**已完成的工作**

1. **数据加载与复制**：
   * 成功读取kobe\_data.csv文件，并创建数据副本确保原始数据不受影响。
2. **探索性数据分析（EDA）**：
   * 完成了单变量和双变量分析，通过多种可视化手段深入理解了数据分布和特征之间的关系。
3. **数据预处理**：
   * **特征删除**：移除了无关或冗余的特征，简化了数据集。
   * **特征工程**：创建了新特征time\_remaining，并处理了season特征。
   * **数据划分**：将数据集划分为训练集（25,697条记录）和测试集（5,000条记录）。
   * **缺失值处理**：在训练集中填补了数值型和分类特征的缺失值，并在测试集中填补了分类特征的缺失值。
   * **异常值处理**：仅对训练集中的shot\_distance进行了Z-score异常值检测与过滤，确保测试集数据的完整性。
   * **编码与标准化**：对分类变量进行了独热编码，并对数值型特征进行了标准化处理。
   * **数据保存**：将处理后的训练集和测试集分别保存为独立的CSV文件，确保原始数据的完整性。

**后续步骤**

1. **模型构建与训练**：
   * 使用预处理后的训练集kobe\_train\_preprocessed\_features.csv和kobe\_train\_preprocessed\_labels.csv进行模型训练。
   * 选择合适的机器学习算法（如随机森林、XGBoost等），并进行参数调优。
2. **模型评估**：
   * 使用交叉验证和性能指标（如准确率、精确率、召回率、F1分数、ROC-AUC）评估模型的表现。
   * 分析特征重要性，理解哪些特征对预测结果影响最大。
3. **预测与结果保存**：
   * 对预处理后的测试集kobe\_test\_preprocessed\_features.csv进行预测。
   * 将预测结果保存为CSV文件，供后续分析和报告使用。
4. **数据可视化与报告**：
   * 构建高级可视化图表和交互式仪表盘，直观展示分析结果和模型性能。
   * 编写详细的项目报告，包含数据探索、预处理、模型构建与评估等各个部分。