


# Project 3: Subreddit NLP Classification

Brandon Griffin  
10/23/2020

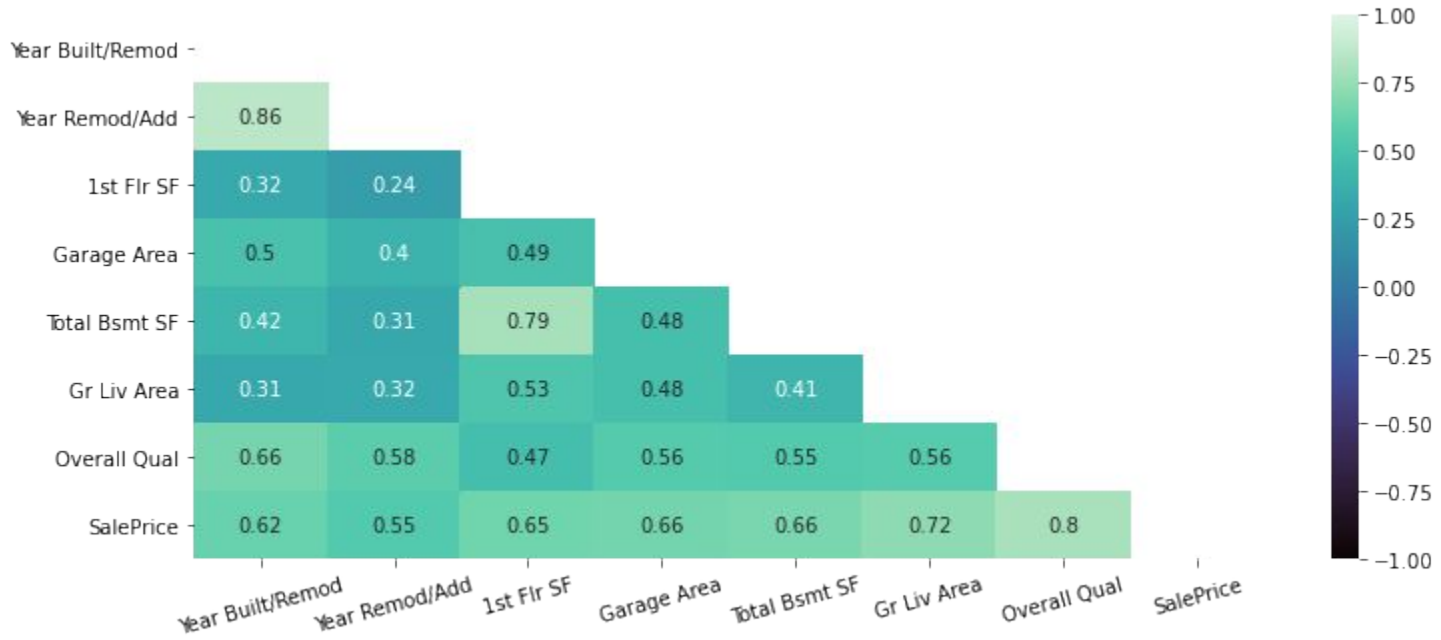


# Introduction

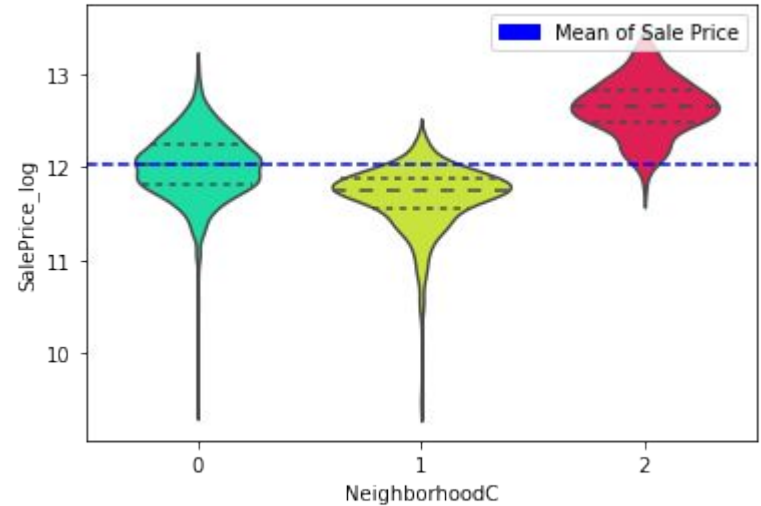
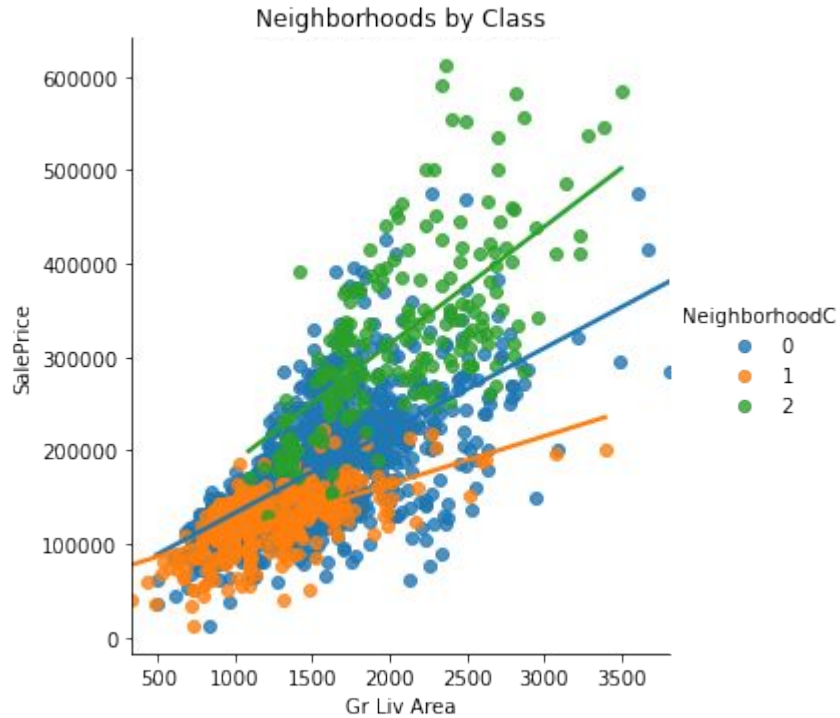
**Problem Statement:** An investigation of the association of suicide ideation and depression, in particular the use of Natural Language Processing to identify suicidality in Redditors seeking support online.

**Target Audience:** A tech savvy daughter of a real estate tycoon has recently inherited her late mother's assets. She hired me to run an OLS regression analysis to help her maximize profit while selling her homes in Iowa as she heads to Florida for early retirement.

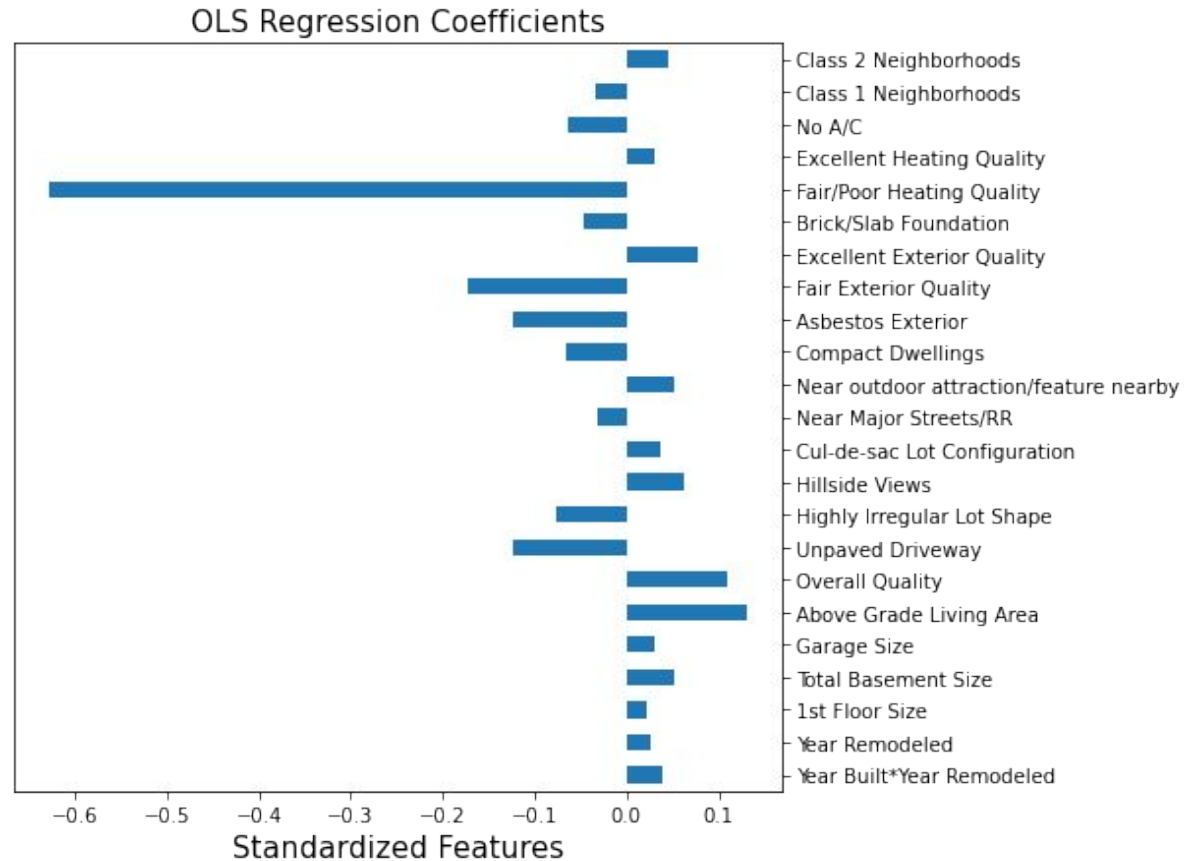
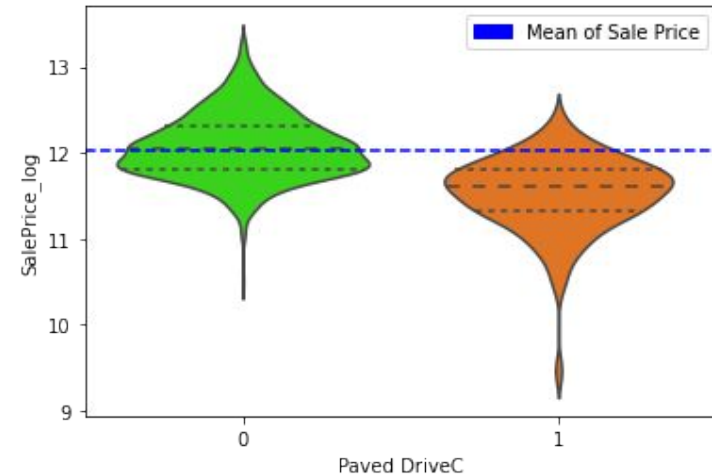
# Methodology/Procedure



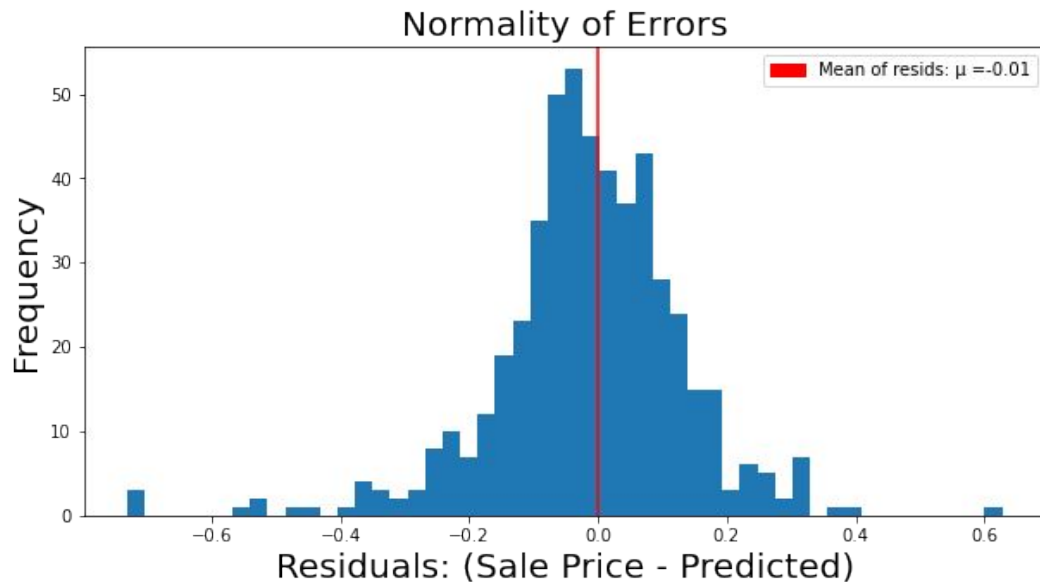
# Data Considered



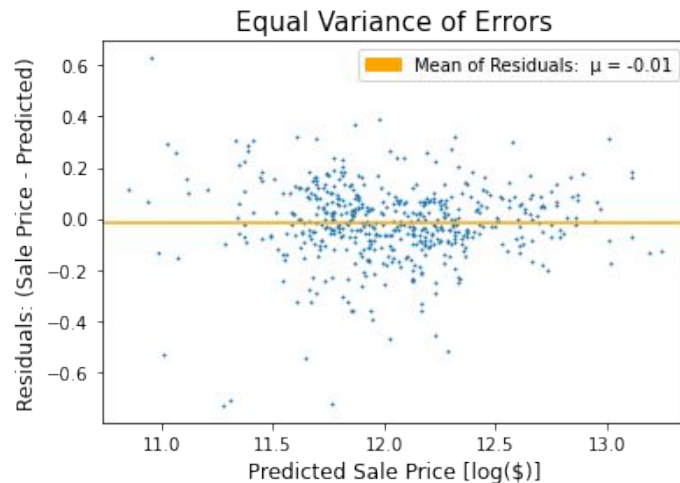
# Primary Findings



# Reliability of Model



Metric	Score
RMSE	0.15 (<30,174)
R <sup>2</sup>	0.876108
Baseline RMSE	83,323



# Recommendations to client

1. Quality over quantity

- Exterior quality: Make necessary repairs!

2. Above ground living area is important

3. Central A/C and quality of heating are surprisingly important

4. Paved driveways have value

5. No asbestos

# Sources

Fernandes, A.C., Dutta, R., Velupillai, S. *et al.* Identifying Suicide Ideation and Suicidal Attempts in a Psychiatric Clinical Research Database using Natural Language Processing. *Sci Rep* 8, 7426 (2018).

<https://doi.org/10.1038/s41598-018-25773-2>