A blue cover with gold text

Description automatically generated

OBJETIVOS  
Objetivo general:  
Construir un benchmark respaldado por una API que calcule el Costo Total Anual (Annual Total Cost) para cuentas vista según un perfil de uso y recomiende el plan de menor costo esperado para el usuario.

Objetivos específicos:

1. Consolidar datos oficiales: ingestión y unificación de CMF y Banco Central (UF) en Parquet.
2. Estandarizar esquema: definir un esquema canónico y reglas de validación para costos mensuales en CLP y valores normalizados en UF.
3. Motor de cálculo: implementar un ATC determinista y generar rankings comparativos.
4. API y visualización: exponer endpoints mínimos y una vista comparativa básica para la toma de decisiones.

JUSTIFICACIÓN  
La transparencia financiera permite a consumidores y PYMEs (Pequeñas y medianas empresas) elegir productos bancarios con menor costo total. La CMF publica un simulador con un “costo neto total” mensual por institución y producto, aunque la información está dispersa para comparaciones masivas. Centralizarla, normalizarla con la UF oficial del Banco Central y exponerla mediante un servicio reproducible reduce la asimetría de información y mejora la comparabilidad. En una fase posterior se incorporarán tablas oficiales del SERNAC para ampliar el alcance a tarjetas.

“**SERNAC** es el organismo público chileno que **protege los derechos de los consumidores**. En pocas palabras: **es la institución que vela porque las empresas cumplan la ley y traten bien a los clientes**, recibiendo reclamos, mediando conflictos y fiscalizando prácticas comerciales”

CODIGOS UNESCO

* 1209.03: Análisis de Datos (Cálculo de ATC y benchmarking).
* 1203.96: Bases de Datos (Almacenamiento y estructuración de datos bancarios).
* 1203.88: Inteligencia de Negocio (Generación de valor para la toma de decisiones).
* 1203.93: Cloud Computing (Arquitectura API y despliegue en contenedores).

ALCANCE  
En esta entrega:  
• Ingesta de CMF simulación promedio de cuentas vista (HTML tabular) y serie UF (BDE o CMF API).  
• Limpieza: parseo de moneda a CLP numérico, estandarización de nombres, derivación de ATC anual, perfilado de columnas.  
• Salida: Parquet canónico, ranking Top-N por menor ATC anual, endpoints de descarga y perfilado.

Fuera de alcance en Entrega 1:  
• Parsing de documentos PDF.  
• Datos personales o sensibles.

Fases siguientes:  
• Incorporación de tablas del comparador de tarjetas SERNAC.  
• Perfiles de uso paramétricos y recomendador.  
• Persistencia relacional y endpoints públicos de ATC y recomendación.

COMPETENCIAS (niveles propuestos 1 a 5)  
Estructurar información: nivel 5. Esquema canónico y validaciones a partir de datos reales.  
Diseñar sistemas de datos: nivel 3. Parquet ahora, ORM y DB listos para la siguiente fase.  
Técnicas de ML: nivel 3. Se deja para fases posteriors, es determinista.  
Modelización estadística: nivel 3. Descriptivos, normalización a UF (Unidad de Fomento) y ranking.  
Gestión de proyectos: nivel 3. Plan, riesgos y hitos definidos.  
Cumplimiento legal y normativo: nivel 4. Solo fuentes públicas oficiales, con trazabilidad.  
Visualización: nivel 3. Tabla comparativa simple y métricas de resumen.

DATOS

CMF simulador de cuentas vista, vista de simulación promedio.  
Tipo: página HTML pública.  
Campos observados en Parquet: Institución, Producto, Costo neto total (cadena con símbolo $), source, extraction\_date.  
Uso: línea base para costo mensual y derivación de ATC anual.

UF oficial para normalización (elegir una de dos rutas, ambas válidas y oficiales):  
a) CMF API v3 UF. Requiere API key y permite JSON o XML para día, mes o periodos.  
b) Banco Central de Chile BDE GetSeries. Requiere credenciales; entrega series por código con rangos de fechas.  
Campos observados en Parquet: fecha, valor\_uf, source, serie\_id, extraction\_date.

SERNAC comparador de tarjetas de crédito.  
Tipo: página HTML pública.  
Estado: pendiente para fase 2, no necesario en E1.

NOTA SOBRE LOS DATASETS ACTUALES  
cta\_cuentavista\_cmf.parquet  
• Estructura recibida: Institución (texto), Producto (texto), Costo neto total (texto con $ y separadores), Unnamed: 3 (nulo), source (texto), extraction\_date (datetime).

• Limpieza mínima aplicada: eliminación de columna vacía, parseo de moneda a float en CLP, renombres canónicos (institucion, producto, cta\_mensual\_clp, fuente, fecha\_captura) y derivación cta\_anual\_clp = cta\_mensual\_clp x 12.

indicadores\_uf.parquet  
• Estructura recibida: fecha (datetime), valor\_uf (float), source (texto), serie\_id (texto), extraction\_date (datetime).  
• Limpieza mínima aplicada: renombres canónicos (fecha, valor, fuente, serie\_id, fecha\_captura).

RESULTADOS PRELIMINARES  
Metodología breve:  
• Los montos mensuales provienen de la simulación promedio CMF.  
• ATC anual se calcula como ATC mensual multiplicado por 12.  
• La UF se utilizará en la fase siguiente como normalizador, manteniendo fecha de referencia y trazabilidad de la serie.

A screenshot of a black screen

Description automatically generatedTabla 1. Top 10 cuentas vista más baratas por ATC mensual y anual (CLP)

Tabla 2. Resumen estadístico del ATC mensual y anual en CLPA screenshot of a graph

Description automatically generated

A screenshot of a computer

Description automatically generatedTabla 3. Resumen de la serie UF observada

RIESGOS Y MITIGACIÓN

1. Cambios de estructura HTML en CMF o SERNAC. Mitigación: selectores robustos y pruebas. Si falla, CSV manual mínimo y trazable para continuidad.
2. Disponibilidad de servicios UF. Mitigación: alternar entre CMF API y BDE según disponibilidad y cachear últimas lecturas.
3. Frescura de datos. Mitigación: trazabilidad con fecha\_captura y recapturas programadas.
4. Formatos numéricos heterogéneos. Mitigación: funciones centralizadas de parseo y validación de tipos.

PLANIFICACIÓN Y CRONOGRAMA. REVISAR  
Hito 1 Limpieza y esquema canónico de CMF y UF. Estado: en curso.  
Hito 2 EDA mínimo y ranking Top-N por ATC anual. Estado: pendiente inmediato.  
Hito 3 Documento corto, checklist de calidad y referencias. Estado: pendiente inmediato.  
Fase 2 SERNAC tarjetas, perfiles de uso y recomendaciones.  
Fase 3 Persistencia relacional, endpoints públicos y tablero simple.

![A screenshot of a computer program

Description automatically generated]()

REFERENCIAS   
[Comisión para el Mercado Financiero. Simulador de cuentas de depósito a la vista: simulación promedio y definición de costo neto total. Accedido en diciembre de 2025.](https://servicios.cmfchile.cl/simuladorcuentavista/simulacionpromedio)

[Comisión para el Mercado Financiero. Documentación API CMF Bancos v3: Unidad de Fomento (UF). Accedido en diciembre de 2025.](https://api.cmfchile.cl/documentacion/UF.html)

[Banco Central de Chile. Statistics Database API BDE: documentación GetSeries y catálogo de series. Accedido en diciembre de 2025.](https://si3.bcentral.cl/Siete/es/Siete/API?respuesta=)

[Servicio Nacional del Consumidor. Comparador de tarjetas de crédito. Accedido en diciembre de 2025.](https://www.sernac.cl/portal/619/w3-article-64916.html)