Business Analyst Intern Assignment

This assignment will test your analytical, problem-solving, and data handling skills across multiple tools (Excel, Python, Tableau). Please attempt each section carefully and submit your work in the required formats.

Section 1: Basic Excel Analysis (Ease into Data Exploration)

🎯Objective:

Work with the provided **Iphone** sales dataset (attached in the email) to explore and analyze trends in pricing, demand, and overall performance.

Direction:

Clean and organize the dataset provided. Ensure there are no duplicates, missing values, or formatting inconsistencies.
Analyze the data in a way that helps you uncover meaningful business insights.

Think about how you would present patterns, compare performance across products and time, and highlight areas that require attention.

1. Use Excel functions such as **VLOOKUP, Pivot Tables, Conditional Formatting, and Data Validation** to process and analyze the data efficiently.
   ✅Create meaningful visualizations (charts, graphs, pivot tables) to represent key insights.

Deliverable:

1. An Excel/Sheets file containing your analysis and supporting visualizations.

2. A concise summary (200–300 words) that explains your findings and the story behind the data.

Section 2: Python-Based Advanced Analysis (Data Handling & Insights)

🎯Objective:

You are given a rich transactional dataset (**Audio dataset** attached in the email) with various fields. Your task is to use Python (Pandas, Numpy, Visualization libraries) to **analyze, structure, and interpret** the data in a way that provides **market-level insights**.
Think of yourself as presenting this to a business leader who wants to understand how the e-commerce market is behaving.

Guidelines for Exploration (Not Step-by-Step Tasks)

- **Category & Brand Dynamics**: Which categories and brands dominate the market, and where do you see concentration vs. fragmentation? What does the **80/20 rule** (Pareto effect) look like here?

- **Price Sensitivity**: How does price variation influence demand? Can you group categories or brands into **high, medium, low elasticity** segments?

- **Customer Behavior**: Use concepts like **RFM analysis** or clustering to define distinct customer groups. How do these groups differ in spending, frequency, and product choices?

- **Geographic Trends**: What patterns emerge across geographies (city, state, pincode)? Are there **regional specializations** in product demand?

- **Cohorts & Retention**: What does repeat purchase behavior look like across cohorts? Do cohorts vary across platforms, sources, or regions?

Expectations

- Move beyond basic summaries; focus on **comparisons, anomalies, and patterns** that tell a story.
- Visualize trends clearly — charts, funnels, plots, heatmaps, etc.
- Connect your analysis back to **strategic insights**: what do these findings imply for the business?

📦 **Deliverable**

- A Python notebook with structured analysis, clean code, and visualizations.
- A report (500–700 words) summarizing:
    - Key findings
    - Market-level insights
    - Actionable recommendations

Section 3: Complex Open-Ended Problem – Deriving Demographics (Age & Gender Prediction)

🎯Problem Statement:

We want to derive user demographics (age and gender) based solely on transactional data.
Explicit demographic details are not available.

📝 Tasks

1. **Design a Methodology:**
    - Propose a framework to infer age groups (<25, 25–40, 40+) and gender (male/female) using only transactional signals.
    - Define weights for product categories and brands. Example:
        - Kids products → higher likelihood of 25–40 group.
        - Electronics (high-value) → skew toward male <40.
        - Beauty & Fashion → stronger gender signals.
    - Consider transaction frequency, spending power, and time of purchase as behavioral indicators.
2. **Scoring Model:**
    - Build a rule-based or ML-driven scoring system that assigns probabilities of belonging to each demographic bucket.
    - Example: If a user's top 3 categories are Beauty, Apparel, and Footwear with >60% share of spend, weigh them toward the female demographic.
3. **Validation Approach:**
    - Define how this model could be validated once demographics are available (precision/recall, confusion matrix, lift over random baseline).

📦 Deliverable

- A detailed methodology document (step-by-step reasoning, weights, assumptions, formulas).
- *(Optional)* A Python prototype applying this framework to classify users into demographic buckets.

✅ **By the end of this assignment, you will demonstrate skills across:**

- **Excel/Sheets** → practical analysis (Section 1)
- **Python for Data Handling** → advanced transactional analysis (Section 2)
- **Analytical Reasoning & Problem Framing** → demographic inference using transaction-only data (Section 3)

The instructions given are not exhaustive in nature. You can use your own creative & innovative ways to solve the problem