**CROSS SELLING RECOMMENDATION: WEEK 9 DELIVERABLES**

***Data Cleansing and Transformation***

*Submit a pdf document and ipynb notebook which should contain following details:*

* *Team member's details : Group Name (give a name to your group), Name, Email, Country, College/Company, Specialization (Data Science, NLP, Data Analyst)*
* *Problem description*
* *Github Repo link*
* *Data cleansing and transformation done on the data.*
* *Try at least 2 techniques to clean the data (for NA values : mean/median/mode/Model based approach to handle NA value/WOE and like this try different techniques to identify and handle outliers as well)*
* *Each member should code and review peers work. (Review comment should be present in the github repo)*
* *Each team member should work on different data cleansing approach.*

***Note****:*

* *If one team member is using mean to impute values then other member should experiment on segmented approach or any other model based approach to impute the null values.*
* *you are allowed to merge the code of each individual and work together to get good result.*
* *Make sure code of each team member is placed at provided URL (single repository for whole team).*

**General Task for All**

* Rename columns
* Check for nulls.
* Delete null values group agreed should be deleted.
* Replace the Spanish values in columns (Gender, Customer type, Customer relation, Employee index, Residence index, Foreigner index, Deceased index) with their English interpretations.
* Replace categorical columns values (Activity index and segmentation) with their equivalent column values.
* Check each column’s value to ensure they are correct and correspond with description.
* Change data types of columns as required to save memory.
* Check for duplicates.
* Trim special characters.
* Please add any other tasks to be done and inform the group, thank you.

**Role Assignment**

|  |  |  |
| --- | --- | --- |
| **Member** | **Task** | **Method** |
| **Ebaghae** | **Data Cleansing and Transformation.** | * **Null Values:** delete, fillna and impute using mode. * **Outliers detection:** Percentile method * **Outliers solution:** Replace outliers by imputing mean values. |
| **Yusuf** | **Data Cleansing and Transformation.** | * **Null Values:** delete, fillna, impute using mean. * **Outliers detection:** Visualizations * **Outliers solution:** Cap/Clip the outliers. |
| **Gladys** | **Data Cleansing and Transformation.** | * **Null Values:** delete, impute using mean and fillna. * **Outliers detection:** IQR method * **Outliers solution:** Cap/Clip the outliers. |
| **Harika** | **Data Cleansing and Transformation.** | * **Null values:** delete, impute using median and fillna. * **Outliers detection:** Std Deviation method * **Outliers solution:** Replace outliers by imputing median. |

**Note: reasons behind each method used should be written .i.e explain why you used a method and results from the method. Also, steps taken in cleansing data should be documented as the cleansing is been carried out for ease of understanding, and to enable group members to comment in Github repo.**