

## Módulo 3 Data Analysis

### Exploratory Data Analysis

- Resumir características de los datos
- Comprender mejor el data set
- Encontrar insights
- Extraer variables importantes

### Estadística descriptiva

- Descripción de la data
- Resumen de la muestra y medidas

Python  $\Rightarrow$  `describe()`

$\hookrightarrow$  para variables numéricas / distribución

`value_counts()`

$\hookrightarrow$  para variables categóricas

$\rightarrow$  Box-plots: ideal para conocer la distribución outliers

$\rightarrow$  Scatter plot: relación entre dos variables

X  $\rightarrow$  variable independiente

Y  $\rightarrow$  variable dependiente / target

### Group By (Python)

- En variables categóricas para conocer comportamiento de nuestra variable dependiente
- Gráficamente se puede utilizar heatmap

### ANOVA: Análisis de varianza

- Comparación estadística de grupos
- Encontrar correlación entre distintos grupos de variables categóricas
- Ftest score:  
valor p: grado de confianza

### Correlación

- Medida de interdependencia entre dos variables

"Correlación no implica causalidad"

↳ Lluvia causa mayor presencia de paraguas, pero mayor paraguas no causan lluvia.

## Estadísticas

→ Correlación Pearson: entre 0 y 1 y con valor  $p$  para determinar efectividad del coeficiente de correlación

→ Usualmente tablas se grafican con heatmap.

## Review Questions (evidencia)

