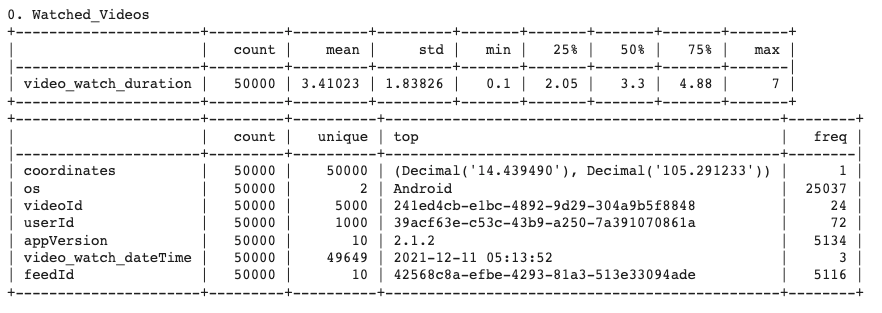
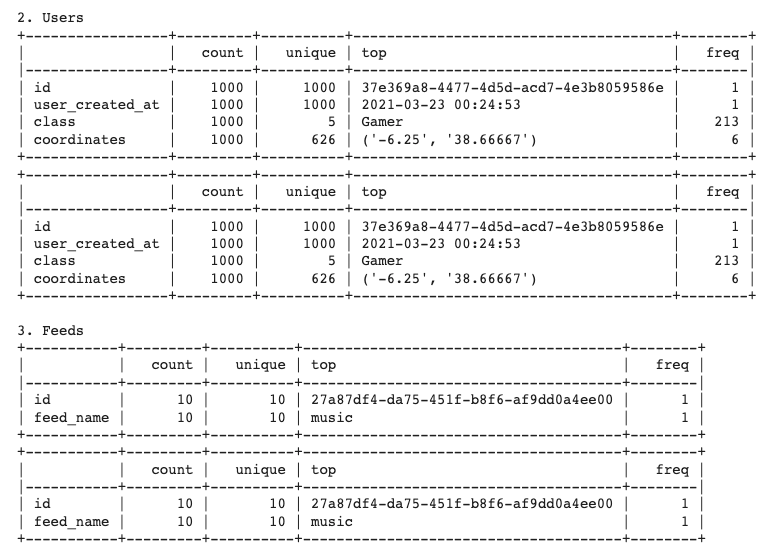
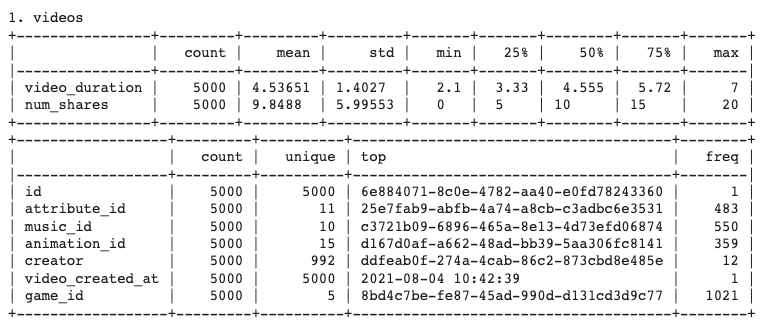
* Product recommendations based on the key findings from EDA

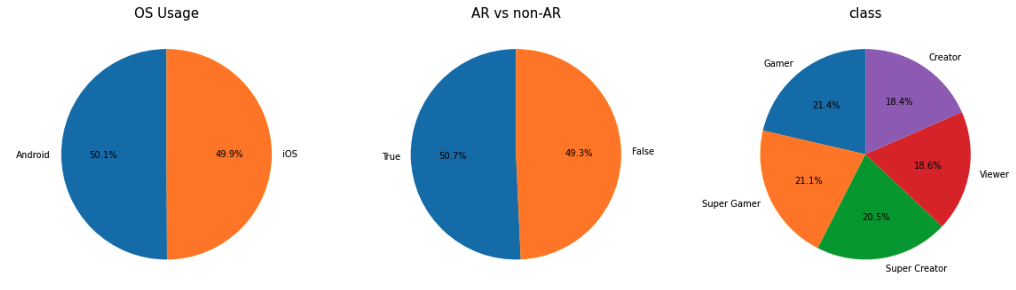
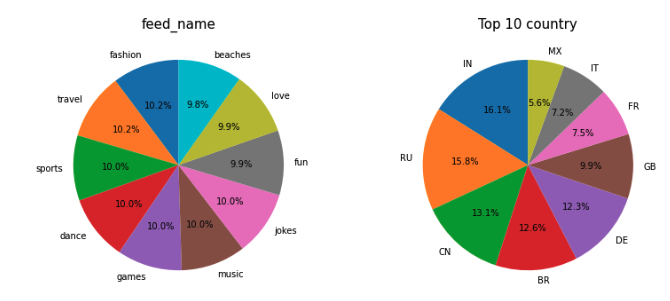
(Refer to “Correlation data analysis for product recommendation” in the next page)

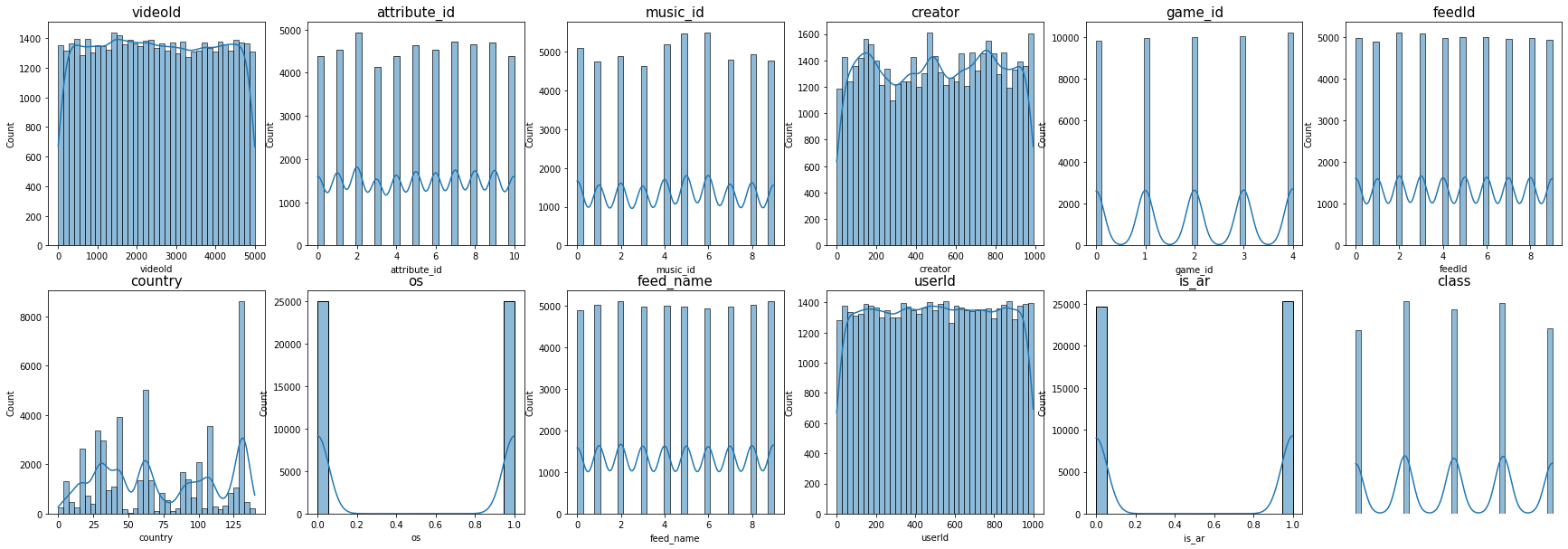
* + Encourage users to create long duration video : 2~4 minutes
    - “Video watch duration” and “Number of share” are the key parameters that indicate the success of the products.
    - “Video watch duration” is positively correlated with “video duration”.
    - “Video watch duration” is positively correlated with “video duration” and video watch duration
  + Recommend specific music and games to users based on the user’s video profile in the product. Most impatient users would take the recommendation
    - Based on the correlation analysis, video id is correlated with game id, music id. Using this information, it is recommended to promote certain music and games to users
  + Minor recommendation: Product is mostly popular in India, Russia, China and a few European countries. There is room for expanding markets in western countries. Recommend to increase the marketing effort in northern America and more European countries.
* Analytic details (Google Colab with Pandas dataframe)
  + Step1: Basic information check based on pandas dataframe
    - Create 4 dataframes based on 4 csv files and run the first level data sanity check

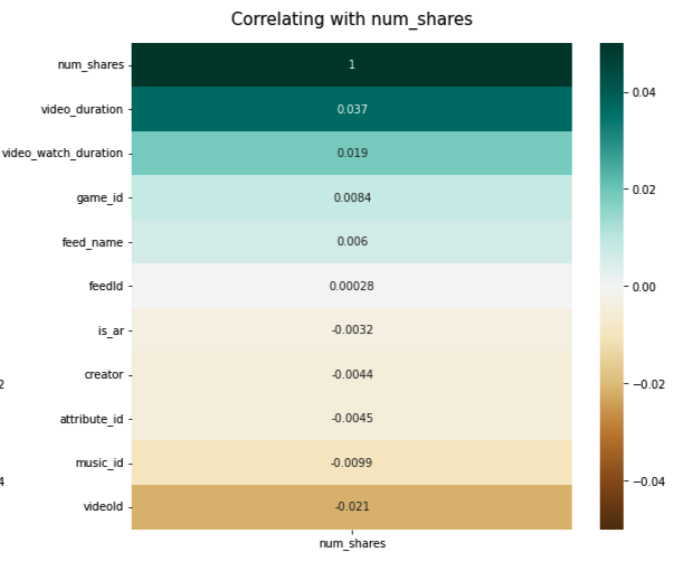
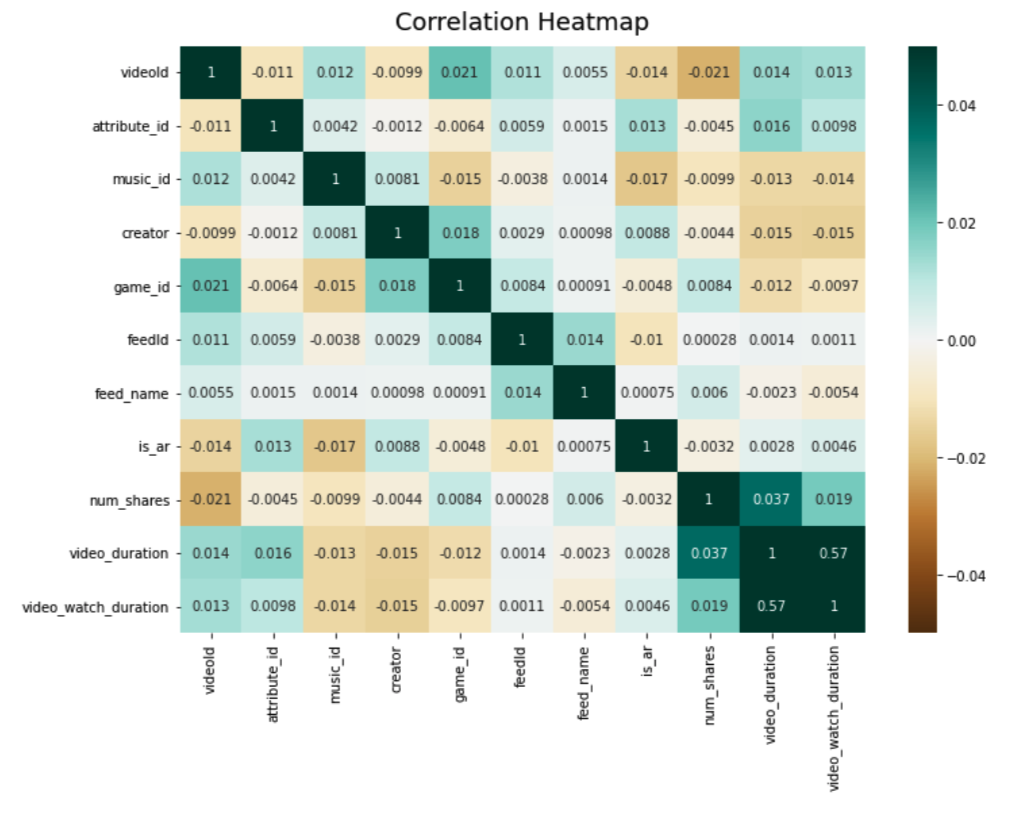
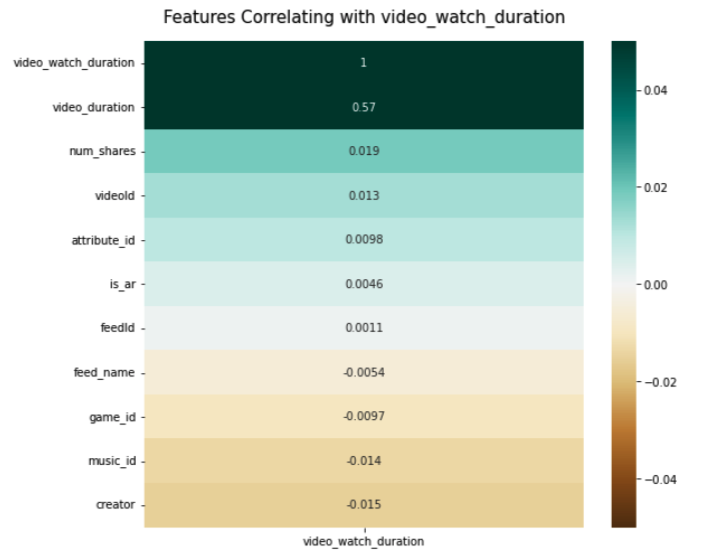
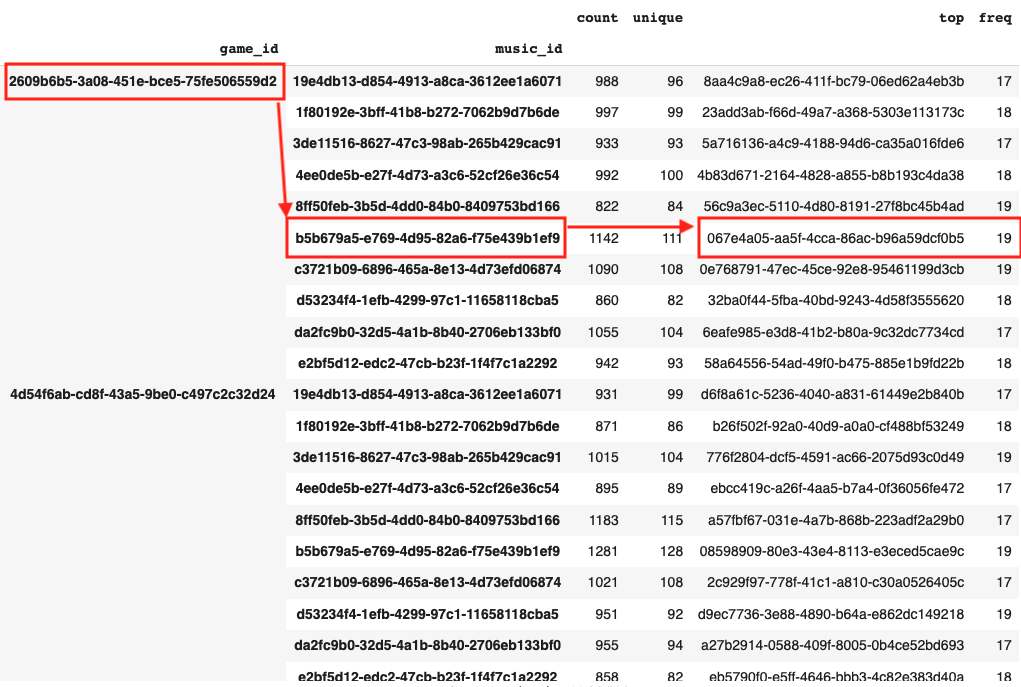




* + Step 2: Merge data based on the inner joint.
    - All data frames have common columns that enables merging all data frames into one universal frame
  + Step 3: Run data visualization and correlation study
    - Simple uni-variable data analysis
      * Pie chart





* + - * Histogram
    - Correlation data analysis for product recommendation
      * 1st plot: Correlation among all the variables. Categorical variables are coded to unique integer based on pandas “dataframe.cat.codes”
      * 2nd plot: Correlation of Video watch duration as dependant variable
      * 3rd plot: Correlation of number of shares as dependant variable
      * Dataframe analysis 

based on groupby

([“game\_id”, “music\_id”])

The operation reveals that certain game id and music are more tightly coupled with a specific video ID.

Product can recommend game and music based on the user data

* Next steps
  + Change the data sampling method to capture more realistic data distribution. Currently all variables show the uniform distribution and it affects the correlation results. Gather data based on the stratified sampling and re-run the EDA
  + Build game and music recommendation algorithm development based on the user data : Classification model based on machine learning algorithm development