Analysis Report & Sources of Additional Data Used

Lopez Atuesta Juan, Karani Mercy, Mudliar Rahul, Khalil Muhammad Faizan

Illinois State University

Author Note

Department of Technology, Illinois State University

Correspondence concerning this report should be addressed to, Department of Technology, Illinois State University

**Analysis Report & Sources of Additional Data Used**

The data (Insurance, 2023) utilized in the preparation of this report includes 40,877 commercial liability closed claims submitted for calendar years 2007-2012. These claim reports account for $7.7 billion in paid settlements. As complement it was considered and additional data source (U.S. Census Bureau, 2020). The total settlement amount might be incomplete according to the Texas Department of Insurance. Trifacta software was used for the process of cleaning, structuring, and enriching the data into a desired format. Tableau software was used as our visual analytics platform to create a cohesive data flow from information to clear insights.

**Story Point 1**

The graph in Story Point One presents an intriguing visual representation of the insurance claim and settlement data over the years. Sum of Final demand of Plaintiff side on the y-axis, while the x-axis represents the years classified into quarters for each year. Overall, the graph in Story Point One provides a clear and compelling visualization of the insurance claim and settlement data over the years. The use of a line chart and quarters allows for easy comparison and analysis of the data, and the steady increase in the final demand of plaintiff.

**Story Point 2**

The graph in Story Point Two presents the settlement amount data over the years, separated by quarters. Settlement amount refers to the actual amount paid out to the plaintiff as a result of a claim, which can differ from the amount initially demanded by the plaintiff.

The graph also allows us to compare settlement amounts between different quarters and years, which can help us identify any significant changes or trends. For example, we can see that the settlement amounts for Q2 of 2017 are significantly higher than in other years, indicating that this quarter may have had more severe claims or a higher number of claims filed.

Overall, the graph in Story Point Two provides an insightful view of the settlement amount data over the years, allowing us to identify trends and patterns in settlement amounts between different quarters and years. The presence of variance between the initial demand and settlement amount highlights the variance between them.

Top of Form

**Story Point 3**

Bottom of Form

Story point three aims to calm the audience and draw them into the presentation.

**Story Point 4**

From story point 3 we identified that the right thing to do is subtract the amount demanded by the plaintiff by the settlement amount in the year 2007 to 20012.

Story point 4 is broken into a bar graph which indicates the difference between the amount demanded by the plaintiff and the settlement amount in the year 2007 to 2012 which is called the variance.

In the X-axis, we have the variance, and in the Y-axis the year the date was claimed divided into quarters.

The highlight of this graph is as follows.

Trend analysis: The chart intends to identify the variance of the data over time. The variance is changing by a big margin over time.

Transition: The next Story point intends to explain in detail if the variance has any impact on the business class and the counties.

**Story Point 5**

Story point 5 is divided into two maps, the first map indicates the claims count cases, and the second map indicates the variance with the additional data of the U.S. Census Bureau, 2020. Using this data, we focused on the business claims to identify the businesses that were most affected.

Trend analysis: The map intends to identify the impact that the variance has on the business classes and the counties with the highest variance with an analysis of whether the population was the cause of the high variance.

Comparison: In the first map counties such as El Paso, Hidalgo, Bexar, and Travis have low populations but have a big difference between the claim counts and variance. There is also an indication that industries such as agriculture, apartments, townhomes, and condos as well as construction firms are the most affected.

This leads to the conclusion that the population in the counties is not responsible for the high variance shown.

**Story Point 6**

Transition: We saw previously in the two maps the count of cases per counties and variances per counties, we also observed that counties with higher population didn’t particularly have either higher number of cases or high variance. Now we try to find how these counties fair when we put both the things against each other, meaning we will do a new calculation by dividing variance with count of cases. This calculation we decided to name “Variable”. Now when we display variable over the map with respect to divisional counties, we get the slide number 6. In this graph we can see that the variable for each county don’t have any difference when we try to observe with just simply viewing it. However, this doesn’t mean that there is nothing wrong with the cases it just means that with respect to counties there is no significant difference. Hence, we can say that counties have no particular impact. Further to know more and dig deep we will take few counties as an example and evaluate the variable for them happening over a period.

**Story Point 7**

Transition: In the below two charts we have taken four counties to test our hypothesis they are Harris, Dallas, Jeff Davis, and Zapata. The first chart shows us the variable for these counties and the count of cases for these counties but the main thing to notice here is that how Harris and Dallas have a positive variable as compared to Jeff Davis and Zapata who have a negative variable meaning Jeff Davis and Zapata are losing money over those number of cases. The money lost by Jeff Davis and Zapata is very big amount being $704,417 and 3,119,600 respectively. The main observation and the reason why they are losing money can be seen in the second chart, the reason being the amount of time spent over lawsuits in the court. As the time spend by Harris and Dallas per case is approximately 13 and 14 months respectively, whereas for Jeff Davis and Zapata it is 20 and 24 months. It can be assumed that clearly the longer a case lasts the more money will be spent by both parties to keep up. It may be in form of paperwork, lawyer fees etc. Now that we have established the relationship, we need to look how much of an impact does this have for all the counties all together over a period of time, which we will see in next slide.

**Story Point 8**

Story point 8 is broken into two graphs. The graph on top displays the Running Sum of the number of Cases Closed from 2007 to 2012 on the Y AXIS and the duration in months on the X AXIS. The graph at the bottom displays the Running Sum of the difference between the amount claimed by the plaintiff and the amount of settlement between 2007 and 2012 on the Y AXIS, and the length in months on the X AXIS.

In addition, a time range (0-60 months) was inserted as a band (grey) in the X AXIS to identify the analysis area to be treated in the subsequent Story point.

The highlights in the Story point are as follows.

Trend analysis: The chart intends to identify trends in the data over time. The upward trend in the running sum indicates that the variables are increasing.

Change point detection: The change in the slope of the line indicates a change point in the data. This could be the result of a new behavior in the data.

Seasonality analysis: In the charts we can identify seasonal patterns in the data.

Comparison: The charts are used to compare the running sum of the different variables.

Transition: The next Story point intends to explain in detail the change point detection and the reasons for seasonal pattern.

**Story Point 9**

Story point 9 overlaps the graphs shown in Story point 8. The graph displays the Running Sum of the number of Cases on the left Y AXIS, in the right Y AXIS displays the Running Sum of the difference between the amount claimed by the plaintiff and the amount of settlement Closed from 2007 to 2012, and the duration in months on the X AXIS.

The highlights in the Story point are as follows.

Change points detection at the month sixty: The top point of the yellow graph represents a total of 40,208 (98%) closed cases, while the bottom point represents a total of 2.5 billion (93%) of the difference between the amount claimed by the plaintiff and the amount of settlement.

Trend analysis: The graphs show similar trends in the data over time. A growing trend in the running total shows that the variables are increasing. The defined changing points in both graphs reach over 93% of the running sum.

In summary, running sum was used as a tool for analyzing the data over time. They provide valuable insights into trends, patterns, and changes in the data that can inform decision-making. In this situation there is a relation between the number of closed claims, difference between the amount claimed by the plaintiff and the amount of settlement from 2007 to 2012 over the time.

The result suggests that there is no reason to continue the claims process after sixty months because there is no significant difference between the amount claimed by the plaintiff and the amount settled. The extension of the process just creates disadvantages for the state of Texas, insurance providers, and customers who will see no change in the eventual result.

Transition: The next Story point show the conclusions and recommendations for the Texas Department of Insurance (TDI)

**Story Point 10**

Conclusion & Recommendations

1. The apparent relationship between the number of cases and the variance indicates that the final amount granted is less in counties with more cases than in counties with fewer cases.

2. Agriculture, apartments townhomes &condos, manufacturers of chemicals, dentists, liquor liability, daycare centers, and construction firms are the most affected in the top 5 counties that have the highest variance.

3. The counties with the longest case resolution duration show the greatest final amount awarded per case (variable).

4. Cases resolved in less than 30 months account for about 90% of the variance. As a result, based on the data, we recommend establishing a method to pay the amount requested in situations that exceed 24 months and the most affected sectors to be given more consideration as the victims suffer for long periods without getting settlements. However, it is important to note that for cases exceeding $1,000,000 USD, there are relevant outliers that require special attention.

5. A breakdown in Business Class will be developed for further study in order to provide more specific recommendations to the TDI.

**References**

Insurance, T. D. (2023, 4 25). TDI. Retrieved from Property and Casualty Reports: https://www.tdi.texas.gov/reports/report4.html#closed

U.S. Census Bureau, P. D. (2020, 01 08). TEXAS STATE LIBRARY and ARCHIVES COMMISSION. Retrieved from Population Estimates for Texas Counties, 2010-2017: Arranged in Descending Order: https://www.tsl.texas.gov/ref/abouttx/popcnty201011.html