IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Andrew

November 8th 2023

Outline

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

Executive Summary

The goal of this research is to analyze SpaceX Falcon 9 data collected through various sources and employ Machine Learning. Models to predict the success of first stage landing that provides other space agencies the ability to decide if they bid against SpaceX.

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory Data Analysis with Data Visualization
  - Exploratory Data Analysis with SQL
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - Exploratory Data Analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- Project background and context
  - SpaceX is the most successful company of the commercial space age, making space travel affordable. The company 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. Therefore, if we can determine the cost of a launch. Based on public information and machine learning models, we are going to predict if SpaceX will reuse the first stage.
- Problems you want to find answers
  - Determine if the first stage of SpaceX Falcon 9 will land successfully
  - Impact of different parameters/variables on the landing outcomes (e.g., launch site, payload mass, booster version, etc.)
  - Correlations between launch sites and success rates

Section 1

Methodology

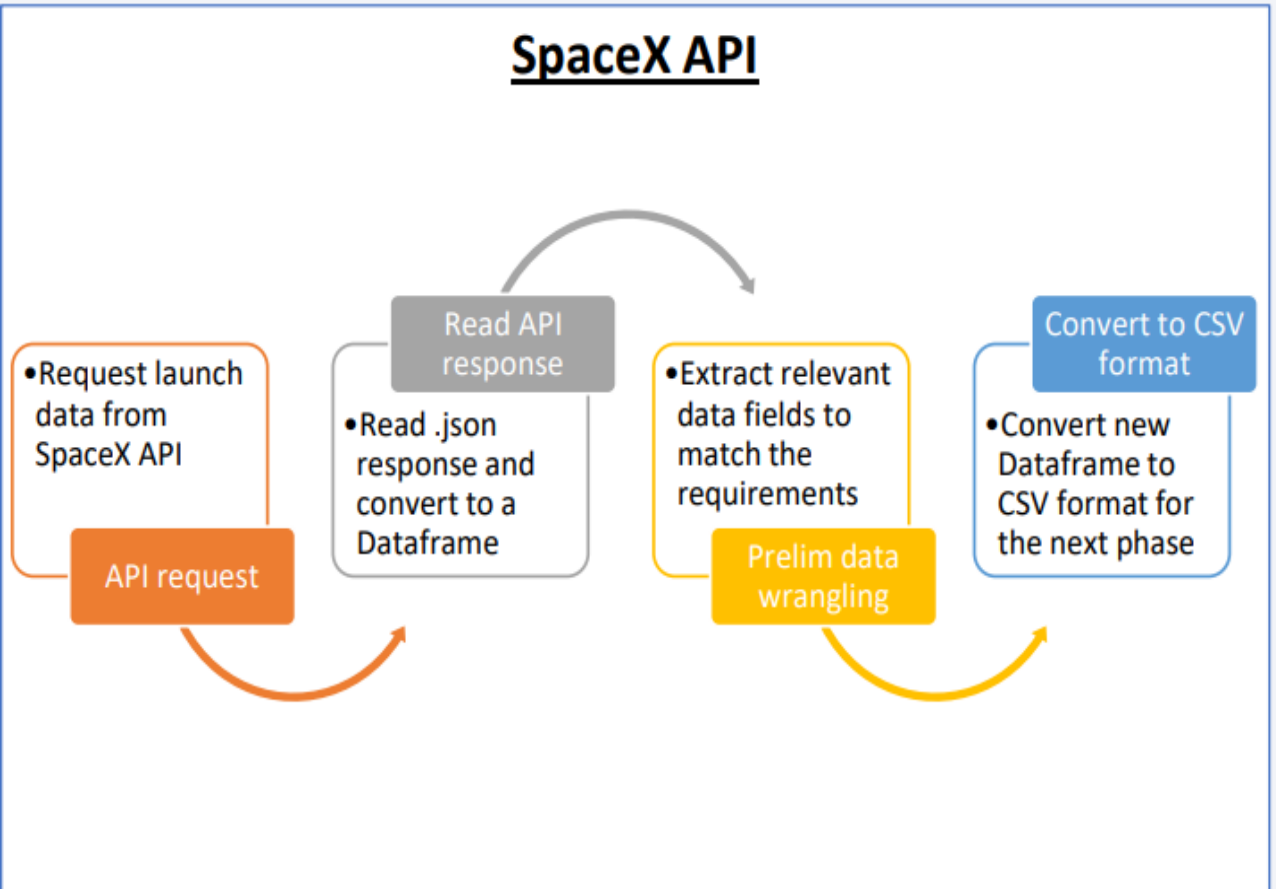
Methodology

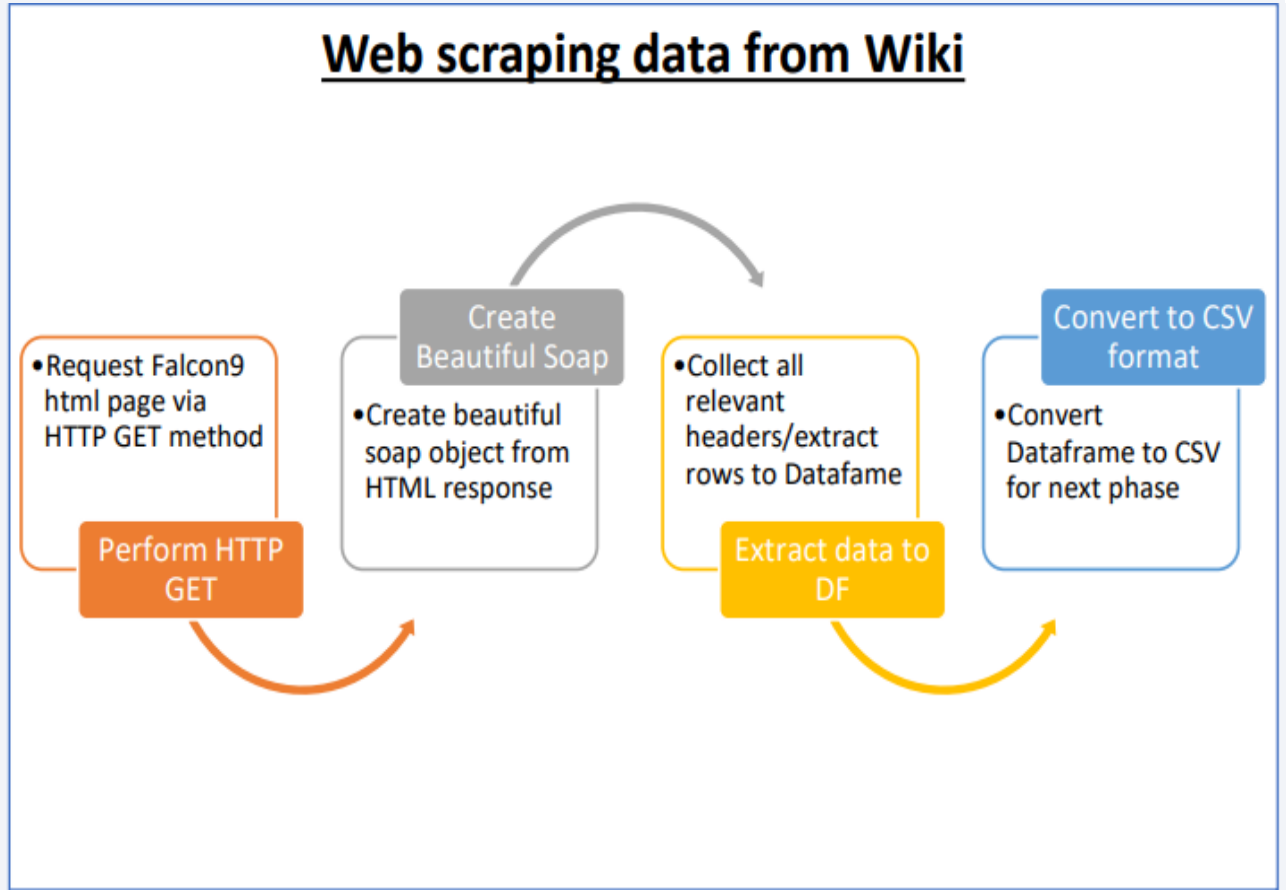
Executive Summary

- Data collection methodology:
  - SpaceX API
  - Using Web Scrapping from Wikipedia(link)
- Perform data wrangling
  - Determined labels for training the supervised models by converting mission outcomes in to training labels(0-unsuccessful, 1-successful)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Created a columns for ‘class’, standardized and transformed data; train/test split data; find best classification algorithm(logistic regression, SVM, decision tree, KNN)

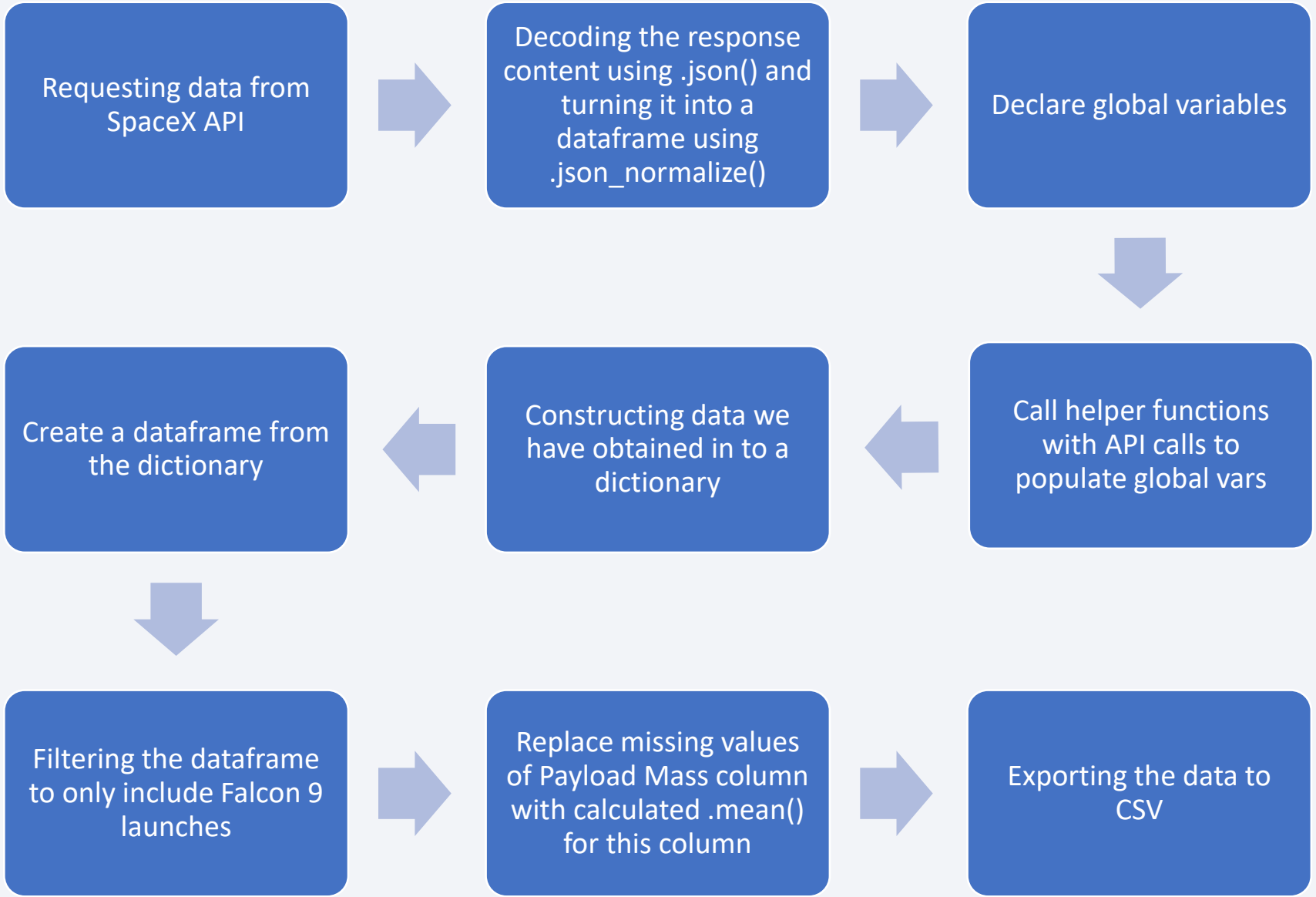
Data Collection

Data collection is the process gathering data from available sources. This data can be structured, unstructured, or semi-structured. For this project, data was collected via SpaceX API and Web scrapping Wiki pages for relevant launch data.



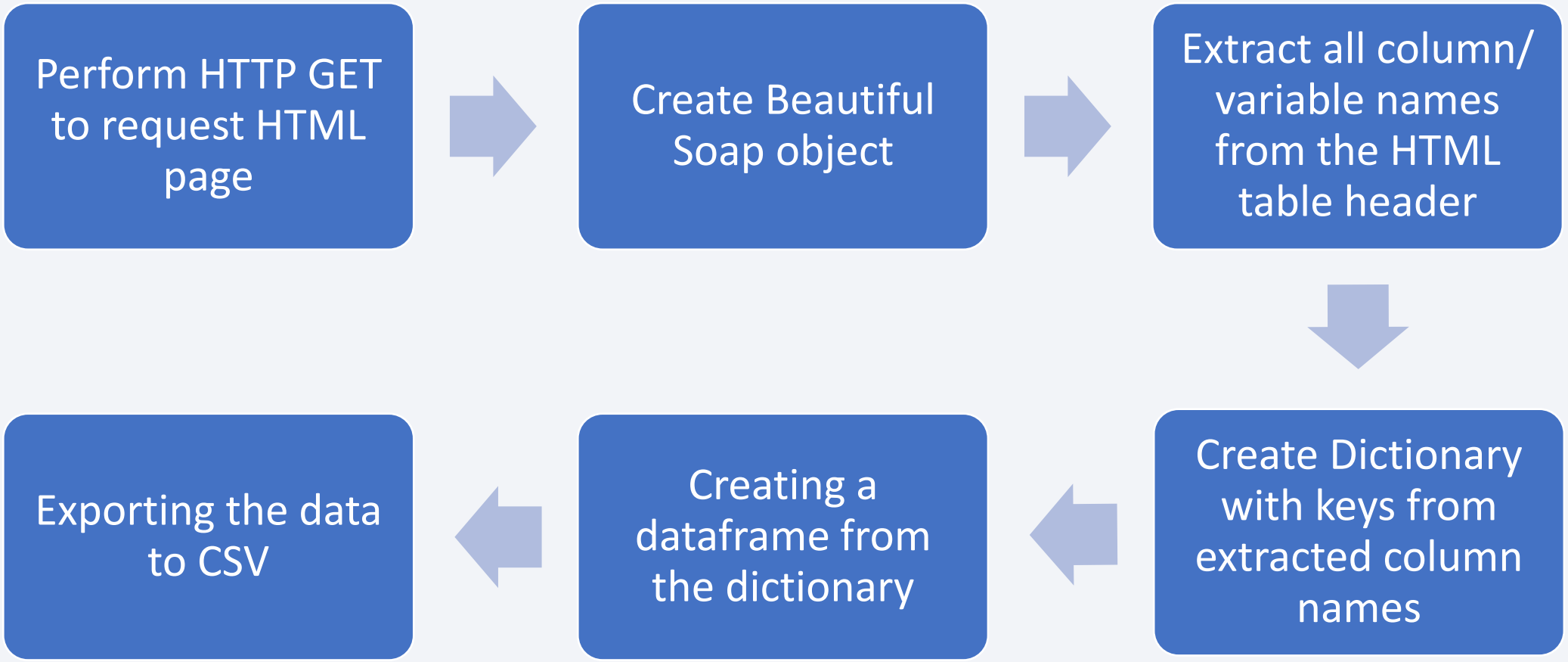


Data Collection – SpaceX API



GitHub 8

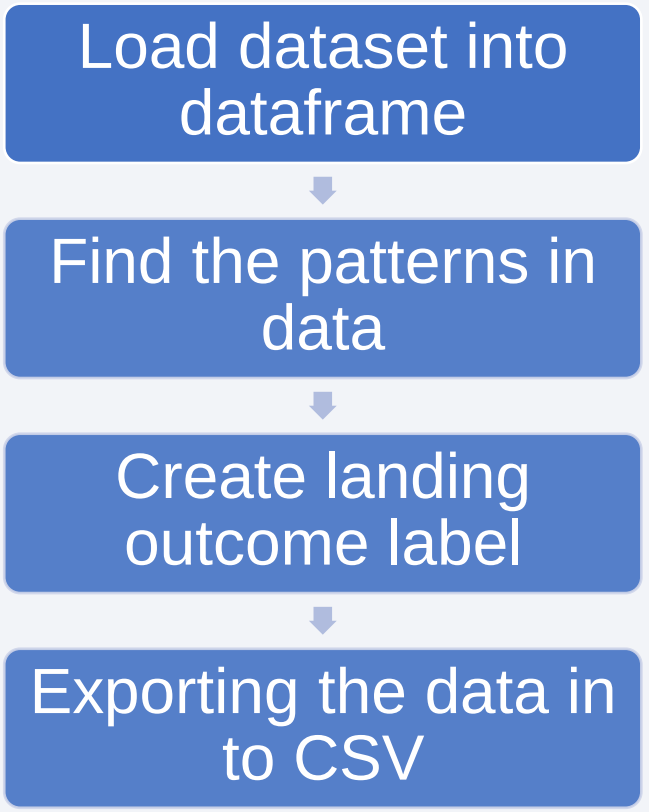
Data Collection - Scraping



GitHub

Data Wrangling

- In the data set, there are a number of different cases where the booster landed successfully or unsuccessfully.
- True Ocean means the mission outcome was successfully landed to a specific region of the ocean
- False Ocean means the mission outcome was unsuccessfully landed to a specific region of the ocean
- RTLS means the mission outcome was successfully landed to a ground pad
- False RTLS means the mission outcome was unsuccessfully landed to a ground pad
- True ASDS means the mission outcome was successfully landed on a drone ship
- False ASDS means the mission outcome was unsuccessfully landed on a drone ship



GitHub

EDA with Data Visualization

- Chart were plotted:
- Flight number vs Payload Mass, Flight Number vs Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend
- Scatter plot show relationship or correlation between two variables making patterns easy to observe
- Bar chart commonly used to compare the values of a variable at a given point in time. Bar charts make it easy to see which groups are highest/common and how other groups compare against each other. Length of each bar is proportional to the value of the items that it represents.
- Line charts show trends in data over time.

EDA with SQL

Performed SQL queries:

- 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 when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship 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 failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking 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

GitHub

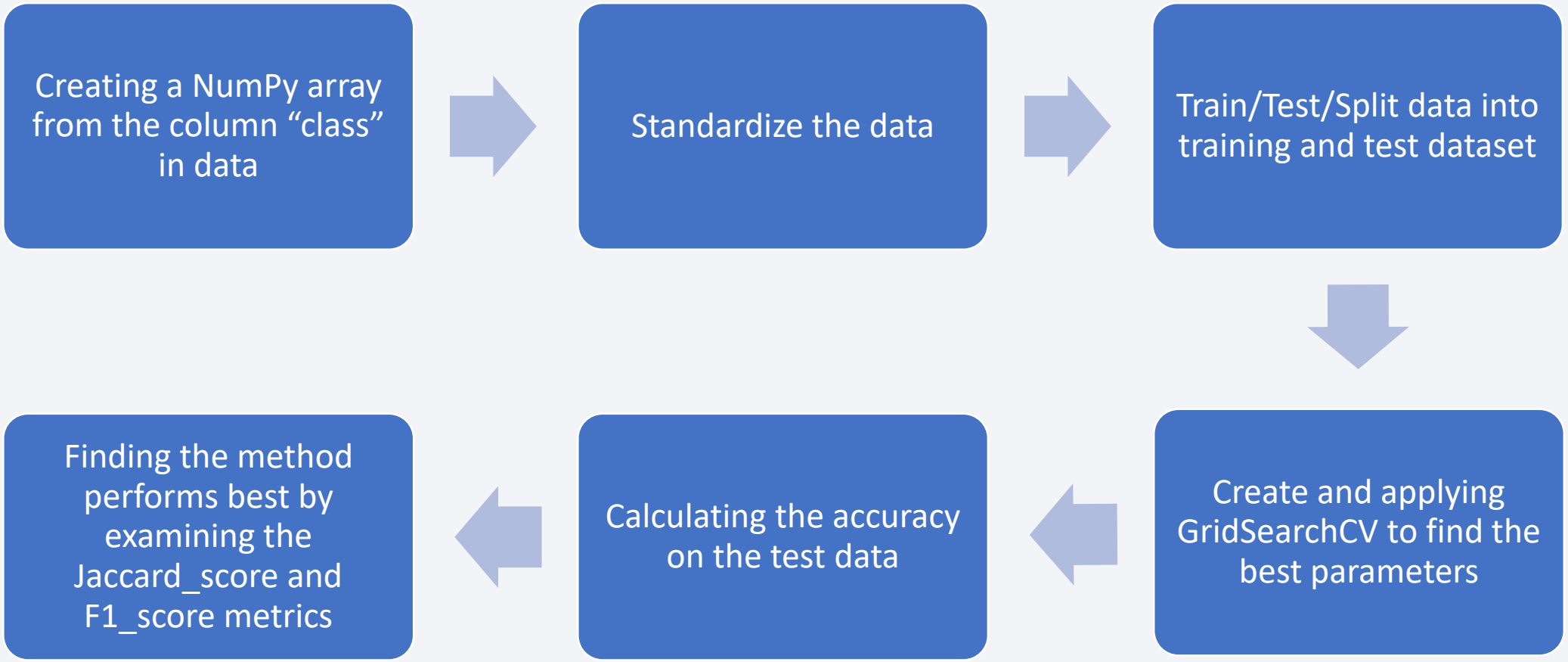
Build an Interactive Map with Folium

- Folium interactive map helps analyze geospatial data to perform more interactive visual analytics and better understand factors such location and proximity of launch sites that impact launch success rate.
- Markers of all Launch Sites:
  - Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location.
  - Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.
- Color Markers of the launch outcomes for each Launch Site:
  - Added color Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.
- Distances between a Launch Site to its proximities: -
  - Added colored Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City.

Build a Dashboard with Plotly Dash

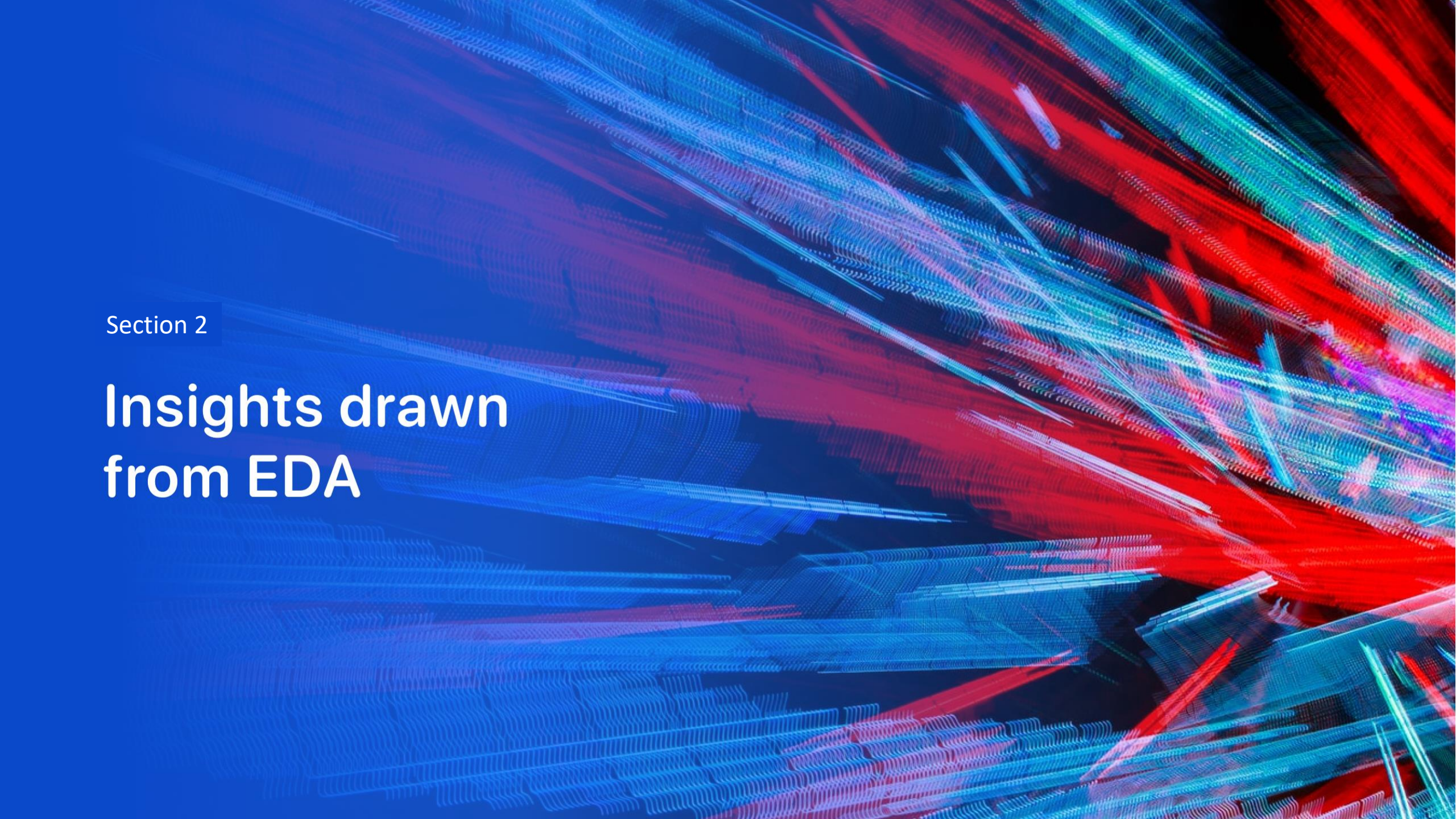
- Launch Sites Dropdown List:
  - Added a dropdown list to enable Launch Site selection
- Pie Chart showing Success Launches (All Sites/Certain Site):
  - Added a pie chart to show the total successful launches count for all sites and the Success vs Failed counts for the site, if a specific Launch Site was selected
- Slider of Payload Mass Range:
  - Added a slider to select Payload range
- Scatter chart of payload mass vs success rate for the different Booster Version:
  - Added a scatter chart to show the correlation between Payload and Launch Success.

Predictive Analysis (Classification)

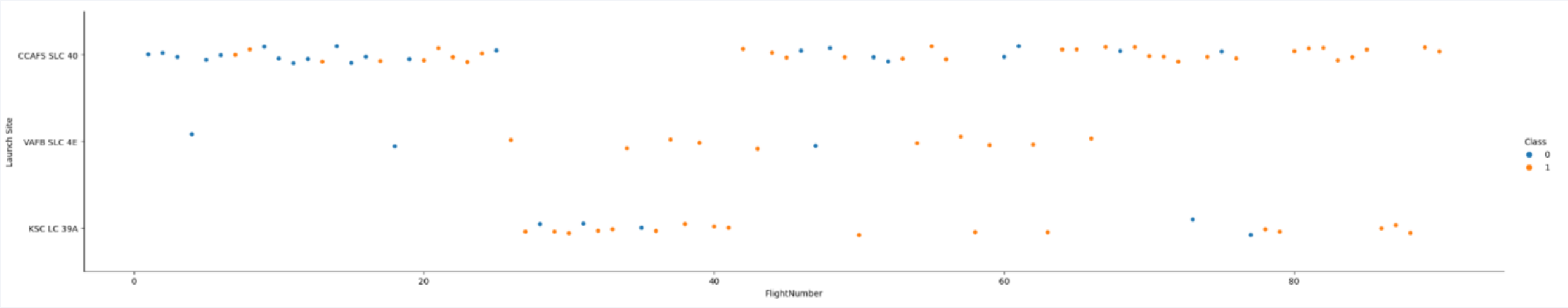


Results

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

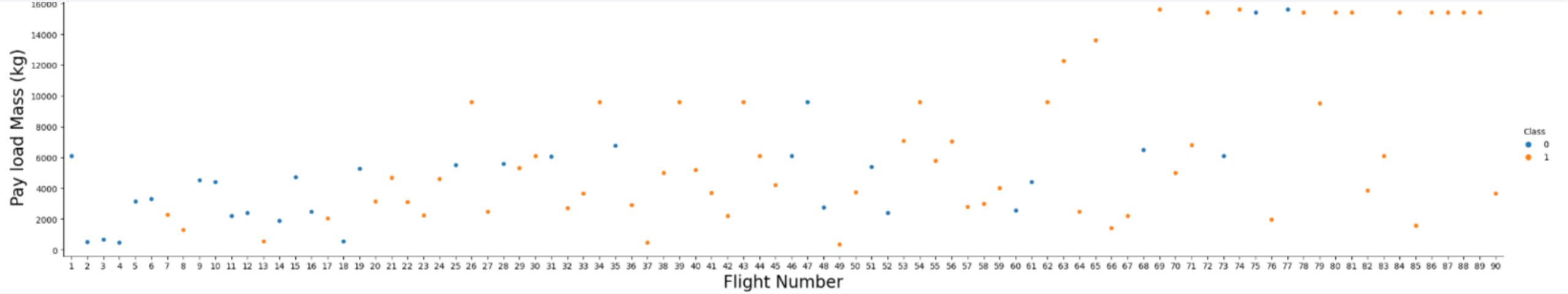

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



- Explanation:
  - The earliest flights failed while the latest flights succeeded.
  - The CCAFS SLC 40 launch site has about a half of all launches
  - VAFB SLC 4E and KSC LC 39A have higher success rates.

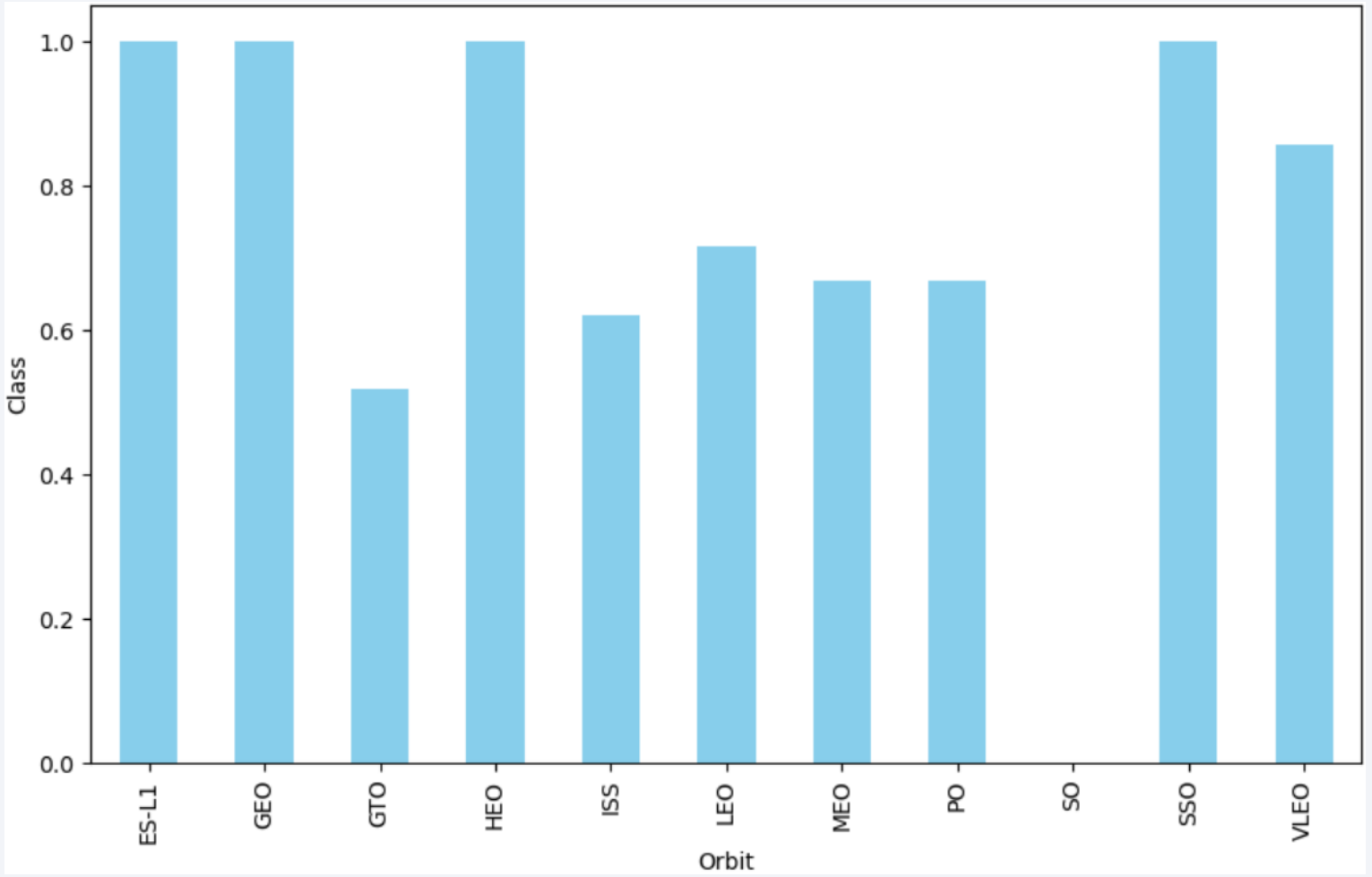
Payload vs. Launch Site



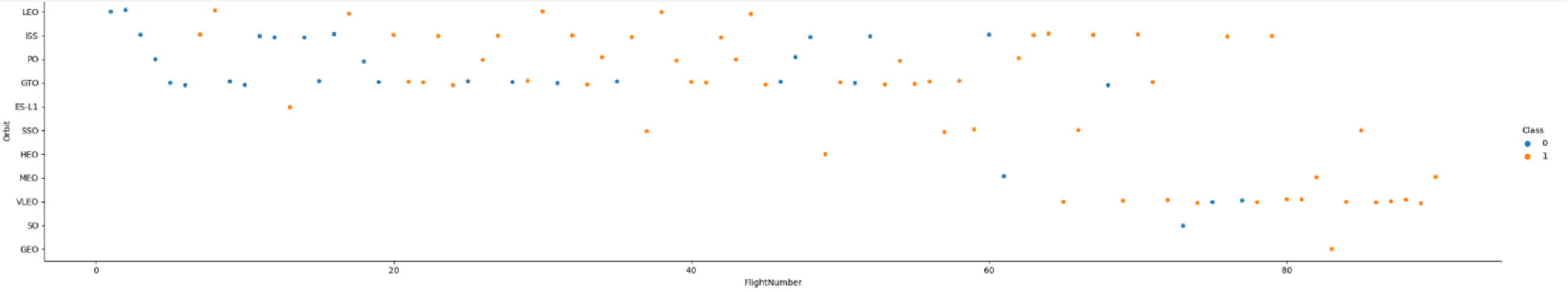
- Explanation:
  - For every launch site the higher the payload mass, the higher the success rate.
  - Most of the launches with payload mass over 7000kg were successful
  - For launch site VAFBSLC 4E no rockets launched for payload greater than 10000kg

Success Rate vs. Orbit Type

- Explanation:
  - The Orbits ES-LI, GEO, have the highest successful rates with 100%
  - SO orbit has the lowest success rate with 0%
  - Orbits with success rate between 50% and 85%: GTO, ISS, LEO, MEO, PO

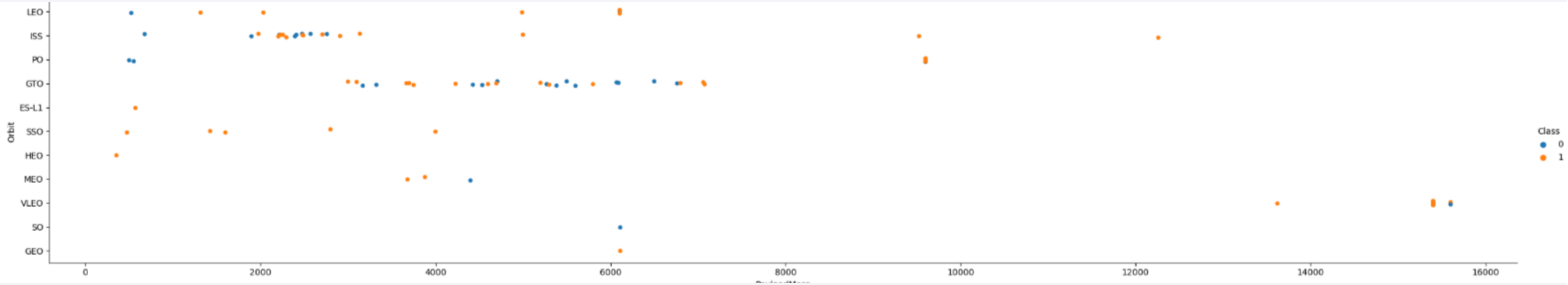


Flight Number vs. Orbit Type



- Explanation:
  - For most orbits (LEO, ISS, PO, SSO, MEO) successful landing rates appear to increase with light numbers.
  - There is no relationship between flight number and orbit for GTO

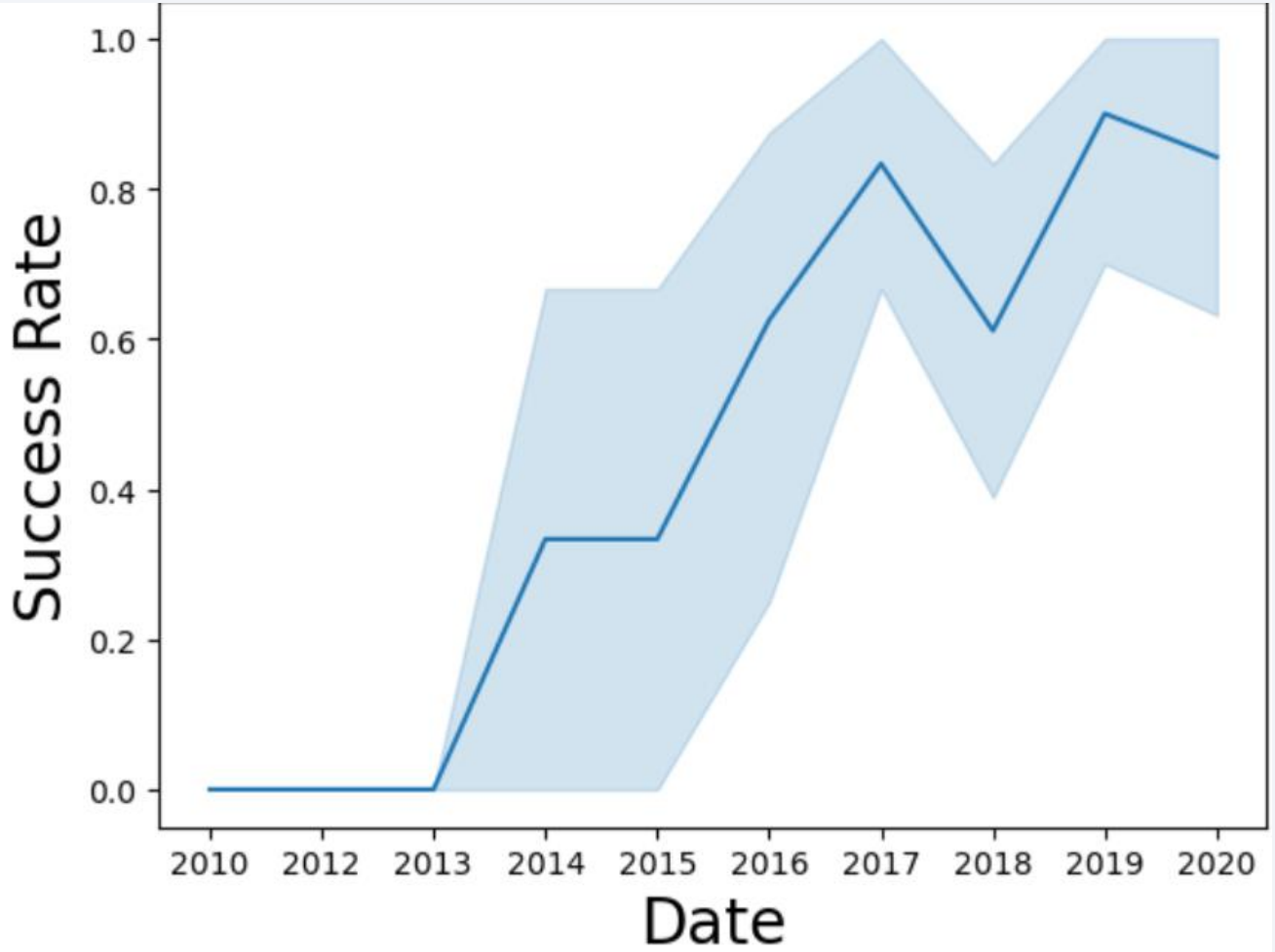
Payload vs. Orbit Type



- Explanation:
  - Successful landing rates appear to increase with payload for orbits LEO, ISS, PO, SSO
  - With GEO orbit is not clear pattern between payload and orbit for successful and or unsuccessful landing

Launch Success Yearly Trend

- Explanation: the success rate since 2013 kept increasing till 2020 but there was a slight decrease in 2018.



All Launch Site Names

In [18]:
```
%sql select distinct(Launch_Site) from SPACEXTABLE
```

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

Out[18]:

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

- Explanation: Displaying the names of unique launch sites in the space mission.

Launch Site Names Begin with 'CCA'

In [21]:
```
%sql select Launch_Site from SPACEXTABLE where Launch_Site like 'CCA%' limit 5
```

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

Out[21]:

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

- Explanation:
  - Displaying 5 launch site names begin with string “CCA”

Total Payload Mass

```
In [26]: %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTABLE WHERE Customer = 'NASA (CRS)'
 * sqlite:///my_data1.db
Done.
Out[26]: sum(PAYLOAD_MASS__KG_)
         45596
```

- Explanation:
  - Displaying the total pay load mass carried by boosters launched by NASA(CRS).

Average Payload Mass by F9 v1.1

```
In [28]: %sql select AVG(PAYLOAD_MASS__KG_) from SPACEXTABLE where Booster_Version = 'F9 v1.1'
 * sqlite:///my_data1.db
Done.
```

Out[28]:

AVG(PAYLOAD_MASS__KG_)
2928.4

- Explanation:
  - Displaying the average pay load mass carried by boosters version F9 v1.1.

First Successful Ground Landing Date

In [32]:
```
%sql select min(Date) from SPACEXTABLE where Landing_Outcome like '%Success%'
```

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

Out[32]:

min(Date)
2015-12-22

- Explanation:
  - Listing the date when the first successful landing outcome in ground pad was achieved.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
In [34]: %sql select Booster_Version from SPACEXTABLE where Landing_Outcome='Success (drone ship)' and PAYLOAD_MAS
 * sqlite:///my_data1.db
Done.
```

Out[34]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- Explanation:
  - Listing the names of the boosters which have success in drone ship and have pay load mass greater than 4000 but less than 6000 .

Total Number of Successful and Failure Mission Outcomes

In [41]:
```
%sql select Mission_Outcome, count(*) from SPACEXTABLE group by Mission_Outcome;
```

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

Out[41]:

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

- Explanation:
  - Listing the total number of successful and failure mission outcomes.

Boosters Carried Maximum Payload

- Explanation:
  - Listing the names of the boosters versions which have carried the maximum payload mass.

```
In [44]: %sql select distinct Booster_Version, (select max(PAYLOAD_MASS__KG_) from SPACEXTABLE) as max from SPACEX
 * sqlite:///my_data1.db
Done.
```

Out[44]:

Booster_Version	max
F9 v1.0 B0003	15600
F9 v1.0 B0004	15600
F9 v1.0 B0005	15600
F9 v1.0 B0006	15600
F9 v1.0 B0007	15600
F9 v1.1 B1003	15600
F9 v1.1	15600
F9 v1.1 B1011	15600
F9 v1.1 B1010	15600
F9 v1.1 B1012	15600
F9 v1.1 B1013	15600
F9 v1.1 B1014	15600
F9 v1.1 B1015	15600

2015 Launch Records

In [49]:
```
%sql select substr(Date,6,2) as month,Landing_Outcome, Booster_Version, Launch_Site from SPACEXTABLE wher
```

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

Out[49]:

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

- Explanation:
  - Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in the year 2015.

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

- Explanation:
  - Ranking the counts of landing outcomes between the date 2010-06-04 and 2017-03-20.

```
In [55]: %sql select Landing_Outcome, count(*) as count from SPACEXTABLE where Date >= '2010-06-04' and Date <= '2
 * sqlite:///my_data1.db
Done.
```

Out[55]:

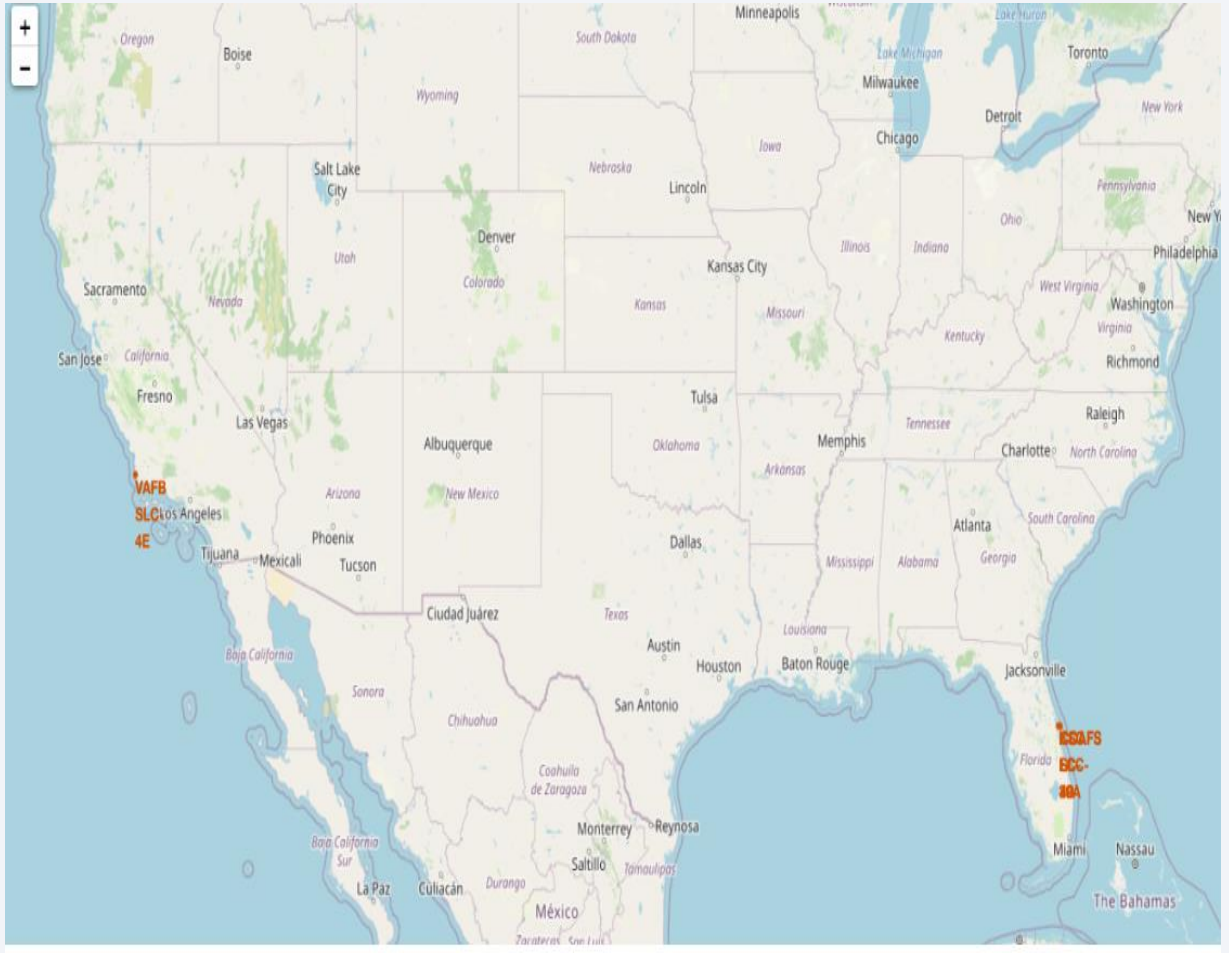
Landing_Outcome	count
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

Section 3

Launch Sites Proximities Analysis

All Launch Sites Map

- Explanation:
  - We can see Falcon 9 launch sites located in US. We have four main launch sites:
    - VAFB SLC-E4 (California)
    - CCAFS LC-40 (Florida)
    - KSC LC-39A (Florida)
    - CCAFS SLC-40(Florida)

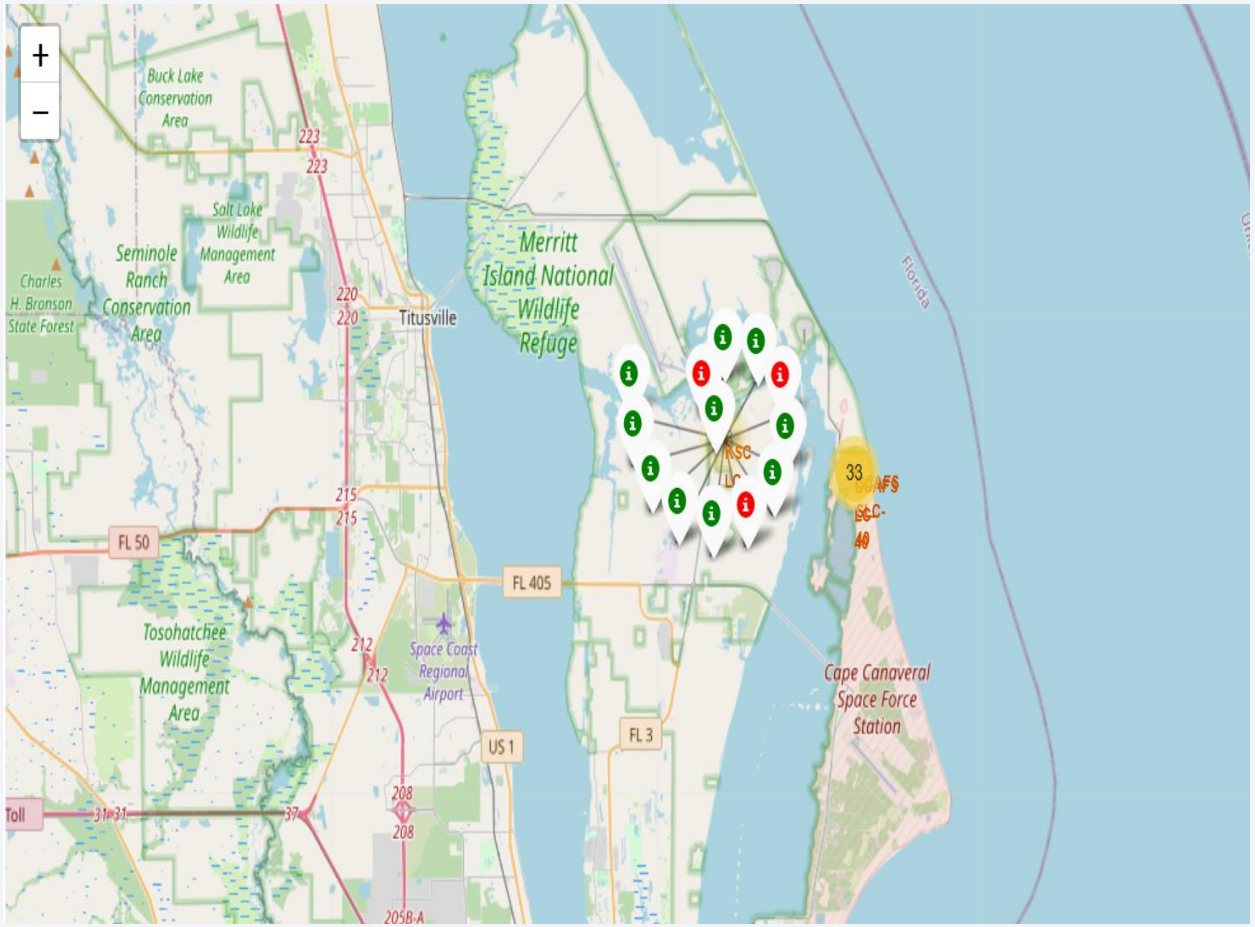


Failed/Success Launch Map

- Explanation:
  - From the color-labeled markers we should be able to easily identify which launch sites have relatively high success rates.

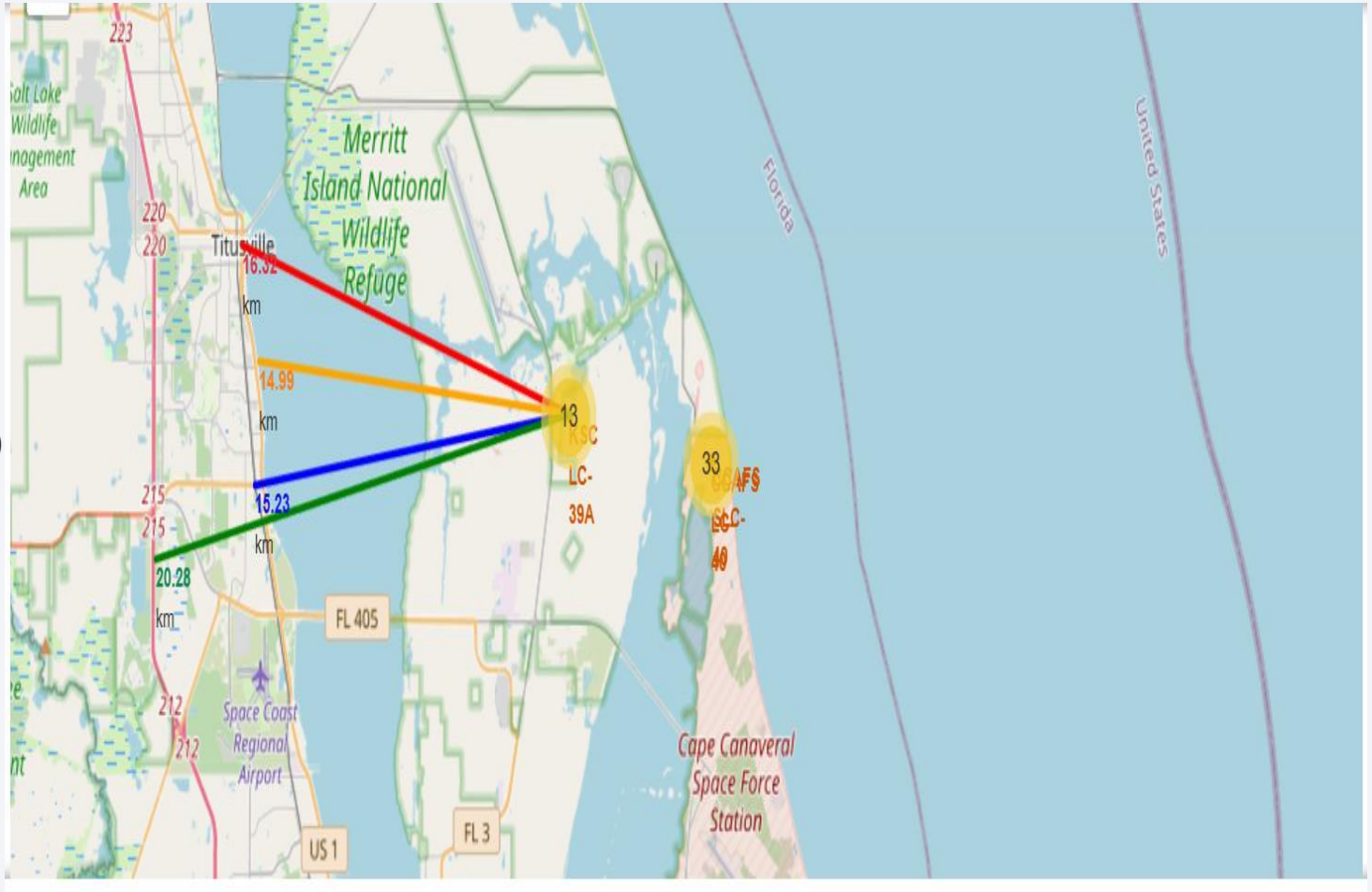
  - Green Marker = Successful Launch

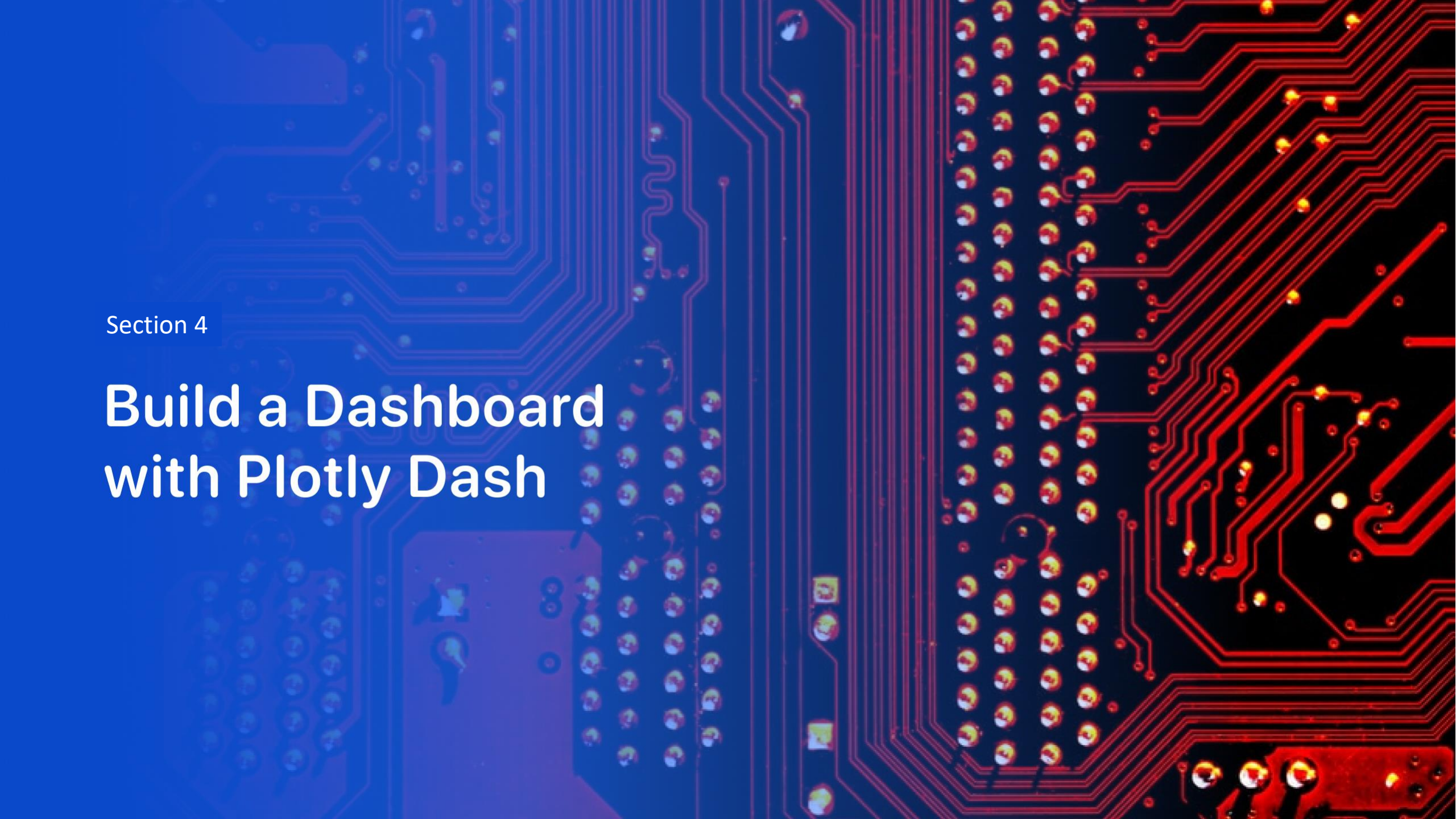
  - Red Marker = Failed Launch



Distance from the launch site KSC LC-39A to its proximities

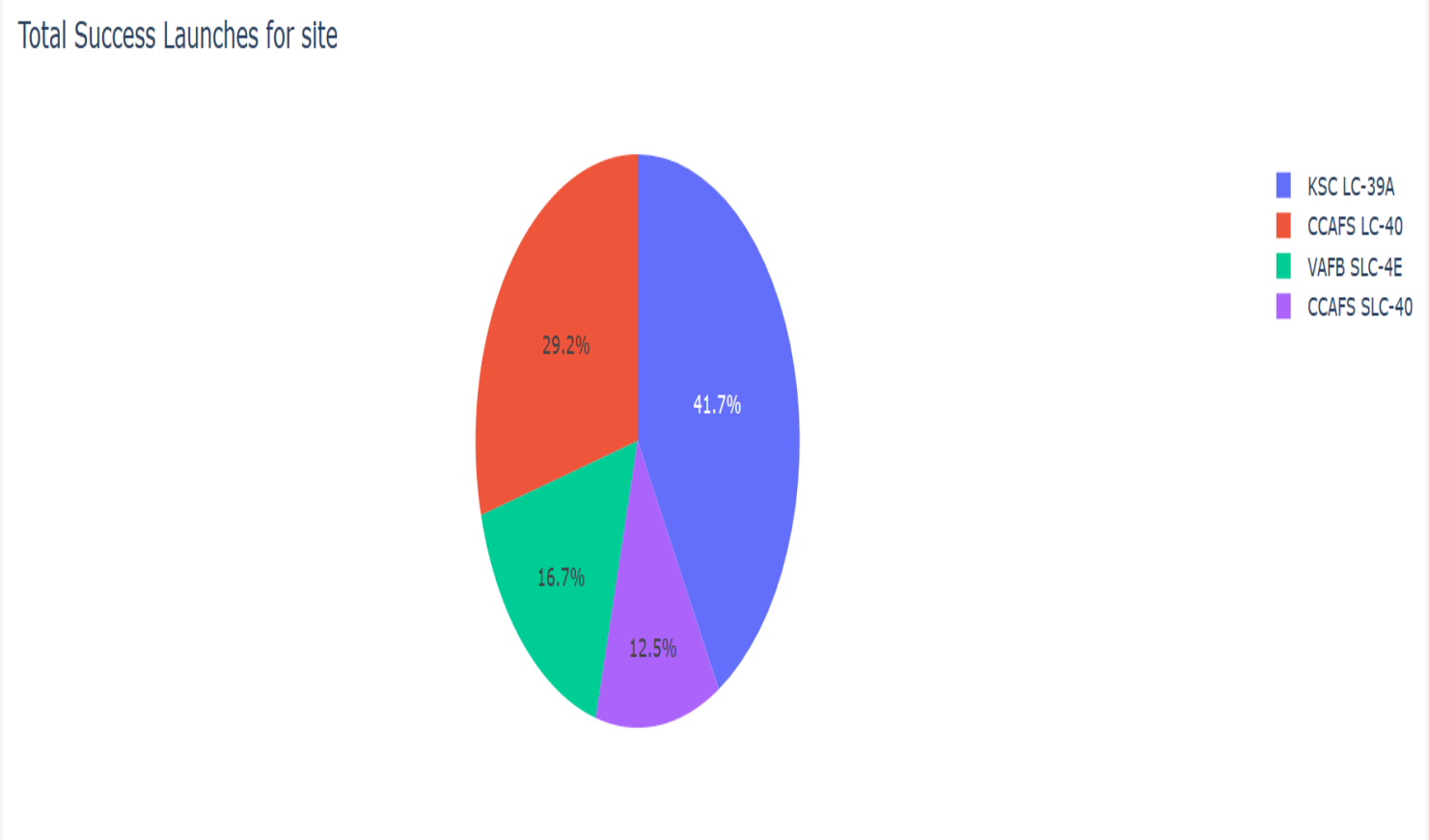
- Explanation:
  - From the analysis of launch site KSC LC 39A we can see:
    - Relative close to railway (15.23km)
    - Relative close to highway (20.28km)
    - Relative close to coastline (14.99km)
    - Relative close to closest city Titusville (16.23km)




Section 4
Build a Dashboard with Plotly Dash

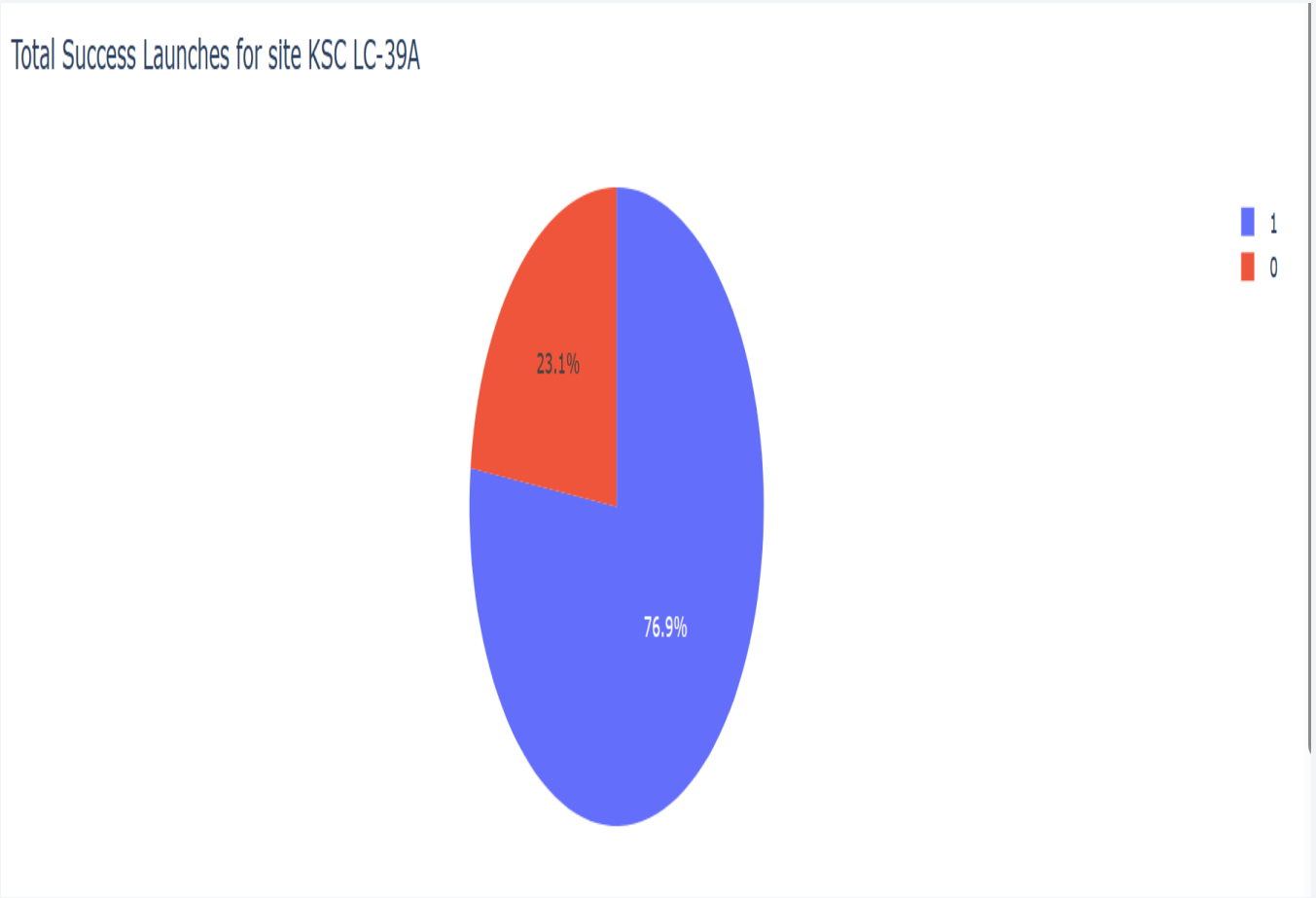
Launch success for all sites

- Explanation:
  - You can see clearly success rate of all launch sites and KSC LC-39A has the most highest successful launches, VAFB SLC-4E is the lowest successful launches.

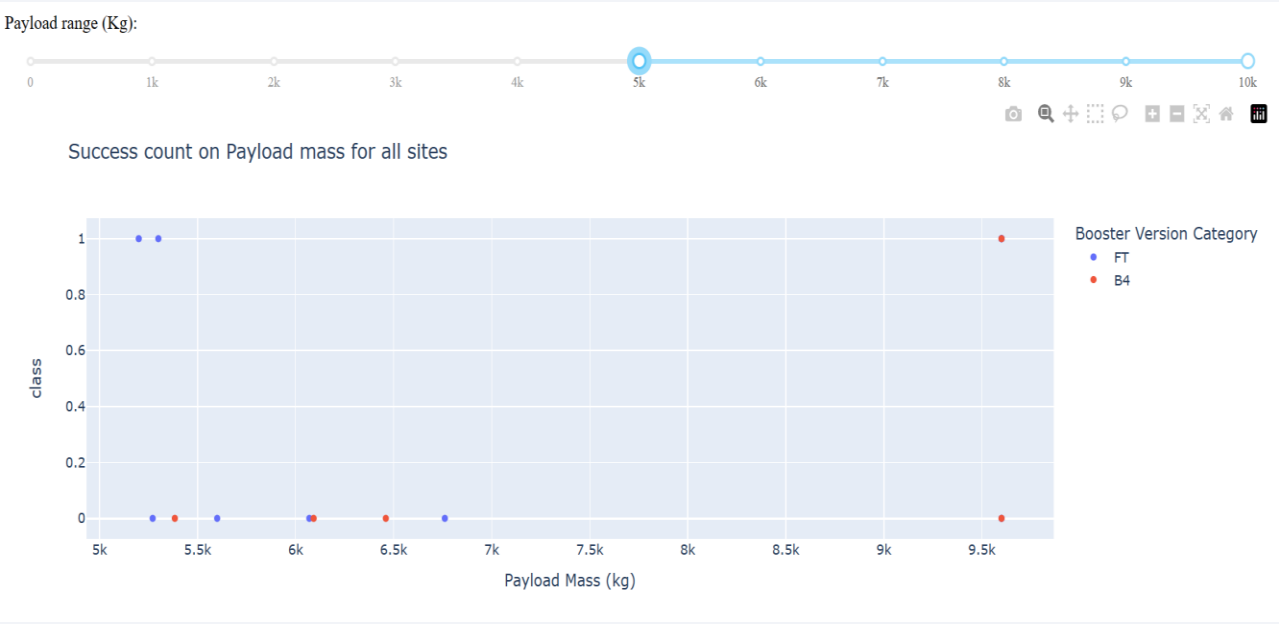


Launch site with highest launch success ratio

- Explanation:
  - KSC LC-39A has the highest successful rate (76.9%) with 10 successful and only 3 failed.



Payload Mass vs Launch Outcome for all sites

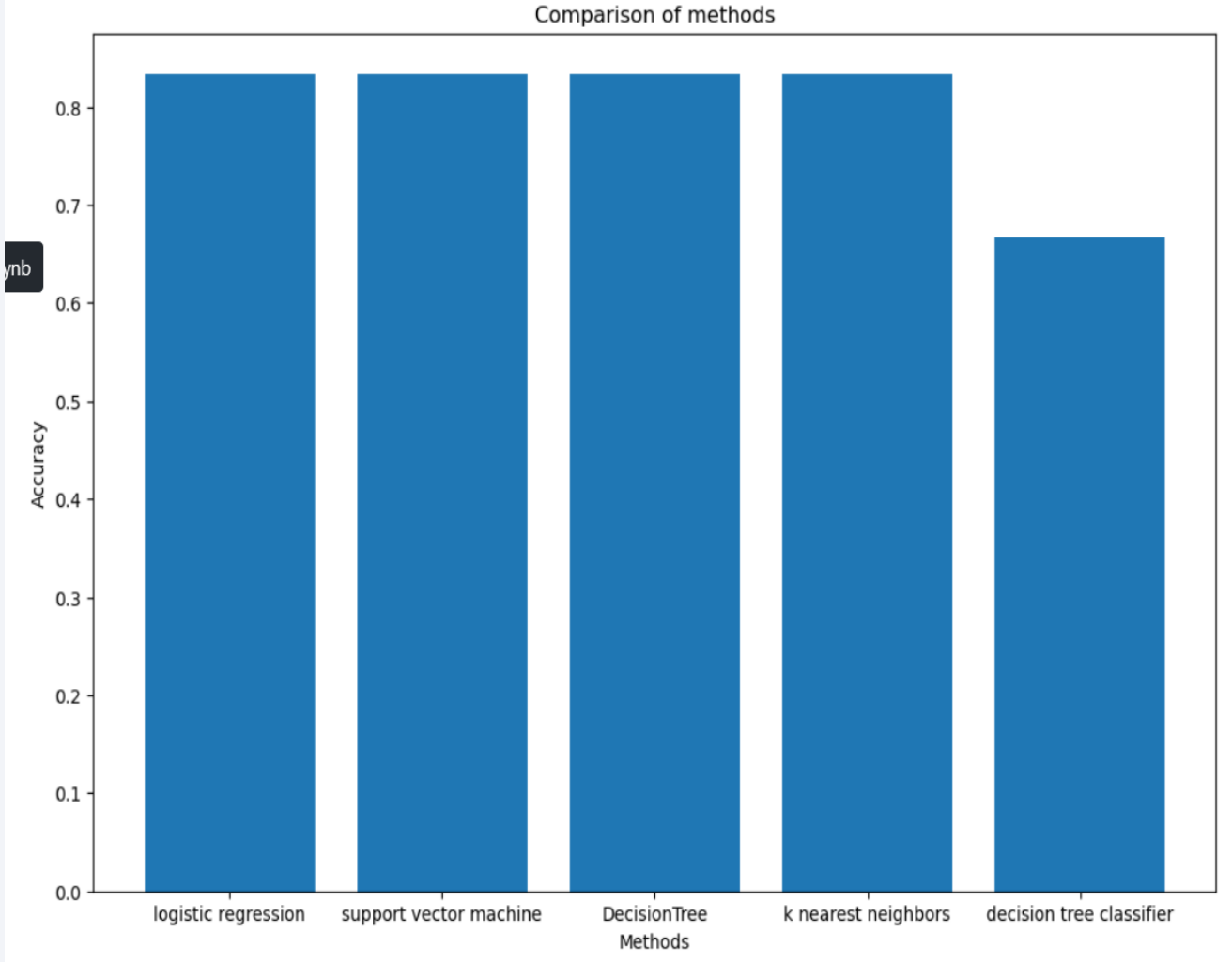


- Explanation:
  - The charts show that between 2000 and 5500kg have highest successful rate.

Section 5

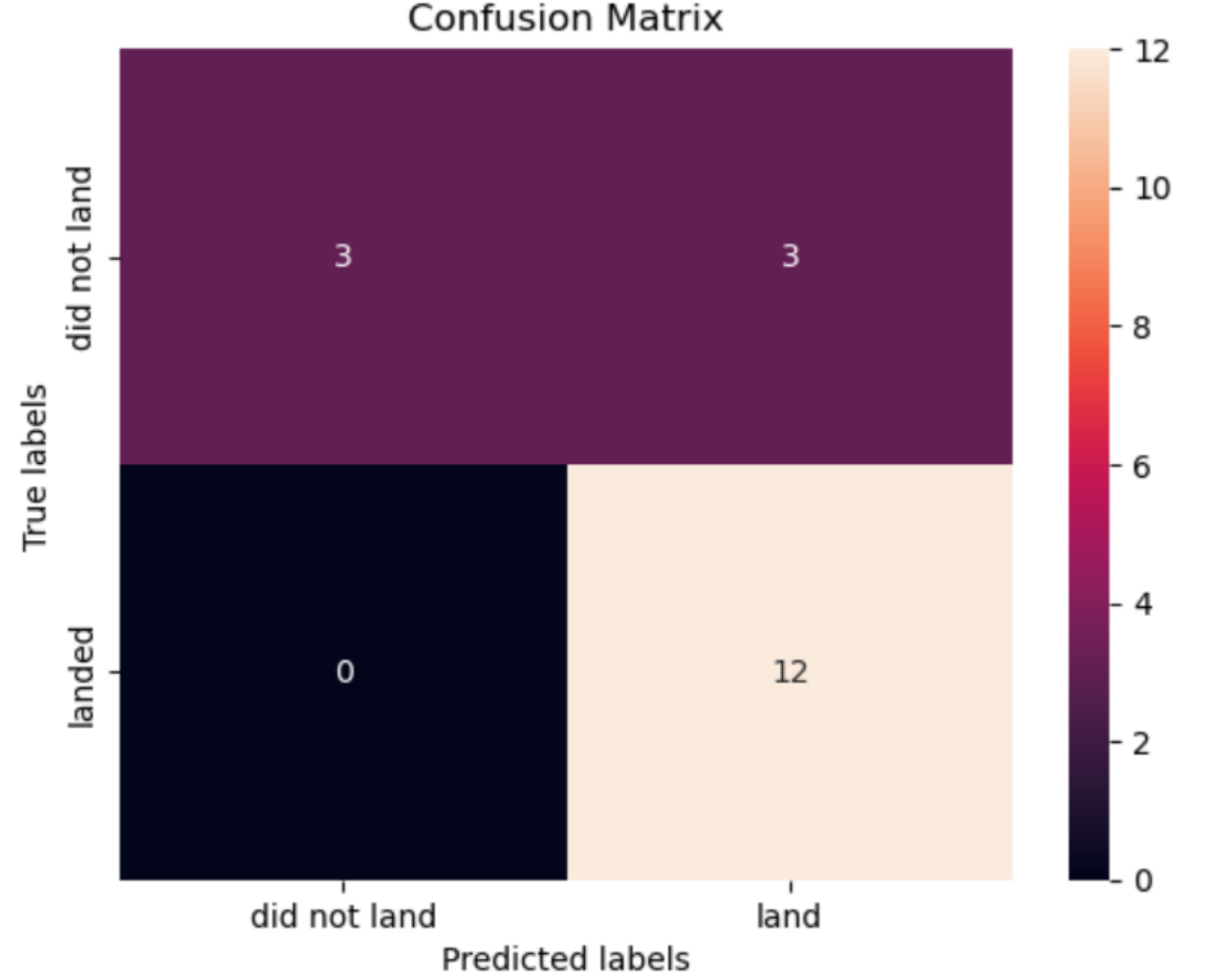
Predictive Analysis (Classification)

Classification Accuracy



- Explanation:
  - Based on bar chart, we can see Logistic Regression, SVM, Decision Tree, kNN have the highest classification score with 0.875.

Confusion Matrix



- Explanation:
  - 12 scenarios were predicted Yes for landing, and they did land successfully
  - 3 scenarios were predicted No for landing, and they did not land
  - 3 scenarios were predicted No for landing, and they did not land successfully
  - Overall, the classifier is correct about 83% in TP and TN.

Conclusions

- As the numbers of flights increase, the first stage is more likely to land successful.
- Launch success rate increased by about 80% from 2013 to 2020.
- Orbits ES-L1, GEO, HEO, and SSO have the highest launch success rate and orbit GTO is lowest rate.
- KSC LC-39A has the highest success rate of the launches for all the sites and CCAFS SLC-40 has the lowest launch success rate.
- Launch site are located nearly from the city, coastline, railroads, and highways.
- The best performance of ML Classification are Logistic Regression, SVM, Decision Tree, kNN with an accuracy 87.5%.

Appendix

- IBM
- Dash
- SpaceX

Thank you!