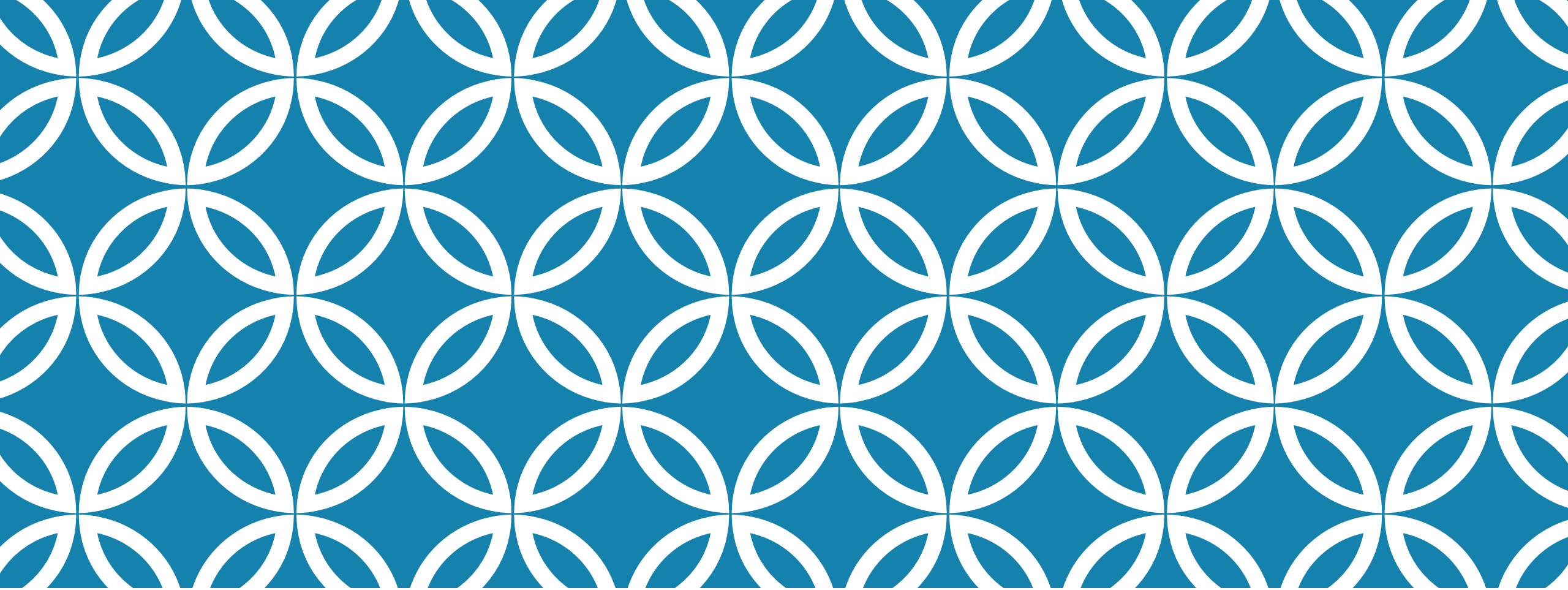


PREDICTING COVID-19 BURDEN IN NYC NEIGHBORHOOD

Chengda Zhang, MD

INTRODUCTION

The New York City (NYC) has faced an outbreak of COVID-19 in the past 2 months. According to the data presented by NYC Health Department, different modified zip code areas (MODZCTA) are affected differently by COVID-19. Distribution of venues in NYC may directly affect the outbreak or indirectly associated with COVID-19 through close connection with local social economic situation.

PROBLEM

More US cities are facing outbreak of COVID-19. Modeling case burden of COVID-19 in US cities will help us allocate proper resources and polices to certain areas based on available location data.

I aim to use location data from Foursquare to model the case burden of COVID-19 in different MODZCTAs of NYC.

DATA SOURCE

NYC Health Department: COVID-19 case burden by MODZCTA and geographic boundary of each MODZCTA

NYC venues category information is queried from Foursquare

DATA CLEANING

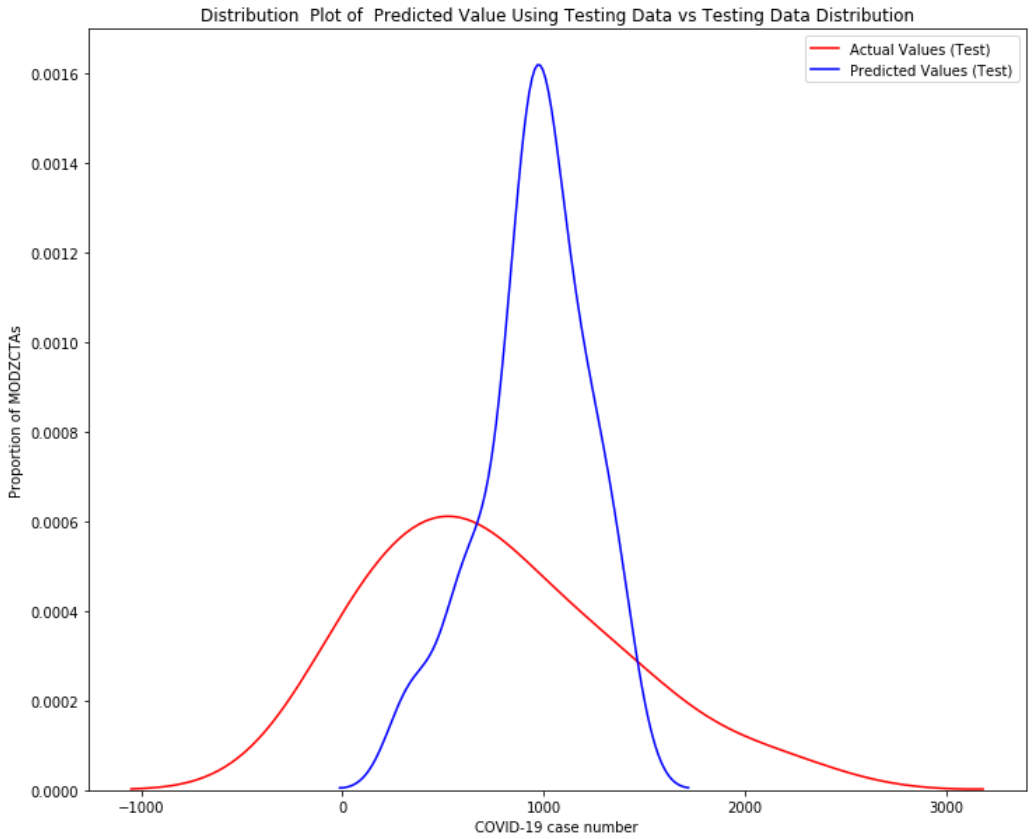
Discarded all venues that are not physically located in an MODZCTA.

EXPLORATORY ANALYSIS

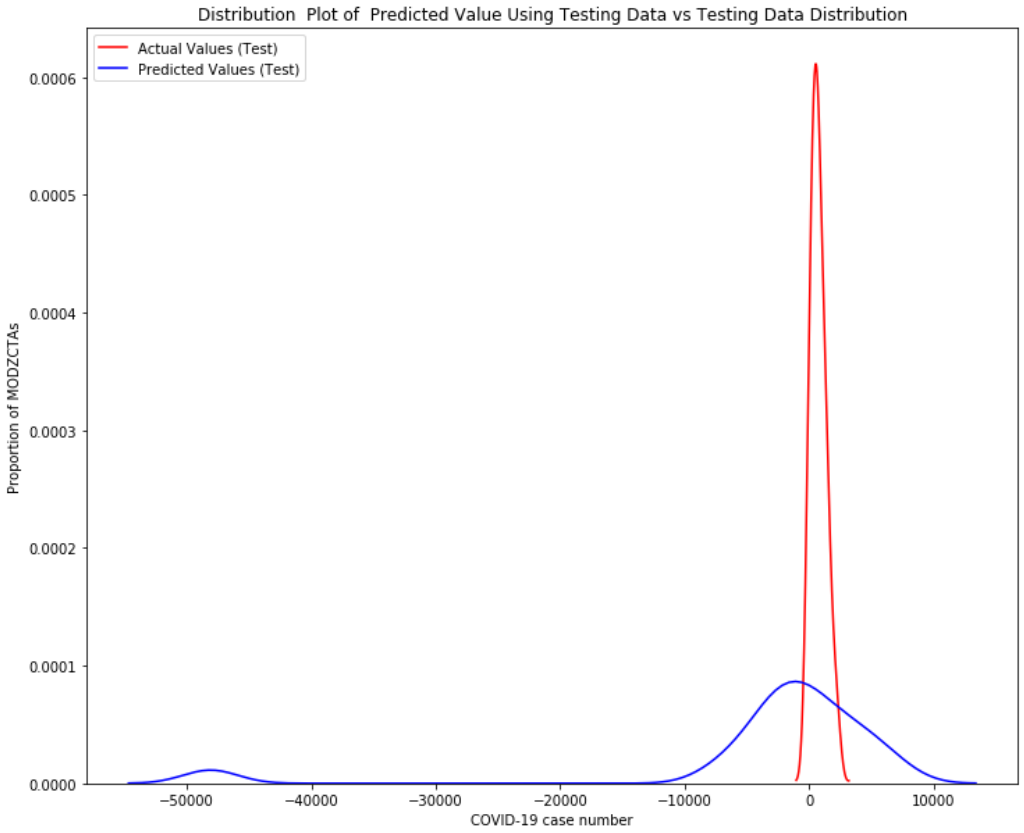
When too many subcategories being used, the model tends to overfit.

In order to avoid over-fitting, we combined sub-categories of venues into 9 large groups. This combining is based on category tree from Foursquare. The 9 large groups that are included are: 'Food', 'College & University', 'Event', 'Arts & Entertainment', 'Nightlife Spot', 'Outdoors & Recreation', 'Professional & Other Places', 'Shop & Service', 'Travel & Transport', 'Residence'.

RESULTS, LINEAR REGRESSION



RESULTS, POLYNOMIAL REGRESSION



CONCLUSION

The distribution of venues may partly explain the COVID-19 case distribution, however is not an ideal predictor for COVID-19 case burdens.