

## PRÁCTICA 1

### Descripción

El objetivo de esta tarea es aplicar técnicas de manipulación de datos en R utilizando `dplyr` y generar visualizaciones informativas con `ggplot2`. Para ello, se utilizarán las funciones `filter()`, `arrange()`, `mutate()`, `select()`, `summarize()`, y `group_by()`, además de generar gráficos de dispersión, barras y facetado con `facet_wrap()`.

Se trabajará con el conjunto de datos **Airbnb NYC 2025**, que contiene información sobre alquileres de Airbnb en Nueva York hasta 02 de marzo de 2025, incluyendo precios, número de reviews, ubicación y tipo de alojamiento.

Descargue el archivo `listings.csv` que se compartió (utilice la función `read_csv()`). Las variables principales a utilizar son las siguientes:

- `host_location`. Ubicación del alojamiento.
- `room_type`. Tipo de alojamiento (entero, privado, compartido).
- `price`. Precio de la noche en dólares.
- `minimum_nights`. Número mínimo de noches requeridas.
- `number_of_reviews`. Cantidad de reseñas del alojamiento.
- `availability_365`. Disponibilidad en días del año.

### Parte I. Manipulación de datos con `dplyr`

- Muestra las primeras filas y la estructura del dataset (`head()`, `str()`).
- Realiza una transformación para la columna `price`, asegurándose de quitar el carácter dólar (\$) y convirtiendo toda la columna a valor numérico.
- Filtra los registros para mantener solo alojamientos con más de 10 reseñas y un precio menor a \$500 por noche. Además, excluye valores faltantes para la columna `host_location`.
- Ordena los datos de mayor a menor disponibilidad (`availability_365`).
- Crea una nueva columna llamada `total_revenue`, que represente el ingreso potencial en un año (`price * availability_365`).
- Selecciona las columnas `host_location`, `room_type`, `price`, `availability_365`, `total_revenue` y `number_of_reviews`.

## Parte II. Resúmenes y agrupaciones

- Agrupa los datos por `host_location` y calcula:
  - Precio promedio por noche.
  - Número total de alojamientos.
  - Ingreso potencial promedio (`mean(total_revenue)`).
- Agrupa los datos por `room_type` dentro de cada `host_location` y calcula la disponibilidad promedio.

## Parte III. Visualización de datos con ggplot2

### Gráfico 1. Dispersión (`scatter plot`)

- Crea un gráfico de dispersión donde el eje x sea el precio y el eje y el número de reseñas (`number_of_reviews`).
- Diferencia los puntos por `room_type`, utiliza el parámetro `color` en la función `aes`.

### Gráfico 2. Barras (`bar plot`)

- Crea un gráfico con la evolución del precio promedio de los alojamientos en función del `room_type`.
- Usa `geom_col()`, asegurándose de que se diferencie el tipo de alojamiento con el parámetro `fill` en la función `aes`.

### Gráfico 3. Facetado (`facet_wrap`)

- Crea un gráfico de dispersión donde se muestre la relación entre disponibilidad (`availability_365`) y precio (`price`).
- Divide el gráfico en paneles por `host_location` usando `facet_wrap(~host_location)`.

## Evaluación

- Uso correcto de `filter()`, `arrange()`, `mutate()`, `select()`, `summarize()`, y `group_by()`.

### Análisis exploratorio.

- Identificación de patrones en precios y disponibilidad.
- Análisis de distribución por ubicación y tipo de alojamiento.

### Visualización de datos

- Gráfico de dispersión correctamente etiquetado.
- Gráfico de barras mostrando tendencias de precios.
- Gráfico facetado comparando disponibilidad vs precio.
- Explicaciones claras de los hallazgos en cada gráfico.

### Consideraciones:

- Construya un R Notebook y desarrolle cada uno de los puntos expuestos anteriormente.
- Cargue el cuaderno a UEDI con el nombre **“r-practica-1.Rmd”**.
- Las copias totales o parciales obtienen nota cero.
- Se envía a más tardar el **lunes 01-septiembre hasta las 23:59**.