

# *UWS Rent Prediction Analysis*

*IEOR 4523 Final Project*

# *Contents*

- ❖ Abstract
- ❖ Data Description
- ❖ Data Preprocessing
- ❖ Visualizations
- ❖ Data Analysis
- ❖ Conclusion & Evaluation

# Data Extraction

- Renthop

Separately displays “Features & Amenities” which makes it easier to extract and quantify.

Little to no Security against Web Scraping.

- Google API

Average Restaurant Rating of the five closest restaurants to the apartment.



323 West 96th Street, Apt 206  
Upper West Side, Upper Manhattan, Manhattan

**\$3,650** No Fee

By Colleen Dixon, Last 30 min

1 Bed / Flex 2 | 1 Bath

[Check Availability](#)

## Features & Amenities

No Fee	Fitness Center
Convertible	Laundry In Building
Dishwasher	Common Outdoor Space
Doorman	Cats Allowed
Elevator	Dogs Allowed

## Nearby Public Transportation

- 1 2 3 at 96 St (900 ft mi • 4 min) • MTA New York City Transit
- 1 at 103 St (0.39 mi mi • 9 min) • MTA New York City Transit
- 1 2 at 86 St (0.51 mi mi • 12 min) • MTA New York City Transit
- A C B at 96 St (0.58 mi mi • 14 min) • MTA New York City Transit
- A C B at 103 St (0.67 mi mi • 16 min) • MTA New York City Transit

# Data Description

	Link	Address	Price	Number of Beds	Number of Bathrooms	Size (sqft)	Distance from Closest Subway	Hop Score	Laundry in Building	Doorman	No Fee	Pool	Dishwasher	Elevator	distance_to_subway	times_to_subway	new_distance
0	<a href="https://www.renthop.com/listings/850-amsterdam...">https://www.renthop.com/listings/850-amsterdam...</a>	850 Amsterdam Avenue, Apt 2C	\$3,300	1 Bed	1 Bath	Nan	(560 ft mi • 3 min) •	100.0	1	1	1	0	0	1	680ftmi	3	680
1	<a href="https://www.renthop.com/listings/808-columbus-...">https://www.renthop.com/listings/808-columbus...</a>	808 Columbus Avenue, Apt 10E	\$4,913	1 Bed	1 Bath	652.0	(0.23 mi • 5 min) •	100.0	0	0	1	0	0	1	0.23mimi	5	1214.4
2	<a href="https://www.renthop.com/listings/801-amsterdam...">https://www.renthop.com/listings/801-amsterdam...</a>	801 Amsterdam Avenue, Apt 07G	\$4,552	1 Bed	1 Bath	671.0	(0.23 mi • 5 min) •	100.0	0	0	1	0	0	1	0.23mimi	5	1214.4
3	<a href="https://www.renthop.com/listings/203-west-108th-...">https://www.renthop.com/listings/203-west-108th...</a>	203 West 108th Street, Apt 12A	\$3,987	3 Bed	2 Bath	Nan	(810 ft mi • 3 min) •	100.0	0	0	1	0	0	0	810ftmi	3	810
4	<a href="https://www.renthop.com/listings/west-end-ave/...">https://www.renthop.com/listings/west-end-ave/...</a>	75 West End Ave, Apt 57F	\$4,100	1 Bed	1 Bath	Nan	(0.37 mi • 9 min) •	100.0	0	1	0	0	1	0	0.37mimi	9	1953.6
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1142	<a href="https://www.renthop.com/listings/central-park-...">https://www.renthop.com/listings/central-park...</a>	Central Park West	\$3,699	1 Bed Flex 2	1 Bath	Nan	(560 ft mi • 3 min) •	63.1	1	1	1	0	1	1	560ftmi	3	560
1143	<a href="https://www.renthop.com/listings/west-end-aven-...">https://www.renthop.com/listings/west-end-aven...</a>	West End Avenue	\$3,350	1 Bed Flex 2	1 Bath	Nan	(550 ft mi • 3 min) •	63.1	1	1	1	0	1	1	650ftmi	3	650
1144	<a href="https://www.renthop.com/listings/west-end-aven-...">https://www.renthop.com/listings/west-end-aven...</a>	West End Avenue	\$3,250	1 Bed Flex 2	1 Bath	Nan	(550 ft mi • 3 min) •	63.1	1	1	1	0	1	1	650ftmi	3	650
1145	<a href="https://www.renthop.com/listings/215-west-101st-...">https://www.renthop.com/listings/215-west-101st...</a>	215 West 101st Street, Apt 5G	\$4,500	2 Bed	1 Bath	Nan	(520 ft mi • 3 min) •	80.2	1	1	1	0	1	1	620ftmi	3	620
1146	<a href="https://www.renthop.com/listings/columbus-aven-...">https://www.renthop.com/listings/columbus-aven...</a>	Columbus Avenue and 97th St	\$4,800	2 Bed Flex 3	1.5 Bath	950.0	(0.25 mi • 6 min) •	83.1	1	1	1	0	1	1	0.25mimi	6	1320.0

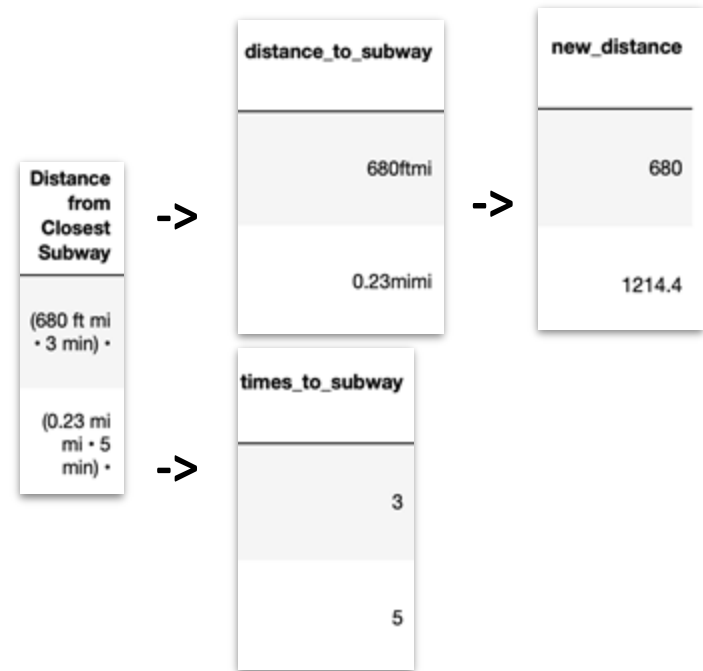
1147 rows x 17 columns

# Data Preprocessing - some examples

- We consider Private Room, Loft, and Studio as rooms with only one bedroom, and rooms with flexible space are calculated based on the number of bedrooms after expansion.

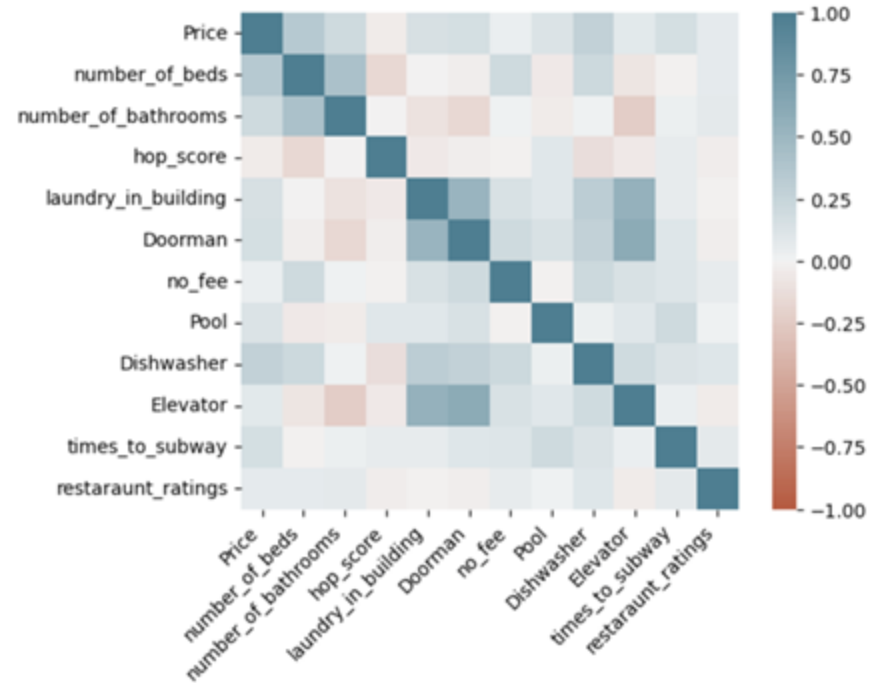
Number of Beds	Number of Bathrooms		Number of Beds	Number of Bathrooms
1 Bed	1 Bath	->	1	1
Studio	1 Bath		1	1
Private Room	1 Bath		1	1
Loft	1 Bath		1	1
2 Bed Flex 3	1.5 Bath		3	1.5

- Divide the column into time and distance variables



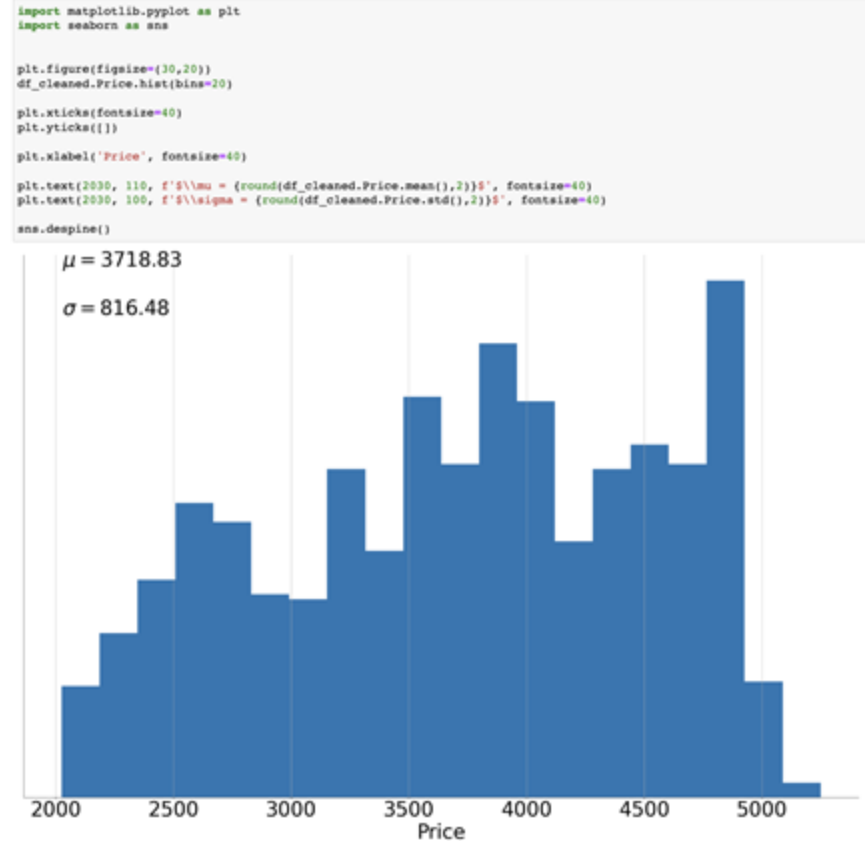
# Data Visualization - correlation

We create a heat map of the correlation coefficients of all the variables. This allows us to visualize if there will be any multicollinearity problems and help us in feature selection for our prediction models.



# Data Visualization-the price range of apartments

To observe the price range of apartments we build a histogram using matplotlib library and the results show us the mean and standard deviation of the prices of the apartments.

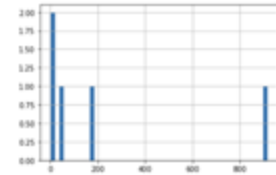


# Data Visualization- frequency of apartment based on attributes

We take different independent variables and see their effect on the apartment prices independently. The histograms built depict the frequency of apartments available on the basis of attributes such as number of bedrooms.

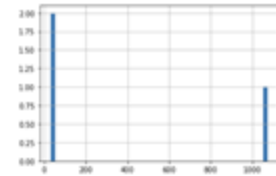
```
df_cleaned.groupby('number_of_beds').size().sort_values(ascending=False).hist(bins=50)
```

<AxesSubplot:>



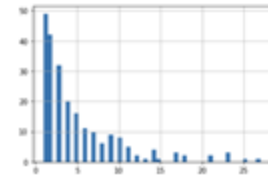
```
df_cleaned.groupby('number_of_bathrooms').size().sort_values(ascending=False).hist(bins=50)
```

<AxesSubplot:>



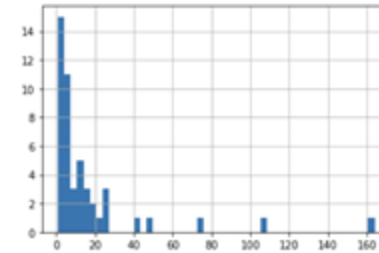
```
df_cleaned.groupby('hwp_score').size().sort_values(ascending=False).hist(bins=50)
```

<AxesSubplot:>



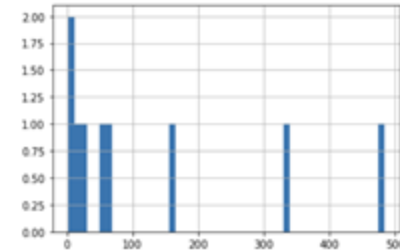
```
df_cleaned.groupby('new_distance').size().sort_values(ascending=False).hist(bins=50)
```

<AxesSubplot:>



```
df_cleaned.groupby('times_to_subway').size().sort_values(ascending=False).hist(bins=50)
```

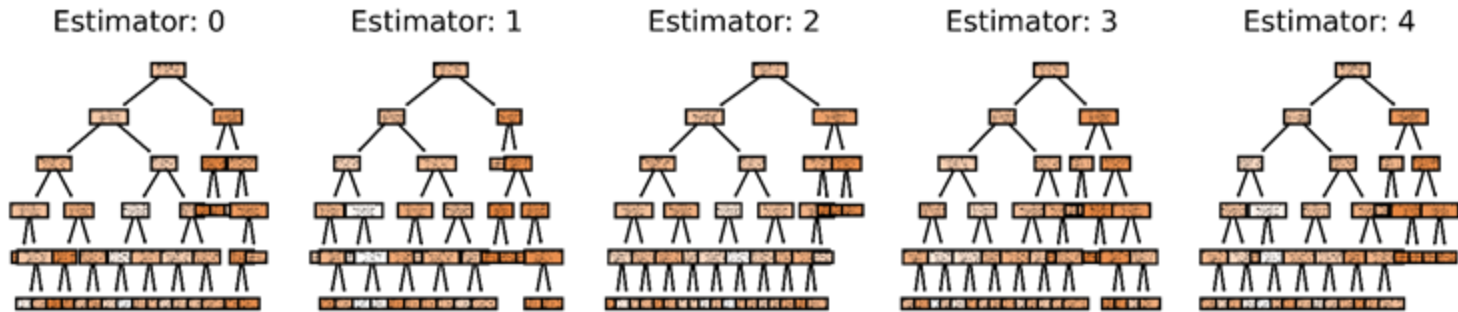
<AxesSubplot:>





# Data Analysis - predicting missing 'size' values

- Many 'size\_sqft' entries are missing values -> predict them using Random Forest Regressor
- Use all features, except Price
- Out-of-sample model score = 66%

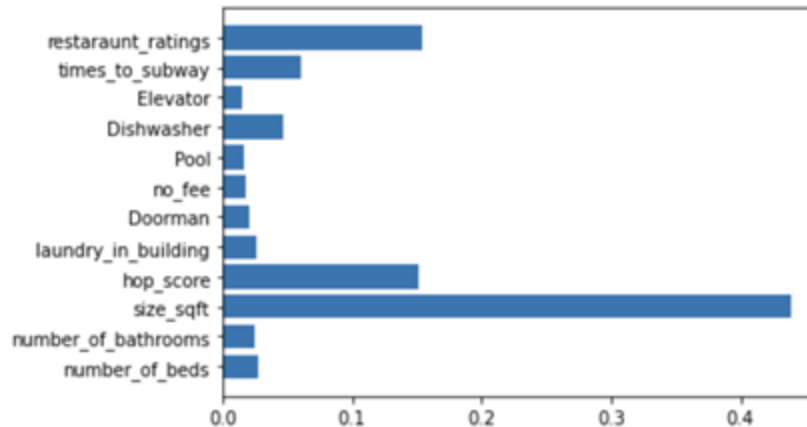


# Data Analysis - prediction models

Model	Parameters	Score on Test Set
<b>Linear Regression</b>		0.30
<b>Ridge Regression</b>	scaled variables	0.31
<b>Random Forest Regressor</b>	n_estimators = 100 max_depth = 5	0.58
<b>K-neighbours Regressor</b>	k = 3 -> optimal	0.52

# Data Analysis - feature importance

We looked at which features were more influential in making predictions in the Random Forest model - mean decrease in impurity (sklearn built-in function)





***Thank you for your listening!***

*UWS Rent Prediction Analysis*

*Group name:  
You can't six with us*