ABC Exploratory Data Analysis

# 1 Basic details

a. Number of product column is 1

b.Total rows are 342 and columns are 3.

c. Total number of products are - 10

d.1 Time duration of sales from 2019-10-01 to 2023-03-01

d.2 Total years 3, months 5 and days 0 in time series

d.3 Total number of the Months are 41

d.4 Total number of the Weeks are 179

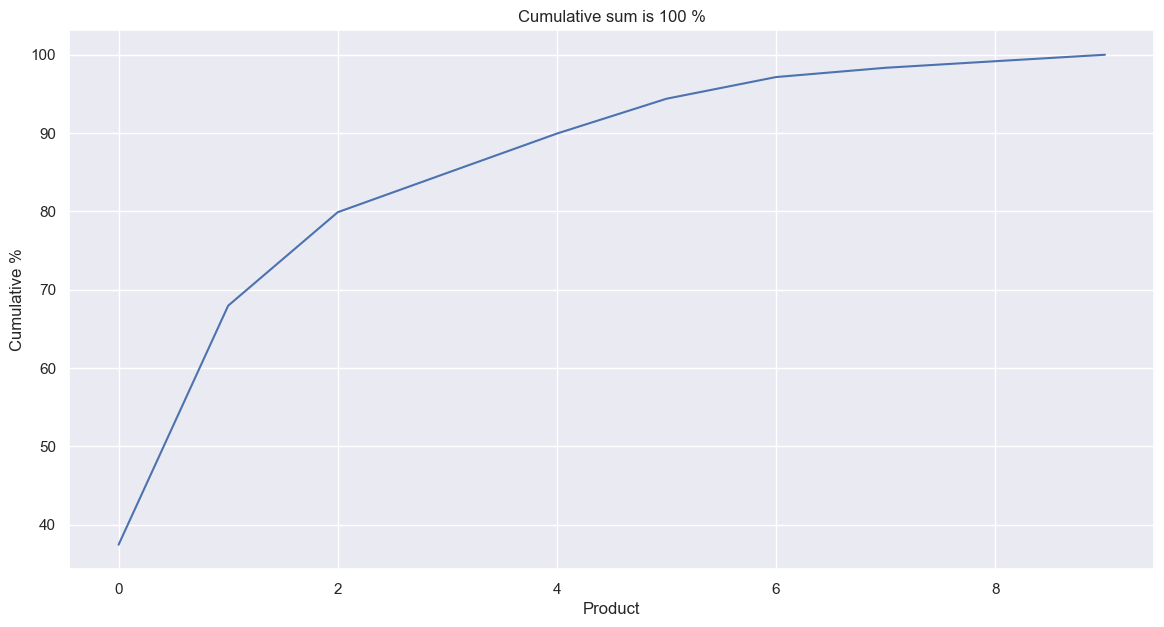
e.1 Max avg sale 1080327.79 and product is Detol-Pune

e.2 Min avg sale 23833.57 and product is -> Book-Delhi

e3. Last Date of Time Series Data is Mar 2023

e3. Last Date of Time Series Data is Mar 2023

# 2.ABC Analysis



Total grains destribution (80,95,100)

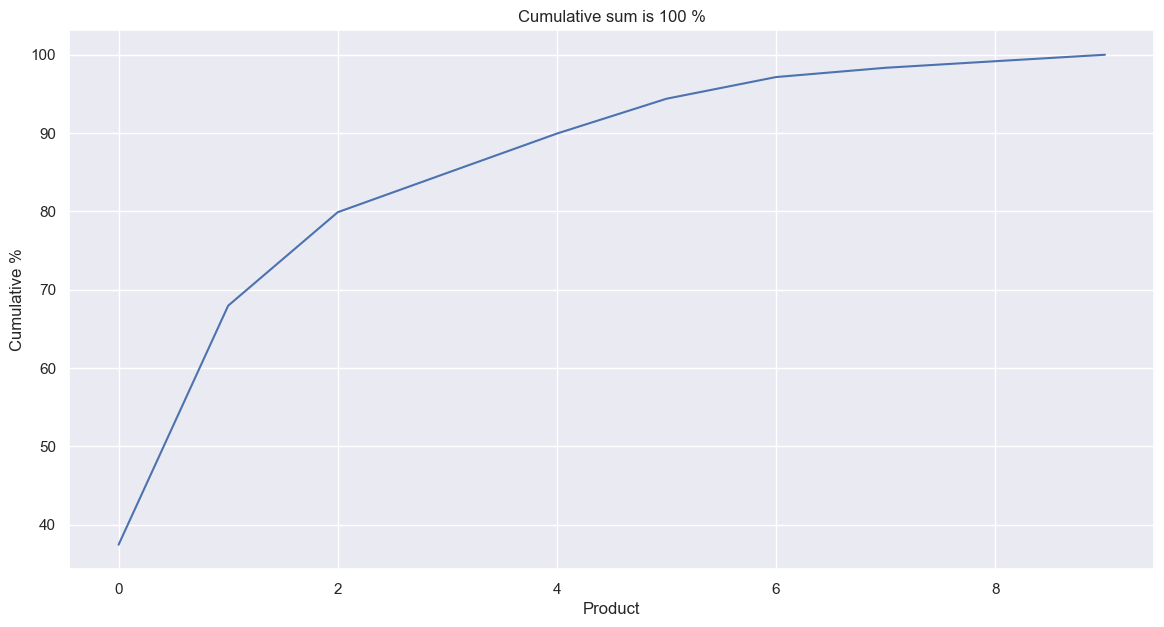
Number of grain in A catetory are 3

Number of grain in B catetory are 3

Number of grain in C catetory are 4

e3. Last Date of Time Series Data is Mar 2023

# 3.ABC Analysis

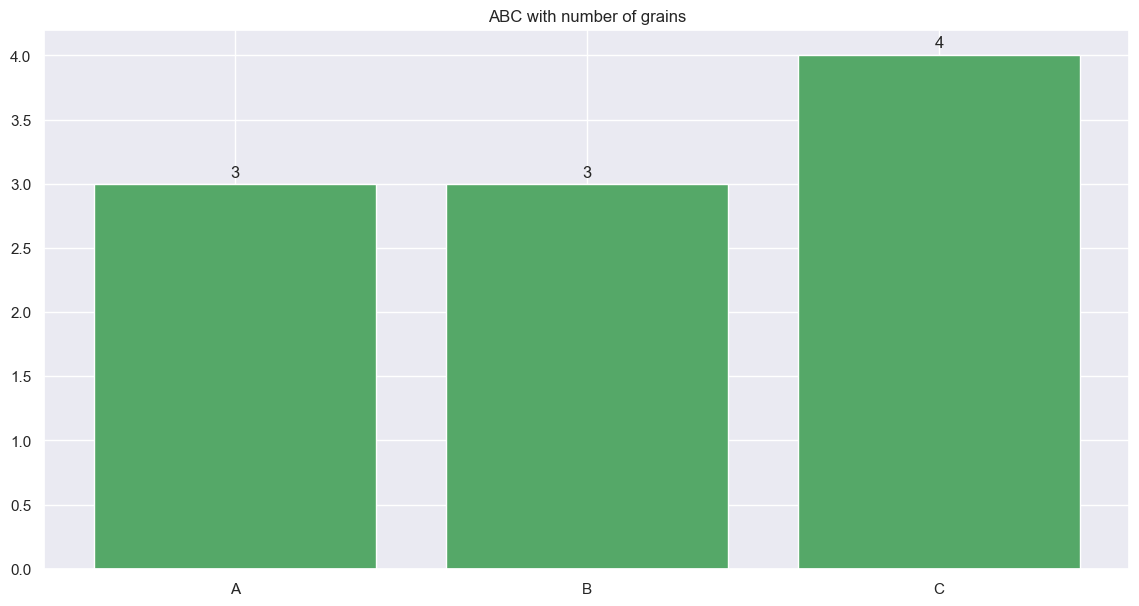


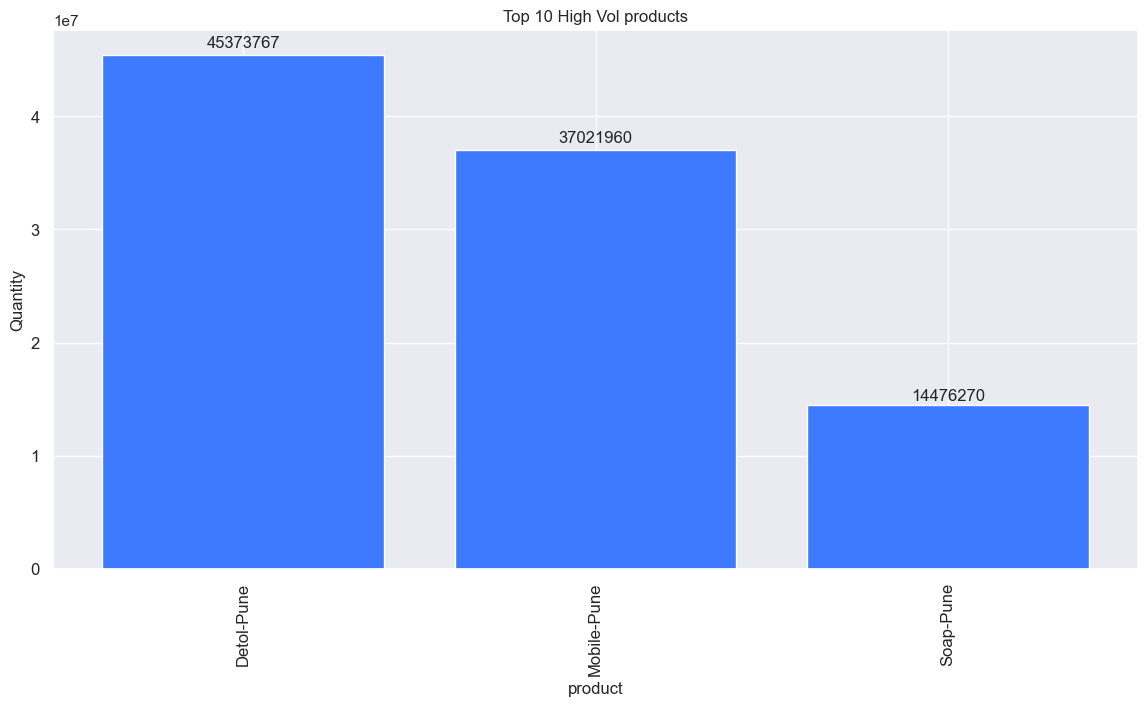
Total grains destribution (80,95,100)

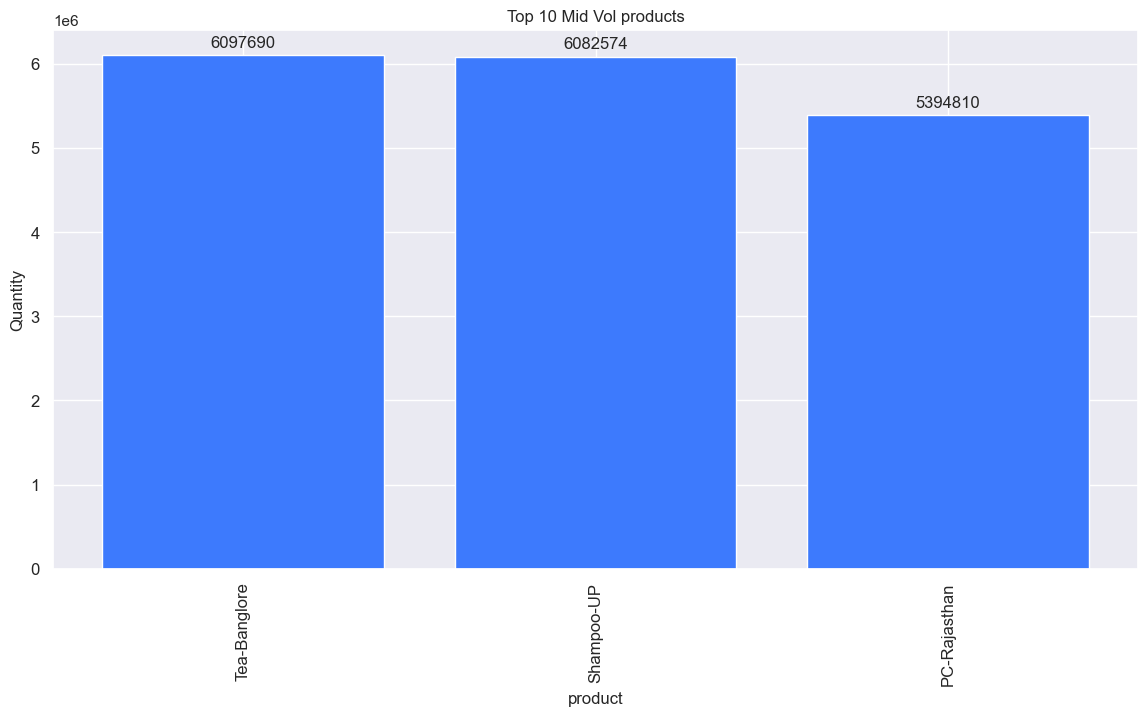
Number of grain in A catetory are 3

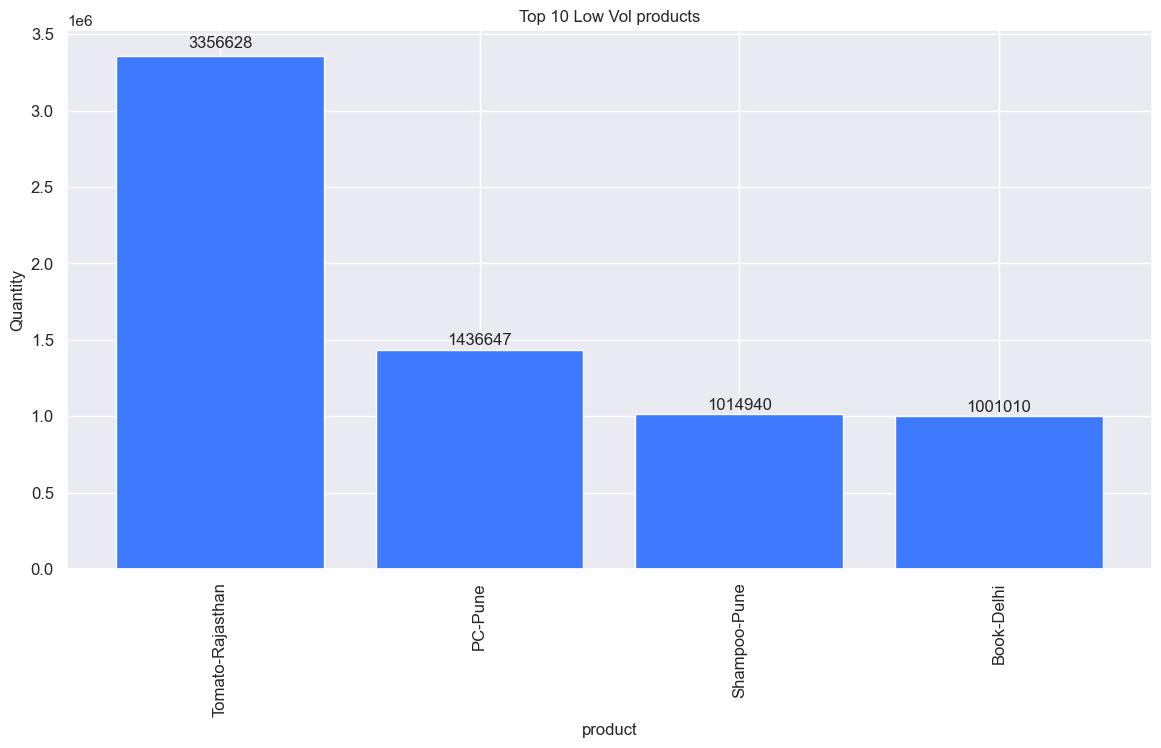
Number of grain in B catetory are 3

Number of grain in C catetory are 4

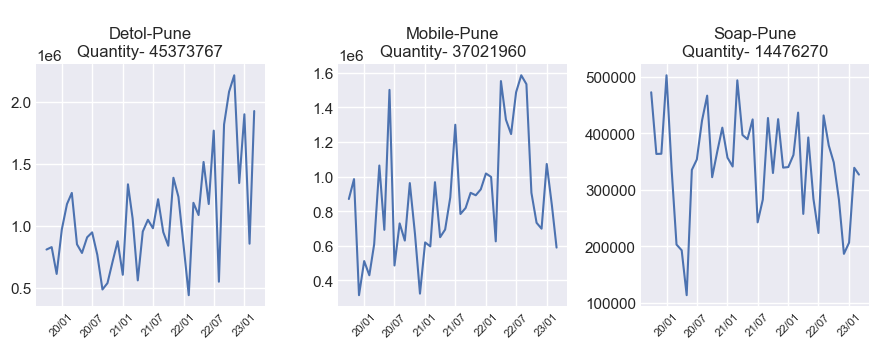




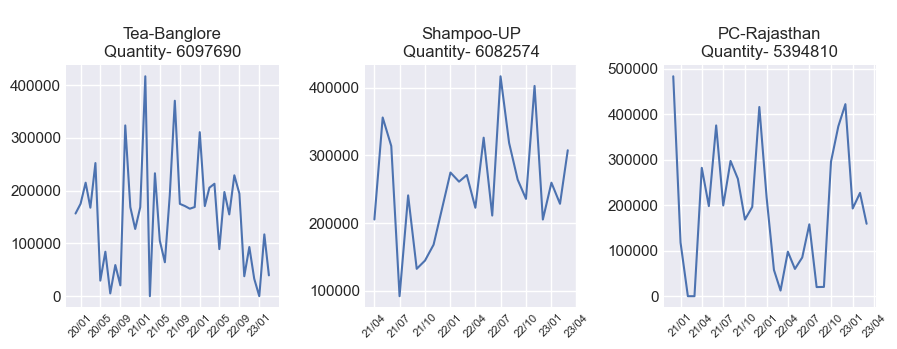




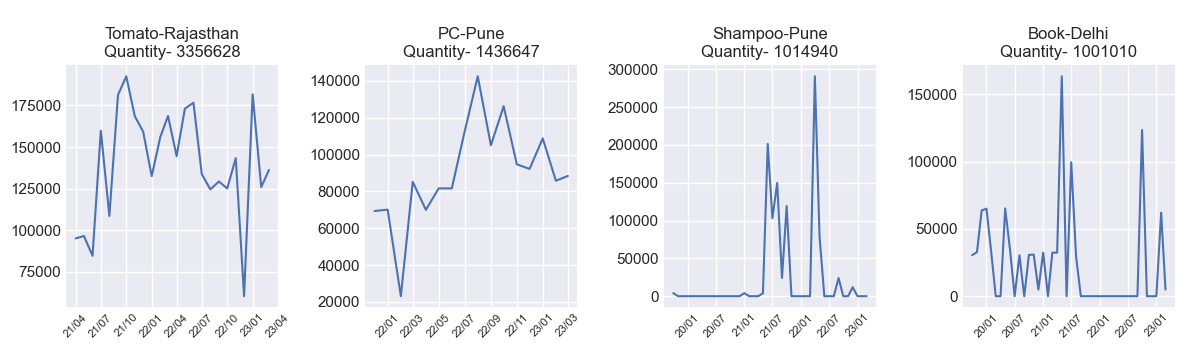
High Vol Data



Mid Vol Data



Low Vol Data

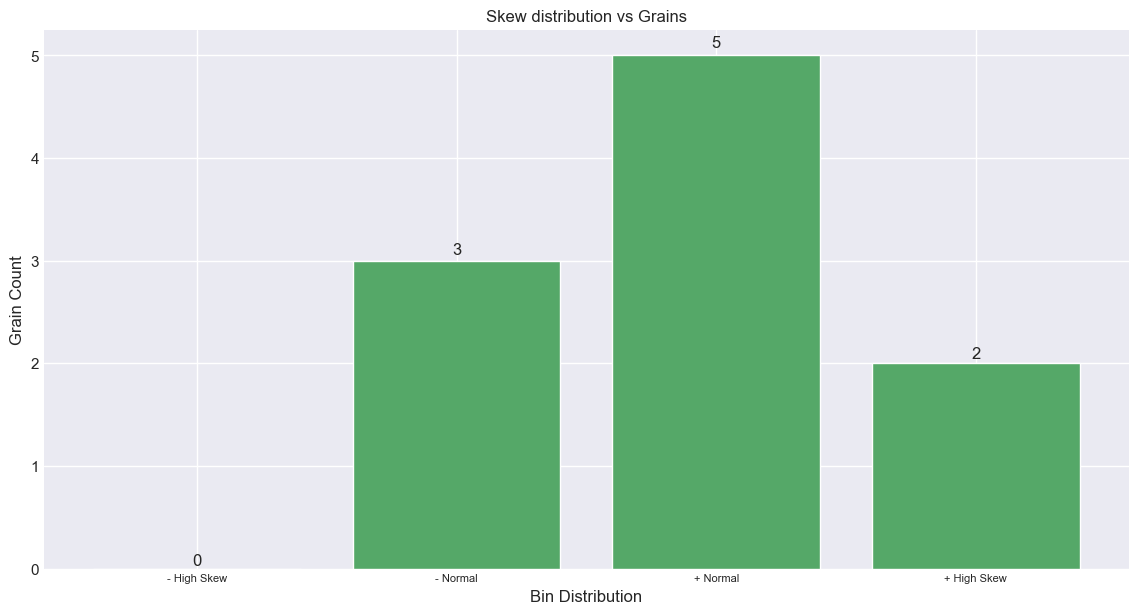


# 4.Sales Missing Values

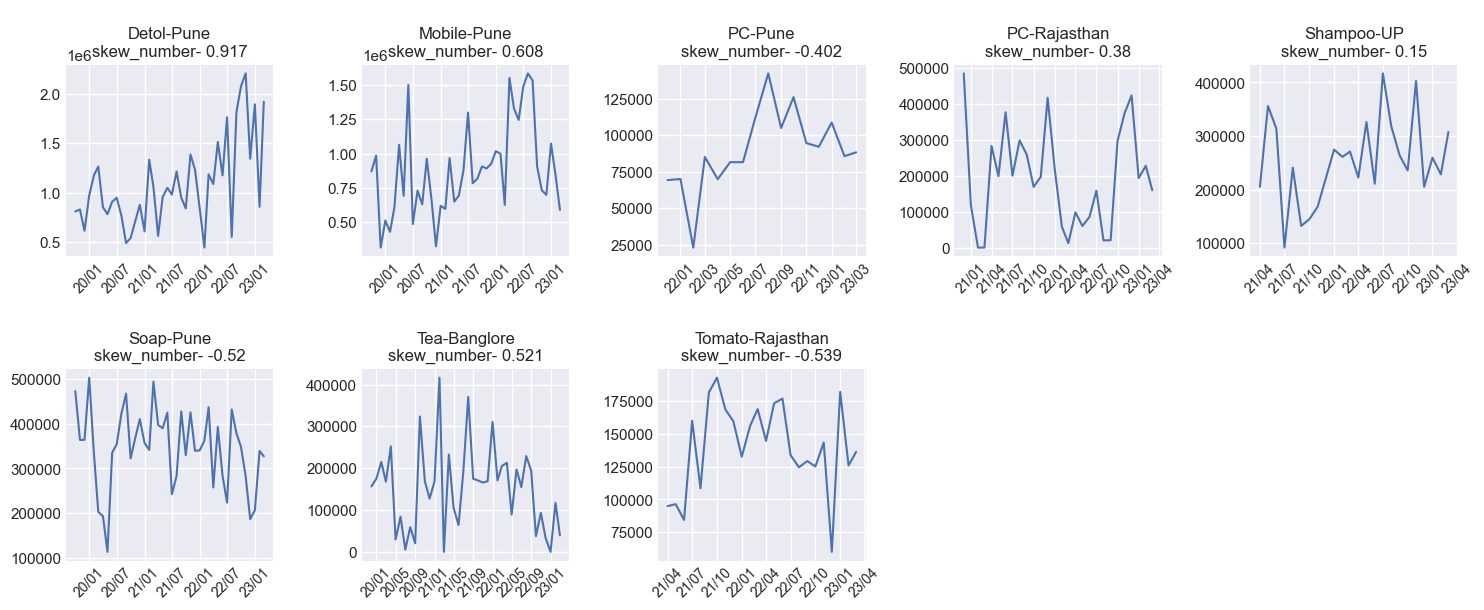
No missing values found

# 5.Skewness

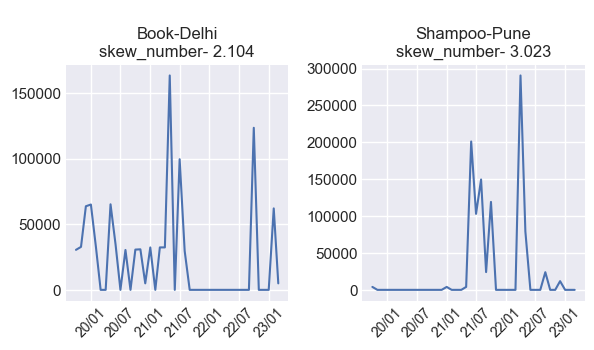
Ideal skewness is less than -1 or greater than 1, the distribution is highly skewed. If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.If skewness is between -0.5 and 0.5, the distribution is approximately symmetric.The Ideal values kurtosis between -2 and +2 are considered acceptable



No Skew/Normal Data



Right Skewed Data



Left Skewed Data

Dataframe is empty - left\_skewed\_data

None

# 5.Creating Time Series Feature

Day, Dayofyear, dayofweek, month, year, weekofyear, start date and end date













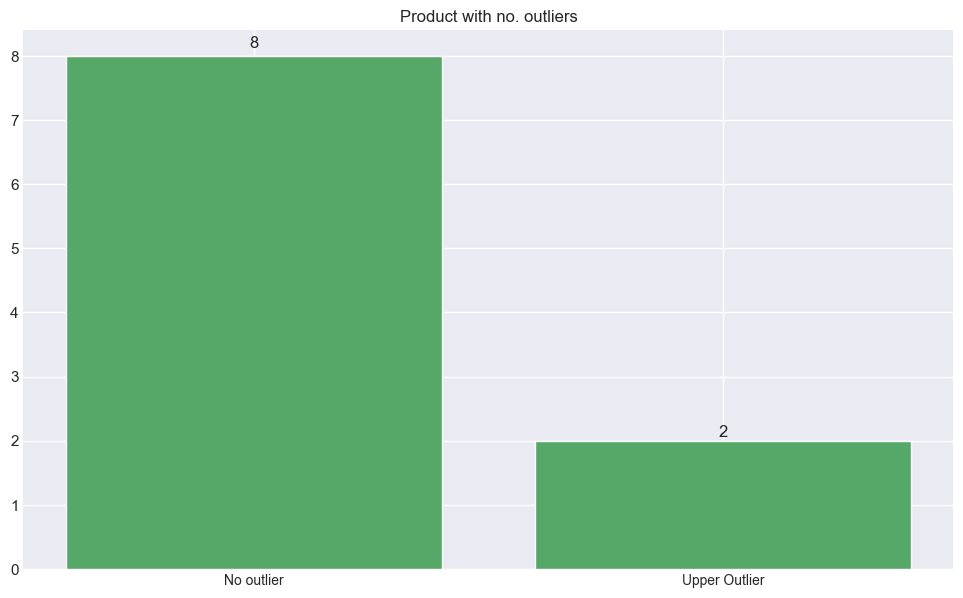


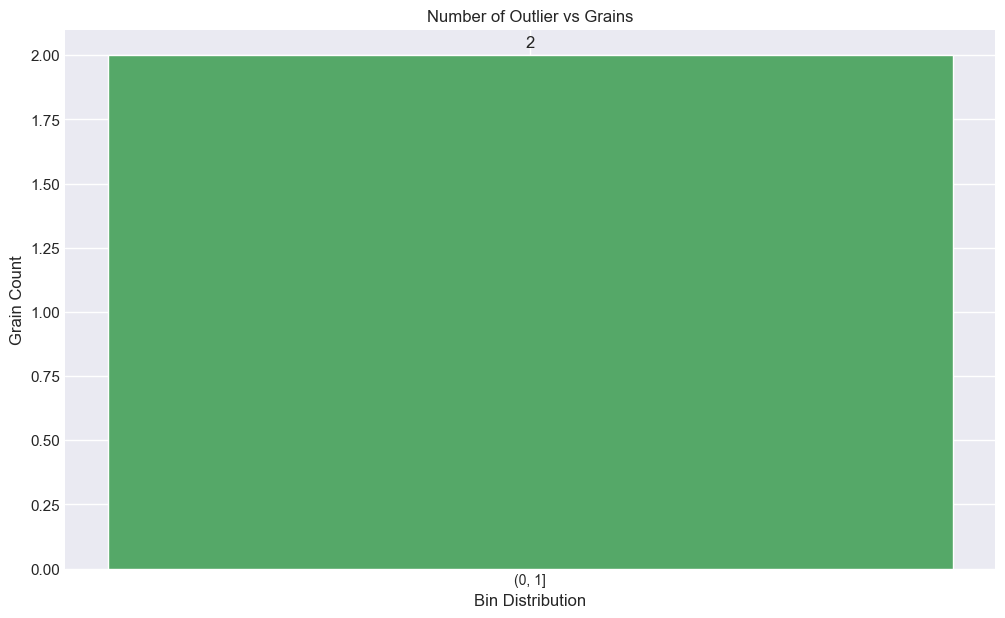
# 7.Outlier

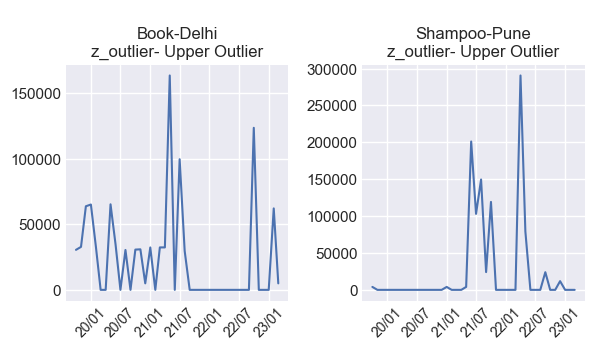
Product 2 has Zscore greater/less then 3, treated as outliers

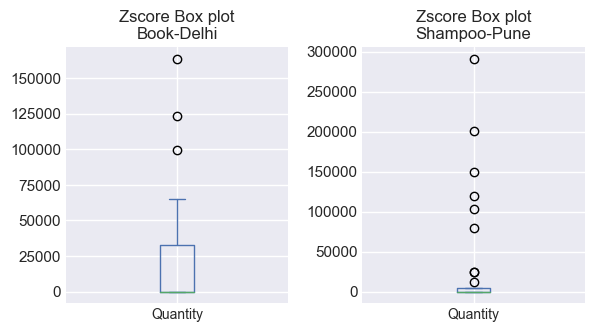
Min Zscore -2.65 is belong to Soap-Pune product

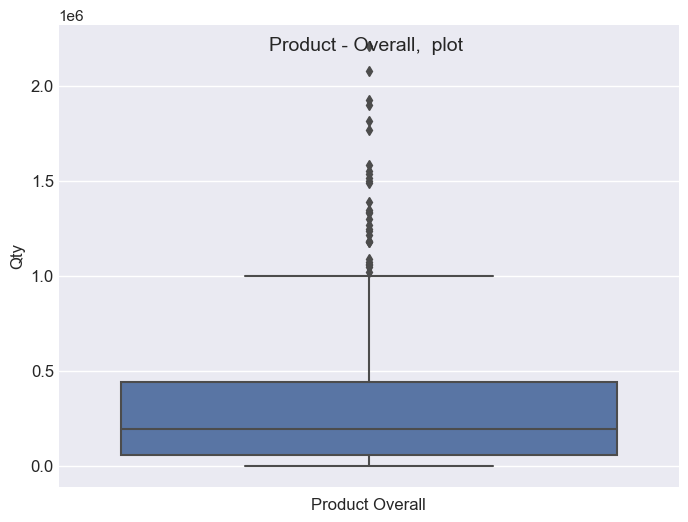
Max Zscore 4.38 is belong to Shampoo-Pune product

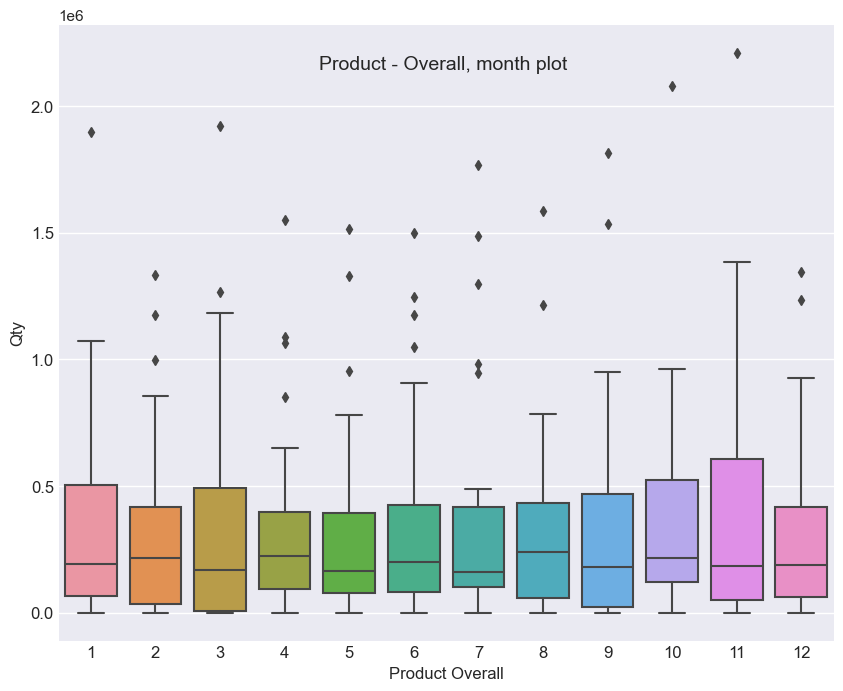


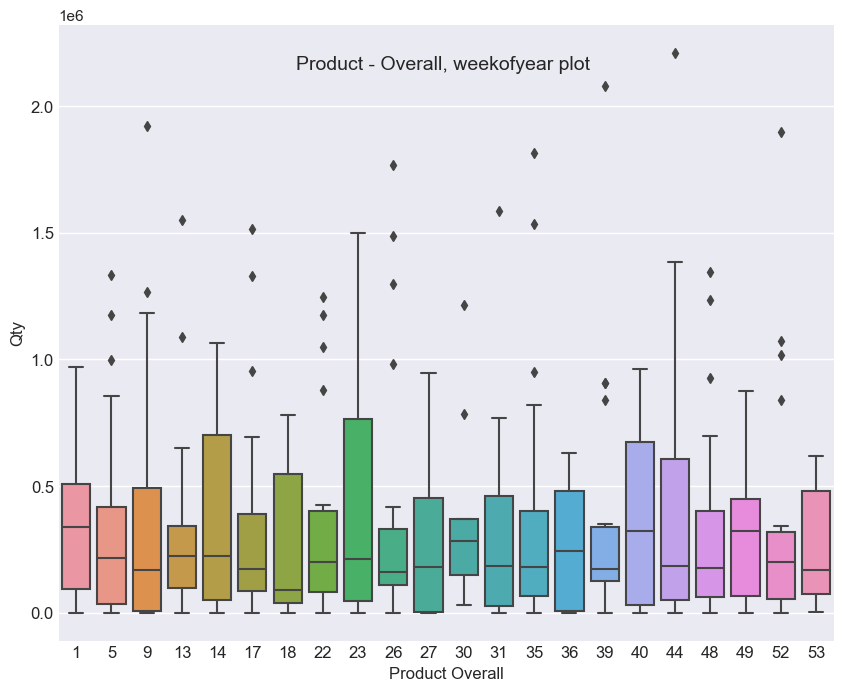


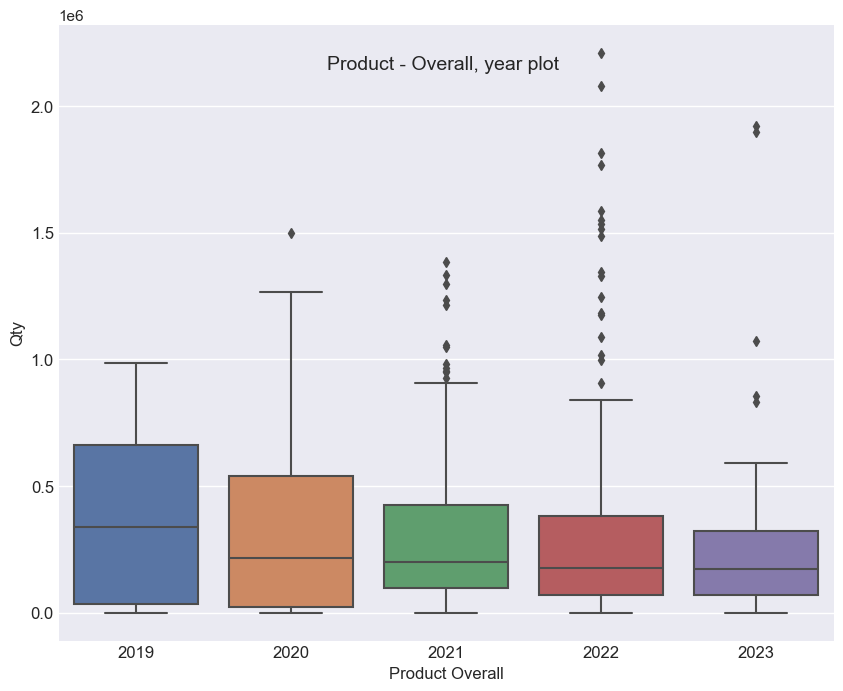










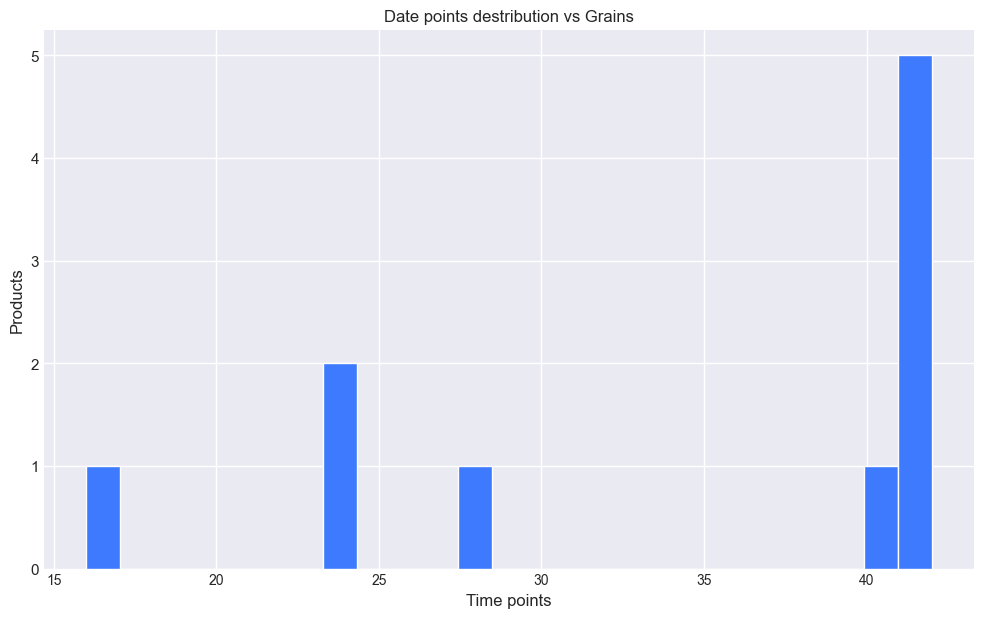


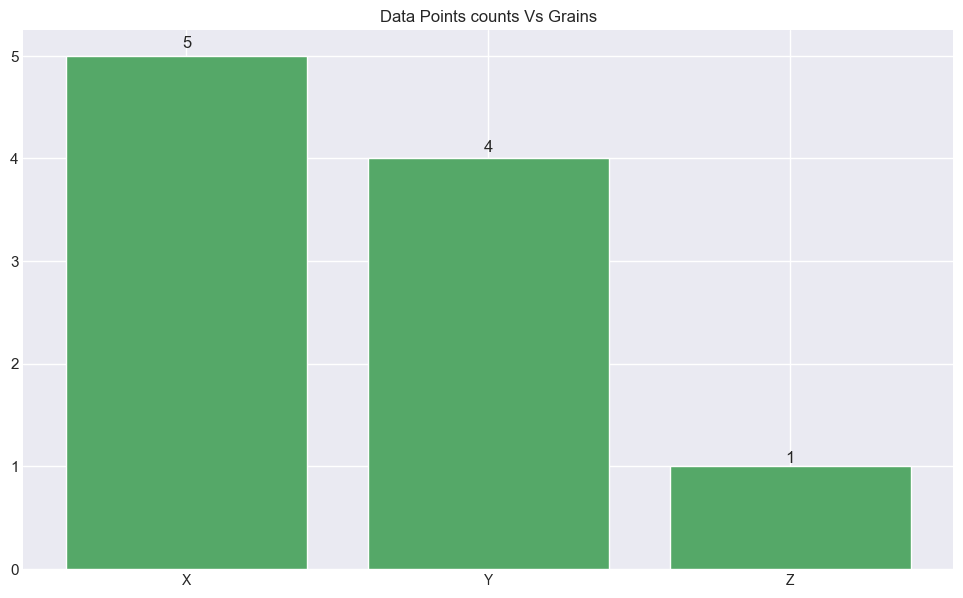
# 8.XYZ Segmentation

Total Products 10

\*\*\*\*\* Number of the Products (20, 40, >40)\*\*\*\*\*

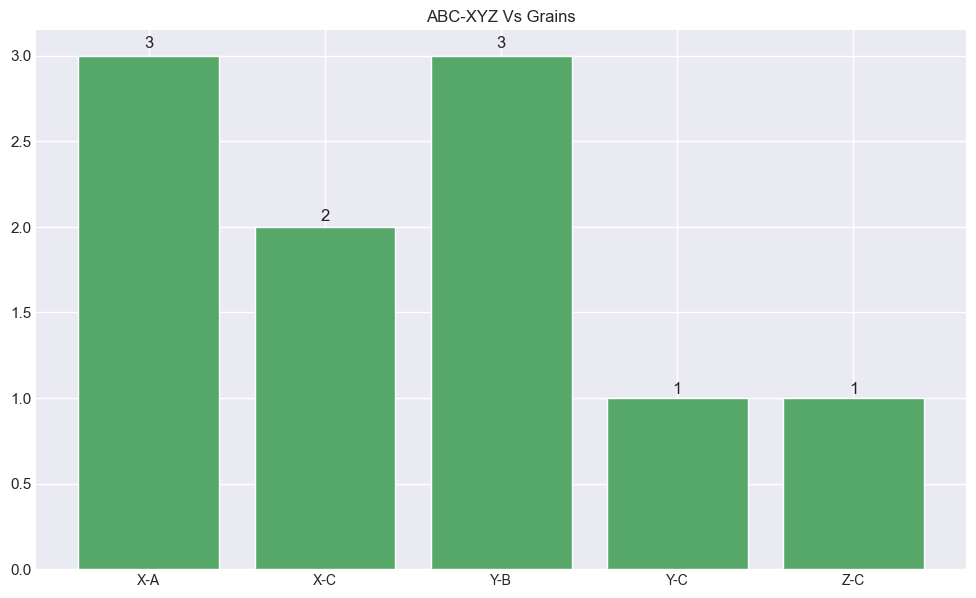
X 5  
Y 4  
Z 1  
Name: data\_points\_flag, dtype: int64





# 9.ABC-XYZ Segmentation

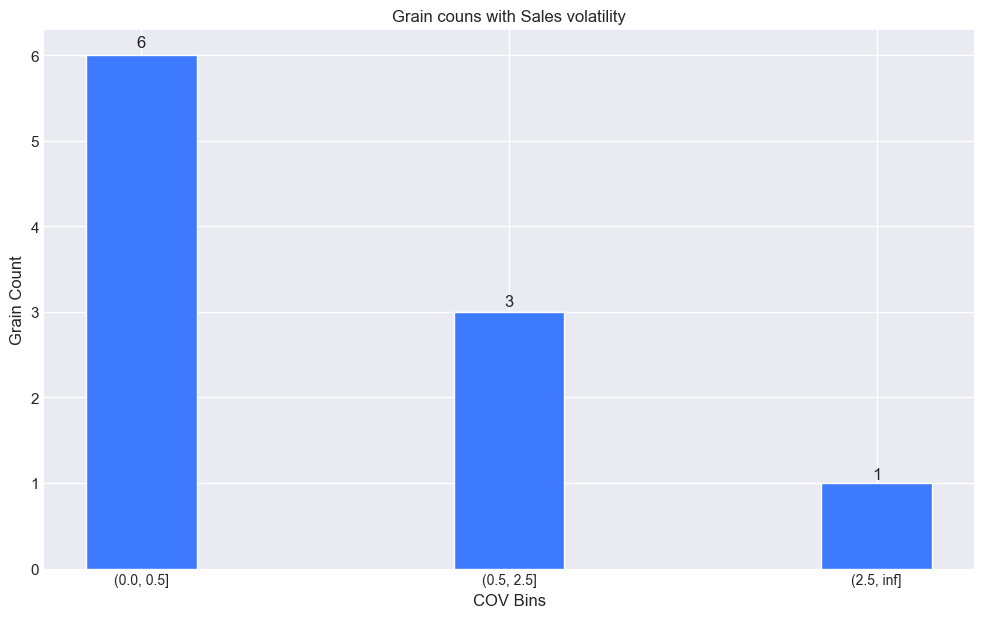
X-A 3  
Y-B 3  
X-C 2  
Z-C 1  
Y-C 1  
Name: abc-xyz, dtype: int64



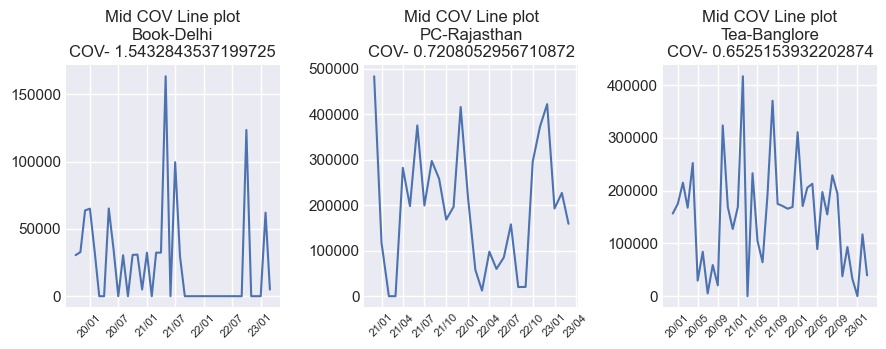
# 10.COV

Max COV of sales is 2.55

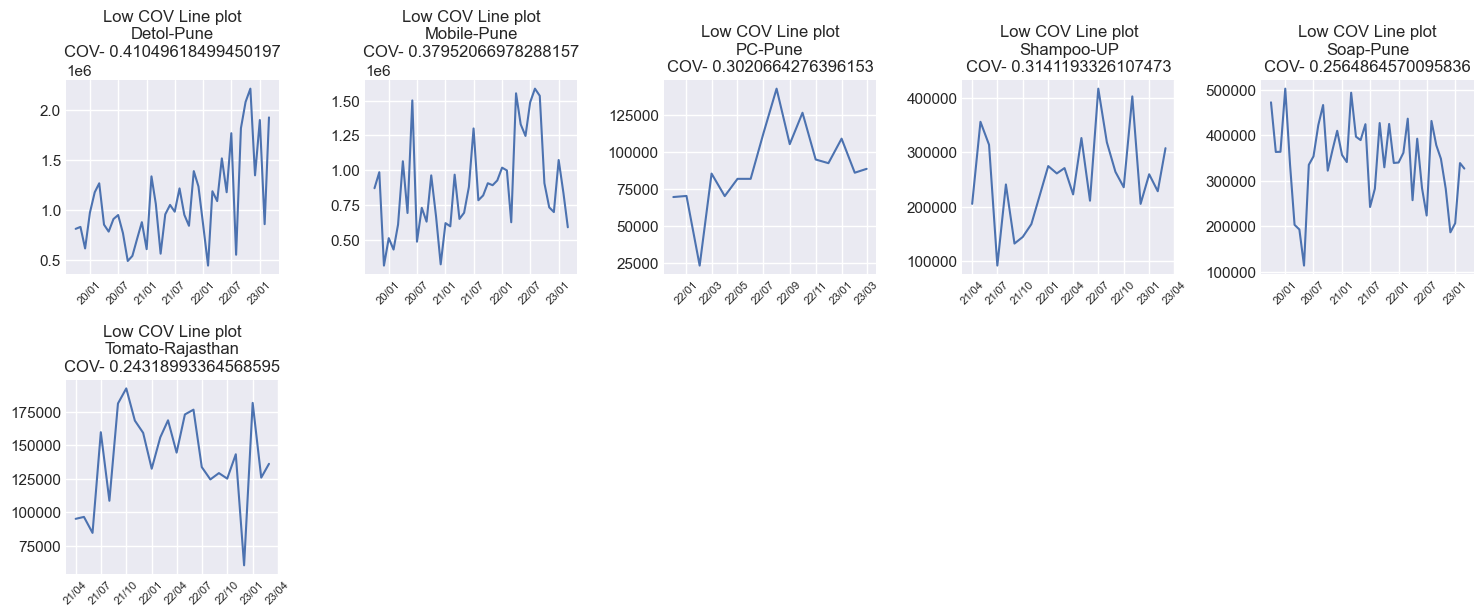
Min COV of sales is 0.24



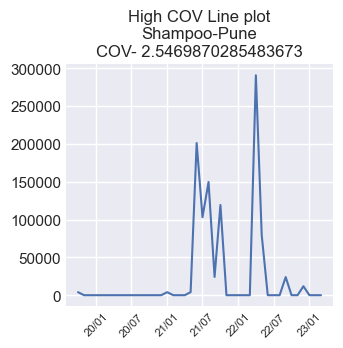
Mid COV Plots



Low COV Plots



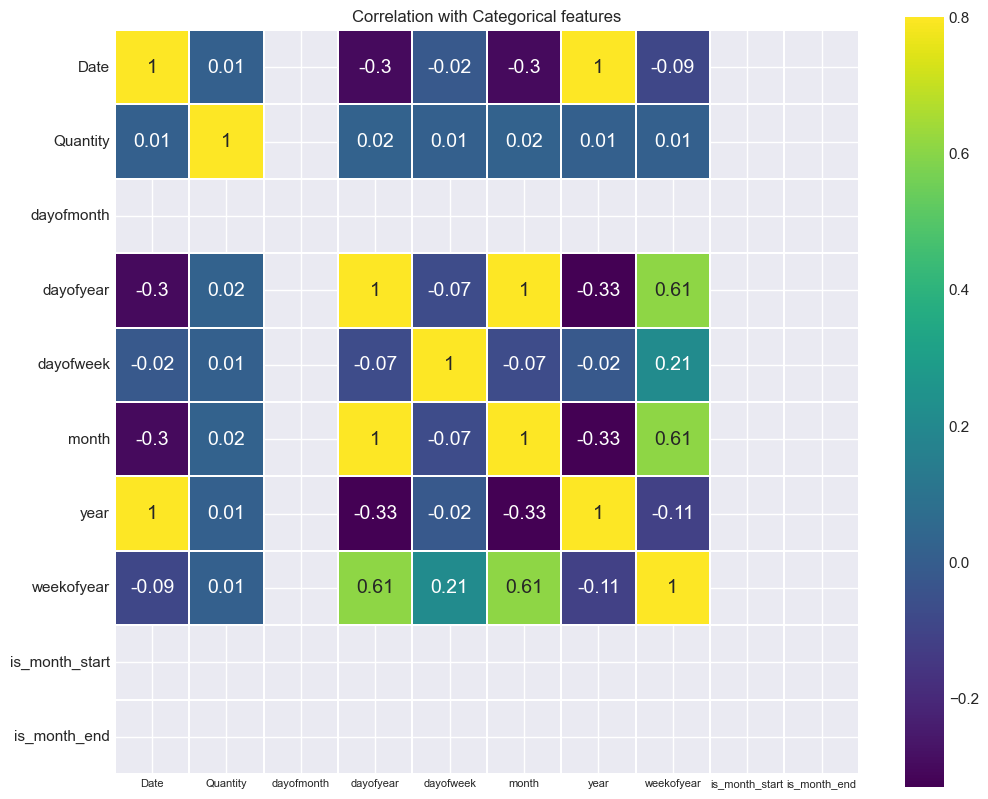
High COV Plots



# 11.Features Correlation

Numerical features are ['Date', 'Quantity', 'dayofmonth', 'dayofyear', 'dayofweek', 'month', 'year', 'weekofyear', 'is\_month\_start', 'is\_month\_end'] and shape of the data (342, 10)

----- No Categorical features ----

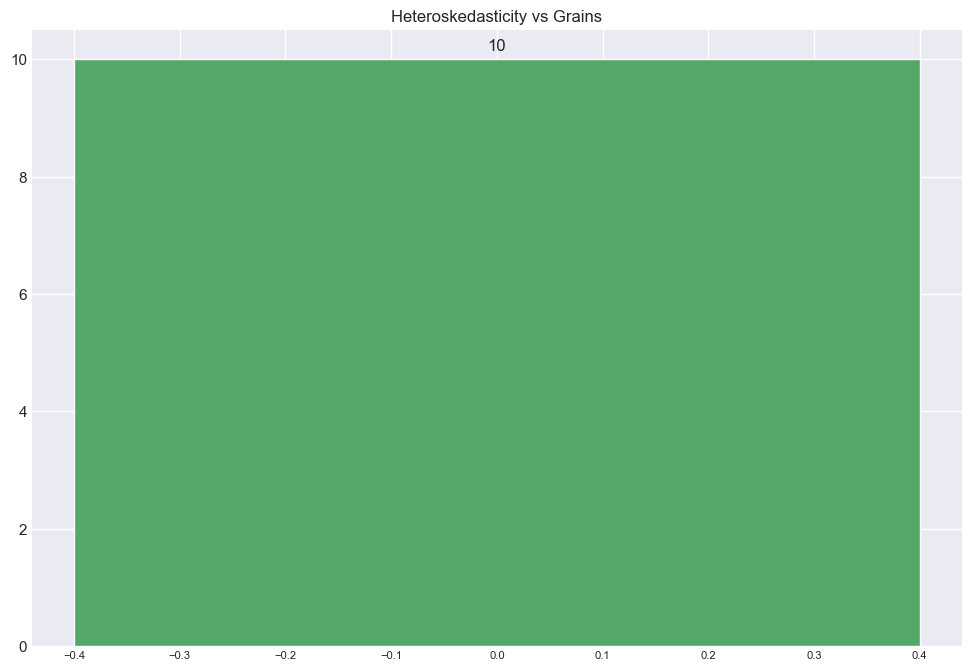


# 12.Negative Sales Check

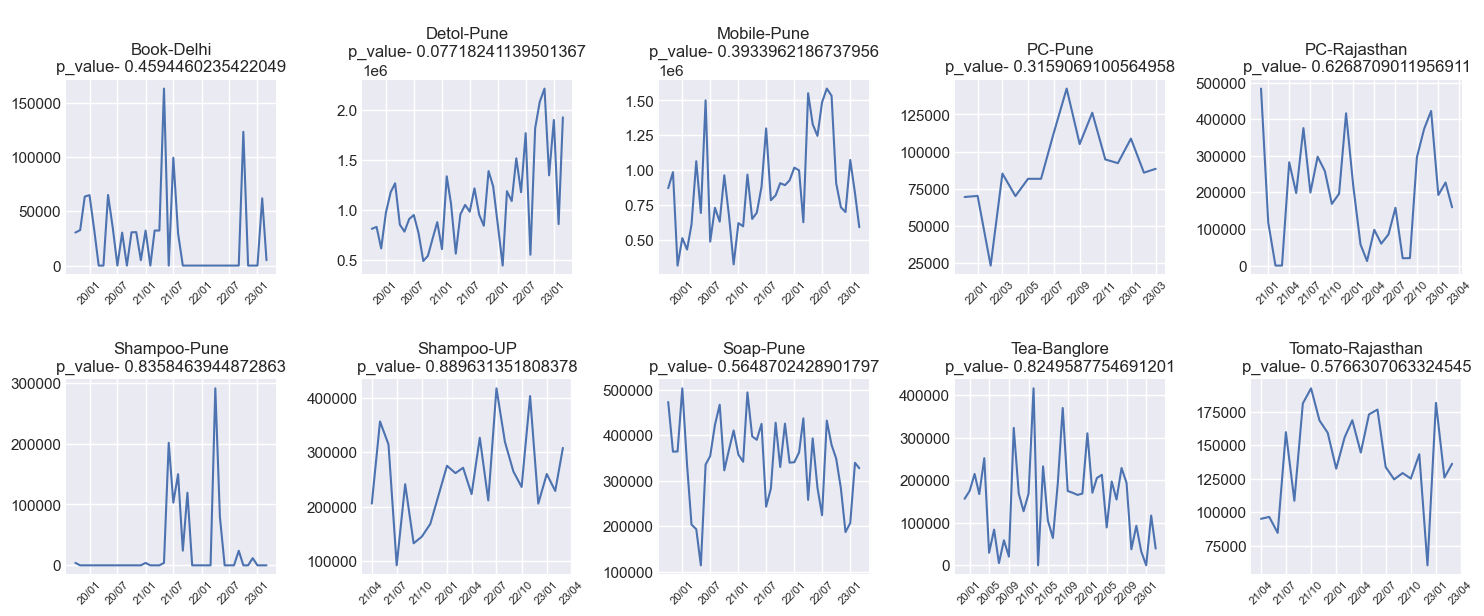
No Negative values in sales

# 13.Heteroscedasticity

17.--> Heteroscedasticity   
  
 Heteroscedasticity refers to the unequal scatter of residuals at differentlevels of a response variable, which violates the assumption that the residuals are equally scattered at each level of the response variable.  
 Heteroscedasticity is mainly due to the presence of outlier in the data. Outlier in Heteroscedasticity means that the observations that are either small or large with respect to the other observations are present in the sample. Heteroscedasticity is also caused due to omission of variables from the model.



False Heteroskedasticity max 10 grains



True Heteroskedasticity Plots

Dataframe is empty - true\_hdcty\_data

# 14.Leading Zeros

No grain found which date start from zero sales

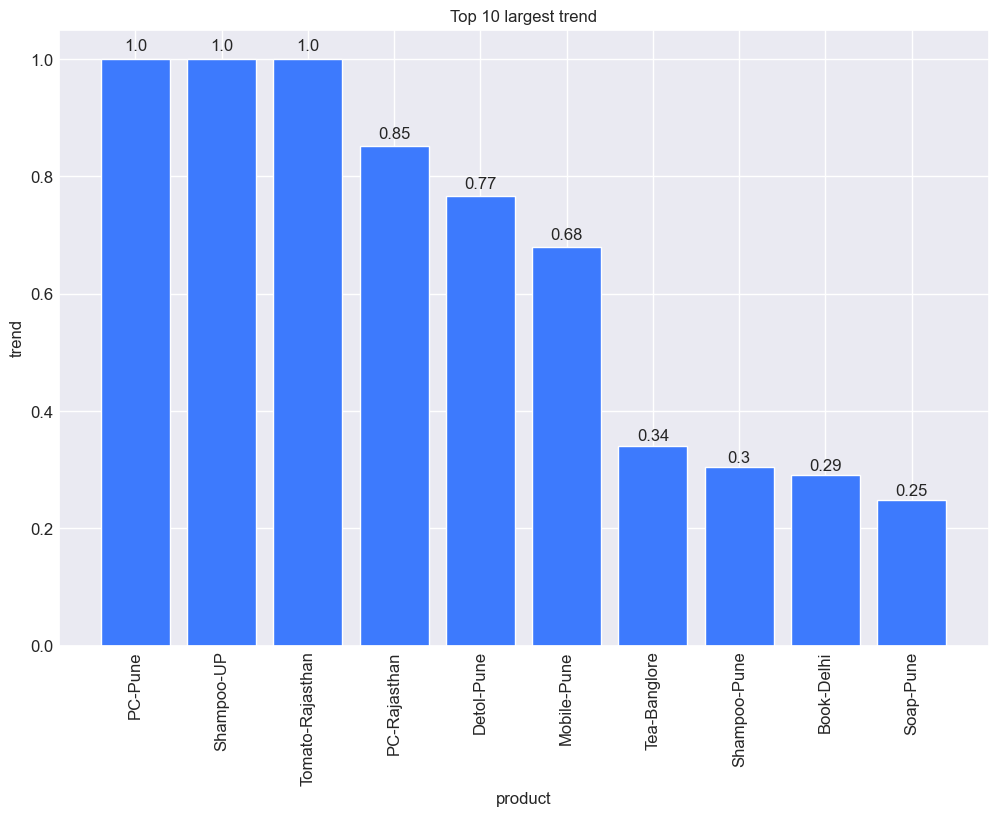
# 15.Trend and Seasonlity

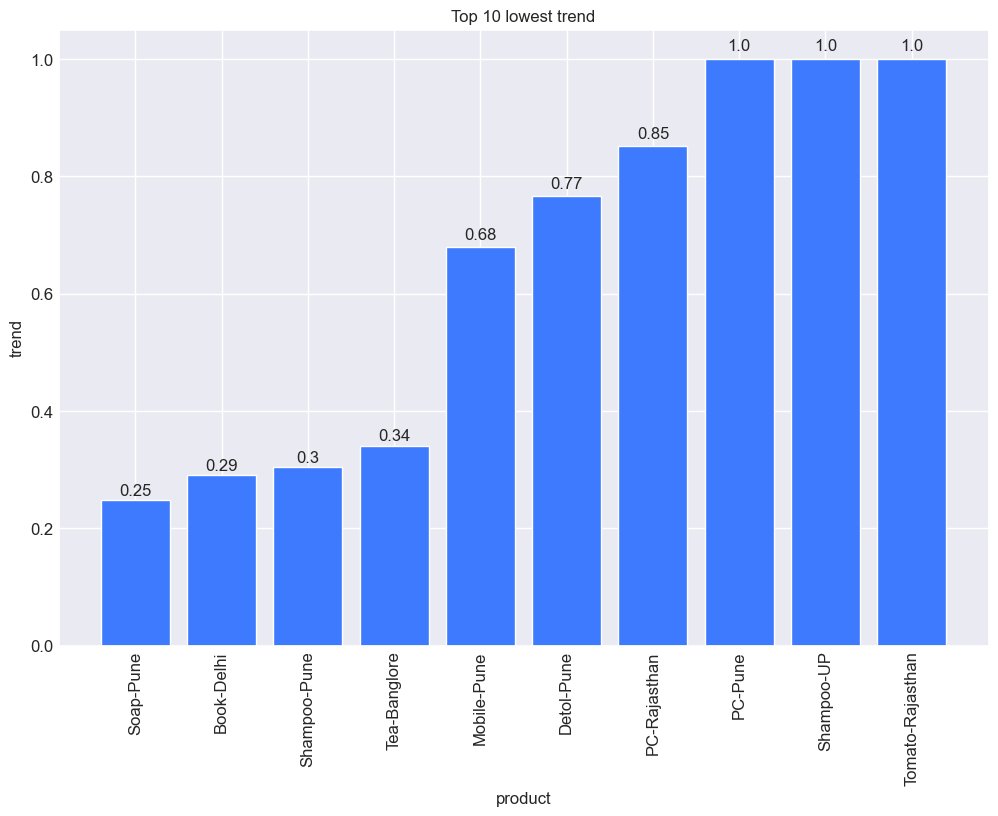
Data types for tsfeatures columns

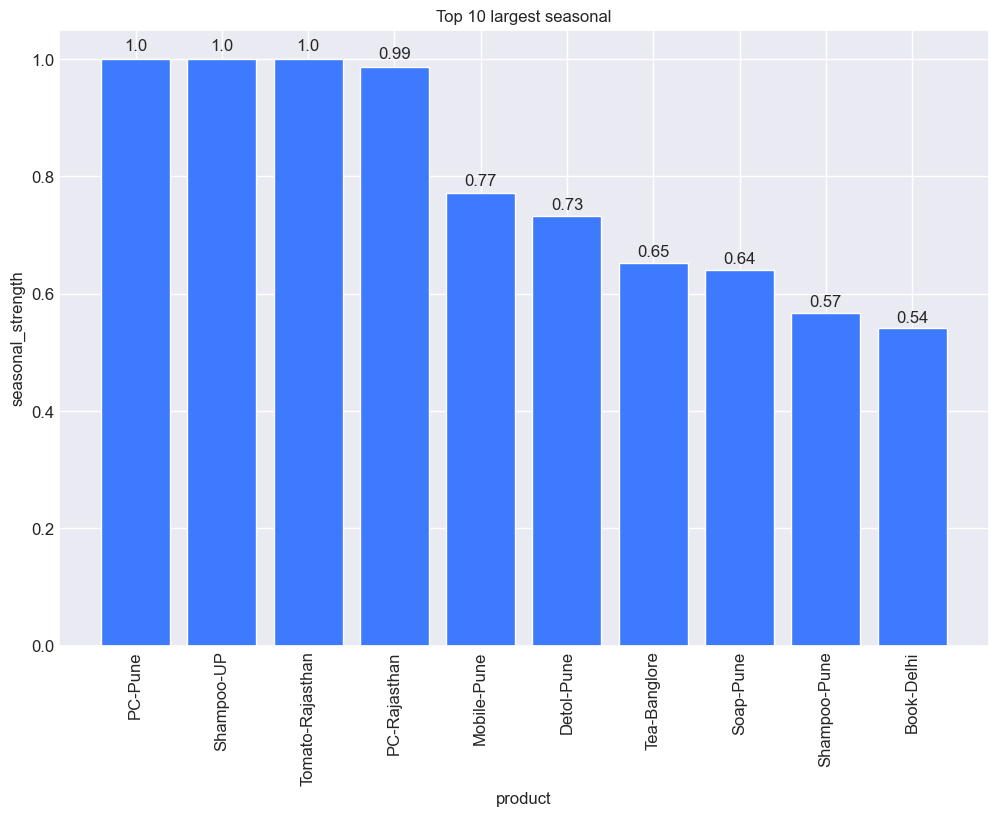
ds datetime64[ns]  
y float64  
unique\_id |S16  
dtype: object

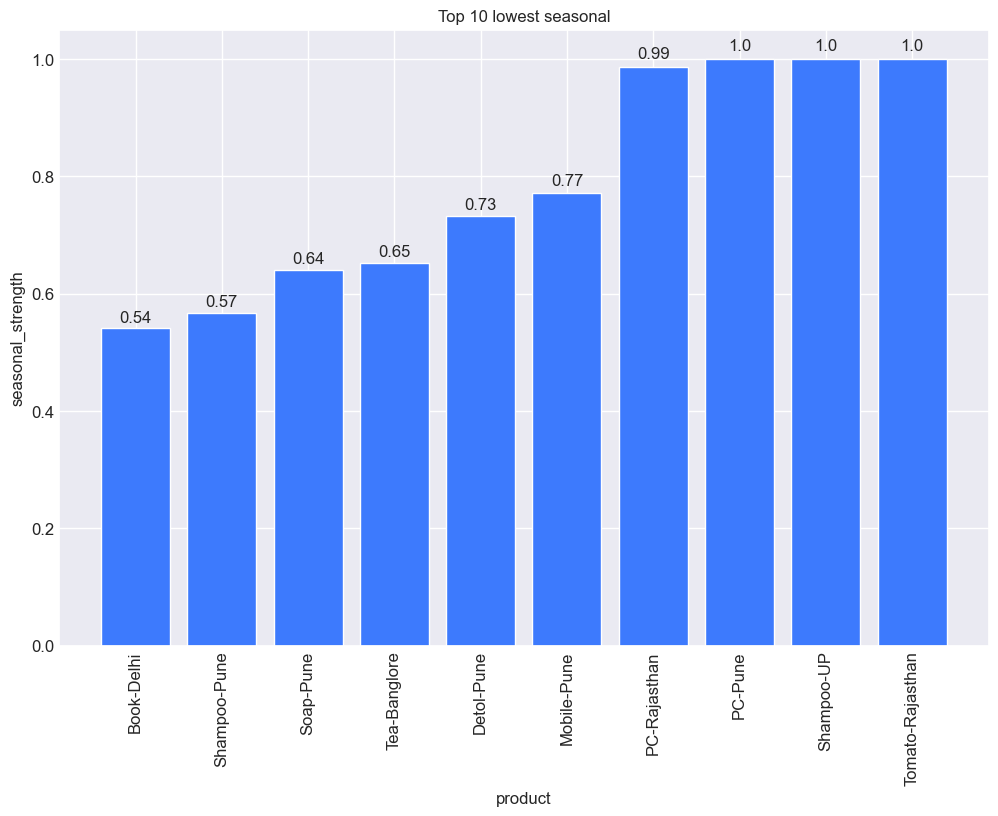
Grains above 0.75 seasonale point are 5

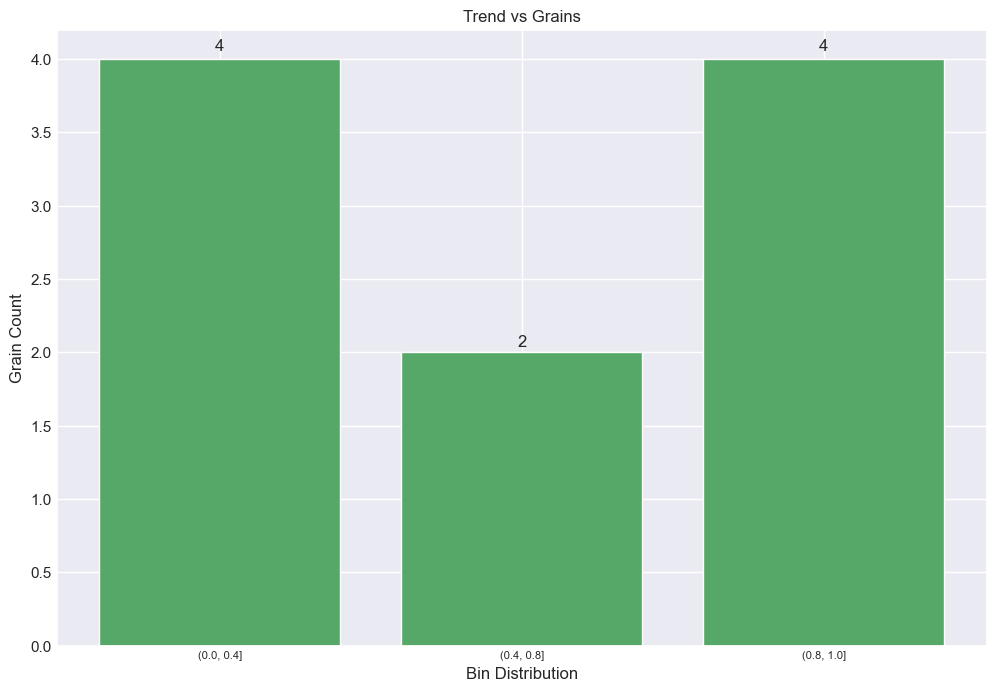
Grains above 0.4 trend point are 6

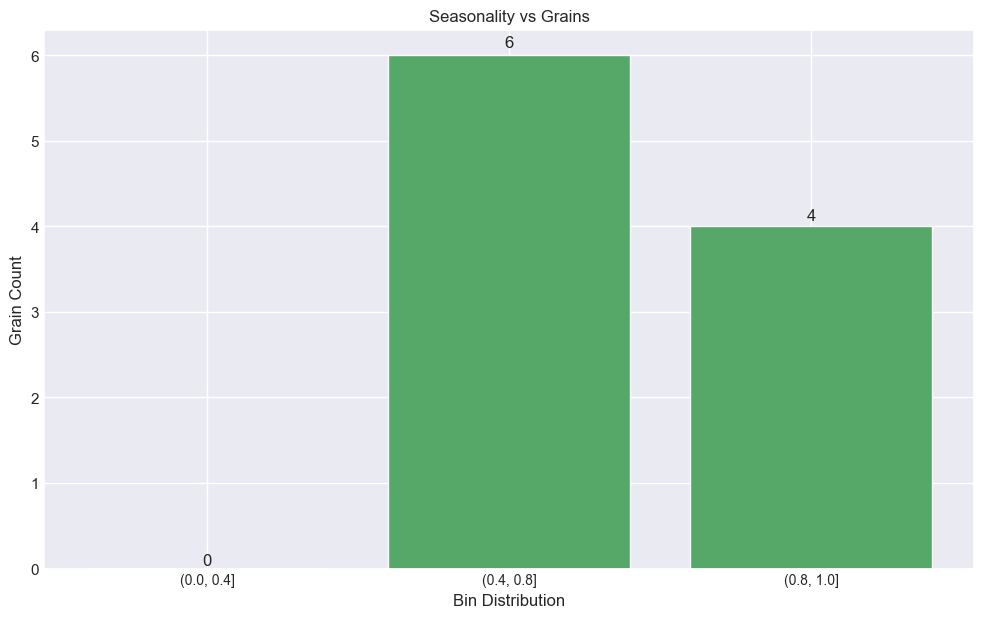






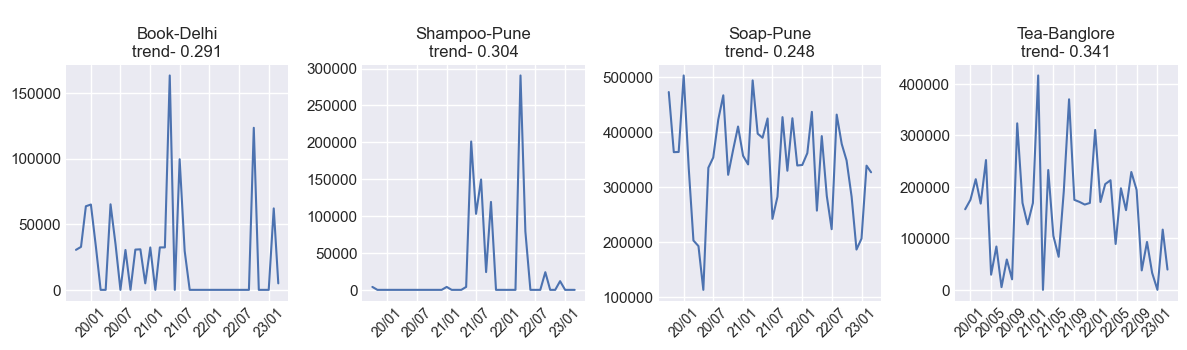




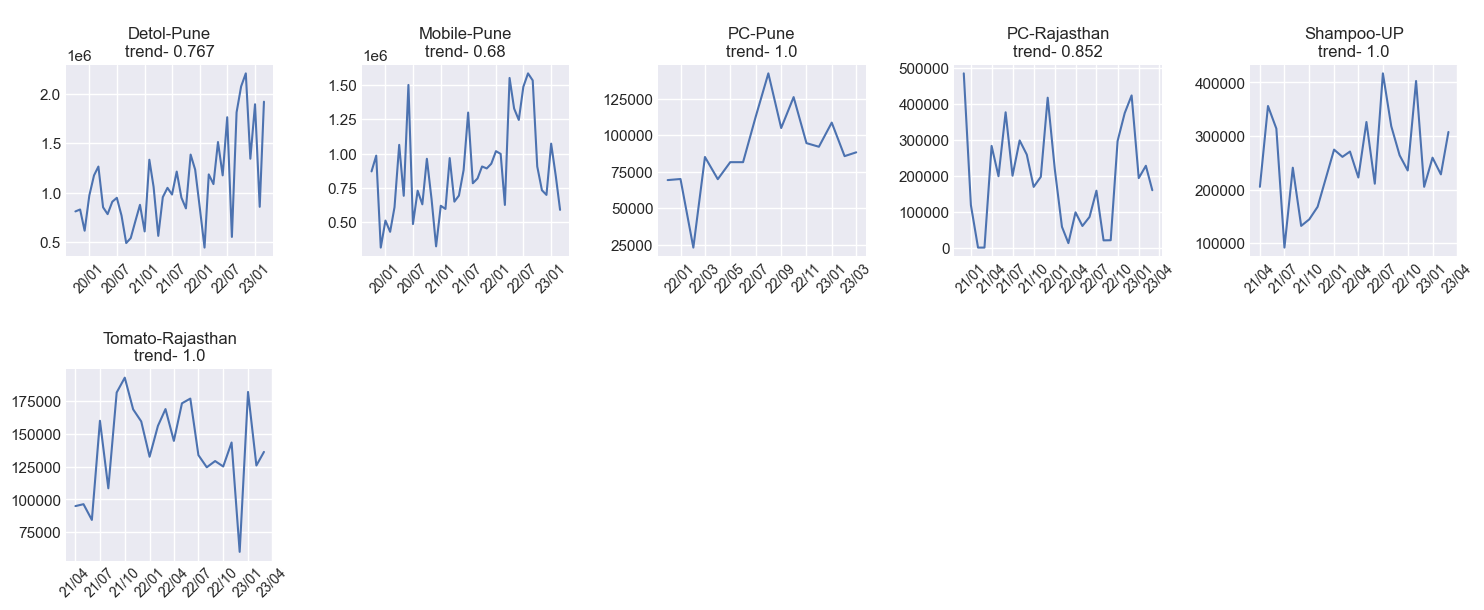


Trend line plots

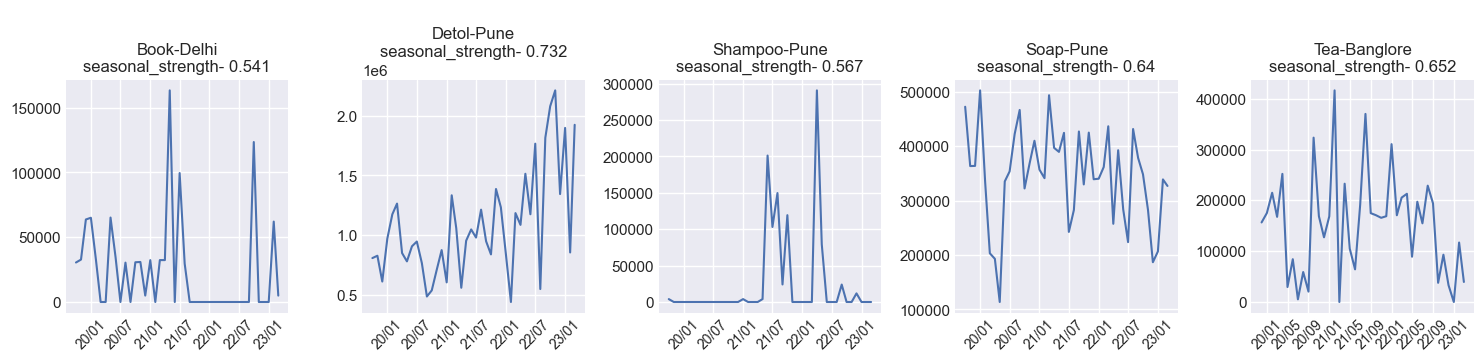
<--Low Trend-->



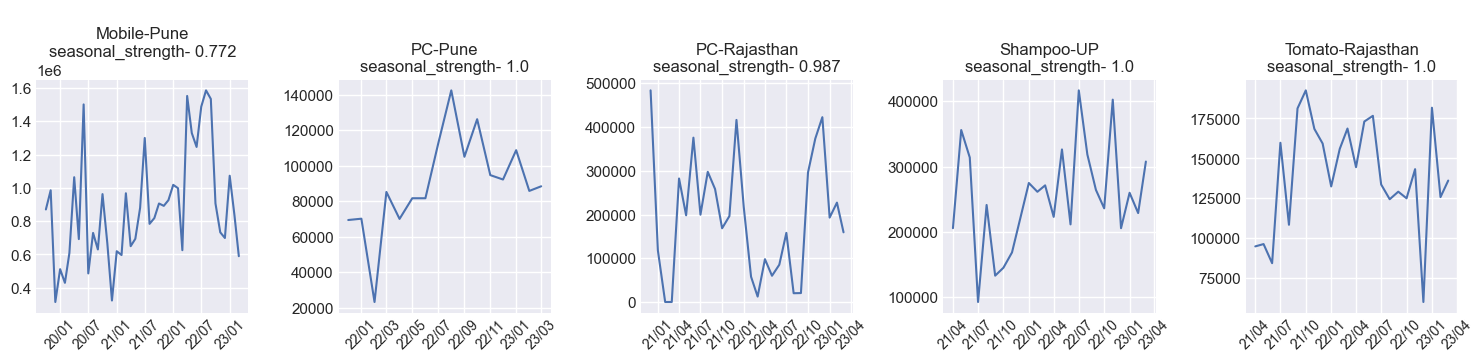
<--High Trend-->



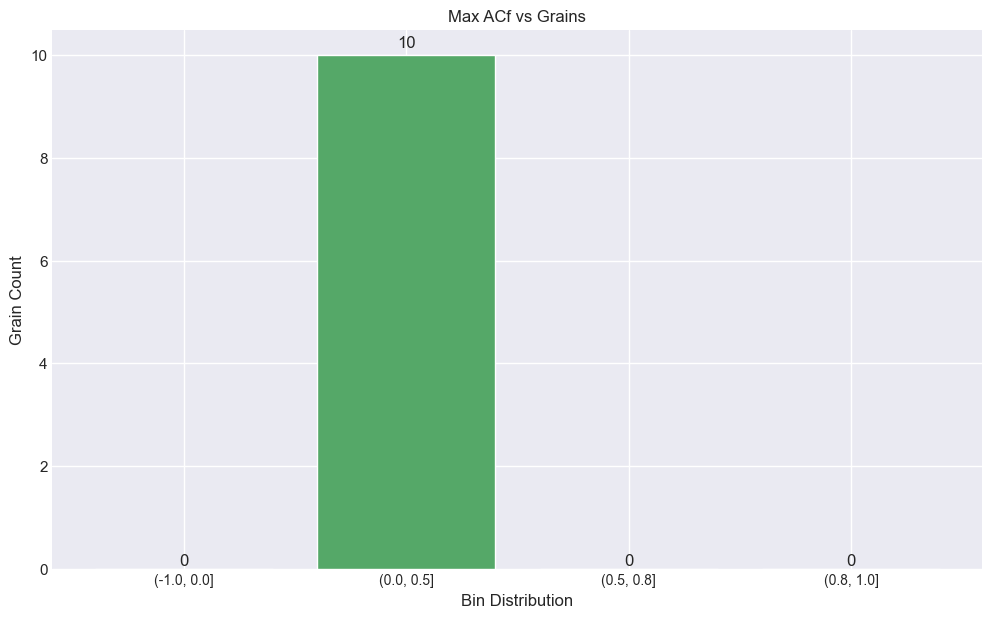
<--Low Seasonality-->



<--High Seasonality-->

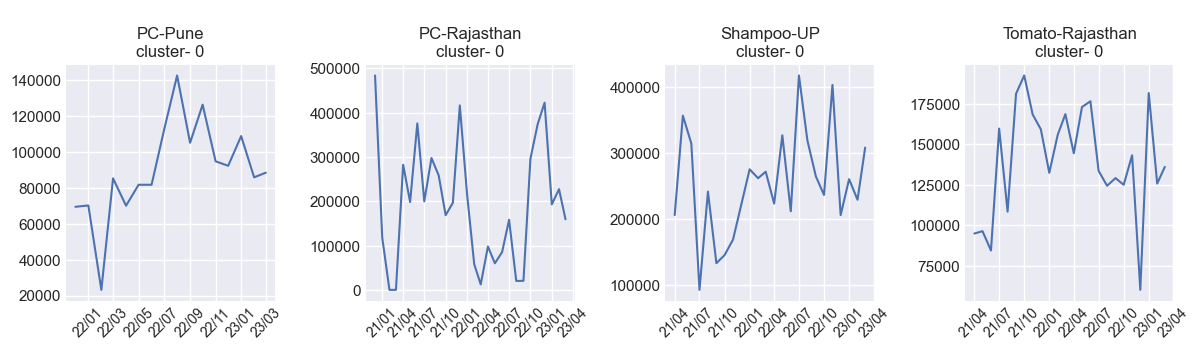


# 16.ACF and PACF Feature

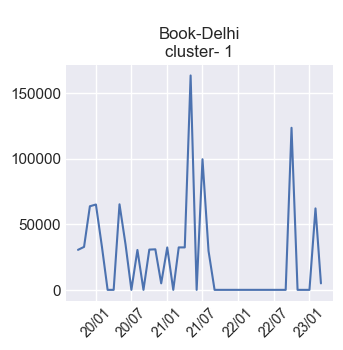


# 17.Clustering and Pattern similarity

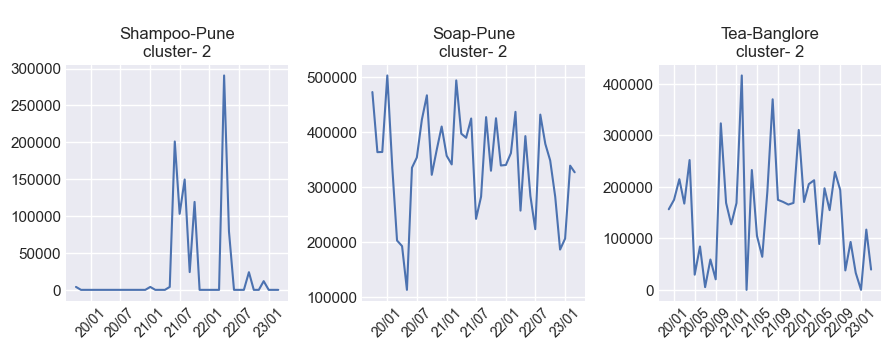
Cluster-0



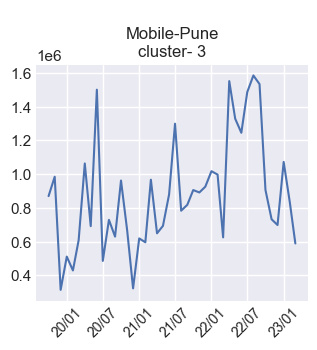
Cluster-1



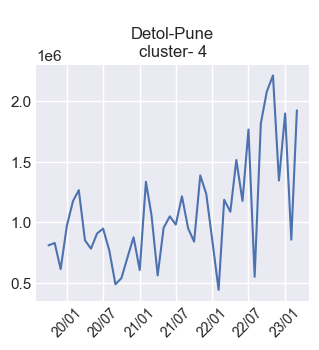
Cluster-2



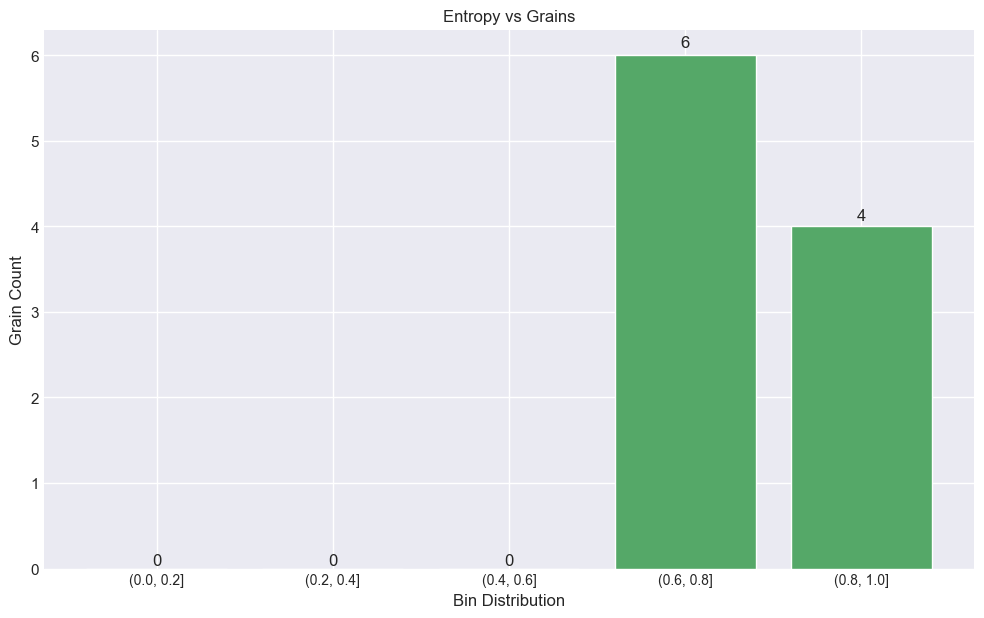
Cluster-3



Cluster-4



# 18.Entropy

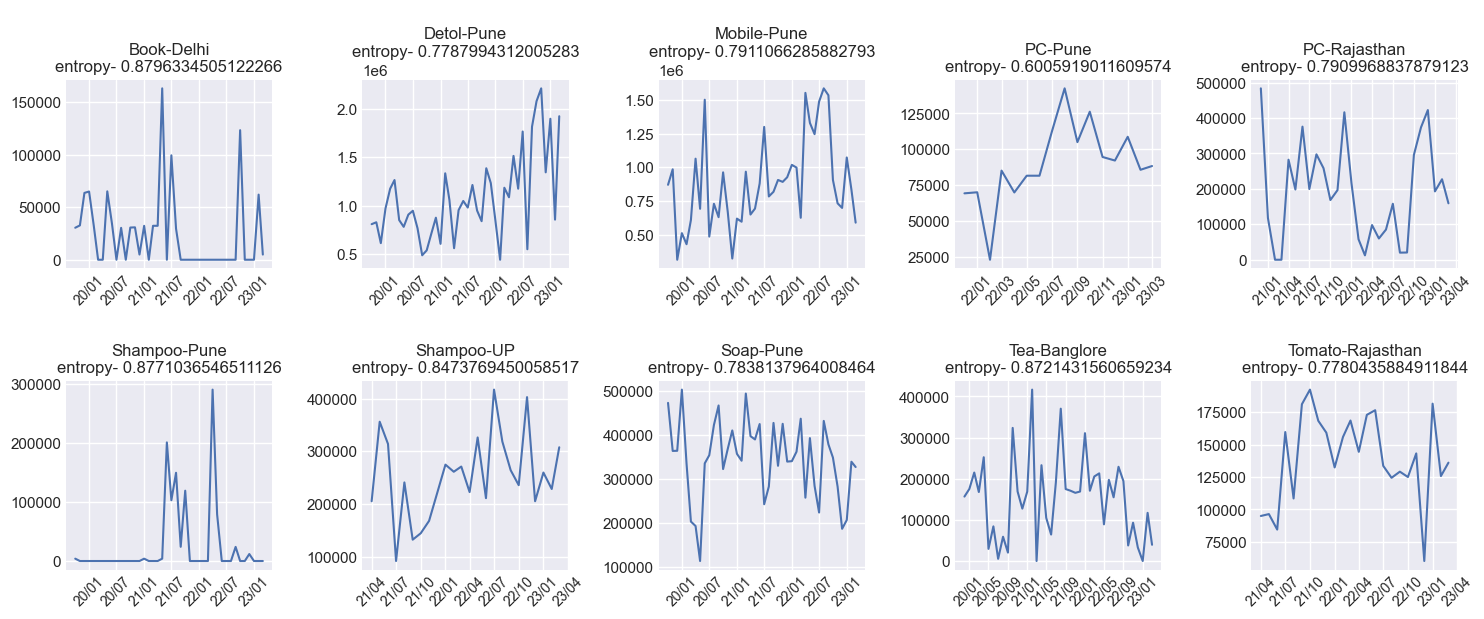


<----Low Entropy Plots--->

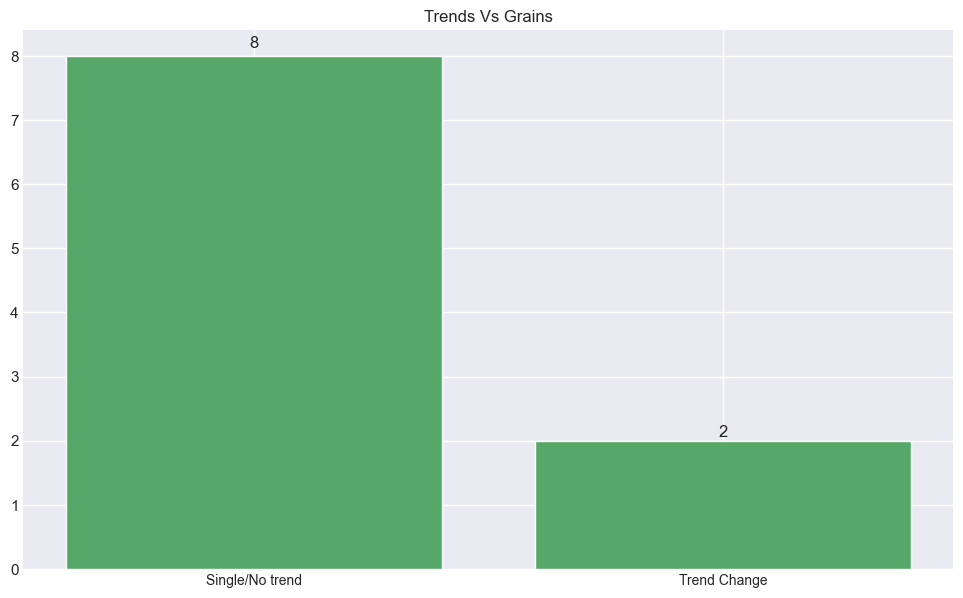
Dataframe is empty - low\_entropy\_data

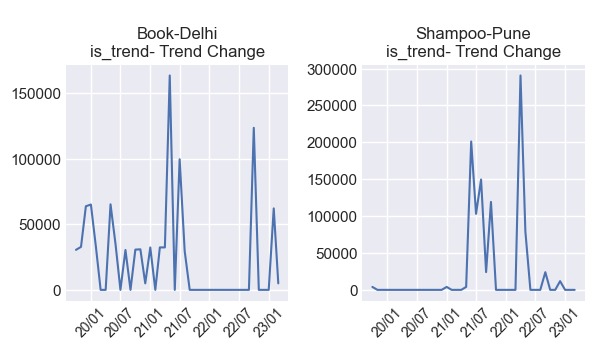
None

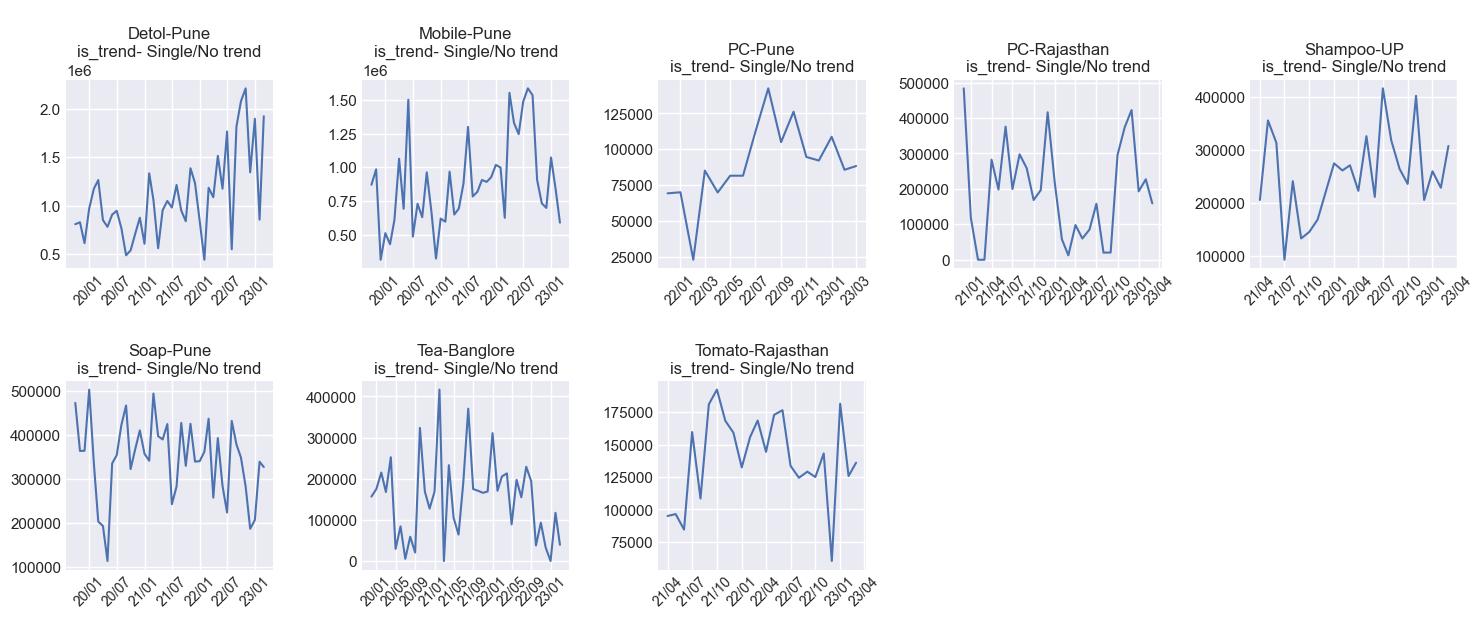
<----High Entropy Plots---->



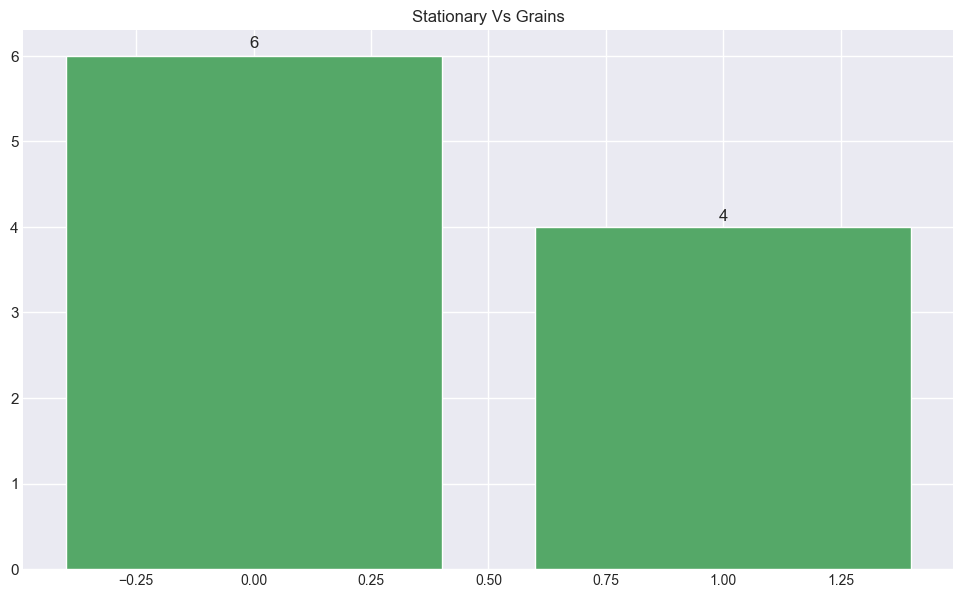
# 19.Trend Line

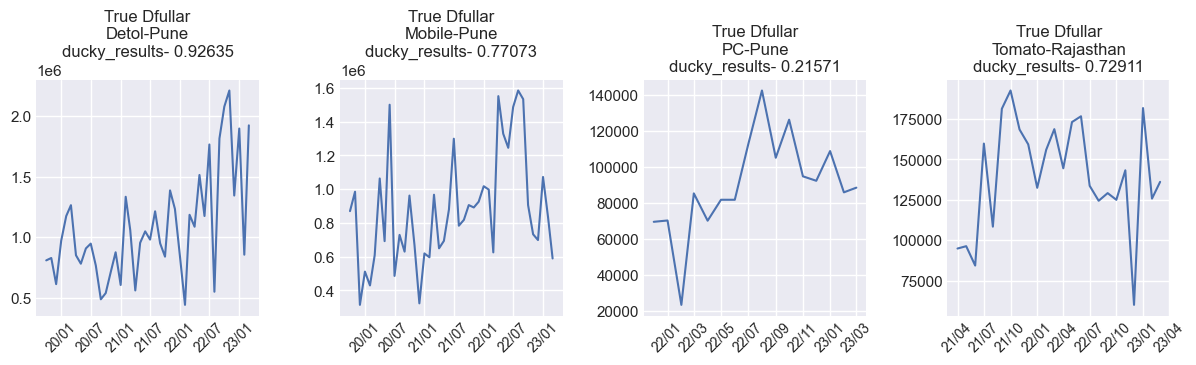


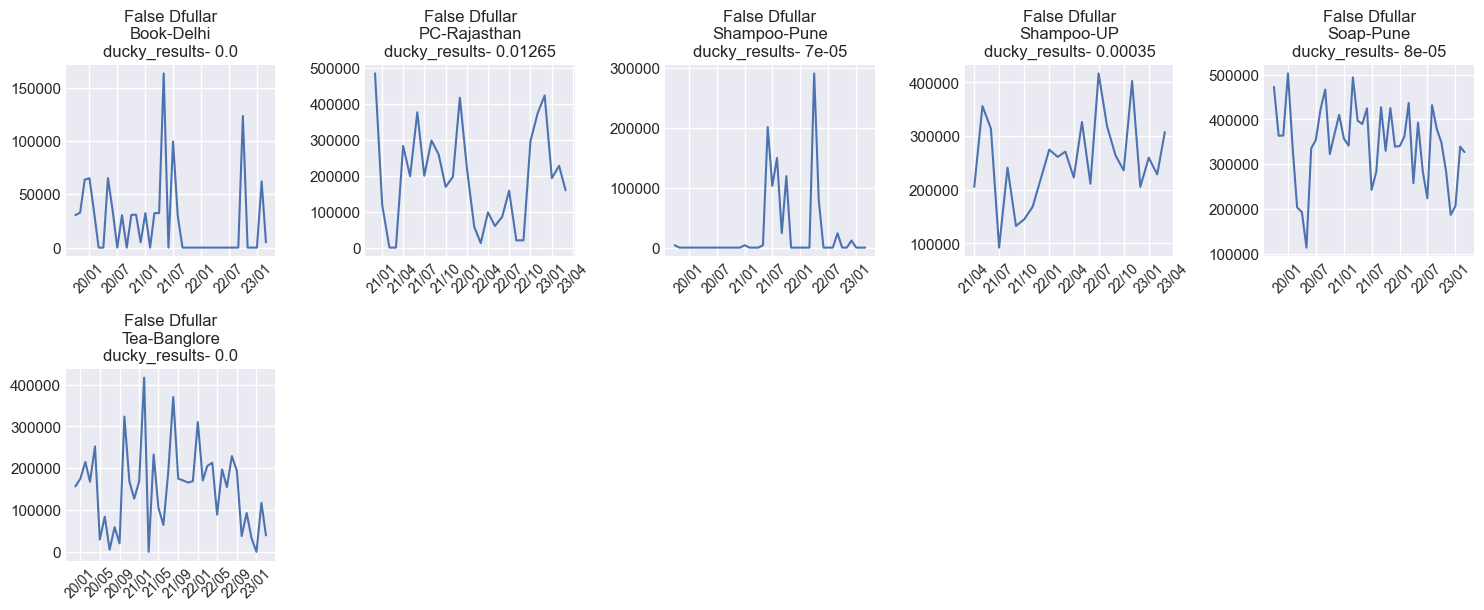




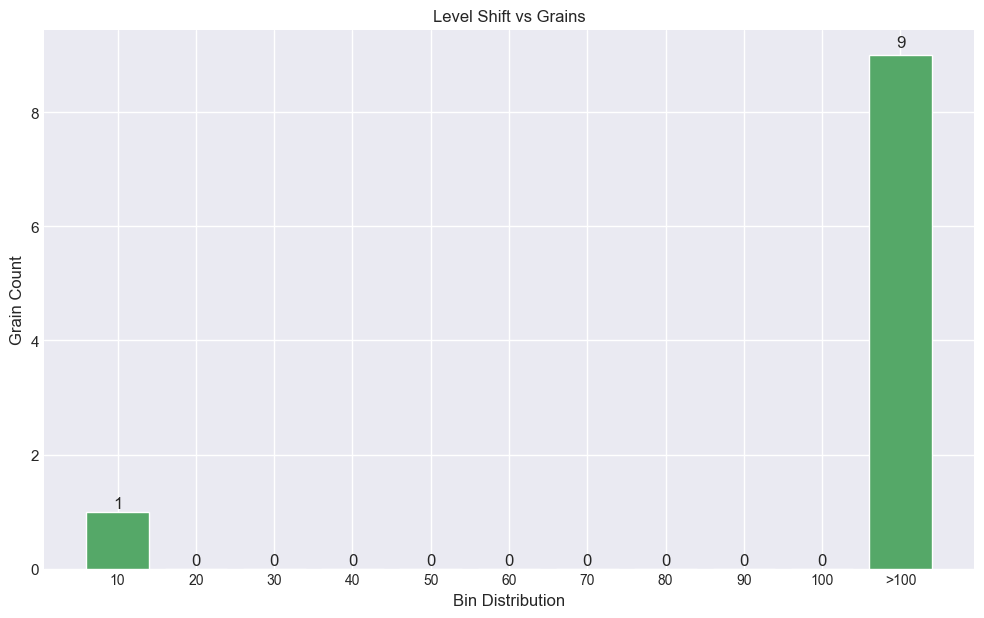
# 20.Stationary Test(Ducky Fullar Test)



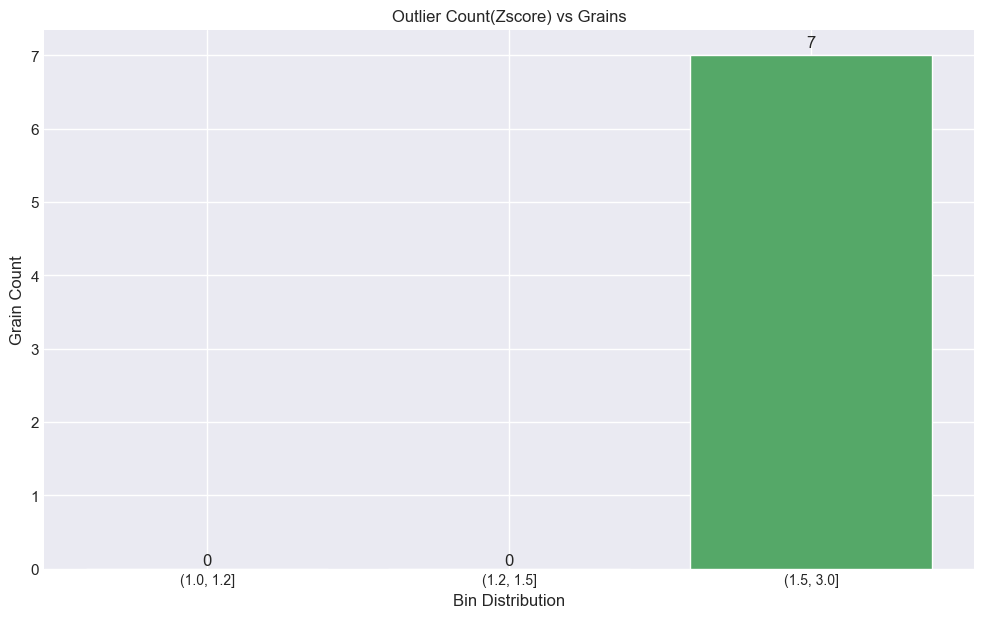


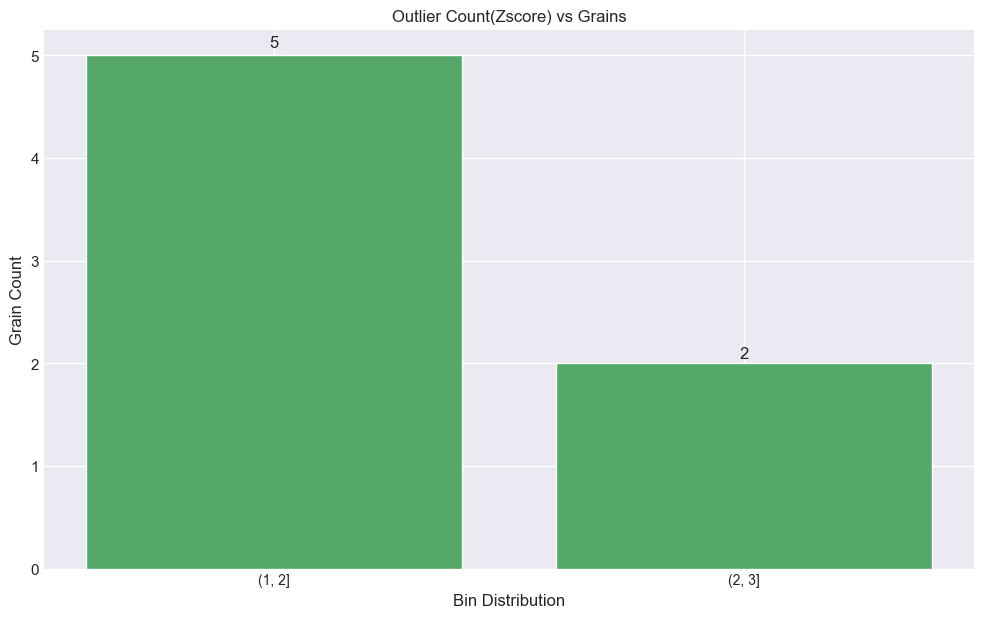


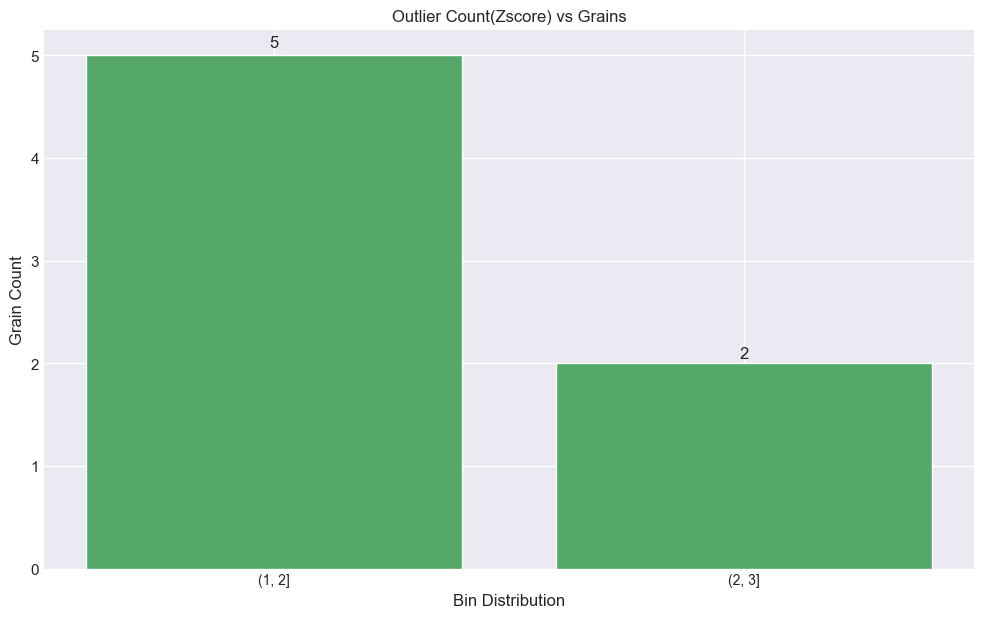
# 21.Level Shift

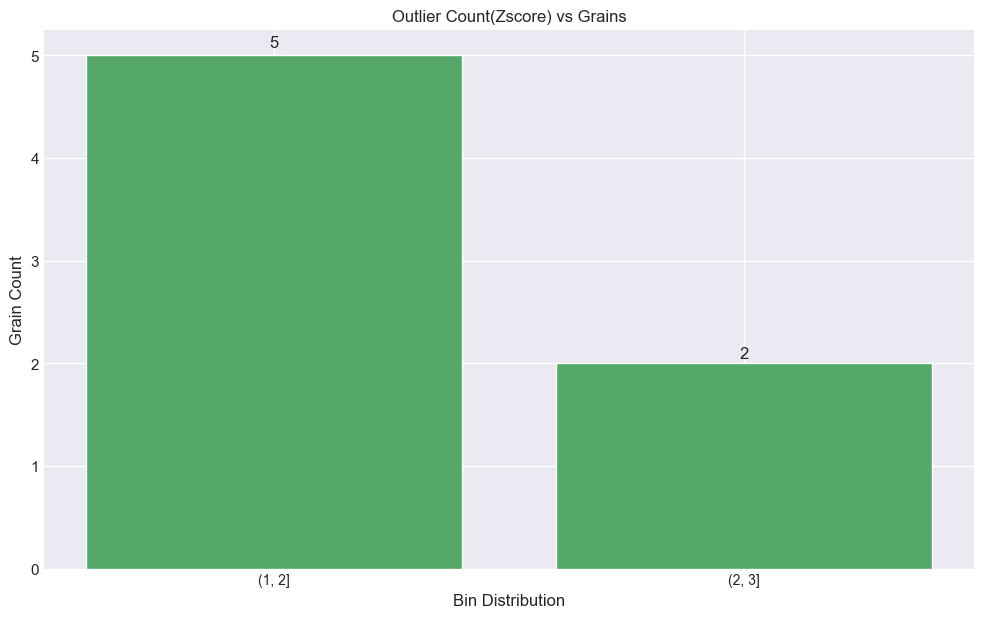


# 22.Outlier in last 1 year

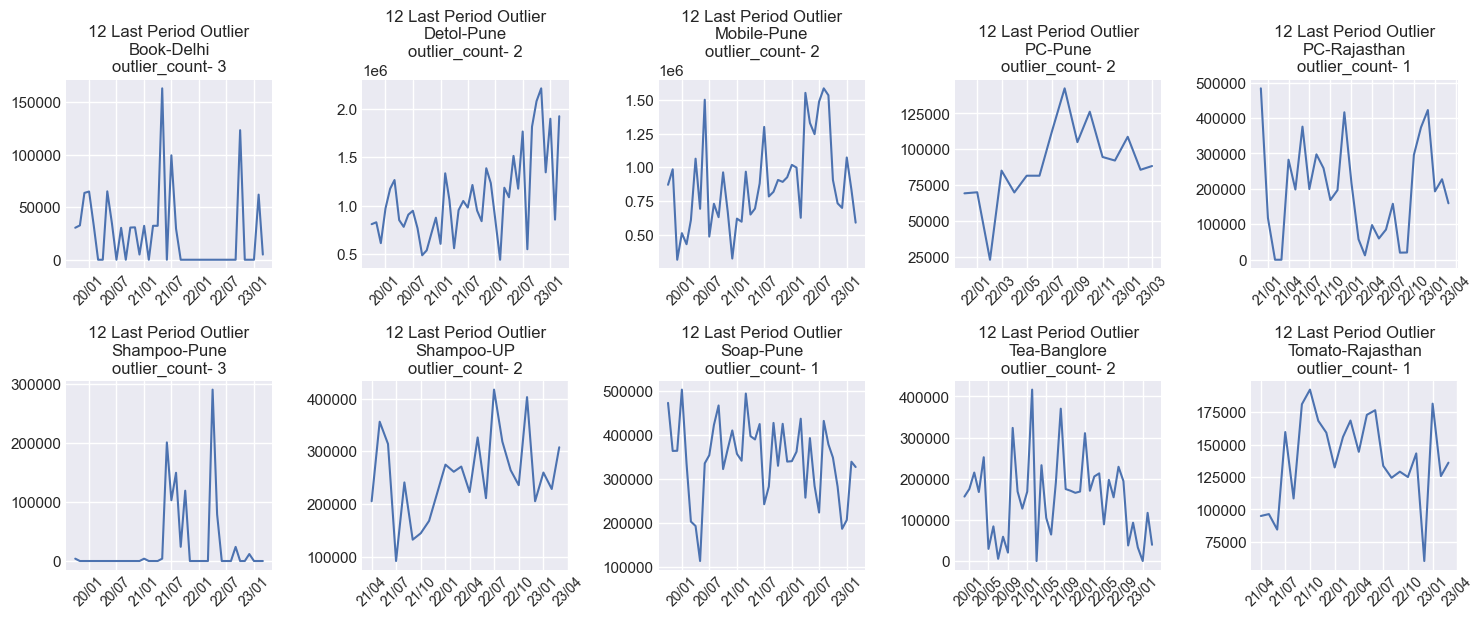




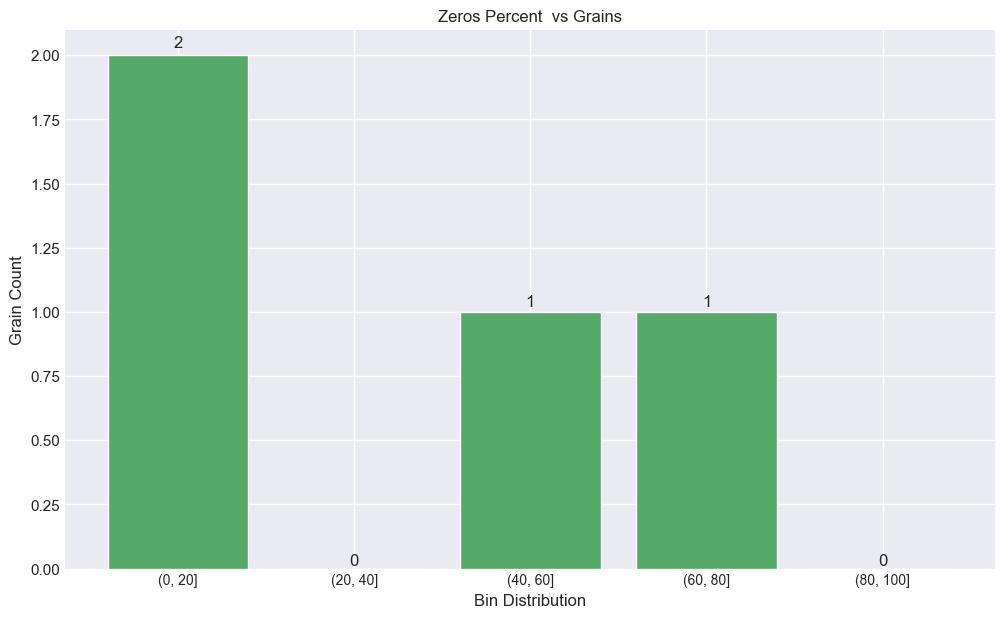




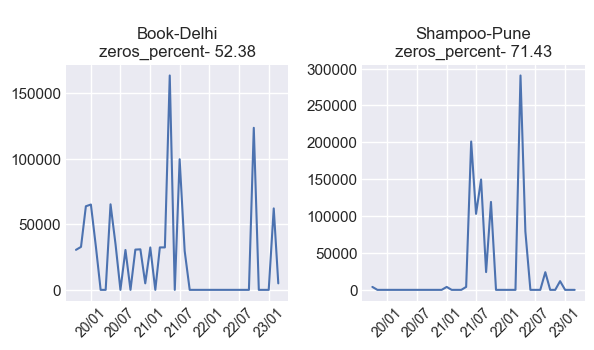




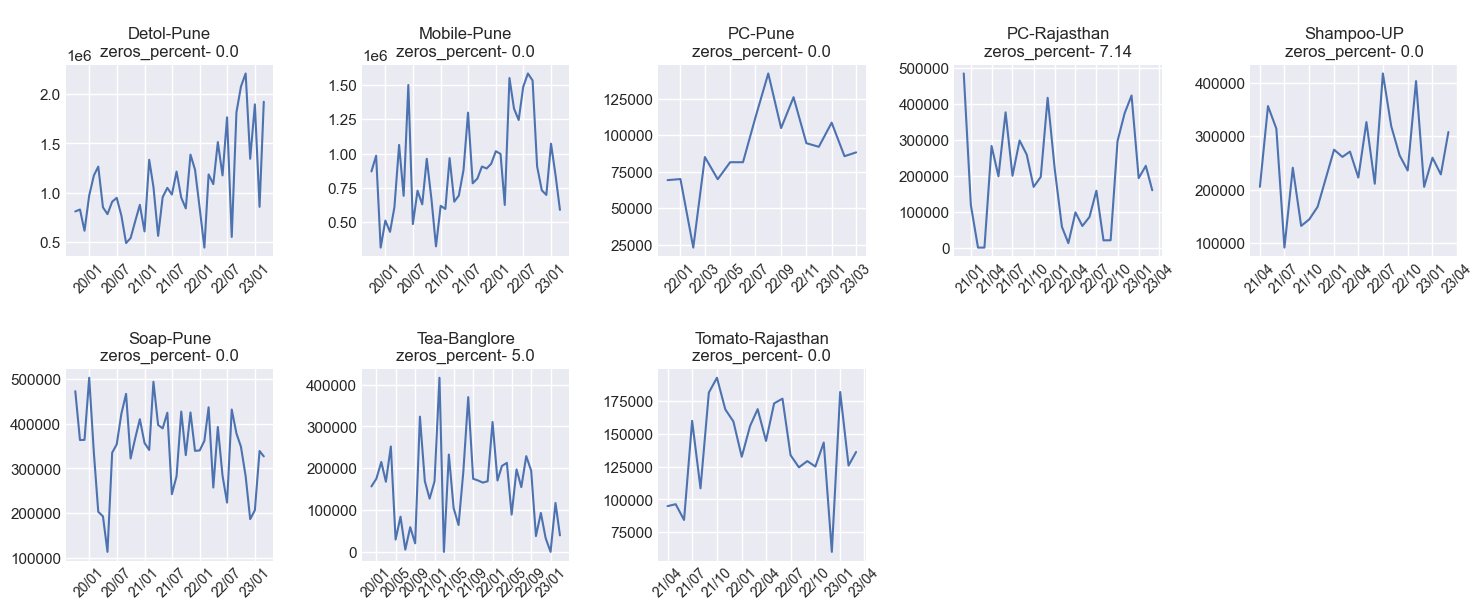
# 23.Number of Zeros



Zeros Examples



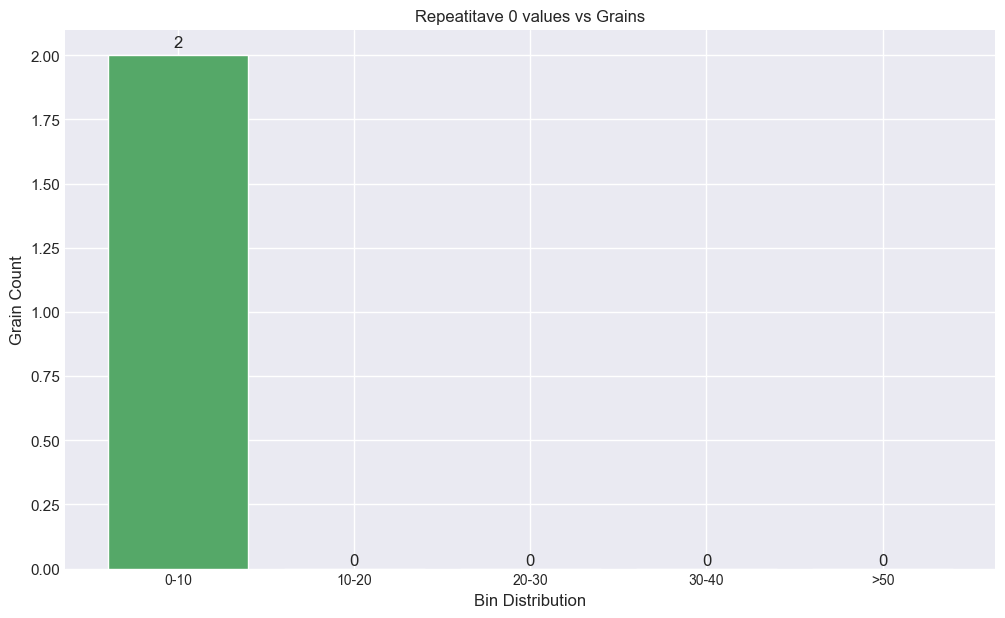
No Zeros Examples



# 24.Repeatitative 1 value

There is no grain having consecuting count as 1

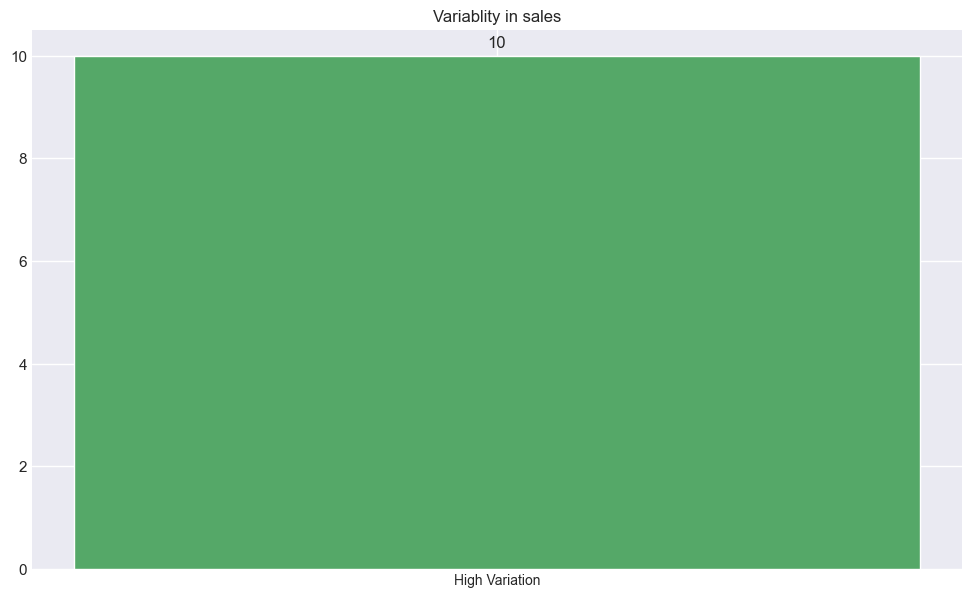
# 25.Repeatitative 0 value

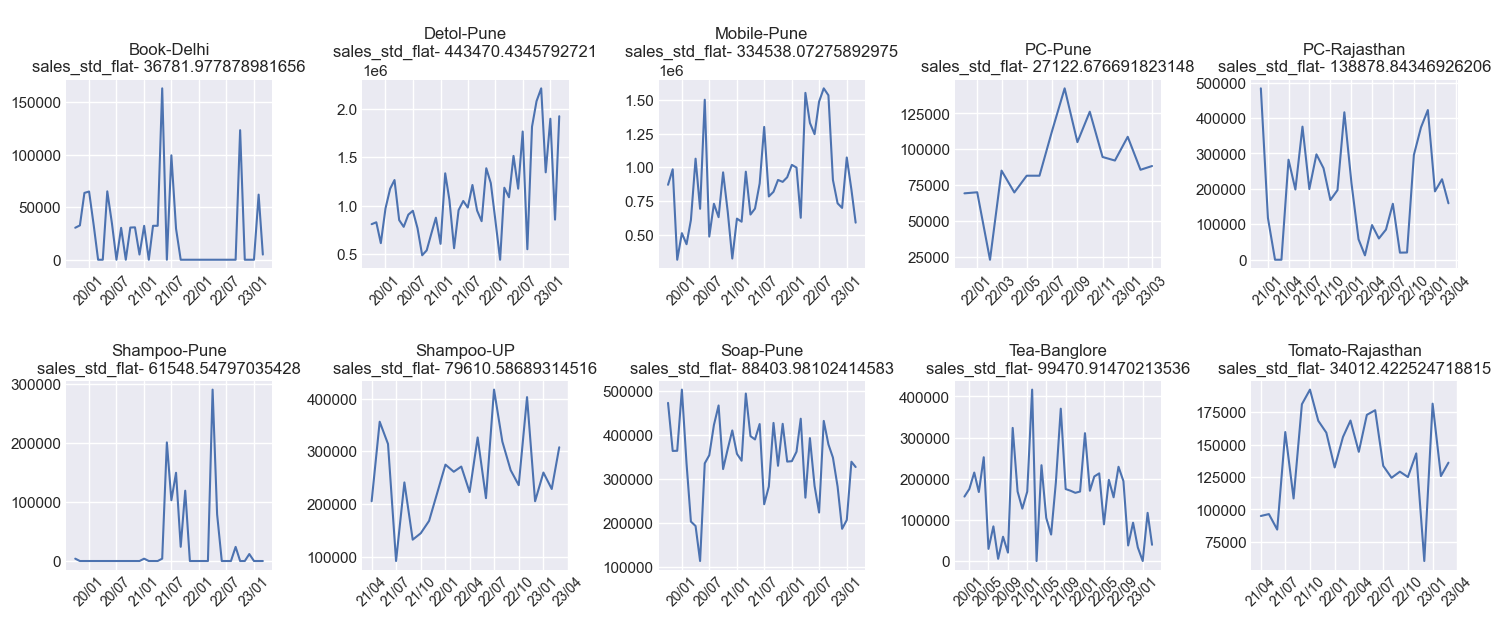


There is no grain having consecuting count as 0

# 26.Flat Sales

No flat grains found and variablity present in sales





Mid Variation Sales Data is empty

Low Variation Sales Data is Empty

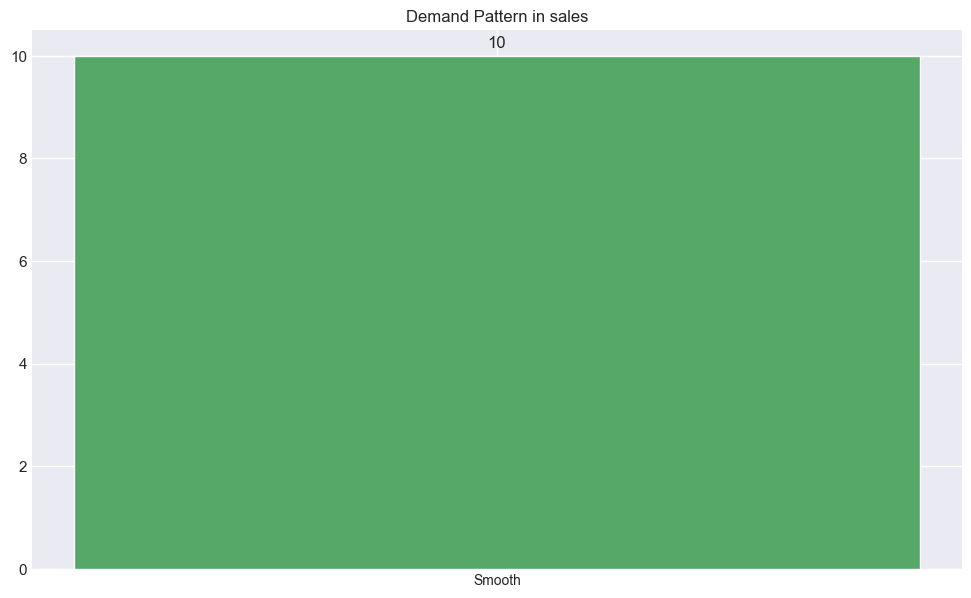
Time taken (in Mins) - 17

# 27.WMAPE and WBIAS

wmape results-Empty DataFrame  
Columns: [Date, sum\_abs\_err, sum\_sales, error\_wmape]  
Index: []

wbias results-Empty DataFrame  
Columns: [Date, sum\_err, sum\_sales, error\_bias]  
Index: []

Time taken (in Mins) - 18



Combine into single CSV..

/Users/umeshnagar/myproject/ABC\_GrainWiseEDA\_Results.csv

Done

Time taken (in Mins) - 18