Homework 1

From the NYC collision data answer the following

1. Compute the number of persons who got injured in each borough (10 points)
2. Compute the number of persons who got killed in each borough (10 points)
3. List the top 10 on street location from Bronx from where most of the injuries were reported (20 points)
4. Create a new column "Address" and populate them with combined values of BOROUGH, ZIP CODE, ON STREET NAME with each string separated by a coma. Use Stringer package for this purpose (10 points)
5. Calculate the average of number of people injured according to time in a day and print the top 10 hours of the day where this average was maximum in the Manhattan borough (20 points)
6. How does the number of person injured, and the number of persons killed vary according to each year for all the boroughs? (20 points)

From the Boston crime dataset answer the following (10 points)

	OFFENSE_CODE_GROUP	DISTRICT
1	Residential Burglary	B3
2	Drug Violation	B2
3	Motor Vehicle Accident Response	C11
4	Missing Person Reported	B2
5	Simple Assault	E13
6	Property Lost	*NA*
7	Police Service Incidents	D4
8	Investigate Person	B2
9	Motor Vehicle Accident Response	C11

	B3	B2	C11	E13
Residential Burglary	0	0	0	0
Drug Violation	0	0	0	0
r Vehicle Accident Response	0	0	0	0
Missing Person Reported	0	0	0	0
Simple Assault	0	0	0	0
Property Lost	0	0	0	0
Police Service Incidents	0	0	0	0
Investigate Person	0	0	0	0

Select only **OFENSE_CODE_GROUP** and **DISTRICT** from the Boston crime dataset (see image to the top left). Convert the two columns into a matrix (as shown in the top right image). Make sure that the matrix is populated. For example, if in district **B3**, **Residential Burglary** occurred 50 times then the cell corresponding to **Residential Burglary and B3** in the matrix should be populated as 50. Compute this for all combinations of **OFENSE_CODE_GROUP** and **DISTRICT**. While creating the matrix make sure that *NA*

from **OFENSE_CODE_GROUP** and **DISTRICT** is not included in either matrix rows or columns.