**RISK ANALYISIS FOR LIFE INSURANCE POLICY WITH MACHINE LEARNING**



**UNIVERSITY OF EMBU**

**BACHELOR OF SCIENCE**

in

**COMPUTER SCIENCE**

by

**(MEGANE ANJAVO MULEMI)**

**(B135/18669/2020)**

Project Supervisor

**Dr. Boaz Kipyego Too**

Department of Computing and Information Technology

60100-06 Embu, Kenya. April – 2022

## DECLARATION

The work contained in this project proposal document has not been previously documented in any other higher education institutions.

I certify that the intellectual content of this project is the product of my work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Student; Supervisor;

Signature………………… Signature………………...

Date…………………

## DEDICATION

I dedicate this project to the Almighty God, my source and inspiration, wisdom, knowledge and understanding. He has been my strength throughout this program an on his wings I soar. I also wish to dedicate this project to my family as they have been my encouragement and made sure I gave it my all.

## ACKNOWLEDGEMENT

I would like to express my sincere gratitude to the Almighty God for giving me strength, life and good health to complete this work. I would also like to thank my supervisor Mr. Boaz Too whose incisive, ready and constructive critiques of the proposal’s progress have been valuable. My supervisor has been remarkably patient, providing consistent guidance, constructive comments, suggestions and helpful advice during successive stages of this project. My special gratitude also extends to my classmates and all research contacts who have been very helpful and cooperated effectively during the research project

# **ABSTRACT**

Life insurance provides trillions of dollars of ﬁnancial security for millions of individuals and families worldwide. Life insurance companies need to accurately assess individuals’ risk to determine their premiums. The traditional underwriting process used to assess this risk is based on manually examining an applicant’s health, behavioral, and ﬁnancial proﬁle. The existence of large historical data sets provides an unprecedented opportunity for machine learning to transform underwriting in the life insurance industry. The main objective of this study is to predict the amount of premiums and individual should be able to pay either monthly or annually depending on factors like age, occupation, lifestyle, health conditions and the environment in which the insured lives. The target population are customers between the ages of 31-45 since it consists of young people most of them being in need of life insurance. Data collection, processing and analysis will be determined through predictive analytics using linear regression to do analysis of large data sets (big data) to make inferences or identify meaningful relationships, and the use of these relationships is to predict the future events. The study recommends that insurance companies should push and market policies that provide for both risk coverage and savings component because that’s what the customers prefer. The insurance companies should also consider lowering the cost of premiums, improve on agent’s integrity, improve on customer service and develop new product varieties.

Table of Contents

[KABARAK UNIVERSITY 1](#_Toc80948090)

[DECLARATION 2](#_Toc80948091)

[DEDICATION 3](#_Toc80948092)

[ACKNOWLEDGEMENT 4](#_Toc80948093)

[ABSTRACT 5](#_Toc80948094)

[CHAPTER 1 8](#_Toc80948095)

[INTRODUCTION 8](#_Toc80948096)

[1.1 BACKGROUND STUDY 9](#_Toc80948097)

[1.2 PROBLEM STATEMENT 11](#_Toc80948098)

[1.3 RESEARCH OBJECTIVES 11](#_Toc80948099)

[1.4 RESEARCH QUESTIONS 12](#_Toc80948100)

[1.5 STUDY JUSTIFICATION 12](#_Toc80948101)

[1.6 SCOPE OF STUDY 12](#_Toc80948102)

[1.7 LIMITATIONS 12](#_Toc80948103)

[CHAPTER 2 13](#_Toc80948104)

[LITERATURE REVIEW 13](#_Toc80948105)

[2.1 INTRODUCTION 13](#_Toc80948106)

[2.2 CRITIQUE OF THE EXISTING LITERATURE RELEVANT TO THE STUDY 14](#_Toc80948107)

[2.2.3 GETTING DATA ISSUES RIGHT 15](#_Toc80948108)

[2.2.4 NEW PRODUCTS AND SERVICES 15](#_Toc80948109)

[2.3 REVIEW OF THEORETICAL LITERATURE 15](#_Toc80948110)

[2.4 THE CONCEPTUAL FRAMEWORK 16](#_Toc80948111)

[2.4.2 INDEPENDENT VARIABLES 17](#_Toc80948112)

[CHAPTER 3 18](#_Toc80948113)

[METHODOLOGY 18](#_Toc80948114)

[3.1 Introduction 18](#_Toc80948115)

[3.2 Research Design 19](#_Toc80948116)

[3.2.1 Quantitative research design 21](#_Toc80948117)

[3.3 Location of the study 22](#_Toc80948118)

[3.4 population of the study 22](#_Toc80948119)

[3.5 sampling procedure and sample size 22](#_Toc80948120)

[3.6 instruments of study 22](#_Toc80948121)

[3.7 data collection procedure 23](#_Toc80948122)

[3.8 data processing and analysis 23](#_Toc80948123)

[3.9 Ethical considerations 24](#_Toc80948124)

[CHAPTER 4: 25](#_Toc80948125)

[SYSTEM IMPLEMENTATION AND DEVELOPMENT 25](#_Toc80948126)

[4.1 Introduction 25](#_Toc80948127)

[4.2 System Architecture 25](#_Toc80948128)

[4.2 User Interfaces 25](#_Toc80948129)

[4.3 User Interface modules 26](#_Toc80948130)

[4.3.1 User interface for sign up 26](#_Toc80948131)

[User interface for login 26](#_Toc80948132)

[User dashboard 27](#_Toc80948133)

[Admins dashboard 28](#_Toc80948134)

[4.6 Back-end development 28](#_Toc80948135)

[4.7 Database design models 30](#_Toc80948136)

[Users table 30](#_Toc80948137)

[Predictions table 31](#_Toc80948138)

[CHAPTER 5 32](#_Toc80948139)

[CONCLUSION AND RECOMMENDATION 32](#_Toc80948140)

[4.8 Conclusion 32](#_Toc80948141)

[4.9 Recommendations 32](#_Toc80948142)

[References 33](#_Toc80948143)

## 

## CHAPTER 1

## INTRODUCTION

**INTRODUCTION**

Life insurance is a contract between an individual or organization (insurance policy holder) and an insurer (insurance company). The policyholder pays the policy premium and the insurer promises to pay a designated beneficiary a sum of money upon the occurrence of an insured event. Depending on the contract, events such as terminal or critical illness may also trigger payment. The motive behind this study is to help millions of people to consider life insurance at an earlier time to support their families a later time. Also, anyone working in a high-risk job is likely to pay more for coverage, since the risks alone are reason enough to get life insurance.  
In the life insurance industry it’s important to employ the use machine learning models to predict the risks affecting individuals thus predicting the premium amounts each should pay. The core benefit of life insurance is that the financial interests of one’s family e.g., children’s education, remain protected from circumstances such as loss of income due to critical illness or death of the policyholder.

### **1.1 BACKGROUND STUDY**

We all must face the inevitability of death and the economic hardship that others might face when we die. Buying life insurance is one way to ease the burden of that economic risk. We can protect surviving family members by paying a relatively small amount, called the premium, to an insurance company. Then, the insurance company will pay a relatively large sum of money to the beneficiaries of the policy when the insured person dies. Certain types of life insurance may provide benefits for you and your family while you’re still living. For example, permanent life insurance offers a cash value component, which can be put to good use during your lifetime.

There are two main categories of life insurance:

**Term life insurance**: This type of policy offers protection for a limited period. It is the simplest and cheapest form of life insurance since it only provides life cover with no investment benefits. The insurance company pays out the full sum assured if the policyholder passes away within the insurance period. No benefits are payable to the policy holder if they are still alive at the time of maturity of the policy.

**Permanent life insurance:** these policies do not expire; they are intended to protect your loved ones permanently, as long as you pay your premiums. Some types of these policies accumulate cash value.

A life insurance policy works with two main components - a premium and a death benefit. Term life insurance has these two components, but permanent or whole life insurance policies also have a cash value component.

1. **Premium**– premiums are the money the policyholder pays for insurance.The insurer must pay the death benefit when the insured dies if the policyholder pays the premiums as required, and premiums are determined in part by how likely it is that the insurer will have to pay the policy’s death benefit based on the insured’s life expectancy. Factors that influence life expectancy include the insured’s age, gender, medical history, occupational hazards, and high-risk hobbies. ﻿ Part of the premium also goes toward the insurance company’s operating expenses. Premiums are higher on policies with larger death benefits, individuals who are higher risk, and permanent policies that accumulate cash value.
2. **Death Benefit**–The death benefit is the amount of money the insurance company guarantees to the beneficiaries identified in the policy when the insured dies. The insured might be a parent, and the beneficiaries might be their children. The insurance company will determine whether there is an insurable interest and if the proposed insured qualifies for the coverage based on the company’s underwriting requirements related to age, health, and any hazardous activities in which the proposed insured participates.
3. **Cash Value**– The cash value of permanent life insurance serves two purposes. It is a savings account that the policyholder can use during the life of the insured; the cash accumulates on a tax-deferred basis. Some policies may have restrictions on withdrawals depending on how the money is to be used. For example, the policyholder might take out a loan against the policy’s cash value and have to pay interest on the loan principal. The policyholder can also use the cash value to pay premiums or purchase additional insurance. The cash value is a living benefit that remains with the insurance company when the insured dies. Any outstanding loans against the cash value will reduce the policy’s death benefit.

Life insurance provides financial support to surviving dependents or other beneficiaries after the death of an insured. Below are some examples of people who may need life insurance:

* **Parents with minor children**– If a parent dies, the loss of their income or caregiving skills could create a financial hardship. Life insurance can make sure the kids will have the financial resources they need until they can support themselves.
* **Adults who own property together**–Married or not, if the death of one adult would mean that the other could no longer afford loan payments, upkeep, and taxes on the property, life insurance may be a good idea. An example would be an engaged couple who took out a joint mortgage to buy their first house.
* **Elderly parents who want to leave money to adult children who provide their care**–Many adult children sacrifice by taking time off work to care for an elderly parent who needs help. This help may also include direct financial support. Life insurance can help reimburse the adult child’s costs when the parent passes away.
* **Families who can**’**t afford burial and funeral expenses**–A small life insurance policy can provide funds to honor a loved one’s passing.
* **Married pensioners**–Instead of choosing between a pension payout that offers a spousal benefit and one that doesn’t, pensioners can choose to accept their full pension and use some of the money to buy life insurance to benefit their spouse.
* **Wealthy families who expect to owe estate taxes**– Life insurance can provide funds to cover the taxes and keep the full value of the estate intact.
* **Young adults whose parents incurred private student loan debt or cosigned a loan for them**–Young adults without dependents rarely need life insurance, but if a parent will be on the hook for a child’s debt after their death, the child may want to carry enough life insurance to pay off that debt.

Before you apply for life insurance, you should analyze your financial situation and determine how much money would be required to maintain your beneficiaries’ standard of living or meet the need for which you’re purchasing a policy. The younger and healthier you are, the easier it will be to qualify for life insurance and the older and less healthy you are, the harder it will be. Certain lifestyle choices, such as using tobacco or engaging in risky hobbies such as skydiving also make it harder to qualify or lead to higher rates. Most people use life insurance to provide money to beneficiaries who would suffer a financial hardship upon the insured’s death. However, for wealthy individuals, the tax advantages of life insurance including tax-deferred growth of cash value, tax-free dividends, and tax-free death benefits, can provide additional strategic opportunities.

### **1.2 PROBLEM STATEMENT**

Lack of trust in the insurance industry is the reason why many individuals don’t bother with insurance. Many insurance firms fail to pay claims, and they don’t own up to offering some benefits. Most people just see insurance as one of the unnecessary expenses and many insurance firms do shut down because of financial challenges and victims of the loss don’t even think twice about purchasing insurance policies.

Economic instability of a country affects all insurance firms. At such situations, the rates can be affected such that the insurance companies might be forced to increase their rates, just like interest rates on credit facilities provided by financial institutions. No client will appreciate this even if it is stated clearly in the contract that the insurance rates might change from time to time. Such situations create a bad image for a company since customers can spread the information about a service or product, they were not happy with very fast.

Mismanagement also poses a challenge in the insurance industry. As the owner of the insurance business, one is responsible for all issues that his or her clients may have regarding the management of the insurance business. All insurance firms that are mismanaged can`t hide their faults for a longer time without the clients noticing. As time moves, there will be a constant increase in the number of clients` complaints, and if his or her insurance firm is not transparent, then he or she will lose more customers. Also, incompetent management may cost the company a lot particularly if they have poor communication with their clients.

### **1.3 RESEARCH OBJECTIVES**

A research study designed to mitigate the risk associated with life insurance policies had the following general and specific objectives:

General objective: To predict the amount of premiums individuals should be able to pay either monthly or annually depending on the risk associated with the individual.

Specific objectives:

1. To develop a Risk analysis system for life insurance based on machine learning
2. To test and evaluate the effectiveness and efficiency of the proposed system
3. To know the various investment alternatives that is mostly preferred by the customers.
4. To find out the criteria that people think about before investing in a life insurance policy
5. To find out the level of awareness of reliance life insurance among people.
6. To know if life policies are suitable for specific people.
7. To find out where policies should be kept.
8. To understand what a rider means.
9. To understand what premium waiver means.
10. To understand the difference between term life insurance and permanent life insurance.

### **1.4 RESEARCH QUESTIONS**

1. How will the premium amount be determined?
2. Which investment alternatives are mostly preferred by you customers?
3. Which criteria is used in life insurance policy?
4. How do you conduct the level of awareness of reliance life insurance?
5. Are whole life policies more suitable for some people than for others?
6. Where should I keep my policy?
7. What is a rider?
8. What does premium waiver mean on my amendment policy?
9. What is term life and permanent life insurance?

### **1.5 STUDY JUSTIFICATION**

Life insurance provides more benefits to your family when needed. The machine learning model will be able to achieve this goal through analysis different kinds of risks. Some of the benefits include: income replacement for years of lost salary, paying off your mortgage, paying off other debts such as loans, providing funds for your kid’s college education and tax advantaged estate planning benefits. Life insurance cover benefits the named beneficiaries of the insured upon death.

### **1.6 APPLICATION OF PROJECT**

The main aim of a life insurance cover is to secure the needs of dependents after one's untimely death. In addition to the emotional suffering, the financial insecurity arising out of losing the primary earner can be immense.

## CHAPTER 2

## LITERATURE REVIEW

### **2.1 INTRODUCTION**

Life insurance policy is a contract between an insurer and a policyholder. The contract is legally binding and for it to be enforced, the life insurance application must accurately disclose the insured’s current and past conditions and the high-risk activities associated with the individual.

The gap between individuals with life insurance and those without is very huge. According to LIMRA International’s 2015 Insurance Barometer Study, 100 million Americans (43% of the adult population) have no life insurance coverage. The rate of household coverage has continued to trend down since 1960 and is now near a 50-year low. At the same time, LIMRA reports that half of all U.S. households (about 59 million) believe they need more life insurance. This is one of the biggest gaps in the history of U.S. financial services in terms of both people and dollars. It represents a huge opportunity for the life insurance industry and you. (LIMRA, Ashley Durham, 2015).

The ability to have a system that can make difficult choices when determining the premium amounts will be a game changing idea to many insurance companies. The data collected would give detailed information on different kinds of risk activities associated with the insured.

### **2.2 CRITIQUE OF THE EXISTING LITERATURE RELEVANT TO THE STUDY**

#### **2.2.1 IMPROVED RISK SELECTION**

Machine learning will be able to help insurers evaluate and price insurance risks through new kinds of modelling and data with even a potential way to replace the statistical models. IGNL participants suggested that artificial intelligence and machine learning systems could enable superior underwriting hence delivering a competitive advantage to those who are better able to deploy them. In the insurance industry, making the right risk decision makes one a winner. Winners and losers will be defined by the much information they have and can apply to making risk judgments.

#### **2.2.2 IMPROVING OPERATIONAL EFFECTIVENESS**

Machine learning applications are helping insurers increase efficiency and drive the costs down by automating underwriting, customer service and claims processing. This will help the insurers operations and insurance processes in the following ways:

#### **2.2.2.1 CUSTOMER EXPERIENCE**

Automated assistants and chat bots cut costs by automating responses to basic questions or handling simple complaints from those individuals who would like to purchase a life insurance policy.

#### **2.2.2.2 FRAUD DETECTION AND PREVENTION**

Machine learning and artificial intelligence technology is increasingly deployed to detect and prevent fraud. Data collected from different individuals is highly sensitive and needs to be kept as a secret. By doing this, the insurance industry is estimated to save up to $20 billion a year.

#### **2.2.2.3 CLAIM MANAGEMENT**

The number of claims that require human analysis and interaction are seemingly to be on the rise each single day. Machine learning can lower the costs and improve the customer experience through speeding the claims resolution.

#### **2.2.2.4 TALENT EVALUATION**

Machine learning systems will improve the assessment and coaching of front-line staff, while the language processing system will enable better compliance and increased sales effectiveness.

### **2.2.3 GETTING DATA ISSUES RIGHT**

Insurers haven’t invested much in core platforms for a long period of time thus they don’t have much confidence in the quality of their data. Data maintenance, cleaning and engineering are crucial here. Data can be stolen or adversaries can poison data sets. The process called adversarial machine learning, involves injecting statistical noise or false information into a system training data to affect outcomes.

### **2.2.4 NEW PRODUCTS AND SERVICES**

The major concern is shifting from the current model of indemnifying against damages to a model focused on risk mitigation and prevention. Machine learning models, using data acquired from the available datasets can anticipate and potentially prevent risks. The service model is aimed to where you need to help people help themselves, which will reduce costs and reduce premium amounts.

### **2.3 REVIEW OF THEORETICAL LITERATURE**

While new life settlement pricing models continue to emerge, some being patented and copyrighted, two fundamental methodologies have emerged in most pricing models: Deterministic Pricing, and Probabilistic Pricing. While these two methodologies differ significantly, they share several common contributing factors, including but not limited to: presenting an expected internal rate of return (IRR) to the policy investor, utilizing the age and life expectancy (LE) of the insured to determine applied mortality rates, reverse engineering the policy’s cost of insurance (COI), discounting the death benefit, and applying account and cash values of the policy. (Dan Zollars, Scott Grossfeld & Deborah Day, 2003)

#### **2.3.1 DETERMINISTIC PRICING MODEL**

According to (Dan Zollars, Scott Grossfeld & Deborah Day, 2003) methodology used to value life insurance policies is the Deterministic Pricing Model. The Deterministic Pricing Model  originated from the early viatical settlement industry where life expectancies of insured were shorter and theoretically more predictable than longer senior life settlements which are more common in today’s secondary life insurance markets.

The basic pricing premise of the Deterministic Pricing Model assumes that, in addition to the purchase price, the purchaser pays the full policy costs until the insured’s estimated life expectancy or probable time of death. The value of the life insurance policy is then determined by calculating the discounted value of the death benefit from the life expectancy (the cash inflow), less the discounted value of the policy costs (premiums) required to maintain the policy (the minimum cash outflows) from date of purchase to life expectancy. The discount rate applied to this cash flow stream is the expected return the investor seeks over the specified period of time.  Some investors add one to two years to the estimated life expectancy as a risk buffer to ensure they obtain their expected rate of return. The methodology of the Deterministic Pricing Model is simple and easy to understand and is based primarily upon the accuracy of the life expectancy prediction. In portfolio theory, the Deterministic Pricing Model assumes a large portfolio of policies is purchased with the insured having somewhat homogenous characteristics and the mortality of those insured approximating the mean expected life expectancy of the entire portfolio.

However, the primary weakness of the deterministic pricing methodology is its failures to directly take into account the probability of deaths occurring before or after the estimated life expectancy. In addition, deterministic methodology does not consider the mortality curves of the insured. Accordingly, researchers have looked for additional life insurance pricing valuation tools in order to better evaluate life insurance policies for purchase.

### **2.4 THE CONCEPTUAL FRAMEWORK**

Life insurance conceptual framework defines the relevant variables for the study and maps out how they might relate to each other. Some the variables addressed here are independent and dependent variables. Regression analysis tool will help us in predicting one variable (dependent variable) based on its relationship with one or more other variables (independent variables).

#### **2.4.1 DEPENDENT VARIABLES**

The dependent variable is the variable being tested and measured in an experiment, and is 'dependent' on the independent variable. Tax and cost of living are our examples:

#### **2.4.1.1 Tax**

There are situations where the beneficiary is taxed on some or all of a policy. If the policyholder elects not to have the benefit paid out immediately upon his death but instead held by the life insurance company for a given period of time, the beneficiary may have to pay taxes on the interest generated during that time. Also, when a death benefit is paid to an estate, the person or persons inheriting the estate may have to pay estate taxes on it.

#### **2.4.1.2 Cost of living**

When estimating the cost of living of an individual, then several factors such as age, salary, marital status, occupation and lifestyle, etc. become independent variables while the cost of living is highly dependent on such factors.

### **2.4.2 INDEPENDENT VARIABLES**

The independent variable is the variable the experimenter changes or controls and is assumed to have a direct effect on the dependent variable.

#### **2.4.2.1 Age**

Age becomes the first factor which comes into mind before a life insurance company decides the premium. An individual at a younger age is at a low risk of getting life-threatening diseases thus the premium amount will be low if you are buying a life insurance at a younger age. Older individuals are associated with a risk of getting life threatening diseases or death increases and thus resulting to higher premiums.

#### **2.4.2.2 Gender**

Gender becomes the second biggest determinant of pricing. Statistical models are used by insurance carriers to approximate how long someone with a specific profile will be around. On averagely, women live nearly five years longer than men. Since they will be paying premiums for long periods of time than males, they enjoy slightly lower rates than men.

#### **2.4.2.3 Profession**

Individuals associated with professions such as shipping, transport, mining, piloting aircraft, oil and gas are at a higher accidental risk which can result to death. As a result, the premium amount charged by the insurer will be much higher compared to someone who is in a profession where the risk to life is minimal such as a secretarial job or desk job.

#### **2.4.2.4 Health conditions**

During the underwriting process, many carriers include a medical exam in which the company records weight, height, blood pressure and other key factors. They may also require an electrocardiogram to check your heart in some cases. It is very important to get any serious conditions like cholesterol and diabetes managed before searching for coverage to ensure a competitive rate.

#### **2.4.2.5 Lifestyle**

Any time you engage in high-risk activities like racing cars and hill climbing, there’s an increased likelihood that you will meet an early end which is a big concern for carriers. Some companies also charge more if you have a relatively dangerous profession, such as mining, fishing or transport with heavy machinery.

#### **2.4.2.6 Smoking and drinking habits**

Smoking and drinking can result into a major setback to your health thus increasing the risk of life-threatening diseases or in even certain conditions like death. Life insurance companies always check such habits with an individual before deciding the premium amounts. Those individuals who are associated with smoking and consume alcohol are likely to pay higher premium amounts than those who don’t.

# 

# **CHAPTER 3**

## METHODOLOGY

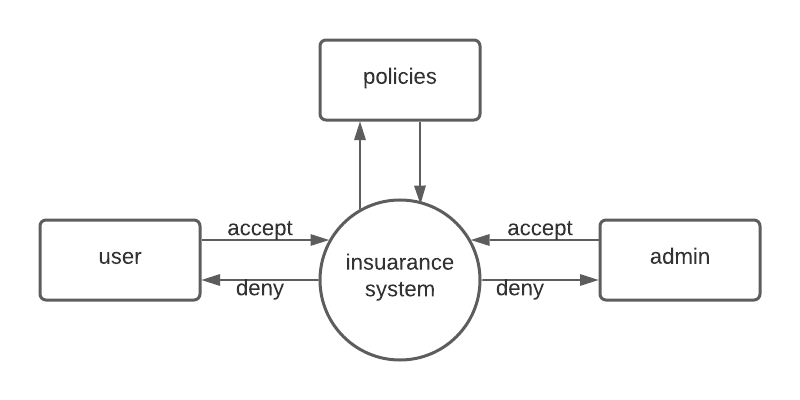
### **3.1 Introduction**

This chapter entails the methodological concerns used in conducting this research and provides a justification for each of the steps taken. It involves the research design, location and population of the study, sampling procedure and sample size, validity and reliability of the research, data collection, processing and analysis, summary of statistical methods and the ethical considerations of the research.

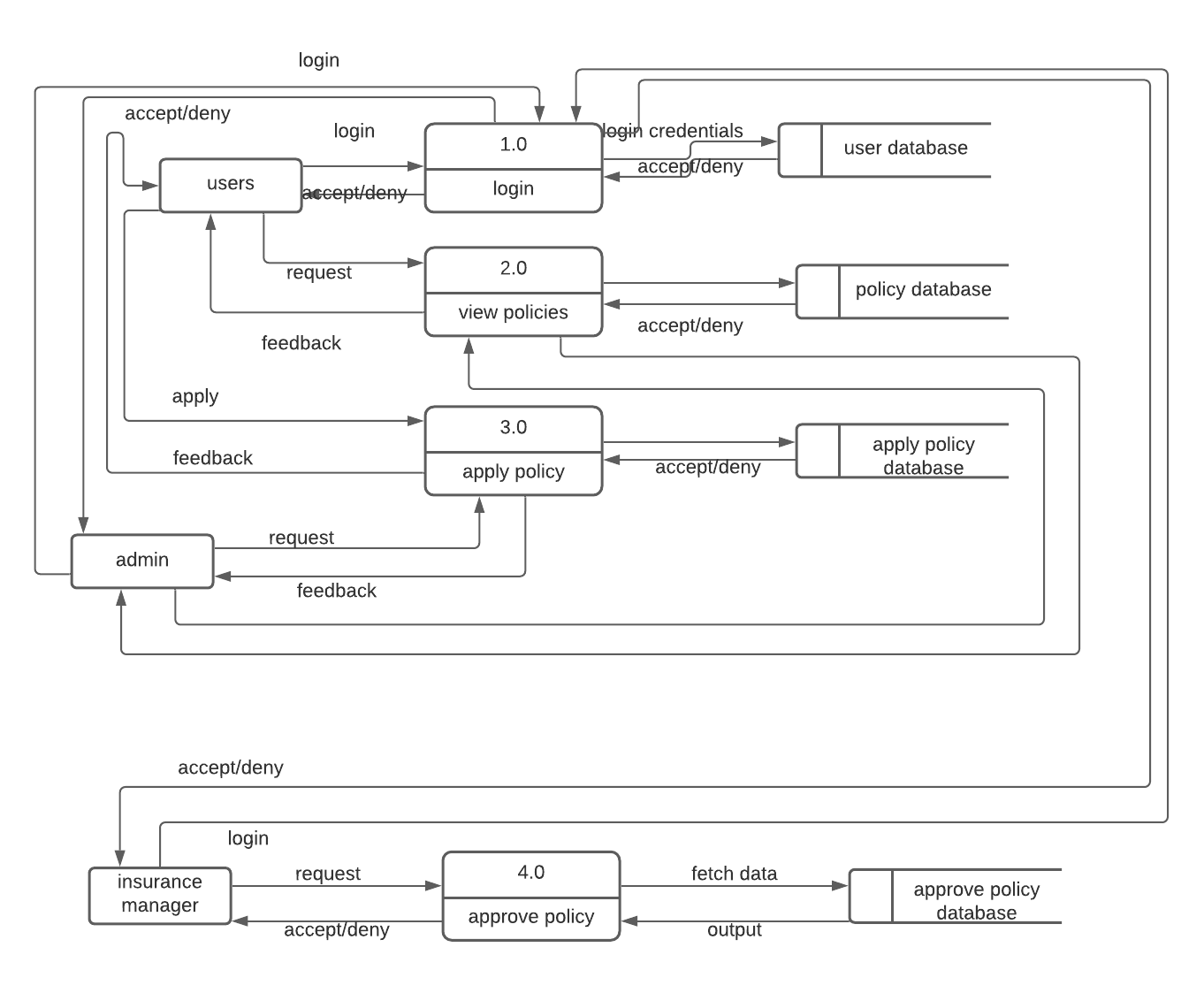
### **3.2 Research Design**

Research design is the framework of research methods and techniques chosen by a researcher. The design allows researchers to look into research methods that are suitable for the study. In this study, we will look into quantitative research design.

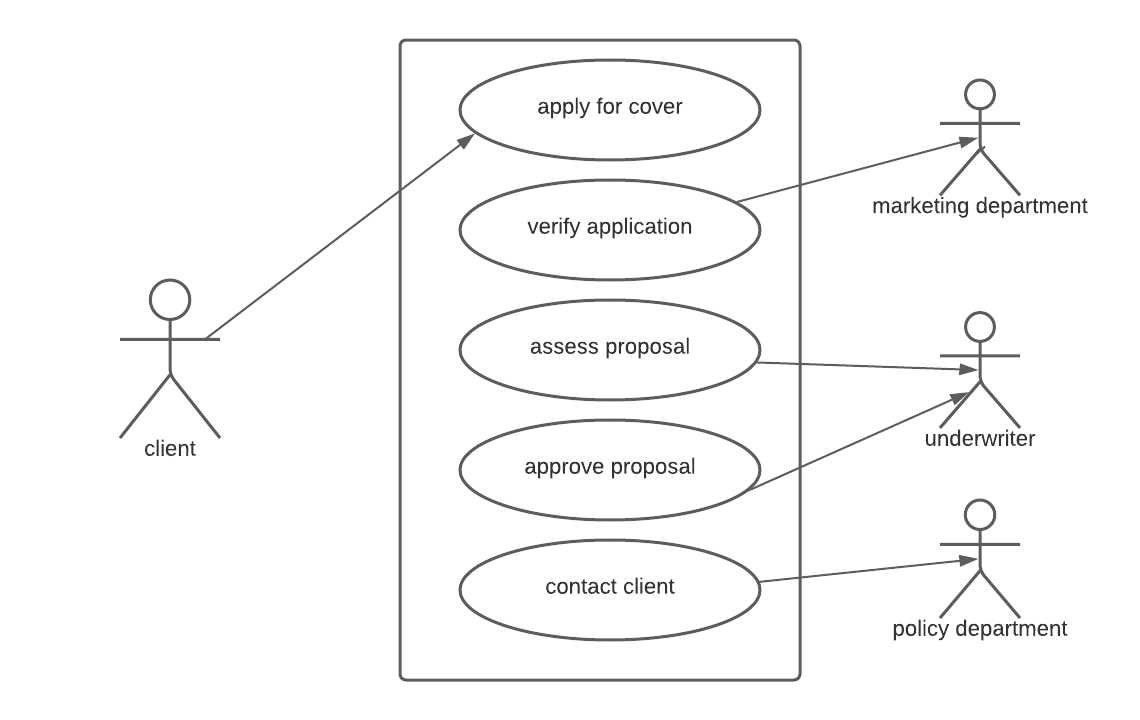
#### **Context Diagram**



#### **Data flow Diagram (DFD)**

****

**Use case diagram**

****



### **3.2.1 Quantitative research design**

This design analyzes a variable relationship in terms of numbers and statistical strategies to review its findings. Insights drawn from hard numerical data and analysis prove to be highly effective when making decisions related to the future of the business. Regression analysis will be used since the study involves modelling and analyzing several variables where the relationship includes a dependent variable and one or more independent variables. Under quantitative research, we shall look into experimental research design.

#### **3.2.1.2 Experimental research design**

This design determines the cause and effects of relationships among various variables in such a way that the independent variable is changed in order to observe its effect on the dependent variable.

### **3.3 Location of the study**

Rural areas in Kenya have become the major hotspot with large numbers of people without life insurance covers. The life insurance will ensure that families living in the rural areas have a safe and secure future so that they can lead a happy life. The insurance will help them to cover risks related to various aspects in their life.

### **3.4 Population of the study**

Polit and Hungler (1999) refer to the population as an aggregate or totality of all the objects, subjects or members that conform to a set of specifications. In this study, the population was Kenyan residents of different age groups, social-economic status and residential areas who need life insurance covers.

### **3.5 sampling procedure and sample size**

The sampling procedure used in this study is stratified random sampling since the researcher tries to understand the existing relationship between the independent and dependent variables. The estimated population in this study is 5000 where the sample size is 10% of the population which becomes 500.

### **3.6 instruments of study**

#### **3.6.1 Pilot study**

Pilot study helps to test the research protocols, data collection instruments and other research techniques. It is conducted to help identify potential problems areas. The pilot study will be carried out through the following techniques;

* Interviews
* Observations
* Questionnaires
* Surveys
* Focus groups

#### **3.6.2 Validity of the instrument**

The system will be able to calculate the correlation between the input data and the results of the criterion enabling its users to enjoy benefits such as; paying off debts and other expenses, protecting the family and loved ones, adding more financial security and leaving an inheritance.

#### **3.6.3 Reliability of the instrument**

Through multiple regression analysis, the model will help us understand how much the dependent variable change when we change the independent variables. Also, the model will be more accurate since it gives us more of the information available to us who estimate the dependent variable. This will build trust among its users since it will predict the outcomes depending on the risk factors associated with an individual.

### **3.7 data collection procedure**

Data collection is a process of collecting information from relevant sources to find solutions to the research problem, test the hypothesis and evaluate the outcomes. In this study we used secondary data collection methods since we will be working with the already available datasets to predict the outcomes. Below is the procedure used in collecting the data ;

* Data is downloaded as .csv file(concurrent version systems).
* Data is analyzed, cleaned and manipulated to match the desired algorithm.
* Exploratory data analysis (EDA) is performed to see the effect of each independent variable on the dependent variable.
* Based on the EDA, the following machine learning algorithm is selected:
* Multiple Linear Regression

### **3.8 data processing and analysis**

#### **3.8.1 Data processing**

Data processing is done when data has been collected and translated into usable information. Below are the steps used in processing the collected data ;

Gender, profession, health condition, lifestyle, smoking and drinking habits are converted to categorical variables first and then they are converted to numerical variables to be compatible with the model building.

Missing values are removed, and the data is cleaned for analysis to build the model.

Some of the columns are scaled for the model building.

5-cross validation is performed to train and test the data and compute the out-of-sample metrics.

#### **3.8.2 Data analysis**

**Data analysis** is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis (Guru99, 2020). Below are the steps of data analysis:

Data collection

Data cleaning

Data analysis

Data interpretation

Data visualization

### **3.9 Ethical considerations**

All researchers have good intentions, but if they are not careful, their studies can place individuals in situations that involve risk (Gall, 1996). Some of the considerations pertaining this research are:

1. Informed consent – the researcher is required to obtain permissions from the specific authority where he/she is to conduct the research.
2. Invasion of privacy and confidentiality - the type of research should be conducted under the assumption that the researcher will keep the findings anonymous and confidential.
3. Deception – happens when the researcher provides misleading information about the project.
4. Data protection – the researcher must put in place measurers to ensure that data is protected from unauthorized or unlawful processing and against accidental loss or destruction.
5. Protection from harm – the researcher should take the responsibility to protect the subjects from whatever harm which might befall them e.g., emotional stress, embarrassment or physical discomfort.

# **CHAPTER 4:**

# **SYSTEM IMPLEMENTATION AND DEVELOPMENT**

## 4.1 Introduction

This chapter addresses the last phase of SDLC: systems implementation. In this phase, the researcher built the code base of the system and tested the various functionalities of the system to ensure that it performed at designed. The researcher used python with naïve bayes model to develop the system. The model contains various types of files. The logic of the system was built using python files, while the user interface was build using HTML and CSS files. The researcher used SQLite3 as a development database, PostgreSQL as a production database, and chrome as the web server. The system was deployed on Heroku; a cloud Platform-as-a-Service used as web application deployment model.

## 4.2 System Architecture

The architecture of the proposed life insurance premium predictor has three main components: a database, a webserver and a web client. The database stores data on properties, users, and preference models of the users each in its own table. The web server receives and processes HTTP requests to perform CRUD (create, read, update and delete) actions on the data stored in the database. The web client provides the user interface that enables users to interact with the system. The system architecture is illustrated below.

Web client

Web server

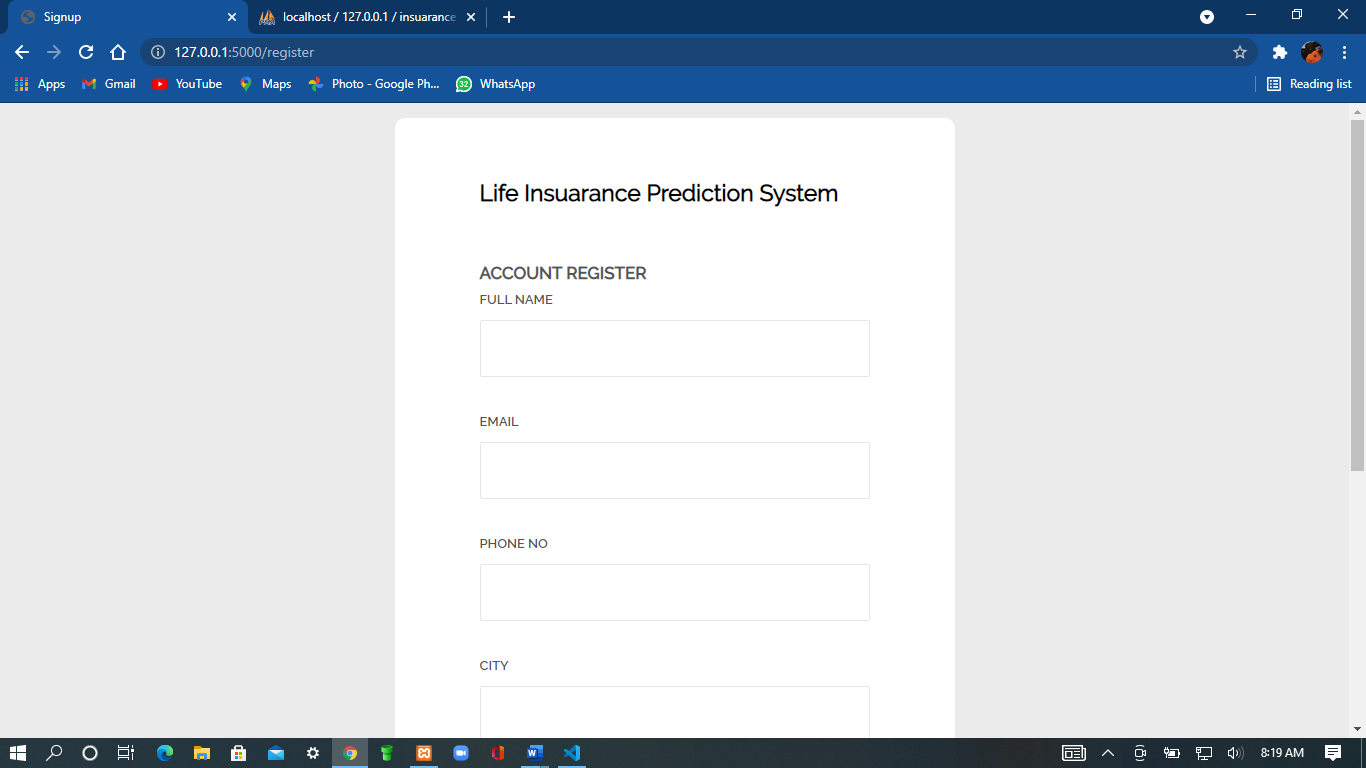
Database

## 4.2 User Interfaces

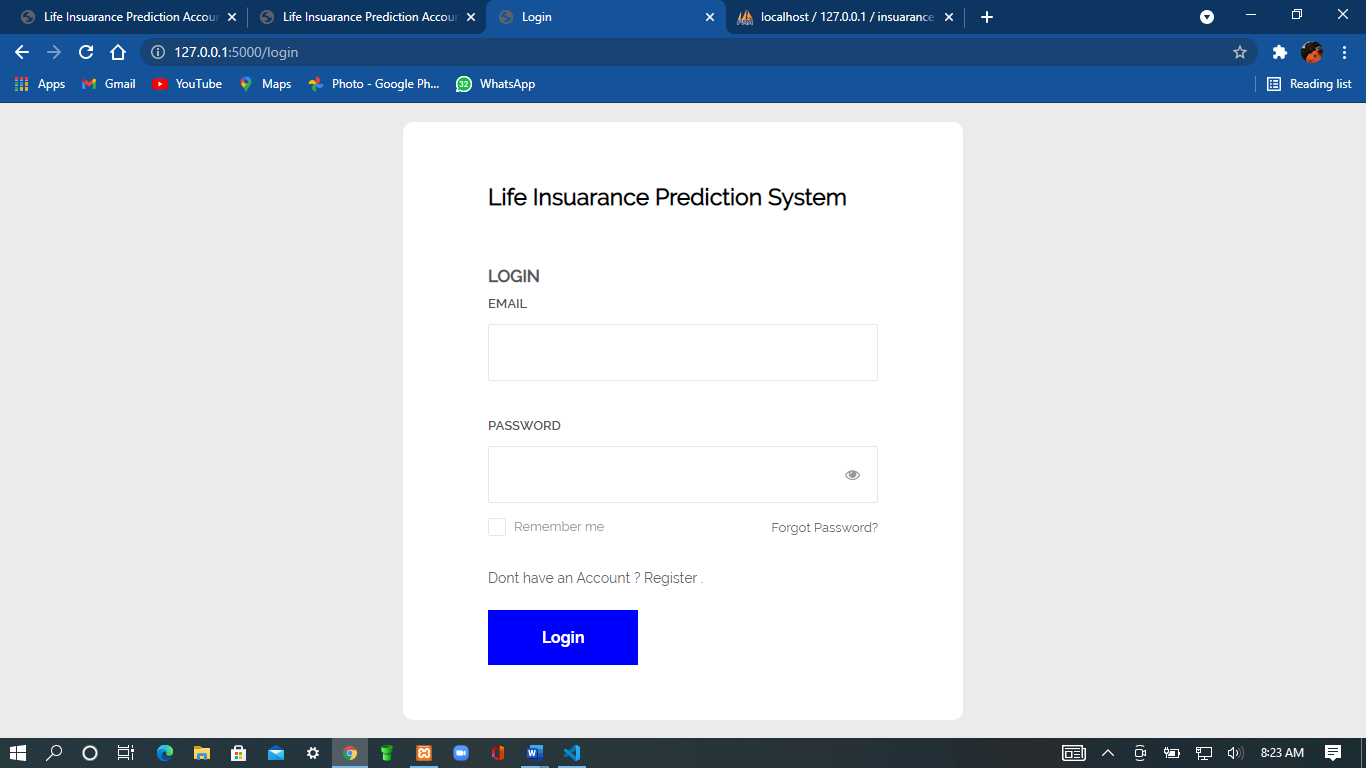
User interfaces were built using HMTL for content and CSS for styling. Use of interactive user interface to enhance user experience and allow even novice users to be familiarized by the system. Elements like bold text style to attract attention and the use of color similarities to provide a nice UI. This is the main functionality of the life insurance system: it enables users to key in their preferences then the system predicts the premium amount that should be paid on a monthly basis.

## 4.3 User Interface modules

# **4.3.1 User interface for sign up**

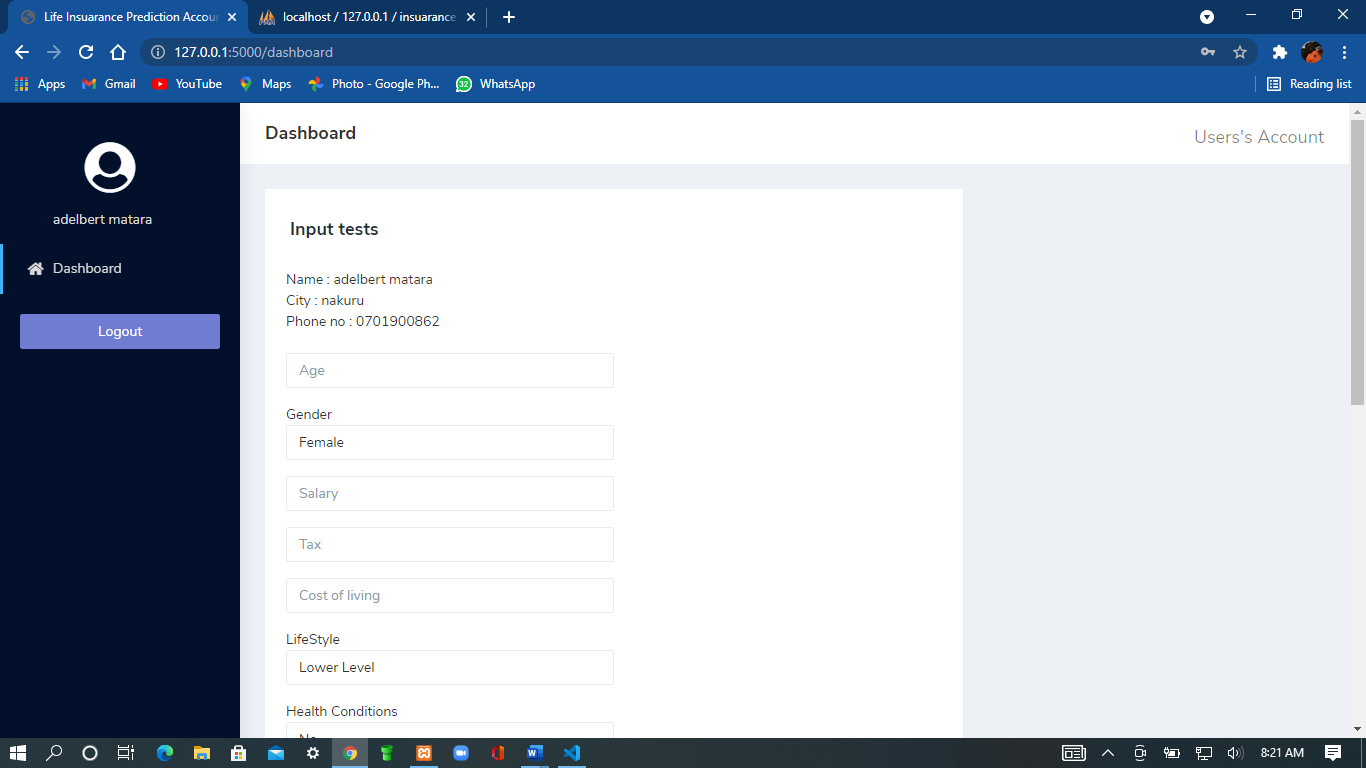


# **User interface for login**

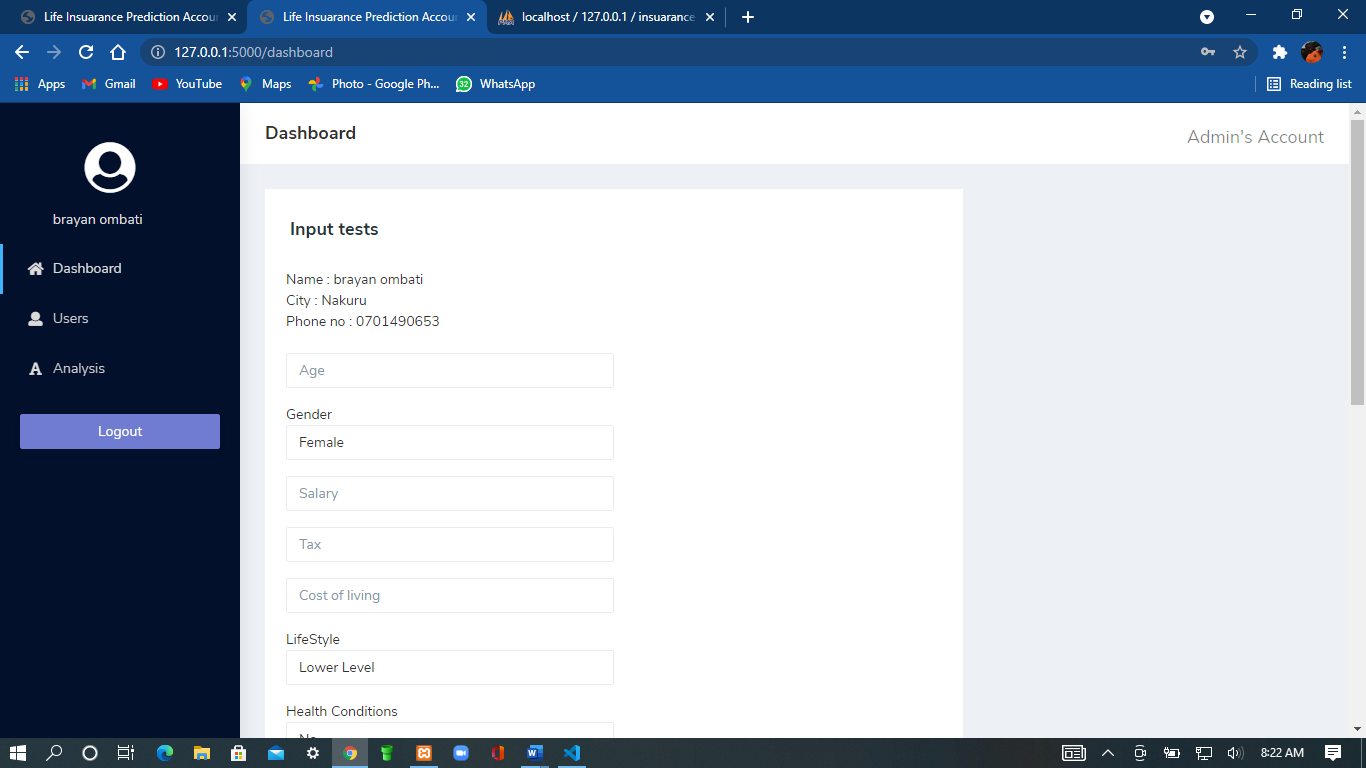


# **User dashboard**

This is the main functionality of the life insurance system: it enables users to know the premium amount the insurance company is likely to ask for from the systems prediction. This interface displays a dashboard with various attributes from which the user is supposed to select according to his or her preferences.

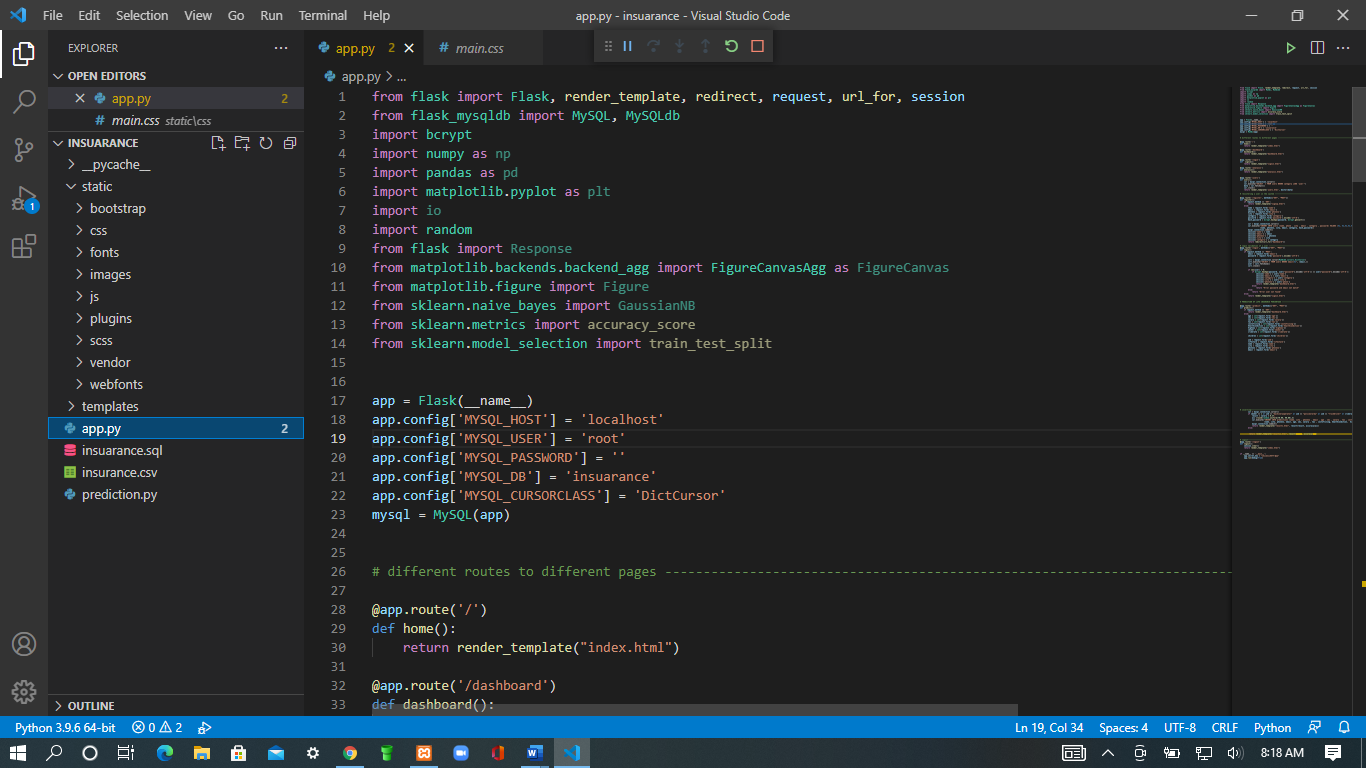


# **Admins dashboard**

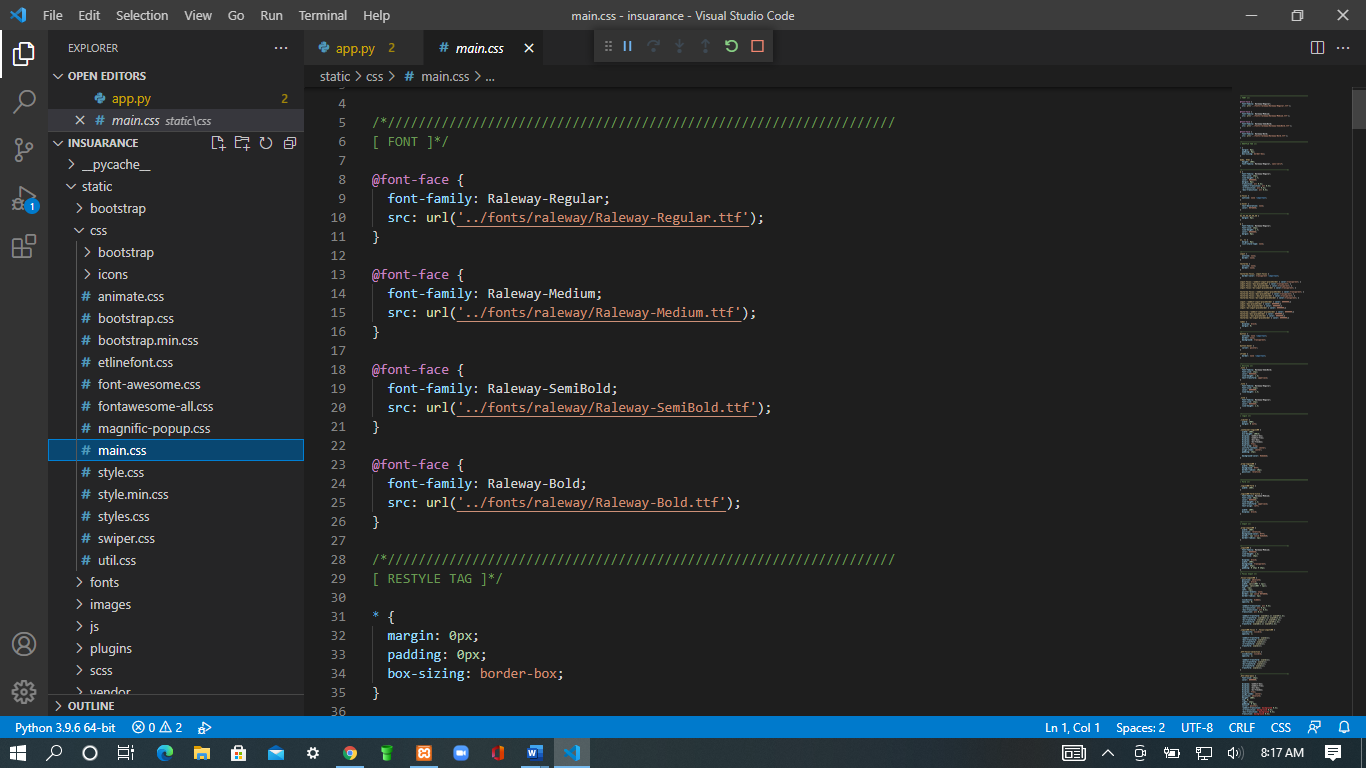


## 4.6 Back-end development

The code that implements the core algorithms used in this prototype is written in the python programming language. Html and php, being MVC (Model-View-Controller) framework, these algorithms are housed in models (that map to classes defined in Chapter Four). In particular, these algorithms that implement the logic of generating results for the search defined in the ‘naïve bayes ’ model. The ‘Views’ are responsible for displaying user interfaces, while controllers are responsible for managing the flow of data between the database, the models and the views. Css was used for styling.

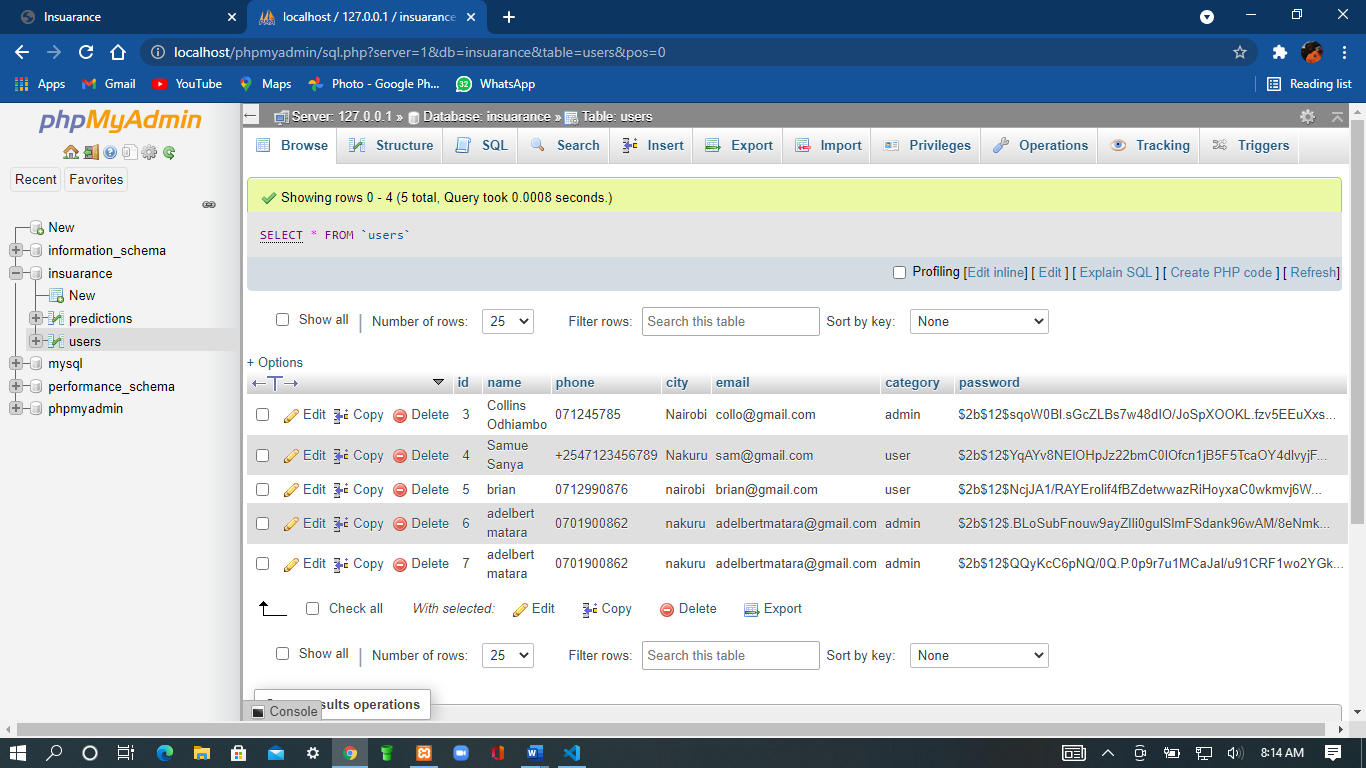


Here, you can see the application of css for styling of the User interface.



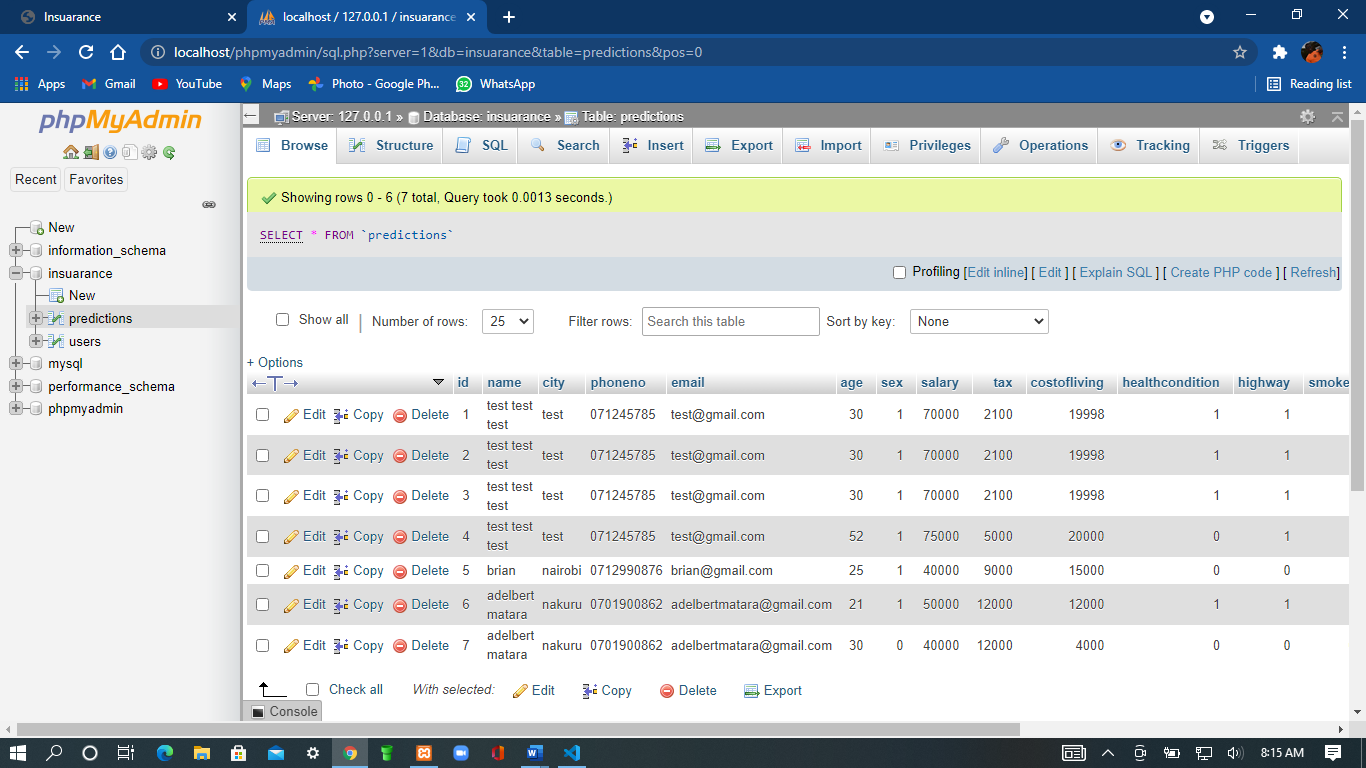
## 4.7 Database design models

## Users table



## 

## Predictions table



## 

## CHAPTER 5

## CONCLUSION AND RECOMMENDATION

## 4.8 Conclusion

This study had one main objective which to predict the amount of premiums individuals should be able to pay either monthly or annually depending on the risk associated with the individual. The next paragraphs indicate how these objectives were achieved in the course of this research endeavor.

The study achieved the first objective (To predict the amount of premiums individuals should be able to pay either monthly or annually depending on the risk associated with the individual) in Chapter 2. In this section, the study examined the currently available life insurance system technologies focusing on the definition and evolution of the insurance systems. It further discussed various functions of the insurance system and reviewed briefly various recommendation approaches. These include collaborative filtering, content-based, knowledge-based, community-based, demographic-based, hybrid approaches.

Section 3 discussed various types of critiquing-based insurance systems including natural language-based systems, system-suggested critiquing systems and user-initiated critiquing systems. These systems use various algorithms to arrive at their recommendations.

Section 4 describes in details the user-system interactions used in user-initiated critiquing systems focusing on the user preference elicitation, system recommendation, user feedback and selection of desired item. These are steps used by algorithms found in this type of insurance system. Section 5 presents one particular algorithm used in this study to generating suggestions for recommendation. The algorithm is presented in the form of pseudo-code.

## 4.9 Recommendations

With the increased availability of Internet connection in Kenya and the reducing costs of accessing the Internet, coupled with increased interest in online commerce, many consumers are turning to online resources to locate items that are of interest to them. It is, therefore, important for businesses to invest in technology that supports their customers to easily access the information they require to find what they are looking for and ultimately make purchases. This analysis applies to the life insurance policy. Most insurance companies say a reasonable amount for life insurance is **six to 10 times the amount of annual salary**. Another way to calculate the amount of life insurance needed is to multiply your annual salary by the number of years left until retirement.

# **References**

1. LIMRA, and Ashley Durham*.*2015 Insurance Barometer Study*. 2015.*
2. Life Settlement Industry Resources the Art of the Deal: Pricing Life Settlements, Contingencies, Zollars, Grossfeld & Day (2003)
3. Murtha, J.A. 1997. Monte Carlo Simulation: Its Status and Future.
4. Carmines, E., & Zeller, R. (1979).  [Reliability and Validity Assessment](https://books.google.com/books?hl=en&lr=&id=o5x1AwAAQBAJ&oi=fnd&pg=PA1&dq=Reliability+and+Validity+Assessment&ots=2J9DktBCCZ&sig=IKQ-54_tnTYHRxRmAmcNvncLr6Y#v=onepage&q=Reliability%20and%20Validity%20Assessment&f=false). Beverly Hills, CA: Sage

Publications.

1. Yuan, X., Lee, J.-H., Kim, S.-J., & Kim, Y.-H. (2013) .Toward a user-oriented recommendation system for real estate websites. Information Systems, 38(2), 231-243.
2. Viappiani, P., Faltings, B., & Pu, P. (2006, 12). Preference-based Search using Example-Critiquing with Suggestions. Journal of Artificial Intelligence Research, 465-503.
3. Viappiani, P., & Faltings, B. (2006). Design and Implementation of Preference-based Search. Ecole Polytechnique Fédérale de Lausanne, Artificial Intelligence Lab, Lausanne.
4. Guru99, 2020. What is data analysis https://www.guru99.com/what-is-data-analysis.html