Assignment 7

Schallner Ludwig 1850413, Wiegnand Andreas 1878334

Task 1)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Birthdate | Type | Treatment | We | He | Exp. d. | Telephone |
| Daniel Tinker | 23.08.1951 | Lung | Radiation | 83 | 182 | 2018-09 | 04917 4939 |
| Michael Riem | 05.01.1967 | Pancreas | Cytostatics | 91 | 182 | 2018-02 | 03492 48217 |
| Claudia Kling | 09.07.1961 | Gastric | Resection | 61 | 165 | Curable | 09201 34219 |
| Kurt Dreyer | 17.01.1956 | Lung | Radiation | 89 | 179 | 2019-01 | 03920 24928 |
| Sabine Kupfer | 17.05.1983 | Pancreas | Cytostatics | 71 | 155 | 2019-05 | 02389 23471 |
| Benjamin Reis | 23.08.1952 | Lung | Radiation | 87 | 177 | 2018-05 | 08704 19756 |
| Richard Kruse | 17.05.1971 | Pancreas | Cytostatics | 98 | 180 | 2019-04 | 04028 50285 |
| Julia Heim | 02.03.1964 | Gastric | Cytostatics | 65 | 172 | Curable | 03721 2391 |
| Karla Fried | 01.10.1965 | Pancreas | Resection | 68 | 169 | 2019-03 | 01294 42921 |
| Patrick Steiß | 04.03.1969 | Gastric | Resection | 95 | 184 | 2018-08 | 07293 1057 |

1. **Identifiers** (Name (because full name), Telephone (each Number is unique))

**Quasiidentifiers** (Birth date, We, He (on them one there are not uniquely identify the record owner but combined the will)

**Sensitive Data** (Type, Treatment, Exp. d. (all very sensitive data, no other then the patient and the doctor should know about)

1. Anyomized Dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Birthdate | Type | Treatment | We | He | Exp. d. | Telephone |
| \* | <1960 | Lung | Radiation | 8\* | >175 | 2018-09 | \* |
| \* | <1960 | Lung | Radiation | 8\* | >175 | 2018-05 | \* |
| \* | <1960 | Lung | Radiation | 8\* | >175 | 2019-01 | \* |
| \* | >1965 | Pancreas | Cytostatics | <80 | <175 | 2019-05 | \* |
| \* | >1965 | Pancreas | Cytostatics | >=90 | >175 | 2018-02 | \* |
| \* | >1965 | Gastric | Resection | >=90 | >175 | 2018-08 | \* |
| \* | >1965 | Pancreas | Cytostatics | >=90 | >175 | 2019-04 | \* |
| \* | 1960-1965 | Gastric | Resection | <80 | <175 | Curable | \* |
| \* | 1960-1965 | Pancreas | Resection | <80 | <175 | 2019-03 | \* |
| \* | 1960-1965 | Gastric | Cytostatics | <80 | <175 | Curable | \* |

**Name**: Identifier, have to be cleared,  
**Birtdate**: Have to generalized to that at least 3 of each generalization is common  
**Type**: Needed to still have information and its 3-anoymity  
**We**: like Birth date  
**He**: the same  
**exp d**. each row with curable have to be deleted because there is know way to 3-anoymized this without clearing the column   
**telephone**: is an Identifiers that’s why it have to be cleared

The **red one** have to be deleted, otherwise a 3 Anonymous Version of the original table isn’t possible

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Name | Birthdate | Type | Treatment | We | He | Exp. d. | Telephone |
| \* | <1960 | Lung | Radiation | 8\* | >175 | 2018-09 | \* |
| \* | <1960 | Lung | Radiation | 8\* | >175 | 2018-05 | \* |
| \* | <1960 | Lung | Radiation | 8\* | >175 | 2019-01 | \* |
| \* | >1965 | Pancreas | Cytostatics | >=90 | >175 | 2018-02 | \* |
| \* | >1965 | Gastric | Resection | >=90 | >175 | 2018-08 | \* |
| \* | >1965 | Pancreas | Cytostatics | >=90 | >175 | 2019-04 | \* |
| \* | 1960-1965 | Gastric | Resection | <80 | <175 | Curable | \* |
| \* | 1960-1965 | Pancreas | Resection | <80 | <175 | 2019-03 | \* |
| \* | 1960-1965 | Gastric | Cytostatics | <80 | <175 | Curable | \* |

1. The homogeneity attack: is an attack which trys to find homogeneity. For example if the attacker Alice knows that Bob is born <1960 or weights 80-89 and that Bobs record is in the table she can conclude that Bob has the Lung type.

Task 2)

Part 1)

1. Requests left: 1.191.923, right 1.192.045
2. Users: 199, Hosts (Webservers): 29469
3. Top 5 Hosts:
   1. [['http://static.cache.l.google.com"', 37780],
   2. ['http://www.google-analytics.com"', 28315],
   3. ['http://www.jetztspielen.de"', 22410],
   4. ['http://tbn0.google.com"', 21406],
   5. ['http://www.vtunnel.com"', 19507]]

Part 2)

Top List for candidates of Müller

**Assumption**. because Müller retrieves web page**s** if have to be at least 2 / day,

5 days >= **10 retrieved pages on that time interval** and filtering out the weekend (2-8 May 2009, 2 and 3 was the weekend) of that time interval

['10.1.2.80', 39] //most likely

['10.1.2.54', 38]

['10.1.2.62', 38]

['10.1.2.77', 38]

['10.1.2.76', 34]

['10.1.2.68', 33]

['10.1.2.60', 33]

['10.1.2.66', 33]

['10.1.2.73', 32]

['10.1.2.65', 32]

['10.1.2.52', 31]

['10.1.2.67', 31]

['10.1.2.61', 31]

['10.1.2.71', 31]

['10.1.2.58', 30]

['10.1.2.72', 30]

['10.1.2.75', 29]

['10.1.2.70', 29]

['10.1.2.64', 29]

['10.1.2.74', 29]

['10.1.2.63', 29]

['10.1.2.55', 28]

['10.1.2.79', 27]

['10.1.2.56', 27]

['10.1.2.59', 26]

['10.1.2.69', 25]

['10.1.2.78', 24]

['10.1.2.53', 23]

['10.1.2.57', 21]

['10.1.2.51', 20]

Part 3)

**Assumtion**: Because Julian informs herself **EACH** evining, count have to be at least close to the observed days (7)

['10.1.2.32', 6] //most likely (because of the assumtion)

['10.1.2.64', 3]

['10.1.2.61', 3]

['10.1.2.75', 3]

['10.1.2.79', 3]

['10.1.2.72', 3]

['10.1.2.52', 3]

['10.1.2.71', 2]

['10.1.2.165', 2]

['10.1.2.51', 2]

['10.1.2.187', 2]

['10.1.2.58', 2]

['10.1.2.66', 2]

['10.1.2.76', 2]

['10.1.2.70', 2]

['10.1.2.55', 2]

['10.1.2.59', 2]

['10.1.2.56', 2]

['10.1.2.69', 1]

['10.1.2.10', 1]

['10.1.2.94', 1]

['10.1.2.99', 1]

['10.1.2.81', 1]

['10.1.2.53', 1]

['10.1.2.37', 1]

['10.1.2.186', 1]

['10.1.2.78', 1]

['10.1.2.68', 1]

['10.1.2.40', 1]

['10.1.2.62', 1]

['10.1.2.140', 1]

['10.1.2.57', 1]

['10.1.2.101', 1]

['10.1.2.65', 1]

['10.1.2.9', 1]