Case Study: Optimizing Discount Strategies for E-commerce Sales

Business Context:

Company ABC, a dominant player in the beauty and personal care market, seeks to leverage the power of data-driven decision-making to optimize its online discount strategies. In an increasingly competitive e-commerce landscape, understanding the impact of discounts on sales, profitability, and customer behavior is critical for achieving revenue goals.

Business Objective:

In the dynamic world of ecommerce, optimal product pricing is crucial for maximizing profitability and maintaining a competitive edge. Manually setting and adjusting prices for numerous products is inefficient and often relies on intuition rather than data-backed insights.

As a Data Scientist at this leading company, you have been tasked with analyzing a comprehensive dataset spanning daily sales, marketing performance, and competitor pricing data across five major online platforms and diverse city tiers in India for a period of 6 months.

Challenge 1: Conduct a comprehensive exploratory data analysis to understand the distribution, central tendencies, and variability of each attribute. Visualize the data to identify any outliers or trends that may require further investigation.

Challenge 2: Perform a correlation analysis among the variables present in the ecommerce sales data. Write your analytical interpretation to explain insights to sales team.

Challenge 3: Build a regression model to predict quantity sold of the products based on the variables in the given dataset. Perform the basic assumptions test to validate the regression model output.

Challenge 4: Based on the data analysis conducted on the ecommerce sales data, create a 2-page powerpoint to present insights and recommendations to the CEO of the ABC company.

Case Study Submission Guideline (Time: 2-3 Hours on Coding & 1 Hour on PPT)

Please provide your complete analysis in the form of a well-documented Python Code i.e., either Python file or Jupyter Notebook along with the code comments and observations made.

Dataset Description:

Here are the definitions of the columns in your dataset:

Product and Pricing Information:

- Product: The name of the product as per the listing on the ecomm platform
- Date: The date on which the discount and sale took place (daily level).
- MRP: Maximum Retail Price - The highest price at which a product can be sold to consumers.
- Discount: The percentage discount being offered on the product for that day (for ex: 0.1 is 10%)
- CLP: Customer Landing Price - The actual price that a customer pays for the product after the discount is applied.

Sales and Marketing Performance:

- Platform: The e-commerce platform where the product is listed (e.g., Amazon, Nykaa).
- City: The city from where the customer order originated.
- City Tier: The tier classification of the city (e.g., Tier 1, Tier 2, Tier 3), typically based on population and economic factors.
- Impressions: The number of times the product was displayed to potential customers on the platform.
- Clicks: The number of times customers clicked on the product listing.
- Spend: The marketing or advertising spend allocated to the product on that platform and day.
- Quantity Sold: The number of units of the product that were sold on that day.

Competitor Information:

- Competitor: The name of a competitor selling a similar product.
- Competitor MRP: The Maximum Retail Price of the competitor's product.
- Competitor Discount: The percentage discount being offered by the competitor on that day (for ex: 0.15 is 15%).
- Competitor Price: The actual selling price of the competitor's product after applying the discount.