Introduction to the problem statement

Conglomerate Inc., has a variety of brands under its umbrella which includes Diapers, Headphones, and Breakfast cereals. In order to understand the effect of price elasticity on a categorial basis, they want to run promotional campaigns, this will eventually help them to strengthen their promotional strategy. The first and foremost motive is to analyze the impact of price on product demand for different products and product categories. At last, a statistical report should be generated to visualize the data and understand the patterns.

Understanding Datasets and Data Exploration

The given file for the dataset contains Tab Separated values containing the data for per day price and sale of a particular product. In the dataset, we're provided with Item Id, Average Price, Date, and Units sold. A total of 462000 entities are provided, including Diapers, Headphones, and Cereals as categories.

As we can see from the data below, there are different IDs for different categories, but even there are other IDs at the item level. For instance, let's consider the Diapers category, there are 110000 entities in total where there are only 100 unique Item IDs, 1100 of each kind.

Column-wise dataset description

```
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
 0   item_id     462000 non-null  object
 1   category    462000 non-null  object
 2   date        462000 non-null  object
 3   avg_price   462000 non-null  float64
 4   units_sold  462000 non-null  int64
dtypes: float64(1), int64(1), object(3)
memory usage: 17.6+ MB
```

No. of unique IDs

```
np.unique(diaper_df['item_id'])
```

```
array(['D1036', 'D1053', 'D1207', 'D1211', 'D1278', 'D131', 'D1322',
       'D1346', 'D1349', 'D1408', 'D1425', 'D1454', 'D1515', 'D1714',
       'D1938', 'D1963', 'D2050', 'D2092', 'D2096', 'D2391', 'D2587',
       'D268', 'D2786', 'D2927', 'D3029', 'D306', 'D3078', 'D3108',
       'D3237', 'D3248', 'D3307', 'D3684', 'D3864', 'D4192', 'D4231',
       'D4263', 'D4389', 'D4436', 'D4564', 'D4595', 'D4670', 'D4687',
       'D4844', 'D506', 'D5111', 'D5119', 'D5156', 'D5374', 'D5491',
       'D5574', 'D559', 'D5674', 'D5826', 'D5991', 'D6035', 'D6230',
       'D6247', 'D6346', 'D6573', 'D6836', 'D6853', 'D6897', 'D6936',
       'D702', 'D7147', 'D718', 'D7274', 'D7328', 'D7329', 'D7511',
       'D7680', 'D7689', 'D7734', 'D7755', 'D7870', 'D7938', 'D796',
       'D80', 'D8075', 'D8085', 'D8139', 'D8141', 'D8200', 'D8421',
       'D8500', 'D8542', 'D8755', 'D8829', 'D8861', 'D8876', 'D9378',
       'D9387', 'D9492', 'D9604', 'D9687', 'D9761', 'D9775', 'D9904',
       'D9927', 'D9976'], dtype=object)
```

Item level data distribution

```
diaper_df['item_id'].value_counts()
```

```
D9775    1100
D2092    1100
D8876    1100
D3108    1100
D7680    1100
         ...
D8500    1100
D6936    1100
D8200    1100
D7511    1100
D1349    1100
Name: item_id, Length: 100, dtype: int64
```

Grouped data in accordance with a particularly unique item id

	item_id	category	date	avg_price	units_sold
0	D9775	Diapers	2017-01-01	12.46	94
420	D9775	Diapers	2017-01-02	14.12	40
840	D9775	Diapers	2017-01-03	11.72	32
1260	D9775	Diapers	2017-01-04	9.37	24
1680	D9775	Diapers	2017-01-05	12.60	35
...	...	...	...	...	...
459900	D9775	Diapers	2020-01-01	10.12	74
460320	D9775	Diapers	2020-01-02	12.49	53
460740	D9775	Diapers	2020-01-03	12.45	73
461160	D9775	Diapers	2020-01-04	14.16	100
461580	D9775	Diapers	2020-01-05	11.10	110

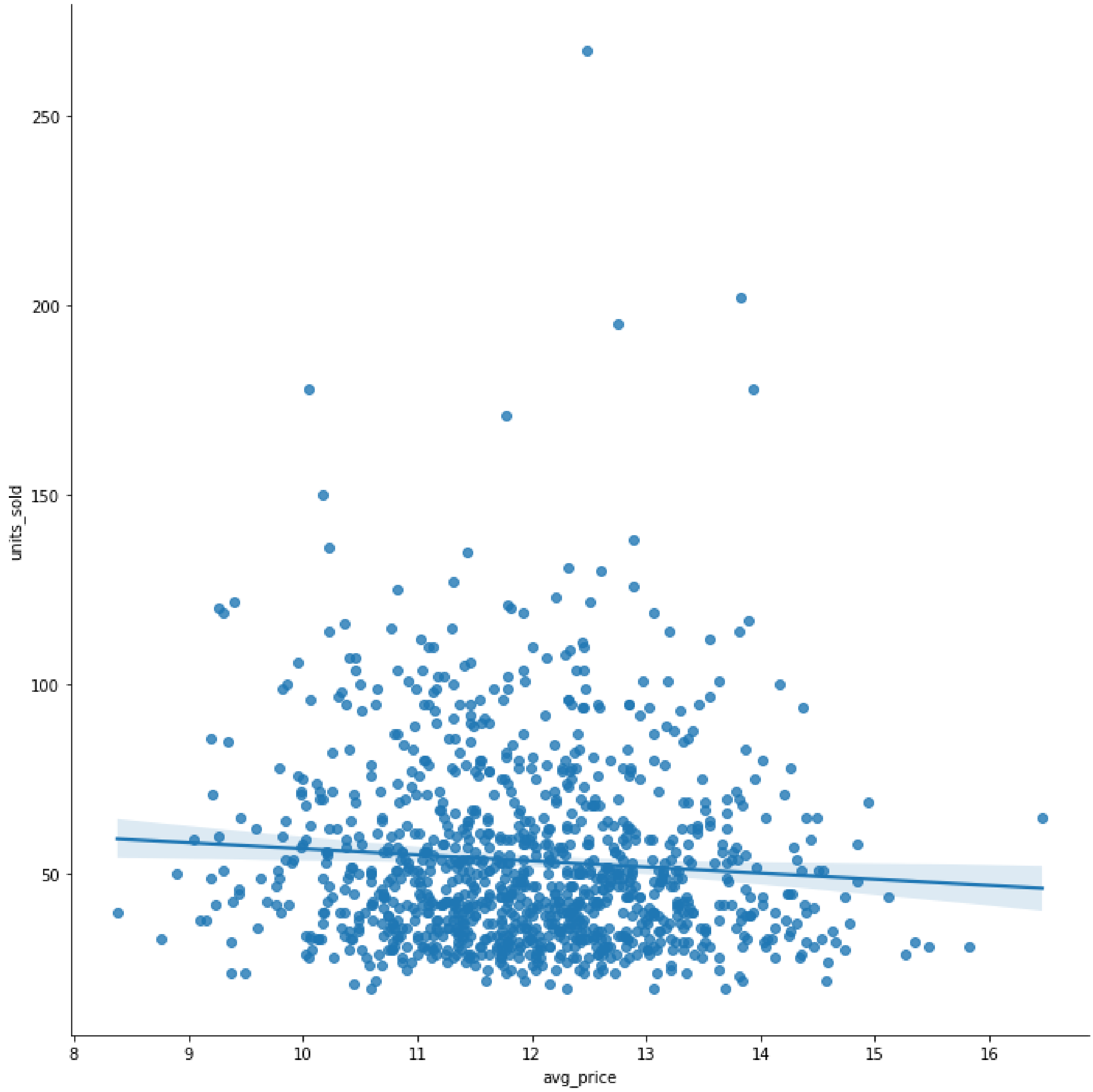
1100 rows × 5 columns

Data Visualization

Scatter Plot

As we know from the Data exploration stage that we've different data at the item level too as we've different item ID's, To draw the relation between *Units sold* and *Average price* and to understand the trends associated with a particular item_id we'll derive a scattered plot of *Average price* against *Units sold* and do the further analysis.

Instance plot for item_id D9775 for Diapers



Ordinary least square estimation

To estimate the unknown parameters other than the average price we'll use this linear regression technique i.e. the ordinary least square estimation.

Instance OLS regression result

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     np.log(units_sold)   R-squared:                       0.260
Model:                            OLS   Adj. R-squared:                  0.260
Method:                 Least Squares   F-statistic:                     386.7
Date:                Wed, 07 Sep 2022   Prob (F-statistic):           5.38e-74
Time:                        09:47:49   Log-Likelihood:                -561.34
No. Observations:                1100   AIC:                             1127.
Df Residuals:                    1098   BIC:                             1137.
Df Model:                           1
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
Intercept            14.2685      0.611     23.368      0.000      13.070      15.467
np.log(avg_price)    -2.9329      0.149    -19.665      0.000      -3.226      -2.640
==============================================================================
Omnibus:                       55.881   Durbin-Watson:                   0.898
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               63.296
Skew:                           0.563   Prob(JB):                     1.80e-14
Kurtosis:                       3.333   Cond. No.                         218.
==============================================================================
```

Price Elasticity

Price elasticity of demand is the ratio of the percentage change in quantity demanded of a product to the percentage change in price. Economists employ it to understand how supply and demand change when a product's price changes.

If there is a negative percentage change in price, it means there is a price decrease. When there is a price decrease, the quantity demanded of goods increases. So, the change in quantity demanded is positive.

	item_id	category	date	avg_price	units_sold	% Change in Demand	% Change in Price	Price Elasticity
0	H1112	Headphones	2017-01-01	68.59	11	NaN	NaN	NaN
1	H1112	Headphones	2017-01-02	59.01	8	-0.272727	-0.139671	1.952648
2	H1112	Headphones	2017-01-03	65.35	6	-0.250000	0.107439	-2.326893
3	H1112	Headphones	2017-01-04	75.08	3	-0.500000	0.148891	-3.358171
4	H1112	Headphones	2017-01-05	59.62	9	2.000000	-0.205914	-9.712807
...	...	...	...	...	...	...	...	...
1095	H1112	Headphones	2020-01-01	61.35	12	0.090909	-0.026963	-3.371658
1096	H1112	Headphones	2020-01-02	53.85	20	0.666667	-0.122249	-5.453333
1097	H1112	Headphones	2020-01-03	64.91	12	-0.400000	0.205385	-1.947559
1098	H1112	Headphones	2020-01-04	65.31	18	0.500000	0.006162	81.137500
1099	H1112	Headphones	2020-01-05	58.65	29	0.611111	-0.101975	-5.992743

1100 rows × 8 columns