Machine Learning project(Drug Consumption Classification.

Link:

Google Colab

https://colab.research.google.com/drive/1tEuPCZgNEKDP9vvSO03Fj8xLm6-kJD_n#scrollTo=WToUI68i0pXb

explanation part 1

import matplotlib.pyplot as pltimport seaborn as snsimport plotly.express as pxfrom plotly.subplots import make_subplotsimport plotly.graph_objects as goimport warningswarnings.filterwarnings("ignore")

Imports:These lines import necessary libraries for data visualization, including Matplotlib, Seaborn, and Plotly.warnings.filterwarnings("ignore") is used to suppress any warnings that may be generated during execution.

- ---

explanation part 2

df = pd.read_csv('/content/drug_consumption.csv')

Data Loading:This line reads the dataset 'drug_consumption.csv' into a pandas DataFrame named df.

- ---

explanation part 3

df.info()df.shapedf.describe()df.duplicated().sum()df.head()

Data Exploration:df.info() provides a concise summary of the DataFrame including column data types and missing values.df.shape returns the dimensions of the DataFrame (number of rows and columns).df.describe() provides summary statistics for numerical columns.df.duplicated().sum() counts duplicate rows in the DataFrame.df.head() displays the first few rows of the DataFrame.

- ---

explanation part 4

age = {...}gender = {...}education = {...}country = {...}ethnicity = {...}usage = {...}df['Age'] = df['Age'].replace(age)df['Gender'] = df['Gender'].replace(gender)df['Education'] = df['Education'].replace(education)df['Country'] = df['Country'].replace(country)df['Ethnicity'] = df['Ethnicity'].replace(ethnicity)df['Alcohol'] = df['Alcohol'].replace(usage)...df['VSA'] = df['VSA'].replace(usage)

Data Transformation:This section appears to be mapping encoded values to their corresponding categorical labels for better readability.Dictionaries (age, gender, education, country, ethnicity, and usage) are created to map encoded values to their corresponding categories.Each column in the DataFrame is transformed using replace() to replace the encoded values with their corresponding labels.

- ---

explanation part 5

num_cols = df.select_dtypes(include=['float64']).columnscat_cols = df.select_dtypes(include=['object']).columns

Feature Selection:This section identifies numeric and categorical columns in the DataFrame, which is essential for preprocessing and modeling.

- ---

explanation part 6

f, ax = plt.subplots(2, 4, figsize=(15, 10))ax = ax.flatten()for index, cols in enumerate(num_cols):sns.histplot(data=df, x=cols, ax=ax[index], kde=True)ax[index].set_title(cols)plt.tight_layout()plt.show()

Data Visualization:This section creates subplots to display histograms for numeric columns using Seaborn.

- ---

explanation part 7

f, ax = plt.subplots(2, 4, figsize=(15, 10))ax = ax.flatten()for index, cols in enumerate(num_cols):sns.boxplot(data=df, y=cols, ax=ax[index])ax[index].set_title(cols)plt.tight_layout()plt.show()

Data Visualization (Boxplots):Similar to the previous section, this code creates subplots to display boxplots for numeric columns using Seaborn.

- ---

explanation part 8

age_alcohol = df.groupby(['Age', 'Alcohol']).size().reset_index(name='count')# Similar grouping for other drugs (not shown in the provided code)Grouping and Aggregation:This section aggregates data by grouping based on age and drug usage (in this case, alcohol), then counts the occurrences.

px.bar(age_alcohol, y='count', x='Age', color='Alcohol', barmode='group', title='Alcohol use by age', text_auto=True)

Data Visualization (Bar Plot):This code creates a bar plot using Plotly Express to visualize the relationship between age and alcohol usage.

Similar code for visualizing drug usage by gender, age, etc. (not shown in the provided code)

Additional Data Visualization:This section likely contains additional code for visualizing drug usage by other demographic factors like gender, ethnicity, etc., using Plotly Express.

- ---

explanation part 9

UK = df[df['Country'] == 'UK']UK_alcohol = UK.groupby(['Alcohol']).size().reset_index(name='count')

Filtering Data:This code filters the DataFrame to include only records where the 'Country' column is 'UK', then aggregates data by 'Alcohol' usage in the UK.

- ---

explanation part 10

fig = px.pie(UK_alcohol, values='count', names='Alcohol', title='UK alcohol use status')fig.update_traces(textposition='outside', textinfo='label + percent + value')

Data Visualization (Pie Chart):This code uses Plotly Express to create a pie chart representing the distribution of alcohol usage in the UK.

- ---

explanation part 11

sns.heatmap(df[num_cols].corr(), annot=True, cmap='Reds')plt.show()

Data Visualization (Heatmap):This code generates a heatmap using Seaborn to visualize the correlation between numeric columns in the DataFrame.

Machine Learning

Introduction:
This project aims to explore and visualize patterns of drug consumption using demographic and behavioral data. By analyzing a dataset containing information on individuals' demographics, personality traits, and drug usage habits, we seek to understand how factors such as age, gender, education, and country influence drug consumption behavior. Through exploratory data

analysis (EDA) and visualization techniques, we aim to gain insights into drug consumption patterns and identify potential trends and correlations.

Data:
The dataset used in this project contains anonymized information collected from individuals regarding their demographics (age, gender, education, country, ethnicity), personality traits (Nscore, Escore, Oscore, Ascore, Cscore), and drug usage habits (Alcohol, Amphet, Cannabis, etc.). Each individual is categorized based on their drug usage frequency, ranging from 'Never Used' to 'Used in Last Day'. The dataset enables us to analyze how different factors influence drug consumption behavior.

Approach:

1. Data Loading and Preprocessing:
    - We start by loading the dataset into a pandas DataFrame and perform initial preprocessing steps such as handling missing values and replacing encoded values with readable labels. This ensures that the data is ready for analysis.
2. Exploratory Data Analysis (EDA):
    - We conduct exploratory data analysis to understand the distribution and characteristics of the dataset. This involves examining summary statistics, detecting outliers, and visualizing the distribution of variables using histograms, boxplots, and other visualization techniques.
3. Visualization of Drug Consumption Patterns:
    - We use various visualization tools such as Matplotlib, Seaborn, and Plotly Express to visualize drug consumption patterns across different demographic groups. This includes creating bar plots, pie charts, and heatmaps to analyze drug usage trends based on factors such as age, gender, education, and country.
4. Insights and Interpretation:
    - Through our analysis and visualization, we aim to uncover insights into drug consumption patterns and identify potential correlations between demographic factors and drug usage habits. We interpret the visualizations to understand which groups may be more susceptible to

certain types of drug consumption and how these patterns vary across different demographics.

Conclusion:
By leveraging exploratory data analysis and visualization techniques, this project provides valuable insights into drug consumption patterns and sheds light on the factors influencing these behaviors. Understanding these patterns can inform public health initiatives, targeted interventions, and further research into substance abuse prevention and treatment strategies. Additionally, the project demonstrates the power of data analysis and visualization in uncovering meaningful insights from complex datasets.

> Thank you so much, Prof.Leonid Zakirov for you efforts with us!!