Problem Statement:

The CRM team of a retail store wants to:

1. Understand which factors influence the revenue generated by the customers
2. Predict the revenue "**Amount**" a given customer is likely to generate

As a team of data science consultants, you have to solve this problem and come up with a solution.

Use the Retail_Train.csv dataset that contains transactional and demographic information of customers of the retail store, to predict revenue.

Variable Description:

- PersonID - ID of the customer
- Amount - Revenue that a customer has generated for the retail store
- Family Size - Number of members in a customer's family
- Distance - Distance between customer's place and the nearest retail store
- Duration - Total time duration of interaction between customer and retail store
- DirectVisit - Number of times the customer has directly visited the retail store
- OnlineVisit - Number of times the customer has visited the store online
- Quantity - Total quantity of the products purchased
- NumberofFrequentItems - Number of most frequently purchased items
- TransactionMode - Mode of payment transaction
- Area - Area code of retail store outlet
- Occupation - Customer's occupation

Approach Hints:

Following are the pointers that may help you in creating the solution:

- Exploratory data analysis and how this analysis would help you in modeling
- Data pre-processing and reasons why have you performed what you have performed
- Decide the error metric you wish to focus on and why
- Create models using different modeling techniques
- Apply dimensionality reduction and regularization techniques
- How different factors affect revenue generated by customers
- What would you suggest to the CRM team ?