# Super Mart Case Study

**Section 01: Business Objective understanding and High-level Approach**
Lay down using a few slides - *refer Industry template.*

- **Overall Business objective** – put in simple words what's the scope of the analysis and what business problem is being solved *(this ensures that you've gone through the client's requirement).*
- **Understanding of the problem** in your own words *(this helps align your understanding of the business problem with client's requirement).*
- **Approach** – Very high-level view of the approach you plan to use to address the problem *(this helps the client see that you have a plan in place to attack the problem).*

**Section 02: Data Health Review**
Report the below results in a few slides *(ensure slide formatting and grammatical correctness).*

- Do the variables get read in Python in the right format (Integer, Float, Boolean, Date, Object)? ***List down what corrective steps were taken (if any) for the affected variables?***
- Do any variables have missing values? ***Create a list of variables for which you found missing values and what % values in these variables were affected?***
- Do any variables have outliers? ***Use suitable plots to show the outliers for all these affected variables.***
- Are there any variables that require cleaning (extra spaces, special characters, unexpected values, etc.) or replacement of values (e.g., Yes/No to 1/0)? ***List down all such variables and the kind of cleaning required.***
- Are there any duplicate records in the data? ***If yes, how many.***

1. ***Generate Extended Data Dictionary (EDD) of the provided dataset to help comment on data quality***.

**Section 03: Exploratory Data Analysis**
Perform Exploratory Data Analysis to comment on what information the dataset conveys and if it's complete/suitable to solve the given business problem. Also explore for any patterns/ insights that might guide in addressing the overall business problem. To be concrete, generate the below:

**A. Observe Univariate distributions on both Object and Numeric variables**
   a. Object variables – *Use suitable plots or visuals to show these distributions. Provide suitable commentary on what you observe for each variable.*
   b. Numeric variables – *Use suitable plots to show these distributions. Provide suitable commentary on what you observe for each variable.*

**B. Observe Bi variate distributions**
   a. Scatter plots to show relationship between relevant Numeric variables, and Cross-tabulation for relevant categorical variables, etc.

**Section 04: KPI/ Metric based questions –** These questions have a specific ask (pin-pointed) and getting to the required outcome is quite straightforward.

1. How many customers accepted the first two campaigns?
2. What is the average and median # Web Purchases for each Web Visit? *While commenting on the findings on this summary also report # of customers to consider how many customers that finding is based upon*.
3. Compare the average and median of 'total spends' of customers based on:
   1. Website visits: Customers with <10 visits versus >=10 visits?
   2. Store purchases: Customers with <10 purchases versus >=10 purchases?
   3. Discount purchases: Customers with <10 purchases versus >=10 purchases?
4. What kind of relationship 'total spends' has with different numerical variables? Use suitable plots to show these results and report your findings. For discrete variables (*numeric in nature but with low number of unique values*) applying binning to create bins with sufficient and equitable number of data points to observe this relationship (*use a different plot type than continuous variables*).
5. Show Average and Median of 'total spends' for different object variables (for their respective labels) and report your findings.

**Section 05: Open-ended questions and recommendations –** These are business-oriented questions which do not tell much about the kind of expected outcome, rather they require you to check if a certain phenomenon is occurring or not, or whether there is plausibility of a certain pattern. Often these questions need to be asked on your own and

to answer them one needs to think through in terms of: 'what kind of output is expected', 'how to get it – which variables and by doing what' and 'whether the achieved outcome helps answer the question'. Since this can be iterative, it requires a lot of brainstorming and asking the right questions (as per the business objective).

**Set A**

1. What kind of customers spend more?
2. What kind of customers had a higher Website conversion?
3. "Website visits helped drive overall sales". Justify or refute this statement using an appropriate summary/analysis/result.
4. Is there any relationship between sales of Meat and Fish products?
5. What kind of customers responded the most in the last campaign?

**Set B**

6. What kind of customers spend more on different product types?
7. What kind of customers made more Website visits?
8. Is there any relationship between sales of Wine and Meat products?
9. Did certain customers prefer Web Purchase over Store Purchases? If yes, how were their profiles different?
10. Is customer complaints an area of concern that should be closely looked at? Show appropriate metrics to support your response.

*End output expected is a PPT. Keep Python codes, analysis excel files, etc. as a back-up for the final presentation.*