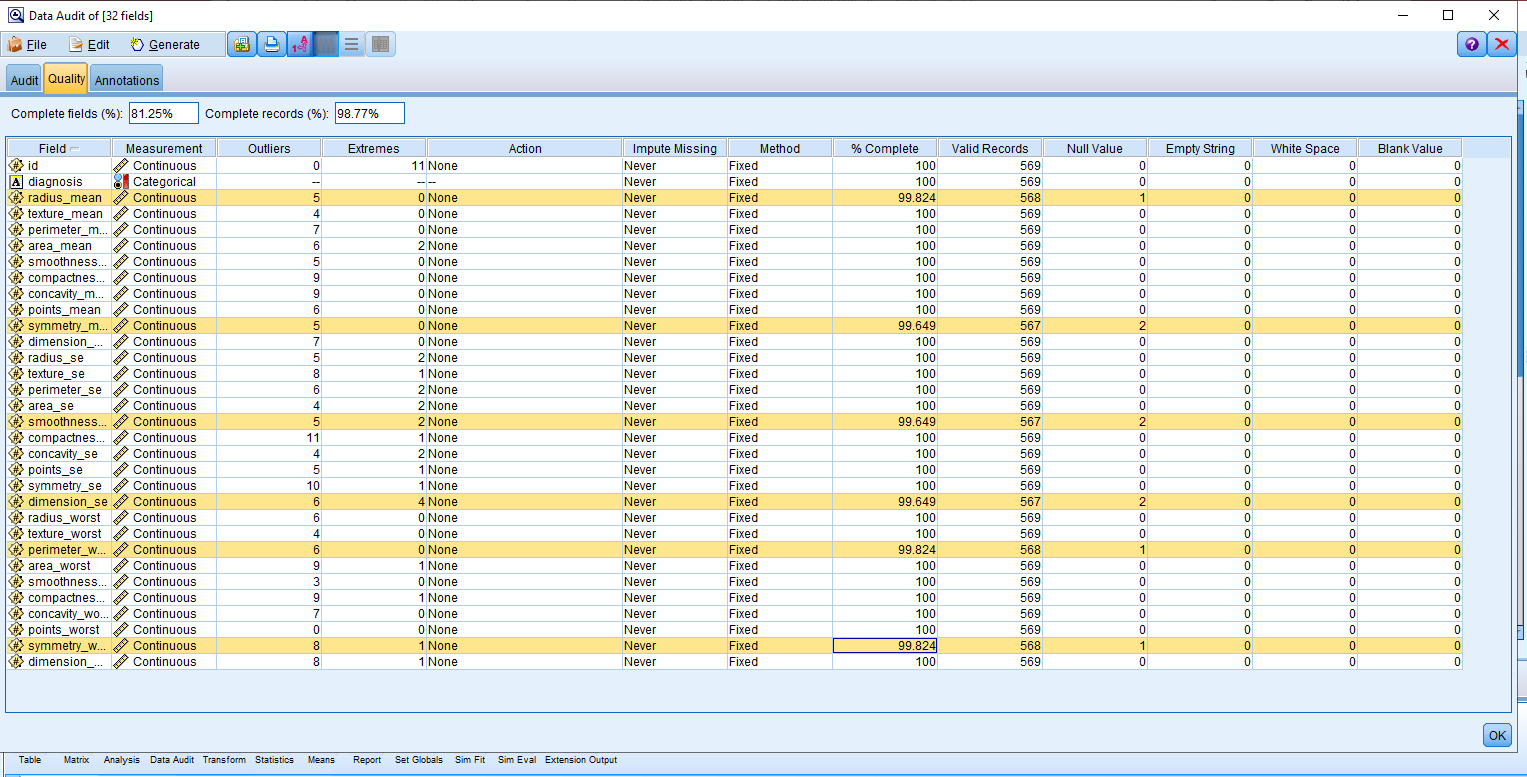
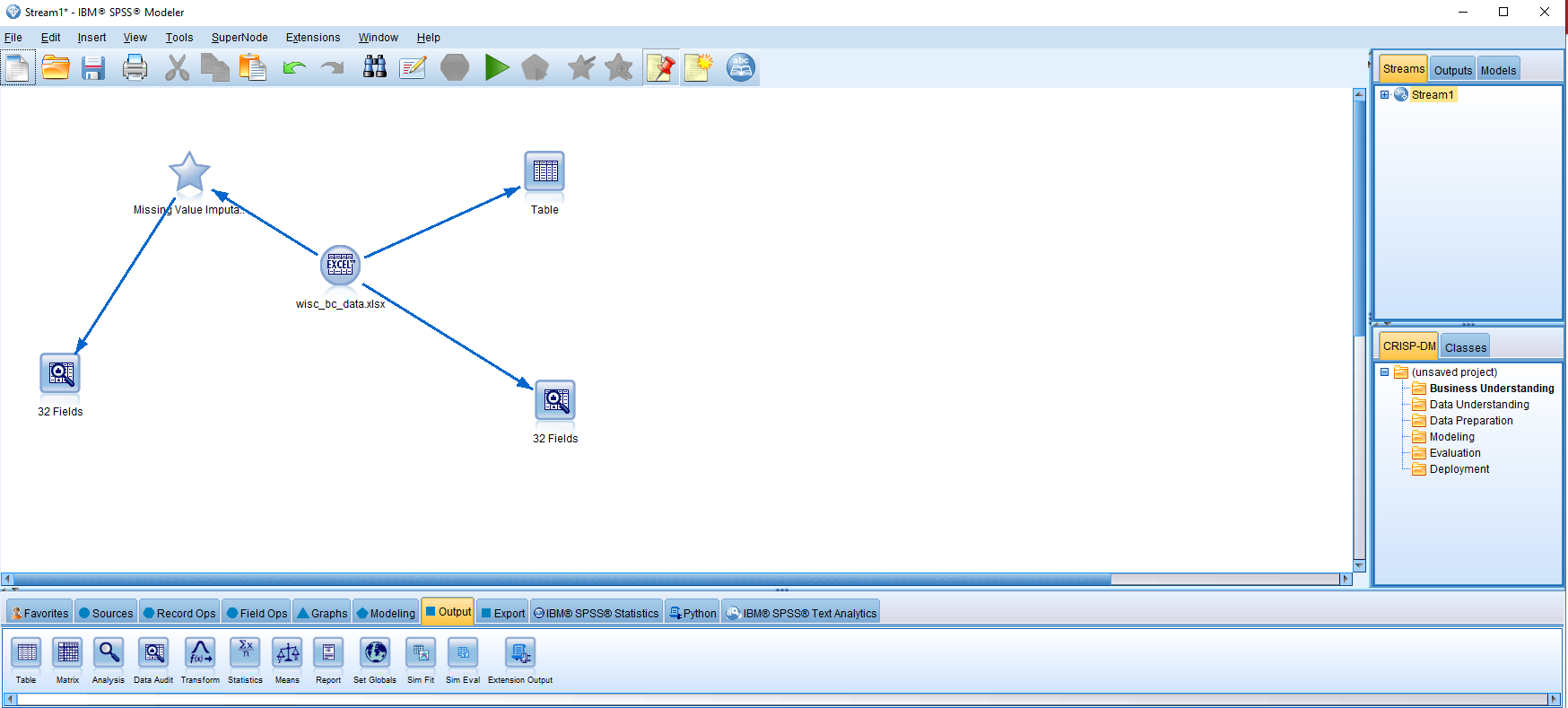
**1. Explore whether there are missing values for any of the variables. Show how you did so in SPSS Modeler**

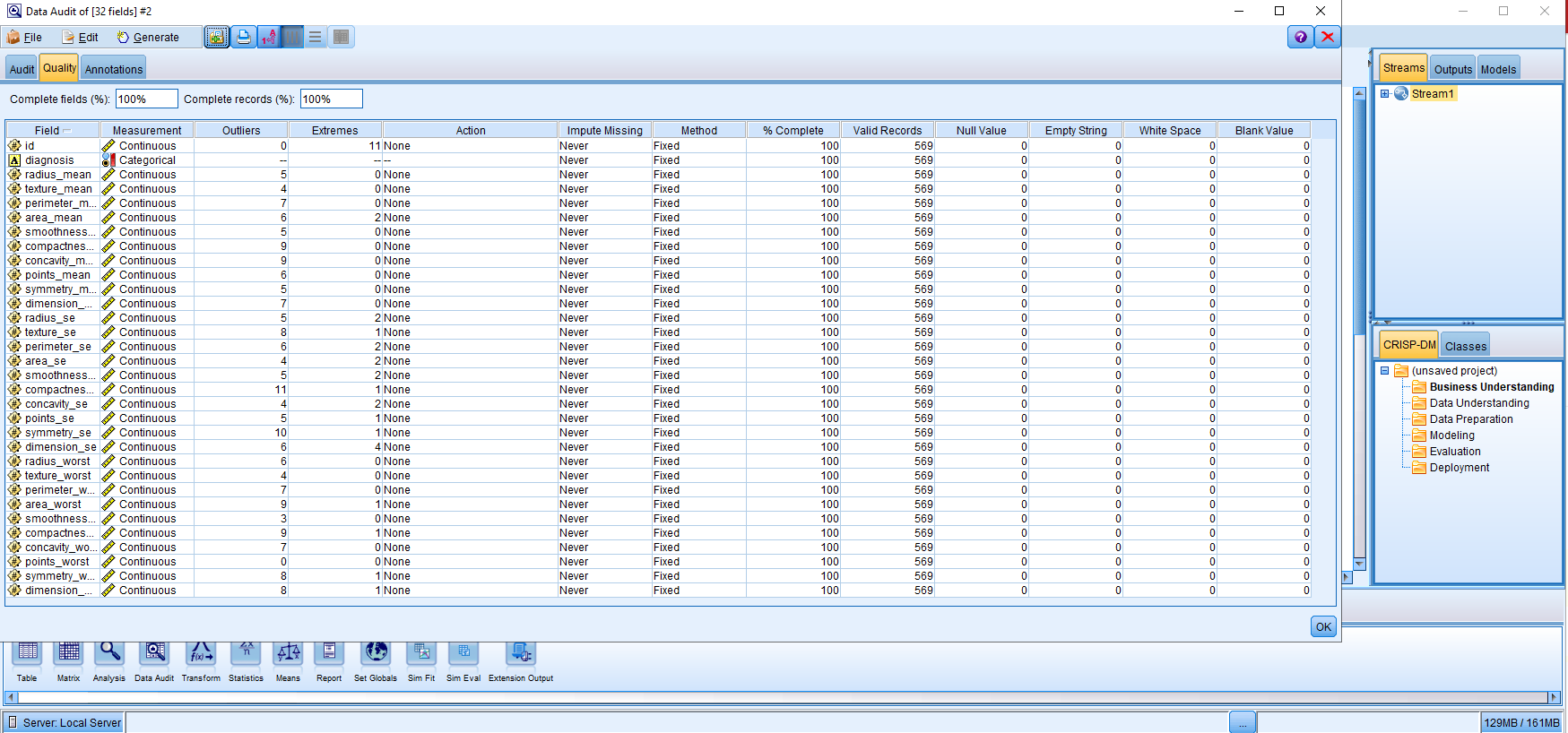
Below is the snapshot for all the missing variables from the given data set, we have used the features of Data Audit output table to get the quantity of missing variables.



Once the missing variables has been handled, then we have generated the missing value node to address these variables from given data set.

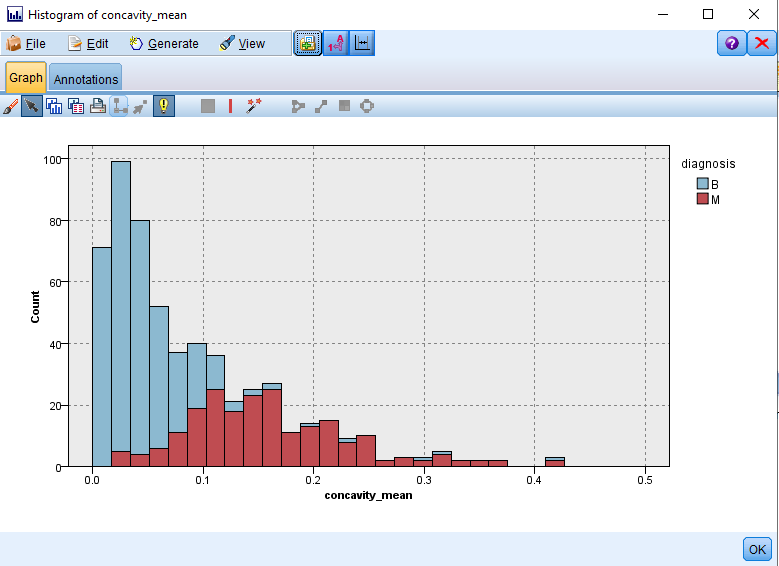


Below is the snapshot of data after handling the missing variables, complete detail of variable is available in submitted stream.

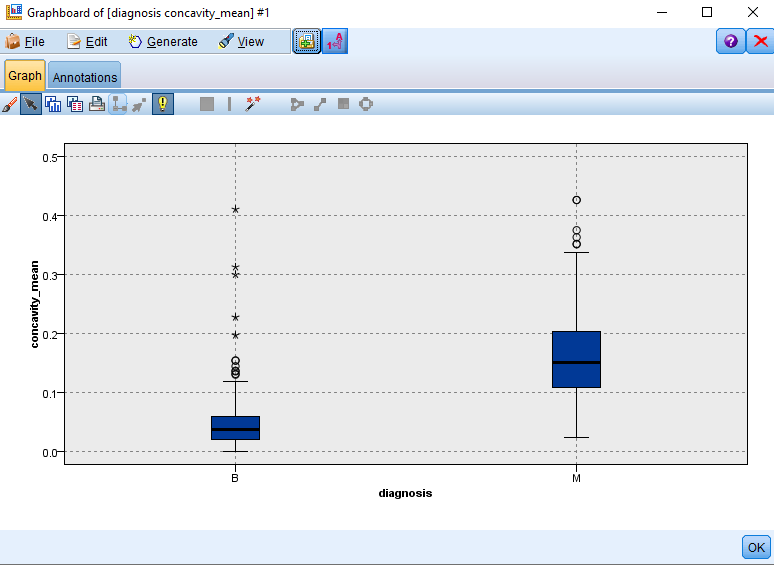


**2. Use a graph to determine visually whether there are any outliers in the average concavity field (concavity\_mean) that measures the severity of concave portions of the contour. What kind of graph would provide better visualization for this task.**

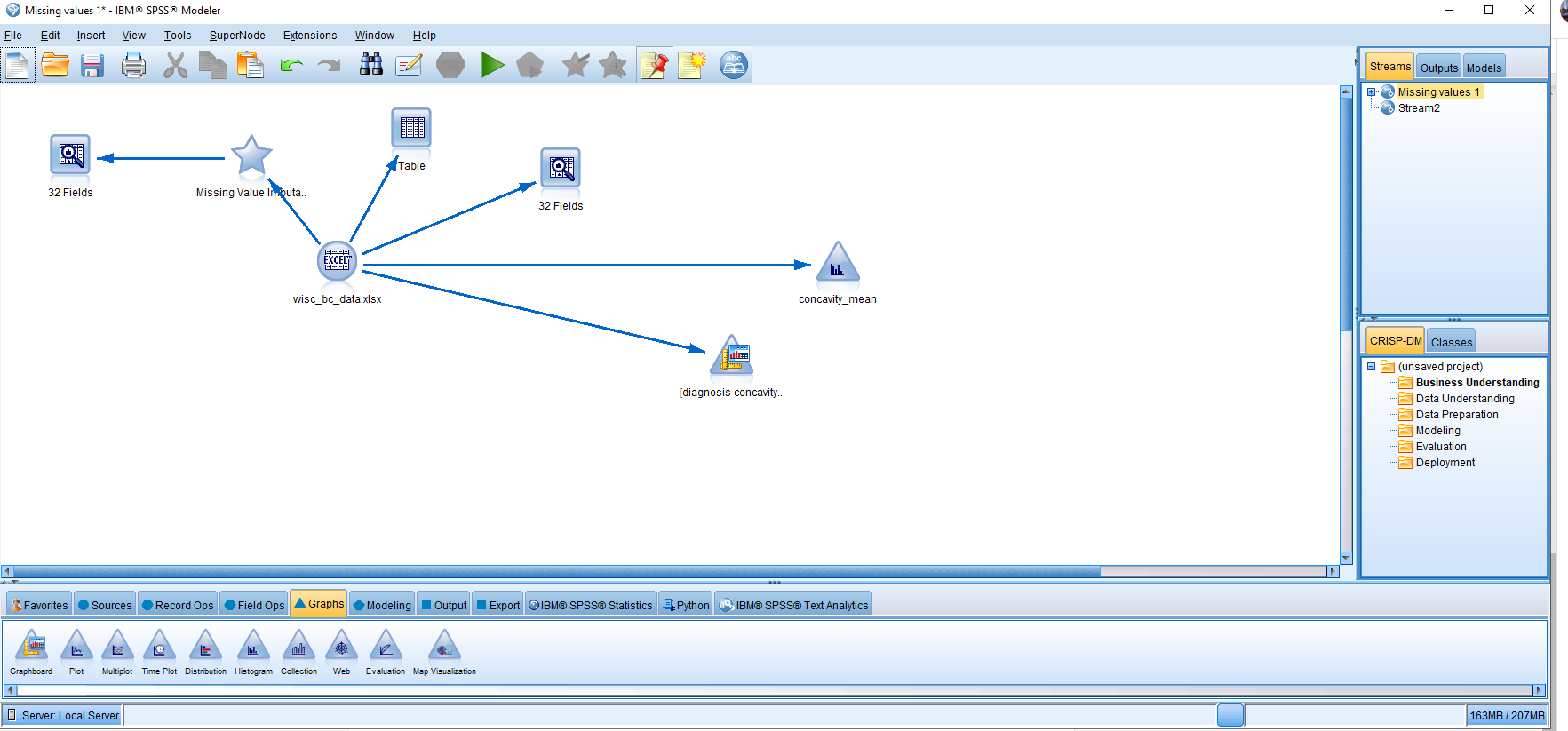
Below is visual representation of the average concavity attribute (concavity\_mean), from the below histogram graph we can say the there is no outliers available, we will also check with box plot graph.



Below is the box plot graph representation of the average concavity attribute (concavity\_mean).



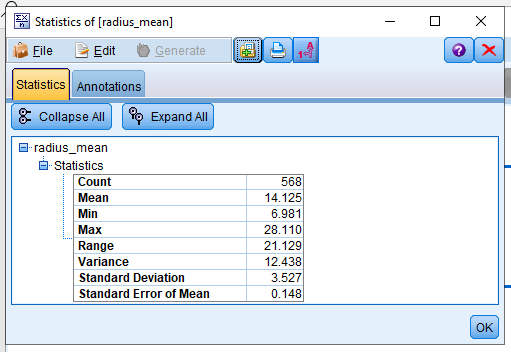
From a above box plot, we can confirm that there are no outliers available.



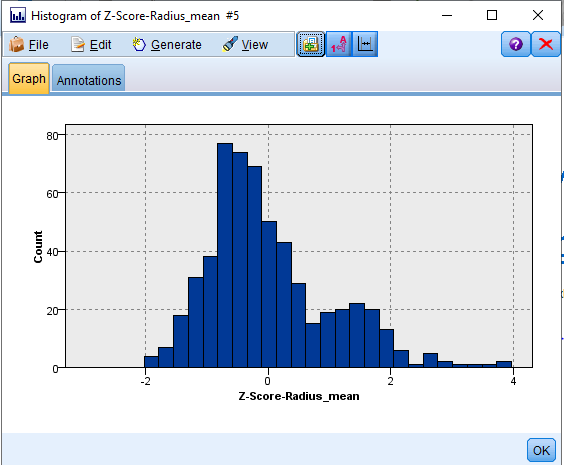
The above is the snapshot of stream till this task.

**3. Transform the radius\_mean attribute using Z-score standardization. Using a graph, describe the range of the standardized values. Hint: consider adding a Derive node to create the Z-score. You can find the required summary statistics (mean and std deviation) to calculate the Z-score using the statistics node.**

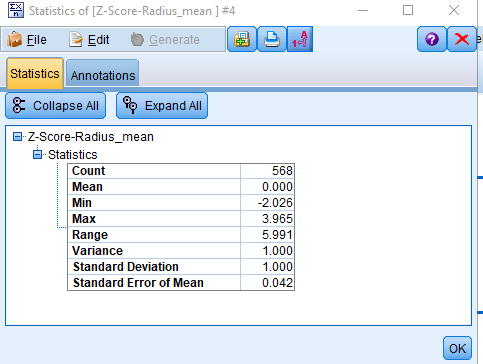
In order to calculate the Z-score standardization, we first have to calculate the mean and standard deviation which can be shown below.



The above is mean and standard deviation for given data using statics, we will be using these values to calculate the asked z-score.

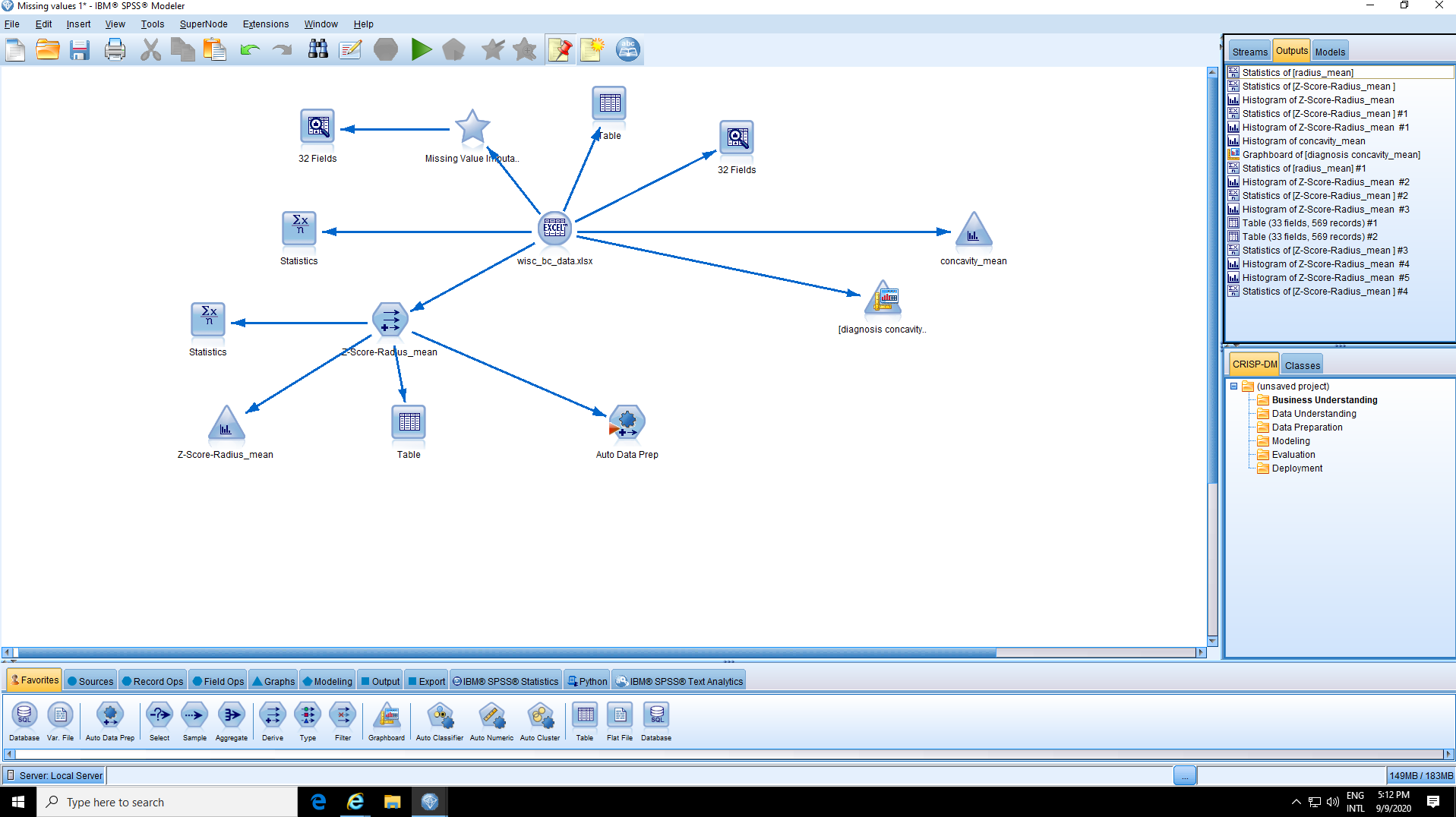


The above is the histogram graph representation of Z-score standardization, which mostly varies from -2 to 2.

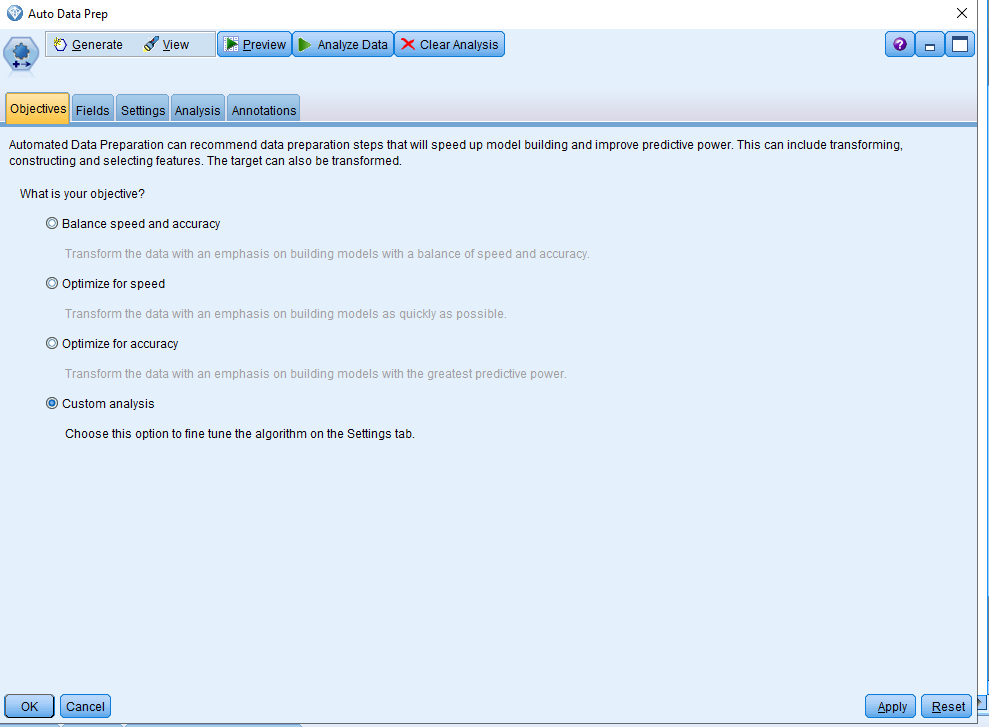


We have transformed the radius\_mean attribute using Z-score standardization (mean μ = 14.125) and standard deviation (σ) = 3.527. The Z-score ranges from about - 2.026 to 3.965 indicating the potential existence for several outliers.

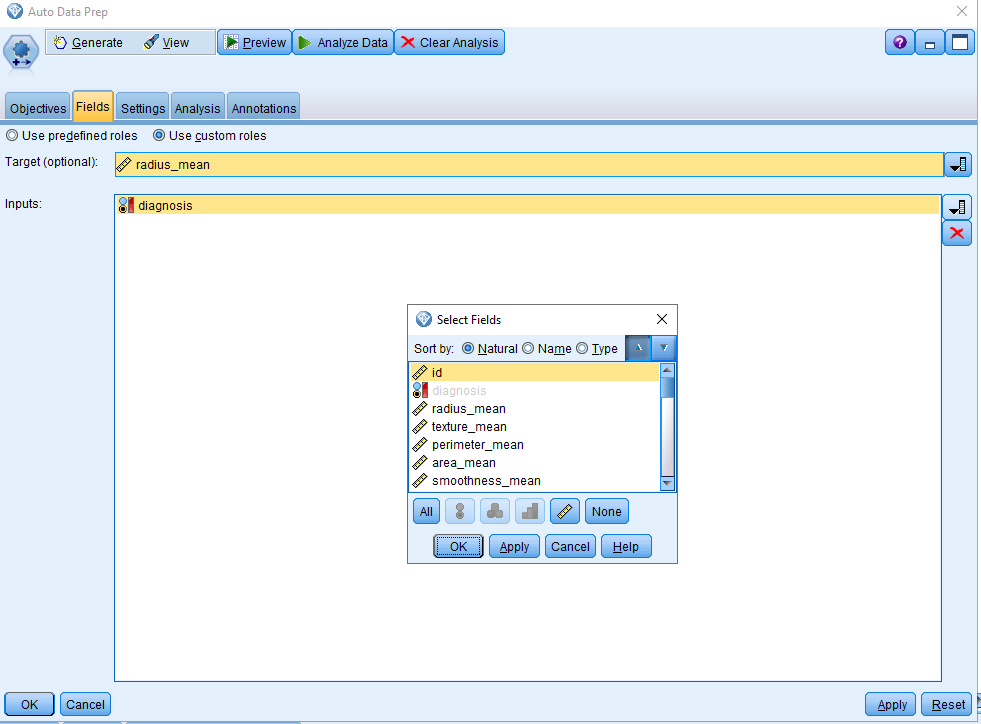
**4.How would you transform all 30 metrics to Z-scores without having to create derived nodes for each of them? Hint: Investigate the Auto Data Prep node. We did not cover this in class, but it should be very simple with a little experimentation**.



In order to calculate the Z-score for all the 30 metrics, we first need to add and connect the Auto Data prep node, which is shown in the snap shot above.

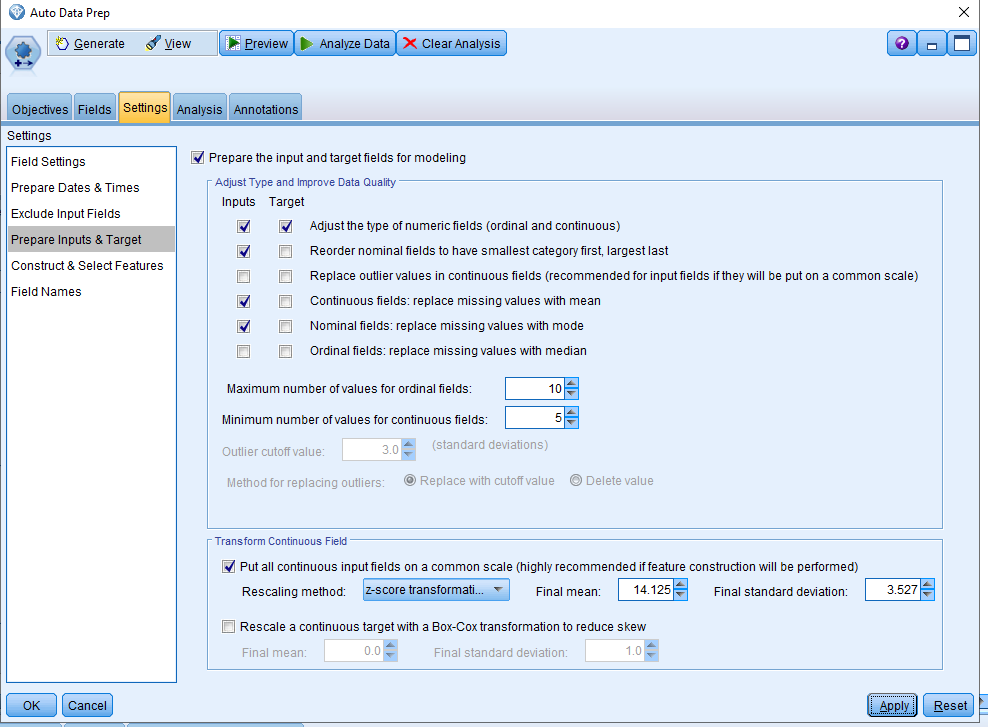


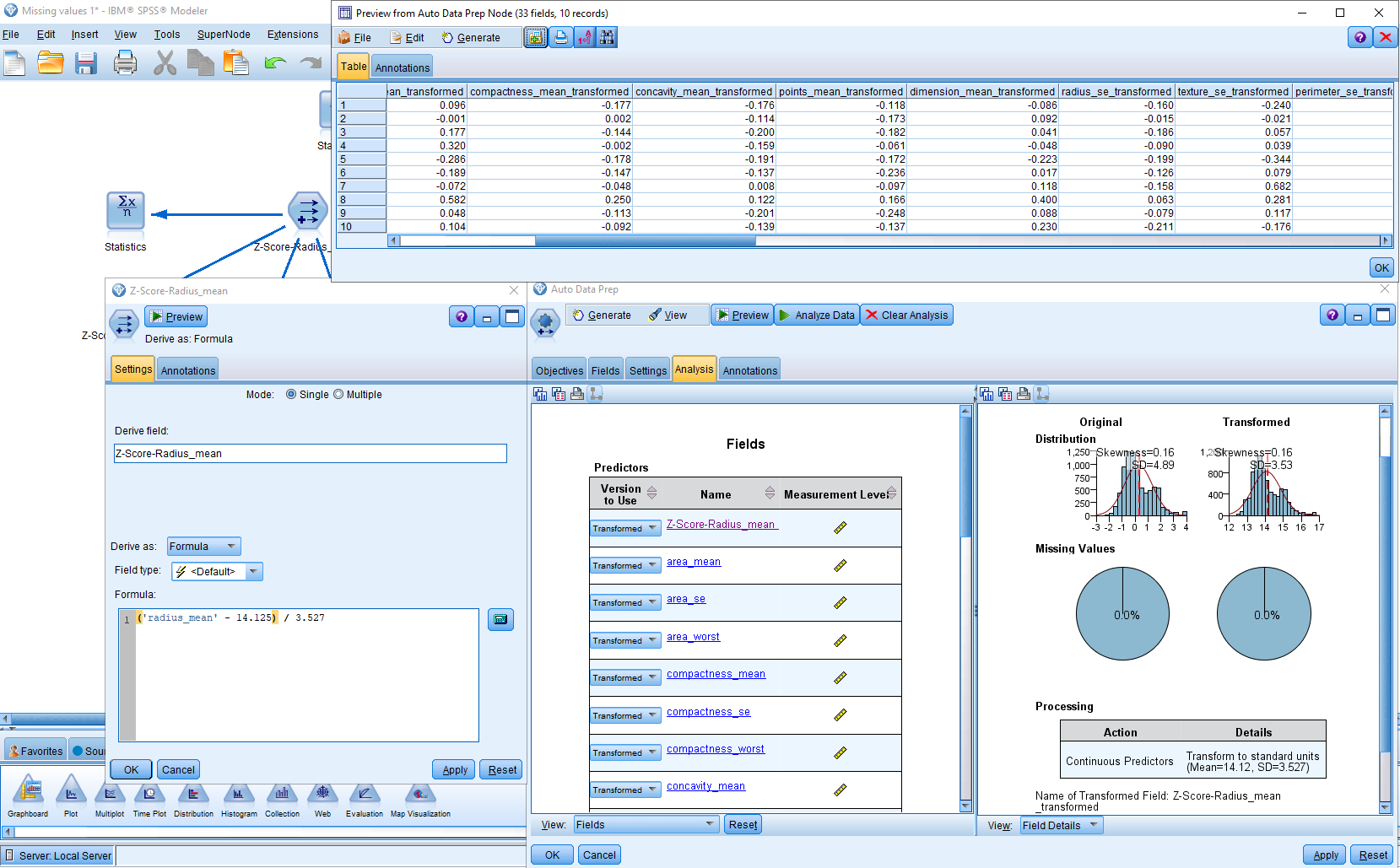
From the edit option in Auto Data prep node, we have to select the custom analysis option, as shown above.



Under the fields option we can select the target and inputs field to calculate the required Z-score, we can select any attribute, to valid our calculation we have taken the radius\_mean attribute whose Z-score we have already calculate in previous part.

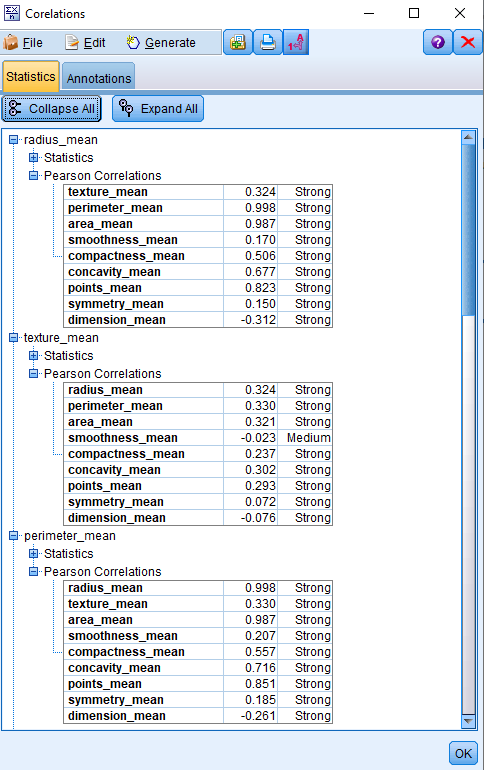
Now under setting tab and in prepare Inputs and target section, we have to insert mean and standard deviation of the selected attribute, for this case we have, mean = 14.125 and standard deviation = 3.527.

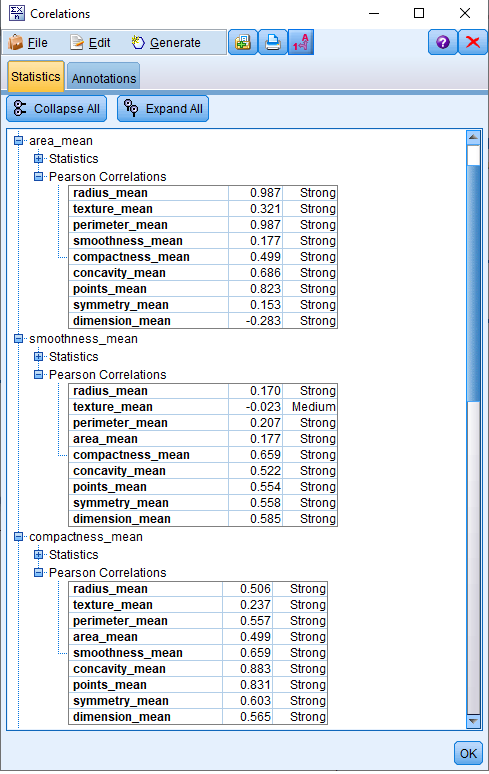


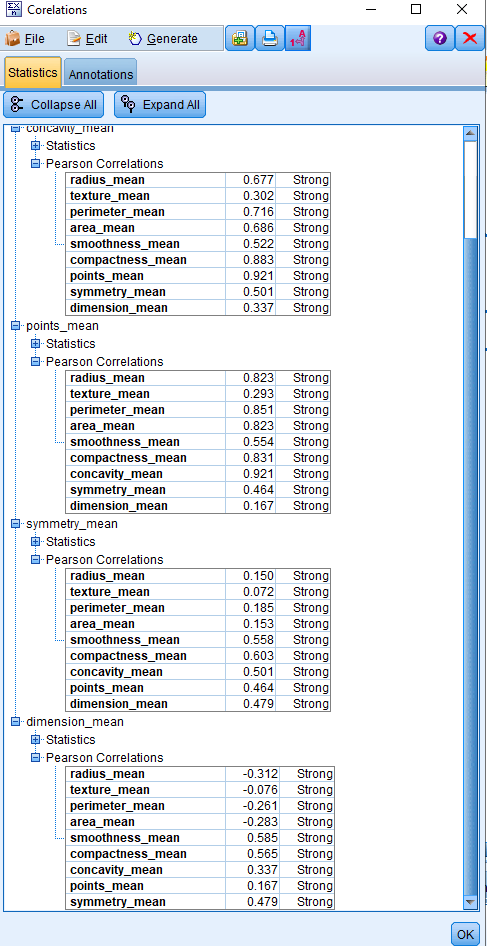


Once we click on Analyze Data from the above tab, we get the required results of selected attribute, this case the Z-score is varies in the same range which we have calculated earlier.

**5. Investigate whether there are any correlated variables among the first 10 numeric fields (the ones with \_mean prefix)**

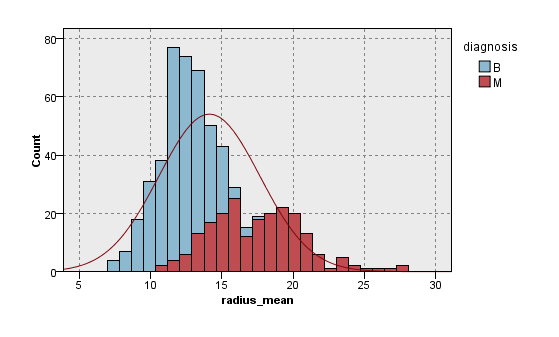


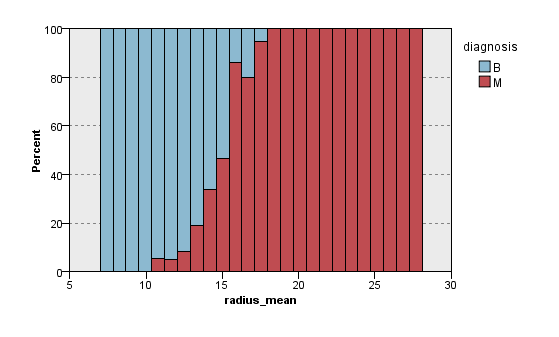




From the above we can say that there is no corelated variables.

**6. Construct a normalized histogram of radius\_mean, with an overlay of the target variable diagnosis. Explain the results.**





The above normalized histogram graph of the required attribute (radius\_mean), the diagnosis is coded as “M” to indicate malignant, or “B” to indicate benign. The graph shows that radius\_mean for (M) malignant are higher then as compared to (B) benign, these features relate to the shape and size of the cell nuclei.