Building Relationships

For those couples who call it quits after a year, the main reason would be among the differences in their opinion, life styles and maybe financial stress. Therefore, ultimately even if the problem is being looked at from different scenarios and angles the main reason for any couple to break up would be because they are experiencing money troubles. After looking at the Mint transaction data, it can be inferred that the people who have broken up would be able to have a more successful current or future relationship based on the purchasing power of any individual.

Scenario A:

Say, if a person A is interested in visiting restaurants quite often but if person B is not quite keen on visiting restaurants, then there is a higher likelihood that the relationship will not sustain for long. But then again if we try to understand the root cause in this situation the frequency of a person visiting a restaurant will be directly proportional to the money that the person invests for it which holds a dependency on the individual's purchasing power.

Scenario B:

Say, if a person A is interested in travelling and travels using public transport or taxi quite often but if person B is not quite keen on travelling, then there is a higher likelihood that the relationship will not sustain for long. This may be as a result of one not being available for the other. But then again if we try to understand the root cause in this situation the frequency of a person travelling professionally or for leisure will be directly proportional to the money that the person invests for it which holds a dependency on the individual's purchasing power.

Anomalies / Outliers: There can be a few people for whom the logic might not hold true and that would be clearly marked as false positives for the confusion matrix generated after using the neural networks machine learning algorithm for supervised learning for building relationships.

I tried extracting features using the **K Nearest Neighbor** search from the scikit learn libarary in python but the results obtained where not as distinguishable and accurate as the neural network algorithm.

Solution:

Compatibility of 2 individuals can be extracted from the features from the given data, so, I have extracted out essential features that would enable us to determine the compatibility of the 2 individuals.

Now,

Purchasing Power = Income of any Individual – Basic requirements (Water, Electricity, Water, Sewer, Loans, Rent)

For instance,

Higher the Income if lesser is the basic requirements => Higher is the purchasing power

Lower the Income if higher is the basic requirements => Lower is the purchasing power

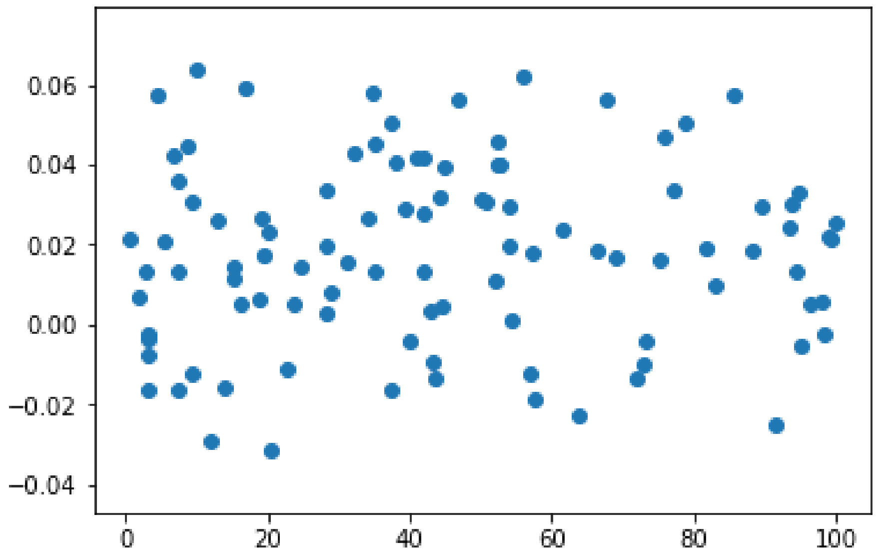

Fig 1: Scatter Plot of Mint Financial Transactions from 100 individuals
X-axis: unique authorization id Vs Y-axis: Purchasing Power (Amount)

The idea behind extracting the purchasing power of each of the individual is that this information can be used to classify individuals into different classes.

For instance,

```
Class 1: Individuals with >= 10000
```

```
Class 2: Individuals with < 10000 and > 0
```

```
Class 3: Individuals with > -10000 and <= 0
```

```
Class 4: Individuals with <= -10000
```

Neural Networks: Used to perform supervised learning by training and then learning to obtain a better accuracy over the recognition rate. The error metric used for recognizing the difference is the **"Sum of Squared Error"** and the hypothesis function used for activation of the multi-layered perceptron is the **"Sigmoid Function"** used for logistic regression in order to classify the required data. The neurons in the hidden layer can be varied and therefore a different error plot can be obtained after each run but this method eliminates the overfitting problem caused during classification of data.

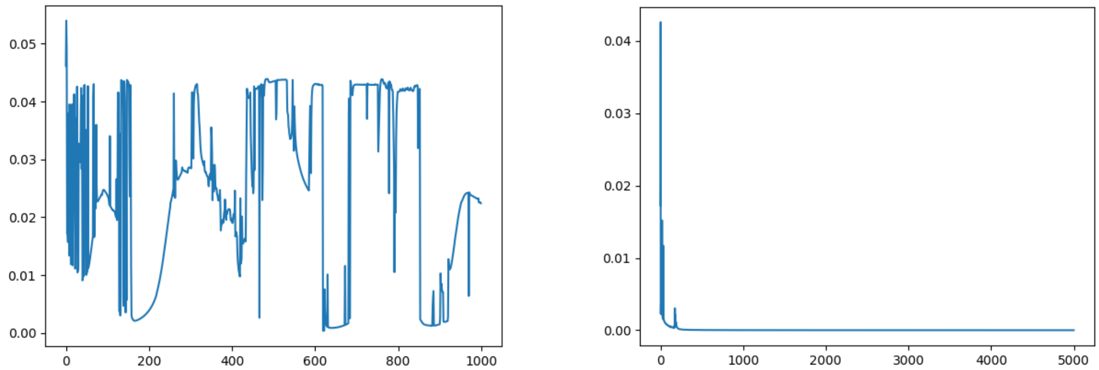

Fig: X-axis [Sum of Squared Differences] V/s Y-axis [Epochs at (1000, 5000)]

From above, it is clear that as the number of epochs increases the neural network trains itself to reduce the sum of squared error in classifying the data into their respective classes thereby making it a probable logic for logistic regression over the Mint Financial Transactions of the individuals.

----------- Epochs = 1000 -------

Recognition rate (% correct) = 85.0

Confusion Matrix: Metric to display positives, negatives, false negatives, and false positives

Actual -->	Class 1	Class 2	Class 3	Class 4
Assigned as Class 1:	64	5	0	0
Assigned as Class 2:	1	7	5	0
Assigned as Class 3:	0	0	0	0
Assigned as Class 4:	0	0	4	14

For instance, here on the first line we get that there were 64 instances which were classified as class 1 which were class 1 in actual. But there were 5 class 2 instances that were classified as class 1 as well. And similarly, even other results lead to a recognition rate of 85% by the machine learning algorithm after training the network for a 1000 epochs.

Future Work: This can be further used in conjunction with other algorithms to classify different individuals into different class and perform a voting over the maximum of the class results that each person has been classified. For e.g., if an individual from one algorithm gets class 1 and gets class 2 from other 2 algorithms then the individual may be classified as class 2 with a more certainty factor.