



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



AVANCE 1 PROYECTO INTEGRADOR

MINERÍA DE DATOS - GRUPO 003 – VIERNES 6 PM -9 PM

EQUIPO 12

1809285 DÍAZ TALAMANTES PERLA MILLET
1819111 HERNÁNDEZ BOLAÑOS LESLYE MARISOL
1824073 SOUSA RIVERA JUAN ANTONIO

PROF. MAYRA CRISTINA BERRONES REYES

MINERÍA DE DATOS

NUEVO LEÓN, MX, OCTUBRE 2020

1) Título de la base de datos.

- a) Nombre con el que se encuentra en la página encontrada: Informe mundial sobre la felicidad
- b) Url de la página: <https://www.kaggle.com/unsdsn/world-happiness>

2) Descripción de los datos.

- a) Tipo de datos

El tipo de dato que presenta nuestra base de datos está compuesto por 5 tablas, que hacen referencia a los años de 2015, 2016, 2017, 2018 y 2019, cada una con 141 datos que contiene en sus columnas: el país, la región donde pertenece, PIB per cápita, Familia, Esperanza de vida, Libertad, Generosidad, Confianza, Corrupción del gobierno.

- b) Descripción

Como lo mencionamos anteriormente cada tabla está compuesta por 8 columnas, cada una de ellas tiene el promedio de la respuesta obtenida a través de una encuesta, cada una de estas columnas describen en qué medida estos factores contribuyen cada variable, representada por el encabezado de cada columna para cada país y región, estas medidas están en un rango del 0-10, siendo el 0 lo que causa menor felicidad y el 10 como máxima felicidad.

3) Justificación del uso de datos.

- a) ¿Cuáles fueron las características que les llamó la atención de los datos? ¿Qué les hizo querer trabajar con ellos?

La primera característica es el tema, el cual, tiene como base principal la felicidad, otra característica es que se puede apreciar cómo las personas perciben la felicidad, además de cómo está distribuida en diferentes rubros y cómo es que para cada país cambia la relación y prioridad de cada variable en cuanto a la felicidad.

- b) ¿Qué beneficio encuentran de trabajar con estos datos?

A nosotros nos beneficia directamente en la obtención de nuevos conocimientos relacionados con la felicidad y cómo se comporta dependiendo del entorno de la persona encuestada, así mismo, logramos aplicar los conocimientos adquiridos a lo largo del semestre en un caso práctico de manera efectiva y lograremos comprender de mejor manera los mismos desde su naturaleza hasta su alcance.

Para nuestro trabajo es beneficioso, ya que, la base está muy bien construida y contamos con la suficiente información como para usar los métodos aprendidos en clase con facilidad.

4) Planteamiento del problema.

a) Problemática

Nuestro problema principal se centra en analizar cuanta importancia tiene el dinero para determinar la felicidad de un país o un conjunto de países, de igual forma, podemos analizar como una segunda problemática, cuál variable, de las presentadas en la base de datos; familia, esperanza de vida, libertad, generosidad, corrupción del gobierno, tiene mayor importancia para determinar la felicidad de un país.

b) Pensando como empresa.

Debido a que las interpretaciones finales de nuestra la base de datos se centran en describir la importancia que tienen las variables para la determinación de la felicidad, podemos visualizar al gobierno de un país desarrollado o subdesarrollado como una empresa, de tal modo que, nuestra conclusión del objetivo principal ayude a dicho país a determinar de qué forma podría aumentar la economía para lograr unos ciudadanos más felices. Si nos enfocamos en nuestro objetivo secundario, la interpretación de las variables pueden ser utilizadas por una empresa de marketing para diseñar estrategias publicitarias, enfocadas hacia la variable x (que corresponde a la de mayor impacto a la felicidad), esto con el fin de crear anuncios donde se asocie a la marca con la variable x, es decir, crear la idea de que el producto presentado se relacione con la variable x, lo que a su vez se asocia con la felicidad con el objetivo de difundir la marca y aumentar más ventas del producto.

5) Objetivo final.

Objetivo principal:

- Identificar la correlación entre la economía y la felicidad de un país.
- Analizar la influencia de la economía, de un país desarrollado, con la felicidad de este y de igual forma analizar esta influencia para los países subdesarrollados.
- Interpretar y concluir qué grupo de países, los desarrollados o los subdesarrollados, son lo que mayor importancia tiene la economía para definir su felicidad.

Objetivo secundario:

- Analizar qué factor/es son los que tienen mayor importancia en la felicidad de un país, esto a través de la correlación entre cada variable con su nivel de felicidad.
- Comparar los factores de mayor importancia en cuanto a la felicidad entre países desarrollados y subdesarrollados.

6) Planeación de la herramienta a utilizar.

a) Técnicas

Utilizaremos las técnicas descriptivas como el clustering, visualización y reglas de asociación.

Clustering: Aplicaremos la técnica del clustering para hacer agrupaciones de los países conforme a un rango en el que se encuentre su PIB, esto nos ayudará a clasificarlos como países desarrollados y subdesarrollados, lo que, a su vez será de gran utilidad para llegar a nuestro objetivo principal, el cual, es analizar la relación entre economía y felicidad para los países desarrollados, así como para los países subdesarrollados.

Algoritmo:

Usaremos datos externos para determinar los países que son desarrollados y subdesarrollados, posteriormente realizaremos una tabla en la que estarán separados estos datos como países subdesarrollados o desarrollados, usaremos el algoritmo de **K-Means** porque consideramos el más adecuado para la técnica del clustering.

Reglas de asociación:

Para la asociación utilizaremos la herramienta de correlación entre cada uno de los componentes del indicador de felicidad para tratar de encontrar cómo se comportan las variables frente a las demás, lo que nos sirve para poder tomar conclusiones sobre qué pasa si se le da más importancia a uno de estos componentes y poder explicar si el dinero está mayormente relacionado con la felicidad.

Algoritmo

Utilizaremos los datos de la base para medir las **correlaciones** entre los parámetros del indicador felicidad y se generara una tabla con los resultados obtenidos.

Visualización:

Para la visualización se tiene pensado utilizar la herramienta de Tableau para generar mapas y gráficos de dispersión con los datos obtenidos de la correlación y lograr tener de una manera más visual los datos obtenidos. De esta forma generaremos que los resultados sean fáciles de entender para todo público.