**制定数据清洗规范（2分）**  
（1）缺失值处理：数值型用均值/中位数或前后向填充；分类型用众数填充；缺失比例过高的特征或样本可删除。  
（2）异常与噪声处理：按业务阈值或统计规则（Z-score>3、IQR法）识别异常；视情况删除或截尾/插补；统一异常编码（如“-”“N/A”）为缺失。

**制定特征工程规范（3分）**  
（3）数值特征缩放：统一采用Z-score标准化或Min-Max归一化；训练集拟合、测试集变换，避免数据泄漏。  
（4）类别特征编码：名义型用One-Hot，有序型用Ordinal；高基数可目标编码/频数编码并做正则化与交叉验证。  
（5）特征选择与构造：基于相关性/模型重要性/正则化（L1）或PCA降维筛选；结合领域知识做分箱、交互项与聚合特征，保留可解释性并通过交叉验证评估收益。