



## **Pharma Data Analysis**

**1. Retrieve all columns for all records in the dataset.**

Select \* from pharma\_data\_analysis;

**2. How many unique countries are represented in the dataset?**

Select count(distinct Country) as no\_of\_unique\_countries from  
pharma\_data\_analysis;

**3. Select the names of all the customers on the 'Retail' channel.**

Select customer\_name from pharma\_data\_analysis where  
sub\_channel="retail";

**4. Find the total quantity sold for the 'Antibiotics' product class.**

Select sum(quantity) as total\_quantity from pharma\_data\_analysis  
Where product\_class="Antibiotics";

**5. List all the distinct months present in the dataset.**

Select distinct month from pharma\_data\_analysis;

**6. Calculate the total sales for each year.**

Select Year,sum(Sales) as total\_sales

From pharma\_data\_sales

Group by year;

**7. Find the customer with the highest sales value.**

Select customer\_name

From pharma\_data\_analysis

Order by sales desc

Limit 1;

**8. Get the names of all employees who are Sales Reps and are managed by 'James Goodwill'.**

Select customer\_name from pharma\_data\_analysis

Where Manager="James Goodwill";

**9. Retrieve the top 5 cities with the highest sales.**

Select city from pharma\_data\_analysis

Order by Sales desc

Limit 5;

**10. Calculate the average price of products in each sub-channel.**

Select Sub-channel,avg(Price) as avg\_price

From pharma\_data\_analysis Group by Sub-channel;

**11. Join the 'Employees' table with the 'Sales' table to get the name of the Sales Rep and the corresponding sales records.**

Select `Name of the Sales Rep`

From Sales

Inner join Employees

On Employees.EmployeeID=Sales.EmployeeID;

**12. Retrieve all sales made by employees from ' Rendsburg ' in the year 2018.**

Select \* from pharma\_data\_analysis

Where City=" Rendsburg";

**13. Calculate the total sales for each product class, for each month, and order the results by year, month, and product class.**

Select Product class,Month,Year,sum(Sales) as total\_sales

From pharma\_data\_analysis

Order by Product class,Month,Year;

**14. Find the top 3 sales reps with the highest sales in 2019.**

Select Name\_of\_sales\_rep

From pharma\_data\_analysis

Where Year=2019

Order by Sales desc

Limit 3;

**15. Calculate the monthly total sales for each sub-channel, and then calculate the average monthly sales for each sub-channel over the years.**

Select Sub-channel,Month,sum(Sales) as total\_sales

From pharma\_data\_analysis

Group by Sub-channel,Month;

Select Sub-channel,Year,Month,avg(Sales) as total\_sales

From pharma\_data\_analysis

Group by Sub-channel,Month,Year;

**16. Create a summary report that includes the total sales, average price, and total quantity sold for each product class.**

Select Product Class,sum(Sales) as total\_sales,avg(Price) as  
avg\_price,sum(Quantity) as total\_quantity

From pharma\_data\_analysis

Group by Product Class;

**17. Find the top 5 customers with the highest sales for each year.**

Select Customer\_name ,year,max(Sales) as highest\_sales

From pharma\_data\_analysis

Group by year

Limit 5;

**18. Calculate the year-over-year growth in sales for each country. 2 of 2**

Select country, [(select sum(sales) from pharma\_data\_analysis where year=2020) - (select sum(sales) from pharma\_data\_analysis where year=2019)]/ (select sum(sales) from pharma\_data\_analysis where year=2019)

From pharma\_data\_analysis

Group by Country;

**19. List the months with the lowest sales for each year**

Select Month

From (Select Year,min(Sales) as min\_sales

From pharma\_data\_analysis Group by Year);

**20. Calculate the total sales for each sub-channel in each country, and then find the country with the highest total sales for each sub-channel**

With Sub\_channelsale as(

Select Sub\_channel,Country,sum(Sales) as total\_sales

From pharma\_data\_analysis

Group by Sub\_channel,Country

)

Select Sub\_channel,country,total\_sales

From

(select Sub\_channel,Country,total\_sales

From Sub\_channelsale

Order by total\_sales desc Limit 1);





## **HR Data Analysis**

**1. Retrieve the total number of employees in the dataset.**

```
select count(distinct EmployeeID) from `general data`;
```

**2. List all unique job roles in the dataset.**

```
select distinct JobRole from `general data`;
```

**3. Find the average age of employees.**

```
select avg(Age) as avg_age from `general data`;
```

**4. Retrieve the names and ages of employees who have worked at the company for more than 5 years.**

```
select `Emp Name`, age from `general data`
```

```
where YearsAtCompany>5;
```

**5. Get a count of employees grouped by their department.**

```
select Department,count(distinct EmployeeID) as Emp_count
```

```
from `general data` group by Department;
```

**6. List employees who have 'High' Job Satisfaction.**

```
select `Emp Name` from `general data`  
  
inner join employee_survey_data  
  
on `general data`.EmployeeId=employee_survey_data.EmployeeId where  
JobSatisfaction=3;
```

**7. Find the highest Monthly Income in the dataset.**

```
select max(MonthlyIncome) as highest_monthly_income from `general data`;
```

**8. List employees who have 'Travel\_Rarely' as their BusinessTravel type.**

```
select `Emp Name` from `general data`  
  
where BusinessTravel="Travel_Rarely";
```

**9. Retrieve the distinct MaritalStatus categories in the dataset.**

```
select distinct MaritalStatus from `general data`;
```

**10. Get a list of employees with more than 2 years of work experience but less than 4 years in their current role.**

```
select `Emp Name` from `general data`  
  
where TotalWorkingYears>2 and YearsSinceLastPromotion<4;
```



**11. List employees who have changed their job roles within the company (JobLevel and JobRole differ from their previous job).**

```
select e1.`Emp Name`,e1.JobRole from `general data` e1,`general data` e2
where e1.JobRole!=e2.JobRole;
```

**12. Find the average distance from home for employees in each department.**

```
select Department,avg(DistanceFromHome) as avg_distance
from `general data`
group by Department;
```

**13. Retrieve the top 5 employees with the highest MonthlyIncome.**

```
select `Emp Name`,MonthlyIncome
from `general data`
order by MonthlyIncome desc
limit 5;
```

**14. Calculate the percentage of employees who have had a promotion in the last year.**

```
select (select count(*) from `general data`where
YearsSinceLastPromotion=1)*100/count(*)
from `general data`;
```

**15. List the employees with the highest and lowest EnvironmentSatisfaction.**

```
select `Emp Name`  
from `general data`  
inner join employee_survey_data  
on `general data`.EmployeeID=employee_survey_data.EmployeeID  
where employee_survey_data.EnvironmentSatisfaction=1  
  
union  
  
select `Emp Name`  
from `general data`  
inner join employee_survey_data  
on `general data`.EmployeeID=employee_survey_data.EmployeeID  
where employee_survey_data.EnvironmentSatisfaction=4;
```

**16. Find the employees who have the same JobRole and MaritalStatus.**

```
select * from `general data`  
where JobRole=MaritalStatus;
```

**17. List the employees with the highest TotalWorkingYears who also have a PerformanceRating of 4.**

```
select * from `general data`  
where PerformanceRating=4  
order by TotalWorkingYears Desc;
```

**18. Calculate the average Age and JobSatisfaction for each BusinessTravel type. 2 of 2**

```
select g.BusinessTravel,avg(g.age) as avg_age,e.JobSatisfaction
from `general data` g
inner join employee_survey_data e
on g.EmployeeID=e.EmployeeID
group by g.BusinessTravel,e.JobSatisfaction;
```

**19. Retrieve the most common EducationField among employees.**

```
select EducationField,count(*) as most_common
from `general data`
group by EducationField
order by most_common desc
limit 1;
```

**20. List the employees who have worked for the company the longest but haven't had a promotion.**

```
select *
from `general data`
where YearsSinceLastPromotion=0
order by YearsAtCompany Desc;
```



## Paytm Mall Epurchase Data Analysis

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

```
select count(Distinct Category_grouped) as unique_categories from paytm epurchase data;
```

**2. Can you list the top 5 shipping cities in terms of the number of orders?**

```
select Shipping_city, count(distinct Name) as no_of_orders from paytm epurchase data
```

```
group by Shipping_city
```

```
order by no_of_orders desc limit 5;
```

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

```
select * from paytm epurchase data where Category="electronics";
```

**4. Filter the data to show only rows with a "Sale\_Flag" of 'Yes'.**

```
select * from paytm epurchase data where sale_flag="on sale";
```

**5. Sort the data by "Item\_Price" in descending order. What is the most expensive item?**

```
select item_nm from paytm epurchase data order by item_price desc;
```

**6. Apply conditional formatting to highlight all products with a "Special\_Price\_effective" value below \$50 in red.**

```
select * from paytm epurchase data where Special_Price_effective < 4157 and color="red";
```

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

```
select category,sum(item_price) as total_sales from paytm epurchase data group by category;
```

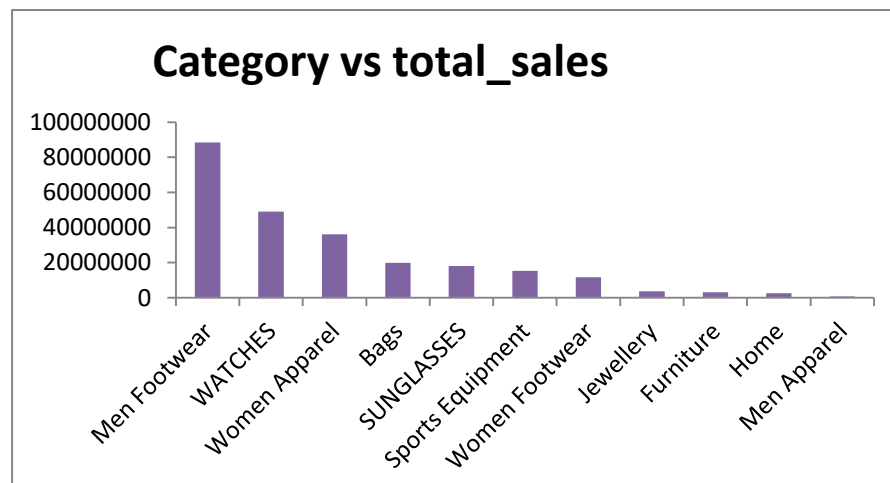
Row Labels	Sum of Item_Price
Bags	23272288
Furniture	3961755
Home	3051213
Jewellery	4549307
Men Apparel	723305
Men Footwear	101245089
Sports Equipment	18085020
SUNGLASSES	21935695
WATCHES	62213793
Women Apparel	44010575
Women Footwear	13408398
<b>Grand Total</b>	<b>296456438</b>

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

```
select category,sum(item_price) as total_sales
```

```
from paytm epurchase data
```

```
group by category;
```

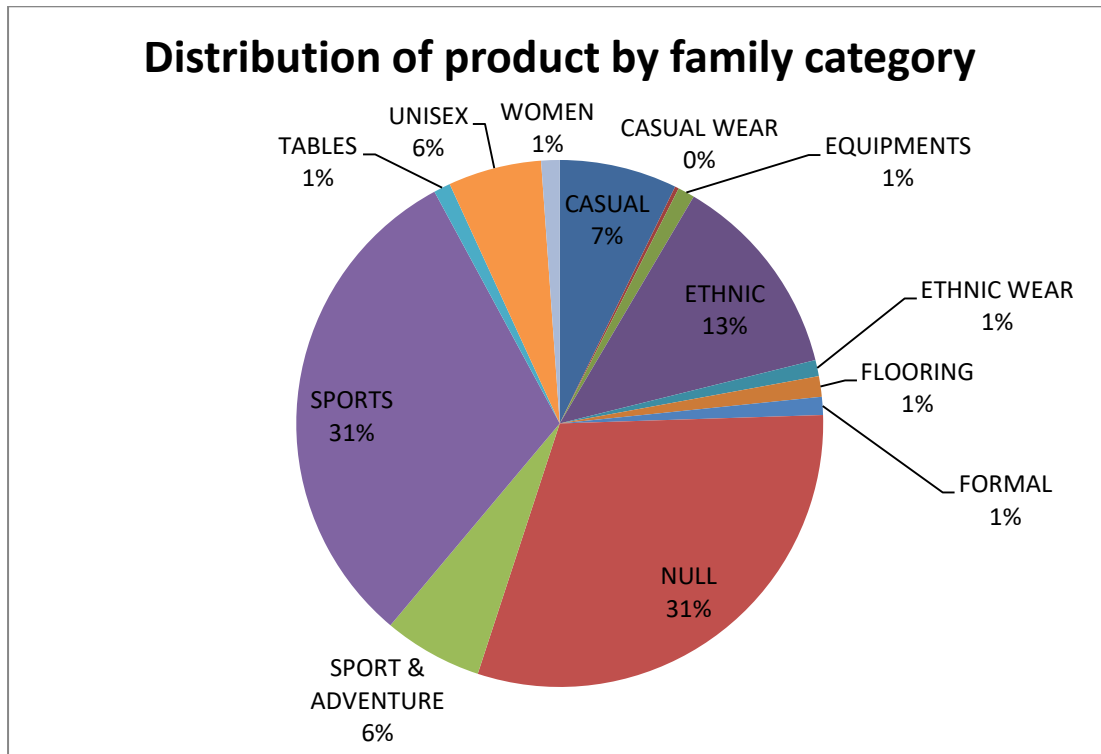


9. Create a pie chart to show the distribution of products in the "Family" category.

```
select Family,count(distinct item_nm)
```

```
from paytm epurchase data
```

```
group by Family;
```



10. Ensure that the "Payment\_Method" column only contains valid payment methods (e.g., Visa, MasterCard).

```
select * from paytm epurchase data where Payment_method in ("COD","Prepaid");
```

11. Calculate the average "Quantity" sold for products in the "Clothing" category, grouped by Product\_Gender."

```
select Product_Gender,avg(Quantity) as avg_qua
```

```
from paytm epurchase data
```

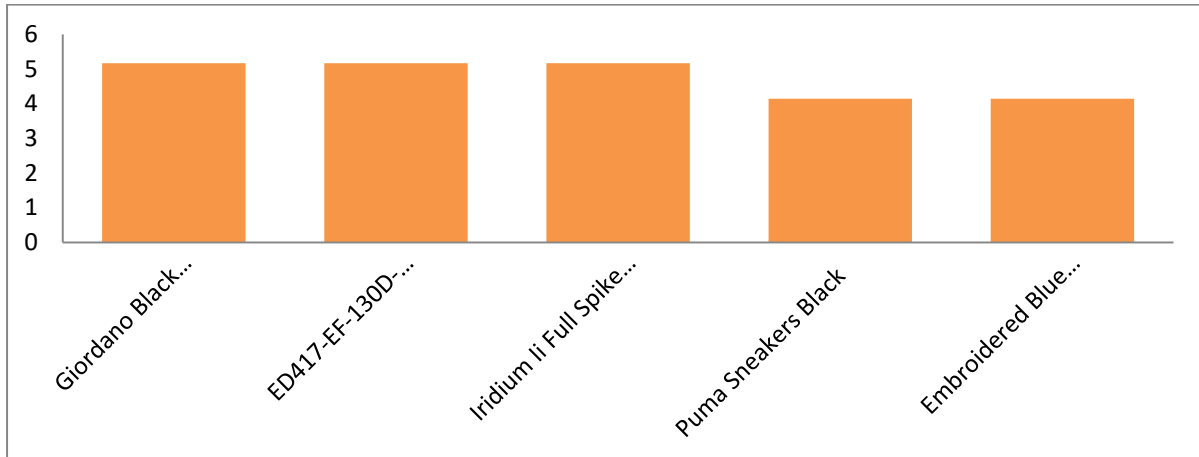
```
group by Product_Gender;
```

12. Find the top 5 products with the highest "Value\_CM1" and "Value\_CM2" ratios. Create a chart to visualize this data.

```
select item_nm,max(value_cm1/value_cm2) as ratio
```

```
from paytm epurchase data
```

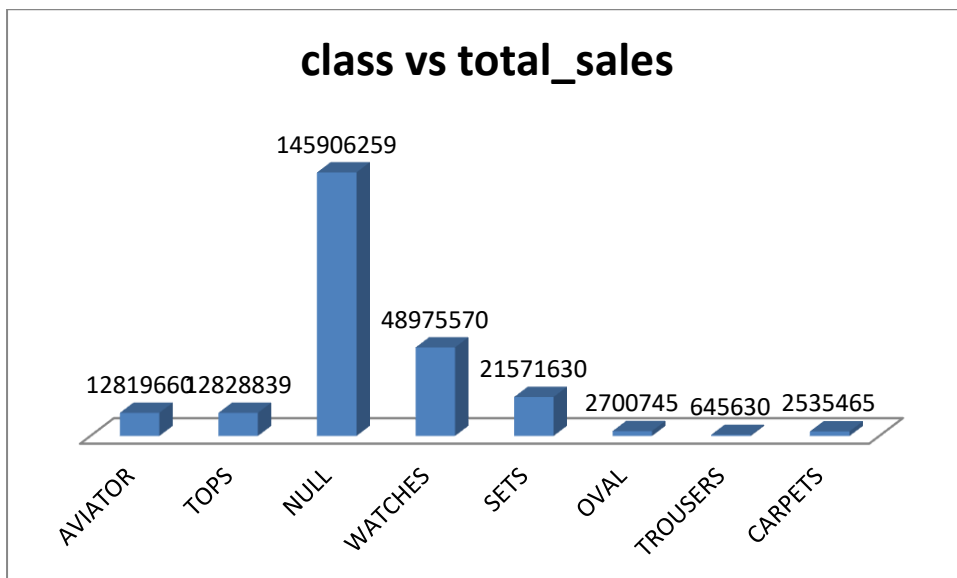
```
group by item_nm order by ratio desc limit 5;
```



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

```
select class,sum(item_price) as total_sales
```

```
from paytm epurchase data group by class;
```



**14. Use VLOOKUP or INDEX-MATCH to retrieve the "Color" of a product with a specific "Item\_NM."**

```
select distinct color  
from paytm epurchase data  
where item_nm="Tan Boots";
```

**15. Calculate the total "coupon\_money\_effective" and "Coupon\_Percentage" for products in the "Electronics" category.**

```
select sum(coupon_money_effective),sum(Coupon_Percentage)  
from paytm epurchase data  
where category="electronics";
```

**16. Perform a time series analysis to identify the month with the highest total sales.**

```
select extract(month from date) as month,sum(item_price)  
from paytm epurchase data  
group by month;
```

**17. Calculate the total sales for each "Segment" and create a scatter plot to visualize the relationship between "Item\_Price" and "Quantity" in this data.**

```
select segment,sum(item_price) as total_sales  
from paytm epurchase data  
group by segment;
```

**18. Use the AVERAGEIFS function to find the average "Item\_Price" for products that have a "Sale\_Flag" of 'Yes.'**

```
select avg(item_price) as avg_price  
from paytm epurchase data  
where Sale_Flag="on sale";
```

**19. Identify products with a "Paid\_pr" higher than the average in their respective "Family" and "Brand" groups.**

```
select * from paytm epurchase data  
where Paid_pr > ALL( select avg(Paid_pr) from paytm epurchase data group by brand,family);
```



**20. Create a pivot table to show the total sales for each "Color" within the "Clothing" category and use conditional formatting to highlight the highest sales.**

```
select color,sum(item_price) as total_sales
```

```
from paytm epurchase data
```

```
group by color;
```