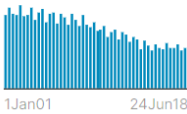# 2. Data Acquisition and Cleaning

2.1 Data Acquisition

The data acquired for this project is a combination of data for one particular location. The data source of the project uses a Chicago crime data that shows the crime per borough in Chicago. The dataset contains the following columns:

- ID
- Case Number
- Date
- Block
- IUCR
- Primary Type
- Description
- Location Description
- Arrest
- Domestic
- Beat
- District
- Ward
- Community Area
- FBI Code
- X Coordinate
- Y Coordinate
- Year
- Updated On
- Latitude

2.2 Data Cleaning

The data preparation is done by taking the 2014-2015 data collected and the major categories of the crime are pivoted to get the total crimes per borough as per the category.

	ID	Case Number	Date	Block	IUCR		Primary Type	
					0820	8%	THEFT	21%
		6635448		59590	0486	8%	BATTERY	18%
		unique values		unique values	Other (400)	84%	Other (33)	61%
			1Jan01 — 24Jun18					

⚠ Whoops, something went wrong loading your data.

	ID	Case Number	Date	Block	IUCR	Primary Type
1	10000092	HY189866	03/18/2015 07:44:00 PM	047XX W OHIO ST	041A	BATTERY
2	10000094	HY190059	03/18/2015 11:00:00 PM	066XX S MARSHFIELD AVE	4625	OTHER OFFENSE
3	10000095	HY190052	03/18/2015 10:45:00 PM	044XX S LAKE PARK AVE	0486	BATTERY
4	10000096	HY190054	03/18/2015 10:30:00 PM	051XX S MICHIGAN AVE	0460	BATTERY
5	10000097	HY189976	03/18/2015 09:00:00 PM	047XX W ADAMS ST	031A	ROBBERY
6	10000098	HY190032	03/18/2015 10:00:00 PM	049XX S DREXEL BLVD	0460	BATTERY
7	10000099	HY190047	03/18/2015 11:00:00 PM	070XX S MORGAN ST	0486	BATTERY

Data will be cleaned in order to find out which area was most affected and how was it reduced in the different years and what type of crime is most likely to occur is also analysed.

	ID	Case Number	Date	Block	IUCR		Primary Type	
					0820	8%	THEFT	21%
					0486	8%	BATTERY	18%
			1Jan01 — 24Jun18		Other (400)	84%	Other (33)	61%