

# NYC apartments...



# NYC apartments...



New studio apartment in NYC  
on the market. 2,500/mo



# **Fantastic apartments &Where to find them?**



# Intro

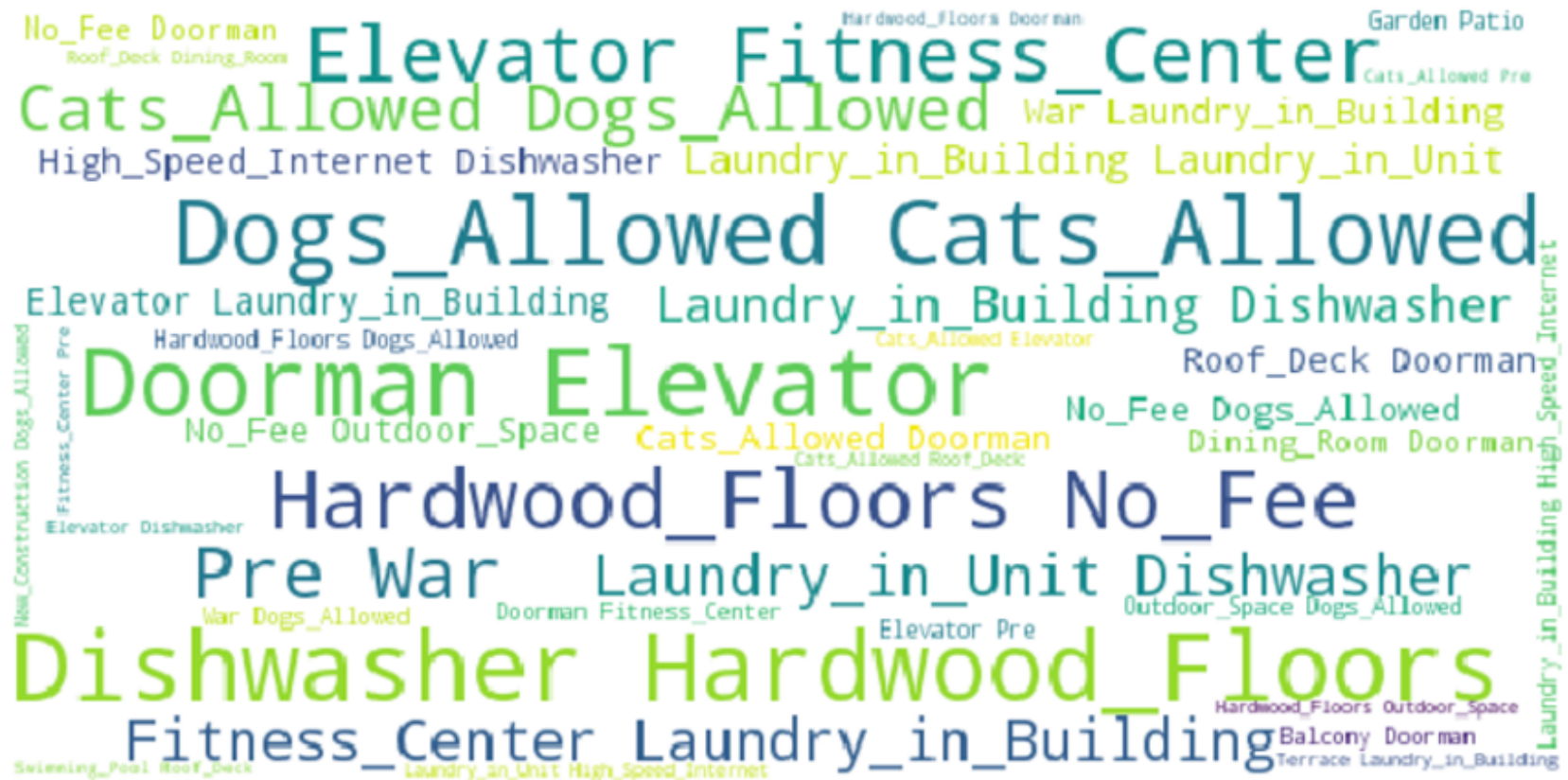
**Objective:** To predict rental price based on given apartment features

**Data:** Renthop manhattan listings through scrapy

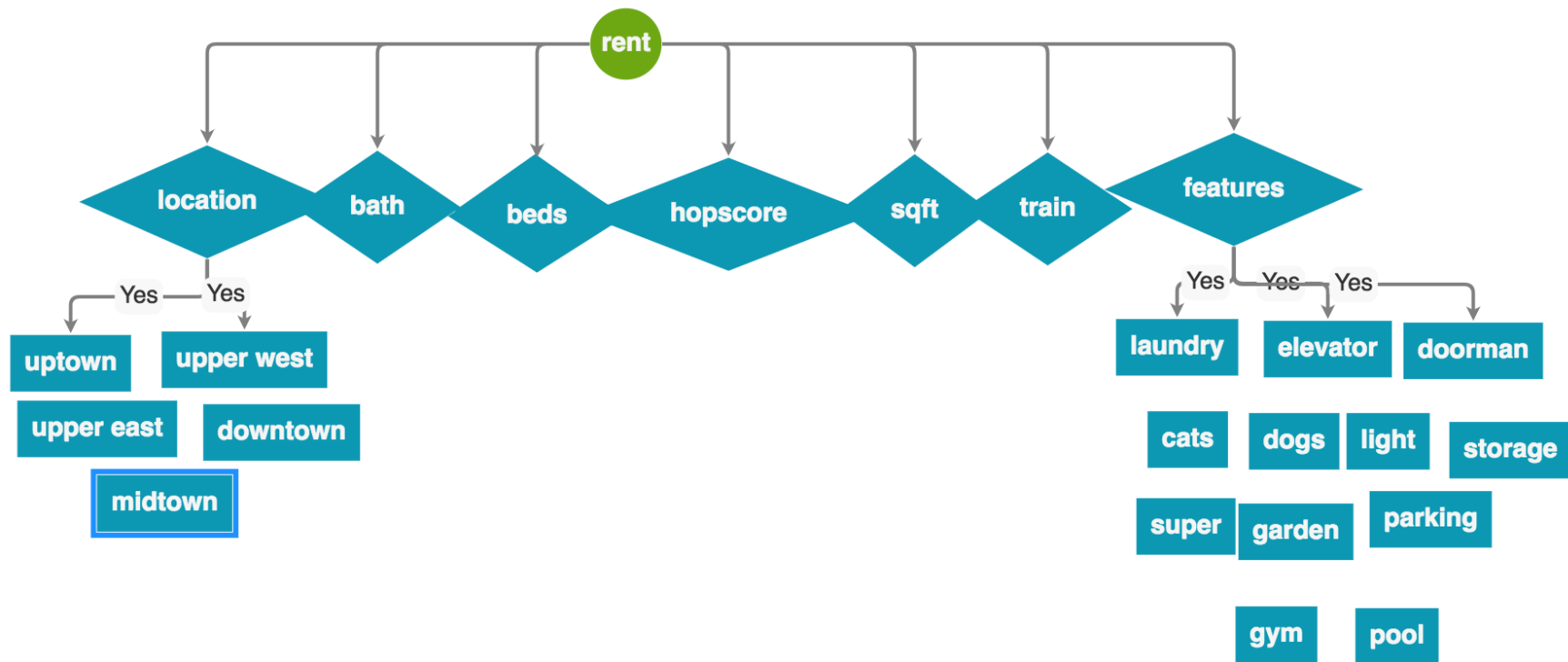
**Methods:** Lasso regression







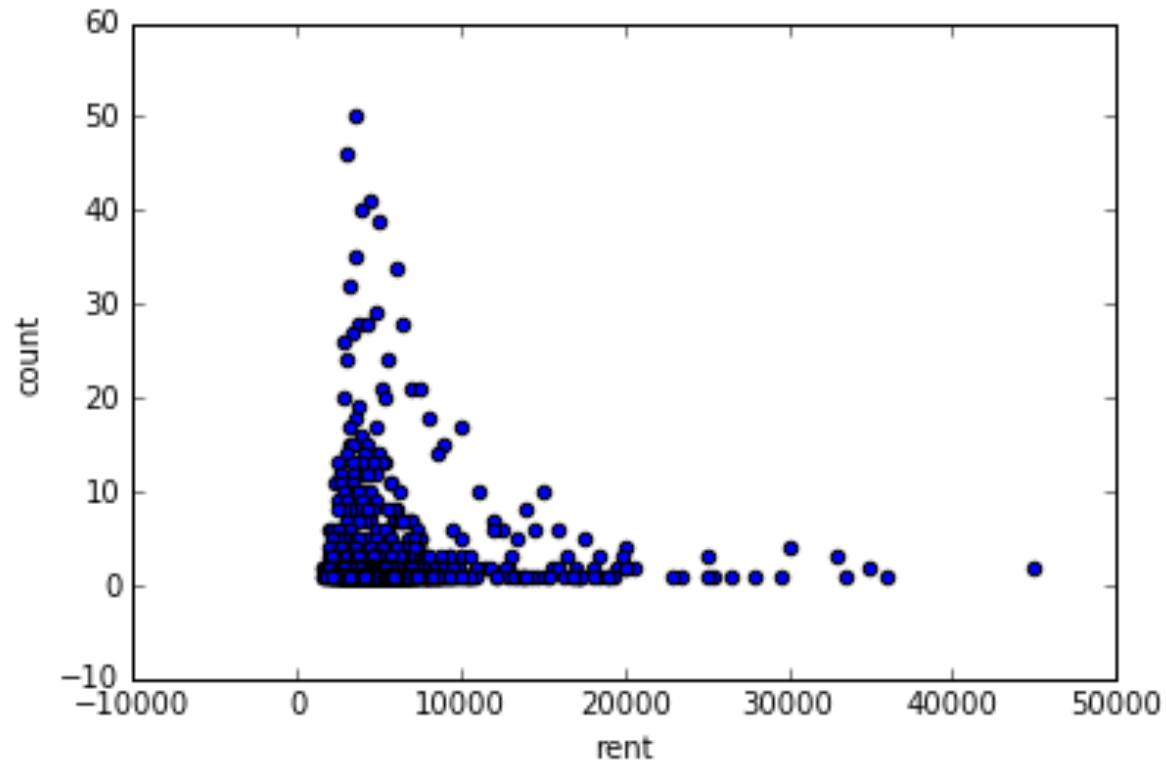
# My features



# Data prep

taking a close look at our y variable... log transform, exclude outliers?

normalization

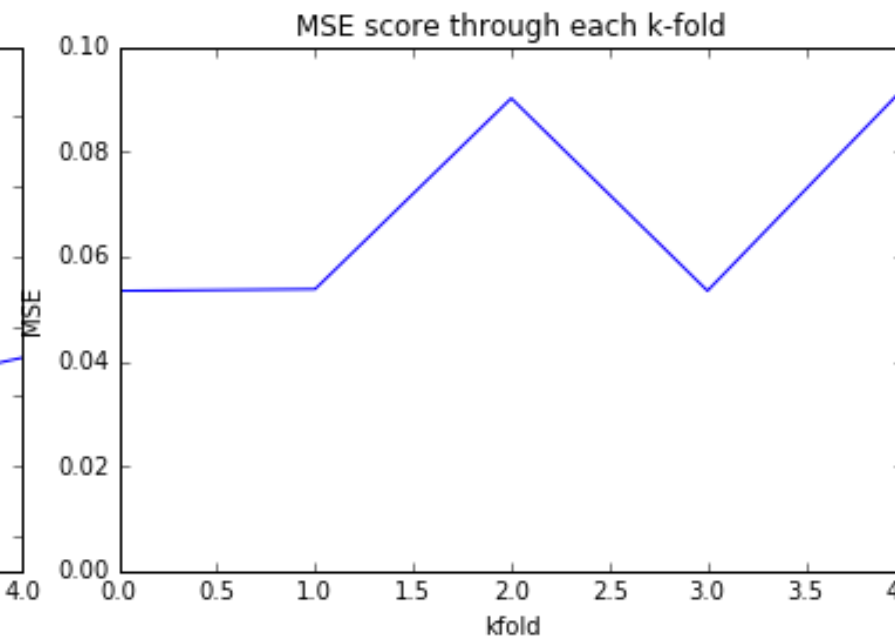
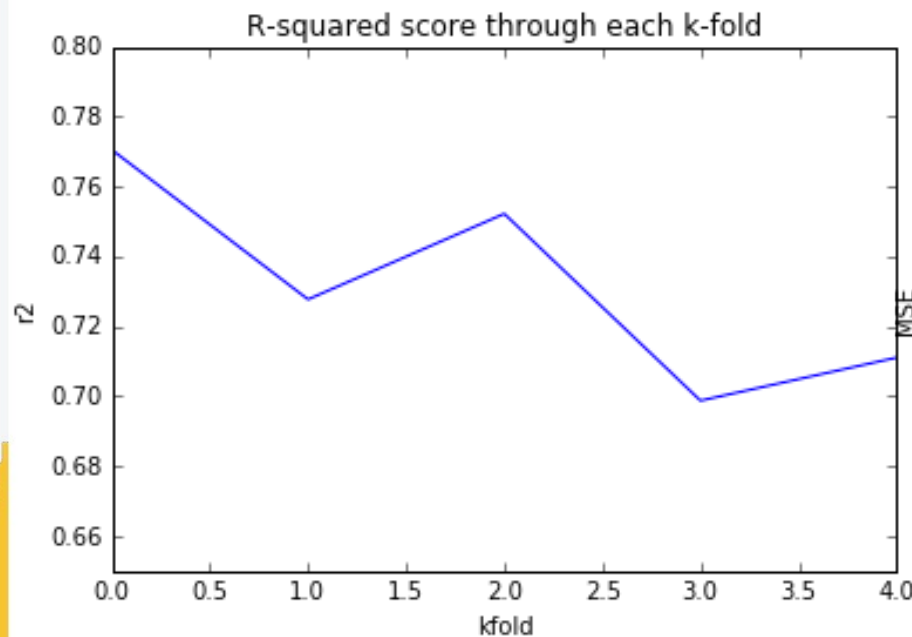


# Lasso regression

$\text{alphas}=[1\text{e-}3, 1\text{e-}1, 0, 1]$      $\text{n\_folds}=5$

average  $r^2 = 0.7$

average MSE = 0.06



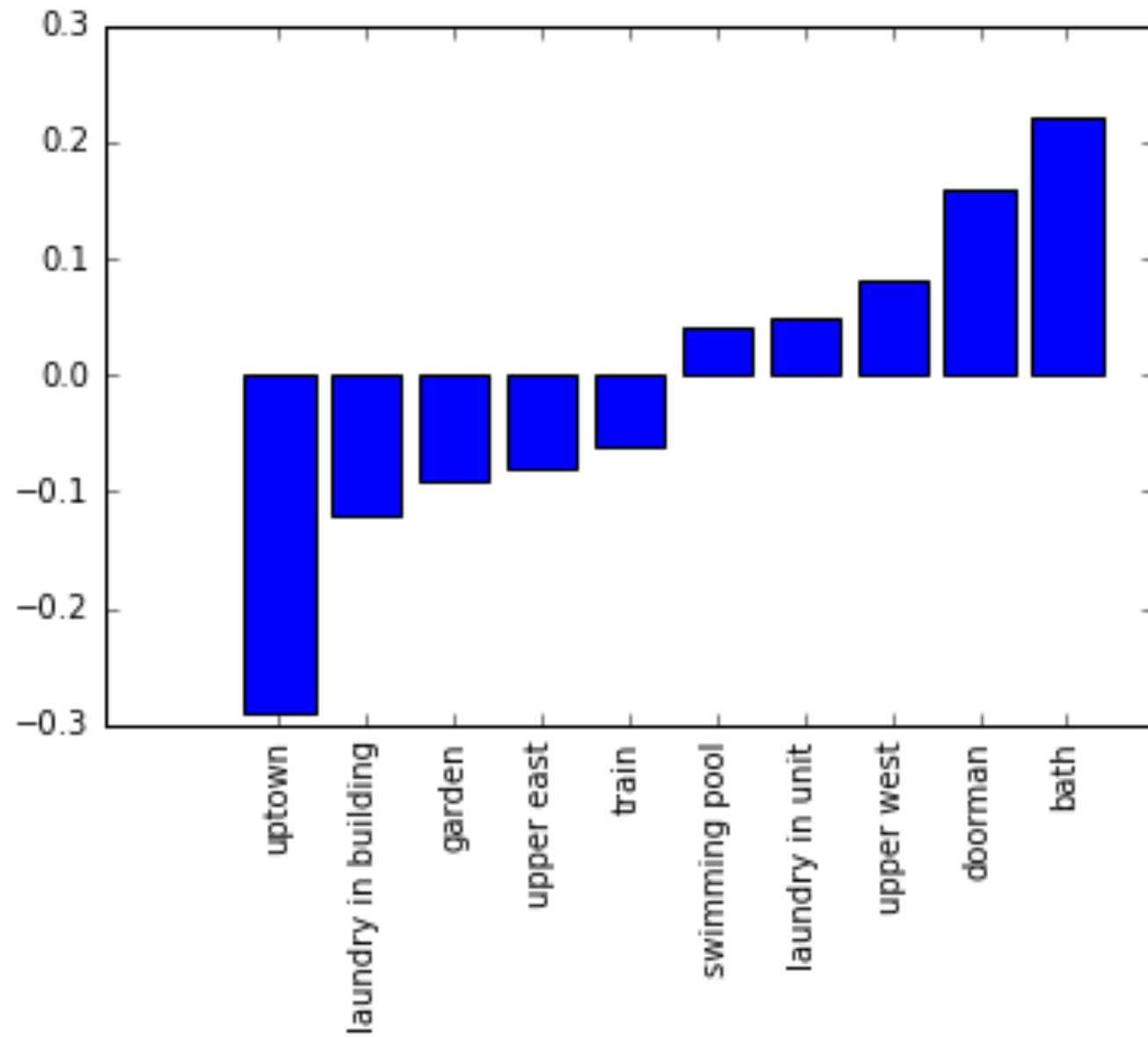


# Result

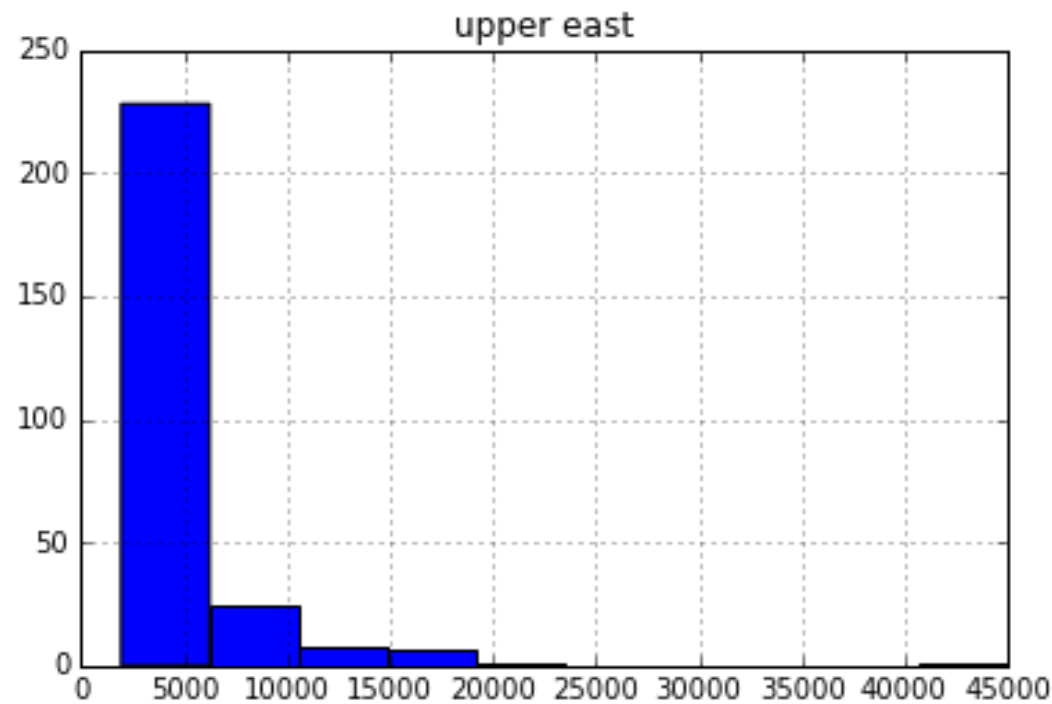
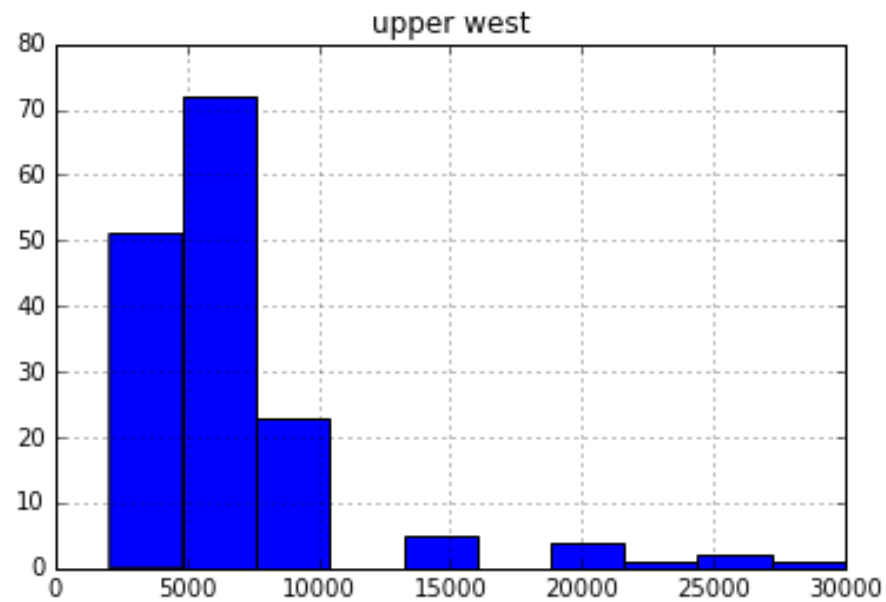
bath	0.22	uptown	-0.29
doorman	0.16	laundry in building	-0.12
upper west side	0.08	garden	-0.09
laundry in unit	0.05	upper east side	-0.08
swimming pool	0.04	train	-0.06



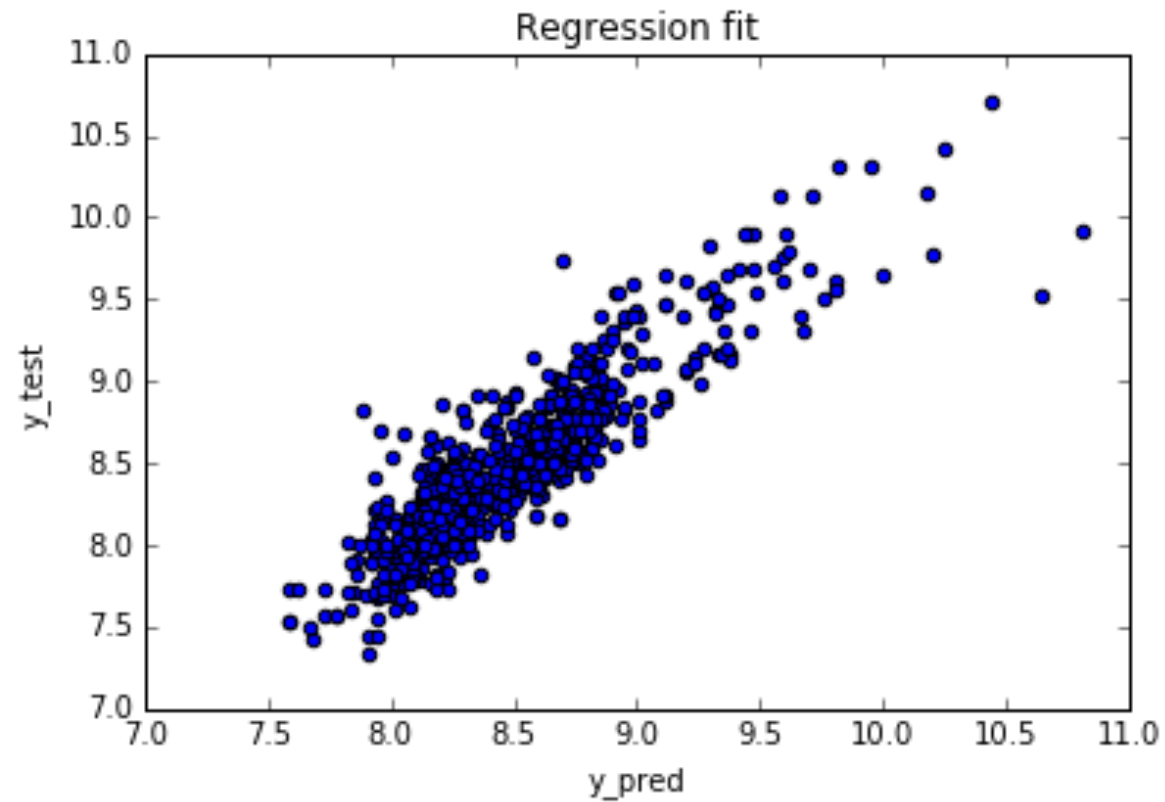
# Result



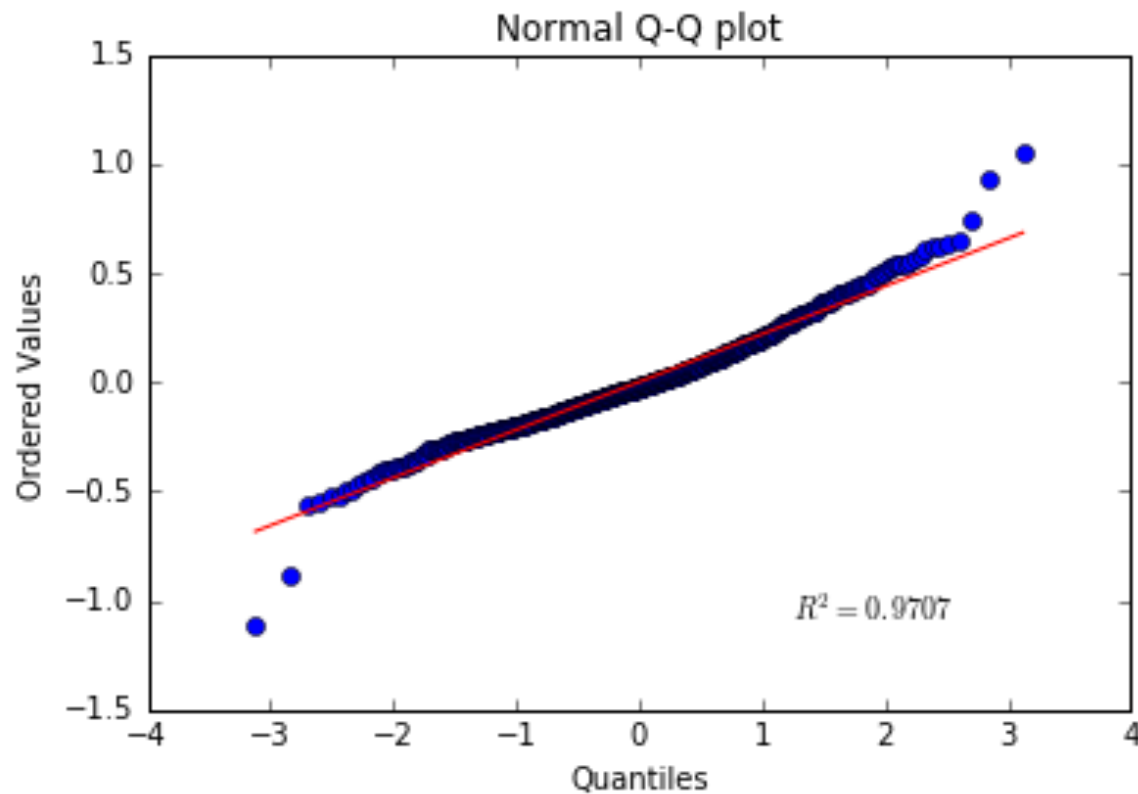
# Upper East vs Upper West



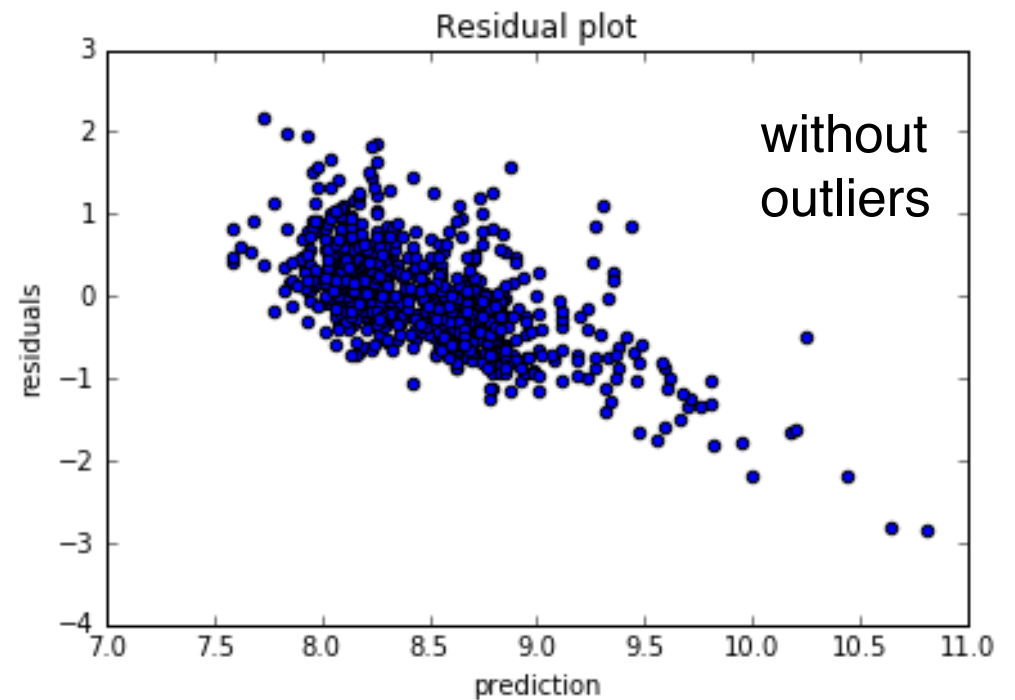
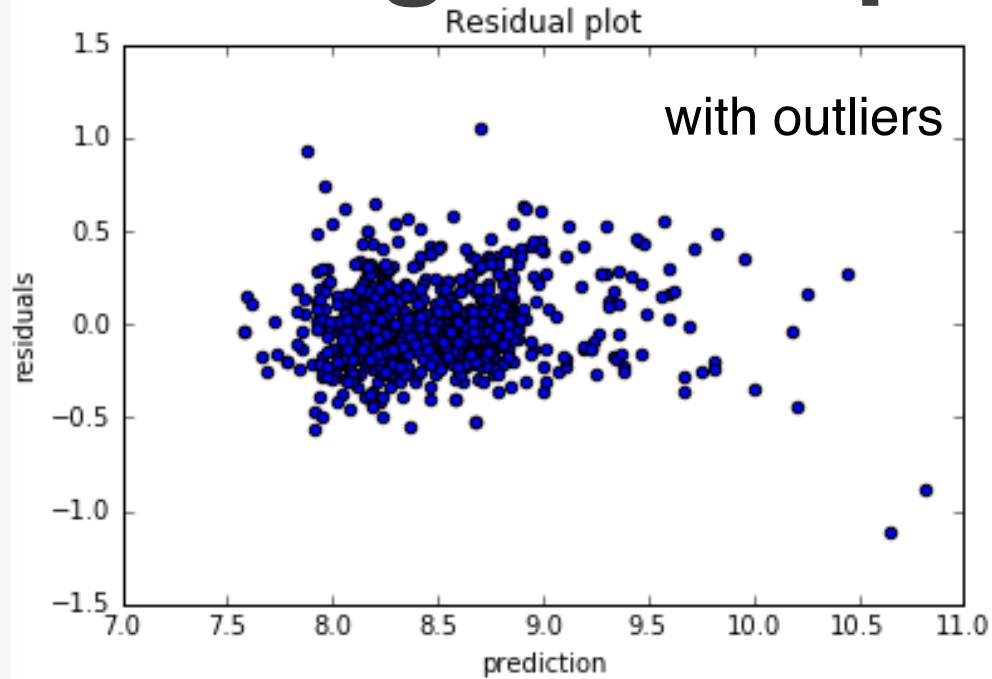
# Testing assumptions



# Testing assumptions



# Testing assumptions





# Some thinking

1. Question: When shall we exclude outliers in regression?
2. Improvement: Incorporate demographic data by zip code
3. Application

Thanks! Questions

