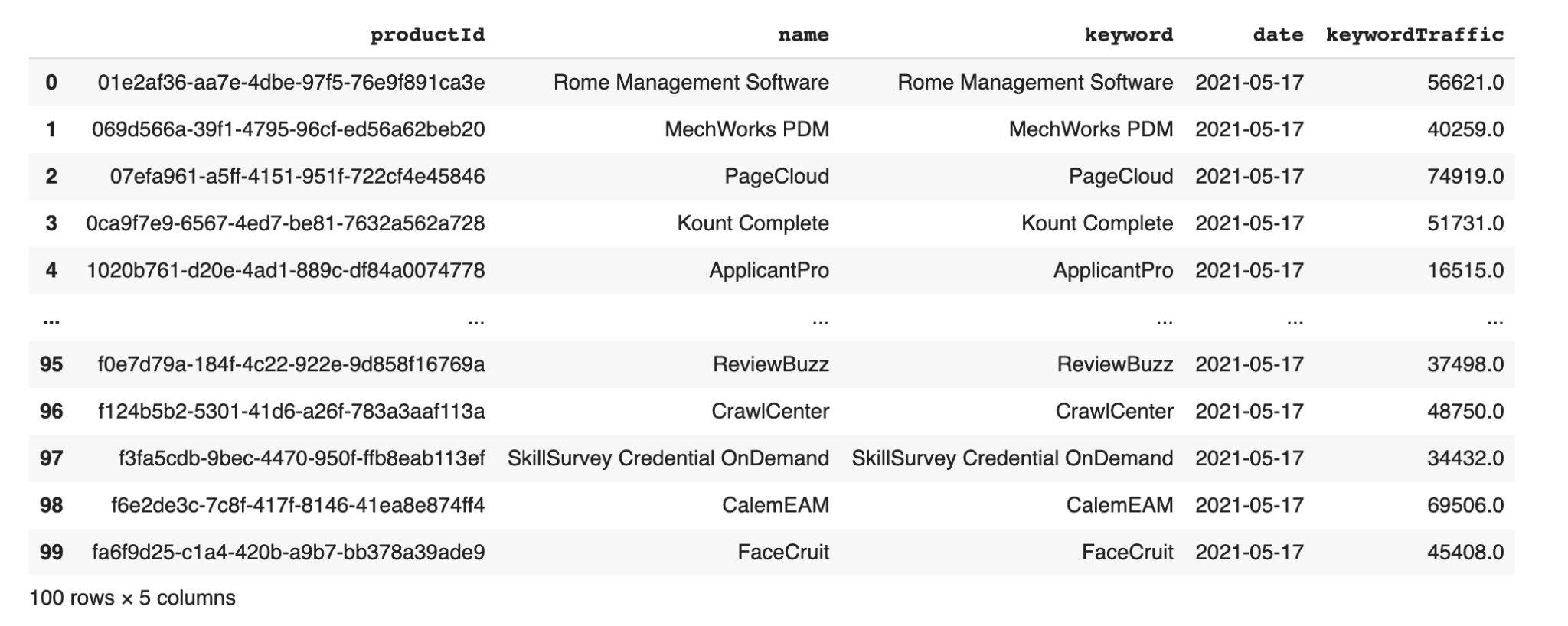
# AIMultiple Data Analysis Assignment

Use <https://colab.research.google.com/> to create your code and comment precisely on your code before submitting the assignment. Create a new notebook and share your code with us after completing the assignment. There will be three tasks in this assignment where each of them will assess your ability to use a different function of python, pandas or sql. We’re aware that there are an infinite number of solutions when it is a coding challenge, but we will ask you to stick with libraries such as pandas, numpy, sqlalchemy, pandasql, datetime and random. If you think that there are better libraries that can be used for this challenge, feel free to use them, but your outputs should be pandas dataframes just like the examples below.

## Task 1

Connect to our demo database with the help of the following instructions: <https://docs.google.com/document/d/1yLxR0BvVDkwaA2BY5kVECAk0anOnoMuwuh2WsMuxU8k/edit?usp=sharing> and import necessary tables to python. (names of necessary tables for these tasks: products & keywords) In this task we ask you to prepare a data frame consisting of IDs of products, names of products, keywords that describe the product (you should retrieve this information from the keywords table and merge products and keywords tables using productId), date of the record and keywordTraffic. You can fill the keywordTraffic column with random numbers between 1-100000.

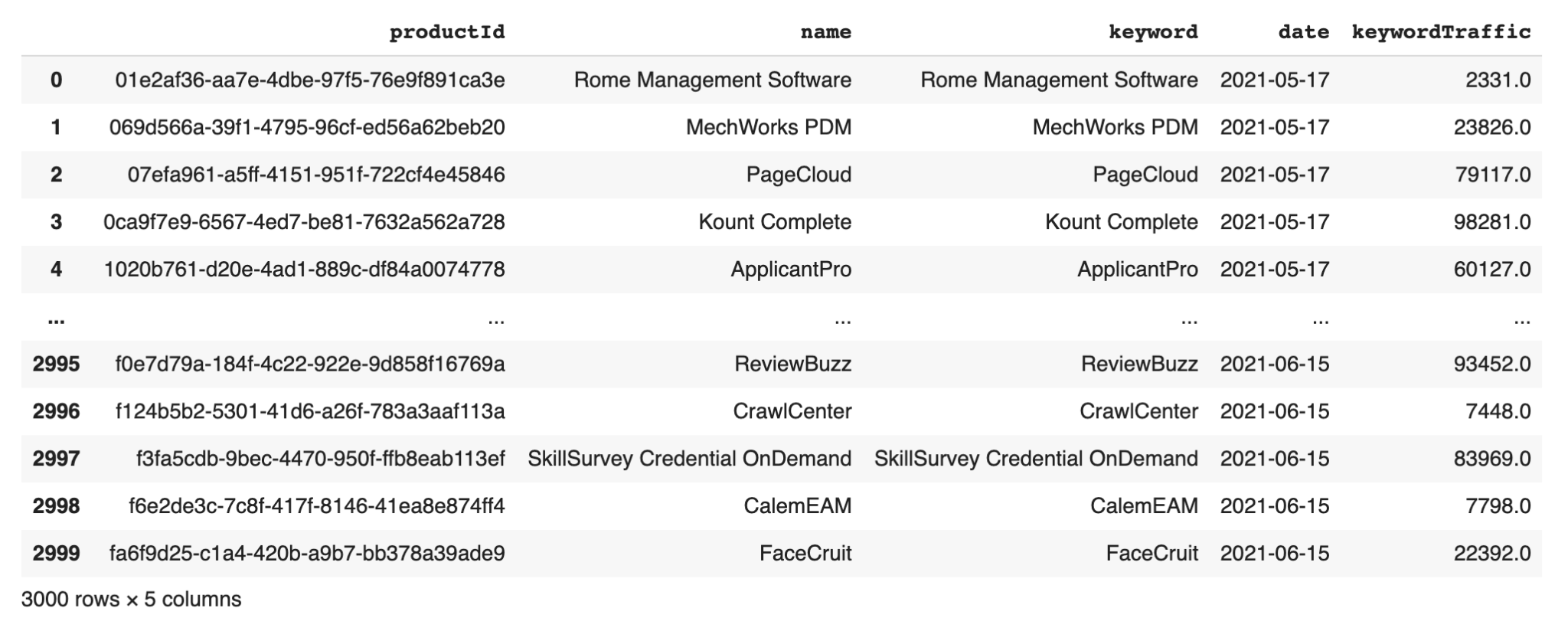
There will be 100 records in the output table as our demo dataset has 100 products and the output table should look like the following:



## Task 2

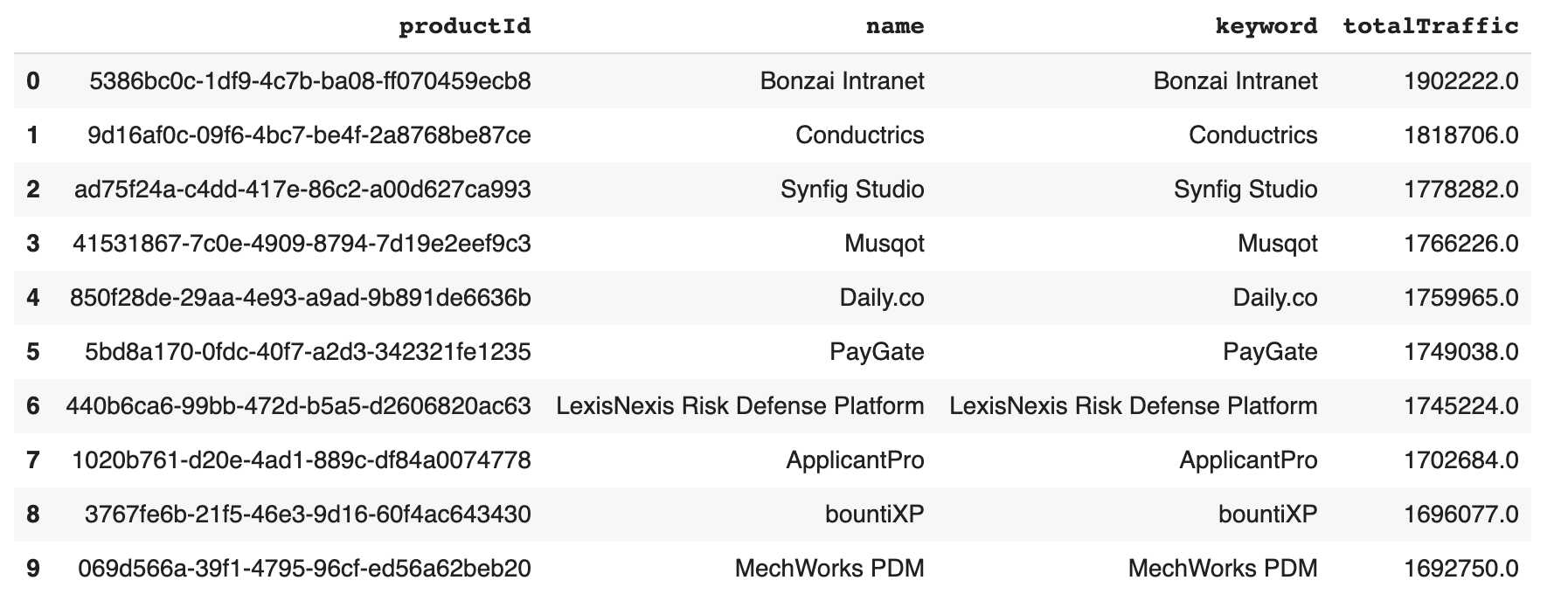
The output of the first task should give you records that belong to a single day. Use the format of the output from the first part and generate records for the next 30 days for each product. You should generate random keywordTraffic for each product for each day. Create a new table in the database with your name (e.g. johndoe\_traffics) and insert the results to the table.

The output data frame should look like the following:



## Task 3

Find top 10 products that had the most keywordTraffic over the 30 days interval in the dataset generated in Task 2. The output should look like the following:



## Task 4

Initiate an exploratory data analysis on the table you created in task 2. Find the mean, standard deviation, min, max, 1st, 2nd, 3rd quartiles. Plot the probability density function of keywordTraffic by using the values generated in task 2. Finally, create a model that estimates keywordTraffic of each product a week ahead based on the historical data.