1. Good morning, I am Rudy. Today, I am going to give a presentation of my project which in the domain of health. The question I am tying to answer is “Could we prevent avoidable death by reducing health risk factors?”
2. Here is the structure of my presentation. I will go through them one by one.
3. First part is the motivation and outcome. There are around 4, 800 people died from avoidable death in Victoria in 2015. It is a large number. And avoidable death can be prevented. Therefore, I want to figure out “Could we prevent avoidable death by reducing health risk factors?”. After data processing and analyzing, my answer is yes. We can prevent avoidable death by reducing health risk factors.

Let me discuss how I did it. First of all, here are the two datasets I used. “LGA\_15\_Health\_Risk\_Factor.csv” and “LGA\_15\_Avoidable\_Death\_by\_Cause.csv”. There are two main reasons for I choosing them. Obviously, first reason is these two datasets have the information I want. And the second one is these two datasets are of the same area at the same period of time which means it is easy to do integration on them.

1. Here are the data wrangling methodologies I used during data processing. For data preprocessing, I did data transforming on the raw data. Then, integrate them on the primary key of both datasets. I deal with missing values by deleting parts of missing values as well as imputation with mean value. Then, I detect outliers by boxplot. I also add two new values, which are total population has health risk factors per 100 population and total avoidable death per 100,000 population. After that, for data visualization, I mainly used four methods which are scatter plot, heat map, pie chart as well as the clustering map. Finally, I use the Pearson correlation for data analyzing in order to determine correlation between features.
2. There 4 interesting and useful findings. The first finding is about an outlier. By inspecting the boxplot of total number of avoidable death in each area in Victoria in 2015. I found that Northern Grampians has an extremely high number of total avoidable death per 100,000 population of 108.9. This means one person would die from avoidable death in every 100 people in Northern Grampians. It is a very serious health problem. So, I would suggest the government to pay more attention to this area depends on my finding.
3. Then, here is the heatmap of Pearson correlation between each kind of avoidable death and each kind of health risk factor. From this map, we can see that population takes adequate fruit has a negative correlation with each kind of avoidable death. On the other head, population has other health risk factors have a positive correlation with each kind of health risk factors. Thus, I would suggest the government to prevent avoidable death by reducing health risk factor and encouraging people take enough fruit every day.

The third finding is from pie charts. The upper side pie chart is about avoidable death. We can see that cancer causes around 48% of avoidable death, thus, I would suggest the government to encourage citizens to diagnosing cancer more regularly. The pie chart in the bottom is about the health risk factors. We can see the population with obese, risk waist measurement and low-exercise occupy almost 70% of total population has avoidable death. Thus, I would suggest the government to encourage people to do more exercise.

1. The last finding is from the map. Here is the clustering map of total avoidable death in each area in Victoria in 2015. There are three clusters and class 3 has the most total number of avoidable death. Therefore, I would suggest the government to take more action on those area.
2. It is quite challengeable to do data processing. There are three main challenges during my data processing. The biggest challenge is finding good datasets. It takes me quite a lot of time to find appropriate datasets. Then, there are some dead findings during my data analyzing stage and some are very strange. One of them implies that we can prevent avoidable death by increasing overweight population, which is not reasonable. Thus, I drop these finding. Finally, it is not easy to code and spend a lot of time on searching coding skills I need.
3. In conclusion, we can prevent avoidable death by reducing health risk factors. Then, the government should pay more attention to those area which has high total avoidable death.
4. That all of my presentation today, thank you for listening!