IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ruby Sharma
11/21/2025

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

Summary of methodologies

- We collected SpaceX data using API and web-scrapping. Performed data wrangling by replacing missing values with mean data.
- Plotted scatter plots to showcase the relationship between different variables.
- Executed SQL queries to extract important information like unique launch site, average payload mass carried by Falcon9, the first successful landing outcome date, unique landing outcomes.
- Bulit and interactive folium map to showcase the geospatial information of all the sites and plotly dashboard to show the proportion of successful mission.
- Performed predictive analysis by dividing the data into x(independent variables) and y (target variable) data. Divided the data into train and test data. Created four models and estimated best parameters using grid search and computed their accuracy using score method on test data.

Summary of all results

- We determined there are four unique launch sites located close to coastline and highway and far from cities.
- KSC LC-39A site has the most successful landings.
- We observed that rockets having heavy payload mass can also have safe landing of the first stage.
- We observed that the increase in the number of attempts of flights increases the chances of safe landings.
- Rockets launched to SSO and VLEO orbits has the most successful landing proportions. These orbits are closer to earth.
- Rockets launched to VELO carries high payload mass, still has successful landings.
- We employed four models, Logistic Regression, Support vector machine, Decision Tree and K Nearest Neighbor.
- All four models has an accuracy of 83.33 % and they all performed equally well.

Introduction

- Companies are building spacecrafts to make space travel affordable.

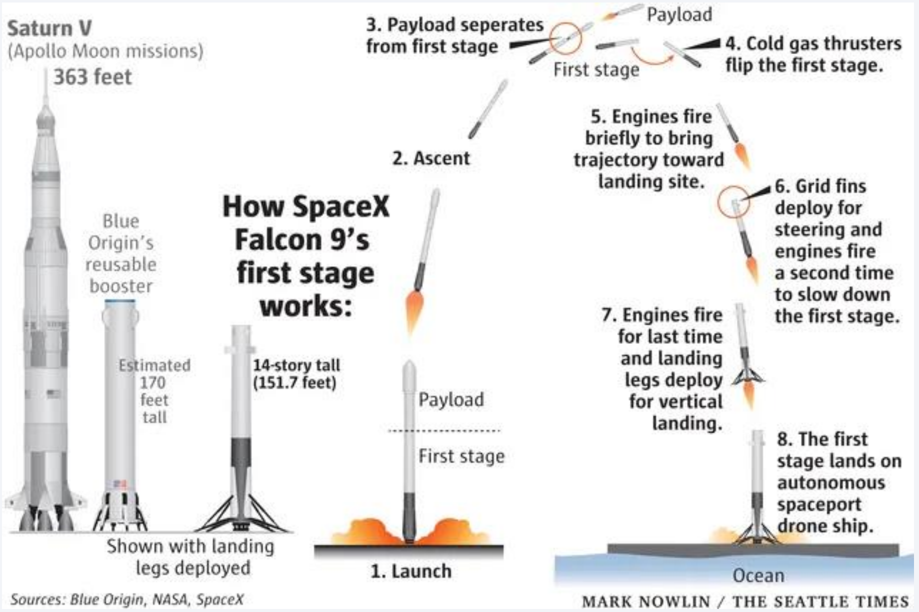
- Some of those companies are Virgin Galactic, Rocket Lab, Blue Origin and SpaceX.

- Out of these companies, SpaceX is most successful, accomplished and offers low cost for the launch.

- SpaceX Falcon9 rocket launch costs around 62 million dollars, whereas other companies offers around 165 million dollars.

- The secret behind SpaceX Falcon9 low cost is the reusage of its first stage.

- On right we can see the launch of SpaceX Falcon 9 rocket and how the first stage works and successfully recovered, that can be used later on.



Saturn V
(Apollo Moon missions)
363 feet
Blue Origin's reusable booster
Estimated 170 feet tall
How SpaceX Falcon 9's first stage works:
14-story tall (151.7 feet)
Shown with landing legs deployed
1. Launch
Payload
First stage
2. Ascent
3. Payload seperates from first stage
Payload
First stage
4. Cold gas thrusters flip the first stage.
5. Engines fire briefly to bring trajectory toward landing site.
6. Grid fins deploy for steering and engines fire a second time to slow down the first stage.
7. Engines fire for last time and landing legs deploy for vertical landing.
8. The first stage lands on autonomous spaceport drone ship.
Ocean

Sources: Blue Origin, NASA, SpaceX

MARK NOWLIN / THE SEATTLE TIMES

As a Data Scientist, we want to predict the successful landing of first stage based on the rocket launch features.

Section 1

Methodology

Methodology

Executive Summary

- **Data collection methodology:**
  - One of the ways, we collected the Data was using API call through request.get. and then normalized the data from json to pandas dataframe.
  - In another way, we performed wed scraping of Wikipedia page using BeautifulSoup library. Extracted all the tables and columns names.
- **Perform data wrangling:**
  - Determined which columns has missing values and replaced it with mean of the rest of the column values.
- **Perform exploratory data analysis (EDA) using visualization and SQL:**
  - Plotted scatter plots between different variables of launch data to observe any relationship. Plotted bar chart and line plot to get estimate of success rate.
  - Executed SQL queries to extract important information like unique launch site, average payload mass carried by Falcon9, the first successful landing outcome date, unique landing outcomes.

Methodology Contd.

Executive Summary

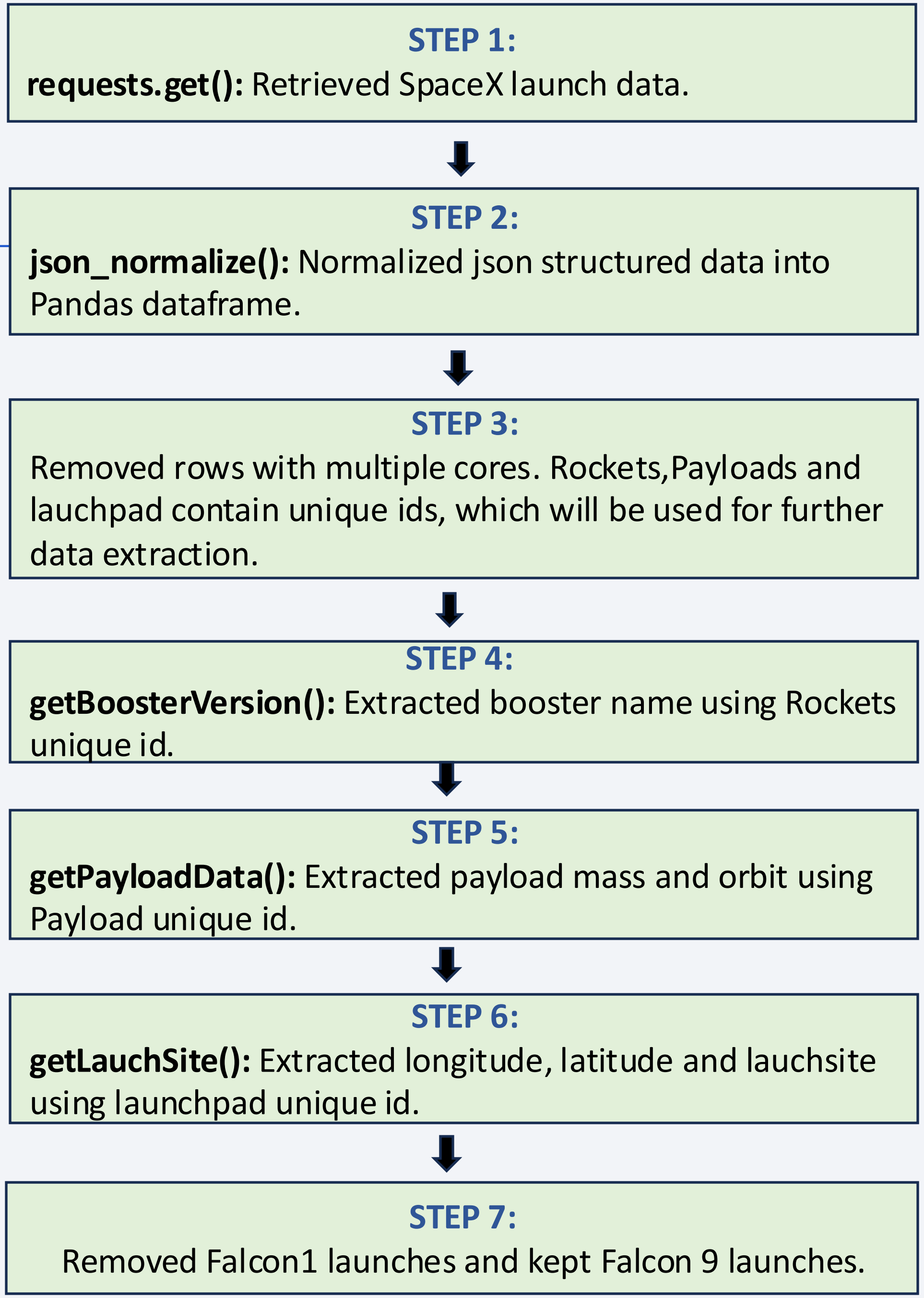
- **Interactive Visual Analytics and Dashboard:**
  - Built an interactive folium map to visualize the geo-spatial location of the launch sites and to see what places are closer to site which might affect the success rate.
  - Used plotly library to create an interactive dashboard to visualize the pie chart showcasing the success rate of all sites together and individually. Also included a scatter plot to show the relationship between payload mass and mission outcome.
- **Predictive analysis (model building and evaluation):**
  - Divided the spaceX launch data into X (input features) and Y (target) variables. Standardized the X variable and split both variables using train_test_split function randomly.
  - We then created model objects, and performed GridSeach using best possible parameters to find the best parameters.
  - Using the best hypermeter values, we determined model accuracy on test data, finally obtained the confusion matrix.
  - We employed four ML algorithms, Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors.

Data Collection – SpaceX API

- We downloaded data using API calls.

- On right we have the flowchart of how we collected the data.

- Each box shows the step no, with methods used (in bold letters) followed by the action.



STEP 1:
requests.get(): Retrieved SpaceX launch data.

STEP 2:
json_normalize(): Normalized json structured data into Pandas dataframe.

STEP 3:
Removed rows with multiple cores. Rockets,Payloads and lauchpad contain unique ids, which will be used for further data extraction.

STEP 4:
getBoosterVersion(): Extracted booster name using Rockets unique id.

STEP 5:
getPayloadData(): Extracted payload mass and orbit using Payload unique id.

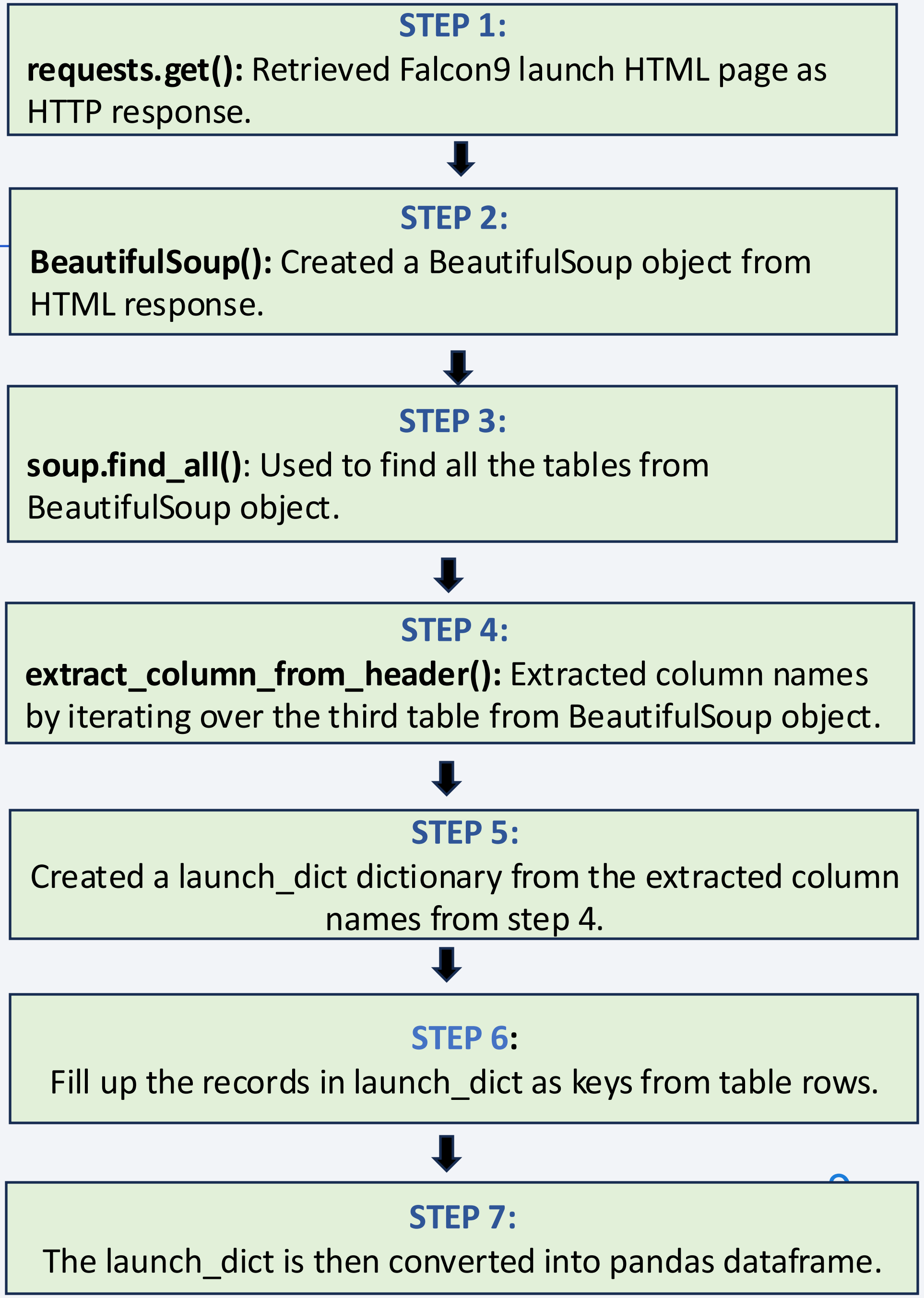
STEP 6:
getLauchSite(): Extracted longitude, latitude and lauchsite using launchpad unique id.

STEP 7:
Removed Falcon1 launches and kept Falcon 9 launches.

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- We used BeautifulSoup for extracting Falcon 9 launch data in the form of HTML table from Wikipedia.
- Next, we parsed the table and converted it into Pandas dataframe.
- On right we have the flowchart of how we collected the data.
- Each box shows the step no, with methods used (in bold letters) followed by the action.



STEP 1:
requests.get(): Retrieved Falcon9 launch HTML page as HTTP response.

STEP 2:
BeautifulSoup(): Created a BeautifulSoup object from HTML response.

STEP 3:
soup.find_all(): Used to find all the tables from BeautifulSoup object.

STEP 4:
extract_column_from_header(): Extracted column names by iterating over the third table from BeautifulSoup object.

STEP 5:
Created a launch_dict dictionary from the extracted column names from step 4.

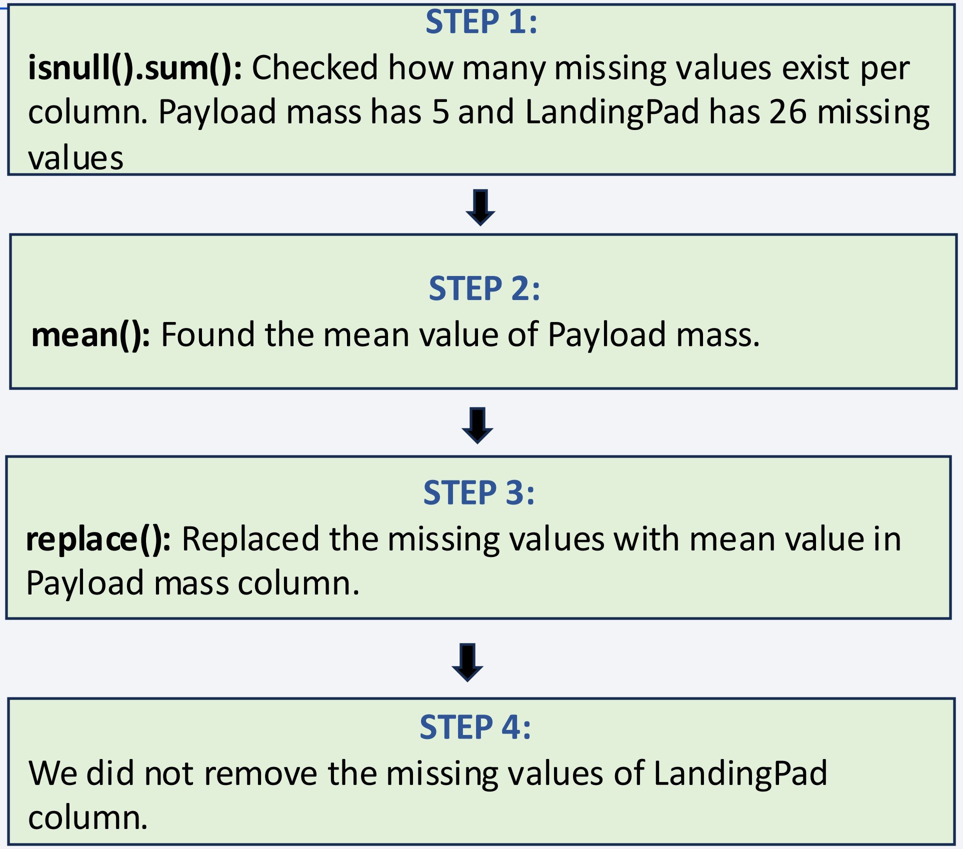
STEP 6:
Fill up the records in launch_dict as keys from table rows.

STEP 7:
The launch_dict is then converted into pandas dataframe.

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- On right we have the flowchart of how we processed the data.
- Each box shows the step no, with methods used (in bold letters) followed by the action.



STEP 1:

isnull().sum(): Checked how many missing values exist per column. Payload mass has 5 and LandingPad has 26 missing values

STEP 2:

mean(): Found the mean value of Payload mass.

STEP 3:

replace(): Replaced the missing values with mean value in Payload mass column.

STEP 4:

We did not remove the missing values of LandingPad column.

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

EDA with Data Visualization

- I have utilized the scatterplot to view the relationship between variable pairs like “Flight Number and Launch Site”, “Payload Mass and Launch Site”, “FlightNumber and Orbit””PayloadMass and Orbit” in relation to success or failure of first stage return

- A bar plot is used to visualize what is the success rate based on each orbit type.

- A line chart is used to see the success rate change from year 2010 to year 2020.

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/edadataviz.ipynb

EDA with SQL

1. I retrieved the unique launch sites of spaceX using Distinct command in SQL.
2. Displayed top five records of launch details of launch_site started with “CCA%”.
3. Displayed the total payload mass carried by the boosters launched by NASA (CRS).
4. Display average payload mass carried by booster version F9 v
5. Observed unique landing outcome of spaceX.
6. I retrieved the date when the first successful landing outcome in ground pad was achieved.
7. Got the list of the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
7. Got the count of total number of successful and failure mission outcomes.
8. Retrived the list of booster_version that carries the maximum payload mass.
9. Extracted the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
10. Ranked 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 link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

The success rate can also depend on the location and proximities of a launch site. Therefore, with the folium map we are Trying to find the location of launch sites and what places are closer to them.

Location of Launch sites

- First the latitude and longitude of each sites was retrieved from the dataset.
- Used Folium circle to circle the location of each four sites.
- Used Folium marker, with site names and color to add the text on each sites.

Success and failure outcomes at each sites

- At each site location, cluster marker was created.
- The cluster markers represents the row or data point of site and its corresponding success and failure.
- These cluster markers are then customized with icon to show the total
- number of success and failures using two different colors.
- The goal was to see which sites has good number of success so that we can learn what is unique about these sites.

Places in proximity to each sites

- Mouse position function was used to find the coordinates of any mouse position on the map.
- Calculate distance function is used to find the difference between two locations.
- Calculated distance between sites and places like coastline, highway and city.
- The distance is then added to the map to see which places are closer to the sited.

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- I have created a dashboard using Plotly library in Python.
- The dashboard has two components one is to visualize key information regarding the total number of success and failures of outcome. Another is to visualize the relationship between payload mass and mission outcome.

Top component

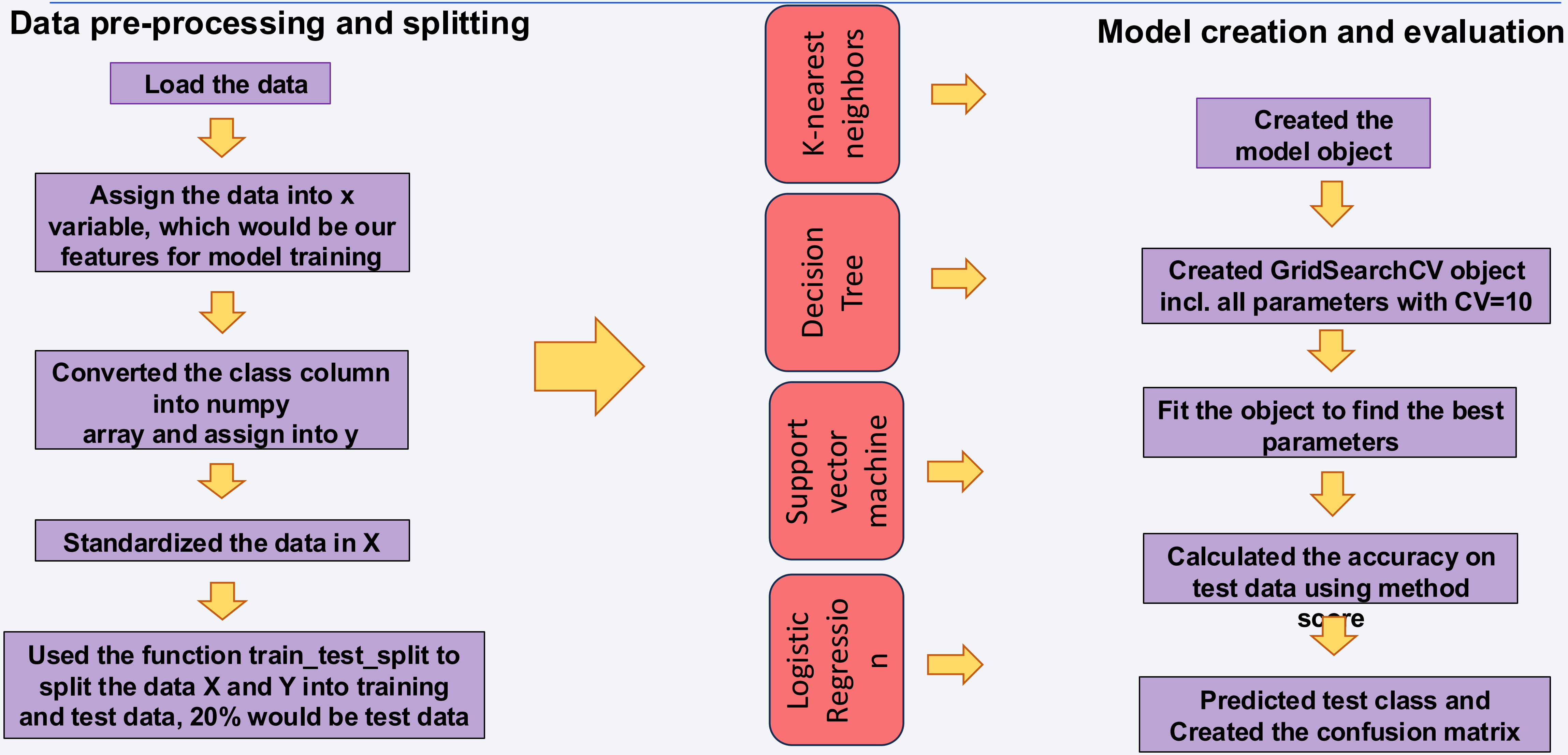
- A pie chart has been created to show the proportion of total success in all the sites.
- Separate pie chart has been created to show the number of failure and success of each site individually.
- The site selection is done using the drop-down menu.

Bottom component

- A scatter plot has been created where x-axis represents the payload mass and y-axis represents the mission outcome.
- A range slider is used to select the range of payload mass. Colors of scatter points are assigned using Booster version.
- The goal of the scatter plot is to see which Booster version with what ranger of payload mass leads to a successful outcome.

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/spacex-dash-app.py

Predictive Analysis (Classification)



Data pre-processing and splitting

Load the data

Assign the data into x variable, which would be our features for model training

Converted the class column into numpy array and assign into y

Standardized the data in X

Used the function train_test_split to split the data X and Y into training and test data, 20% would be test data

K-nearest neighbors

Decision Tree

Support vector machine

Logistic Regression

Model creation and evaluation

Created the model object

Created GridSearchCV object incl. all parameters with CV=10

Fit the object to find the best parameters

Calculated the accuracy on test data using method score

Predicted test class and Created the confusion matrix

Github link: https://github.com/RubySharma49/Space-X-Falcon-9-First-Stage-Landing-Prediction/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

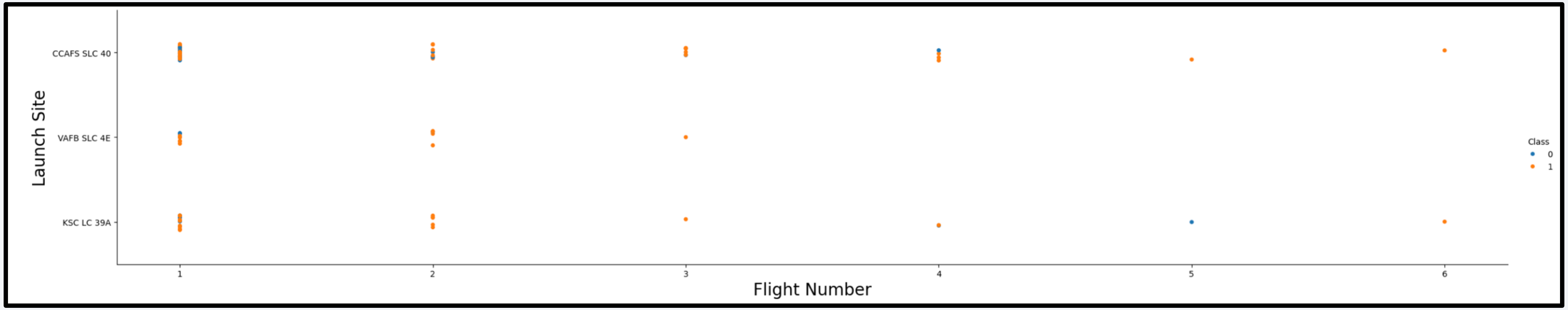
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

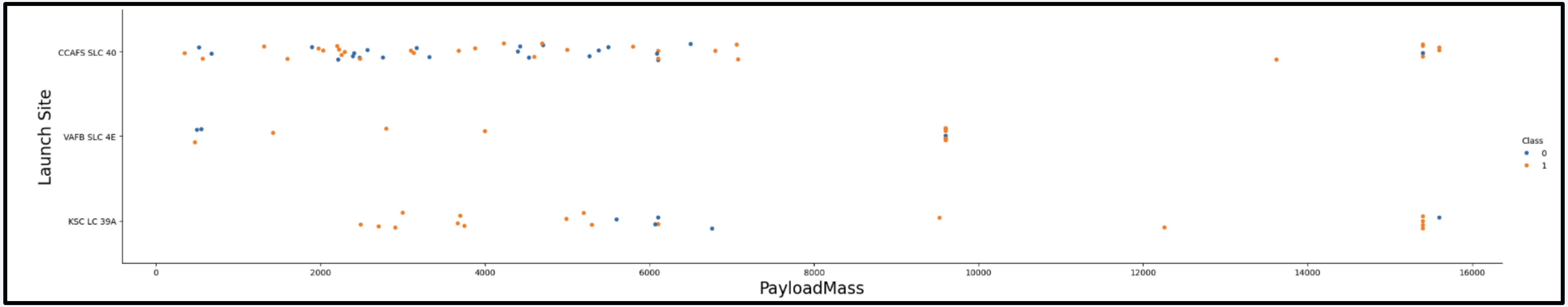
Flight Number vs. Launch Site

- Below is the scatter plot to show the relationship between flight number and launch site.
- The color points show the success or failure of the mission.
- We can see from the plot as the flight number increases, the proportion of mission success is also increasing.
- This trend is observed in all the sites.



Payload vs. Launch Site

- Below is the scatter plot to show the relationship between Payload mass and Launch site.
- The color points show the success or failure of the mission.
- We can see from the plot most of the launch happens with low payload mass < 8000 kgs. We see both success and failure of the mission.
- There are instances of spacecraft launching with high payload mass and successful mission.



Success Rate vs. Orbit Type

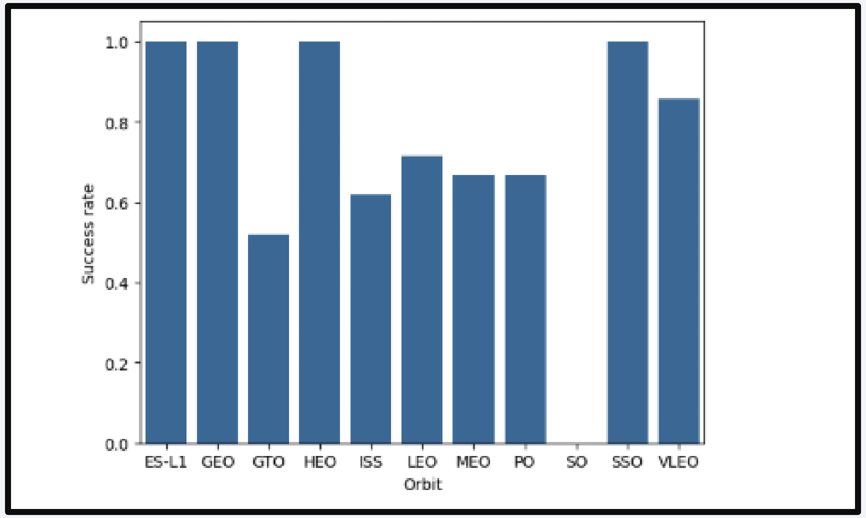
- On the top right, we have a bar plot to show the success rate of the mission based on different orbit types. Bottom we total count of launches and total count of successful launches.

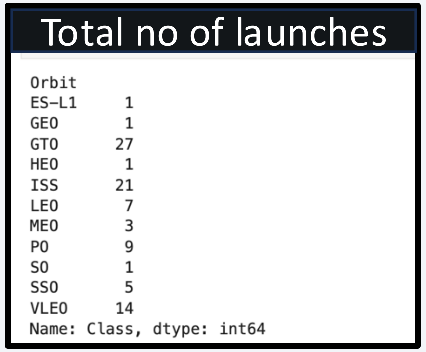
- From the plot, it looks like, ES-L1, GEO, HEO have 100% success rate. They all are high distant orbits, EL-L1 (1.5 million km from earth), GEO(~35,786 km from earth) and so on. One would think far distant orbit types have goo success rate. However, we can not say that with confidence as the bottom plot show the number of launches taken for these orbits are 1. Not enough data, to make this observation.

- SSO is the only orbit type (within 600-800 km) which has five launches and all are successful.

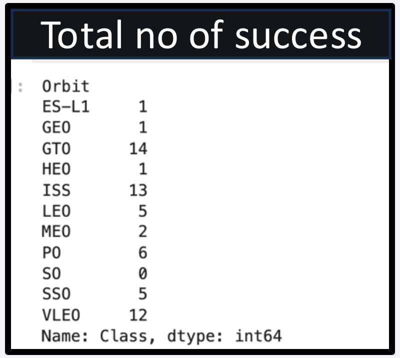
- VLEO also has a good success rate of ~80%, 12 out of 14 launches are successful.

- From the data, it seems GTO, ISS has most of the launches, their success rat is ~50% and ~62% respectively.




Total no of launches

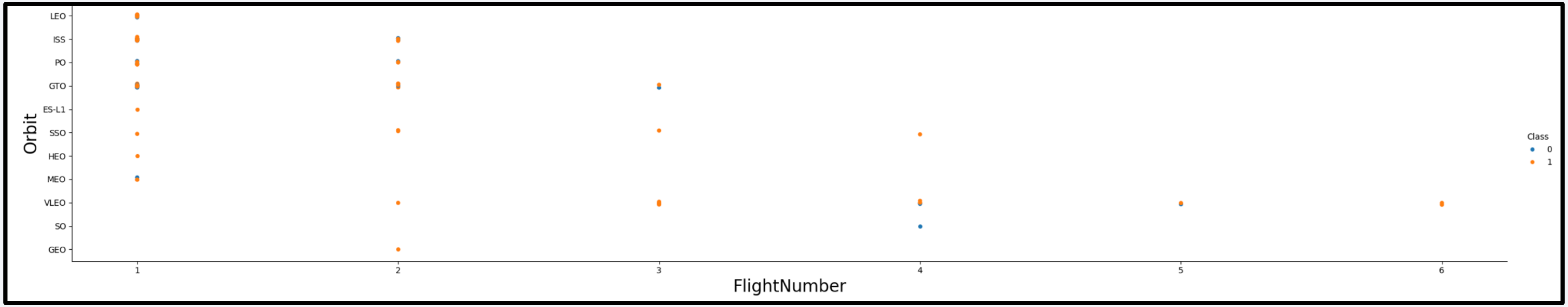
```
Orbit
ES-L1      1
GEO        1
GTO       27
HEO        1
ISS       21
LEO        7
MEO        3
PO         9
SO         1
SSO        5
VLEO      14
Name: Class, dtype: int64
```


Total no of success

```
Orbit
ES-L1      1
GEO        1
GTO       14
HEO        1
ISS       13
LEO        5
MEO        2
PO         6
SO         0
SSO        5
VLEO      12
Name: Class, dtype: int64
```

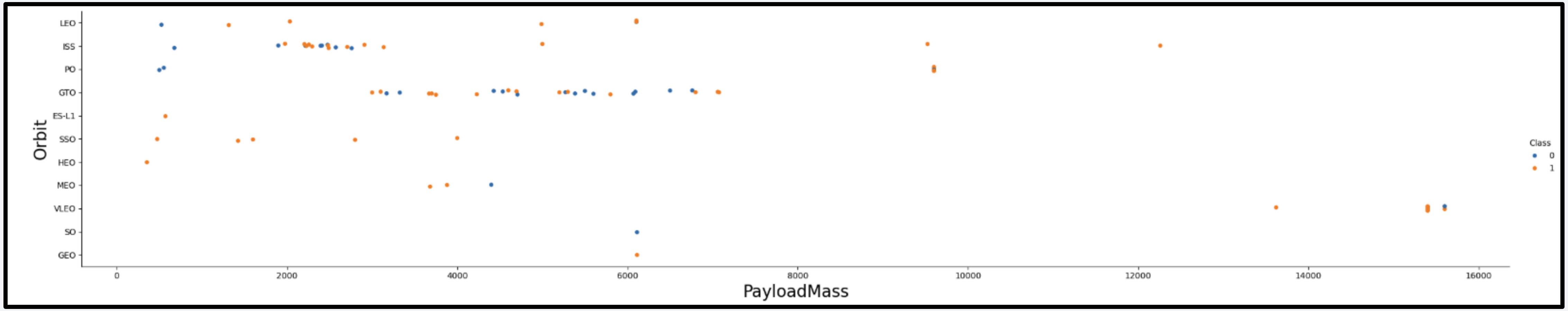
Flight Number vs. Orbit Type

- Below is the scatter plot to show the relationship between Payload mass and Launch site.
- The color points show the success or failure of the mission.
- We can see from the plot most of the number of flights increase the number of successful return also increases.



Payload vs. Orbit Type

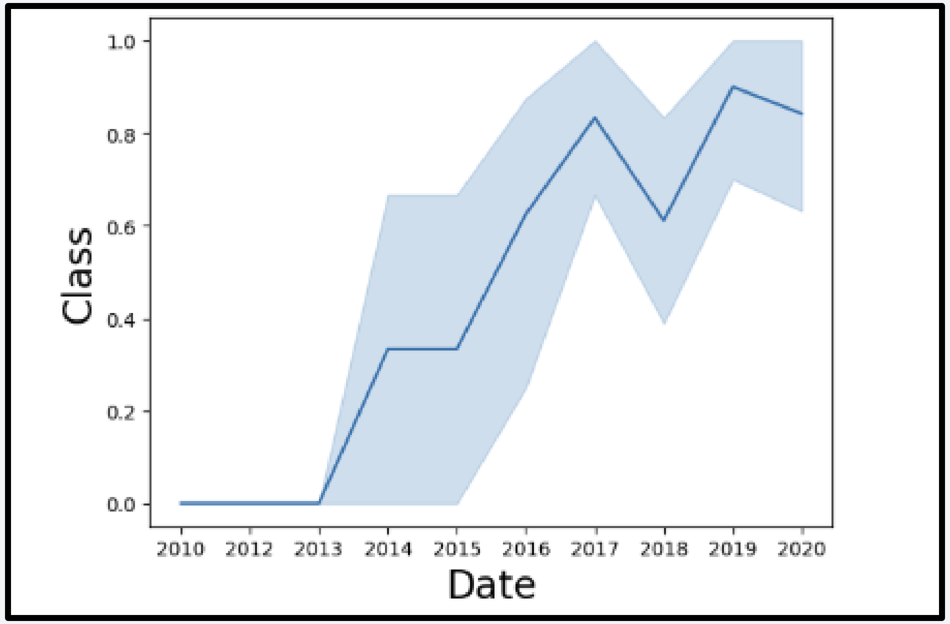
- Below is the scatter plot to show the relationship between Payload mass and Orbit Type.
- The color points show the success or failure of the mission.
- We can see from the plot SSO orbit which has all success launch has payload mass <4000 kgs, they are lighter weight.
- VLEO orbit type close to earth within 160km from earth, all has heavy payload mass > 13000 kgs and 80% of them are successful.
- PO polar orbit launches, have 7 launches > 8000 kgs and 6 of them are successful and 2 with low payload mass are unsuccessful.
- High distant orbit types has low payload mass.
- Other orbit types has <6000kgs payload mass.
- It is observed that heavy payload mass rockets are used for closer orbit types.



Launch Success Yearly Trend

- On right we have line plot to show the trend of successful launches from year 2010 to 2020.

- It can be seen clearly that with passing year the number of successful launches increasing.



All Launch Site Names

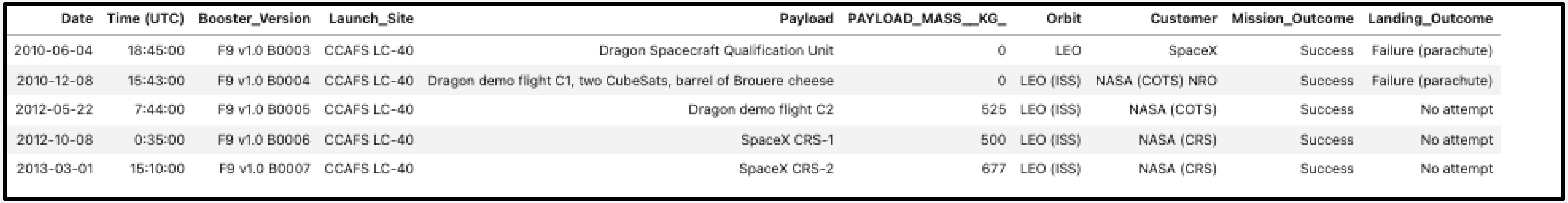
From our sql data exploration, we extracted the following unique rocket launch sites:

Unique site names are:

1) **CCAFS LC40**
2) **VAFB SLC-4E**
3) **KSC LC-39A**
4) **CCAFS SLC-40**

Launch Site Names Begin with 'CCA'

- Following are the launch details of Falcon 9 rockets having launch site begins with ‘CCA’.
- These are top five records and all of them has mission outcome “success”.



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

The total payload mass carried by the boosters launched by NASA is **45596 kgs**.

Average Payload Mass by F9 v1.1

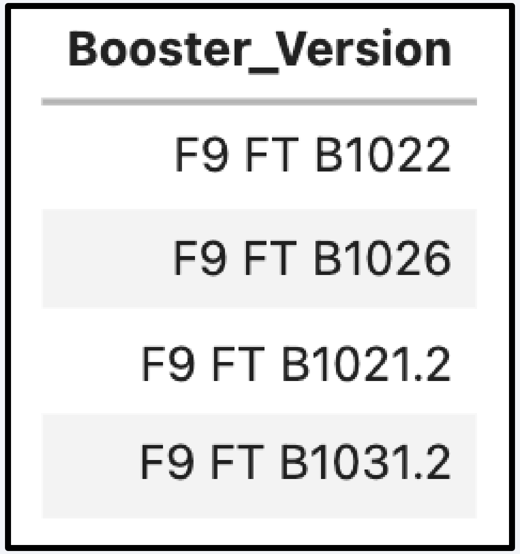
The average payload mass carried by booster version F9 v1.1 is **2928.4 kgs** which is quite less than NASA one.

First Successful Ground Landing Date

The first date, when spaceX landing outcome was successful was on **22nd December 2015**.

Successful Drone Ship Landing with Payload between 4000 and 6000

On right we have names of booster versions which has successful landing, having
Payload mass between 4000 and 6000 kgs.



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

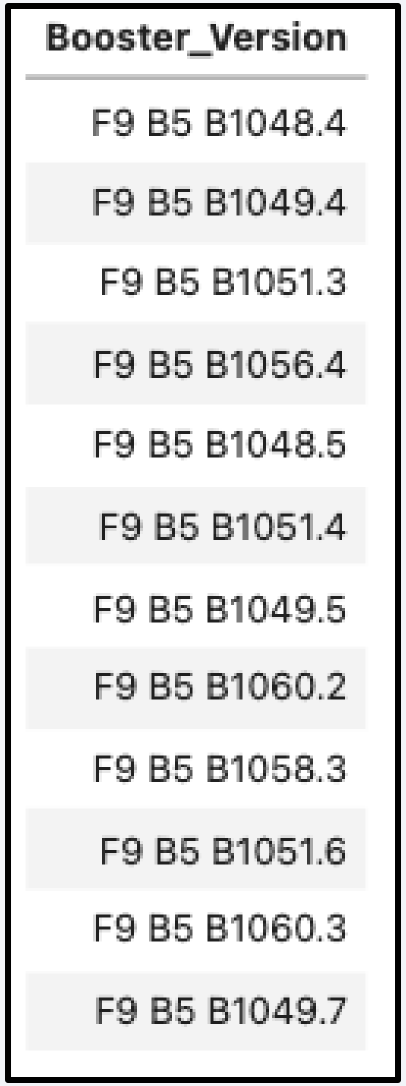
Total Number of Successful and Failure Mission Outcomes

Total number of **successful missions** outcomes are **700**.

Total number of **failure missions** outcomes are **7**.

Boosters Carried Maximum Payload

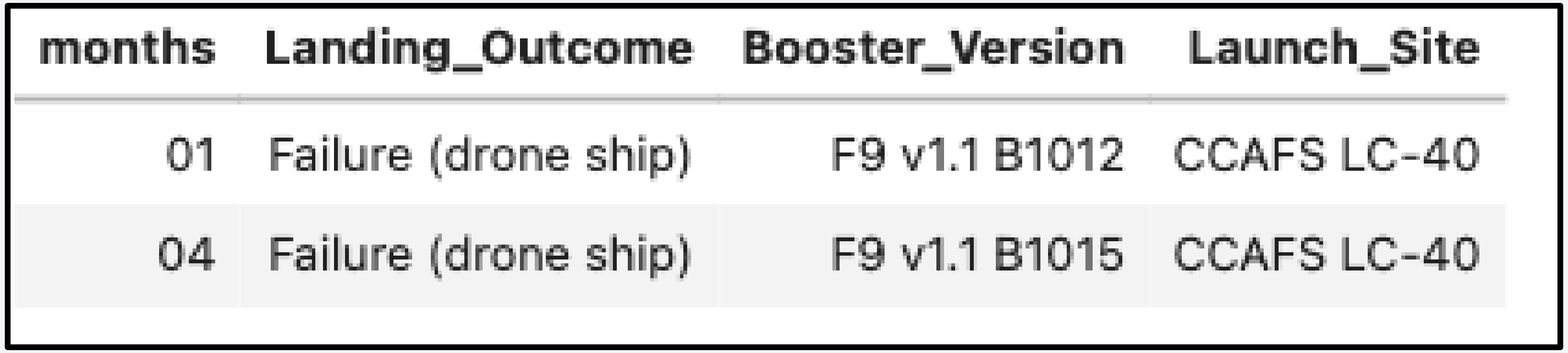
On right we have names of booster versions which carried maximum payload.



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

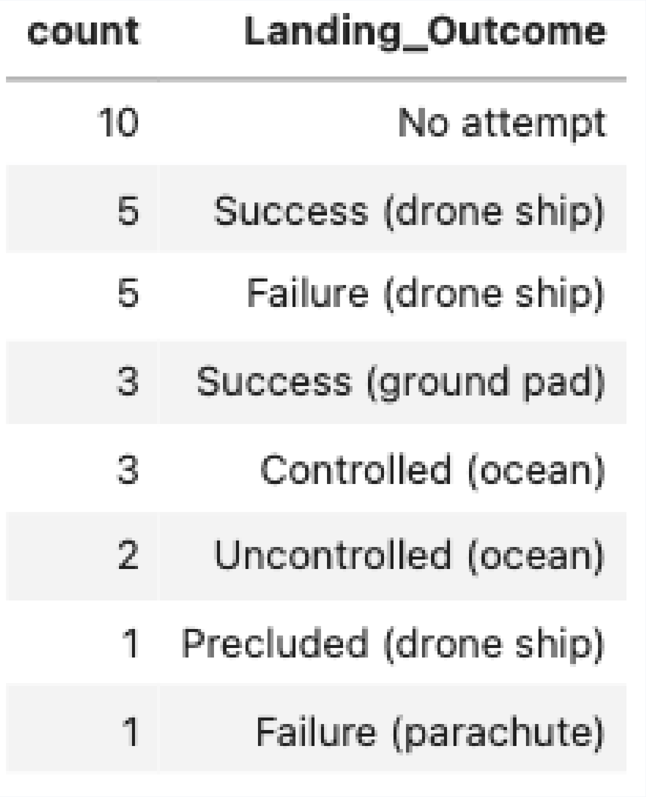
Following the are launch records of year 2015 for months January to April.



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

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

On right we have different landing outcomes with their occurrence count.



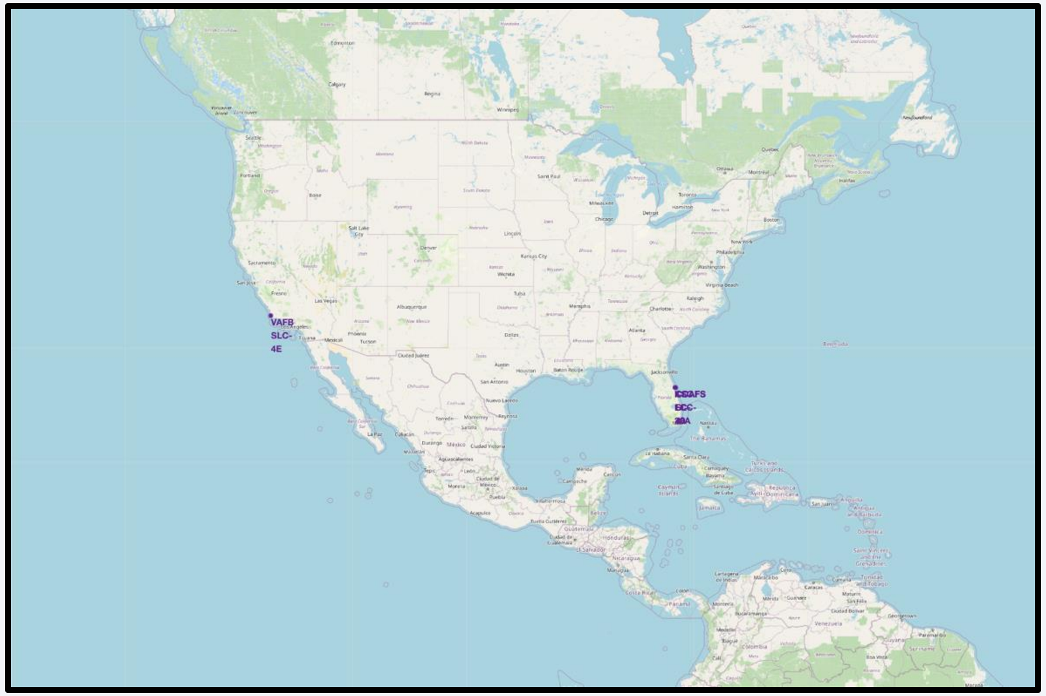
count	Landing_Outcome
10	No attempt
5	Success (drone ship)
5	Failure (drone ship)
3	Success (ground pad)
3	Controlled (ocean)
2	Uncontrolled (ocean)
1	Precluded (drone ship)
1	Failure (parachute)

Section 3

Launch Sites Proximities Analysis

Location of Launch sites

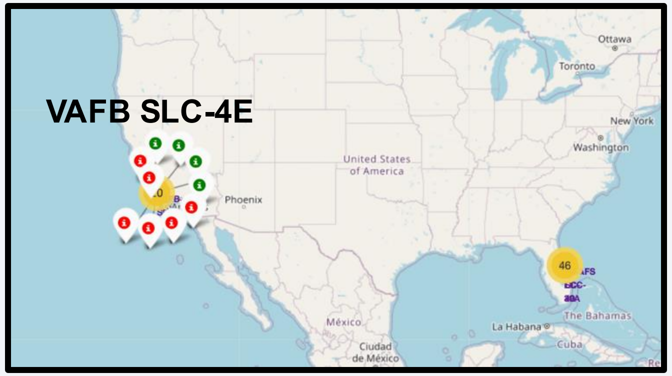
- On right we have a map of united states.
- The text in blue shows the location and name of different launch sites.
- All these launch sites are in close proximity to ocean.
- Three sites are situated close to each other at east coast of Florida.
- The fourth site is at the west coast of California.

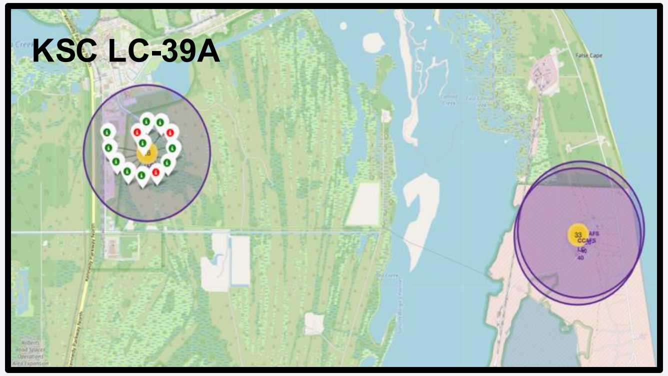


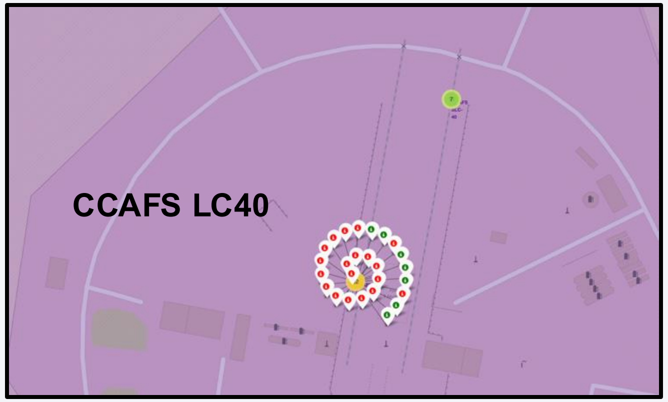
Successful and failed missions

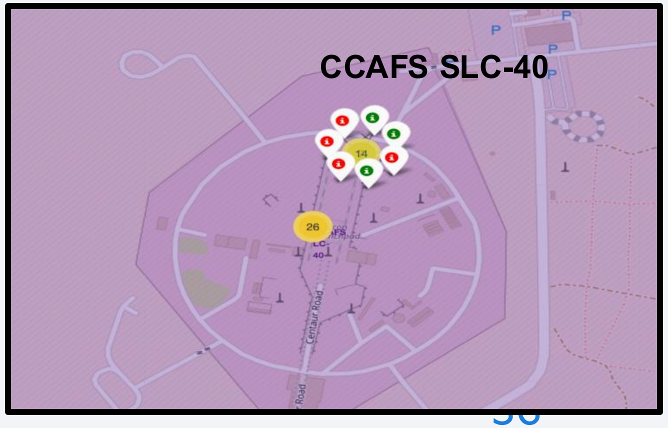
- The maps on right shows the red and green circle markers on top of each sites.
- Red marker shows failed missions.
- Green markers shows successful missions.
- KSC LC-39A has most successful missions.

Four launch sites marked by yellow circle



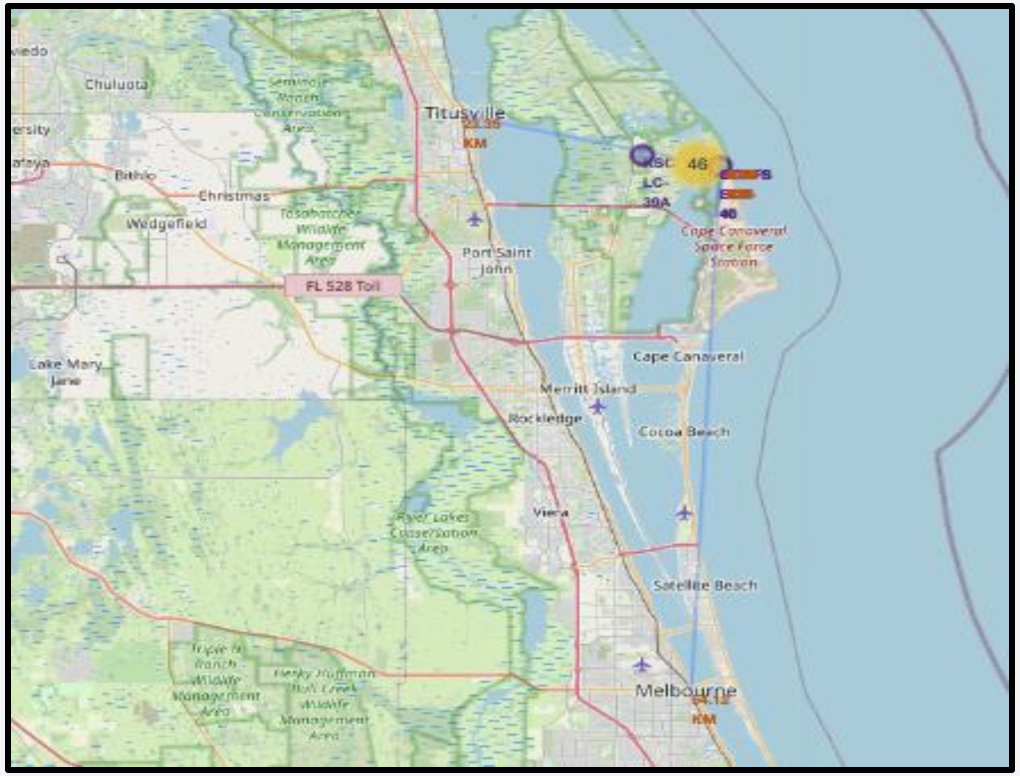




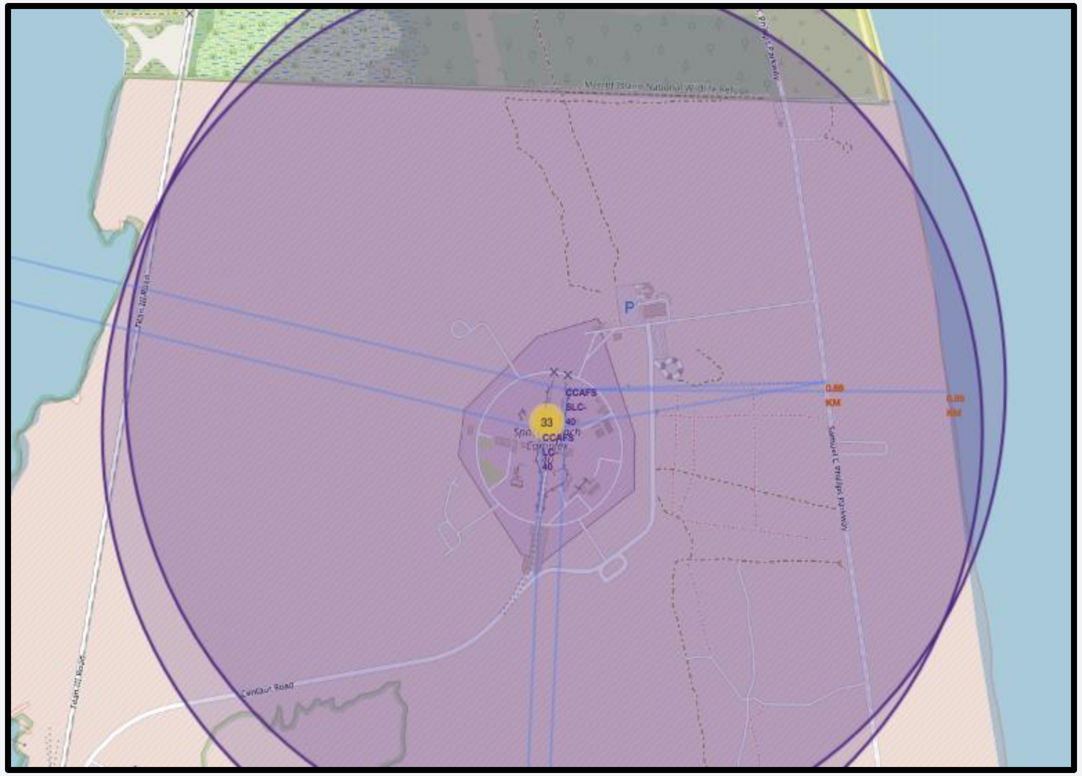


Places in close proximity to launch sites

CCAFS SLC-40 & CCAFS LC40 distance from Melbourne and Titusville cities



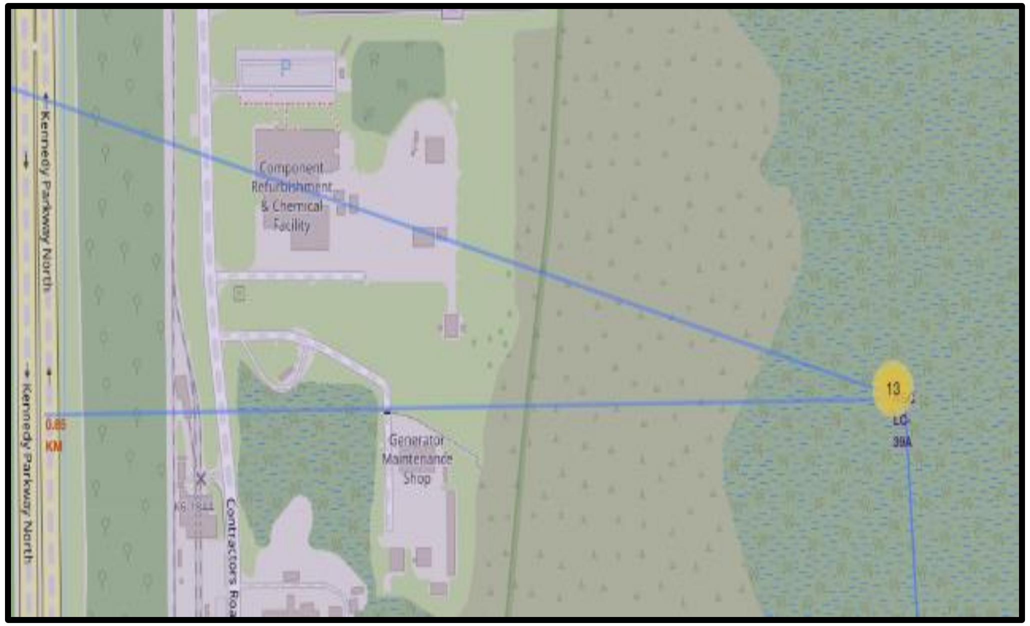
CCAFS SLC-40 and CCAFS LC40 distance from coastline and highway



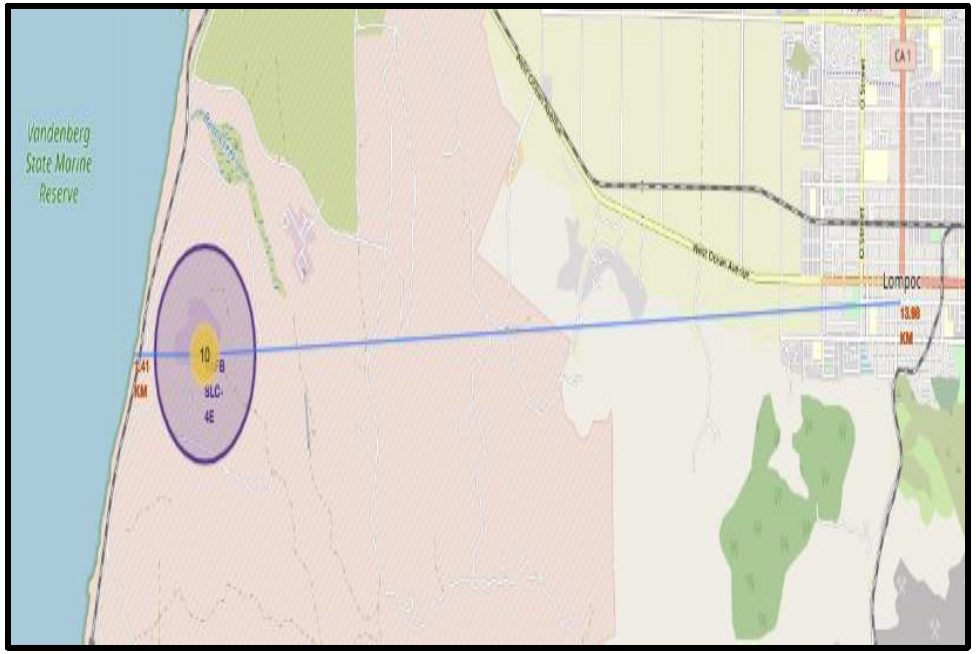
- On top right, we have the distance of coastline from CCAFS SLC-40 and CCAFS LC40 and top left from two cites.
- The calculated distance is labelled in red color.
- We can see that the coastline and highway is closer compared to cities.
- Closeness to ocean and distance from cities helps in the reduction of risk towards human lives during rocket launch.

Places in close proximity to launch sites

KSC LC-39A distance from Kennedy highway



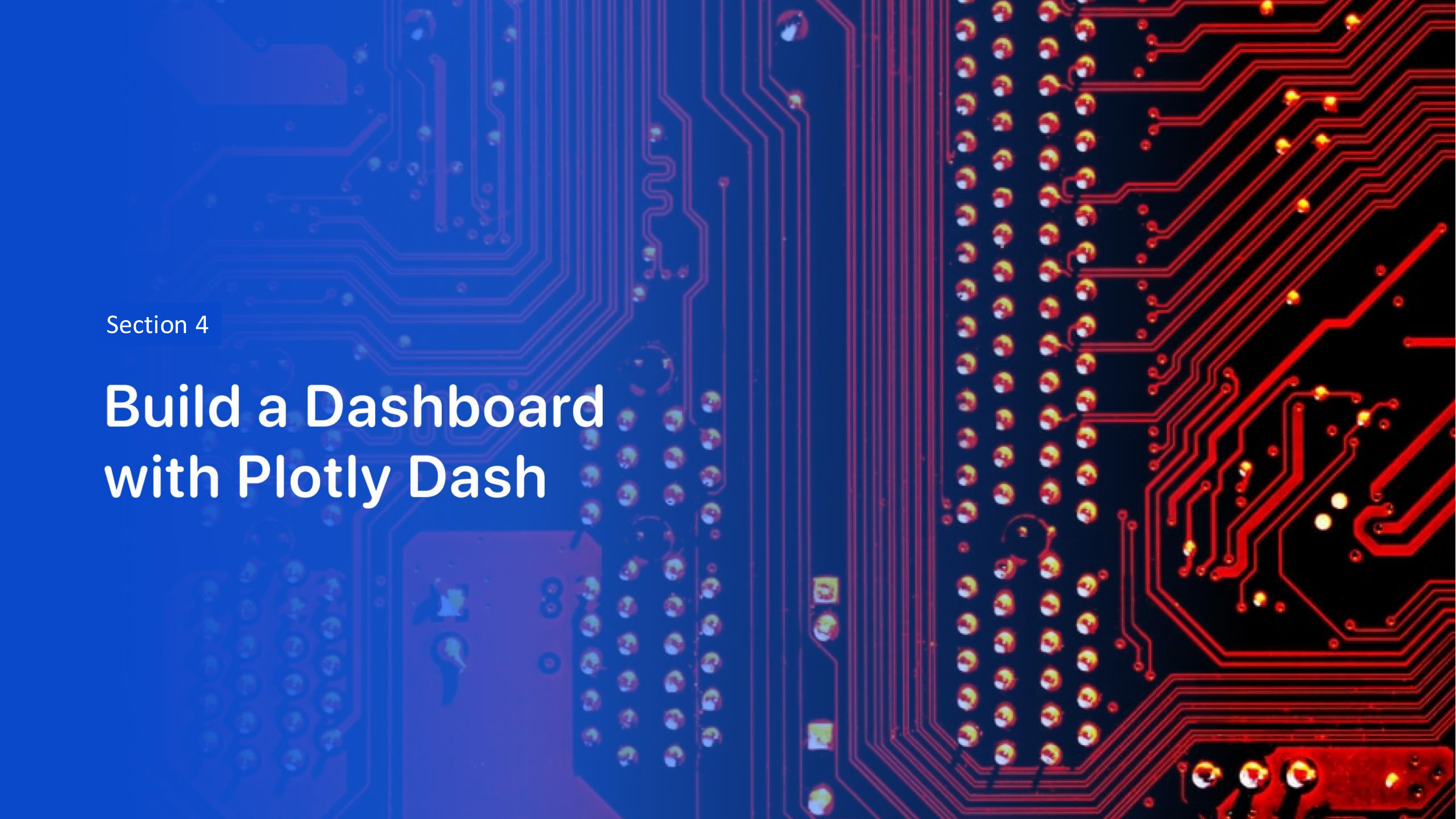
VAFB SLC-4E distance from coastline and highway



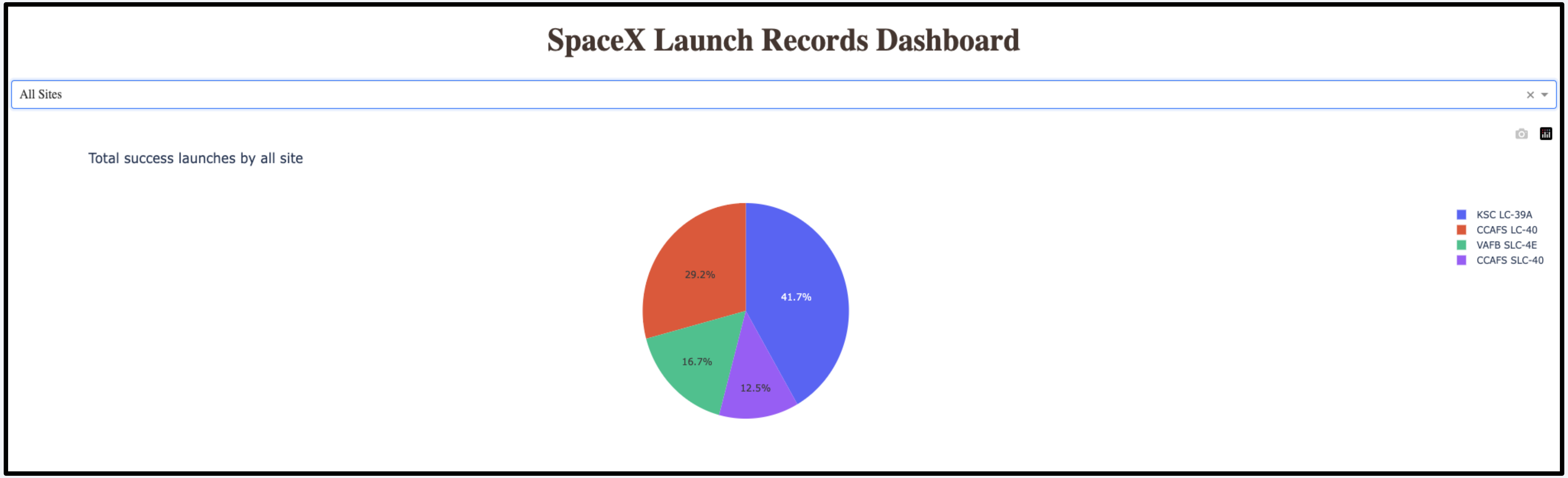
- We made the same observation from other sites as well that the coastline and highway are closer compared to cities.
- Highways help in the smooth transportation of equipment to the launch sites.
- Lauch of rocket close to open water far from cities is safe and risk-free towards human population.

Section 4

Build a Dashboard with Plotly Dash



SpaceX launch records for all sites



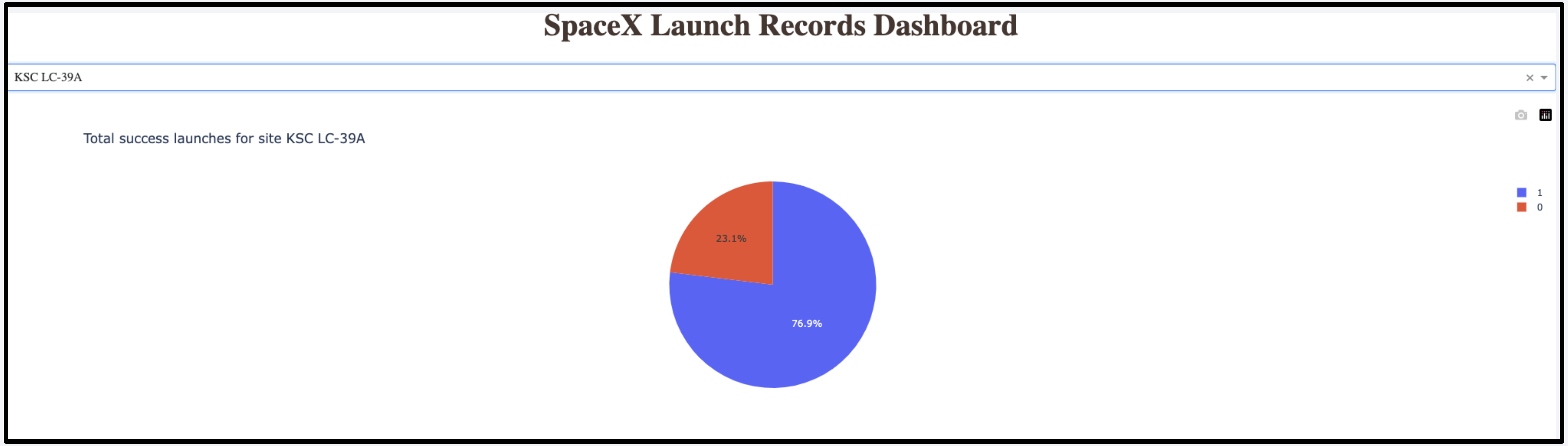
SpaceX Launch Records Dashboard

All Sites

Total success launches by all site

- The above dashboard shows the drop-down menu to select the site and pie chart show the successful launches.
- Above screenshot shows the selection of ‘all-sites’ with pie chart which shows the successful launches of all sites.
- KSC LC-39A has highest proportion of successful launches followed by CCAFS LC-40, then VAFB SLC-4E and lowest is CCAFS SLC-40.

Most successful launch record of site



SpaceX Launch Records Dashboard

KSC LC-39A

Total success launches for site KSC LC-39A

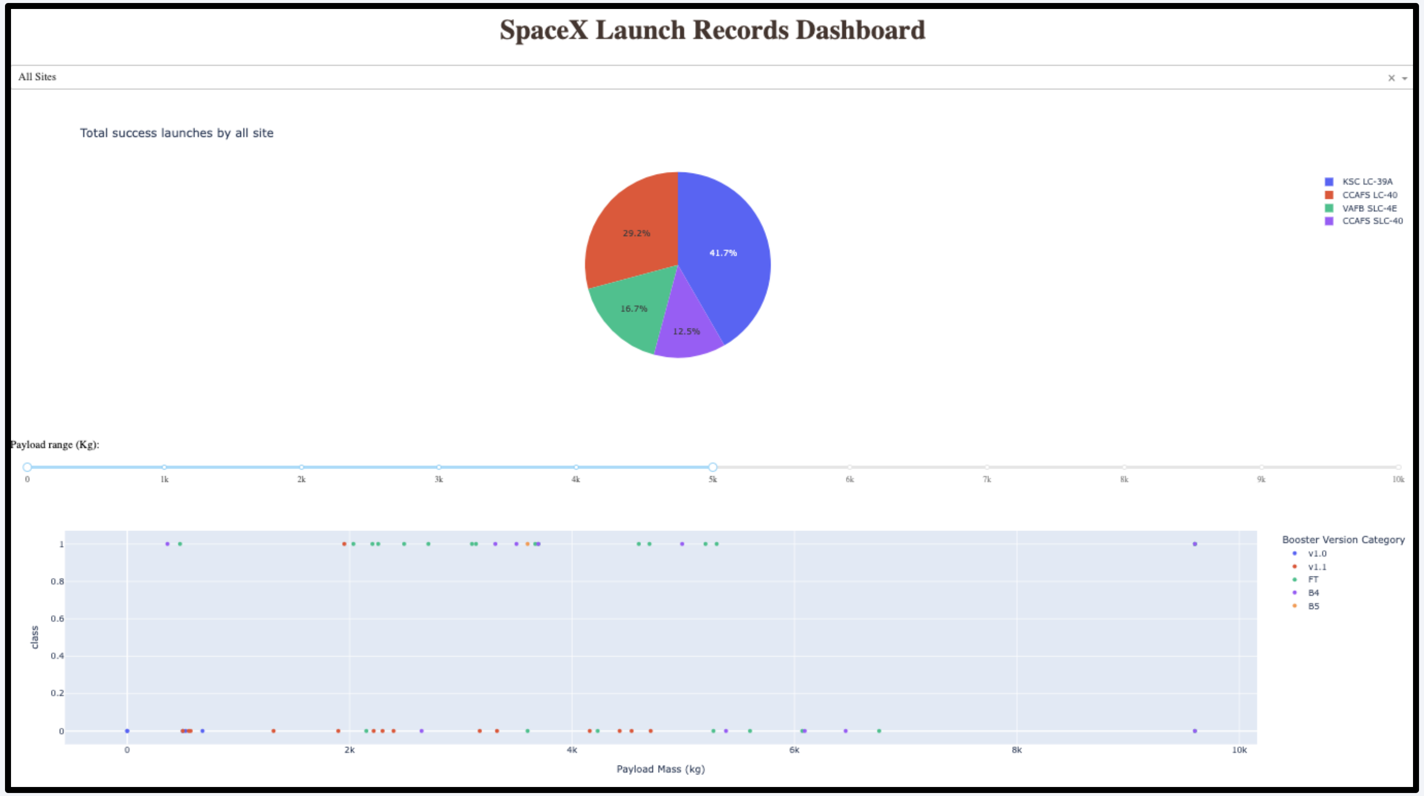
- KSC LC 39-A has the most successful launches among all sites.
- The pie-chart shows the ration of successful and failed launches.

Payload mass vs launch outcome

- On right, we have a screenshot of scatterplot which shows the relationship of payload mass and launch outcome for all sites.

- The points are colored by the booster version.

- We can see that FT booster has the most successful launches followed by B4 booster.



SpaceX Launch Records Dashboard

All Sites

Payload range (Kg):

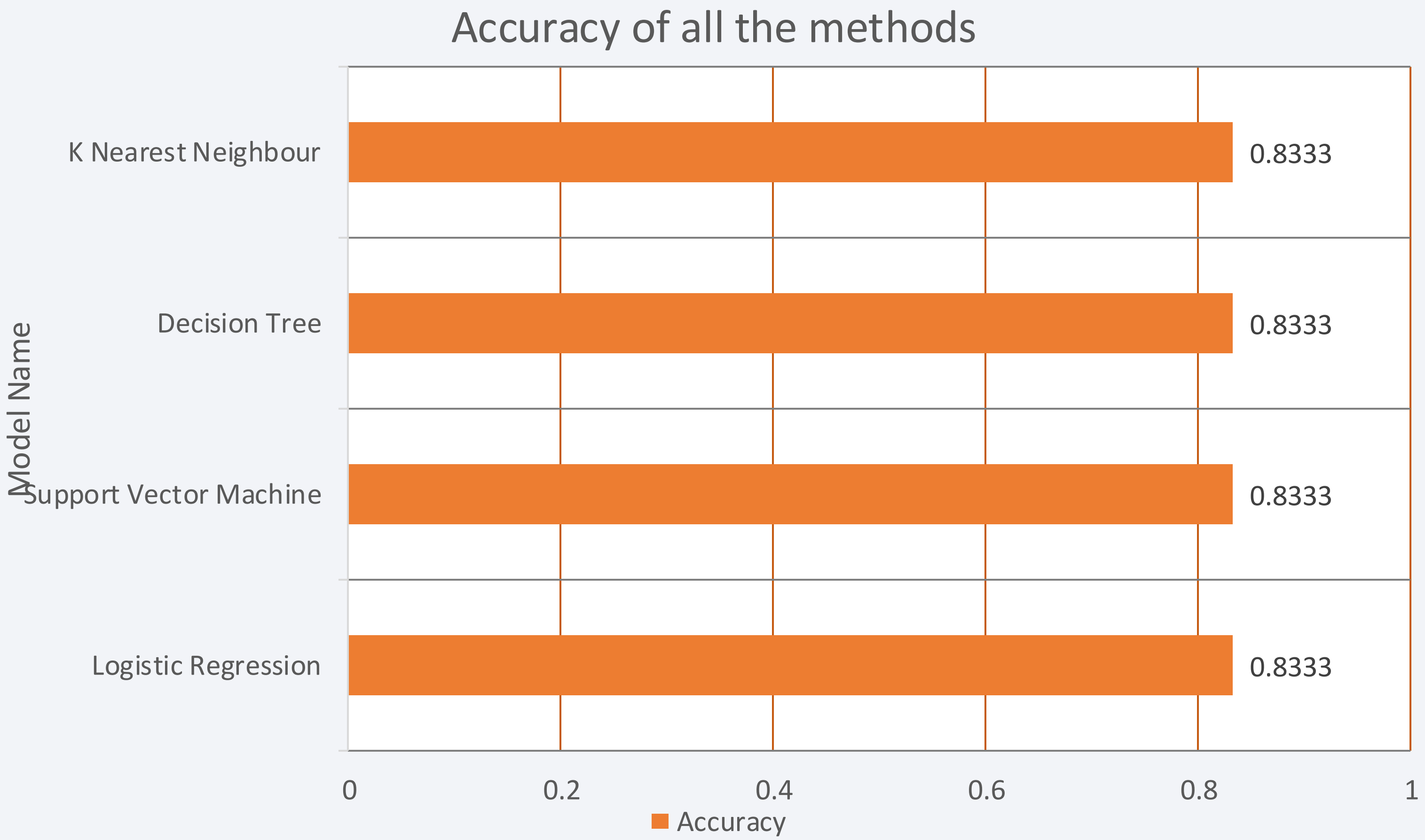
Section 5

Predictive Analysis (Classification)

Classification Accuracy

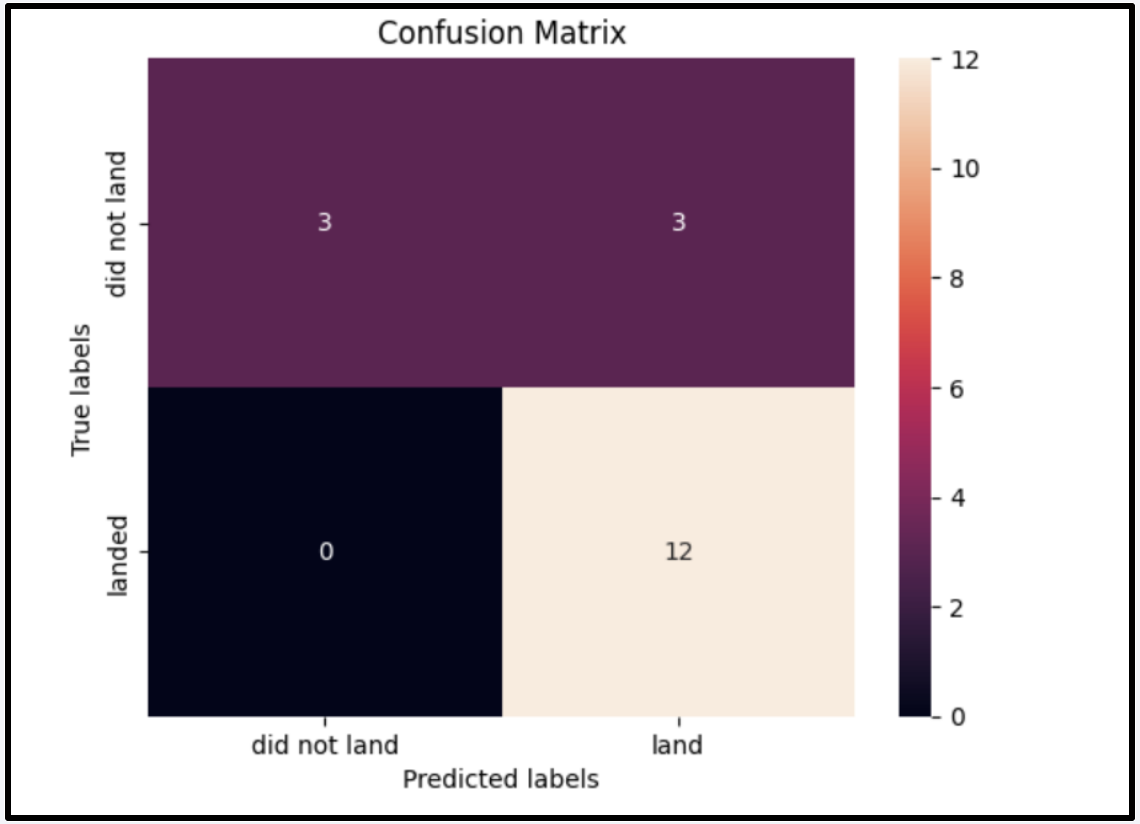
- On right we have classification accuracy of four models on test data.

- We can see all four models performed equally with accuracy of 83.33 %.



Confusion Matrix

- On right we have confusion matrix which shows how well the model can differentiate the classes.
- All four models had same confusion matrix.
- We can see True positives are 12 and they are correctly predicted by the models.
- However, the false positives are 6 but the model incorrectly classifies three of them false negatives.



Conclusions

1. In total, we have four unique launch sites for the spaceX.
2. We observed these sites are closer to coastlines, highways and far from cities or populated areas.
3. Highways help in easy transport; coastline helps in safe launches.
4. We observed that rockets having heavy payload mass can also have safe landing of the first stage.
5. Rockets launched to SSO and VLEO orbits has the most successful landing proportions. These orbits are closer to earth.
6. Rockets launched to VELO carries high payload mass, still has successful landings.
7. We observed that the increase in the number of attempts of flights increases the chances of safe landings.
8. With passing year from 2010 to 2020, the proportion of safe landing has increased a lot.
9. The average payload mass carried by booster version F9 v1.1 is 2928.4 kgs.
10. Booster FT and B4 has more successful record compared to other booster versions.
11. We employed four models, Logistic Regression, Support vector machine, Decision Tree and K Nearest Neighbor.
12. All four models has an accuracy of 83.33 % and they all performed equally well.

Thank you!