**Report 4**

**Part 1: Course exercises**

1.1:

1.2:

1.3:

1.4:

1.5:

2.1:

2.2:

2.3:

2.4:

2.5:

3.1:

3.2:

3.3:

3.4

3.5:

4.1:

4.2:

5.1:

5.2:

5.3:

**Part 2:**

**Download the files from the Moodle and complete the following Pandas practices and screenshot your solutions and results.**

1. 读取2004年的数据，并展示查看数据的格式.

2. 对第一年的数据进行预处理：删除第一行，选取真实列名所在的第一行，更改列名，删除多余的行.

3. 第二步删除行后，index被打乱：重新设置被打乱的index，添加年份变量.

4. 批量读取连接数据：自定义一个函数对后面所有年份进行批量预处理（删除第0列，删除前朗行后后一行，重塑年份变量）.

5. 定义函数用于批量读取及凭借数据，调用第四步中定义的函数进行批量处理：建立空列表用于存放数据，通过循环遍历读取文件，读取数据，一次累加年份，进行预处理及其重塑变量等.

6. 重塑列名，链接数据，并填充缺失值.

7. 检查数据：对数据进行计数，替换文字中的空格，删除多余类型数据，替换有歧义数据，再次检查.

8. 人口数据的清洗与重塑：修改列名，删除多余的行，统一格式.

9. 拼接数据：填充数据，转换数据类型，最终查看数据并保存输出.

*Ref：Pandas数据预处理与数据重塑案例（*[*https://zhuanlan.zhihu.com/p/44677396*](https://zhuanlan.zhihu.com/p/44677396)*）*