**Below data file is the raw data downloaded from the Qualtrics survey platform.**

* Main\_Personalization\_Preferences\_12-18-21-raw\_data\_Qualtrics.csv: Original file directly downloaded from Qualtrics, includes all 530 participants. Treatment and control group responses are recorded in separate columns coded with “T” and “C” respectively e.g. benefit\_4 variable is recorded as “11\_C\_benefit\_4” for the control group and “11\_T\_benefit\_4” for the treatment group.

**Below data files are processing files that are not directly used in the analysis.**

* Filtered\_Personalization\_Preferences\_12-18-21-raw\_data\_Qualtrics.csv: Three respondents were filtered based on the understanding check questions. At the very minimum, both the treatment and control group were expected to understand that their data would be used to recommend songs to themselves. 3 participants responded “no” to this question, and were filtered out.
* Cleaned\_Modified\_Personalization\_Preferences\_12-18-21-Qualtrics.csv: Using MS Excel tools, a column to identify control and treatment groups was added. Control and treatment columns were consolidated together e.g. column benefit\_4 includes answers for both the treatment and control group, identified by the column “group”.

**Below data files are directly used in the data analysis.**

* Cleaned\_Modified\_Substituted\_Personalization\_Preferences\_12-18-21-Qualtrics.csv: Main data set used in the analysis. Some blank cells (e.g. participants who never used PSPs in real life skipped relevant demographic questions) were substituted with NAs.
* Modified\_Platforms\_Used\_Personalization\_Preferences\_12-18-21-Qualtrics.csv: A version that is used only for demographic questions relevant to participants’ experience with real life PSPs e.g. which PSPs they use in real life. These answers were recorded as a list in the original data folder and were tokenized in this modified version e.g. original answer “Netflix, Hulu” was tokenized as “Netflix:1, Hulu:1, Spotify:0…”