

# 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 29, 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 sólida para la toma de decisiones en el ámbito de la hostelería en esta región. Los puntos que se abordan en este taller son permiten entender y analizar los datos disponibles, ayudando a identificar patrones, relaciones y características clave de las empresas de hostelería.

## 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 y los ingresos anuales. 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 realizar un análisis exhaustivo 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.

## Punto 1: Filtrar Empresas de Hostelería

Filtrar los datos para trabajar únicamente con las empresas que pertenecen a la subcategoría de hostelería.

## Punto 2: Listas y Estadísticas Descriptivas

1. Crear listas por categoría a partir de las columnas del DataFrame:
  - Categoría
  - Número de empleados
  - Habitaciones
  - Camas
2. 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 los Habitaciones
  - Media, mediana y desviación estándar de las camas.

## Punto 3: Atípicos

Utilizar diagramas de caja (boxplots) para detectar valores atípicos por categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO en:

- Número de Camas.
- Número de Habitaciones

## Punto 4: EDA

1. Crear histogramas por categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO para analizar la distribución de:
  - Número de Camas.
  - Número de empleados.
2. Crear gráficos de dispersión para analizar la relación de la categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO entre:
  - Número de Camas.
  - Número de empleados.
3. Calcular y visualizar la matriz de correlación entre las variables numéricas del DataFrame.

## Punto 5: Teorema del Límite Central

1. Generar una distribución no normal (por ejemplo, exponencial).
2. Tomar muestras de diferentes tamaños (por ejemplo, 10, 30, 100) y calcular sus medias para la categoría ESTABLECIMIENTOS DE ALOJAMIENTO TURISTICO.
3. Visualizar las distribuciones de las medias muestrales para demostrar el Teorema del Límite Central.

## Punto 6: Cálculo del Tamaño de Muestra para Estimación de Media y Proporción

1. Calcular el tamaño de muestra necesario para estimar la media con un nivel de confianza específico.
2. Calcular el tamaño de muestra necesario para estimar una proporción con un nivel de confianza específico.

## Punto 7: Intervalos de Confianza para Media y Proporción

1. Calcular intervalos de confianza para la media usando una muestra de datos.
2. Calcular intervalos de confianza para una proporción basada en una muestra.
3. Calcular intervalos de confianza para la varianza y el cociente de varianzas.
4. Calcular intervalos de confianza para la diferencia de medias con varianzas conocidas y no conocidas.

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.