

Problem 1: Sales prediction:

Idea: we will build a linear regression model to predict Sales using month data for dummy Shampoo dataset.

We will divide our problem into:

- 1.Loading data and cleaning**
 - 2.Splitting data to training and testing sets**
 - 3.Building Linear Regression model**
 - 4.Calculate mean square error**
-

First: Loading data and cleaning:

- We will read the data using pandas.
- We will inspect data and check for null values
- After checking that there are no null values, we must convert data to Timestamp for the sake of building Linear Regression model
- A timestamp is a sequence of information used to find when a particular event occurred, giving the date and time of the day

Second: Splitting data to training and testing sets:

- We must split data into training and testing sets.
- We will reshape the data to a shape that the model would like to see before it can do any predictions.
- It is usually good to keep 70% of the data in train dataset and the rest 30% in test dataset.

Third: Building Linear Regression model:

- We will define **Linear Regression model** and fit train data and test data to it.
- After plotting, we realized that our model is not good.

Fourth: Calculate mean Square error:

- we will calculate mean square error for both training and testing data

Resources:

1. **<https://realpython.com/linear-regression-in-python>**
2. **[/https://towardsdatascience.com/the-complete-guide-to-linear-regression-analysis-38a421a89dc2](https://towardsdatascience.com/the-complete-guide-to-linear-regression-analysis-38a421a89dc2)**
3. **<https://www.analyticsvidhya.com/blog/2021/03/a-practical-tutorial-to-simple-linear-regression-using-python/>**
4. **<https://www.geeksforgeeks.org/>**