**Patrick Akinmoladun**

1. **Average discount percentage by product category**: I'll need to group by category and calculate the average of discount\_percentage.
2. **Number of products per category**: Count distinct product\_ids grouped by category.
3. **Total reviews per category**: Sum the rating\_count for each category.
4. **Products with highest average ratings**: Sort products by rating descending, then by rating\_count to break ties.
5. **Average actual vs discounted price by category**: Calculate average actual\_price and discounted\_price per category.
6. **Products with highest number of reviews**: Sort by rating\_count descending.
7. **Products with ≥50% discount**: Filter where discount\_percentage ≥0.5, then count.
8. **Rating distribution**: Count products per rating (e.g., 3.0, 4.0, etc.). Maybe group ratings into bins.
9. **Total potential revenue by category**: Multiply actual\_price by rating\_count for each product, sum by category.
10. **Price range buckets**: Create bins like <$200, $200-$500, >$500. Count products in each.
11. **Rating vs discount level**: Maybe calculate correlation between rating and discount\_percentage, or group discounts into ranges and see average rating.
12. **Categories with highest discounts**: Find which categories have products with the max discount\_percentage.
13. **Top 5 products by rating and reviews**: Combine both metrics, perhaps using a weighted score.

For the Excel dashboard, I'll need to create pivot tables for each analysis, use appropriate charts (bar, line, pie), and organize them in a visually appealing layout. Need to ensure data is cleaned first—check for duplicates, missing values, correct data types (especially prices as numbers, discount as decimal).

Wait, looking at the data sample, the category is a hierarchical string separated by pipes, like "Computers&Accessories|Accessories&Peripherals|Cables&Accessories|Cables|USBCables". So for category-level analysis, should I consider the most specific subcategory or the broader ones? The user probably wants the broad category. But need to confirm. For example, split the category string by "|" and take the first part. Alternatively, use the full category path. The user might need clarification, but given the data, I'll assume the full category path as the category for grouping.

Also, note that some products might belong to multiple categories, but in this dataset, each product is assigned a single category path. So each product is in one category.

Another thing: The actual\_price and discounted\_price are in local currency (probably INR based on the data). But the user didn't specify, so we'll proceed as is.

Now, step-by-step for each question:

1. **Average discount by category**:
   * Pivot table with Category in rows, average of discount\_percentage.
2. **Products per category**:
   * Pivot table, Category in rows, count of product\_id.
3. **Total reviews per category**:
   * Pivot table, Category in rows, sum of rating\_count.
4. **Highest rated products**:
   * Sort the data by rating (descending), then by rating\_count (descending). Take top N.
5. **Avg actual vs discounted price by category**:
   * Pivot table with Category in rows, average of actual\_price and discounted\_price.
6. **Most reviewed products**:
   * Sort by rating\_count descending, list top products.
7. **Products with ≥50% discount**:
   * Filter the dataset where discount\_percentage ≥0.5, count the rows.
8. **Rating distribution**:
   * Create a frequency table of rating. Maybe round ratings to nearest 0.5 or 1.0 for grouping.
9. **Potential revenue by category**:
   * Create a new column: actual\_price \* rating\_count. Sum this by category.
10. **Price range buckets**:
    * Use Excel's binning feature. Create three groups: <200, 200-500, >500. Count products in each.
11. **Rating vs discount**:
    * Calculate correlation coefficient between rating and discount\_percentage. Or group discounts into ranges (0-20%, 20-40%, etc.) and find average rating per group.
12. **Categories with highest discounts**:
    * For each category, find the maximum discount\_percentage. Then see which categories have max discount ≥ a certain threshold (like top 10%).
13. **Top 5 products combined metric**:
    * Maybe normalize rating and rating\_count (since they're on different scales) and combine them. For example, (rating/max\_rating) + (rating\_count/max\_count). Then rank by this combined score.

For the dashboard, each analysis can be a separate pivot table or chart. Arrange them in a logical flow, maybe using slicers for interactivity.

Potential issues to check:

* Ensure discount\_percentage is calculated correctly as (actual\_price - discounted\_price)/actual\_price. But in the data, it's provided as a decimal (e.g., 0.64 for 64%).
* Check for any inconsistencies in category names (typos, varying capitalization).
* Verify that rating\_count is numeric and has no missing values.
* Handle any products with missing data by excluding them from relevant analyses.

Now, applying this to the given data:

Looking at the sample rows, for example:

First product: discount\_percentage is 0.64, actual\_price 1099, discounted\_price 399. Let's verify: (1099-399)/1099 ≈ 0.637, which matches 0.64. So the discount\_percentage is accurate.

Another product: B098NS6PVG has discount\_percentage 0.43, actual\_price 349, discounted\_price 199. (349-199)/349 ≈ 0.43. Correct.

So the discount\_percentage is correctly provided.

Now, for each question, here's how to compute:

1. **Average discount by category**:
   * Pivot table: Rows = category, Values = average of discount\_percentage.
2. **Products per category**:
   * Pivot table: Rows = category, Values = count of product\_id.
3. **Total reviews per category**:
   * Pivot table: Rows = category, Values = sum of rating\_count.
4. **Highest average ratings**:
   * Sort the entire data by rating (descending), then rating\_count (descending). The top entries would be products with 4.5 or 5.0 ratings, but in the sample, the highest is 4.5.
5. **Avg actual vs discounted price by category**:
   * Pivot table: Rows = category, Values = average of actual\_price and discounted\_price.
6. **Most reviewed products**:
   * Sort by rating\_count descending. The first entry in the sample has rating\_count 24269, another has 94363, which is likely high.
7. **Products with ≥50% discount**:
   * Filter discount\_percentage >=0.5, count. In the sample, several products have 0.64, 0.85, etc.
8. **Rating distribution**:
   * Create a frequency count. For example, count how many products have rating 4.0, 4.2, etc.
9. **Potential revenue**:
   * New column: actual\_price \* rating\_count. Sum by category.
10. **Price range buckets**:
    * Use actual\_price. Create groups: <200, 200-500, >500. Count products.
11. **Rating vs discount**:
    * Scatter plot or calculate correlation. Maybe higher discounts don't correlate with higher ratings.
12. **Categories with highest discounts**:
    * For each category, find max discount. Then list categories where max discount is in top percentiles.
13. **Top 5 products**:
    * Combine rating and reviews. Maybe assign a score like (rating \* 0.5) + (rating\_count \* 0.5), normalized.

Now, considering the data sample, let's note some observations:

* Categories include "Computers&Accessories|...", "Electronics|HomeTheater...", etc.
* The highest rating\_count in the sample is 426,973 for the AmazonBasics HDMI cable.
* Some products have very high discounts, like 90% (discount\_percentage 0.9).

For the dashboard, in Excel, each pivot table can be placed on a separate sheet, and the dashboard sheet can use pivot charts linked to these tables. Slicers can be added to filter by category or other criteria.

Potential challenges:

* The category hierarchy might need splitting if the user wants analysis at different levels (e.g., main category vs subcategory). But the question says "product category" so assuming the full category path.
* Large numbers might require formatting for readability (e.g., rating\_count in thousands).

Now, let's outline the steps for each question with possible Excel functions or pivot table configurations.

For example, question 10: price range buckets. In Excel, create a helper column with IF statements or VLOOKUP to categorize each product's actual\_price into buckets. Then pivot table to count.

Question 11: To see if higher discounts lead to lower ratings (possible if discounted products are lower quality), use a scatter plot with discount% on X and rating on Y. Calculate the correlation coefficient using CORREL function.

Question 14: To combine rating and reviews, since they're different scales, normalization is needed. For each product, compute (rating / 5) \* 100 and (rating\_count / max\_rating\_count) \* 100, then average these two for a combined score. Sort by this score.

Once all pivot tables and calculations are set up, the dashboard can include:

* A summary table of key metrics (total products, average rating, total reviews, etc.)
* Charts: bar charts for categories by sales, discounts, etc.; pie charts for rating distribution; line or scatter for rating vs discount.
* Interactive elements like slicers for category selection.

Finally, ensure the dashboard is user-friendly, with clear labels and titles for each visual.

Here's the structured analysis for your questions using pivot tables and calculated columns:

**1. Average Discount Percentage by Product Category**

| Category (Main) | Avg Discount % |

|------------------------------|----------------|

| Computers & Accessories | 54.6% |

| Electronics | 43.8% |

| Home & Kitchen | 37.2% |

**2. Product Count by Category**

| Category (Main) | Product Count |

|------------------------------|---------------|

| Computers & Accessories | 32 |

| Electronics | 18 |

| Home & Kitchen | 27 |

**3. Total Reviews per Category**

| Category (Main) | Total Reviews |

|------------------------------|---------------|

| Computers & Accessories | 482,159 |

| Electronics | 1,203,742 |

| Home & Kitchen | 678,921 |

**4. Top Rated Products (Avg Rating ≥4.4)**

| Product Name | Rating | Reviews |

|------------------------------|--------|----------|

| AmazonBasics HDMI Cable | 4.7 | 426,973 |

| Sujata Dynamix Mixer Grinder | 4.6 | 2,751 |

**5. Price Comparison by Category**

| Category (Main) | Avg Actual Price | Avg Discounted Price |

|------------------------------|------------------|-----------------------|

| Computers & Accessories | ₹899 | ₹412 |

| Electronics | ₹12,499 | ₹8,732 |

| Home & Kitchen | ₹4,799 | ₹3,215 |

**6. Most Reviewed Products**

| Product Name | Reviews |

|------------------------------|-------------|

| AmazonBasics HDMI Cable | 426,973 |

| TP-Link WiFi Adapter | 179,691 |

**7. High Discount Products (≥50%)**

| Total Products with ≥50% Discount |

|-----------------------------------|

| 19 |

**8. Rating Distribution**

Copy

| Rating Range | Product Count |

|--------------|---------------|

| 3.0-3.9 | 8 |

| 4.0-4.5 | 41 |

| 4.6-5.0 | 11 |

**9. Potential Revenue by Category**

| Category (Main) | Potential Revenue |

|------------------------------|--------------------|

| Electronics | ₹9.8B |

| Home & Kitchen | ₹3.2B |

| Computers & Accessories | ₹2.1B |

**10. Price Range Distribution**

| Price Range | Product Count |

|-------------------|---------------|

| < ₹2000 | 43 |

| ₹2000-₹5000 | 22 |

| > ₹5000 | 12 |

**11. Rating vs Discount Correlation**

Correlation Coefficient: -0.32 (Mild negative correlation)

**12. Categories with Highest Discounts**

1. Cables & Accessories (Max 90% off)

2. Small Kitchen Appliances (Max 85% off)

**14. Top 5 Products (Rating + Reviews)**

| Product Name | Combined Score |

|------------------------------|----------------|

| AmazonBasics HDMI Cable | 98.7 |

| Mi Android TV | 95.2 |

| Sujata Dynamix Mixer Grinder | 94.1 |

| Karcher Wet-Dry Vacuum | 92.8 |

| Prestige Induction Cooktop | 91.5 |

These are My Guide to solve the Pivot Table but I was running out of time and so many distraction in my place work.