

Argoid Analytics Internship

Assignment2

Summary:

I have clearly mentioned all the steps in my ipython Notebook.

So here is the brief of the whole Project.

I analyzed the data first and differentiated the data according to the Country_basis, Sales-basis and number_of_sales basis.

After analyzing and visualizing the whole data I decided to go through frequency-Matrix approach.

I created my own Machine Learning Model from scrap using the TransactionId and ItemCode.

Frequency_Matrix is an NXN matrix. (where N is the number of unique items)

First,

I created a Dictionary which Assigns each Item Code to a simpler code between 0 to 3406 [Total Number of Unique Items is 3407]

Then Using my Fit Function I Created the Matrix.

Matrix stores the frequency of each item bought together.

i.e. $\text{Matrix}[3][4] = 11$

means Items with Simpler code 3 and 4 are bought together 11 times.

Second,

I added to new features month_year and Total Cost which helped in better understanding of the data. Description of each Function is provided in the notebook also.

My Model used 3 functions:

1. Fit (Takes two Panda data Frame columns as input)
2. Predict (Takes one item at a time)
3. Name

The overall Model works on the Principle of common ness between two products which is analyzed using the vast dataset. Each time a Transaction Contains 2 or more items it changes the values in the matrix and that matrix in the end is used to determine or recommend the next two products for that item
