**Capstone Project Data Science: Customer Segmentation**

**Project description:** Customer segmentation is important for businesses to understand their target audience. Different advertisements can be curated and sent to different audience segments based on their demographic profile, interests, and affluence level.

**Context of the Problem:**

Distributors use our sales force automation tools to help them track their sales, improve their operations and increase their top line.

We would like to help these distributors maximize their profits by allowing them to identify which merchants they should target.

The specific distributor that we will be analysing is Cow and Buffalo Milk company. They produce dairy products for the entire country.

Below is a list of merchants that they service along with meta data that could be useful in determining which of the merchants they should invest resources in.

**Dataset features:**

The dataset has the following features:

*Merchant Id – This is the unique Identification number that is given to a merchant.*

*Annual Revenue – This is the annual income of the merchant.*

*Spending score – It is the score (out of 100) given to a merchant, based on the money spent on distributor products and the behaviour of the customer.*

*City – The city that the merchant is located in.*

*Most Purchased Product – In terms of money spent on a particular product, this is the most popular product for that specific merchant.*

*Therefore, this is the most purchased product by that merchant from Cow and Buffalo Milk company.*

*Payment score – It is the score (out of 5) given to a merchant by a consultant, based on the ability for the merchant to repay inventory that is purchased on credit.*

*A score of 5 is great, a score of 1 is poor.*

**Problem:**

Make use of the merchant data set below to develop models that will help Cow and Buffalo Milk company target the right customers.

1. **The goal is to help Cow and Buffalo Milk company not only increase sales but also to become more efficient in allocating advertising spend.**
2. **Given the payment history that the merchant has and the cities that the merchant operates in, create a credit scoring algorithm that will help the distributor.**
3. **Figure out which are the most creditworthy merchants and which ones are not.**
4. **Create a parameter based on their nature from the cluster group they belong to, such as Careful, Spendthrift, General, Target and Miser**

**Submission:**

The submission should include a section on how to run the code as well as a write up on what you discovered.

1. **How you would you advise Cow and Buffalo Milk company on which merchants to target.**
2. **Download the final analysed data into CSV file with the below sample format having to which cluster they belong to and their nature columns for all the merchants in the data.**

|  | Merchant Id | Annual\_Revenue | Spending\_Score | City | Most\_Purchased\_Product | Cluster\_Number | Cluster\_Nature |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 1 | 15 | 39 | Mtwara | Mtindi 500ml | 2 | Careful |
| 1 | 2 | 15 | 81 | Zanzibar City | Fresh Milk 250ml | 3 | Spendthrift |
| 2 | 3 | 16 | 6 | Mtwara | Mtindi 500ml | 2 | Careful |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**Evaluation Scheme:**

**Total marks:** **100**

**Deliverables [Total marks - 95]:**

1. Loading the data into python environment and performing the data quality checks with statistical analysis – 5 marks.
2. For all the 6 tasks mentioned above, each carries 15 marks – 90 marks.

**Project Submission [Total marks - 5]:**

1. Once the project has been created, upload all the files on GitHub & commit (save) all the changes, make sure you add a readme file containing detailed description of your thoughts during the project creation. **[3 marks]**
2. Once done, kindly copy the GitHub link of your project & submit the same using your dashboard. **[2 mark]**