University of Science and Technology of Hanoi

Fundamentals of Data Science

S&P 500 Crisis Analysis

Group Members:

Nguyen Tuan Thanh   22BA13289   thanhnt.22ba13289@usth.edu.vn
Nguyen Chi Quang   22BA13262   quangnc.22ba13262@usth.edu.vn

Hanoi, Sep 2025

Contents

ACKNOWLEDGEMENTS

ABSTRACT

Chapter 1

Introduction

Chapter 2

Business Analysis

2.1 Background

2.2 Problem Statement

2.3 Analytical Approach

Chapter 3

Data Collection and Management

3.1   Introduction

The dataset employed in this project is the Standard& Poor's (S&P) 500 Index data. It contains monthly historical records with key financial indicators such as Dividend, Earnings, and Price-to-Earnings (P/E) Ratio. The dataset consists of 1833 entries and 10 features, and is provided in CSV (Comma Separated Values) format.

Date	SP500	Dividend	Earnings	CPI	LIR	Real Price	Real Dividend	Real Earnings	PE10
2023-06-01	4345.37	68.71	181.17	305.11	3.75	4359.88	68.94	181.77	32.41
2023-05-01	4146.17	68.54	179.17	304.13	3.57	4173.45	68.99	180.35	31.14
2023-04-01	4121.47	68.38	177.17	303.36	3.46	4159.03	69.00	178.78	31.15
...	...	...	...	...	...	...	...	...	...
2017-09-01	2492.84	48.17	107.08	246.82	2.20	3091.85	59.74	132.81	32.97
2017-08-01	2456.22	47.85	106.06	245.52	2.21	3062.56	59.67	132.24	32.71

Table 3.1: S&P 500 Dataset

The data types and detailed discussion of each feature in the dataset are provided below:

Features	Type	Data Type	Explanation
Date	datetime	Qualitative	Timestamp of the record
SP500	float	Qualitative	The nominal value at that time.
Dividend	float	Qualitative	The average dividend payout of companies.
Earnings	float	Qualitative	The average earnings of companies.
CPI	float	Qualitative	A measure of inflation and purchasing power.
LIR	float	Qualitative	The 10-year government bond yield.
Real Price	float	Qualitative	The inflation-adjusted S&P 500 index value (using CPI).
Real Dividend	float	Qualitative	Dividend values adjusted for inflation.
Real Earnings	float	Qualitative	Earnings adjusted for inflation.
PE10	float	Qualitative	The cyclically adjusted price-to-earnings ratio.

Table 3.2: Features Explanation

3.2   Problem Context

3.3   Overall Architecture

3.4   Data Collection

3.5   Data Management

3.6   Data Processing

Chapter 4

Statistics

4.1 Descriptive Statistics

4.1.1 Full Sample

4.1.2 Period Analysis

4.2 Distribution Analysis

4.3 Correlation

4.4 Hypothesis Testing

Chapter 5

Visualization

5.1 Market Performance

5.2 Valuation Metrics

5.3 Risk

5.4 Dividends

Conclusion and Future Work

Chapter 6

References