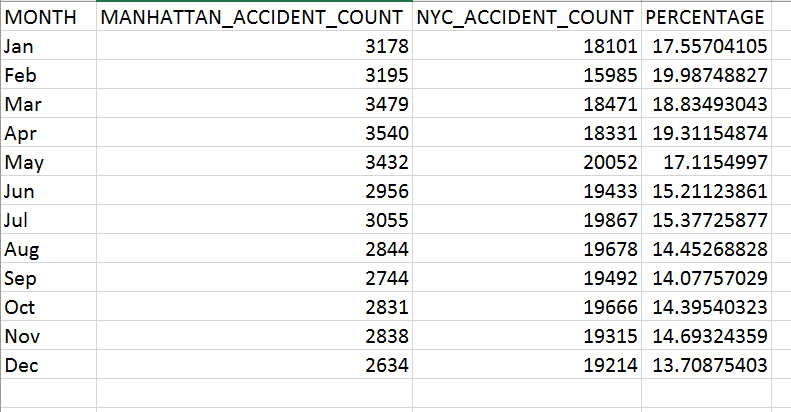
**Assignment 3**

**Question 1 Part 1:**

I have performed analysis on New York Accident Data set and calculated the percentage of accidents occurred for each month of 2016 in MANHATTAN to the overall accidents occurred in New York city.

I have used Strftime() function to extract the month name from the MONTH column and applied groupby() function to group the data based on the BOROUGH names. After performing the analysis I have stored the data to CSV file using to\_csv() function.

**Output:**

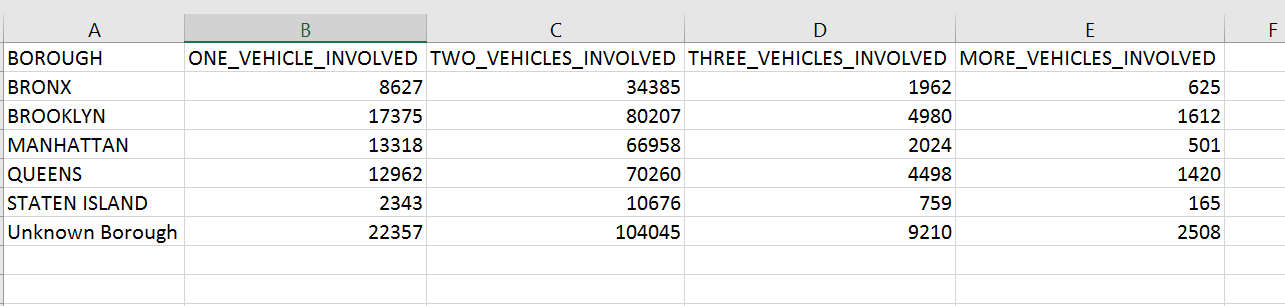


**Question 1 Part 2:**

For part 2 I have performed the analysis on the same data set as of part 1 and calculated the distribution of each accidents scale based on the number of cars involved in an accident for example, 1 car was involved, 2 cars were involved etc.

I have assigned values 1 to the columns ‘Vehicle 1 type’, ‘Vehicle 2 Type’….’Vehicle 5 Type’ where ever the values are not null and for the null values I have assigned values 0 using lambda function. Added those five columns and collected the data into a new column which gives the results as 0,1,2,3,4 or 5. Which signifies the number of cars involved in an accident.

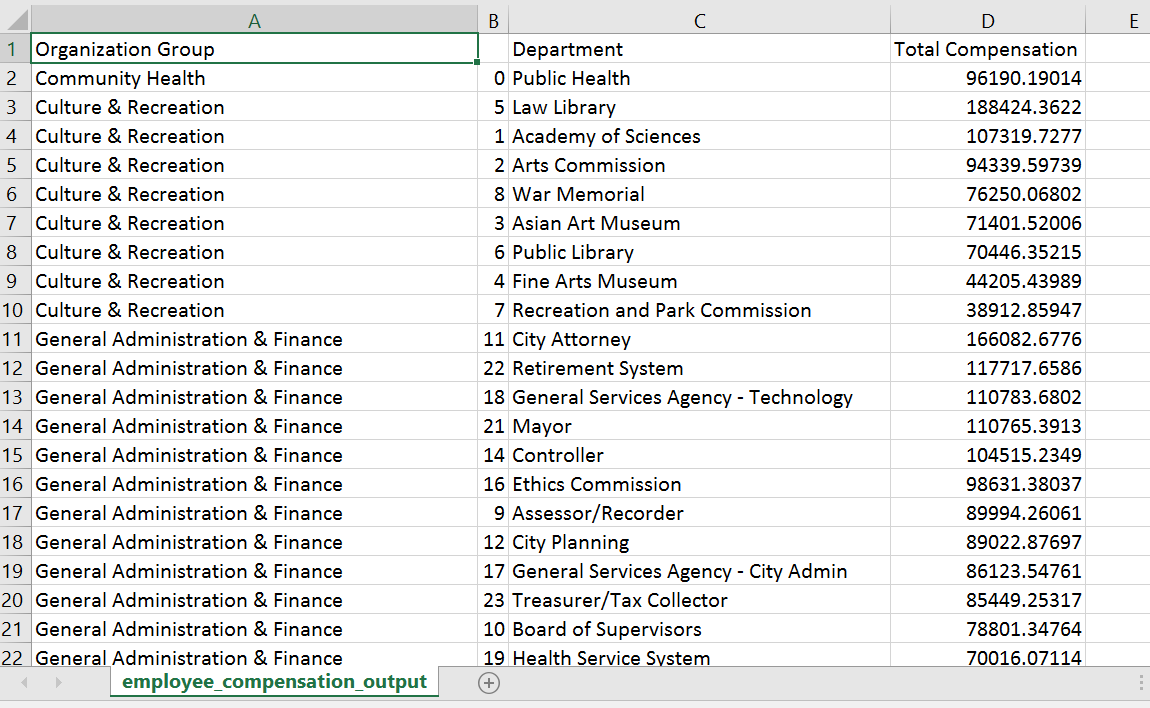
**Output:**



**Question 2 Part 1:**

Performed analysis based on employee compensation data and calculated the highest paid department in each organization. Used groupby() function to group the organization and Department and then calculated the mean of Total compensation columns also sorted the values of mean\_total\_compensation columns in descending order.

**Output:**

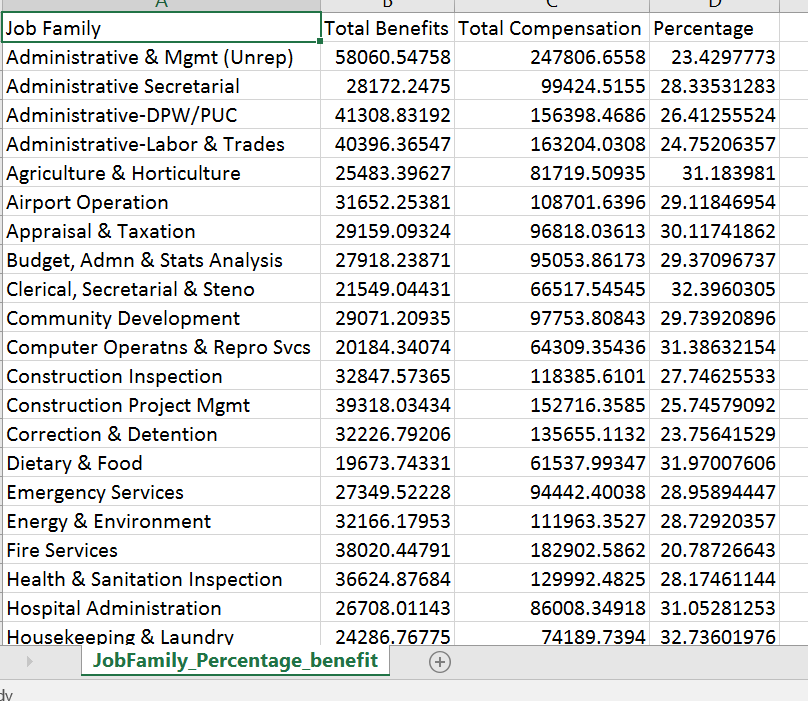


**Question 2 Part 2:**

Performed analysis based on employee compensation data and calculated average salary of each employee for the calendar year type.

Filtered the data frame for the employees whose overtime salary is more than 5% of their salary, also calculated percentage of total benefit with respect to total compensation for each job family

**Output:**



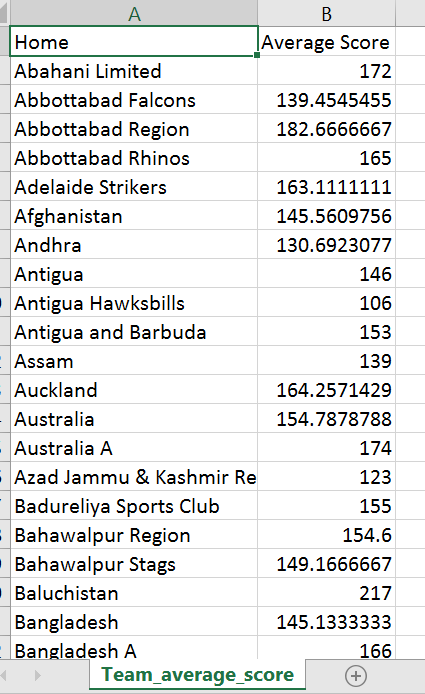
**Question 3 part 1:**

Calculated average scores of teams who hosts and wins the games based on the cricket match dataset.

Applied filter on the data frame to give only those rows where column ‘home’ matches with the column ‘winner’ after that, compared if the column ‘winner’ matched with the column ‘innings1’ or ‘innings2’.

If winner matched with the columns innings1 then added the scores from innings1\_runs else added the scores from innings2\_runs and then stored the added score in a new columns and calculated the average.

**Output:**



**Question 4 Part 1:**

Based on the movie awards dataset performed analysis, extracted the data from the award columns and split it into several columns and then calculated the total awards and total nominations count.

**Output:**

