Module-03, Python for Data Analysis

Real Data Example(911 Calls)

Dostdar Ali
Instructor

Data science and Artificial Intelligence
3-Months Course
at
Karakaroum international Univrsity

January 14, 2024

Table of Contents

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
  - 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Dataset 911 Calls

- We will be analyzing some 911 call data from Kaggle.
- The data contains the following features.
- 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)

Basic Steps for Data load and View

- Import fundamental libraries Like numpy and pandas.
- Import visualization libraries like matplotlib and seaborn.
- Read in the csv files as dataframe called df.
- Check the info or shape of the df.
- Check the head of df.

Basic Steps for Data load and View

- Import fundamental libraries Like numpy and pandas.
- Import visualization libraries like matplotlib and seaborn.
- Read in the csv files as dataframe called df.
- Check the info or shape of the df.
- Check the head of df.

Basic Steps for Data load and View

- Import fundamental libraries Like numpy and pandas.
- Import visualization libraries like matplotlib and seaborn.
- Read in the csv files as dataframe called df.
- Check the info or shape of the df.
- Check the head of df.

Basic Steps for Data load and View

- Import fundamental libraries Like numpy and pandas.
- Import visualization libraries like matplotlib and seaborn.
- Read in the csv files as dataframe called df.
- Check the info or shape of the df.
- Check the head of df.

Basic Steps for Data load and View

- Import fundamental libraries Like numpy and pandas.
- Import visualization libraries like matplotlib and seaborn.
- Read in the csv files as dataframe called df.
- Check the info or shape of the df.
- Check the head of df.

Basic Questions

- What are the top 5 zipcodes for 911 calls?
- What are the top 5 townships (twp) for 911 calls?
- Take a look at the 'title' column, how many unique title codes are there?
- Now use seaborn to create a countplot of 911 calls by Reason.
- Now use seaborn to create a countplot of the Day of Week column with the hue based off of the Reason column.

Basic Questions

- What are the top 5 zipcodes for 911 calls?
- What are the top 5 townships (twp) for 911 calls?
- Take a look at the 'title' column, how many unique title codes are there?
- Now use seaborn to create a countplot of 911 calls by Reason.
- Now use seaborn to create a countplot of the Day of Week column with the hue based off of the Reason column.

Basic Questions

- What are the top 5 zipcodes for 911 calls?
- What are the top 5 townships (twp) for 911 calls?
- Take a look at the 'title' column, how many unique title codes are there?
- Now use seaborn to create a countplot of 911 calls by Reason.
- Now use seaborn to create a countplot of the Day of Week column with the hue based off of the Reason column.

Basic Questions

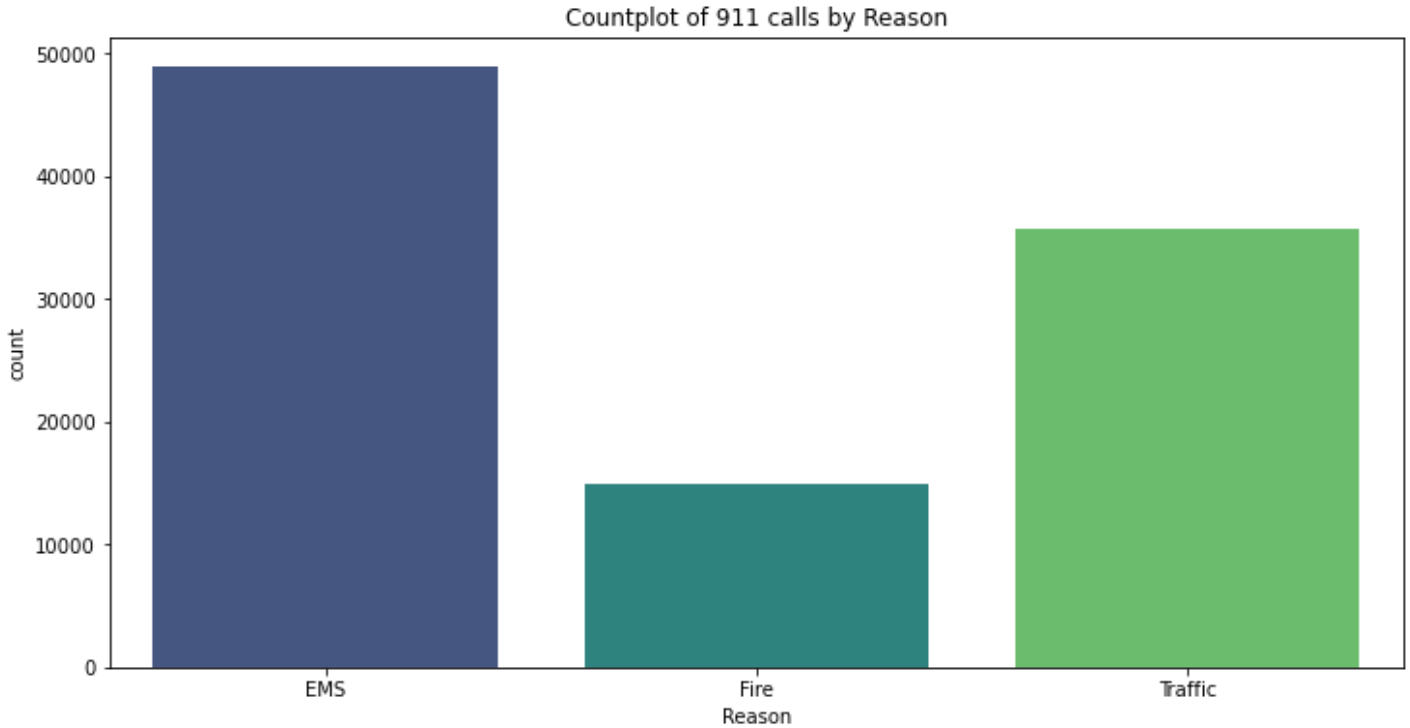
- What are the top 5 zipcodes for 911 calls?
- What are the top 5 townships (twp) for 911 calls?
- Take a look at the 'title' column, how many unique title codes are there?
- Now use seaborn to create a countplot of 911 calls by Reason.
- Now use seaborn to create a countplot of the Day of Week column with the hue based off of the Reason column.

Basic Questions

- What are the top 5 zipcodes for 911 calls?
- What are the top 5 townships (twp) for 911 calls?
- Take a look at the 'title' column, how many unique title codes are there?
- Now use seaborn to create a countplot of 911 calls by Reason.
- Now use seaborn to create a countplot of the Day of Week column with the hue based off of the Reason column.

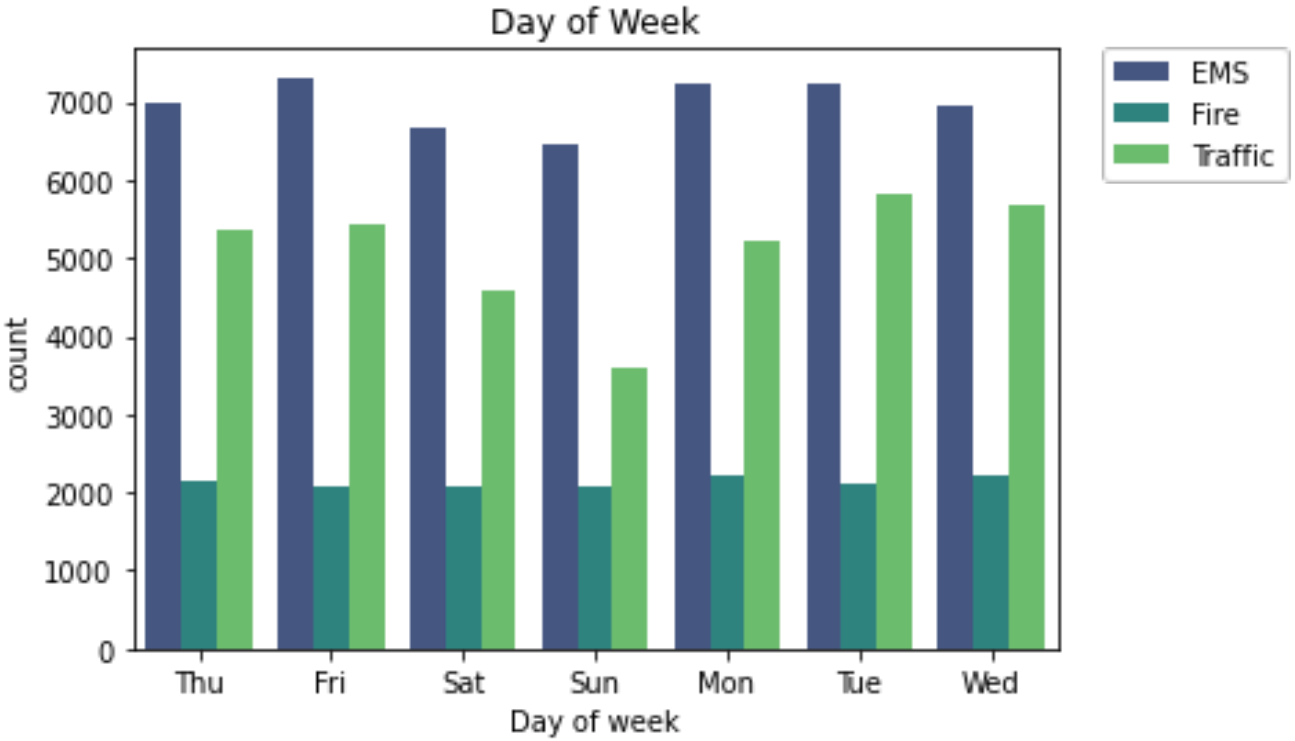
Countplot Reason

- Countplot by Reason



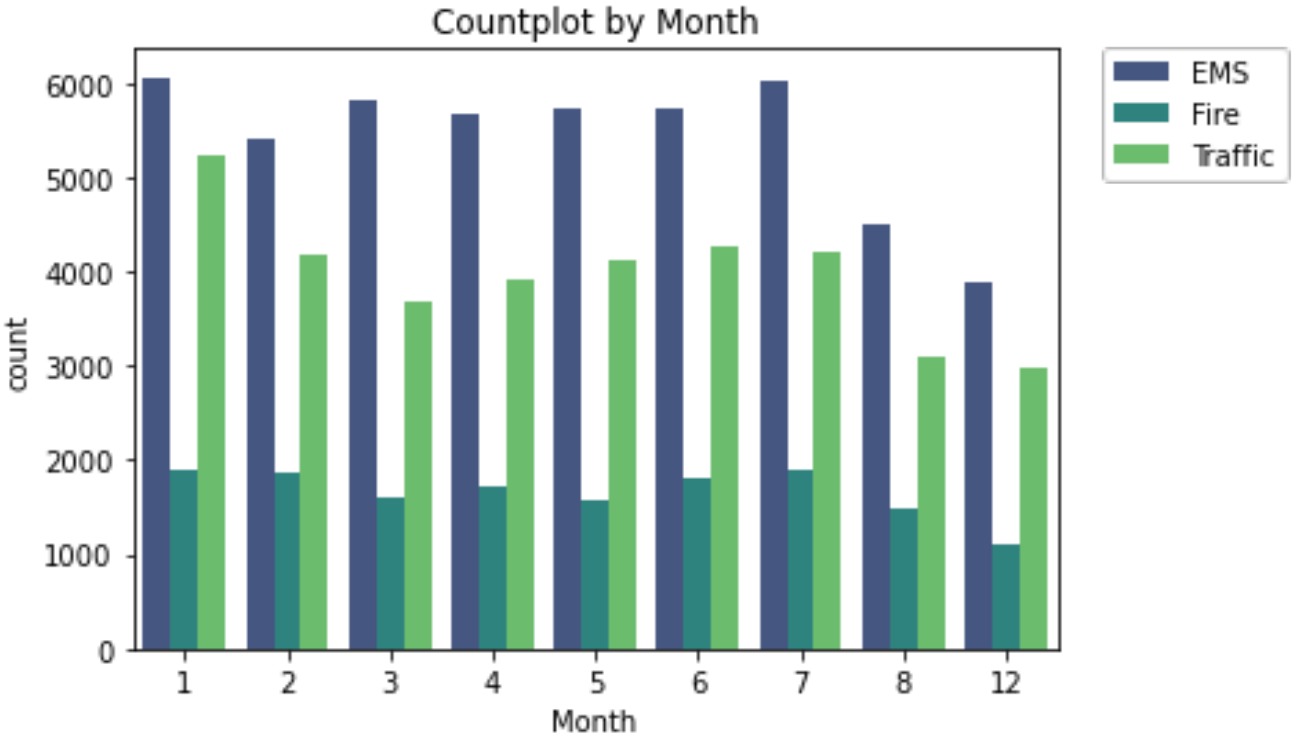
Countplot Day of Week

- Countplot by Day of Week



Countplot by Month

- Countplot by Month



Inside Data

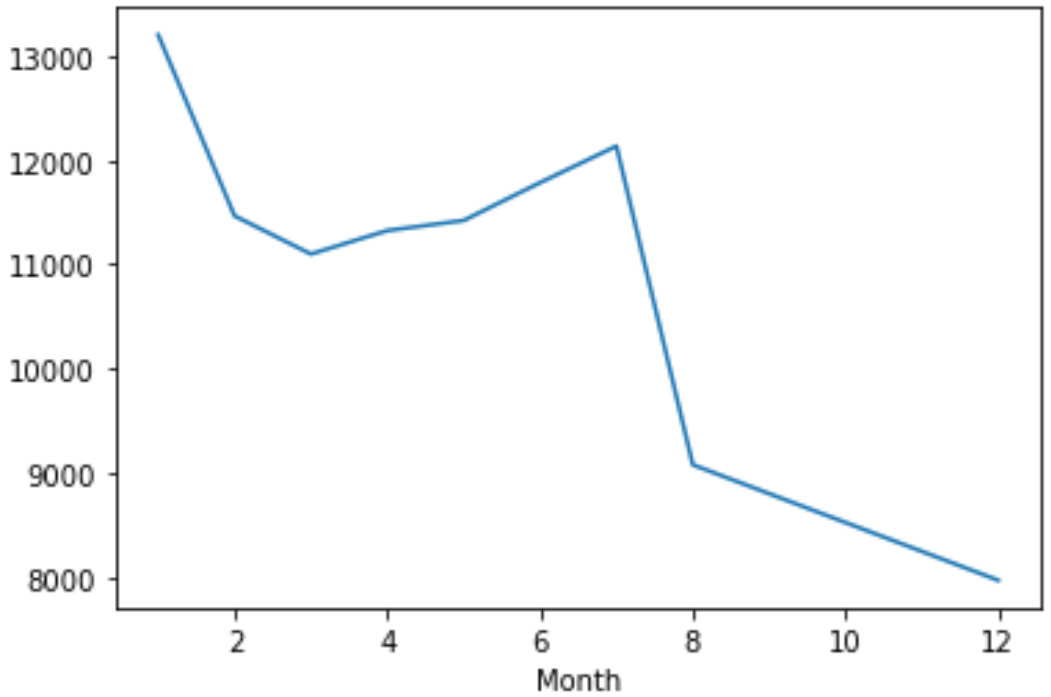
We have noticed it was missing some Months, let's see if we can maybe fill in this information by plotting the information in another way, possibly a simple line plot that fills in the missing months, in order to do this, we'll need to do some work with pandas...

Now create a gropuby object called byMonth, where you group the DataFrame by the month column and use the count() method for aggregation. Use the head() method on this returned DataFrame.

Month	lat	lng	desc	zip	title	timeStamp	twp	addr	e	Reason	Hour	Day of week
1	13205	13205	13205	11527	13205	13205	13203	13096	13205	13205	13205	13205
2	11467	11467	11467	9930	11467	11467	11465	11396	11467	11467	11467	11467
3	11101	11101	11101	9755	11101	11101	11092	11059	11101	11101	11101	11101
4	11326	11326	11326	9895	11326	11326	11323	11283	11326	11326	11326	11326
5	11423	11423	11423	9946	11423	11423	11420	11378	11423	11423	11423	11423
6	11786	11786	11786	10212	11786	11786	11777	11732	11786	11786	11786	11786
7	12137	12137	12137	10633	12137	12137	12133	12088	12137	12137	12137	12137
8	9078	9078	9078	7832	9078	9078	9073	9025	9078	9078	9078	9078
12	7969	7969	7969	6907	7969	7969	7963	7916	7969	7969	7969	7969

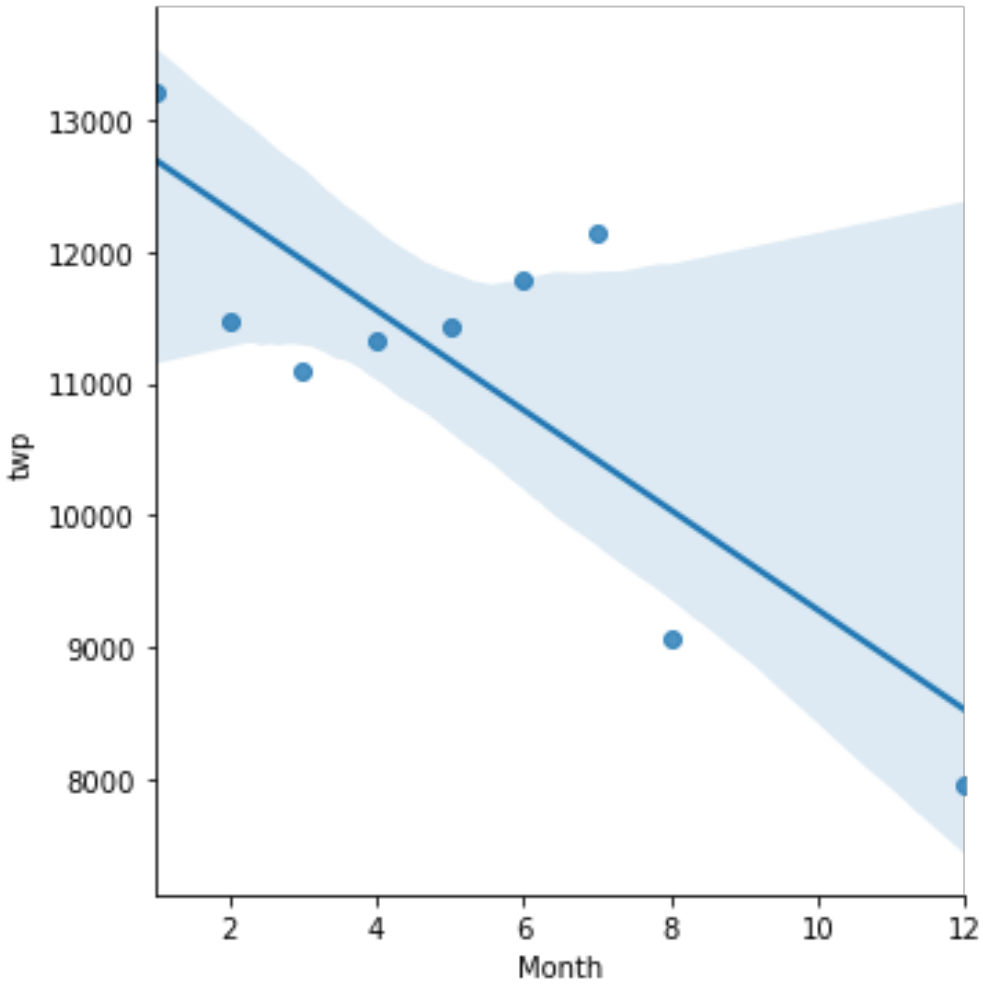
Line plot

- Line plot



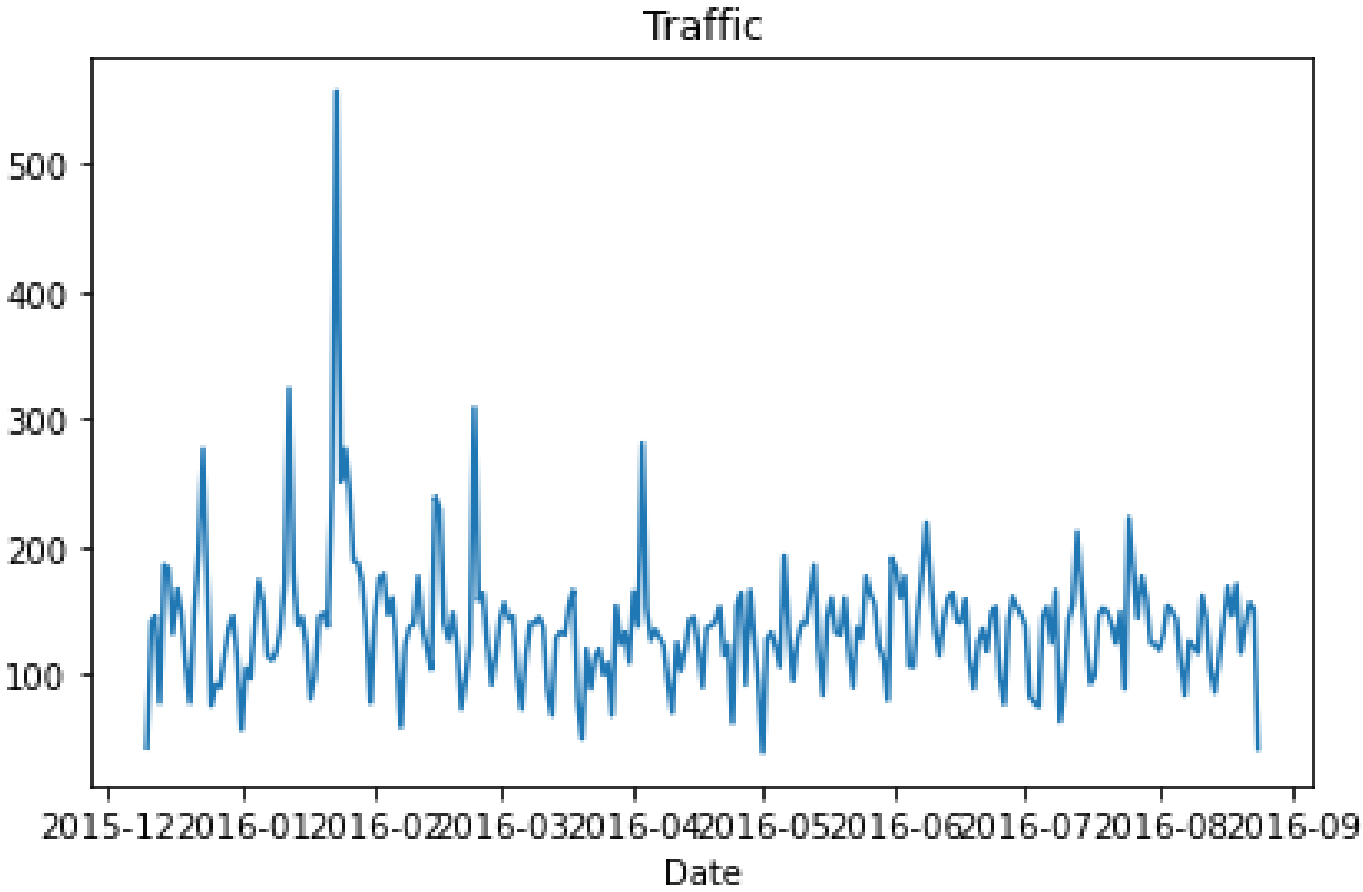
Regression plot

- Regression plot



Calls for a day in a reason

- Calls for a Reason



Great Job
Thank you