El análisis de datos es el proceso de inspeccionar, limpiar, transformar y modelar datos con el objetivo de descubrir información útil, llegar a la conclusiones y apoyar la toma de decisiones. En pocas palabras uno se tiene que convertir grandes volúmenes de datos en **información útil** para la **toma de decisiones**.

El grafico representa el porcentaje de las ventas totales (en pesos) por frecuencia de compra, en cada ubicación y tipo de cliente.

Que es lo que representan las líneas en el grafico

* **Líneas más largas:** Mayor variación dentro del grupo. Es decir, algunos clientes en ese grupo compran mucho más o mucho menos que el promedio.
* **Líneas más cortas:** Los clientes del grupo se comportan de manera más parecida (menos variación).
* Si las barras de error de dos grupos se superponen mucho, puede que **no haya una diferencia estadísticamente significativa** entre ellos.
* Gráfico

  El contenido generado por IA puede ser incorrecto.

**🧩 ¿Qué elementos incluye?**

* **Eje X (horizontal)**: Las **ubicaciones** donde se realizan ventas (Corral, Las Ánimas, Niebla, Valdivia, Los Molinos).
* **Eje Y (vertical)**: El **porcentaje de ventas (%)** realizadas en cada ubicación.
* **Colores**: Cada color representa una **frecuencia de compra**:
  + 🟦 Azul = **Mensual**
  + 🟧 Naranjo = **Quincenal**
  + 🟩 Verde = **Semanal**
* **Barras**: Altura = porcentaje de ventas para esa frecuencia de compra en esa ubicación.
* **Líneas negras encima de las barras** = *intervalos de error o desviación estándar*, indicando **variabilidad** en los datos.

**🔍 ¿Qué puedes observar?**

* En **Niebla**, la mayoría de los clientes compra **mensualmente** (barras azules son más altas).
* En general, **la frecuencia mensual domina en todas las ubicaciones**.
* Las compras **semanales** (verde) son menos frecuentes en todas las ubicaciones.
* Algunas ubicaciones como **Las Ánimas** y **Los Molinos** tienen una distribución más pareja entre mensual y quincenal.
* Las **barras con líneas más largas** indican **más variación** entre los clientes en esa categoría.

**🧠 ¿Cómo te sirve esto?**

* Puedes usar esta info para **segmentar promociones**:
  + Clientes **mensuales** → promociones por volumen o fidelización.
  + Clientes **semanales o quincenales** → promociones de urgencia o combos rápidos.
* Identificar zonas donde podrías incentivar **más frecuencia de compra** (por ejemplo, clientes quincenales que podrían volverse semanales).

**“¿Qué puedes analizar con este gráfico?**

1. **Preferencias de compra por zona:**
   * ¿En qué lugares se concentra más la compra semanal? ¿O la mensual?
   * Por ejemplo, si en *Valdivia* el 80% de las ventas son de clientes **mensuales**, puedes adaptar tus promociones a ese ritmo.
2. **Oportunidades de fidelización:**
   * Si en *Niebla* hay muchos clientes que compran quincenalmente, quizás puedas incentivarlos a comprar semanal con descuentos.”

Gráfico, Gráfico de barras

El contenido generado por IA puede ser incorrecto.

**1. Ventas totales por tipo de cliente**

* **Gráfico:** Barplot
* **Objetivo:** Ver quién compra más (minimarkets, panaderías, restaurantes, etc.)

**2. Mapa de calor de ventas por ubicación y tipo de cliente**

* **Gráfico:** Heatmap
* **Objetivo:** Ver dónde estás vendiendo más y qué tipo de cliente predomina por zona.

**📊 Ingreso Total por Tipo de Cliente**

**🧩 ¿Qué elementos incluye?**

* **Eje X (horizontal)**: Muestra los distintos **tipos de cliente**:
  + Consumidor minorista
  + Minimarket
  + Panadería
  + Restaurante
  + Tienda Pequeña
* **Eje Y (vertical)**: Representa el **ingreso total en pesos ($)** que generó cada tipo de cliente.
* **Altura de las barras**: Refleja cuánto ingreso total aportó ese segmento de cliente.

**🔍 ¿Qué puedes observar?**

* **Panaderías** generan el ingreso más alto, con más de **7 millones de pesos**.
* Le sigue **Tienda Pequeña**, también con un ingreso significativo (más de 6 millones).
* **Consumidor minorista** es el que **menos aporta** al ingreso total.
* **Minimarket y Restaurante** están en un rango medio de contribución.

**🧠 ¿Cómo te sirve esto?**

* Te permite **priorizar esfuerzos comerciales**:
  + Las **panaderías** son tu **cliente más rentable**: conviene fidelizarlas, entender qué compran, y potenciar ese segmento.
  + Las **tiendas pequeñas** también son un buen mercado: podrías ofrecer paquetes personalizados o descuentos por volumen.
  + **Consumidores minoristas** generan poco ingreso, por lo que invertir mucho en ese segmento podría no ser tan rentable.

1. **Ventas de harina por ubicación y tipo de cliente**

Gráfico

El contenido generado por IA puede ser incorrecto. **Mapa de calor**

**🧩 ¿Qué elementos lo componen?**

* Eje Y (filas) → Representa las ubicaciones donde se realizan las ventas:  
  Corral, Las ánimas, Los molinos, Niebla, Valdivia
* Eje X (columnas) → Representa los distintos tipos de cliente:  
  Consumidor minorista, Minimarket, Panadería, Restaurante, Tienda Pequeña
* Color de las celdas:  
  Cuanto más oscuro el color (más hacia el azul marino), mayor es el valor de las ventas de harina ($).  
  Colores más claros indican ventas más bajas.
* Números dentro de las celdas:  
  El valor exacto en pesos ($) de las ventas de harina para esa combinación de ubicación y tipo de cliente.

🔍 **¿Qué se observa?**

* La panadería en Corral es, con diferencia, el cliente que más harina compra: $2,142,000.
* Valdivia y Niebla también tienen fuertes ventas de harina a panaderías.
* En general, las panaderías son los mayores compradores de harina, sin importar la ubicación.
* Otros tipos de cliente tienen ventas mucho más bajas y similares entre ubicaciones.

🧠 **¿Para qué te sirve este gráfico?**

* Para identificar oportunidades geográficas y comerciales.
* Por ejemplo:
  + Los molinos tienen ventas bajas en general: podrías investigar por qué.
  + Corral y Valdivia son puntos fuertes para ventas de harina, especialmente a panaderías: podrías fidelizar esos clientes y replicar la estrategia en otras zonas.
  + Restaurantes y minoristas no son grandes compradores de harina. Tal vez puedas diseñar una estrategia diferenciada o nuevos productos para ellos.

Gráfico, Gráfico de cajas y bigotes

El contenido generado por IA puede ser incorrecto.

**4. Comparación de márgenes por producto**

* **Gráfico:** Boxplot o caja
* **Objetivo:** Ver si ganas más con harina o con huevos (y ver la variabilidad del margen por cliente).
* **🔍 ¿Qué podemos leer en este gráfico?**
* **Ganancia Harina ($):**
* 📦 La mayoría de las ganancias están entre aproximadamente **$0 y $40.000** (eso es el "cuerpo" de la caja).
* ➖ La línea dentro de la caja es la **mediana** (valor medio), está cerca de los $15.000.
* 🧨 Hay muchos **valores atípicos** (puntos sobre la caja), incluso uno llegando a cerca de **$500.000**. Esto indica que **algunos clientes generan ganancias muy altas**, pero la mayoría no.
* **Ganancia Huevos ($):**
* 📦 La caja es más grande, lo que indica mayor **variabilidad**.
* ➖ La **mediana** está más alta que la de la harina (alrededor de $60.000), lo que sugiere que, en promedio, **los huevos generan más ganancia por cliente**.
* ❌ También hay valores negativos, lo que podría indicar **pérdidas** en algunos casos.
* 💡 Hay menos outlier’s que en la harina, lo que sugiere **menos clientes extremos**.

**🧠 ¿Qué puedes concluir como analista?**

1. **Huevos son más rentables en promedio**, pero también más variables (mayor dispersión).
2. **Harina tiene muchos clientes pequeños**, pero **unos pocos muy rentables**, lo que podrías aprovechar con **estrategias personalizadas** para ellos.
3. Hay **posibles pérdidas** en ambos productos que sería bueno investigar (¿malos precios? ¿Errores de costos?).

**Conclusión General del Análisis de Ventas**

**1. Frecuencia de Compra por Ubicación**

* Los clientes en **Niebla**, **Las Ánimas** y **Corral** muestran una **alta preferencia por compras mensuales**.
* Las compras **semanales** son menos frecuentes en todas las ubicaciones, lo que indica que la mayoría de los clientes **no requieren reposición constante** o hacen **compras más grandes y espaciadas**.

**Estrategia**: Ofrecer promociones o suscripciones para fomentar compras más frecuentes en ubicaciones clave.

**2. Ganancia por Producto (Boxplot)**

* Hay **mayor variabilidad en la ganancia por Harina**, con muchos **outliers** (casos excepcionales de mucha ganancia).
* Las **ganancias por Huevos** son más consistentes, aunque con una mediana superior.

**Estrategia**: Optimizar la venta de Harina, ya que tiene más potencial para grandes ganancias en casos puntuales. Podrías identificar qué clientes generan esos outliers para enfocarte en ellos.

**3. Ingreso Total por Tipo de Cliente**

* Las **Panaderías** generan el mayor ingreso total, seguidas por las **Tiendas Pequeñas** y **Restaurantes**.
* El **Consumidor minorista** representa un ingreso mucho menor.

**Estrategia**: Priorizar acciones comerciales con panaderías y tiendas pequeñas, que son los **clientes más rentables**.

**4. Mapa de Calor de Ventas por Ubicación y Tipo de Cliente**

* Las panaderías de **Corral**, **Niebla** y **Valdivia** son responsables de las **ventas más altas de harina**.
* Algunas combinaciones (como "Consumidor minorista" en "Los Molinos") tienen **cero o muy pocas ventas**.

**Estrategia**: Reforzar relaciones con panaderías en ubicaciones fuertes y **explorar zonas o segmentos con ventas bajas** para entender la causa (¿falta de cobertura? ¿baja demanda?).

**5. Frecuencia de Compra por Tipo de Cliente (Countplot)**

* Ciertos tipos de clientes (como **Minimarket** o **Tienda pequeña**) tienen patrones de compra repetidos, lo que puede indicar oportunidades de automatización o campañas personalizadas.

**Estrategia**: Usar la frecuencia de compra para planificar visitas, entregas y promociones según comportamiento.

**Modelos de predicciones sugeridos**

**1. Modelo de Predicción de Demanda (Regresión)**

**Objetivo:** Predecir cuánto va a comprar un cliente en el futuro.

* **Modelo recomendado:**
  + Random Forest Regressor (muy robusto con datos no lineales y con muchas variables)
  + Alternativas: XGBoost, Linear Regression si buscas algo más simple.
* **Variables de entrada (features):**
  + Tipo de cliente
  + Ubicación
  + Frecuencia de compra
  + Temporada o mes
  + Compras anteriores (volumen, monto)
  + Método de pago
* **¿Para qué sirve?**
  + Anticiparte al stock necesario
  + Ofrecer promociones justo antes de que el cliente compre
  + Detectar caídas de consumo

**Modelo de Segmentación de Clientes (Clustering)**

**Objetivo:** Agrupar clientes similares para ofrecer promociones personalizadas.

* **Modelo recomendado:**
  + K-Means Clustering
* **Variables de entrada:**
  + Volumen comprado (harina y huevos)
  + Monto gastado
  + Frecuencia de compra
  + Método de pago
* **¿Para qué sirve?**
  + Identificar grupos como:
    - Clientes grandes y frecuentes
    - Clientes ocasionales pero rentables
    - Clientes sensibles al precio
  + Diseñar estrategias distintas para cada segmento

**3. Modelo de Propensión a Comprar (Clasificación)**

**Objetivo:** Predecir si un cliente está cerca de volver a comprar o no.

* **Modelo recomendado:**
  + Logistic Regression
  + Random Forest Classifier
* **¿Para qué sirve:**
  + Enviar campañas de WhatsApp/email a clientes que están por quedarse inactivos
  + Prevenir la pérdida de clientes

**📣 4. Modelo de Recomendación de Productos (Recomendador tipo Amazon)**

**Objetivo:** Sugerir productos complementarios.

* **Ejemplo:**
  + Si compra harina 25kg → sugerir caja de huevos con descuento
* **Modelo recomendado:**
  + Collaborative Filtering o Market Basket Analysis (Apriori o FP-Growth) para descubrir reglas como:

"El 70% de los que compran harina de 10 kg también compran bandejas de 30 huevos."

**🧠 Estrategia combinada:**

| **Modelo** | **Mejora ventas** | **Aumenta clientes** |
| --- | --- | --- |
| Predicción de demanda | ✅ | 🟡 (te ayuda a retenerlos) |
| Segmentación | ✅ | ✅ |
| Propensión a comprar | ✅ | ✅ |
| Recomendación de productos | ✅ | 🟡 |