DATA VISUALIZATION

Datasets Preprocessing, Transformation Process and Dashboard Wireframe

Documented By:

Team – 3C

Contents:

1. **Data Preprocessing**

2. **Understanding the Requirements of Dashboard**

3. **Dashboard Wireframe**

Welcome to Excelerate's dynamic journey, a visual world where insights emerge and data comes to life. Our goal in creating our Week 2 dashboard wireframe was to turn unprocessed data into a story of potential. Imagine a beautiful fabric of analytics created by a symphony of carefully designed through hand-drawn lines. Our canvas is an opportunity to investigate the beating heart of user involvement rather than merely a tableau. Shows how participation and sign-ups fluctuate in the User Activity Overview, reflecting the pulse of our community's expansion.

Demographic Insights show a diverse picture of our Excelerate family, highlighting the various shades of gender distribution, student status, and majors. Get ready for an exploration through the landscapes of popularity, completion rates, and regional subtleties as you explore Opportunity Metrics. The Comparative Analysis Section serves as a forum for contrasting accomplishments and promoting a constructive environment of rivalry and cooperation. Lastly, take a look at the Skill Development Trends crystal ball to see how our users' talents are changing as they seize opportunities. This wireframe is more than simply a design; it's a sneak peek of the composition we're creating, one in which information flows from vision to data and creativity to insight.

In essence, the wireframe is not just a static representation but a dynamic storyboard that unfolds the narrative of Excelerate's community engagement. It caters to diverse stakeholders, offering them a visual gateway to interpret data trends, make informed decisions, and foster continuous growth and development.

Data Preprocessing:

Steps are:
1. Handling Outliers and Anomalies
2. Normalize or Scale Relevant Features
3. Addressing Data Quality Issues
4. Feature Engineering
5. Data Transformation

Step 1: Handling Outliers and Anomalies

User Data: As there is no numerical or quantitative indicator column in this dataset, so we couldn't apply any statistical methods or visualizations like box plot or histogram to identify outliers.

Though we transform the <mark>Gender Column</mark> 'Don't want to Specify', Blank values & Null attributes set to Unknown and kept rest of the others as it is.

For <mark>Degree Column</mark> Null & Blank values set to Unknown and kept rest of the others as it is.

For missing values in <mark>City & Zip Column</mark>, set them Unknown.

We removed some of the rows from the dataset because of the corresponding Column's values were Blank. In the given dataset we had <mark>total 27562 rows</mark>, after cleaning process we have <mark>total 27491 rows</mark> of that dataset. (<mark>Total Deleted rows 71</mark>)

Opportunity Sign Up and Completion Data:

In this dataset there are two numerical column, Reward Amount and Skill Points Earned. Both of these two columns got value depending on the Status Description column. Those who have Rewards Award in Status Description got the Rewards Amount and have earned Skill Points while others are not.

For Rewards Amount the values are 50,80,100,120,150,200,500,800,1000 and 2500. Based on the standard statistical method (percentile 25 & 75, with determining the lower and upper bound) of identifying outliers using the interquartile range (IQR), the outlier is 2500. Here it is represented in three graphs

| Box Plot | Scatter Plot | Histogram |
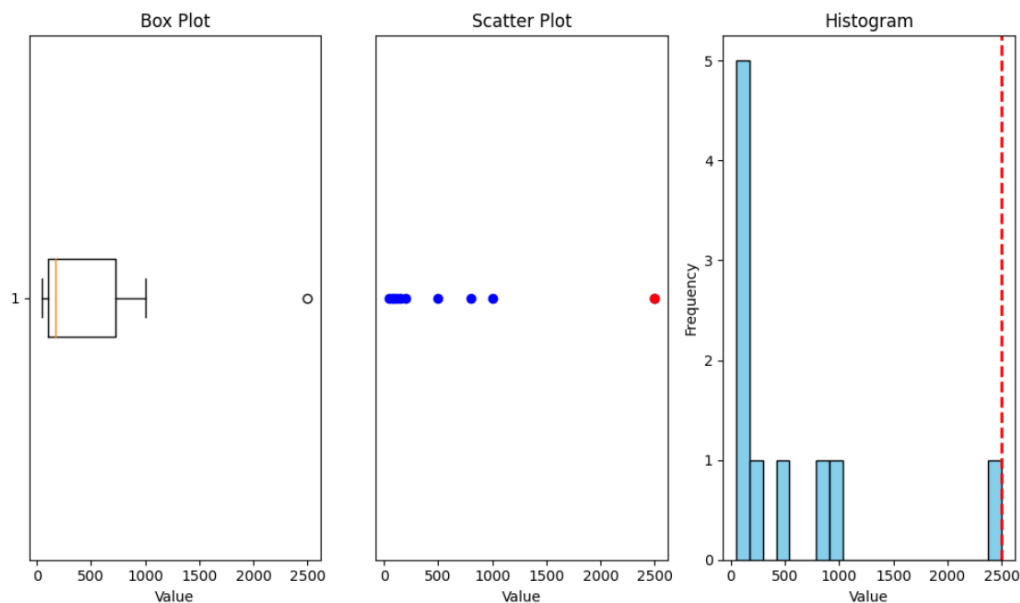
For Skill Points Earned values are 10, 30, 105, 137, 210, 219, 267, 277, 403, 425, 454, 491, 534, 656, 704, 735, 854, 920, 986, 1052, 117, 1182, 1183, 1283, 1497, 1678, 1708, and 1776. Based on the standard statistical method of identifying outliers using the interquartile range (IQR) there is no Outlier detected.

Step 2: Normalize or Scale Relevant Features

User Data: As there is no Numerical Column in this dataset so we didn't apply any normalization here.

Opportunity Sign Up and Completion Data:
We Normalized numerical columns Reward Amount and Skill Points Earned, Min-Max-Difference Scaling to figure out the standardize range of these features.

For Reward Amount

Min	50
Max	2500
Difference	2450

For Skill Points Earned

Min	10
Max	1776

Difference	1766

Also, we append three additional columns at the right side of the dataset named R Amount, S P Earned and Total to further normalize the Reward Amount and Skill Points Earned Column. For R Amount, S P Earned (Value minus minimum divided by the difference) and sum of R Amount, S P Earned represents the Total. Values in Total column which are greater than equal to 1 are represents Reward Amount and Skill Points Earned are relatively high for these two columns.

Step 3: Addressing Data Quality Issues

User Data: There are some missing City in this dataset and we set them Unknown.

Opportunity Sign Up and Completion Data:

For City Column, we replaced the Blank City names with Unknown and for the US city names there were some typos and we fix them (by Google and Tableau)

For Current/Intended Major: We did the following for better visualization,
all Computer Science, Computer Science & Engineering related subjects to Computer Science and Computer & Information Systems and Information Systems to Information Systems and other missing values or subjects set to Other.

For Reward Amount Column: We found those who have the Rewards Amount status in Status Description column has got the Reward Amount, so we replace the Blank Cells of Reward Amount by 'No Reward Amount'.

For badge ID and badge Name: We set Unknown for missing entries.

For Skill Points Earned we set "No Points Earned" and set "No Skills Earned" for Skills Earned column.

Step 4: Feature Engineering

User Data: No new feature is added.

Opportunity Sign Up and Completion Data:

Current Student Status column's attributes converted by the following numerical values for better representation

Attributes	Number
Unknown	1
High School Student	2
Undergraduate Student	3
Graduate Program Student	4
Not in Education	5

Country	Zip Code	Graduation Date(YYYY MM)	Current Student Status	Current Student Status Numeric	Current/Intended Major	Status Description	
Pakistan	60650	2022-08		5	Computer Science	Not Started	Oct
India	503001	2020-01		5	Computer Science	Not Started	Oct
India	522601	2024-12		5	Information Systems	Rewards Award	Oct
India	507111	2020-03		5	Information Systems	Rewards Award	Oct
Kenya	200	2018-10		5	Other	Started	Mar
Nigeria	220101	2020-12		5	Electrical and Electronic Engineering	Started	Mar
Ghana	233	2022-09		5	Statistics	Started	Jun
India	742236	2020-07		5	Data Science	Team Allocated	Jul
Nigeria	240001	2020-08		5	Other	Started	Jun
Nigeria	740102	2022-02		5	Philosophy	Started	Feb
India	400101	2018-12		5	Public Health	Started	May
India	440032	2025-08		5	Electrical and Electronic Engineering	Started	May
Nigeria	100001	2021-09		5	Architecture	Started	Jun
Cameroon	0	2007-06		5	Economics	Started	Jun
India	500036	2021-06		5	Computer Science	Started	Jan
Nigeria	100234	2016-10		5	Pharmacy and Pharmacology	Team Allocated	Sep
India	202001	2016-01		5	Computer Science	Started	Jun
China	723500	2023-06		5	Communication	Team Allocated	Jul
Ghana	233	2017-05	Not in Education	5	Accounting and Finance	Started	Jun
India	431001	2022-09	Not in Education	5	Bioinformatics	Started	May

(Filter dropdown shown over the Current Student Status column: Sort A to Z, Sort Z to A, Sort by Color, Sheet View, Clear Filter From "Current Student S...", Filter by Color, Text Filters, Search box, (Select All), Graduate Program Student, High School Student, Not in Education (checked), Undergraduate Student, Unknown, OK / Cancel)

Opportunity Completion Rate (%) Column for each of the opportunity (those who completed and those who started).

Life Beyond Saint Louis University's Campus	0.48%	Rewards Award	Oct 26, 2022, 09:22:38	Oct 28, 2022, 13:30:00
Life Beyond Saint Louis University's Campus	0.48%	Rewards Award	Oct 26, 2022, 11:33:39	Oct 28, 2022, 13:30:00
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Jun 14, 2023, 23:54:04	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Apr 23, 2023, 06:37:09	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Jun 15, 2023, 16:32:47	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Mar 29, 2023, 21:40:23	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Feb 24, 2023, 08:02:50	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Apr 18, 2023, 09:54:28	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Jun 17, 2023, 07:49:38	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Jun 19, 2023, 12:41:36	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Jun 16, 2023, 11:10:23	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Mar 28, 2023, 08:50:56	Nov 03, 2022, 18:30:39
Career Essentials: Getting Started with Your Professional Journey	4.48%	Rewards Award	Jul 05, 2023, 15:42:54	Nov 03, 2022, 18:30:39
Slide Geeks: A Presentation Design Competition	0.36%	Rewards Award	Dec 26, 2022, 11:09:40	Jan 05, 2023, 18:52:39
Digital Marketing	10.23%	Rewards Award	Apr 24, 2023, 02:17:28	May 08, 2023, 02:30:00
Digital Marketing	10.23%	Rewards Award	Jul 02, 2023, 23:14:32	Jul 24, 2023, 02:30:00
Digital Marketing	10.23%	Rewards Award	Jul 03, 2023, 16:28:29	Jul 24, 2023, 02:30:00
Digital Marketing	10.23%	Rewards Award	May 06, 2023, 14:39:39	Jun 12, 2023, 02:30:00
Digital Marketing	10.23%	Rewards Award	Jun 15, 2023, 19:55:39	Jul 24, 2023, 02:30:00
Digital Marketing	10.23%	Rewards Award	Aug 03, 2023, 19:47:05	Aug 21, 2023, 04:30:00

Understanding the Requirements of Dashboard:

1. **Platform Activity:**
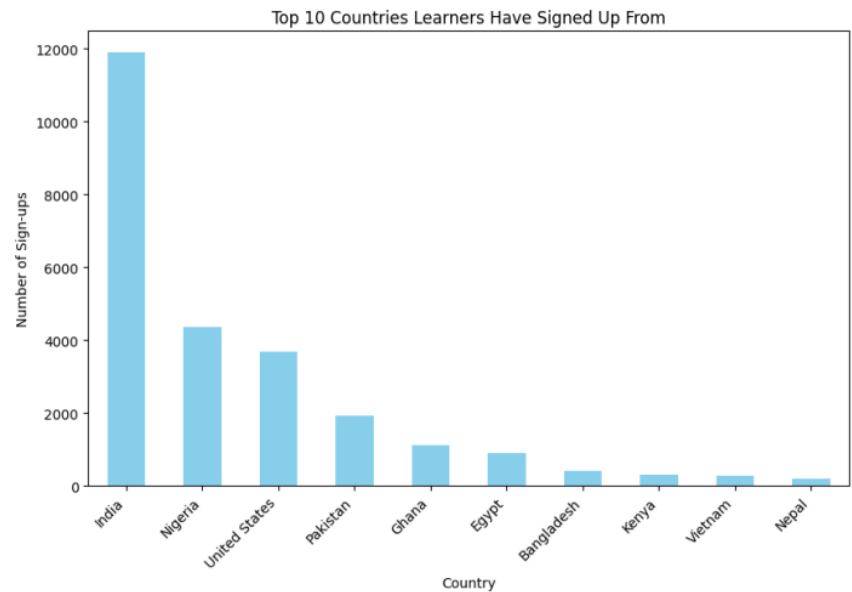   Total 27,491 People signed up in the platform

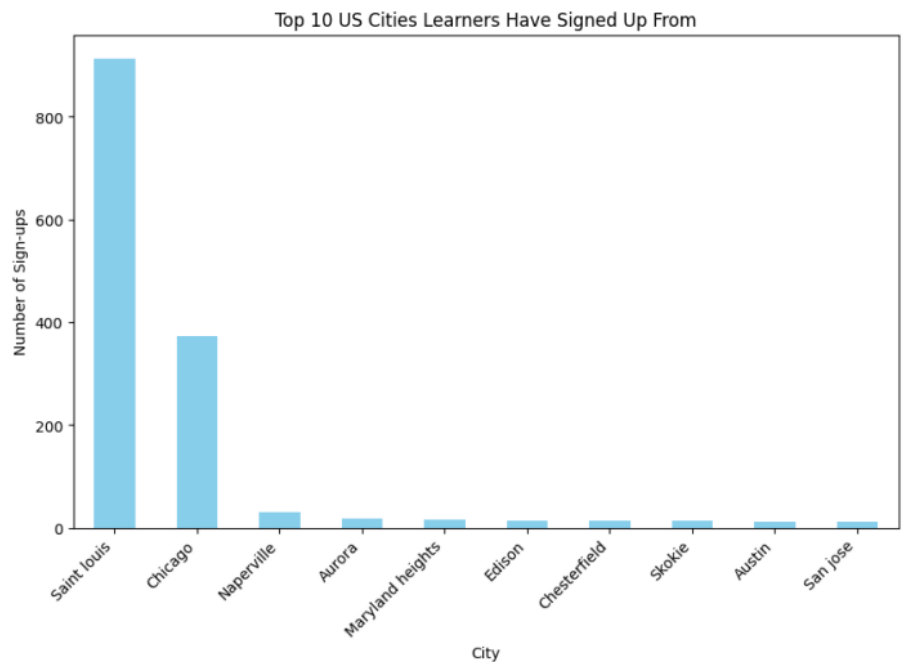   Signed up for Opportunity 11,481
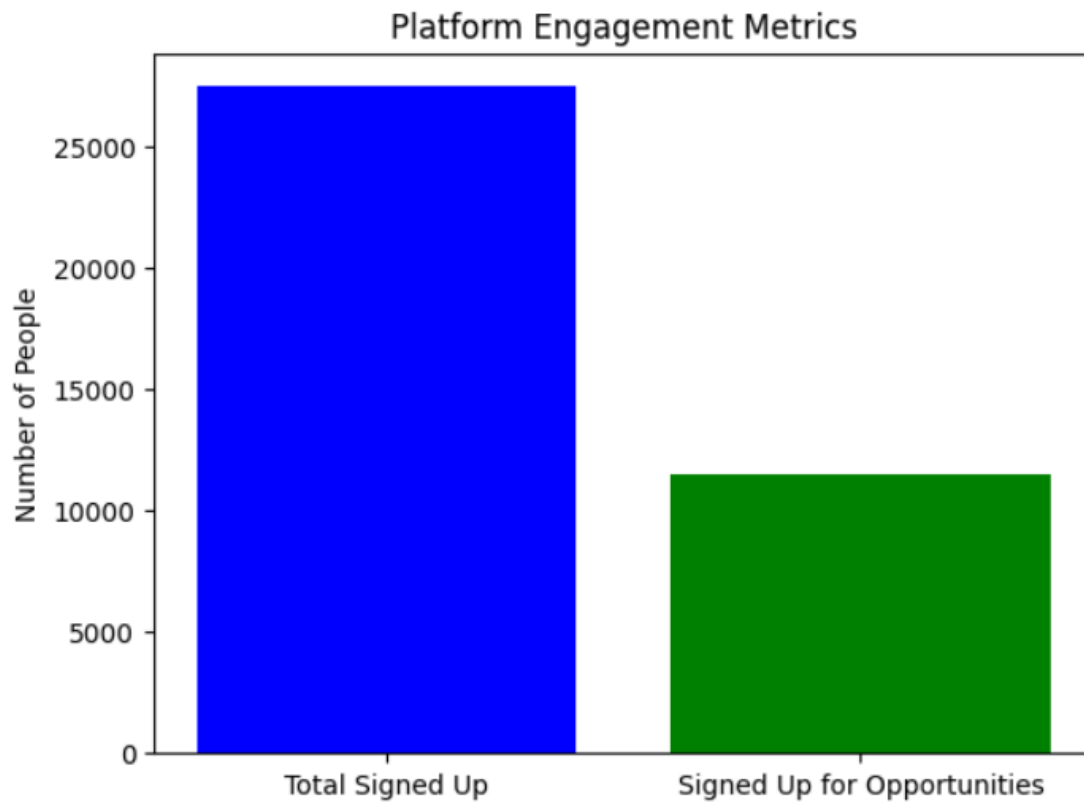
2. Global Reach:

Here are Top 10 Countries Learners have signed up from

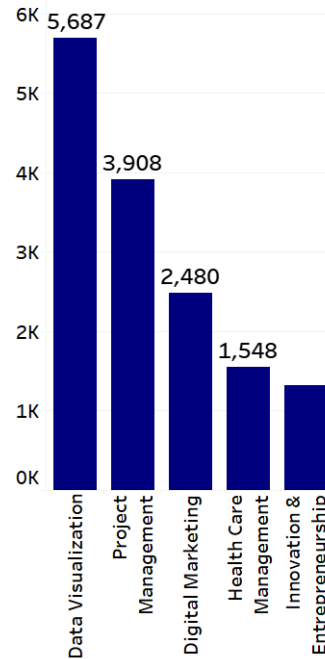Top 10 Countries Learners Have Signed Up From

3. US City Insights:
List of Cities in the US, from where learners have signed up

Top 10 US Cities Learners Have Signed Up From

4. Opportunity Popularity:

The most popular opportunities learners have signed up for

4. Opportunity Popularity

5. Completion Trends:

Here are the top 5 most popular opportunities learners have completed

5. Completion Trends
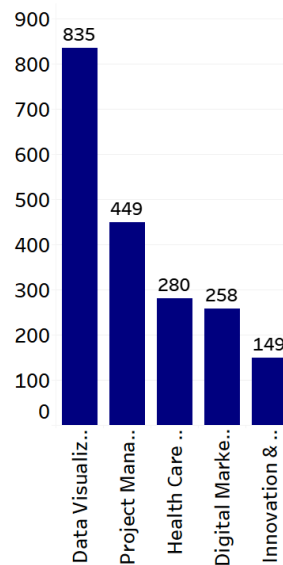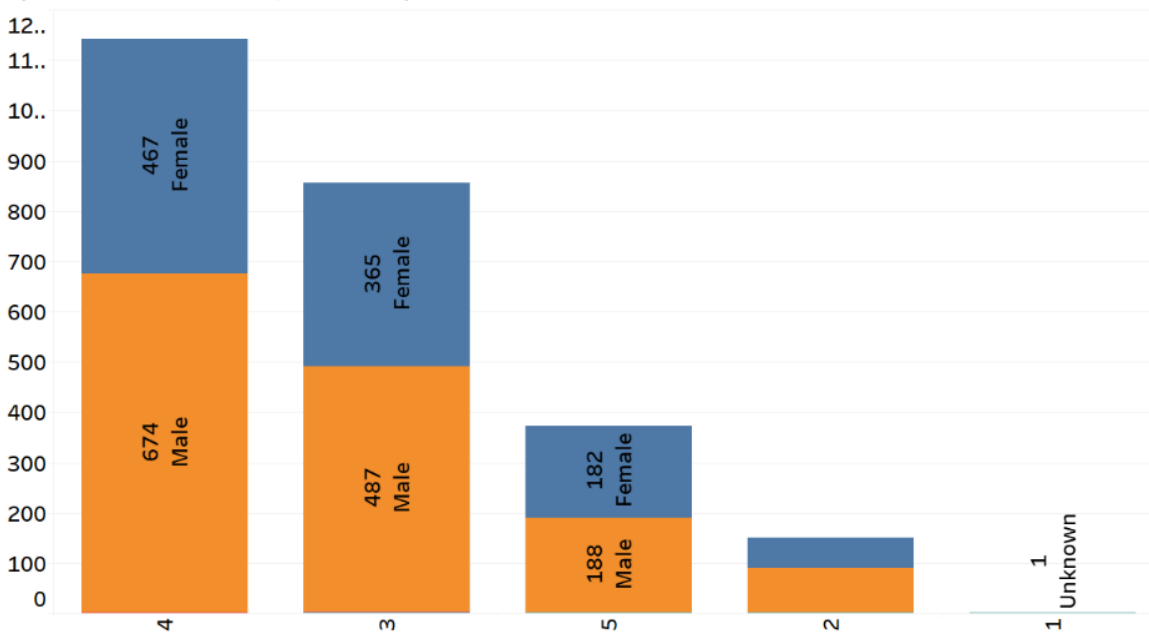
6. Demographic Analysis:

Demographic breakdown of Gender of those who have signed up in the platform
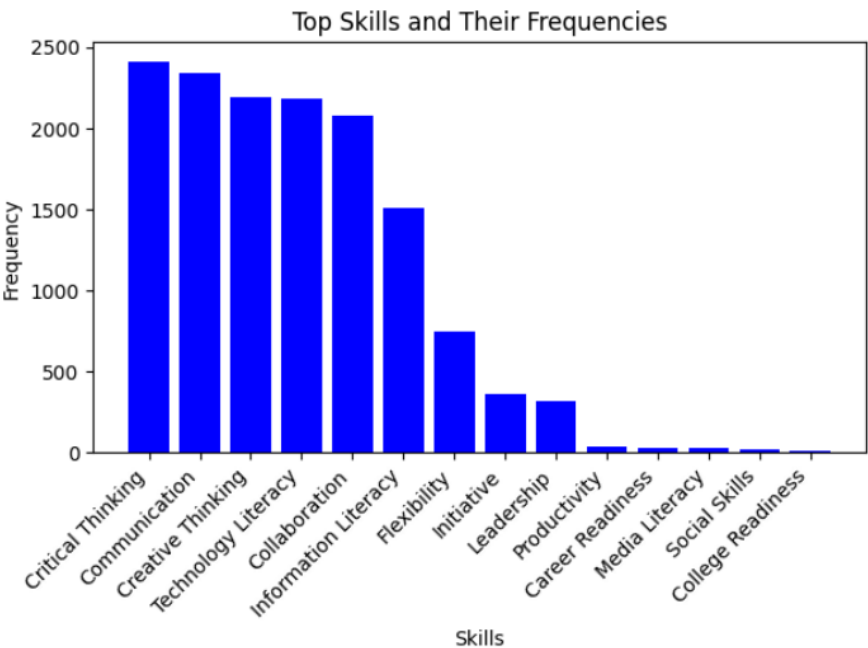
6. Demographic Breakdown (Signed up Gender)
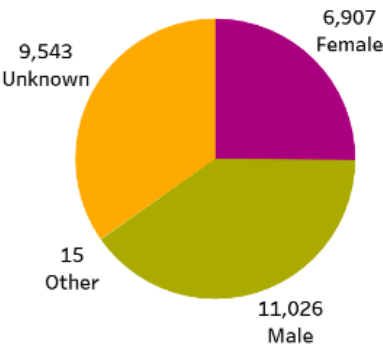
Demographic breakdown of Student Status by Gender those who have completed.

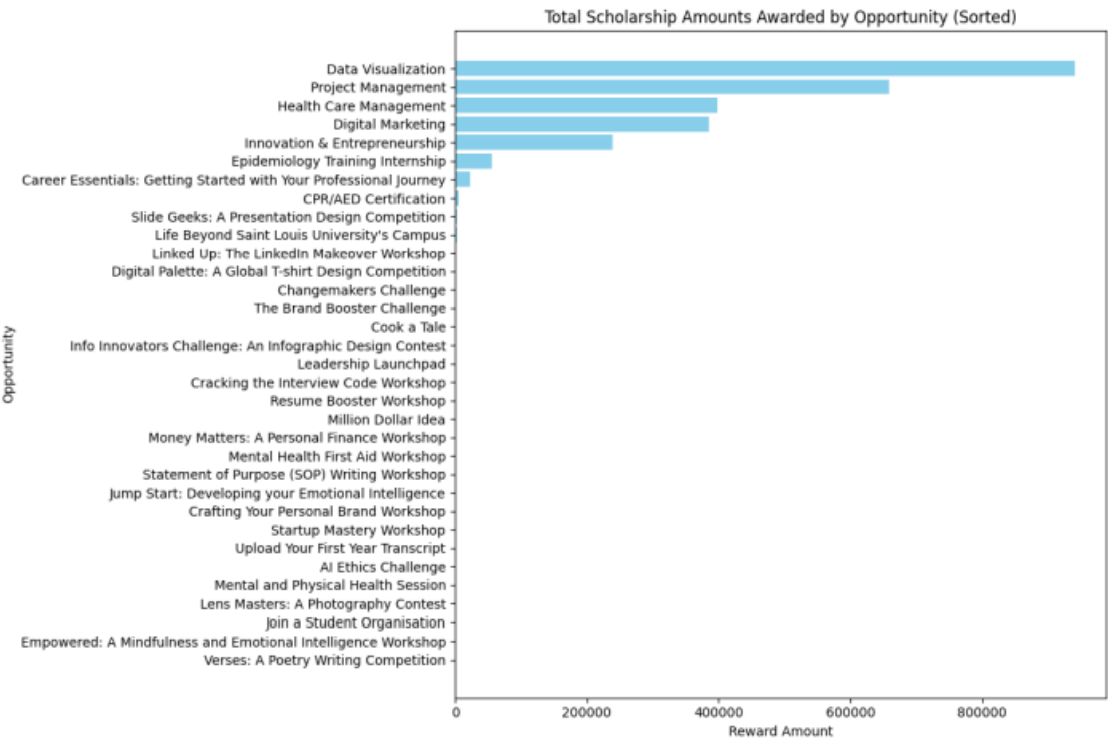**6. Demographic breakdown of (Student Status by Gender)**

7. Skill Development:

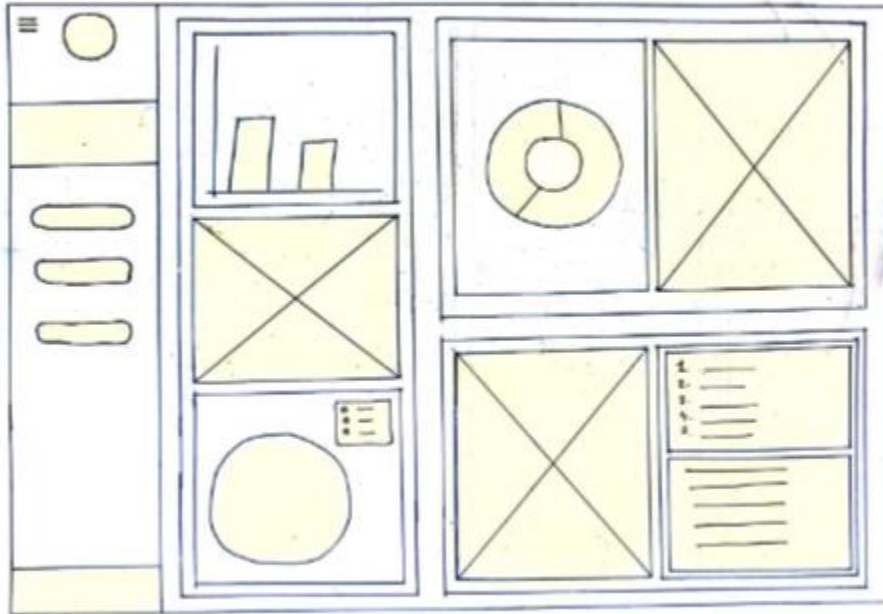Here are the most gained skills on Excelerate Platform

8. Scholarship Impact:

Total Scholarship amount **2,725,860** awarded on the following opportunities

Dashboard Wireframe:

Below is the hand-drawn dashboard wireframe:

User Activity Overview:

Time-Based Trends: Capture the pulse of user engagement over time, showcasing spikes in sign-ups and peaks in participation. A bar chart could elegantly narrate this story, emphasizing growth patterns and identifying pivotal moments.

Participation Metrics: Highlight key metrics such as total sign-ups, active users, and participation rates to provide a comprehensive overview of user engagement.

Demographic Insights:

Pie Charts for Gender Distribution: Using visually appealing pie charts to represent the distribution of users based on gender, providing an at-a-glance understanding of the community's composition.

Bar Charts for Student Status and Majors: We utilized bar charts to illustrate the distribution of students versus non-students and showcase the prevalent majors within the user base.

Opportunity Metrics:

Bar Graph for Popular Opportunities: Display the most popular opportunities through a bar graph, allowing stakeholders to identify trends and focus on activities that resonate most with the community.

Geographic Heatmap: Illuminate geographical insights using a heatmap, visualizing the regional impact of opportunities and helping tailor future initiatives to specific locations.

Comparative Analysis Section:

Stacked Bar Charts for Completion Rates: Construct stacked bar charts to compare completion rates across different opportunities, enabling a nuanced understanding of success factors.

Side-by-Side Comparisons: Incorporate side-by-side visualizations to juxtapose demographic groups, allowing for insightful comparisons in completion rates and participation levels.

Skill Development Trends:

Line Graphs for Skill Acquisition: Illustrate skill development trends through line graphs, showcasing the evolution of skills acquired by users over time as they engage with various opportunities.

Bubble Charts for Skill Popularity: Implement bubble charts to highlight the popularity of specific skills within the community, providing a visual hierarchy of skill acquisition trends.

Conclusion:

Excelerate's transformative journey unfolds in a symphony of data preparation, requirement comprehension, and dashboard wireframing. The basis was laid by the careful preparation of the data, which guaranteed the accuracy and consistency of our conclusions. Understanding the intricate requirements of our dynamic dashboard was like trying to harmonize opportunity measurements, demographic information, user engagement, comparison analysis, and skill development trends with a complex symphony. Our hand-drawn wireframe, a visual masterpiece that captured the spirit of Excelerate's vibrant community, emerged as the climax of this orchestration. This week deliverable has demonstrated the ability to combine technical accuracy with artistic intuition, giving stakeholders the clarity they need to traverse the symphony of data. As we reveal the conclusion of this chapter, we start a new movement in which innovation takes center stage in Excelerate's never-ending pursuit of excellence and insights transcend knowledge.