**Informal Reflection**

**Tiancheng Kuang**

When I was coding, one difficulty I met is how to define which incidents records can be considered as solved. I back to original data and carefully checked data. Then I found as long as one incident solved, it will be marked ‘Arrest’, ‘Booked’, ‘Cited’, ‘Prosecuted’, etc. Only unsolved incidents be marked as ‘NONE’ or ‘UNFOUNDED’. Thus, I came up with an idea to calculate the resolution rate. If ‘NONE’ or ‘UNFOUNDED’ exist in Resolution column, I mark this incidents records as 0 in a new column. To do that, I created two new columns, one is ‘NONE’ and another is ‘UNFOUNDED’. I used lambda and apply to Resolution column. Then I create a new column, ‘Solved or Not’, by adding ‘NONE’ and UNFOUNDED’. Final step is minus ‘Solved or Not’ by one because if incidents solved, it will be marked 2 and if it unsolved will be marked as 1. Grouping by district and letting sum of ‘Solved or Not’ divided by sum of all incidents, we get the resolution rate.

Our analysis shows there is larger number of incidents records during 18:00 to 20:00. And during 4:00 to 6:00 has the lowest number of incidents. However, there is probability that during 18:00 to 20:00, the most common incidents could be theft or fraud, which are not fatal and dangerous incidents. But during 4:00 to 6:00, although number of incidents is relatively low, they are most dangerous incidents, such as kidnapping, murder and weapon related crimes. Thus I further analyzed the data by grouping time period (1 hour frequency) and types of incidents, calculating the sum of total number of incidents. Now we can carefully analyze which type of incidents are most common in different time periods.

**Qianer Wu**

In the related work section, it is difficult to find the crime rate analysis exactly in San Francisco for the past ten years. Then I changed the approach method, finding crime studies which used similar analysis methods / similar demographic and geographic sample / similar purpose of study. Then I compared several studies with ours and had some conclusions.

**Shenqian Wen**

During our first meeting, we want to evaluate the reaction speed in each police district. But we cannot get reliable and accurate data. Then we turn to evaluate each police district’s ability of resolving incidents by measuring the rate of success resolution. We always talk about data-driven analysis. However, data is a tool, but it is still a problem.