Rising Stars – Benchmarking Model

Python programs manual

# Introduction:

In this document we will introduce the 4 python programs used for benchmarking model, namely,

* *Convert\_publication\_date.py*:
  + Standardize/convert “Publication Date” variable, a string type variable found in “Oncology Profiles – Full pubs.csv”, provided by Yves and his team
* *Estimate\_age.py*:
  + Estimate physicians age based on “profile\_education.csv”, a csv version of “profile\_education.txt”, provided by Yves and his team
* *Accumulative\_number\_of\_publication.py*:
  + Compile accumulative number of publication for each physician from their first year of publication
* *Benchmarking\_model.py*
  + Identify rising stars by comparing candidate’s accumulative number of publication to benchmark

# The Model:

The way benchmarking works are as follows:

1. Determine the number of years, take 15 for example. *Accumulative\_number\_of\_publication.py* will generate track record in terms of accumulative number of publication for all physicians for their first 15 years since their first year of publication
2. Order KOL by the number of publication published during their first 15 years since fir year of publication
3. Ping down a group of bottom 30% KOL from step2
4. Calculate average accumulative number of publication published, for group from step2, during their first to first 15 years since the first year of publication, the result is the benchmark
5. Compare nonKOL with benchmark, if x% (x is a user defined number between 0 and 1) of years an nonKOL outperformed benchmark, and he/she is younger than y-year-old (y is a user defined number used as age cutoff), he/she is a rising star

# How to run the python programs:

A specific sequence must be followed: please make sure change xxxxxx to your designated directory and execute each line of code

1. Open *convert\_publication\_date.py* in any python IDE (spyder, notebook), run through each line of code, and the end, “Oncology Profiles – Full pubs.csv”, the input, will have one extra column named “Year”, denoting the year in which the paper is published
2. Open *estimate\_age.py*, it takes “profile\_education.csv”, and maps out estimated age for each physician
3. Open *accumulative\_number\_of\_publication.py*, it takes the output of *convert\_publication\_date.py and estimate\_age.py*, outputs accumulative number of publication since the first year of publication
4. Open *benchmarking\_model.py*, input the product of *accumulative\_number\_of\_publication.py*, outputs list of physicians marked as rising star