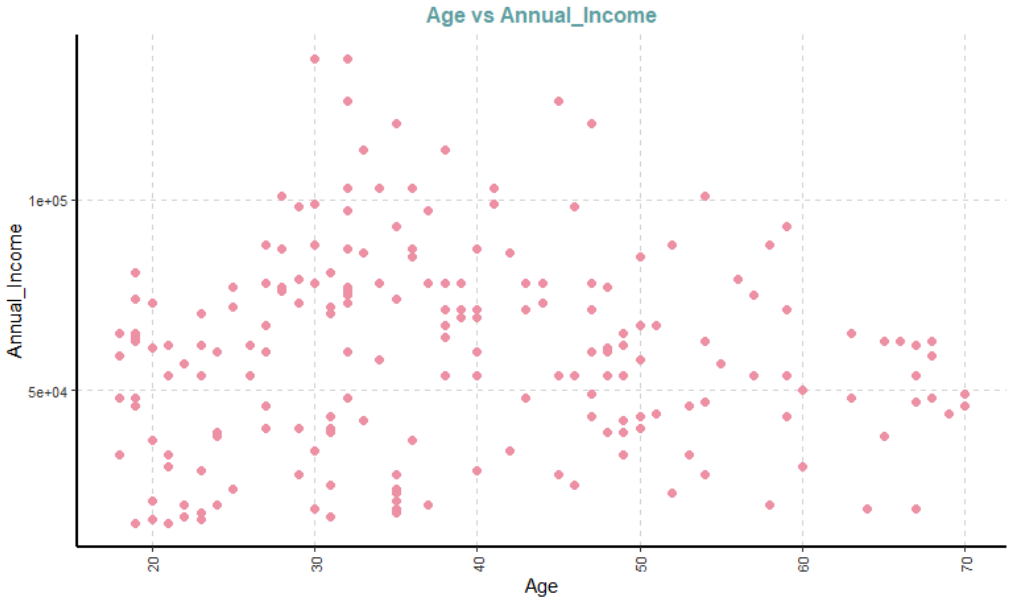
# Sentiment Analysis Case Studies

- **Influence of age on Annual income in market basket dataset**

  At the first glance of the head of the dataset, it is seen that any variable can be used as a target variable for the analysis although Spending Score is preferred. When the numerical summary is generated, it is noticed that salary range from 15000 to 137000. The skewness of the salary is 0.319 which means it is little bit on the positive side and the mean is 61500. Inspecting the customers age summary, it is seen that mean age is 38.85 and it ranges between 18-70. It proves that spending score is given to only people above 18. The interquartile range is 41500 to 78000 which means that most of the salaries are between this range and similarly interquartile range of age is between 28.75 and 49. This is further proved by the scatter plot as shown below. When the scatter plot is generated of the age with spending score, it can be noticed that although the shape shows no relation at first glimpse but on close inspection it is seen that there are two income peaks. First peak is at around the age of 30-32 and the second peak is seen at around 45-47. This proves that highest income is of mid-aged people between 30-47 and it tends to decrease after this range.



- **Influence of age on Spending score in market basket dataset**

  As mentioned earlier the Spending Score is given by the mall to the customer according to their buying habits. Higher the score means higher the buying power is of the customer and the spending score ranges between 0-100. From the statistical results, it is seen that mean of spending score is 50.20 which means it is almost in the middle of the range. The interquartile range is 34.75 to 73 which means that scores above 25% and below 75% are in this range. When spending score summary is seen of each gender separately, it is seen that mean is 48.511 for male gender and 51.527 for female. It is further explored from the scatter plot shown below. The scatter plot shows a very sharp decrease in age after spending score crosses value of 62.5. This clearly proves that scores of 0 or above are received by only people below age of 40. There is a higher number of scores between age 37.5 and 62.5 which proves the interquartile range values. There is another outcome seen on scrutinizing the plot, which is the distribution of scores of each gender separately. It is noticed that out of top three scores, two are female and scores received by customers of age 20 or below are generally men. Although there are other noticeable outcomes of each gender score distributions, but they are not mentioned as it needs greater amount of data to be analysed.