Forecasting Project

#	Name	ID
1	Omar Ahmed Abdulhameed	20190653
2	Shady Mohamed Sobhy	20200248
3	Ahmed Khaleel Abdullah	20200012

Import Libraries

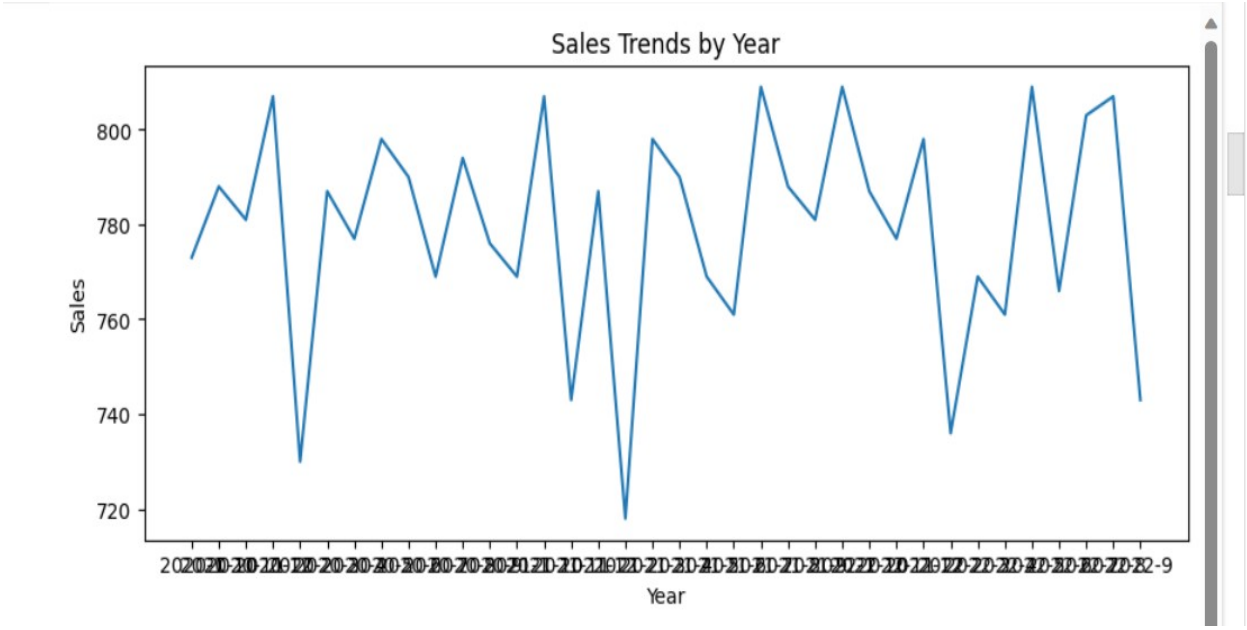
```
[76] import pandas as pd
     import numpy as np
     import tensorflow as tf
     import matplotlib.pyplot as plt
     import statsmodels.api as sm
     from statsmodels.tsa.holtwinters import ExponentialSmoothing
     from sklearn.model_selection import train_test_split
     from sklearn.preprocessing import StandardScaler,LabelEncoder, MinMaxScaler
     from tensorflow.keras.models import Sequential
     from tensorflow.keras.layers import Dense
```

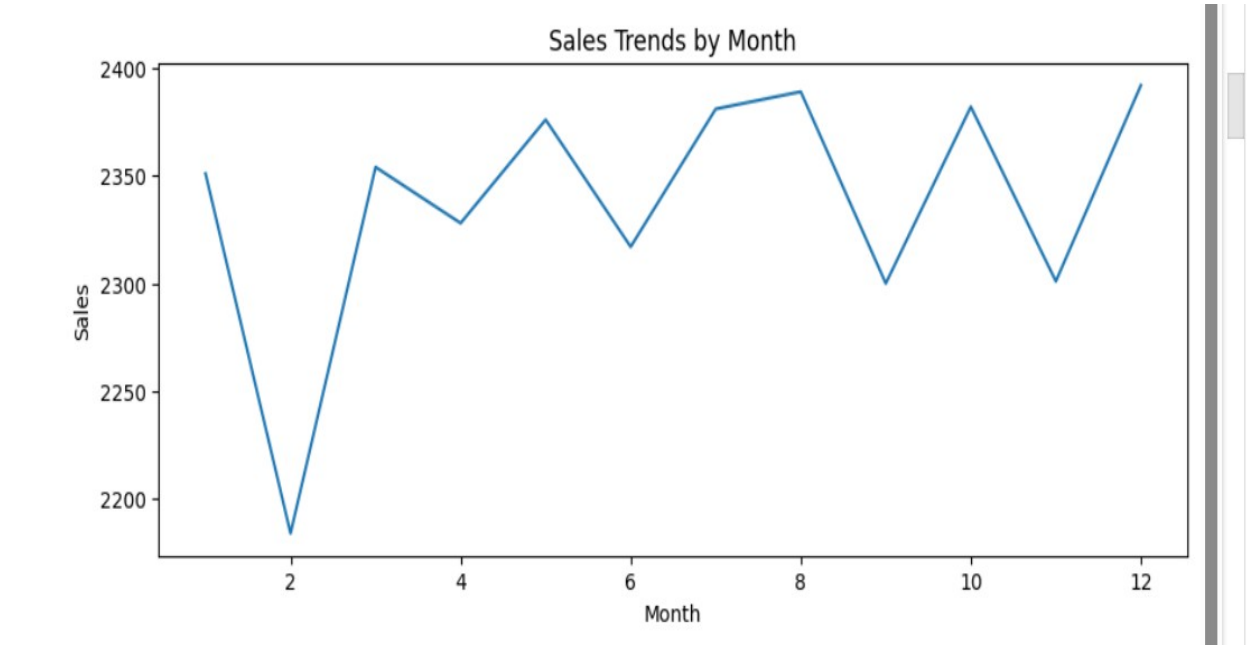
Exploratory Data Analysis (EDA)

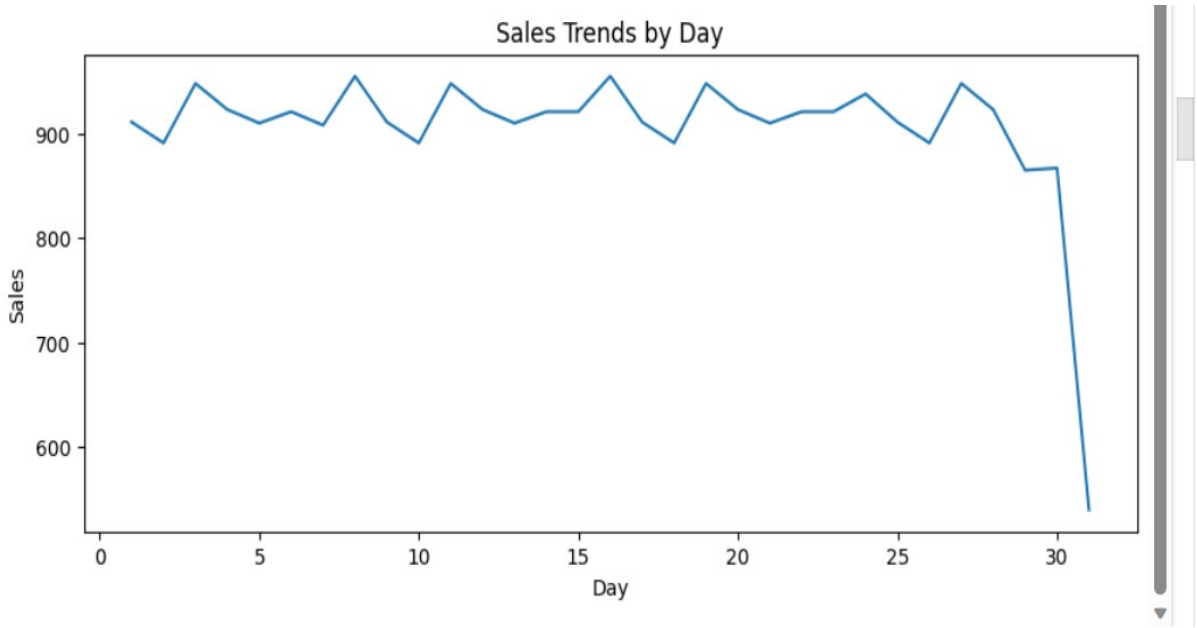
```
# Summary statistics
print(df.describe())
```

```
        product_ID  quantity_sold  sales_price  customer_id          age  \
count  1094.000000    1094.000000  1096.000000  1096.000000  1096.000000
mean     25.644424     121.383912    69.321168    12.961679    40.343066
std      12.254867     167.717350    78.228076     7.199213    13.737826
min      11.000000       7.000000     7.000000     1.000000    23.000000
25%      17.000000      33.000000    26.000000     7.000000    29.000000
50%      26.000000      51.000000    44.000000    13.000000    36.000000
75%      32.750000     112.000000    74.000000    19.000000    52.000000
max      51.000000     643.000000   323.000000    25.000000    65.000000

       purchase_frequency  total_purchase_amount
count         1096.000000            1093.000000
mean             4.647810            3371.989936
std              2.512808            1672.811649
min              1.000000            1132.000000
25%              2.000000            2255.000000
50%              5.000000            2356.000000
75%              7.000000            5432.000000
max              9.000000            5544.000000
```

Plotting sales


Sales Trends by Year
Sales
800
780
760
740
720
Year


Sales Trends by Month
Sales
2400
2350
2300
2250
2200
2
4
6
8
10
12
Month



Customer behavior analysis

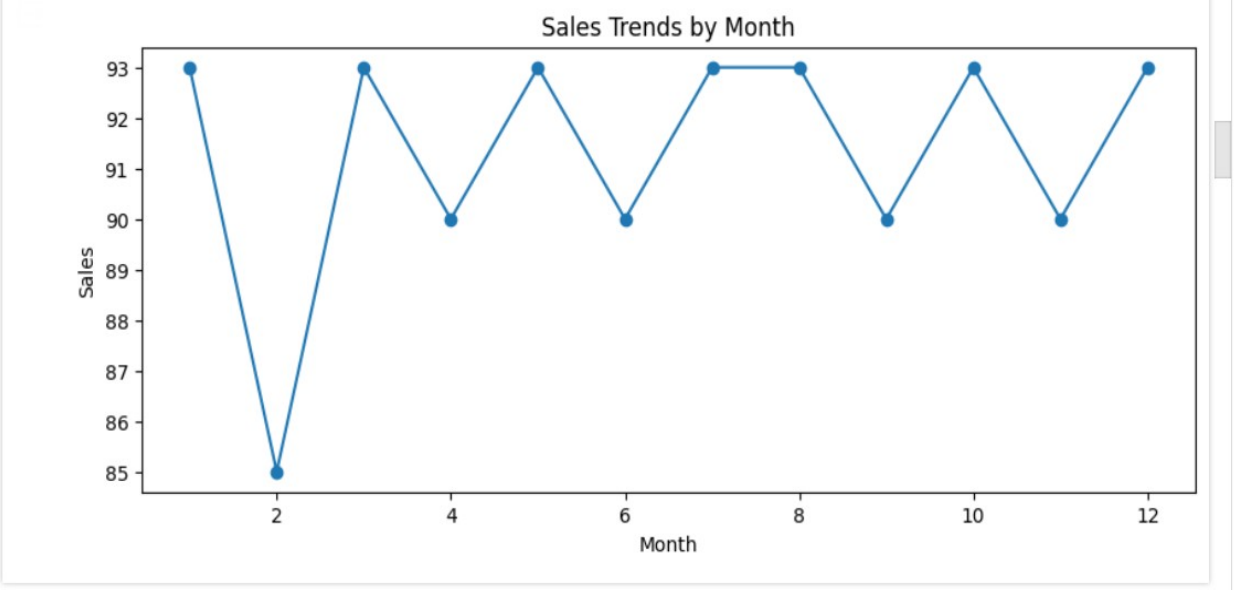
```
Total number of customers: 25
Average purchase by customer:
 customer_id
1     1.0
2     1.0
3     1.0
4     1.0
5     1.0
6     1.0
7     1.0
8     1.0
9     1.0
10    1.0
11    1.0
12    1.0
13    1.0
14    1.0
15    1.0
16    1.0
17    1.0
18    1.0
19    1.0
20    1.0
21    1.0
22    1.0
23    1.0
24    1.0
25    1.0
Name: sales, dtype: float64
```

Popular products

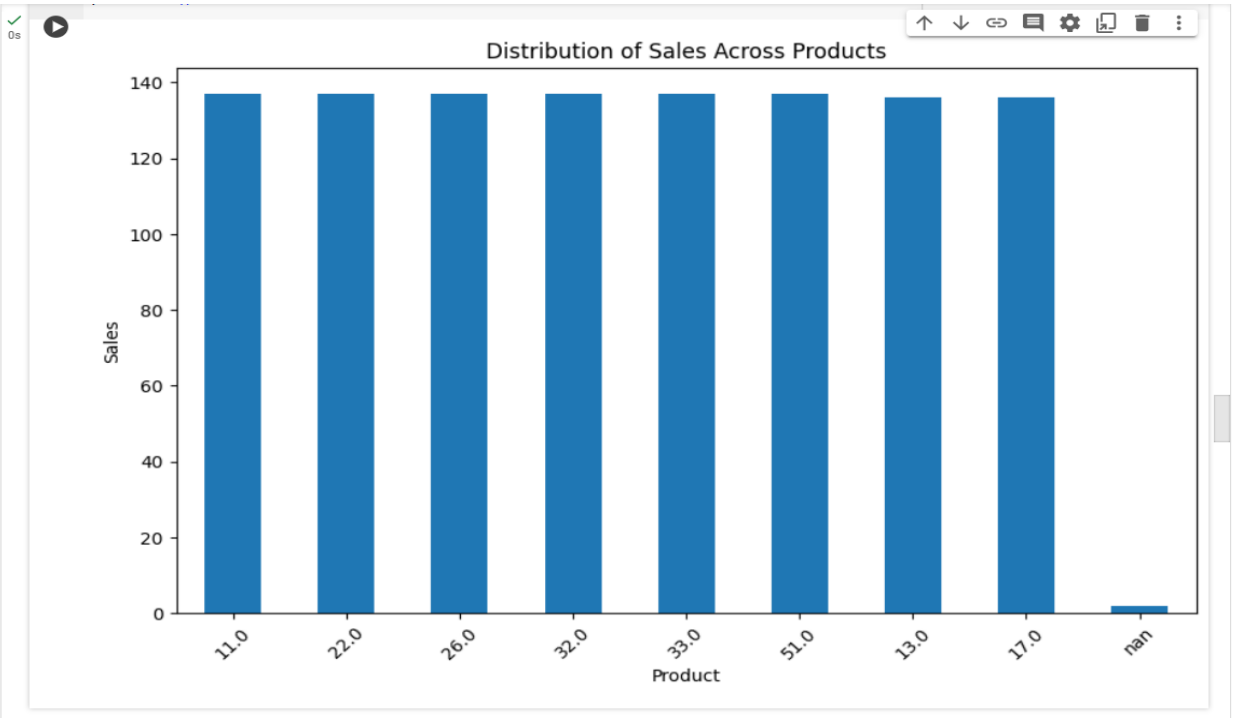
```
[105] # Popular products
      popular_products = df['product_ID'].value_counts().head(10)

      print("Top 10 popular products:\n", popular_products)
```

```
Top 10 popular products:
 22.0    137
32.0    137
11.0    137
33.0    137
26.0    137
51.0    137
17.0    136
13.0    136
0.0       2
Name: product_ID, dtype: int64
```

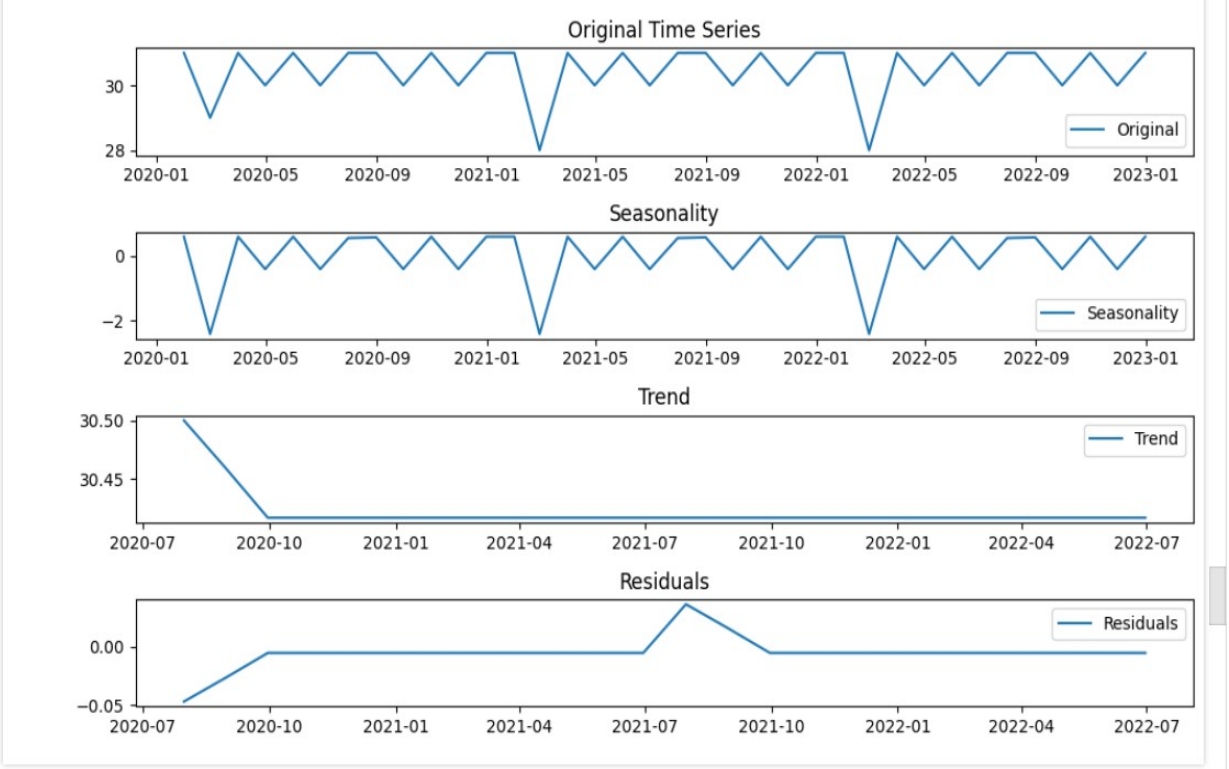
Sales trends over time (line plot)



Distribution of sales across products (bar plot)



Plot the original time series, seasonality, trend, and residuals



Plot the actual sales and forecasted sales