IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
  - Perform Data collection
  - Perform data wrangling
  - Perform exploratory data analysis (EDA) using visualization and SQL
  - Perform interactive visual analytics using Folium and Plotly Dash
  - Perform predictive analysis using classification models
- Summary of all results
  - Flight number increase, success rate increases as well.
  - ES-L1, GEO, HEO, SSO has very high success rate; GTO has very low success rate
  - VLEO has data for very high flight number.
  - For LEO, ISS, PO, GTO, the success rate is higher when flightnumber is higher.
  - Sucess rate since 2013 kept increasing till 2020
  - LC-39A has the highest success count, Success Rate for LC-39A is 76.9%
  - Booster version FT seems to have the highest success rate
  - All 4 models have the same performance for the test accuracy
  - For confusion matrix:
  - When true label is landed, predicted label is all correct as landed
  - When true label is not landed, predicted label is 50% correct

Introduction

- Project background and context
  - In this capstone, we will predict if the Falcon 9 first stage will land successfully.
  - 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.
  - Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.
- Problems you want to find answers
  - predict if the Falcon 9 first stage will land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected either from an API or using Web Scraping method.
- Perform data wrangling
  - Determine Training Labels by converting Landing Outcomes into 1 (success) and 0 (failure)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Create a column for class, Standardize the data, Split into Train/Test set
  - Find best hyperparameters for Logistic Regression/SVM/Decision Tree/KNN models
  - Evaluate classification models with accuracy and confusion matrix metrics.

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

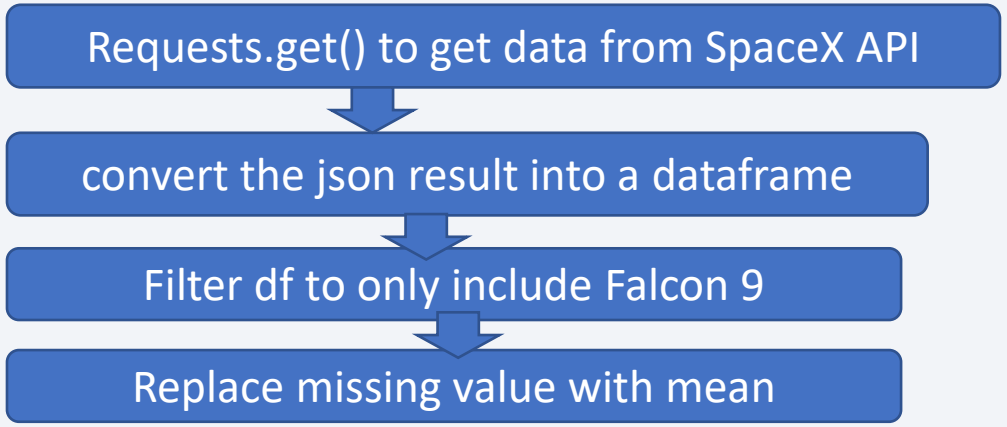
Method 1: API Method
Method 2: Web Scraping

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

- https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week1.%20Data%20Collection%20and%20Wrangling/1a.%20Data%20Collection%20using%20API%20Method.ipynb

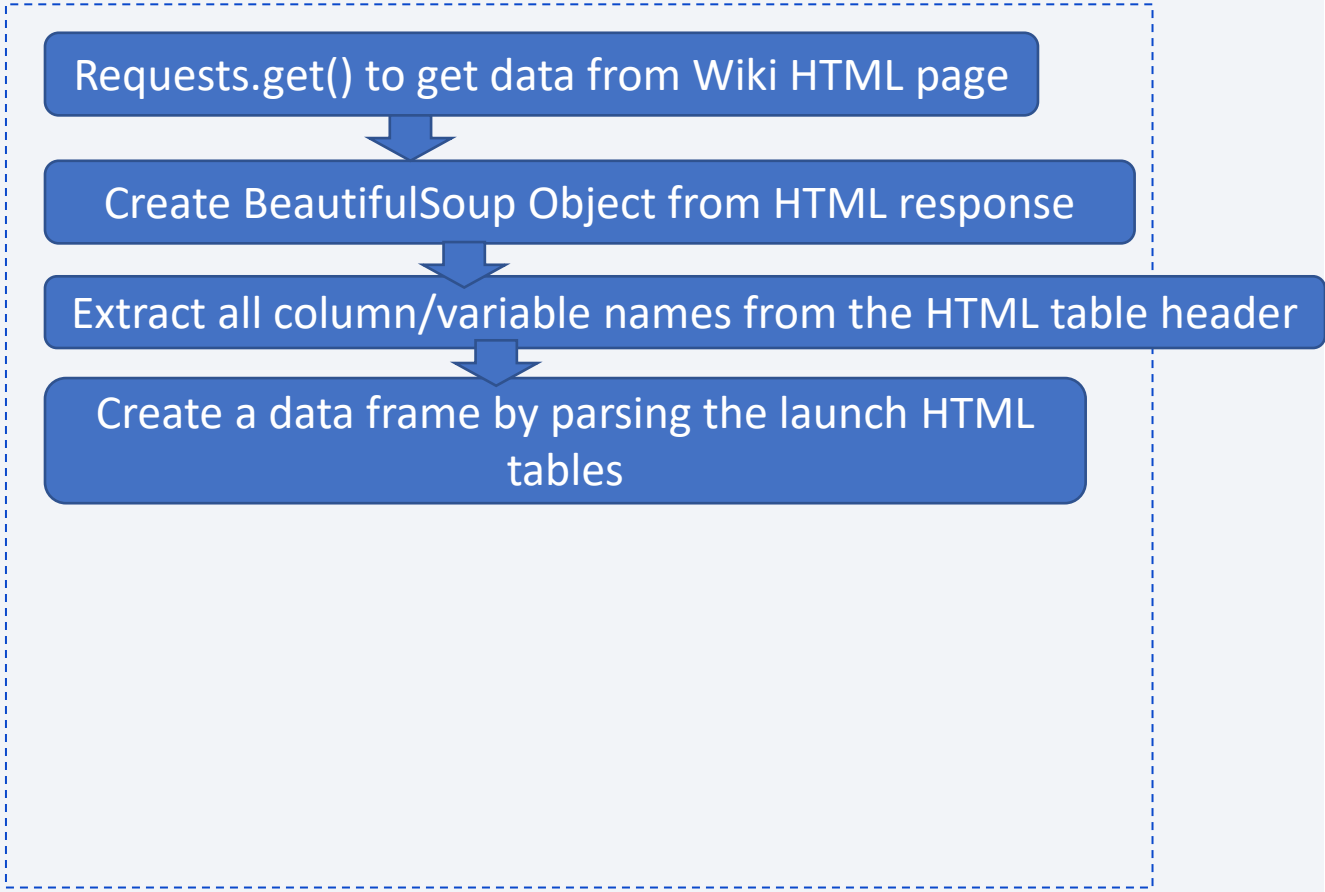


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

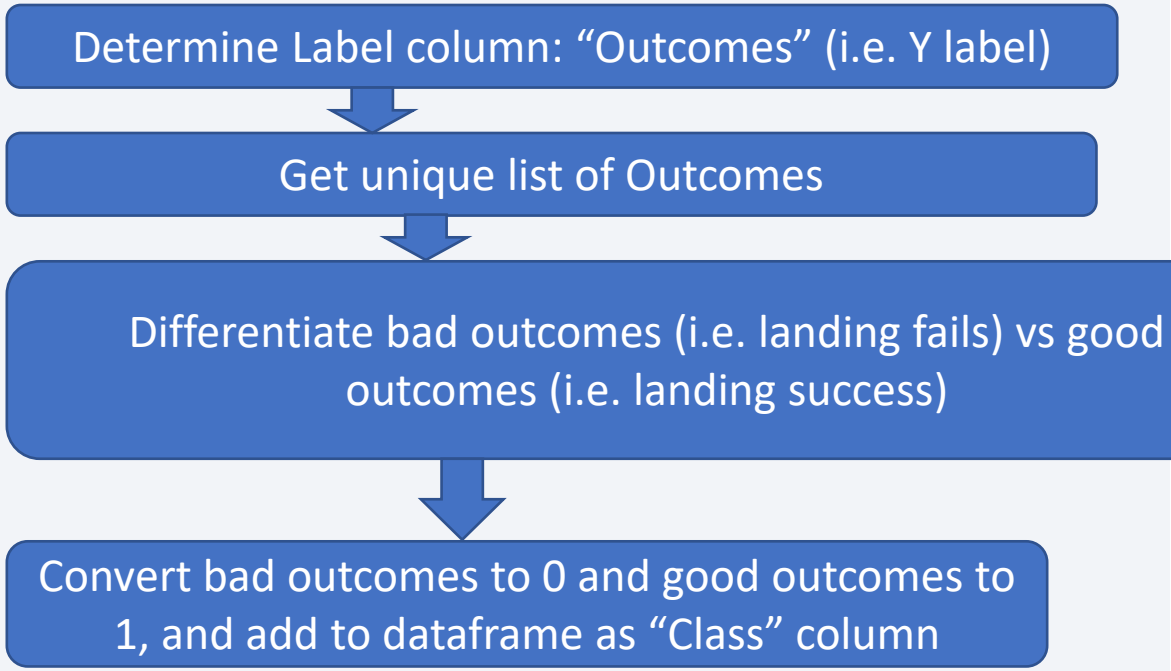
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

- https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week1.%20Data%20Collection%20and%20Wrangling/1b.%20Data%20Collection%20using%20Web%20Scraping%20Method.ipynb



Data Wrangling

- Describe how data were processed
  - Determine Training Labels by converting Landing Outcomes into 1 (success) and 0 (failure)
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
- https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week1.%20Data%20Collection%20and%20Wrangling/2.%20Data%20Wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
    - 1. PayloadMass vs. FlightNumber and outcomes of the launch. To observe if launch success will have positive or negative relationship with either flight number or pay load mass.
    - 2. Launch Site vs. Flight Number and outcomes of the launch. To observe which site have more success rate and its relationship with flight number
    - 3. Launch Site vs. Payload and outcomes of the launch. To observe which site have more success rate and its relationship with payload mass
    - 4. Average Success Rate of each orbit type
    - 5. Orbit vs. Flight Number and outcomes of the launch. To observe which orbit has more success rate and its relationship with flight number
    - 6. Avg success rate of each year. To observe if overall success rate is increasing.
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose
    - https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week2.%20EDA%20(Exploratory%20Data%20Analysis)/Lab%202.%20EDA%20with%20Pandas%20and%20Matplotlib_Seaborn.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - ***Display the names of the unique launch sites in the space mission***
  - ***Display 5 records where launch sites begin with the string 'CCA'***
  - ***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 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 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***
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
  - https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week2.%20EDA%20(Exploratory%20Data%20Analysis)/Lab%201.%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

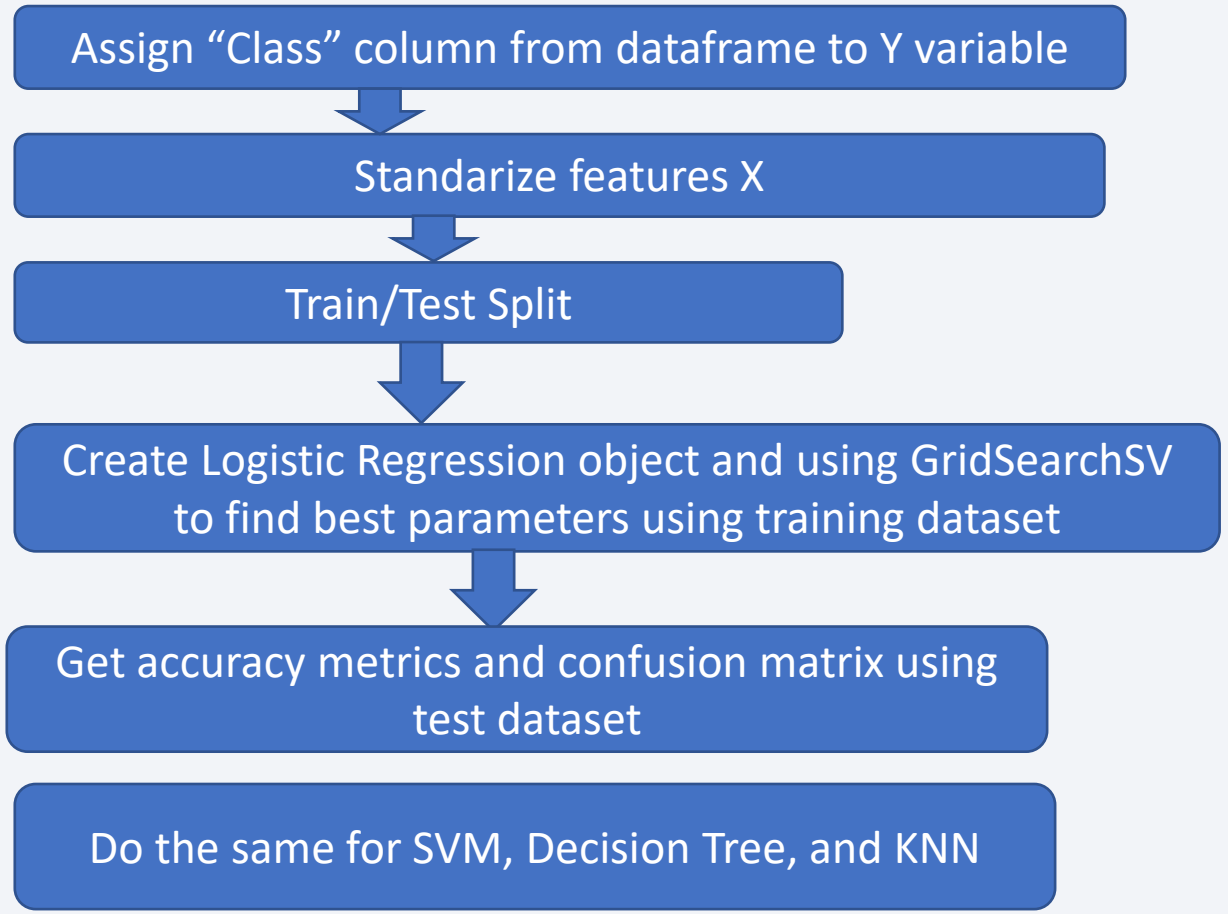
- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map; Explain why you added those objects
  - Circle object to indicate NASA Johnson Space Center
  - Marker object as a text marker to show NASA Johnson Space Center's name on map
  - Circle and Marker objects for each launch site
  - Marker_Cluster object to indicate success vs fail launch outcomes with different marker color
  - PolyLine object to show the distance between two points on the map
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
  - https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week3.%20Interactive%20Visual%20Analytics%20and%20Dashboard/Lab1_Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard; Explain why you added those plots and interactions
  - Add a dropdown list to enable Launch Site selection
  - Add a pie chart to show the total successful launches count for all sites
  - If a specific launch site was selected, show the Success vs. Failed counts for the site
  - Add a slider to select payload range
  - Add a scatter chart to show the correlation between payload and launch success, based on selected site and payload range
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose
  - https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week3.%20Interactive%20Visual%20Analytics%20and%20Dashboard/Lab2_spacex_dash_app.py

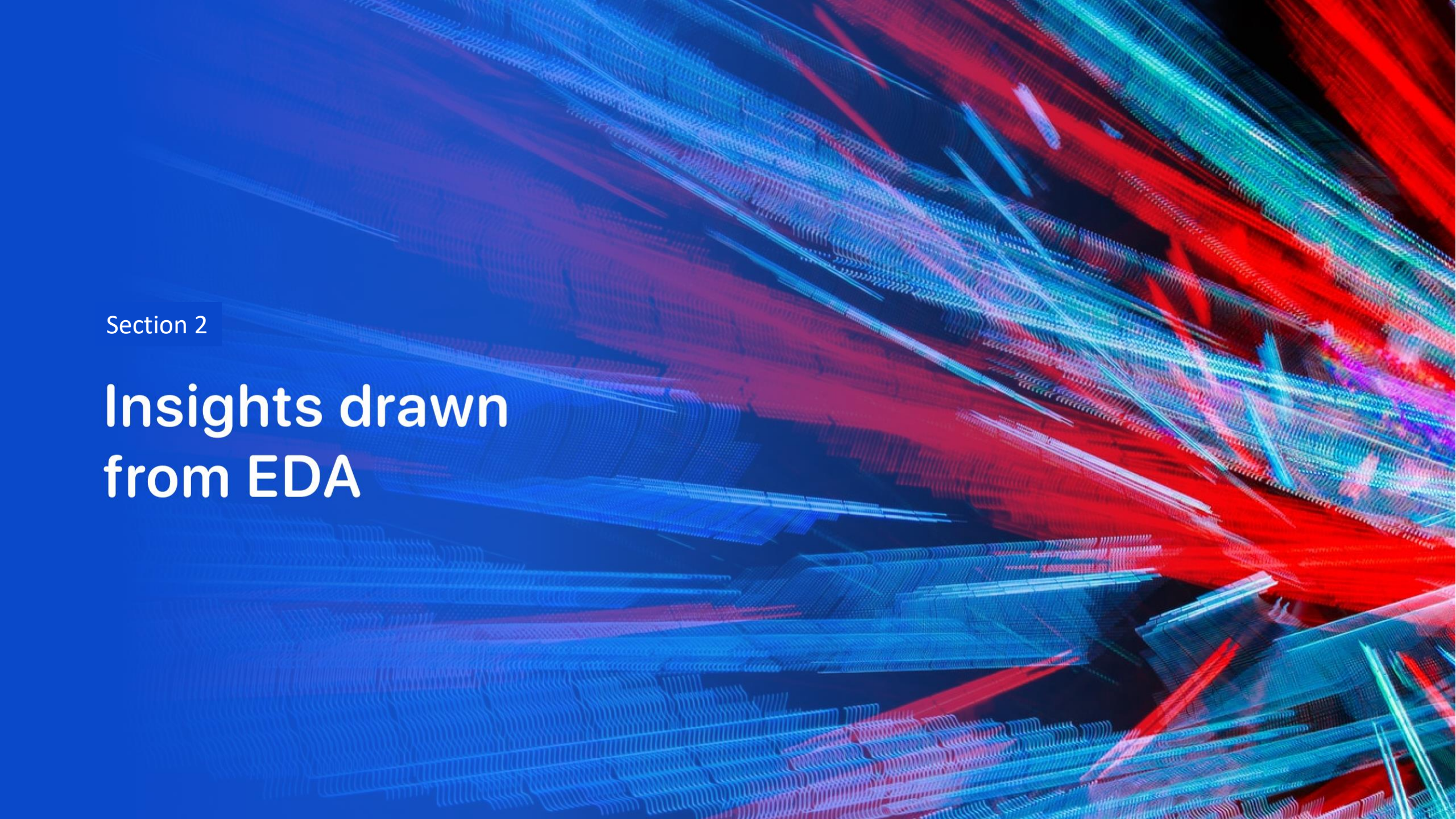
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
  - Create a column for class, Standardize the data, Split into Train/Test set
  - Find best hyperparameters for Logistic Regression/SVM/Decision Tree/KNN models
  - Evaluate classification models with accuracy and confusion matrix metrics.
- You need present your model development process using key phrases and flowchart
  - See right
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose
  - https://github.com/Haihan-W/IBMDSProfessional/blob/main/Course10_Applied%20Data%20Science%20Capstone/Week4.%20Predictive%20Analysis%20using%20ML/Machine%20Learning%20Prediction.ipynb

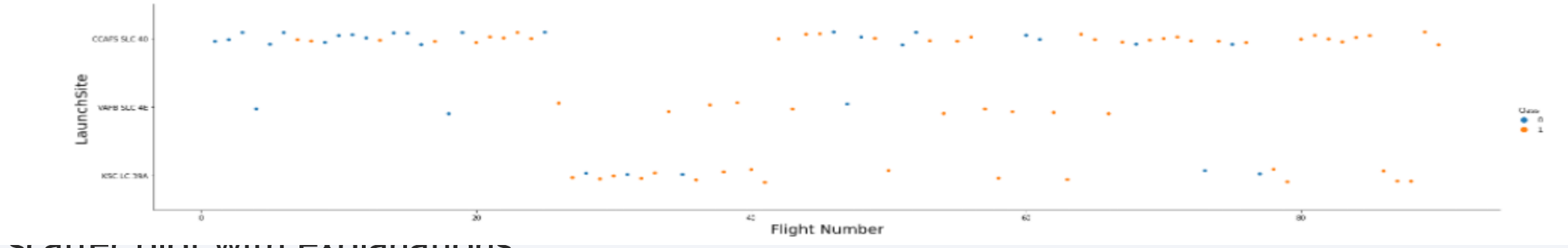


Results

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


Section 2
Insights drawn from EDA

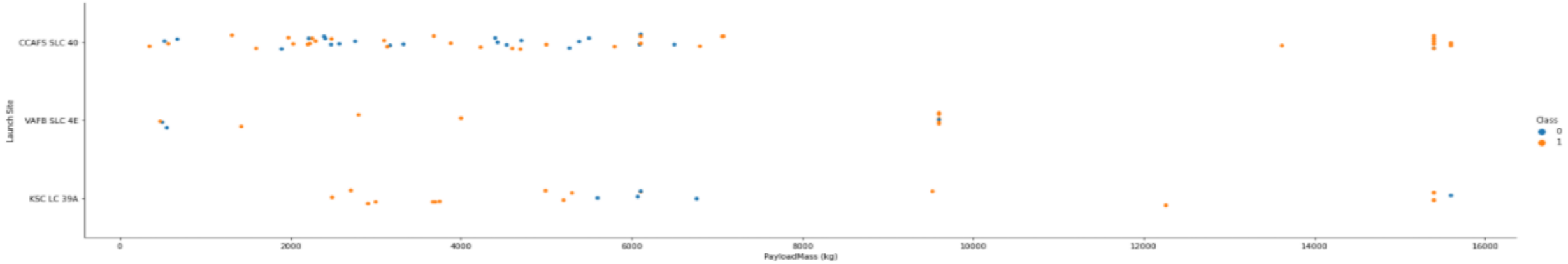
Flight Number vs. Launch Site

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

- scatter plot with explanations
  - Flight number increase, success rate increases as well.
  - This applies for all three lauch sites

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site



- Show the screenshot of the scatter plot with explanations
  - For these three sites, it is observed that there is no strongly relationship between payload and success rate.

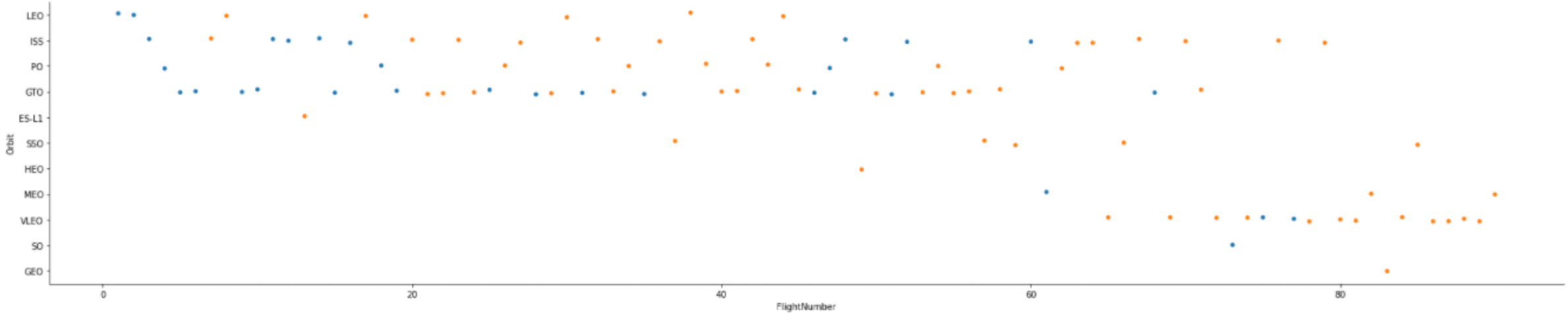
Success Rate vs. Orbit Type

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

- Show the screenshot of the scatter plot
  - ES-L1, GEO, HEO, SSO has very high success rate
  - GTO has very low success rate
  - SO does not have success rate data or its success rate is 0.

Flight Number vs. Orbit Type

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

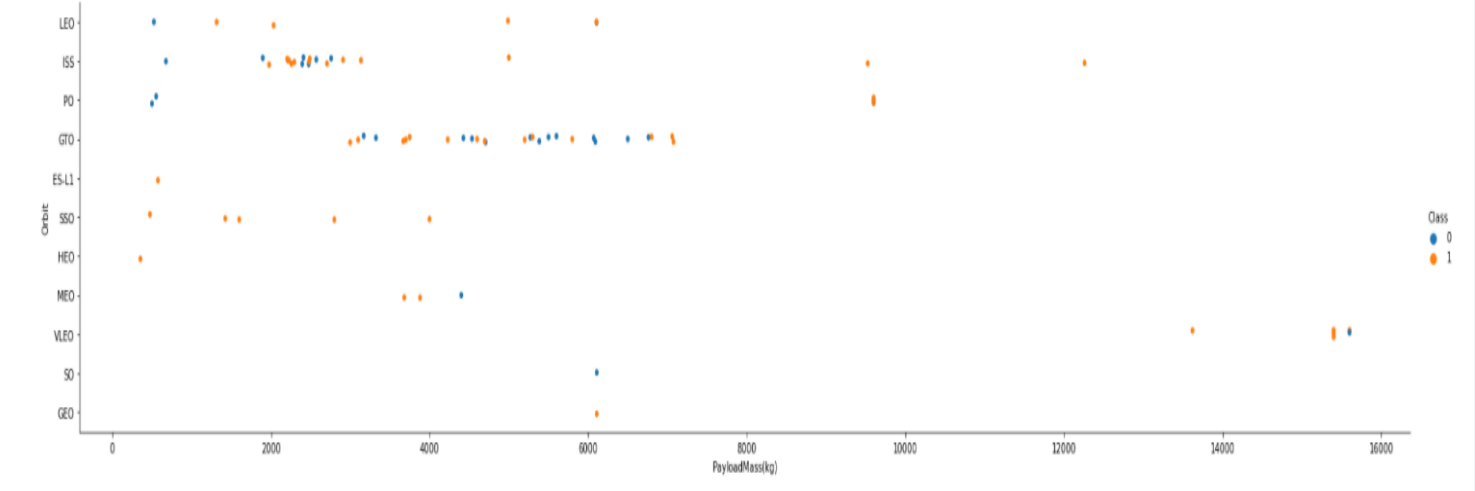


- VLEO has data for very high flight number.
- For LEO, ISS, PO, GTO, the success rate is higher when flightnumber is higher.

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

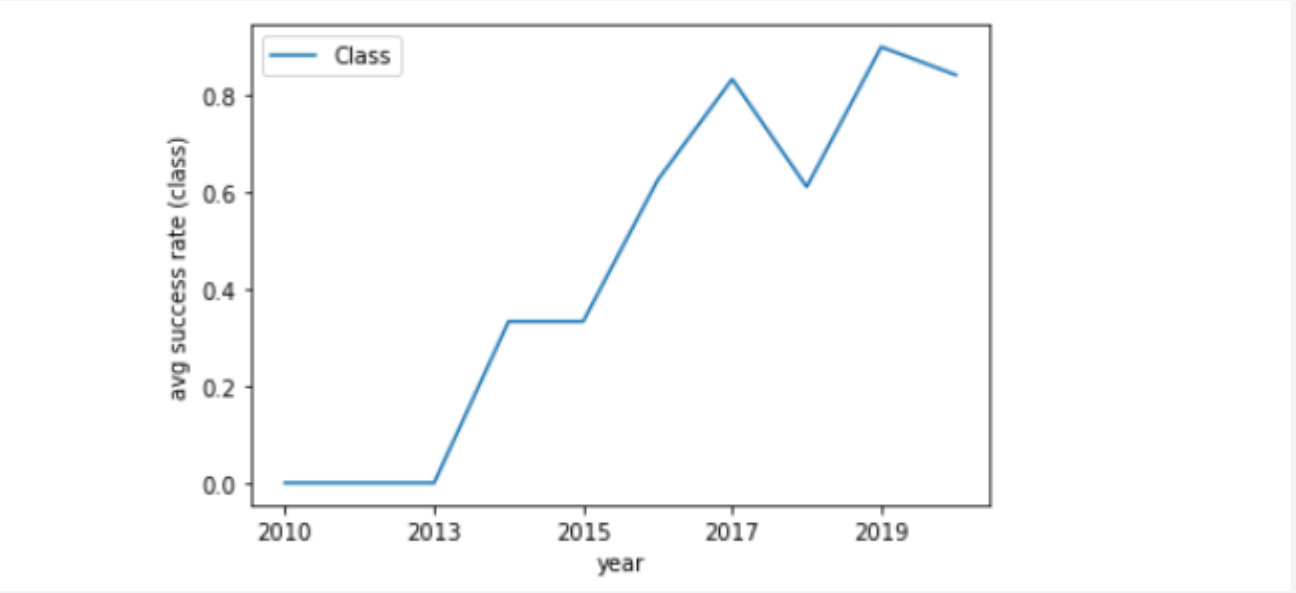
- Show the screenshot of the scatter plot with explanations
  - VLEO has higher payload mass than other orbits
  - For LEO, ISS, PO orbit types, payload increases, success rate decreases.



Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations
  - Sucess rate since 2013 kept increasing till 2020



All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Task 1

Display the names of the unique launch sites in the space mission

In [8]:
```
%sql SELECT DISTINCT(LAUNCH_SITE) as "Unique_LAUNCH_SITE" from SPACEXDATASET
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[8]:

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

Launch Site Names Begin with 'CCA'

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

Task 2

Display 5 records where launch sites begin with the string 'CCA'

In [20]:
```
%sql SELECT * from SPACEXDATASET WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[20]:

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

In [14]:
```
%sql SELECT SUM(payload_mass__kg_) from SPACEXDATASET WHERE customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[14]:

1
45596

Average Payload Mass by F9 v1.1

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

Task 4

Display average payload mass carried by booster version F9 v1.1

In [15]:
```
%sql SELECT AVG(payload_mass__kg_) from SPACEXDATASET WHERE BOOSTER_VERSION LIKE 'F9 v1.1%'
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[15]:

1
2534

First Successful Ground Landing Date

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

Task 5

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In [17]:
```
%sql SELECT MIN(DATE) from SPACEXDATASET where LANDING__OUTCOME = 'Success (ground pad)'
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[17]:

1
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

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [18]:
```
%sql SELECT DISTINCT(BOOSTER_VERSION) from SPACEXDATASET where LANDING__OUTCOME = 'Success (drone ship)' AND (PAYLOAD_MASS__KG_ b
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[18]:

```
ATASET where LANDING__OUTCOME = 'Success (drone ship)' AND (PAYLOAD_MASS__KG_ between 4000 and 6000)
```

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Task 7 ¶

List the total number of successful and failure mission outcomes

In [25]:
```
%sql SELECT COUNT(MISSION_OUTCOME) from SPACEXDATASET where MISSION_OUTCOME like 'Success%' OR MISSION_OUTCOME like 'Failure%'
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[25]:

1
101

Boosters Carried Maximum Payload

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

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [29]:
```
%sql SELECT DISTINCT(BOOSTER_VERSION) from SPACEXDATASET where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXDATA
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[29]:

booster_version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

Task 9

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
[32]: %sql SELECT LANDING__OUTCOME,BOOSTER_VERSION,LAUNCH_SITE from SPACEXDATASET where LANDING__OUTCOME = 'Failure (drone ship)' and Y
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

```
SPACEXDATASET where LANDING__OUTCOME = 'Failure (drone ship)' and YEAR(DATE) = '2015'
```

ut[32]:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Task 10

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

In [46]:
```
%sql SELECT LANDING__OUTCOME, COUNT(*) as Count FROM SPACEXDATASET where (DATE BETWEEN '2010-06-04' AND '2017-03-20') group by LA
```

```
 * ibm_db_sa://xyh74690:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[46]:

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

```
ere (DATE BETWEEN '2010-06-04' AND '2017-03-20') group by LANDING__OUTCOME order by Count DESC
```

Section 3

Launch Sites Proximities Analysis

Folium Map - all launch sites' location markers

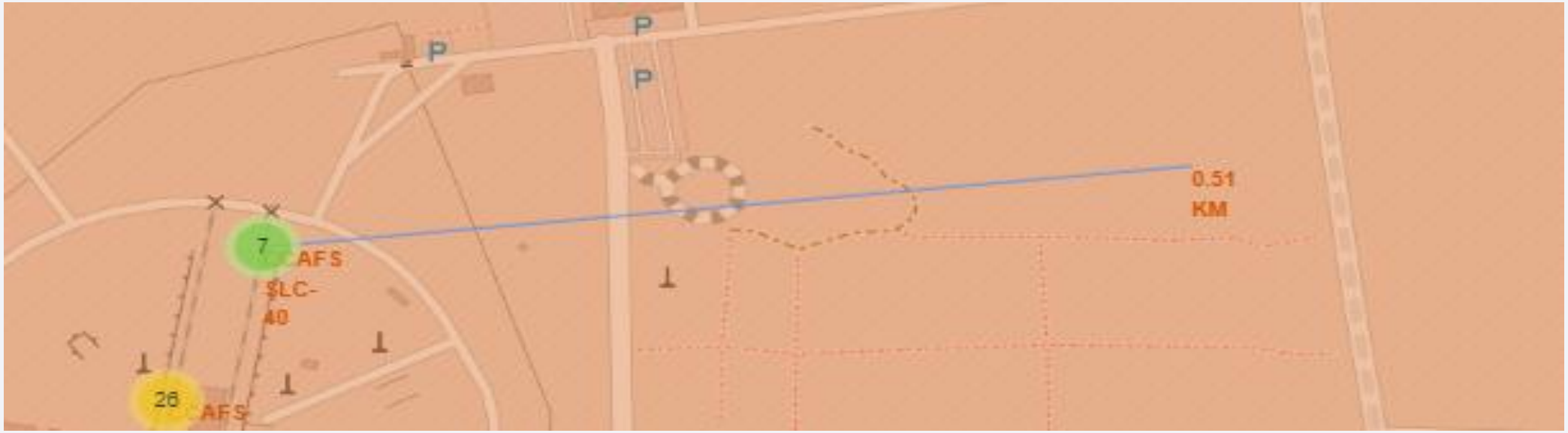
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

- Explain the important elements and findings on the screenshot
  - VAFB SLC-4E is on west coast
  - All other sites are on east coast

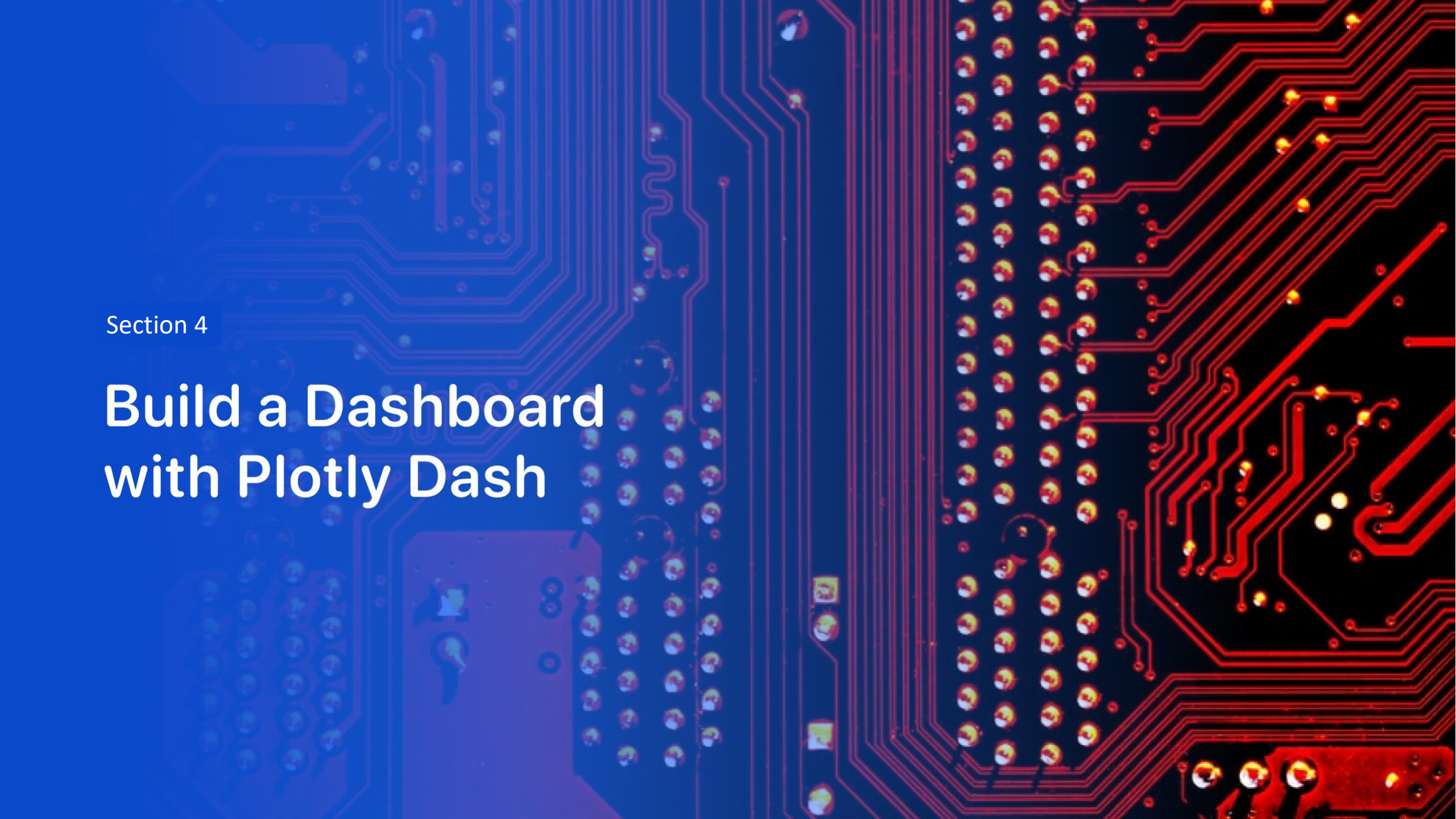
Folium Map - all launch sites' location markers and color-labeled launch outcomes on the map

- Replace <Folium map screenshot 2> title with an appropriate title

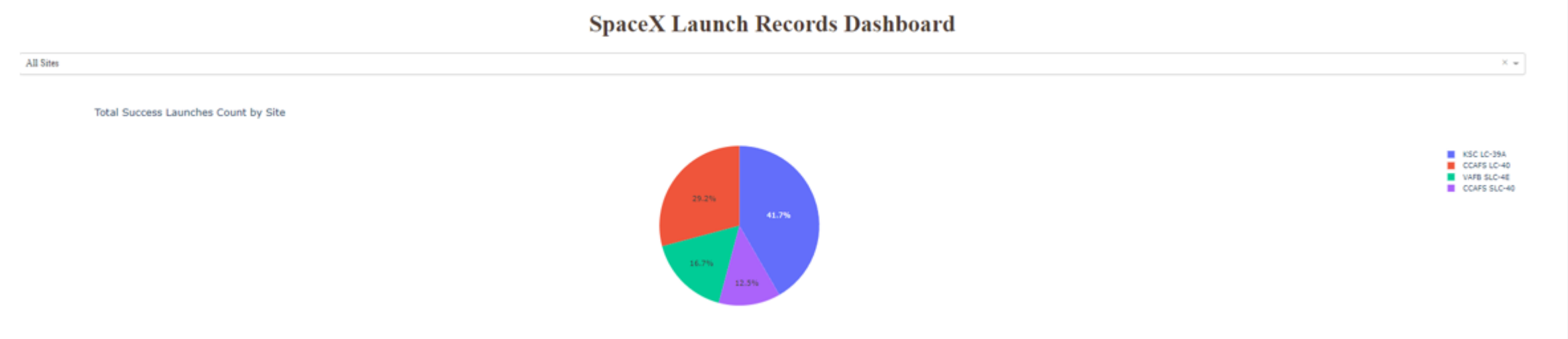
Folium Map with distance indication between SLC-40 Site to Coastline



- Explain the important elements and findings on the screenshot
  - 0.51 km between SLC-40 and coastline

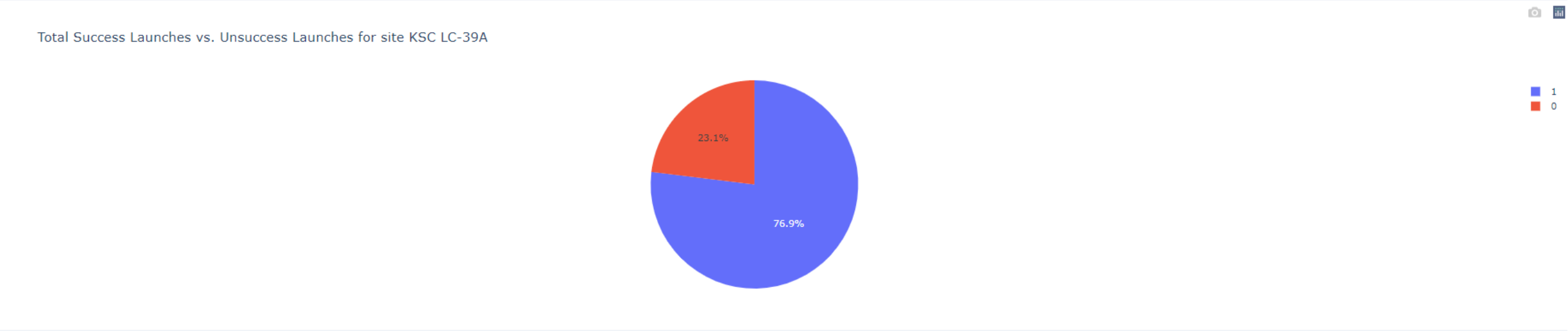

Section 4
Build a Dashboard with Plotly Dash

Launch Success Count in a Pie Chart for all Sites



- Explain the important elements and findings on the screenshot
  - LC-39A has the highest success count

Piechart for the launch site with highest launch success ratio



- Explain the important elements and findings on the screenshot
  - Success Rate for LC-39A is 76.9%

Launch Outcome vs. Payload for all sites

- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.
  - FT seems to have the highest success rate from the chart

Section 5

Predictive Analysis (Classification)

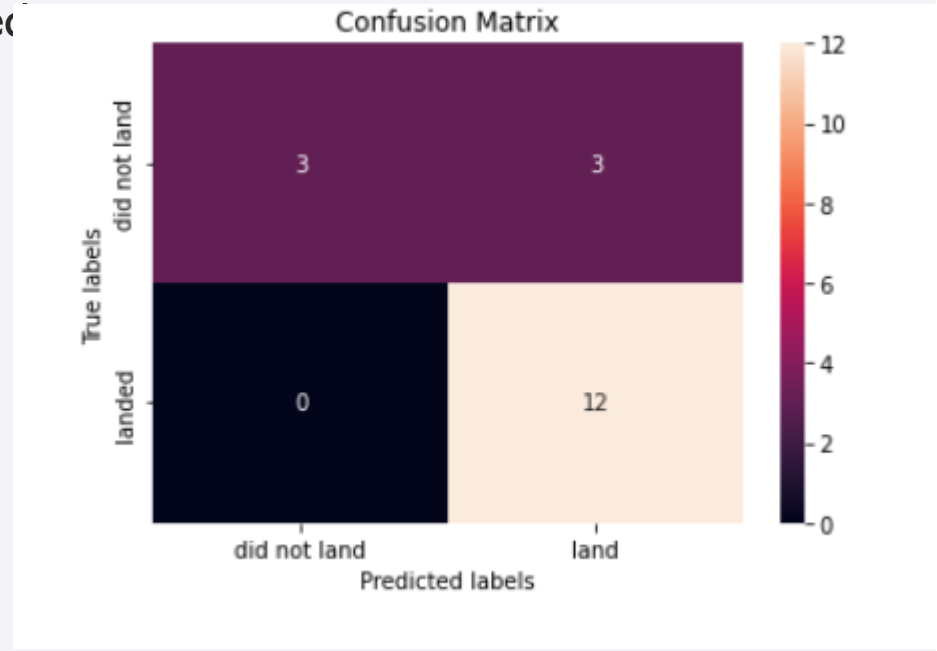
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy
  - All 4 algorithm have the same performance for the test accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
  - When true label is landed, predicted label is all correct as landed
  - When true label is not landed, predicted label is 50% correct.



Conclusions

- Flight number increase, success rate increases as well.
- ES-L1, GEO, HEO, SSO has very high success rate; GTO has very low success rate
- VLEO has data for very high flight number.
- For LEO, ISS, PO, GTO, the success rate is higher when flightnumber is higher.
- Sucess rate since 2013 kept increasing till 2020
- LC-39A has the highest success count, Success Rate for LC-39A is 76.9%
- Booster version FT seems to have the highest success rate
- All 4 models have the same performance for the test accuracy
- For confusion matrix:
    - When true label is landed, predicted label is all correct as landed
    - When true label is not landed, predicted label is 50% correct

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!