



\$ RESTAURANT COUNTS

MACHINE LEARNING ON 1 DOLLAR SIGN RESTURANTS

Null Hypotheses

1\$ sign restaurant distribution is not related to:

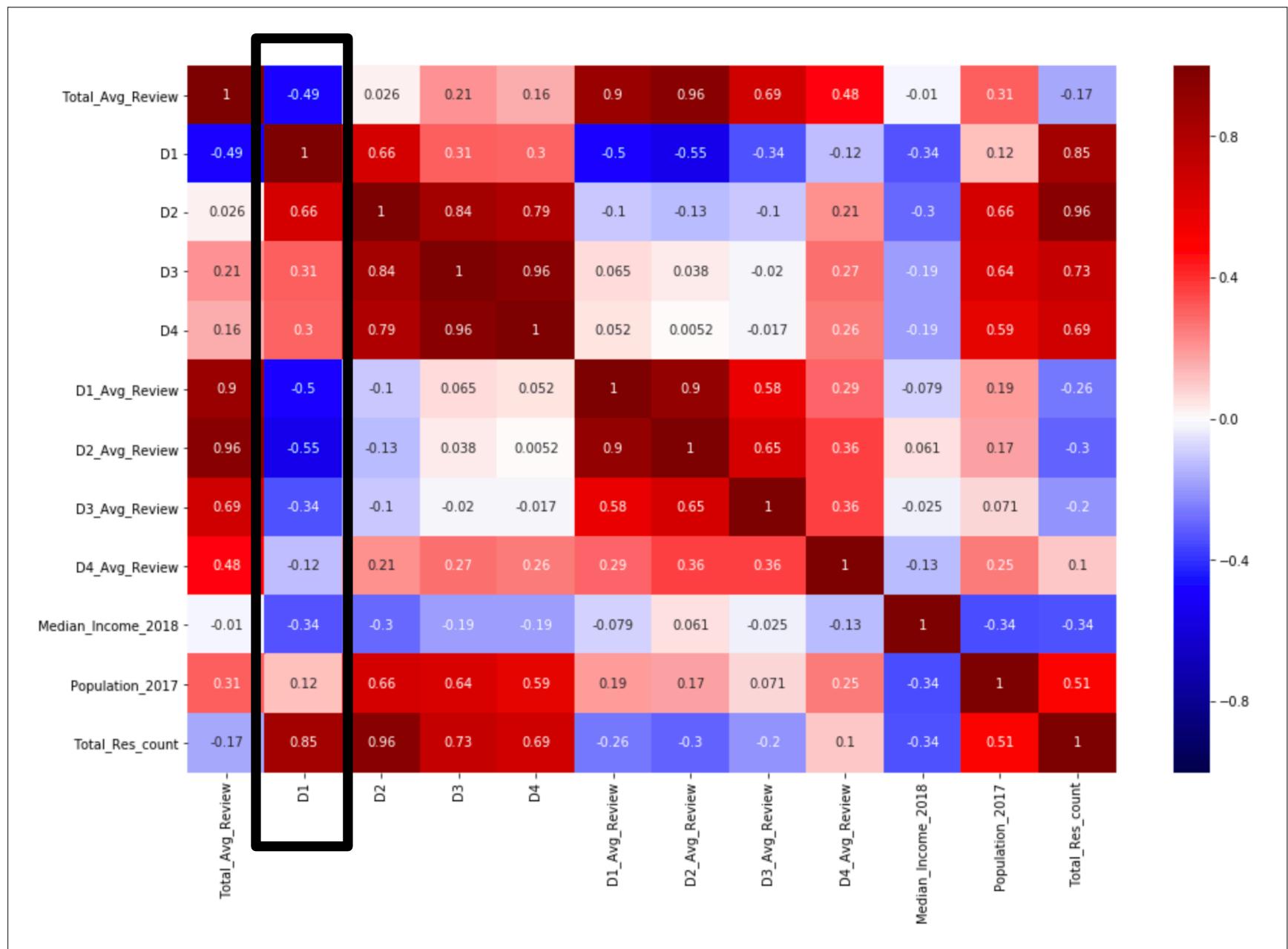
- Income
- Population
- Local eating habits

What to collect

- Restaurants name
 - \$ signs
 - Review counts
 - Address
- Median Household Income based on Cities
- Population on Cities



Heat Map Correlation

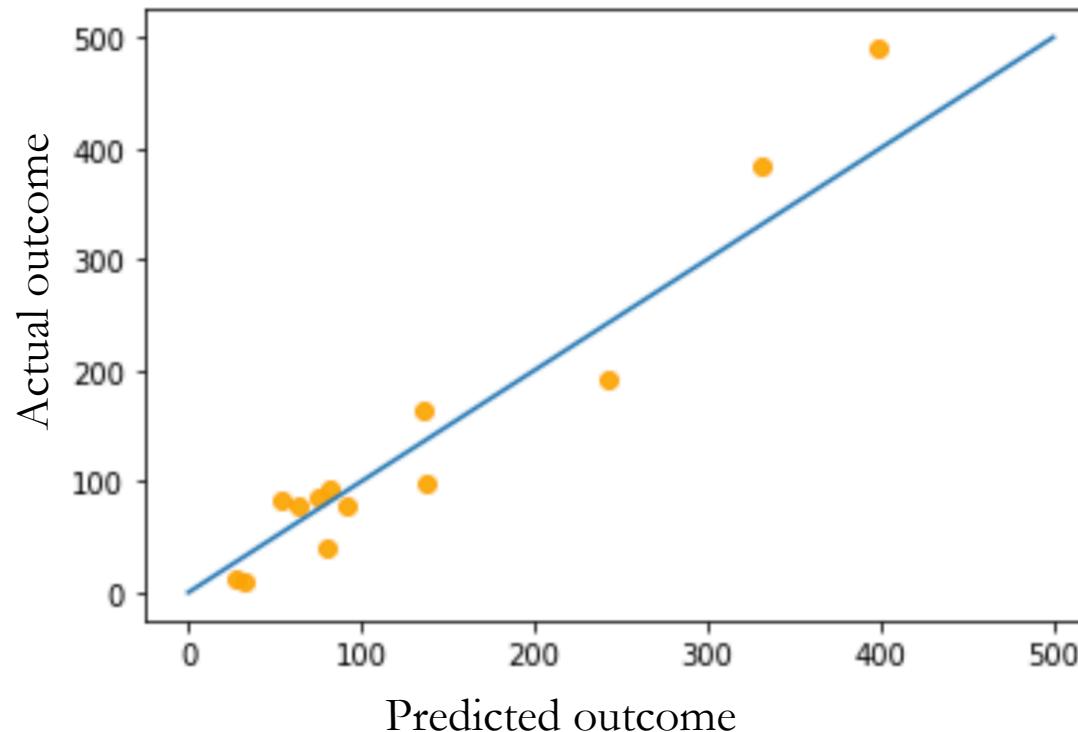


Result Lasso

- Total Restaurants count
- Population
- Income
- Eating out

Validation R² score was: 0.9207542467825832
Feature coefficient results:

Total_Res_count : 89.14826045314489 >> 0.3540757438437085
Population_2017 : -20.875107613133466 >> -0.08291097569786565
Median_Income_2018 : -15.021466475202494 >> -0.05966170162822312
Total_Avg_Review : -25.211535814489856 >> -0.10013423987840563



Conclusion

Cheap food economy is tied to

- Total restaurants per area
- Area population
- Local eating habits
- Income

1

Increase data sets

2

Add more features

- Unemployment rate
- Demographic features (college town)

3

Breakdown cities by size
Breakdown by income

Future work

Questions

Appendix

Result OLS

- X1 : Restaurant counts
- X2 : Population
- X3 : Income
- X4 : Avg Reviews

OLS Regression Results

Dep. Variable: D1 R-squared: 0.905
Model: OLS Adj. R-squared: 0.899
Method: Least Squares F-statistic: 138.6
Date: Thu, 10 Oct 2019 Prob (F-statistic): 5.68e-29
Time: 16:19:07 Log-Likelihood: -314.75
No. Observations: 63 AIC: 639.5
Df Residuals: 58 BIC: 650.2
Df Model: 4
Covariance Type: nonrobust

	coef	std err	t	P> t	[0.025	0.975]
const	112.2698	4.697	23.904	0.000	102.868	121.671
x1	107.9174	6.030	17.895	0.000	95.846	119.989
x2	-38.2096	6.286	-6.078	0.000	-50.793	-25.627
x3	-16.1438	5.107	-3.161	0.003	-26.368	-5.920
x4	-26.7095	5.379	-4.965	0.000	-37.477	-15.942

Omnibus: 12.612 Durbin-Watson: 1.513
Prob(Omnibus): 0.002 Jarque-Bera (JB): 40.687
Skew: 0.217 Prob(JB): 1.46e-09
Kurtosis: 6.913 Cond. No. 2.37

Setups

- Tools
 - Pandas
 - Numpy
 - Statsmodel
 - Sklearn
 - Beautiful soup
 - Selenium
- Data
 - Yelp
 - US Bureau



State	City	Total_Avg_Review	D1	D2	D3	D4	D1_Avg_Review	D2_Avg_Review	D3_Avg_Review	D4_Avg_Review	Median_Income_2018	Population_2017	Total_Res_count
AZ	Surprise	89.965854	99.0	105.0	1	0	45.0	78.5	0.0	0.0	70280.0	138144.0	205.0
CA	San Leandro	238.528205	100.0	94.0	1	0	128.5	236.0	1231.0	0.0	81722.0	89683.0	195.0
CA	San Diego	865.789941	164.0	478.0	31	3	349.0	615.5	1055.0	1256.0	79646.0	1425999.0	676.0
WA	Seattle	642.889145	29.0	331.0	59	14	347.0	398.0	605.0	590.0	93481.0	744949.0	433.0
CA	Pasadena	610.387417	85.0	195.0	19	3	286.0	517.0	705.0	409.0	77549.0	141374.0	302.0
CA	Alhambra	510.532895	77.0	74.0	1	0	239.0	355.0	524.0	0.0	57432.0	84650.0	152.0
CA	Fremont	398.602837	104.0	174.0	4	0	223.5	353.5	303.5	0.0	136718.0	237815.0	282.0
WA	Bellevue	326.574468	40.0	174.0	18	3	144.0	239.5	485.5	636.0	113698.0	147595.0	235.0

Features

- Restaurants counts
- Avg Review counts
- 2018 Median Household income
- 2017 Population

	Total_Avg_Review	D1	D2	D3	D4	D1_Avg_Review	D2_Avg_Review	D3_Avg_Review	D4_Avg_Review	Median_Income_2018	Population_2017	Total_Res_count
count	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000	87.000000	8.700000e+01	87.000000
mean	434.185190	104.264368	167.356322	11.310345	2.080460	308.500000	348.390805	388.091954	191.068966	90784.620690	3.10022e+05	285.011494
std	278.607513	225.639631	232.755435	20.742783	4.109598	756.842127	202.326188	435.102905	324.628833	31220.842689	5.670787e+05	460.417745
min	52.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	45938.000000	6.322600e+04	1.000000
25%	232.027543	9.500000	23.500000	0.000000	0.000000	83.500000	217.250000	0.000000	0.000000	67697.000000	8.287700e+04	34.000000
50%	413.080000	60.000000	112.000000	3.000000	0.000000	175.000000	333.500000	315.000000	0.000000	81652.000000	1.122620e+05	191.000000
75%	575.442032	99.500000	193.500000	13.500000	3.000000	272.250000	449.250000	572.500000	354.250000	112706.000000	2.430420e+05	316.500000
max	1507.000000	1918.000000	1634.000000	1 108.000000	23.000000	6591.000000	950.000000	2407.000000	1652.500000	183038.000000	3.990469e+06	3665.000000

.Describe ()

- D# : restaurant counts of \$ sign
- D#_Avg_Reviews : Mean reviews count of all restaurants under each \$ sign

