*Agata Julia Rapiej, Student Number 2086695, a.j.rapiej@tilburuniversity.edu Johan van Boven, Student Number 2072347,j.k.vanboven@tilburguniversity.edu Kas Kniesmeijer, Student Number 2081456, k.l.kniesmeijeruniversity.edu*

*Team Project*

Interactive Data Transformation Group 18



Table of Contents

[Data Description 2](#_Toc104305231)

[Description of Analytic Tasks 3](#_Toc104305232)

[Task 1 3](#_Toc104305233)

[Task 2 3](#_Toc104305234)

[Task 3 3](#_Toc104305235)

[Data Collection 4](#_Toc104305236)

[Results 5](#_Toc104305237)

[Result Task 1 5](#_Toc104305238)

[Result Task 2 5](#_Toc104305239)

[Result Task 3 5](#_Toc104305240)

[Bibliography: 7](#_Toc104305241)

# Data Description

Nowadays, healthcare plays a significant role of our prosperous society. Health insurance helps people from the financial risks of unanticipated health expenses. In the 21st century, healthcare is increasingly dependent on technology. The cost of medical technology keeps on rising and also its use increases, which contributes to the ever-growing healthcare costs (Kumar, 2011). As a result, health insurance companies are looking for new insights into healthcare costs using Big Data.

The selected data constitutes to Big Data because it fulfills three key components:

**Volume**: In the field of healthcare, the speed at which new data is being generated is increasing rapidly.  
**Velocity**: The need for generated healthcare data to be processed and analyzed is also increasing extremely fast.   
**Variety**: Healthcare institutions track and store all sorts of data types: structured, unstructured and semi-structured data.

# Description of Analytic Tasks

## Task 1

Description: Health insurance companies want to gain insight into the effect of heavy smoking and obesity on healthcare costs for people below 30.

Analytic Research Question: Does heavy smoking ( > 10 cigarettes a day) and obesity (BMI > 30) result in healthcare costs for people below the age of 30?

*Variables: Amount of cigarettes smoked per day, BMI, Age, Healthcare costs*

## Task 2

Description: Health insurance companies want to gain insight into the effect of high blood pressure and high heart rate on healthcare costs for people with a low monthly income.

Analytic Research Question: Does high blood pressure ( >140/mm Hg ) and a high heart rate ( > 100bpm ) result in higher healthcare costs for people with a low monthly income ( < €2000 )?

*Variables: Blood pressure, Heart rate, Monthly income, Healthcare costs*

## Task 3

Description: Health insurance companies want to see if physical activity impacts healthcare costs compared to non-sporters.

Analytic Research Question: Does exercising (minimal of 1 hour per week) have a positive impact on the average monthly healthcare costs, compared to non-sporters? Is monthly income a possible reason why non-sports don’t sport (is the monthly income for sports sig. Higher than for non-sporters).

*Variables: Weekly sporting hours, Healthcare costs*

# Data Collection

The data generated from the medical database consists of qualitative and quantitative data. To perform a successful analysis of health insurance costs based on several diseases and lifestyles, certain data attributes are needed.

These include gender, age, children, cigarettes per day, weekly sporting hours, monthly charges for health insurance, monthly income, BMI, SBP, and BPM.

These factors contribute to the research because they provide a pattern, such as a high number of cigarettes a day or low weekly hours spent on exercising, resulting in probable high health insurance costs due to the increase in bad lifestyle behavior.

By exploring these patterns, we can investigate the effect of these data attributes on people’s health insurance costs.

# Results

## Result Task 1

The results returned by the code from task 1 consist of two means of monthly healthcare costs over two groups.

* Group 1 has an age below 30, a BMI-value of over 30 and smokes more than 10 cigarettes per day.
* Group 2 has an age below 30, a BMI-value of 30 or lower, and smokes 10 or fewer cigarettes per day.

Group 1 has an average monthly cost of €151,56.

Group 2 has an average monthly cost of €144,60.

These results show that heavy smoking and obesity have an negative effect on healthcare costs for people below the age of 30.

## Result Task 2

The results from task 2 consist of two groups with different health insurance costs.

* Group 1 showed that people that have an SPB larger than 140, a BPM higher than 100, and a monthly income lower than €2000 a month, pay on average €149,16 per month for health insurance.
* Group 2 showed that people with an SBP greater than 140, a BPM lower than 100, and a monthly income lower than €2000, paid on average €147,65 for health insurance.

It can be concluded that people with high SBP and BPM pay more for health insurance in comparison with people with low SBP and BMP with income lower than €2000 per month.

## Result Task 3

The code for task 3 for the average “Monthly\_Charges” is splited into two groups:

1. Sporters (weekly\_sporting\_hours > 0)
2. Non sporters (weekly\_sporting\_hours = 0)

The group sporters have an average of €148,38 of monthly healthcare costs (Monthly\_Charges).

The non-Sporters group have an average of €147,85 of monthly healthcare cost (Monthly\_Charges).

The results show that the average monthly healthcare costs are higher for sporters than for non-sporters.

# Bibliography:

Kumar R. K. (2011). Technology and healthcare costs. *Annals of pediatric cardiology*, *4*(1), 84–86. <https://doi.org/10.4103/0974-2069.79634>