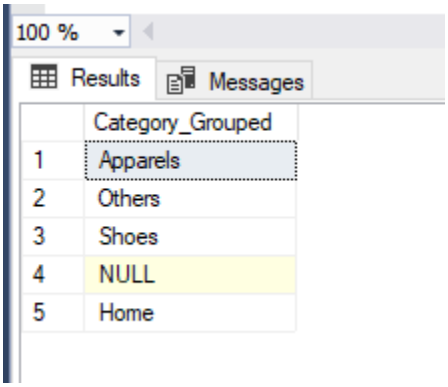
PAYTM MALL EPURCHASE DATA ANALYSIS ASSESSMENT QUESTIONS

1. **What does the "Category_Grouped" column represent, and how many unique categories are there?**

```
SELECT DISTINCT Category_Grouped
FROM Purchase_data$;
```

OBTAINED RESULT:



100 %

Results | Messages

	Category_Grouped
1	Apparels
2	Others
3	Shoes
4	NULL
5	Home

2. **List the top 5 shipping cities in terms of the number of orders.**

```
SELECT TOP 5
    [Shipping_city],
    COUNT(*) AS Order_Count
FROM
    Purchase_data$
GROUP BY
    [Shipping_city]
ORDER BY
    Order_Count DESC;
```

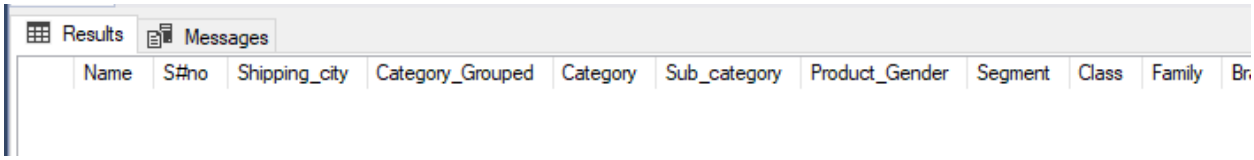
OBTAINED RESULT:

	Shipping_city	Order_Count
1	New Delhi	4560
2	Chennai	4254
3	Bangalore	3974
4	Mumbai	3159
5	Hyderabad	2849

3. Show me a table with all the data for products that belong to the "Electronics" category.

```
SELECT *
FROM Purchase_data$
WHERE [Category_Grouped] = 'Electronics';
```

OBTAINED RESULT:



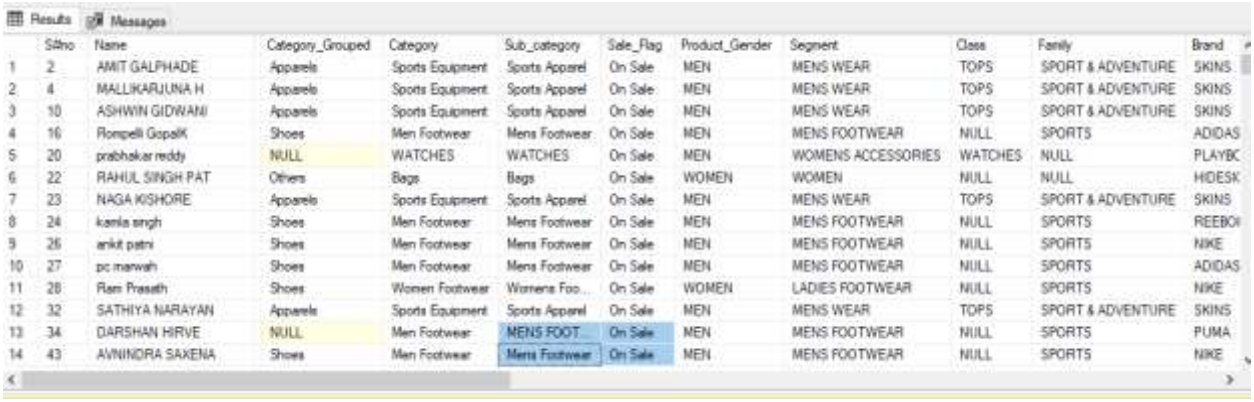
Results | Messages

Name	S#no	Shipping_city	Category_Grouped	Category	Sub_category	Product_Gender	Segment	Class	Family	Br

4. Filter the data to show only rows with a "Sale_Flag" of 'Yes'.

```
SELECT *
FROM Purchase_data$
WHERE [Sale_Flag] = 'On Sale';
```

OBTAINED RESULT:



Results | Messages

	S#no	Name	Category_Grouped	Category	Sub_category	Sale_Flag	Product_Gender	Segment	Class	Family	Brand
1	2	AMIT GALPHADE	Apparels	Sports Equipment	Sports Apparel	On Sale	MEN	MENS WEAR	TOPS	SPORT & ADVENTURE	SKINS
2	4	MALLIKARJUNA H	Apparels	Sports Equipment	Sports Apparel	On Sale	MEN	MENS WEAR	TOPS	SPORT & ADVENTURE	SKINS
3	10	ASHWIN GIDWANI	Apparels	Sports Equipment	Sports Apparel	On Sale	MEN	MENS WEAR	TOPS	SPORT & ADVENTURE	SKINS
4	16	Rompelli GopalK	Shoes	Men Footwear	Mens Footwear	On Sale	MEN	MENS FOOTWEAR	NULL	SPORTS	ADIDAS
5	20	prabhakar reddy	NULL	WATCHES	WATCHES	On Sale	MEN	WOMENS ACCESSORIES	WATCHES	NULL	PLAYBC
6	22	RAHUL SINGH PAT	Others	Bags	Bags	On Sale	WOMEN	WOMEN	NULL	NULL	HIDESK
7	23	NAGA KISHORE	Apparels	Sports Equipment	Sports Apparel	On Sale	MEN	MENS WEAR	TOPS	SPORT & ADVENTURE	SKINS
8	24	kamla singh	Shoes	Men Footwear	Mens Footwear	On Sale	MEN	MENS FOOTWEAR	NULL	SPORTS	REEBOI
9	26	ankit patni	Shoes	Men Footwear	Mens Footwear	On Sale	MEN	MENS FOOTWEAR	NULL	SPORTS	NIKE
10	27	pc marwah	Shoes	Men Footwear	Mens Footwear	On Sale	MEN	MENS FOOTWEAR	NULL	SPORTS	ADIDAS
11	28	Ram Prasath	Shoes	Women Footwear	Womens Foo...	On Sale	WOMEN	LADIES FOOTWEAR	NULL	SPORTS	NIKE
12	32	SATHIYA NARAYAN	Apparels	Sports Equipment	Sports Apparel	On Sale	MEN	MENS WEAR	TOPS	SPORT & ADVENTURE	SKINS
13	34	DARSHAN HIRVE	NULL	Men Footwear	MENS FOOT...	On Sale	MEN	MENS FOOTWEAR	NULL	SPORTS	PUMA
14	43	AVNINDRA SAXENA	Shoes	Men Footwear	Mens Footwear	On Sale	MEN	MENS FOOTWEAR	NULL	SPORTS	NIKE

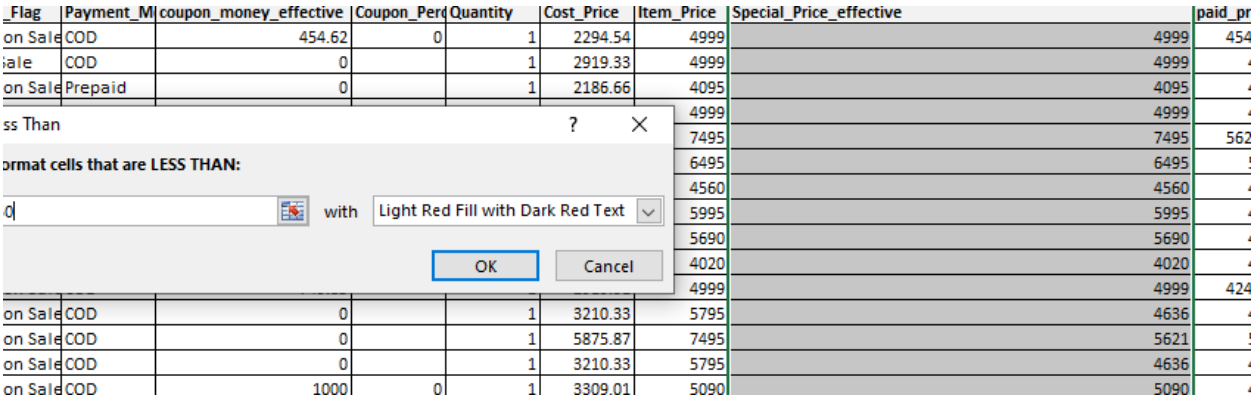
5. Sort the data by "Item_Price" in descending order. What is the most expensive item?

```
SELECT * FROM  Purchase_data$
order by Item_Price desc;
```

Results | Messages

	S#no	Name	Shipping_city	Category_Grouped	Category	Sub_category	Product_Gender	Segment	Class	Family	Item_Price	Branc
21	2155	vaidyanathan ve	Coimbatore	NULL	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	ADIC
22	1564	Pavan Kumar Mun	Chennai	NULL	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	ADIC
23	1579	shashi chavali	Hyderabad	Shoes	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	PUM
24	1595	Gonuguntla Gout	Secunderabad	Shoes	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	PUM
25	1692	MUKESH AGARWAL	Lucknow	NULL	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	ADIC
26	1709	Rajesh Lal Das	Jamshedpur	NULL	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	ADIC
27	1711	Pramod Verma	Jabalpur	NULL	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	ADIC
28	1853	ajay CH	Bangalore	NULL	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	ADIC
29	2558	GAUTAM KATHUR...	Rudrapur	Shoes	Men Footwear	Mens Footwear	MEN	MENS FOOTWEAR	NULL	SPORTS	13500	PUM
30	2617	manas mondal	berhampur	NULL	Furniture	LIVING	UNISEX	LIVING	NULL	TABLES	13500	TRU

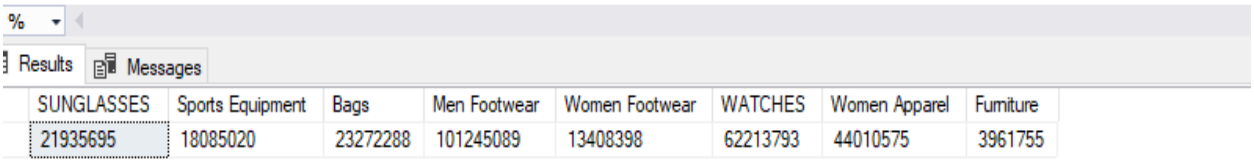
6. Apply conditional formatting to highlight all products with a "Special_Price_effective" value below $50 in red.



_Flag	Payment_M	coupon_money_effective	Coupon_Per	Quantity	Cost_Price	Item_Price	Special_Price_effective	paid_pr
on Sale	COD	454.62	0	1	2294.54	4999	4999	454
ale	COD	0		1	2919.33	4999	4999	
on Sale	Prepaid	0		1	2186.66	4095	4095	
						4999	4999	
						7495	7495	562
						6495	6495	
						4560	4560	
						5995	5995	
						5690	5690	
						4020	4020	
						4999	4999	424
on Sale	COD	0		1	3210.33	5795	4636	
on Sale	COD	0		1	5875.87	7495	5621	
on Sale	COD	0		1	3210.33	5795	4636	
on Sale	COD	1000	0	1	3309.01	5090	5090	

7. Create a pivot table to find the total sales value for each category.

```
SELECT
    SUM(CASE WHEN Category = 'SUNGLASSES' THEN Item_Price ELSE 0 END) AS SUNGLASSES,
    SUM(CASE WHEN Category = 'Sports Equipment' THEN Item_Price ELSE 0 END) AS [Sports Equipment],
    SUM(CASE WHEN Category = 'Bags' THEN Item_Price ELSE 0 END) AS Bags,
    SUM(CASE WHEN Category = 'Men Footwear' THEN Item_Price ELSE 0 END) AS [Men Footwear],
    SUM(CASE WHEN Category = 'Women Footwear' THEN Item_Price ELSE 0 END) AS [Women Footwear],
    SUM(CASE WHEN Category = 'WATCHES' THEN Item_Price ELSE 0 END) AS WATCHES,
    SUM(CASE WHEN Category = 'Women Apparel' THEN Item_Price ELSE 0 END) AS [Women Apparel],
    SUM(CASE WHEN Category = 'Furniture' THEN Item_Price ELSE 0 END) AS Furniture
FROM
    [dbo].[Purchase_data$];
```

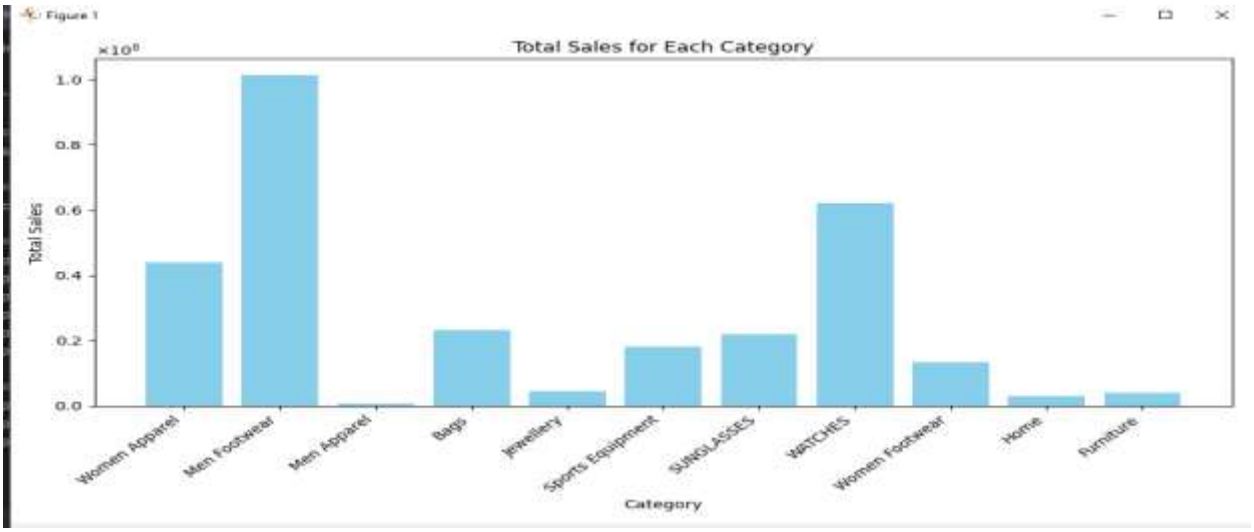


SUNGLASSES	Sports Equipment	Bags	Men Footwear	Women Footwear	WATCHES	Women Apparel	Furniture
21935695	18085020	23272288	101245089	13408398	62213793	44010575	3961755

8. Create a bar chart to visualize the total sales for each category.

```python
import matplotlib.pyplot as plt
import pandas as pd
import pyodbc
from matplotlib.ticker import ScalarFormatter

server_name = 'localhost\\SQLEXPRESS01'
database_name = 'ita'
trusted_connection = 'yes'

connection_string = f'DRIVER={{SQL Server}};SERVER={server_name};DATABASE={database_name};Trusted_Connection={trusted_connection};'

connection = pyodbc.connect(connection_string)

sql_query = """
SELECT
    Category,
    SUM(Item_Price) AS Total_Sales
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Category;
"""

sales_df = pd.read_sql(sql_query, connection)
```

```python
connection.close()

plt.figure(figsize=(10, 6))
plt.bar(sales_df['Category'], sales_df['Total_Sales'], color='skyblue')
plt.xlabel('Category')
plt.ylabel('Total Sales')
plt.title('Total Sales for Each Category')
plt.xticks(rotation=45, ha='right')

plt.gca().yaxis.set_major_formatter(ScalarFormatter(useMathText=True))
plt.tight_layout()
plt.show()
```



9. Calculate the average "Quantity" sold for products in the "Clothing" category, grouped by "ProductGender”

```
SELECT
    Product_Gender,
    AVG(Quantity) AS Average_Quantity
FROM
    [llk].[dbo].[Purchase_data$]
WHERE
    Category_Grouped = 'Apparels'
GROUP BY
    Product_Gender;
```

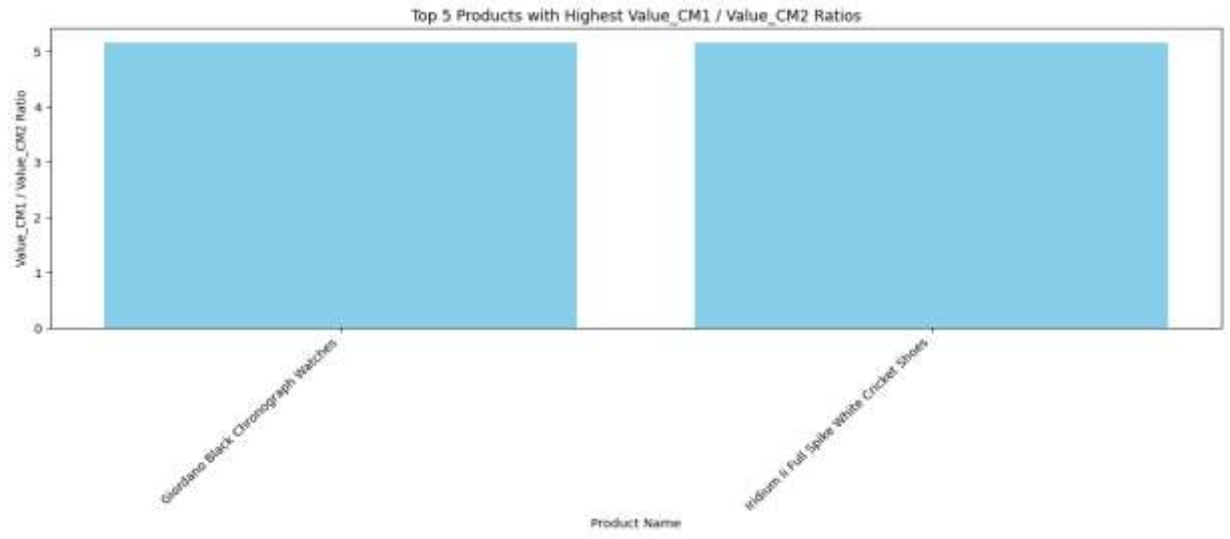
100 %

Results | Messages

	Product_Gender	Average_Quantity
1	WOMEN	1
2	MEN	1

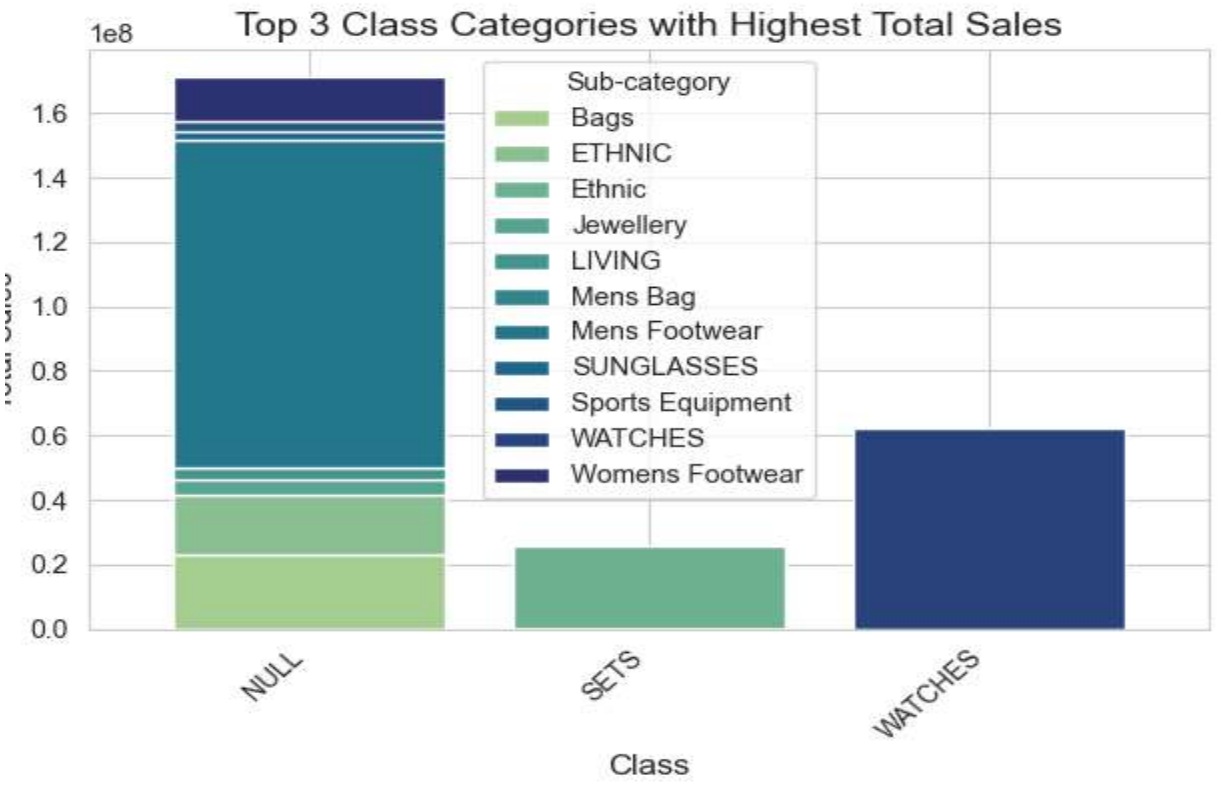
10. Find the top 5 products with the highest "Value_CM1" and "Value_CM2" ratios. Create a chart to visualize this data.

```
sql_query = """
SELECT TOP 5
    Item_NM,
    Value_CM1 / NULLIF(Value_CM2, 0) AS Ratio
FROM
    [dbo].[Purchase_data$]
WHERE
    Value_CM2 <> 0
ORDER BY
 Ratio DESC;
"""
top_products_df = pd.read_sql(sql_query, connection)

connection.close()

# Plotting the bar chart
plt.figure(figsize=(10, 6))
plt.bar(top_products_df['Item_NM'], top_products_df['Ratio'], color='skyblue')
plt.xlabel('Product Name')
plt.ylabel('Value_CM1 / Value_CM2 Ratio')
plt.title('Top 5 Products with Highest Value_CM1 / Value_CM2 Ratios')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```



11. Identify the top 3 "Class" categories with the highest total sales. Create a stacked bar chart to represent this data.

```
sql_query = """
SELECT
    Class,
    Sub_category,
    SUM(Item_Price) AS Total_Sales
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Class, Sub_category;
"""
sales_df = pd.read_sql(sql_query, connection)
class_sales = sales_df.groupby('Class')['Total_Sales'].sum().sort_values(ascending=False)
top_3_classes = class_sales.head(3).index.tolist()
top_3_sales = sales_df[sales_df['Class'].isin(top_3_classes)]
pivot_table = top_3_sales.pivot_table(index='Class', columns='Sub_category', values='Total_Sales', aggfunc='sum', fill_value=0)

plt.figure(figsize=(10, 6))
pivot_table.plot(kind='bar', stacked=True, colormap='crest', width=0.8)
plt.xlabel('Class', fontsize=12)
plt.ylabel('Total Sales', fontsize=12)
plt.title('Top 3 Class Categories with Highest Total Sales', fontsize=14)
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.yticks(fontsize=10)
plt.legend(title='Sub-category', fontsize=10)
plt.tight_layout()
plt.show()
```



12. Find the total sales for each "Brand" and display the top 3 brands in terms of sales.

```
SELECT Top 3
    Brand,
    SUM(Item_Price) AS Total_Sales
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Brand
ORDER BY
    Total_Sales DESC
```

Results | Messages

	Brand	Total_Sales
1	NIKE	43406505
2	SANGRIA	22614700
3	PUMA	20832027

13. Calculate the total revenue generated from "Electronics" category products with a "Sale_Flag" of 'Yes'.

```
SELECT
    SUM(Item_Price) AS Total_Revenue
FROM
    [dbo].[Purchase_data$]
WHERE
    Category = 'Electronics'
    AND Sale_Flag = 'Yes';
```

100 %

Results | Messages

	Total_Revenue
1	NULL

14.Identify the top 5 shipping cities based on the average order value (total sales amount divided by the number of orders) and display their average order values

```
SELECT Top 5
    Shipping_city,
    SUM(Item_Price) / COUNT(*) AS Average_Order_Value
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Shipping_city
ORDER BY
    Average_Order_Value DESC
```

100 %

Results | Messages

	Shipping_city	Average_Order_Value
1	Dalhousie	13500
2	Panagudi	13500
3	Vallabh Vidyanagar	13500
4	laxminagar	13500
5	Seoni Malwa	13500

15.Determine the total number of orders and the total sales amount for each "Product_Gender" within the "Clothing" category.

```
SELECT
    Product_Gender,
    COUNT(*) AS Total_Number_of_Orders,
    SUM(Item_Price) AS Total_Sales_Amount
FROM
    [dbo].[Purchase_data$]
WHERE
    Category_Grouped = 'Apparels'
GROUP BY
    Product_Gender;
```

Results | Messages

Product_Gender	Total_Number_of_Orders	Total_Sales_Amount
WOMEN	6426	41169012
MEN	3087	15039825

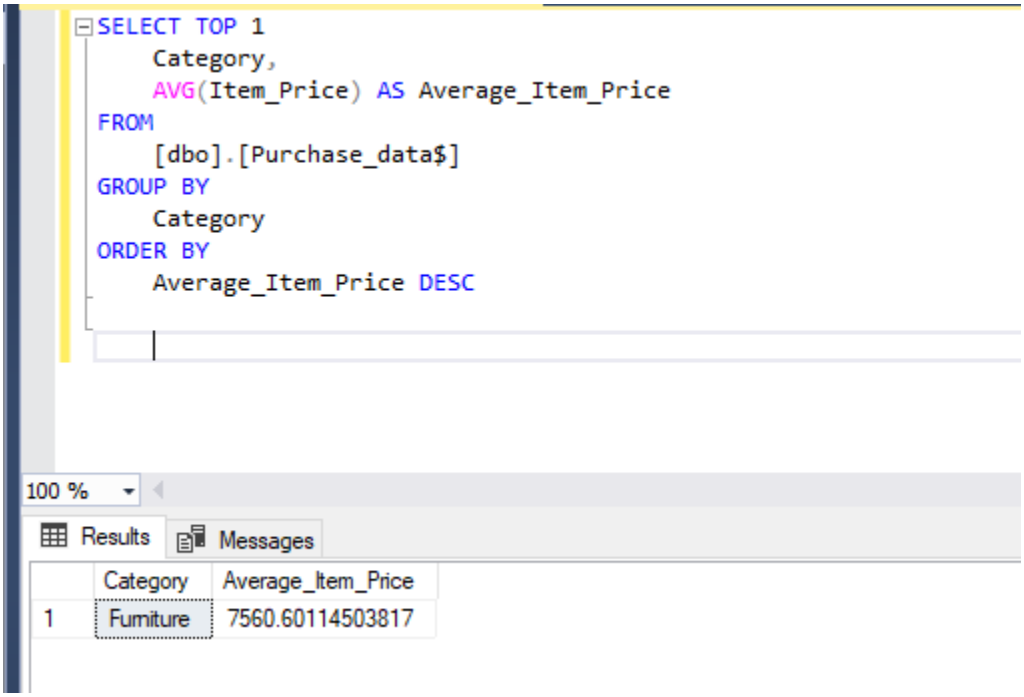
16. Calculate the percentage contribution of each "Category" to the overall total sales.

```
SELECT
    Category,
    SUM(Item_Price) / (SELECT SUM(Item_Price) FROM [dbo].[Purchase_data$]) * 100 AS Percentage_Contribution
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Category;
```

Results | Messages

	Category	Percentage_Contribution
1	Women Apparel	14.8455453681191
2	Men Footwear	34.1517592544237
3	Men Apparel	0.243983569687227
4	Bags	7.85015436230803
5	Jewellery	1.53456171526961
6	Sports Equipment	6.10039711804134
7	SUNGLASSES	7.3992979029182
8	WATCHES	20.9858127621435
9	Women Footwear	4.52288980143518

17.Identify the "Category" with the highest average "Item_Price" and its corresponding average price.



```
SELECT TOP 1
    Category,
    AVG(Item_Price) AS Average_Item_Price
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Category
ORDER BY
    Average_Item_Price DESC
```

100 %

Results | Messages

	Category	Average_Item_Price
1	Furniture	7560.60114503817

18. Find the month with the highest total sales revenue.

There is no details about data

19. Calculate the total sales for each "Segment" and the average quantity sold per order for each segment.

```
SELECT
    Segment,
    SUM(Item_Price) AS Total_Sales,
    AVG(Quantity) AS Average_Quantity_Per_Order
FROM
    [dbo].[Purchase_data$]
GROUP BY
    Segment;
```

100 %

Results | Messages

	Segment	Total_Sales	Average_Quantity_Per_Order
1	MENS ACCESSORIES	13279676	1
2	OUTDOOR & HIKING	3045195	1
3	MENS WEAR	15039825	1
4	WOMENS ACCESSORIES	7386940	1.01057501652346
5	UNISEX	15519256	1
6	LADIES FOOTWEAR	13408398	1
7	WOMEN	30013131	1.00186760738742
8	WOMENS JEWELLERY	4549307	1.00498338870432
9	MENS APPARELS	723305	1