Here is a notebook for Ridgewood NJ Home Prices

The Notebook does the following things

1. Gets home sales information from 2012 through September 2019 a.The data comes from the Village of Ridgewood Website b.The data is in PDF format
   A. Cleans the data from the Village of Ridgewood. a. The notebook imports the various PDF files and converts them to a combined Pandas Data Frame. b. The data is cleaned of issues with spaces column header breaks, Nan values in lieu of zeros.
   B. The cleaned data is saved as CSV a. The CSV is to be consumed by Tableau for Data Visualizations

In [ ]:

```
#install and import of the required modules.
#tabula is used to parse the PDF file format into Pandas
```

In [3]:

```
pip install tabula-py
```

```
Requirement already satisfied: tabula-py in c:\users\khoppe\anaconda3\lib\sit
e-packages (1.4.2)
Requirement already satisfied: distro in c:\users\khoppe\anaconda3\lib\site-p
ackages (from tabula-py) (1.4.0)
Requirement already satisfied: pandas in c:\users\khoppe\anaconda3\lib\site-p
ackages (from tabula-py) (0.24.2)
Requirement already satisfied: numpy in c:\users\khoppe\anaconda3\lib\site-pa
ckages (from tabula-py) (1.16.4)
Requirement already satisfied: python-dateutil>=2.5.0 in c:\users\khoppe\anac
onda3\lib\site-packages (from pandas->tabula-py) (2.8.0)
Requirement already satisfied: pytz>=2011k in c:\users\khoppe\anaconda3\lib\s
ite-packages (from pandas->tabula-py) (2019.1)
Requirement already satisfied: six>=1.5 in c:\users\khoppe\anaconda3\lib\site
-packages (from python-dateutil>=2.5.0->pandas->tabula-py) (1.12.0)
Note: you may need to restart the kernel to use updated packages.
```

In [4]:

```
import pandas as pd
import tabula as tb
print ('Install and Imports Complete')
```

```
Install and Imports Complete
```

In [5]:

```
#Get the data
df_2019 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2019_all.pdf",pages="all")
df_2018 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2018_all.pdf",pages="all")
df_2017 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2017_all.pdf",pages="all")
df_2016 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2016_all.pdf",pages="all")
df_2015 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2015_all.pdf",pages="all")
df_2014 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2014_all.pdf",pages="all")
df_2013 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2013_all.pdf",pages="all")
df_2012 = tb.read_pdf("http://mods.ridgewoodnj.net/pdf/Assmt/2012_all.pdf",pages="all")
```

```
Got stderr: Nov 27, 2019 4:01:18 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:20 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:20 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:21 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:22 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
```

```
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:23 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:24 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
```

```
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:25 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:26 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
```

```
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:27 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:28 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
```

```
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:29 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:30 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
```

```
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:31 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:32 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
```

```
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:33 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:34 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:35 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:36 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:36 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:36 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
```

```
Nov 27, 2019 4:01:36 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
Nov 27, 2019 4:01:36 PM org.apache.pdfbox.pdmodel.font.PDType1Font <init>
WARNING: Using fallback font ArialMT for MS Sans Serif

Got stderr: Nov 27, 2019 4:01:38 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif

Got stderr: Nov 27, 2019 4:01:44 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif

Got stderr: Nov 27, 2019 4:01:49 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif

Got stderr: Nov 27, 2019 4:01:54 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif

Got stderr: Nov 27, 2019 4:02:00 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif

Got stderr: Nov 27, 2019 4:02:05 PM org.apache.pdfbox.pdmodel.font.PDType1Fon
t <init>
WARNING: Using fallback font ArialMT for MS Sans Serif
```

In [6]:

```
#the data needs to be cleaned up.  The pdf import is not perfect and the sourc
e data as nulls

#clean up the data df_2019
df_2019=df_2019[df_2019.Block !='Block']
df_2019=df_2019.drop('Reason',axis=1)
df_2019['Sales Price'] = df_2019['Sales Price'].str.replace("$"," ")
df_2019['Assessment'] = df_2019['Assessment'].str.replace("$"," ")
df_2019['Sales Price'] = df_2019['Sales Price'].str.replace(" ","")
df_2019['Assessment'] = df_2019['Assessment'].str.replace(" ","")
df_2019.fillna(0,inplace=True)

#clean up the data df_2018
df_2018=df_2018[df_2018.Block !='Block']
df_2018=df_2018.drop('Reason',axis=1)
df_2018['Sales Price'] = df_2018['Sales Price'].str.replace("$"," ")
df_2018['Assessment'] = df_2018['Assessment'].str.replace("$"," ")
df_2018['Sales Price'] = df_2018['Sales Price'].str.replace(" ","")
df_2018['Assessment'] = df_2018['Assessment'].str.replace(" ","")
df_2018.fillna(0,inplace=True)

#clean up the data df_2017
df_2017=df_2017[df_2017.Block !='Block']
df_2017=df_2017.drop('Reason',axis=1)
df_2017['Sales Price'] = df_2017['Sales Price'].str.replace("$"," ")
df_2017['Assessment'] = df_2017['Assessment'].str.replace("$"," ")
df_2017['Sales Price'] = df_2017['Sales Price'].str.replace(" ","")
df_2017['Assessment'] = df_2017['Assessment'].str.replace(" ","")
df_2017.fillna(0,inplace=True)

#clean up the data df_2016
df_2016=df_2016[df_2016.Block !='Block']
df_2016=df_2016.drop('Reason',axis=1)
df_2016['Sales Price'] = df_2016['Sales Price'].str.replace("$"," ")
df_2016['Assessment'] = df_2016['Assessment'].str.replace("$"," ")
df_2016['Sales Price'] = df_2016['Sales Price'].str.replace(" ","")
df_2016['Assessment'] = df_2016['Assessment'].str.replace(" ","")
df_2016.fillna(0,inplace=True)

#clean up the data df_2015
df_2015=df_2015[df_2015.Block !='Block']
df_2015=df_2015.drop('Reason',axis=1)
df_2015['Sales Price'] = df_2015['Sales Price'].str.replace("$"," ")
df_2015['Assessment'] = df_2015['Assessment'].str.replace("$"," ")
df_2015['Sales Price'] = df_2015['Sales Price'].str.replace(" ","")
df_2015['Assessment'] = df_2015['Assessment'].str.replace(" ","")
df_2015.fillna(0,inplace=True)

#clean up the data df_2014
df_2014=df_2014[df_2014.Block !='Block']
df_2014=df_2014.drop('Reason',axis=1)
df_2014['Sales Price'] = df_2014['Sales Price'].str.replace("$"," ")
df_2014['Assessment'] = df_2014['Assessment'].str.replace("$"," ")
df_2014['Sales Price'] = df_2014['Sales Price'].str.replace(" ","")
df_2014['Assessment'] = df_2014['Assessment'].str.replace(" ","")
df_2014.fillna(0,inplace=True)
```

```
#clean up the data df_2013 note header name change this year and 2012
df_2013=df_2013[df_2013.Block !='Block']
df_2013=df_2013.drop('NU\rCode',axis=1)
df_2013['Sales Price'] = df_2013['Sales Price'].str.replace("$"," ")
df_2013['Assessment'] = df_2013['Assessment'].str.replace("$"," ")
df_2013['Sales Price'] = df_2013['Sales Price'].str.replace(" ","")
df_2013['Assessment'] = df_2013['Assessment'].str.replace(" ","")
df_2013.fillna(0,inplace=True)

#clean up the data df_2012
df_2012=df_2012[df_2012.Block !='Block']
df_2012=df_2012.drop('NU\rCode',axis=1)
df_2012['Sales Price'] = df_2012['Sales Price'].str.replace("$"," ")
df_2012['Assessment'] = df_2012['Assessment'].str.replace("$"," ")
df_2012['Sales Price'] = df_2012['Sales Price'].str.replace(" ","")
df_2012['Assessment'] = df_2012['Assessment'].str.replace(" ","")
df_2012.fillna(0,inplace=True)
```

```
C:\Users\khoppe\Anaconda3\lib\site-packages\pandas\core\ops.py:1649: FutureWa
rning: elementwise comparison failed; returning scalar instead, but in the fu
ture will perform elementwise comparison
  result = method(y)
```

In [13]:

```
#clean up the names of the columns

df_2019.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Bsmt Area',
 'Fin Bsmt',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2018.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Bsmt Area',
 'Fin Bsmt',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2017.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Bsmt Area',
 'Fin Bsmt',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2016.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2015.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2014.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2013.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
df_2012.columns = ['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
```

In [14]:

```
#The basement area and finished basement stats are not availble from 2016-2012
#Although nice to have I dropped them so I can compare all data across all yea
rs

df_2019=df_2019.drop('Bsmt Area',axis=1)
df_2018=df_2018.drop('Bsmt Area',axis=1)
df_2017=df_2017.drop('Bsmt Area',axis=1)

df_2019=df_2019.drop('Fin Bsmt',axis=1)
df_2018=df_2018.drop('Fin Bsmt',axis=1)
df_2017=df_2017.drop('Fin Bsmt',axis=1)
```

In [17]:

```
#Merge all years into one dataframe
df_All = pd.concat([df_2019, df_2018,df_2017,df_2016,df_2015,df_2014,df_2013,d
f_2012], axis=0)
```

In [19]:

```
#confirm the merge worked
df_All.head()
```

Out[19]:

	Block	Lot	Location	Sales Date	Book	Page	Sales Price	Assessment	Ratio	Use	...	
0	1104	1	801 N MONROE ST	8/5/2019	3338	976	820,000	785,100	95.74%	Single Family	...	:
1	1105	1	1034 HILLCREST RD	8/22/2019	3359	731	840,000	907,700	108.06%	Single Family	...	2
2	1106	6	288 RICHARDS RD	6/27/2019	3302	883	790,000	629,200	79.65%	Single Family	...	2
3	1202	13	915 HILLCREST ROAD	8/15/2019	3367	1594	680,000	791,300	116.37%	Single Family	...	2
4	1202	22	869 HILLCREST RD	1/17/2019	3165	2132	965,000	922,200	95.56%	Single Family	...	2

5 rows × 24 columns

In [20]:

```
#Noticed a problem where the Flood column has both N and No for no
df_All.Flood.replace(['N'], ['No'], inplace=True)
```

In [30]:

```
#Noticed issues with the data quality in the style of house

df_All.Style.replace(['Tear Down'], ['Ranch'], inplace=True)
df_All.Style.replace(['Bi Level'], ['Bilevel'], inplace=True)
df_All.Style.replace(['Bungalo'], ['Bungalow'], inplace=True)
df_All.Style.replace(['Ccape Cod'], ['Cape Cod'], inplace=True)
df_All.Style.replace(['Williamsburg Col'], ['Colonial'], inplace=True)
df_All.Style.replace(['Colonial/Raised Ranc'], ['Colonial'], inplace=True)
df_All.Style.replace(['Colonial/Condo'], ['Condo'], inplace=True)
df_All.Style.replace(['Raised Ranch'], ['Ranch'], inplace=True)
df_All.Style.replace(['2 Fam Condo'], ['Townhouse'], inplace=True)
df_All.Style.replace(['Duplex'], ['Townhouse'], inplace=True)
df_All.Style.replace(['Apt Condo'], ['Condo'], inplace=True)
df_All.Style.replace(['2 Family Condo'], ['Condo'], inplace=True)
df_All.Style.replace(['Expanded Ranch'], ['Ranch'], inplace=True)
df_All.Style.replace(['Exp Ranch'], ['Ranch'], inplace=True)
```

In [33]:
```
#Check to see how the Style column looks now
df_All['Style'].value_counts()
```

Out[33]:
```
Colonial          1605
Cape Cod           330
Split Level        193
Ranch              146
Tudor              138
Cape Ranch          86
Cape Colonial       67
Bungalow            43
Townhouse           39
Bilevel             38
Condo               31
Contemporary         7
Manor Home           2
BiLevel              1
Name: Style, dtype: int64
```

In [34]:
```
#More Sytyle clean up
df_All.Style.replace(['BiLevel'], ['Bilevel'], inplace=True)
df_All.Style.replace(['Contemporary'], ['Other'], inplace=True)
df_All.Style.replace(['Manor Home'], ['Other'], inplace=True)
```

In [35]:
```
#Check to see how the Style column looks now
df_All['Style'].value_counts()
```

Out[35]:
```
Colonial          1605
Cape Cod           330
Split Level        193
Ranch              146
Tudor              138
Cape Ranch          86
Cape Colonial       67
Bungalow            43
Bilevel             39
Townhouse           39
Condo               31
Other                9
Name: Style, dtype: int64
```

In [39]:

```
list(df_All.columns)
```

Out[39]:

```
['Block',
 'Lot',
 'Location',
 'Sales Date',
 'Book',
 'Page',
 'Sales Price',
 'Assessment',
 'Ratio',
 'Use',
 'No Units',
 'Year Built',
 'Style',
 'Story Height',
 'Bldg Area',
 'Total Rooms',
 'Bdrms',
 'Full Baths',
 'Half Baths',
 'Lot Size',
 'Zone',
 'Neigh',
 'Elem School',
 'Flood']
```

In [41]:

```
#Spaces in column Names are hard to deal with
df_All.columns = ['Block',
 'Lot',
 'Location',
 'Sales_Date',
 'Book',
 'Page',
 'Sales_Price',
 'Assessment',
 'Ratio',
 'Use',
 'No_Units',
 'Year_Built',
 'Style',
 'Story_Height',
 'Bldg_Area',
 'Total_Rooms',
 'Bdrms',
 'Full_Baths',
 'Half_Baths',
 'Lot_Size',
 'Zone',
 'Neigh',
 'Elem_School',
 'Flood']
```

In [46]:

```
#Yet more clean up
df_All.Elem_School.replace(['sp'], ['Somerville'], inplace=True)
df_All.Full_Baths.replace(['32'], ['2'], inplace=True)
df_All.Flood.replace(['Y'], ['Yes'], inplace=True)
```

```
C:\Users\khoppe\Anaconda3\lib\site-packages\pandas\core\generic.py:6586: Sett
ingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/st
able/indexing.html#indexing-view-versus-copy
  self._update_inplace(new_data)
```

In [47]:

```
#drop a row with poor data in it
df_All=df_All[df_All.Location != '310 HEIGHTS RD']
```

In [48]:

```
df_All
```

Out[48]:

	Block	Lot	Location	Sales_Date	Book	Page	Sales_Price	Assessment	Ratio
0	1104	1	801 N MONROE ST	8/5/2019	3338	976	820,000	785,100	95.74%
1	1105	1	1034 HILLCREST RD	8/22/2019	3359	731	840,000	907,700	108.06%
2	1106	6	288 RICHARDS RD	6/27/2019	3302	883	790,000	629,200	79.65%
3	1202	13	915 HILLCREST ROAD	8/15/2019	3367	1594	680,000	791,300	116.37%
4	1202	22	869 HILLCREST RD	1/17/2019	3165	2132	965,000	922,200	95.56%
5	1202	27	839 HILLCREST RD	4/19/2019	3248	150	751,000	694,900	92.53%
6	1202	30	819 HILLCREST ROAD	5/30/2019	3283	2307	1,149,000	758,300	66.00%
7	1203	8	828 MORNINGSIDE RD.	7/18/2019	3331	1180	1,160,000	1,148,800	99.03%
8	1205	3	230 RICHARDS RD	7/1/2019	3294	663	1,185,000	916,500	77.34%
9	1205	9	219 HAMILTON RD	6/27/2019	3297	802	892,000	925,000	103.70%
10	1206	7	848 HILLCREST RD	3/11/2019	3205	1141	999,000	1,126,500	112.76%
11	1301	9	447 SHELBOURNE TERR	6/20/2019	3369	187	1,100,000	950,000	86.36%
12	1303	13	750 PARSONS ROAD	6/27/2019	3346	1685	1,170,000	904,400	77.30%
13	1306	20	691 N MONROE ST	8/15/2019	3376	1469	650,000	579,800	89.20%
14	1308	8	714 PARSONS RD	6/24/2019	3300	2159	1,410,000	1,367,700	97.00%
15	1308	15	399 GLENWOOD RD	6/25/2019	3372	2035	1,300,000	955,200	73.48%
16	1309	23	241 BEDFORD RD	4/12/2019	3230	2373	1,550,000	1,226,200	79.11%
17	1311	3	302 GLENWOOD RD	7/26/2019	3347	2325	1,640,000	1,439,800	87.79%
18	1312	14	385 MANCHESTER RD	1/8/2019	3154	773	950,000	943,200	99.28%
19	1313	18.01	562 MORNINGSIDE RD	4/9/2019	3241	993	1,450,000	1,097,800	75.71%

	Block	Lot	Location	Sales_Date	Book	Page	Sales_Price	Assessment	Ratio
20	1313	28	341 FAIRMOUNT RD	5/15/2019	3269	835	750,000	538,900	71.85%
21	1401	12	81 AVONDALE RD	3/14/2019	3212	2435	1,325,000	1,340,900	101.20%
22	1403	7	718 HILLCREST RD	1/23/2019	3165	850	799,000	706,200	88.39%
23	1404	12	45 GLENWOOD RD	8/14/2019	3338	846	735,000	706,800	96.16%
24	1405	2	769 UPPER BLVD	4/5/2019	3230	147	715,000	687,500	96.15%
25	1406	10	606 HEIGHTS ROAD	3/20/2019	3231	1521	775,000	848,400	109.47%
26	1406	12	119 CALIFORNIA ST	4/18/2019	3248	293	950,000	799,100	84.12%
27	1406	13	125 CALIFORNIA ST	6/10/2019	3293	1875	1,137,500	1,000,000	87.91%
28	1407	4	636 HEIGHTS RD	8/19/2019	3348	522	750,000	615,200	82.03%
29	1408	10	33 SHERWOOD RD	6/17/2019	3303	2069	1,299,000	969,200	74.61%
...	...	...	...	...	...	...	...	...	...
321	4707	18	640 KENWOOD RD	5/17/2012	1057	2152	580,000	653,000	112.59%
322	4707	30	685 TERHUNE RD	2/8/2012	965	107	455,000	583,000	128.13%
323	4708	3	524 E SADDLE RIVER RD	8/29/2012	1161	850	550,000	617,600	112.29%
324	4709	9	706 TERHUNE RD	8/3/2012	1129	1603	740,000	761,700	102.93%
325	4709	10	705 KINGSBRIDGE LA	11/1/2012	1234	830	525,000	625,800	119.20%
326	4709	15	655 KINGSBRIDGE LA	8/15/2012	1132	1560	480,000	515,900	107.48%
327	4801	6	310 EASTBROOK RD	7/26/2012	1153	249	770,000	807,200	104.83%
328	4804	3	281 EASTBROOK RD	1/31/2012	951	868	685,000	795,200	116.09%
329	4804	4	291 EASTBROOK RD	9/7/2012	1172	323	780,000	846,400	108.51%

	Block	Lot	Location	Sales_Date	Book	Page	Sales_Price	Assessment	Ratio
330	4805	1	311 EASTBROOK RD	3/29/2012	1010	2058	690,000	725,000	105.07%
331	4903	13	877 NORGATE DR	5/15/2012	1054	1692	595,500	689,600	115.80%
333	4906	5	812 NORGATE DR	7/9/2012	1109	1957	517,000	625,700	121.03%
334	4906	11	813 BINGHAM RD	11/15/2012	1235	1151	635,000	700,200	110.27%
335	4907	3	583 EASTBROOK RD	5/2/2012	1049	366	519,000	581,800	112.10%
336	4908	6	914 NORGATE DR	3/8/2012	998	1705	440,000	662,400	150.55%
337	4908	7	926 NORGATE DR	11/30/2012	1242	1019	700,000	747,800	106.83%
338	4908	34	622 EASTBROOK RD	9/19/2012	1178	1019	555,000	604,300	108.88%
339	4908	42	542 EASTBROOK RD	4/30/2012	1049	2208	603,000	668,200	110.81%
340	4908	53	577 WESTBROOK RD	10/16/2012	1260	1938	560,000	527,700	94.23%
341	4912	7	376 WILLIAM ST.	3/20/2012	1001	2325	465,000	522,300	112.32%
342	4912	14	393 JEFFERSON ST	12/14/2012	1259	1548	467,500	515,800	110.33%
343	5001	12.02	326 JEFFERSON ST	9/7/2012	1180	1456	670,000	746,800	111.46%
344	5003	15	256 VAN EMBURGH AVE	10/9/2012	1200	776	419,000	512,200	122.24%
345	5003	27	837 AUBURN AVE	5/24/2012	1193	2012	350,000	486,400	138.97%
346	5004	21	255 VAN EMBURGH AVE	1/9/2012	929	198	422,500	437,400	103.53%
347	5004	22.01	257 VAN EMBURGH AVE	7/30/2012	1117	2292	405,000	546,200	134.86%
348	5004	22.04	27 THEYKEN PL	2/1/2012	968	95	810,000	876,000	108.15%
349	5005	8	210 GATEWAY RD	5/24/2012	1063	196	452,000	544,700	120.51%
350	5006	22	237 GATEWAY RD	5/17/2012	1049	265	792,500	820,800	103.57%

	Block	Lot	Location	Sales_Date	Book	Page	Sales_Price	Assessment	Ratio
351	5006	27	259 GATEWAY RD	5/23/2012	1062	454	595,000	667,500	112.18%

2724 rows × 24 columns

In [53]:

```
#Check to see how many Locations were sold more than once in dataset
df_multi = (df_All['Location'].value_counts())
```

```
In [54]: df_multi
```

```
Out[54]: 140 BELLAIR RD            9
         64 PARK SLOPE             8
         7 LIBERTY ST              5
         355 GLENWOOD RD           3
         523 UPPER BLVD            3
         214 FAIRFIELD AVE         3
         430 BOGERT AVE            3
         176 COTTAGE PL            3
         419 UPPER BLVD            3
         235 DEMAREST ST           3
         816 PARSONS RD            3
         344 GRANDVIEW CIRCLE      3
         210 ORCHARD PL            3
         186 MC KINLEY PL          3
         208 DOREMUS AVE           3
         315 WALTHERY AVE          3
         1023 HILLCREST RD         3
         221 EMMETT PL             3
         323 WALTHERY AVE          3
         387 BERKSHIRE RD          3
         178 N PLEASANT AVE        3
         615 GROVE ST              3
         458 SHEFFIELD RD          3
         621 ALANON RD             3
         560 VAN BUREN ST          3
         643 MIDWOOD RD            3
         249 LOCKWOOD RD           3
         61 WARREN PL              3
         36 RICHMOND AVE           3
         135 SUNSET AVENUE         3
                                  ..
         635 N MONROE ST           1
         451 GOFFLE ROAD           1
         441 GEORGE ST             1
         14 MAYNARD CT             1
         241 HOPE ST               1
         126 SUNSET AVE            1
         475 DORCHESTER RD         1
         286 HIGHLAND AVE          1
         244 CANTERBURY PL         1
         746 FERNWOOD CT           1
         244 S PLEASANT AVE        1
         265 GOFFLE ROAD           1
         437 UPPER BLVD            1
         452 DORCHESTER RD         1
         111 WALTHERY AVE          1
         190 ORCHARD PL            1
         209 S BROAD ST            1
         472 BEVERLY RD            1
         97 MADISON PL             1
         35 GARFIELD PL.           1
         259 HIGHWOOD AVE          1
         371 GILBERT ST            1
         341 FAIRMOUNT RD          1
         311 ALLEN PL              1
         47 ETHELBERT PL           1
         160 FAIRMOUNT RD          1
```

```
238 OLIVIA ST           1
297 MOUNTAIN AVE        1
231 PHELPS RD           1
943 E RIDGEWOOD AVE     1
Name: Location, Length: 2363, dtype: int64
```

In [55]:

```
#drop another location with poor data in it
df_All=df_All[df_All.Location != '64 PARK SLOPE']
```

In [56]:

```
#save as csv so I can have a look in Excel if I want and save a nice clean cop
y somewhere
df_All.to_csv(r'C:\Testing\Ridgewood_House_Prices.csv')
```

In [57]:

```
#Decided some folks might be angry to know the price of their home is listed p
ublicly so removed
#columns pinpointing precise house location
#and saved a clean version to csv
df_All_Clean = df_All.drop(['Location','Block','Lot','Book','Page','Zone','Nei
gh'],axis=1)
df_All_Clean.to_csv(r'C:\Testing\Ridgewood_House_Prices_Clean.csv')
```

In [58]:

```
#check if it looks good now
df_All_Clean.head()
```

Out[58]:

	Sales_Date	Sales_Price	Assessment	Ratio	Use	No_Units	Year_Built	Style	Story_
0	8/5/2019	820,000	785,100	95.74%	Single Family	1	1974	Colonial	
1	8/22/2019	840,000	907,700	108.06%	Single Family	1	1941	Cape Cod	
2	6/27/2019	790,000	629,200	79.65%	Single Family	1	1947	Cape Cod	
3	8/15/2019	680,000	791,300	116.37%	Single Family	1	1961	Cape Cod	
4	1/17/2019	965,000	922,200	95.56%	Single Family	1	1931	Colonial	