Week-2 Report

NAME : Syed Hur Abbas Naqvi

DATE : 30-June-2025

INTRODUCTION

Dataset Overview : The dataset draws on anonymized details from every individual who has registered on the Excelerate platform. It reveals notable patterns in user demographics and academic focus. The dataset includes information such as the user's name, date of birth, country, and various other important points such as the date they signed up for the platform, the name of their opportunity, as well as the start and end date of their opportunity.

Male users, numbering over 5000, outnumber their female counterparts, who exceed 3000, reflecting a gender distribution that is relatively balanced but still tilted toward males. This suggests that while the platform appeals to both genders, it has a stronger pull among males.

Academic interests are concentrated in technology-driven fields, with Information Systems emerging as the predominant major, closely followed by Computer Science. Most users were born between 2000 and 2002, with 2001 being the most common birth year. This indicates that the typical user is a young adult, likely at the beginning of their higher education or early in their professional journey.

Geographically, the user base is primarily concentrated in the United States, closely followed by India. There is also representation from Africa, Australia, and various Asian countries, though these numbers are comparatively lesser. Collectively, the data illustrates Excelerate’s appeal to a youthful and tech-oriented audience.

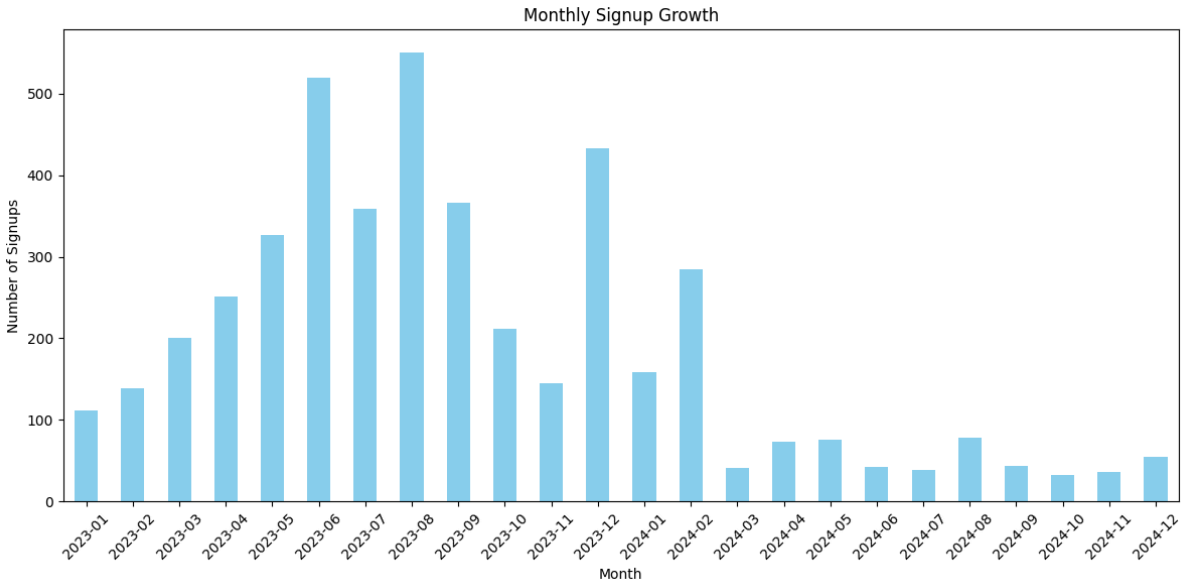
Analysis Goal : The primary objective of analyzing trends in signup and completion data through Exploratory Data Analysis (EDA) is to acquire

insights about user behaviour and engagement. By systematically examining these trends, businesses can identify patterns and anomalies among all the users, which are important for optimising conversion rates and enhancing the overall customer experience.

Understanding signup and completion trends enables businesses to pinpoint factors that drive or hinder user activation and retention. This knowledge is crucial to meet user needs. Ultimately, leveraging these insights supports improved customer satisfaction, higher engagement, and sustained business growth.

Additionally, tracking these trends over time allows businesses to assess the effectiveness of recent initiatives, marketing campaigns, or product changes made based on user engagement. Segmenting the data by demographics or academic focus can reveal specific groups that may require targeted support. Such an analysis helps ensure the platform remains responsive to evolving user expectations and needs.

SIGNUP TRENDS



Observations:

- Strong upward trend from January to August 2023, peaking in August 2023.
- A sharp decline follows from September 2023 onward but sudden peaks in December 2023 and February 2024, with lower but sustained signups into 2024.

Explanation:

- The initial rise likely corresponds to new outreach efforts, academic cycles (e.g., spring/summer course enrollments), or successful marketing campaigns.
- The sudden decline coincides with semester start (students stop exploring new opportunities).
- The drop around beginning in September might also reflect academic breaks and holiday seasons.
- A smaller secondary bump in December 2023 may be tied to end-of-year promotions or a final enrollment window.

Conclusion:

- This chart supports the need to time outreach efforts around late spring and summer, when engagement is highest.

Observations:

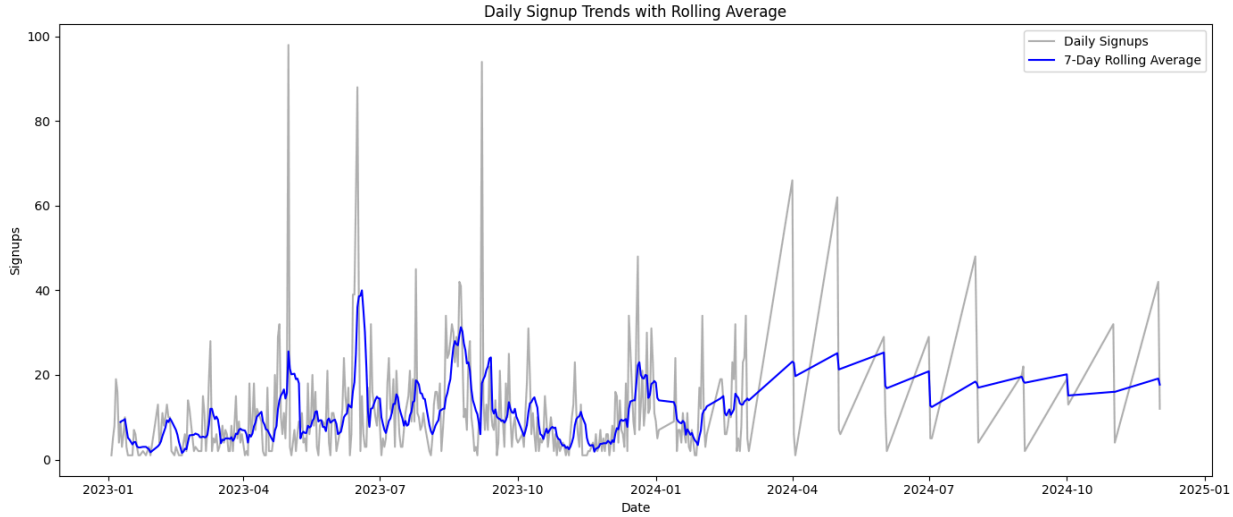
- Signups occur across all days, but most concentrated between 2 – 6pm.
- Weekends show more dispersed and generally lower activity.

Explanation:

- This popularity of afternoons aligns with typical availability patterns of students (after classes) or working professionals (after work or during lunch).
- The Monday spike could be related to weekly newsletter pushes, internal deadlines, or team-based campaigns launching on Mondays.
- Lower activity on weekends suggests signups are tied more to institutional or structured contexts (e.g., courses or mentorship programs) than personal browsing.

Conclusion:

- To maximize engagement, schedule email campaigns or promotional pushes early in the week (esp. Monday afternoon) and around lunch-to-evening hours.



Observations:

- There is clear daily volatility with several sharp spikes in signup activity.
- Spikes occur around June, August, December 2023.
- The 7-day rolling average smooths out the trend and shows overall relatively constant signups

Explanation:

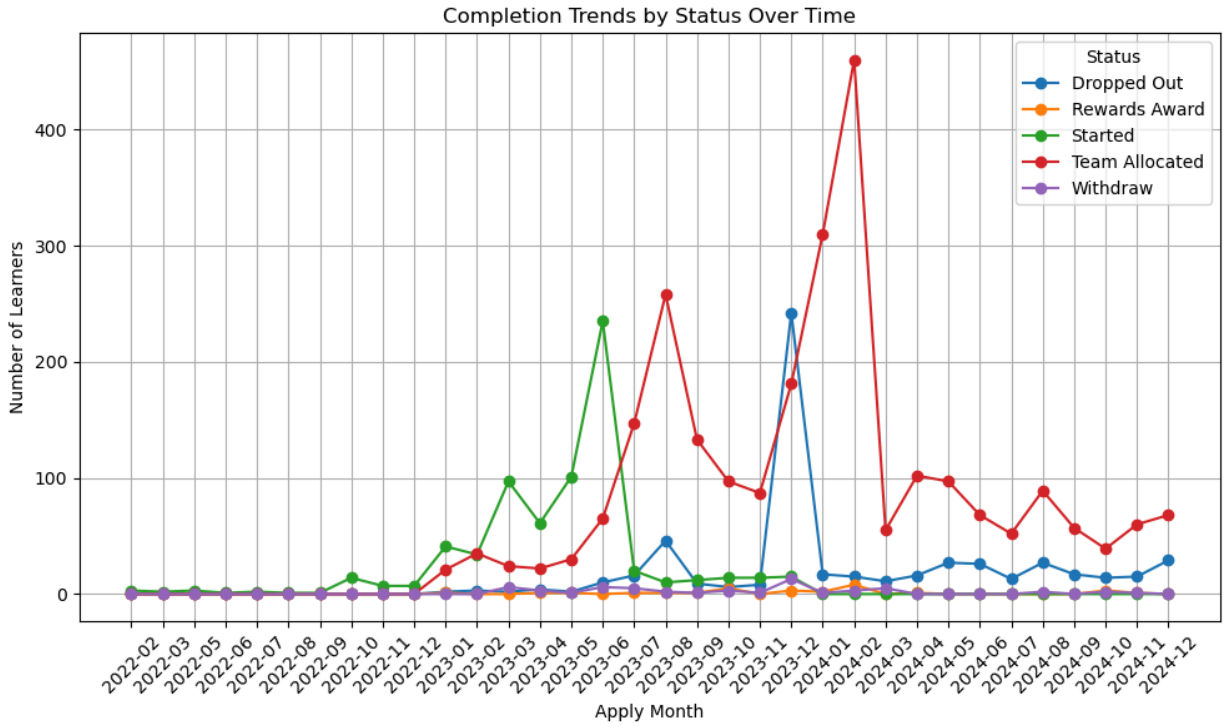
- Sharp spikes might come from the launch of new programs, targeted promotional events, or internal enrollment deadlines
- Sustained, steady, and less volatile interest in 2024, even if lower in overall volume, suggests a good level of retained user acquisition strategy.

Conclusion:

- Daily spikes indicate the Excelerate platform's responsiveness to events or campaigns. Consider analyzing those high-traffic dates to identify what worked well.

COMPLETION TRENDS

- Stability: Show trends in completion rates with line graphs, noting any instability.



Observation:

- A lot of learners started between Jan and June 2023, with the highest number in June.
- In Jan 2024, a big jump happened in team assignments, reaching around 450 learners.
- Many learners dropped out in Nov 2023, just after team allocation increased.
- "Withdraw" and "Rewards Award" stayed low and didn’t change much.

Explanation:

- The rise in learners starting in early 2023 likely means a new batch began.
- A large number of learners were moved into teams in early 2024.
- Soon after that, more learners dropped out—possibly because teams weren't working well or the tasks were too tough.
- Overall, learner numbers kept going up and down during busy periods, showing a lack of stability.

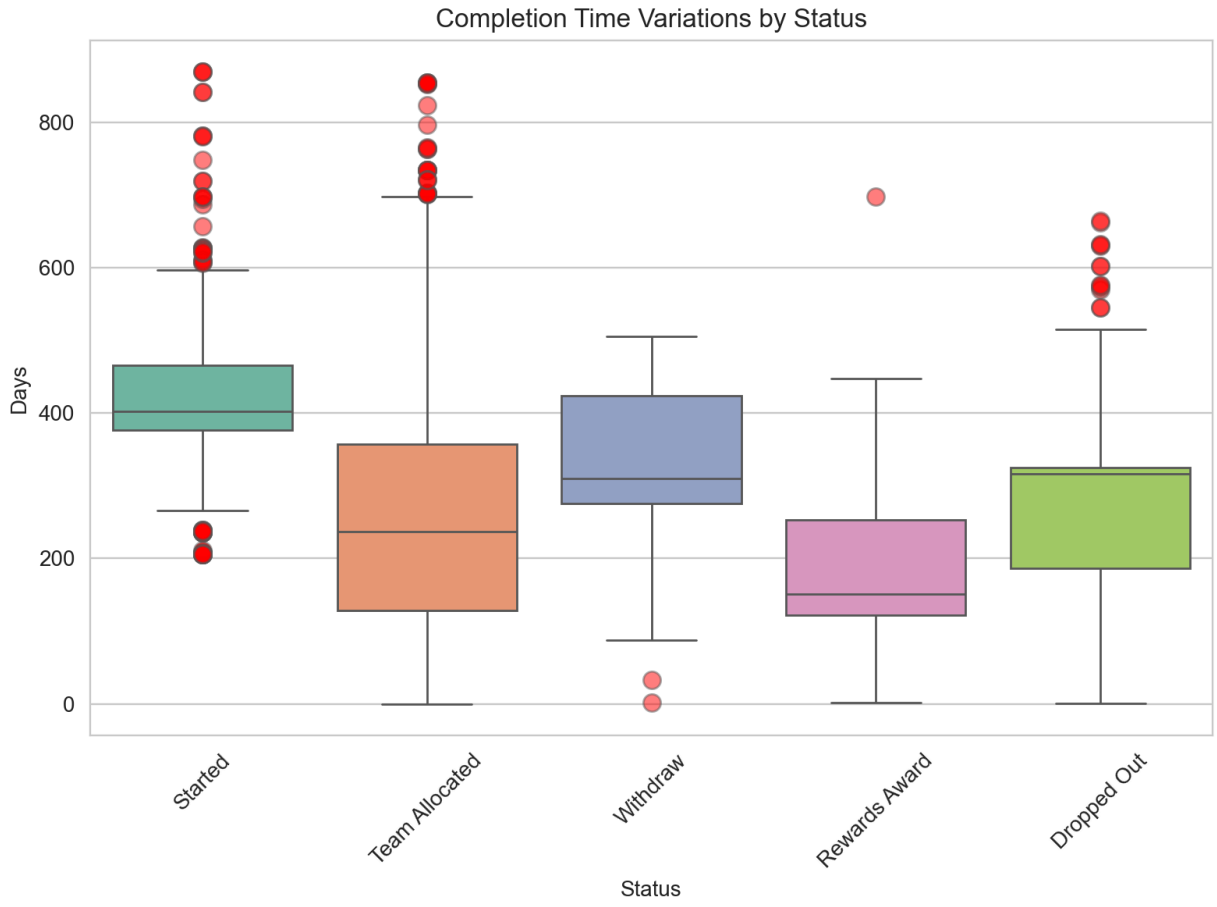
Conclusion:

- The completion process seems unstable, especially when learners are assigned to teams or dropping out.
- Many learners started, but only a few made it to the end.

Instability in Completion Trends:

- Sudden jumps in "Started", "Team Allocated", and "Dropped Out" show an uneven flow of learners.
- Completion trends were not steady—they kept rising and falling each month.

- Time Variations: Use box plots to illustrate changes in completion times and outliers.



Observation:

- Started and Team Allocated stages have many unusual cases, with some projects taking more than 800 days.
- Withdraw and Dropped Out also show a lot of variation in time, though with fewer extreme cases.

- Rewards Award is the most stable, with fewer delays and more consistent timing.

Explanation:

- The delays in Started and Team Allocated suggest that early stages face issues like poor planning or lack of resources.
- The ups and downs in Withdraw and Dropped Out stages show that exits are not handled in a steady way.
- Rewards Award is smoother, which means it is likely better planned and managed.

Conclusion:

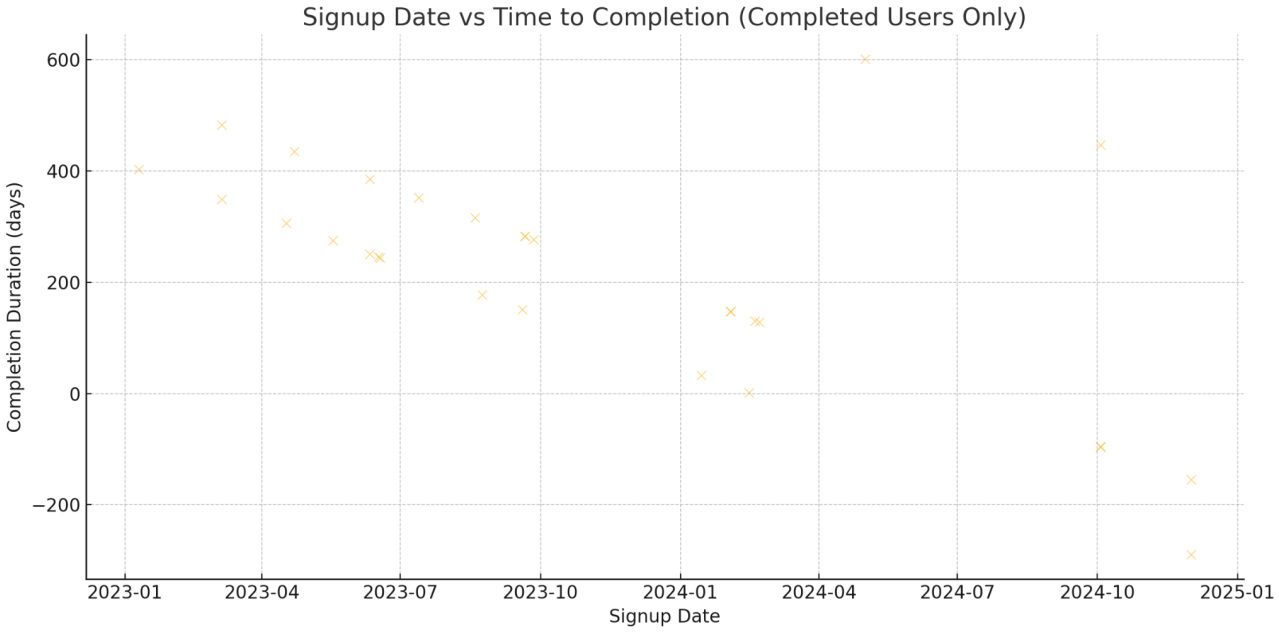
- The early parts of the project are not stable and often face long delays.
- Later stages like Rewards Award are more reliable, showing the value of good planning.
- To avoid delays, it's important to improve how projects are managed from the beginning.

Outliers :

- Started and Team Allocated stages have many extreme cases, with times going over 800 days
- This shows delays and irregularities in the early parts of the project
- Rewards Award stage has very few such delays, showing smoother execution
- Highlights the need for better planning and management in the initial project phases

PATTERNS AND CORRELATIONS

The signup date of each completed user is plotted against the corresponding completion date in this scatter plot. The fact that every point is above the diagonal (y = x) as predicted verifies that completions come after signups.



The scatter plot above visualizes the relationship between **signup date** and **time to completion** (in days) for users who completed their opportunities which is marked as **Rewards Award.**

Some important insights revealed:

Activity concentration: In line with the mid-2023 signup spike, the majority of completions occur between June 2023 and March 2024.

This implies that users are more likely to complete their registration if they do it during periods of high engagement.

Consistency of program duration: For the majority of users, the time between the signup and completion dates is quite constant, suggesting a rather uniform duration for numerous opportunities.

Lagging completions: A few outliers show extended intervals between registration and completion. These might have to do with delayed engagement or longer or more complicated opportunities.

The platform successfully encourages completions after significant signup waves, especially during the summer and end-of-year cycles, as this visualization confirms. In order to assist users after they sign up and guide them gradually toward completion, it also emphasizes the significance of ongoing engagement strategies.

Performance Across different groups

1. **Gender Based Performance**

   Based on the data, completion rates are extremely low across all genders, though male and female users show nearly equal performance.

Gender	**Completion Rate**
Male	0.65%
Female	0.62%

Rather not say	0.00%

Engagement Strategy:

- Frequent and Customised reminders could increase continuation.
- A more gender-inclusive strategy to increase engagement.

2. Major Based Performance

Based on the Top 10 majors, Tech-related majors dominate but completion remains low overall.

Major	**Completion Rate**
Information Systems	1.19%
Computer Science	1%
Cyber Security	4.21%
Others	0%

Engagement Strategy:

- Leaderboard: A friendly competition element that ranks users based on activity or achievement.
- Progress Badges: Users earn badges as they complete tasks.

3. **Country Based Performance**

   Users in the United States show the highest completion rates. Some countries, despite the high signup volumes, show no completions.

Country	**Completion Rate**
Others	0.00%
Nigeria	0.22%
India	0.39%
Ghana	0.56%
United States	1.09%

Engagement Strategy:

- Run country based adverts based on users from different countries alumni success stories.
- Introduce **c**ountry based cohorts.

OUTLIERS AND ANOMALIES

In [1]:
```
pip install pandas --quiet
```

```
Note: you may need to restart the kernel to use updated packages.
```

In [2]:
```
pip install numpy --quiet
```

```
Note: you may need to restart the kernel to use updated packages.
```

In [3]:
```
import pandas as pd
import numpy as np

print('Pandas Version:', pd.__version__)
print('\nNumpy Version:', np.__version__)
```

```
Pandas Version: 2.2.2

Numpy Version: 1.26.4
```

In [5]:
```
import matplotlib.pyplot as plt
```

In [8]:
```
df = pd.read_csv('Downloads/Cleaned_Preprocessed_Dataset_Week1 - SLU Opportunity Wi
```

In [10]:
```
df['Opportunity Start Date'] = pd.to_datetime(df['Opportunity Start Date'], errors=
df['Opportunity End Date'] = pd.to_datetime(df['Opportunity End Date'], errors='coe
```

In [13]:
```
df['Completion Duration'] = (df['Opportunity End Date'] - df['Opportunity Start Dat
df[['Opportunity Start Date', 'Opportunity End Date', 'Completion Duration']]
```

Out[13]:

	Opportunity Start Date	Opportunity End Date	Completion Duration
0	2022-11-03 18:30:39	2024-06-29 18:52:39	604.0
1	2022-11-03 18:30:39	2024-06-29 18:52:39	604.0
2	2022-11-03 18:30:39	2024-06-29 18:52:39	604.0
3	2022-11-03 18:30:39	2024-06-29 18:52:39	604.0
4	2022-11-03 18:30:39	2024-06-29 18:52:39	604.0
...	...	...	...
4567	2024-03-08 14:00:00	2024-03-08 11:30:00	-1.0
4568	2024-03-08 14:00:00	2024-03-08 11:30:00	-1.0
4569	2024-03-08 14:00:00	2024-03-08 11:30:00	-1.0
4570	2024-03-08 14:00:00	2024-03-08 11:30:00	-1.0
4571	2024-03-08 14:00:00	2024-03-08 11:30:00	-1.0

4572 rows × 3 columns

In [14]:
```
Q1 = df['Completion Duration'].quantile(0.25)
Q3 = df['Completion Duration'].quantile(0.75)
```

```
IQR = Q3 - Q1

lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Find anomalies
anomalies = df[(df['Completion Duration'] < lower_bound) | (df['Completion Duration
```

```
print("🔍 Detected Anomalies:")
anomalies[['First Name', 'Completion Duration', 'Opportunity Start Date', 'Opportun
```

```
🔍 Detected Anomalies:
```

Out[14]:

	First Name	Completion Duration	Opportunity Start Date	Opportunity End Date

In [15]:

```
# Group by end date only (not time)
df['Completion Date Only'] = df['Opportunity End Date'].dt.date

# Count completions per day
daily_counts = df['Completion Date Only'].value_counts().sort_index()
daily_df = daily_counts.reset_index()
daily_df.columns = ['Date', 'Completions']

# Define low days as below the 25th percentile
low_threshold = daily_df['Completions'].quantile(0.25)
low_days = daily_df[daily_df['Completions'] < low_threshold]

print("📉 Low Completion Days:")
low_days
```

```
📉 Low Completion Days:
```

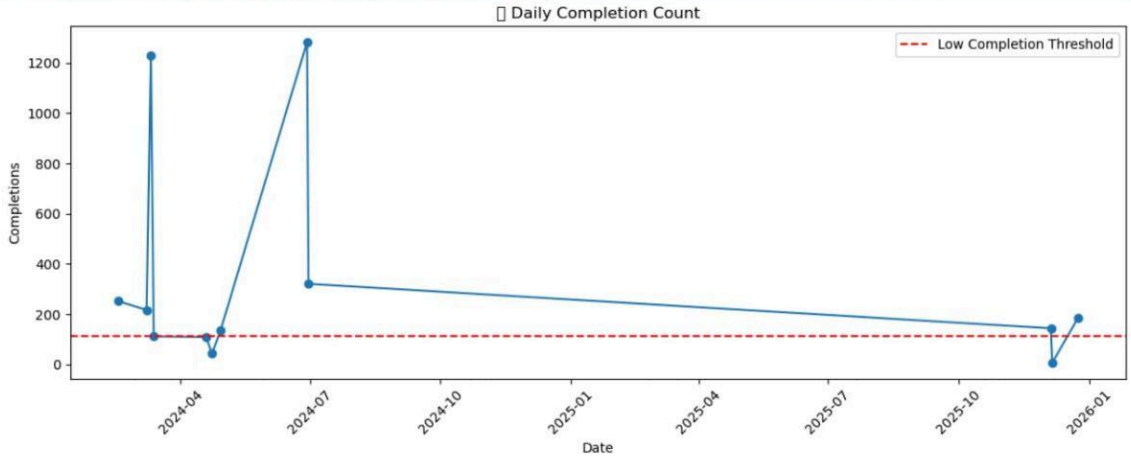
Out[15]:

	Date	Completions
4	2024-04-19	109
5	2024-04-23	43
10	2025-12-06	7

In [16]:

```
plt.figure(figsize=(12, 5))
plt.plot(daily_df['Date'], daily_df['Completions'], marker='o')
plt.axhline(low_threshold, color='red', linestyle='--', label='Low Completion Thres
plt.title('📈 Daily Completion Count')
plt.xlabel('Date')
plt.ylabel('Completions')
plt.xticks(rotation=45)
plt.legend()
plt.tight_layout()
plt.show()
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_4932\2510839720.py:9: UserWarning: Glyph
128200 (\N{CHART WITH UPWARDS TREND}) missing from font(s) DejaVu Sans.
  plt.tight_layout()
C:\ProgramData\anaconda3\Lib\site-packages\IPython\core\pylabtools.py:170: UserWarni
ng: Glyph 128200 (\N{CHART WITH UPWARDS TREND}) missing from font(s) DejaVu Sans.
  fig.canvas.print_figure(bytes_io, **kw)
```




```
In [17]: pip install nbconvert
```

Brief Summary of the Steps:

1. Loaded the dataset using pandas.
Read the data (usually in CSV format) into a DataFrame.

2. Converted string date columns into actual datetime objects.
Specifically for: Opportunity Start Date and Opportunity End Date.

3. Calculated the duration of completion.
Subtracted Start Date from End Date to get the number of days it took to complete each opportunity.

4. Detected anomalies in completion time.
Used the IQR (Interquartile Range) method to find outliers — entries that took unusually short or long times to complete**.**

5. Identified low completion days.
Grouped the data by End Date and counted how many completions happened each day, then identified the days with completion counts below the 25th percentile as “low completion” days.

6. Visualized daily completions using a line plot.
This helps visually spot patterns, spikes, or low-activity periods.

RECOMMENDATIONS

Target peak days

- Intensify marketing efforts eg social media ads and email campaigns during peak months which are mainly june august and december
- Offer limited time incentives such as early bird discounts bonus content or referral rewards to encourage immediate signups
- Highlight success stories or testimonies during these peak days to reinforce trust and motivation to students hence more signups

Drops in signups and completions

- Conduct user feedback surveys and analyze user behaviour data to identify common drop-off points
- Implement reminder systems ,such as automated emails or in-app notifications to re-engage inactive users
- Since most students drop-off is December avoid major transitions or deadlines in december. Also launch engagement campaigns before the holiday season to keep students motivated

Support long tail users

- Provide progress tracking dashboard, personalized timelines and motivational nudges to keep engagement
- Offer flexible pacing options including the ability to pause and resume or opt-in for deadline extensions
- Create a support community or mentorship program to connect slower users with peers ar mentors for encouragement

Strategies for different demographic groups

- Offer multilingual support and culturally relevant examples where appropriate to improve inclusivity and accessibility
- Tailor communication styles and content formats based on age location or user goals
- Introduce country based cohorts

CONCLUSION

Key Insights : The analysis highlights clear patterns in user engagement and completion across the platform. Signup activity peaked from January to August 2023, especially in August, then declined with smaller spikes in December and February. Most signups occurred on weekday afternoons, particularly Mondays. Completion trends showed that many learners started in early 2023, with the highest numbers in June, but dropouts increased after team allocation, making the process unstable and resulting in few learners finishing.

Early project stages, such as "Started" and "Team Allocated," experienced the most delays and irregularities, while the "Rewards Award" stage was more consistent. Outliers with extreme delays were mostly seen in the initial phases. Completion rates were low across all genders and majors, with tech majors dominating signups. U.S. users had the highest completion rates, while some countries had none.

Next Steps : For further analysis, we can begin by collecting more detailed data on user engagement, dropout reasons, and completion barriers across different groups. Surveys or interviews can be used to gain qualitative insights, and segment the data by time, user demographics, and project stages to identify specific patterns.

Next, we can apply advanced analytics and visualisation tools to track user journeys and measure the impact of interventions such as reminders or gamification. Regular review of the effectiveness of these strategies can help refine them as needed. Ensuring data privacy and security throughout the process is of utmost importance, as well as treating analysis as an ongoing effort to drive continuous improvement.