Assumptions

In this short exercise we make many assumptions about the data and its underlying meanings. Some of them, if overturned, may require us to start from ground zero, but most of them do not have dramatic consequences.

User Activities

- **Resources Clicked**: We assume here also that resources_clicked came exclusively from recommendations, as we do not have information on how a user came upon a resources. However, web analytics data could be helpful in the future to fine-tune the resources recommender.
- **Feedback Rating**: We assume feedback_rating for user activities reflects overall chat experience, and is not a rating of moderators.
- **Timestamp**: Throughout, we assume that this sample only has users from EST timezone. This is not a risky assumption practically speaking. If needed, we can split the analysis for timezones.

Recommendations

- We assume that each recommendation represents a unique result from the recommendation engine, rather than a unique piece of content. A content-centric dataset can be instrumental for analyzing recommendation effectiveness in the future.
- **Feedback Score**: Without specific information on the feedback survey, we interpret the feedback_score as a rating of how relevant the recommendation was, which reflects user preference.

Moderator Performance

- **User Satisfaction Score**: If the assumption about feedback_rating in the user_activity_data holds, we assume further that the user_satisfaction_score is the average satisfaction score from all users helped.
- **Time period and aggregation**: we assume that chat_sessions_moderated and avg_response_time are gathered over the same time period. However, much of our analysis will not stand if, say some data were aggregated over a year and some over a month; or if the time periods are staggered. If and when there's further information regarding time, we can conduct further analysis on performance based on tenure, time of year, etc.

Concerns

For moderators, We were able to that every user that is in recommendation_data is definitely in user_activity_data. However, there is still a significant amount of users in user_activity_data that have clicked on resources, for whom we cannot find the click through rate. The analysis can be more comprehensive if the dataset is complete with recommendation data for the same 1000 users in user_activity_data.