## Analysis on Police Department dataset containing crimes and their description

**Purpose of Analysis**

It’s seen that with the help of previous data, one can make better decisions in order to maximize the output. Similar is the case with Police Departments also who always want to ensure maximum safety citizens. I am analysis dataset provided by Police Department in order to make sure the safety of citizens. It is 2016 dataset which contains 9 columns as following:

* **incident\_id** - A number assigned to each incident reported.
* **category** - Category of the incident reported
* **crime\_description** - Description explaining the nature of the crime.
* **crime\_date** - date on which the crime was reported.
* **department\_district** - district in which the police department is located.
* **resolution** - Details of resolution (if any).
* **address** - Address where the crime occurred.
* **department\_id** - police department id.
* **location** - lat-long location where the crime was committed.

Thus by using all these features, let us see what an all insights we can extract so that we can help police in safety standards of people. I am making use of python language for the purpose where I will be using pandas library for data analysis , matplotlib and seaborn libraries for plotting graphs and visualizations.

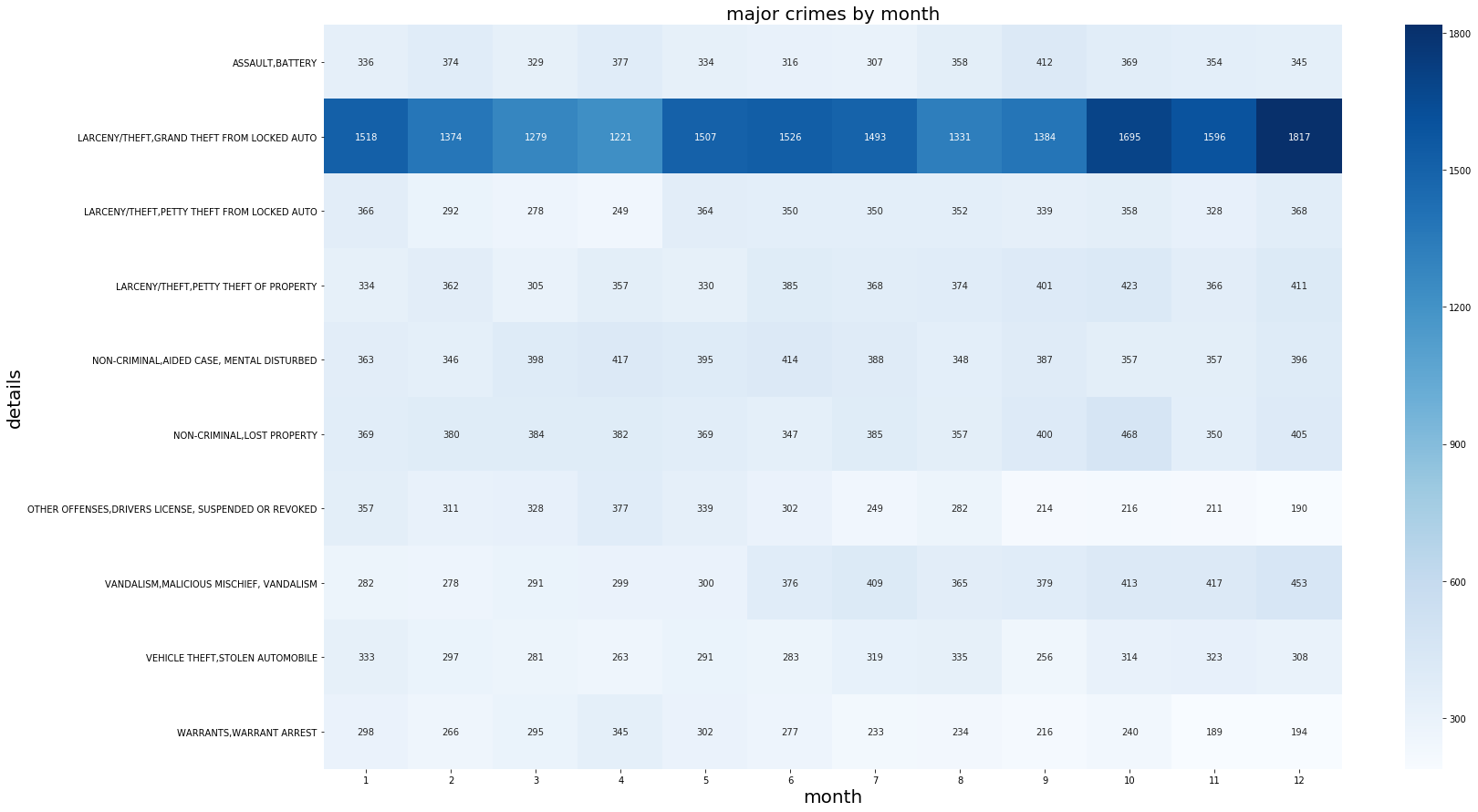
**First Step**

Lets install all necessary libraries that will be used during analysis. After reading the dataset file, there is extraction of “hour”, “year” and “month” from “crime\_date” feature. Moreover there is dropping of “address” column as it is not useful for analysis. Since this is police dataset, there are no null values.

Now using groupby and sorting of “crime\_description” with respect to “incident\_type” has given the total of all crimes separately. It has been seen that **GRAND THEFT FROM LOCKED AUTO** had maximum number of crimes equal to **17741.** And in this category also, there is only one kind of crime associated which is **LARCENY/THEFT**. Thus it becomes equal to 17741.

Combination of “crime\_description” and “category” makes one feature which is “details”. Thus after grouping and “details” and “resolution”, top 10 crimes has been collected under “top\_crimes”.

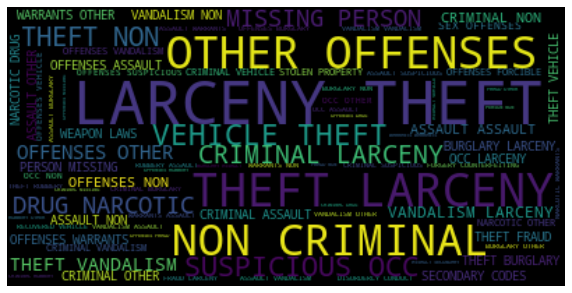
Heatmap of these top 10 crimes with respect to month looks like following:



This clearly shows that **LARCENY/THEFT AND GRAND THEFT FROM LOCKED AUTO**  having the most number of crimes and peak months are **October ,December and January.**

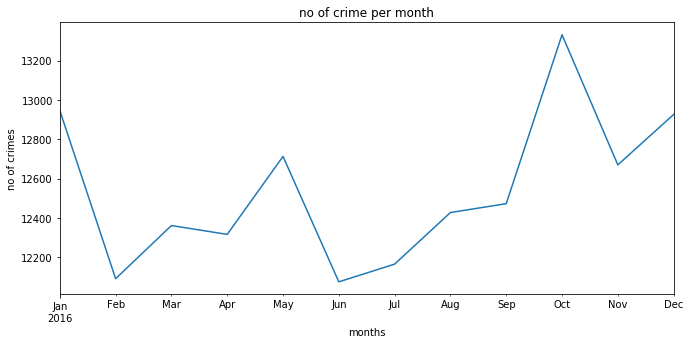
**Second Step**

In this step checking of most frequently occurring crime has been shown by using wordcloud.



It is clearly shown that **LARCENY/THEFT**  has maximum number of crimes.

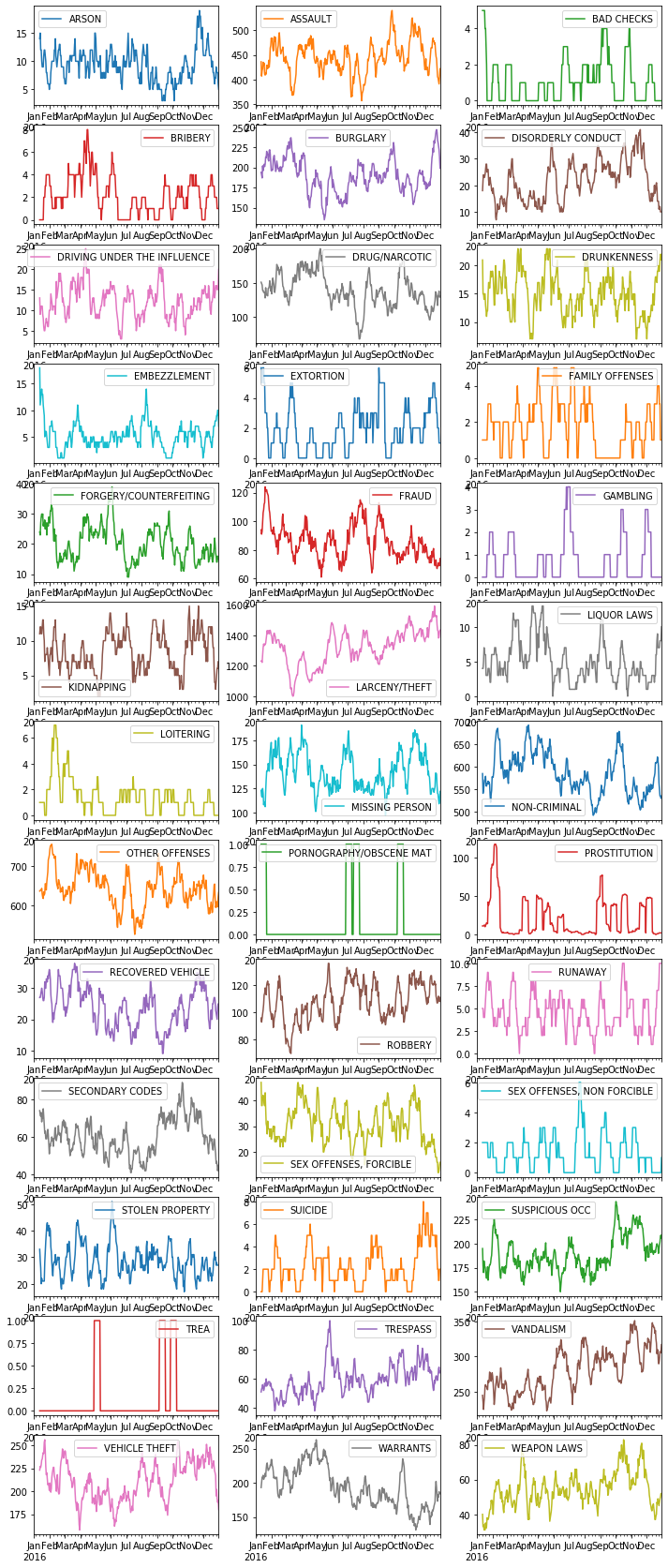
Moving ahead, lets see **number of crimes per month**.



Can clearly see that **May** and **October** is the **peak** of crimes and **February** and **June** is minimum.

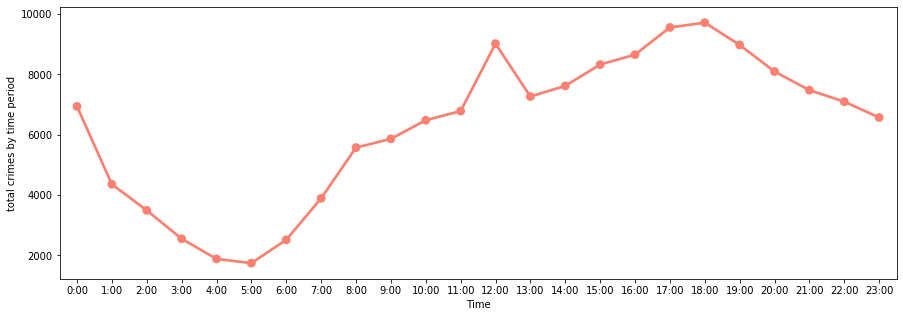
**Third Step**

Here pivot table is created using “incident\_id” and “category” and all the null values has been filled with 0. And now we can see frequency trends for each crime which is following:



**Fourth Step**

There is also another useful insight where total crimes can be seen with respect to time.

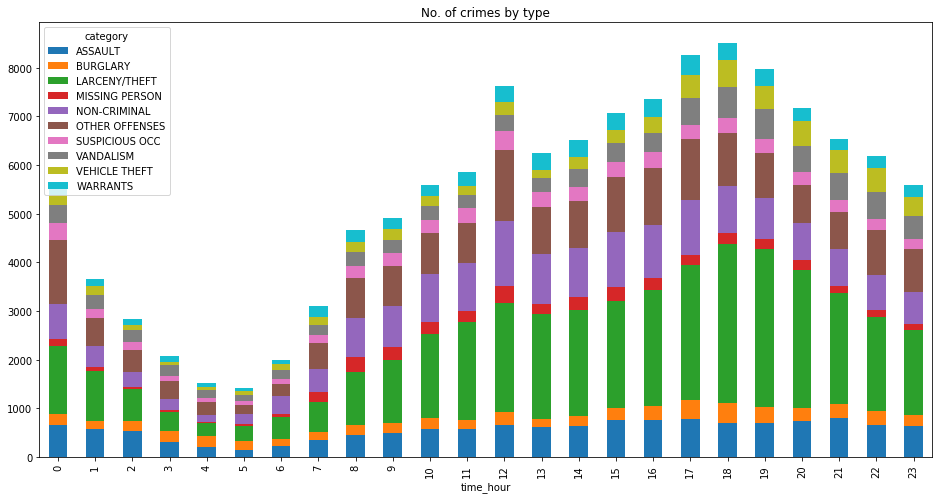


This shows that there is decrease in crimes after midnight and early in the morning its minimum and towards afternoon it again start rising and reaches to the peak and starts decreasing toward the night.

Hence most of the crimes happen during daytime which suggests that police and citizens need to be careful during daytime.

**Fifth Step**

There is extremely powerful tool which is known as stacked bar plot. Using this plot we can clearly see top 10 crimes with respect to hour where every hour shows all the crimes altogether.

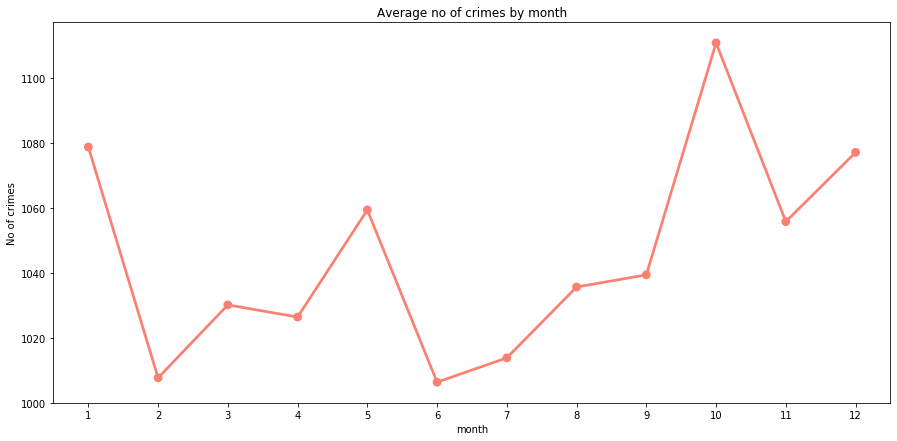


Taking only top crimes into account and plotting the bar stacked plot shows the no of crimes at different hour of the day. And among all of them **larceny**/**theft** is still the most frequent one.

Moreover in the morning crime rate is less and its at peak at 6 in the evening and then starts decreasing.

**Sixth Step**

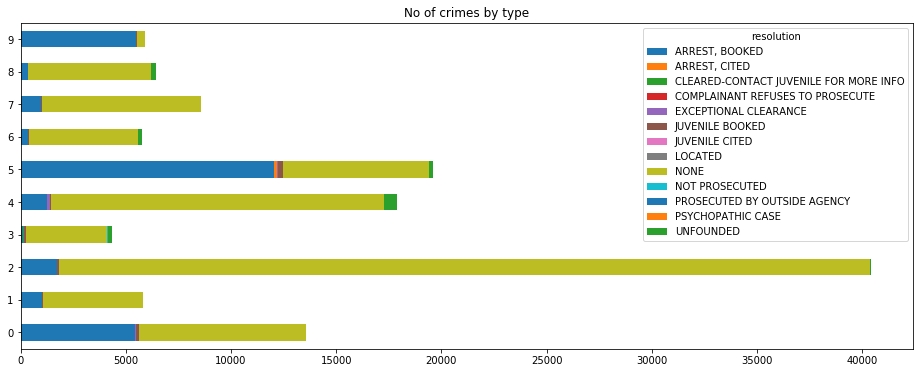
In this step, “Average number of crimes per month” is shown which is quite similar to number of crimes per month.



Again average no of crimes are highest in October. This shows that crimes dips in february and spikes in October. This can be very useful info for travellers and for police department for which months there should be high alert.

**Seventh Step**

Here horizontal bar plot is plotted of top 10 crimes with resolutions taken for each crime.



Highest number of arrest, booked happened in case of vandalism which is index no 5.

**Conclusion**

Data analysis with the help of python helps us drew most important insights such as maximum number of crimes are Grand theft from locked auto. Thus this insight helps us to make such decisions which helps in reduction of this crime.

**Moreover it can be seen in seventh step that there are no steps taken for most of the crimes except missing person crime.** This alerts police to take necessary and appropriate actions in order to reduce the crime rate and ensure more safety standards in the city.