

Universidad Técnica Nacional

Minería de datos - ISW 911

Laboratorio # 3

Asignación Individual

- Ingrese a <https://www.kaggle.com/> si no tiene una cuenta, cree una, descargue el dataset **Women's E-Commerce Clothing Reviews**

The screenshot shows the Kaggle interface. On the left, there's a sidebar with options like 'Create', 'Home', 'Competitions', 'Datasets', 'Models', 'Code', and 'Discussions'. The main area displays the dataset 'Women's E-Commerce Clothing Reviews'. It includes a search bar at the top, followed by tabs for 'Data Card', 'Code (208)', 'Discussion (8)', and 'Download (3 MB)'. Below these are category filters: Business, Earth and Nature, Internet, NLP, Ratings and Reviews, and Text Mining. A large button labeled 'Download (3 MB)' is prominent. To the right, there's a 'Data Explorer' section showing the file 'Womens Clothing E-Commerce Reviews.csv' (8.48 MB), with a preview of the first few rows and columns.

- Guarde el archivo en su repositorio GITHUB de forma pública.
- Use Python o R Studio para realizar las siguientes tareas:
 - Cargue el set de datos en su ambiente de trabajo preferido
 - Conviértalo a dataframe
 - Utilice las funciones correspondientes para explorar la estructura de datos del dataset (nombres de columnas, tipos y cantidad de datos, ect)
 - Cree un nuevo set de datos que contenga únicamente las siguientes columnas

Age

Rating

Recommended IND

Positive Feedback Count

Division Name

Department Name

Class Name

- Determine cuales de las anteriores son variables categorías

- Genere los scripts para determinar los valores únicos para las columnas

Rating

Recommended IND

Positive Feedback Count

- Determine si el conjunto de datos contiene valores nulos, cuantos y en cuales columnas

- Cambie el nombre de las columnas al español.

- Genere los gráficos boxplot para cada uno de los elementos de cada variable categórica, para detectar valores atípicos presentes

- Genere los histogramas de la edad para cada una de las variables categóricas.

- Genere las siguientes agrupaciones de datos
 - Cantidad de datos por División
 - Cantidad de datos por Departamento
 - Cantidad de datos por Clase
- Convierta los resultados anteriores a dataframes, recuerde convertir los índices en columnas.
- Genere los gráficos de tipo PIE y de Barras. para los tres grupos anteriores
- Determine el promedio de la edad por cada una de las variables categóricas
- Documente apropiadamente las líneas de código que genera.
- Adjunte el archivo con el nombre **lab3_nombre.ipynb** o **lab3_nombre.R**, según su la plataforma de su elección, y adjúntelo al repositorio correspondiente del campus virtual, ***archivos con el formato diferente al solicitado no se revisaran y se calificarán con 0.***

----- Ultima Linea -----