Data Acquisition and Management

Name of the Student

Name of the University

Authors note

# Part A

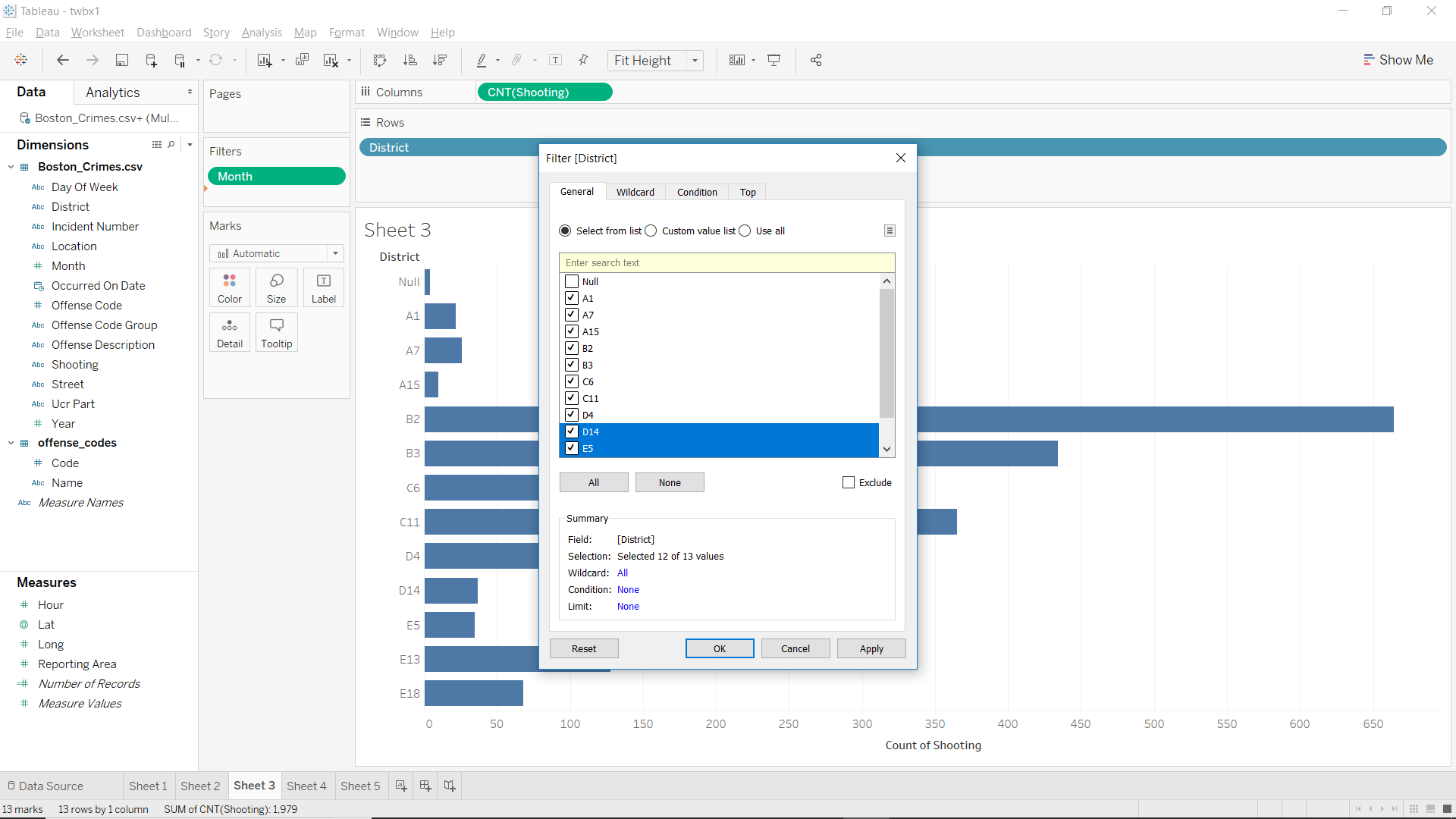
In this course, there are mainly three types of data which includes the structured, semi structured and unstructured data which are collected from the internal or external sources. Internal data can be customer feedback data, product sales data or customer data. Through the use of the tools like the google analytics or Hadoop like tools it is possible to solve different kind of business problems like Shell improved their business revenues by collecting data from seismic waves. From the web too using different kind of scraping tools.

The data life cycle is another important factor in data analysis. while collecting and analyzing data it is important to keep the metadata about the data. In the different phases such as planning, execution and closure phases it is important to have clearly defined roles and responsibilities. There are multiple ways in order to store data such as traditional and dynamic schema.

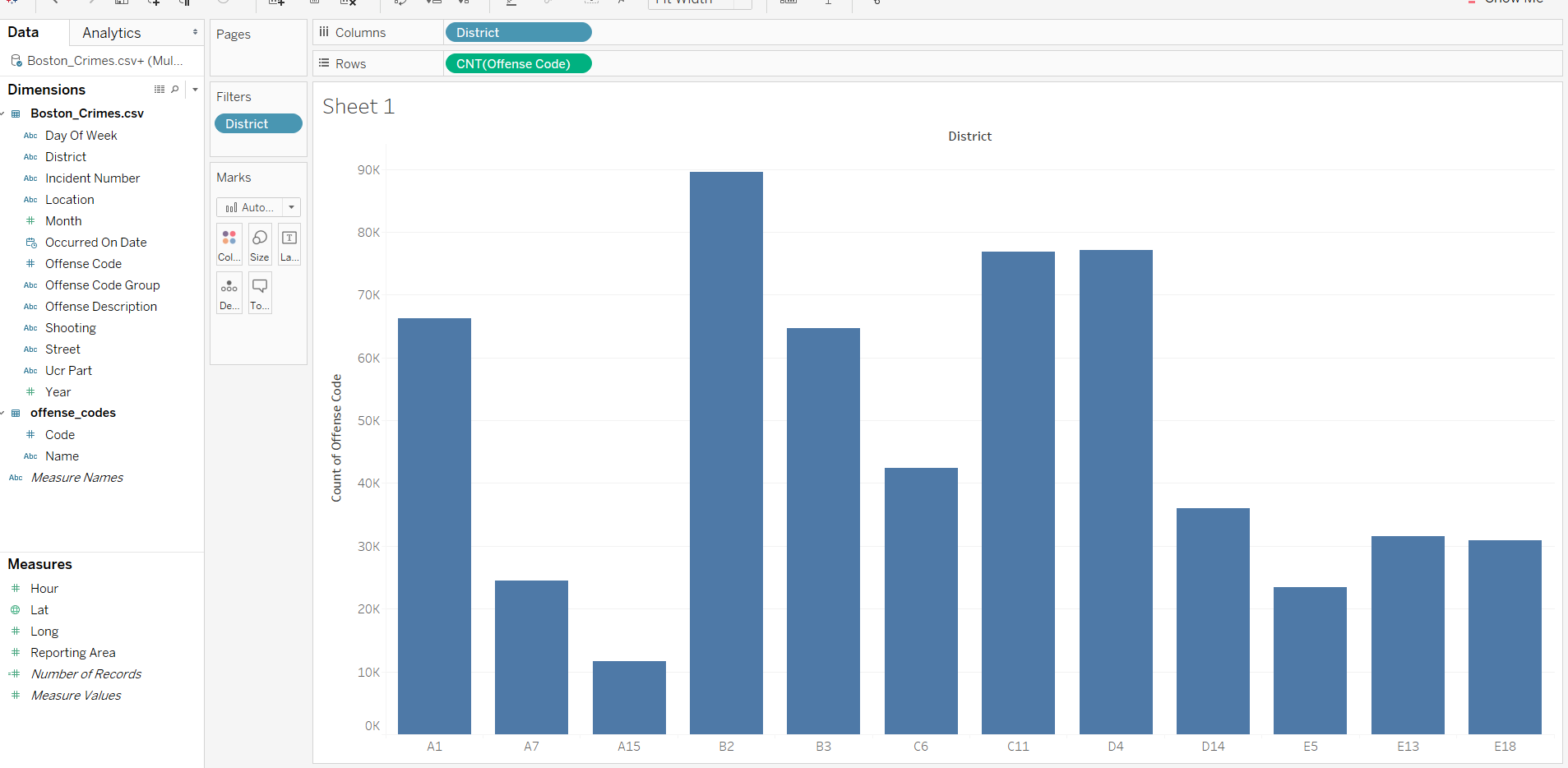
Storage of the collected data is also an important for managing and analyzing data. The storage can be of the following types; database, data lake, data hub, data warehouse and data mart. Furthermore, the data pre processing and data cleansing are also important stages for making the data to be analyzed in order to get the insights from the collected data. Most of the tools such as excel, tableau comes with the filters for analyzing data. With the help of the data dictionary it is also possible to get better sorting and filtering capabilities. Using the SQL, it is possible to Filter, delete and retrieve and update to get the insights.

# Part B

In order to avid the null values or missing values, they are filtered out using the tools available in Tableau.

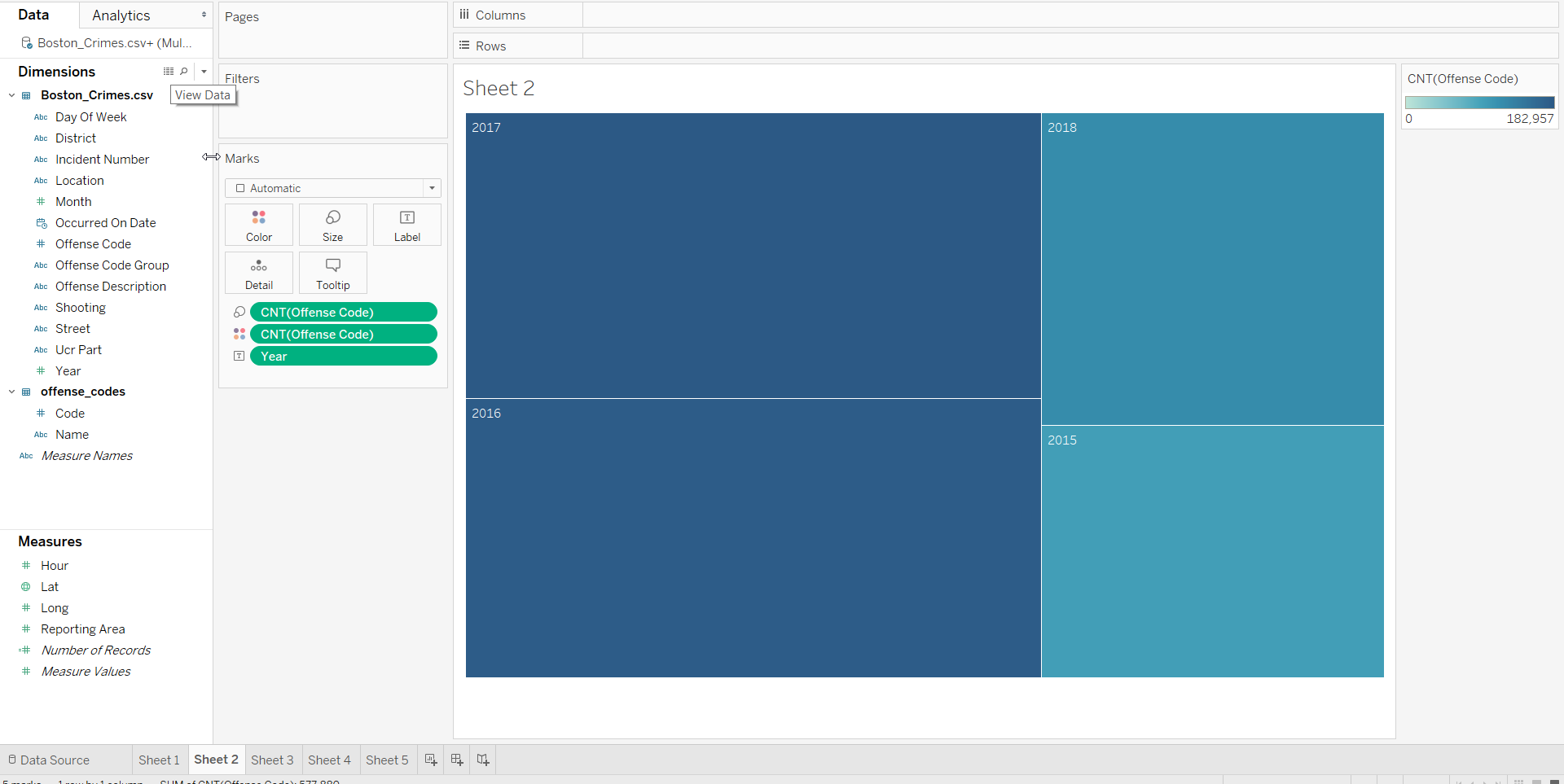


At first the number of the offence codes from different district are plotted using the bar chart.



From the above bar plot, it is clear to see that, B2 district is the district that reported the highest number of committed crimes and offences. Lowest number of crimes /offences are recorded from the A15 district. Other districts that have comparatively high number of committed crimes in the districts A1, D4, C11.

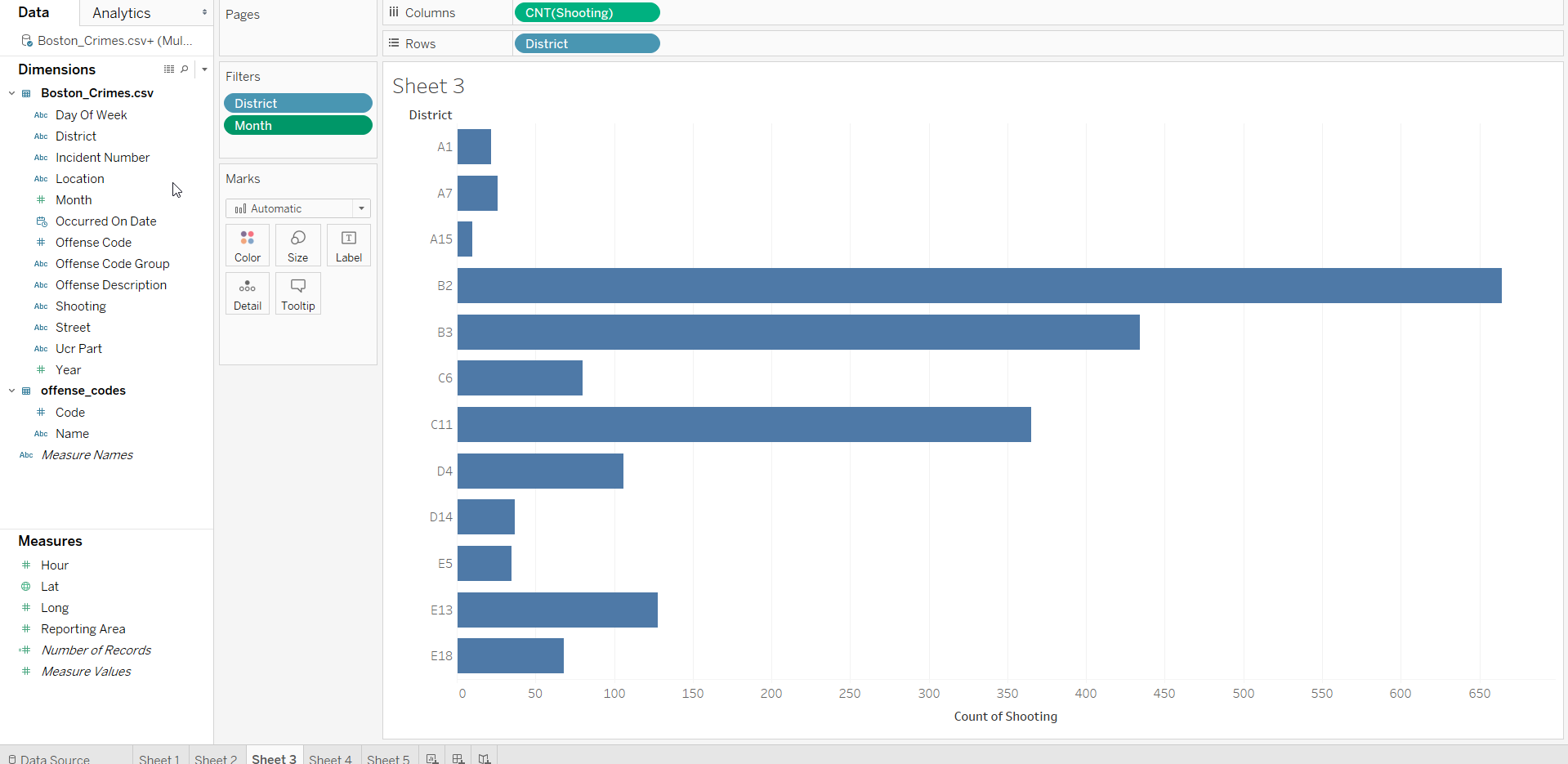
In the next stage the year with the maximum number of committed crimes are plotted using heatmap.



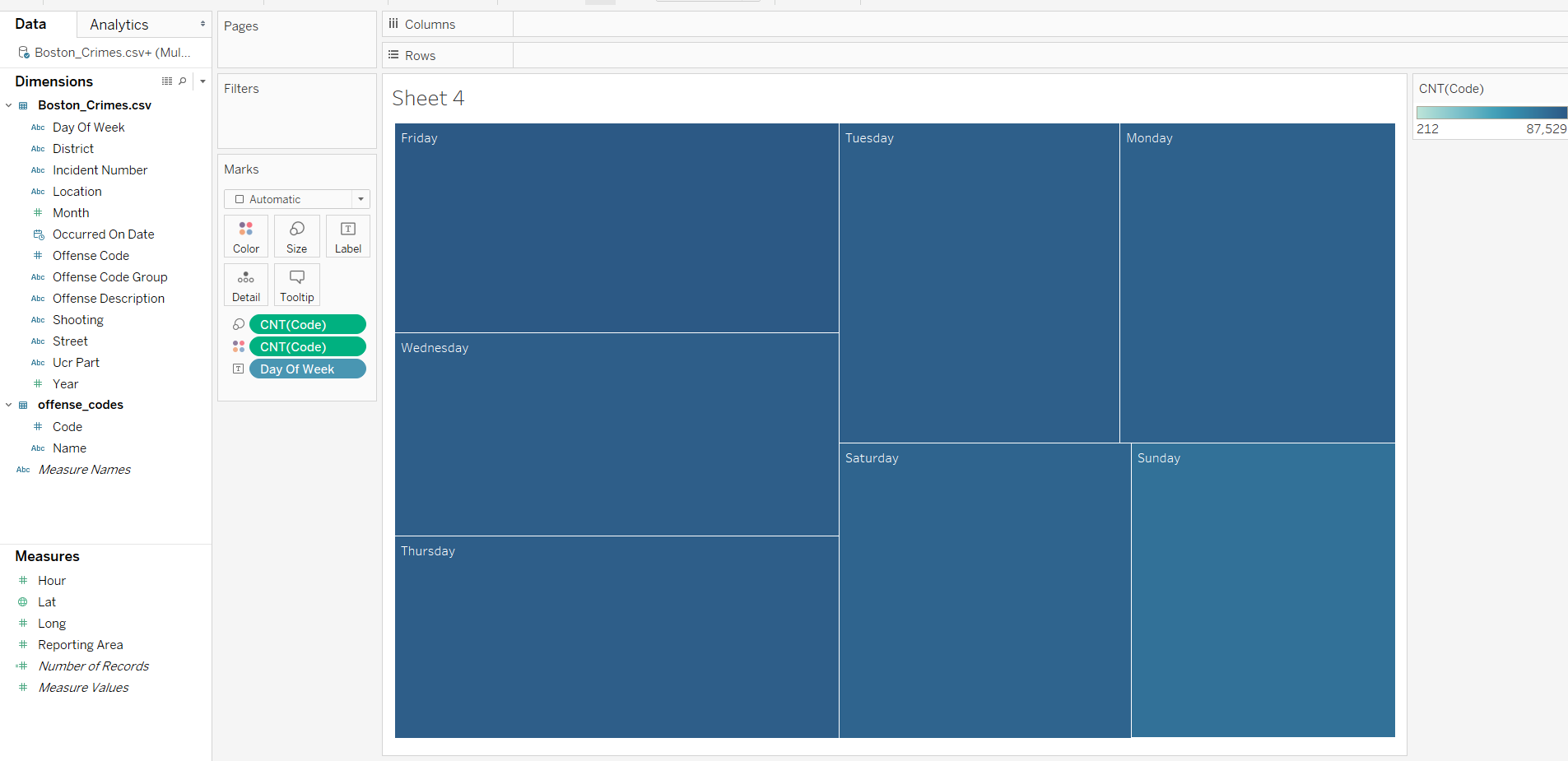
Here, it can be stated that maximum number of crimes are recorded in the year of 2017 and lowest number of crimes are committed in the year of 2015 as depicted in the above heat map plotting.

# Part C

For the next plot the number of the shooting incidents are compared from the different districts are also compared to find out the severity of the crimes in the districts.



Here, it can be stated that the B2, B3 and C11 are the districts that are facing severe crimes that includes shooting in the areas.

Now the most violent days is investigated from the available data sets. In order to do so the heat map is utilized. 

Thus, from the plot it is evident that maximum number crimes are recorded on the Fridays and least number of crimes are recorded for Sunday. Other significant days that recorded higher number of crimes are Wednesday’s and Thursdays.