|  |
| --- |
| Memorandum |

# Improved Pricing & Product Offerings

To: CEO of General Casualty Kansas City (GCKC)

From: Michael Rawat

Date: 11/22/2017

### Core Findings

. ………………………………………

The most common profile of an individual involved in a fatal automobile accident is a Male, over the age of thirty, driving in urban areas in the south. Therefore, premiums for individuals who fit characteristics of this profile should be higher than individuals who do not. Furthermore, it is apparent that fatal accidents are less likely to occur in the North-Eastern region, therefore if approved for expansion, states such as Connecticut and Maine should be heavily considered.

### Fundamental Questions

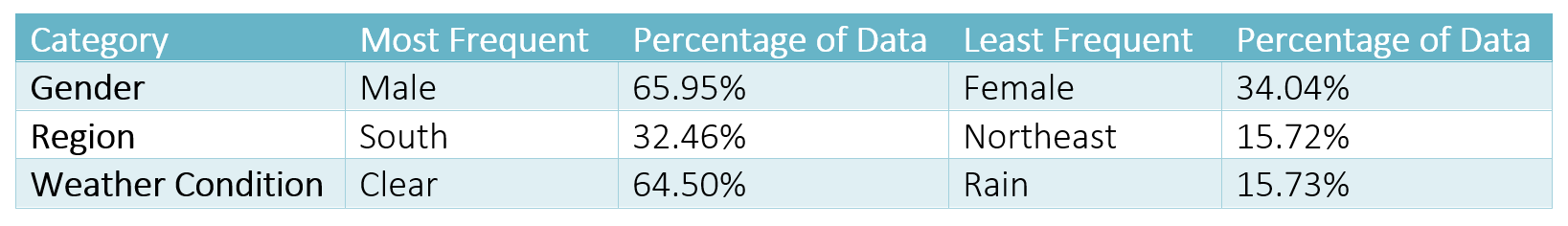
Is it possible to improve the pricing and product offerings for automobile insurance using data provided on fatal automobile accidents from 2011? Furthermore, is there evidence to support the expansion of automobile insurance into another region as a positive strategy, if so, which region?

### Analysis and Commentary

. ………………………………………

Initially the data set contained missing records and incorrect values across many categories. For accurate analysis, records with missing values were either discarded or the values were calculated based on common practices. For example, records that had a missing value for age were filled by the average age of individuals in the data based on the person type. This preserves all the information captured in other categories whilst not harming the integrity of the category being filled. For a more detailed explanation of how every variable was handled in this stage, refer to my R Markdown file attached.

Following the handling of incorrect and missing data, frequency diagrams were created to ascertain a surface level understanding of what the data set contains. The following table contains the major insights provided by frequency charts:



The result for Gender is not surprising as males are considered more aggressive drivers than females and therefore more prone to accidents. However, the fact that this data set shows more fatalities when the weather is clear as opposed to more challenging conditions is interesting. One justification may be that individuals take extra precautions when the weather is poor, and become over confident when the weather does not appear to be an issue. To look further into profiles of individuals more prone to characteristics, the data needs to be grouped because currently the variables in the table above are independent of one another.

Beyond these initial findings, the data was broken down into groups whereby records within a group are similar and records from different groups are dissimilar. It was clear that the data set is naturally formed by four main groups. The characteristics of each group are shown in the table below:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Group | Gender | Weather | Road Type | Age | Involvement | Injury | Region |
| 1 | Male | Clear | Urban | 30 | Driver | Fatal | South |
| 2 | Female | Clear | Rural | 19 | Passenger | Various | West |
| 3 | Male | Clear | Rural | 39 | Driver | Fatal | West |
| 4 | Male | Clear | Urban | 64 | Driver | Fatal | South |

These groups highlight the most important factors that contribute to being involved in an accident. For example, Males are a dominant characteristic of groups that are involved in fatal accidents, as also indicated by the frequency charts. However, these groups help establish links between multiple categories. For example, the table also shows that Males driving in Urban areas in the South is a good indication of a higher risk individual since this combination of characteristics appear in two groups.

### Formal Recommendations

. ………………………………………

Based on the initial frequency charts and group findings, a more detailed hierarchy can be established regarding pricing for automobile insurance. This is necessary because not everyone driving in the south will be male, over the age of 30, and driving in an urban area. Instead there will be customers that match only some of these characteristics. Therefore, based on the dominant characteristics determining groups, insurance prices should be staggered based on Gender first, followed by Age and Region. Whereby Gender has the largest impact on an insurance price and Region has the least impact on an insurance price. For example, a 30-year-old Male driving in the West should have a higher insurance premium than a 50-year-old female driving in the South, because Gender is the most significant characteristic.