

实验一

学号： 220111028 姓名： 许辰涛

一、 运行截图（配合截图结果，可以添加一些文字说明）

合并两张表格

使用 0 填充 NaN

计算总成绩与平均成绩

将数据按总成绩降序排序

| | id | sex | politics | chinese | math | biology | physics | chemistry | \ |
|----|-------|-----|----------|---------|------|---------|---------|-----------|---|
| 59 | 92101 | 2.0 | 96.0 | 96.0 | 87.5 | 72.0 | 93.0 | 65.0 | |
| 22 | 92202 | 1.0 | 78.0 | 89.0 | 83.5 | 81.0 | 91.0 | 77.0 | |
| 58 | 92102 | 1.0 | 94.0 | 97.0 | 86.5 | 61.0 | 93.0 | 64.0 | |
| 51 | 92104 | 2.0 | 89.0 | 97.0 | 69.5 | 86.0 | 83.0 | 62.0 | |
| 48 | 92106 | 2.0 | 88.0 | 88.0 | 78.0 | 60.0 | 90.0 | 70.0 | |
| 57 | 92110 | 1.0 | 92.0 | 94.0 | 71.0 | 65.0 | 78.0 | 62.0 | |
| 14 | 92203 | 2.0 | 74.0 | 93.0 | 84.5 | 50.0 | 89.0 | 72.0 | |
| 47 | 92204 | 2.0 | 88.0 | 81.0 | 87.5 | 60.0 | 84.0 | 63.0 | |
| 26 | 92205 | 2.0 | 81.0 | 79.0 | 84.0 | 60.0 | 91.0 | 64.0 | |
| 31 | 92207 | 2.0 | 83.0 | 91.0 | 70.5 | 60.0 | 84.0 | 64.0 | |
| 29 | 92105 | 1.0 | 82.0 | 85.0 | 79.5 | 60.0 | 88.0 | 66.0 | |
| 38 | 92108 | 2.0 | 84.0 | 90.0 | 69.5 | 50.0 | 80.0 | 60.0 | |
| 27 | 92112 | 2.0 | 81.0 | 75.5 | 76.5 | 43.0 | 78.0 | 83.0 | |
| 39 | 92116 | 2.0 | 84.0 | 87.0 | 67.5 | 52.0 | 82.0 | 60.0 | |
| 53 | 92209 | 2.0 | 90.0 | 91.0 | 70.5 | 39.0 | 89.0 | 53.0 | |
| 9 | 92113 | 1.0 | 70.0 | 85.0 | 66.0 | 63.0 | 86.0 | 65.0 | |
| 35 | 92117 | 1.0 | 83.0 | 91.0 | 80.5 | 44.0 | 89.0 | 56.0 | |
| 41 | 92206 | 2.0 | 86.0 | 86.0 | 77.5 | 40.0 | 92.0 | 60.0 | |
| 40 | 92115 | 1.0 | 85.0 | 91.0 | 72.5 | 63.0 | 70.0 | 50.0 | |
| 28 | 92213 | 1.0 | 82.0 | 76.0 | 65.0 | 60.0 | 75.0 | 60.0 | |
| 55 | 92208 | 1.0 | 91.0 | 88.0 | 63.0 | 48.0 | 75.0 | 54.0 | |
| 54 | 92120 | 1.0 | 90.0 | 84.0 | 55.0 | 50.0 | 82.0 | 60.0 | |
| 11 | 92211 | 2.0 | 71.0 | 73.0 | 69.0 | 42.0 | 95.0 | 61.0 | |
| 42 | 92214 | 1.0 | 86.0 | 86.0 | 62.0 | 50.0 | 73.0 | 65.0 | |
| 36 | 92212 | 1.0 | 84.0 | 85.0 | 61.5 | 60.0 | 78.0 | 41.0 | |
| 6 | 92111 | 1.0 | 61.0 | 86.0 | 74.0 | 51.0 | 74.0 | 61.0 | |
| 25 | 92122 | 1.0 | 80.0 | 88.5 | 63.5 | 44.0 | 75.0 | 70.0 | |

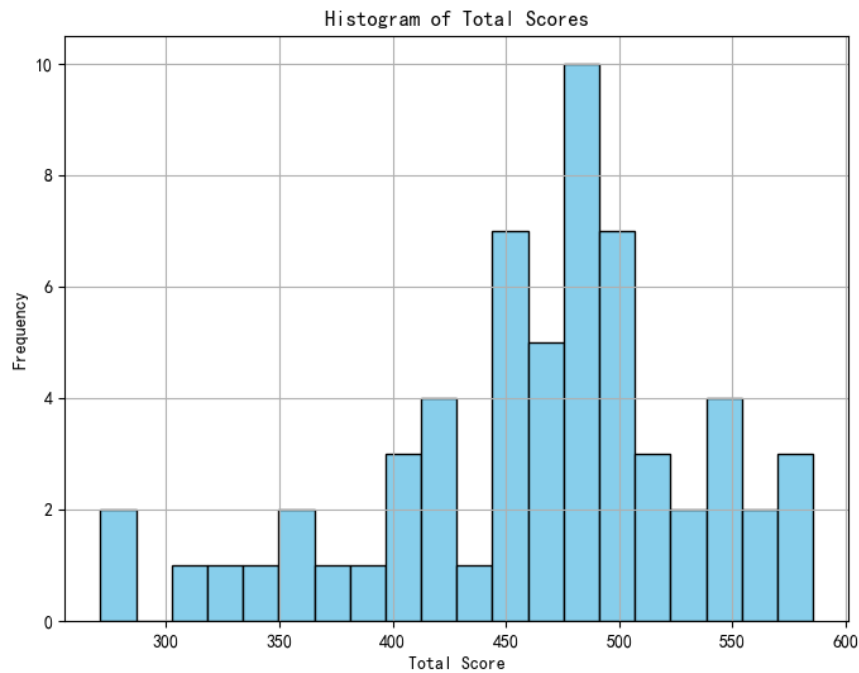
| | geography | history | 总成绩 | 平均成绩 | 成绩等级 |
|----|-----------|---------|-------|-----------|------|
| 59 | 76.0 | 92.0 | 585.5 | 83.642857 | 良 |
| 22 | 81.0 | 93.0 | 580.5 | 82.928571 | 良 |
| 58 | 79.5 | 95.0 | 575.0 | 82.142857 | 良 |
| 51 | 83.0 | 94.0 | 569.5 | 81.357143 | 良 |
| 48 | 81.5 | 77.0 | 555.5 | 79.357143 | 良 |
| 57 | 83.0 | 87.0 | 545.0 | 77.857143 | 中 |
| 14 | 82.5 | 92.0 | 545.0 | 77.857143 | 中 |
| 47 | 79.0 | 92.0 | 542.5 | 77.500000 | 中 |
| 26 | 81.0 | 92.0 | 540.0 | 77.142857 | 中 |
| 31 | 81.5 | 62.0 | 534.0 | 76.285714 | 中 |
| 29 | 72.5 | 98.0 | 533.0 | 76.142857 | 中 |
| 38 | 86.5 | 94.0 | 520.0 | 74.285714 | 中 |
| 27 | 78.0 | 91.0 | 515.0 | 73.571429 | 中 |
| 39 | 79.0 | 71.0 | 511.5 | 73.071429 | 中 |
| 53 | 68.5 | 81.0 | 501.0 | 71.571429 | 中 |
| 9 | 64.0 | 84.0 | 499.0 | 71.285714 | 中 |
| 35 | 55.0 | 62.0 | 498.5 | 71.214286 | 中 |
| 41 | 57.0 | 76.0 | 498.5 | 71.214286 | 中 |
| 40 | 65.0 | 82.0 | 496.5 | 70.928571 | 中 |
| 28 | 78.0 | 76.0 | 496.0 | 70.857143 | 中 |
| 55 | 76.0 | 95.0 | 495.0 | 70.714286 | 中 |
| 54 | 67.5 | 81.0 | 488.5 | 69.785714 | 中 |
| 11 | 76.5 | 76.0 | 487.5 | 69.642857 | 中 |
| 42 | 63.5 | 81.0 | 485.5 | 69.357143 | 中 |
| 36 | 74.0 | 91.0 | 483.5 | 69.071429 | 中 |

按性别分别计算各门课程的成绩

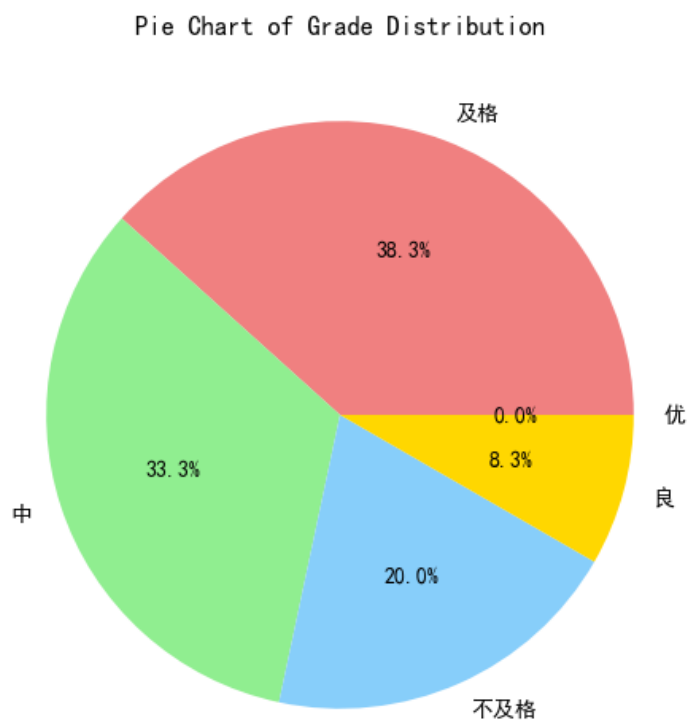
| | id | politics | chinese | math | biology | physics \ |
|-----|--------------|-----------|-----------|-----------|-----------|-----------|
| sex | | | | | | |
| 1.0 | 92171.066667 | 78.866667 | 83.500000 | 60.033333 | 51.200000 | 72.733333 |
| 2.0 | 92169.133333 | 75.100000 | 80.283333 | 60.266667 | 48.633333 | 77.666667 |

| | chemistry | geography | history | 总成绩 | 平均成绩 |
|-----|-----------|-----------|-----------|------------|-----------|
| sex | | | | | |
| 1.0 | 52.366667 | 62.533333 | 78.933333 | 461.233333 | 65.890476 |
| 2.0 | 55.800000 | 67.950000 | 78.433333 | 465.700000 | 66.528571 |

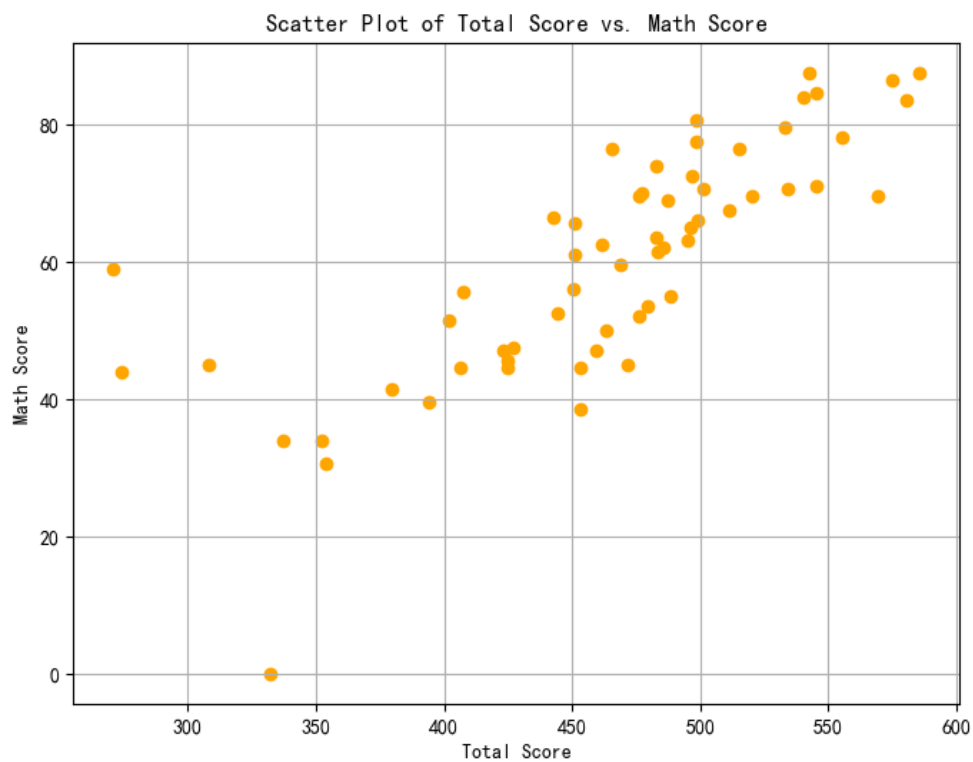
绘制总成绩的直方图



绘制平均成绩的优、良、中、及格和不及格的饼图



提取数学（math）成绩，绘制总成绩和数学成绩的散点图



二、 代码（文本式粘贴，重要代码处，需添加注释）

```
import pandas as pd

import matplotlib.pyplot as plt

plt.rcParams['font.sans-serif']=['SimHei'] #解决中文显示乱码问题

plt.rcParams['axes.unicode_minus']=False


df1 = pd.read_table("ReportCard1.txt")

df2 = pd.read_table("ReportCard2.txt")

df = pd.merge(df1, df2, on='id')


# 用 0 填充缺考成绩为 NaN 的情况

df = df.fillna(0)


# 计算每个同学所有课程的总成绩和平均成绩

df['总成绩'] = df.iloc[:, 2:9].sum(axis=1)

df['平均成绩'] = df.iloc[:, 2:9].mean(axis=1)


# 将数据按总成绩的降序排序
```

```
df = df.sort_values(by='总成绩', ascending=False)

# 按性别分别计算各门课程的平均成绩

avg_score_by_gender = df.groupby('sex').mean()

# 分组计算平均成绩的优、良、中、及格和不及格

bins = [0, 59, 69, 79, 89, 100]

labels = ['不及格', '及格', '中', '良', '优']

df['成绩等级'] = pd.cut(df['平均成绩'], bins=bins, labels=labels)

grade_counts = df['成绩等级'].value_counts()

# 绘制总成绩的直方图

plt.figure(figsize=(8, 6))

plt.hist(df['总成绩'], bins=20, color='skyblue', edgecolor='black')

plt.title('Histogram of Total Scores')

plt.xlabel('Total Score')

plt.ylabel('Frequency')

plt.grid(True)

plt.show()
```

绘制平均成绩的优、良、中、及格和不及格的饼图

```
plt.figure(figsize=(8, 6))
```

```
plt.pie(grade_counts, labels=grade_counts.index, autopct='%1.1f%%',  
colors=['lightcoral', 'lightgreen', 'lightskyblue', 'gold',  
'lightgrey'])
```

```
plt.title('Pie Chart of Grade Distribution')
```

```
plt.show()
```

提取数学（math）成绩，绘制总成绩和数学成绩的散点图

```
plt.figure(figsize=(8, 6))
```

```
plt.scatter(df['总成绩'], df['math'], color='orange')
```

```
plt.title('Scatter Plot of Total Score vs. Math Score')
```

```
plt.xlabel('Total Score')
```

```
plt.ylabel('Math Score')
```

```
plt.grid(True)
```

```
plt.show()
```