A Machine learning approach in the analysis of Texas consumer complaints on insurance companies

Lenin Kamma

Bellevue University

Professor Catie Williams

September 20, 2020

# Abstract

# Texas Department of Insurance (TDI) accepts written complaints against insurance companies, health maintenance organizations (HMOs), insurance agents or agencies, and other persons or entities regulated by TDI. Complaints generally involve such matters as claims and benefits, false advertising, misrepresentation of policies, and HMO quality of care1. In this paper, the author investigates the complaints from the TDI website, filed by consumers in the state of Texas to identify the reason and the type of complaints, research the insurance companies involved, and identify the resolution time for each type of complaint. This study also highlights the insurance problems that consumers are facing and predicts the dispute resolution period for different types of complaints. Along with this, it investigates complaints in different zones of Texas to predict patterns in the dispute resolution.

*Keywords*: Texas Insurance Complaints, Consumer, Insurance

# Background

Customer insurance complaints and the administration of complaints resolution are very important for businesses, especially in the insurance industry where the long-term relationship with the customers is of paramount importance. There are several federal agencies that collect consumer complaints and use that information to monitor the industry they regulate4. The Texas Department of Insurance is the official state agency charged with regulating the insurance industry in Texas. This department also keeps an eye out on insurance scams and deceptive insurance practices2 3. In order to retain customers and avoid litigation, the insurance industry must act on each complaint filed by consumers diligently so that the resolution and agreement can be reached with the consumers as soon as possible. However, it is observed that insurance companies often delay the process of resolution and take 4-6 months to resolve a dispute. It is also observed that complaints of the same nature are filed against certain companies because of their bad practices towards claims. This study is to focus on the complete view on consumer complaints and provide some important data points to improve the complaint resolution strategy.

# Data

The data used for this project is extracted from the insurance complaints information system stored and maintained by the Texas Department of Insurance (TDI). As required by law, TDI stores consumer complaints data and facilitates the resolution of complaints. These complaints are submitted by consumers to report unresolved claims or bad practices from insurance agents or agencies. This project uses the data from the years 2018 and 2019. The total number of data records extracted for the year 2019 was 6240 and for 2018 was 12406.

# Data Features

The datasets extracted for this study have 11 different data elements which are given below.

**Complaint number**: This is a unique number identifying a specific complaint owner:

**Received date**: The date that TDI received the complaint from the consumer

**Closed date**: The date the complaint was closed

**Correspondent location**: The correspondent location is the region of Texas

**Subject of complaint**: The description of the person/entity against whom the complaint is filed.

**Subject name**: The name of the person or entity against whom the complaint was filed.

**Subject ID**: A unique number assigned to each subject person or entity.

**Line of coverage**: The type of coverage about which the correspondent complains.

**Reason for complaint**: The description of the reason the correspondent is complaining

**Disposition of complaint**: The resolution of the complaint by TDI

**Type of complaint:** The ICIS summaries and download files include confirmed complaints

Data Analysis and Methods

The Author uses various exploratory data analysis techniques to identify the distribution of each data feature. After applying factor analysis, it is determined that the subject ID, complaint number, received date or closed date were of not much importance to this study. However, the resolution time period, which is the difference between the received date and closed date, is a very important feature to be used for prediction purpose. As part of the initial data cleanup, many duplicate entries are removed from the dataset and date fields were converted into the appropriate month and year values. One hot encoding is done to convert the categorical features into numeric values. Various histograms, scatter plots and bubble charts were created to find the relation between the type of complaints, zones and the subject insurance company. Aggregate resolution times were prepared for companies where the number of complaints were the highest. K-means clustering was used to categorize the different clusters of the complaints. Multiple linear regression model was used to predict the resolution time based on the type of complaint, line of coverage, insurance company and geographical location.

Questions

This study answers the following research questions.

* What type of complaints that consumers usually file?
* Which insurance companies are targeted in these complaints?
* What the average resolution period?
* Is there any correlation between the state zone and complaint type?
* Can we identify the practices of insurance companies based on the cluster of complaints?
* Is there any correlation between the company and the number of complaints?
* Which line of coverage is subjected to the most complaints?
* Can we classify complaints based on the reason for complaint?
* Are there any difference between zones in the number of complaints?
* Any correlation between the type of complaint and the resolution period?

Results

* As shown in Figure 1, there were 1831 consumer complaints reported in the Year 2019 compared to the complaints received in the Year 2018, which was at 3881. The month by month numbers have been on the decline since 2018.
* Blue Cross Blue Shield (BCBS) of Texas has received the highest (595) number of complaints followed by the United HealthCare Insurance Company (490) (Figure.2)
* There are 13 different insurance companies who have received more than 100 complaints in two years (Figure 2)
* Across all regions of Texas, Accident and Health category received the highest number of complaints (2359) followed by Automobile insurance (2091). (Figure 4)
* Unsatisfactory settlement or offer from the insurance companies is the main reason for many complaints across Texas. Delay in claim handling is the second topmost reason for consumers to file a complaint (Figure 3)
* While consumers in South East Texas has filed more complaints (around 32%) compared to any other zone in Texas, consumers in West Texas filed a fewer complaints(1.14%) on automobile insurance (Figure 5)
* It is surprising to see the same average resolution time(190 days) for major insurance companies like Blue Cross Blue Shield of Texas and United Health Care

Conclusions

Understanding the nature of complaints can help the insurance companies improve their products and services. However, analyzing a huge number of complaints is not an easy task for human11 McKinsey’s global research across industries shows that improving the customer experience can do far more to drive profitable growth than raising advertising spending or lowering prices10. This paper provides a computational approach to analyze and visualize the complaints filed by consumers across the state of Texas. The results indicate that the consumer complaints are declining month by month and year by year, which is a good sign that indicates that insurance companies are working on resolving the complaints. Two of the major insurers, BCBS of Texas and United Health Care, received more complaints than any other companies. This may be due to the fact that they handle a far greater amount of insurance policies than other companies. It is observed that Automobile and Accident Health are two major categories in which customers are not satisfied with their insurance companies. The same relationship is found in complaints filed in Oregon State7 and California State9 and it is also evident from the fact that those two categories have a major share of the insurance complaints across the different zones of Texas. One interesting observation is that consumers in North Central Texas and North East Texas filed more Auto and Home complaints compared to Accident and Health complaints. The Insurance companies has to look into this statistic to discover why auto complaints are higher in these zones. West Texas reported the least number of complaints than any other zones in Texas, which may indicate a higher satisfaction rate in that zone. The average resolution period is found to be at 178 days. However, one important observation for both the BCBS of Texas and United Health Care is that they both have average resolution time of 190 days. Along with this, the maximum resolution period is found to be at 600 days for both companies. There may be some regulatory deadlines which may have caused this correlation in the number of days. Also, based on the linear regression, it is observed that none of the input fields are correlated to resolution period. So the insurance company seems to have no bearing on the complaint resolution period. Another observation is that there is no correlation found between the Texas zones where the complaints were filed and the resolution period. The data indicates that insurance companies have to improve their customer service to reduce the number of complaints. If insurance companies pay more attention on the settlement process, they would save time and money.

Limitations:

The author of this study retrieved the data from the TDI website and it was observed that the last quarter of 2019 had a lower number of records compared to the last quarter of 2018. This study did not make any correlation between the number people insured and the number of complained in a particular zone. This research was done only based on the data available in the Texas Department of Insurance website.

Acknowledgements

I would like to thank Professor Catie Williams for her help in reviewing this paper and thank all my classmates for their reviews.

# References

# 1 Internet Complaints Information System, https://www.tdi.texas.gov/consumer/icis/index.html

2 Insurance. File a consumer complaint online. https://www.texasattorneygeneral.gov/consumer-protection/financial-and-insurance-scams/insurance

3 Filing a consumer complaint. <https://www.usa.gov/consumer-complaints>

4 Consumer Complaints.<https://consumerfed.org/consumer-complaints/>

5 Market analysis handbook. 2000. https://www.naic.org/documents/prod\_serv\_marketreg\_mah\_hb.pdf

6 Consumer complaint database. https://www.consumerfinance.gov/data-research/consumer-complaints/

7 Oregon insurance complaints. 2014. https://dfr.oregon.gov/help/Documents/2311-14.pdf

8 Background on insurance fraud. 2020. https://www.iii.org/article/background-on-insurance-fraud

9 Consumer complaint study. <http://www.insurance.ca.gov/01-consumers/120-company/03-concmplt/>

10 The growth engine: Superior customer experience in insurance. https://www.mckinsey.com/industries/financial-services/our-insights/the-growth-engine-superior-customer-experience-in-insurance#

11 Computational analysis of insurance companies: GEICO case study. http://sbp-brims.org/2018/proceedings/papers/latebreaking\_papers/LB\_7.pdf

Appendix (Figures)

The appendix has various graphs as listed below

Figure 1: Texas consumer complaints on Insurance

Figure 2: Top 25 insurance companies with the highest consumer complaints in Texas

Figure 3: Complaint reason vs Count

Figure 4: Different types of consumer complaints in Texas

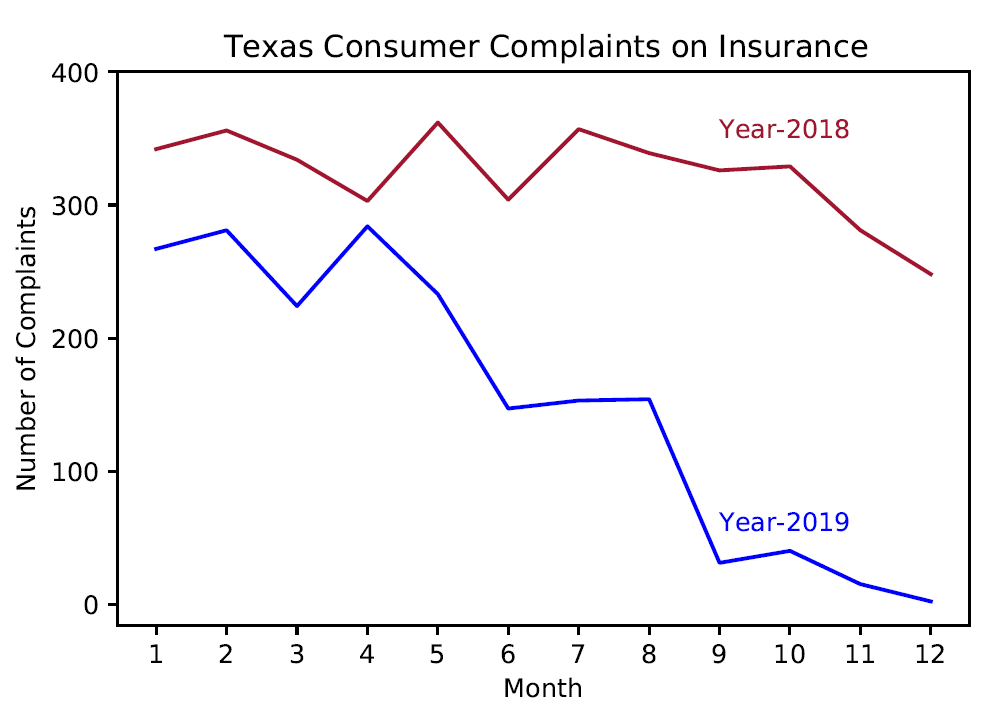
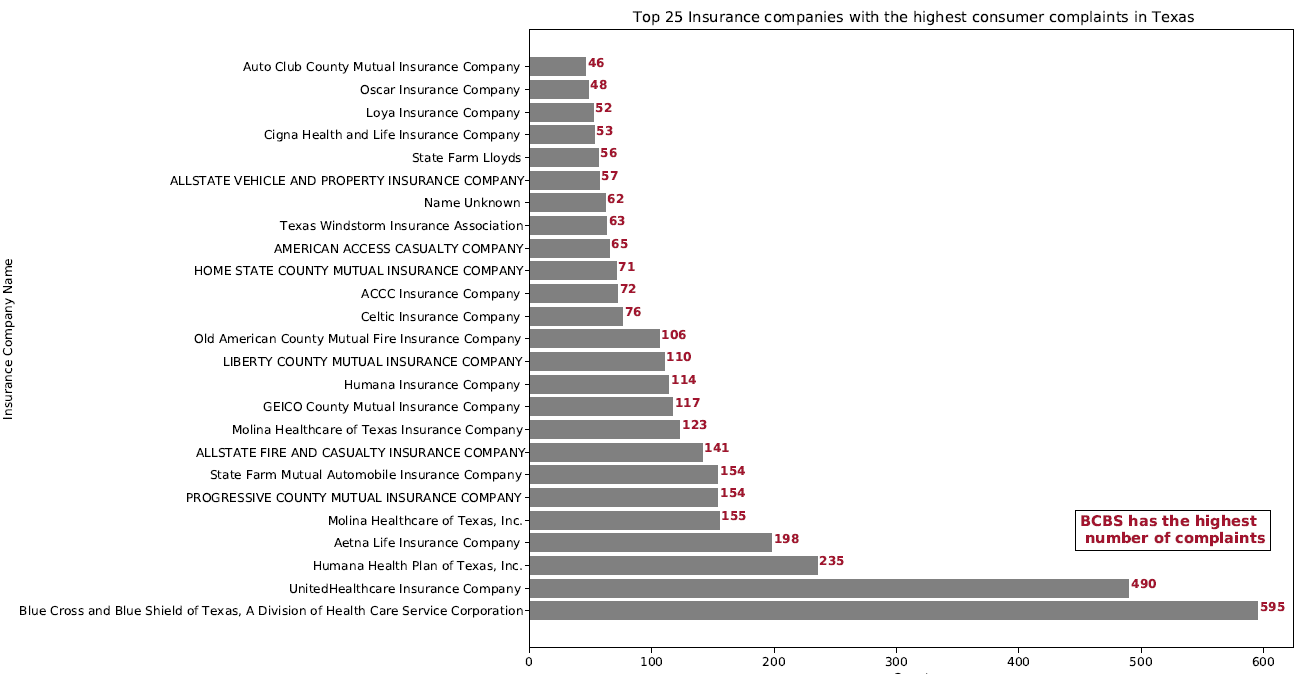
Figure 5: Percentage of Consumer Complaints vs. ZonesFigure-1

Figure 2

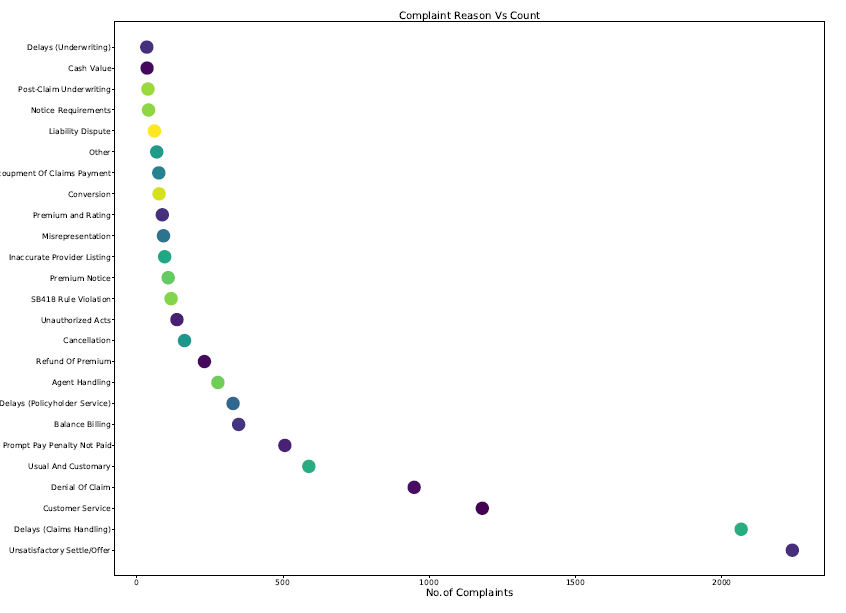


Figure 3

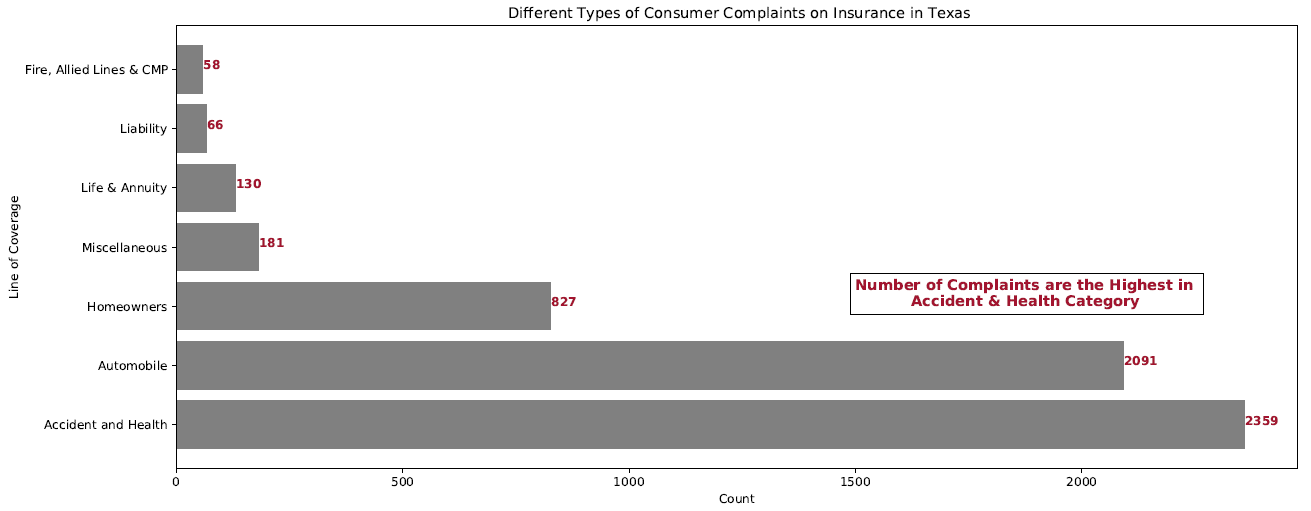


Figure 4

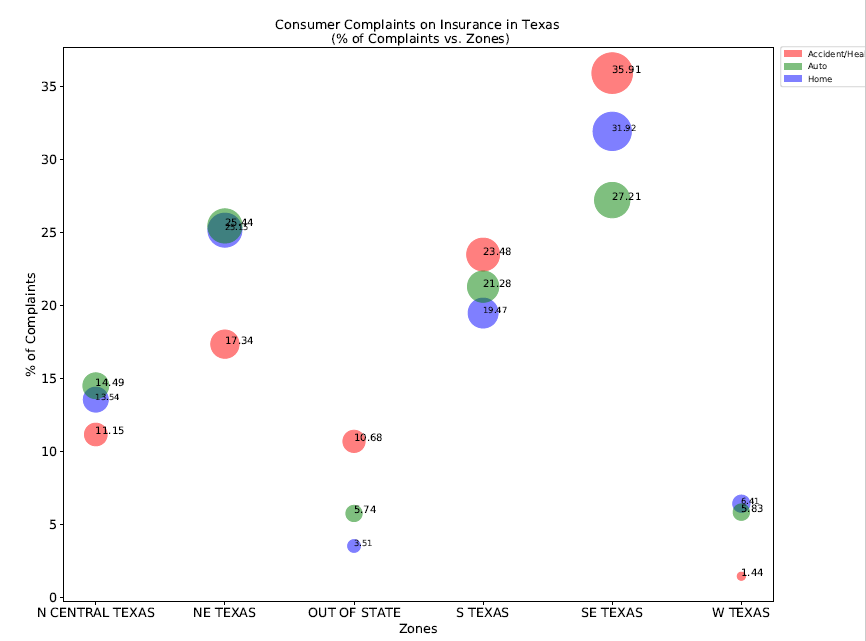


Figure 5