

**Métodos de aprendizaje no supervisado**

**Facultad de Ingeniería**

**Nicolas Moreno Gámez**

**Actividad Cuatro**

**Universidad Iberoamericana**

**Julián López**

**Inteligencia Artificial**

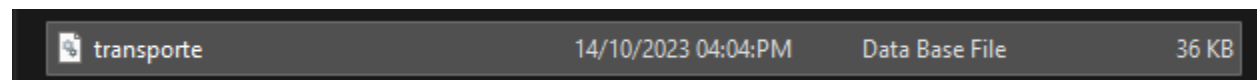
**2023**

### 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).
- **Día:** 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 Conexión a la Base de Datos
try:
    conn = sqlite3.connect('transporte.db')
    print("Conexión a la base de datos establecida con éxito.")
except Exception as e:
    print(f"Error al conectar a la base de datos: {e}")
```

```
PS D:\ZiCk\Downloads\Inteligencia Artificial\IntArtUni\Actividad 4> python scrip1.py
Conexión a la base de datos establecida con éxito.
```

### 2. 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.

```
# 2. Prueba de Lectura de Datos
try:
    df = pd.read_sql_query("SELECT * FROM ocupacion_data", conn)
    print(f"Datos leídos correctamente. {df.shape[0]} registros cargados.")
except Exception as e:
    print(f"Error al leer datos: {e}")
finally:
    conn.close()
```

```
PS D:\ZiCk\Downloads\Inteligencia Artificial\IntArtUni\Actividad 4> python scrip1.py
Conexión a la base de datos establecida con éxito.
Datos leídos correctamente. 1000 registros cargados.
```

Database Structure   Browse Data   Edit Pragmas

Table: **ocupacion\_data**

	Hora	Dia	Ocupacion
	Filter	Filter	Filter
1	0	0	27
2	1	1	43
3	2	2	79
4	3	3	74
5	4	4	86
6	5	5	95
7	6	6	74
8	7	0	40
9	8	1	65
10	9	2	31
11	10	3	97
12	11	4	81
13	12	5	46
14	13	6	69
15	14	0	10
16	15	1	62
17	16	2	60

17 of 1000

### 3. 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.

```
# 3. Prueba de Clustering con KMeans
try:
    X = df[['Hora', 'Ocupacion']]
    kmeans = KMeans(n_clusters=3)
    df['cluster'] = kmeans.fit_predict(X)
    print("Clustering realizado exitosamente.")
except Exception as e:
    print(f"Error durante el clustering: {e}")
```

```
PS D:\ZiCk\Downloads\Inteligencia Artificial\IntArtUni\Actividad 4> python scrip1.py
Conexión a la base de datos establecida con éxito.
Datos leídos correctamente. 1000 registros cargados.
C:\Users\ZiCk\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning:
The default value of 'n_init' will change from 10 to 'auto' in 1.4. Set the value of 'n_init' explicitly to suppress the
warning
  super()._check_params_vs_input(X, default_n_init=10)
Clustering realizado exitosamente.
```

#### 4. 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.

```
# 4. Prueba de Visualización
try:
    plt.figure(figsize=(12, 7))
    sns.barplot(data=df_grouped, x='Hora', y='Ocupacion', hue='cluster', palette='viridis', ci=None)
    plt.xlabel('Hora del día')
    plt.ylabel('Ocupación promedio')
    plt.title('Ocupación promedio por hora del día segmentada por clusters')
    plt.legend(title='Cluster')
    plt.savefig('results/ocupacion_promedio_clusters.png') # Guardamos la visualización en un archivo
    print("Visualización generada y guardada en 'results/ocupacion_promedio_clusters.png'.")
    plt.show()
except Exception as e:
    print(f"Error en la visualización: {e}")
```

```
PS D:\Zick\Downloads\Inteligencia Artificial\IntArtUni\Actividad 4> python scrip1.py
Conexión a la base de datos establecida con éxito.
Datos leídos correctamente. 1000 registros cargados.
C:\Users\Zick\AppData\Local\Programs\Python\Python311\Lib\site-packages\sklearn\cluster\_kmeans.py:1416: FutureWarning:
The default value of 'n_init' will change from 10 to 'auto' in 1.4. Set the value of 'n_init' explicitly to suppress the
warning
  super()._check_params_vs_input(X, default_n_init=10)
Clustering realizado exitosamente.
Cálculo de promedios realizado exitosamente.
D:\Zick\Downloads\Inteligencia Artificial\IntArtUni\Actividad 4\scrip1.py:47: FutureWarning:
The 'ci' parameter is deprecated. Use 'errorbar=None' for the same effect.
  sns.barplot(data=df_grouped, x='Hora', y='Ocupacion', hue='cluster', palette='viridis', ci=None)
```



### 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://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\\_ML5qtr04B0FCyWDn81ID0LC24qNVEEnQ?nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAIoiJPbmVEcmI2ZUZvckJ1c2luZXNzliwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldiYilsluZlZmVycmFsTW9kZSI6InZpZXciLCJyZWZlcnJhbFZpZXciOiJNeUZpbGVzTGlua0RpcmVjdCJ9fQ&e=2W1vNJ](https://laiberocol-my.sharepoint.com/:v/g/personal/nmoren14_ibero_edu_co/EZbDt7cYmyBPvv_ML5qtr04B0FCyWDn81ID0LC24qNVEEnQ?nav=eyJyZWZlcnJhbEluZm8iOnsicmVmZXJyYWxBcHAIoiJPbmVEcmI2ZUZvckJ1c2luZXNzliwicmVmZXJyYWxBcHBQbGF0Zm9ybSI6IldiYilsluZlZmVycmFsTW9kZSI6InZpZXciLCJyZWZlcnJhbFZpZXciOiJNeUZpbGVzTGlua0RpcmVjdCJ9fQ&e=2W1vNJ)

### Repositorio

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