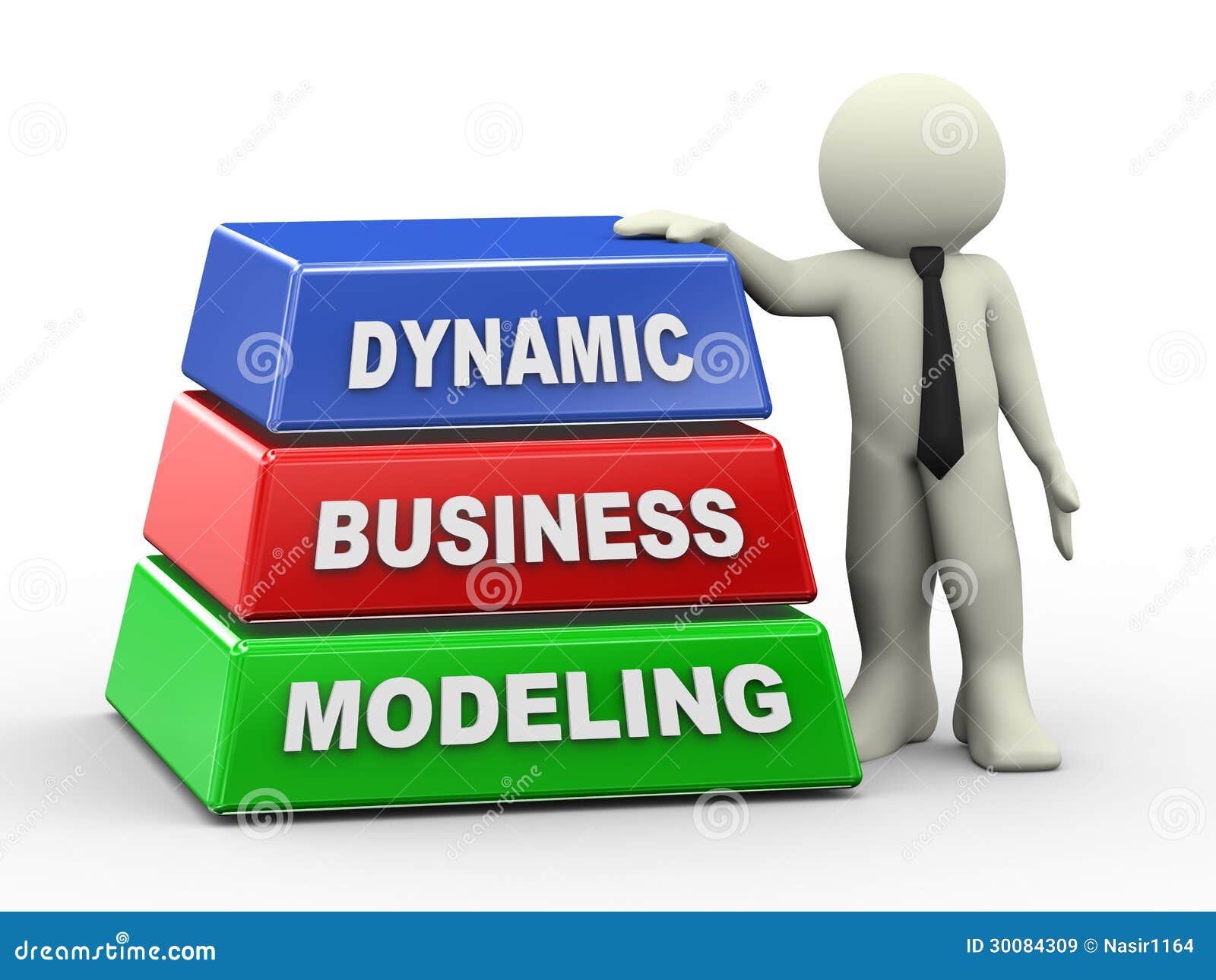
PROJECT 3

**Prototype Development, Business & Financial Modelling for Health Insurance Companies**



**Date of Submission**: 09.07.2023

**Submitted By:**  Sandeep Ghodeswar

# Table of Contents

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Content** | **Page. No** |
| 1 | Problem Statement | 3 |
|  |  |  |
| 2 | Prototype Selection | 4 |
|  |  |  |
| 3 | Exploratory Data Analysis | 5 |
|  |  |  |
| 4 | Prototype Development | 11 |
|  |  |  |
| 5 | Business Modelling | 13 |
|  |  |  |
| 6 | Financial Modelling | 15 |
|  |  |  |
| 7 | Implementation | 17 |

**1.0 Problem Statement**

Step 1: Prototype Selection

We are required to choose one prototype idea from the prototype reports of all teammates in my team on the following criteria:

1. Feasibility: The product/service can be developed in short term future. (2-3 years)
2. Viability: The product/service should be relevant or able to survive in long term future. (20-30 years)
3. Monetization: The product/service should be monetizable directly. (Indirectly monetizable product/service should be dropped for this project)

Step 2: Prototype Development – this step is optional

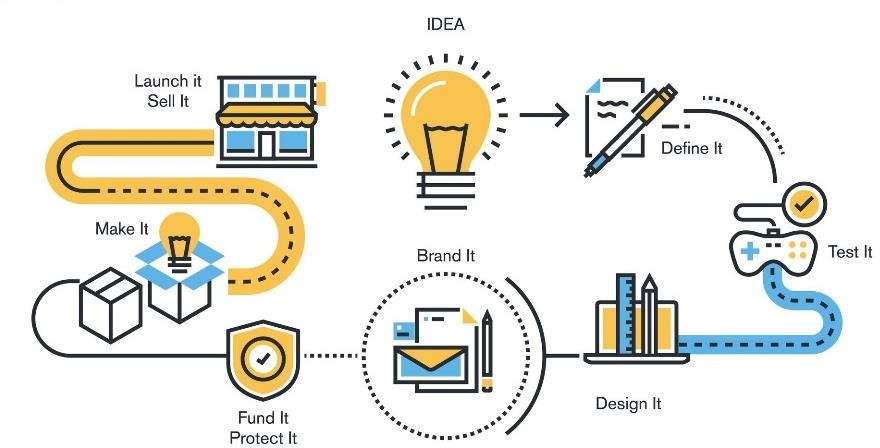
* 1. Small scale code implementation/model building of the prototype
  2. A working replica of the product/service needs to be developed using web or app development.

Step 3: Business Modelling

Business models are important for both new and established businesses. They help new, developing companies attract investment, recruit talent, and motivate management and staff. We need to come up with a business model for making profit which identifies the product/service the business plans to sell, its target market, and any anticipated expenses.

Step 4: Financial Modelling

We need to identify in which market our product/service will be launched and design a financial equation corresponding to that market trend by collecting some data/statistics regarding the market online, performing forecasting/predictions on that market using regression models or time series forecasting.



# 2.0 Prototype Selection

We chose the prototype idea of AI in Health Insurance Companies which would utilize artificial intelligence and natural language processing capabilities to determine whether the health insurance claim is guaranteed or not based on fraud detection.

Let us examine how our idea meets various criteria:

1. Feasibility:

The development of a health insurance claim chatbot is feasible within a span of 2-3 years. The technology required for natural language processing, machine learning, and development of a chatbot is already well-established and continuously improving.

1. Viability:

As technology advances, the chatbot can evolve and adapt to changing customer needs and preferences. With ongoing improvements, the chatbot can become more intelligent, capable of handling complex inquiries, and provide personalized support. It can be updated to accommodate new claim processing regulations and requirements. Hence this concept is viable in the long-term future of 20-30 years as well.

1. Monetization:

Various insurance companies can offer this product as a premium feature of their value- added services to policyholders, or charge a subscription fee for access to the chatbot. They can also explore partnerships with third-party service providers to offer the chatbot’s service in collaboration with their offerings which can create additional revenue streams. Hence a health insurance claim chatbot can be monetized directly

# 2.1 About the Prototype

This AI product is dedicated to processing health insurance claims. It incorporates a chatbot powered by natural language processing techniques to promptly gather data from users.

The 24x7 chatbot provides round-the-clock-support, allowing policyholders to get assistance whenever they need it. It efficiently collects, and analyzes the data such as medical reports provided by individuals to determine the eligibility of the health insurance policy, which reduces the need for human agents to handle routine claim inquiries. Its primary function is to swiftly detect potential fraud within the submitted data and provide real-time feedback.

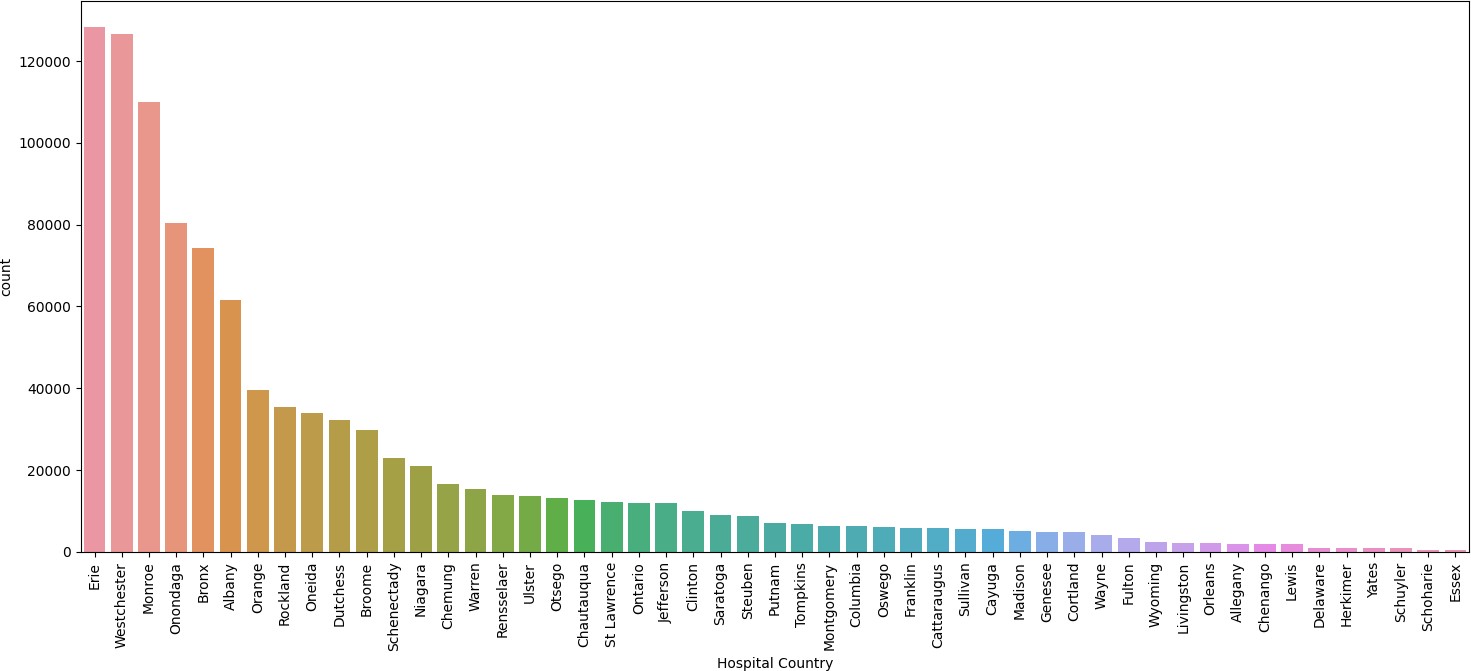
It is essential to ensure data security and privacy compliance, especially when dealing with sensitive health information. Regular updates and improvements to the chatbot’s capabilities are also crucial to maintain its effectiveness over time.

# Exploratory Data Analysis

The dataset has been taken from Kaggle <https://www.kaggle.com/datasets/nikitkashyap/hospital-insurance-claim-fraud>

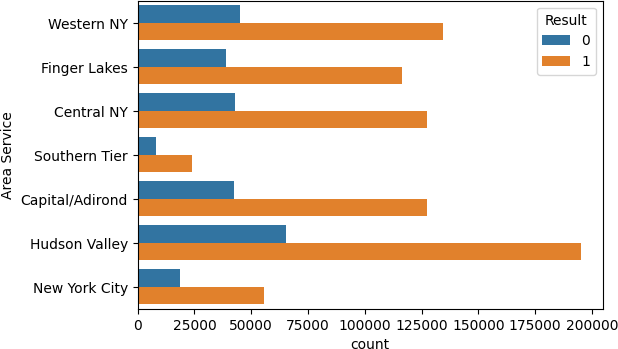
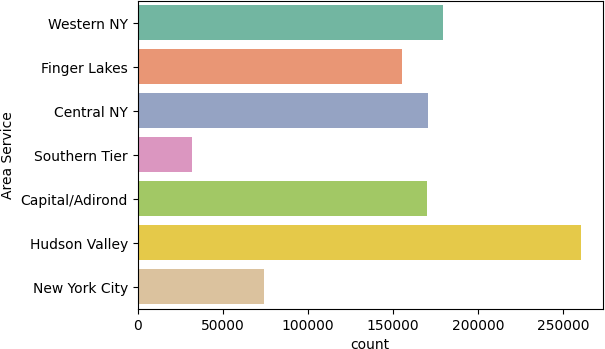
# Geographical & Demographic Analysis

## Hospital Country



**Observation:** Erie and Westchester are the two countries with 75% insurance frauds, while Essex is the country having minimum cases, but still accounts for 71% insurance fraud cases.

## Area of Service

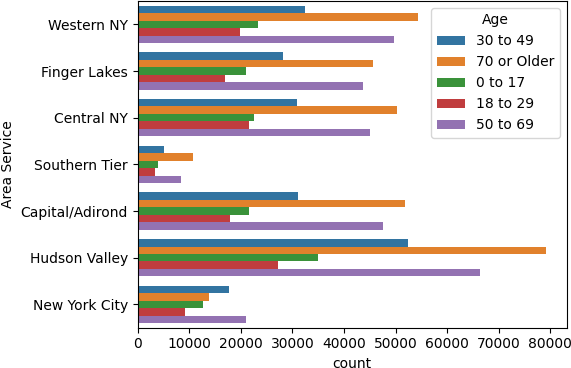
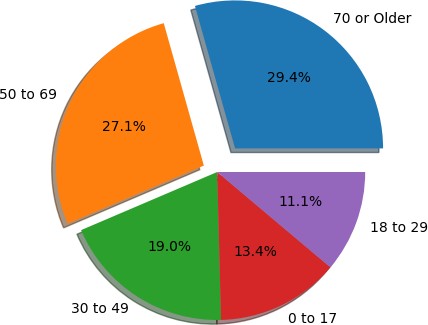


**Observation**: Hudson Valley is the area of service having the greatest number of cases, overall, all area of services has 75% insurance frauds.

## Age

Usually, older adults, particularly those eligible for Medicare, may be at a higher risk of being targeted for health insurance fraud. They often require more healthcare services, have more complex medical conditions, and may be less familiar with new technologies.

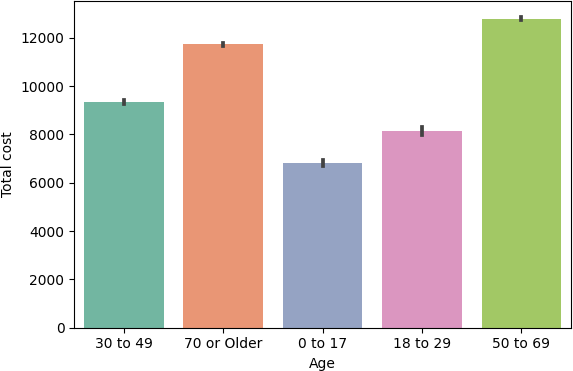
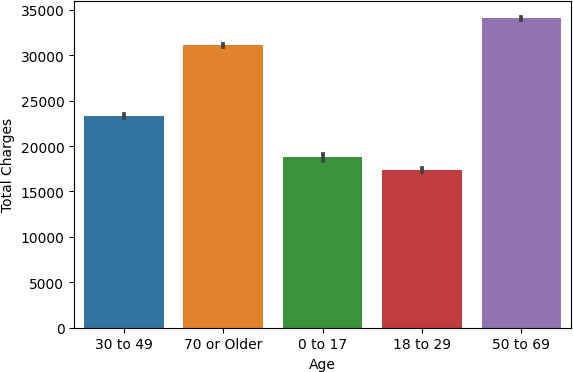
Fraudsters may exploit these vulnerabilities by billing for unnecessary services, identity theft etc. While not as prevalent as in the elderly population, frauds can also occur among young individuals.



**Observation**: People in their early 50s to late 70s have been targeted the most in each Area of Service.

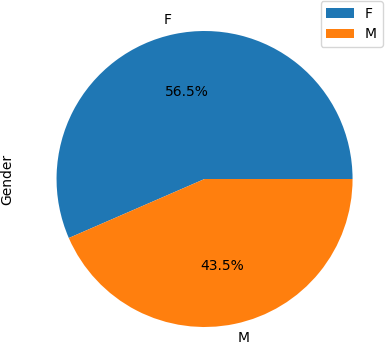
## Age and Total charges, cost

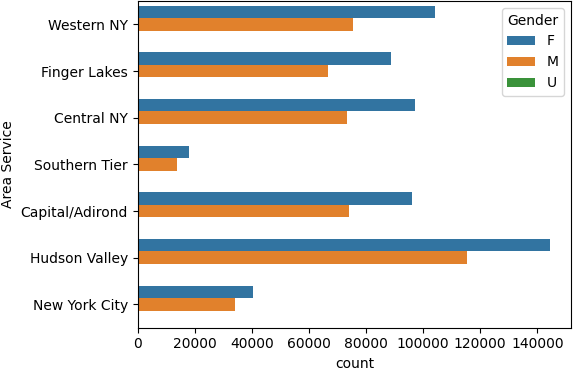
As, we all know our health insurance premium will increase with age as it is believed that as you get older, the chances of health problems are more likely to increase, you will need larger sums of money to cover your rising medical expenses, including healthcare, treatments, and even surgeries



**Observation**: It is quite evident from both the plots as well, that there is a linear trend between Age and the total charges, cost. As the Age is increasing, the total charges and cost is also increasing**.**

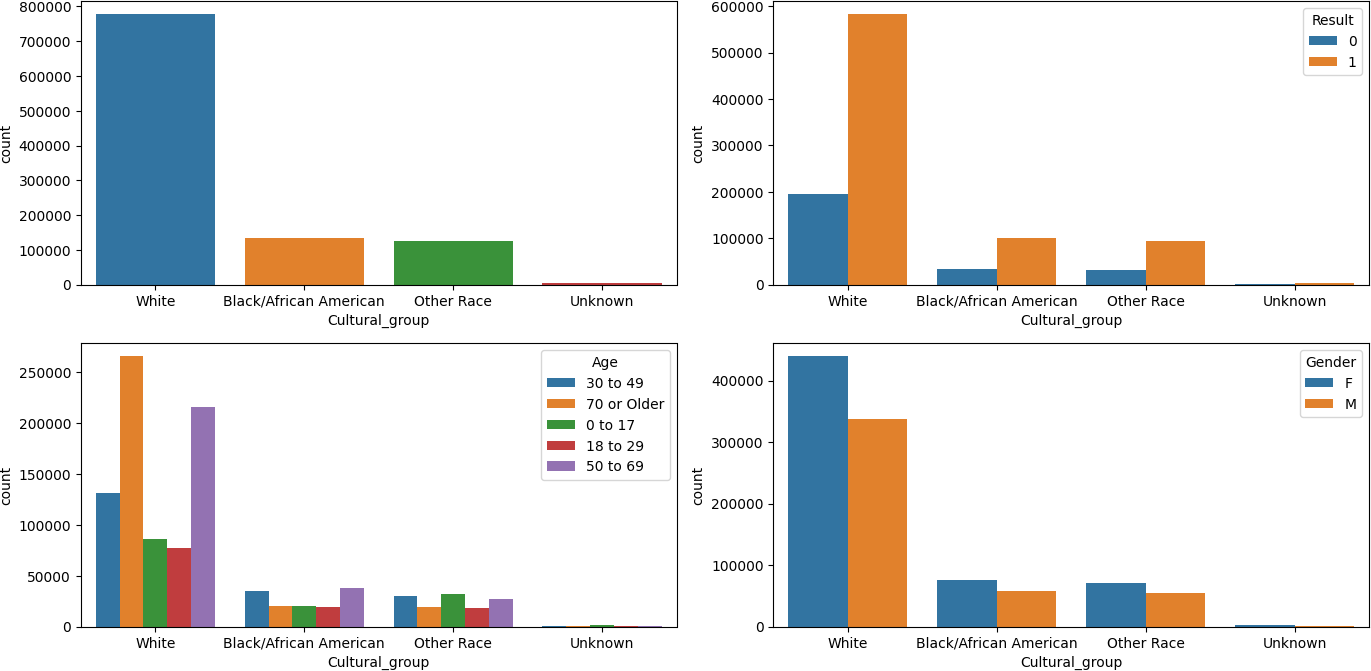
## Gender





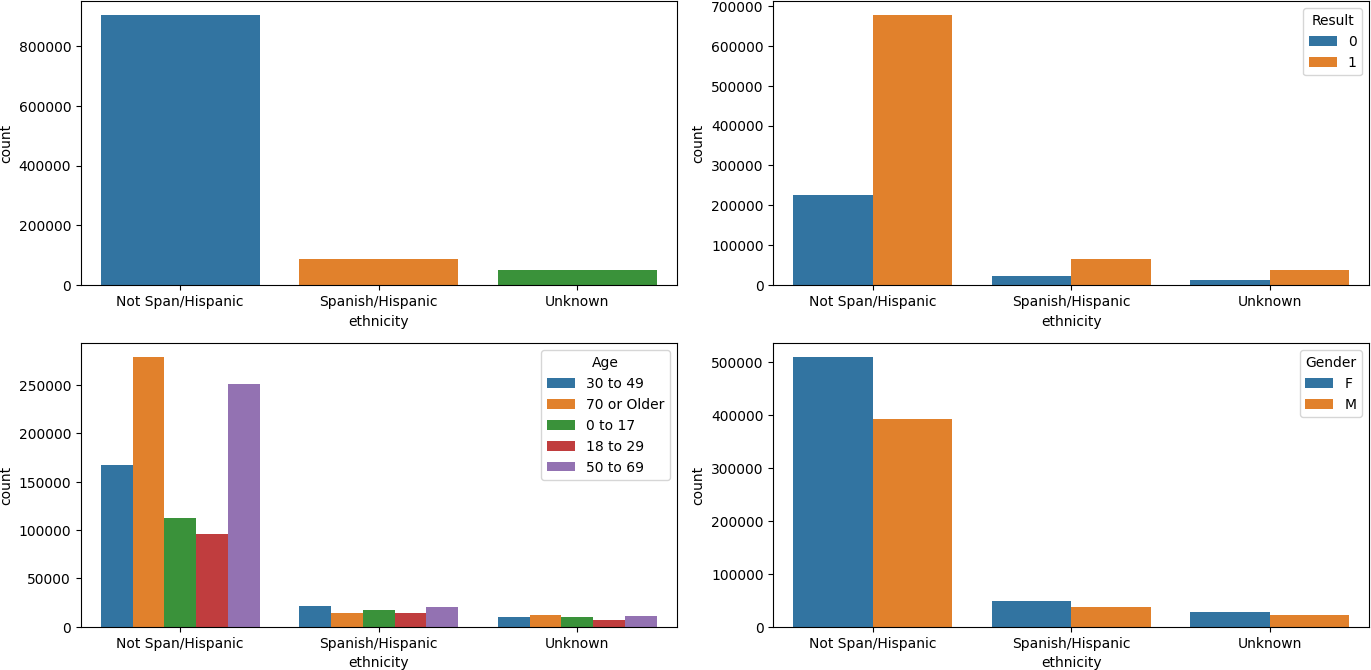
**Observation**: A lot of females have been targeted over male for insurance fraud cases.

## Cultural groups



**Observation**: Whites seem to be the most dominant cultural group where there is a huge number of insurance fraud claim cases. Whites in their early 50s to late 70s have been targeted the most, whereas other age groups have been targeted for different cultural groups. In most cases, females have been targeted over male.

## Ethnicity



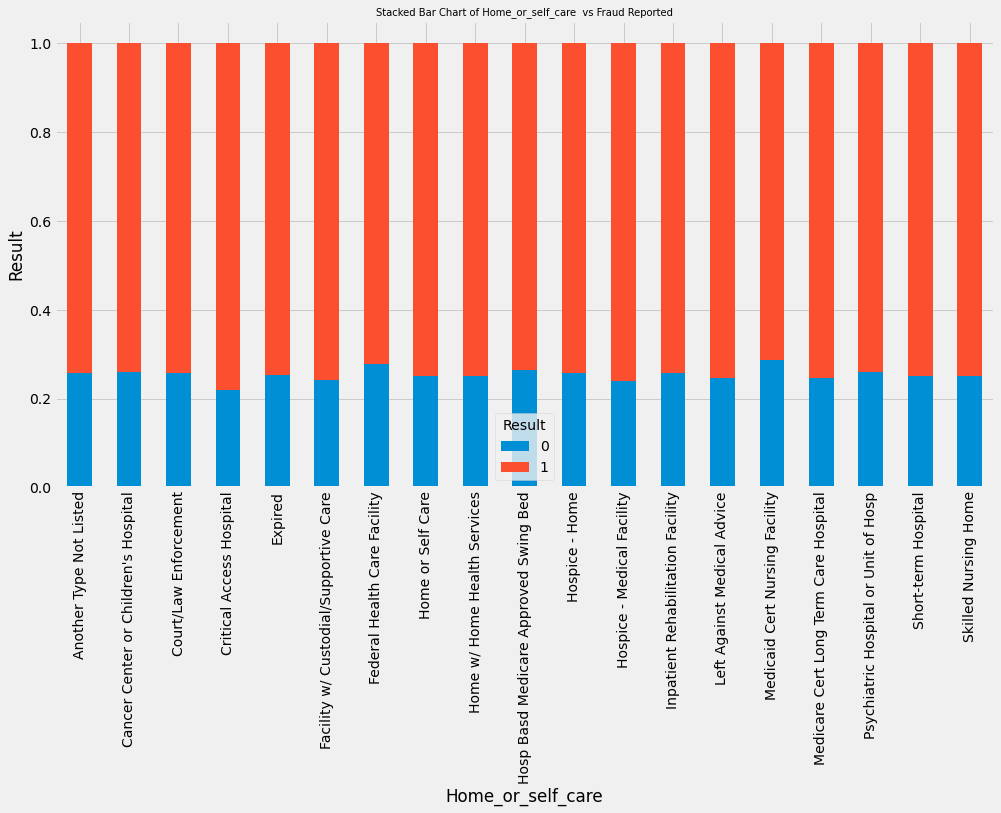
**Observation**: Not Span/Hispanic seems to be the most dominant cultural group where there is a huge number of insurance fraud claim cases. People in their early 50s to late 70s have been targeted the most, whereas other age groups have been targeted for different ethnicities. In most cases, females have been targeted over male

# 4.2 Psychographic & Behavioral Analysis

## Service and facilities acquired

The facility and services provided also impact fraudulent activities since there is no direct supervision over the facility. Considering associated factors, such as the frequency, duration, and nature of services claimed, it is possible to detect suspicious behaviors that may point to hospital insurance fraud.

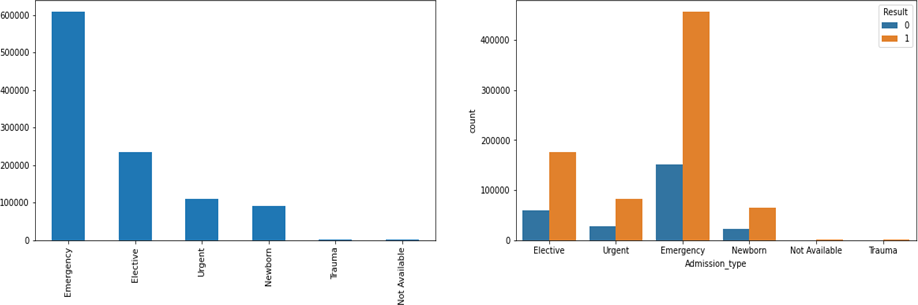




**Observation**: The graph shows that the majority of the patients opt for home or self- care while the other major options are home health services and nursing homes. These factors are also the one that contribute majorly to fraudulent cases. Rehabilitation facility has almost all their cases as fraudulent whereas children’s hospital and cancer enter have very low fraudulent cases put of its total.

## Admission type

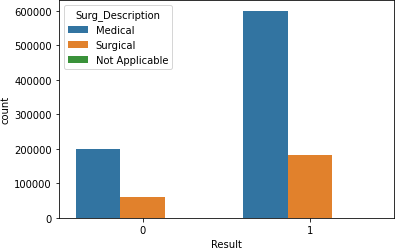
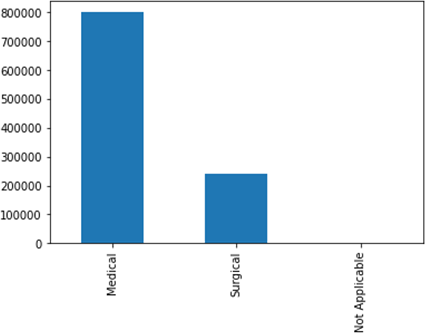
Analysis of elective admissions can help identify any patterns of unnecessary or excessive procedures being claimed, indicating potentially fraudulent activities. Analyzing emergency admissions can help detect any suspicious patterns, such as frequent and repetitive emergency claims for the same individual. Analyzing the patterns and characteristics of trauma admissions can help identify any suspicious trends or anomalies that may indicate potential fraud, such as inflated claims or fraudulent reporting of traumatic incidents.



**Observation**: As we can see from the left graph, the majority of the admission type are emergency cases while the least are related to trauma. Considering the large number of emergency cases, it has also resulted as the highest fraud count as well as the highest non-fraudulent count. This shows that there is a very good chance of fraudulent activities during an emergency visit of a patient.

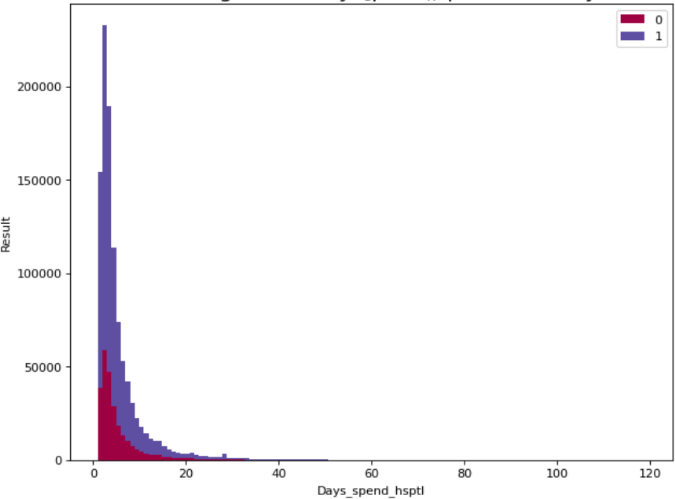
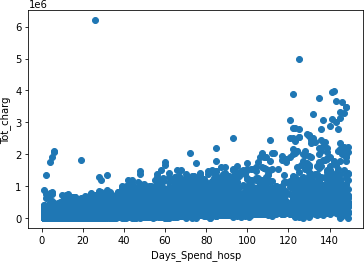
## Surgical Description

There are many factors involved in the decision of whether to perform surgery on a patient. The hospital and management can try to fraud the patient who has very little knowledge of what is going on. There can be many chances to include some fraudulent or extra charges under the pretext of surgery.



**Observation**: The graph on the left shows that majority of the treatment were non- surgical. There are many unknown factors which may lead to increased fraudulent activities in medical cases. As a result, it has high count of fraudulent cases.

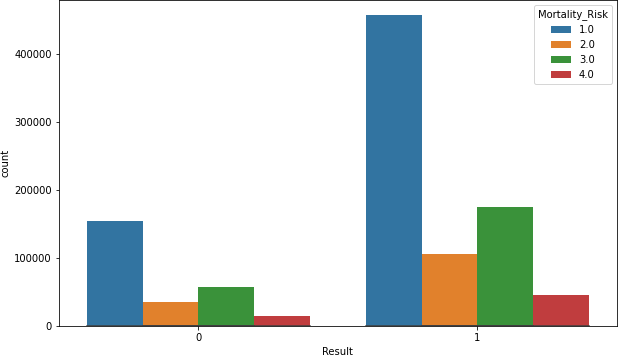
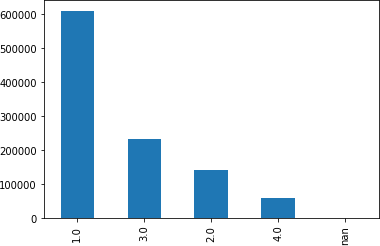
## Number of days spent in the hospital



**Observation**: The left graph shows the distribution of the number of days spent in the hospital by the patients and the total charge they had to pay for their treatment. The number of days ranges from 0 (same-day leave) to 150 with higher number of patients present in the starting range. The stacked plot on the right shows the histogram

distribution for the days spent while also displaying the fraudulent and non-fraudulent count.

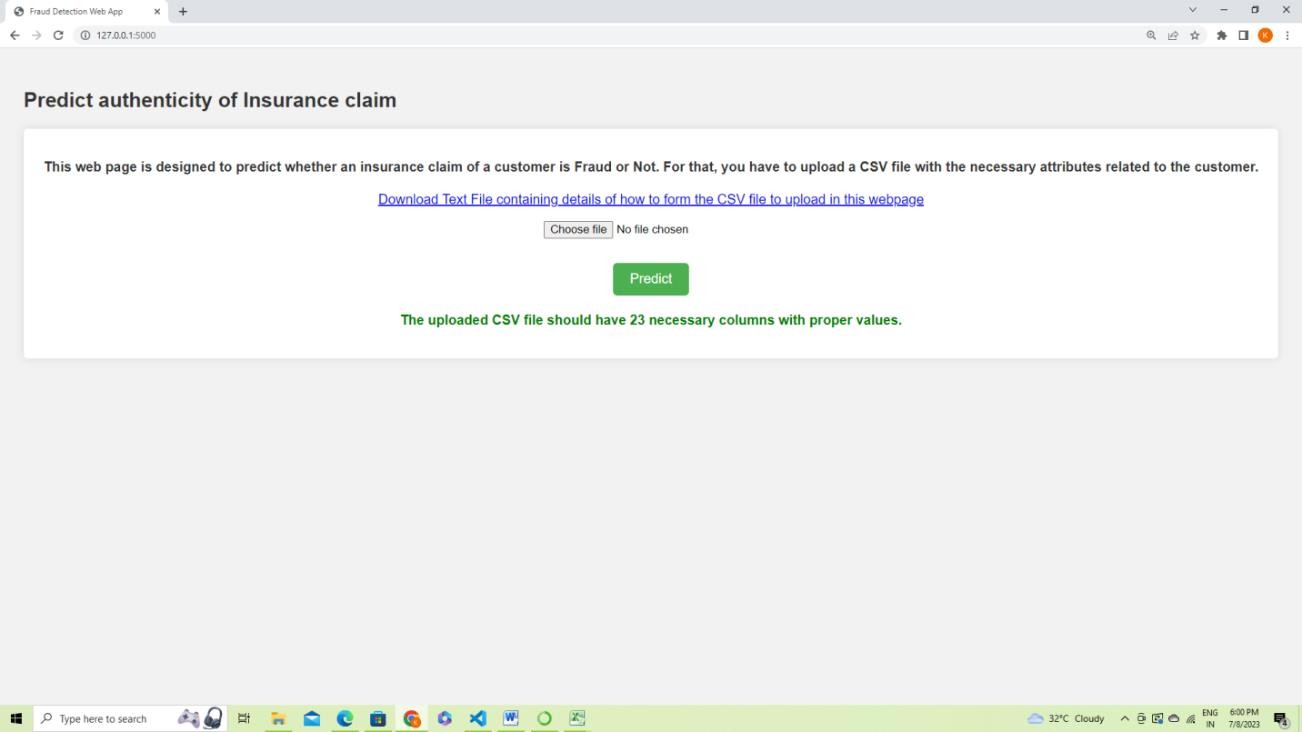
## Mortality risk

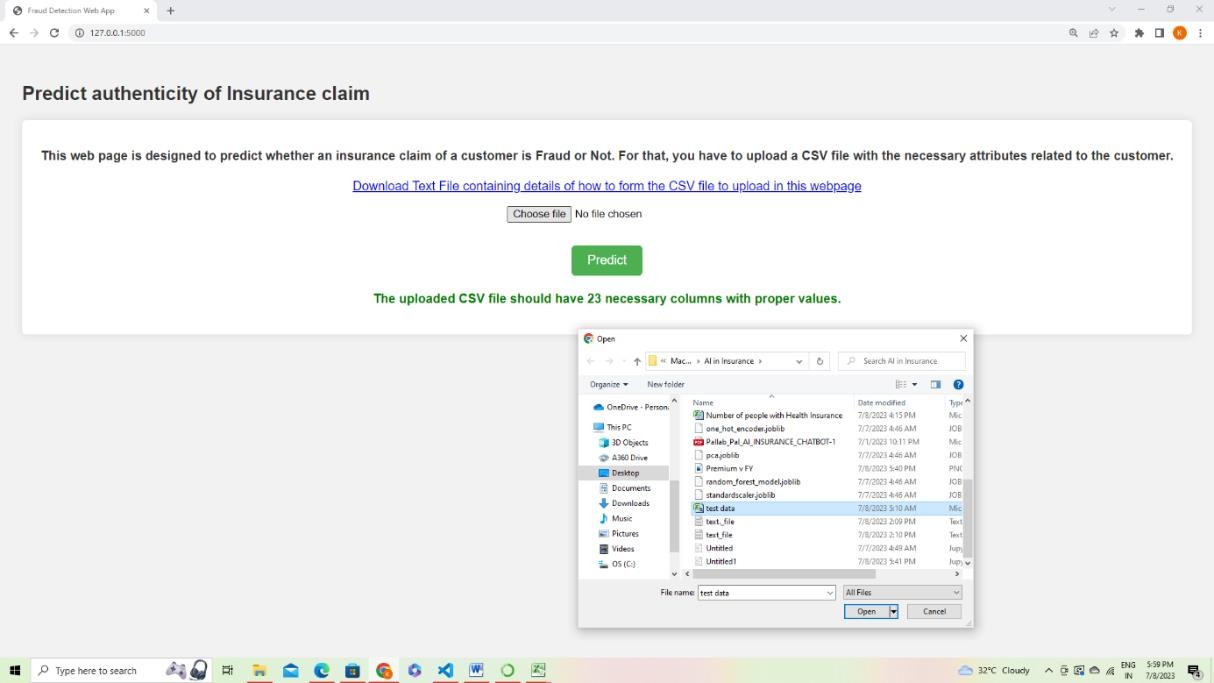


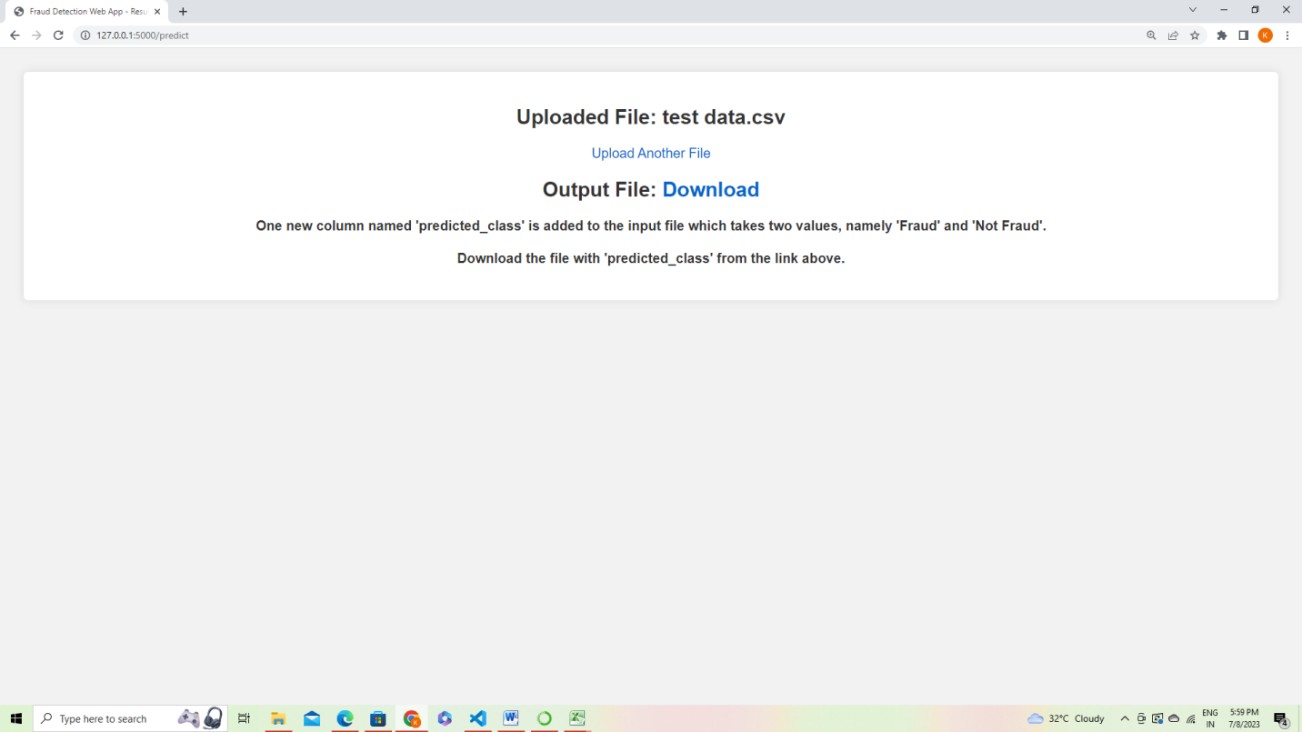
**Observation**: There are 4 mortality risks numbered 1 to 4. It is clear from the left graph that the highest count comes from mortality risk 1 and then 3. Following the same pattern, mortality risks 1 and 3 have the highest number of fraudulent cases.

# 4.0 Prototype Development

Developing a chatbot is a complex task that typically involves creating a conversational artificial intelligence system that can interact with users in a human-like-manner. This requires substantial time, effort, and expertise in natural language processing and machine learning. However, considering the short time frame available for the project, we decided that developing a fully functional chatbot would not be feasible. Therefore, we decided to develop a working webpage of our product idea which would determine whether the health insurance claim is fraudulent or not. Below are some screenshots for the same.





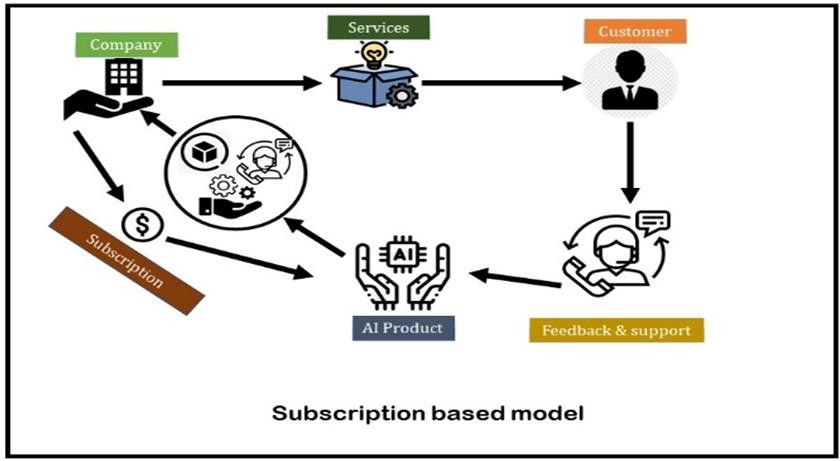


# 5.0 Business Modelling

This section of the report focuses on exploring the proposed business model for the presented idea. While there are numerous business models to consider, we have specifically selected the 'Subscription-based Business Model' as the most suitable option for our idea.

## Subscription-Based Business Model

In this model, insurance companies would subscribe to AI-powered insurance assessment and eligibility service on a recurring basis.



Here are some key aspects of the subscription-based model:

1. Subscription Tiers

Offer different subscription tiers with varying levels of features and benefits. For example, a basic tier with essential functionality and a premium tier with advanced features and priority support.

1. Pricing Structure

Define subscription pricing based on factors such as the number of users or claims processed, the size of the insurance company, or the level of customization required. Consider offering flexible pricing options, such as monthly or annual billing cycles.

1. Feature Differentiation

Clearly outline the features and benefits of each subscription tier. The basic tier might include core functionalities like damage assessment and coverage evaluation, while the premium tier could offer additional features like detailed reports, advanced analytics, or integration with third-party systems.

1. Scalability

Ensure that the subscription-based model is designed to accommodate the scalability needs of insurance companies. As their business grows, they should have the flexibility to upgrade their subscription tier or adjust their pricing accordingly.

1. Value Proposition

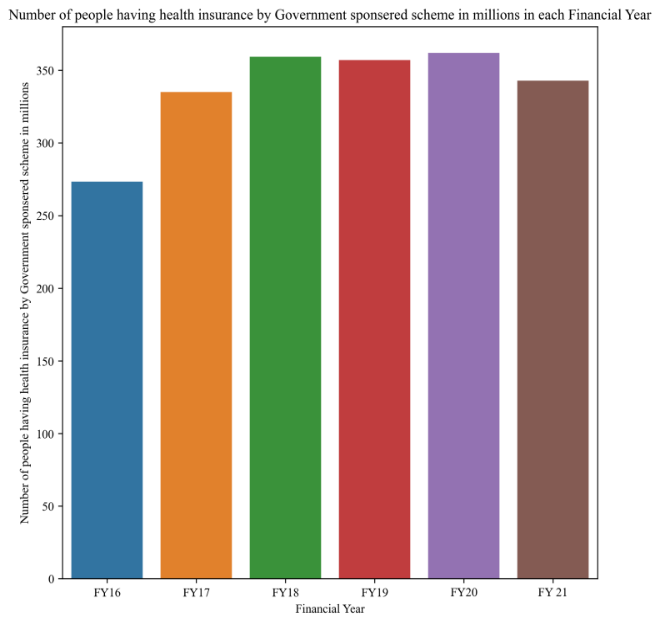
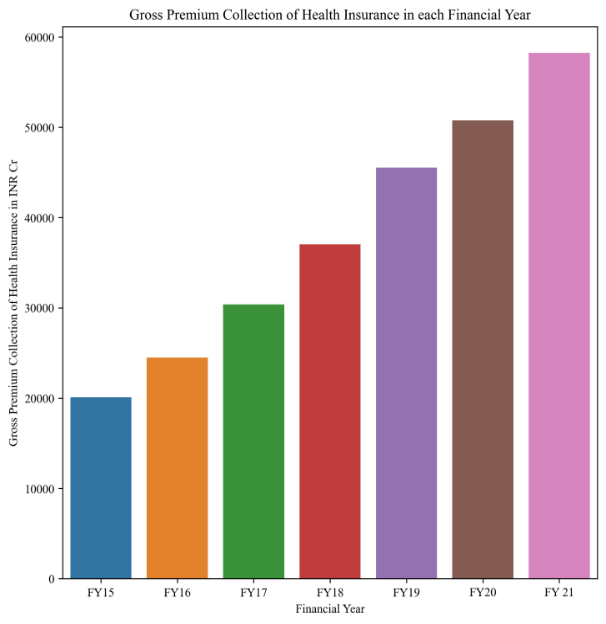
Communicate the value and cost savings that insurance companies can gain by subscribing the AI product. Emphasize how it can streamline claim processes, improve accuracy, reduce manual effort, and expedite claim settlements, resulting in enhanced operational efficiency and customer satisfaction.

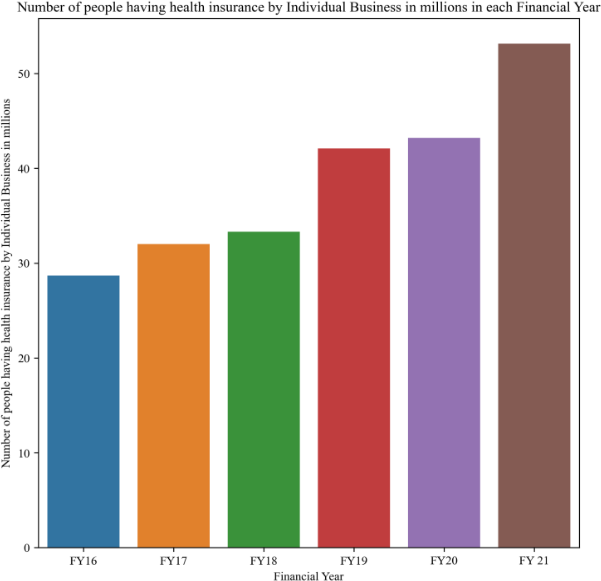
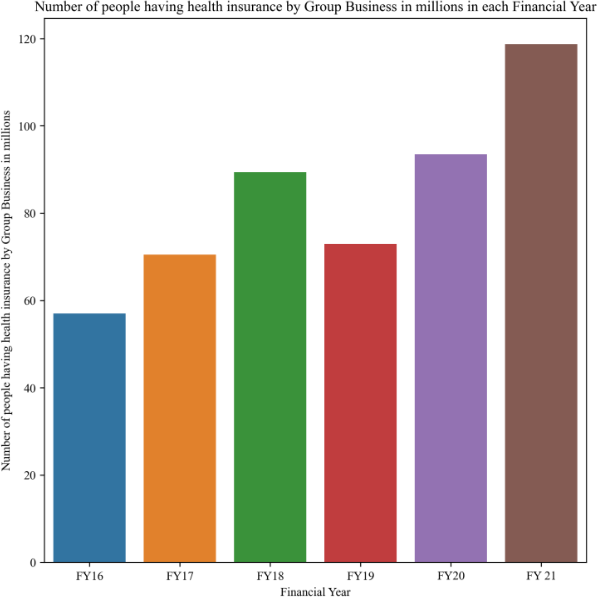
1. Customer Success and Support

Provide robust customer support to address any issues or questions that insurance companies may have during their subscription period. Offer training materials, documentation, and access to a dedicated support team to ensure a positive user experience.

# 6.0 Financial Modelling

The various areas where our product can be launched are as follows:

1. Insurance Companies: Health insurance companies themselves can use this product to assist their customers in reporting potential fraud and to provide information about various fraudulent activities. This can help companies improve their fraud detection processes and reduce fraudulent claims.
2. Healthcare Providers: Hospitals, clinics, and medical practitioners, can use this product to educate their staff and patients about insurance frauds, which can provide guidance on identifying and reporting suspicious activities, ensuring both providers and patients are vigilant against fraudulent practices.
3. Government Agencies: Various government agencies who are responsible for overseeing health insurance programs can use this product, which can aid in reducing fraud and protecting the integrity of government-funded insurance programs.
   1. **(b)**



**(c) (d)**

Data/Statistics were gathered from the sources mentioned below:

* [https://www.forbes.com/advisor/in/health-insurance/health-insurance-](https://www.forbes.com/advisor/in/health-insurance/health-insurance-statistics/#%3A~%3Atext%3DHealth%20Insurance%20Coverage%2DRelated%20Statistics%2Cthe%20people%20in%20the%20country.%26text%3DNearly%20400%20million%20individuals%20in%20India%20have%20zero%20access%20to%20health%20insurance) [statistics/#:~:text=Health%20Insurance%20Coverage%2DRelated%20Statistics,t](https://www.forbes.com/advisor/in/health-insurance/health-insurance-statistics/#%3A~%3Atext%3DHealth%20Insurance%20Coverage%2DRelated%20Statistics%2Cthe%20people%20in%20the%20country.%26text%3DNearly%20400%20million%20individuals%20in%20India%20have%20zero%20access%20to%20health%20insurance) [he%20people%20in%20the%20country.&text=Nearly%20400%20million%20in](https://www.forbes.com/advisor/in/health-insurance/health-insurance-statistics/#%3A~%3Atext%3DHealth%20Insurance%20Coverage%2DRelated%20Statistics%2Cthe%20people%20in%20the%20country.%26text%3DNearly%20400%20million%20individuals%20in%20India%20have%20zero%20access%20to%20health%20insurance) [dividuals%20in%20India%20have%20zero%20access%20to%20health%20insur](https://www.forbes.com/advisor/in/health-insurance/health-insurance-statistics/#%3A~%3Atext%3DHealth%20Insurance%20Coverage%2DRelated%20Statistics%2Cthe%20people%20in%20the%20country.%26text%3DNearly%20400%20million%20individuals%20in%20India%20have%20zero%20access%20to%20health%20insurance) [ance](https://www.forbes.com/advisor/in/health-insurance/health-insurance-statistics/#%3A~%3Atext%3DHealth%20Insurance%20Coverage%2DRelated%20Statistics%2Cthe%20people%20in%20the%20country.%26text%3DNearly%20400%20million%20individuals%20in%20India%20have%20zero%20access%20to%20health%20insurance)
* [https://www.statista.com/statistics/657244/number-of-people-with-health-](https://www.statista.com/statistics/657244/number-of-people-with-health-insurance-india/) [insurance-india/](https://www.statista.com/statistics/657244/number-of-people-with-health-insurance-india/)

Once the data were collected, the next step was to analyze them to uncover patterns, relationships, trends, or insights which typically involves employing statistical techniques, data visualization tools, or other analytical methods.

Figure 1: a) Gross Premium Collection of Health Insurance in each Financial Year. b) Number of people having health insurance by Government sponsored scheme in millions in each Financial Year. c) Number of people having health insurance by Group Business in millions in each Financial Year. d) Number of people having health insurance by Individual Business in millions in each Financial Year.

From the Figure 1a we can see that there is continuous upward and linear trend in Gross Premium Collection of Health Insurance in between Financial Year 2015 to 2021. Also, from Figure 1c and Figure 1d we can deduce that number of people buying Health insurance premium from Group business and Individual business is increased within Financial Year 2016 to 2021.

Our goal now is to estimate the amount of profit our company can make in each year by integrating AI in insurance fraud detection. Because it will be a new branch in our company, so we have to bear some fixed cost at the start in the form of acquiring a new location to set up a department, set up desktops and other associated things for employees. The variable cost will also come in the form of acquiring right number of skilled Machine Learning Engineers as the customer base grows, as well as paying subscription fees to the cloud service provider for hosting our server.

Considering the average price of Insurance premium as 20 thousand rupees in India, we can charge an additional 10% on the premium for the faster resolve of claims which will be achieved by integrating AI in insurance fraud detection. Charging the customer more than 10% of the premium might push them away whereas charging too low might not benefit the company at all. That is why 10% looked like plausible number to charge the customer for the service.

So, the cost of this extra service will be around two thousand rupees per year. If we estimate a linear growth of the number of our customer then we can write as follow,

y = ax + b (1)

In the above equation x represents the financial year (2016 to 2021) and y represent the number of customers of our company. An estimate has been made about the values of the co-efficient a and intercept b by fitting linear regression curve to the data shown in Figure 1c and Figure 1d. The values of a and b are taken as 4.7 and -9454. So, our final equation for estimation of size of our customer base is as follows,

y = 4.7x - 9454 (2)

If the fixed cost is considered as F and variable cost is considered as V, then the profit P can be said to be as follows,

P = 2000y - Vy - F (3)

After putting the expression for y from equation (2) we get the final equation for profit as,

P = 4.7 (2000 - V) x - 9454(2000 - V) - F

# 7.0 Implementation

1. Prototype Development & Financial Modelling
2. Exploratory Data Analysis