Correlation Matrix Explanation:

	Item_Weight	Item_Visibility	Item_MRP	Item_Outlet_Sales
Item_Weight	1.000000	0.021314	0.307007	0.201944
Item_Visibility	0.021314	1.000000	-0.127025	-0.136765
Item_MRP	0.307007	-0.127025	1.000000	0.993958
Item_Outlet_Sales	0.201944	-0.136765	0.993958	1.000000

1. **Correlation Coefficients**:

    - Values range from -1 to 1.
    - A positive value indicates a direct relationship (as one variable increases, so does the other).
    - A negative value indicates an inverse relationship (as one variable increases, the other decreases).
    - Values close to 0 suggest little to no linear relationship.

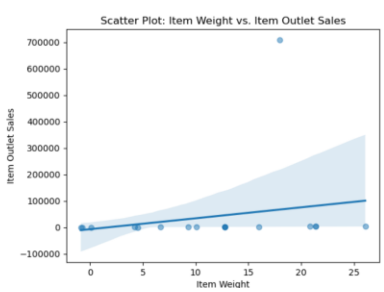
Potential Problem Statements and Analysis:

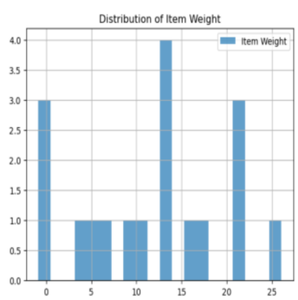
1. **Impact of Item Weight on Sales:**

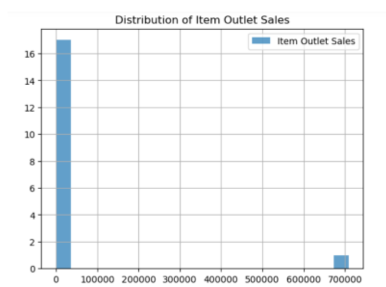
    1.Analyze the correlation between Item_Weight and Item_Outlet_Sales.

    ➢ **Item_Weight and Item_Outlet_Sales**: Correlation = 0.202, showing a weak positive relationship. Heavier items may sell slightly better, but the relationship isn't strong

    2.Visualize the relationship using scatter plots with regression line and histograms as below:-







    3.Test the hypothesis that heavier items have higher sales using statistical tests.

    ➢ To test whether there is a significant relationship between **Item_Weight** and **Item_Outlet_Sales**, the **Pearson correlation test** is the most suitable statistical test.

Pearson correlation coefficient: 0.202

- This indicates a **weak positive correlation** between **Item_Weight** and **Item_Outlet_Sales**. As the weight of the item increases, there is a slight increase in the sales, but the relationship is not very strong.

P-value: 0.422

- Since the **p-value** is greater than **0.05**, **fail to reject the null hypothesis**. This means that the observed correlation between **Item_Weight** and **Item_Outlet_Sales** is **not statistically significant** at the 5% significance level. In other words, the weak positive correlation you observed could be due to random chance, and there's not enough evidence to suggest a significant relationship between the two variables.
- While there is a positive correlation between **Item_Weight** and **Item_Outlet_Sales**, the correlation is weak and not statistically significant. Therefore, you cannot conclude that heavier items consistently lead to higher sales based on this analysis.

```python
from scipy.stats import pearsonr

# Assuming dataset is your DataFrame with 'Item_Weight' and 'Item_Outlet_Sales' columns
correlation, p_value = pearsonr(dataset['Item_Weight'], dataset['Item_Outlet_Sales'])

print(f"Pearson correlation coefficient: {correlation}")
print(f"P-value: {p_value}")
```

```
Pearson correlation coefficient: 0.2019441898937217
P-value: 0.4216289445198158
```

<u>Hypothesis Testing for the Relationship Between Item Weight and Item Outlet Sales</u>:

- **Null Hypothesis (H_0)**: There is no significant relationship between **Item_Weight** and **Item_Outlet_Sales**. (The correlation coefficient is zero.)
- **Alternative Hypothesis (H_1)**: There is a significant relationship between **Item_Weight** and **Item_Outlet_Sales**. (The correlation coefficient is not zero.)

Results:

- **Pearson correlation coefficient** = 0.202 (weak positive correlation)
- **P-value** = 0.422 (greater than 0.05)

Conclusion:

- **Since the p-value is 0.422, which is greater than 0.05**, **fail to reject the null hypothesis**.
- This means you do **not have sufficient evidence** to conclude that heavier items are significantly related to higher sales.

Final Answer:

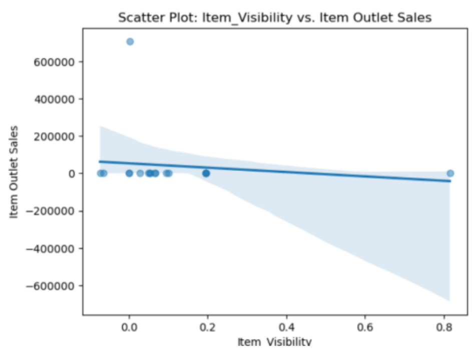
- **Reject the null hypothesis?** No (p-value > 0.05).
- **Fail to reject the null hypothesis?** Yes (there's insufficient evidence against the null).
- **Accept the null hypothesis?** No, simply fail to reject it.
- **Accept the alternative hypothesis?** No, because failing to reject the null means cannot support the alternative hypothesis

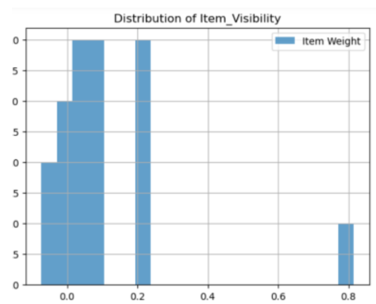
2. **Impact of Item Visibility on Item_Outlet_Sales:**

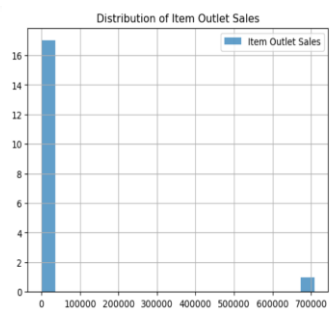
1.Analyze the correlation between Item_Visibility and Item_Outlet_Sales.

- **Item_Visibility and Item_Outlet_Sales**: Correlation = -0.137, showing a weak negative relationship. Items with higher visibility tend to have slightly lower sales.

2.Visualize the relationship using scatter plots with regression line and histograms as below:-







3.Test the hypothesis that higher visibility leads to higher sales.

```python
from scipy.stats import pearsonr

# Assuming dataset with 'Item_Weight' and 'Item_Outlet_Sales' columns
correlation, p_value = pearsonr(dataset['Item_Visibility'], dataset['Item_Outlet_Sales'])

print(f"Pearson correlation coefficient: {correlation}")
print(f"P-value: {p_value}")
```

```
Pearson correlation coefficient: -0.1367650653337633
P-value: 0.5884094446668455
```

Pearson Correlation Coefficient (-0.1368): This value measures the strength and direction of the linear relationship between two variables. A value close to 0 suggests a weak linear relationship. Here, the correlation is slightly negative, indicating a weak inverse relationship, but it's very close to zero, so the relationship is minimal.

P-value (0.5884): The p-value tests the null hypothesis that there is no linear correlation between the two variables. A high p-value (commonly above 0.05) indicates insufficient evidence to reject the null hypothesis. In this case, 0.5884 is well above 0.05, meaning the correlation is not statistically significant

Hypothesis Testing for the Relationship Between Item Weight and Item Outlet Sales:

- **Null Hypothesis (H_0)**: There is no significant relationship between **Item_Visibility** and **Item_Outlet_Sales**. (The correlation coefficient is zero.)
- **Alternative Hypothesis (H_1)**: There is a significant relationship between **Item_Visibility** and **Item_Outlet_Sales**. (The correlation coefficient is not zero.)

Results:

- **Pearson correlation coefficient** = -0.1368 (weak negative correlation)
- **P-value** = 0.5884 (greater than 0.05)

Conclusion:

- **Since the p-value is** 0.5884 , **which is greater than 0.05**, **fail to reject the null hypothesis**.
- This means you do **not have sufficient evidence** to conclude that items visibility is not significantly related to outlet sales.

Final Answer:

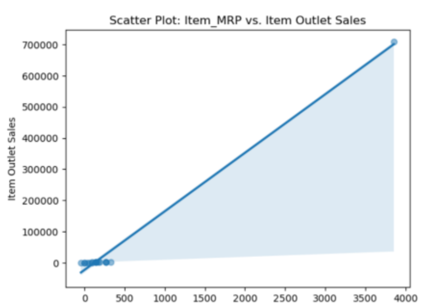
- **Reject the null hypothesis?** No (p-value > 0.05).
- **Fail to reject the null hypothesis?** Yes (there's insufficient evidence against the null).
- **Accept the null hypothesis?** No, simply fail to reject it.
- **Accept the alternative hypothesis?** No, because failing to reject the null means cannot support the alternative hypothesis

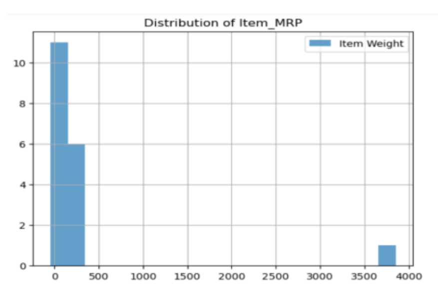
3. **Impact of Item Price on Item_Outlet_Sales:**

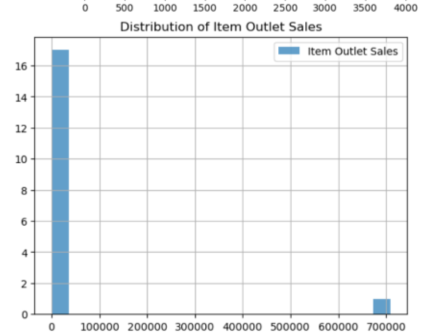
    1.Analyze the correlation between Item_MRP and Item_Outlet_Sales.

    - **Item_MRP and Item_Outlet_Sales**: Correlation = 0.994, which is very close to 1. This suggests a very strong positive relationship, meaning that higher-priced items are strongly associated with higher sales.

    2.Visualize the relationship using scatter plots with regression line and histograms as below







    3.Test the hypothesis that higher prices lead to higher sales.

```python
from scipy.stats import pearsonr

# Assuming with two coloum
correlation, p_value = pearsonr(dataset['Item_MRP'], dataset['Item_Outlet_Sales'])

print(f"Pearson correlation coefficient: {correlation}")
print(f"P-value: {p_value}")
```

```
Pearson correlation coefficient: 0.9939575116574986
P-value: 8.767802493516824e-17
```

Hypothesis Testing for the Relationship Between Two Variables:

1. **Hypotheses**:

   - **Null Hypothesis (H_0)**: There is no significant relationship between the Item MRP and Item Outlet Sales (correlation coefficient = 0).
   - **Alternative Hypothesis (H_1)**: There is a significant relationship between the two variables (correlation coefficient ≠ 0).

2. **Results**:

   - **Pearson Correlation Coefficient = 0.99396**:
     This indicates an extremely strong positive linear relationship between the two variables.
   - **P-value = 8.77×10^{-17} (essentially 0)**:
     The p-value is far smaller than the standard significance level of 0.05, meaning this result is highly statistically significant.

3. **Conclusion**:

   - Since the p-value is extremely small (<< 0.05), reject the null hypothesis (H_0).
   - This means there is sufficient evidence to conclude that the two variables are significantly related.

Final Answer:

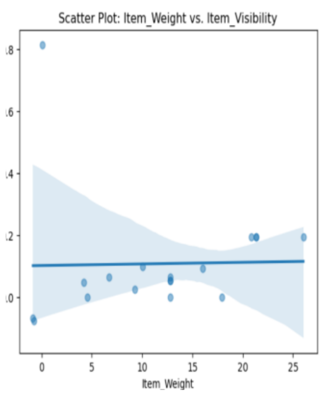
- **Reject the null hypothesis?** Yes (p-value << 0.05).
- **Fail to reject the null hypothesis?** No.
- **Accept the null hypothesis?** No, because it has been rejected.
- **Accept the alternative hypothesis?** Yes, there is strong evidence supporting the alternative hypothesis.

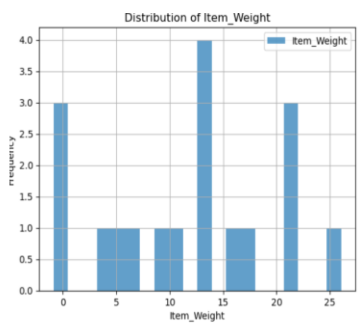
4.The Impact of Item_Weight and Item_Visibility:

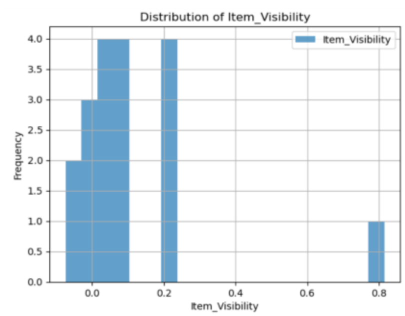
1.Analyze the correlation between Item_Visibility and Item_Outlet_Sales.

- Correlation = 0.021, which is very close to 0. No linear relationship between these two variables.

2.Visualize the relationship using scatter plots with regression line and histograms as below:-







1. **Pearson Correlation Coefficient**: **Value**: 0.0213

    - **Interpretation**: This is a very small positive number, suggesting an almost negligible linear relationship between Item_Weight and Item_Visibility. Essentially, there is no meaningful correlation.

2. **P-value**:
    - **Value**: 0.9331
    - **Interpretation**: The p-value is much greater than the typical threshold (0.05). This indicates that the observed correlation is not statistically significant.

Hypothesis Testing:

- **Null Hypothesis (H_0)**: There is no significant relationship between Item_Weight and Item_Visibility (correlation coefficient = 0).
- **Alternative Hypothesis (H_1)**: There is a significant relationship between Item_Weight and Item_Visibility (correlation coefficient ≠ 0).

Final Answer:

- **Reject the null hypothesis?** No (p-value > 0.05).The p-value (0.9331) is much greater than 0.05, fail to **reject** the null hypothesis.

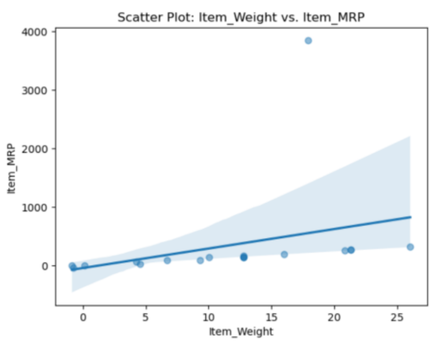
- **Fail to reject the null hypothesis?** Yes. **fail to reject** the null hypothesis. There is no evidence to suggest a significant relationship.
- **Accept the null hypothesis?** No.in hypothesis testing, we **never "accept"** the null hypothesis. We can only fail to reject it, which is what we are doing here.
- **Accept the alternative hypothesis?** No.Since we failed to reject the null hypothesis, we **cannot accept** the alternative hypothesis.

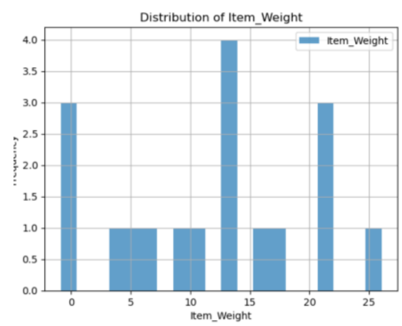
5.Impact of Item_Weight and Item_MRP

1.Analyze the correlation between Item_Visibility and Item_MRP.

➢ Correlation = 0.307, indicating a weak positive relationship. Heavier items might be slightly more expensive.

2.Visualize the relationship using scatter plots with regression line and histograms as below:-





Pearson Correlation Coefficient:**Value**: 0.3070

❖ **I**This indicates a moderate positive linear relationship between Item_Weight and Item_MRP. While not very strong, there is some degree of correlation, meaning that as Item_Weight increases, Item_MRP tends to increase as well.

P-value:**Value**: 0.2153

❖ The p-value is greater than the typical significance level of 0.05, indicating that the result is **not statistically significant**. There isn't enough evidence to claim that the correlation is significantly related to price .

```
from scipy.stats import pearsonr

# Assuming with two coloum
correlation, p_value = pearsonr(dataset['Item_Weight'], dataset['Item_MRP

print(f"Pearson correlation coefficient: {correlation}")
print(f"P-value: {p_value}")
```

```
Pearson correlation coefficient: 0.30700714534558093
P-value: 0.21526838438077486
```

Hypothesis Testing:

- **Null Hypothesis (H_0)**: There is no significant relationship between Item_Weight and Item_MRP (correlation coefficient = 0).
- **Alternative Hypothesis (H_1)**: There is a significant relationship between Item_Weight and Item_MRP (correlation coefficient ≠ 0).Final Answer:

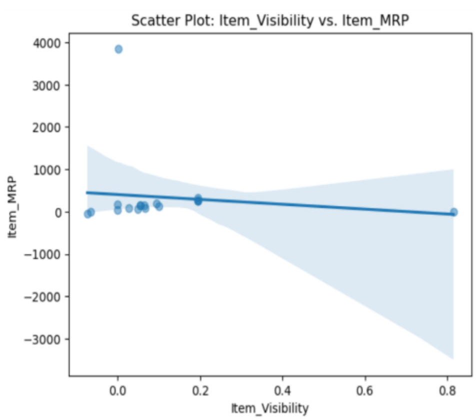
- **Reject the null hypothesis?** No (p-value > 0.05).Since the p-value (0.2153) is greater than 0.05, we **do not reject** the null hypothesis.
- **Fail to reject the null hypothesis?** Yes.We **fail to reject** the null hypothesis because the evidence is insufficient to suggest a significant relationship.
- **Accept the null hypothesis?** No.In hypothesis testing, we **do not accept** the null hypothesis; we only fail to reject it when there is insufficient evidence.
- **Accept the alternative hypothesis?** No.Since we fail to reject the null hypothesis, we **cannot accept** the alternative hypothesis

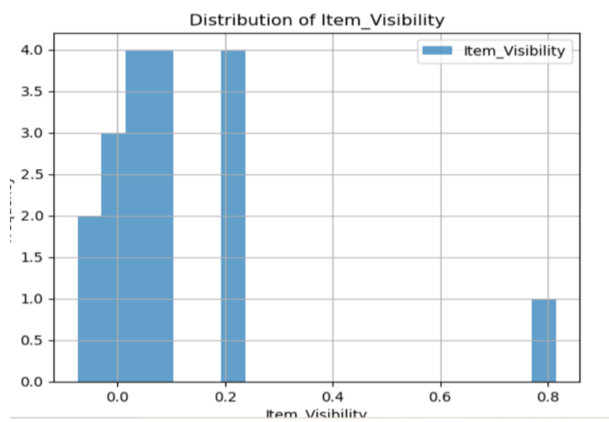
6.Impact on the Item_Visibility and Item_MRP

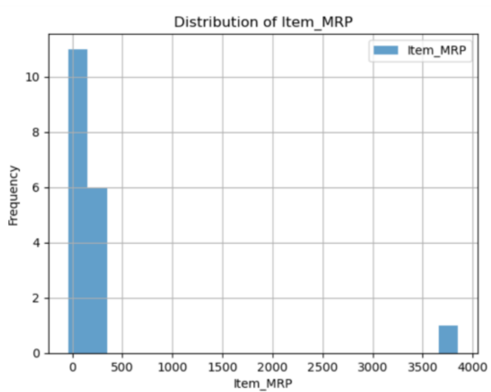
1.Analyze the correlation between Item_Visibility and Item_MRP.

- Correlation = -0.127, indicating a weak negative relationship. More visible items might be slightly less expensive.

2.Visualize the relationship using scatter plots with regression line and histograms as below:-







Pearson Correlation Coefficient:**Value**: -0.1270

- **Interpretation**: This indicates a very weak negative linear relationship between Item_Visibility and Item_MRP. The correlation is close to zero, so it suggests that there is little to no meaningful relationship between the two variables.

P-value:**Value**: 0.6155

- **Interpretation**: The p-value is much greater than 0.05, indicating that the result is **not statistically significant.** There isn't enough evidence to claim that the correlation is significantly different from zero.

```
sns.pairplot(dataset)
```

```
from scipy.stats import pearsonr

# Assuming with two coloum
correlation, p_value = pearsonr(dataset['Item_Visibility'], dataset['Item_MRP'])

print(f"Pearson correlation coefficient: {correlation}")
print(f"P-value: {p_value}")
```

```
Pearson correlation coefficient: -0.12702483145249066
P-value: 0.6154750382791915
```

Hypothesis Testing:

- **Null Hypothesis (H_0)**: There is no significant relationship between Item_Visibility and Item_MRP (correlation coefficient = 0).
- **Alternative Hypothesis (H_1)**: There is a significant relationship between Item_Visibility and Item_MRP (correlation coefficient ≠ 0).

Final Answer:

- **Reject the null hypothesis?** No (p-value > 0.05).Since the p-value (0.6155) is greater than 0.05, we **do not reject** the null hypothesis.
- **Fail to reject the null hypothesis?** Yes.We **fail to reject** the null hypothesis because the evidence is insufficient to suggest a significant relationship.
- **Accept the null hypothesis?** No.We **do not accept** the null hypothesis; we simply fail to reject it.
- **Accept the alternative hypothesis?** No.Since we fail to reject the null hypothesis, we **cannot accept** the alternative hypothesis

Summary of Findings:

1. **Impact of Item Weight on Sales**:
   - ❖ Item weight alone is not a key drive of sales.
2. **Impact of Visibility on Sales**:
   - ❖ Visibility does not have a meaningful impact on sales, and the negative correlation is too weak to be actionable.
   - ❖ Need to improve in this area
3. **Impact of Price (MRP) on Sales**:
   - ❖ Higher-priced items are strongly associated with higher sales, suggesting a robust relationship.
   - ❖ Pricing strategies are good. Ensure premium products are positioned with features and marketing that justify their price.
4. **Impact Between Weight and Visibility**:
   - ❖ There is no meaningful relationship between item weight and visibility. These factors operate independently.

5. **Impact Between Weight and MRP**:

    - ❖ Heavier items tend to have slightly higher prices, but the relationship is weak and not statistically significant.
    - ❖ While there's some association, pricing strategies should rely on additional factors like demand trends, product features, and market competition.

6. **Impact Between Visibility and MRP**:

    - ❖ Visibility does not meaningfully affect item pricing.

Business Implications:

1. **Position High-Price Items Strategically**:

    - Highlight premium features (e.g., durability, uniqueness) in marketing campaigns.
    - Create narratives around high-quality and exclusivity to justify pricing.

2. **Segment Customers**:

    - Target premium-priced items to customers with higher purchasing power.
    - Consider bundling high-priced products with complementary items to increase overall sales value.

3. **Optimize Shelf Placement and Visibility for Premium Products**:

    - Although visibility doesn't correlate strongly with sales, ensuring premium products are well-placed could further boost their performance.

4. **Expand Product Mix**:

    - Introduce more premium-tier items bundle if your customer base responds well to higher-priced offerings.

Your Covariance Matrix Explanation:

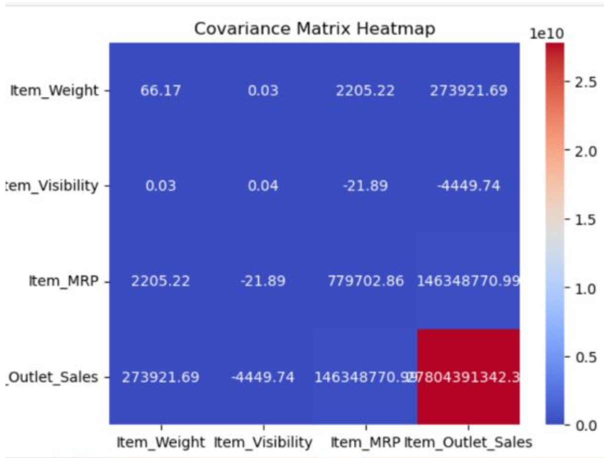
	Item_Weight	Item_Visibility	Item_MRP	Item_Outlet_Sales
Item_Weight	66.172375	0.033830	2.205217e+03	2.739217e+05
Item_Visibility	0.033830	0.038072	-2.188546e+01	-4.449736e+03
Item_MRP	2205.217380	-21.885455	7.797029e+05	1.463488e+08
Item_Outlet_Sales	273921.687772	-4449.736240	1.463488e+08	2.780439e+10

Diagonal Elements (Variance):

- **Item_Weight**: Variance = 66.17 (moderate spread of weight).
- **Item_Visibility**: Variance = 0.0381 (low spread, most items have similar visibility).
- **Item_MRP**: Variance = 779,702.9 (high spread, a wide range of prices).
- **Item_Outlet_Sales**: Variance = 27,804,390,000 (extremely high, indicating large variability in sales).

- **Off-Diagonal Elements** (Covariance):

  - **Item_Weight vs. Item_MRP (2205.22)**: Positive covariance means that **heavier items** tend to be associated with **higher prices**.
  - **Item_Weight vs. Item_Outlet_Sales (273,921.69)**: Positive covariance suggests that **heavier items** are slightly linked to **higher sales**, though the relationship isn't very strong.
  - **Item_Visibility vs. Item_MRP (-21.89)**: Negative covariance implies a very weak inverse relationship between **visibility** and **price**, suggesting visibility doesn't strongly influence pricing decisions.
  - **Item_Visibility vs. Item_Outlet_Sales (-4,449.74)**: Negative covariance shows a weak inverse relationship between **visibility** and **sales**.
  - **Item_MRP vs. Item_Outlet_Sales (146,348,800)**: Strong positive covariance, indicating that **higher-priced items** tend to have **higher sales**. This is a strong indicator that **pricing** plays a critical role in sales performance.



.

1. **Positive Covariance**:

❖ When two variables have a positive covariance, they tend to increase or decrease together. For example, if **Item_Weight** and **Item_MRP** have a positive covariance, heavier items tend to have higher prices.

2. **Negative Covariance**:

❖ A negative covariance means that as one variable increases, the other tends to decrease. For instance, if **Item_Visibility** and **Item_Outlet_Sales** have negative covariance, as the visibility of an item increases, its sales may decrease (though this relationship could be weak).