Proyecto del Módulo 7: Técnicas avanzadas para ciencia de datos y empleabilidad

BASE DE DATOS: **UNDERSTANDING CAREER ASPIRATIONS OF GEN Z**

Adriana Acosta

Bootcamp: Ciencia de Datos e Inteligencia Artificial

Contenido

- Planteamiento del problema
- Análisis exploratorio de los Datos
- Preparación de los datos
- Aplicación del modelo y resultados
- Conclusiones

Comprender las aspiraciones de carrera profesional de la Generación Z

Contenido

- **Planteamiento del problema**
- Análisis exploratorio de los Datos
- Preparación de los datos
- Aplicación del modelo y resultados
- Conclusiones

Comprender las aspiraciones de carrera profesional de la Generación Z

Planteamiento del Problema

- Identificar grupos de jóvenes de la generación Z en la India con preferencias similares en cuanto a sus aspiraciones laborales y entornos de trabajo preferidos.

Planteamiento del Problema

Contexto y Variables de Interés:

- **Factores que contribuyen más en las aspiraciones laborales:** Esta variable incluyen factores como el salario, la influencia de diferentes personas en el estudiante como sus padres, personas que han cambiado el mundo, su circulo social, las oportunidades de crecimiento profesional, el impacto social de la empresa, entre otros.

- **Entorno de trabajo preferido:** Esta variable podría incluir preferencias sobre el tipo de ambiente de trabajo, como trabajar en un entorno colaborativo, trabajar de forma remota desde casa, trabajar en un entorno corporativo tradicional, etc.

- **Configuración de trabajo preferida:** Esta variable podría incluir preferencias sobre el tipo de configuración de trabajo, si se desea trabajar de manera individual, con un grupo pequeño o grande de personas, trabajar en una startup, una empresa establecida,etc.

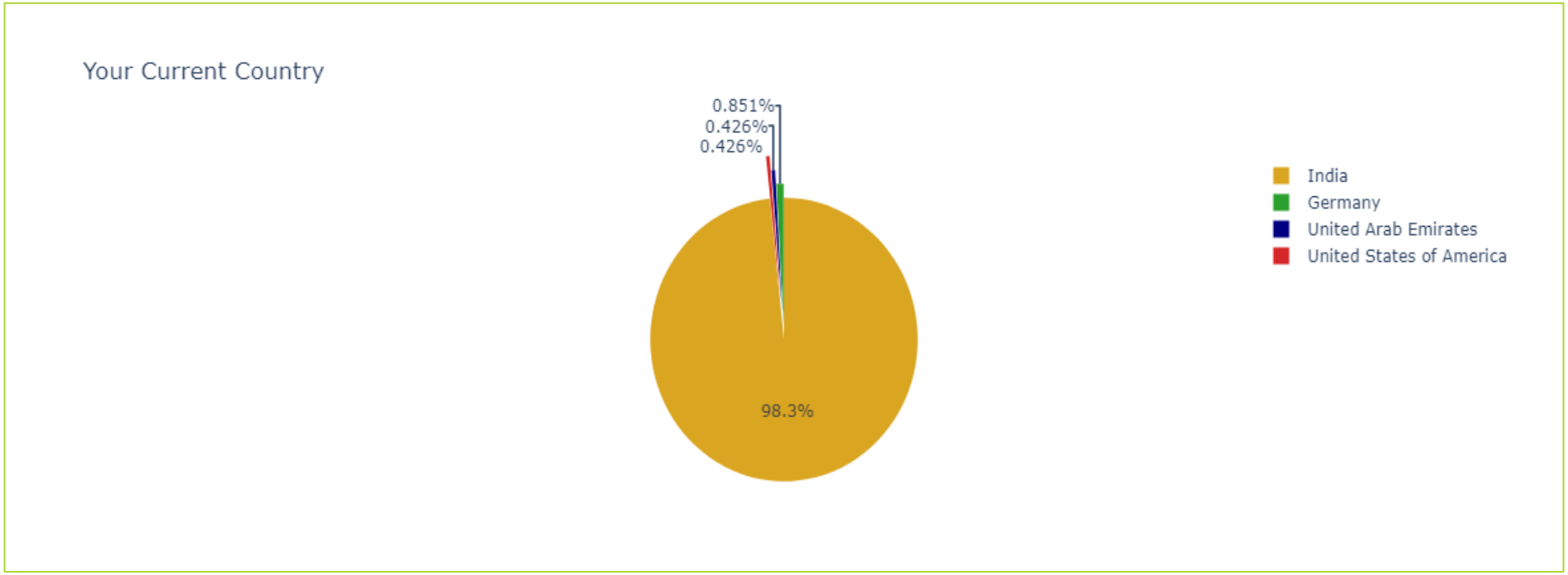
Contenido

- Planteamiento del problema
- **Análisis exploratorio de los Datos**
- Preparación de los datos
- Aplicación del modelo y resultados
- Conclusiones

Comprender las aspiraciones de carrera profesional de la Generación Z

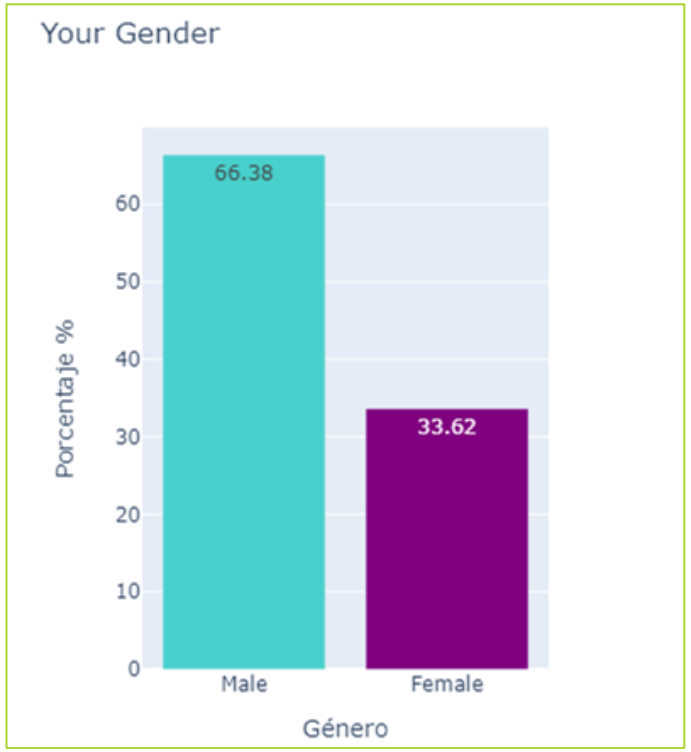
Análisis exploratorio y limpieza de los datos (EDA)

- Porcentaje de personas por país

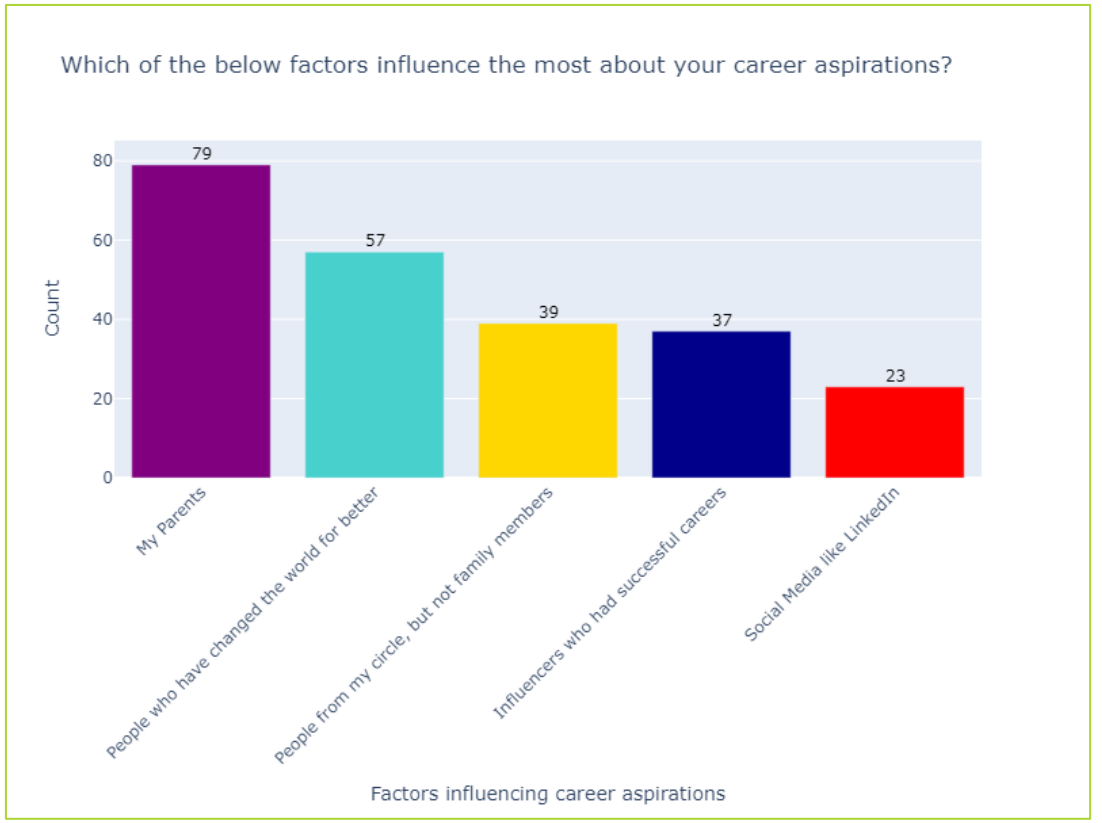


Análisis exploratorio y limpieza de los datos (EDA)

- Distribución por género

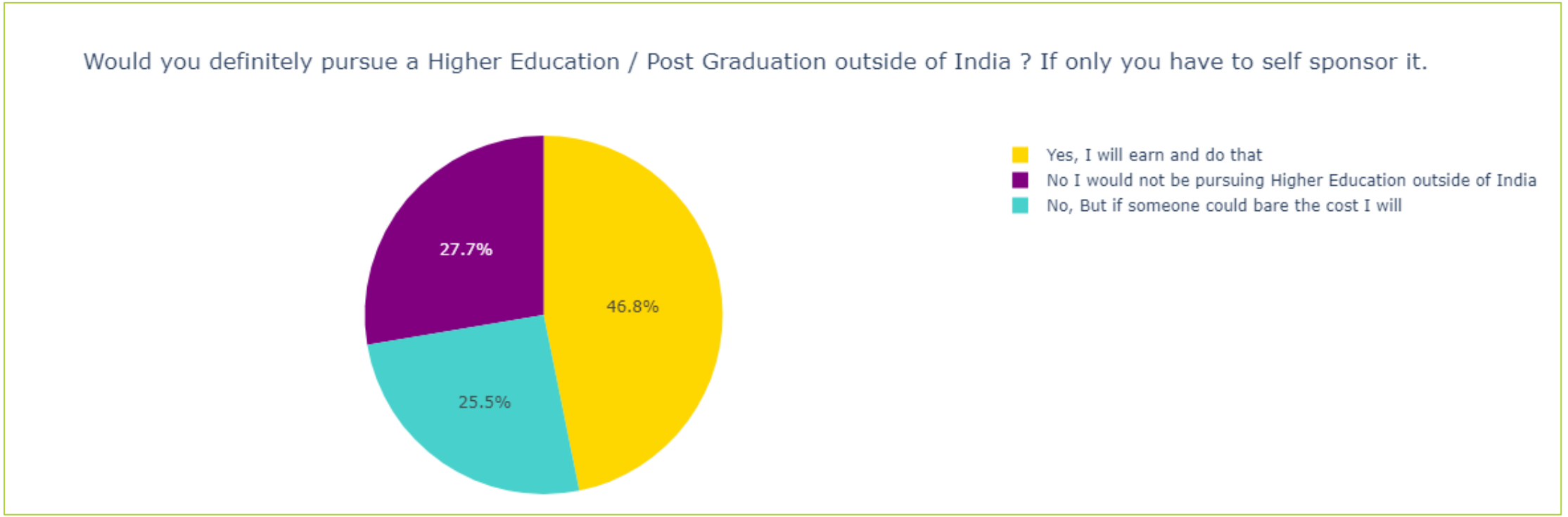


- Factores que influyen en la aspiración profesional



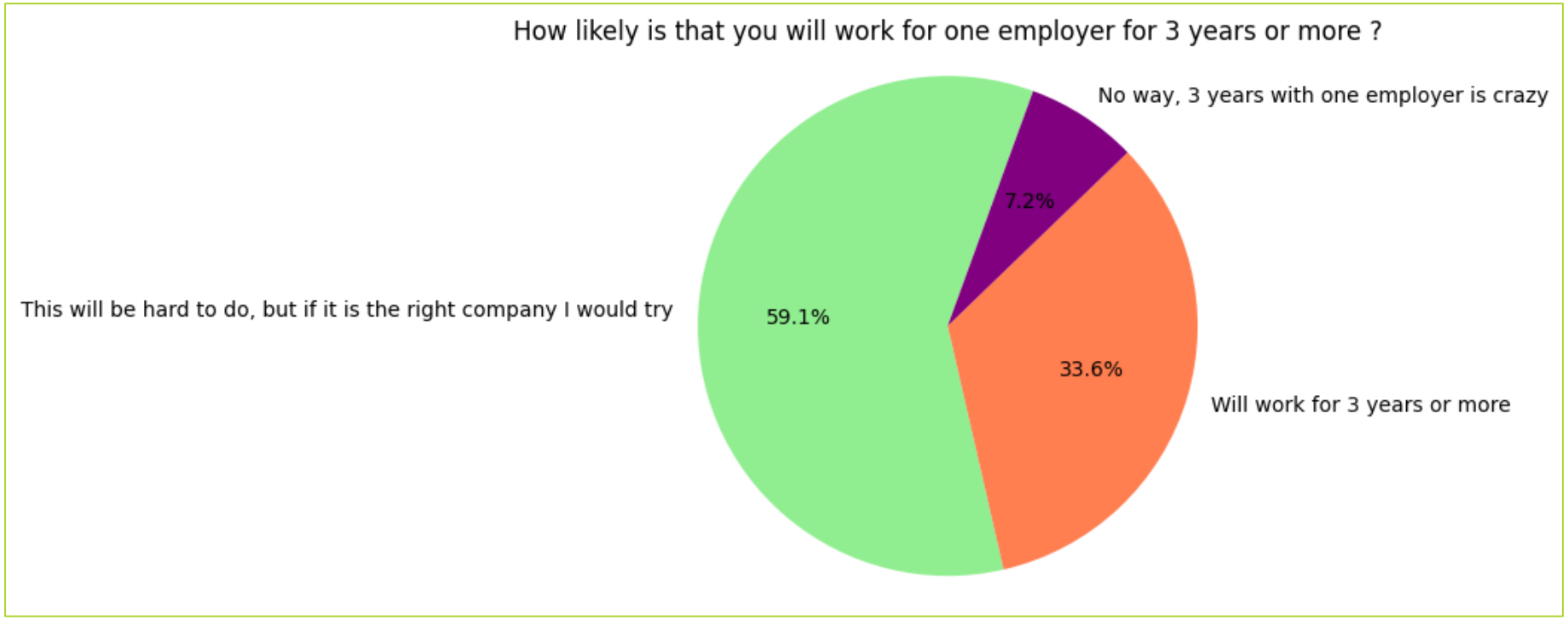
Análisis exploratorio y limpieza de los datos (EDA)

- Elección de educación superior o un posgrado fuera de la India si el estudiante se lo tuviera que costear



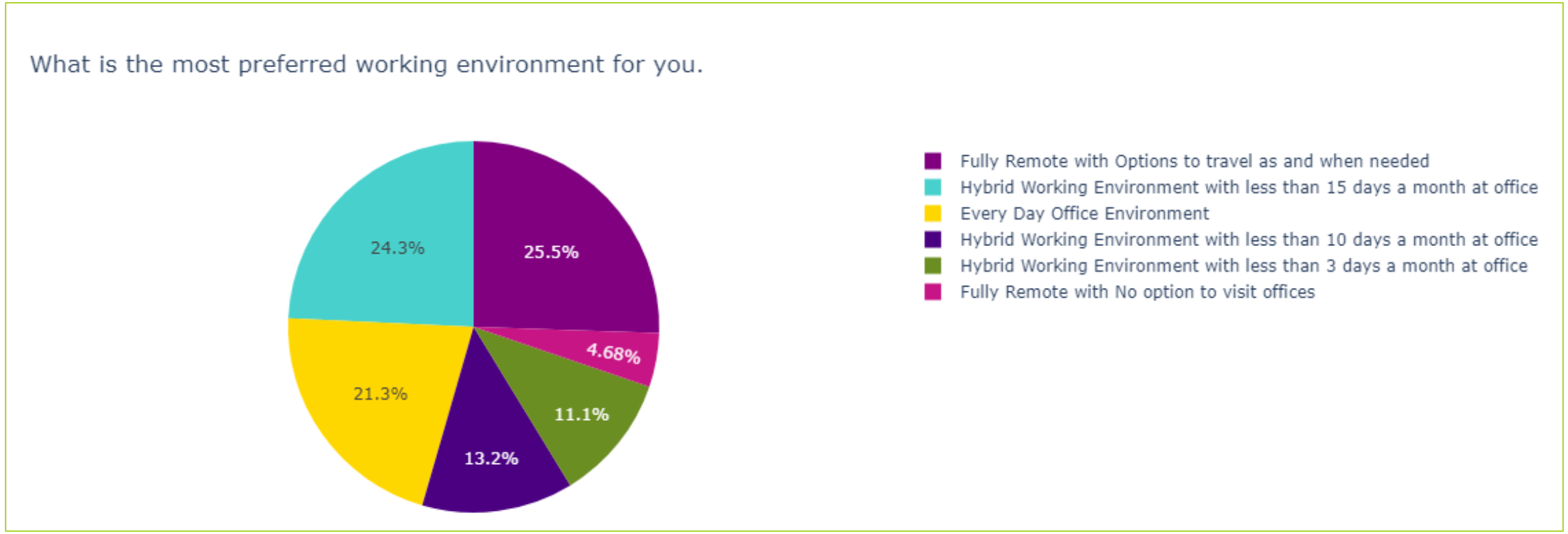
Análisis exploratorio y limpieza de los datos (EDA)

- Posibilidades de trabajar para un mismo empleador durante 3 o más años



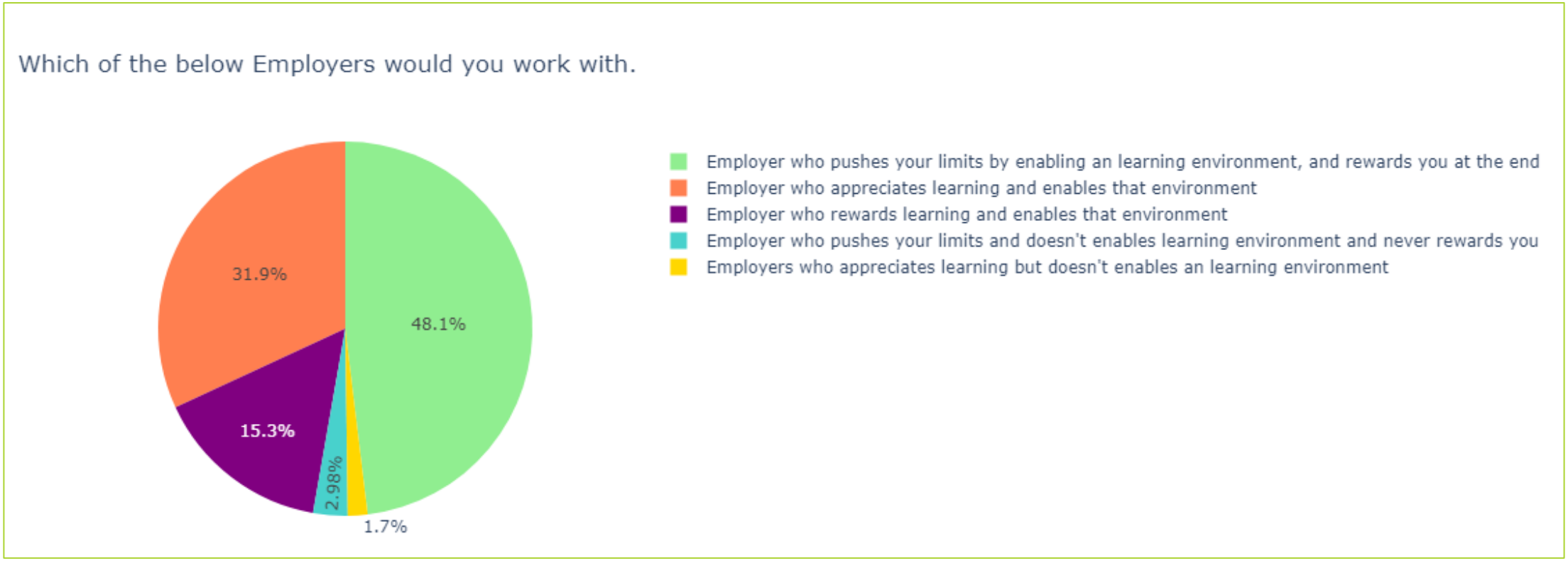
Análisis exploratorio y limpieza de los datos (EDA)

- Entorno de trabajo preferido de los estudiantes, remoto, tradicional o hibrido



Análisis exploratorio y limpieza de los datos (EDA)

- Tipo de empleadores con el cuál preferiría trabajar

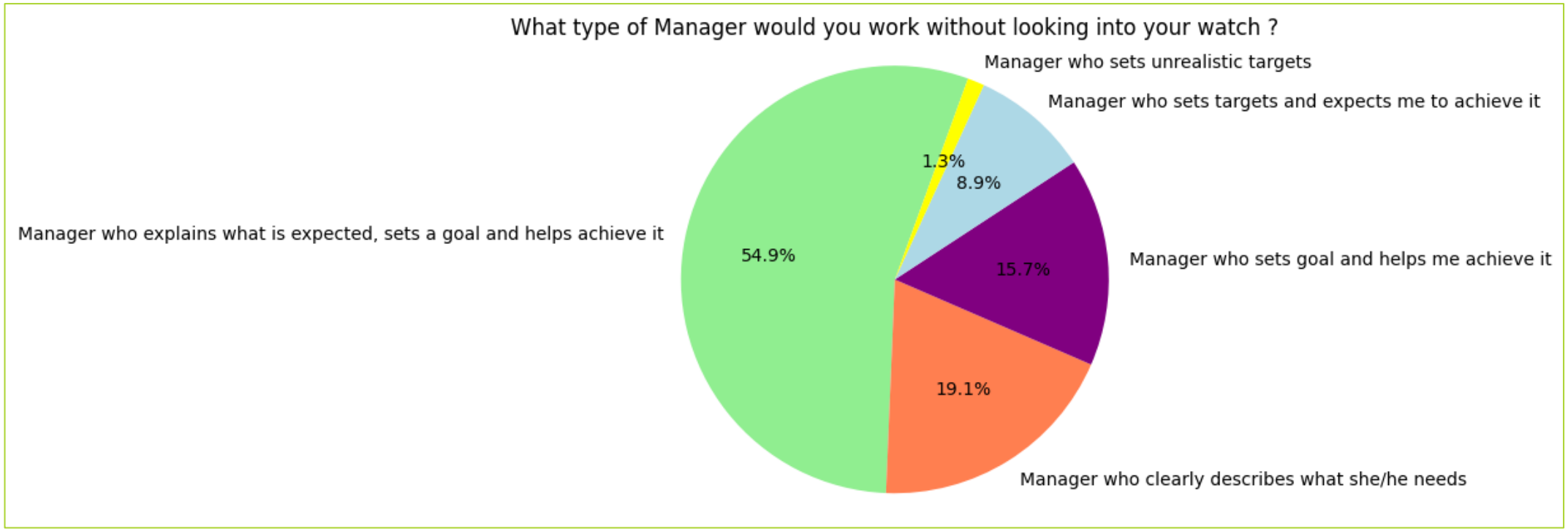


Análisis exploratorio y limpieza de los datos (EDA)

- Cantidad de personas con las que preferiría trabajar

Análisis exploratorio y limpieza de los datos (EDA)

- Tipo de gerente con el que podría trabajar sin mirar el reloj



Análisis exploratorio y limpieza de los datos (EDA)

- Elementos duplicados

```
Número de filas duplicadas: 0
```

- Completitud

	columna	total	completitud
0	Your Current Country.	0	100.0
1	Your Current Zip Code / Pin Code	0	100.0
2	Your Gender	0	100.0
3	Which of the below factors influence the most ...	0	100.0
4	Would you definitely pursue a Higher Education...	0	100.0
5	How likely is that you will work for one emplo...	0	100.0
6	Would you work for a company whose mission is ...	0	100.0
7	How likely would you work for a company whose ...	0	100.0
8	How likely would you work for a company whose ...	0	100.0
9	What is the most preferred working environment...	0	100.0
10	Which of the below Employers would you work with.	0	100.0
11	Which type of learning environment that you ar...	0	100.0
12	Which of the below careers looks close to your...	0	100.0
13	What type of Manager would you work without lo...	0	100.0
14	Which of the following setup you would like to...	0	100.0

Contenido

- Planteamiento del problema
- Análisis exploratorio de los Datos
- **Preparación de los datos**
- Aplicación del modelo y resultados
- Conclusiones

Comprender las aspiraciones de carrera profesional de la Generación Z

Preparación de los datos

- Variables de interés y conversión a variables categóricas

```
# Seleccionar las variables relevantes para el análisis de tendencias
variables_interes = ['Which of the below factors influence the most about your career aspirations ?',
                     'What is the most preferred working environment for you.',
                     'Which of the following setup you would like to work ?']

data_interes = df[variables_interes]

# Codificar variables categóricas utilizando one-hot encoding
data_encoded = pd.get_dummies(data_interes)
data_encoded
```

Preparación de los datos

- Escalamiento de datos y reducción de dimensionalidad

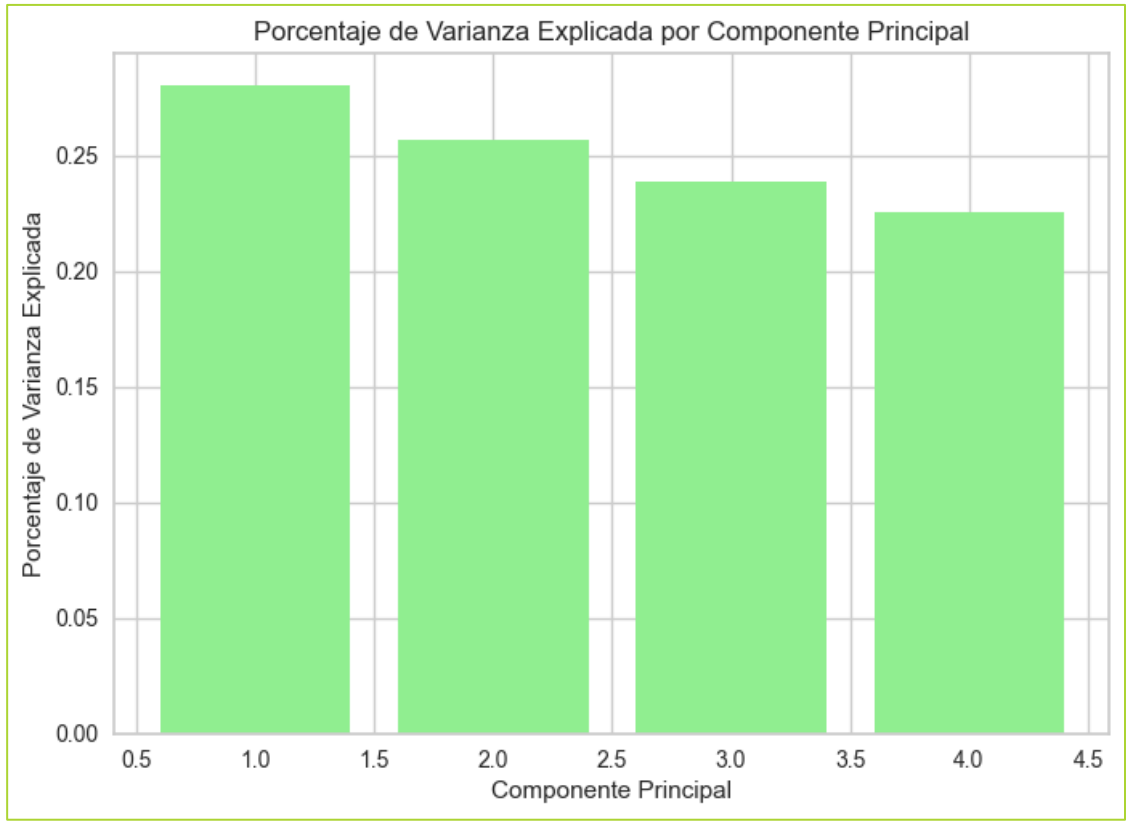
```
# Escalar los datos
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data_encoded)

# Aplicar PCA para reducir la dimensionalidad
pca = PCA(n_components=4)
data_pca = pca.fit_transform(data_scaled)
data_pca
```

PCA-1: 28.00%
PCA-2: 53.64%
PCA-3: 77.49%
PCA-4: 100.00%

Preparación de los datos

- Porcentaje de Varianza Explicada por componente principal

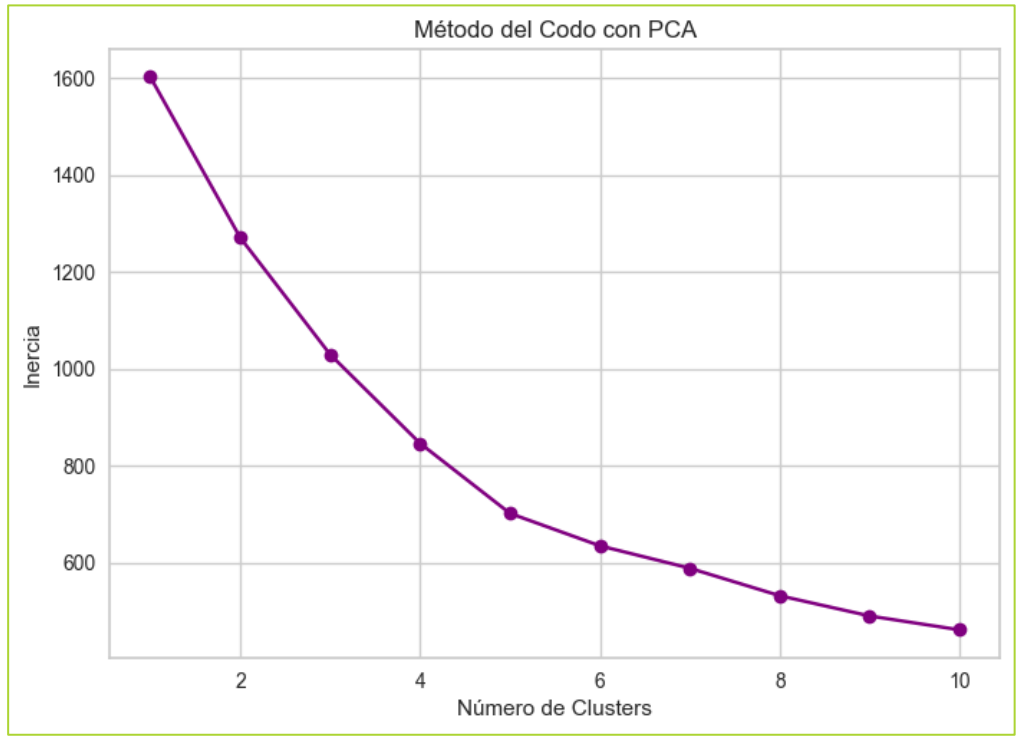


PCA-1: 28.00%
PCA-2: 53.64%
PCA-3: 77.49%
PCA-4: 100.00%

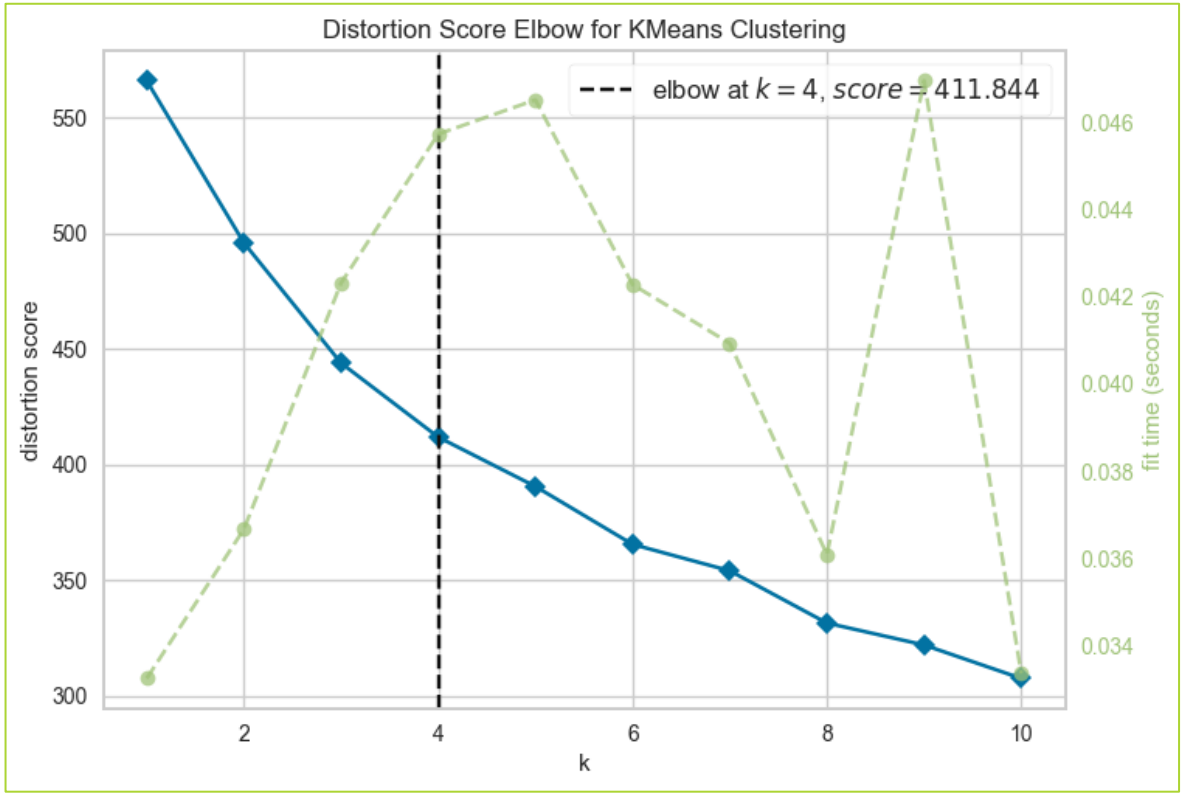
Preparación de los datos

- Método del codo con PCA,

para obtener cantidad de clústeres



- Ratificación de cantidad de clústeres usando **KElbowVisualizer**



Contenido

- Planteamiento del problema
- Análisis exploratorio de los Datos
- Preparación de los datos
- **Aplicación del modelo y resultados**
- Conclusiones

Comprender las aspiraciones de carrera profesional de la Generación Z

Aplicación del modelo y resultados

Solución Propuesta:

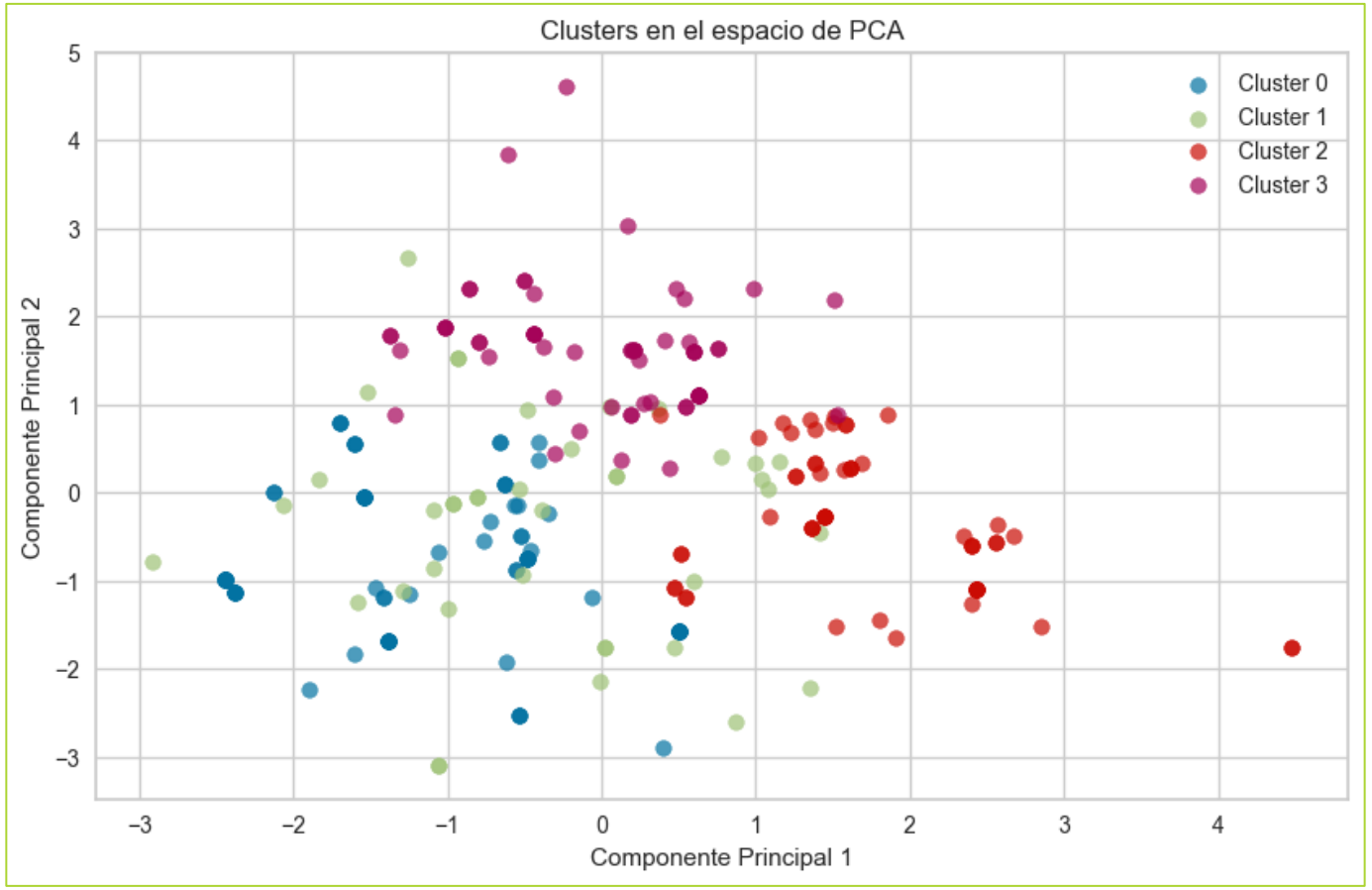
- Utilizando técnicas de clustering como K-means y DBSCAN, se podrían agrupar a los jóvenes de la generación Z en India en clústeres que compartan preferencias similares en cuanto a sus aspiraciones laborales y entornos de trabajo preferidos.

- Permitiendo identificar grupos de personas con características y preferencias similares, lo que podría ser útil para diseñar estrategias de reclutamiento, planificación de recursos humanos y desarrollo de carreras adaptadas a las necesidades y preferencias específicas de cada grupo.

Aplicación del modelo y resultados

- Aplicar K-means con el numero optimo de Clústeres

- Coeficiente de silueta para K-means: **0.258605061235615 4**
- Coeficiente de silueta para DBSCAN: **0.037078661876398085**



Contenido

- Planteamiento del problema
- Análisis exploratorio de los Datos
- Preparación de los datos
- Aplicación del modelo y resultados
- **Conclusiones**

Comprender las aspiraciones de carrera profesional de la Generación Z

Conclusiones

- **Coeficiente de silueta para K-means:** 0.258605061235615 4 . Este valor indica que los clústeres generados por K-means tienen una separación relativamente aceptable entre ellos. Sin embargo, aún puede haber cierta superposición o se requiere hacer una mejor clasificación.

- **Coeficiente de silueta para DBSCAN:** 0.037078661876398085. Este valor indica que los clústeres generados por DBSCAN tienen una separación bastante baja entre ellos. Esto podría sugerir que DBSCAN no ha sido tan efectivo para separar claramente los datos en clústeres distintos, o que los datos son intrínsecamente difíciles de agrupar con este algoritmo.

El coeficiente de silueta para K-means es significativamente mayor que para DBSCAN, lo que podría sugerir que K-means ha generado clústeres con una mejor separación en comparación con DBSCAN.


¡Gracias!