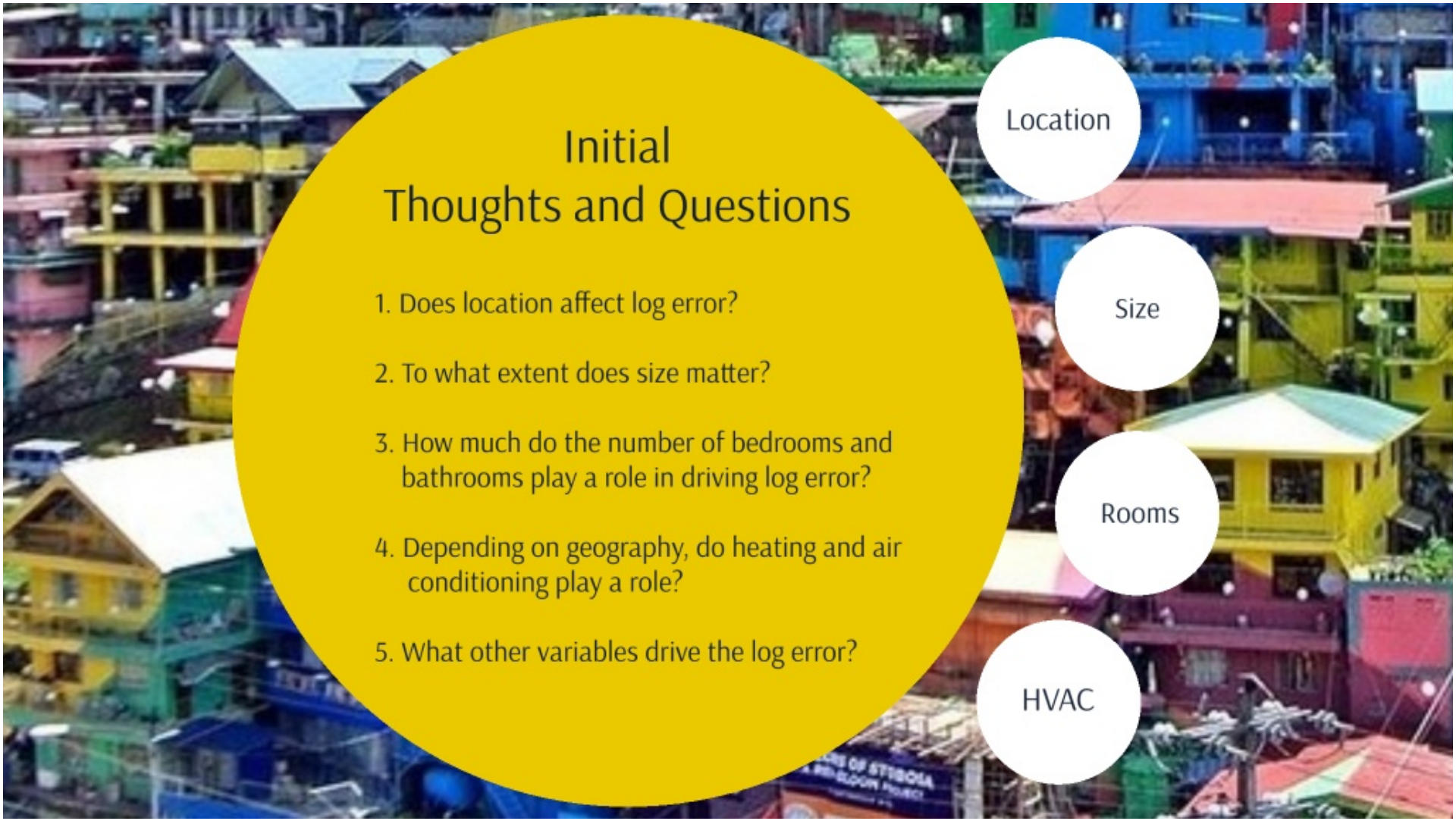
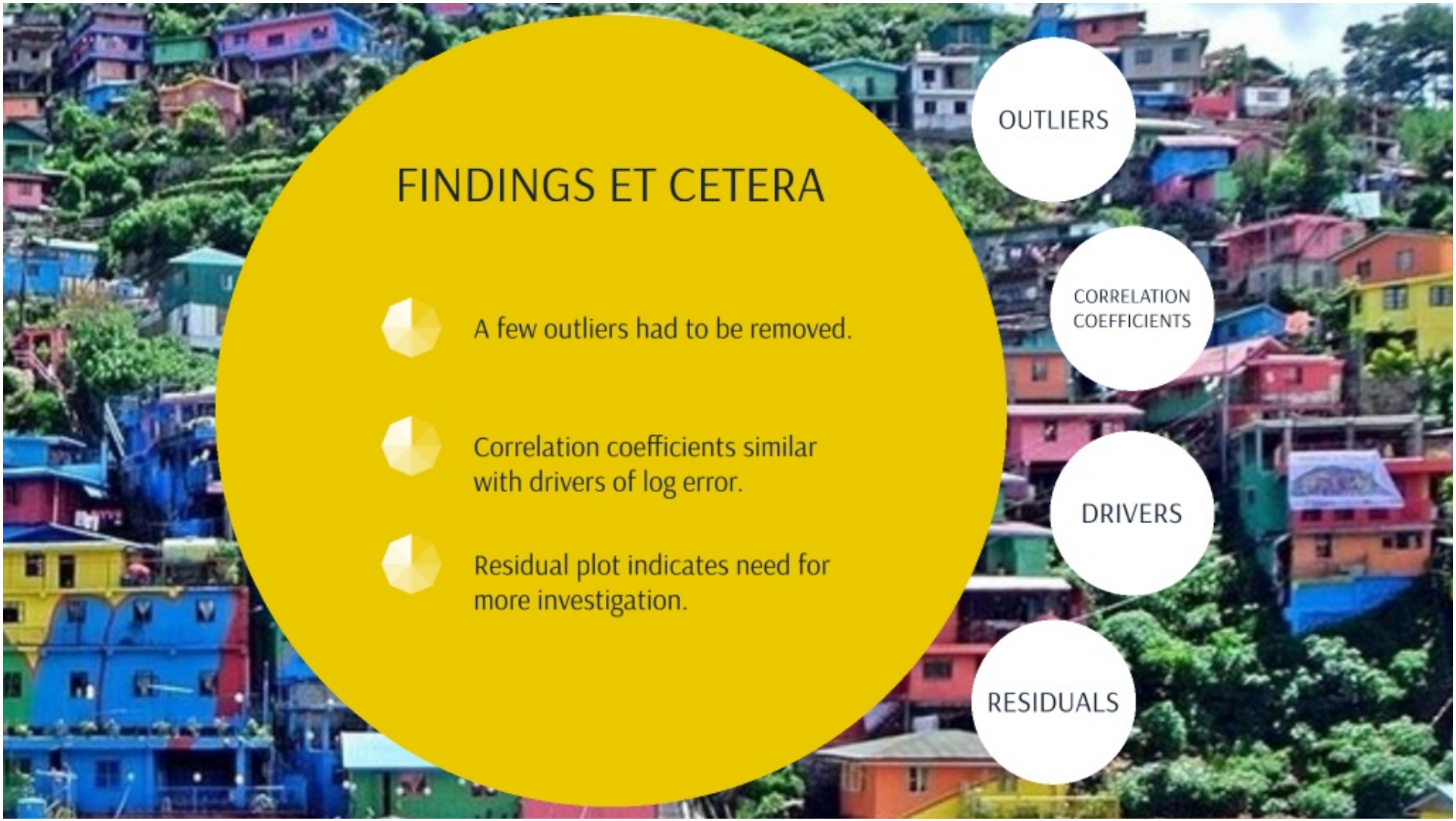
Zillow
Zingzestimate
Project
Presented by Ednalyn C. De Dios, aka "Dd"
FINDINGS
ETC.
ANSWERS.
QUESTIONS
???
DATA
DICTIONARY
https://bit.ly/ZingDd


Initial
Thoughts and Questions
1. Does location affect log error?
2. To what extent does size matter?
3. How much do the number of bedrooms and bathrooms play a role in driving log error?
4. Depending on geography, do heating and air conditioning play a role?
5. What other variables drive the log error?
Location
Size
Rooms
HVAC

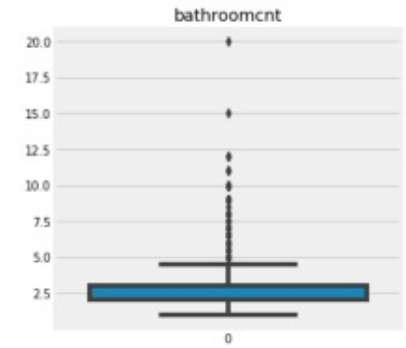
Zillow
Zingzestimate
Project
Presented by Ednalyn C. De Dios, aka "Dd"
FINDINGS
ETC.
ANSWERS.
QUESTIONS
???
DATA
DICTIONARY
https://bit.ly/ZingDd

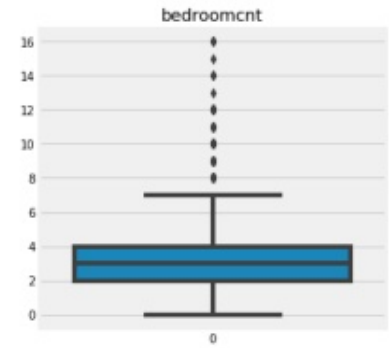
DATA DICTIONARY
We will be using 2016 data.
logerror: log function of the error between Zestimate and sale price of a home.
bathroomcnt: Number of bathrooms in home including fractional bathrooms.
bedroomcnt: Number of bedrooms in home.
calculatedfinishedsquarefeet: Calculated total finished living area of the home.
fullbathcnt: Number of full bathrooms (sink, shower + bathtub, and toilet) in home.
garagecarcnt: Total number of garages on the lot including an attached garage.
roomcnt: Total number of rooms in the principal residence.
yearbuilt: The Year the principal residence was built.
taxvalueddollarcnt: The total tax assessed value of the parcel.
taxamount: The total property tax assessed for that assessment year.

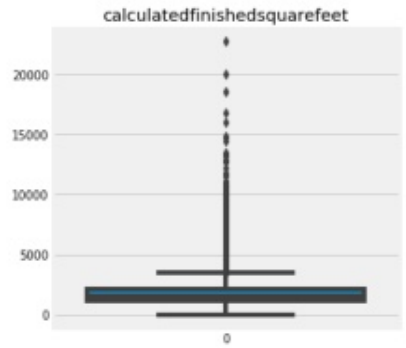
Zillow
Zingzestimate
Project
Presented by Ednalyn C. De Dios, aka "Dd"
FINDINGS
ETC.
ANSWERS.
QUESTIONS
???
DATA
DICTIONARY
https://bit.ly/ZingDd

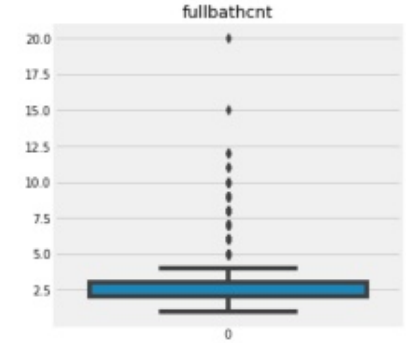

FINDINGS ET CETERA
A few outliers had to be removed.
Correlation coefficients similar with drivers of log error.
Residual plot indicates need for more investigation.
OUTLIERS
CORRELATION COEFFICIENTS
DRIVERS
RESIDUALS

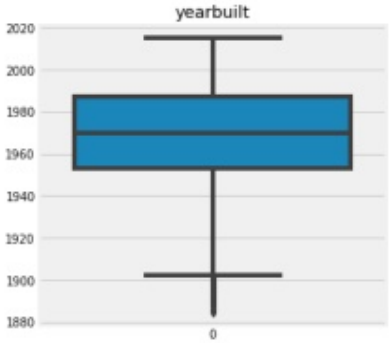
WITH OUTLIERS

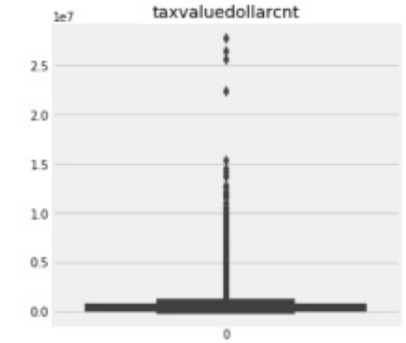


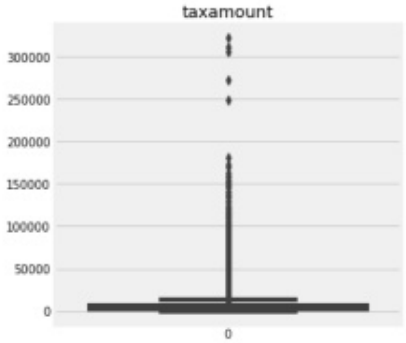




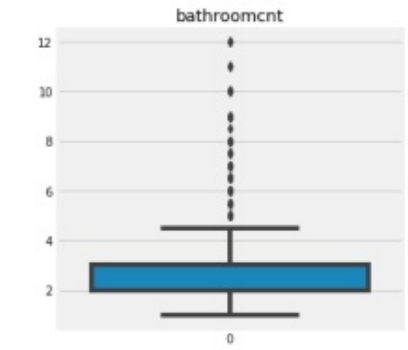


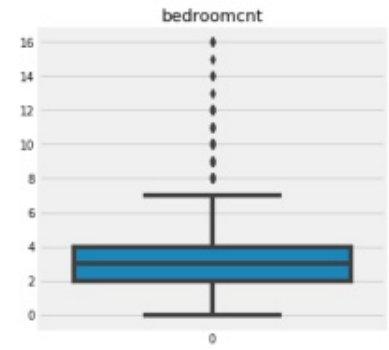


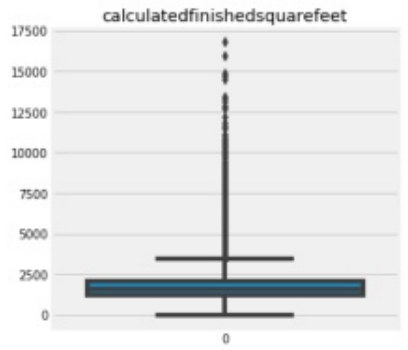


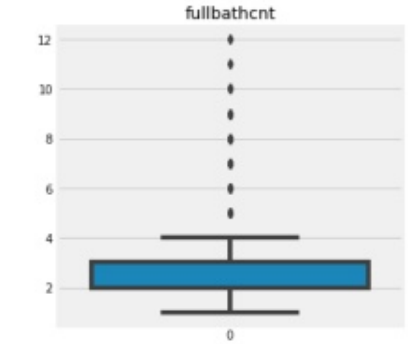


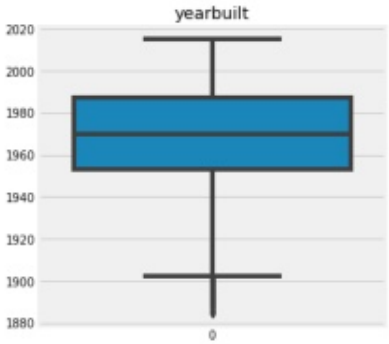
WITHOUT OUTLIERS

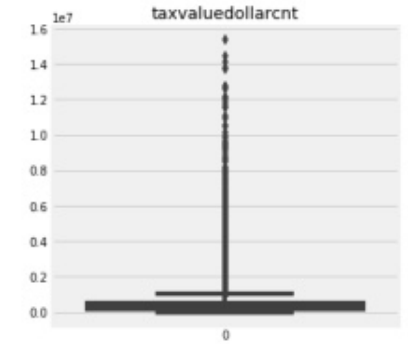


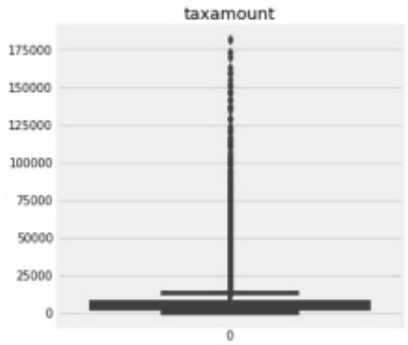












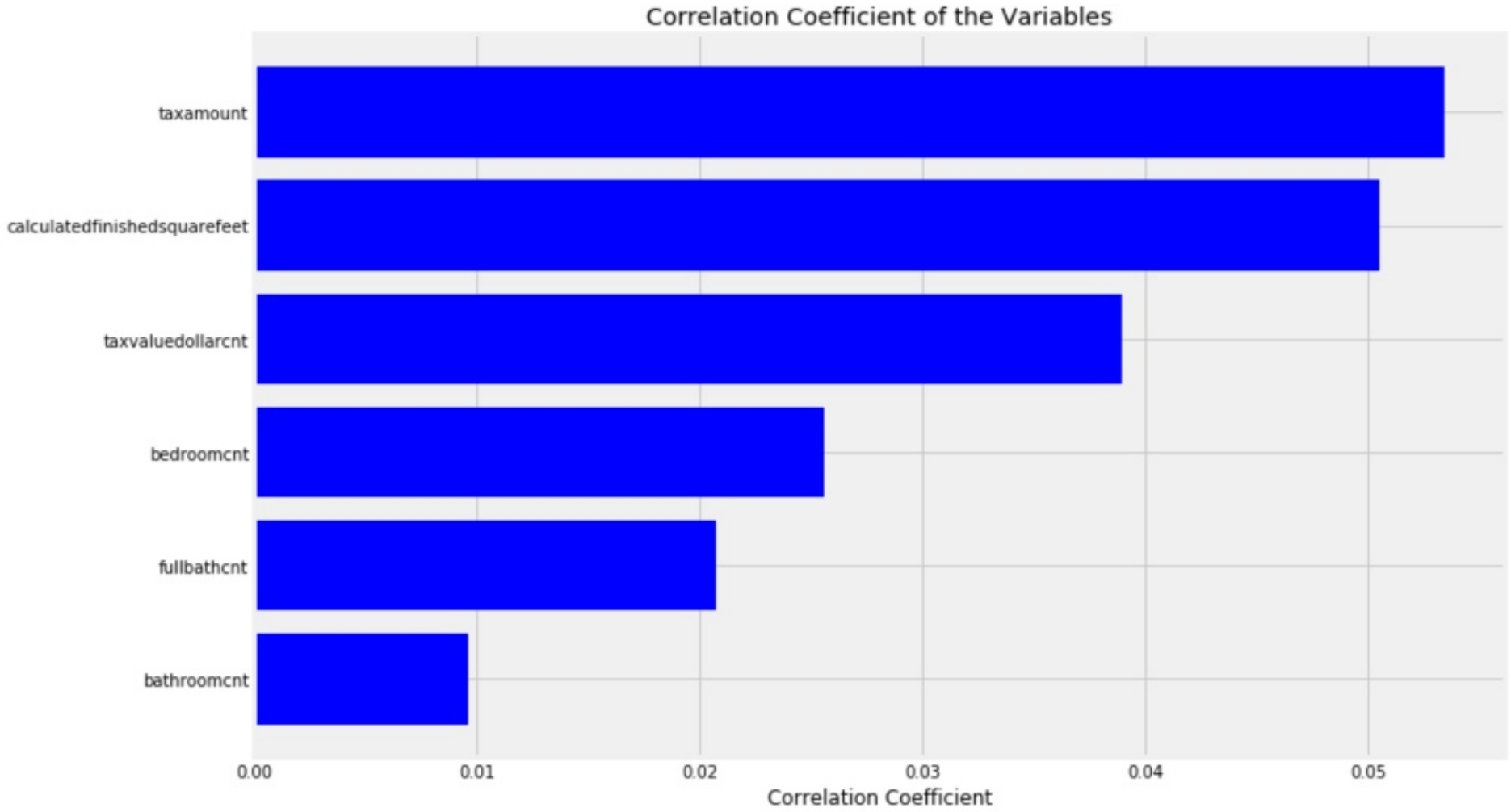
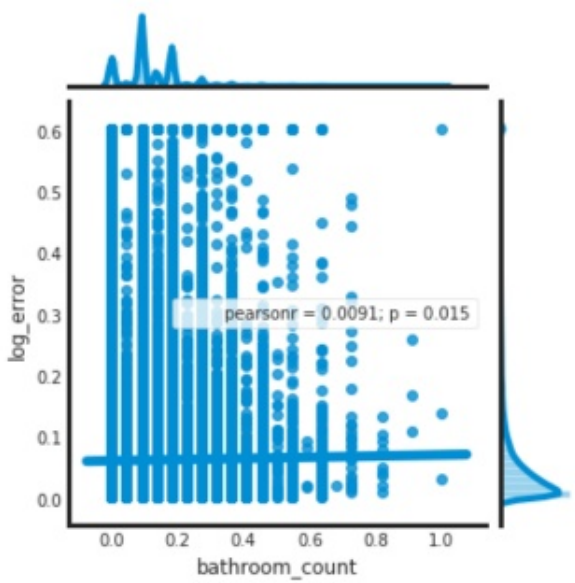
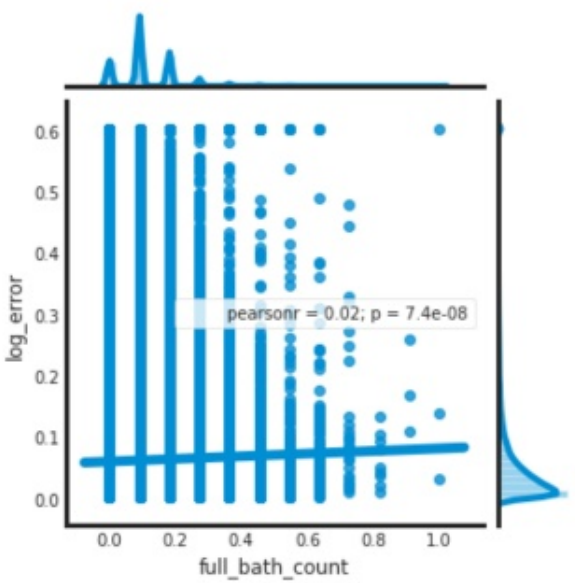

Correlation Coefficient of the Variables
taxamount
calculatedfinishedsquarefeet
taxvaluedollarcnt
bedroomcnt
fullbathcnt
bathroomcnt
0.00
0.01
0.02
0.03
0.04
0.05
Correlation Coefficient

EXHIBIT A


log_error
bathroom_count
pearsonr = 0.0091; p = 0.015


log_error
full_bath_count
pearsonr = 0.02; p = 7.4e-08

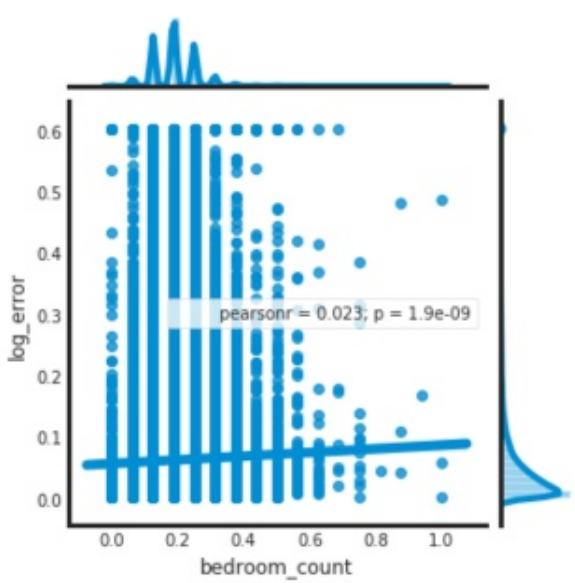
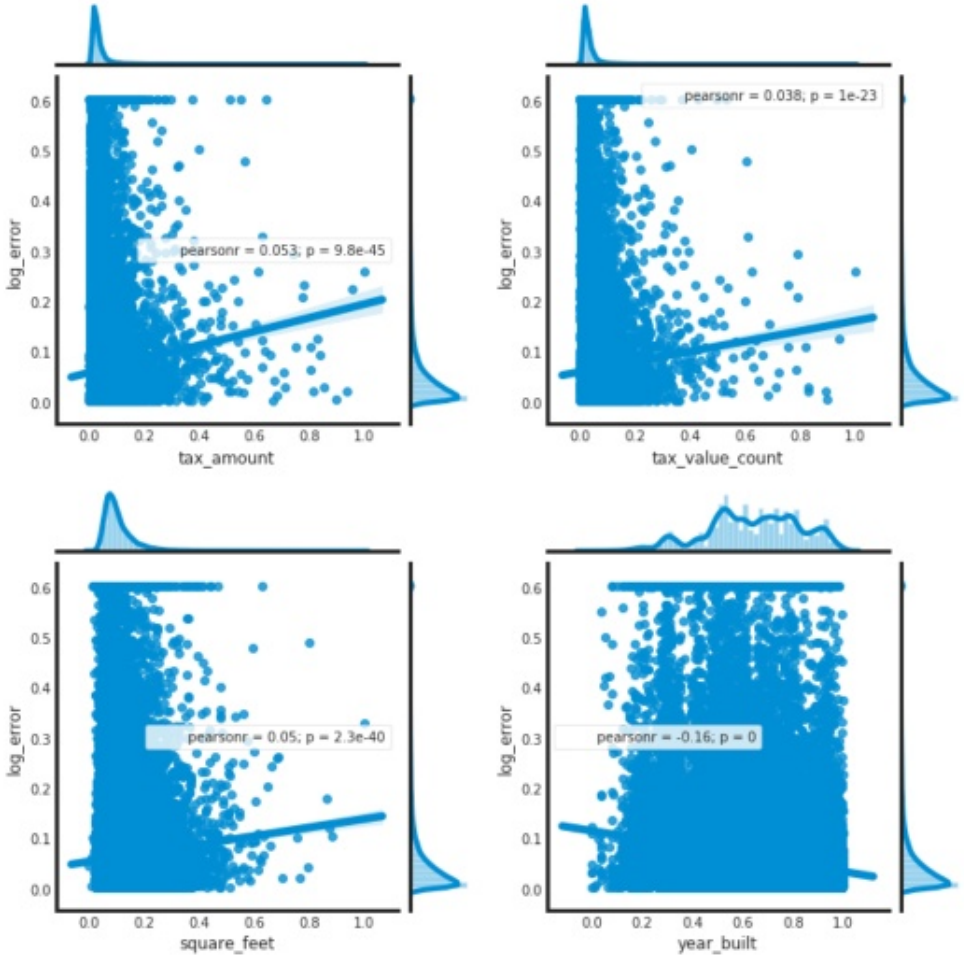
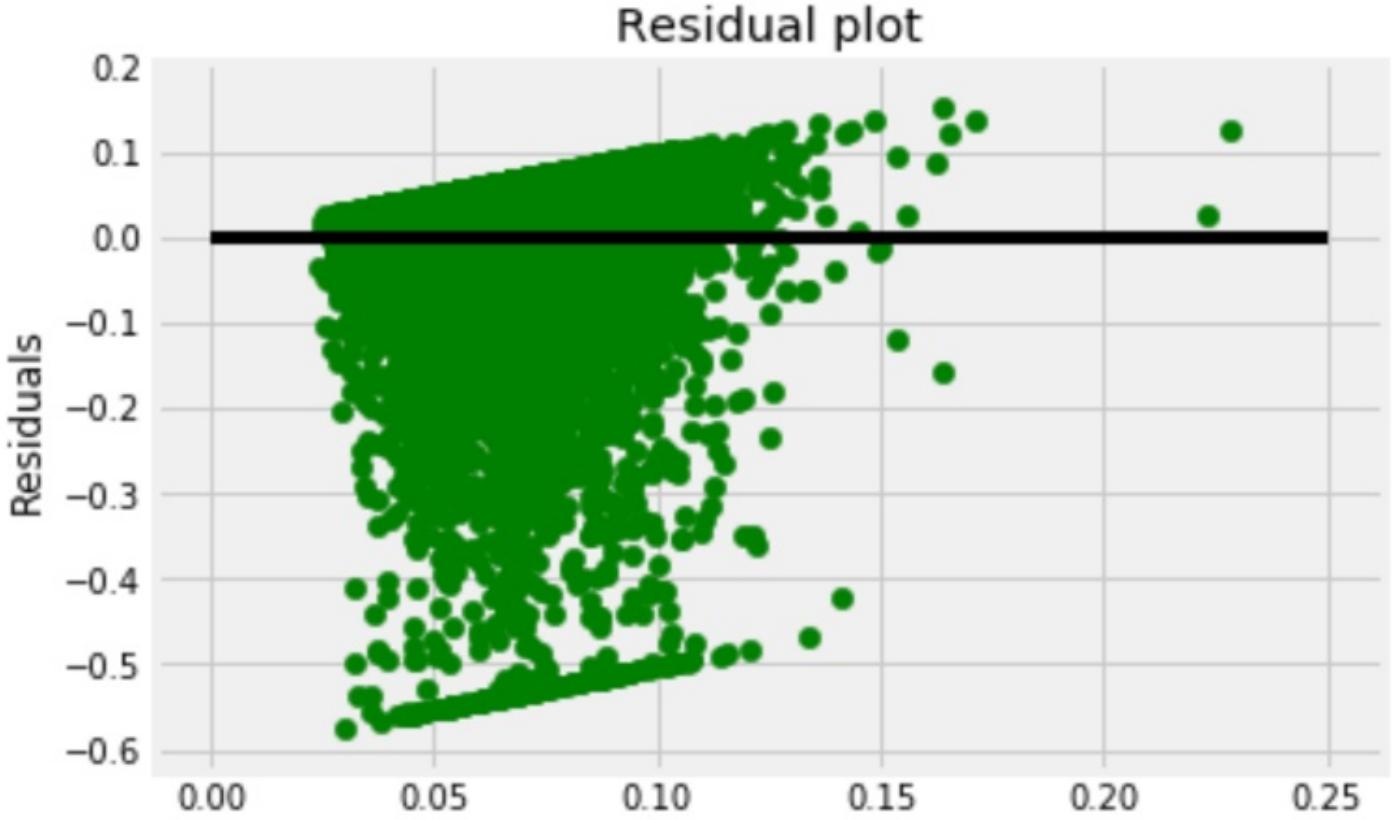

log_error
bedroom_count
pearsonr = 0.023; p = 1.9e-09

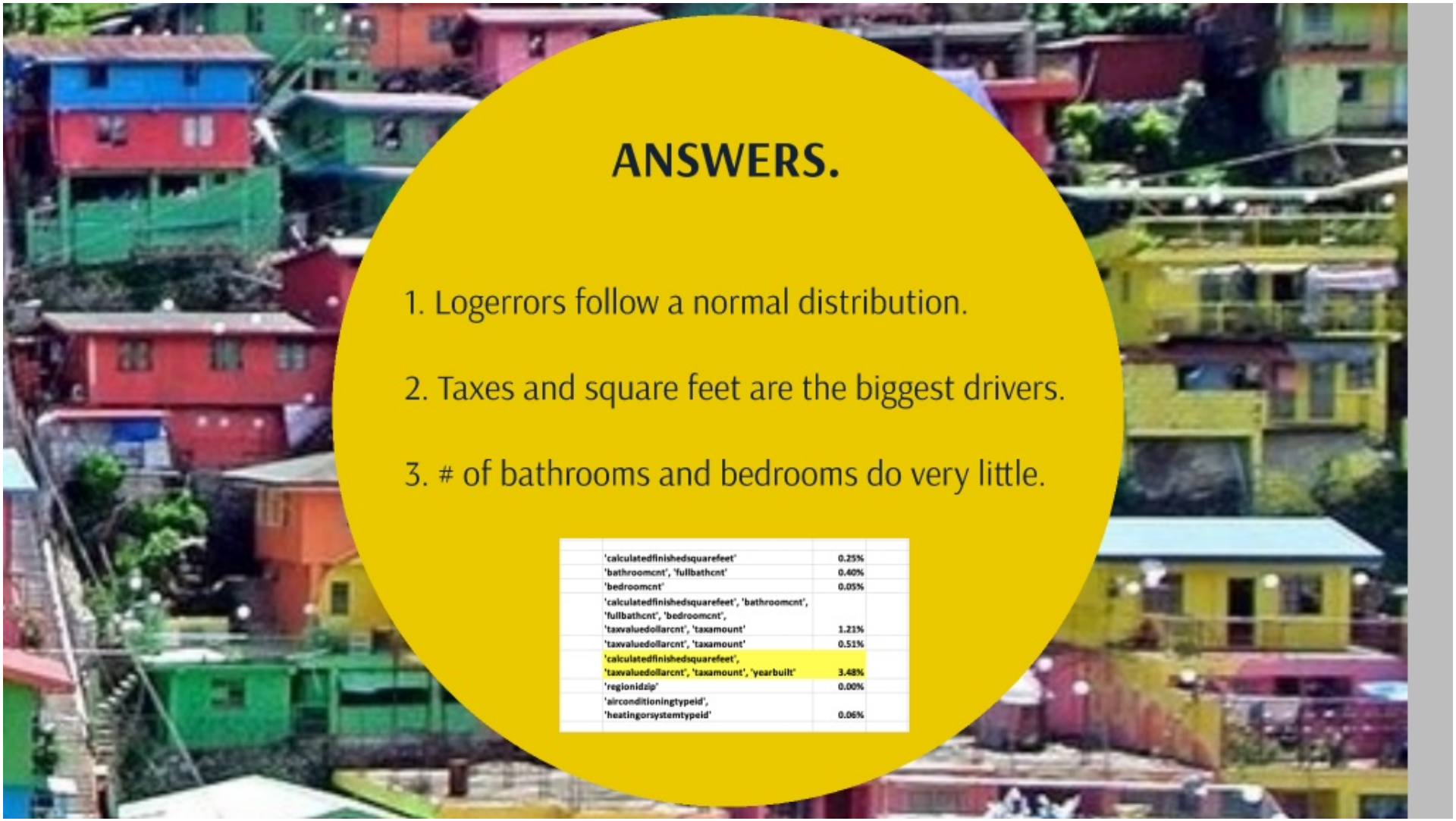
EXHIBIT B


log_error
tax_amount
pearsonr = 0.053; p = 9.8e-45
log_error
tax_value_count
pearsonr = 0.038; p = 1e-23
log_error
square_feet
pearsonr = 0.05; p = 2.3e-40
log_error
year_built
pearsonr = -0.16; p = 0

RESIDUALS


Residual plot
Residuals
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
−0.5
−0.6
0.00
0.05
0.10
0.15
0.20
0.25

Zillow
Zingzestimate
Project
Presented by Ednalyn C. De Dios, aka "Dd"
FINDINGS
ETC.
ANSWERS.
QUESTIONS
???
DATA
DICTIONARY
https://bit.ly/ZingDd


ANSWERS.
1. Logerrors follow a normal distribution.
2. Taxes and square feet are the biggest drivers.
3. # of bathrooms and bedrooms do very little.
'calculatedfinishedsquarefeet' 0.25%
'bathroomcnt', 'fullbathcnt' 0.40%
'bedroomcnt' 0.05%
'calculatedfinishedsquarefeet', 'bathroomcnt', 'fullbathcnt', 'bedroomcnt', 'taxvaluedollarcnt', 'taxamount' 1.21%
'taxvaluedollarcnt', 'taxamount' 0.51%
'calculatedfinishedsquarefeet', 'taxvaluedollarcnt', 'taxamount', 'yearbuilt' 3.48%
'regionidzip' 0.00%
'airconditioningtypeid', 'heatingorsystemtypeid' 0.06%

Zillow
Zingzestimate
Project
Presented by Ednalyn C. De Dios, aka "Dd"
FINDINGS
ETC.
ANSWERS.
QUESTIONS
???
DATA
DICTIONARY
https://bit.ly/ZingDd