

```
In [3]: import os
import pandas as pd
import numpy as np
import pickle
import json
```

```
In [4]: Path = "C:/Users/schen414/Desktop/\\DataDive_2019/"
```

```
In [5]: NYC_311_2018_2019_SEP_DF = pd.read_csv(Path+'data\\NYC_311_20180101_through_20190930.csv')
```

C:\Users\schen414\AppData\Local\Continuum\Anaconda3\envs\tf\lib\site-packages\IPython\core\interactiveshell.py:3049: DtypeWarning: Columns (6,17,20,31,34) have mixed types. Specify dtype option on import or set low\_memory=False.

interactivity=interactivity, compiler=compiler, result=result)

```
In [6]: import pandas_profiling
```

```
In [7]: NYC_311_2018_2019_SEP_DF.head()
```

	Unique Key	Created Date	Closed Date	Agency	Agency Name	Complaint Type	Descriptor	Location Type	Incident Zip	Incident Address	...	Vehicle Type
0	38060564	12/31/2017 12:00:26 AM	01/08/2018 02:05:27 PM	DOT	Department of Transportation	Street Sign - Damaged	Bus Stop	Street	11356	NaN	...	NaN
1	38058568	12/31/2017 12:01:00 AM	01/03/2018 12:00:00 AM	DSNY	Department of Sanitation	Request Large Bulky Item Collection	Request Large Bulky Item Collection	Sidewalk	11221	487 MONROE STREET	...	NaN
2	38062203	12/31/2017 12:01:03 AM	01/13/2018 12:05:43 AM	HPD	Department of Housing Preservation and Develop...	HEAT/HOT WATER	ENTIRE BUILDING	RESIDENTIAL BUILDING	10468	2845 UNIVERSITY AVENUE	...	NaN
3	38059708	12/31/2017 12:01:03 AM	12/31/2017 01:40:08 AM	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	Store/Commercial	11237	195 STANHOPE STREET	...	NaN
4	38063401	12/31/2017 12:01:07 AM	12/31/2017 07:12:19 AM	NYPD	New York City Police Department	Noise - Commercial	Loud Music/Party	Store/Commercial	10040	65 SHERMAN AVENUE	...	NaN
5 rows x 41 columns												

```
In [8]: NYC_311_2018_2019_SEP_DF.columns
```

Index(['Unique Key', 'Created Date', 'Closed Date', 'Agency', 'Agency Name', 'Complaint Type', 'Descriptor', 'Location Type', 'Incident Zip', 'Incident Address', 'Street Name', 'Cross Street 1', 'Cross Street 2', 'Intersection Street 1', 'Intersection Street 2', 'Address Type', 'City', 'Landmark', 'Facility Type', 'Status', 'Due Date', 'Resolution Description', 'Resolution Action Updated Date', 'Community Board', 'BBL', 'Borough', 'X Coordinate (State Plane)', 'Y Coordinate (State Plane)', 'Open Data Channel Type', 'Park Facility Name', 'Park Borough', 'Vehicle Type', 'Taxi Company Borough', 'Taxi Pick Up Location', 'Bridge Highway Name', 'Bridge Highway Direction', 'Road Ramp', 'Bridge Highway Segment', 'Latitude', 'Longitude', 'Location'], dtype='object')



## Overview

### Dataset info

Number of variables	41
Number of observations	4658607
Total Missing (%)	31.4%
Total size in memory	1.4 GB
Average record size in memory	328.0 B

### Variables types

Numeric	4
Categorical	34
Boolean	0
Date	0
Text (Unique)	0
Rejected	2
Unsupported	1

### Warnings

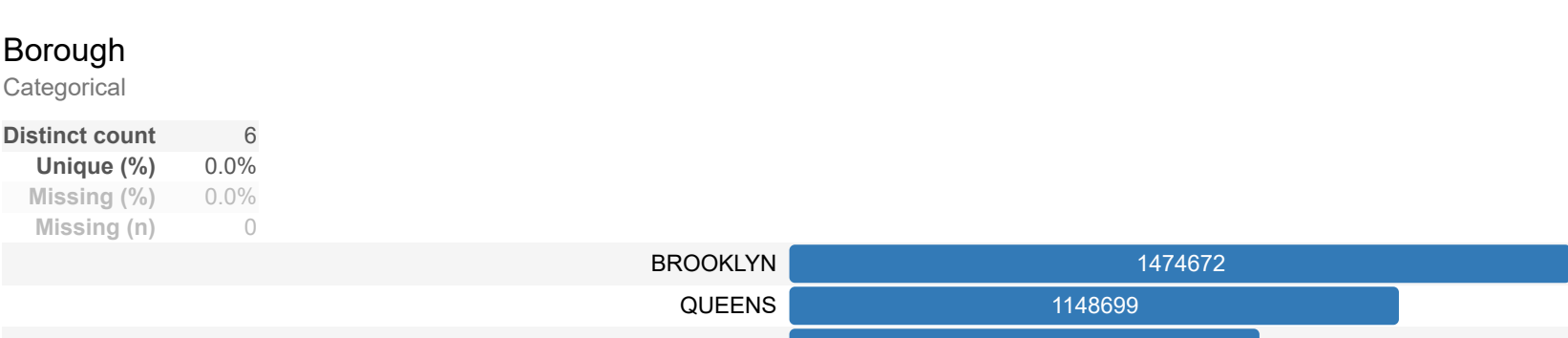
- [Address\\_Type](#) has 457798 / 9.8% missing values [Missing](#)
- [Agency\\_Name](#) has a high cardinality: 1004 distinct values [Warning](#)
- [BBL](#) has 884935 / 19.0% missing values [Missing](#)
- [Bridge\\_Highway\\_Direction](#) has 4651000 / 99.8% missing values [Missing](#)
- [Bridge\\_Highway\\_Direction](#) has a high cardinality: 51 distinct values [Warning](#)
- [Bridge\\_Highway\\_Name](#) has 4651697 / 99.9% missing values [Missing](#)
- [Bridge\\_Highway\\_Name](#) has a high cardinality: 79 distinct values [Warning](#)
- [Bridge\\_Highway\\_Segment](#) has 4648257 / 99.8% missing values [Missing](#)
- [Bridge\\_Highway\\_Segment](#) has a high cardinality: 4004 distinct values [Warning](#)
- [City](#) has 222263 / 4.8% missing values [Missing](#)
- [City](#) has a high cardinality: 646 distinct values [Warning](#)
- [Closed\\_Date](#) has 174010 / 3.7% missing values [Missing](#)
- [Closed\\_Date](#) has a high cardinality: 1777563 distinct values [Warning](#)
- [Community\\_Board](#) has a high cardinality: 77 distinct values [Warning](#)
- [Complaint\\_Type](#) has a high cardinality: 386 distinct values [Warning](#)
- [Created\\_Date](#) has a high cardinality: 3442409 distinct values [Warning](#)
- [Cross\\_Street\\_1](#) has 1848169 / 39.7% missing values [Missing](#)
- [Cross\\_Street\\_1](#) has a high cardinality: 19487 distinct values [Warning](#)
- [Cross\\_Street\\_2](#) has 1859363 / 39.9% missing values [Missing](#)
- [Cross\\_Street\\_2](#) has a high cardinality: 19757 distinct values [Warning](#)
- [Descriptor](#) has 65912 / 1.4% missing values [Missing](#)
- [Descriptor](#) has a high cardinality: 1322 distinct values [Warning](#)
- [Due\\_Date](#) has 2988945 / 64.2% missing values [Missing](#)
- [Due\\_Date](#) has a high cardinality: 1563855 distinct values [Warning](#)
- [Facility\\_Type](#) has 3385624 / 72.7% missing values [Missing](#)
- [Incident\\_Address](#) has 842547 / 13.8% missing values [Missing](#)
- [Incident\\_Address](#) has a high cardinality: 771408 distinct values [Warning](#)
- [Incident\\_Zip](#) has 186600 / 4.0% missing values [Missing](#)
- [Incident\\_Zip](#) is an unsupported type, check if it needs cleaning or further analysis [Warning](#)
- [Intersection\\_Street\\_1](#) has 3677294 / 78.9% missing values [Missing](#)
- [Intersection\\_Street\\_1](#) has a high cardinality: 14422 distinct values [Warning](#)
- [Intersection\\_Street\\_2](#) has 3680138 / 79.0% missing values [Missing](#)
- [Intersection\\_Street\\_2](#) has a high cardinality: 15136 distinct values [Warning](#)
- [Landmark](#) has 4347691 / 93.3% missing values [Missing](#)
- [Latitude](#) is highly correlated with [Y\\_Coordinate\\_\(State\\_Plane\)](#). ( $\rho = 1$ ) [Rejected](#)
- [Location](#) has 305071 / 6.5% missing values [Missing](#)
- [Location](#) has a high cardinality: 770374 distinct values [Warning](#)
- [Location\\_Type](#) has 1035025 / 22.2% missing values [Missing](#)
- [Location\\_Type](#) has a high cardinality: 162 distinct values [Warning](#)
- [Longitude](#) is highly correlated with [X\\_Coordinate\\_\(State\\_Plane\)](#). ( $\rho = 1$ ) [Rejected](#)
- [Park\\_Facility\\_Name](#) has a high cardinality: 3022 distinct values [Warning](#)
- [Resolution\\_Action\\_Updated\\_Date](#) has 89156 / 1.9% missing values [Missing](#)
- [Resolution\\_Action\\_Updated\\_Date](#) has a high cardinality: 1773880 distinct values [Warning](#)
- [Resolution\\_Description](#) has 382186 / 8.2% missing values [Missing](#)
- [Resolution\\_Description](#) has a high cardinality: 1030 distinct values [Warning](#)
- [Road\\_Ramp](#) has 4651054 / 99.8% missing values [Missing](#)
- [Street\\_Name](#) has 642800 / 13.8% missing values [Missing](#)
- [Street\\_Name](#) has a high cardinality: 20788 distinct values [Warning](#)
- [Taxi\\_Company\\_Borough](#) has 4654844 / 99.9% missing values [Missing](#)
- [Taxi\\_Pick-Up\\_Location](#) has 4621825 / 99.2% missing values [Missing](#)
- [Taxi\\_Pick-Up\\_Location](#) has a high cardinality: 15323 distinct values [Warning](#)
- [Vehicle\\_Type](#) has 4658209 / 100.0% missing values [Missing](#)
- [X\\_Coordinate\\_\(State\\_Plane\)](#) has 305071 / 6.5% missing values [Missing](#)
- [Y\\_Coordinate\\_\(State\\_Plane\)](#) has 305071 / 6.5% missing values [Missing](#)

## Variables

### Address\_Type

Categorical

Distinct count	6
Unique (%)	0.0%
Missing (%)	9.8%
Missing (n)	457798

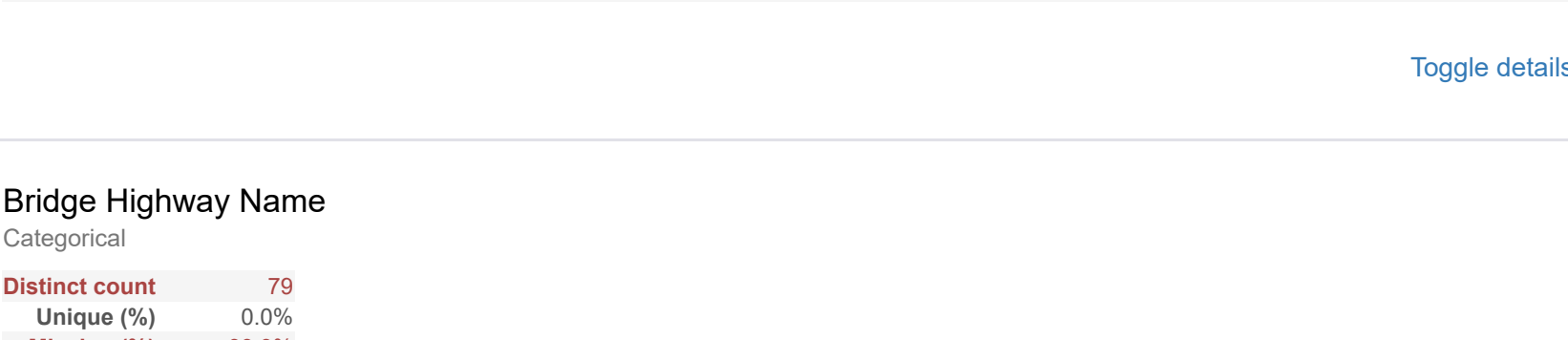


[Toggle details](#)

### Agency

Categorical

Distinct count	29
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

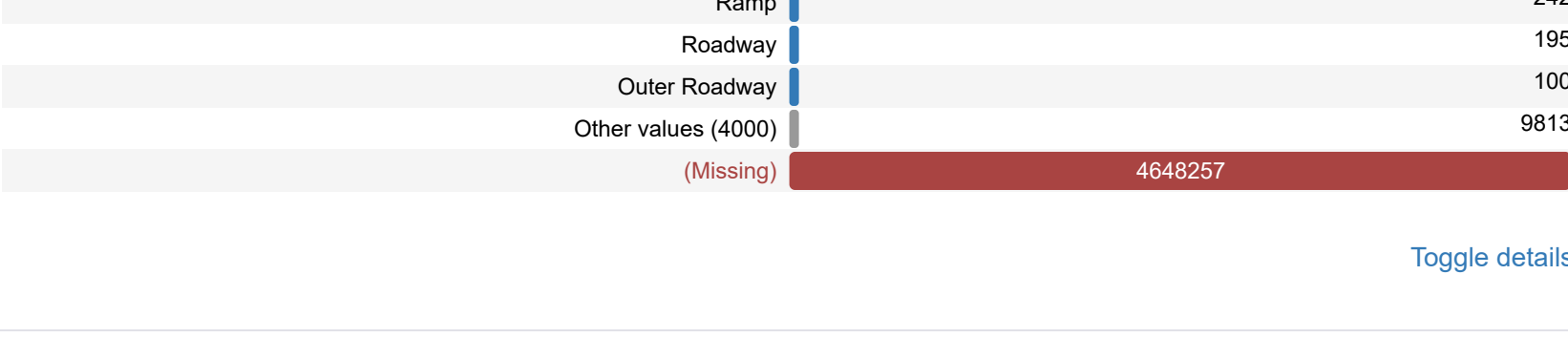


[Toggle details](#)

### Agency\_Name

Categorical

Distinct count	1004
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

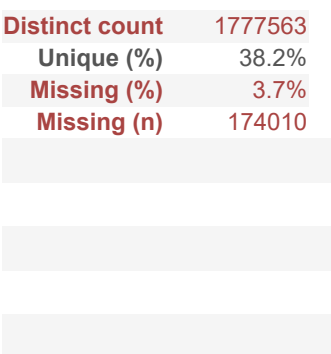


[Toggle details](#)

### BBL

Numeric

Distinct count	544239
Unique (%)	11.7%
Missing (%)	19.0%
Missing (n)	884935
Infinite (%)	0.0%
Infinite (n)	0
Mean	2830800000
Minimum	0
Maximum	5270000000
Zeros (%)	0.0%

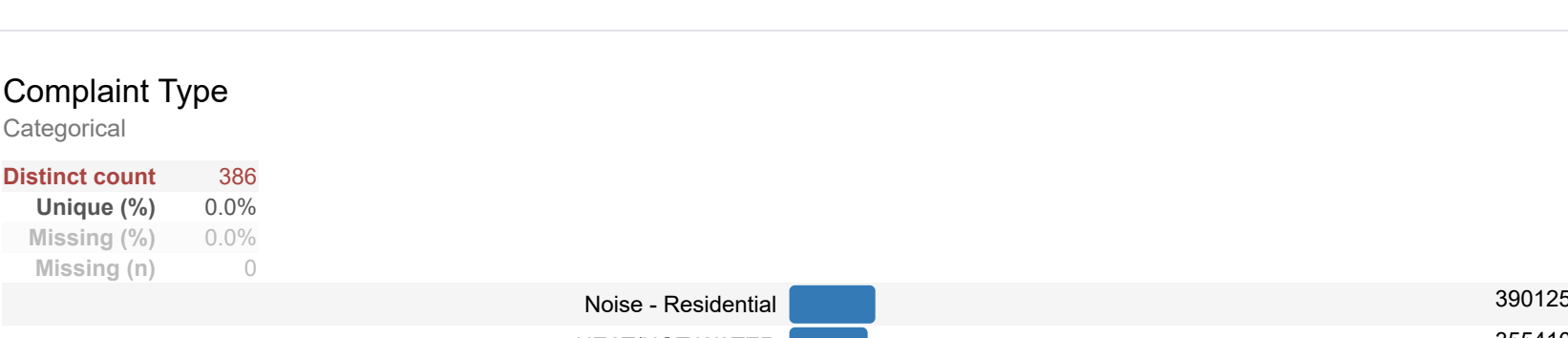


[Toggle details](#)

### Borough

Categorical

Distinct count	6
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

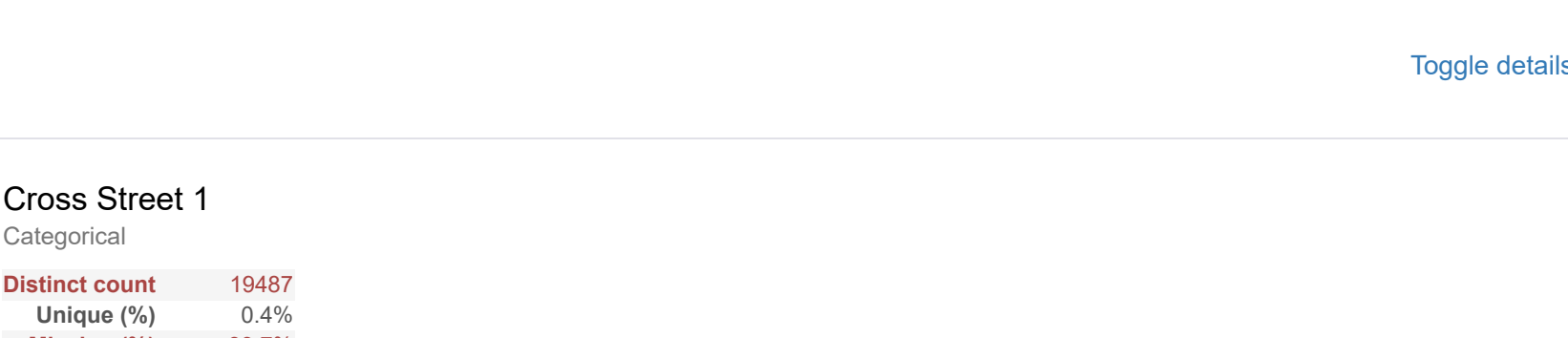


[Toggle details](#)

### Bridge\_Highway\_Direction

Categorical

Distinct count	51
Unique (%)	0.0%
Missing (%)	99.8%
Missing (n)	4651000

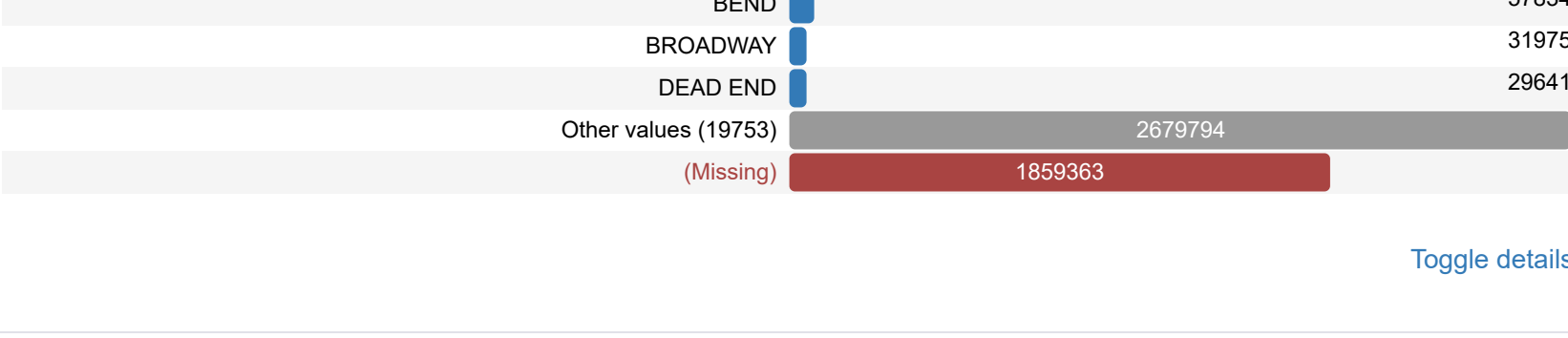


[Toggle details](#)

### Bridge\_Highway\_Name

Categorical

Distinct count	79
Unique (%)	0.0%
Missing (%)	99.8%
Missing (n)	4651697

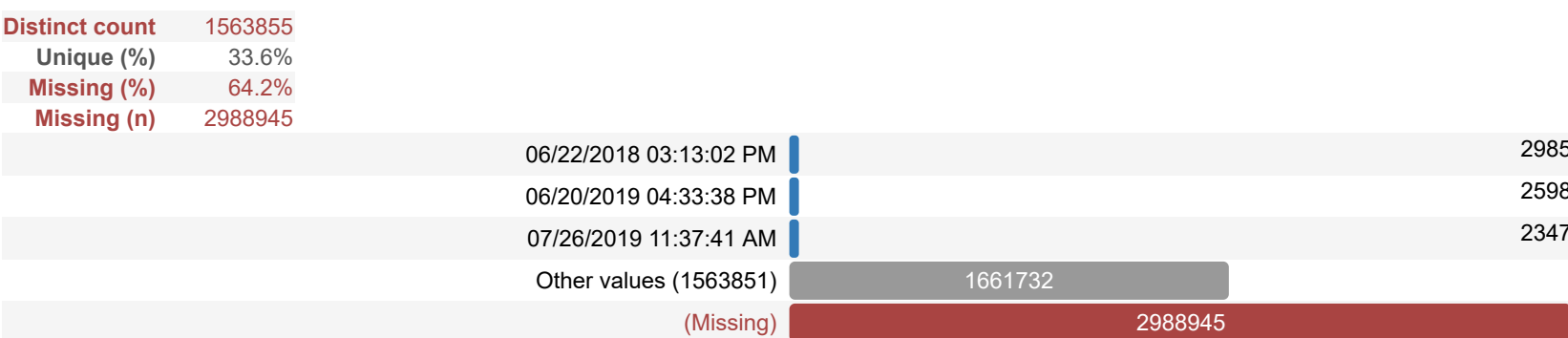


[Toggle details](#)

### Bridge\_Highway\_Segment

Categorical

Distinct count	4004
Unique (%)	0.1%
Missing (%)	99.8%
Missing (n)	4648257

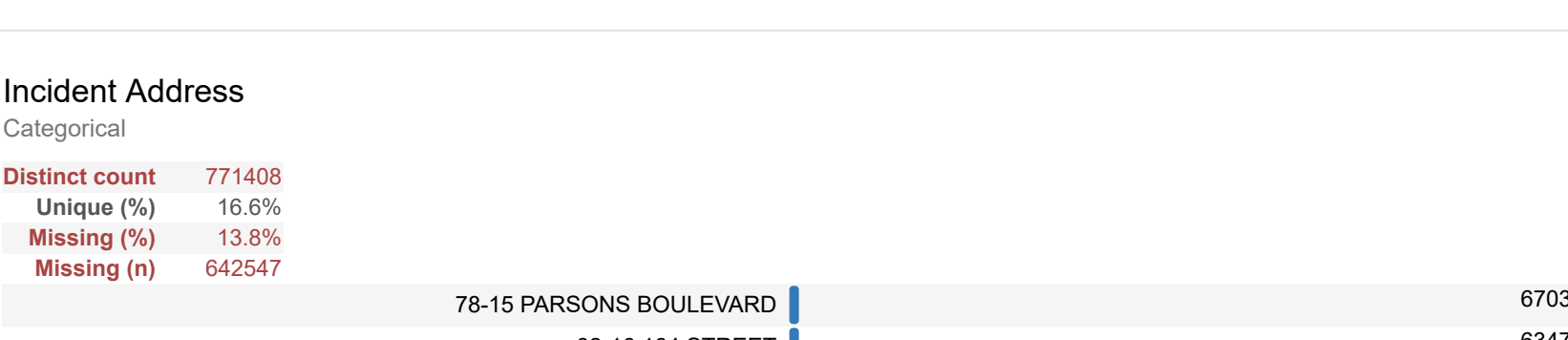


[Toggle details](#)

### City

Categorical

Distinct count	646
Unique (%)	0.0%
Missing (%)	4.8%
Missing (n)	222263

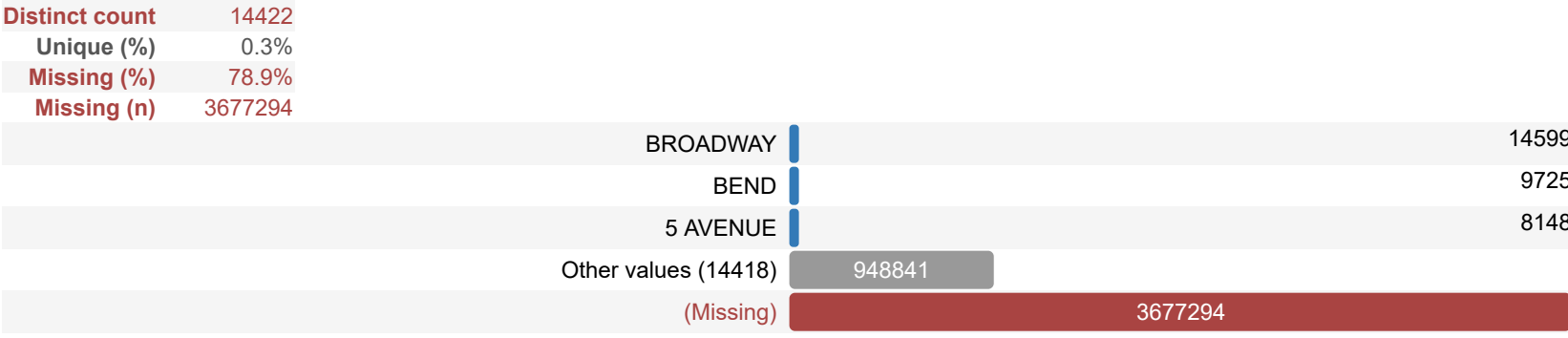


[Toggle details](#)

### Closed\_Date

Categorical

Distinct count	1777563
Unique (%)	38.2%
Missing (%)	13.3%
Missing (n)	174010

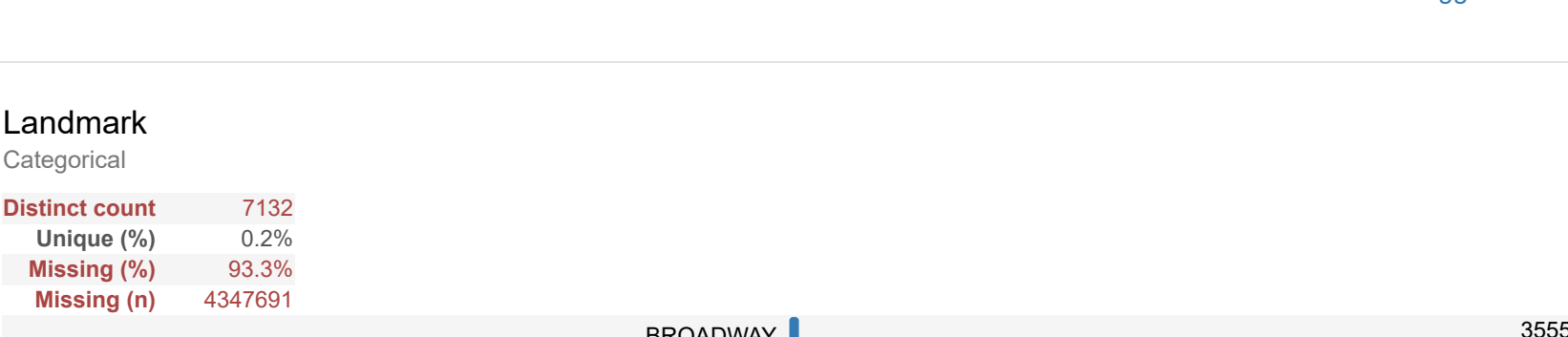


[Toggle details](#)

### Community\_Board

Categorical

Distinct count	77
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

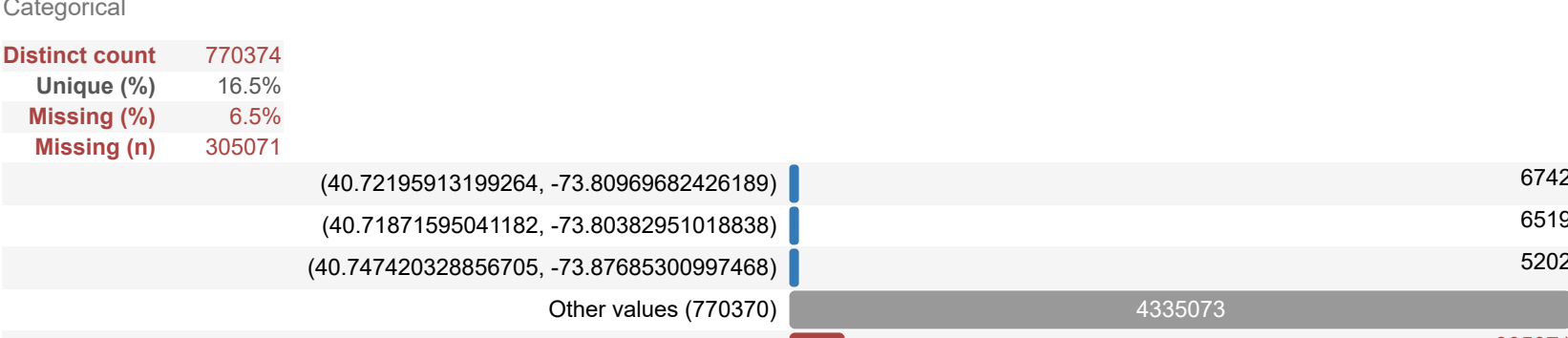


[Toggle details](#)

### Complaint\_Type

Categorical

Distinct count	386
Unique (%)	0.0%
Missing (%)	0.0%
Missing (n)	0

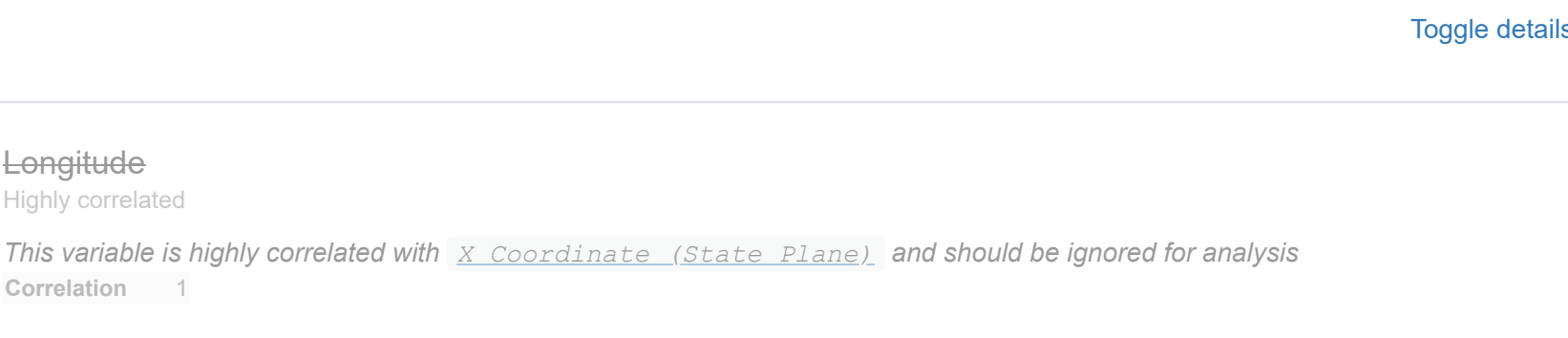


[Toggle details](#)

### Created\_Date

Categorical

Distinct count	3442409
Unique (%)	73.9%
Missing (%)	0.0%
Missing (n)	0

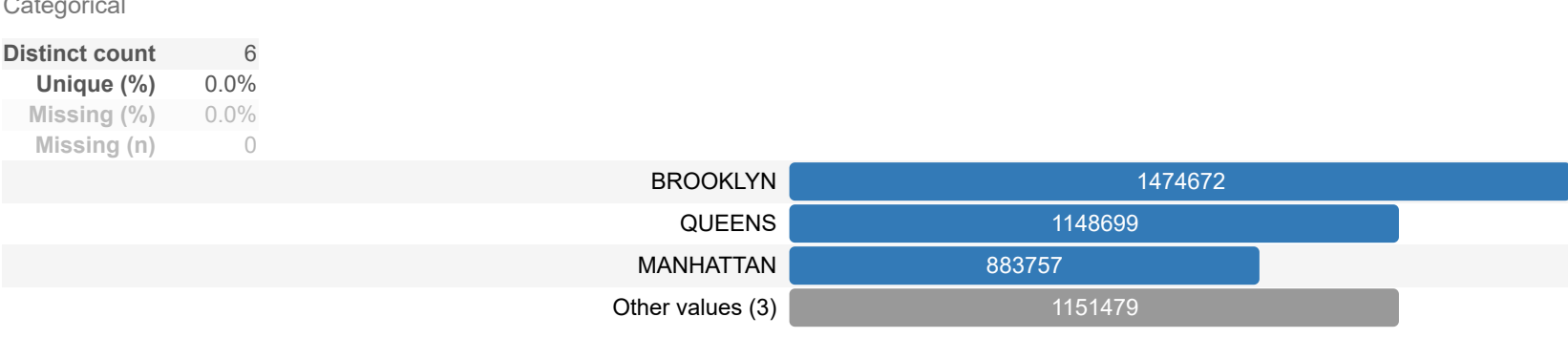


[Toggle details](#)

### Cross\_Street\_1

Categorical

Distinct count	19487
Unique (%)	0.4%
Missing (%)	39.7%
Missing (n)	1848169

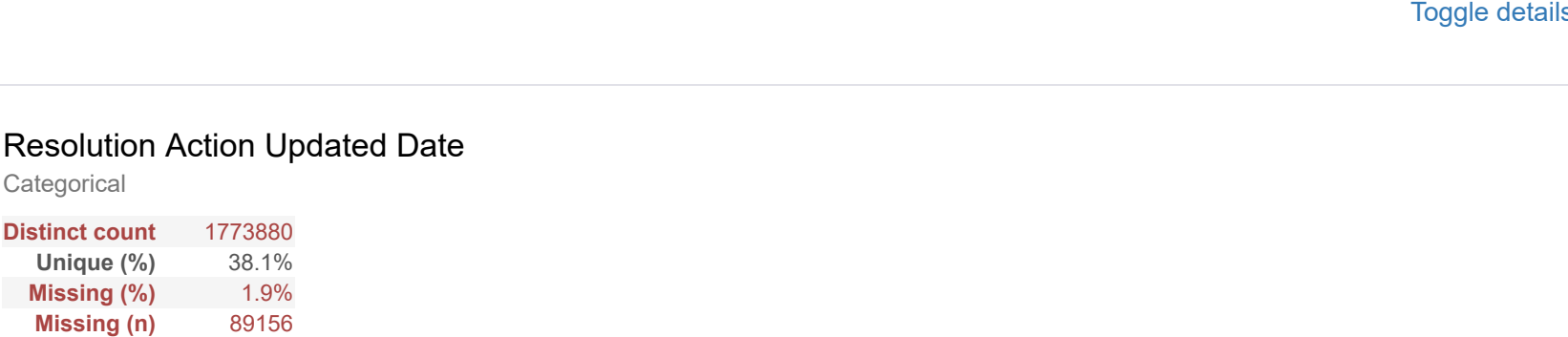


[Toggle details](#)

### Cross\_Street\_2

Categorical

Distinct count	19757
Unique (%)	0.4%
Missing (%)	38.9%
Missing (n)	1859363

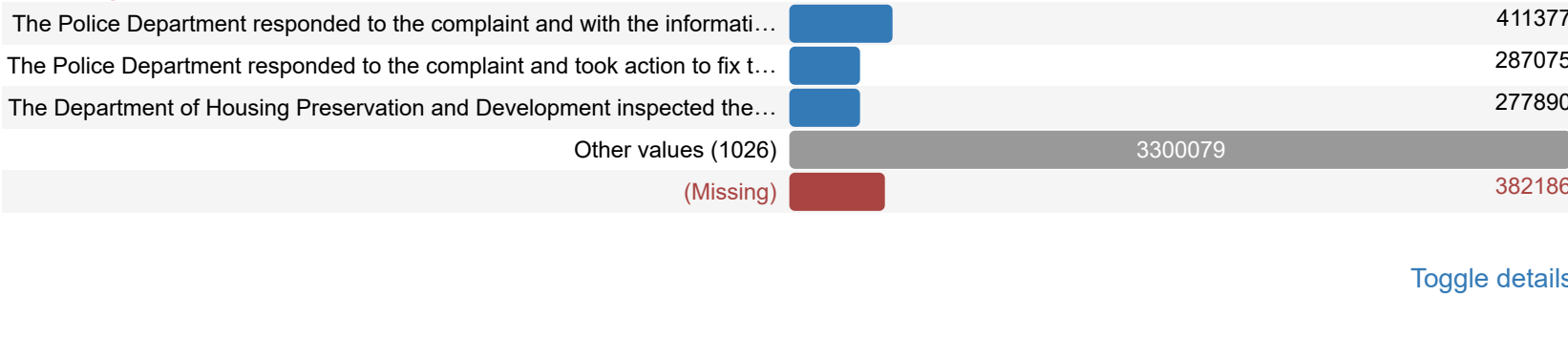


[Toggle details](#)

### Descriptor

Categorical

Distinct count	1322
Unique (%)	0.0%
Missing (%)	1.4%
Missing (n)	65912

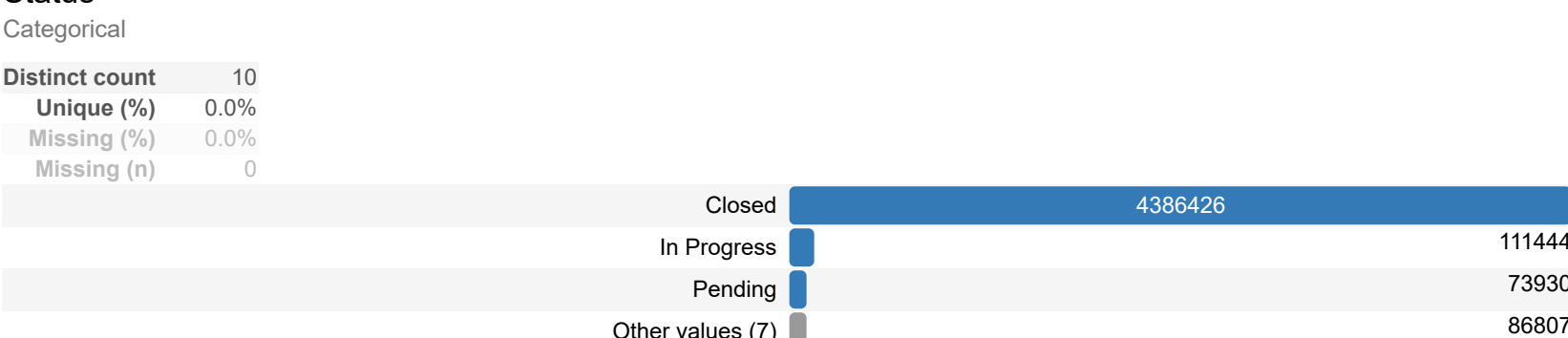


[Toggle details](#)

### Due\_Date

Categorical

Distinct count	1563855
Unique (%)	33.6%
Missing (%)	64.2%
Missing (n)	2988945

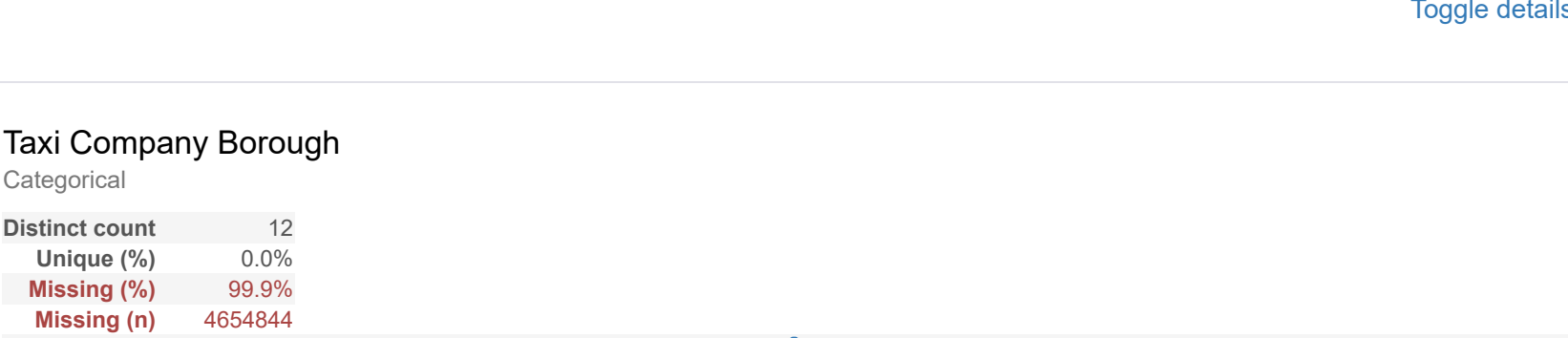


[Toggle details](#)

### Facility\_Type

Categorical

Distinct count	4
Unique (%)	0.0%
Missing (%)	72.7%
Missing (n)	3385624

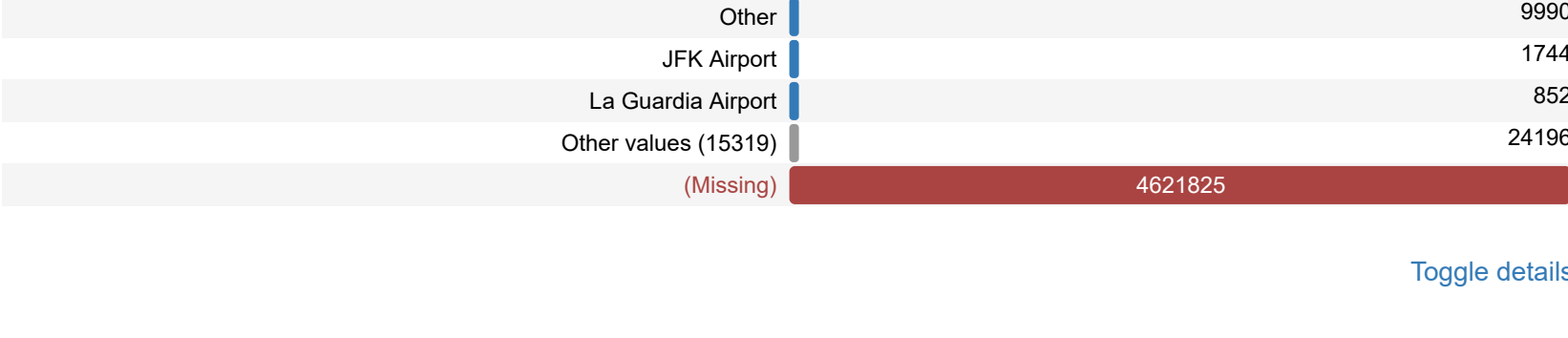


[Toggle details](#)

### Incident\_Address

Categorical

Distinct count	771408
Unique (%)	16.0%
Missing (%)	13.8%
Missing (n)	642547



[Toggle details](#)

### Incident\_Zip

Unsupported

This variable is an unsupported type, check if it needs cleaning or further analysis

Unsupported values: 0

### Intersection\_Street\_1

Categorical

Distinct count	14422
Unique (%)	0.3%
Missing (%)	78.9%
Missing (n)	3677294



[Toggle details](#)

### Intersection\_Street\_2

Categorical

Distinct count	15136
Unique (%)	0.3%
Missing (%)	79.0%
Missing (n)	3680138



[Toggle details](#)

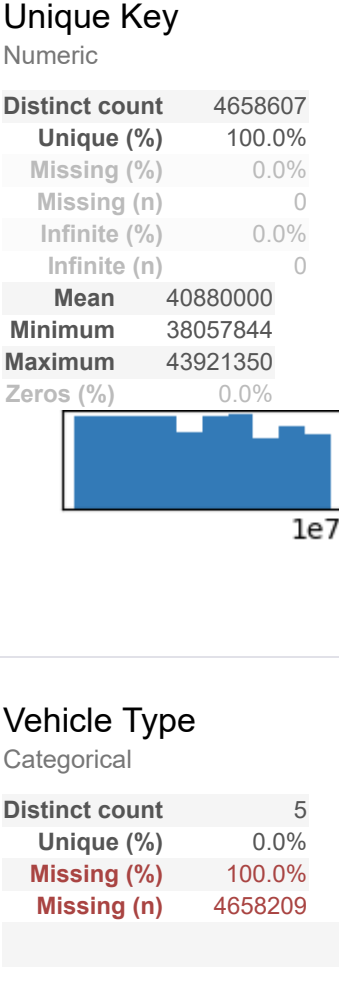
### Landmark

Categorical

Distinct count	7132
Unique (%)	0.2%
Missing (%)	93.3%
Missing (n)	4347691



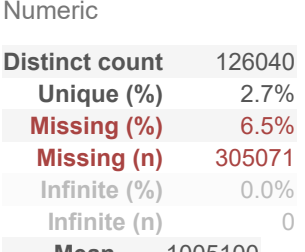




Toggle details

## Vehicle Type

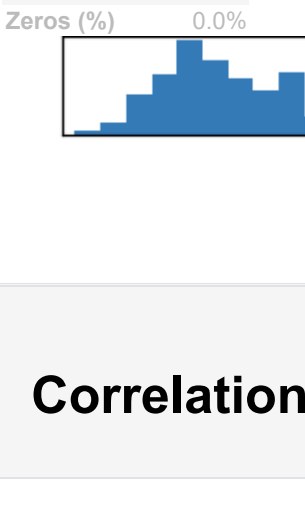
Categorical



Toggle details

## X Coordinate (State Plane)

Numeric



Toggle details

## Y Coordinate (State Plane)

Numeric



Toggle details

## Correlations



## Sample

	Unique Key	Created Date	Closed Date	Agency	Agency Name
0	38060564	12/31/2017 12:00:26 AM	01/08/2018 02:05:27 PM	DOT	Department of Transportation
1	38068568	12/31/2017 12:01:00 AM	01/03/2018 12:00:00 AM	DSNY	Department of Sanitation
2	38062203	12/31/2017 12:01:03 AM	01/13/2018 12:05:43 AM	HPD	Department of Housing Preservation and Develop...
3	38059708	12/31/2017 12:01:03 AM	12/31/2017 01:40:08 AM	NYPD	New York City Police Department
4	38063401	12/31/2017 12:01:07 AM	12/31/2017 07:12:19 AM	NYPD	New York City Police Department

## Quick look in Tableau:

[https://public.tableau.com/profile/shuang.chen7143#!/vizhome/DataDive2019\\_Project\\_311\\_EDA\\_Shuang/ComplaintsbyType-Agency-Division](https://public.tableau.com/profile/shuang.chen7143#!/vizhome/DataDive2019_Project_311_EDA_Shuang/ComplaintsbyType-Agency-Division)