

Trabajo Práctico 2

Juan Manuel Basso
Emiliana Verdun
Mariana Zunino

Tecnología Digital VI
Universidad Torcuato Di Tella

7 de octubre de 2024

1 Análisis exploratorio de los datos

1.1 Descripción de las variables principales

El conjunto de datos proporcionado incluye características referidas al usuario como así de la subasta. Las principales variables son:

- **Label:** Es la variable objetivo y representa si un usuario hizo clic (1) o no (0) en una publicidad.
- **auction_bidfloor:** Es una variable numérica y puede influir en la propensión de los usuarios a interactuar con anuncios de alto o bajo costo.
- **auction_time:** Representa la hora en que se realizó la subasta. Esta variable es importante ya que el comportamiento del usuario puede variar a lo largo del día y de la semana.
- Categorías de negocio (**action_categorical_0 a action_categorical_7**): Estas variables categóricas identifican diferentes aspectos del negocio y pueden ser claves para identificar patrones de comportamiento en ciertos segmentos de usuarios o tipos de anuncios.
- **auction_age:** Representa la edad del usuario al momento de la subasta, lo que podría estar relacionado con las tendencias de interacción con las publicidades.

1.2 Análisis y Visualización de Patrones

1.2.1 Relación entre *auction_time* y *Label*

La hora del día en que ocurre la subasta puede influir en el comportamiento del usuario. A continuación, se presenta la distribución de impresiones por hora del día, seguido por la distribución de clicks (*Label = 1*) por hora.

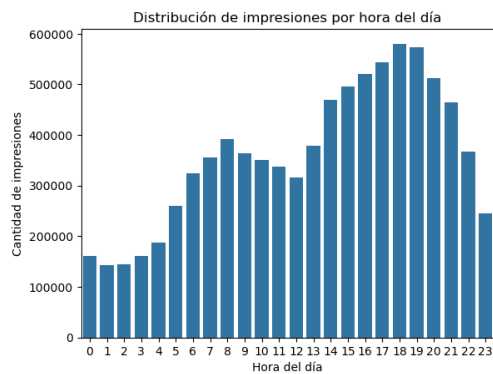


Figure 1: Cantidad de impresiones por hora del día.

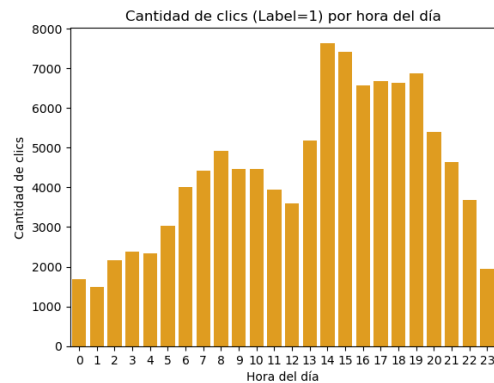


Figure 2: Cantidad de clics ($Label = 1$) por hora del día.

1.2.2 Correlación de Pearson entre las Variables y *Label*

Para identificar las variables más relevantes para predecir los clics, se calculó el coeficiente de correlación de Pearson entre las principales variables numéricas y la variable *Label*. En el siguiente gráfico se muestran las 8 variables con mayor correlación.

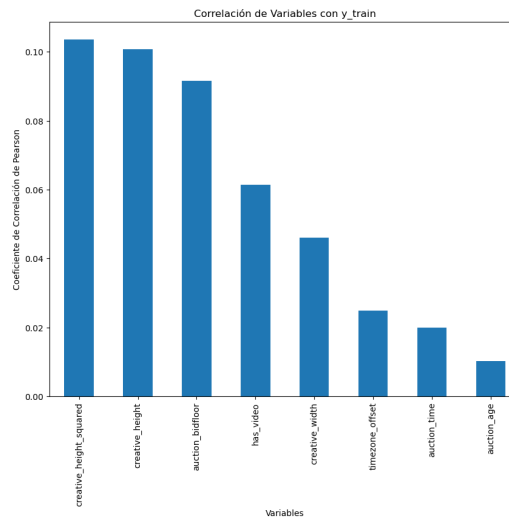


Figure 3: Correlación de Pearson de las variables con *Label*.