Premium Payment Behaviour in an Insurance Company

Author Name: Mr. Hardik Wadhwa
B.Tech, Department of Information Technology
Maharaja Agrasen Institute of Technology, Sector-22, Rohini, Delhi –110086

Abstract

As the number of Indians who understand the value of investing in insurance increases, the study for understanding the basic behaviour of customers has taken a precedent to understand the needs of customers as well as to allow the company to support its customer base more effectively. This paper focuses on the payment receipts of customers to understand the policy premium paying pattern followed by an insurance companies customer over a period of time. The main objective of this paper to identify customers behaviour of paying the policy premiums and will they pay their next premium on time or not and in doing so help the company in understanding where additional effort has to be paid to help the customer and improve the rate of policy premium payment. Data was collected by various sites and some previous years data of some policy companies. Frequencies, Tabulation and some Data Science models have been used for the analysis. The objective of this project is to convert the generated data to graphs and plots, to create comparisons between various factors that can be used to analyse the data effectively and generate an algorithm that can be used in real-life applications to give accurate findings based on the previously generated data.

Keywords: Insurance, Customer's Behaviour, Premium paid, Defaulters

INTRODUCTION

Health insurance in a narrow sense would be 'an individual or group purchasing health care coverage in advance by paying a fee called the premium. The health insurance market in India is very limited covering about 10% of the total population. Health insurance expenditure in India is roughly 6% of GDP, much higher than most other countries with the same level of economic development.

Of that, 4.7% is private and the rest is public. In private insurance, buyers are willing to pay a premium to an insurance company that pools people with similar risks and insures them for health expenses. The key distinction is that the premiums are set at a level, which provides a profit to a third party and provider institutions. Premiums are based on an assessment of the risk status of the consumer (or of the group of employees) and the level of benefits provided, rather than as a proportion of the consumer's income.

For example, you may pay a premium of Rs. 5000 each year for a medical insurance cover of Rs. 200,000/- so that if God forbid, you fall ill and need to be hospitalized in that year, the insurance provider company will bear the cost of hospitalization etc. for up to Rs. 200,000. Now if you are wondering how can the company bear such high hospitalization cost when it charges a premium of only Rs. 5000/-, that is where the concept of probabilities comes into the picture. For example, like you, there may be 100 customers who would be paying a premium of Rs. 5000 every year, but only a few of them (say 2-3) would get hospitalized that year and not everyone. This way everyone shares the risk of everyone else. So, if a large number of customers do not pay the premium on time, it might disrupt the cash flow and smooth operation for the company. A customer may stop making regular premium payments for a variety of reasons - some may forget, some may find it expensive and not worth the value, some may not have money to pay the premium etc.

Researching whether a customer would make the premium payment can be extremely helpful for the company because it can then accordingly plan its communication strategy to reach out to those customers who are less likely to pay and convince them to continue making timely payment.

Now, to predict, whether the customer would pay the next premium or not, we have information about past premium payment history for the policyholders along with their demographics (age, monthly income, area type).

Literature Review

Literature Review is required to take the matter into considerations that can't be cleared in the past researches. Many researchers try to interpret various kind of conclusions and to improve those past results literature review is needed. The present literature serves many varied interesting features, which forms the vital background for the study and conducted a consideration

Jagendra Kumar (2005) by his study, revealed that The Life Insurance penetration, in India is Just about 2% of GDP. The life insurance premium per capita is just Rs. 550/- The LIC is the largest player with over 2000 officers. After liberalization, it has improved efficiency and customer services, among the private life insurance companies ICICI Prudential Life Insurance and Birla Sunlife are the first and second-largest players. Other prominent companies in competition are – Bajaj -Allianz, HDFC Standard Life, Kotak Mahindra, ING Vysya Aviva Life, and Metlife etc.

Viswanadham (2005) studied claims settlement operations of insurance companies with the objectives of evaluating performance in terms of both maturity and death claims before and after the IRDA period. Claim settlement processing time expressed in speed ratios and adjudicatory measures of the corporation to redress the grievances of policyholders in settlement of claims. The study concluded that corporation should provide efficient service with courtesy in the matters of claim settlements. It should create the highest trust in the minds of policyholders by establishing an open and transparent grievance redressal procedure. As a satisfied customer will be a brand ambassador for the insurance company; claim settlement should be given more importance.

Jagendra Kumar (2005) by his study, revealed that customers of different age level and maturity levels, take different approaches to pay the premium. Customers at the starting years of the policy pay the premium timely and frequently and as the year's pass, they become less responsible.
Moreover, people of the age group of 30-45 take the policy more seriously. In all these responses emotional factor of people to their family and their working schedules are the main reasons for their approach of payment of premiums.

Yusuf, Gbodamasi and Hamadu (2009) conducted an empirical study on the customers that income of these customers somehow plays a role in their manner of paying the premium. It is observed customers having low income sometimes deliberately tries to avoid the payment of premium. On the other hand, people having high income or we can say, constantly having a higher income rate pays the premium on time and sometimes in advance. This can be observed that income plays a vital role in the payment of premiums of the policy.

Neuman (2007) worked on various factors on which a company evaluates the customers. Customers mostly forget about the premium payment as after the due date they usually ignore it or postpone it for a time. But after some time, they didn't bother about the payment as well. Due to this companies usually marked this kind of customers as defaulters. This study concluded that the irresponsible nature of customers is the major cause for the customers not to pay the premium on time.

Albert (2013) noted in his article that the age of the customers played the most important role. As the customers of older age forget to pay the premium on time. As age is the factor, the memory becomes weaken day by day. Even at the time, they forget about details and sometimes forget about what kind of policy they have registered. This study concluded that customers with higher age forget to pay policy because of the priority issues and sometimes mental conditions also considered.

Objectives of the Research

The main objective of the research paper is to identify the customer's likelihood of paying premium continuously in the near future thus showing their value for the company, allowing for a more focus based growth in the future.

Research Design

Research has been done on various stages as listed below.

- The decision of necessary factors required for the research
- Creating the database
- Data Visualization
- Data Extrapolation

1. Factors Required: Deciding the various factors and listing them as necessary and unnecessary for our project based on our hypothesis. Hypothesis testing is the part where we picturise all our requirements from the data before looking at the actual data. Below factors are taken into consideration.

i. Unique ID of the customers:- this will serve as the primary identifier of the customer and help in mapping the various activities of a customer in our case it will be their premium payment receipts.

ii. Age in days of the policyholder:- the number of days the policy has been taken by the customer it is done to make the data continuous rather than discrete(which would be the case if the number of years was considered)

iii. Monthly income of the policyholder

iv. No. of premiums late by 3-6 months

v. No. of premiums late by 6-12 months

vi. No. of premiums late by more than 12 months

vii. Total premiums paid on time:- it gives the total number of premium that has been paid in the specified time, without any delays.

2. Creating the database: 'Creating the database' is the most important step in the working of this particular research. This study aims to predict the premium payment of customers in the near future so that the company has a more focus driven curriculum. Customers who aren't beneficial in long term for the company are marked as defaulters. This study was based on primary information and the data were quantitative. Samples were taken by adopting a convenient sampling technique. Various insurance companies' websites and informative closed-end databases were used as a tool of data collection. Descriptive studies of individuals have been done. Standard factors are taken into consideration for the collection of data. Permission for accessing databases is asked by respective companies, whenever needed. People of every category, age group and sex are included and not any particular kind of group is excluded on any basis. Data, to as much extent, is cleaned and there are not many missing values taken into consideration while doing the research.

3. Data Visualization: The raw data that has been collected needs editing and processing as a lot of errors are produced in the database creation step which needs to be removed so that there is more accuracy to the developed algorithm. In the first step, information has been carefully edited and the incomplete answer has been evaluated according to the research theme. This collected information has been analyzed according to the objectives of the research. Meaningful tables have been generated from the process of data. The data has been analyzed by using both qualitative and quantitative techniques. In this process description, explanation and generalization have been made using statistical tests. Similarly, discourse and statistical analysis have been linked with tables and figures have been used to achieve the objectives. For this study, the descriptive statistics: frequency, correlation, and cross-tabulation were used to meet the research objectives and thereby to answer the defined research.

4. Data Extrapolation: The data thus collected was tabulated, interpreted & analyzed to make the study meaningful. In the present study, percentage, a frequency & cross tabulation method has been used for analysis. For this study, the descriptive statistics: frequency, correlation, and cross-tabulation were used to meet the research objectives and thereby to answer the defined research.

Analysis of Data

1. Explanation of factors and sample data

	0	1	2	3	4	5	6	7	8	9	10
id	104167.0	52420.000	42421.000	31301.000	21543.000	87645.000	78654.0	54231.000	98465.000	110976.000	74653.000
perc_premium_paid_by_cash_credit	1.0	0.111	0.027	0.786	0.098	0.543	1.0	0.465	0.274	0.352	0.047
age_in_days	16436.0	19723.000	16430.000	21354.000	6753.000	17432.000	19853.0	15623.000	18375.000	18436.000	14295.000
income	27290.0	136000.000	114110.000	85790.000	43190.000	147382.000	57463.0	483972.000	74653.000	184922.000	47538.000
Count_3-6_months_late	1.0	0.000	0.000	0.000	2.000	0.000	0.0	5.000	3.000	0.000	3.000
Count_6-12_months_late	0.0	0.000	0.000	0.000	1.000	1.000	0.0	2.000	4.000	0.000	2.000
Count_more_than_12_months_late	1.0	0.000	1.000	0.000	1.000	1.000	0.0	0.000	0.000	1.000	0.000
no_of_premiums_paid	4.0	9.000	13.000	12.000	3.000	7.000	9.0	11.000	13.000	15.000	22.000
target	1.0	1.000	1.000	1.000	0.000	0.000	1.0	0.000	0.000	1.000	0.000

```
1
```

Table 1

Table 1 depicts the sample data of our complete dataset. In the table, we have taken a sample dataset of all the factors on which our research is done. These are the top 6 rows of all the factors. We have considered our target as the required result. When the target is 0 it means the customer will not pay the next premium and when the target is 1, it means the customer is likely to pay his/her next premium.

The thing that is to be noted here we haven't considered the health issues and particular gender of any individual. The complete calculations and observations have been done only on this dataset (complete one).

	count	mean	std	min	25%	50%	75%	max
id	524.0	126669.103053	1.524557e+05	154.0	47892.75000	86655.50	158954.25	1955646.0
perc_premium_paid_by_cash_credit	523.0	0.612570	3.742997e-01	0.0	0.31945	0.66	1.00	1.0
age_in_days	524.0	18465.559160	5.235509e+03	1764.0	14674.25000	17782.50	22678.50	31400.0
income	524.0	914468.049618	1.292250e+06	16541.0	283943.00000	754265.00	954437.00	8745630.0
Count_3-6_months_late	524.0	2.223282	2.166526e+00	0.0	0.00000	2.00	4.00	9.0
Count_6-12_months_late	524.0	1.780534	1.873515e+00	0.0	0.00000	1.00	3.00	8.0
Count_more_than_12_months_late	524.0	0.677481	9.177392e-01	0.0	0.00000	0.00	1.00	4.0
no_of_premiums_paid	524.0	11.337786	4.155489e+00	1.0	9.00000	11.00	14.00	40.0
target	524.0	0.887405	3.163996e-01	0.0	1.00000	1.00	1.00	1.0

```
1
```

Table 2

Table 2 depicts various values as explained. The count is the total number of values i.e. 500 are taken into consideration for the observation. The mean of all the factors is calculated to find some particular average values of all the factors. And the minimum and maximum values exist in the dataset.

Moreover, we also calculated the quartile deviation at 25,50 & 75 percentile. These all values are calculated to find the basic scenario of the data. For instance, we can observe that age of customers here has an average value of 18465.55 i.e. of 50 years. The maximum age of the customer is 86 years. Now based on his age various things will be considered and various observations have been done. Based on all this data, the part of research that this individual is likely to pay the premium or not is being done.

2. Explanation of range of customers who paid the percentage premium amount by cash credit or full payment at once

S. No	% premium paid by customers by cash	Frequency
1	0-10	60
2	10-20	40
3	20-30	25

S. No	% premium paid by customers by cash	Frequency
4	30-40	60
5	40-50	20
6	50-60	12
7	60-70	45
8	70-80	20
9	80-90	28
10	90-100	214

Table 3

This frequency table 3 shows that the maximum number of customers have paid the full amount in cash. More precisely, 214 customers are the one who prefers to pay the 100% amount by cash credit or at once. Even there are 100 customers are from the category who paid less than 20% amount in cash credit. This data concluded that the company have more of the cash flow so that they can pay the amount to another customer if someone claims. Cash flow is manageable by the company. It shows a positive sign for the company's requirements.

3. Explanation of income of customers

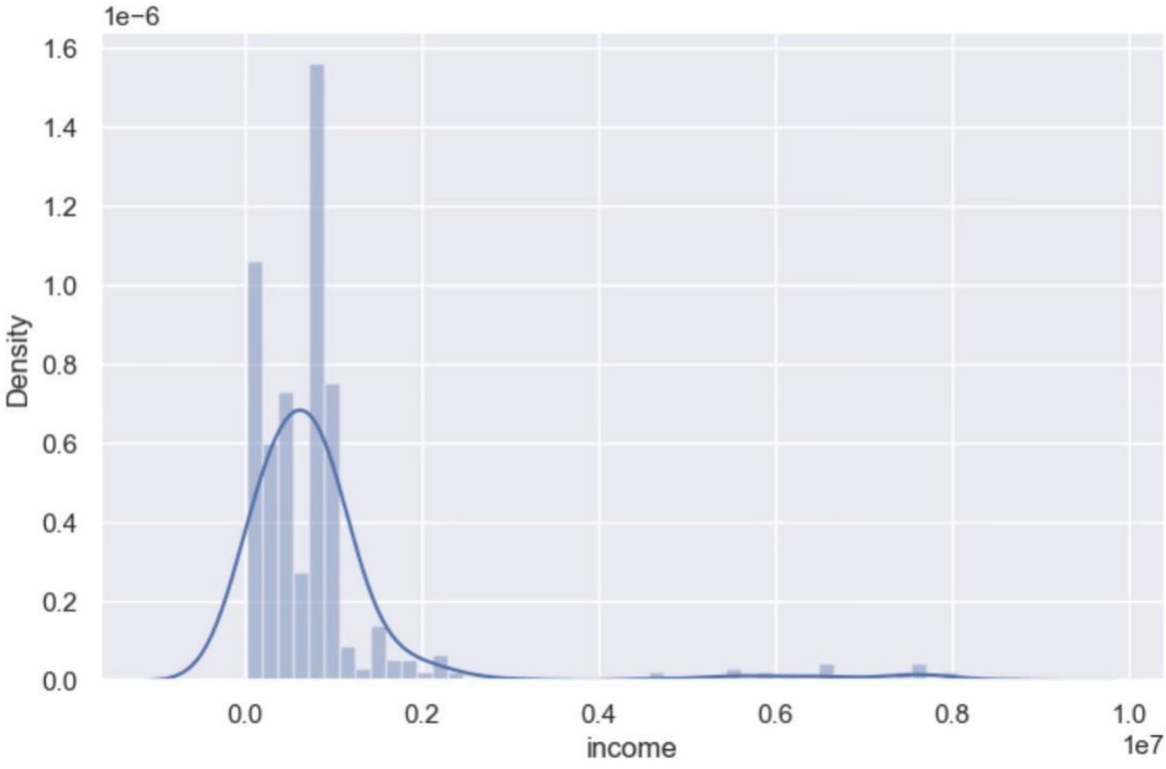

Figure 1

Table 4

```
In [31]:  1 pd.options.display.float_format
          2 df.income.describe(percentiles=[
          3

Out[31]: count       524.0000
         mean     914468.0496
         std     1292249.5587
         min       16541.0000
         0%        16541.0000
         10%      128911.8000
         20%      210355.8000
         30%      409544.1000
         40%      496554.8000
         50%      754265.0000
         60%      788884.6000
         70%      874456.0000
         80%      985641.0000
         90%     1456321.0000
         99%     7538603.7700
         99.9%   8301546.5160
         100%    8745630.0000
         max     8745630.0000
         Name: income, dtype: float64
```

In the above graph i.e. figure 1, we have observed that the income of most of the customers is within the amount of 10 lakh. Some of the customers lie in the category of outliers where it can be observed that the value of these outliers is much higher than the normal range. These outlier values are the reason that some of the values lie in the area of 80 lakh. These uneven values in the income can lead to misinterpretation of the data. Data analysis can't be more accurate with these outlier values. The exact values of these outliers and how they can lead to misinterpretation is shown in table 4.

In table 4, we have observed that in total 524 entries we have taken into consideration the mean is 914468, whereas most of 90% values lie in the range of 10 lakhs. This is because of the outlier values. From the above table, we can observe that 80% of the values lie within the 985641. 90% lies within 15 lakh approx. From the above table, we can conclude that most of the customer's company have had the income within 15 lakhs approximately. The company have to set their policies and target accordingly. It isn't necessary that the customer having a higher income will pay the next premium. Income can be a factor but it isn't the most important factor.

So far, we have considered the age and income of the customers as a crucial factor but it simply doesn't provide us with the required results. Moreover, the percentage of the premium paid by cash by the customers can also show the number of customers who contribute to the cash flow of the company but they will pay their next premium or not is commendable. So, for that, we need some important correlations between the factors so that we can find some important research from these factors. These correlations must be between target values and other factors as the target is the ultimate factor that we need to examine. Dependencies of target values on other factors can conclude our research proficiently. To find these dependencies we prepare a Heatmap of the correlation between other factors with the target values.

Heat Map

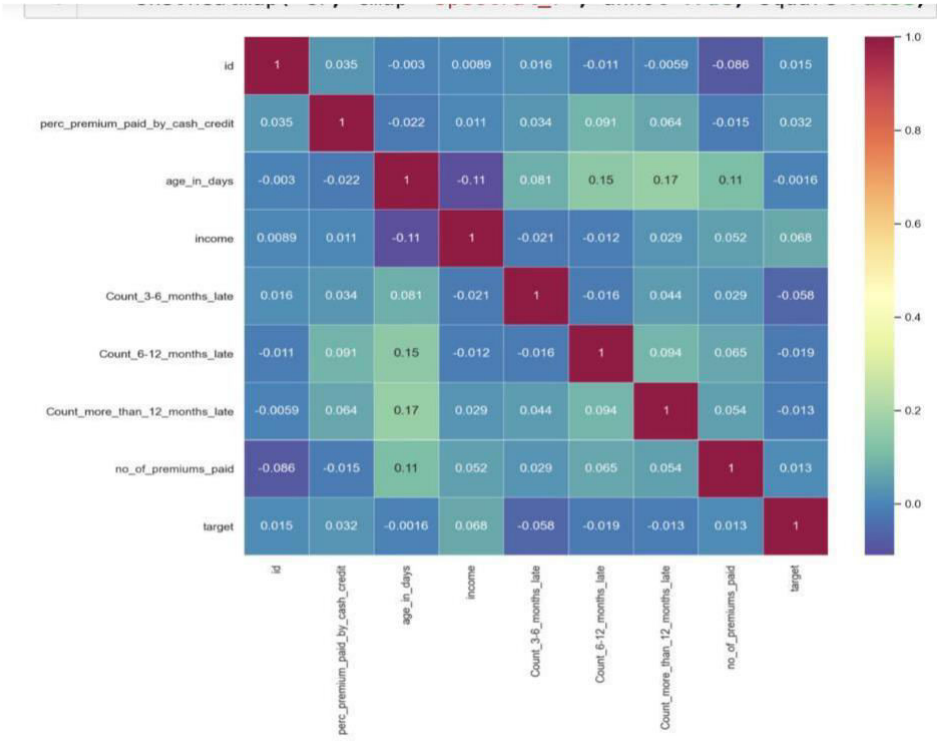

Figure 2

Conclusions

From the heat map we can conclude the following points:

1. The ID of the customers is nearly 0.015 correlated i.e. almost 0. So, customers having different id numbers is not going to affect the target of the company.
2. Age in days is negatively (-0.001) correlated with the target. It implies that there are fewer chances that a customer with higher age will pay the premium on time. But the correlation is close to zero so this factor doesn't impact the company to any significant extent.
3. Income of the customers is positively related w.r.t. target. Customers with higher income have higher chances of paying the premium on time. This correlation is of significant importance to the company due to its larger positive value.
4. Premium paid by cash credit is positively correlated (0.032) w.r.t. target value. It simply implies that the company is not bothered by much extent as this only help them to improve the cash flow for the business.

5. Count 3-6 months is negatively correlated w.r.t. target. This factor plays an important role as customers who haven't paid their premium on time in past have the image that they will not pay the premium on time or not at all in future.

6. In the same manner, count 6-12 months late and count more than 12 months late have negative w.r.t. target values. These factors can give the prediction of customers who will not pay the premium on time in the near future.

7. The values generated here can be used to establish that the past behaviour of a customer is most likely to be repeated only when a combination of factors are considered i.e a customer with low income and a history of defaulting in its payment may have higher chances of repeating the behaviour.

References

1. Yusuf, H. O., Gbadamosi, A. &Hamadu, D. (2009). Attitudes of Nigerians towards insurance services: An empirical study. African Journal of Accounting, Economics, Finance and Banking Research, 4 (4), 334- 346.

2. Viswanadham, P. (2005). Claims settlement operations performance evaluation of LIC of India, The Indian Journal of Commerce, Vol. 58 (2), 80-90.

3. Yadav, R. K. &Mohania, S. (2014). Claim settlement process of life insurance services – A case study of ICICI prudential life insurance Application of Cost Reduction ... 44 company, International Letters of Social and Humanistic Sciences, 24, 26-32.

4. Kalani, M., Salunkhe, H. A., &Ahirrao, M. B. (2013). Comparative study of claim settlement ratio of LIC with other insurance companies in India, Indian Journal of Applied Research, 3 (5), 389-391.

5. Bates, I. & Atkins, B. (2007). Management of Insurance Operations. London, Global Professional Publishing.

6. Brear, S. (2004). Chartered insurance institute (CII) course book, UK, Personal lines Insurance, CII learning solutions, pp 14/9-14/17. Butler, S. & Francis, P. (2010). Cutting the Cost of Insurance Claims, taking control of the process. Booz & Co.

7. Namasivayam, N., S. Ganesan and S. Rajendran (August 2006), "Socioeconomic Factors Influencing the Decision in Taking Life Insurance Policies", Insurance chronicle (The ICFAI University Press), page 65-70

8. Rao, Bn. Vankateswara (June 2006), "LIC- New Business Lacks Vigor", Insurance chronicle (The ICFAI University Press), page33-40

9. Ms Sunayna Khurana, Lecturer, "Customer Preferences in Life Insurance Industry in India", ICFAI National College

10. "WHO South-East Asia Region: India statistics summary (2002 - present)". World Health Organization. Retrieved 13 January 2014.

11. http://www.cppr.in/article/health-insurance-and-telecom-markets-a-comparative-study