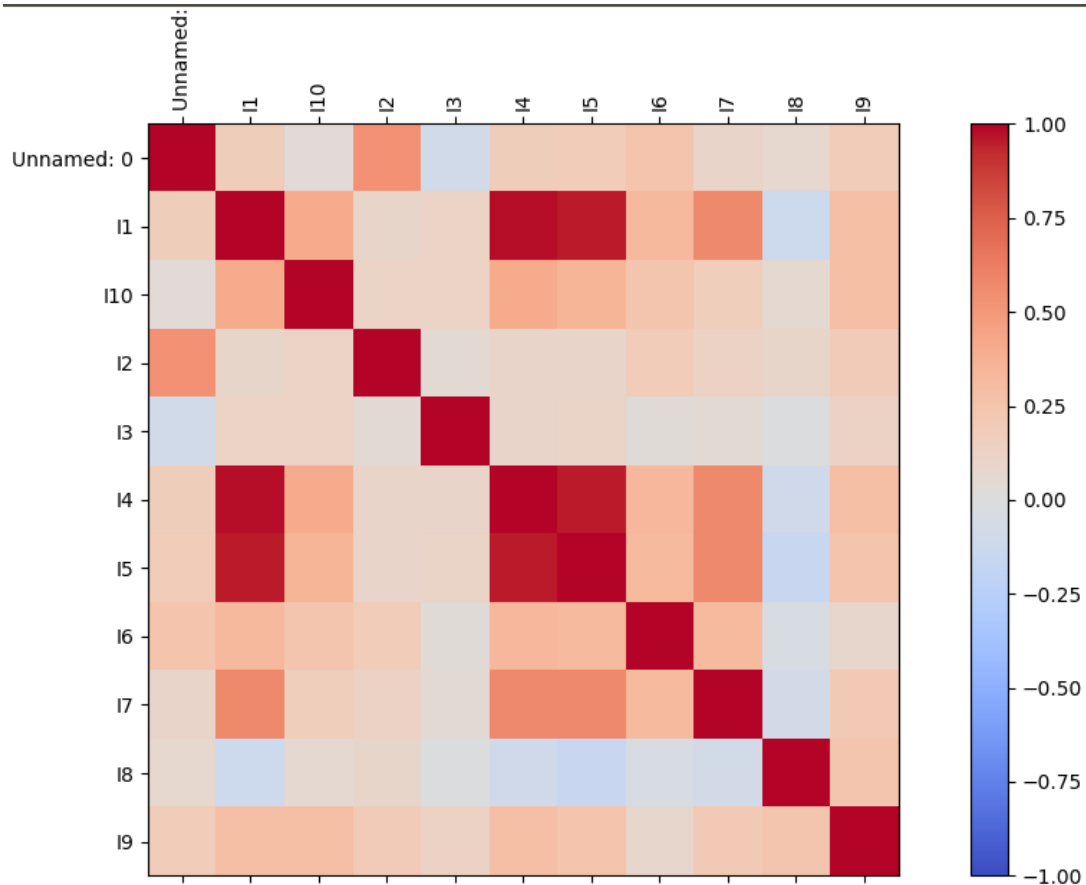


- Because of the fact that our dataset mostly contains floating point numbers, there's no chance to find exactly the same values for a feature. Therefore we were not be able to analyze frequency/counts for the features and create graphics about it.
- For the pairwise correlation part, since we have 7598 features in our dataset, we were supposed to make feature selection in order to work with them. We searched from the open soruce functions that applies feature selection to a given database. We found 4 different recipies which is provided by scikit-learn. One of them is Univariate Selection, but giving floating point numbers causes some errors, then we round our values in our datasets and work on rounded values just for feature selection. We chose 10 feature to analyze. The function returns us the selected columns' values. Then we try to find these columns order and save them to another excel file. At the end, we created correlation graph for these features.

1. Matrix that shows the correlation values between columns

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10
I1	1.0	.101	.118	.981	.958	.325	.573	-.112	.285	.400
I2		1.0	.503	.105	.105	.183	.131	.101	.189	.124
I3			1	.108	.111	.027	.050	-.015	.132	.120
I4				1.0	.956	.333	.573	-.099	.282	.406
I5					1.0	.318	.575	-.142	.245	.347
I6						1.0	.314	-.043	.090	.237
I7							1.0	-.074	.206	.170
I8								1.0	.237	.068
I9									1.0	.288
I10										1.0

2. Correlation Graph:



- There are both negative and positive correlations between our features in our dataset. The intense red ones and the close values to 1.0 means there is a strong positive correlation among two feature which means the distribution of the values are quite same. When one of them has an increasing trend, the other one also has. If there's no correlation among two features, in other words, the correlation value is close to 0, we can see that the distribution of the feature values are not dependent. On the other hand, in the blue ones and the ones which has negative numbers, we can say that the values of the features have opposite trends. If one of them is increasing, then the other one is decreasing or vice versa.