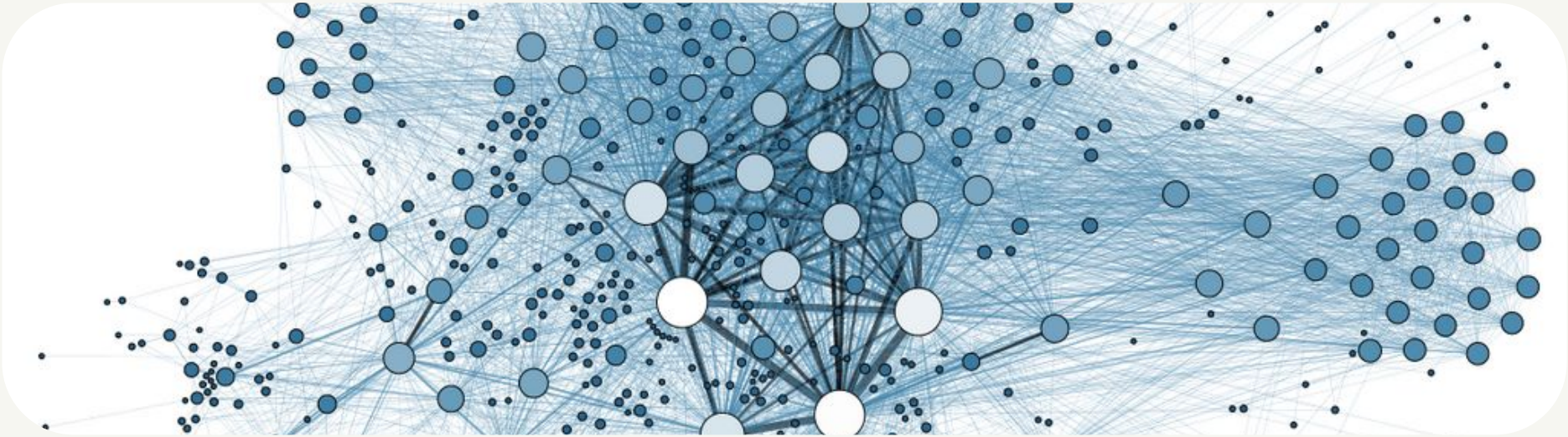


Excelerate User Data & Opportunity Wise Data Analysis

0212 DVA Team 4A

Team Members

Abdullah Imran	Akshaya Cheruku	Nwabueze Victor	Chirag Pawaskar	Omootemi Modupe
Team Lead	**Meeting Facilitator**	**Project Scribe**	**Project Manager**	**Assistant Team Lead**

Overview

1. Introduction
2. Data Visualization
3. AI Model Training
4. Dashboards
5. Conclusion

1. Intro

1.1: Dataset Overview
1.2: Data Cleaning & Validation

1.1 Dataset Over-view

1. 27,000+ user data records and 20,000+ Opportunity Wise Data records
2. Both datasets were of different instances and conveyed different meanings, hence, could not be combined
3. User Data represents User Demographics and Sponsorship Preferences
4. Opportunity Wise Data represents User Engagement and Opportunity Participation
5. Dataset had missing and duplicated data, non-related data types, making it unsuitable for analysis
6. It was necessary to clean and validate the data for a good analysis

1.2 Data Cleaning & Validation

Data Cleaning

Filling in the missing data

- Unrewarded people were filled with “no reward”

Dropping the missing data where the data could not be replaced

- Gender, Country, city, Zip Code

Dropping the duplicated data

Data Validation

Handling the data outliers through boxplots and Z scores > 3 or < -3

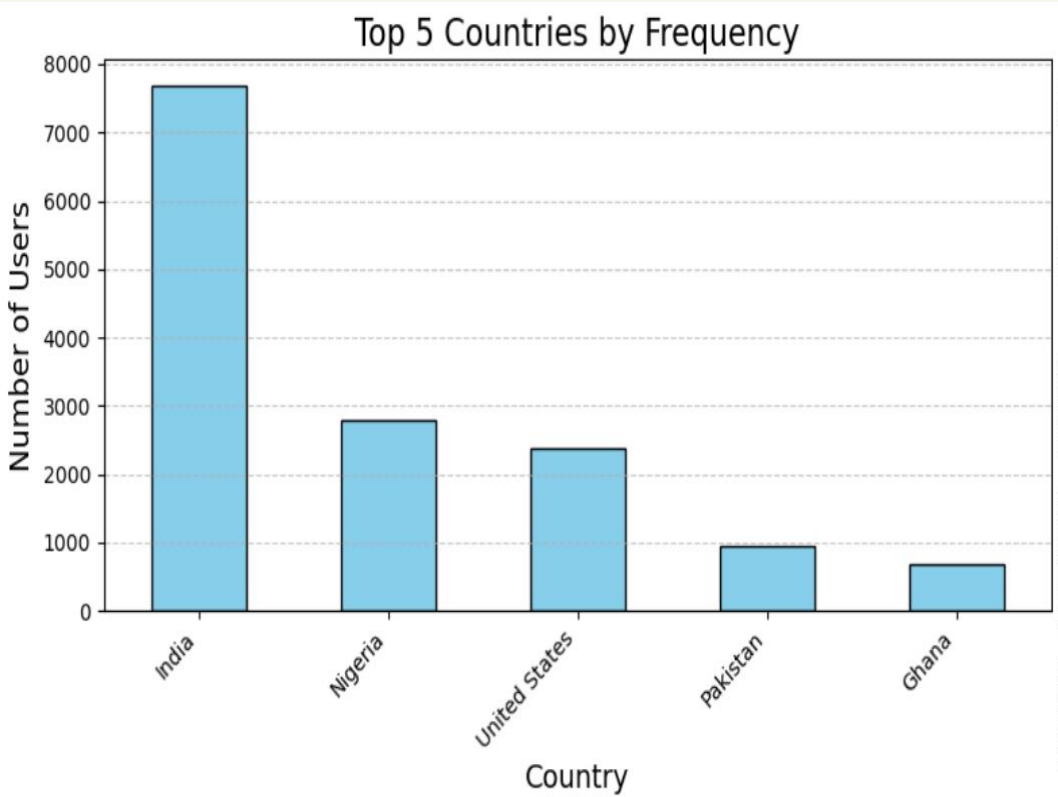
One Hot encoding the textual data and normalizing the numerical data

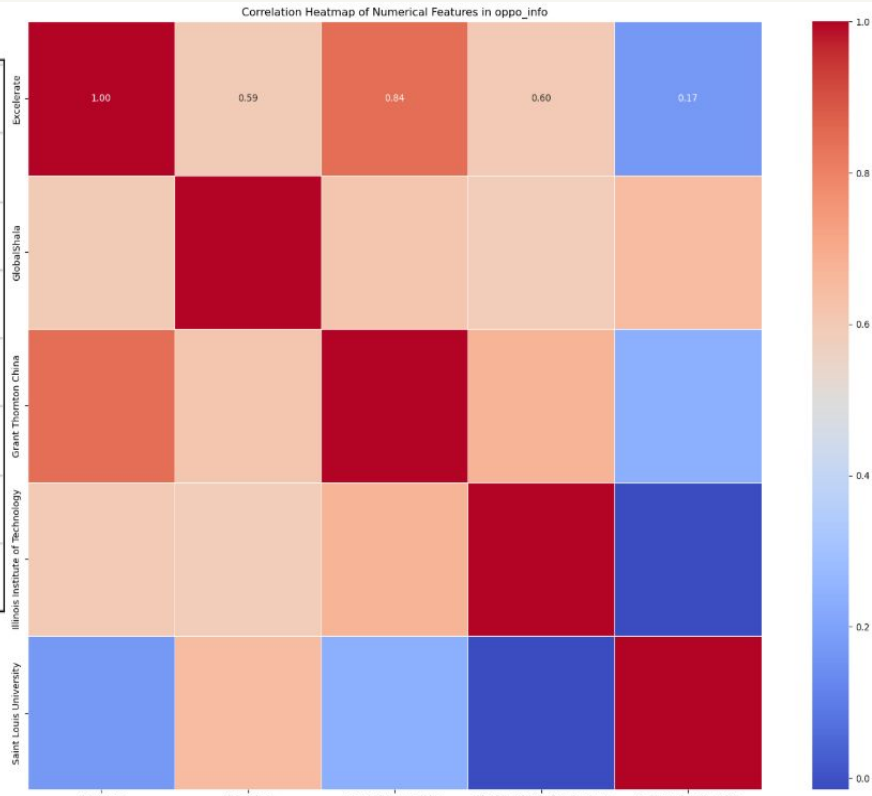
Validating the data types such as datetimens[64] for date based columns

2. Visuals

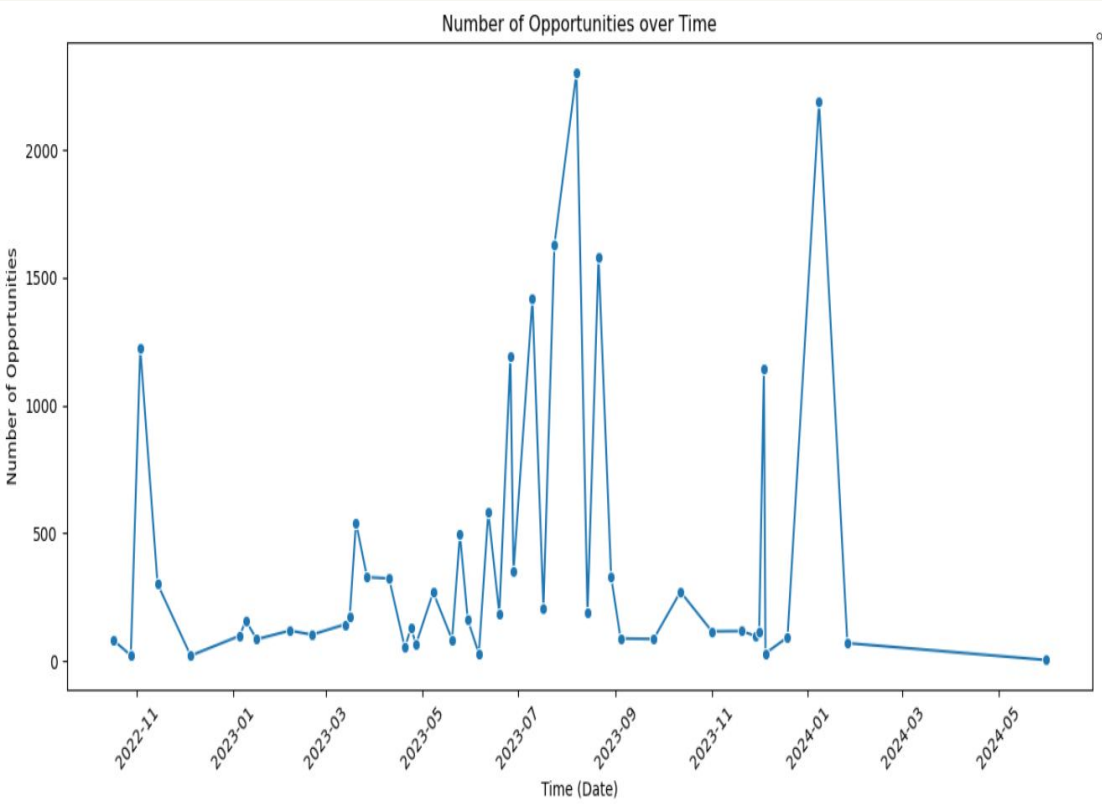
2.1: User Data Visualization
2.2: Opportunity Data Visualization

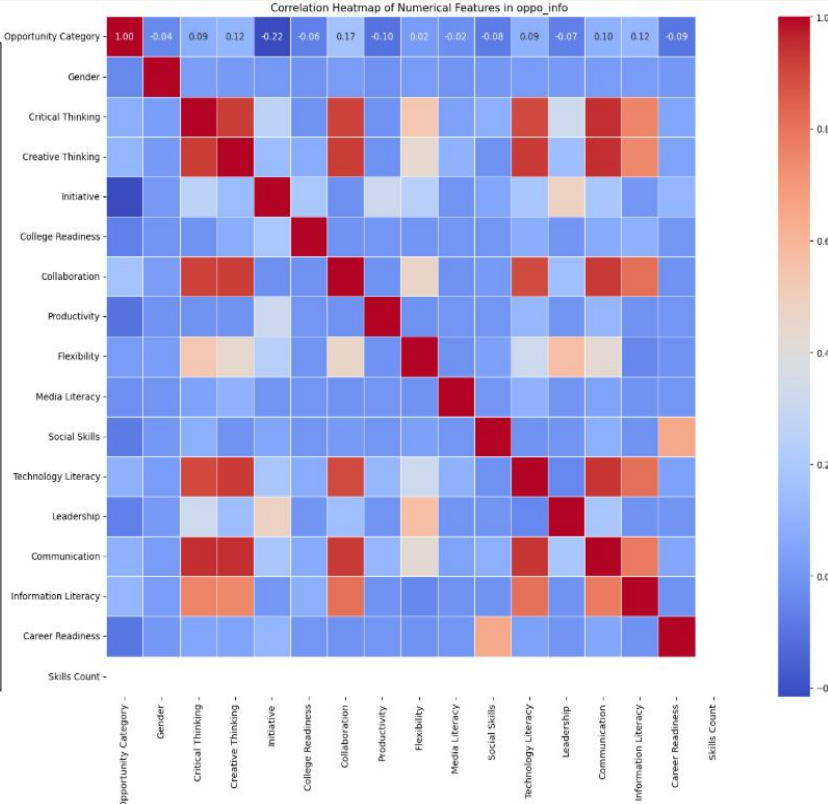
2.1 User Data Visualization





2.2 Opportunity Visualization

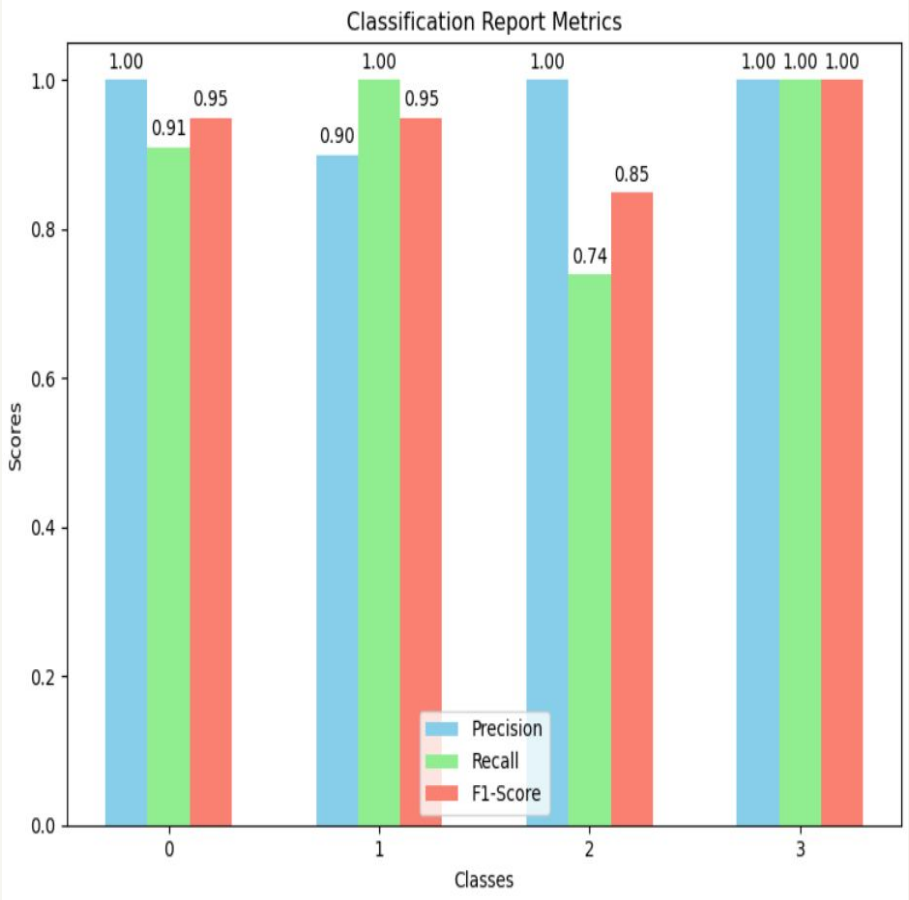




3. AI Model Training

3. AI Model Training

- Training Random Forest on Opportunity Wise Data for 7 features such as skill points, gender, location, etc. and predicting it against the target class of Opportunity Category
- Top feature/class after the model training was Critical Thinking
- Model predicted with an accuracy of 99%



4.Dashboards

Python was used for Data Cleaning and Validation purposes, while PowerBI was used for dashboard creation. Two Dashboards have been created for Opportunity Data for better visualization

User Data Dashboard

Opportunity Dashboard 1

Opportunity Dashboard 2

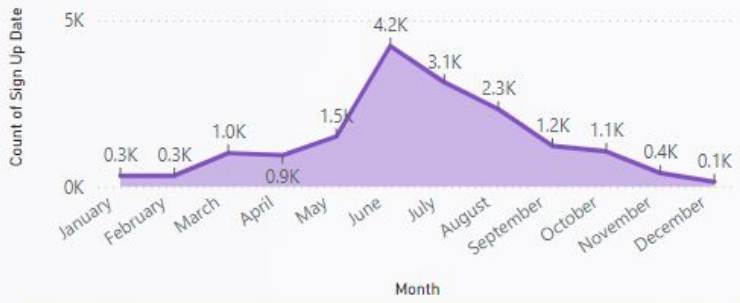
User Data

16.63K
Count of Sign Up Date

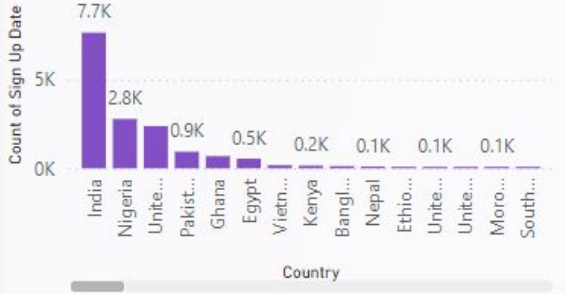
3688
Count of city

129
Count of Country

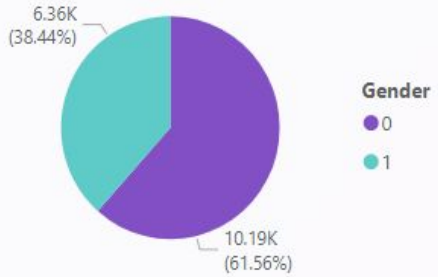
Count of Sign Up



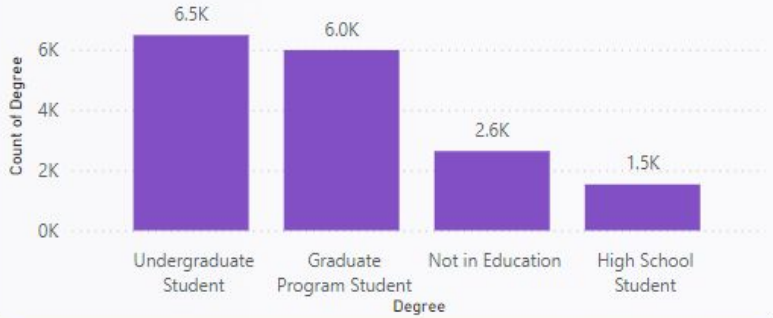
Count of Sign Up Date by Country



Count of Gender



Count of Degree by Degree



Count of isFromSocialMedia

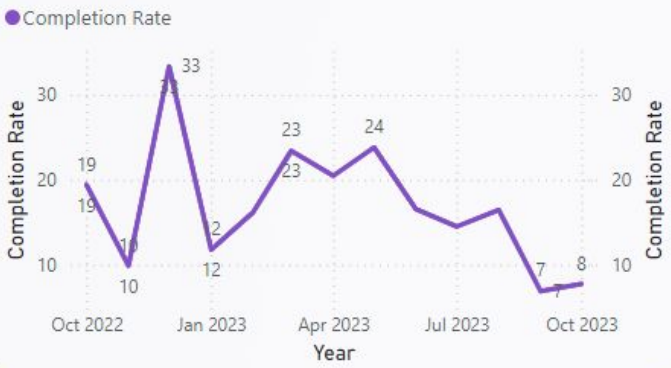
Opportunity Data

2.71M
Sum of Reward Amount

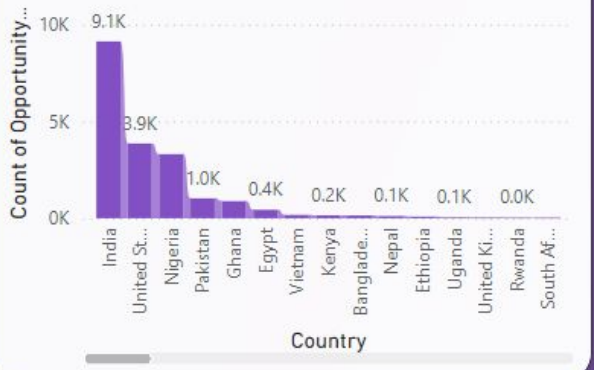
20.19K
Count of Profile Id

19.39K
Count of Opportunity Star...

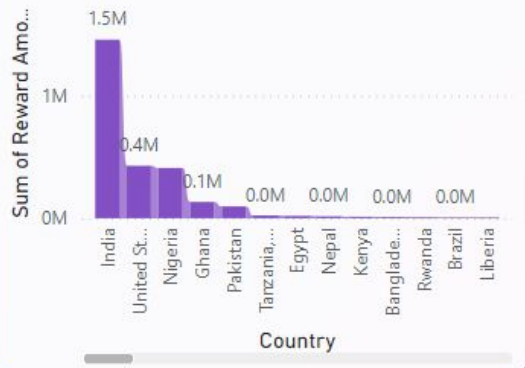
Completion Rate of Opportunity



Count of Opportunity Name by Country

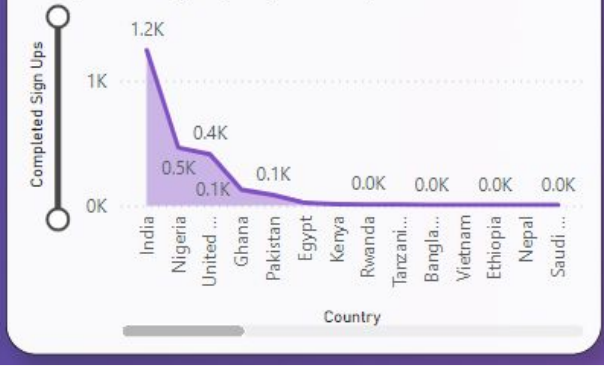


Reward Amount

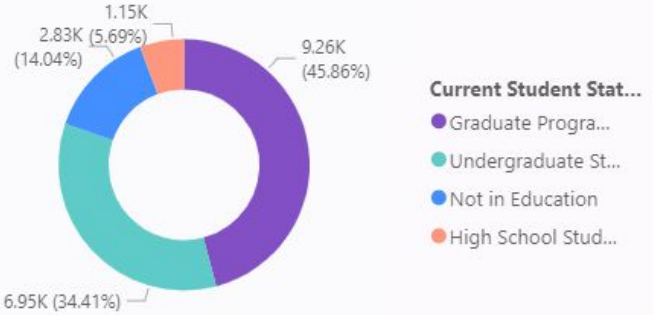


Top Opportunity

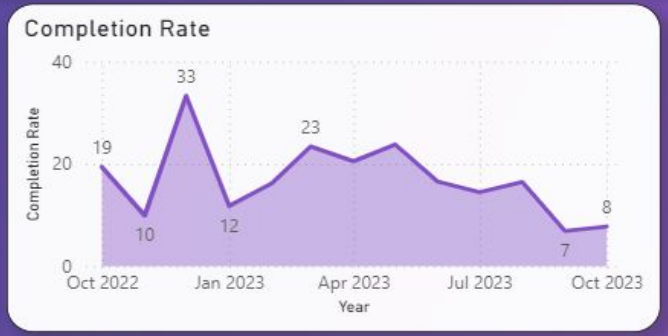
Completed Sign Ups by Country

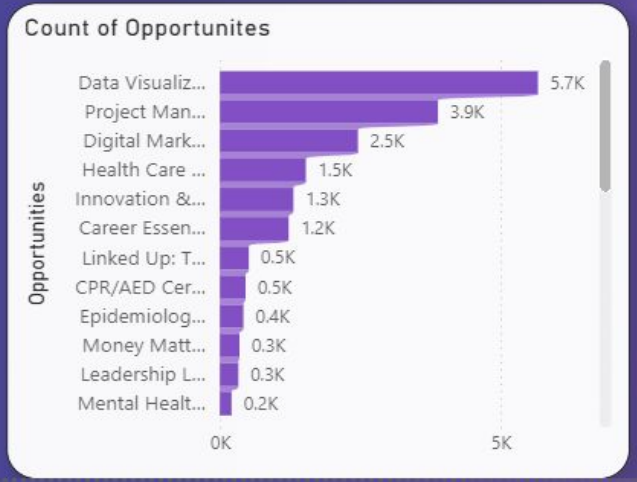


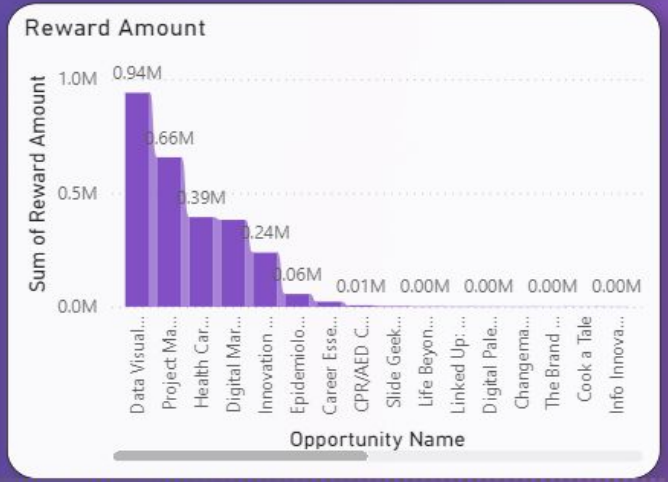
Current Student Status



Opportunity Data







5. Wrap up

5. Conclusion

- We thoroughly cleaned the data and validated it using Python
- We then trained AI model for Opportunity Data Prediction
- Then we created three dashboards that were refined for better visualization
- These dashboards are relevant to stakeholders such as senior management, project sponsors, partners, employees and other key stakeholders which can help them in making informed data driven decisions about the future and how company can be governed
- Excelerate can use these insights and demographics to train AI models and reduce the churn rate of its target audience through targeted campaigns and more.

Thank You