**Trabajo Práctico Módulo 3**

**Aprendizaje no supervisado**

**Segmentación de clientes**

Descripción:

La segmentación de clientes es un proceso a través del cual se clasifica a los mismos en diferentes grupos que comparten características similares de formas tales que los resultados sean relevantes para campañas de marketing. Las características utilizadas para describir a cada cliente pueden ser edad, sexo, intereses, hábitos de consumo, entre otros.

Para este proyecto, se provee de un conjunto de datos con las siguientes características:

* Origen: Exposys Data Labs
* Columnas:
  + **CustomerID**: identificador único del cliente
  + **Gender**: género del cliente
  + **Age**: edad del cliente
  + **Annual Income (k$)**: ingresos mensuales del cliente
  + **Spending Score (1-100)**: puntaje relacionado al nivel de gastos del cliente

Consigna:

Los objetivos principales del proyecto son dos:

1. identificar la base de clientes con mayor potencial para la venta de producto/s
2. implementar algoritmos de clustering para segmentar la base de clientes

El trabajo debe incluir:

1. Análisis exploratorio de datos
   1. Descripción de las variables
   2. Verificar si hay valores faltantes y outliers
   3. Visualizaciones necesarias para comprender los datos
   4. Análisis de Distribución y Análisis de correlación de las variables
2. Modelización
   1. Seleccionar las variables para la posterior clusterización
   2. Evaluar y efectuar las transformaciones de variables necesarias (ej: estandarización, normalización, etc)
   3. Aplicar K-Means, DBSCAN y clustering jerárquico (incluir dendograma)
   4. Analizar la factibilidad de aplicar reducción de dimensionalidad. Probar el efecto en los algoritmos mencionados
   5. Describir los clusters obtenidos y ponerles nombres/etiquetas significativas
   6. Reportar los resultados de ambos modelos finales, incluyendo justificación de los hiperparámetros escogidos y del por qué de las variables utilizadas
3. Se debe entregar:
4. Jupyter notebook o un archivo .py con el código organizado. Deben incluirse comentarios y breve interpretaciones de los resultados.
5. Una presentación de máximo 15 diapositivas explicando los hallazgos obtenidos (no más de 5 minutos).

El trabajo se puede hacer en grupos de hasta 4 personas.

**Fecha de Presentación del TP:**

**Cohorte 1: Jueves 19/08.**

**Cohorte 2: Viernes 27/08.**