Hi, we are the first group, and we are going to present the last lab of AI. PASA DIAPO

In the first section, as the X set is composed just with the 2 features of glucose and bmi, we proceed to apply the computing of inertia in order to know the number of clusters, so that we can perform k means. Here we conclude the importance from 1 to 70 interval. Thanks to the second metric of the average of silhouette coefficient, we conclude that 10 is the optimal. PASA DIAPO

Here are the plots from the silhouette metric and the interpretation of the model with the cluster labels above, and down the y values. PASA DIAPO

In the second section, we compute the same steps, but the x set having all features. The interval of interest coincides with the previous point, so we apply silhouette coefficient again. As a result, we get 10 as the best number for k means. PASA DIAPO

The first interpretation that we perform is the means from all clusters in each feature. Some data to stand out are the age from clusters 9 and 7. Or the high value of insulin from cluster 5. But this are just some examples. In order to provide a better visual conclusion, we extract the histograms of each type of group. PASA DIAPO

Now we can see clear peaks from some of the types, and some important distributions like the age mentioned before. PASA DIAPO