Analysis of Customer Behaviour in Automobile Insurance Industry

Smit Jain
x18135340
Analytical CRM
MSc in DA - NCI

***Abstract*— The competition amongst the customers is judged upon the understanding the companies have about their customers. The better the understanding, more are the chances of making bigger business with the customers. The automobile industry in United States of America is a big industry and the insurance companies have a huge database of the records of their customers. The industry has been growing since ever and the need to make more efficient decisions to acquire and maintain more number of customers has been a competition amongst the insurance providers. This project aims to find a relationship in the patterns of the customer behaviour in response to the marketing campaigns in the automobile industry. With the help of various visualizations, the relationships have been analysed and based on this analysis the decisions can be made.**

Keywords - Insurance Industry, Customer Behaviour Analysis, Customer Relationship Management (CRM), Data Visualization, Tableau.

I. INTRODUCTION

The automobile industry in the United States is growing year by year and is earning great profits. The insurance providers have different types of policies like Corporate Auto, Personal Auto, Special Auto. As per a report, biggest automobile industries in United States are The Allstate Corporation, Berkshire Hathaway Inc., State Farm, Progressive Corp. The total revenue of the automobile industry is $285 billion and its annual growth from 2014-2019 has been 3.3% and the industry has employed 279,001 people within the industry. With the raising profit levels, theres another thing that is raising Customer Expectations. Hence to market more efficiently and sell the insurance products more, the insurance companies invest a huge amount to develop a Customer Relationship Management system.

Most of the companies, now-a-days rely heavily upon the CRM system which gives them the potential to attract and acquire new customers which will help the company grow. The companies maintain the customers data which helps them to maintain relationship with the customer and helps the marketing department to improve their goals of customers retention for a long-term [1].

The most important points to be considered for customer retention are commitment and the trust [2], [3], [4], [5]. For a relationship to blossom between two parties, there must be a certain level of trust in between. It is the lookout of the company to maintain this trust with the customers, because once the trust is broken, it is hard to re-build and the relationship ends. Trust sometimes can be seen as the most important investment to build a relationship. This in turn results in a long-term relationship.

Whenever there is an interaction with customers, a relationship is established and the details related to all the transactions of the customer with the company are recorded and periodically managed [6]. The CRM in this area plays the most important role. The companies which have CRM system developed can easily run their business process. All the customer details are stored on the CRM systems and whenever a new transaction is done, all the details for the customer can be seen on the system. Hence, it creates a chance for the customer to purchase the product again in the future by making it easier for the sales agents to market the right product to the right customer. With this, the customers interest, loyalty and trust can be increased and at the same time, make the sales and profits increase.

The organization of the paper is as follows: the next, section 2, states the hypothesis made in the project. After that, a background of the dataset (section 3) is there which is followed by the data pre-processing (section 4). The next section 5 is about the previous researches and work done related to the topic of this paper. Then comes the methodology in section 6, where the methodology used has been discussed in detail. Section 7 contains details about the experiment done, which is followed by the results in section 8. Section 9 is about the conclusion drawn after performing the project.

II. HYPOTHESIS

The null hypothesis is defined as follows:

H0: Customers response pattern is same for all the marketing plans.

By critically analyzing the dataset, this project will test the hypothesis given above by performing experiments and based on the result, the conclusion will be provided.

III. DATASET BACKGROUND

The dataset that has been used in this project has been sourced from IBM Watsons Analytics community. The dataset consists of 9134 rows and 24 columns. It has data about the customers insurance policy, its value, response to the advertising and other details related to the customer which are briefly detailed below:

Attribute	Description
Customer	A unique field having the Customer ID of the customer.
State	The state in which the customer lives. There are 5 states in this column.
Customer Lifetime Value (CLV)	This contains CLV of each customer.
Response	This is a binary field which contains response from the user in either yes or no.
Coverage	This column contains the type of coverage the customer is having under the policy. There are 3 types of coverage : Basic, Extended and Premium.
Education	This field shows the education qualification of the customer.
Effective to date	This contains the date in the format of mm/dd/yy.
Employment status	The employment status of the customer Employed or unemployed.
Gender	Gender of the customer: F or M.
Income	This field contains income of the customer.
Location code	This contains location of the customer: Rural, Suburban or Urban.
Marital Status	The Marital status of the customer: Married, Divorced or single.
Monthly premium auto	This contains the premium paid by the customer each month.
Months since last claim	This contains the number of months since the last time claim was made.
Months since policy inception	Number of months since the policy started for each customer.
Number of open complaints	Number of complaints logged by the customer
Number of policies	This contains the number of policies owned by the customer.
Policy Type	This contains type of policy that the customer has : Corporate, Personal or Special.
Policy	This contains the sub-category of the type of policy that the customer owns.
Renew offer type	This field contains the offer type that was presented to the user for the renewal of the policy : offer1, offer2 or offer3.
Sales channel	This contains the channel through which the customer was contacted for sales: Agent, branch, call center, web.
Total claim amount	Total amount of claim made by the customer.
Vehicle class	This contains the class of the car that the customer has: Four door car, Two door car, luxury car, luxury SUV, Sports car, SUV.
Vehicle size	The size of the vehicle Large, Medium, Small.

TABLE I: Dataset Background

IV. Data Pre-processing

When the dataset is extracted or downloaded from any external source, the data is in its raw format. There are possibilities of having missing values within the dataset. So, for dealing with the irregularities in the dataset the data pre-processing is done. The dataset that has been taken for this project is almost in its cleanest form. There are no missing values in the entire dataset.

V. Literature review

The topic chosen for this project is about the automobile insurance, hence it is necessary to have a brief idea about the insurance industry and how CRM works in that sector. There are a lot of researchers who have done researches in the insurance related field. The major issue that the insurance industry is facing right now is obtaining the clients and then keeping record of them in an organized and managed way.
The companies who are into financial sectors make use of the CRM systems, processes and the strategies and are striving hard to increase the total value/profit they make per customer [7]. The companies invest heavily in strategy planning and in return they get better financial outputs.
Major countries in the world are having a tight competition amongst the insurance industry and companies follow one mantra to success managing the portfolios. These companies have portfolios of all the customers, but the pay major attention on the most profitable customers [8]. The most profitable customers are the ones who have trust on the companies and stick to one company and never fall in trap of other marketing schemes by other companies.
Most of the companies strive to gain trust of the customers. Once the trust of the customer is acquired, the companies really dont have to worry about the competitive brands, the prices and the advertising of the product at higher levels. To gain the trust, the customer has to be satisfied with the product or the services that the company is providing [9].
There are 3 types of distinct categories in the consumers, as per the study done [10], and those are the upper class, the middle class and the lower class. The customers who are giving the highest profits to the the business and the organization fall under this category (top 10%). Hence, to retain those customers, they have to be provided with the best of the services and offers so that they stay loyal and never go to the competitors. The other group is made up of customers who undoubtedly provide a big share of profit to the organization (40-50 %) but are also at a higher risk of working with the competitor companies. Hence CRM has to identify the needs of the customers correctly. The last, lower class (40-50 %), group is very slow growing in terms of business profit and is beneficial for the company to recognize this chunk and put the least efforts on them. This divides the total customer base in to 3 parts, first the most profitable ones, second the profitable ones yet they do explore new offers and the third the least profitable ones.
In the article [7], the author has discussed about the retaining of the customers and has ascertained that having satisfied customers is having advantage more than acquiring new

customers. The acquiring of new customers is a very big task and involves a lot of investment for a company. Whereas, the efforts spent for retaining the old customers takes much lesser efforts and much more profitable.
This tells us about the loyalty of the customer. Establishing and maintaining long term relationships with the customers is the motto of the relationship marketing [11]. The significance of the retention, CLV and the relationship between the customer and the organization has been emphasized in the work done by [12].// The customer satisfaction plays a major role in defining the success of any organization. It is because once the customer gets satisfied, it is helpful for the organization to understand the customer better and they can create better CLVs. The improved satisfaction also promises the intention of re-purchase by the customer to the organization which obviously influences the purchasing behaviour of the customer and in return shows its impact on the performance of the organization [6]. All of the above researches motivate to do an analysis on the behaviour of the customers, how they respond to marketing campaigns and their buying behaviours.

VI. Methodology

This project aims to find out the behavior of the customer and try to find out a pattern in the response to the marketing campaigns held by the insurance company. There are 24 columns in the data set chosen for this project and a correlation between all the attributes has been checked. To check the correlation between the variables (columns) in this dataset, SPSS has been used which clearly outlines the correlation matrix and also it is easy to understand. Since the number of records in this dataset are more than 9000, it is best to use data visualization techniques to visualize the trends and patterns in the dataset. For visualizing the data, Tableau has been used which is an effective way of visualizing the data and is easy for the readers to understand.

VII. Experiment

The dataset chosen for this project is having a large amount of data, that is why data visualization technique has been chosen and beautiful and meaningful insights have been developed with the help of visualizations to help the business make decisions that can help the sales grow in figures. The dataset consists of the location details of each and every customer. Using this, the company can figure out which area is having maximum number of its customer and based on that marketing campaign decisions can be made. From the visualization below 1, it can be seen that the maximum number of customers are from California.

The bubble chart below 2 shows the percentage of the customers who respond to the marketing campaigns. It signifies whether or not the customers have responded to the efforts spent for marketing.

Now, to drill down further into the response to the marketing campaigns, the offer that was given to the

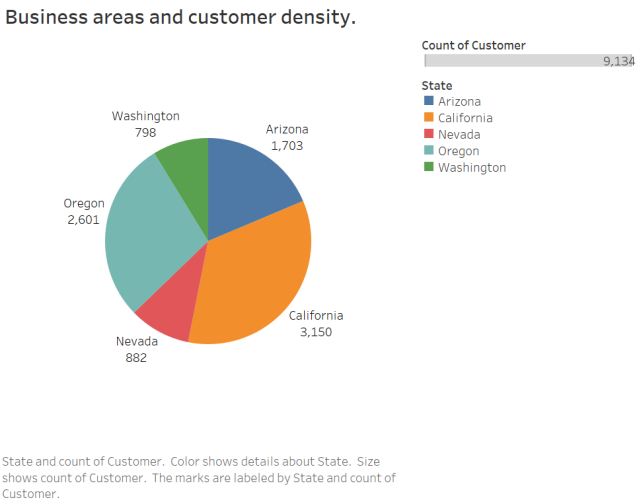

Fig. 1: Business areas and customer density

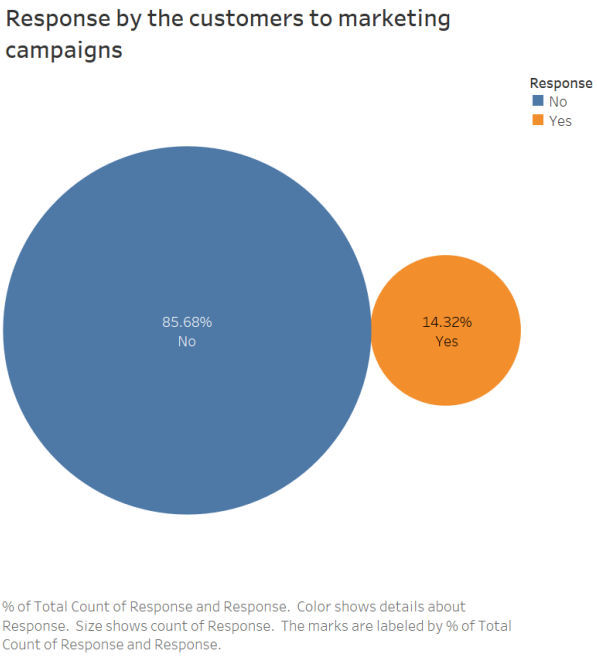

Fig. 2: Customer's response

customers can be traced, which will further tell about why the customer responded yes or no. Based on this, the customers can be categorized and the marketing actions can be amended for future. The visualization below 3 shows about what type of offers resulted more into a negative response.

The next visualization shows the percentage of the customers belonging to each class categorized by the type of offer. This will help the business in segmenting the customers based on the type of offer. The previous visualization has one more critical information, that is, maximum number of positive responses are from offer 1 and offer 2. Here with the help of next visualization 4, we can drill down to the vehicle class, and it is evitable that the customers with four door vehicles were more responsive as compared to others. If

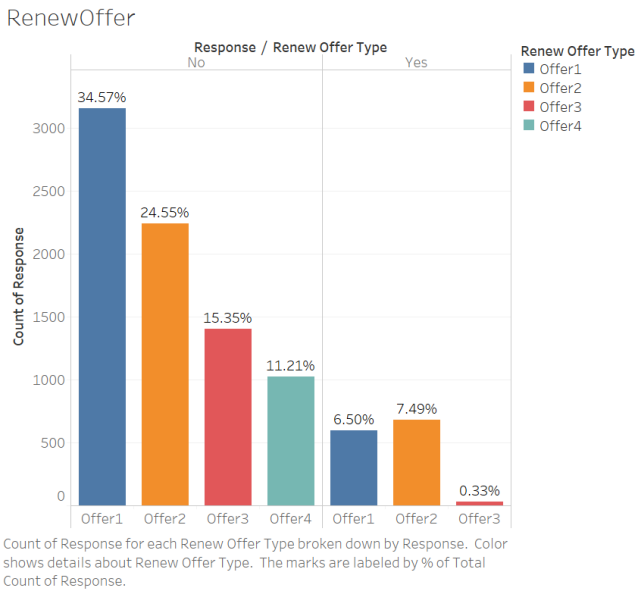

Fig. 3: Renew Offer type

the difference in the response by the customers is significant, the business can amend and make changes in the offers to suit the customers.

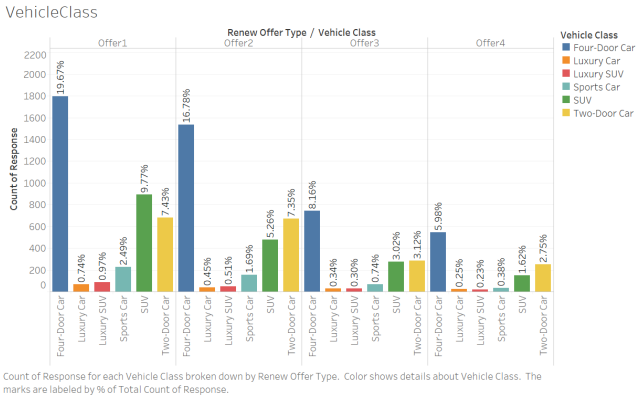

Fig. 4: Vehicle Class

The next visualization 5 talks about the channel using which the customer is communicated with. It shows that the most number of responses are from the agents. This will give the company a brief idea about where to channelize the strength and also to train the staff for a better result.

The visualization below 6 shows that the customers who are having medium sized vehicles have responded the best as compared to others. The agents have got the maximum responses in comparison to all other sales channel and of those responses, maximum are from the medium size vehicle owners.

The next visualization 7 shows that the maximum number of responses are from the employed people. This will help the business shape the future marketing campaigns which

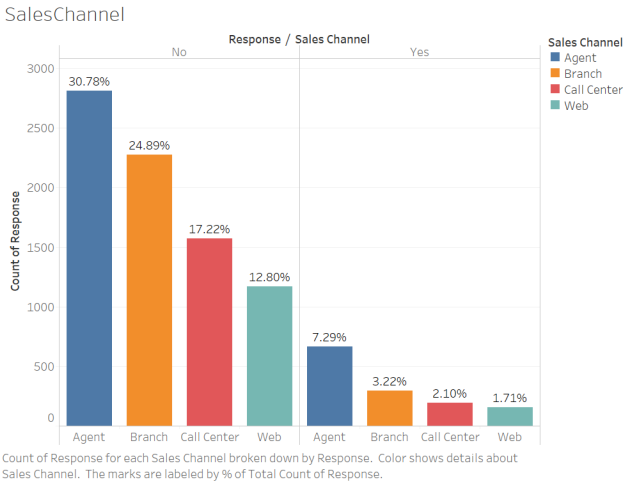

Fig. 5: Communication channel

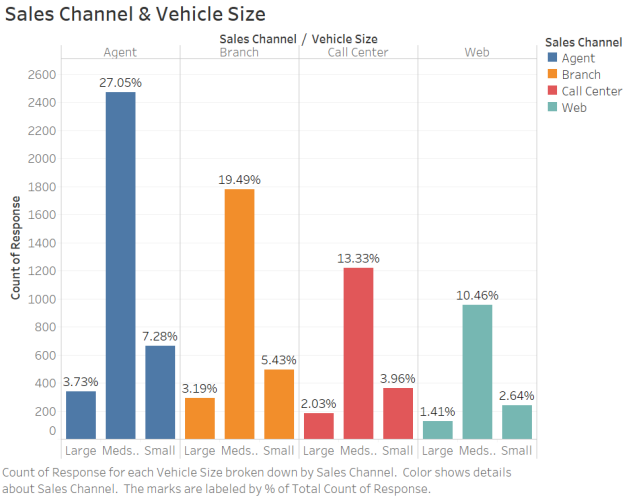

Fig. 6: Vehicle vs Channel

will influence more sales.

The visualization shown 8 below shows the number of responses belonging to each policy type and policy. It is clear that most number of responses are from people who have personal auto policy type and the looking deeper in to it, the policy personal L3 received most number of responses.

The next visualization 9 talks about the employment status and the coverage provided to the customer. This tells what type of coverage is being opted more by the people with different employment status. This can help the business make important decisions like which customer has to be targeted for specific coverage type.

Another interesting fact that can be captured from the visualization below 10 is that the maximum number of people who responded are having four door car are having personal auto policy type and the specific policy is Personal L3.

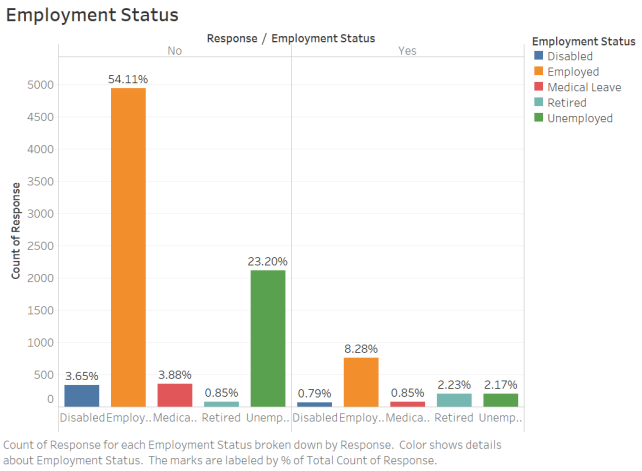

Fig. 7: Employment status

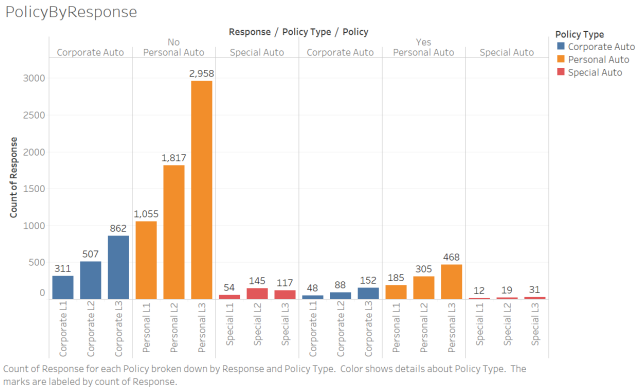

Fig. 8: Policy by response

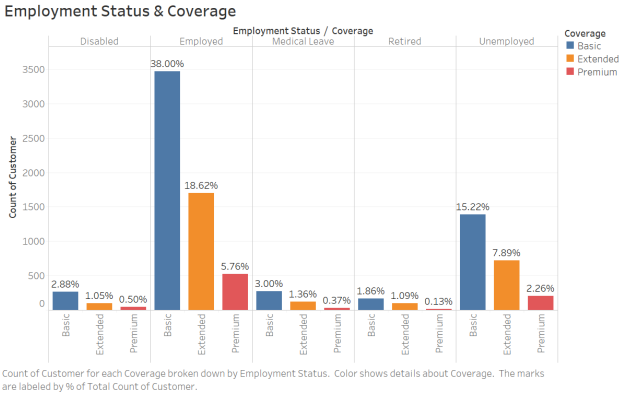

Fig. 9: Coverage type

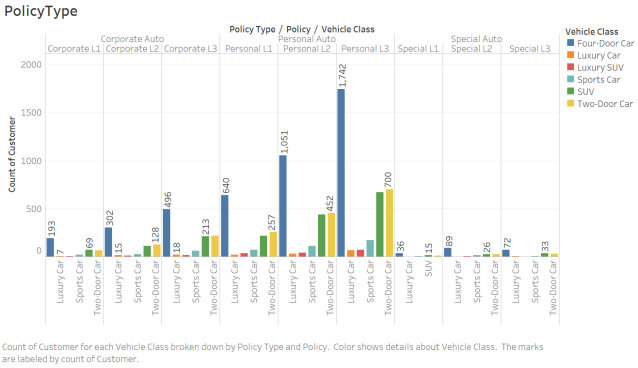

Fig. 10: Policy type

The correlations matrix is described below 11. It tells us about how much the variables are correlated to each other.

VIII. Result

As seen in the visualizations done in the above section, it is clear that the null hypothesis is rejected. The behavior of the customers varies across all types of the policies. The most critical and business impacting observations that were made in the visualizations above are as follows. The customers who were given the offer 1 for the renewal came with the highest number of negative responses (34.5% of the total responses). But in the same visualization, of all the positive responses, the offer 2 for the renewal got maximum positive responses and is better than other two offers. Another visualization tells about the response by the class of the cars owned by the customers. Most of the customers who are having four door cars have responded the most in all types of renewal offers. The next visualization is about the sales channel that is used to communicate with the customers. It is quite evident from the visualization that the agents got the maximum number of responses, followed by the branches, call center and the websites. This can make the company make decisions where the customers are coming from, which channel is having maximum number of customers and then they can improve the sales procedures by introducing some experimental changes. The employment status is another factor that has driven the responses from customers. As it is depicted in the visualization, the most number of responses are from the employed customers. In the visualizations above it can be clearly seen that the responses from all the customers are varied and it depends on the class of the vehicle they use, the policy that they have taken and the type of coverage. The correlation matrix on the other hand shows that the variables are not highly correlated. Hence, with the help of visualizations the company can make effective decisions that can help generate more sales and leads and in return earn greater profits.

IX. Result

This project aims to analyze the relationships and patterns of the response by the customers in the automobile insurance industry. To conclude, after carefully analyzing the visualizations made from the customers response, the response of the customers is different based on the type of policy and the class of the car. The visualizations provided can help the business shape their policies better and make the marketing campaigns aligned with the business goals. In this way, the organizations can make better offers and make better relationships with the customers.

Correlations

		Customer Lifetime Value	Response	Income	Monthly Premium Auto	Months Since Last Claim	Months Since Policy Inception	Number of Open Complaints	Number of Policies	Total Claim Amount
Customer Lifetime Value	Pearson Correlation	1	-.009	.024*	.396**	.012	.009	-.036**	.022*	.226**
	Sig. (2-tailed)		.393	.020	.000	.271	.368	.001	.036	.000
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Response	Pearson Correlation	-.009	1	.012	.011	-.017	.003	-.010	-.021*	.017
	Sig. (2-tailed)	.393		.254	.295	.113	.778	.345	.046	.107
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Income	Pearson Correlation	.024*	.012	1	-.017	-.027*	-.001	.006	-.009	-.355**
	Sig. (2-tailed)	.020	.254		.111	.011	.933	.540	.408	.000
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Monthly Premium Auto	Pearson Correlation	.396**	.011	-.017	1	.005	.020	-.013	-.011	.632**
	Sig. (2-tailed)	.000	.295	.111		.631	.053	.210	.283	.000
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Months Since Last Claim	Pearson Correlation	.012	-.017	-.027*	.005	1	-.043**	.005	.009	.008
	Sig. (2-tailed)	.271	.113	.011	.631		.000	.609	.383	.470
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Months Since Policy Inception	Pearson Correlation	.009	.003	-.001	.020	-.043**	1	-.001	-.013	.003
	Sig. (2-tailed)	.368	.778	.933	.053	.000		.912	.203	.750
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Number of Open Complaints	Pearson Correlation	-.036**	-.010	.006	-.013	.005	-.001	1	.001	-.014
	Sig. (2-tailed)	.001	.345	.540	.210	.609	.912		.886	.174
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Number of Policies	Pearson Correlation	.022*	-.021*	-.009	-.011	.009	-.013	.001	1	-.002
	Sig. (2-tailed)	.036	.046	.408	.283	.383	.203	.886		.822
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134
Total Claim Amount	Pearson Correlation	.226**	.017	-.355**	.632**	.008	.003	-.014	-.002	1
	Sig. (2-tailed)	.000	.107	.000	.000	.470	.750	.174	.822	
	N	9134	9134	9134	9134	9134	9134	9134	9134	9134

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Fig. 11: Correlation matrix

REFERENCES

[1] R. Jain and S. Bagdare, "Determinants of customer experience in new format retail stores.," *Journal of Marketing Communication*, vol. 5, no. 2, pp. 34 – 44, 2009.

[2] F. R. Dwyer, P. H. Schurr, and S. Oh, "Developing buyer-seller relationships," *Journal of Marketing*, vol. 51, no. 2, pp. 11–27, 1987.

[3] G. T. Gundlach, R. S. Achrol, and J. T. Mentzer, "The structure of commitment in exchange.," *Journal of Marketing*, vol. 59, no. 1, p. 78, 1995.

[4] R. M. Morgan and S. D. Hunt, "The commitment-trust theory of relationship marketing.," *Journal of Marketing*, vol. 58, no. 3, p. 20, 1994.

[5] D. T. Wilson, "An integrated model of buyer-seller relationships," *Journal of the Academy of Marketing Science*, vol. 23, no. 4, pp. 335–345, 1995.

[6] M. Alhabeeb, "On consumer trust and product loyalty," *International Journal of Consumer Studies*, vol. 31, no. 6, pp. 609–612.

[7] M. Stone and B. Foss, *CRM in financial services : a practical guide to making customer relationship management work.* London : Kogan Page, 2002., 2002.

[8] M. B. D Biswamohan, "E-crm practices and customer satisfaction in insurance sector," *Research Journal of Management Sciences*, vol. 1, no. 1, pp. 2–6, 2012.

[9] M. Guilln, J. P. Nielsen, T. H. Scheike, and A. M. Prez-Marn, "Time-varying effects in the analysis of customer loyalty: A case study in insurance.," *Expert Systems With Applications*, vol. 39, no. 3, pp. 3551 – 3558, 2012.

[10] M. Lloyd Susan, "Loyalty.comcolon; customer relationship management in the new era of internet marketing.," *Journal of Consumer Marketing*, no. 4, p. 368, 2001.

[11] C. K. Anu Helkkula, "Circularity of customer service experience and customer perceived value," *Journal of Customer Behaviour*, vol. 9, no. 1, pp. 37–53, 2010.

[12] G. Botha and A. van Rensburg, "Proposed business process improvement model with integrated customer experience management," *South African Journal of Industrial Engineering*, vol. 21, pp. 45 – 58, 00 2010.

X. APPENDIX

The time log for the project is given in the table II which shows the efforts spent week-wise and the tasks performed in the respective weeks.

Time	Task
Week 1 to Week 3	Topic and data set hunt.
Week 4 to Week 5	Topic finalization and project proposal submission.
Week 5 to Week 7	Found technical research papers.
Week 8 to Week 9	Data Analysis and Visualization.
Week 10	Report making and converting into Latex format.

TABLE II: Project log

XI. ACKNOWLEDGEMENTS

This project has been done under the guidance of Vikas Sahni and with the help of Nilay, who helped me shape the idea.

XII. VISUALIZATIONS FILE

The tableau sheets for this project can be found on the github repository (https://github.com/smitj92/CRMproject).