**Assignment 1: Data Analysis**

**Submitted to:** Amad Mumtaz  
**Submitted by:** Huzaifa Munir  
**Dated:** 28th October 2024

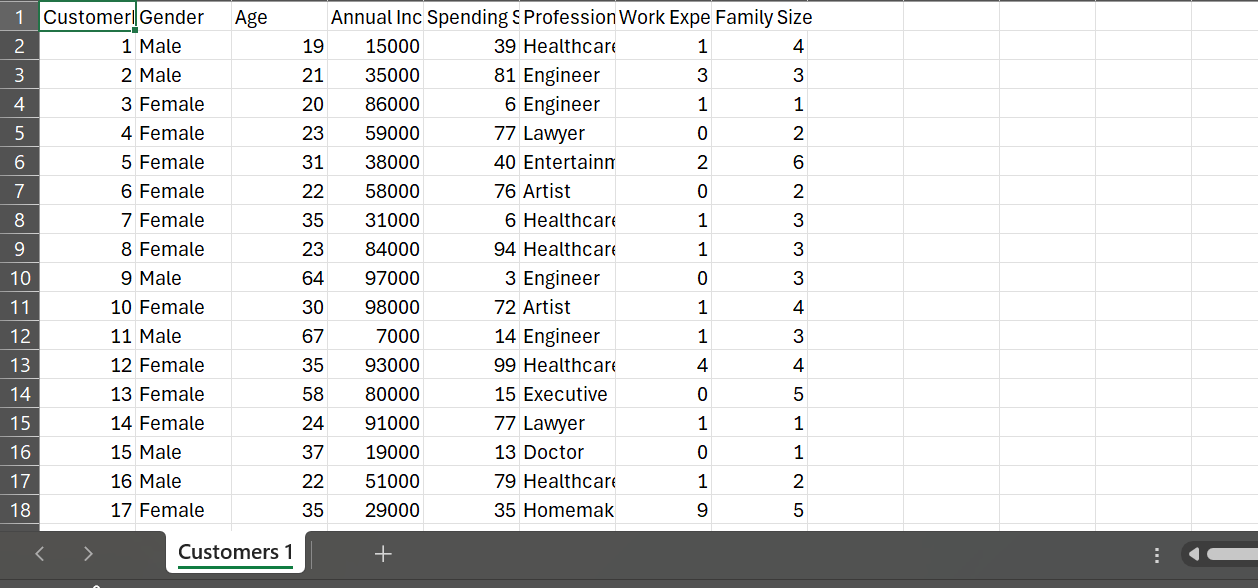
**Problem Statement:** A customer's dataset has been attached. Your task here is to analyze this data and provide your feedback/ view via exploration.

**Instructions:**

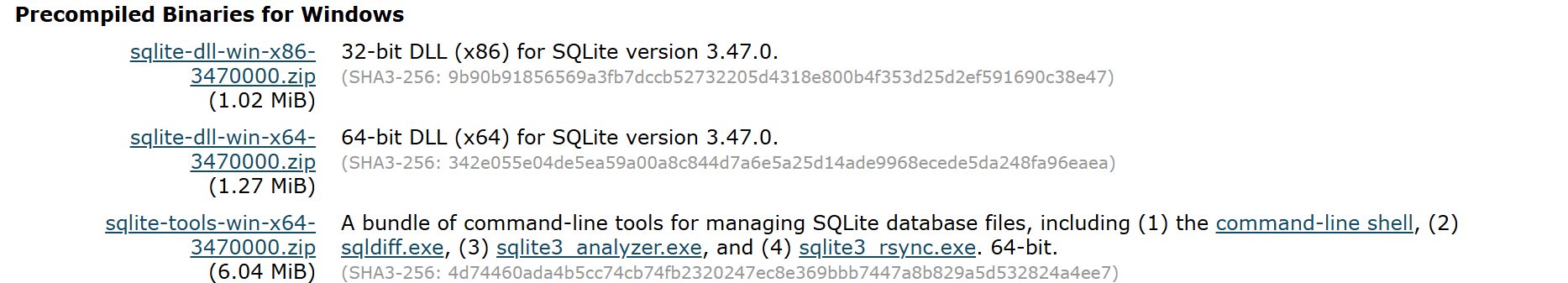
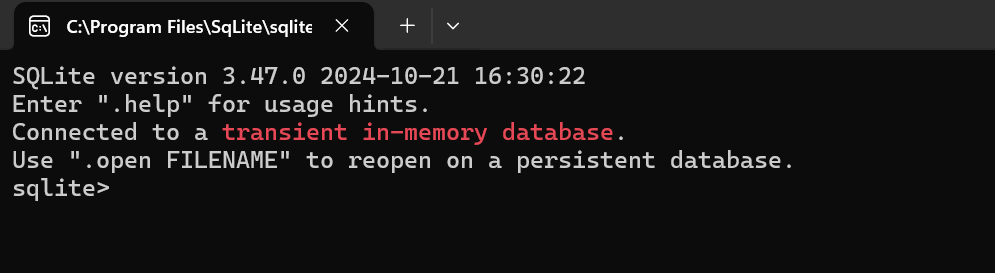
Follow the following instructions provided:

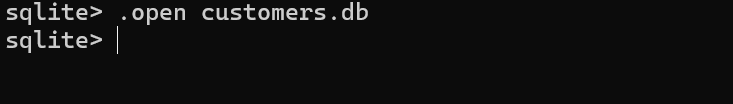
1. Open the CSV file, you can use any database and import this file as a table with the name of your choice.
2. You can use SQL to search for your answers or to explore the dataset.
3. Provide some key points like types of professions mentioned etc. Average spending etc.
4. Thus provide a detailed analysis against this data (At least 5 points).
5. You can collaborate with other team members and can have the same points.
6. Submit your solution via file share (excel or word). DM me once completed.
7. You can you SQLITE DB here which is a serverless database.
8. Install it from the web following instructions and then import the CSV files as a table.
9. Formulate your queries and run them to collect answers.

**Provided Dataset:**

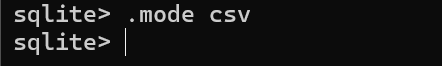
****

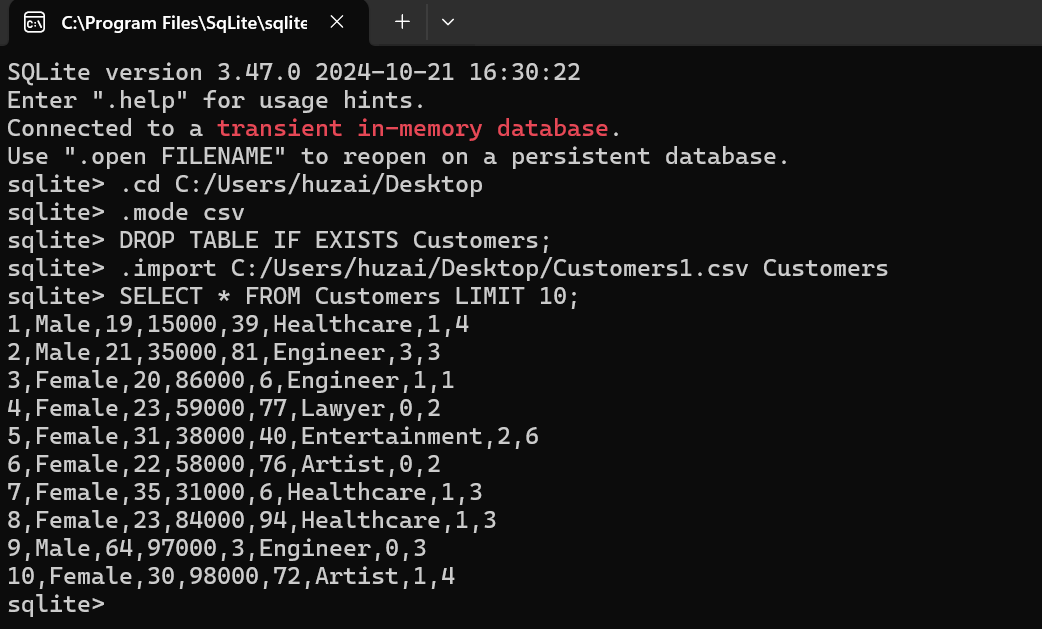
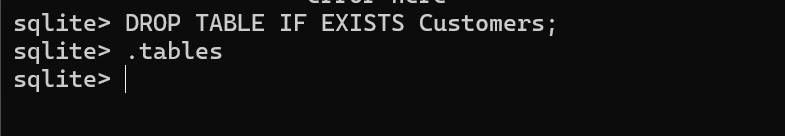
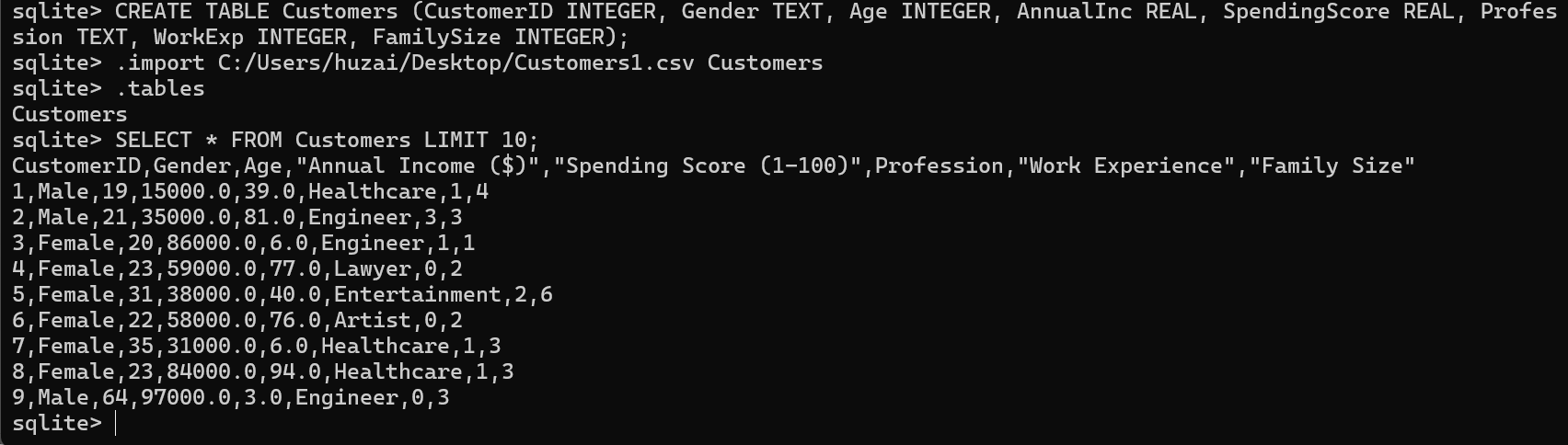
**Data Analysis:**SQLite is a lightweight serverless database that is easy to use for small scale tasks.

1. **Step1:**
   * + Install SQLite from the web if it is not installed on your PC.
     + Go to the official SQLite website and according to your operating system select the respective option.
     + Under the **Precompiled Binaries for Windows** section, download the zip file for [sqlite-tools-win-x64-3470000.zip](https://www.sqlite.org/2024/sqlite-tools-win-x64-3470000.zip) (3rd one).  
       
     + After downloading it, now let’s set up SQLite. Create a folder by the name of SQLite in the program files (recommended, not necessary) of your C:/ drive and then extract the zip file to this folder.
     + Open the folder where you extracted the files. Here, you will see 4 files: sqlite3\_rsync.exe, sqlite3\_analyzer.exe, sqlite3.exe, sqlite3.exe.
     + Double Click to open the sqlite3.exe file and it will open an SQLite command line interface.   
       ****  
       **Note:** The above work is for Windows. Kindly, check your OS requirements before downloading and setting.
2. **Step2:**   
   Now create a persistent database.
   * + To do this, first make sure you are in the right folder where you have full permissions.
     + Firstly, change the directory to **Desktop.** For this, type the following command in the SQLite command line interface.  
       A screenshot of a computer

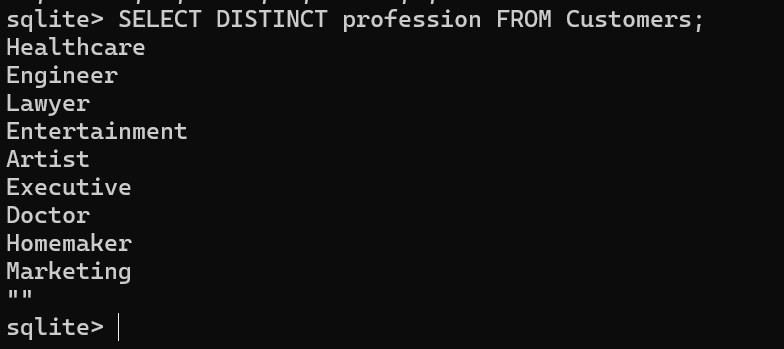
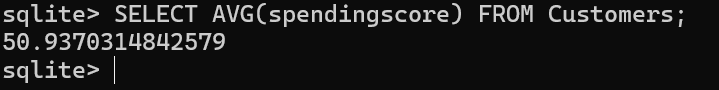
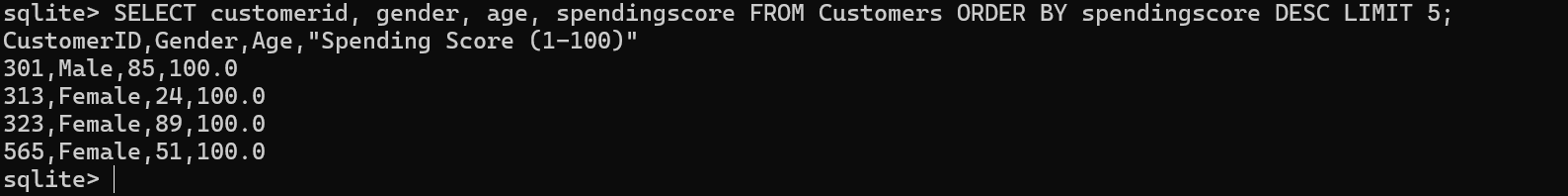
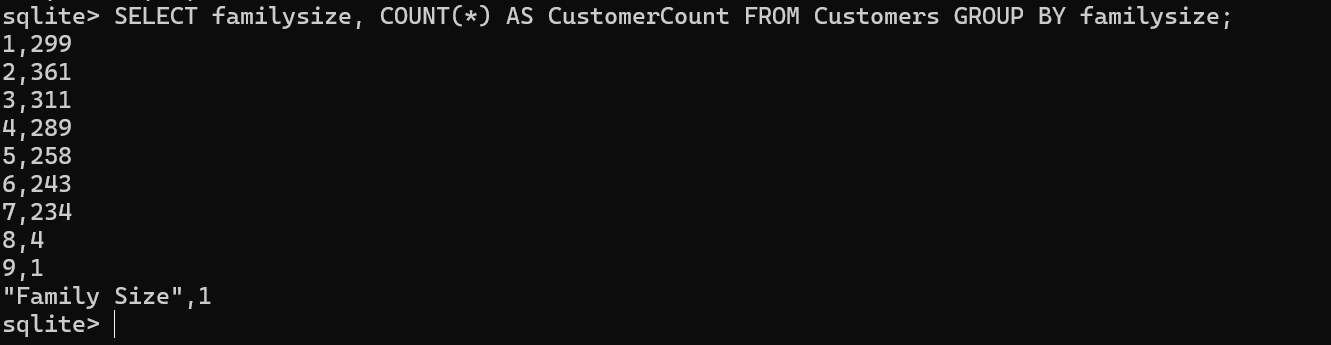
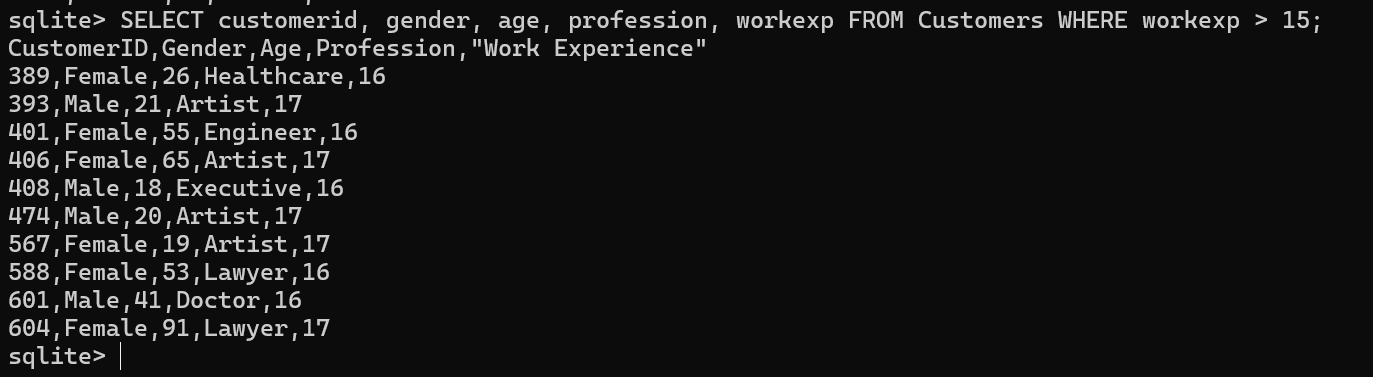
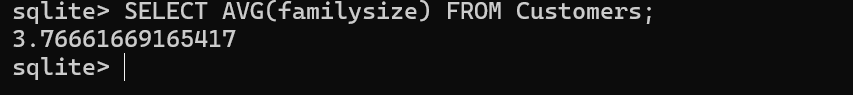
       Description automatically generated
     + Now, create the database by typing the following command.   
       
     + Once the command runs successfully, a file called customers.db should appear on your **Desktop** or the specified directory. You can name the file anything you like.
     + You can also perform the same task with the following command.  
       

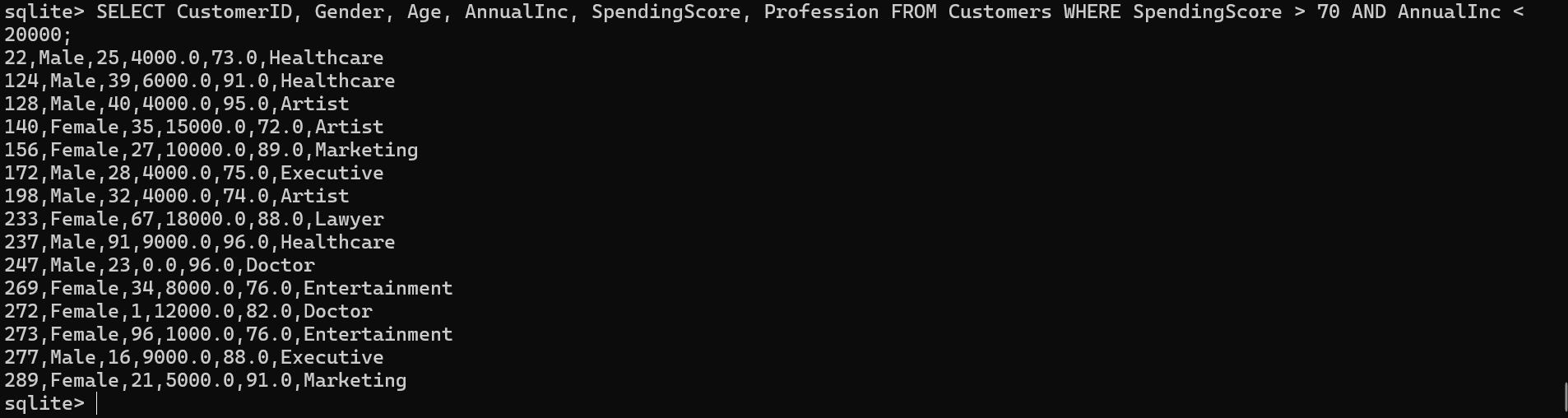
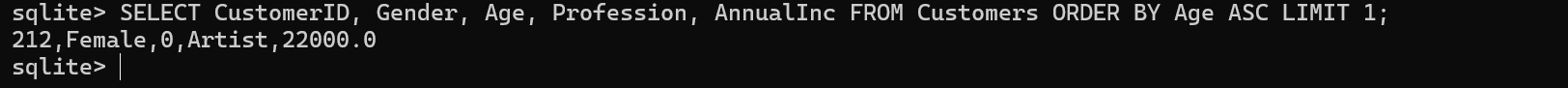
**Note:** Make sure that you are in a directory where you have permission to create files. The simplest way to avoid this issue is to navigate to a folder like your **Desktop** or **Documents** where you have full access.

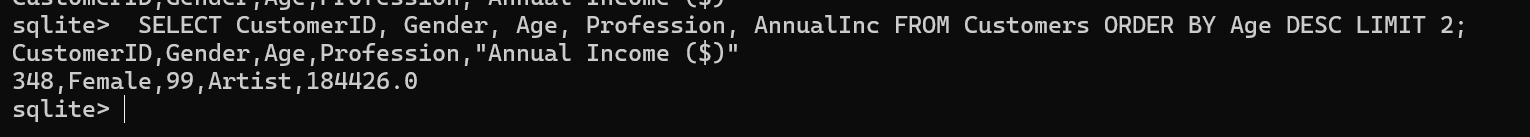
1. **Step3:**   
   Now, import the CSV file.
   * + To perform this task, first set the mode to CSV by following command.  
         
       
     + Now, import your CSV file. (Since, I kept it in my Desktop, you can do the same or choose your preferred directory.)  
       ****
     + The CSV is imported into a **table** (in this case, named Customers) inside your SQLite database (customers.db). The original CSV file(Customers1.csv) itself remains unchanged, but its data is copied into the SQLite table(Customers).  
       **A screenshot of a computer

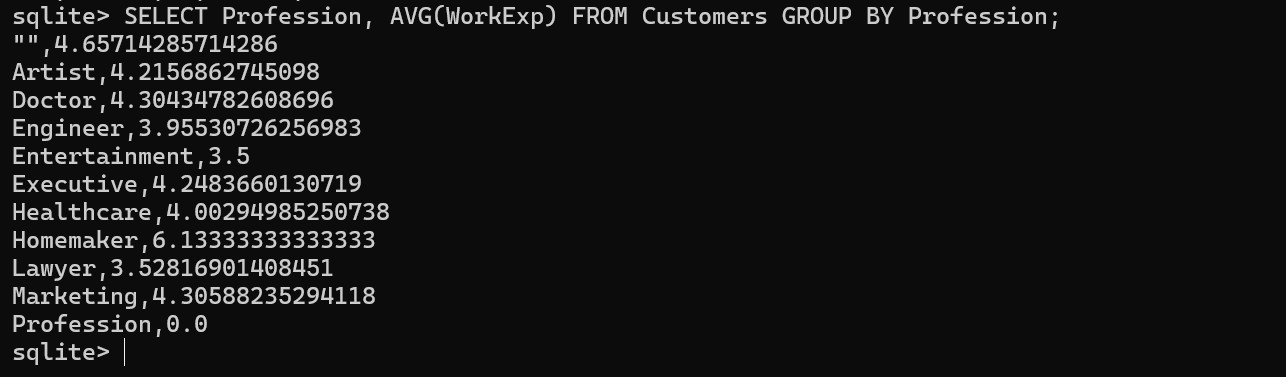
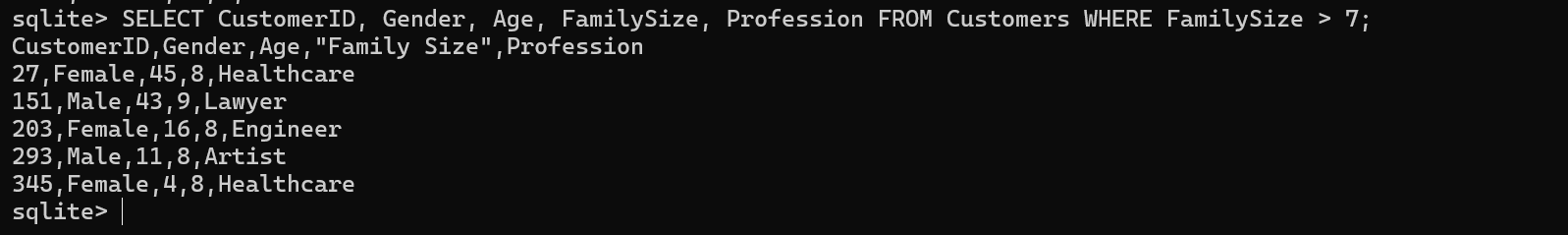
       Description automatically generated**
2. **Step4:**Now, verify if the above commands were successful.   
   
   * + If it doesn’t work, you can try creating the table manually and then copying the data to it. I will first drop the table and then check if there are any tables.   
       ****  
       If there were, the output would have been “Customers”.
     + Now, we will be manually creating the table and importing all of the data into it. Also, check with ‘.tables’ command to verify.  
       ****
3. **Step5:**
   * + Check the structure of the imported table.  
       A computer screen with white text

       Description automatically generated
     + To view the first 10 rows of the table  
       A computer screen shot of a black screen

       Description automatically generated
4. **Step6:** 
   * + If you want to find the number of unique professions in the data, perform the following.   
       
     + If you want to find the average spending of the customers, then do the following.  
       
     + If you want to find the highest spenders based on their genders and age, then do this.  
       
     + If you want to find the customer count based on the family size, run this query.  
       
     + If you want to find customers with a certain number of years of experience, e.g., 15 years.  
       
     + If you want to find the average family size, then write this query in SQLite CLI.  
       
     + If you want to find the highest earners based on their genders and age, then do this.  
       A black screen with white text

       Description automatically generated
     + If you want to find customers with a high spending score but low income based on their genders and age, then do this.  
       
     + If you want to find the youngest & oldest customers, then do this.  
       



* + - If you want to find the average work experience by profession, then do this.  
      
    - If you want to find customers with large families, then do this.  
      

**Conclusion:**

The data analysis reveals key insights into customer demographics, spending behavior, and income distribution across various professions, age groups, and family sizes. Notably, high earners tend to have higher spending scores, and significant variations exist in work experience and family size across professions.  
The analysis shows that healthcare professionals have the highest average spending score, while customers aged 26-35 have the highest spending across all age groups. Additionally, customers with large families (5+ members) tend to have lower average spending scores despite higher annual incomes.  
  
