**Data Science deviated people into groups by clustering:**

* Created a tool that to deviated people into groups by clustering to help company

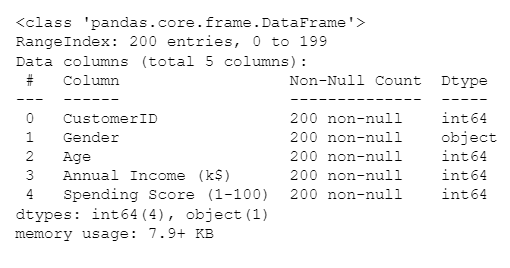
**Packages:** pandas, numpy, sklearn, matplotlib, seaborn

**The solving mechanism**

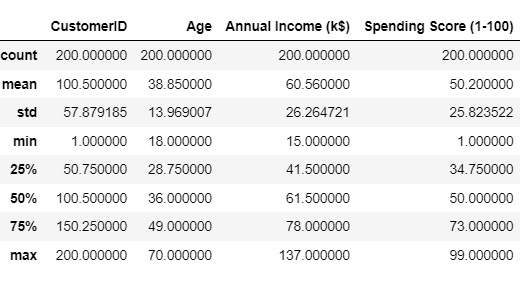
* build machine learning model using python

**Describe the dataset**

* Data source:
  + [Mall Customer Segmentation Data | Kaggle](https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-python?resource=download)
* Data description
* I use pandas library to description dataset



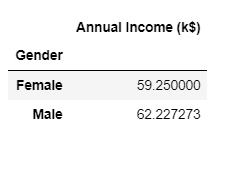
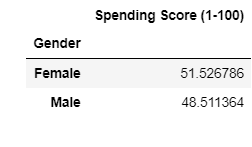
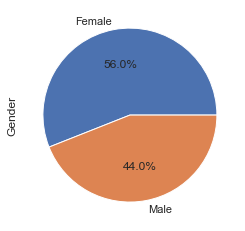
* + form output I know number of rows (13320) and num of columns (9)
  + name of columns and data type for each column
  + number of null values in columns (Ex: bath has 53 sell null)



* + I conclude from this table count , mean , min , median , max , standard deviation
  + From this information I know count of value in each column
  + Std mean standard deviation it help us to know the spread of values
  + Max , Min , mean , Median of each column

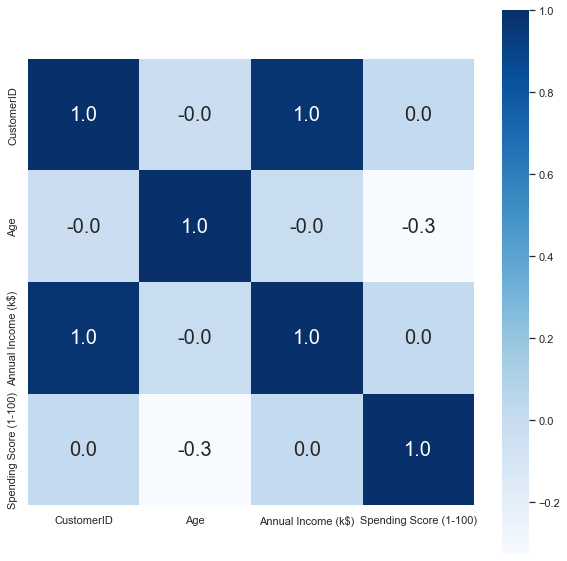
**descriptive statistics and data distribution charts**

I looked at the distributions of the data and the value counts for the various categorical variables. Below are a few highlights from the pivot tables.

**Table 1**

**Figure 1**



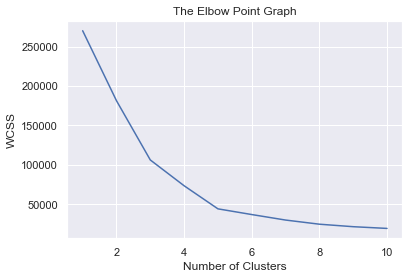
**Figure 2**

**التعليق**

* (Table 1) pivot table that continent tow columns fistr column name location oher column price home in this location
* (figure 1) barplot in X name location Y count house in location
* ( figure 2 ) this chart dis correlation between [ total\_sqft , bath , price , bhk ]

**Model Building**

* First, Choosing the Annual Income Column for X variable.
* Then , Spending Score column , and Finding wcss value for different number of clusters



* Optimum Number of Clusters = 5
* Training the k-Means Clustering Model
* return a label for each data point based on their cluster for Y variable.
* 5 Clusters - 0,1,2,3,4
* Visualiing all the clusterrs

