IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

SNEHA MONDAL
17/10/2021

Outline

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

Executive Summary

This is all about -

- ❑ Data collection
- ❑ Data wrangling
- ❑ Exploratory Data Analysis(EDA) using SQL as well as visualization libraries
- ❑ Building an interactive dashboard
- ❑ Predictive Analysis(Classification)

All the above mentioned methodologies will collectively determine if the first stage of Falcon 9 will land successfully and along the way recognize the factors or attributes impacting the landing of the rocket successfully in one piece.

Introduction

SpaceX 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.

Will the first stage of Falcon 9 land successfully? What are some of the factors contributing to a successful landing of the rocket?

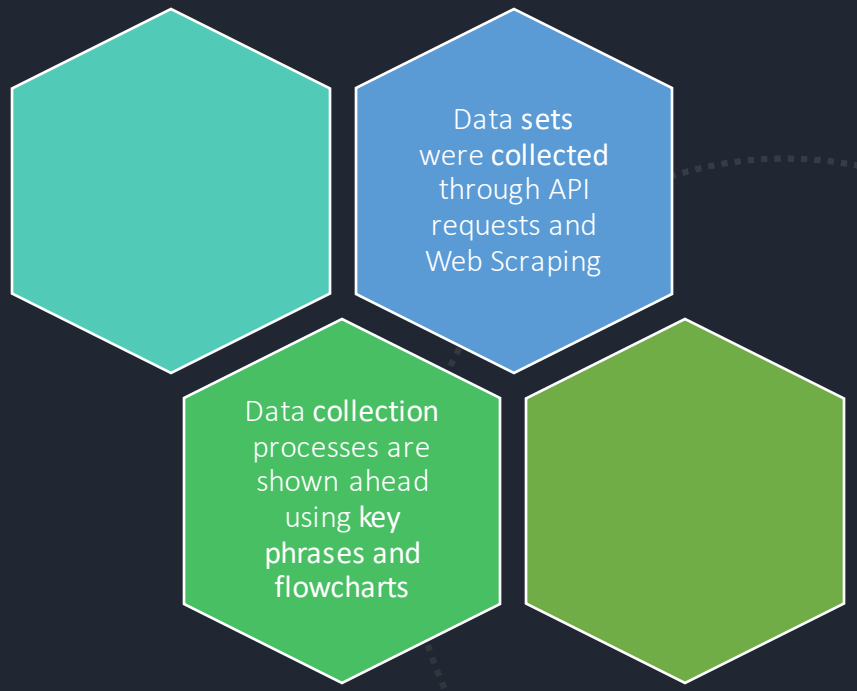
Section 1

Methodology

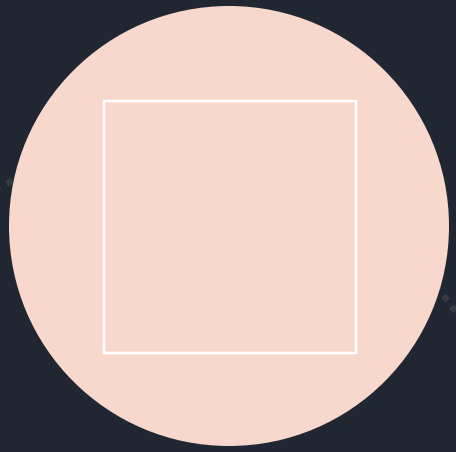
Methodology

- Executive Summary
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection



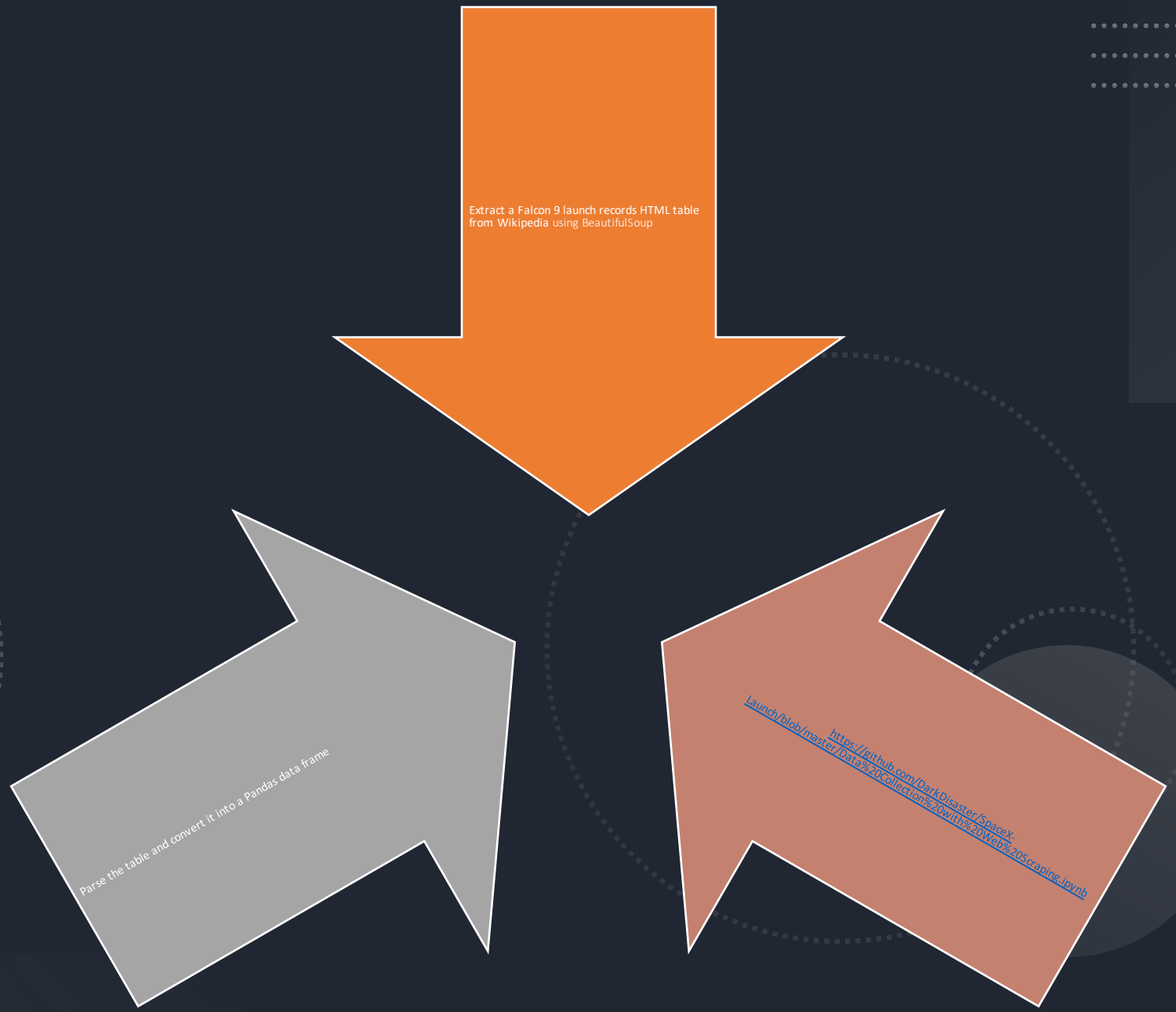
Data Collection – SpaceX API



GET REQUEST TO THE SPACEX API AND CLEAN THE REQUESTED DATA. DECODE THE RESPONSE CONTENT AS JSON AND THEN TURN IT INTO A DATA FRAME FOR FURTHER ANALYSIS.

HTTPS://GITHUB.COM/DARKDISASTER/SPACEX-LAUNCH/BLOB/MASTER/DATA%20COLLECTION%20API.IPYNB

Data Collection - Scraping



Data Wrangling

Description	We perform **Exploratory Data Analysis (EDA)** to find some patterns in the data and determine what would be the label for training supervised models
Idea	In the data set, there are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident; we will mainly convert those outcomes into Training Labels with 1 : the booster successfully landed, `0` : landing was **unsuccessful.**
GitHub URL	https://github.com/DarkDisaster/SpaceX-Launch/blob/master/EDA.ipynb

EDA with Data Visualization

Scatter plots – identify relationships between various features in the data set

Bar charts – identify success rates of particular features

Line charts – visualize average launch success trends yearly

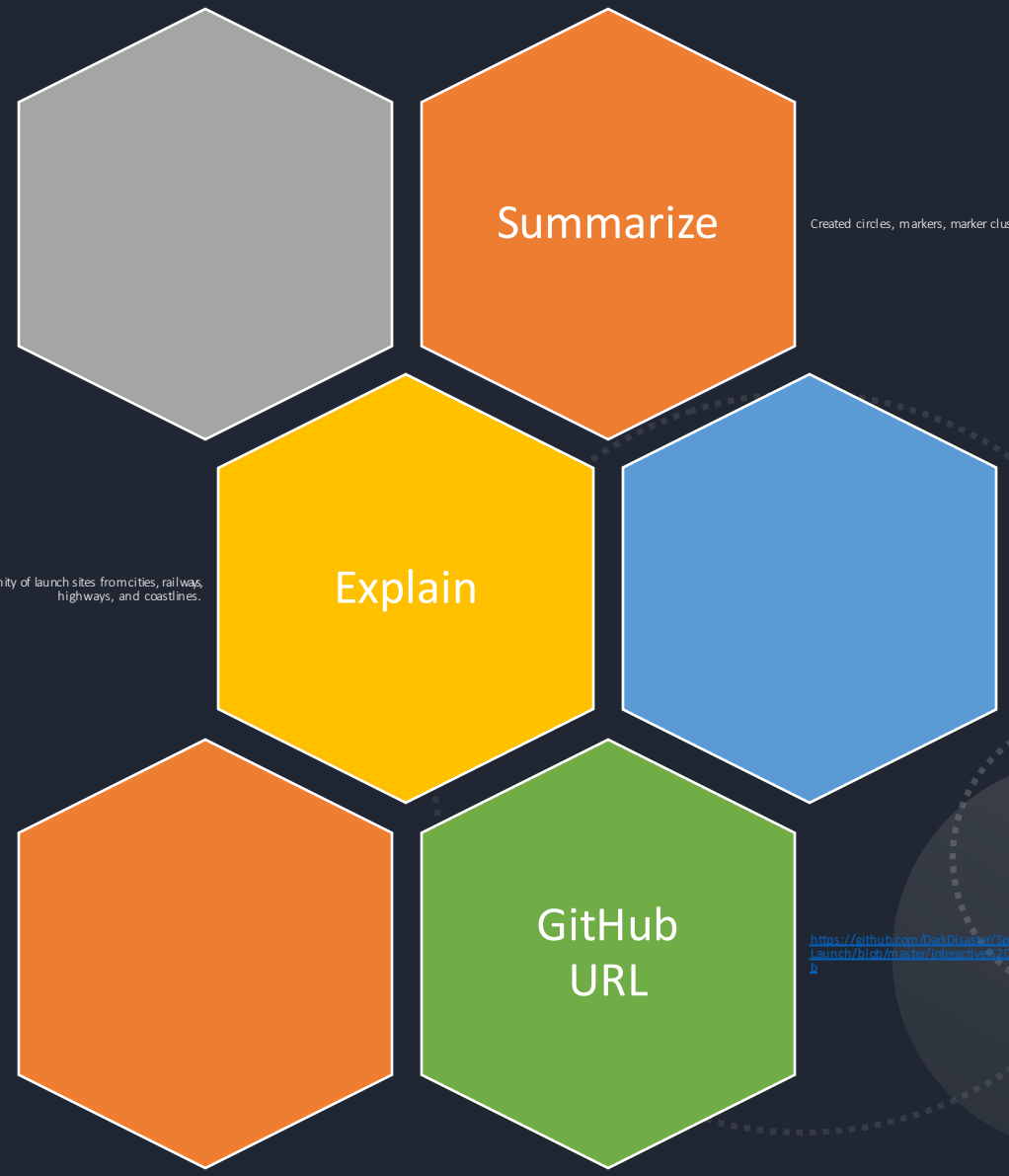
https://github.com/DarkDisaster/SpaceX-Launch/blob/master/EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- *Display the names of the unique launch sites in the space mission*
- *Display the total payload mass carried by boosters launched by NASA (CRS)*
- *Display average payload mass carried by booster version F9 v1.1*
- *List the date when the first successful landing outcome in ground pad was acheived*
- *List the total number of successful and failure mission outcomes*
- *List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015*
- *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*

https://github.com/DarkDisaster/SpaceX-Launch/blob/master/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium



Created circles, markers, marker clusters, mouse position, PolyLine

The objects were created to understand proximity of launch sites from cities, railways, highways, and coastlines.

https://github.com/ParkOjasa-r/SpaceX-Launch/blob/master/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

Summarize

- Dropdown list, Slider, Pie chart and Scatter plot has been made use of.

Explain

- Dropdown list : Launch Site selection
- Pie chart : Success vs. Failed launches
- Slider : Select payload range
- Scatter plot : Correlation between payload and launch success

GitHub URL

- https://github.com/DarkDisaster/SpaceX-Launch/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

Perform exploratory Data Analysis and determine Training Labels

Create a column for the class

Standardize the data

Split into training data and test data

Find best Hyperparameter for SVM, Classification Trees and Logistic Regression using Grid Search

Find the method performs best using test data

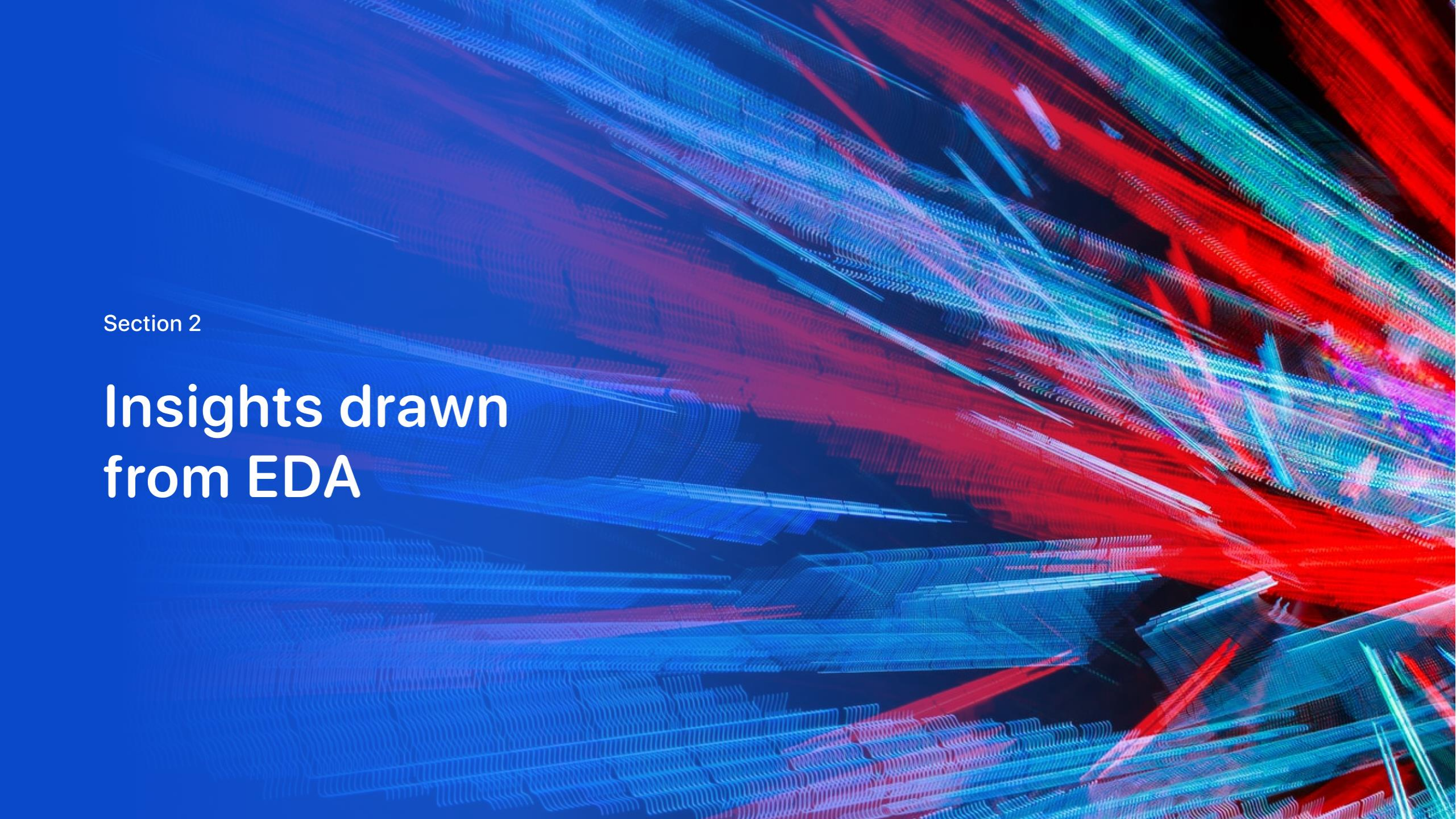
https://github.com/DarkDisaster/SpaceX-Launch/blob/master/Machine%20Learning%20Prediction.ipynb

Results

Exploratory data analysis results

Interactive analytics demo in screenshots

Predictive analysis results

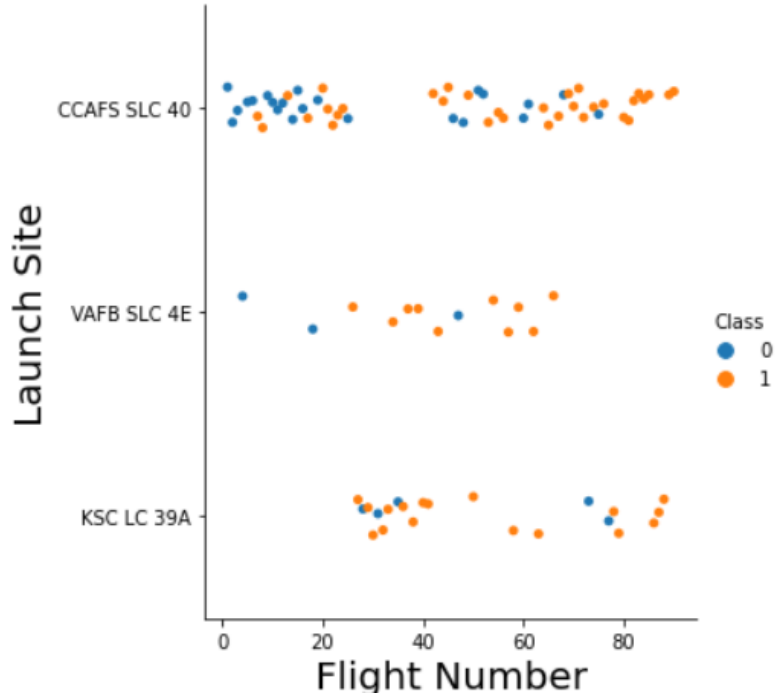


Section 2

Insights drawn from EDA

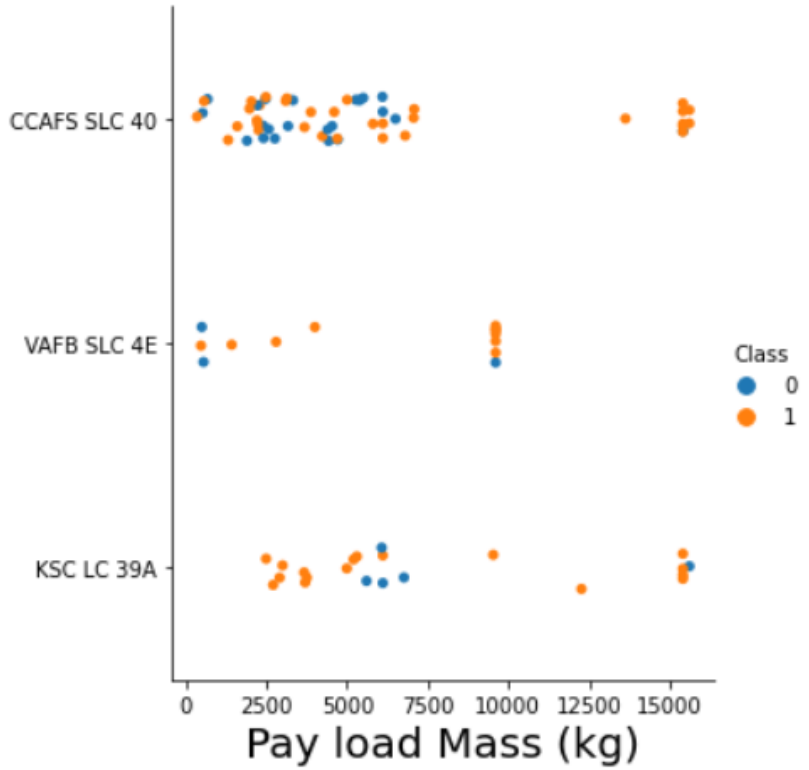
Flight Number vs. Launch Site

- VAFB SLC 4E have higher successful landings for flight numbers > 20
- CCAFS SLC 40 mostly have failed landings irrespective of flight numbers
- KSC LC 39A can't be concluded upon based on just flight numbers



Payload vs. Launch Site

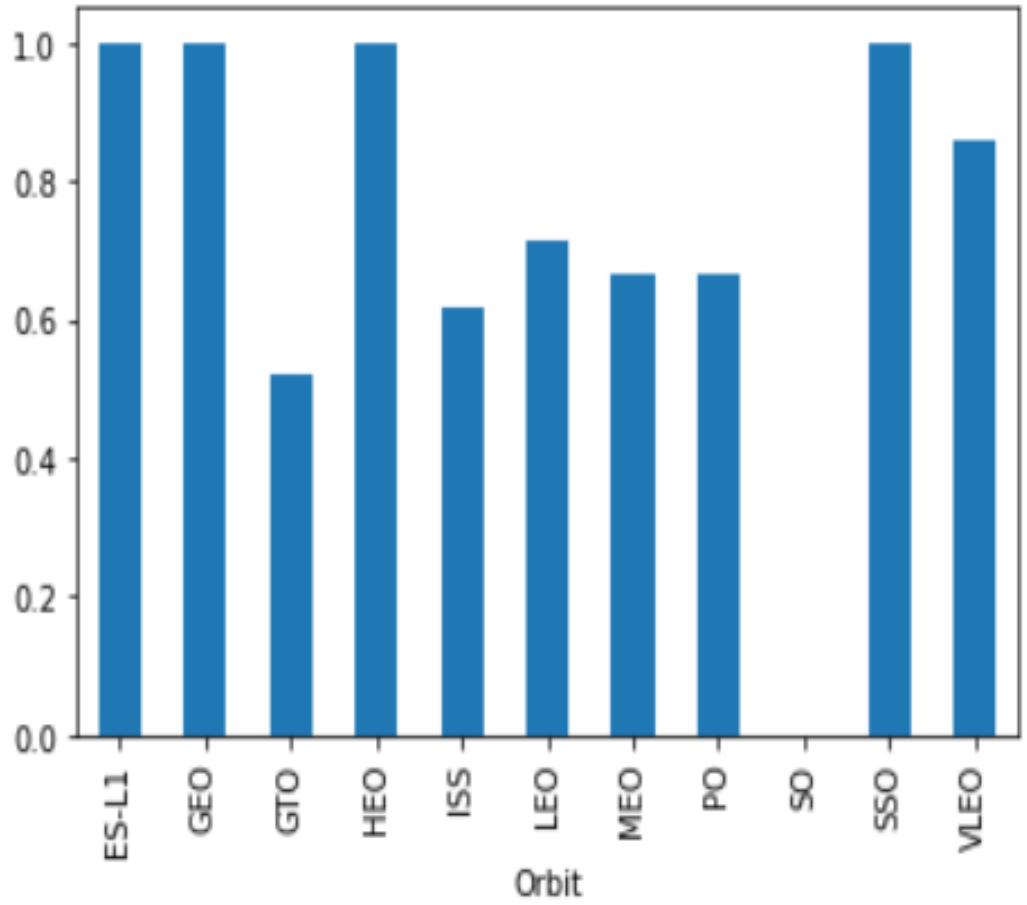
- For the VAFB SLC 4E Launch site there are no rockets launched for heavy payload mass(greater than 10000).
- CCAFS SLC 40 gives successful launches when payload mass > 12500



Success Rate vs. Orbit Type

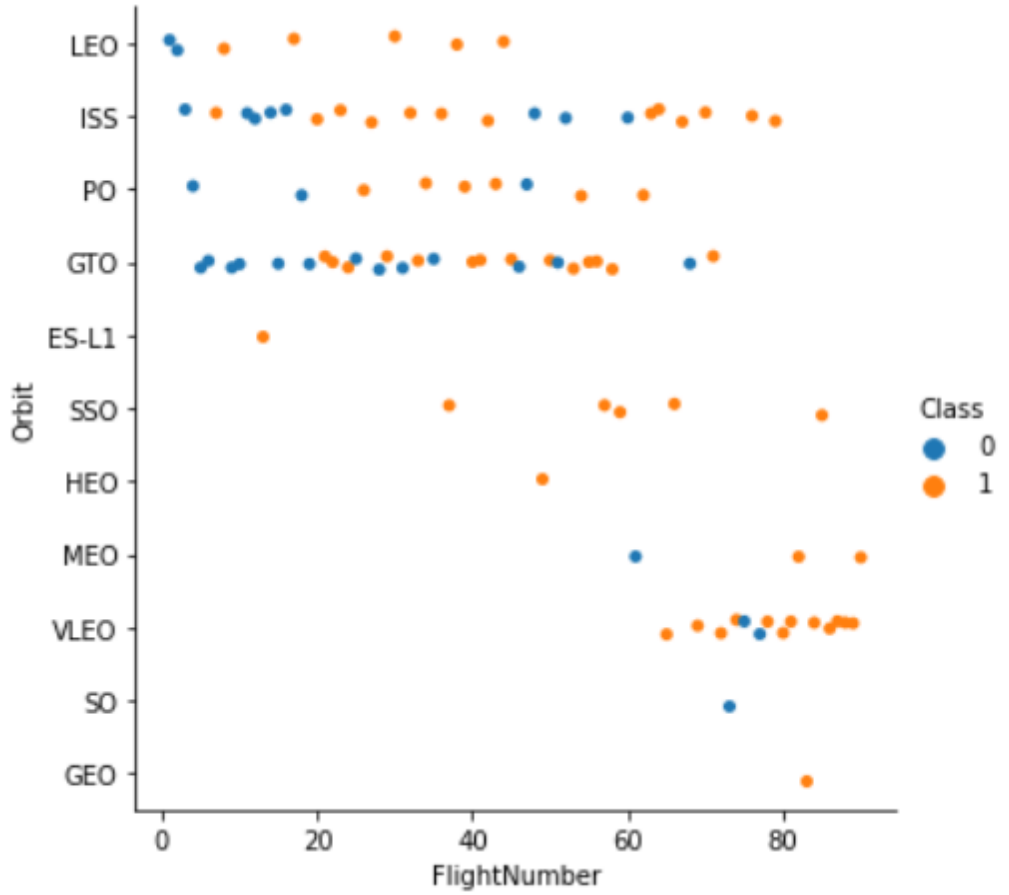
The following orbits have higher success rates :

- ES-L1
- GEO
- HEO
- SSO
- VLEO



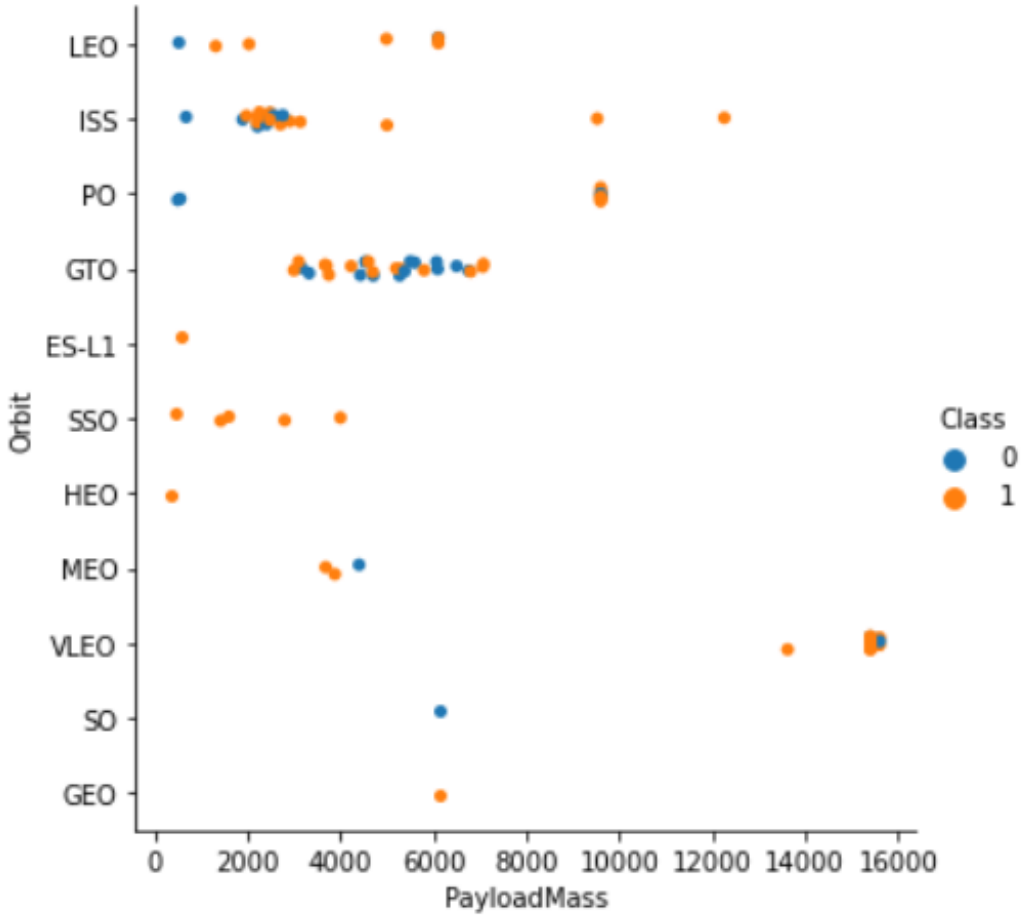
Flight Number vs. Orbit Type

- In the LEO orbit the Success appears related to the number of flights.
- There seems to be no relationship between flight number when in GTO orbit.



Payload vs. Orbit Type

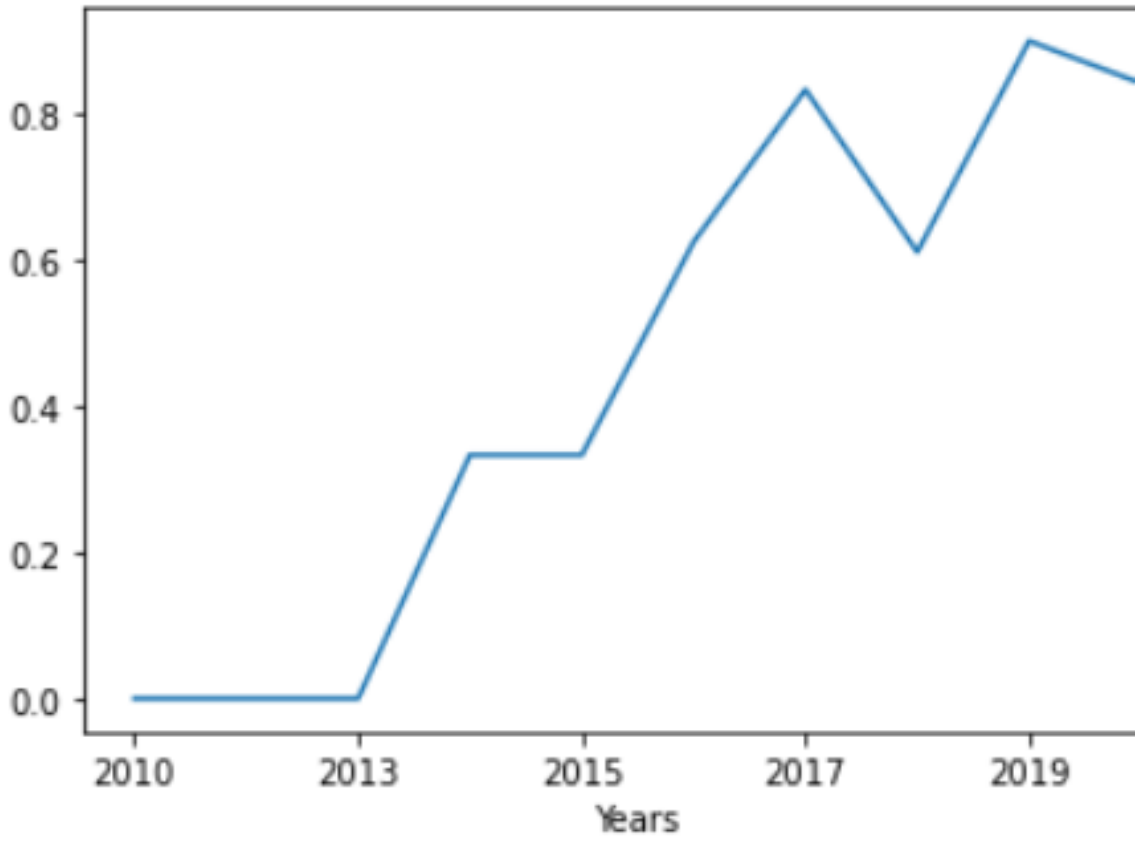
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.



Launch Success Yearly Trend

- Shows a line chart of yearly average success rate

- The sucess rate since 2013 kept increasing till 2020



All Launch Site Names

- Find the names of the unique launch sites
- `%sql SELECT DISTINCT(LAUNCH_SITE) FROM SPACEXDATASET`

CCAFS LC-40
CCAFS SLC-40
KSC LC-39A

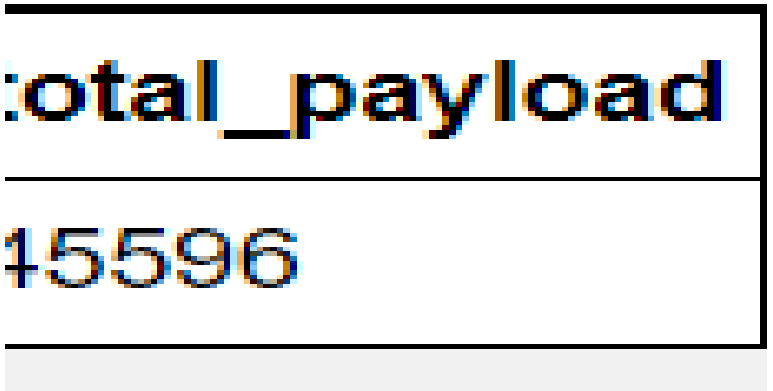
DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- `%sql SELECT * FROM SPACEXDATASET WHERE LAUNCH_SITE LIKE 'CCA%' limit 5`

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
-
```
%sql SELECT sum(payload_mass__kg_) as total_payload FROM SPACEXDATASET WHERE customer = 'NASA (CRS)'
```



Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
-
```
%sql SELECT avg(payload_mass__kg_) as average_payload FROM SPACEXDATASET WHERE booster_version = 'F9 v1.1'
```

Done.

average_payload
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- %sql SELECT MIN(DATE) AS First_date FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (ground pad)'

first_date
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
- %%sql
- SELECT BOOSTER_VERSION FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000

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

9]:

mission_outcome	total_count
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- `%sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) AS Total_count FROM SPACEXDATASET GROUP BY MISSION_OUTCOME`

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- %%sql
- SELECT BOOSTER_VERSION FROM SPACEXDATASET
- WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXDATASET)

[0]:

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

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015.
- %%sql
- SELECT BOOSTER_VERSION, LANDING__OUTCOME, LAUNCH_SITE, DATE FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Failure (drone ship)' **and** DATE LIKE '2015%'

booster_version	landing__outcome	launch_site	DATE
F9 v1.1 B1012	Failure (drone ship)	CCAFS LC-40	2015-01-10
F9 v1.1 B1015	Failure (drone ship)	CCAFS LC-40	2015-04-14

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(LANDING__OUTCOME) AS Count FROM SPACEXDATASET WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__OUTCOME ORDER BY Count DESC

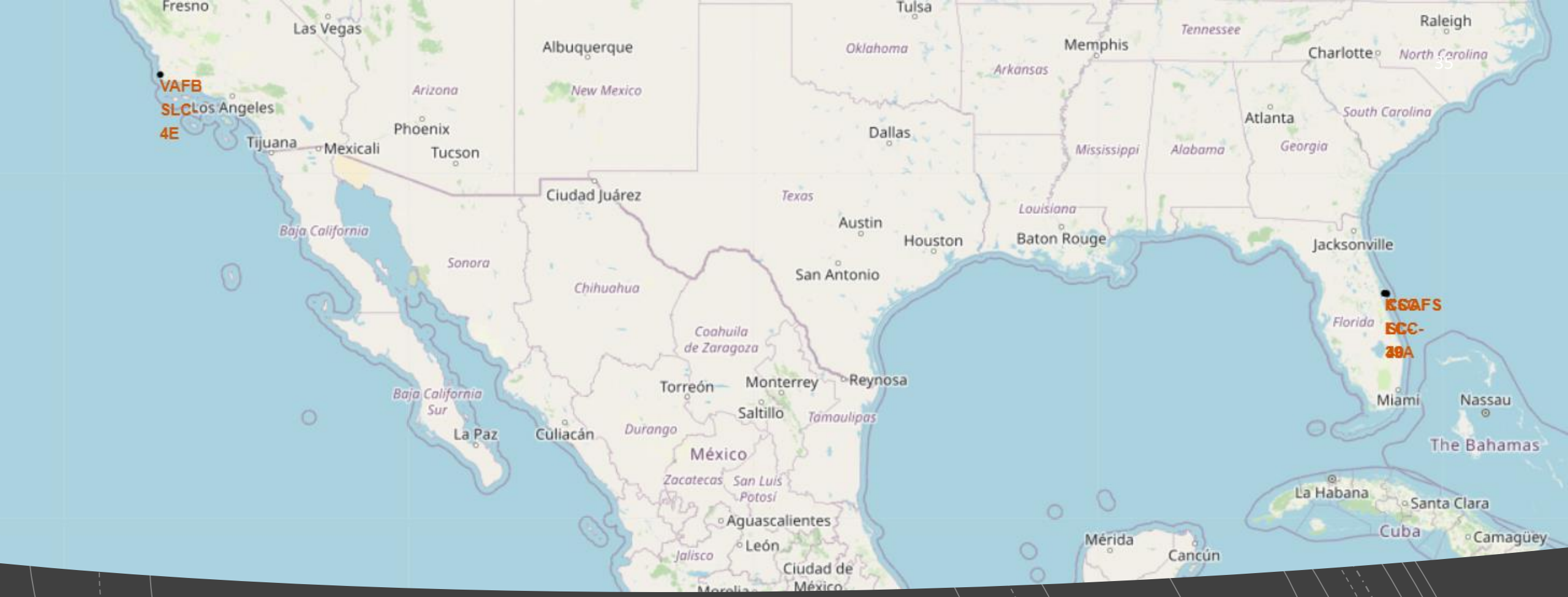
Done.

]:

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

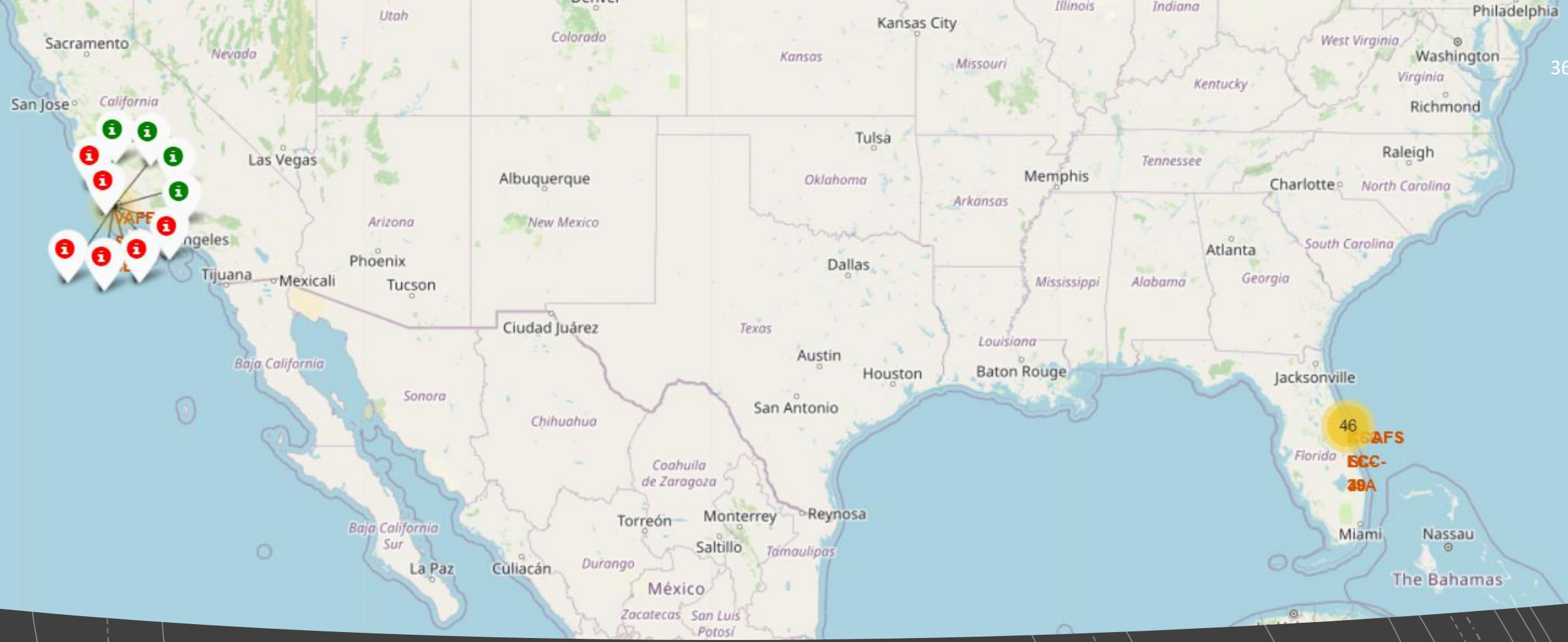
Section 4

Launch Sites Proximities Analysis



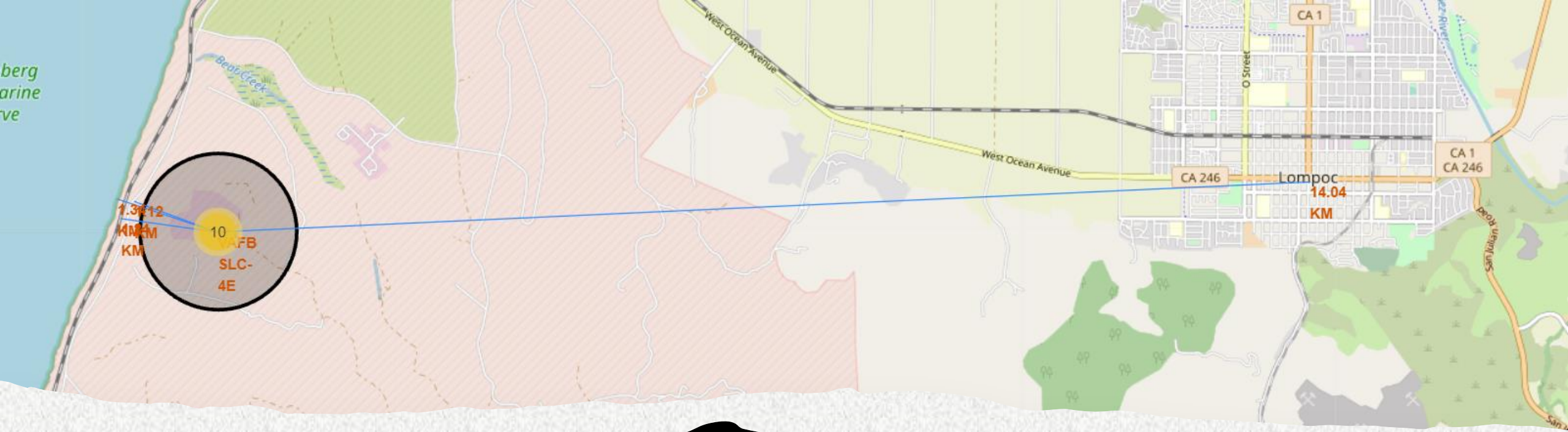
Marking all Launch Sites on a map

- All 4 launch sites are shown in the map above.
- One thing very obvious from the map is that all the Launch Sites are in very close proximity to the coastline.



Success/Failed launches for each site on a map

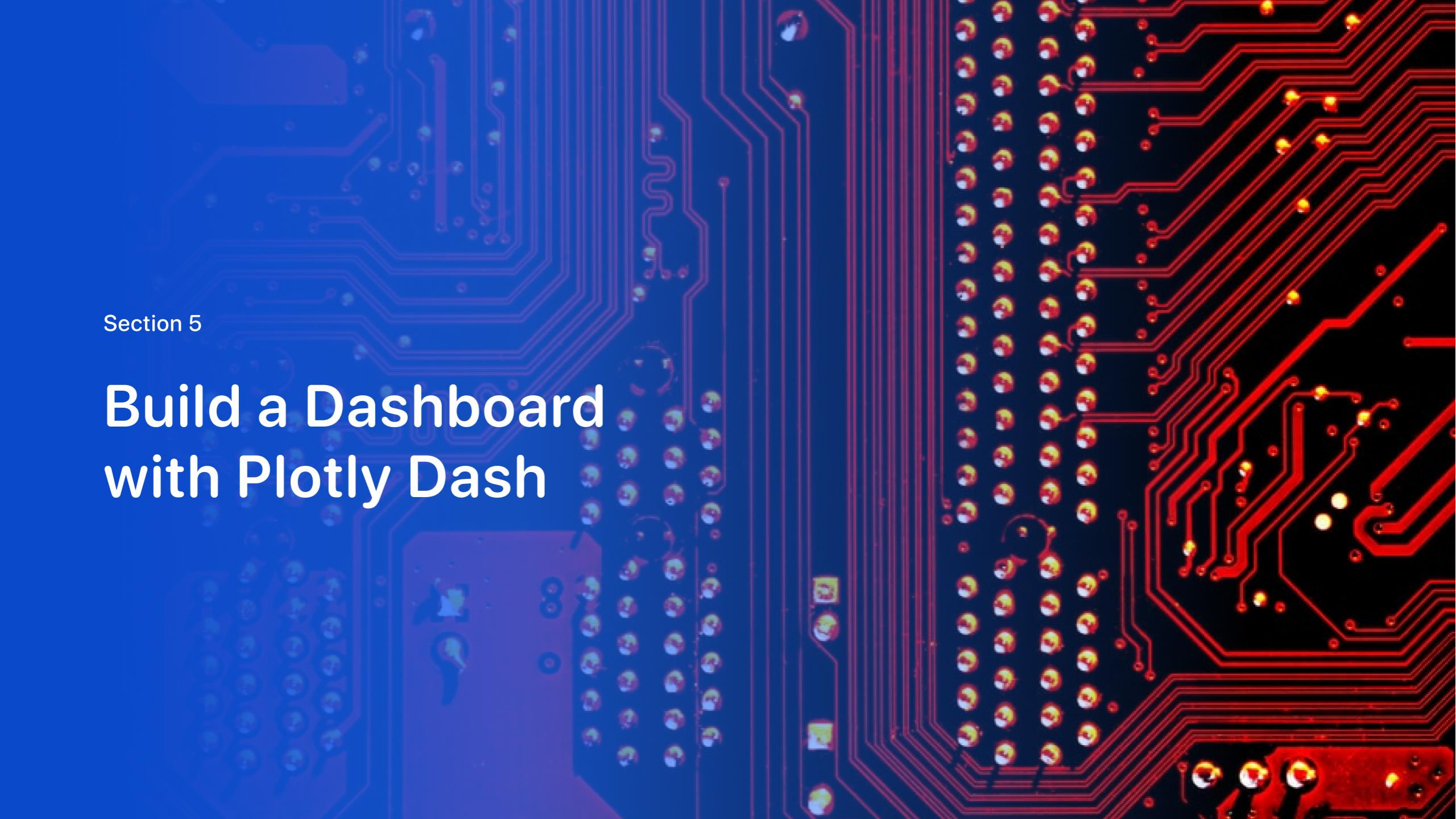
- The above map shows VAFB SLC 4E having been expanded to show the successful and failed launches.
- The red markers are the failed ones and the green markers represent the successful ones.
- Similarly the other sites can be expanded to show their success/failed launches as well.



Distance of a Launch Site to its proximities

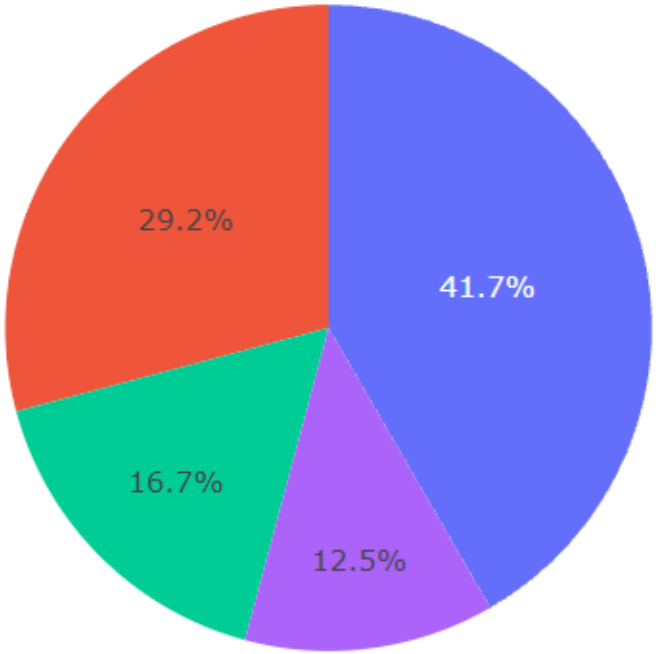
The Launch Site VAFB SLC 4E has been considered here and its distance from the folowing is shown in the map.

- Closest Highway : Lasalle Canyon Road (1.12 km)
- Closest Railway : Santa Barbara Subdivision MT1 (1.24 km)
- Closest City : Lompoc (14.04 km)
- Closest Coastline : Vandenberg State Marine Reserve (1.64 km)
- Therefore, its clear that launch sites keep a distance from cities as they don't want to cause destruction to cities in case of failed landings, although they maintain close proximities to railways, highways and coastlines.



Section 5

Build a Dashboard with Plotly Dash



Launch success counts for all sites

- The launch success counts of all sites are shown above.
- KSC LC-39A shows the highest launch success count.(41.7%)
- It is followed by CCAFS LC-40 (29.2%).

Launch Site with highest launch success ratio

- As seen in the previous slide, Launch Site KSC LC-39A has the highest launch success ratio.
- Therefore, the pie chart above shows the success (1) and failed (0) launches of the site KSC LC-39A.
- Successful launch percentage : 76.9%
- Failed launch percentage : 23.1%

Correlation between success and payload for all sites using payload slider

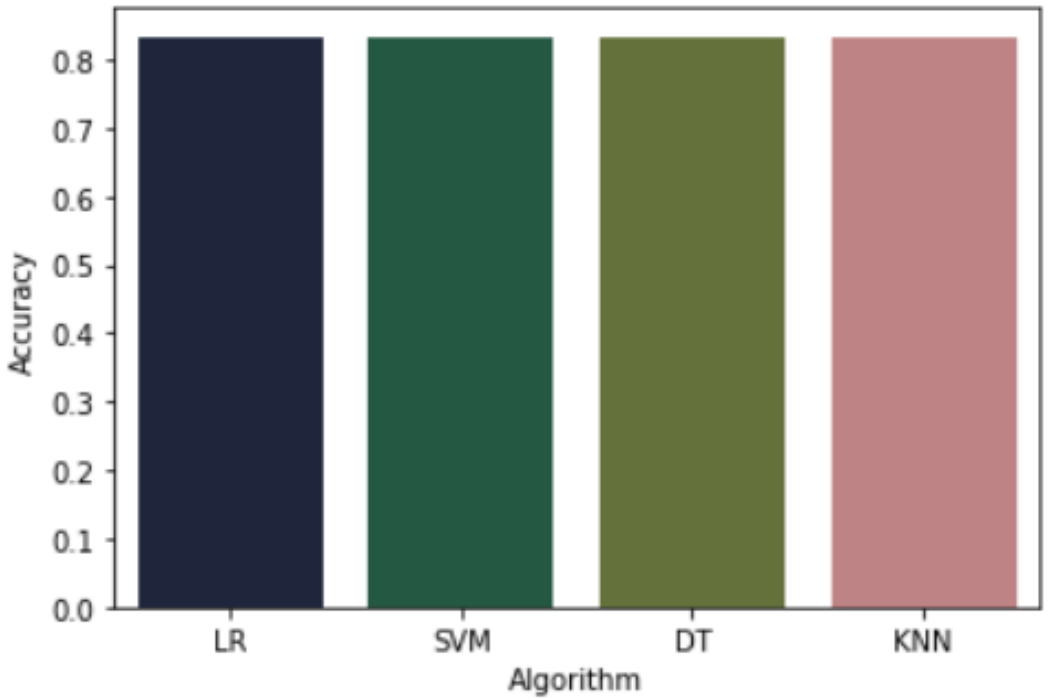
Show above is Payload vs. Launch Outcome scatter plot for all sites, with points colored as per Booster Version Category and different payload selected in the range slider.

- FT mostly gives successful launches when payload mass : 2000-5500 kg.
- B4 gives only successful launches when payload mass : 3000-5000 kg. Beyond this range B4 only has failed attempts.
- V1.1 hardly makes any contribution towards succesful launches as it only gave class 0 (failed attempts) irrespective of payload mass meaning there is no correlation between v1.1 booster category version and payload mass.

Section 6

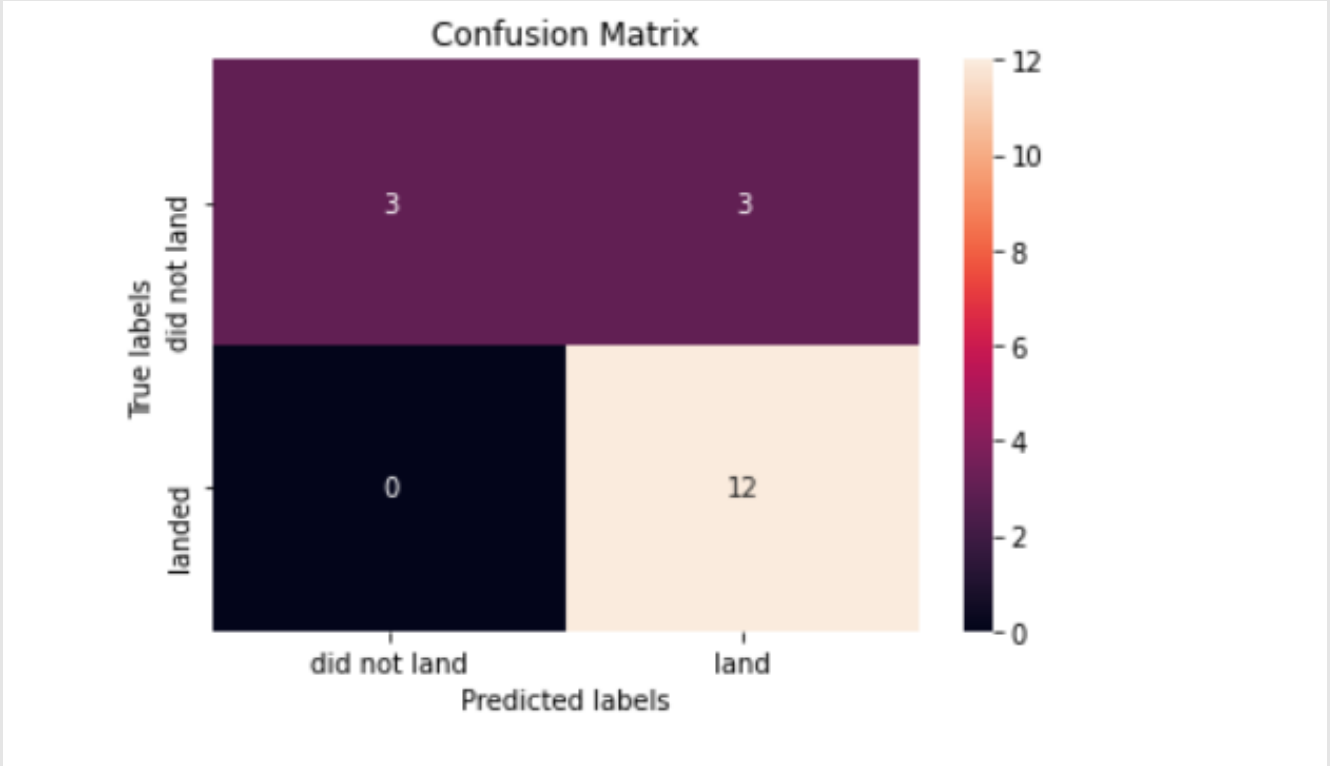
Predictive Analysis (Classification)

Classification Accuracy



- Visualize the built model accuracy for all built classification models, in a bar chart
- LR – Logistic Regression
- SVM – Support Vector Machine
- DT – Decision Tree
- KNN – K Nearest Neighbors
- All the algorithms show the same accuracy.

Confusion Matrix



- Since all the classification algorithms have the same accuracy on the test data we have only one common confusion matrix for all
- Shown on the left is the Confusion Matrix.

Conclusions

Some features that affect success rate are Flight No., Pay load mass, Orbits, Launch Site, Flights, Reused, Legs, Landing Pad and Serial.

Launch Sites are in close proximities to railways, highways and coastlines.

Pay load mass and Booster version category together also play a role in determining successful launches.

The classification algoritms namely,

- Logistic Regression
- Support Vector Machine
- Decision Trees
- K Nearest Neighbors

All the above algorithms perform well on the test data when tuned to find the best parameters.

Appendix

As for Predictive Analysis (Classification) the best options are -

- Logistic Regression with

  {'C':0.01, 'penalty':'l2', 'solver':'lbfgs'}

- Support Vector Machine with

  {'C':1.0, 'gamma':0.031622776601683793, 'kernel':'sigmoid'}

- Decision Tree with

  {'criterion':'gini', 'max_depth':6, 'max_features':'auto', 'min_samples_leaf':2, 'min_samples_split':5, 'splitter':'random'}

- K-Nearest Neighbors with

  {'algorithm':'auto', 'n_neighbors':10, 'p':1}

All the above tuned parameters are found by Grid Search CV.

Thank you!