IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

MANASA KASULA

29-05-2024

Outline

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

Executive Summary

❖ Summary Of Methodologies

- ✓ Data Collection Using API
- ✓ Scraping Relevant data from Websites
- ✓ Data Wrangling
- ✓ Exploratory Data Analysis with Data Visualization
- ✓ EDA with SQL
- ✓ Building an interactive map with Folium
- ✓ Building a Dashboard with Plotly Dash
- ✓ Predictive analysis (Classification)

❖ Summary Of All Results

- ✓ Exploratory Data Analysis Results
- ✓ Prediction analysis Results
- ✓ Visual Representation of the data indicating the success rate.

Introduction

❑ Project background and context

SpaceX is the first Company to develop a liquid propellant rocket that has reached the orbit.

SpaceX Accomplishments include: Sending spacecraft to the International Space Station. Starlink , a satellite internet constellation providing satellite Internet access.

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.

Determining the price of each launch of Space Y by gathering the information about SpaceX and creating Dashboards. Finally, Building a Machine Learning Model using the public Information to predict the successful landing of the rocket

.

❑ Problems you want to find answers

Can we reduce the reduce the cost of launching a Rocket?

Is there any possibility to split/move the massive work done at first step to later stages?

Which Classification algorithm is best suitable for prediction Analysis?

Section 1

Methodology

Methodology

Data collection methodology:

- Data Collection using an SpaceX REST API
- Using Web Scraping to get the data from Website.

➢ Perform data wrangling

- Using REST API to get into the actual information.
- Dealing with the Missing Values.

➢ Perform exploratory data analysis (EDA) using visualization and SQL

- Analyzing the data whether the data can be directly used to build a machine Learning Model for successful landing.

Methodology (cont..)

- Perform interactive visual analytics using Folium and Plotly Dash
  - Building an interactive maps.
  - Creating Dashboards from the resultant data for Visualization.
- Perform predictive analysis using classification models
  - Using the data we create a Machine Learning Model to predict the success rate of Landing.
  - Using Various Classification Algorithms to test the model with best Accuracy.

Data Collection

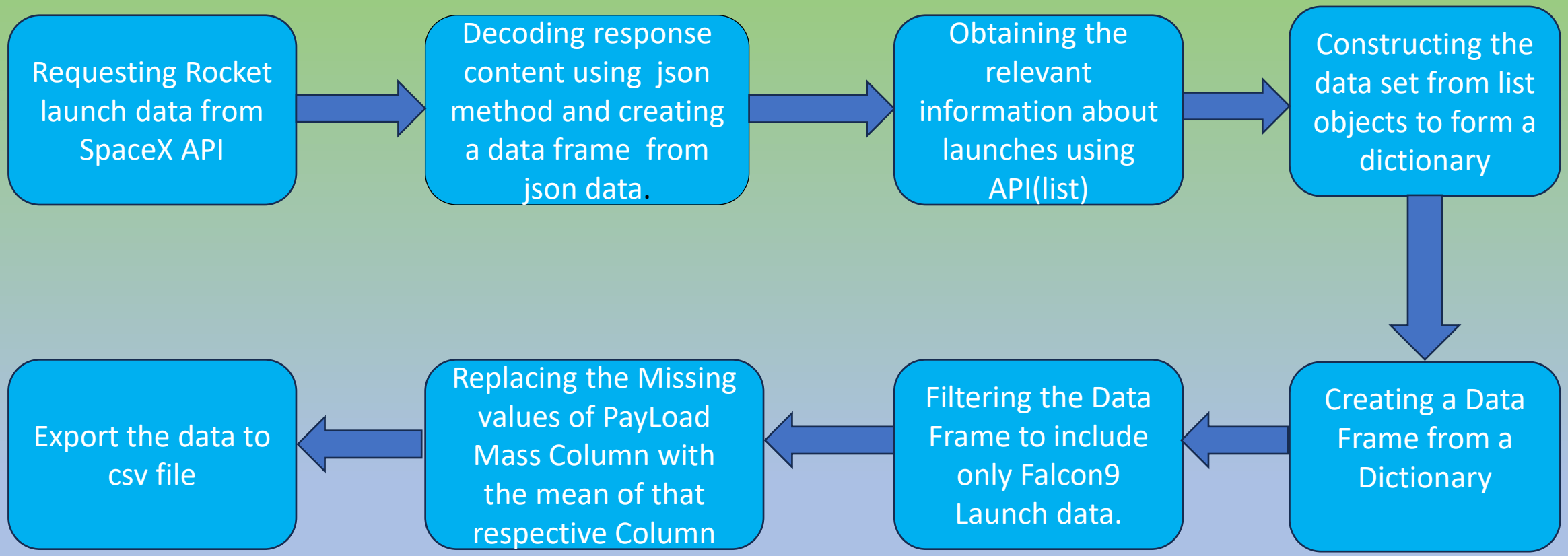
- This is the Initial step in Data Analysis .
- Data collection process involved a combination of using SpaceX REST APIs and Web Scraping of data from the website.
- Here we use this both data collection Methods in order to get complete information about the launches for a detailed analysis.
- Data Columns obtained by using SPACEX REST API:

  Flight Number, Date, Booster Version, Payload Mass, Orbit, Launch site, Outcome, Flights, Grid Fins, Reused, Legs, landing Pad, Block, Reused Count, Serial, Longitude, Latitude.

- Data Columns obtained by using Web Scraping

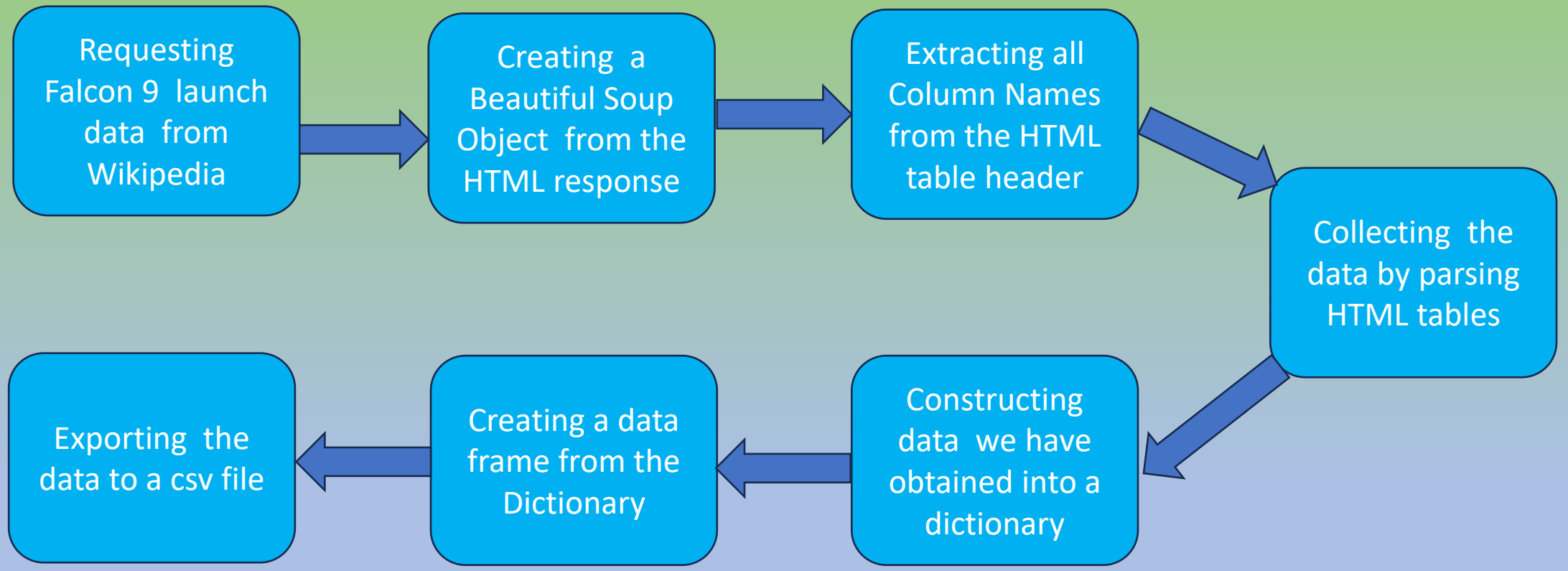
  Flight no, Launch Site, Payload, Payload Mass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API



Git Hub URL: Data Collection API

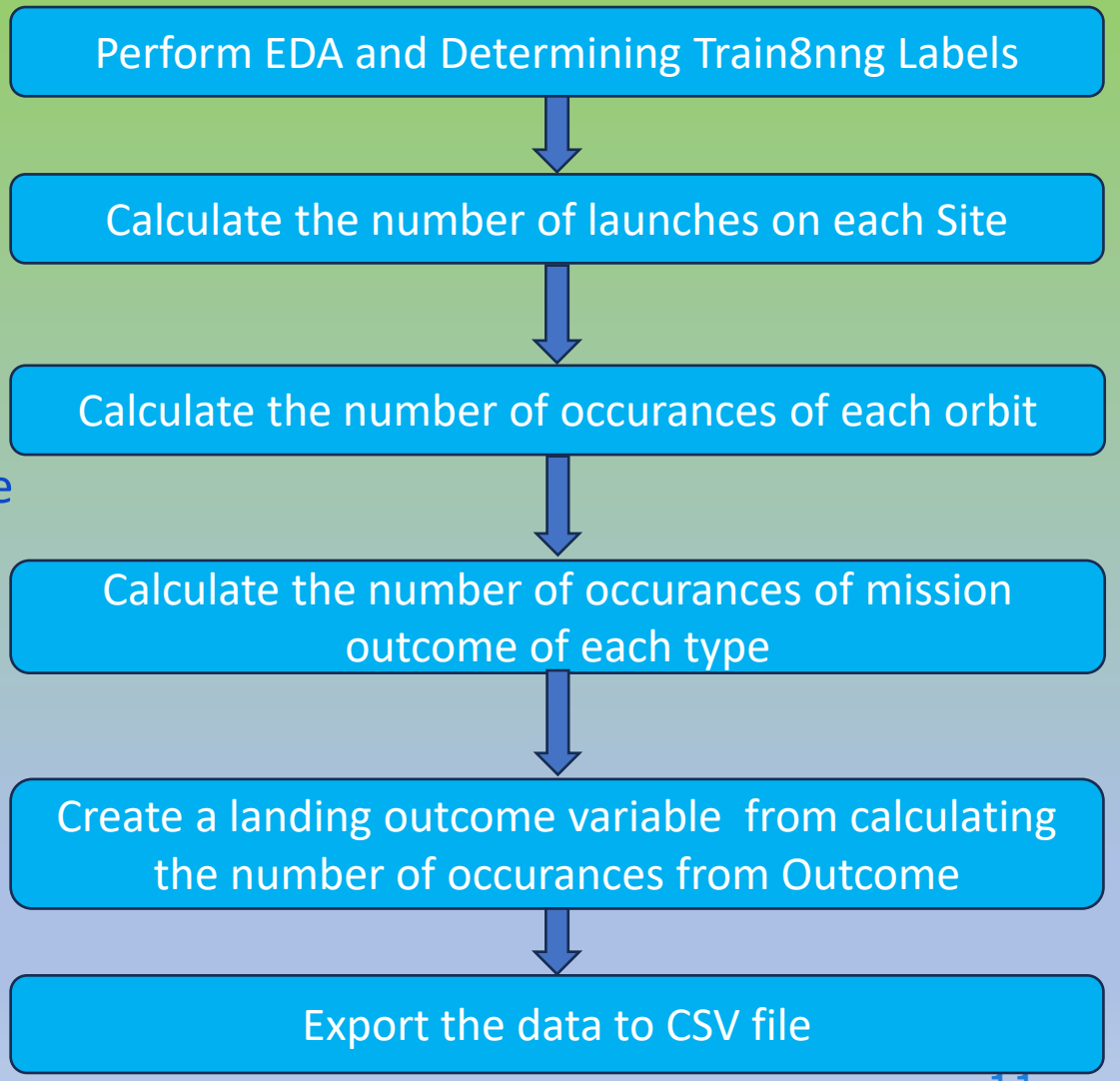
Data Collection - Scraping



Git Hub URL: Data Collection-Web Scraping

Data Wrangling

- In the data set there are cases where the booster does not land successfully.
- Sometimes the landing was attempted but failed due to accidents.
- Calculating all the required information to categorize the data into successful and unsuccessful landings.
- Converting the outcomes into Training Labels with ‘1’ means successfully landed and ‘0’ with unsuccessful landing.



Git Hub URL: Data Wrangling

EDA with Data Visualization

Scatter Plots:

Scatter Plots show the relationship between variables. If a relationship exists, they can be used in the Machine Learning Model.

Charts Plotted:

Flight Number vs Payload Mass, Flight Number vs Launch site, Payload Mass vs Launch Site, Flight Number vs Orbit Type, Payload Mass vs Orbit Type

- ## Bar Plot:

Bar Charts show comparisions among discrete categories. The Goal is to show the relationship between the specific categories being compared with a measured value.

Charts Plotted: Orbit Type vs Success Rate

Line Plot:

Line Charts show the trends in data over time (Time series)

Charts Plotted: Year Vs Success rate

Git Hub URL: EDA with Data Visualization

EDA with SQL

- Displaying the names of the unique launch sites in the SpaceX 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 the 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 pay load mass using a sub query
- Listing the records displaying the month names, failure landing outcomes in drone ship booster versions, launch site for the months in the year 2015
- Ranking the count of landing outcomes between the date 2010-06-04 and 2017-03-20 in descending order

Git Hub URL: EDA with SQL

Build an Interactive Map with Folium

❑ 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 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.

❑ Colored Marked of the launch outcomes for each Launch Site:

- Added colored colored 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 Proximites:

- Added colored lines to show distances between the Launch site KSC LC-39A()example) and its proximities like Railway, Highway, Coastline and Closest City
- Git Hub URL: Interactive Visual Analytics with Folium

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:

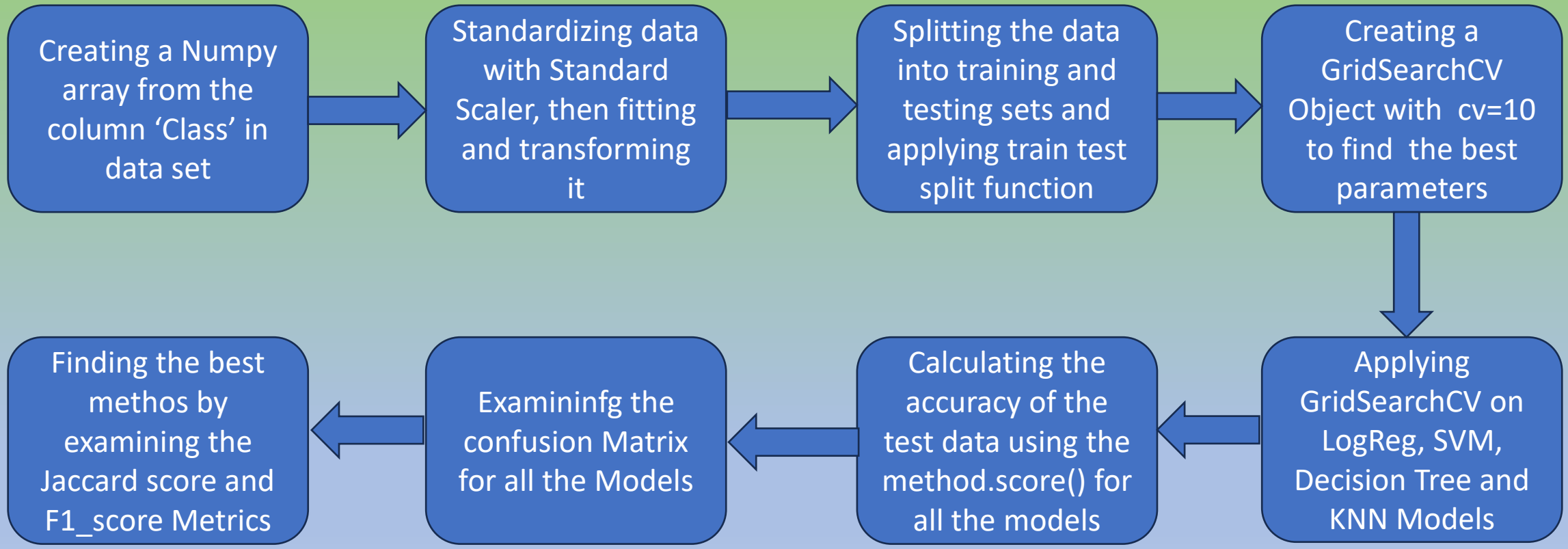
  Added a slider to select to select Payload range.

- Scatter Chart of Payload Mass vs Success Rate for the different Booster Versions:

  Added a scatter chart to show the correlation between Payload and Launch Success.

Git Hub URL: Plotly Dashboard

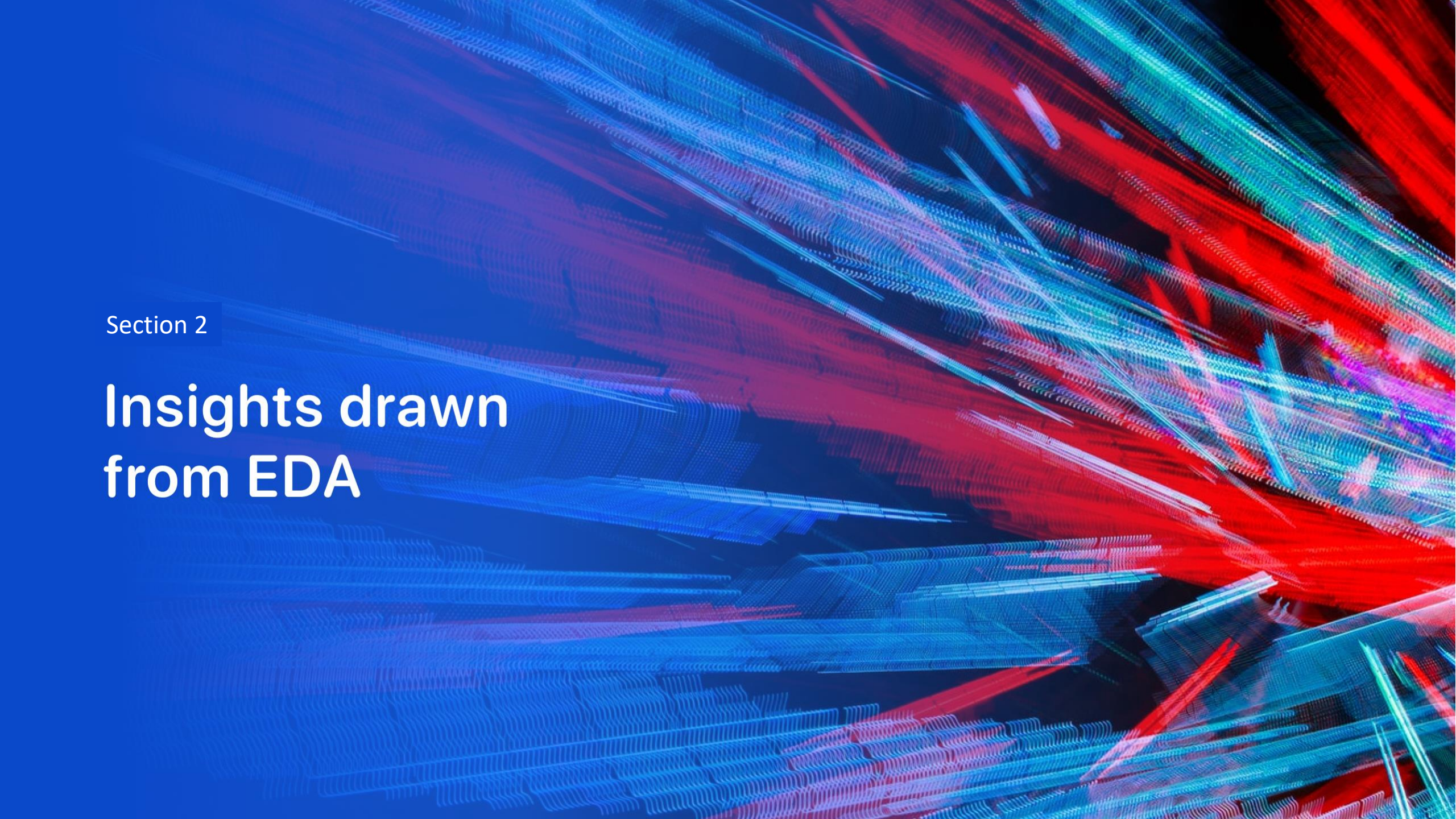
Predictive Analysis (Classification)



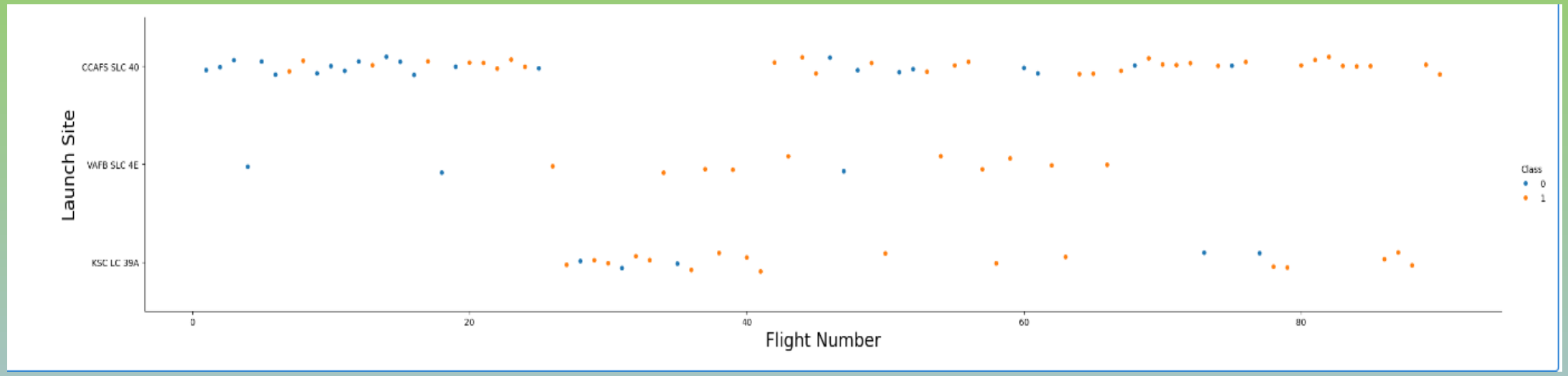
Git Hub URL: Machine Learning Prediction

Results

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


Section 2
Insights drawn from EDA

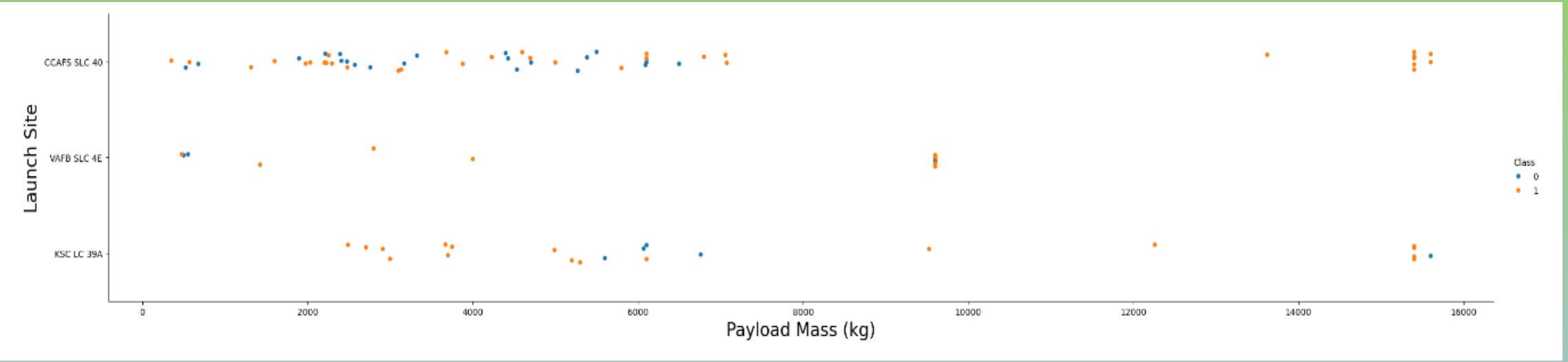
Flight Number vs. Launch Site



Explanation:

- The Earliest Flights failed while the latest flights all succeeded.
- The CCAFS SLC 40 has launched more when compared to the other sites.
- VAFB SLC 4E and KSC LC 39A have higher success rates.
- It can be assumed that each new launch have a higher success rate.

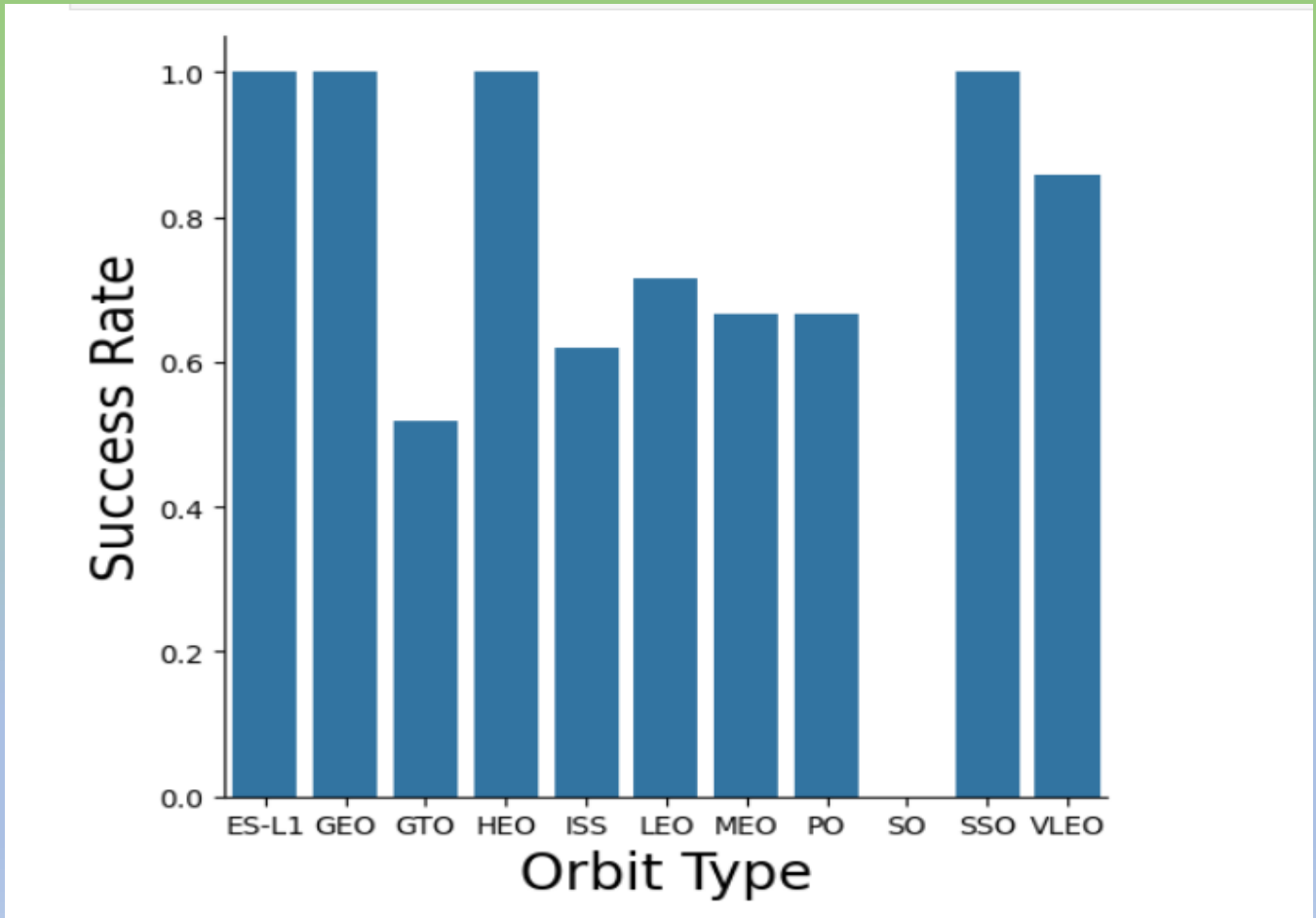
Payload vs. Launch Site



Explanation:

- The Launches were successful for most of the cases of KSC LC 39A and VAFB SLC 4E for any Payload Mass applied in the attempt.
- All the launches of KSC LC 39A were successful with Payload less than 5500kg.

Success Rate vs. Orbit Type



Explanation:

- Orbits having 100% success rate:

  ES-L1, GEO, HEO, SSO

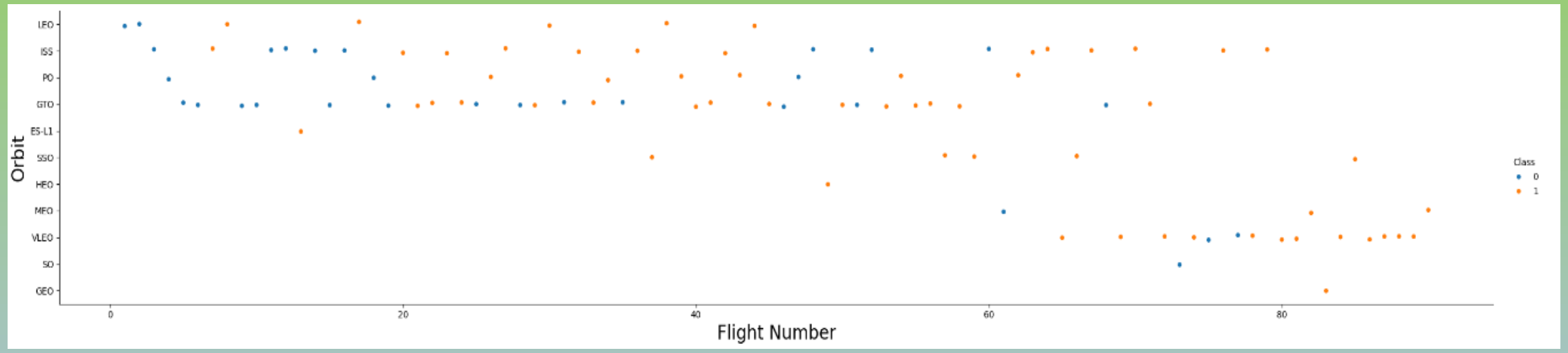
- Orbits having 0% success rate:

  SO

- Orbits having success rate greater than 50% and less 75%

  GTO, ISS, LEO, MEO, PO

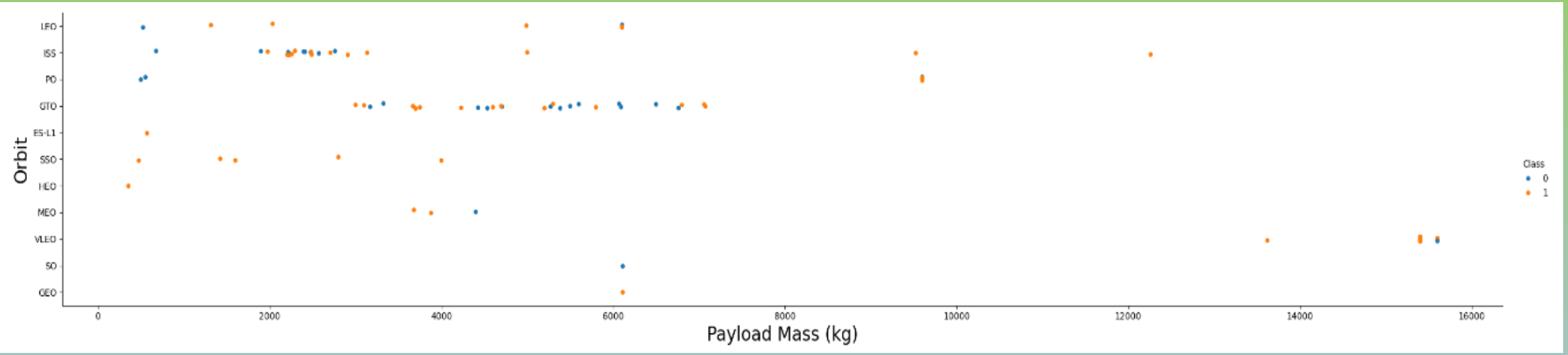
Flight Number vs. Orbit Type



Explanation:

- The Maximum number of flights were launched at the orbits LFO, ISS, PO, GTO, ES-L1
- Most of the Initial ’20’ Flight Launches were unsuccessful.
- Later on , we can observe most of the flight launches were successfully landed.

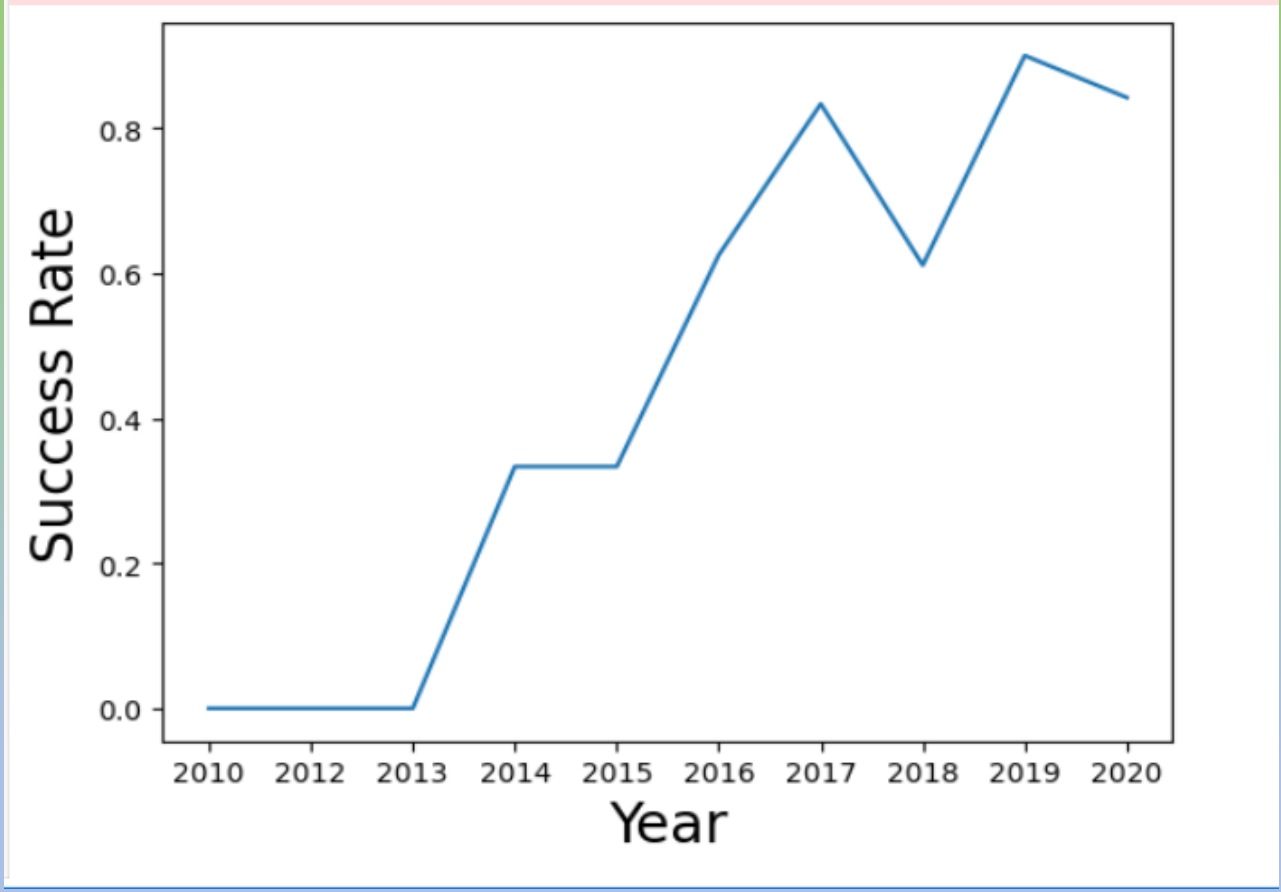
Payload vs. Orbit Type



Explanation:

- Most of the Launches were made with Payload Mass less than 8000.
- SSO orbit has 100% success rate with Payload less than 4000.
- We can see that many launches were attempted with Payload grater than 3000 and less than 7000 at GTO orbit.

Launch Success Yearly Trend



Explanation:

- The Successful Launches of Flights were started from the year ‘2013’
- There is tremendous increase in the success rate from ‘2015’ to ‘2017’
- The Success rate is constant in the period 2014-2015.

EDA With SQL

All Launch Site Names

- Find the names of the unique launch sites

  There are 4 unique Launch Sites included in the Project.

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

```
%sql select distinct launch_site from SPACEXTABLE;
```

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

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`

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

[14]:

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

- The Image above showing the 5 records where the launch site begin with ‘CCA’.

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
[17]: %sql Select sum(PAYLOAD_MASS__KG_) as Total_Payload_Mass from SPACEXTABLE where Customer='NASA (CRS)'
 * sqlite:///my_data1.db
Done.
[17]:
```

Total_Payload_Mass
45596

- The Total Payload carried by the boosters from NASA (CRS) is ‘45596’.

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

```
[20]: %sql Select AVG(PAYLOAD_MASS__KG_) as Average_Payload_mass from SPACEXTABLE where Booster_Version like '%F9 v1.1%'
```

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

[20]:

Average_Payload_mass
2534.6666666666665

- The Average Payload Mass carried by the booster version F9 v1.1 is 2534.67

First Successful Ground Landing Date

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

```
[22]: %sql select min(date) as first_successful_landing from SPACEXTABLE where landing_outcome = 'Success (ground pad)';
 * sqlite:///my_data1.db
Done.
[22]: first_successful_landing
------------------------------
                    2015-12-22
```

- The First successful landing outcome on ground pad is ‘2015-12-22’.

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

```
[23]: %sql select booster_version from SPACEXTABLE where landing_outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;
```

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

[23]:

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

- The Boosters F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2 had successfully landed on drone ship, which are having Payload mass greater than ‘4000’ and less than ‘6000’

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
[26]: %sql select mission_outcome, count(*) as total_number from SPACEXTABLE group by mission_outcome;
 * sqlite:///my_data1.db
Done.
```

[26]:

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

- Total number of Successful mission Outcomes is ‘99’
- Total number of failure mission outcomes is ‘2’

Boosters Carried Maximum Payload

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

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

```
[27]: %sql Select  booster_version from SPACEXTABLE where PAYLOAD_MASS__KG_= (Select max(PAYLOAD_MASS__KG_) from SPACEXTABLE);
 * sqlite:///my_data1.db
Done.
```

[27]:

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

- The Image showing the booster versions which have carried the maximum payload mass.

2015 Launch Records

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

[36]:
```
%%sql select substr(Date, 6,2) as month, date, booster_version, launch_site, landing_outcome from SPACEXTABLE
      where landing_outcome = 'Failure (drone ship)' and substr(Date,1,4)= '2015';
```

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

[36]:

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)

- There was unsuccessful landing(drone ship) on 10th of January and 14th April in the year 2015.

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

- The Image Displaying the count of the landing outcomes in the descending order between 2010-06-04 and 2017-03-20.

[34]:
```
%%sql select landing_outcome, count(*) as count_outcomes from SPACEXTABLE
      where date between '2010-06-04' and '2017-03-20'
      group by landing_outcome
      order by count_outcomes desc;
```

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

[34]:

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

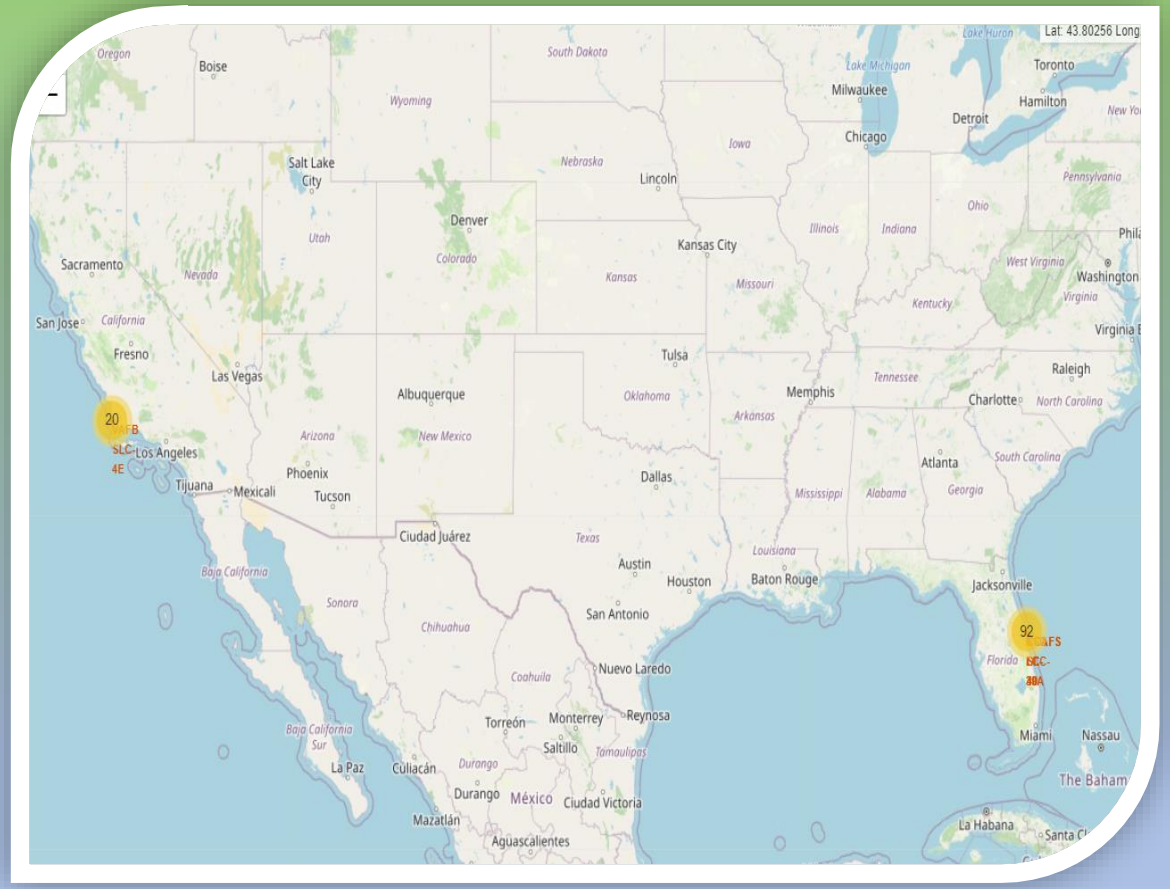
Section 3

Launch Sites Proximities Analysis

All Launch Sites location on a Global Map

Most of the Launch sites are in proximity to the Equator line. If a ship is launched from the equator it goes up into space, and also moving around the Earth at the same speed as before when it has launched. This is because of inertial. This speed will help the spacecraft keep up a good enough speed to stay at orbit.

All Lauch Sites are in very close proximity to the coast,while launching rockets towards the ocean it minimizes the risk of having any debris dropping or exploding near people.

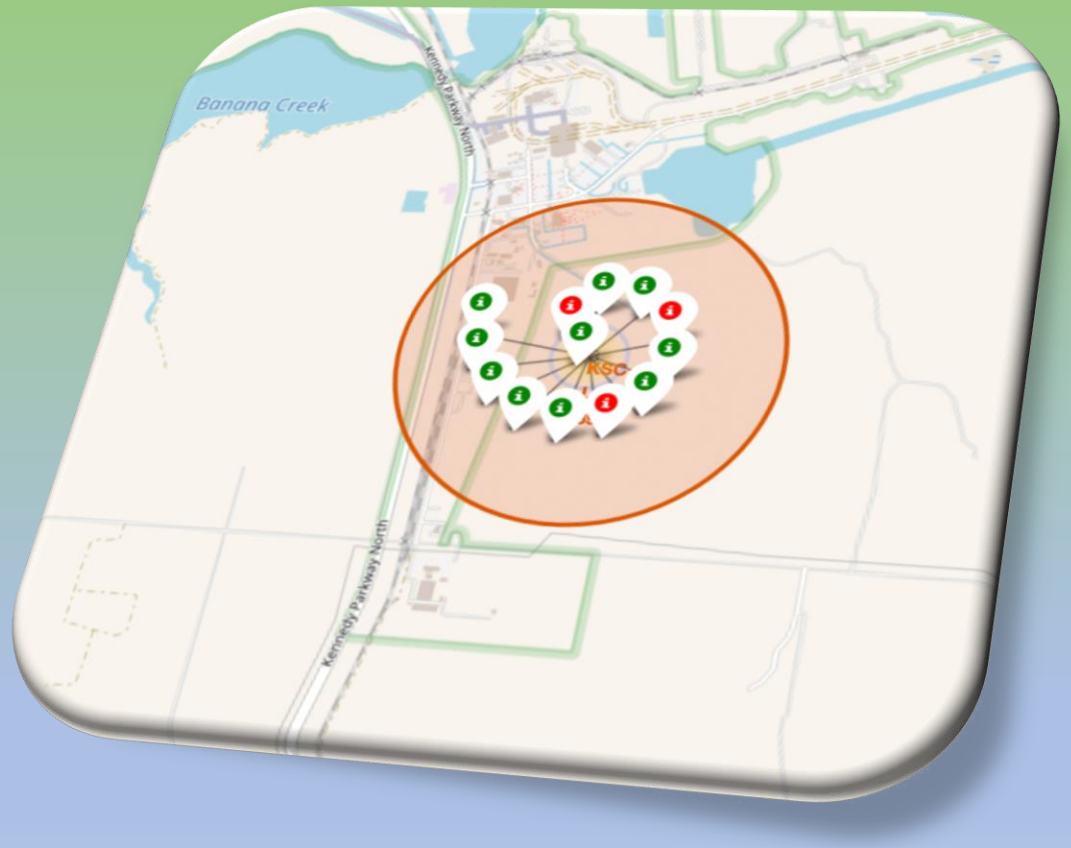


Color-launched launch records on the Map

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

Launch Site KSC LC-39A has very high Success rate.

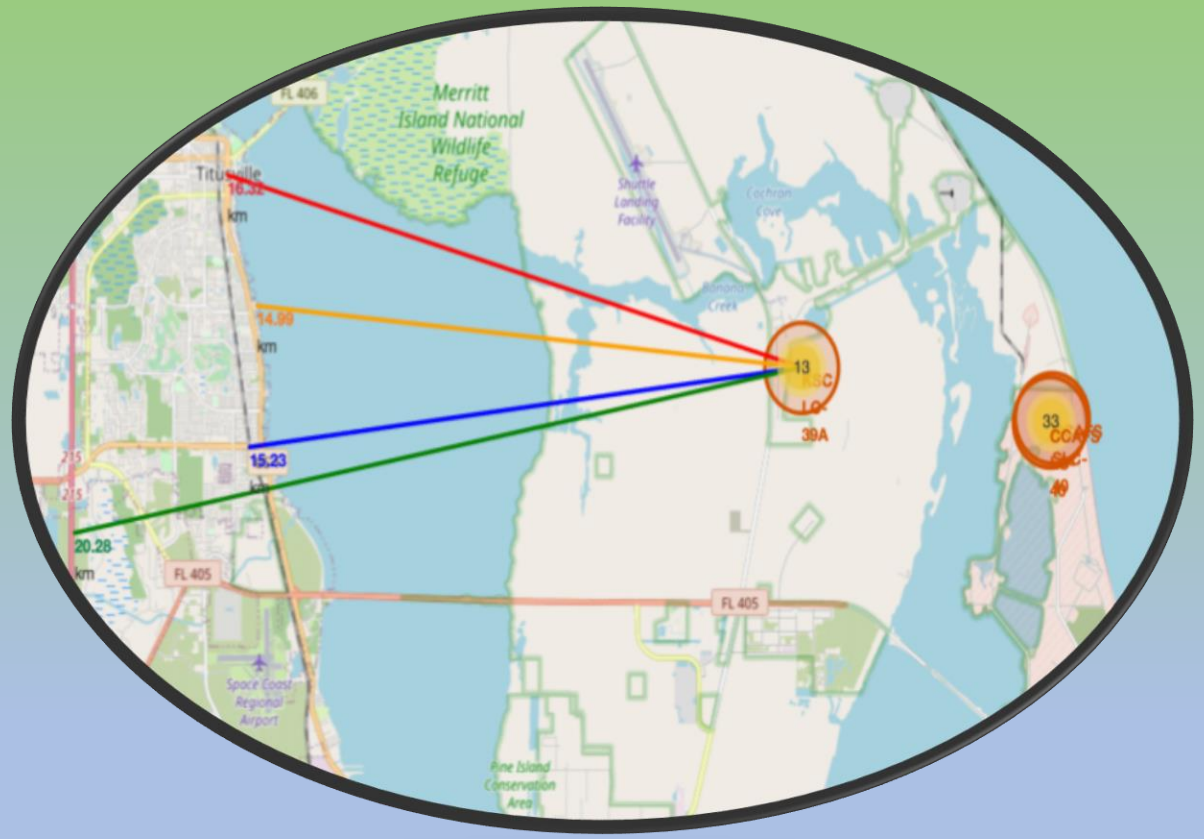


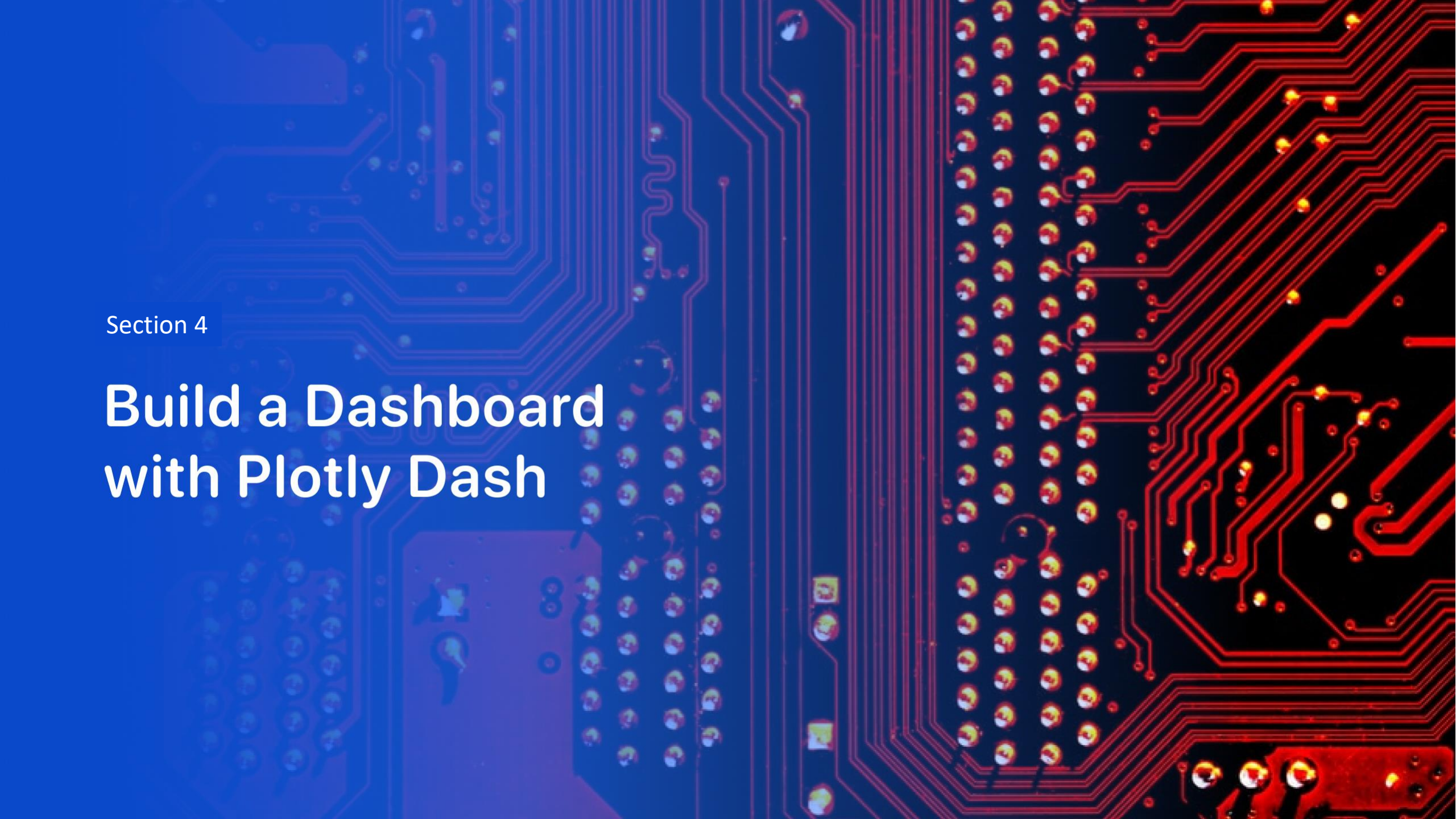
Distance from the launch site KSC LC-39A to its proximities

- From the Visual analysis of the Launch site KSC LC-39A we can clearly see that it is:
  - relative close to railway (15.23km)
  - relative close to highway (20.28km)
  - relative close to coastline (14.99km)

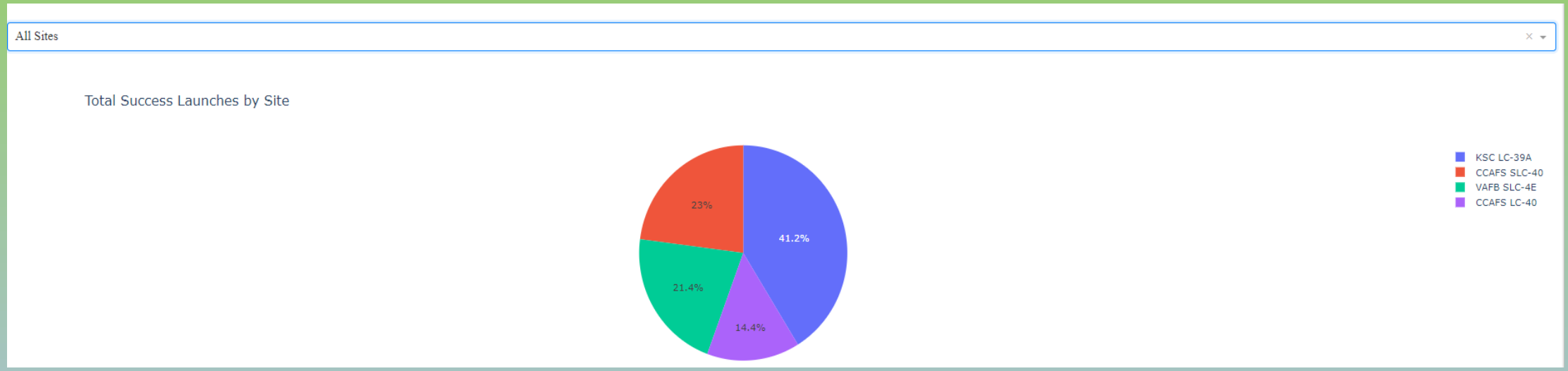
Also the Launch site KSC LC 39-A is relative close to closest city Titusville.

Failed rocket with its high speed can cover distances like 15-20 km in few seconds. It could be potentially dangerous to populated areas.



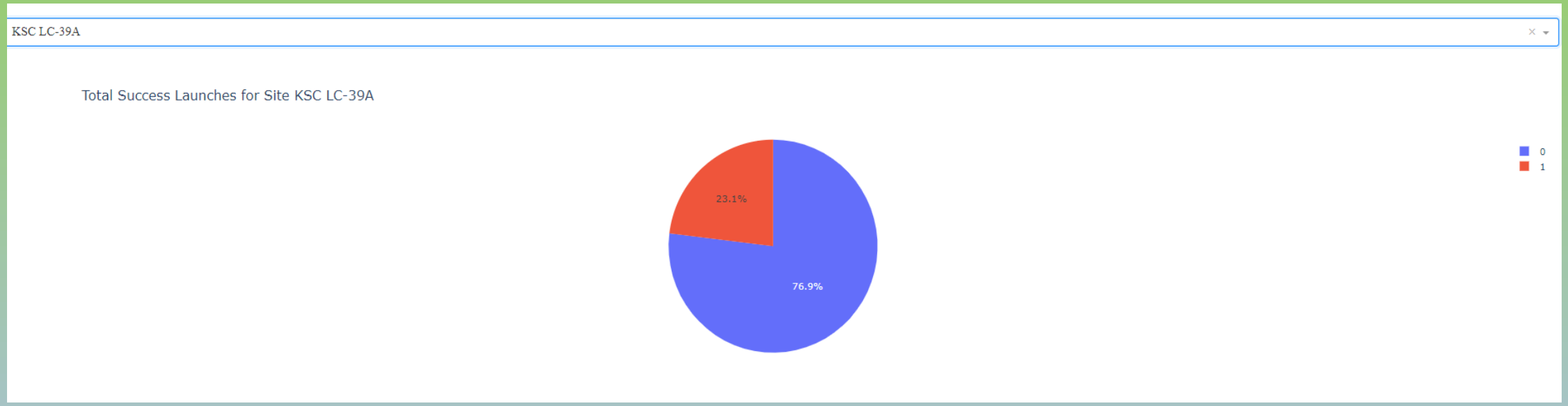

Section 4
Build a Dashboard with Plotly Dash

Launch Success rate for all Sites



- Explanation:
  - Success Rate of KSC LC-39A is '41.2%'
  - Success Rate of CCAFS SLC-40 is '23%'
  - Success Rate of VAFB SLC-4E is '21.4'
  - Success Rate of CCAFS LC-40 is '14.4%'

Launch Site With Highest Launch Success Ratio



Explanation:

- KSC LC 39-A is the launch site with highest launch success ratio.
- From the pie chart we can observe the success rate for KSC LC 39-A is about 76.9%

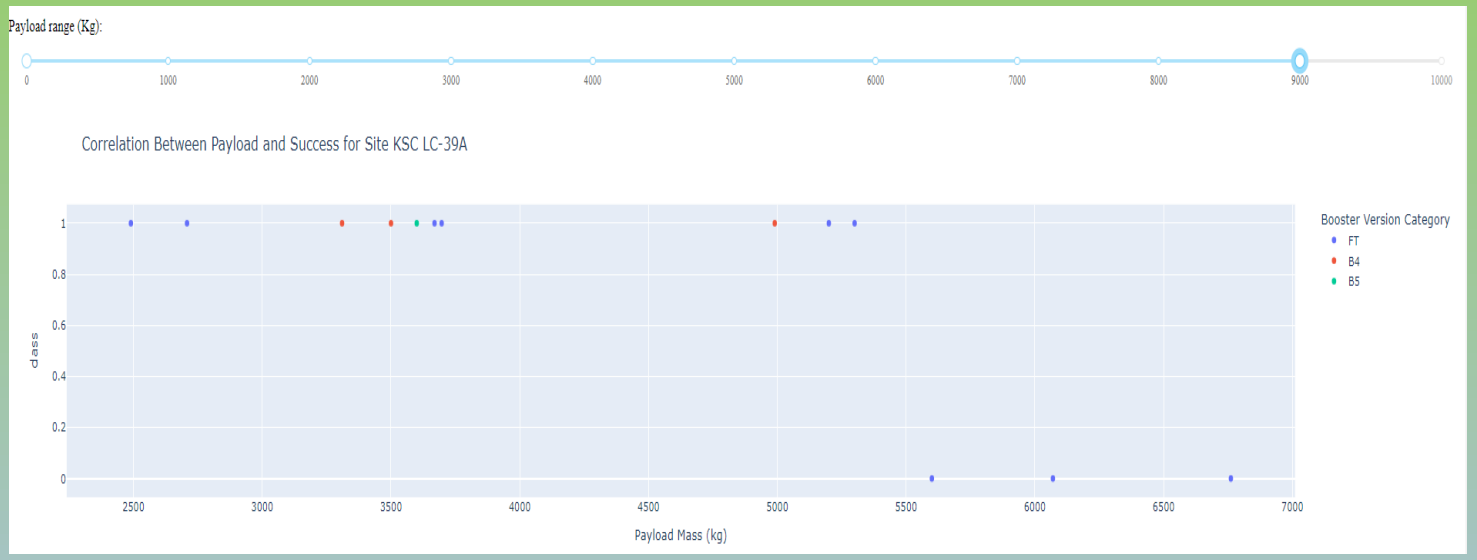
Payload Mass vs. Launch Outcome for all Sites

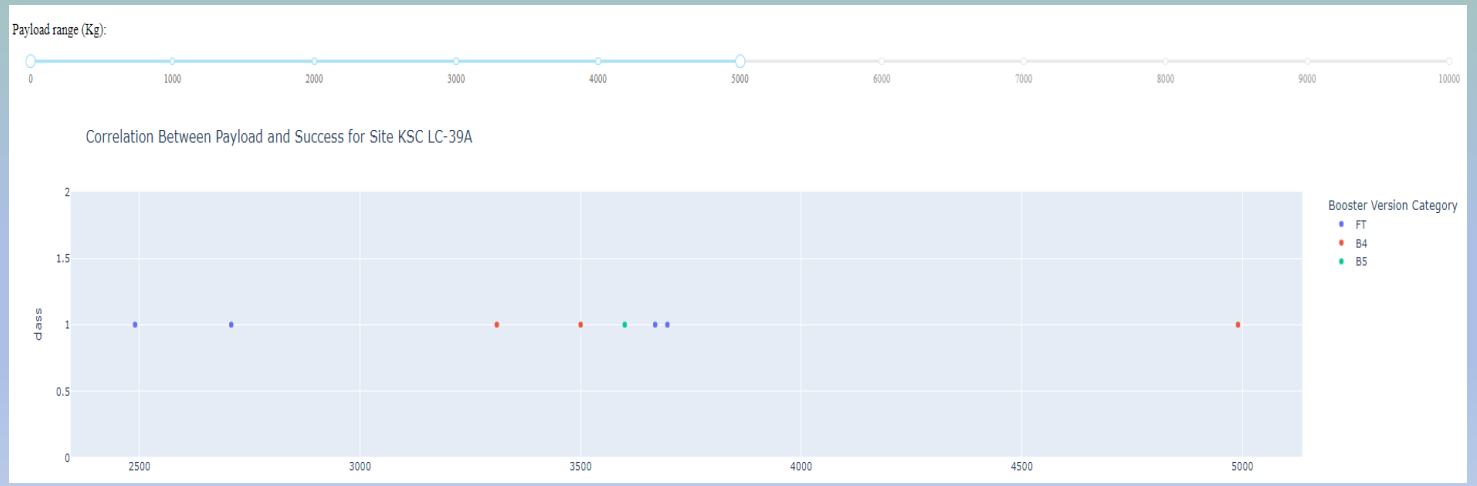
Explanation:

The Charts showing the scatter plot view, the relation between the variables ,Payload Mass and Class .

We can observe the success rate changes for every payload Mass Range.

As the Payload increases there is an increase in the success rate.



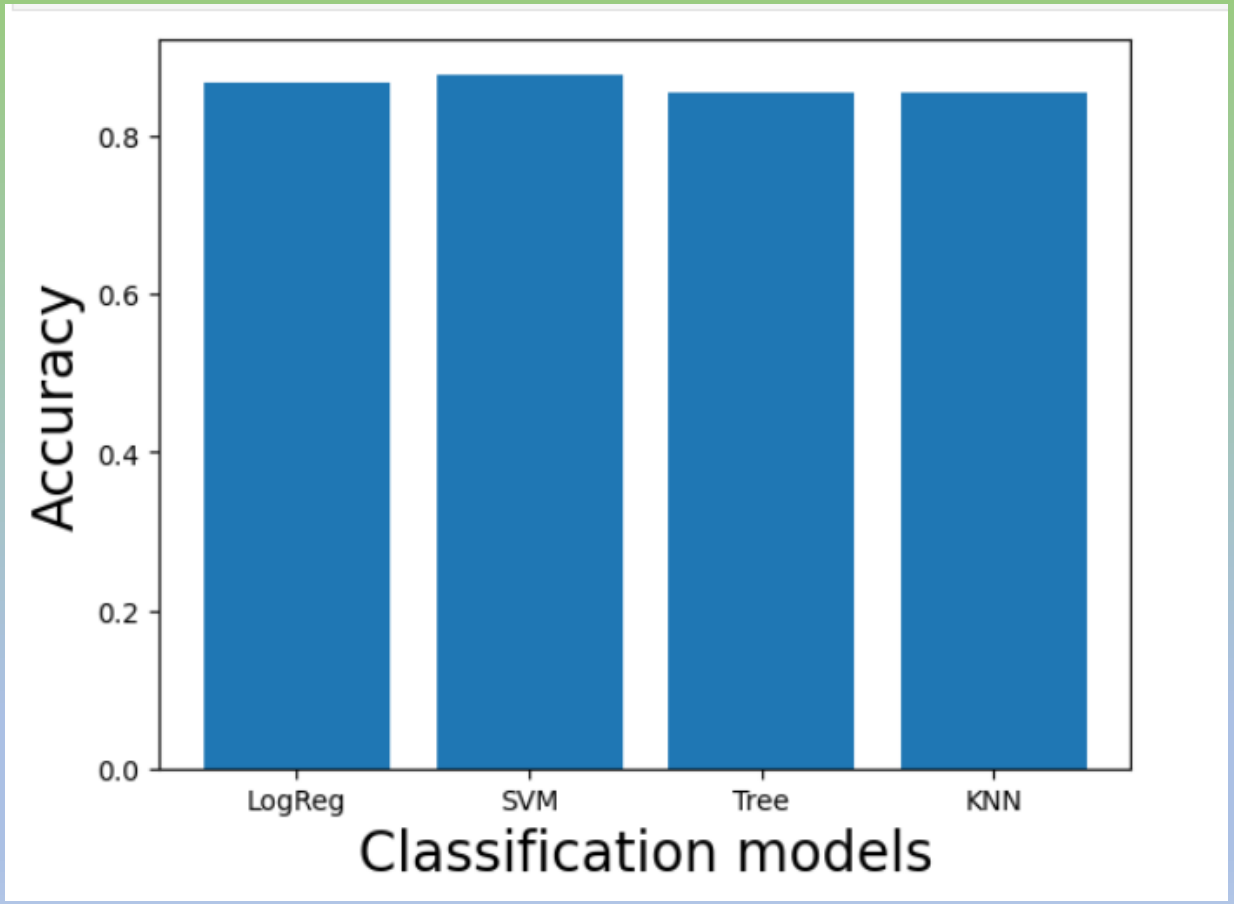


Section 5

Predictive Analysis (Classification)

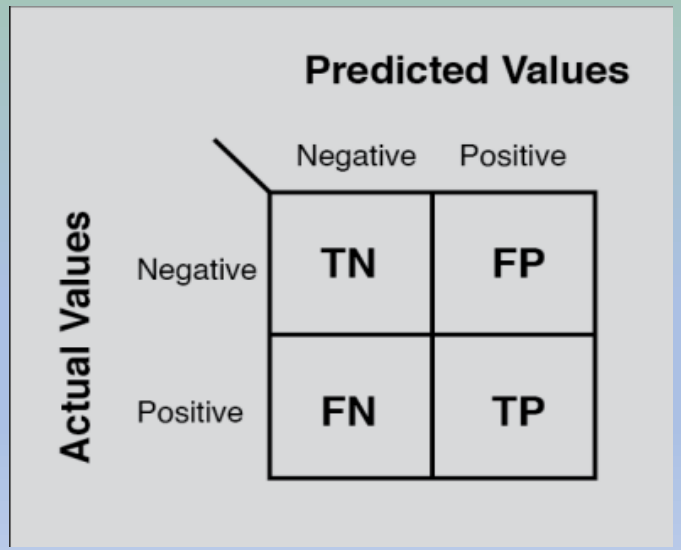
Classification Accuracy

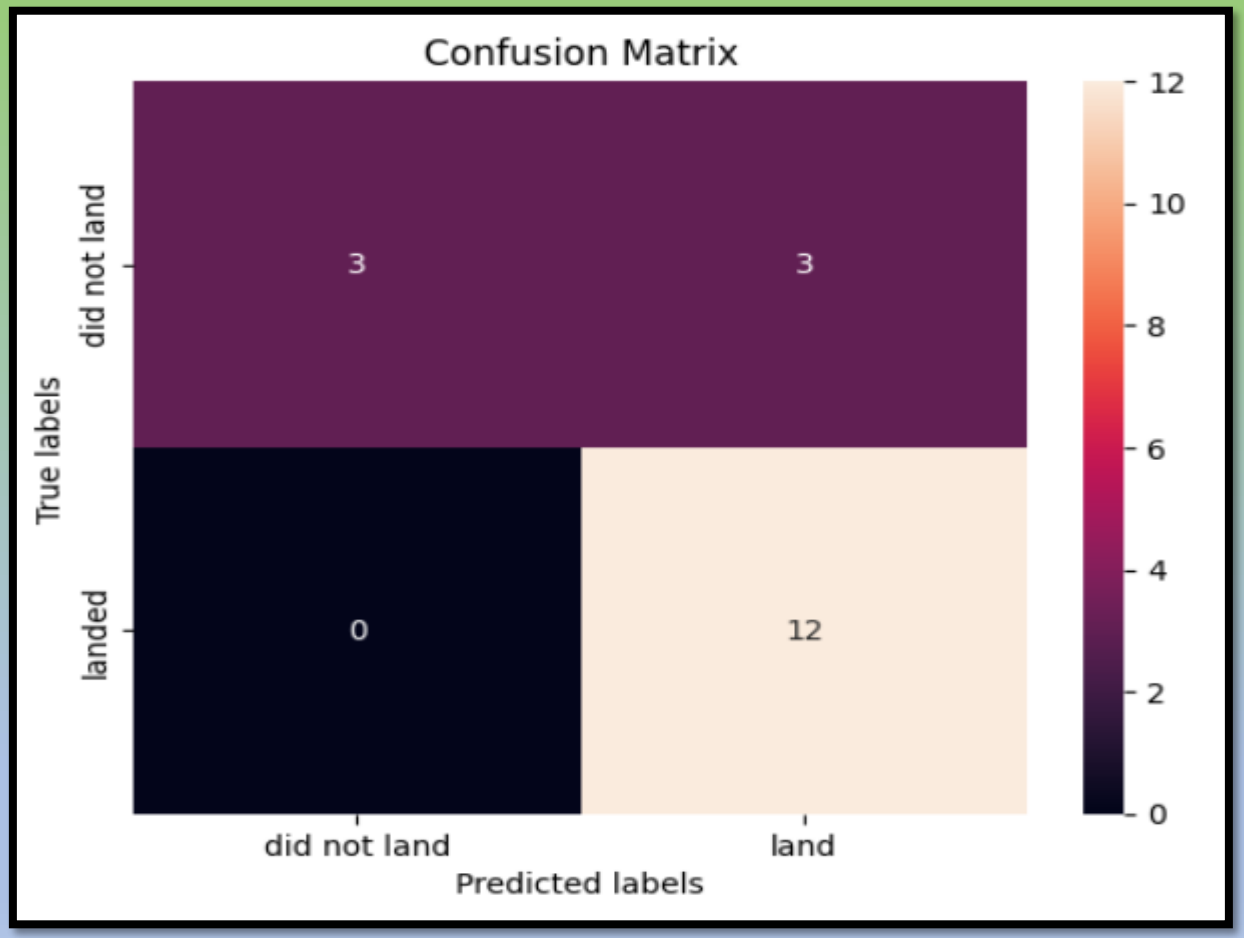
- All the models has the same accuracy when the model is test with only the test data.
- In this Case, we can find the best model when the model is tested with the entire dataset.
- This results in SVM is the best suitable classification algorithm for prediction with accuracy about 90%.



Confusion Matrix

- Examining the Confusion matrix, we see that the logistic regression can distinguish between the different class labels.
- This matrix is constructed with True/Actual labels and Predicted labels





Conclusions

- Data is collected using API's and Web Scraping.
- CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, AACFS SLC-40 are the four launch sites.
- Most of the launch sites are in proximity to the Equator and all the sites are in very close proximity to the coast.
- The Success rate of launches increased over the years.
- KSC LC 39-A has the highest success rate of the launches among all the launch sites.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.

Appendix

- Complete Data Analysis from Data Collection to Prediction:
  - Applied Data Science Capstone - Complete analysis

- Data collected from the website for analysis:
  - Data Collection - Web Scraping

Thank you!