**Análisis de Base de Datos**

**Objetivo:**  
Cada grupo deberá seleccionar una base de datos, realizar un análisis exploratorio básico, comprender su estructura y contenido, y presentar un informe con los hallazgos más relevantes.

Cada grupo debe realizar las siguientes tareas con la base de datos seleccionada:

1. **Selección de la Base de Datos**

* Nombre de la base de datos: Hotel-bookings
* Fuente (URL, repositorio, etc.): [Kaggle](https://www.kaggle.com/datasets/qucwang/hotel-bookings-analysis-dataset)
* Formato del archivo (.csv, .xlsx.): CSV (delimitado por comas)

1. **Estructura de la Base de Datos**

* Número total de filas (registros): 119391
* Número total de columnas (atributos o variables): 32

1. **Descripción de las Columnas**

Por cada columna se debe especificar:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |

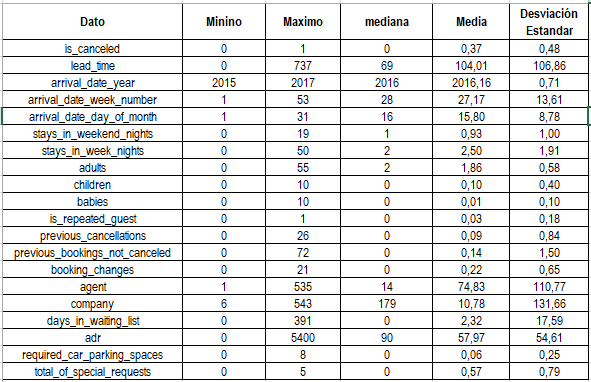
1. **Calidad de los Datos**

* ¿Hay valores faltantes en alguna columna? ¿Cuántos?: Country: 1, Agent: 16340, Company: 112593
* ¿Hay columnas con valores repetidos o inconsistentes? Si, en las 32 columnas hay datos repetidos
* ¿Existen datos fuera de rango o errores evidentes? No
* ¿Conoce la última fecha de actualización de los datos? Si, volumen 22, Febrero 2019

1. **Análisis Cuantitativo Básico**

Para columnas numéricas:

* Mínimo, máximo, media, mediana, desviación estándar.



Para columnas categóricas:

* Frecuencia de cada categoría. Hotel: City hotel: 0.66 Resort Hotel: 0.33
* Número de categorías únicas.

1. **Conclusiones Iniciales del Grupo**

* ¿Qué información relevante puede extraerse de esta base de datos? Podemos evidenciar que las personas prefieren reservar en hotel de ciudad embargo presenta una cancelación del 42%
* Patron de los países origen de los clientes
* Meses en los que tienen mayor número de huéspedes
* ¿Qué problemas de calidad o estructura podrían afectar un análisis más profundo? El hecho de no contar con criterios que me permitan identificar si hay información repetida (cedula, teléfono, email) afecta el análisis profundo de la información
* ¿Qué preguntas podría ayudar a responder esta base de datos?
* Qué tipo de hoteles prefieren los clientes de acuerdo a su lugar de procedencia
* Cuáles son los precios más buscados - a qué segmentos se debe dirigir la publicidad
* Que habitaciones son las preferidas para reservar según el tipo de hotel
* ¿Qué mejoras o limpiezas serían necesarias?
* Revisar los campos vacíos – agregar el país donde se encuentra ubicado el hotel

1. **Entrega**

Cada grupo debe entregar este documento diligenciado y, si es posible, el archivo de base de datos con sus observaciones o anotaciones.

Realizado por: Ana Maria Ramirez Acevedo, Santiago Quintero Ramirez, Johanna Gonzalez Franco