

## Caso de Uso: “Segmentación de clientes aplicado a empresas de telecomunicaciones”

**Contexto:** una empresa líder en el sector de telecomunicaciones enfrenta el desafío de **mejorar su estrategia de retención de clientes y optimizar sus ofertas de servicios**. La compañía cuenta con una extensa base de datos de clientes que incluye información demográfica, historial de uso de servicios y patrones de gasto.

Dado que opera en un mercado altamente competitivo y busca utilizar sus datos de clientes de manera más efectiva para mantener su posición de liderazgo. La empresa reconoce que una segmentación adecuada de su base de clientes podría llevar a estrategias más personalizadas y eficientes.

**Objetivo:** Implementar un modelo de segmentación no supervisado que **agrupe a los clientes en clusters significativos, permitiendo:**

- Identificar patrones de comportamiento de los clientes
- Calcular las componentes principales
- Estimar clusters y analizar el perfil de cada cluster

### Requisitos funcionales:

1. **Preprocesamiento de datos:** Los datos requieren limpieza y preparación, incluyendo el manejo de valores faltantes, codificación de variables categóricas y normalización de variables numéricas.
2. **Selección de características:** Determinar qué variables son más relevantes para la segmentación de clientes.
3. **Elección del algoritmo:** Seleccionar el algoritmo de clustering más apropiado para este conjunto de datos y objetivo de negocio.
4. **Determinación del número óptimo de clusters:** Utilizar métodos estadísticos para determinar el número ideal de segmentos de clientes.
5. **Interpretación de resultados:** Analizar y etiquetar los clusters resultantes de manera que sean accionables para el negocio.
6. **Validación del modelo:** Asegurar que la segmentación sea robusta y significativa para los objetivos de negocio.

### Entregables Esperados:

1. PCA de dos dimensiones con la varianza explicada.
2. Un modelo de segmentación no supervisado que agrupe eficazmente a los clientes.
3. Análisis detallado de las características de cada cluster identificado.

4. Recomendaciones estratégicas basadas en los insights obtenidos de la segmentación.
5. Visualizaciones que ilustren la distribución de los clusters y sus características principales.

Columna	Significado	Tipo de Dato
ID	Identificador único del cliente	Texto
Genero	Género del cliente	Categórico (M/F)
Casado	Estado civil del cliente	Categórico (Si/No)
Edad	Edad del cliente	Numérico (Entero)
Graduado	Si el cliente es graduado universitario	Categórico (Si/No)
Profesion	Ocupación del cliente	Categórico
Experiencia laboral	Años de experiencia laboral	Categórico
Gastoscore	Nivel de gasto del cliente o de apalancamiento financiero	Categórico (Bajo/Medio/Alto)
Family_Size	Tamaño de la familia del cliente	Numérico (Entero)
anno_nacimiento	Año de nacimiento del cliente	Numérico (Entero)
Generación	Generación a la que pertenece el cliente	Categórico
Ingresos anuales brutos	Ingresos anuales del cliente	Numérico (Decimal)
Provincia	Provincia de residencia del cliente	Categórico
Digital_encuesta_index	Índice de encuentro digital	Numérico (Entero)
Gasto_medio_mensual_cuota	Gasto medio mensual en cuota	Numérico (Entero)
Abandono	Si el cliente ha abandonado	Categórico (0/1)
Dias_cliente	Número de días como cliente	Numérico (Entero)
camp_anno	Tipo campaña aplicada anualmente.	Numérico (Decimal)

Los datos se encuentran en azure la tabla se llama: IA\_Client\_Segment\_MODIFICADO.

Requisitos para la realización del trabajo:

- Grupos de máximo 1 personas
- Se puede usar cualquier plataforma con python
- Se debe de presentar un documento de máximo 5 diapositivas (sin incluir portadas). Si se presenta más de 5 hojas el trabajo no será evaluado.
- El trabajo debe de contener una solución de viable. No intensificar en la parte técnica.
- El plagio será sancionado con el suspenso automático de la tarea.