**Problem statement**

The challenge presented is to develop a robust predictive model capable of forecasting future sales figures for these companies. The objective is to mine the dataset to identify trends, patterns, and potential causative factors that significantly impact sales outcomes. By integrating historical sales data with predictive analytics, the goal is to achieve accurate sales forecasts that companies can utilize to make informed strategic decisions, optimize inventory management, and allocate resources effectively.

**Submission Instructions**

1. Rename the submission notebook as "NUS\_DATATHON\_SINGLIFE\_<TEAM NAME>.ipynb".
   1. [For example, if your team name is "NUS 1", your submission notebook should be named as "NUS\_DATATHON\_SINGLIFE\_NUS 1.ipynb".]
2. The notebook already contains a template function, named testing\_hidden\_data, to take in the test data and output the predictions. You are to fill in the function with your model and any preprocessing function that you may have.
   1. Note: Each team should only submit one notebook and the notebook should be in the master branch.
3. You can save your model and upload it in the GitHub repository provided that it does not exceed the file size limit on GitHub (see Section 3 on instructions for large models). You are to provide a script the load the model. Please name the loading function as load\_model.
4. If the model is not saved, we will assume that the model is trained using the training function and the output of the training model is a model that can be is trained using the template function, testing\_hidden\_data, provided in the notebook. Please name the training function as train\_model.
5. If your model is too large to be uploaded to GitHub, you can upload it to Google Drive and share the link with us. Please ensure that the link is accessible by anyone with the link.
6. Please provide a LOAD\_MODEL\_INSTRUCTIONS.md file in the GitHub repository to explain how to load the model. Note that you are strongly discouraged to use such large models.

**Judging Criteria**

Datathon Comm will judge all submissions then submit top 10 to the judges

| Data Analysis (60%) | * Data Cleaning   + Effectiveness of Null Value Management, Data Consistency, Data Accuracy, Robustness. * Feature Selection   + Appropriateness of technique used, Rationalization, Impact on Model, Innovativeness * Feature Engineering   + Selection of Methods, Impact on Analysis, Creativity, Scalability * Depth of Analysis   + Level of complexity, Coverage of various the data and problem, Insightfulness, Clarity, Relevance * Leveraging Domain Knowledge   + Effective use of domain knowledge in framing and conducting analysis, Creativity, Integration and Impact. * Visualizations of the Data/Clustering   + Clarity, Relevance, Clustering Techniques, Aesthetics and Insightfulness. |
| --- | --- |
| Modelling (30%) | * Handling Highly Imbalanced Samples   + Techniques Used, Effectiveness and Impact * Choice of Model   + Model Fit, Justification, Complexity vs Performance, Data Suitability |
| Performance (10%) | * Performance of Model   + AIC, Adjusted R^2, Balance and Context Relevance |
| Additional Considerations (non-scored, for tiebreakers) | * Insightfulness   + The degree to which the insights derived from the analysis are interesting or novel. |