Instructions:

The below task is to check your general familiarity with dashboarding, databases, logic, and approach to solve simple ETL and SQL tasks. It will also give you a glimpse of data from the TV streaming world and help you decide your interest and curiosity for the same. It should not take you more than 3 hours to complete this.

Send us all your code and answers back as a zipped file. Feel free to make assumptions (please state them explicitly)

All the best and we hope you enjoy the task!

Data Description:

Whenever a user watches something on the Zattoo platform using any application on any device a unique session is created . This session has the metadata information about the watch history of the user like device used, time at which the watch event started, region, live channel, duration etc.

A sample file with 200,000 such watch sessions data is provided . Assume the sessions were generated on the same day.

File description:

raw_watch_sessions.csv

The file has following columns:

1: user_id : User ID of the customer. (Integer)

2: regionid : Region id from which watch request generated. (Varchar(4))

3: duration : Duration of the watch session (in seconds). (Integer)

4: channel: Name of the live channel watched. (Varchar(64))

5: unique_sess_id : Unique ID of the watch session. (Varchar(32))

6: app_id: Id of the application used to watch. (Integer)

7: device: Name of the device. (Varchar(64)
8: cats: Category of the program being watched. (Varchar(64)

9: zattoo_internal_id: internal zattoo ID (Varchar(320))

9: backend_id: zattoo backend ID (Varchar(100))

Problem

Please be aware of query optimization, performance and possible costs associated with the processing of large quantities of data.

Task 1: Write an SQL query that answers the following :

- Which channel is most popular per region?
    - What is the device share for this channel?
- What is the maximum duration per channel per user id?
- What are the top 3 most popular channels per category (cats)?

Task 2: Using EDA (Exploratory Data Analysis), what kind of statements can you make about the data? Use a tool of your choice to solve this task and draw conclusions about the data set. We would like to see that you can make simple dashboards that are intuitive for stakeholders