

Proyecto - aprendizaje no supervisado

El objetivo del proyecto final consiste en permitirle a los estudiantes poner en práctica los conocimientos teóricos y prácticos adquiridos durante la clase. El proyecto se realizará de forma grupal y se desarrollará con datos de ventas una tienda en línea.

Toda la información del caso la podrán encontrar en:

<https://www.kaggle.com/jihyeseo/online-retail-data-set-from-uci-ml-repo>

Deberá responder los incisos que se enlistan a continuación, preparar una presentación con sus resultados, la cual deberá subir a youtube.com (el tiempo del video debe estar entre 5 y 10 minutos).

Incisos:

1. Explorar el dataset train
 - a. ¿Qué tipos de datos identificamos?
 - b. ¿Cuál es la distribución de las variables univariadas?
 - c. Presentar por lo menos 2 tablas de contingencia que relacionen las variables.
 - d. Presentar un listado de preguntas que ayuden a descubrir los patrones en la base de datos de ventas.

Ejemplo:

 - ¿Los clientes son recurrentes o solo compran en una ocasión?
 - ¿Los precios se comportan diferentes según la región?
 - ¿La cantidad de productos diferentes que compran los clientes varían por región?
 - ¿Las ventas presentan alguna estacionalidad por mes?
 - e. Presentar gráficas para responder las preguntas planteadas en el punto anterior (puede utilizar Powerbi o ggplot):
 - i. De barras
 - ii. De dispersión
 - iii. Histograma
 - iv. Boxplot
 - v. Treemap
2. Aplicar los modelos de unsupervised learning
 - a. Clustering (recomendación aplicarlo a clientes)
 - i. Utilizar la gráfica de codo como referencia.

- ii. Medir la silueta para identificar que tan bien funcionó la asignación de sus clústers.
 - b. Association rules (recomendación aplicarlo a productos)
3. Generar una visualización utilizando métodos de geografía analítica (1 en Google y 1 en Powerbi o R).
 - a. Ventas
 - b. Precio promedio
 - c. Cantidad de clientes
 - d. Cantidad de productos diferentes
4. Presenten los resultados de los clústers que identificaron, colóquenles un nombre que los identifique y explicar cuales son las características que comparten.
5. Presente las reglas de asociación que identifica con los productos y presente sus recomendaciones (ofertas o promociones para incentivar las ventas).
6. Conclusiones y recomendaciones para el dueño del negocio

Rubrica:

EVALUACIÓN	CATEGORIA	ASPECTO EVALUADO	PUNTOS
CONTENIDO	EXPLORACIÓN	TIPOS DE DATOS	1.00%
		DISTRIBUCIÓN UNIVARIADA	1.00%
		TABLAS DE CONTIGENCIA	1.00%
		IDENTIFICACIÓN DE PATRONES Y RELACIONES	1.00%
		PREGUNTAS	2.00%
		GRAFICAS DE EXPLORACIÓN	4.00%
	MODELOS	CLUSTERING	10.00%
		REGLAS DE ASOCIACIÓN	10.00%
		ANÁLITICA GEOGRÁFICA	7.00%
	METRICAS DE VALIDACIÓN DE RESULTADO	GRÁFICA DE CODO	2.00%
		SILUETA	2.00%
		SUPPORT, LIFT Y CONFIDENCE	4.00%
	ANÁLISIS	ARGUMENTOS DE ELECCIÓN	2.50%
		CONCLUSIÓN Y RECOMENDACIÓN	2.50%
PRESENTACIÓN		DOMINIO DEL TEMA	20%
		ORGANIZACIÓN Y SECUENCIA	10%
		CALIDAD DE PRESENTACIÓN	10%
		INICIATIVA / INNOVACIÓN	10%
			100%

Entregables:

- Rscript
- Powerbi (si aplica)
- Presentación PowerPoint con procedimiento
- Video en www.youtube.com