EXPLORATORY DATA ANALYSIS REPORT

Group 5B|0801-DV Associate internship.

INTRODUCTION

The main reason for this Exploratory Data Analysis Report is to find out why the majority of the users who applied for the opportunities are not always completing the opportunities and what makes few complete the opportunities.
The datasets that will be used for this find to draw insight are as follows; User Data and Opportunity Wise Data
User data: this dataset encompasses non-identifying information about every user who has ever created an account on Excelerate. The data is comprehensive, covering all users, regardless of their engagement with specific opportunities.

Opportunity signup and completion data: This dataset focuses on non-identifying user information about learners who have engaged with specific opportunities on the Excelerate platform.

DATA OVERVIEW

USER DATA

The User data which includes every user that has ever created an account on the Excelerate platform involves datasets with 8 column/column headings; Preferred sponsors of the user, gender of the user, country of residence during signup, degree or academic level of user, signup date and time of the user, city of residence, zip or postal code and the last column showing whether the user got the information through social media or not. The data has a total of 27,563 rows which is imperative to the same number of users with an account on Excelerate.

OPPORTUNITY WISE DATA

The Opportunity Wise Data comprises 17 different columns and 20323 rows. The rows were later reduced to 11482 after duplicates were removed. The profile ID column was used to remove the duplication because it is the most unique field among other fields.

COLUMN ANALYSIS

USER DATA

Preferred sponsors: a list of sponsors the user can choose from which are GlobalShala (GLS), Grant Thornton (GT), Illinois University(ILS), Saint Louis University (SLU) and Excelerate. A user can pick one or more sponsors from the list according to the dataset.

Gender: shows whether a user is a male or female and can be left blank (null) because the field is not mandatory when signing up. The frequency ratio of male to female to null is

Country: çountry of residence of user during signup which spans across the globe. The frequency of the countries using Excelerate.

Degree: The level of education of the user at the time of signup. This includes; High school students, Undergraduate, Graduate, and Not in education.

Signup date: The signup date and time when signing up

City: City of residence when signing up

Zip: Zip code or postal code of City of residence when signing up

Is from social media: This can be true or false depending on whether the user gets the information from a Google search or not.

OPPORTUNITY WISE DATA

The Opportunity Wise Data has a wide data structure with 17 Columns which are;
Profile ID: The profile ID is the unique Identity of the user on the Accelerate.
Opportunity ID: The opportunity ID is the unique ID that is particular to the opportunity a user applied for.
Opportunity Name: This is the name given to each opportunity on the Excelerate platform.
Opportunity Category; This is the category of all the available opportunities.
Opportunity End Date: This is the date on which an available opportunity will end.
Gender: This is the given gender for each user.

City: City of residence when signing up.

State: This is the state of residence of the user in his or her country and it is presented when signing up for the Excelerate platform.

Country: çountry of residence of user during signup which spans across the globe. The frequency of the countries using Excelerate.

Zip: Zip code or postal code of City of residence when signing up

Graduating Date (yyyy mm): This is the graduating date of the user for the opportunity applied for.

Current Student Status: This is the current level of education of each user. Whether the user is still in High School, an undergraduate, or a graduate student or not in education.

Current/Intended Major: This is the current academic status of the user before applying for any opportunity.

Status Description: This is the status of the user's opportunity applied for. Where Rejected =340, Applied = 26, Not Started = 732, Team Allocation= 8077, Drop Out = 17, Reward Award = 1285, Started = 693, Withdraw = 311.

Apply Date: This is the date the user applied for the opportunity.

Opportunity Start date: This is the date the users start the opportunity they applied for.

Reward Amount: This amount is given to the user who completed the opportunity.

Badge ID: This is the unique badge ID given to the user who earned the badge for the opportunity.

Badge name: This is the name of the badge given to the users.

Skill points earned: This is the point earned by the users during the opportunity.

Skills Earned: This is the number of skills gotten by the user after completing the opportunity.

PROFILE ID ANALYSIS

The profile ID is the unique ID for a particular user in the Excelerate platform. During the cleaning of the dataset, there were a lot of duplicates and the duplicates were removed to ensure the dataset's integrity. Profile ID is our unique value in the opportunity dataset. It was very fundamental in our data-cleaning process to use this column for the identification of duplicates. During the cleaning process, 8841 duplicates were found in this column. Remaining a total of 11481 unique value.

OPPORTUNITY STATUS DESCRIPTION

The status description column contains a lot of team allocations to users and few users started the opportunities. Among the few that started, it small number of users got the rewards. There is an insight we need to draw here to know why very few users got the reward at the end of the opportunity. Profile ID is our unique value in the opportunity dataset. It was very fundamental in our data-cleaning process to use this column for the identification of duplicates. During the cleaning process, 8841 duplicates were found in this column. Remaining a total of 11481 unique value.

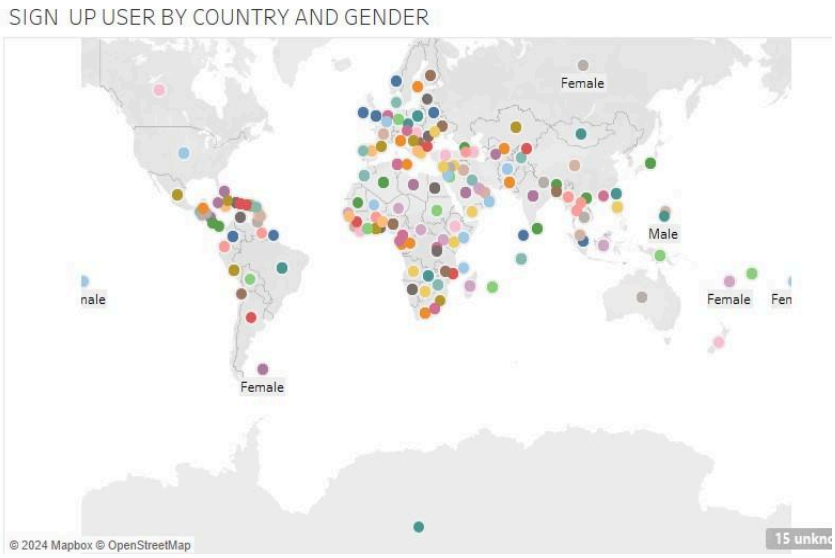
BASIC STATISTICS

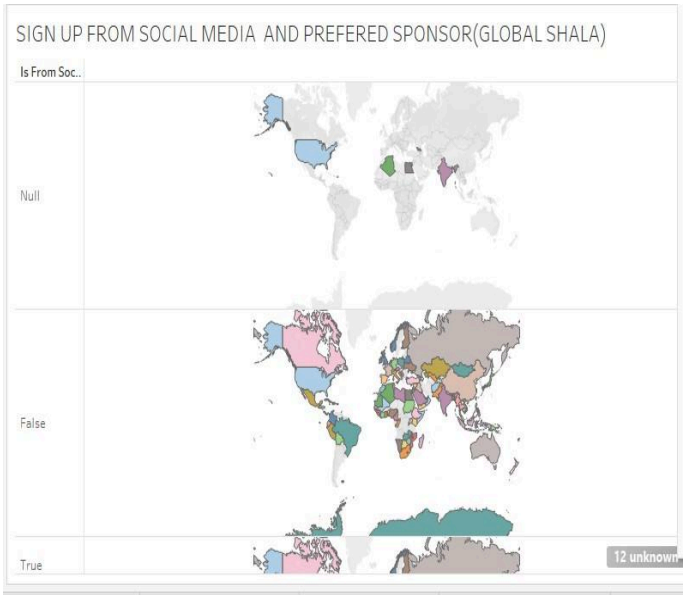
From the exploration so far, we see that there were a lot of users who applied for various opportunities. However, very few users were able to get the rewards at the end of the opportunity. The number of users that got

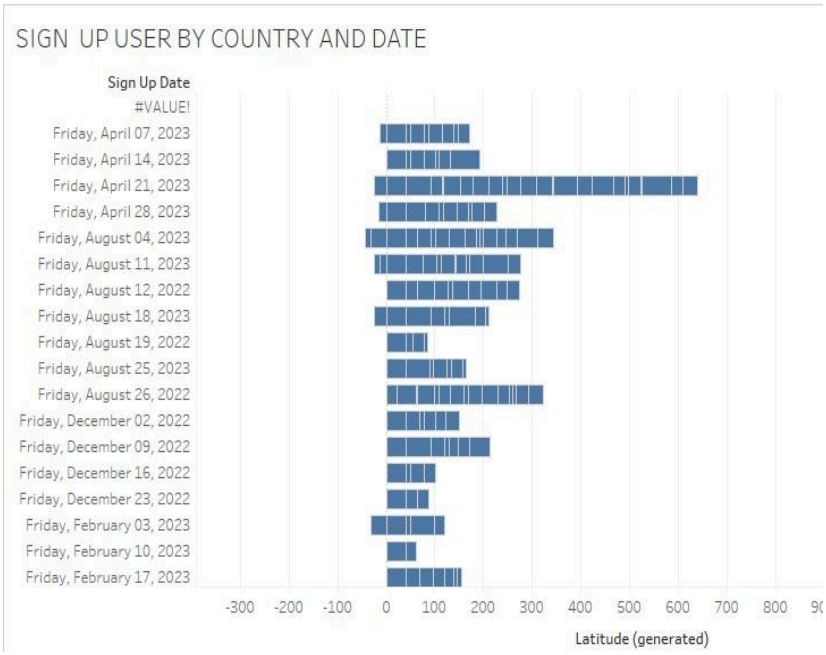
INITIAL OBSERVATIONS

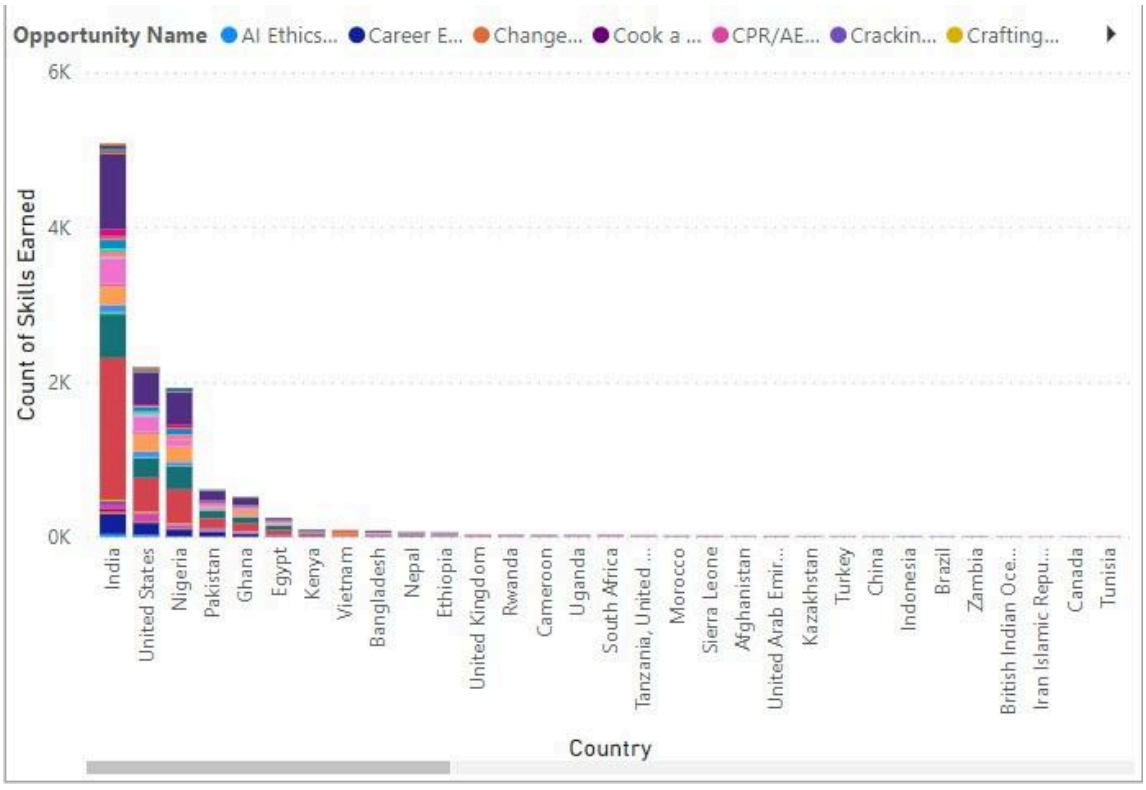
The initial observation we saw here is that the more users that applied to the opportunities the fewer users that got the rewards. Some users allocated to the team couldn’t start and some that started couldn't get the rewards for the opportunity.

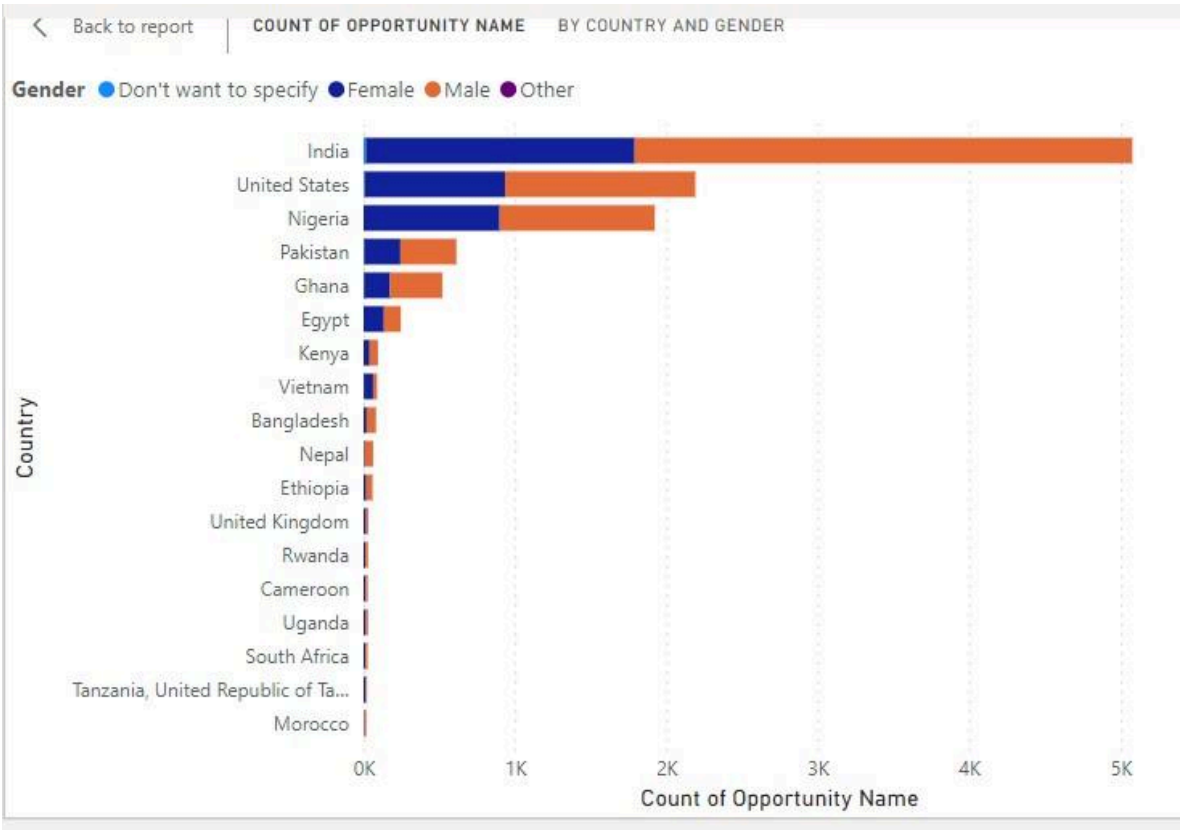
VISUALIZATION

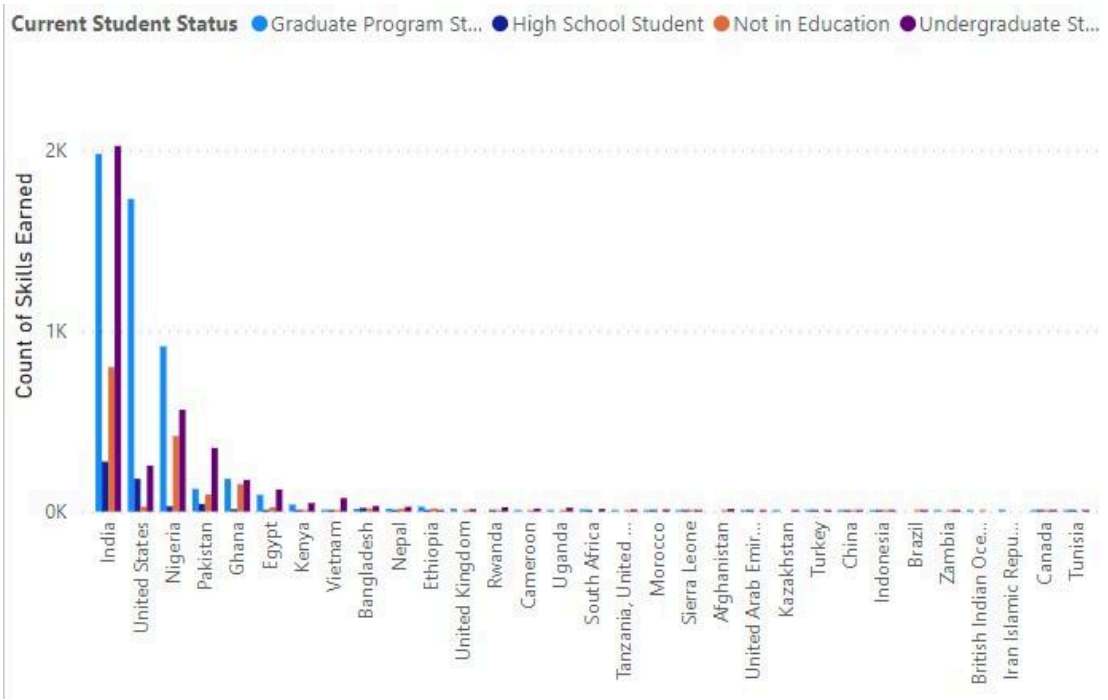





SIGN UP USER BY COUNTRY AND DATE
Sign Up Date
#VALUE!
Friday, April 07, 2023
Friday, April 14, 2023
Friday, April 21, 2023
Friday, April 28, 2023
Friday, August 04, 2023
Friday, August 11, 2023
Friday, August 12, 2022
Friday, August 18, 2023
Friday, August 19, 2022
Friday, August 25, 2023
Friday, August 26, 2022
Friday, December 02, 2022
Friday, December 09, 2022
Friday, December 16, 2022
Friday, December 23, 2022
Friday, February 03, 2023
Friday, February 10, 2023
Friday, February 17, 2023
-300 -200 -100 0 100 200 300 400 500 600 700 800
Latitude (generated)


Opportunity Name
AI Ethics...
Career E...
Change...
Cook a ...
CPR/AE...
Crackin...
Crafting...
6K
4K
2K
0K
Count of Skills Earned
India
United States
Nigeria
Pakistan
Ghana
Egypt
Kenya
Vietnam
Bangladesh
Nepal
Ethiopia
United Kingdom
Rwanda
Cameroon
Uganda
South Africa
Tanzania, United ...
Morocco
Sierra Leone
Afghanistan
United Arab Emir...
Kazakhstan
Turkey
China
Indonesia
Brazil
Zambia
British Indian Oce...
Iran Islamic Repu...
Canada
Tunisia
Country


Back to report
COUNT OF OPPORTUNITY NAME BY COUNTRY AND GENDER
Gender Don't want to specify Female Male Other
India
United States
Nigeria
Pakistan
Ghana
Egypt
Kenya
Vietnam
Bangladesh
Nepal
Ethiopia
United Kingdom
Rwanda
Cameroon
Uganda
South Africa
Tanzania, United Republic of Ta...
Morocco
Country
0K
1K
2K
3K
4K
5K
Count of Opportunity Name


Current Student Status Graduate Program St... High School Student Not in Education Undergraduate St...
Count of Skills Earned
2K
1K
0K
India
United States
Nigeria
Pakistan
Ghana
Egypt
Kenya
Vietnam
Bangladesh
Nepal
Ethiopia
United Kingdom
Rwanda
Cameroon
Uganda
South Africa
Tanzania, United ...
Morocco
Sierra Leone
Afghanistan
United Arab Emir...
Kazakhstan
Turkey
China
Indonesia
Brazil
Zambia
British Indian Oce...
Iran Islamic Repu...
Canada
Tunisia

CHALLENGES FACED

There were a lot of challenges cleaning the dataset because of the irregularity of the Zip code column and also the visualizations were not easy to generate or design.

NEXT STEPS

Our next now is to focus on Visualization so that we can have a designed dashboard.