In this mini-project, you will analyze data from the Google Play Store. Below, you can familiarize yourself with the data description. We have broken down the tasks into steps to make them easier to check. Some tasks also include small hints. Good luck! :)

**Data Description**

● **App** — The name of the application.

● **Category** — The category the application belongs to.

● **Rating** — The user rating for the application.

● **Reviews** — The number of user reviews for the application.

● **Size** — The size of the application.

● **Installs** — The number of user downloads/installations of the application. ● **Type** — Whether the application is paid or free.

● **Price** — The price of the application.

● **Content Rating** — The age group the application is targeted at.

● **Genres** — The genres the application belongs to.

● **Last Updated** — The date of the last application update in the Play Store. ● **Current Ver** — The current version of the application in the Play Store. ● **Android Ver** — The minimum required Android version.

**Tasks**

**Task 1:** Save the first 3 rows of the DataFrame to the variable data\_head and the last 3 rows to data\_tail, respectively.

**Task 2:** Save the number of columns and rows in the DataFrame to the variables n\_col and n\_row, respectively.

**Task 3:** Now, let's see how many unique applications are in our data (the App column).

**Task 4:** Let's count the number of missing values in the application rating column. Save the number of missing values in the Rating column to the variable rating\_missing.

**Task 5:** Let's learn how to combine DataFrames and get acquainted with the concat function. Suppose we want to create a DataFrame that includes: the first three rows of the original DataFrame, rows 6-8 (inclusive), rows 16-19 (inclusive), and the columns: App, Size, Genres, Current Ver in that order. Essentially, this will look like combining three smaller DataFrames, which are slices of the original data. As your answer, upload the resulting DataFrame in CSV format. When saving the final DataFrame, the index column should be kept.

**NB!** Remember zero-based indexing and the rules for including/excluding boundaries when slicing.

**Task 6:** For further analysis, we need to remove duplicate applications (from the App column) and reset the index. Save the result as a DataFrame in the variable unique\_playstore.

**Task 7:** In this task, the column names need to be brought to a standard format — all letters should be lowercase, and spaces should be replaced with underscores. Methods that might be useful:

● rename

● str.lower

● str.replace

● columns

**Task 8:** Let's look at the proportion of paid and free applications. As your answer, provide the proportion of free applications. Round the answer to two decimal places, using a period as the decimal separator.

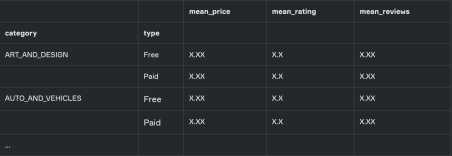
**NB!** Don't forget to remove duplicates before calculating the proportion!

**Task 9:** From the DataFrame, select only the applications that belong to the EDUCATION category and have more than 1000 user reviews. Save the new DataFrame to the variable education\_playstore. The data has been cleaned of duplicates, and the column names have been transformed according to the previous tasks and saved in the playstoreDataFrame.

**NB!** Don't forget to reset the index in the resulting DataFrame!

**Task 10:** Let's remove the extra symbols from the price column (price) to make it easier to work with, and convert the data to the float type.

**Task 11:** For the final task, you need to create a **pivot table** from the data. The resulting table must match the template shown in the photo:



The columns must have the corresponding names, and the table values should be rounded to the specified number of decimal places. As your answer, upload the resulting table in CSV format (using a comma , as the separator).

**NB!** You will need to search online for "Pandas pivot table".

**NB!** The table should be created from the data with duplicates removed.

**Task 12:** The Installs and Size columns are not in a numeric format. Clean both columns to make them usable for calculations.

**Task 13:** The last\_updated column contains date information. Let's use it to find out how many apps were updated each year.

**Task 14:** Let's find the top 5 most expensive app categories.

**Task 15:** The genres column can contain multiple categories for a single app (e.g., "Art & Design;Creativity"). This format is not ideal for analysis. Your task is to determine the most common (top 10) *individual* genres. Sort them in the final table by their