**Solution Approach to Problem Statement**

* Exploratory data analysis is performed as illustrated in the presentation. No null values are present in the data, so no imputation is required. No obvious arithmetic/logical relations exist between columns, so no new features have been derived.
* Outlier removal is not performed since outliers may provide unique insights (due to nature of data), i.e., high sales during festivals/holidays, or low sales for a newly opened outlet, as well as to maximize data available for training.
* Numerical columns are scaled to range [0, 1].
* Categorical columns are encoded as a one-hot numeric array. If an unknown categorical feature is present during transformation, it is ignored.
* Data is partitioned by splitting into 10 consecutive folds (without shuffling). Each fold is then used once as a test set while the 9 remaining folds form the training set.
* Initially, all features are selected and the model is fit and evaluated. Less relevant features are later removed using permutation importance.
* Boosting methods almost always outperform other regression techniques, so only boosting techniques are tried, namely CatBoostRegressor, LGBMRegressor and XGBRegressor, using all 8 features.
* Next, the ensemble model is built using a Voting Regressor, comprising the 3 models, and is observed to outperform individual models (lowest RMSE over 10 folds of training data).
* Hyperparameter Tuning is performed for VotingRegressor, and the only hyperparameter set is iterations for CatBoostRegressor, set to 100. For LGBMRegressor and XGBRegressor, default models are used.
* Permutation Importance (PI) is used to evaluate feature importance and remove features with low/negligible impact on model performance. Initially, all 8 features are used to train the model. Here, column Item\_ID has the lowest PI, so it is removed.
* Permutation Importance is again calculated after training the model using 7 features this time. Item\_W has the lowest PI score, so it is removed.
* Next, PI scores are calculated after training using 6 features. Outlet\_Location\_Type has the lowest PI, so it is removed. Model is then evaluated using 5 features, however, RMSE increases, so Outlet\_Location\_Type is added back to the feature set and 6 features are finalised.
* Model is now trained using all training data, and this final model is used to make predictions for the test set.