

# Cryptocurrency Data Pipeline

## Project Report

### Team Members

Student ID	Name
22B031332	Shim Daniil
22B030340	Jakupov Dias
22B030372	Kabiyev Zhanbek

### Project Overview

This project implements a real-time cryptocurrency market data pipeline using Apache Airflow, Apache Kafka, and SQLite. The system collects global market metrics from the CoinMarketCap API, processes the data through a streaming pipeline, and generates daily analytics with visualizations.

#### Technology Stack:

- Apache Airflow 2.7.0 - Workflow orchestration
- Apache Kafka - Message streaming
- PostgreSQL - Airflow metadata database
- SQLite - Application data storage
- Docker Compose - Container orchestration
- Python (pandas, matplotlib) - Data processing and visualization

## Job 1: Data Ingestion (Producer)

**Purpose:** Fetches real-time cryptocurrency market data from CoinMarketCap API and publishes to Kafka.

**Schedule:** Runs every 5 minutes via Airflow DAG

**Process:**

- Connects to CoinMarketCap Global Metrics API
- Fetches data every 30 seconds for 5 minutes per DAG run
- Publishes raw JSON responses to Kafka topic 'raw\_events'
- Handles API errors gracefully with retry logic

**Data Collected:**

- Total market cap and 24h volume
- BTC and ETH dominance percentages
- DeFi, Stablecoin, and Derivatives metrics
- Active cryptocurrencies and exchanges count

## Job 2: Data Cleaning and Storage

**Purpose:** Consumes raw data from Kafka, cleans it, and stores in SQLite database.

**Schedule:** Runs every hour via Airflow DAG

**Cleaning Operations:**

- Parses nested JSON structure from CoinMarketCap API
- Converts timestamps to datetime format
- Removes records with null timestamp or market cap
- Deduplicates records based on timestamp
- Clips negative values in numeric columns to 0
- Forward-fills missing dominance values

**Output:** Cleaned records stored in 'events' table with 17 metrics per record

## Job 3: Daily Analytics and Visualization

**Purpose:** Computes daily aggregations and generates visualization charts.

**Schedule:** Runs daily at midnight via Airflow DAG

### **Analytics Computed:**

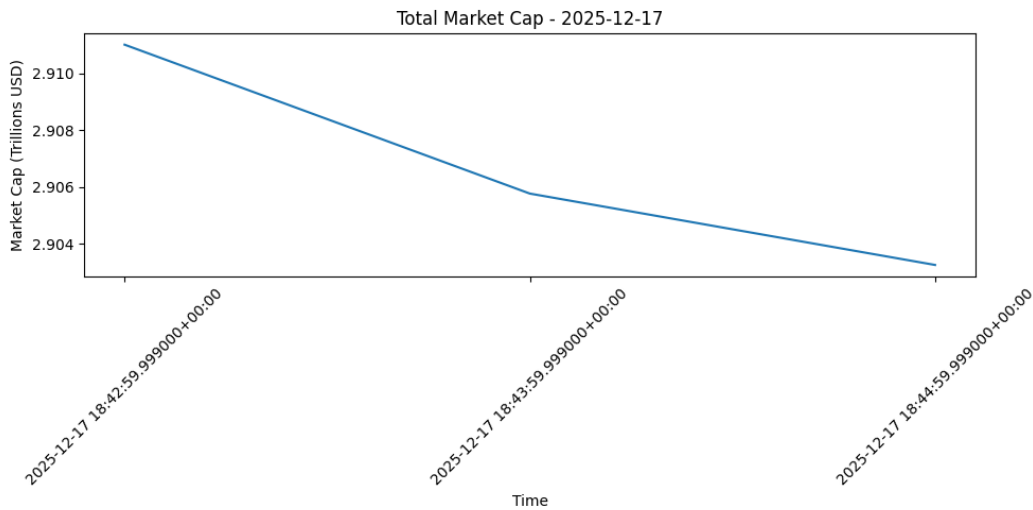
- Average, min, max market cap and volume
- BTC/ETH dominance statistics (avg, min, max, std dev, percentiles)
- Daily BTC dominance change
- DeFi, Stablecoin, and Derivatives averages
- High volatility event count (>5% daily change)
- Records per hour metric

### **Visualizations Generated:**

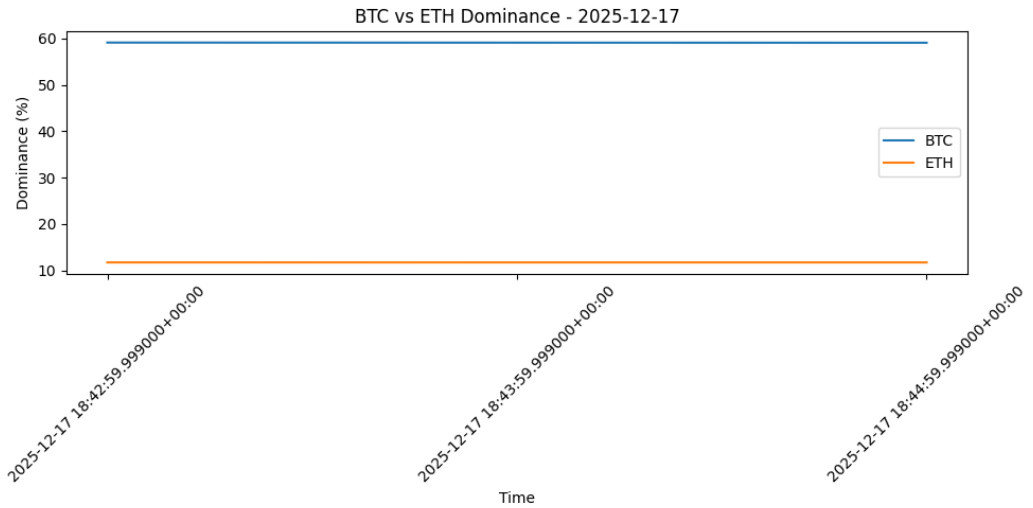
- Market Cap Timeline - Line chart showing intraday market cap movement
- BTC vs ETH Dominance - Dual line chart comparing dominance trends
- Market Segments Pie - Distribution of DeFi, Stablecoins, and Altcoins

# Generated Charts

## Total Market Cap Timeline

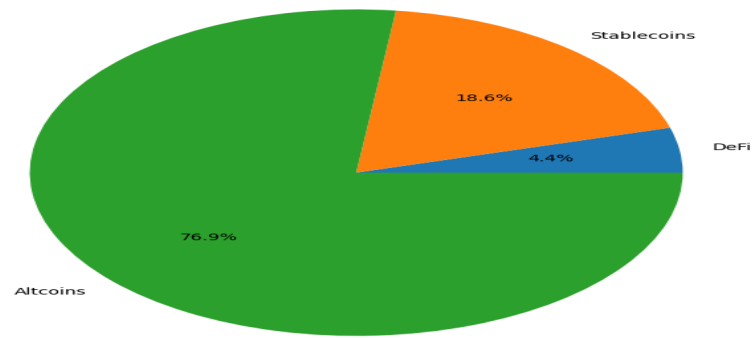


## BTC vs ETH Dominance



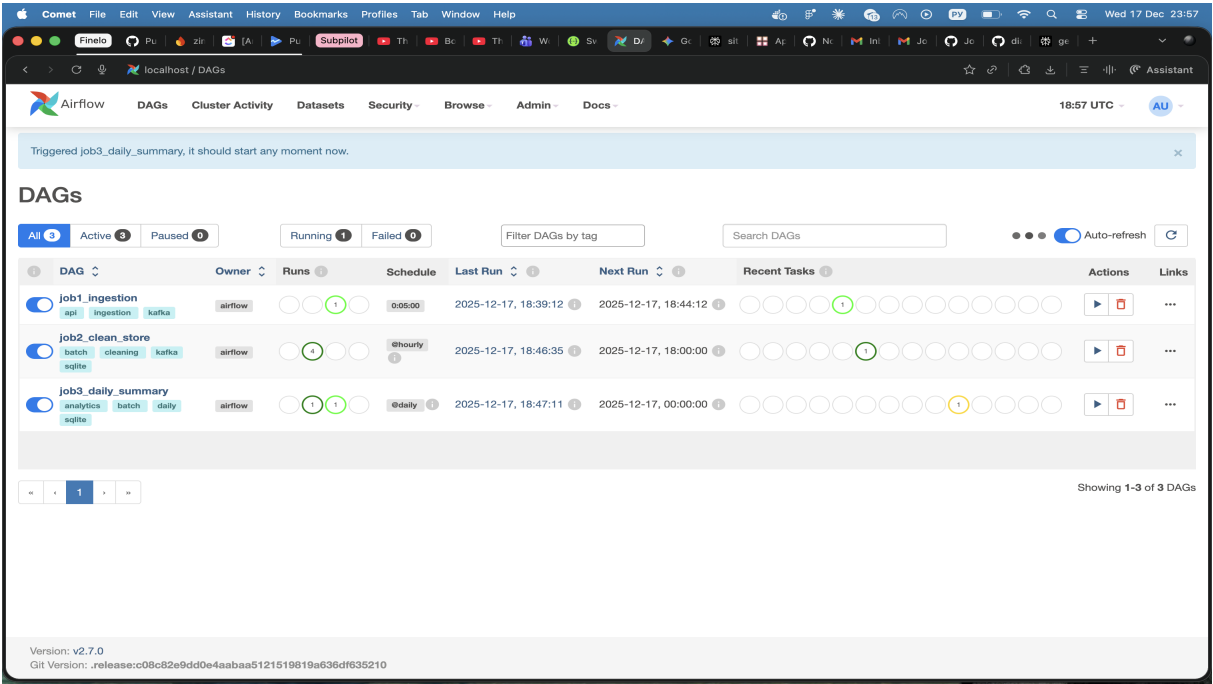
## Market Segments Distribution

Market Segments - 2025-12-17



# System Screenshots

## Airflow DAGs Dashboard - Shows all three pipeline jobs



# Daily Summary Table - Aggregated analytics data

AntigravityFileEditSelectionViewGoRunTerminalWindowHelp

final - app.dbOpen Agent ManagerRestart to Update

data > app.dbFilter 3 tables...

TABLES

daily\_summary

ROWID

id

date

total\_re...

avg\_total\_ma...

min\_total\_ma...

max\_total\_m...

avg\_total\_vol...

avg\_btc\_do...

min\_btc...

total\_records

avg\_total\_marke...

min\_total\_marke...

max\_total\_marke...

avg\_total\_volum...

avg\_btc\_domina...

min\_btc\_domina...

max\_btc\_domin...

avg\_eth\_domina...

btc\_dominance...

avg\_defi\_market...

avg\_defi\_volum...

avg\_stablecoin...

avg\_stablecoin...

avg\_derivatives...

avg\_defi\_24h\_c...

avg\_stablecoin...

avg\_derivatives...

avg\_active\_cryp...

avg\_active\_exch...

std\_total\_marke...

std\_btc\_domina...

p25\_btc\_domina...

p75\_btc\_domina...

median\_total\_m...

high\_volatility\_c...

records\_per\_hour

created\_at

events

ROWID

id

1

2025-12-17

3

2906678657946.1323

2903248121135.7666

2911021160079.5693

107093885174.23999

59.07640051861767

59.8575

1

2

SQLITE VIEWER FREE v26.12.3

main\*

Git Graph

Copy as Excel

Go LiveAntigravity - Settings

## Daily Summary Schema - Database structure

The screenshot displays the Antivray application interface. At the top, the title bar reads 'Antivray | File | Edit | Selection | View | Go | Run | Terminal | Window | Help'. The main window is titled 'final - app.db' and shows a table of data. The table has columns: 'id', 'date', 'total\_re...', 'avg\_total\_ma...', 'min\_total\_ma...', 'max\_total\_m...', 'avg\_total\_vol...', 'avg\_btc\_d...', and 'min\_btc...'. The data is filtered for '2025-12-17'. The table shows two rows of data. The first row has 'id' 1 and 'date' 2025-12-17. The second row has 'id' 2 and 'date' 2025-12-17. The table is part of a larger application window that includes a sidebar with a file explorer, a top bar with various icons, and a bottom bar with a status bar and a 'Go Live' button. The status bar at the bottom indicates 'v25.12.3' and 'Page 1/1'.



# Events Table - Raw cleaned data records

Antigravity File Edit Selection View Go Run Terminal Window Help

final - app.db Open Agent Manager Restart to Update

data > app.db

Filter 3 tables. Rows: 3

TABLES

# avg\_stablecoin...

# avg\_derivatives...

# avg\_active\_cryp...

# avg\_active\_exch...

# std\_total\_marke...

# std\_btc\_domin...

# p25\_btc\_domin...

# p75\_btc\_domin...

# median\_total\_m...

# high\_volatility\_c...

# records\_per\_hour

# created\_at

events

ROWID

# id

# timestamp

# btc\_dominance

# eth\_dominance

# active\_cryptocu...

# active\_market\_p...

# active\_exchanges

# total\_market\_cap

# total\_volume\_24h

# altcoin\_market...

# altcoin\_volume...

# defi\_market\_cap

# defi\_volume\_24h

# defi\_24h\_perce...

# stablecoin\_mark...

# stablecoin\_volu...

# stablecoin\_24h...

# derivatives\_volu...

# derivatives\_24h...

# created\_at

sqlite\_sequence

id

timestamp

btc\_domi...

eth\_domi...

active\_c...

active...

active\_e...

total\_marke...

total

1

2025-12-17 18:42:59.999000+0...

59.095709217090

11.742650061168

9002

117064

907

2911021160079.5693

107

2

2025-12-17 18:43:59.999000+0...

59.075905923826

11.735568311558

9002

117064

907

2905766692623.061

106

3

2025-12-17 18:44:59.999000+0...

59.057586414929

11.7348080607588

9002

117064

907

2903248121135.7666

107

Page 1 / 1

Copy