Applied data Science Capstone

Part 1

By Tapiwa Terrence Munodawafa

27/07/2025

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

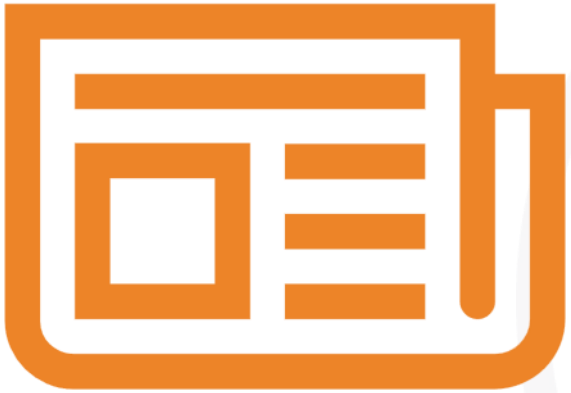
EXECUTIVE SUMMARY

- Several APIs were used for:
  - Data collection
  - Web scrapping
- EDA was performed
- ML techniques were applied

INTRODUCTION

- Prediction of Falcon 9’s first stage successful landing
- This data is crucial for performance enhancement and safety
- Exploratory Data Analysis (EDA) was also made
  - SQL
  - Visualization

METHODOLOGY



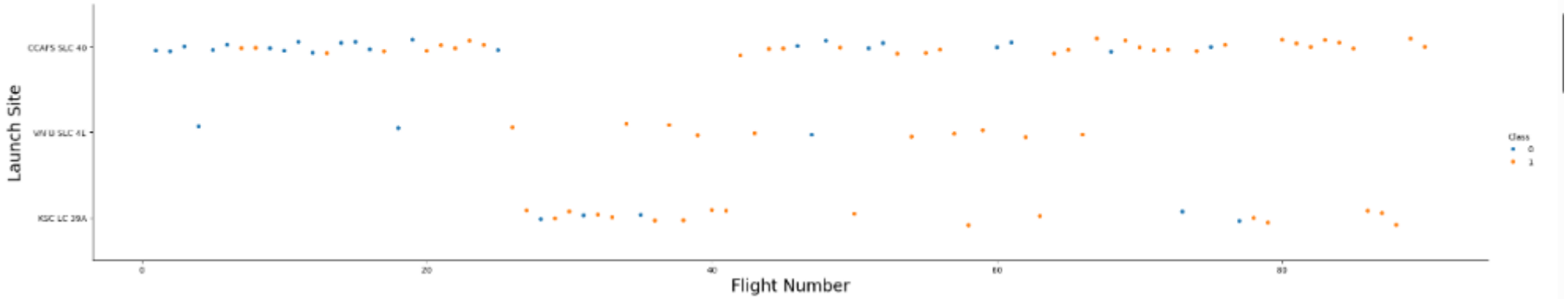
- Data collection
  - SpaceX API
  - Web scrapping
- Data wrangling
  - Dropping irrelevant columns
- EDA using SQL and visualization
- Interactive visual analytics using Plotly Dash and Folium
- Performed prediction using classification models

RESULTS – EDA Visualization

TASK 1: Visualize the relationship between Flight Number and Launch Site

Use the function `catplot` to plot `FlightNumber` vs `LaunchSite`, set the parameter `x` parameter to `FlightNumber`,set the `y` to `Launch Site` and set the parameter `hue` to `'class'`

5]:
```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the Launch site, and hue to be the class value
sns.catplot(x='FlightNumber', y='LaunchSite', hue='Class', data=df, aspect=5)
plt.xlabel('Flight Number', fontsize=20)
plt.ylabel('Launch Site', fontsize=20)
plt.show()
```



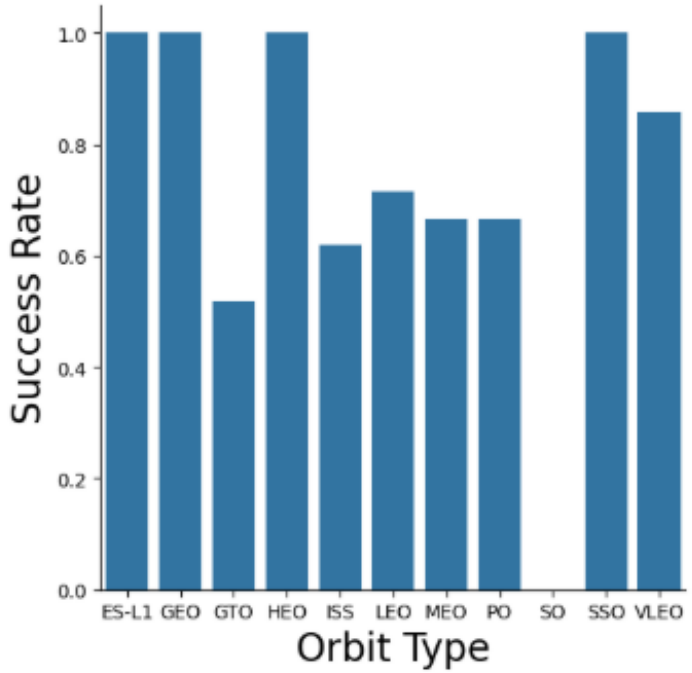
RESULTS – EDA Visualization

TASK 3: Visualize the relationship between success rate of each orbit type

Next, we want to visually check if there are any relationship between success rate and orbit type.

Let's create a `bar chart` for the sucess rate of each orbit

[7]:
```
# HINT use groupby method on Orbit column and get the mean of Class column
sns.catplot(x= 'Orbit', y = 'Class', data = df.groupby('Orbit')['Class'].mean().reset_index(), kind = 'bar')
plt.xlabel('Orbit Type',fontsize=20)
plt.ylabel('Success Rate',fontsize=20)
plt.show()
```



RESULTS – Sample SQL

Task 1

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

```
%sql select distinct(launch_site) from SPACEXTABLE
```

```
 * sqlite:///my_data1.db
Done.
```

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

RESULTS – All SQL queries

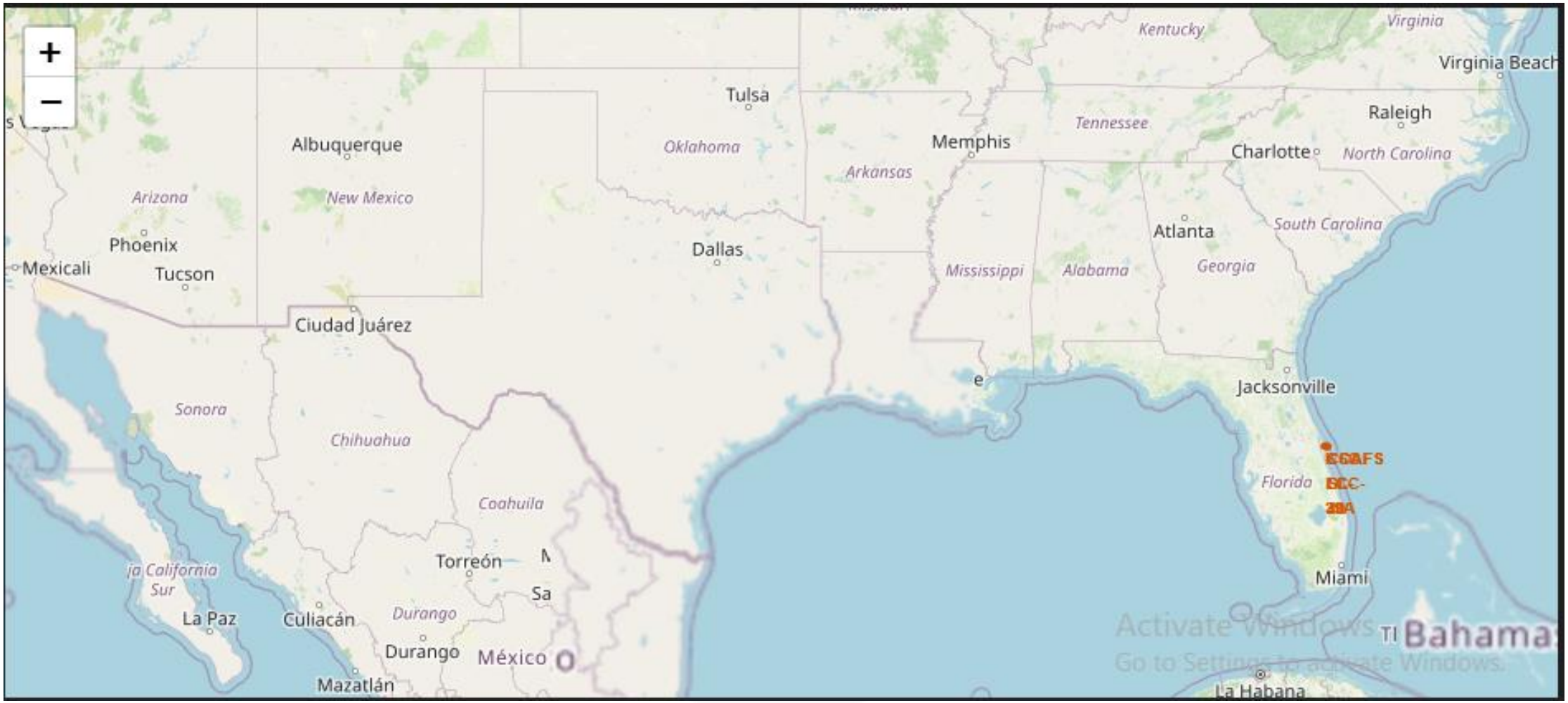
- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful 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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the

months 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/ttmunoz/Spacex-data-collection/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

RESULTS – Sample Folium



PROGRAMMING LANGUAGE TRENDS

Current Year

Next Year

PROGRAMMING LANGUAGE TRENDS - FINDINGS & IMPLICATIONS

Findings

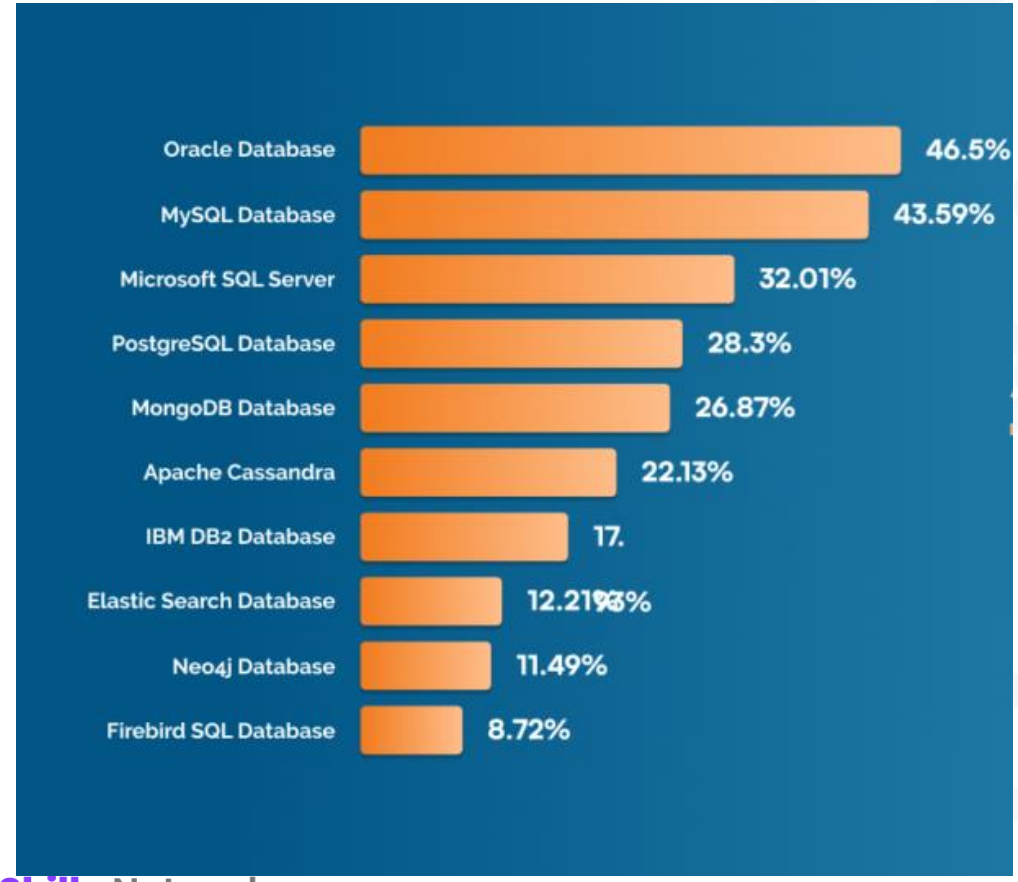
- Python will be the top
- PHP, Swift etc now not favoured
- JavaScript still favoured

Implications

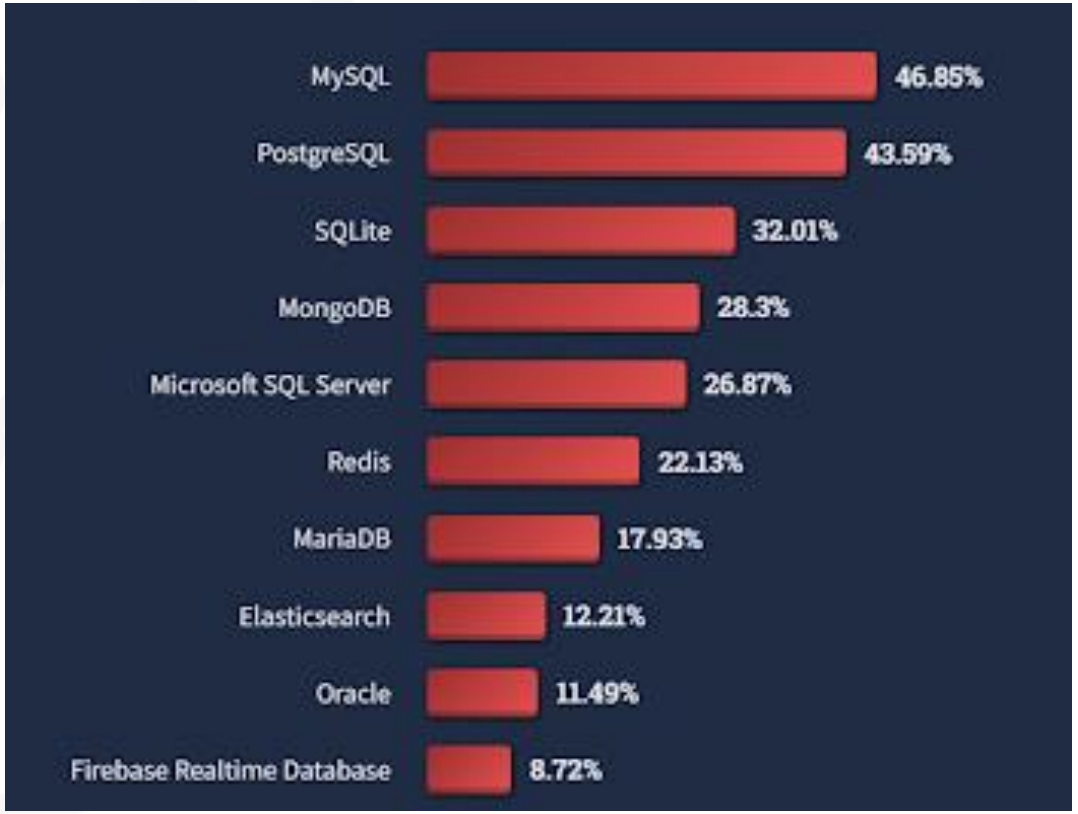
- Programmers will have to learn Python
- Programmers use alternative modern languages
- Web pages still need JS

DATABASE TRENDS

Current Year



Next Year



DATABASE TRENDS - FINDINGS & IMPLICATIONS

Findings

- MySQL will be the top
- NoSQL databases are increasing
- Less usage of Oracle

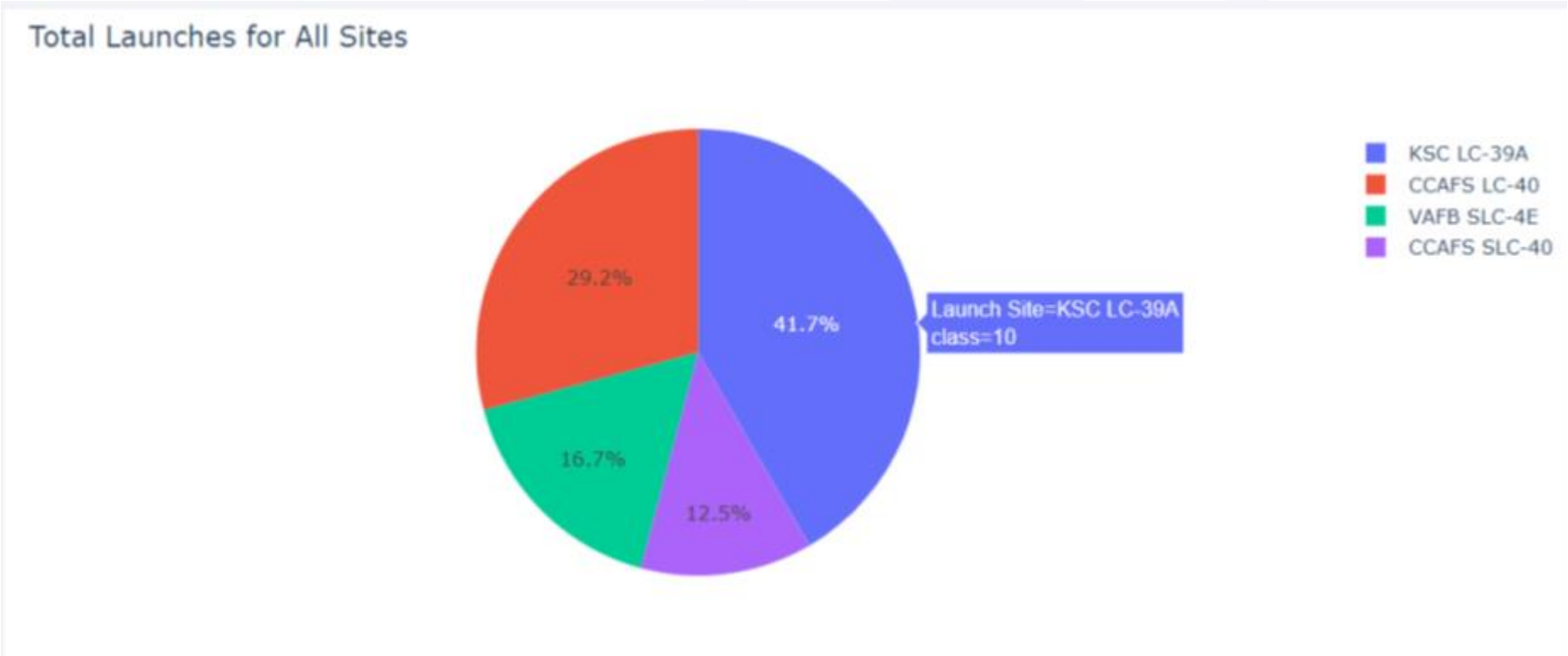
Implications

- MySQL is easy and open source so many programmers will use it
- More applications can be developed to store unstructured content
- Promotion of MySQL since it is open source

DASHBOARD

https://github.com/ttmunoz/Spacex-data-collection

DASHBOARD TAB 1



DASHBOARD TAB 2

Screenshot of dashboard tab 2 goes here

DASHBOARD TAB 3

Screenshot of dashboard tab 3 goes here

DISCUSSION

- This data analysis will help future companies and competitors to find suitable launch sites and parameters to prevent failures

OVERALL FINDINGS & IMPLICATIONS

Findings

- Best payloads not 0-6000 kg Launch Sites are close to sea
- ML work good

Implications

- Payloads of 0-6000 kg are quite risky
- It is not over populated
- More fine tuning needed on ML algorithms for best results

CONCLUSION

- SpaceX data was collected from open sources
- Best launch site among four launch sites is KSC LC-39A
- Payloads of 0-6000 kg are quite risky
- Launch Sites are close to sea, not over populated
- • All ML algorithms work as expected

APPENDIX

- Github link
- https://github.com/ttmunoz/Spacex-data-collection

JOB POSTINGS

In Module 1 you have collected the job posting data using Job API in a file named “job-postings.xlsx”. Present that data using a bar chart here. Order the bar chart in the descending order of the number of job postings.

POPULAR LANGUAGES

In Module 1 you have collected the job postings data using web scraping in a file named “popular-languages.csv”. Present that data using a bar chart here. Order the bar chart in the descending order of salary.