# Portfolio 2 — Investment Data Platform

This document describes the scope, objectives, and delivery roadmap for the Investment Data Platform portfolio project, built using Microsoft Fabric.

## Project Objective

Design and implement an end-to-end data engineering platform for financial market data, focused on ingestion, transformation, modeling, and analytics-ready consumption.

## Core Principles

1. End-to-end data engineering focus (not a simple tracker)
2. Production-oriented architecture using Microsoft Fabric
3. Clear separation of raw, curated, and analytics layers
4. Business-oriented financial use case

## Architecture Overview

The project follows a Lakehouse architecture with Medallion layers: Bronze for raw ingestion, Silver for cleaned and standardized data, and Gold for analytics-ready datasets consumed by Power BI.

## Technology Stack

1. Microsoft Fabric (Lakehouse, Data Pipelines, Notebooks)
2. PySpark and SQL for transformations
3. Public financial market APIs
4. Power BI for analytics consumption

## Roadmap

1. **Week 1 — Foundation:** Data source selection, repository setup, Lakehouse creation, Bronze ingestion pipeline.
2. **Week 2 — Processing:** Silver layer transformations, data normalization, historical handling.
3. **Week 3 — Analytics:** Gold layer modeling, financial metrics calculation, BI-ready structures.
4. **Week 4 — Delivery:** Power BI dashboard, documentation, final review.

## Expected Outcomes

1. A complete and documented data engineering project
2. Clear demonstration of Microsoft Fabric capabilities
3. Analytics-ready financial datasets
4. Interview-ready portfolio artifact