****

**Universidad Internacional de La Rioja**

**Facultad de Ingeniería y Tecnología**

# **Máster Universitario en Análisis y Visualización de Datos Masivos / Visual Analitics & Big Data**

**INFORME ESTADISTICO**

|  |  |
| --- | --- |
| **Actividad de estudio presentado por:** | **Juan David Escobar Escobar** |
| **Tipo de trabajo:** | **Actividad** |
| **Modalidad:** | **Individual** |
| **Director/a:** | **Yamila García Martínez** |
| **Fecha:** | **Enero 2022** |

**Índice de contenidos**

[Máster Universitario en Análisis y Visualización de Datos Masivos / Visual Analitics & Big Data 1](#_Toc92481530)

[1. Planteamiento del problema 4](#_Toc92481531)

[1.1 Formulación del problema 4](#_Toc92481532)

[1.1.2 Problema general 4](#_Toc92481533)

[1.2 Objetivo general 5](#_Toc92481534)

[2. Marco Metodológico 5](#_Toc92481535)

[2.1 Variables 5](#_Toc92481536)

[2.2 Metodología 5](#_Toc92481537)

[2.3 Tipo de estudio 5](#_Toc92481538)

[2.4 Población 5](#_Toc92481539)

[2.5 Métodos de análisis de datos 6](#_Toc92481540)

[a. Análisis descriptivos, 6](#_Toc92481541)

[3. Tablas, graficas, métricas e interpretación 6](#_Toc92481542)

[4. Análisis discusión de resultados 9](#_Toc92481543)

[Se hace una comparación entre los resultados y el problema planteado 9](#_Toc92481544)

[5. Conclusiones 9](#_Toc92481545)

[En función a los objetivos formulados 9](#_Toc92481546)

[6. Recomendaciones 9](#_Toc92481547)

[En función a los objetivos formulados 9](#_Toc92481548)

[7. Anexos 9](#_Toc92481549)

[En función a los objetivos formulados 9](#_Toc92481550)

[7.1 Encuesta 9](#_Toc92481551)

[En función a los objetivos formulados 9](#_Toc92481552)

[7.2 Base de datos 9](#_Toc92481553)

[En función a los objetivos formulados 9](#_Toc92481554)

[8. Bibliografía 10](#_Toc92481555)

[9. Referencias bibliográficas 12](#_Toc92481556)

**Índice de tablas**

# **Planteamiento del problema**

La situación problemática de la epidemia por COVID-19 a generado un gran impacto a nivel mundial, no solo en la cantidad de fallecimientos por el mismo, si no también en diferentes sectores de la sociedad, tales como la economía, el político, la educación, los hábitos, la salud física y mental, entre otros.

Debido al brote de Covid-19 muchos gobiernos tuvieron que imponer la medida de cuarentena por seguridad y para evitar expandir el riesgo de infección. El encierro genero grandes cambios y complicaciones a nivel de salud física y mental, lo cual genero problemas de ansiedad, estrés y depresión.

No bastando con lo anterior, estudios indican que los efectos de la COVID-19 sobre la salud mental y el bienestar de los niños y jóvenes podría prolongarse durante muchos años (UNICEF, 2021). Los confinamientos y las restricciones de momento impuestas por la pandemia alejaron a ese colectivo de componentes clave para el desarrollo de los niños, como la familia, los amigos, las aulas y el juego, lo que dejará una “huella imborrable” en ellos (Naciones Unidas, 2021).

Aunque las prácticas presenciales tienen componentes irremplazables para el alumnado de ciencias de la salud, la crisis actual nos ha hecho plantearnos en nuestro entorno tanto la adaptación de los métodos docentes como los sistemas de gestión de las prácticas (Revista de Psiquiatría y Salud Mental - Journal of Psychiatry and Mental Health, 2021).

Todos estos aspectos mencionados anteriormente me conllevan a realizar la presente investigación **¿Cómo impacta el COVID-19 la educación, el estilo de vida social y la salud mental de los estudiantes?** (Encuesta en línea desde 13 julio a 17 de julio 2020).

# **Formulación del problema**

# **Problema general**

**¿Cuál es la incidencia del COVID-19 con la educación, el estilo de vida social y la salud mental de los estudiantes, en la India?** El COVID-19 incide significativamente con el impacto en los cambios de estilo de vida, creación de nuevas formas educativas y la afectación de la salud mental de los estudiantes y las personas en la India debido a diversos cambios obligados de comportamiento y hábitos inusuales.

# **Objetivo general**

Determinar como se relaciona la COVID-19 con la educación, el estilo de vida social y la salud mental de los estudiantes, en la India.

# **Marco Metodológico**

# **Variables**

**Tabla 1. Descripción de variables de la encuesta.**

**Graphical user interface, application

Description automatically generated**

Como instrumento de recogida de datos, se utilizo una encuesta tomada de la base de datos abierta Kaggle (Nidhi, 2020), dicha encuesta divide los datos en 4 grupos o categorías, los cuales se describen a continuación:

1. **Datos demográficos**: Edad y Región.
2. **Datos rutina diaria**: Promedio tiempo en clases online, promedio de tiempo utilizado para el auto aprendizaje.
3. **Valoración aprendizaje en línea:** Calificación o Raiting de las clases en línea.
4. **Valoración hábitos saludables o estilo de vida**: Tiempo promedio horas de sueño, cambio de peso, tiempo promedio de ejercicio físico, numero de comidas al día, supresores de estrés, acercamiento con los miembros de la familia o amigos, entre otros.

# **Metodología**

El método utilizado en la presente investigación es cuantitativo observacional (Encuesta 19 variables Liker Scale).

# **Tipo de estudio**

Básico descriptivo correlacional, Experimental longitudinal.

# **Población**

La población esta constituida por **1182** individuos de diferentes edades y centros de educativos en la capital de India Delhi y fuera de la capital. En la **Figura 1** se describe el detalle demográfico de la población. La edad media de las personas es de 20 años, de las cuales el 61% residen en Delhi capital de la India y el 39% en regiones diferentes. Se agrupan las personas en 3 grupos de edades (6,95 – 24,3], (24,3 – 41,7], (41,7 – 59 - 1], los cuales representan jóvenes, adultos y adultos mayores, la población mas representativa es la de los jóvenes con un 90.1% de participación, seguida por la de los adultos con un 9.1% y los adultos mayores con un 0.8%. Por ultimo y no menos importante se encuentra que esta población esta en un rango de [7-59] años.

**Figura 1. Detalle distribución demográfica.**

**A picture containing table

Description automatically generated**

# **Métodos de análisis de datos**

# **Análisis descriptivos,**

Para hallar las tablas, graficas, medidas de tendencia central, dispersión y análisis de datos se utilizo el Software R-Studio.

# **Tablas, graficas, métricas e interpretación**

A continuación, se muestra un análisis descriptivo y correlacional entre las variables de cada uno de los ámbitos que se describen en los grupos de datos de la base de datos. Análisis de rutina diaria académica y los individuos clasificados en 3 grupos de edades (6,95 – 24,3], (24,3 – 41,7], (41,7 – 59 - 1].

**Tabla 1.** diferencias entre las variables del grupo de datos de rutina diaria (Tiempo clases online, autoestudio, tiempo de sueño, tiempo en redes sociales y ejercicio físico) distribuidos en 3 rangos de edad (jóvenes, adultos, adulto mayor).

**A picture containing text, receipt, battery

Description automatically generated**

De acuerdo con los resultaos obtenidos en la **Tabla 1**, se puede observar que los estudiantes invierten un mínimo de 2 horas y máximo de 12 horas en auto aprendizaje, el tiempo medio de aprendizaje para todas las edades es de 2,7 horas diarias. El tiempo promedio para todas las edades de las clases en línea fue de 2,8 horas diarias, sin embargo, el rango mas juvenil presenta una media un poco mas alta de 3,25 horas diarias dedicadas y menor para la población adulta mayor con una media de 2.33 horas diarias.

El rango medio de horas de sueño se comprende entre 4 y 15 horas, la población total invierte 7,5 horas de sueño o descanso diario en promedio para recuperarse de sus actividades de rutina diaria.

El habito de ejercicio físico es uno de los hábitos saludables mayormente afectados, los estudiantes de todas las edades, solo invierten 0.039 horas de ejercicio físico, solo un 12,3% de los estudiantes invierte un tiempo considerado como saludable entre 1.5 - 3.33 horas diarias, y solo un 0.2% de la población total de los estudiantes practica ejercicio de alta intensidad, lo anterior tiene mucho sentido debido a que muchos estudiantes han concentrado mas su tiempo en actividades que implican mayor responsabilidad y se pueden efectuar de manera virtual o remota a lo que se refiere el estudio en línea, auto-estudio y un poco de ascio y socialización virtual por medio de las redes sociales con una media de 2,35 horas. También se debe tener presente que la salud tanto física como mental de los estudiantes se puede afectar por otras variables o factores como el número de comidas al día, falta de espacios de socialización, restricciones para hacer ejercicio al aire libre, situación critica de salud mundial, y los extremos que desequilibran cada uno de los hábitos de las personas.

**Figura 2. Preferencia red social.![A picture containing timeline

Description automatically generated]()**

Escuchar música es la actividad favorita de los estudiantes para reducir el estrés, un 23,4% de los estudiantes prefieren esta opción, otras actividades como jugar en línea, leer libros, redes sociales y dormir se encuentran en el Top de las mas populares, dichas actividades se realizan de manera individual, las cuales son mas accesibles en tiempos de pandemia, según la Dirección general de comunicación social cuando las personas no tienen interacción ni relaciones subjetivas que proporcionen seguridad y predictibilidad puede causar un impacto que inhibe las conexiones cerebrales que los jóvenes requieren para poder desarrollar después otro tipo de habilidades más complejas, construir humanidad, y situarse en la sociedad, lo anterior (Dirección general de Comunicación Social, 2021).

**Figura 3. Proporción de calificación clases online por Edades.**

**Chart, bar chart

Description automatically generated**

**La Figura 2** muestra 2 gráficos, en el primero se detalla de manera general la percepción de los estudiantes con respecto a la calidad de las clases Online, el porcentaje mas alto de calificación es del 34,9% el cual indica la calificación mas baja “**Very Poor**”, por otro lado el valor de la media para la calificación obtenido fue de apenas del 13,9%, lo cual indica un descontento por parte de los estudiantes y quizá también falta de calidad en las clases por parte de las instituciones, ya que es una nueva modalidad y no se cuenta con experiencia y adaptación a las mismas.

El grupo de estudiantes mas jóvenes califican en mayor proporción “**Very Poor**” comparado con la población de mediana edad y los adultos mayores, de hecho los adultos mayores tienen a conformarse, esto porque ya cuentan con experiencia previa y se adaptan fácil al cambio o tienden a ser mas resilientes, mientras que los jóvenes tienden a ser mas dependientes, la población de edad media hoy en día esta mas digitalizada y se adapta mejor que los otros dos grupos ya que se han adaptado fácilmente a trabajar mediante dispositivos electrónicos, ya que esto optimiza sus tiempos para cumplir con la alta demanda de responsabilidades que hoy en día demandan las organizaciones.

**Figura 4. Grafica de lo que mas extrañan los estudiantes**

**Chart, histogram

Description automatically generated**

De la grafica de la **Figura 4** cabe destacar que lo que mas extrañan los estudiantes es el colegio, y es natural ya que este es el segundo hogar de ellos, donde interactúan con sus compañeros y profesores de clase, en el top 5 de cosas que extrañan los estudiantes se las instituciones educativas, encuentran los amigos, los viajes, aprender libremente y comer afuera. Un porcentaje debajo del 5% no extraña nada o solo los amigos y ejercitarse, quizá están acoplados a un estilo de vida tal como lo plantea la situación.

**Figura 5. Diferencias entre las variables de cercanía familiar, problemas de salud, productividad y edad**

**Chart, treemap chart

Description automatically generated**

En **Figura 5** se analizan 3 variables binarias Familia, Salud y productividad. Con respecto a la salud un porcentaje relativamente bajo del 13,7% sufre problemas de salud, considerando la situación del Covid-19, un hallazgo importante es que la población que esta entre los (41.7 y 59.1) dice no tener ningún problema de salud, mientras que las poblaciones mas jóvenes un pequeño porcentaje si. Como resultado de las restricciones, situación de salud publica, las personas han estado mas cercanas de su familia arrojando un promedio del 70.3% de la población. El tiempo de productividad esta en un 48.6% y para la población de edad mayor alcanza el 75%, gracias a la virtualidad las personas lograron tener mayor productividad, al optimizar tiempo y espacio en un solo lugar, reducir actividades ineficientes y tener mas tiempo de concentración y productividad.

# **Conclusiones**

# En base a los resultados obtenidos se encuentra un alto impacto en los hábitos saludables de los estudiantes, también algunos factores a nivel de estado anímico se encontraron afectados, tales como la falta de socialización, el poder salir al aire libre y practicar actividades físicas, visitar restaurantes, compartir con los amigos, viajar, conocer el mundo y disfrutar de nuevas experiencias.

# Los estudiantes adquirieron un buen habito de estudio gracias a la virtualidad y al enfoque en estas actividades que implican un grado de responsabilidad, teniendo en cuenta que muchos estudiantes y profesores no se adaptaron aún de la mejor manera a las clases online, ya que es primera vez que se enfrentan a ello.

# Un factor importante por mencionar es que las personas recuperaron el valor de las familias, los resultados indican que las personas sienten mas confianza de sus personas mas cercanas y tienden a estar mas unidos, esto también es un efecto de la situación de salud publica, ya que las familias se cuidan entre si para superar la situación.

# Para finalizar se debe mencionar que esta situación afecto gravemente la vida social, física y emocional de las personas, dejando en algunos situaciones de estrés, ansiedad y depresión, algunas categorías y porcentajes nos muestran que hay personas poco motivadas y algunas poblaciones de rangos de edad mayor que se adaptaron mas fácil que las mas jóvenes, estos problemas de salud mental deberán ser tratados para evitar que los estudiantes y las personas sigan pasando un mal momento por mucho tiempo.

# **Anexos**

# Se adjuntan diferentes artefactos en repositorio personal de GitHub utilizados para llevar a cabo el informe estadístico entre los cuales están:

# **Base de datos:** COVID-19 Survey Student Responses.csv.

# **Informe estadístico:** Informe.docx.

# **Script R-Studio:** Actividad 1 R.

# **Herramientas:** R-Studio (<https://www.rstudio.com/>)

# **Librerias:** readr, testthat, plyr, dplyr, ggplot2, Hmisc, corrplot, plyr, fdth, berryFunctions, tidyverse, PerformanceAnalytics, rpivotTable, sqldf, reshape2, ggeasy, grid, gridBase, forecast.

# **Base de datos (Encuesta)**

# A picture containing window, building, indoor, shoji Description automatically generated

# **Bibliografía**

**https://www.youtube.com/watch?v=CuKr7GzohbI**

**https://www.youtube.com/watch?v=VWOvYPTrrNA**

**https://datos.gob.es/es/documentacion/guia-**

[**https://datos.gob.es/es/documentacion/guia-practica-de-introduccion-al-analisis-exploratorio-de-datos**](https://datos.gob.es/es/documentacion/guia-practica-de-introduccion-al-analisis-exploratorio-de-datos)

[**https://www.youtube.com/watch?v=CIIzowIm9ZQ**](https://www.youtube.com/watch?v=CIIzowIm9ZQ)

[**https://miasesordetesis.com/como-identificar-y-definir-las-dimensiones-de-las-variables/**](https://miasesordetesis.com/como-identificar-y-definir-las-dimensiones-de-las-variables/)

[**https://www.youtube.com/watch?v=788WtRruUlU**](https://www.youtube.com/watch?v=788WtRruUlU)

[**https://www.researchgate.net/publication/347935769\_COVID-19\_and\_its\_impact\_on\_education\_social\_life\_and\_mental\_health\_of\_students\_A\_Survey**](https://www.researchgate.net/publication/347935769_COVID-19_and_its_impact_on_education_social_life_and_mental_health_of_students_A_Survey)

[**https://www.kaggle.com/kunal28chaturvedi/covid19-and-its-impact-on-students**](https://www.kaggle.com/kunal28chaturvedi/covid19-and-its-impact-on-students)

[**https://soka.gitlab.io/blog/post/2019-02-14-diagramas-de-tarta-en-r/**](https://soka.gitlab.io/blog/post/2019-02-14-diagramas-de-tarta-en-r/)

[**https://rdrr.io/cran/plotrix/man/histStack.html**](https://rdrr.io/cran/plotrix/man/histStack.html)

**https://www.odiolaestadistica.com/estadistica-r/covarianza/**

**Herramientas: Jupyter Nootebooks Versión 6.4.2 (**[**https://jupyter.org/**](https://jupyter.org/)**), Python versión 3.0 (**[**https://www.python.org/**](https://www.python.org/)**), MongoDB**

# **Referencias bibliográficas**

**datos.gob.es. (22 de 9 de 2021). *datos.gob.es.* Obtenido de Guía Práctica de Introducción al Análisis Exploratorio de Datos: https://datos.gob.es/es/documentacion/guia-practica-de-introduccion-al-analisis-exploratorio-de-datos.**

**Rocio Chavez Ciencia de Datos. (2020). Histogramas en Python [video]. YouTube.**

**AIEngineering. (2020). Data Cleaning and Analysis using Apache Spark [video]. YouTube.**

**Un Analista de Datos Peruano. (2019). Para qué me sirve la Desviación Estandar? [video]. YouTube.**

**orvizar TV. (2020). Pandas desde cero #3 |Limpiar Datos | Valores nulos | Eliminar datos | tutorial español [video]. YouTube.**

**Dattos.org. (2019). Gráficos en Python con Matplotlib, Seaborn y Ploty. YouTube.**

**Tomas\_IA. (2020). Limpieza de datos con Python [video]. YouTube.**

**Escuela de datos. (28 de 6 de 2017). *Escuela de datos*. Obtenido de Introducción a Pandas y Jupyter Notebook de Python: http://es.schoolofdata.org/2017/06/28/introduccion-a-pandas-y-jupyter-notebook-de-python/**

**Invarato, R. (20 de 4 de 2021). *jarroba.com*. Obtenido de Instalar Apache Spark en cualquier sistema operativo y aprender a programarlo con Python, Scala o Java: https://jarroba.com/instalar-apache-spark-en-cualquier-sistema-operativo-y-aprender-a-programarlo-con-python-scala-o-java/**