

Mapa de ruta del caso práctico: Cuestionario PHQ9

1.- Preguntar

Mapa de ruta de caso práctico - Preguntar

La salud mental es una preocupación a nivel mundial, sobre todo a raíz de la pandemia de Covid-19. Para diagnosticar score de PHQ9, un especialista en el área de la salud utiliza diversas herramientas tales como cuestionarios, entrevistas y actividades que lo llevan a determinar no solamente si un paciente tiene score de PHQ9, sino el nivel que, en los casos más graves puede representar un riesgo para la vida del paciente. Una de esas herramientas es el cuestionario PHQ-9 (Patient health questionnaire), el cual es un cuestionario de 9 preguntas para evaluar la presencia de síntomas de depresión en un paciente.

Analizar scores de PHQ-9 y encontrar o no posibles relaciones con las características de los pacientes puede llevar a los especialistas de la salud mental a trazar estrategias que conlleven a prevenir niveles altos de score de PHQ9 en los pacientes.

Entregable

Una declaración clara de la tarea a realizar con el dataset elegido.

NOTA: Este es un ejercicio de carácter **académico**, como parte de la **tarea final del Certificado de carrera de Google: Análisis de datos.**

2.- Preparar

Mapa de ruta de caso práctico - Preparar

Se utiliza el dataset:

Dataset_14-day_AA_depression_symptoms_mood_and_PHQ9.csv,
el cual es de dominio público, y fue descargado de:

<https://www.kaggle.com/datasets/thedevastator/phq-9-depression-assessment>

by Sebastian Burchert; André Kerber; Johannes Zimmermann; Christine Knaevelsrud

EL utilizar un dataset de dominio público siempre conlleva a considerar algunos elementos importantes:

- El sesgo en la recopilación: el no haber sido parte de la recopilación de los datos, no se puede tener la garantía de que han sido recolectados sin sesgo alguno. Sin embargo, el respaldo que Kaggle le otorga a los datos es parte de lo que se debe considerar al momento de utilizar un dataset de este sitio.
- La autorización para utilizar los datos. Kaggle menciona en la descripción del dataset que es de dominio público, por lo que es posible ser utilizado para esta tarea.
- El formato en el cual los datos, se encuentran disponibles. Estos datos fueron descargados en un archivo csv, lo que facilita que puedan ser puestos en un dataframe para que puedan ser procesados en Rstudio.

Entregable

Una descripción de todas las fuentes de datos utilizadas.

3.- Procesar

Mapa de ruta de caso práctico - Procesar

1. Se crea una carpeta llamada PHQ9-MH Project para mantener organizados los archivos que se vayan generando durante el desarrollo del proyecto.
2. A continuación, se asegura que los datos que se utilizarán son del dominio público, para lo cual, se descarga el archivo en formato csv y se abre con Excel. Se revisan las columnas y se corrobora que no existe manera de identificar a los pacientes que respondieron la encuesta. No existen en el dataset datos sensibles, y los pacientes son identificados con un ID numérico consecutivo:1,2,3...etc.
3. Se identifica en el sitio de descarga de Kaggle un diccionario de datos, el cual facilitará la identificación de los datos de cada variable (Columna).
4. Las columnas q1 a la q14, q16, q46 y q47 contienen valores NA en su mayoría, por lo que no se analizarán en una primera etapa del análisis. La ausencia de valores en estas columnas puede significar que el paciente no quiso seguir respondiendo las encuestas, no pudo por algún motivo o su puntuación no resultó significativa para el estudio. El análisis de estas columnas, en una segunda etapa del proyecto, nos contestará estas preguntas. Por lo tanto, la generación de un segundo dataset (subset), nos permite organizar los datos para su posterior análisis.
5. Se trabaja en una primera etapa del proyecto con los primeros 9 días de la encuesta, se nota que en un mismo día se levanta varias veces la encuesta por lo que existen varias puntuaciones para un mismo día pero en diferentes horarios.
6. Además de las puntuaciones de PHQ9, se cuenta con un score de felicidad el cual representa un mayor score cuando la felicidad es mayor. Este valor puede ser importante más adelante ya que podría existir una correlación entre los porcentajes de PHQ9 y este mencionado score de felicidad.
7. Se cuenta con 16,160 encuestas realizadas a 185 pacientes.

Entregable

Documentación de la limpieza y manipulación de los datos.

4.- Analizar

Mapa de ruta de caso práctico - Analizar

1. Se procede a calcular las medidas de tendencia central.
2. Se añade una columna adicional con el promedio de la puntuación del score durante los 9 días. Este promedio también se grafica con el objetivo de encontrar patrones en la información.
3. Al calcular el promedio de los scores de los 9 días de PHQ9 por paciente y se encuentra como patrón que el promedio es el mismo a lo largo del día, y sin variaciones en los 9 días.
4. No se cuenta con la edad en algunos pacientes, ni el sexo, por lo que se deduce que son datos opcionales y que el paciente prefirió no proporcionar.
5. Construimos 3 datasets para visualizar: uno para score de PHQ9, otro para felicidad y un tercero con ambas. Todos los valores que se tienen en estos datasets están agrupados por paciente. Se utiliza el promedio generado en el paso 2 y el score de felicidad.
6. Se calcularon algunos datos para conocer algunas características de la población encuestada.

Entregable

Un resumen de los resultados del análisis de los datos del nuevo dataset (solo 9 días de encuestas)

5.- Compartir

Mapa de ruta de caso práctico - Compartir
<p>Se generan los siguientes gráficos:</p> <ul style="list-style-type: none">-Diagrama de barras por sexo-Boxplots de los scores de felicidad y PHQ9-Gráfica de Wrap con scores de felicidad, PHQ9 por sexo-Gráfica de dispersión con la correlación entre los promedios de PHQ9 y felicidad.
<p>Entregable</p> <p>Visualizaciones de respaldo y hallazgos clave</p>

6.- Actuar

Mapa de ruta de caso práctico - Actuar

Se concluye que:

- Existe una relación entre el promedio de lo que obtiene un paciente en el cuestionario PHQ9 y el promedio de lo que obtiene en el score de felicidad a lo largo de los 9 días.
- La correlación es de -0.6233787, lo que significa que a mayor score de PHQ9, menor felicidad.
- Así mismo, se concluye que existen 129 pacientes femeninos, 32 pacientes masculinos, 1 paciente transgénero y 24 pacientes omitieron comentar.
- En cuanto a las edades oscilan entre 12 y 57 años, con una edad promedio de 29.7778.
- Las pacientes femeninas obtienen mayor puntuación de score de PHQ9 y los pacientes masculinos mayor puntuación en felicidad.

Entregable

- Mapa de ruta del caso práctico
- Script en R para el análisis y la creación de gráficas
- Presentación en pdf con los resultados obtenidos.