**Report 3**

**Part 1: Course exercises**

1.1:DateFrame

1.2:索引 数据

1.3:DateFrame

1.4:read\_csv

1.5:multiIndex

2.1:T

2.2:T

2.3:T

2.4:F

2.5:F

3.1:C

3.2:D

3.3:A

3.4:A

3.5:A

4.1:

**Series：**

**数据结构维度**：是一维标签化数组，可简单理解为类似于列表，但功能更强大，适用于处理单列数据，比如时间序列数据中的某一列数据 。

**数据类型**：能包含任何数据类型，如整数、字符串、浮点数等，但一个 Series 中通常为单一数据类型。

**索引**：每个 Series 都带有一个索引，索引可以是自定义标签，也可以是默认的整数索引，通过索引能方便地访问和操作数据。

**功能**：功能相对简单，主要用于一维数据的处理和分析，如简单的统计计算、数据筛选等。

**DataFrame：**

**数据结构维度**：属于二维标签化数据结构，可看作是由多个 Series 组成的表格，适合处理多列数据构成的表格型数据。

**数据类型**：能够容纳不同类型的数据，比如同一 DataFrame 中可以同时包含文本、数值、布尔值等多种数据类型。

**索引：**具有行索引和列索引，这种双索引结构让数据的组织和访问更加灵活，可以从行和列两个方向对数据进行操作。

**功能**：提供了更多高级功能，诸如复杂的数据合并、分组聚合运算、数据透视表创建、数据清洗与转换等，在处理复杂的数据分析任务时优势明显。

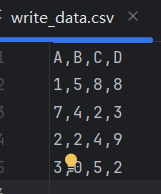
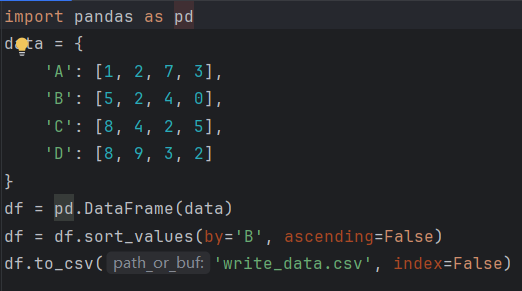
4.2:

它允许在一个轴向上拥有两个或两个以上的索引层级 ，提供了一种在更低维度形式中处理更高维度数据的方式。可以将其理解为给数据建立了多层标签体系，让数据组织更立体。比如在时间序列数据中，可以外层索引为年份，内层索引为月份，方便管理和分析数据。

5.1:

5.2:

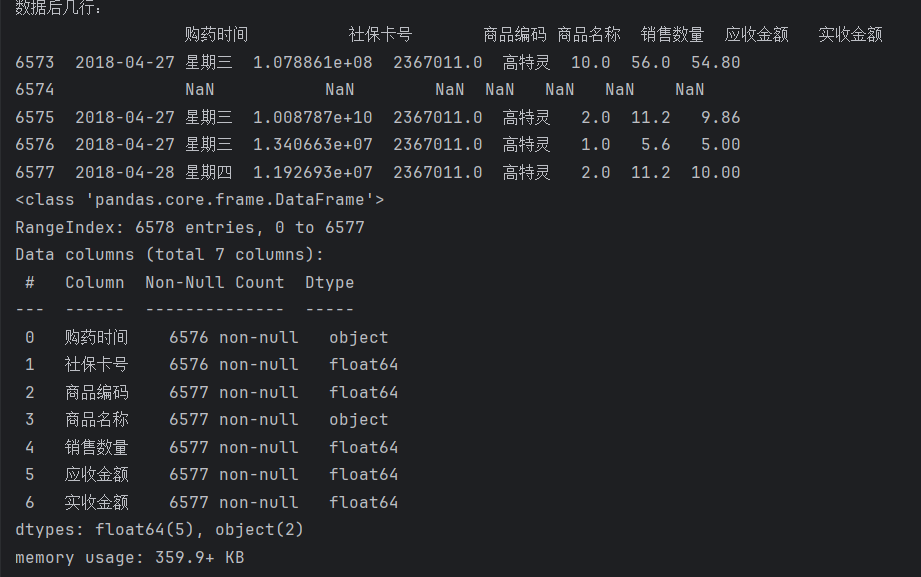
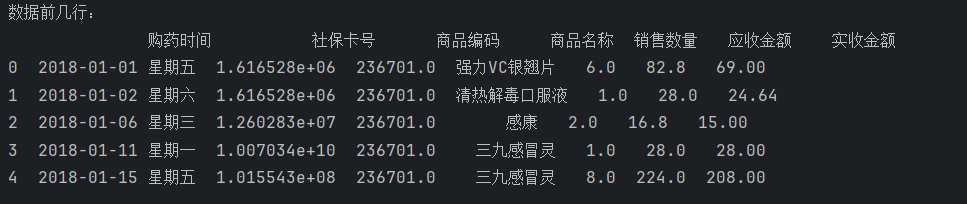
5.3:



**Part 2:**

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

1. 读取文件并查看文件中数据的基本信息.



2. 把获取数据中“购药时间”改为售药时间.

3. 删除数据中的缺失值.

4. 将字符串转换为浮点型数据.

5. 将“售药时间”中的星期去除，获取日期，并将“售药时间”这一列设置成所获取的日期.

6. 将获取的日期转换为时间格式.

7. 按照“售药时间”进行降序排列.

8. 将“销售数量”、“应收金额”、“实收金额”这三列中的异常值排除掉.

9. 将最终的数据写入到新的.xlsx文件，以学号命名.