IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hiren Kanojia
09-06-2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Data Visualization and SQL
  - Interactive visual analytics using Folium
  - Predictive analysis using classification models
- Summary of all results
  - Correlations variables
  - Build a predictive model

Introduction

- Project background and context
  - The era of commercial spaceflight has arrived, and businesses are making spaceflight accessible to all. Suborbital spaceflights are offered by a different company. SpaceX has achieved the most success. Sending spacecraft to the International Space Station is just one of SpaceX's accomplishments. Satellite Internet access is provided by Starlink, a satellite internet constellation. sending manned spacecraft into orbit. The affordable rocket launches are one factor in SpaceX's ability to do this. On its website, SpaceX promotes Falcon 9 rocket launches for 62 million dollars; other providers charge upwards of 165 million dollars for each launch. A large portion of the savings is due to SpaceX's ability to reuse the first stage. So, if we can figure out whether the first stage will land, we can figure out how much a launch will cost. The Falcon 9 from SpaceX launches like a typical rocket.

- Problems you want to find answers
  - The aim of this project is to determine or predict if the first stage will land or not using the Machine learning Model.

Section 1

Methodology

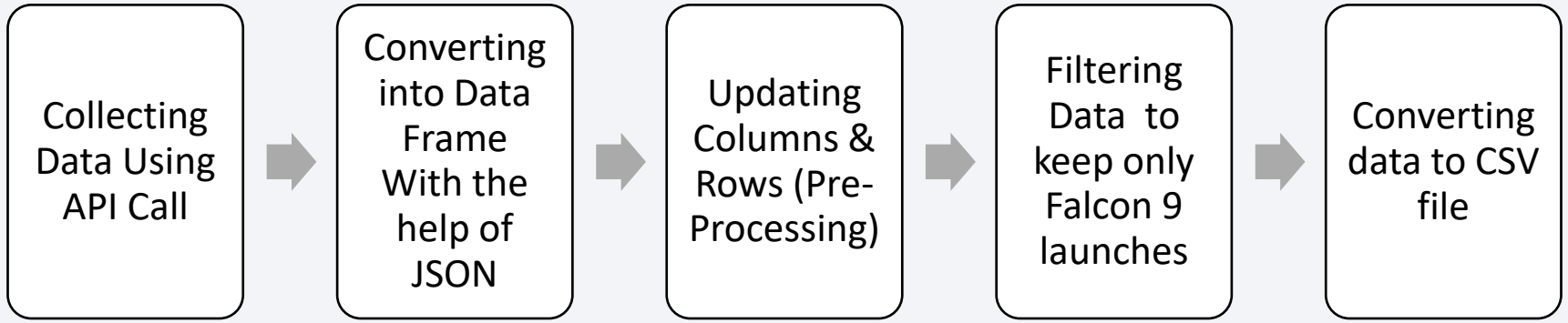
Methodology

Executive Summary

- Data collection methodology:
  - Data Collection is done using Web Scrapping.
- Perform data wrangling
  - Exploratory Data Analysis is done to find the patterns in the data & determine what would be the label for the training supervised model.
- Perform exploratory data analysis (EDA) using visualization and SQL
  - Discovering new patterns and insights from the data using python's visualization libraries (Matplotlib, Seaborn) & SQL
- Perform interactive visual analytics using Folium and Plotly Dash
  - Dash & Folium python libraries were used to create the interactive dashboards & visualization on the maps.
- Perform predictive analysis using classification models
  - Different classification machine learning models were built and tested. The model with highest accuracy score is selected.

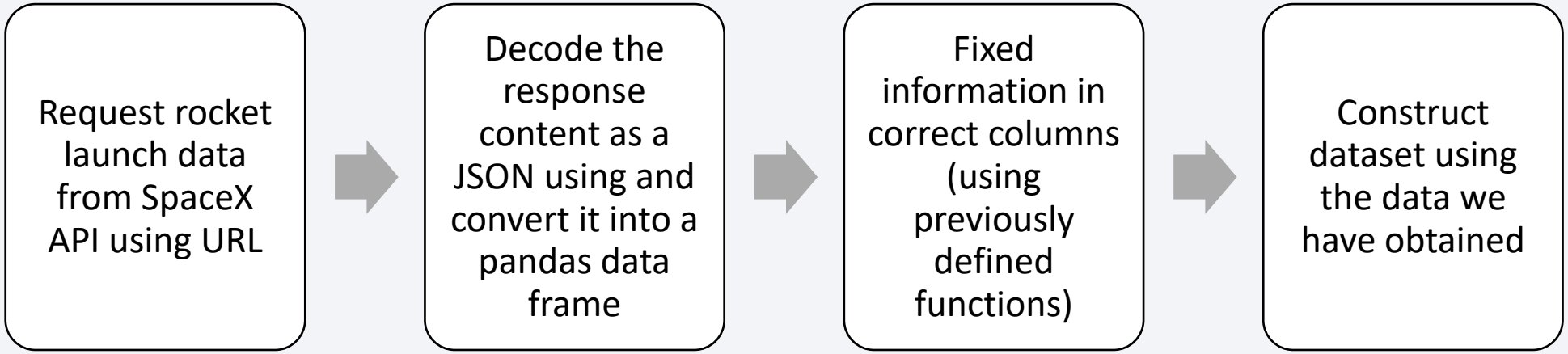
Data Collection

- Data sets were collected using the API call from several websites, I collected rocket, launchpad, payloads, and cores data from https://api.spacexdata.com/v4 website



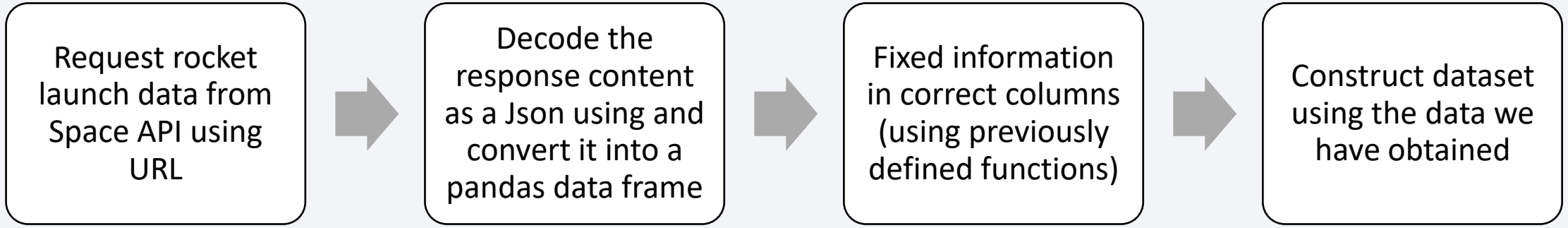
Data Collection – SpaceX API

- GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/1-data-collection-api.ipynb



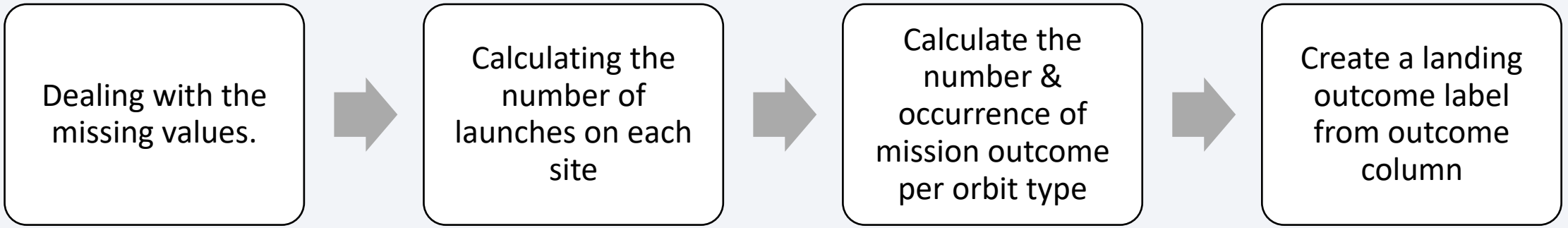
Data Collection - Scraping

- GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/jupyter-labs-webscraping.ipynb



Data Wrangling

GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/data_wrangling_jupyterlite.jupyterlite.ipynb



EDA with Data Visualization

- The graphs used are :
  - Scatter Plot
  - Bar Chart
  - Line Chart

- GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- SQL queries performed
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CA'
  - 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 achieved
  - List the names of the boosters which have success in drone ship 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. Use a subquery
  - List the failed landing outcomes in drone ship, their booster versions, and launch site names for the 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
- GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Summary of map objects:
  - Markers: Show a geo location from latitude and longitude data
  - Cluster: Show a group of markers
  - Circles: Show a single location
  - Lines: Show distance between two
  - I have added object to find some geographical patterns about launch site

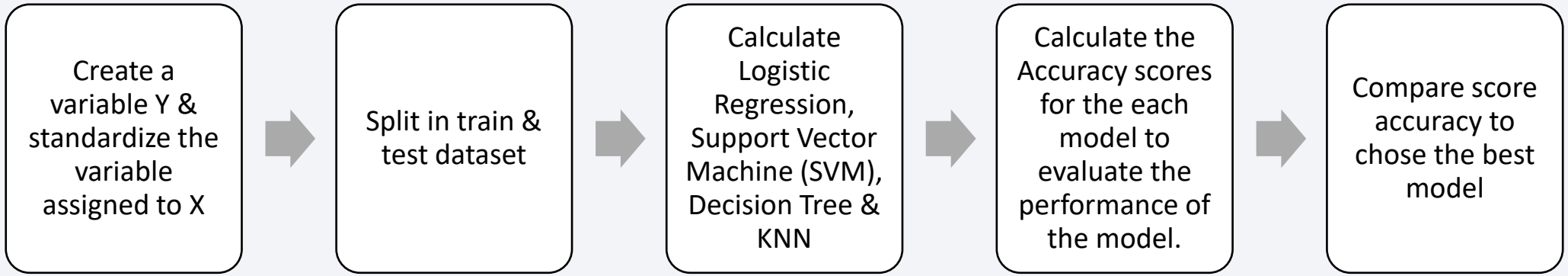
- GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/launch_site_location.jupyterlite.ipynb

Build a Dashboard with Plotly Dash

- Summary of plots:
  - Bar: Show categories differences
  - Line: Reports time series changes
  - Pie: Shows the percentage of events
  - Tree: Shows complex relationship of variables in interactive way
  - Map: Shows variables of states on a map

- GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/spacex_dash_app.py

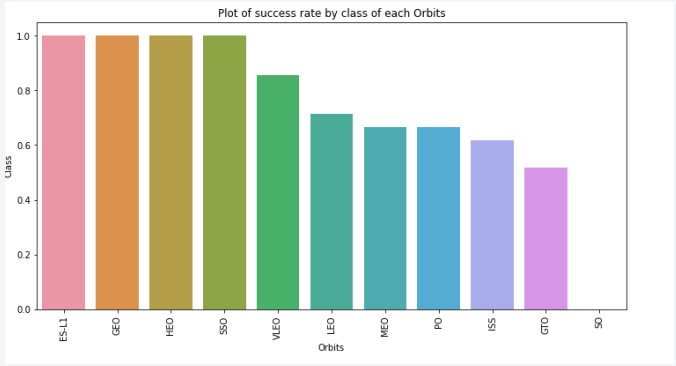
Predictive Analysis (Classification)

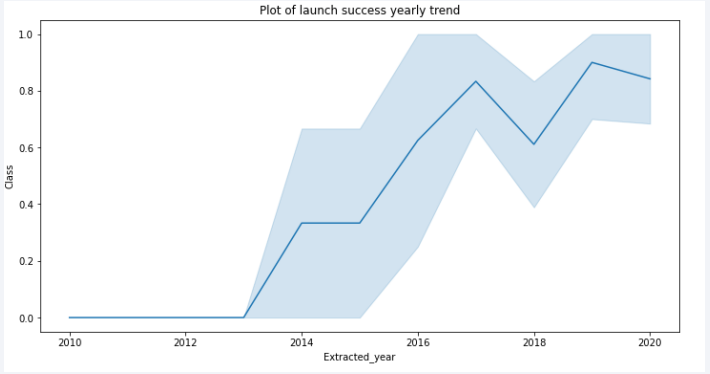


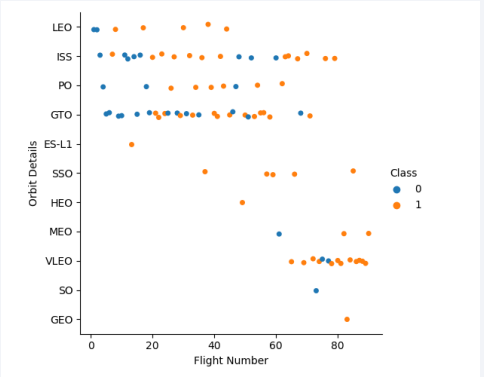
GitHub Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project/blob/f97c78395fa8e9e4302509a9631910b7bbcd6db9/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

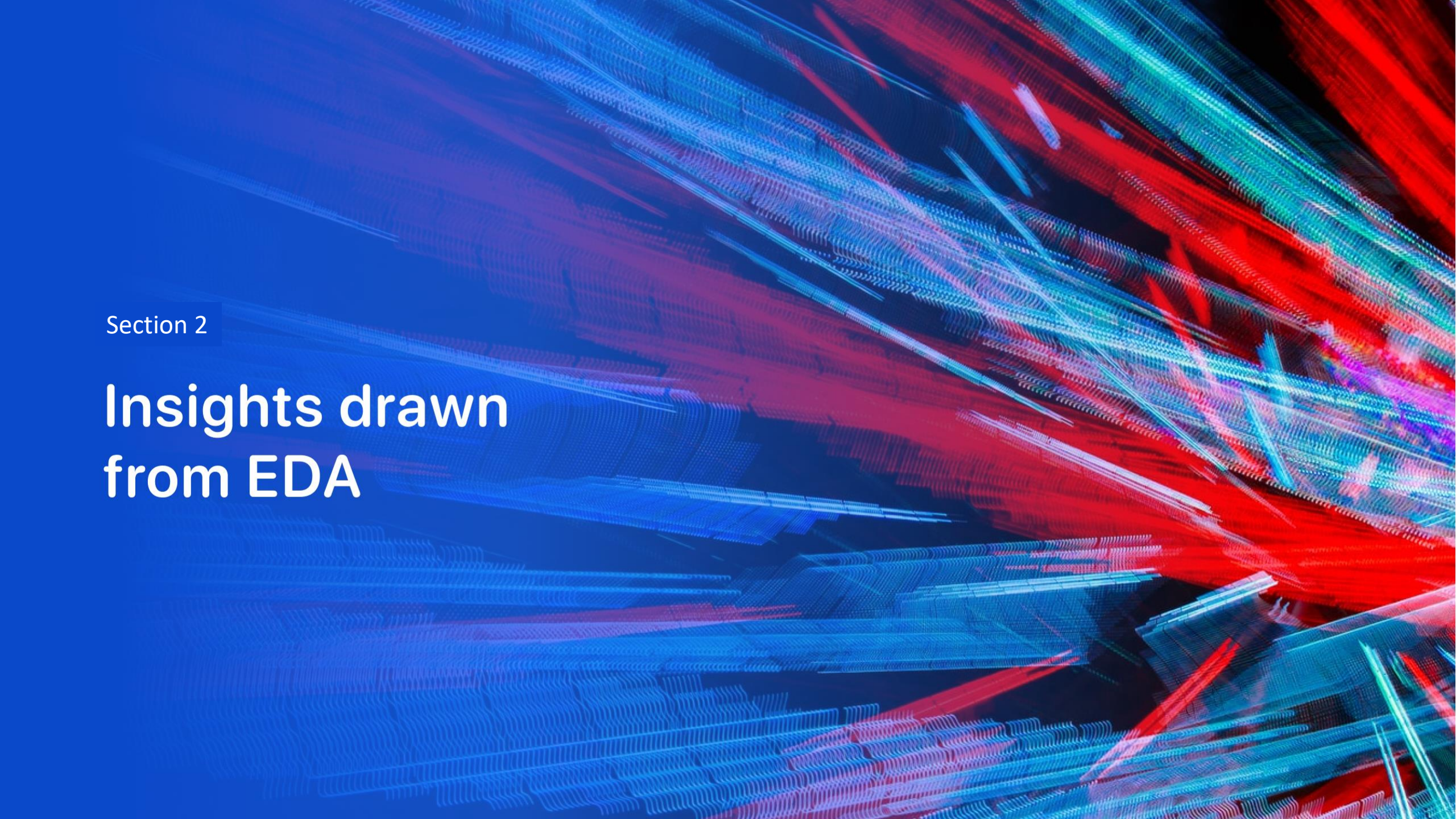
- Exploratory data analysis results
  - Both API & web scrapping are capable to collect XSpace data
- Interactive analytics demo in screenshots
  - EDA with SQL is effective for data filtering
  - EDA with interactive visualization provides informative information
  - Plotly Dash is powerful to show instant data change





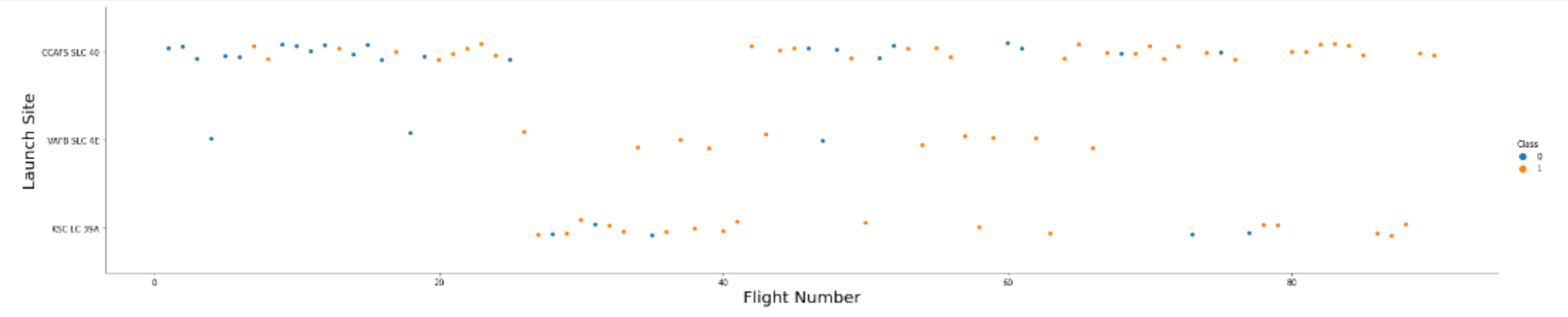


- Predictive analysis results
  - Decision Tree Classifier Algorithm has the best accuracy of predicting.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

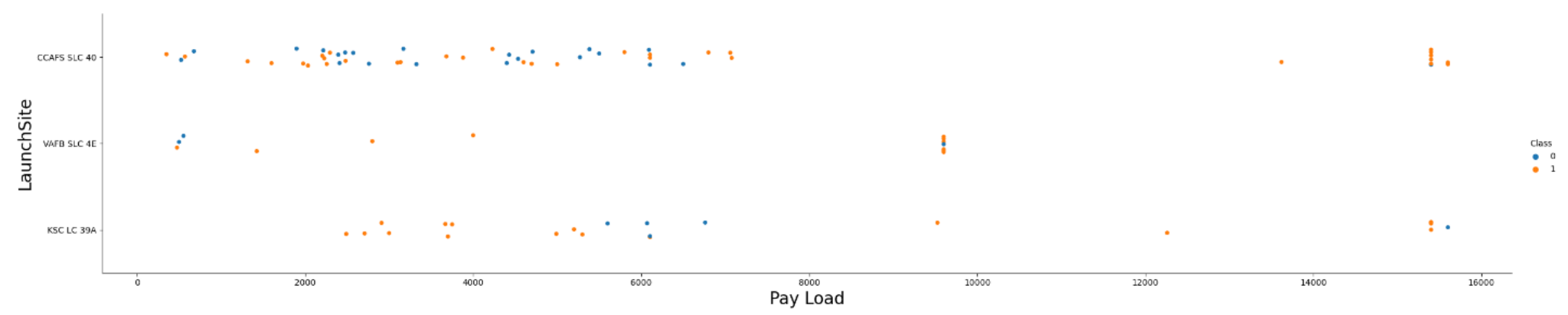
- Show a scatter plot of Flight Number vs. Launch Site



- For the CCFS SLC 40 category there seems to be a higher concentration than class 1 for flights with Flight Number high.

Payload vs. Launch Site

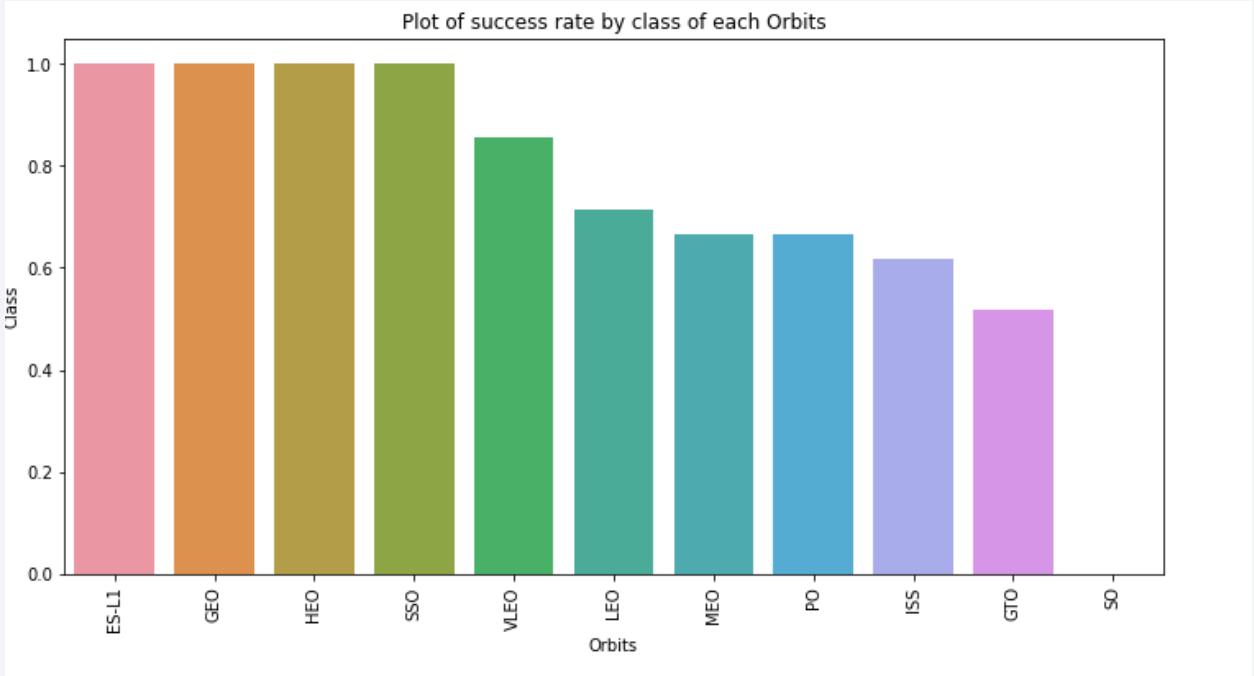
- Show a scatter plot of Payload vs. Launch Site



- No strong correlation is appeared in this case.

Success Rate vs. Orbit Type

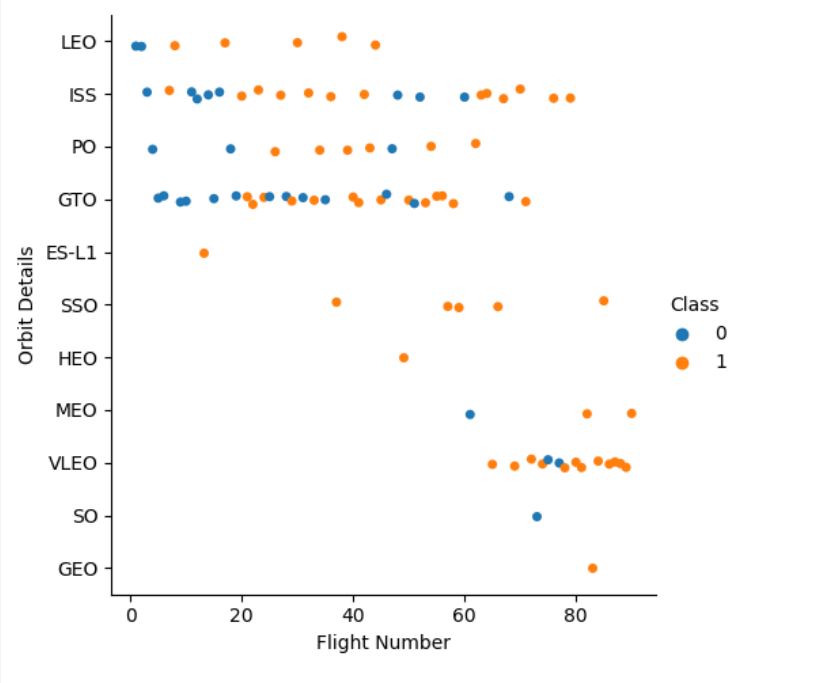
- Show a bar chart for the success rate of each orbit type



- We can see that the orbits with the highest success rate are: SSO, HEO, GEO, ES-L1
- While the GTO Orbit it is the one with lowest rate

Flight Number vs. Orbit Type

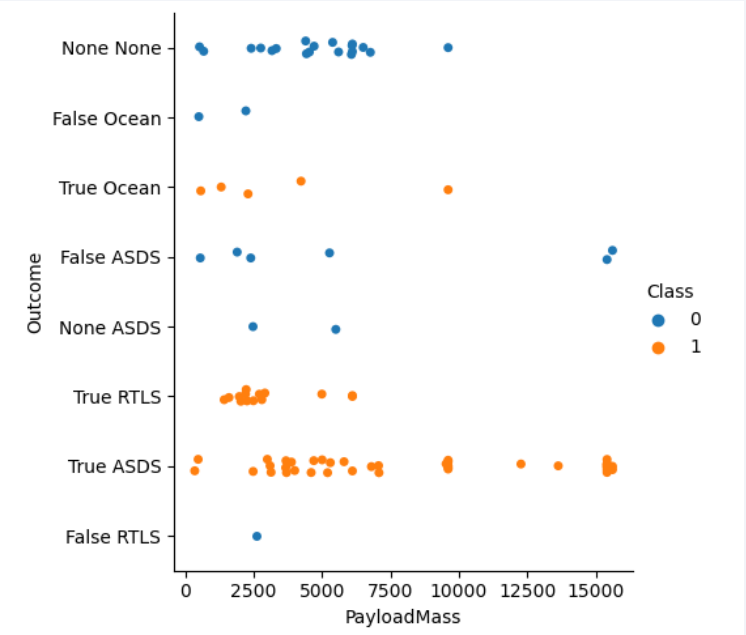
- Show a scatter point of Flight number vs. Orbit type



- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type

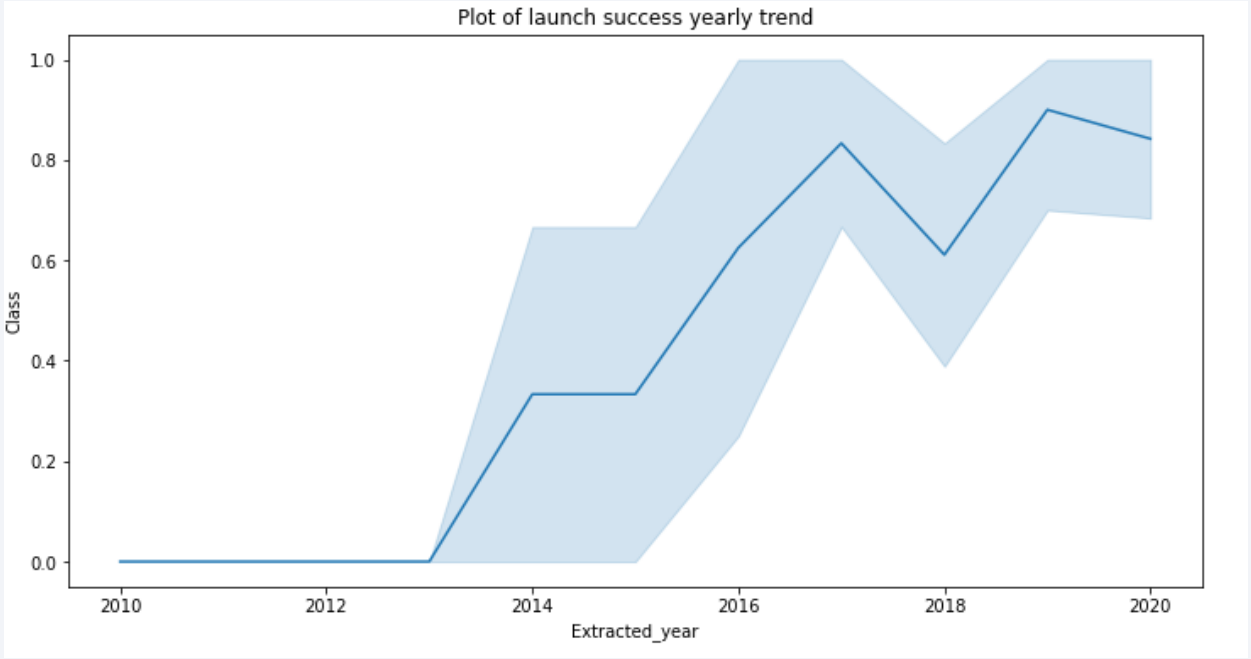
- Show a scatter point of payload vs. orbit type



- You should observe that Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend

- Show a line chart of yearly average success rate



- You can see that the success rate since 2013 kept increasing till 2020.

All Launch Site Names

- Find the names of the unique launch sites
  - CCAFS LC -40
  - VAFB SLC -4E
  - KSC LC-39A
  - CCAFS SLC -40

```
query = "SELECT DISTINCT Launch_Site FROM SPACEXTBL"
result = con.execute(query)
unique_launch_sites = [row[0] for row in result.fetchall()]
for i in unique_launch_sites:
    print(i)
```

```
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None
```

- Present your query result with a short explanation here
  - SQL query : "SELECT DISTINCT Launch Site FROM SPACEXTBL“
  - Using the DISTINCT keyword we have obtained the desired result.

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

	Date	Time_(UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
1	12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0.0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
3	10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
4	03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt

- Present your query result with a short explanation here
  - SQL Query : "SELECT * FROM SPACEXTBL WHERE Launch_Site LIKE 'CCA%%' LIMIT 5"

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

	total payload mass carried by boosters	Customer
0	45596.0	NASA (CRS)

- Present your query result with a short explanation here
  - SQL query : "SELECT SUM(PAYLOAD_MASS__KG_), Customer FROM SPACEXTBL WHERE Customer='NASA (CRS)'"

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

	Average payload mass carried by boosters	Booster_Version
0	2534.666667	F9 v1.1 B1003

- Present your query result with a short explanation here
  - SQL query : "SELECT AVG(PAYLOAD_MASS__KG_), Booster_Version FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%%%'"

First Successful Ground Landing Date

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

	Date	Time_(UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	01/08/2018	1:00:00	F9 B4 B1043.1	CCAFS SLC-40	Zuma	5000.0	LEO	Northrop Grumman	Success (payload status unclear)	Success (ground pad)

- Present your query result with a short explanation here
  - SQL query1 = "create view View1 as SELECT * FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)'"
  - SQL query2 = "select * from View1 where [Date] = (SELECT MIN(Date) FROM View1)"

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

- Present your query result with a short explanation here
  - SQL query : "select distinct Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ between 4000 and 6000"

	boosters
0	F9 v1.1
1	F9 v1.1 B1011
2	F9 v1.1 B1014
3	F9 v1.1 B1016
4	F9 FT B1020
5	F9 FT B1022
6	F9 FT B1026
7	F9 FT B1030
8	F9 FT B1021.2
9	F9 FT B1032.1
10	F9 B4 B1040.1
11	F9 FT B1031.2
12	F9 B4 B1043.1
13	F9 FT B1032.2
14	F9 B4 B1040.2
15	F9 B5 B1046.2
16	F9 B5 B1047.2
17	F9 B5 B1046.3
18	F9 B5B1054
19	F9 B5 B1048.3
20	F9 B5 B1051.2
21	F9 B5B1060.1
22	F9 B5 B1058.2
23	F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

	Mission_Outcome	Count
0	None	0
1	Failure (in flight)	1
2	Success	98
3	Success	1
4	Success (payload status unclear)	1

- Present your query result with a short explanation here
  - SQL query : "select Mission_Outcome, COUNT(Mission_Outcome) from SPACEXTBL GROUP BY Mission_Outcome"

Boosters Carried Maximum Payload

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

- Present your query result with a short explanation here
  - SQL query = "select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL)"

	Booster_Version
0	F9 B5 B1048.4
1	F9 B5 B1049.4
2	F9 B5 B1051.3
3	F9 B5 B1056.4
4	F9 B5 B1048.5
5	F9 B5 B1051.4
6	F9 B5 B1049.5
7	F9 B5 B1060.2
8	F9 B5 B1058.3
9	F9 B5 B1051.6
10	F9 B5 B1060.3
11	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

	Month	Landing_Outcome	Booster_Version	Launch_Site
0	10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
1	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- Present your query result with a short explanation here
  - SQL query = "SELECT substr(Date, 4, 2) AS Month, Landing_Outcome, Booster_Version, Launch_Site FROM SPACEXTBL WHERE substr(Date, 7, 4) = '2015' AND Landing_Outcome LIKE '%Failure (drone ship)%';"

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

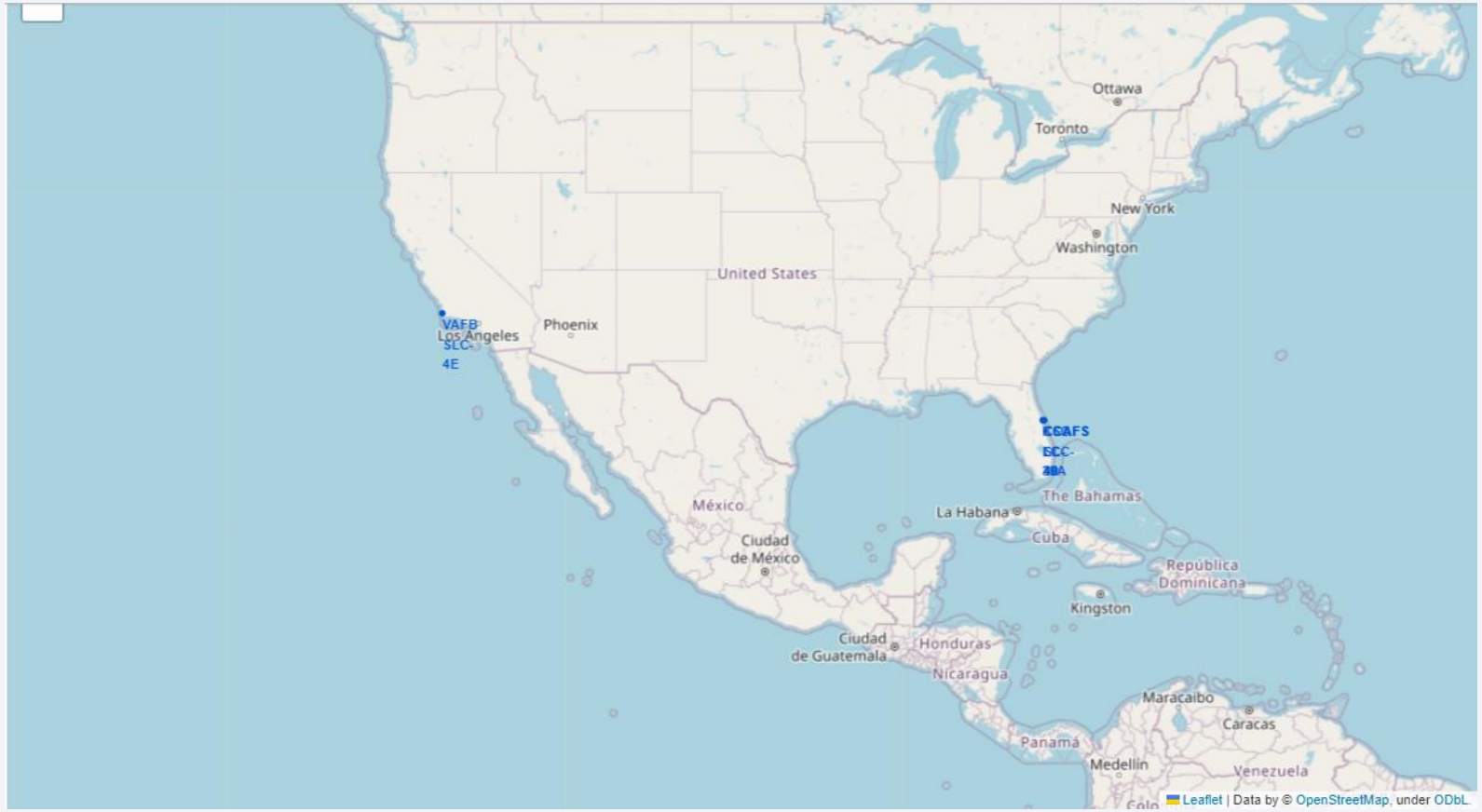
	landingoutcome	count
0	No attempt	10
1	Success (drone ship)	6
2	Failure (drone ship)	5
3	Success (ground pad)	5
4	Controlled (ocean)	3
5	Uncontrolled (ocean)	2
6	Precluded (drone ship)	1
7	Failure (parachute)	1

- Present your query result with a short explanation here
  - SQL query : SELECT Landing_Outcome, COUNT(Landing_Outcome) FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome ORDER BY COUNT(Landing_Outcome) DESC

Section 3

Launch Sites Proximities Analysis

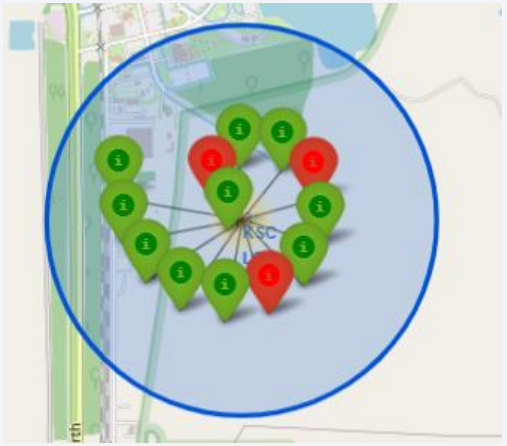
Launch Sites



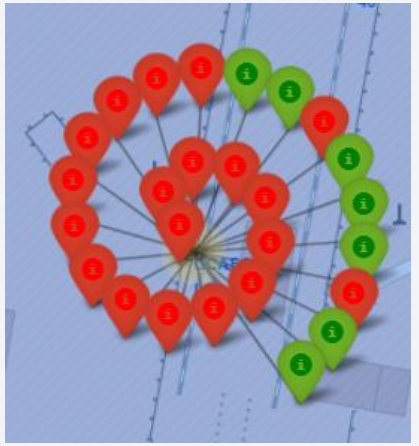
- Both ground and sea surface sites are necessary also south areas maybe a proper areas.
- The transportation base chosen maybe important

Outcome of Launch

VAFB SLC -4E



KSC LC-39A



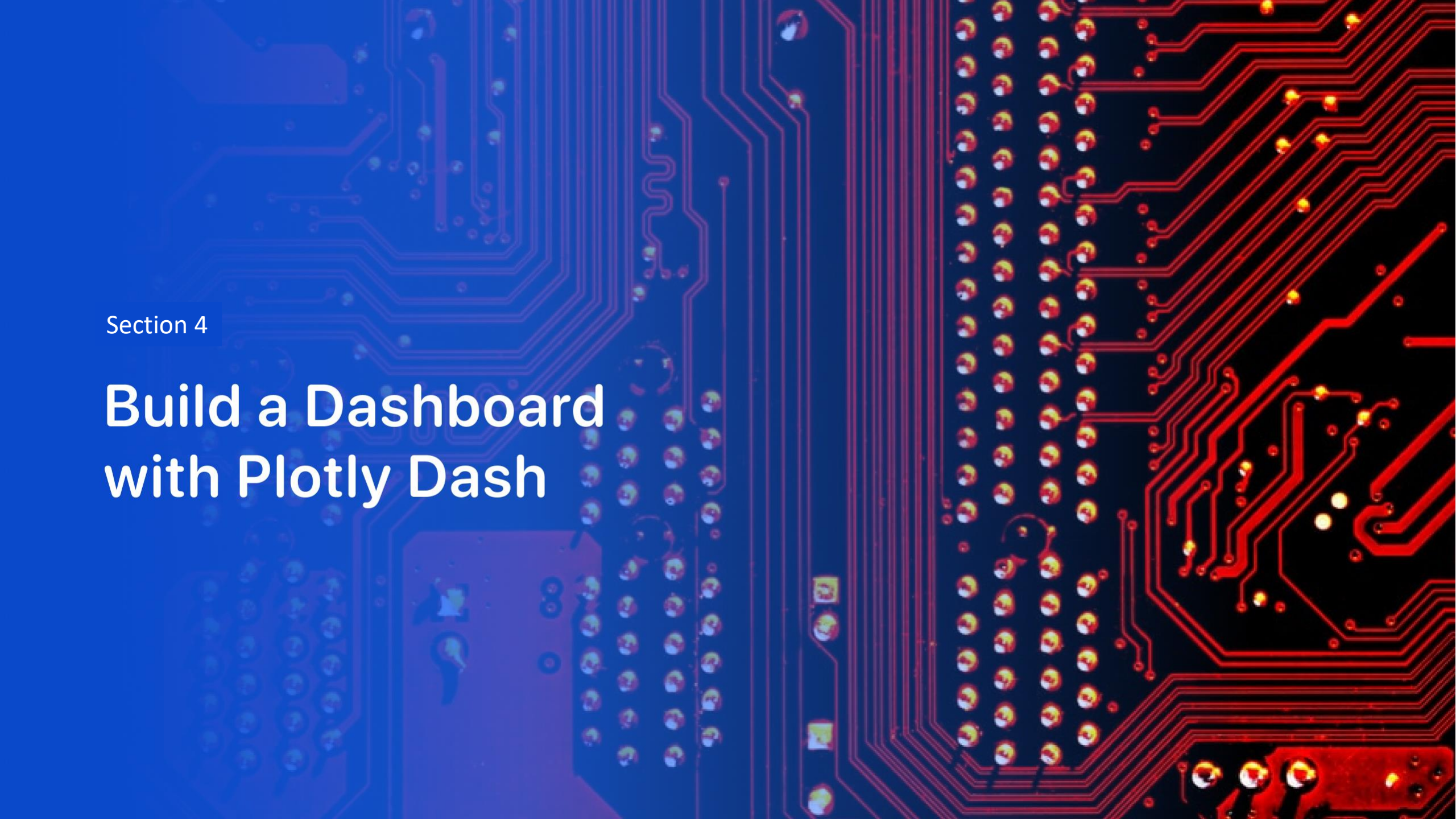
CCAFS LC -40

CCAFS SLC -40

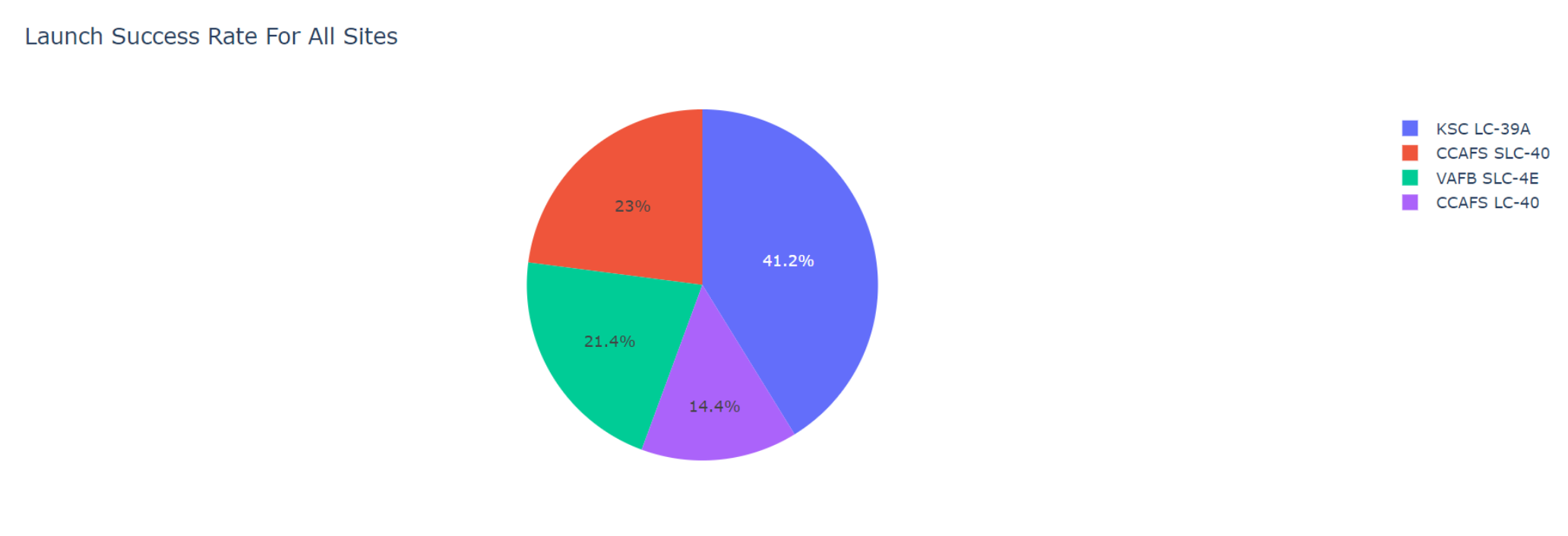
Color icons is impressive way to show the rate of success. For example, in KSC there is the best rate of success.

Important Places & Their Distance

This map shows the distance from points of interest. In this case the distance from coast to the launch site is marked and calculated as may be it will be important for further analysis

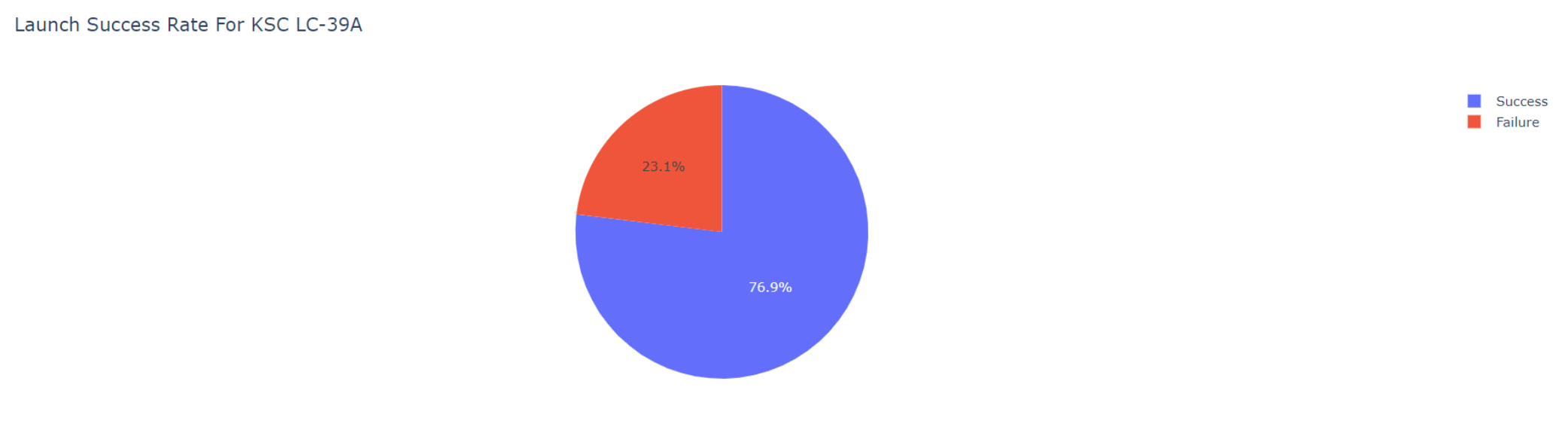

Section 4
Build a Dashboard with Plotly Dash

Success Rate of Launch Sites



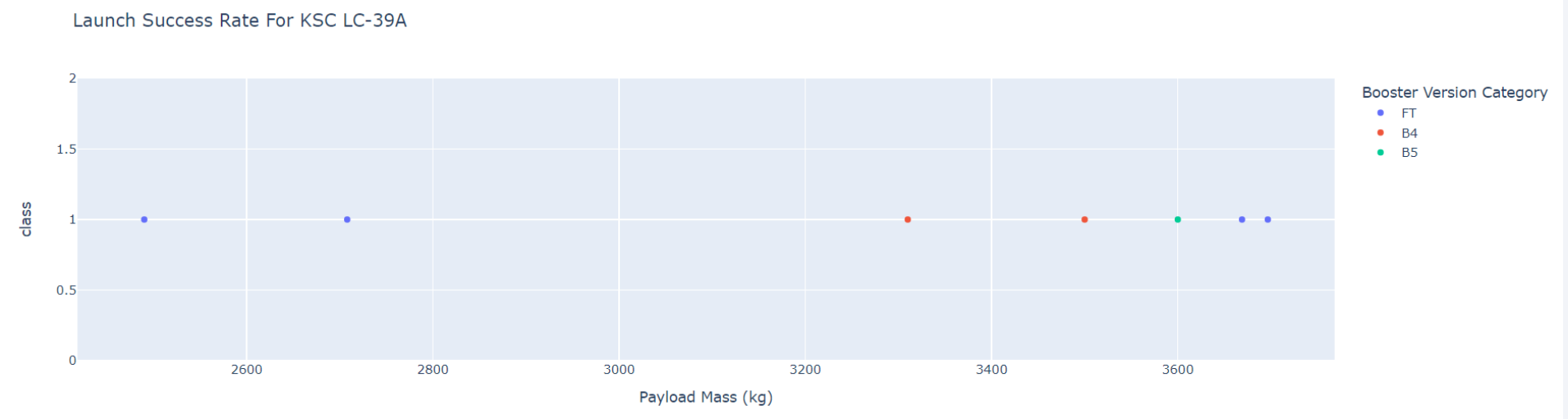
- The launch site with the most Success rate is KSC LC-39A
- The launch site with least success rate is CCAFS LC-40

KSC LC-39A

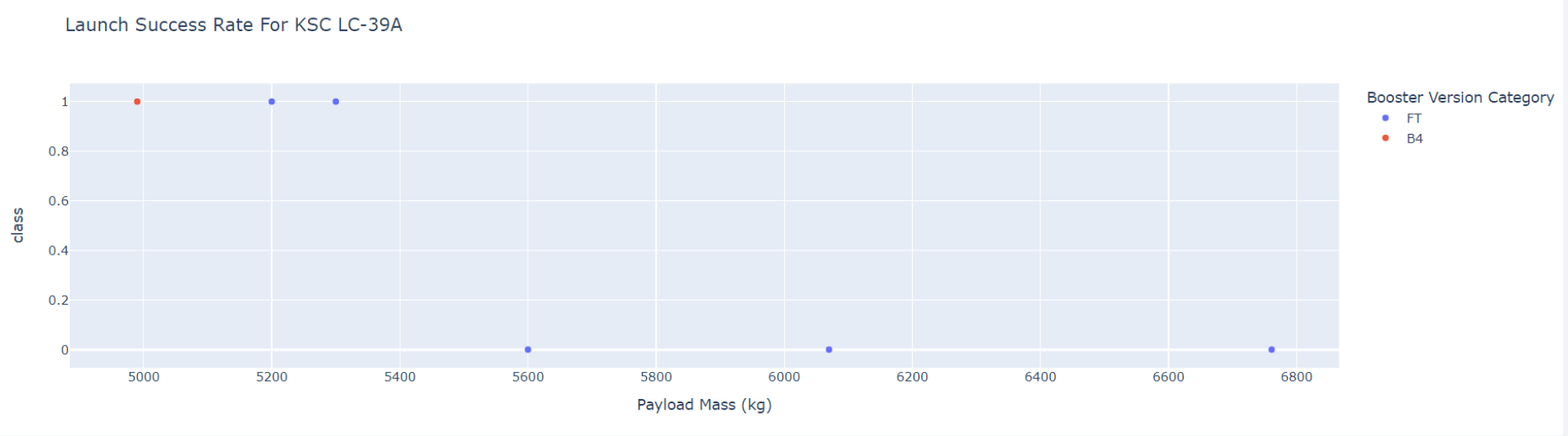


- This image shows ：
  - 76.9% of launches from this launch site are resulted as successful
  - 23.1% of launches from this launch site are resulted as Failure

Payload vs. Launch Outcome



Payload vs. Launch Outcome (0-4000)

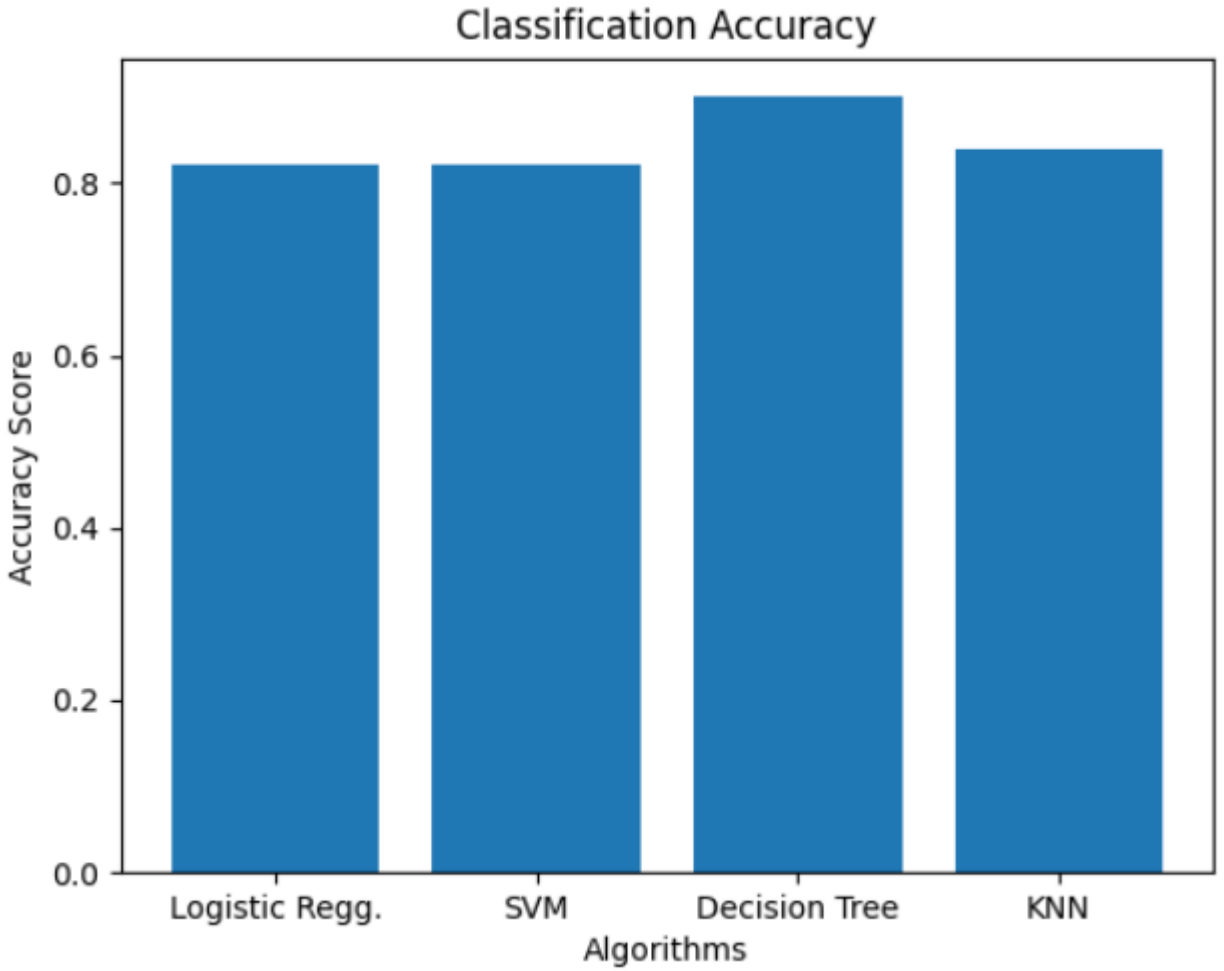


Payload vs. Launch Outcome (4000-10000)

Section 5

Predictive Analysis (Classification)

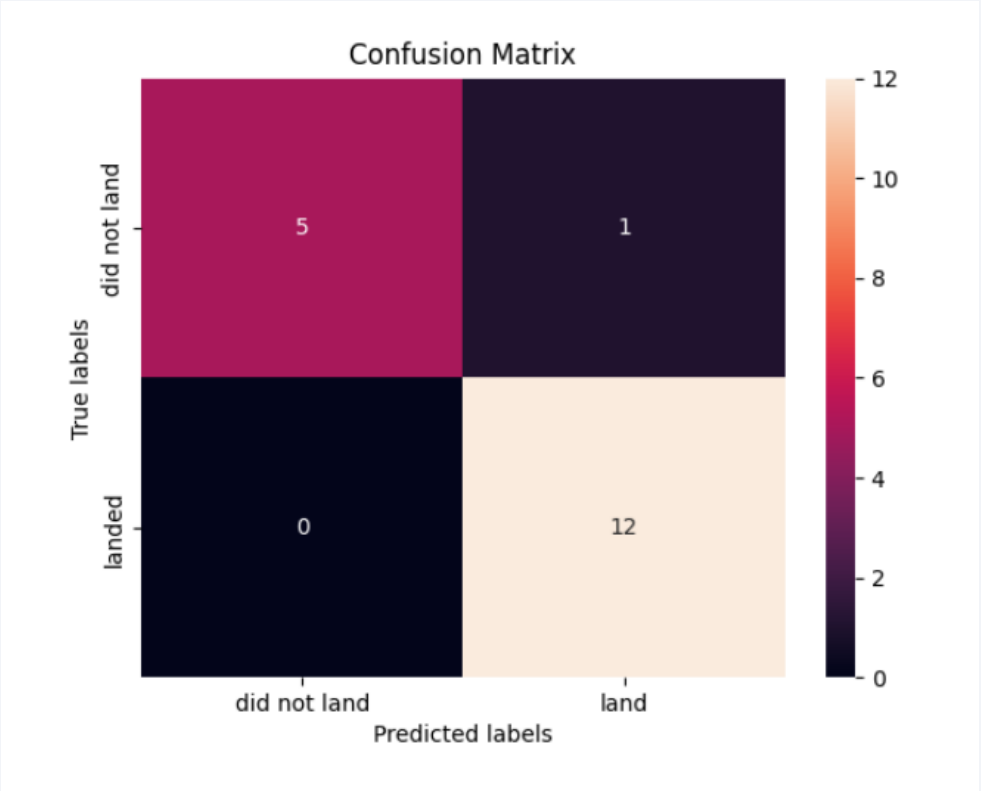
Classification Accuracy



Decision Tree has the highest accuracy score of 90%. So, Decision tree will be the main model to predict the outcome.

Confusion Matrix

- Confusion matrix of Decision Tree algorithm.
- From this we can say that our model has predicted :
  - 17 outcomes correctly or accurately out of total 18 outcomes.
  - 1 outcome is predicted wrong out of 18 outcomes.
  - This shows that our model is working properly with high accuracy.



Conclusions

- We can conclude that:
  - The larger the flight amount at a launch site, the greater the success rate at a launch site.
  - Launch success rate started to increase in 2013 till 2020.
  - Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
  - KSC LC-39A had the most successful launches of any sites.
  - The Decision tree classifier is the best machine learning algorithm for this task

Appendix

- For more details or code files please visit my GitHub Repository
- Link : https://github.com/hirenkanojia/Applied-Data-Science-Capstone-Project

Thank you!