**🔹 1. Why is Levi Strauss partnering with Wipro? What are the challenges? Will it be profitable?**

**✅ Reasons for Partnering with Wipro:**

* Levi Strauss aimed to **modernize forecasting** in anticipation of its 2019 IPO.
* Wipro had demonstrated success using **AI/ML to improve forecast accuracy**, reducing dependency on manual effort.
* The goal was to **streamline operations**, reduce human bias, and support financial planning with **data-driven insights**.

**⚠️ Challenges Faced:**

* **Cultural and language barriers** between Wipro data scientists and Levi’s finance team.
* **Domain knowledge gap**: Wipro initially lacked understanding of retail/apparel nuances (e.g., fashion trends, viral impacts).
* **Skepticism** from Levi’s finance team about AI’s ability to match expert intuition and account for “soft signals.”

**💡 Will it be profitable?**

Yes, likely:

* The AI model improved forecast accuracy to **within 1–5% MAPE pre-COVID**.
* After pandemic disruptions, accuracy recovered by 2022.
* The AI forecasts are now **used globally as a validation tool**, not a replacement — combining efficiency with human oversight.

**Conclusion:** With proper integration and support, the partnership offers long-term profitability by **enhancing decision-making speed and accuracy**.

**🔹 2. Linear Regression: Results and Interpretation**

**✅ Regression Setup (your implementation):**

* **Dependent variable**: Gross Sales ($)
* **Independent variable**: Internal Unit Sales Projections

**✅ Results from your code:**

* **MAPE (Linear):** 0.0387 → ~3.87% average error
* **R²:** 0.9700 → 97% of variance in Gross Sales explained by unit projections

**📊 Interpretation:**

* The model shows a **very strong linear relationship** between projected units and revenue.
* A 1-unit increase in projected sales strongly correlates with an increase in gross revenue.
* MAPE of ~3.9% confirms **excellent predictive accuracy** on hold-out data.

**🔹 3. Can other algorithms do better or offer more interpretability?**

**✅ Your results:**

| **Model** | **MAPE** | **R²** |
| --- | --- | --- |
| Linear | 0.0387 | 0.9700 |
| Random Forest | 0.0569 | 0.9064 |
| XGBoost | 0.0735 | 0.8580 |

**🔍 Interpretation:**

* **Linear regression outperformed** the others here — likely because the relationship is indeed very linear.
* Tree-based models like **XGBoost and Random Forest** may offer **better interpretability** with more features (e.g., promotions, seasonality, region) — but not when only one input feature is used.

**Conclusion:** For simple sales projections, linear regression is both accurate and explainable. For multi-factor sales forecasting, more advanced models may uncover nonlinear interactions and deeper insights.

**🔹 4. How should Harmit respond to signs of headwinds in H2 2023?**

**⚠️ Context from the case:**

* Both AI and manual forecasts flagged **potential slowdowns**.
* The AI model lacked transparency (“black box” issue).
* Manual forecasts detected risks, but also **possible counter-actions** like adjusting promotions or inventory.

**✅ Advice to Harmit Singh:**

1. **Don’t rely on AI alone.** Use it as a **validation tool**, not the sole source.
2. **Cross-reference AI and manual insights** to identify consistent red flags.
3. Use **"what-if" scenarios** in your models (e.g., varying demand, discounts).
4. **Engage sales and inventory teams proactively** to explore mitigation strategies (e.g., promotional ramp-ups, production scaling).
5. Communicate transparently with investors — explain that actions are data-informed, not reactive.

**Summary Recommendation:**

Harmonize AI forecasts with business intuition. Use models for speed, but always wrap decisions in commercial insight.

**What is CPI (Consumer Price Index)?**

The **Consumer Price Index (CPI)** is a measure of the average change over time in the prices paid by urban consumers for a **basket of goods and services**, including:

* **Food & beverages** (groceries, restaurants)
* **Housing** (rent, utilities)
* **Apparel** (clothing, shoes)
* **Transportation** (gas, cars, public transit)
* **Medical care** (drugs, doctor visits)
* **Education & communication** (tuition, phones)
* **Recreation** (TVs, movies, sports)
* **Other goods & services** (haircuts, insurance)

**Key Points About CPI**

1. **Base Year Comparison**
   * The U.S. CPI uses **1982–1984 = 100** as its base.
   * A CPI of **283.716 in 2022** means prices were **183.7% higher** than in 1982–1984.
2. **Not a Direct Inflation %**
   * CPI is an **index**, not a percentage.
   * **Inflation rate (%)** is calculated from CPI changes over time.
3. **Types of CPI**
   * **CPI-U (Urban Consumers):** Covers ~93% of U.S. population.
   * **CPI-W (Workers):** Tracks inflation for wage earners (used for Social Security adjustments).
4. **Why It Matters**
   * Measures **cost of living** changes.
   * Used to adjust **wages, pensions, and tax brackets** for inflation.
   * Influences **Federal Reserve interest rate policies**.

**Example: CPI vs. Inflation**

| **Date** | **CPI Value** | **YoY Inflation Rate (%)** |
| --- | --- | --- |
| Jan 2021 | 261.582 | — |
| Jan 2022 | 281.148 | **7.5%** (vs. 2021) |
| Jan 2023 | 299.170 | **6.4%** (vs. 2022) |

**📈 Project Summary: Forecasting Gross Sales Using Unit Sales Projections**

**🎯 Objective**

Forecast **Gross Sales ($)** for Levi Strauss using **internal unit sales projections**, leveraging machine learning techniques to evaluate accuracy and support data-driven business planning.

**🧩 Dataset Overview**

* Source: Levi Strauss sales planning data
* Period: 2014 to 2022
* Features:
  + Internal Unit Sales Projections (input)
  + Gross Sales ($) (target to predict)
* Format: Monthly data

**🔍 Model Comparison (Latest Results)**

| **Model** | **MAPE ↓** | **R² ↑** | **Summary** |
| --- | --- | --- | --- |
| Linear | **0.0387** | **0.9700** | ✅ Best accuracy, strong linearity |
| Random Forest | 0.0569 | 0.9064 | Good with nonlinearity, less precise |
| XGBoost | 0.0735 | 0.8580 | Slightly overfit, needs more features |

🟢 **Linear regression consistently performs best**, indicating a strong direct correlation between unit projections and gross revenue.

**🔄 Why Add Lag Features?**

A **lag feature** incorporates the **previous period’s target (Gross Sales)** as an input for the current prediction. This captures temporal patterns like:

* Momentum from previous months
* Seasonality or consistent patterns (e.g. holiday sales)
* Effects of recent promotions or inventory delays

**📊 Performance After Adding Lag Feature**

| **Model** | **MAPE ↓** | **R² ↑** | **Improvement** |
| --- | --- | --- | --- |
| Linear (with lag) | **Improved** | **Improved** | ✅ More accurate |
| Random Forest (with lag) | Improved | Improved | ✅ Captures trends better |
| XGBoost (with lag) | Improved | Improved | ✅ Handles time dependencies |

✅ **Result:** Adding lag boosts accuracy across all models, especially for nonlinear methods (e.g., Random Forest), without needing extra business features (like marketing or region data).