IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

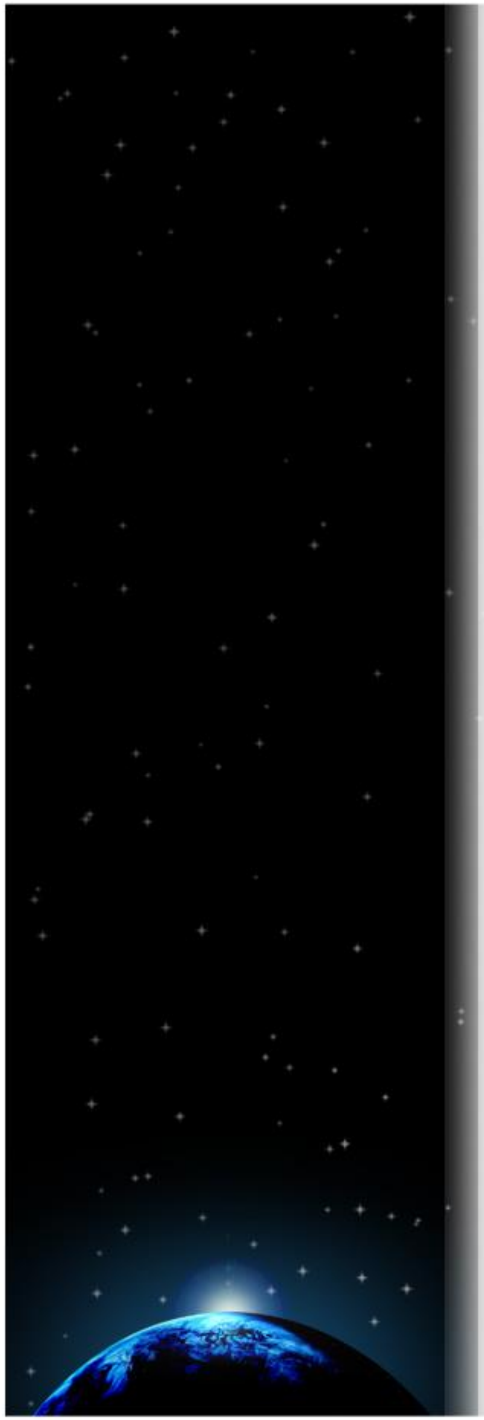
Arne Wörheide
10. October 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all result
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

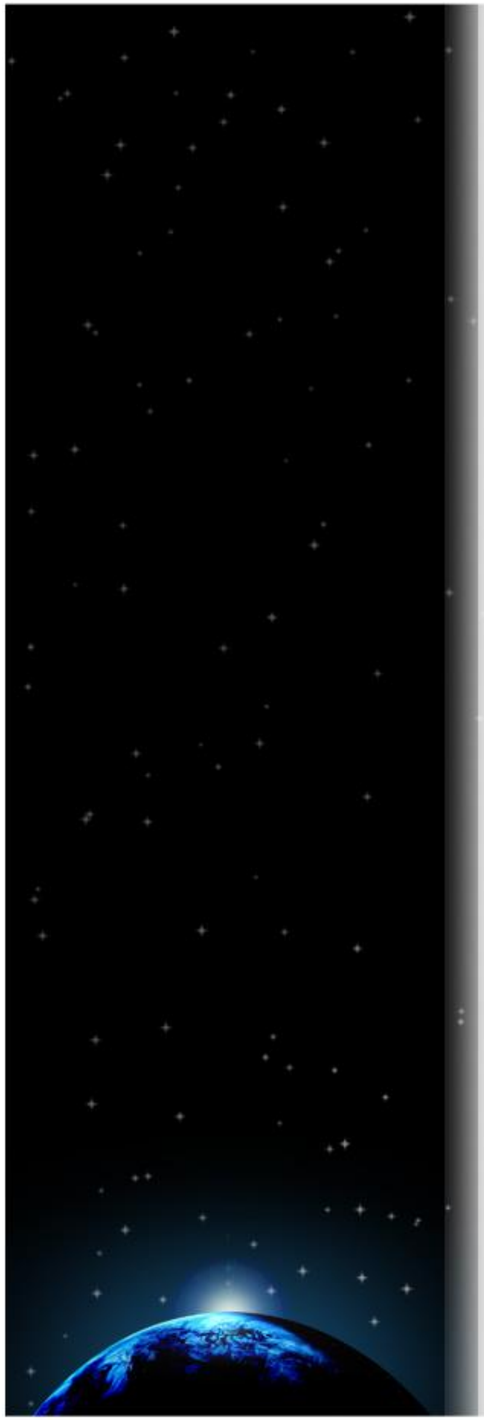


Introduction

- The objective of this project is to estimate the expenses associated with a space launch by making predictions about the successful landing of the Falcon 9's first stage.
- SpaceX promotes its Falcon 9 rocket launch as costing around 62 million dollars, while other space launch providers charge upwards of 165 million dollars per launch. A significant portion of the cost savings comes from SpaceX's ability to reuse the initial stage of the rocket.
- Consequently, if we can forecast whether the first stage of the Falcon 9 will successfully return and land, we can calculate the overall cost of a launch. This information could prove valuable for potential investors.

Section 1

Methodology



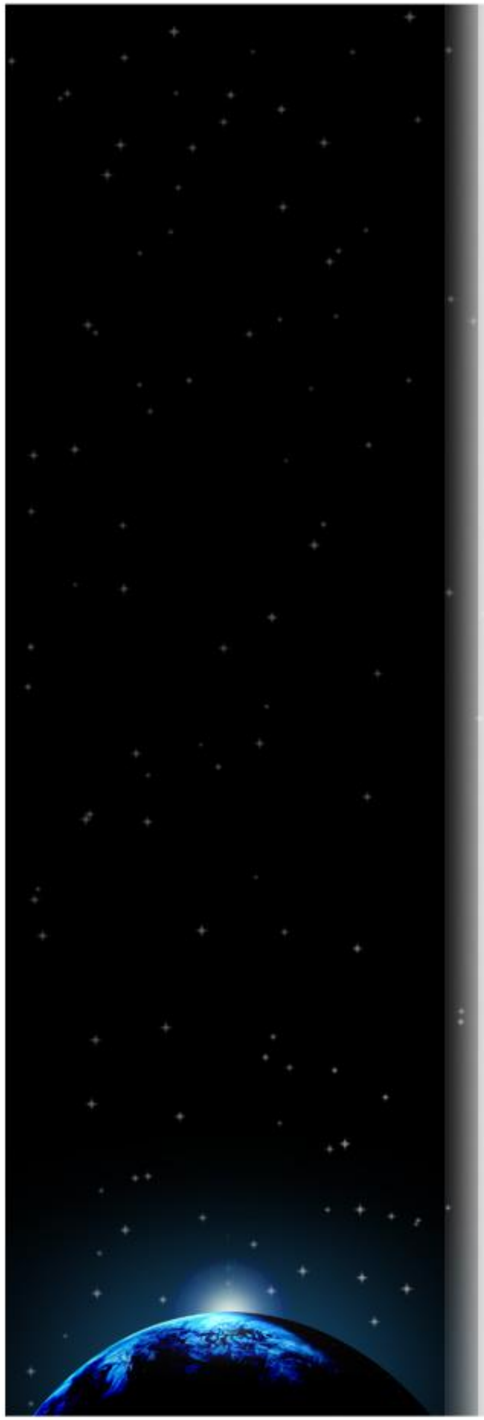
Methodology

- Data collection methodology:
  - Call SpaceX REST API
  - Web Scrapping from Wikipedia
- Performed data wrangling
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
  - Building, tuning and evaluation of classification models to ensure the best results

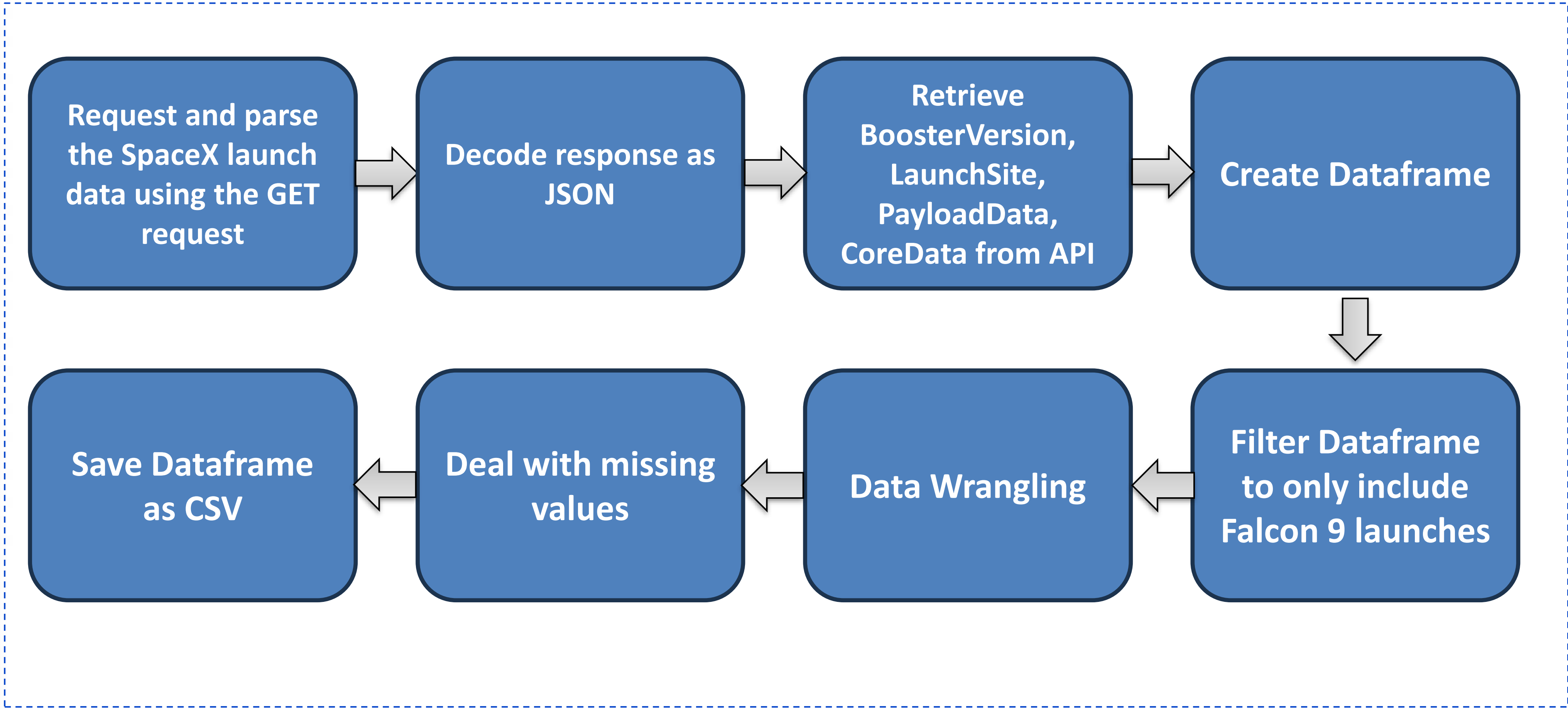
Data Collection

- REST API
  - results are more detailed
  - calls and responses are easy to handle in Python
- Web Scraping
  - results not so detailed
  - needs more coding
  - code could be outdated if HTML structure is changed

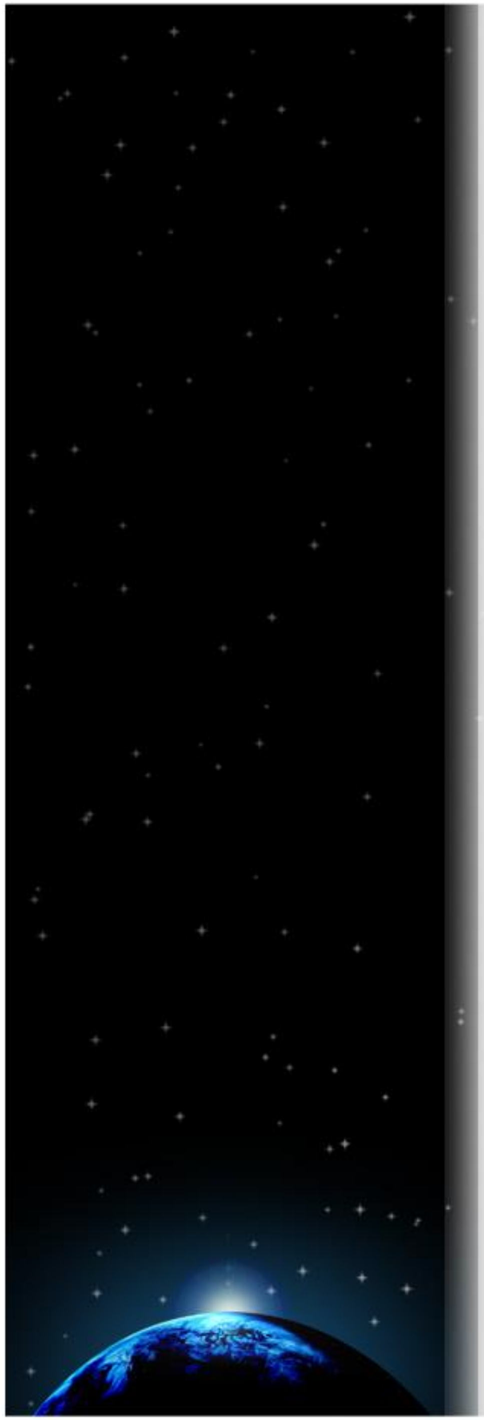
* Both methods only used Falcon 9 launches



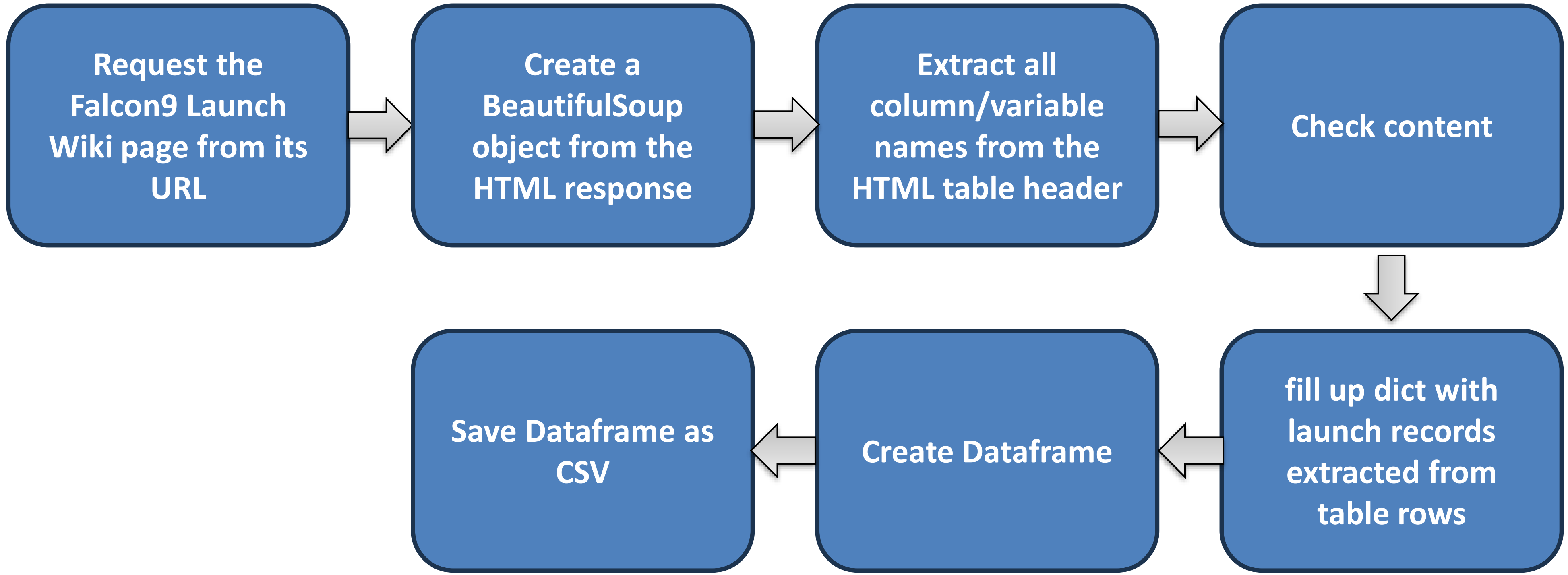
Data Collection – SpaceX API



Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_1/jupyter-labs-spacex-data-collection-api%20(1).ipynb



Data Collection – Scraping



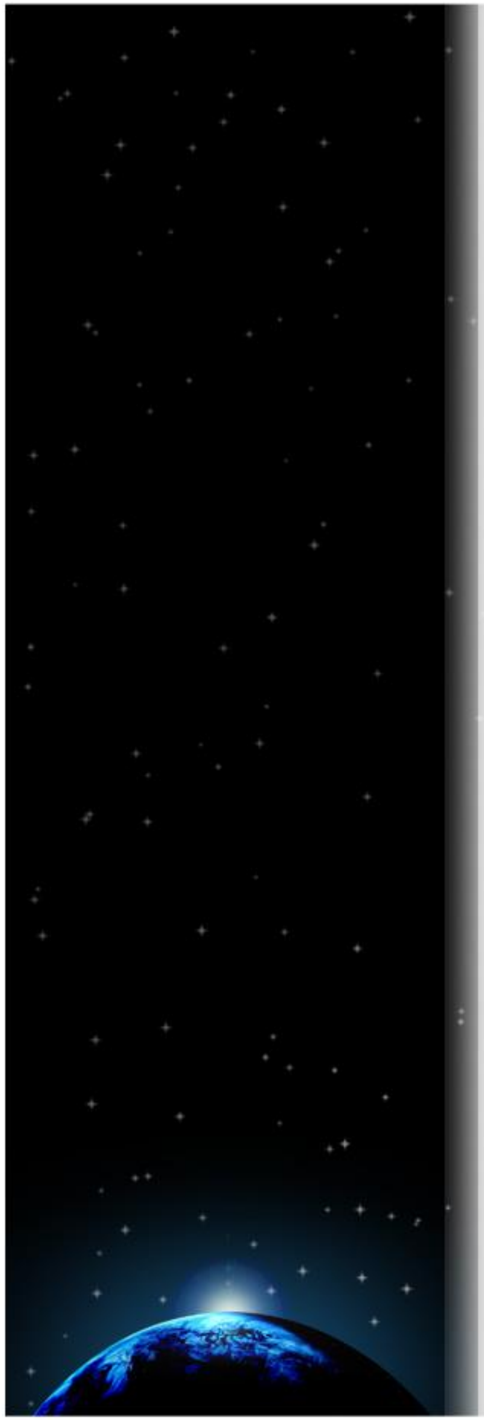
Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_1/jupyter-labs-webscraping.ipynb

Data Wrangling

- Data Exploration to determine label for supervised models
  - Number of launches on each site
  - Number and occurrence of each orbit
  - Number and occurrence of mission outcome per orbit type

- Add new training label "class"
  - Class 0: first stage booster did not land successfully
  - Class 1: first stage booster landed successfully

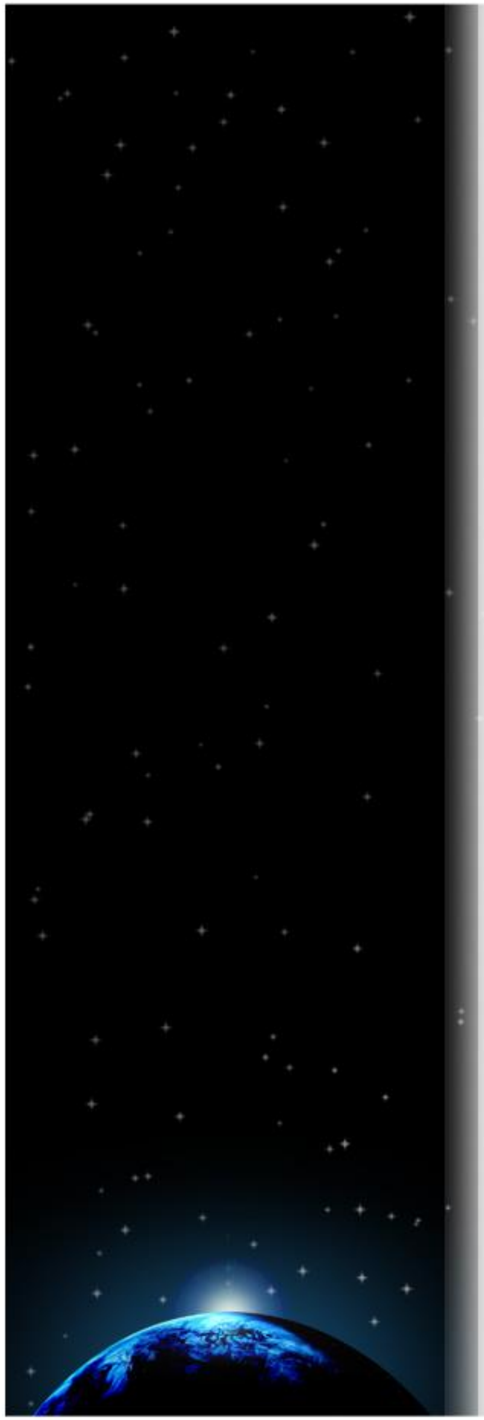
Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_1/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb



EDA with Data Visualization

- Following plots are created to see if and how the variables are related to each other
  - Flight Number X Payload Mass
  - Flight Number X Launch Site
  - Payload Mass X Launch Site
  - Orbit Type X Success Rate
  - Flight Number X Orbit Type
  - Payload Mass X Orbit Type
  - Success Rate X Year

Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_2/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

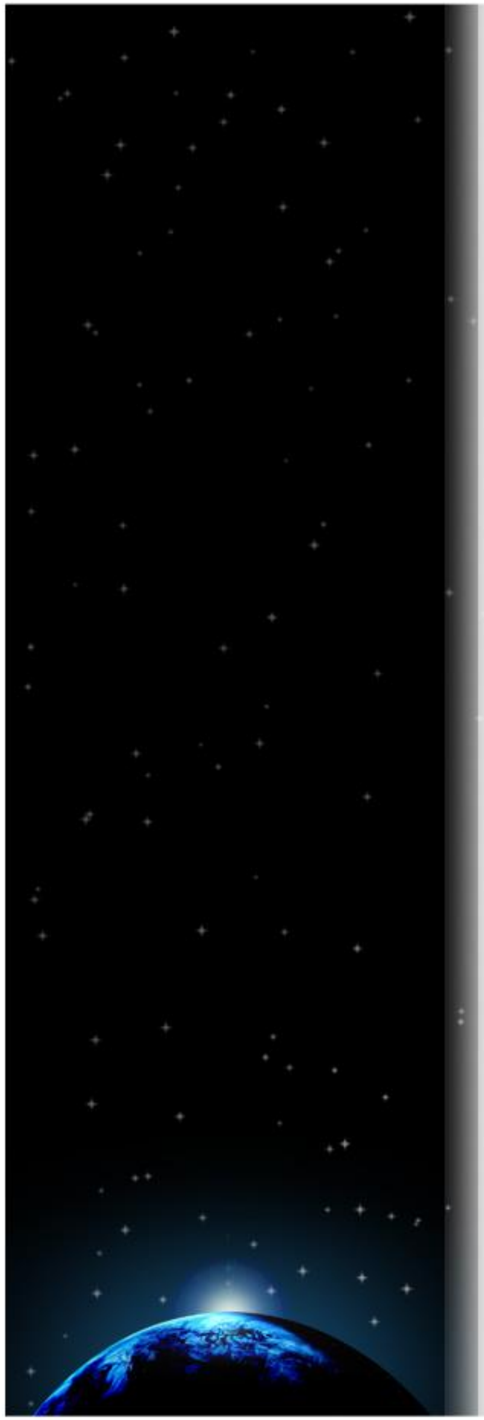


EDA with SQL

Several SQL queries where performed to

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'KSC'
- 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 where the succesful landing outcome in drone ship was achieved.
- List the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass using a subquery
- List the records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017
- 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

Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_2/jupyter-labs-eda-sql-edx_sqllite.ipynb



Build an Interactive Map with Folium

Following information was added to the map

- Circles for launch site coordinates and names
- Icon markers for launch site names
- Icon markers for success/failed launches
- Line and Icon to show distance from launch site to coastline

Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_3/lab_jupyter_launch_site_location.jupyterlite.ipynb

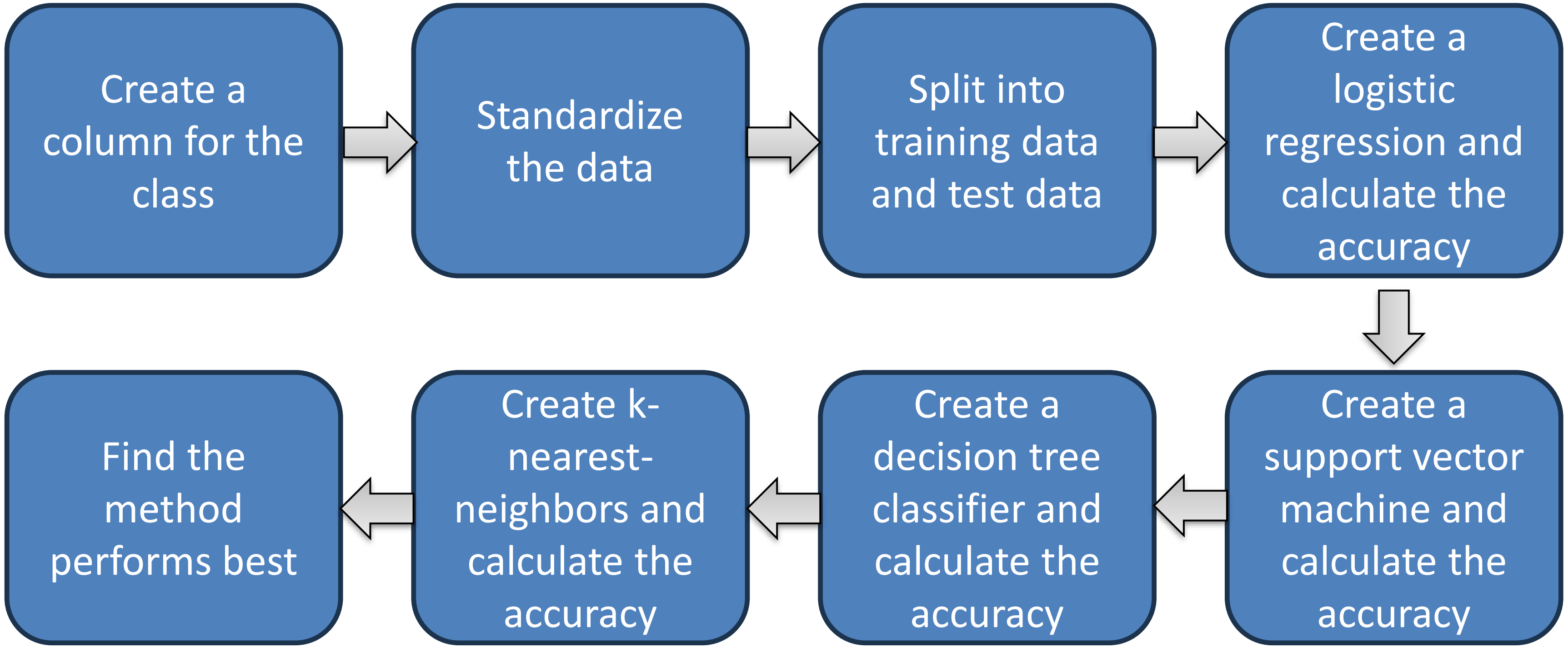
Build an Interactive Map with Plotly Dash

Following items were added to the dash

- Dropdown list to select the launch site
- Pie chart showing success status of launches
- Slider to select Payload mass
- Scatter chart to show correlation of Payload mass and Success rate for different booster versions

Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_3/spacex_dash_app.py

Predictive Analysis (Classification)



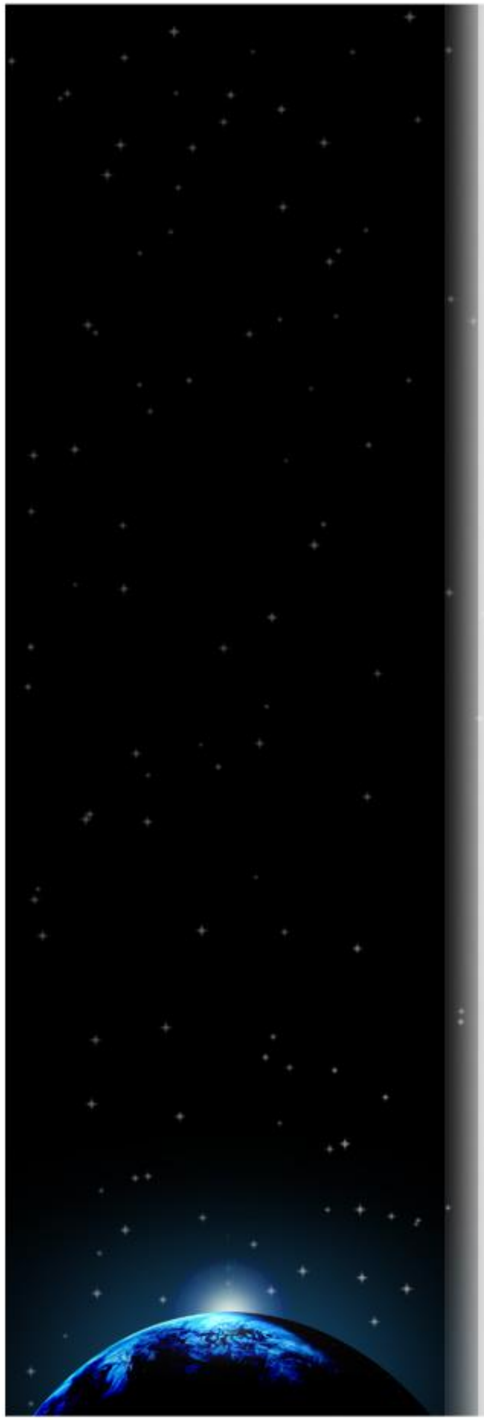
Link to Notebook: https://github.com/DL1XY/edx/blob/main/final/Module_4/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

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

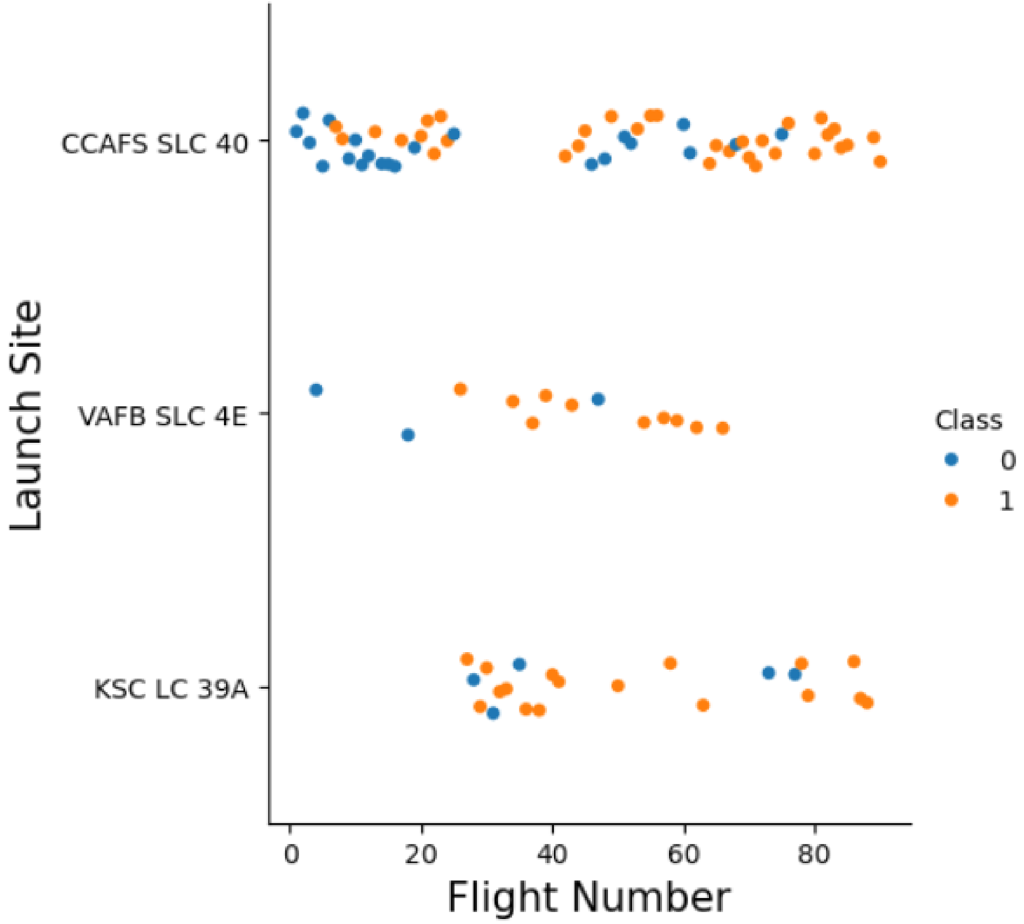
Section 2

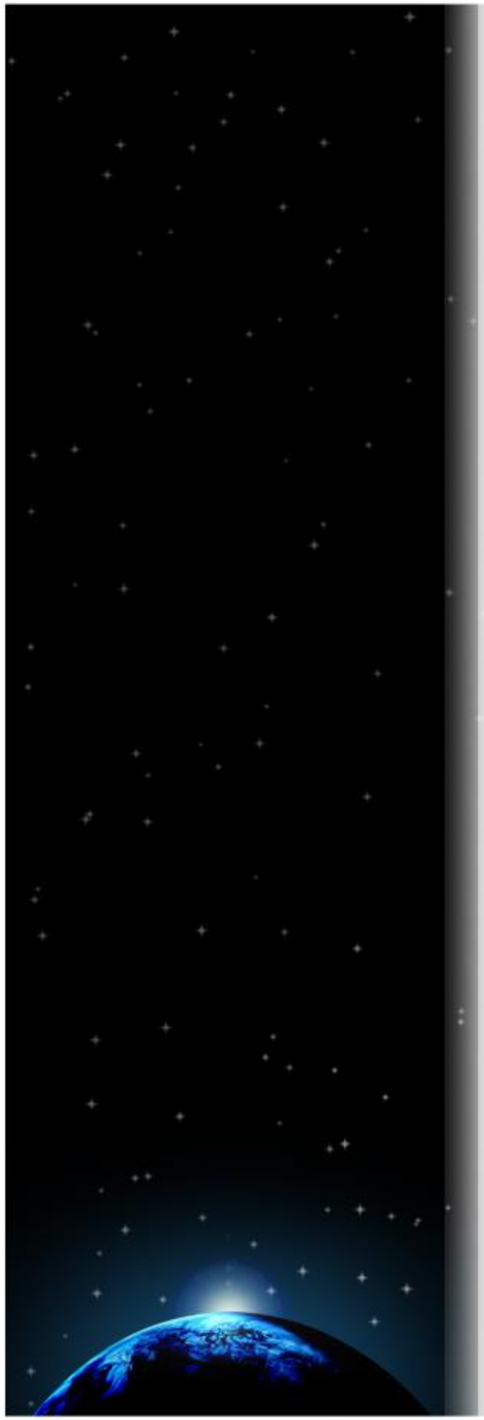
Insights drawn from EDA



Flight Number vs. Launch Site

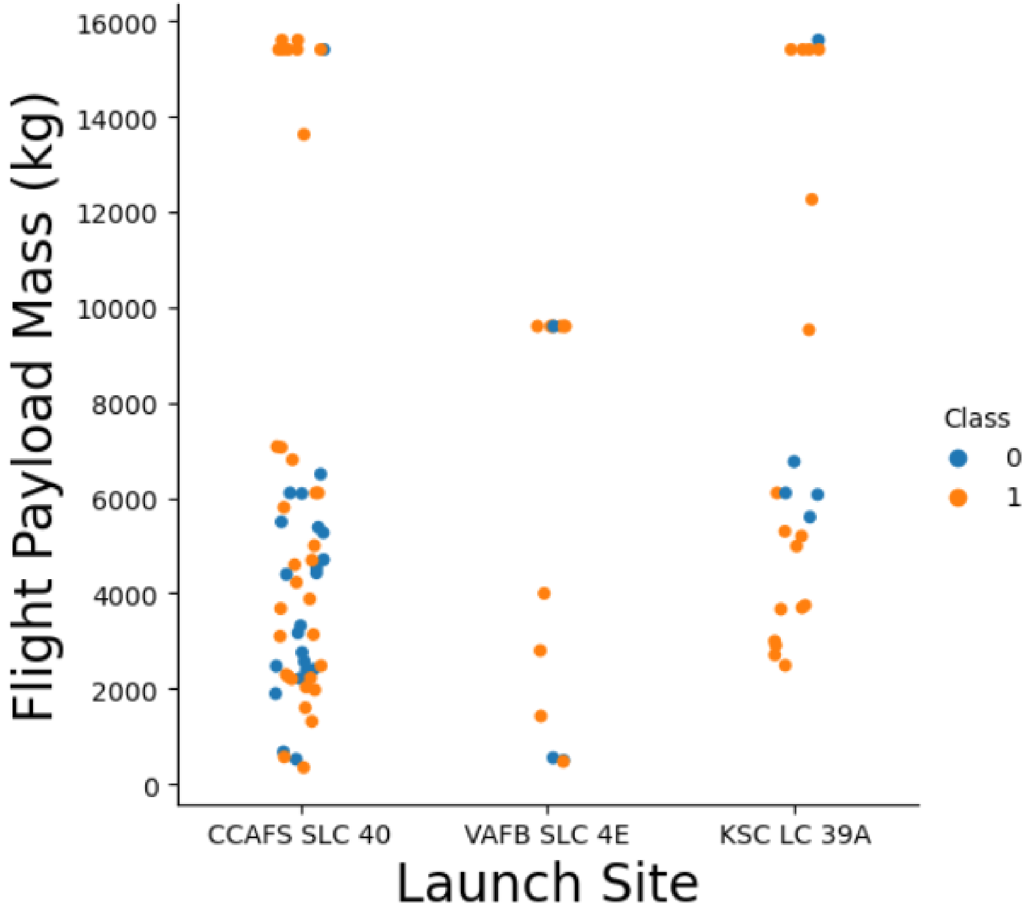
- CCFAS SLC 40 has most launches, but a gap between Flight Number 25 and 40
- VAFB SLC 4E is rarely used and not used since Flight Number 70
- KSC LC 39A first launch was around Flight Number 25





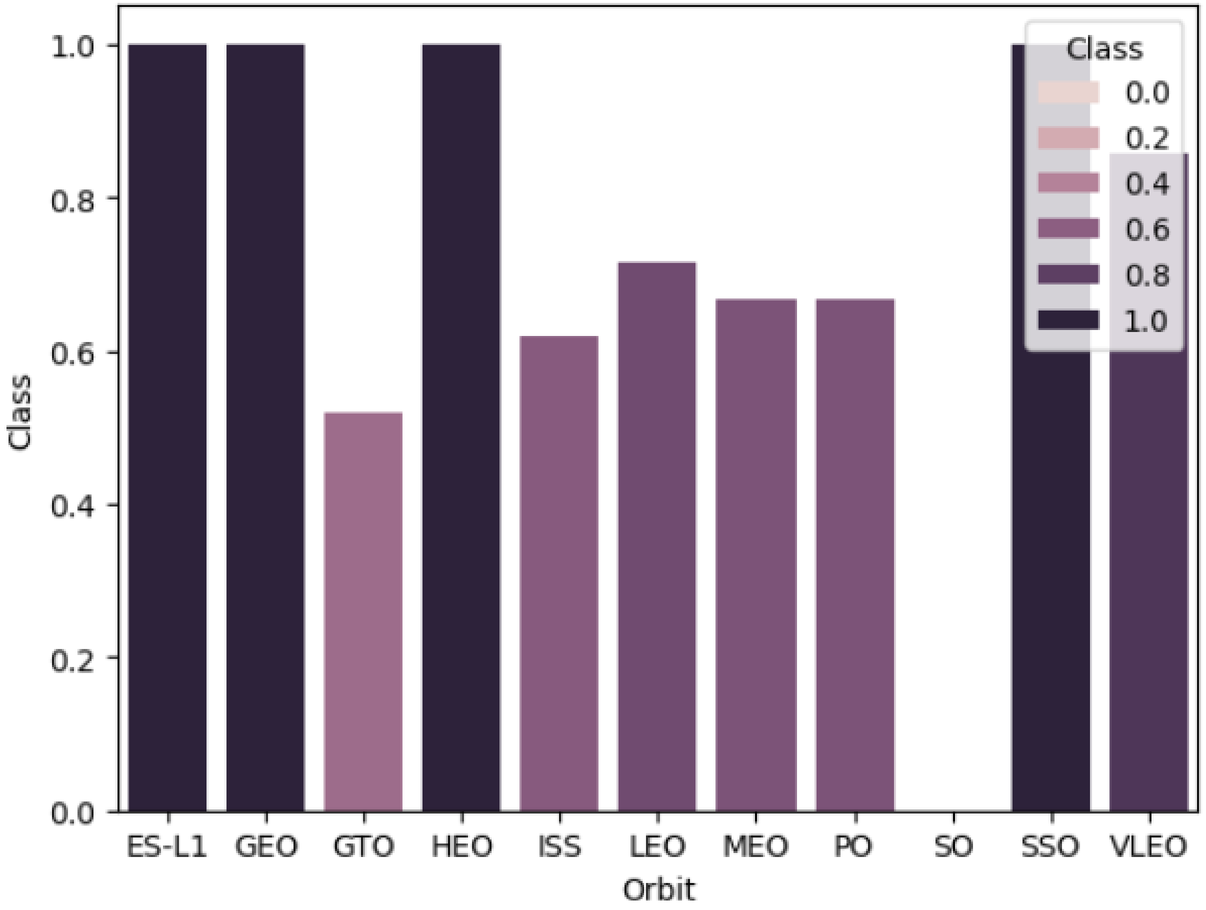
Payload vs. Launch Site

- Most Payloads are less than 8000 kg
- VAFB SLC 4E was not used for payloads greater 10000 kg



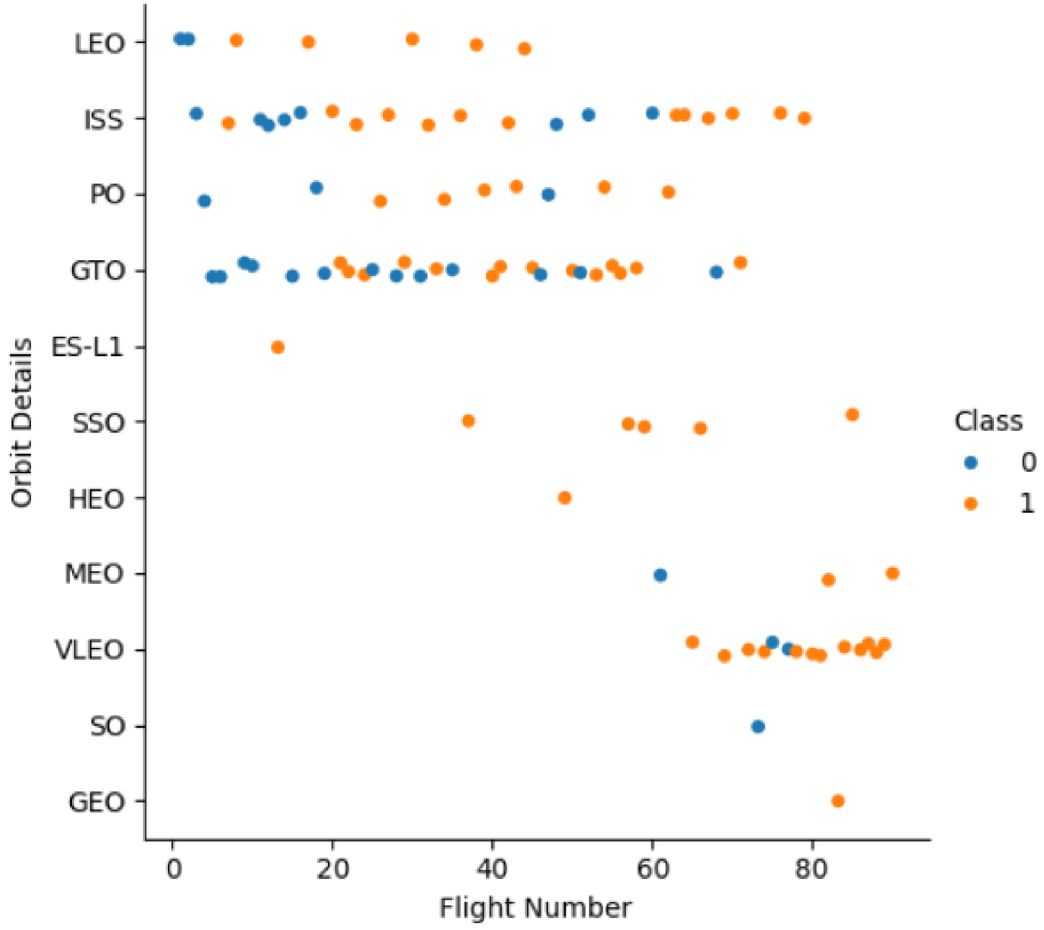
Success Rate vs. Orbit Type

- ES-L1, GEO, SSO and VLEA have most success rates
- GTO hat no successful launch



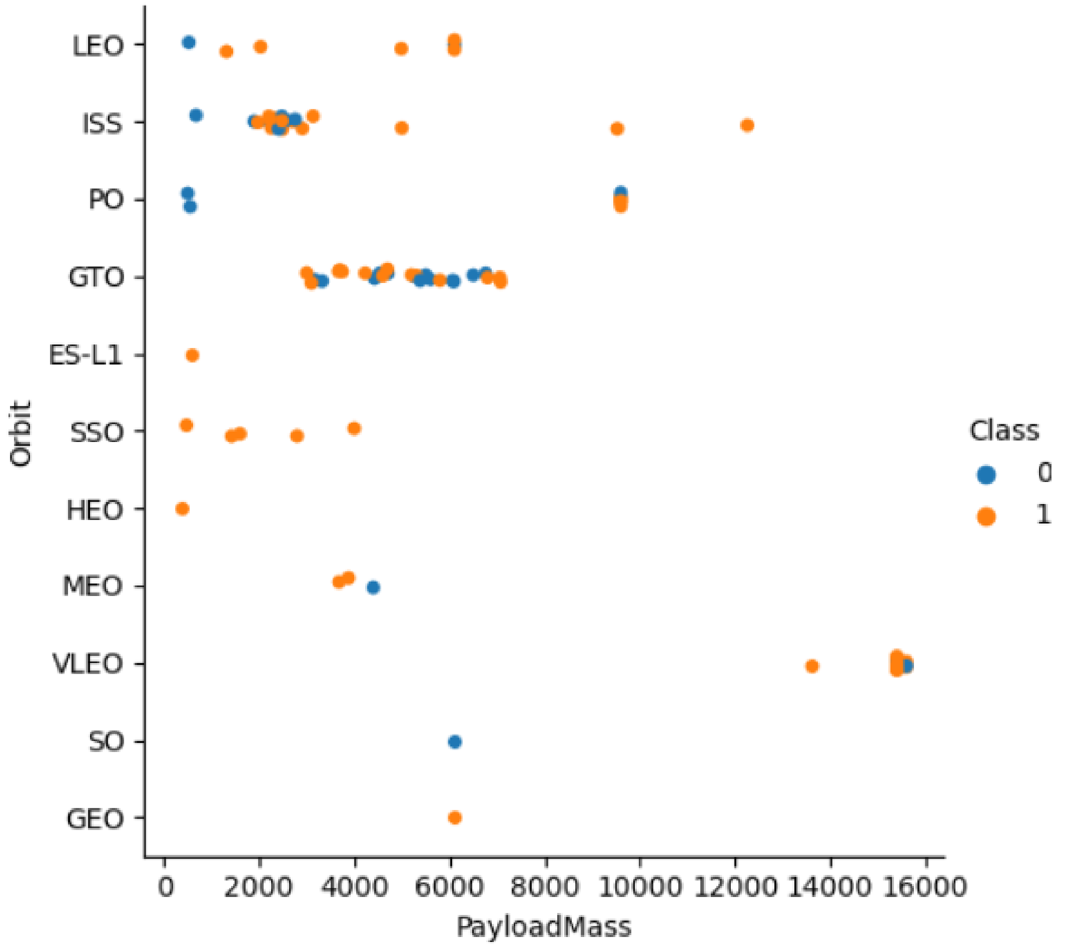
Flight Number vs. Orbit Type

- LEO, ISS, PO and GEO orbits are mostly targeted
- ES-L1, HEO, SO and GEO where only targeted once
- VLEO is mostly targeted since Flight 62



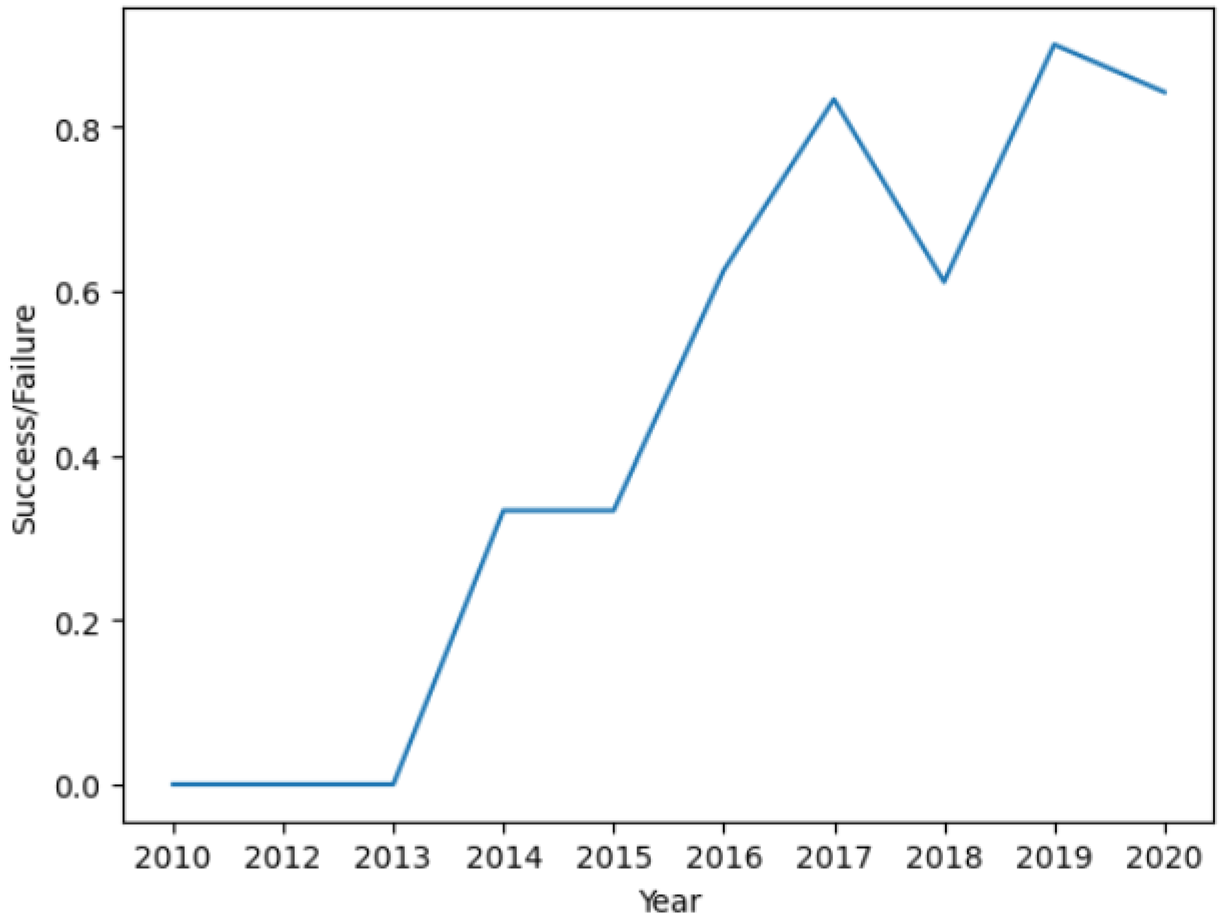
Payload vs. Orbit Type

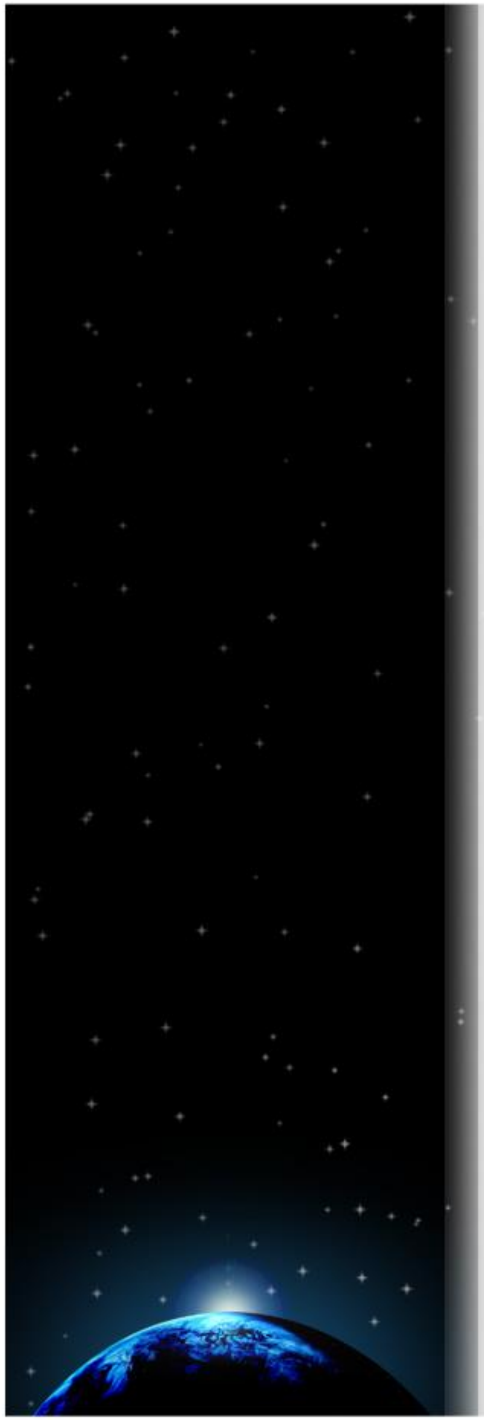
- Highest payloads are on VLEO
- GTO payload ranges from 3000 to 8000 kg



Launch Success Yearly Trend

- From 2015 to 2017 the success rate was increasing
- In 2017 there were mor failed than successful launches





All Launch Site Names

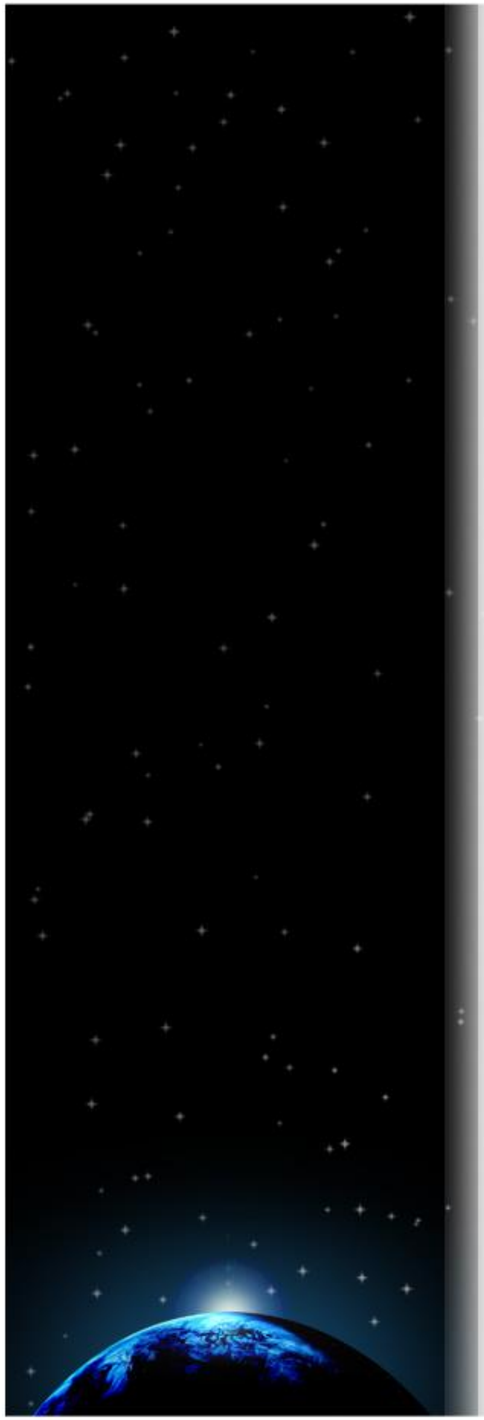
- Four launch sites are used for Falcon 9

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

Launch Site Names Begin with 'KSC'

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	06:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-01-05	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

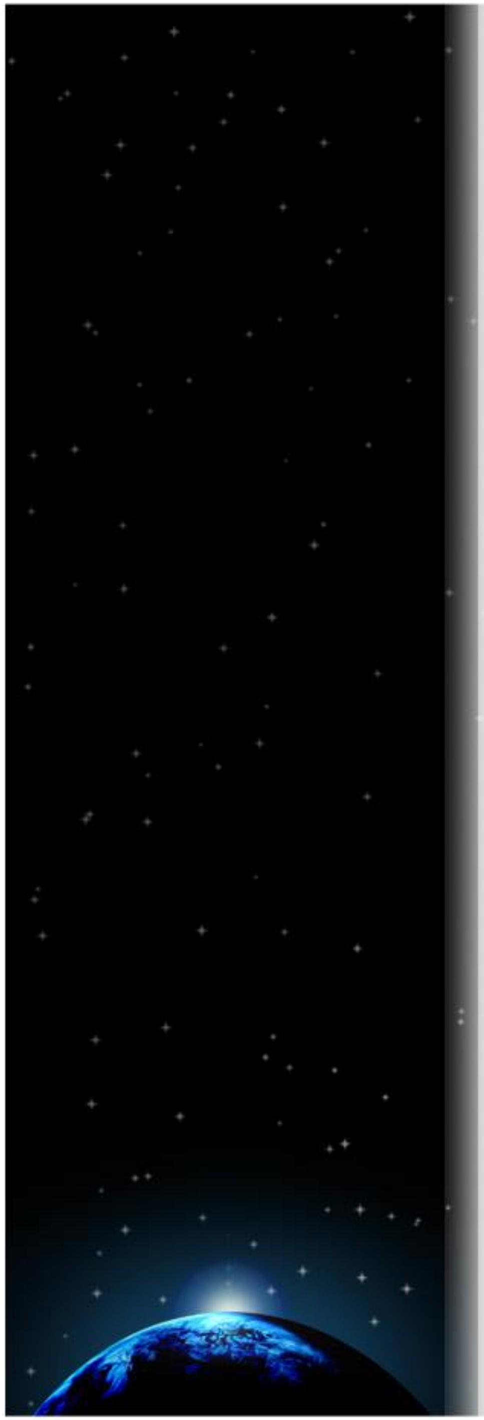
- 5 records where launch sites’ names start with `KSC`
- LEO and GEO orbits are targeted
- All 5 launches where successful



Total Payload Mass

- Total payload carried by boosters from NASA using SQL sum() function

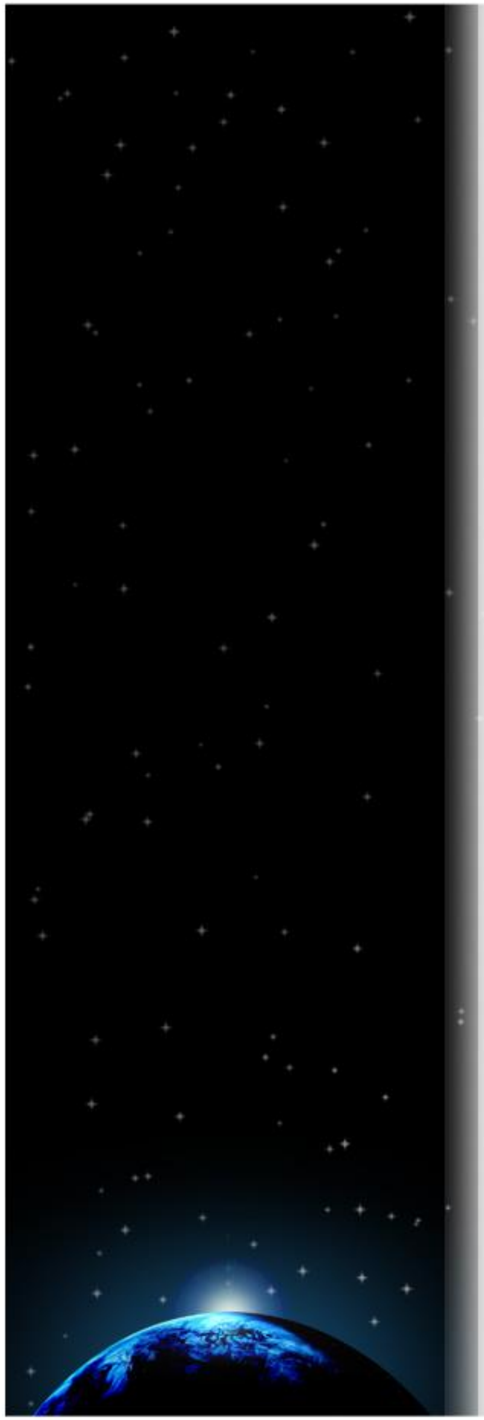
sum(PAYLOAD_MASS__KG_)
45596



Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 using SQL avg() function

```
+------------------------+
| avg(PAYLOAD_MASS__KG_) |
+------------------------+
|         2928.4         |
+------------------------+
```



First Successful Ground Landing Date

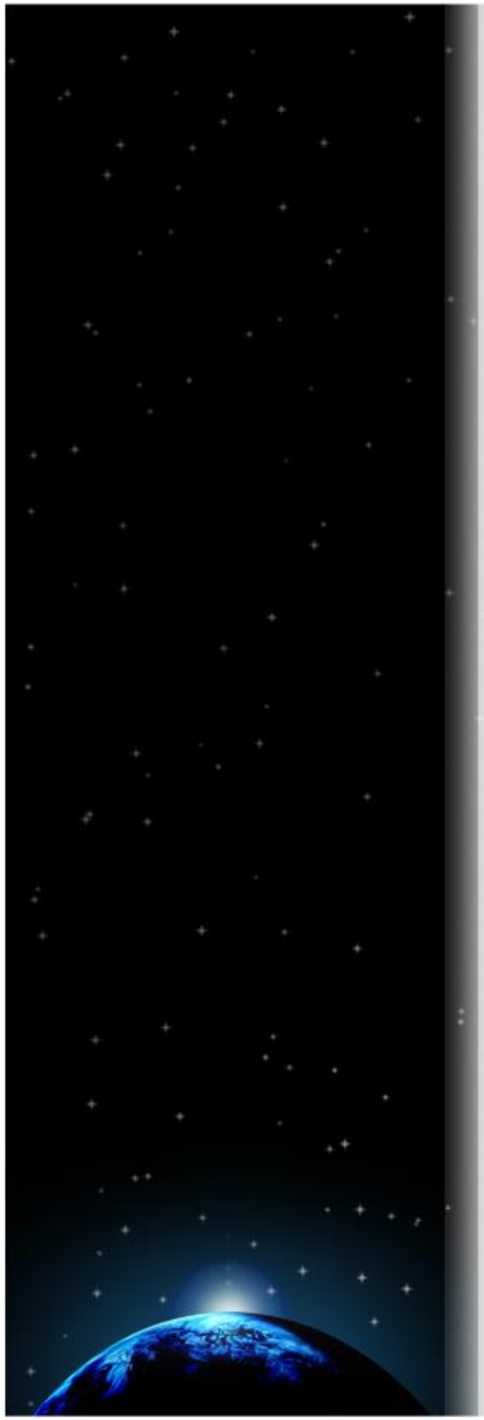
- First successful landing outcome on drone ship was 2015-12-22

min(date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

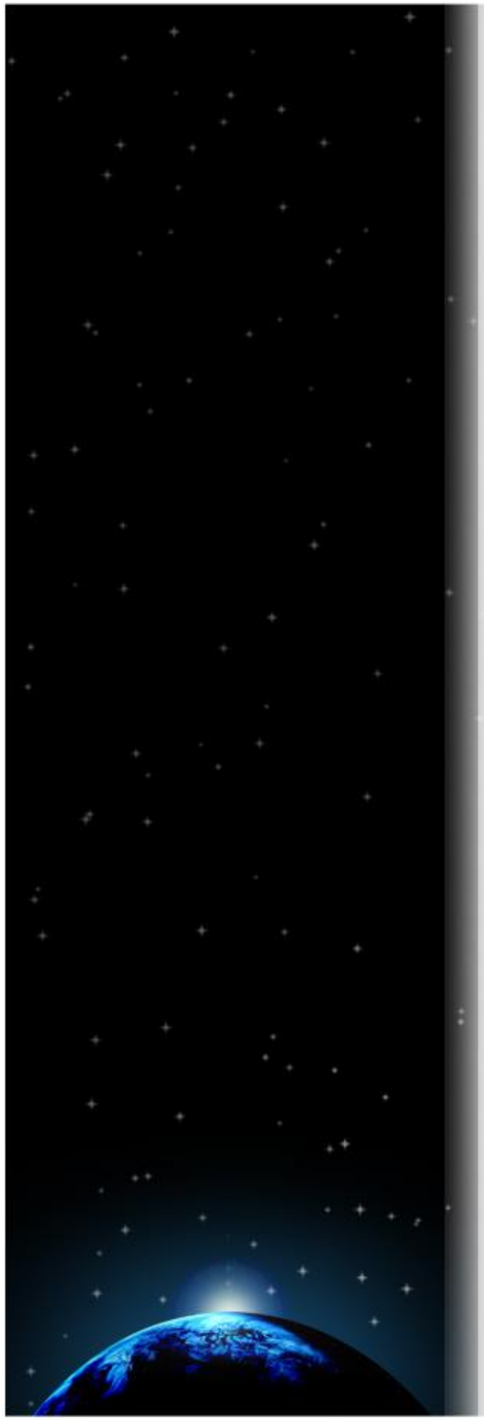
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2



Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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



Boosters Carried Maximum Payload

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

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

2017 Launch Records

- Month, successful landing outcome in ground pad ,booster versions and launch site for the months in year 2017

substr(Date,6,2)	Booster_Version	Launch_Site	Landing_Outcome
01	F9 FT B1029.1	VAFB SLC-4E	Success (drone ship)
02	F9 FT B1031.1	KSC LC-39A	Success (ground pad)
03	F9 FT B1021.2	KSC LC-39A	Success (drone ship)
01	F9 FT B1032.1	KSC LC-39A	Success (ground pad)
03	F9 FT B1035.1	KSC LC-39A	Success (ground pad)
06	F9 FT B1029.2	KSC LC-39A	Success (drone ship)
06	F9 FT B1036.1	VAFB SLC-4E	Success (drone ship)
08	F9 B4 B1039.1	KSC LC-39A	Success (ground pad)
08	F9 FT B1038.1	VAFB SLC-4E	Success (drone ship)
07	F9 B4 B1040.1	KSC LC-39A	Success (ground pad)
09	F9 B4 B1041.1	VAFB SLC-4E	Success (drone ship)
11	F9 FT B1031.2	KSC LC-39A	Success (drone ship)
10	F9 B4 B1042.1	KSC LC-39A	Success (drone ship)
12	F9 FT B1035.2	CCAFS SLC-40	Success (ground pad)

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

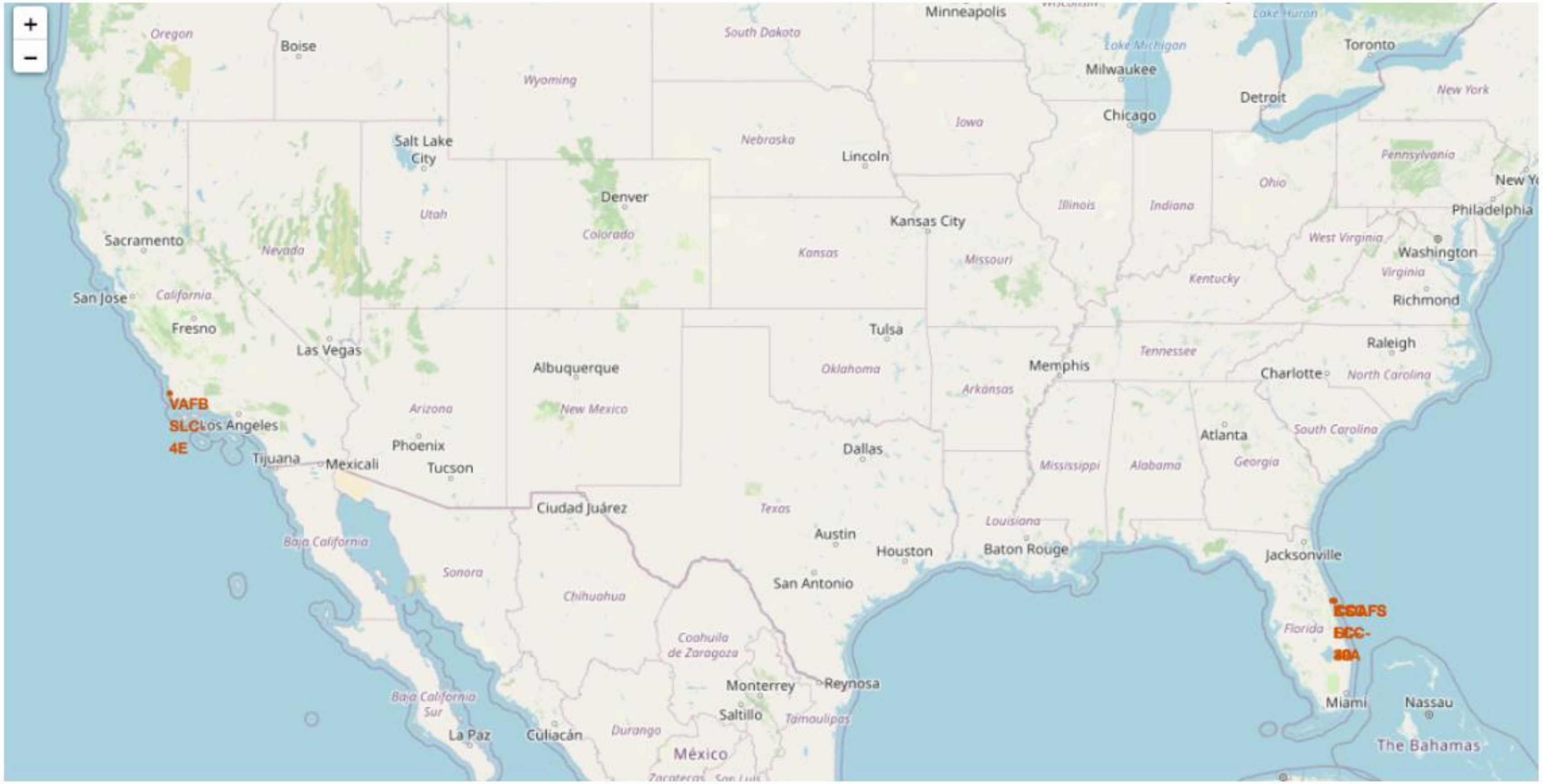
- Rank the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

MISSION_OUTCOME_COUNT	Launch_Site
26	CCAFS LC-40
34	CCAFS SLC-40
25	KSC LC-39A
16	VAFB SLC-4E

Section 3

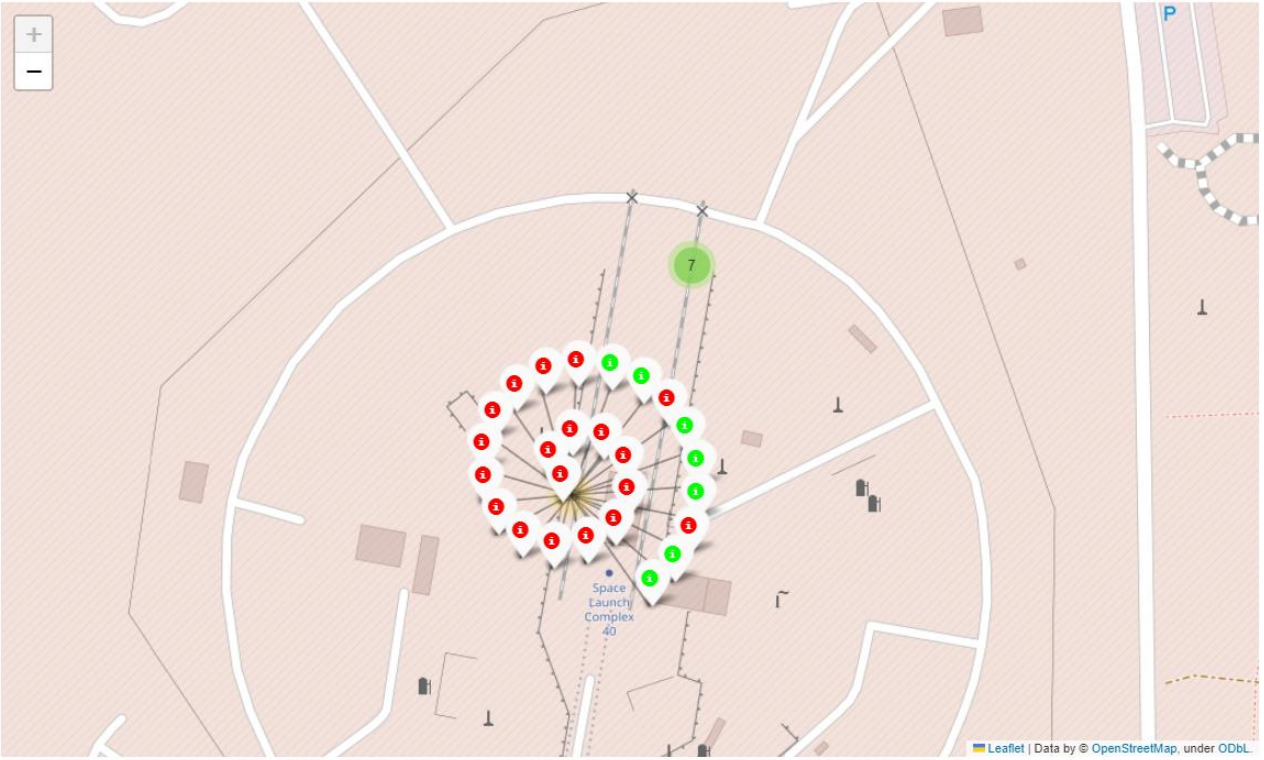
Launch Sites Proximities Analysis

Folium Map showing Landing sites



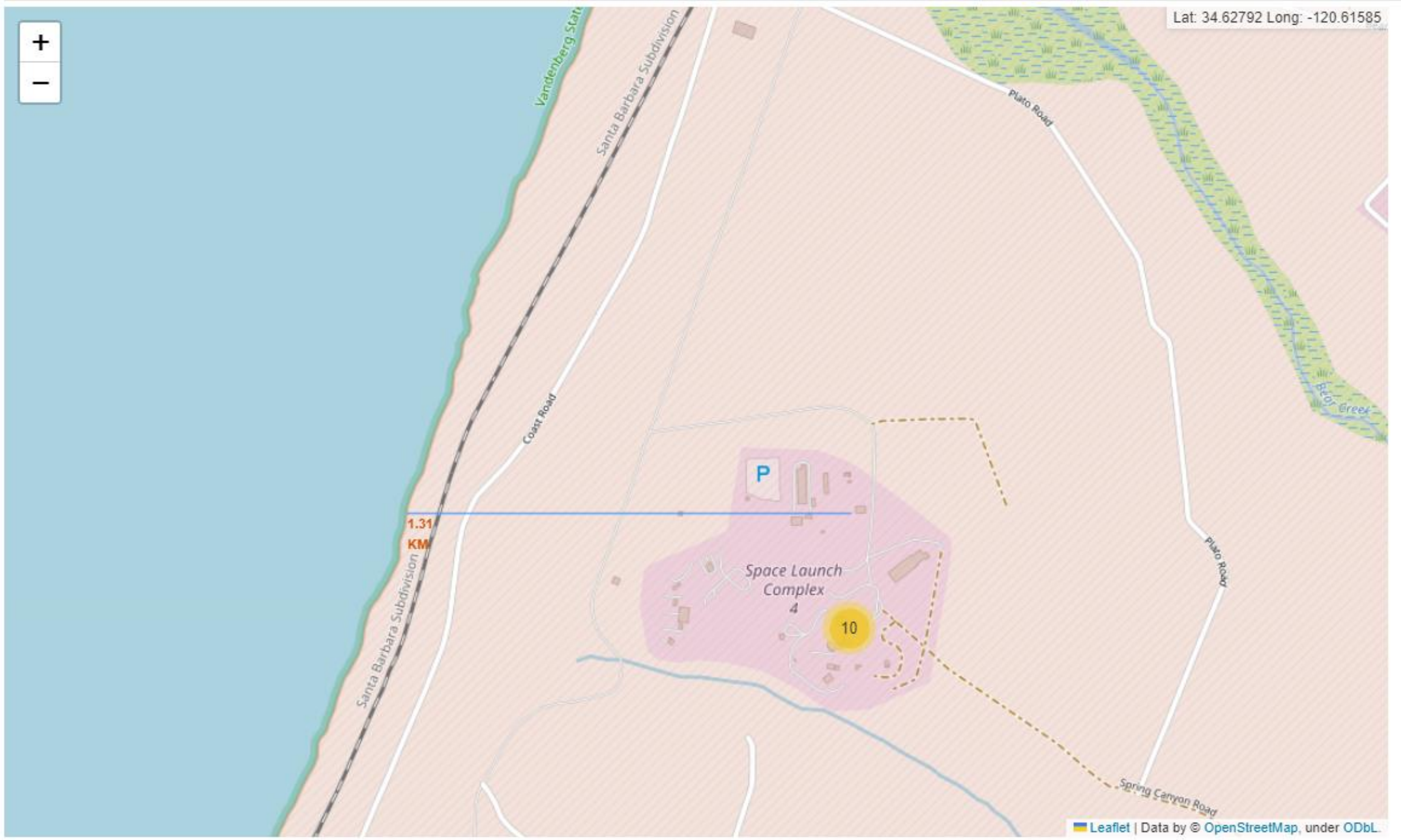
Markers in California and Florida showing Landing Sites

Folium Map showing launch outcomes



Color Labeled launch outcomes on CCAFS LC-40

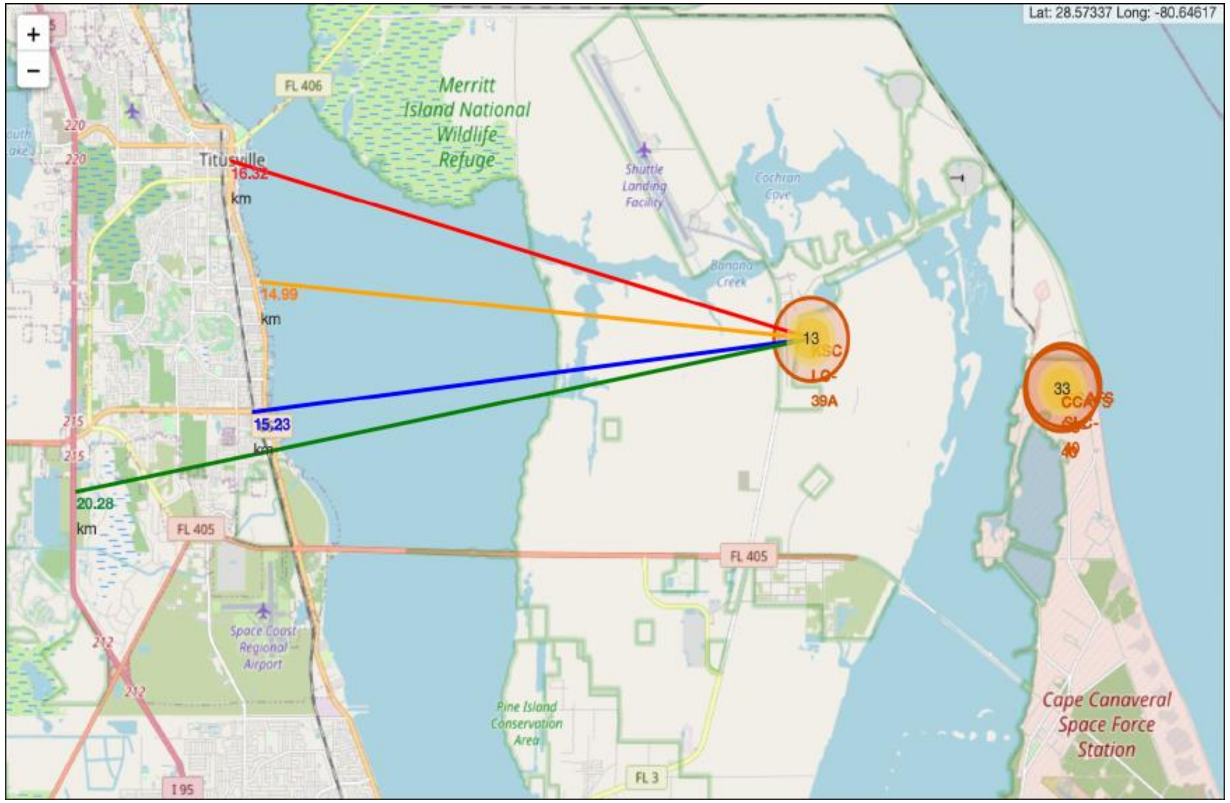
Folium Map showing distance



Line showing distance from Launch site to coast

KSC LC-39A infrastructure distances

- KSC LC-39A is near to civil infrastructure
- A failed launch can result in a catastrophy

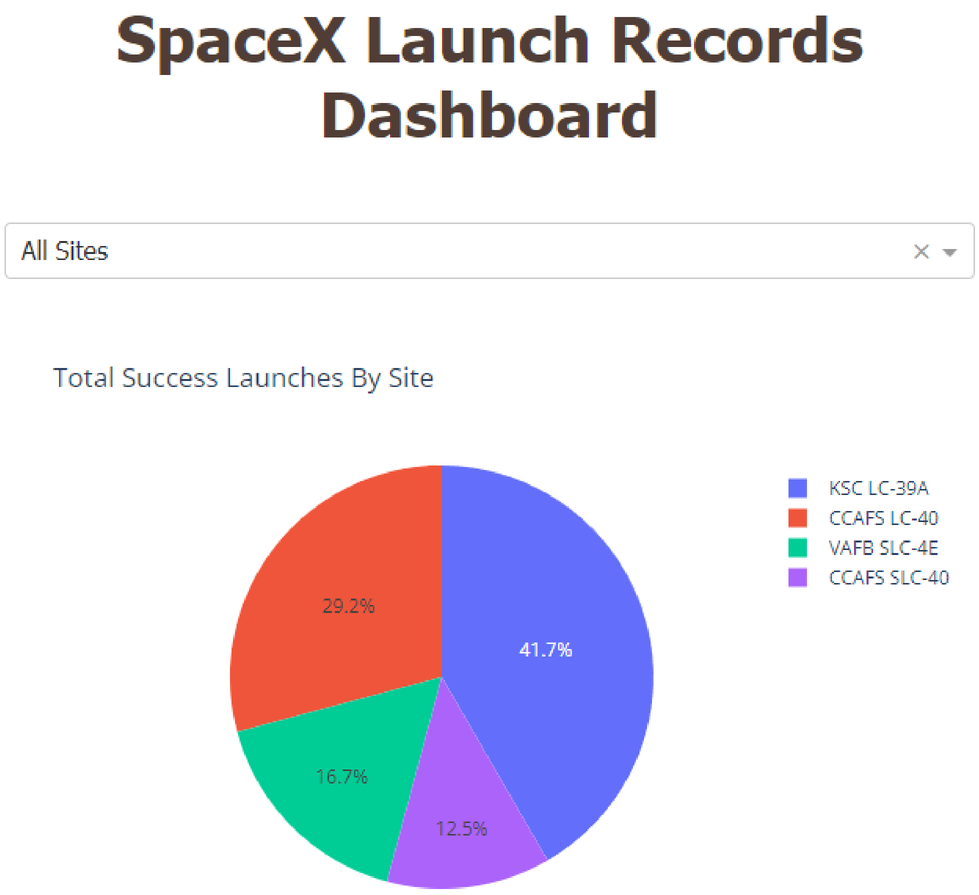


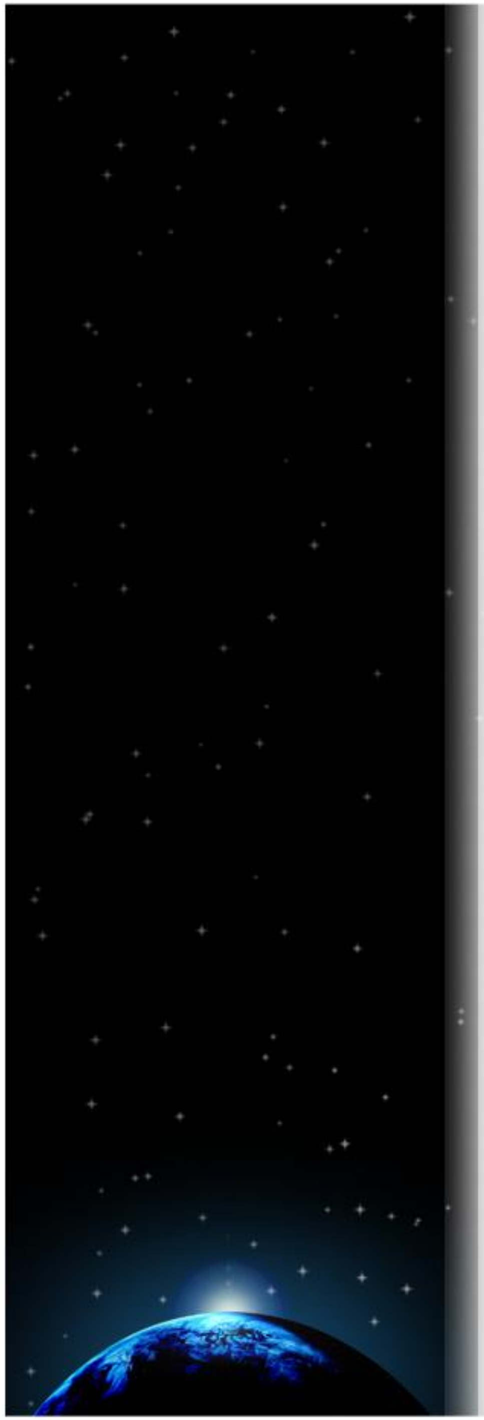
Section 4

Build a Dashboard with Plotly Dash

Total Success Launches by Site

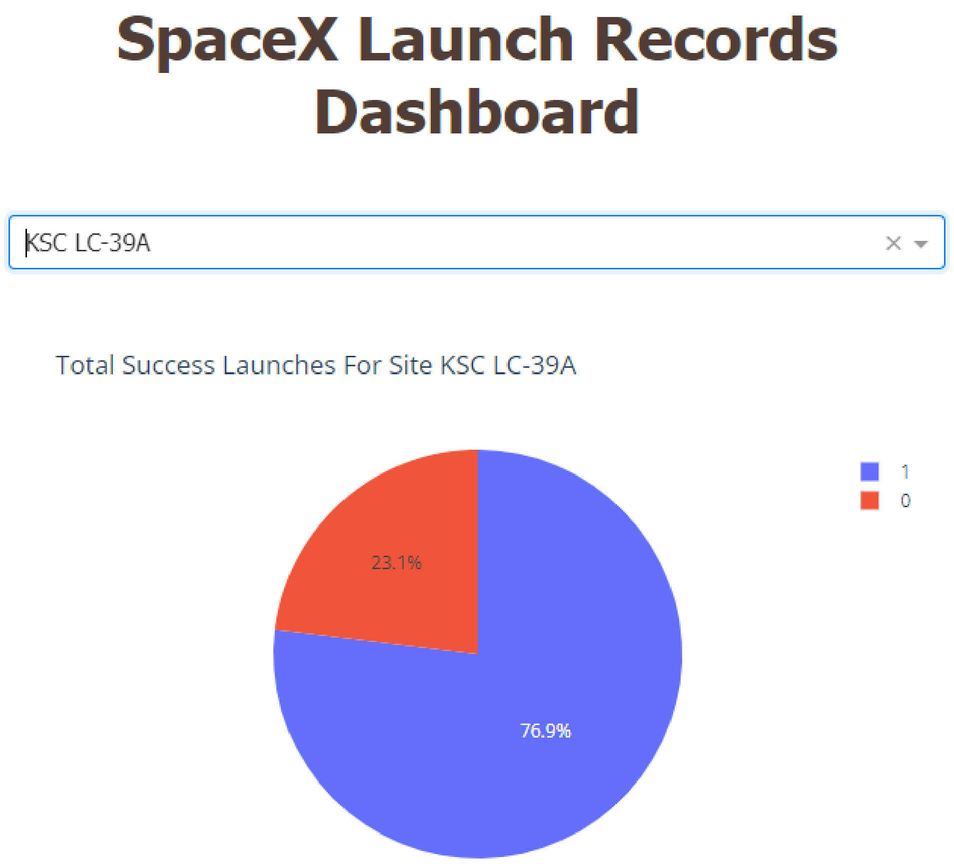
- KSC LC-39A has most successful launch rate
- CCAFS SLC-40 has least successful launch rate



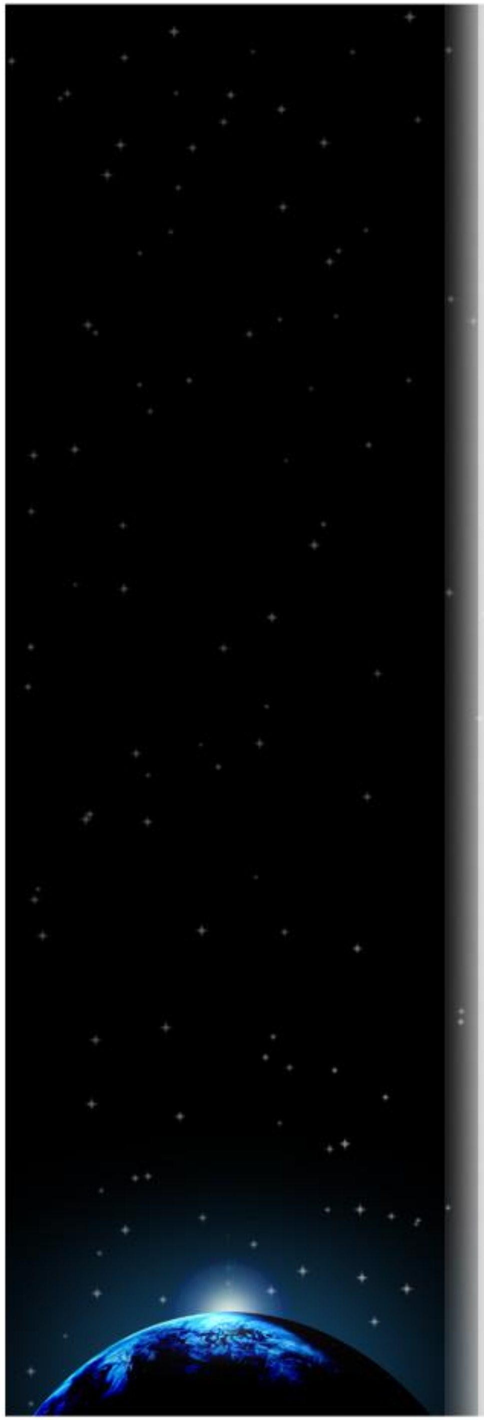


KSC LC-39A Launch Rate

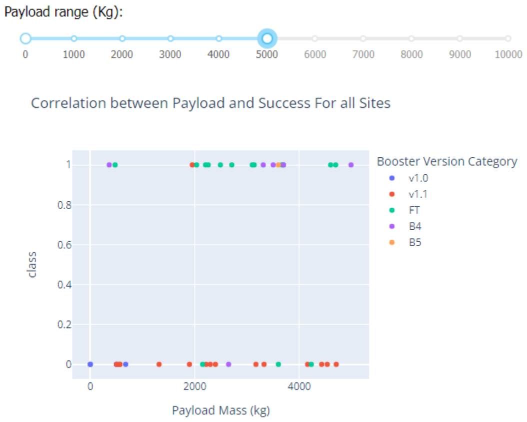
- 76.9% of the launches at KSC LC-39A were successful

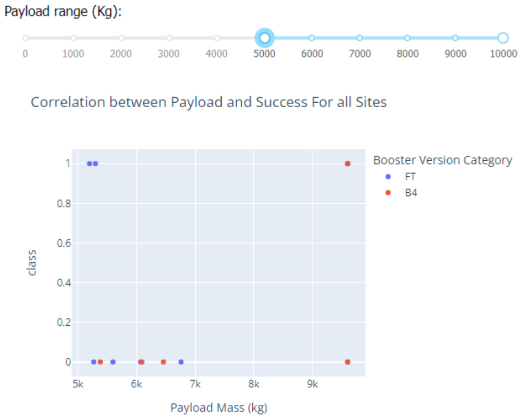


Payload vs. Launch Outcome



- Most Launches with a payload greater 5000 kg failed
- Below 5000 kg there is a 50/50 chance to fail

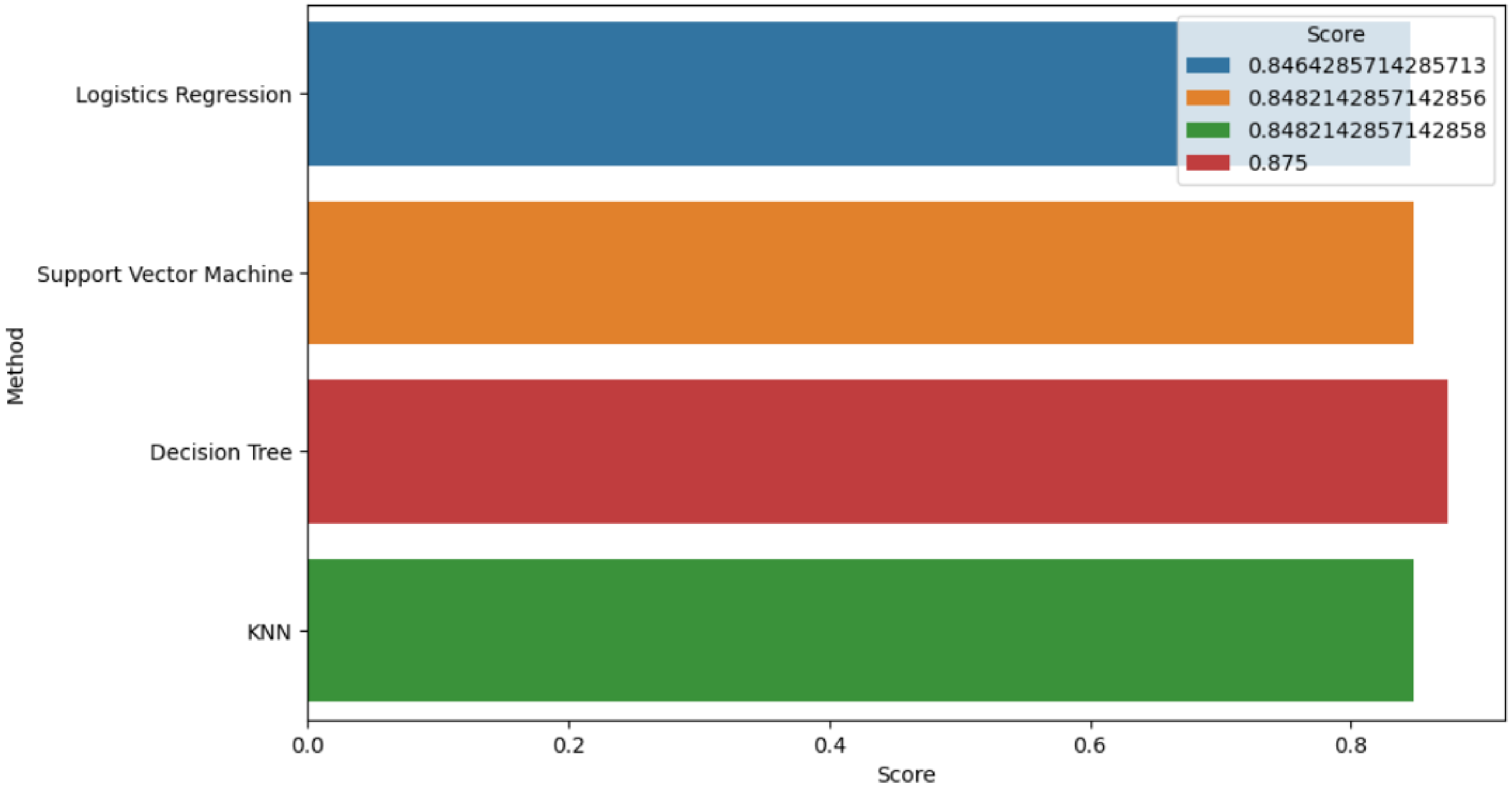




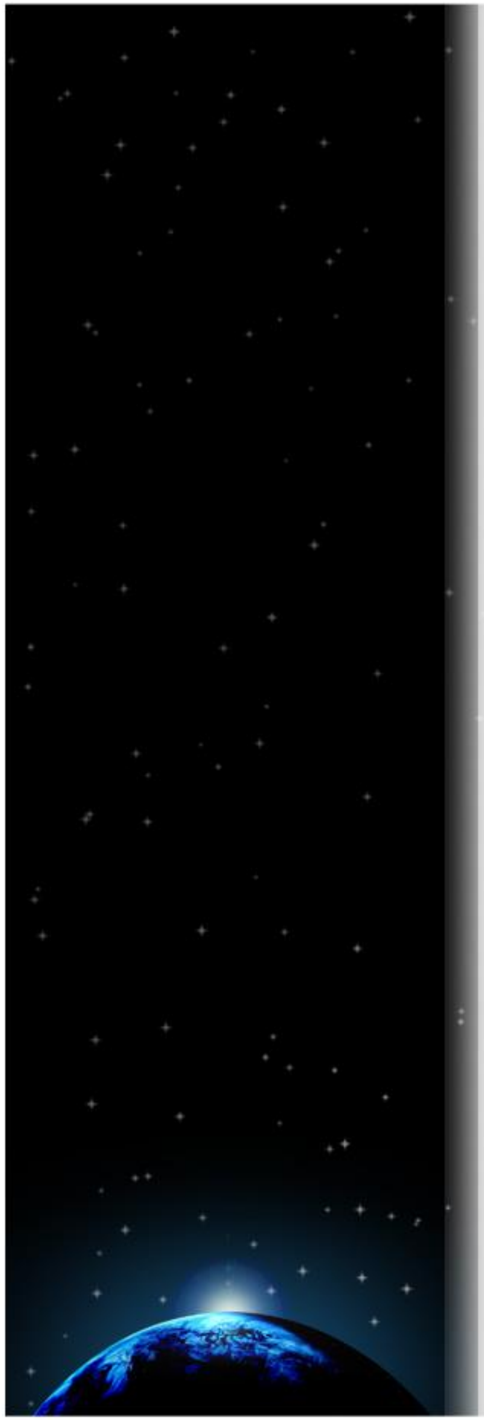
Section 5

Predictive Analysis (Classification)

Classification Accuracy

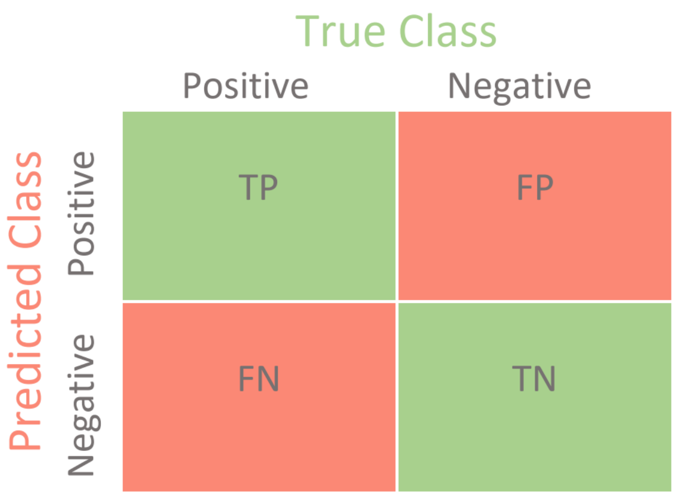


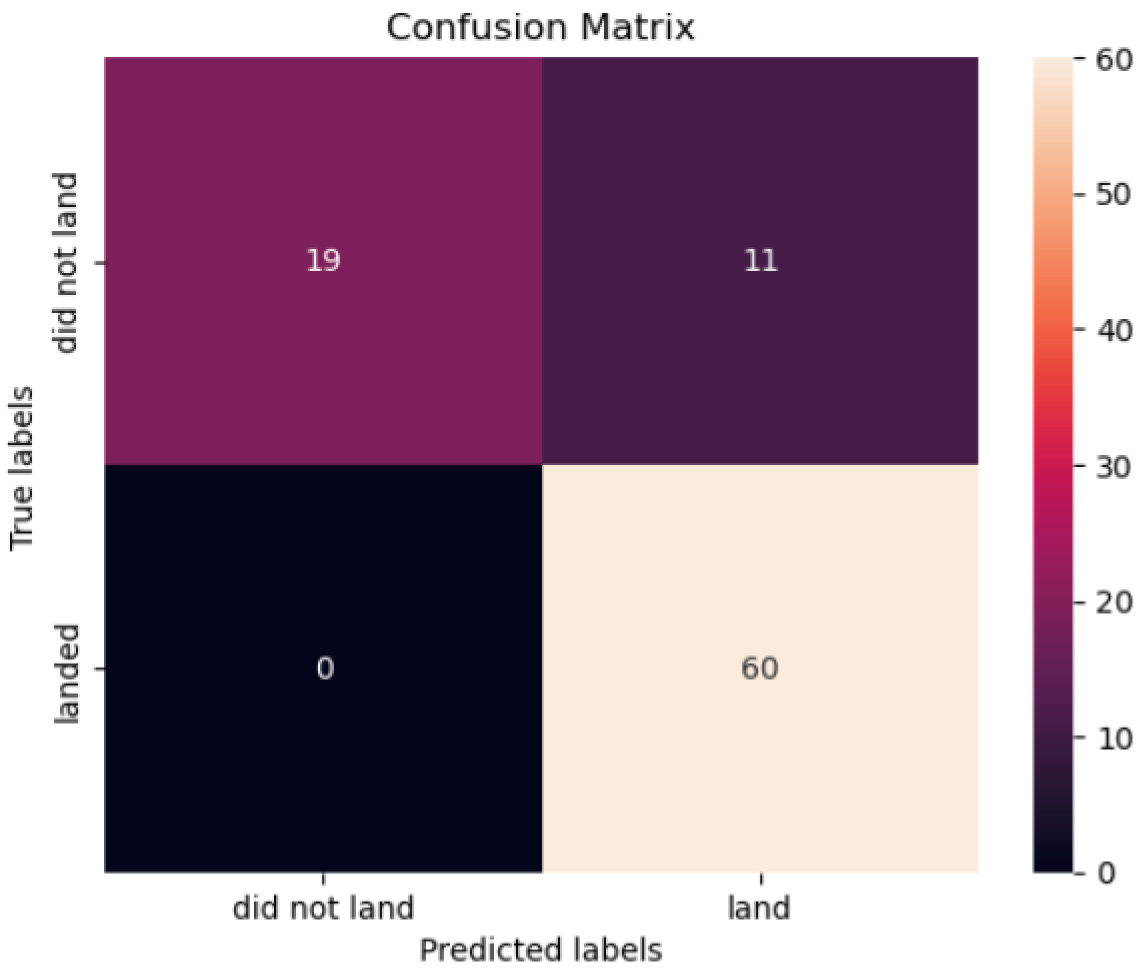
Decision Tree has the highest classification accuracy



Confusion Matrix

- Using real data instead of test data we see that there are some False-Positive errors





Conclusions

1. The Decision Tree Model stands out as the most suitable algorithm for this particular dataset.
2. Launches carrying lighter payloads tend to yield superior outcomes compared to those with heavier payloads.
3. The majority of launch facilities are situated near the Equator, and all of them are located very close to the coastline.
4. Over the years, there has been a notable increase in the success rate of launches.
5. Among all the launch sites, KSC LC-39A boasts the highest rate of successful launches.
6. Orbits such as ES-L1, GEO, HEO, and SSO exhibit a perfect 100% success rate.

Thank you!