

Proyecto

El presente proyecto tiene como objetivo realizar un exhaustivo análisis de datos relacionados con las órdenes, clientes (afiliados por medio de una membresía) y tipos de comida de un grupo de restaurantes, con el fin de identificar patrones, tendencias y áreas de mejora que contribuyan a optimizar las operaciones y mejorar la experiencia del cliente.

Objetivos

- Realizar un análisis descriptivo para comprender la distribución y características de las órdenes, clientes y tipos de comida.
- Identificar patrones de consumo de los clientes, preferencias alimenticias y tendencias de ventas.
- Evaluar la rentabilidad de cada tipo de comida y su impacto en las ventas.
- Identificar los platos más populares y aquellos que presentan oportunidades de mejora en términos de rentabilidad y satisfacción del cliente.
- Proponer recomendaciones y estrategias basadas en los hallazgos del análisis para mejorar la eficiencia operativa, aumentar las ventas y fidelizar a los clientes.

Descripción

En este proyecto, se llevará a cabo un análisis de información recopilada de varios restaurantes, centrándose en datos relacionados como, el tipo de restaurante, ubicación del restaurante, información de los clientes, información de la orden, fecha y hora en que se realiza la orden. El objetivo principal es comprender el comportamiento de los clientes, identificar tendencias en las órdenes y obtener información valiosa sobre las preferencias de comida en diferentes ubicaciones.

Datos disponibles

En términos generales los datos disponibles incluyen:

- **Información del restaurante:** Tipo de restaurante, ciudad en donde se ubica y porcentaje de ingresos.
- **Información del cliente:** Datos demográficos de los clientes (género y ciudad de origen), recuento de pedidos al mes y presupuesto (límite de crédito por la membresía adquirida).
- **Información de la orden:** Detalles de las órdenes realizadas, incluyendo los productos pedidos, precio, fecha y hora de la orden.

Metodología

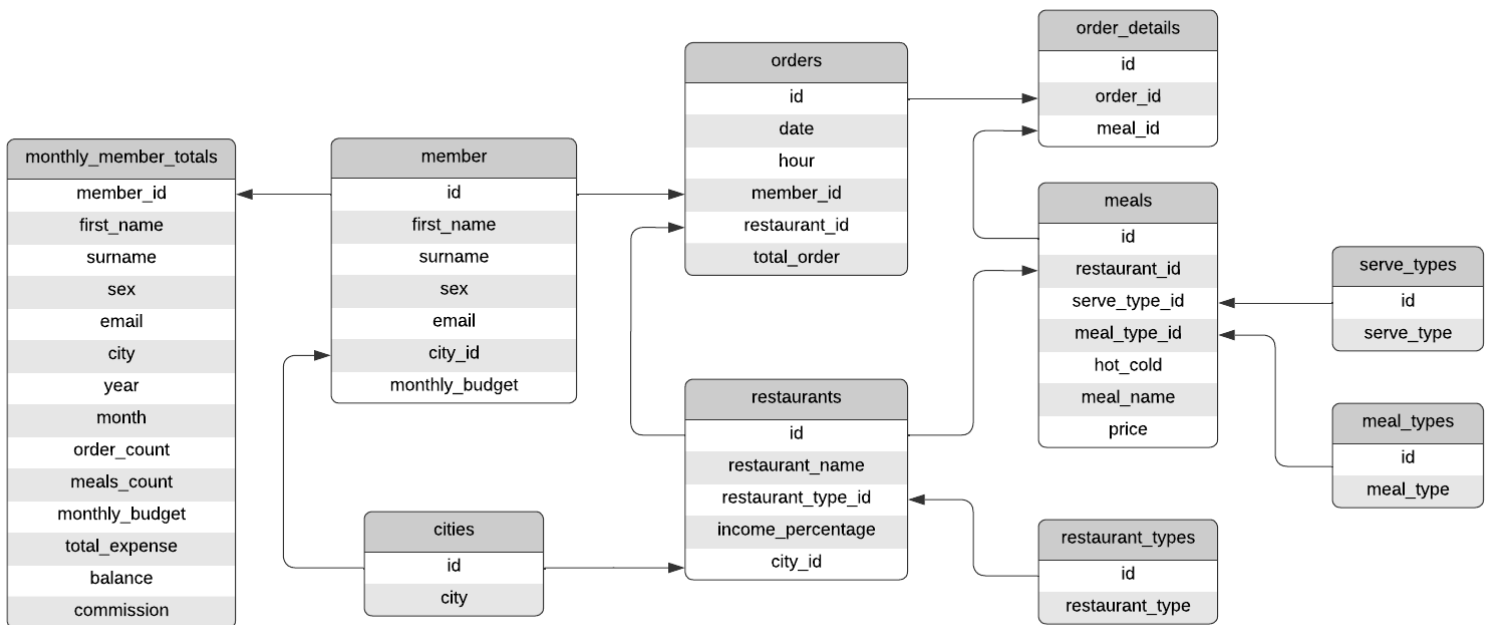
- **Preparación de datos:** Se cargarán los datos en un Notebook de R desde los diferentes conjuntos de datos proporcionados. Clic [aquí](#) para descargar.
- **Análisis descriptivo:** Se llevará a cabo un análisis descriptivo de los datos para comprender la distribución de variables, frecuencia de órdenes y tipos de comida más populares. Se utilizarán técnicas estadísticas para resumir y visualizar la información de manera efectiva.
- **Identificación de tendencias:** Se identificarán tendencias de consumo a lo largo del tiempo.

Pasos a desarrollar

1. Conociendo los datos

- 1.1. Cargue los paquetes `tidyverse`, `dplyr`, `ggplot2`.
- 1.2. Lea cada uno de los archivos `.CSV` que se compartieron usando la función `read_csv` y para cada uno defina una variable.
- 1.3. Imprima cada una de las variables que almacenan los datos, estos poseen una estructura como la que se presenta en la **Figura 1**. Como puede notar existe una relación entre los datos. Analice toda la información para comprender el contexto. En la sección de **Anexos** también puede encontrar un detalle de cada una de las variables.
- 1.4. También utilice las funciones `head` y `str` en cada variable para tener más contexto.

Figura 1. Estructura relacional de los datos



2. Transformando los datos

- 2.1. Luego de cargar los datos y asignarlos a variables, use aquella variable donde asigno el archivo "members.csv" y por medio del concepto de Factores transforme la columna `sex` a un factor y renombre los niveles "F" y "M" por "Femenino" y "Masculino" respectivamente.
- 2.2. Use la variable donde asigno el archivo "meals.csv" y por medio del concepto de Factores transforme la columna `hot_cold` a un factor y aplique un ordenamiento donde "cold" es menor que "hot". Valide usando una comparación (`cold < hot`).
- 2.3. Use la variable donde asigno el archivo "monthly_member_totals.csv" y por medio del concepto de Factores transforme la columna `sex` a un factor y renombre los niveles "F" y "M" por "Femenino" y "Masculino", haga lo mismo con la columna `month` y renombre los niveles "1", "2", "3", "4" y "5" por "Enero", "Febrero", "Marzo", "Abril" y "Mayo" respectivamente. Por último, agregue una nueva columna por medio de la función `mutate`, que calculará la proporción de órdenes (usando la variable "order_count") por la cantidad de comidas (usando la variable "meals_count").
- 2.4. Imprima la variable con los cambios aplicados y con la nueva columna.

3. Comportamiento de los datos

Evaluando tendencias lineales

Si ahora graficamos la proporción `order_count/meals_count`, (columna construida en el inciso 2.3) veremos un patrón curioso.

- 2.1. Utilice `ggplot` para hacer una gráfica lineal. Tome de referencia la variable creada en el inciso 2.3 (eje y) versus la variable `month` (eje x).
- 3.2. Utilice la variable `month` para definir un color distinto en la estética de la gráfica.

Relación, selección y ordenamiento de datos

- 3.3. Tomando como referencia las variables en donde asigno los siguiente archivos:

- "cities.csv"
- "member.csv"
- "orders.csv"
- "restaurants.csv"

- 3.4. Utilice la función `inner-join` y construya las relaciones correspondientes (tome en cuenta cada uno de los identificadores de cada archivo), agregue sufijos a las columnas coincidentes por medio del parámetro `suffix`. Asigne la relación construida a una variable.

- 3.5. Use la función `select` en la variable creada anteriormente para presentar los siguientes campos:

- `first_name`
- `surname`
- `sex`
- `email`
- `city` (campo perteneciente a miembro)
- `date`
- `hour`
- `total_order`
- `name`
- `city` (campo perteneciente a restaurante)

- 3.6. Cree un ordenamiento ascendente para el campo `date_order`.

- 3.7. Por último imprima el resultado.

Obteniendo el conteo por tipo de comida

- 3.8. Tomando como referencia las variables en donde asigno los siguiente archivos:

- "meals.csv"
- "serve_types.csv"
- "meal_types.csv"
- "restaurants.csv"
- "cities.csv"
- "restaurant_types.csv"

- 3.9. Utilice la función `inner-join` y construya las relaciones correspondientes, agregue sufijos a las columnas coincidentes por medio del parámetro `suffix`. Asigne la relación construida a una variable.

- 3.10. Use la función `select` en la variable creada anteriormente para presentar los siguientes campos:
- `restaurant_name`
 - `city`
 - `restaurant_type`
 - `income_persentage`
 - `serve_type`
 - `meal_type`
 - `meal_name`
 - `price`
- 3.11. Use la función `count` en la variable creada anteriormente para obtener el conteo por la variable `meal_type`, luego asigne el resultado a una variable.
- 3.12. Imprima la variable del inciso anterior.

Comportamiento de ingresos por ciudad

- 3.13. Use la función `group_by` en la variable creada en el inciso 3.10 y agrupe por `city`, luego aplique la función `summarize` para obtener la sumatoria de `income_persentage`, asigne a la sumatoria la etiqueta `sumIncomePersentaje`, el resultado deberá almacenarlo en una variable.
- 3.14. Imprima la variable del inciso anterior.
- 3.15. Tome de referencia la variable anterior y aplique la función `ggplot` para construir un gráfico de barras, use `city` para el eje `x` y `sumIncomePersentaje` para el eje `y`.
- 3.16. Utilice la variable `city` para definir un color distinto en la estética de la gráfica.

Conociendo la tendencia de consumo en un restaurante en particular

- 3.17. Use la función `filter` en la variable creada en el inciso 3.10 para obtener los datos que cumplan con las siguientes condiciones:
- `meal_type = "Vegan", "Chicken"`
 - `restaurant_name = "Restaurant 2"`
 - `price > 30`
- 3.18. Guarde el resultado en una variable
- 3.19. Utilice la variable del inciso anterior para aplicar la `ggplot` y construir un gráfico de barras, use `serve_type` para el eje `x` y `price` para el eje `y`.
- 3.20. Investigue cómo utilizar la función `labs` para agregar un título, un subtítulo y etiquetas al eje `x` y `y`.

Tendencias de comida por restaurante y ubicación (ciudad)

- 3.21. Tomando como referencia las variables en donde asigno los siguiente archivos:
- `"cities.csv"`
 - `"restaurants.csv"`
 - `"restaurant_types.csv"`
 - `"orders.csv"`
 - `"meals.csv"`
 - `"serve_types.csv"`
 - `"meal_types.csv"`
 - `"order_details.csv"`

- 3.22. Utilice la función `inner-join` y construya las relaciones correspondientes, agregue sufijos a las columnas coincidentes por medio del parámetro `suffix`. Asigne la relación construida a una variable.
- 3.23. Use la función `select` en la variable creada anteriormente para presentar los siguientes campos:
- `restaurant_name`
 - `city`
 - `restaurant_type`
 - `income_percentage`
 - `serve_type`
 - `meal_type`
 - `meal_name`
 - `price`
 - `date`
 - `hour`
 - `total_order`
 - `id` (Identificador de `order_details`)
- 3.24. Tome la variable del inciso anterior y aplique la función `group_by` para agrupar por `city`, `restaurant_name`, `meal_name`.
- 3.25. Luego utilice la función `summarise` para obtener la sumatoria de `total_order` y asigne la etiqueta `totals`.
- 3.26. Por último guarde el resultado en una variable e imprima.
- 3.27. Toma la variable del inciso anterior y construye una gráfica de barras por medio de la función `ggplot`. Para ello use `restaurant_name` en el eje `x` y `totals` en el eje `y`.
- 3.28. Por último, toma de referencia el inciso anterior y aplica `faceting` tomando de referencia la variable `city`.

Análisis estadístico propio

- 3.29. Utilice la variable donde asigno el archivo "`monthly_member_totals.csv`" y haga un análisis estadístico propio considerando lo siguiente:
- Aplique las funciones `filter`, `group_by`, `summarise` y `count` para descubrir comportamientos particulares.
 - Use un tipo de gráfica para representar comportamientos o tendencias en los datos.
- 3.30. Para cada uno de los dos incisos anteriores concluya en función de los resultados obtenidos.

Condiciones

- El proyecto se debe trabajar en parejas.
- Todos los incisos deben ser desarrollados en un Notebook de R con el nombre **Proyecto.Rmd**.
- En el Notebook debe figurar el nombre de cada uno de los estudiantes.
- Para dudas concernientes al proyecto se utilizará el foro creado en el aula virtual de manera que todos los estudiantes puedan ver las preguntas y posteriores respuestas.
- Las copias totales o parciales obtendrán una nota de 0 puntos.
- No habrá prórroga.

Entrega

- La entrega será el martes **16 de abril a las 23:59**.
- La entrega se hará en el aula virtual y solo 1 de los integrantes debe hacer la entrega.
- Enviar únicamente el .Rmd.

Anexos

cities

- **id**: Identificador único para cada ciudad.
- **city**: Nombre de la ciudad.

restaurant_types

- **id**: Identificador único para el tipo de restaurante.
- **restaurant_type**: Nombre del tipo de restaurante.

serve_types

- **id**: Identificador único para el tipo de servicio.
- **city**: Nombre del tipo de servicio.

meal_types

- **id**: Identificador único para el tipo de comida.
- **city**: Nombre del tipo de comida.

restaurants

- **id**: Identificador único del restaurante.
- **restaurant_name**: Nombre del restaurante.
- **restaurant_type_id**: Identificador único para el tipo de restaurante.
- **income_persentage**: Porcentaje de ingresos del restaurante.
- **city_id**: Identificador único para cada ciudad.

orders

- **id**: Identificador único para cada pedido.
- **date**: Fecha del pedido.
- **hour**: Hora del pedido.
- **member_id**: Identificador único para cada miembro.
- **restaurant_id**: Identificador único para cada restaurante.
- **total_order**: Precio total del pedido.

members

- **id**: Identificador único para cada miembro.
- **first_name**: Nombre del miembro.
- **surname**: Apellido del miembro.
- **sex**: Sexo del miembro.
- **email**: Correo electrónico del miembro.
- **city_id**: Identificador único de la ciudad.
- **monthly_budget**: Presupuesto mensual del miembro.

meals

- **id**: Identificador único para cada comida.
- **restaurant_id**: Identificador único para cada restaurante.
- **serve_type_id**: Identificador único para cada tipo de servicio.
- **hot_cold**: Comida fría o caliente.
- **meal_name**: Nombre de la comida.
- **price**: Precio de la comida.

order_details

- **id:** Identificador único para cada detalle del pedido.
- **order_id:** Identificador único para cada pedido.
- **meal_id:** Identificador único para cada comida.

monthly_member_totals

- **member_id:** Identificador único para cada miembro.
- **first_name:** Nombre del miembro.
- **surname:** Apellido del miembro.
- **sex:** Sexo del miembro.
- **email:** Correo electrónico del miembro.
- **city:** Ciudad del miembro.
- **year:** Año relevante.
- **month:** Mes relevante.
- **order_count:** Conteo de pedidos en el mes correspondiente para cada miembro.
- **meals_count:** Conteo de comidas en el mes correspondiente para cada miembro.
- **monthly_budget:** Presupuesto mensual del miembro.
- **total_expense:** Gasto total mensual del miembro.
- **balance:** Saldo mensual del afiliado.
- **commission:** Comisión mensual del socio.