

## IE 360 Statistical Forecasting and Time Series

### Homework 1, due November 20<sup>th</sup>, 2020

**Instructions:** Please solve the following exercises using R (<http://www.r-project.org/>) or Python (<https://www.python.org/>). You are expected to use GitHub Classroom and present your work as an html file (i.e. web page) on your progress journals. There are alternative ways to generate an html page for you work:

- A Jupyter Notebook including your codes and comments. This works for R and Python, to enable using R scripts in notebooks, please check:
  - <https://docs.anaconda.com/anaconda/navigator/tutorials/r-lang/>
  - <https://medium.com/@kyleake/how-to-install-r-in-jupyter-with-irkernel-in-3-steps-917519326e41>

Things are little easier if you install Anaconda (<https://www.anaconda.com/>). Please export your work to an html file. Please provide your \*.ipynb file in your repository and a link to this file in your html report will help us a lot.

- A Markdown html document. This can be created using RMarkdown for R and Python-Markdown for Python

Note that html pages are just to describe how you approach to the exercises in the homework. They should include your codes. You are also required to provide your R/Python codes separately in the repository so that anybody can run it with minimal change in the code. This can be presented as the script file itself or your notebook file (the one with \*.ipynb file extension).

The last and the most important thing to mention is that academic integrity is expected! Do not share your code (except the one in your progress journals). You are always free to discuss about tasks but your work must be implemented by yourself. As a fundamental principle for any educational institution, academic integrity is highly valued and seriously regarded at Boğaziçi University.

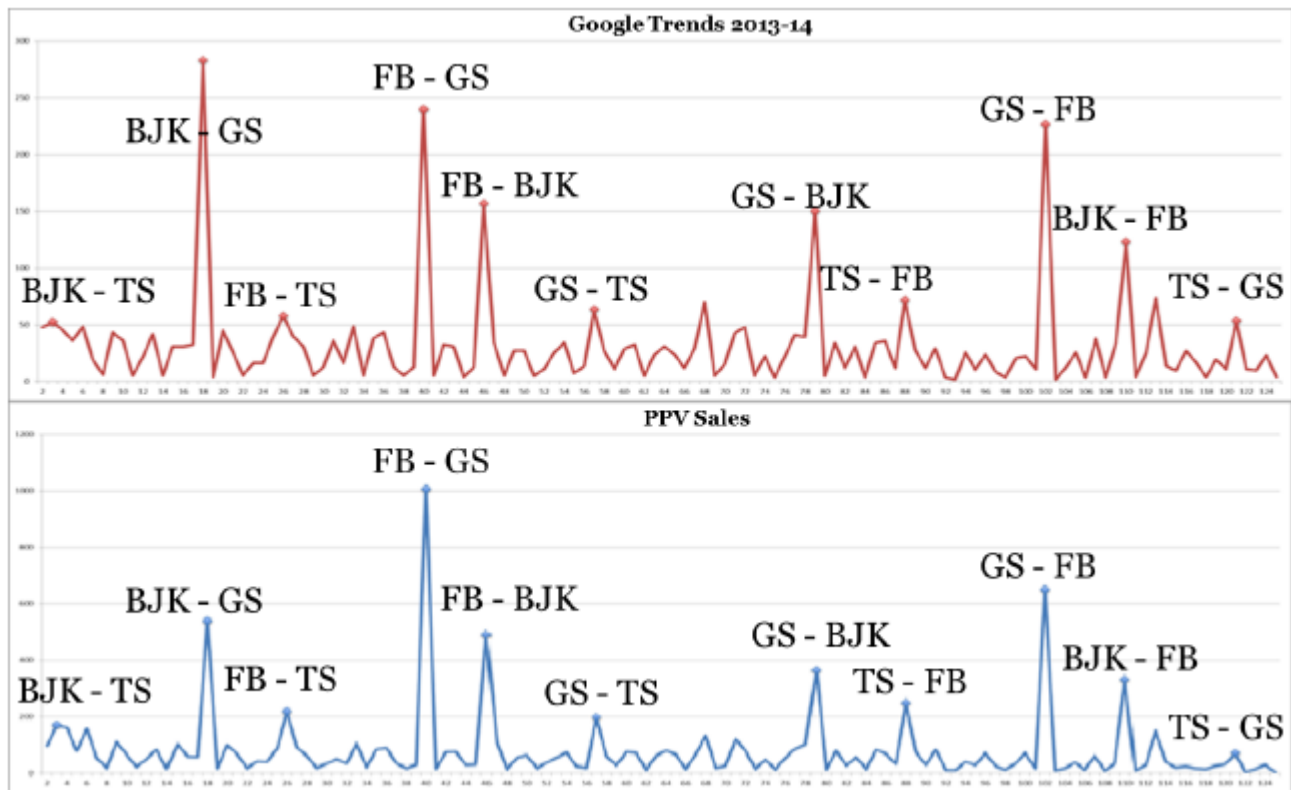
### Exercise: Google Trends

In this exercise, you are expected to obtain a dataset in which you are interested and provide plots to understand the distribution of the data. The data to be worked on should be related to Turkey. Some of the potential resources are made available on the following link: [http://r338.github.io/ab-2016/00\\_verikaynaklari/](http://r338.github.io/ab-2016/00_verikaynaklari/)

Another interesting data source is Google Trends (<http://trends.google.com/>). The Wikipedia definition for Google Trends is:

“Google Trends is a public web facility of Google Inc., based on Google Search that shows how often a particular search-term is entered relative to the total search-volume across various regions of the world, and in various languages. The horizontal axis of the main graph represents time (starting from 2004), and the vertical is how often a term is searched for relative to the total number of searches, globally.[1] Below the main graph, popularity is broken down by countries, regions, cities and language. Note that what Google calls "language", however, does not display the relative results of searches in different languages for the same term(s). It only displays the relative combined search volumes from all countries that share a particular language (see "flowers" vs "fleurs"). It is possible to refine the main graph by region and time period.”

Sometimes it is interesting to check if search volume of a certain keyword has some relation to an event. For example, number of searches for “flu” can give you an idea about the number of people affected by the flu over the world. Another example is from a senior design project we have worked on last year. It was about understanding if Google Trends provide information about the number of pay-per-view (PPV) customers of Digiturk. To be more specific, Lig TV is a football channel of Digiturk holding rights to broadcast Turkish Super League live since 2001. It serves its customer through different channels such as the Internet platform “Digiturkplay” or satellite TV “Digiturk Plus”. All channels include a service called pay-per-view (PPV) providing customer to pay for the particular programs that they watch via private telecast. The aim here is to see if it is possible to forecast demand for PPV through the Internet search data, namely Google Trends Data. To achieve this, we obtained PPV data from Digiturk. For each soccer game, we obtained the daily search volume information for each team playing a game (i.e. if the game is between Beşiktaş and Fenerbahçe, we obtained the search volume for each team on the game day and sum these values to find out a measure that will quantify the popularity). The plot of the PPV sales versus the search volume information for each game is schematized in Figure 1. As you can see, Google Trends data is highly correlated with the sales.



**Figure 1.** Search volume versus PPV sales

Please perform the following tasks:

- Please select 3 different measures of interest from the databases in the link, [http://r338.github.io/ab-2016/00\\_verikaynaklari/](http://r338.github.io/ab-2016/00_verikaynaklari/), (i.e. unemployment rate over years, stock prices of a certain company). Any other source would be fine.
- Plot the histogram of the measures for each year (i.e. you can obtain the unemployment rate of each city for different years). Check if the shape of histogram is changing significantly over the years (visual inspection will be enough).

- c) Determine a search keyword that can be relevant to the measures you consider in the earlier parts (i.e. if you are interested in unemployment rate, you can check the search volume for “iş ilanı”). Plot the boxplot of the search volume and boxplot of the data you obtained for each year. Note that you need to provide these plots for each measure of interest. Is there any trend you can observe?

**Important Note:** Although time frame is given as years in the questions, you can use any time frame you want. For example you may be interested in intraday prices of Garanti Bankası stocks. In that case you will have price information for each day. Then, you can compare the daily distributions in part (b) for five days and your keyword can be “GARAN” (abbreviation for Garanti Bankası in Borsa İstanbul) for Google Trends data in part (c). You can obtain the intraday search volume for “GARAN” and draw the boxplots accordingly.