## **Python Questions - Classification**

Create a Jupyter notebook, name it `python\_questions.ipynb` to load the data present on file `creditcardmarketing.xlsx` for `credit\_card\_classification`. Fill the notebook with the Python code needed to answer each question. This last file must be delivered.

1. Select all the data from table `credit\_card\_data` to check if the data was imported correctly.

2. Select all the data from the table to verify if the command worked. Limit your returned results to 10.

3. Use code to find how many rows of data you have.

4. Now find the unique values in some of columns:

- What are the unique values in the column `Offer Accepted`?

- What are the unique values in the column `Reward`?

- What are the unique values in the column `Mailer Type`?

- What are the unique values in the column `# Credit Cards Held`?

- What are the unique values in the column `household\_size`?

5. Arrange the data in decreasing order by the `average\_balance`. Return only the `customer\_number` of the top 10 customers with the highest `average\_balances` in your data.

6. What is the average of `Average Balance` of all the customers in your data?

7. In this exercise use `groupby` to check the properties of some of the categorical variables in our data. Note wherever `average\_balance` is asked in the questions below, please take the average of the column `average\_balance`. Show the results in a data frame and a plot.

- What is the average balance of the customers grouped by `Income Level`? The returned result should have only two columns, `Income` and `Average Balance` of the customers.

- What is the average balance of the customers grouped by `number\_of\_bank\_accounts\_open`? The returned result should have only two columns, `number\_of\_bank\_aaccounts\_open` and `Average Balance` of the customers.

- What is the average number of credit cards held by customers for each of the credit card ratings? The returned result should have only two columns, `rating` and `average number of credit cards`.

- Is there any correlation between the columns `credit\_cards\_held` and `number\_of\_bank\_accounts\_open`? You can analyze this by grouping the data by one of the variables and then aggregating the results of the other column. Visually check if there is a positive correlation or negative correlation or no correlation between the variables.

- Check the number of customers in each category (ie number of credit cards held) to assess if that category is well represented in the dataset to include it in your analysis. For eg. If the category is under-represented as compared to other categories, ignore that category in this analysis

8. Your managers are only interested in the customers with the following properties:

- Credit rating medium or high

- Credit cards held 2 or less

- Owns their own home

- Household size 3 or more

For the rest of the things, they are not too concerned. Write code to find what are the options available for them.

Can you filter the customers who accepted the offers here?

9. Your managers want to find out the list of customers whose average balance is less than the average balance of all the customers in the database. Write a query to show them the list of such customers.

10. What is the number of people who accepted the offer vs the number of people who did not?

11. Your managers are more interested in customers with a credit rating of high or medium. What is the difference in average balances of the customers with high credit card rating and low credit card rating?

12. In the database, which all types of communication (`Mailer Type`) were used and with how many customers?

13. Provide the details of the customer that is the 11th least `Q1\_balance` in your database.