**Cleaining and Transformation of data**

* The dataset is loaded into power BI by selecting text as the dataset is in the csv format.
* All the cleansing and transformation of data was done in the Query Editor view of Power BI.
* The data set has column headers but power Bi did not recognize this. Promoted the first row of the data set as it contains the column headers.
* Added an index column that starts from 1 and named it as Record\_no.
* Check for null values or NA values of the fields that have irrelevant data types.
* Before changing the data type of a field check if the column has any values that made Power Bi to assume the data type. The year was a text field as it had NA values. The NA values were filtered before changing the datatype. If this is not done in this order Power BI will return error values when it tries to convert NA to number or date.
* Several columns were created using the conditional and logical statements to provide categorzationn that helps with the views.
* The DAX formulas in Power Bi are case sensitive and the syntax should be followed strictly while creating the formulas.
* The formula to categorize age groups

if [DRIVERAGE]<=2 then "Infants"

else if(([DRIVERAGE]>2) and ([DRIVERAGE]<=12)) then "Children"

else if(([DRIVERAGE]>12) and ([DRIVERAGE]<=17)) then "Teenagers"

else if(([DRIVERAGE]>17) and([DRIVERAGE]<=25)) then "Youth"

else if(([DRIVERAGE]>25) and ([DRIVERAGE]<=40)) then "Adults"

else if (([DRIVERAGE]>40) and ([DRIVERAGE]<=60)) then "Middle Age"

else "Old"

* The formula to calculate road types by speed limit

if [SPEEDLIMIT] <= 20 then "Low speed"

else if [SPEEDLIMIT] >20 and [SPEEDLIMIT] <= 30 then "Urban(30mph)"

else if [SPEEDLIMIT] >30 and [SPEEDLIMIT] <= 35 then "Unpaved(35mph)"

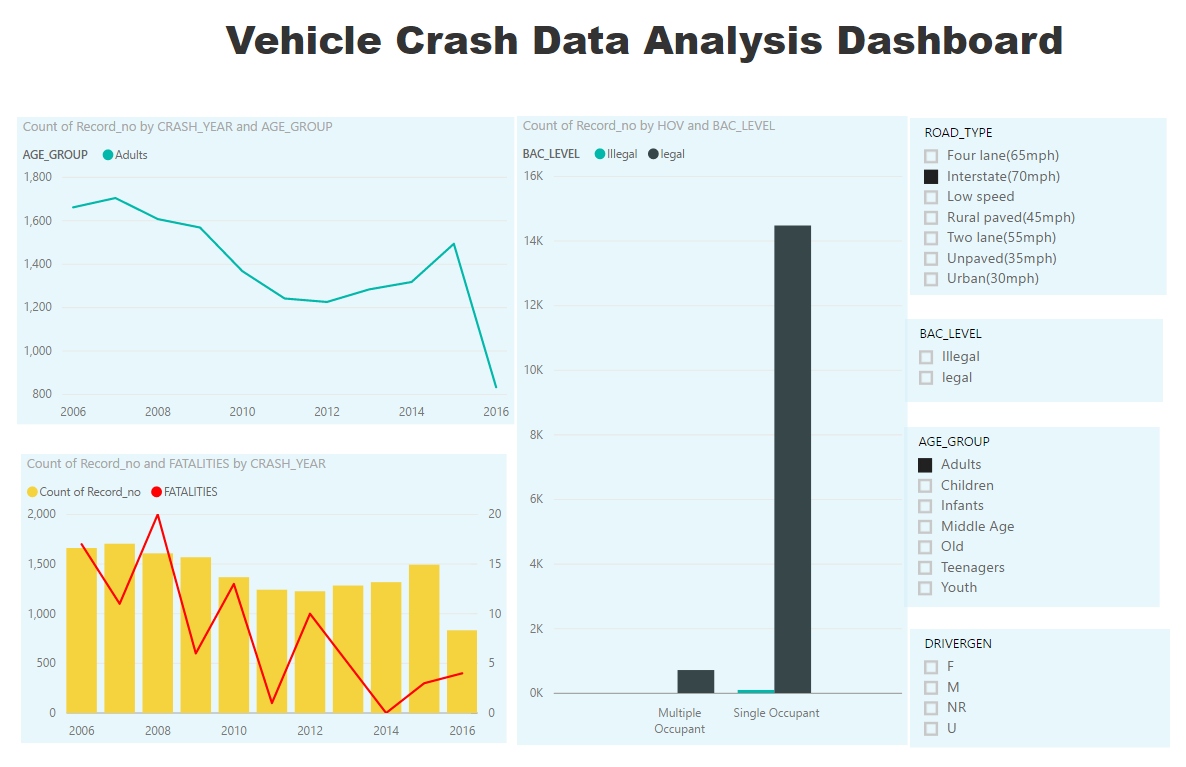
else if [SPEEDLIMIT] >35 and [SPEEDLIMIT] <= 45 then "Rural paved(45mph)"

else if [SPEEDLIMIT] >45 and [SPEEDLIMIT] <= 55 then "Two lane(55mph)"

else if [SPEEDLIMIT] >55 and [SPEEDLIMIT] <= 65 then "Four lane(65mph)"

else "Interstate(70mph)"

* Power Bi also allows to add columns by specifying the condition in the Condiitonal column button in the Add column tab og the Query Editor. Bac\_level and HOV were added that way.



Vehicle Crash Analysis Dashboard using Power BI

It has three charts

* A line chart that shows the number of crashes from 2006 to 2016 by age groups.
* A line and column combination chart that shows the number of crashes from 2006 to 2016 in the column and the number of fatalities from 2006 to 2016 in the line.
* A clustered column chart that shows the number of crashes with legal and illegal Blood Alcohol levels. It also splits the data by single and multiple occupants at the time of the crash.
* There are filters for road types, Bac levels , age groups and driver gender that allows the user to filter the data in the visualization.