**任务一、精灵宝可梦数据分析**

**简介：**

本数据集（Pokemon.csv）中涵盖了从第一代到第七代共801只宝可梦小精灵的信息，其中包括了宝可梦精灵的名称，基础能力值，身高，体重，类型，传奇性等等。具体完整字段含义参考pokemon\_data\_description.txt.，其中is\_legendary代表是否为传奇宝可梦。

**问题：**

1. 对宝可梦数据集进行数据分析。

Tips: 可参考以下问题进行分析，也可自行选取角度进行分析。

（1）宝可梦的整体平均性别占比为何？

（2）宝可梦体型（身高，体重）分布为何？

（3）不同世代的宝可梦数量分布为何？  
（4）比较不同世代的宝可梦水平（基础攻击，基础防御，特殊攻击，特殊防御，速度等）。  
（5）宝可梦的类型（type1，type2）分布如何？

（6）传奇宝可梦的数量是多少？

1. 建立一个分类器来**识别传奇宝可梦**，并且选取合适的指标分析与评估模型效果。

**任务二、房价预测**

**简介：**

当请购房者们描述他们理想中的房子，他们可能会从房子占地面积多大，房间有几个卧室等方面开始谈起，但是事实证明，一间房子成交价格的影响因素，远远比这要复杂得多。在本数据集（house\_prices.csv）中，有 79 个特征变量，来描述房子的各个方面，具体完整字段含义参考houseprice\_data\_description.txt.，其中SalePrice代表房子价格。

**问题：**

**1.对房价数据进行分析**

**2.请建立模型，预测房子价格，并且选取合适的指标，分析与评估模型效果。**

Tips：可参考以下步骤进行建模：

1. 读取数据集
2. 数据预处理&特征工程
3. 数据集（训练集、预测集）划分
4. 模型训练
5. 模型预测与评估

**要求：**

**二选一，运用所学的知识完成任务**

1. 提交完整代码，代码需要可以运行。
2. 提交一份分析报告，报告中应包含两部分内容：
   1. 整体以及代码各部分的思路，功能。
   2. 数据分析的相关结果和结论。
3. 主要评价点（分值构成比例暂定，还会调整）：
   1. 方案的合理性（30%）
   2. 方案的难度（20%）
   3. 代码质量（15%）
   4. 分析结果和结论（包括丰富程度，呈现的方式，合理性、预测准确度等）（25%）
   5. 报告撰写的质量（10%）