



**Data Exploration and Visualization  
Programming Exercise - 1**  
**Name: Subbulakshmi Natarajan**  
**Student ID: 34069178**

## 1. Data checking and cleaning (i.e., Steps 1 to 3)

- **Image of loading the dataset**

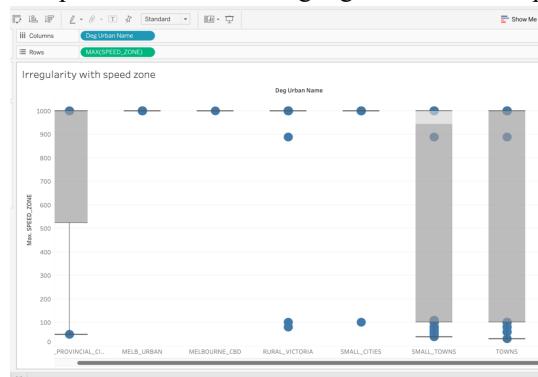
I changed my CSV file format to Excel format since I got an error when I was uploading the CSV file through text file format. We can see that the accident time data type is changed to date and time it can be said to be one of errors since we have a separate column as accident date which is there in the dataset.

This is image of what it looks like when I load it into tableau is below:

- **Irregularities and errors**

### 1) Speed Zone (Not on scale limit 0-110)

Using Tableau, I found the irregularities or errors through a box plot for the speed zone to depict the maximum speed that each city has taken. It can clearly be seen that the speed limit has crossed the scale like 888,999 and 777 which is not included in the speed scale. In the diagram, we can see that the speed zone I have taken is the maximum speed zone according to the urban name we can see that there are some data points which say beyond the speed scale as I have highlighted them in the picture. Below is the diagram for the following:



To correct this error, I have used Excel for clearing this unwanted data and made it null values now after the data is cleared here is how the figure looks:

**Step 1:** I filtered the irrelevant data that is 777, 888 and 999.

**Step 2:** Filtered the 777,888,999.

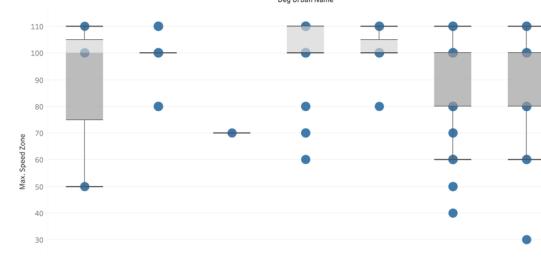
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
T2012000051	1/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WARRAMBOOL	SMALL CITIES	5 Not at intersection	999	-38.37474	142.023041				
T2012000054	4/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	777	-38.4536	140.214629				
T2012000055	5/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	777	-38.4536	140.214629				
T2012000073	3/1/2012	17:20:00	1	Tuesday	4	Collision with fixed object	3 Day	DARIBIN	MELB. URBAN	2 T-intersection	999	-37.72050	144.991546				
T2012000076	5/1/2012	15:30:00	1	Sunday	4	Collision with fixed object	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000079	4/1/2012	15:30:00	1	Sunday	4	Collision with fixed object	3 Day	WARRAMBOOL	SMALL CITIES	5 Not at intersection	999	-38.37474	142.023041				
T2012000081	3/1/2012	17:20:00	1	Tuesday	4	Collision with fixed object	3 Day	WARRAMBOOL	SMALL CITIES	5 Not at intersection	999	-38.37474	142.023041				
T2012000083	5/1/2012	20:10:00	1	Thursday	3	Struck pedestrian	3 Day	RURAL VICTORIA	RURAL VICTORIA	5 Not at intersection	999	-37.8446	147.055228				
T2012000086	1/1/2012	05:30:00	1	Sunday	4	Collision with fixed object	3 Day	DAIRYLAND	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000087	7/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	CARLTON	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000088	1/1/2012	1:00:00	1	Sunday	4	Collision with fixed object	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000089	1/1/2012	11:00:00	1	Sunday	4	Collision with fixed object	3 Day	MITCHELL	RURAL VICTORIA	5 Not at intersection	999	-37.7716	145.083872				
T2012000090	1/1/2012	11:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000091	1/1/2012	11:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000094	3/1/2012	18:00:00	1	Tuesday	6	Vehicle overturned (no collision)	2 Day	HEPHERN	TOWNS	5 Not at intersection	999	-37.8078	144.957074				
T2012000095	1/1/2012	15:00:00	1	Sunday	4	Collision with vehicle	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000097	3/1/2012	10:15:00	1	Tuesday	4	Collision with vehicle	3 Day	WARRAMBOOL	SMALL CITIES	5 Not at intersection	999	-38.37474	142.023041				
T2012000098	3/1/2012	10:15:00	1	Tuesday	4	Collision with vehicle	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.37474	142.023041				
T2012000099	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	BRUNSWICK	MELB. URBAN	2 T-intersection	999	-37.73050	144.782616				
T2012000100	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	MONASH	MELB. URBAN	5 Not at intersection	999	-37.66234	145.000000				
T2012000101	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	MELBOURNE	MELB. URBAN	5 Not at intersection	999	-37.8078	144.957074				
T2012000102	1/1/2012	15:30:00	1	Sunday	1	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	1 Cross intersection	999	-37.8078	144.957074				
T2012000103	1/1/2012	17:45:00	1	Sunday	1	Collision with vehicle	3 Day	PORT MELBOURNE	MELB. URBAN	5 Not at intersection	999	-37.64023	144.935070				
T2012000104	1/1/2012	18:00:00	1	Monday	1	Collision with vehicle	3 Day	WHITEHORN	MELB. URBAN	4 Multiple intersections	999	-37.7716	145.083872				
T2012000105	1/1/2012	18:00:00	1	Monday	1	Collision with vehicle	3 Day	WHITEHORN	MELB. URBAN	5 Not at intersection	999	-37.64023	145.083872				
T2012000106	1/1/2012	18:30:00	1	Monday	1	Collision with vehicle	3 Day	WHITEHORN	MELB. URBAN	5 Not at intersection	999	-37.64023	145.083872				

### Step 3: Removed the irrelevant data and make as null values.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
T2012000051	1/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WARRAMBOOL	SMALL CITIES	5 Not at intersection	999	-38.37474	142.023041				
T2012000054	4/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	777	-38.4536	140.214629				
T2012000055	5/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	777	-38.4536	140.214629				
T2012000073	3/1/2012	17:20:00	1	Tuesday	1	Collision with vehicle	3 Day	DARIBIN	MELB. URBAN	2 T-intersection	999	-37.72050	144.991546				
T2012000076	5/1/2012	15:30:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000079	1/1/2012	20:10:00	1	Thursday	3	Struck pedestrian	3 Day	RURAL VICTORIA	RURAL VICTORIA	5 Not at intersection	999	-37.8446	147.055228				
T2012000081	1/1/2012	05:30:00	1	Sunday	4	Collision with vehicle	3 Day	CARLTON	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000083	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000086	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000087	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000088	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000089	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000094	3/1/2012	18:00:00	1	Tuesday	6	Vehicle overturned (no collision)	2 Day	HEPHERN	TOWNS	5 Not at intersection	999	-37.8078	144.957074				
T2012000095	1/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000097	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	BRUNSWICK	MELB. URBAN	2 T-intersection	999	-37.73050	144.782616				
T2012000098	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	MONASH	MELB. URBAN	5 Not at intersection	999	-37.66234	145.000000				
T2012000099	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	MELBOURNE	MELB. URBAN	5 Not at intersection	999	-37.8078	144.957074				
T2012000100	1/1/2012	15:30:00	1	Sunday	1	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	1 Cross intersection	999	-37.8078	144.957074				

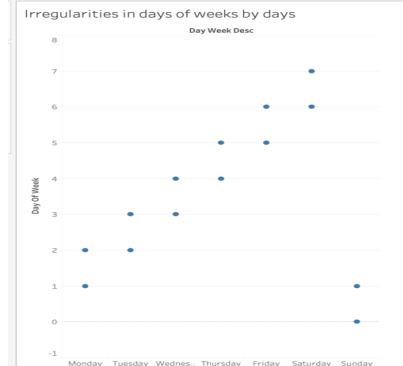
After the data is cleaned the figure below in Tableau:

Speed zone vs urban name (cleaned data)



## 2) Days of week numerically the data is wrongly numbered according to the numerical order.

In the tableau, I checked the day of the week with the day of the week desc that is Sunday, Monday etc. When I noticed that it says 0 to 7. According to me, the week should start from 1- Sunday till 7 - Saturday we have a total of 7 days in a week. Below is the graph which shows irregularity in days of weeks by days which clearly shows that Sunday and Monday have week 1 so we can technically say it is wrong which is showing in the graph below:



To clean this data, we will be using Excel to adjust the number of days in the column of day of the week.

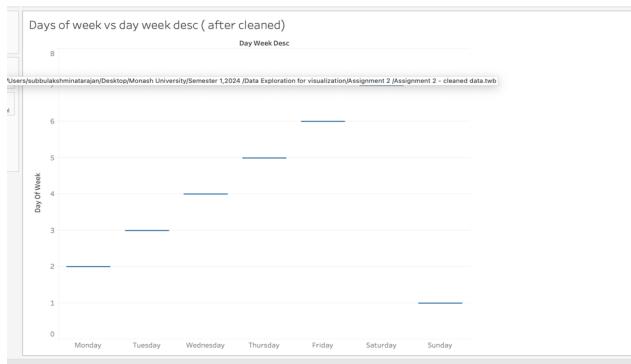
### Step 1: Numerical order starts with 0 so we will be filtering with each day of the week and removing the 0.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
T2012000051	1/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.37474	142.023041				
T2012000054	4/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	777	-38.4536	140.214629				
T2012000055	5/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	777	-38.4536	140.214629				
T2012000073	3/1/2012	17:20:00	1	Tuesday	1	Collision with vehicle	3 Day	DARIBIN	MELB. URBAN	2 T-intersection	999	-37.72050	144.991546				
T2012000076	5/1/2012	15:30:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000079	1/1/2012	20:10:00	1	Thursday	3	Struck pedestrian	3 Day	RURAL VICTORIA	RURAL VICTORIA	5 Not at intersection	999	-37.8446	147.055228				
T2012000081	1/1/2012	05:30:00	1	Sunday	4	Collision with vehicle	3 Day	CARLTON	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000083	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000086	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000087	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000088	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000089	1/1/2012	11:00:00	1	Sunday	4	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	2 T-intersection	999	-37.8078	144.957074				
T2012000094	3/1/2012	18:00:00	1	Tuesday	6	Vehicle overturned (no collision)	2 Day	HEPHERN	TOWNS	5 Not at intersection	999	-37.8078	144.957074				
T2012000095	1/1/2012	15:00:00	1	Sunday	4	Collision with fixed object	3 Day	WAIS COAST	TOWNS	5 Not at intersection	999	-38.4536	140.214629				
T2012000097	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	BRUNSWICK	MELB. URBAN	2 T-intersection	999	-37.73050	144.782616				
T2012000098	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	MONASH	MELB. URBAN	5 Not at intersection	999	-37.66234	145.000000				
T2012000099	4/1/2012	8:15:00	1	Saturday	4	Collision with fixed object	2 Day	MELBOURNE	MELB. URBAN	5 Not at intersection	999	-37.8078	144.957074				
T2012000100	1/1/2012	15:30:00	1	Sunday	1	Collision with vehicle	3 Day	MELBOURNE	MELB. URBAN	1 Cross intersection	999	-37.8078	144.957074				

### Step 2: Replace 0 to 1 for Sunday to make it a numerical order.

**Step 3:** You can see that from Monday onwards there is Tuesday which is in numerical order, so I adjusted accordingly for the following week so that it doesn't have irregularities.

Data Cleaned after this graph which shows clearly:



### 3) Removing brackets in the urban name

**Step 1:** Remove the bracket for Fall Creek, Lake Mountain, Mount Baw Baw, Mount Buller, and Mount Hotham.

**Step 2: Clean the bracket.**

### 3) Spelling error for Baw Baw in the column LGA\_NAME

To clean I used Excel for it.

**Step 1:** You can see Baw Baw filtered in the dataset.

ACCIDENT_NO	ACCIDENT_DATE	DAY_OF_WEEK	ACCIDENT_TYPE_DESC	ACCIDENT_TYPE_CODE	SEVERITY_CODE	LGA_NAME	DEG_URBAN_NAME	ROAD_ID	ROAD_ZEOMETRIC_CODE	SPEED_ZONE	LATITUDE	LONGITUDE
1010000009	1/1/2012	2:25:00	1 Sunday	4 Collision with a fixed object	3 Dark No stree	BAW BAW	RURAL_VICTORIA			100	-38.254857	145.79579
1010000010	2/1/2012	16:00:00	1 Monday	4 Collision with a vehicle	2 Day	BAW BAW	RURAL_VICTORIA			100	-38.02512	145.90606
1010000011	4/1/2012	11:00:00	4 Wednesday	4 Collision with a fixed object	2 Day	BAW BAW	RURAL_VICTORIA			100	-38.02512	145.90606
1010000014	7/1/2012	4:50:00	7 Saturday	4 Collision with a fixed object	3 Dark Street	BAW BAW	RURAL_VICTORIA			100	-38.254857	145.79579
1010000047	7/1/2012	12:00:00	7 Saturday	8 No collision and no object struck	3 Day	BAW BAW	RURAL_VICTORIA			100	-38.17988	145.90254
1010000158	21/1/2012	11:00:00	7 Saturday	4 Collision with a fixed object	2 Day	BAW BAW	RURAL_VICTORIA			100	-38.02512	145.90606
1010000160	21/1/2012	11:00:00	7 Saturday	4 Collision with a fixed object	2 Day	BAW BAW	RURAL_VICTORIA			100	-38.02512	145.90606
1010000169	26/1/2012	13:00:00	5 Thursday	6 Vehicle overtaking/colliding	2 Day	BAW BAW	RURAL_VICTORIA			100	-38.02512	145.90606
1010000205	27/1/2012	9:00:00	6 Friday	4 Collision with a fixed object	3 Day	BAW BAW	RURAL_VICTORIA			100	-38.261134	145.71991
1010000206	28/1/2012	9:00:00	6 Friday	4 Collision with a fixed object	3 Day	BAW BAW	RURAL_VICTORIA			100	-38.261134	145.71991
1010000207	4/2/2012	15:00:00	7 Saturday	5 Collision with some other object	3 Day	BAW BAW	RURAL_VICTORIA			100	-38.261134	145.71991
1010000337	9/2/2012	2:27:00	6 Friday	4 Collision with a fixed object	3 Dark No stree	BAW BAW	RURAL_VICTORIA			100	-38.262021	145.89646
1010000892	11/2/2012	13:00:00	7 Saturday	8 No collision and no object struck	3 Day	BAW BAW	RURAL_VICTORIA			100	-37.859193	145.423751

**Step 2:** Now change Mount Baw Baw in the dataset.

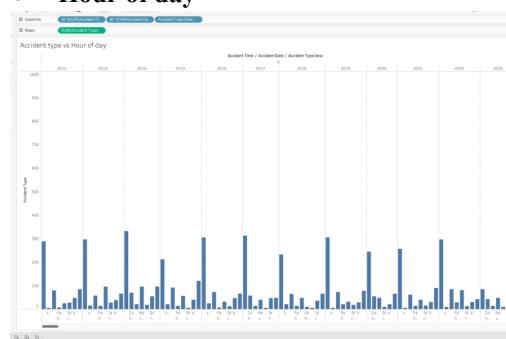
ACCIDENT_NO	ACCIDENT_DATE	DAY_OF_WEEK	ACCIDENT_TYPE_DESC	ACCIDENT_TYPE_CODE	SEVERITY_CODE	LGA_NAME	DEG_URBAN_NAME
1012000009	1/1/2012	2:25:00	1 Sunday	4 Collision with a fixed object	3 Dark No stree	BAW BAW	RURAL_VICTORIA
1012000010	2/1/2012	16:00:00	1 Monday	1 Collision with vehicle	2 Day	BAW BAW	RURAL_VICTORIA
1012000016	4/1/2012	9:10:00	4 Wednesday	4 Collision with a fixed object	2 Day	BAW BAW	RURAL_VICTORIA
10120000414	7/1/2012	4:50:00	7 Saturday	4 Collision with a fixed object	3 Dark Street	BAW BAW	SMALL_CITIES
10120000447	7/1/2012	12:00:00	7 Saturday	8 No collision and no object struck	3 Day	BAW BAW	RURAL_VICTORIA
1012000156	21/1/2012	11:20:00	7 Saturday	4 Collision with a fixed object	2 Day	BAW BAW	RURAL_VICTORIA
1012000186	24/1/2012	0:00:00	2 Tuesday	5 collision with some other object	3 Unk.	BAW BAW	RURAL_VICTORIA

From this screenshot, we can see that we have removed all the errors from the column from LGA\_name to have clean data.

### 2. Data exploration (i.e., Step 4)

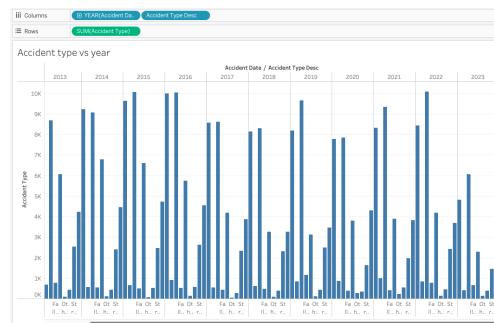
Q1.

#### • Hour of day



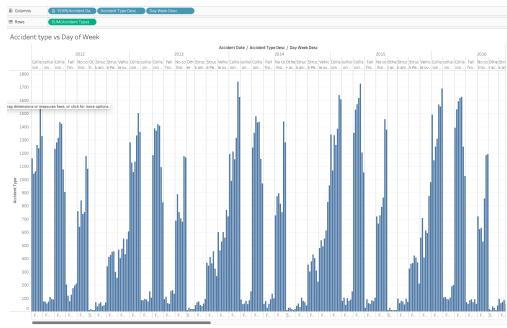
Each year the accident type for each year fluctuated according to accident type in each year. It has been said that the year said to me the highest according to other years and accident types. When it comes to accident type it can be said that vehicles have been recorded the highest. But this data clearly says that we cannot see each hour how many accident types have happened in each year that has been recorded.

#### • Year



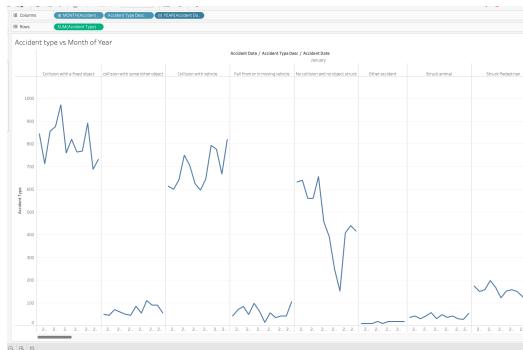
Each year the accident type for each year fluctuated according to accident type in each year. It has been said that the year said to me the highest according to other years and accident types. When it comes to accident type it can be said that vehicles have been recorded the highest.

#### • Day of week



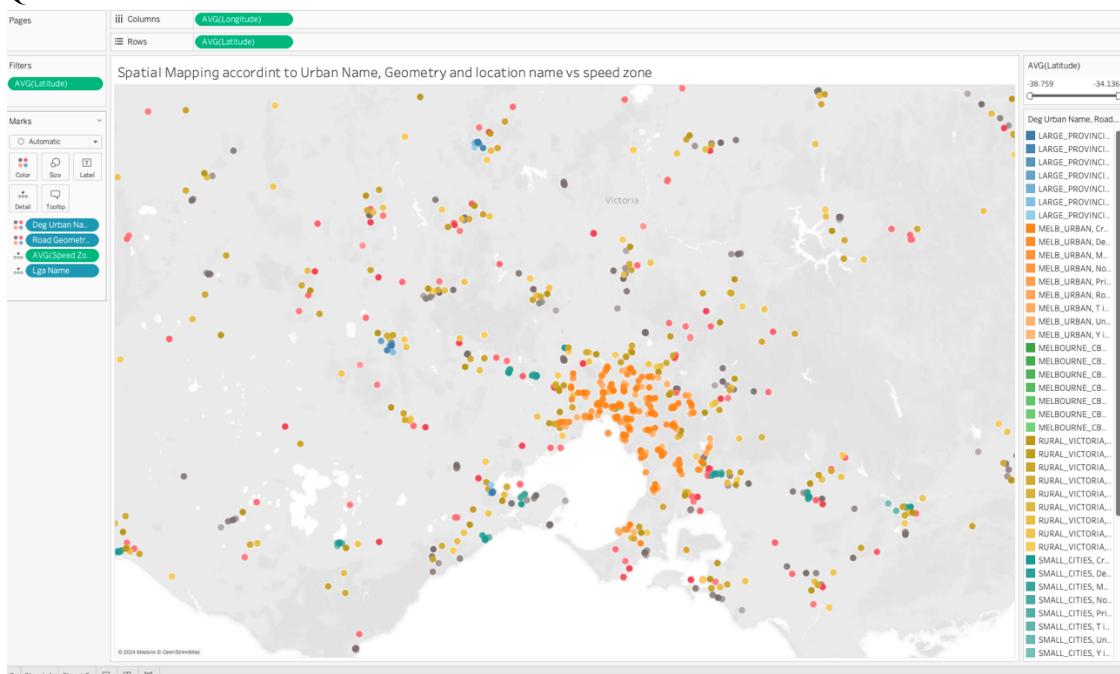
Each week in the years it says the same as above that year 2022 is the same.

## ● Month of Year



Each month of year by accident type it can be said by the scatter plot we can fluctuation of the graph due to increase and decrease of the accident.

Q2.



From the spatial graph, we can clearly identify the speed zone in each region and cities that can be seen clearly in this graph.

Challenge:

We cannot identify due to the data quality of the dataset we cannot specify the road support for this map.