

# Factors Influencing Renthop.com Apartment Listing Interest

Bill Grieser

The George Washington University

Machine Learning I, DATS 6202

Final Assignment, Spring 2017

# This Project

- Renthop is an apartment listing website, matching renters with available apartments and their rental agents
- They sponsored a Kaggle competition: given the information for a set of listings on the website for NYC, **predict if there would be High, Medium, or Low interest in the apartment**
- That is what this project undertakes



renthop

Search

Resources

Favorites

Find Roommates

Post Rental

Log In



New York City, NY

Studio

1

2

3

4+

Lft

Rm

\$ 0

to

\$ 50000

Filters

SEARCH

## Apartments for Rent in NYC

New York, NY

Updated - April 23, 2017

## Upcoming Open Houses

Sun, Apr 23	4:00pm - 5:00pm	\$2,750	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$2,704	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$3,478	1BR, 1BA at 247 N 7th Street

« Back | Page 1 of 366 (7,312 Rentals) | Next »

Sort: HopScore | Price



Posted 7 mins ago

# Background

Check Availability

No Fee · Doorman · Elevator · Laundry in Unit · Dishwasher · Loft



Posted 6 mins ago

**100** 1BR, 1BA at 58 Linden Blvd, Apt 2H  
Flatbush, Central Brooklyn, Brooklyn

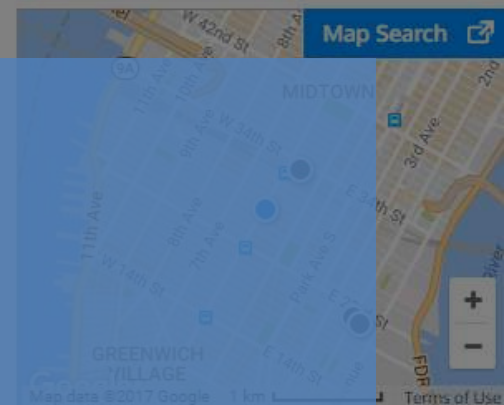
\$1,800

Per Month

By NYC Leasing

Check Availability

By Owner · Exclusive · No Fee · Elevator · Dishwasher



## New York City, NY

New York City is the world's greatest city. Of course, that is open to debate in many parts of the world, but among the residents long and new, there is simply no contest. Whether you seek gourmet dining, legendary Broadway shows, luxury brands, or more humble interests; you'll certainly find what you are looking for somewhere within the five boroughs of NYC.

Probably the most difficult part of living in New York is actually moving here, both because of the difficulty in finding NYC...

[Area Guide](#) · [Popular Buildings](#)**100** 1BR, 1BA at Water St.

# The Renthop Website

Apartment for Rent in New York City

Secure | [https://www.renthop.com/search/nyc?location\\_search=&min\\_price=0&max\\_price=50000&q=&has\\_floorplan=on&sort=hopscore&page=1&use](https://www.renthop.com/search/nyc?location_search=&min_price=0&max_price=50000&q=&has_floorplan=on&sort=hopscore&page=1&use)

Apps | Mail - Bill Grieser - O... | Nat Geo Map | Gmail | 3rd-party Analytics In... | Mail - Bill Grieser - O... | Using The EGit Eclipse... | YouTube | My Applications | nohup command - m... | Other bookmarks

renthop

Search Resources Favorites Find Roommates Post Rental Login

New York City, NY

Studio 1 2 3 4+ Lft Rm

\$ 0 to \$ 50000

Filters

SEARCH

Apartments for Rent in NYC

New York, NY

Updated - April 23, 2017

Upcoming Open Houses

Sun, Apr 23 4:00pm - 5:00pm \$2,750 Studio, 1BA at 247 N 7th Street

Sun, Apr 23 4:00pm - 5:00pm \$2,704 Studio, 1BA at 247 N 7th Street



Sun, Apr 23 4:00pm - 5:00pm \$3,478 1BR, 1BA at 247 N 7th Street

« Back

Page 1 of 366 (7,312 Rentals)

Next »

Sort: HopScore | Price



100 2BR, 2BA at 6th Ave.

NoMad, Midtown Manhattan, Manhattan

\$7,200



Per Month

By Tal Cohen

Check Availability

No Fee · Doorman · Elevator · Laundry in Unit · Dishwasher · Loft

Posted 7 mins ago



100 1BR, 1BA at 58 Linden Blvd, Apt 2H

Flatbush, Central Brooklyn, Brooklyn

\$1,800


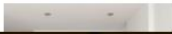
Per Month

By NYC Leasing

Check Availability


By Owner · Exclusive · No Fee · Elevator · Dishwasher

Posted 6 mins ago



100 1BR, 1BA at Water St.

Map Search



New York City, NY

New York City is the world's greatest city. Of course, that is open to debate in many parts of the world, but among the residents long and new, there is simply no contest. Whether you seek gourmet dining, legendary Broadway shows, luxury brands, or more humble interests; you'll certainly find what you are looking for somewhere within the five boroughs of NYC.

Probably the most difficult part of living in New York is actually moving here, both because of the difficulty in finding NYC...

[Area Guide](#) · [Popular Buildings](#)

# Search

Bathrooms					Keywords
<input type="button" value="1"/>	<input type="button" value="2"/>	<input type="button" value="3"/>	<input type="button" value="4"/>	<input type="button" value="5+"/>	<input type="text" value="WiFi, Parking (Comma Separated)"/>

Must Haves	Unit Features	Building Features
<input type="checkbox"/> Has Photos	<input type="checkbox"/> Furnished	<input type="checkbox"/> Doorman
<input checked="" type="checkbox"/> Has Floorplan	<input checked="" type="checkbox"/> Laundry In Unit	<input type="checkbox"/> Elevator
<input type="checkbox"/> Open House	<input type="checkbox"/> Private Outdoor Space	<input type="checkbox"/> Fitness Center
	<input type="checkbox"/> Parking Space	<input checked="" type="checkbox"/> Laundry In Building
		<input type="checkbox"/> Common Outdoor Space
		<input type="checkbox"/> Storage Facility

Listing Type	Pet Policy
<input type="checkbox"/> By Owner	<input type="checkbox"/> Cats Allowed
<input type="checkbox"/> Exclusive	<input type="checkbox"/> Dogs Allowed
<input type="checkbox"/> Sublet / Lease-Break	
<input checked="" type="checkbox"/> No Fee	
<input type="checkbox"/> Reduced Fee	
<input type="checkbox"/> Short Term Allowed	

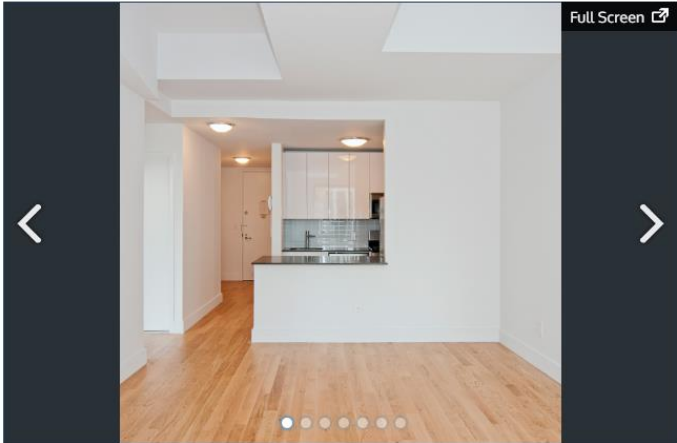
Studio at Wall Street pos x

Secure | https://www.renthop.com/listings/wall-street/305/8708521

Apps Mail - Bill Grieser - O Nat Geo Map Gmail 3rd-party Analytics Mail - Bill Grieser - O Using The EGit Eclipse YouTube My Applications nohup command - m Other bookmarks

renthop Search Resources Favorites Find Roommates Post Rental Login

« Back to Search New to NYC? Check out our [NYC Renter's Guide](#) Create an Email Alert Create



Full Screen

Mike Hassan  
BLU Realty Group - Headquar

Name

Phone Number

Email

Hi, I'm interested in this apartment. Is it still available for a showing?

Check Availability

Report Listing

Studio at Wall Street  
Financial District, Downtown Manhattan, Manhattan

\$2,595 Per Month 98.0 HopScore

Studio 1 Bath Immediate Move-In  
Listing Posted 12 mins ago

Heart Twitter Facebook Pinterest

A listing has:

(Data available in **Bold**)

- **Photos**
- **Size**
- **Display Address**
- Listing Agent Information (**Id only**)
- **Price**
- "Hop Score"

## Studio at Wall Street

Financial District, Downtown Manhattan, Manhattan

\$2,595

Per Month

98.0

HopScore

Studio | 1 Bath | Immediate Move-In

Listing Posted 12 mins ago



### Description

Large Studio with Open Kitchen and Lots of Closet Space. Available Immediately.


### Features & Amenities

- |                     |                        |                |
|---------------------|------------------------|----------------|
| ✓ No Fee            | ✓ Floorplans Available | ✓ dishwasher   |
| ✓ doorman           | ✓ gym                  | ✓ storage      |
| ✓ elevator          | ✓ garage               | ✓ High Ceiling |
| ✓ Oversized Windows | ✓ Abundant Closets     | ✓ Concierge    |
| ✓ Lounge            | ✓ Laundry Facility     | ✓ Valet        |
| ✓ Roofdeck          | ✓ Billiards Room       | ✓ Bike Room    |



Mike Hassan

BLU Realty Group - Headquar

Name	
Phone Number	
Email	
Hi, I'm interested in this apartment. Is it still available for a showing?	

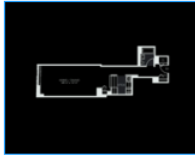
Check Availability

[Report Listing](#)

A listing has:

- How long on site (**Creation Date**)
- **Description**
- **Bulleted Feature List**
- **Check Availability** (probably our target)

## Floorplans



## On the Map



**Mike Hassan**

BLU Realty Group - Headquar

Name



Phone Number

Email

Hi, I'm interested in this apartment. Is it still available for a showing?

**Check Availability**

[Report Listing](#)

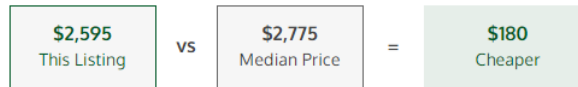
A listing has:

- Optional Floorplan (as picture)
- Map (Longitude / Latitude)



## Price Comparison

Comparing **this listing** against median prices for [Studio / 1BA apartments in Financial District](#).



The price of this apartment is **\$180** cheaper than the median price.

## HopScore Breakdown


This listing has a HopScore of **98.0** and was posted **12 mins ago**. The listing quality and manager score is **fair**. Some of the contributing factors to the HopScore are listed below.

GOOD	Manager has registered with RentHop
GOOD	Manager has good availability for showings
GOOD	Manager logged in within the last 24 hours
GOOD	Manager recently posted/featured new listings
NEUTRAL	Street number not provided
NEUTRAL	Manager does not yet have user reviews



**Mike Hassan**

BLU Realty Group - Headquar

Name 

Phone Number

Email

Hi, I'm interested in this apartment. Is it still available for a showing?

**Check Availability**

[Report Listing](#)

A listing has:

- Price Comparison
- Hop Score (really a manager score)

### Similar Apartments Nearby



**Studio at Financial District**  
Financial District, Downtown Manhat...

**\$2,550** | **100**  
Per Month | HopScore

Posted 49 mins ago



**Studio at Wall Street**  
Financial District, Downtown Manhat...

**\$2,745** | **100**  
Per Month | HopScore

Posted 13 mins ago



**Studio at FiDi Water**  
Financial District, Downtown Manhat...

**\$2,389** | **100**  
Per Month | HopScore

Posted 18 mins ago



**Studio at Water**  
Financial District, Downtown Manhat...

**\$2,390** | **99.9**  
Per Month | HopScore

Posted 22 mins ago



**Studio at Pine Street**  
Financial District, Downtown Manhat...

**\$2,820** | **99.0**  
Per Month | HopScore

Posted 53 mins ago



**Studio at Water**  
Financial District, Downtown Manhat...

**\$2,398** | **99.1**  
Per Month | HopScore

Posted 25 mins ago



**Studio at Exchange Place**  
Financial District, Downtown Manhat...

**\$2,420** | **98.2**  
Per Month | HopScore

Posted 33 mins ago

A listing has:

- Comparables



renthop

Search

Resources

Favorites

Find Roommates

Post Rental

Log In



New York City, NY

Studio

1

2

3

4+

Lft

Rm

\$ 0

to

\$ 50000

Filters

SEARCH

## Apartments for Rent in NYC

New York, NY

Updated - April 23, 2017

## Upcoming Open Houses

Sun, Apr 23	4:00pm - 5:00pm	\$2,750	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$2,704	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$3,478	1BR, 1BA at 247 N 7th Street

« Back | Page 1 of 366 (7,312 Rentals) | Next »

Sort: HopScore | Price



Posted 7 mins ago



Posted 6 mins ago



100 2BR, 2BA at 16th Ave  
\$2,200 Per Month  
By NYC Leasing

Check Availability

No Fee · Doorman · Elevator · Laundry in Unit ·  
Dishwasher · Loft

100 1BR, 1BA at 58 Linden Blvd, Apt 2H  
Flatbush, Central Brooklyn, Brooklyn

\$1,800

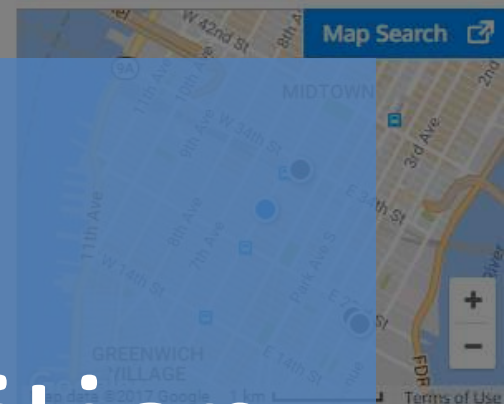
Per Month

By NYC Leasing

Check Availability

By Owner · Exclusive · No Fee · Elevator · Dishwasher

100 1BR, 1BA at Water St.



New York City is the world's greatest city. Of course, that is open to debate in many parts of the world, but among the residents long and new, there is simply no contest. Whether you seek gourmet dining, legendary Broadway shows, luxury brands, or more humble interests; you'll certainly find what you are looking for somewhere within the five boroughs of NYC.

Probably the most difficult part of living in New York is actually moving here, both because of the difficulty in finding NYC...

[Area Guide](#) · [Popular Buildings](#)

# The Question

Can we identify factors that influence the interest level specific New York City listings receive on the Renthop.com apartment listing web site . . .

. . . by fitting a Machine Learning model to predict listing interest level, using only data provided by Renthop through the Kaggle “Two Sigma Connect: Rental Listing Inquiries” competition?



# S.M.A.R.T.

- Specific
  - Defined by the Kaggle Competition
- Measurable
  - Use goodness-of-fit measures of the classification model
  - Kaggle score submissions
  - Use other statistical measures (chi squared test)
- Actionable
  - Where possible, choose techniques that are interpretable, even at the expense of model accuracy
- Relevant
  - Real-life problem
- Time Bound
  - Strictly Bounded – see next chart



# Bounds of the Problem

- The main bound is the due date
  - Kaggle and class coincide
- Kaggle ground rule: No outside data

*It became essential to time-bound the phases in order to not get lost*





# What does Success look like?

- A fitted model using most of the data provided that has a reasonable goodness of fit, as measured by Log Loss and F1 measures against held out test data
- Some insights into successful listings supported by the data and model
- A Kaggle entry in the competition



renthop

Search

Resources

Favorites

Find Roommates

Post Rental

Log In



New York City, NY

Studio

1

2

3

4+

Lft

Rm

\$ 0

to

\$ 50000

Filters

SEARCH

## Apartments for Rent in NYC

New York, NY

Updated - April 23, 2017

## Upcoming Open Houses

Sun, Apr 23	4:00pm - 5:00pm	\$2,750	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$2,704	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$3,478	1BR, 1BA at 247 N 7th Street

« Back | Page 1 of 366 (7,312 Rentals) | Next »

Sort: HopScore | Price

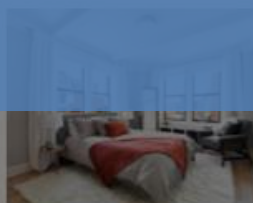


Posted 7 mins ago

# The Data

Check Availability

No Fee · Doorman · Elevator · Laundry in Unit · Dishwasher · Loft



Posted 6 mins ago

100 1BR, 1BA at 58 Linden Blvd, Apt 2H  
Flatbush, Central Brooklyn, Brooklyn

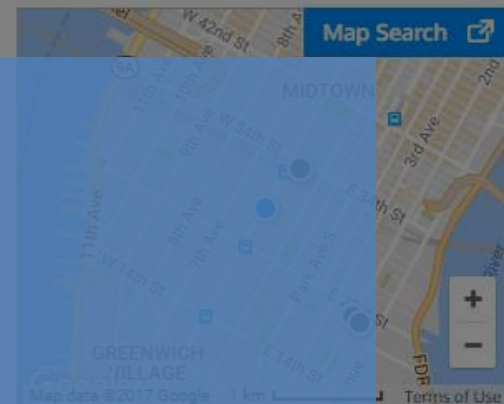
\$1,800

Per Month

By NYC Leasing

Check Availability

By Owner · Exclusive · No Fee · Elevator · Dishwasher



## New York City, NY

New York City is the world's greatest city. Of course, that is open to debate in many parts of the world, but among the residents long and new, there is simply no contest. Whether you seek gourmet dining, legendary Broadway shows, luxury brands, or more humble interests; you'll certainly find what you are looking for somewhere within the five boroughs of NYC.

Probably the most difficult part of living in New York is actually moving here, both because of the difficulty in finding NYC...

[Area Guide](#) · [Popular Buildings](#)

100 1BR, 1BA at Water St.



# Original Data Fields

- **Listing\_id** (int): Unique ID for the listing
- **Bedrooms** (int): number of bedrooms
- **Bathrooms** (decimal): number of bathrooms. Half-baths are 0.5
- **Building Id** (GUID): Unique ID of building, or 0 if no building id
- **Created** (date): Creation data of the listing
- **Description** (string): Free-form text description entered by the listing agent
- **Display Address** (string): Address of property as seen in the listing
- **Features** (list of strings): a list of features entered by the listing agent about this apartment
- **Longitude, Latitude** (decimal, degrees): Location of the apartment
- **Manager Id** (GUID): Unique ID of the listing agent
- **Photos** (list of URLs): The photos associated with the listing. Also provided in a file grouped by Listing ID.
- **Price** (int): Monthly rent in USD
- **Street Address** (string): Address of the apartment; not displayed in the listing
- **Interest Level** (string, categorical): This is the target variable. It has three categories: 'high', 'medium', 'low'

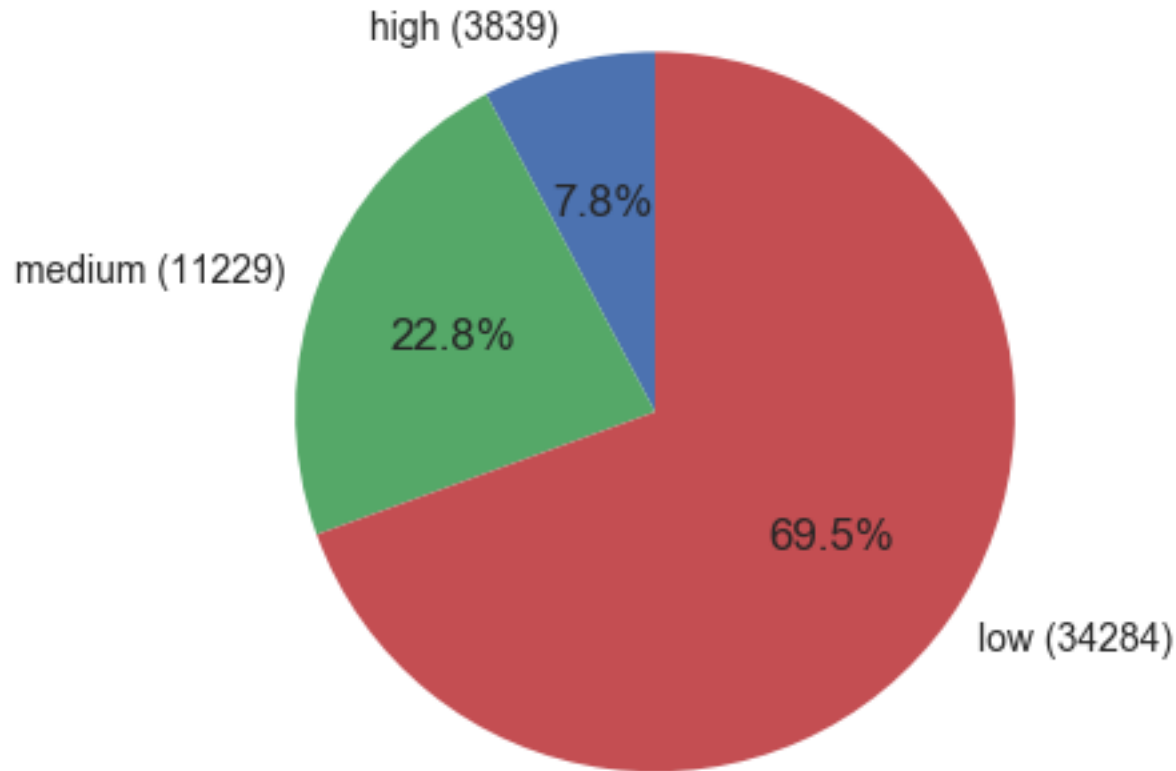
# Approach and Strategy

- Explore Data
- Create features usable for analysis
  1. **Variable length list of listing features**
  2. **Comparable listings**
  3. **Location**
  4. **Images**
  5. **Text description**
- Use mRMR, Random Forest, and home-grown feature selector to identify important features
- Fit competing models using different classifiers attempting to predict Interest Level
- Determine insights
- Report Results



# Data Summary

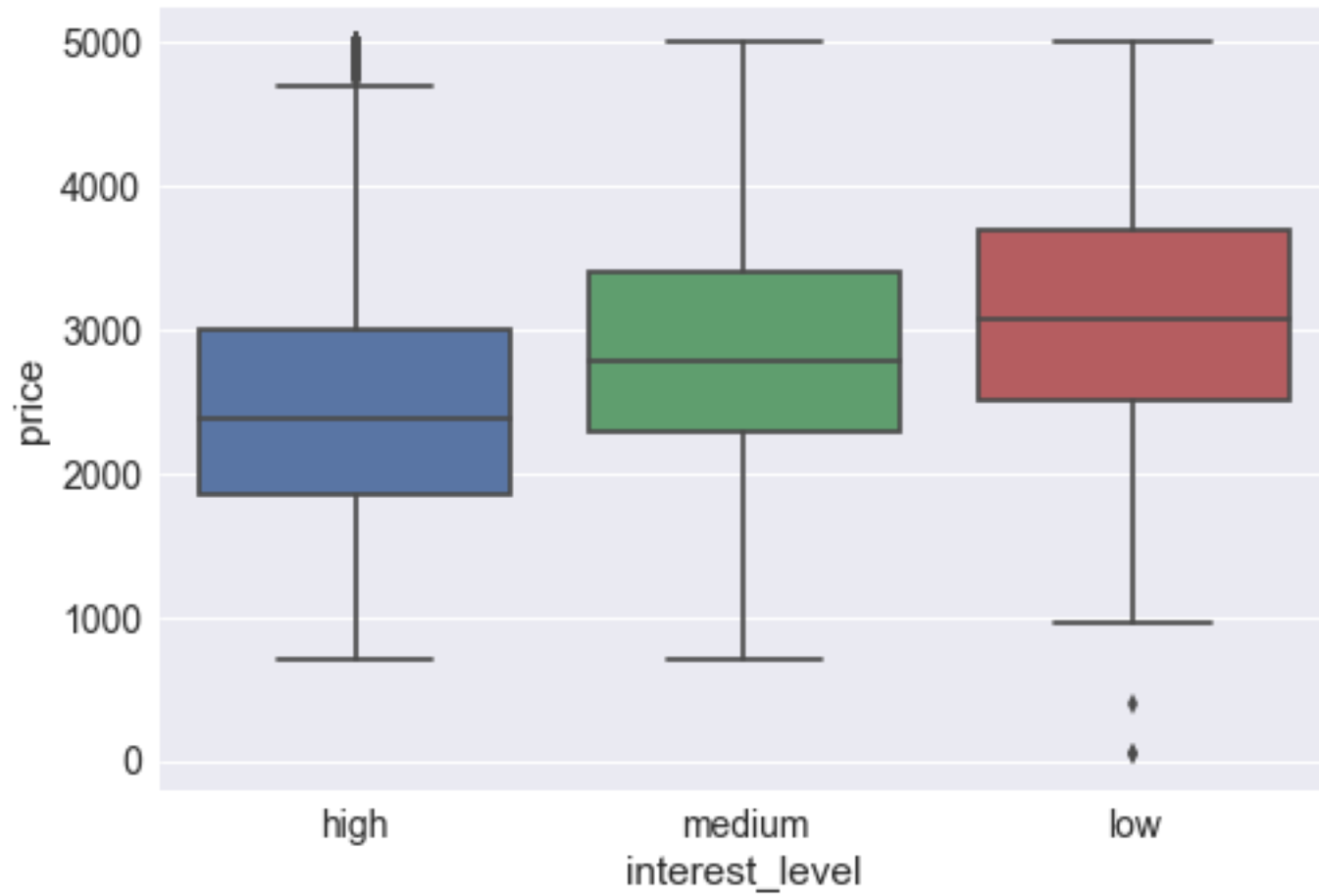
Percentage of Interest Categories: All Training Data



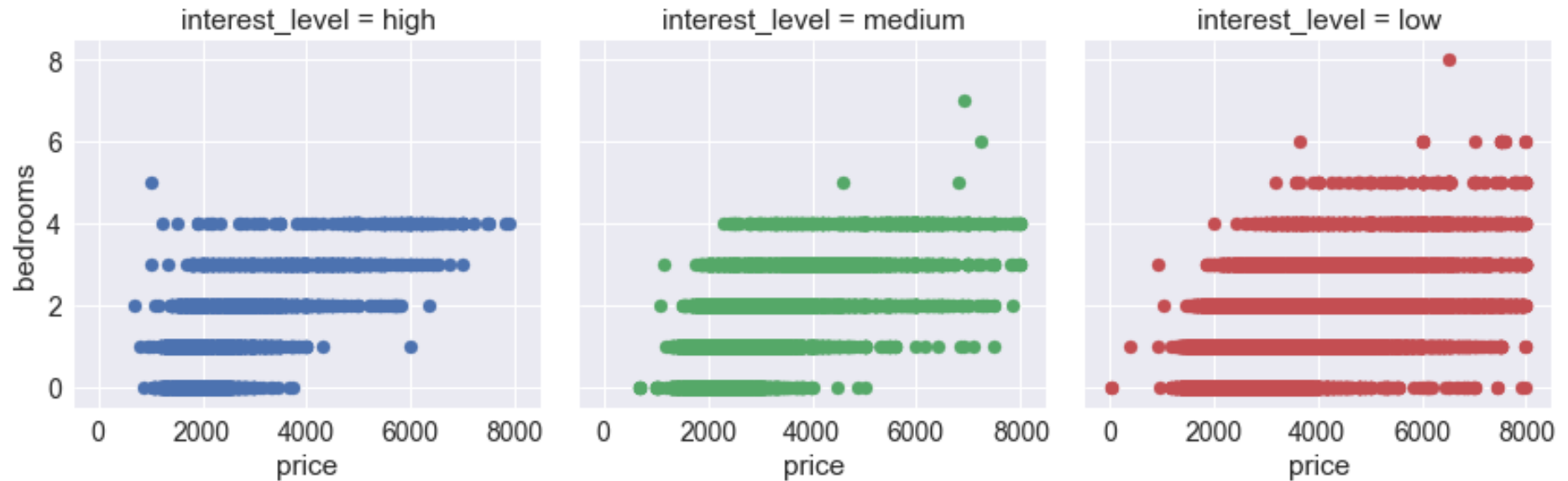
49352 Samples x 15 columns

Date range: April 1 – June 30 2016

# Price Summary

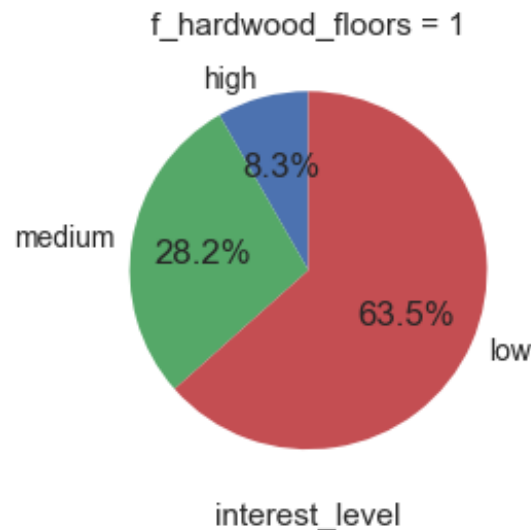
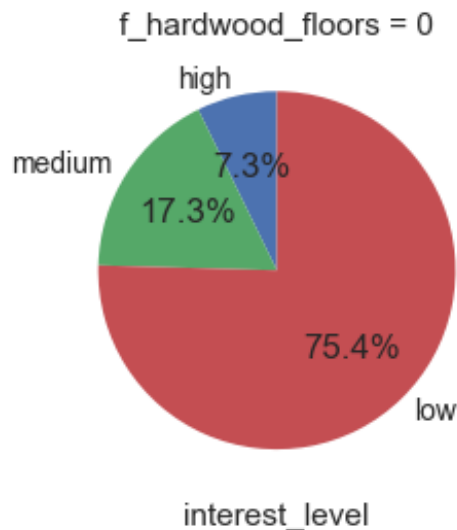


# Price and Bedrooms by Interest Level



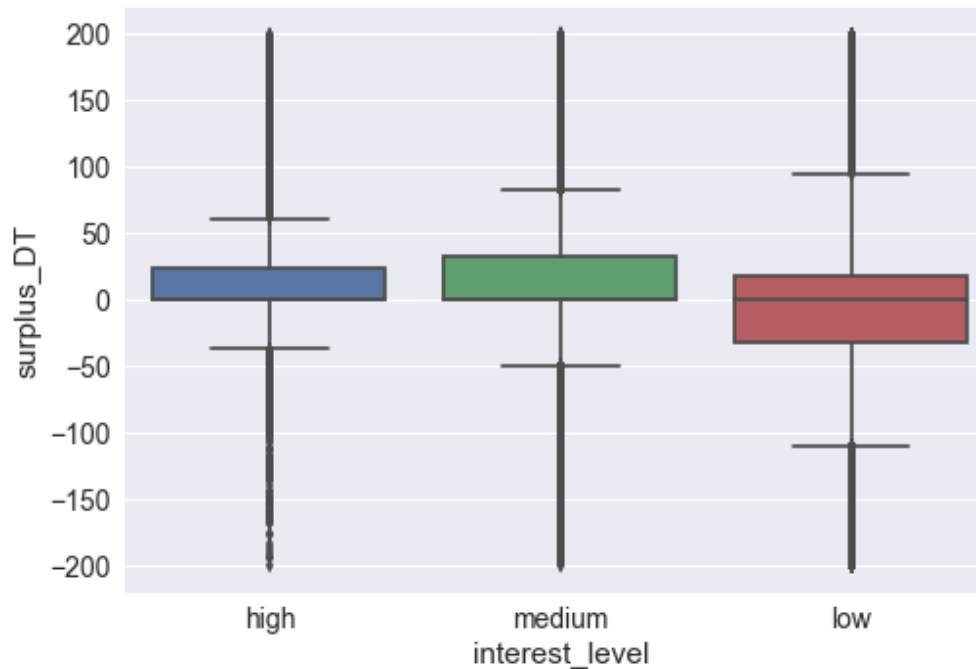
# Task 1: Listing Features

- Found all features occurring  $> 100$  times
- Used Tfidf Vectorizer to identify similar feature strings; collapsed them together
- Manually collapsed others
- Created dummy variables for remaining features (57)



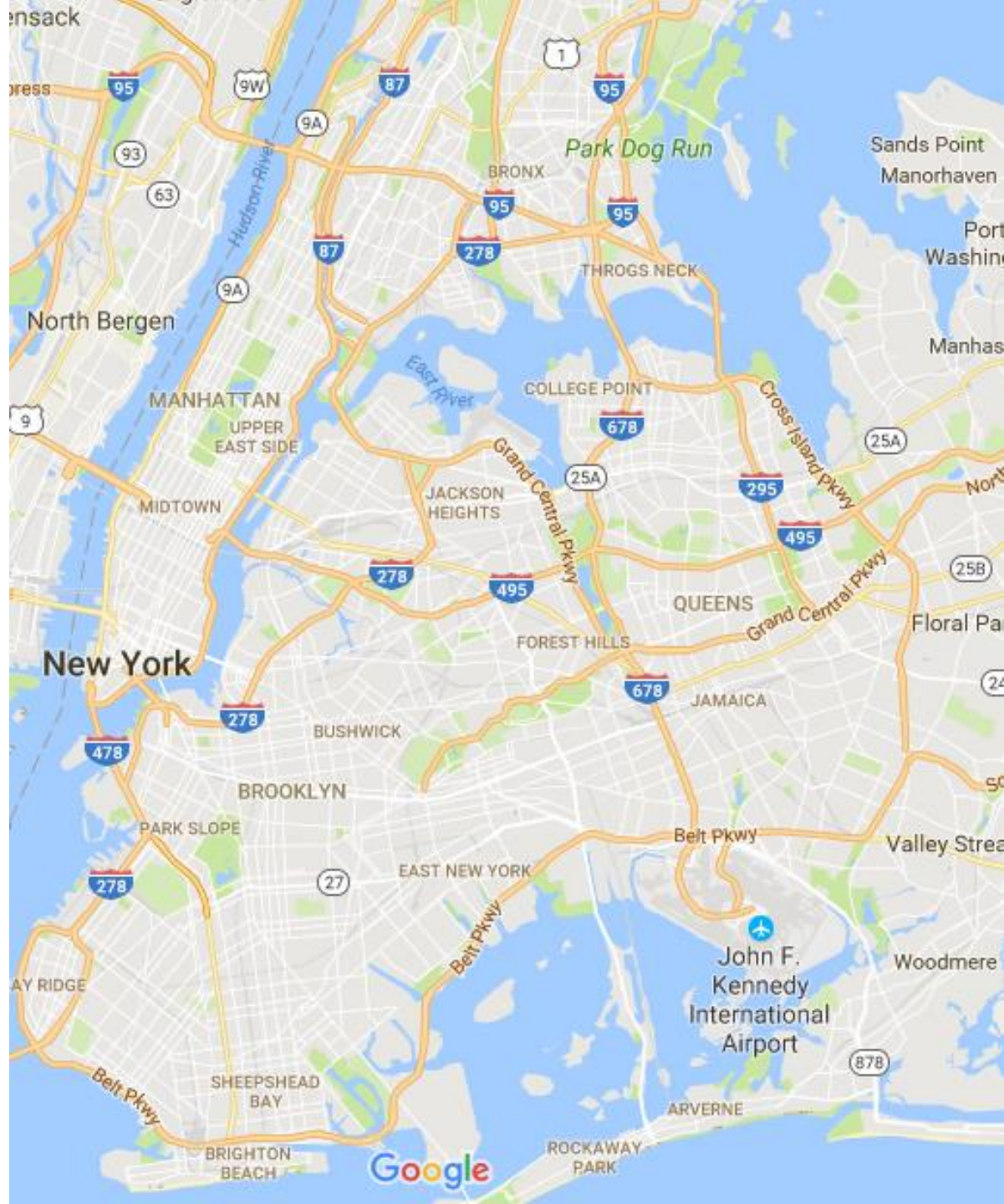
# Task 2: Comparables

- In the training data, ran a regression to predict rent for an apartment that considered only:
  - Bedrooms
  - Bathroom
  - Lon/Lat
- Scored each listing's rent against their predicted rent and calculated a renter's surplus (+ is bargain, - is pricey)



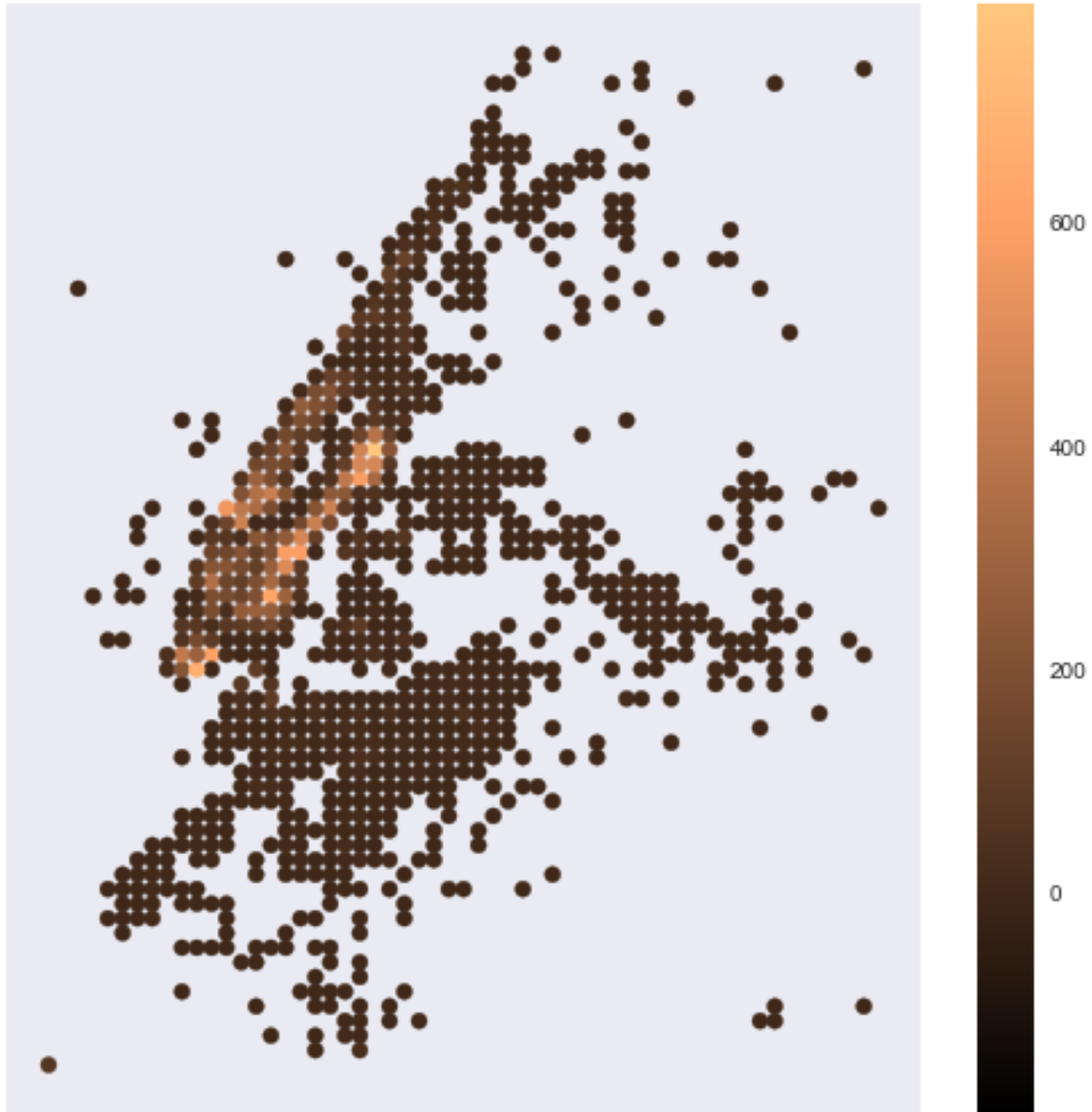
# Task 3: Location

- Overlayed a lon/lat grid on NYC and determined a grid cell for each listing
- Calculated an average neighborhood desirability and used it to boost/penalize the listings in the cell

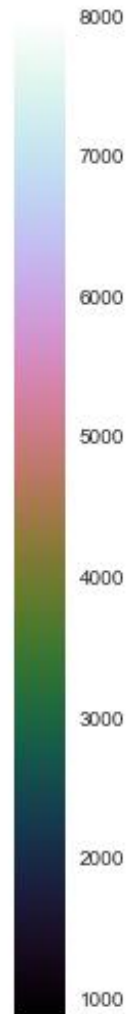




Total Listings by Geo Square



Average Price by Geo Square



Average Interest by Geo Square



# Task 4: Images

- Goal: Determine if a listing has a floorplan
- Converted every image (millions) to same size (160 x 140) and 256 colors
- Tried training images, running them through PCA and then training on floorplans: No success
- **Found a method that worked:** Train a neural network on the image color **histograms**
- Manually bootstrapped the training data 50 listings at a time until model could process 150 listings with no errors
- Used an MLP Classifier (1 hidden layer with 50 nodes)
- Created a dummy variable for each listing: "Has Floorplan"

# Positive Training Data

OneDrive - gwmail.gwu.edu > DATS 6202 Machine Learning I > code > renthop > floorplans > train > positive

Search positive

mail.gwu.edu

Files

Intro to Data Science

Data Warehouse





































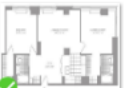


















Machine Learning

its

in

in1

ercises

										
6811966_8b83c24 d1e43d49907fd4e 5d35d9aa56.bmp	6812437_0f4c3333 f0a6ff000343b179 0d045961.bmp	6812493_b83b6e0 94dad1151deb53 c49a57b0812.bm p	6812711_aeb870f 782511a27bd2717 7bfd463e25.bmp	6812933_0fb6e4fa 9cdde35b308fae2 ee5b37387.bmp	6820665_2df8b3e cba310801d65107 bf6d92594e.bmp	6824234_3bd5b1b 0dc28e255617e78 e6b8b2f579.bmp	6826770_8d2c4cb 850f64339ffe82fd b97654fd7.bmp	6826814_86cd733 28cac5529686fa8 db9d4ab16c.bmp	6831188_7e13cb2 92c1874802b244f cb5015cae8.bmp	6860654_2bad3b0 fed57b2a96a25bc 30af9a98d5.bmp
										
6863668_abe375b f89aff3d8d8186ff ebc7a0fe2.bmp	6866345_b8c6fd4 dbb11f1893c0bc1 c2f83f851b.bmp	6882524_5b794a5 d23063baf6b1773 f4e6cde9b5.bmp	6902015_3d19ec9 019836a63d236a0 6212f9999a.bmp	6907177_1ec8b57 debe1f65f2365de 712675d028.bmp	6927662_c03c652 5c7f0a4276baffcb 75f67d4b3.bmp	6928889_2000e9ef 28b9f3c3bb27777 878886b84.bmp	6929419_36dac10 df30888baa5832b 821a44296b.bmp	6958766_fc4eb39 5e287bb7f58cf86 9befd0dd32.bmp	6959042_bf65384 1081c5597bce146 b236723308.bmp	6988818_c7dd666 0623671eeef061bf 6f1dea90a.bmp
										
6989856_e0e3ac4 a651d4f16b04357 57806a173b.bmp	6991070_ab1fbda d7b4323de787e67 b32c148689.bmp	6995737_a864933 eb34ae994c723ea e1900f61b4.bmp	7002508_0dd0962 a74ef687ea6a7b8 57a74a3abc.bmp	7002599_f5cfee61 5abb3d1f54f2257 28cb274c7.bmp	7019362_652264c c6c9b6ef85840de f259bd5dd8.bmp	7030756_a88ede4 87057400be5289e 7fc18c8603.bmp	7041220_fa1baff5 32c8894428d7faf9 db2e34af.bmp	7043911_17dff1ba a17b52e057eaa5b ffc3584ec.bmp	7043994_826a124 10f67da306759b4 d246f5cad4.bmp	7046487_9cde90a d898d8d7d5a8e8 7b9acc3e384.bm p
										
7046575_ae011a1 4122df1146405c7 a8dbd6fc4f.bmp	7050514_dba3e4c ba4d3a9db0b2115 c20070f7ef1.bmp	7058281_115351d 18add5ef798ad4b 2d5b537421.bmp	7060550_f05a6e1a 6ca029eea293f38 1afe2b1d1.bmp	7069502_3458857 27e516b026be083 dab52e2931.bmp	7095936_21bb000 c94597c5bb26147 a08b2dc879.bmp	7102722_e848985 96be72c3e7a41af 36e9a5ea50.bmp	7110275_c477453 da06281043421ed 79ec615548.bmp	7119655_c392b40 27a77c375e0a8d4 f0da77b337.bmp	7121165_86583dd 7d858e475e14c53 34c14e9ccf.bmp	7125623_bf189b8 68c563c2093f3d5 3b73d9251c.bmp
										



# Negative Training Data

gwmail.gwu.edu > DATS 6202 Machine Learning | > code > renthop > floorplans > train > negative

Search negative



6812234\_4100635  
bf753bbd099f642  
a0b849db83.bmp



6812258\_90ef0dd  
81ada1243fdab2f  
051b471d86.bmp



6812275\_525ca8b  
21348483202b447  
e1956d3921.bmp



6812275\_969d285  
4a1625c85adc0a3  
72cabda8f6.bmp



6812524\_4fe179e5  
8f782e16b40bf33  
9c39583da.bmp



6812524\_e5db5ed  
70e73697ba547a6  
71402cc197.bmp



6812711\_0b6ed50  
cf8b59d4f77b39f5  
6b68bb7ca.bmp



6812836\_75836d0  
4cdd25562590fe7  
29ff87b061.bmp



6812836\_282812d  
1a69a00658136eb  
ce919db42a.bmp



6812836\_3663278  
76bd8ce5f961e35  
b80f1b523f.bmp



6814958\_cf06c5a8  
c727d780dd10c68  
e3d1127cc.bmp



6816264\_032895d  
cb455a8462adb55  
3fc20a1c12.bmp



6823346\_a752e0b  
5896ad376d26729  
504ec87b7a.bmp



6825871\_be8946f  
70b0b2f755dcb7  
95ecbdb1a4.bmp



6826770\_0ff63ae5  
db963c10b702745  
7e181fcec.bmp



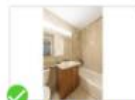
6826770\_897dc40  
08c1519d473da93  
56ea02abce.bmp



6827241\_bfd60da  
a7eb86931af5d66  
d41e090f2d.bmp



6830946\_8dcf6e3  
6434107e71ee87c  
2f14f1172b.bmp



6830946\_3963b09  
c8c79e1a3d84df9  
a16e224bb7.bmp



6844926\_76927d3  
1032203a8a428ed  
6bea628392.bmp



6851303\_7f54ae91  
6f7dd3447fd6466  
70022ecb7.bmp



6851303\_210d535  
5bfee2af2e9728b  
9d64aa86e.bmp



6852129\_41c7e4a  
3b492dc1899e7e6  
f0488a67da.bmp



6856484\_bf43a79  
4b856b499bbb63  
b5afbe4bb6a.bm  
p



6859799\_73d7a58  
bcce8af3f388117e  
561f57484.bmp



6859799\_b4c3a93  
3b2b900a6ba2386  
8ed4e5f151.bmp



6866032\_5508f924  
30f7c57a1adfe5d  
db361d23f.bmp



6866345\_8a22218  
48638e7660f6927  
cc5fa61cbf.bmp



6875098\_eb165ef  
c032b541c48b8ec  
6eee22f8b4.bmp



6886592\_4288953  
ac43f7fd7f74e96e  
5b3da7996.bmp



6899629\_08934e8  
a87d80320557657  
a5d35d6ea5.bmp



6899629\_a44234a  
401d978d0237c5a  
e7a6d0c2e5.bmp



6899629\_d920303  
ab983eb3bc421c4  
76697816b9.bmp



6899629\_e45d955  
37f821806536d53  
8aa2e012f5.bmp



6899629\_f050acef  
8f9ff17dced4bea6  
fe104780.bmp



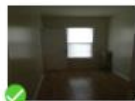
6901378\_04b4fd8  
d0b4e01edb7f744  
c68cab8a48.bmp



6901378\_cd39ca7  
22b7b1c635d897  
d27b73a36eb.bm  
p



6941336\_cced59c  
3f4d0b3755fca73  
c861d91a7a.bmp



6942047\_e2e60e5  
a8c2eade82845d3  
4bf19793e0.bmp



6962003\_9a4d37d  
f17379fccc9840b3  
e8be99e60.bmp



6972829\_1d804d4  
fb9b09b4bb5f036  
52b39c5a56.bmp



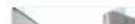
6976219\_0d9928ff  
56ff44912a52da87  
c5f5d929.bmp



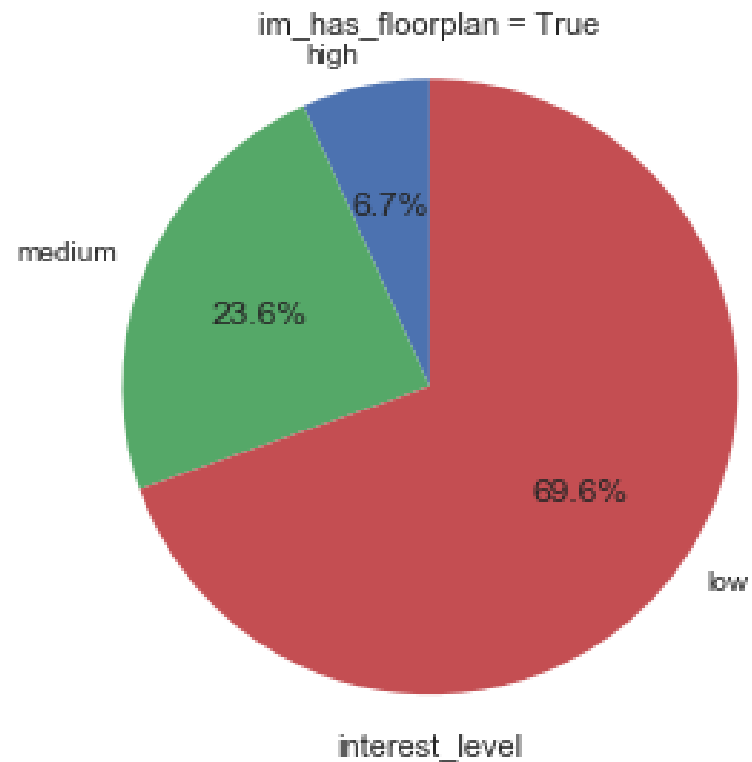
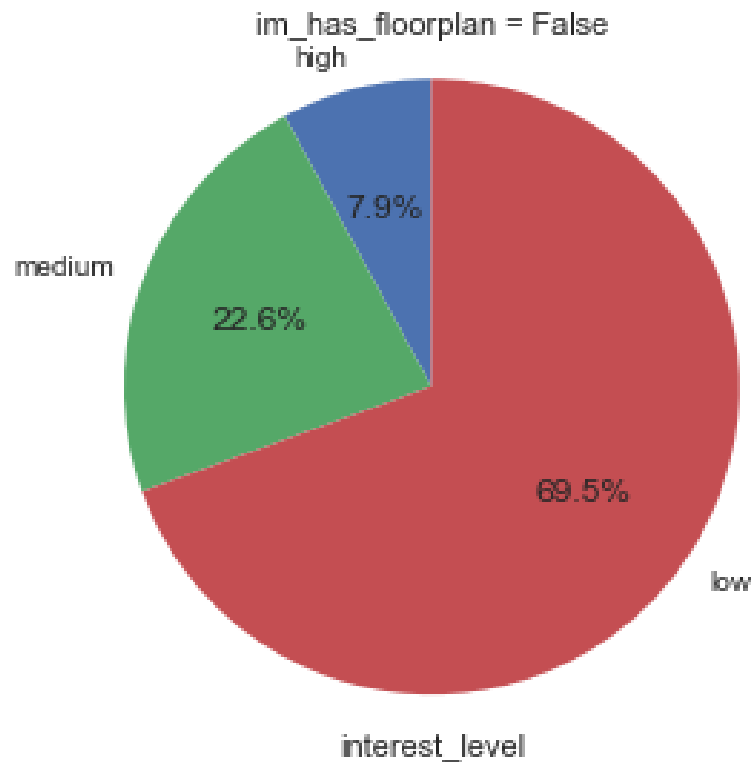
7046575\_727a05c  
cf678dfaed27808  
667ab52ca7.bmp



7057712\_247beeb  
9e760ce659d281f  
3f82b4d7ea.bmp



# Floor Plan Results

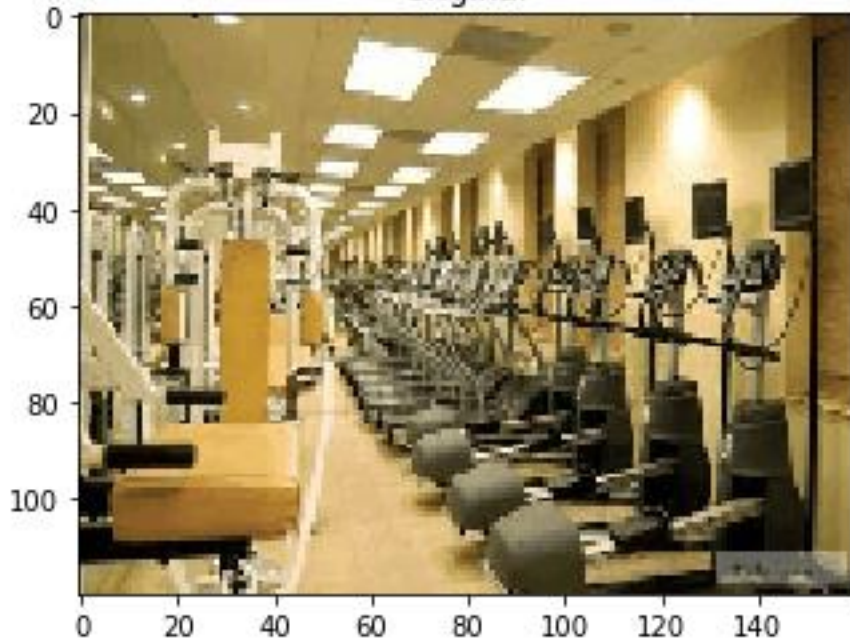


# Busy vs. Plain Pictures

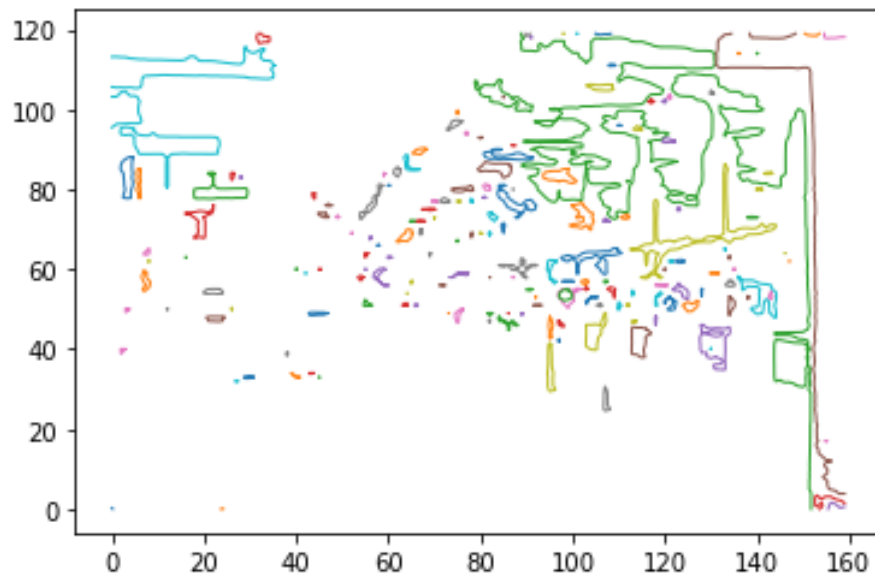
- What works better? Dull pictures of the apartment rooms or vibrant pictures of the view, the gym, the building itself, the neighborhood, etc?
- For each image, used sci-kit image to determine:
  - Number of contours
  - Number of points in edges
- Fed that to a KMeans unsupervised learner
  - Fit it with a subset of the training data
  - Selecting for 2 clusters gave sets of images with a perceived difference when run against larger sets of images
  - Better results not running through PCA first
  - Clustering distinguished “Plain” vs. “Busy” images
- Totals in training data after clustering:
  - 30,032 Floorplans (excluded from the clusters)
  - 1,252,704 “Plain” images
  - 930,136 “Busy” images

# Busy Example

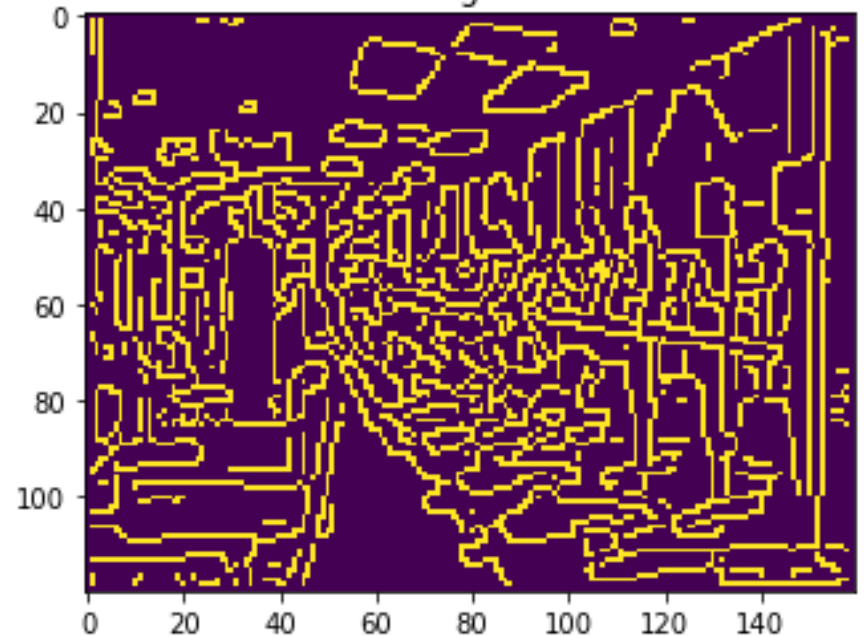
Original



Contours



Edges



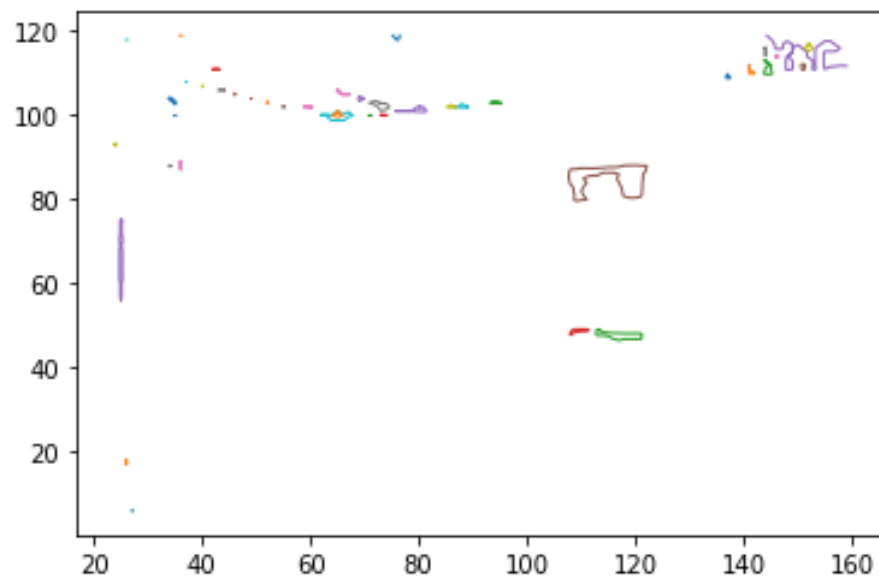


# Plain Example

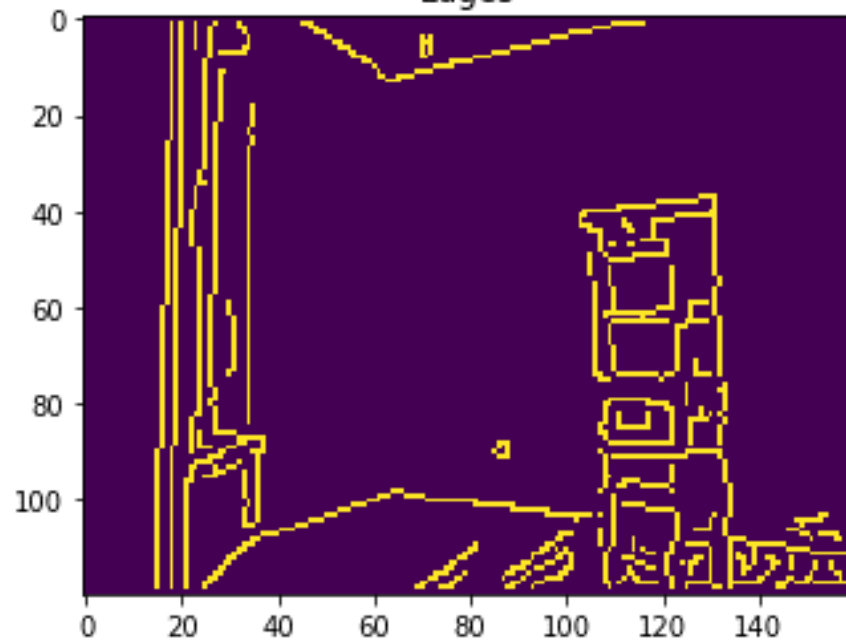
Original



Contours



Edges



# Some Busy Pictures



6933963\_98b2e22  
7b2eeaed0b4a090  
401a84a0a2.bmp



6936114\_42846b9  
1ee215aa3303f3a  
dc90a78969.bmp



6954470\_ac5247e  
0c46e48ed31aac8  
6da73ba42a.bmp



6954470\_d7f9afce  
55d12cc99b1ede0  
7f938b799.bmp



6957326\_5a85f163  
3216d288d1386b6  
09170278a.bmp



6957326\_8f5f2b6a  
400166c48294035  
0199225af.bmp



6957326\_8f750f92  
dbc29c5859dfbc0  
cd140066f.bmp



6957326\_407ea1a  
02d8ce1cf6d5c1d  
8011e3131b.bmp



6957326\_466fa6a9  
b15c17d99b62b1  
71baac5671.bmp



6971257\_44f4f1c0  
64df3edcd42812  
d4eb54ced.bmp



6971257\_13782c1  
6718a42b221a668  
d14276476c.bmp



6971257\_da9d41c  
1d0e07c4b4ebe97  
b1b2bcd7f.bmp



6971257\_f43b284  
c19a7c770ba582f  
e0845b9d31.bmp



6971257\_f468d77  
4f1ead4fc8b71db  
09742d4614.bmp



6977976\_ea87971  
b58618a880f22ed  
55fd79259f.bmp



6989039\_59bfb5ea  
1adfd29ad063436  
f596a80906.bmp



6989039\_80daed4  
4a383b589d7e820  
3134798c98.bmp



6989039\_141e1ca  
471cbbac2c10015  
478d13b236.bmp



6989039\_542c8f2e  
7ae2133efcb0892  
e96153ff8.bmp



6989039\_9617678  
af762c0563c6c8c  
0b57f800df.bmp



6989039\_b31d852  
47ce324a8c49768  
0fdc0486a9.bmp



6994323\_e0f9a1b  
d3383df7722e489  
5f3deedeb3.bmp



6999912\_3e5b1a4  
292f856731c81b7  
ad772f381c.bmp



6999912\_933238fa  
a51b78f62fa34be  
1d74df880.bmp



6999912\_a6a2b0b  
9d70df8848d5f14  
1723760881.bmp



7003486\_3f516d7f  
78d65bd958464b  
3b2b644785.bmp



7003486\_bbde8fc  
6623d9fd1d81cc  
affe4404e5.bmp



7022453\_4fed1d7  
b112fe7d0790cff7  
a63cc30ea.bmp



7022453\_4ff9adad  
f24a3169cff532b5  
f273567c.bmp



7022453\_8f695465  
91dddef286962547  
fdaadfb75.bmp



7022453\_22e8a05  
40664b6a5310d4e  
e823d88981.bmp



7022453\_24d3448  
ffd7aa17794a0d4  
c89939c502.bmp



7022453\_365fb06  
0e90e933d2e0252  
d90879caf6.bmp



7022453\_9829d7d  
07cf70762533be3  
84aca1b9c3.bmp



7022453\_4272049  
644b9c06bf7c6cf  
723e40afe8.bmp



7022453\_e78ba00  
2c4034faf4d26984  
eb26cbbd0.bmp



7026019\_6a73ff4f  
a790c5dc54d2972  
931cc52a1.bmp



7026019\_92e35a9  
413352ba22f8ab0  
2f8ef52aef.bmp



7026019\_adad21c  
54adaa2f8ab7a9f  
4ea3269e38.bmp



7026019\_b08ac0d  
8ef850023acff21  
e4a891a8a.bmp



7026019\_d1b0a4d  
3b97386766deed8  
008c1d16f7.bmp



7026019\_d72a02f  
5a5c64e93cc22f4  
b733d786a3.bmp



7026532\_3ebc991  
1252840696b2d40  
2a8760ef6f.bmp



7026532\_a1e53d9  
de5760530155d63  
f569e88249.bmp





# Some Plain Pictures



6814059\_07f79a6c  
166ed349f8787ca  
85e0e85df.bmp



6814059\_8b08f4c  
595fb2f601a684ce  
6ca317725.bmp



6814059\_3826168  
caad5416ed49258  
958720012d.bmp



6814059\_a0b51a8  
2aa83f359514d85  
5de82a3784.bmp



6814059\_cd36dc8  
6933cf6aea4f2c3c  
7453c6464.bmp



6814337\_1c2f7912  
3dfb3e64069762  
495b02bf4.bmp



6814337\_5e5348cf  
d3387022cdc6838  
b8249c92f.bmp



6814337\_cbc2b2c  
d68263675dded2  
54672cee5fe.bmp



6814337\_efda68d  
4245a6b1299729d  
0644cf6e43.bmp



6815669\_c19e00e  
294f62eb93204f0  
b9bdbb1b6d.bm  
p



6815669\_d3408c1  
400e3bf8d2f8b54  
d41ed764e7.bmp



6815669\_ec75792  
9bf238011660521  
3fe48ade3.bmp



6815906\_35edcff  
cc1319bcbf8e48b  
1406bb26e.bmp



6816406\_92e10be  
2c196334b3462f7  
e5985867a4.bmp



6816406\_221de75  
66c1f9833bca6c8  
589330b43f.bmp



6816406\_d6a0fc6  
daf72bb8c80eae4  
d5ab48ab38.bmp



6816406\_f290255f  
cefc2b2b567a6cd9  
8b5da4201.bmp



6821240\_7a1381b  
9e0989611248d07  
c8ad00e62e.bmp



6821240\_09a6222  
1f92c8d5b29f57fd  
f28fd5003.bmp



6821240\_80b8697  
65d8273de42985b  
4e1dba5140.bmp



6821240\_83b1128  
12d4766f9d5d83b  
6790a0a94e.bmp



6821240\_ac5333fe  
7371ffda07e7fb76  
47e14088.bmp



6829112\_4a8bbb2  
c082651fb4a5926  
b6a3ef6378.bmp



6829112\_82dfd05  
c7933a129d5bf1  
918c86f08d.bmp



6829112\_2328d48  
8643cbcb8de57fa  
a6dee057a0.bmp



6829112\_a963636  
2d6d4b907cd83b  
2eabc2776f0.bmp



6829112\_c9faedd  
8a0520c870df100  
c9e1dd456b.bmp



6832285\_2c8070b  
5b03f790a5b0efd  
4028e7f4a.bmp



6832285\_3aa347b  
31997fd9a5f1227  
36f06aa10.bmp



6832285\_3d25c0d  
6fafa3895727fbf  
16fb94a6d.bmp



6832285\_3ece5e8  
9b1eb5f2af8b191  
22e2503e68.bmp



6832285\_4ec6121  
754a4ed4cce20fc  
d73cb98396.bmp



6832285\_7330a40  
003bb516d37f4ee  
262a3a1c81.bmp



6832285\_4211314f  
dbb9f37a7d1d563  
c7e073625.bmp



6832285\_ac46da0  
2727cb831a8c599  
db7058d18c.bmp



6832285\_b13c8a6  
eee3a183584e987  
45451b69a3.bmp



6832285\_c6df1b0  
786018fc303cbbd  
6d36e56940.bmp



6832285\_ee3ea8d  
db55c038fe8bf42  
5532da1294.bmp



6832285\_fe7870c  
dda63137e2c01e7  
a3c6b1f38e.bmp



6833599\_6a62706  
7b3512244edcf8  
4db375d806.bmp



6833599\_0376c6fc  
650c4f2a1bfa9651  
1739fc6d.bmp



6833599\_52140a6  
40b39447d08b832  
3f03b6e0f9.bmp



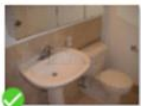
6833599\_c4cbb61  
40b39447d08b832  
e7bec456fe.bmp



6835078\_dc7bc37  
42c5f87b25cfddb  
efd7b43abc.bmp



6838445\_02b61f5  
e5be59a4801b273  
85008f4540.bmp



6838445\_e7be91c  
7eaded1de1db6c  
95fc45ad90b.bm



6845240\_0f35c9d  
6b91d5982b18239  
91e8a3cb44.bmp



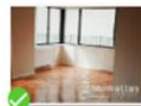
6845240\_719c1e1  
86ff39a57ac1113  
f8d5453ce.bmp



6845240\_7830ed0  
4daff920a76c7af5  
0803e570e.bmp



6845240\_cf16fd0f  
5cfb8c0ff4f41e8e  
9c0a8c02.bmp



6847952\_e646dbe  
76512d755c245da  
a882ff72c3.bmp



6848173\_5190f53  
b975949f6413248  
2302900c22.bmp



6849058\_5ffe4611  
796eed2f9f7d5b1  
9ac3c09e2.bmp



6849058\_7ff73acc  
7c45e8a16aa1b3b  
df3a7257a.bmp

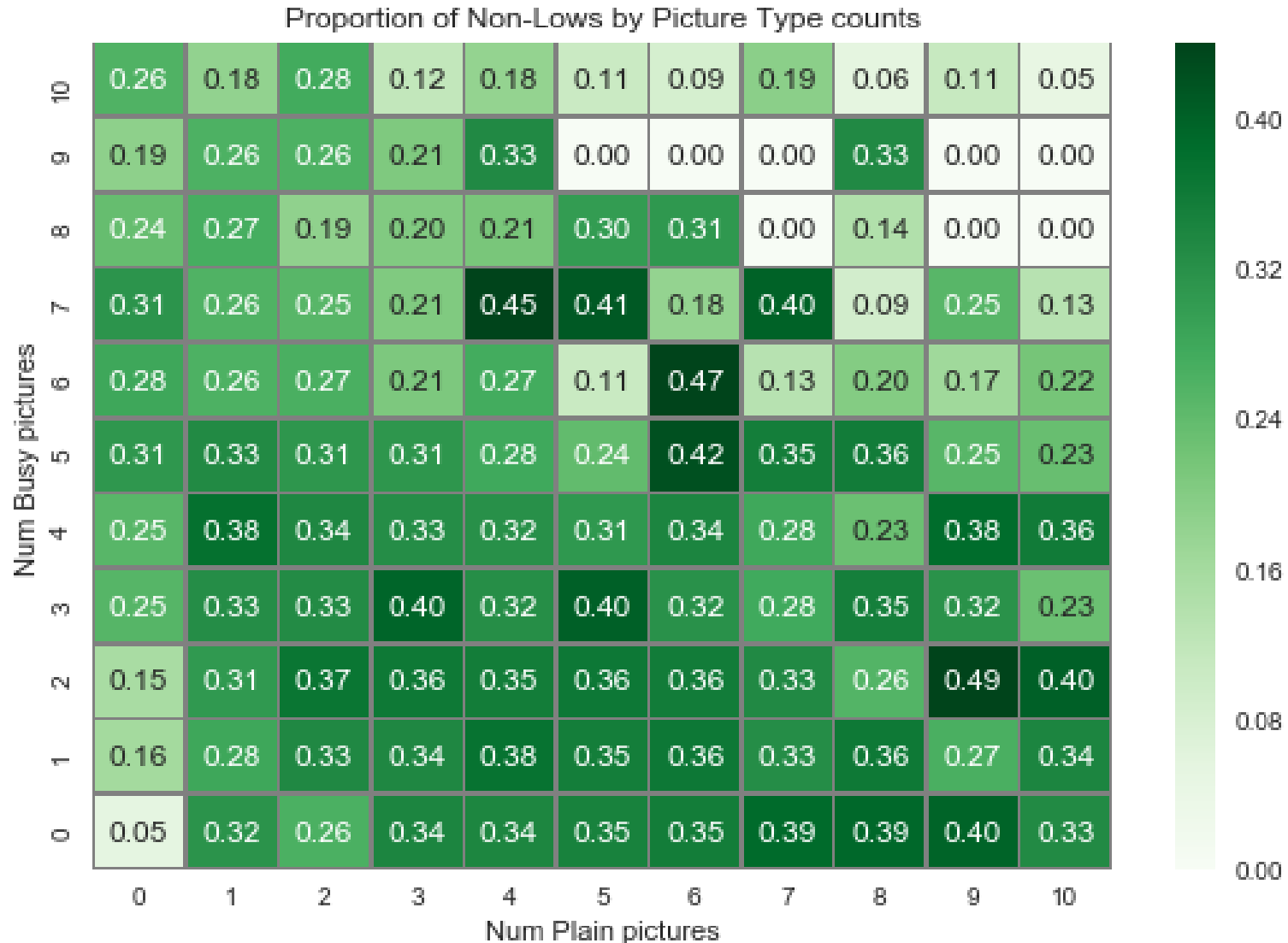


6849058\_42ddb79  
bc2da9a30419920  
4646aab5b0.bmp

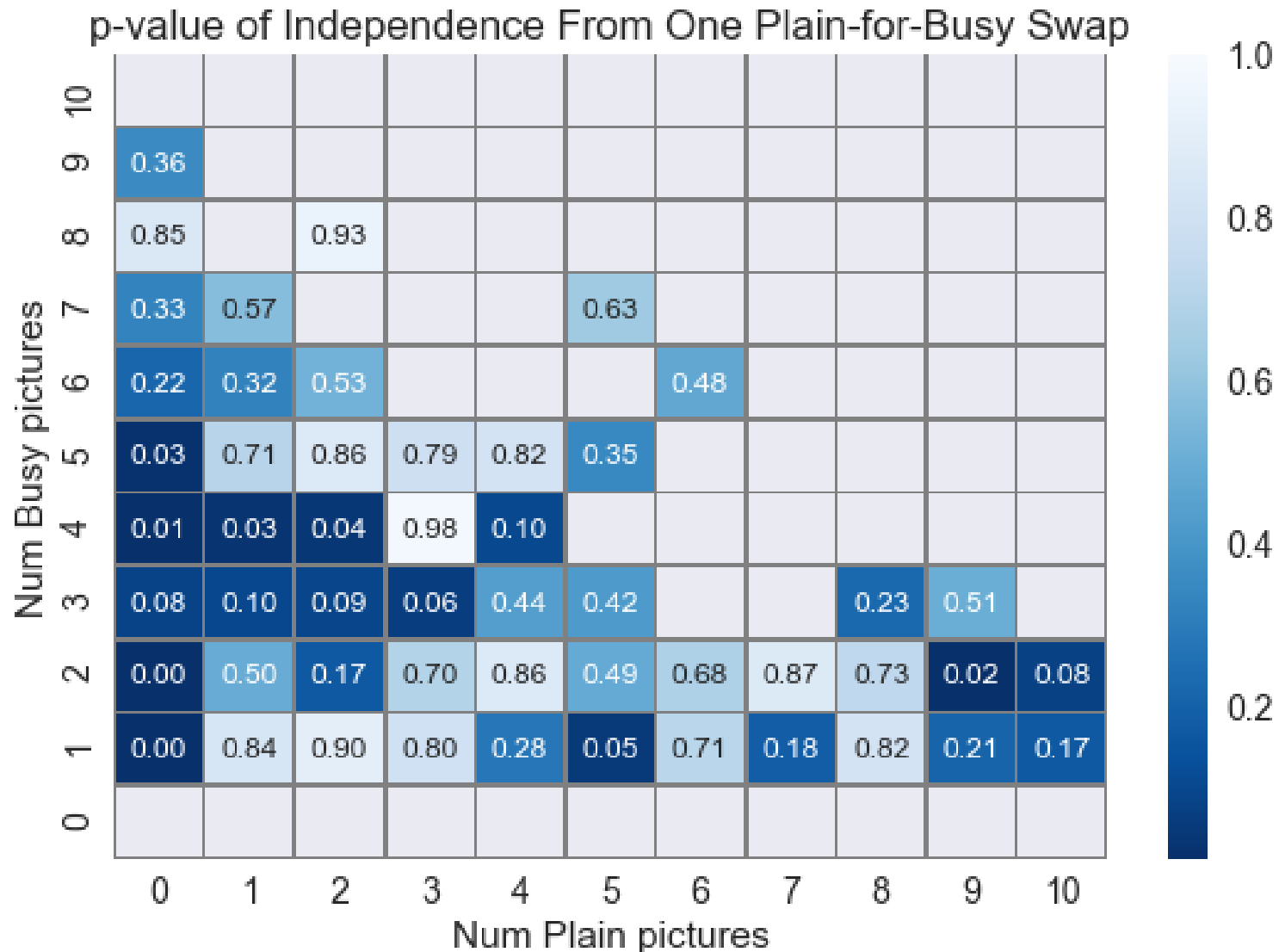
# How Pictures Were Used



# Success Metric for Picture Combos



# Plain-for-Busy Chi Squared Test

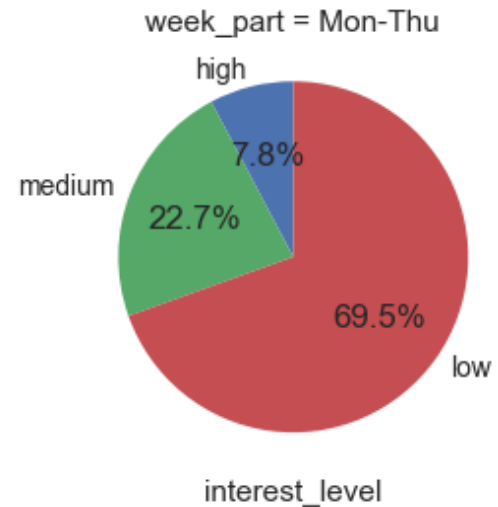
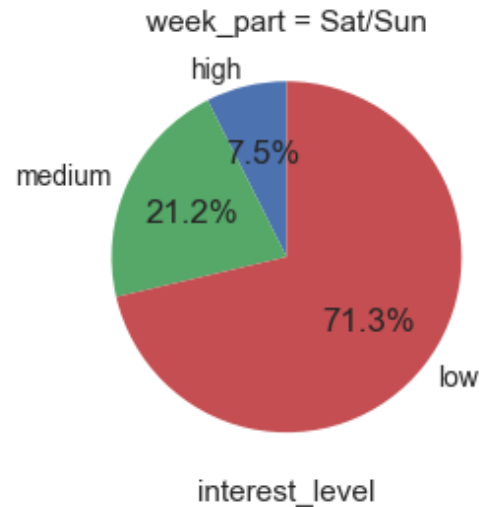
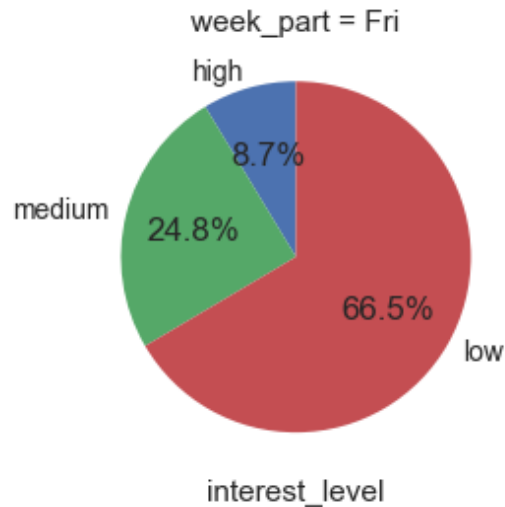


# Task 5: Text Description

- Used a tf-idf Vectorizer to identify 1-, 2-, 3-grams in the text description
- Min frequency: 1%, max Frequency 98%
- Approximately 800 words in the vocabulary
- Using only the description vector to predict the label – log loss score of 0.74
  - Null Log loss model: scored 0.787
- Created a feature for predicted non-low based on text alone and included it in the main model

# Miscellaneous

- Created dummies for day-of-week the listing was created



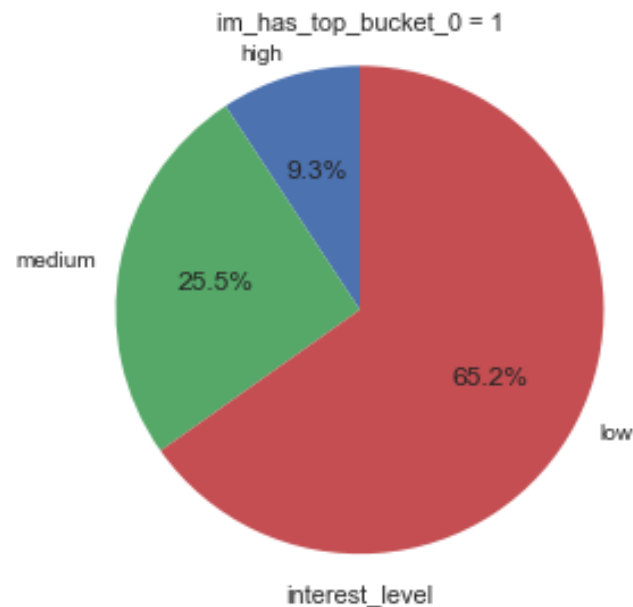
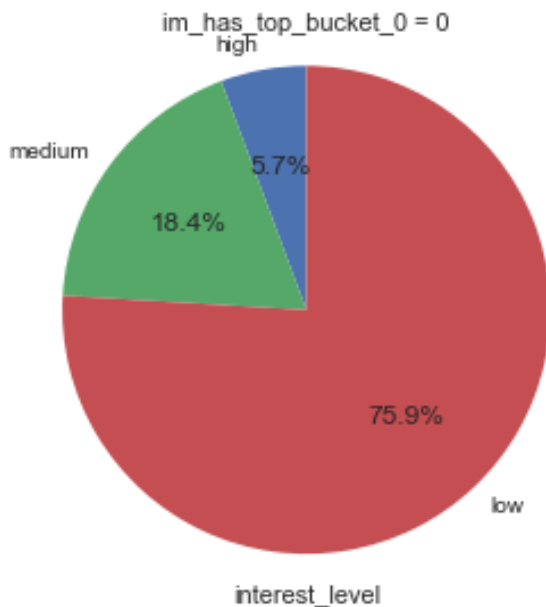
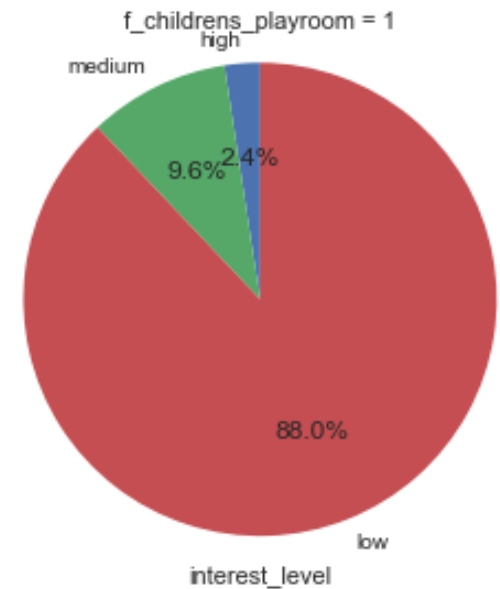
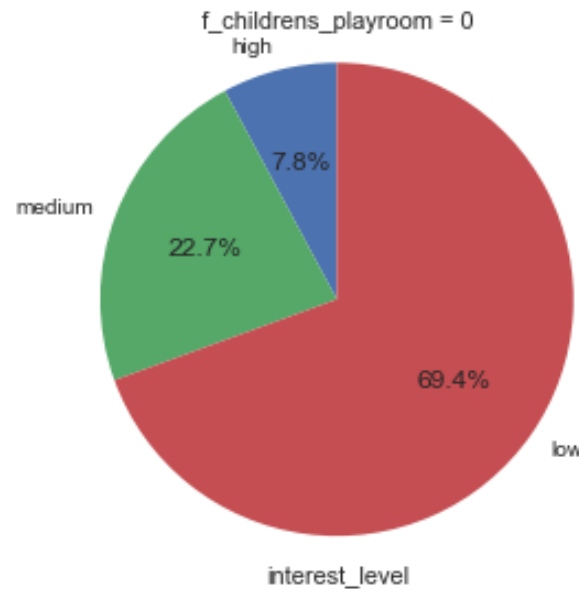
- Created a boost for successful property managers – related to the HOP score and user ratings of the agents
- Created metadata about the listing such as:
  - Percentage Upper Case in the description
  - Log of description length
  - HTML formatting in description; website URL in description
- After feature engineering, new training data shape: 49352 rows x 123 columns + 2 label columns (original categorical converted to integer)



# Top Binary Features Selected by MRMR

Feature	Feature
Has Building ID	Has More Plain Pictures
No Rental Fee	Display Address Starts with a Digit
Reduced Rental Fee	Has 5 bedrooms
Furnished	Has 1 bedroom
Dogs Allowed	Laundry Room
Renovated	Public Outdoor Space
Hardwood Floors	Created on Friday
Washer in Unit	Fireplace
Has 6 Bedrooms	Newly Renovated
Has 8 Bedrooms	Children's Playroom

Playroom  
is a  
penalty



Plain  
pictures  
are a boost



renthop

Search

Resources

Favorites

Find Roommates

Post Rental

Login



New York City, NY

Studio

1

2

3

4+

Lft

Rm

\$ 0

to

\$ 50000

Filters

SEARCH

## Apartments for Rent in NYC

New York, NY

Updated - April 23, 2017

## Upcoming Open Houses

Sun, Apr 23	4:00pm - 5:00pm	\$2,750	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$2,704	Studio, 1BA at 247 N 7th Street
Sun, Apr 23	4:00pm - 5:00pm	\$3,478	1BR, 1BA at 247 N 7th Street

« Back | Page 1 of 366 (7,312 Rentals) | Next »

Sort: HopScore | Price



Posted 7 mins ago



Posted 6 mins ago



100 2BR, 1BA at 100 Ave.  
Midtown Manhattan, Manhattan  
\$7,200 Per Month  
By Tal Cohen

Check Availability

No Fee · Doorman · Elevator · Laundry in Unit ·  
Dishwasher · Loft

100 1BR, 1BA at 58 Linden Blvd, Apt 2H  
Flatbush, Central Brooklyn, Brooklyn

\$1,800

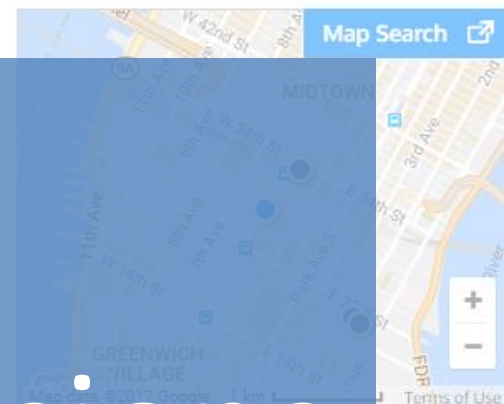
Per Month

By NYC Leasing

Check Availability

By Owner · Exclusive · No Fee · Elevator · Dishwasher

100 1BR, 1BA at Water St.



Map Search

New York City is the world's greatest city. Of course, that is open to debate in many parts of the world, but among the residents long and new, there is simply no contest. Whether you seek gourmet dining, legendary Broadway shows, luxury brands, or more humble interests; you'll certainly find what you are looking for somewhere within the five boroughs of NYC.

Probably the most difficult part of living in New York is actually moving here, both because of the difficulty in finding NYC...

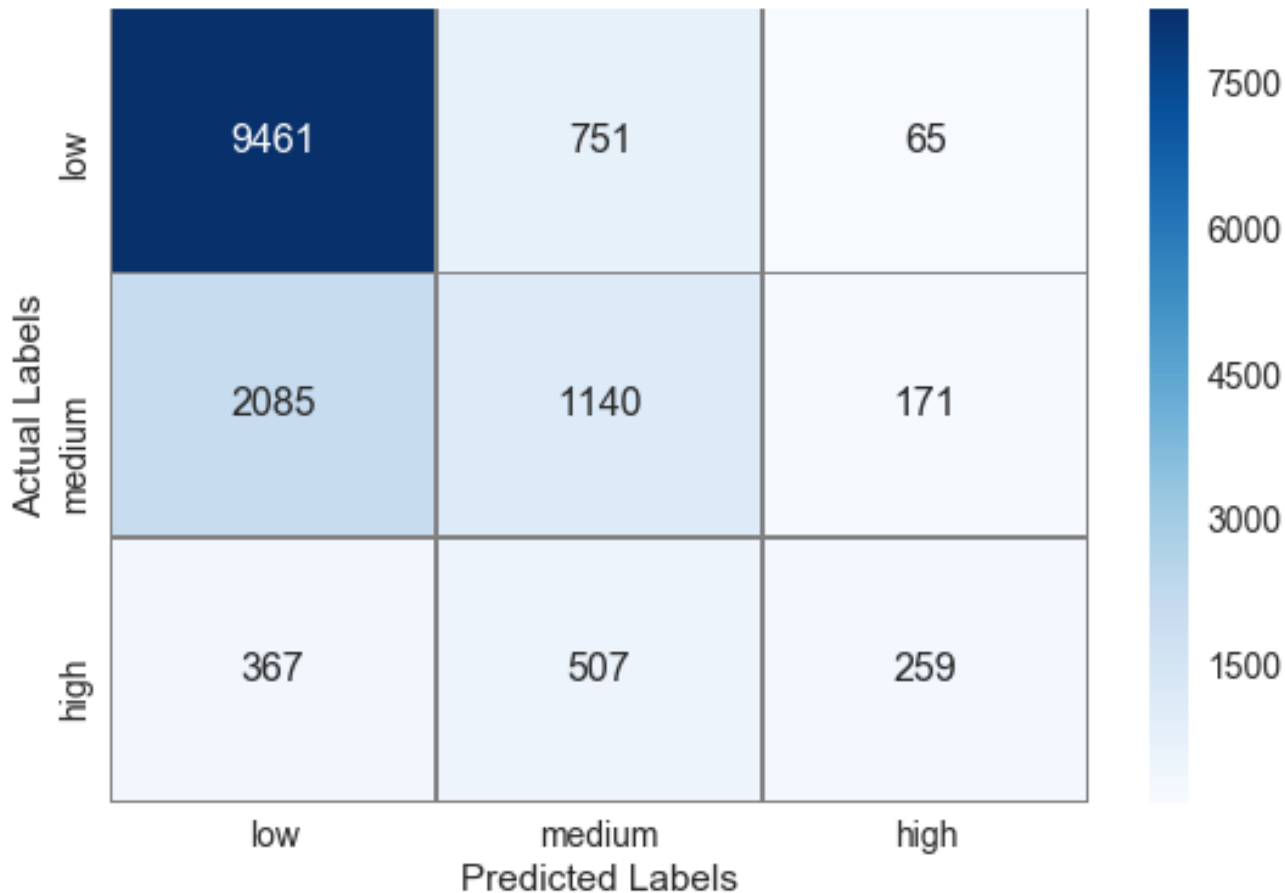
[Area Guide](#) · [Popular Buildings](#)

# Results of Modeling

- Three best performing models:
  - Random Forest (1000 classifiers, auto n\_features)
  - MLP Neural Network (1 hidden layer 10 nodes)
  - Logistic Regression (C=1, l2 normalization, multinomial)

# Random Forest

Results for Hold Out Test Data (Random Forest)



# Random Forest

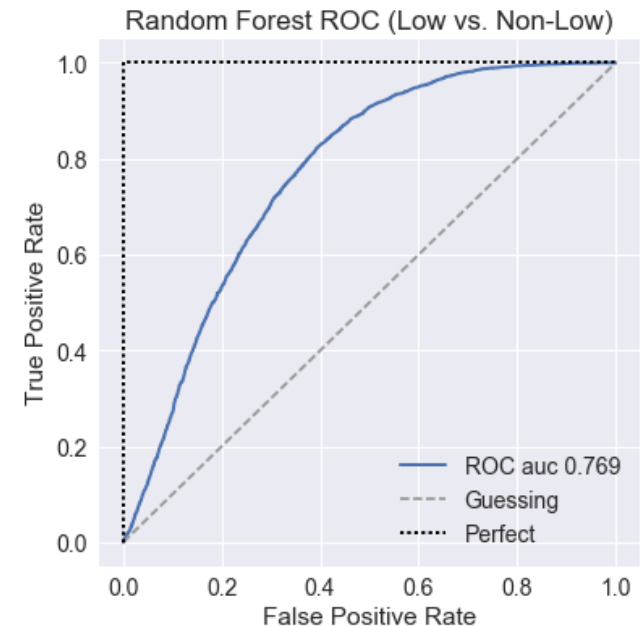
```
Feature importance:
manager_boost 0.207346114229
price 0.094255241664
desc_prob_0 0.0934663652871
neighborhood_boost 0.0674976774421
surplus_DT 0.0663217209287
m_description_length_modlog 0.0658757370221
m_description_UC_pct 0.0649307769437
grid_y 0.0468302964799
listing_feature_count_modlog 0.0453032392492
m_photos_count_modlog 0.0440506134713
grid_x 0.0434702193913
m_has_building_id 0.0208387698602
created_on_wed 0.0110780265638
bedroom_2 0.0102545084573
created_on_tue 0.0102157632852
created_on_thu 0.00992834795607
created_on_fri 0.00974111421575
bathrooms 0.0094823578633
im_has_bucket_1 0.00941038349117
bedroom_1 0.00940995126481
created_on_sat 0.00894835240923
im_has_top_bucket_0 0.00780621686129
m_display_addr_digit_start 0.00768078551956
im_has_top_bucket_1 0.00728999076254
bedroom_3 0.00717192996728
created_on_sun 0.00652371449416
im_has_bucket_0 0.00636962516371
im_has_floorplan 0.00483609621372
bedroom_4 0.00314354338473
bedroom_5 0.000476242460198
bedroom_6 4.61494481599e-05
bedroom_8 1.28249479584e-07
bedroom_7 0.0
```

- Cross-validating Grid Search selected the non-zero probability from the description over the Listing Features list – but it was very close
- Best run against hold-out test data was with the list of features

```
Random Forest

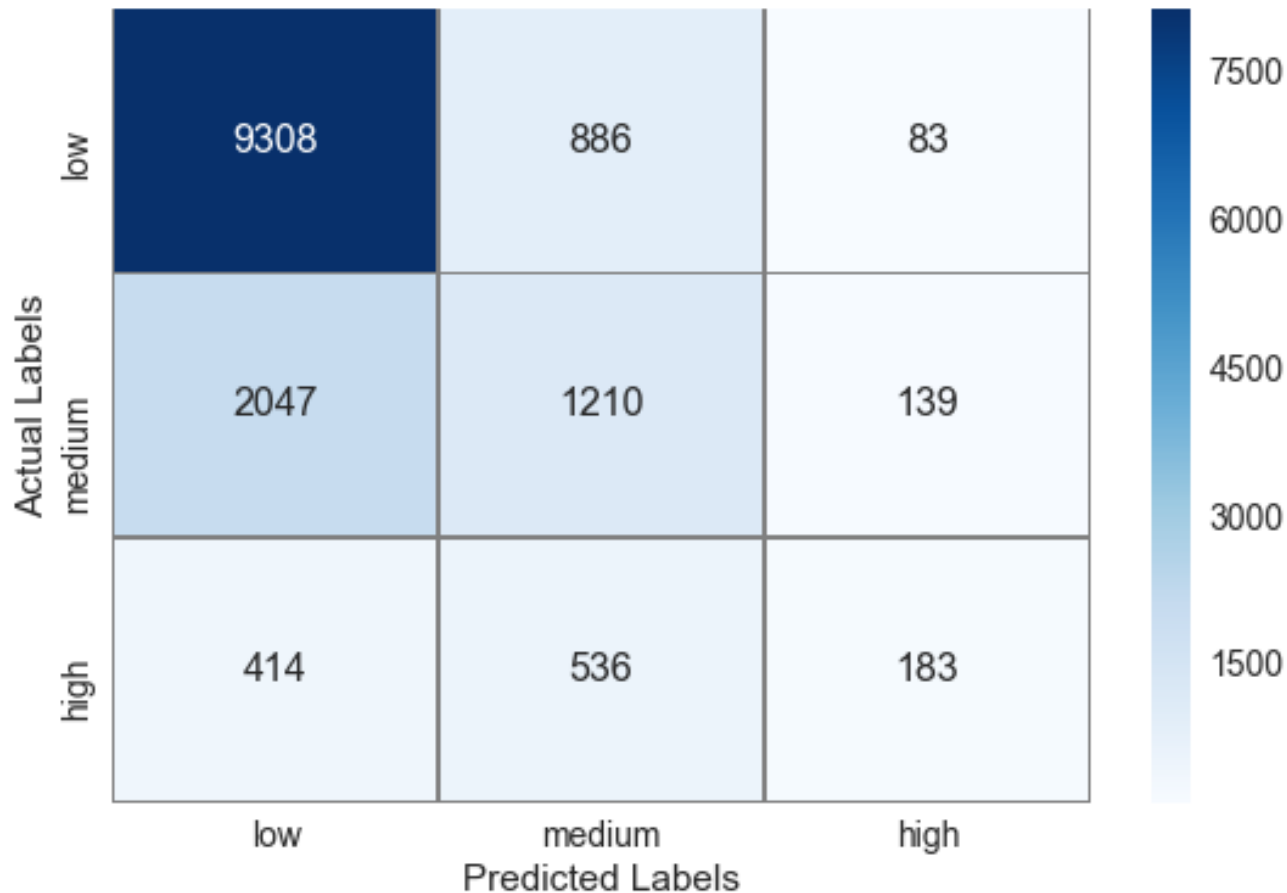
Log loss: 0.593958277915
F1 score: 0.706492740422

[[9461  751   65]
 [2085 1140  171]
 [ 367  507  259]]
Random Forest Adjusted)
```



# Logistic Regression

Results for Hold Out Test Data (Logistic Regression)





# Logistic Regression

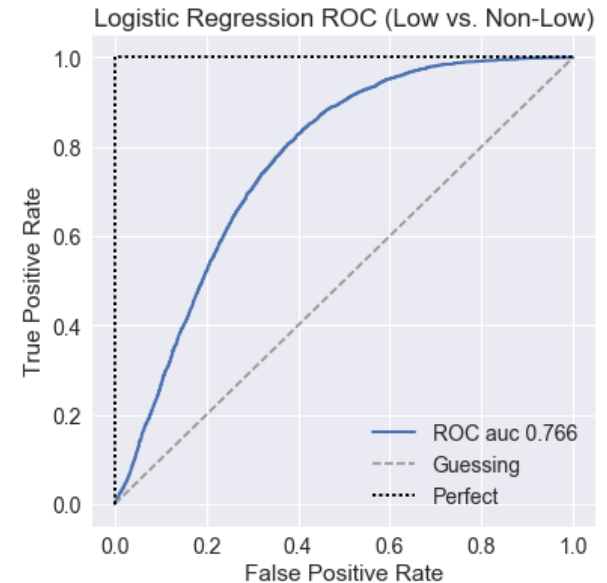
```
Coefficients:
bedroom_4      -1.38585392671
manager_boost  -1.24141179436
f_subway       1.13222479516
f_furnished    -0.978486168256
price         0.837056826108
bedroom_3     -0.783683585396
im_has_top_bucket_1 -0.769325882987
im_has_top_bucket_0 -0.722079317938
m_has_building_id -0.704392244677
f_simplex      0.530237447817
bedroom_5      0.451675279184
f_onsite_garage 0.409496586251
f_newly_renovated 0.407701458373
f_reduced_fee  -0.383760570101
bedroom_2     -0.368601305702
f_multilevel   -0.339567376893
f_no_fee       -0.31637979972
f_common_outdoor_space -0.305412302998
f_renovated    -0.295512834513
f_stainless_steel_appliances -0.288331906157
f_exclusive    -0.287969207776
neighborhood_boost -0.26352562922
f_concierge    0.240088720421
f_laundry_room 0.226048576103
f_laundry_in_building -0.211937225452
f_private_outdoor_space -0.208576042196
f_bike_room    0.20063921899
f_live_in_super 0.190363592592
f_washer_in_unit 0.188912537562
f_washerdryer  -0.188478297162
```

- Predicting the negative case, so negative coefficients are related improved success

```
Logistic Regression

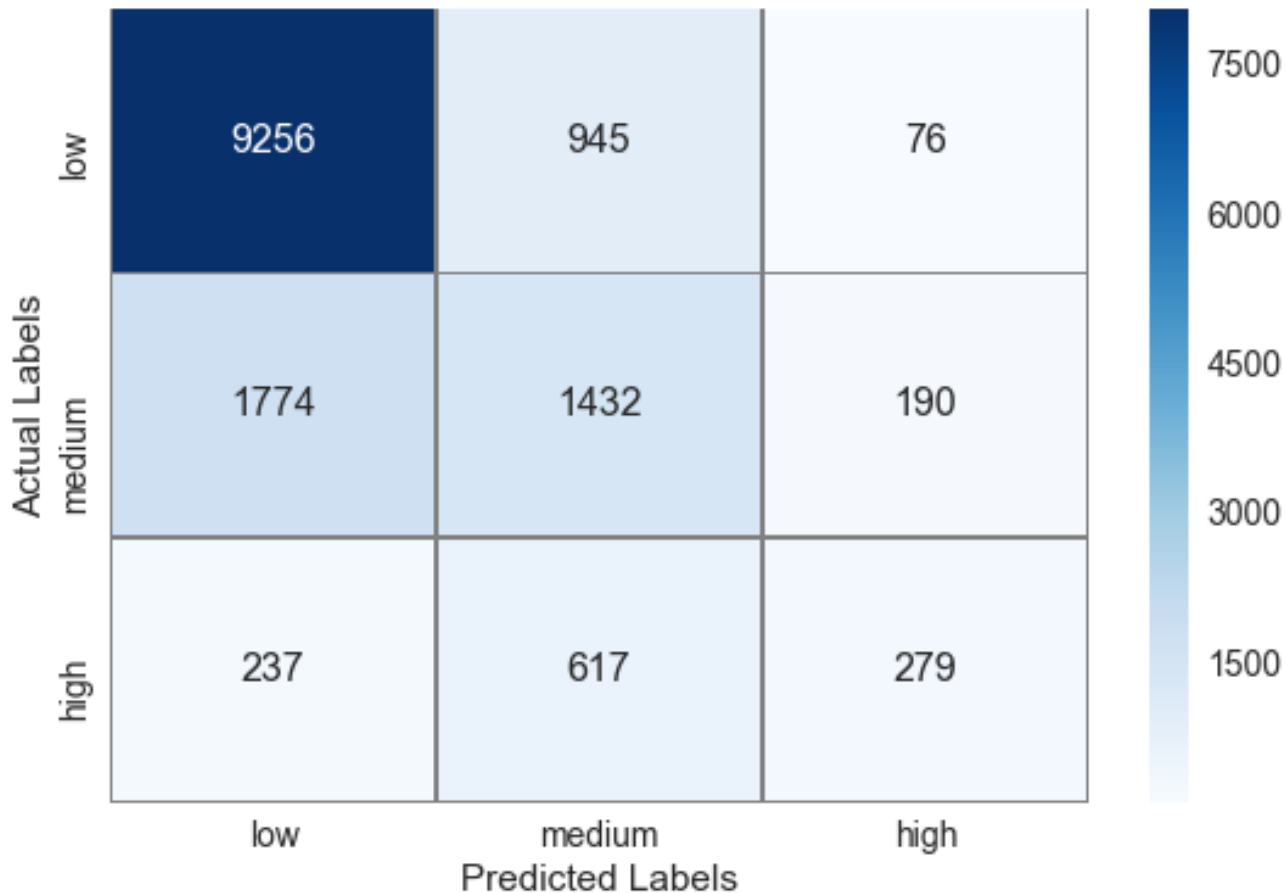
Log loss: 0.620872752432
F1 score: 0.696409884514

[[9308  886   83]
 [2047 1210  139]
 [ 414  536  183]]
```



# MLP Neural Network

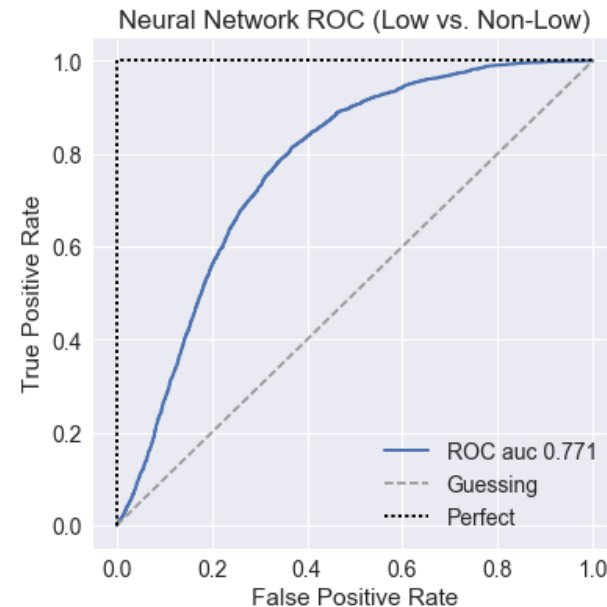
Results for Hold Out Test Data (MLP Neural Network)



# MLP Neural Network

- Log Loss almost as good as Random Forest
- Did the best on F1 score
- Grid Search selected the simplest network (1 hidden layer, 10 nodes)

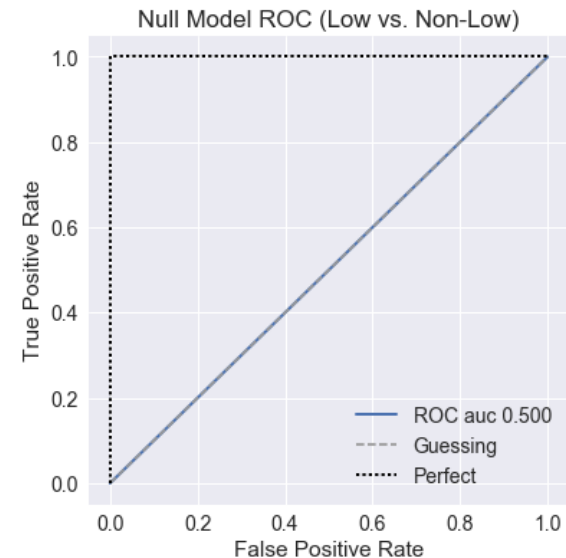
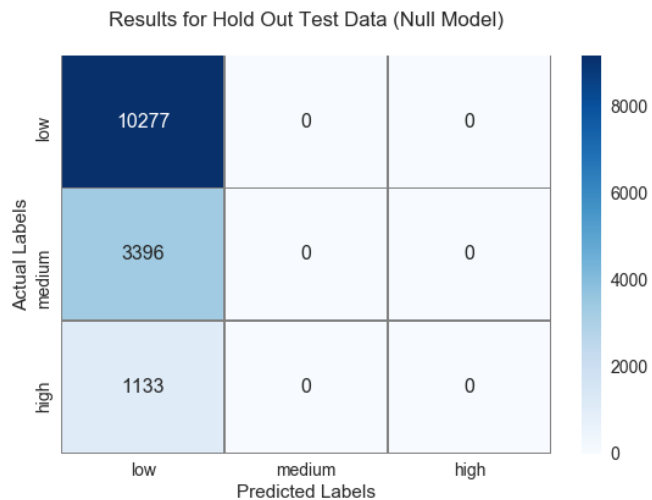
```
Neural Network  
  
Log loss: 0.605959498991  
F1 score: 0.724673686007  
  
[[9256  945   76]  
 [1774 1432  190]  
 [ 237  617  279]]
```



# Null Model

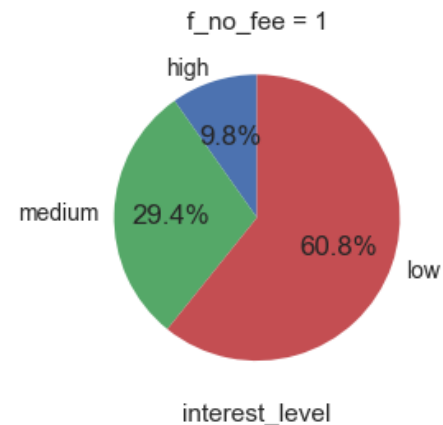
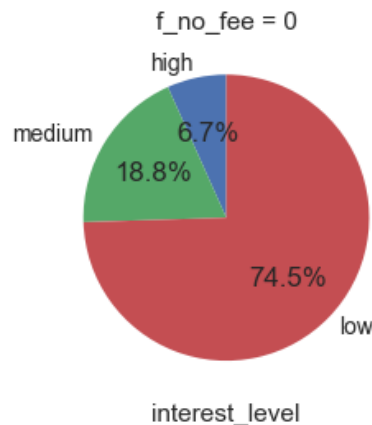
(Predicting every listing has the same probabilities as the population average)

```
Null Model  
  
Log loss: 0.787858212727  
F1 score: 0.5687815305  
  
[[10277    0    0]  
 [ 3396    0    0]  
 [ 1133    0    0]]
```



# Insights

- For listings, pictures are important. At least one should be a picture of the apartment
- Try and have the listing hit on a Friday
- Networking / Social Profile is important for agents
- Keep the price down, especially the rental fee
- The description may not be that important if the pictures and feature list tell the story. There were very close results substituting description score for the list of binary features



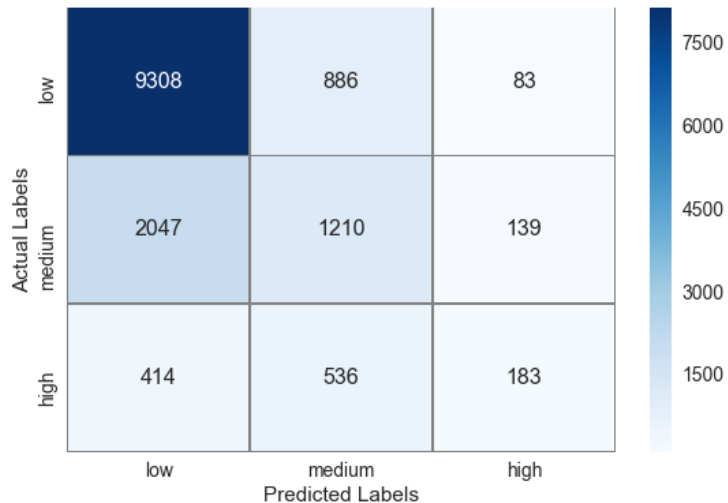
# Further Study

- Images
  - Identify kitchens and bathrooms
- Description
  - More sophisticated analysis needed
- More data on the relationship to the property manager would be desirable
- Model of Models

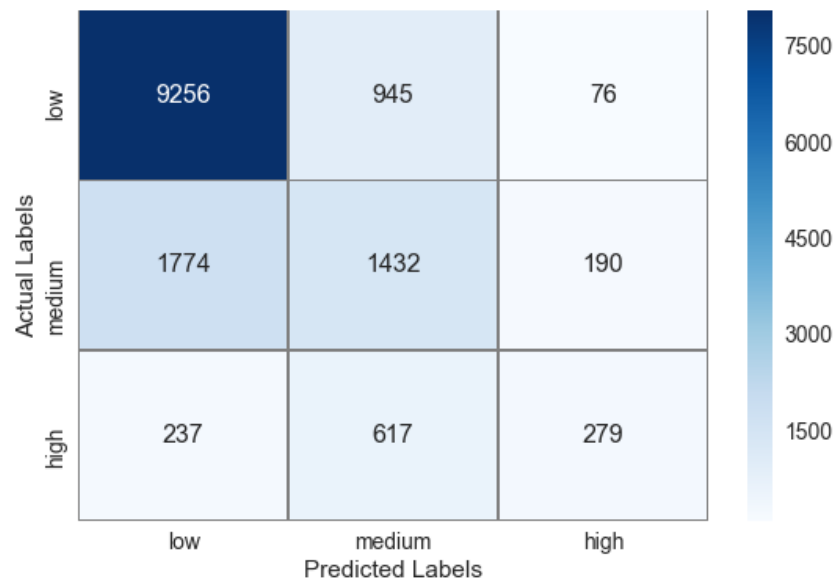
# Backup Charts



Results for Hold Out Test Data (Logistic Regression)



Results for Hold Out Test Data (MLP Neural Network)



Results for Hold Out Test Data (Random Forest)

