Mercado Libre

Advanced Analytics Business Case

Business Context

You are a data scientist at a growing e-commerce platform that sells **pharmaceutical products** and **consumer electronics** supplied by multiple third-party vendors. The company has enjoyed steady growth, but recently there are concerns that growth is slowing down in certain areas.

Some vendors have reported **decelerating sales trends** for their products over the past year. Your task is to investigate the data, quantify any slowdown, and recommend a data-driven plan of action. This exercise simulates an end-to-end project: from diagnosing the problem through analysis, to forecasting its impact, and finally designing an experiment to address it.

Data Overview

You are provided with a **single-table dataset (dataset.csv)** containing **historic order data** for two years (January 2022 – December 2023). The dataset has one row per order line item (each product in an order is a row) with the following columns:

- **order_id** – Unique identifier for the order.
- **order_date** – Date of the order (timestamp format).
- **customer_id** – Unique customer identifier.
- **product_id** – Unique product identifier.
- **product_category** – Category of the product, either "Pharma" or "Electronics".
- **vendor_id** – Identifier for the vendor (third-party seller).
- **vendor_name** – Name of the vendor (e.g., Vendor A, Vendor B, etc.).
- **price** – Unit price of the product in the order (in USD).
- **quantity** – Quantity of the product purchased in the order.
- **revenue** – Total revenue for that line (price * quantity).

The data spans at least 24 months, capturing seasonal peaks and troughs in sales (for example, holiday season upticks, or seasonal demand for pharmaceuticals). Starting around mid-2023, one product category shows signs of a **declining trend** in sales. Multiple vendors (Vendor A, B, C, D) are represented, each with products in one or both categories. There are approximately 5,000–10,000 rows of data in total. Use this data to fuel your analysis and insights.

Your Challenge

Imagine you've been asked to investigate why overall sales growth is slowing and to propose a plan to address it. This will involve **exploratory analysis, forecasting, and experiment design**:

- **Exploratory Analysis:** Start by examining the historical sales data. Look for trends, seasonal patterns, and any anomalies or deviations. Identify if any specific product category (Pharma vs. Electronics) or any vendor is experiencing a **slowdown or decline in sales**. For example, are electronics sales dropping relative to last year? Is a particular vendor's performance deteriorating? Summarize key insights with supporting data (e.g., charts or statistics). This EDA will help you form a hypothesis about what is happening and why.

- **Forecasting:** Based on what you discover, choose a reasonable approach to forecast future sales for the affected area. For instance, if you suspect the **Electronics category** is declining, build a model to forecast the next 12 months of electronics sales (you may use any tool or model you think is appropriate, such as time-series modeling). The goal is to **quantify the trend** – confirm if the category is indeed declining and estimate the potential impact over the next year if no action is taken. State any assumptions and document your methodology (e.g., did you seasonally adjust the data, which model you used, how you evaluated forecast accuracy, etc.).

- **Intervention Experiment Design:** Management is considering actions to counteract the slowdown (for example, a pricing discount on electronics or a targeted marketing campaign). Design an **A/B test** to evaluate the effectiveness of such an intervention. Choose one concrete intervention (e.g., a 10% price cut on a subset of products, or an email marketing promotion) aimed at improving the slowing trend. Your test design should include:

    - A clear **hypothesis** (e.g., "Offering a 10% discount on electronics will increase the average weekly sales by 15% for that category").

    - Identification of **key metrics** to measure success (e.g., revenue uplift in electronics, conversion rate, etc.).

    - **Experiment setup** details: how you would split the groups (treatment vs control), ensure randomization, and run the test (for example, which products or customers get the discount).

    - A **power analysis / sample size calculation** to determine how long or how large the experiment needs to be to detect a meaningful effect. Consider the expected effect size and variability in sales.

    - Discussion of **costs or risks** (e.g., revenue lost due to discounting during the test, or potential impact on customer perception).

    - Potential **outcome scenarios** and how you would interpret them. For instance, what would you conclude if the test shows no effect vs. a positive effect? How would it influence next steps?

Throughout the exercise, make sure to connect your analysis to the **business context**: Why do these trends matter? How would your findings inform business decisions? We encourage you to think about **possible causes** for the trend (e.g., competition, market saturation in electronics, changes in customer behavior, etc.) and incorporate those considerations into your recommendations.

Deliverables

Prepare the following deliverables as if you were presenting your findings to both technical colleagues and business stakeholders:

- **Data Analysis Code:** Provide any SQL queries (or well-documented pseudocode) you used to extract or aggregate data for your analysis. Additionally, include your analysis code (Python notebooks, R scripts, etc.) used for EDA and forecasting. This should show how you analyzed the data, including any visuals you created (you may include rendered charts in the notebook or as separate image files if applicable).

- **Forecast Model Documentation:** Briefly document your forecasting approach and results. This could be included in the notebook or a separate markdown/pdf. Explain which model or method you chose and why, and include the 12-month forecast results (table or chart) quantifying the expected decline or trend. If you tried multiple approaches, you can mention how you selected the final model.

- **Presentation Summary (PDF or Slides):** A succinct slide deck or report (PDF) that summarizes your key findings and recommendations. This should be presentation-quality:

    - Introduce the problem and context.

    - Show key insights from the EDA (e.g. trend graphs, seasonal patterns, the identified decline).

    - Highlight the forecasted impact (e.g., "Projected 12-month sales for Electronics will be X% lower if the trend continues").

    - Propose the A/B test intervention and detail the experiment design (hypothesis, duration, expected outcome).

    - Include your recommendations on what the company should do, based on these findings. For example, if the test is successful, how should it be rolled out? If the forecast looks

bad, what strategies might improve the situation?

The slide deck should be understandable to a **non-technical executive**, focusing on the "so what" — the business implications and your proposed action. You can assume the audience will trust that your technical analysis was done correctly, so focus on insights and recommendations.

Tips for Success

- **Start with the data:** Let the data guide you. There may be many things you could analyze, but focus on what seems most pertinent (e.g., declining category sales). Investigate the timing and magnitude of any slowdown.

- **Think end-to-end:** Show how the EDA informs the forecasting, and how the forecast informs the need for an intervention. Each step should build on the previous.

- **Justify your choices:** There is no single "right" model or test design. What matters is that you explain why your approach is reasonable and what assumptions you made.

- **Practical considerations:** When designing the experiment, be realistic. Consider constraints like seasonality (e.g., avoid running a short test over an abnormal period like Black Friday if it skews sales), and operational limitations (can all vendors apply a discount? what if only some products are discounted?).

- **Communication:** Clear communication is part of the evaluation. Structure your analysis and presentation in a logical way. Define any technical terms and avoid jargon in the executive summary.

- **Business recommendations:** Remember to tie your results back to business decisions. If you conclude a category is declining, the business wants to know "what should we do about it?" Your A/B test design and any other suggestions should address that.

Good luck, and have fun with the analysis! We are looking forward to seeing how you connect data insights to actionable business outcomes.

---

dataset.csv

Below is a **preview of the dataset** (the first few rows) to illustrate the format and content. The full dataset file contains all orders from Jan 2022 through Dec 2023.

```
Unset
order_id,order_date,customer_id,product_id,product_category,vendor_id,vendor_name,price,quanti
ty,revenue
10001,2022-01-01,C1097,P205,Pharma,V004,Vendor D,29.60,5,148.00
10002,2022-01-01,C1051,P261,Pharma,V002,Vendor B,39.01,2,78.02
10003,2022-01-02,C1017,E179,Electronics,V002,Vendor B,409.73,5,2048.65
10004,2022-01-02,C1039,E193,Electronics,V004,Vendor D,83.18,3,249.54
10005,2022-01-03,C1012,P255,Pharma,V004,Vendor D,35.04,2,70.08
10006,2022-01-03,C1056,P207,Pharma,V001,Vendor A,81.48,5,407.40
10007,2022-01-03,C1011,P290,Pharma,V004,Vendor D,83.36,1,83.36
10008,2022-01-04,C1031,P290,Pharma,V002,Vendor B,87.69,2,175.38
10009,2022-01-04,C1030,E169,Electronics,V003,Vendor C,251.59,1,251.59
10010,2022-01-05,C1062,E138,Electronics,V003,Vendor C,298.07,1,298.07
... (continued for thousands more rows) ...
```

Each order_id may appear multiple times if an order contains multiple products (each line is a product in the order). The order_date is the date of purchase. price is the unit price at time of order, and revenue is pre-calculated for convenience. The data is already cleaned and ready for analysis, but feel free to do any additional preprocessing as needed (e.g., converting dates, aggregating by month, etc.).

Hint: To analyze trends, you might want to aggregate this data by month, category, or vendor. For example, summing revenue by month for each product_category could help reveal overall trends and seasonal patterns.