PLAYING AROUND TED TALK DATASET

Data Source: https://www.kaggle.com/datasets/rounakbanik/ted-talks

Context:

The dataset contains information about all audio-video recordings of TED Talks uploaded to the official TED.com website until September 21st, 2017. The data includes information about all talks including the number of views, number of comments, descriptions, speakers, and titles.

Description of columns:

comments: The number of first-level comments made on the talk

description: A blurb of what the talk is about

duration: The duration of the talk in seconds

event: The TED/TEDx event where the talk took place

film_date: The Unix timestamp of the filming

languages: The number of languages in which the talk is available

main_speaker: The first named speaker of the talk

name: The official name of the TED Talk. Includes the title and the speaker.

num_speaker: The number of speakers in the talk

published_date: The Unix timestamp for the publication of the talk on TED.com

ratings: A stringified dictionary of the various ratings given to the talk (inspiring, fascinating, jaw dropping, etc.)

related_talks: A list of dictionaries of recommended talks to watch next

speaker_occupation: The occupation of the main speaker

tags: The themes associated with the talk

title: The title of the talk

url: The URL of the talk

views: The number of views on the talk

Solve these questions

1. Read the ***ted.csv*** dataset in your Google Colab as ***ted***.
2. Determine the size of ***ted***.
3. Display the last 3 rows of ***ted***.
4. Identify the data type of each column of ***ted*** and determine if there are any null values in them.
5. Determine whether ***ted*** contains any missing observation, if so, identify it.
6. Which three talks provoke the most online discussion?
7. Generate a histogram and a boxplot to visualize the distribution of ***comments.*** You can make some modifications for ease of decision-making.
8. Visualize the number of ted talks that took place each year.
9. What are the "best" five events in ted history to attend?
10. Unpack the ***ratings*** column and determine the frequent ratings for the first talk.
11. Count the total number of ***ratings*** received by each talk and print the minimum, maximum, mean, standard deviation, and quartiles for the same.
12. Identify which occupations deliver the funniest and least funny ted talks on average. Also, reanalyze the funny rate for occupations occurring at least five times.

Important Guidelines:

1. Group has to prepare one Presentation with 5-10 Slides answering all the questions asked above (maximum 10 mins).
2. Slides should include the followings (1-2 Slide(s) for each): Project Title and Student Names, Problem Statement, Data Description and Source, Answers to the Questions asked (2-4 slides), Name of the Methods used, What did you learn from this.
3. Present the project in front of a Jury followed by Question and Answer Session. Evaluation will be based on your work, presentation and answers to questions raised by the Jury.

BEST OF LUCK !