CE/CZ4073 : Data Science for Business

2017-2018, Semester 2 | Nanyang Technological University, Singapore

Assignment 4

Posted on 9 April 2018 · Clarify by 17 April 2018 · Submit by 23 April 2018

Submission : Submit the `R Markdown` and `R Shiny` files for the assignment within a single compressed folder – `assign4_FullName.zip` – containing all relevant details/items in connection with the two problems, where `FullName` is your full name, as per the official records at NTU.

Your submission should be emailed to `sg.sourav@ntu.edu.sg` by midnight of 23 April 2018 from your official NTU Email Address (do not use your personal email address). Late submissions will be penalized with lower grades, and submissions after 27 April 2018 will not be graded.

Problem 1 [10 points]

Background : The dataset `https://www.kaggle.com/jessevent/all-crypto-currencies/` on Kaggle records the historical (daily) financial indices corresponding to all cryptocurrencies.

Task : Download the dataset `https://www.kaggle.com/jessevent/all-crypto-currencies/` from Kaggle, extract the following fields/columns for just the following cryptocurrencies (top five as per `https://coinmarketcap.com/`), and convert the data to suitable time series format(s).

```
Fields / Columns : "symbol", "date", "open", "close"
Cryptocurrencies : "BTC", "ETH", "XRP", "BCH", "LTC"
```

Perform time-series modelling of `"open"` and `"close"` prices of the chosen five cryptocurrencies during 2017-04-01 to 2018-03-31 using suitable choice(s) of ARIMA model. Predict the `"open"` and `"close"` prices of the cryptocurrencies for the period 2018-04-01 to 2018-04-15. Present the complete time-series analysis, including comments on the choices you made for ARIMA, in the form of a single `R Markdown` document, ready to be `knit` into a corresponding `HTML` website.

Problem 2 [10 points]

Background: Quandl (`https://www.quandl.com/`) is one of the largest sources for economic and financial data. World Development Indicators dataset from the World Bank is an interesting one – `https://www.quandl.com/data/WWDI-World-Bank-World-Development-Indicators` – that tells the tales of countries in terms of the historical values of their development indicators.

Task : Build an `R Shiny` app that "compares" the historical timeline of Singapore with that of *five* other countries of your choice, during 1966–2015, in terms of any *five* indicators of your choice from the WWDI dataset. The user should be able to choose the `"country"` and the `"indicator"` to compare. Your app should *forecast* for the years 2016–2017 in each such case.

This is an individual assignment. Properly acknowledge every source of information that you refer to, including discussions with your fellow students, if any. Verbatim copy from any source is strongly discouraged, and plagiarism will be heavily penalized. It is strongly recommended that you write the codes completely on your own. Please write these codes in R and not in Python.