**数据预处理的流程**

数据获取([数据获取](note://WEBc2eb11e0ad924d55bcb90ff0f2351410))

数据清洗([数据清洗](note://WEBcdce726a62820d20dd75631325ef63d6))

数据集成([数据集成](note://WEBd3523b884ebbe07e57f9ddf8690b7b39))

数据变换([数据变换](note://WEB59beb4e1103f5672722d2b21c207406f))

数据规约([数据规约](note://WEBa05c34aaed72ccd96090958d9e351fe1))

数据转换([数据转换](note://WEB2a28e6b7818bf89a3b19e8f68d890c6f))

数据统计([数据统计](note://WEBfc1199a51c475ca5c56a9b878b81367a))

**数据问题**

数据缺失(属性值为空)

数据重复(同一条数据出现多次)

数据异常(个别数据远离数据集)

数据冗余(数据中有无意义的部分)

数据值冲突(同一属性存在不同值类型)

数据噪音(属性值不符合常理)