문항 1. 데이터 분석과 빅데이터 분석의 차이점에 4가지에 대해서 설명하시오.

1. 사후처리와사후처리

정보관리시스템을이용해데이터분석을수행하기위해서데이터에대해서사전처리가필요했습니다. 여기서사전처리란필요한정보만추려내고필요없는정보는제거해서데이터를축소하는것을말합니다. 오늘날에는빅데이터처리기술이등장함에따라사전처리보다사후처리가더욱중요해졌습니다. 데이터를저장하여모아놓고데이터에서직접의미를찾아내기위해서무엇을해야할까요? 데이터유형에따라저장방법을결정하는저장관리, 신뢰성있는데이터를확보하는품질관리, 데이터의유출방지와안전한사용을보장하는보안관리등의역할의수행이중요해진것입니다.

1. 표본조사와전수조사

통계분석과같은전통적인데이터분석법도표본데이터에서의미를찾는것에초점을두고발전해왔습니다. 왜냐하면수집, 처리비용, 분석능력등의부담이존재했기때문입니다. 하지만이제조사대상이되는자료전체를빠짐없이조사하는, 전수조사가능해졌습니다. 전수조사의장점은표본조사가주지못하는패턴과같은정보를제공한다는점입니다.

1. 양과질

세번째는질과양입니다. 데이터의질보다양이더강조되는시대입니다. 빅데이터란말에도나타나듯이데이터의양이크지않다면빅데이터가아닐것입니다. 실시간으로분석할수있는데이터의양이증가하면사소한몇몇오류는분석결과에큰영향을미치지않습니다.분석할수있는데이터가많으면많을수록결과의정확성은증가하게됩니다.

1. 인과관계와상관관계

기존의데이터분석은이론적인틀과정해진목적에따라서만진행되었습니다. 하지만비즈니스의상황에서중요한것은인과관계보다도상관관계파악입니다. 빅데이터분석을통해다양한상관관계를빠르게저렴하게도출할수있게되었다고말할수있습니다.

## 문항2. 데이터 전처리 기술 3가지가 무엇인지 명칭을 작성하고, 그 개념에 대해 설명하시오.

1. 데이터 클리닝 : 결측치 대체, 잡음 데이터 평활, 이상치의 확인 및 제거. 불일치 해결
2. 데이터 통합 : 다양한 로그파일 및 데이터베이스의 통합, 일관성 있는 데이터 형태로 변환
3. 데이터 변환 : 정규화, 집합화, 요약, 계층생성
4. 데이터 축소 : 축소된 데이터도 원래 데이터와 같은 분석 결과를 얻을수 있어야 함. 컴퓨팅 시간 등 고려 위해 데이터 축소가 필요.