



# ETL Report

---

**Cy Edmonds, Francis Kabongo, Huda Awad**

## ETL Report

### Extract:

Pulled down the files into an excel document from NYC Department of Finance, <https://www1.nyc.gov/site/finance/taxes/property-annualized-sales-update.page>. In the Detailed Annual Sales Reports by Borough, we grabbed the 2016, 2017, and 2018 data files for the five boroughs: Manhattan, Bronx, Brooklyn, Queens, and Staten Islands.

### Transform:

Took the excel sheets and pulled them into python. We concatenated all of the files for the same boroughs of the different years. Then to clean the data, we used pandas. We identified the number of the boroughs with the names and switched out all the numbers for the names (i.e 1 being Manhattan, 2 being Bronx, 3 Brooklyn, etc.). In all, for this step we changed the data type of the borough column from an integer to a string.

We continued by finding the different data types to ensure that we have a consistent run through with all of the data. Noticing that the data sets had nulls, we changed the nulls to zeros. We've found that the data types were similar for the data pulled from 2016 and 2017, but then in 2018 the columns Zip Codes to Years Built were in as float64, and to keep the uniformity we changed it from float64 to int64.

To remove any data that we were not using, we dropped the Easement columns from each of the borough's data set as it did not have any information in the rows and did not provide anything for the data sets.

Cleaning up some more, we took the column names and unified them with the code: `rolling_2018.columns = rolling_2018.columns.str.strip().str.replace(' ', '_')`. This easily took all the columns for each of the files and added an underscore( \_ ) for every space in between the words.

### Load:

To load our data, we have decided to use a Relational Database (SQL), since the data was being represented in tables and rows. With it being a very large dataset, more than 80,000 rows, we could simply import the data into a SQL database and quickly query anything out that we would like to identify from combining the different datasets.

## ETL Report

This is the data set that we have retrieved from the NYC Department of Finance for 2018.

BOROUGH	NEIGHBORHOOD	BUILDING CLASS CATEGORY	TAX CLASS AS OF FINAL ROLL 18/19	BLOCK	LOT	EASEMENT OF FINAL ROLL 18/19	ADDRESS	APARTMENT NUMBER	ZIP CODE	RESIDENTIAL UNITS	COMMERCIAL UNITS	TOTAL UNITS
1	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	390	61	A4	189 EAST 7TH STREET		10009	1	0	1
1	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	390	61	A4	189 EAST 7TH STREET		10009	1	0	1
1	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	400	19	A4	526 EAST 5TH STREET		10009	1	0	1
1	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	390	35	B1	113 AVENUE C		10009	2	0	2
1	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	404	1	B9	166 AVENUE A		10009	2	0	2
1	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	404	1	B9	166 AVENUE A		10009	2	0	2
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	379	37	C7	133 AVENUE D		10009	20	2	22
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	385	2	C7	21-23 AVENUE B		10009	12	4	16
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	389	18	C1	200 EAST 7TH STREET		10009	18	0	18
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	390	50	C1	209 EAST 7TH STREET, 1W		10009	29	0	29
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	390	54	C3	203 EAST 7TH STREET		10009	4	0	4
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	390	64	C4	187 EAST 7TH STREET		10009	8	2	10
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	391	12	C1	610 EAST 9TH STREET		10009	24	0	24
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	392	22	C4	374 EAST 10TH STREET		10009	17	0	17
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	392	27	C4	384 EAST 10TH STREET		10009	45	0	45
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	392	48	C1	627 EAST 9TH STREET		10009	35	0	35
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	23	C7	632 EAST 11TH STREET		10009	33	2	35
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH		10009	12	0	12
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 1B		10009	12	0	12
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 5B		10009	12	0	12
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 5A		10009	12	0	12
1	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 3A		10009	12	0	12

We pulled in each of the years to python to do some of the cleaning.

```
import pandas as pd
import numpy as np

# Python SQL toolkit and Object Relational Mapper
import sqlalchemy
from sqlalchemy.ext.automap import automap_base
from sqlalchemy.orm import Session
from sqlalchemy import create_engine, func

destination1 = "XLS/2018/2018_manhattan.xlsx"
man_2018 = pd.read_excel(destination1, skiprows=4)
man_2018.head()

man_2018.dtypes

destination2 = "XLS/2018/2018_bronx.xlsx"
bronx_2018 = pd.read_excel(destination2, skiprows=4)
bronx_2018.head()

bronx_2018.dtypes

destination3 = "XLS/2018/2018_brooklyn.xlsx"
brooklyn_2018 = pd.read_excel(destination3, skiprows=4)
brooklyn_2018.head()
```

## ETL Report

Importing the data into SQL for NYC 2016.

The screenshot shows the MySQL Workbench interface with the 'nyc\_rolling\_2016' table selected in the 'Schemas' pane. The 'Table: nyc\_rolling\_2016' information pane on the left lists the columns: NUMS, BOROUGH, NEIGHBORHOOD, BUILDING\_CLASS\_CATEGORY, TAX\_CLASS\_FINAL, BLOCK, LOT, BUILDING\_CLASS\_FINAL, ADDRESS, APARTMENT\_NUMBER, ZIP\_CODE, RESIDENTIAL\_UNITS, and COMMERCIAL\_UNITS. The main window displays the 'Result Grid' for the query 'SELECT \* FROM nyc\_property\_db.nyc\_rolling\_2016;'. The table has 24 rows and 12 columns. The data includes various Manhattan neighborhoods like Alphabet City, Morningside Heights, and Morningside Park, with building classes such as 'RENTALS - WALKUP APARTMENTS' and 'COOPS - WALKUP APARTMENTS'.

NUMS	BOROUGH	NEIGHBORHOOD	BUILDING_CLASS_CATEGORY	TAX_CLASS_FINAL	BLOCK	LOT	BUILDING_CLASS_FINAL	ADDRESS	APARTMENT_NUMBER
0	MANHATTAN	ALPHABET CITY	03 THREE FAMILY DWELLINGS	1	376	24	C0	264 EAST 7TH STREET	
1	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	385	2	C7	21 AVENUE B	
2	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	389	53	C1	615 EAST 6TH STREET	
3	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	26	C7	234 EAST 4TH STREET	
4	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	39	C7	197 EAST 3RD STREET	
5	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	399	39	C7	197 EAST 3RD STREET	
6	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	400	20	C4	528 EAST 5TH STREET	
7	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	400	61	C7	193 EAST 4TH STREET	
8	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	401	15	C5	510 EAST 6TH STREET	
9	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	401	27	C4	534 EAST 6TH STREET	
10	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	402	21	C4	154 EAST 7TH STREET	
11	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	404	17	C7	526 EAST 11TH STREET	
12	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	404	41	C4	329 EAST 10TH STREET	
13	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	404	55	C2	301 EAST 10TH STREET	
14	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2B	406	32	C4	210 AVENUE B	
15	MANHATTAN	ALPHABET CITY	08 RENTALS - ELEVATOR APARTMENTS	2	387	153	D9	629 EAST 5TH STREET	
16	MANHATTAN	ALPHABET CITY	08 RENTALS - ELEVATOR APARTMENTS	2B	400	21	D1	530-532 EAST 5TH STREET	
17	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 3E	
18	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 1C	
19	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	46	C6	317 EAST 3RD STREET, 9	
20	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	49	C6	311 EAST 3RD STREET, 24	
21	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	49	C6	311 EAST 3RD STREET, 17	
22	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2C	374	8	C6	702 EAST 5TH STREET, 5E	
23	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	375	30	C6	742 EAST 6TH STREET, 3	
24	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	16	C6	248 EAST 7TH STREET, 15/16	

Importing the data into SQL for NYC 2017.

The screenshot shows the MySQL Workbench interface with the 'nyc\_rolling\_2017' table selected in the 'Schemas' pane. The 'Table: nyc\_rolling\_2017' information pane on the left lists the columns: NUMS, BOROUGH, NEIGHBORHOOD, BUILDING\_CLASS\_CATEGORY, TAX\_CLASS\_FINAL, BLOCK, LOT, BUILDING\_CLASS\_FINAL, ADDRESS, APARTMENT\_NUMBER, ZIP\_CODE, RESIDENTIAL\_UNITS, and COMMERCIAL\_UNITS. The main window displays the 'Result Grid' for the query 'SELECT \* FROM nyc\_property\_db.nyc\_rolling\_2017;'. The table has 24 rows and 12 columns. The data includes various Manhattan neighborhoods like Alphabet City, Morningside Heights, and Morningside Park, with building classes such as 'RENTALS - WALKUP APARTMENTS' and 'COOPS - WALKUP APARTMENTS'.

NUMS	BOROUGH	NEIGHBORHOOD	BUILDING_CLASS_CATEGORY	TAX_CLASS_FINAL	BLOCK	LOT	BUILDING_CLASS_FINAL	ADDRESS	APARTMENT_NUMBER	ZIP_CODE
0	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	375	28	C4	738 EAST 6TH STREET		10009
1	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	385	36	C7	27 AVENUE C		10009
2	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	392	5	C2	151 AVENUE B		10009
3	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2A	392	6	C2	153 AVENUE B		10009
4	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	400	61	C7	193 EAST 4TH STREET		10009
5	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	405	16	C4	516 EAST 12TH STREET		10009
6	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	407	18	C7	520 EAST 14TH STREET		10009
7	MANHATTAN	ALPHABET CITY	08 RENTALS - ELEVATOR APARTMENTS	2	379	34	D5	141 AVENUE D		10009
8	MANHATTAN	ALPHABET CITY	08 RENTALS - ELEVATOR APARTMENTS	2	394	44	D7	181 AVENUE C		10009
9	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 1C		10009
10	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 3A		10009
11	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 5A		10009
12	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 2E		10009
13	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	40	C6	327 EAST 3RD STREET, 1B		10009
14	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	46	C6	317 EAST 3RD STREET, 12		10009
15	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	49	C6	311 EAST 3RD, 22		10009
16	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	373	49	C6	311 EAST 3RD STREET, 12		10009
17	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	375	30	C6	742 EAST 6TH STREET, 3		10009
18	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	375	30	C6	742 EAST 6TH STREET, 1		10009
19	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	5	C6	98 AVENUE C, 3D		10009
20	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	14	C6	244 EAST 7TH STREET		10009
21	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	14	C6	244 EAST 7TH STREET		10009
22	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	16	C6	248 EAST 7TH STREET, 24		10009
23	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	16	C6	248 EAST 7TH STREET, 2		10009
24	MANHATTAN	ALPHABET CITY	09 COOPS - WALKUP APARTMENTS	2	376	19	C6	254 EAST 7TH STREET, 4		10009



## ETL Report

Importing the data into SQL for NYC 2018.

The screenshot displays the MySQL Workbench interface. The left sidebar shows the 'SCHEMAS' tree with 'nyc\_property\_db' selected. The main window shows the 'nyc\_rolling\_2018' table with a query result grid. The query executed is 'SELECT \* FROM nyc\_property\_db.nyc\_rolling\_2018;'. The result grid shows columns: NUMS, BOROUGH, NEIGHBORHOOD, BUILDING\_CLASS\_CATEGORY, TAX\_CLASS\_FINAL, BLOCK, LOT, BUILDING\_CLASS\_FINAL, ADDRESS, APARTMENT\_NUMBER, ZIP\_CODE, and RESIDENTIAL\_UNITS. The data is filtered to show rows 0 through 26.

NUMS	BOROUGH	NEIGHBORHOOD	BUILDING_CLASS_CATEGORY	TAX_CLASS_FINAL	BLOCK	LOT	BUILDING_CLASS_FINAL	ADDRESS	APARTMENT_NUMBER	ZIP_CODE	RESIDENTIAL_UNITS
0	MANHATTAN	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	390	61	A4	189 EAST 7TH STREET	0	10009	1
1	MANHATTAN	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	390	61	A4	189 EAST 7TH STREET	0	10009	1
2	MANHATTAN	ALPHABET CITY	01 ONE FAMILY DWELLINGS	1	400	19	A4	526 EAST 5TH STREET	0	10009	1
3	MANHATTAN	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	390	35	B1	113 AVENUE C	0	10009	2
4	MANHATTAN	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	404	1	B9	166 AVENUE A	0	10009	2
5	MANHATTAN	ALPHABET CITY	02 TWO FAMILY DWELLINGS	1	404	1	B9	166 AVENUE A	0	10009	2
6	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	379	37	C7	133 AVENUE D	0	10009	20
7	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	385	2	C7	21-23 AVENUE B	0	10009	12
8	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	390	50	C1	200 EAST 7TH STREET	0	10009	18
9	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	390	50	C1	209 EAST 7TH STREET, 1W	0	10009	29
12	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	391	12	C1	610 EAST 9TH STREET	0	10009	24
13	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	392	22	C4	374 EAST 10TH STREET	0	10009	17
14	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	392	27	C4	384 EAST 10TH STREET	0	10009	45
15	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	392	48	C1	627 EAST 9TH STREET	0	10009	35
16	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	23	C7	632 EAST 11TH STREET	0	10009	33
17	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH	0	10009	12
18	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 1B	0	10009	12
19	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 5B	0	10009	12
20	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 5A	0	10009	12
21	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 3A	0	10009	12
22	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 3B	0	10009	12
23	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 2A	0	10009	12
24	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 4A	0	10009	12
25	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 6A	0	10009	12
26	MANHATTAN	ALPHABET CITY	07 RENTALS - WALKUP APARTMENTS	2	393	47	C4	377 EAST 10TH STREET, 6B	0	10009	12