Contents

[Project Documentation for Vrinda Store Sales Data Analysis 2](#_Toc151713350)

[1.0 Problem Statement 2](#_Toc151713351)

[2.0 Sample Questions 2](#_Toc151713352)

[3.0 Gathering Data 2](#_Toc151713353)

[4.0 Data Cleaning 2](#_Toc151713354)

[5.0 Data Processing 3](#_Toc151713355)

[6.0 Data Analysis 3](#_Toc151713356)

[6.0.1 Show sum of amount and count of orders by month 3](#_Toc151713357)

[6.0.2 Which month got the highest sales and orders? 4](#_Toc151713358)

[6.0.3 Who purchased more- men or women in 2022? 4](#_Toc151713359)

[6.0.4 What are different order status in 2022? 4](#_Toc151713360)

[6.0.5 List top 10 states contributing to the sales? 5](#_Toc151713361)

[6.0.6 Relation between age and gender based on number of orders. 5](#_Toc151713362)

[6.0.7 Which channel is contributing to maximum sales? 6](#_Toc151713363)

[7.0 Final Report 6](#_Toc151713364)

[8.0 Sample Insights 7](#_Toc151713365)

[9.0 Final Conclusion to improve Vrinda Store sales: 7](#_Toc151713366)

# Project Documentation for Vrinda Store Sales Data Analysis

## Problem Statement

Vrinda Store wants to create an annual sales report for 2022. So that, Vrinda can understand their customers and grow more sales in 2023.

## 2.0 Sample Questions

1. Show number of orders and sum of amount sold in same diagram
2. Which month got the highest sales and orders?
3. Who purchased more- men or women in 2022?
4. What are different order status in 2022?
5. List top 10 states contributing to the sales?
6. Relation between age and gender based on number of orders.
7. Which channel is contributing to maximum sales?
8. Highest selling category?

## 3.0 Gathering Data

Provided by Vrinda Store in form of excel data. So there is no need to build ETL pipelines to get data.

## 4.0 Data Cleaning

1. First, apply filters in all columns to see inconsistencies and null values in the columns. Typically used to check spelling errors in categories mostly.
2. Found problem in gender column i.e. In some rows the representation for Men is M and in some there is full length word i.e., Men itself and vice versa for women. Now to solve this issue, we choose an approach i.e., Select only M from filters, select whole column -> then Find and Replace then Find (M) and replace with (Men). Ctrl + F for find and replace modal.
3. Next issue, is in Quantity column there number of quantity is represented as some in numeric and some in alphabet e.g. 1 and one. Cleaning approach is select column Find and replace then One to 1.

## 5.0 Data Processing

* Added New column F, with name Age group and calculated teenage, adult or senior based on age with this formula =IF(E11>=50,"Senior",IF(E11>=30,"Adult","Teenager")) . Then paste as values so that system wont use resource more for calculations .
* Added new Column Month, here we extract month from the date using this function in excel =TEXT(G2,"mmm") … mmm stands for short form like DEC if there was mmmm i.e 4 m’s then it would be December.

## 6.0 Data Analysis

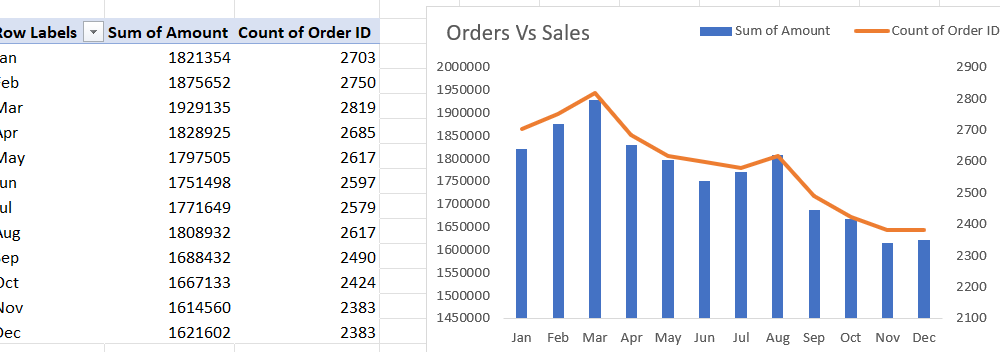
We are using Excel, so we are using pivot tables for Data Analysis.

### 6.0.1 Show sum of amount and count of orders by month

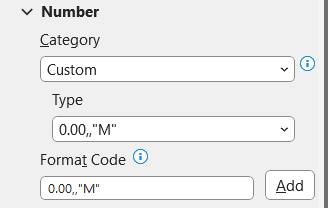
Drag the Month field to Rows, Order ID and Amount to Values in pivot table. Remove the Grand total from table using Design -> Grand Totals -> Off for Rows and Columns.

For data visualization we are using Combo Chart i.e a pivot chart and problem here is orders are in thousand and amount in millions so to make it visible at bottom .. secondary axis .. Count of Order ID to Line and Tick the secondary axis.

Right click on buttons and then Hide all field buttons on Chart.



After this in left side the sales number are hard to read so double click on it and format axis will open then do this



### 6.0.2 Which month got the highest sales and orders?

So, from the chart we can see that the orange line is highest in march and sales is also highest in march. Thus, we can state that the sales and order was highest in March.

**NOTE – In future if the source data changes, because we are using pivot charts.. go to pivot chart analyze -> refresh and then chart will get updated.**

### 6.0.3 Who purchased more- men or women in 2022?

For this we created new sheet with men vs women name.

In Rows select Gender then select Amount(sum) in values.

For data visualization use pie chart, Data labels – data callout, go to format and then remove background, go to home and edit color and make it bold.

**Ans**: Women purchased more than the men i.e., 64% to 36% for men.

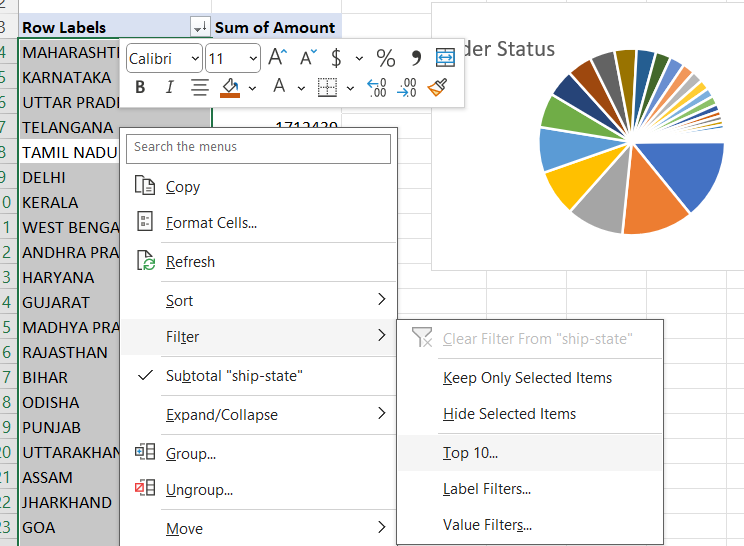
### 6.0.4 What are different order status in 2022?

So, for order status in pivot table status and count or order id is needed. Then do same process as before for pie chart but double click on pie chart and change angle so that title and legends don’t mess around with each other.

**Ans**: There were 92% delivered, 2% Refunded, 3% Returned, 3% Cancelled.

### 6.0.5 List top 10 states contributing to the sales?

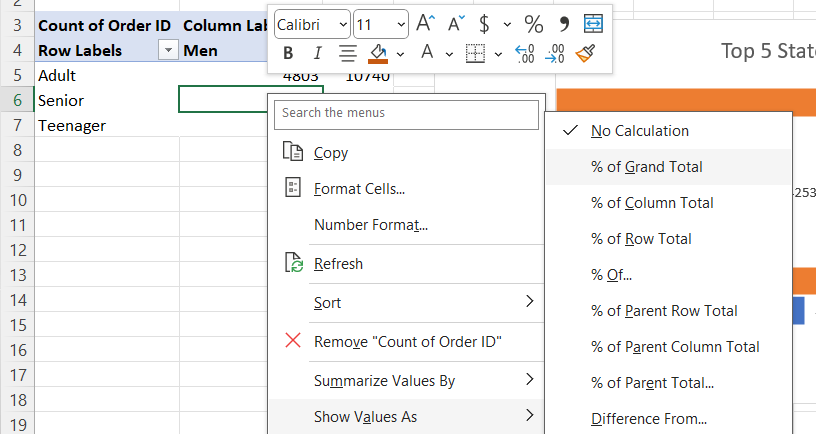
In pivot table we will do simply add ship state to rows and amount to values.. sort sum of amount to descending and filter top 5 for now.



Use Bar Chart here, and make format in millions.  
**Ans**: Top 5 States are Tamil Nadu, Telangana, Uttar Pradesh, Karnataka and Maharashtra.

### 6.0.6 Relation between age and gender based on number of orders.

We are using Age group i.e., calculated field in Rows, Gender in Columns and Count of Order ID in values. We will then calculate % of grand total using following process.



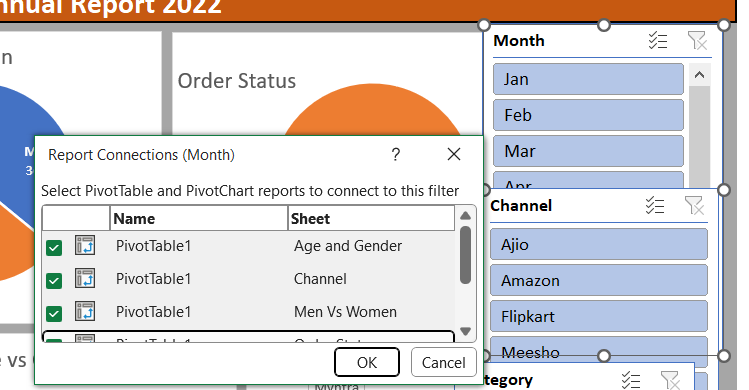
**Ans**: Relation between age and gender are Adult Women Did 34% shopping where as Adult Men were only 15% and so on.

### 6.0.7 Which channel is contributing to maximum sales?

For this we keep Channel in Rows and Count of Order ID in Values

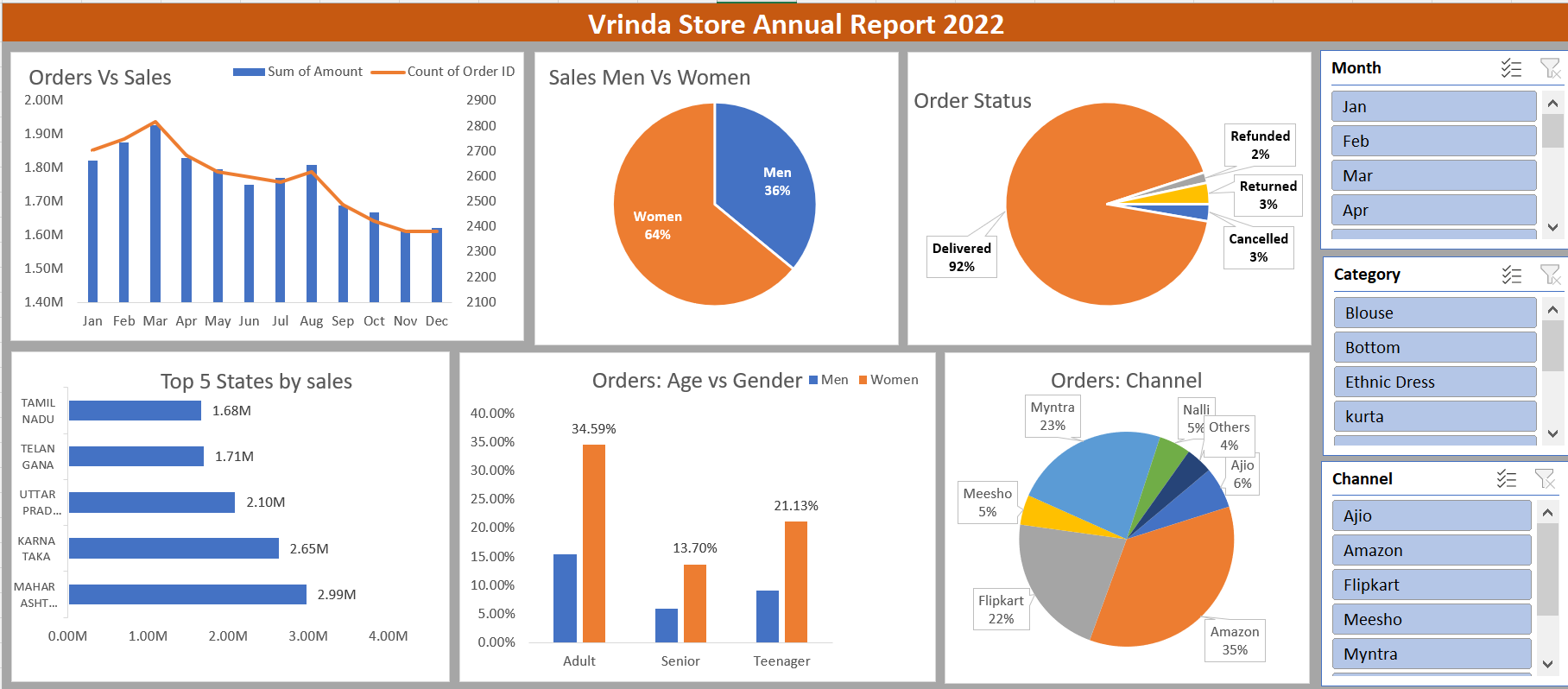
We used pie chart here because when we have 360-degree overview we use pie chart.

**Ans**: 35% is highest from Amazon



Connect slicer to all pivot tables

## 7.0 Final Report



## 8.0 Sample Insights

1. Women are more likely to buy compared to men (~65%)
2. Maharashtra, Karnataka and Uttar Pradesh are top three states (~35%)
3. Adult age group (30-49) years is max contributing (~50%)
4. Amazon, Flipkart and Myntra Channels are max contributing (~80%)

## 9.0 Final Conclusion to improve Vrinda Store sales:

Target women customers of age group (30-49) yrs. living in Maharashtra, Karnataka, and Uttar Pradesh by showing ads/offers/coupons available on Amazon, Flipkart and Myntra.