

# Explorando Alternativas de Conjuntos de Datos para el Proyecto BIT

Paula Sofía Torres

21 de Marzo de 2024

## 1 Introducción

En este documento, se detallan los conjuntos de datos seleccionados para el análisis en el marco del Proyecto BIT. Se proporciona una justificación clara de la elección de estos conjuntos, así como una descripción de las variables de interés que serán consideradas en el estudio

## 2 Datasets

### 2.1 Tenis de Falabella

URL: [Falabella](#)

**Justificación de selección:** Este conjunto de datos es esencial para entender las variables que impactan en el precio de los tenis distribuidos por la empresa Falabella. Además, se busca identificar patrones subyacentes en cuanto a la proporción de descuento aplicada a estos productos. ¿Existen marcas o tipos de tenis que tienden a tener descuentos más frecuentes o significativos?

**Variables de interés:**

- Tipo
- Marca
- Precio
- Precio con descuento
- Género
- Modelo
- Material
- Material forro
- Material suela

- Disciplina
- Material plantilla
- Hecho en

## 2.2 Mercado Inmobiliario

URL: [Metrocuadrado](#)

URL: [Percepción de seguridad en barrios de Bogotá](#)

**Razón de elección:** Este dataset es relevante para analizar el impacto de diversas características en el valor de venta de un inmueble. Se pretende cruzarlo con información de la percepción de la seguridad en barrios - localidades en Bogotá de la Camara de Comercio.

**Variables de interés (Metro Cuadrado):**

- Precio
- Área construida
- Área privada
- Habitaciones
- Estrato
- Baños
- Barrio
- Estado
- Exteriores
- Zonas comunes
- Características del sector

**Variables de interés (Percepción de seguridad):**

- Localidad
- Barrio
- P102. El barrio en el que usted habita es.... (segura - insegura)
- P204 De la siguiente lista indíqueme de qué delito o delitos fue víctima

## 2.3 Evaluación de mercado laboral

**URL:** [Data that captures industry and country-specific employment growth based on updates to LinkedIn member profiles](#)

**Razón de elección:** Este conjunto de datos es parte de la asociación LinkedIn - Banco Mundial, que ayuda a gobiernos e investigadores a comprender los mercados laborales en rápida evolución con datos detallados y dinámicos. Permite a los líderes comparar y comparar los mercados laborales de todo el mundo; analizar habilidades, ocupaciones, migración e industrias; y apalancamiento...

Sería interesante evaluar qué habilidades son más necesarias en determinados países según su nivel de ingresos, locación e industria. Asimismo, poder identificar tendencias o cambios radicales en el tiempo. (Se están buscando otros datasets que puedan agregar valor al análisis, aunque este tiene 15 columnas y 7335 filas)

**Variables de interés:**

- country name
- wb income
- industry name
- growth rate 2015 - 2019