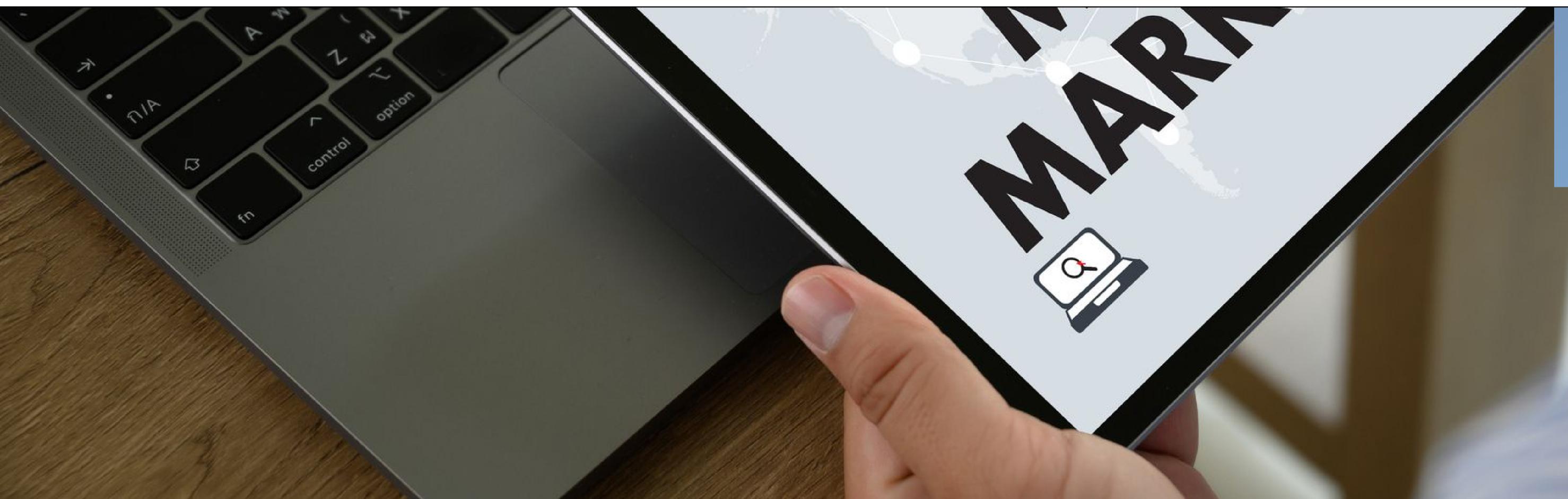


July, 14 2022

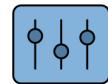
WORD  
MOUTH  
MARKETING

# Best business configuration for the restaurant with the effect of eWOM on check-in intention: Yelp data study



Victoria Sibatorkina, 7738742  
Zurya Jeronimo Valencia, 7737125  
Gil Zalcberg, 6857556

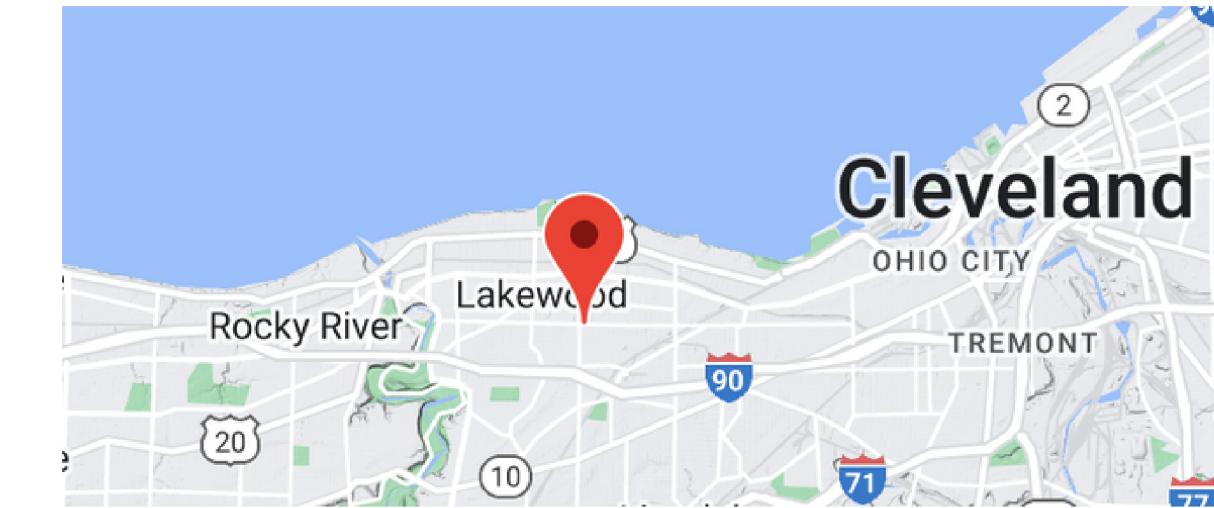
# Introduction



## Dataset

The sample concentrates in:

- Lakewood, OH, USA
- 10 months period January - October 2018



The initial dataset totals 7,600 observations with 29 variables:  
different and unique Restaurants with different cuisine, price range, and business attributes



## Problem

Finding the optimal business model configuration for opening a new restaurant in Lakewood, OH with eWOM considered.

- Identifying important variables for check-in behavior
- Finding the best predictive machine learning model to find the best business configuration

# Variables

<b>Business attributes</b>					
Statistic	N	Mean	St. Dev.	Min	Max
Check-in	7,600	0.095	0.293	0	1
Reviews	7,600	233.480	201.305	22	870
WiFi	7,600	0.440	0.496	0	1
Credit card	7,600	1.000	0.000	1	1
Parking	7,600	0.960	0.196	0	1
Price level	7,600	1.800	0.490	1	3
Functioning	7,600	0.840	0.367	0	1
American	7,600	0.480	0.500	0	1
Mexican	7,600	0.120	0.325	0	1
Others	7,600	0.400	0.490	0	1

Statistic	N	Mean	St. Dev.	Min	Max
ch_in	7,600	0.095	0.293	0	1
cum_n_tips	7,600	4.058	2.921	1	13
cum_max_friends	7,600	1,034.960	1,751.483	2	6,506
cum_max_u_elite	7,600	6.320	2.054	3	9
cum_max_us_fans	7,600	171.880	305.434	0	1,150
cum_max_us_tip	7,600	788.600	540.878	10	1,523
male	7,600	1.200	0.800	0	3
female	7,600	1.615	1.395	0	6

Statistic	N	Mean	St. Dev.	Min	Max
ch_in	7,600	0.095	0.293	0	1
n_photo	7,600	6.360	9.381	0	38
business_stars	7,600	3.960	0.467	2.500	4.500

# Variables

**External Twitter data**

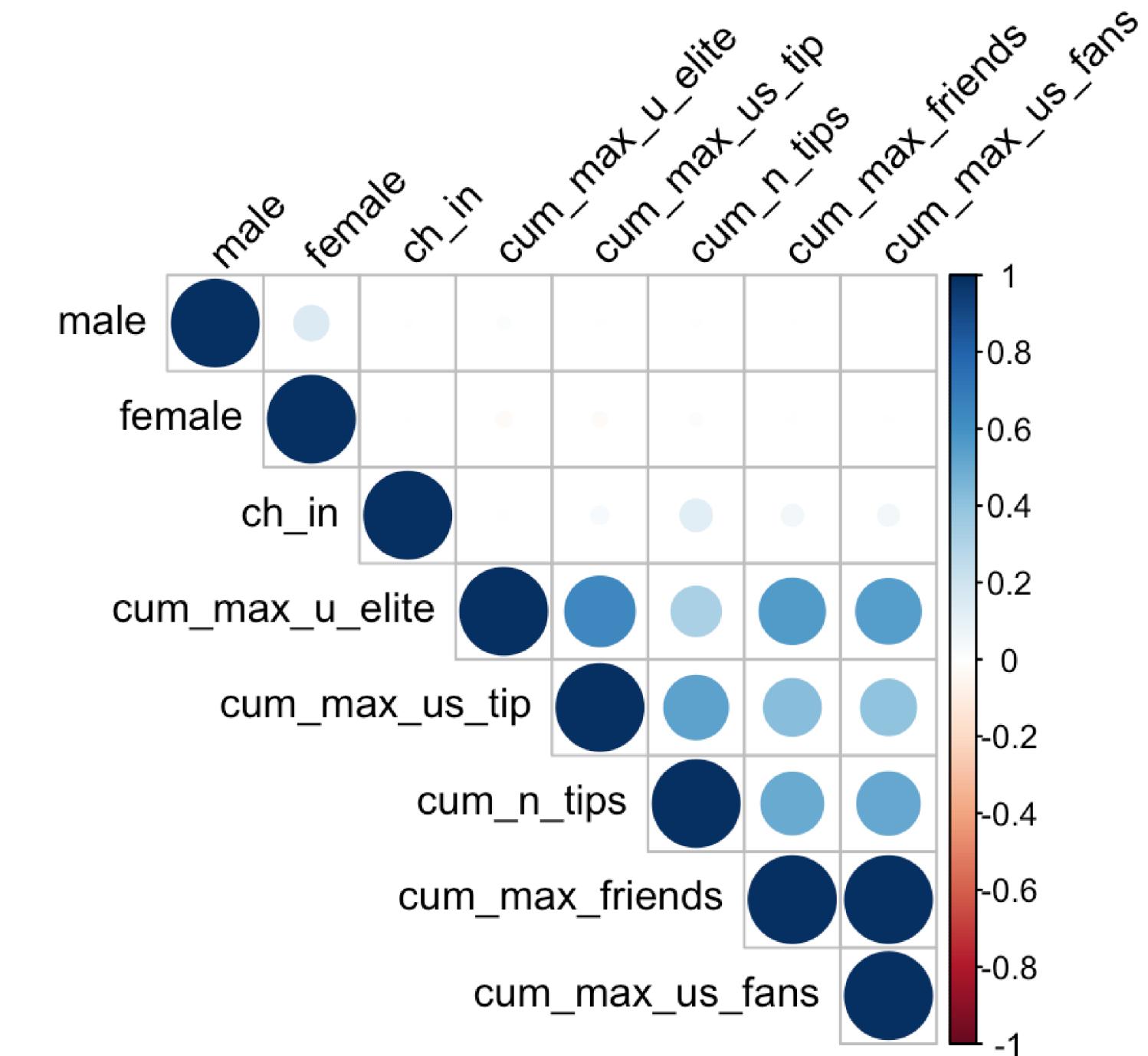
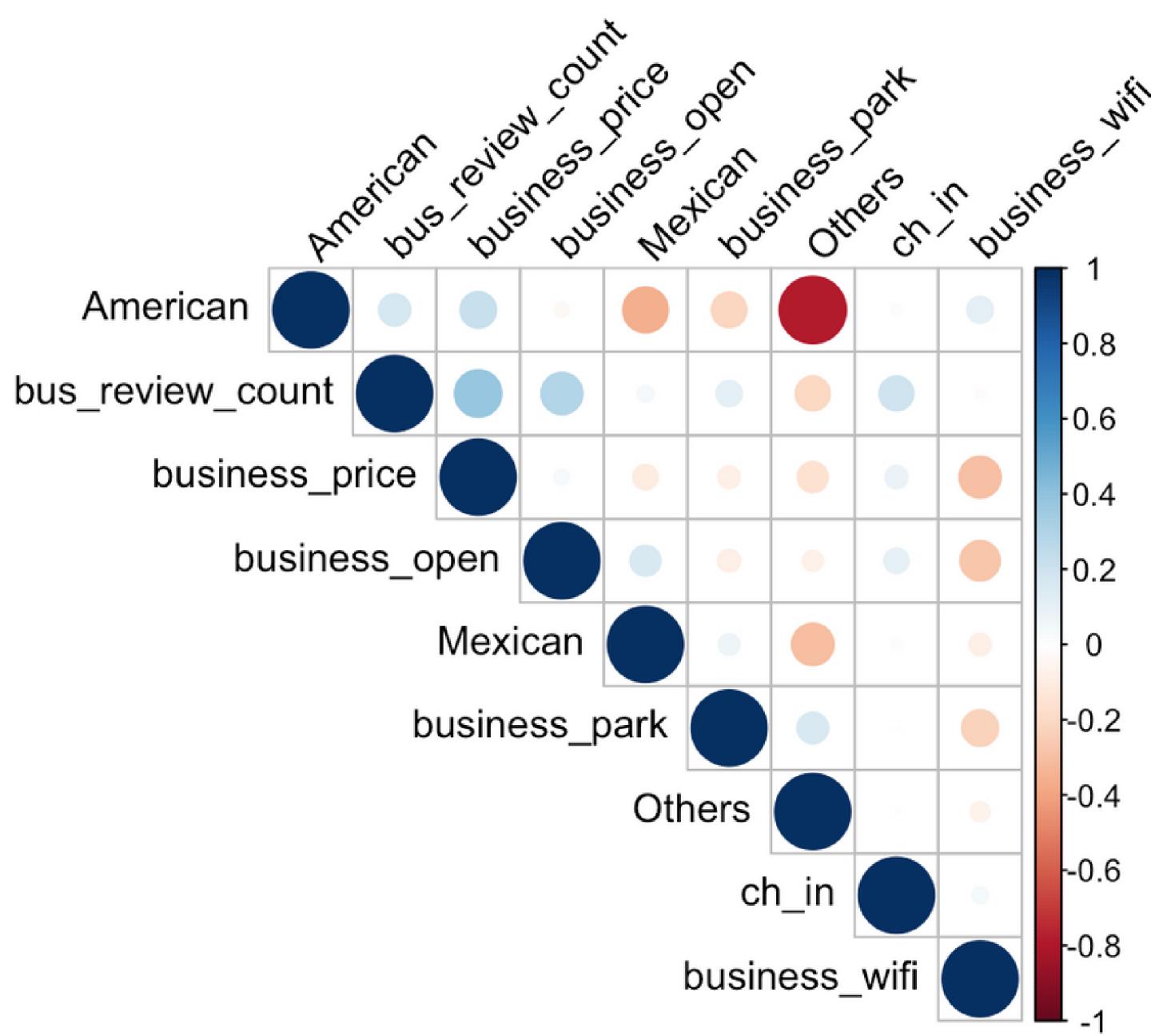
Statistic	N	Mean	St. Dev.	Min	Max
ch_in	7,296	0.095	0.293	0	1
Twitts_num	7,296	854.625	3,126.549	1	15,753
cum_followers	7,296	5,895,644.000	21,744,184.000	266	109,152,271
Positive_words_twitt	7,296	251.250	880.584	0	4,421
positive_words_review	7,296	8.958	6.586	1	22

**External weather data**

Statistic	N	Mean	St. Dev.	Min	Max
ch_in	6,080	0.095	0.293	0	1
PRCP	6,080	26.039	68.362	0	602
TMAX	6,080	165.980	114.010	-121	350
TMIN	6,080	93.332	107.587	-160	261
AWND	6,080	48.577	19.554	13	128
PGTM	6,080	1,215.687	693.538	5	2,359
WDF2	6,080	187.385	109.457	10	360
WDF5	6,080	186.543	106.039	10	360
WSF2	6,080	94.822	30.638	36	206
WSF5	6,080	122.452	39.164	45	246

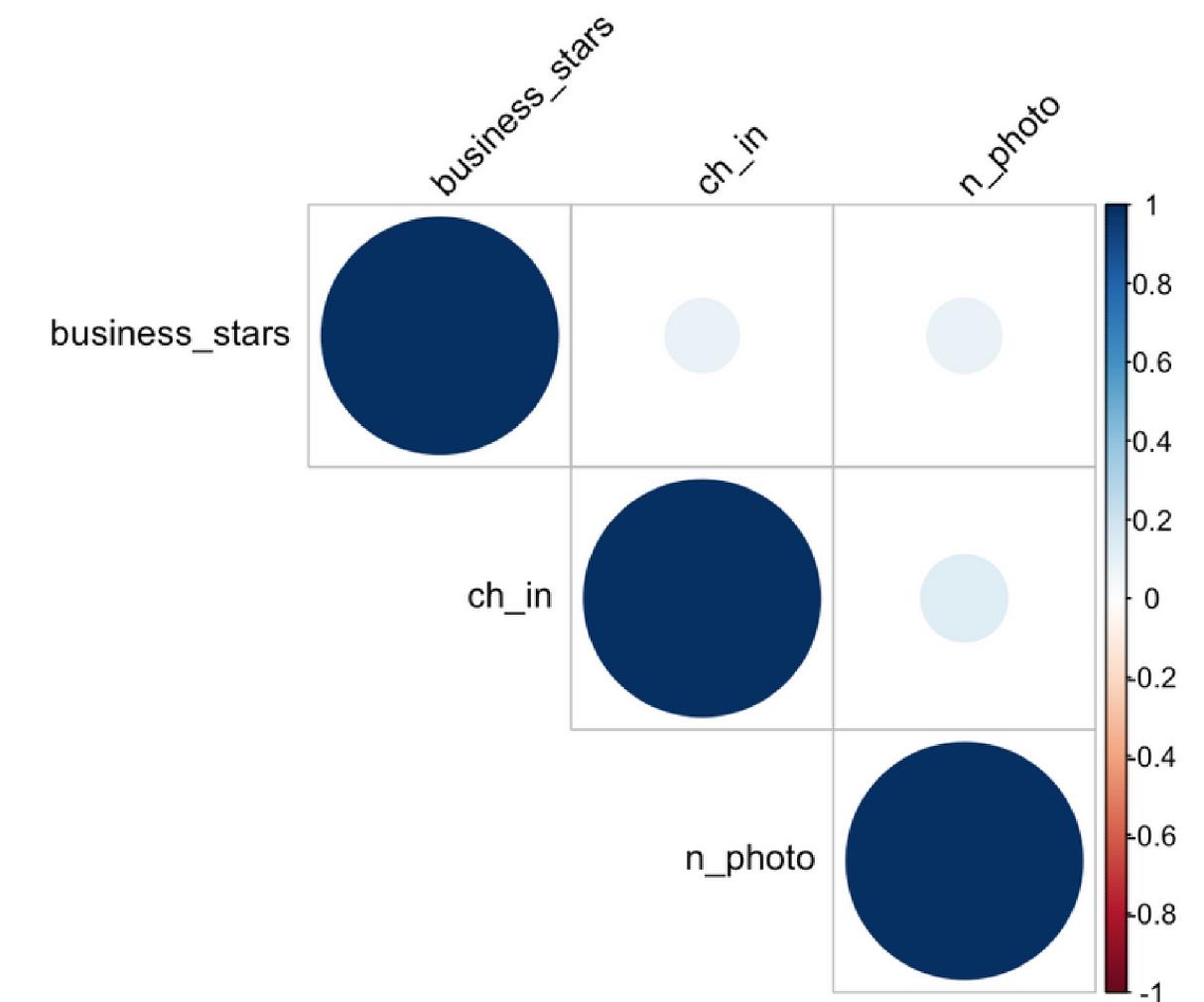
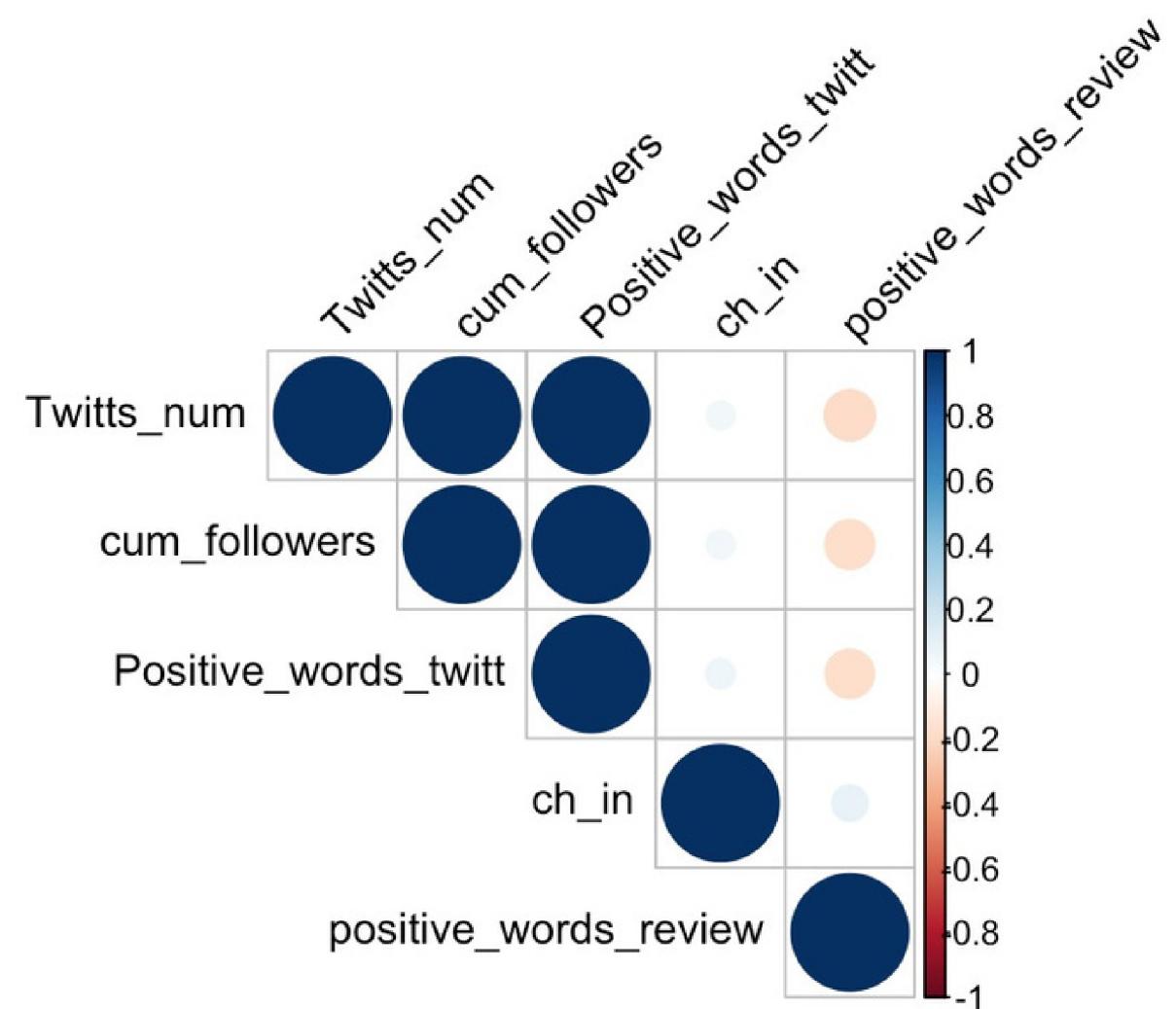
# Correlations

Correlation plots of business attributes (left) and user characteristics variables subset (right)



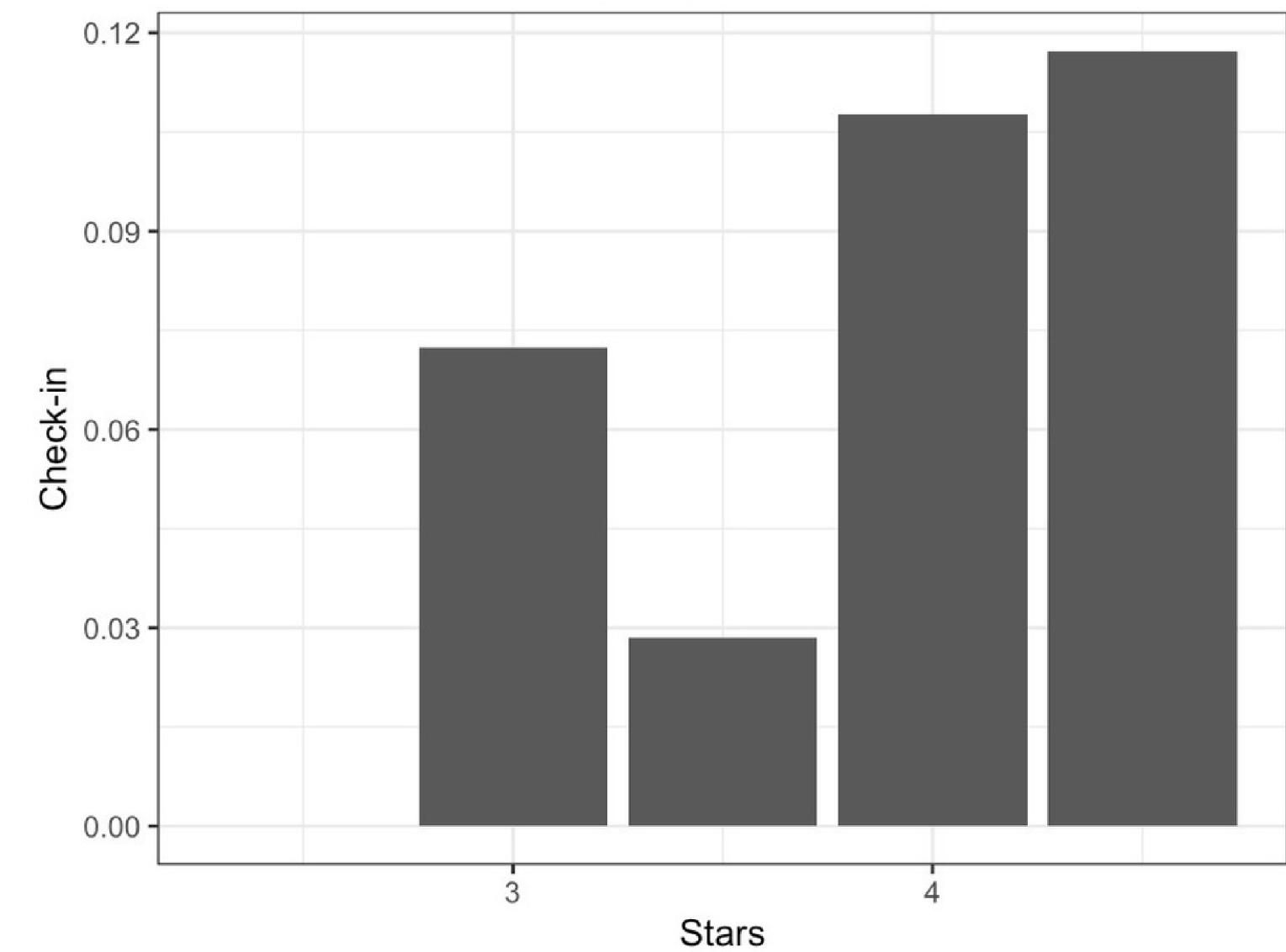
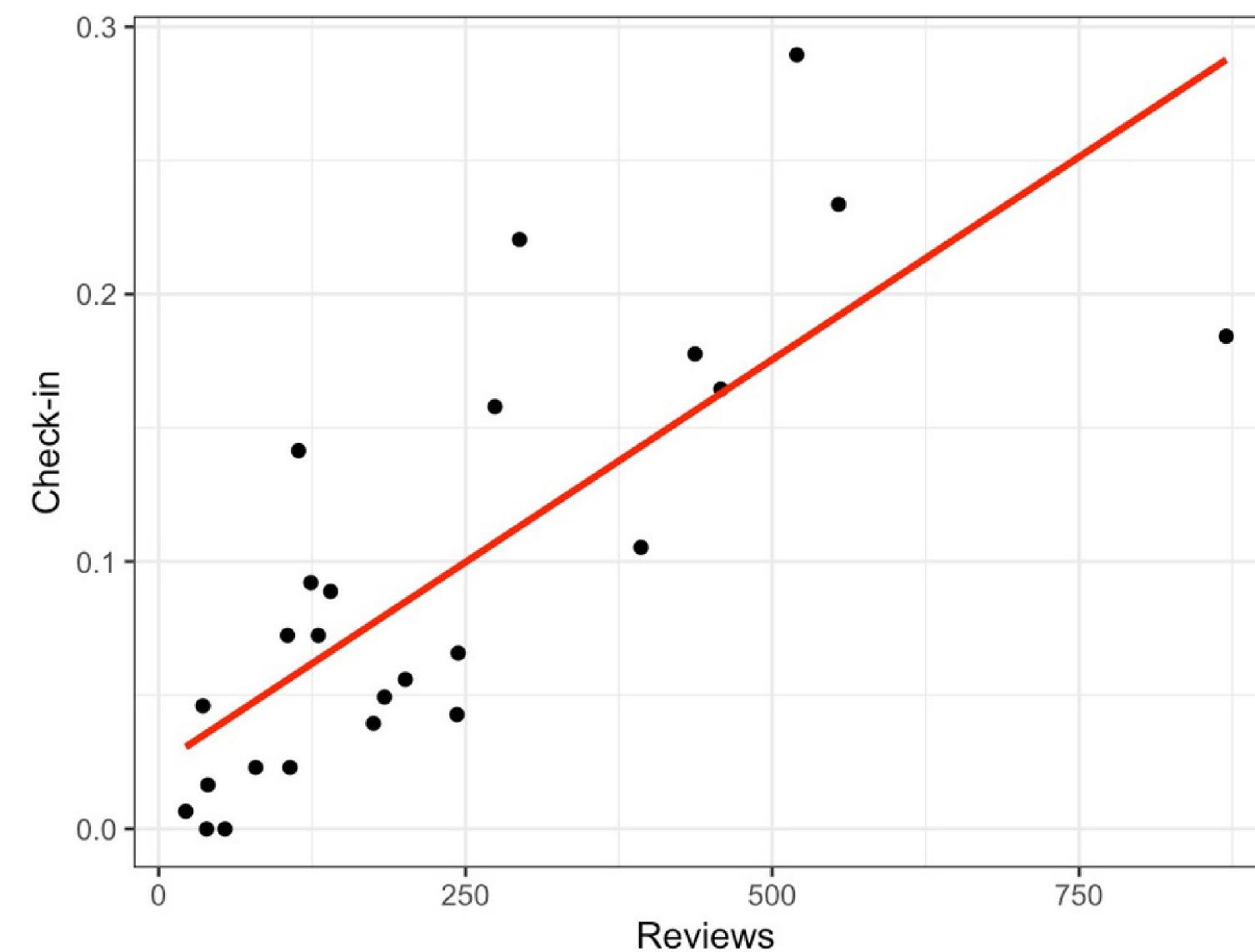
# Correlations

Correlation plots of review data (left) and weather variables subset (right)



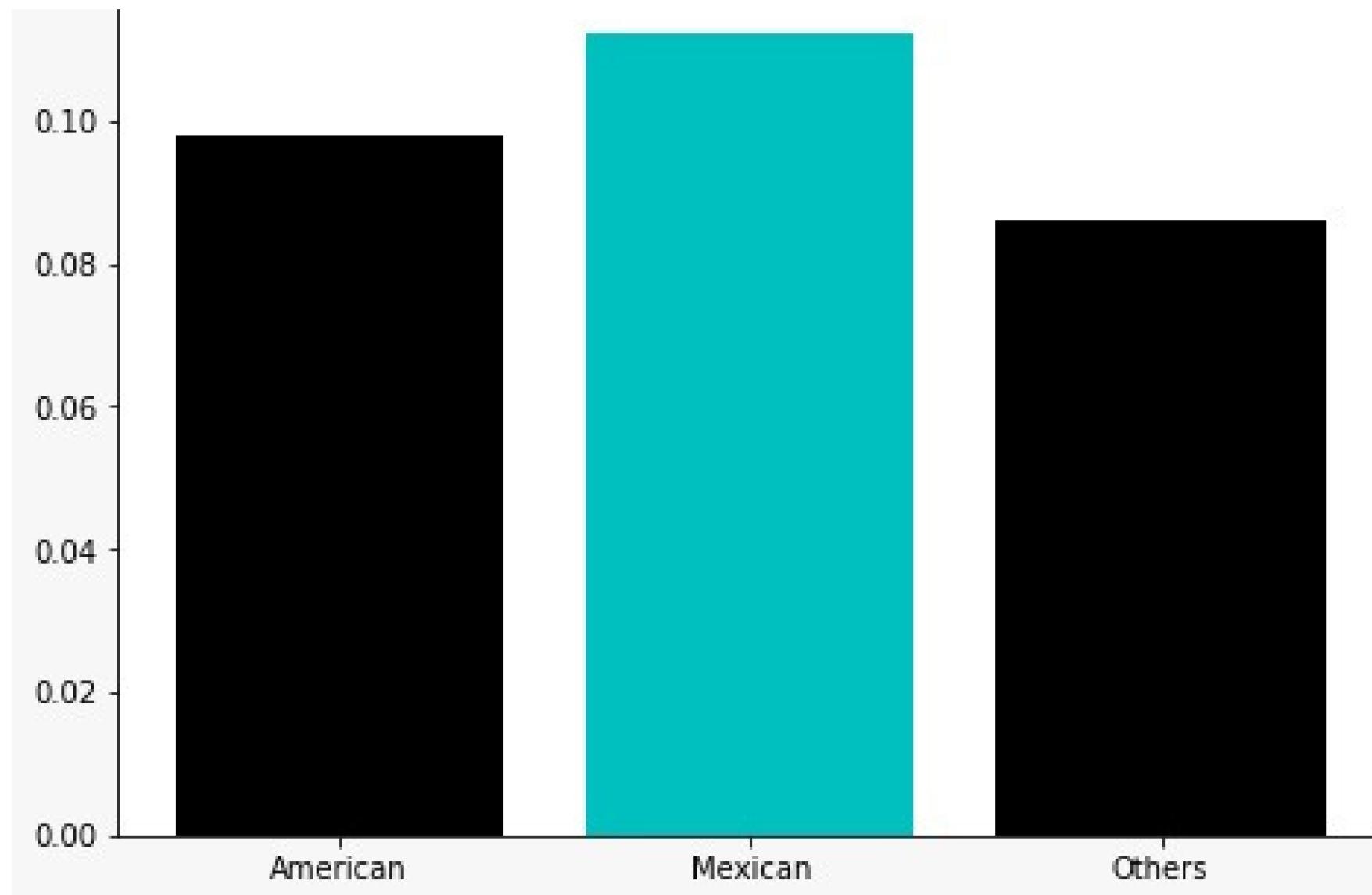
# Density plots

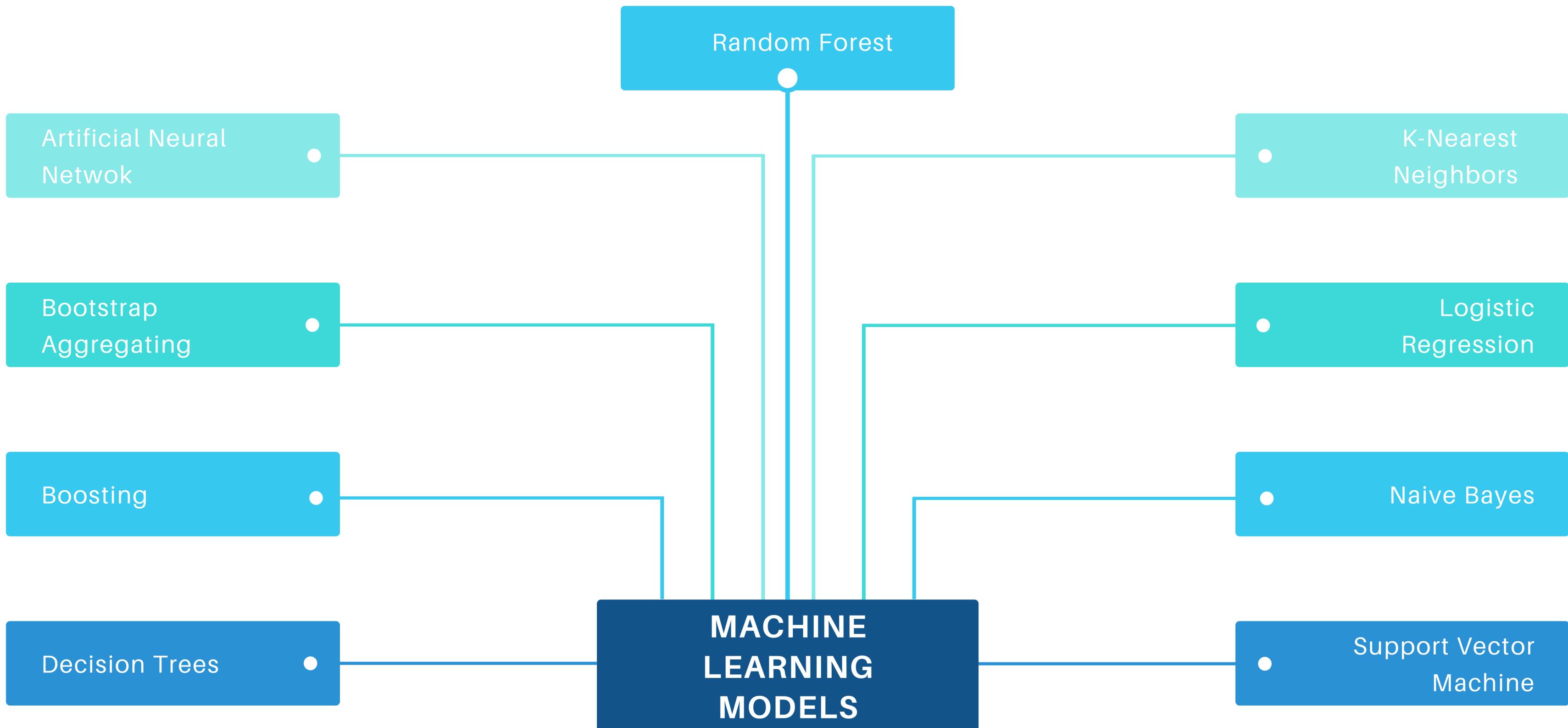
Check-in likelihood split by number of Reviews (left) and by Rating (right)



# Density plots

Check-in likelihood split by number of Restaurant category





# Logistic Regression

## Significant business variables

- business\_wifi
- business\_price
- WE1

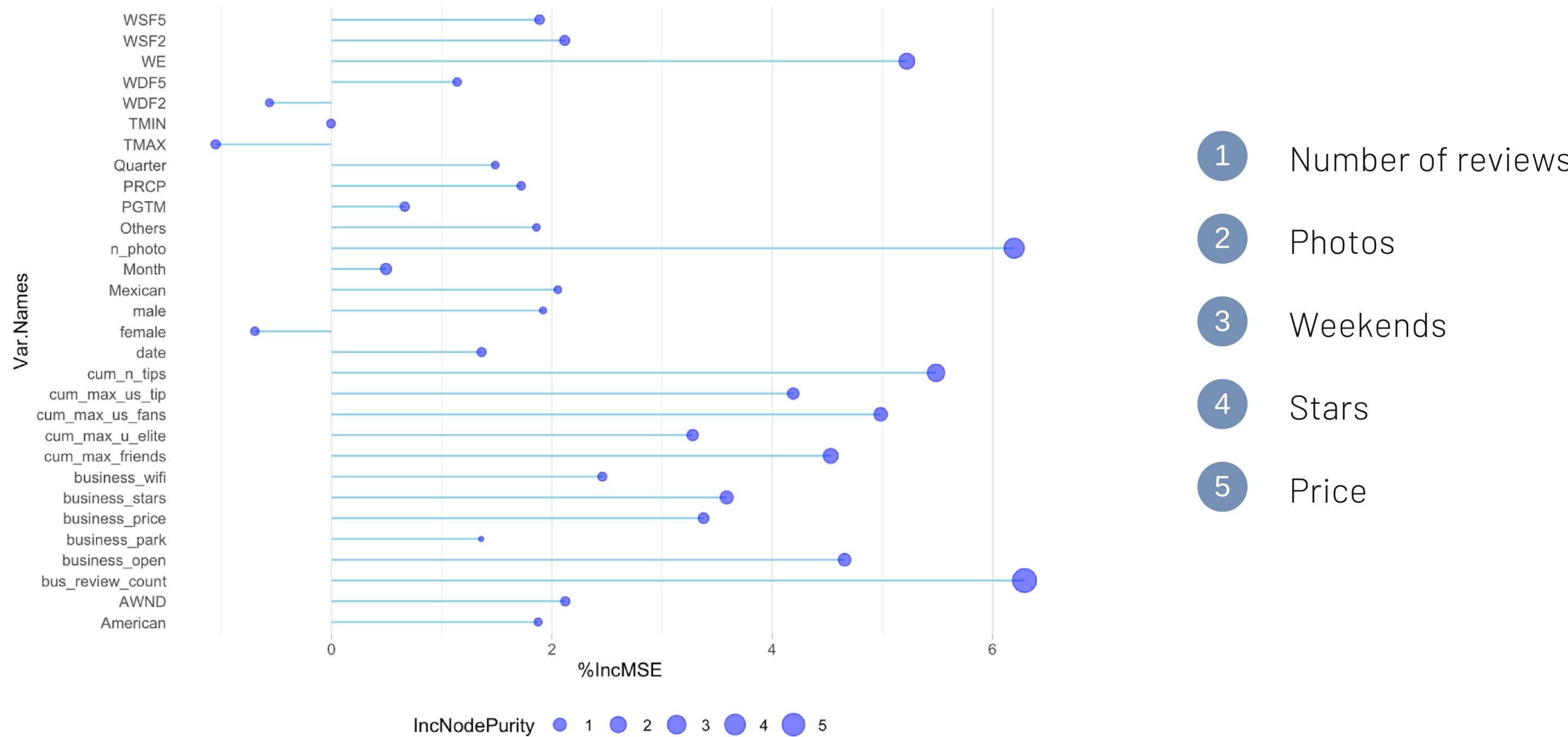
## The effect

Increases the check-in probability:

- The availability of Wi-Fi
- The higher price ranges
- Open on weekend

<i>Dependent variable:</i>	<i>ch_in</i>	
	Coefficients	St. dev.
<i>Constant</i>	276.236***	93.435
<i>business_stars</i>	-0.113	0.222
<i>bus_review_count</i>	-0.0003	0.001
<i>business_wifi</i>	1.222***	0.164
<i>business_park</i>	-0.021	0.387
<i>business_price</i>	1.090***	0.172
<i>business_open</i>	1.748***	0.339
<i>n_photo</i>	0.062***	0.010
<i>cum_n_tips</i>	0.130***	0.036
<i>cum_max_friends</i>	0.001	0.001
<i>cum_max_u_elite</i>	0.005	0.061
<i>cum_max_us_fans</i>	-0.003	0.004
<i>cum_max_us_tip</i>	0.00000	0.0002
<i>PRCP</i>	0.0002	0.001
<i>TMAX</i>	-0.001	0.002
<i>TMIN</i>	0.002	0.002
<i>AWND</i>	-0.005	0.005
<i>PGTM</i>	0.0001*	0.0001
<i>WDF2</i>	0.0003	0.001
<i>WDF5</i>	0.001	0.001
<i>WSF2</i>	0.010	0.006
<i>WSF5</i>	-0.006	0.005
<i>male</i>	0.027	0.058
<i>female</i>	-0.019	0.034
<i>date</i>	-0.016***	0.005
<i>WE1</i>	0.941***	0.097
<i>QuarterQ2</i>	0.342	0.269
<i>QuarterQ3</i>	2.744***	0.989
<i>QuarterQ4</i>	3.282***	1.144
<i>MonthAugust</i>	-0.481*	0.274
<i>MonthFebruary</i>	-0.283	0.261
<i>MonthJanuary</i>	-0.791**	0.370
<i>MonthJuly</i>	-0.654*	0.391
<i>MonthJune</i>	0.957**	0.404
<i>MonthMay</i>	0.424	0.289
<i>American</i>	-1.280***	0.240
<i>Others</i>	-0.159	0.275

# Variable importance



# Models Comparison

	Accuracy	Sensitivity	Specificity	AUC
Artificial Neural Network	0.7783	0.19444	.83939	54.0%
Bootstrap Aggregating	0.7382	0.22149	0.95959	78.3%
Boosting	<b>0.7013</b>	<b>0.76389</b>	<b>0.69477</b>	<b>79.8%</b>
Decision Tree	0.6928	0.78472	0.68314	79.7%
K-Nearest Neighbors	0.552	0.76389	0.76389	67.9%
Logistic Regression	0.7072	0.70139	0.70785	76.9%
Naive Bayes	0.825	0.41667	0.86773	64.2%
Random Forest	0.698	0.77083	0.69041	77.9%
Support Vector Machine	0.9053	0.00000	1.00000	58.1%

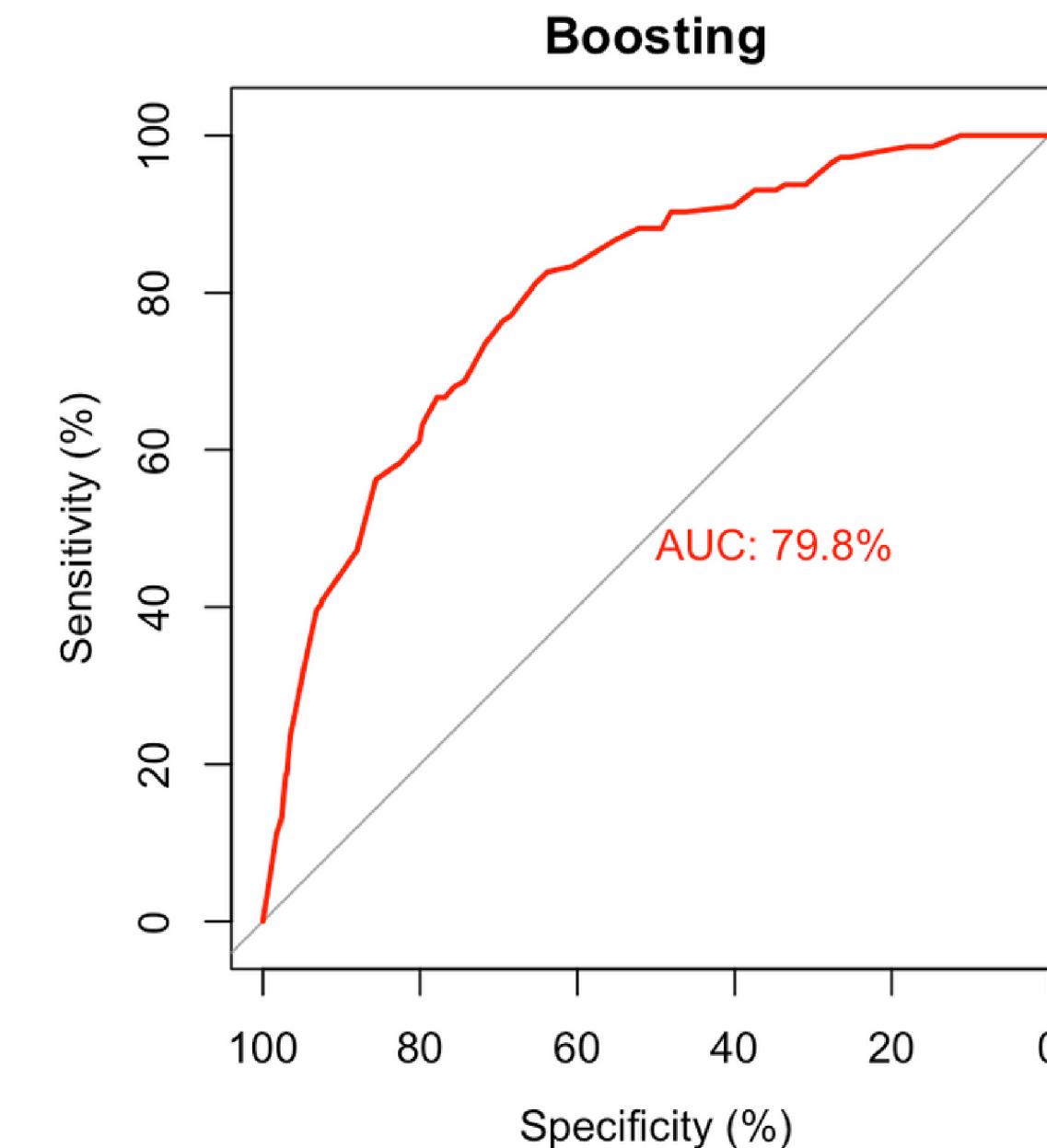
# Gradient Boosting

Accuracy: 70.13%

Sensitivity: 76.38%

Specificity: 69.48%

AUC: 79.8%



# Market Simulation

Given 3 different business configurations:

American



Reviews: 299



Mexican



Reviews: 299



Asian



Reviews: 299



# Market Simulation

Given 3 different business configurations:

American



Reviews: 299



Check-in prediction: 0.119

Mexican



Reviews: 299



Check-in prediction: 0.128

Asian

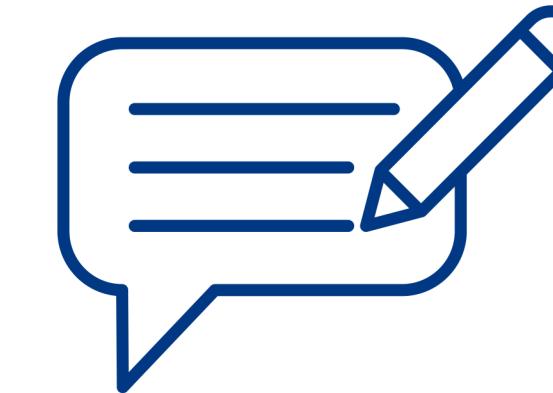
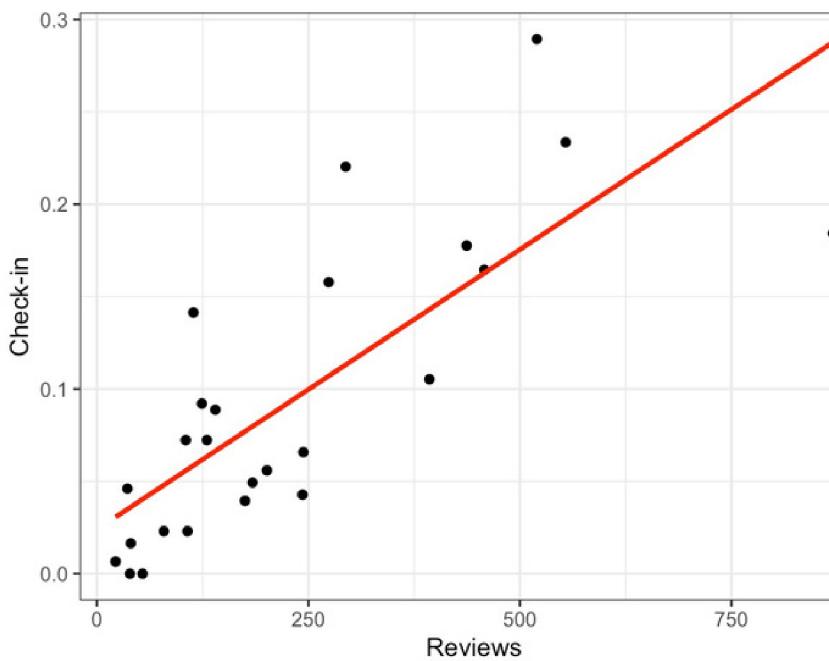


Reviews: 299



Check-in prediction: 0.079

# Ways to increase positive Reviews for the Restaurant on Yelp:



- Build a strong online presence - establish brand's presence on all major platforms.
- Encourage Reviews from loyal customers
- Respond to the Reviews
- Incentivize customers - Rewarding these customers with coupons, meal vouchers, discounts, etc. would help your restaurant to get more reviews and encourage repeat business.
- Repost the UGC - Share the positive restaurant reviews across Restaurant's social media channels.

# Conclusion

## Work done

- Mined the data
- Cleaned and transformed the data
- Built regression and created machine learning models
- Identified important variables
- Found best performing algorithm to predict check-in behavior on Yelp
- Determined best restaurant attributes configuration
- Suggested marketing campaign to induce reviews and ratings



## Outcome

Best restaurant in town with high check-in score can be Mexican restaurant with price range 2, Wi-Fi, parking place and credit card payment.

Important to encourage customers to leave online reviews and create better Restaurant Rating.

## Critical discussion

- Not all data that might potentially effect on check-in behavior were used in the analysis
- The effect of marketing campaigns was not included in the analysis
- The sample and recommendations are limited to Lakewood, OH, USA



**Thank you!**

Victoria Sibatorkina, 7738742  
Zurya Jeronimo Valencia, 7737125  
Gil Zalcberg, 6857556