How often do you hear about incidents in your city?

Do you know type of the incidents and how quick they are taken care?
Do you consider it as safety in your city and neighborhood?

- So I am trying to show you level of safety in our city(New York City) based on the number and type of incident in each zip code.
- It requires a dataset that includes different type of incident, response time, number of incident and also population.
- Two datasets are taken from www.cityofnewyork.us
- The features are shown in the next slide

‘STARFIRE_INCIDENT_ID', 'INCIDENT_DATETIME', ‘ALARM_BOX_BOROUGH',

'ALARM_BOX_NUMBER', 'ALARM_BOX_LOCATION', 'INCIDENT_BOROUGH', ‘ZIPCODE',

'POLICEPRECINCT', 'CITYCOUNCILDISTRICT', ‘COMMUNITYDISTRICT',

'COMMUNITYSCHOOLDISTRICT', ‘CONGRESSIONALDISTRICT',

'ALARM_SOURCE_DESCRIPTION_TX', ‘ALARM_LEVEL_INDEX_DESCRIPTION',

'HIGHEST_ALARM_LEVEL', ‘INCIDENT_CLASSIFICATION',

'INCIDENT_CLASSIFICATION_GROUP', ‘DISPATCH_RESPONSE_SECONDS_QY',

'FIRST_ASSIGNMENT_DATETIME', ‘FIRST_ACTIVATION_DATETIME',

'FIRST_ON_SCENE_DATETIME', ‘INCIDENT_CLOSE_DATETIME',

'VALID_DISPATCH_RSPNS_TIME_INDC', ‘VALID_INCIDENT_RSPNS_TIME_INDC',

'INCIDENT_RESPONSE_SECONDS_QY', ‘INCIDENT_TRAVEL_TM_SECONDS_QY',

'ENGINES_ASSIGNED_QUANTITY', ‘LADDERS_ASSIGNED_QUANTITY',

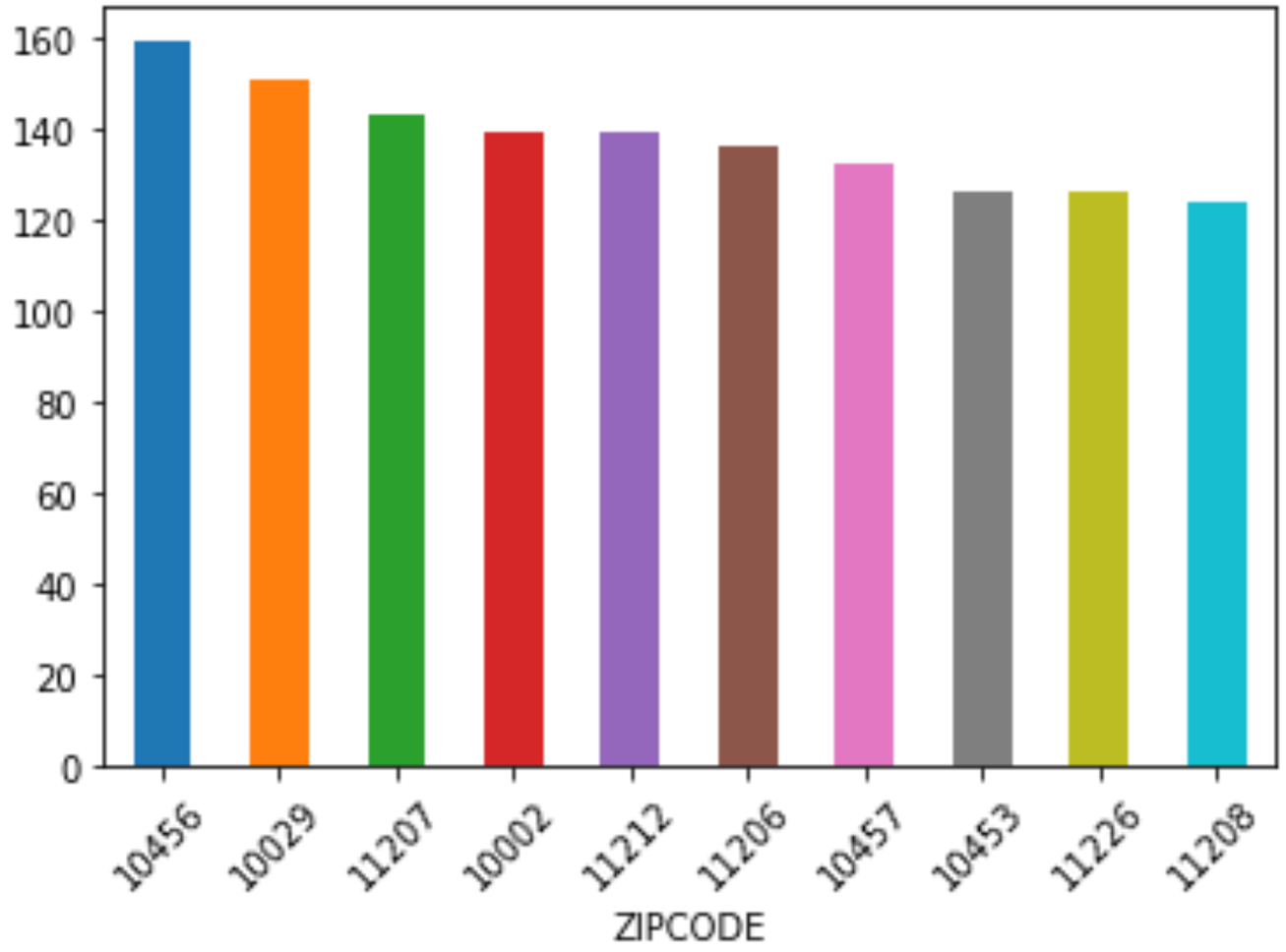
‘OTHER_UNITS_ASSIGNED_QUANTITY’

‘Population’

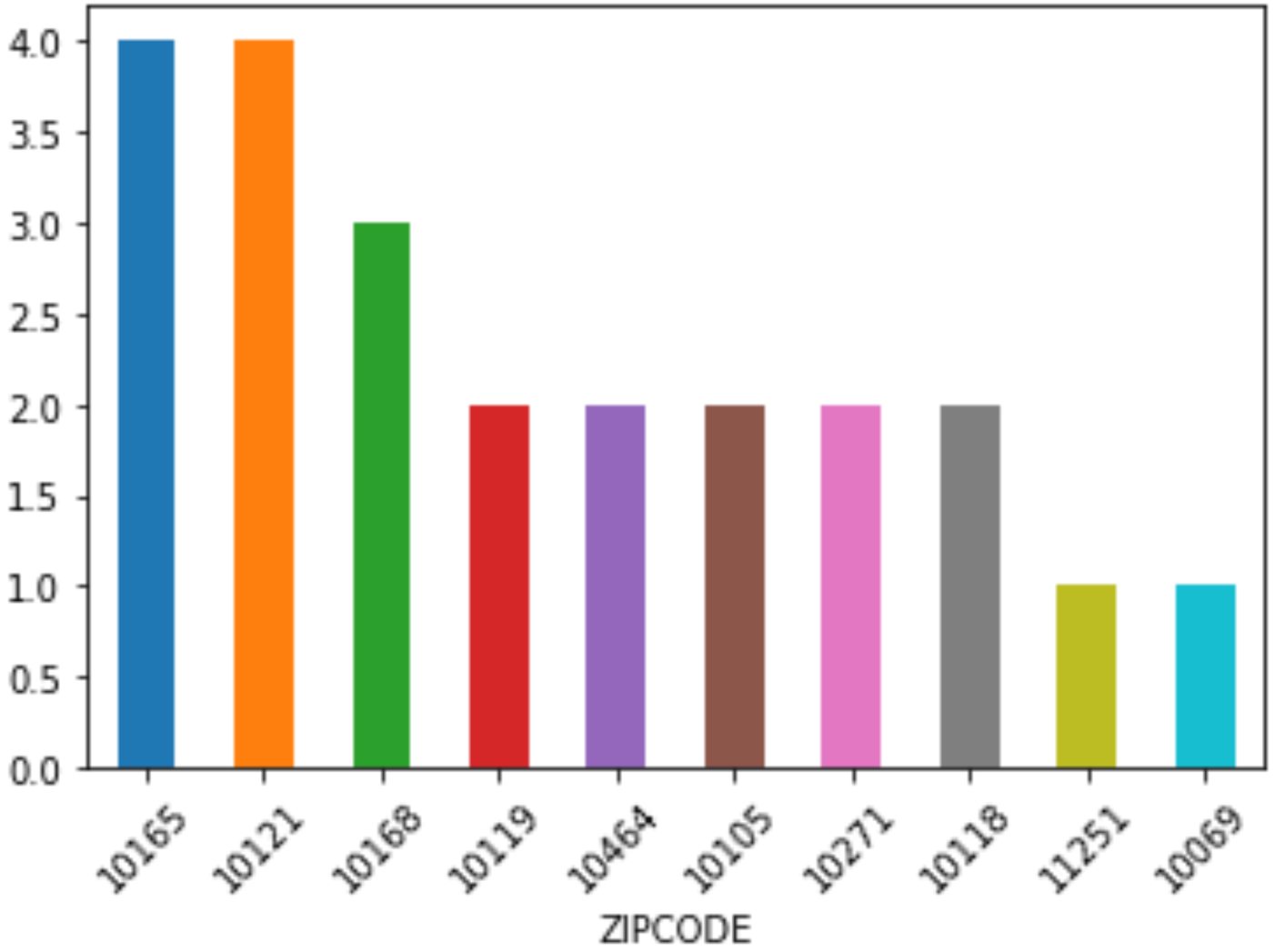
Let’s do exploratory

- Not all of incident should be taken care by fire departments
- Not all of incident are real
- It makes some null values in the dataset
- Real incidents be determined based on dispatch response time

Let’s see 10 top zip code area based on number of incident



Let’s see 10 bottom zip code area based on number of incident



- Incident_Datetime; The date and time of the incident.
- First_Onscene_Datetime; The date and time of the first unit at the scene of the incident.
- First_Assignment_Datetime; The date and time of the first unit assignment.
- Incident_Response_Second; The elapsed time in seconds between the incident_datetime and the first_onscene_datetime.
- Incident_Response_Travel; The elapsed time in seconds between the first_assignment_datetime and the first_onscene_datetime.
- Dispatch_Response_Second; The elapsed time in seconds between the incident_datetime and the first_assignment_datetime.

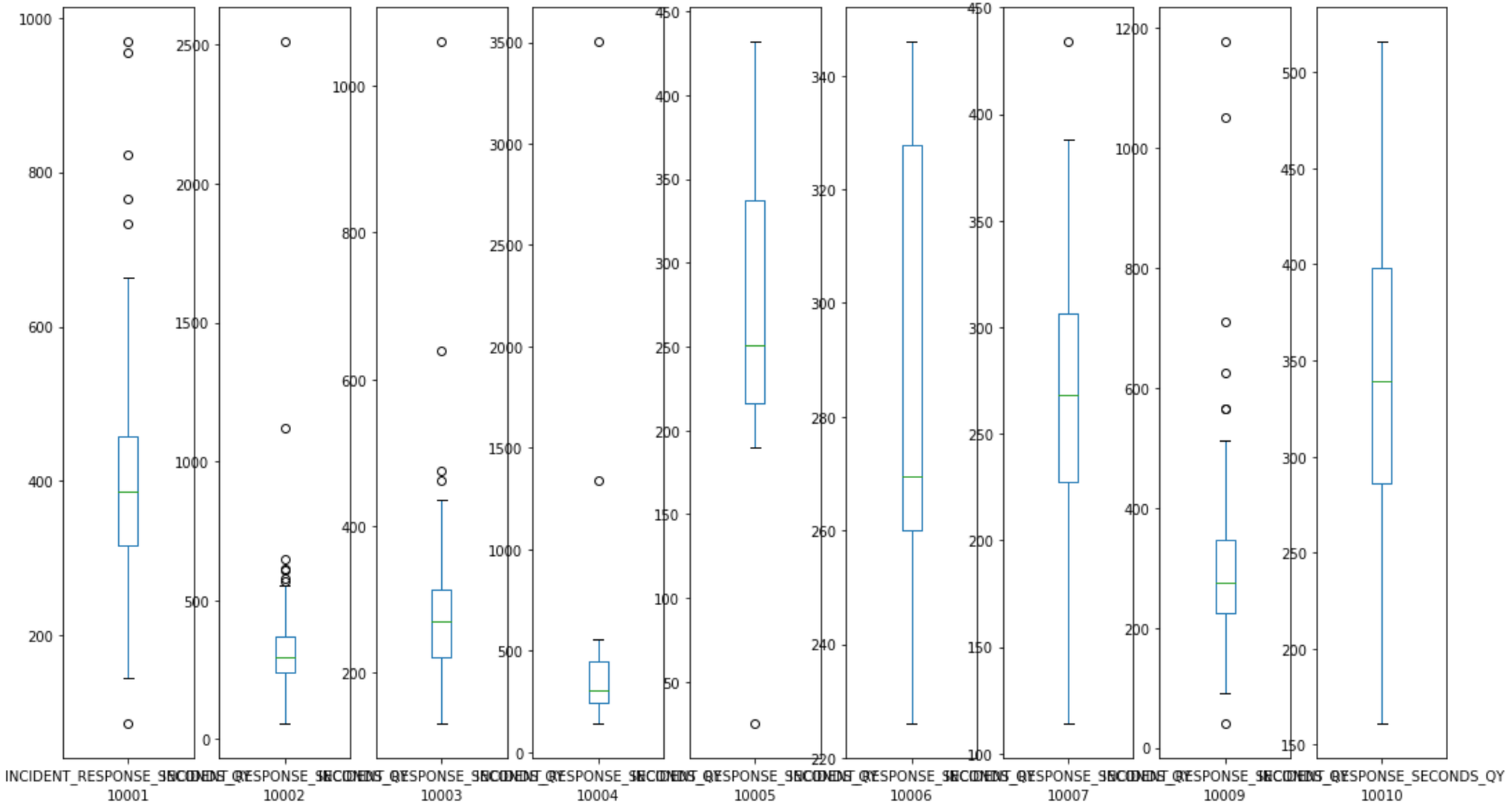
I am going to use :

1. Incident_Response_Second, Incident_Response_Travel and Dispatch_Response_Second features in terms of how quick a fire department response to an incident.(they should be plotted)

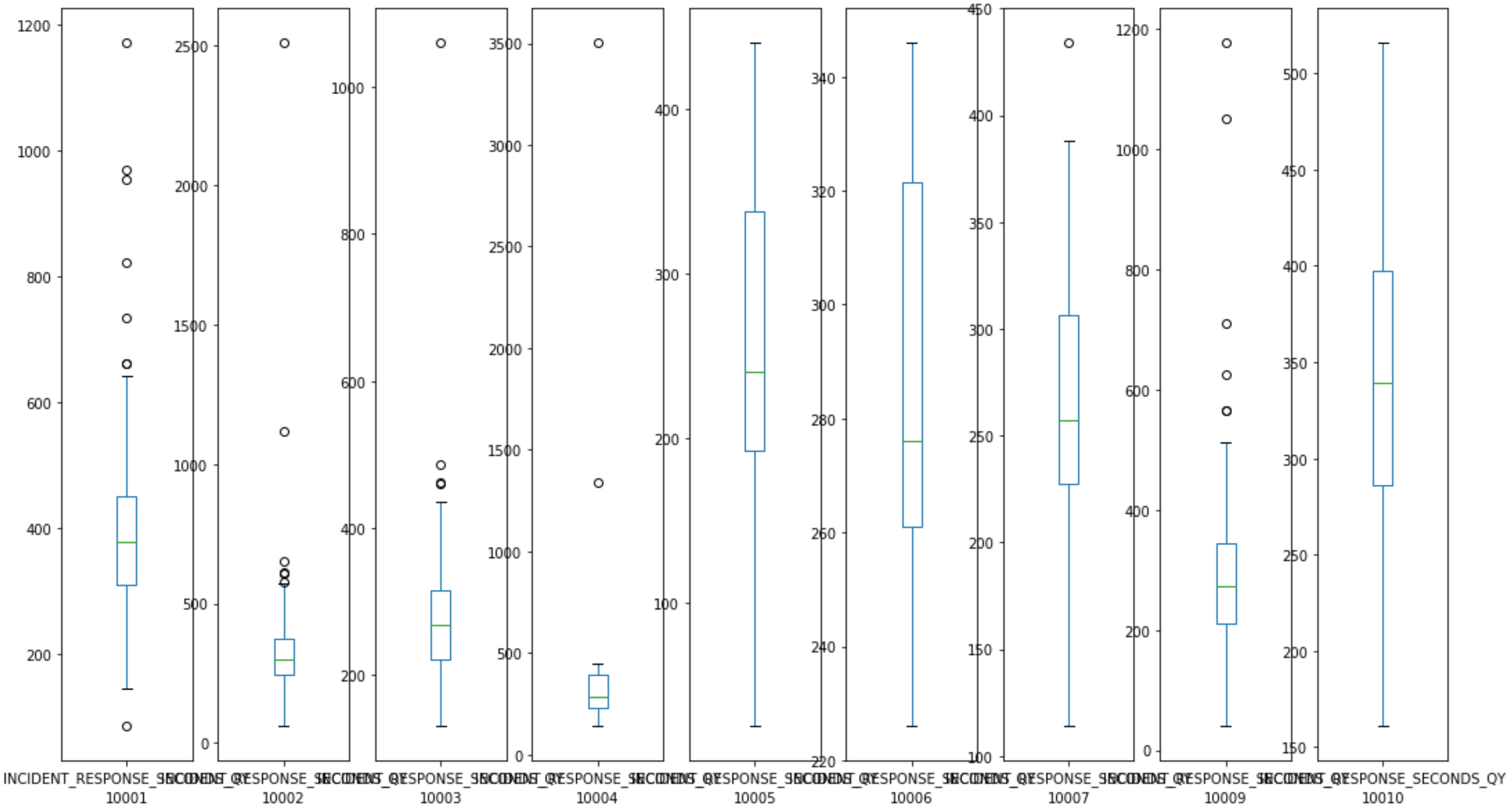
2. STARFIRE_INCIDENT_ID

3. INCIDENT_CLASSIFICATION

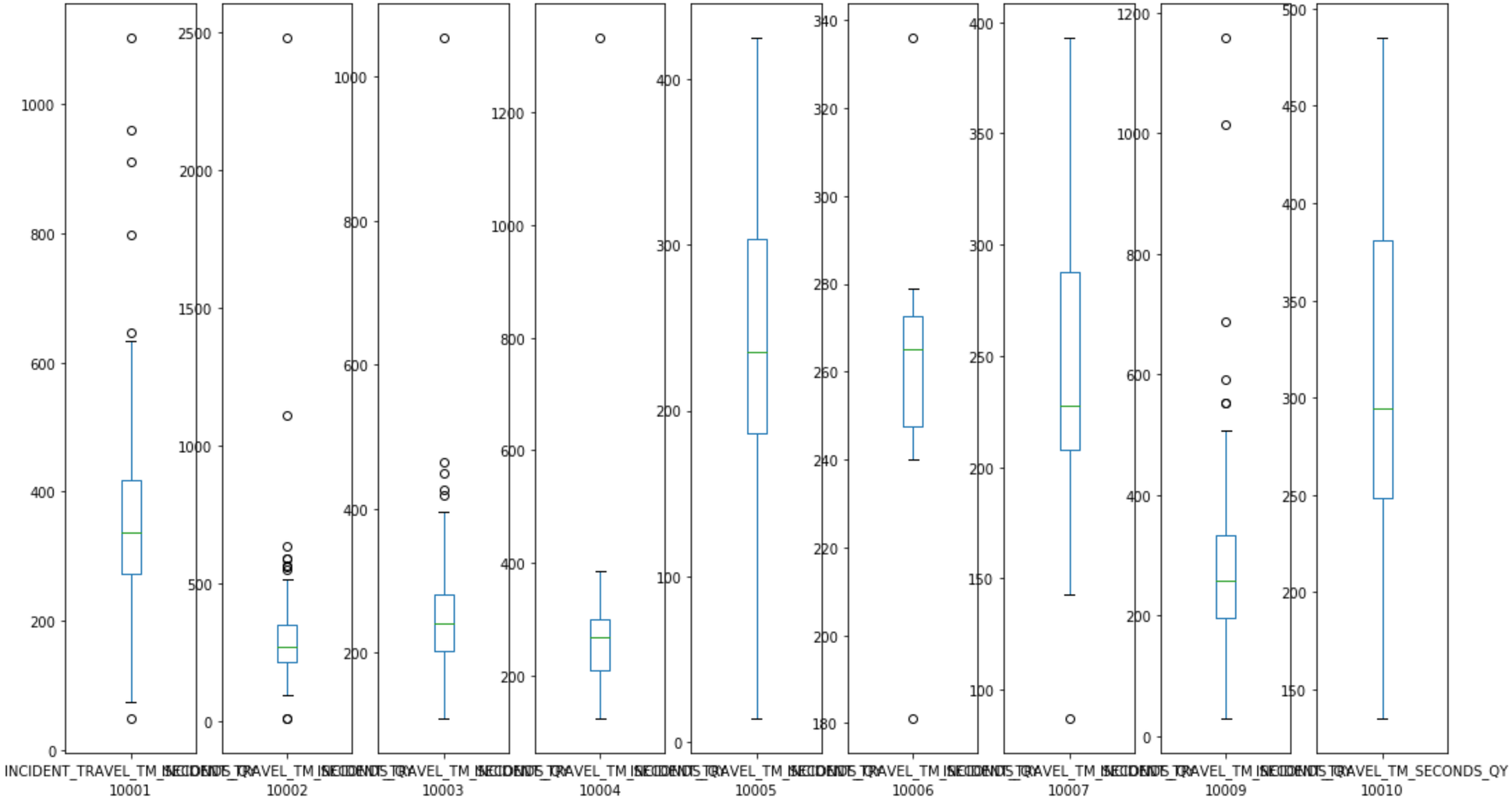
Dispatch_Response_Second



Incident_Response_Second

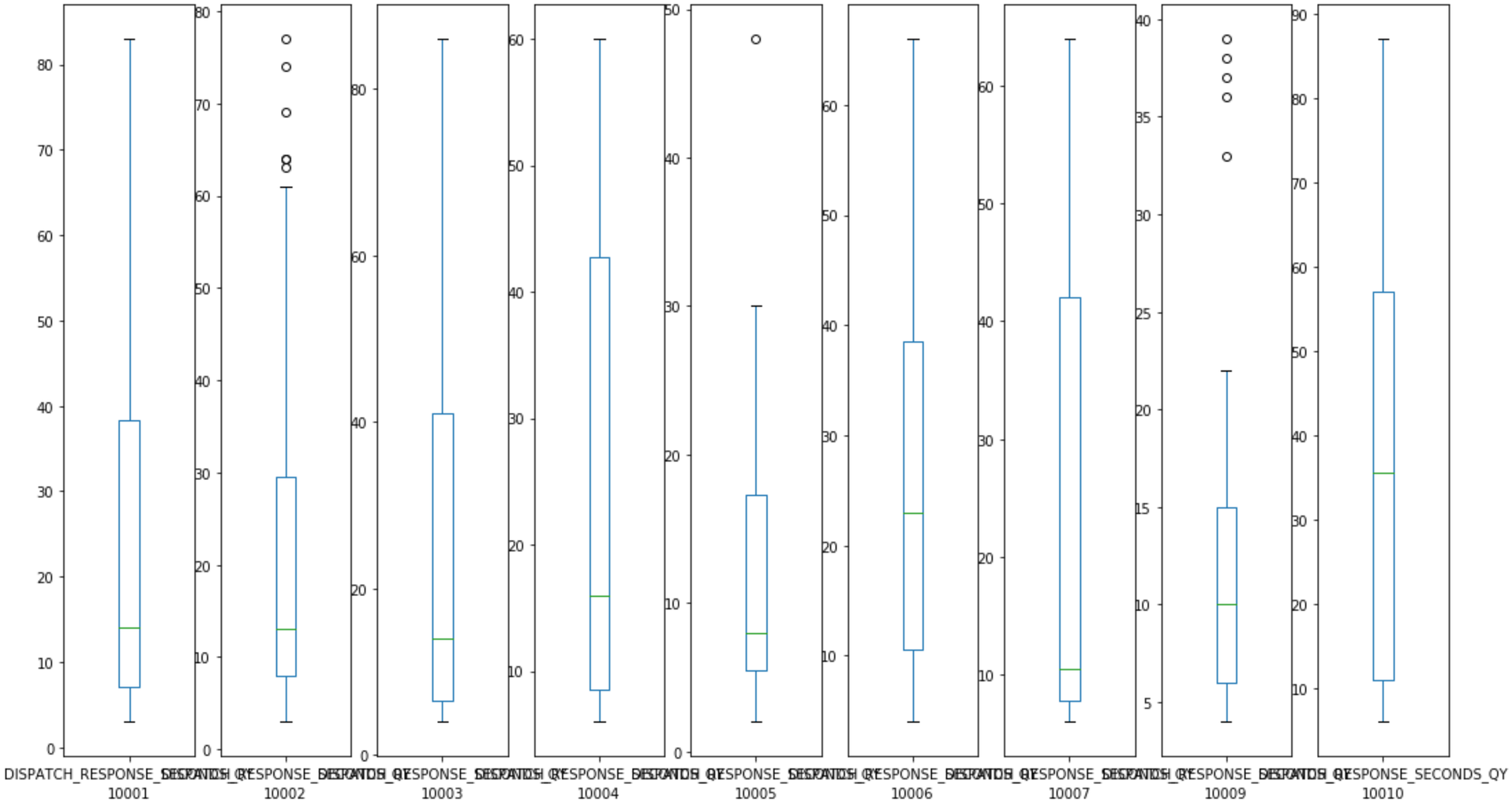


Incident_Response_Travel

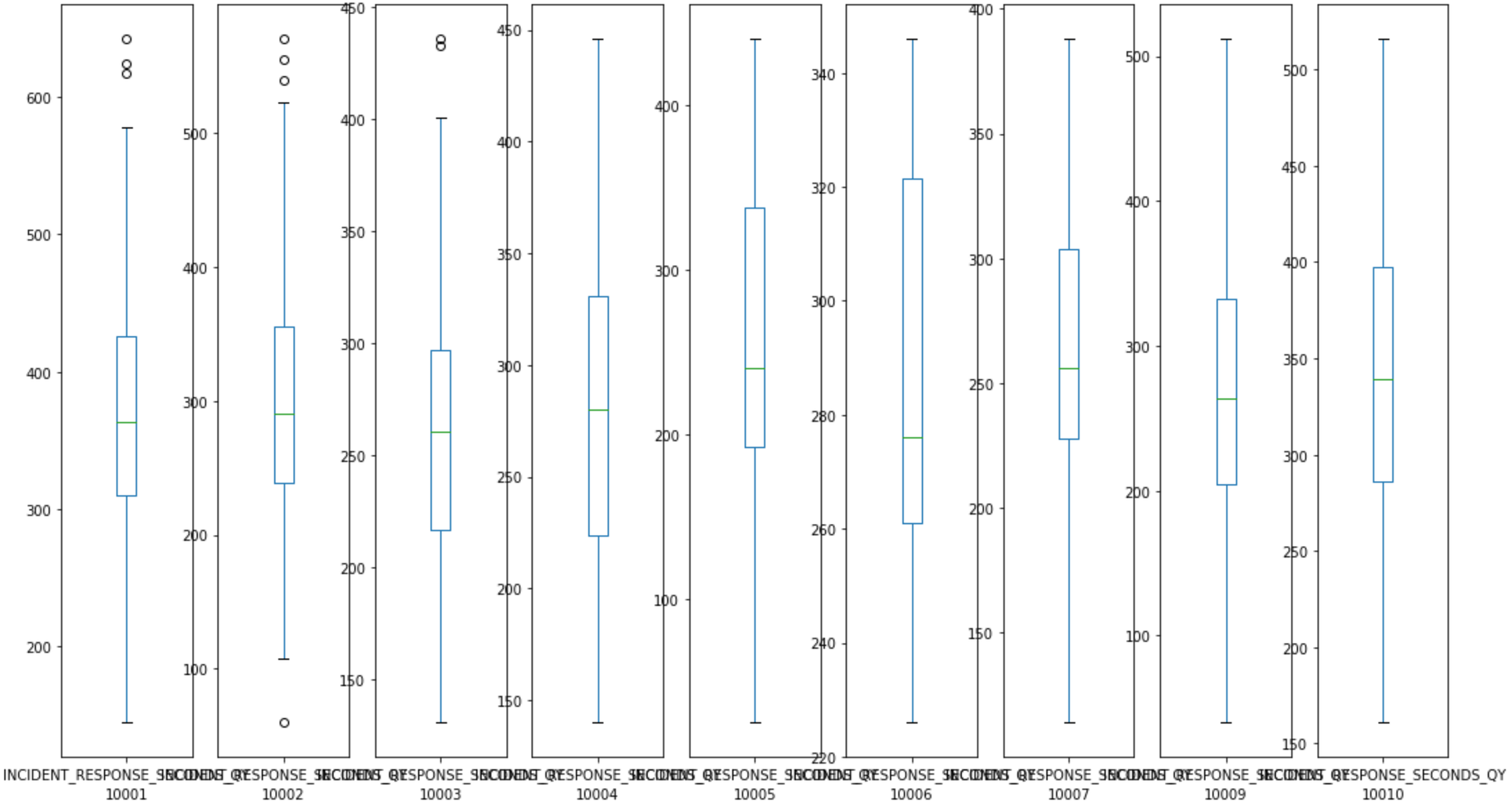


How the plots look like after moving outliers

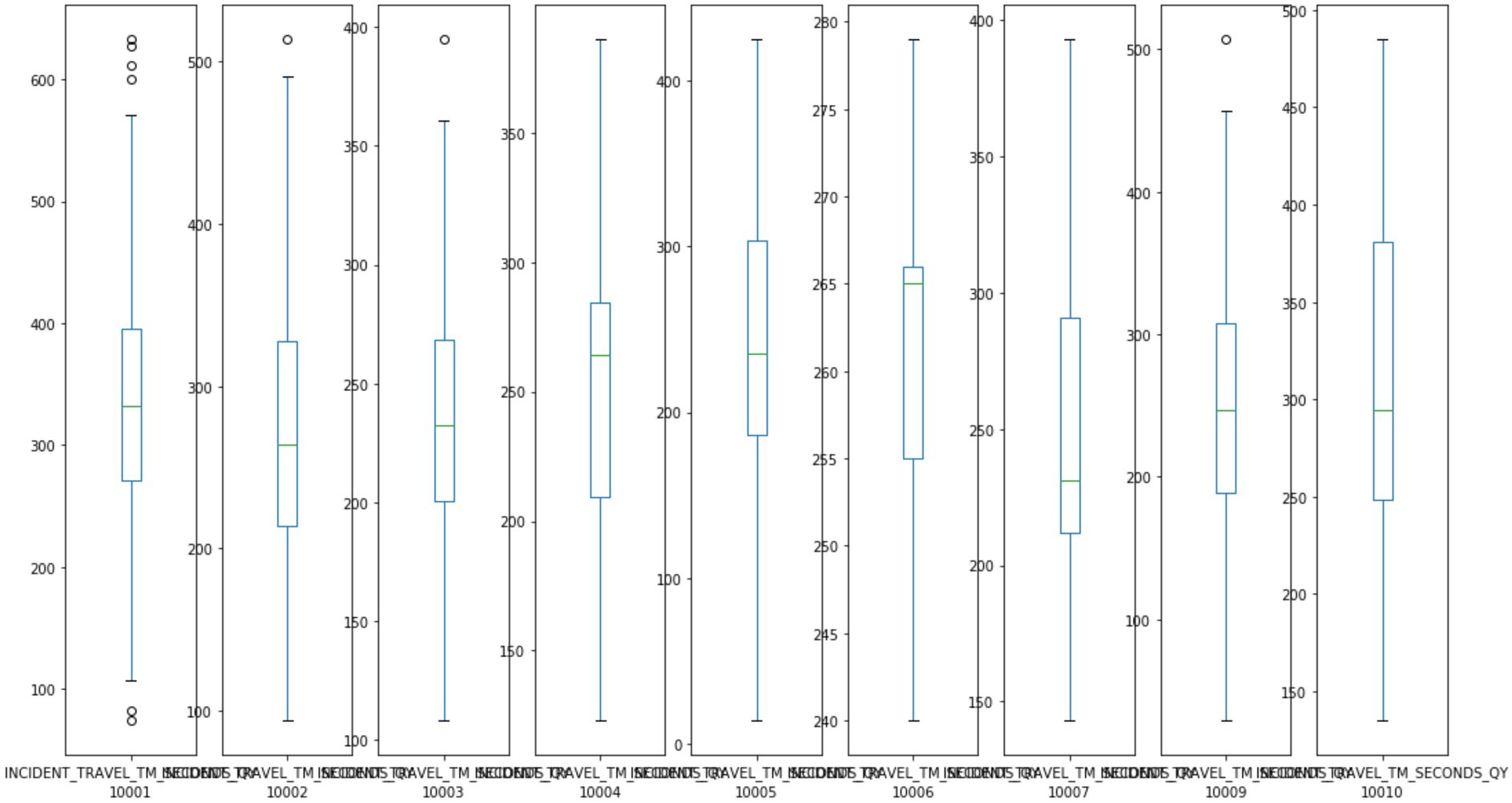
Dispatch_Response_Second



Incident_Response_Second



Incident_Response_Travel



Feature engineering, making new features

- Average of Dispatch_Response_Second
- Average of Incident_Response_Second
- Average of Incident_Response_Travel
- Number of incident
- Classified incident

Merging data frames and series with new features to ‘Population’ data set;

5 rows × 63 columns

	ZIPCODE	STARFIRE_INCIDENT_ID	AVE_INCIDENT_TRAVEL_SECOND	AVE_INCIDENT_RESPONSE_SECOND	AVE_DISPATCH_RESPONSE_SECONDS	Abandoned Derelict Vehicle Fire	Alarm System - Defective	Alarm System - Testing	Alarm System - Unnecessary	Assist Civilian - Non-Medical	...	Undefined Emergency	Untenanted Building Fire	Utility Emergency - Electric	Utility Emergency - Gas	Utility Emergency - Steam	Utility Emergency - Undefined	Utility Emergency - Water	Vehicle Accident - Other	Vehicle Accident - With Extrication	2010 Census Population
0	10001	78	334.291667	364.457143	23.918919	0	5	1	5	23	...	0	0	2	1	0	0	4	1	0	21102
1	10002	110	277.580000	303.951923	20.783019	0	4	0	2	31	...	0	0	2	7	3	0	7	3	0	81410
2	10003	61	235.000000	266.789474	24.728814	0	7	0	5	6	...	0	0	2	2	2	0	5	3	0	56024
3	10004	15	256.857143	284.692308	26.500000	0	4	0	0	1	...	0	0	0	0	0	0	0	1	0	3089
4	10005	11	229.545455	249.545455	14.100000	0	2	0	0	2	...	0	0	0	0	0	0	0	1	0	7135

Making target column (safe, unsafe) based on number of incident by following function and code

```
def binning(col, cut_points, labels=None):
#Define min and max values:
minval = col.min()
maxval = col.max()

#create list by adding min and max to cut_points
break_points = [minval] + cut_points + [maxval]

#if no labels provided, use default labels 0 ... (n-1)
if not labels:
labels = range(len(cut_points)+1)

#Binning using cut function of pandas
colBin = pd.cut(col,bins=break_points,labels=labels,include_lowest=True)
return colBin

cut_points = [40]
labels = ['safe', 'unsafe']
df4['LEVEL_OF_SAFETY'] = binning(df4["STARFIRE_INCIDENT_ID"], cut_points, labels=labels)
```

And of course scaling the data set