Lab 5 – SECTION A , BATCH 1 Date: 6th Dec 2021

Exer 1: Time Series Analysis

Use the "employment.csv" data set and perform time series analysis and visualization through the following questions.

1. Convert datestamp column to a datetime object and Set the datestamp columns as the index of your DataFrame. Check if there are missing values in each column.
2. Generate a boxplot to find the distribution of unemployment rate for every industry .
3. Using line chart Visualize the unemployment rate of workers by industry .
4. Plot the monthly and yearly trends .
5. Apply time series decomposition to your dataset to visualize the trend and seasonality .
6. Visualize the seasonality of Agriculture, Health and Finance sector.
7. Visualize the seasonality of multiple time series and the correlation between each time series in the dataset.

Exer 2: Association Rule Mining

Use the "groceries.csv" dataset and answer the following:

1. How many transactions and items are there in the data set?
2. Prepare the data for finding association rules. Each transaction will contain a list of item in the transaction.
   [['citrus fruit', 'semi-finished bread', 'margarine', 'ready soups'],
   ['tropical fruit', 'yogurt', 'coffee'],…….
   ['whole milk']]

3. Use Python library *mlxtend* and convert the transactions into a format that can be used in the Apriori method for finding frequent itemsets.
   pip install mlxtend
   from mlxtend.preprocessing import TransactionEncoder
   from mlxtend.frequent_patterns import apriori, association_rules
4. Find top selling items with minimum support of 2%.
5. Find all frequent itemsets with minimum support of 5%.
6. Find all frequent itemsets of length 2 with minimum support of 2%.
7. Find the top 10 association rules with minimum support of 2%, sorted by confidence in descending order.
8. Find association rules with minimum support of 2% and lift of more than 1.0.