

Proyecto final: el precio de Café Altura

Durante el curso exploraste el panel de ventas, estimaste regresiones y las diagnosticaste, modelaste series de tiempo, respuestas binarias y datos panel. Ahora juntas todo para responder una pregunta de negocio con una recomendación defendible. Café Altura es una cadena ficticia y todos los datos son simulados. Se usan Python 3.11 o superior, pandas, statsmodels 0.14 y linearmodels. El notebook proyecto-precio-optimo contiene una solución de referencia: consúltalo al terminar, no antes.

Contexto

Café Altura tiene 40 sucursales en 8 ciudades de México. En 2025 subió sus precios y vendió menos tazas que en 2024. La dirección te encarga un análisis con una pregunta central: **¿qué precio conviene cobrar?** Quiere saber cuánto cae la demanda cuando sube el precio, si la subida de 2025 fue un error y qué hacer en adelante. Producir cada bebida cuesta **18 MXN** (supón ese costo unitario constante).

Datos

Copia las funciones de datos de la primera celda de cualquier notebook del curso (generar_tiendas, generar_ventas, generar_panel, generar_mensual y generar_clientes), sin cambiar las semillas:

```
df = generar_panel()
print(df.shape)                # (4160, 13)
print(df["tienda"].nunique())   # 40
print(df["semana"].min(), df["semana"].max())
```

El panel cubre 104 semanas, del 2024-01-01 al 2025-12-22. Si no obtienes esas cifras, revisa las semillas. generar_mensual() trae las ventas mensuales de la cadena de 2014 a 2023 y generar_clientes() una encuesta de 2500 clientes. No uses el diccionario VERDAD: tu análisis debe llegar a las cifras por sí mismo.

Preguntas obligatorias

1. **Diagnóstico descriptivo.** Cuantifica el cambio de precio y de tazas entre 2024 y 2025, total y por ciudad. Muestra la relación entre precio y ventas entre tiendas y dentro de cada tienda, y explica por qué difieren.

2. **La regresión ingenua.** Estima $\log_{\text{ventas}} \sim \log_{\text{precio}}$ por MCO e interpreta el coeficiente. Explica, con el concepto de sesgo por variable omitida, por qué no sirve para decidir el precio.
3. **Una elasticidad confiable.** Estima la elasticidad con datos panel y efectos fijos por tienda (`PanelOLS` con `EntityEffects`), con controles de promoción, temperatura y publicidad. Reporta el coeficiente, su intervalo de confianza al 95 % y su interpretación.
4. **Diagnóstico del modelo.** Justifica la elección de efectos fijos frente a pooled y efectos aleatorios con al menos dos pruebas (por ejemplo, la F de los efectos y el enfoque de Mundlak). Usa errores estándar por cluster de tienda y muestra cuánto cambian frente a los clásicos.
5. **Cuánto explica el precio.** Con tu elasticidad, calcula qué parte de la caída de tazas de 2025 se debe a la subida de precio.
6. **Recomendación de precio.** Aplica la regla de Lerner, $(p^* - c) / p^* = -1 / e$, con un costo unitario $c = 18$ MXN. Calcula el precio óptimo p^* y el rango que resulta de usar los extremos del intervalo de la elasticidad. Compara con el precio medio de 2025.
7. **Ingreso contra utilidad.** Simula al menos cuatro cambios de precio (por ejemplo, -10 %, -5 %, +5 % y +10 %) y reporta el efecto en tazas, ingreso y utilidad. Explica por qué un escenario puede subir el ingreso y bajar la utilidad.

Extras opcionales

- **Pronóstico 2024.** Con `generar_mensual()` en logaritmo, identifica un SARIMA con ACF/PACF y AIC, valida en un holdout de 12 meses (2023) contra un naive estacional con MAPE, revisa Ljung-Box y pronostica 2024 con intervalos.
- **Clientes y café en grano.** Con `generar_clientes()`, estima logit y probit de `compra_grano`, reporta efectos marginales promedio y propone a qué clientes dirigir una campaña.

Entregables

1. `proyecto.ipynb`, ejecutado de principio a fin sin errores tras reiniciar el kernel, con una celda de texto por pregunta.
2. **Informe de 2 páginas** (PDF o Markdown) para la dirección: pregunta, datos, método en lenguaje no técnico, cifras clave, recomendación y tres limitaciones.
3. **Presentación de 5 diapositivas** para la dirección, con la estructura de abajo.

Estructura sugerida de la presentación

1. **La pregunta y el hallazgo:** qué preguntó la dirección y la respuesta en una frase con cifra.

2. **Por qué el dato crudo engaña:** una gráfica de precio contra ventas entre tiendas y dentro de cada tienda.
3. **La elasticidad:** el valor con efectos fijos, su intervalo y qué significa en tazas.
4. **Escenarios:** tabla de tazas, ingreso y utilidad para cada cambio de precio, con el precio de Lerner.
5. **Recomendación y riesgos:** qué hacer, cómo comprobarlo (por ejemplo, un piloto en algunas tiendas) y qué supuestos podrían fallar.

Escribe para alguien que no conoce la econometría: una idea por diapositiva y cifras redondeadas.

Rúbrica (100 puntos, más 10 de extras)

- **Diagnóstico descriptivo (10 puntos):** cifras 2024-2025 correctas y la paradoja entre y dentro de tiendas explicada.
- **Regresión ingenua y sesgo (10 puntos):** interpretación correcta y explicación del sesgo por variable omitida.
- **Elasticidad con efectos fijos (20 puntos):** especificación correcta, intervalo e interpretación.
- **Diagnóstico del modelo (15 puntos):** pruebas pertinentes, errores por cluster y conclusión clara.
- **Precio y escenarios (20 puntos):** regla de Lerner bien aplicada, rango por incertidumbre e ingreso separado de utilidad.
- **Informe (10 puntos):** claro, con cifras y limitaciones honestas, sin inventar causas.
- **Presentación (10 puntos):** cinco diapositivas que se entienden sin el notebook.
- **Reproducibilidad (5 puntos):** el notebook corre completo y las semillas no cambiaron.
- **Extras (hasta 10 puntos):** pronóstico validado o modelo de clientes con efectos marginales.