Command Line – Task 7

BY ROHAN RAJ

Task (7)

→ Now it's your turn.

1. Select a dataset (you can use the links in Moodle)
2. Download the file using `curl`.
3. Analyze the dataset using the commands learnt in this lesson.
4. Some ideas:
   - How big is the dataset? How many lines?
   - Filter specific information
   - Compute Min/Max Values, Average Values
5. Use three new commands (not introduced in the lecture)
6. Sum up your results in a small presentation.

Selecting a Dataset

- https://www.kaggle.com/datasets/wardabilal/spotify-global-music-dataset-20092025

Downloading the file

- #!/bin/bash
- curl -L -o ~/Downloads/spotify-global-music-dataset-20092025.zip https://www.kaggle.com/api/v1/datasets/download/wardabilal/spotify-global-music-dataset-20092025
- unzip ~/Downloads/spotify-global-music-dataset-20092025.zip

Analyzing the dataset

- Number of lines, words and character count: wc track_data_final.csv
- Tracks filtered as getting only getting from Britney Spears: grep "Britney Spears" track_data_final.csv
- Most popular track: awk -F',' 'NR==1 {print; next} {if ($4 > max) {max=$4; line=$0}} END {print line}' track_data_final.csv
- Longest track (max duration): awk -F',' 'NR>1 {if ($5 > max) {max=$5; line=$0}} END {print line}' track_data_final.csv
- Most followed artist: awk -F',' 'NR>1 {if ($9+0 > max) {max=$9; line=$0}} END {print line}' track_data_final.csv

New commands

- To use those commands, we need to install csvkit using "pip install csvkit"
- Once installed, we can use some csv functions like:
- Render a CSV file in the console as a Markdown-compatible, fixed-width table: csvlook track_data_final.csv
- Sort CSV files: csvsort track_data_final.csv
- Print descriptive statistics for each column in a CSV file: csvstat track_data_final.csv
- These commands were introduced by Christopher Groskopf (2020). More information on https://csvkit.rtfd.org

Thank you