Pulse of the Nation:
A statistical investigation into the vital signs of Armenian healthcare

Anush Aghinyan, Yeva Avetisyan, Gurgen Hovakimyan

BS in Data Science

American University of Armenia

May 9, 2024

Contents

ABSTRACT

The exploration of Armenia's healthcare system proved to be both fascinating and a handful of challenges. In our capstone project, we set the ambitious goal to thoroughly investigate health insurance in Armenia.

This required us to navigate through a lot of obstacles, however having the clear goal in mind we were able to pass right through these difficulties. Over several months, our team dedicated itself to gathering relevant data, which involved extensive research and collaborations with healthcare insurance companies. Due to all of the effort we were not only able to get the needed data but were successful in the data cleaning stage - another crucial part of the capstone project.

Through our persistent work, we were able to reach the needed results and identify patterns and trends within the data. Our analysis included data visualizations, which were crucial in making the complex information more accessible and understandable. These visualizations were not only instrumental in our analysis but also enhanced our final R shiny dashboard.

The R shiny dashboard are effectively showcasing our findings, allowing us to present our findings in a dynamic and interactive format. The ability to manipulate the visual data in real-time will our audience with a clearer understanding of the health insurance landscape in Armenia. This capstone project is not only showing our analytical and programming skills but also deepens our understanding of Armenia's healthcare system. This was an excellent opportunity to handle real-world data and present it in an impactful, user-friendly manner.

In this report, we will dig deeper into each stage of our capstone project. We will take a look at how the data was gathered and how it was cleaned and analyzed. We will go over our visualizations and findings, and we'll talk about how the R shiny dashboard was created.

Our Capstone Project GitHub link: https://github.com/Yevaave/healthcare_analysis

Considering that our capstone project is done on very private data, which we got based on an NDA contract with the insurance company, our dataset is not going to be published.

1 INTRODUCTION

In Armenia's healthcare sector, the strategic usage of data would not only enhance operational efficiencies but might also initiate some policy reforms. Our capstone project centers around the exploration and analysis of health insurance data within Armenia, since the created framework can be used by various health insurance companies. As Jesse Sullivan wrote in their "Big Data and You: A Guide to Data Analytics in the Health Insurance Industry," using data science and frameworks similar to ours, health insurance companies in Armenia can provide the proper care and insurance plans to the right people at the right time, identify frauds before they happen, deliver more personalized insurance plans and much more. In our project, by using the data gathered from one of Armenia's largest health insurance providers, we will find some hidden patterns, discrepancies, and potential areas for improvement within the system.

Through the long process of data collection, cleaning, and analysis, our plan is to provide an overall view of the healthcare insurance sector in Armenia. In this project, we used many data visualization techniques and developed an interactive R Shiny dashboard to communicate the findings effectively, making it accessible to users involved in healthcare planning and policy making. The ultimate goal is to contribute actionable insights that could guide improvements in healthcare accessibility and efficiency across Armenia.

2 DATA GATHERING

Before submitting our capstone project idea, we knew that we need to address a critical step: data gathering. Understanding the potential complexities involved in finding the necessary data, we went ahead and started working on this task without delay. The process of finding the correct data was not the easiest part, but we were ready to spend time and effort on it.

After extensive research and many collaborations, we explored many sources to secure the essential data for our project. While we were able to find some readily available online datasets, they did not quite go into the insurance field as deeply as we had planned. Our research involved reaching out to key stakeholders within the healthcare sector in Armenia. One of our first wins was managing to secure a meeting with one of the largest health insurance providers in Armenia.

We talked about our project's goals and our plan to the director, mentioning that we want to find some patterns and potential issues within Armenia's healthcare system. Inspired by our initiative and the potential framework, the director agreed to collaborate with us. We had a long discussion to select the most relevant data attributes, including specific columns, date ranges, and the scope of data necessary for our study. After careful selection and negotiations, the company gave us the private dataset based on an NDA contract, considering that the data is a part of private company information. With the data in hand, we were now ready to dig deep into the next phase - data cleaning and analysis- and excited to extract meaningful insights that could potentially influence healthcare practices and policies in Armenia.

3 DATA CLEANING

With our objectives clearly defined and our data on hand, we started the next critical phase of our capstone project: data cleaning and analysis. This stage required massive attention to details and a lot of work to ensure clean data, which is very important for a meaningful analysis. We did encounter many challenges when working on this phase, including missing values, suspicious inputs, and inconsistencies that could have prevented us from finding valuable insights and creating visualizations. Due to our programming skills and knowledge of necessary Python commands, we successfully navigated these obstacles.

One of the significant challenges we faced was the language issue. Since it was taken from an Armenian health insurance company, our dataset was fully in Armenian. This required us to find and use Python dictionaries to translate our data, transforming the dataset into a more suitable format for our further analysis.

Another thing that was not making our job easier was the geographic part of our data. Our initial dataset included various small villages and cities, some with only one or two data entries, which would affect our visualizations and analysis, making these harder to read. To address this, we decided to reorganize the data by grouping these smaller entries into Armenia's regions. Our team got the Armenia's map and the necessary Python libraries, grouping all the small cities and villages into the 11 Armenian regions. This reclassification also helped us to have a clearer visualization and deeper understanding of regional health insurance trends. With a cleaner and more organized dataset, we were well-prepared to move into the deeper analytical phases of our capstone project, equipped with data that was both accessible and analytically valuable.

4 ANALYSIS, VISUALIZATIONS AND OUR FINDINGS

Having the data all clean and ready, we were excited to start the main part of our capstone project, which was analysis and visualizations. We decided to start with statistical analysis in Python. We did a descriptive analysis, which showed us the mean, std, min, and max values of claim status and insured individuals' ages. Descriptive analysis shows that out of 249755 values in the work sphere column, 117449 are from the Information and Connection fields. For the Sex column, out of 249755 are Females, the rest are Males and Others. Out of the Illness Categories with 249755 values, Dentistry is at the top with 42750 rows. For regions, we noticed that Yere-

van has 219568 values out of 249754 overall values. Which means that while Yerevan has the 60% of Armenia's population, it has the 87% of overall insured people.

We did a grouped analysis, showcasing mean claim amounts by region.

```
Mean claim amount by region:
Region
Aragatsotn     25151.633317
Ararat         27716.370055
Armavir        25872.933764
Artsakh        28104.226131
Gegharkunik    25986.911421
Kotayk         25934.171422
Lori           28517.681994
Shirak         28808.555624
Syunik         29601.977915
Tavush         25362.423184
Vayots Dzor    26837.892050
Yerevan        29445.130930
```

Fig. 1. Mean claim amounts by region.

Next, we tried Frequency analysis, showing the frequency of each illness category. The top 5 being Dentistry (42750 values), Family Doctor (30033 values), Ophthalmology (25324 values), Annual Medical Examination (23514 values), and Gynecology (15164 values).

After using Python commands for the statistical analysis, we moved from Jupyter Notebook to R studio for visualizations using R.

In the paper "Decisions Through Data: Analytics in Healthcare," Mary J. Wills says, "The amount of data in healthcare is increasing at an astonishing rate. However, in general, the industry has not deployed the level of data management and analysis necessary to make use of those data". With our visualizations, we want to show how all this data can be used to make better decisions in healthcare insurance in Armenia.

Our analysis of Armenia's healthcare insurance data first goes into the demographic and regional distribution of insured people. It also goes into insured individuals' claim patterns and the diverse illness categories experienced and treated during our data's time range.

The age distribution graph illustrates a bimodal age distribution with peaks around the ages of 25 and 50, which means that mostly people between these ages are the ones using healthcare insurance.

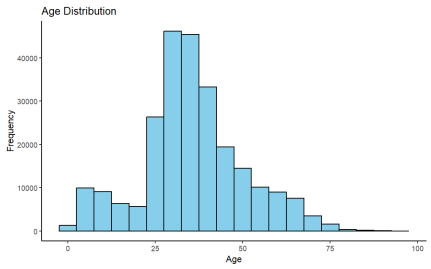


Fig. 2. Illustration of age distribution of insured individuals.

We also found out that males show a more extended distribution, implying a broader age range of coverage compared to females.

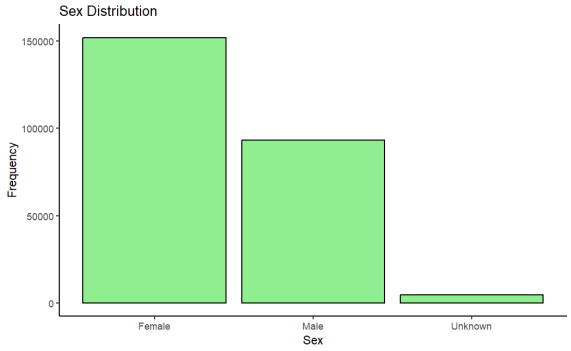


Fig. 3. Illustration of sex distribution of insured individuals.

The regional analysis highlights Yerevan's absolute dominance in insurance claims. The overwhelming majority of cases are from Yerevan, which proves the centralization of healthcare services in the capital. Other regions like Kotayk and Lori also show activity but are still far away from Yerevan, indicating potential accessibility and service availability issues in rural areas.

Work sphere distribution data depict a concentration

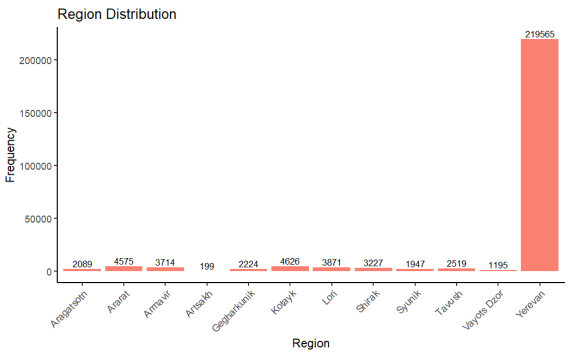


Fig. 4. Illustration of regional distribution of insured individuals.

of claims in sectors such as education, healthcare, and professional, scientific, and technical activities. This concentration is most likely because these sectors are better covered by insurance plans or are more aware of the importance of healthcare insurance.

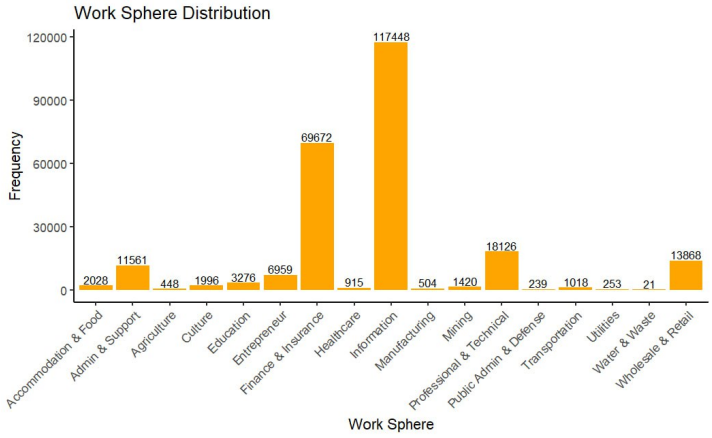


Fig. 5. Illustration of work sphere distribution of insured individuals.

In terms of healthcare claims, the majority are satisfactorily resolved, as shown by the claim status distribution. However, here we see a significant number of claims are either unsatisfied or canceled, which suggests changes in the healthcare insurance processes.

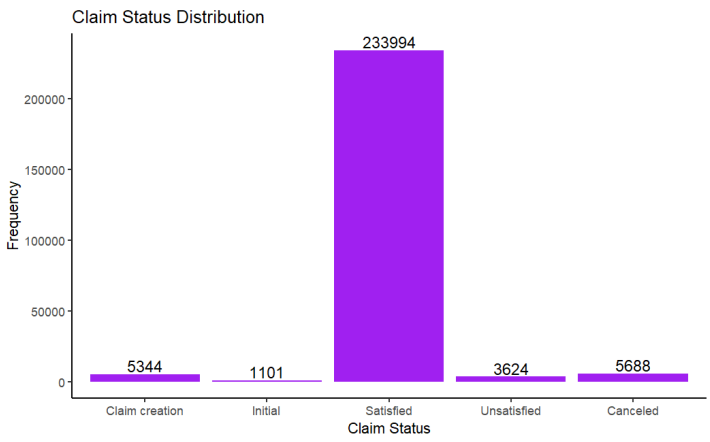


Fig. 6. Claim status distribution.

Distribution of illness category indicates a high frequency of claims in categories such as general surgery, internal medicine, and dermatology. This distribution not only reflects health issues within the insured population that happen more often than other, but also indicate areas where insurance coverage is most commonly used.

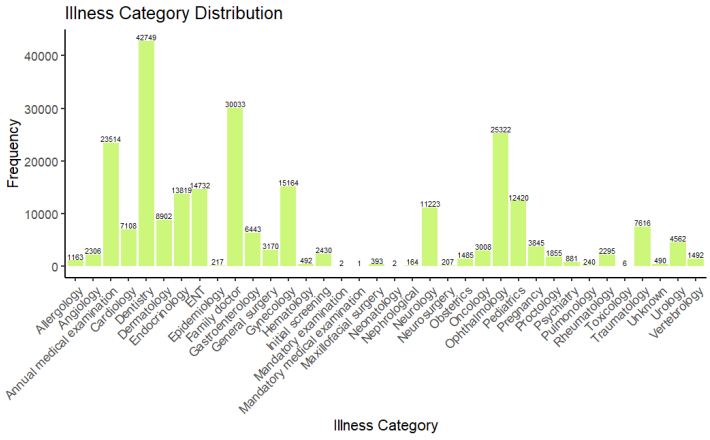


Fig. 7. Distribution of illness categories.

The analysis of claim amounts by sex shows that males have higher median claim amounts than females.

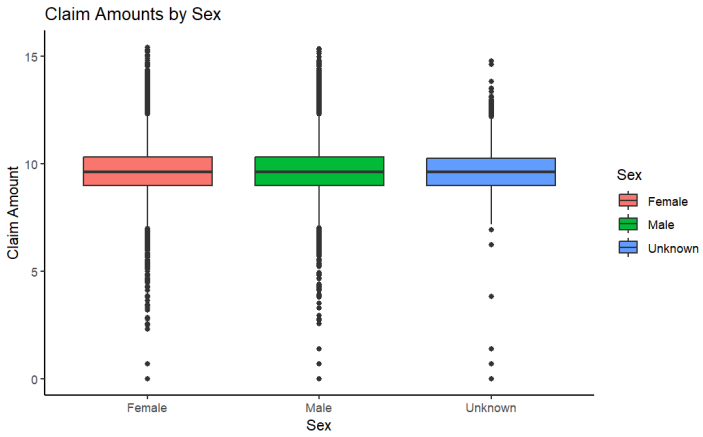


Fig. 8. Distribution of claim amounts by sex.

We also can see that while claim amounts are somewhat consistent across the regions, Yerevan exhibits not only the highest frequency of claims but also a significantly wider range of claim amounts, which might indicate a larger variance in the types of healthcare services accessed. Regions like Aragatsotn, Ararat, and Armavir show a narrower range of claim amounts. This shows how different healthcare service availability and type of healthcare facilities are across regions.

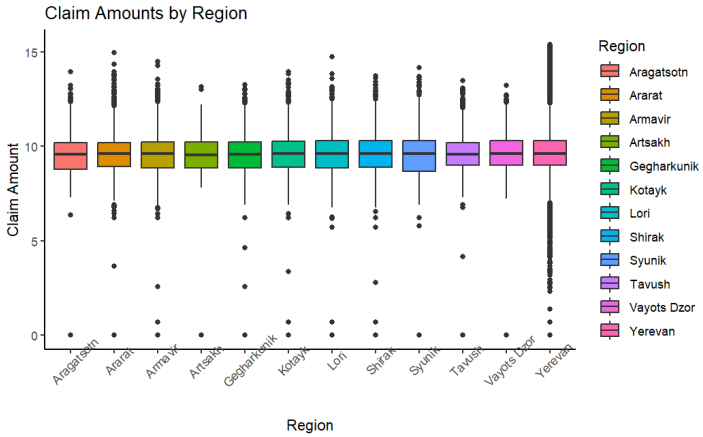


Fig. 9. Distribution of ilness categories.

As already during statistical analysis, for better visibility, we created view of the most frequent illnesses reported in healthcare claims. Dentistry is in the first position, followed by visits to family doctors and ophthalmology, annual medical examination and gynecology. The high frequency of dental and eye care issues highlights their prominence in healthcare claims, suggesting areas where public health interventions or insurance coverage adjustments may be necessary to meet the population's needs.

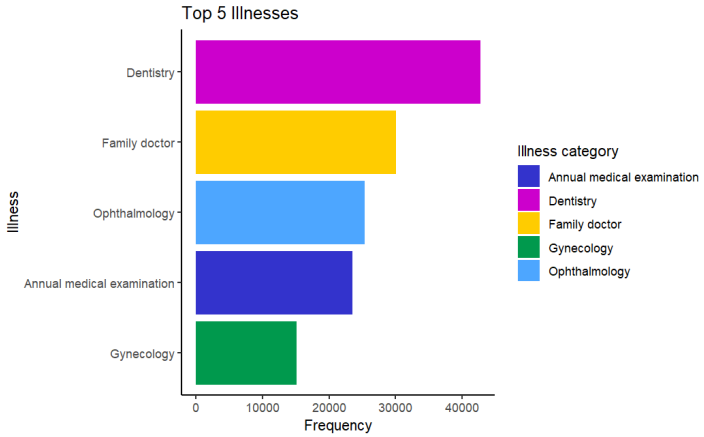


Fig. 10. Top 5 illnesses.

Seeing that dentistry is the top illness category we decided to take a look at ages most often having dental claims. Our analysis are showing a prominent number of claims among individuals aged 28 to 34. The relatively young poeple being at the top underscores the potential emphasis on dental health issues prevalent in this age range. However, this trend may indicate specific health awareness or benefits targeting younger adults in Armenia's dental care sector.

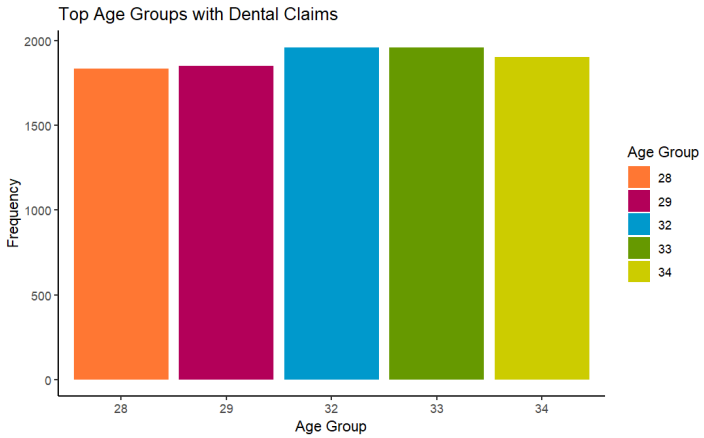


Fig. 11. Dentistry top age groups.

Next, we went to distribution of eye doctor visits across different work spheres, considering it was also one of the top illness categories. We see the Information sector leading, followed by Finance and Insurance. This suggests that individuals in this

work spheres are required to have significant visual focus and higher screen time.

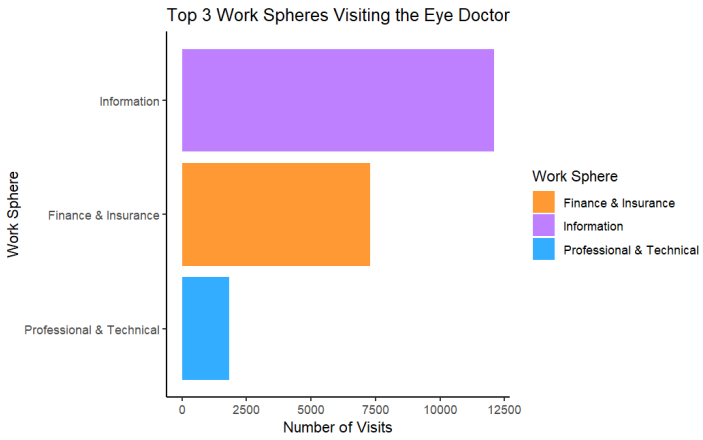


Fig. 12. Ophthalmology top work spheres.

We also decided to take a deeper look at oncological claims. Our analysis indicate a higher occurrence of claims among middle-aged adults, specifically those between 33 to 40 years old. This suggests the potential need for targeted cancer prevention and screening programs within these age groups.

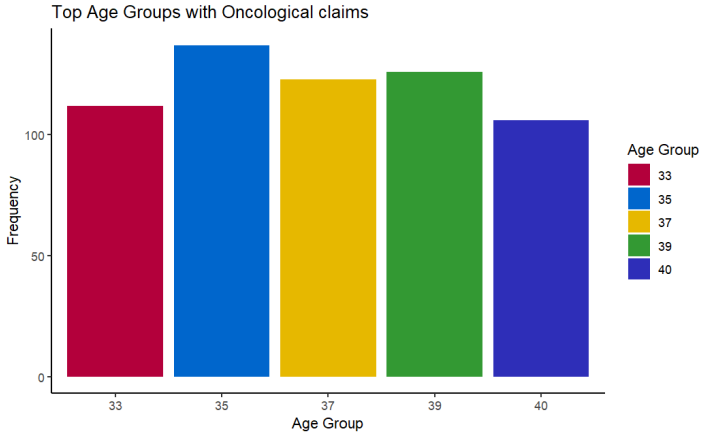
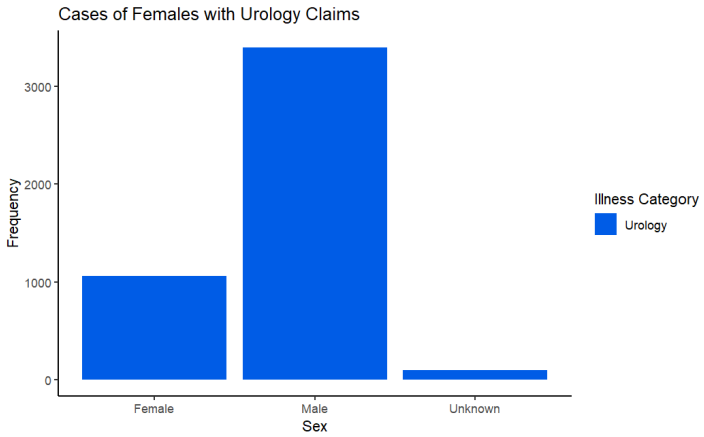
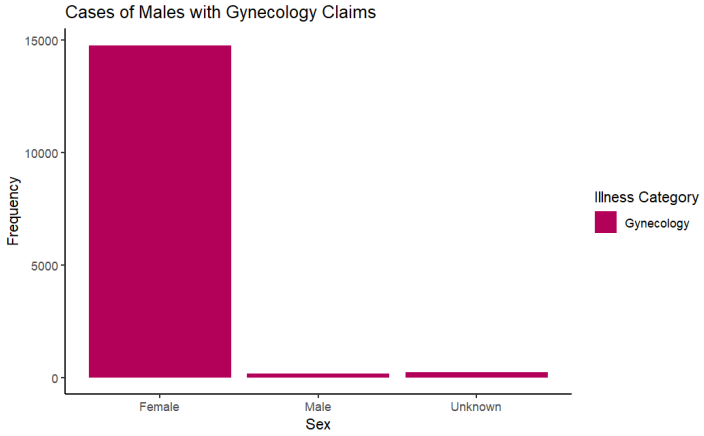
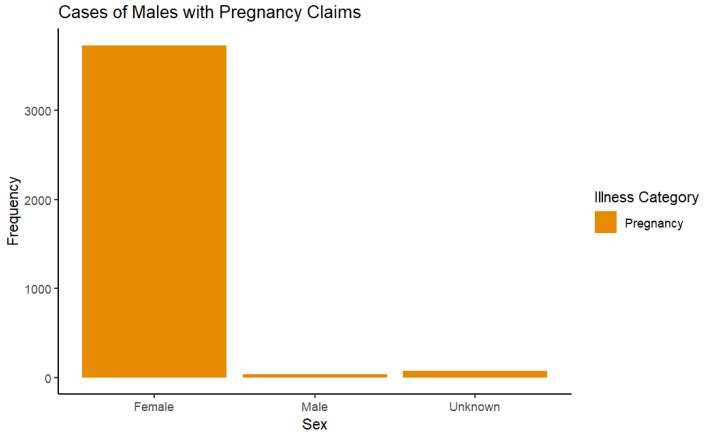


Fig. 13. Oncology top age groups.

We went ahead and investigated a bit deeper values with work spheres ”Gynecology” and ”Urology”. A fun fact we found here is that many males attended gynecology doctor during our data range and many females attended urology doctor. Out of curiosity, we dug a bit more and found a few male pregnancy claims.







The data visualization and analysis not only provide a clear picture of the healthcare insurance landscape in Armenia but also pinpoint several areas for policy intervention, particularly in improving access and satisfaction with healthcare insurance outside the capital and among older age groups.

5 R SHINY DASHBOARD

With the completion of our visualizations, we finally started the capstone's final stage: developing a dashboard that would showcase our analysis in a more user-friendly way. Coming out of previous experiences, we decided to go with the R Shiny dashboard and immediately started our work.

Clarity and organization were some of our top priorities here, so we distributed our visualizations across several tabs. The first of these, the "Demographic Insights" tab, offers an overview of the data we did our analysis on to make sure our users have a deeper understanding of our findings. This tab shows that our analysis was done on a diverse group of individuals—varying by gender, age, geographic location throughout Armenia, and work spheres. This initial introduction of the data prepares the users for more detailed examinations in the next tabs.

As we are done with the tab of the demographic overview, we encounter more specialized analysis in the following tabs: "Healthcare Claim Analysis," "Box Plot Analysis," "Interesting Insights," and "Fun Facts." Each tab is designed to take the reader deeper into different parts of our analysis, providing different insights and highlighting significant patterns and trends within the data.

The tabs "Healthcare Claim Analysis," "Box Plot Analysis," and "Interesting Insights" contain the biggest part of our analysis, being the tabs with most of the patterns we found and questions we raised. However, with the amount we had, we could not help but notice a few interesting claims that we decided to showcase under the "Fun Facts" tab.

Considering that our capstone project is done on very private data, which we got based on an NDA contract with the insurance company, our dashboard is not going to be published.

6 CONCLUSION

The journey through data gathering, cleaning, and detailed analysis culminates in our capstone project, which has revealed to us many insights into Armenia's health insurance dynamics. The project shows us the variation in insurance coverage and service satisfaction across different regions and demographic groups.

Our findings suggest critical intervention points for policy makers, particularly our project is showcasing that a huge work should be done on healthcare accessibility and satisfaction outside our capital Yerevan, and addressing the insurance needs of the elderly population. The R Shiny dashboard developed as part of this project can help the stakeholders to better see and understand the data analysis made. Due to interactive data visualizations, such complex data can be comprehensible and engaging to a broad audience.

This capstone project stands as a bridge between data analysis and practical healthcare insurance improvements. In their article "Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance," the authors discuss how data analysis should help insurance companies in making decisions on what services should be covered and how they can be sustained to meet the health needs of people. Our project once more reflects the potential of data-driven decisions in crafting a more inclusive and efficient healthcare system in Armenia. We hope that the insights provided will initiate action towards optimizing health insurance practices and policies in our country.

REFERENCES

[1] Ho CWL, Ali J, Caals K. Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance. Bull World Health Organ. 2020 Apr 1.

[2] Sullivan, J. (2024, April 2). Big Data and You: A Guide to Data Analytics in the Health Insurance Industry. Hitachi Solutions. https://global.hitachi-solutions.com/blog/big-data-in-health-insurance/

[3] Decisions through Data: Analytics in Healthcare: Journal of Healthcare Management. (n.d.). LWW.