

Proyecto de Análisis de Datos

Autor: Agustín Peri

Fecha: 16/04/2025

Introducción y contexto

"Empanadas Don Pedro" es una cadena familiar de empanadas fundada por Doña Carmen y Don Pedro hace 15 años. Lo que comenzó como un pequeño negocio local se ha expandido a 8 sucursales en diferentes barrios de la ciudad. Sus hijos Miguel (gerente general), Lucía (chef principal) y Daniel (responsable de marketing) se han unido al negocio familiar.

En los últimos dos años, enfrentaron diversos desafíos:

1. Fluctuaciones inesperadas en las ventas entre sucursales.
2. Dificultad para predecir la demanda de diferentes tipos de empanadas.
3. Dificultad para identificar la causa de las malas puntuaciones.
4. Resultados mixtos de sus campañas promocionales.
5. Incapacidad de tener un registro centralizado del estado de la compañía.

La familia decidió implementar un enfoque basado en datos para optimizar sus operaciones y mejorar su rentabilidad. Sin embargo, encontraron dificultades en la carga y unificación de datos debido a la descentralización de los mismos.

Problemática principal: Optimizar la gestión de la cadena de empanadas utilizando análisis de datos para mejorar la rentabilidad, predecir la demanda, optimizar el inventario y personalizar estrategias de marketing.

Preguntas específicas a resolver:

- ¿Cuál es la sucursal con mayor recaudación?
- ¿Cómo han evolucionado las ventas a lo largo del tiempo?
- ¿Cuál es el método de pago más utilizado?

- ¿Existen diferencias en las calificaciones entre sucursales?
- ¿Qué tipo de empanada se vende en mayor cantidad?

Análisis y exploración del dataset

El dataset consta de 500 registros y 10 columnas: 4 numéricas y 6 categóricas.

Columna	Tipo de dato	Descripción
fecha_venta	object	Fecha en que se realizó la venta. Fundamental para análisis temporal.
tipo_empanada	object	Sabor o variedad de la empanada vendida. Permite conocer preferencias.
sucursal	object	Local donde se realizó la venta. Permite evaluar rendimiento por tienda.
medio_pago	object	Forma de pago. Permite entender hábitos de consumo.
metodo_compra	object	Canal de compra: Presencial, Delivery o Web.
cliente	object	Tipo de cliente: Nuevo o Habitual.
unidades	int64	Cantidad de empanadas vendidas. Variable objetivo secundaria.
precio_total	float64	Importe total pagado. Clave para ingresos y facturación.
calificacion	float64	Puntuación del cliente (1 a 5). Indica satisfacción.
tiempo_entrega_min	float64	Tiempo de entrega en minutos. Indica eficiencia.

Tratamiento de variables

- Conversión de `fecha_venta` a formato datetime para facilitar análisis temporal.
- **Estandarización de etiquetas:**
 - "Sucursal Palermo" → "Sucursal Norte".
 - "MercadoPago" → "Transferencia".
- **Imputación de valores faltantes:**
 - `calificacion`: media por `metodo_compra`.
 - `tiempo_entrega_min`: media por `sucursal`.

Análisis exploratorio

Calificaciones: mayoría entre 3 y 5.

Evolución de ventas: tendencia estable.

Recaudación por sucursal: la más alta fue Sucursal Norte.

Tipos de empanadas más vendidas: Humita y Carne picante.

Respuestas a preguntas claves:

Sucursal más rentable: Sucursal Norte.

Evolución de ventas: Estabilidad con fluctuaciones bimensuales.

Relación entre espera y calificaciones: No hay relación significativa.

Promedio de puntuaciones por sucursal: Entre 3.1 (Centro) y 3.7 (Oeste).

Método de compra más elegido: Presencial.

Medio de pago más elegido: Tarjeta de débito.

Preprocesamiento y desarrollo del modelo

- Objetivo: Predecir cantidad de unidades vendidas.
- Variables predictoras: tipo_empanada, sucursal, medio_pago, cliente, metodo_compra.
- Modelo: **Random Forest Regressor**.
- Métrica de evaluación:
 - **RMSE:** ≈ 1.58 (error promedio aceptable).
 - **R²:** ≈ 0.777 (modelo explica el 77.7% de la variabilidad).

Conclusión: El modelo es efectivo para estimar unidades vendidas. Con más datos, validación cruzada y enriquecimiento de variables se podría mejorar la precisión de la predicción.