```
# 1. The following questions use the data collected from the anonymous survey
collected at the beginning of the course and from previous semesters. The dataset
can be downloaded from the module on Canvas.
# Apply Pandas queries, manipulation, grouping, and aggregating to answer these
questions.
!pip install pandas
import pandas as pd
survey = pd.read_csv("surveydataFall2024.csv")
survey
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages
(2.2.2)
Requirement already satisfied: numpy>=1.22.4 in
/usr/local/lib/python3.10/dist-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.10/dist-packages (from pandas) (2024.2)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas)
(1.16.0)
```

	Semester	ComputerScience	Math	Statistics	MachineLearning	DomainExpertise	Communication
0	Spring2019	1.0	1.0	1.0	1.0	1.0	1.0
1	Spring2019	6.0	5.0	3.0	1.0	1.0	1.0
2	Spring2019	7.0	7.0	7.0	6.0	4.0	8.0
3	Spring2019	8.0	8.0	6.0	6.0	9.0	9.0
4	Spring2019	8.0	7.0	6.0	7.0	7.0	8.0
...	...	...	...	...	...	...	...
611	Fall2024	6.0	6.0	4.0	2.0	3.0	6.0
612	Fall2024	6.0	7.0	5.0	1.0	1.0	6.0
613	Fall2024	7.0	8.0	5.0	7.0	2.0	3.0
614	Fall2024	6.0	7.0	6.0	5.0	3.0	8.0
615	Fall2024	5.0	7.0	8.0	7.0	3.0	7.0

```
# a. Show only rows where Statistics skill is 10
survey[survey["Statistics"] == 10]
```

	Semester	ComputerScience	Math	Statistics	MachineLearning	DomainExpertise	Communication
51	Spring2019	7.0	9.0	10.0	7.0	8.0	10.0
64	Spring2019	10.0	10.0	10.0	10.0	7.0	10.0
85	Fall2019	8.0	9.0	10.0	8.0	5.0	8.0

	Semester	ComputerScience	Math	Statistics	MachineLearning	DomainExpertise	Communicati
116	Spring2020	8.0	9.0	10.0	1.0	5.0	7.0
134	Spring2020	8.0	8.0	10.0	5.0	7.0	9.0
321	Spring2022	7.0	10.0	10.0	5.0	7.0	10.0
353	Spring2022	4.0	10.0	10.0	1.0	8.0	9.0
389	Spring2022	3.0	7.0	10.0	1.0	2.0	5.0
395	Summer2022	10.0	10.0	10.0	10.0	10.0	10.0
430	Spring2023	5.0	8.0	10.0	1.0	3.0	6.0
474	Spring2023	7.0	10.0	10.0	1.0	2.0	10.0
517	Summer2023	8.0	10.0	10.0	1.0	3.0	8.0
568	Fall2024	10.0	10.0	10.0	8.0	3.0	8.0
596	Fall2024	10.0	10.0	10.0	10.0	5.0	10.0

```python
# b. Show only rows where both Statistics and Math skills are 10
survey[survey["Statistics"] == 10][survey["Math"] == 10]
```

```
UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  survey[survey["Statistics"] == 10][survey["Math"] == 10]
```

	Semester	ComputerScience	Math	Statistics	MachineLearning	DomainExpertise	Communicati
64	Spring2019	10.0	10.0	10.0	10.0	7.0	10.0
321	Spring2022	7.0	10.0	10.0	5.0	7.0	10.0
353	Spring2022	4.0	10.0	10.0	1.0	8.0	9.0
395	Summer2022	10.0	10.0	10.0	10.0	10.0	10.0
474	Spring2023	7.0	10.0	10.0	1.0	2.0	10.0
517	Summer2023	8.0	10.0	10.0	1.0	3.0	8.0
568	Fall2024	10.0	10.0	10.0	8.0	3.0	8.0
596	Fall2024	10.0	10.0	10.0	10.0	5.0	10.0

```python
# c. Show only the Semester, Communication, and Visualization columns where both
Statistics and Math skills are 10
survey[survey["Statistics"] == 10][survey["Math"] == 10][["Semester",
"Communication", "Visualization"]]
```

```
UserWarning: Boolean Series key will be reindexed to match DataFrame index.
  survey[survey["Statistics"] == 10][survey["Math"] == 10][["Semester",
  "Communication", "Visualization"]]
```

	Semester	Communication	Visualization
64	Spring2019	10.0	10.0
321	Spring2022	10.0	9.0
353	Spring2022	9.0	8.0
395	Summer2022	10.0	10.0

	Semester	Communication	Visualization
474	Spring2023	10.0	10.0
517	Summer2023	8.0	6.0
568	Fall2024	8.0	8.0
596	Fall2024	10.0	10.0

```python
# d. Show only the Semester, Communication, and Visualization columns in
decreasing order of Statistics and then Math skills
survey.sort_values(by=["Statistics", "Math"], ascending=False)[["Semester",
"Communication", "Visualization"]]
```

	Semester	Communication	Visualization
64	Spring2019	10.0	10.0
321	Spring2022	10.0	9.0
353	Spring2022	9.0	8.0
395	Summer2022	10.0	10.0
474	Spring2023	10.0	10.0
...	...	...	...
0	Spring2019	1.0	1.0
432	Spring2023	1.0	1.0
507	Summer2023	7.0	9.0
296	Fall2021	1.0	1.0
346	Spring2022	NaN	NaN

```python
# e. Show only rows where Statistics skill is missing
survey[survey["Statistics"].isna()]
```

	Semester	ComputerScience	Math	Statistics	MachineLearning	DomainExpertise	Communication
346	Spring2022	NaN	NaN	NaN	NaN	NaN	NaN

```python
# f. Show for every student, only their Semester, Math skill, Statistics skill,
and the maximum of their Math and Statistics skills (Hint: use axis=1)
survey[["Semester", "Math", "Statistics"]].assign(Max=survey[["Math",
"Statistics"]].max(axis=1))
```

	Semester	Math	Statistics	Max
0	Spring2019	1.0	1.0	1.0
1	Spring2019	5.0	3.0	5.0
2	Spring2019	7.0	7.0	7.0
3	Spring2019	8.0	6.0	8.0

	Semester	Math	Statistics	Max
4	Spring2019	7.0	6.0	7.0
...	...	...	...	...
611	Fall2024	6.0	4.0	6.0
612	Fall2024	7.0	5.0	7.0
613	Fall2024	8.0	5.0	8.0
614	Fall2024	7.0	6.0	7.0
615	Fall2024	7.0	8.0	8.0

```python
# g. Show the median value of Computer Science skills
survey["ComputerScience"].median()
```

7.0

```python
# h. Show the median value of Computer Science skills in Fall 2023
survey[survey["Semester"] == "Fall2023"]["ComputerScience"].median()
```

7.0

```python
# i. Show the median values of (each of) Computer Science, Math, and Statistics
# skills in every semester
survey.groupby("Semester")[["ComputerScience", "Math", "Statistics"]].median()
```

Semester	ComputerScience	Math	Statistics
Fall2019	7.0	7.0	5.0
Fall2020	7.0	7.0	7.0
Fall2021	8.0	7.0	6.0
Fall2023	7.0	7.0	6.0
Fall2024	6.0	7.0	6.0
Spring2019	7.0	7.0	6.0
Spring2020	8.0	7.0	7.0
Spring2021	7.0	7.0	6.0
Spring2022	7.0	7.0	6.0
Spring2023	7.0	7.0	6.0
Summer2021	7.0	7.0	6.0
Summer2022	7.0	6.5	5.0
Summer2023	7.0	7.0	6.0

```python
# j. Show the median values of (each of) Computer Science, Math, and Statistics
# skills in every semester of students who have taken CPSC483 (TakenCPSC483 is Yes)
survey[survey["TakenCPSC483"] == "Yes"].groupby("Semester")[["ComputerScience",
"Math", "Statistics"]].median()
```

	ComputerScience	Math	Statistics
Semester			
Fall2019	8.0	7.0	5.0
Fall2020	6.0	6.0	4.0
Fall2021	8.0	8.0	7.0
Fall2023	9.0	8.0	8.0
Fall2024	5.5	6.0	4.5
Spring2019	8.0	6.0	6.0
Spring2020	8.0	6.0	5.5
Spring2021	7.0	7.5	6.0
Spring2022	8.0	6.5	7.0
Spring2023	8.0	7.0	7.0
Summer2021	7.0	6.0	5.0
Summer2022	10.0	10.0	10.0
Summer2023	9.0	7.0	6.0

```python
# k. Show the median values of (each of) Computer Science, Math, and Statistics
skills for groups of students who have taken and not taken CPSC483 (Hint:
groupby)
survey.groupby("TakenCPSC483")[["ComputerScience", "Math",
"Statistics"]].median()
```

	ComputerScience	Math	Statistics
TakenCPSC483			
No	7.0	7.0	6.0
Yes	8.0	7.0	6.0

```python
# l. Show the number of students who took the survey in every semester
survey.groupby("Semester").size()
```

	0
Semester	
Fall2019	26
Fall2020	48
Fall2021	40
Fall2023	27
Fall2024	55
Spring2019	69
Spring2020	60
Spring2021	40
Spring2022	78
Spring2023	86

	0
Semester	
Summer2021	34
Summer2022	22
Summer2023	31