AICP Internship Task

User Profiling and Segmentation

User profiling and segmentation are powerful techniques that enable data professionals to understand their user base in-depth and tailor their strategies to meet diverse user needs. Below is the process we can follow for the task of User Profiling and Segmentation:

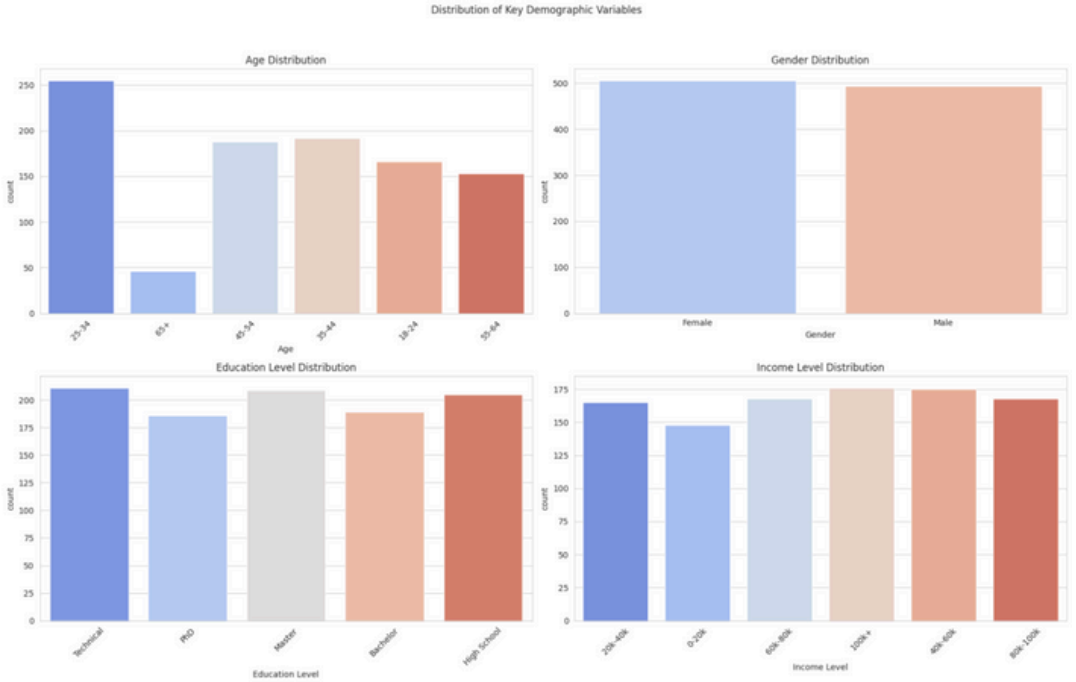
1. Determine what you aim to achieve with user profiling and segmentation, such as improving customer service, personalized marketing, or product recommendation.
2. Collect data from various sources, including user interactions on websites/apps, transaction histories, social media activity, and demographic information.
3. Create new features that capture relevant user behaviours and preferences. It may involve aggregating transaction data, calculating the frequency of activities, or extracting patterns from usage logs.
4. Select appropriate segmentation techniques.
5. For each segment identified, create user profiles that summarize the key characteristics and behaviours of users in that segment.

Find dataset “user_profiles_for_ads.csv”
Here’s an overview of all the columns in the dataset:

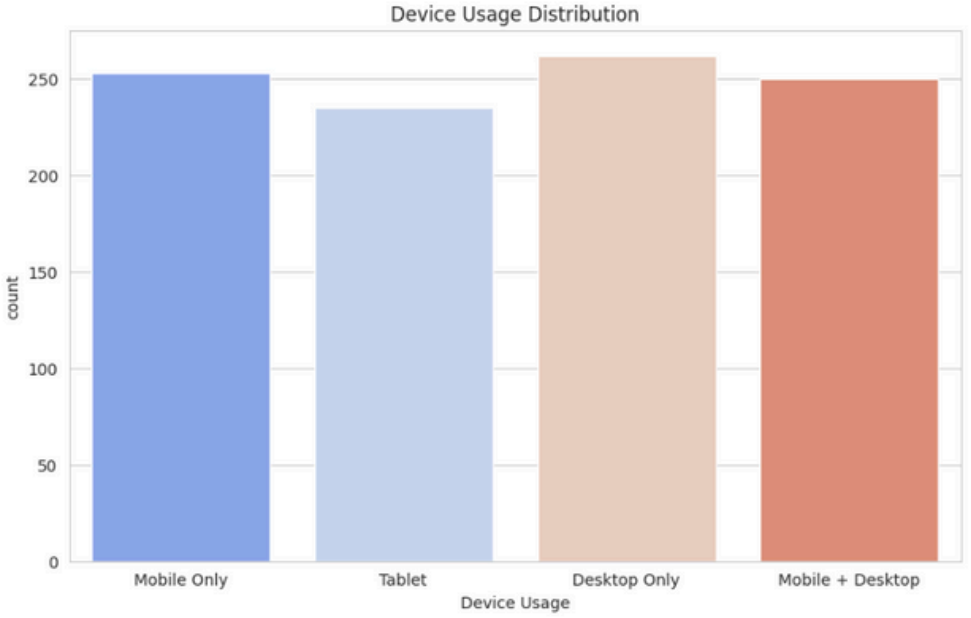
- **User ID**: Unique identifier for each user.
- **Age**: Age range of the user.
- **Gender**: Gender of the user.
- **Location**: User’s location type (Urban, Suburban, Rural).
- **Language**: Primary language of the user.
- **Education Level**: Highest education level achieved.
- **Likes and Reactions**: Number of likes and reactions a user has made.
- **Followed Accounts**: Number of accounts a user follows.
- **Device Usage**: Primary device used for accessing the platform (Mobile, Desktop, Tablet).
- **Time Spent Online (hrs/weekday)**: Average hours spent online on weekdays.
- **Time Spent Online (hrs/weekend)**: Average hours spent online on weekends.
- **Click-Through Rates (CTR)**: The percentage of ad impressions that lead to clicks.
- **Conversion Rates**: The percentage of clicks that lead to conversions/actions.
- **Ad Interaction Time (sec)**: Average time spent interacting with ads in seconds.
- **Income Level**: User’s income level.
- **Top Interests**: Primary interests of the user.

Q.1: Import data and check null values, column info, and descriptive statistics of the data.

Q.2: Begin EDA by visualizing the distribution of the key demographic variables (age, gender, education level, and income level)



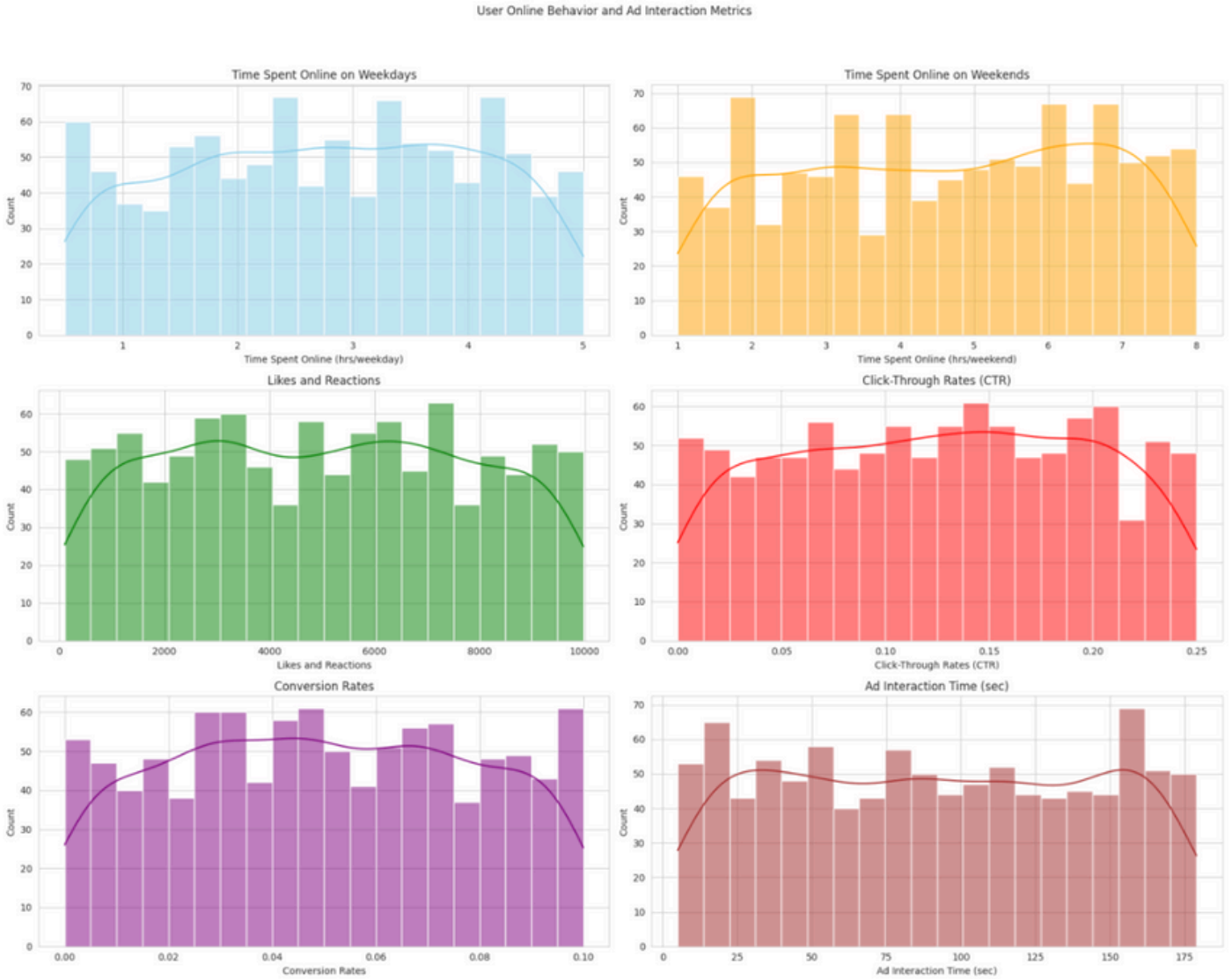
Q.3: Now examine device usage patterns to understand the primary means by which users access the platform. Additionally, explore users' online behaviour, including their engagement with content and ads, and identify the most common interests among users.



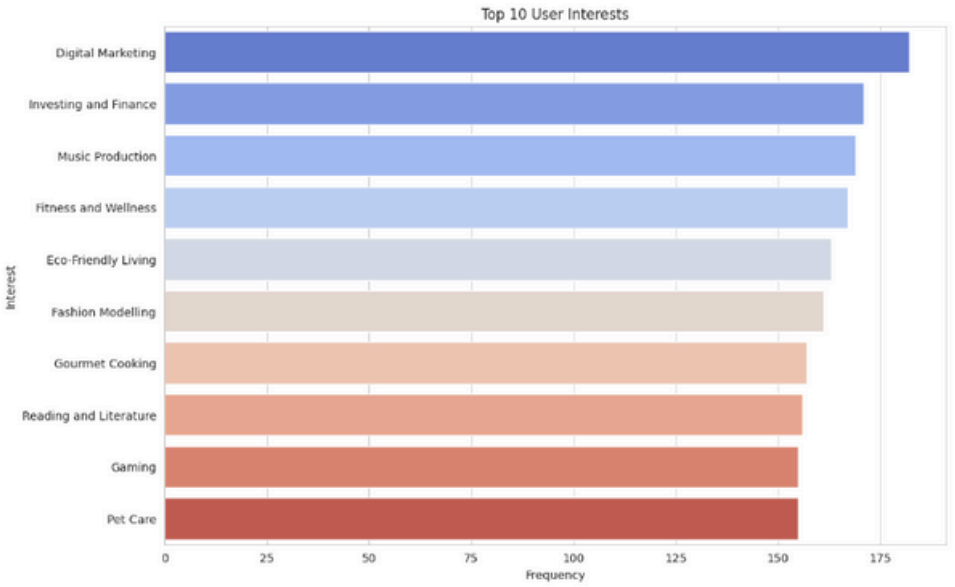
Q.4:

- Analyze the average time users spend online on weekdays versus weekends.
- Investigate user engagement metrics, such as likes and reactions.
- Delve into ad interaction metrics, including Click-Through Rates (CTR), Conversion Rates, and Ad Interaction Time.

It will help to understand the users' activity patterns and their interaction with ads, which is crucial for effective ad targeting and optimization.



Q.5: Identify the most common interests among users, which is vital for segmenting the audience and tailoring ad campaigns to match user preferences



Q.6: You can now segment users into distinct groups for targeted ad campaigns. Segmentation can be based on various criteria, such as:

- Demographics: Age, Gender, Income Level, Education Level
- Behavioural: Time Spent Online, Likes and Reactions, CTR, Conversion Rates
- Interests: Aligning ad content with the top interests identified

To implement user profiling and segmentation, apply clustering techniques or develop personas based on the combination of these attributes.

Start by selecting a subset of features that could be most indicative of user preferences and behaviour for segmentation and apply a clustering algorithm to create user segments:

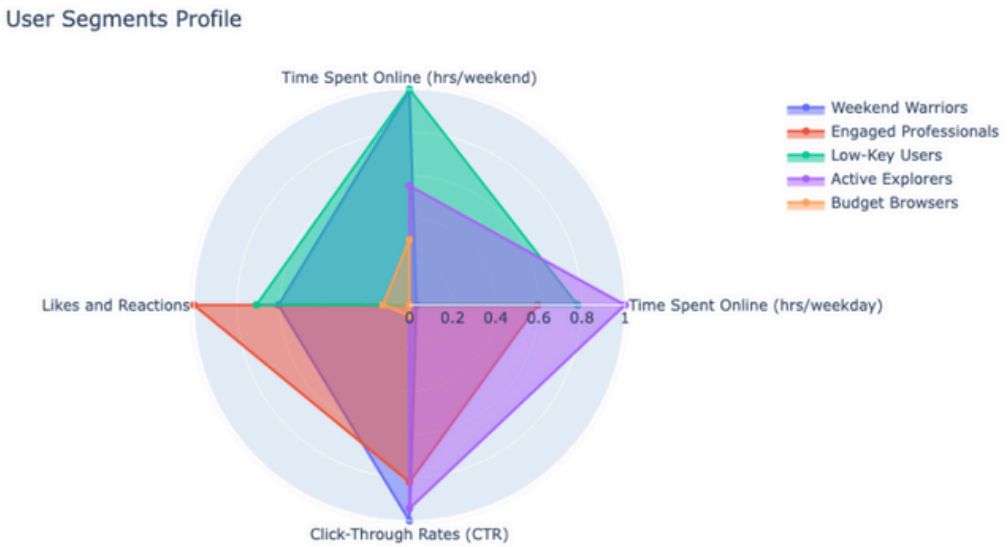
Q.7: Now compute the mean values of the numerical features and the mode for categorical features within each cluster to get a sense of their defining characteristics

Q.8: Now assign each cluster a name that reflects its most defining characteristics based on the mean values of numerical features and the most frequent categories for categorical features.

Based on the cluster analysis, summarize and name the segments as follows:

1. **Cluster 0** – "Weekend Warriors": High weekend online activity, moderate likes and reactions, predominantly male, age group 25-34, income level 80k-100k.
2. **Cluster 1** – "Engaged Professionals": Balanced online activity, high likes and reactions, predominantly male, age group 25-34, high income (100k+).
3. **Cluster 2** – "Low-Key Users": Moderate to high weekend online activity, moderate likes and reactions, predominantly male, age group 25-34, income level 60k-80k, lower CTR.
4. **Cluster 3** – "Active Explorers": High overall online activity, lower likes and reactions, predominantly female, age group 25-34, income level 60k-80k.
5. **Cluster 4** – "Budget Browsers": Moderate online activity, lowest likes and reactions, predominantly female, age group 25-34, lowest income level (0-20k), lower CTR.

Q.9: Now, create a visualization that reflects these segments, using the cluster means for numerical features and highlighting the distinctive characteristics of each segment. Create a radar chart that compares the mean values of selected features across the clusters, providing a visual representation of each segment's profile.



Q.10: Write down the summary of your above experience.