# Informe Sprint 1 — Proyecto Crypto ETL + EDA

Universidad de Ibagué — Ingeniería de Sistemas

Materia: Business Intelligence / Análisis de Datos

Fecha: 27/10/2025

Versión: 1.0

Autores:

• Juan David Reyes Cure

• Julio David Suarez Olaya

• Adriana Michelle Diaz Suarez

## 1. Introducción

El presente informe corresponde al Sprint 1 del proyecto denominado 'Crypto ETL + EDA', cuyo propósito es construir un pipeline reproducible para la recolección, limpieza y análisis exploratorio de datos históricos de criptomonedas. Este sprint se centró en crear una base sólida de datos limpios, confiables y listos para modelado predictivo. Los activos analizados fueron Bitcoin (BTC), Ethereum (ETH) y Binance Coin (BNB).

## 2. Objetivos del Sprint 1

• Identificar fuentes de datos confiables (CoinGecko y Binance).

• Diseñar y probar métodos de extracción mediante APIs públicas y autenticadas.

• Procesar y limpiar datos (manejo de nulos, duplicados y consistencia temporal).

• Almacenar los datos en formato estructurado (CSV).

• Elaborar un informe exploratorio con estadísticas y visualizaciones.

• Definir el stack tecnológico para los siguientes sprints.

## 3. Tecnologías y Fuentes de Datos

El proyecto utiliza el lenguaje Python (versión 3.10 o superior) con librerías orientadas al análisis de datos y visualización, como pandas, numpy, requests, matplotlib y seaborn. Las fuentes principales fueron las APIs de CoinGecko (requiere API Key) y Binance (pública), que permiten obtener información histórica diaria sobre precios, volumen y capitalización de mercado.

## 4. Proceso de Desarrollo

• Extracción de datos: Se implementaron dos rutas, una con CoinGecko (con API Key) y otra con Binance (sin autenticación). Los scripts `extract\_real\_data.py` y `extract\_data.py` automatizan la recolección y almacenamiento de la información en formato CSV.

• Limpieza y feature engineering: Se eliminaron duplicados, se estandarizaron fechas, se trataron valores nulos y se generaron variables derivadas como `daily\_return`, `log\_return`, `roll\_vol\_30d` (volatilidad anualizada) y `roll\_mean\_30d` (promedio móvil de 30 días). Todo el proceso está automatizado en `clean\_data.py`.

• Análisis exploratorio (EDA): Se generaron estadísticas descriptivas y se analizaron las tendencias de precios, volatilidad y retornos para los tres activos, con salidas tabulares y gráficas generadas mediante `eda\_report.py`.

## 5. Resultados del Sprint 1

Como resultado, se obtuvo un dataset consolidado con más de 30.000 registros limpios, abarcando información histórica diaria de BTC, ETH y BNB. Este dataset está listo para ser utilizado en el modelado predictivo y visualización avanzada. Los principales entregables son:

- data/crypto\_clean\_BTC\_ETH\_BNB.csv — Dataset limpio consolidado.

- reports/eda/EDA\_summary.csv — Estadísticas descriptivas por activo.

- docs/Sprint1\_EDA\_Report.md — Informe técnico del sprint.

- reports/eda/\*.png — Visualizaciones generadas durante el EDA.

## 6. Conclusiones

El Sprint 1 cumplió con todos los objetivos establecidos, entregando un pipeline funcional y automatizado para la extracción, limpieza y análisis exploratorio de datos de criptomonedas. Se logró obtener un dataset limpio, estructurado y validado, que será la base para el entrenamiento de modelos predictivos en los siguientes sprints.

El análisis exploratorio permitió identificar comportamientos de tendencia y volatilidad en los activos digitales, brindando una comprensión inicial del mercado que servirá para afinar las estrategias de modelado y predicción.

## 7. Próximos Pasos

• Validar outliers y huecos temporales en el dataset.

• Incorporar indicadores técnicos (RSI, EMA, MACD) para enriquecer el análisis.

• Implementar modelos de predicción de precios (Prophet, ARIMA, LSTM).

• Desarrollar una API con FastAPI para servir los resultados.

• Crear un dashboard interactivo con Streamlit.

## 8. Anexos

Estructura del proyecto:  
BI\_3/  
├─ scripts/  
│ ├─ extract\_data.py  
│ ├─ extract\_real\_data.py  
│ ├─ clean\_data.py  
│ └─ eda\_report.py  
├─ data/  
│ ├─ raw\_crypto.csv  
│ └─ crypto\_clean\_BTC\_ETH\_BNB.csv  
├─ reports/eda/  
│ ├─ \*.png  
│ └─ EDA\_summary.csv  
└─ docs/Sprint1\_EDA\_Report.md

Checklist de cumplimiento:  
✓ Fuentes de datos identificadas  
✓ Extracción y limpieza automatizadas  
✓ Dataset limpio almacenado en CSV  
✓ EDA generado con estadísticas  
✓ Stack tecnológico definido para Sprint 2