Trending Threads: Can a Local Fashion Store Compete with Influencers?

In a cozy corner of Ahmedabad, Gujarat, there is a small but stylish boutique named "Threadora", owned by 29-year-old entrepreneur Riya Mehta. Over the past year, Riya has seen her once-steady foot traffic fluctuate wildly. Some days, people pour in non-stop. Other days, it's crickets.

After speaking with her younger cousin who runs a fashion blog, Riya realized something: Influencer posts on Instagram and YouTube seem to be driving a lot of unexpected interest in certain items, especially the 'Kurti & Palazzo' sets and Indo-western collections. Her cousin explained how even micro-influencers (those with 5K to 50K followers) can massively impact buying decisions in Tier-2 and Tier-3 cities.

Riya decided to gather data over the next 15 days on:
- Number of daily customer footfalls
- Total sales
- Ad spending (online + local print)
- Number of Instagram posts that mention her store or tag products
- Number of WhatsApp product inquiries
- Whether a discount or deal was running that day
- Day of the week

She believed a little data and some basic analysis could give her an edge. She reached out to her college friend Priyansh, who just finished a basic course in Machine Learning, to help her explore the data.

Objective
You are Priyansh. Your job is to help Riya make sense of her store's recent performance using Machine Learning and Data Analysis techniques. She wants to understand:
- What influences her daily sales and customer visits the most?
- Do influencers truly matter for her boutique's performance?
- How effective is her ad spending?
- Is there a predictable pattern over time?
- Can she predict or improve sales using basic modeling?

Your Task

You will perform a complete ML workflow based on Riya's boutique data. Here's how each concept from Unit 1 fits in:

1. Data Acquisition and Pre-processing
- Read the provided `.csv` dataset containing Riya's 15-day store activity.
- Perform data cleaning: Check for missing or inconsistent entries.
- Apply pre-processing using Pandas and NumPy.

2. Data Transformation
- Normalize numerical columns (like sales, ad spend, inquiries) using Min-Max or Z-score normalization.
- Convert categorical data (like Day or Deal Offered) into numerical format for modeling.

3. Exploratory Data Analysis (EDA)
- Plot summary statistics: mean, median, mode, etc.
- Identify any outliers in the data (e.g., a sudden spike in inquiries or footfall).
- Check correlations between:
  • Influencer mentions and sales
  • Ad spend and footfall
  • WhatsApp inquiries and actual purchases

4. Data Visualization
- Use Matplotlib and Seaborn to:
  • Plot footfall and sales over time
  • Visualize relationships using scatterplots or pair plots
  • Show distribution of daily footfall/sales using histograms or boxplots

5. Data Modeling and Analysis
- Build a Linear Regression model to predict sales using independent variables (ad spend, footfall, inquiries, influencer posts).
- Build a Logistic Regression model to predict the probability of 'High Sales Day' (1 if sales > ₹5000, 0 otherwise).

- Use summary outputs to explain what variables have the most influence.

6. Bonus – Time Series Element
- Try plotting sales and footfall over days.
- Discuss whether there's a weekly pattern or trend (e.g., higher sales on weekends).

Riya's Final Question

At the end of your analysis, Riya asks:
"If I want to increase my daily sales and footfall, should I invest more in influencer marketing, WhatsApp marketing, or just go with print ads? Can you show me with numbers?"
Use your analysis and modeling to answer this in a simple conclusion.

Dataset Info

Filename: threadora_boutique_data.csv
Contains 15 rows of realistic data capturing:
- Daily footfall
- Sales (₹)
- Ad spending (₹)
- Instagram mentions
- WhatsApp product inquiries
- Deal offered (Yes/No)
- Day and Date