## 制定数据清洗规范（2分）

（1）**缺失值处理**：对数值型数据采用均值/中位数填补或前后向填充；对分类变量使用众数填补；缺失率过高的特征可直接删除。  
（2）**异常值处理**：根据业务规则或统计方法（如箱线图 1.5IQR、Z-score > 3）标记异常值，并选择删除或替换为合理值。

## 制定特征工程规范（3分）

（3）**特征标准化/归一化**：对数值特征进行 Z-score 标准化或 Min-Max 归一化，消除量纲影响。  
（4）**特征编码**：对类别型变量采用合适的编码方式（如 One-Hot 编码、Label Encoding）。  
（5）**特征选择与构造**：通过相关性分析、特征重要性、降维（PCA）等方法筛选关键特征；同时可构造交互特征或分箱特征提升模型表现。