**Module: Digital Health (WS 23/24)**
**Lecturer: Prof. Dominik Böhler**

**Deggendorf Institute of Technology**
**Faculty European Campus Rottal-Inn**
**Master of Global Public Health**

**Student Name: Sahar Shawqi Mohammed Azhar Ramadhan**

**Date of Submission: 12th July, 2024**

## The Use Case: Problem Description and Stakeholders

- **Medical Insurance Cost Prediction**

A vital component of controlling individual health expenses and mitigating financial risks associated with medical care is health insurance. Insurers take into account a variety of lifestyle and demographic factors when determining premiums in order to guarantee equitable pricing and accessibility. This study examines a dataset that includes characteristics like geography, insurance costs, number of children, age, sex, body mass index (BMI), and smoking status. Our objective is to examine these factors to understand how they affect insurance prices, spot important patterns, and offer information that will help legislators and insurers make decisions.

- Justification:

Understanding the relationship between lifestyle and demographic characteristics and insurance costs is essential for several reasons:

Fair Pricing: Rather than basing fairness on presumptions or broad guidelines, insurers must establish premiums that accurately represent the risk associated with everyone.

Policy Development: By using the data's insights, policymakers can create policies that improve accessibility and equity in the insurance industry. Reforms in policy may result, for example, from the identification of differences in rates among various locations or demographic groupings.

Public health insights: Public health programs can be targeted to reduce high-risk behaviors and encourage healthier lifestyles, potentially lowering overall healthcare costs. This is possible by researching the effects of characteristics like smoking and BMI on insurance prices.

Consumer Awareness: Encouraging healthier habits and well-informed decision-making can be achieved by educating consumers on how their demographics and lifestyle choices affect their insurance costs.

- Stakeholders:

Insurance Companies: these organizations have a strong interest in learning how different variables affect risk and expense because it helps them establish fair premiums and efficiently manage financial risk.

Policyholders: By understanding the variables influencing their insurance prices, both present and potential insurance clients can make better financial and health-related decisions.

Regulators and Policymakers: To create and implement regulations that guarantee equitable procedures in the insurance sector and shield customers from unfair pricing practices, governmental organizations and regulatory bodies need comprehensive data.

Healthcare Providers: By understanding the financial effects of patient demographics and lifestyle decisions, physicians, hospitals, and other healthcare organizations may be able to better inform their recommendations for services and care.
Public health officials: By identifying patterns and connections between insurance prices, health behaviors, and demographics, they can utilize this data to guide public health programs and policies.

Researchers and Analysts: To learn more about health economics, insurance markets, and the socioeconomic aspects impacting healthcare costs, academics and market analysts can make use of this dataset.

Data Description: The following variables are included in the dataset:
Age: The person's years of age. Sex: The person's gender (male or female). Body Mass Index, or BMI, is a weight-and-height-based indicator of body fat. Children: The total number of kids that the insurance plan covers. Smoker: The person's status as a smoker (yes or no). Region: The geographic area (northeast, northwest, southeast, southwest) in which the person dwells. Charges: The individual's insurance-related costs.

- Conclusion:

The objective of this study is to examine the dataset and identify the correlations between these factors and insurance costs. By doing this, we seek to support a just and efficient health insurance system by providing insightful information to insurers, legislators, and other stakeholders.

## ➢ The Pathway to the Solution

- Identify a problem:

Accurately predicting insurance costs is a major issue with broad ramifications for policyholders and insurance providers. For that reason, we need to understand the problem, study the prevalence, determine the associated factors and find a solution by making a root cause analysis and collaborative cooperation among all stakeholders.

Accurate cost projections can lessen the burden of audits, shield the public from negative effects, and keep insurers solvent. It enables insurers to appropriately set premium prices depending on personal risk variables. Complex statistical models and techniques that insurance companies keep confidential are used to calculate insurance premiums. Predicting these premiums with accuracy can help insurers become more accountable and transparent.

- Conducting research

A literature review for scholarly articles was conducted using electronic databases such as Kagel, ResearchGate, Google Scholar, and PubMed. The key words and phrases used for the search included: medical insurance cost prediction", "common involved variables, such as, age, sex, BMI, region," "policy makers and regulators", "root cause analysis", "rates" " insurance markets" and "health behaviors".

- Assessing the data:

A more comprehensive review of the publications and bibliographies that demonstrated an increase in the cost of medical insurance in relation to various factors like age, sex, BIM., children, geography, and smoking status.

-Understand the role of socioeconomic, cultural factors and lifestyle in determining insurance costs and managing the global increased cost prediction.

-Examine the features of the dataset, such as age, sex, BMI, children, smoking status, region, and the goal variable (insurance charges), to gain an understanding of it.

To find trends and connections between features and the target variable, visualize the data. To see how each feature affects insurance costs, for instance, make scatter plots, bar plots, and histograms.
Execute feature scaling, categorical variable encoding, and any other required data modifications.

- Identifying the stakeholders:

It is crucial to identify stakeholders in medical insurance costs to comprehend the range of viewpoints and interests present. The goals of various stakeholders, including employers, insurers, physicians, consumers, and government agencies, all affect how much medical care costs and distribute available resources.

- Prioritizing stakeholders and allocating funds wisely to meet their issues are made easier with the use of stakeholder analysis.

- The data assessment also identified problems like research gaps and current inequities, which emphasize the necessity for methodical approaches to stakeholder management that balance a range of interests to make the best possible policies.

The healthcare system is complicated, with various internal and external players spanning different sectors and levels and fostering excellent connections and open communication with key stakeholders is critical for effective healthcare projects.

- Solve the problem:

Studies have shown that there is a global increase in medical insurance cost particularly after the COVID-19 pandemic, and a need for a holistic approach to address the challenges. Here a few steps that need to be taken to achieve the collaborative cooperation between the stakeholders, to face the burden of the current increasing problem:

- The combination of strategies like, guaranteeing effective use of healthcare resources and minimizing needless or low-value care, utilization management programs include prior permission requirements, step therapy protocols, and case management can help limit costs.

-Although Costs can be controlled by negotiating reduced payment rates with providers, designing tier-specific or constrained provider networks, and directing patients toward affordable, high-quality providers.

-Higher cost-sharing leads to lower healthcare spending without significantly impacting health outcomes for most populations.

- In relation to the health sector, preventing or effectively managing chronic illnesses, investments in population health efforts, wellness programs, and chronic disease management can lower long-term healthcare expenditures and enhance overall health.

-In general, the collaborative work between the associated partners including health services consumers and promoting the usage of telemedicine in order to provide more convenient and cost-effective access to health care services can play a significant role in managing the growing health insurance costs.

## ➢ The Implementation Process

Description of the detailed steps and settings in the tools used to show how the interactions and data generation are facilitated by the system you developed: do also talk about the journey from your first prototype to the final version, incl. all wrong pathways you took.

- Part One:

1. First step was to study all tutorials available during the course about R language ( As a beginner who does not have any background or enough experience in the field of programming, studying the tutorial courses was a frustrating thing and it kind of formed a barrier for me, but through continuous effort and following the existing steps, and besides looking at YouTube, I was able to overcome this barrier and found myself interested in learning R language, especially when I was getting results that made me feel constantly motivated.
2. Downloading and installing RStudio and R from the interactive R tutorial, as well as all packages needed progressively during the journey
3. Selecting a suitable data set from Kaggle and downloading it onto a desktop folder to create a working directory. (In the beginning, for the purpose of learning and training, I chose a dataset from Kaggle, concerned with a public health issue, regarding to the

level of obesity, and I used chat GPT to get a database limited to five countries, although to obtain codes of visualization.

4. For the exam I chose from "Kaggle "in CSV file form, the dataset of " Medical Insurance Cost Prediction".
5. I studied the data to get a preliminary review of what may be suitable to visualize the data with (e.g.: bar chart, scatterplot or other tolls of visualization).
6. I imported the medical insurance cost prediction CSV file on RStudio or we can use the function read.csv() to import the dataset.
7. I used the ChatGPT to get the preliminary codes of visualization, that took from me a long time to decide which visualization is the best to provide a suitable clarification, although to do modification based on my study of tutorial, handbook, and YouTube classes.
8. On my way i did a mistake by creating more than one project in the same folder on the desktop ,trying to avoid errors which caused failures in file reading when I was restarting the RStudio to work again, as well as at end made the running of visualization on shiny impossible, indicating that the csv file directory and path is not correct ,so I started from the beginning and created a new project in the document folder and I changed the path directory to read the csv from : medical_insurance <- read.csv("C:/Users/Sarah/Desktop/Digital Health/Digital health3/medical_insurance.csv") to medical_insurance <- read.csv("C:\Users\Sarah\Documents\New Rproject\medical_insurance.csv") and I changed (\) to (/).

Here is my data visualization steps, as the script is named "Draft1" and saved in new project folder:

1- Installing different packages and loading them using the library function.
2- Reading in the data.
3- Interacting with data functions and packages to create data frames
4- Creating plots and scatter plots with regression line.

```
# Load necessary libraries

library(ggplot2)

  # Load the data from a CSV file

medical_insurance <- read.csv("C:\\Users\\Sarah\\Documents\\New Rproject\\medical_insurance.csv")

# View the data

View(medical_insurance)

 # Create the data frame directly for demonstration purposes


# Visualizations

ggplot(data, aes(x = charges)) +
```

```
+  geom_histogram(binwidth = 5000, fill = "blue", color = "black")
  labs(title = "Distribution of Insurance Charges", x = "Charges", y = "Frequency")

ggplot(data, aes(x = smoker, y = charges)) +
 + geom_boxplot(aes(fill = smoker))
  labs(title = "Insurance Charges by Smoker Status", x = "Smoker", y = "Charges")

ggplot(data, aes(x = bmi, y = charges, color = smoker))  +
geom_point () +
  labs(title = "BMI vs Insurance Charges by Smoking Status", x = "BMI", y = "Charges")

ggplot(data, aes(x = region, y = charges))  +
+  geom_boxplot(aes(fill = region))
  labs(title = "Insurance Charges by Region", x = "Region", y = "Charges")

 Scatter plot with regression line#
ggplot(data, aes(x = bmi, y = charges))  +
 + () geom_point
  + geom_smooth(method = "lm", col = "blue")
  labs(title = "Relationship between BMI and Insurance Charges", x = "BMI", y = "Insurance
Charges")
```

- Part Two: " Shiny ":

- Creating an account on Shiny.io and linking it to Rstudio through the Tokens.

- Creating a Shiny web application file on Rstudio called " Shiny " and saving it in the new project folder, in document folder with the original csv file from Kaggle.

-Loading the required packages including the shiny package and preparing the data to plot a graph.

- Defining the UI for the Shiny application and defining the server logic.

- Produce a comparative scatterplot showing the relation between the insurance charges and different variables like age, sex and BMI.

Runing the application and publishing the app.R code in the script file to Shiny. -

- Testing the link generated in the Shiny application to check that it opens on the browser or not ( I faced at the beginning this problem that the link after publishing is not opening ,but I tried several times with the help of ChatGPT to discover the reason ,and recommended the installation of RTools (RTools: Toolchains for building R and R packages from source on Windows) . At the end after several times and starting again to create app.R file it worked successfully.

- Creating a repository on github account and uploading all the documents and URL.

- References: -

1-Kaushik, K., Bhardwaj, A., Dwivedi, A.D. and Singh, R., 2022. Machine learning-based regression framework to predict health insurance premiums. International Journal of Environmental Research and Public Health, 19(13), p.7898.

2-Orji, U. and Ukwandu, E., 2024. Machine learning for an explainable cost prediction of medical insurance. Machine Learning with Applications, 15, p.100516.

3- https://www.kaggle.com/code/shubhamptrivedi/health-insurance-price-predict-linear-regression

4-https://datauab.github.io/medical_insurance/

5-Mohamadi, E., Olyaeemanesh, A., Rashidian, A., Hassanzadeh, A., Razavi, M. and Foroushani, A.R., 2018. Stakeholders analysis of health insurance benefit package policy in iran. Health Scope, 7(2).

6- Xian, Y. and Yaeghoobi, E., 2018. Utilising LinkedIn as an effective marketing channel to engage with and attract potential customers

7-Frech III, H.E., 2009. Cost-Sharing Enhances Cost Control. American Health & Drug Benefits, 2(6), p.237.

8-Morreim, E.H., 1987. Cost containment and the standard of medical care. Calif. L. Rev., 75, p.1719.

9-Temple, J. and Ludwig, B., 2010. Implementation and evaluation of carousel dispensing technology in a university medical center pharmacy. American Journal of Health-System Pharmacy, 67(10), pp.821-829.