**INTRO**

HI MY NAME IS ALLANTE AND MY PARTNER IS ZACH

WE ARE GOING TO DISCUSS THE DRIVERS WE IDENTIFIED THAT CONTRIBUTE

TO THE LOGERROR OF SINGLE FAMILY HOMES IN THE 2017 ZILLOW DATASET

ACQUIRE

THIS DATA WAS ACQUIRED FROM THE CODEUP DATABASE USING MYSQL WORKBENCH

PREP

TO PREPARE THE DATA WE REMOVED ROWS WITH NULLS WHILE MAINTAINING AT LEAST 75% OF OUR ROWS

THEN WE REMOVED COLUMNS WITH A NON-NULL COUNT UNDER APPROXIMATELY 51K

WE DID NOT INTENTIONALLY REMOVE OUTLIERS

WE ALSO IMPUTED THE AGE OF THE HOMES BY SUBTRACTING CURRENT DATE FROM THE YEAR BUILT

**MEDIAN OF LOGERROR**

THE FIRST QUESTION WE ASKED IS

WHAT IS THE MEDIAN AND MEAN LOGERROR OF THE POPULATION TO FIND THE CENTER OF THE LOGERROR

STATISTICAL QUESTIONS 1

NEXT WE WONDERED HOW THE MEAN OF EACH COUNTIES LOGERROR COMPARES TO EACH OTHER

YOU CAN SEE THE MEAN IS HIGHER IN ORANGE COUNTY COMPARED TO LOS ANGELES AND VENTURA COUNTY.

THE STATISTICAL TEST SHOWS THERE IS A SIGNIFICANT DIFFERENCE IN THE MEAN OF THE LOGERRORS ACROSS COUNTIES

**STATISTICAL QUESTION 2**

NEXT WE ASKED IF THERE IS A CORRELATION BETWEEN THE TAX AMOUNT AND THE LOGERROR.

HERE WE CAN SEE AS THE TAX AMOUNT INCREASES THE LOGERROR APPEARS TO MOVE CLOSER TO 0.

THE RESULTS OF THE STATISTICAL TEST SUPPORT A CORRELATION BETWEEN THESE TWO VARIABLES

**MENTIONING CLUSTERS**

DURING THIS PROJECT WE CREATED THREE COMBINATIONS OF FEATURES TO IDENTIFY CLUSTERS.

THE FIRST WAS CREATED USING THE FEATURES OF LATITUDE, LONGITITUDE AND HOME AGE

THE SECOND CLUSTER WAS CREATED USING BEDROOMS AND BATHROOMS

THE LAST CLUSTER WAS CREATED USING SQUARE FEET AND LOT SIZE SQUARE FEET

NOW I WILL HAND IT OVER TO ZACH TO DISCUSS THE QUESTIONS ASKED ABOUT OUR CLUSTERS.

**Zach**

show /mention 3 type of clusters created before talking about stat questions

Statistical question 3

Statistical question 4

Exploration summary

Models

* Mention variables were scaled)

Model summary

* Update README after presentation (Allante)
* Upload latest Wrangle, Explore, Model py files

**Allante**

Intro

Acquire data

Prep data

Median of logerror

Statistical Questions 1

Statistical question 2

Mention 3 combinations of clusters created and two selected for statistical testing

**Zach**

Statistical question 3

Statistical question 4

Exploration summary

Models

* Mention variables were scaled

Model summary

Recommendation

Next steps

**Remaining tasks**

* Make sure all cells have comment code / make sure comment code in py files
* Upload latest Wrangle, Explore, Model py files
* Create video
* Update README after presentation
* Submit link to repo for project to be submitted to be graded

**Workarounds for git hub issue**

* Is Zach able to curl -O “file link” into his git repo?
  + If he is able to do this is he able to still get push after doing so?
    - Maybe this avoids getting the same error (just a thought I had)