Project in Time Series Analysis Course

Project Overview: The final project entails the practical application of time series analysis and forecasting principles to a real-world dataset. With a focus on SARIMA, Prophet, Hidden Markov Models, State Space Models, and Regression with Fourier terms, groups are tasked with

- Selecting a time series dataset
- Conducting a comprehensive analysis
- Presenting their findings in a detailed report

The project's goal is to establish a significant connection between theoretical concepts and empirical data, necessitating a dual thought process that integrates subject understanding with thorough dataset examination.

The primary objective centers around executing a detailed analysis and forecast of time series data. This begins with pivotal steps such as dataset selection and the formulation of a research question, which serves as the project's primary goal for understanding and addressing. The analysis unfolds through stages encompassing data exploration, stationarity testing, and advanced modeling using techniques such as SARIMA and Prophet. The project concludes with a forecasting task utilizing classical and machine learning tools on time series data, followed by a summary of key findings.

Part 1 - Data Selection, Visualization, and Introduction (15% of the Grade)

In this initial phase of the project, each group is tasked with selecting a time series dataset consisting of at least 300 data points. We encourage exploration across diverse domains such as finance, economics, engineering, or social sciences to ensure a broad spectrum of analysis possibilities. It is advisable to choose a dataset that aligns either with the group's existing knowledge or with a topic the group is eager to explore further.

The goal of Part 1 is to conduct a preliminary analysis to identify any underlying seasonality, trends, and patterns within the data, without the application of a specific model. Visualizations are key in this phase, offering insights into the behavior of the data. Creativity in presenting these visualizations is encouraged.

Part 2 - Model Fitting (35% of the Grade)

In Part 2 your goal is to fit models to the selected dataset. Groups are required to:

- **Select at least three methods** for modeling and forecasting, ensuring that one of them is the SARIMA model. If the dataset contains a large number of observations and fitting SARIMA becomes challenging, consider aggregating the data (e.g., converting an hourly series into a daily series by calculating averages).
- For each model, provide a clear explanation of the rationale for fitting the model and selecting its parameters.
- Generate forecasts for future time points.
- Compare the performance of methods, and quality of their forecast, and evaluate their goodness of fit.

Part 3 - Incorporating an Exogenous Variable (15% of the Grade)

In Part 3, introduce an explanatory variable or vector, exogenous to the original time series, which is related to the series subject. Incorporate this variable into the analysis to demonstrate its impact on the primary dataset. For example, if the time series is the level of the Sea of Galilee, you can add a vector of the weather conditions, including temperature and precipitation. This addition should showcase how external factors can influence the behavior of the time series data. Use visualization and modeling to explain the effect. Using an indicator of a change in series as the exogenous variable is strictly prohibited.

Part 4 – Change-Point Detection (10% of the Grade)

Many time-series exhibit changes in distribution due to external factors (e.g., COVID-19) or regulatory changes (e.g., taxation reforms). Is there a natural point, caused by an external factor, in which the time-series may have major change in the distribution? If so, use statistical tools to analyze the effect of the change. If there is no natural time point like that, use statistical tools for checking if there is a change in the distribution of the series over time.

Part 5 – Summary Report (25% of the Grade)

Summarize your results in a final report. The summary report should cover all aspects outlined in the instructions. Key components of the report should include:

- Introduction: Providing background information on the dataset, stating analysis questions, and setting the context for the project.
- Methodology: Justifying the chosen statistical methodologies, explaining model fitting procedures, and detailing parameter selection criteria.
- Results and Discussion: Presenting findings from the analysis, including visualizations, model comparisons, and insights gained.
- Conclusion: Summarize the project's findings
- References: Ensuring scholarly referencing of all sources consulted.

Initial Submission

Submission of a one-page concise report that describes the dataset of both the original series you would like to analyze and the corresponding exogenous series (see Part 3). Illustrate a basic breakdown of the original series data into trend, seasonality, and noise. This should be completed by 11:59 pm on February 4, 2025.

Final Submission Instructions

The submission should include the report in a PDF file, written either in Hebrew (preferred) or English, using a word processor such as Word or LaTeX. Additionally, provide the code used for the analysis and the dataset.

I. Submit the report file in PDF format. The first page should include the names and IDs of the students, along with the respective universities or academic institutions which they are associated with. Use a font size of 11 or 12. The report length can be up to 12 pages but may be shorter.
II. Submit two code files: one in a Jupyter Notebook format and the other in HTML.
III. The dataset, including the exogenous variable, should be included in the submission as a .csv file. Additionally, a link to the original data site should be provided in the report.
IV. Only **one** member of the group should submit the project
V. Anyone with submission extension or submission in a group of three must attach a Miluim/ belonging to group 3 certificate.

Ensure that each of the four files is submitted individually, not in a zipped folder (PDF, IPYNB, HTML, and CSV). Include the IDs of the submitting individuals in the filename of each file. Failure to adhere to this format may result in a deduction of points for the project.

Grading will be based on how effectively you addressed the project requirements, the clarity of the report and visualizations, as well as your level of creativity.

Submission Deadline

The project report and required files must be submitted by 11:59 PM on April 6, 2025.

Possible datasets can be found in:

- https://www.kaggle.com/search?q=time+series+in%3Adatasets
- https://github.com/liaoyuhua/open-time-series-datasets?tab=readme-ov-file
- https://catalog.data.gov/dataset/?tags=time-series&_tags_limit=0
- https://info.data.gov.il/home/
- https://www.cbs.gov.il/he/Statistics/Pages/%d7%9b%d7%9c%d7%99%d7%9d-%d7%95%d7%9e%d7%90%d7%92%d7%a8%d7%99-%d7%a0%d7%aa%d7%95%d7%a0%d7%99%d7%9d.aspx