



Data Visualization Associate Early Internship

*Sponsored by **EXCELERATE***

Presented by Sub Team: 34A



Agenda

1. Introduction
2. Team Charter
3. Exploratory Data Analysis
4. Dashboard Overview
5. Key Decisions and Design Choices
6. Challenges Faced
7. Solutions Implemented
8. Insights Derived
8. Addressing Key Questions
9. User Interaction and Guidance
10. Visual Highlights
11. Impact on Decision-Making
12. Recommendations
13. Conclusion

Presentation Team 34A



Introduction

1. Aishwarya Kore aishwaryakore81@gmail.com
2. Uchenna Okolo 2ucheokoll@gmail.com
3. Prajwal Ghadage prajwaldg4@gmail.com
4. Sangam Chettri chettrisangam356@gmail.com
5. Pearl Rubyth Thomas R pearlrubyththomas@gmail.com
6. Vani Bajaj vanibajaj0001@gmail.com
7. Abhinay Matam mabhi4184@gmail.com

EXPLORATORY DATA ANALYSIS (EDA)

- The exploratory Data Analysis report presents a comprehensive overview and summarization of the User Data and Opportunity Wise datasets used in building the dashboard. It highlights key features and uncovers patterns, detect anomalies and areas that require further cleaning, as well as check assumptions using statistical and graphical methods. The outcomes set the ground work for building a compelling and an interactive dashboard that provides invaluable insights to support strategic decision-making.

- **EDA summary of the User-Data reveals the following:**

- Total Rows: 27,562
- Total Columns: 8
- Total Missing Values: 39,496
- Data Type: A mix of categorical, and datetime types
- Sign up Date is in the object dtype format instead of datetime format

Summary of Missing Values

Coulum	Missing Value Count	Missing Value %
Gender	9,535	24.14
Country	62	0.16
Degree	10,612	27.40
City	9,534	24.14
zip	9,544	24.14
From Social Media	9	0.02





TEAM CHARTER

Our Team Charter showcases:

- **The team roles & responsibilities**
- **The Mission Statement:** To deliver actionable insights and a thorough project plan that meets the needs of our stakeholders, ensuring the successful completion of the project.
- **The Vision:** To cultivate a collaborative and results-driven team environment that supports both project success and the professional growth of every team member.
- **Objectives:** Ensure Data Quality, Perform an in-depth analysis, develop interactive dashboard that would provide actionable insight to support the sponsor's strategic decision making.
- **Core Values:** Integrity, Accountability, Teamwork, Respect, Innovation, and Discipline.
- **The Internal Checks, Balances, and Reviews:** This covers everything about assignments, meetings, communication guidelines, status updates and deadlines for deliverables
- **Project Time lines & Task Assignments**

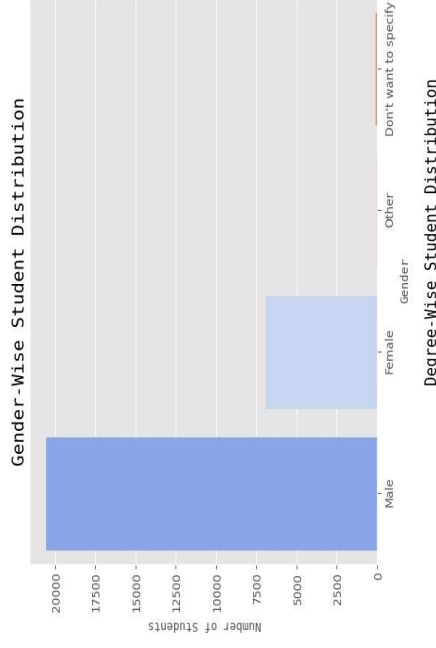


User Data: Data Cleaning and Missing Value Handling Procedure

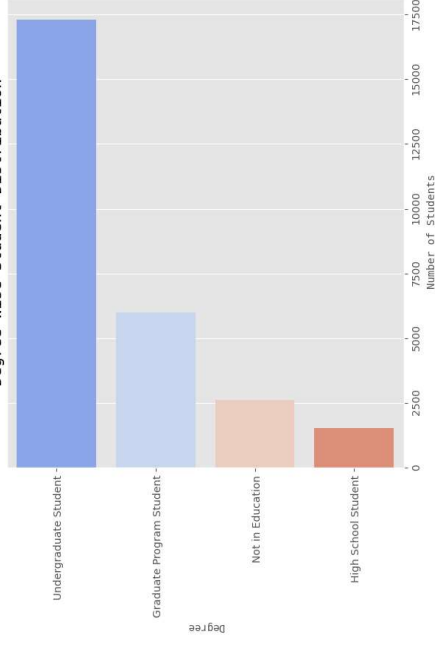
- The team reviewed its earlier approach to handling of the User Data missing values and incorporated the feedback from Excelerate Management Team.
- Imputation strategy was then used in replacing the missing values.
- 1. Gender Column: The most common gender “Male” was used to fill the missing values in the gender column
- 2. Country column: Missing data points were filled with “India”, it being the mode in the column.
- 3. Degree column: Missing values were replaced with “undergraduate degree’ because of its highest count in the column.
- 4. City Column: Missing values were replaced with the most prevalent city, Hyderabad
- 5. Zip column: Missing data points in the zip column were replaced with a placeholder “unknown”
- 6. Is From Social Media column: “True” was used to fill up the missing data points

Visual Insights from the User Data Exploration:

Gender Distribution: Male users are the most represented genders and followed closely by Female. Non datapoints are filled up by Male.

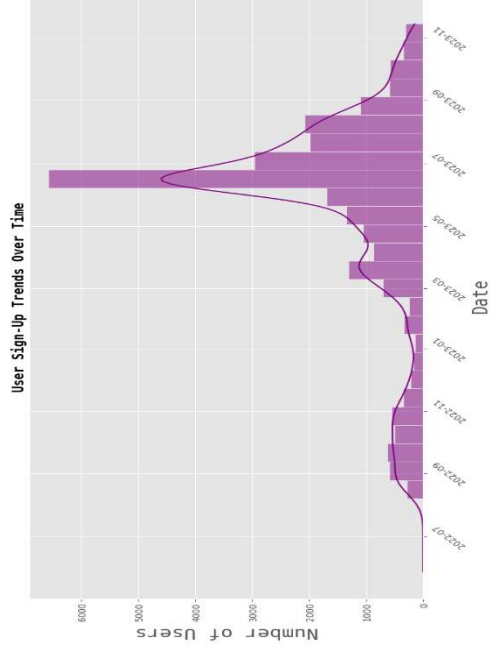


Degree Distribution: Most users are undergraduate students, accounting for 39% of the total, with graduate program students following closely at 36%.

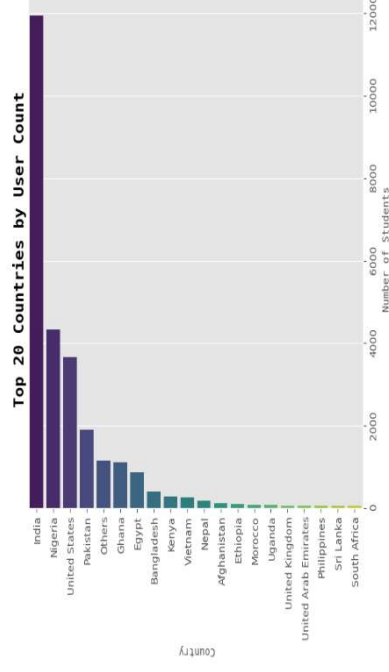


Visual Insights from the User Data Exploration:

Sign-up Trend: The peak sign-up occurred in the month of June 2023



Country Representation: Dominated by users from 'India' (>94%), and followed by Nigeria and United States in the second and third places respectively. Other countries have sparse representation (Sri Lanka, South Africa, etc.).



Opportunity Sign-Ups and Completion Data

- EDA of the Opportunity Sign-Ups and completion dataset reveals the dataset has:

- Total Rows: 20,322
- Total Columns: 21
- Total Missing Values count: 89,884
- Data Type: A mix of categorical, numerical and datetime types
- Opportunity Start Date, Graduation Date (YYYY MM), Apply Date, and Opportunity End Date are all in the object dtype format instead of datetime format.

Summary of Columns with Missing Values

Column	Missing Value Count	Missing Value %
Gender	1	<(0.01)
City	1	<(0.01)
State	14	0.02
Zip Code	13	0.01
Graduation Date(YYYY MM)	1	<(0.01)
Current Student Status	1	<(0.01)
Current/Intended Major	44	0.05
Opportunity Start Date	804	0.894
Reward Amount	17,801	19.8
Badge Id	17,801	19.8
Badge Name	17,801	19.8
Skills Points Earned	17,801	19.8
Skills Earned	17,801	19.8



Data Cleaning and Handling of Missing Values

Column	Error correction	Missing Values Replaced with
Profile Id/Opportunity Id	Leading or trailing spaces removed	NA
Gender	Standardized	Unknown
Opportunity End Date/Opportunity Start Date, Apply Date and Graduation Date		Error correction applied
Opportunity Name and Opportunity Category	Converted date format to Datetime	Unknown
City Name, State Name and Country Name	Opportunity and Category Names were Capitalized and typos fixed	Unknown
Zip Code	Names were capitalized and typos fixed	Unknown
Current Student Status	Missing zip codes replaced with a placeholder "unknown"	Unknown
Current/Intended Major	Capitalized Status Name	Unknown
Reward Amount	Capitalized Major Name	Undecided
Badge Id	Convert to Numeric and error correction applied to fix NaN introduced	0
Badge Name	Ensures Ids are unique identifiers	0
Skills Points Earned	Capitalized	Unknown
Skills Earned	Converted to Numeric	0
	Capitalized Skill Name	None

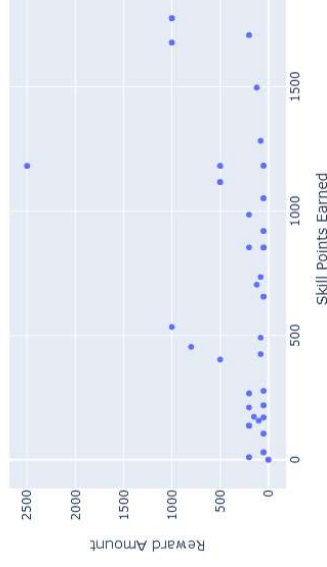
Visuals Insights from the opportunity Sign-Ups and Completion dataset

Scatter Plot of Reward Amounts versus the Skill Points Earned:

The scatter plot visualization shows that most users or participants earned reward amounts of 500 or less. The highest reward amount earned skill points of 1182 while skill points of 1776 only earned 1000 in reward amount.



Scatter Plot of Reward Amount vs Skill Points Earned

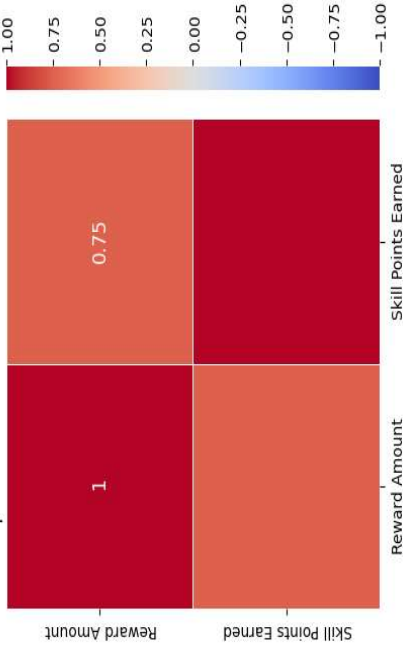


Heatmap showing correlation between reward amount and skill points earned.

The dark red color with a correlation coefficient of 1 shows a very strong positive correlation between the Rewards Amount and the Skill Points Earned, indicating Skill Points Earned increases proportionally as the Reward Amount increases. At the opposite end of the spectrum, the dark blue, the value of -1 shows a strong negative correlation between the two variables, depicting a proportionally inverse relationship between the two variables.. The values nears the value 'zero' with lighter shades of colors suggest little or no correlation between the Skill Points Earned and the Reward Amount.



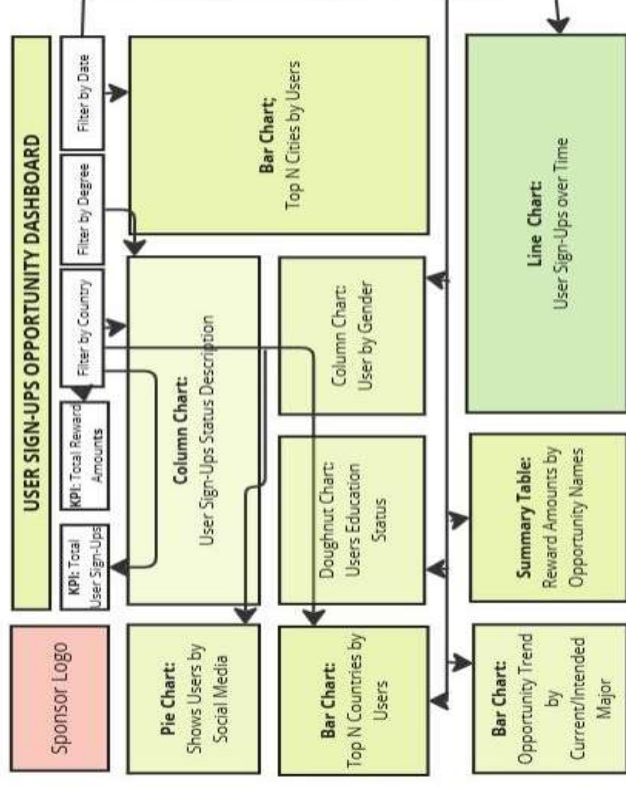
Correlation Heatmap between Reward Amount and Skill Points Earned



DASHBOARD WIREFRAME

The dashboard wireframe depicts a low-fidelity representation of the dashboard, showing how different chart elements are interconnected. Every chart element in the frame is selected to address a specific key question and provide actionable insights to aid strategic decision making

USER SIGN-UPS OPPORTUNITY DASHBOARD - WIREFRAME



Every filter will affect each chart element but a few relationships are depicted for brevity

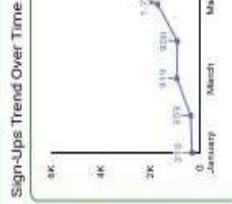
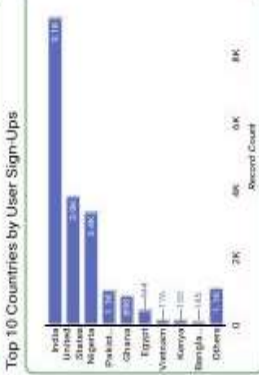
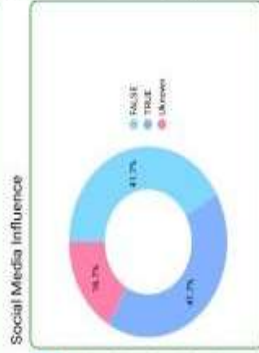
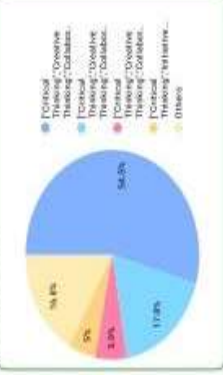
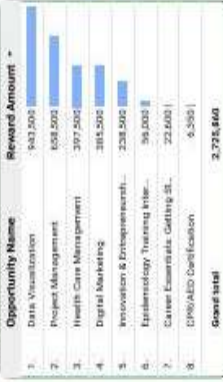
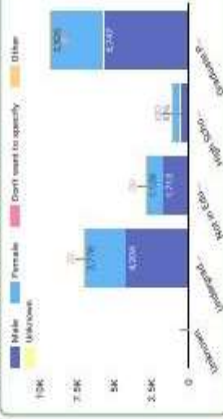
...xcelerate

Select date range -

Degree Major	Country
•	•

	20.3K	27.6K
20.3K		
27.6K		

2.7M





Key Decisions and Design Choices

1

Separate Sections dedicated to Platform Signups, Opportunity Signups, Opportunity Completed.

2

Use of Bar charts for visualizing ranked records within the dashboard.

3

Use of Pie charts to communicate demographic insights within the dashboard.

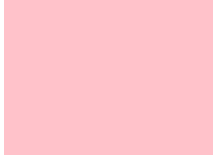
4

Use of tables to display top records.

Challenges Faced



- Inconsistent values in the city column (e.g., “Saint Louis”, “St Louis”, “SAINT LOUIS”, “st louis”).
- A high number of missing values in the Gender column.
- The presence of "Other" in the City column.
- Difficulty in extracting and analyzing the top skills efficiently.
- The Skills column contains entries in a list format, which makes analysis challenging.
- A high number of missing values in several opportunity-related columns.



Solutions Implemented

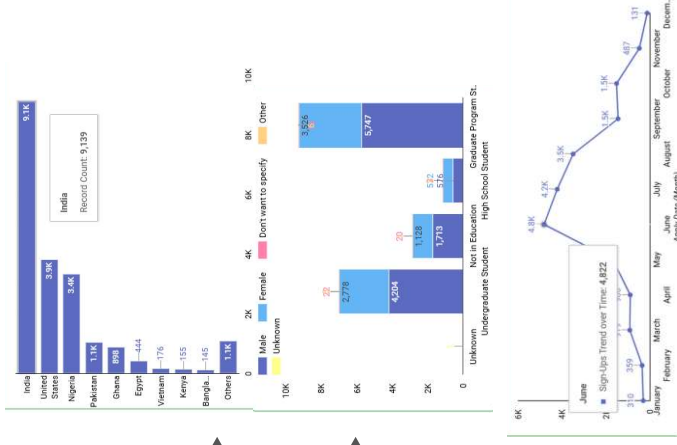


- The implementation of data standardization and normalization techniques resulted in a clean and consistent City column.
- To address the issue of null values, we applied advanced imputation methods.
- The City column contained a significant number of entries labeled as "Other," which were excluded from the insights analysis.
- For the dedicated skills dataset, we utilized normalization techniques to ensure consistency and standardization of skill entries.
- To handle the challenge of multiple opportunity-related columns with a high number of null values, we first separated the dataset and then applied a selective deletion strategy.



Insights Derived

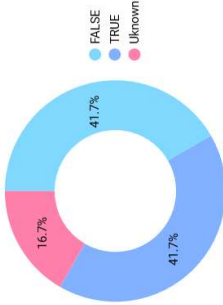
- India has the highest number of sign-ups (9.1K), followed by the USA (3.9K), indicating strong adoption in these regions.
- More males have signed up compared to females.
- The highest sign-ups occurred between June and July, peaking in June (~4.8K sign-ups), indicating seasonal trends in user engagement.



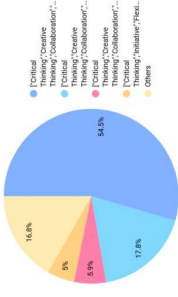


Insights Derived

- 41.7% of users were influenced by social media in signing up, showing the platform's dependency on social media for outreach.



- The most earned skills are Critical Thinking, Problem-Solving, and Data Analysis, aligning with the high demand for analytical and cognitive skills in today's job market.



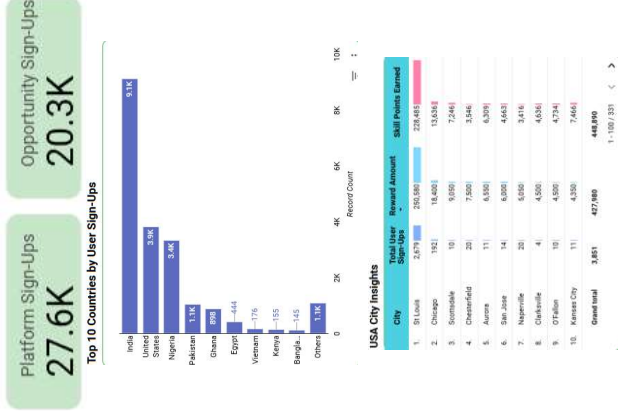
- Data Visualization Opportunity Leads with the Highest Rewards Awarded i.e 943,500 , Contributing to a Total of 2,725,860 Across All Opportunities.

Scholarship Insights		
	Opportunity Name	Reward Amount
1.	Data Visualization	943,500
2.	Project Management	658,500
Grand total		2,725,860



Addressing Key Questions

1. How many people are signed up on the platform, and how many of those have signed up for opportunities?
2. What are the top 10 countries learners have signed up from?
3. What are the cities in the US learners have signed up from?

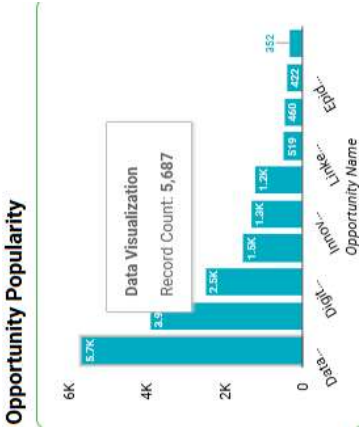




Addressing Key Questions

- 1. Which is the most popular opportunity learners have signed up for?
- 2. Which is the most popular opportunity learners have completed?

	Opportunity Name	Reward Amount
1.	Data Visualization	943,500
2.	Project Management	658,500
3.	Health Care Management	397,500
4.	Digital Marketing	385,500
5.	Innovation & Entrepreneurship	238,500
6.	Epidemiology Training Intern...	56,000
7.	Career Essentials: Getting St...	22,600
8.	CPR/AED Certification	6,350
9.	Slide Geeks: A Presentation ...	3,500
10.	Life Beyond Saint Louis Univ...	2,400
	Grand total	2,725,860



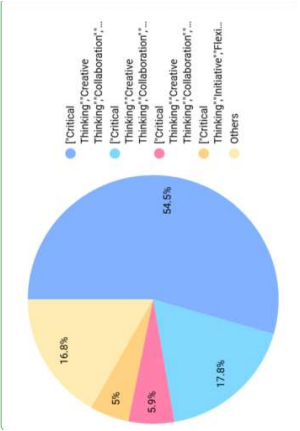


Addressing Key Questions



- 1. What are the most gained skills on Excelerate?
- 2. How much is the total scholarship awarded and through which opportunities?

Most Earned Skills



Scholarship Insights

	Opportunity Name	Reward Amount
1.	Data Visualization	943,500
2.	Project Management	658,500
3.	Health Care Management	397,500
4.	Digital Marketing	385,500
5.	Innovation & Entrepreneurship	238,500
6.	Epidemiology Training Intern...	56,000
7.	Career Essentials: Getting St...	22,600
8.	CPR/AED Certification	6,350
9.	Slide Geeks: A Presentation ...	3,500
10.	Life Beyond Saint Louis Univ...	2,400
	Grand total	2,725,860



User Interaction and Guidance



- The dashboard incorporates an intuitive navigation system, allowing users to seamlessly explore different sections and insights.
- Visualizations within the dashboard are designed to be interactive, enabling users to drill down into specific data points for a more detailed analysis.
- Robust filtering options are integrated, allowing users to customize their view based on specific criteria.
- To foster continuous improvement, a user feedback mechanism is embedded within the dashboard.



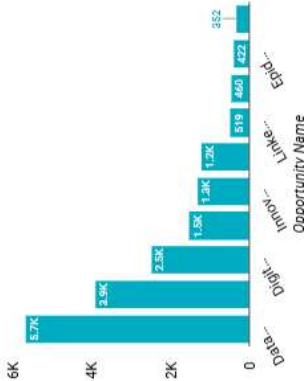


Visual Highlights

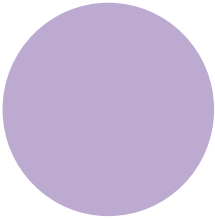
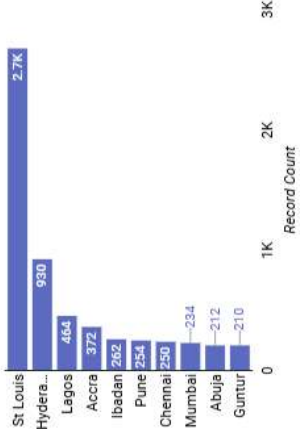
Scorecard



Column Chart



Bar Chart





Visual Highlights

Bar Chart

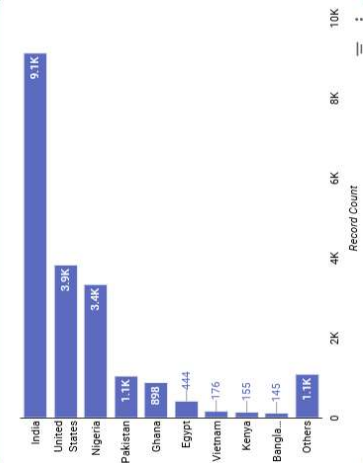
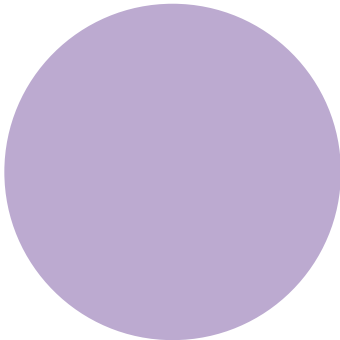


Table With Heatmap

	Opportunity Name	Reward Amount
1.	Data Visualization	943,500
2.	Project Management	658,500
3.	Health Care Management	397,500
4.	Digital Marketing	385,500
5.	Innovation & Entrepreneurship	238,500
6.	Epidemiology Training Intern...	56,000
7.	Career Essentials: Getting St...	22,600
8.	CPR/AED Certification	6,350
9.	Slide Geeks: A Presentation ...	3,500
10.	Life Beyond Saint Louis Univ...	2,400
	Grand total	2,725,960





Visual Highlights

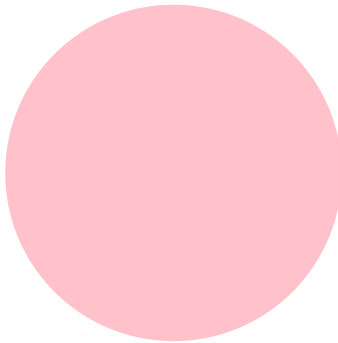
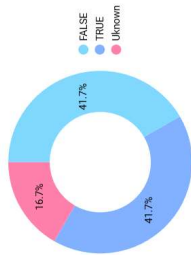
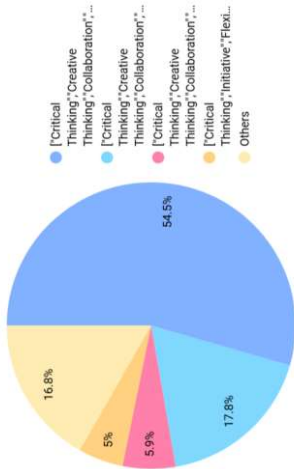


Pie Chart

Doughnut Chart

Most Earned Skills

Social Media Influence



Impact on Decision-Making



1

The dashboard's metrics facilitate data-driven evaluations of program effectiveness.

2

The insights derived from the dashboard allow for efficient management and optimization of internship programs.

3

Understanding skill acquisition trends enables Excelerate to adapt its skill development initiatives proactively.



RECOMMENDATIONS:

1. To ensure balanced geographical reach, the management team should implement targeted marketing campaign to enhance presence and credibility in low-engagement regions.
2. To address the varying popularity of different programs, the sponsor might consider
 - allocating additional resources to popular program to further strengthen their appeal.
 - revamping underperforming programs through modification to align them more closely with the current market needs and student interest.
 - increasing promotional efforts by enhancing marketing and outreach for underperforming programs to raise awareness and attract potential students
3. The platform sign-ups seem slightly skewed toward the male gender. There may be a need to adjust the outreach strategy to create a more balanced inclusive environment, attracting a diverse user base and enhancing overall engagement.
4. Some opportunities received higher scholarship amounts than others. To ensure a fair distribution of scholarships that supports a diverse range of programs, it is recommended that the sponsor assess its funding allocation across various initiatives to ensure alignment with organizational goals and desired outcomes.
5. Given the seasonal spike in the user sign-ups over time, it is important to:
 - optimize program reach by targeting high-performing regions and majors.
 - analyze dropout rates to refine the user journey.
 - improve outreach strategies for gender balance and social media effectiveness.
 - ensure equitable scholarship distribution to enhance accessibility.
 - plan promotions strategically to coincide with peak engagement times.

CONCLUSION:

The exploration and in-depth analysis of the two datasets presented for this program provide invaluable insights into the platform engagement, key factors potentially influencing user sign-ups for opportunities, and overall participation trends. The analysis reveals a platform sign-ups conversion rate of approximately 41.7%, indicating that a significant proportion of users transitioned from registration to active participations in opportunities. By leveraging these insights and the recommendations outlined in the team's reports, the Excelerate management team can make informed decisions to enhance user engagement, optimize resource allocation, and improve the overall effectiveness of its programs.



Thank You



Regards: Team 34A