

Rent in San Francisco

TidyTuesday Challenge (7/5/22)

Alan Wang

INSERT DATE

Set Up

```
# load packages
library(tidyverse)
library(tidymodels)
library(knitr)
library(rms)
library(patchwork)
library(dplyr)
library(ggplot2)
library(ggmap)

# load data
rent <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidyuesday/master/data/rent/rent.csv')
```

Overview of Data

This data were collected by Dr. Kate Pennington, who scraped Craigslist posts for rents of properties in San Francisco from 2000 to 2018. The dataset contains 200,796 observations, each of which represents an individual unit, identified by a unique ID, Craigslist posting date, neighborhood, city, county, monthly rent price, number of beds/baths, size in terms of square feet, whether there is room in the apartment, the address, and a description.

Data Preparation

The following analysis will primarily focus on the rent price, number of beds, number of baths, size, year, and whether there is room in the apartment. Thus, before any visualizations

are made or any analysis is conducted, any observations that do not have values for the aforementioned variables will be dropped. The `room_in_apt` variable will also be made into a categorical variable. This new dataset has 14,629 observations.

```
rents <- rent |>
  drop_na(year | nhood | price | beds | baths | sqft | room_in_apt) |>
  mutate(room_in_apt = as.factor(room_in_apt))
```

Linear Regression

This is a linear regression model that will be used to predict monthly rent price for rental units in San Francisco with year posted on Craigslist, number of beds, number of paths, number of square feet, and whether or not there is room in the apartment.

First, w

```
set.seed(1)

rents_split <- initial_split(data = rents)
rents_train <- training(rents_split)
rents_test <- testing(rents_split)

rents_spec <- linear_reg() |>
  set_engine("lm")

rents_rec <- recipe(price ~ year + beds + baths + sqft + room_in_apt, data = rents) |>
  step_center(all_numeric_predictors()) |>
  step_dummy(all_nominal_predictors()) |>
  step_zv(all_predictors())

rents_wflow <- workflow() |>
  add_model(rents_spec) |>
  add_recipe(rents_rec)

rents_fit <- rents_wflow |>
  fit(data = rents_train)

rents_fit |>
  tidy() |>
```

```
kable(digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	2804.232	12.623	222.158	0.000
year	189.227	5.026	37.650	0.000
beds	54.187	18.483	2.932	0.003
baths	474.582	24.595	19.296	0.000
sqft	0.765	0.025	30.318	0.000
room_in_apartment_X1	-390.499	241.506	-1.617	0.106

**Insert interpretations for model

Visualizations

To get the visualizations, the following command `install.packages(c("leaflet", "sp"))` was run

```
rent_maps <- rent |>  
  drop_na(lat | lon) |>  
  filter(lon >= -180 || lon <= 180)
```