School of Computer Science Engineering and Technology

|  |  |
| --- | --- |
| Course- B**. Tech** | Type- Specialization Elective |
| Course Code**- CSET369** | Course Name- **Time Series Analysis** |
| Year- **3rd Year** | Semester- **V** |
| Date- **Week 2** |  |

**Lab Assignment -2**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Experiment | CO1 | CO2 | CO3 | CO4 |
| Visualization and Statistical Tests for Non-Stationarity |  |  |  |  |

1. You have been provided with a dataset that records various air quality parameters, including the PM2.5 index in the air (https://www.kaggle.com/datasets/fedesoriano/air-quality-data-in-india). Your task is to perform a basic analysis of the PM 2.5 over time such as:

* Visualize PM 2.5 Concentration Over Time:
* Daily Average PM 2.5 Concentration
* Highlight High PM 2.5 Concentration Events (PM 2.5> threshold)

2. Develop a comprehensive set of visualizations to explore and analyze the weather data of Delhi from January 1, 2013, to April 24, 2017 (https://www.kaggle.com/datasets/sumanthvrao/daily-climate-time-series-data). The goal is to uncover trends, seasonal patterns, anomalies, and relationships between different weather parameters, providing a deep understanding of Delhi's climate dynamics.

* Create individual time series plots for mean temperature, humidity, wind speed, and mean pressure to observe trends over the entire date range.
* Plot the time series data to observe its non-stationary nature, identifying trends and seasonality.
* Combine multiple weather parameters (e.g., mean temperature, humidity, wind speed, and mean pressure) in a single time series plot to observe how these variables interact over time. What relationships or simultaneous patterns do you observe?
* Generate histograms for each weather parameter to analyze their distributions. What are the most common values for each parameter, and how frequent are extreme values?
* Plot two related weather parameters (e.g., temperature and humidity) on dual axes to explore how they vary together over time. What insights can you gain from this dual-axis comparison?
* Discuss potential implications of these findings for understanding Delhi's climate and how they could be useful for weather forecasting or climate studies.

3. Download the following time series dataset and read them as a python dataframe and print the heads.

* Air quality Data (<https://www.kaggle.com/datasets/fedesoriano/air-quality-data-in-india)>)

**Perform all the following tasks on the above dataset.**

1. Load a time series dataset containing air quality measurements () for a specific location. Plot the time series data. What initial observations can you make regarding the trends, seasonality, and fluctuations in air quality over time?
2. Conduct an Augmented Dickey-Fuller (ADF) test on the original air quality time series. What do the test statistic and p-value suggest about the stationarity of the data?
3. Conduct a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test on the original air quality time series. What do the test statistic and p-value suggest about the stationarity of the data?