Tasks

Task 1: (sample)

1) Open table Job_Large and select k-Anonymity as anonymization method.
2) Click Next.
3) Define attribute types. Read the instructions (Introduction and Getting Started) on right hand side. Think what type should each attribute (column) be and select right attributes. In case you are confused, here is the list of attribute types: name - ID, last name - ID, age - QID, gender - QID, city - QID and job - Sensitive.
4) Click next
5) You are now asked to specify attribute actions for each attribute. Note that some attributes are already preselected based on your choices in previous step. However you still have to specify attribute actions for age gender and city. Again, first thing that you need to do is read the explanations on right hand side. Think what action should be done to each attribute. In case you are stuck, here are actions: age - Generalize, gender - Keep as is (can you guess why?) and city - generalize.
6) Click next
7) You are now asked to specify interval size for values in column age. Again first thing to do is to read instructions of right hand side to get a better overview of what you need to do. Enter 10 and click Next.
8) Now you are asked to generalize city names. In real life situation, a more general value should be something that is common for all specific values. However in these tasks we care more about getting few groups that are about equal in size than about good general value.
8.1) Click on city names that could belong to one group and then click on Set generalization rule and enter some general value.
8.2) Repeat step 8.1 until you have no city names left.
In this sample task we are going to define the following rules:
Tallinn,Tartu,Paide->Large city
Viljandi,Türi,Jõgeva->Medium city
Vändra,Rapla,Põltsamaa,Elva->Small city
8.3) Click Next.
9) You should now see the resulting table. Since this is the sample task you do not have to submit result of this task. If you did everything correctly, you should see the following information about QIDs
Unique QIDs: 29
Smallest QID group: 2
Largest QID group: 14
Since the smallest QID group contains only 2 different people this table is only 2-anonymous.

question 1:
Does 2-anonymous table offer sufficient protection?
1) Yes, because generalizing QID values make it difficult to find specific person from given table.
2) No, because with 2-anonymous table sensitive attribute value can be easily guessed (with 50% confidence).

Task 2:
After solving first task, you ended up with having 2-anonymous table. However 2-anonymous table may not offer sufficient protection to privacy. Your task is to come up with a better solution. For example 5-anonymous table would be already better.
1) Follow steps 1-8 from first task. This time, let’s generalize gender values male and female to any gender.
Task 3:
By now you should have learned what to do when you need to increase the protection to privacy. Now let’s see an other factor that plays part in protection for table. In task 2, you anonymized table that had 200 records. Now let’s look at table that has 20 records.
1) Open table Job (not Job_Large)
2) Do everything else like you did in task 2 to increase protection to privacy (select same types, same actions and specify same generalization rules as you did in task 2.

Task 4:
We have so far anonymized tables where each row represents one person. However in some cases it could be that one person is present in multiple rows.
1) Go to Open table view.
2) On right side, scroll down to (X, Y)-Anonymity and read about (X, Y)-Anonymity.
3) Choose table Disease_xy.
4) This table is about different diseases that people have. One person can have one or more diseases. Using this hint choose proper anonymization method and click Next.
5) By now we expect that you already know what column types to choose. Choose right column type for each column and click next. If you have forgotten them, you can read about attribute types in right panel.
6) Choose proper actions for each attribute.
7) Generalize the values. Interval size for age: 10, all cities to estonia.

Task 5:
So far we have looked data sets that consist of single table, however in real life data is often stored in relational database.
1) Go to Open table view and on the right panel read about MultiRelational k-anonymity.
2) Select Job_Relational and Person_Relational.
3) MultiRelational k-anonymity is automatically selected for you. Click Next.
4) You are now asked to Join selected tables. Read the instructions on right hand side. Getting started will tell you what to do.
5) Add required rule(s). Click Next.
6) Specify attribute types and attribute actions. Hint: id is unique identifier, you dont have to worry about name and last name (Not Defined), you can choose to remove them (Remove column).
7) Generalize age to interval size of 10.

Task 6:
Introducing additional requirement for better privacy protection.
1) In Open table view, read about l-Diversity.
2) Select table Job_Large_Monotonous and anonymization method l-Diversity.
3) Specify attribute types and actions.
4) Use these generalization rules:
   Interval for age: 10
   City names:
   Tallinn,Tartu,Paide->Large city
   Viljandi,Türi,Jõgeva->Medium city
   Vändra,Rapla,Põltsamaa,Elva->Small city
   Hint: if you don’t wish to click on city names that much you can copy-paste these three lines above.

Task 7:
Introducing additional requirement for better privacy protection 2.
1) Do steps 1-4 from Task 6 but this time choose Job_Large instead of Job_Large_Monotonous.
Task 8:
Do the same as you did in previous task, except now select t-closeness instead of l-diversity. Read about t-closeness if you haven’t already.

Task 9
Sometimes generalizing QID attributes is not enough. For example in a company salary might be a confidential information. It is possible, that boss who knows QID and salary of his employees and wants to find out if any of his employees have participated in questionnaire about salaries. Since salary was declared confidential, participation in such list could harm employee’s relationship with employee.
1) In Open table view, read about $\epsilon$-Differential privacy.
2) Select table Job_large_Salary and anonymization method $\epsilon$-Differential privacy.
3) In attribute types specification view look at salaries.
4) Specify attribute types and actions.
5) generalize age and city as in previous tasks.
Hint: Job can be Non-sensitive this time.
Feedback questions

Thank you for using this software to learn privacy preserving data publishing. I would like to collect your feedback for this tool to evaluate its usefulness and usability. Please circle the option you would like to choose.

<table>
<thead>
<tr>
<th>Please indicate the extent to which you agree with following statement.</th>
<th>Strongly disagree</th>
<th>Strongly agree</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Using program helped to learn about privacy preserving data publishing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. Program helped to understand the difference between anonymization methods.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. Program was easy to learn.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4. Program was easy to understand.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. Program was easy to remember.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6. You would like to use this program to prepare for exam.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7. Instructions given by this program helped to solve the tasks.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8. Instructions given by this program were easy to follow.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. Instructions given by this program were easy to remember.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10. Tasks were easy to follow.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11. Solving tasks helped to understand privacy preserving data publishing better.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>12. Questions in quiz were easy to understand.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>13. Questions in quiz helped to learn privacy preserving data publishing.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>14. Statistics at the right of anonymized table were helpful for answering the questions in quiz.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>15. Statistics in this program were easy to understand.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>