

Market Sales Analysis for Client *ABC*: ML Models

Problem Statement

The client *ABC* is a US based large retailer of apparel and accessories. *ABC* management is planning to expand their business by opening more shops in different locations across the US. Details of the existing shops are available in the data source shared.

This is a continuation of the Set 1 exercise, but now *ABC* management want you to use Machine Learning methods to answer some of the business questions listed below.

Dataset

The data is a subset of set 1 and it contains the sales of and other characteristics of 691 stores. The sales column corresponds to average monthly sales at the store during the period Jan 2019 to Dec 2019.

Dataset: **DS Internship – Modeling – Data.xlsx**

Machine Learning

With this data, *ABC* management want to identify potential locations for opening new stores. One way we can do this is to predict the sales of new stores given the characteristics of the new store and use the predicted sales (based on ML algorithms) to make the decision. Please follow the instructions below to solve the problem:

1. Build a linear regression model to predict the sales of the stores and identify key drivers of sales.
2. What is the MAPE of the regression model? What other metrics can be used to determine model performance?
3. Report model performance and model summary of final model in an excel file. Present the results in an easy-to-read format.

Guidelines for ML Model

1. Create train-test datasets to avoid overfitting.
2. Build a Linear Regression model using OLS and summarize the inference.
3. Use appropriate variable selection methods to find the significant features.
4. Try to address multicollinearity through feature engineering.
5. Consider doing residual analysis to see any patterns in residuals and see where the model is not predicting well or check model performance in store groups.