

DSE 2141– Data Analytics Lab
Lab 7 – Date: 20th September 2023

EXERCISE 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.

EXERCISE 2 : Text Analysis

Download the amazon_baby.zip file and answer the following:

1. Check the number of the reviews received for each product.
2. Check the products that have more than 15 reviews.
3. Find any missing review are present or not, If present remove those data.
4. Clean the data and remove the special characters and replace the contractions with its expansion by converting the uppercase character to lower case. Also, remove the punctuations.
5. Add the Polarity, length of the review, the word count and average word length of each review.
6. Visualize the distribution of the word count, review length, and polarity.
7. Visualize polarity considering the rating.
8. Visualize the count of the reviews of each rating available in the dataset.
9. List the Top 20 products based on the polarity.
10. Visualize to check whether the review length changes with rating.
11. Visualize the distribution of Top 25 Unigram, Bigram and Trigram.