Winning Space Race with Data Science

Muhibur Rahman
Chowdhury Shafy
09/26/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data was collected using SpaceX REST API and Using Web scraping method.
  - Analyzed the success rate of the landing outcome of the first stage using Data Wrangling
  - Visualized the data using Matplotlib, Seaborn, Folium, Plotly and Dash.
  - Performed EDA using SQL, Calculated payload Mass, Successful Launches and % of Successful Launches.
  - Build Models to predict landing outcomes.
- Summary of all results
  - Launch success has improved over time, KSC LC 39A has the highest success rate, Most Launch sites are located near coastal area, and they’re near equator.
  - All models performed Almost similarly but Decision Tree model has a good success rate.

Introduction

- Project background and context
  - SpaceX, strives to make space travel affordable for everyone. They have successfully reused their First Stage of Falcon9 rocket lowering the cost of space travel by more than half.
- Problems you want to find answers
  - How payload mass, launch site, number of flights, and orbits affect first-stage landing success.
  - Rate of successful overtime
  - Best predictive model for successful landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using SpaceX Rest Api and Web Scrapping
- Perform data wrangling
  - Data was wrangling by filtering the data, handling missing value and applying one hot encoding to prepare the data for analysis and modeling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - To predict landing outcomes using classification models. Tune and evaluate models to find best model and parameters

Data Collection – SpaceX API

- Request Data from SpaceX Api
- Decode Response using .json_normalize() to convert to data frame
- Request information about launches
- Create Dictionary from the data
- Create data frame from the dictionary
- Filter data frame to contain only Falcon 9 launches
- Replace missing value of Payload Mass
- Export Data to a csv file

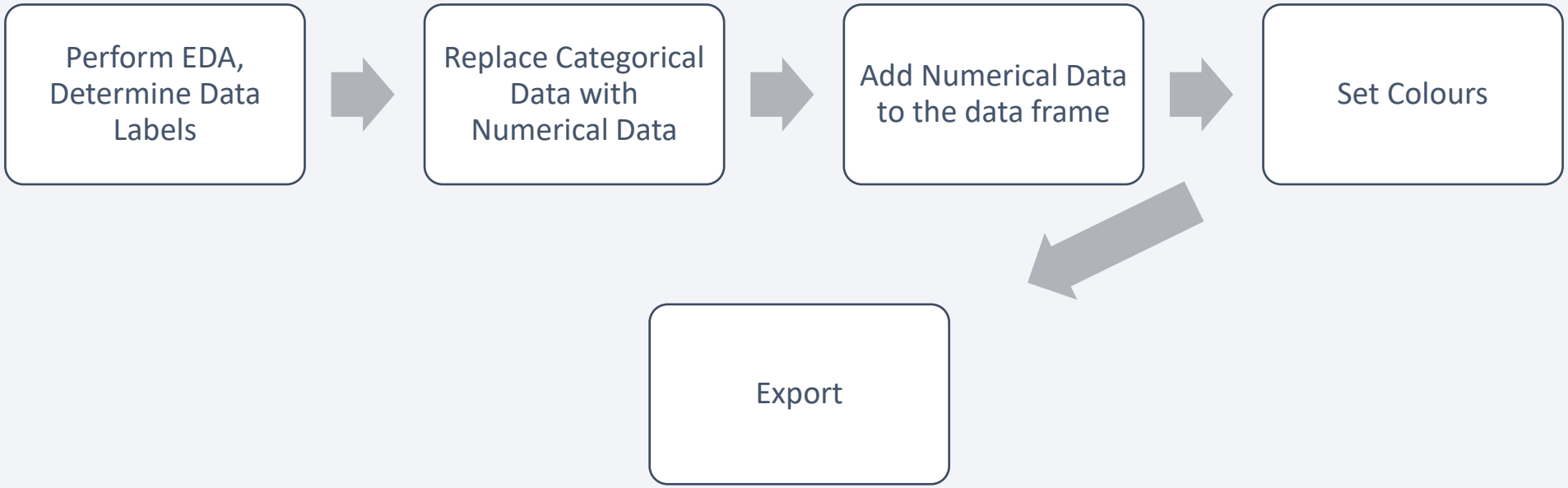
- https://github.com/mrshafy13/Applied_Data_Science_Capstone/blob/main/jupyter-labs-spacex-data-collection-api-v2.ipynb

Data Collection - Scraping

- Request data from Wikipedia
- Create BeautifulSoup object
- Extract Column Name
- Collect Data
- Create Dictionary
- Create Dataframe
- Export Data

- https://github.com/mrshafy13/Applied_Data_Science_Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling



- https://github.com/mrshafy13/Applied_Data_Science_Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling-v2.ipynb

EDA with SQL

- Names of unique launch sites
- 5 records where launch site begins with ‘CCA’
- Total payload mass carried by boosters launched by NASA(CRS)
- Average payload mass carried by booster version F9 v1.1
- Date of first successful landing on ground pad
- Names of boosters which had success landing on drone ship and have payload mass greater than 4000 bur less than 6000
- Total number of successful and failed missions.
- Names of booster versions which have carried the max payload
- https://github.com/mrshafy13/Applied_Data_Science_Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

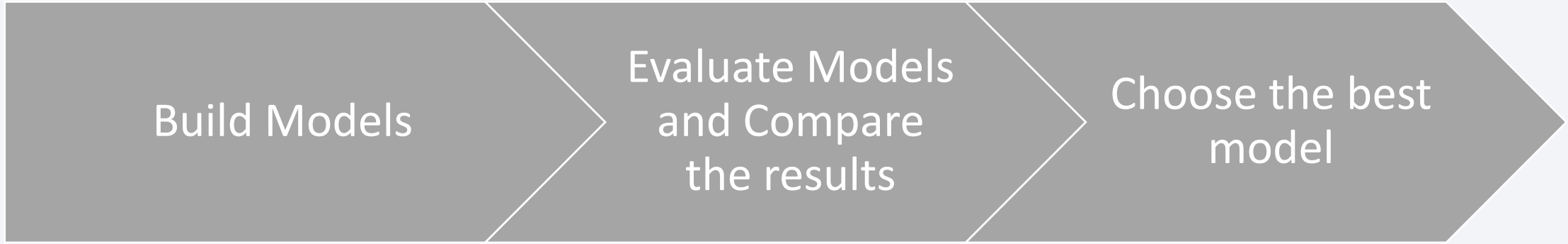
Build an Interactive Map with Folium

- Markers indicating Launch Sites
- Added colored circle at NASA JSC coordinate with a popup label showing it's description.
- Added circles at all the launch sites.
- Added colored markers of successful and unsuccessful launches to show which have higher success rate.
- Showed distance between launch sites and nearby relevant locations.
- https://github.com/mrshafy13/Applied_Data_Science_Capstone/blob/main/lab-jupyter-launch-site-location-v2.ipynb

Build a Dashboard with Plotly Dash

- Allow users to select all launch sites or a certain launch site
- Allow user to see successful and unsuccessful launches as a percent of total
- Allow user to select payload mass range
- Allow user to see the correlation between Payload and Launch Success
- https://github.com/mrshafy13/Applied_Data_Science_Capstone/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)



Results

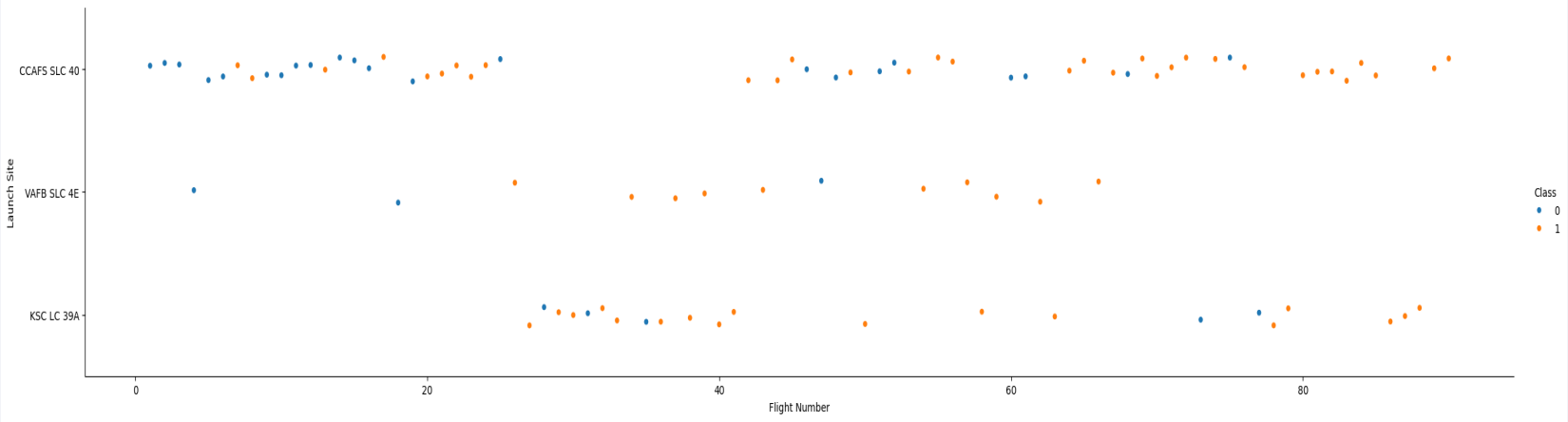
By performing various tests we saw that 3 machine learning models performed well. We can use them for our future predictions.
EDA Helped us with creating dataset with relevant data.

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

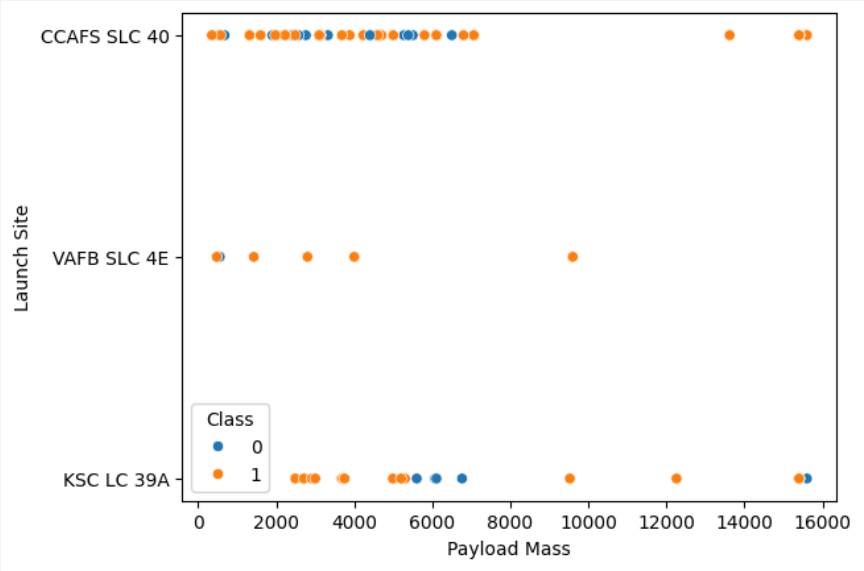
Flight Number vs. Launch Site



From here we can say CCAFS SLC 40 is the most used Launch site but compared to others it has less success rate.

Payload vs. Launch Site

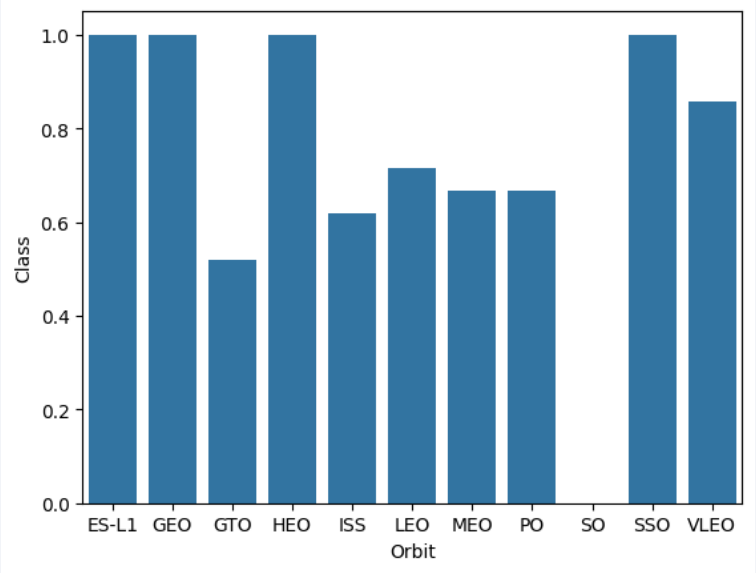
Payload vs. Launch Site



Here we can see VAFB SLC 4E launch site is used for launching rockets with lower payload

Success Rate vs. Orbit Type

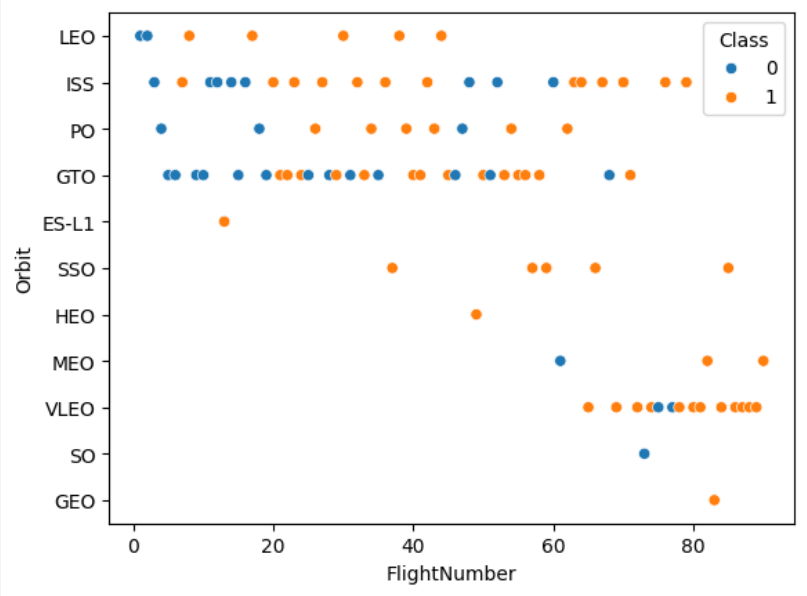
- Show a bar chart for the success rate of each orbit type



- Here we can see that ES-L1 HEO and SSO orbit has 100% success rate but GTO ISS LEO MEO and PEO has 50-75% success rate

Flight Number vs. Orbit Type

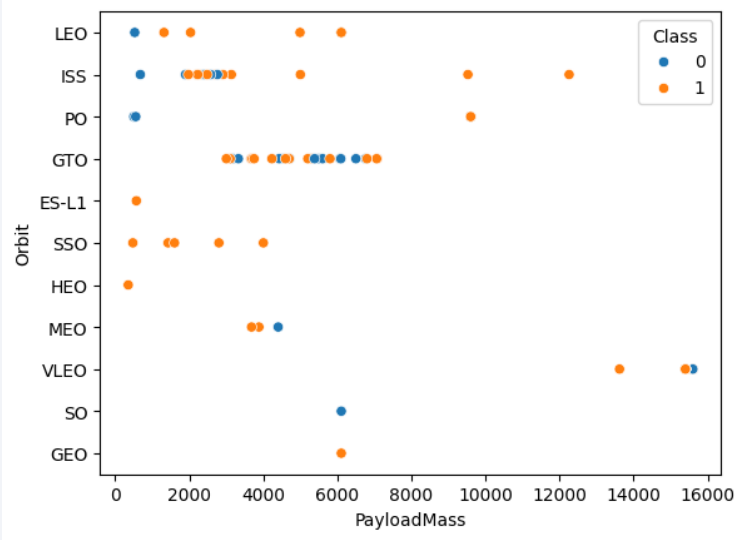
- Flight number vs. Orbit type



- We can see most of the flights are for GTO and ISS orbital but VLEO is getting popular.

Payload vs. Orbit Type

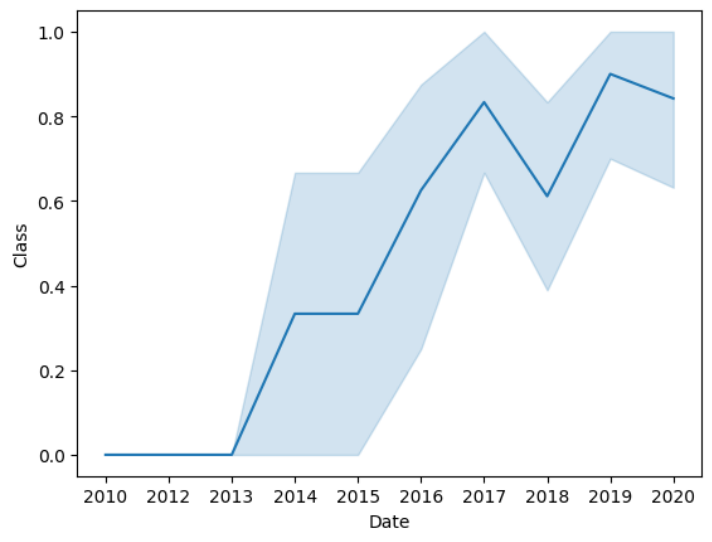
- Payload vs. orbit type



- From here we can say that Payload around 3000 to 8000 are good suit for GTO orbital.

Launch Success Yearly Trend

- yearly average success rate



- We can say that year by year the success rate has increased but a drop was on 2018 but was recovered by 2019

All Launch Site Names

- Find the names of the unique launch sites

Task 1

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

```
%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
```

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

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

- Using SQL query found all the unique launch site

Launch Site Names Begin with 'CCA'

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

Task 2

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

```
%sql SELECT * \
    FROM SPACEXTBL \
    WHERE LAUNCH_SITE LIKE'CCA%' LIMIT 5;
```

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

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

Five records where launch sites begin with ‘CCA’ was shown

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)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Total payload carried by boosters from NASA are shown.

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';
```

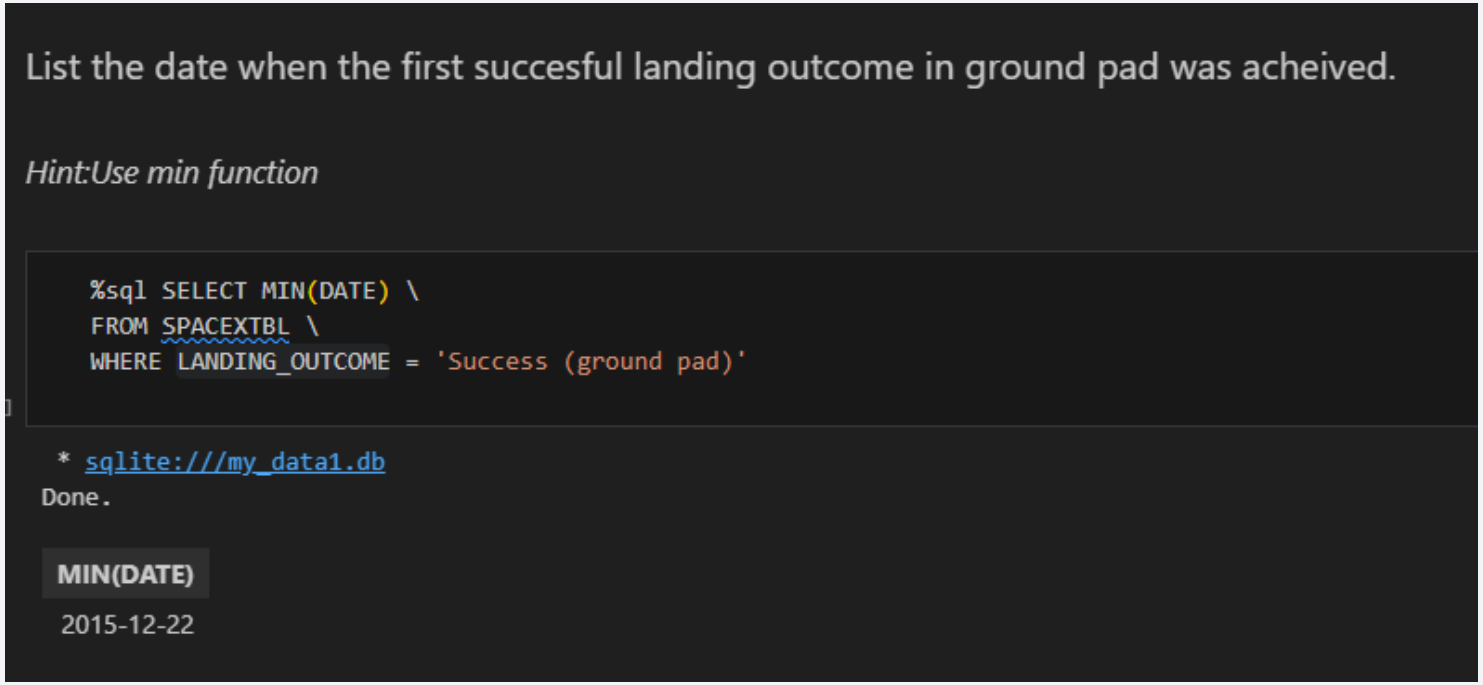
```
 * sqlite:///my_data1.db
Done.
```

AVG(PAYLOAD_MASS__KG_)
2928.4

- Here's the total payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

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



2015-12-22 was the first date of successful landing outcome in ground pad

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

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 PAYLOAD \
FROM SPACEXTBL \
WHERE LANDING_OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

Payload
JCSAT-14
JCSAT-16
SES-10
SES-11 / EchoStar 105

Here's the list and query of names of booster which have successfully landed on drone ship

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(*) as total_number \
FROM SPACEXTBL \
GROUP BY MISSION_OUTCOME;
```

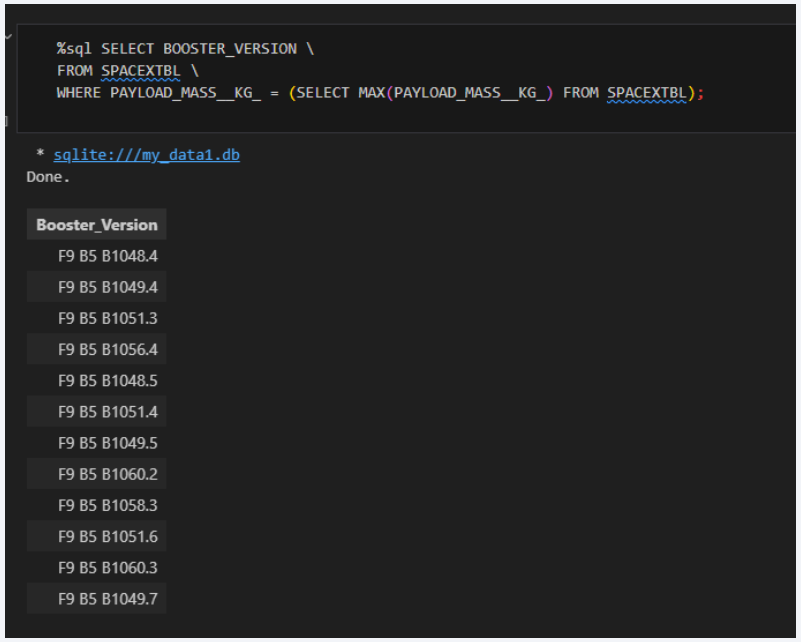
```
 * sqlite:///my_data1.db
Done.
```

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

- 98 of 101 mission was successful fully but some were successful with some Issues only 1 has failed

Boosters Carried Maximum Payload

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



```
%sql SELECT BOOSTER_VERSION \
FROM SPACEXTBL \
WHERE PAYLOAD_MASS__KG_ = (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

- List of the booster version are shown

2015 Launch Records

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

```
%sql SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, [LANDING_OUTCOME] \
FROM SPACEXTBL \
where [LANDING_OUTCOME] = 'Failure (drone ship)' and substr(Date,0,5)='2015';
```

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

month	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- Only 2 has failed to land

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

```
%sql SELECT [LANDING_OUTCOME], count(*) as count_outcomes \
FROM SPACEXTBL \
WHERE DATE between '2010-06-04' and '2017-03-20' group by [LANDING_OUTCOME] order by count_outcomes DESC;
```

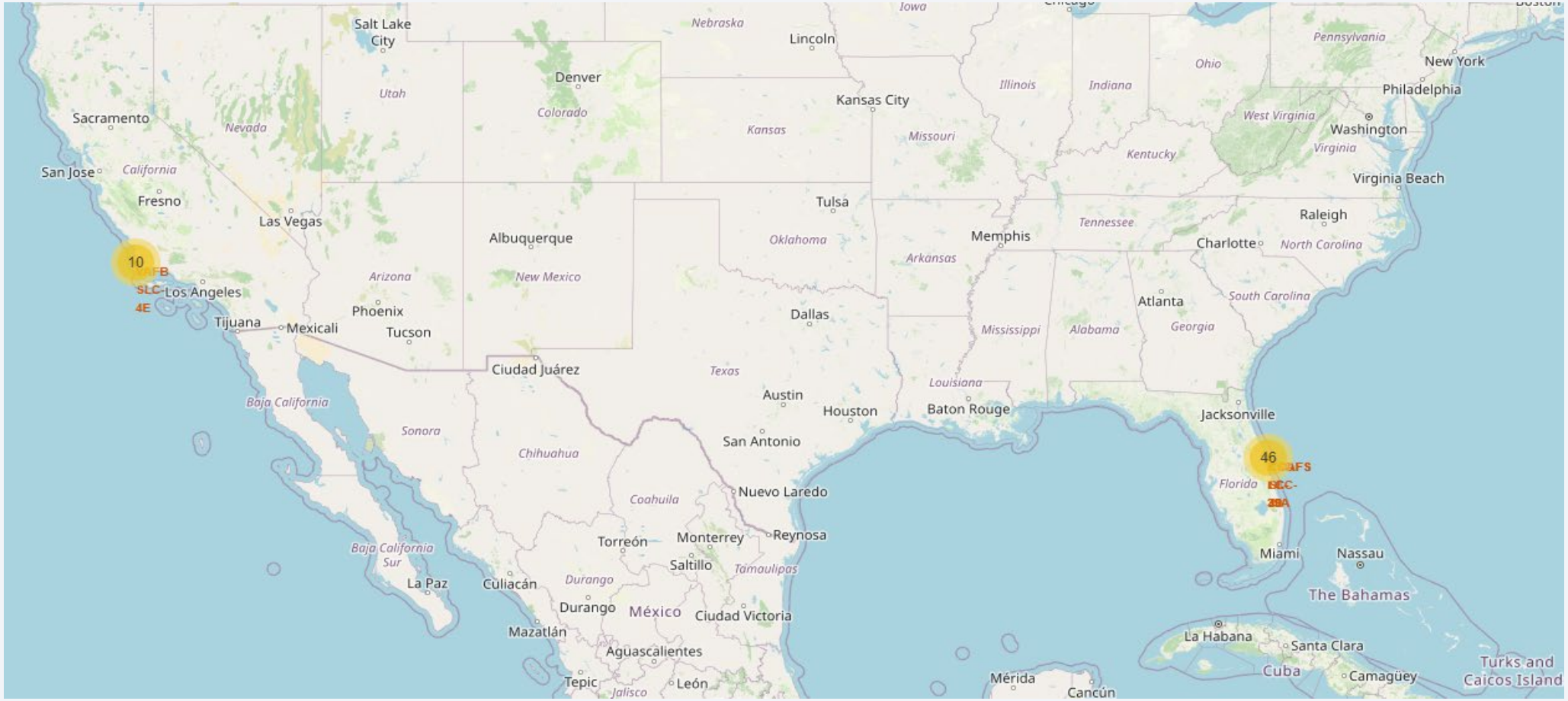
```
 * sqlite:///my_data1.db
Done.
```

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

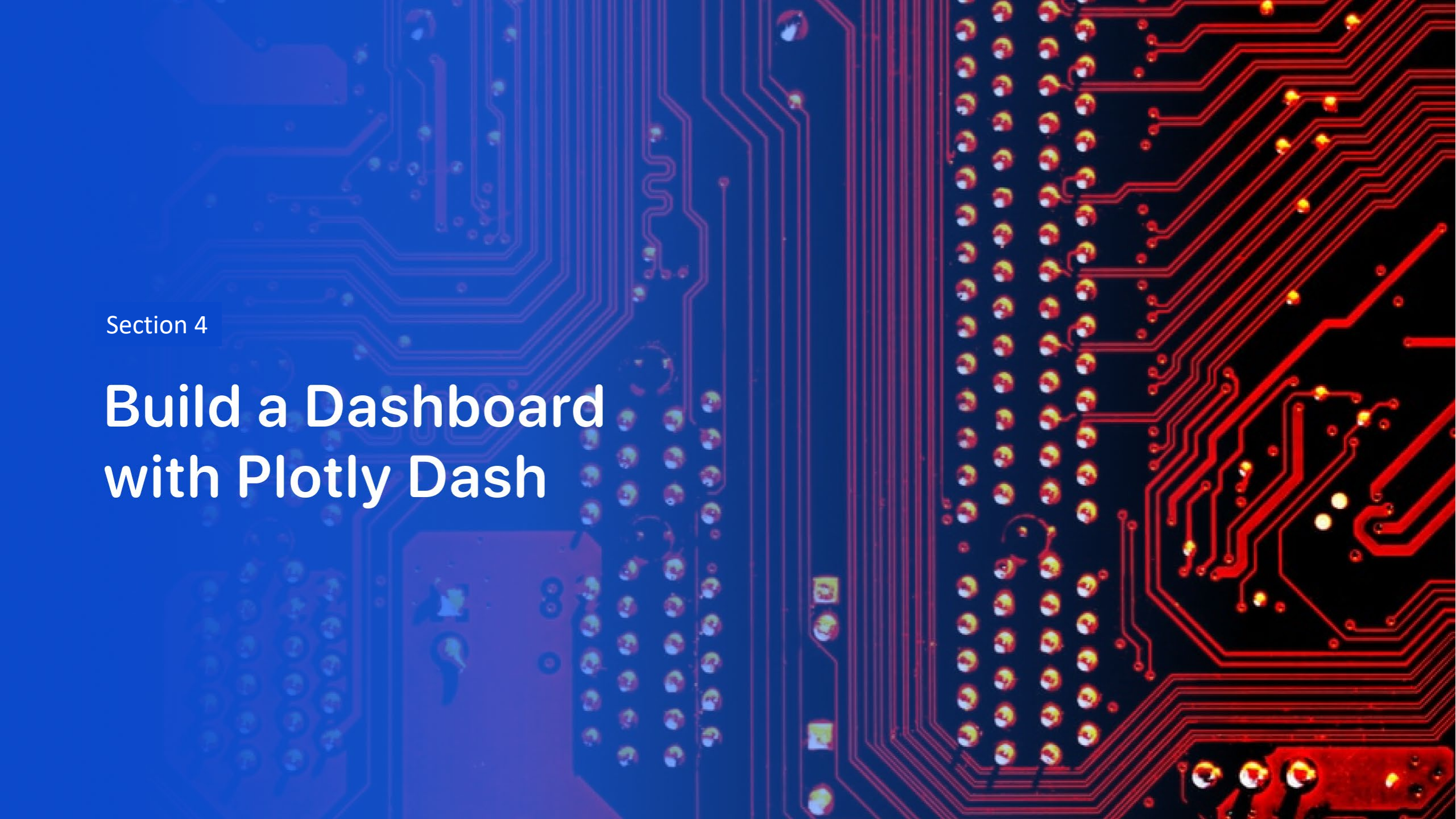

Section 3
Launch Sites
Proximities Analysis

Location of Launch Site

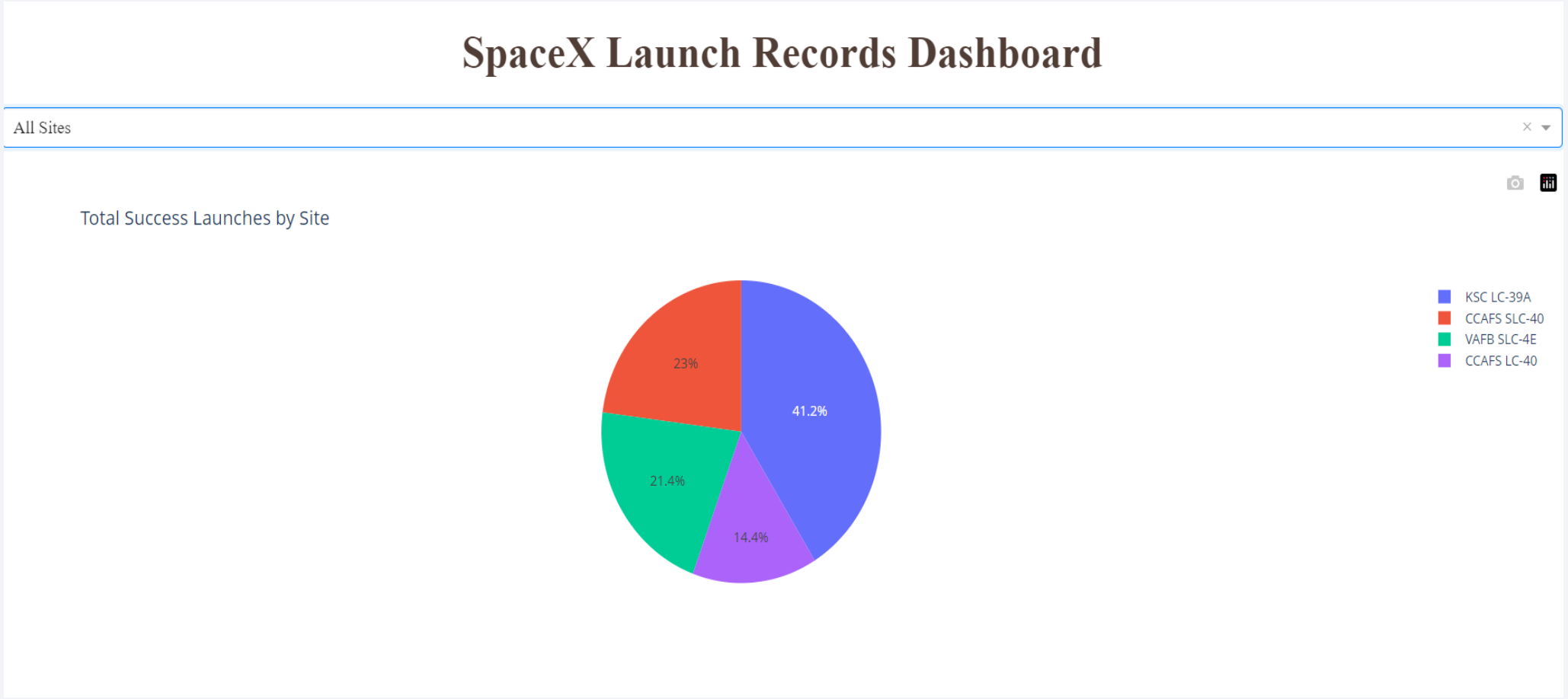
Outcome Map


Salt Lake City
Nebraska
Lincoln
Iowa
Pennsylvania
New York
Ohio
Denver
Utah
Illinois
Indiana
Philadelphia
Sacramento
Nevada
Colorado
Kansas City
Kansas
Missouri
West Virginia
Washington
Kentucky
Virginia
San Jose
California
Virginia Beach
Fresno
Tulsa
Raleigh
Las Vegas
Tennessee
Albuquerque
Oklahoma
Memphis
Charlotte
North Carolina
10
Arkansas
Arizona
New Mexico
Atlanta
South Carolina
Los Angeles
Phoenix
Dallas
Tijuana
Mexicali
Tucson
Mississippi
Alabama
Georgia
Ciudad Juárez
Texas
Louisiana
Austin
Houston
Baton Rouge
Jacksonville
Baja California
Sonora
San Antonio
46
Chihuahua
Florida
Nuevo Laredo
Coahuila
Torreón
Monterrey
Reynosa
Miami
Nassau
Baja California Sur
Saltillo
Tamaulipas
La Paz
Culiacán
Durango
The Bahamas
Durango
México
Ciudad Victoria
Mazatlán
La Habana
Santa Clara
Aguascalientes
Cuba
Camagüey
Turks and Caicos Island
Tepic
León
Jalisco
Mérida
Cancún

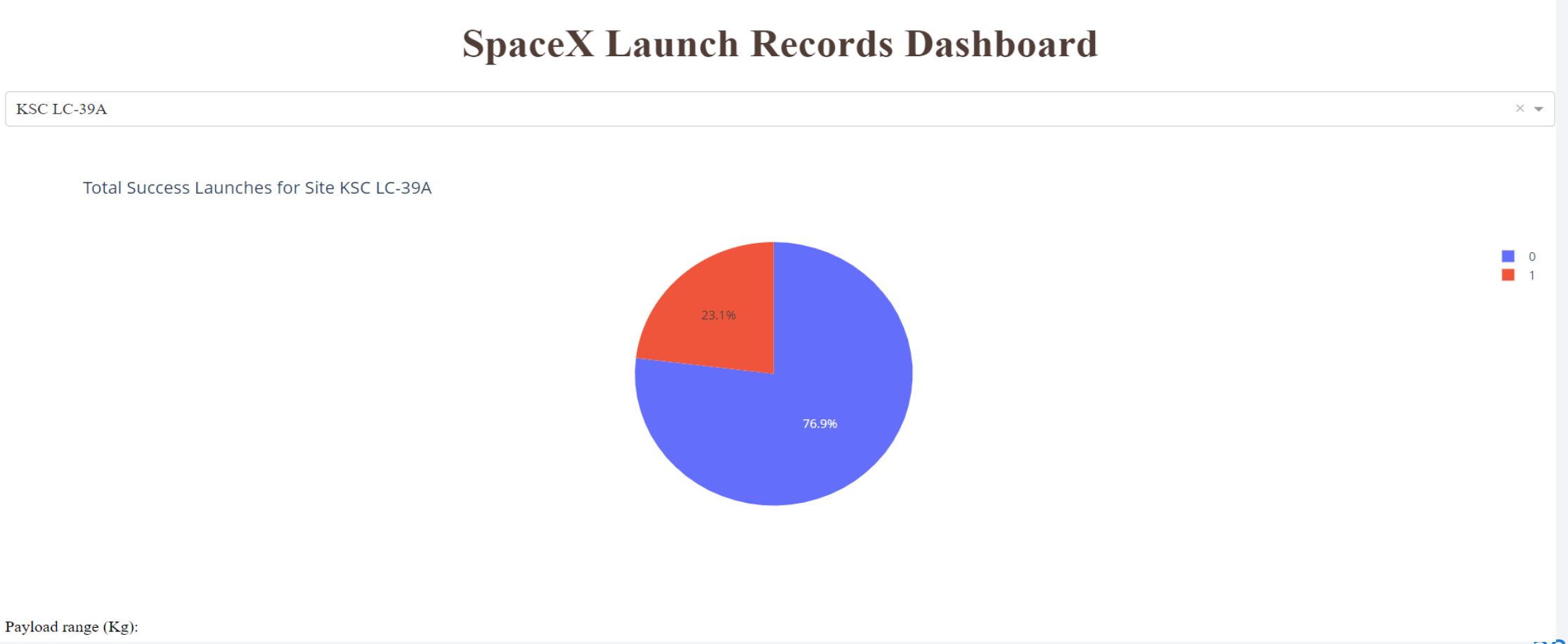
Marked important locations


Section 4
Build a Dashboard with Plotly Dash

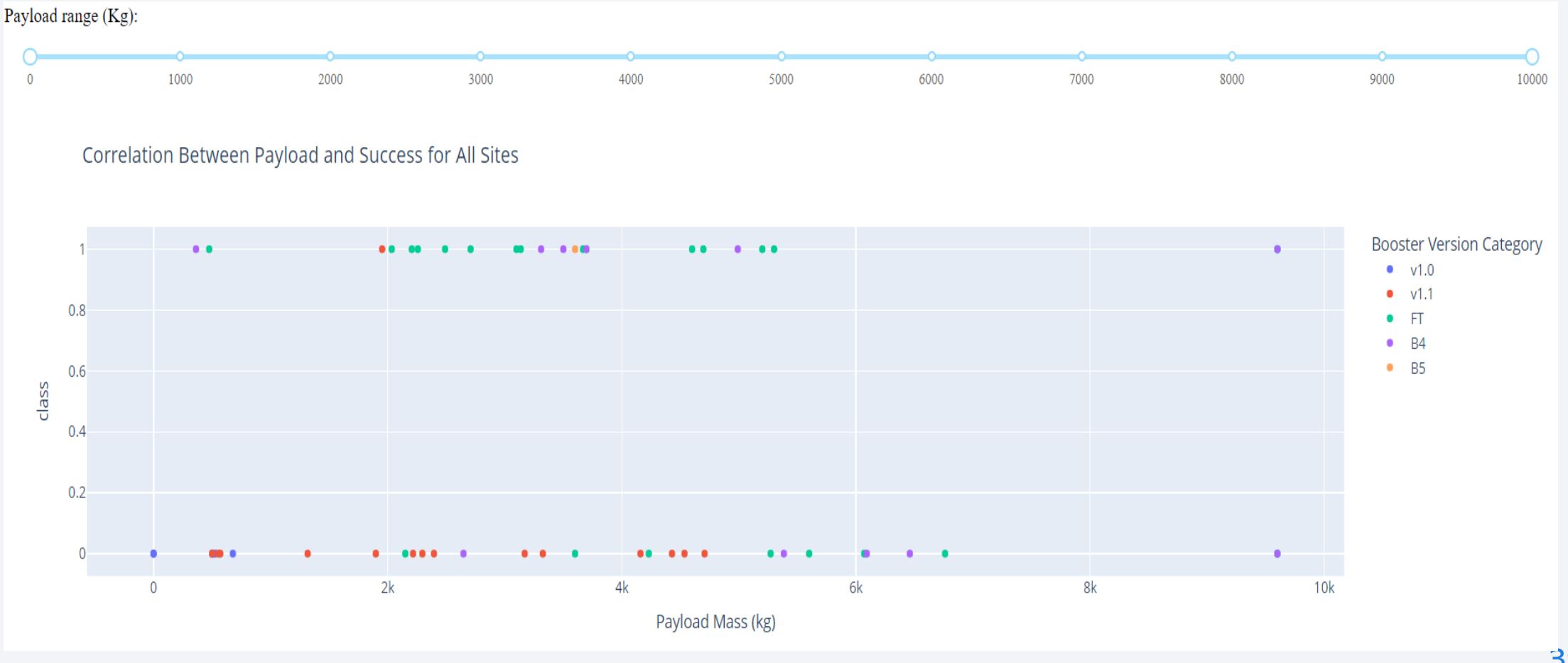
PieChart for All Sites



Launch site with highest Success Ratio



Payload VS Launch Outcomes

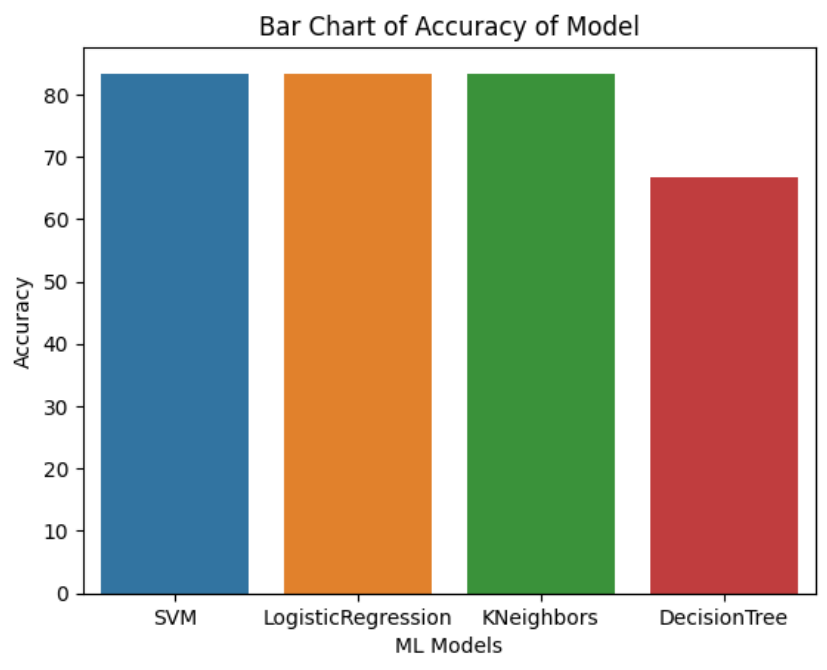


Section 5

Predictive Analysis (Classification)

Classification Accuracy

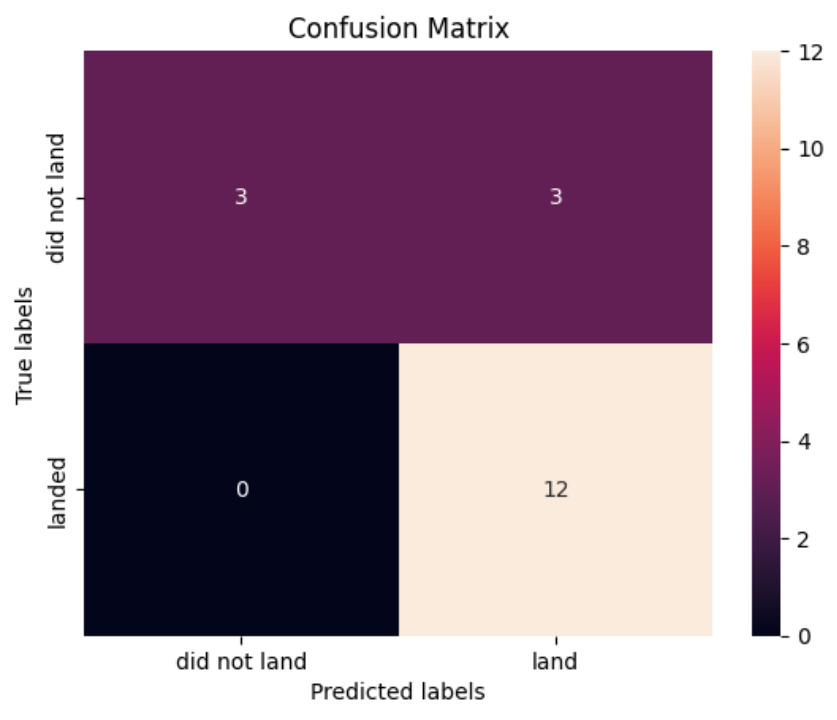
- Visualize the built model accuracy for all built classification models, in a bar chart



- Here we can see All the model has performed similarly but DecisionTree Lacks accuracy more than others.

Confusion Matrix

- Confusion matrix of SVM model



- Here we can see this model is capable of predicting landing quite accurately which 83.33%.

Conclusions

- We can use API and Web Scrapping to collect data
- We can convert our data to our desire using data wrangling
- Performing EDA and Visualizing data helps us understand the relation between various factors
- Machine Learning Models can help us predicting the possibilities of various tasks.

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!