**Universidad San Francisco de Quito**

**Maestría en Ciencia de Datos**

**Fundamentos de Ciencia de Datos**

# Informe Proyecto 1

# Objetivo General

**Tema del proyecto:** “Factores determinantes de la inclusión financiera de mujeres en América Latina: Un análisis basado en la base WWBI del Banco Mundial Introducing the Worldwide Bureaucracy Indicators”.

* 1. **Relevancia del proyecto:**

# Pregunta de investigación: ¿Cómo se relacionan las características demográficas y socioeconómicas de mujeres emprendedoras con el acceso a servicios financieros y su bienestar económico, basándonos en la base de datos WWBI del Banco Mundial?

# Hipótesis inicial: Factores como el nivel educativo, el tipo de empleo, el ingreso mensual y la ubicación geográfica influyen significativamente en el acceso a productos financieros y, en consecuencia, en el bienestar económico de las mujeres.

# Relevancia: Este análisis proporciona una base empírica para diseñar políticas públicas y programas de inclusión financiera enfocados en mujeres, especialmente en contextos de desigualdad económica o exclusión financiera. Puede ser especialmente útil para ONGs, instituciones financieras inclusivas y organismos multilaterales como el Banco Mundial.

# Organización del proyecto basada en la metodología CRISP-DM

# Fase 1: Business Understanding

# Problema: Baja inclusión financiera entre mujeres, lo que limita su desarrollo económico y social.

# Valor del proyecto: El MVP será un dashboard interactivo con insights clave sobre factores que impulsan o limitan el acceso a servicios financieros en mujeres. Esto permitirá a decisores de política pública actuar de manera informada.

# Fase 2: Data Understanding

# Fuente de datos: World Bank’s World-Wide Bureau of Indicators (WWBI) – [ver enlace](https://documentos.bancomundial.org/es/publication/documents-reports/documentdetail/099062623201028153/p1687031c943fa1313f6814bcf1aeb51371a5665a9ac).

# Frecuencia de actualización: Los datos corresponden a encuestas transversales (última edición: 2023) y no se actualizan anualmente.

# Variables clave:

# country, region, gender, age, education\_level, employment\_status, monthly\_income, has\_bank\_account, uses\_mobile\_money, borrowed\_last\_year, savings\_account.

# Desafíos esperados:

# Valores nulos las columnas años.

# Inconsistencias categóricas por país o idioma.

# Fase 3: Data Preparation

# Tareas previstas:

# Imputación de valores faltantes (valores Nan en las columnas).

# Normalización de variables categóricas (employment\_status, education\_level).

# Detección y manejo de outliers (por ejemplo, ingresos extremadamente altos o extremedamente bajos).

# Generación de variables derivadas en relación al filtrado de los datos como: “Región”, “Tipo de indicador”).

# 1.3. Análisis Exploratorio de los Datos (EDA)

# Descripción de las variables:

# Filas (observaciones): 61,004

# Columnas (variables): 27

# Tipo de datos:

# 4 columnas categóricas:

# Country Name, Country Code, Indicator Name, Indicator Code

# 23 columnas numéricas (años del 2000 al 2022)

# A partir del año 2000:

# Más del 90% de los datos están ausentes para los primeros años (2000–2005).

# Por ejemplo:

# 2000: 57,377 valores faltantes (~94%)

# 2022: 60,573 valores faltantes (~99%)

# Esto indica que la cobertura temporal es incompleta y variable según el indicador y el país.

# Visualizaciones iniciales:

# 1.3.2.1 Evolución del indicador: Females, as a share of public paid employess by occupational group: Mangers en Latinoamérica

# 

# 

# Descripción:

# A partir de la imagen presentada, que muestra la evolución del indicador "Females, as a share of public paid employees by occupational group: Managers" en países de Latinoamérica entre 2010 y 2021, se pueden destacar las siguientes conclusiones:

# Variabilidad significativa entre países:

# El porcentaje de mujeres en cargos de gerencia del sector público varía notablemente entre países latinoamericanos.

# Algunos países muestran trayectorias ascendentes (como Uruguay, Costa Rica y Ecuador), mientras que otros tienen tendencias más irregulares o estancadas (como Perú, Nicaragua o Honduras).

# Presencia femenina superior al 50% en varios países:

# Países como Uruguay, Colombia, Ecuador y Argentina superan consistentemente el 50% en ciertos años, lo cual sugiere avances importantes en paridad de género en puestos directivos públicos.

# Estabilidad o estancamiento en otros casos:

# Algunos países como Perú, Nicaragua y Paraguay muestran un porcentaje femenino que oscila alrededor del 30%–40%, sin una tendencia clara de mejora.

# Tendencia general al alza, con fluctuaciones:

# Aunque hay fluctuaciones interanuales, la mayoría de los países muestran una tendencia positiva o estable en el tiempo, lo que podría reflejar políticas públicas de inclusión laboral o mejoras institucionales.

# 1.3.2.2 Indicador "Females, as a share of public paid employees by occupational group: Managers" en Ecuador en comparación con la media regional de Latinoamérica entre 2010 y 2021

# 

# Descripción:

# Variabilidad significativa en Ecuador:

# Ecuador muestra fluctuaciones marcadas a lo largo del período.

# En 2014, se observa una caída abrupta en la participación femenina en cargos gerenciales del sector público, seguida de una recuperación entre 2015 y 2016.

# En algunos años (2012 y 2016), Ecuador supera levemente la media regional, pero en la mayoría del período se encuentra por debajo del promedio latinoamericano.

# Especialmente en 2014 y 2017, la brecha con la media regional es notable.

# La media regional es más estable que la de Ecuador y presenta una tendencia leve al alza hasta 2019, con una caída en 2020 que probablemente se relacione con efectos de la pandemia.

# Nota: las gráficas restantes se pueden observar directamente en el notebook

# Outliers:

# Female to male wage ratio in the public sector (using mean)" en el año 2015 para países de Latinoamérica.

# A graph with a blue rectangle AI-generated content may be incorrect.

# Descripción:

# Relación salarial femenina promedio por debajo de la paridad:

# La mediana de la razón salario mujer/hombre en el sector público está por debajo de 1, lo que indica que, en promedio, las mujeres ganan menos que los hombres en la mayoría de los países.

# Desigualdad salarial entre países:

# Países como Brasil (BRA) y Chile (CHL) tienen ratios cercanos a 0.75, lo cual evidencia una brecha salarial significativa.

# En contraste, la República Dominicana (DOM) aparece como un outlier positivo, con una razón superior a 1.2, lo que indica que las mujeres ganaron más que los hombres en promedio ese año.

# Distribución relativamente concentrada:

# La mayoría de los países se agrupan entre valores de 0.85 y 1.0, indicando cierto grado de consistencia en la desigualdad salarial dentro de la región, aunque sin alcanzar paridad.

# Países por debajo de 0.9:

# Bolivia (BOL), Perú (PER), Paraguay (PRY), entre otros, están por debajo de 0.9, lo que refleja una brecha persistente y preocupante.

# 1.4. Plan de limpieza y manipulación de datos (Data Wrangling)

Notebook para limpieza y tratamiento de datos del dataset de indicadores de género en LATAM.Cada fila del dataset representa una serie temporal asociada a un indicador particular para un país, con valores anuales desde el año 2000 hasta el 2017. El objetivo principal es limpiar el dataset eliminando columnas y filas no informativas, y luego aplicar imputación por regresión lineal sobre las series temporales incompletas.

Esta notebook asume que el archivo df\_latam\_genero.csv está en la carpeta .[/data/clean/parciales](https://colab.research.google.com/drive/1fiHNwApNzPoCSexWd4VtbuAoRjpkDdpm?usp=sharing)/

# Nota: el detalle se encuentra en la notebook: 02\_DataWrangling.ipynb

**1.5 Estructuctura del repositorio de GitHub**

WWBI\_InclusionFinanciera/

│

├── data/

│ ├── raw/ <- Datos originales de WWBI

│ ├── clean/ <- Datos con imputación inicial

│ └── curated/ <- Datos listos para análisis y modelado

│

├── notebooks/

│ ├── 01\_EDA.ipynb

│ └── 02\_DataWrangling.ipynb

│

├── src/

│ ├── data\_prep.py

│ └── utils.py

│

├── reports/

│ ├── figures/

│ └── reporte.pdf

│

├── README.md

├── .gitignore

└── requirements.txt