

# Analysis Report: Podcast Listening Time

This document summarizes the results of an exploratory data analysis on podcast listening behavior based on a dataset containing various features such as episode length, number of ads, genre, sentiment, and time of publication.

## Q1. What is the distribution of the target variable (Listening\_Time\_minutes)?

The distribution of Listening\_Time\_minutes appears to be right-skewed, as observed from the histogram and boxplot. This indicates that while most listeners tune in for shorter periods, a few episodes achieve significantly longer engagement.

## Q2. Are there outliers or skewness in numerical features?

The skewness scores of numerical features reveal varying degrees of skew. The boxplot and the use of an outlier detection method confirm that there are outliers present, especially in the listening time and possibly episode length.

## Q3. Which features are most correlated with Listening\_Time\_minutes?

A correlation matrix and heatmap show that Episode\_Length\_minutes is the most strongly correlated with Listening\_Time\_minutes. Other features have low or negligible correlation.

## Q4. How does Episode\_Length\_minutes affect listening time?

A scatterplot shows a positive relationship — longer episodes tend to have longer listening times. However, the relationship may not be strictly linear, and diminishing returns could be present.

## Q5. Do more ads decrease listening time?

A boxplot suggests that episodes with more ads tend to have slightly lower listening times, although the variation is not drastic. There might be a negative trend.

## Q6. Which genres get the most listening time?

A boxplot by genre shows variability across genres, but overall no single genre dominates listening time clearly.

## Q7. Does publication day impact listening time?

Boxplots of Listening\_Time\_minutes grouped by Publication\_Day show no strong trend. A one-way ANOVA test confirms that the day of the week is not a statistically significant factor ( $p > 0.05$ ).

## Q8. Does sentiment of the episode relate to listening time?

Violin plots show that sentiment categories (Positive, Neutral, Negative) do not significantly affect listening time. A one-way ANOVA test again supports this with  $p > 0.05$ .

**Q9. Does the genre of a podcast affect listening time?**

One-way ANOVA across genres yields  $p > 0.05$ , indicating that genre does not have a significant effect on how long users listen.

**Q10. Does the day of the week on which a podcast episode is published influence listening time?**

Again, ANOVA results show no significant difference across days ( $p > 0.05$ ).

**Q11. Does the time of day the podcast is published affect listening time?**

A one-way ANOVA for Publication\_Time also shows no statistically significant effect ( $p > 0.05$ ).

**Q12. Does the sentiment of a podcast episode significantly influence listening time?**

ANOVA testing shows no significant difference among sentiment types (Positive, Neutral, Negative), suggesting sentiment does not affect listening behavior.

**Q13. Is there a difference in listening time between weekday and weekend episodes?**

A t-test comparing weekday and weekend episodes finds no significant difference ( $p > 0.05$ ). Publishing on weekends does not significantly increase or decrease listening time.

### Summary:

Most traditional content variables (genre, sentiment, ads, day/time of publication) do not significantly influence listening time. Episode length is the most relevant predictor, and data indicates some outliers and skewness in listening patterns.