```
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
from matplotlib import pyplot as plt
import seaborn as sns
from scipy.stats import ttest_1samp
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import statsmodels.api as sm

df=pd.read_csv("Warehouse_and_Retail_Sales.csv")

df.dropna(inplace=True)

df.drop("ITEM DESCRIPTION",axis=1,inplace=True)

df["ITEM TYPE"]=df["ITEM TYPE"].astype("category")

df["RETAIL SALESI IN LAKHS"]=df["RETAIL SALES"] * 100000
df["WAREHOUSE SALES IN LAKHS"]=df["WAREHOUSE SALES"] * 100000

df.head()
```

```
   YEAR  MONTH                           SUPPLIER ITEM CODE ITEM TYPE
\
0  2020      1  REPUBLIC NATIONAL DISTRIBUTING CO    100009      WINE

1  2020      1                          PWSWN INC    100024      WINE

2  2020      1            RELIABLE CHURCHILL LLLP      1001      BEER

3  2020      1          LANTERNA DISTRIBUTORS INC    100145      WINE

4  2020      1               DIONYSOS IMPORTS INC    100293      WINE


   RETAIL SALES  RETAIL TRANSFERS  WAREHOUSE SALES  RETAIL SALESI IN
LAKHS  \
0          0.00               0.0              2.0
0.0
1          0.00               1.0              4.0
0.0
2          0.00               0.0              1.0
0.0
3          0.00               0.0              1.0
0.0
4          0.82               0.0              0.0
82000.0

   WAREHOUSE SALES IN LAKHS
0                  200000.0
```

```
1                  400000.0
2                  100000.0
3                  100000.0
4                       0.0
```

```python
df.drop(columns=["RETAIL SALES","WAREHOUSE SALES"],inplace=True)

df["RETAIL TRANSFERS"]=df["RETAIL TRANSFERS"].astype(int)

df["TOTAL SALES"]=df["RETAIL SALESI IN LAKHS"] + df["WAREHOUSE SALES IN LAKHS"]

df["RETAIL SALESI IN LAKHS"]=df["RETAIL SALESI IN LAKHS"].astype(int)
df["WAREHOUSE SALES IN LAKHS"]=df["WAREHOUSE SALES IN LAKHS"].astype(int)
df["TOTAL SALES"]=df["TOTAL SALES"].astype(int)

df.head()
```

```
   YEAR  MONTH                           SUPPLIER ITEM CODE ITEM TYPE
\
0  2020      1  REPUBLIC NATIONAL DISTRIBUTING CO    100009      WINE

1  2020      1                          PWSWN INC    100024      WINE

2  2020      1            RELIABLE CHURCHILL LLLP      1001      BEER

3  2020      1          LANTERNA DISTRIBUTORS INC    100145      WINE

4  2020      1               DIONYSOS IMPORTS INC    100293      WINE


   RETAIL TRANSFERS  RETAIL SALESI IN LAKHS  WAREHOUSE SALES IN LAKHS
\
0                 0                       0                    200000

1                 1                       0                    400000

2                 0                       0                    100000

3                 0                       0                    100000

4                 0                   82000                         0


   TOTAL SALES
0       200000
1       400000
2       100000
3       100000
4        82000
```

```
df=df.set_index("ITEM CODE")
df.head()
```

```
           YEAR  MONTH                           SUPPLIER ITEM TYPE  \
ITEM CODE
100009     2020      1  REPUBLIC NATIONAL DISTRIBUTING CO      WINE
100024     2020      1                          PWSWN INC      WINE
1001       2020      1            RELIABLE CHURCHILL LLLP      BEER
100145     2020      1          LANTERNA DISTRIBUTORS INC      WINE
100293     2020      1               DIONYSOS IMPORTS INC      WINE

           RETAIL TRANSFERS  RETAIL SALESI IN LAKHS  WAREHOUSE SALES
IN LAKHS  \
ITEM CODE

100009                    0                       0
200000
100024                    1                       0
400000
1001                      0                       0
100000
100145                    0                       0
100000
100293                    0                   82000
0

           TOTAL SALES
ITEM CODE
100009          200000
100024          400000
1001            100000
100145          100000
100293           82000
```

```
df.iloc[100145]
```

```
YEAR                                             2017
MONTH                                              10
SUPPLIER                    STE MICHELLE WINE ESTATES
ITEM TYPE                                        WINE
RETAIL TRANSFERS                                   20
RETAIL SALESI IN LAKHS                        1810000
WAREHOUSE SALES IN LAKHS                      1100000
TOTAL SALES                                   2910000
Name: 78239, dtype: object
```

```
df.head()
```

```
           YEAR  MONTH                           SUPPLIER ITEM TYPE  \
ITEM CODE
100009     2020      1  REPUBLIC NATIONAL DISTRIBUTING CO      WINE
```

```
100024     2020      1                         PWSWN INC      WINE
1001       2020      1           RELIABLE CHURCHILL LLLP      BEER
100145     2020      1         LANTERNA DISTRIBUTORS INC      WINE
100293     2020      1              DIONYSOS IMPORTS INC      WINE

           RETAIL TRANSFERS  RETAIL SALESI IN LAKHS  WAREHOUSE SALES
IN LAKHS  \
ITEM CODE

100009                    0                       0
200000
100024                    1                       0
400000
1001                      0                       0
100000
100145                    0                       0
100000
100293                    0                   82000
0

           TOTAL SALES
ITEM CODE
100009          200000
100024          400000
1001            100000
100145          100000
100293           82000
```

```
df["Z SCORE"]=(df["TOTAL SALES"]-df["TOTAL SALES"].mean())/df["TOTAL
SALES"].std()
```

```
df.head()
```

```
           YEAR  MONTH                            SUPPLIER ITEM TYPE  \
ITEM CODE
100009     2020      1  REPUBLIC NATIONAL DISTRIBUTING CO      WINE
100024     2020      1                         PWSWN INC      WINE
1001       2020      1           RELIABLE CHURCHILL LLLP      BEER
100145     2020      1         LANTERNA DISTRIBUTORS INC      WINE
100293     2020      1              DIONYSOS IMPORTS INC      WINE

           RETAIL TRANSFERS  RETAIL SALESI IN LAKHS  WAREHOUSE SALES
IN LAKHS  \
ITEM CODE

100009                    0                       0
200000
100024                    1                       0
400000
1001                      0                       0
```

```
100000
100145                    0                       0
100000
100293                    0                   82000
0

           TOTAL SALES   Z SCORE
ITEM CODE
100009          200000 -0.114193
100024          400000 -0.106675
1001            100000 -0.117951
100145          100000 -0.117951
100293           82000 -0.118628
```

```
new_df=df[df["Z SCORE"] <=3]
```

```
new_df.drop("Z SCORE",axis=1,inplace=True)
```

```
new_df.shape
```

```
(305723, 8)
```

STATISTICAL ANALYSIS

TTEST

HYPOTHESIS TESTING H0: The sample mean is equal to the hypothesized population mean. H1: The sample mean is not equal to the hypothesized population mean.

```
sample=new_df.sample(n=200)
sample
```

```
           YEAR  MONTH                           SUPPLIER ITEM TYPE  \
ITEM CODE
10572      2019     10                        LEGENDS LTD      KEGS
71650      2019      2                   PALM BAY IMPORTS      WINE
166292     2019     11  REPUBLIC NATIONAL DISTRIBUTING CO      WINE
50759      2019      2            BOSTON BEER CORPORATION      BEER
47236      2017     12              A VINTNERS SELECTIONS      WINE
...         ...    ...                                ...       ...
348474     2020      1                           RVWC LLC      WINE
329220     2019      5                     PUNTO VINO LLC      WINE
234148     2019     10                ELITE WINES IMPORTS      WINE
328266     2019      6  PRESTIGE BEVERAGE GROUP OF MD LLC      WINE
64564      2020      1               CONSTELLATION BRANDS      WINE

           RETAIL TRANSFERS  RETAIL SALESI IN LAKHS  WAREHOUSE SALES
IN LAKHS  \
ITEM CODE
```

```
10572                     0                        0
100000
71650                     2                   318000
300000
166292                    0                        0
300000
50759                     1                   350000
10800000
47236                     0                   323000
0
...                     ...                      ...
...
348474                    0                        0
100000
329220                    0                        0
200000
234148                    0                        0
200000
328266                    0                    16000
0
64564                     0                   187000
0

           TOTAL SALES
ITEM CODE
10572           100000
71650           618000
166292          300000
50759         11150000
47236           323000
...                ...
348474          100000
329220          200000
234148          200000
328266           16000
64564           187000

[200 rows x 8 columns]
```

```
samp_mean=sample["RETAIL SALESI IN LAKHS"].mean()
samp_mean
```

```
780824.98
```

```
pop_mean=df["RETAIL SALESI IN LAKHS"].mean()
pop_mean
```

```
700364.352634506
```

```
t_statistic,p_value=ttest_1samp(sample["RETAIL SALESI IN
LAKHS"],pop_mean)

t_statistic,p_value

(0.41668002023265704, 0.6773615671327835)
```

A positive t-statistic suggests that the sample mean is slightly higher than the hypothesized population mean.

This suggests that the observed difference between the sample mean and the hypothesized population mean (0.416 standard errors) is likely due to random variation and not statistically significant. In other words, the data does not provide strong evidence to support a significant difference between the sample mean and the hypothesized population mean.

p-value (greater than the significance level) indicates that there is insufficient evidence to reject the null hypothesis, suggesting that the sample mean is not significantly different from the hypothesized value.

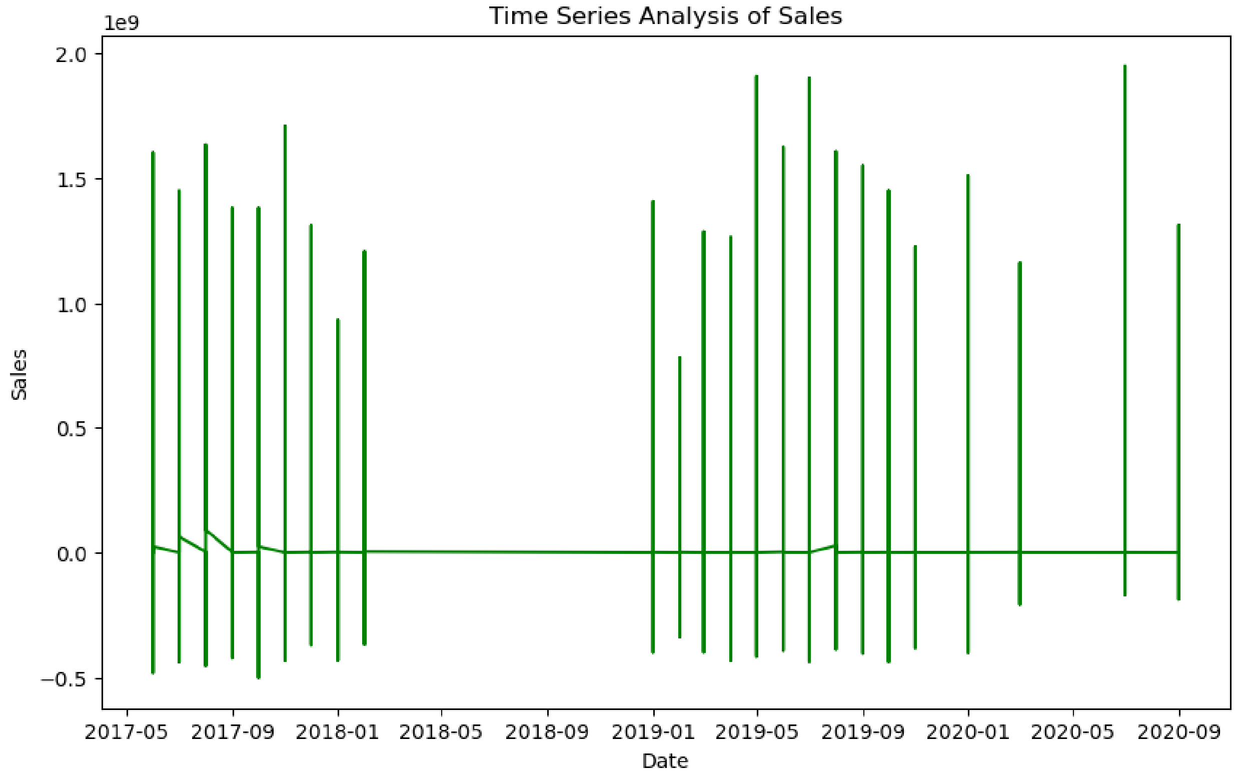
CORRELATION ANALYSIS

```
numerical_columns=["RETAIL TRANSFERS","RETAIL SALESI IN
LAKHS","WAREHOUSE SALES IN LAKHS","TOTAL SALES"]
numerical_columns

['RETAIL TRANSFERS',
 'RETAIL SALESI IN LAKHS',
 'WAREHOUSE SALES IN LAKHS',
 'TOTAL SALES']

correlation_matrix=df[numerical_columns].corr()
correlation_matrix

                          RETAIL TRANSFERS  RETAIL SALESI IN LAKHS  \
RETAIL TRANSFERS                  1.000000                0.979426
RETAIL SALESI IN LAKHS            0.979426                1.000000
WAREHOUSE SALES IN LAKHS          0.493279                0.501256
TOTAL SALES                       0.574493                0.584324

                          WAREHOUSE SALES IN LAKHS  TOTAL SALES
RETAIL TRANSFERS                          0.493279     0.574493
RETAIL SALESI IN LAKHS                    0.501256     0.584324
WAREHOUSE SALES IN LAKHS                  1.000000     0.995104
TOTAL SALES                               0.995104     1.000000
```

Strong positive correlation between RETAIL TRANSFERS and RETAIL SALESI IN LAKHS (0.96). Moderate positive correlations between RETAIL TRANSFERS and WAREHOUSE SALES IN LAKHS (0.49) and TOTAL SALES (0.57). Moderate positive correlations between RETAIL SALESI IN LAKHS and WAREHOUSE SALES IN LAKHS (0.49) and TOTAL SALES (0.58). Very strong positive correlation between WAREHOUSE SALES IN LAKHS and TOTAL SALES (0.99).

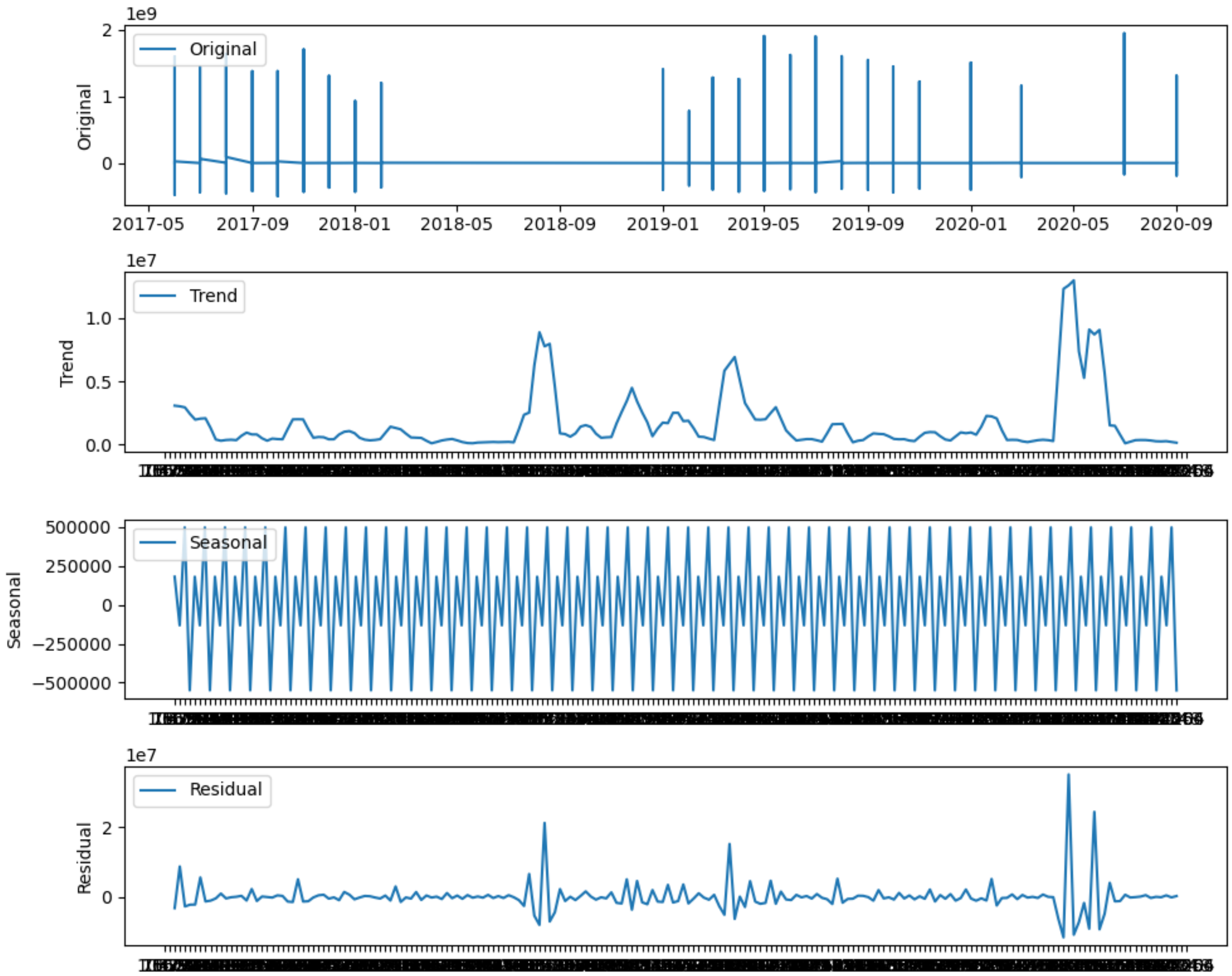
Time Series Analysis

```
df["Date"] = pd.to_datetime(df["YEAR"].astype(str) + '-' +
df["MONTH"].astype(str))

df.set_index("Date", inplace=True)

df.sort_index(inplace=True)

plt.figure(figsize=(10, 6))
plt.plot(df.index, df["TOTAL SALES"], label="TOTAL
SALES",color="green")
plt.xlabel("Date")
plt.ylabel("Sales")
plt.title("Time Series Analysis of Sales")
plt.show()
```



MACHINE LEARNING TECHNIQUES

```
df.columns

Index(['YEAR', 'MONTH', 'SUPPLIER', 'ITEM TYPE', 'RETAIL TRANSFERS',
       'RETAIL SALESI IN LAKHS', 'WAREHOUSE SALES IN LAKHS', 'TOTAL
SALES',
       'Z SCORE'],
      dtype='object')

result = sm.tsa.seasonal_decompose(sample["TOTAL SALES"],
model="additive",period=4)

ig, (ax1, ax2, ax3, ax4) = plt.subplots(4, 1, figsize=(10, 8))
ax1.plot(df['TOTAL SALES'], label='Original')
ax1.legend(loc='upper left')
ax1.set_ylabel('Original')

ax2.plot(result.trend, label='Trend')
ax2.legend(loc='upper left')
ax2.set_ylabel('Trend')

ax3.plot(result.seasonal, label='Seasonal')
ax3.legend(loc='upper left')
ax3.set_ylabel('Seasonal')

ax4.plot(result.resid, label='Residual')
ax4.legend(loc='upper left')
ax4.set_ylabel('Residual')

plt.tight_layout()
plt.show()
```


1e9
Original
Original
2
1
0
2017-05
2017-09
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09
2020-01
2020-05
2020-09
1e7
Trend
Trend
1.0
0.5
0.0
Seasonal
Seasonal
500000
250000
0
−250000
−500000
1e7
Residual
Residual
2
0