**4과목 1절**

1절 데이터 분석 기법의 이해

1. 데이터 처리

* 대기업은 DW, DM 을 통해 분석 데이터를 가져와서 사용
* DW에 포함되지 못한 자료의 경우 기존 운영시스템 Legacy, Staging 영역, ODS (operational Data store) 에서 데이터를 가져와서 DW에서 결합

1. 시각화 : 효율적으로 분석, SNA분석 (사회연결망분석) 할 때 주로 사용
2. 공간분석 (GIS) : 공간분석은 공간적 차원과 관련된 속성들을 시각화 하는 분석. 지도위에 관련 속성들을 생성하고, 크기, 모양, 선 굵기 등으로 구분하여 인사이트를 얻는다
3. 탐색적 자료 분석 (EDA)  
   개요 : 다양한 차원과 값을 조합해 가며 특이한 점이나 의미 있는 사실을 도출하고 분석의 최종 목적을 달성해가는 과정. 데이터의 특징과 내재하는 구조적 관계를 알아내기 위한 기법의 통칭.   
   EDA 4가지 주제 : 저항성의 강조, 잔차계산, 자료변수의 재표현, 그래프를 통한 현시성
4. 통계분석   
   기술통계 : 모집단으로부터 표본을 추출하고 표본이 가지고 있는 정보를 쉽게 파악 할 수 있도록 데이터를 정리하거나 요약하기 위해 하나의 숫자 또는 그래프의 형태로 표현하는 절차  
   추론 통계 : (추측) 통계, 모집단으로부터 추출된 표본의 표본 통계량으로부터 모집단의 특성인 모수에 관한 통계적으로 추론하는 절차
5. 데이터 마이닝  
   개요 : 대표적인 고급 데이터 분석법, 대용량의 자료로부터 정보를 요약하고 미래에 대한 예측을 목표로 자료에 존재하는 관계, 패턴, 규칙 등을 탐색하고 이를 모형화 함으로써 이전에 알려지지 않은 유용한 지식을 추출하는 방법   
   방법론 : 데이터베이스에서 지식 탐색, 기계학습, 패턴 인식 (장바구니 분석, 연관규칙)