911 Calls Capstone Project

For this capstone project we will be analyzing some 911 call data from Kaggle. The data contains the following fields:

- lat : String variable, Latitude
- lng: String variable, Longitude
- desc: String variable, Description of the Emergency Call
- zip: String variable, Zipcode
- title: String variable, Title
- timeStamp: String variable, YYYY-MM-DD HH:MM:SS
- twp: String variable, Township
- addr: String variable, Address
- e: String variable, Dummy variable (always 1)

Just go along with this notebook and try to complete the instructions or answer the questions in bold using your Python and Data Science skills!

Data and Setup

** Import numpy and pandas **

```
import numpy as np
import pandas as pd
```

** Import visualization libraries and set %matplotlib inline. **

```
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

** Read in the csv file as a dataframe called df **

```
df=pd.read_csv('911.csv')
```

** Check the info() of the df **

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 99492 entries, 0 to 99491
Data columns (total 9 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   lat        99492 non-null  float64
 1   lng        99492 non-null  float64
 2   desc       99492 non-null  object 
 3   zip        86637 non-null  float64
 4   title      99492 non-null  object 
 5   timeStamp  99492 non-null  object 
 6   twp        99449 non-null  object 
 7   addr       98973 non-null  object 
 8   e          99492 non-null  int64  
dtypes: float64(3), int64(1), object(5)
memory usage: 6.8+ MB
```

** Check the head of df **

```
df.head()
```

	lat	lng	desc	zip	title	timeStamp	
0	40.297876	-75.581294	REINDEER CT & DEAD END; NEW HANOVER; Station ...	19525.0	EMS: BACK PAINS/INJURY	2015-12-10 17:40:00	HAN
1	40.258061	-75.264680	BRIAR PATH & WHITEMARSH LN; HATFIELD TOWNSHIP...	19446.0	EMS: DIABETIC EMERGENCY	2015-12-10 17:40:00	HAT TOW
2	40.121182	-75.351975	HAWS AVE; NORRISTOWN; 2015-12-10 @ 14:39:21-St	19401.0	Fire: GAS-ODOR/LEAK	2015-12-10 17:40:00	NORRIS

Basic Questions

** What are the top 5 zipcodes for 911 calls? **

```
df['zip'].value_counts().head(5)
```

```
19401.0    6979
19464.0    6643
19403.0    4854
19446.0    4748
19406.0    3174
Name: zip, dtype: int64
```

** What are the top 5 townships (twp) for 911 calls? **

```
df['twp'].value_counts().head(5)
```

```
LOWER MERION    8443
ABINGTON        5977
NORRISTOWN      5890
UPPER MERION    5227
CHELTENHAM      4575
Name: twp, dtype: int64
```

** Take a look at the 'title' column, how many unique title codes are there? **

```
df['title'].nunique()
```

```
110
```

Creating new features

[ ] ↳ *47 cells hidden*