



Case Study Assessment

Phase 01 Deliverables *(End output expected is a PPT. Keep Python codes, analysis excel files, etc. as a back-up for the final presentation)*

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** – refer *Industry template*

Section 03: Exploratory Data Analysis

Perform Exploratory Data Analysis to comment on what information does the dataset convey 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.

Phase 02 Deliverables *(End output expected is a PPT. Keep Python codes, analysis excel files, etc. as a back-up for the final presentation)*

These deliverables will comprise of Two types of questions

Section 04: KPI/ Metric based questions – These questions have a specific ask (pin-pointed) and getting to the required outcome is quite straightforward. **Examples** *(with reference to the campaign data covered in lectures):*

- *What were the Top 3 months with the highest number of customers contacted?*
 - *For which months the # customers contacted were > 200 AND the % conversion was > 11%*
 - *What is the average call duration for the converted customers?*
1. How many unique values do each of these variables have - Customer ID, Order ID, and Product ID? What does this information tell us about the granularity of input data?

Answer all the remaining questions (both section 04 and 05) by taking into consideration the 'data granularity' understanding you just developed

2. How many unique customers has the furniture company catered to in the last 4 years? Report what % of these customers gave single orders versus repeat orders (over 4 years)?
3. Based on the order count in previous question, show distribution of % customers to show how many customers gave 1 order, 2 orders, 3 orders,.... Report any insights relevant for the firm.
4. For different Product Sub-Categories prepare a summary to report the below metrics:
 - I. Total Sales
 - II. Total Quantity
 - III. Average % discount
 - IV. Total profit
 - V. Total Shipping Cost

What relevant insights can be drawn from this summary?

5. Which is the biggest market for the firm? What are the top 3 products in terms of Sales for each of the markets?
6. Compute # days it takes between order being placed and shipping. Using a frequency distribution comment on “what % of orders (unique orders) take how long in shipping”
7. What kind of relationship ‘Sales’ 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*)
8. Show Average and Median of ‘Sales’ for different object variables (for their respective labels) and report your findings. Create these summaries only for relevant object variables (decide based on your understanding of data)

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). **Examples** (with reference to the campaign data covered in lectures):

- Does the previous campaign have any impact on customers?
- Did mode of contact have a different effect on conversion?
- Did more contacts result in a higher conversion?

Set A

1. How can one interpret each record in the provided data? What kind of information does it capture?
2. Are there any customers who have maintained or increased order volumes over the 4 years of ordering? Show appropriate summaries and provide your recommendations on what the firm can do with these insights
3. If x \$profit for every y \$sales is used as a metric to measure profitability (for e.g., \$0.20 profit for every \$1 sales), which products have given the highest profit returns over the years?
4. Usually, companies offer discounts to boost product sales and in a way profit. How has discounts on orders worked for this furniture firm- show these results using a suitable chart? Based on what you observe from your analysis, make relevant recommendations on strategic actions that the firm can take on offering discounts

Set B

5. Which markets, customer segments, product categories are doing well - in terms of \$sales and \$profits (over time)
6. The firm's portfolio comprises different products which contribute to the total yearly sales. How has the firm's portfolio changed over time in terms of product sales and which products have done better/worse?
7. How much time does the firm take on average to deliver different product categories and for different countries? Does this comparison need to be done on an overall level or for shipping modes separately? Is there any scope of improvement that you can suggest?
8. Do certain products see a high demand (in terms of # orders) in certain months of the year? For the products (Phones, Chair, Furnishings and Art) check for any cyclic patterns over the duration of 48 months. If yes, provide relevant recommendations so that the firm can boost or reduce its production accordingly