

HackForGood - El síndrome de BurnOut (Segunda limpieza)

MAD09 - Maslash

2024-03-16

SEGUNDA LIMPIEZA DE DATOS

En esta segunda limpieza de datos lo único que hay que hacer es, a la base de datos limpiada anteriormente, añadimos la variable 'Cluster' para poder ver cómo afecta ser de un cluster u otro en los datos visualmente. Pues esta

```
library(readxl)

datos_encuesta <- read_excel("archivo_modificado0.xlsx")
encuesta_clusters <- read_excel("Clusters.xlsx")
head(encuesta_clusters)

## # A tibble: 6 x 21
##   Estudiante.42 Motivacion Eleccion.estudios Gustar.estudios Aprovechar.tiempo
##   <chr>          <chr>          <chr>          <chr>          <chr>
## 1 No           No           Si           No           Si
## 2 No           Si, mucho    Si           Si           Si
## 3 No           No           Si           Si           No
## 4 Si           Si, mucho    Si           Si           Si
## 5 Si           Sí, pero po~ Si           Si           Si
## 6 Si           Sí, pero po~ No           Si           No
## # i 16 more variables: Organizacion <chr>, Comodo.compañeros <chr>,
## #   Horas.sueño <chr>, Dormir.suficiente <chr>, Comidas.diarias <chr>,
## #   Bebidas.energéticas <chr>, Actividad.fisica <chr>, Tiempo.estudios <chr>,
## #   Metodología <chr>, Acompañado <chr>, Ambiente.competitivo <chr>,
## #   Comodo.competicion <chr>, Autonomía <chr>, Presion.tareas <chr>,
## #   Burnout <chr>, cluster_label <dbl>
```

Añadimos la variable de a qué cluster pertenece cada individuo, para esto, primero la seleccionamos del un excel y se la incorporamos al que habíamos modificado anteriormente.

```
# Extraemos la variable
cluster <- encuesta_clusters$cluster_label

# Unimos la base de datos anterior con la variable 'Cluster'
datos_encuesta4 <- cbind(datos_encuesta, cluster)
head(datos_encuesta4)
```

```
##   Estudiante.42   Motivacion Eleccion.estudios Gustar.estudios
```

## 1	No	No	Si	No
## 2	No	Si, mucho	Si	Si
## 3	No	No	Si	Si
## 4	Si	Si, mucho	Si	Si
## 5	Si	Si, pero poco	Si	Si
## 6	Si	Si, pero poco	No	Si
##	Aprovechar.tiempo	Organizacion	Comodo.compañeros	Horas.sueño
## 1	Si	Si, mucho	Si	6
## 2	Si	Si, mucho	Si	3
## 3	No	No, nada	Si	7
## 4	Si	Si, un poco	Si	6
## 5	Si	Si, un poco	Si	8
## 6	No	No, nada	Si	7
##	Dormir.suficiente	Comidias.diarias	Bebidas.energéticas	Actividad.fisica
## 1	No	Mas de 3	Nada	5 o mas
## 2	No	Menos de 3	1	1 - 2 dias
## 3	Si	Mas de 3	1	3 - 4 dias
## 4	Si	Menos de 3	Nada	1 - 2 dias
## 5	Si	3	1	1 - 2 dias
## 6	Si	3	Nada	3 - 4 dias
##	Tiempo.estudios	Metodología	Acompañado	Ambiente.competitivo
## 1	6	No	No	No
## 2	4	Si	Si	Si
## 3	2	No	No	Si
## 4	8	Si	Si	No
## 5	8	Si	Si	No
## 6	6	Si	No	Si
##	Comodo.competicion	Autonomía	Presion.tareas	Burnout cluster
## 1	Si	Si	No	Si 4
## 2	Si	Si	Si	No 2
## 3	Si	Si	No	No 3
## 4	No	Si	Si	Si 1
## 5	Si	Si	No	No 1
## 6	Si	Si	Si	Si 3

Lo descargamos como un excel. Este dataframe lo usaremos para realizar todos los gráficos del PowerBI

```
library(writexl)
```

```
## Warning: package 'writexl' was built under R version 4.3.3
```

```
write_xlsx(datos_encuesta4, "archivo_modificado.xlsx")
```