Recommeding the products and their sales to the sales team

Jupyter Notebook

Approach Of Solving Dataset

1. Data Loading
2. Data Preprocessing
3. Exploratory Data Analysis
4. Label Encoding
5. Model Building
6. Model Input
7. Note

Data Loading

- Here in these step I have just load dataset using pandas library
- See dataset using head function and check the dataset using shape function

Data Preprocessing

- In these analysis I do some analysis on this dataset
- Attribute Information:-This method give us the basic information of all the attributes like count of attributes, number of rows & columns , numerical attributes and categorical attributes and so on
- get_missing_values:-It is a private method ,so it cannot accesed by object outsid ethe class.This function will give us a basic information like count of missing values
- Agg_Tabulation:-This method is a extension of additional inforamtion like about the data like Entrophy value,missing value percentage and some observations

- outlier_count :-It is a private method which return you a outliers present in a interquartile range
- num_count_summary :-This method which returns you basic information about Numerica variable like Positive values,Negitive Vlues,Unique Count,Zero count positive and negitive,infiniyt-count and count of outliers
- statstical_summary :-This method which return you a various percentile of the data including count and mean

Exploratory Data Analysis

- The Purchase of product 'B' is comparatively higher than other products, on the other side the sales of the Product 'E' shows dull in sales.
- In these plot we says that origin of sales through social media is very high compared to others and disply through sales is very less
- Sales through East region is very high and North region is very less
- From the above plot we can observe that all Customers are not tried to using all the products, their using the products with their own reference.
- The trending of the Product 'A' in all region's is good, as it produces a more sales.Also Product 'B' shows a same trend in all regions.
- The regions Known as 'Central','East','North','South' are customers tried to refund the products, but in the west region no Refundant is happen
- Did u see that some of the products like 'A','B','C','F' products are tried to refunded but 'G','D' & 'E' are not refunded.

Label Encoding

- **Label Encoding** refers to converting the labels into numeric form so as to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way on how those **labels** must be operated.
- So in this dataset we have converted object columns into integer

Model Building & Model Input

- In these we have created RandomForestRegressor and use cross_val_score to predict the sales
- Give some model input to the dataset so it is very usable to build a model.

Note

1.Not all the Customers are using the all products, their using products on their own preference.

2.Refundent of products happen is all regions except west region.

3.Product A,B are showing higher trends in all regions , in other hand product 'G' shows less impact.

4.In the Dataset we can see that the 'origin' column different for different customer so we need to handle that.