Group: Coding | Section: Language Coding

Q : 1 | Mark:10

After opening their shops at various location of Hyderabad in last 2 years, Bigbasket wants to revamp their existing shopping malls. Lest the efforts would go in vein, they hired a data scientist and provided him the sales-data. This data for 10,8405 products sold from various stores tells about the product and the store selling them in the city (details are mentioned in Data section). The scientist aims to build a model to predict sales for a product at a particular store. Using this model, Bigbasket will try to understand the properties of products and stores which play a key role to increase their sales.

Data
The product related data is given in Product.csv (Click on the file name to start the download)

Variable	Description
Product_id	Unique product ID
Product _Weight	Weight of product
Product _Fat	Whether the product is low fat or not
Product _Type	The category to which the product belongs
Product _MRP	Maximum Retail Price (list price) of the product

The store related data is given in Store.csv (Click on the file name to start the download)

Variable	Description
Store_id	Unique store ID
Store_establishment_Year	The year in which store was established
Store_area	The size of the store in terms of ground area covered
Store_City	City in which the store is located
Store_Citypopulation	
Store_Locality	Locality type
Store_type	Whether the outlet is just a grocery store or some sort of supermarket
Store_Parking	Number of parkings
Store_hasPump	Number of pumpstations

The data connecting the products with the stores is given in ProductStore.csv (Click on the file name to start the download)

The data connecting the products with the stores is given in ProductStore.csv (Click on the file name to start the download)

Variable	Description
Product_id	Unique product ID
Store_id	Unique store ID
Product_Stock	Number of products available at store
Product_Visibility	The % of total display area of all products in a store allocated to the particular product

whereas, the sale related data is provided in Saleslog.txt (Click on the file name to start the download)

Variable	Description
Product_id	Unique product ID
Store_id	Unique store ID
Sale_Date	Date of sales
Sale_Amount	Amounts of products sold

Integrate the datasets properly; divide them into two parts (training data 80% and testing data 20%), predict sales for a product for a particular store based on the details on the products and the stores.

Task
1. Predict Item_Outlet_Sales for the training and testing data.

Evaluation
The performance for the task will be evaluated using RMSE (Root Mean Square Error) measure.

Deliverables:
1) Your Code - If you are using Python, select your version from language drop down then upload your .py file using the "*Upload code as file*" - else if you are using markdown, Jupyter notebook, knitpy, Java etc. - Use "*Upload Attachments*" button at the bottom.
2) A write up outlining your approach. The file should be named as "*approach*" with appropriate extension (approach.txt or approach.pdf or approach.docx)