**Exploratory Data Analysis (EDA)**

**Background**

What is EDA?

Exploratory Data Analysis (EDA) is the foundation step before we perform any data analysis. Not only it reveals the true nature of data, but it also serves as an open-minded exploration to uncover valuable information.

Why is EDA important?

First, we will need to confirm if the data makes sense in the context of business problems Next, by getting the topline summary of the data (i.e. using descriptive statistics), we can uncover the underlying data quality issues, hence able to detect outliers & anomalies. Most importantly, understand data patterns and the correlation between variables. Finally, we can drop redundant columns to efficiency and derive new variables for further insights.

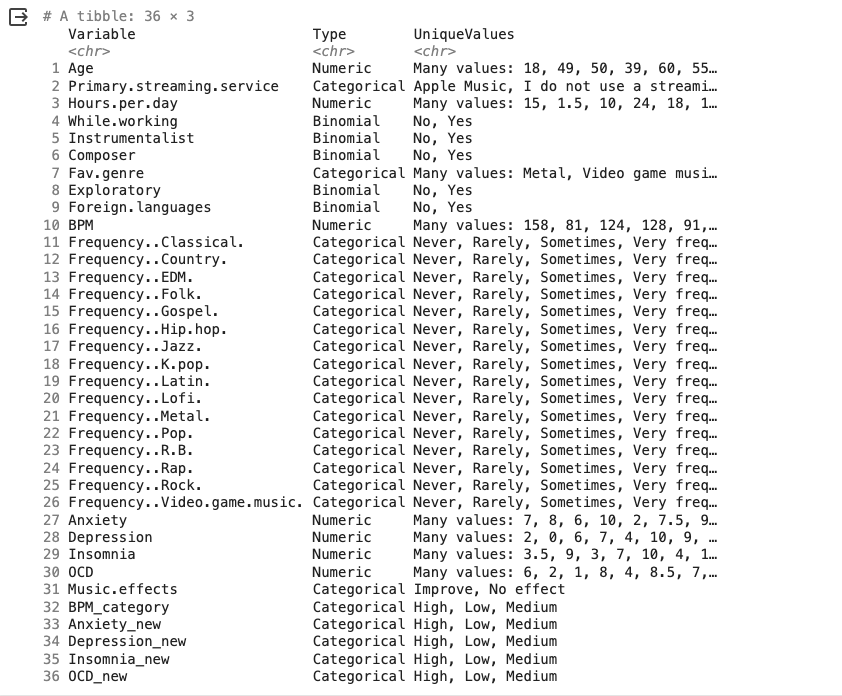
Types of EDA

For univariate data (single variable), the data can be either categorical or quantitative. For categorical data, we can use either tabulation (non-graphical) or bar chart (graphical) to represent them. As for numerical data, we can use either descriptive summary statistics (non-graphical) or histogram (graphical) to represent them.

For data that involves multiple variables (multivariate data), similarly, we can use a cross-tabulated report (non-graphical) to present categorical data, and scatterplot (graphical) to represent numerical data.

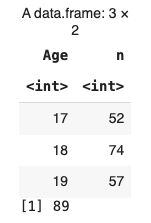
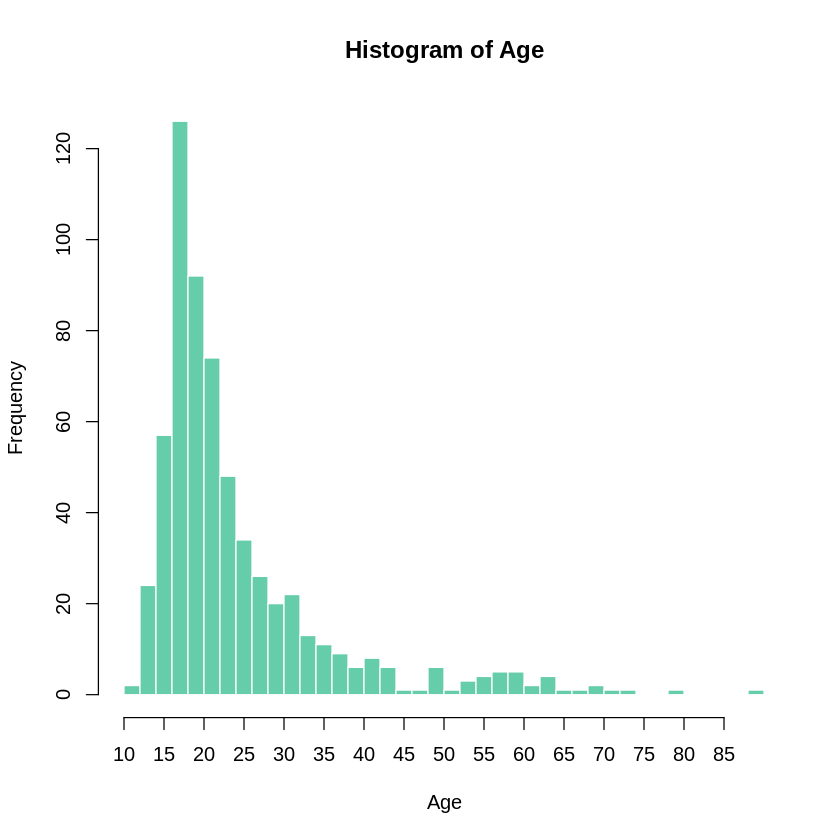
[Link to original code](https://colab.research.google.com/drive/1tBEX1cKcBxCb89-EFsdMdYoenb7Dhnaz?usp=sharing)

Overview of Data Types



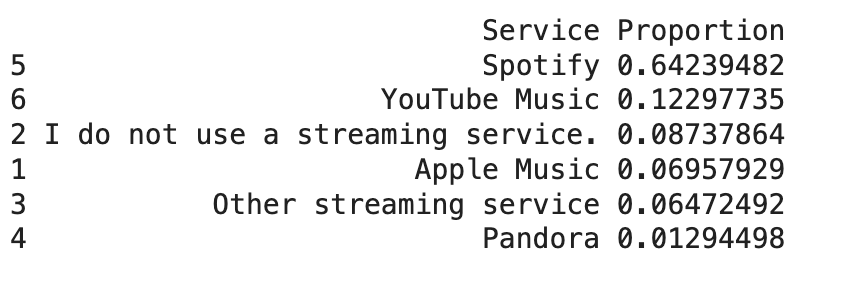
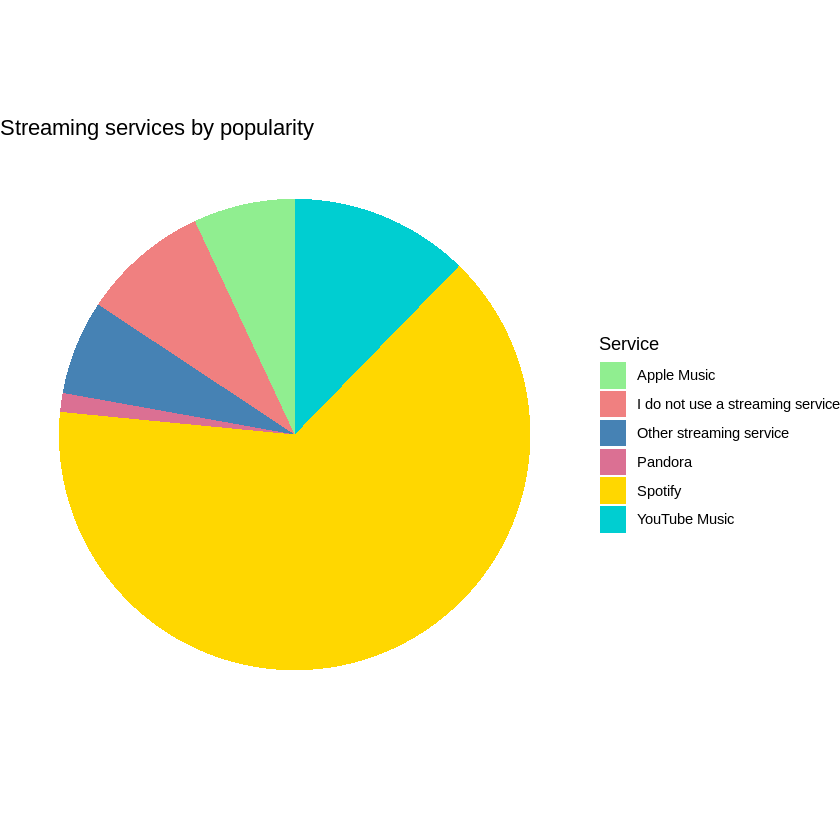
EDA: Questions, Visualisations & Insights

1. Age



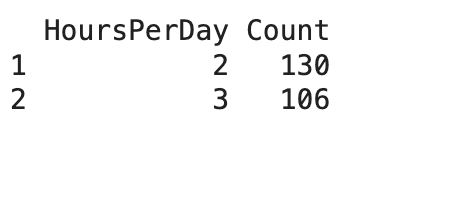
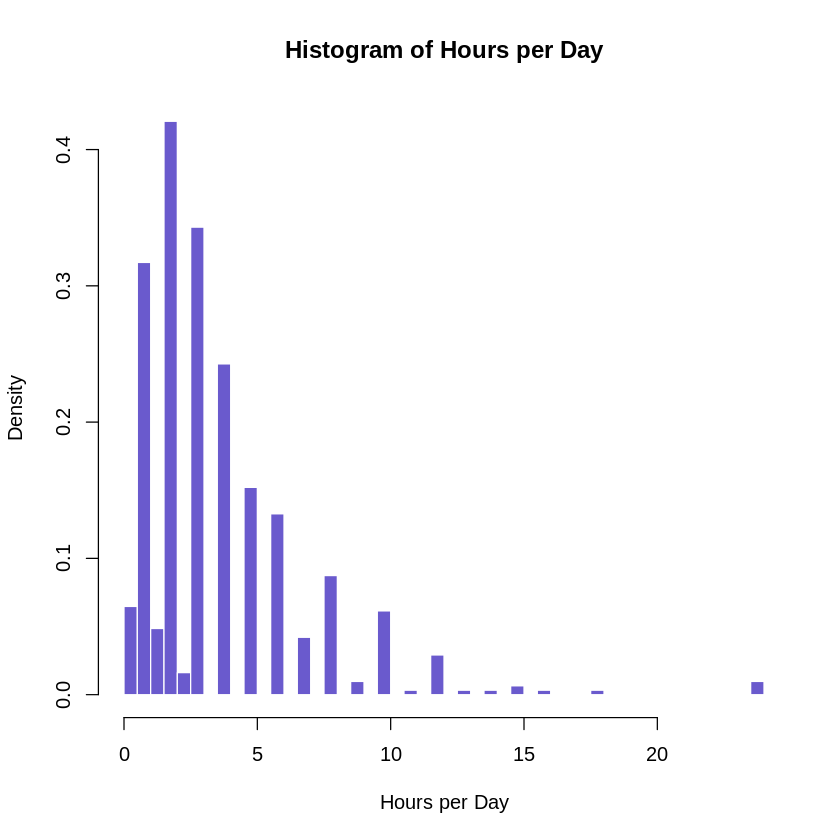
* Histogram, showing distribution of age
* Participants (Aged: 17~19) as per table
* Max Age is 89

2. Streaming Service



* Pie cart, showing percentage of streaming device
* Top device
  + Spotify (65%)
  + YouTube Music (12%)
  + No streaming service (8%)
  + Apple Music (7%)

3. Hours per day



* Histogram, showing hours used, mostly < 5hours
* Exact hours: Mostly 2 hours

