

# TRABAJO - Fundamentos de Big Data

“FinPlus Analytics Challenge”

## 1. Contexto

La empresa **FinPlus**, una entidad financiera en proceso de transformación digital, ha lanzado una convocatoria para seleccionar un **equipo de analítica de datos** que desarrolle una solución integral para comprender y optimizar el comportamiento de sus clientes.

FinPlus cuenta con información anonimizada de su base de clientes y de sus patrones de comportamiento digital y transaccional. La dirección busca contratar al equipo capaz de desarrollar la **propuesta analítica más completa, eficiente y visualmente poderosa**, capaz de ofrecer *insights accionables* y un plan de mejora basado en datos.

Cada grupo competirá como si fuera una **consultora de Big Data**. Solo una será seleccionada como ganadora del contrato.

## 2. Objetivo general

Diseñar y presentar una **solución analítica end-to-end** que permita a FinPlus:

1. Comprender mejor su base de clientes (perfilado, segmentación y hábitos).
2. Identificar patrones de comportamiento clave (uso de productos, actividad, interacción).
3. Detectar riesgos de abandono o pérdida de rentabilidad.
4. Encontrar oportunidades de crecimiento (cross-selling o engagement).
5. Proponer indicadores (KPIs) y visualizaciones que faciliten la toma de decisiones.

## 3. Datos disponibles

- **CLIENTS.csv** → información demográfica y contractual.
- **BEHAVIOURAL.csv** → comportamiento digital y transaccional.
- **Data\_Dictionary.xlsx** → explicación de las variables.

Los datos están anonimizados, pero reflejan el tipo de información real con la que trabajaría un departamento de *Customer Analytics*.

## 4. Fases del trabajo

Cada grupo deberá desarrollar un proyecto completo con una estructura profesional y justificada.

#### 4.1. Fase de análisis de negocio

- Comprensión del contexto y definición del problema.
- Identificación de los objetivos específicos de análisis (qué quiere descubrir o predecir FinPlus).
- Recogida de **requisitos funcionales y técnicos**:
  - Funcionales: qué debe conseguir la solución. *Ej.: segmentar, detectar patrones, monitorizar KPIs, etc.*
  - Técnicos: entorno, herramientas, arquitectura, librerías, dependencias, etc.

Entregable 1: *Documento de propuesta inicial* (máx. 2 páginas).

#### 4.2. Fase de diseño del proyecto

- Creación del **esquema general del trabajo**: arquitectura de datos, flujo ETL, almacenamiento, visualización, etc.
- Definición de **roles y responsabilidades** dentro del grupo: Data Engineer, Analyst, Architect, Project Lead, etc.
- Planificación del trabajo: timeline, milestones, control de versiones.

Entregable 2: *Diagrama de arquitectura + planificación de tareas*.

#### 4.3. Fase de construcción del entorno colaborativo

- Creación de un entorno Docker con PySpark y dependencias configuradas.
- Repositorio en GitHub / GitLab / Bitbucket con control de versiones y documentación ([README .md](#)).
- Estructura clara de carpetas y notebooks de análisis.

Entregable 3: Evidencias del trabajo colaborativo: repositorio, commits, issues, pull requests.

#### 4.4. Fase de desarrollo técnico (ETL y análisis)

- Implementar un **pipeline ETL completo en PySpark**: Ingesta → Limpieza → Transformación → Almacenamiento → Agregación.
- Generar datasets de análisis bien estructurados (*curated layer*).
- Controlar la calidad de datos: nulos, duplicados, consistencia, tipos de datos, etc.
- Elaborar una serie de **indicadores y análisis de comportamiento**, como mínimo:
  - **Actividad del cliente**: frecuencia, intensidad, recencia...

- **Valor económico:** rentabilidad, compras, volumen...
- **Interacción y fidelidad:** uso de canales, repetición, respuesta a estímulos...
- **Riesgo potencial:** indicadores de abandono, inactividad...
- **Oportunidades comerciales:** clientes valiosos con baja interacción...

La empresa valorará positivamente cualquier aportación extra que se lleve a cabo.

Entregable 4: Código con ETL y Análisis.

## 4.5. Fase de visualización

- Crear un **dashboard en Tableau / PowerBI** que incluya:
  - Un mínimo de 4 KPIs relevantes.
  - Un mínimo de 5 visualizaciones explicativas (gráficos comparativos, heatmaps, tendencias, etc.).
  - Filtros interactivos y narrativa visual coherente.

¡IMPORTANTE! Enfocar el dashboard a la dirección de FinPlus, no a perfiles técnicos.

Entregable 5: Dashboard de visualización (enlace).

## 4.6. Fase de análisis avanzado (opcional)

- Si algún equipo quiere o lo considera interesante, puede aplicar análisis predictivos:
  - Segmentación por comportamiento o valor.
  - Cálculo de scores (riesgo, propensión, engagement).
  - Clasificación por tipos de clientes.
  - etc.

Entregable 6 (opcional): Código con el Análisis Avanzado.

# 5. Presentación

- Elaborar una **presentación ejecutiva** (10 minutos máximo).
  - Introducción: problema de negocio y enfoque.
  - Arquitectura: pipeline, tecnologías y decisiones.
  - Resultados y visualizaciones clave.
  - Insights y recomendaciones para la empresa.
- Presentar ante el resto de la clase y el profesor (como “comité de dirección de FinPlus”).
- La mejor propuesta será elegida como la **solución ganadora del contrato**.

## 6. Rúbrica del Trabajo (total 100 puntos)

Área	Descripción	Ponderación
Definición y enfoque de negocio	Claridad en el problema, realismo, coherencia con los datos	15%
Diseño técnico y arquitectura	Pipeline, estructura de carpetas, planificación y roles	15%
Implementación técnica (PySpark + ETL)	Limpieza, calidad de datos, modularidad y funcionalidad	25%
Análisis y KPIs	Pertinencia y solidez de los indicadores, consistencia analítica	15%
Dashboard y comunicación visual	Claridad, diseño, narrativa visual, storytelling ejecutivo	20%
Originalidad, innovación y creatividad	Argumentación, creatividad, profesionalidad del equipo	10%

>> La **defensa oral** se evaluará como un elemento a parte tal y como se indica en la Guía Docente de la Asignatura.

## 7. Herramientas obligatorias

- **Python / PySpark** para el procesamiento y análisis.
- **Docker (o semejante)** para el entorno de trabajo reproducible.
- **Github (o semejante)** para control de versiones.
- **Tableau (o semejante)** para la visualización de resultados.
- **Documentación en Markdown o PDF** con esquema del proyecto.

## 8. Criterios de victoria

El jurado valorará especialmente:

- Originalidad en el enfoque y claridad del storytelling.
- Coherencia entre los objetivos iniciales y los resultados obtenidos.
- Solidez técnica del pipeline (código limpio, reproducible, documentado).
- Capacidad para transformar datos en *insights de negocio reales*.
- Impacto visual y comunicativo del dashboard.

El equipo con la propuesta mejor valorada recibirá el reconocimiento como  
**“FinPlus Analytics Partner 2025”** 

## 9. FinPlus Assistant (ChatBot de Interacción)

Enlace a [FinPlus Assistant](#)

**FinPlus Assistant** es un chatbot que representa oficialmente a la empresa y será su único canal de comunicación para resolver dudas sobre la convocatoria y la documentación entregada. El propósito de interactuar con FinPlus es obtener **aclaraciones sobre:**

- Contexto.
- Objetivos.
- Criterios de evaluación.
- Información disponible en los archivos.

### ✓ Ejemplos de preguntas aceptadas

- ¿Cuál es el principal objetivo del proyecto de FinPlus?
- ¿Qué tipo de información contiene la base de datos proporcionada?
- ¿Qué resultados espera la dirección con esta iniciativa?
- ¿Qué criterios utiliza FinPlus para evaluar las propuestas?
- ¿Cuál es el alcance de la transformación digital que están impulsando?

### ✗ Ejemplos de preguntas NO aceptadas

- Cualquier consulta sobre **métodos analíticos, herramientas o modelos de machine learning**.
- Preguntas relacionadas con **código, scripts o interpretaciones técnicas**.
- Solicitudes de **información no contenida en la documentación o de datos adicionales**.
- Consultas que impliquen **revelar soluciones o estrategias que dependen del trabajo del equipo**.

En estos casos, FinPlus responderá de forma cortés pero firme que la información solicitada **no puede ser compartida** o que **corresponde a la interpretación del equipo participante**.

>> **Nota:** el chatbot está pensado para ayudaros a entender el contexto de negocio. Si queréis ayuda con código, datos, etc. usad vuestro propio LLM (chatGPT, Gemini, Perplexity, etc.).

## 10. Formato del Trabajo

- **Formato:** Máximo 4 personas.
- **Definición del Equipo:** se enviará al profesor un email con:
  - Nombre del equipo.

- Nombre completo de cada miembro.
- **Rol de cada miembro** del equipo: debéis definir el rol que cada uno desempeñará en el proyecto, así como una explicación de las funciones principales (ej.: data scientist, tech lead, etc.).
- **Repositorio GitHub o GitLab** con una estructura parecida a la siguiente (no tiene por qué ser la misma):

```
/FinPlusAnalytics_TeamX/
├── data/          # Datasets (sin modificar)
├── notebooks/    # Notebooks de análisis
├── src/           # Scripts ETL o funciones auxiliares
├── docker/        # Configuración del entorno
├── docs/          # Documentación (PDF, diagramas, informes)
├── dashboards/   # Capturas o enlaces a Tableau
├── README.md      # Guía de instalación y uso
└── requirements.txt # Dependencias del entorno
```

- **Entregables:**
  - Entregable 1: *Documento de propuesta inicial* (máx. 2 páginas).
  - Entregable 2: *Diagrama de arquitectura + planificación de tareas*.
  - Entregable 3: Evidencias del trabajo colaborativo: repositorio, commits, issues, pull requests.
  - Entregable 4: Código con ETL y Análisis.
  - Entregable 5: Dashboard de visualización (enlace).
  - Entregable 6 (opcional): Código con el Análisis Avanzado.
- **Requisitos adicionales:**
  - Todo el código debe ejecutarse correctamente dentro del contenedor Docker.
  - Los notebooks y scripts deben estar documentados y comentados.
  - Cada commit debe reflejar un cambio claro y trazable.
  - El repositorio debe incluir al menos un **archivo README** con:
    - Descripción del proyecto.
    - Instrucciones para replicar el entorno y ejecutar el pipeline.
    - Roles y tareas del equipo.