引言

数据库系统，关系数据库

数据仓库：一个从多个数据源收集的信息存储库，存放在一致的模式下，并且通常驻留在单个站点上。数据仓库通过数据清理、数据变换、数据集成、数据装入和定期刷新来构造。

预测分析，分类：决策树、神经网络、相关分析、聚类、离群点分析

统计学：统计假设检验。直方图，散点图。

中心趋势度量：均值、中位数、众数

相似性度量。

数据预处理

数据清理：缺失值。使用最可能的值填充缺失值。

数据集成

数据规约（编码）

数据压缩

数据规范化。

关联和相关性