Case Study

Time Series Analysis of Daily Steps Data

Introduction: In this case study, we will perform a time series analysis on a fitness dataset to gain insights into the patterns and trends in an individual's daily steps over time. The dataset contains information about daily steps taken by an individual and some superfluous data. Our goal is to analyze the data and identify any recurring patterns, seasonal variations, or anomalies that could help optimize the individual's fitness routine.

Dataset: The fitness dataset comprises daily steps measurements for 15 months. It includes the following variables:

1. Date: The date of the measurement.
2. Steps: The number of steps taken by the individual on a given day.
3. Goal: A daily steps goal

You can download the dataset in CSV format from Blackboard. There are 15 different files that need to be analysed.

Objectives:

1. Explore the overall trends in daily steps taken by the individual.
2. Identify any seasonality or recurring patterns in the data.
3. Forecast the time series and test the fitness of it.

Methodology:

1. Data Preparation [20]:
    - Load the dataset into a Python programming environment).
    - Convert the "Date" column to the appropriate date format.
    - Set the index
    - Handle missing values, if any, through imputation or removal.
2. Exploratory Data Analysis (EDA) [10]:
    - Visualize the time series of daily steps to understand the overall trends and patterns.
    - Calculate descriptive statistics such as mean, median, and standard deviation.
3. Test for seasonality and if the data is stationary [20]:
    - Perform a seasonal decomposition of the time series to separate the data into trend, seasonal, and residual components.
    - Plot and interpret the decomposed components to identify any patterns or seasonality.
    - Perform an Augmented Dickey-Fuller test to test if the data is stationary.
4. Model selection [20]:
    - Pick an appropriate time-series analysis model(s).
    - Estimate the model parameters
5. Forecasting [20]:
    - Forecasting should start at 25 June 2023.

- Make a visual comparison of the actual vs the forecasted steps.
- Tests and/or compare different models and see how close your forecast is to the actual data. (e.g., Mean Squared Error, R-squared).

6. Reflection [10]
    - Outline your decision-making process, i.e. why you made the choices you made.
    - Outline what you learned from the data and provide some insights into it.
    - Outline what you will do differently next time.

Please note that all steps should be well documented, and the code should be submitted as a single Jupyter 3 Notebook.