



UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE CIENCIAS FÍSICO
MATEMÁTICAS



MINERÍA DE DATOS

Avance I Proyecto Integrador Equipo 3

Maestro: Mayra Cristina Berrones Reyes

Grupo: 02 / miércoles

Matricula:

1810461

1805800

1798528

Nombres:

Rossel Castillo Tania Sarahi

Martínez Jiménez Gloria Nohemí

Noriega Ortega Miguel Alejandro

Página:

World Happiness Report <https://www.kaggle.com/unsdsn/world-happiness>

Los datos encontrados en la base de datos son múltiples tablas las cuales consisten en el puntaje o nivel de felicidad por país según una encuesta realizada por el “Gallup World Poll”, el puntaje de felicidad es la suma de ciertos factores los cuales están detallados en la tabla individualmente

La base de datos a utilizar posee 10 columnas las cuales son:

- **PIB per cápita:** Cuanto influye el ingreso per cápita en la felicidad. (Número)
- **Familia:** Que tanto influye la familia en la felicidad. (Número)
- **Esperanza de vida:** Cuanto influye la esperanza de vida en la felicidad. (Número)
- **Libertad:** Que tanto influye la libertad en la felicidad. (Número)
- **Generosidad:** Que tanto influye la generosidad en la felicidad. (Número)
- **Confianza en el Gobierno/Corrupción:** Que tanto influye la confianza en el gobierno en la felicidad. (Número)
- **Puntaje de felicidad:** Suma de las columnas ya mencionadas. (Número)
- **País.** (Texto)
- **Regiones.** (Texto)
- **Ranking General:** Orden del país con mayor puntaje de felicidad al país con menor puntaje. (Número)

La base de datos nos interesó mucho ya que estudia la felicidad a través del mundo lo cual es un tema bastante interesante y que puede dar pie a varios cambios, una vez que se analiza la base de datos se puede identificar qué factores son los que más influyen en la felicidad de los habitantes de un país específico.

El beneficio que encontramos al trabajar con esta información es que al analizar esta base de datos se puede trabajar directamente en un área o factor para que de esta manera pueda aumentar más felicidad de los ciudadanos, viéndolo de otra forma también podemos analizar qué impacto puede tener en una persona el que su área o factor que le da más felicidad este siendo descuidada, si esto solo causa una menor felicidad o si puede llegar a causar problemas mentales o ser un punto de alto riesgo para suicidio.

Esta segunda idea fue la que más nos interesó y es en la cual nos concentraremos durante el proyecto.

La problemática que encontramos es que en la actualidad la tasa de suicidios va en aumento alrededor del mundo y no se le está tomando la debida importancia para reducirla y/o eliminarla a tal grado que la tasa de suicidio es mayor que los muertos por el virus de COVID-19, por lo que es necesario darle su atención requerida.

El objetivo Principal que tenemos es el de encontrar cuales factores son los que influyen más en el nivel de felicidad de la población para de esa manera aislar ese factor como el más importante ya que si en dado caso llegara a faltar o reducirse podría afectar la salud mental de las personas siendo un alto potencial de suicidio.

Un objetivo secundario podría ser el de encontrar que áreas y grupos son los más vulnerables a sufrir de depresión y problemas mentales, este objetivo se puede analizar más a fondo utilizando la base de datos secundaria “Mental Health and Suicide Rates” (<https://www.kaggle.com/twinkle0705/mental-health-and-suicide-rates?select=Crude+suicide+rates.csv>), la cual posee datos sobre los hospitales y personal preparado para ayudar en esos casos, así como estadísticas sobre los suicidios divididos por edades y países para una identificación más fácil de grupos vulnerables.

Las técnicas que utilizaremos serán las de Clustering y Visualización.

- Clustering será nuestra principal herramienta ya que nos será útil para identificar qué factores son los que influyen más en el nivel de felicidad de la población para de esa manera aislar ese factor como el más importante ya que si en dado caso llegara a faltar o reducirse podría llegar a afectar la salud mental de las personas ocasionando el suicidio.
Los factores se tomarán como proporciones de acuerdo con su puntaje de felicidad para que de esta manera todos sean homogéneos y se analicen de forma adecuada.
- Visualización será utilizada para presentar la información y hallazgos de interés de manera más accesible y entendible.

Nota: Los algoritmos no fueron incluidos debido a que aún no tenemos el conocimiento suficiente sobre estas técnicas y queremos esperar a conocerlas para decidir qué técnicas nos serían más útiles.