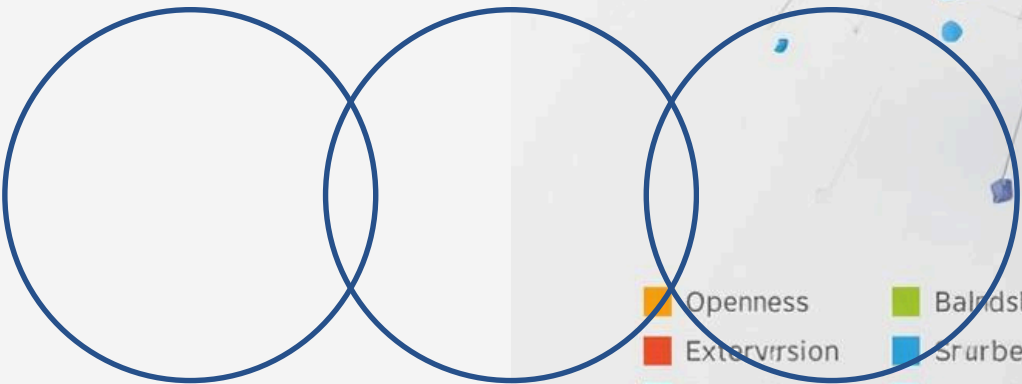


Clustering con K-Means

Guzman Alexander Quintero Osorio



Simaric's Clustering Analysis

Personalitus clustering based on the OCEAN model, based, the Calleanpes agee enalling ons.

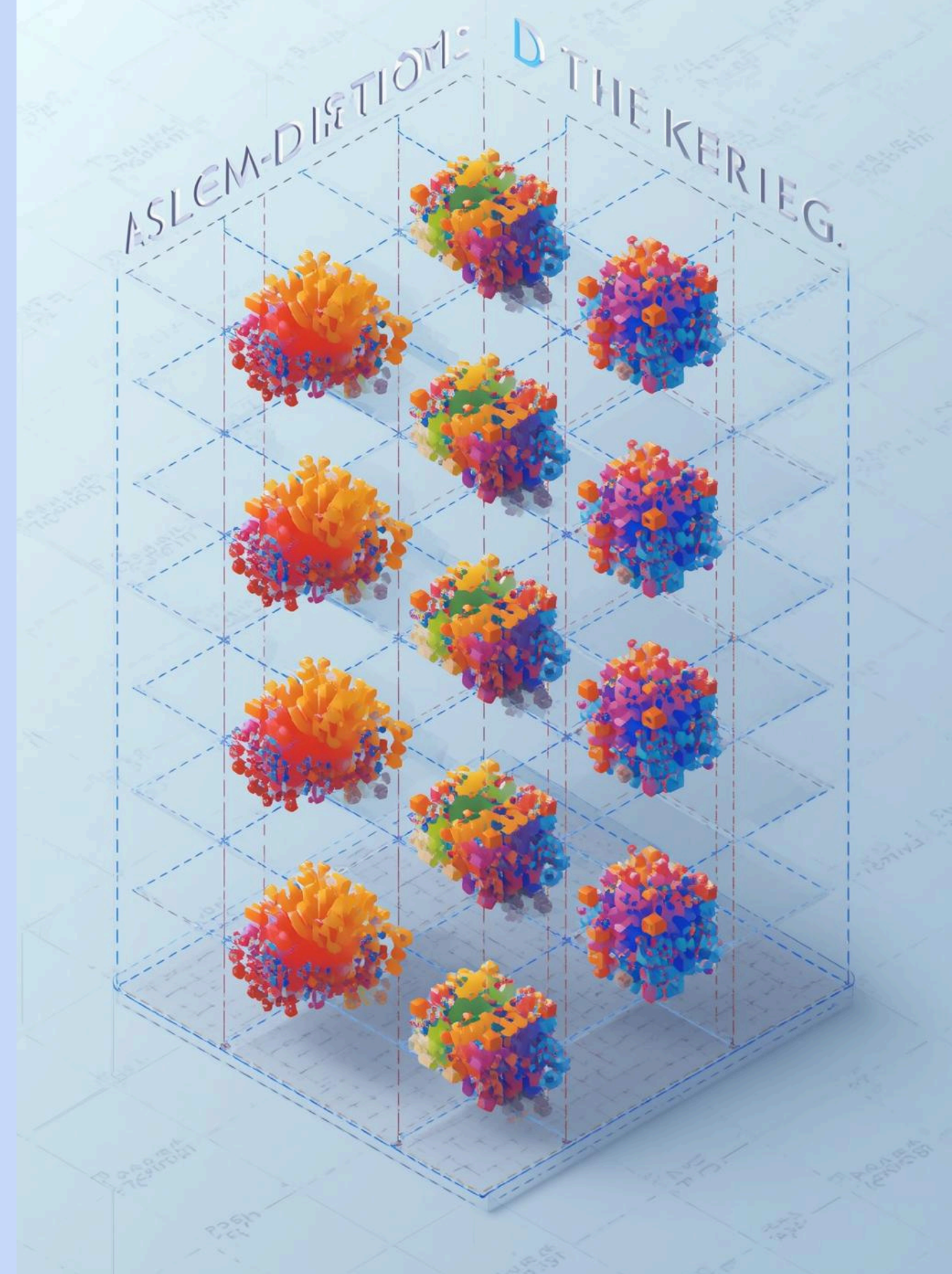


- | | | | | |
|-------------|-------------|----------|----------|----------|
| Openness | Balndsfulac | Call | Git% | Goclubes |
| Extervrsion | Srurbel | Cultotic | Redilunr | Liomges |
| Extivelocel | Gcppls | Mosoey | Fajo | Sriner |

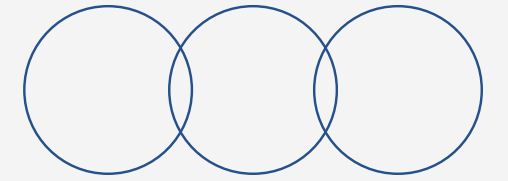
Metodología de Clustering

Uso de K-Means para el análisis

El algoritmo K-Means es una técnica no supervisada que permite identificar grupos naturales en los datos. Utilizamos perfiles de figuras públicas con rasgos de personalidad para determinar estas agrupaciones.



Resultados del Clustering



Evaluación de K=5

Cluster 1

El **Cluster 1** incluye figuras públicas con altos niveles de Openness y Extraversión, caracterizándose por su creatividad y sociabilidad. Estos individuos suelen destacar en campos artísticos y de entretenimiento.

Cluster 2

El **Cluster 2** agrupa a personalidades más reservadas, con un enfoque en Agreeableness. Tienden a ser empáticos y solidarios, encontrándose comúnmente en roles de servicio y cuidado comunitario.

Cluster 3

En **Cluster 3**, se observan figuras públicas con una mezcla equilibrada de rasgos OCEAN. Estos individuos son adaptables y versátiles, destacándose en diversas áreas profesionales y de liderazgo.

COMPARACION ENTRE K=4 Y K=5

```
Mapeo de Cluster a Categoría (K=4):  
{np.int32(0): np.int64(4), np.int32(1): np.int64(2), np.int32(2): np.int64(8), np.int32(3): np.int64(2)}  
  
--- Métricas de Clasificación (K=4, después del mapeo) ---  
Exactitud (Accuracy): 0.3429  
Precisión (Precision): 0.1717  
Recall: 0.3429  
F1-Score: 0.2265
```

Exactitud(accuracy)

Al comparar la '**Exactitud**' de ambos escenarios, observamos que no hay diferencia en este valor: ambos modelos, con K=4 y K=5, obtienen una Exactitud de 0.3429

```
Mapeo de Cluster a Categoría (K=5):  
{np.int32(0): np.int64(8), np.int32(1): np.int64(2), np.int32(2): np.int64(2), np.int32(3): np.int64(4), np.int32(4): np.int64(7)}  
  
--- Métricas de Clasificación (K=5, después del mapeo) ---  
Exactitud (Accuracy): 0.3429  
Precisión (Precision): 0.2561  
Recall: 0.3429  
F1-Score: 0.2493
```

Precisión

Observamos una mejora notable en la **Precisión** con K=5 (0.2561) en comparación con K=4 (0.1717). Esto significa que cuando el modelo con K=5 'predice' una categoría (a través del mapeo del cluster dominante), es más confiable en que esa predicción sea correcta que el modelo con K=4

RECALL(SENcIBILIDAD)

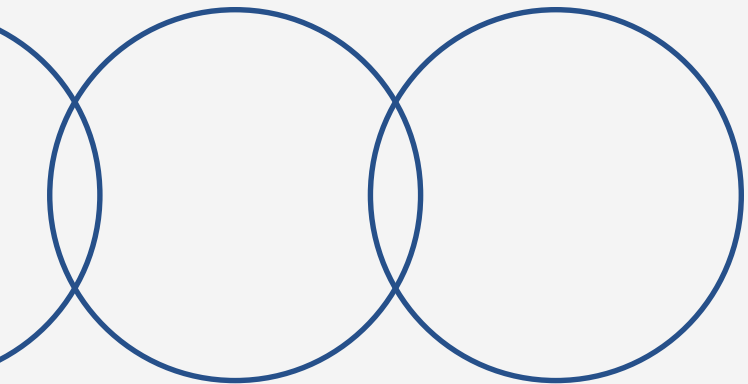
El Recall se mantuvo idéntico en 0.3429 para ambos valores de K. Esto indica que la capacidad del modelo para identificar correctamente todos los puntos de una categoría determinada fue la misma para K=4 y K=5, después de realizar el mapeo.

F1-SCORE

El F1-Score para K=5 fue de 0.2493, mientras que para K=4 fue de 0.2265. Esto indica una ligera mejora en el F1-Score para K=5.

Conclusiones y Futuras Direcciones

Reflexiones sobre el análisis y los
próximos pasos



Más dimensiones

Incluir rasgos como **Conscientiousness** y **Neuroticism** podría ofrecer un análisis más completo de las figuras públicas.

Nuevas técnicas

Explorar métodos alternativos de clustering, como **DBSCAN** o jerárquico, podría revelar patrones adicionales en los datos.

Distintividad laboral

Evaluar cómo las categorías laborales son percibidas desde la **perspectiva de personalidad** puede enriquecer el entendimiento del análisis.

Gracias por su atención

