-----------------------------------------------------------------------------------------------------------

Memorandum

-----------------------------------------------------------------------------------------------------------

To: The CEO of General Casualty Kansas City

From: Headley Appollis

Subject: Clustering Analysis

Date: December 5, 2017

-----------------------------------------------------------------------------------------------------------

**Abstract:**

How can the clients be segmented in the crash dataset that can ultimately assist us in making better pricing decisions? How can the data assist us in choosing which region to penetrate next as we would like to go nationwide with this product?

**Overview of the data:**

I have analyzed the crash data set and due to the nature of the data, I had to perform data analysis by using various techniques. There were some interesting trends in the data, e.g. most accidents happened in the first quarter of the year during the winter months and in rural areas. The more obvious trends that were evident, i.e. 65% of drivers were male, and most accidents happened to drivers younger than 36 years, and predominantly over weekends.

**Analysis summary and key findings:**

By conducting a cluster analysis on the crash data, the following were concluded. I found that was five (5) statistically distinct groups within the data. The groups were relatively the same size except for one smaller group, however there are still distinct differenced amongst the groups and strong similarities within the groups.

The characteristics of the five clusters is:

* Cluster 1 (534 crashes): Accidents happened on rural roads, in cloudy conditions, with young female drivers age 36 and below, the driver was injured; the group is associated with serious injuries that happened around the middle of the month.
* Cluster 2 (877 crashes): Accidents happened on rural roads, in clear conditions, with young male drivers age 36 and below, the passenger was injured; the group is associated with serious injuries that happened around the 24th of the month.
* Cluster 3 (1374 crashes): Accidents happened on urban roads, in clear conditions, with young male drivers age 36 and below, the driver was injured; the group is associated with serious injuries that happened around the 5th of the month.
* Cluster 4 (1136 crashes): Accidents happened on rural roads, in clear conditions, with young male drivers age 36 and below, driver was injured; the group is associated with serious injuries that happened around 7th of the month.
* Cluster 5 (1078 crashes): Accidents happened on rural roads, in clear conditions, with young male drivers age 36 and below, the driver was injured; the group is associated with serious injuries that happened around the 24th of the month.

There are geographical differences between the groups as per the table below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cluster | Midwest | Northeast | South | West |
| 1 | 19.1% | 37.6% | 18.9% | 21.2% |
| 2 | 18.6% | 14.5% | 16.2% | 48.3% |
| 3 | 9.0% | 12.0% | 55.1% | 23.1% |
| 4 | 48.2% | 7.5% | 15.7% | 26.9% |
| 5 | 21.4% | 9.5% | 41.6% | 24.4% |

Midwest is strongly associated with cluster 4, Northwest with cluster 1, The South with both clusters 3 and 5, and the West with cluster 2.

**Recommendations:**

Based on the characteristics of each cluster per region and based on the distinguishing factors I’d like to make a few recommendations that can help with differential pricing. The South and/or West region can be considered first or both can be attempted with the expansion strategy plan.

The South region is strongly associated with cluster 3 and 5. A pricing preference can be given to female drivers and all drivers older than 36 years of age. In addition a low premium with high deductible can be presented if customers claim in January – this will hopefully deter reckless driving during the peak winter months.

The West region is strongly associated with the cluster 2. A pricing preference can be given to customers that do not make use of rural roads on a day-to-day basis and customers that are older than 36 years as well as all female drivers. Fewer accidents happened in the first two weeks of the month for this cluster. In addition the lower premium with a high deductible can be presented if customers claim in the month of January – this will hopefully deter reckless driving and encourage cautious driving, during the peak winter month.

The Northwest region is strongly associated with the cluster 1. A pricing preference can be given to customers that do not make use of rural roads on a day-to-day basis and customers that are older than 36 years as well as all male drivers. In addition the lower premium with a high deductible can be presented if customers claim in the month of January – this will hopefully deter reckless driving and encourage cautious driving, during the peak winter month.

The current Midwest region is strongly associated with cluster 4. So I will use the same pricing strategy as I’ve mentioned above.