

**EMERGENCY MANAGEMENT SIT764, T2 2019 DEAKIN UNIVERSITY, BURWOOD**

HAND OVER DOCUMENT – “AMBULANCE DATA”

**Contents**

[Project Description￼ ￼](bookmark://_Toc21019459#_Toc21019459)

[Initial Requirements:￼ ￼](bookmark://_Toc21019460#_Toc21019460)

[Milestone objectives on Project Initiation:￼ ￼](bookmark://_Toc21019461#_Toc21019461)

[What was agreed with Customer?￼ ￼](bookmark://_Toc21019462#_Toc21019462)

[Team Member Details￼ ￼](bookmark://_Toc21019463#_Toc21019463)

[Technologies and Tools￼ ￼](bookmark://_Toc21019464#_Toc21019464)

[Development process￼ ￼](bookmark://_Toc21019465#_Toc21019465)

[1.Analysing client requirement and analysis of suitability of client provided data for the Project completion:￼ ￼](bookmark://_Toc21019466#_Toc21019466)

[2.Data Collection for the features:￼ ￼](bookmark://_Toc21019467#_Toc21019467)

[3.Data Cleansing:￼ 7](bookmark://_Toc21019469#_Toc21019469)

[4.Data Visualisation:￼ 7](bookmark://_Toc21019470#_Toc21019470)

[SOURCE CODE ￼ 11](bookmark://_Toc21019474#_Toc21019474)

[Machine Learning Codes￼ 1￼](bookmark://_Toc21019475#_Toc21019475)

[R Code for Data Collection, Cleansing and Visualization￼ 1￼](bookmark://_Toc21019476#_Toc21019476)

[Datasets￼ 1￼](bookmark://_Toc21019477#_Toc21019477)

[KNOWN DEFECTS￼ 1￼](bookmark://_Toc21019478#_Toc21019478)

[INCIDENT AND PROBLEM RECORDS￼ 11](bookmark://_Toc21019479#_Toc21019479)

# Project Description

Emergency management tribe is a Data science team which Identify seasonal variations of calls, patterns of call locations, correlations between characteristics of local populations (age groups, gender, occupation) and ambulance services used. Identify and download suitable datasets and make precise software and data requirements and development plan. Explore existing work in the literature in this area. Complete Ambulance Victoria analysis and use it to predict response time and other performance indicators in the next three years. Build data analytics and machine learning algorithms for Ambulance Victoria and visualise the emerging patterns.

Project involves entire data science life cycle – Data Collection, Data Cleansing, Data Aggregation, Data Visualisation, Model Building and Forecasting based on Data Analysis.

## Initial Client Requirements:

A) Identify patterns of

* Call locations
* Age groups
* Gender
* Occupation
* Ambulance services used

1. Find Correlation among features mentioned above
2. Create Machine Learning algorithms for Ambulance Victoria
3. predict response time and other performance indicators in the next 3

## Milestone objectives on Project Initiation:

***22nd July 2019 to 11th August 2019*** -: Identify and download suitable datasets and make clear software and data requirements and development plan. Explore existing work in the literature in this area.

***12th August 2019 to 25th August 2019*** -: Identify appropriate data analysis and machine learning packages/algorithms for the project.

***26th August 2019 to 15th September 2019*** -: Build data analytics and machine learning algorithms for Ambulance Victoria and visualise the emerging patterns.

***16th September 2019 to 6th October 2019*** -: Complete Ambulance Victoria analysis and use it to predict response time and other performance indicators in the next three years. Write a detailed technical report and make a demo for the presentation.

## What was agreed with Customer?

Client provided data is available in the following link : <https://data.gov.au/search?q=ambulance%20victoria>

. On further analysis during Iteration 1(from ***22nd July 2019 to 11th August)***, this data was identified as not sufficient to meet the client requirements. Team did research and data collection to identify suitable datasets meeting the client requirements. [www.data.gov.au](http://www.data.gov.au)

On analysing the available data and discussing with technical advisors, it was identified that finding correlation among features and building a model including all the identified features was beyond the scope of the project. On further discussion with the client,the following requirements were finalised:

1. Collect data from a website, cleanse it and convert it to a suitable format to perform data analysis. Data is to be obtained for the below features:

* Call locations
* Age groups
* Gender
* Occupation
* Ambulance services used

1. Find Correlation among features mentioned above.
2. Create Machine Learning algorithms for Ambulance Victoria
3. predict response time and other performance indicators in the next 3

Below Timelines were agreed upon discussion with Client:

***12th August 2019 to 25th August 2019*** -: Identify suitable datasets and make clear software and data requirements and development plan.

***26th August 2019 to 8th September 2019*** -: Perform data cleansing and conversion to a suitable format for analysis.

***9th September 2019*** ***to 22nd September 2019:*** -: Identify appropriate data analysis and machine learning packages/algorithms for the project.

***23rd September 2019 to 6th October 2019*** -: prediction for three months and Technical report presentation.

Remaining requirements which were present in the initial client requirement was agreed to be completed in the next stage of the project:

* Develop Machine Learning Models.
* Visualize the emerging patterns.

# Team Member Details

***Squad Supervisor:*** ***Maia Angelova Turkedjieva***

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Role** | **Phone Number** | **Email ID** |
| Nikita | Team Lead   * Machine learning model * Prediction & Analysis | 0414612037 | [ndighe@deakin.edu.au](mailto:ndighe@deakin.edu.au) |
| Pavan Kumar Vadla | Deputy team Lead   * Data sourcing * Visualization * Hand over documentation | 0451225828 | [pvadla@deakin.edu.au](mailto:pvadla@deakin.edu.au) |
| Vijay bathini | * Data sourcing * Cleaning | 0433884945 | [vbathini@deakin.edu.au](mailto:vbathini@deakin.edu.au) |
| Varsha Sankar | * Data Pre-processing, Data Visualization | 0433953196 | [vsankar@deakin.edu.au](mailto:vsankar@deakin.edu.au) |
| Piyush Gandhi | * Machine learning model * Prediction & Analysis * Visualization * Documentation | 0402041560 | [gandhipi@deakin.edu.au](mailto:gandhipi@deakin.edu.au) |
| Bala Tharun Reddy Shyamala | * Data Sourcing * Data Cleaning | 0490357774 | [bshyamala@deakin.edu.au](mailto:bshyamala@deakin.edu.au) |
| Nikhil Reddy Viraati | * Data sourcing * Data Visualization | 0452426575 | [nviraati@deakin.edu.au](mailto:nviraati@deakin.edu.au) |
| Aiswarya Subrahmanian | * Machine learning model * Prediction & Analysis | 0432247154 | [Asubrahmanian@deakin.edu.au](mailto:Asubrahmanian@deakin.edu.au) |
| Simran Rashpal Dhillon | * Data sourcing * Visualization | 0410425500 | [Srdhillon@deakin.edu.au](mailto:Srdhillon@deakin.edu.au) |
|  |  |  |  |

# Technologies and Tools

* Tableau and Python for visualisation
* R – “**tidyverse**” – package for Data cleansing
* Python – Machine Learning model building and Data cleaning.
* Html –
* JavaScript –

# Development process

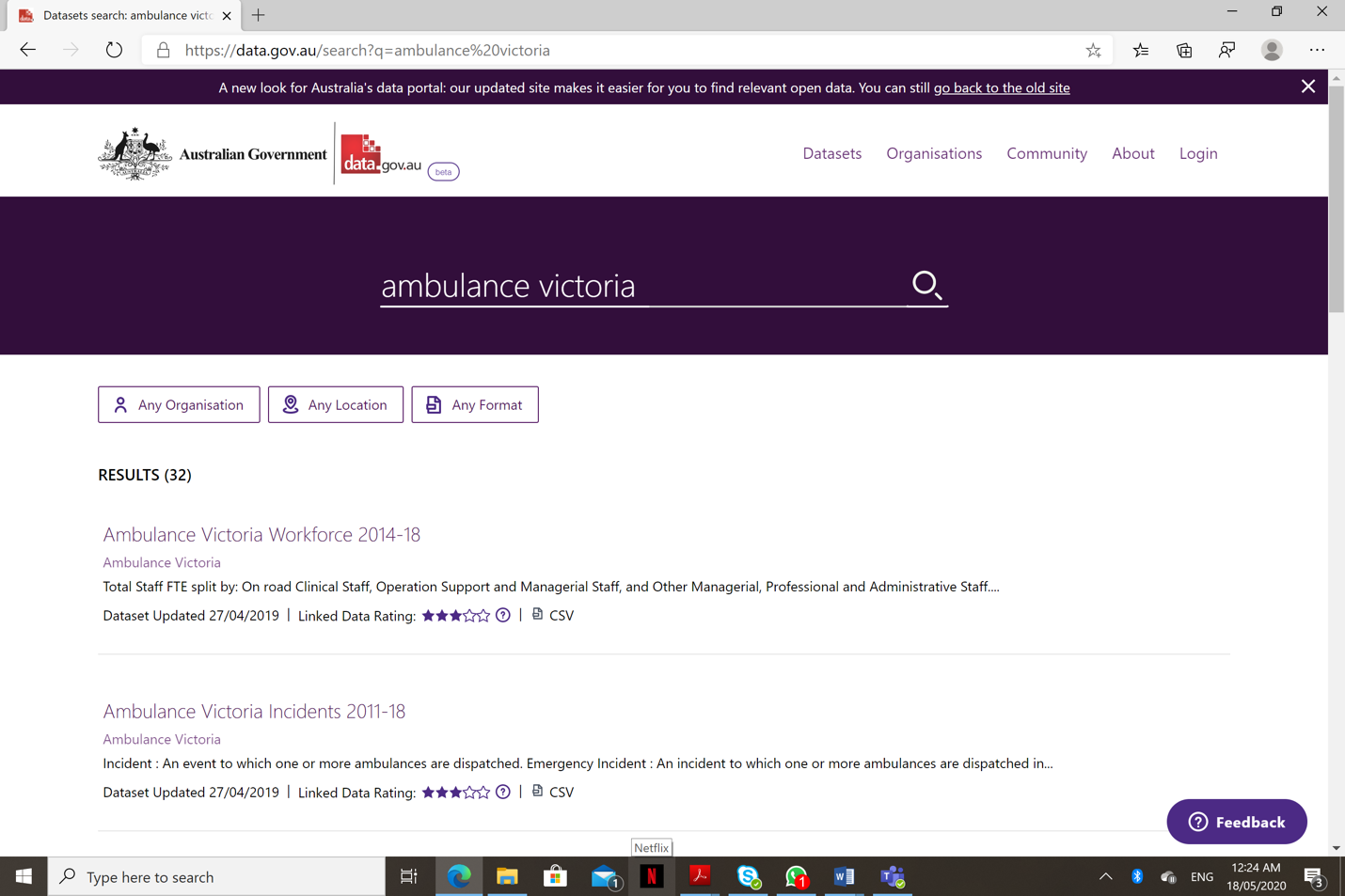
Project Involves all the different stages in Data Science Lifecycle that is data collection, data cleansing, data aggregation, data visualisation, model building and prediction based on analysis. As it was agreed with the client, in the first phase of the project, the team completed visualisation and pattern analysis based on each feature.

Along with this team also started with prediction based on the analysis model built using Logistic Regression.

The prediction was made based on predicting the Crash Area Types as well as the Gender Involved in the accident. Prediction based on other features will be completed in the next phase of the project. Also, the next step will be involving building of machine learning models based on each feature.

Various stages in the development lifecycle of this project are as follows:

## 1.Analysing client requirement and analysis of the suitability of client-provided data for the Project completion:

Initial client requirement as, mentioned earlier ,included finding out features of Ambulance Victoria. Suggested features were call locations, age group, gender, ambulance services and other essential characteristics. The team needed to analyse the correlation among the mentioned features and predict rate of accidents using machine learning model built based on the features mentioned. The client had provided following link to a dataset to be used for analysis. <https://data.gov.au/search?q=ambulance%20victoria> 

The data available from this link included ambulance used, this did not have any information based on weather data

On discussion with the client, Sergiy, regarding the dataset being unfit for analysis, it was decided to research and obtain better datasets for analysis.

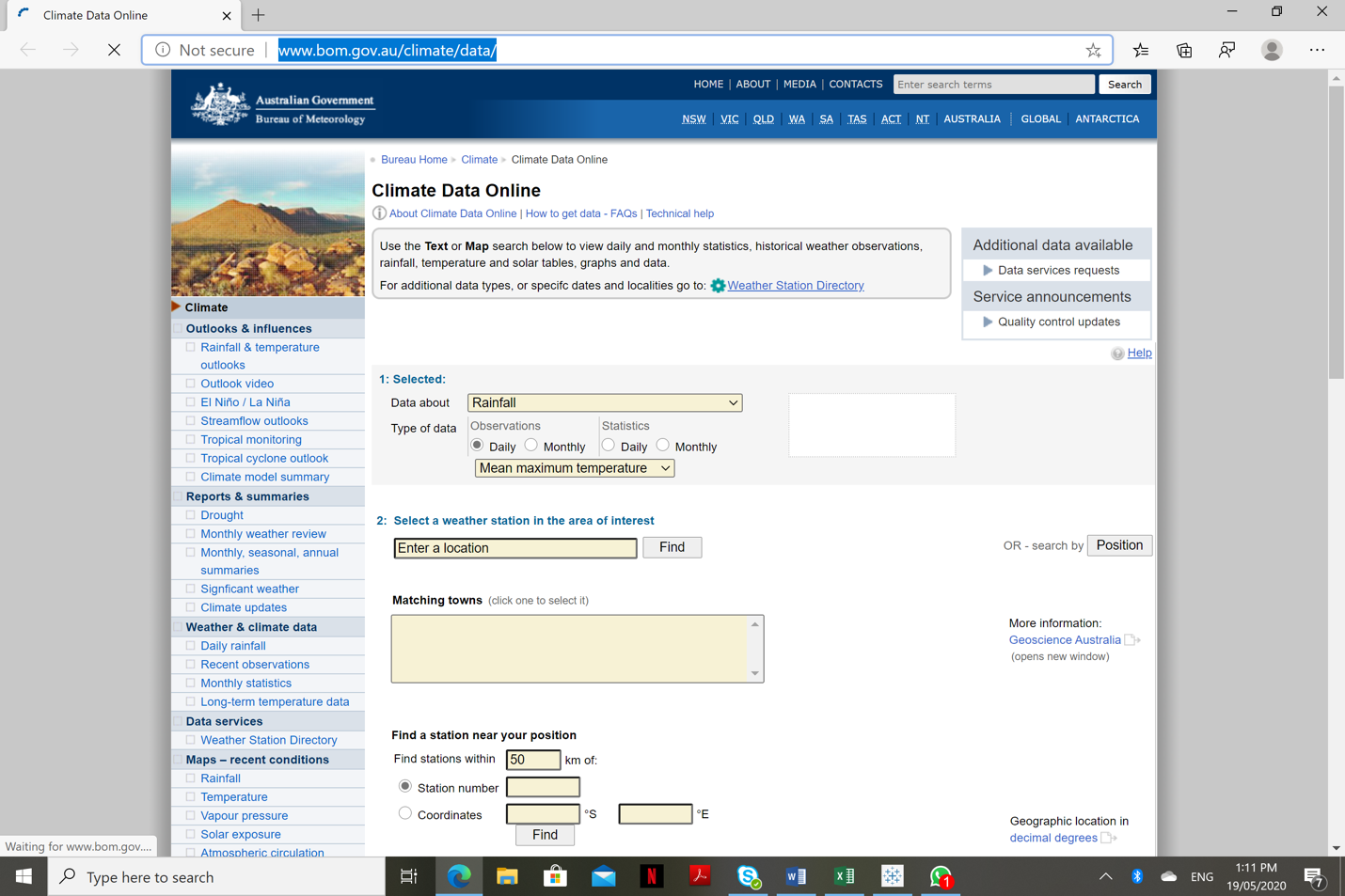
## 

## 2.Data Collection for the features:

After so much research, our team has been able to find all the data required for prediction.

**Ambulance data** has Location, type of code used for the call, response type and other various categories.

**Weather data** has 2015-18 years rainfall data in all different required data which is acquired from the following website: <http://www.bom.gov.au/climate/data/>



## 3.Data Cleansing:

Data downloaded for each of the features were analysed to check for the presence of outliers, null values and invalid fields. Data obtained also had to be converted to suitable formats for further processing. We used python as our primary tool by importing packages of pandas and NumPy to the working directory to clean the data. Our team skipped unnecessary rows and columns which are not useful for our prediction model. Road accident data which was in excel format where combined with Ambulance data of csv file to see relevant columns to drop or skip columns that are not needed. Data cleaning also involved renaming columns with destined names. After this data has been pre-processing the dataset is passed to data visualization

## 4.Data Visualisation:

Data visualisation was performed on the cleaned and formatted datasets. Tableau Software and plotly package in R programming were used in performing interactive visualisation.

Tableau Software provides an option to upload the dashboards generated to Tableau public or private servers and then the HTML code for the same can be generated. This required team to have access to licensed tableau software. Team members obtained student access using Deakin University mail id and used it for dashboard generation and hosting. Below sample, steps are for creating visualisation.

**Step 1:**

Start the Tableau software, just like any other software. You can download a trial version of Tableau from http://www.tableau.com/

A screenshot of a social media post

Description automatically generated

**Fig 1: TABLEAU opening Version**

**Step 2:**

Next,use the ‘Connect’ feature in the left panel to open the required document.

A screenshot of text

Description automatically generated

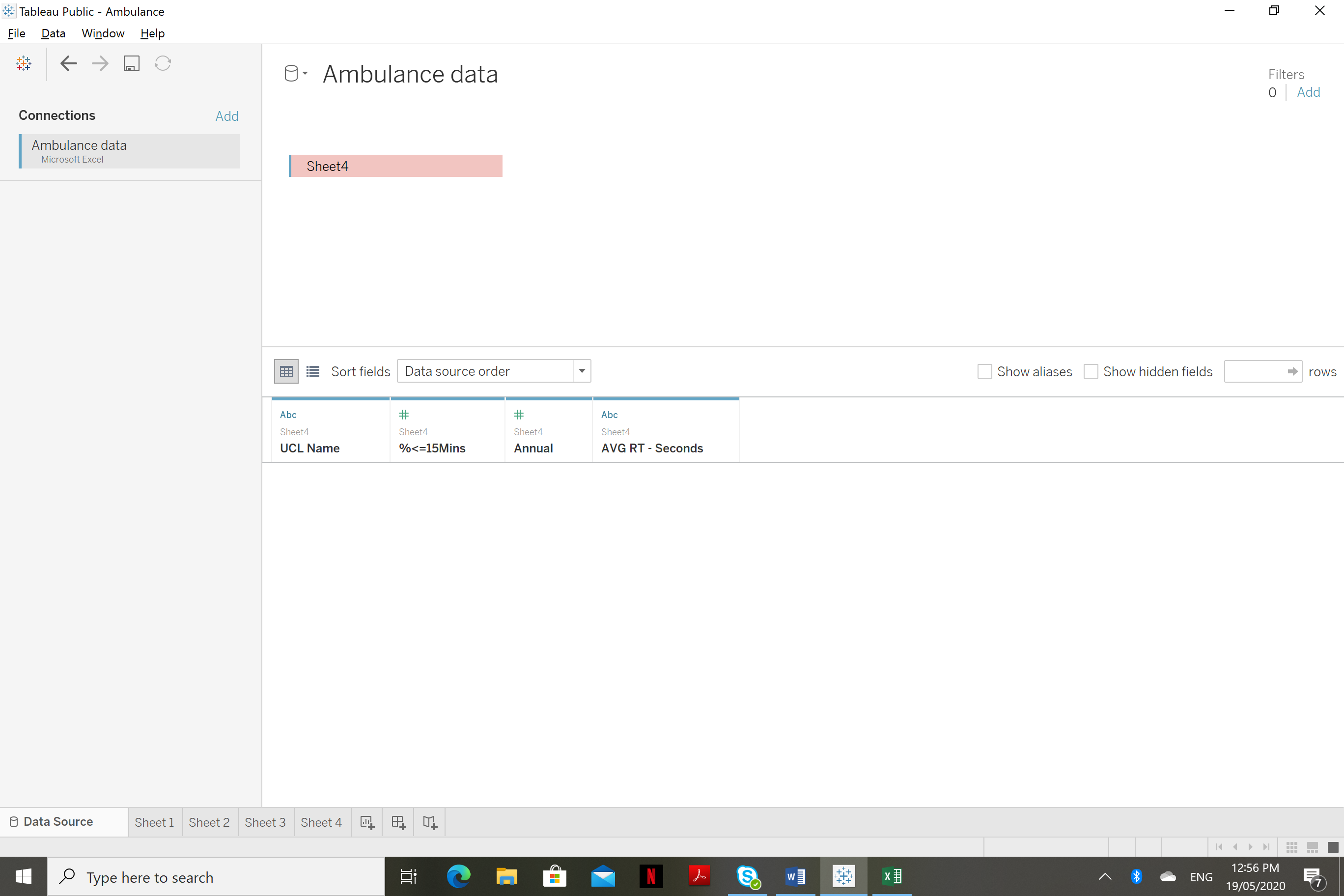
**Fig 2a: Connect to Data**

A screenshot of a cell phone

Description automatically generated

**Fig 2b: Selecting required Excel file**

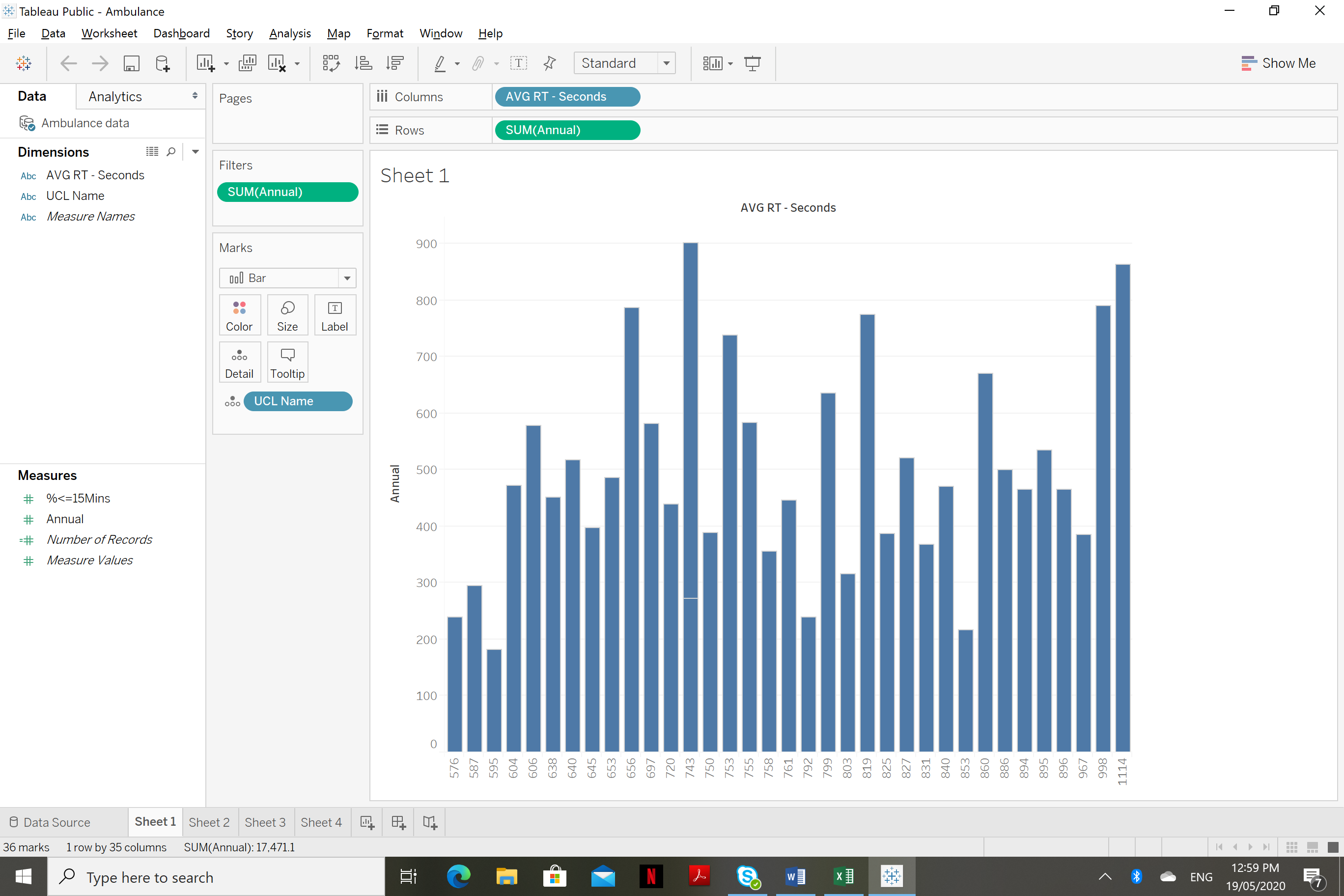
**Step 3:**‘Connect live’. Ensure that the ‘Data’ worksheet is selected and displaying in the Preview, then click the ‘Sheet 1’ tab at the bottom of the screen.



**Fig 3: Tableau connected to our Excel File**

**Step 4:**

We are identifying the labels for the Categorical data and Continuous data.

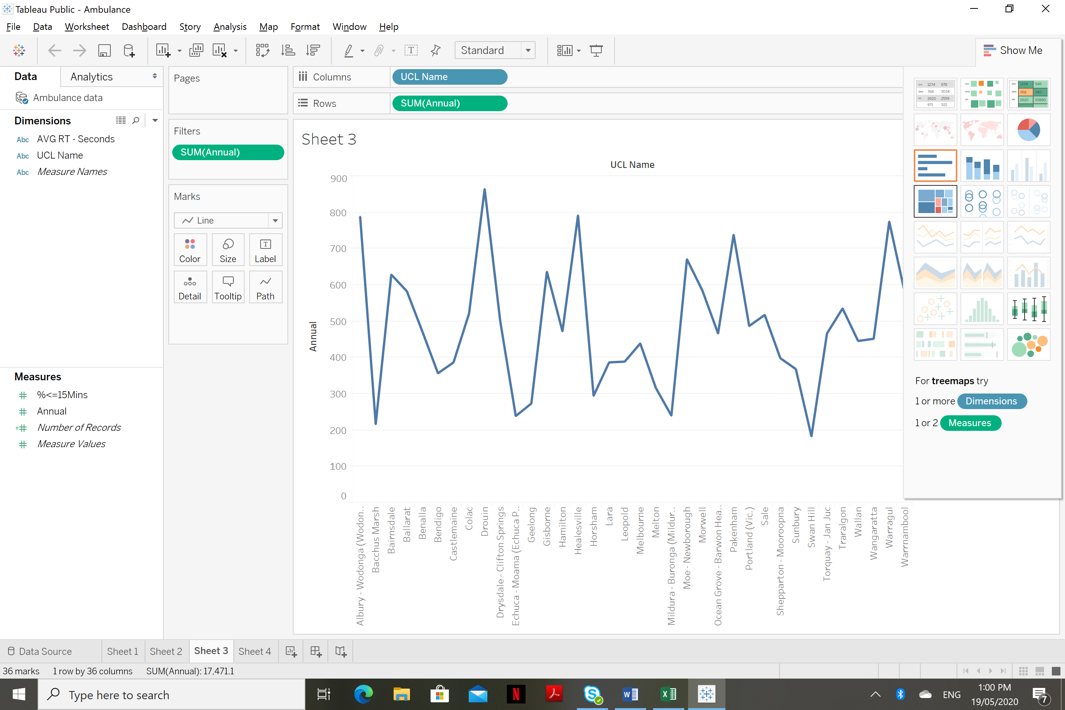


**Fig 4: Tableau work sheet.**

**Step 5**

**Building a tableau workspace**

Begin by placing the ‘UCL Name /Dimension on to the Pages Panel(drag-and-drop)



**Fig 5 : Dragging and Dropping the required variables**

# SOURCE CODE

We have uploaded the codes and datasets to Deakin Sharepoint. Links to the same are as below:

## Machine Learning Codes & Code for Visualization:

[Link to ipynb](https://deakin365.sharepoint.com/sites/Data2IntelligenceConsulting/Shared%20Documents/D2I%20(Emergency%20Management)/Ambulance%20data/Piyush/Analysis_1.ipynb)

## 

## Datasets:

[Link to the Dataset](https://teams.microsoft.com/l/file/79588746-B1E0-4A2E-B45C-6527259772C6?tenantId=d02378ec-1688-46d5-8540-1c28b5f470f6&fileType=xlsx&objectUrl=https%3A%2F%2Fdeakin365.sharepoint.com%2Fsites%2FData2IntelligenceConsulting%2FShared%20Documents%2FD2I%20(Emergency%20Management)%2FAmbulance%20data%2FPiyush%2FRoadAccidents.xlsx&baseUrl=https%3A%2F%2Fdeakin365.sharepoint.com%2Fsites%2FData2IntelligenceConsulting&serviceName=teams&threadId=19:5e24cce19aaa43a4b4f31094a2cc040d@thread.tacv2&groupId=10567a03-e6be-45e6-b917-f7b8920e9c9b)

# KNOWN DEFECTS

Dataset obtained was not sufficient enough to conduct analysis or to build a proper machine learning model involving all the features which were in consideration. It limits the analysis to each feature separately and a hindrance to performing correlation analysis among the independent variables.

# INCIDENT AND PROBLEM RECORDS

None till date.