

# Exploración y Visualización de datos - Examen

28 de mayo de 2025

## 1. Introducción

El contenido de la base de datos **googleplaystore** proporciona información detallada sobre las descargas de aplicaciones móviles realizadas a través de la plataforma Google Play Store.

En el cuadro 1 se describen las variables correspondientes.

Variable	Descripción
App	Nombre de la aplicación
Category	Categoría a la que pertenece la aplicación
Rating	Calificación general de los usuarios de la aplicación (en el momento de la extracción)
Reviews	Número de reseñas de los usuarios de la aplicación (en el momento de la extracción)
Size	Tamaño de la aplicación (en el momento de la extracción)
Installs	Número de descargas/instalaciones de la aplicación (en el momento de la extracción)
Type	Pago o gratuita
Price	Precio de la aplicación (en el momento de la extracción)
Content Rating	Grupo de edad al que se dirige la aplicación: Niños / Adultos mayores de 21 años / Adultos
Genres	Una aplicación puede pertenecer a varios géneros (además de su categoría principal)
Last Updated	Última actualización
Current Ver	Versión actual
Android Ver	Versión de Android

Cuadro 1: significado de las variables

Realizar un informe, en formato **de RMarkdown HTML**, en donde se analice la performance de estas aplicaciones. Colocar los códigos utilizados.

Fecha máxima de entrega: **miércoles 4 de junio a las 23:59 hs.**

## 2. Resolver las siguientes consignas:

### 1. Acondicionamiento de la base de datos

- a) Verificar si hay filas duplicadas, y si es así eliminarlas.
- b) Detectar datos atípicos, analizarlos, y si es pertinente eliminarlos.
- c) Verificar los tipos de variables, y si es necesario modificarlos.
- d) Agrupar las categorías que consideres necesarias.

### 2. Análisis descriptivo

- a) Calcular el promedio, desvío, mediana, y cuartiles para las variables numéricas (cuantitativas) por categoría. Nota: si la variable tiene un rango muy amplio separar por grupo y realizar estos cálculos dentro de cada grupo.
- b) Armar una función, que como argumento de entrada reciba la base datos, y como salida entregue una lista, solo para el grupo de jóvenes (Teen), en donde cada nivel sea para cada categoría. La función debe entregar lo siguiente:
  - nombre de la categoría
  - un dataframe con las columnas App y Reviews
  - mínimo, máximo y promedio de Reviews

### 3. Análisis exploratorio y visualización

- a) Comparar por año el total de actualizaciones que tuvo cada categoría.
- b) ¿Cuál fue la aplicación más cara de cada categoría?
- c) Analizar y comparar la distribución del rating entre las aplicaciones pagas y gratuitas.
- d) Realizar una visualización que muestre las diferencias en las categorías referidas salud y las referidas al deporte, con respecto al tamaño de la aplicación, diferenciando entre aplicaciones gratuitas y pagas.
- e) Realizar un gráfico que muestre la evolución de la cantidad de actualizaciones por mes para el año 2018, comparando las categorías sobre salud, educación y entretenimiento.
- f) Mediante diagramas de boxplots comparar los precios de las aplicaciones referidas a la familia, al juego, y a la salud.
- g) Dentro del grupo adolescente (Teen) realizar una comparación entre las categorías referidas al entretenimiento y el juego con respecto a las que no lo son.
- h) ¿Hay relación entre el tamaño y el precio? Realizar el análisis de manera global y por categoría (elegir 4)

Se puede agregar cualquier otro análisis que consideres necesario, o incluso realizar una segmentación mayor en los ítems pedidos.