

Tarea 1 Suicide

Marcelo Alberto Sanchez Zaragoza

8 de febrero de 2021

1. Introducción

De la base de datos proporcionada se pretende realizar un análisis exploratorio y descriptivo de los datos para poder dar respuesta a algunas preguntas que nos hemos planteado al principio y que al tener los resultados de dicho análisis vamos a poder dar alguna respuesta. Las preguntas que nos hemos planteado son las siguientes:

- ¿Qué país presenta más casos de suicidio a lo largo del tiempo?
- ¿Qué rango de edad presenta un mayor número de suicidios?
- ¿Qué generación presenta mayor número de suicidios?

- ¿El número de suicidios tiene alguna relación con el ingreso per capita?
- ¿Hay alguna otra relación entre las variables que nos proporcionan?

Al final del reporte se presentara las conclusiones y daremos respuestas a las preguntas que se pudieran responder, cabe resaltar que no solo nos limitaremos a responder estas preguntas, también se pretende buscar relaciones entre las distintas variables para entender mejor el fenómeno. Para realizar el análisis antes mencionado nos apoyamos del lenguaje de programación Python.

2. Descripción de los datos

La base de datos que lleva por nombre suicide.data cuenta con 2780 filas a las que llamamos registros y con 12 columnas a las que llamaremos variables. Entre las variables que podemos observar en la base de datos son las siguientes: 'country', 'year', 'sex', 'age', 'suicides_no', 'population', 'suicides/10k', 'country-year', 'HDI for year', 'gdp_for_year(\$)', 'gdp_per_capita(\$)' y 'generation'. La variable HDI nos dice el índice de Desarrollo Humano, GPD el producto interno bruto y per capita.

La base de datos cuenta con información de 101 países, dentro de cada país la información se encuentra dividida en años y los correspondientes años están divididos en dos grupos, estos dos grupos corresponden al sexo, masculino y femenino. En los grupos de sexos tenemos intervalos en las edades, estos intervalos que nos presentan son seis. Los intervalos de edades(años) son las siguientes: 5 a 14, de 15 a 24, de 25 a 34, de 35 a 54, de 55 a 74 y de 74 a más.

También la base de datos cuenta a que generación pertenece el registro, entre las generaciones que nos presenta la base de datos están: Boomers(1949-1968), G.I.Generation(1901-1927), Generation X(1969-1980), Generation Z(1994-2010), Millenials(1980-2000) y Silent(1930-1948).

Cada registro nos proporciona información de un determinado grupo de individuos con cierta edad, nacionalidad, sexo, generación, año del registro ,etc. Algunos países cuentan con registros que van desde 1987 hasta 2016 y otros no tienen el mismo número de años registrados.

La variable del índice de Desarrollo Humano(HDI) no cuenta con algunos registros y más adelante se explicara como se trabajo la falta de datos.

3. Exploración de los datos

Se realizo una primer consulta para agrupar el número de suicidios por país y poder encontrar aquel que tuviera el mayor número de registros. Después encontramos que en el cuadro 3.1 el país que presento el mayor número de suicidios a lo largo de los años fue la Federación Rusa con un total de 1,209,742 suicidios.

suicides_no	
country	
Albania	1970
Antigua and Barbuda	11
Argentina	82219
Armenia	1905
Aruba	101
...	...
United Arab Emirates	622
United Kingdom	136805
United States	1034013
Uruguay	13138
Uzbekistan	34803

Cuadro 3.1: Número de suicidios por país.

En el cuadro 3.2 nos proporciona el intervalo de edad que cuenta con el mayor número de suicidios a lo largo de los años que tiene la base de datos. El intervalo de edad que tiene el mayor número de suicidios es el intervalo que va de los 35 a los 54 años de edad, con un total de 2,452,141.

suicides_no	
age	
15-24 years	808542
25-34 years	1123912
35-54 years	2452141
5-14 years	52264
55-74 years	1658443
75+ years	653118

Cuadro 3.2: Número de suicidios por intervalo de edad.

Otro dato que se puede encontrar es que generación presento una mayor número de suicidios a lo largo de los años, el cuadro 3.3 nos dice que la generación que presenta el mayor número de casos de suicidio fue la generación Boomers con un total de 2,284,498 suicidios. También encontramos que el sexo que presento más suicidios fue el sexo masculino con un total de 5,188,910 casos de suicidios y lo podemos observar en el cuadro 3.4.

suicides_no	
generation	
Boomers	2284498
G.I. Generation	510009
Generation X	1532804
Generation Z	15906
Millenials	623459
Silent	1781744

Cuadro 3.3: Número de suicidios por generación.

Para poder analizar un poco más a detalle la relación que se pudiera dar entre ciertas variables que contiene la base de datos, nos vamos a enfocar

suicides_no	
sex	
female	1559510
male	5188910

Cuadro 3.4: Número de suicidios por sexo.

en año 2010. En este año queremos observar si existe cierta correlación entre las variables. En la figura 3.1 observamos que hay una correlación paqueña entre la variable indice de Desarrollo Humano y el número de suicidios en el 2010, esta correlación es de 0.1609, otra correlación interesante es entre la variable PIB per capital y la tasa de suicidios ya que es muy pequeña, la correlación entre estas dos variables es de 0.0133.

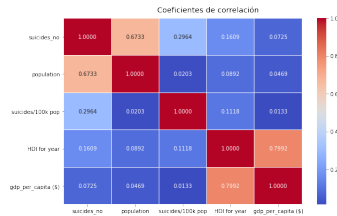


Figura 3.1: Coeficientes de correlación del 2010.

Nuestra siguiente figura 3.2 nos muestra los coeficientes de correlaciones del 2013, los valores no cambiaron mucho. Comparamos estas dos figuras para observar si hay muchos cambios al pasar unos años pero los coeficientes sigue siendo casi iguales.

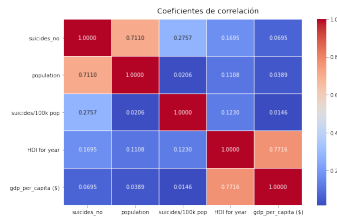


Figura 3.2: Coeficientes de correlación del 2014.

Los anteriores datos no ayudaron a poder contestar algunas de las preguntas que nos formulamos al principio, muchas de ellas se contestaron al momento de presentar cada una de las figuras anteriores.

3.1. Analisis de México y Estados Unidos

En este pequeño apartado vamos a realizar un análisis de dos países: México y Estados Unidos. Se va a observar el comportamiento de algunas de las variables que tienen los registros y se van a comparar entre ellas.

La figura 3.3 que estamos presentando nos muestra como ha aumentado el número de suicidios a lo largo de los años en el país de México. También mostramos el comportamiento que ha tenido el país de Estados Unidos, en la figura 3.4 se puede ver como ha ido creciendo el número de suicidios a lo largo del tiempo.

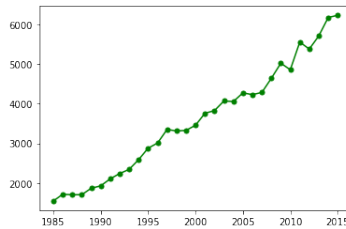


Figura 3.3: Gráfico de número de suicidios en México de 1985 a 2015.

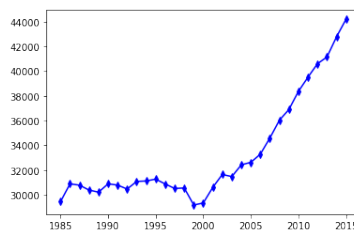


Figura 3.4: Gráfico de número de suicidios en Estados Unidos de 1985 a 2015.

Al comparar ambas figuras se puede observar que el país de México ha tenido un crecimiento constante a lo largo de los años mientras que en Estados Unidos se puede observar que el comportamiento es distinto ya que durante los años de 1998 a 2003 se ve que la curva tiene un descenso en el número de suicidios pero después crece de forma acelerada y supera en cantidad a México.

Los siguientes cuadros 3.5 y 3.6 nos dan el número total de suicidios a lo largo de los años(1985-2015) dividido por sexo. Se puede observar que para ambos países el sexo masculino ha presentado un número superior al sexo femenino, pero el país que tiene más casos de suicidios es Estados Unidos y se ve presente en el sexo masculino.

suicides_no	
sex	
female	19334
male	91805

Cuadro 3.5: Número de suicidios por sexo en México.

suicides_no	
sex	
female	213797
male	820216

Cuadro 3.6: Número de suicidios por sexo en Estados Unidos.

Algo que se puede mencionar o tomar en cuenta es que la población total de cada país es diferente. Lo siguiente que se va a mostrar son las correlaciones que existe entre cada una de las variables de ambos países.

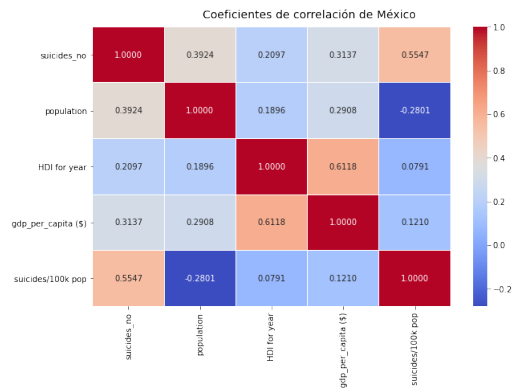


Figura 3.5: Coeficientes de correlación en México.

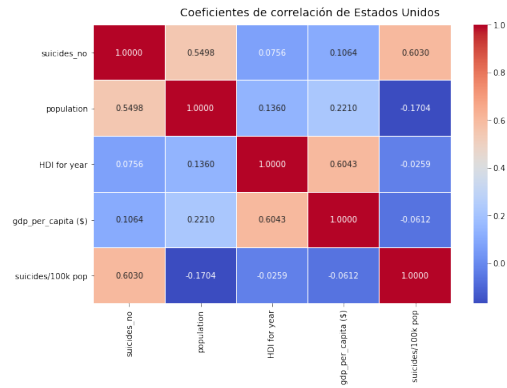


Figura 3.6: Coeficientes de correlación en Estados Unidos.

Observamos que para México(figura 3.5) hay una correlación de 0.2097 entre el número de suicidios y el índice de Desarrollo Humano, también hay una correlación entre el número de suicidios y la población, aquí tenemos un valor de 0.3924. Para Estados Unidos(figura 3.6) no pasa lo mismo ya que notamos que el coeficiente para estos casos son ligeramente distintos, para el coeficiente de correlación entre el número de suicidios y el índice de Desarrollo Humano es 0.0756 y para el coeficiente de número de suicidios y población es de 0.5498.

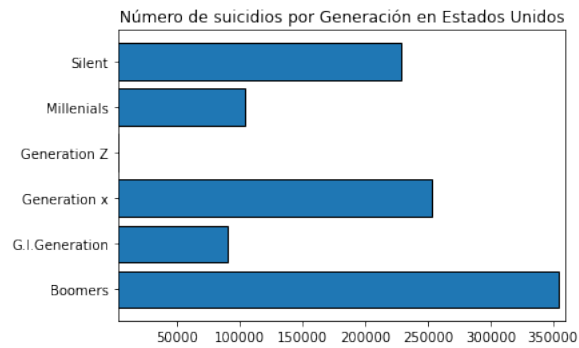


Figura 3.7: Número de suicidios por Generación en Estados Unidos.

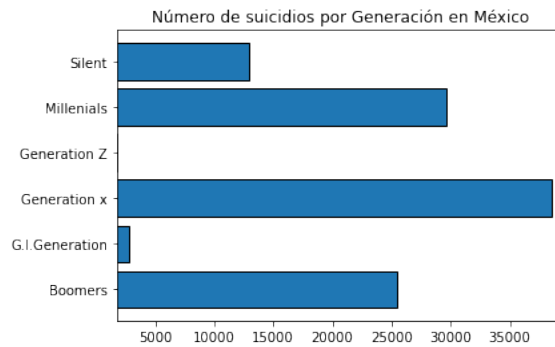


Figura 3.8: Número de suicidios por Generación en México.

En la figura 3.7 y figura 3.8 mostramos la cantidad de suicidios a lo largo del tiempo dividida por generación, se observa que para México la generación que realizó un mayor número de suicidios fue la generación X y para Estados Unidos fue la generación Boomers. Si se comparan ambas cantidades vemos que Estados Unidos tiene un número bastante grande comparado contra el número que tiene México con la generación X.

4. Conclusión

Después de analizar los datos para el caso donde tomamos la información de todos los países y cuando tomamos el caso donde solo comparamos entre México y Estados Unidos, se observó una cierta relación entre las variables del número de suicidios y el índice de Desarrollo Humano. Observe que para poder trabajar con los datos de la variable "HDI for year" se hizo primero una búsqueda de la media para poder sustituirla en aquellos registros vacíos. Para el caso donde se tomaron todos los países primero se buscó la media de todos los países y este dato fue el que se colocó en los lugares vacíos. Para el caso donde se trabajó solamente con dos países se buscó la media de cada uno de ellos y ese se colocó en los lugares vacíos. Las preguntas que al principio se plantearon se pudieron resolver conforme se fueron presentando los datos, también se respondieron las preguntas para el caso de México con Estados Unidos. En este último caso se hicieron algunas comparaciones entre ambos países y se mencionaron algunas diferencias importantes como la población que presenta cada uno y sus índices de Desarrollo Humano.

La información que nos podría haber ayudado a entender el número de suicidios en cada país pudo haber sido algún indicador de consumo de drogas, indicador de algún tipo de acoso o un indicador de acceso a productos o artículos letales como armas de fuego, etc. Con estas variables propuestas se podría buscar alguna correlación que pudiera existir entre las que ya se mencionaron al principio y el número total de suicidios.

Bibliografía Consultada

- T. Hastie, R. Tibshirani, J. Friedman. The elements of statistical learning. Data mining, inference and prediction. 2nd. edition. Springer, 2009.
- Henry, A. F. and J. F. Short, Suicide and Homicide (New York: Free Press, 1954)