



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM



FACULTAD DE CIENCIAS FÍSICO-MATEMÁTICAS

# Universidad Autónoma de Nuevo León

## Facultad de Ciencias Físico-Matemáticas

### Avance I Proyecto Integrador

Equipo: 4

Grupo: 02

Integrantes:

- 1804029 Castillo Trujillo Sabina Alejandra
- 1821920 Alanís Mares Víctor Hugo
- 1887860 Castro Hernández Jesús Ramón

Mayra Cristina Berrones Reyes

Minería de Datos

### 1) Título de la base de datos:

- a) World Development Indicators
- b) URL: <https://www.kaggle.com/worldbank/world-development-indicators?select=Indicators.csv>

### 2) Descripción de los datos:

- a) Los datos contenidos en esta tabla son indicadores que se obtuvieron de la página del Banco Mundial (una organización multinacional especializada en finanzas y asistencia), la cual a su vez compiló esta información de fuentes internacionales oficialmente reconocidas
- b) La base de datos consta únicamente de 6 columnas:
  - **CountryName**: contiene los nombres de los países que se van a estudiar, los cuales son alrededor de 247 países (Texto)
  - **CountryCode**: a cada país le corresponde un código o una abreviatura para poder diferenciarlo de los demás países e identificarlo de manera rápida (Texto)
  - **IndicatorName**: esta columna guarda los nombres de todos los indicadores que el Banco Nacional ha registrado de los diferentes países (los cuales son alrededor de 1344). Estos indicadores contienen información sobre agricultura, clima, crecimiento económico, salud, población, pobreza, etc. (Texto)
  - **IndicatorCode**: a cada indicador le corresponde una abreviatura única que lo diferencia de los demás indicadores (Texto)
  - **Year**: año al que corresponde el resultado del indicador de un determinado país (Número)
  - **Value**: resultado que el indicador arroja dependiendo de sus unidades (ya sea en porcentajes, en miles, millones, etc.) (Número)

### 3) Justificación de los datos:

- a) Una de las características que nos llamó la atención y que nos hizo querer trabajar con estos datos es la gran cantidad de información que esta base tiene, en especial el gran número de indicadores con los que cuenta y la variedad de rubros a los que estos indicadores de desarrollo están enfocados. A simple vista nos pudimos dar cuenta que se podían sacar muchas conclusiones con base en esta información, ya que contamos con factores muy

valiosos que afectan el desarrollo de una nación y, el hecho de poder cuantificar estos factores y el desarrollo como tal, nos causó aún más emoción

- b) Existen muchos beneficios al trabajar con esta base de datos. Primero que nada, tenemos información suficiente en cuanto a cantidad de datos, por lo que si queremos sacar conclusiones o hacer predicciones éstas van a estar bien fundamentadas.

Segundo, debido a la enorme cantidad de indicadores de desarrollo con los que contamos, nos podemos dar la libertad de desechar algunos que no sean tan relevantes en nuestro estudio, pero sin caer en la falta de información.

Por último, pero no menos importante, sentimos que las conclusiones a las que podemos llegar y la solución al planteamiento de nuestro problema pueden ser de interés tanto nacional como internacional, ya que estaríamos viendo el desarrollo de los países de una manera objetiva y cuantificable.

#### **4) Planteamiento del Problema**

¿Cómo ser un país de primer mundo?

Imagínense las posibilidades para un país en crecimiento de conocer los indicadores principales que comparten los países desarrollados, esto impactaría directamente en sus estrategias de gobierno, repartición de recursos, visiones futuras, y en general el desarrollo humano.

Como problemática principal, nos planteamos como una organización internacional (sea el BIS, Banco Mundial, ONU, etc.) solicita un análisis de los indicadores que se tienen en la base de datos para ver los que impactan en mayor escala y los que se consideran un factor común entre las tres clasificaciones (Desarrolladas, Vías de desarrollo, Tercer Mundo), para así, dar recomendaciones o poner condiciones sobre los países deficientes y lograr un cambio.

Como expectativa del estudio es mejorar el entorno mundial, así como analizar los países que, a pesar de no cumplir con lo analizado, forman parte de cierto grupo, y viceversa.

#### **5) Objetivo final**

El objetivo final de este programa (Y de este estudio) es determinar las variables a las que un país le debe poner atención para así aumentar sus posibilidades de ser un país de primer mundo (o en su defecto simplemente aumentar la calidad de vida de sus habitantes) hay variables que parecen obvias por ejemplo es fácil ver que si el producto interno bruto per cápita aumenta es probable que la calidad de vida general

aumente, pero ¿Que tanto afecta? ¿Es la variable más importante? ¿Hay maneras alternas de mejorar la calidad de un país?

Todas estas preguntas son las que buscamos responder, y el objetivo final es más que nada representar dos tipos de variables:

- 1- Las que tienen en común países de primer mundo
- 2- Las que tienen en común países de tercer mundo

Para así poder ver la comparativa general, también se buscará buscar si existen datos atípicos respecto a si existe algún país que pertenezca al primer mundo, pero tenga algunas cualidades que se aprecian normalmente en el tercer mundo y viceversa.

## **6) Planeación de la herramienta a utilizar**

El proceso general de trabajo se tiene pensado como una línea de tiempo:

1. Ordenamiento de los datos: Actualmente los datos están dispersos en 4 variables donde cada fila determina un año y un país, se planea hacer una transposición de los datos para que cada variable sea una columna y posiblemente partir los datos por país.
2. Creación de variable indicadora: Esto se debe hacer a mano y representa la creación de una variable que nos indique la calidad de vida general del país evaluado (1 = primer mundo, 3=tercer mundo, 2= calidad promedio, 0= No se sabe/no aplica) esto se hace con el fin de ser esta la variable que debemos “predecir”
3. Análisis de componentes principales: Debido a la cantidad masiva de variables que tenemos debemos primero hacer una depuración para solo quedarnos con las variables que sí importan.
4. Bosque aleatorio para clasificación: De nuevo, esto es debido a la cantidad de variables que consideramos se quedaran incluso después del PCA, aquí buscamos que se nos arroje un modelo para así poder identificar las variables relevantes (incluso después del PCA).
5. Análisis de clústeres: Aquí identificamos los países parecidos en general y compararemos sus variables indicadoras para determinar la coherencia interna de los datos.
6. Conclusiones: Una vez tengamos las respectivas visualizaciones de los análisis hechos se procederá a simplemente ponerlo todo en un “resumen” del análisis total en formato póster.