**Report 5**

**Part 1: Course exercises**

1.1:拆分

1.2:by

1.3:可迭代

1.4:groups

1.5:DataFrameBy

2.1:F

2.2:F

2.3:F

2.4:T

2.5:F

3.1:A

3.2:A

3.3:B

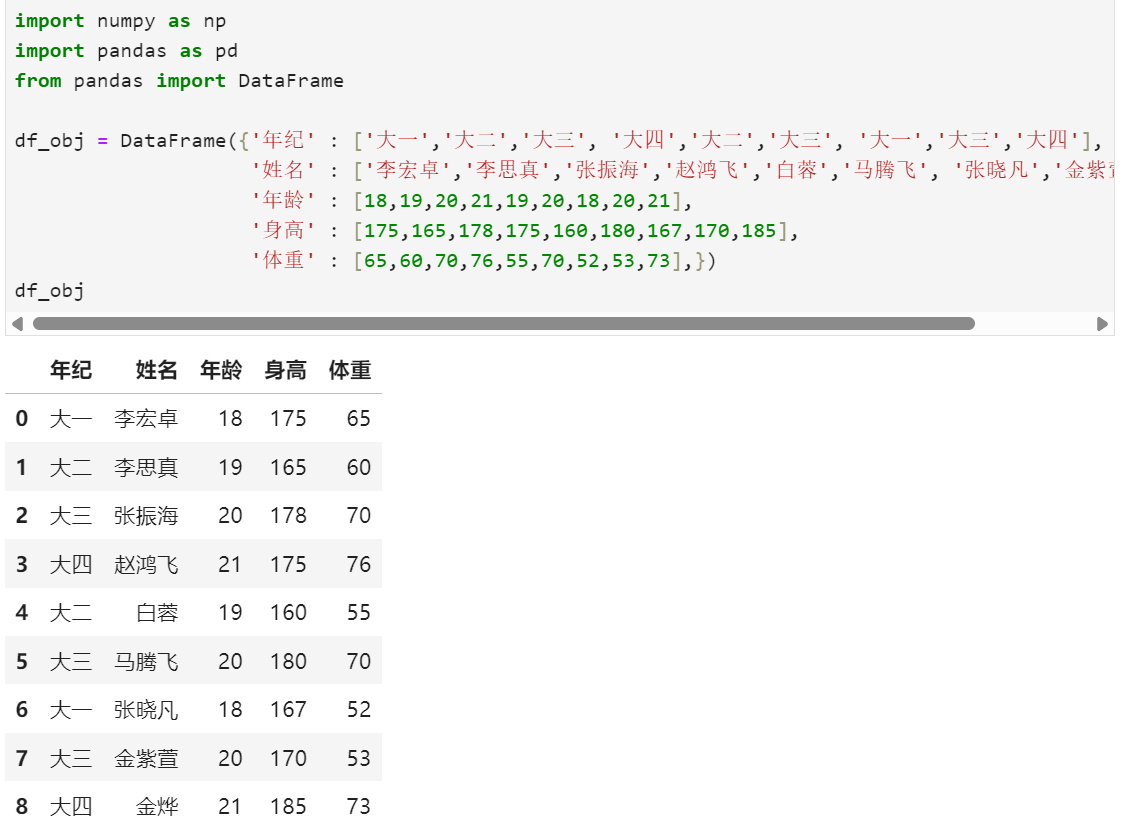
3.4D

3.5:C

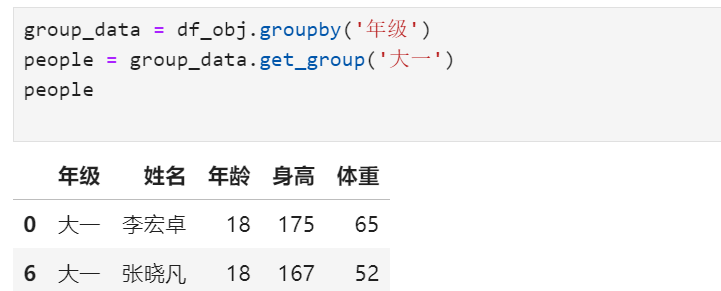
4.1： 分组的流程主要是根据，其对象的列行索引名称、key值、自定义的series对象索引、字典、函数等进行拆分开的，然后分组展示出来；聚合的流程是将每个分组应用统计运算，并把运算后的结果合并到一起的。

4.2: 其对象的列行索引名称、key值、自定义的series对象索引、字典、函数进行分组的。

5:（1）：



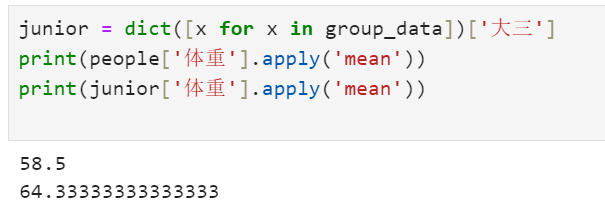
（2） ：



（3）：



（4） ：

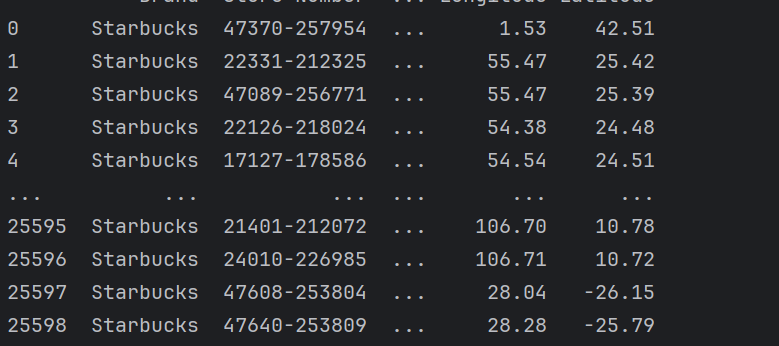


**Part 2:**

**Download the ‘Starbucks.csv’ file from the Moodle and complete the following Pandas practices and screenshot your solutions and results.**

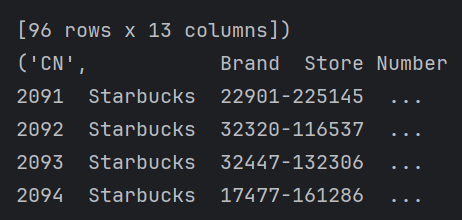
1. print所有宁波星巴克门店信息.

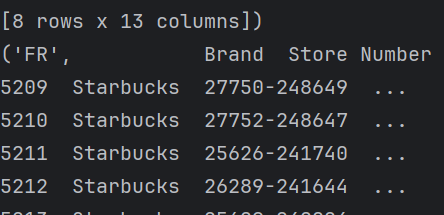
import numpy as np  
import pandas as pd  
from pandas import DataFrame  
df = pd.read\_csv('./Report 5-Starbucks.csv')  
print(df)



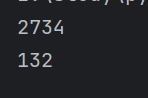
2. 分析法国的星巴克数量和中国的哪个多.

df = pd.read\_csv('./Report 5-Starbucks.csv')  
df = df.groupby(by='Country')  
  
for dfi in df :  
 print(dfi)



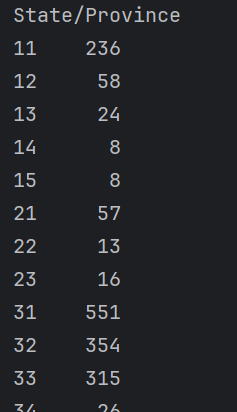


cou = df["Brand"].count()  
cn = cou['CN']  
fr = cou['FR']



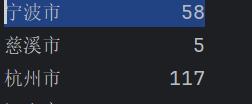
3. 分析中国每个省份星巴克的数量的情况.

df\_prv = df\_CN.groupby(by='State/Province')  
cou = df\_prv["State/Province"].count()  
print(cou)



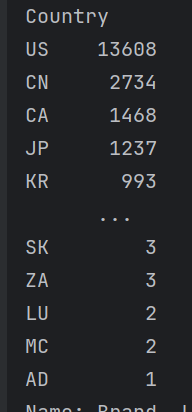
4. 分析宁波的星巴克和杭州的星巴克数量情况.

df\_prv = df\_CN.groupby(by='State/Province')  
df\_pro = df\_prv.get\_group('33')  
df\_city = df\_pro.groupby(by='City')  
cou = df\_city['City'].count()  
  
hangzhou = cou['杭州市']  
ningbo = cou['宁波市']  
  
  
print(hangzhou)  
print(ningbo)

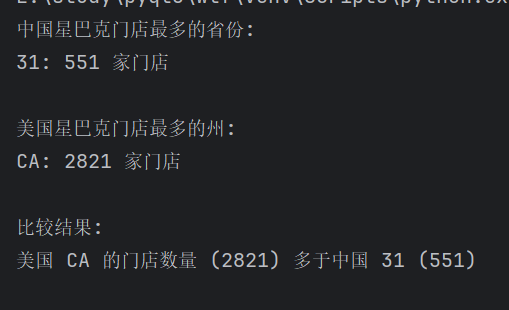
5. 哪个国家星巴克门店数量最多？哪个国家最少？.

df = pd.read\_csv('./Report 5-Starbucks.csv')  
df = df.groupby(by='Country')  
cou = df['Brand'].count()  
cou\_sorted = cou.sort\_values(ascending=False)  
print(cou\_sorted)



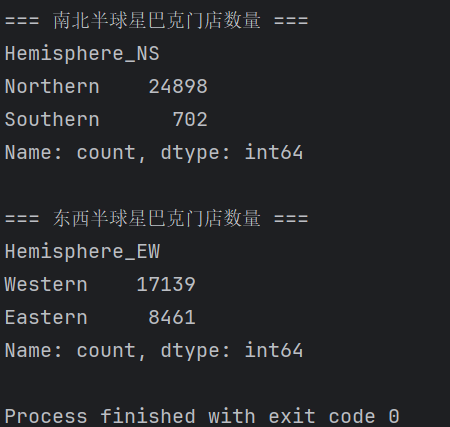
6. 比较中国星巴克门店最多的省份和美国星巴克门店最多州的数量.

china\_data = df[df['Country'] == 'CN'] # 中国的数据  
us\_data = df[df['Country'] == 'US'] # 美国的数据  
  
  
  
# 3. 统计中国各省的门店数量  
china\_province\_counts = china\_data['State/Province'].value\_counts()  
most\_province\_cn = china\_province\_counts.idxmax() # 门店最多的省份  
max\_count\_cn = china\_province\_counts.max() # 该省份的门店数量  
  
# 4. 统计美国各州的门店数量  
us\_state\_counts = us\_data['State/Province'].value\_counts() # 注意列名可能是 'State/Province' 或 'State'  
most\_state\_us = us\_state\_counts.idxmax() # 门店最多的州  
max\_count\_us = us\_state\_counts.max() # 该州的门店数量  
  
# 5. 比较结果  
print("中国星巴克门店最多的省份:")  
print(f"{most\_province\_cn}: {max\_count\_cn} 家门店")  
  
print("\n美国星巴克门店最多的州:")  
print(f"{most\_state\_us}: {max\_count\_us} 家门店")  
  
print("\n比较结果:")  
if max\_count\_cn > max\_count\_us:  
 print(f"中国 {most\_province\_cn} 的门店数量 ({max\_count\_cn}) 多于美国 {most\_state\_us} ({max\_count\_us})")  
elif max\_count\_cn < max\_count\_us:  
 print(f"美国 {most\_state\_us} 的门店数量 ({max\_count\_us}) 多于中国 {most\_province\_cn} ({max\_count\_cn})")  
else:  
 print(f"中国 {most\_province\_cn} 和美国 {most\_state\_us} 的门店数量相同 ({max\_count\_cn})")



7. 比较南北半球和东西半球星巴克门店数量.

df['Hemisphere\_NS'] = df['Latitude'].apply(lambda x: 'Northern' if x >= 0 else 'Southern')  
# 计算东西半球  
df['Hemisphere\_EW'] = df['Longitude'].apply(lambda x: 'Eastern' if x >= 0 else 'Western')  
# 统计南北半球的门店数量  
ns\_counts = df['Hemisphere\_NS'].value\_counts()  
# 统计东西半球的门店数量  
ew\_counts = df['Hemisphere\_EW'].value\_counts()  
print("=== 南北半球星巴克门店数量 ===")  
print(ns\_counts)  
print("\n=== 东西半球星巴克门店数量 ===")  
print(ew\_counts)



8. 哪个时区的星巴克门店最多？哪个时区最少？.

df\_timezone = df.groupby(by='Timezone')  
dfcou = df\_timezone["Brand"].count()  
# print(dfcou)  
most\_time = dfcou.idxmax()  
# most\_time = dfcou  
min\_time = dfcou.idxmin()  
print(f"最多数量的时区是 ： {most\_time}, 最少的失时区是{min\_time}")



*Ref：Pandas数据分组聚合案例（*https://www.isolves.com/hlw/dsj/2020-06-23/21345.html*）*