|  |
| --- |
| **E-commerce Sales Data Analysis and forecasting future sales using Excel**  **About Data:** |
| I used real-world data from Kaggle and supplemented it with a dataset created by ChatGPT. I made sure the data matched industry trends and made sense. Since the Kaggle dataset was quite small, I expanded it to work with a larger dataset, similar to real-world data. This helped me practice data cleaning, analysis, and forecasting in a structured way. |

The dataset contains sales data from May to Oct 2024. It includes the following columns:

**Order\_id:** Unique Id which identifies a single order placed.

**Order\_date:** Date and time of the order

**SKU**: Product number used by the retailer

**Color**: The color of the product ordered

**Size**: The size of the product ordered

**Unit\_price:** Price per product

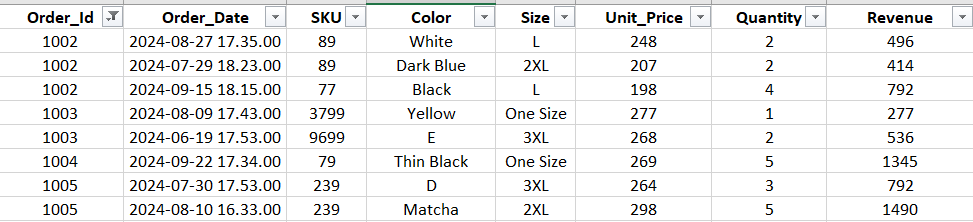
**Quantity**: Number of Quantities Purchased

**Revenue:** unit\_price\*quantity

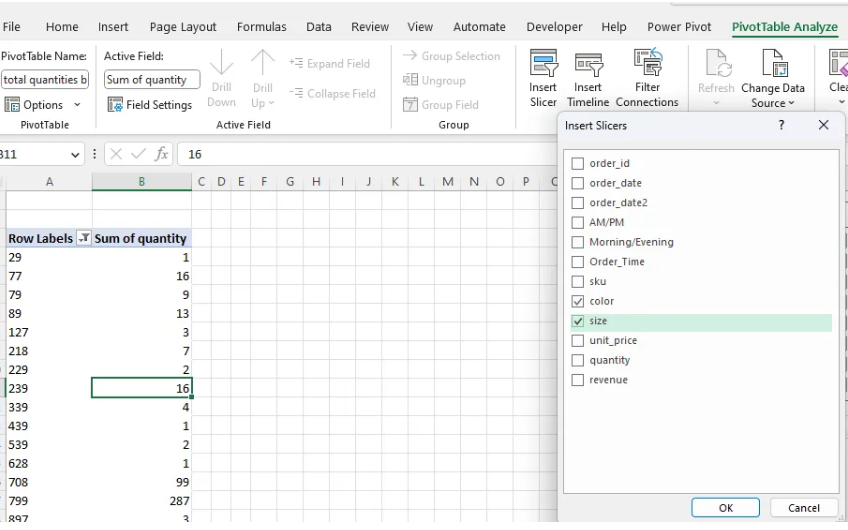
After analyzing the data we need to answer some questions:

**Business Questions**

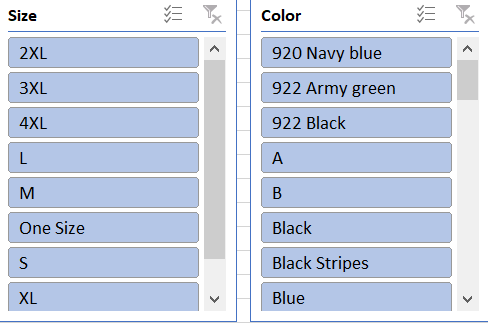
|  |  |
| --- | --- |
| **What are the best and worst-selling SKU items? by color? size?** | |
| **What is the average order value?** | |
| **What are the peak days or time periods with the highest sales? Do sales follow a trend or a seasonality?** | |
| ***Advanced:* Predict sales in the next months.** | |
|  | |
| **Note:** Order\_id is not unique. Multiple products ordered with the same order Id but different SKU/size/color are represented as different rows in the dataset. The quantity column always has values 1, 2,3,4 & 5. | |



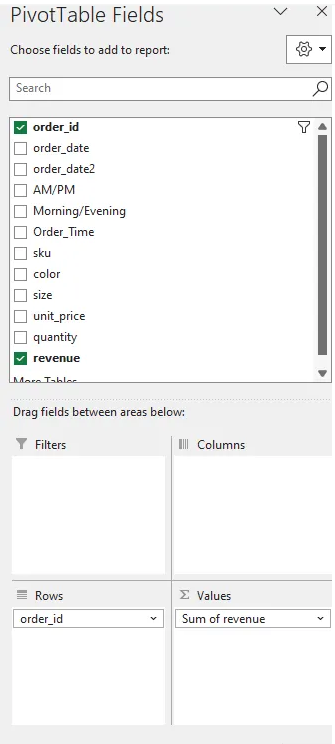
|  |  |
| --- | --- |
| **Data Cleaning:** | |
| Here we will perform some data cleaning to prepare the data for further analysis. | |
|  | |
| **1) Removing Duplicates:** I selected the entire data and then clicked Remove Duplicates in the Data Tab. There were 7 duplicate values in this case that were removed. | |
|  | |
| **2)  Checking for BLANK cells***:* Select the entire dataset and click Find and Select and finally select Go To Special. | |
|  | |
| Next, select Blanks and click Ok.     |  | | --- | | All blank cells should be highlighted. | |  | | In this dataset, there were blank cells only in size. There were 41 blank cells and I do not want to lose that much data. Since the data in column Size is categorical, it is best to replace the blank cells with Mode. I calculated the Mode by plotting a bar chart of the "Size" column. As evident from the below bar chart, the mode is "One Size". |     To replace all blanks with One Size, select the entire size column, go to Find and Select, and click replace. Replace all Blanks with One Size.    After selecting replace all, we see that 41 blank cells have been replaced.    **3. Spell Check***:* Next I wanted to spell check my data to maintain uniformity. I used Spelling from the Review pane after selecting the color column.    In this case, we need a space between Skintone. When we click Change All, all Skintone will be replaced by Skin tone      Similarly, I replaced Spelling errors from the size column. One SIze was replaced with One Size in Size Column.      **4) Separate Date and time***:* To analyze the data separately with date and time, I decided to separate the date and time. We can do this easily by using the Trunc function in Excel. | |
| Then I have used the below-mentioned formula to get only the Time. This will remove the date and keep only the time. | |
| |  | | --- | | Both columns can now be formatted appropriately for better analysis. | |  | | Now I want to get the AM/PM in column K from the Time from column J using this formula: |     I used a simple IF formula to create another column Morning/Evening in column L to represent AM by Morning and PM by Evening.    Finally, I like to align my data so that it becomes uniform and easier to read. In this case, I used center alignment. | |
| |  | | --- | | **Exploratory Data Analysis** | | 1. **What are the best and worst-selling SKU items? by color? size?** | |  | |  | | I selected the entire dataset to answer this question and created a Pivot table in a new sheet.    For finding the best and worst selling SKU items, I used SKU in rows and Sum of Quantity as values.    To create a chart, I selected the entire Pivot table and selected the Column chart from the Insert tab. | | |
| **Best and Worst Selling SKU items**   |  | | --- | | **Ans- The best-selling SKU is bobo and the worst-selling SKU is 799.** | |  | | Now we need to find the Best and Worst Selling SKU items by size and color. I decided to use Slicer for this so that we can see the best and worst-selling items for each color and size. | |  | |  | | Go to Pivot Table Analyze, select Insert Slicer, and then select color and size. | | |



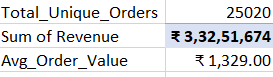
Since the best and worst-performing SKUs changed by color and size, I made an interactive chart where we can see the performance of SKUs based on size and color.



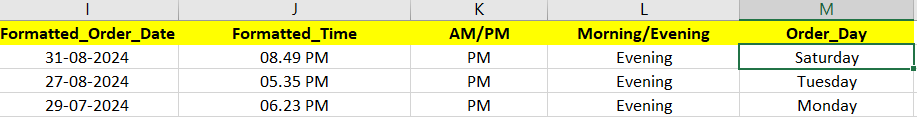
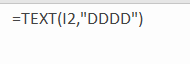
|  |
| --- |
| **2. What is the average order value?** |
|  |
|  |
| Since order Ids are repeated. First I created a pivot table that calculates the Sum of revenue per order Id. |



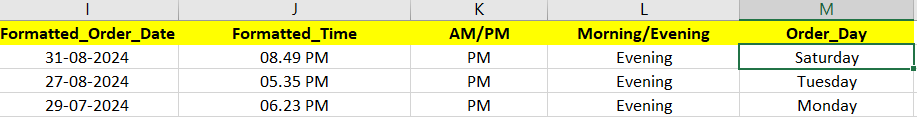
From the Pivot table, we can get the Total number of unique orders and Total Revenue from all Orders.



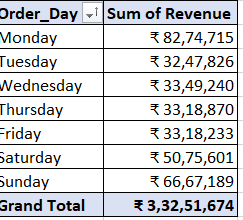
|  |
| --- |
| 3. What are the peak days or time periods with the highest sales? Do sales follow a trend or a seasonality? |
|  |
|  |
| To find out if sales are higher/lower on any particular day, I extracted the day (column M) from Order\_date (column I) using the TEXT function. |



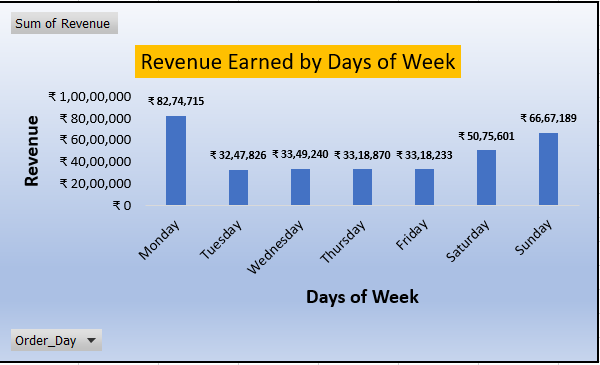
Since the column has a formula, we will not get it as a field in the pivot table. Hence I copied and pasted(Paste Special) the values of the column into another blank column. I named the column Order\_day (column M)



To find days with the highest sales, I created a pivot table with Order\_day in rows and the sum of Revenue as values. After that, I created a column chart from the Insert tab.

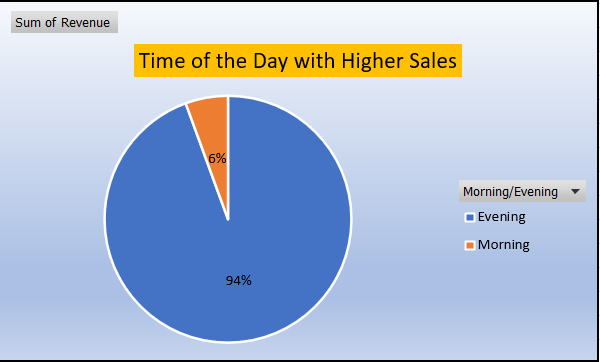


The chart looks like the following

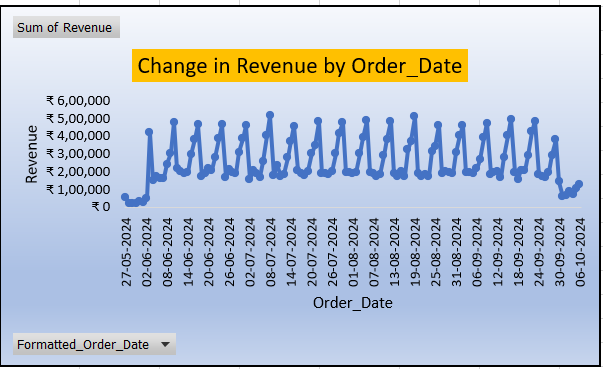


**Revenue Earned by Day of the Week**

|  |
| --- |
| **Interpretation & Insights** |
| Revenue is higher on weekends than on weekdays. |
| Monday has the highest revenue (₹82,74,715). |
| Sunday (₹66,67,189) and Saturday (₹50,75,601) also show higher revenue compared to weekdays. |
| Tuesday to Friday have relatively stable and lower revenue, ranging from ₹32,48,726 (Tuesday) to ₹33,49,240 (Wednesday). |
| **Trend and Seasonality:** |
| A clear pattern emerges where revenue increases significantly on weekends, peaking on Monday. |
| This suggests that orders placed over the weekend might be processed or recorded on Monday, leading to the spike. |
| The weekday revenue remains steady without major fluctuations. |
| This indicates a seasonal trend where weekend sales drive higher revenue, spilling over into Monday. |
|  |
| Similarly, to find the time of the day with the highest sales, I used the Morning/Evening field and the Sum of revenue and created a pie chart. |

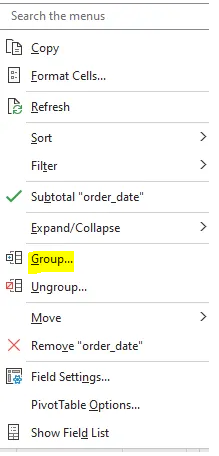


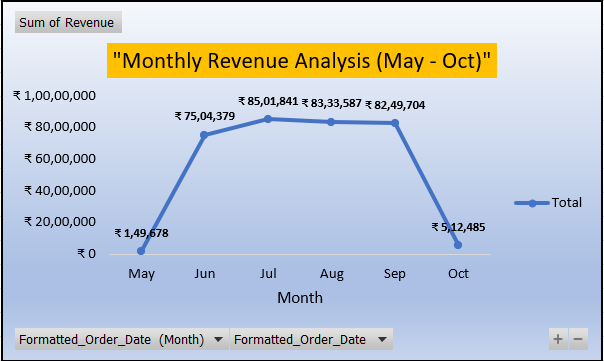
|  |  |
| --- | --- |
| **Interpretation & Insights** | |
| Evening Sales Dominate – The blue section, representing Evening sales, accounts for 94% of the total revenue. Morning Sales Are Minima – The orange section, representing Morning sales, contributes only 6% of the total revenue. Conclusion – The business sees significantly higher sales in the evening compared to the morning. This could indicate customer preferences or peak shopping hours. | |
| **Evening sales dominate (94%)**, while **Morning sales are very low (6%)**. | |
| **Key Insights:** | |
| **1. Peak sales occur in the evening**, indicating higher customer activity. | |
| **2. Morning sales are weak**, suggesting a need for promotions or discounts to boost them. | |
| **3. Operational impact** – More staff & inventory may be needed in the evening. | |
| **4. Marketing strategy** – Focus ads and offers in the evening for maximum impact. | |
|  | |
| **Seasonality in data:** | |
|  | |
|  | |
| I created a line chart of Order date and Sum of revenue to examine if we have seasonality in the data. | |



**Change in Revenue by Order\_Date**

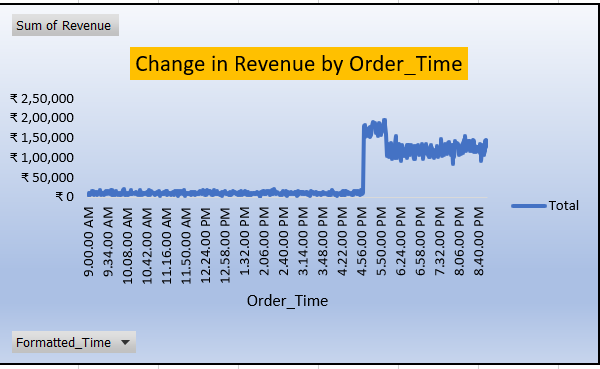
|  |
| --- |
| This chart represents the **Change in Revenue by Order\_Date** over time. Here are some key insights: |
| After analyzing the peak sale dates I found that on 08/07/2024,19/08/2024,16/09/2024 highest number of the orders were placed. |
| **1. Strong Seasonal Pattern**: The revenue shows a repeating pattern with **regular peaks and troughs**, indicating **weekly or periodic fluctuations** in sales. |
| **2. Initial Growth**: There is a **sharp rise** in revenue at the beginning (end of May 2024), suggesting a **launch, promotional event, or start of data collection**. |
| **3. Cyclical Spikes**: The peaks are quite consistent, likely corresponding to **high-demand days** (e.g., weekends, sales promotions, or paydays). |
| **4. Sharp Drop in Late September**: A **significant decline** is observed at the end of September 2024, potentially due to: |
| a) End of a sales period. |
| b) Seasonal slowdown. |
| c) External market conditions. |
| **5. Recent Uptick**: After the sharp decline, a **slight recovery** is visible in early October 2024. |
|  |
| To get a clearer picture I right-clicked on the order date of the previous pivot table and grouped dates in groups of months to get the "Monthly Revenue Analysis". |





**Monthly Revenue Analysis (May-Oct)**

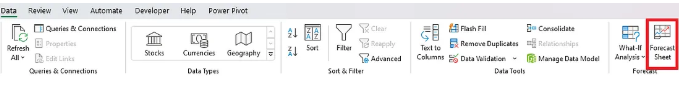
|  |
| --- |
| **Key Observations-**  **Strong Growth (May–July):** Sales surged from ₹1.5L to ₹85L, likely due to seasonal demand or promotions. |
| **Peak Stability (July–Sept):** Revenue remained high (~₹83L-₹85L), indicating sustained demand. |
| **Sharp Decline (Oct):** Sales dropped to ₹5.1L, possibly due to seasonality or market changes. |
|  |
| **Then I plotted the following line graph that highlights these below mentioned key points-** |
|  |
| Evening time is a clear winner for higher revenue. |
| **Before 4:50 PM:** Revenue remains consistently low. |
| **At 4:50 PM:** There is a **sharp spike** in revenue, indicating a sudden increase in sales. |
| **After 4:50 PM:** Sales remain **higher but fluctuating** until at least **8:40 PM** (the last recorded time). **Possible reasons include order processing delays, cancellations, or refunds affecting revenue.** |
|  |
| So, the **peak sales period starts at 4:50 PM and continues until the end of the recorded data (8:40 PM) It may be because of any promotional event, flash sale, or bulk purchase at that time/ or may be because of the evening rush in the business.** |



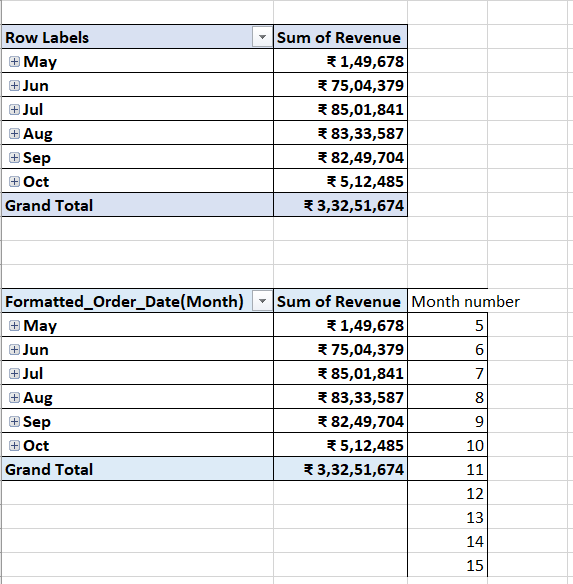
**Change in Revenue by Order\_Time**

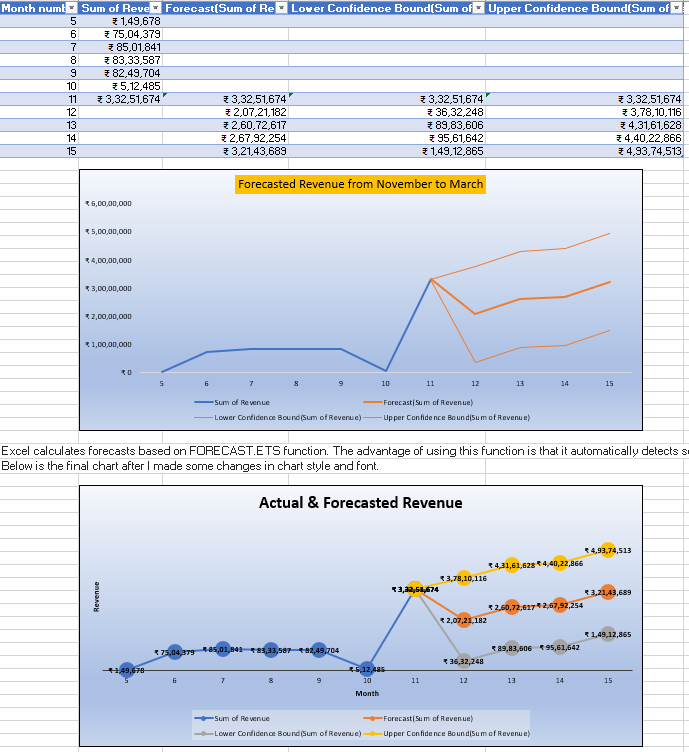
|  |
| --- |
| **Question- During which time-period does the sales volume reach its peak?** |
| **Answer- Massive revenue spike at 5 PM, peaking between 5 PM - 8 PM.** |
|  |
| I plotted a **line chart** to show the trend of revenue over different hours of the day. The line connects data points to display how sales change over time, making it easy to identify the **peak hours for higher sales** (5 PM - 8 PM) |

|  |
| --- |
| **Key Observations:** |
|  |
| Sales are **very low** from **10 AM to 4 PM**. |
| Sales **suddenly increase sharply at 5 PM** and stay high until **8 PM**. |
| After 8 PM, sales **drop back to low levels**. |
|  |
| This means that the **best time for higher sales is between 5 PM and 8 PM**. |
|  |
| ***4. Advanced:* Predict sales in the next months.** |
|  |
|  |
| We will use Forecast Sheet from the Data tab to predict monthly sales from November till March. |
| ***Forecast Period:*** Since we need to predict from November till March, So Forecast start period will be from 11 and Forecast end period will be 15. |



|  |
| --- |
| I created a pivot table of month and sum of revenue. I pasted the value from the pivot table in separate columns in order to make further calculations. |
| Excel’s forecast function does not recognize non-numeric input for months. Hence, I inserted a column month number and used the month number instead of the month name. |
|  |
| Next, select the month number and total revenue and click on Forecast sheet from the Data pane. |

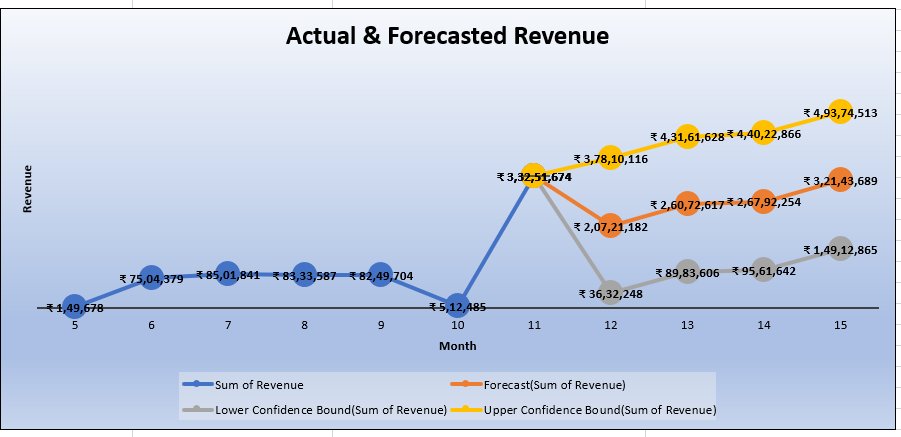




|  |
| --- |
| When the forecast sheet worksheet is expanded by clicking on Options we get various choices. |
| ***Forecast End:*** I changed the forecast end period to 12 from 10 since we need to predict till December. |
| I kept other options as default. |
| ***Confidence Interval:*** Selecting a 95% confidence interval means that 95% of the future values will fall in the range of upper and lower confidence boundaries. |
| ***Interpolation:*** Interpolation is used to fill in the missing points based on the weighted average of neighboring points where values for certain months are missing. |
| ***Aggregate duplicates using:***If there are multiple values for some months, then an average value is calculated by default. |
| Now when we click on create, Excel will create forecasted data in a separate sheet. |
|  |
|  |
| Excel calculates forecasts based on **FORECAST.ETS** function. The advantage of using this function is that it automatically detects seasonality in the data and leads to better prediction. |



Below is the final chart after I made some changes in chart style and font.



**Actual & Forecasted Revenue**

|  |
| --- |
| **Insights:** |
| In my analysis, I observed a stable revenue trend initially, followed by a sharp spike at point 11, likely due to a major event or seasonal demand. |
| The forecast indicates potential fluctuations, with an upper confidence bound suggesting growth opportunities and a lower bound indicating possible decline. |
| To optimize revenue, I would recommend targeted marketing strategies and operational improvements to stay within the upper forecasted range." |
|  |
| **Revenue Optimization Strategies** |
|  |
| **Targeted Marketing:** Use **personalized promotions**, **seasonal campaigns**, and **retargeting ads** to attract and retain customers. Collaborate with **influencers & affiliates** for brand visibility. |
| **Recommendation** |
| Let’s revisit the business questions and provide answers. |

|  |  |
| --- | --- |
|  | |
|  | |
|  | |
| **Question 1) What are the best and worst-selling SKU items? by color? size?** | |
| **Answer**- The best-selling SKU is bobo and the worst-selling SKU is 799. It is advisable to keep a larger stock of SKU bobo and minimize storage of SKU 799. A similar decision should be made by analyzing each SKU item by size and color from the interactive chart. Eg. SKU 8499 should be stocked more for Thin Black color and SKU 708 should be stocked most for "One Size". The worst selling item for Cream color is SKU 2499 and the worst selling item for 2XL size is 229. | |
|  | |
| **Question 2) What is the average order value?** | |
| **Answer**-The average order value is Rs. 1329. Adding personalized product recommendations and providing discount coupons are great ideas to increase the average order value. | |
|  | |
| **Question 3) What are the peak days or time periods with the highest sales? Do sales follow a trend or a seasonality?** | |
|  | |
| **Answer - Evening Sales Dominance:** Sales are significantly higher in the **evening (94%)** compared to the morning (6%), making it a key time for promotions. Running ads in the morning may help attract potential buyers. | |
| **Sales Trend Over the Month:** Revenue peaks in **June, July, August and September**, followed by a decline in **October**, indicating possible seasonality or supply chain limitations. | |
| **Weekly Patterns:** **Monday has the highest sales, followed by Sunday,** while other weekdays see relatively lower revenue. This suggests a possible payday effect or customer shopping habits early in the week. Further analysis is needed to check if inventory shortages affect sales at the end of the month.  **-----End of Reports------** | |