

# Universidad Externado

## Proyecto Aplicado

### Exploración de la Industria de Hostelería en Riohacha

#### Introducción Ciencia de Datos

Mayo 2024

Departamento de Matemáticas

# Exploración de la Industria de Hostelería en Riohacha

May 31, 2024

## Introducción

En este taller, se realizará un análisis estadístico de las empresas de hostelería en Riohacha utilizando la base de datos disponible en el siguiente enlace: [Base](#). El objetivo es proporcionar una base para la toma de decisiones en el ámbito de la hostelería en esta región. Los puntos que se abordan en este taller permiten entender y analizar los datos disponibles, ayudando a identificar patrones, relaciones y características clave de las empresas de hostelería.

## Descripción de la Base de Datos

La base de datos utilizada en este análisis contiene información detallada sobre las empresas de turismo en Riohacha. Las columnas incluidas en la base de datos son:

- **CODIGO\_RNT**: Código RNT de la empresa.
- **DEPARTAMENTO**: Departamento donde se encuentra la empresa.
- **MUNICIPIO**: Municipio donde se encuentra la empresa.
- **CATEGORIA**: Categoría de la empresa.
- **SUB\_CATEGORIA**: Subcategoría de la empresa.
- **HABITACIONES**: Número de habitaciones disponibles en la empresa.
- **CAMAS**: Número de camas disponibles en la empresa.
- **NUM\_EMP**: Número de empleados que trabajan en la empresa.
- **Numero\_Turistas**: Número de turistas atendidos por la empresa (generado aleatoriamente y proporcional al número de empleados).
- **Numero\_Cancelaciones**: Número de cancelaciones registradas por la empresa (generado aleatoriamente y proporcional al número de empleados).
- **Tipo\_Transporte**: Tipo de transporte más común utilizado por los turistas para llegar a la empresa (puede ser "avión", "auto propio" o "terrestre").

- **WEB\_VISIT**: Visitas a la web.

Esta información permite realizar un análisis del sector de hostelería en Riohacha, que sirve para identificar tendencias y áreas de oportunidad para mejorar la competitividad de las empresas locales.

## Historia del Caso de Estudio

Imagina que eres un analista de datos contratado por una prestigiosa empresa de consultoría, "Innovative Tourism Advisors" (ITA). Esta empresa ha sido contratada por el gobierno de Riohacha para realizar un análisis detallado del sector de hostelería local con el fin de revitalizar esta industria y atraer más visitantes nacionales e internacionales.

Riohacha, una ciudad costera con un gran potencial turístico gracias a su belleza natural y su rica cultura, ha experimentado una fluctuación en el número de visitantes en los últimos años. El gobierno local desea entender mejor el panorama actual del sector de hostelería para implementar estrategias efectivas que impulsen el crecimiento económico y mejoren la competitividad de las empresas locales.

ITA ha recopilado una base de datos que incluye información detallada sobre las empresas de turismo en Riohacha. Esta base de datos contiene columnas como el nombre de la empresa, el tipo de servicio que ofrecen, el número de empleados, el número de habitaciones y camas. Para los fines de este análisis, nos centraremos exclusivamente en las empresas de hostelería.

Como analista principal, tu misión es utilizar tus habilidades de ciencia de datos para comprender la composición de esta base de datos. Los resultados de tu análisis ayudarán al gobierno de Riohacha a tomar decisiones informadas y a diseñar políticas que fomenten el desarrollo sostenible de la hostelería en la región.

1. Filtrar los datos para trabajar únicamente con las empresas que pertenecen a la subcategoría de hostelería.
2. Listas y Estadísticas Descriptivas
  - (a) Crear listas por categoría a partir de las columnas del DataFrame:
    - Categoría
    - Número de empleados
    - Habitaciones
    - Camas
  - (b) Calcula las estadísticas descriptivas (media, desviación estándar, min, max) de cada una de las categorías.
  - (c) Crea tres gráficos de dispersión diferentes, donde la variable del eje vertical siempre será Número de Turistas y las variables del eje x son:
    - Número de Cancelaciones.
    - Número de Empleados.
    - Número de visitas en las web.

El color de los puntos debe ser igual a las categorías.

- (d) Filtra la categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO y calcula las estadísticas descriptivas:
- Media, mediana y desviación estándar del número de empleados
  - Media, mediana y desviación estándar de las habitaciones
  - Media, mediana y desviación estándar de las camas
  - Media, mediana y desviación estándar de Número de Turistas.
  - Media, mediana y desviación estándar de número de cancelaciones.
3. Crear diagramas de barras para visualizar:
- Número de Cancelaciones.
  - Número de Empleados.
  - Número de visitas en las web.
4. Utilizar diagramas de caja (boxplots) para detectar valores atípicos por categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO en:
- Número de Cancelaciones.
  - Número de Empleados.
  - Número de visitas en las web.
5. Crear histogramas por categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO para analizar la distribución de:
- Número de Cancelaciones.
  - Número de Empleados.
  - Número de visitas en las web.
6. Calcular y visualizar la matriz de correlación entre las variables numéricas del DataFrame.
7. Vamos a hacer un resumen por subcategoría
- Genera una base por cada subcategoría (Filtrar)
  - Obten la moda de las variables categóricas y las media de las numéricas.
  - Crea una nueva tabla con los resultados anteriores. La base debe tener por subcategoría la moda de las variables categóricas y la media de las numéricas.

Dados los análisis anteriores, sugiere cuál puede ser el problema que se debe analizar y cómo se puede abordar a partir de la información que se obtiene de la base de datos, bajo el esquema de alguna de las metodologías de la Ciencia de Datos.