Week 9 Delivereables

Members: Uthej Goud
Email: goudtigulla@gmail.com
Country: Ireland
College: Griffith College Dublin
Specialization: Data Science
Submitted To: Data Glacier
Date: 24/07/2021
Batch Name: LISUM01

Problem Description:

XYZ bank wants to roll out Christmas offers to their customers. But Bank does not want to roll out same offer to all customers instead they want to roll out personalized offer to particular set of customers. If they manually start understanding the category of customer then this will be not efficient and also they will not be able to uncover the hidden pattern in the data (pattern which group certain kind of customer in one category). Bank approached ABC analytics company to solve their problem. Bank also shared information with ABC analytics that they don't want more than 5 group as this will be inefficient for their campaign.

Github Repo link: https://github.com/uthej12/Customer_Segmentation

Data Cleaning Approaches:

- The columns ult_fec_cli_1t and conyuemp have 99% null values so the best action would be to drop these two columns as there are not many insights that can be gained for this problem, and if these columns are used it can create a skewed model.
- The null values in renta column can be replaced by the average of column.
- There are null values in other columns but they amount to around 1% of the dataset so I have decided to remove them from the dataset, as they are overlapping null values and replacing them is not possible since they are incomplete observations.
- All columns have been assigned correct dtypes.
- Outilers from columns such as antiguedad have been removed.

Final Dataset Ready:

- The final dataset has 982105 and 14 columns.
- The outliers have been dealt with and NA values handled
- The column names have changed
- The dataset was originally in Spanish and relevant translation of column names and observation names have been retained.
- The final dataset looks like shown below

	Customer Code	Employee Index	Country	Gender	Age	New Customer	Customer Seniority(months)	Customer type	Customer relation type	Residence index	Foreigner index	Channel joined	Province	Gross Income
0	1375586	N	ES	F	35	0	6	1	A	S	N	KHL	MALAGA	87218.10000
1	1050611	N	ES	M	23	0	35	1	I	S	S	KHE	CIUDAD REAL	35548.74000
2	1050612	N	ES	M	23	0	35	1	I	S	N	KHE	CIUDAD REAL	122179.11000
3	1050613	N	ES	F	22	0	35	1	I	S	N	KHD	ZARAGOZA	119775.54000
4	1050614	N	ES	M	23	0	35	1	A	S	N	KHE	ZARAGOZA	139646.15094