Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- In a market where there is very little amount of competition space X has revolutionized space exploration with their reusable rockets
- Falcon9 booster successful recovery depends on features such as:
  - orbit
  - payload mass
  - booster versions
  - Launching sites
- Many different machine learning algorithms are used to predict the likely hood of Falcon9 booster recovery.
- Based on these features, the best Machine Learning supervised classification model developed in this report, predicted booster recovery outcome with an accuracy close to 94%.

Introduction

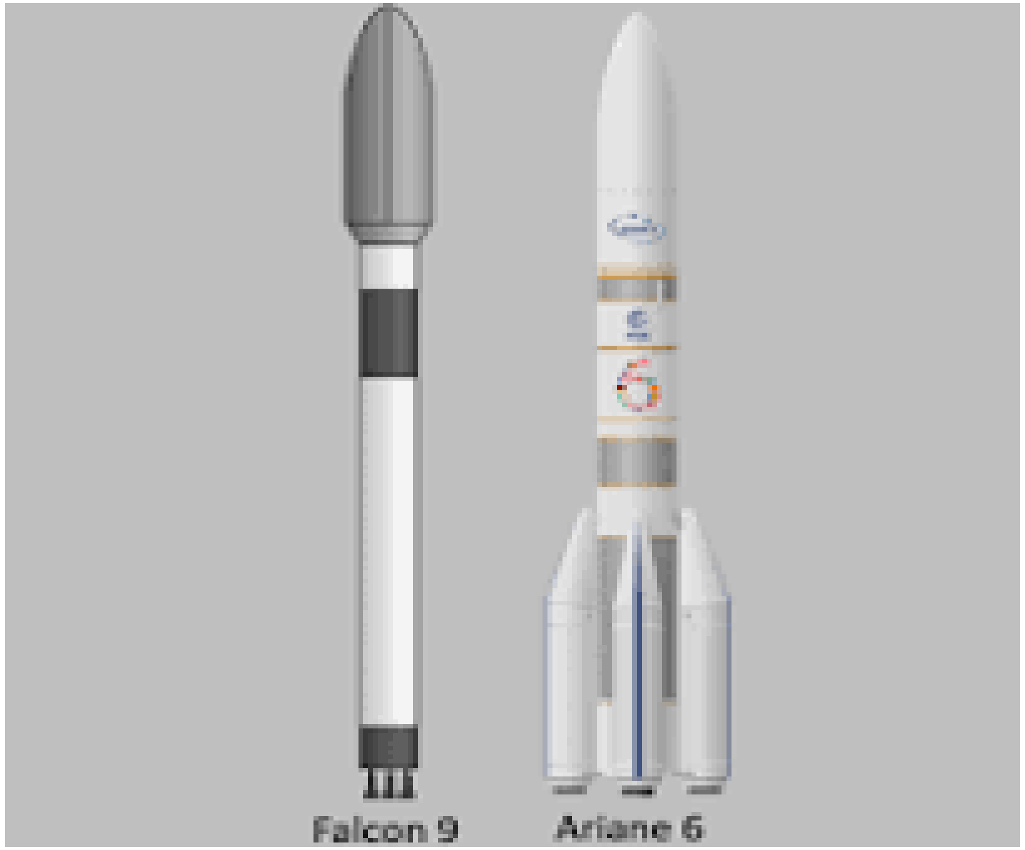
- In the 1970s and 1980s, non-military commercial satellites began to be launched in large numbers.
- From 2000 to 2017, Ariane 5 (ESA - Arianespace) dominated the market with: lower cost, high payload mass, reliability.
- In terms of cost, China, India, Russia are tough competitors with less reliable classic launchers like LongMarch, Proton, Soyuz-2, PSLV. New players like BlueHorizon are still in early stage.
- Since 2017, SpaceX Falcon9/Falcon9 Heavt is increasingly dominating the market with better cost per kg thanks to the reusable booster concept.
- Competition may heat up again with the advent of Ariane 6 and SpaceX Starship.

Introduction

Falcon 9/Falcon9 heavy v. Ariane 6

- Falcon9 / Falcon 9 Heavy can easily compete with new Ariane 6 (2 or 4 boosters) in terms of:
    - Cost per kg
    - Payload mass
- SpaceX Starship will further crush the competition in both domains.
- Falcon9 and Starship costs are way lower due to reuse of boosters
- Ariane-6 will be a classic launcher for reliability purposes
- Falcon9 reliability can’t compete with Ariane-5G record
- Falcon9 booster recovery success rate is still an issue.

Introduction



Section 1

Methodology

Methodology

Executive Summary

1. Data collection from open Data base and Wikipedia (Falcon9, Ariane-5).

2. Data wrangling.

3. Exploratory data analysis using SQL query and visualization of correlation between parameters.

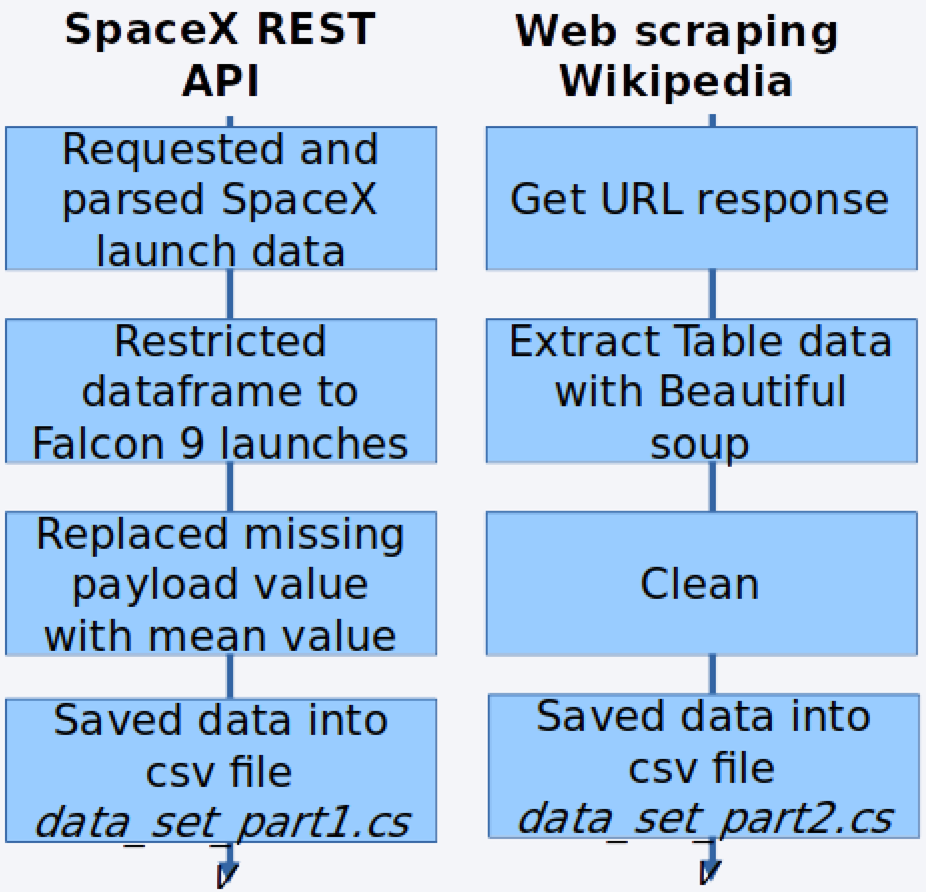
4. Visual analytics: launch sites with Folium, success rates with Plotly Dash.

5. Classification Models development and validations. Selection of best predictive model

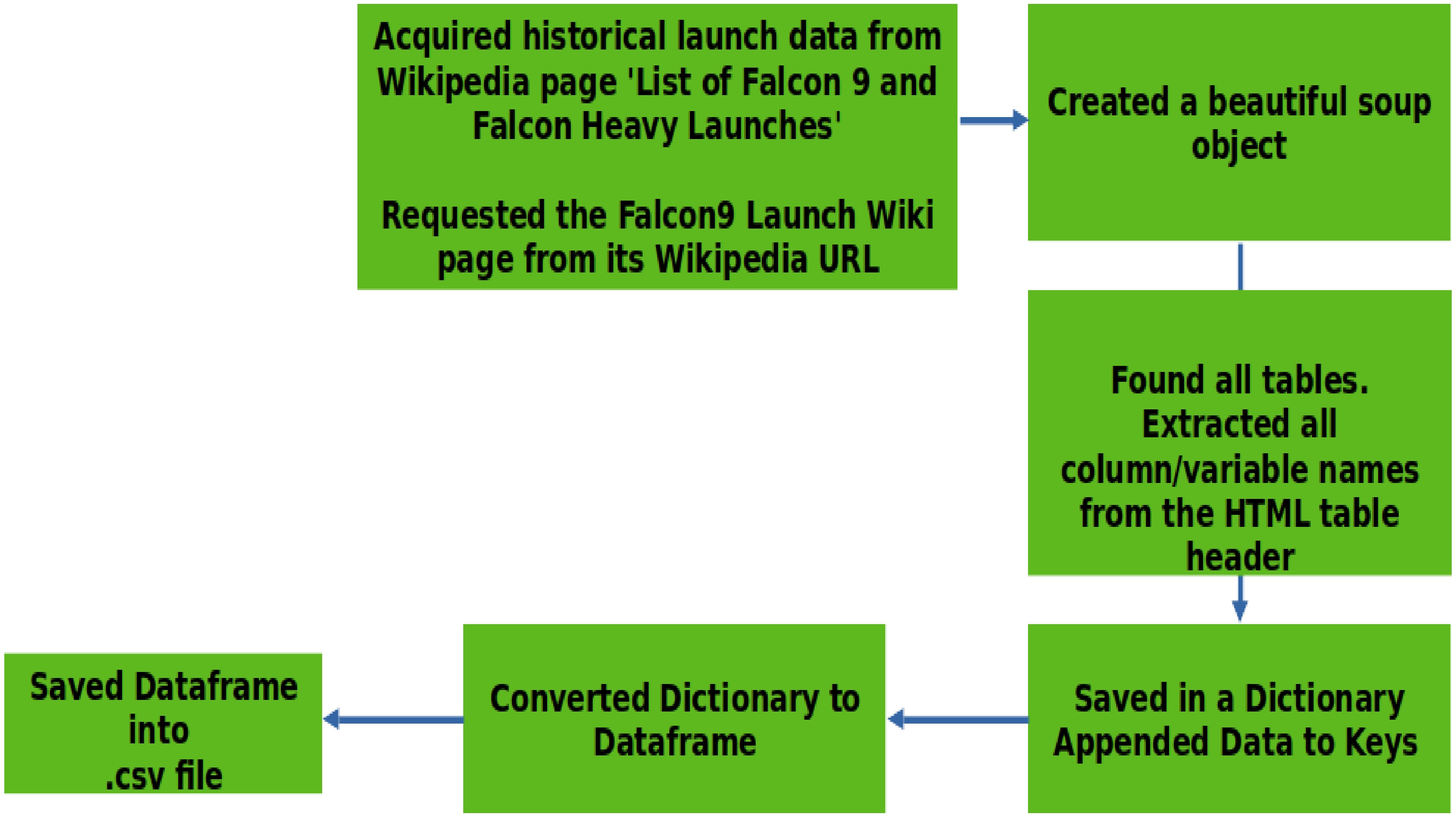
Data Collection

Data was collected from:

- open source SpaceX REST API
- webscraping Falcon9 launch data in Wikipedia



Data Collection – SpaceX API



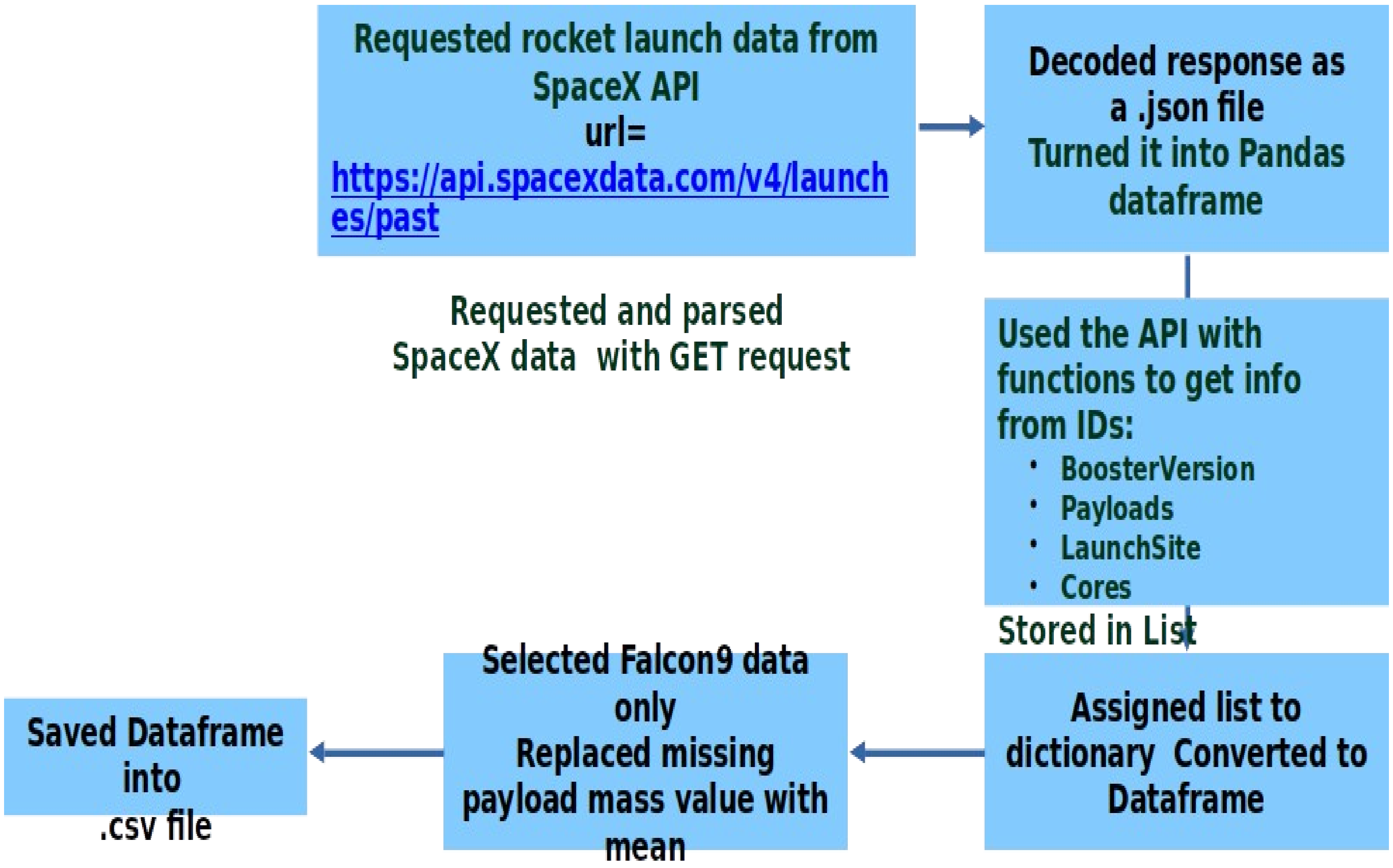
Data Collection - Scraping

Preliminary Stage

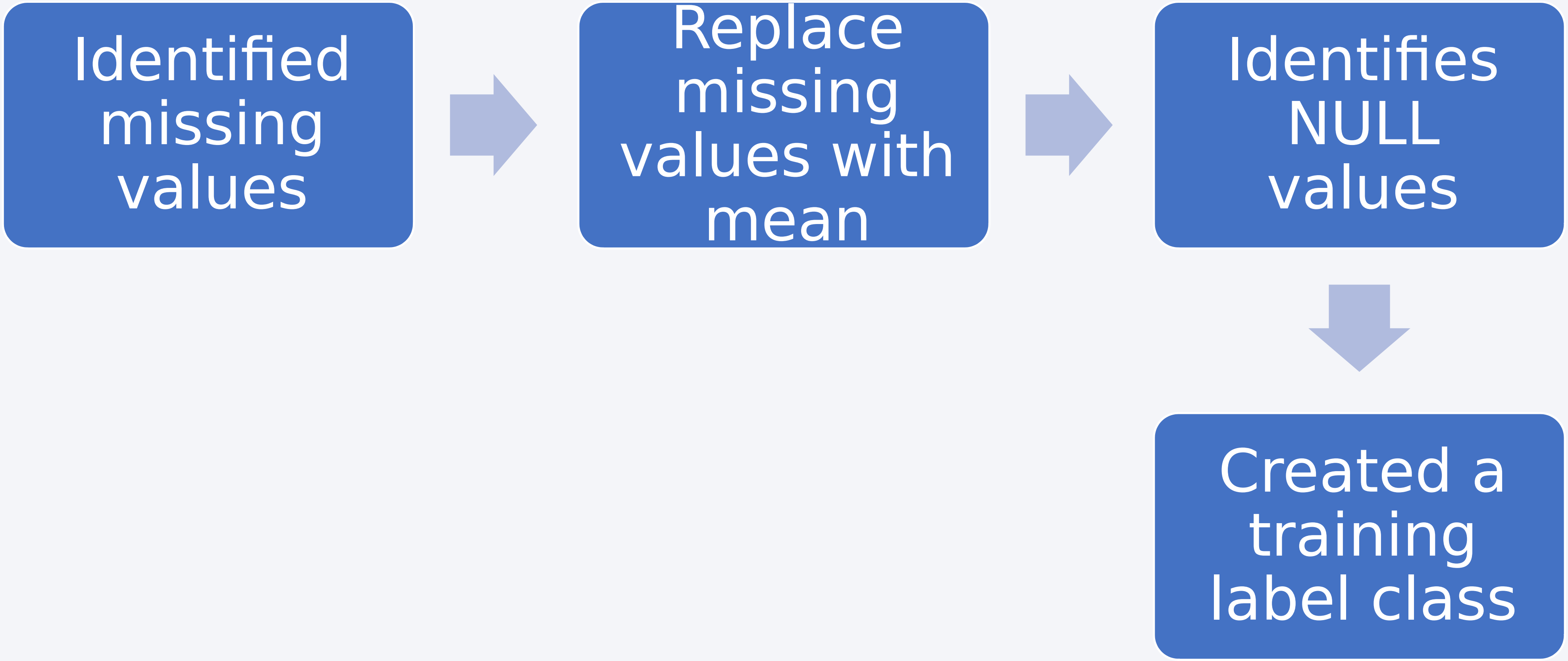
We defined a series of functions helping the use the API to extract information

- getBoosterVersion(data)
- getLaunchSite(data)
- getPayloadData(data)
- getCoreData(data)

Data Collection - Scraping



Data Wrangling



EDA with Data Visualization

Scatter Plots

- Flight Number VS. Payload Mass
- Flight Number VS. Launch Site
- Payload VS. Launch Site
- Orbit VS. Flight Number
- Payload VS. Orbit Type
- Orbit VS. Payload

Bar Graphs

- Mean VS. Orbit

Line Graphs

- Success Rate VS. Year

EDA with SQL

Performed SQL queries to gather information about the dataset. Some of the questions we were asked about the data is

•Displaying the names of the unique launch sites in the space mission

•Displaying 5 records where launch sites begin with the string 'CCA’

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

•Displaying average payload mass carried by booster version F9 v1.1

• Listing the date where the successful landing outcome in drone ship was achieved.

• Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000

• Listing the total number of successful and failure mission outcomes

• Listing the names of the booster versions which have carried the maximum payload mass.

• Listing the records which will display the month names, successful landing outcomes in ground pad ,booster versions, launch site for the months in year 2017

• Ranking the count of successful landing outcomes between the date 2010-06-04 and 2023-03-20 in descending order.

Build an Interactive Map with Folium

- Launch success rate may depend on the location and proximity of a launch site.
- Folium Interactive Map was used for visualizing and analyzing SpaceX Launch Sites.
- Used Interactive mapping library called Folium , Identified all SpaceX launch sites on a map: Florida, California , Included longitude and latitude info.
- Identified successful/failed launches for each site on map Calculated the distance between a launch site (CCAFS_SLC40 in Cape Canaveral, FL) and:
    - Closest coastline
    - Closest high traffic density railway:
    - Florida East Coast Railway
    - Closest high traffic density highway: Interstate I95
    - Closest high density urban area: Orlando (FL)

Build a Dashboard with Plotly Dash

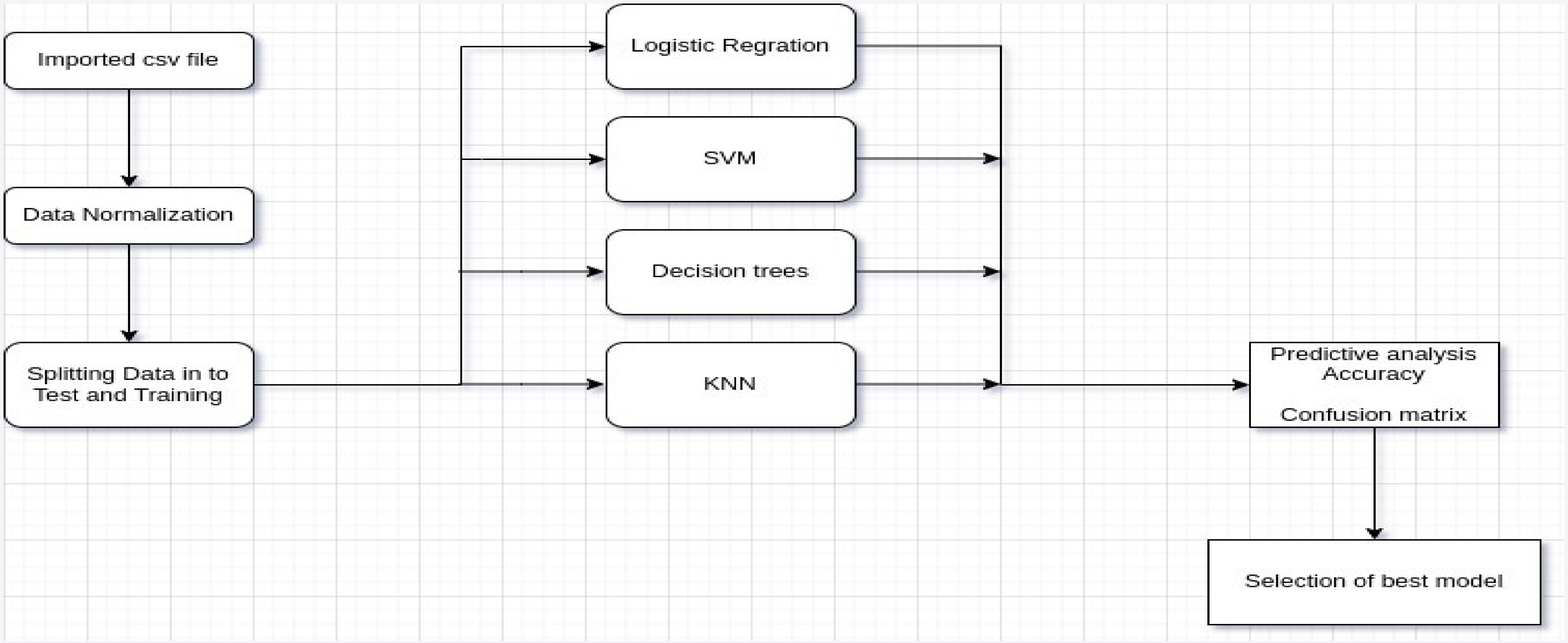
We built an interactive dashboard with Plotly including:

- Dropdown menu for selecting launch sites
- Pie charts displaying success rate.
- Scatter chart displaying launch site, payload mass, success/failure
- Range slider for selecting range of payload mass (kg).

For analyzing SpaceX launch records features:

- site with largest successful launches.
- site with highest launch success rate
- payload range(s) with highest launch success rate
- payload range(s) with lowest launch success rate
- F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) with highest launch success rate.

Predictive Analysis (Classification)

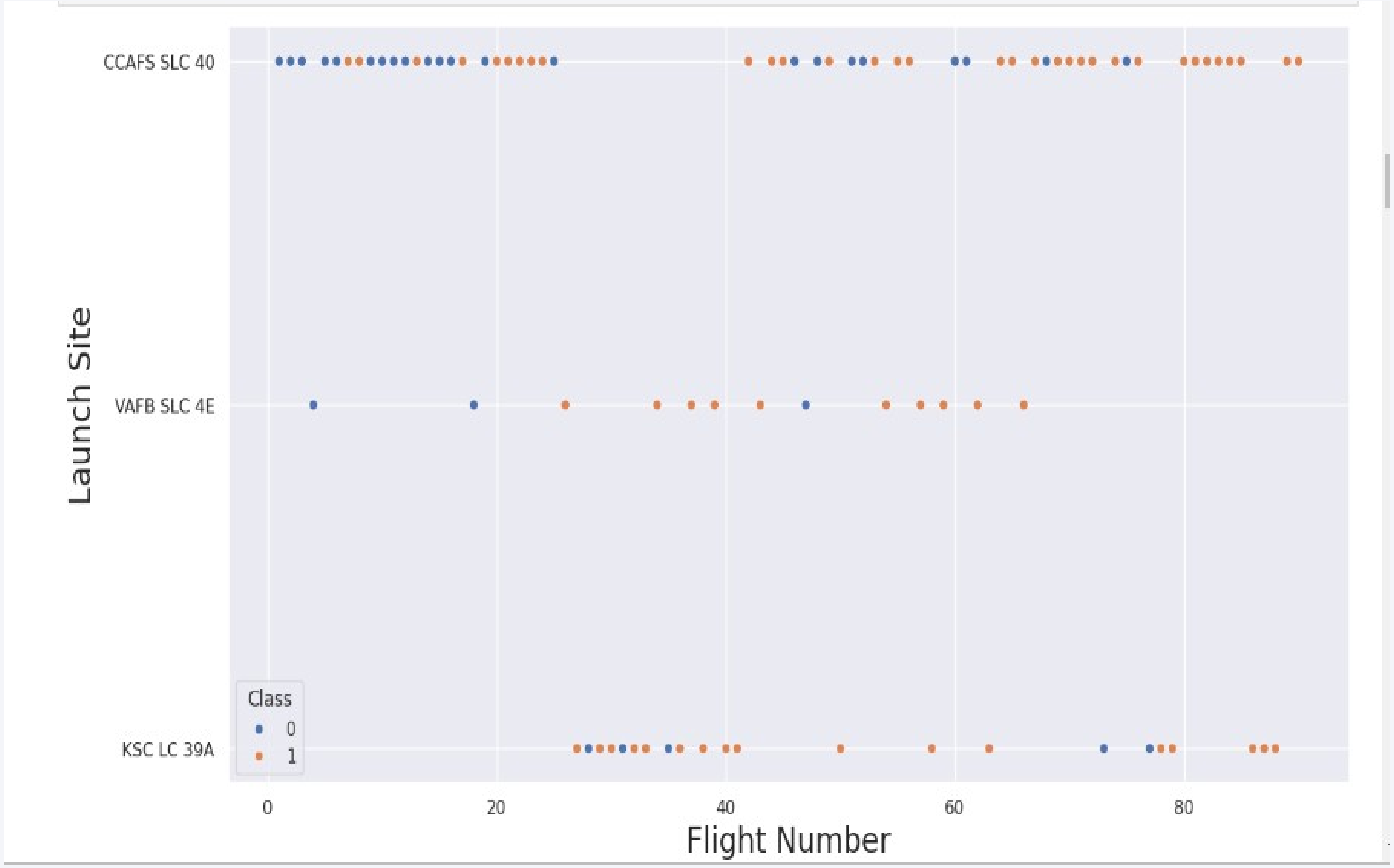


Section 2

Insights drawn from EDA

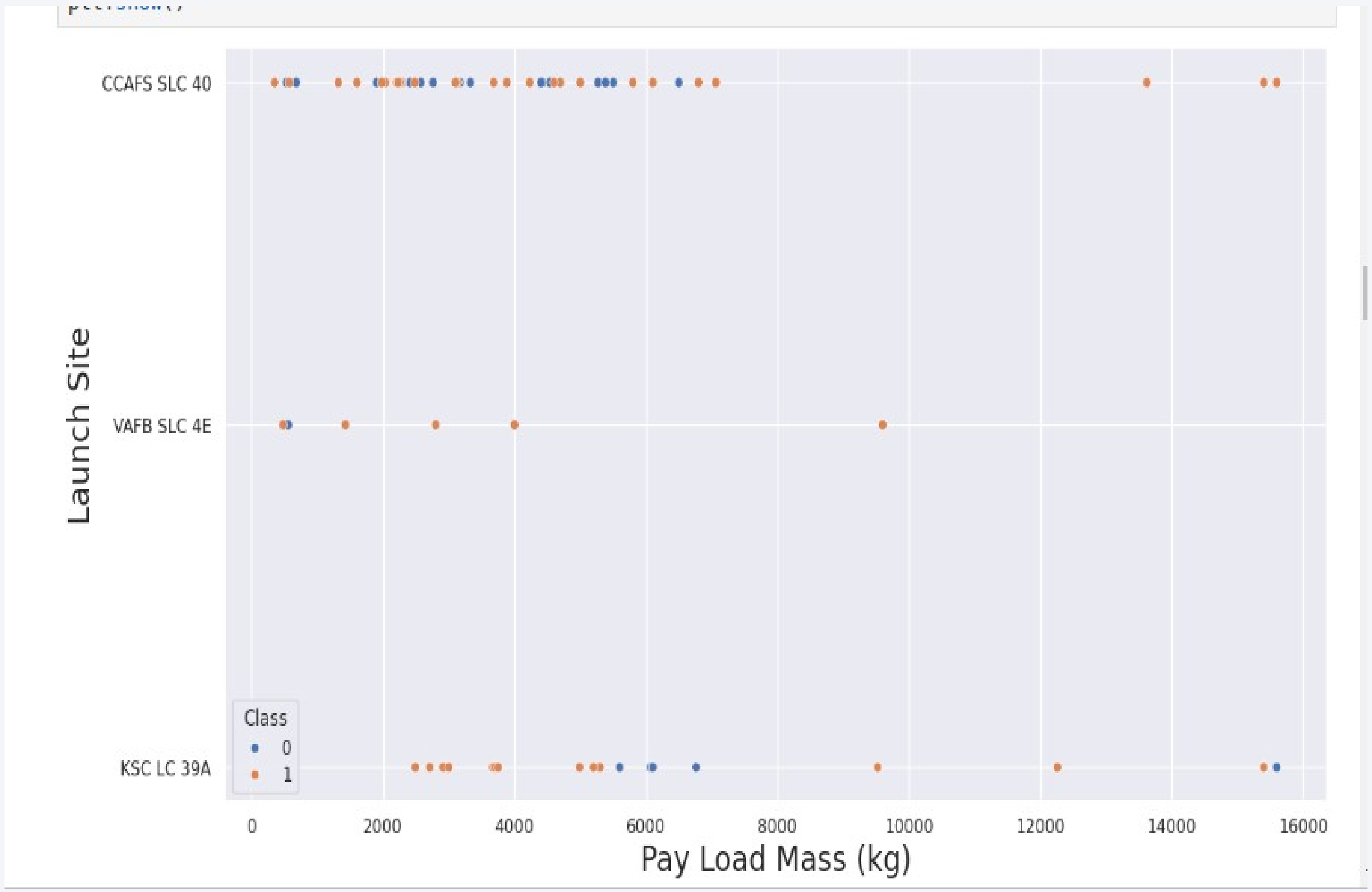
Flight Number vs. Launch Site

- CCAFS-SLC 40 is the most used launch site.
- CCAFS-SLC 40 has most of failures , particularly in the early stage of Falcon9 project



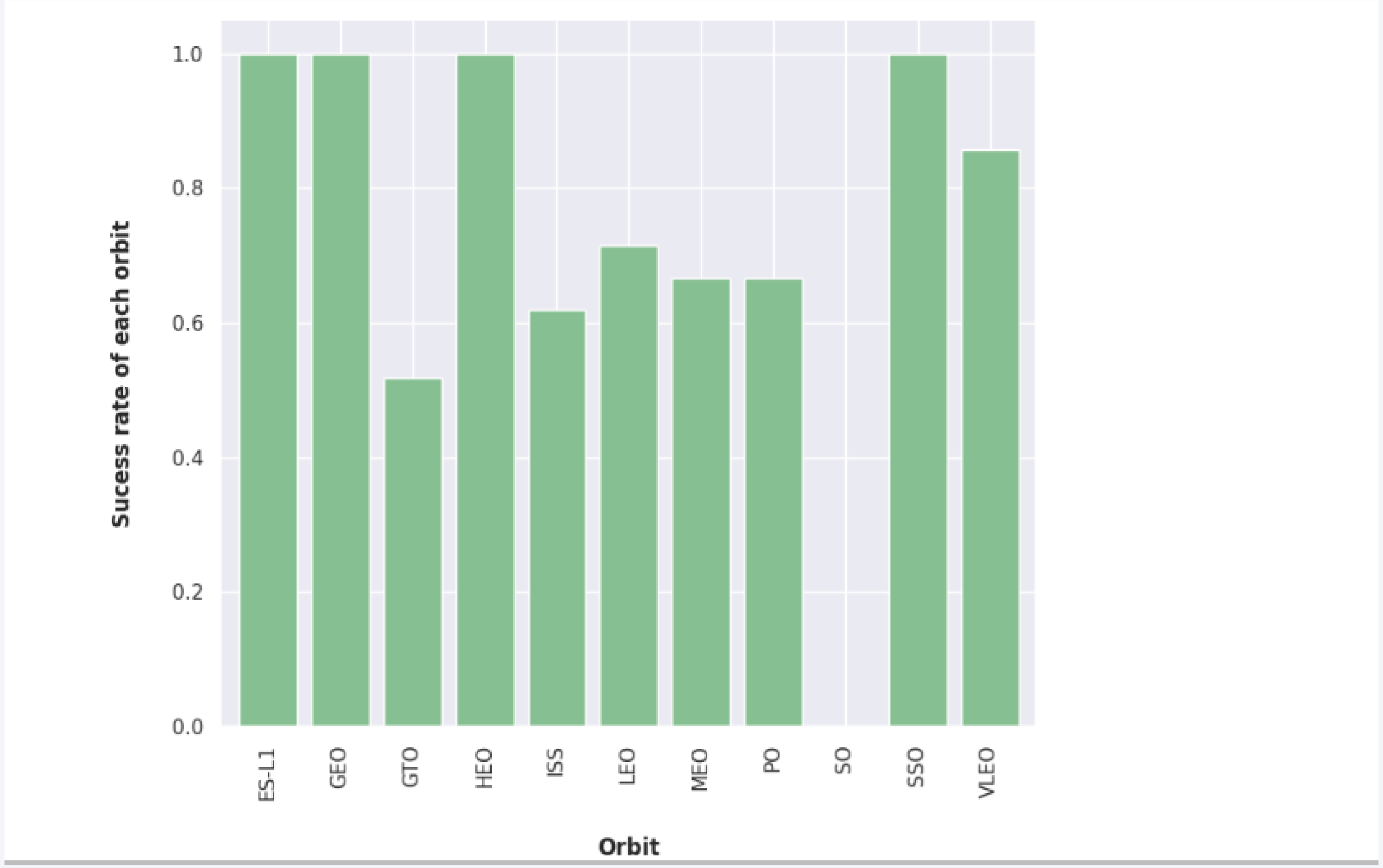
Payload vs. Launch Site

- Given Falcon9 specifications, heavy payloads are sent to low/medium orbits
- It looks like the percentage of failures is lower for heavy payload.
- More information is needed for extracting some correlation: success rate v. payload/orbit



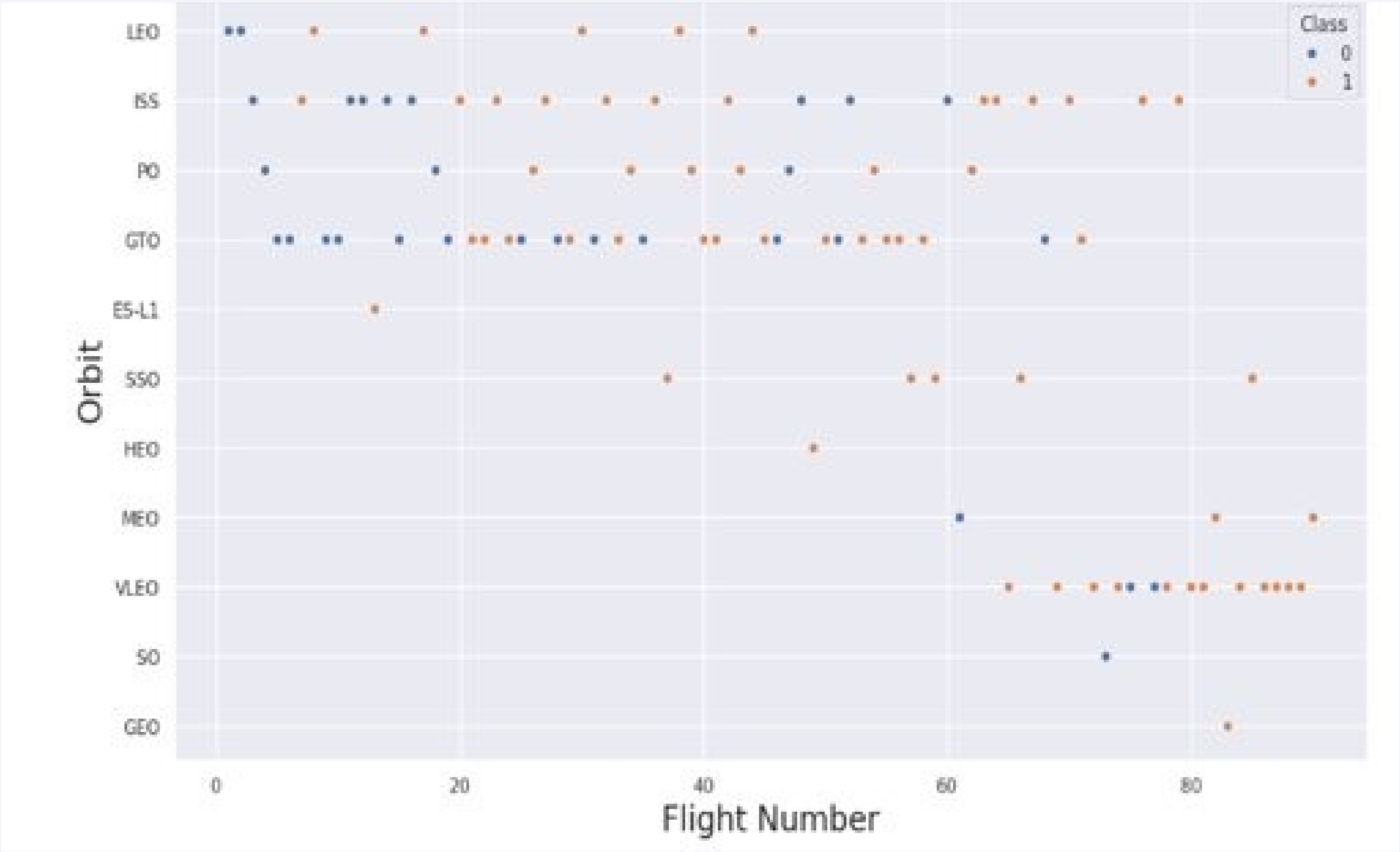
Success Rate vs. Orbit Type

- GTO sees the lowest success rate.
- SSO (polar low orbit) the highest one.
- Success rate may strongly depend on both: payload mass and orbit.



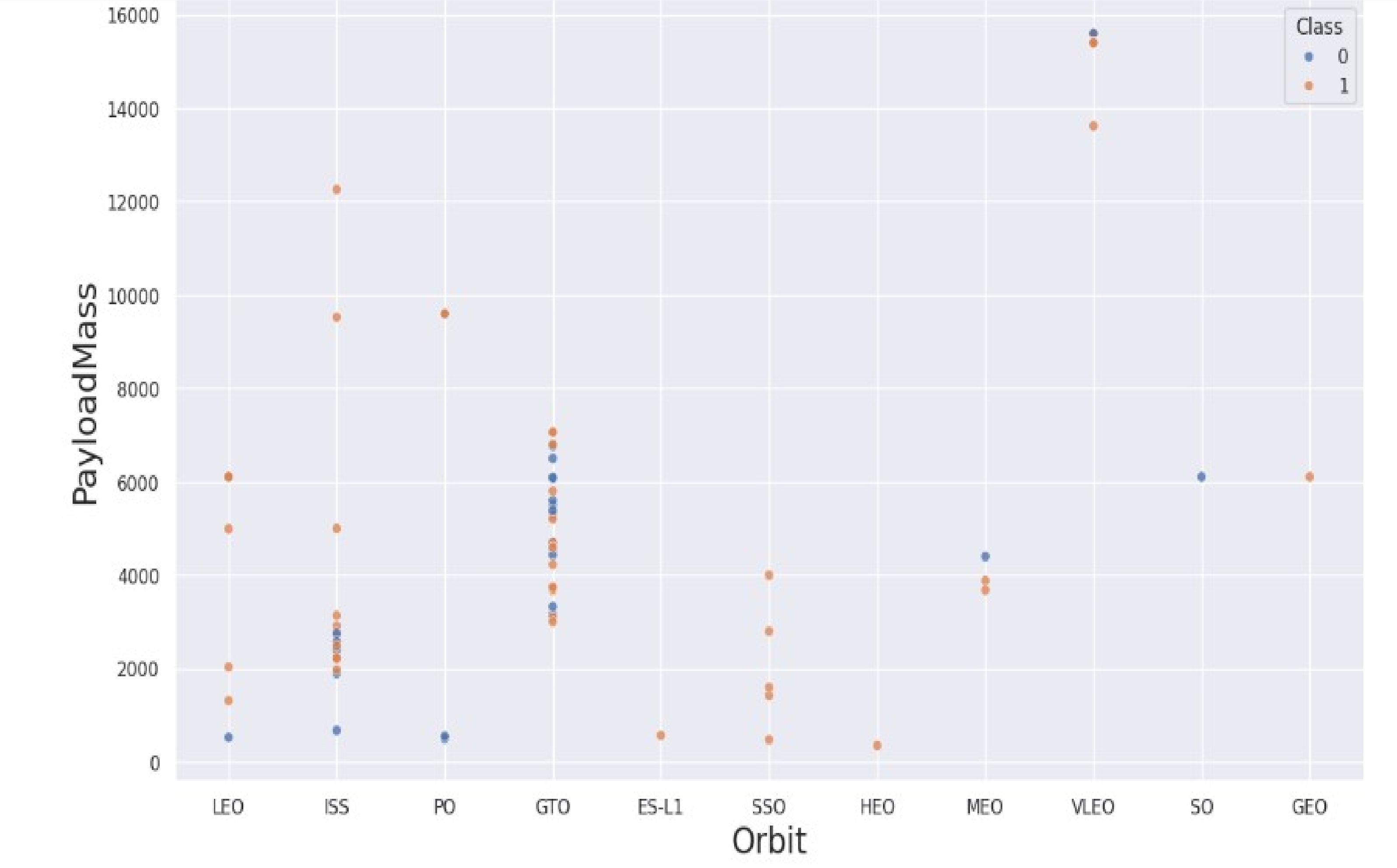
Flight Number vs. Orbit Type

- The number of flights for: GEO, SO, HEO, ESL-1, MEO is not significant for concluding about success rate.
- It looks like GTO are higher risk missions, low orbits are lower risk.



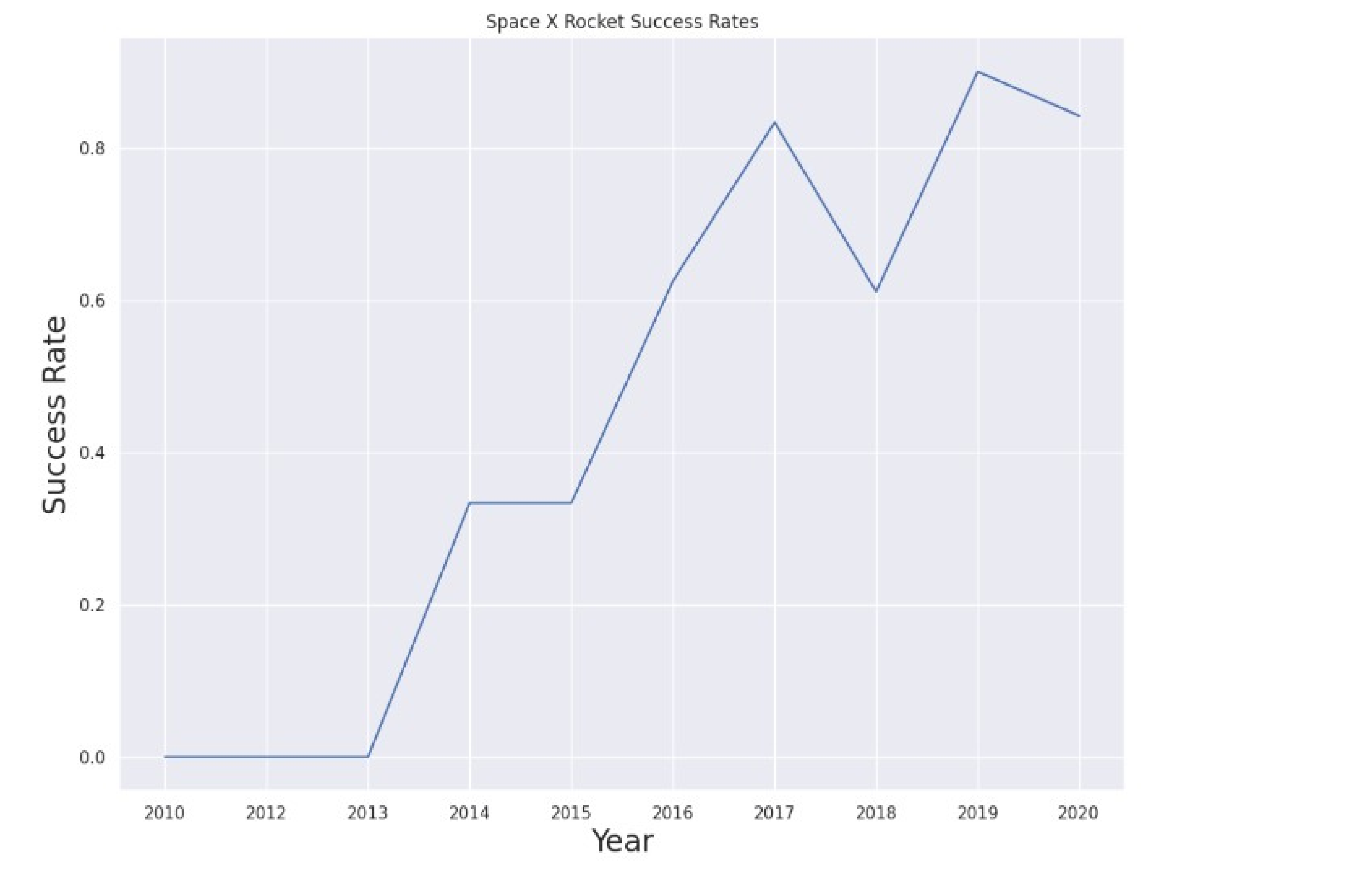
Payload vs. Orbit Type

- Maximum success rate with: low orbit except (ISS) and low payload mass
- Between 2000 and 7500 kg, success rate seems to be evenly distributed for GTO.
- Independently of payload mass, GTO is a risky “orbit” affecting missions success rate. Falcon 9 reliability improves over time, but there are still recent failed booster recovery after GTO launches



Launch Success Yearly Trend

- Falcon 9 reliability significantly improves over time .
- Falcon9 average booster recovery success rate si 66%.
- Success rate currently sufficient for SpaceX financial viability.



All Launch Site Names

```
[8]: %sql SELECT DISTINCT LAUNCH_SITE as "Launch_Sites" FROM SPACEXTBL;

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

[8]:

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

Launch Site Names Begin with 'CCA'

```
[10]: %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;

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

[10]:

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

```
[11]: %sql SELECT SUM(PAYLOAD_MASS__KG_) AS "Total Payload Mass by NASA (CRS)" FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';

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

[11]:

Total Payload Mass by NASA (CRS)
45596

Average Payload Mass by F9 v1.1

```
[12]: %sql SELECT AVG(PAYLOAD_MASS__KG_) AS "AVERAGE" FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';

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

[12]:

AVERAGE
2928.4

First Successful Ground Landing Date

```
[13]: %sql SELECT MIN(DATE) AS "First Succesful Mission Outcome" FROM SPACEXTBL WHERE MISSION_OUTCOME = 'Success';

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

[13]:

First Succesful Mission Outcome
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

[14]:
```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE MISSION_OUTCOME = 'Success' \
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

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

[14]:

Booster_Version
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1014
F9 v1.1 B1016
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1030
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 FT B1032.2
F9 B4 B1040.2
F9 B5 B1046.2
F9 B5 B1047.2

Total Number of Successful and Failure Mission Outcomes

[17]:
```
%sql SELECT sum(case when MISSION_OUTCOME LIKE '%Success%' then 1 else 0 end) AS "Successful Mission", \
    sum(case when MISSION_OUTCOME LIKE '%Failure%' then 1 else 0 end) AS "Failure Mission" \
FROM SPACEXTBL;
```

 * sqlite:///my_data1.db
Done.

[17]:

Successful Mission	Failure Mission
100	1

Boosters Carried Maximum Payload

[18]:
```
%sql SELECT DISTINCT BOOSTER_VERSION AS "Booster Versions which carried the Maximum Payload Mass" FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

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

[18]:

Booster Versions which carried the Maximum Payload Mass
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

```
[19]: %sql SELECT  (DATE,6,2) as "Month", BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE (DATE,0,5) = '2015' AND \
MISSION__OUTCOME = 'Failure';

 * sqlite:///my_data1.db
(sqlite3.OperationalError) row value misused
[SQL: SELECT  (DATE,6,2) as "Month", BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE (DATE,0,5) = '2015' AND  MISSION__OUTCOME = 'Failure';]
(Background on this error at: http://sqlalche.me/e/13/e3q8)
```

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

[20]:
```
%sql SELECT MISSION_OUTCOME  as "Landing Outcome", COUNT(MISSION_OUTCOME ) AS "Total Count" FROM SPACEXTBL \
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' \
GROUP BY  MISSION_OUTCOME \
ORDER BY COUNT(MISSION_OUTCOME) DESC ;
```

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

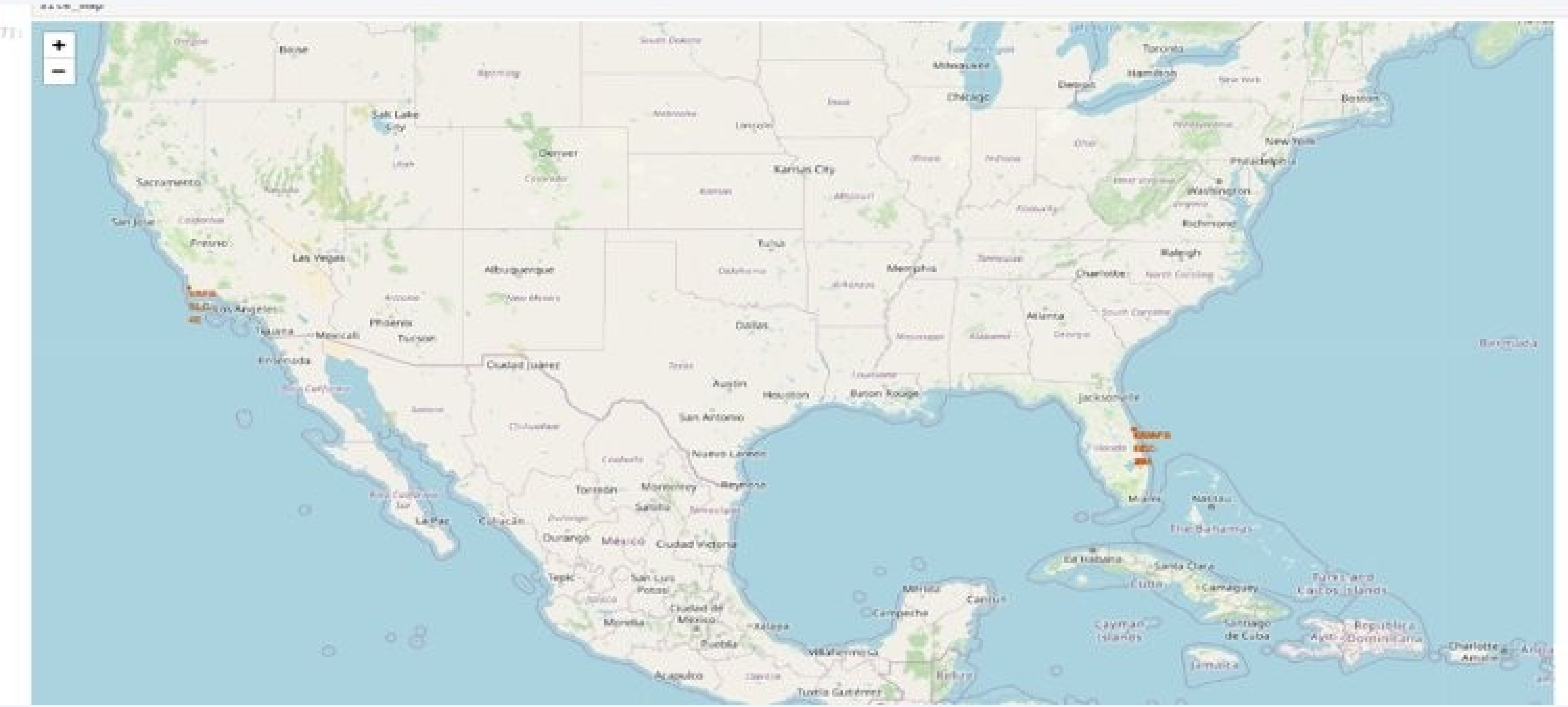
[20]:

Landing Outcome	Total Count
Success	30
Failure (in flight)	1

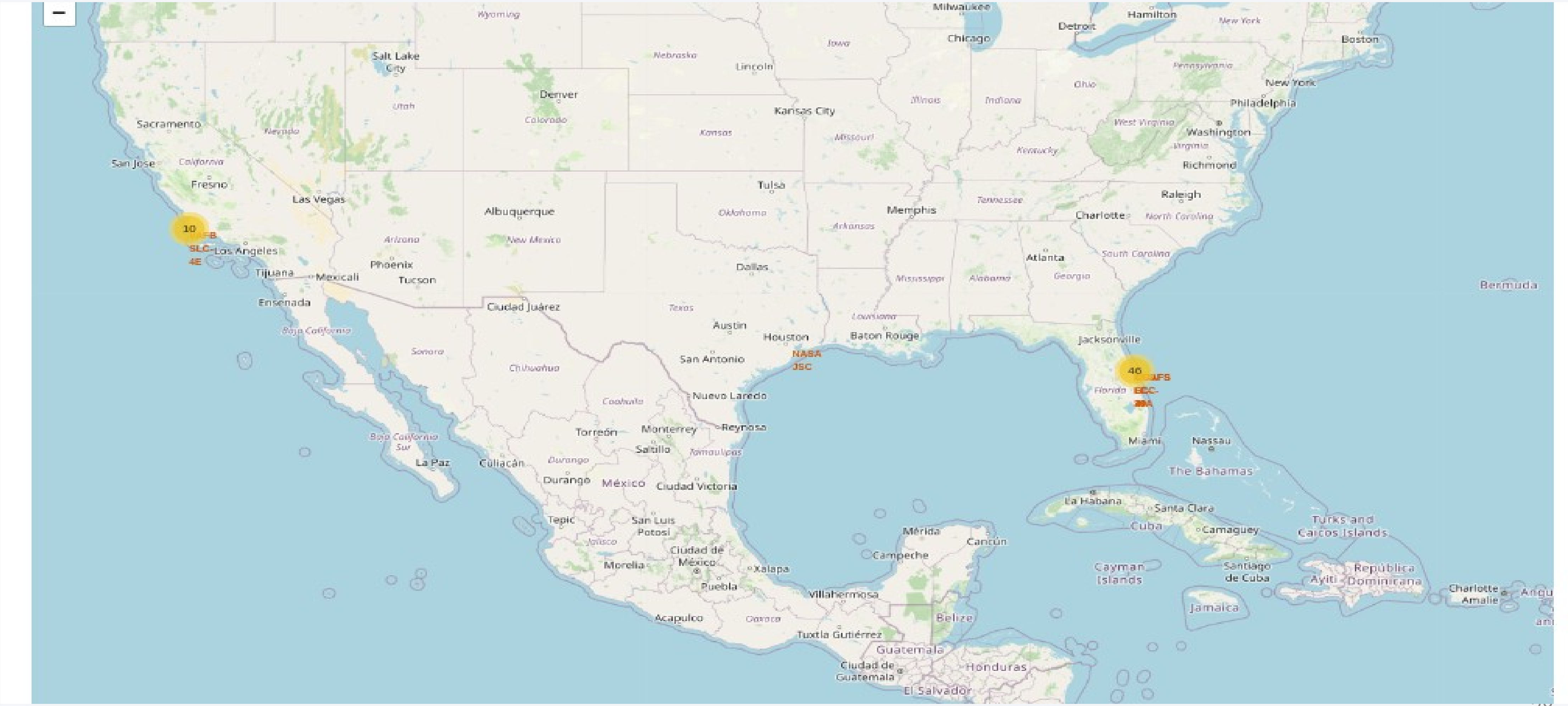
Section 3

Launch Sites Proximities Analysis

Folium Map 1

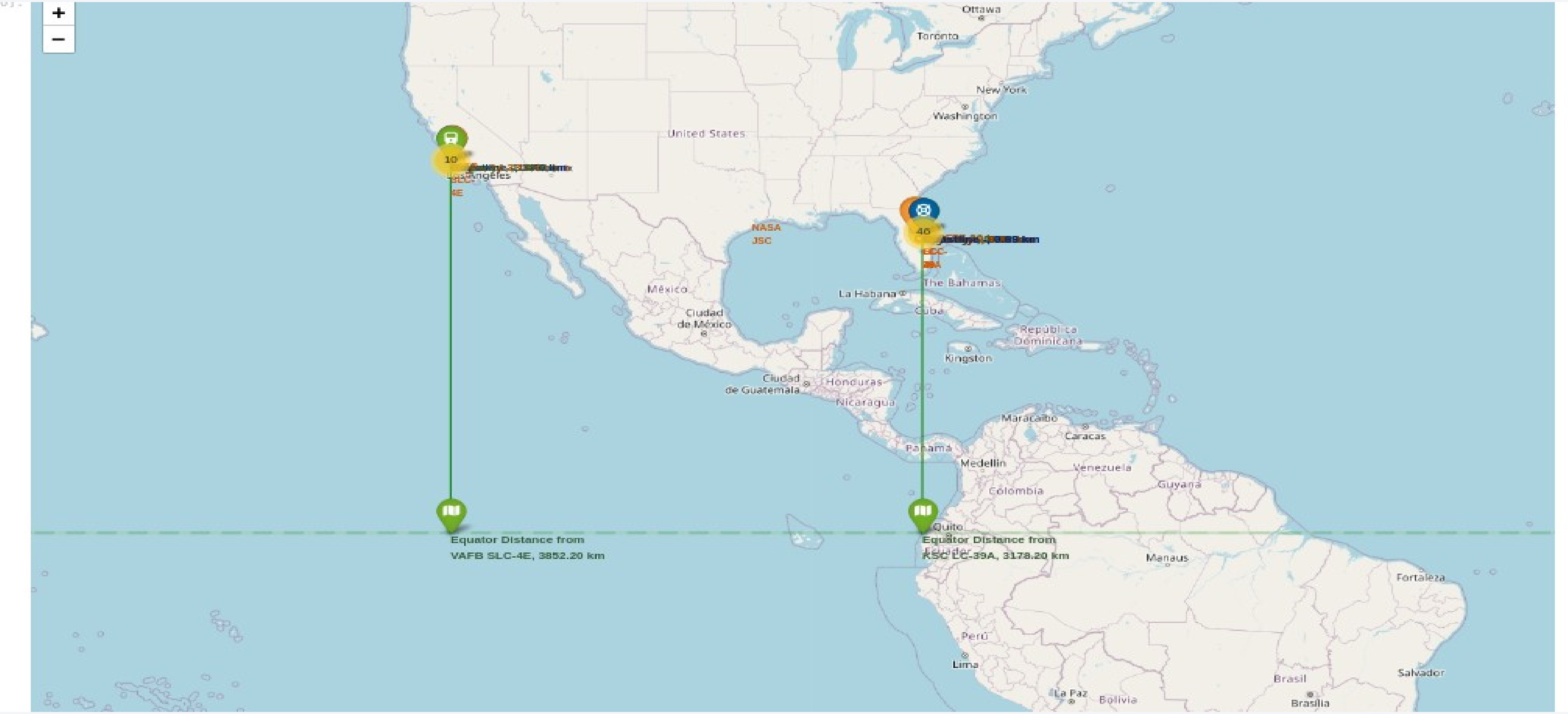


Folium Map 2



10
SLC-4E
Los Angeles
46
NASA JSC
Sacramento
San Jose
Fresno
Las Vegas
Salt Lake City
Denver
Albuquerque
Phoenix
Tucson
Tijuana
Mexicali
Ensenada
Ciudad Juárez
Dallas
Austin
Houston
San Antonio
Baton Rouge
Kansas City
Lincoln
Tulsa
Memphis
Chicago
Milwaukee
Detroit
Hamilton
Boston
New York
Philadelphia
Washington
Richmond
Raleigh
Charlotte
Atlanta
Jacksonville
Miami
Nassau
The Bahamas
Bermuda
La Habana
Santa Clara
Cuba
Camaguey
Santiago de Cuba
Cayman Islands
Jamaica
Turks and Caicos Islands
República Dominicana
Ayiti
Nuevo Laredo
Reynosa
Monterrey
Saltillo
Torreón
Ciudad Victoria
Durango
México
La Paz
Culiacán
Tepic
San Luis Potosí
Ciudad de México
Morelia
Puebla
Xalapa
Acapulco
Villahermosa
Campeche
Mérida
Cancún
Tuxtla Gutiérrez
Belize
Guatemala
Ciudad de Guatemala
Honduras
El Salvador

Folium Map 3



Ottawa
Toronto
New York
Washington
United States
NASA
JSC
The Bahamas
La Habana
Cuba
México
Ciudad
de México
República
Dominicana
Kingston
Ciudad
de Guatemala
Honduras
Nicaragua
Maracaibo
Caracas
Panamá
Medellín
Venezuela
Colombia
Guyana
Quito
Equator Distance from
VAFB SLC-4E, 3852.20 km
Equator Distance from
KSC LC-39A, 3178.20 km
Manaus
Fortaleza
Perú
Lima
La Paz
Bolivia
Brasil
Brasília
Salvador

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

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

- Point 1
- Point 2
- Point 3
- Point 4
- ...

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!