21st October 2020

TEST 1
ESTIMATING PUBLISHING DATES

1. BASIC INFORMATION

The relevant indices from a given portfolio and their respective publication dates from inception to 31 December 2016 are recorded on the file 'Indices_Publication_Dates.csv'.

The first row of the file contains the codes for each of the indices, whereas the remaining rows contain the publication dates in 'dd/mm/yyyy' format. These are aligned by column to the indices.

Each index - 40 in total - is defined by its primary market location, currency and tenor. However, these attributes may be only partially reflected within the codes as each index was partially encrypted from its original name.

2. YOUR TASKS

Using solely the data provided above, your tasks are:

A. To develop a script in Python 3 that can be used to project/predict the future publishing dates for each index

  i. As a minimum, the model must produce the following outputs
      1. *A file structured in a similar manner to the input file above containing all predicted publication dates for the years 2017 to 2023*
      2. *Two complimentary files containing (a) the weekend dates and (b) holiday dates associated with each of the indices for the years 2017 to 2023. Differently from item 1, whereas there is full flexibility with regards to the structure of the files, these must be developed with two specific goals in mind:*
          a. *Files to be optimised to minimal sizes*
          b. *To avoid/reduce data recording redundancies*
          c. *Files can be easily read in other languages such as Javascript, Ruby, etc.*

B. To propose and develop a classifier/clustering algorithm in Python 3 that places the indices into different families based on their patterns and similarities.

  i. A simple txt file containing the proposed family reference and index should be produced as output.

C. To develop a short model specification and analysis document for the related programmes and data above (1, 2A and 2B)

   i. Whereas the structure of the document is adaptable, it must contain the minimum elements (not necessarily in the same order or grouping) below:

1. *Scope of application and limitations[1]*
   a. *Potential data issues*
   b. *Data quality assessment for each index*
   c. *Proposed indices within and out of scope*
   d. *Accuracy and error expectations[2]*

2. *Model Structure*
   a. *Overview*
   b. *Schematic logic flow*
   c. *Inputs and outputs*
   d. *Description of key objects, functions, variables and attributes*

3. *Methods / Methodology*
   a. *Description of the methods and methodology used*
   b. *Supporting equations*
   c. *Brief and basic academic and/or industrial references*

4. *Brief user guide / instructions*

3. GENERAL GUIDANCE

A. On the scripts

   i. A single script is to be produced for both tasks A and B.

   ii. The script is to be recorded in a .py extension file which can be opened and run in any shell. (i.e. no Jupyter notebook).

   iii. Pre-existing basic modules and/or packages may be used (such as datetime, os, csv, matplotlib, etc.). However, these should be avoided if not strictly effective or necessary. Specialised modules should also be avoided.

   iv. The script should contain all necessary comments to facilitate understanding, but verbose is to be avoided. (i.e. Be short and sharp).

   v. Readability is more important than technical coding elegancy (for example, multi-layered nesting of list comprehensions should be avoided).

---

[1] A limited number of graphs or tables is expected to be produced to illustrate rationale and to support conclusions
[2] Description, including speculation, of the potential factors driving accuracy and errors is encouraged.

vi. Use of effective structures such as functions, classes, dictionaries, frozensets, namedtuples, etc. are encouraged. However, these need to be applied consistently and most effectively.

vii. There should be consistency in naming variables, as well as good logic and convention to the names, including the use of upper/lower cases.

B. On the model specification and analysis document

i. There are no limits to the size of the document, but one would expect to be between 7-15 pages long.

ii. It is important to be succinct and clear, avoiding long descriptions.

iii. Graphs and tables are encouraged to facilitate reading but avoiding overuse.

iv. Equations (or group of) should be cross-referenced to the location within the script.

v. Avoid too many colours. Focus on content and clean style rather than aesthetics.

4. TIMING AND DELIVERABLES

The script, output files and document should be electronically delivered two days before the deadline. The original input file does not need to be returned.