Day 10 Data Project

For this assignment, I used a dataset about users' reviews about different types of phones in different countries, different languages, date of reviews, source of reviews, author, and product type. The dataset didn't have too many tuples, so I used one SQL table to represent the data. The table has eleven tuples which most of which are in the format of VARCHAR. At first, I decided to assign a format of DECIMAL for some of the tuples. After going through the data, I realized that the data is mostly a mix of characters and numbers. It was hard to decide one format of data for a specific tuple, so I decided to declare the format of most tuples as VARCHAR, which is easier to fit in mixed data. As a visual representation of the data, I have written some python codes and SQL queries that count how many reviews of customers about the phones are written in a specific language and in a particular country. Then, I have compared those two visualizations to see customers from which country and language have written the most reviews about a particular phone type.

Following are the two visualization that I have produced from the dataset using python codes and SQL queries;

Reviews from different countries in various languages