In [1]:

**import** pandas **as** pd

**import** numpy **as** np

**import** matplotlib.pyplot **as** plt

**%**matplotlib inline

**%**reload\_ext autoreload

**%**autoreload 2

# Reading the Data and looking through it

In [2]:

df **=** pd.read\_csv('apartment\_cost\_list.csv') df.head(3)

Out[2]:

**Job # Borough House #**

**Street Name**

**Block Lot Bin # Job**

**Type**

**Community**

**- Board**

**Curb Cut**

**0** 123768976 MANHATTAN 1598 2ND

AVENUE

1545 49 1049907 A2 108 NaN

**1** 140922893 MANHATTAN 160 E 89 ST 1517 43 1048094 A2 108 NaN

**2** 301773579 BROOKLYN 65 GATES

AVENUE

1962 86 3056039 A1 302 NaN

In [3]:

df.columns

Out[3]:

Index(['Job #', 'Borough', 'House #', 'Street Name', 'Block', 'Lot', 'Bin #',

'Job Type', 'Community - Board', 'Curb Cut', 'Fully Permitted', 'Initial Cost', 'Proposed Zoning Sqft', 'Horizontal Enlrgmt',

'Vertical Enlrgmt', 'Enlargement SQ Footage', 'Zoning Dist1', 'Job Description'],

dtype='object')

|  |  |
| --- | --- |
| In [4]: |  |
| df.shape |  |
| Out[4]: |  |
| (19613, 18) |  |
| In [5]: |  |
| pd.isnull(df).sum() |  |
| Out[5]: |  |
| Job # | 0 |
| Borough | 0 |
| House # | 0 |
| Street Name | 0 |
| Block | 0 |
| Lot | 0 |
| Bin # | 0 |
| Job Type | 0 |
| Community - Board | 0 |
| Curb Cut | 19597 |
| Fully Permitted | 3988 |
| Initial Cost | 0 |
| Proposed Zoning Sqft | 0 |
| Horizontal Enlrgmt | 19450 |
| Vertical Enlrgmt | 19371 |
| Enlargement SQ Footage | 0 |
| Zoning Dist1 | 2096 |
| Job Description  dtype: int64 | 0 |

In [6]:

df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 19613 entries, 0 to 19612 Data columns (total 18 columns):

# Column Non-Null Count Dtype

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 |  | Job # | 19613 | non-null |  | int64 |
| 1 |  | Borough | 19613 | non-null |  | object |
| 2 |  | House # | 19613 | non-null |  | object |
| 3 |  | Street Name | 19613 | non-null |  | object |
| 4 |  | Block | 19613 | non-null |  | int64 |
| 5 |  | Lot | 19613 | non-null |  | int64 |
| 6 |  | Bin # | 19613 | non-null |  | int64 |
| 7 |  | Job Type | 19613 | non-null |  | object |
| 8 |  | Community - Board | 19613 | non-null |  | int64 |

1. Curb Cut 16 non-null object
2. Fully Permitted 15625 non-null object
3. Initial Cost 19613 non-null object
4. Proposed Zoning Sqft 19613 non-null int64

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 13 Horizontal Enlrgmt | | 163 non-null | | object |
| 14 Vertical Enlrgmt | | 242 non-null | | object |
| 15 Enlargement SQ Footage | | 19613 | non-null | int64 |
| 16 Zoning Dist1 | | 17517 | non-null | object |
| 17 Job Description | | 19613 | non-null | object |
| dtypes: int64(7), object(11) memory usage: 2.7+ MB | | | | |
| In [7]: |  | | | |
| df.nunique(axis**=**0) |  | | | |
| Out[7]: |  | | | |
| Job # | 13403 | | | |
| Borough | 5 | | | |
| House # | 1959 | | | |
| Street Name | 2580 | | | |
| Block | 2493 | | | |
| Lot | 265 | | | |
| Bin # | 6819 | | | |
| Job Type | 3 | | | |
| Community - Board | 59 | | | |
| Curb Cut | 1 | | | |
| Fully Permitted | 4392 | | | |
| Initial Cost | 46 | | | |
| Proposed Zoning Sqft | 331 | | | |
| Horizontal Enlrgmt | 1 | | | |
| Vertical Enlrgmt | 2 | | | |
| Enlargement SQ Footage | 229 | | | |
| Zoning Dist1 | 145 | | | |
| Job Description  dtype: int64 | 12871 | | | |

Out[8]:

**Job # Block Lot Bin # Community -**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | **Board** | **Zoning Sqft** |
| **count** | 1.961300e+04 | 19613.000000 | 19613.000000 | 1.961300e+04 | 19613.000000 | 19613.000000 |
| **mean** | 1.624266e+08 | 1551.725794 | 1261.811197 | 1.506365e+06 | 151.269005 | 719.625809 |
| **std** | 9.157722e+07 | 1550.780753 | 2764.626109 | 9.319992e+05 | 90.115653 | 10859.926278 |
| **min** | 1.021331e+08 | 0.000000 | 0.000000 | 1.000043e+06 | 101.000000 | 0.000000 |
| **25%** | 1.048244e+08 | 816.000000 | 17.000000 | 1.028758e+06 | 105.000000 | 0.000000 |
| **50%** | 1.217112e+08 | 1245.000000 | 36.000000 | 1.046983e+06 | 107.000000 | 0.000000 |
| **75%** | 1.408258e+08 | 1576.000000 | 66.000000 | 1.084564e+06 | 111.000000 | 0.000000 |
| **max** | 5.401803e+08 | 16278.000000 | 9007.000000 | 5.860016e+06 | 503.000000 | 724475.000000 |

**Proposed**

In [9]:

df['Initial Cost'] **=** df['Initial Cost'].str.replace('$', '')

C:\Users\HP\AppData\Local\Temp\ipykernel\_11480\1069720601.py:1: FutureWarn ing: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will \*not\* be t reated as literal strings when regex=True.

df['Initial Cost'] = df['Initial Cost'].str.replace('$', '')

In [10]:

df['Initial Cost'] **=** pd.to\_numeric(df['Initial Cost'])

<class 'pandas.core.frame.DataFrame'> RangeIndex: 19613 entries, 0 to 19612 Data columns (total 18 columns):

# Column Non-Null Count Dtype

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 |  | Job # | 19613 | non-null |  | int64 |
| 1 |  | Borough | 19613 | non-null |  | object |
| 2 |  | House # | 19613 | non-null |  | object |
| 3 |  | Street Name | 19613 | non-null |  | object |
| 4 |  | Block | 19613 | non-null |  | int64 |
| 5 |  | Lot | 19613 | non-null |  | int64 |
| 6 |  | Bin # | 19613 | non-null |  | int64 |
| 7 |  | Job Type | 19613 | non-null |  | object |
| 8 |  | Community - Board | 19613 | non-null |  | int64 |

1. Curb Cut 16 non-null object
2. Fully Permitted 15625 non-null object
3. Initial Cost 19613 non-null float64
4. Proposed Zoning Sqft 19613 non-null int64

|  |  |  |  |
| --- | --- | --- | --- |
| 13 Horizontal Enlrgmt | 163 non-null | | object |
| 14 Vertical Enlrgmt | 242 non-null | | object |
| 15 Enlargement SQ Footage | 19613 | non-null | int64 |
| 16 Zoning Dist1 | 17517 | non-null | object |
| 17 Job Description | 19613 | non-null | object |

dtypes: float64(1), int64(7), object(10) memory usage: 2.7+ MB

In [12]:

df.describe()

Out[12]:

**Job # Block Lot Bin # Community -**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | | | | **Board** |  |
| **count** | 1.961300e+04 | 19613.000000 | 19613.000000 | 1.961300e+04 | 19613.000000 | 19613.000000 |
| **mean** | 1.624266e+08 | 1551.725794 | 1261.811197 | 1.506365e+06 | 151.269005 | 59442.155713 |
| **std** | 9.157722e+07 | 1550.780753 | 2764.626109 | 9.319992e+05 | 90.115653 | 26687.304059 |
| **min** | 1.021331e+08 | 0.000000 | 0.000000 | 1.000043e+06 | 101.000000 | 40000.000000 |
| **25%** | 1.048244e+08 | 816.000000 | 17.000000 | 1.028758e+06 | 105.000000 | 45000.000000 |
| **50%** | 1.217112e+08 | 1245.000000 | 36.000000 | 1.046983e+06 | 107.000000 | 57000.000000 |
| **75%** | 1.408258e+08 | 1576.000000 | 66.000000 | 1.084564e+06 | 111.000000 | 70000.000000 |
| **max** | 5.401803e+08 | 16278.000000 | 9007.000000 | 5.860016e+06 | 503.000000 | 490000.000000 |

**Initial Cost**

Index(['Job #', 'Borough', 'House #', 'Street Name', 'Block', 'Lot', 'Bin #',

'Job Type', 'Community - Board', 'Curb Cut', 'Fully Permitted', 'Initial Cost', 'Proposed Zoning Sqft', 'Horizontal Enlrgmt',

'Vertical Enlrgmt', 'Enlargement SQ Footage', 'Zoning Dist1', 'Job Description'],

dtype='object')

In [14]:

df.corr()

Out[14]:

**Job # Block Lot Bin #**

**Community**

**- Board**

**Initial Cost**

**Proposed En Zoning S**

**Sqft**

**Job #** 1.000000 0.476468 -0.155658 0.989301 0.990595 0.004833 0.074731

**Block** 0.476468 1.000000 -0.128026 0.508842 0.501237 -0.005337 0.043344

**Lot** -0.155658 -0.128026 1.000000 -0.149478 -0.160610 0.017577 -0.010634

**Bin #** 0.989301 0.508842 -0.149478 1.000000 0.997911 0.002150 0.073943

**Community**

**- Board** 0.990595 0.501237 -0.160610 0.997911 1.000000 0.001022 0.073728

**Initial Cost** 0.004833 -0.005337 0.017577 0.002150 0.001022 1.000000 -0.004520

**Proposed**

**Zoning Sqft** 0.074731 0.043344 -0.010634 0.073943 0.073728 -0.004520 1.000000

**Enlargement**

**SQ Footage** 0.054551 0.050316 -0.021025 0.063730 0.063738 0.000681 0.010333

|  |  |  |
| --- | --- | --- |
| In [15]: |  | |
| *#Curb Cut* | *19597* |  |
| *#Fully Permitted* | *3988* |  |
| *#Horizontal Enlrgmt* | *19450* |  |
| *#Vertical Enlrgmt* | *19371* |  |
| *#Zoning Dist1* | *2096* |  |
| *#Job Description*  df **=** df.drop(['Curb | *0*  Cut','Fully Permitted','Horizontal Enlrgmt','Vertical | Enlrgmt','Zonin |
| In [16]: |  |  |
| df.shape |  |  |
| Out[16]: |  |  |
| (19613, 12) |  |  |

df.head(3)

Out[17]:

**Job # Borough House #**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | **Type** | **- Board** | **Cos** |
| 1545 | 49 | 1049907 | A2 | 108 | 65000.0 |
| 1517 | 43 | 1048094 | A2 | 108 | 85000.0 |
| 1962 | 86 | 3056039 | A1 | 302 | 60000.0 |

**Street Name**

**Block Lot Bin # Job**

**Community**

**Initia**

**0** 123768976 MANHATTAN 1598 2ND

AVENUE

**1** 140922893 MANHATTAN 160 E 89 ST

**2** 301773579 BROOKLYN 65 GATES

AVENUE

|  |  |  |
| --- | --- | --- |
| In | [18]: | |
| df.nunique(axis**=**0) | |  |
| Out[18]: | |  |
| Job # | | 13403 |
| Borough | | 5 |
| House # | | 1959 |
| Street Name | | 2580 |
| Block | | 2493 |
| Lot | | 265 |
| Bin # | | 6819 |
| Job Type | | 3 |
| Community - Board | | 59 |
| Initial Cost | | 46 |
| Proposed Zoning Sqft | | 331 |
| Enlargement SQ Footage  dtype: int64 | | 229 |
| In | [19]: | |

df **=** df.drop(['Job #','House #','Street Name','Block','Bin #'],axis**=**1)

In [20]:

df.head()

Out[20]:

**Borough Lot Job**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Type** | **Board** | **Cost** | **Zoning Sqft** | **Footage** |
| **0** MANHATTAN 49 | A2 | 108 | 65000.0 | 0 | 0 |
| **1** MANHATTAN 43 | A2 | 108 | 85000.0 | 0 | 0 |
| **2** BROOKLYN 86 | A1 | 302 | 60000.0 | 0 | 0 |
| **3** BROOKLYN 35 | A2 | 314 | 75000.0 | 0 | 0 |
| **4** BROOKLYN 10 | A1 | 301 | 40000.0 | 0 | 0 |

**Community -**

**Initial**

**Proposed**

**Enlargement SQ**

|  |  |
| --- | --- |
| df.shape |  |
| Out[21]: |  |
| (19613, 7) |  |
| In [22]: |  |
| df.isnull().sum() |  |
| Out[22]: |  |
| Borough | 0 |
| Lot | 0 |
| Job Type | 0 |
| Community - Board | 0 |
| Initial Cost | 0 |
| Proposed Zoning Sqft | 0 |
| Enlargement SQ Footage  dtype: int64 | 0 |
| In [23]: |  |
| df.dtypes |  |
| Out[23]: |  |
| Borough | object |
| Lot | int64 |
| Job Type | object |
| Community - Board | int64 |
| Initial Cost | float64 |
| Proposed Zoning Sqft | int64 |
| Enlargement SQ Footage  dtype: object | int64 |