D9 report: ICT SECURITY ANALYSIS

Kristiina Köster, Tõnis Rahe, Mai Ristioja

Project instructions: https://courses.cs.ut.ee/2020/ids/fall/Main/FirstStepsOfTheProject
Project repository: https://github.com/mairistioja/ICTsecurityAnalysis

Task 2. Business understanding

Identifying your business goals

Our team's motivation behind this project idea is to understand better the area of ICT security and it's aspects, how it is organized now and if there is any value at all to stress this field. Security breaches may lead to great losses to companies – losses of information, losses of money, losses of trust, etc. There is a possibility that different companies from different fields of life may benefit from our project if it can show to them what security measures the other companies have applied and how it correlates to the number of security breaches that they have experienced.

The business goal is that the companies would not bear losses because of deficient and low standard security measures.

Business success criteria might be the substantially smaller number of security breaches.

Assessing your situation

Our resources are us – a group of three students who have little experience this far (but we also have support from our instructors) and two datasets from Statistics Estonia database: "ICT security in enterprises by economic activity and number of persons employed" and "Presence of ICT specialists by economic activity of enterprise and number of employed persons". If necessary, we can also use "IV kv 2019. a tasutud maksud, käive ja töötajate arv" (2019 IV quarter paid taxes, revenues and number of workers) dataset from www.emta.ee website. We plan to work on our laptops and use Jupyter notebooks and Tableau Desktop (version 2020.3.3) software.

Requirement is to finish the project by 17th December and it must have sufficient technical and presentational quality.

Risks for the project involve lack of skills and lack of understanding, to alleviate that we plan to finish earlier rather than at the last minute and ask for help if needed. Another risk is just lack of time because of other large projects at university and as far as our team can, we use similar measures as before.

Important terminology:

- Information and communications technology (ICT) is an extensional term for information technology (IT) that stresses the role of unified communications and the integration of telecommunications and computers, as well as necessary enterprise software, middleware, storage and audiovisual systems, that enable users to access, store, transmit, and manipulate information.
- ICT security: the security of communication and information systems that handle, store or transfer classified information. It is determined by the protective measures that were taken to ensure the availability, integrity and confidentiality of these systems.
- Aggregate data is high-level data which is acquired by combining individual-level data.
- Correlation is a statistic that measures the degree to which two variables move in relation to each other.

The cost of this project is only time, it is planned to take up approximately 90 hours. Benefits are knowledge about ICT security measures and their effectiveness which may substantially reduce losses.

Defining your data-mining goals

Our project goal is to study how much the size and fields of activity influence practices of ICT security. We try to find correlations between different security measures and the volume of security breaches and also to show them in a visually pleasing and understandable way.

Our project is successful when it finds and shows correlations between different security measures and the volume of security breaches, thus gaining insight how effective different security measures might be. Final assessment is made by course instructors.

Task 3. Data understanding

Gathering data

Data requirements:

The values we use are floating point numbers, as time range we mainly use year 2019 (we might compare values from the ICT specialists dataset over the years 2014-2020) and data formats are percent and year.

Data availability:

Required data exists and is usable. At the moment, we cannot access raw values of individual companies. But even though we use only aggregated data, we can still statistically analyse it and show relevant results.

Selection criteria:

In this project we will use two main databases. The data in the both databases is aggregated and all values are shown as percentages. Specifically, the ICT security in enterprises and presence of ICT specialists are grouped by number of employed persons and by economic activity.

Details about the database:

Name: ICT security in enterprises by economic activity (emtak 2008) and number of persons employed

Link: http://andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=IT147

- Economic activity (28 activities + total of all activities)
- Number of employed persons: total; 10-19; 20-49; 50-99; 100-249; 250 and more
- Reference period: 2019
- Indicator: percentage of enterprises
- ICT security: 35 indicators

Details about the database:

Name: Presence of ICT specialists by economic activity of enterprise and number of employed persons

Link: http://andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=IT138

- Reference period: 2014-2019(2020)

- Economic activity: 28 activities + total of all activities
- Numeric data format: percentage of enterprises

Describing data

In both datasets fields of economic activities contain 29 categories:

- Total
- Agriculture, forestry and fishing
- Mining industries
- Processing industries
- Food and drink production
- Textile, garment and leather production
- Wood, paper and printing industries
- Coke, chemical, main pharmacy, rubber, plastic and non-metallic mineral production
- Metal and metal products industries
- Computer, electronics and optical appliance production
- Electrical and non-categorized machine and appliance production
- Motorized vehicles and other transportational devices production
- Furniture and other production, machine and appliance repair and installation
- Energy, gas, steam and conditioned air provision
- Water, sewage, waste and pollution management
- Construction
- Wholesale and retail trade, repair of motorized vehicles and motorbikes
- Transport and storage
- Accommodation and catering
- Information and communication technologies
- Info and communication
- Financial and insurance activities
- Real estate activities
- Vocational, scientific and technology related activities
- Management and helping activities
- Education

- Healthcare and social welfare
- Art, entertainment and free time
- Other service activities

First database: ICT security in enterprises by economic activity and number of persons employed

- **Link:** http://andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=IT147

Unfiltered dataset has 5391 rows altogether. In our team project we filter the rows in several ways, group the size of the companies together showing values by fields of activities separately (981 rows) as well as group the fields of activities together showing values by size of the companies (176 rows). Size of the companies is shown in column "Tööga hõivatud isikute arv" (Number of people occupied) and the ranges are total; 10-19; 20-49; 50-99; 100-249; 250 and more. Estonian companies with more than 50 people are all represented in the dataset, but are, however, grouped together. Samples were taken from smaller companies with at least 10 employees. Thus if we want to compare sizes of different groups, we can find the numbers out for bigger companies only from Estonian tax authority dataset (this might contain some imperfections as the timeranges coincide but are not exactly the same). It would be helpful if we want to compare the rows, because otherwise one row may represent 2 companies and other row 50 companies and comparison may likely be unreliable.

"Security" column contains 35 indicators:

- Strong password authentication
- Using the most recent software versions
- Using biometrical method to identify and authenticate user
- Encrypting data, documents and emails
- Data backup in different location
- Network access control
- Using VPN

- Copying and storing log files about security incident for analysing
- Executing ICT risk analysis
- ICT security tests
- Use of ICT security methods
- Voluntary participation in ICT security trainings, forwarding internal knowledge
- Compulsory trainings, reading of compulsory materials
- ICT security aspects are fixed in contracts
- Employees are being informed about ICT security in different ways
- Employees are not being informed about ICT security concerns
- ICT security activities are performed by company's own employees
- ICT security activities are performed by service provider outside of the company
- ICT security activities are performed by company's own employees and by service provider outside of the company
- ICT activities including ICT security activities are performed by company's own employees
- ICT activities are performed by company's own employees, ICT security activities are performed by service provider outside of the company
- ICT activities are performed by service provider outside of the company, ICT security activities are performed by company's own employees
- ICT activities including ICT security activities are performed by service provider outside of the company
- Documents about evaluation, practices and procedures of ICT security exist
- There are ICT specialists and documentation about evaluation, practices and procedures of ICT security
- Security policy was last changed in last 12 months
- Security policy was last changed 12 to 24 months ago
- Security policy was last changed more than 24 months ago
- Security policy was last changed in last 24 months
- Experienced ICT security incident in last year, ICT services were not available
- Experienced ICT security incident in last year, data destruction or corruption
- Experienced ICT security incident in last year, leaking of confidential data
- Experienced at least one ICT security incident in last year
- Experienced no ICT security incidents in last year

- Insurance against ICT security incidents

Second database: Presence of ICT specialists by economic activity of enterprise and number of employed persons

- **Link:** http://andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=IT138

Unfiltered dataset has 4097 rows altogether. We also use filtering of the rows based on size and fields of activity.

The column "Indicator" contains five indicators:

- Enterprises with ICT specialists
- Enterprises which provided ICT training for ICT specialists in the previous year
- Enterprises which provided ICT training for employees in the previous year
- Enterprises which recruited or tried to recruit ICT specialists in the previous year
- Enterprises that had difficulty recruiting ICT specialists the previous year

Altogether our data is suitable for our project's goals, it contains values and indicators that we need and sufficient amounts of rows.

Exploring data

- It specialists have a greater role to play in companies in the fields of information and communications and information and communications technology.
- Companies with ICT specialists have been involved in a bit more than 20% of all companies in different fields.
- Most companies that tried to recruit ICT specialists were having difficulty recruiting.
- Companies in the areas of information and communications technology, together, tried most to recruit ICT specialists, more than half of the companies, and therefore they were also most in difficulty in recruiting.

- The share of ICT specialists in the total number of information and communication, information and communication technology and computer, electronics and optical equipment manufacturing companies is more than 50%.
- In 2018, at least 10% of businesses in most areas offered their employees ICT training.
- Among the companies in the field of information and communications, information and communication technologies and financial and insurance activities are the most companies that offer ICT training to their employees. The share of companies is more than 40%.
- Nearly 90% of all companies in different sectors were exposed to no ICT security incident last year.

Hypothesis: Information and communication and the role of companies in the financial and insurance sectors in the use of various security measures is greater than the use of security measures in all other areas.

Hypothesis: The impact of ICT security incidents is greater in companies with low security measures.

Verifying data quality

The data we use in this project is high-quality. Statistics have been made by specialists from The Statistics Office and Estonian state uses this data for state management.

Task 4. Planning your project (0.5 points)

Project plan and list of tasks:

- Preparing data in Microsoft Excel 2016 for use in Tableau Desktop: 5h (mostly done by Mai Ristioja).

- Getting the numbers of different companies from the Estonian Tax and Customs Board (Eesti Maksu- ja Tolliamet) dataset and grouping them by size and by fields of activity using Jupyter Notebook: 5h (Tõnis Rahe).
- Doing visualisations using Tableau Desktop: 27h (Mai Ristioja 15h, Tõnis Rahe 12h).
- Finding correlations and testing them using permutation tests: 39h (Mai Ristioja 11h, Kristiina Köster 22h, Tõnis Rahe 6h).
- Preparing the final report (possibly using Canva): 18h (Kristiina Köster 9h, Tõnis Rahe 9h)

Methods and tools that we plan to use:

Our team plans to use permutation tests to find correlations of attributes.
Our team plans to build strong visual analysis of existing data using Tableau.