**Instructions for Student Query Tool**

**Yu Zheng**

**A20580847**

**Problem statement:**

The overall goal of your SQT program is to read student records from a file, and let the user enter queries to see a detail or summary report.

**1st step:**

I opened the txt file(student.txt), and then used pandas to transfer it into excel.

Table

Description automatically generated

Table

Description automatically generated

The result shows that this dataset contains 100 records of students, including ID, name, GradYear, GradTerm, and Program.

**2nd step:**

Query 1: Use pandas to display all student records.

**Table

Description automatically generated**

**3rd step:**

Query 2: Display students whose last name begins with a certain sting.

I used ‘lower()’ to convert all data in the ‘Last’ column and the strings that the user input to lowercase. The following result shows all the information of students whose last name begins with ‘s’.

Table

Description automatically generated

**4th step:**

Query 3: Display all records for students whose graduating year is a certain year.

The following result shows all students who graduated in 2021.

Table

Description automatically generated

**5th step:**

Query 4: Display a summary report of number and percent of students in each program, for students graduating on/after a certain year.

First, I used the "query()" function to select all MSBA students who graduated in 2019. Then, I used these students as "row2" and all 2019 graduates as "row1" and calculated the proportion of MSBA students among all 2019 graduates by dividing row1 by row2.

**Graphical user interface, text, application, email

Description automatically generated**

**Appendix:**

*# open txt file*

f\_path=r'/Users/zhengyu/Desktop/课件-S1/python/project/student.txt'

with open(f\_path,encoding = 'UTF-8') as f:

data = f.read()

print(data)

import pandas as pd

df = pd.read\_csv('/Users/zhengyu/Desktop/课件-S1/python/project/student.txt',sep='\t')

df.to\_excel('/Users/zhengyu/Desktop/课件-S1/python/project/studentframe.xlsx')

df

*# Display all student records*

students\_df=pd.DataFrame(df)

print(students\_df)

*#Display students whose last name begins with a certain string*

students\_df['Last'] = students\_df['Last'].str.lower()

search = input("Enter the name to search for: ")

search\_name = search.lower()

mask = students\_df['Last'].str.startswith(search\_name)

selected\_students\_df = students\_df[mask]

print(selected\_students\_df)

*#Display all records for students whose graduating year is a certain year*

import pandas as pd

search\_year = float(input('Enter the gradyear to search for: '))

student\_records = students\_df.loc[students\_df['GradYear'] == search\_year]

print(student\_records)

*#Display a summary report of number and percent of students in each program, for students graduating on/after a certain year*

search\_year = float(input('Enter the gradyear to search for: '))

search\_program = input('Enter the program to seach for: ')

print(students\_df.query(f'GradYear == {search\_year} and DegreeProgram == "{search\_program}"').groupby('GradYear')['ID'].count())

grads\_certain\_year = students\_df.query(f'GradYear == {search\_year}')

rows\_year,cols\_year = grads\_certain\_year.shape

grads\_certain\_program = students\_df.query(f'GradYear == {search\_year} and DegreeProgram == "{search\_program}"')

rows\_program,cols\_program = grads\_certain\_program.shape

value = rows\_program/rows\_year

percentage = value\*100

print("{:.2f}%".format(percentage))