Coffee Shop Expansion Workshop

Introduction

Welcome to the **Coffee Shop Expansion Workshop**

You're stepping into the shoes of a **data analyst** at a rapidly growing coffee shop chain. The executive team is preparing to open a **new store**, but they need your help answering a critical question:

> **Which city should we expand to next?**

Using real business data, your goal is to analyze store performance and external market conditions to make a **clear, data-backed recommendation**. This is the kind of problem you'll face in real analytics jobs where your analysis drives strategy, and your ability to uncover insights and communicate them effectively is key.

Workshop Objectives:

- Thinking critically about business problems
- Structuring your analysis with clear objectives
- Creating and interpreting KPIs (key performance indicators)
- Aggregating, visualizing, and comparing performance
- Making persuasive, data-driven recommendations

Dataset Information

You'll work with **two datasets**, simulating how real-world data comes from multiple sources that you need to bring together.

1. `Coffee_Shop_Transaction_Data.csv`

This dataset contains **daily transaction-level data** for 15 coffee shop locations over a 2-year period.

Key Features:

- `Store ID` and `City`: identifies the store and its location
- `Date`: the transaction day (used for trend analysis)
- `Foot Traffic`: number of people who visited the store
  `Customers`: number of actual paying customers
- `Avg Order Value ($)`: average revenue per transaction
- `Daily Sales ($)`: total sales for the day
- `Avg Customer Rating`: customer satisfaction on a 5-star scale
- `Local Competition Count`: how many other coffee shops operate nearby

2. `Coffee_Shop_Demographics.csv`

This file contains **external market data** for each city where the stores operate.

Key Features:

- `City`: matches to the store data
- `Population`: total residents in the city
- `Median Income ($)`: a proxy for spending power
- `Young Adult % (20–35)`: represents the primary target customer group

Why This Setup Matters

Real data analysis often requires **merging internal performance metrics with external market factors**. You'll need to:

- Join these datasets correctly
- Create your own KPIs to evaluate performance
- Use trends, averages, and ratios to tell the full story
- Think beyond the numbers: "*What would I tell a stakeholder based on this data?*"

— Directions

Step 1: Load & Explore the Data

Why: Analysts need to understand the **structure and health** of data before making conclusions. This is your first look at what you're working with.

You need to:

- Load both CSVs (sales and demographics)
- Explore shape, columns, types, and summary stats

Hints:

- Use `pd.read_csv()` to load files
- Use `.info()` to check data types and nulls
- Use `.describe()` to get numeric summaries
- Use `.head()` to visually inspect rows

Step 2: Merge Datasets

Why: To evaluate a store's performance **in context**, you must combine internal (sales) and external (market) data.

You need to:

- Join demographic info into your sales dataset using the `City` column

Hints:

- Use `.merge()` with `on='City'`
- Use `.info()` afterward to ensure the merge worked as expected (no missing columns)
- Check the result with `.head()`

Step 3: Engineer Key Metrics (KPIs)

Why: Raw numbers don't tell the full story. KPIs help you understand **behavior and efficiency**.

You need to create:

- **Conversion Rate**: How many visitors become paying customers?
- **Revenue per Visitor**: How much value does each visitor generate?
- **Revenue per Customer**: How much does each customer spend?

Hints:

- Use basic math (`/`) between columns
- Name your new columns clearly (`'Conversion Rate'`, etc.)
- Check for any potential division by zero
- Use `.head()` or `.describe()` to validate your results

Step 4: Analyze Performance at the Store Level

Why: Stores are your units of execution. Understanding which are thriving (or failing) tells you what works.

You need to:

- Group your data by `Store ID`
- Calculate the average for sales, traffic, KPIs
- Sort stores to identify top performers

Hints:

- Use `.groupby()` + `.agg()` to calculate multiple KPIs
- Use `.reset_index()` to flatten the result
- Use `.sort_values()` to rank stores by a key metric

Step 5: Aggregate to the City Level

Why: Your decision is about cities not stores. You need to **roll up** your metrics to that level.

You need to:

- Group by `City`
- Calculate average KPIs across stores in that city
- Include **demographic data** (population, income, young adult %)

Hints:

- Use `.groupby('City')` and `.agg()`
- For demographic data, use `'first'` if the value is the same for every row
- Look for high `Revenue per Visitor`, high `Customer Value`, and growing customer base

Step 6: Visualize Monthly Trends

Why: Averages are helpful, but **trends show growth, decline, or volatility**. You don't want to expand into a declining market.

You need to:

- Analyze how sales change month-to-month in each city
- Create a line chart of average sales over time

Hints:

- Use `.dt.to_period('M')` to extract month
- Group by both `Month` and `City`
- Use `.unstack()` to pivot the data so each city is its own line
- Plot the result with `.plot()`

Step 7: Make a Recommendation

Why: This is where you bring it all together just like you would in a real job.

You need to:

- Choose the **best city** for a new store
- Support your choice using:
    - 2–3 key KPIs
    - Trend insights (growing, stable, declining)
    - Demographic fit (income, age group, population)
    - Competition risk

Hints:

- Revisit your aggregated metrics
- Look at your visualizations, what patterns stand out?
- Think beyond numbers: Which city is **sustainable** and **scalable**?