Time Series Project Proposal

Student Name: Ginu Varghese **Student ID:** D00251842

Project Title: Time Series Analysis on Weather data

1. Project Description

Weather refers to the day-to-day atmospheric conditions including the temperature, precipitation, humidity. Most of our daily decisions are impacted by the weather conditions. The first thing most people do when they wake up is to check the weather conditions to plan the rest of the day. In general, the daily weather conditions can affect the feelings and insight of people towards their life and world. So, it is important to analyze and predict the patterns of weather for the easy going of life.

The main components of weather are temperature, wind, humidity, and atmospheric pressure. Temperature refers to how cold or hot the atmosphere is, and the humidity is the amount of water vapor in the atmosphere usually referred as the atmospheric moisture. Atmospheric pressure is the pressure within the Earth's atmosphere. These components can be used to describe weather at any time. This data with change in time can be used by meteorologists to find out the weather conditions and to forecast the weather patterns for the future. It also helps the people to use the daily forecast to plan their day-to-day activities and life according to the weather patterns of the day.

The objective of this project is to study the daily weather patterns of different variables such as temperature, pressure, and humidity from 2009 to 2011 and determine the behavior of data from time to time and to train models on the data and to learn how to forecast the weather data for future.

The weather dataset for the project was downloaded from the Harvard dataverse. The selected dataset was available to the public for download and use with proper citation to the site. The dataset gives the information about the temperature, humidity, and pressure along with the wind direction of the day and the weather condition of the recorded hours in each day from 2009 to 2011.

2. Research questions

1. Which model will give the best r squared value for the available data?
2. How does the window shifting affect the model?
3. Which transform methods will be better for the dataset for better analysis?
4. Which resampling method is suitable for the time series analysis with the available data?

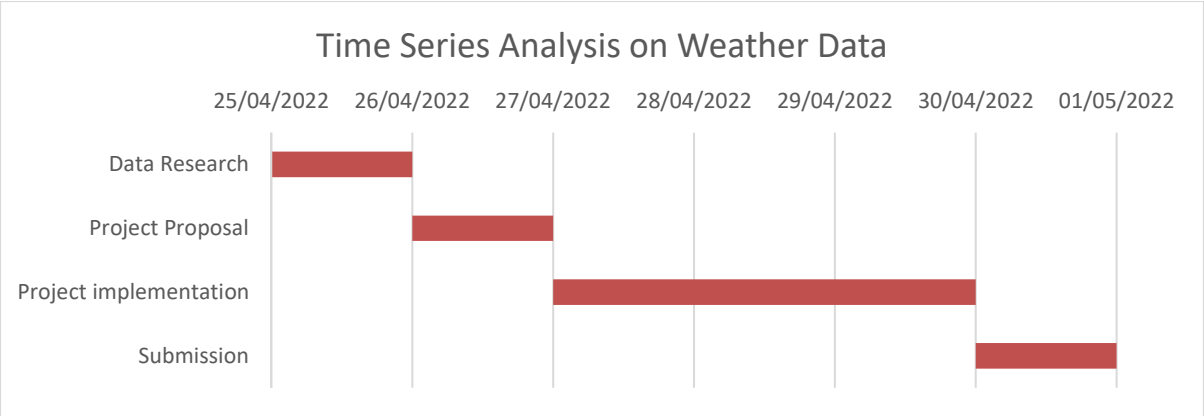
3. Data description, tools and software, references, link to data sources and ethics consideration.

The weather dataset for the project was downloaded from the Harvard Dataverse. The selected dataset was available to the public for download and use with proper citation to the site. The dataset gives the information about the temperature, humidity, and pressure along with the wind direction of the day and the weather condition. The data is recorded at each hour daily from 2009 to 2011. There are 9861 observations and 9 columns in the dataset.

Dataset download link:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2K9FFE

The project will be done using the python libraries and scikit learn libraries in the python. Since the dataset is available to the public for downloading and using with proper citation, there is no other ethical issues with the dataset.

4. **Gantt Chart**



5. **Special requirements and deliverables of the project**

No special requirements are required for the project. The deliverable of the project will be an analysis, its outcomes, and a report on the project.

6. **Risks – What risks can you identify? What will be the impact if the risk becomes a reality? What can you do to minimize the impact?**
    - Loss of data – In order to avoid the loss of data, all the data have been backed up to cloud.
    - Ethics risks – Data is downloaded from a public website and proper citation will be given for using the dataset.
    - Unexpected output - All outputs will be reported.