

Milestone 3 – Peer Review

1. Project Title

Predicting Follower Growth on Twitch: The Impact of Hours Streamed

2. Research Question / Problem Statement

- **Main Question:** Does the number of hours a streamer spends streaming predict the total number of followers gained on Twitch?
- **Hypothesis:** Streamers who spend more hours streaming tend to gain more followers due to increased visibility and audience engagement.

3. Peer Review

The peer review started with explaining to Xiao Li (the reviewer) my project problem statement and then briefly explaining the terms/columns that might be unknown to him, like “Twitch”, “Stream Time”, and “Watch Time”. Then I described my approach to this problem, that a simple linear regression model will be used to predict the followers gained based on the stream time, stream time being the independent variable and followers gained being the dependent variable.

The review continued with me explaining my analysis and results of the EDA, through plots like scatter plots, histograms, and boxplots, which so far had resulted in no identification of any linear (positive or negative) relation between the “stream time” and “followers gained”, which was not expected. The scatter plot was non linear/random and presented no conclusive evidence of any relationship between the two variables. Upon review, it was suggested to try plotting the relation between “percentage of followers gained” and “stream time” instead of “followers gained” vs “stream time”. Despite doing this, the results were not much different. The plot presented no explanation for Streamers with really low stream times having really high follower gained percentages or number of followers gained.

It was then suggested to filter for data with low stream time and high followers gained to identify if there is anything special about these data points. My explanation for this behavior was that those particular streamers were either really famous previously and only streamed from time to time, making their followers gained really high despite the short stream times. Another peer suggestion was that “stream time” was not the only variable influencing the followers gained and that there must be another variable that can help predict the followers gained. As such, a multivariable linear regression model should be considered.

These two suggestions, filtering the data and considering a multi-regression model were really helpful, as otherwise the model would have been trained on only one variable with limited generalization ability.