

Exploratory Data Analysis

CONTENTS

1- INTRODUCTION

2- DATA OVERVIEW

3- COLUMN ANALYSIS

4- DATA CLEANING

5- PROFILE ID ANALYSIS

6- OPPORTUNITY STATUS DISTRIBUTION

7- BASIC STATISTICS

8- INITIAL OBSERVATIONS

9- VISUALIZATIONS

10- CHALLENGES FACED

11- NEXT STEPS

1- Introduction

Purpose of the EDA

Our purpose of this EDA report is to gain comprehensive understanding of the datasets, assess the data quality, identify patterns and relationships and detect outliers. This will help us perform informed data processing and facilitate data-driven decision making.

Datasets

- **User Data:** Covers non-identifying information about all the users who ever created an account on Excelerate regardless of whether they're engaged in a specific opportunity or not.
- **Opportunity Sign Up and Completion Data:** Shows non-identifying information of users who signed up for specific opportunities on Excelerate.

2- Data Overview

User Data

- Each row represents a unique user account on Excelerate.
- The dataset contains 27,562 rows excluding the header rows.
- It contains 8 columns
- There're no unique identifiers for each row on the dataset.
- There were no duplicate rows found.

Opportunity Sign Up Data

- This dataset contains the opportunities that users signed up to. Each user may sign up for multiple opportunities so there could be more than one row with the same Profile ID.
- The dataset contains 20,322 rows excluding the header rows.
- A combination of Profile ID and Opportunity ID columns can be used as a unique identifier for each row in this dataset.

- It contains 21 columns.
- There were no duplicate rows found.

3- Column Analysis

User Data

Column Heading	Analysis
PreferredSponsors	<ul style="list-style-type: none"> • This column shows the different sponsors users can choose on Excelerate platform. Users can choose more than one sponsor. • Data Type: String • No missing values
Gender	<ul style="list-style-type: none"> • This column shows the gender of the users. • Data type: String • Distribution of each gender fields: <ul style="list-style-type: none"> - Female 6,910 - Male 11,027 - Don't want to specify 75 - Other 15 • Not a mandatory field so there are 9,535 missing values
Country	<ul style="list-style-type: none"> • This shows the country each user indicated they live in. • It contains 170 different countries • Data Type: String • Missing Values: 62
Degree	<ul style="list-style-type: none"> • Shows the academic level of each user. • Data Type: String • This is also not a mandatory field so there are 9,585 missing fields • Unique Values: <ul style="list-style-type: none"> - High School Student 1,562 - Not in Education 2,646 - Graduate Program Student 6,015 - Undergraduate Student 6,527
Sign Up Date	<ul style="list-style-type: none"> • Shows the date each user created their Excelerate account. • Data Type: Date and time • No missing values

City	<ul style="list-style-type: none"> Shows the city each user indicated they live in. Data Type: String Not a mandatory field so there are 9,533 missing values. This column contains a lot of misspelled and unclear values.
Zip	<ul style="list-style-type: none"> Shows the zip code of the city each user indicated they live in. Data Type: String This column contains missing and unclear values
IsFromSocialMedia	<ul style="list-style-type: none"> This shows whether the user signed up via a social media account or not. Data Type: Boolean Missing Values: 9 Values: <ul style="list-style-type: none"> - True if they signed up via Google: 13,811 users - False if they signed up manually: 13,742 users

Opportunity Sign Up Data

Column Heading	Analysis
Profile ID	<ul style="list-style-type: none"> Unique identifier for the user account on Excelerate. Data Type: Alpha-numeric No missing values 11,481 unique values that may be repeated across multiple columns
Opportunity ID	<ul style="list-style-type: none"> Unique identifiers for the opportunities on the platform. Data Type: Alpha-numeric No missing values 33 unique opportunity IDs that are repeated across multiple tables
Opportunity Name	<ul style="list-style-type: none"> The name of the opportunity the user applied to. Data Type: String No missing values There are 33 opportunity names that are repeated in multiple rows

Opportunity Category	<ul style="list-style-type: none"> • The category of the opportunity the user applied to. • Data Type: String • No missing values • Each category distribution: <ul style="list-style-type: none"> - Event 2007 - Internship 15360 - Engagement 104 - Competition 1121 - Course 1730
Opportunity End Date	<ul style="list-style-type: none"> • This shows the date on which the opportunity ends. • Data Type: Date and Time • No missing values
Gender	<ul style="list-style-type: none"> • Shows the gender of the user. • Data Type: String • 1 missing value • Distribution of each gender fields: <ul style="list-style-type: none"> - Don't want to specify 63 - Males 12240 - Females 8004 - Others 14
City	<ul style="list-style-type: none"> • Shows the city in which the user indicated they live in. • Data Type: String • 1 missing value • This column contains a lot of misspelled and unclear values.
State	<ul style="list-style-type: none"> • Shows the state in which the user indicated they live in. • Data Type: String • 1 missing value • This column contains a lot of misspelled and unclear values.
Country	<ul style="list-style-type: none"> • Shows the country in which the user indicated they live in. • Data Type: String • No missing values • 108 countries repeated in multiple rows
Zip Code	<ul style="list-style-type: none"> • Shows the zip code of the city in which the user indicated they live in. • Data Type: String • This column contains missing and unclear values
Graduation Date	<ul style="list-style-type: none"> • Shows the graduation date of each user. • Data Type: Date • 1 missing value

Current Student Status	<ul style="list-style-type: none"> Shows the academic level of each user. Data Type: String 1 missing value Academic levels are distributed as: <ul style="list-style-type: none"> - Undergraduate Student 7009 - Graduate Program Student 9297 - High School Student 1153 - Not in Education 2862
Current/Intended Major	<ul style="list-style-type: none"> This column shows the major the user is currently pursuing or wants to pursue. Data Type: String 1 missing value This column has misspelled and unclear values
Status Description	<ul style="list-style-type: none"> This column describes the status of all the users' applications right now. Data Type: String It has no missing values There are 8 different statuses and are distributed as: <ul style="list-style-type: none"> - Dropped Out 24 - Applied 89 - Withdraw 622 - Rejected 726 - Started 810 - Not Started 1324 - Rewards Award 2521 - Team Allocated 14206
Apply Date	<ul style="list-style-type: none"> Shows the date on which the user applied for the opportunity. Data Type: Date No missing values
Opportunity Start Date	<ul style="list-style-type: none"> Shows the date each learner started their opportunity. Data Type: Date 804 missing values which indicate the number of users who didn't start the opportunities, so they don't have corresponding dates.

Reward Amount	<ul style="list-style-type: none"> • This shows the reward amounts that are given to each user. • Data Type: Numeric • 17,801 missing values • The number of users who were rewarded each amount are: <ul style="list-style-type: none"> - 50 305 users - 80 6 users - 100 1 users - 120 4 users - 150 1 users - 200 138 users - 500 886 users - 800 1 user - 1000 473 users - 2500 706 users
Badge ID	<ul style="list-style-type: none"> • Shows a unique identifier for each badge • Data Type: Alpha-numeric • 17,801 missing values
Badge Name	<ul style="list-style-type: none"> • Shows the badge name each user is given. • Data Type: String • 17,801 missing values • 56 unique badge names repeated in multiple rows
Skill Points Earned	<ul style="list-style-type: none"> • Shows the number of skill points each user earned. • Data Type: Numeric • 17,801 missing values
Skills Earned	<ul style="list-style-type: none"> • Shows the skills each user accomplished. • Each field may show more than one skill. • Data Type: String • 17,801 missing values

4- Data Cleaning

USER DATA

- Duplicate rows: No duplicate rows were found on the dataset

Reference/Sponsors	Gender	Country	Degree	Sign Up Date	City	Zip	isFromMobile
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male	Nigeria	Undergrad	2023-07-20T04:09:58.800Z	Owerri	400100	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male	India	Undergrad	2023-04-24T09:57:07.405Z	Kottayam	686501	FALSE
["GlobalShala","Winois Institute of Technology","Saint Louis University","Grant Thornton China","Excelerate"]	null	India	null	2022-10-14T19:13:36.303Z			FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Alabama			2023-08-08T12:29:01.772Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	Ghana	Not in Edu	2023-06-15T16:31:42.719Z	Kumasi	AT-1214-9	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	India		2023-07-06T18:49:16.691Z	Chennai	600033	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male			2023-05-15T21:20:04.370Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male	India		2023-07-26T17:01:59.361Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male			2023-07-27T18:02:17.535Z	Lagos	100278	TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male			2023-05-05T04:47:28.446Z	NAS	386570	TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male			2023-05-26T17:05:18.742Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female			2023-05-07T07:06:50.500Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	United States	Graduate	2023-06-13T06:16:48.909Z	Pattola	147001	TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	United States	Graduate	2023-05-01T09:20:26.705Z	SPARROW	21219	TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	United States	Graduate	2023-10-15T01:46:05.107Z	Saint Louis	63043	TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	Ghana	Graduate	2023-06-15T16:22:07.196Z	Accra	233	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male	India		2023-04-05T12:37:13.852Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]				2023-06-15T18:53:01.568Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	Nigeria	Not in Edu	2023-03-03T12:22:28.947Z	Abuja	234	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	India			2023-07-18T14:44:25.551Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Male	India	Graduate	2023-08-28T19:21:02.450Z	Saint Louis	63106	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	India	Undergrad	2023-06-14T12:30:47.640Z	Peshawar	923	TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	India			2023-06-18T16:33:18.751Z			FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	India	Graduate	2023-05-01T09:29:16.814Z	vikarabad	501106	FALSE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	India		2023-03-13T02:03:26.649Z			TRUE
["GlobalShala","Grant Thornton China","Saint Louis University","Winois Institute of Technology","Excelerate"]	Female	Kenya	Undergrad	2023-03-28T20:28:06.290Z	Busia	50400	TRUE

- Categorizing the null and invalid values in City, Gender, Country, Zip Code and all of the categorical columns into “others”/”not specified” to avoid deleting the entire rows.
- Deleting the Zip Code column as it’s full of invalid and incorrect values and won’t be used for either data processing or visualizations.
- Splitting the date and time columns to separate date and time as only the date will be used.

- Ensuring the correct data type is applied to all the columns and that are no null or invalid values present.

USER ID	PreferredSponsors	Gender	Country	Degree	Sign Up Date	City	State
1	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Male	Nigeria	Undergraduate Student	23/07/2023	Dweret	
2	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Male	India	Undergraduate Student	24/04/2023	Kottayam	
3	"GlobalShala", "Illinois Institute of Technology", "Saint Louis University", "Grant Thornton China", "Excelerate"	Others	India	Others	14/10/2022	Others	
4	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	Albania	Not Specified	06/06/2023	Others	
5	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	Ghana	Not in Education	15/06/2023	Kumasi	
6	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	India	Not Specified	06/07/2023	Chennai	
7	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	Nigeria	Not Specified	15/05/2023	Others	
8	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	United States	Not Specified	26/07/2023	Others	
9	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Male	Nigeria	Undergraduate Student	27/07/2023	Lagos	
10	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Male	India	High School Student	05/05/2023	RAS	
11	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	India	Not Specified	26/05/2023	Others	
12	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	India	Not Specified	07/05/2023	Others	
13	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	India	Undergraduate Student	13/06/2023	Patiala	
14	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	United States	Graduate Program Student	01/05/2023	SPARROWS PO	
15	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	United States	Graduate Program Student	15/10/2023	Saint Louis	
16	"Saint Louis University"	Male	Ghana	Graduate Program Student	15/06/2023	Accra	
17	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	India	Not Specified	05/04/2023	Others	
18	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	Pakistan	Not Specified	15/06/2023	Others	
19	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	Nigeria	Not in Education	03/03/2023	Abuja	
20	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	India	Not Specified	19/07/2023	Others	
21	"Saint Louis University"	Male	India	Graduate Program Student	26/08/2023	Saint Louis	
22	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	Pakistan	Undergraduate Student	14/06/2023	Peshawar	
23	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	India	Not Specified	19/06/2023	Others	
24	"Saint Louis University"	Female	India	Graduate Program Student	01/05/2023	vikarabad	
25	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Others	India	Not Specified	13/03/2023	Others	
26	"GlobalShala", "Grant Thornton China", "Saint Louis University", "Illinois Institute of Technology", "Excelerate"	Female	Kenya	Undergraduate Student	26/03/2023	Busia	

OPPORTUNITY SIGN UP AND COMPLETION DATA

- Duplicate rows: No duplicate rows were found on the dataset

Opportunity ID	Opportunity Name	Opportunity Category	Opportunity Start Date	Opportunity End Date	Opportunity Status	Opportunity Type	Opportunity Location	Opportunity Country	Opportunity City	Opportunity State	Opportunity Zip Code	Opportunity Duration (YYYY-MM)	Opportunity Interest Status	Opportunity Interest Message	Opportunity Status
311e44c2-90000000-004f-1	Statement of Purpose (Event)		Jan-05, 2023, 18:58:39		Male	Dhaka	Saint	Bangladesh			1342	2023-06	High School Student	Data Science	Not Star
36814900-00000000-004f-1	Statement of Purpose (Event)		Jan-05, 2023, 18:58:39		Female	Amritsar	Punjab	Alphandstar			123242	2021-01	Undergraduate Student	Otherassass	Reward
81543208-00000000-004f-1	Statement of Purpose (Event)		Jan-05, 2023, 18:58:39		Female	Sanna	Madhya pr	India			485001	2023-04	High School Student	Computer Science and Eng	Not Star
483a3a4d-00000000-004f-1	Statement of Purpose (Event)		Jan-05, 2023, 18:58:39		Male	Hydrabad	Telangana	India			500039	2024-12	Graduate Program Student	Computer Science	Not Star
c28a15f0-00000000-004f-1	Statement of Purpose (Event)		Jan-05, 2023, 18:58:39		Male	Bangalore	Karnataka	India			560015	2022-06	Undergraduate Student	jinhk	Not Star
061380b9-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	Saint Louis Missouri	United States				63108	2024-12	Graduate Program Student	Computer Science and Info	Reward
26398489-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male						252901	2025-01	Graduate Program Student	Aerospace Engineering	Not Star
265a5251-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male						No zip	2019-06	Undergraduate Student	English Language and Litera	Team All
314e6a76-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female						500102	2024-12	Graduate Program Student	Project Management	Reward
481e1a5f-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male						500074	2024-12	Graduate Program Student	Software Engineering	Reward
444e6a76-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male						46609	2022-06	Undergraduate Student	Computer Science and Eng	Team All
5174680f-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male						560070	2025-01	Graduate Program Student	Cyber Security	Reward
57a20005-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male						517501	2022-08	Graduate Program Student	Computer Information Sys	Not Star
9f50a3b0-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	Hydrabad	Telangana	India			500068	2022-05	Graduate Program Student	Computer Science and Info	Reward
79a6a4e3-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	GUNTUR	ANDHRA P	India			524486	2020-07	Graduate Program Student	Information technology	Reward
a61615ab-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	Madanapa Andhra Pri	India				517325	2022-08	Graduate Program Student	Computer Science and Info	Not Star
483a3a4d-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	Hydrabad	Telangana	India			500039	2024-12	Graduate Program Student	Computer Science	Not Star
a58a730a-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	Hydrabad	Telangana	India			500077	2014-12	Graduate Program Student	Computer Science and Info	Reward
4a206f44-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	Multan	Punjab	Pakistan			60650	2022-08	Not in Education	Bsc in Computer Science	Not Star
b6828fbc-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	Krishna	Andhra Pri	India			521190	2024-12	Graduate Program Student	Computer Science	Reward
8a18e74f-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	Hydrabad	Telangana	India			500087	2021-06	Graduate Program Student	Computer Information Sys	Reward
c2a6e77c-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	Nazamabad	Telangana	India			503001	2020-01	Not in Education	Computer Science	Not Star
e7729a57-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	KANINADA ANDHRA P	India				533005	2022-08	Undergraduate Student	Computer Science	Not Star
57350306-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	Narasarao Andhra Pri	India				522901	2024-12	Not in Education	Computer Science and Info	Reward
8b6518fc-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Female	Hydrabad	Telangana	India			500010	2011-08	Graduate Program Student	Project Management	Reward
f8a628fa-b6-00000000-0078-1	Life Beyond Saint Louis Event		Oct-27, 2022, 18:29:00		Male	Vijayawada Andhra Pri	India				521151	2024-12	Graduate Program Student	Computer Science and Info	Not Star

- Categorizing the null and invalid values in City, State, Country and most of the categorical columns into "others"/"not specified" to avoid deleting the entire rows.

- Deleting the Zip Code column as it's full of invalid and incorrect values and won't be used for either data processing or visualizations.
- Replacing the null and blank values for Badge ID, Badge Name and Skill Earned with N/A, and Rewards Earned and Skill Points Earned with 0 as those missing values are for the students who didn't start their opportunities yet.
- Splitting the date and time columns to separate date and time as only the date will be used.
- Ensuring the correct data type is applied to all the columns and that there are no null or invalid values present.

Profile ID	Opportunity Name	Opportunity Category	Opportunity End Date	Gender	City	Country
00004118-8a98-4b64-a47e-6c8d989f5335	00000000-00WQ- Data Visualization	Internship	01/01/2024	Male	Owerri	Nigeria
001c3520-e936-4e4c-1543-51347791e1d9	00000000-00WQ- Data Visualization	Internship	01/01/2024	Male	Lagos	Nigeria
002a7d71-e532-4bdf-8779-bc96a7925792	00000000-00NT-f Digital Marketing	Internship	01/01/2024	Female	Patiala	India
002c6227-035d-4fee-a485-cb51029a3c5b	00000000-00NT-f Digital Marketing	Internship	01/01/2024	Female	Saint Louis	United States
002c6227-035d-4fee-a485-cb51029a3c5b	00000000-00ZY-h Innovation & Entrepreneurship	Internship	01/01/2024	Female	Saint Louis	United States
00311692e-5c72-4e08-8f2f-a74be42138f1	00000000-00WQ- Data Visualization	Internship	01/01/2024	Male	Accra	Ghana
00311692e-5c72-4e08-8f2f-a74be42138f1	00000000-00NT-f Digital Marketing	Internship	01/01/2024	Male	Accra	Ghana
00311692e-5c72-4e08-8f2f-a74be42138f1	00000000-00ZY-h Innovation & Entrepreneurship	Internship	01/01/2024	Male	Accra	Ghana
00311692e-5c72-4e08-8f2f-a74be42138f1	00000000-00HB-h Project Management	Internship	01/01/2024	Male	Accra	Ghana
003a3965-d82b-4827-9675-1eac43b6387e	00000000-10EY-Y Linked Up: The LinkedIn Makeover Workshop	Event	30/01/2024	Female	Abuja	Nigeria
0049929a-793d-4e3a-904e-43447b662817	00000000-00WQ- Data Visualization	Internship	01/01/2024	Male	Saint Louis	India
0049929a-793d-4e3a-904e-43447b662817	00000000-00NT-f Digital Marketing	Internship	01/01/2024	Male	Saint Louis	India
0049929a-793d-4e3a-904e-43447b662817	00000000-00HB-h Project Management	Internship	01/01/2024	Male	Saint Louis	India
004b8a25-174d-4e9a-a41b-b5336327e170	00000000-00N2-j Career Essentials: Getting Started with Your Professional Journey	Course	30/12/2023	Female	Peshawar	Pakistan
00628747-9775-4a85-a179-4d5007cc58b7	00000000-00N2-j Career Essentials: Getting Started with Your Professional Journey	Course	30/12/2023	Female	Vikarabad	India
00628747-9775-4a85-a179-4d5007cc58b7	00000000-00HB-h Project Management	Internship	01/01/2024	Female	Vikarabad	India
006d9e06-ed3c-4d46-beed-96ac9a0374f1	00000000-00WQ- Data Visualization	Internship	01/01/2024	Male	Hyderabad	India
006d9e06-ed3c-4d46-beed-96ac9a0374f1	00000000-00NT-f Digital Marketing	Internship	01/01/2024	Male	Hyderabad	India
006d9e06-ed3c-4d46-beed-96ac9a0374f1	00000000-00HB-h Project Management	Internship	01/01/2024	Male	Hyderabad	India
00728a84-55c8-47cb-942d-61498ba7cd7	00000000-00WQ- Data Visualization	Internship	01/01/2024	Female	San Jose	United States
0078e24-2d05-4cab-a908-09f705894ad	00000000-00WQ- Data Visualization	Internship	01/01/2024	Female	St Louis	India
0078e24-2d05-4cab-a908-09f705894ad	00000000-00NT-f Digital Marketing	Internship	01/01/2024	Female	St Louis	India
0078e24-2d05-4cab-a908-09f705894ad	00000000-00GA-f Health Care Management	Internship	01/01/2024	Female	St Louis	India
0078e24-2d05-4cab-a908-09f705894ad	00000000-00ZY-h Innovation & Entrepreneurship	Internship	01/01/2024	Female	St Louis	India
0078e24-2d05-4cab-a908-09f705894ad	00000000-00HB-h Project Management	Internship	01/01/2024	Female	St Louis	India
0079144d-09d4-4501-a737-2b2dc571616c	00000000-00WQ- Data Visualization	Internship	01/01/2024	Male	Coimbatore	India

5- Profile ID Analysis

- There's no Profile ID column in **User Data**.
- As for the **Opportunity Sign Up Data**, there are 11,481 unique profile IDs in the dataset and some of them are duplicated or repeated across multiple rows as it identifies the ID of each user and many users have applied to many different opportunities.

6- Opportunity Status Description

This column describes the status of all the users' applications right now.

It has no missing values

There are 8 different statuses, and their occurrences are distributed as:

- Dropped Out 24
- Applied 89
- Withdraw 622
- Rejected 726
- Started 810
- Not Started 1324
- Rewards Award 2521
- Team Allocated 14,206

7- Basic Statistics

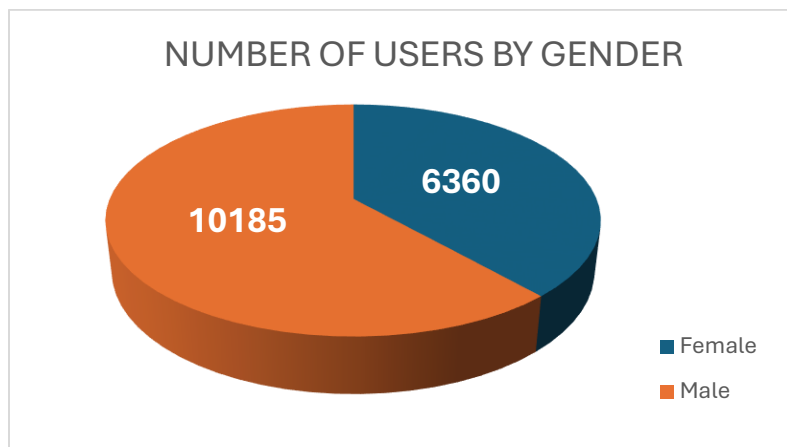
Column	Mean	Median	Mode	Minimum	Maximum	Standard Deviation
Reward Amount	1081.261	200	500	50	2500	926.68
Skill Points Earned	1186.91	1182	1182	10	1776	416.49

8- Initial Observation

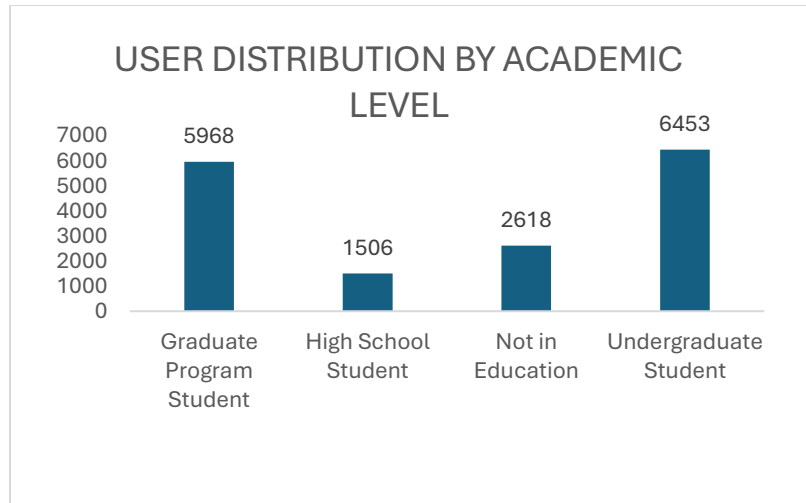
- Reviewing the number of unique profile IDs in the **Opportunity Sign Up Data** which is 11,481 and the total number of 27,562 users from the **User Data** means that a total number of 16,081 users only signed up on the platform and didn't continue to apply to any of the available opportunities.
- The number of male users is significantly higher than the female users.
- There are significantly more applications for the internship category type than the other types.

9- Visualizations

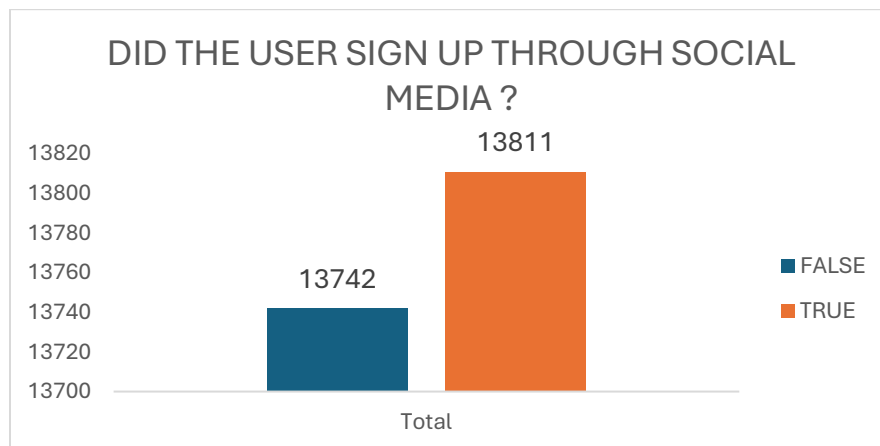
User Data Table Visualizations:



This visualization compares the number of male to the female users who have signed up on Excelerate.

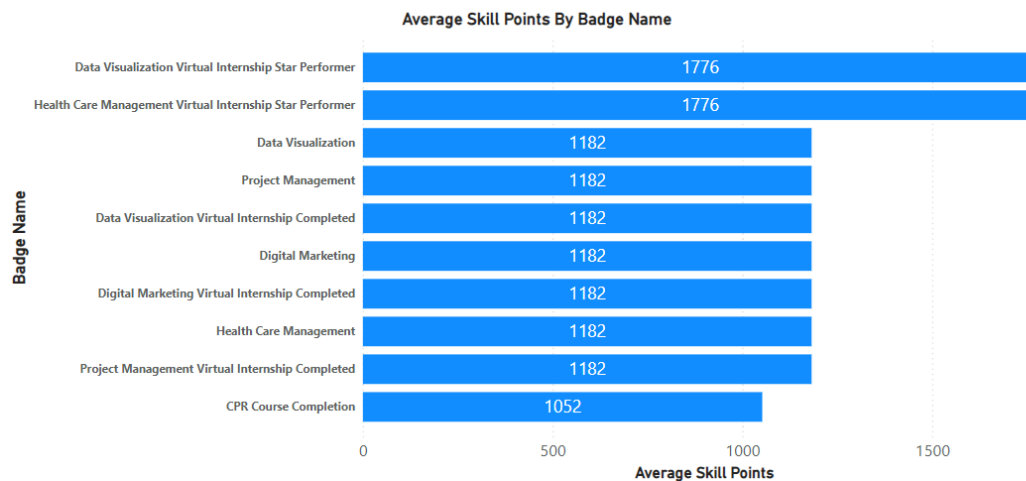


This visualization shows the account users on Excelerate by their educational degree.



This visualization shows the number of users who used a Google account sign up vs the number of users who signed up manually

Opportunity Sign Up Data Visualizations:



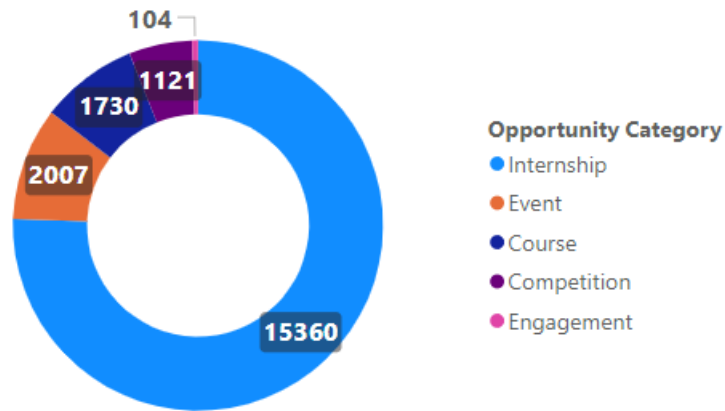
This visualization shows the top average skill points that are earned by badge names.

Average Reward Amounts By Opportunity Names

Opportunity Name	Average of Reward Amount
Career Essentials: Getting Started with Your Professional Journey	200
Changemakers Challenge	63
Cook a Tale	69
CPR/AED Certification	50
Cracking the Interview Code Workshop	50
Data Visualization	1130
Digital Marketing	1494
Digital Palette: A Global T-shirt Design Competition	59
Epidemiology Training Internship	609
Health Care Management	1420
Info Innovators Challenge: An Infographic Design Contest	75
Innovation & Entrepreneurship	1601
Jump Start: Developing your Emotional Intelligence	50
Leadership Launchpad	50
Life Beyond Saint Louis University's Campus	200
Linked Up: The LinkedIn Makeover Workshop	50
Mental Health First Aid Workshop	50
Million Dollar Idea	86
Money Matters: A Personal Finance Workshop	50
Project Management	1467
Resume Booster Workshop	50
Slide Geeks: A Presentation Design Competition	389
Statement of Purpose (SOP) Writing Workshop	200
The Brand Booster Challenge	63

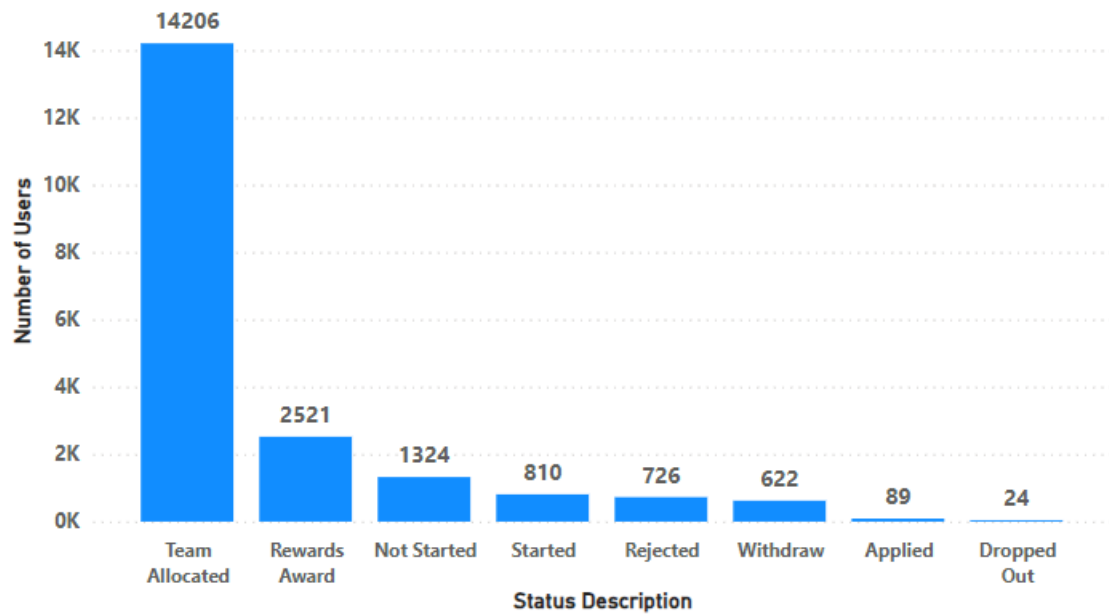
This visualization shows reward amounts for every opportunity the users applied to.

Distribution of Users by Opportunity Category



This shows the number of learners per each opportunity category.

Status Description Distribution



This shows the distribution of users by their opportunity status.

10- Challenges Faced

- Although each entire row in the **User Data** is unique, there's no unique/identifier column entry to make it easier to locate any specific rows.
- The lack of foreign key to connect the 2 datasets to each other.
- PreferredSponsor in **User Data** and Skills Earned in **Opportunity Sign Up Data** column fields contain multiple values that will be easier to analyze if separated in a different table or grouped together by categories.
- City and zip code columns in both datasets contain a lot of errors, misspellings and unclear data that makes it difficult to analyze or draw conclusions about those column values.

11- Next Steps

- Data processing and transformation to ensure the quality and reliability of the given datasets. This could include standardizing data types, applying normalizations, and identifying and handling missing values and outliers.
- Identifying the key questions that need to be answered during the analysis of the given dataset.
- Performing data analysis by identifying potential areas of interest, patterns, trends and data relationships.
- Communicating key findings through the proper choice of visualizations and interactive elements.