**PROJECT CASE 1 – GROUP 8**

**THE UK AUTOMOBILE INSURANCE COMPANY**

**Team Members:**

**Animesh Johri**

**Manoj Natarajan**

**Navashree Venkatesan**

**Nikhil Salian**

**Vivek Chander Selva Kumar**

**Objective of Case Study**

We are part of the data analytics team working for The UK Automobile Insurance company based in UK. We have a division that handles auto insurance and we are analyzing data related to accidents and mapping it to the conditions and demographics related to the accident site. The objective of the case study is to gain insights such as main reasons for accidents the circumstances like weather and road conditions to plan our accident coverage. We are also planning to run a campaign for road safety and will use insights from the study to demonstrate it better

**Dataset Source**

The dataset used for this case study was referenced from the following source

https://www.kaggle.com/silicon99/dft-accident-data

<https://data.gov.uk/dataset/road-accidents-safety-data>

**Dataset Overview**

This data is about vehicle collisions that took place in UK in the year of 2006 and has been collected by UK police force. There are 3 files “*Accidents vlookup.xlsx*”, “*vehicles vlookup.xlsx*” and “*Road-Accident-Safety-Data-Guide.xls”.***“*Accidents vlookup.xlsx*”** is the primary data source and “*vehicles vlookup.xls” is the secondary data source.* ***Accident\_Index*** is the key which is used to link the two data sources

**Data Manipulation**

The data source file had numeric data in it and this numeric data was referenced in the file “***Road-Accident-Safety-Data-Guide.xls****”* as we could not gain any insights using this numeric data we performed VLOOKUP on the data source file to map the numeric values to the exact text present in the lookup file. We also filtered out data and performed analysis for the year 2006

**Inspiration for Insights**

**Issues to be addressed using this data**-

* What are the trends for accidents involving different road users e.g. motorcycles, peds, cyclists?
* Do road safety campaigns really work?
* What are the chances of accidents for different groups or vehicles ?

**Insights**

1. **The type of vehicle that has met with the maximum number of accidents:**

The first insight obtained from the data is to find the type of vehicle that has met with the maximum number of accidents. We can find this by visualizing the number of vehicles that have met with an accident in each case and along with the type of vehicle. From this we have found that the type of vehicle which has met with the maximum number of accidents is Cars and the maximum number of vehicles that have met with the same accident is 16. This allows us to negotiate the insurance value for the different type of vehicles.

**Visualization 1: Packed bubbles chart**

The bubble chart shows the different types of vehicles that were involved in the accidents and the size of the bubble represents the vehicle which has met with the maximum number of accidents. In our visualization, it is clearly seen that car is the type of vehicle that has met with the maximum number of accidents and second the pedal cycle and van is the third type of vehicle that has met with the maximum number of accidents.

**Visualization 2: Box-and-whisker plot**

The box-and-whisker plot shows us the average of each type of vehicles involved in an accident which is represented by the box and it also shows the outliers for each type of vehicle; the maximum and minimum number of each type of vehicle involved in a single accident. The visualization shows that the median number of cars involved in a single accident is one and the upper whisker is 3 and lower whisker is 1. The farthest outlier value for cars is 16.

**Visualization 3: Pie chart**

The pie chart shows the percentage of each type of vehicle involved in accidents against the total number of accidents. The type of vehicle which is involved in the maximum number of accidents is the car which is ----% of the total value.

1. **Does Light condition relate to features like age in case of accident:**

This insight is used to understand if there is any direct co-relation between light condition and other factors like age, road type. The findings will help the organization by planning the policies and the premium to be paid accordingly on basis of age or place of residence for the people buying insurance from our organization. We have achieved this using following three visualizations.

**Visualization 1: Bar Chart**

From this visualization, we can infer that the top 2 light conditions during which maximum accident take place are

1. Daylight
2. Darkness Light lit

The age group involved in most of these accidents were between 25 – 35 years. We can see here that age and light conditions are not directly proportional i.e. If a person is old and is driving under dark conditions, then it does not mean that there will be higher probability of accident and we don’t have to charge higher premiums to people based on age

**Visualization 2: Stacked Bars**

Using this visualization, we can see that the top 2 objects that are being hit during an accident are

1. Other Permanent Objects
2. Trees

Also, here we see that that age group that hits permanent objects during an accident is 16-35.

This may indicate rash driving in the younger age group and more matured and sensible driving

as the age group moves higher

**Visualization 3: Area Chart**

Using this visualization, we can see that the proportion is maintained between the age and the sex of the driver as the age increases and there is no difference i.e. female drivers do not commit more accidents as their age increases