# Problema a resolver

Rest-on es una guía turística que está armando en su web una sección de restoranes, que sirva para asesorar a sus clientes en las características y calidad de los restoranes. A estos efectos, los contratan a ustedes para apoyarlos en los algoritmos de soporte de su web, con dos objetivos:

1. Poder recomendarles a sus clientes restoranes similares a los que han consultado
2. Rankear a los restoranes, en términos relativos a los restoranes comparables.

A estos efectos, cuenta con una base de datos de restoranes, a la cual se puede acceder en el siguiente link: <https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data>

# Entendimiento del caso de negocio y planificación del trabajo

## Discuta el desafío planteado en términos de la importancia que tiene para una empresa cualquiera y para Rest-on en particular.

Ofrecer recomendaciones que generan valor para los clientes es un objetivo para Rest-ON. No es solo recomendar una opción al cliente, sino que esta desarrolle una conexión con la marca y los clientes vean en la empresa un asesoramiento fundamentado. Por tanto, es objetivo realizar un este análisis permitiendo que los clientes no solo encuentren lo que están buscando, sino que se identifiquen con el servicio.

Este desafió está presente en todas aquellas empresas en su valor de negocio esté vinculado con la fidelidad de los clientes que trabajan con la marca.

## Identifique conceptual y técnicamente qué tipo de modelos serían necesarios para atender este problema

Para ofrecer una información que genere valor para el cliente se deben clasificar correctamente los restaurantes a partir de las características que estos poseen y los servicios que ofrece, también se puede tener en cuenta las características propias de los usuarios que visitan los restaurantes que permita realizar una segmentación en función de las características de los usuarios.

Todo dependerá del análisis del negocio detallando donde se identifique las variables que pueden formar parte de la segmentación de los restaurantes y poder ofrecer una calidad del servicio superior a los turistas que vienen a recibir la experiencia de Rest-ON

## Plantee hipótesis respecto a qué dimensiones de análisis y variables pueden ser relevantes para lograr el objetivo

Analizando el modelo de negocio de la empresa planteamos como hipótesis que:

Para ofrecer mejores recomendaciones es necesario conocer las características de los restaurantes, entender que tipos de cliente asisten y establecer las peculiaridades de la comida que ofrecen

**Dimensiones propuestas:**

* Características de los restaurantes
* Características de la comida
* Datos de los clientes que concurren

**Como variables dentro de cada dimensión identificamos las siguientes:**

* Características de la comida: Esta dimensión abarca las particularidades de la comida respecto al tipo de menú, la forma en que se sirve, restricciones alimenticias, entre otros.
* Características de los restaurantes – aquí se recoge la tipología del restaurante, en relación a aspectos operativos y físicos, tales como ubicación, disponibilidad de estacionamiento, medios de pagos aceptados, y otras cualidades que lo hagan atractivo.
* Datos del cliente: Deberíamos conocer el perfil y las costumbres de los clientes que concurren al restaurante, que nos permita clasificarlos teniendo en cuenta características de los clientes que concurren y en qué modo lo hacen.

*Ejemplo: Si es con familia, o amigos, si prefieren un ambiente relajado, etc*

## Identifique, en base a lo anterior, qué tipo de datos precisaría obtener para llevar a cabo este trabajo

|  |  |  |
| --- | --- | --- |
| Dimensión | Atributos | Observaciones |
| Características de la comida | Servicios ofrecidos. Almuerzo, Cena, Merienda, Menú Ejecutivo | Se consultará por separado por cada uno de los servicios que ofrece. |
| Consumo de bebida alcohólica (Ejemplo Vino, cerveza, tragos) | Cuál es la bebida alcohólica que más se consume. |
| Estilo de la comida: italiana, Mediterránea |  |
| Tipo de servicio: A la mesa, Bufet, fast-food |  |
| Restricciones alimenticias | Se consultará por separado disponibilidad de menú para celiacos, vegetarianos, veganos, diabéticos |
| Características de los restaurantes | Medios de pagos | Listado de opciones de medios de pago con los que cuenta el restaurante |
| Ubicación Geográfica | Coordenadas geográficas para medir cercanías entre restaurantes |
| Ambiente / Estilo | Se refiere al diseño o ambiente. Ej. Con espacios abiertos, vintage, naturalista, entre otros. |
| Música ambiente o No |  |
| Rango de precios | Se clasificarán los precios en categorías. |
| Área de juegos para niños |  |
| Área para fumadores |  |
| Calificación otorgada al servicio (rating 5 estrellas) |  |
| Disponibilidad de parking |  |
| Datos del cliente | Distribución de edades | Por rangos |
| Horarios de mayor concurrencia |  |
| Estado civil |  |
| Con quien concurren: Con familia, con hijos, con pareja, con amigos |  |
| Sexo de mayor concurrencia |  |
| Frecuencias de visitas (promedio de repetición mensual) |  |

## Arme un plan de trabajo, identificando las diferentes etapas y actividades para lograr el objetivo

Se propone para el desarrollo de esta investigación utilizar la metodología CRISP–DM que abarca desde el entendimiento del negocio hasta el despliegue de la solución sin dejar de lado las tareas de gestión de proyectos que permiten el monitoreo y control del tiempos, costos y riesgos asociados

## ¿Cómo será utilizado el resultado del trabajo por el cliente?

El resultado o solución va a ser utilizado por la agencia turística para incluirlo en su motor de búsqueda y ofrecer un servicio que beneficie a sus usuarios, generando más volumen de visitas a su sitio y fidelizándolos. También será usado por los asesores de la empresa utilizarán el sitio para ofrecer a los vacacionistas mejores recomendaciones según sus necesidades y las características de los restaurantes.

Los restaurantes en la medida que este motor de búsqueda y recomendaciones tenga mayor difusión se beneficiaran por tener mayor publicidad y el perfil de clientes que se sienta más cómodo respecto la propuesta que ofrece.

El usuario en la medida que el motor de búsqueda sea reconocido y fiable en cuanto a las recomendaciones podrá disfrutar de opciones más adecuadas a sus preferencias.

## En base a lo anterior, ¿Qué áreas se deberían involucrar en el proyecto y qué rol cumplirían?

Para la ejecución exitosa del proyecto tendría que participar todoas las áreas que estarán en el ciclo completo del proyecto, desde la definición hasta su puesta en producción y mantenimiento del proyecto. En tre las áreas identificadas están comercial, marketing y el área de soporte técnico que son quienes más conocen a los usuarios de la guía. Por otro lado el área de operaciones y IT al conocer como manejar la gestión de la empresa realizarán la captura de datos y recopilación de datos de los diferentes restaurantes. Finalmente, el área financiera deberá verificar los impactos económicos de las iniciativas y el equipo de Analistas de Datos para interactuar con los restantes interlocutores.

# Extracción, Transformación y Carga de datos

## Diseñe la estructura de tabla datos analítica: ¿qué va a ser cada fila? cuáles van a ser las columnas.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Atributos | Tipo de datos | Nombre Dataset | Transformación de datos | Fuentes de datos | Comentarios |
| Identificador del restaurante | Numeric | placeID | None |  |  |
| ¿Lunes a viernes? | Boleano | days |  | chefmozhours4.csv | chefmozhours4.csv |
| ¿Sabados? | Boleano | days |  | chefmozhours4.csv | chefmozhours4.csv |
| **¿Domingo?** | Boleano | days |  | chefmozhours4.csv | chefmozhours4.csv |
| **cash** | Boleano | Rpayment | cash == YES cuando Rpayment = cash | chefmozaccepts.csv | Vamos a seleccionar solo los siguientes medios de pagos que tienen la tabla de clientes **(userpayment.csv)** |
| **bank\_debit\_cards** | Boleano | Rpayment | bank\_debit\_cards == YES cuando Rpayment = bank\_debit\_cards | chefmozaccepts.csv | Vamos a seleccionar solo los siguientes medios de pagos que tienen la tabla de clientes **(userpayment.csv)** |
| **MasterCard-Eurocard** | Boleano | Rpayment | MasterCard-Eurocard == YES cuando Rpayment = MasterCard-Eurocard | chefmozaccepts.csv | Vamos a seleccionar solo los siguientes medios de pagos que tienen la tabla de clientes **(userpayment.csv)** |
| **VISA** | Boleano | Rpayment | VISA == YES **cuando** Rpayment = VISA | chefmozaccepts.csv | Vamos a seleccionar solo los siguientes medios de pagos que tienen la tabla de clientes **(userpayment.csv)** |
| **American\_Express** | Boleano | Rpayment | American\_Express == YES cuando Rpayment= American\_Express | chefmozaccepts.csv | Vamos a seleccionar solo los siguientes medios de pagos que tienen la tabla de clientes **(userpayment.csv)** |
| **Tipo de comida** | Boleano | Rcuisine |  |  | Cafeteria,California,Caribbean,Chinese,Contemporary,Continental-European,Deli-Sandwiches,Dessert-Ice\_Cream,Diner,Dutch-Belgian,Eastern\_European,Ethiopian,Family,Fast\_Food,Fine\_Dining,French,,Game,German,Greek,Hot\_Dogs, |
| **¿Tiene parking?** | Boleano | parking\_lot | parking\_lot = yes,valet\_parking, free, validated\_parking | chefmozparking.csv | **Se considera como parking los valores(yes,valet\_parking, free, validated\_parking)** |
| **Ciudad** | Nominal | city | Realizar transformaciones de datos | geoplaces2.csv | |
| **estado** | Nominal | state | Realizar transformaciones de datos | geoplaces2.csv | |
| **Codigo Postal** | Nominal | zip | Realizar transformaciones de datos | geoplaces2.csv | |
| **latitud** | Numeric | latitude |  | geoplaces2.csv | |
| **longitud** | Numeric | longitude |  | geoplaces2.csv | |
| **alcohol** | Nominal | alcohol |  | geoplaces2.csv | |
| **Fumadores?** | Nominal | smoking\_area | | geoplaces2.csv | Este valor es positivo cuando: **only\_at\_bar/permitted / section** |
| **vestimenta** | Nominal | dress\_code |  | geoplaces2.csv | |
| **Acceso?** | Nominal | accessibility |  | geoplaces2.csv | |
| **Precio** | Nominal | price |  | geoplaces2.csv | |
| **ambiente** | Nominal | Rambience |  | geoplaces2.csv | |
| **franquicia** | Nominal | franchise |  | geoplaces2.csv | |
| **area\_abierto\_cerrado** | Nominal | area |  | geoplaces2.csv | |
| **¿internet?** | Nominal | other\_services | | geoplaces2.csv | Descartamos este valor porque no es representativo para la muestra de datos |
| **food\_rating** | Numeric | food\_rating |  | rating\_final.csv | |
| **rating** | Numeric | rating |  | rating\_final.csv | |
| **service\_rating** | Numeric | service\_rating | | rating\_final.csv | |