

< [Return to "Data Scientist Nanodegree" in the classroom](#)

# Identify Customer Segments

REVIEW

HISTORY

## Meets Specifications

Congratulations on passing this project on very first attempt! You were flawless throughout ! You have very good hand in python. Thoroughly enjoyed going through your code. So keep it up ! For your future reference, you could use following links,

<https://www.analyticsvidhya.com/blog/2016/07/practical-guide-data-preprocessing-python-scikit-learn/>

<https://onlinecourses.science.psu.edu/stat505/node/54/>

<https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>

<https://www.analyticsvidhya.com/blog/2018/08/dimensionality-reduction-techniques-python/>

<https://www.kaggle.com/cgump3rt/investigate-missing-values>

<https://medium.com/greyatom/why-how-and-when-to-scale-your-features-4b30ab09db5e>

## Preprocessing

✓	All missing values have been re-encoded in a consistent way as NaNs.
Great work in encoding of missing values :)	
✓	Columns with a large amount of missing values have been removed from the analysis. Patterns in missing values have been identified between other columns.
Good work in identification of six outlier columns :)	
✓	The data has been split into two parts based on how much data is missing from each row. The subsets have been compared to see if they are qualitatively different from one another.
Nice work in setting the threshold to 20. Optimal threshold value lies between 9 and 32 :)	
✓	Categorical features have been explored and handled based on if they are binary or multi-level.
Brilliant work in encoding of categorical variables :)	
✓	Mixed-type features have been explored, resulting in re-engineered features.
Good work in engineering mixed type variables :)	
✓	Dataset includes all original features with appropriate data types and re-engineered features. Features that are not formatted for further analysis have been excluded.
Good work here in including only the columns that are required for further analysis :)	
✓	A function applying pre-processing operations has been created, so that the same steps can be applied to the general and customer demographics alike.
Nice work in putting everything together in a function :)	

## Feature Transformation

✓	Feature scaling has been properly applied to the demographics data. Imputation has been performed to remove remaining missing values.
Nice work in feature scaling :) Alternatively, you could also take a look at sklearn's 'RobustSclaer' <a href="https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html">https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html</a>	
✓	Principal component analysis has been applied to the data to create transformed features. A variability analysis has been performed to justify a decision on the number of features to retain.
Good work here in implementation of PCA :)	
✓	Weights on at least three principal components are used to make inferences on correlations between original features of the data. General meanings are ascribed to principal components where applicable.
Nice work in explaining about the weights of principal components :)	

## Clustering

✓	Multiple cluster counts have been tested on the general demographics data, and the average point-centroid distances have been reported. A decision on the number of clusters to use is made and justified.
Nice work in implementation of Kmeans. Alternatively, you could also take a look at 'minibatchkmeans' <a href="https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MiniBatchKMeans.html">https://scikit-learn.org/stable/modules/generated/sklearn.cluster.MiniBatchKMeans.html</a>	
✓	Cleaning, feature transformation, dimensionality reduction, and clustering models are applied properly to the customer demographics data.
Good work in applying same steps to customer data :)	
✓	A comparison is made between the general population and customers to identify segments of the population that are central to the sales company's base as well as those that are not.
Brilliant analysis !	

 [DOWNLOAD PROJECT](#)

RETURN TO PATH

Rate this review

