python 데이터 분석 실무(<https://wikidocs.net/book/1867>)

**프로세스 이해**

일반적인 실무에서의 분석 프로세스는 아래와 같다. 상황에 따라 특정 과정이 생략되거나 반복되는 경우가 빈번하다. 실제 분석가는 대부분의 시간을 **전처리와 데이터 이해/탐색에 대부분의 시간을 사용**한다. 복잡한 모델링이나 최적화된 파라메터 튜닝, 멋진 시각화 작업과 같은 업무는 극히 일부만 차지할 뿐이다.

* 문제 정의 및 가설 수립
* 데이터 설계/수집/처리/검수
* 데이터 추출 및 전처리
* 데이터 탐색 및 모델링 (평가, 모델 개선)
* 보고서 작성 및 커뮤니케이션

### 소프트 스킬의 중요성 이해

위와 같은 기술적 내용과 더불어, 아래 나열한 소프트 스킬을 제고하여 분석 결과의 전달력과 효과성을 높이는 것 역시 중요하다. 자동차 엔진으로 비유하자면, 위와 같은 기술적 내용은 엔진의 기계 파트라고 할수 있다. 이러한 기계 파트가 유연하게 작동하기 위해 엔진오일과 냉각수 등이 필요한 것처럼 아래 소프트 스킬이 그러한 역할을 한다.

* 데이터 분석가로서 마인드셋/태도/원칙
* 커뮤니케이션 방식 / 보고서 작성법
* 비즈니스 문제 정의/해결 방안 고민
* 비즈니스 본질 및 주요 고객에 대한 이해

<기초 지식>

* Python(for문, if문, lambda함수, list comprehension, slicing 등)을 통해 원하는 형태의 데이터셋을 만들 수 있음
* Numpy, Pandas, Scipy, Statsmodel, Scikit-learn, Matplotlib 등의 라이브러리 기본 사용법을 알고 있음
* 통계 및 머신러닝에 대한 기초 지식을 보유하고 있음
  + ANOVA, T-test 등 통계적 가설 검정 방법 및 해석 방법, 개념 등에 대해 이해함
  + Ridge/LASSO Regression, Logistic Regression, RandomForest, SVM, ANN 등 머신러닝 모델을 이해함
  + Confusion Matrix, AUC, R-sqaured, RMSE 등 모델 평가 방법을 알고 있음
  + Grid Search 및 K-fold CV 등을 통해 모델의 파라메터 튜닝 방법을 알고 있음
  + 표준화, 로그 transformation, Polynomial/Interaction항 등 주요 Feature Engineering 방법을 알고 있음
* SQL 코드를 작성하여 db에서 원하는 테이블을 추출/조인할 수 있음
* 위의 기초 능력을 3개월동안 갖추고, 예제를 통해 증명해내기!

7/19) 점프 투 파이썬 => Python(for문, if문, lambda함수, list comprehension, slicing 등)을 통해 원하는 형태의 데이터셋을 만들 수 있음