**Documentation Flow**

**Project - 2**

**EDA-1**

**A) Problem Statement->**

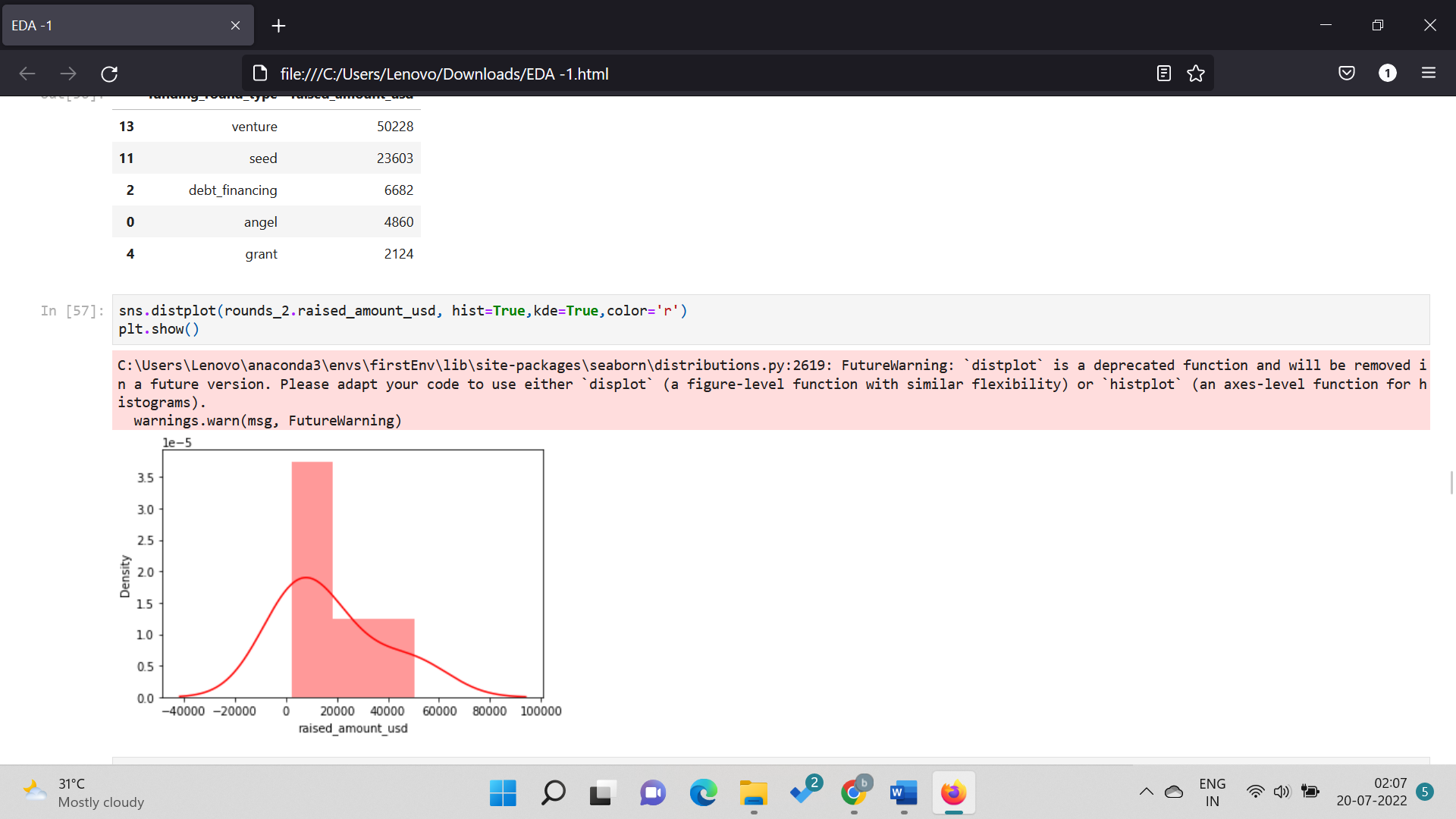
We are provided with real time companies dataset and we have to perform EDA on that to find out solution to some needed excel questions which would be beneficial for the company to make some decisions in future.

**B) Overview ->**

Firstly we did the analysis of the data and then checked for null values and cleared them, then we found out the solution to the questions using different command and then later we made some plots to visualize the data in a better way.

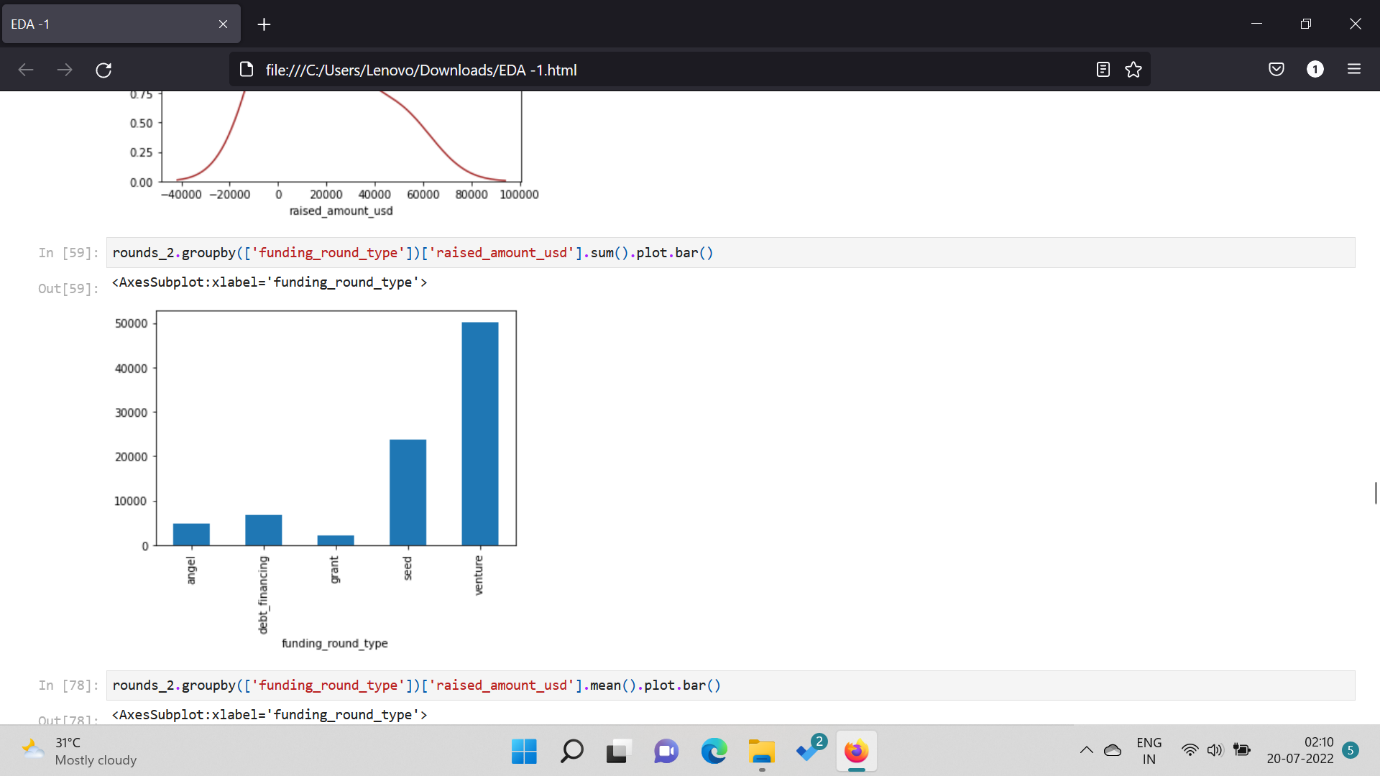
**C)Step-wise (plots, inferences, preprocessing)->**

**Step-1->**

****

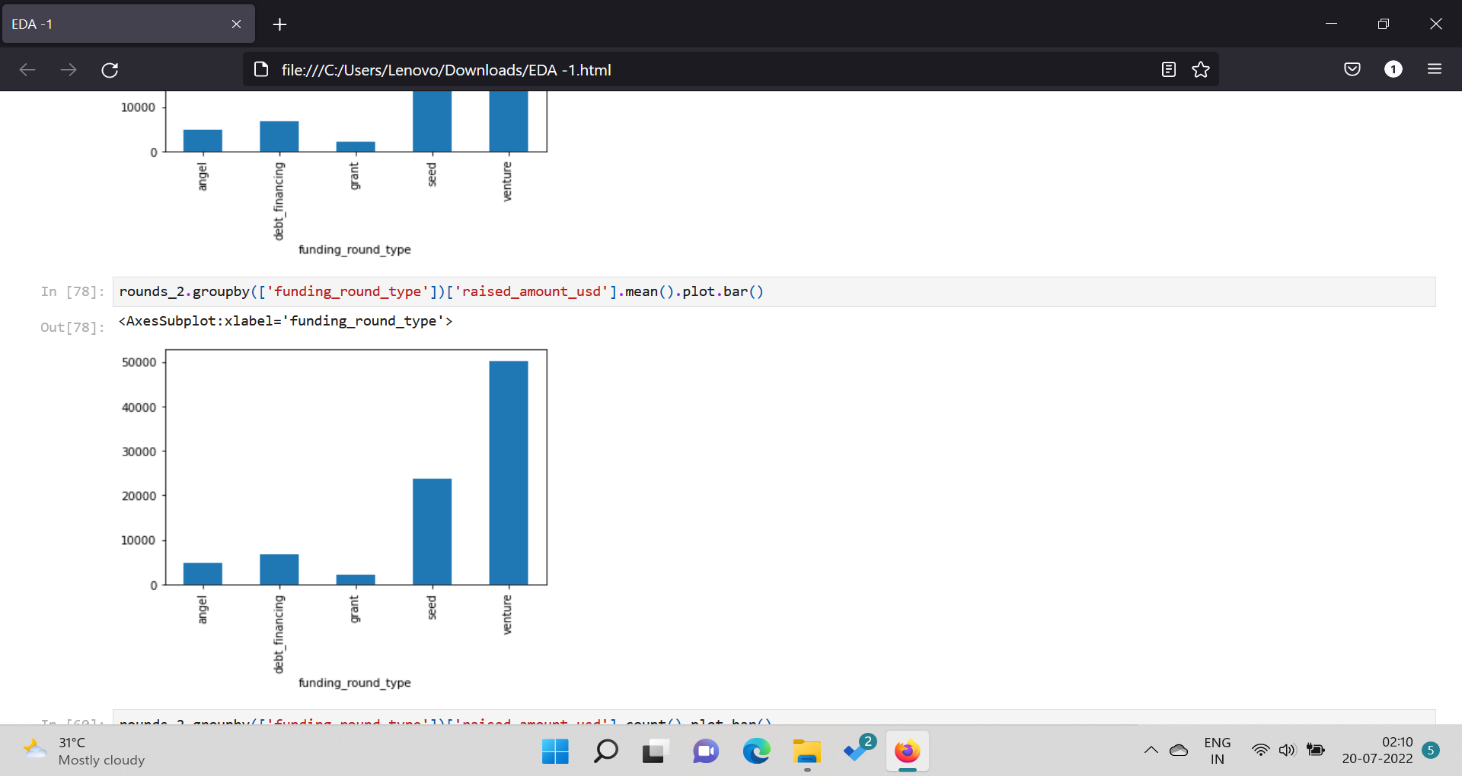
**Inference from this ->** how is our plot distributed, kde depicts the probability density function of the continuous or non-parametric data variables

**Step-2->**



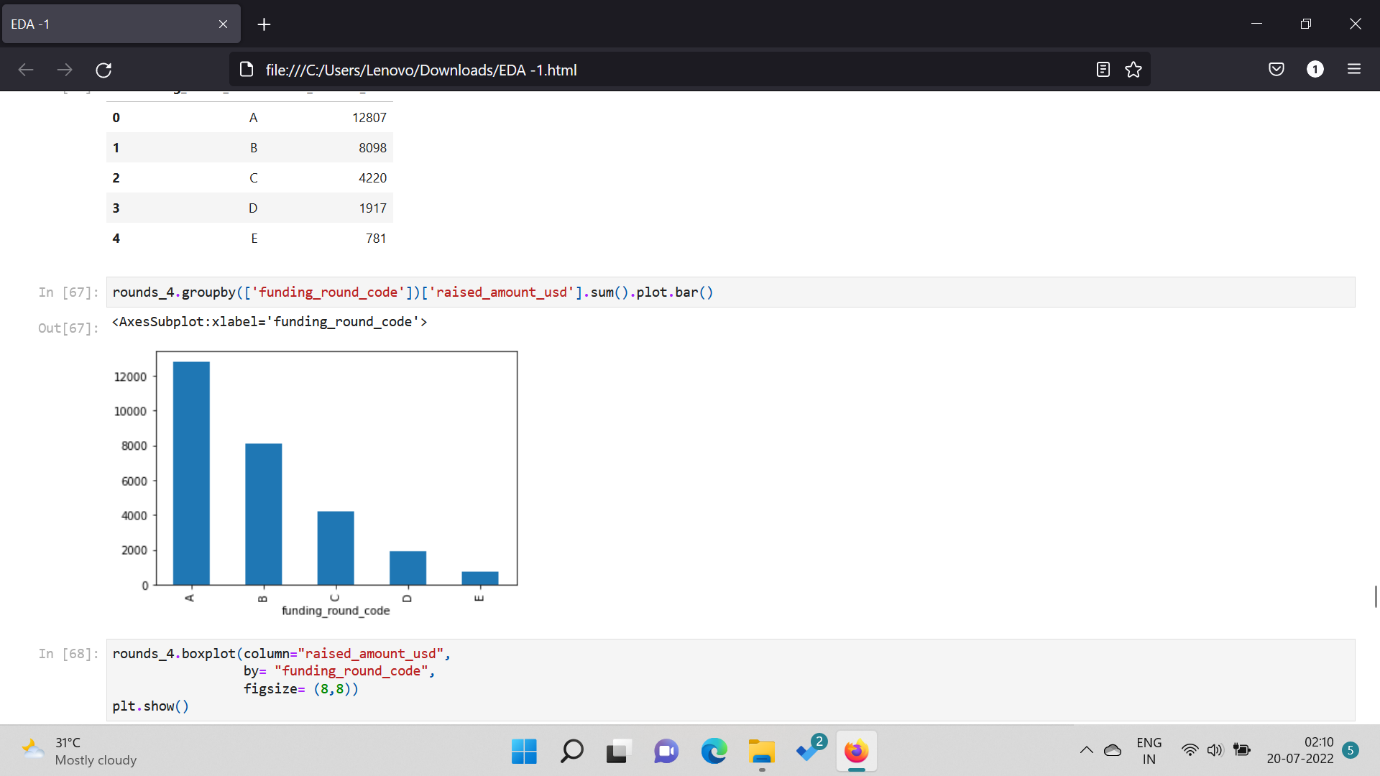
**Inference from this ->** how is funding round type related with raised amount usd.

**Step-3->**



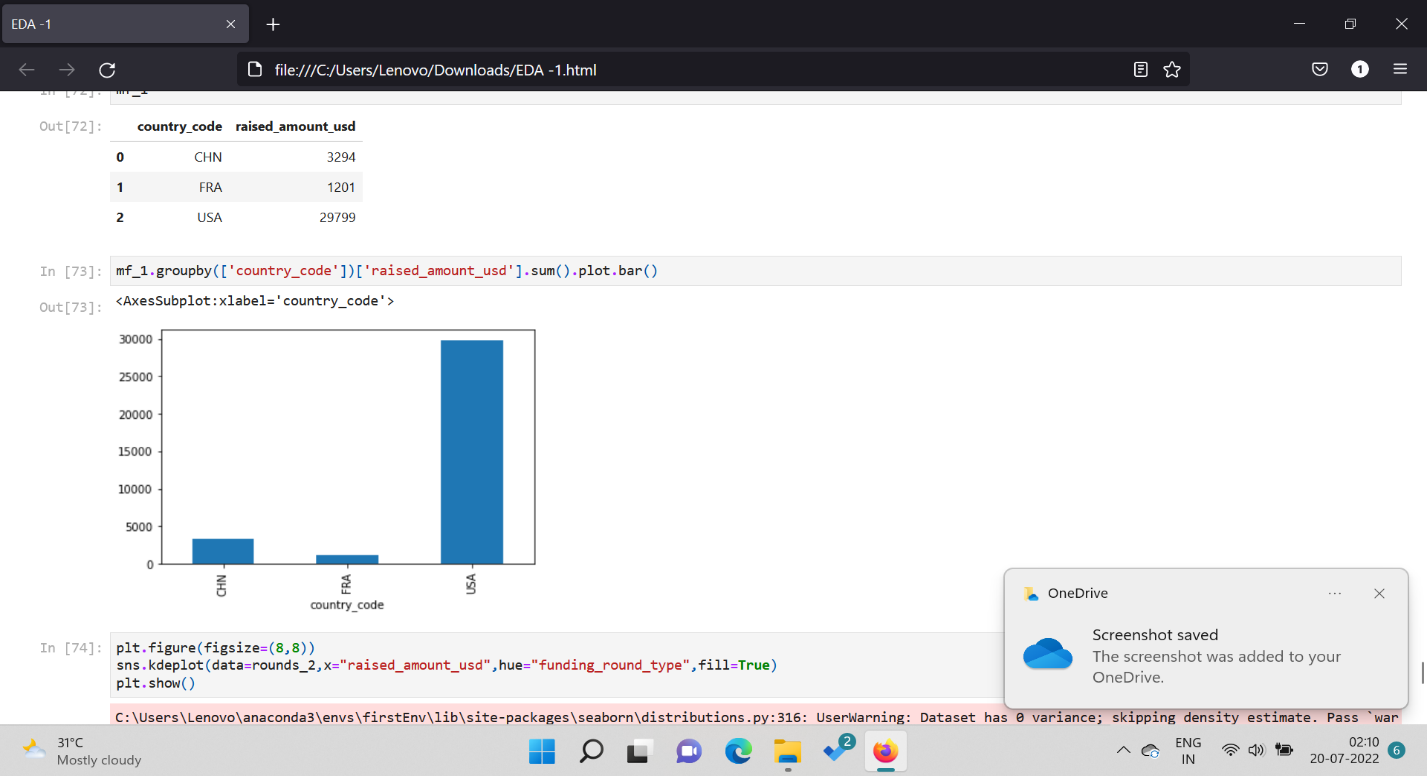
**Inference from this ->** Mean values plot between round type related with raised amount usd.

**Step-4->**



**Inference from this ->** how is funding round code related with raised amount usd.

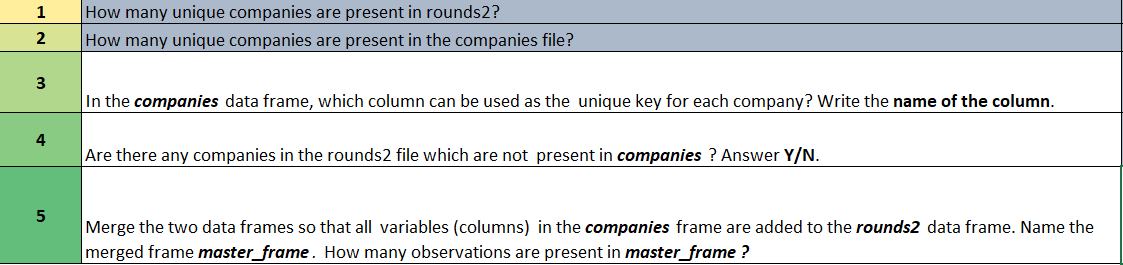
**Step-5->**



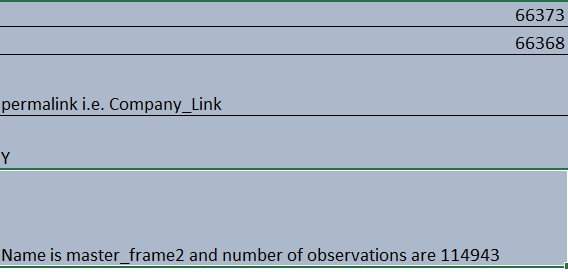
**Inference from this ->** how is country code related with raised amount usd.

**A) Jump to the questions** {High level view for each exercise} -

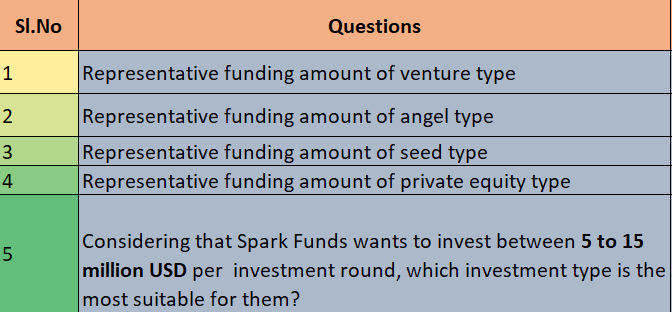
**Table-1: Questions**

****

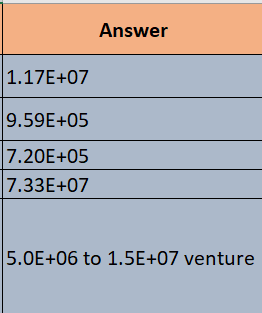
**Table-1: Answers**

****

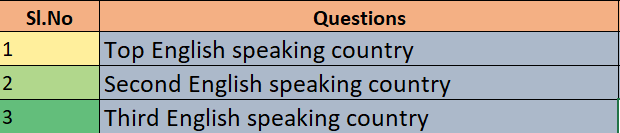
**Table-2 : Questions**

****

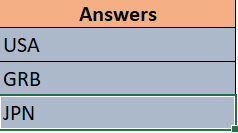
**Table-2 : Answers**

****

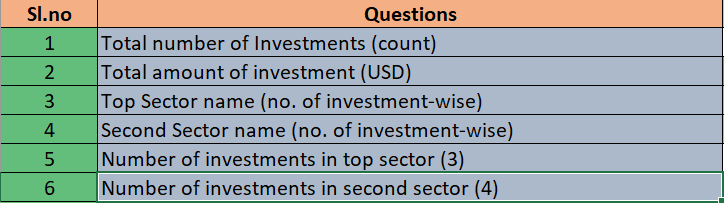
**Table-3 : Questions**

****

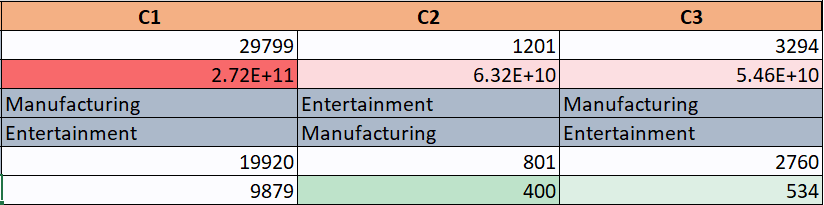
**Table-3 : Answers**

****

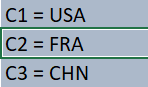
**Table-4 : Questions**

****

**Table-4 : Answers**

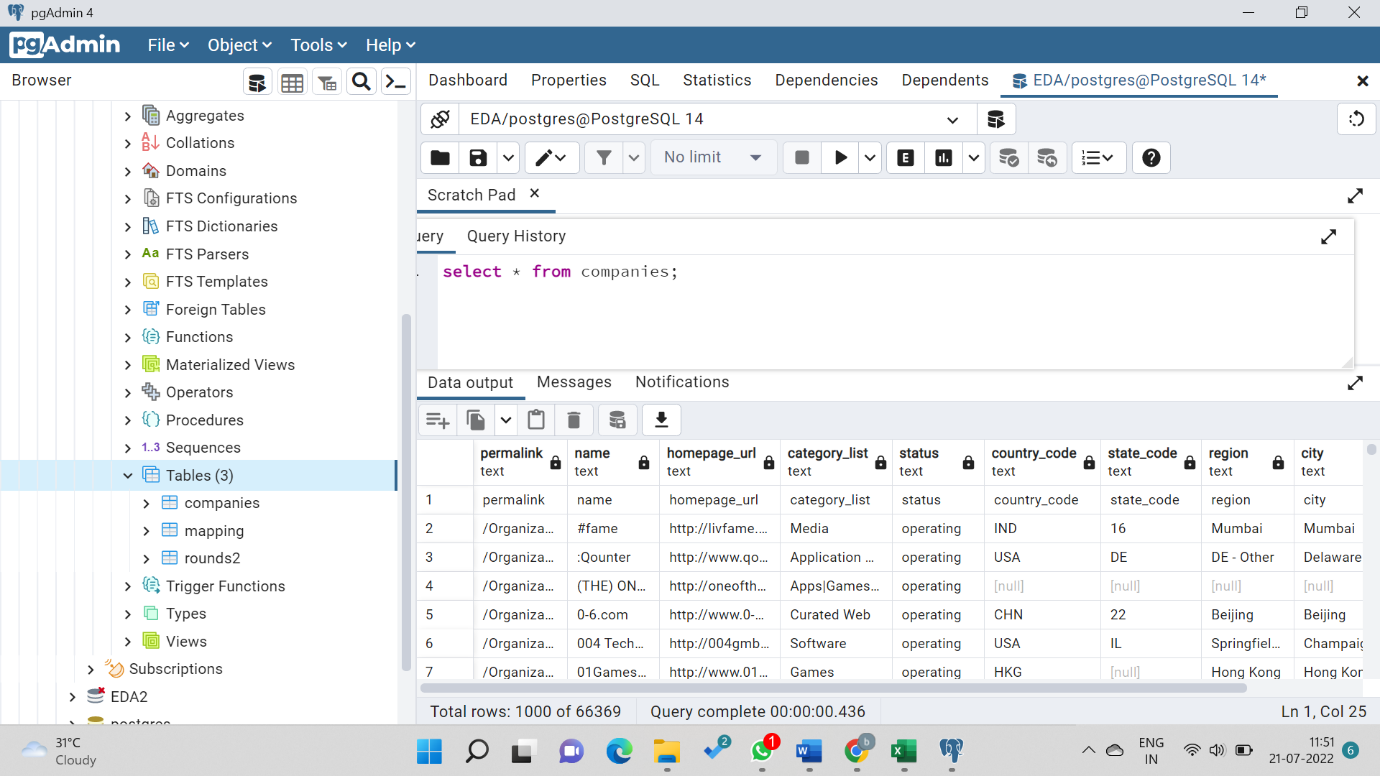
****

In Table-4 we assume let C1, C2 and C3 according to the top highest raised amount countries, that are:

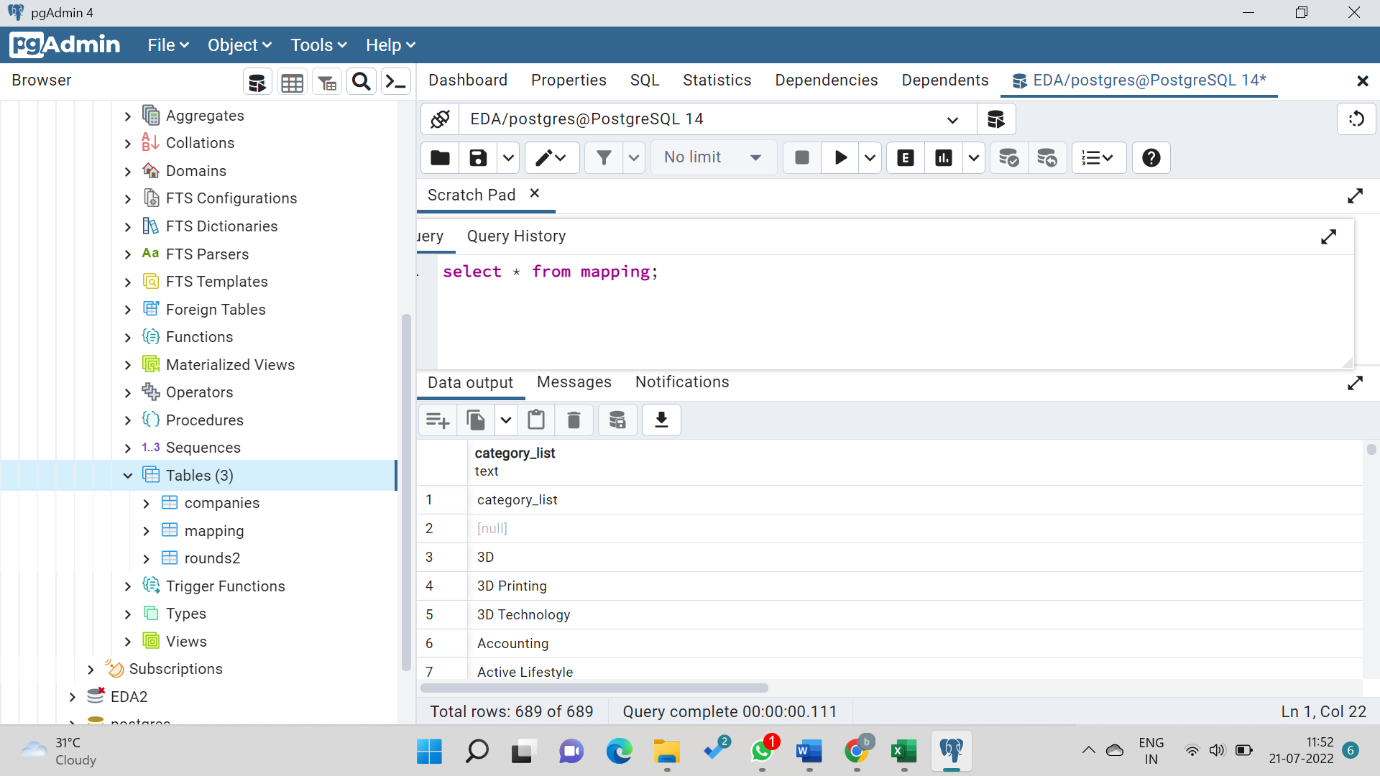
****

**Connection with progress->** Fetching the dataset in PgAdmin by creating tables and importing csv files into that.

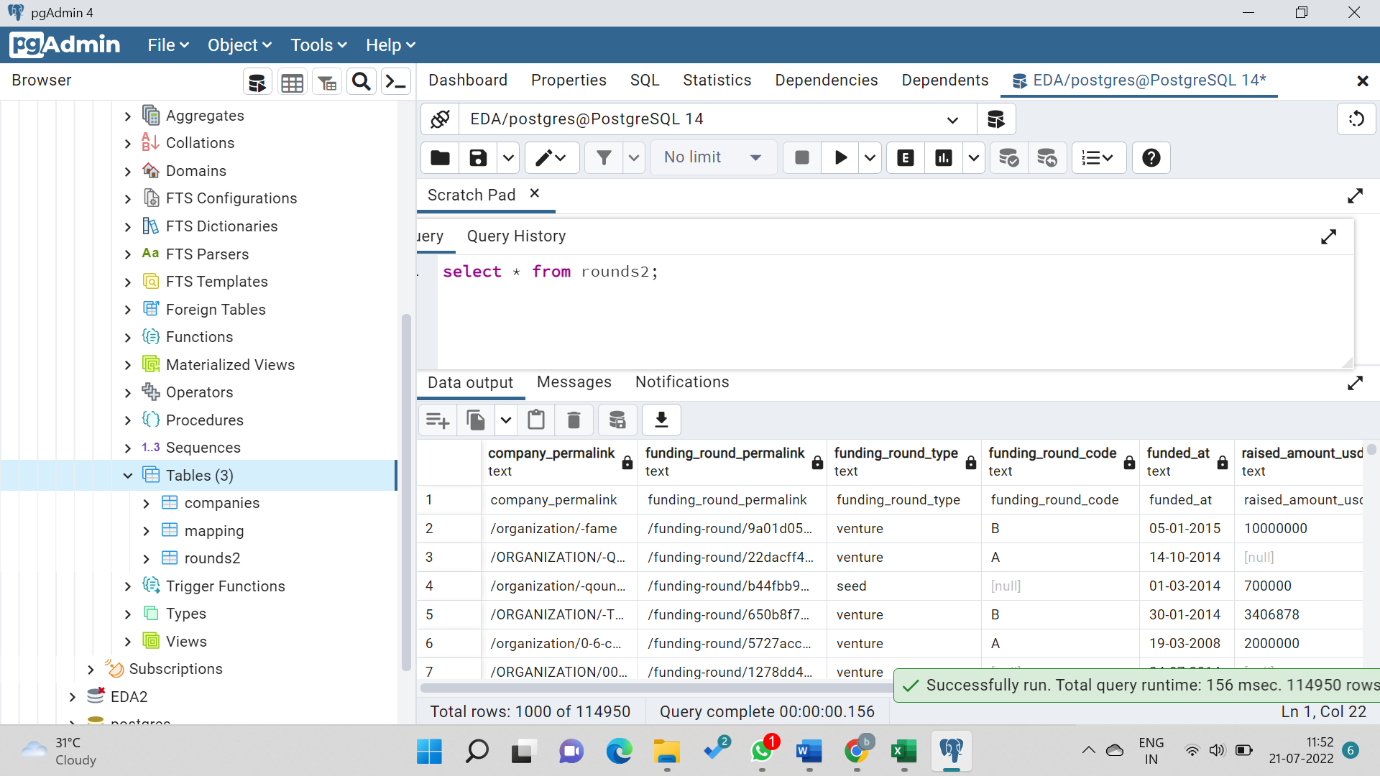
**Companies dataset**



**Mapping dataset**



**Rounds dataset**



**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**