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 visualization libraries and set %matplotlib inline.

Read in the csv file as a dataframe called df

Check the info() of the df

In [106…

```
<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
```

Data Cleaning

Drop Column 'e' inplace

In [107…

Which features contain blank, null or empty values?

We can check for missing values with pandas isnull(). This indicates whether values are missing or not. Then we can sum all the values to check every column.

In [108…

Out[108]:
```
lat              0
lng              0
desc             0
zip          12855
title            0
timeStamp        0
twp             43
addr           519
dtype: int64
```

Feature Engineering and Creating new Feature

Reason feature and Title Code

In the titles column there are "Reasons/Departments" specified before the title code. These are EMS, Fire, and Traffic. We are going to use .apply() with a custom lambda expression to create a new column called "Reason" that contains this string value.

For example, if the title column value is EMS: BACK PAINS/INJURY , the Reason column value would be EMS and titile_code column value would be BACK PAINS/INJURY

In [110…

Out[110]:

	Reason	title_code
0	EMS	BACK PAINS/INJURY
1	EMS	DIABETIC EMERGENCY
2	Fire	GAS-ODOR/LEAK
3	EMS	CARDIAC EMERGENCY
4	EMS	DIZZINESS

You should have seen that these timestamps are still strings. Use pd.to_datetime to convert the column from strings to DateTime objects.

In [111...

You can now grab specific attributes from a Datetime object by calling them. For example:

time = df['timeStamp'].iloc[0]

time.hour

You can use Jupyter's tab method to explore the various attributes you can call. Now that the timestamp column are actually DateTime objects, use .apply() to create 3 new columns called Hour, Month, and Day of Week. You will create these columns based off of the timeStamp column, reference the solutions if you get stuck on this step.

In [124...

Out[124]:

	Hour	Month	Day of Week
0	17	12	3
1	17	12	3
2	17	12	3
3	17	12	3
4	17	12	3

Notice how the Day of Week is an integer 0-6. Use the .map() with this dictionary to map the actual string names to the day of the week:

```
dmap = {0:'Mon',1:'Tue',2:'Wed',3:'Thu',4:'Fri',5:'Sat',6:'Sun'}
```

Code:

```
df['Day of Week']=df['Day of Week'].map(dmap)
```

In [125...

Out[125]:

```
0    Thu
1    Thu
2    Thu
3    Thu
4    Thu
Name: Day of Week, dtype: object
```

EDA

What is the most common Reason for a 911 call based on the Reason Column created

The number one reason for 911 calls are Emergency Medical Services. Almost half of the reasons are for EMS.

Draw a seaborn countplot for each Reason

In [116…

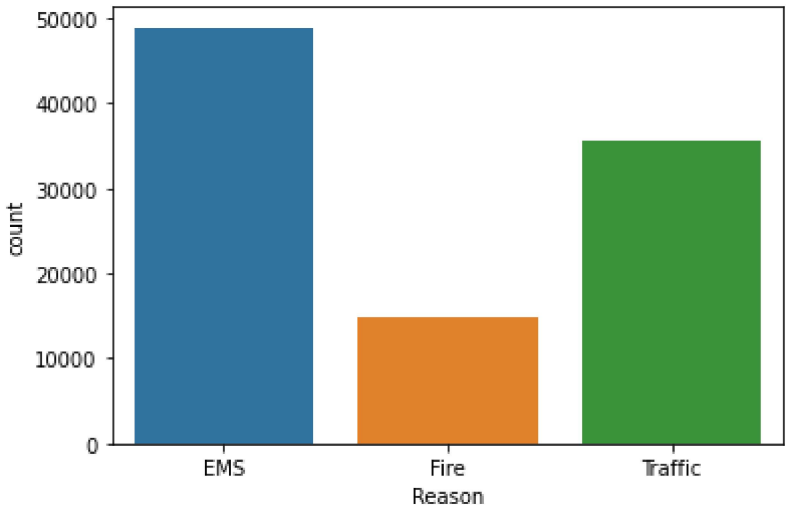
```
'EMS'
```

Out[116]:

In [117…

```
<AxesSubplot: xlabel='Reason', ylabel='count'>
```

Out[117]:



What are the top 5 zip codes for 911 calls?

In [118…

Out[118]:

	zip
19401.0	6979
19464.0	6643
19403.0	4854
19446.0	4748
19406.0	3174

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

In [119…

In [120…

Out[120]:

	twp
LOWER MERION	8443
ABINGTON	5977
NORRISTOWN	5890
UPPER MERION	5227
CHELTENHAM	4575

In [ ]:

Which were the top 5 title codes under EMS Reason that was caled for 911 Emergency

In [122…

Out[122]:
```
 RESPIRATORY EMERGENCY    5112
 CARDIAC EMERGENCY        5012
 FALL VICTIM              4863
 VEHICLE ACCIDENT         3935
 SUBJECT IN PAIN          2687
Name: title_code, dtype: int64
```

Now create a simple plot off of the dataframe indicating the count of calls per month.

In [123…

Out[123]:
```
<AxesSubplot: xlabel='Month'>
```