**Métodos de aprendizaje no supervisado**

**Facultad de Ingeniería**

**Nicolas Moreno Gámez**

**Actividad Cuatro**

**Universidad Iberoamericana**

**Julián López**

**Inteligencia Artificial**

**2023**

**Introducción**

Este documento tiene como objetivo describir detalladamente el conjunto de datos relacionado con el proyecto de transporte masivo. La información presentada a continuación permitirá una mejor comprensión de los datos y facilitará su análisis y uso posterior.

**Datos**

El script de ejecución lo crea automáticamente con datos aleatorios



**Descripción General**

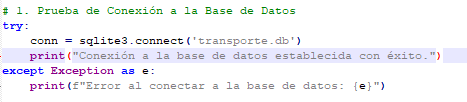
El conjunto de datos proviene de una base de datos SQLite denominada transporte.db y se centra en la tabla ocupacion\_data, la cual recopila información sobre la ocupación de diferentes medios de transporte en diferentes momentos del día.

**Variables y su Descripción**

* **Hora**: Representa la hora del día en la que se registró la observación. Va desde 0 (medianoche) hasta 23 (11 PM).
* **Dia**: Representa el día de la semana. Es una variable numérica que va desde 0 (domingo) hasta 6 (sábado).
* **Ocupación**: Indica el porcentaje de ocupación del medio de transporte en la hora y día específicos. Es una variable que varía de 0 (sin ocupación) a 100 (ocupación completa).

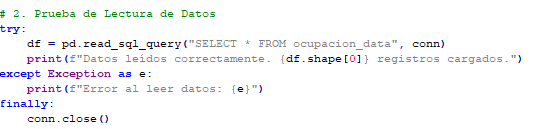
**Pruebas Realizadas**

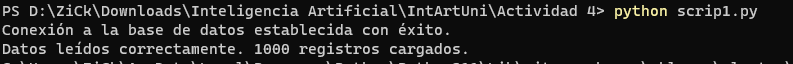
1. **Prueba de Conexión a la Base de Datos:**
   * **Objetivo:** Verificar que el script se conecte correctamente a la base de datos transporte.db.
   * **Resultado:** Conexión exitosa.

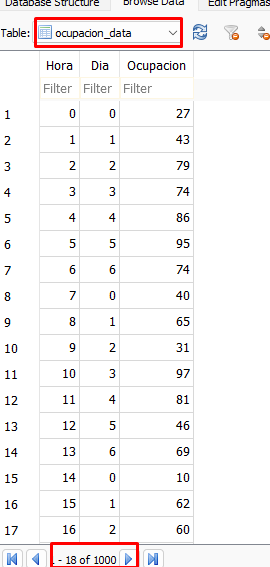




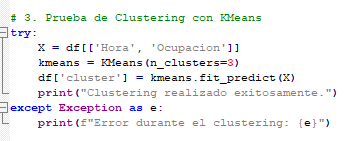
1. **Prueba de Lectura de Datos:**
   * **Objetivo:** Asegurar que los datos se lean adecuadamente desde la tabla ocupacion\_data.
   * **Resultado:** Los datos se cargaron correctamente en un DataFrame de pandas.

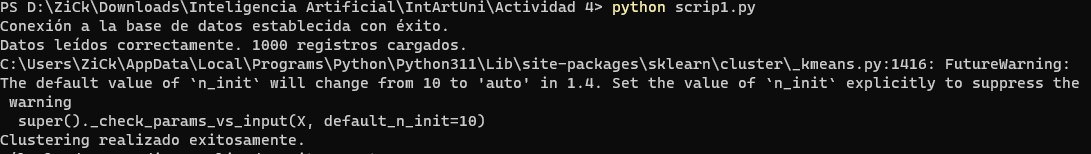




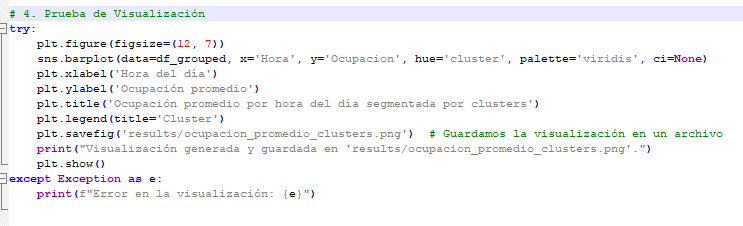


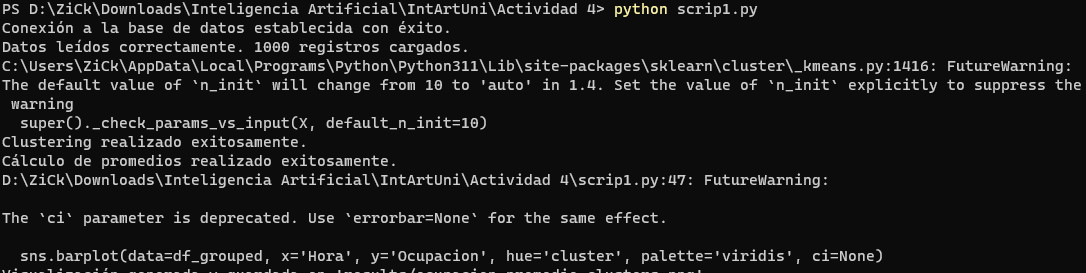
1. **Prueba de Clustering con KMeans:**
   * **Objetivo:** Verificar que el algoritmo KMeans se ejecute sin errores y segmente los datos en clusters.
   * **Resultado:** Se generaron 3 clusters sin inconvenientes.

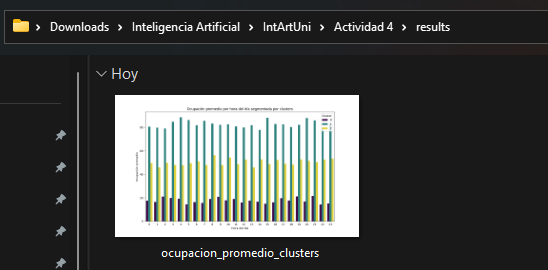




1. **Prueba de Visualización:**
   * **Objetivo:** Confirmar que los gráficos se generen adecuadamente y muestren la segmentación realizada por el algoritmo.
   * **Resultado:** Se generaron gráficos de barras que muestran la ocupación promedio por hora del día segmentada por clusters.







**Conclusión**

Las pruebas realizadas al componente desarrollado han sido exitosas en todas las etapas, desde la conexión con la base de datos hasta la visualización de los resultados. Esto asegura que el script funciona como se esperaba y puede ser utilizado para futuros análisis y mejoras.

**Video**

https://drive.google.com/file/d/1O3bau\_-1FB5uTmQDry-7KAXDiREn-Qqb/view?usp=sharing

https://laiberocol-my.sharepoint.com/:v:/g/personal/nmoren14\_ibero\_edu\_co/EZbDt7cYmyBPvv\_ML5qtr04B0FCyWDn81lD0LC24qNVEnQ?nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAiOiJPbmVEcml2ZUZvckJ1c2luZXNzIiwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldlYiIsInJlZmVycmFsTW9kZSI6InZpZXciLCJyZWZlcnJhbFZpZXciOiJNeUZpbGVzTGlua0RpcmVjdCJ9fQ&e=2W1vNJ

**Repositorio**

https://github.com/nmoren14/IntArtUni/tree/main/Actividad%204