CSE514, Fall 2022, HW 1 Name: Student ID:
Note: This homework is worth a total of 15 points

Q1 (2pts): Refer to the "Lunch features" dataset to give an example of each data type:

Q1a: Numerical and Discrete:

Q1b: Numerical and Continuous:

Q1c: Categorical and Nominal:

Q1d: Categorical and Ordinal:

Q2 (5pts): You've been tasked with inputting the "Lunch features" dataset into a new database that can only accept numerical feature values. You must keep a minimum of 5 features in addition to price, but it's fine to leave null values for samples that do not have a feature value recorded. List the features you'll choose to keep and how you would process them for input:

Feature	Processing
Price	No processing necessary, just input the decimal value in dollar units
Q2a:	
Q2b:	
Q2c:	
Q2d:	
Q2e:	

Q3 (2pts): Identify a data quality problem in the Spring subset of the “Lunch features” dataset. Propose a method to handle it.

Q4 (6pts): Within the “Lunch features” dataset, the Spring subset has many more features than the Fall subset. To integrate the two into a single matrix, you could either drop all extra features from the Spring samples or add all the features to the Fall samples.
Answer three of the following with unique reasons:

Q4a: Why would dropping all extra features from the Spring samples would be a good idea?

Q4b: Why would dropping all extra features from the Spring samples would be a bad idea?

Q4c: Why would adding all extra features to the Fall samples would be a good idea?

Q4d: Why would adding all extra features to the Fall samples would be a bad idea?