

|  |
| --- |
| EPAM Systems, RD Dep.  INTRODUCTION TO DATA QUALITY |
| Exit Task |

# Business requirements

Input data set contains 2 csv files: googleplaystore, googleplaystore\_user\_reviews



Business description and requirements for source data can be found in Google\_Play\_Store\_Data\_Set\_Requirements.xlsx



NOTE! AFTER WORK WITH ENCLOSED FILES DON’T FORGET TO SAVE (SAVE AS) CHANGES BEFORE CLOSING. OTHERWISE YOUR CHANGES WILL BE LOST.

# TASK

Data Engineering team is working on ETL to upload data from source files into Database (PostgreSQL) applying necessary rules to remove invalid rows. Targets for these files will be 2 tables in a DB: google\_play\_store and google\_play\_store\_user\_reviews

Your task is to implement data quality check list to ensure that data is fully and correctly processed. Please use the following excel format\* for this purpose:



\*- Columns **#, Table, DQ check, Column Name, Description** are mandatory.

Column **SQL** is optional, you can fill it in for additional points.