Assignment Case Study 2

Steps Followed:

- Clean **Start-up Name** and **Investment Type** in the original data frame
- Make one data frame for "**Crowd Funding**" and "**Seed Funding**" only
- Make another data frames for " **Private Equity**"
- Define **city function** to return a sorted frequency dictionary for **cities (among NCR, Bangalore and Mumbai)**
- Define **investor function** to return a sorted frequency dictionary for **investment type**
- Call the investor function created above to get **investor name** frequency dictionary for:
    - Dataframe with **all** investment types
        - Use this dictionary to check the **top 5** most times funded investor **(including repeat startups)**
- Call the **city function** to return a sorted frequency dictionary for:
    - Dataframe with all cities
    - The dictionary returned with this function call will have the **first element** as the city which has **startups most times funded**

- Create two new data frames from (one with **crowd funding and seed funding** rows and other with **Private Equity** funding rows only)
- Create corresponding child **dataframes** with **investor name** split by **comma** with expand=True (from the above created data frames)
    - Clean up undisclosed investors
- The new child data frames will be concatenated with the **corresponding** parent dataframes to get each investor in a new row
- Remove duplicates in the startupmname column and investor column to remove investors that have invested in the same startup multiple times
- Now with the duplicates removed group by and count the investor name to get the unique investor list