DAY 6
Time Series

What is a Time Series:

- A dependent variable that is observed at different time intervals usually equal intervals
- Examples
    - U.S.'s GDP over time
    - Customers Reviews during March 2019
    - GPAs of all sophomore students during the summer of '18.
    - Sales and revenue
    - Height of a boy over time.
- Time Series Plots
    - Possible Patterns
        - Horizontal Patterns
            - Where the data seems to fluctuate randomly around a mean
            - An adult's height. (Constant but only may change due to reading error / posture)
        - Trend Pattern
            - When random fluctuation occur, but has a constant trend about it.
            - Eg. Temperature at March (It keep increasing)
            - Types of Trends
                - Uptrend
                - Downtrend
                - Stationary Trend
        - Seasonal Pattern
            - Deals with **seasonality** - short term periodic relationship
            - Identifying a common pattern in the data
            - Seasonal Patterns can exist without trends
                - Eg. Tides, and how far they recede throughout a day.
            - Eg. Low-Med-High is a constant shape shift of the data throughout the months.
            - Common with monthly data
            - Employment graph, Patio Furniture Sales graph
        - Cyclical Patterns
            - Has high runs and low runs (ups and downs)
            - Eg. Employment, Gasoline Prices
        - Irregular Patterns
            - Unobservable, but time dependant
            - E.g. Sales of Medicine due to floods
    - How to Plot
        - Take a variable that you care about
        - Take a series of Y's that are influence by time
    - Terms used in Plotting

        - **Seasonality**
            - Short term periodic relationship
        - **Cycles**
            - Long term periodic relationship
        - **Irregular**
            - Unobservable, but time dependant
            - Eg. Oil spills

What can one achieve with Time-Series:

- Business Forecasting
- Understanding Past Behavior
- Plan the future
- Evaluating a current accomplishment.

When do we not use Time-Series:

- When values are constant.
    - Number of coffee sales are constant across months
- Values are in the form of functions
    - No point as you can use a function to calculate.

What is Stationarity:

- Dataset has a constant mean, one that doesn't vary across time
- Dataset has a constant variance, one that doesn't vary across time
- Autocovariance doesn't depend on time
    - Correlation between values of the different time frames shouldn't be there.
    - E.g. t-1's values and t's values don't have any correlation between them
- When a dataset has these 3 points it is said to be stationary, and then can time-series can be used upon it.

Checking Stationarity in Python:

- Python has two popular tools to check the stationarity of a dataset
- **Rolling Statistics**
    - Plot the moving average, variance and check whether that is moving or not.
    - A moving average is one, that is taken in a specific time window only
    - Visual Technique
        - Not deployable on production, but it is useful for Proof of Concept
- **Augmented Dickey Fuller Test**
    - Statistical Test to check for stationarity
    - Data Science related
    - Null Hypothesis is that the Time Series is non-stationary
    - Result consists of Test Statistics and Critical Values

Models to work with Time Series Data

Arima

- The ARIMA model shows great promise
- ARIMA = AR + I + MA
    - **AR -** Auto Regression
        - Describes co-relation between previous time period and current time period
    - **I -** Integration
    - **MA -** Moving Average
        - The noise in the time series is averaged out and smoothened out and calculated.
        - The crest / throughs of noise are smoothened out to bring an average forecast
- **Parameters of ARIMA (to predict)**
    - P
        - Autoregressive lags
        - Prediction done using PACR graph (Partial Autocorellation)
    - Q
        - Moving average
        - Prediction done using ACF Plot (Auto Correlation Plot)
    - D
        - Order of differentiation
            - If the order of integration is 1, then D = 1
            - To make data stationary, we use differentiation