

# **Exploratory Data Analysis Report (EDA) for User Data and Opportunity and Completion Data for Team 2H.**

## **PART 1: - USER DATA**

### **1. Introduction**

#### **1.1 Brief Background**

Excelerate is the world's first platform for gaining 21st-century skills through personalized experiences and learnership opportunities from anywhere in the world. Uniquely positioned, it serves as the sole platform uniting employers, universities, and high schools, curating a diverse array of experiences, skills, and support crucial in preparing learners for future career triumphs and success within the global economy.

For that purpose, the learners had to do a comprehensive Exploratory Data Analysis (EDA), Preprocessing, and Wireframe on two key datasets: User Data and Opportunity Sign-Up and Completion Data. These datasets hold non-identifying user information, offering distinct narratives waiting to be unveiled. The primary focus of this analysis was the Opportunity Data, aiming to uncover patterns, perform feature engineering, and provide decision-making support.

The challenge revolved around interpreting non-identifiable user data related to participants engaging with targeted opportunities within the Excelerate platform. This encompassed demographics, skill acquisition, and geographical distribution, painting a diverse engagement picture. The objective was to dissect and analyze data across demographics, earned skills and achieved reward amounts do a comparative analysis, and much more.

The specific objectives entailed gathering and analyzing non-identifiable Opportunity data, exploring diverse facets such as demographics, skill acquisition, and geographical distribution, and conducting a thorough examination across demographics, skills, and reward amounts to extract actionable insights. Additionally, identifying trends and correlations within the data aimed to inform strategic decisions and optimize user engagement on the Excelerate platform and also to do a comparative analysis of the two datasets.

#### **1.2 Problem Statement**

The challenge revolves around leveraging a dataset containing multiple column variables such as 'PreferredSponsors,' 'Gender' 'Country' 'Degree' 'Sign Up Date' 'City' 'ZIP', and 'IsFromSocialMedia'. The task involves analyzing this dataset to extract meaningful insights and correlations among these variables. The aim is to understand the relationships between preferred sponsors, demographic information (gender, country), educational background (degree), temporal data (sign-up date), geographical factors, and the influence of social media on user registration. The objective is to uncover patterns and trends within this dataset that can

aid in optimizing sponsorship strategies, understanding user demographics, and refining marketing approaches for increased user engagement and platform growth.

### **1.3 Aim and Objectives**

The analysis aimed to determine the number of users registered on the Excelerate platform and discern trends, particularly focusing on social media applicants and various market strategies. This holistic approach sought to provide a comprehensive understanding of user behavior and engagement dynamics within Excelerate (Sponsor).

### **Agenda**

1. Business Understanding
  - Define the problem and objectives of the project
  - Determine the sponsors and their requirements
  - Identify the available data sources and their limitations
2. Data Understanding
  - Collect the relevant data from various sources
  - Perform initial data exploration and visualization to gain insights
  - Clean and preprocess the data, handle missing values and outliers

### **Objectives:**

1. Develop a visualization dashboard that accurately convey the sponsor on SWOT
2. Clean, preprocess, and transform the data to make it suitable for the visualization
3. Evaluate various metrics in the preprocessing process to gain insights on rate of conversion of the applicants, and will focus on the marketing gaps, skills and other gaps.
4. Develop a wireframe as a blueprint for dashboard creation
5. Develop documentation on the project evolution and process for the sponsors to explain the outputs and limitations.

## **2. Data Overview**

The User Data encompasses non-identifying information about every user who has ever created an account on Excelerate such as variables such as 'PreferredSponsors,' 'Gender' 'Country' 'Degree' 'Sign Up Date' 'City' 'ZIP', and 'IsFromSocialMedia'. The dataset is comprehensive, covering all users, regardless of their engagement with specific opportunities. Each row represents a unique user, and the dataset provides a holistic view of the user base.

The following are the features of the dataset:

- The dataset consists of 27562 rows and 8 columns.
- The dataset includes mainly categorical data types.
- Categorical features include Gender, Country, City etc.
- The dataset has 7763 missing values across 5 columns.

- Columns with missing values: Gender, City, Country, Degree, and IsFromSocialMedia.

## 2.1 Data Cleaning Procedure

### Reading And Understanding the User Data:

As stated earlier, the dataset comprises **27562 rows** and **8 columns** containing information about people who had created accounts on Excelerate.

#### Sample Data:

Preferr	Gender	Country	Degree	Sign Up	city	zip	isFrom	cialMedia
1	["GlobalSI Male	Nigeria	Undergrad	2023-07-21:09:58.602Z	Owerri	460103	FALSE	
2	["GlobalSI Male	India	Undergrad	2023-04-24:09:57:07.405Z	kottayam	686501	FALSE	
3	["GlobalSI null	India	null	2022-10-14T17:13:36.303Z			FALSE	
5	["GlobalShala","Grant Thornton China"	Albania		2023-06-06T12:29:01.772Z			TRUE	
6	["GlobalSI Female	Ghana	Not in Edu	2023-06-11:09:58.602Z	Kumasi	AT-1214-9	FALSE	
7	["GlobalSI Female	India		2023-07-01:09:58.602Z	Chennai	600033	FALSE	
8	["GlobalShala","Grant Thornton China"	Nigeria		2023-05-15T21:30:04.370Z			TRUE	
9	["GlobalShala","Grant Thornton China"	United States		2023-07-26T17:01:59.361Z			TRUE	
10	["GlobalSI Male	Nigeria	Undergrad	2023-07-21:09:58.602Z	Lagos	100278	TRUE	
11	["GlobalSI Male	India	High Schc	2023-05-01:09:58.602Z	RAS	388570	TRUE	
12	["GlobalShala","Grant Thornton China"	India		2023-05-26T17:05:18.742Z			TRUE	
13	["GlobalShala","Grant Thornton China"	India		2023-05-07T07:06:50.500Z			TRUE	
14	["GlobalSI Female	India	Undergrad	2023-06-11:09:58.602Z	Patiala	147001	TRUE	
15	["GlobalSI Female	United Sta	Graduate	2023-05-01:09:58.602Z	SPARROW	21219	TRUE	
16	["GlobalSI Female	United Sta	Graduate	2023-10-01:09:58.602Z	Saint Loui	63043	TRUE	
17	["Saint Loi Male	Ghana	Graduate	2023-06-11:09:58.602Z	Accra	233	FALSE	
18	["GlobalShala","Grant Thornton China"	India		2023-04-05T12:37:13.652Z			TRUE	
19	["GlobalShala","Grant Thornton China"	Pakistan		2023-06-15T18:53:01.568Z			TRUE	
20	["GlobalSI Female	Nigeria	Not in Edu	2023-03-01:09:58.602Z	Abuja	234	FALSE	
21	["GlobalShala","Grant Thornton China"	India		2023-07-19T18:44:25.551Z			TRUE	
22	["Saint Loi Male	India	Graduate	2023-08-21:09:58.602Z	Saint Loui	63108	FALSE	
23	["GlobalSI Female	Pakistan	Undergrad	2023-06-11:09:58.602Z	Peshawar	923	TRUE	
24	["GlobalShala","Grant Thornton China"	India		2023-06-19T10:33:19.751Z			FALSE	

#### Dropping Unwanted Variable:

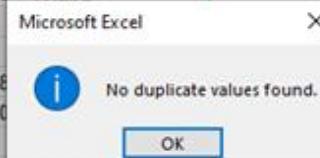
By selecting all the variables, activate the filter option.

The “Zip” column was deleted due to non-uniformity and non-accuracy. Since State and Country variables were present, the zip variable was deleted.

B	C	D	E	F	G	H
1	Gender	Country	Degree	Sign Up Date	city	isFromSocialMedia
2	Male	Nigeria	Undergraduate Student	2023-07-23T08:05:58.602Z	Owerri	FALSE
3	Male	India	Undergraduate Student	2023-04-24T09:57:07.405Z	kottayam	FALSE
4	null	India	null	2022-10-14T17:13:36.303Z		FALSE
5	Albania			2023-06-06T12:29:01.772Z		TRUE
6	Female	Ghana	Not in Education	2023-06-15T16:31:42.719Z	Kumasi	AT-1214-9090
7	Female	India		2023-07-06T18:49:16.691Z	Chennai	600033
8		Nigeria		2023-05-15T21:30:04.370Z		TRUE
9		United States		2023-07-26T17:01:59.361Z		TRUE
10	Male	Nigeria	Undergraduate Student	2023-07-27T18:02:17.535Z	Lagos	100278
11	Male	India	High School Student	2023-05-05T04:47:25.446Z	RAS	388570
12		India		2023-05-26T17:05:18.742Z		TRUE
13		India		2023-05-07T07:06:50.500Z		TRUE

#### Finding Duplicates:

Preferr	Gender	Country	Degree	Sign Up	city	zip	isFrom	cialMedia
["GlobalShala","Grant Thornton China"]	Male	Nigeria	Undergrad	2023-07-21T10:00:00Z	Owerri	460103	FALSE	
["GlobalShala","Grant Thornton China"]	Male	India	Undergrad	2023-04-21T10:00:00Z	kottayam	686501	FALSE	
["GlobalShala","Grant Thornton China"]	null	India	null	2022-10-14T17:13:36.303Z			FALSE	
["GlobalShala","Grant Thornton China"]	Albania			2023-06-06T12:29:01.772Z			TRUE	
["GlobalShala","Grant Thornton China"]	Female	Ghana	Not in Edu	2023-06-11T10:00:00Z	Kumasi	AT-1214-9	FALSE	
["GlobalShala","Grant Thornton China"]	Female	India		2023-07-01T10:00:00Z	Chennai	600033	FALSE	
["GlobalShala","Grant Thornton China"]	Nigeria			2023-05-15T21:30:04.370Z				
["GlobalShala","Grant Thornton China"]	United States			2023-07-26T17:01:59.361Z				
["GlobalShala","Grant Thornton China"]	Male	Nigeria	Undergrad	2023-07-21T10:00:00Z	Lagos	100278		
["GlobalShala","Grant Thornton China"]	Male	India	High Schc	2023-05-01T10:00:00Z	RAS	388570		
["GlobalShala","Grant Thornton China"]	India			2023-05-26T17:05:18.742Z				
["GlobalShala","Grant Thornton China"]	India			2023-05-07T07:06:50.500Z			TRUE	
["GlobalShala","Grant Thornton China"]	Female	India	Undergrad	2023-06-11T10:00:00Z	Patiala	147001	TRUE	
["GlobalShala","Grant Thornton China"]	Female	United Sta	Graduate	2023-05-01T10:00:00Z	SPARROW	21219	TRUE	
["GlobalShala","Grant Thornton China"]	Female	United Sta	Graduate	2023-10-11T10:00:00Z	Saint loui	63043	TRUE	
["Saint Lohit Male"]	Male	Ghana	Graduate	2023-06-11T10:00:00Z	Accra	233	FALSE	
["GlobalShala","Grant Thornton China"]	India			2023-04-05T12:37:13.652Z			TRUE	
["GlobalShala","Grant Thornton China"]	Pakistan			2023-06-15T18:53:01.568Z			TRUE	
["GlobalShala","Grant Thornton China"]	Female	Nigeria	Not in Edu	2023-03-01T10:00:00Z	Abuja	234	FALSE	
["GlobalShala","Grant Thornton China"]	India			2023-07-19T18:44:25.551Z			TRUE	
["Saint Lohit Male"]	Male	India	Graduate	2023-08-21T10:00:00Z	Saint Loui	63108	FALSE	
["GlobalShala","Grant Thornton China"]	Female	Pakistan	Undergrad	2023-06-01T10:00:00Z	Peshawar	923	TRUE	
["GlobalShala","Grant Thornton China"]	India			2023-06-10T10:22:10.751Z			FALSE	



No duplicate was found as seen hence; no action required.

### Handling Missing Values:

I. **Preferred Sponsor column** - No action required as there were no missing values.  
However, the sponsors were split to identify the five categories of the preferred sponsors.

### II. **Gender Column:**

Male	11027
Female	6910
Other	15
Don't want to specify	75
Null	1282
Blank	8253
<b>Action:</b>	
i.	Other attached to "Don't want to specify".
ii.	Blank and Null assigned "Not specified".

### III. Country Column:

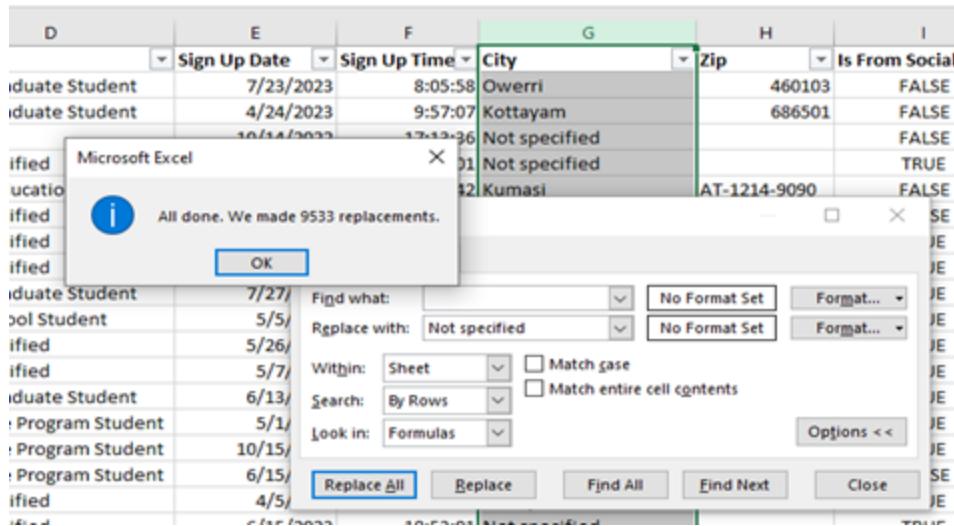
Null = 26; Blank = 36; replaced with “Not specified”.

### IV. Degree Column:

Null = 1442; Blank = 9370; replaced with “Not specified”.

### V. City Column:

Blank = 9533; replaced with “Not specified”. Also, 5 cities were represented with “A” which cuts across all countries and does not represent any city hence, they were deleted. Also, the city names were standardized.



## VI. Is From Social Media:

Blank = 9; replaced with “Not specified”.

Is From Social-Media	Action
TRUE	No changes
FALSE	No changes
Blanks	9 blanks replaced with “Not Specified” so as to enhance other areas “Not Specified” has been included.

VII. Finally, the “Sign-up date” column was separated into **Sign-up date** (with format changed to “MM-DD-YYYY”) and **Sign-up time** columns respectively

E	F	G	H	I	J	K	L
Preferred Sponsor	Gender	Country	Degree	Sign Up Date	Sign Up Time	City	IsFromSocialMedia
Excelerate	Male	Nigeria	Undergraduate Student	7/23/2023	08:05:58.4622	Owerri	FALSE
Excelerate	Male	India	Undergraduate Student	4/24/2023	09:57:07.4652	Kottayam	FALSE
Excelerate	Not specified	India	Not specified	10/14/2023	17:13:26.4726	Not specified	FALSE
Excelerate	Not specified	Albania	Not specified	8/6/2023	12:29:05.7722	Not specified	TRUE
Excelerate	Female	Ghana	Not in Education	8/15/2023	16:31:42.7197	Kumasi	FALSE
Excelerate	Female	India	Not specified	7/6/2023	18:49:18.6912	Chennai	FALSE
Excelerate	Not specified	Nigeria	Not specified	5/15/2023	21:30:04.8702	Not specified	TRUE
Excelerate	Not specified	United States	Not specified	7/26/2023	17:05:59.3612	Not specified	TRUE
Excelerate	Male	Nigeria	Undergraduate Student	7/27/2023	18:02:17.5352	Lagos	TRUE
Excelerate	Male	India	High School Student	5/5/2023	04:47:25.4462	Ras	TRUE
Excelerate	Not specified	India	Not specified	5/26/2023	17:05:18.7422	Not specified	TRUE
Excelerate	Not specified	India	Not specified	5/7/2023	07:06:50.5002	Not specified	TRUE
Excelerate	Female	India	Undergraduate Student	6/13/2023	09:16:48.0292	Pattala	TRUE

Therefore, there are a total of **27557 rows** and **12 columns** after data cleaning.

### Validating the categorical data.

The screenshot shows a Microsoft Excel spreadsheet with data across multiple columns including PreferredSponsor, Gender, Country, Degree, Sign Up Date, city, isFromSo, and lMedia. A 'Find and Replace' dialog box is open over the spreadsheet. The 'Find what' field is empty ('null'). The 'Replace with' field contains 'Others'. The 'Look in' dropdown is set to 'Formulas'. The 'Replace All' button is highlighted in blue.

PreferredSponsor	Gender	Country	Degree	Sign Up Date	city	isFromSo	lMedia
["GlobalShala"]	"Male	Nigeria	Undergrad	2023-07-23T08:05:58.602Z	Owerri	FALSE	
["GlobalShala"]	"Male	India	Undergrad	2023-04-24T09:57:07.405Z	kottayam	FALSE	
["GlobalShala"]	"Others	India	Others	2022-10-14T17:13:36.303Z	Others	FALSE	
["GlobalShala"]	"Others	Albania	Others	2023-06-06T12:29:01.772Z	Others	TRUE	
["GlobalShala"]	"Female	Ghana	Not in Edu	2023-06-15T16:31:42.719Z	Kumasi	FALSE	
["GlobalShala"]	"Female	India	Others	2023-07-06T18:49:16.691Z	Chennai	FALSE	
["GlobalShala"]	"Others	Nigeria	Others	2023-05-15T21:30:04.370Z	Others	TRUE	
["GlobalShala"]	"Others	United States	Others	2023-07-26T17:01:59.361Z	Others	TRUE	
["GlobalShala"]	"Male	Nigeria	Undergrad	2023-07-27T18:02:17.535Z	Lagos	TRUE	
["GlobalShala"]	"Male	India	High Sch	2023-05-05T04:47:25.446Z	RAS	TRUE	
["GlobalShala"]	"Others	India	Others	2023-05-26T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Others	India	Others	2023-05-07T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Male	India	Undergrad	2023-06-13T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Female	United States	Graduate	2023-05-01T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Female	United States	Graduate	2023-10-15T17:05:18.742Z	Others	TRUE	
["Saint Louis University"]	"Male	Ghana	Graduate	2023-06-15T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Others	India	Others	2023-04-05T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Others	Pakistan	Others	2023-06-15T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Female	Nigeria	Not in Edu	2023-03-03T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Others	India	Others	2023-07-19T17:05:18.742Z	Others	TRUE	
["Saint Louis University"]	"Male	India	Graduate	2023-08-26T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Female	Pakistan	Undergrad	2023-06-14T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Others	India	Others	2023-06-19T17:05:18.742Z	Others	TRUE	

### Replaced Others with “Not specified”:

The screenshot shows a Microsoft Excel spreadsheet with data across multiple columns including PreferredSponsor, Gender, Country, Degree, Sign Up Date, city, isFromSo, and lMedia. A 'Find and Replace' dialog box is open over the spreadsheet. The 'Find what' field contains 'Others'. The 'Replace with' field contains 'Not specified'. The 'Look in' dropdown is set to 'Formulas'. The 'Replace All' button is highlighted in blue.

PreferredSponsors	Gender	Country	Degree	Sign Up Date	city	isFromSo	lMedia
["GlobalShala"]	"Grant Thc Male	Nigeria	Undergrad	2023-07-23T08:05:58.602Z	Owerri	FALSE	
["GlobalShala"]	"Grant Thc Male	India	Undergrad	2023-04-24T09:57:07.405Z	kottayam	FALSE	
["GlobalShala"]	"Illinois In Not specified	India	Not specified	2022-10-14T17:13:36.303Z	Not specified	FALSE	
["GlobalShala"]	"Grant Thc Not specified	Albania	Not specified	2023-06-06T12:29:01.772Z	Not specified	TRUE	
["GlobalShala"]	"Grant Thc Female	Ghana	Not in Edu	2023-06-15T16:31:42.719Z	Kumasi	FALSE	
["GlobalShala"]	"Grant Thc Female	India	Not specified	2023-07-06T18:49:16.691Z	Chennai	FALSE	
["GlobalShala"]	"Grant Thc Not specified	Nigeria	Not specified	2023-05-15T21:30:04.370Z	Not specified	TRUE	
["GlobalShala"]	"Grant Thc Not specified	United States	Not specified	2023-07-26T17:01:59.361Z	Not specified	TRUE	
["GlobalShala"]	"Grant Thc Male	Nigeria	Undergrad	2023-07-27T18:02:17.535Z	Lagos	TRUE	
["GlobalShala"]	"Grant Thc Male	India	High School	2023-05-05T04:47:25.446Z	RAS	TRUE	
["GlobalShala"]	"Grant Thc Not specified	India	Not specified	2023-05-26T17:05:18.742Z	Not specified	TRUE	
["GlobalShala"]	"Grant Thc Not specified	India	Not specified	2023-05-07T07:06:30.500Z	Not specified	TRUE	
["GlobalShala"]	"Grant Thc Female	India	Undergrad	2023-06-13T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Female	United States	Graduate	2023-06-15T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Gram Thc Female	United States	Graduate	2023-06-15T17:05:18.742Z	Others	TRUE	
["Saint Louis University"]	"Male	Ghana	Graduate	2023-06-15T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Not specified	India	Not specified	2023-07-19T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Not specified	Pakistan	Not specified	2023-05-26T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Female	Nigeria	Not in Edu	2023-06-15T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Not specified	India	Not specified	2023-07-06T18:49:16.691Z	Others	TRUE	
["Saint Louis University"]	"Male	India	Graduate	2023-08-26T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Female	Pakistan	Undergrad	2023-06-14T17:05:18.742Z	Others	TRUE	
["GlobalShala"]	"Grant Thc Not specified	India	Not specified	2023-06-19T17:05:18.742Z	Others	TRUE	

### 3. Column Analysis

- The columns ‘Sign Up Date’ and ‘Sign Up Time’ are of date data type having no missing values.

- The columns ‘Preferred Sponsors’, ‘Preferred Sponsors\_1’, ‘Preferred Sponsors\_2’, ‘Preferred Sponsors\_3’, ‘Preferred Sponsors\_4’, ‘Preferred Sponsors\_5’, ‘Gender’, ‘Country’, ‘City’, ‘Degree’ and ‘isFromSocialMedia’ are of categorical data type with missing values mentioned below:
  - i.Preferred Sponsors\_1 = 3
  - ii.Preferred Sponsors\_2 = 934
  - iii.Preferred Sponsors\_3 = 1057
  - iv.Preferred Sponsors\_4 = 2735
  - v.Preferred Sponsors\_5 = 3033
  - vi.Gender = 0
  - vii.Country = 0
  - viii.City = 0
  - ix.IsFromSocialMedia = 0.

The next section of this EDA report is profile id analysis, followed by opportunity status distribution and basic statistics. These sections are only characterized by the Opportunity Sign Up and Completion Data which will be discussed in the second part of this report.

#### **4. Initial Observations**

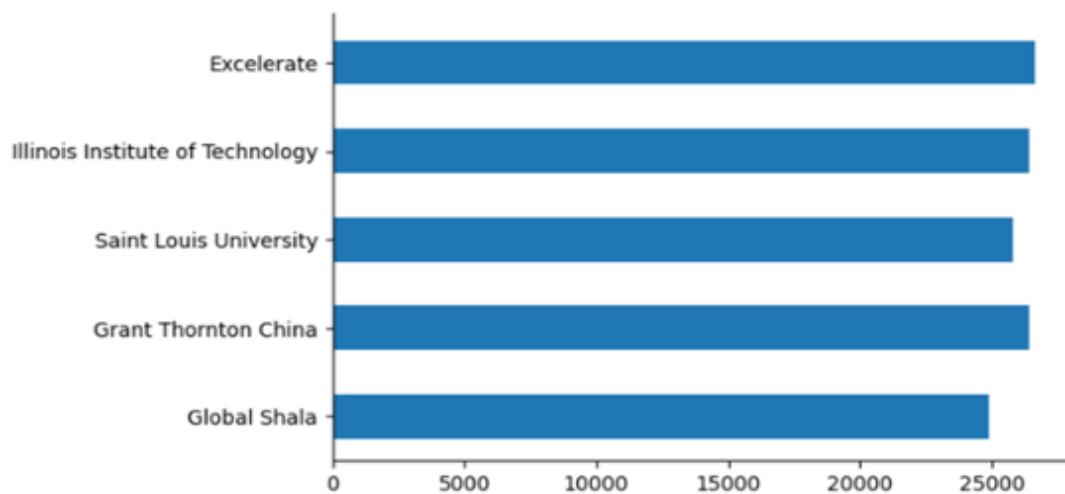
During the EDA process, the following observations were identified.

- Applicants were observed to have chosen Global Shala as the most preferred sponsor. Also, a large percentage of applicants chose up to five sponsors. As can be seen, Excelerate has a greater number of applicants of 26628, however, Excelerate was chosen as the least preferred sponsor. This is an area of interest to be explored in the upcoming weeks. Figure5.1 shows the distribution of sponsors by applicants.
- Three different categories of gender were grouped with 6909 identified as female, 11023 as male while 9535 were not specified. This is shown in Figure5.2 and to be explored in the upcoming weeks. Despite the lower inquiry count for females compared to males, females exhibit a higher percentage of individuals with disclosed degrees (see Figure5.9). This trend might be attributed to the possibility that a smaller proportion of males chose to disclose their degrees in comparison to females.
- Most of the applications/inquiries were made in 2023 with a total of 24241 and were mostly applied in June. 3316 applications were made in 2022. In contrast, most months in 2022 had significantly lower inquiry counts compared to the heightened interest observed in 2023. Figures 5.7 and 5.8 show the visualization of applicants by year and month respectively.
- Figure5.3 shows that India has the highest number of applicants as shown below with a total of 11892. There was a substantial difference in the inquiry counts between the countries in the first and second positions. This is also an area to explore in the upcoming weeks following the top ten countries of application.
- Figure5.4 shows that the city with the highest number of applicants – 9533 were not specified. This implies that the largest number of inquiries came from individuals who chose not to disclose their city - a potential area to be investigated. Among the disclosed cities, Hyderabad, followed by Saint Louis, had the highest number of inquiries.

6. Figure5.5 shows that the highest number of applicants, 10810 did not specify their degree level. Figure5.9 also confirms the analysis with few female applicants who failed to specify their degree. A potential area to also explore.
7. There is a very close relationship between those who applied via social media and those who did not, 13807 and 13741 respectively. This is shown in Figure5.6 and to be explored in the subsequent weeks.

## 5. Visualizations

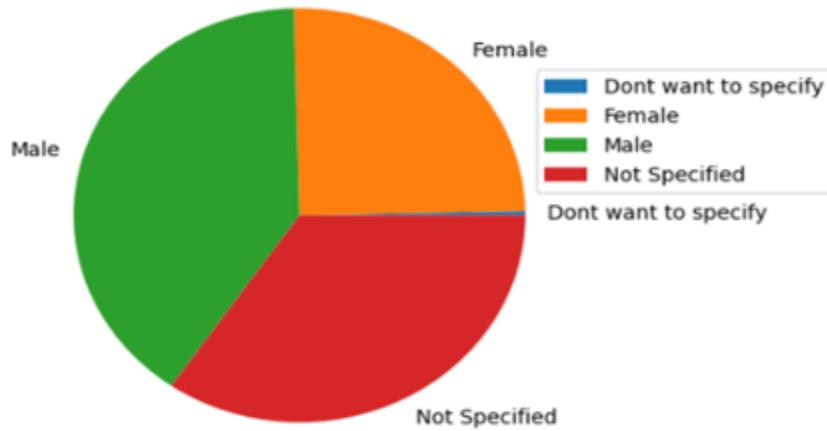
### 5.1 Analyzing Distribution of Preferred Sponsor



**Figure. 5.1: Distribution of Preferred Sponsors**

It has been observed that Excelerate is the most favored sponsor, with the Illinois Institute of Technology and Grant Thornton China following closely in preference.

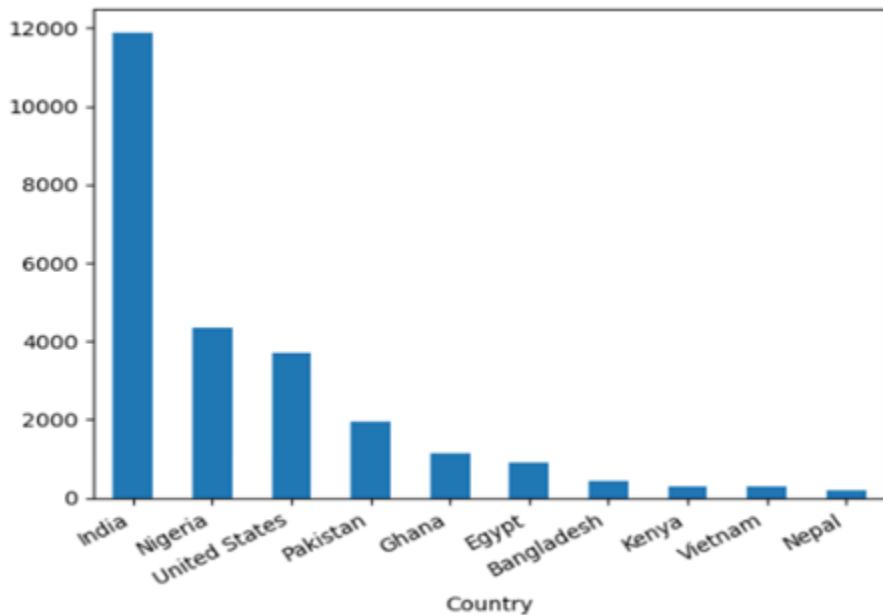
### 5.2 Analyzing Distribution of Applicants by Gender



**Figure. 5.2: Applicants by Gender**

We have found that the highest number of inquiries came from males, with females ranking third. The count of individuals choosing not to disclose their gender fell between the counts of males and females.

### 5.3 Analyzing Distribution of Top 10 Applicants by Country

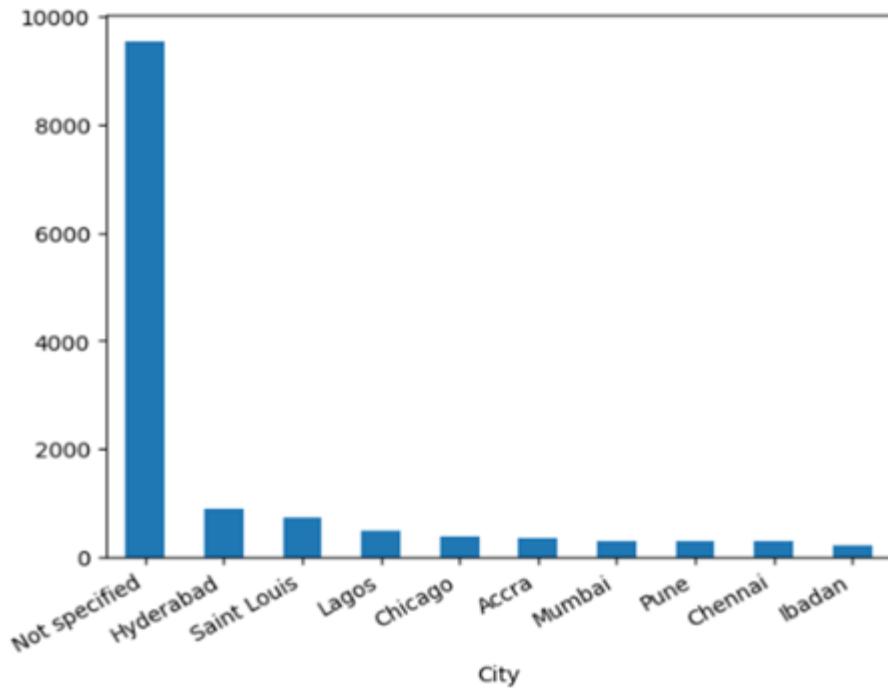


**Figure. 5.3: Top Ten Applicants**

It was noted that the highest number of inquiries originated from India, followed by Nigeria. There was a substantial difference in the inquiry counts between the countries in the first and second positions.

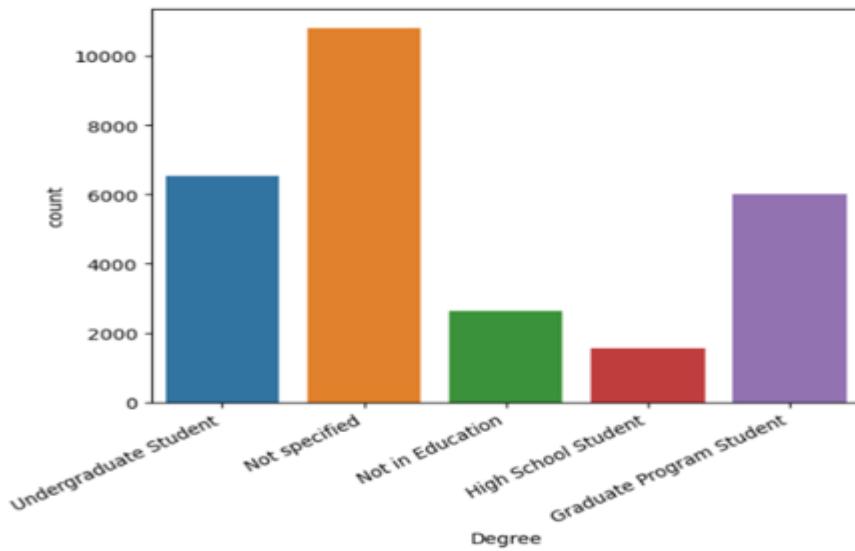
### 5.4 Top 10 Cities based on Number of applicants.

From the graph below, it was observed that the largest number of inquiries came from individuals who chose not to disclose their city. Hyderabad, followed by Saint Louis, had the highest number of inquiries.



**Figure. 5.4: Top Ten Cities of Applicants**

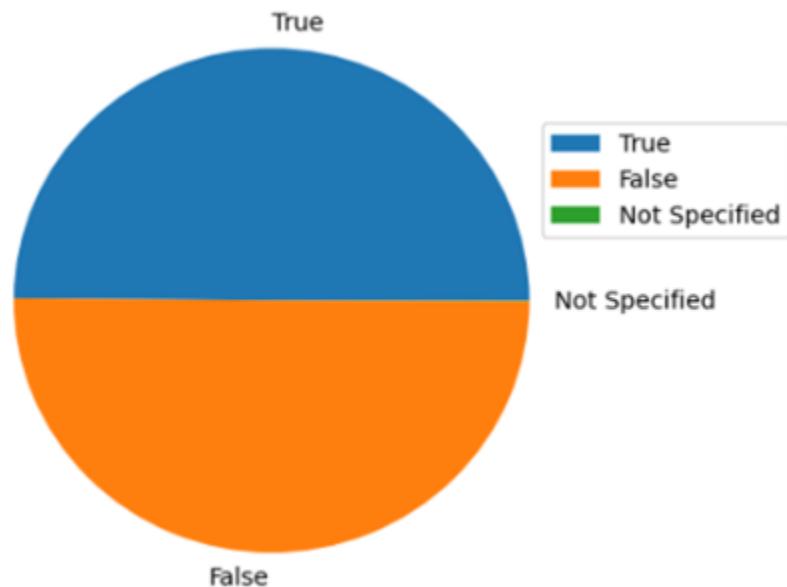
#### 5.4 Count of enquires based on Degree.



**Figure. 5.5: Applicants by Degree**

It was noted that the majority of inquiries were from individuals who opted not to disclose their degree. Among those who did disclose their degree, undergraduate and graduate program students ranked highest in the list.

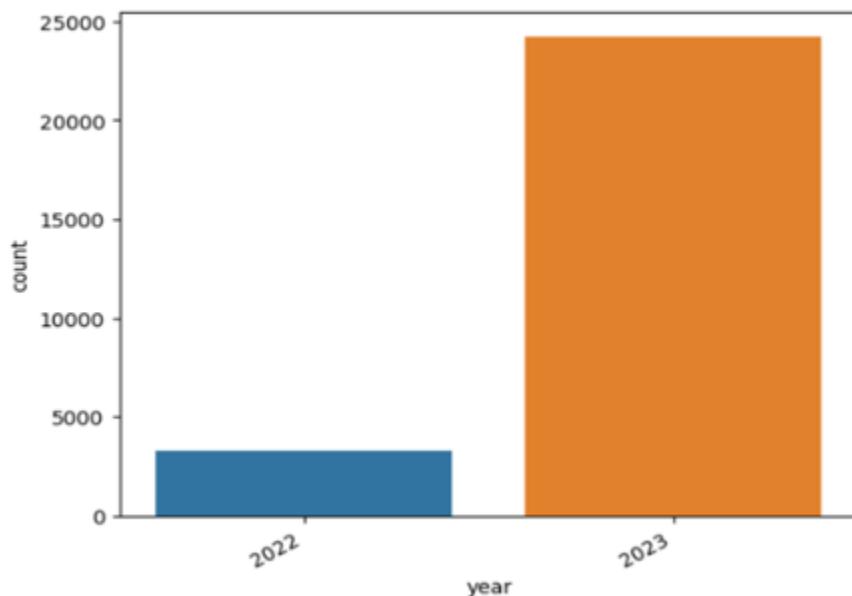
### 5.5 Pie chart based on whether the applicant is from Social Media.



**Figure. 5.6: Social Media Applicants**

The distribution of applicants was consistent, whether they originated from social media channels or other sources. There was an even spread of applicants across both social media and non-social media channels.

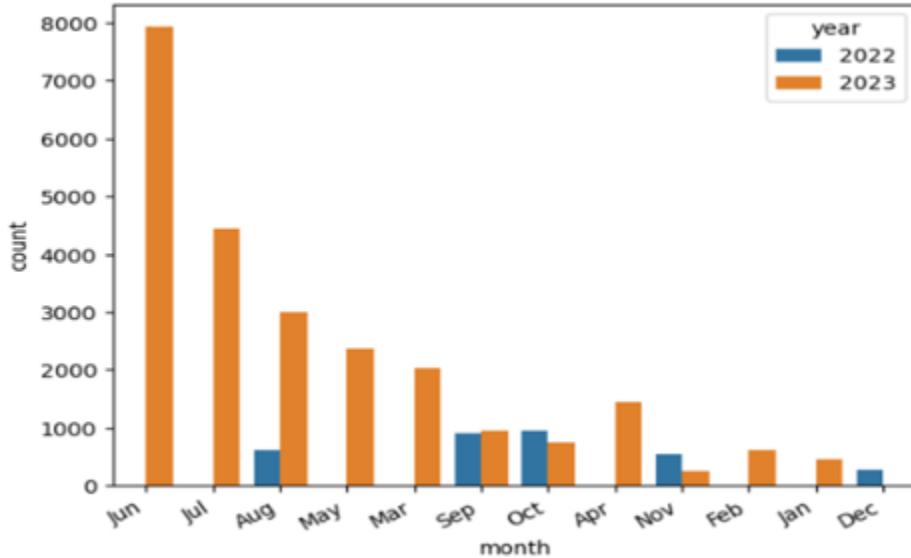
### 5.5 Distribution of Yearwise Applicant count



**Figure. 5.7: Year of Application**

In 2023, the highest volume of inquiries was observed, indicating a peak in interest or engagement. The year experienced the maximum number of inquiries.

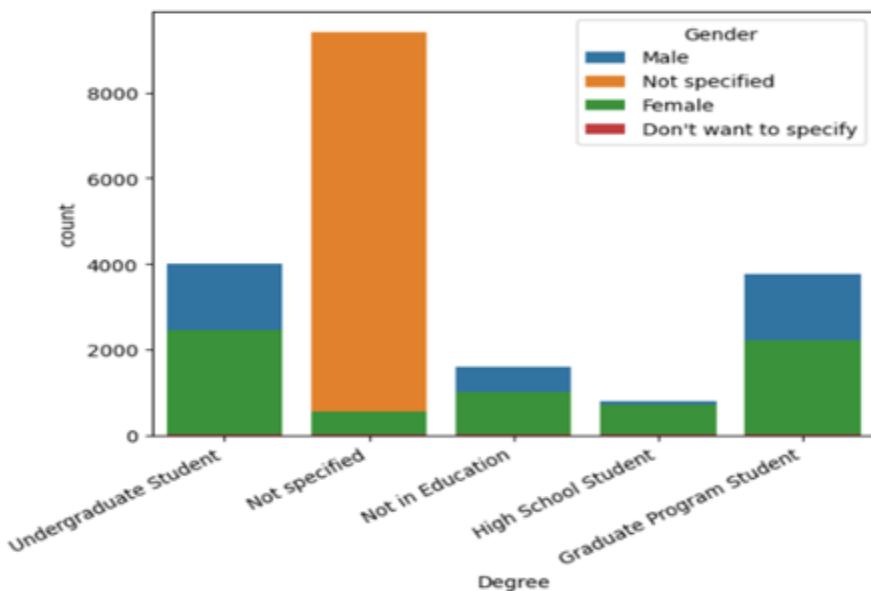
### 5.5 Distribution of month-wise Applicant count



**Figure. 5.8: Month of Application**

June 2023 recorded the highest number of inquiries, with July closely following suit. In contrast, most months in 2022 had significantly lower inquiry counts compared to the heightened interest observed in 2023.

### 5.5 Distribution of Education Based on Gender



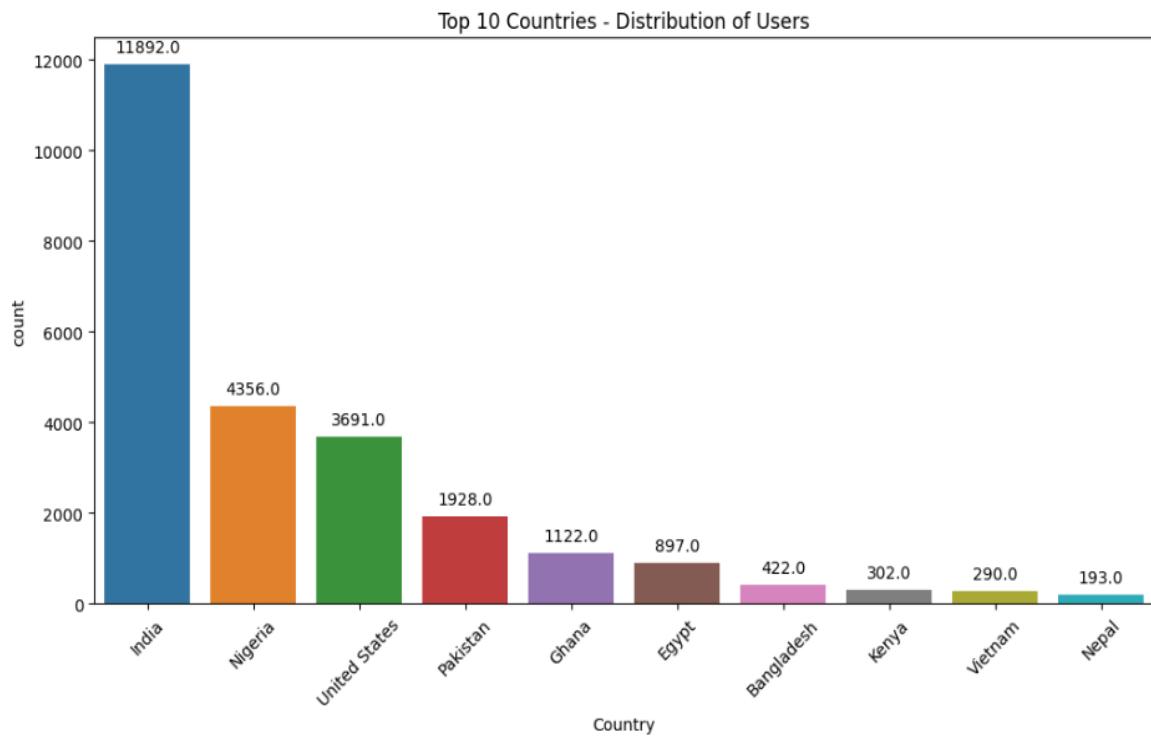
**Figure. 5.9: Relationship between Degree and Gender**

Despite the lower inquiry count for females compared to males, females exhibit a higher percentage of individuals with disclosed degrees. This trend might be attributed to the possibility that a smaller proportion of males chose to disclose their degrees in comparison to females.

## 6.0 Data Preprocessing and Further Visualization

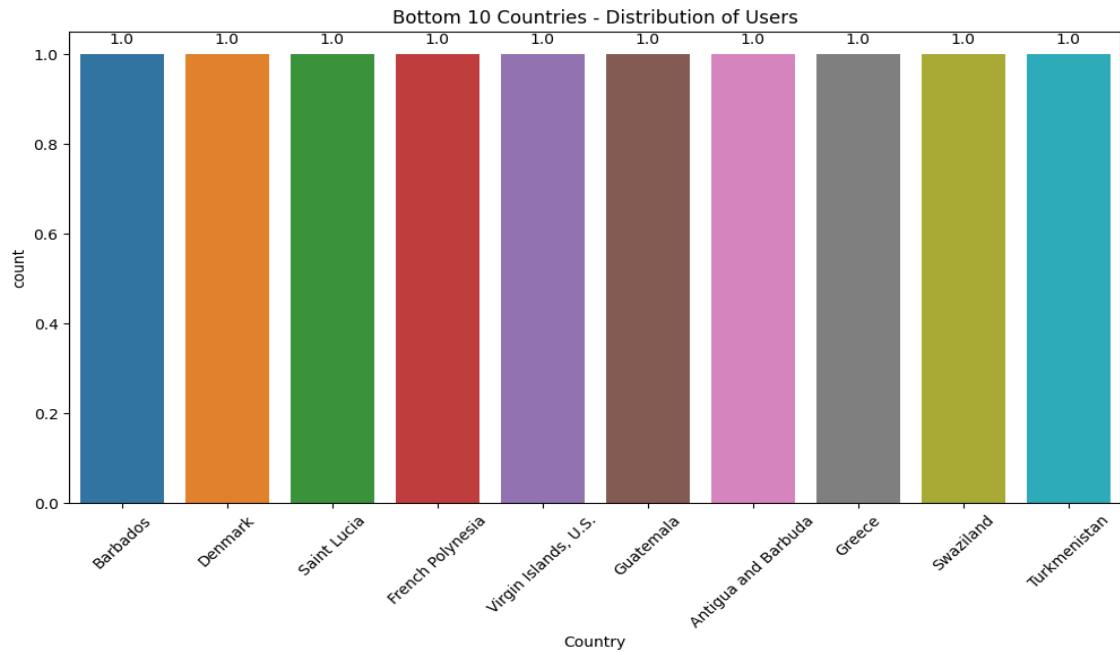
In continuation of the data cleaning and exploratory data analysis, the data was preprocessed and further visualized to derive more insight for the analysis. Since the User Data is categorical and transforming variables to numerical would not give the actual representation of the dataset, features like Day, Week and Month were extracted from the sign up date column to derive insight of the user activity over time.

### 6.1 Country-Wise Distribution



**Figure. 6.1: Top 10 Countries- Distribution of Users**

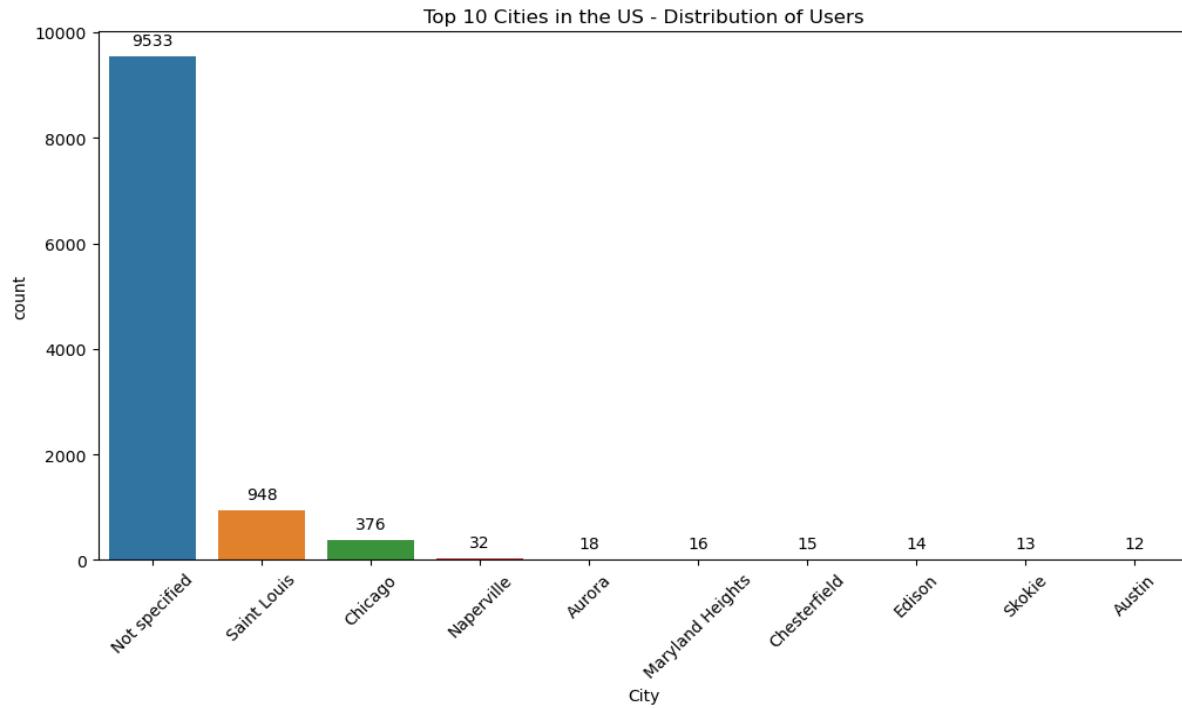
From the chart here, it shows Indians has the highest distribution of users on the platform followed by Nigerians which is then followed by users from the United States.



**Figure. 6.2: Bottom 10 Countries- Distribution of Users**

From the chart here, it shows Turkmenistan has the lowest distribution of users on the platform followed by Swaziland, which is then followed by users from Greece, though there is no significant difference among the bottom ten countries as seen.

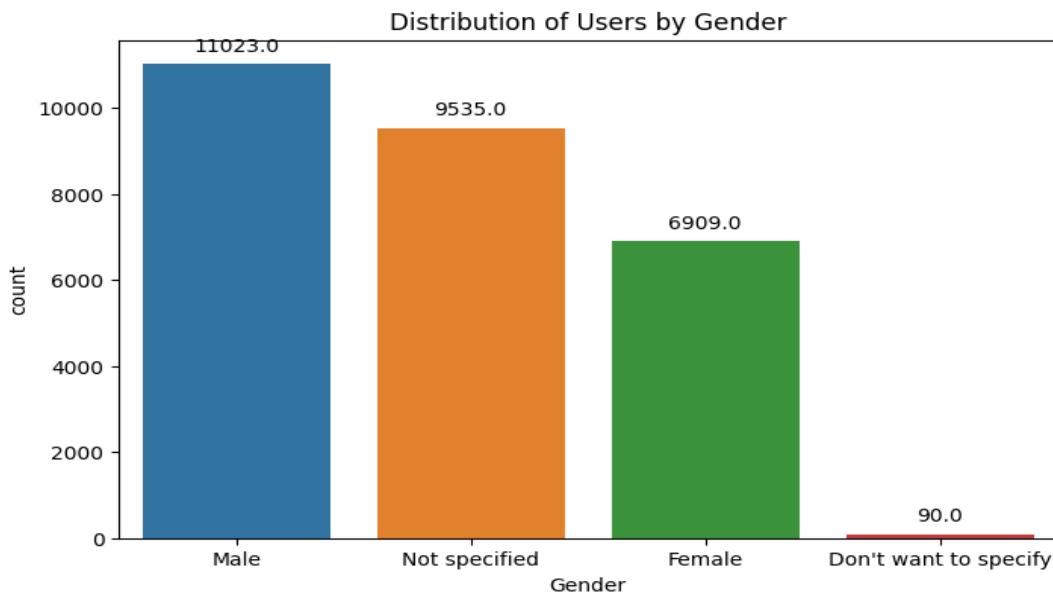
## 6.2 Distribution of the Top 10 Cities in the US



**Figure. 6.3: Top 10 Cities in the US**

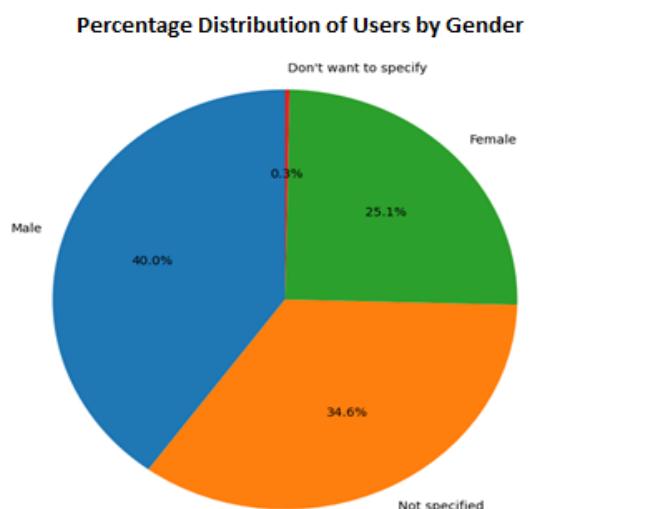
A large number of users in the top 10 US city didn't specify the City where they originate from, among those that specify, users from Saint Louis has shown higher distribution in the United States, followed by Chicago and Naperville.

### 6.3 Distribution of User by Gender



**Figure. 6.4: Gender distribution**

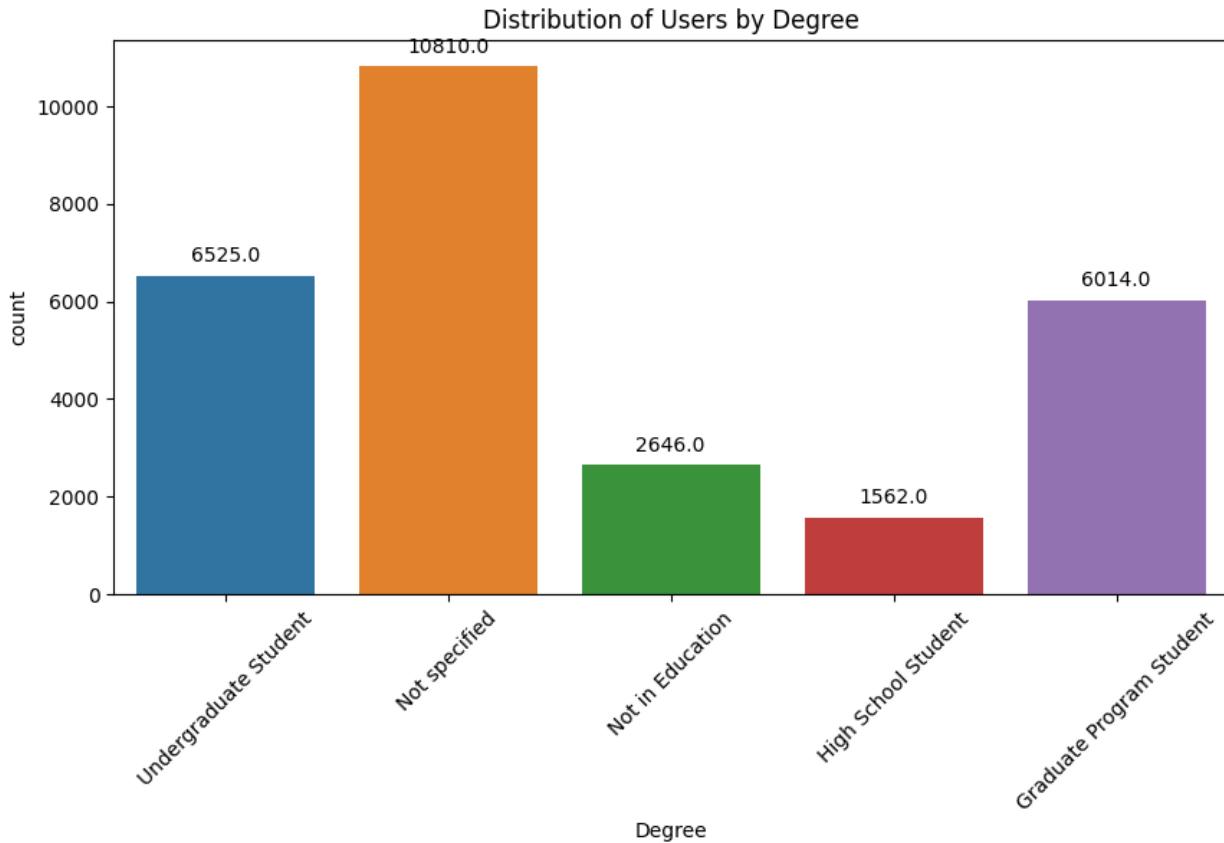
The distribution of users by Gender has shown that a significant number of Males has signed up on the platform, and this is followed by a good number of users who did not specify their Gender.



**Figure. 6.5: Distribution of Users by Gender (%)**

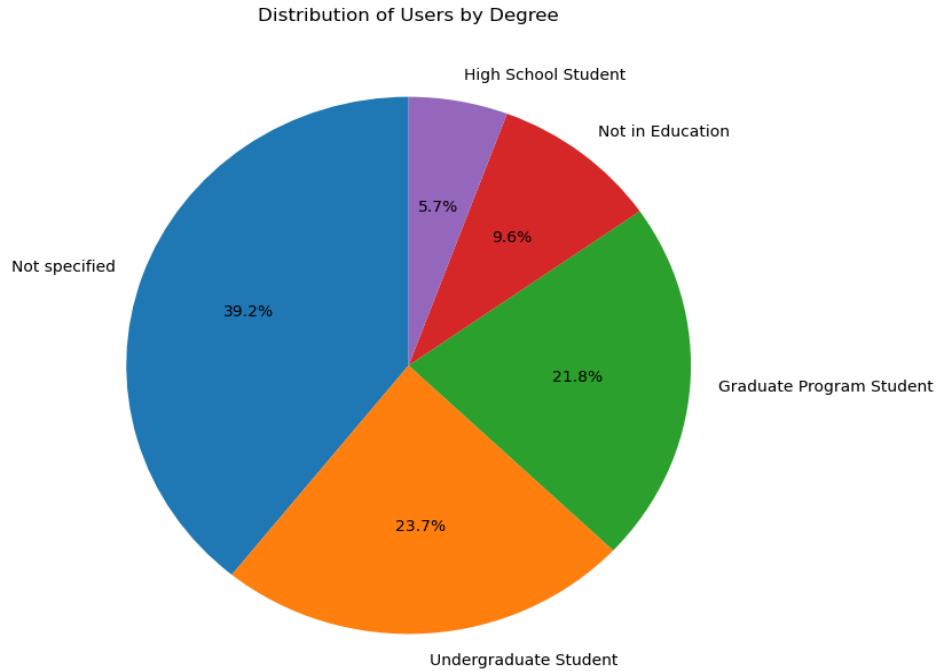
From Figure 6.5 above, 40.0% of users who are Male and 25.1% of users who are Females have signed up on the platform while 34.6% and 0.3% of users are ‘Not specified’ and ‘Don’t want to specify’ respectively. Meanwhile, those who do not want to specify have no significant impact on the users as 0.3% could be negligible.

#### 6.4 Distribution of Users by Degree



**Figure. 6.6: Degree Distribution**

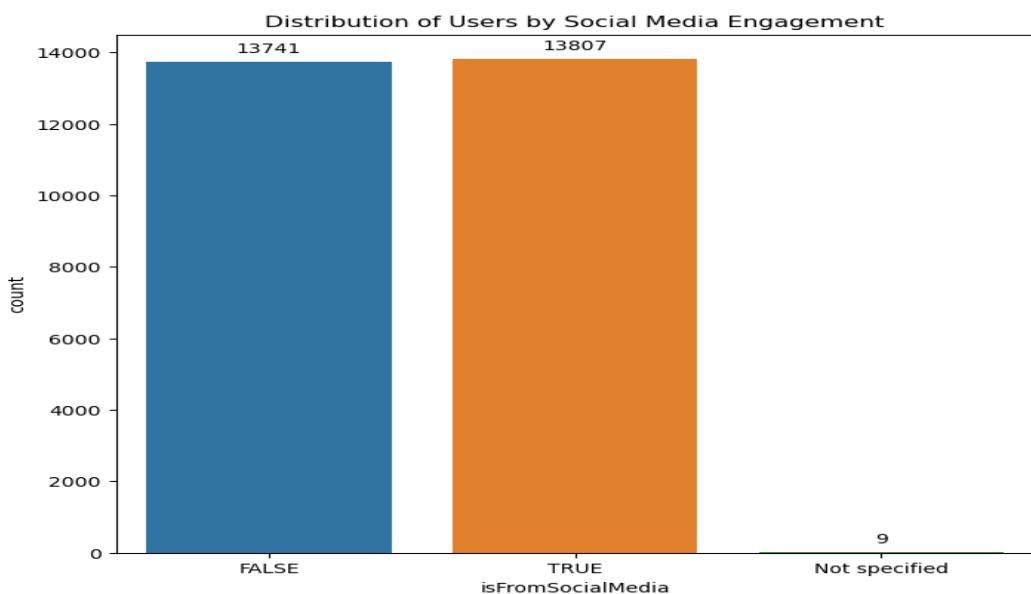
There are 6525 undergraduate students who signed up on the platform compared to the 6014 Graduate Program Students who have signed up. And a high number of 10810 students did not specify their degree. The plot of the degree distribution indicates that a substantial portion of users did not specify their degree. This category likely includes individuals who either chose not to disclose their educational background or have not yet determined their academic status. Following this, among those who did provide information, there is a higher representation of undergraduate students compared to graduate program students. This is further represented in the pie chart below (Figure. 6.7).



**Figure. 6.7: Percentage of Users by Degree**

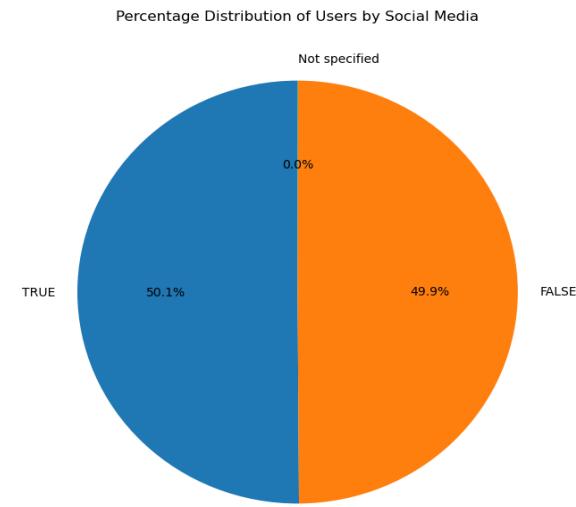
The degree distribution, expressed as percentages, reveals that 32.9% of users did not specify their degree. Among those who provided information, 23.7% identified as undergraduate students, while 21.8% identified as graduate program students. It's notable that 5.7% of users identify as High School Students, indicating a presence of individuals in the pre-college stage. Additionally, 9.6% of users fall into the category of "not in education."

## 6.5 Distribution of Users by Social Media Engagement



**Figure. 6.8: Users Social Media Engagement**

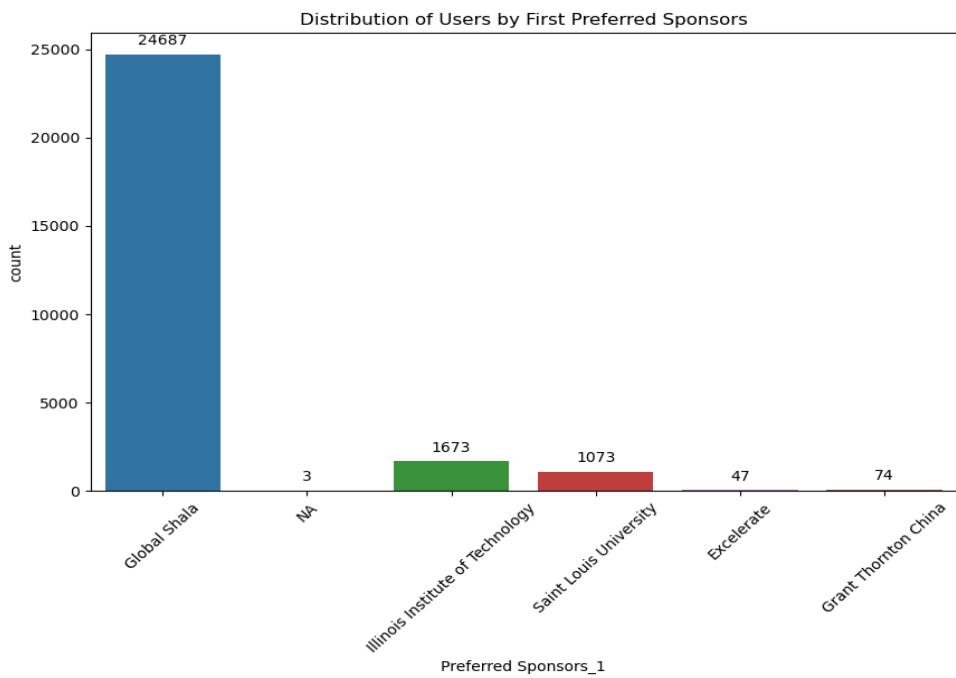
This distribution has shown that more users signed up using social media than other sources. However, there is a close relationship between those who utilized social media and those from other sources. Additionally, those who did not indicate a specific source of application do not affect the application source as it has approximately zero percentage effect as shown in Figure6.9 below.



**Figure. 6.9: Percentage Users by Social Media.**

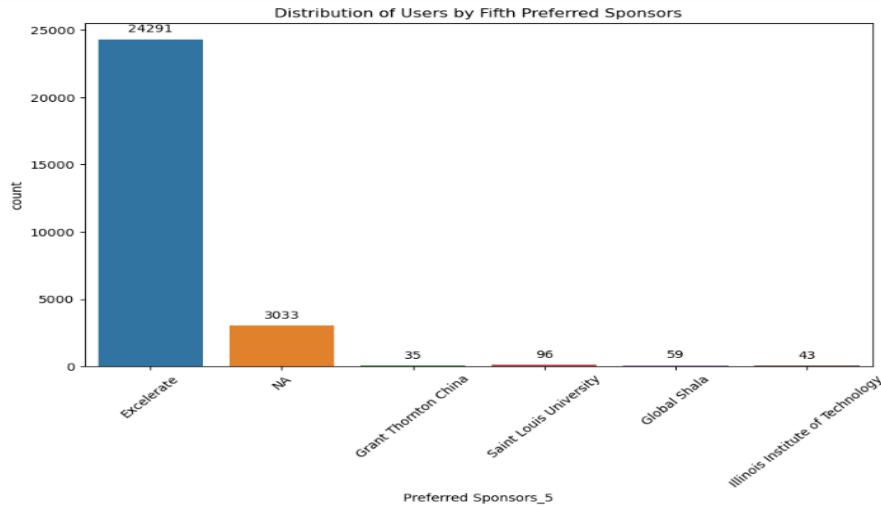
50.1% of users signed up through social media while 49.9% engaged via other sources.

## 6.6 Distribution of Users by Preferred Sponsors.



**Figure. 6.10: First Preferred Sponsors Distribution**

The first preferred sponsors distribution appears to prioritize Global Shala as the primary sponsor, with the highest level of support. Following closely behind, the second and third preferred sponsors are the Illinois Institute of Technology and Saint Louis University, respectively.

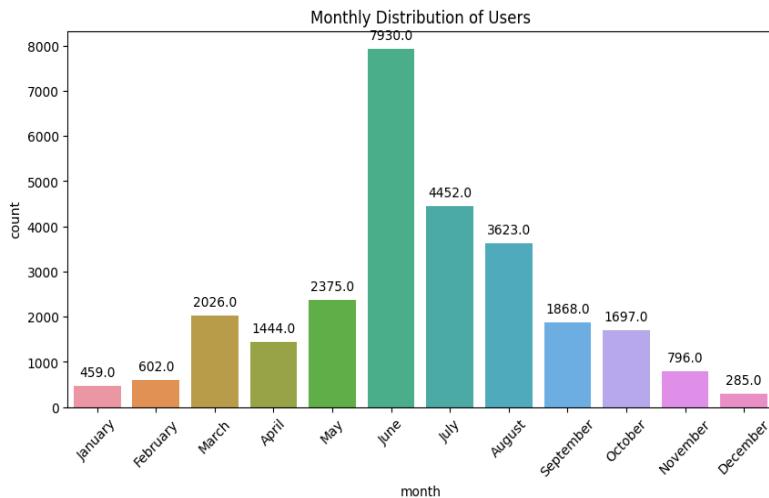


**Figure. 6.11: Fifth Preferred Sponsors Distribution**

In the fifth preferred sponsors' distribution, Excelsrate takes the lead as the highest supporter, indicating a significant level of endorsement and engagement from this entity. Grant Thornton China and Saint Louis University follow as the second and third preferred sponsors, respectively, suggesting substantial contributions and support from these organizations within the Global Shala community.

## 6.7 Time Base User Distribution Trend.

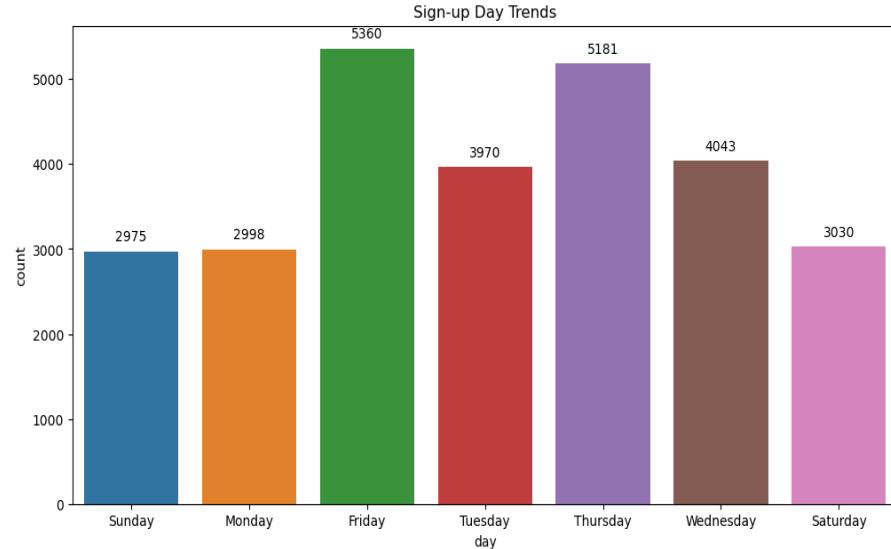
### 6.7.1 Distribution of Users by Month



**Figure. 6.12: User Distribution by Month**

The monthly distribution of users has shown that June has more traffic of 7930 users, followed by July with a traffic of 4452 users, and 3623 users for August.

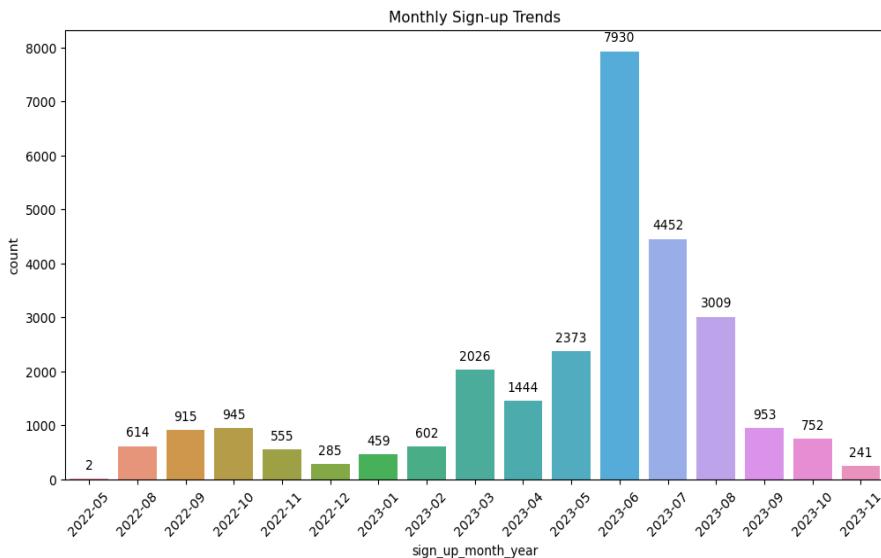
### 6.7.2 Distribution of Sign-up Day Trend



**Figure. 6.13: Sign-up Day Trends**

Based on the observed sign-up day trends, it appears that Friday consistently emerges as the most popular day for user registrations. This inference is supported by the data indicating that a significant number of individuals choose to sign up on Fridays. Following closely behind, Thursday, and Wednesday also exhibit higher sign-up rates compared to other days of the week.

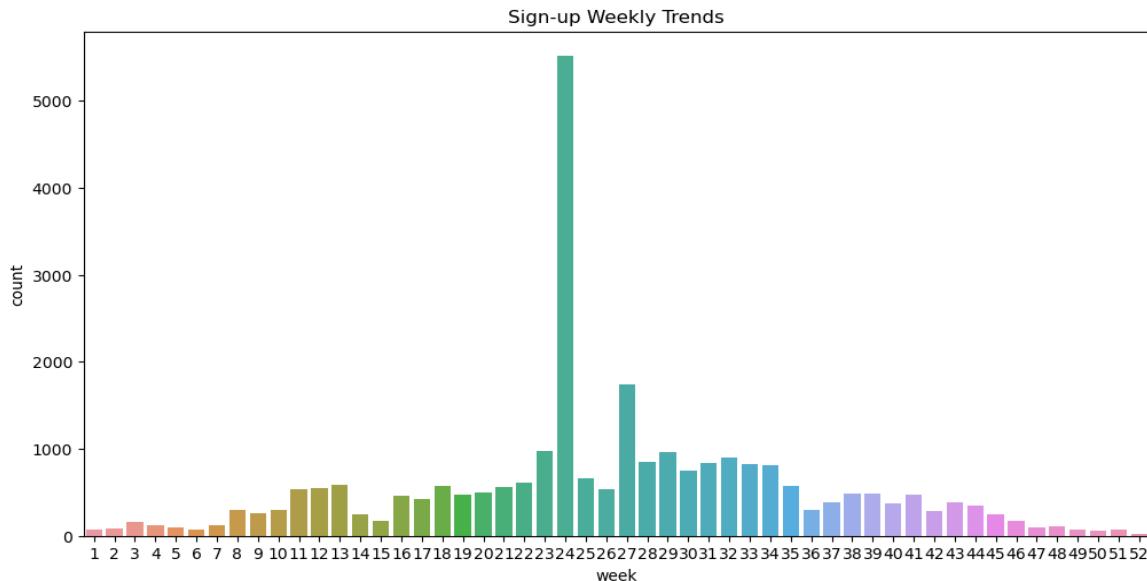
### 6.7.3 Distribution of Month of the Year Sign-up Trends



**Figure. 7.14: Monthly Sign-up Trends**

The monthly signup trend reveals a clear pattern, with June standing out as the month with the highest number of signups. Following closely behind, July and August demonstrate the second and third-highest signup rates, respectively. This trend suggests a peak in user engagement and interest during the early summer months.

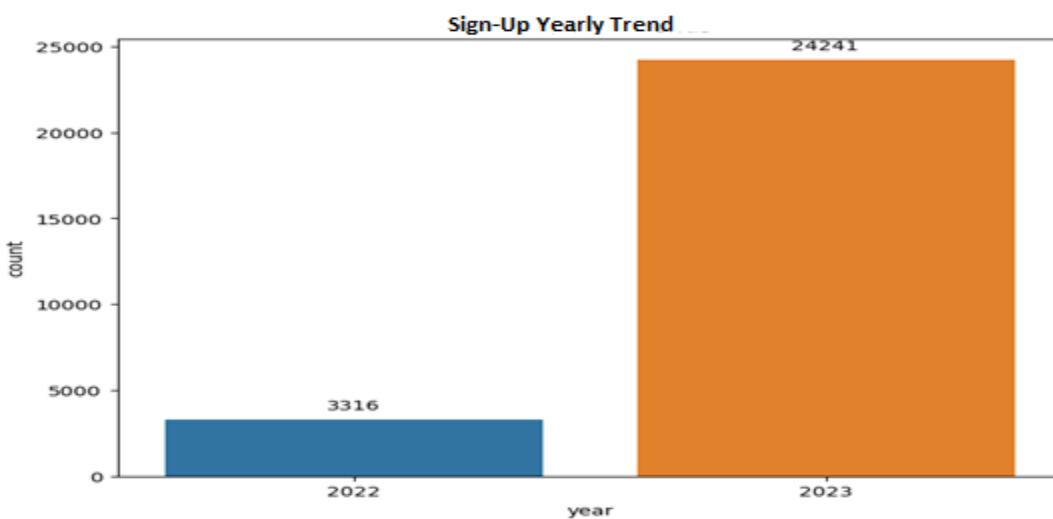
#### 6.7.4 Distribution of Sign-up Weekly Trends



**Figure. 6.15 Sign-up Weekly Trends**

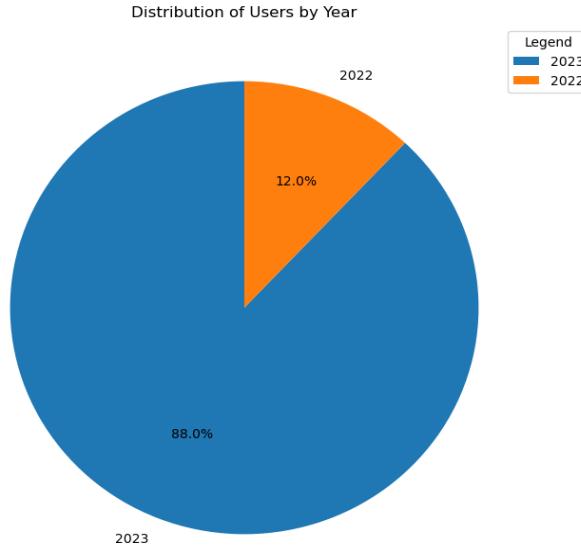
The sign-up weekly trends indicate that Week 24 experienced the highest number of sign-ups, followed by Week 27 and Week 29. This pattern suggests specific peaks in user engagement and registration activity during these weeks.

#### 6.7.5 Distribution of Sign-up Yearly Trends



**Figure. 6.16 Sign-up Year Trends**

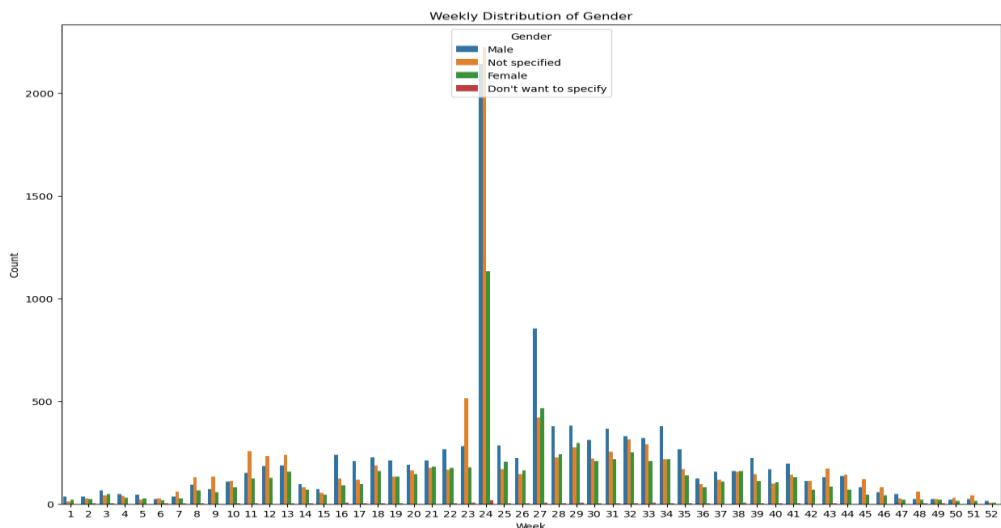
The sign-up yearly trends reveal a significant increase in user registrations from 2022 to 2023. In 2022, the platform garnered 3,316 users, while in 2023, the number of sign-ups surged to 24,241 users. The percentage distribution is shown in Figure 6.17 below.



**Figure. 6.17 User by Year (%)**

The distribution by year indicates a substantial shift in user sign-ups, with 88.0% of registrations occurring in 2023 and the remaining 12.0% taking place in 2022. This distribution reflects a significant increase in user engagement and adoption of the platform in the more recent year.

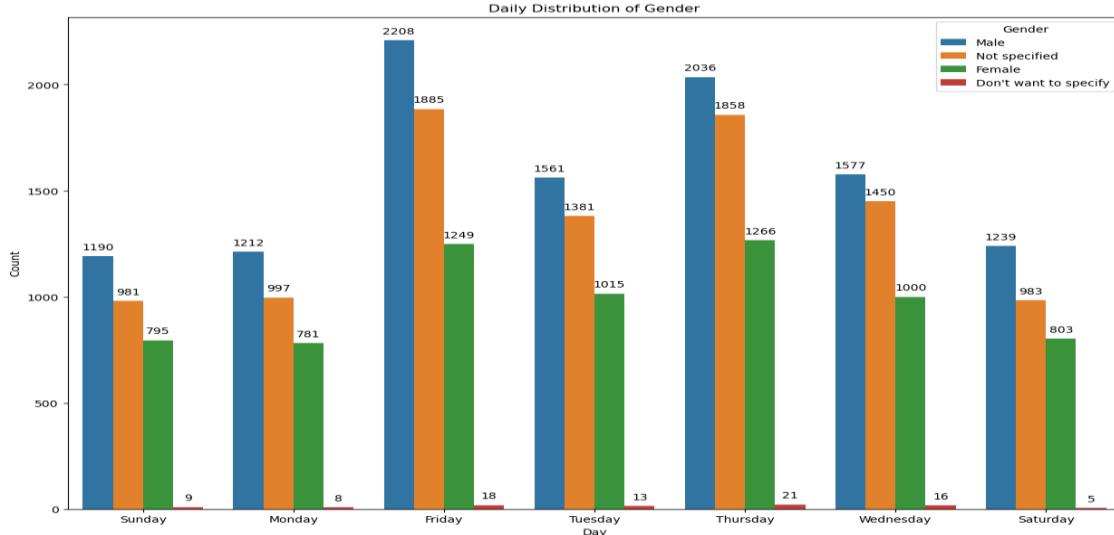
#### 6.7.6 Weekly Distribution of Gender



**Figure. 6.18 Weekly Distribution of Gender**

The gender distribution over time reveals specific patterns, particularly in Week 24. In this week, there appears to be a higher number of male users compared to females. Additionally, a notable portion of users did not specify their gender.

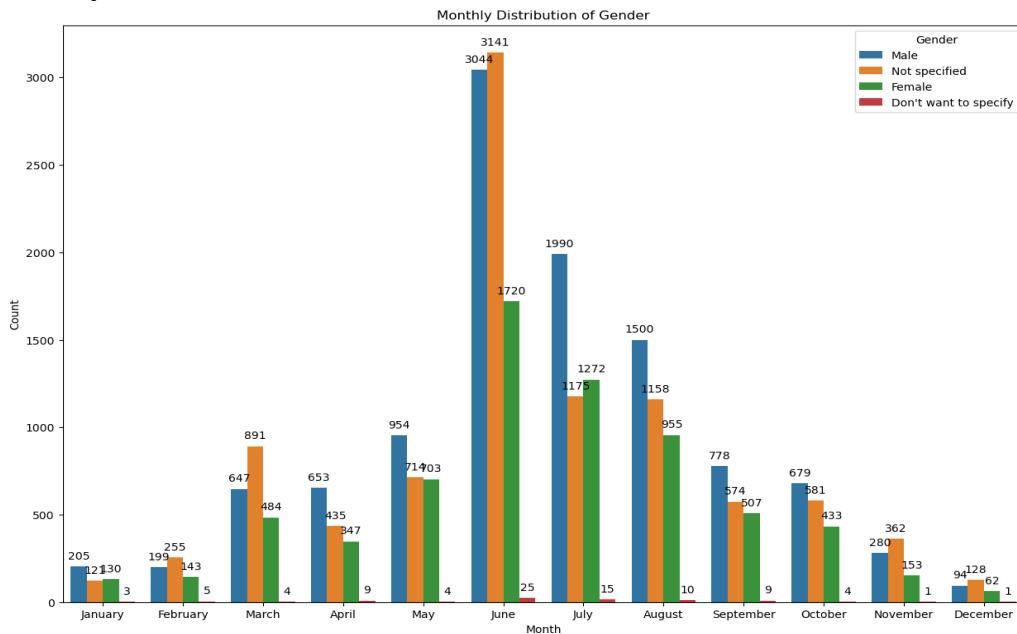
### 6.7.7 Daily Distribution of Gender



**Figure. 6.19: Daily Distribution of Gender**

The daily gender distribution indicates that on Fridays, there is a higher number of male users who have signed up compared to females. Additionally, it is noteworthy that the number of users who didn't specify their gender surpasses the number of female users.

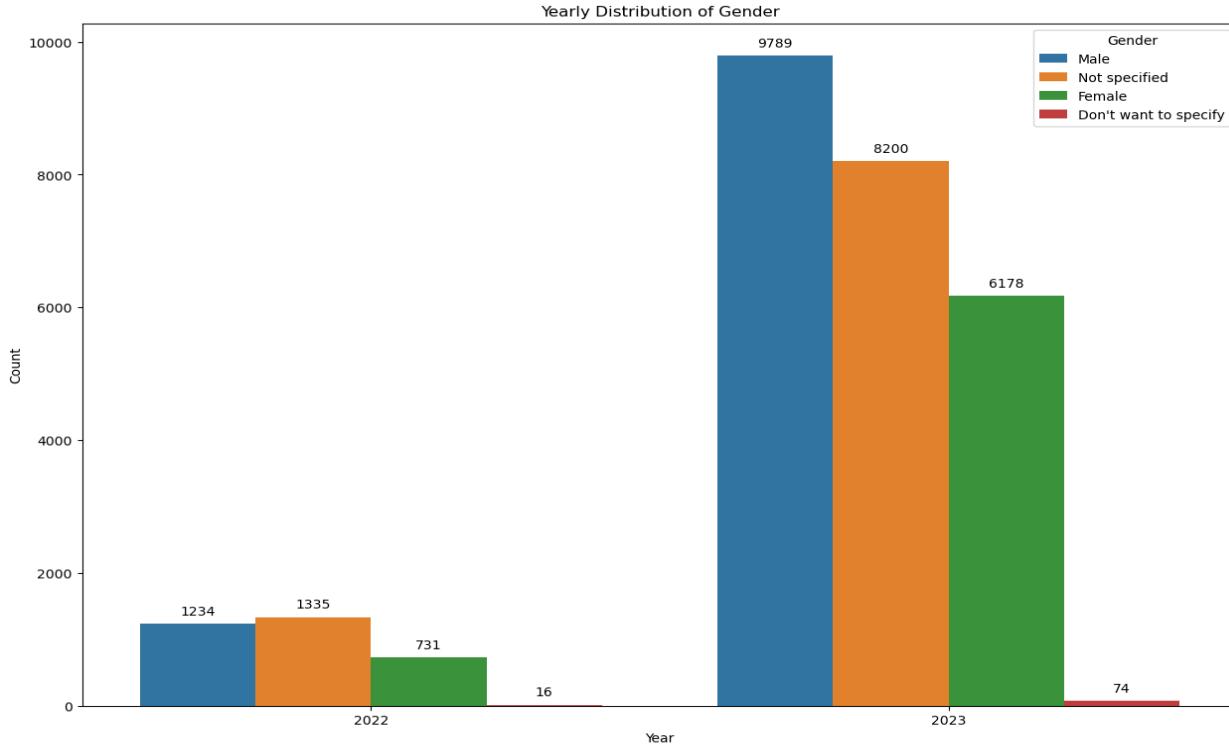
### 6.7.8 Monthly Distribution of Gender



**Figure. 6.20: Monthly Distribution of Gender**

The gender distribution by month reveals that during June, there is a higher number of male users compared to females. Additionally, it's noteworthy that the number of users who didn't specify their gender surpasses the number of female users.

### 6.7.9 Yearly Distribution of Gender

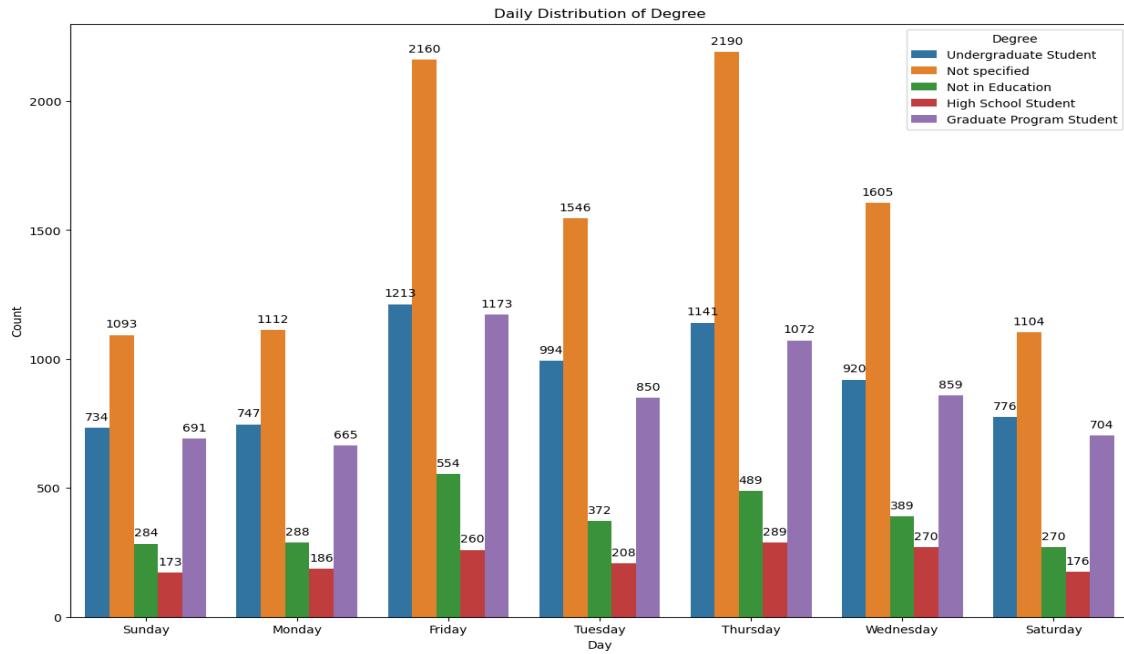


**Figure. 6.21 Yearly Distribution of Gender**

The yearly gender distribution highlights interesting trends between 2022 and 2023. In 2022, users who didn't specify their gender are more numerous, followed by a higher number of male users, while female users are the least represented. On the other hand, in 2023, there is a shift in the gender distribution, with male users being the most numerous, followed by those who didn't specify their gender, and then female users.

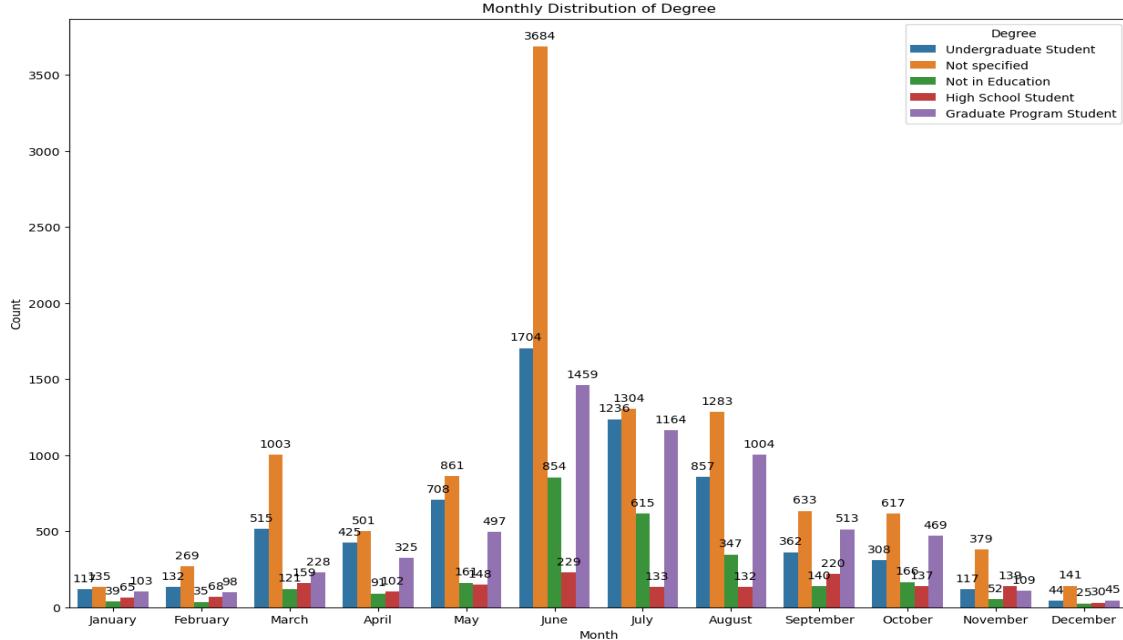
### 6.7.10 Daily Distribution of Degree

In the graph below, the daily distribution of degrees on Friday, the peak of user signups indicates that a significant proportion of users in this period were undergraduate students. Following closely behind are graduate program students, suggesting a notable representation of individuals pursuing higher education. In comparison, those not in education and high school students show lower participation levels during this peak signup period.



**Figure. 6.22: Daily Distribution of Degree**

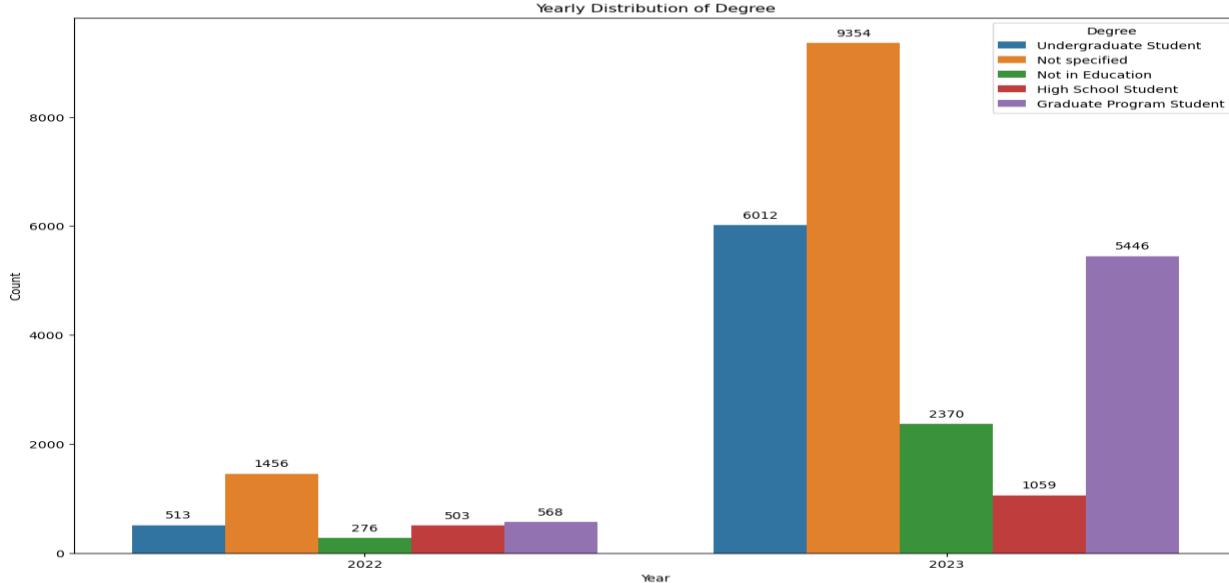
#### 6.7.11 Daily Distribution of Degree



**Figure. 6.23: Monthly Distribution of Degree**

The monthly distribution of degrees highlights distinctive patterns in user demographics. Specifically, in the month of June, there is a notable prevalence of undergraduate users, surpassing the number of high school students. Additionally, users who chose not to specify their educational status stand out as the largest group during this period.

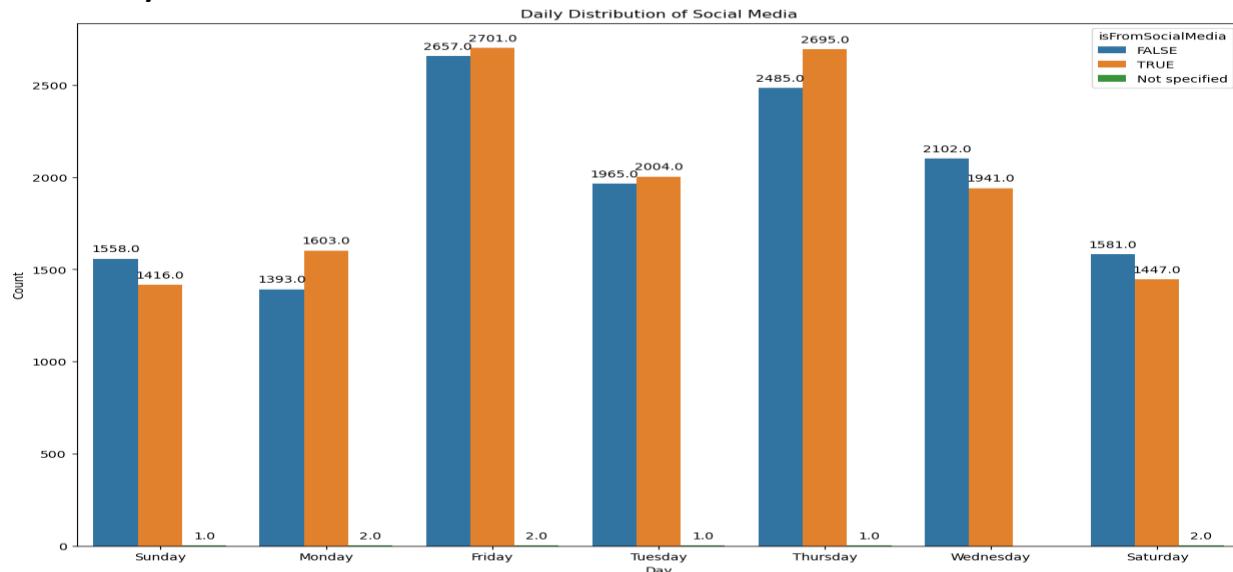
### 6.7.12 Yearly Distribution of Degree



**Figure.6.24: Yearly Distribution of Degree**

The yearly distribution of degrees reveals distinct patterns in user demographics between the years 2022 and 2023.

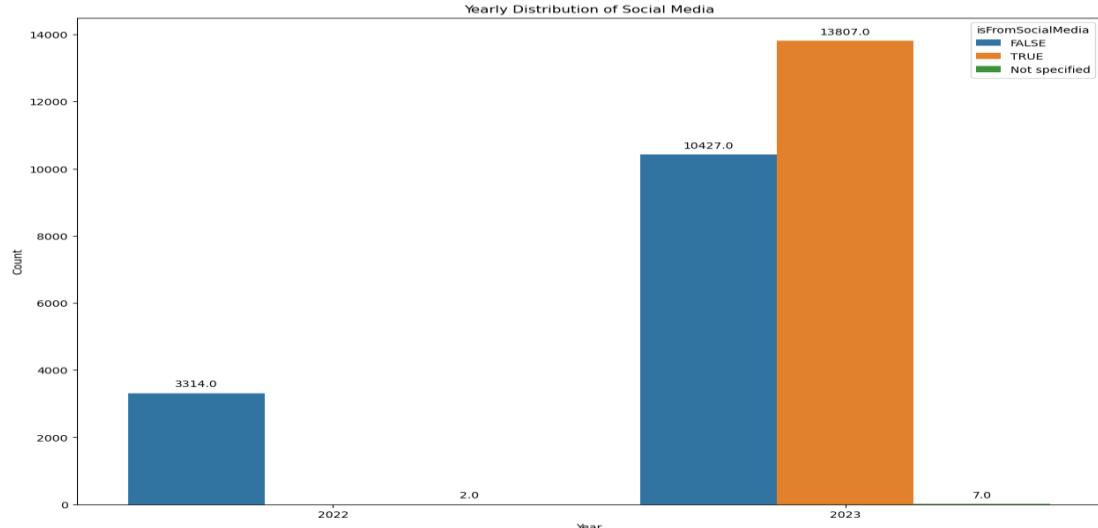
### 6.7.13 Daily Distribution of Social Media



**Figure. 6.25: Daily Distribution of Social Media.**

The daily distribution of social media signups indicates distinct patterns, with Monday, Wednesday, and Saturday standing out as the days when a significant number of users choose to sign up through social media channels. In contrast, other days exhibit lower levels of signups through social media, with more users opting for manual signup methods.

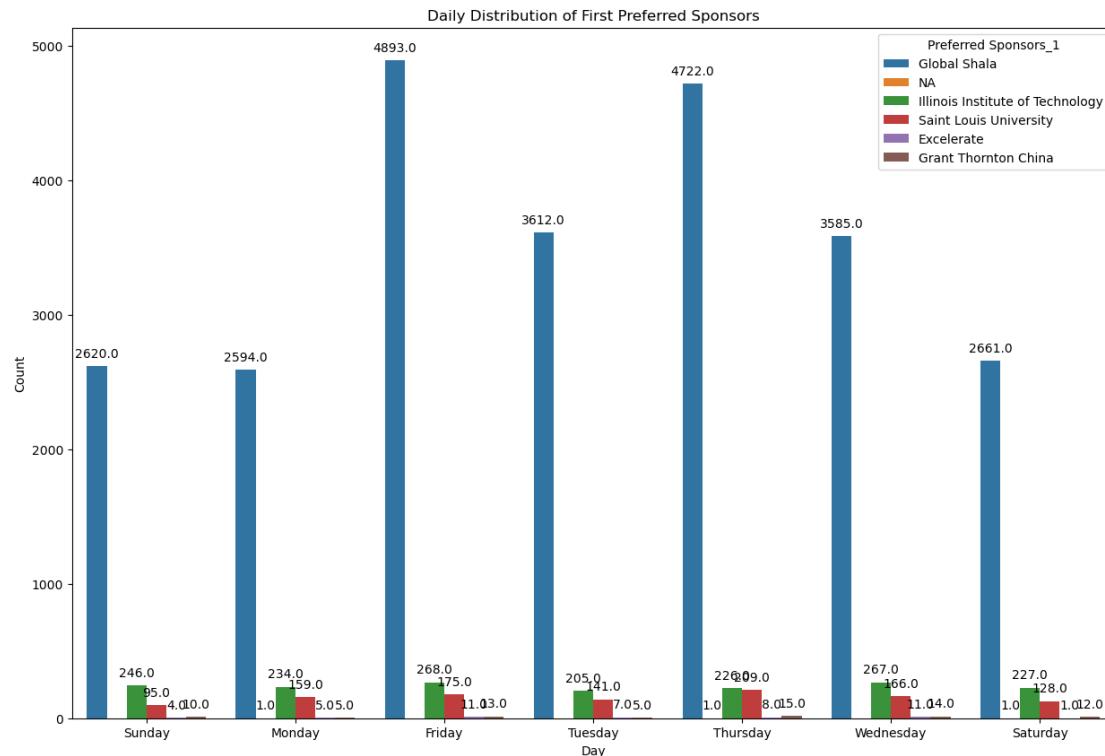
#### 6.7.14 Weekly Distribution of Social Media



**Figure. 6.26: Weekly Distribution of Social Media**

The yearly comparison between 2022 and 2023 in terms of signup methods reveals a notable shift in user behavior. In 2022, a higher number of users opted to sign up through social media channels, suggesting a preference for this method. However, in 2023, there was a change in this trend, with more users choosing to sign up manually compared to social media.

#### 6.7.15 Daily Distribution of First Preferred Sponsors



**Figure. 6.27: Weekly Distribution of Preferred Sponsors**

The daily distribution of first preferred sponsors indicates a consistent pattern, with Global Shala having the highest number of distributions among users on each day of the week. This suggests a prominent and continuous endorsement of Global Shala as the primary or preferred sponsor across all days.

#### 6.7.16 Monthly Distribution of First Preferred Sponsors

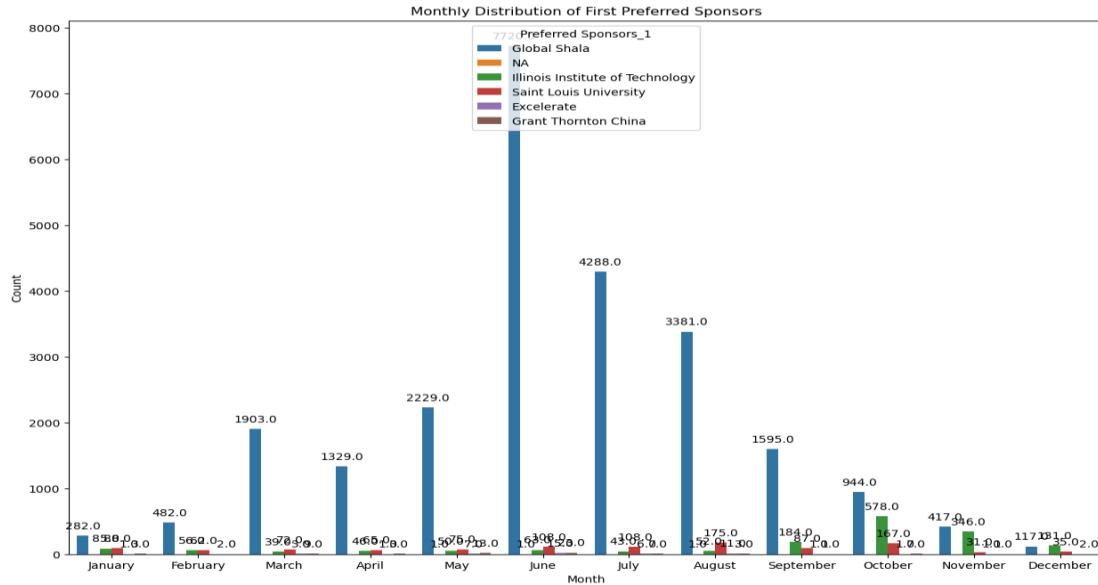


Figure. 6.28: Monthly Distribution of Preferred Sponsors

The monthly distribution of first preferred sponsors reveals a consistent trend, with Global Shala consistently having the highest number of distributions among users in each month of the year. Additionally, the mention of a peak in all months suggests that the trend is not only consistent but also experiences a notable surge in June.

#### 6.7.17 Yearly Distribution of First Preferred Sponsors

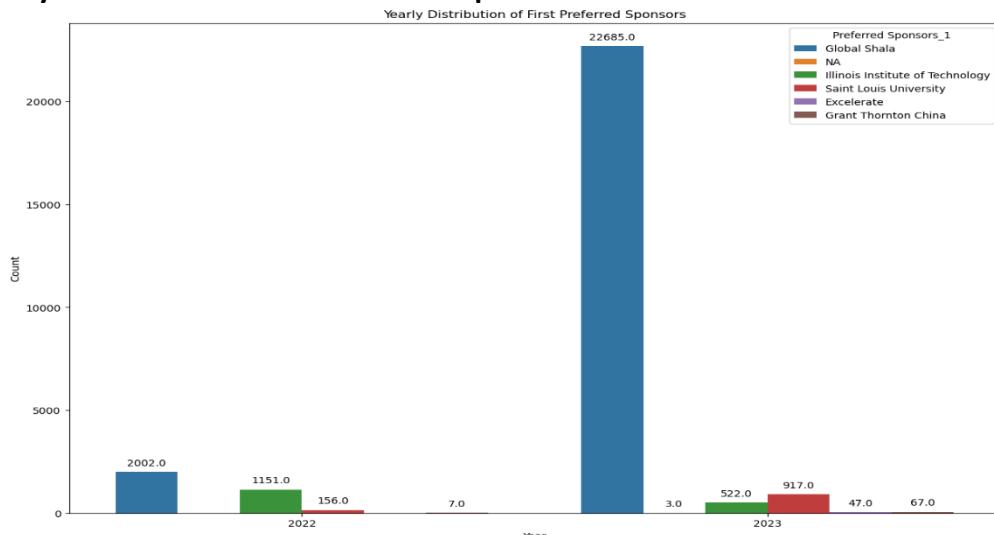
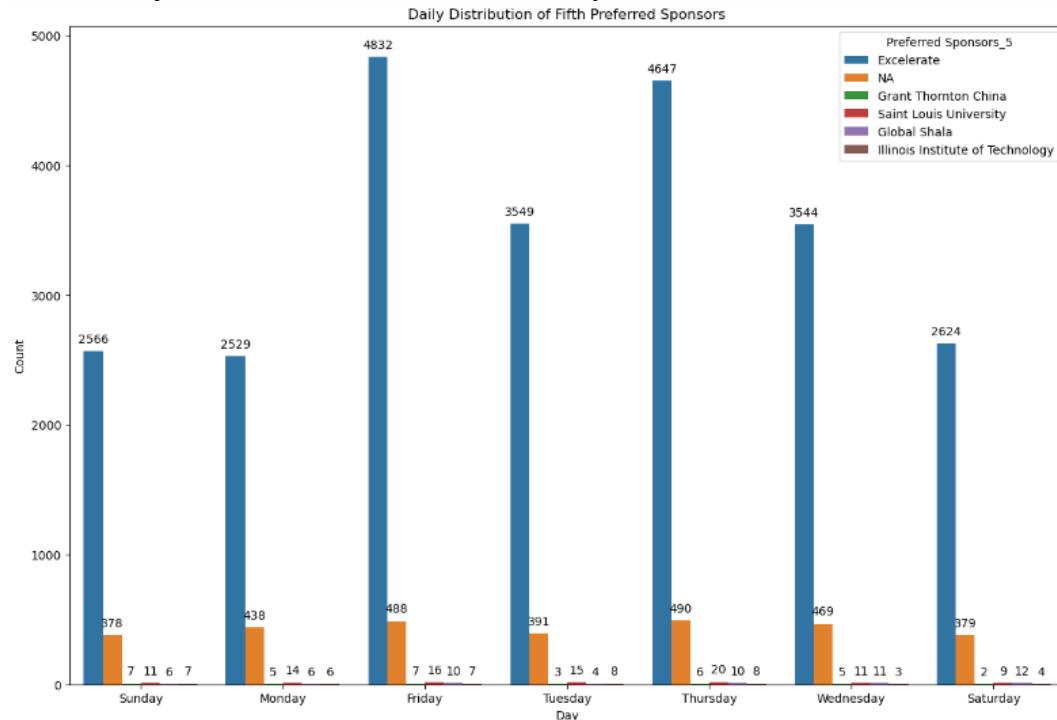


Figure. 6.29: Yearly Distribution of Preferred Sponsors

The yearly distribution of first preferred sponsors indicates a notable trend, with Global Shala having a higher number of distributions among users in the year 2023 compared to 2022. The mention of a peak in 2023 further suggests that Global Shala's prominence as the first preferred sponsor reached its zenith during that year.

### 6.7.18 Daily Distribution of Fifth Preferred Sponsors

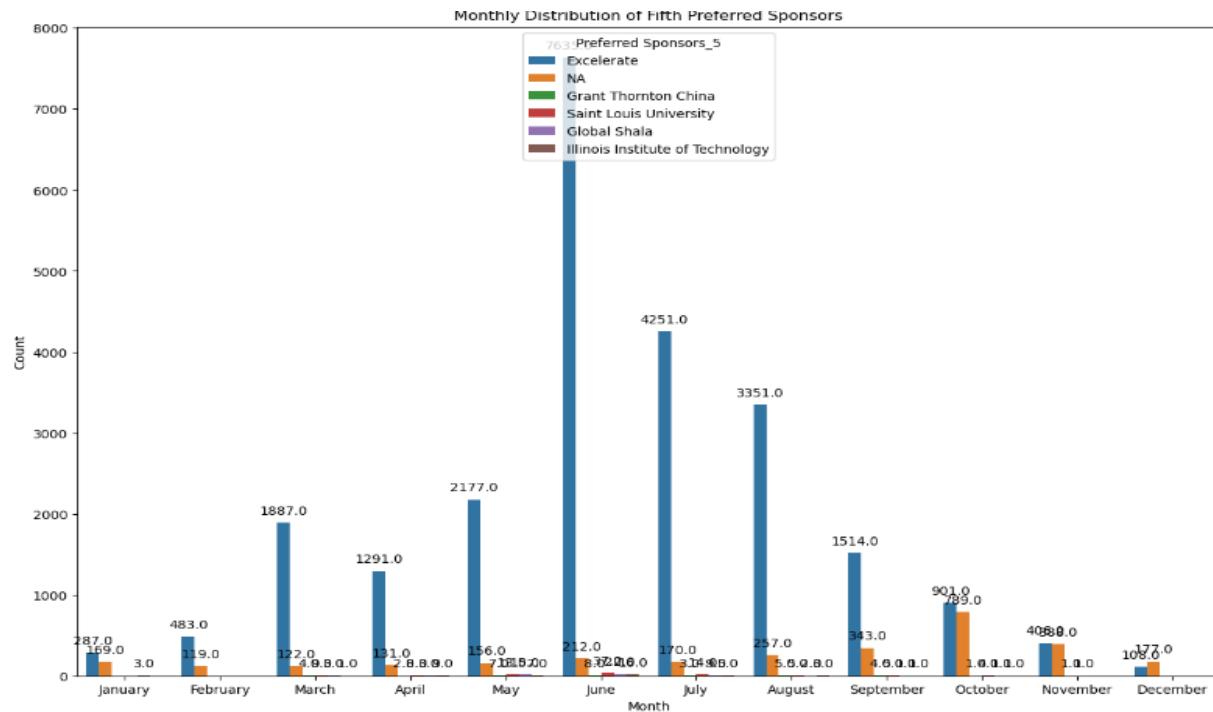


**Figure. 6.30: Daily Distribution of Fifth Preferred Sponsors**

The daily distribution of the fifth preferred sponsor indicates that Excelerate consistently surpasses all other sponsors on each day. This suggests that Excelerate holds a significant and consistent position as the fifth preferred sponsor, with a higher number of distributions compared to other sponsors in this.

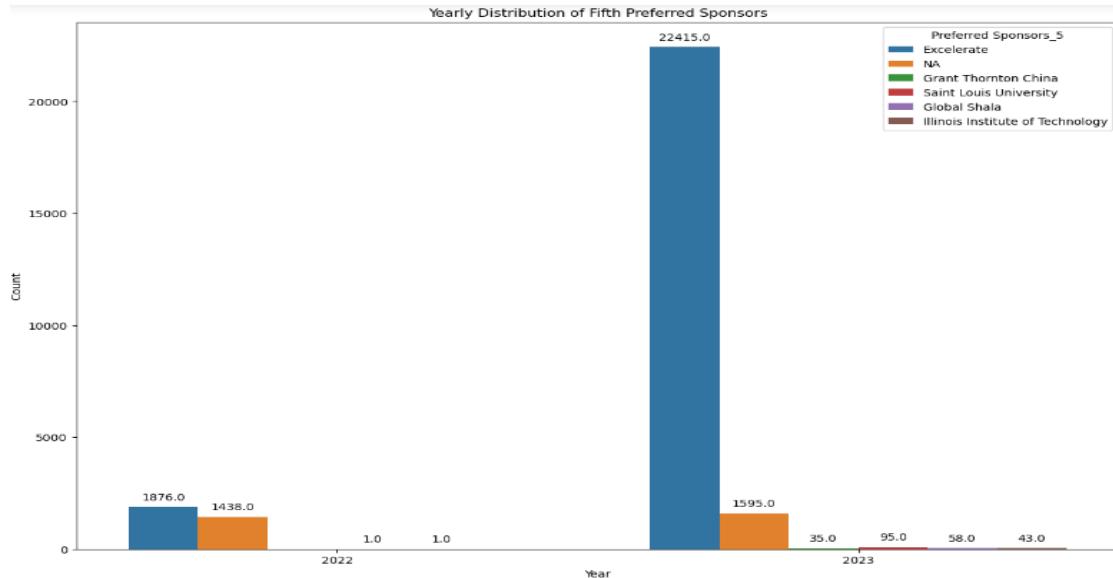
### 6.7.19 Monthly Distribution of Fifth Preferred Sponsors

In the analysis below, the monthly distribution of the fifth preferred sponsor reveals a consistent and enduring pattern, indicating that Excelerate surpasses all other sponsors for whole months except in December. This suggests that Excelerate maintains a strong and sustained position as the fifth preferred sponsor across the entire duration of the month.



**Figure. 6.31: Monthly Distribution of Fifth Preferred Sponsors**

#### 6.7.20 Yearly Distribution of Fifth Preferred Sponsors

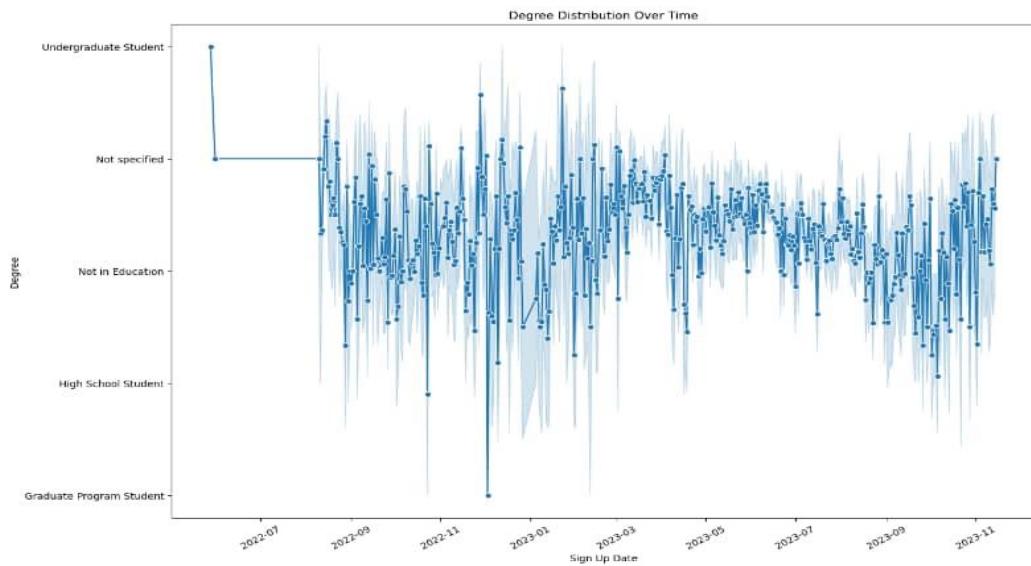


**Figure. 6.32 Yearly Distribution of Fifth Preferred Sponsors**

The yearly distribution of the fifth preferred sponsor reveals a consistent and enduring pattern, indicating that Excelserate surpasses all other sponsors for both the years 2022 and 2023. This suggests that Excelserate maintains a strong and sustained position as the fifth preferred sponsor across the entire duration of both years.

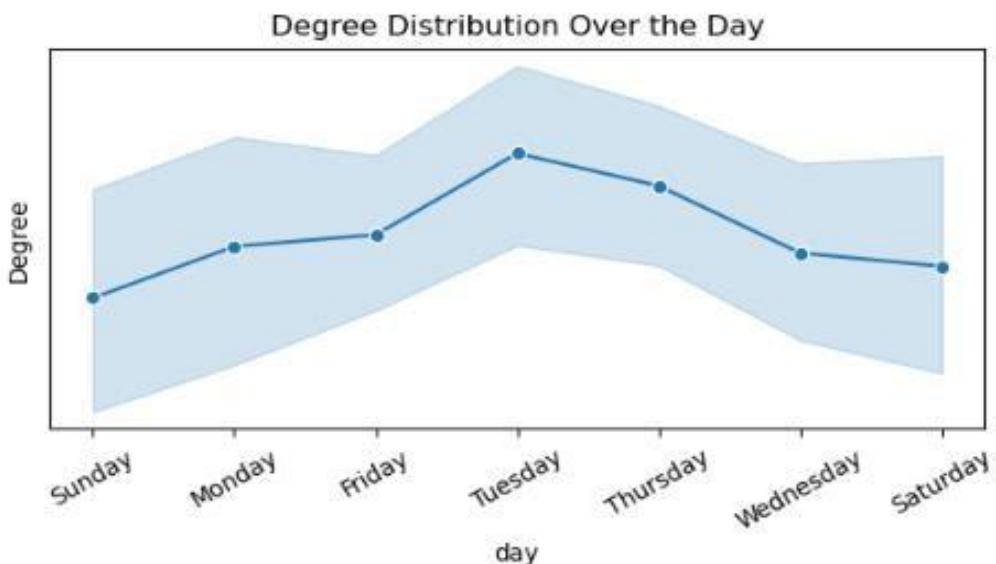
## 6.8 Line plot for User Distribution over Time

### 6.8.1 Degree Distribution Over Time

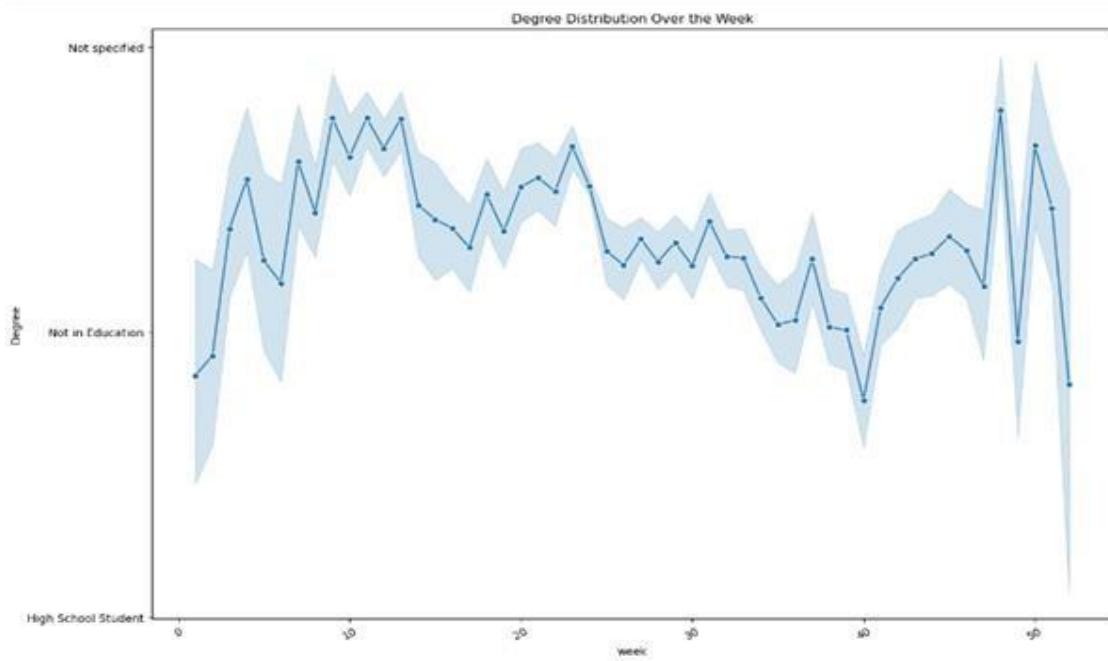


**Figure. 6.33(a) Degree Distribution Over Time**

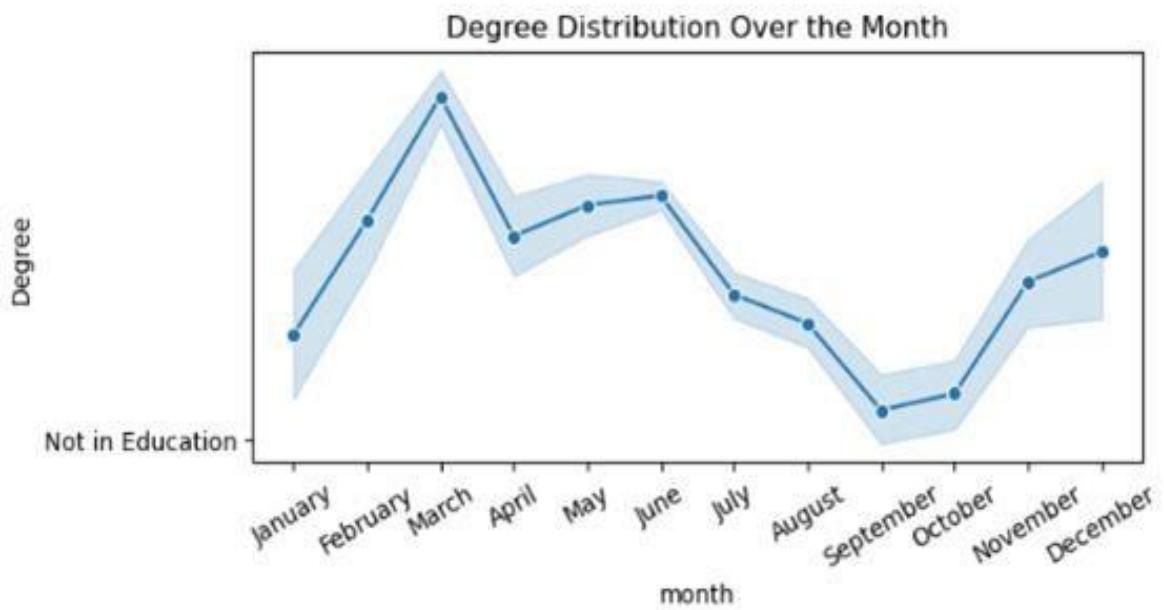
Figure33(a) to (d) show the trend of degree distribution over time. While Figure33(a) displays the complete time trend, 33(b) to (d) show the trends extracted on daily, weekly, and monthly basis. The Figures demonstrated an inconsistent pattern and fluctuations of users over the period and areas of increased activities were also identified based on the degree of users.



**Figure. 6.33(b): Daily Degree Distribution**



**Figure. 6.33(c): Weekly Degree Distribution**



**Figure. 6.33(d): Degree Distribution by Month.**

### 6.8.2 Gender Distribution Over Time

Just as shown in Figures 33(a) to (d), Figures 34(a) to (d) show the trend of gender distribution over time and daily, weekly, and monthly activities were also extracted. The Figures demonstrated an inconsistent pattern over the period and areas of increased activities were also identified based on the gender of users.

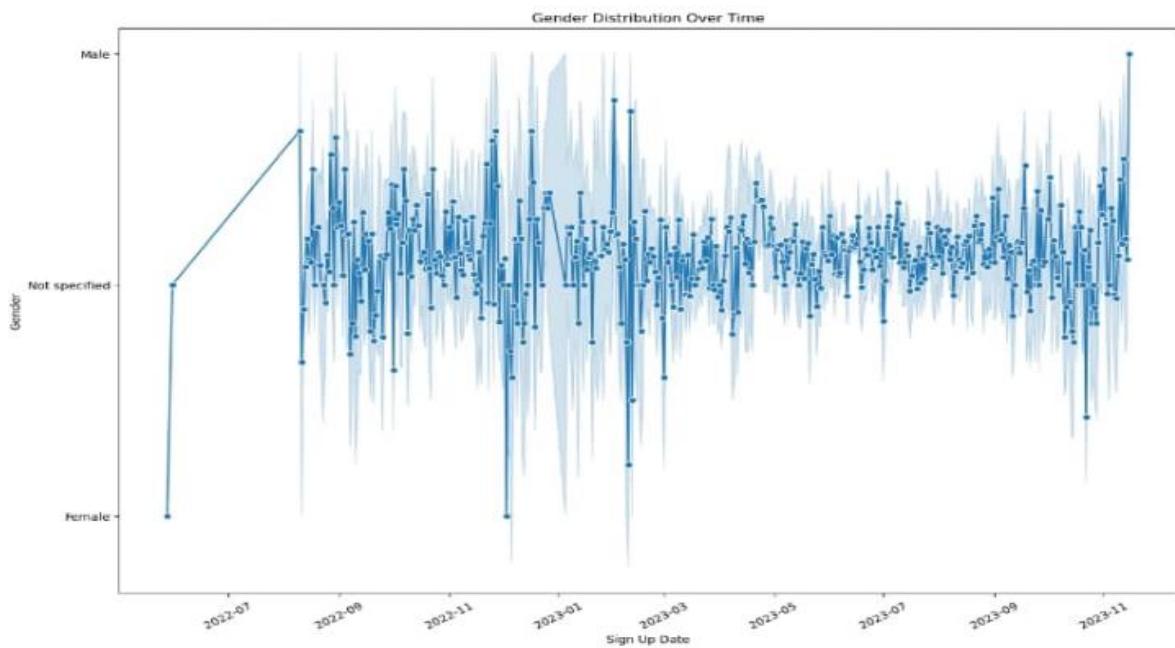


Figure. 6.34(a): Gender Distribution Over Time

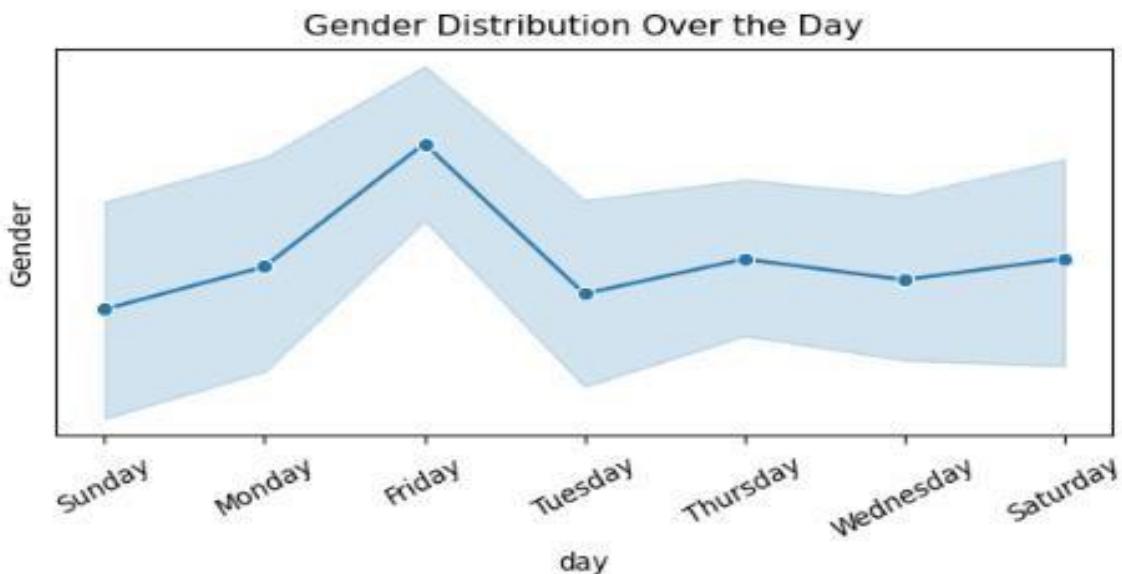
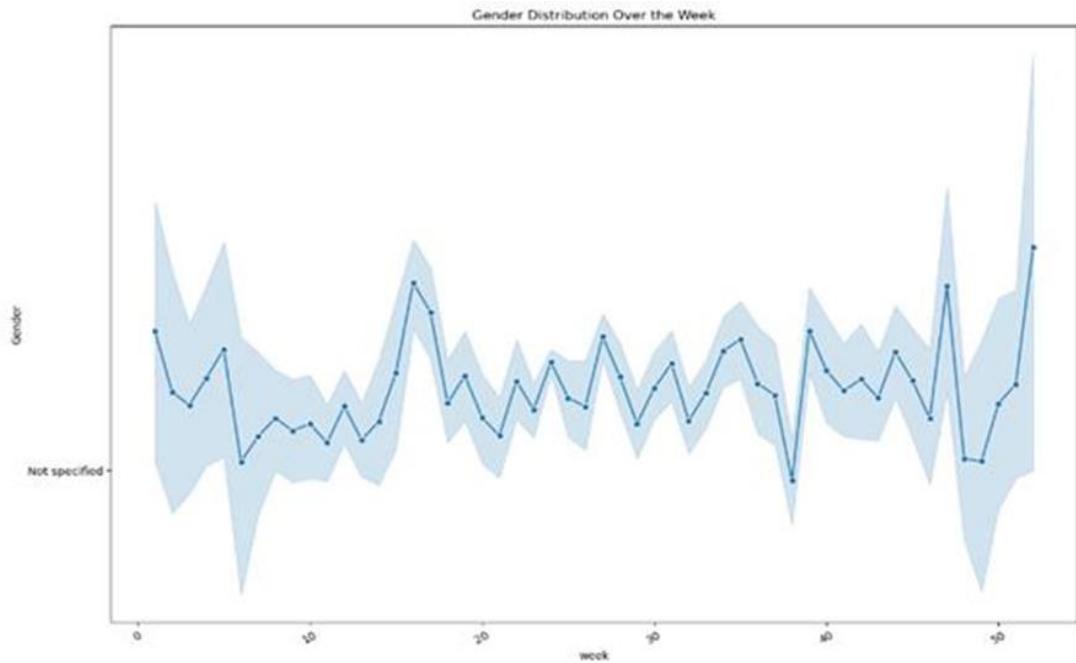
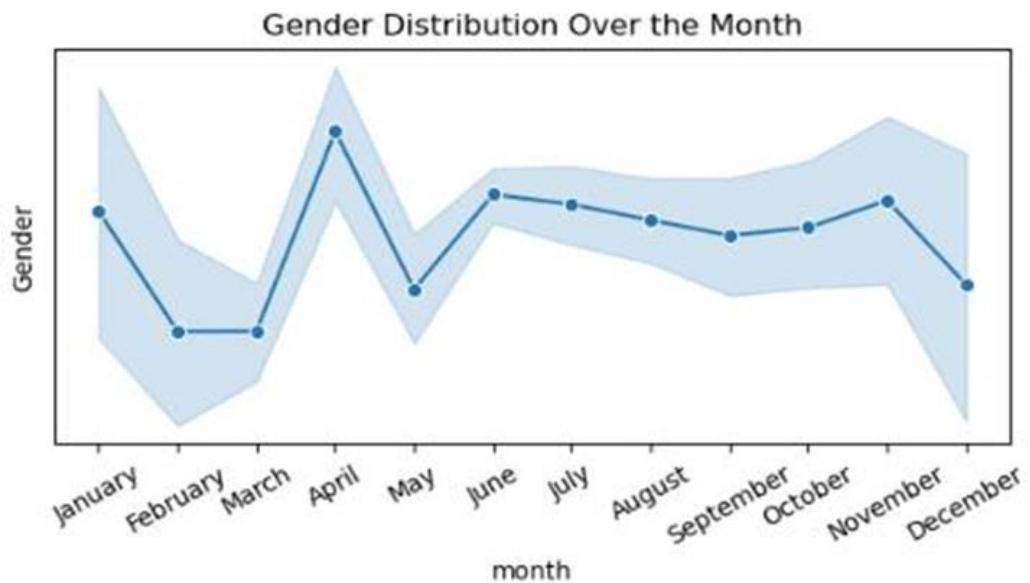


Figure. 6.34(b): Daily Gender Distribution



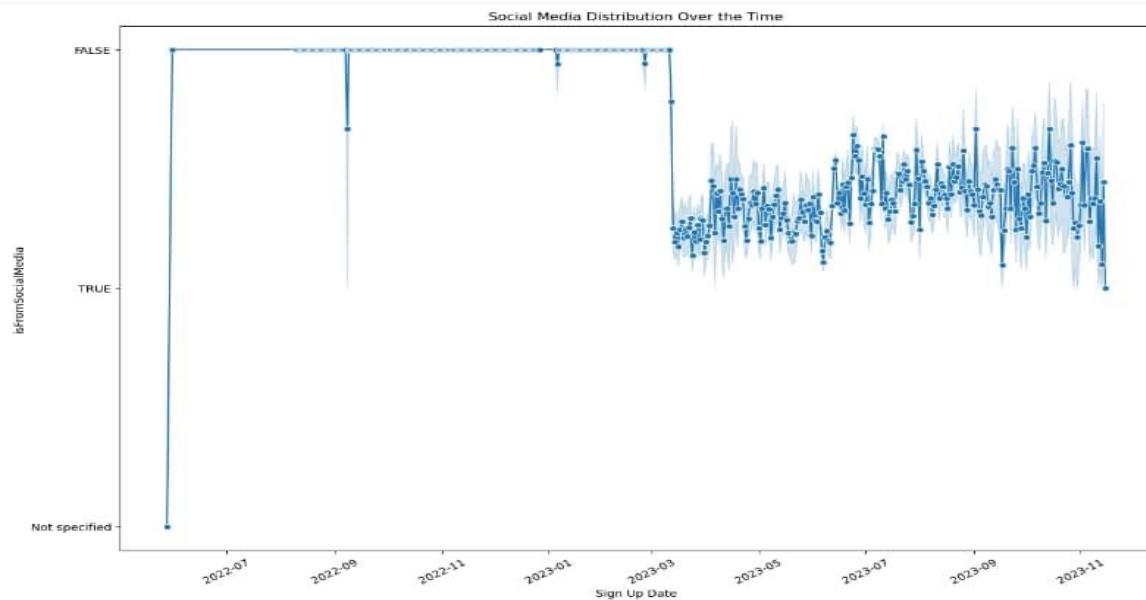
*Figure. 6.34(c): Weekly Gender Distribution*



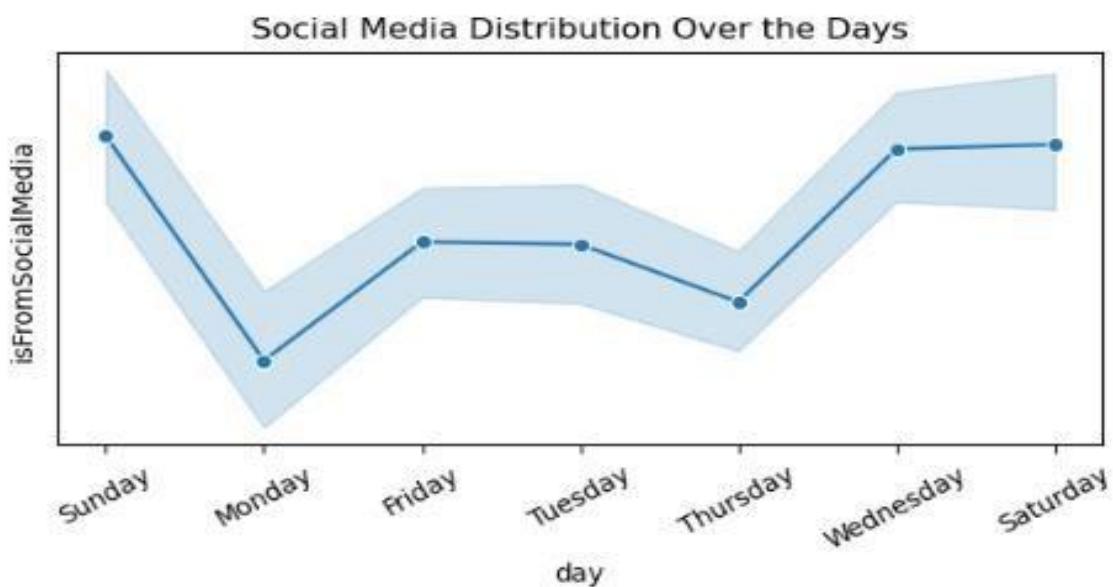
*Figure. 6.34(d): Gender Distribution by Month*

### 6.8.3 Social Media Distribution Over Time

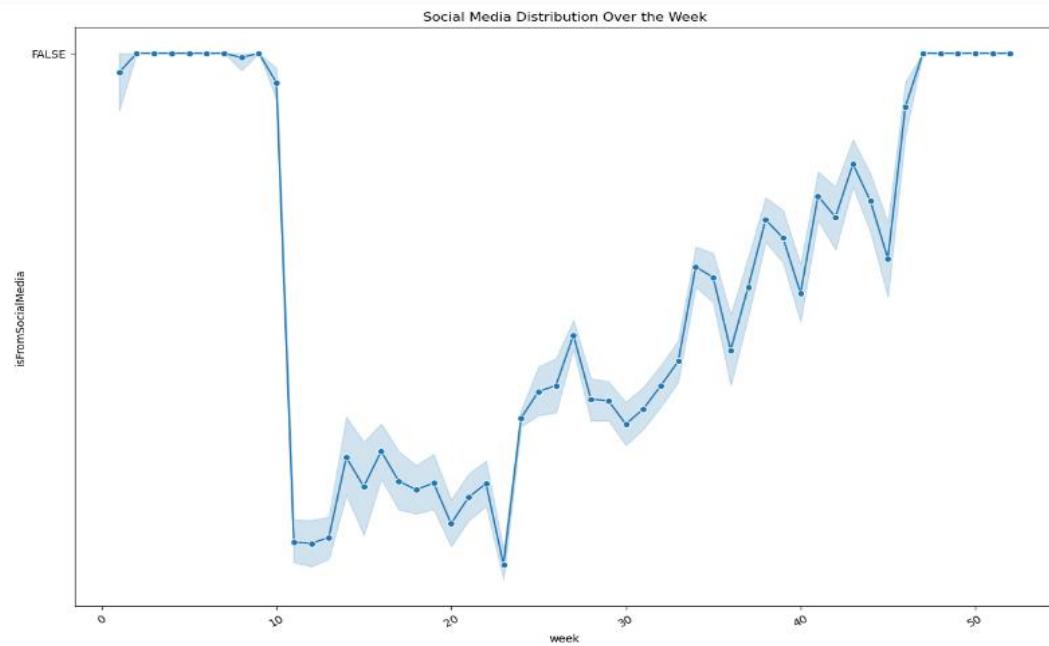
Figures 35(a) to (d) are the same as Figures 33 and 34, represented for social media users.



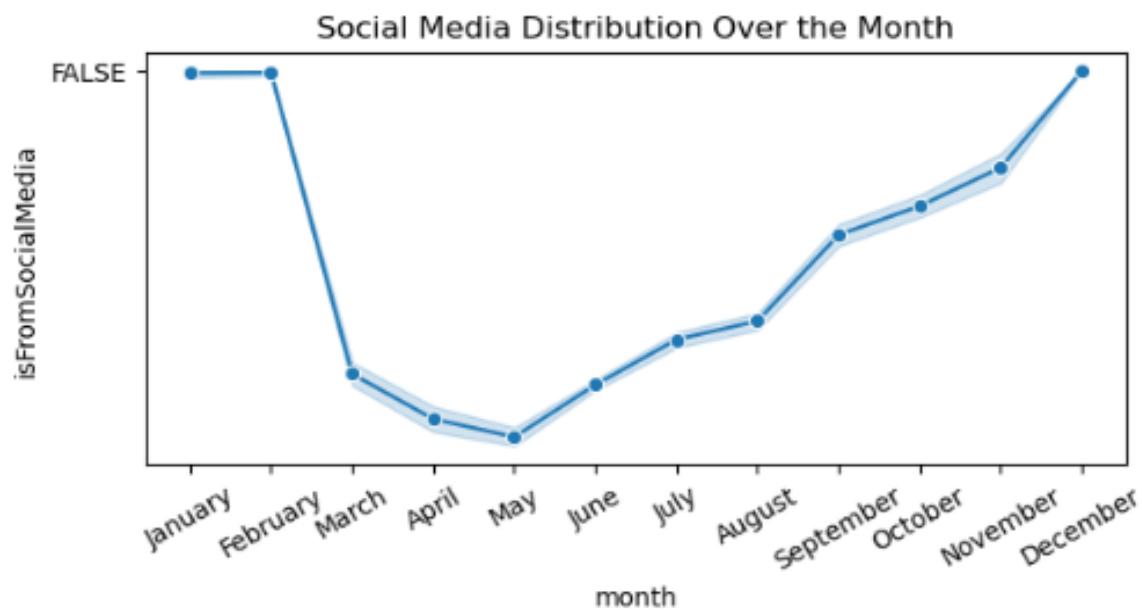
**Figure. 6.35(a): Social Media Distribution Over Time**



**Figure. 6.35(b): Daily Social Media Distribution**



**Figure. 6.35(c): Weekly Social Media Distribution**



**Figure. 6.35(d): Social Media Distribution Over Month**

## 7. Challenges Faced

- **Handling Large Date-set:** The processing time of the large dataset was time consuming and the computational power also affected the analysis of the large data set.
- **Visualization Challenge:** The first plot gave some little difficulty which was later fixed.
- **Data Quality Issues:** Some blanks in many certain rows. They were replaced so as not to affect the data analysis.
- **Categorical Data Handling:** Converting categorical variables into a suitable format for analysis (e.g., one-hot encoding) required careful consideration.
- **Data Transformation Techniques:** Applying appropriate data transformation techniques, such as one-hot encoding or encoding ordinal variables, was complex one.
- **Documentation and Interpretability:** Clearly documenting each step of the preprocessing process and ensuring that the rationale is well-understood was challenging. This documentation is crucial for reproducibility and collaboration.
- **Feature Engineering:** Creating new features or transforming existing ones requires a deep understanding of the data. Choosing relevant features and avoiding overfitting or underfitting was a challenge in feature engineering.

## 8. Next Steps

- Understanding the dashboard and its requirement.
- Build more insights into many relationships from the User Data.
- Look into trends and create more insights.
- Getting started with Google Data Studio
- Creating a Prototype of the dashboard using the Google Data Studio

### Conclusion:

Little trends in degree disclosure habits, geographical preferences, and the efficiency of various marketing channels are all shown by the thorough study. By comprehending the fundamental causes of these patterns, strategic choices about sponsorships, outreach initiatives, and program exposure may be made. The following conclusion was derived from the Trend analysis. Which include:

**Monthly Trends:** June stands out as the peak month for sign-ups, indicating heightened user engagement during the early summer months. July and August follow closely behind in terms of sign-up rates.

**Degree Distribution:** A significant portion of users did not specify their degree, suggesting a diverse user base. Among those who provided information, there is a higher representation of undergraduate students compared to graduate program students.

**Weekly and Daily Trends:** Specific weeks, particularly Week 24, experience higher sign-up rates, and Fridays consistently emerge as the most popular day for user registrations. There is also a notable pattern in daily and weekly distributions of gender, degree, and signup methods.

**Yearly Trends:** The platform experienced a substantial increase in user registrations from 2022 to 2023, with a shift in the distribution of sign-ups. In 2023, there was a significant surge in user engagement, representing 88% of total registrations.

**Gender Distribution:** Males dominate sign-ups, with Fridays showing a higher number of male users. There is a noticeable increase in male representation from 2022 to 2023.

**Signup Methods:** While social media was more popular in 2022, there was a shift in 2023, with more users opting for manual signup methods.

**Preferred Sponsors:** Global Shala consistently holds the position of the first preferred sponsor across days, months, and years. Excelerate stands out as the fifth preferred sponsor, maintaining a strong position consistently.

#### **Recommendations:**

Looking at the number of individuals choosing not to disclose their gender, city, and degree, providing enhanced privacy options during the inquiry process is to be considered.

Educational programs will be channeled to offer and market messages to align with the preferences and aspirations of this demographic.

Looking into the in-depth analysis of peak periods to investigate the factors contributing to the peak in inquiries during 2023, especially in June and July.

Establishing a data security mechanism to communicate the organization's commitment to data privacy and security to build trust among users. This will help to safeguard user information and help users not to worry about their information.

A user feedback mechanism will be helpful to understand user preferences, concerns, and suggestions. This can help to make informed decisions.

The platform witnessed substantial growth in user engagement, especially in 2023. Understanding monthly, weekly, and daily trends, as well as user demographics and preferences, is crucial for tailoring the platform's strategies and offerings to meet the diverse needs of its user base. Additionally, the prominence of specific sponsors suggests the importance of strategic partnerships in driving user adoption and satisfaction.

Based on the trends and patterns observed in the provided data, here are some key recommendations:

**Leverage Peak Months:** Given the peak in sign-ups during June, capitalize on this heightened user engagement by launching targeted marketing campaigns, promotions, or special events during this month to attract and retain users.

**Enhance Degree Information Gathering:** Since a significant portion of users did not specify their degree, consider implementing strategies to encourage users to provide more detailed information about their educational background. This could involve offering incentives, improving data collection processes, or emphasizing the benefits of sharing degree information.

**Optimize Weekly and Daily Strategies:** Recognizing the popularity of Fridays for user registrations, focus on optimizing promotional activities and user engagement initiatives on this day. Additionally, analyze the trends on other peak days, such as Week 24, to tailor specific strategies for those periods.

**Adapt Gender-Specific Marketing:** Acknowledge the gender distribution trends, especially the higher number of male users. Consider tailoring marketing messages and outreach efforts to better resonate with the predominant user demographic.

**Evaluate and Adjust Signup Methods:** With a shift in user preferences from social media to manual signup methods in 2023, assess the reasons behind this change. Conduct user surveys or feedback sessions to understand user preferences and improve the signup experience based on the gathered insights.

**Strengthen Strategic Partnerships:** Given the consistent preference for Global Shala as the first preferred sponsor, strengthen and expand partnerships with this entity. Additionally, explore opportunities to enhance collaborations with Excelerate, the fifth preferred sponsor, to further leverage their positive influence on user engagement.

**Continuous Monitoring and Adaptation:** The yearly and monthly trends indicate a dynamic environment. Implement a system for continuous monitoring of user behaviors and preferences, allowing the platform to adapt quickly to changing trends and evolving user needs.

**Enhance Communication:** Encourage users to provide more information, especially those who did not specify their gender or degree. Clear communication about the benefits of sharing such information can help create a more personalized and engaging platform experience.

## PART 2: - OPPORTUNITY SIGN UP AND COMPLETION DATA

### 3. Introduction

#### 3.1 Brief Background

Excelerate's Opportunity Sign-Up dataset has a rich collection of data on the learner's status. This dataset chronicles the engagement journey between learners and the diverse array of opportunities available within Excelerate's ecosystem. Through the lens of this dataset, the sponsors want to delve into the nuanced interactions, achievements, and skill acquisitions, paving the way for profound insights into learner behavior, opportunity preferences, skill acquisition patterns, and the overall impact of sponsored engagements on learner development and advancement.

To fulfill this goal, as learners we need to undertook a comprehensive task involving Exploratory Data Analysis (EDA), Preprocessing, and Wireframing for two pivotal datasets: User Data and Opportunity Sign-Up and Completion Data. These datasets, containing non-identifying user details, harbor unique stories waiting to be unraveled. The primary focus lay on Opportunity Data, striving to reveal patterns, conduct feature engineering, and offer decision-making support.

The challenge revolves around deciphering non-identifiable user data linked to individuals engaging with targeted opportunities within Excelerate. This encompassed demographics, skill acquisition, and geographic spread, painting a diverse engagement picture. The aim was to dissect and scrutinize data spanning demographics, acquired skills, and rewarded achievements.

The overarching goal of this EDA and preprocessing is to encapsulate vital features from both datasets using statistical methods and visual aids. Unveiling valuable insights aligning with Excelerate's mission to enrich user insights and enhance the overall user experience.

Specific objectives encompassed gathering and analyzing non-identifiable Opportunity data, exploring diverse facets like demographics, skill acquisition, and geographic distribution. A comprehensive review across demographics, skills, and rewards aimed to extract actionable insights. Moreover, identifying trends and correlations within the data served strategic decisions and user engagement optimization within the Excelerate platform, including a comparative analysis between the datasets.

#### 3.2 Problem Statement

The challenge entails conducting Exploratory Data Analysis (EDA), preprocessing and creating a wireframe on Excelerate's Opportunity Sign-Up dataset. These datasets contain non-identifying user details and offer distinct narratives. Focusing primarily on Opportunity Data, the aim is to unveil patterns, engineer features, and aid decision-making. Interpreting user engagement across demographics, skills, and rewards presents the challenge. Extracting insights for informed

decision-making and identifying trends and correlations to optimize user engagement and decision-making aligns with Excelerate's goal of enhancing user experience and insights.

### **3.3 Aim and Objectives**

The primary aim of this Exploratory Data Analysis (EDA) is to summarize key features of the two given datasets through statistical methods and visualizations, extracting valuable insights, and laying the foundation for informed decision-making in subsequent dashboard development stages. Thereby contributing to Excelerate's goal of enhancing user insights and improving the overall user experience.

## **Agenda**

### **Business Understanding**

- Define the problem and objectives of the project
- Determine the sponsors and their requirements
- Identify the available data sources and their limitations

### **Data Understanding**

- Collect the relevant data from various sources
- Perform initial data exploration and visualization to gain insights
- Clean and preprocess the data, handle missing values and outliers

### **Objective:**

1. Develop a visualization dashboard that accurately convey the sponsor on SWOT on Opportunity Dataset
2. Clean, preprocess, and transform the data to make it suitable for the visualization
3. Evaluate various metrics in the preprocessing process to gain insights on rate of conversion of the applicants, and will focus on the marketing gaps, skills and other gaps.
4. Develop a wireframe as a blueprint for dashboard creation
5. Develop documentation on the project evolution and process for the sponsors to explain the outputs and limitations.

### **3.4. Data Overview**

The Opportunity dataset within Excelerate encapsulates vital information crucial to understanding user engagement and interaction with various opportunities offered by the platform. It comprises diverse columns such as 'Profile ID,' 'Opportunity ID,' 'Opportunity Name,' 'Opportunity Category,' 'Opportunity End Date,' 'Gender,' 'City,' 'State,' 'Country,' 'Graduation Date (YYYY MM),' 'Current Student Status,' 'Current/Intended Major,' 'Status Description,' 'Apply Date,' 'Opportunity Start Date,' 'Reward Amount,' 'Badge ID,' 'Badge Name,' 'Skill Points Earned,' and 'Skills Earned.'

This dataset serves as a rich repository documenting user engagement, preferences, and achievements within Excel rate's ecosystem. It chronicles the lifecycle of engagements, from application to completion, showcasing user demographics, geographic spread, educational pursuits, and the rewards garnered through participation. Analyzing this dataset unveils trends, patterns, and correlations critical for optimizing opportunity offerings, refining user experiences, and guiding strategic decisions within the Excelerate platform.

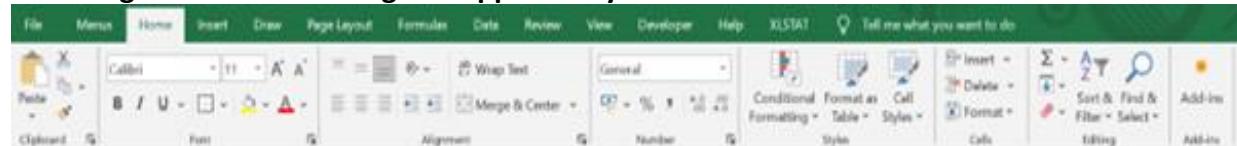
The following are the features of the dataset:

- The Opportunity Sign-Up and Completion Data encompasses 20,258 rows and 37 columns, housing extensive information on learner engagement within Excelerate.
- Each entry in this dataset is identified by a profile ID, enabling tracking of individual learner activities, despite multiple rows potentially sharing the same ID due to learners signing up for multiple opportunities.
- This repetition in rows indicates instances where a learner has actively engaged with multiple opportunities, enriching the dataset with diverse interactions.
- This dataset serves as a pivotal resource, offering insights into the wide array of opportunities learners interact with on the platform.

### Data Cleaning Procedure:

#### Step 1:

##### I. Reading and Understanding the Opportunity Data Set:



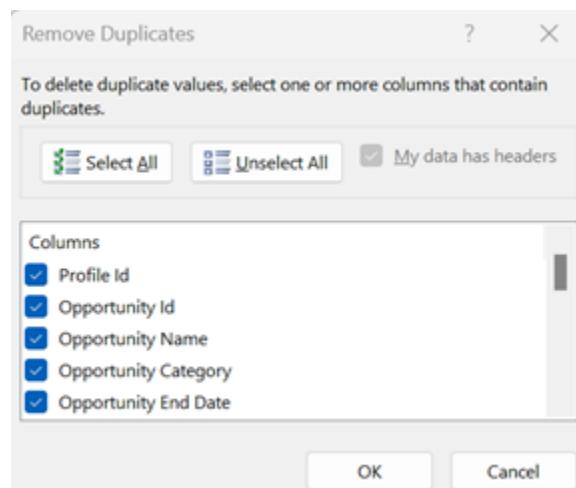
	A	B	C	D	E	F	G
1	Profile Id	Opportunity Id	Opportunity Name	Opportunity Date	Opportunity	State	Country
2	31ce84c2-2bd1-40ba-b2d8-f164fe125306	00000000-064f-19xb-expw-ksbf3n	Statement of Purpose (SOP) Writing Workshop	Event	05-Jan-23	Savar	Bangladesh
3	36814990-f854-4f76-8c63-91f27567d080	00000000-064f-19xb-expw-ksbf3n	Statement of Purpose (SOP) Writing Workshop	Event	05-Jan-23	Punjab	Afghanistan
4	8154328c-ff8e-4bd1-a05-783e340668b5	00000000-064f-19xb-expw-ksbf3n	Statement of Purpose (SOP) Writing Workshop	Event	05-Jan-23	Madhya pradesh	India
5	a3labad6-dble-44c4-a8f4-9e397e252d73	00000000-064f-19xb-expw-ksbf3n	Statement of Purpose (SOP) Writing Workshop	Event	05-Jan-23	Telangana	India
6	c28a815f-2ba3-41e4-a553-7ca68b034a54	00000000-064f-19xb-expw-ksbf3n	Statement of Purpose (SOP) Writing Workshop	Event	05-Jan-23	Karnataka	India
7	061389be-8094-4f73-a7b7-8bf1ff0ed1a8	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Missouri	United States
8	2b39f489-0bb7-4ea2-9af5-de98868cde4c3	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Uttar Pradesh	India
9	2f5a5a25-1c68-4c7b-abef-9c0c8c71e7d8	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Serekunda	Gambia
10	31e6a5e7-e126-4604-b12b-6f6b37ac05e	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Rivers	Nigeria
11	401e1a5f-a0f6-4457-a364-b3fe166988a	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Telangana	India
12	4dedda3e-6c16-4457-bac2-13dc5764fae	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Madhesh Prades	Nepal
13	517b65f5-c413-4791-b926-971b0a5e293b	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Karnataka	India
14	57a20005-0803-4858-8a22-8191976794bd	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Andhra Pradesh	India
15	5ff5dd3e-dc75-477d-91c2-f0988fe59e2c	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Telangana	India
16	7b46dce2-05bc-4507-9fd9-93a87fd0d275	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	ANDHRA PRADESH	India
17	a56155ab-4608-4883-b5ca-h1211ra5fc2	00000000-0678-hcvb-01ae-6qeipby	Life Beyond Saint Louis University's Campus	Event	27-Oct-22	Andhra Pradesh	India

The dataset contains 20,322 rows and 21 columns, denoted as 20322 \* 21 in a row\*column format.

## II. Finding Duplicates:

Select the table > Data > Remove duplicates

1	Status Description	Apply Date	Opportunity Start Date	Reward	Badge I	Badge T	Skill Pol	Skills Ex
2	Not Started	Oct 03, 2022, 17:29:17	Jan 05, 2023, 18:52:39					
3	Rewards Award	Jan 05, 2023, 18:58:37	Jan 05, 2023, 18:52:39	200 00000000-Statement	10	["Critical Thinking", "Creative Thinking", "Communication", "Career Readiness"]		
4	Not Started	Oct 04, 2022, 08:10:07	Jan 05, 2023, 18:52:39					
5	Not Started	Oct 05, 2022, 13:41:47	Jan 05, 2023, 18:52:39					
6	Not Started	Oct 05, 2022, 11:07:17	Jan 05, 2023, 18:52:39					
7	Rewards Award	Oct 27, 2022, 15:05:04	Oct 28, 2022, 13:30:00	200 00000000-Life Beyond	10	["Creative Thinking", "Technology Literacy", "Initiative", "Career Readiness"]		
8	Not Started	Oct 26, 2022, 05:20:31	Oct 28, 2022, 13:30:00					
9	Team Allocated	Oct 16, 2022, 03:29:18	Oct 28, 2022, 13:30:00					
10	Rewards Award	Oct 27, 2022, 06:34:05	Oct 28, 2022, 13:30:00	200 00000000-Life Beyond	10	["Creative Thinking", "Technology Literacy", "Initiative", "Career Readiness"]		
11	Rewards Award	Oct 26, 2022, 17:51:37	Oct 28, 2022, 13:30:00	200 00000000-Life Beyond	10	["Creative Thinking", "Technology Literacy", "Initiative", "Career Readiness"]		



A	B	C	D	E	F	G	H	I	J	K	L
1	Profile	Opport.	Opport.	Opportunity C	Opportunity End Date	Gender	City	State	Country	Zip Cod	Graduate
2	31ce84c2-2bd1-40ba-b2d8-f164fe125306			00000000-Statement Event	Jan 05, 2023, 18:58:39	Male	Dhaka	Savar	Bangladesh	1342	2023-06
3	36814990	00000000-Statement Event			Jan 05, 2023, 18:58:39	Female	Amritsar	Punjab	Afghanistan	1233242	2021-01
4	8154328c	00000000-Statement Event			Jan 05, 2023, 18:58:39	Female	Satna	Madhya pradesh	India	485001	2023-04
5	a81ab4d6	00000000-Statement Event			Jan 05, 2023, 18:58:39	Male	H	Microsoft Excel	India	500039	2024-12
6	c2b8a15f	00000000-Statement Event			Jan 05, 2023, 18:58:39	Male	B		India	560105	2022-06
7	0613899e	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	Si		United States	63108	2024-12
8	2b39f489	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	A		India	282001	2025-01
9	2f5a5a25	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	B		Gambia	No zip	2019-06
10	31e6a5e7	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Female	P		Nigeria	500102	2024-12
11	461e1a5f	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	Hyoerassar	rajangana	India	500074	2024-12
12	4redddale	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	Janakpurd	Madhes Prades	Nepal	456000	2022-06
13	517b0b5f	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	Bangalore	Karnataka	India	560070	2025-01
14	57a20005	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	Tirupati	Andhra Pradesh	India	517901	2022-08
15	5ff5dd3e	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Female	Hyderabad	Telangana	India	500668	2022-05
16	7b4fdde2	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	GUNTUR	ANDHRA PRADESH	India	522426	2020-07
17	a56155ab	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Female	Madanap	Andhra Pradesh	India	517325	2022-08
18	a81ab4d6	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Male	Hyderabad	Telangana	India	500039	2024-12
19	a81e730a	00000000-Life Beyond Event			Oct 27, 2022, 18:29:00	Female	H	Hyderabad,Telangana	India	500072	2024-12

The dataset indicates that learners applied for multiple opportunities, yet there are no duplicate entries present.

### III.Column variables:

Profile Id
Opportunity Id
Opportunity Name
Opportunity Category
Opportunity
State
Country
Zip Code
Graduation Date(YYYY MM)
Current Student Status
Current/Intended Major
Status Description
Apply Date
Opportunity Start Date
Reward Amount
Badge Id
Badge Name
Skill Points Earned
Skills Earned

Some column variables namely "Profile ID", "Opportunity ID," "Opportunity Name", "Opportunity Category", and, " County" do not necessitate any alterations as there were no missing values present and the data was intact.

### Step 2:

#### Data Elimination

##### I.Dropping Unwanted Variables:

A	B	C	D	E
Profile Id	Opportunity Id	Opportunity Name	Opportunity Category	Opportunity
2: 31ce94c2-2bd1-409a-9c2-6f1e2233080	00000000-0049-1398-EXPW-K38F3N	Statement of Purpose (SOP) Writing Workshop	Event	
3: b6814090-0554-4f70-9c63-91217567d080	00000000-0049-1398-EXPW-K38F3N	Statement of Purpose (SOP) Writing Workshop	Event	
4: 815432bc-0ff4-4b31-a05-783e14095685	00000000-0049-1398-EXPW-K38F3N	Statement of Purpose (SOP) Writing Workshop	Event	
5: a834a9d0-0f1c-4444-aef4-8e397e20d773	00000000-0049-1398-EXPW-K38F3N	Statement of Purpose (SOP) Writing Workshop	Event	
6: c00115-0000-4343-9444-000000000000	00000000-0049-1398-EXPW-K38F3N	Statement of Purpose (SOP) Writing Workshop	Event	
7: 213907-0004-4713-a71c-000000000000	00000000-0049-HCVB-03AE-0QEPBY	Life Beyond Saint Louis University's Campus	Event	
8: 2b39f480-0004-4ea2-9a5f-dc988bdc4ec3	00000000-0049-HCVB-03AE-0QEPBY	Life Beyond Saint Louis University's Campus	Event	
9: 219ca25-3c8d-4c7b-abeb-9c0d6f1e-7f8f	00000000-0049-HCVB-03AE-0QEPBY	Life Beyond Saint Louis University's Campus	Event	
10: 31e4a5e7-ed26-4404-812b-6bfbb37ae05e	00000000-0049-HCVB-03AE-0QEPBY	Life Beyond Saint Louis University's Campus	Event	

By selecting all the variables, activate the filter option.

Profile Id	Opportunity Id	Opportunity Name	Opportunity Category	Opport
31ce84c2-2bd1-409a-b2d8-f164fe125306	00000000-004f-10xb-expw-k5bf3n	Statement of Purpose (SOP) Writing Workshop	Event	

In all variable columns, there's a drop-down button visible after applying filter. Upon validation, it became apparent that certain variables, such as Opportunity ID and Badge ID, didn't offer additional insights, so they were removed. Additionally, "Zipcode" was eliminated due to inconsistencies and inaccuracies in its data.

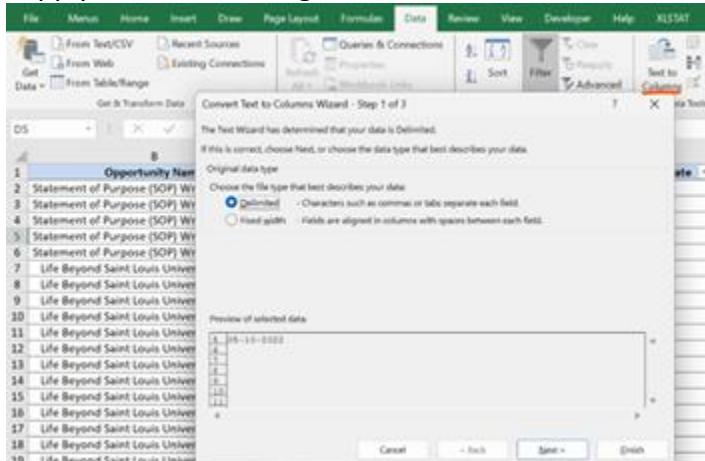
### STEP 3:

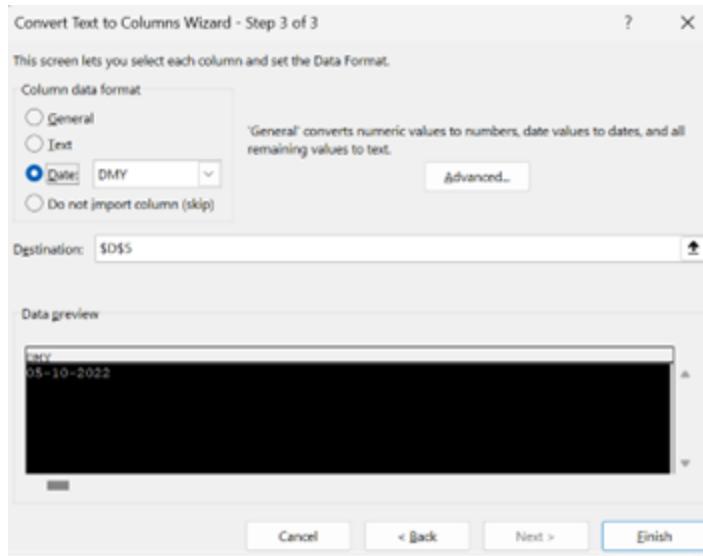
#### Data Validation:

##### I.Re-Arranging the Columns:

The dates columns were rearranged for better readability and validating with the "Apply Date" by calculating the distance between the "Opportunity Start Date" and "Opportunity End Date". The negative distances were removed and the null values in the "Opportunity Start Date" were carefully validated and mapped with "Apply Date + 1". Here, we assumed by approximating the distances.

From the "Apply Date" column, time was extracted and made into a separate column called "Apply time" and changed the format from "Text" to "Date" format.





Apply Date	Apply Time
03-10-2022	17:29:17
05-01-2023	18:58:37
04-10-2022	08:10:07
05-10-2022	13:41:47
05-10-2022	11:07:17
27-10-2022	15:05:04
26-10-2022	05:20:31

The time details were trimmed using the formula from the “Opportunity Start Date” and “Opportunity End Date” variables and unwanted symbols also it has been changed from text to date format.

#### Step 4:

##### I.Outlier Deduction:

The column for "Graduation Date" exhibited outliers. Assuming that current learners also submitted their applications assuming a duration of 5 or more years from the present year and, >2028 years were found 0.25% and hence deleted from the column.

##### II.Splitting the column:

There were 14 unique skills present in the “Skills Earned” column.

Critical Thinking	Communication	Media Literacy	Collaboration	Social Skills
Creative Thinking	Career Readiness	Leadership	Flexibility	College Readiness
Technology Literacy	Initiative	Information Literacy	Productivity	–

To aid visualization, distinct columns were generated for individual skills, with a mapping of 1s denoting the presence of a specific skill set and 0s indicating its absence.

**Data Analyzation:**

**Step 4:**

**I.Opportunity Category**

Opportunity Category	
Competition	Event
Course	Learnership
Engagement	–

A total of five opportunity categories were available. Subsequently, the analysis aimed to identify the top category.

**II.Opportunity End Date**

Opportunity End Date
Oct 2022 – Dec 2025

The dataset comprised learners who applied for opportunities ending between the years 2022 and 2025.

**III.Gender**

Gender	Action
Don't Want to Specify	No change
Female	No change
Male	No change
Other	No change
Blanks	Remove Blanks

The blank columns were removed.

**IV.City**

City	Action
A	Delete
aa	Delete
K	Delete
City	Delete
K	Delete

Cities with a character count of  $\leq 2$  were removed due to limited potential for extracting insights. A total of 22 rows, accounting for 0.11% of the dataset, were eliminated after restricting the field length to  $\geq 2$  characters, which was deemed negligible.

## V.Country

<b>State</b>	<b>Replace</b>
A	Delete
P	Delete
T	Delete
Blanks	Delete

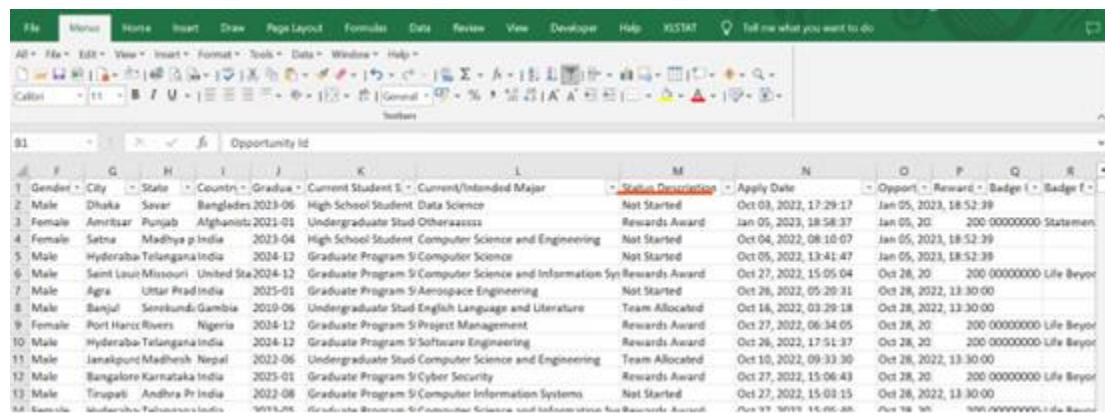
A total of 25 rows, representing 0.12% of the dataset, were removed due to having a field length restricted to  $\geq 1$  character, which was considered negligible.

## **VI. Current Student Status**

Current Student Status	
Graduate Program student	No change
High School Student	No change
Not in education	No change
Undergraduate Student	No change
Blanks	

## **VII.Current/Intended Major**

Here, 0.68% of irrelevant entry was deleted. The other Major with similar specialization was bucketed for easier analysis



The screenshot shows an Excel spreadsheet with data in columns G through O. A filter menu is open over column O, titled 'Status Description'. The menu includes options like 'Sort A to Z', 'Sort Z to A', 'Filter by Color', and 'Text Filters'. Under 'Text Filters', there is a search bar and a list of categories with checkboxes. Several checkboxes are checked, including '(Select All)', 'Accounting', 'Actuarial Science', 'Adult Education', 'Adult Nursing', 'Advertising', 'Aeronautics', and 'Aerospace Engineering'. The 'Rewards Award' checkbox is also checked. At the bottom of the filter menu are 'OK' and 'Cancel' buttons.

Opportunity End Date	Gender	City	State	Country	Graduation Date	Current Student Status	Current/Intended Major	Status Description
05-02-2023	Male	Dhaka	Savar	Bangladesh	2023-06	High School Stu		Not Started
04-02-2023	Female	Amitsar	Punjab	Afghanistan	2021-01	Undergraduate St		Rewards Award
04-02-2023	Female	Satna	Madhya Pradesh	India	2023-04	High School Stu		Not Started
04-02-2023	Male	Hyderabad	Telangana	India	2024-12	Graduate Program		Not Started
04-02-2023	Male	Bangalore	Karnataka	India	2022-06	Undergraduate St		Not Started
27-11-2022	Male	Saint Louis	Missouri	United States	2024-12	Graduate Program		Rewards Award
27-11-2022	Male	Agra	Uttar Pradesh	India	2025-01	Graduate Program		Not Started
27-11-2022	Male	Banjul	Serekunda	Gambia	2019-06	Undergraduate St		Team Allocated
27-11-2022	Female	Port Harcourt	Rivers	Nigeria	2024-12	Graduate Program		Rewards Award
27-11-2022	Male	Hyderabad	Telangana	India	2024-12	Graduate Program		Rewards Award
27-11-2022	Male	Janakpurdham	Madhesh Pradesh	Nepal	2022-06	Undergraduate St		Team Allocated
27-11-2022	Male	Bangalore	Karnataka	India	2025-01	Graduate Program		Rewards Award
27-11-2022	Male	Tirupati	Andhra Pradesh	India	2022-08	Graduate Program		Not Started
27-11-2022	Female	Hyderabad	Telangana	India	2022-05	Graduate Program		Rewards Award
27-11-2022	Male	Guntur	Andhra Pradesh	India	2020-07	Graduate Program		Rewards Award
27-11-2022	Female	Madanapalli	Andhra Pradesh	India	2022-08	Graduate Program		Not Started
27-11-2022	Male	Hyderabad	Telangana	India	2024-12	Graduate Program		Not Started
27-11-2022	Female	Hyderabad	Telangana	India	2024-12	Graduate Program		Rewards Award
27-11-2022	Female	Multan	Punjab	Pakistan	2022-08	Not in Education		Not Started
27-11-2022	Male	Krishna	Andhra Pradesh	India	2024-12	Graduate Program		Rewards Award
27-11-2022	Female	Hyderabad	Telangana	India	2021-06	Graduate Program		Rewards Award

### Data Cleaning Conclusion:

The dataset underwent a comprehensive cleaning process, which involved the removal of duplicates, rearrangement of entries, and elimination of null values and bucketing using Excel. Additionally, it underwent cross-validation to ensure its reliability and readiness for the exploratory data analysis (EDA) phase.

## 4. EDA -PART 2

### Column Analysis:

```
opportunity_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20258 entries, 0 to 20257
Data columns (total 45 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Profile Id      20258 non-null   object  
 1   Opportunity Name 20258 non-null   object  
 2   Opportunity Category 20258 non-null   object  
 3   Apply Date      20258 non-null   datetime64[ns]
 4   Apply Time       19795 non-null   datetime64[ns]
 5   Opportunity Start Date 20258 non-null   datetime64[ns]
 6   Opportunity End Date 20258 non-null   datetime64[ns]
 7   Gender          20258 non-null   object  
 8   City            20258 non-null   object  
 9   State           20258 non-null   object  
 10  Country         20258 non-null   object  
 11  Graduation Date 20258 non-null   datetime64[ns]
 12  Current Student Status 20258 non-null   object  
 13  Current/Intended Major 20258 non-null   object  
 14  Status Description 20258 non-null   object 
```

```

15 Reward Amount      20258 non-null int64
16 Badge Name        20258 non-null object
17 Skill Points Earned 20258 non-null int64
18 Skills Earned     2003 non-null object
19 Skill1            20258 non-null object
20 Skill2            20258 non-null object
21 Skill3            20258 non-null object
22 Skill4            20258 non-null object
23 Skill5            20258 non-null object
24 Skill6            20258 non-null object
25 Creative Thinking 20258 non-null int64
26 Technology Literacy 20258 non-null int64
27 Communication     20258 non-null int64
28 Career Readiness  20258 non-null int64
29 Initiative         20258 non-null int64
30 Media Literacy    20258 non-null int64
31 Leadership          20258 non-null int64
32 Information Literacy 20258 non-null int64
33 Collaboration      20258 non-null int64
34 Flexibility         20258 non-null int64
35 Productivity        20258 non-null int64
36 Social Skills       20258 non-null int64

37 College Readiness  20258 non-null int64
38 Apply Year          20258 non-null int64
39 Apply Month         20258 non-null int64
40 Opportunity Start Year 20258 non-null int64
41 Opportunity Start Month 20258 non-null int64
42 Opportunity End Year 20258 non-null int64
43 Opportunity End Month 20258 non-null int64
44 Apply Hour          19795 non-null float64
dtypes: datetime64[ns](5), float64(1), int64(21), object(18)
memory usage: 7.0+ MB

```

## 5.Summarizing the unique values and their frequencies for categorical columns:

```
#Frequency tables for all the categorical variables
cat_ods.describe()
```

	Profile Id	Opportunity Name	Opportunity Category	Gender	City	State	Country	Current Student Status	Current/Intended Major	Status Description	Badge Name	Skills Earned	Skill1	Skill2
count	20258	20258	20258	20258	20258	20258	20258	20258	20258	20258	20258	2003	20258	20258
unique	11447	33	5	4	2649	1080	108	4	585	8	57	3	6	8
top	c2245f7e-2e9d-42c9-b5af-550be9eae1c8	Data Visualization	Internship	Male	Saint Louis	Missouri	India	Graduate Program Student	Computer Science	Team Allocated	NA	,	NA	NA
freq	22	5676	15312	12203	2138	2615	9113	9289	4053	14161	17746	1990	17746	17746

## 6.Profile ID Analysis:

```
##Examine the uniqueness of Profile IDs in both datasets.  
# Total Applications  
total_application = opportunity_data["Profile Id"].count()  
total_application
```

20258

```
# Unique Profile IDs  
opportunity_data["Profile Id"].nunique()
```

11447

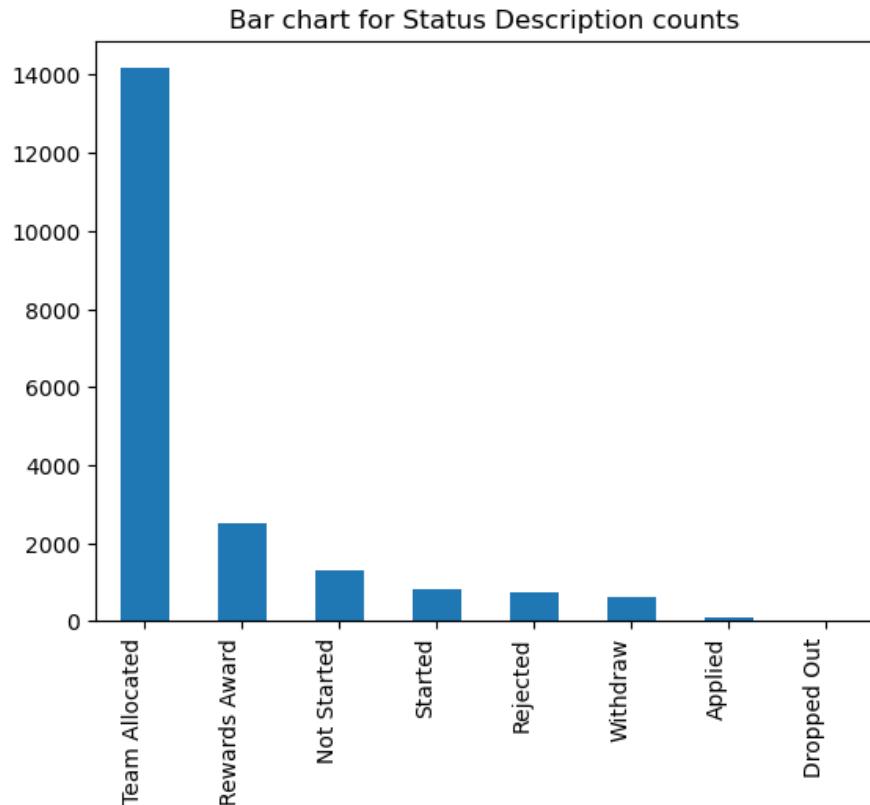
```
#Checking for duplicated profile ID  
opportunity_data['Profile Id'].duplicated().sum()
```

8811

A total of 8,811 profile IDs exhibit duplications, indicating that a corresponding number of learners have engaged with multiple opportunities on the Excelerate website.

## 7.Opportunity Status Distribution:

Among the learners, the predominant status is "teams allocated," indicating that a significant majority have successfully had their teams assigned. The count for statuses such as "rewards awarded" exceeds 2000, while statuses like "not started" and "started" follow suit. The occurrence of the "dropped out" status is minimal. This is shown in the Figure below.



**Figure7.1 Status description count**

## 8.Basic Statistics:

Calculating basic statistics (mean, median, min, max) for relevant numeric columns (e.g., Reward Amount, Skill Points Earned).

```
# Finding the data spread  
opportunity_data.describe()
```

	Reward Amount	Skill Points Earned	Creative Thinking	Technology Literacy	Communication	Career Readiness	Initiative	Media Literacy	Leadership	Information Literacy
count	20258.000000	20258.000000	20258.000000	20258.000000	20258.000000	20258.000000	20258.000000	20258.000000	20258.000000	20258.000000
mean	133.641524	147.182940	0.107957	0.107365	0.115312	0.001432	0.017721	0.001234	0.015549	0.074489
std	482.066680	415.734528	0.310334	0.309584	0.319407	0.037809	0.131940	0.035109	0.123727	0.262572
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	2500.000000	1776.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

**Figure 8.1 Statistics view of the data spread**

## 9.Initial Observations:

### Initial observations/patterns noticed during the exploratory analysis:

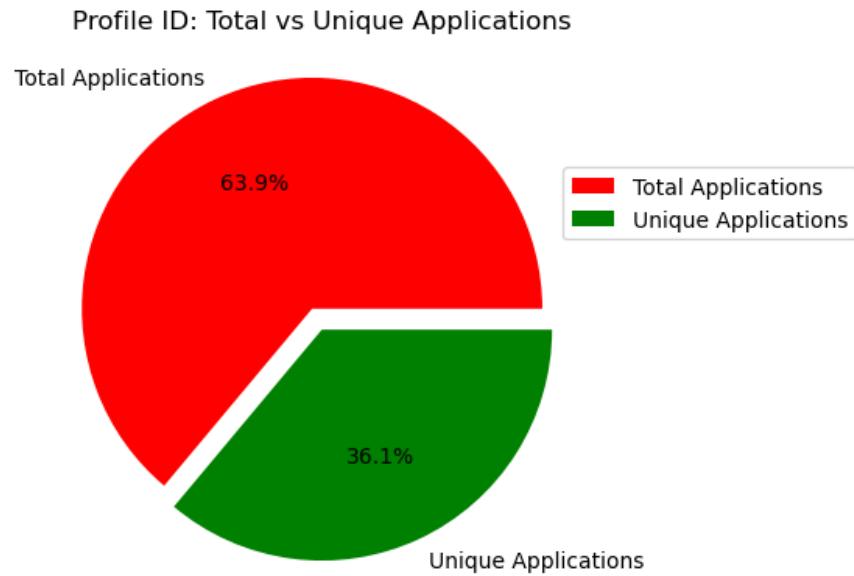
- Only a small percentage of learners, approximately 15%, succeed in acquiring the skills or skill points out of the vast pool of over 20,000 engaged learners.
- Most learners are engaged for learnerships than any other opportunity.
- In the realm of opportunities, the count of learners specializing in Data Visualization takes the lead.

Potential areas of interest for deeper investigation in the upcoming weeks is basically on the skills earned feature as regards this dataset.

## 10.Visualizations

### 10.1 Total Applicant Vs Unique Applicants Analysis

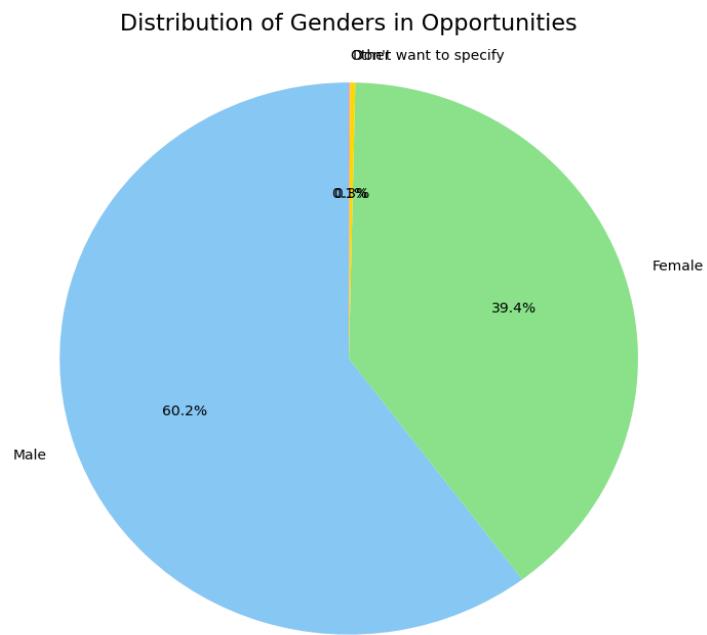
This analysis is required to fetch the ratio of the learner's who register for dual/multiple opportunities available in the Exelerate's portal.



**Figure10.1:** Pie chart to depict the Total applicants Vs Unique Applicants

The graph shows that 63.9% of the total applications contribute, while 36.1% represent unique applications.

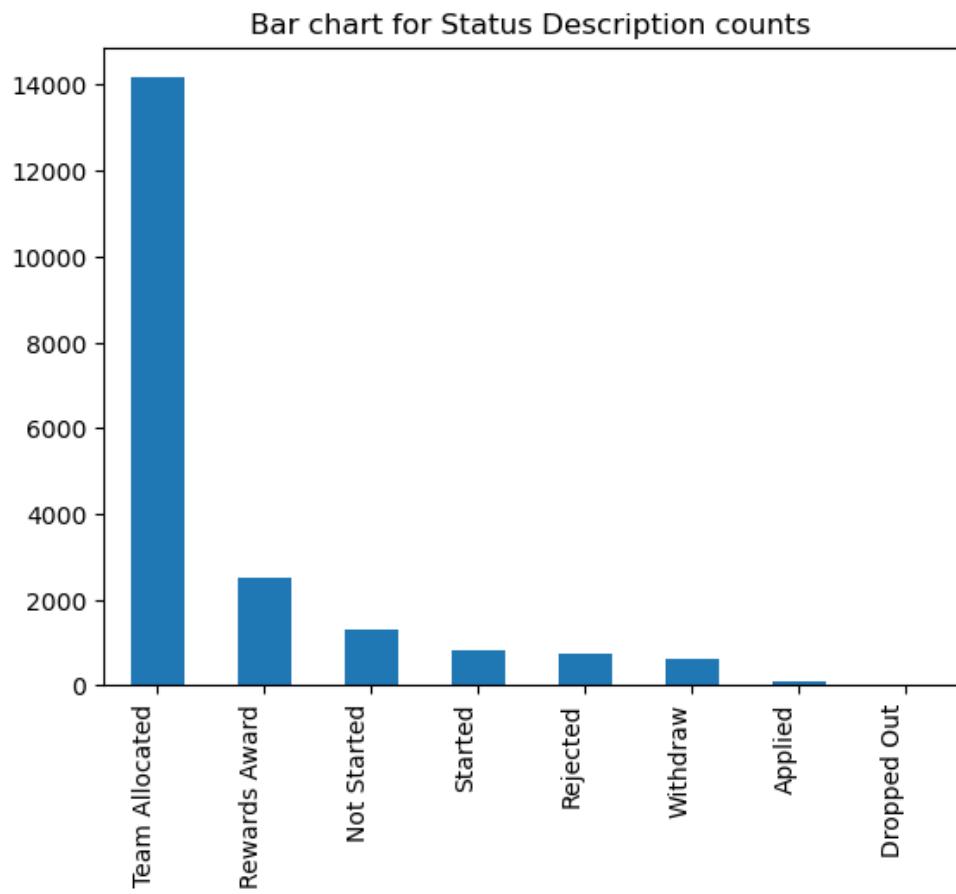
## 10.2 Gender Distribution in Opportunities



**Figure10.2:** Pie Chart for Gender

Among the learners who engaged on the website, 60.2% are males, 39.4% are females, and the remaining 0.4% comprises individuals categorized as others or those who choose not to specify their gender as shown in pie chart above.

### 10.3 Opportunity Status Description

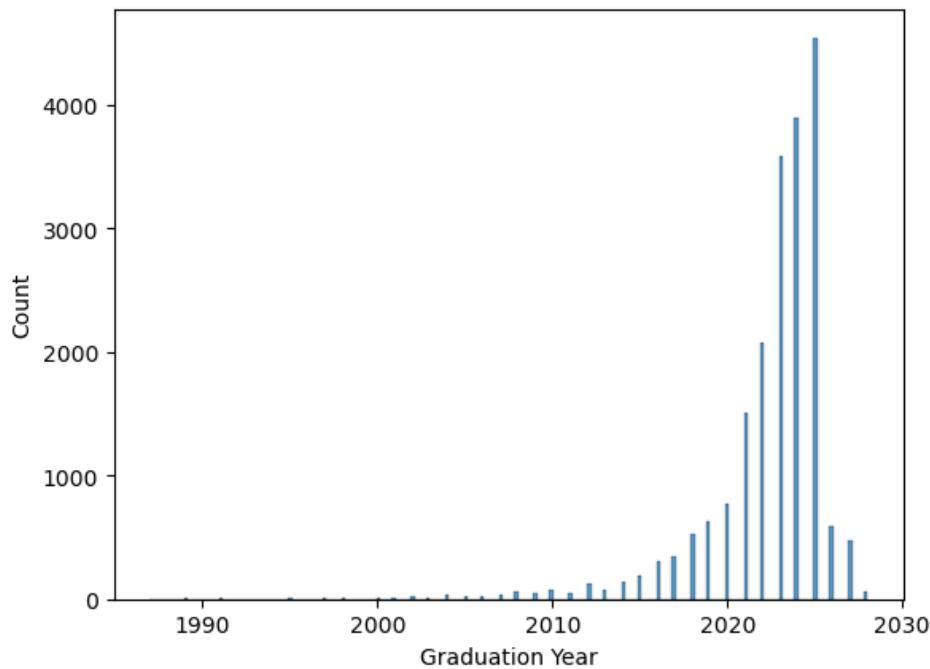


**Figure10.3 : Bar Chart for Status description counts**

In the status of learners, the majority of those who actively participated had their teams allocated.

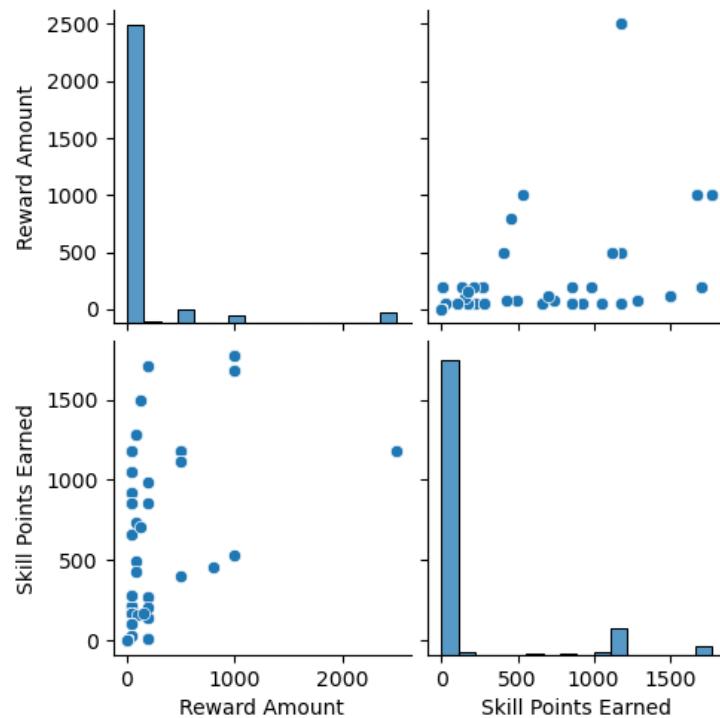
### 10.4 Learners Graduation Year

The graph reveals a notable trend: applicants from 2020 and those set to graduate in 2026 show a significant number of learners, with 2024 recording the highest count at 4800.



**Figure10.4: Learners Graduation Year Spread**

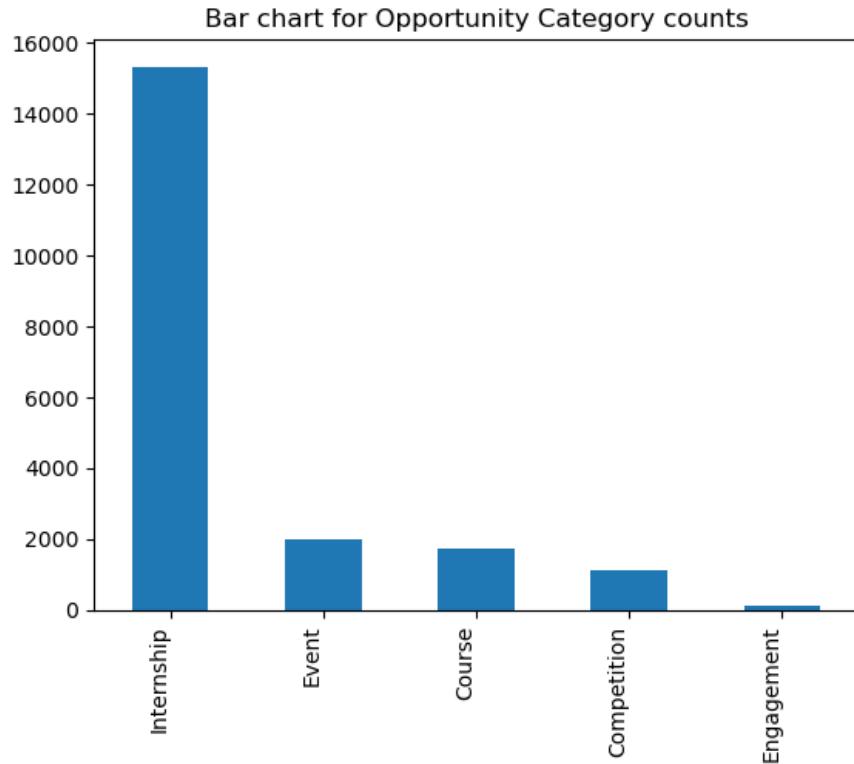
### 10.5 Learners Reward Amount and Skill Points



**Figure10.5: Histogram and Scatter plot for Reward Amount & Skill Points**

This indicates that a small number of learners, less than 1000, received Reward Amounts of \$2500, \$1000, and \$500. This is also a critical indicator that reveals that learners are unable to fetch the scholarship easily due to various circumstances. Learners need to be more serious while taking learnership because it is a critical booster for their career. Only a minimal number of learners are truly benefiting from the learnership, indicating a need for learners to put in more effort to acquire skill points.

#### **10.6 Bar Charts for knowing the distribution of categorical data:**



**Figure10.6 : Bar Chart for Opportunity category counts**

The graph shows that most learners are engaged for internships compared to Events, Course, Competitions and Engagement with the minimum registrations.

#### **10.7 Opportunity Name Counts**

The graph depicts that in the realm of opportunities, the count of learners specializing in Data Visualization takes the lead.

Bar chart for Opportunity Name counts

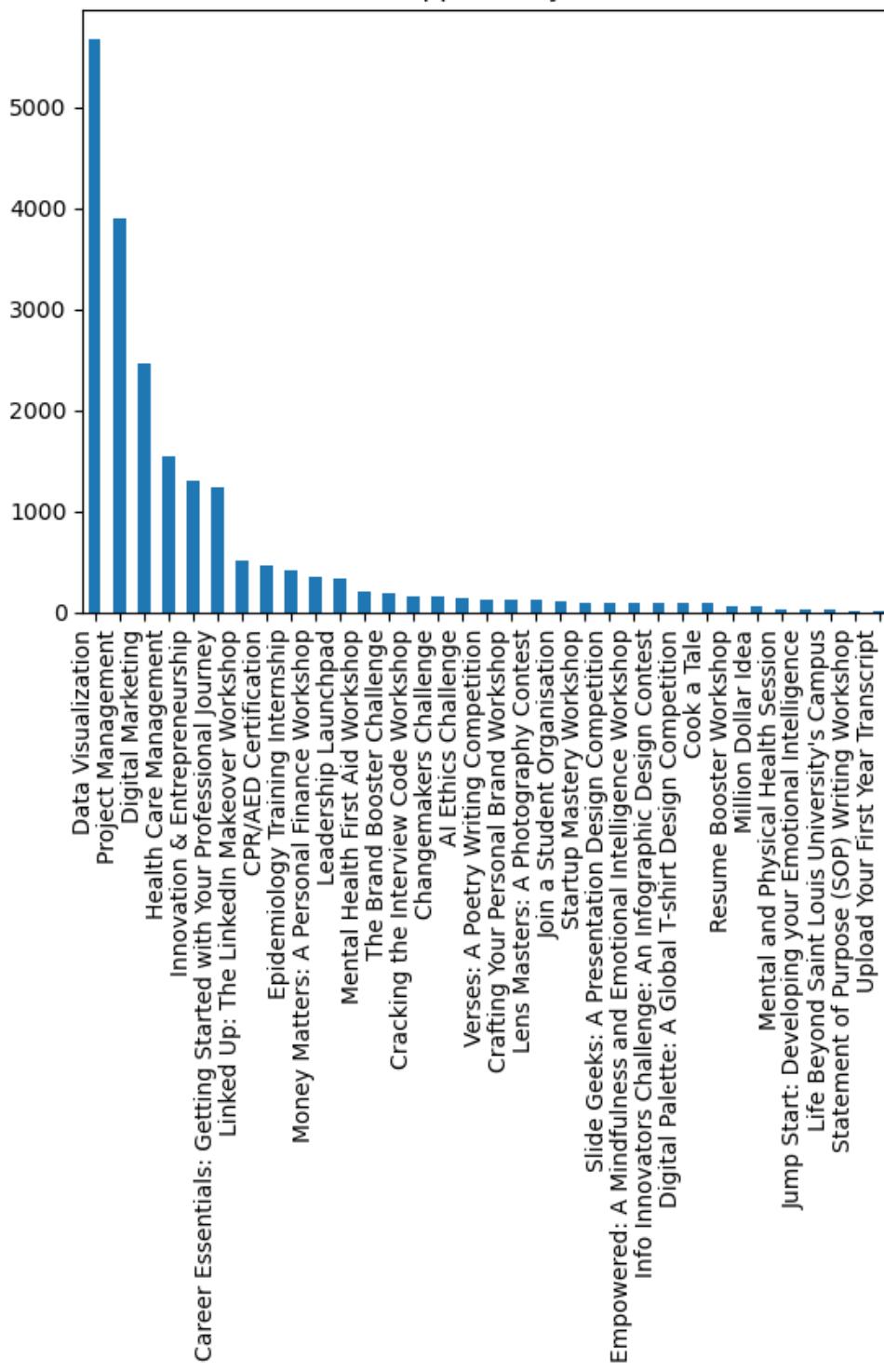


Figure10.7: Bar Chart for Opportunity courses

## 10.8 Monitoring Apply Months

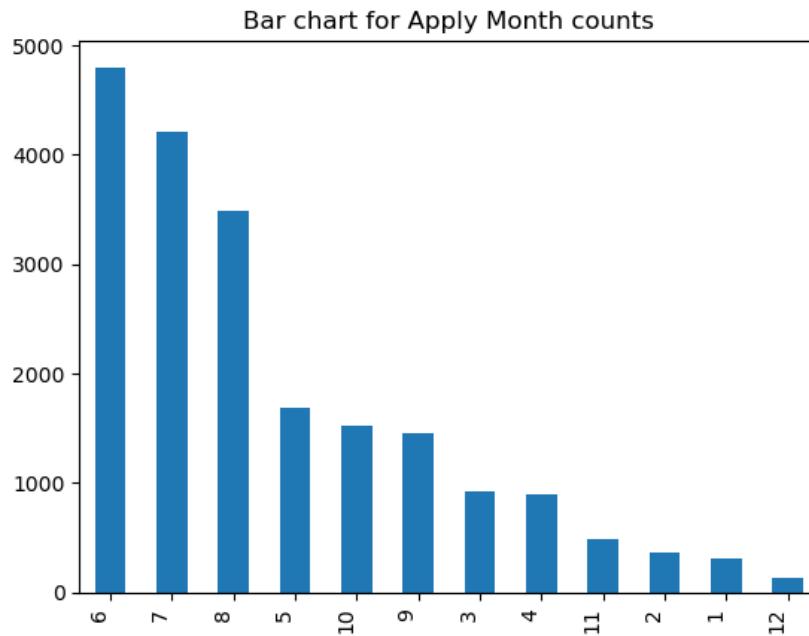


Figure10.8: Bar Chart for Apply Month counts

The highest enrollment numbers typically occur in June, July, and August, making these three months the most popular among learners. This is also evident in the User Data visualization.

## 10.9 Monitoring Apply Year

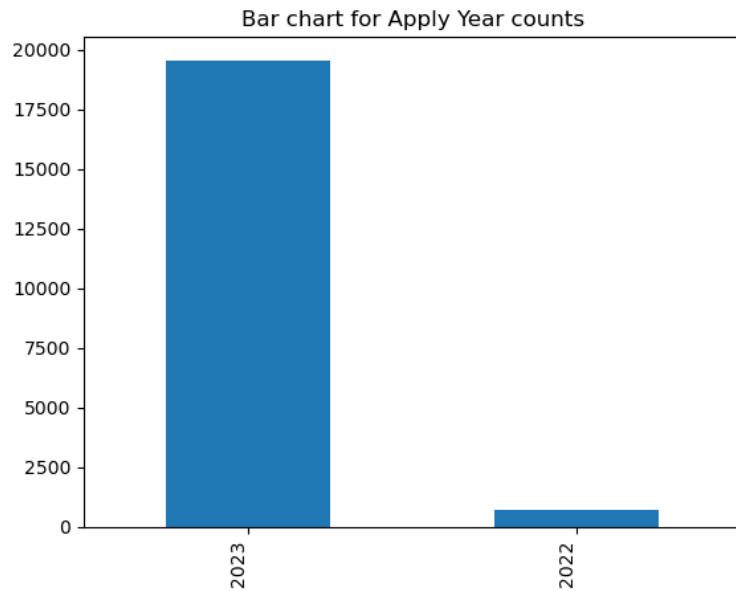


Figure10.9 Bar Chart for Apply year counts

Majority of the learners applied in the year 2023, which conforms with the result of the User Data analysis.

### 10.10 Top 10 countries with most learners

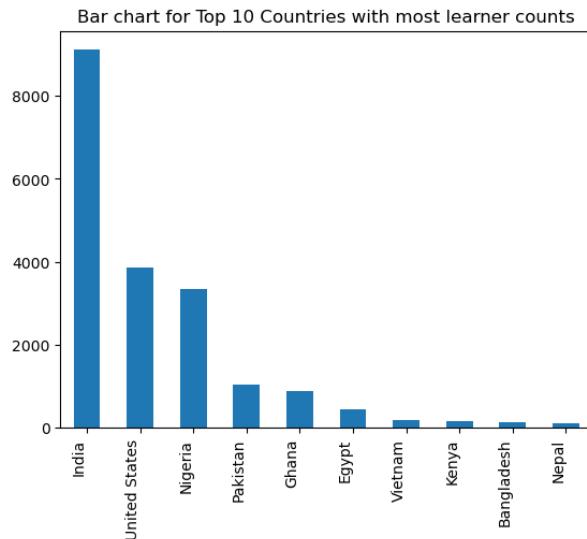


Figure10.10: Bar Chart for countries with most learners

The highest number of engaged learners is found in India, with the United States, Nigeria, Pakistan, and other countries following suit.

### 10.11 Bottom 50 Countries with the least learner counts

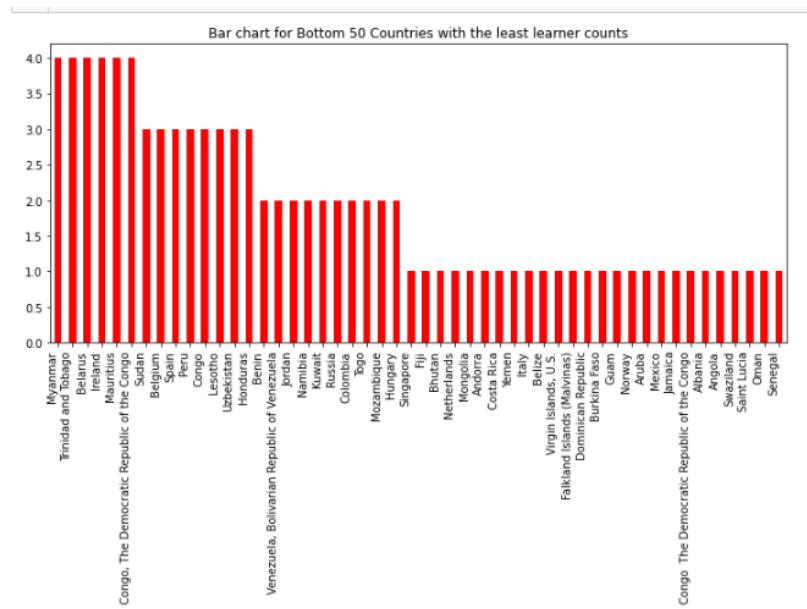
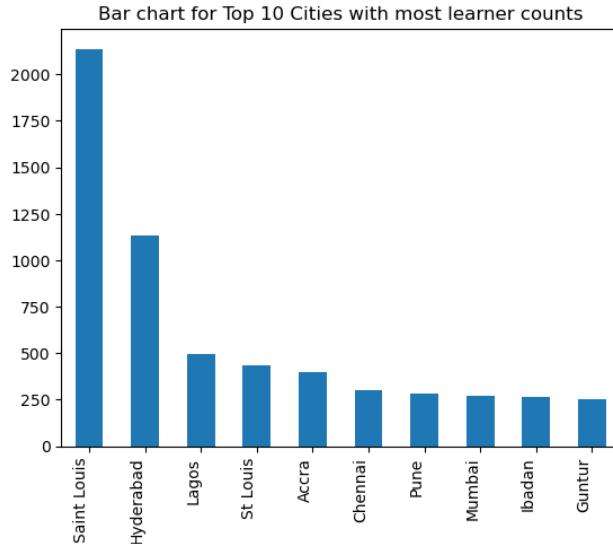


Figure. 10.11: Bar chart for Bottom 50 Countries with the least learner counts

The countries ranking in the bottom 50 have a learner count of **four** or fewer. This graph serves as a valuable tool to enhance Excelerate's outreach in these specific countries.

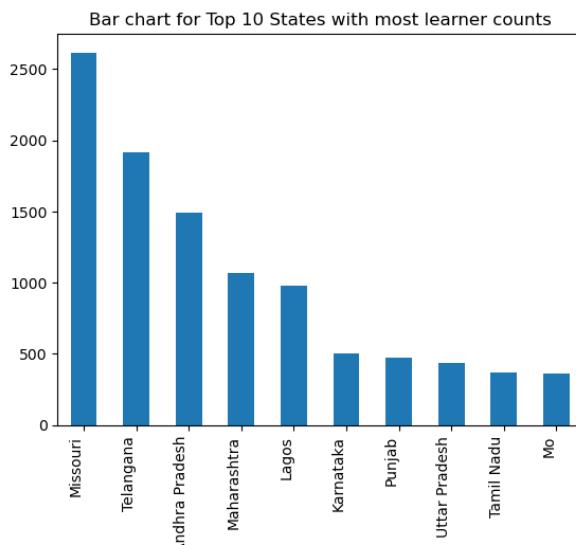
### 10.12 Top 10 Cities with top learners



**Figure10.12: Chart for cities with most learners**

From the graph, Saint Louis City boasts the highest number of learners, securing the top position. Hyderabad follows closely in second place, with Lagos taking the third spot. Remarkably, among the top 10 positions, more than four Indian cities are prominently featured as shown in the Figure above.

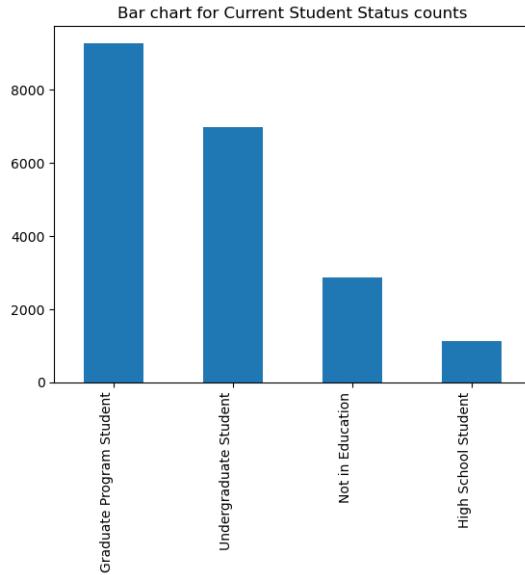
### 10.13 Top 10 States with most learners



**Figure10.13: Bar Chart for states with most learners**

From our analysis, Missouri takes the lead as the top state with the highest number of learners, closely followed by Telangana. Notably, within the top 10 positions, seven out of the ten are occupied by Indian states.

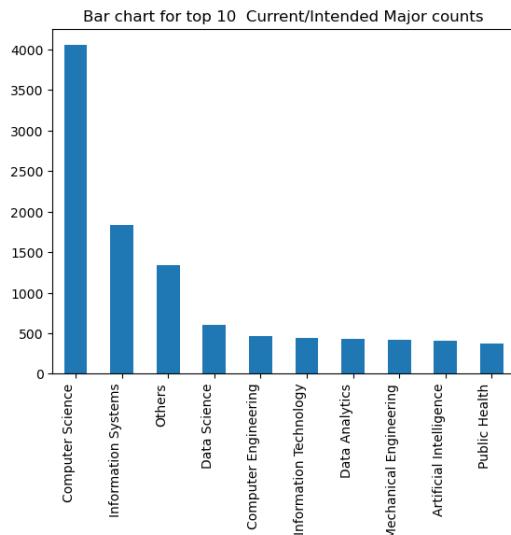
#### 10.14 Current Student Status



**Figure10.14: Bar Chart for student current status.**

The majority of learners who actively participated on the Excelerate website are enrolled in graduate programs, with undergraduate learners, individuals not currently in education, and high school learners following suit.

#### 10.15 Top 10 Current/Intended Majors

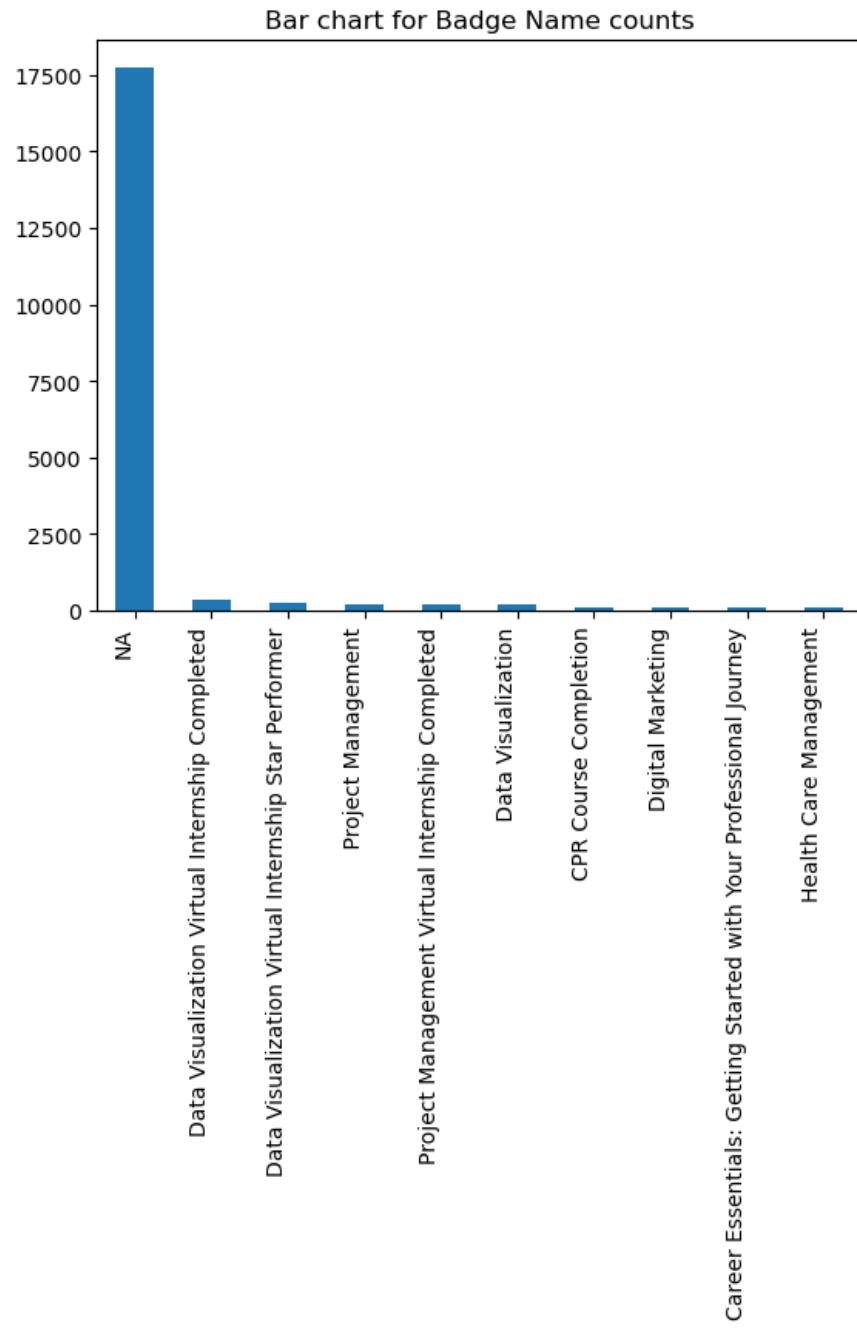


**Figure10.15: Bar Chart for top 10 current/intended majors**

From the analysis, Computer Science stands out as the primary field from which the majority of learners either graduated or plan to pursue as their intended major, with Information Systems following closely.

### 10.16 Badge Counts

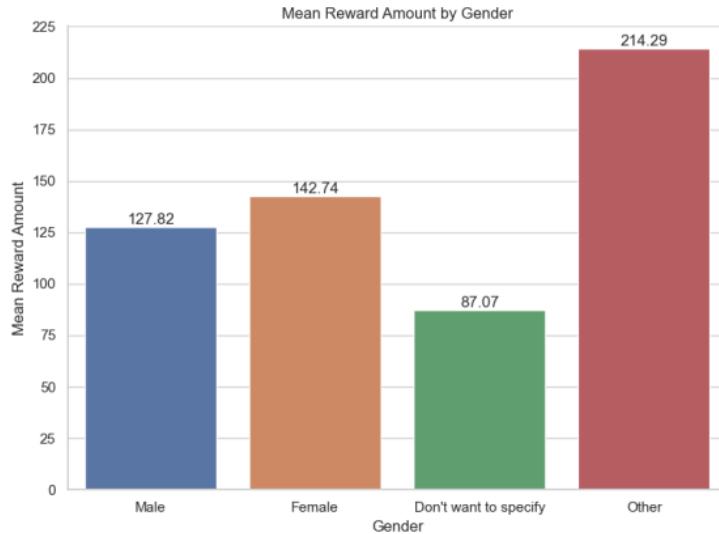
Similar to the observed pattern in the acquisition of skills, the majority of learners did not earn any badges. Only a small percentage achieved this recognition.



**Figure10.16: Bar Chart for badge name counts**

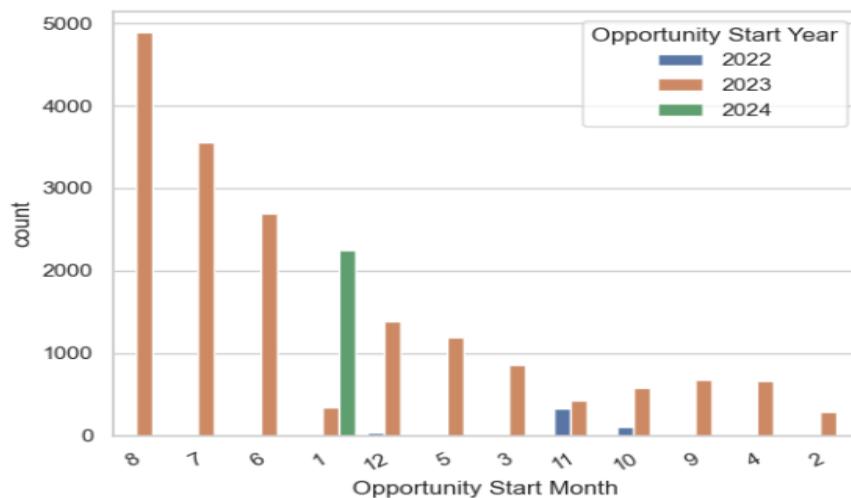
### 10.17 Average Reward Amount by Gender

When breaking down the Reward Amount by gender, the "others" category has the highest average reward amount at 214, followed by female candidates at 143, males at 128, and those opting not to specify at 87.



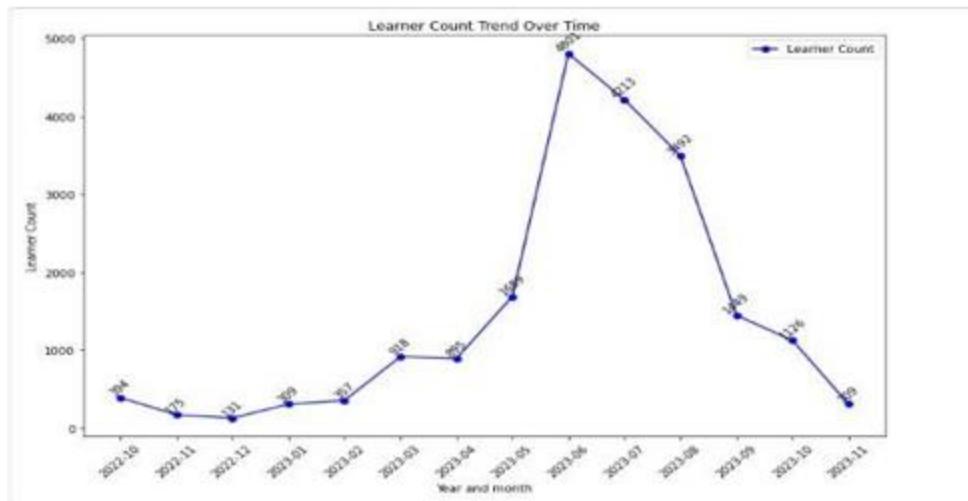
**Figure10.17 : Bar Chart for average reward amount by gender**

### 10.18 Opportunity Start Year



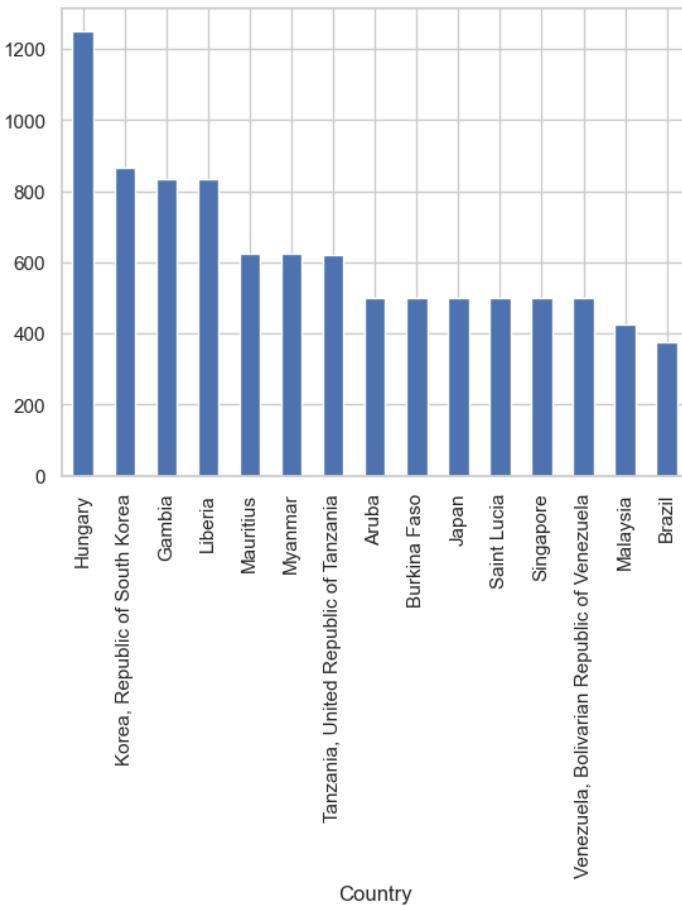
**Figure10.18a: Bar Chart for opportunity start month.**

Throughout 2023, numerous opportunities emerged almost every month. However, 2022 saw only two recorded opportunities, and for 2024, there is currently one scheduled in January. Figure10.18b below shows the fluctuations in learners' count. Notably, there is a sharp increase in learner enrolment from April to July 2023, then a decline from July 2023.



**Figure10.18b: learners Count Trend Over Time**

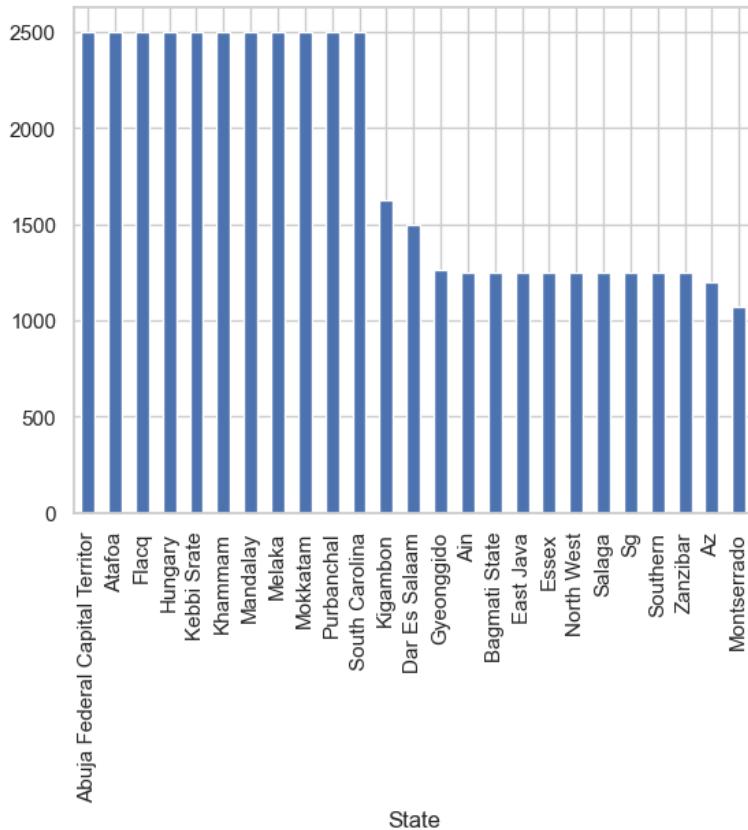
#### 10.19 Top Rewards Earned by Countries



**Figure10.19: Bar Chart for top rewards earned by countries**

Hungary secured the highest reward points among the countries, followed by Korea, Gambia, and Liberia in descending order.

### 10.20 Top Rewards Earned by States

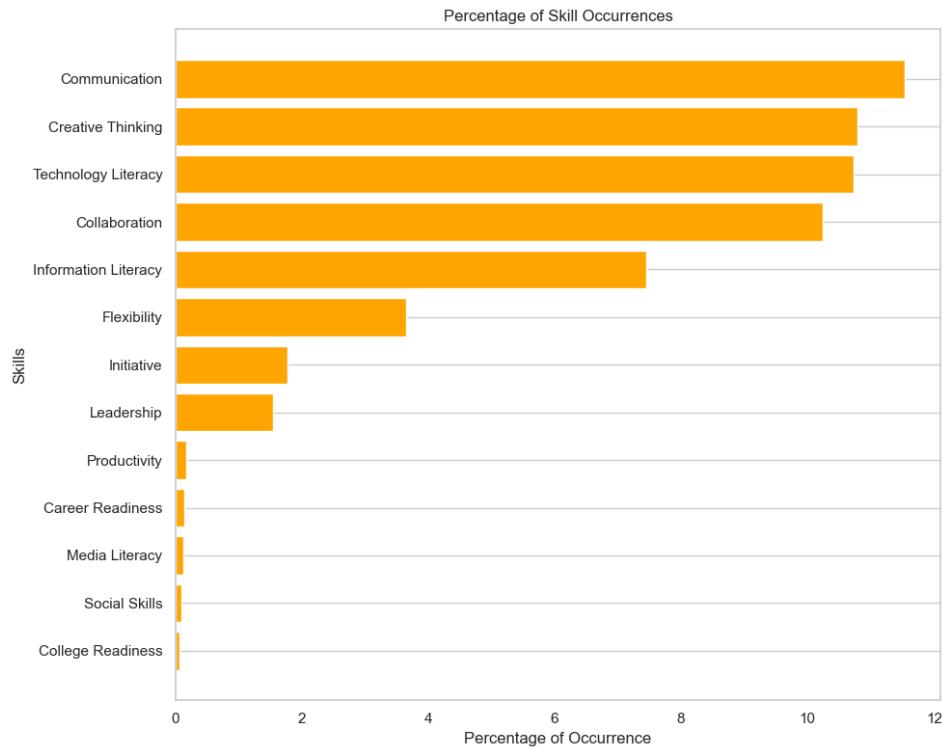


**Figure 10.20: Bar Chart for top rewards earned by state**

The analysis shows that learners from Abuja Federal Capital Territory, Atafoa, Flacq, Kebbi State, Khamman, Mandalay, Melaka, Mokkatam, Purbanchal, and South Carolina achieved the highest reward points, totaling \$2500.

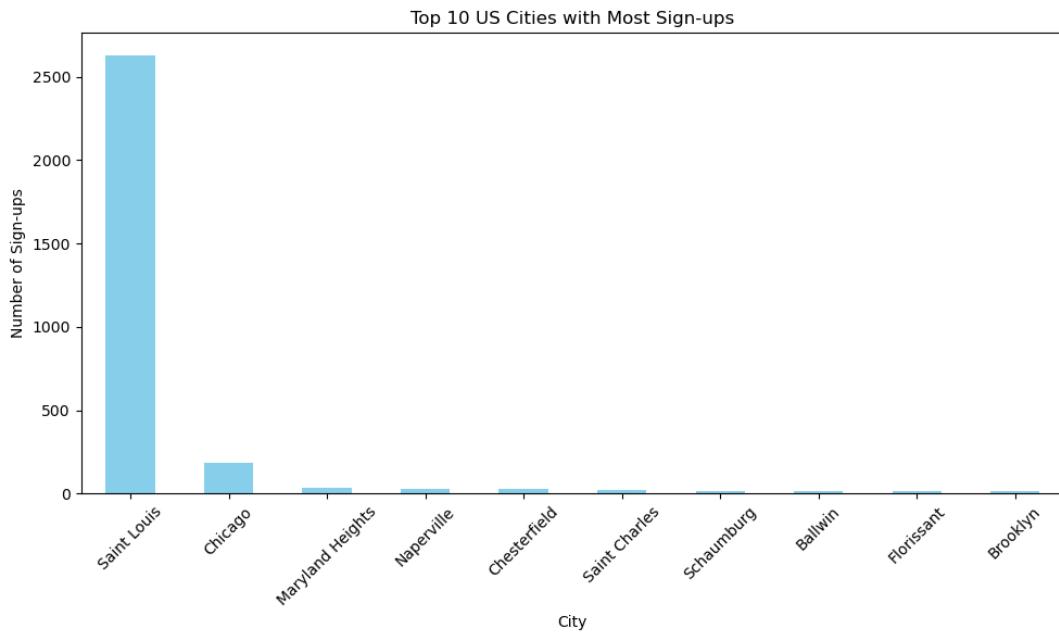
### 10.21 Overall Percentage of Skills Earned by Learners

When considering the skills acquired by learners, Communication stands as the foremost skill attained, trailed by Creative Thinking, Technology Literacy, Collaboration, Information Literacy, and more. The graph shows that Learners have predominantly acquired skills in Communication, Creative Thinking, and Technology Literacy, ranking them as the top three. Conversely, skills such as Media Literacy, Social Skills, and College Readiness have been less frequently gained, positioning them as the least acquired by learners.



**Figure10.21 : Bar Chart for skills earned by applicants**

## 10.22 US City Insights



**Figure10.22 Top 10 US City Sign-Up**

The US city insights shows that Saint Louis tops with the count of more than 2500 learners followed by Chicago with less than 250 learners.

### **10.23 Top opportunities signed by learners by country**

Top completed opportunities in India:

	Opportunity Name	Completed
241	Career Essentials: Getting Started with Your P...	482
246	Data Visualization	3288
247	Digital Marketing	961
253	Innovation & Entrepreneurship	558
264	Project Management	1727

Top completed opportunities in United States:

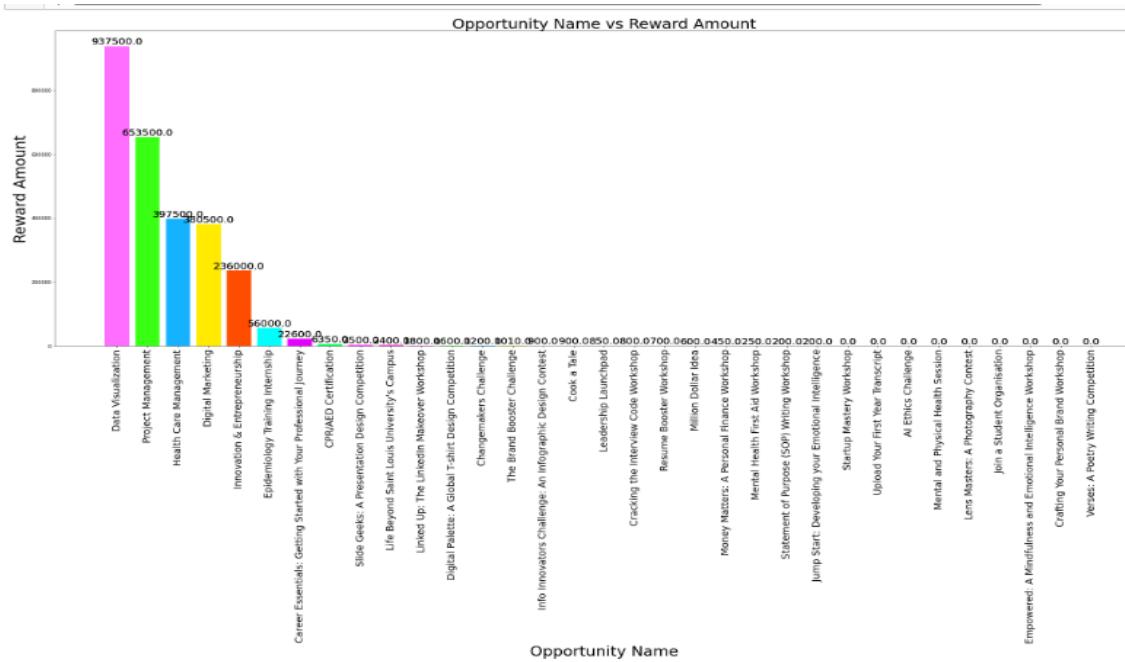
	Opportunity Name	Completed
668	Data Visualization	756
669	Digital Marketing	453
673	Health Care Management	402
675	Innovation & Entrepreneurship	352
686	Project Management	718

Top completed opportunities in Nigeria:

	Opportunity Name	Completed
432	Data Visualization	768
433	Digital Marketing	512
437	Health Care Management	344
439	Innovation & Entrepreneurship	154
447	Project Management	736

India records the highest learners followed by US, Nigeria and Pakistan

### **10.24 Opportunity Vs Reward Amount**



**Figure. 10.23: Bar chart for Opportunity Vs Reward Amount representation**

Total Reward given is \$2707310. It appears that there is a significant variation in the reward amounts awarded to learners across different opportunities. The top opportunities, such as "Data Visualization," "Project Management," "Health Care Management," and "Digital Marketing," have relatively high reward amounts, ranging from \$236,000 to \$937,500. This suggests that learners participating in these programs receive substantial financial support.

On the other hand, there are opportunities, such as "Startup Mastery Workshop," "Upload Your First Year Transcript," "AI Ethics Challenge," and others, where learners have not been awarded any rewards (reward amount is \$0.0). This could imply that either these opportunities do not involve financial incentives or that learners participating in these programs did not receive financial rewards.

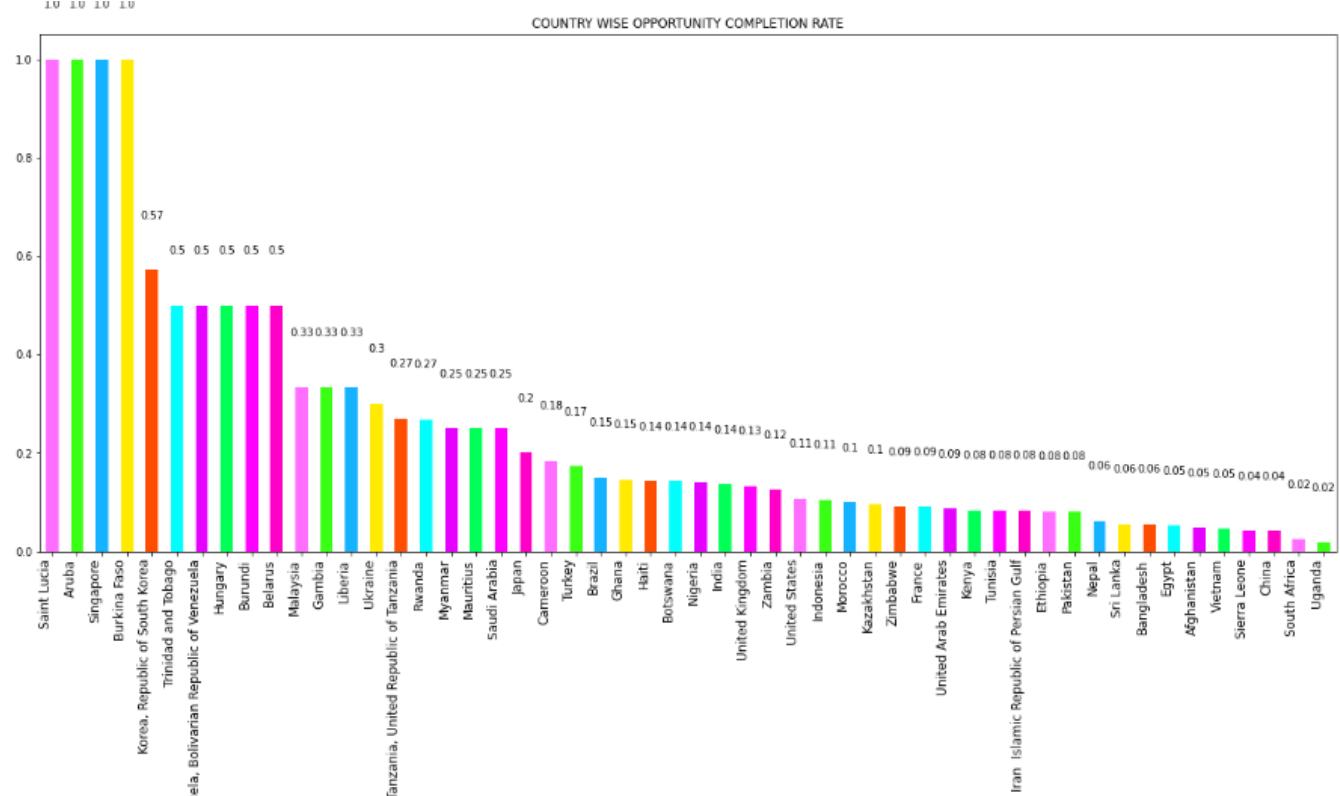
The absence of rewards in certain opportunities does not necessarily indicate a lack of impact. Some programs, such as "Join a Student Organisation," "Crafting Your Personal Brand Workshop," and "Verses: A Poetry Writing Competition," may be designed to provide non-financial benefits, such as skill development, networking, or personal growth.

#### Inferences:

- 1.The distribution of rewards varies widely among different opportunities.
- 2.Opportunities in "Data Visualization," "Project Management," etc., have been particularly lucrative for learners in terms of financial support.
- 3.Opportunities with zero reward amounts may focus on non-monetary benefits or might not have a financial incentive structure.
- 4.Assessing the impact of scholarships involves considering both the financial and non-financial benefits offered by different opportunities.

It's important to consider the goals and objectives of each opportunity, as well as the broader impact on learners' education, skills, and personal development, to draw comprehensive conclusions about the effectiveness of scholarship programs.

### 10.25 Country Wise Opportunity Completion Rate

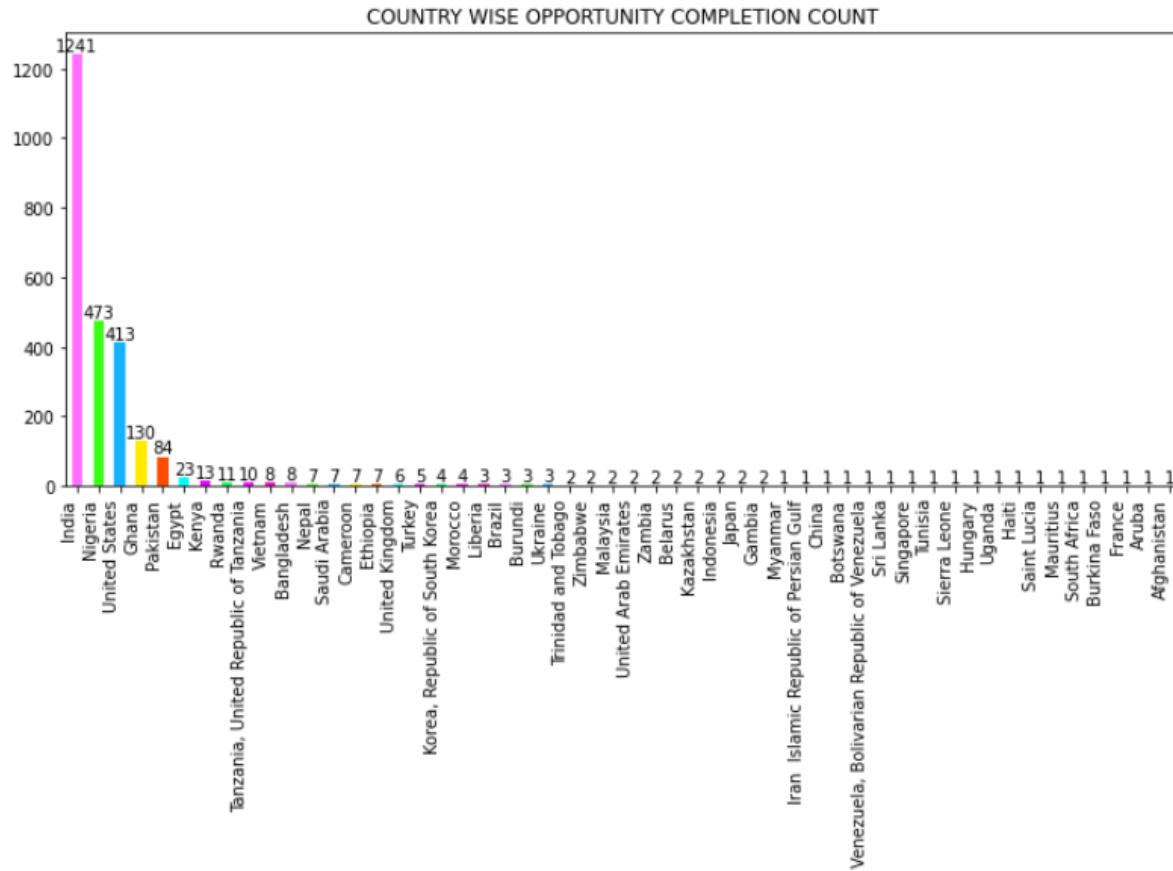


**Figure. 10.24: Bar Chart For Country Wise Opportunity Completion Rate**

While India, the United States, and Nigeria secure top positions in opportunity application rates, only a modest percentage of students, ranging from 11 to 14, succeed in completing the opportunity. In contrast, learners from countries such as Saint Lucia, Aruba, Singapore, Burkina Faso, and South Korea demonstrate a higher completion rate.

### 10.26 Country Wise Opportunity Completion Count

Indian learners lead the chart in terms of opportunity completion counts. Despite Nigeria ranking third in the number of learners who applied for the opportunities, it secures the second position in terms of opportunity completion.



**Figure. 10.25: Bar Chart For Country Wise Opportunity Completion Count**

### 10.3 Data Preprocessing

Preprocessing involves preparing the data for analysis or modeling by cleaning, transforming, and organizing it. The goal of preprocessing is to ensure that the data is in a suitable format for analysis or modeling. It enhances data quality, reduces noise, and prepares the dataset for machine learning algorithms or statistical analyses.

- Handling missing values
- Adding new columns
- Dealing with outliers
- Creating Dummy Variables
- Creating Scaling Features (not included)
- Encoding categorical variables (not included)
- Standardizing or normalizing data (not included)

## 1. Handling Missing Values:

Removing missing values isn't always the best approach. Sometimes, missing values contain valuable information or might represent meaningful patterns in the data. In such cases, imputation methods or other strategies to handle missing values without eliminating them might be more appropriate. The decision to remove missing values depends on the specific dataset, the nature of the missing values, and the analysis or modeling goals.

```
#Finding the null values
opportunity_data.isnull().sum()

Profile Id          0
Opportunity Name    0
Opportunity Category 0
Apply Date          0
Apply Time           0
Opportunity Start Date 0
Opportunity End Date 0
Gender               0
City                 0
State                0
Country              0
Graduation Date      0
Current Student Status 0
Current/Intended Major 0
Status Description    0
Reward Amount         0
Badge Name           17746
Skill Points Earned   0
Creative Thinking      0
```

## Imputing with missing values:

```
#Imputing null values with 'NA'
opportunity_data["Badge Name"] = opportunity_data["Badge Name"].fillna("NA")
```

We have imputed the “Badge Name” column with NA.

## 2. Adding New Columns

### Creating new columns

```
# Extract month from the Apply date
opportunity_data['Apply Year'] = pd.DatetimeIndex(opportunity_data['Apply Date']).year

# Extract year from the Apply date
opportunity_data['Apply Month'] = pd.DatetimeIndex(opportunity_data['Apply Date']).month

# Extract month from the opportunity start date
opportunity_data['Opportunity Start Year'] = pd.DatetimeIndex(opportunity_data['Opportunity Start Date']).year

# Extract year from the opportunity start date
opportunity_data['Opportunity Start Month'] = pd.DatetimeIndex(opportunity_data['Opportunity Start Date']).month

# Extract month from the opportunity end date
opportunity_data['Opportunity End Year'] = pd.DatetimeIndex(opportunity_data['Opportunity End Date']).year

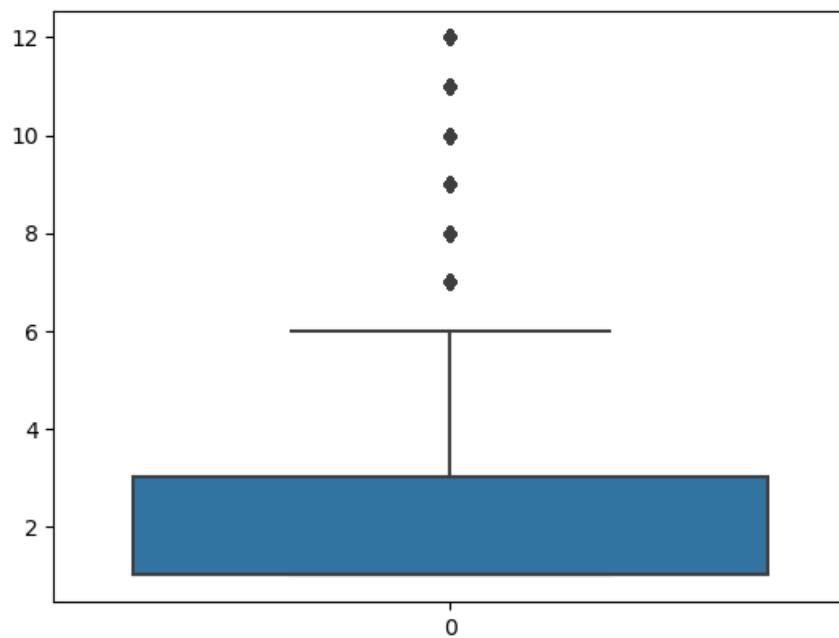
# Extract year from the opportunity end date
opportunity_data['Opportunity End Month'] = pd.DatetimeIndex(opportunity_data['Opportunity End Date']).month
```

New columns were extracted for the analyzation purpose and the original column is dropped.

## 3. Dealing with outliers:

Analyzing outliers is critical for ensuring data quality, understanding data distributions, improving model performance, and gaining valuable insights that might not be apparent in the regular data distribution. Outliers can be found only for the numerical data.

### 3.1 Graduation Date:

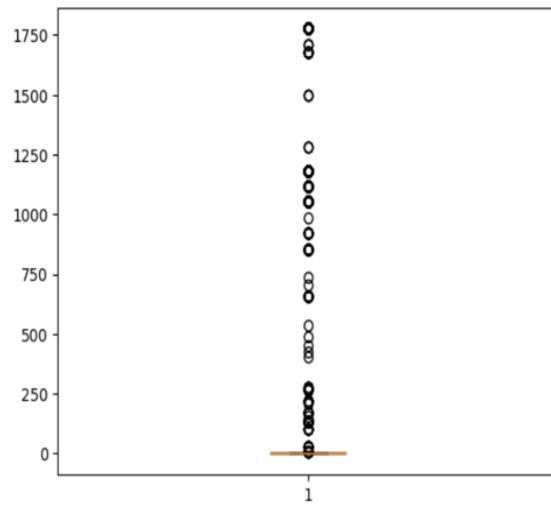
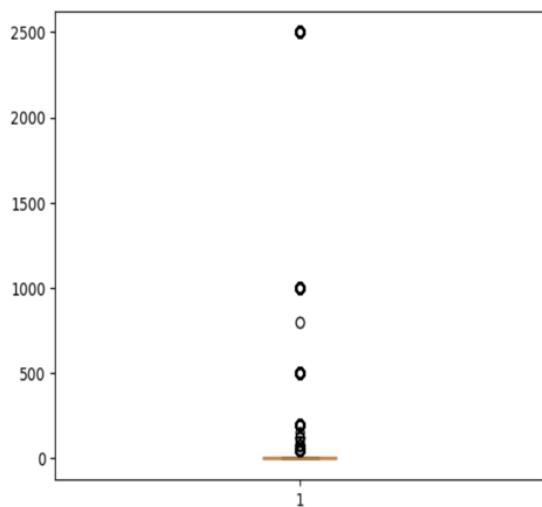


**Figure 26: Boxplot for Graduation Year**

The "Graduation Year" column contains outliers as it includes applicants who graduated before 2017. However, these outliers have not been removed as this learnership opportunity has no specific restriction on the year of graduation, remaining open to applicants from varied graduation years.

### 3.2 Reward Amount and Skill Points

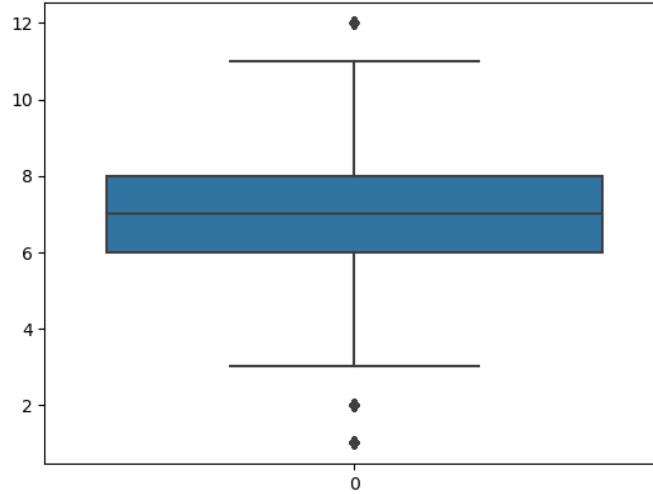
The figures below show box plot of the Reward Amount and Skills Points Earned. This exhibits that not all candidates were able to get the Reward Amount and Points



**Figure 10.27.a and 10.27.b Box-plot for Reward Amount and Skills Points Earned**

Skill Points and Reward Amount indicate the scholarship earned and intern eligibility spread. The presence of numerous outliers is notable, largely due to a higher count of learners with no Reward or Skill Points. This suggests a need for learners to invest additional effort in leveraging the sponsorship for their benefit.

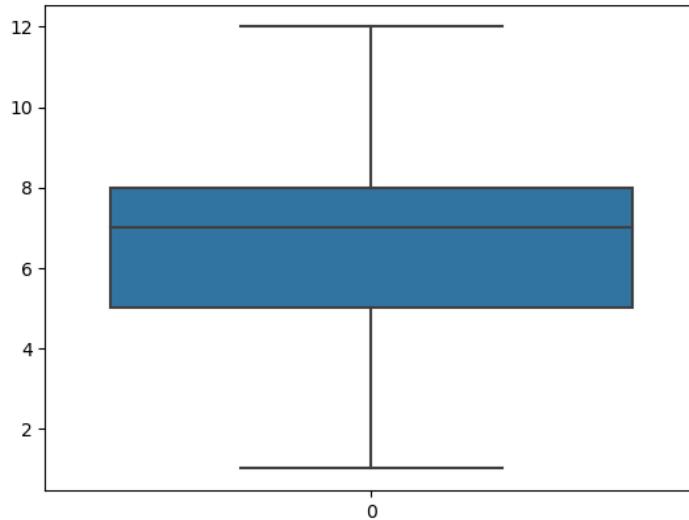
### 3.3 Opportunity Start Year



**Figure10.28:** Boxplot for opportunity start year

A few outliers in the Opportunity Start year have been excluded from the analysis.

### 3.4 Opportunity End Year



**Figure10.29:** Boxplot for opportunity End Year

The opportunity end year did not show any outliers.

## 4 Creating Dummy Variables

Creating the Dummy variables We need to convert the categorical variable to numeric.

1.Create the dummy variable

2.Drop the original variable

3.Drop one dummy variable for each dummy variable (n-1)

```
opportunity_data = pd.get_dummies(opportunity_data, drop_first=True)
opportunity_data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20258 entries, 0 to 20257
Columns: 15990 entries, Reward Amount to Badge Name_Statement of Purpose (SOP) Writing Workshop
dtypes: float64(1), int64(24), uint8(15965)
memory usage: 312.3 MB
```

### Creating dummy variables using get\_dummies

Reward Amount	Skill Points Earned	Creative Thinking	Technology Literacy	Communication	Career Readiness	Initiative	Media Literacy	Leadership	Information Literacy	Collaboration	Flexibility	Productivity
0	0	0	0	0	0	0	0	0	0	0	0	0
1	200	10	1	0	1	1	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0

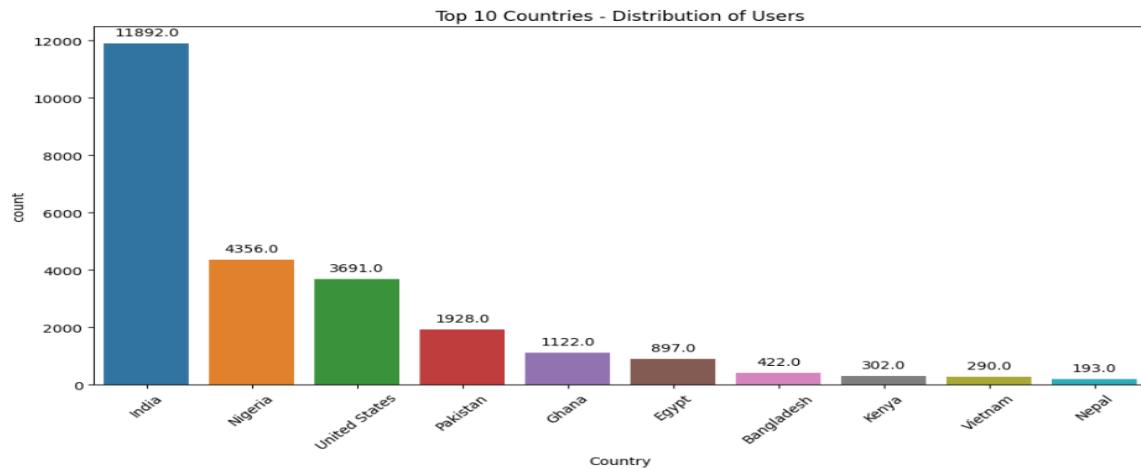
Figure10.30: Dataset after mapping with dummy variables

## 5. Comparative Analysis

### 1. Comparison on the overall rates of users signed up on Excelerate platform Vs Learners completion

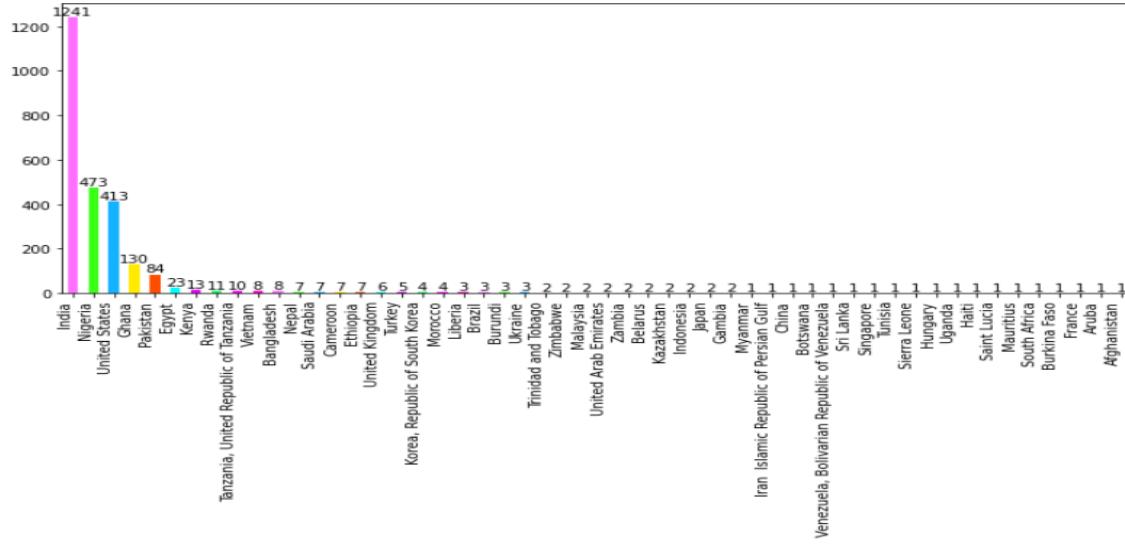
In the context of the Excelerate platform, a total of 27,557 users have signed up. When examining the opportunity application rates, India, the United States, and Nigeria emerge as the top contributors, showcasing high application numbers. However, the completion rates present a nuanced perspective, revealing that only a modest percentage of students, ranging from 11 to 14, successfully complete these opportunities.

Contrastingly, learners from countries like Saint Lucia, Aruba, Singapore, Burkina Faso, and South Korea demonstrate significantly higher completion rates. Notably, Indian learners lead the chart in terms of opportunity completion counts, despite Nigeria securing the second position in terms of opportunity completion ranking third in the number of learners who applied for opportunities.



### Top 10 Countries among Users

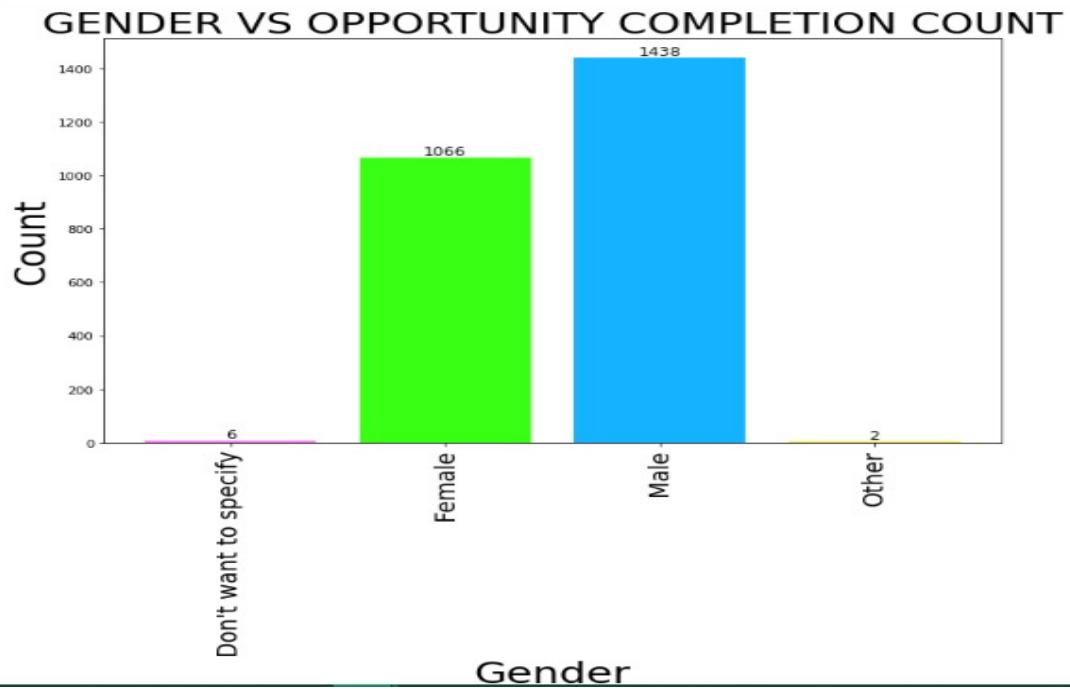
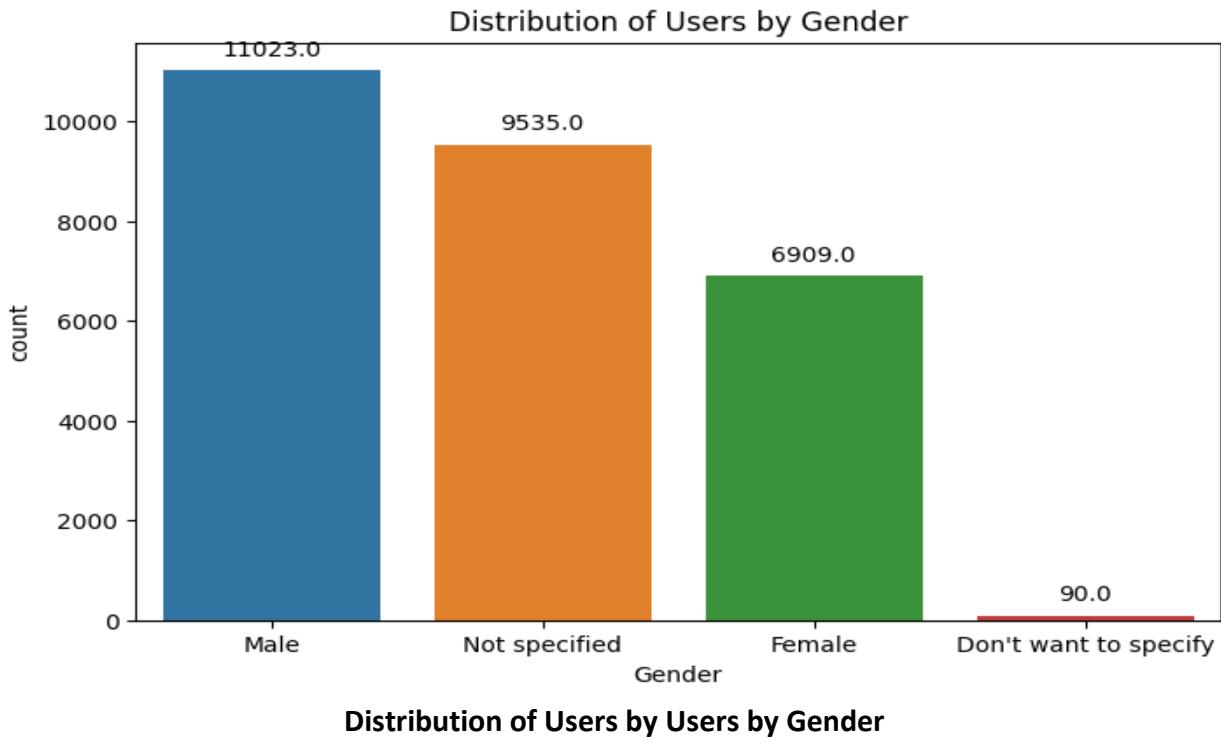
COUNTRY WISE OPPORTUNITY COMPLETION COUNT



### Country-wise Completion Count

## 2. Comparison on the users signed up on Excelerate Platform Vs Learners completion based on gender

As shown in the Figure below, the following are the distribution of gender who have signed up for learning opportunity on the Excelerate platform: 11023 are Males, 6909 are Females, while 9535 and 90 are not specified and don't want to specify respectively. However, there are 1438 Males, 1066 Females learners who have completed the opportunities while there are 6 and 2 who don't want to specify and others respectively.



### **3. Comparison on the registered learners from different countries Vs Learners completion**

As shown in the graph above, India, the United States, and Nigeria have high opportunity application rates, suggesting significant interest and engagement from learners in these

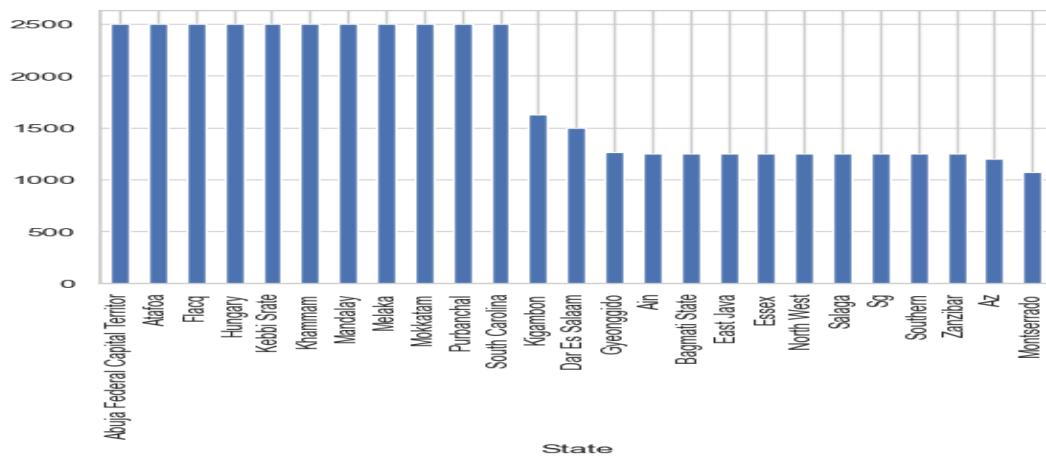
countries. However, the completion rates are modest, ranging from 11 to 14 percent. This indicates a challenge in converting application interest into successful opportunity completions. Learners from countries such as Saint Lucia, Aruba, Singapore, Burkina Faso, and South Korea demonstrate higher completion rates, suggesting a relatively better success rate in completing the opportunities they apply for.

This variation across countries indicates potential differences in factors influencing completion, such as educational systems, learner commitment, or platform-specific features tailored to certain regions. Indian learners stand out by leading the chart in terms of opportunity completion counts. This suggests that while the completion rates might be modest, the sheer volume of Indian learners translates into a significant number of completed opportunities. Despite ranking third in the number of learners applying for opportunities, Nigeria secures the second position in terms of opportunity completion. This could imply that Nigerian learners, proportionally, have a higher likelihood of completing opportunities compared to their application rates.

#### **4. Comparison on the user sign up Vs learners completion Vs reward Amount**

There's a disparity between application rates and completion rates, highlighting the need for a closer look at factors influencing learner success. The success of Indian learners in completing opportunities indicates a potential area for analysis to understand contributing factors.

Nigeria's performance, securing a higher completion position than its application rank, emphasizes the effectiveness of Nigerian learners in completing opportunities. The analysis highlights the significant variation in reward amounts across different opportunities, with specific high-reward programs and notable achievements by learners from particular locations. Understanding these patterns is crucial for optimizing reward structures, program offerings, and strategic planning to enhance learner engagement and success. Learners from specific locations, including Abuja Federal Capital Territory, Atafoa, Flacq, Kebbi State, Khamman, Mandalay, Melaka, Mokkamat, Purbanchal, and South Carolina, achieved the highest reward points. The total reward points accumulated by learners from these locations amount to \$2,500 as shown in the Figure below.



**Figure 5. State wise count on learners**

The variation in reward amounts suggests that the platform has an opportunity-specific reward structure. Understanding the factors influencing reward amounts can guide future program design and financial planning.

## **6. Challenges Faced**

- Handling Large Date-set: The processing time of the large dataset was time-consuming and the computational power also affected the analysis of the large data set.
- Visualization Challenge: Slicing and dicing of data in various aspects were intellectually challenging
- Data Quality Issues: Some blanks in many certain rows. They were replaced so as not to affect the data analysis.
- Comparative analysis was intellectually challenging as we need to think in the business perspective.

## **12. Next Steps**

- Understanding the dashboard and its requirements.
- Narrowing down the research on fresh insights
- Learning curve for new data visualization tool (Google Data Studio)

## **Conclusion**

The comprehensive analysis reveals compelling insights into learner engagement and demographics on the Excelerate platform. The majority of applications contribute to the total, yet unique applications represent a substantial portion. Males constitute a higher percentage of engagement compared to females, with a small percentage categorized as 'others' or unspecified.

Most actively participating learners have their teams allocated. Notable trends include a significant number of applicants from 2020 and those graduating in 2026, highlighting a concentrated count in 2024. While a small number received high reward amounts, it's evident that accessing scholarships might be challenging for many learners. This underscores the importance of taking leadership seriously as a career booster. Efforts in skill acquisition are crucial, considering only a minimal number benefit substantially.

Internships draw the most engagement, while Data Visualization emerges as a prevalent field among learners. The peak months for enrollment are June, July, and August, consistent with User Data visualization. India hosts the highest number of engaged learners, followed by the United States, Nigeria, and Pakistan. Within cities, Saint Louis, Hyderabad, and Lagos lead in learner engagement. Missouri and Telangana rank highest among states, notably featuring Indian states predominantly in the top positions.

Graduate programs attract the majority of learners, with Computer Science being the primary field of study. Skills acquisition and badge attainment show limited success among learners.

Reward amounts vary by gender, with 'others' receiving the highest average reward, followed by females, males, and unspecified. Opportunities emerged consistently throughout 2023, while 2022 had fewer records, and 2024 indicates one scheduled opportunity.

Hungary leads in reward points, and learners from various locations globally achieved top reward points. Communication emerges as the primary skill attained, followed by Creative Thinking, Technology Literacy, and Collaboration. Saint Louis dominates US city engagement, followed by Chicago.

These insights offer a comprehensive view of learner engagement, shedding light on trends, geographic concentrations, and key areas for platform enhancement and user experience optimization.

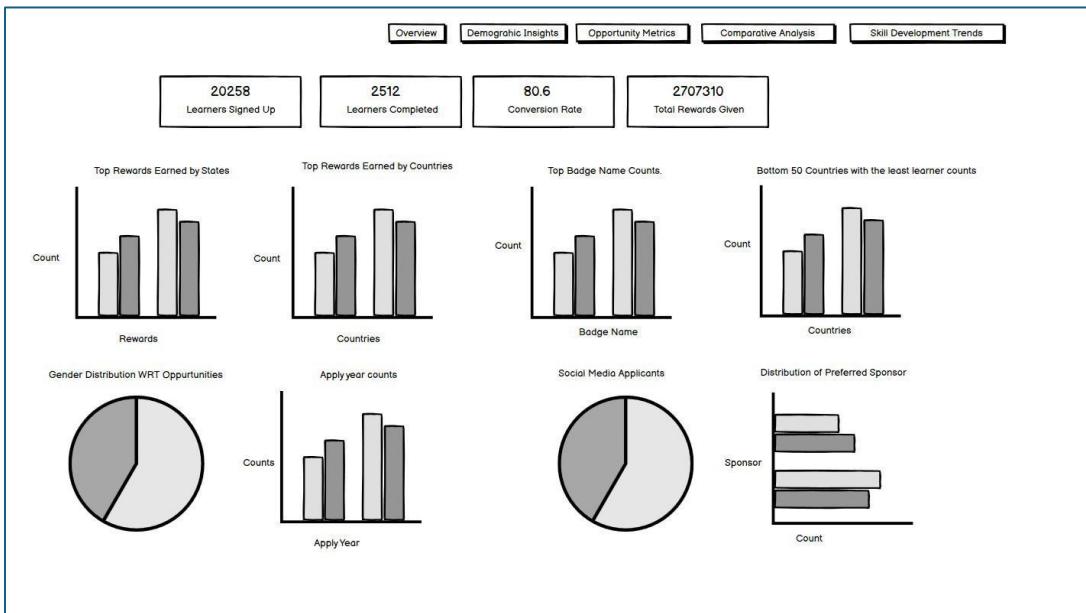
## Recommendations

- **Scholarship Accessibility:** Address the challenge of limited access to scholarships by providing more guidance and resources for learners to maximize their benefits from learnerships.
- **Skill Acquisition:** Encourage and support efforts in skill acquisition among learners, ensuring a broader base benefits substantially from the platform's opportunities.
- **Internship Opportunities:** Leverage the popularity of internships and consider expanding offerings in this area to meet the high engagement observed.
- **Field Focus:** Acknowledge the prominence of Data Visualization and consider further development and support in this field to cater to learner interests.
- **Geographic Focus:** Acknowledge and appreciate the significant learner engagement from India, the United States, Nigeria, and Pakistan, and tailor platform enhancements to cater to these regions' needs.
- **Educational Focus:** Given the strong enrollment in graduate programs and the prevalence of Computer Science as a field of study, tailor content and resources to align with these educational preferences.
- **Enhanced Guidance:** Provide additional support and encouragement for learners to pursue badge attainment and skill development to maximize their learning outcomes.
- **Gender-based Support:** Consider specific initiatives or programs to address gender-based differences in reward amounts, ensuring equitable opportunities across all categories.
- **Opportunity Planning:** Plan and facilitate a more consistent stream of opportunities throughout the year, considering the observed peaks in enrollment during June, July, and August.
- **Localized Engagement:** Recognize the significance of specific cities and states such as Saint Louis, Hyderabad, Lagos, Missouri, and Telangana, focusing on localized strategies for engagement and support.

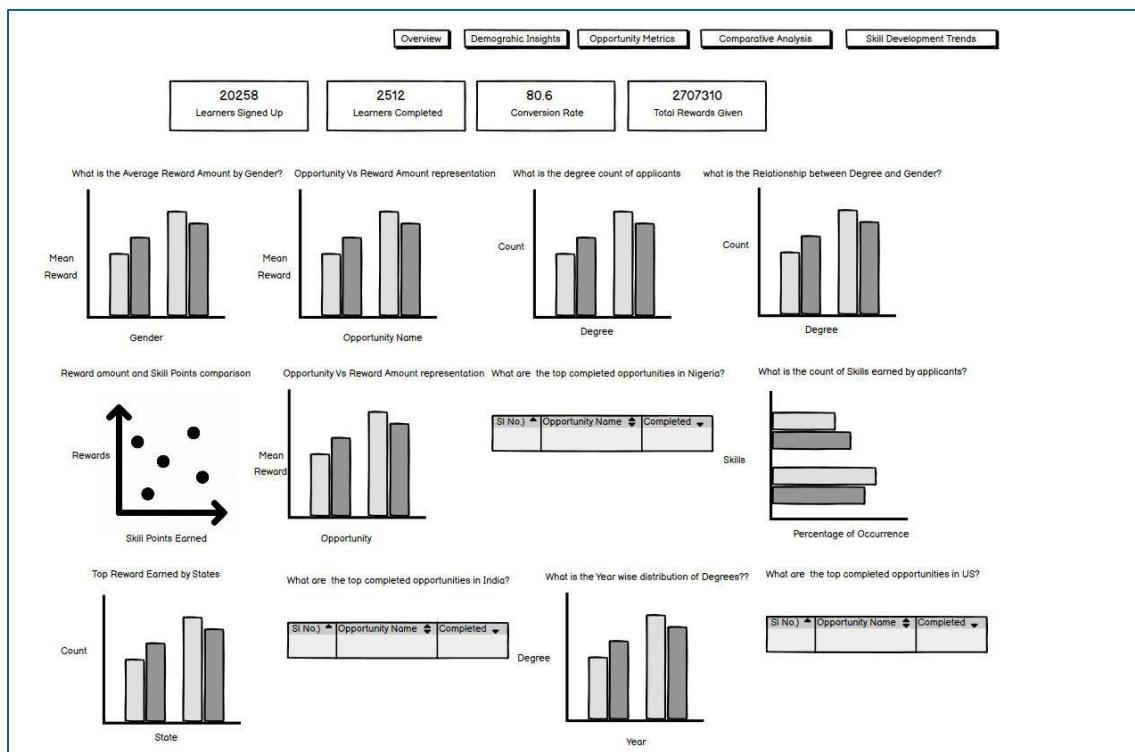
## PART 3: - WIREFRAME FOR DASHBOARD DESIGN

### 3.1 Wireframe Dashboard Design

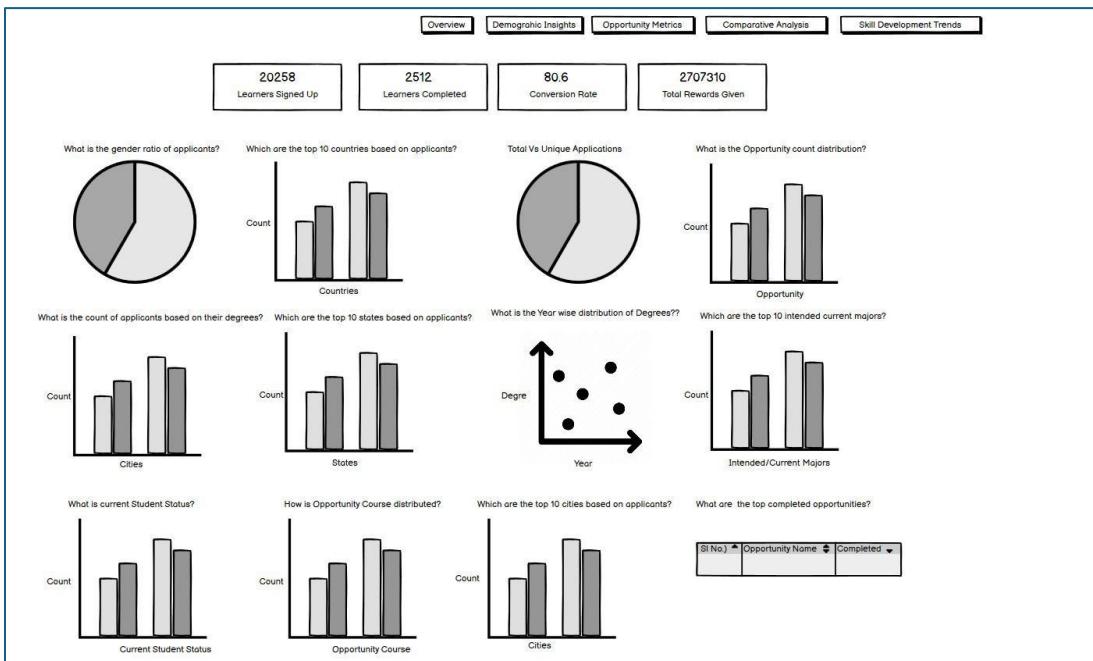
**Dashboard-1**



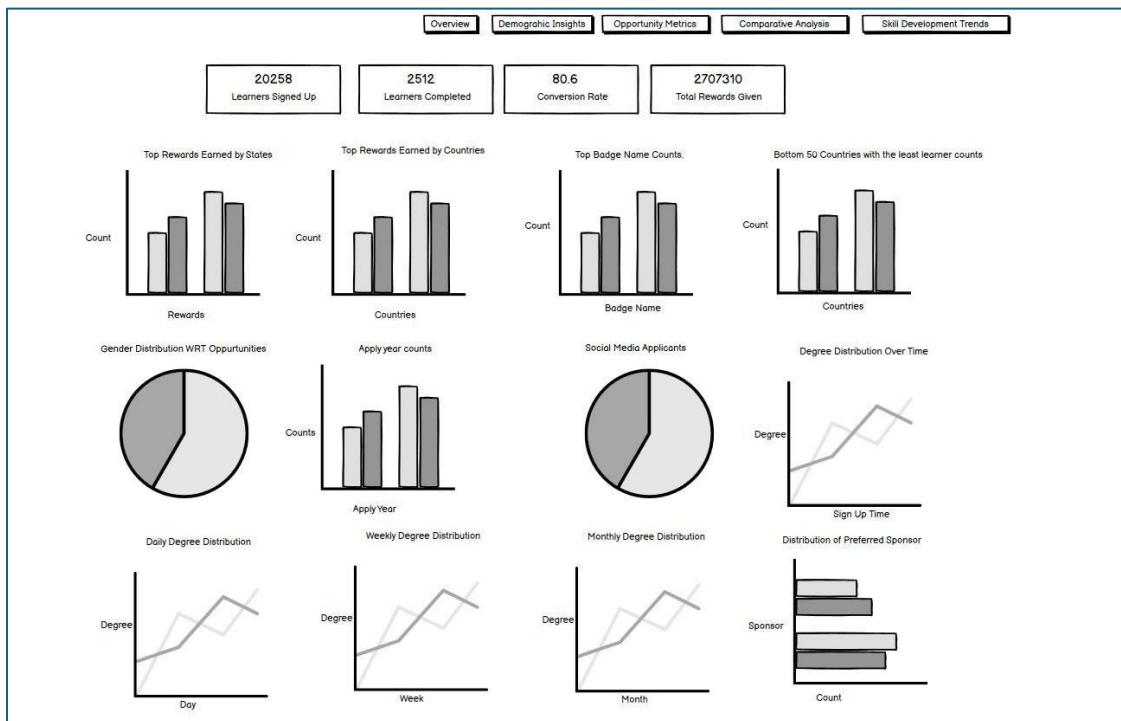
**Dashboard-2**



## Dashboard-3

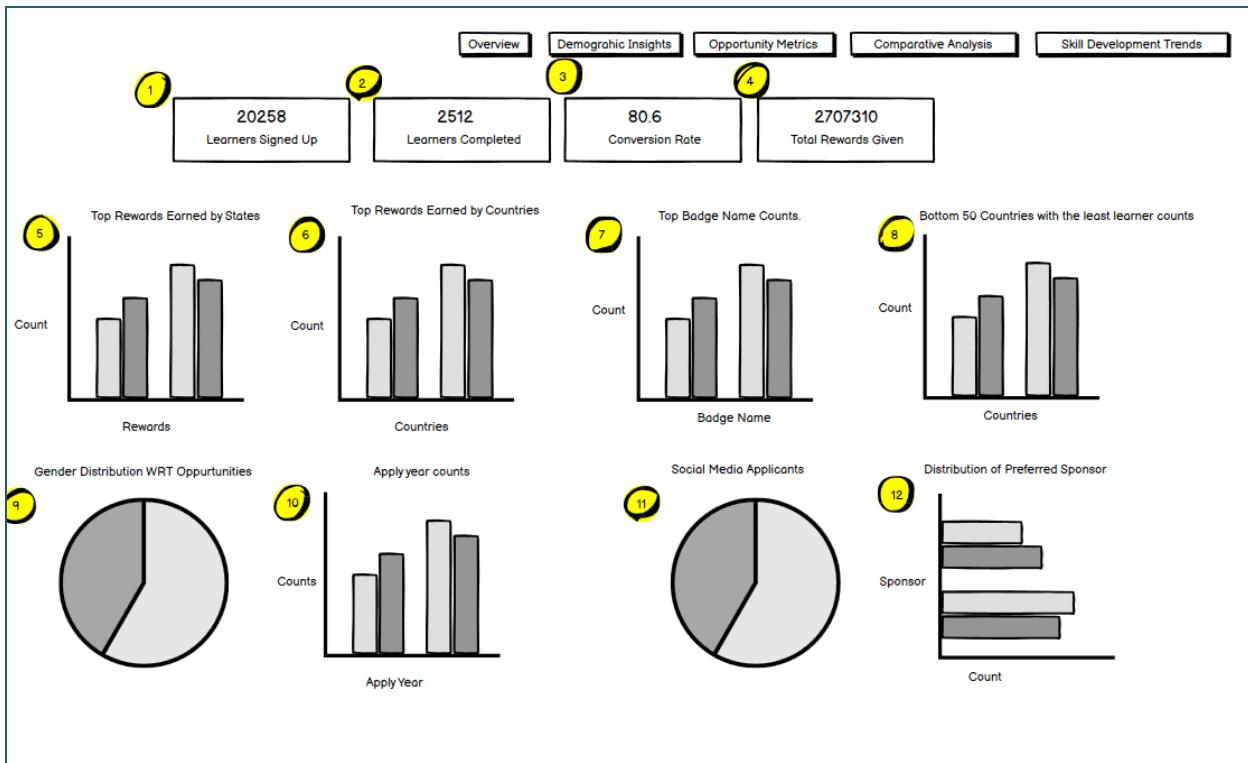


## Dashboard-4



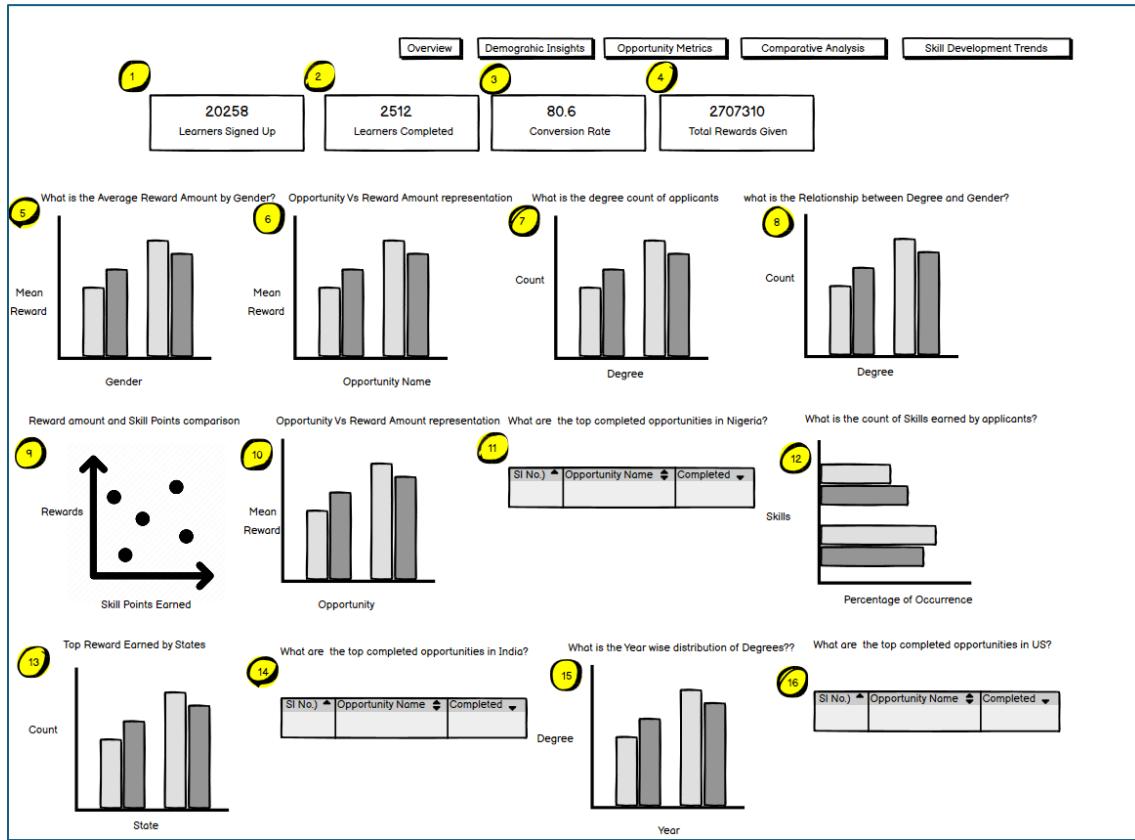
### 3.2 Wireframe Annotation:

#### Dashboard-1



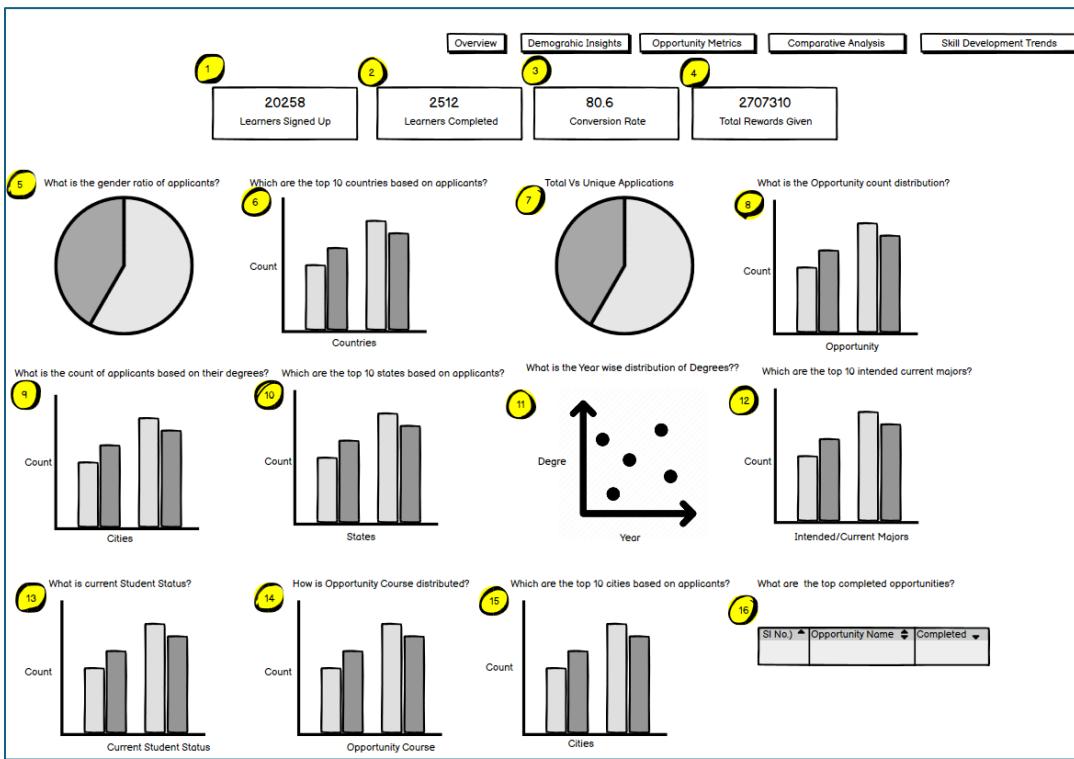
1	Total learners sign up
2	Total learners completed the opportunity
3	Learner's conversion rate
4	Total Rewards given by the sponsor
5	Analysing top reward earned by states
6	Analyzing top Rewards earned by Countries
7	Top Badge earned by Learners
8	Analysis based on least learner's country
9	Analyzing Gender Distribution and opportunities
10	Total application count based on Year
11	Analysing digital marketing strategies
12	Analysing sponsor preference

## Dashboard-2



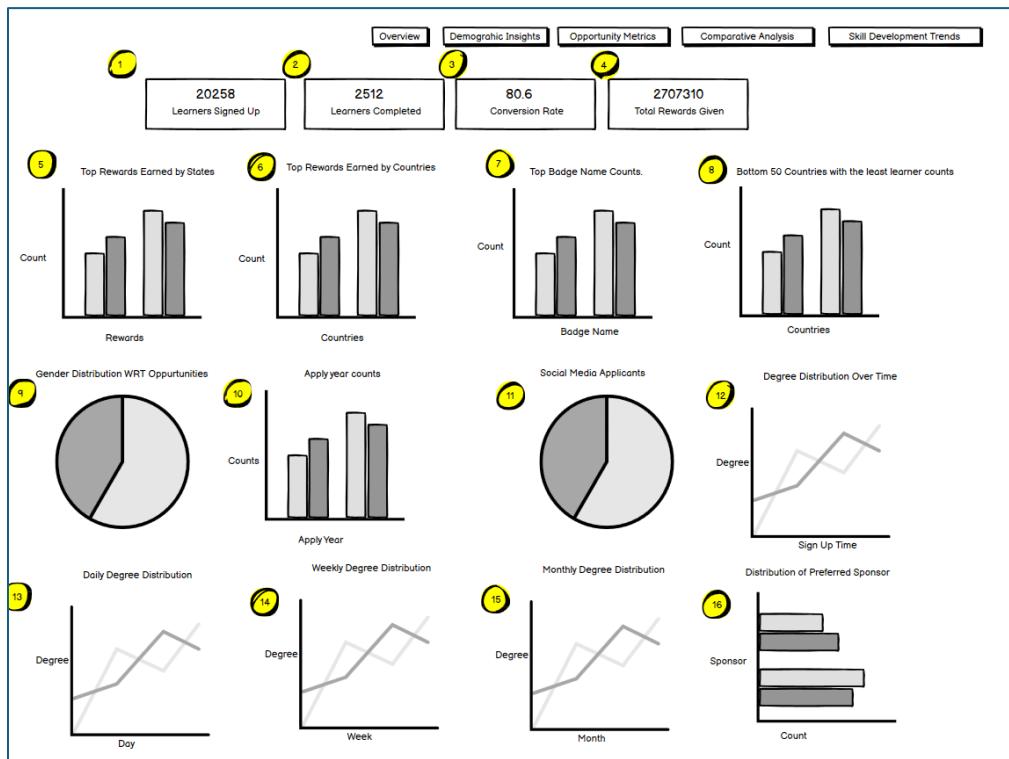
1	Total Learners signed up for the opportunity
2	Total learners completed the opportunity successfully
3	Total learners conversion rate
4	Total Reward Amount issued by sponsor
5	Gender analysis on average Reward Amount gained
6	Analyzing gaps between opportunity and Reward Amount
7	Number of applicants based on degree
8	Analysing relationship between Degree earned by the learner and Gender
9	Trend analysis between Reward Amount and Skill Points earned
10	Analysing Opportunity and Reward Amount representation
11	Analysing popular completed opportunities in Nigeria
12	Percentage count of skills earned by applicants
13	Top Reward Amounts earned by States
14	Analysing popular opportunities in India
15	Analysing Year wise distribution of Degrees
16	Analysing top completed opportunities in US

## Dashboard-3



1	Total Learners signed up in Excelerate Platform
2	Total learners who has successfully completed the opportunity
3	Total conversion rate of the learners
4	Total Rewards given by the sponsor
5	Categorizing gender-based applicants
6	Top 10 countries with maximum count of application
7	Comparison ratio between total applicant and unique applicants
8	Bar graph representing the Opportunity count
9	Applicants count based on their Cities
10	Top 10 states based on application count is monitored
11	Trend analysis over year wise distribution
12	Top 10 popular intended/current majors to analyze the students specialization and interest applying for opportunity
13	Analyze the current status of the student
14	Analyze popular Opportunity Course distribution
15	To find the top 10 popular cities based on applicants
16	Table to analyze the popular opportunities completed

## Dashboard-4



1	Total learners' signup count
2	Total learners completed
3	Total Conversion rate of the learners
4	Total Rewards given by the sponsor
5	State-wise analysis on Rewards earned
6	County-wise analysis on Rewards earned
7	Total Badges earned by learners
8	Least 50 countries with learners counts
9	Gender-wise distribution of learners and opportunity availed
10	Apply Year analysis
11	Digital marketing analysis
12	Analysis of Degeree distribution over time
13	Analysis on Degree distribution on daily basis
14	Analysis on Degree distribution on weekly basis
15	Analysis on degree distribution on monthly basis
16	Analysis of preferred sponsor distribution among learners

## **Overall Conclusion**

The report meticulously outlines the critical phases in data processing, beginning with data cleaning, encompassing exploratory data analysis (EDA), preprocessing, and culminating in the design of a wireframe dashboard with detailed annotations. The data cleaning segment delineates the rigorous steps taken to ensure data accuracy by rectifying missing values, eliminating duplicates, and standardizing formats. Subsequently, the EDA section illuminates the in-depth analysis conducted to uncover patterns, correlations, and outliers, crucial for informing subsequent preprocessing steps. Preprocessing details the transformations applied to the data post-EDA, including feature scaling and categorical variable encoding, refining it for model readiness. The pinnacle of the report lies in the wireframe dashboard illustration, presenting a visual representation of data insights through charts, filters, and interactive components. The accompanying annotations expound upon each element's purpose and functionality, enabling stakeholders to envisage the dashboard's layout and comprehend how it encapsulates the project's objectives and data-driven insights. This comprehensive approach not only reinforces the importance of visualization in data comprehension but also equips sponsors with a clear visualization of the final dashboard's potential efficacy in conveying complex data insights.