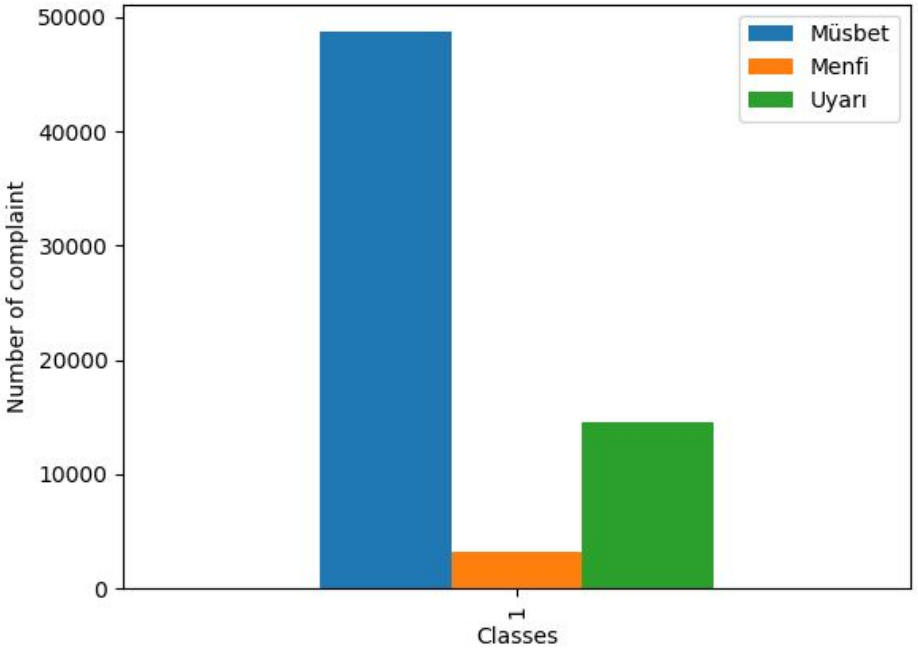

Figure 1 : A bar chart that shows ratio of classes

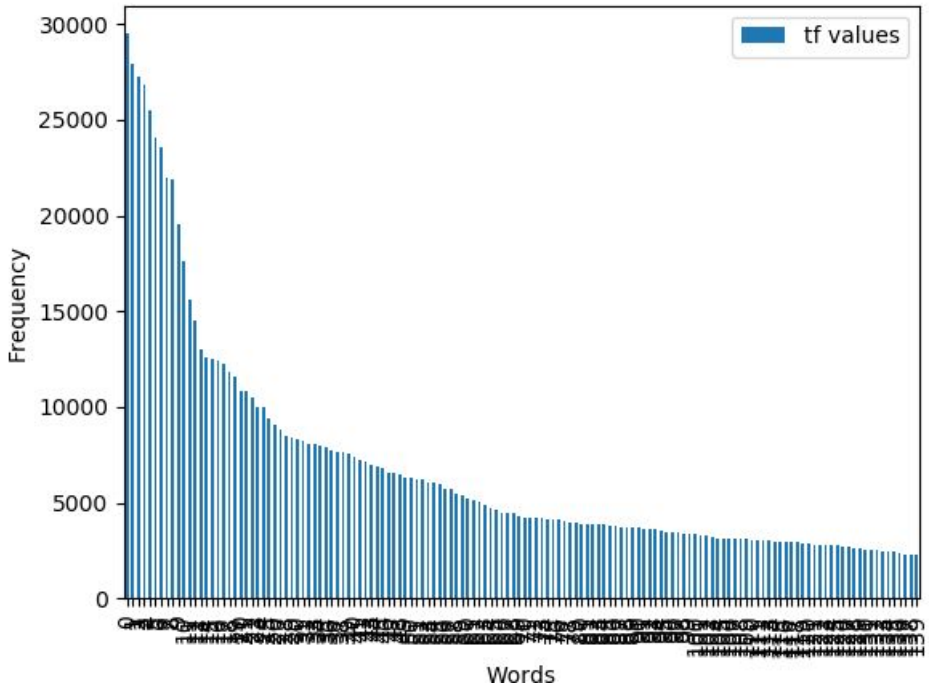

Figure 2: A bar chart that shows first 140 words with corresponding frequencies

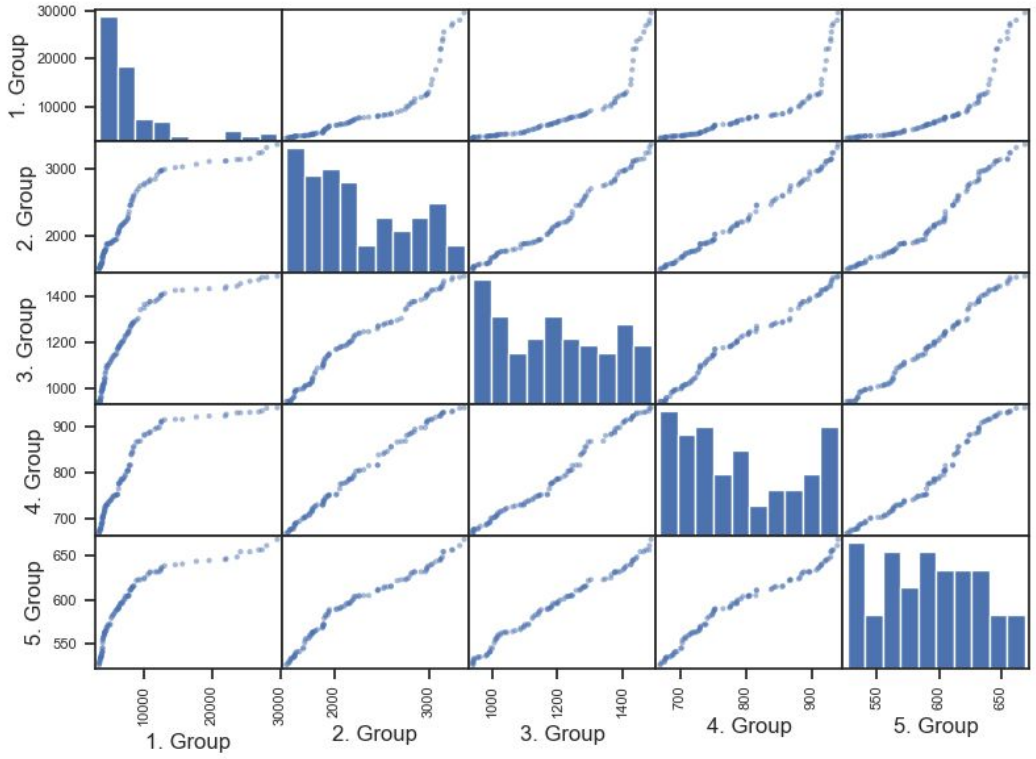

Figure 3: Scatter plot matrix of first 500 words group by 5

Results

In Figure 1 we can see that most of the complaints are classified as Müsbet and very small amount of them is classified as Menfi. And some of them classified as Uyarı which can give penalty to the driver if he get another complaints.

In Figure 2 we plot the graph of first 500 words with their frequencies and it is very compatible with Zipf's curve. It says some words repeated in most of the sentences in dataset and they don't help us to classify the sentences.This words are called as Stop words and we must eliminate them to get better results. Lastly we saw that only a small part of all words are in this Stop words class.

In Figure 3 we divide the words and frequencies data into five groups so that each group contain a hundred words. Then we plot each pair in a five to five matrix which called as scatter plot matrix. In these lines we saw that the linearity decreases when one of the axis contains first group that contains Stop words whose frequencies are high valued and highly variable.