

```
In [1]: import pandas as pd
```

```
In [2]: # https://openpolicing.stanford.edu/data/
# https://github.com/stanford-policylab/opp/blob/master/data_readme.md
# url='https://raw.githubusercontent.com/justmarkham/pycon-2018-tutorial/master/police.csv'
# df=pd.read_csv(url)
df=pd.read_csv('police.csv')
```

```
In [3]: df.head(2)
```

```
Out[3]:
```

	stop_date	stop_time	county_name	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search
0	2005-01-02	01:55	NaN	M	1985.0	20.0	White	Speeding	Speeding	
1	2005-01-18	08:15	NaN	M	1965.0	40.0	White	Speeding	Speeding	

```
In [4]: df.shape
```

```
Out[4]: (91741, 15)
```

```
In [5]: df.dtypes
```

```
Out[5]: stop_date          object
stop_time          object
county_name        float64
driver_gender       object
driver_age_raw      float64
driver_age          float64
driver_race         object
violation_raw       object
violation           object
search_conducted    bool
search_type         object
stop_outcome        object
is_arrested         object
stop_duration       object
drugs_related_stop  bool
dtype: object
```

```
In [6]: df.isna().sum()
#df.isnull().sum()
```

```
Out[6]: stop_date          0
stop_time          0
county_name        91741
driver_gender       5335
driver_age_raw      5327
driver_age          5621
driver_race         5333
violation_raw       5333
violation           5333
search_conducted    0
search_type         88545
stop_outcome        5333
is_arrested         5333
stop_duration       5333
drugs_related_stop  0
dtype: int64
```

#### Removing columns which contain only missing values

```
In [7]: # ذخیره در دیتافریم جدید
df_drop_method_one = df.drop(['county_name'],axis=1)
df_drop_method_one.head(1)
```

```
Out[7]:
```

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search_conducted	search_type
0	2005-01-02	01:55	M	1985.0	20.0	White	Speeding	Speeding	False	

```
In [8]: # حذف ستون هایی که تماماً خالی می باشند
df_drop_method_two = df.dropna(axis='columns',how='all')
df_drop_method_two.head(1)
```

```
Out[8]:
```

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search_conducted	search_type
0	2005-01-02	01:55	M	1985.0	20.0	White	Speeding	Speeding	False	

```
In [9]: # حذف از دیتافریم اصلی و جایگزینی
df.drop(['county_name'],axis=1,inplace=True)
df.head(1)
```

```
Out[9]:
```

	stop_date	stop_time	driver_gender	driver_age_raw	driver_age	driver_race	violation_raw	violation	search_conducted	search_type
0	2005-01-02	01:55	M	1985.0	20.0	White	Speeding	Speeding	False	