Winning Space Race with Data Science

Stefan Civelek
26.06.2025

Outline

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

Executive Summary

Summary of methodologies

- Data Collection
- Data wrangling
- EDA visualisation
- EDA SQL queries
- Interactive Map with Folium
- Dashboard with Plotly
- Data Classification

Summary of all results

- EDA Results
- Interactive analytics
- Predictive analysis

Introduction

Project background and context

SpaceX is innovative Aerospace Company. There Falcon 9 rocket launches coast 62 Million, way less then other providers.

Problems you want to find answers

We will try to predict if a SpaceX Falcon 9 rocket will land successfuly

Section 1
Methodology

Methodology

Data collection methodology:

- Web Scrapping from Wikipedia SpaceX / Falcon 9
- SpaceX Rest API

Perform data wrangling

- Data Cleaning : null values and One Hot Enocding data fields

Perform exploratory data analysis (EDA) using visualisation and SQL

Perform interactive visual analytics using Folium and Plotly Dash

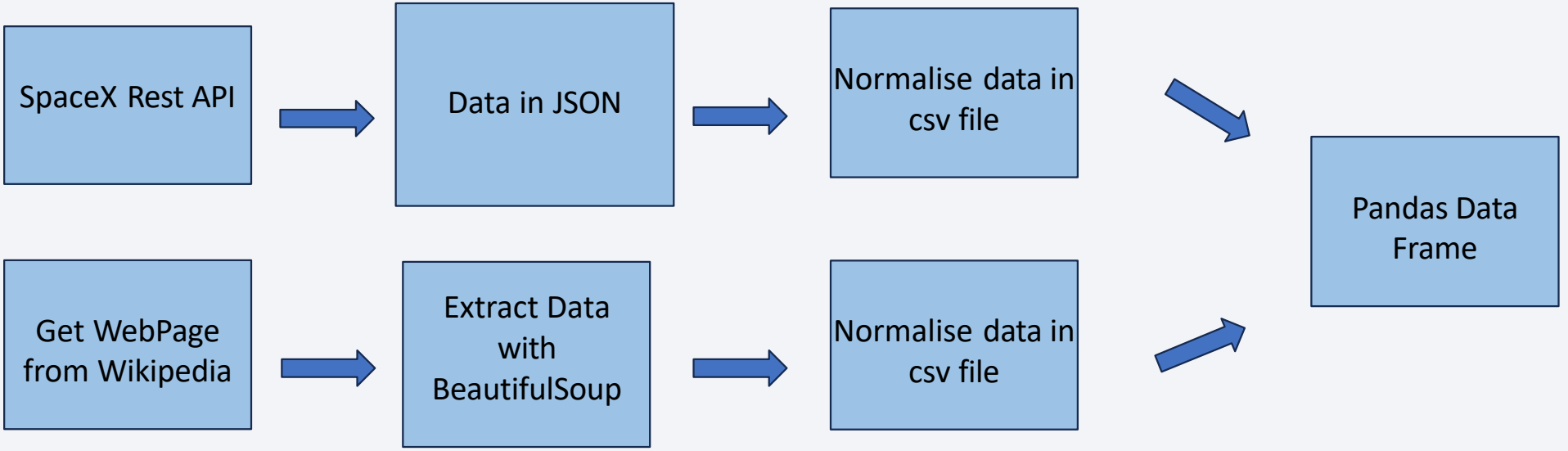
Perform predictive analysis using classification models

- KNN, SVM, Decision Trees

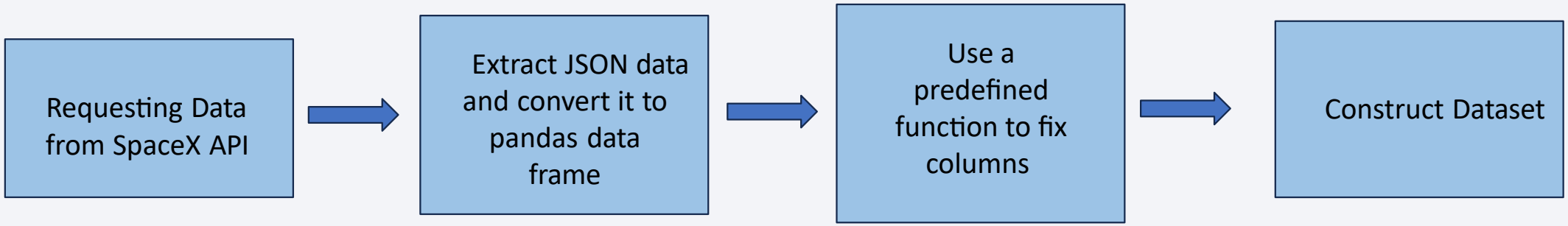
Data Collection

Describe how data sets were collected

- SpaceX Rest API: SpaceX launch data (Rocket Type, payload, landing & launch details,)
- Web Scraping from Wikipedia with BeautifulSoup

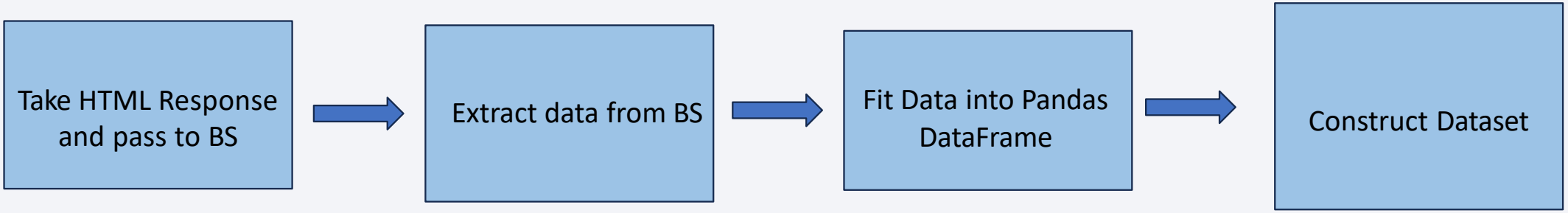


Data collection - SpaceX REST API



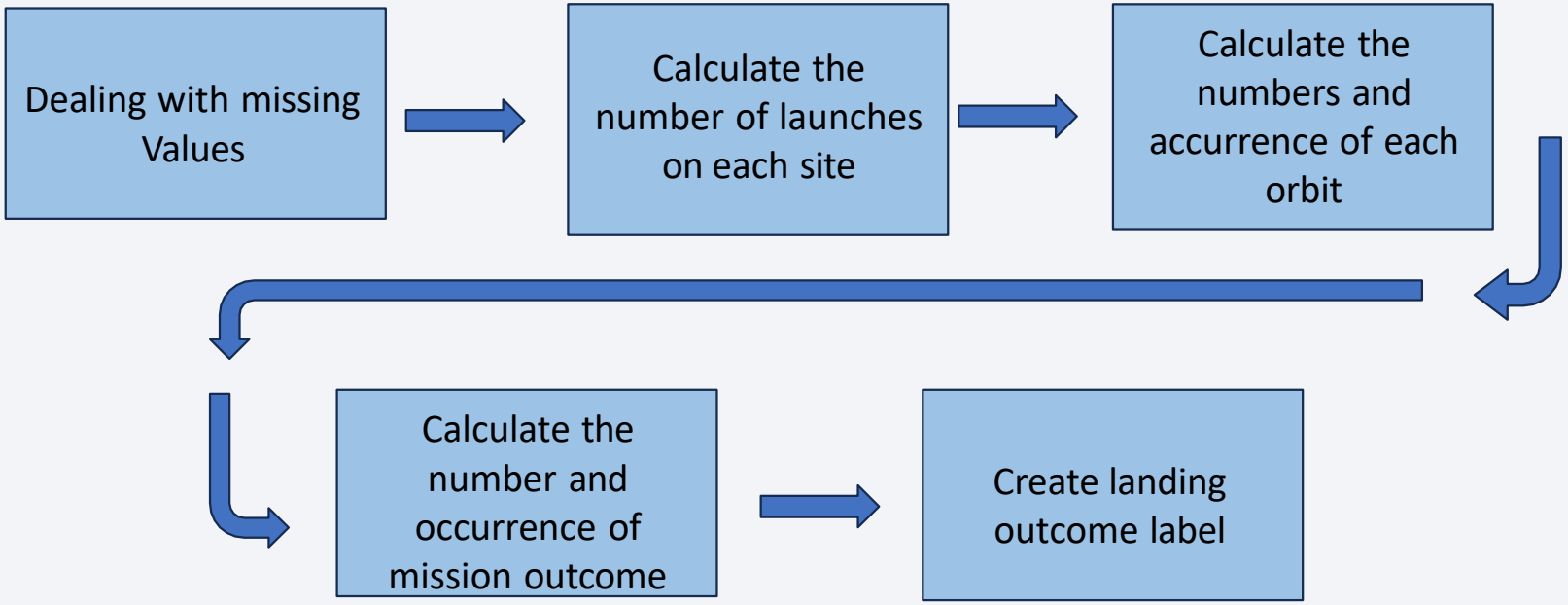
https://github.com/StefanC05/DataScience_Capstone/blob/main/1_1_jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Web Scraping



https://github.com/StefanC05/DataScience_Capstone/blob/main/1_2_jupyter-labs-webscraping.ipynb

Data Wrangling



EDA with Data Visualisation

Used Chart:

- Scatter plot: to display the relationships between different variables.
- Bar charts: for quantities of different types of values.
- Visualising development over time with line charts

https://github.com/StefanC05/DataScience_Capstone/blob/main/2_2_jupyter-labs-eda-visualization.ipynb

EDA with SQL

SQL queries:

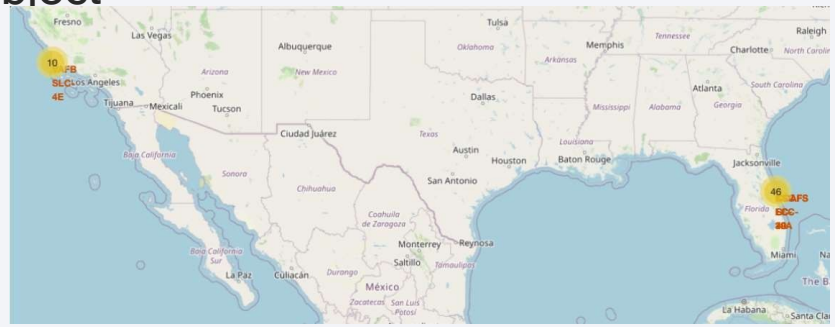
- Display of unique launch sites
- Display 5 records for launch sites begin with “CCA”
- Show total payload by boosters by “NASA (CRS)”
- Display average payload for boosters version “F9 c1.1”
- Date of first successful landing outcomes
- Boosters with payload between 4000 and 6000kg who successful land on ground pad
- Total number of successful and failure mission outcome
- Versions and launch site for each month in the year 2017
- Ranking the count of successful landing outcome between 04.06.2010 and 20.03.2017

https://github.com/StefanC05/DataScience_Capstone/blob/main/2_1_jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Summarise what map object:

- Marker: Geo location of the launch side by latitude and longitude
- Cluster: Show a group of markers
- Circles: Show a single location
- Lines: Show distance between launch side an areal object



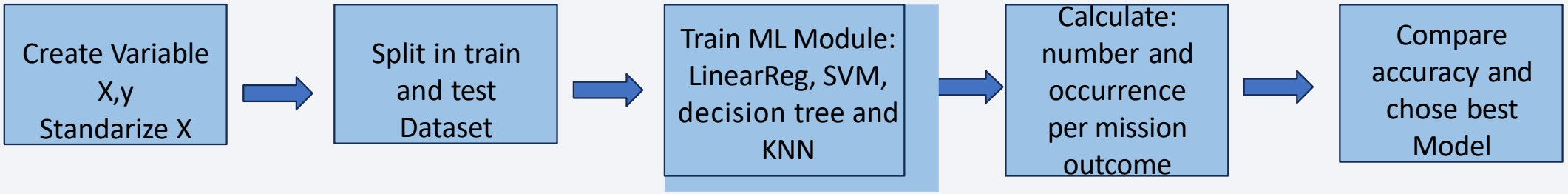
https://github.com/StefanC05/DataScience_Capstone/blob/main/3_1_lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

Summarise what plots/graphs from the Dash

- Bar Chart: difference categories of Rocket type
- Line Chart: Time series changes
- Pie Chart: Percentage of failure an successful landing
- Tree: Shows complex relationship of variable
- Map: Show variable of state on a map

Predictive Analysis (Classification)

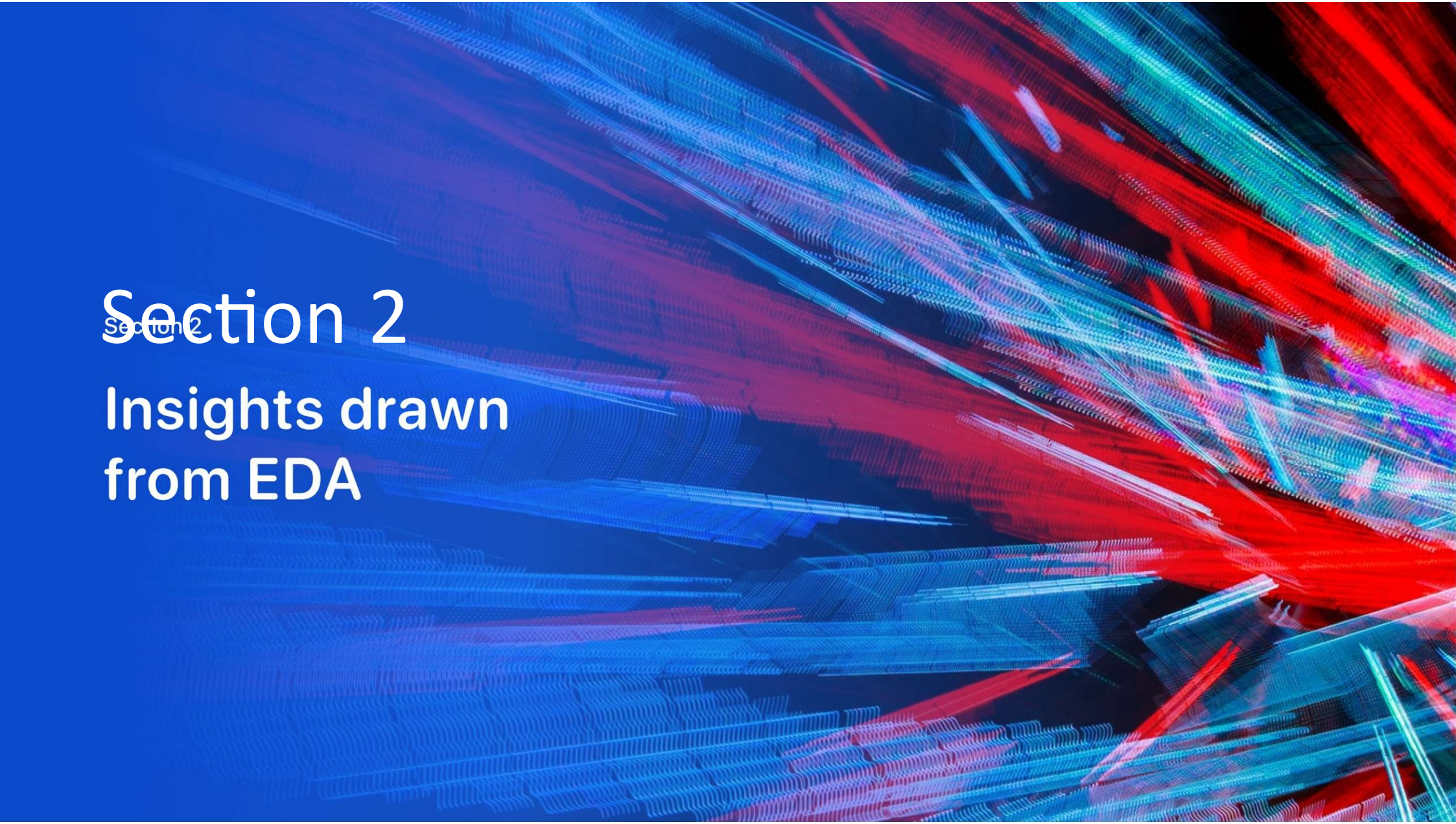


https://github.com/StefanC05/DataScience_Capstone/blob/main/3_1_lab_jupyter_launch_site_location.ipynb

Results

Exploratory data analysis results:

- Low weighted payloads perform better than heavier payload
- The success rate for SpaceX launches is directly proportional time in year they will eventually perfect the launches
- KSC LC 39A had the most successful launches from all the sites
- Orbit GEO, HEO, SSO, ES L1 has the best Success Rate

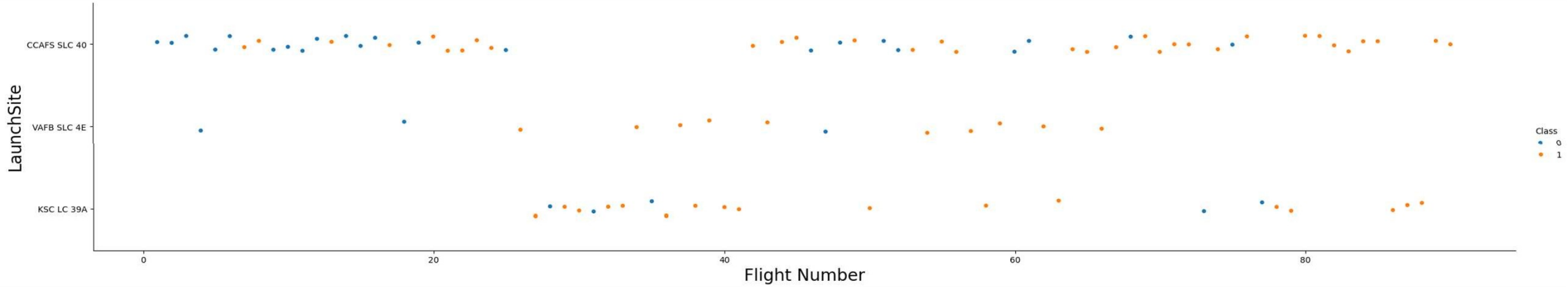


Section 2

Insights drawn from EDA

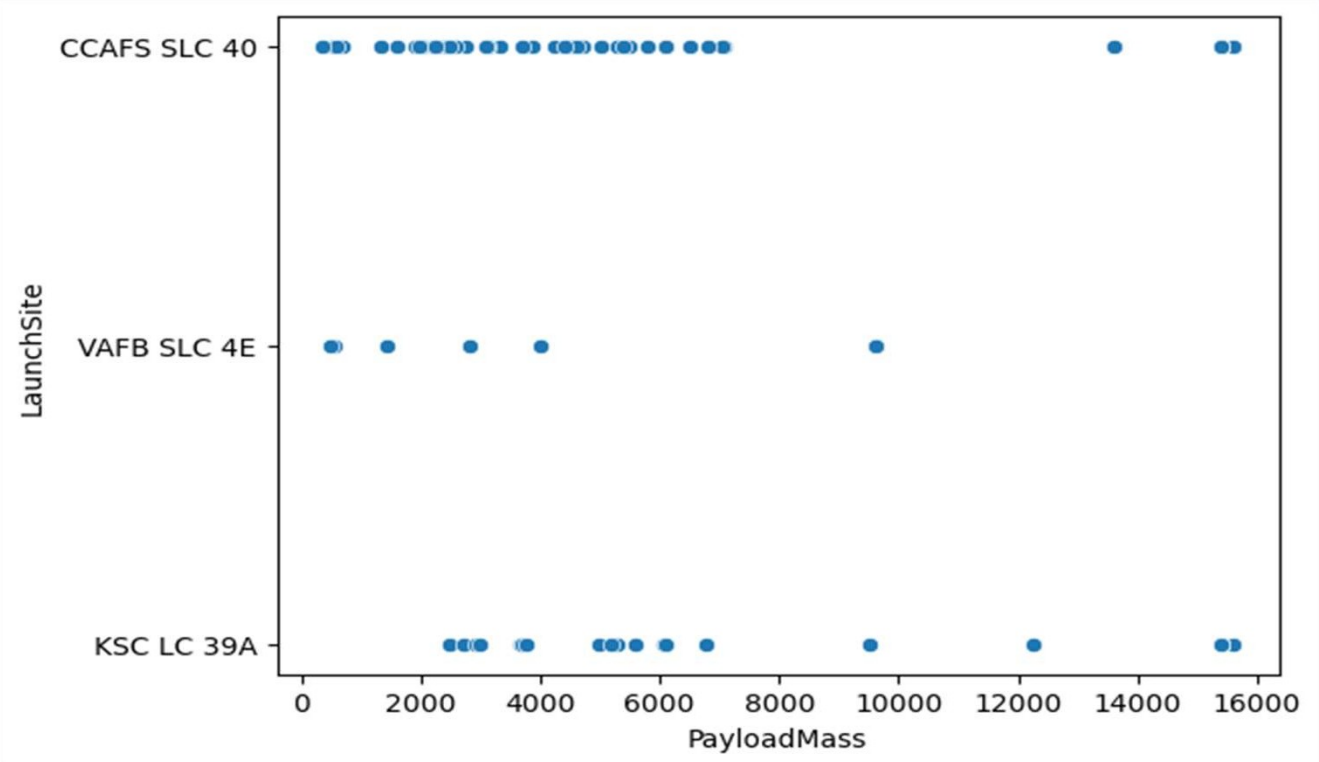
Flight Number vs. Launch Site

- Point = Launch site start (Blue = fails, red = success l)
- Most flight start from CC AFS SLC 40



Payload vs. Launch Site

- KSC LC 39A ha most start with payload over 8000 KG
- CCAFS SLC 40 has payloads from 250 - 16000



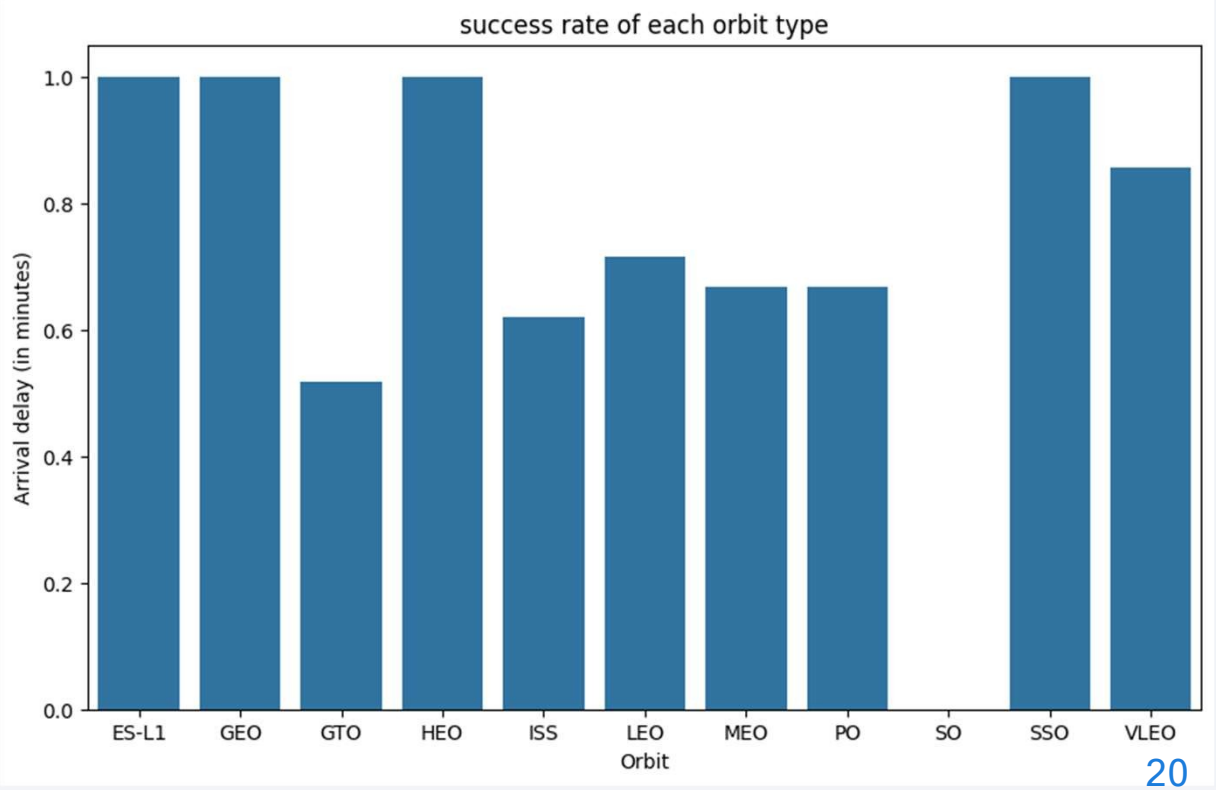
Success Rate vs. Orbit Type

- Most success full Orbit type are :

  ES-L1, GEO, HEO, SSO

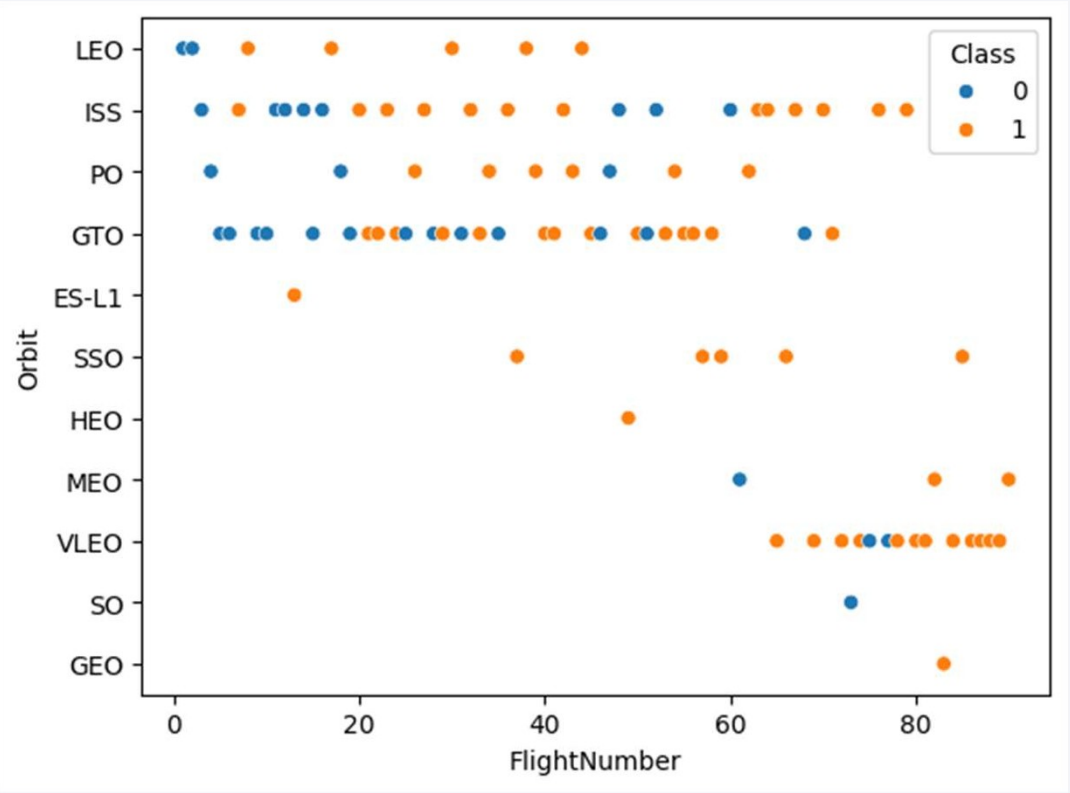
- Most failure Orbit type are

  GRO and ISS



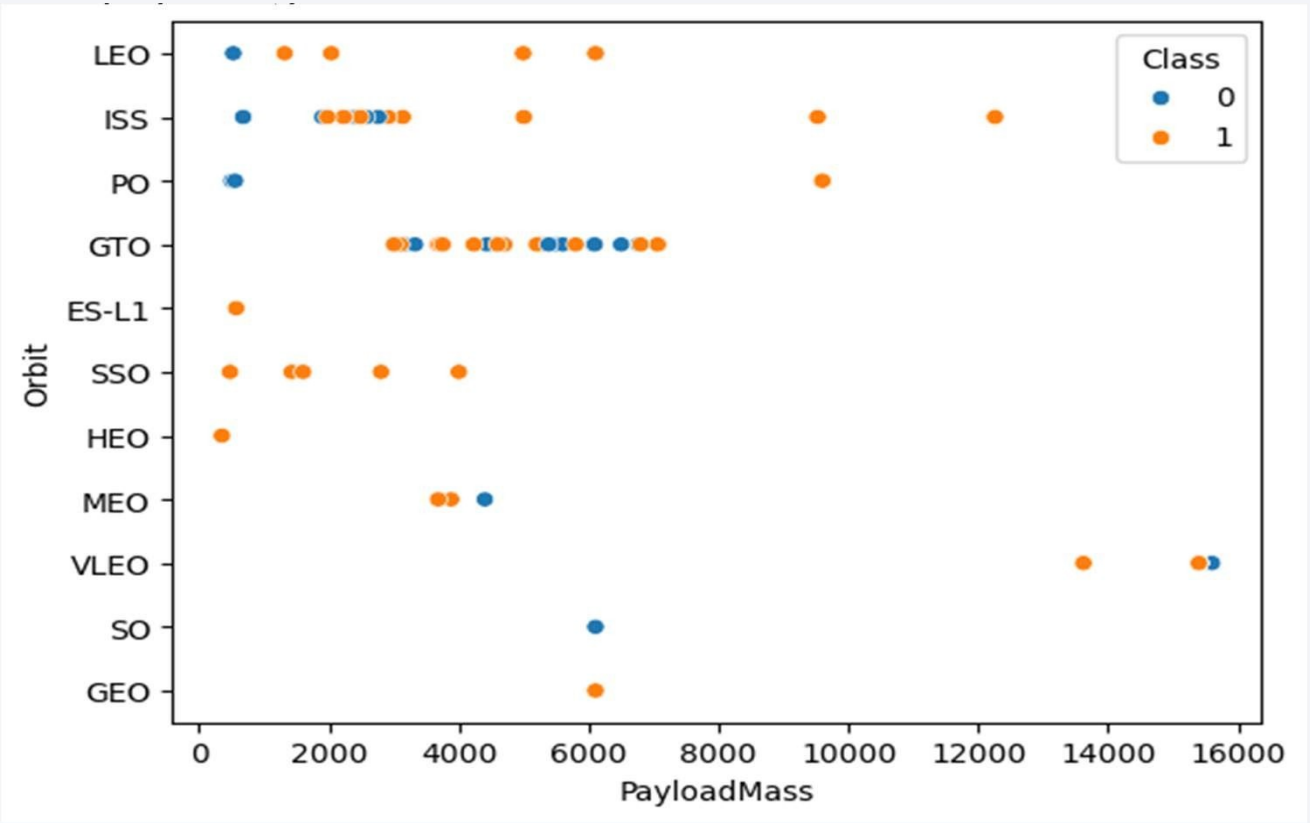
Flight Number vs. Orbit Type

- On ISS the start are continuos over the hole time
- Starts from VLEO: start and end of Test flight



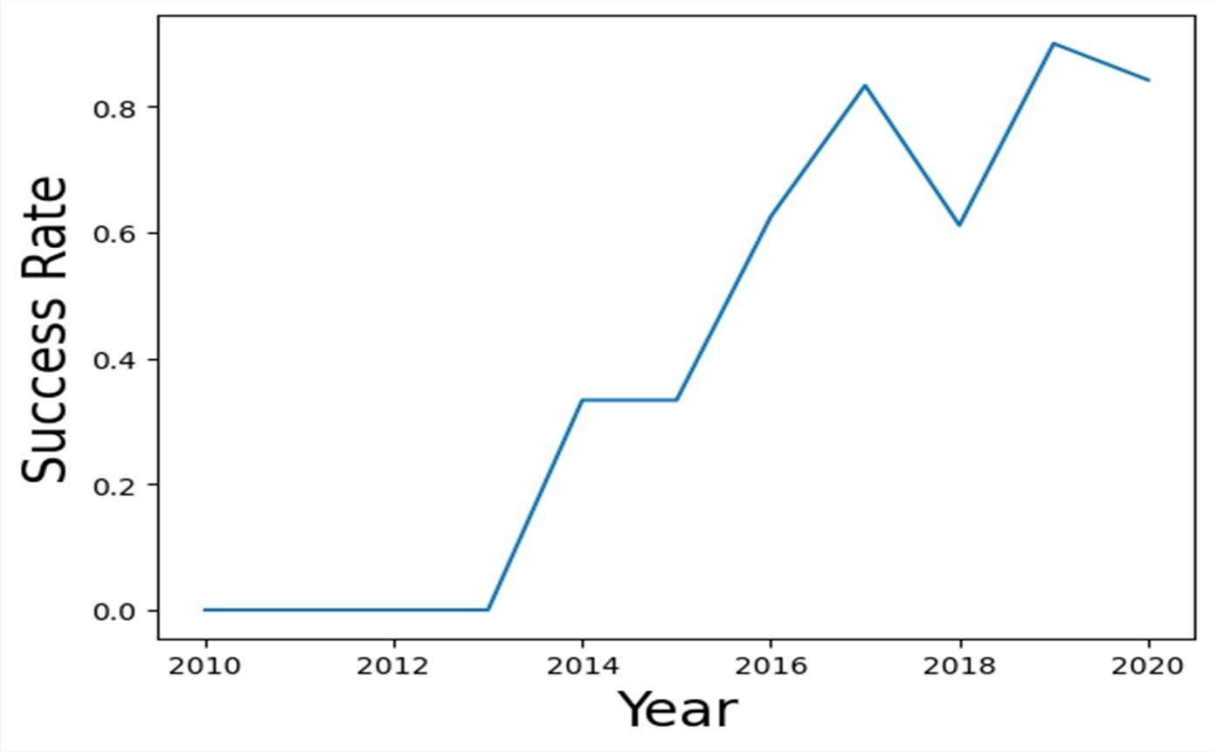
Payload vs. Orbit Type

- Highest Payload are go to VLEO
- And the most divers payload goes to ISS



Launch Success Yearly Trend

We clearly see that the Success raise over time



All Launch Site Names

SpaceX use 4 different Launch Site over the year to start there Rockets.

Out[21]:

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

Launch Site Names Begin with 'CCA'

The first 5 Dataset with a Launch_Site how begin with a CCA

Out[56]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (paracl
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 (paracl
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No att
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No att
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No att

Total Payload Mass

The complete Payload of “NASA (CRS) is over 45 t

Out[57]:

SUM("PAYLOAD_MASS__KG_")
45596

Average Payload Mass by F9 v1.1

The average Payload of Booster F9 v1.1 is nearly 3 t

Out[58]:

AVG("PAYLOAD_MASS__KG_")
2928.4

First Successful Ground Landing Date

The first successful Ground Landing was on 04.06.2010

Out[72]:

Date
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

The following 4 Booster versions carrying a payload between 4 and 6 tons

Out[38]:

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

Total Number of Successful and Failure Mission Outcomes

One hundred Missions were successful and only 1 was a failure

Out[69]:

Count("Mission_Outcome")
100

Out[70]:

Count("Mission_Outcome")
1

Boosters Carried Maximum Payload

Twelve Booster versions have carried the maximum payload

Out[84]:

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

Only 2 missions failed in 2015

Out[59]:

month	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

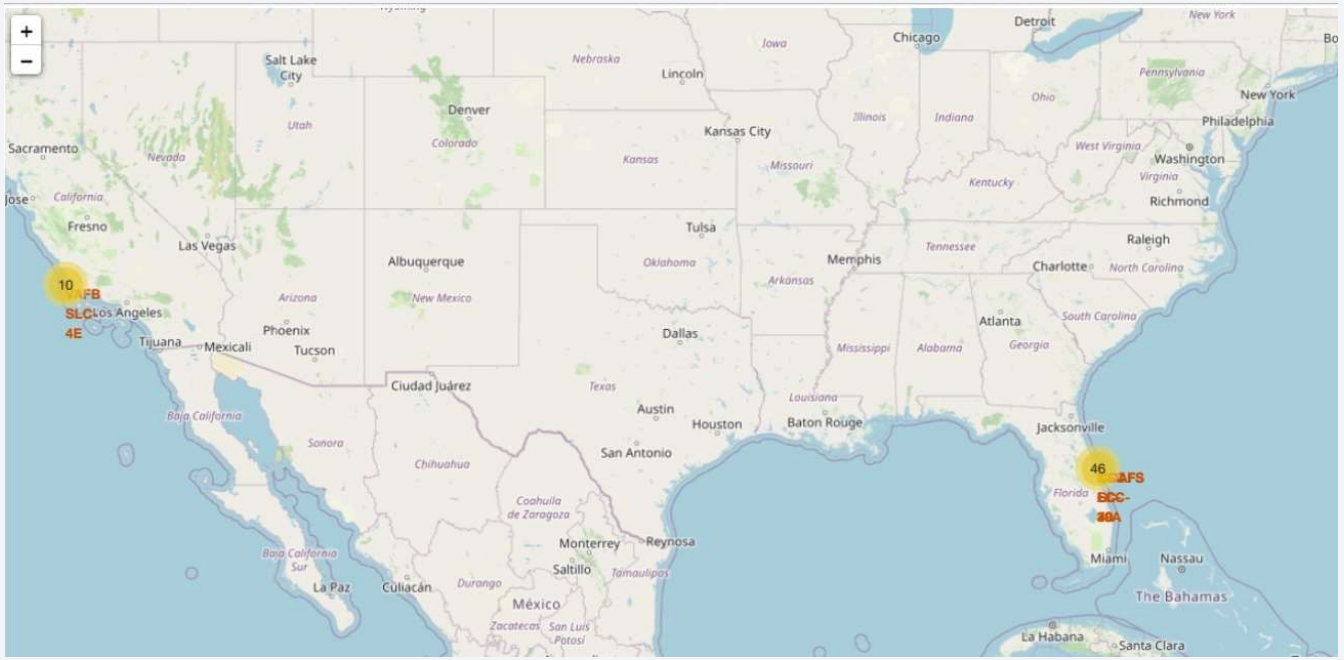
Rank the count of landing outcomes ,in descending order, between 04.06.2010 - 20.03.2017

Out[81]:

Landing_Outcome	Count
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

Launch Site on Folium map

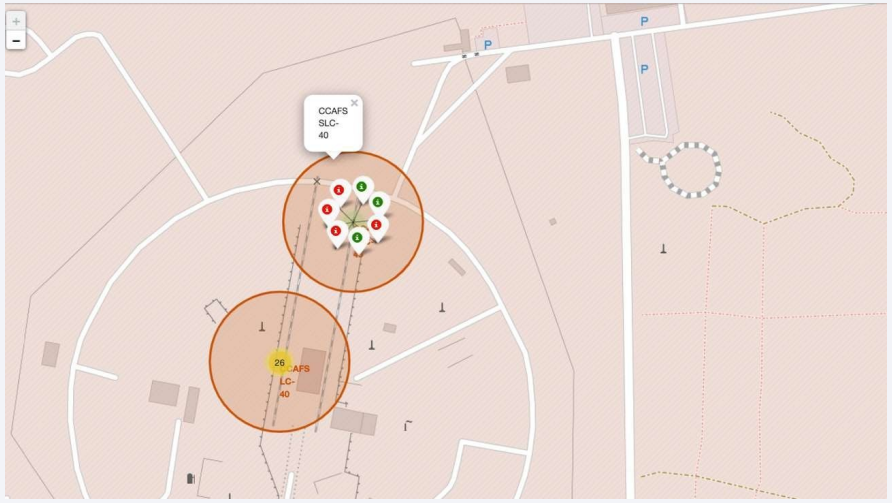


Areas near the equator are grantee easier and fuel saving launch sites

The transportation base chosen are important.

Success of launch sites

CC AFS SLC - 40

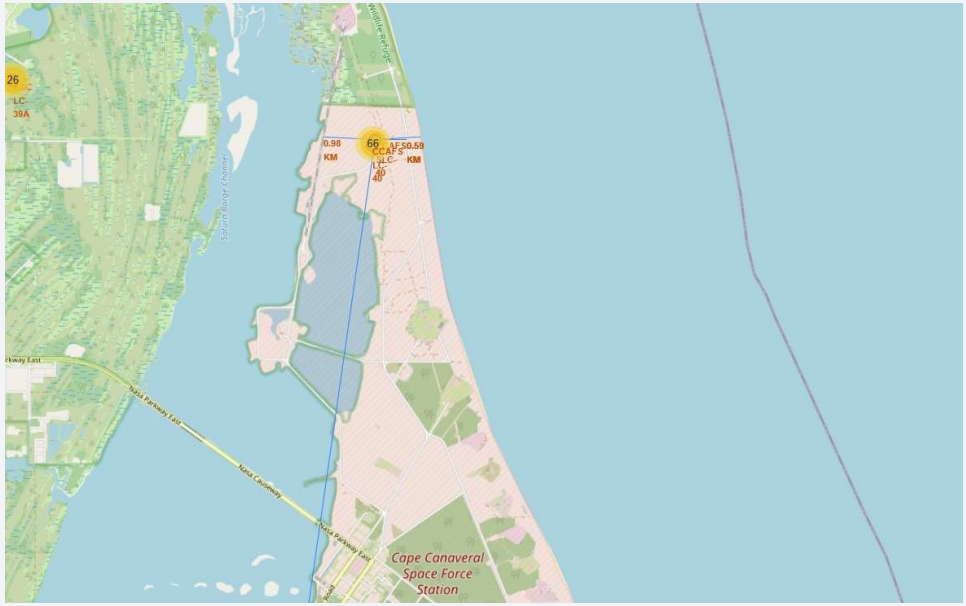


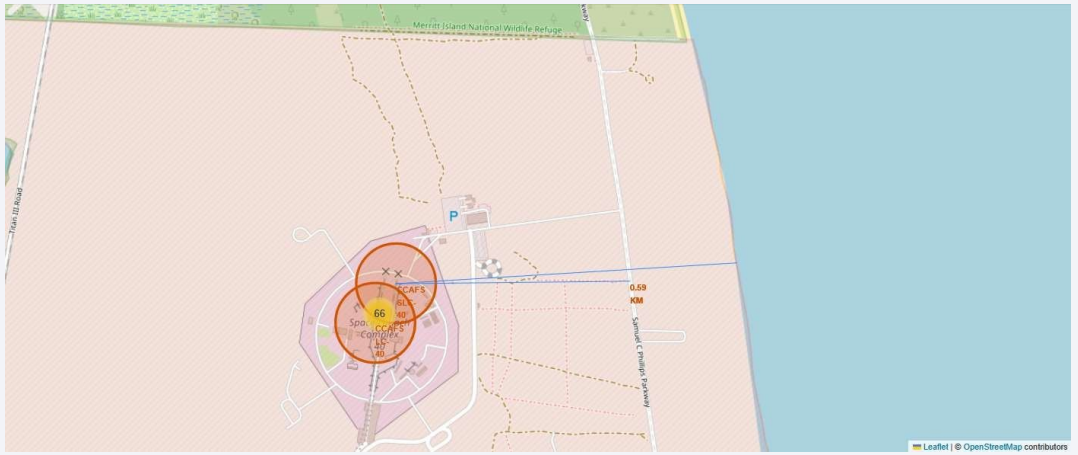
KSC LC - 39 A

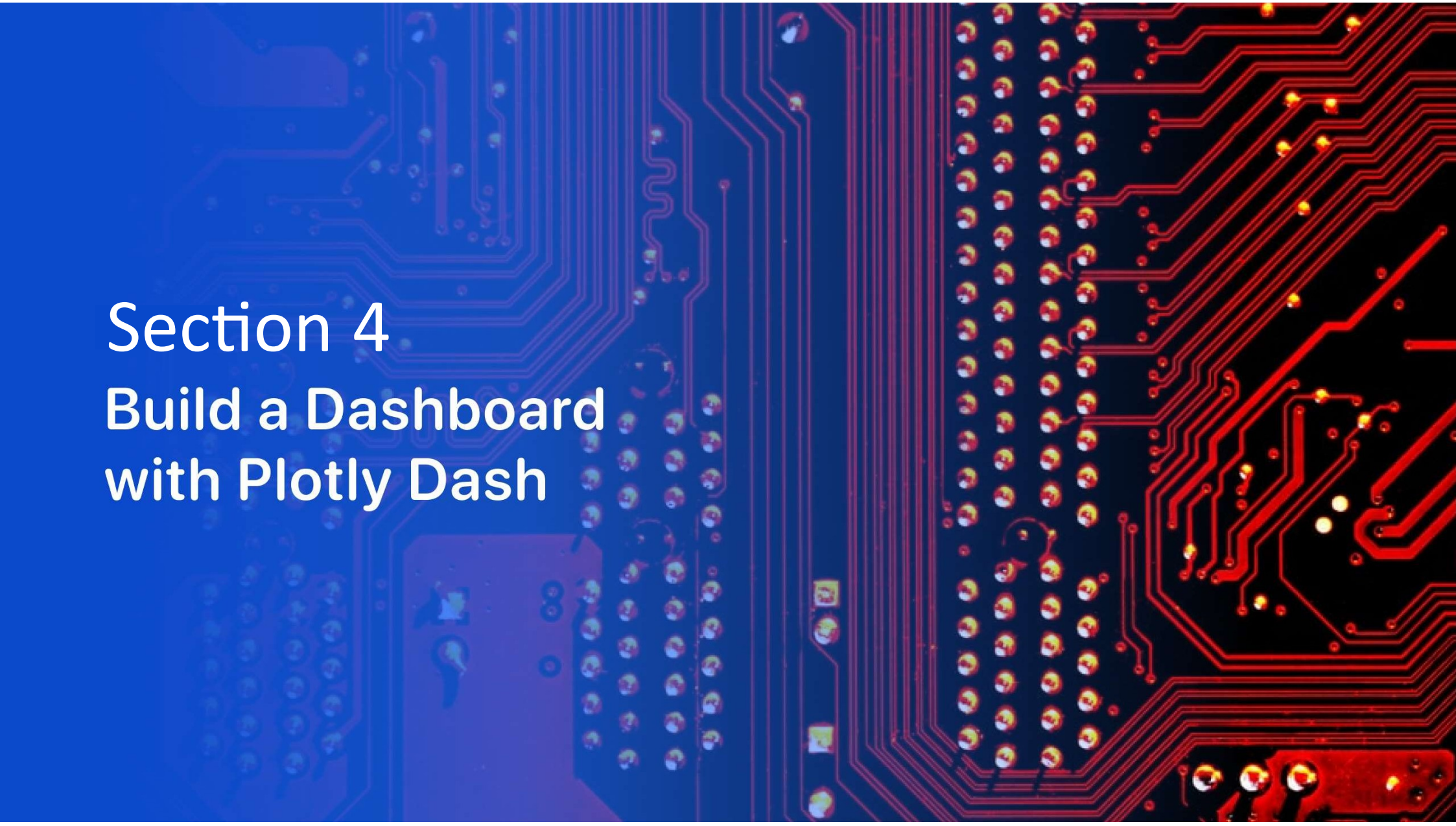
Coloured icons to show the rate of success.

Folium Map : Distance

This map show distances from launch site to coast point, railway tracks and other importance object

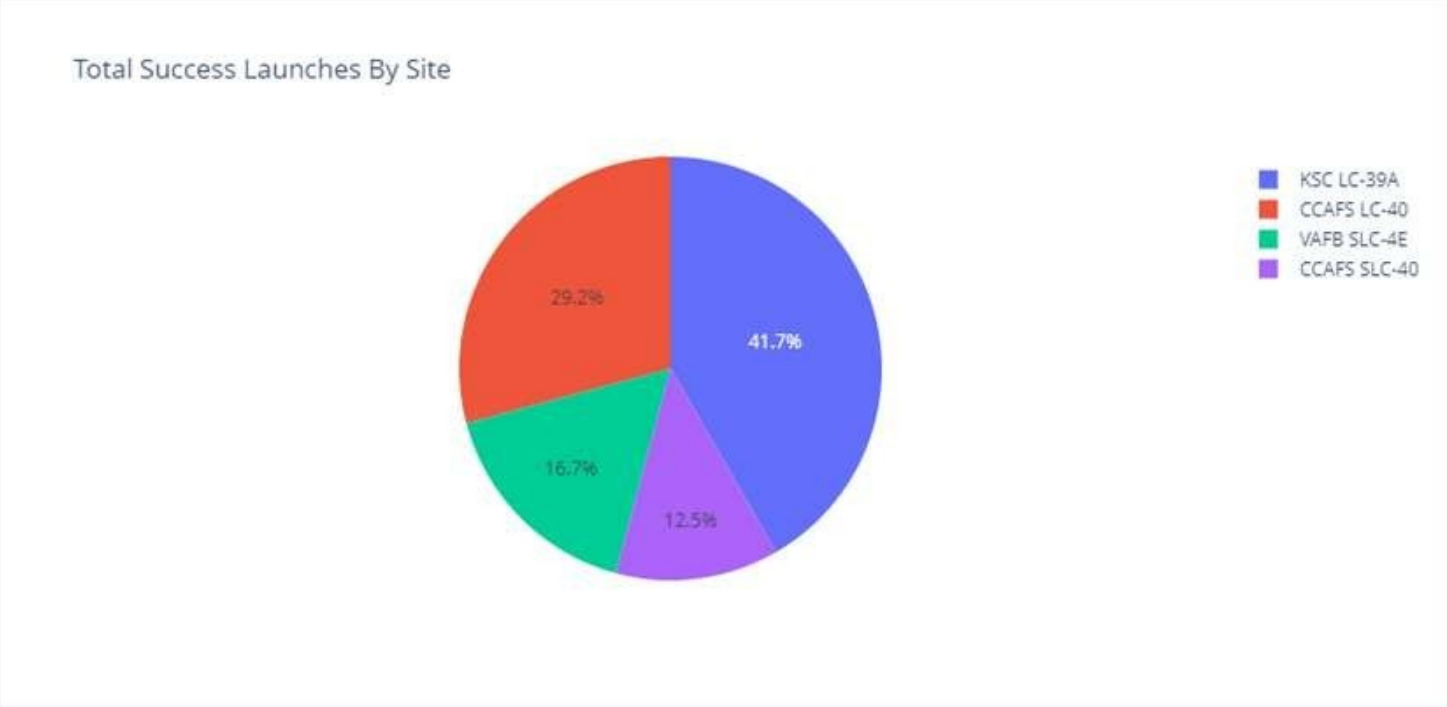





Section 4
Build a Dashboard
with Plotly Dash

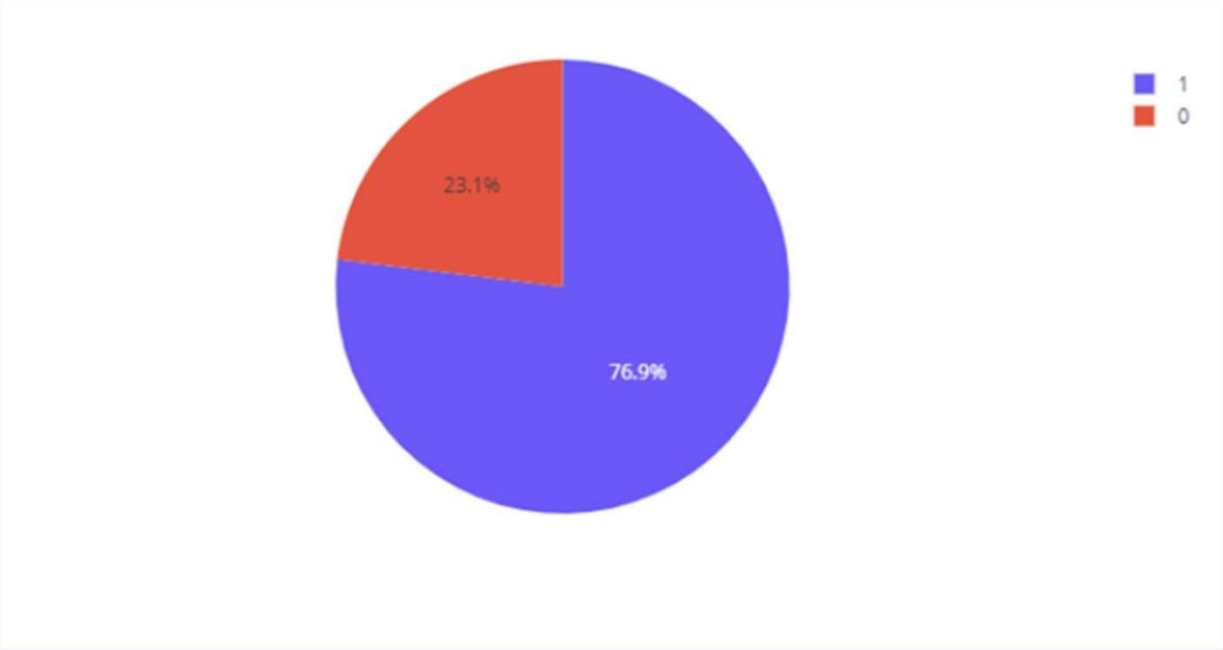
Most successful launch site

KSC LC-39A is the most successful launch site

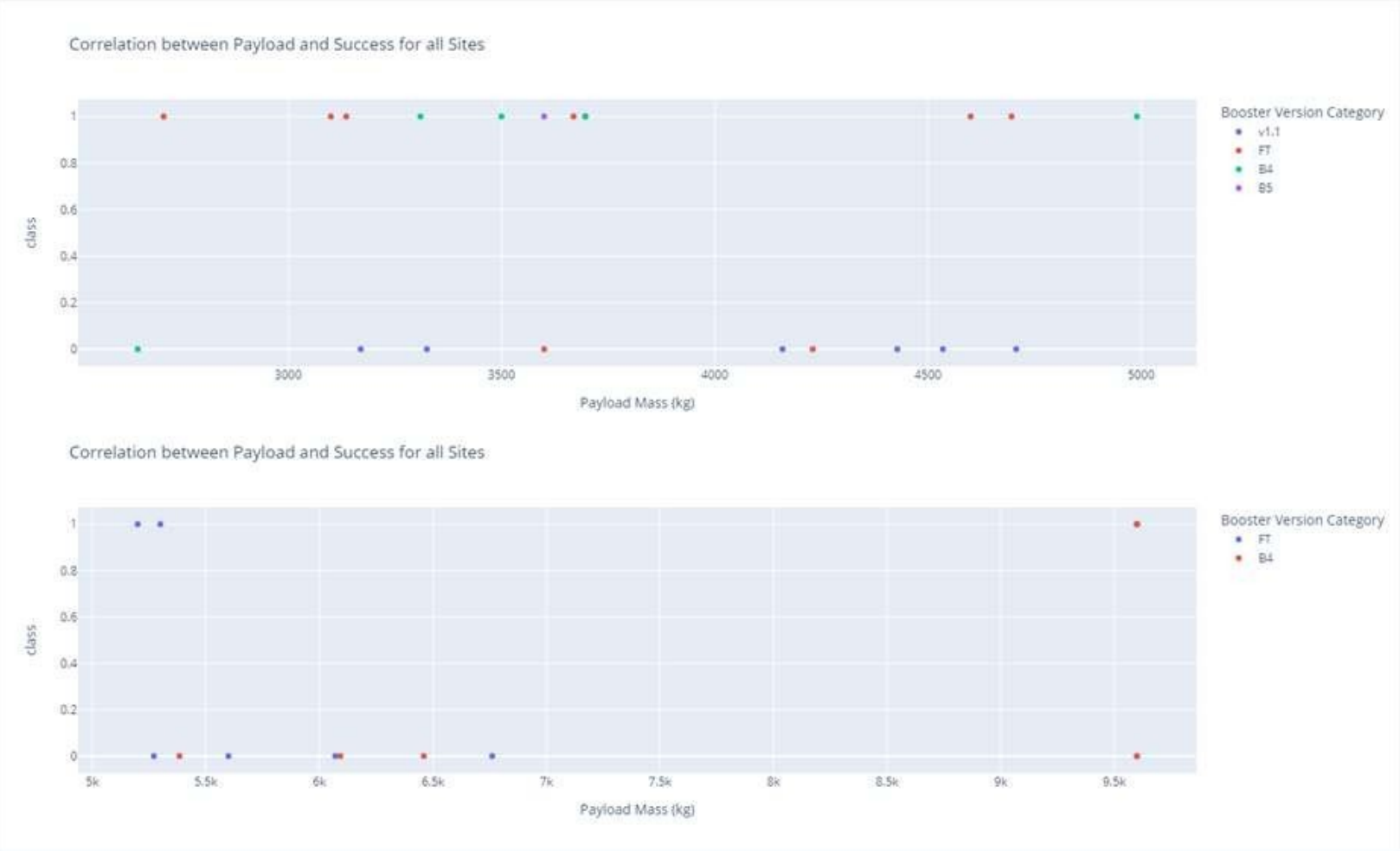


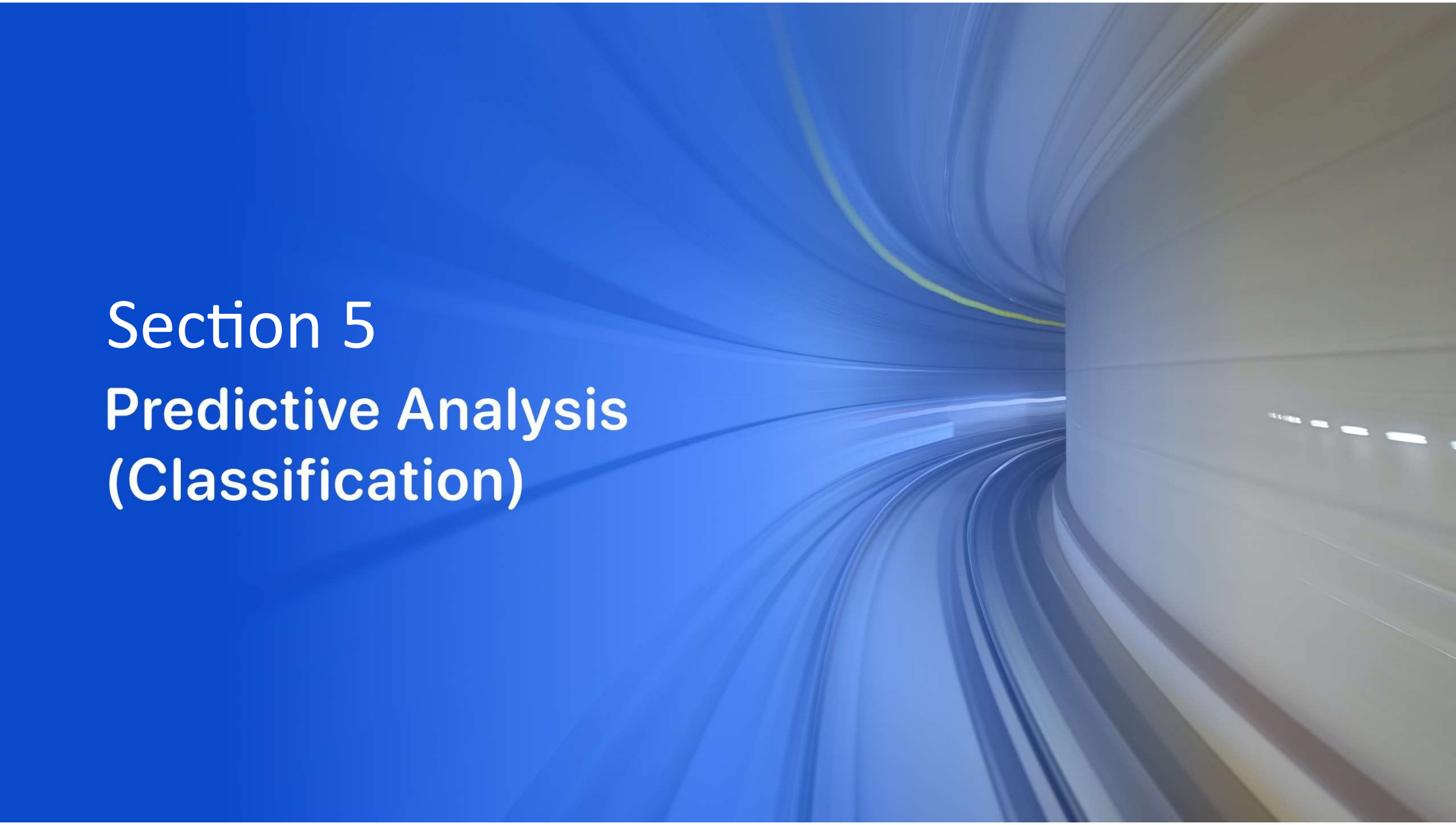
KSC LV-39 A success ratio

76.9% of all launches are Successful and only 23.1 % fails



Payload - Launch outcome



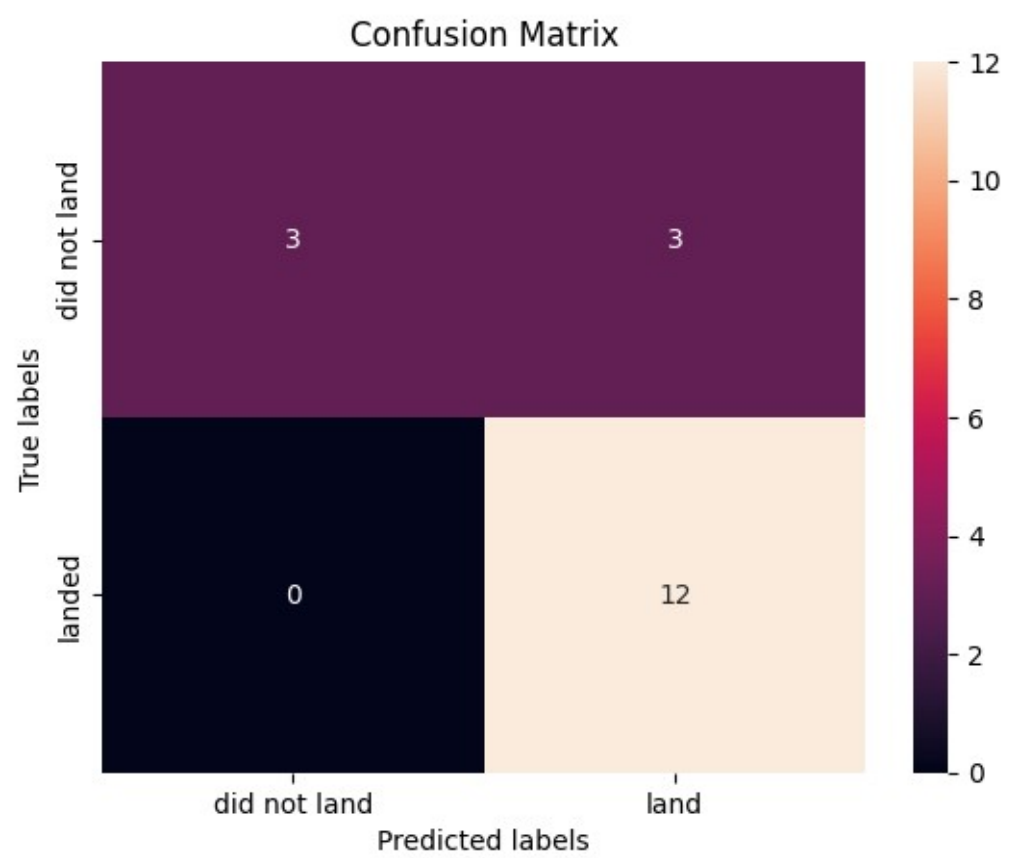


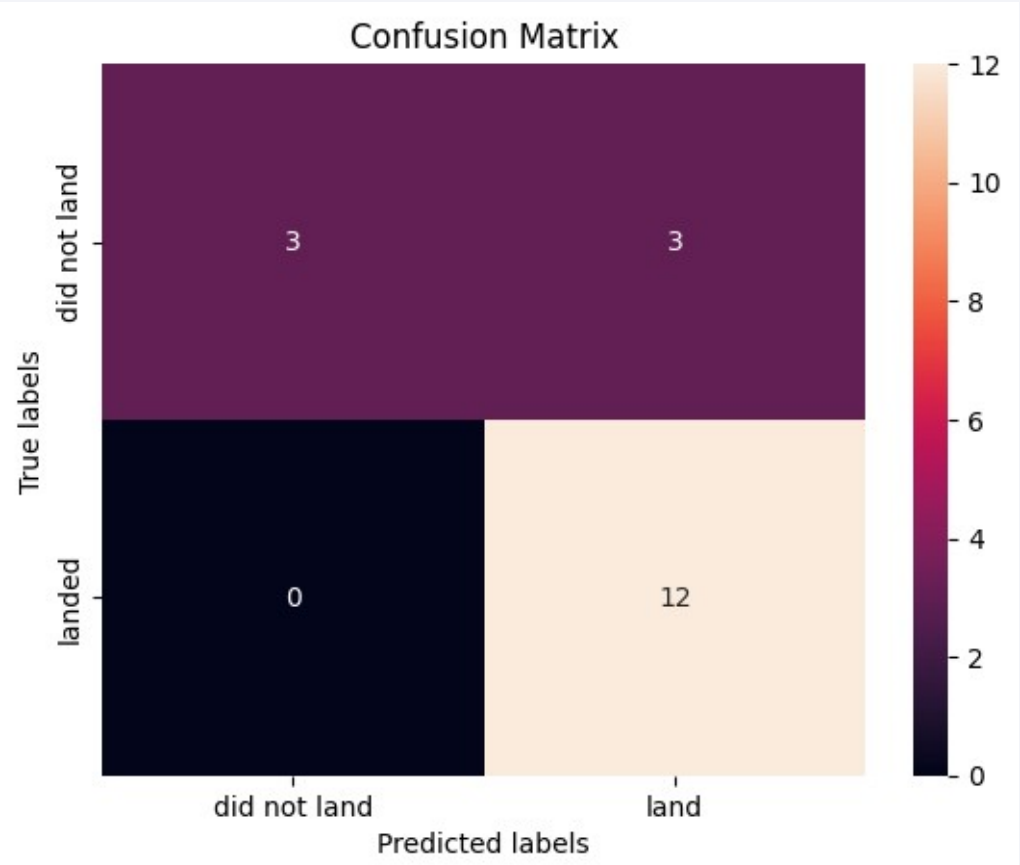
Section 5

Predictive Analysis (Classification)

Confusion Matrix

The ML Model “LogisticRegression” and “SVMclassifier” have both score of .833. Both Models have ascore of .833





Conclusions

- Obits ES-L1, GEO , HEO, SSO has the highest success rate
- Over time the SpaceX increasing the success rate
- The best ML Model is the Decision Tree Classifier
- The most successful launch site is KSC LC 39 A

Appendix

- https://github.com/StefanC05/IBM_DataScience_Capstone/tree/main

Thank you!