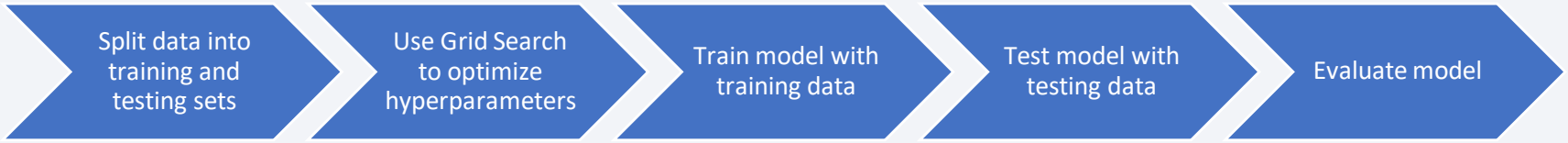
- Map launching locations with markers for each launch color coded based on success and a line from CCAFS launch sites to the coastline.
- Creates a visual representation of how many launches were successful and failures for each launch site
- Interactive Visual Analytics with Folium notebook

Build a Dashboard with Plotly Dash

- Pie chart with distribution of successful launches by site with a drop down to see a breakdown of successes and failures by launch site
- Scatter plot with payload by success colored by booster with slider to change the payload window on the x-axis
- Visual representations of successes by launch site and payload.
- Plotly Dashboard

Predictive Analysis (Classification)

- Logistic Regression was built using Grid Search to optimize hyperparameters, evaluated with a confusion matrix and accuracy score.



- Machine Learning notebook

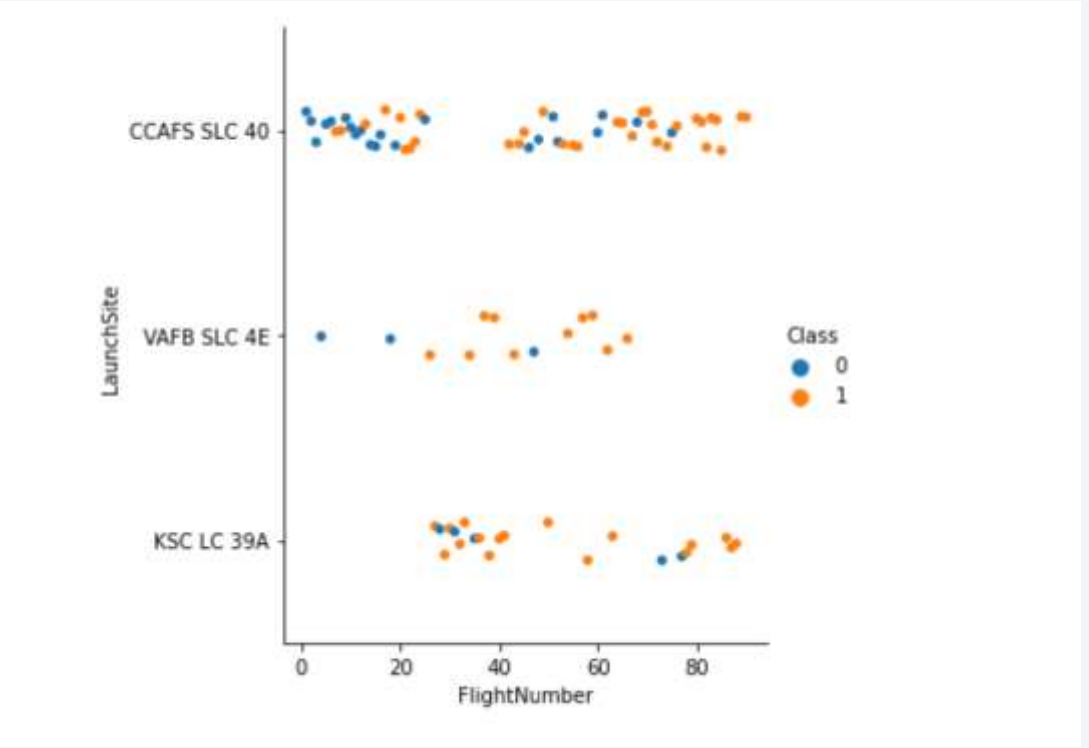
Results

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


Section 2
Insights drawn from EDA

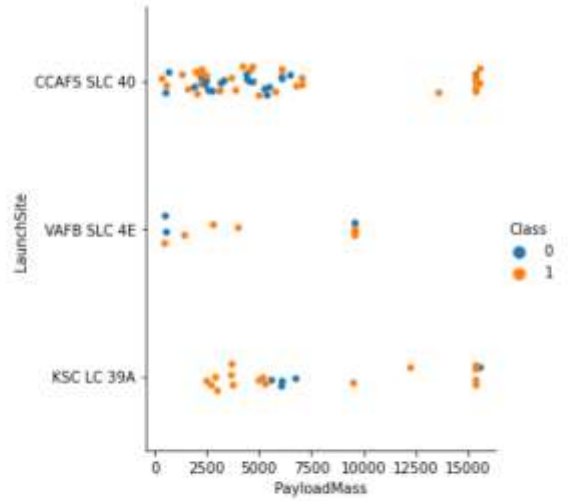
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations



Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations

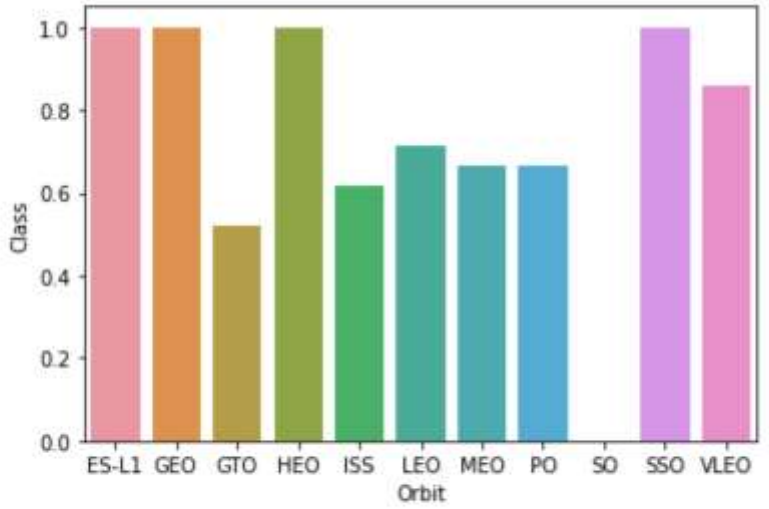


Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

- Show the screenshot of the scatter plot with explanations

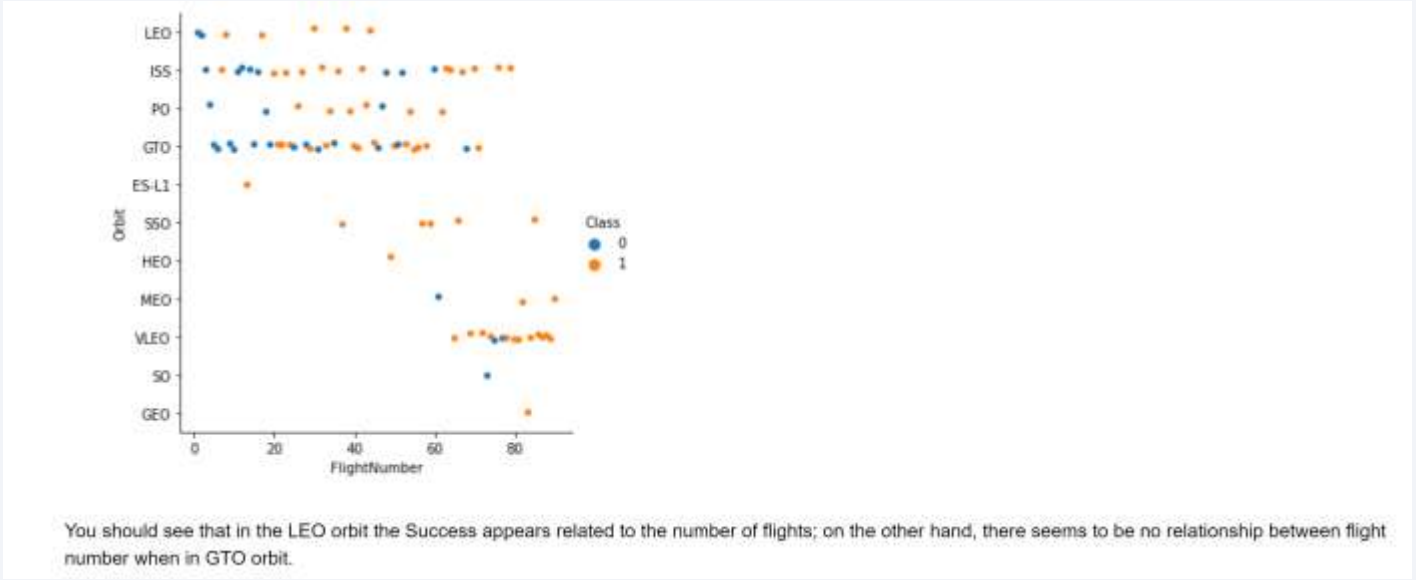


Analyze the ploted bar chart try to find which orbits have high sucess rate.

Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type

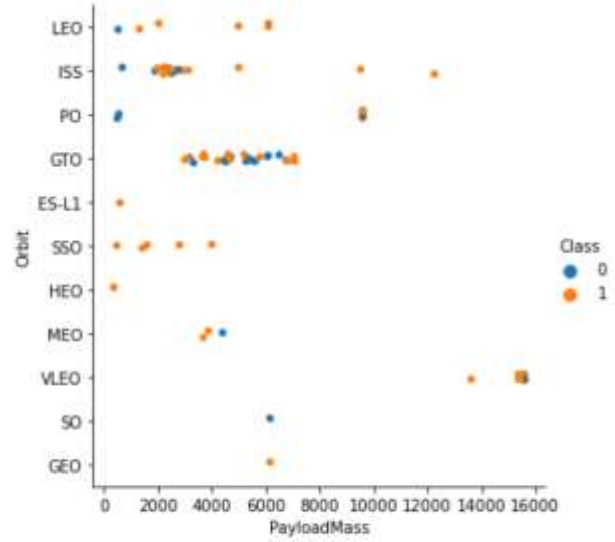
- Show the screenshot of the scatter plot with explanations



You should 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

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

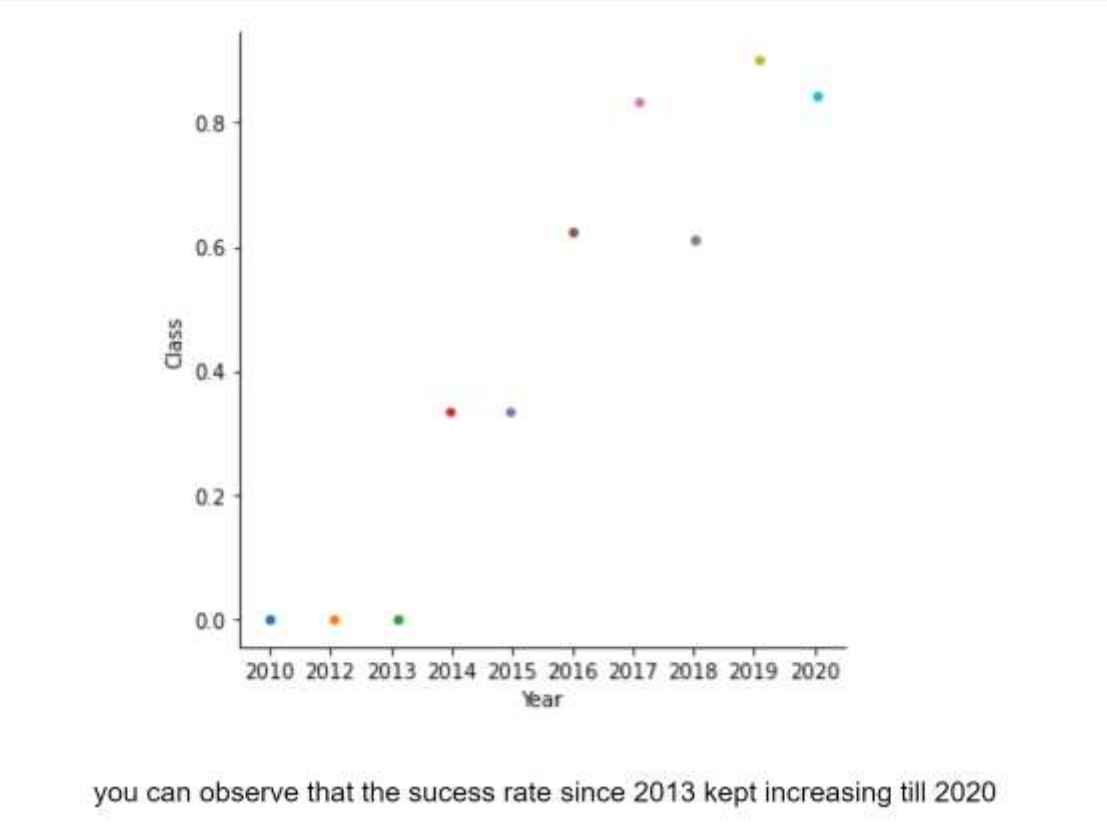


With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

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

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Display the names of the unique launch sites in the space mission

```
In [10]: %%sql
         select distinct launch_site from spacextbl;
```

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

Out[10]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

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

```
In [11]: %sql select * from spacextbl where launch_site like 'CCA%' limit 5;
 * sqlite:///my_data1.db
Done.
```

Out[11]:

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

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

```
In [10]: %%sql select customer, sum(payload_mass__kg_) as total_payload_mass
         from spacextbl where customer = 'NASA (CRS)' group by customer;
```

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

Out[10]:

Customer	total_payload_mass
NASA (CRS)	45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

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

In [11]:
```
%%sql select booster_version, avg(payload_mass__kg_) as avg_payload_mass
from spacextbl where booster_version = 'F9 v1.1' group by booster_version;
```

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

Out[11]:

Booster_Version	avg_payload_mass
F9 v1.1	2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

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

Hint:Use min function

In [20]:
```
%sql select min(date) from spacextbl where "Landing _Outcome" = 'Success (ground pad)';
```

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

Out[20]:

min(date)
01-05-2017

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

- Present your query result with a short explanation here

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [21]:
```
%sql select * from spacextbl where "Landing _Outcome" = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;
```

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

Out[21]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
11-10-2017	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

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

List the total number of successful and failure mission outcomes

In [21]:
```
%%sql select Mission_Outcome, count(*) as total_number_of_outcomes
from spacextbl group by Mission_Outcome;
```

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

Out[21]:

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

Boosters Carried Maximum Payload

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

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

In [22]:
```
%%sql select distinct booster_version
from spacextbl where payload_mass__kg_ =
(select max(payload_mass__kg_) from spacextbl);
```

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

Out[22]:

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
- Present your query result with a short explanation here

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

[31]:
```
%sql select substr(date,4,2) as month, "Landing _Outcome", booster_version, launch_site from spacextbl where substr(date,7,4) = '2015'
```

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

t[31]:

month	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	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
- Present your query result with a short explanation here

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

In [36]:
```
%sql select "Landing _Outcome", count, rank() over (order by count DESC, "Landing _Outcome") as rank from (select "Landing _Outcome", count(*) as
```

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

Out[36]:

Landing _Outcome	count	rank
Success	38	1
No attempt	21	2
Success (drone ship)	14	3
Success (ground pad)	9	4
Controlled (ocean)	5	5
Failure (drone ship)	5	6
Failure	3	7
Failure (parachute)	2	8
Uncontrolled (ocean)	2	9
No attempt	1	10
Precluded (drone ship)	1	11

Section 3

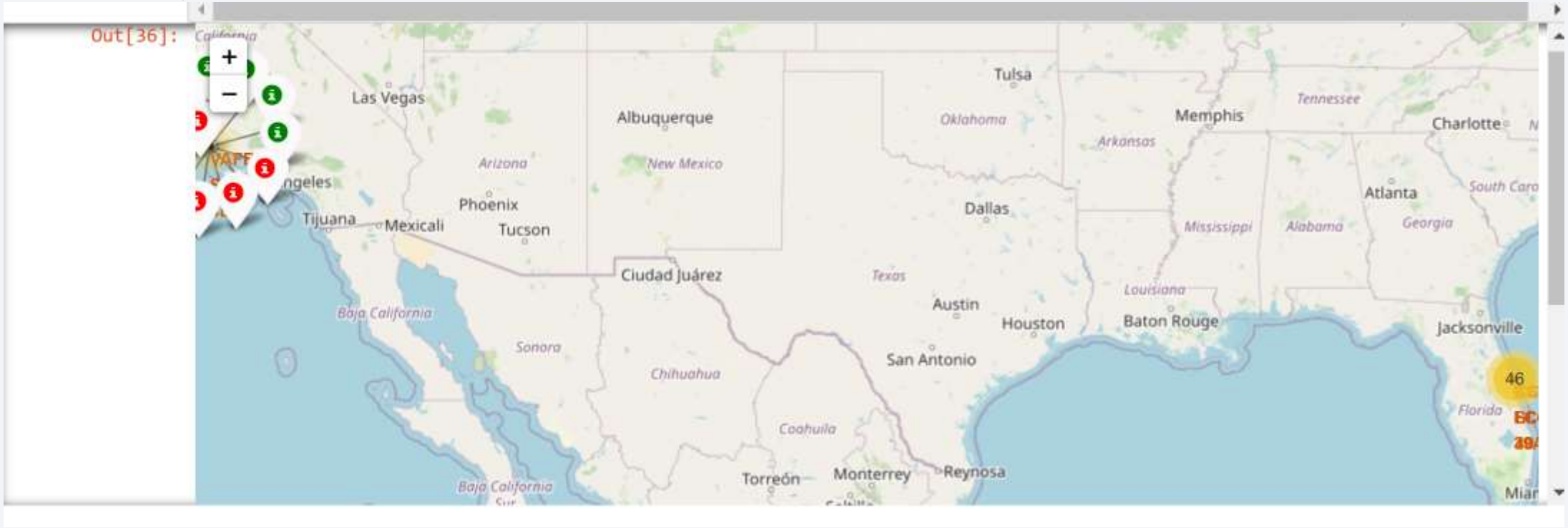
Launch Sites Proximities Analysis

Map of Launch Sites

- Location of launch sites

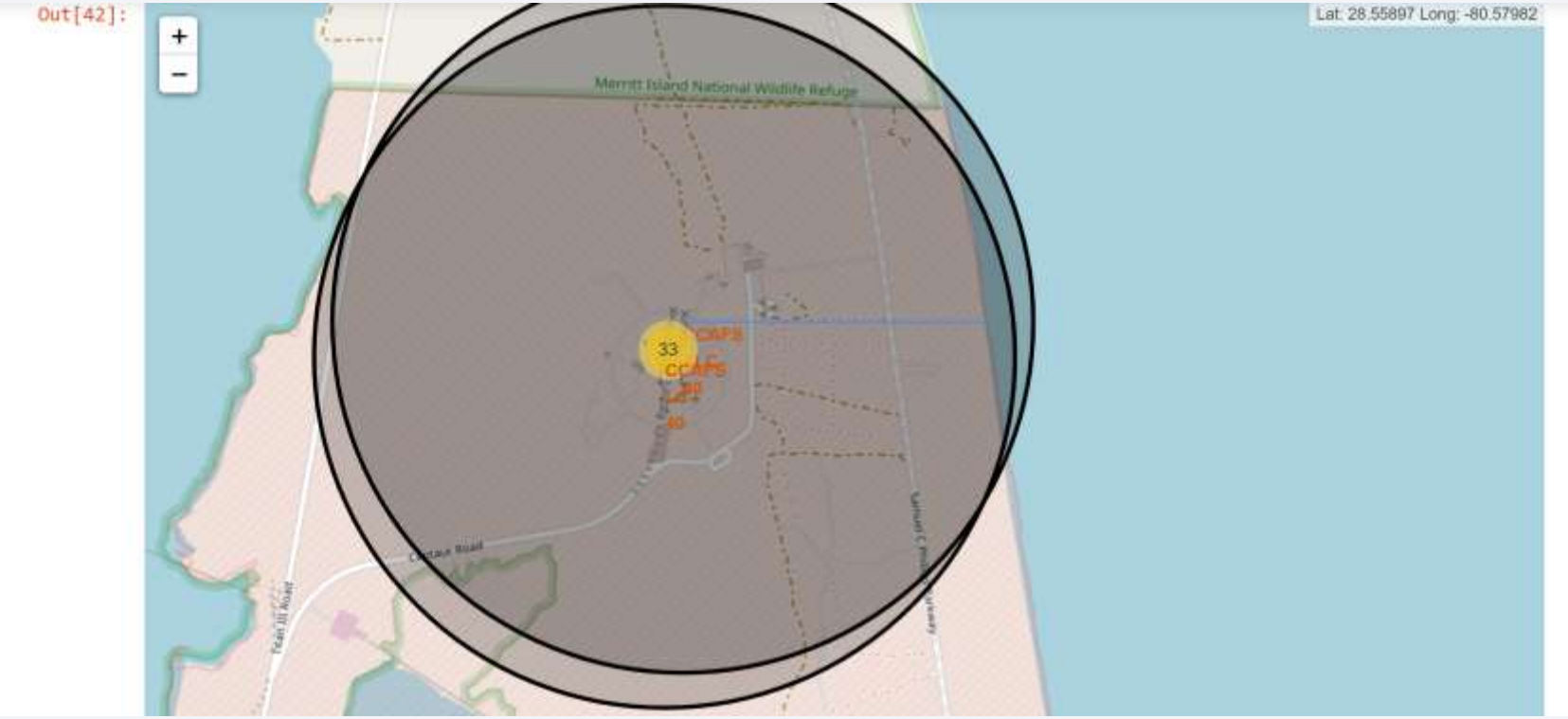
Map of Launch Site per launch by success

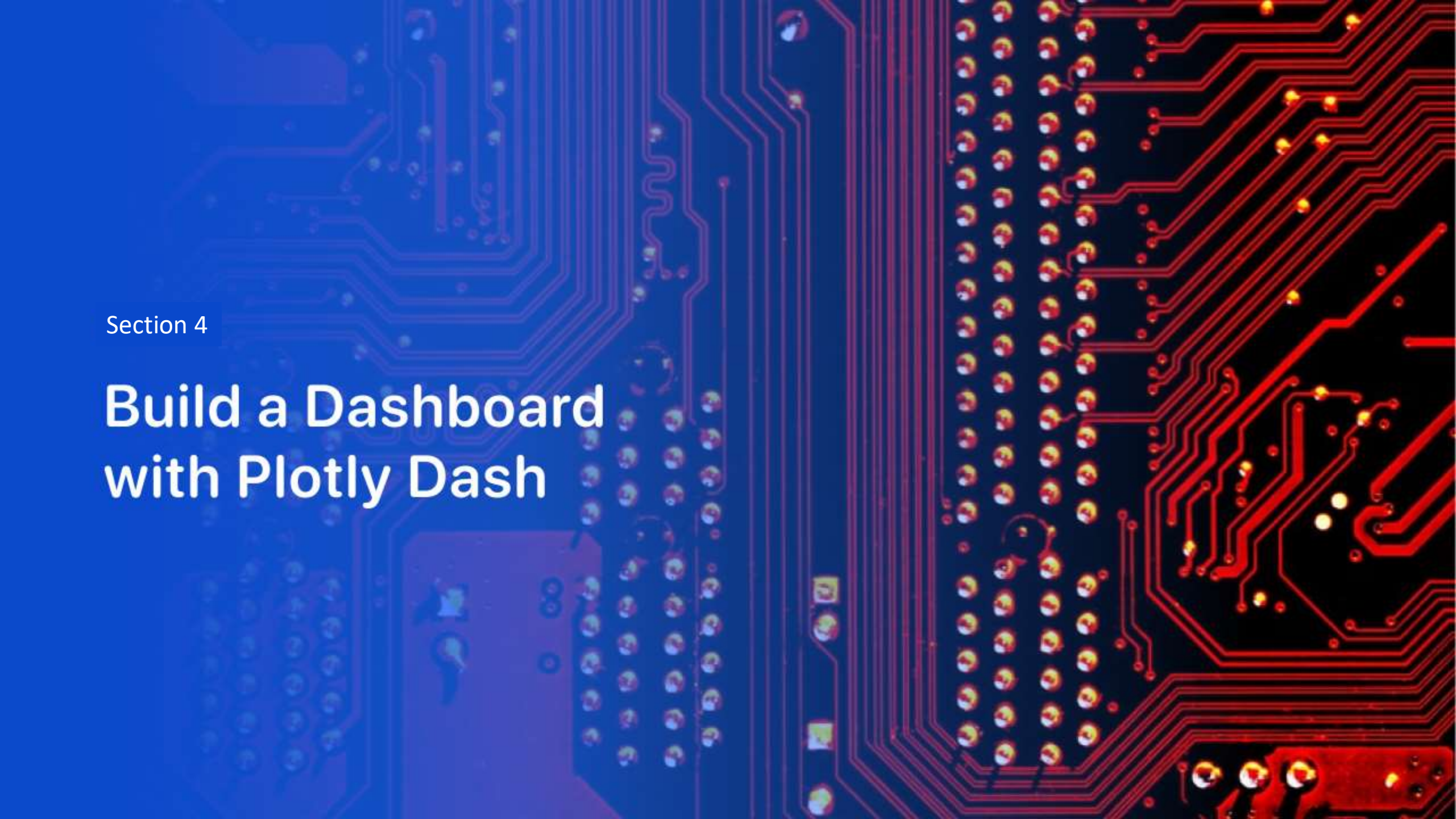
- A pin for each launch at the appropriate launch site colored green for success, red for failure



Launch Site to Coastline

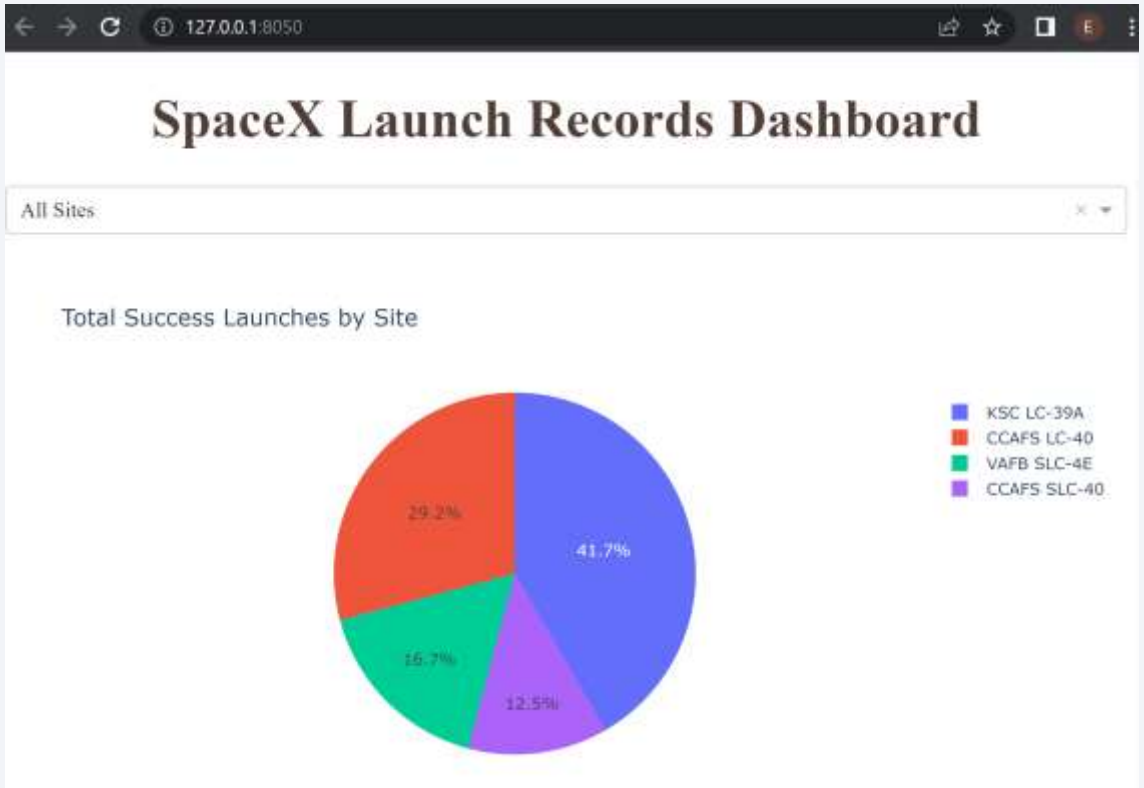
- Includes line to the coastline




Section 4
Build a Dashboard with Plotly Dash

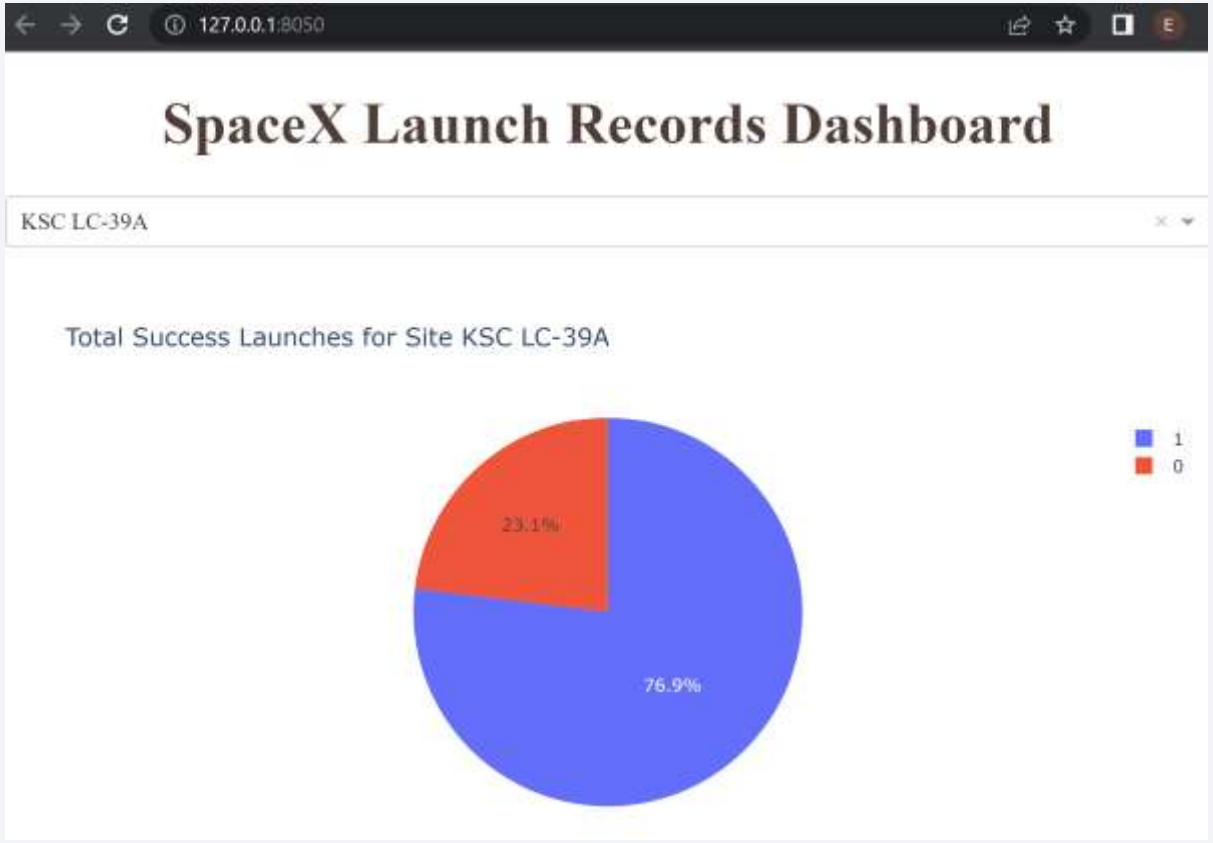
Total Success Launches by Site

- Percentage of total successful launches by site



Success Ratio for KSC LC-39A

- Ratio of successful launches for KSC LC-39A



Payloads Between 2,000 kg and 8,000 kg for All Launch Sites

- Smaller payloads have a higher success rate

Section 5

Predictive Analysis (Classification)

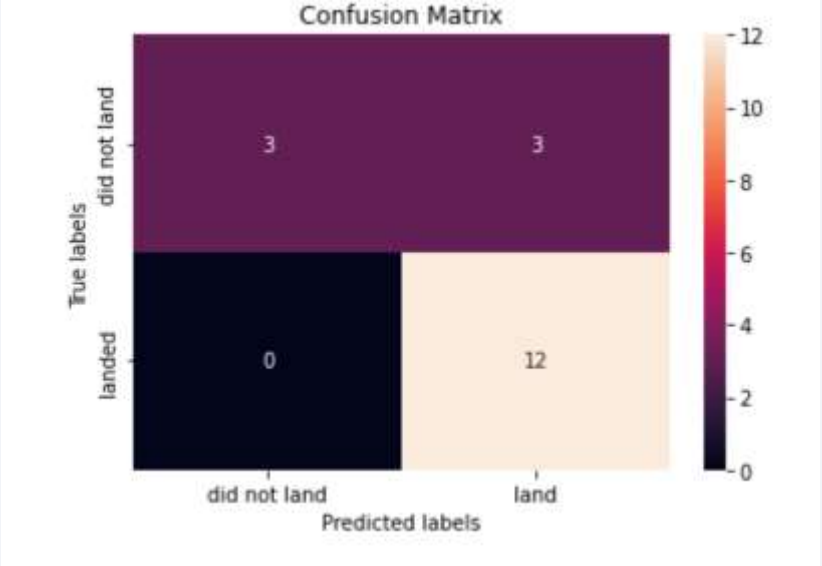
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- All models performed the same
- Lower payloads had better success

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!