

## Facultad de Ingeniería y Ciencias Básicas

## Maestría en Inteligencia Artificial y Ciencia de Datos

**Curso:** Visualización de Datos

**Estudiante:**

- Soren Fabricius Acevedo (22500566).
- Ricardo Muñoz Bocanegra (22500246).
- Nicolas Lozano Mazuera (22500565).
- Juan José Bonilla Pinzón (22502052).
- Juan Manuel García Ortiz (22502268).

## 1. Resumen

En el presente trabajo, desarrollamos la base técnica y estratégica para un dashboard de inteligencia competitiva destinado al **Banco de Occidente**. A través del procesamiento de datos abiertos de la Superintendencia Financiera de Colombia (SFC), donde buscamos transformar registros masivos de tasas y desembolsos en un activo estratégico. El objetivo es anticipar las dinámicas del mercado, realizar benchmarking frente a competidores directos y monitorear el *market share* con una granularidad territorial (Municipios) y por producto (Libre Inversión, Libranza, Vivienda, Tarjeta de Crédito, Vehículos), permitiendo una toma de decisiones ágil en políticas de *pricing* (Tasas) y despliegue comercial (Fuerza Comercial).

## 2. Contexto Organizacional y Usuarios

El proyecto se sitúa en el sector de la banca de consumo y vivienda en Colombia.

**Usuario Líder:** Gerencia de Planeación Comercial

### Decisiones de negocio (Aporte Visual al negocio):

- Ajustes dinámicos de tasas de interés (*pricing*) basados en las Tasas Promedio Ponderadas (TPP).
- Focalización de campañas de marketing en municipios con baja penetración.
- Reasignación de metas comerciales regionales según el comportamiento de la competencia.

## 3. Pregunta de Negocio y Justificación

**Planteamiento del Problema:** ¿Cómo fluctúa la sensibilidad de los desembolsos ante los cambios en las tasas de interés de la competencia por región, y en qué productos del Banco de Occidente presenta brechas de participación frente al benchmark del sector?

**Justificación de la Analítica Visual:** Dada la magnitud del dataset (~7.82 millones de registros), los reportes tabulares tradicionales resultan insuficientes para identificar patrones complejos. La analítica visual es indispensable para:

- Detectar Outliers:** Identificar anomalías en tasas que podrían indicar guerras de precios regionales.
- Análisis Geoespacial:** Comprender la distribución del crédito en los 32 departamentos de Colombia mediante coropletas.

- Tendencias Temporales:** Observar la evolución semanal de las tasas EA o MV% y su correlación con el volumen de crédito colocado.

#### 4. Hipótesis Analíticas

Para guiar nuestra exploración de datos, hemos formulado las siguientes hipótesis:

- H1 (Elasticidad):** Existe una correlación negativa significativa entre la tasa de interés activa y los montos desembolsados, cuya pendiente varía según el tipo de producto (ej. Vivienda vs. Consumo).
- H2 (Competitividad):** El Banco de Occidente mantiene un Market Share superior en territorios donde su tasa promedio se sitúa al menos un decil por debajo de la mediana del mercado.
- H3 (Brecha de Género):** Existen disparidades significativas en el volumen de desembolsos por sexo que no están correlacionadas con la tasa de interés, sugiriendo oportunidades de inclusión financiera.

#### 5. Descripción y Calidad de los Datos

El conjunto de datos proviene de la SFC y abarca información semanal de tasas y desembolsos.

- Volumen:** Aproximadamente 7.82 millones de filas.
- Variables a Observar:** Registro semanal por Entidad, Producto, Fecha y Municipio.
- Calidad:** Estimamos un nivel inicial de limpieza de **60/100**. Se requiere una fase de normalización de nombres de entidades y la validación de integridad en los códigos DANE de municipios para asegurar un mapeo geográfico preciso.

#### 6. Diccionario de Datos (Variables Seleccionadas)

A continuación, definimos las variables críticas para la construcción de nuestros KPIs:

Variable	Descripción	Tipo	Unidad	Dominio / Rango	Periodicidad	Agregaciones
Fecha_Corte	Fecha de referencia del reporte	Temporal	N/A	Fechas válidas	Semanal	Trimestre, Año
Nombre_Entidad	Nombre del banco o entidad	Categórica	N/A	Catálogo SFC	Semanal	Conteo, Grupos
Producto_Credito	Tipo de cartera (Vivienda, etc.)	Categórica	N/A	Lista cerrada	Semanal	Filtro por Producto
Tasa_Interes_Act	Tasa de interés informada	Numérica	% E.A.	0% - 100%	Semanal	Promedio ponderado
Montos_Desemb	Monto total colocado	Numérica	COP	COP\$	Semanal	Sumatoria (Total)
Codigo_Mun	Código DANE del municipio	Categórica	N/A	5 dígitos	Semanal	Agregación a Dpto
Sexo	Género del titular	Categórica	N/A	M, F, N/A	Semanal	% Participación
Nombre_Tipo_Entidad	Tipo de Compañía	Categórica	N/A	Lista cerrada	Semanal	Grupos

#### 7. Variables No Utilizadas y Justificación

Teniendo en cuenta que el dataset cuenta con 23 Variables y para cumplir los objetivos del análisis o Dashboard, hemos seleccionado las principales variables (8), por lo cual justificamos a continuación la exclusión de las otras 15 Variables.

- Redundancia Operativa (Tipo\_Entity, Código\_Entity, Número\_de\_creditos\_desembolsados):** Para el análisis de competencia, el nombre comercial de la entidad es suficiente.
- Fuera del Alcance Estratégico (Tipo\_de\_Persona, Tamaño\_de\_Empresa, Clase\_deudor, Código\_CIIU, Tipo\_de\_crédito, Grupo\_Etnico, Antigüedad\_de\_la\_empresa, Tipo\_de\_garantía):** Estas variables corresponden a una micro-segmentación operativa. Nuestro enfoque es la planeación comercial macro y regional; incluir estas dimensiones añadiría una complejidad innecesaria que no responde a la pregunta de negocio planteada.
- Complejidad y Ruido Visual (Plazo, Margen\_adicional, Tipo\_de\_Tasa, Rango\_monto\_desembolsado):** El factor decisor en la estrategia de pricing es la Tasa Promedio Ponderada. Los componentes internos de la tasa o los plazos son detalles tácticos que no aportan valor a la visualización de la tendencia principal de mercado y el volumen de desembolsos.

## 8. Reglas de Negocio y Definición de KPIs

Para asegurar la reproducibilidad del análisis, aplicaremos las siguientes fórmulas:

- Market Share (MS):** 
$$MS = \frac{\sum \text{Monto Entidad}}{\sum \text{Monto Mercado}} \times 100$$
- Benchmark de Tasa:** Se creará una Columna que tome (Tasa de interés EA \* Monto Desembolsado), con esto calcularemos la TPP de la siguiente forma: 
$$TPP = \frac{(Tasa \times Desembolso)}{\text{Monto Entidad}} \times 100$$
- Integridad Territorial:** Dado que contamos con el código de los municipios, pero no el nombre, se extraerá una base de datos del **DANE**, la cual contenga 3 Columnas (Código Municipio – Nombre Municipio – Nombre Departamento) y se creará una relación dentro del tablero o Dashboard.