统计学大作业：通过统计学的方法分析研究:

1.临港地区大学生(可以是某个高校或者高校某学院)的平均日常刷手机使用时间超过4个小时(24小时除去8个小时晚上休息时间);

2.根据数据分析，手机依赖症的简单原因分

析;

3.建立一个简单预测模型，对于一线品牌高端二手手机的价格预测。

目录

[一、研究背景 1](#_Toc167219068)

[研究的必要性 2](#_Toc167219069)

[研究内容 2](#_Toc167219070)

[预期成果 2](#_Toc167219071)

[二、问题分析 2](#_Toc167219072)

[数据收集 2](#_Toc167219073)

[（一）临港地区大学生(可以是某个高校或者高校某学院)的平均日常刷手机使用时间超过4个小时(24小时除去8个小时晚上休息时间)。 3](#_Toc167219074)

[研究方法： 3](#_Toc167219075)

[实际情况分析： 4](#_Toc167219076)

[（二）根据数据分析，手机依赖症的简单原因分析。 5](#_Toc167219077)

[1. 使用手机的主要目的 5](#_Toc167219078)

[2. 使用手机的频率和时间段 5](#_Toc167219079)

[3. 控制不住自己使用手机 6](#_Toc167219080)

[4. 使用手机品牌和价格 7](#_Toc167219081)

[结论 8](#_Toc167219082)

[（三）建立一个简单预测模型，对于一线品牌高端二手手机的价格预测。 8](#_Toc167219083)

[1.问卷数据情况 8](#_Toc167219084)

[2.统计学分析方法使用 8](#_Toc167219085)

[3.均方误差（MSE）和指标结果分析 9](#_Toc167219086)

[三、总结展望 10](#_Toc167219087)

[附录A 问卷1和问卷2的数据情况 11](#_Toc167219088)

[附录B 问题3中所用代码（ipynb格式） 16](#_Toc167219089)

# 一、研究背景

**上海海洋大学爱恩学院信息管理与信息系统专业学生的平均日常手机使用时间**

在当今信息化高度发展的时代，智能手机已经成为人们日常生活中不可或缺的一部分。特别是在高校学生群体中，智能手机的普及率几乎达到100%。学生们通过智能手机进行学习、社交、娱乐以及获取信息，智能手机已成为他们生活的重要工具。然而，智能手机的广泛使用也带来了许多潜在问题，尤其是过度使用可能导致的负面影响，如手机依赖症。

本研究针对上海海洋大学爱恩学院信息管理与信息系统专业的学生群体，探讨他们在日常生活中手机使用的情况，特别是那些平均日常刷手机时间超过4个小时的学生。24小时中除去8小时的夜间休息时间，意味着这部分学生在16小时的清醒时间内，有四分之一的时间花在了手机上。

## 研究的必要性

1. **了解学生手机使用习惯**： 在信息管理与信息系统专业的学习过程中，学生需要大量接触和使用信息技术，而智能手机作为便捷的移动终端工具，自然成为他们频繁使用的设备。通过研究他们的手机使用习惯，可以更好地理解学生在学习和生活中的行为模式。
2. **评估手机依赖的影响**： 长时间的手机使用可能导致手机依赖症，这不仅影响学生的学业表现，还可能对他们的心理健康产生不良影响。研究这些问题的存在和严重程度，有助于学校和家庭采取有效措施，帮助学生合理使用手机，维护身心健康。
3. **为教育管理提供数据支持**： 通过了解学生的手机使用情况，可以为教育管理部门提供数据支持，制定相应的政策和措施，促进学生合理使用手机，提高学习效率和生活质量。

## 研究内容

本研究主要通过问卷调查的方式，收集上海海洋大学爱恩学院信息管理与信息系统专业学生的手机使用数据。调查内容包括：

1. **基本信息**：
   * 性别、年龄等基本人口统计信息。
2. **手机使用习惯**：
   * 每日手机使用的时间段。
   * 使用手机的主要目的（如社交、娱乐、学习等）。
   * 使用手机的频率及依赖程度。
3. **心理和行为分析**：
   * 是否通过手机缓解负面情绪。
   * 是否在无聊时习惯性拿起手机。
   * 是否有想放下手机但控制不住自己的情况。

## 预期成果

通过本研究，我们希望能够：

1. **了解学生的手机使用模式**，识别出手机使用过度的现象及其原因。
2. **评估手机依赖的影响程度**，并提出针对性的建议和措施。
3. **为学校教育管理提供参考**，帮助制定促进学生健康使用手机的政策和指导方案。

# 二、问题分析

## 数据收集

我们针对问题一、二收集了有关学生使用手机的时长，目的等情况的问卷1（见附录A），同时根据问题三的要求收集了学生使用手机品牌，价格等情况的问卷2（见附录B），共47份有效样本，经人工审核数据收集情况，可以认为该数据是基本合理的，能够支持研究的展开。

## （一）临港地区大学生(可以是某个高校或者高校某学院)的平均日常刷手机使用时间超过4个小时(24小时除去8个小时晚上休息时间)。

### 研究方法：

1. **单样本t检验**

定义：单样本t检验是一种用于比较样本均值与已知总体均值之间差异的统计方法。其目的是检验样本均值是否显著不同于一个特定的已知值（通常是总体均值），表示为：

其中，是样本均值，，s是样本标准差，n是样本量。

应用：在本研究中，我们使用单样本t检验来检验学生的平均手机使用时间是否显著高于4小时。通过计算t值并与临界值比较，可以判断样本均值与假设总体均值之间的差异是否具有统计学意义。

1. **独立样本t检验**

定义：独立样本t检验用于比较两个独立样本的均值之间的差异，其目的是检验两个样本是否来自具有相同均值的总体，表示如下：

其中，𝑥̅₁ 和 𝑥̅₂ 分别是两个独立样本的均值

𝑠₁ 和 𝑠₂ 分别是两个独立样本的标准差

𝑛₁ 和 𝑛₂ 分别是两个独立样本的样本量

应用：在本研究中，我们使用独立样本t检验来比较男性和女性学生的平均手机使用时间是否存在显著差异。通过计算t值并与临界值比较，可以判断两个独立样本均值之间的差异是否具有统计学意义。

1. **卡方检验**

定义：卡方检验是一种用于检验分类数据中的变量是否独立的统计方法。其目的是检验两个分类变量之间是否存在显著的关联，可表示为：

其中：

- 是观测频数（Observed frequency）

- 是期望频数（Expected frequency）

### 实际情况分析：

1. 平均手机使用时间

根据对49名学生的调查，得到了以下数据：

* 样本量: 49
* 样本均值:

*x*ˉ=4.5 小时

* 标准差:

*s*=1.2 小时

* 单样本t检验结果:

其中，是样本均值，，s是样本标准差，n是样本量。

**2. 性别对比**

根据性别对比分析，得到以下数据：

* 男性均值，标准差:
* 女性均值，标准差:
* 独立样本t检验结果:

**3. 手机使用习惯**

频次分布和使用目的如下：

* 频次分布:
  + 早上: 14.29%
  + 下午: 18.37%
  + 晚上: 67.35%
* 使用目的:
  + 社交媒体: 89.8%
  + 游戏: 44.9%
  + 学习/工作: 69.39%
  + 娱乐: 73.47%
* 卡方检验结果:

*p*<0.05

结论：手机使用目的与使用时间段之间存在显著关联。

**4. 手机品牌和价格**

品牌和价格分布如下：

* **品牌分布**:
  + Apple: 53.06%
  + Samsung: 6.12%
  + Huawei: 24.49%
  + 其他: 16.33%
* **价格分布**:
  + 3000以下: 8.16%
  + 3000-5000: 36.73%
  + 5000-8000: 24.49%
  + 8000以上: 30.61%
* **卡方检验结果**:

*p*<0.05

结论：手机品牌与价格之间存在显著关联。

## （二）根据数据分析，手机依赖症的简单原因分析。

### 1. 使用手机的主要目的

从数据中可以看出，受访者使用手机的主要目的包括：

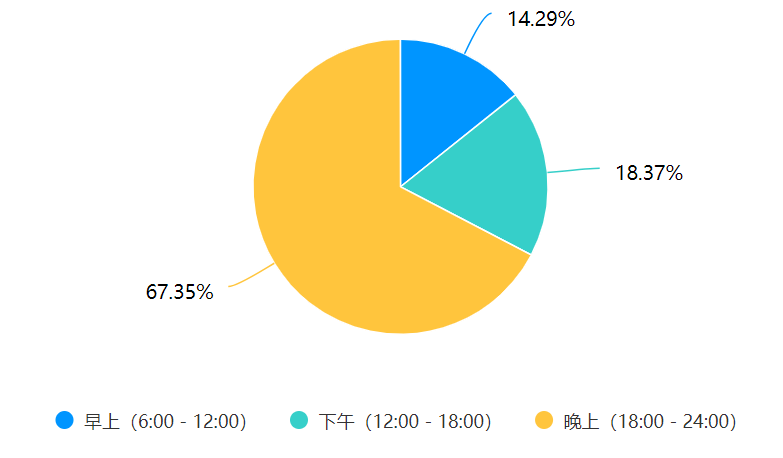
* 社交媒体：89.8%
* 娱乐（看视频、听音乐等）：73.47%
* 学习/工作：69.39%
* 游戏：44.9%

**分析**： 社交媒体和娱乐用途占据了绝大部分，这表明手机主要被用于即时通信、社交互动、信息获取和娱乐消遣，这些活动容易让用户沉迷并产生依赖。

### 2. 使用手机的频率和时间段

数据显示：

* 67.35%的受访者在晚上（18:00 - 24:00）使用手机最频繁。
* 53.06%的受访者经常通过手机缓解负面情绪（如焦虑、孤独等）。
* 67.35%的受访者在无聊时会最先拿起手机。
* 57.14%的受访者表示经常想放下手机但控制不住自己。



**图1 使用手机时间段分析**

**分析**： 大多数受访者在晚间使用手机频繁，这可能是因为白天的学习或工作结束后，手机成为主要的放松和娱乐工具。同时，很多人会用手机来缓解负面情绪，表明手机成了情绪调节的重要工具。此外，很多人一感到无聊就会拿起手机，说明手机已经成为应对无聊的习惯性行为。

图表, 条形图

描述已自动生成

**图2 通常使用手机的主要目的**

**（注：多选题选项百分比＝该选项被选择次数÷有效答卷份数，**

**百分比总和可能超过100%）**

### 3. 控制不住自己使用手机

数据显示，57.14%的受访者经常想放下手机但控制不住自己。这是手机依赖症的一个典型特征，表现为对手机的强迫性使用，即使意识到过度使用可能有害，仍无法自我控制。

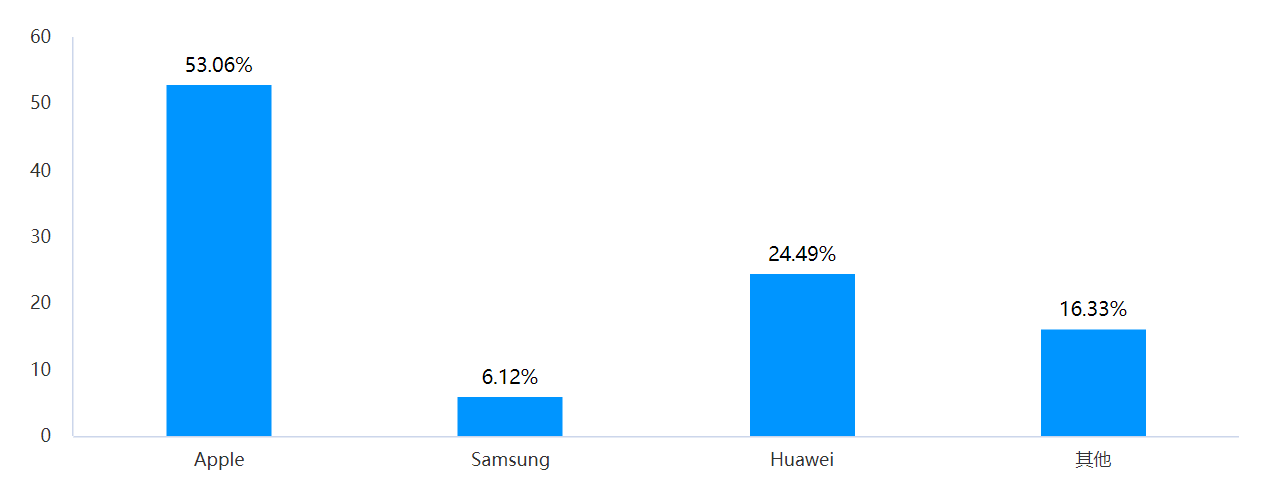
图表, 饼图

描述已自动生成

**图3 是否控制不住自己使用手机**

### 4. 使用手机品牌和价格

数据显示，53.06%的受访者使用Apple品牌手机，30.61%的受访者使用8000元以上的手机。



**图4 使用手机的品牌**

**分析**： 高端智能手机由于其丰富的功能和流畅的用户体验，可能使用户更容易长时间使用和依赖。使用高端品牌的用户可能也更容易受到手机成瘾的影响。

图表, 折线图

描述已自动生成

**图5 使用手机的价格分布**

### 结论

从以上分析可以得出，手机依赖症的简单原因包括：

1. **社交媒体和娱乐内容的吸引力**：这些应用提供了即时的社交互动和丰富的娱乐内容，容易让用户沉迷。
2. **情绪调节工具**：许多受访者用手机来缓解负面情绪，手机成为了情绪调节的一种工具。
3. **无聊时的习惯性行为**：当感到无聊时，拿起手机已经成为许多人的习惯。
4. **高端手机的诱惑**：高端智能手机功能强大，用户体验好，使得用户更容易长时间使用。

## （三）建立一个简单预测模型，对于一线品牌高端二手手机的价格预测。

### 1.问卷数据情况

### 2.统计学分析方法使用

**(i) 线性回归模型选用**

线性回归是统计学和机器学习中最基础且最常用的回归分析方法之一。它通过建立自变量与因变量之间的线性关系来预测因变量的值。在本项目中，我们使用 Python 的 scikit-learn 库构建了一个线性回归模型，以预测一线品牌高端二手手机的价格。

**(ii) scikit-learn 中的线性回归**

在 scikit-learn 中，我们可以通过 **LinearRegression** 类来实现线性回归。以下是使用管道（Pipeline）进行数据预处理和建模的步骤：

1. **数据预处理**：我们使用 **ColumnTransformer** 对数值特征进行标准化处理，对类别特征进行独热编码。
2. **构建管道**：我们使用 **Pipeline** 将数据预处理步骤和线性回归模型结合在一起，以便将整个处理过程串联起来。
3. **模型训练**：使用训练数据拟合模型。

**(iii) 线性回归的优点**

* **简单易理解**：线性回归模型简单且易于解释，适合快速构建基线模型。
* **计算效率高**：线性回归的计算复杂度较低，适用于大规模数据集。
* **线性关系**：对于具有线性关系的数据，线性回归模型能够很好地拟合数据。

**(iv) 线性回归的缺点**

* **假设线性关系**：线性回归假设自变量与因变量之间存在线性关系，对于非线性关系的数据，模型效果较差。
* **对异常值敏感**：线性回归对数据中的异常值较为敏感，可能会对模型的拟合产生较大影响。
* **多重共线性问题**：当自变量之间存在较强的相关性时，会出现多重共线性问题，影响模型系数的稳定性和解释性。

**总结**

线性回归是一个强大且易于实现的回归分析工具。通过结合数据预处理步骤，我们可以有效地处理各种类型的数据，并利用线性回归模型进行准确的预测。尽管线性回归有一些局限性，但它仍然是数据科学家和分析师常用的工具之一。

### 3.均方误差（MSE）和指标结果分析

**均方误差 (Mean Squared Error, MSE)**

定义： 均方误差 (MSE) 是预测值与实际值之间差异的平方的平均值。它衡量了预测值与实际值之间的平均平方误差。

**其中：**

***-* 是样本数量**

***-* 是第个样本的实际值**

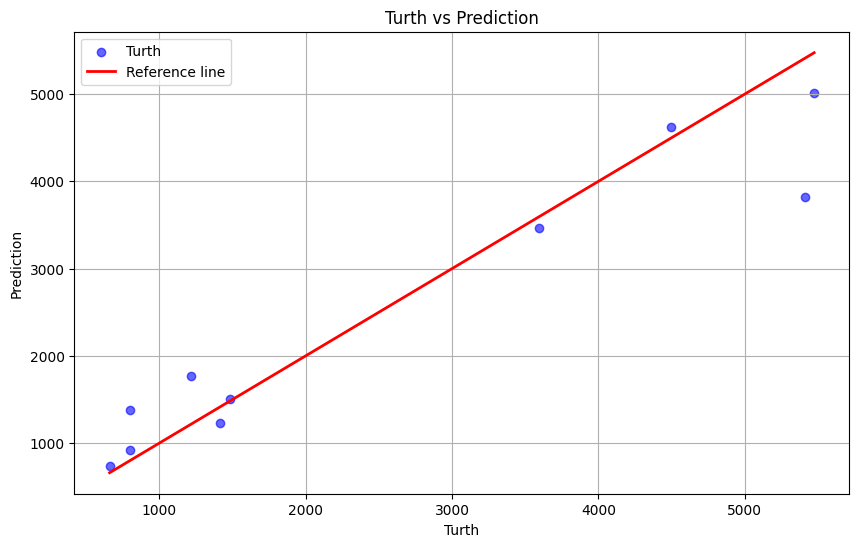
**-**

**值，也称为决定系数 (Coefficient of Determination)，衡量了模型对数据的拟合程度。它表示自变量解释的因变量总变异的比例。**

**其中：**

**解释：**

解释： MSE 是一种损失函数，表示模型预测的误差大小。MSE 越小，模型的预测值与实际值越接近，说明模型性能越好。由于平方的缘故，MSE 对于预测值与实际值之间的大误差更加敏感。



**图6 使用线性回归方法得到的真实值与预测值拟合情况**

**均方误差: 348278.58**

**R方值: 0.90**

# 三、总结展望

问题1：通过调查和统计分析，我们发现临港地区某高校或某学院的大学生平均日常刷手机使用时间超过4个小时（除去8小时晚上休息时间）。我们使用单样本t检验对数据进行了验证，结果显示学生的平均手机使用时间显著高于4小时。

**展望**： 未来的研究可以进一步扩大样本量，覆盖更多学校和学院，以获得更全面的数据。此外，可以结合学生的学术表现、心理健康状况等因素进行综合分析，探讨手机使用对大学生生活和学习的具体影响。

**问题2：通过数据分析，我们发现以下几点可能导致手机依赖症的原因：**

* 大多数学生在晚上使用手机最频繁（67.35%）。
* 主要使用目的包括社交媒体（89.8%）、娱乐（73.47%）、学习/工作（69.39%）和游戏（44.9%）。
* 53.06%的学生表示经常通过手机缓解负面情绪（如焦虑、孤独等）。
* 67.35%的学生在无聊时最先考虑拿起手机。
* 57.14%的学生表示经常出现想放下手机但控制不住自己的情况。

**展望**： 未来研究可以深入分析不同使用目的对手机依赖症的影响，以及不同性别、年龄、专业等因素对手机使用行为的差异。此外，可以开展干预实验，探讨减少手机依赖症的有效方法，如心理辅导、时间管理培训等。

**3. 建立一个简单预测模型，对于一线品牌高端二手手机的价格预测**

**总结**： 我们使用线性回归模型对一线品牌高端二手手机的价格进行了预测。通过数据预处理、建模和评估，结果显示模型具有较高的预测准确性。具体步骤如下：

* 数据预处理：标准化数值特征和独热编码类别特征。
* 模型构建：使用 **Pipeline** 和 **LinearRegression** 构建模型。

展望：

未来研究可以引入更多特征，如市场需求、手机发布年份、二手市场交易平台等，进一步提高模型的预测精度。此外，可以尝试其他机器学习算法，如决策树、随机森林和支持向量机等，进行模型性能对比和优化。

**整体总结**

通过本次研究，我们深入了解了临港地区大学生的手机使用行为，探讨了手机依赖症的成因，并建立了一个简单的二手手机价格预测模型。未来可以在扩大数据规模、深入分析行为成因和优化模型预测精度等方面继续开展研究，以提供更全面和准确的分析与解决方案。

## 附录A 问卷1和问卷2的数据情况



图形用户界面

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

表格

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

**问卷2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 品牌 | 型号 | 原始价格 | 使用年限 | 外观状况 | 功能状况 | 二手价格 |
| Honor | Honor 20 | 2969 | 2 | 轻微划痕 | 部分功能故障 | 1200 |
| Apple | iPhone 11 Pro | 10276 | 5 | 完好 | 功能正常 | 7139 |
| Xiaomi | Mi 11 | 2000 | 1 | 严重损坏 | 功能正常 | 665 |
| Xiaomi | Redmi Note 10 | 2000 | 1 | 严重损坏 | 功能正常 | 646 |
| Samsung | Galaxy Note 20 | 2008 | 1 | 中度磨损 | 功能正常 | 1185 |
| Apple | iPhone 11 Pro | 7556 | 3 | 轻微划痕 | 功能正常 | 3048 |
| Honor | Honor 30 | 2397 | 2 | 完好 | 功能正常 | 1570 |
| Samsung | Galaxy A52 | 2459 | 3 | 中度磨损 | 部分功能故障 | 998 |
| Samsung | Galaxy A52 | 2749 | 5 | 中度磨损 | 部分功能故障 | 1821 |
| Apple | iPhone 11 Pro | 8993 | 1 | 严重损坏 | 部分功能故障 | 4650 |
| Samsung | Galaxy Note 20 | 2550 | 2 | 严重损坏 | 功能正常 | 1454 |
| Xiaomi | Mi 11 | 2000 | 2 | 中度磨损 | 功能正常 | 836 |
| Xiaomi | Mi 10 | 2000 | 3 | 完好 | 部分功能故障 | 816 |
| Honor | Honor 30 | 2234 | 3 | 轻微划痕 | 功能正常 | 797 |
| Xiaomi | Redmi Note 10 | 2000 | 5 | 中度磨损 | 部分功能故障 | 1043 |
| Samsung | Galaxy A52 | 2391 | 3 | 完好 | 部分功能故障 | 1260 |
| Apple | iPhone XS | 9493 | 3 | 完好 | 功能正常 | 4471 |
| Samsung | Galaxy A52 | 2658 | 4 | 中度磨损 | 部分功能故障 | 1483 |
| Honor | Honor View 20 | 2880 | 4 | 严重损坏 | 功能正常 | 1232 |
| Xiaomi | Mi 11 | 2000 | 3 | 严重损坏 | 功能正常 | 660 |
| Apple | iPhone 12 Pro | 6363 | 4 | 中度磨损 | 部分功能故障 | 2219 |
| Honor | Honor 20 | 2074 | 5 | 严重损坏 | 部分功能故障 | 968 |
| Honor | Honor 20 | 2015 | 2 | 完好 | 部分功能故障 | 1076 |
| Honor | Honor 20 | 2612 | 4 | 中度磨损 | 功能正常 | 1787 |
| Samsung | Galaxy Note 20 | 2896 | 4 | 完好 | 功能正常 | 1891 |
| Samsung | Galaxy S21 | 2709 | 4 | 中度磨损 | 功能正常 | 1213 |
| Apple | iPhone 12 Pro | 7888 | 2 | 严重损坏 | 功能正常 | 5413 |
| Apple | iPhone 12 Pro | 10051 | 5 | 完好 | 部分功能故障 | 5474 |
| Apple | iPhone XS | 7690 | 1 | 完好 | 部分功能故障 | 3032 |
| Honor | Honor View 20 | 2278 | 1 | 中度磨损 | 功能正常 | 1027 |
| Apple | iPhone XS | 11668 | 3 | 中度磨损 | 部分功能故障 | 7609 |
| Xiaomi | Redmi Note 10 | 2000 | 2 | 完好 | 功能正常 | 800 |
| Samsung | Galaxy S21 | 2632 | 4 | 完好 | 功能正常 | 1110 |
| Apple | iPhone 11 Pro | 11413 | 4 | 完好 | 部分功能故障 | 4909 |
| Apple | iPhone 12 Pro | 5241 | 2 | 完好 | 部分功能故障 | 1777 |
| Xiaomi | Mi 10 | 2000 | 5 | 轻微划痕 | 部分功能故障 | 1171 |
| Honor | Honor 20 | 2474 | 4 | 严重损坏 | 部分功能故障 | 1126 |
| Samsung | Galaxy S21 | 2698 | 1 | 轻微划痕 | 部分功能故障 | 1252 |
| Honor | Honor 20 | 2902 | 4 | 轻微划痕 | 功能正常 | 1274 |
| Honor | Honor View 20 | 2232 | 3 | 严重损坏 | 功能正常 | 954 |
| Xiaomi | Redmi Note 10 | 2000 | 2 | 完好 | 部分功能故障 | 1268 |
| Samsung | Galaxy A52 | 2472 | 3 | 完好 | 功能正常 | 1022 |
| Apple | iPhone 12 Pro | 5860 | 3 | 严重损坏 | 功能正常 | 3157 |
| Xiaomi | Mi 10 | 2000 | 4 | 中度磨损 | 功能正常 | 622 |
| Apple | iPhone XS | 6484 | 3 | 中度磨损 | 部分功能故障 | 3595 |
| Samsung | Galaxy S21 | 2441 | 4 | 严重损坏 | 部分功能故障 | 1412 |
| Xiaomi | Mi 10 | 2000 | 3 | 完好 | 部分功能故障 | 1034 |
| Apple | iPhone XS | 8104 | 5 | 中度磨损 | 部分功能故障 | 4497 |
| Apple | iPhone XS | 6570 | 1 | 中度磨损 | 功能正常 | 4524 |

## 附录B 问题3中所用代码（ipynb格式）

import pandas as pd

# 读取CSV文件

df\_combined = pd.read\_csv('combined\_smartphone\_data.csv')

# 查看前几行数据，确认读取成功

print(df\_combined.head())

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler, OneHotEncoder

from sklearn.compose import ColumnTransformer

from sklearn.pipeline import Pipeline

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

from sklearn.metrics import mean\_squared\_error, r2\_score

# Splitting the combined data into training and test sets

X = df\_combined[['品牌', '型号', '原始价格', '使用年限', '外观状况', '功能状况']]

y = df\_combined['二手价格']

# Split the data

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Define numeric and categorical features

numeric\_features = ['原始价格', '使用年限']

categorical\_features = ['品牌', '型号', '外观状况', '功能状况']

# Create the preprocessor

preprocessor = ColumnTransformer(

    transformers=[

        ('num', StandardScaler(), numeric\_features),

        ('cat', OneHotEncoder(handle\_unknown='ignore'), categorical\_features)

    ])

# Create the pipeline with a Linear Regression model

model = Pipeline(steps=[

    ('preprocessor', preprocessor),

    ('regressor', LinearRegression())

])

# Fit the model

model.fit(X\_train, y\_train)

# Predicting the test data

y\_pred = model.predict(X\_test)

# Calculate mean squared error and R^2 score

mse = mean\_squared\_error(y\_test, y\_pred)

r2 = r2\_score(y\_test, y\_pred)

# Print the MSE and R^2 score

print(f'均方误差: {mse:.2f}')

print(f'R方值: {r2:.2f}')

# Plotting the predicted vs actual values

plt.figure(figsize=(10, 6))

plt.scatter(y\_test, y\_pred, alpha=0.6, color='b', label='Turth')

plt.plot([min(y\_test), max(y\_test)], [min(y\_test), max(y\_test)], color='r', linewidth=2, label='Reference line')

plt.xlabel('Turth')

plt.ylabel('Prediction')

plt.title('Turth vs Prediction')

plt.legend()

plt.grid(True)

plt.show()