Aditya Jain – [Jainadi5@msu.edu](mailto:Jainadi5@msu.edu) , A181237526

Question 4

Answer

1. Nominal
2. Ordinal
3. Ratio
4. Interval
5. Ordinal
6. Nominal
7. Ordinal
8. Interval

Question 5

Answer

1. SELECT ReleaseYear, Title FROM Movie WHERE Title = ‘Jurassic Park’ ;
2. SELECT Person.FirstName, Person.LastName

FROM Person, Movie

WHERE Person.ID = Movie.Director AND Movie.Title = ‘Jurrasic Park’ ;

1. SELECT Movie.Title

FROM Movie, Person

WHERE Movie.Director = Person.ID

AND Person.FirstName = ‘Steven’ AND Person.LastName = ‘Spielberg’ ;

1. SELECT count(\*) FROM Acts GROUP BY ActorID ;

Question 6

Answer

1. The query is selecting all the titles in the movie table that were released in the year 2017
   1. Number of columns = 1 (title)
   2. Minimum and maximum number of rows = 0 & 150 (as all values can or cannot be equal to ‘2017’)
2. The query is selecting first name and last name from the table person. Where person ID is equal to movie director and title of the film is Star Wars
   1. Number of columns = 2
   2. Minimum and maximum number of rows = 0 and 1
3. The query is selecting last name and first name from the table person. Where person ID is equal to actor ID and movie ID is equal to Acts movie ID and movie director is equal to person ID
   1. Number of columns = 2
   2. Minimum and maximum number of rows = 150,200
4. The query is selecting distinct first name and last name from the table person. Where person ID is equal to actor ID from first A1 and second A2 table, but movie ID is not equal to movie ID in the other table. That means, it produces all the names of the actors who has played role in more than one movie
   1. Number of columns = 2
   2. Minimum and maximum number of rows = 0,75

Question 7

Answer

1. Applying X → X − X¯ transformation will not affect both the discretization process and the bins will remain as it is. This is due to the reason that we are just subtracting a constant value mean thus actual range might change but the distribution will remain same.
2. Applying standardization transformation will affect equal width discretization process and will bring change in the assignment of elements, however there will be no change in the frequency discretization method as it solely depends on number of values
3. Applying standardize and exponentiated transformation will affect equal width discretization however there will still be no changes in equal frequency discretization.

Question 8

Answer

1. Using approach two average of weights = 1700/10 = 170

So, 6 and 12 will be assigned 170 weight using approach two

Using approach three

1. average weight of males = 1000/5 = 200 (assigned to patient 6)
2. average weight of females = 700/5 = 140 (assigned to patient 6)
3. Approach two will have same average value as approach one both equal to 170
4. Approach three is the most reasonable one out of the three approaches because of the following reasons:
   1. It considers other attributes into consideration while calculating averages, thus providing a better estimation criterion
   2. The average calculation considers to cluster data points on basis of gender giving us a better sense of estimation
5. By statistical literature, if the number of missing value cases are less than five % of the sample then the researchers can drop them.

Or else, in practical terms if the missing data is too limited in quantity, that removing them will hardly cause any changes in the dataset and in the expected outcomes then it is wise to remove missing data rather than imputing values.