

Report Title*

(COMP3125 Individual Project)

*Note: Do not used sub-title

Given Name Surname
dept. name of organization

Abstract—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. ***CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.** (provide a short abstract)

Keywords—example1, example2, example3, example 4, example 5 (provide 3-5 keywords)

I. INTRODUCTION (HEADING 1)

Stock market analysis is important for understanding trends, guiding investment decisions, and identifying high-performing sectors. This project focuses on analyzing S&P 500 stock sectors to determine which is performing the best, examine price trends over the past year, identify the most stable companies, and predict closing prices using historical data. By combining descriptive analysis and predictive modeling, this study provides insights into sector performance and short-term stock behavior.

II. DATASETS

A. Source of dataset (Heading 2)

The dataset used in this project was obtained from Kaggle, a reliable platform widely used for data analysis and academic research. It contains historical daily prices and sector information for S&P 500 companies and is updated daily, making it suitable for trend analysis and predictive modeling.

S&P 500 Stocks (daily updated)

B. Character of the datasets

The dataset used in this analysis is composed of three CSV files: sp500_companies.csv, sp500_stocks.csv, and sp500_index.csv, although the study primarily focuses on the first two. All files are in standard comma-separated format, with sp500_companies.csv containing approximately 505 rows of company-level metadata, such as ticker symbols, long names, sectors, and subsectors, while sp500_stocks.csv contains several hundred thousand rows of daily price data, including open, high, low, close, volume, and date for each S&P 500 stock. The dataset reflects the structure of the S&P 500 index, which includes 500 companies but 505 stocks due to multiple share classes (e.g., GOOG and GOOGL). Before analysis, the stock and company datasets were merged using the ticker symbol to attach each price record to its respective sector and company name. Basic cleaning steps included removing rows with missing closing prices. Several engineered features were created to support forecasting, such as lagged closing prices and moving averages, computed using rolling window formulas. These additions enhanced the dataset’s analytical value and allowed for sector comparison, stability measurement, and predictive modeling.

III. METHODOLOGY

In this part, you should give an introduction of the methods/model. First, what's the method/model. What's the assumption of this method/model. What's the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extrawork did you adjust to make the results better. If you have multiple methods, feel free to use subsection A., B. to separate them.

Example: Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

A. Method A: Sector Performance Calculation

To compare the performance of all sectors in the S&P 500, the percentage change in the average closing price was computed for each sector. This method is straightforward and provides a simple measure of long-term growth.

$$\text{Percentage Change} = ((\text{Last Close}/\text{First Close}) - 1) * 100$$

Reason for Choosing: This method is easy to interpret and allows sectors to be ranked based on overall price appreciation. Since the task only requires comparing performance, a simple percentage change is sufficient.

- Pandas for grouping data and performing calculations.

B. Method B: Stability Analysis Using Standard Deviation

Standard deviation measures how much prices fluctuate around the average.

- Low standard deviation = more stable stock
- High standard deviation = more volatile stock

Reason for Choosing: Standard deviation is one of the most used volatility metrics in finance. It provides a direct way to identify companies with steady price movement, which is helpful for determining stability within a sector.

- Pandas for groupby operations and computing standard deviation.
- NumPy for numerical support.

Identify applicable funding agency here. If none, delete this text box.

C. *Method C: Price Forecasting (Regression + Exponential Smoothing)*

1. Linear Regression Model

A regression model uses past data to predict future prices based on relationships between variables. In this analysis (using NVDA as an example), the model includes:

- Lagged closing prices
- Moving averages (5-day, 10-day, 20-day)
- Basic features such as Open, High, Low, and Volume

2. Exponential Smoothing (ETS Model)

Exponential Smoothing is a time-series method that gives more weight to recent observations. It is effective for forecasting short-term trends because it adjusts smoothly as new data arrives.

3. To increase reliability, the final prediction is an average of both models:

$$\text{Combined Forecast} = 0.5 * (\text{Regression}) + 0.5 * (\text{ETS})$$

Reason for Choosing: Regression incorporates technical indicators, while ETS captures time-series trend behavior. By combining them, the model benefits from both approaches, improving stability and reducing error for short-term stock forecasting.

- scikit-learn (Linear Regression)
- statsmodels (Exponential Smoothing)
- Pandas and NumPy for feature engineering

IV. RESULTS

In this section, present your findings using an appropriate method, such as equations, numerical summaries, or visualizations like charts and graphs. Clearly explain all results and provide guidance on how to interpret them. If any unexpected results arise, discuss possible reasons or contributing factors. To improve clarity and organization, consider using subsections (e.g., A, B) to separate different aspects of your results.

Example: After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

A. Result A

Example: XXX

1) *For papers with more than six authors:* Add author names horizontally, moving to a third row if needed for more than 8 authors.

2) *For papers with less than six authors:* To change the default, adjust the template as follows.

a) *Selection:* Highlight all author and affiliation lines.

b) *Change number of columns:* Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

c) *Deletion:* Delete the author and affiliation lines for the extra authors.

B. Results B

Example: Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

C. Results C

a) *Positioning Figures and Tables:* Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

TABLE I. TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	Subhead	Subhead
copy	More table copy ^a		

^a Sample of a Table footnote. (Table footnote)

Fig. 1. Example of a figure caption. (figure caption)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

V. DISCUSSION

Every method/project has its shortage or weakness. Please discuss the unsatisfied results in your project. And discuss the feasible suggestions of future work to revise/improve your result.

Example: xxx

VI. CONCLUSION

In this part, you should summarize your project. What important results did you find for your topic and what's the effect of this result on the real-world?

Example: xxx

ACKNOWLEDGMENT (Heading 5)

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Use the IEEE format for the citation. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first...” Unless there are six authors or

more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (*references*)
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, "Title of paper if known," unpublished.

- [5] R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
- [7] M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.

We suggest that you use a text box to insert a graphic (which is ideally a 300 dpi TIFF or EPS file, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on your frame, use the MSWord "Format" pull-down menu, select Text Box > Colors and Lines to choose No Fill and No Line.