

SAnálisis Exploratorio de Datos 2023

U1 - 1A

Uso básico de librerías de graficación

Nombre: David Aaron Ramirez Olmeda

Programa: Maestría en Ciencia de Datos e Información



Introducción

Esta tarea se basa en la exploración y visualización de datos generados a partir de un código que encontré en internet, tras haber sido inspirado por un instructivo video sobre información relacionada con aplicaciones de citas. La oportunidad de emplear estos datos en la creación de visualizaciones y poner en práctica los conceptos aprendidos en esta unidad resultó sumamente atractiva. En caso de tener el tiempo disponible, recomiendo ver el video mencionado, el cual se encuentra en la sección de referencias, ese material está relacionado al análisis exploratorio de datos y a la ciencia de los datos en general.

Desarrollo

Sin tantas complicaciones, veremos primero un fragmento del pequeño, pero de gran potencial, set de datos y lo usaremos para crear, en este caso, dos visualizaciones que darán gran información sobre el comportamiento de las personas en las apps de citas.

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import ipywidgets as widgets
from ipywidgets import interactive
```

```
In [6]: ruta_archivo = '/Users/aaron/Documentos/MCDI/Semestre 2/Análisis Explorati
user_data = pd.read_csv(ruta_archivo + '/user_data.csv')
print(user_data.head(2))
```

	gender	attractiveness	likes	matches
0	FEMALE	0.255020	46	3
1	FEMALE	0.239791	40	3

```
In [15]: def plot_2(x_column, y_column, gender_filter):
    filtered_data = user_data[user_data['gender'] == gender_filter]
    plt.figure(figsize=(10, 6))
    plt.scatter(filtered_data[x_column], filtered_data[y_column])
    plt.xlabel(x_column)
    plt.ylabel(y_column)
    plt.title(f'{y_column} vs. {x_column} ({gender_filter})')
    plt.grid(True)
    plt.show()

x_widget = widgets.Dropdown(options=user_data.columns, description='X:')
y_widget = widgets.Dropdown(options=user_data.columns, description='Y:')
gender_filter_widget = widgets.Dropdown(options=user_data['gender'].unique())

interactive_plot = interactive(plot_2, x_column=x_widget, y_column=y_widget)
output = interactive_plot.children[-1]

widgets.VBox([gender_filter_widget, x_widget, y_widget, output])
```

Filtrar por ...

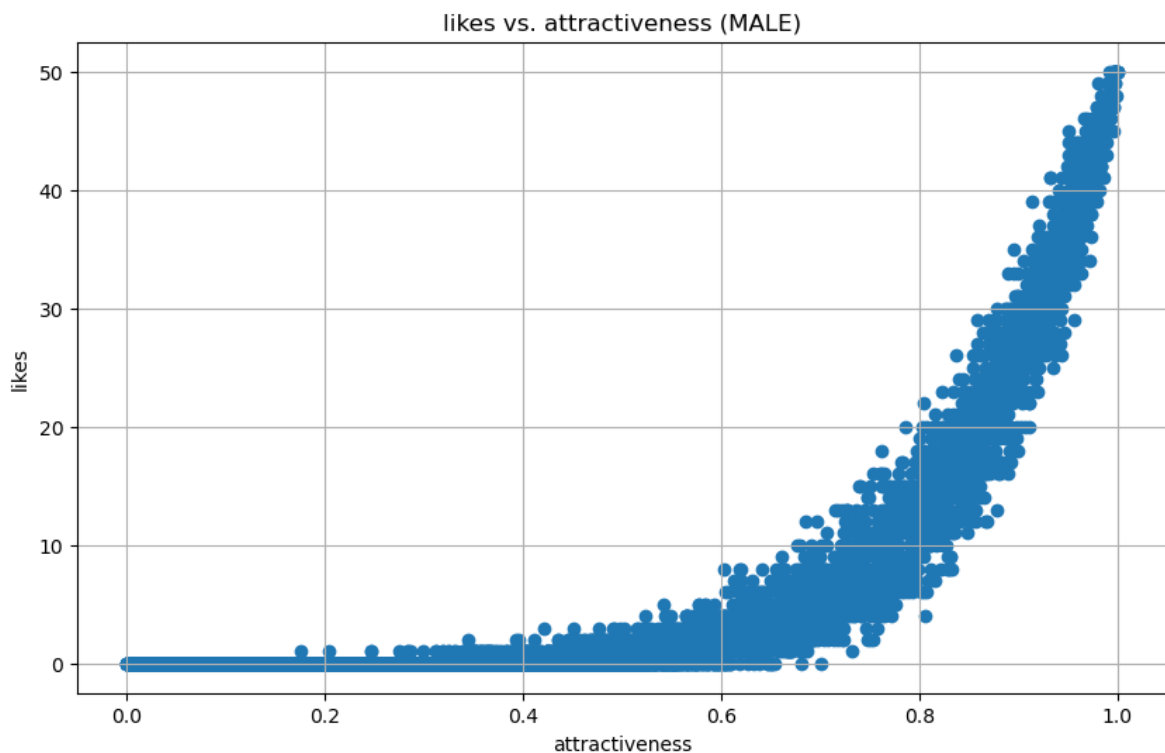
MALE

X:

attractiveness

Y:

likes



```
In [14]: def plot_1(group_column, y_column):
    data = user_data.groupby([group_column, pd.cut(user_data['attractiven
        [y_column].sum().unstack()
    data.plot(kind='bar', stacked=True, figsize=(10, 6))
    plt.xlabel(group_column)
    plt.ylabel(y_column + ' (Suma)')
    plt.title(f'{y_column} por {group_column} y Nivel de Atractivo')
    plt.grid(True)
    plt.legend(title='Nivel de Atractivo')
    plt.show()

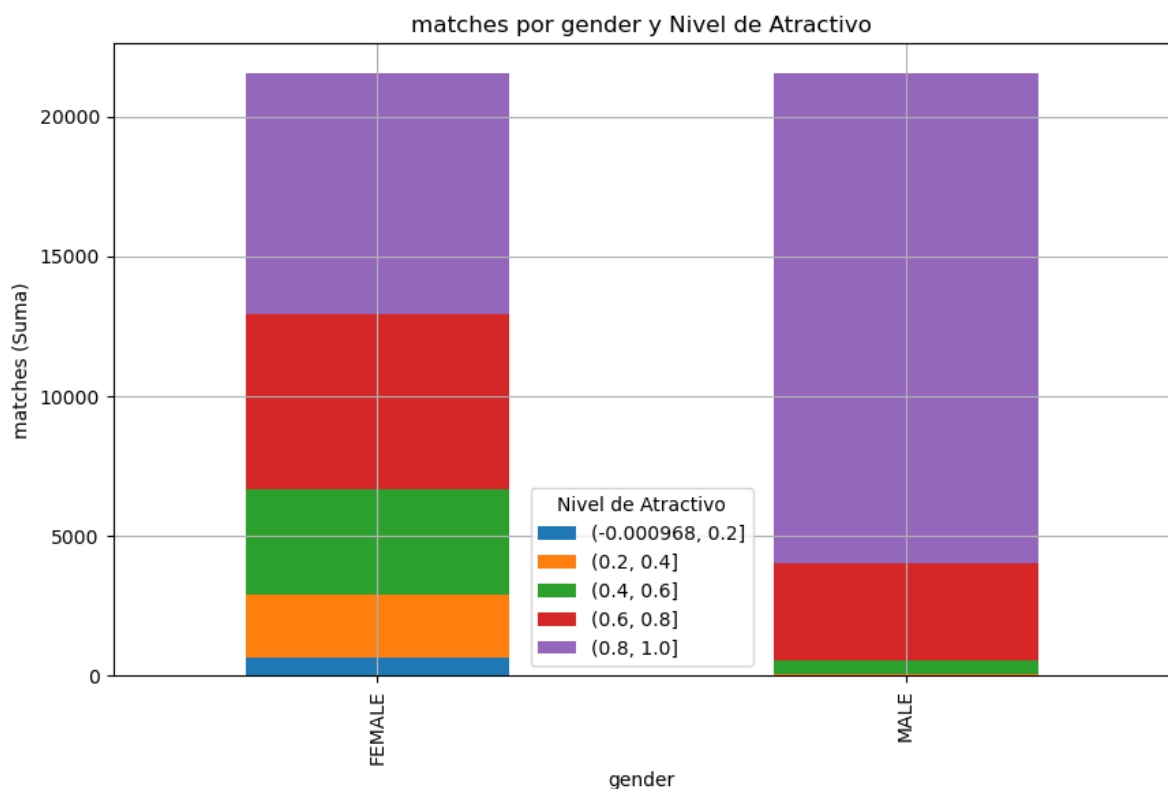
group_widget = widgets.Dropdown(options=['gender'], description='Agrupar por')
y_widget = widgets.Dropdown(options=['likes', 'matches'], description='Y:')

interactive_plot = interactive(plot_1, group_column=group_widget, y_column=y_widget)
output = interactive_plot.children[-1]

widgets.VBox([group_widget, y_widget, output])
```

Agrupar por:

Y:



Conclusión

En esta tarea, nuestro enfoque se centró en la exploración y aplicación de herramientas de visualización y widgets. A través de esta hemos obtenido un mayor entendimiento sobre cómo utilizar estas herramientas para analizar y presentar datos. Tuvimos la oportunidad de examinar dinámicamente los patrones y comportamientos generales de las personas en la aplicación de citas.

Referencias

Memeable Data. (2023, Julio 12). Why Men Get So Few Matches on Dating Apps. Recuperado de <https://memeabledata.simple.ink/why-men-get-so-few-matches-on-dating-apps-3e239cfee5a74c0f8746f39a9c1b00a7> (<https://memeabledata.simple.ink/why-men-get-so-few-matches-on-dating-apps-3e239cfee5a74c0f8746f39a9c1b00a7>).