Context

Supplement Sales Prediction

Your Client WOMart is a leading nutrition and supplement retail chain that offers a comprehensive range of products for all your wellness and fitness needs.

WOMart follows a multi-channel distribution strategy with 350+ retail stores spread across 100+ cities.

Effective forecasting for store sales gives essential insight into upcoming cash flow, meaning WOMart can more accurately plan the cashflow at the store level.

Sales data for 18 months from 365 stores of WOMart is available along with information on Store Type, Location Type for each store, Region Code for every store, Discount provided by the store on every day, Number of Orders everyday etc.

Your task is to predict the store sales for each store in the test set for the next two months.

Content

Train Data

Variable Definition

ID Unique Identifier for a row

Store\_id Unique id for each Store

Store\_Type Type of the Store

Location\_Type Type of the location where Store is located

Region\_Code Code of the Region where Store is located

Date Information about the Date

Holiday If there is holiday on the given Date, 1 : Yes, 0 : No

Discount If discount is offered by store on the given Date, Yes/ No

#Orders Number of Orders received by the Store on the given Day

Sales Total Sale for the Store on the given Day

Test Data

Variable Definition

ID Unique Identifier for a row

Store\_id Unique id for each Store

Store\_Type Type of the Store

Location\_Type Type of the location where Store is located

Region\_Code Code of the Region where Store is located

Date Information about the Date

Holiday If there is holiday on the given Date, 1 : Yes, 0 : No

Discount If discount is offered by store on the given Date, Yes/ No

Sample\_Submission

Variable Definition

ID Unique Identifier for a row

Sales Total Sale for the Store on the given Day

Evaluation

The evaluation metric for this competition is MSLE \* 1000 across all entries in the test set.

Public and Private Split

Test data is further divided into Public (First 20 Days) and Private (Last 41 Days). You will make the prediction for two months (61 days).

Your initial responses will be checked and scored on the Public data.

The final rankings would be based on your private score which will be published once the competition is over.

The sales column that we submit would be compared to the actual answer similar to the following. Instead of 8 items it is 22266 items(the function is avable in sklearn).

Sample Input :

actual = [27.5, 55.9, 25.8, 17.7, 27.6, 55.9, 25.7, 17.8]

predicted = [24.0, 49.1, 21.0, 16.2, 23.3, 47.0, 12.1, 15.2]

(meansquaredlog\_error(actual, predicted))\*1000

Sample Output:

82.9949678377161

Public and Private Split

Test data is further divided into Public (First 20 Days) and Private (Last 41 Days). You will make the prediction for two months (61 days).

Acknowledgements

We wouldn't be here without the help of others. If you owe any attributions or thanks, include them here along with any citations of past research.

Inspiration

Your data will be in front of the world's largest data science community. What questions do you want to see answered?