IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Kei Kondo>

<2022. 12.7>

Outline

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

Executive Summary

- Summary of methodologies:
    - Get data from webs and make a dataframe
    - Get some insights by EDA like data wrangling and visualizing
    - Use Machine learning and make a prediction model

- Summary of all results
    - Success launch keep increasing by years.
    - There are the highest success rate launch site and he highest success rate payloadmass range.
    - Decision Tree is the most accurate model.

Introduction

Context:

- 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.

Purpose:

To predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

Methodology

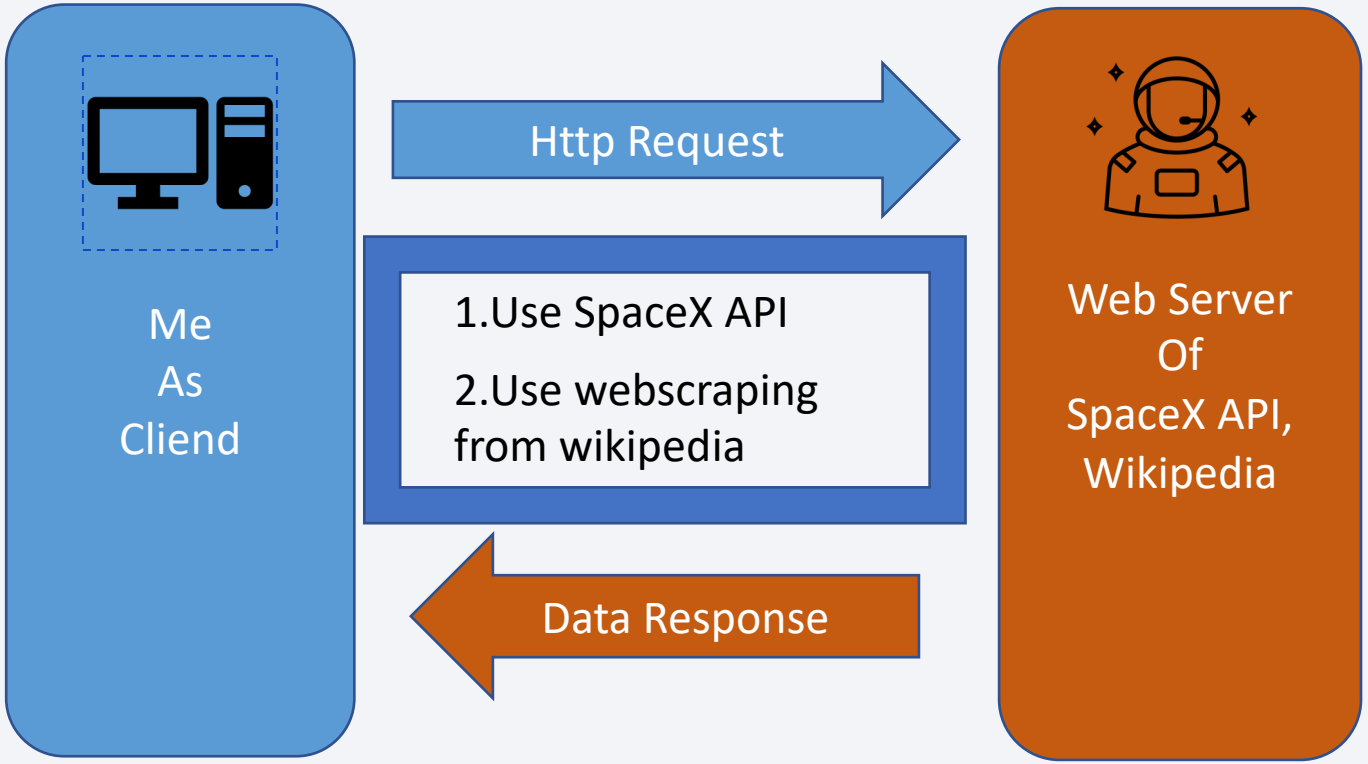
Executive Summary

- 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

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

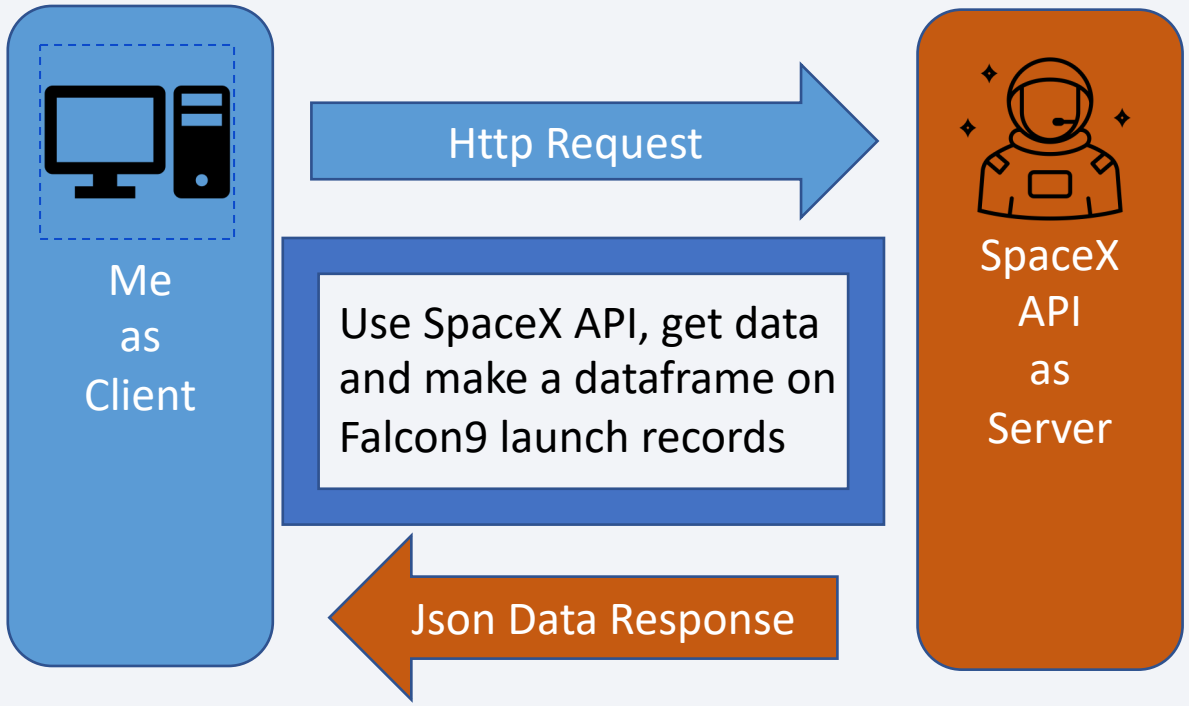
Data collection from web server by REST API and Web Scraping



Data Collection – SpaceX API

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

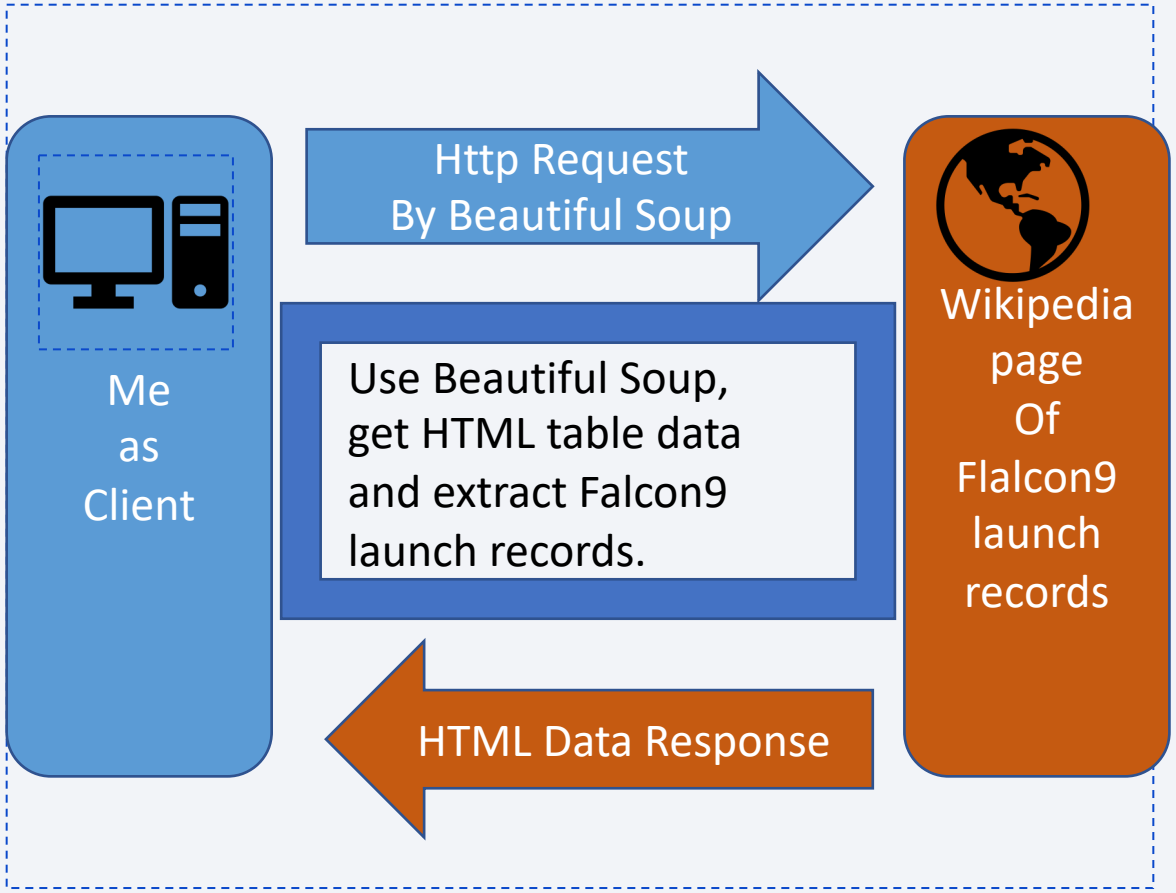
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>" 1_jupyter-labs-spacex-data-collection-api.ipynb"

Data Collection - Scraping

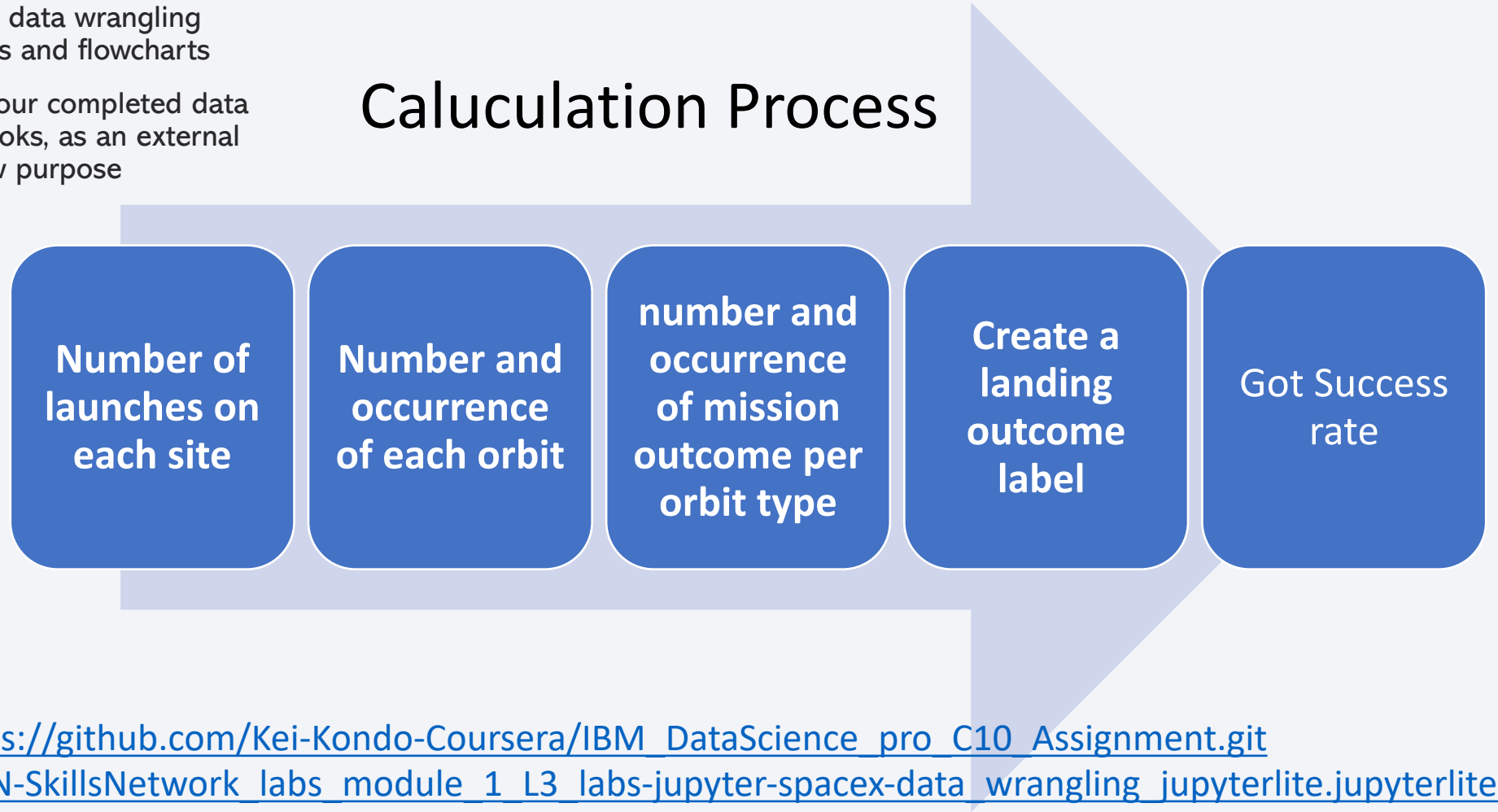
- Present your web scraping process using key phrases and flowcharts
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>"2_jupyter-labs-webscraping.ipynb"

Data Wrangling

- Describe how data were processed
- 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



Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>"3_IBM-DS0321EN-SkillsNetwork_labs_module_1_L3_labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb"

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose

Used plots types and what to see

Scatterplot

- FlightNumber vs PayloadMass
- FlightNumber vs LaunchSite
- LaunchSite vs PayloadMass
- Orbit vs FlightNumber
- Orbit vs PayloadMass

Barplot

- Sucess rate of each orbit

Lineplot

- Success rate for each year and its trend

Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>" 5_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb"

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Performed SQL

- Unique launch sites
- Launch sites begin with the string 'CCA
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- First successful landing outcome in ground pad was achieved
- Booster names which have success in drone ship and have 4000 < payload mass <6000
- Total number of successful and failure mission outcomes
- Booster versions which have carried the maximum payload mass
- Failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015
- Count of successful landing outcomes between the date 04-06-2010 and 20-03-2017

Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>" 4_jupyter-labs-eda-sql-coursera_sqllite.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
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Added map objects and its purposes

Circles	Markers	Line
• Mark all launch sites on a map	• Mark the success/failed launches for each site on the map	• Calculate the distances between a launch site to its proximities

Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>" 6_lab_jupyter_launch_site_location.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 the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Added Plots and what to see

Pie chart

- Success rate for all launch by each site
- Success rate for each launch site by interaction

Scatterplot

- Payloadmass and Success/Failure for boosters
- Add payloadmass interaction slider

Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>" 7_spacex_dash_app.py"

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Classification modeling process

Data Cleaning

- Set prediction target to class(Launch Success or not, binary)
- Normalize by Standardscaler
- Split data to training and test data

Model Fit

- Try LogisticRegression, SVM, Decision Tree and KNN

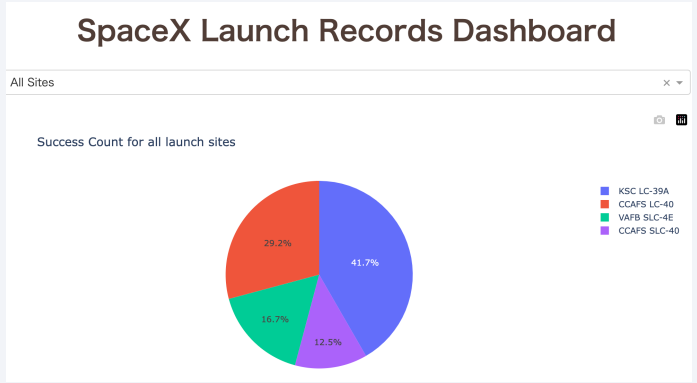
Evaluation

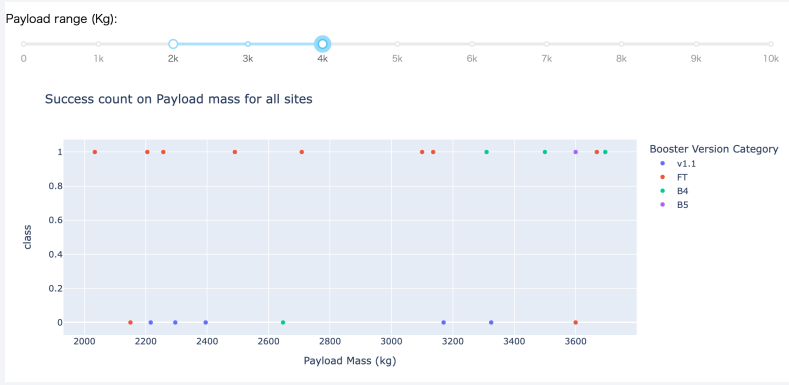
- Get Accuracy score by test data and F1 score
- Observe which model gets the highest score

Notebook URL : https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git
=>" 8_IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb"

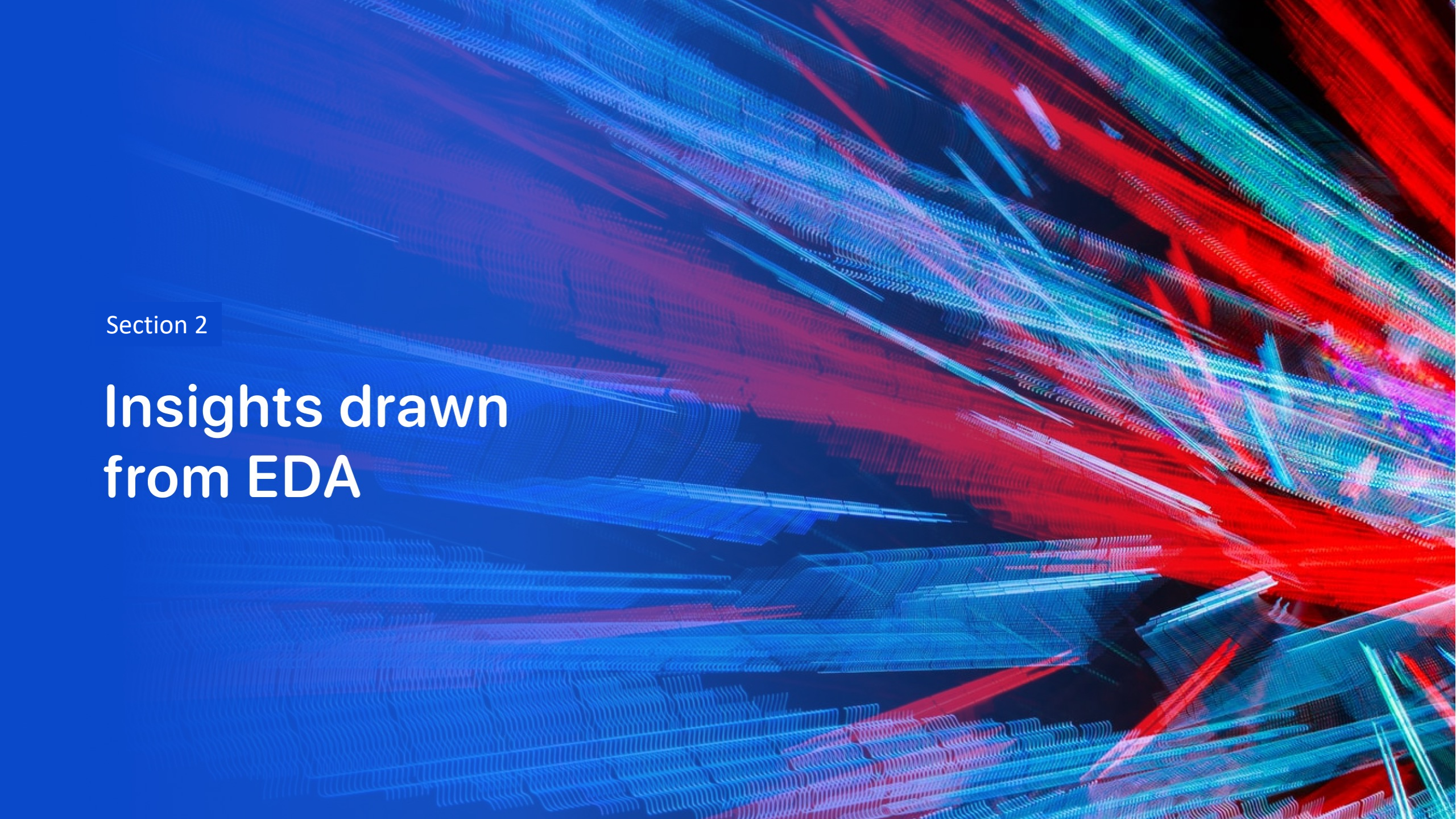
Results

- Exploratory data analysis results:
  - The highest success rate launch site is KSC LC-39A
  - The highest success rate payloadmass is in the range of 2000kg and 4000kg
- Interactive analytics demo in screenshots:

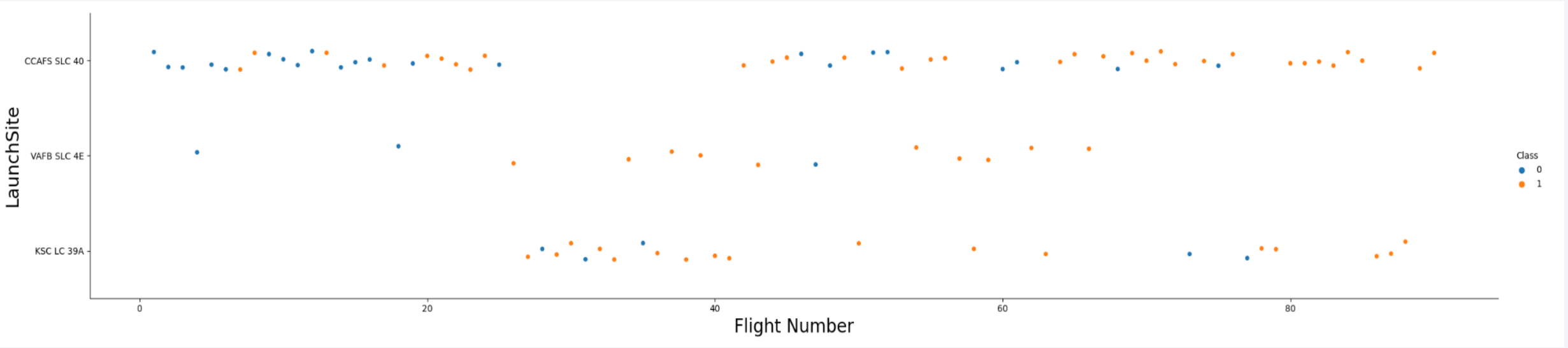




- Predictive analysis results:
  - Decision Tree is the best model this time.


Section 2
Insights drawn from EDA

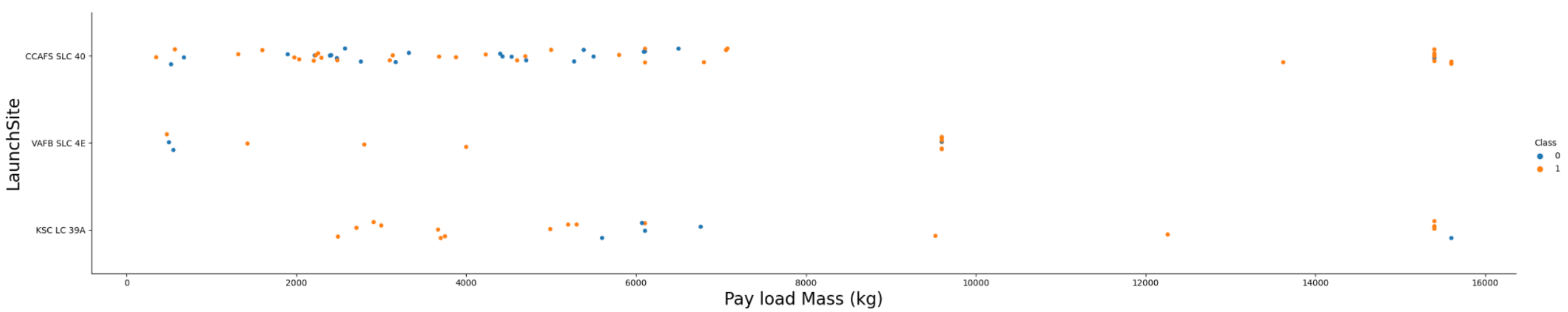
Flight Number vs. Launch Site



CCAFD SLC-40 is the most used site.

- Show a scatter plot of Flight Number vs. Launch Site
- Show the screenshot of the scatter plot with explanations

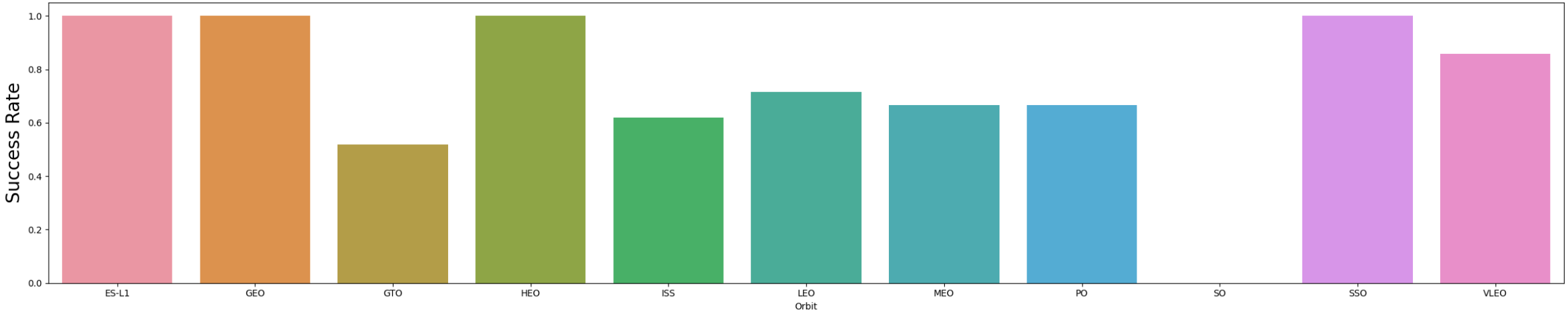
Payload vs. Launch Site



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

- Show a scatter plot of Payload vs. Launch Site
- Show the screenshot of the scatter plot with explanations

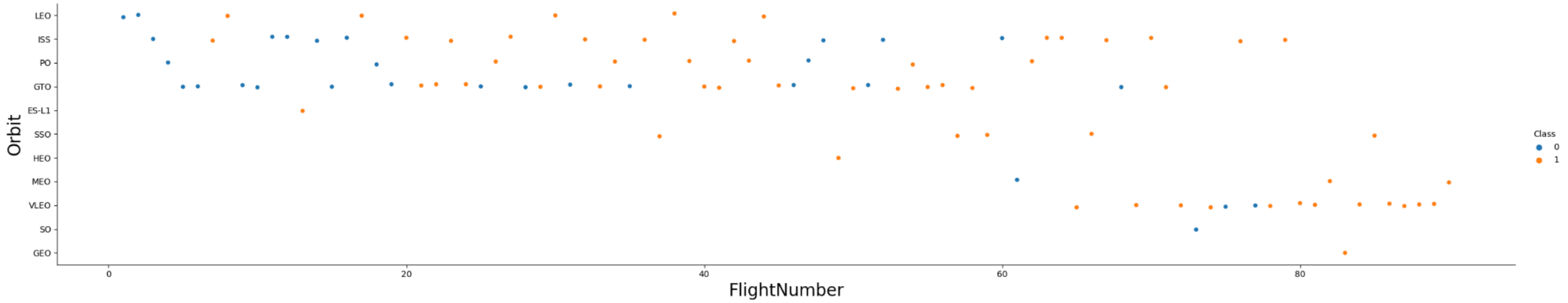
Success Rate vs. Orbit Type



Success rates of some orbit types are higher .

- Show a bar chart for the success rate of each orbit type
- Show the screenshot of the scatter plot with explanations

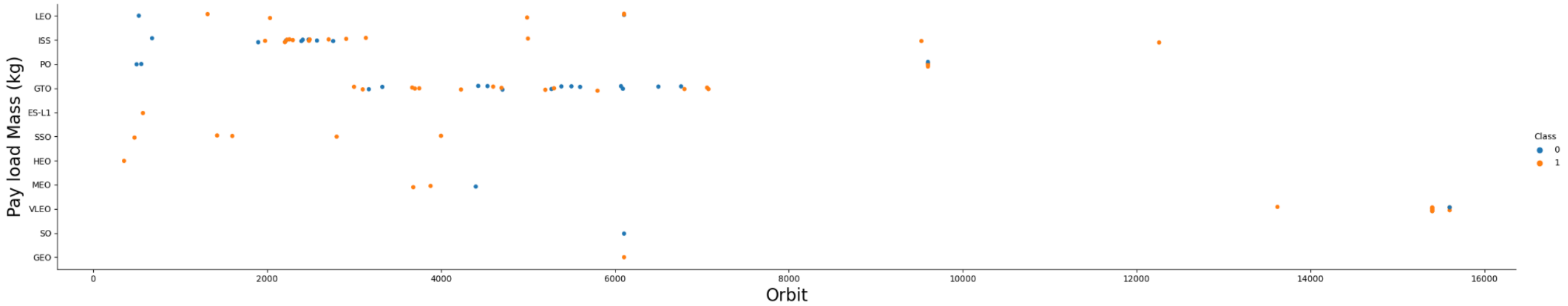
Flight Number vs. Orbit Type



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.

- Show a scatter point of Flight number vs. Orbit type
- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type

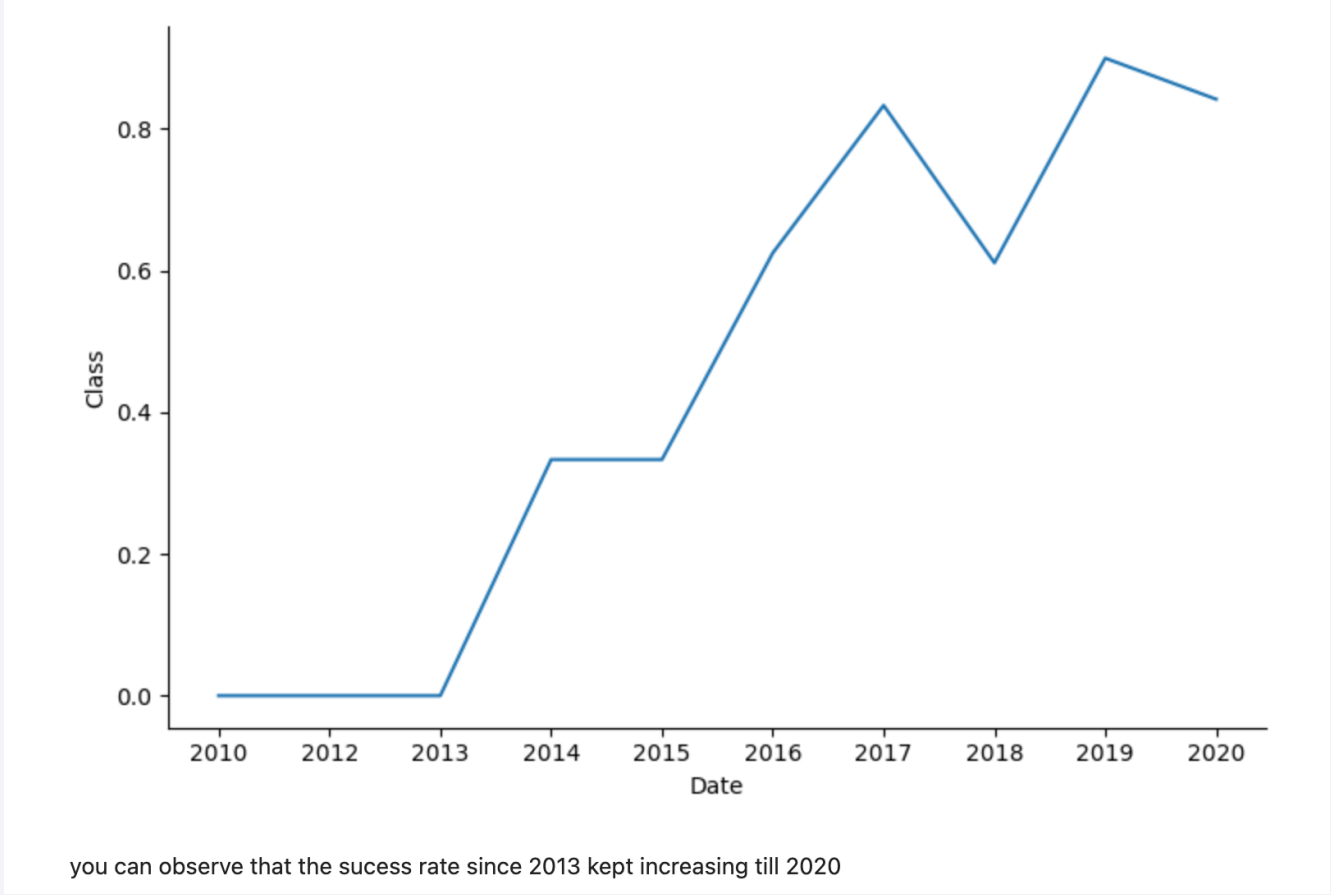


With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

- Show a scatter point of payload vs. orbit type
- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

- Show a line chart of yearly average success rate
- Show the screenshot of the scatter plot with explanations

All Launch Site Names

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

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

These are all Launch Sites (Unique Records from the data)

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

5 records where launch sites begin with `CCA`

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

SUM(PAYLOAD_MASS__KG_)	Customer
45596	NASA (CRS)

Total payload carried by boosters from NAS

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

AVG(PAYLOAD_MASS__KG_)	Booster_Version
2928.4	F9 v1.1

Average payload mass carried by booster version F9 v1.1

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
22-12-2015	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

The dates of the first successful landing outcome on ground pad

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
06-05-2016	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
14-08-2016	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
11-10-2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Count(Date)	Mission_Outcome
98	Success

Total number of successful mission outcomes

Count(Date)	Mission_Outcome
1	Failure (in flight)

Total number of failure mission outcomes

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

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

Names of the booster which have carried the maximum payload mass

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Present your query result with a short explanation here

Date	Month	Landing _Outcome	Booster_Version	Launch_Site
10-01-2015	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
14-04-2015	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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
- Present your query result with a short explanation here

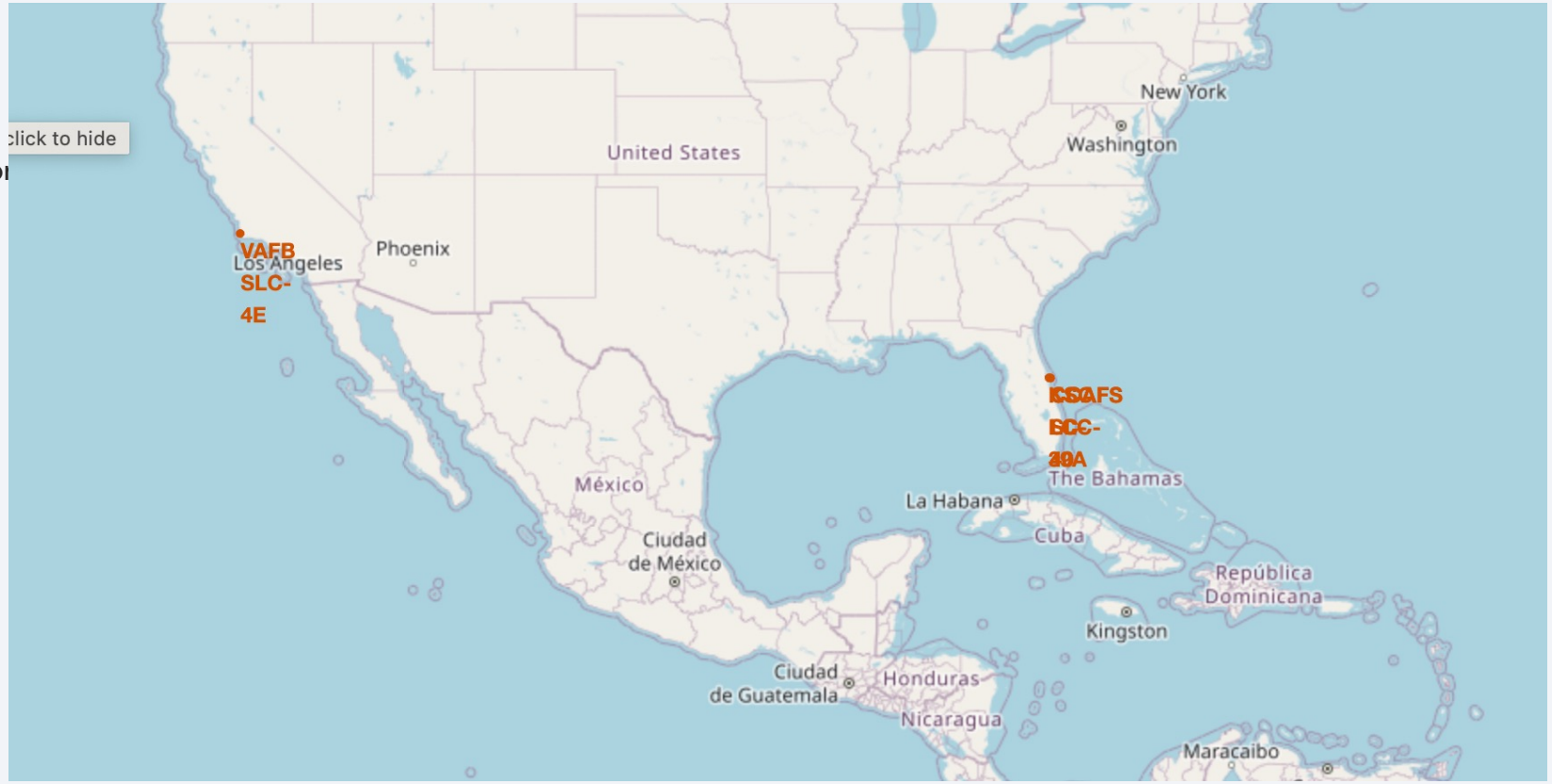
Landing _Outcome	COUNT
Success	20
No attempt	10
Success (drone ship)	8
Success (ground pad)	6
Failure (drone ship)	4
Failure	3
Controlled (ocean)	3
Failure (parachute)	2
No attempt	1

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


Section 3
Launch Sites
Proximities Analysis

< All Launch Sites' Location >

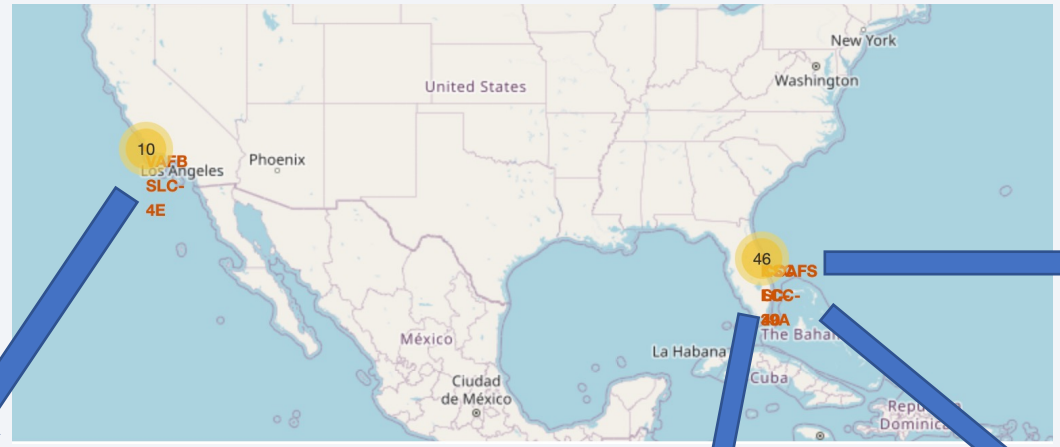
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

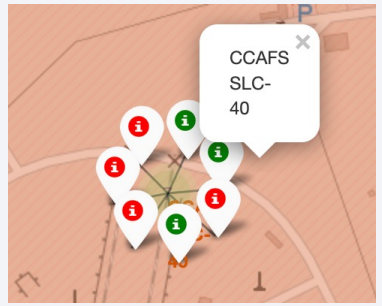


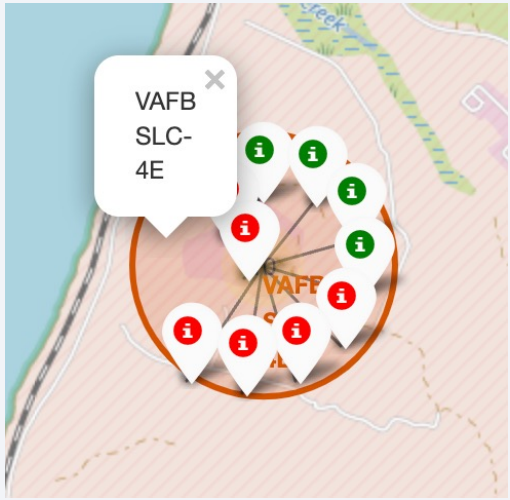
- Launch Sites are in two areas of Texas or Los Angels near the coast lines.

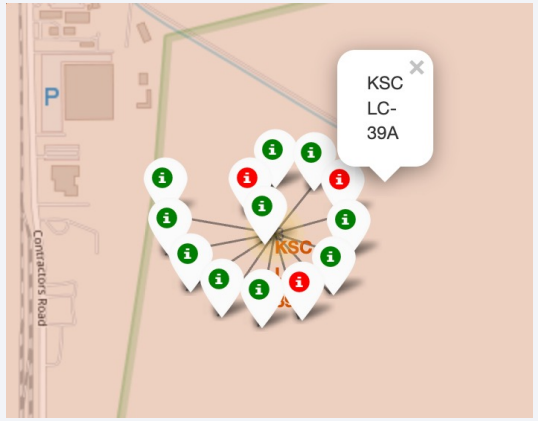
<Color-labeled Launch Outcomes on the Map>

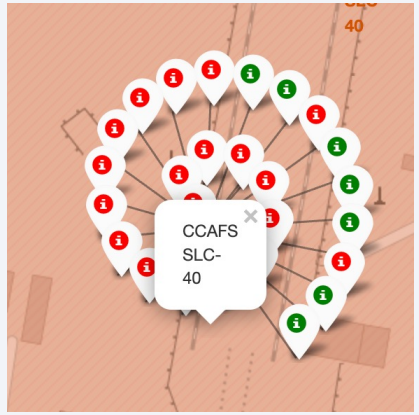
- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot







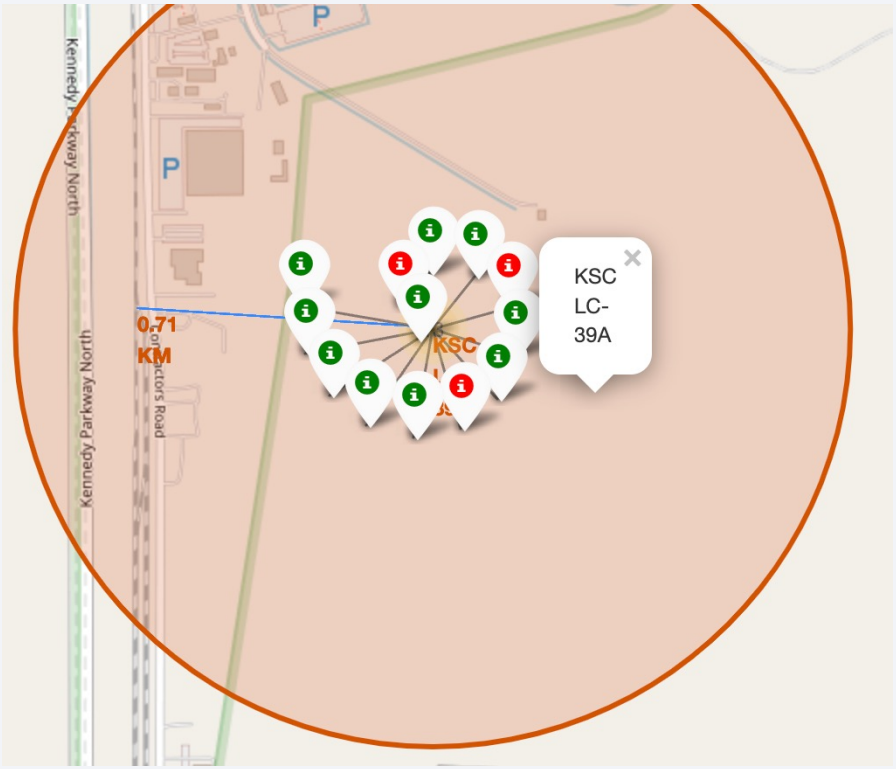




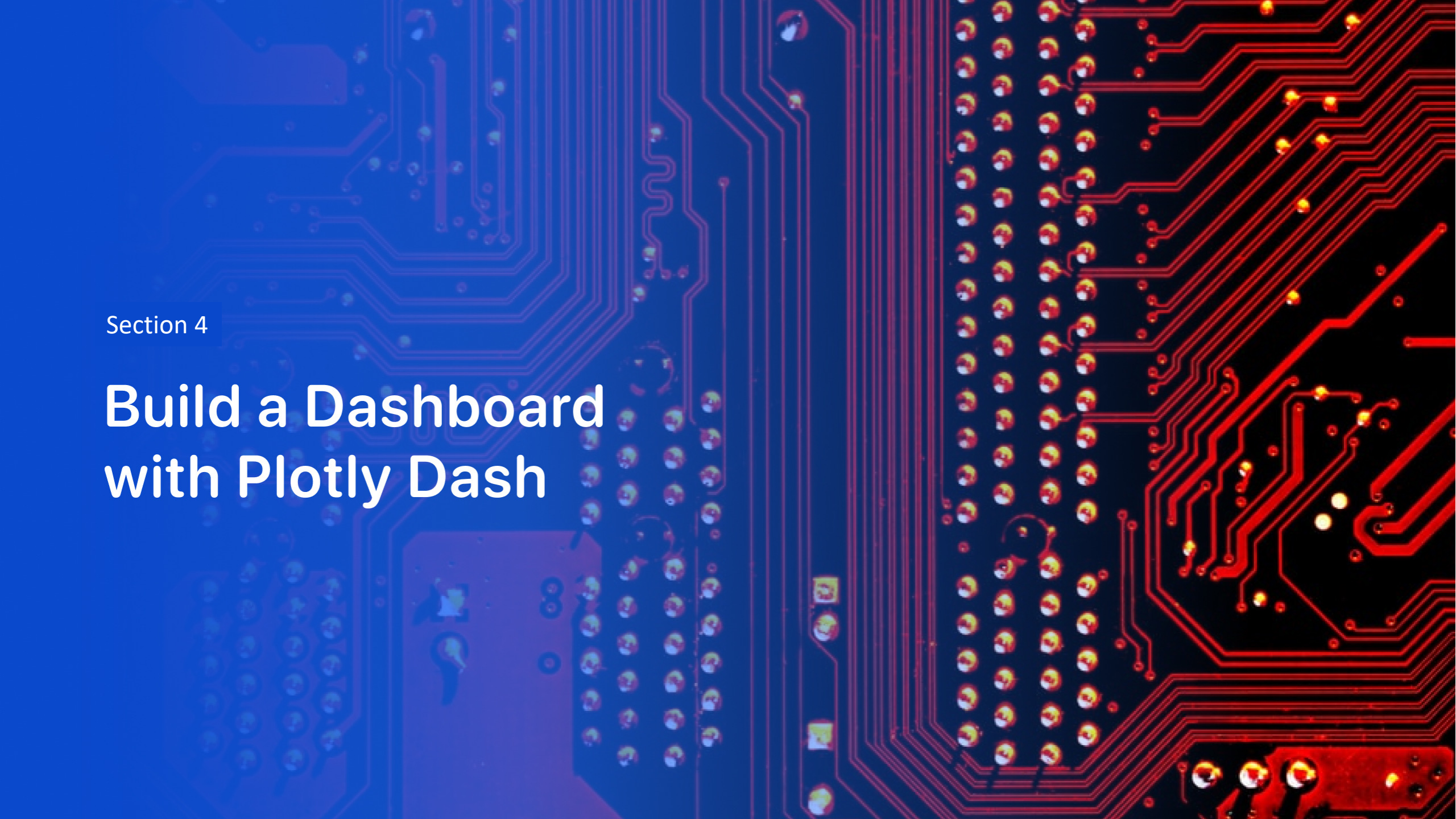
The Launch Site “KSC LC-39A” has the highest success rate of the launch outcome.

<Launch Site Proximity>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot

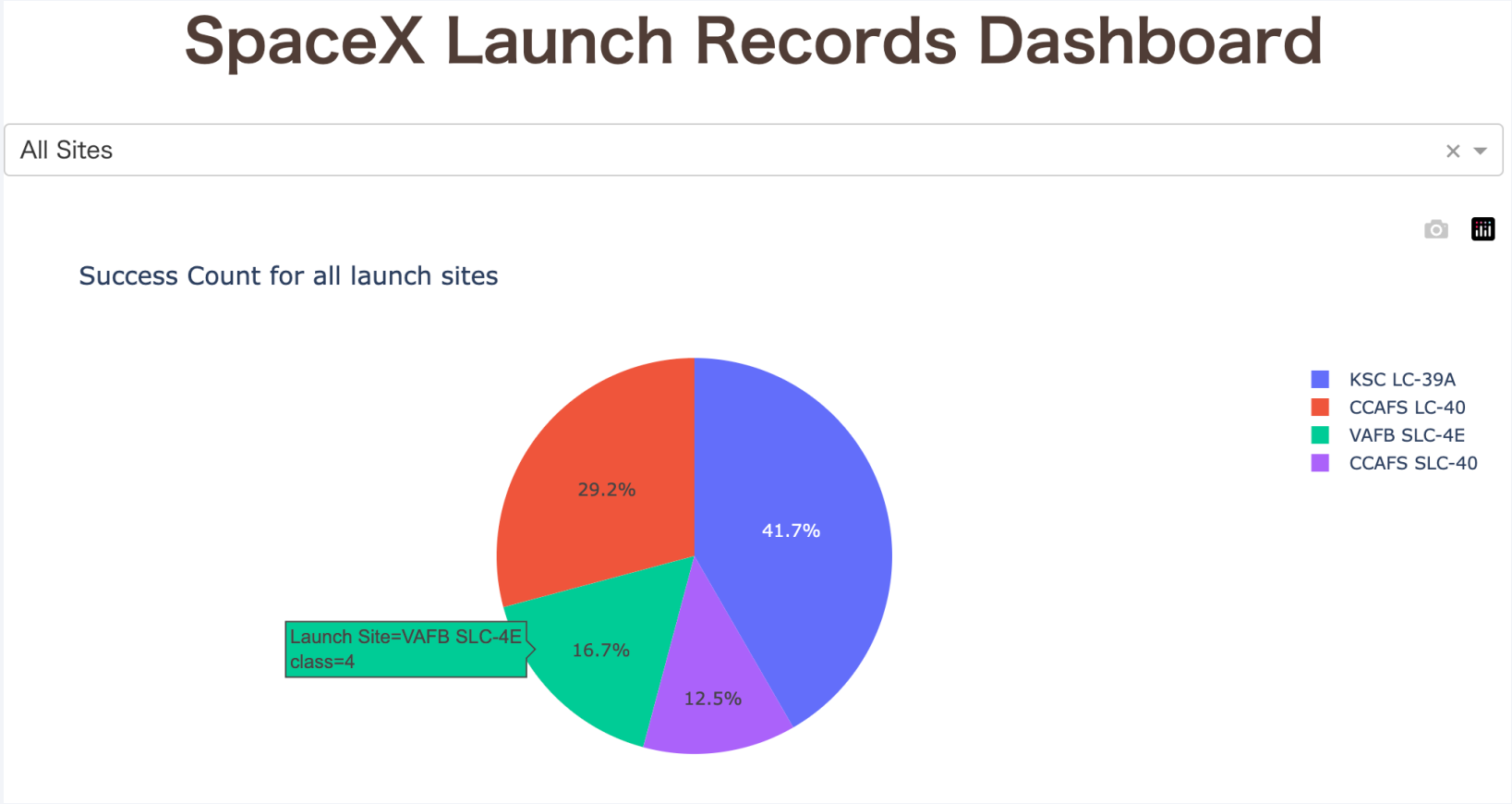


KSC LC-39A, the highest success rate launch site, is at 0.71km from the nearest railway.


Section 4
Build a Dashboard with Plotly Dash

<Sccess Count for all Launch Sites>

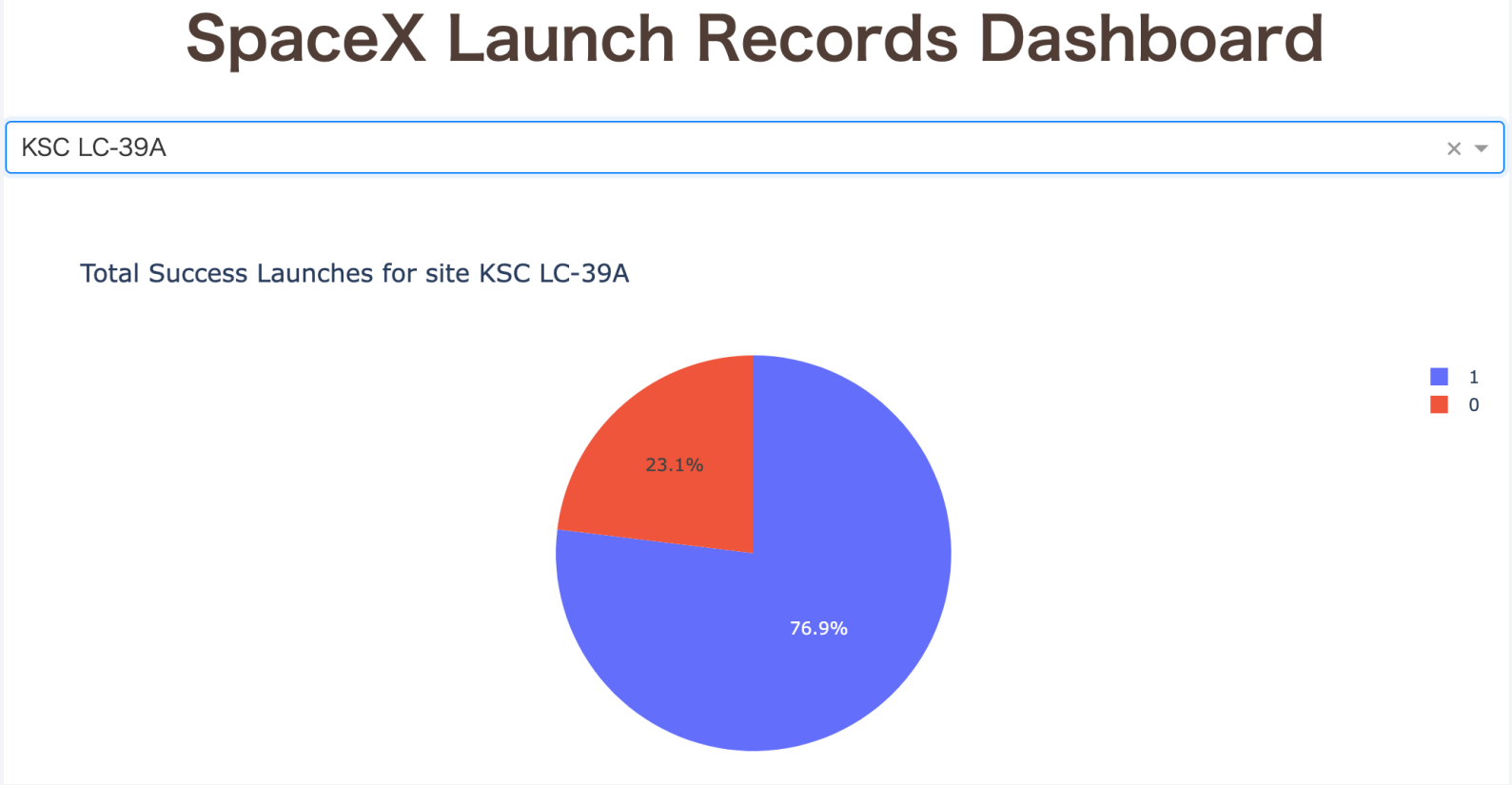
- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot



KSC LC-39A is the highest success count site.

<Total Success Launches for KSC KSC LC-39A>

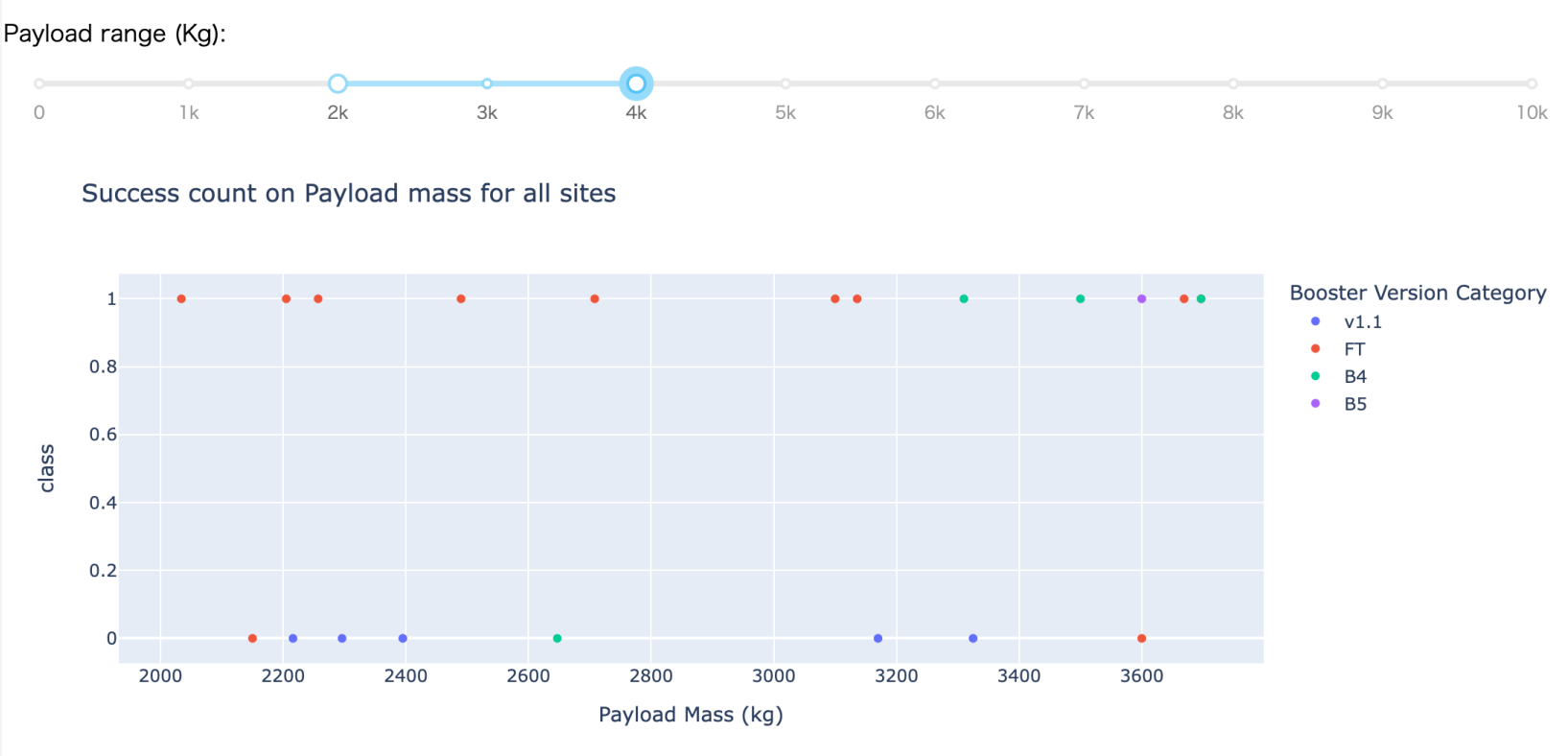
- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot



Success Rate is 76.9 % at KSC LC-39A

<Success Count on Payload Mass for all Sites>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.



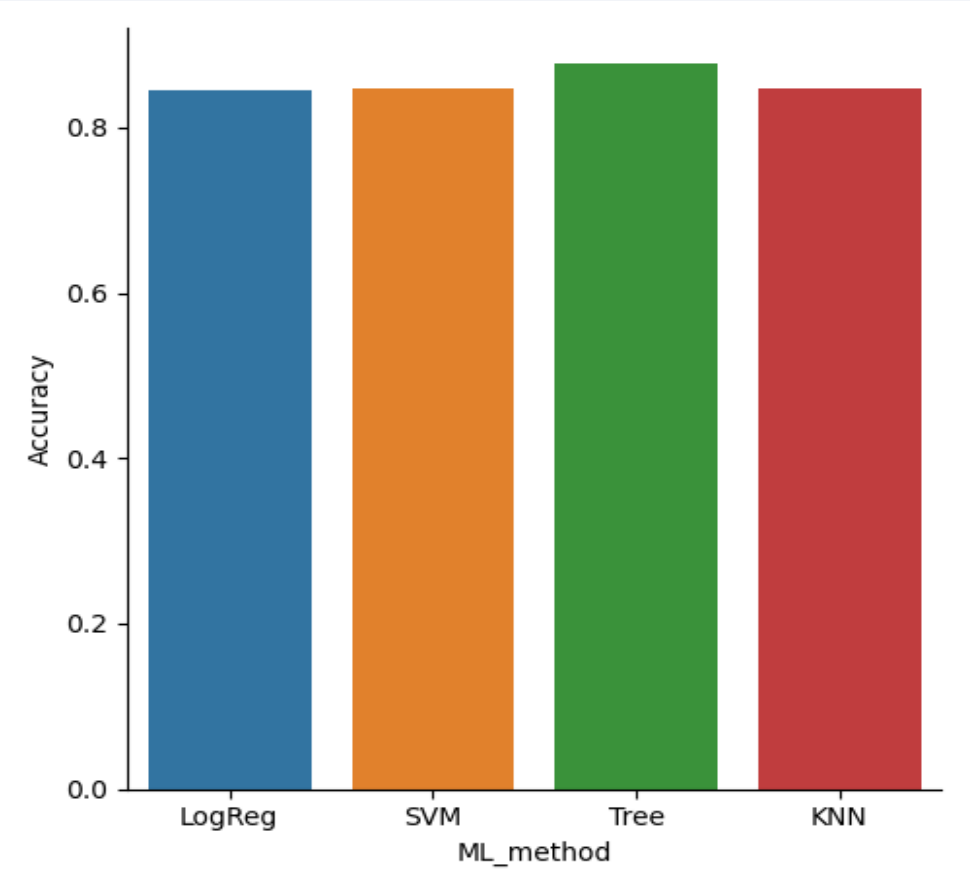
Success Rate is high in the Payload range of 2000-4000 by FT booster.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

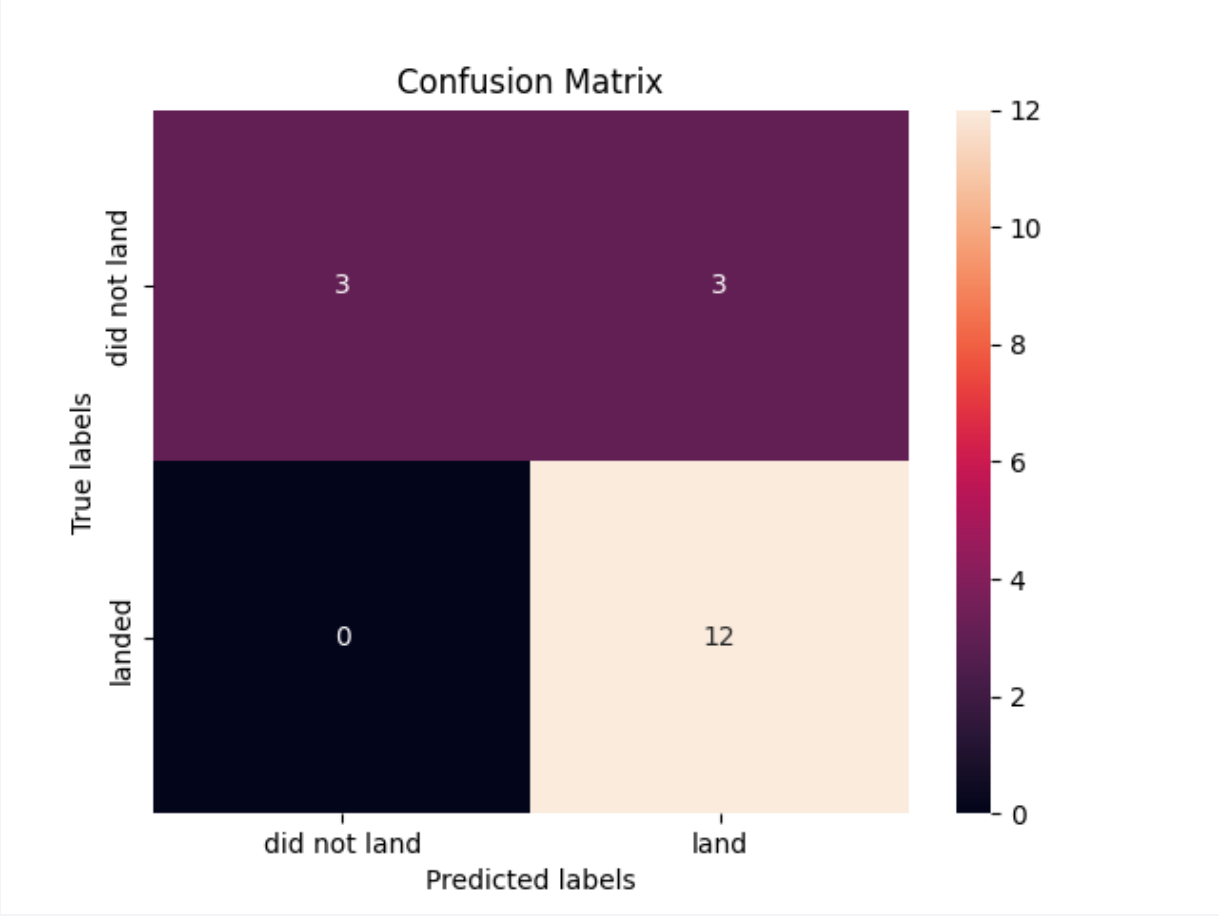
- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy



Decision Tree marks the highest accuracy, 87.7%.

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



True Positive:12
True Negative:3
False Positive:3
False Negative:0

=>F1 score : 0.78

Confusion Matrix of Decision Tree

Conclusions

- Success rate is increasing by years
- Success launcn is increasing as flight trial increases
- Some orbit types are inclined to success
- KSC LC-39A is the highest success rate launch site
- Too heavy rocket fails. The range of better payload mass is more than 2000kg and less than 4000kg
- The best prediction model is decision tree this time and its accuracy is 87.7%

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

- All codes are in the github. See the URL.

  https://github.com/Kei-Kondo-Coursera/IBM_DataScience_pro_C10_Assignment.git

Thank you!