Thinking Process

Section 1- data assembly

- Inspect the data visual to get an initial understanding of what is in each CSV file after downloading.
- To create a model-ready table with the specific features: match the order numbers with the relevant customer ID and relevant seller ID to merge all the CSV files correctly.
- Extract relevant information for the project from the data frame
- Define binary target: if the actual delivery date is before the estimated delivery date = 0; if the actual delivery date is after the estimated delivery date = 1

Section 2 – Exploration Analysis

- Based on each required insight, decide which calculations are most relevant
- Rate: Of the total orders, what is the average amount experiencing a delay (in the relevant category)
- Calculate the number of delays per product category
- Calculate the correlation between freight cost and speed (delayed or not delayed).
- Additional visualisations for interpretation.

Section 3 – Predictive Modeling

- Extract relevant features for modelling: the goal is to predict whether the order will be on time (0) or delayed (1) based on various features
- Removed orders not yet delivered to create a final modelling table
- Run the training of the model (split the dataset into train and test) and obtain results
- Optimization and summary
- Final visualization of data for easier interpretation and to ‘mess around’ with
- Geographic map of delay rate may give a more overall view of information – try this