**Description of Toy – Project**   
  
**[ “Toy\_data.csv” ]**는 각 국가의 주식지수, 원자재 등 **Macro 지표**이다.   
해당 데이터를 Input data로 활용하며,   
**[ “Toy\_data\_target.csv” ]**내 자산의 횡적 리스크 모델 중 1개를 활용한   
자산의 최적 비중을 Target으로 학습/예측을 하고 해당 Portfolio의 성과를 분석한다.

건물, 실외, 대지, 파란색이(가) 표시된 사진

자동 생성된 설명**- Constraints**  
〮 횡적 리스크 모델 : MVO(Sharpe Maximized Portfolio)  
〮 매월 말까지 데이터로 매월 초 Rebalancing, monthly 예측  
〮 Return, volatility, Cov-matrix 등 추정 값은 향후 1개월 값을 활용(1개월 forwardlooking)  
〮 거래비용, Turnover ratio 무시  
〮 학습모델 : 딥러닝 모델 (CNN, GRU, Transformer 또는 SOTA 딥러닝 모델)  
 (Target data가 시계열 특징을 지니고 있는지 확인하고, 특징에 맞는 모델 사용)  
〮 학습/테스트 : 학습 70%, 이후 one-step ahead cross-validation (testing)  
 (학습에 사용할 데이터의 시작 Date는 자율)  
  
〮 성과분석은 Test 구간에서의 포트폴리오 운용 성과 분석, 모델의 예측 성과 분석  
  
**- Python 사용, 이외 기재되어 있지 않은 [가정, 제약]은 자율   
- 제출 및 발표**   
 〮해당 프로젝트의 Python source code 제출, 모델링/성과 방법 설명자료 준비 (자유 양식)  
 〮포트폴리오 운용성과의 우월성을 평가하지 않음.  
 〮발표일 (최종 면접 일) 전날까지 제출  
 〮제출 및 기타 문의사항 : [taehee.cho@qraftec.com](mailto:taehee.cho@qraftec.com)

과제 내용에 대해 부연 설명드리면,

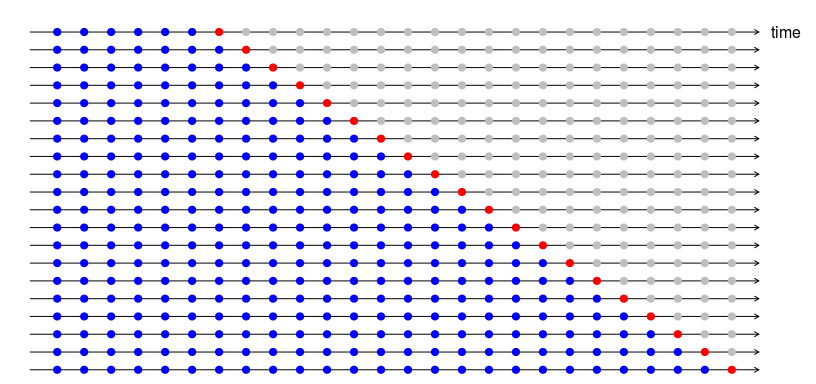
1. 이상적인 MVO 포트폴리오를 만들고, 포트폴리오 비중을 딥러닝 모델의 Label로 사용합니다.
   1. 이상적인 MVO 포트폴리오를 만들기 위해 미래 1개월 데이터 사용 가능하며, 과거 데이터는 모두 사용 가능합니다.

〮 각 자산 별 최대비중 : 25%, 최소비중 : 0% ( long only : 사는 것만)

* 1. 적절한 구간을 정하여 Return, Volatility, Covariance 를 계산하고, 최적의 비중을 결정합니다.

1. 딥러닝 모델 Input 을 정의하기 위해 EDA 를 수행합니다.

〮 학습 전 각 Features, Target의 data EDA 필수  
 (필요시 결측치 보간, 이상치 제거, 수치 변형, 다중공선성제거, feature engineering 등 적용)

1. [2] 에서 얻은 Input 과 [1] 에서 얻은 Label 을 사용하는 딥러닝 모델을 구축합니다.
2. One-step Ahead Cross-Validation (참고: <https://otexts.com/fpp3/tscv.html>) 으로 딥러닝을 평가합니다.
   1. 
   2. Train set은 누적되게 해야 되나?
3. 모델 평가 시에 Loss등 메트릭만 확인하기 보다는 특정 구간 별로 어떤 점 때문에 잘 됐고, 어떤 점 때문에 안 되었는지 확인해주신다면 좋습니다.

I. NA 값 처리 (Input 에 대해서)

