***Project\_1: Financial Data Exporting to CSV***

* 이번 주는 토요일까지 Break 이기 때문에, 시간적인 여유가 있을 것으로 생각되어, Financial Data Handling Project를 시작으로 긴 여정을 시작하는 것이 좋을 것 같다는 생각이 듭니다.
* 개인적인 생각으로는 첫 번째 목표 (i. Data 수집, ii. Data 가공, iii. Strategy 검정 iv) GUI 개발)의 target date는 7월 말로 설정하고, 8월 말부터는 Data Science에서 배우는 ML & DL을 적용하는 것이 좋다고 생각됩니다만, 다른 분들의 의견이 있으시면, 알려주시면 감사하겠습니다.
* 이번 주에는 기초인 Data Handling Project를 시작하겠습니다.
  + Due Date: 2018/06/08 22:00
  + Project 목표
    - Sise2local.ALL\_CMDT.0213을 활용하여, 각자가 원하는 format으로 csv file을 만들 수 있는 능력을 만들기 🡺 Data exporting to csv file.
    - Why csv file로 만드는 가?
      * 우리는 어떠한 명제 (strategy)를 만들고, 이러한 명제가 과거 Historical Data를 바탕으로 된다/안된다를 검정해야 한다 -> Quant의 Main Task 중 하나
      * Raw Data는 하루치 데이터를 쌓으면, 100 MB가 넘어간다, 지금 우리가 보는 Sise2\_local.ALL\_CMDT파일은 Commodity Product (상품선물) Data만 저장한 것이고, Commodity의 Data건수는 파생상품의 Main Product인 Kospi200지수 선물 또는 주식선물에 비하면, 엄청 작은 Size이다.
      * 따라서, 우리가 어떠한 명제를 검정하기 위해서, 그 명제의 검정에 필요한 데이터만을 가져와서 새로운 파일을 저장하고, 그 새로운 파일을 읽어가면서 그 명제를 검증해야 한다.
      * 그렇지 않고, raw data file을 명제의 검정 단계에서 사용하게 된다면, 1년치 data를 읽힐 때, 비효율적으로 서버를 혹사 시키는 행위가 될 것이다.
      * 그렇기 때문에, 우리는 data\_transformed\_KM3.csv파일과 같이 raw data에서 우리가 원하는 대로 customizing을 할 수 있어야 한다. 따라서, 아래의 방법을 실습해보자.
    - I) Sise2local.ALL\_CMDT.0213 file에서 3년 국채 (K165N3000) code를 찾는다.  
      II) 3년 국채의 최근월물 (BM3) code 를 찾는다.  
      III) (K165N3000 & BM3 & A3016 모두 포함) 또는 (165N3000 & BM3 & G7016를 모두 포함)하고 있는 row만 순서대로 filtering하여, 새로운 csv file을 생성하여, 첫 row에는 “시간,현재가,체결수량,매도1호가,매수1호가,최종매도매수,총잔량매도,총잔량매수”를 write하고, 그 아래 row부터, filtering된 A3016및 G7016이 순서대로 write되는 code를 만들어 보기.
    - File Name: 165N3000\_20180213.csv
      * 165N3000: No Hard Coding, read product name data from the raw file.
      * 02180213: No Hard Coding, read date data from the raw file.
* 최종 결과물은, data\_transformed\_KM3.csv와 완벽히 똑같이 나와야 합니다.
* 이번 Project를 하는 Main 목적은, Financial Data에 익숙해지는 것이므로 각자 Project로 진행하여, 각자 Data와 친숙해지는 시간을 갖는 것이 필요할 것 같습니다.
* 다음 Project는 개별로 할 것인지, Group으로 할 것인지 face to face로 이야기 하면서 결정하도록 하겠습니다.
* 저는, 최근에 이번 Project를 목표로 만든 것이 있기에, Pass하고, 제가 만들었던 것을 Google Drive에 올려두도록 하겠습니다. 참고용으로 보시라고 올려두겠지만, 사실 저는 아예 보지 않고, 스스로 처음부터 끝까지 생각하셔서 code를 짜시는 것을 강력히 추천 드리도록 하겠습니다. 자신이 짠 code를 실행했을 때, data\_transformed\_KM3.csv와 똑같은 내용이 출력이 된 것을 확인 후, 각자의 code를 비교해 보는 시간을 갖는 것이 Best 할 것 같다는 생각입니다.
* 한가지 확실한 것은, 모든 분들께서 Data를 처음 접해보시기에, 감이 안 잡히시거나 하다가 이해가 안 되는 부분이 무조건 생길 것이라고 생각합니다.
* 저는 목요일, 금요일: 08:30 ~ 22:00까지 강의실에 있을 것입니다. 궁금한 점이 있으시면 오셔서 물어보셔도 되고, Kakao Talk 또는 Slack으로 언제든지 연락 주십시오.