

Tarea 2

Introducción

La Municipalidad de Arica vendió alrededor de 2300 permisos de circulación en enero de 2024. Para vender más permisos de circulación en enero de 2025, quieren saber si hay grupos de autos (clientes) específicos a los que puedan dirigirse en una futura campaña publicitaria.

Luego de entregar el primer informe (Tarea 1), donde realizó los grupos de autos con k-means, usted cree que quizás se pueden hacer mejores grupos de autos usando otros algoritmos de aglomeración. Para comprobar su hipótesis, compara los valores silueta de tres algoritmos de aglomeración: K-Means, Jerárquico y DBSCAN.

El conjunto de datos de ventas de permisos de circulación de enero 2024 de Arica se encuentra en el siguiente link:

<https://datos.gob.cl/en/dataset/permisos-de-circulacion-enero-2024>

Entregables

- La Tarea se realiza en los mismos grupos que la Tarea 1.
- En webcursos, usted debe subir el código Python utilizado, en formato **Jupyter Notebook (ipynb)**.
- El nombre del Archivo debe ser **Tarea2_iniciales1_iniciales2.ipynb**
 - Ejemplo, el Notebook de Alfonso Tobar y Sofía Álvarez se debe entregar como **Tarea2_AT_SA.ipynb**.

Código

El código Python escrito en Jupyter Notebook debe incluir:

Parte 1: Preparativos (1.5 punto)

- Abrir el Excel (.xlsx). Al momento de la revisión los Notebooks se revisan utilizando el Excel original descargado por lo que el archivo no debe ser modificado manualmente por usted.
- Seleccionar las columnas: Año de fabricación, Monto pagado, Tipo vehículo.
- Normalizar las columnas Año de fabricación y Monto pagado (Puede utilizar StandardScaler o MinMaxScaler). **Este proceso se debe realizar sólo en Scikit-Learn.**
- Transformar Tipo de vehículo a solo 3 categorías: moto, auto pequeño, auto grande.

- Luego transformar Tipo de vehículo de categórico a **One Hot Encoder sólo utilizando Scikit-Learn**.

Parte 2: K-Means (0.5 punto)

- Hacer clusters usando el algoritmo k-means. Usted elija el número de clusters (k) que encuentre correcto utilizando la curva del codo.
- Calcular el valor silueta del algoritmo k-means.

Parte 3: Cluster Jerarquico (1 punto)

- Hacer clusters usando el algoritmo jerárquico. Usted elija el número de clusters (k) y el linkage que encuentre más apropiado mediante el uso de un Dendograma.
- Calcular el valor silueta del algoritmo jerárquico.

Parte 3: DBSCAN (1 punto)

- Hacer clusters usando el algoritmo DBSCAN. Usted elija el **eps** y **min_samples** que encuentre correcto utilizando la curva del codo y las sugerencias vistas en clases.
- Calcular el valor silueta del algoritmo DBSCAN.

Parte 4: Evaluación y Análisis: (2 puntos)

- Crear un gráfico de barras con los valores silueta de los tres algoritmos.
- Basado en el gráfico, explicar cuál algoritmo da los mejores resultados.
- Etiquetar y describir los clusters encontrados sólo del mejor algoritmo.

El Jupyter Notebook debe explicar:

- Cómo se aplicaron los pre-procesamientos y el uso general de cada trozo de código.
- Cómo se eligió el número de clusters (**k**), **eps** y **min_samples**.
- Cómo se etiquetó y describió en detalle los clusters encontrados del **mejor algoritmo**.

Si se entrega un código sin su explicación, entonces dicha parte obtendrá ponderación 0.

Fecha de Entrega:

- **Domingo 28 de Abril 23:59 hrs.**
- Los atrasos son penalizados con 1 punto por hora o fracción de retraso.