IMDE – CA1

[Document subtitle]

Eddy ekofo - X00119959

2017

# Summary:

I am hoping to learn how to acquire data, learn to manage and clean data, visualize data appropriately for the user, meanwhile how to use Tableau. I am hoping to discover how does age, weather, road surface condition etc. has an impact between men and women when it comes to driving.

# Background:

I had many ideas and I was looking for many datasets but I could not find any that interested me enough to do this assignment. First, I had the data about the number of students who entered 3rd Level Education in Ireland and the Institute in which they enrolled in. Because I didn’t have enough numerical columns to work with therefore I searched and changed and found this dataset.

I found this dataset I thought it would be nice to play around with and see what I could come out with. Therefore, I am delighted to have used this for my assignment.

# Dataset:

|  |  |
| --- | --- |
| Name | From |
| 2014 Leeds Road Traffic | <https://data.gov.uk/dataset/road-traffic-accidents> |

# Seven Stages:

## Acquire & Parse:

I acquired from [data.gov.uk](https://data.gov.uk/dataset/road-traffic-accidents) which is a dataset based on Leeds, England. There are many datasets on road collisions alike, but they didn’t have which gender were the cause of the accidents.

The highlighted green columns show the UK Grid Reference for Noting & Easting. The datatype which Tableau consumes are longitude and latitude. After a long online research on how to convert grid reference to longitude and latitude I ended up with UK Grid Reference Finder (<http://gridreferencefinder.com/batchConvert/batchConvert.php)> which does the conversion.

Table 1: Parse

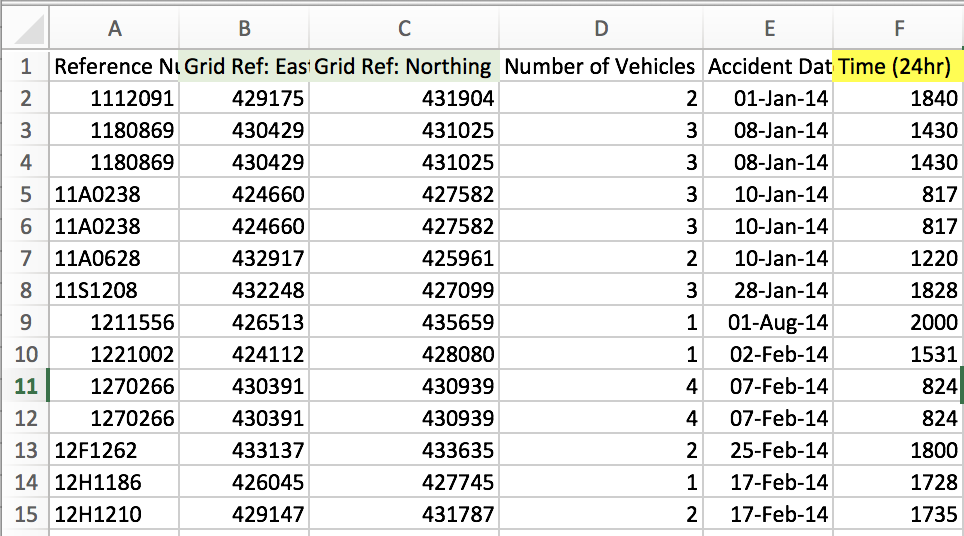
|  |  |  |  |
| --- | --- | --- | --- |
| Grid Ref: Easting | Grid Ref: Northing | Latitude | Longitude |
| 429175 | 431904 | 53.78265 | -1.5587125 |
| 430429 | 431025 | 53.774678 | -1.5397685 |
| 430429 | 431025 | 53.774678 | -1.5397685 |
| 424660 | 427582 | 53.744037 | -1.6275753 |

The first two columns is the are the UK Grid Reference, then the latter two are them converted to longitude and latitude.

## Filter:

I removed unwanted columns that didn’t help me with the visualisation. Like the time (Table 2 highlighted yellow) which each accident occurred which doesn’t have a direct impact on the story I am trying to tell.

Table 2: Filter



### Mine:

In this phase I did some calculations on the total number of people who are hurt during the casualties and the percentage between them both. In my findings, there seems to be more men who crash but more women overall are hurt during the accidents.

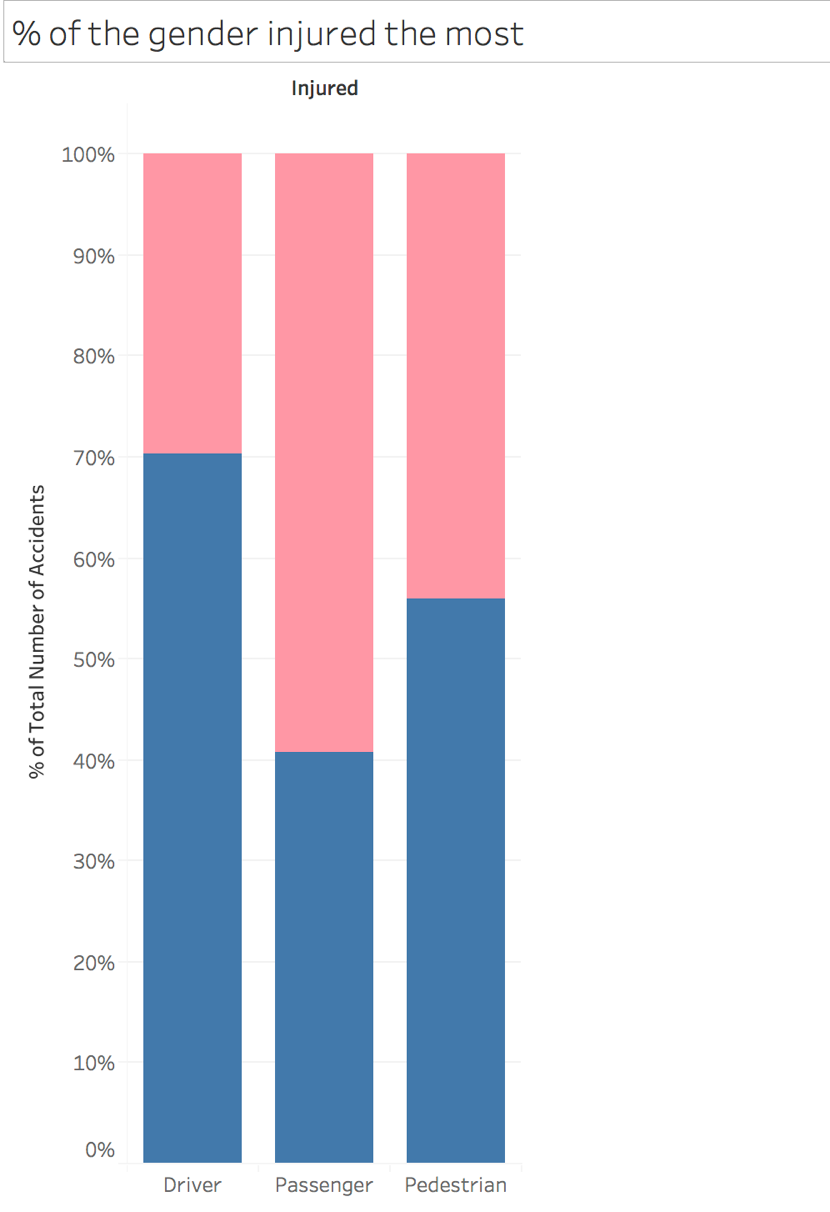
This figure shows us that the majority if the injured Drivers are male. This can mean that most of the drivers are men so there are more men likely to be hurt during the accidents, or men are careless drivers. Age would play a part also, since most young men are known for reckless driving.

Figure 1: Percentage between Male & Female

## Represent:

This stage after all the needed calculations I finally represented the data by making many sheets on Tableau then finally I made a dashboard.

Firstly, with the longitude and latitude columns I converted from UK grid reference (see Acquire & Parse), I used the coordination to make a map, which identifies the area with the most accidents occurred.

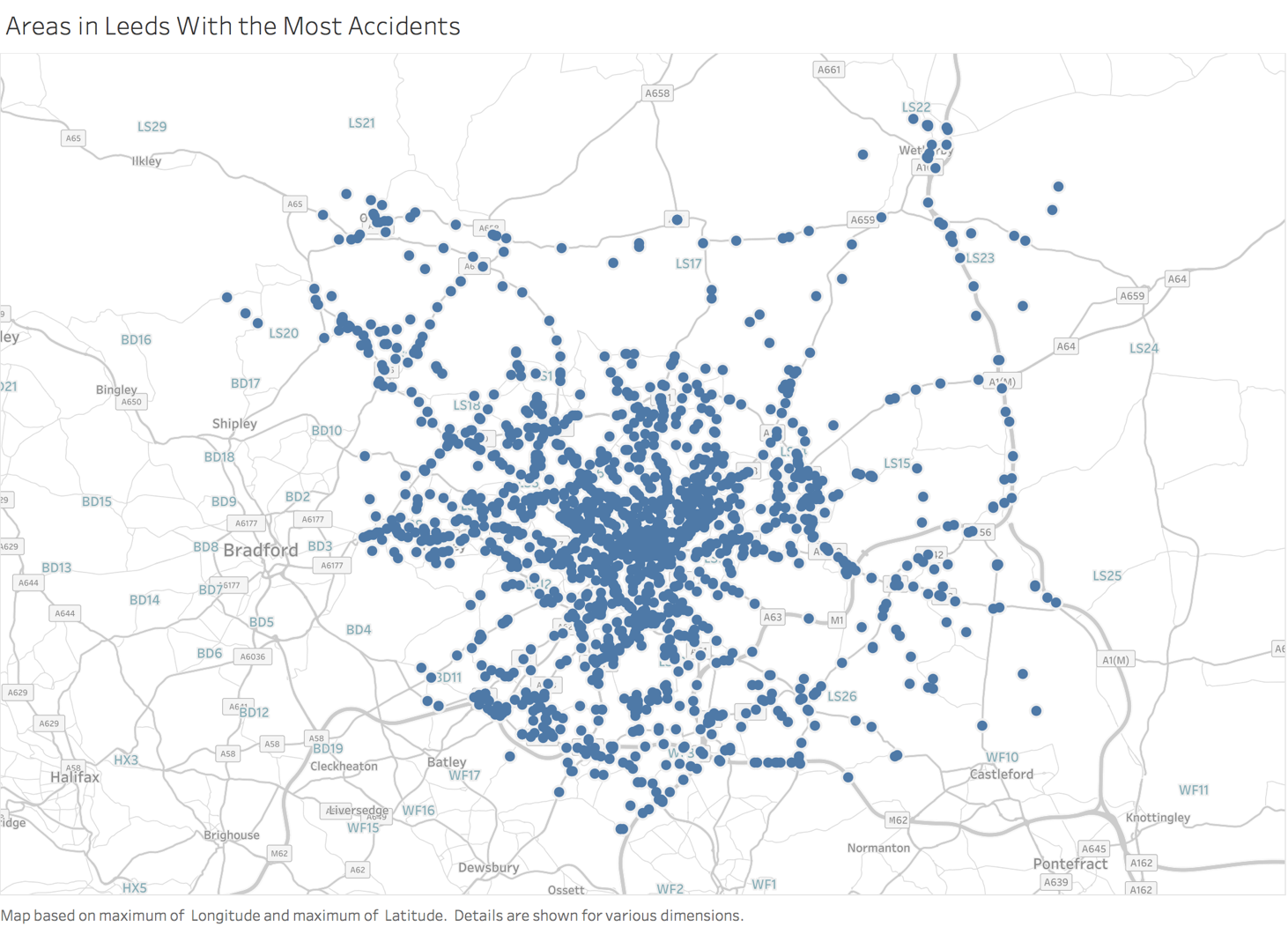


Figure 2: Each accident's occurrence.

By looking at this map then I was able to see there are deeply concentrated areas with plenty of accidents this could possibly mean that it is the centre of the town which gets many of the traffic of the city. and what was the weather condition like when the accidents occurred. The legend indicates the reoccurrence of accidents in one area.

Another worksheet shows which age of the drivers and the number of accidents they’re involved.

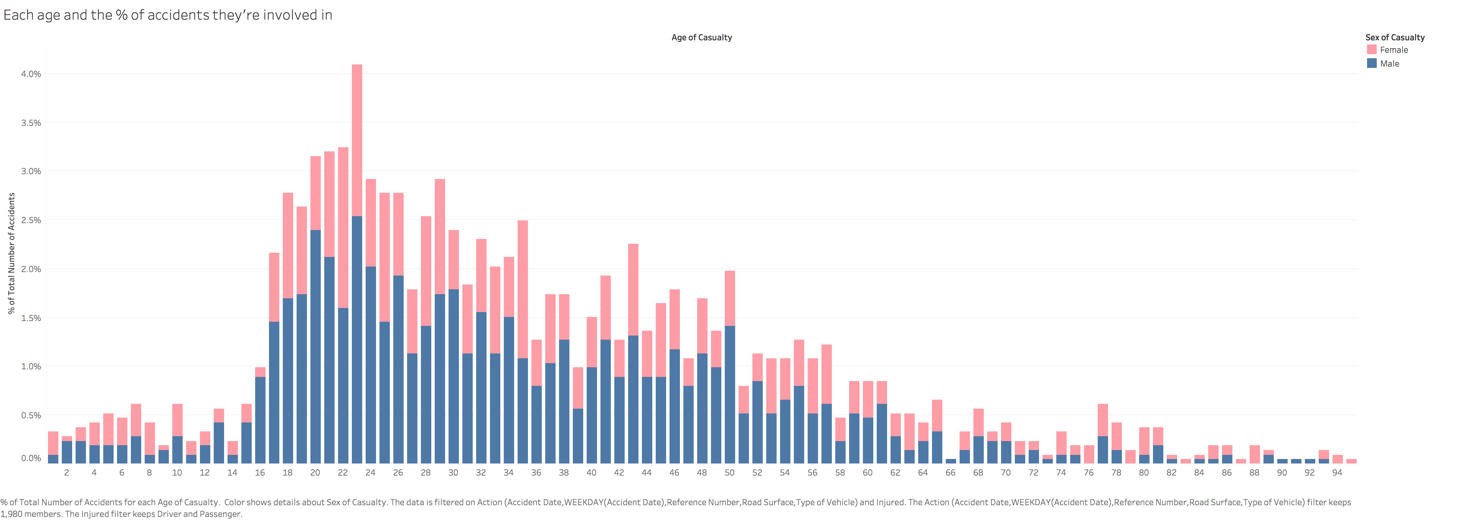


Figure 3: % of age with most accidents

From this I was able to identify which age and gender had the most causalities out of the whole.

## Refine:

I then refined many of my worksheets to get them to tell the overall meaning of the story I wish to tell.

On the map of Leeds, where there occurred more than one accident the shape of the bubble increased in size, signifying the danger or possible a busy street.

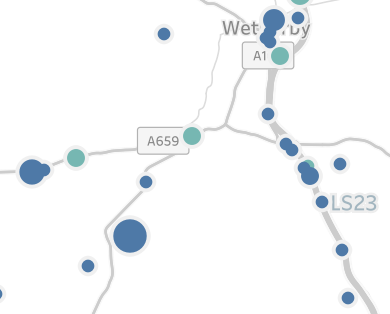
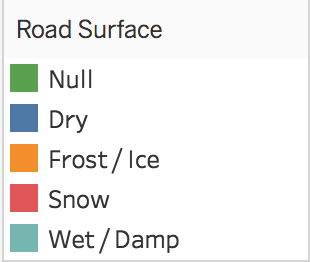


Figure 4: Difference Sizes indicates more accidents Figure 5: Colours represent weather conditions

The colour code indicates which condition was the road when the accidents happened.

This is the whole man refined for the users.

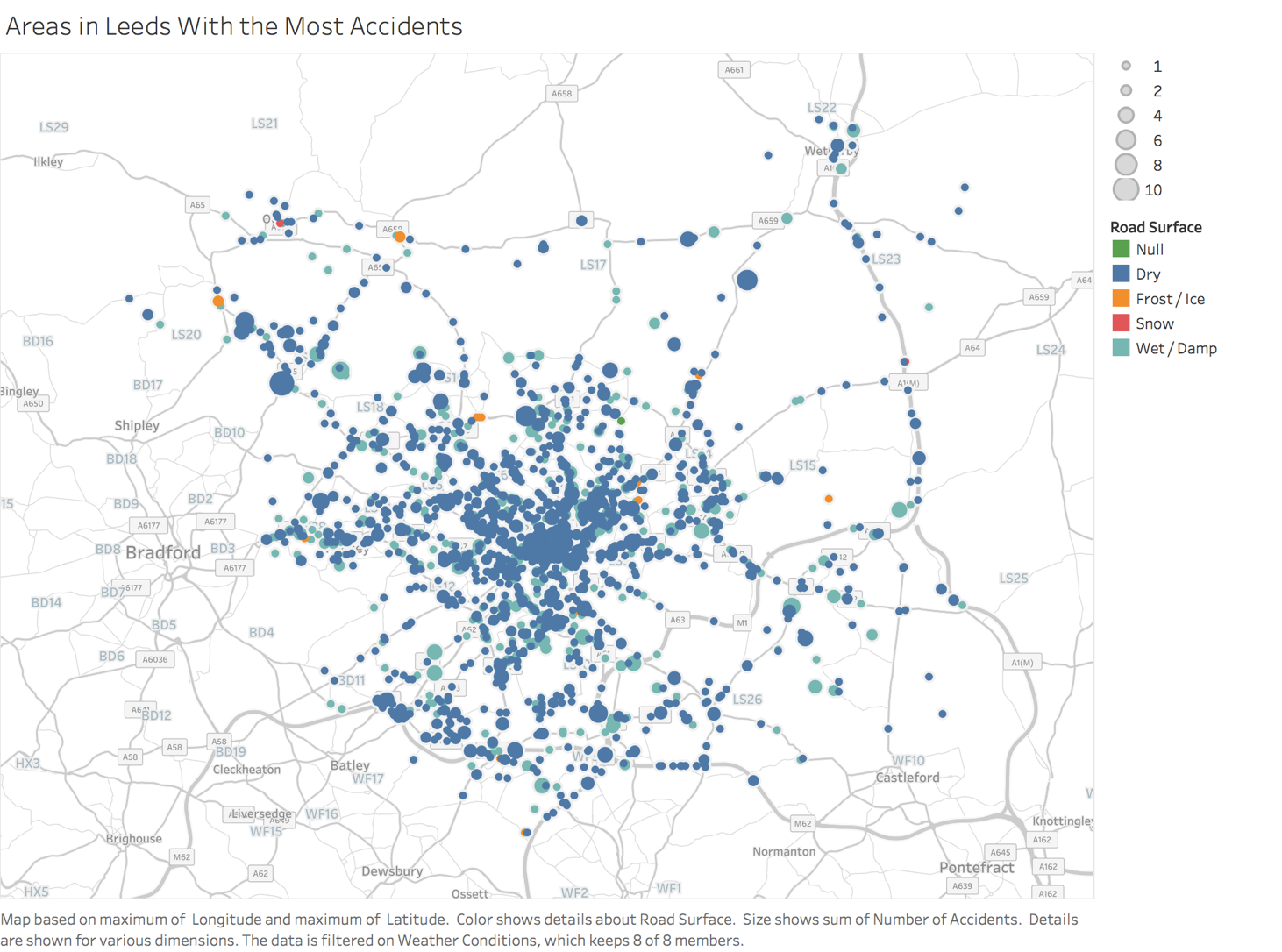


Figure 6: Refined Leeds map for more readability

## Interact:

I then made a dashboard which allows users to interact and explore the data. I have added filters in which when the user hovers over a certain data they would be able to see a narrow version of the dashboard.

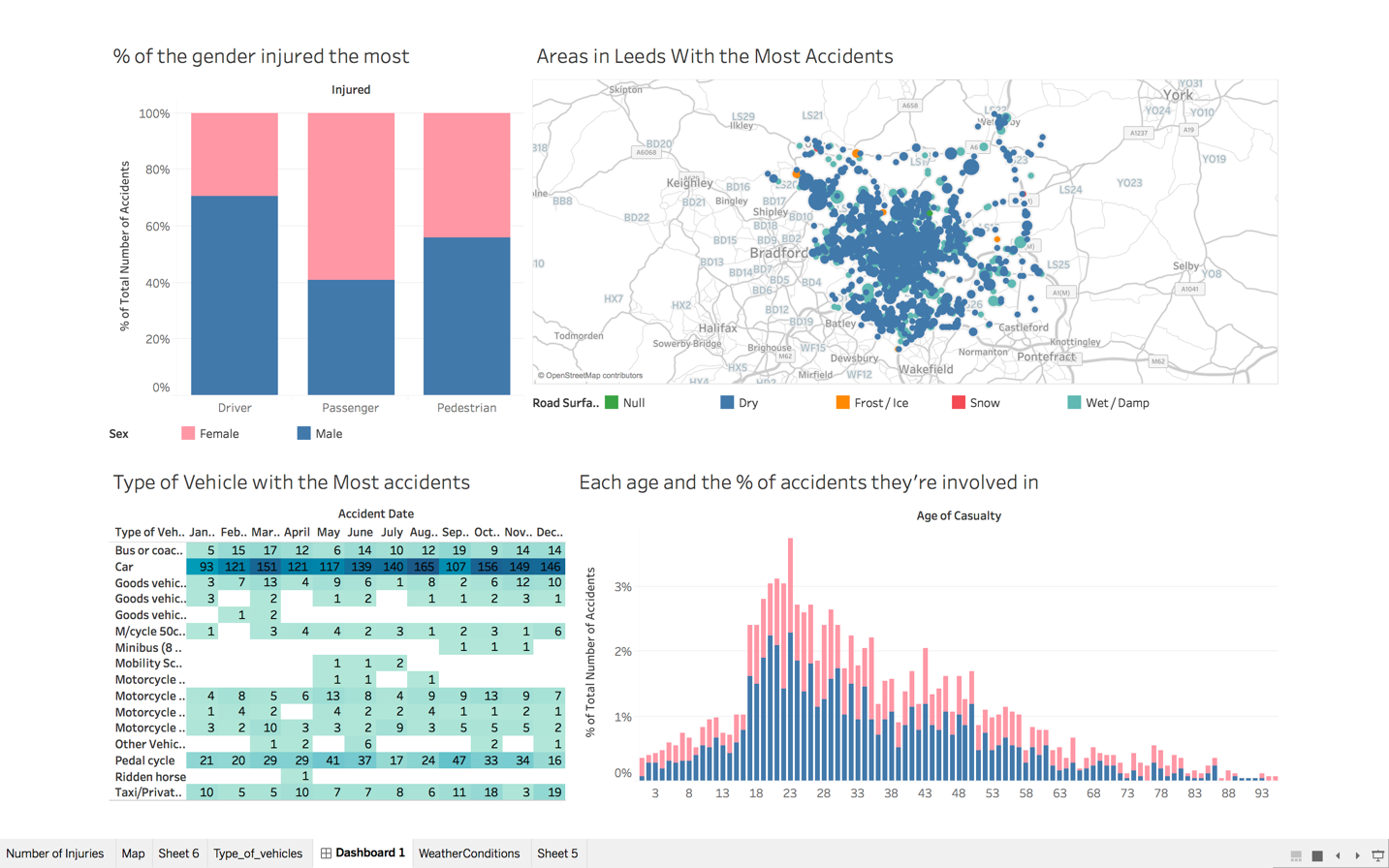


Figure 7: The dashboard

Here we see all the worksheets in one dashboard. These tell a story; each graph has a role to play in the whole story telling. Since I am trying to tell of the impact of road accident to each gender and age group, these graphs help me to do so.

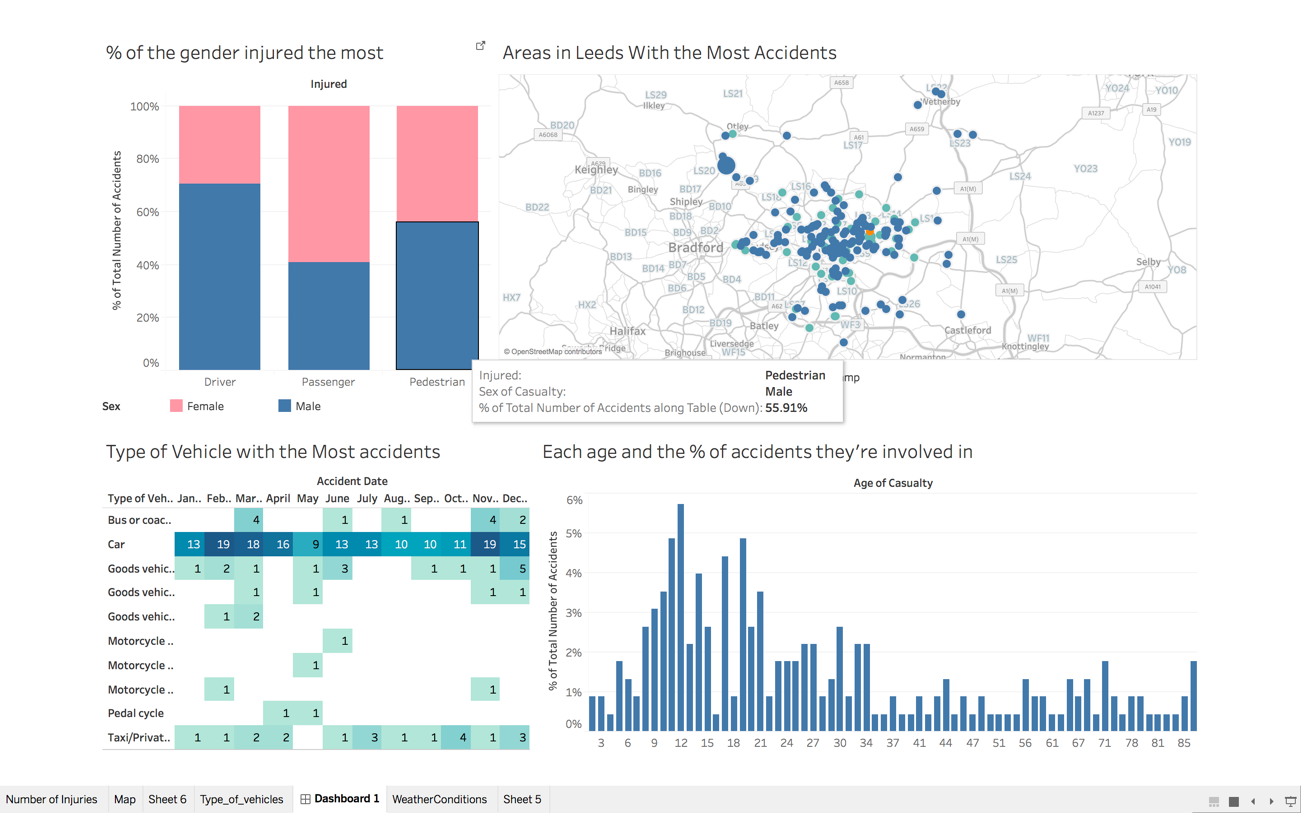


Figure 8: The male pedestrians injured in road accidents in Leeds

In figure 8, the user hovers over the “% of the gender injured the most” graph at male pedestrians, the whole graph narrows it down to when it occurred and which vehicle was involved in the collision. Which age-group were mostly injured at the incident. We see that young men from age nine - mid-twenties are the most involved in the accidents. Uses can interact to the data to explore it more. These filters are an option given by Tableau for a better user-experience.

# Problems and Solutions:

Finding the right dataset to accomplish this assignment was a big problem at the beginning. I finally ended up with this dataset because it was an interesting one and also it had plenty of numerical columns to create many visualisations with.

I first had the background of the Map to dark, but then I changed it to light because I wished to have consistency throughout my dashboard. This made the dashboard look neater and aesthetic.

I tried to create visualisation with time each accident occurred but I couldn’t format the time properly therefore I removed the column altogether from my dataset. This I was going through each stage again and again to make sure I had only used everything I needed.

I encountered a problem as mentioned in the Acquire & Parse stage concerning the UK Grid Reference. After an intensive search on the web, I was directed to a website which converts a large amount at once.

* Paste the data which to convert

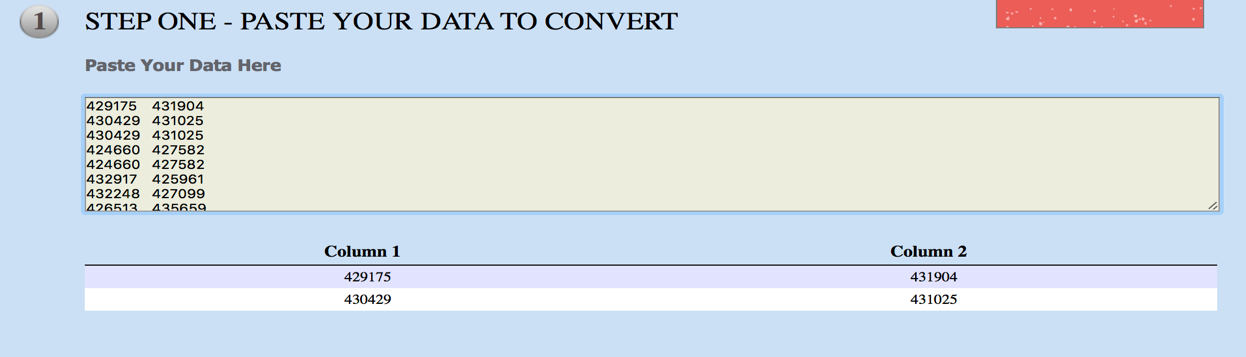


Figure 9: UK Grid reference to Longitude & Latitude

I copied 2 columns (Grid Ref: Easting, Grid Ref: Northing), pasted them in the text field. The rest of the steps are in the website.

* The sixth step is copying the converted data back to the excel.

# Summary:

If you live in Leeds in 2014, you’re most likely to be in a car accident if you’re a male during mid-twenties.

Why? it could be that:

* Most young men drive, the more of them in the streets the more likelihood they would have accidents.

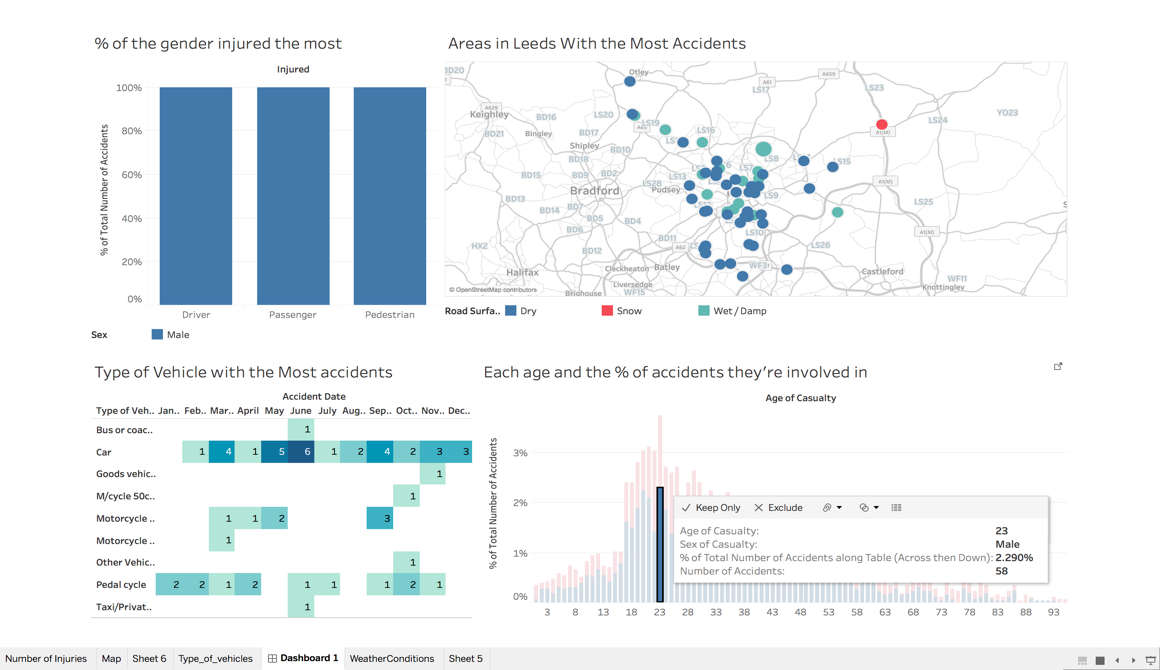


Figure 10: Male, at 23

The figure 10 shows how that most people involved in car accident were 23 years old, these shared the highest percentage of accidents in Leeds. On the sixth of June, there were 6 car crashes by men at the same age, maybe there is a peculiar reason for that.

* Although I didn’t have the speed in which many would’ve crashed yet I would say that the excess of speed by the young men may also be a reason that they’re in so many collisions.
* Something I never knew before, I though most accidents were committed by the female drivers, but this story tells me that young men in Leeds are the cause of most crashes. This might also tell the story not just for Leeds but throughout the UK.
* Most crashes happed during the time when the weather was “fine without high winds” and the roads were “Dry”. Maybe if I had the average rainfall that year I would have to estimate the frequency of crashes during “wet/damp”.

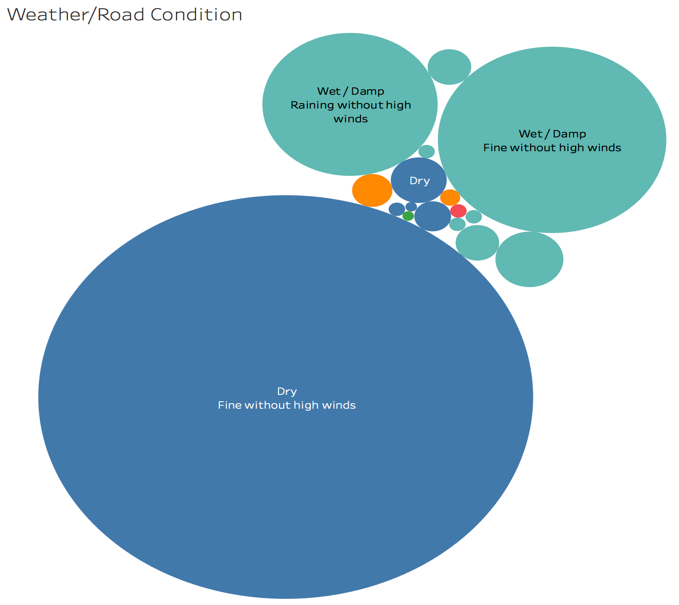


Figure 11: Weather condition