Podcast Listening Behavior: Key Insights and Actionable Recommendations

Prepared by: Shahad Bakhsh

Dataset Coverage: ~18+ user sessions across multiple podcast episodes, categories, countries, and demographics.

1. Introduction:

This exploratory data analysis reveals clear patterns of user engagement, age- and gender-specific content preferences, high-performing episodes, and regional listening trends. It also highlights opportunities for targeted content promotion and areas needing UX improvement. The findings are intended to inform data-driven content strategies, enhance listener retention, and support focused market expansion.

2. Data Health Overview

Key Steps:

- Merged 3 datasets (users, episodes, listens)
- Removed duplicates and corrected data types
- Checked for invalid values, including:
  - 0-second sessions
  - Negative values (none found)
  - Outliers in durations and age (none found)

- Did not apply imputation:
  - No missing or null values were present in the required columns after merging.
- Skipped scaling and normalization:
  - Not needed since the analysis was descriptive, not machine learning-based
- Skipped categorical encoding:
  - Categorical variables (gender, country, category) were used directly for grouping and filtering.
  - Encoding is only necessary for machine learning models, which were not part of this stage.

Note: 13 sessions had zero duration, often due to app bugs or accidental taps. A possible UX red flag.

3. Data Analytics:

## 3.1 Most Listened-To Episode Categories

Analysis:

Listening sessions were grouped by category, and total listening time was calculated in hours. Based on the visualized data, the top three categories with the highest engagement are Sports, Society, and News. Sports leads significantly, followed by Society, and then News.

Insight:

Listeners show strong interest in event-driven and socially relevant content. Sports content likely benefits from time-sensitive narratives such as match commentary or highlights. Society-related episodes may reflect listeners' preferences for culturally relevant discussions or interviews. News content suggests engagement with current affairs and daily updates.

Recommendation:

Refocus content development strategies toward Sports, Society, and News. For Sports, consider expanding into pre- and post-match analysis, athlete interviews, or fan-generated commentary. For Society, explore collaborations with influential speakers and community leaders. For News, produce brief, well-paced updates or region-specific recaps. These efforts can strengthen content alignment with audience interests and increase total listening time across sessions.

Top Categories by Total Listening Time (in Hours)
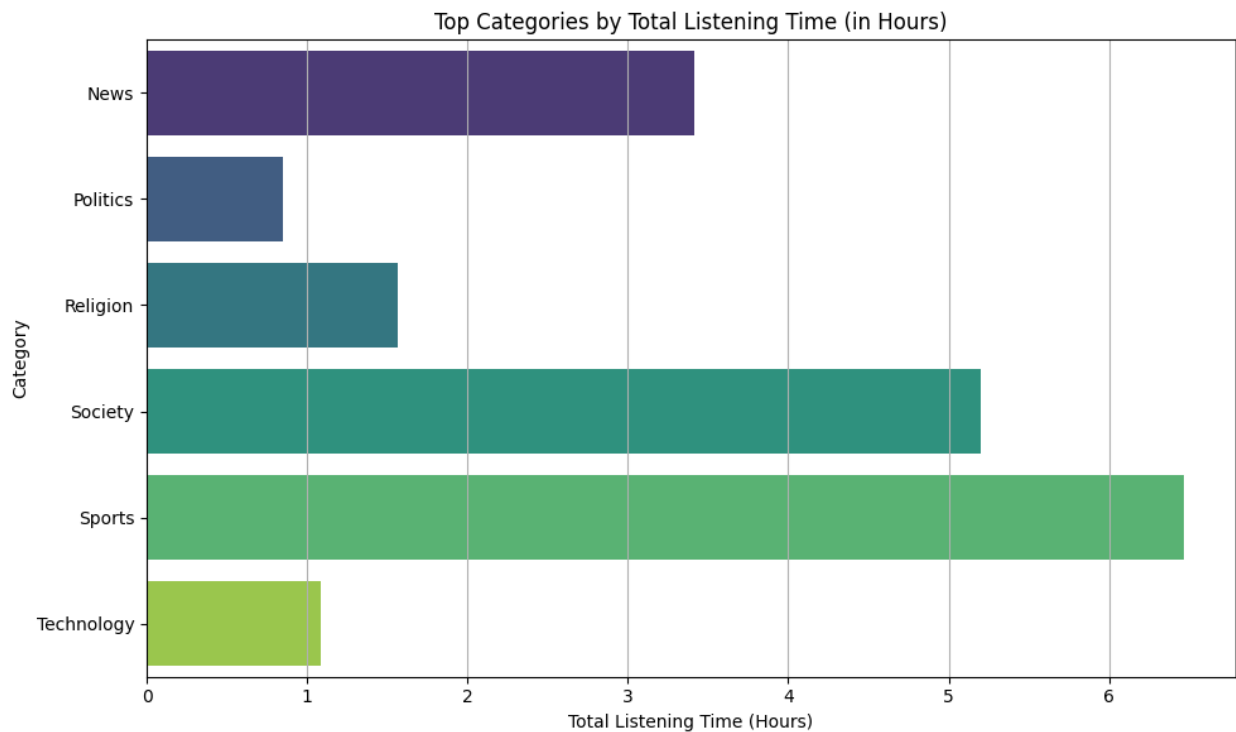
3.2 Average Listening Duration by Gender

Analysis:

The dataset was grouped by gender to calculate two metrics:

- The number of unique users by gender
- The average listening duration (in minutes) per session

The visualized data shows that the number of male and female users is nearly equal (10 females vs. 9 males), ensuring a fair basis for comparison. However, female users exhibit a notably higher average listening duration, around 15 minutes per session, compared to approximately 10.5 minutes for male users.

Insight:

This difference suggests that female listeners tend to stay engaged longer during each listening session. The pattern may reflect stronger emotional or cognitive investment in the content, potentially due to topic relevance, pacing, or preferred formats.

Recommendation:

Consider tailoring content formats by gender. For the female segment, prioritize the development of longer, narrative-driven episodes or serial storytelling formats. Additionally, test promotional messages that encourage deep listening, such as themed playlists or multi-part series. For male users, explore whether shorter, high-impact episodes or dynamic formats (e.g., interviews, summaries) perform better in terms of retention.
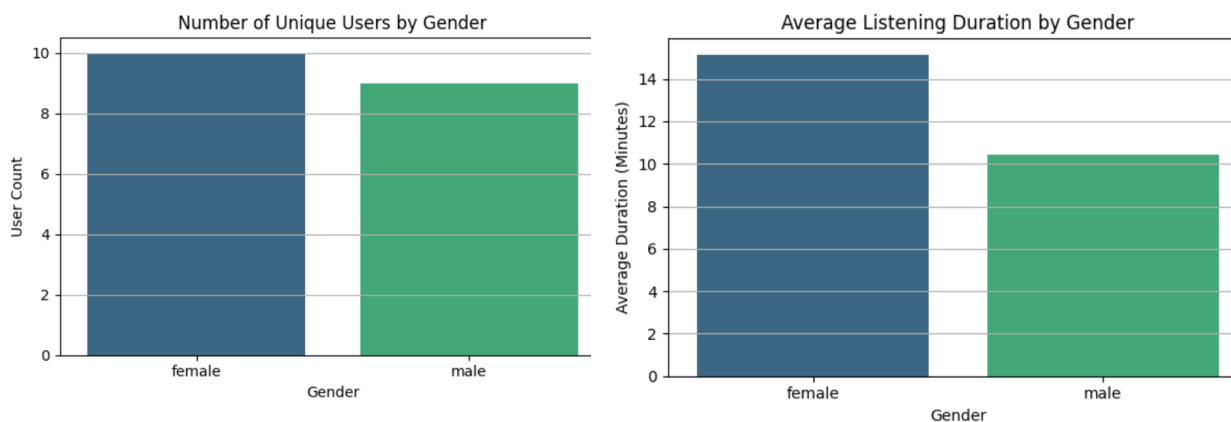
3.3 Average Episodes per User

Analysis:

The analysis measured how many unique episodes each user listened to by grouping data by user ID and counting distinct episode IDs. Sessions with a duration of zero seconds were removed to ensure data accuracy. The distribution shows a concentration around 4–6 episodes, with the mean number of unique episodes per user at approximately 4.42.

Insight:

This metric reflects moderate user engagement. While some users explore many episodes, others show limited interaction. The histogram reveals that most users consume a small set of content, suggesting the need for retention strategies after initial exposure.

Recommendation:

Introduce engagement features such as personalized "You Might Also Like" carousels or prompts triggered after users complete 3–5 episodes. These can encourage content discovery and increase listening volume per user. Consider nudging users who stop at 1–2 episodes with notifications or follow-up suggestions to improve retention.

Distribution: Unique Episodes per User
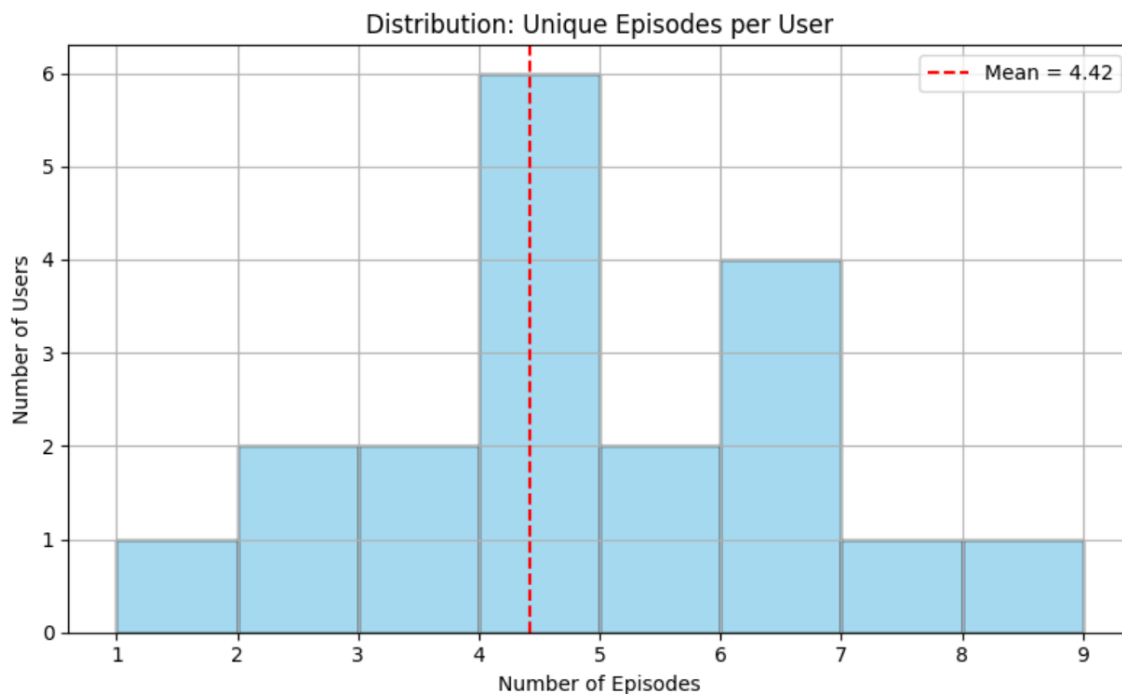
## 3.4 Average Listening Duration by Age Group

Analysis:

Listening session durations were analyzed across individual user ages and grouped by decade into five age groups: teens (10s), 20s, 30s, 40s, and 50s. The boxplot visualization shows a spread of listening times within each age group, with some notable variations. Users in their 30s and early 40s consistently demonstrate longer and more stable listening durations, with several users reaching the maximum observed session duration of 30 minutes.

Insight:

The data suggests that older users, particularly those in their 30s and 40s, tend to engage more deeply with podcast content. These users maintain longer session durations with less variation, implying higher sustained interest. Conversely, teenagers and users in their early 20s have shorter and more inconsistent listening times, possibly indicating lower content retention or preference for shorter formats.

Recommendation:

Content strategies should be tailored to match these behavioral trends. For users under 25, consider producing shorter, high-impact episodes or teaser-style content that encourages further exploration. For users aged 30 and above, invest in deeper, narrative-rich formats such as interviews, thematic series, or long-form discussions that align with their listening preferences. This segmentation approach can improve engagement and increase completion rates across different audience cohorts.

Listening Duration by Age (in Minutes), Colored by Age Group
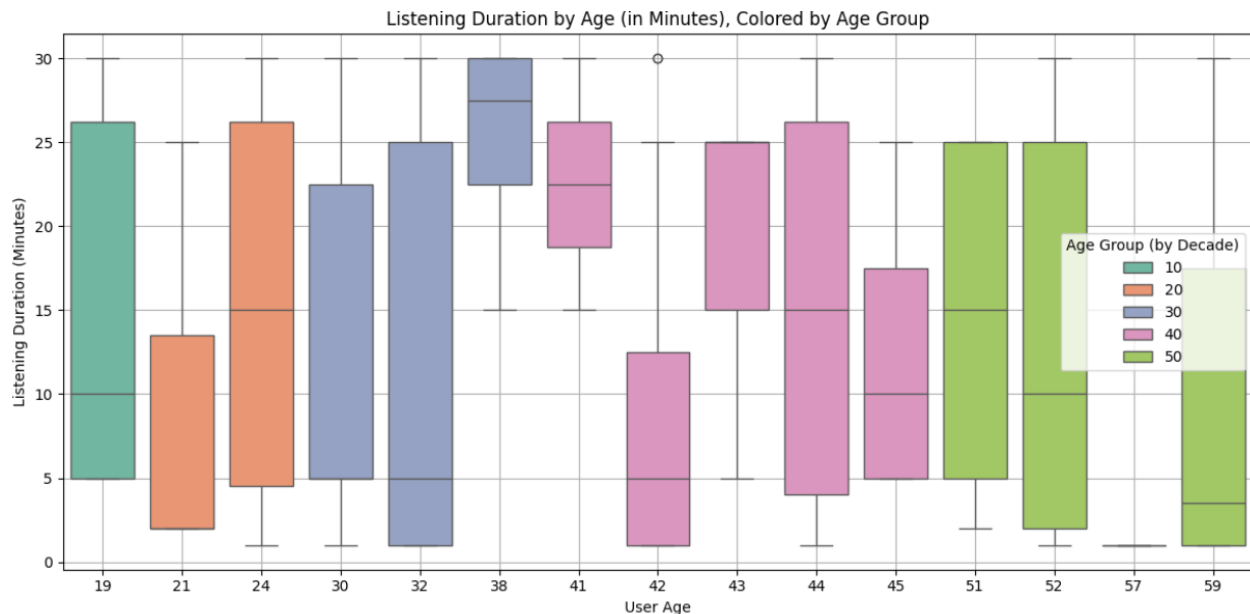
3.5 Gender-Based Podcast Category Preferences and Shared Interests

Analysis: This visualization compares total listening time (in seconds) across podcast categories by gender. Each bar represents the aggregate listening time contributed by either male or female users for a specific category. From the data:

- Females spent the most time listening to Sports, Society, and News.
- Males also favored Society, Sports, and News, though with slightly different proportions.
- Categories like Religion, Technology, and Politics saw lower engagement, with noticeable gender differences in Religion (higher among females) and Politics (higher among males).

Insight: Both genders share a strong interest in Society, Sports, and News, making these common favorite categories. However, gender-specific differences are also evident:

- Females show higher engagement with Religion and Technology.
- Males lean more toward Politics.

These distinctions suggest that while there is substantial overlap in high-level interests, secondary preferences vary and could inform more targeted content strategies.

Recommendation: Focus future podcast production and marketing efforts around Society, Sports, and News as shared top categories. Simultaneously, personalize recommendations and featured content based on secondary interests, such as highlighting Religion and Technology for female users, and Politics for male users. Additionally, consider designing gender-aware campaign messaging that reflects these nuanced preferences.
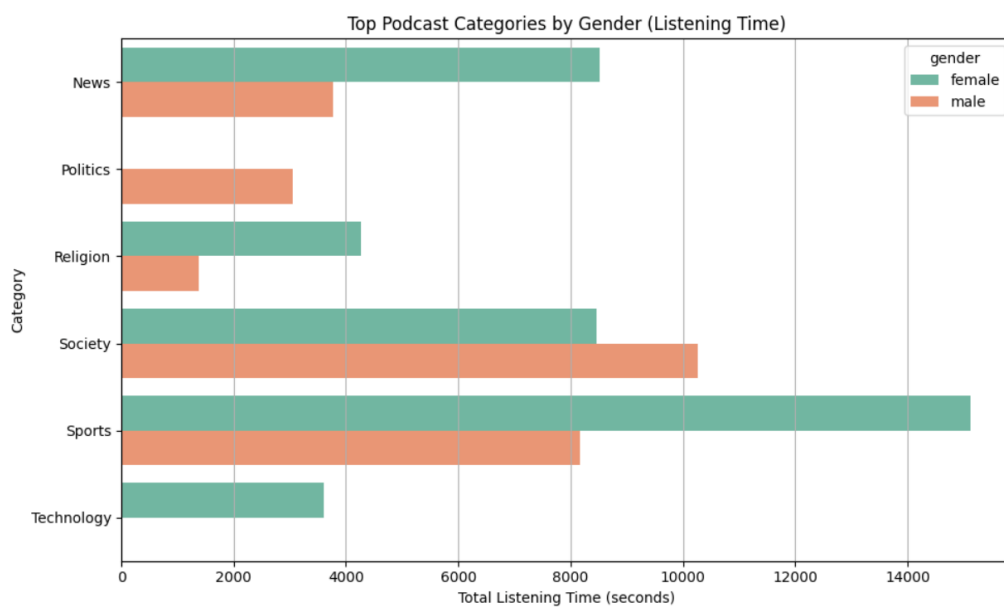
3.6 High Repeat Listens by Age Group

Analysis: This chart displays the number of unique users who replayed specific podcast episodes, segmented by age group. The data reveals that:

- Episode 9 was replayed by a listener in their 20s.
- Episodes 1 and 3 were each replayed by users in their 40s.

Insight: Repeat listening is a strong indicator of content that resonates. For users in their 20s, Episode 9 may reflect motivational or career-aligned themes. Episodes 1 and 3, which were repeatedly played by users in their 40s, likely appeal to themes such as personal reflection, family, or life balance.

Recommendation: Use repeat-listen behavior to guide targeted content promotion. Feature high-replay episodes in segmented newsletters or app banners labeled "Most Replayed by Your Peers." Additionally, assess the structure and themes of these episodes to inform future production strategies tailored to age-specific audience preferences.

Top Repeated Episodes by Age Group
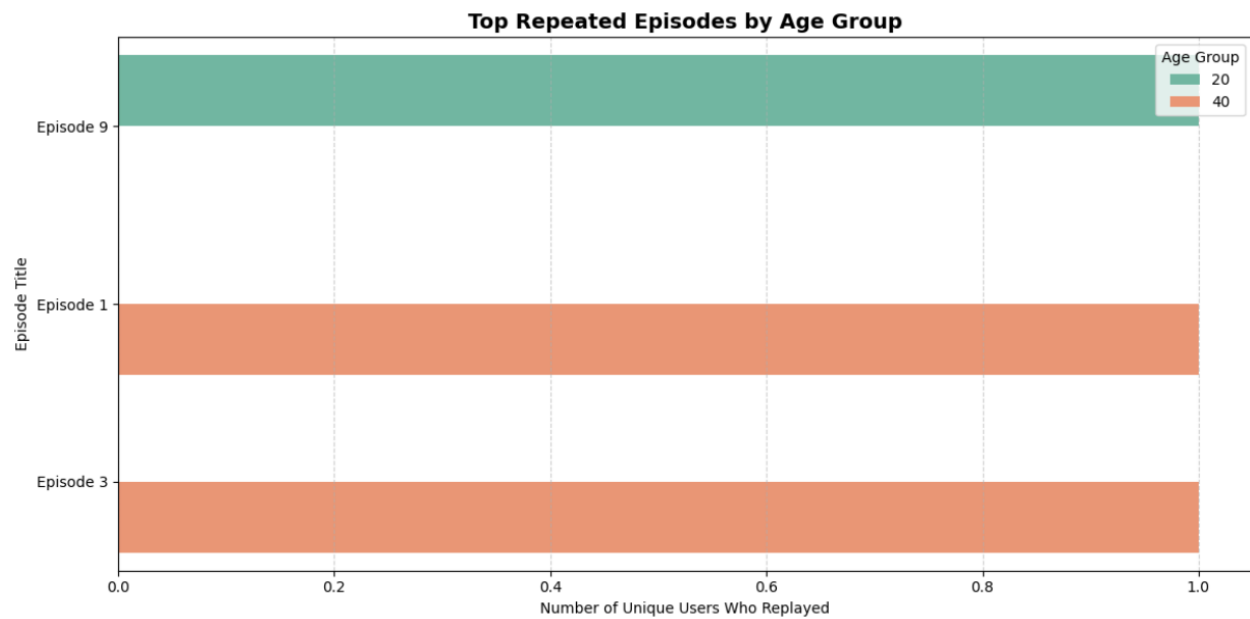
3.7 Listening Time by Country

Analysis:
Total listening time was aggregated by country and converted to hours for clearer interpretation. The five countries featured in this chart demonstrate differing levels of podcast engagement:
- Jordan: 5.5 hours
- Morocco: 5.2 hours
- Saudi Arabia: 3.4 hours
- Egypt: 2.8 hours
- UAE: 1.7 hours

This analysis highlights significant regional variation in listening behavior.

Insight:
Jordan and Morocco show the highest total listening time, indicating strong podcast engagement in those markets. Saudi Arabia and Egypt follow with moderate levels of engagement, while the UAE reflects comparatively lower listening activity. These variations could be influenced by differences in content relevance, platform visibility, language preferences, or cultural listening habits.

Recommendation:
Strengthen content presence in high-performing countries like Jordan and Morocco through regionally tailored episodes, guest speakers, and dialect-specific delivery. In Saudi Arabia and Egypt, focus on increasing engagement through push notifications, culturally aligned episode themes, and improved app features. For the UAE, invest in awareness campaigns, influencer collaborations, or platform partnerships to boost discovery and usage. Tailoring strategy by country will help increase overall reach, engagement, and monetization potential.

Countries by Total Listening Time
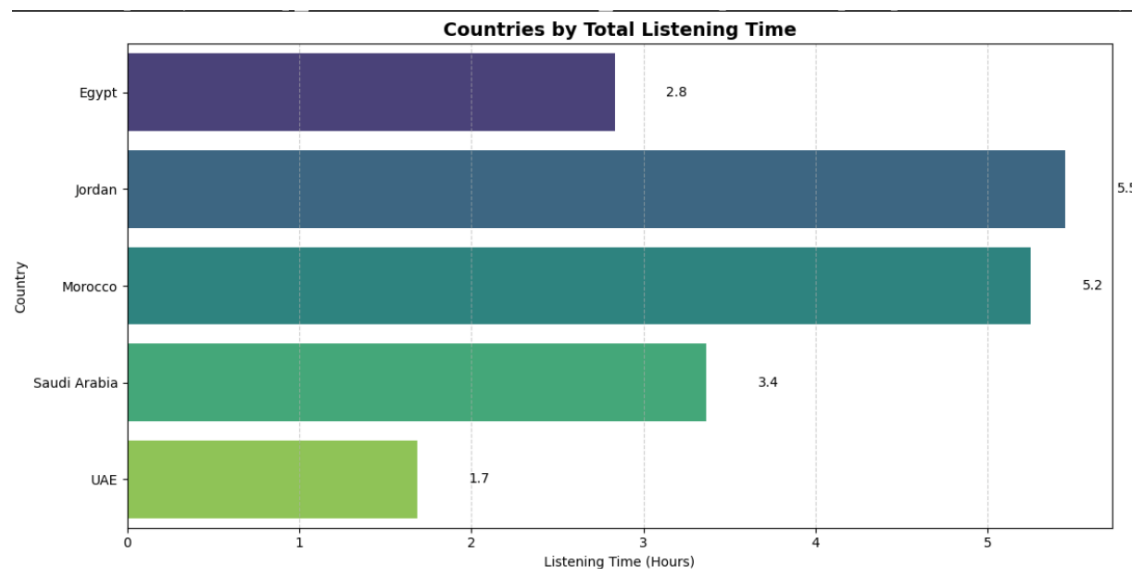
3.8 Short Sessions (Drop-off Detection)

Analysis: Listening sessions lasting two minutes or less were identified as short sessions, commonly signaling premature exits. The analysis was segmented by both country and gender:

- **By Country:** Egypt and Saudi Arabia have the highest number of short sessions (over 10 each), followed by Jordan and Morocco. The UAE recorded the fewest.
- **By Gender:** Males average nearly 2.5 short sessions per user, while females average fewer than 2. This suggests male listeners are more likely to disengage early.

Insight: Short sessions may reflect issues in episode structure such as lengthy intros, unclear value propositions, or poor alignment with listener expectations. High occurrence in specific countries may also point to localization or UX design problems.

Recommendation: To reduce early drop-offs:

- Strengthen the first 30–60 seconds of episodes with compelling hooks, clear summaries, or emotional cues.
- Experiment with trimming introductory segments and monitor engagement metrics.
- Consider A/B testing localized intros in high-drop-off regions like Egypt and Saudi Arabia.
- For male listeners, test more direct, fast-paced content formats or provide early previews to set expectations.

This data should also guide platform messaging, such as introducing "What to Expect" snippets before playback or adding skip-intro buttons.
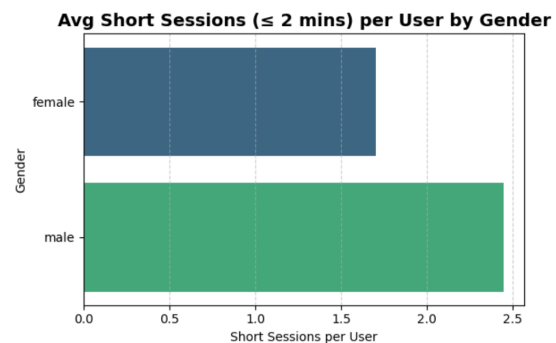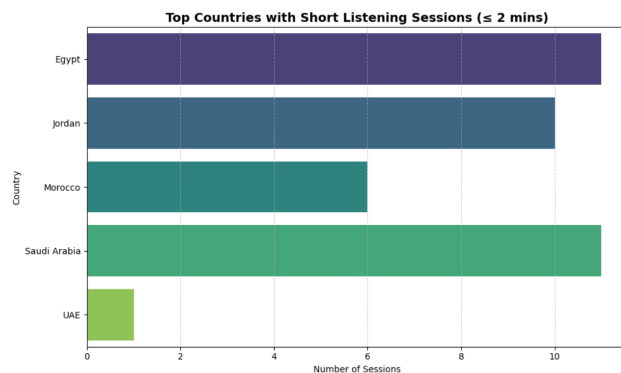
3.9 Zero-Duration Sessions (Technical/UX Issues)

Analysis: Zero-duration sessions, where playback was initiated but no time was recorded, were analyzed by country and gender. These are often associated with technical issues or accidental user behavior such as mis-taps or interface lags.

- **By Country:** Jordan shows the highest count of zero-duration sessions (6), followed by Saudi Arabia (3), Morocco (2), and Egypt (2). The UAE recorded none.
- **By Gender:** Female users had over twice as many zero-duration sessions (9) compared to male users (4), suggesting a possible discrepancy in platform responsiveness or user flow by demographic.

Insight: The presence of zero-duration sessions across multiple countries and disproportionately among female users may signal UX design flaws, compatibility issues, or performance problems specific to certain devices or OS configurations. It may also reflect a mismatch between user expectations and actual interface behavior (e.g., unclear play button feedback).

Recommendation:

Initiate QA testing focused on:

- Tap-to-play response accuracy
- Device/OS combinations prevalent in high-zero-duration regions (e.g., Jordan, Saudi Arabia)
- Gender-based UI/UX testing to ensure usability across user segments

Additionally, consider implementing event tracking to log failed or aborted playback attempts. This data can provide greater visibility into the sequence of user actions leading up to zero-duration sessions and help isolate bugs in the playback system or interface responsiveness.

10
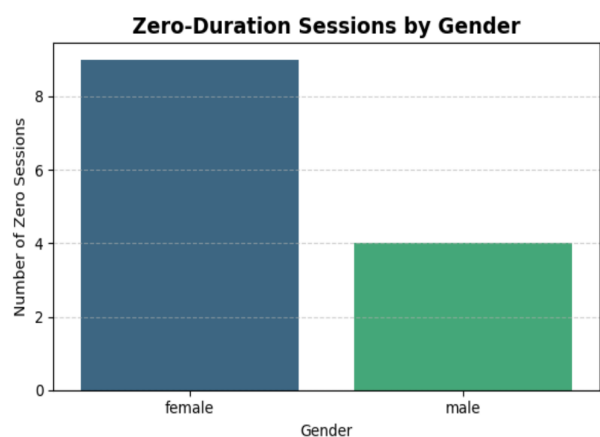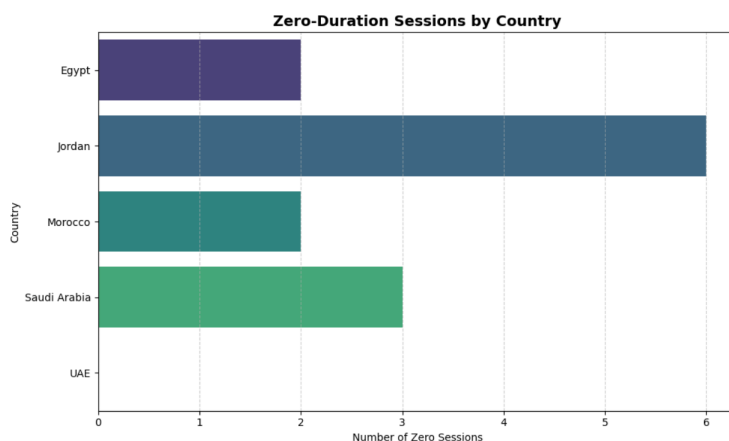
3.10 Basic Recommendation System

Analysis: Based on top categories per user, the model recommends unplayed episodes in the same themes.

Insight: Provides basic but meaningful personalization. Can evolve into dynamic recommendations.

Recommendation: Use this logic to create a "Because You Listened To..." section. Upgrade to ML model for scalability.