Ejercicio 3: EDA con DataSet más extenso

instituto tecnológico de morelia

Ejercicios de práctica

M.I.A. WILFRIDO CORTES OROZCO

2023

|  |  |  |
| --- | --- | --- |
| Nombre del Alumno | Número de control | Número de actividad |
|  |  |  |

# Descripción y objetivos de la actividad.

Objetivos:

* Que el estudiante pueda aplicar el análisis exploratorio, elimine las anomalías y los datos nulos de manera efectiva y adecuada.
* Que el estudiante pueda visualizar los datos.
* Que el estudiante pueda responder preguntas específicas sobre su dataset.
* Que el estudiante siga practicando sus conocimientos teóricos del análisis exploratorio.

Descripción de la actividad:

Para esta actividad se hará uso de un dataset más grande “**Practica\_Youtube.csv**”.

Se usarán las herramientas adquiridas para realizar el análisis de este dataset

1. Muestre los tipos de cada columna.
2. Identifique los nulos que hay por columna.
3. Trate los datos nulos.
4. Analice si una columna puede ser categorizada, si es así, realice la categorización.
5. Dependiendo del tamaño de suscriptores del canal, cree una columna con “size” y proponga las etiquetas que pueden usarse para la categorización, por ejemplo:

“mini: 10k, small: 100k, amateur: 1M, professional: 10M, star: +30M”

Como recomendación, primero analice los rangos de los datos, establezca rangos acorde a esa información.

1. Con la categorización general realizada, revise si hay inconsistencias, es decir, si se definieron 4 categorías para una columna, analice los datos que están fuera de esas categorías.
2. Haga una columna nueva llamada “average\_monthly\_earnings”, producto del promedio las columnas “lowest\_monthly\_earnings” y “highest\_monthly\_earnings”.
3. Identifique las anomalías para los datos numéricos y realice el tratamiento de las mismas.
4. Busque volver numéricas las columnas (tantas como sean posibles) y haga el análisis de correlaciones, puede imprimir la matriz; hay que identificar qué variables están más correlacionadas con average\_monthly\_earnings.
5. Responda las siguientes preguntas:
   1. ¿Qué tipo de correlación tiene el tipo de canal con las ganancias mensuales promedio?
   2. ¿Cuál país (abbreviation) tiene mayor ingreso mensual en promedio?
   3. ¿Qué Youtuber tiene menor ingreso mensual promedio?

tEntregue sobre este documento o sobre un documento de Jupyter-Notebook (preferido), no olvide incluir sus datos como son:

1. Nombre completo.
2. Número de control.
3. Número de actividad.