

mc7787@peel.hpc.nyu.edu's password:

Last login: Tue Apr 26 19:22:19 2022 from 10.16.213.128

```
[base] [mc7787@hlog-2 ~]$ module load python/gcc/3.7.9
(base) [mc7787@hlog-2 ~]$ PYTHONSTARTUP=team3/ana_code/acode.py pyspark --deploy-mode client
[Python 3.7.9 (default, Feb 5 2021, 09:29:06)
[GCC 4.8.5 20150623 (Red Hat 4.8.5-39)] on linux
Type "help", "copyright", "credits" or "license" for more information.
[Setting default log level to "WARN".
[To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/04/26 22:59:18 WARN util.Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Welcome to
```

```
    / _/ \
   / \_ \_ `/_/ \
  /__ / .__/\_\_/_/ \_\_ \
   /_/
version 2.4.0-cdh6.3.4
```

Using Python version 3.7.9 (default, Feb 5 2021 09:29:06)

SparkSession available as 'spark'.

22/04/26 23:00:35 WARN util.Utils: Truncated the string representation of a plan since it was too large. This behavior can be adjusted by setting 'spark.debug.maxToStringFields' in SparkEnv.conf.

22/04/26 23:00:39 WARN scheduler.TaskSetManager: Stage 9 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.

number_of_reviews	review_scores_rating	reviews_per_month	crime_counts	distance	Parking	Tax	Transient Occupancy Tax	review price	features
2.7855841276421556	4.87	0.2687839508529928	0.9088216503511483	1.0829787777748314	0	0	0	0 161.0 [2.78558412764215...	
1.5864300852456286	4.75	0.7230772906063304	0.9011719515200468	-0.1876087286261371	0	0	0	0 88.0 [1.58643008524562...	
-0.39055360627296987	4.9	-0.23424583914208236	0.8949131070218728	-0.4885735904875884	0	0	0	1 99.0 [-0.3905536062729...	
-0.4013567958441097	4.95	-0.2203211390730145	0.8949131070218728	-0.8853743832790563	0	0	0	1 83.0 [-0.4013567958441...	
0.5925366447007594	4.88	-0.22554290159891494	0.5437223879576667	-0.47494512220547364	0	0	0	0 235.0 [0.59253664470075...	
-0.4013567958441097	4.2	-0.3282375646082902	0.2738966473697223	0.4123838002247539	0	0	0	0 56.0 [-0.4013567958441...	
-0.5201918811266484	4.63	-0.34042167716872446	0.2738966473697223	0.4123838002247539	0	0	0	0 56.0 [-0.5201918811266...	
-0.4877823124132288	4.36	-0.3369405021514575	0.2738966473697223	0.4123838002247539	0	0	0	0 56.0 [-0.4877823124132...	
-0.16368662527903233	4.95	-0.2760199393492858	0.31353599585815745	0.20555549287002228	0	0	0	1 200.0 [-0.1636866252790...	
-0.5850110185534877	2.5	-0.35260578972915885	0.28015549186789623	0.8609949411755426	0	0	0	0 58.0 [-0.5850110185534...	
0.12799949314174447	4.91	-0.1733252763399106	0.36499760617647675	0.865149489257108	0	0	0	1 175.0 [0.12799949314174...	
0.24683457842428316	4.9	-0.09325825094277063	0.379601576672216	2.9580630436604625	0	0	0	1 400.0 [0.24683457842428...	
-0.5742078289823478	4.0	-0.34216226467735794	0.3232719761886503	1.8653699693470505	0	0	0	0 105.0 [-0.5742078289823...	
-0.5850110185534877	5.0	-0.34912461471189193	0.28085091903436005	-1.0313305036766882	0	0	0	1 127.0 [-0.5850110185534...	
-0.5850110185534877	4.5	-0.3334593271341906	0.32118569468925895	-0.01515622005177798	0	0	0	0 90.0 [-0.5850110185534...	
0.08478673485718494	4.83	0.007695824557971097	0.36499760617647675	0.8684953668491024	0	0	0	0 298.0 [0.08478673485718...	
-0.5958142081246276	5.0	-0.35260578972915885	0.3872512755033175	3.4884014628534614	0	0	0	1 110.0 [-0.5958142081246...	
-0.5850110185534877	5.0	-0.3439028521859914	0.3142314230246212	1.0986273783540599	0	0	0	1 90.0 [-0.5850110185534...	
-0.32573446884613055	4.92	0.06165403732560892	0.2676378028715483	2.143025000034152	0	0	0	1 271.0 [-0.3257344688461...	
-0.5958142081246276	5.0	-0.32997815211692366	0.4387128858216368	3.675622114376483	0	0	0	1 115.0 [-0.5958142081246...	

only showing top 20 rows

22/04/26 23:00:40 WARN scheduler.TaskSetManager: Stage 10 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.

features price
[2.78558412764215... 161.0
[1.58643008524562... 88.0
[-0.3905536062729... 99.0
[-0.4013567958441... 83.0
[0.59253664470075... 235.0
[-0.4013567958441... 56.0
[-0.5201918811266... 56.0
[-0.4877823124132... 56.0

caoming — mc7787@hlog-2:~ — ssh mc7787@peel.hpc.nyu.edu — 204x63

```
22/04/26 23:00:40 WARN scheduler.TaskSetManager: Stage 10 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.
+-----+
|       features|price|
+-----+
|[2.78558412764215...|161.0|
|[1.58643008524562...| 88.0|
|[-0.3905536062729...| 99.0|
|[-0.4013567958441...| 83.0|
|[0.59253664470075...|235.0|
|[-0.4013567958441...| 56.0|
|[-0.5201918811266...| 56.0|
|[-0.4877823124132...| 56.0|
|[-0.1636866252790...|200.0|
|[-0.5850110185534...| 58.0|
|[0.12799949314174...|175.0|
|[0.24683457842428...|400.0|
|[-0.5742078289823...|105.0|
|[-0.5850110185534...|127.0|
|[-0.5850110185534...| 90.0|
|[0.08478673485718...|298.0|
|[-0.5958142081246...|110.0|
|[-0.5850110185534...| 90.0|
|[-0.3257344688461...|271.0|
|[-0.5958142081246...|115.0|
+-----+
only showing top 20 rows
```

```
22/04/26 23:00:40 WARN scheduler.TaskSetManager: Stage 11 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:41 WARN scheduler.TaskSetManager: Stage 12 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:41 WARN netlib.BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeSystemBLAS
22/04/26 23:00:41 WARN netlib.BLAS: Failed to load implementation from: com.github.fommil.netlib.NativeRefBLAS
22/04/26 23:00:42 WARN scheduler.TaskSetManager: Stage 13 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:42 WARN scheduler.TaskSetManager: Stage 14 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:43 WARN scheduler.TaskSetManager: Stage 15 contains a task of very large size (140 KB). The maximum recommended task size is 100 KB.
```

Coefficients: [-15.90438019874966, 14.900020704571334, -4.288642880835673, 19.598146072087253, 3.840327014837024, 1052.8957102496656, 5.040568970498173, 61.35023682464522]

Intercept: 105.74246698025671

RMSE: 463.552281

r2: 0.037663

```
22/04/26 23:00:57 WARN scheduler.TaskSetManager: Stage 23 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
```

```
+-----+
|       features|review|
+-----+
|[161.0,2.78558412...| 0|
|[88.0,1.586430085...| 0|
|[99.0,-0.39055360...| 1|
|[83.0,-0.40135679...| 1|
|[235.0,0.59253664...| 0|
|[56.0,-0.40135679...| 0|
|[56.0,-0.52019188...| 0|
|[56.0,-0.48778231...| 0|
|[200.0,-0.1636866...| 1|
|[58.0,-0.58501101...| 0|
|[175.0,0.12799949...| 1|
|[400.0,0.24683457...| 1|
|[105.0,-0.5742078...| 0|
|[127.0,-0.5850110...| 1|
|[90.0,-0.58501101...| 0|
|[298.0,0.08478673...| 0|
|[110.0,-0.5958142...| 1|
|[90.0,-0.58501101...| 1|
```

caoming — mc7787@hlog-2:~ — ssh mc7787@peel.hpc.nyu.edu — 204x63

	features review	
[[161.0, 2.78558412...]	0	
[[88.0, 1.586430085...]	0	
[[99.0, -0.39055360...]	1	
[[83.0, -0.40135679...]	1	
[[235.0, 0.59253664...]	0	
[[56.0, -0.40135679...]	0	
[[56.0, -0.52019188...]	0	
[[56.0, -0.48778231...]	0	
[[200.0, -0.1636866...]	1	
[[58.0, -0.58501101...]	0	
[[175.0, 0.12799949...]	1	
[[400.0, 0.24683457...]	1	
[[105.0, -0.5742078...]	0	
[[127.0, -0.5850110...]	1	
[[90.0, -0.58501101...]	0	
[[298.0, 0.08478673...]	0	
[[110.0, -0.5958142...]	1	
[[90.0, -0.58501101...]	1	
[[271.0, -0.3257344...]	1	
[[115.0, -0.5958142...]	1	

only showing top 20 rows

22/04/26 23:00:57 WARN scheduler.TaskSetManager: Stage 24 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:57 WARN scheduler.TaskSetManager: Stage 26 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:58 WARN scheduler.TaskSetManager: Stage 27 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:58 WARN scheduler.TaskSetManager: Stage 28 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:00:59 WARN scheduler.TaskSetManager: Stage 30 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 32 contains a task of very large size (183 KB). The maximum recommended task size is 100 KB.
22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 34 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 36 contains a task of very large size (183 KB). The maximum recommended task size is 100 KB.
22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 38 contains a task of very large size (183 KB). The maximum recommended task size is 100 KB.
22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 40 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.
22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 41 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.

	features review	rawPrediction	probability prediction
[[36.0, -0.55260144...]	0	[[15.0127335878535...]	[[0.75063667939267...], 0.0]
[[38.0, 1.791690687...]	0	[[16.5601174708180...]	[[0.82800587354090...], 0.0]
[[38.0, 2.958435160...]	0	[[16.9588102812755...]	[[0.84794051406377...], 0.0]
[[40.0, -0.52019188...]	0	[[15.7180542988576...]	[[0.78590271494288...], 0.0]
[[40.0, -0.36894722...]	0	[[15.6267716162897...]	[[0.78133858081448...], 0.0]
[[41.0, -0.41215998...]	0	[[15.7237881827385...]	[[0.78618940913692...], 0.0]
[[42.0, -0.57420782...]	1	[[9.38730546005839...]	[[0.46936527300291...], 1.0]
[[44.0, -0.33653765...]	0	[[15.6760559664360...]	[[0.78380279832180...], 0.0]
[[44.0, 0.462898369...]	0	[[15.9086838213536...]	[[0.79543419106768...], 0.0]
[[44.0, 0.841010004...]	0	[[14.7051143788414...]	[[0.73525571894207...], 0.0]
[[45.0, -0.59581420...]	0	[[7.78340874955858...]	[[0.38917043747792...], 1.0]
[[45.0, -0.55260144...]	0	[[13.5771484099209...]	[[0.67885742049604...], 0.0]
[[46.0, -0.49858550...]	0	[[13.8027721859519...]	[[0.69013860929759...], 0.0]
[[48.0, -0.55260144...]	1	[[13.6578560132873...]	[[0.68289280066436...], 0.0]
[[49.0, 0.182015440...]	0	[[14.6542632135253...]	[[0.73271316067626...], 0.0]
[[55.0, -0.40135679...]	0	[[14.6529874581943...]	[[0.73264937290971...], 0.0]
[[56.0, -0.48778231...]	0	[[14.7493579809483...]	[[0.73746789904741...], 0.0]
[[58.0, -0.55260144...]	0	[[13.4391167887229...]	[[0.67195583943614...], 0.0]
[[59.0, -0.42296317...]	0	[[14.3220839654410...]	[[0.71610419827205...], 0.0]
[[60.0, -0.59581420...]	1	[[8.23355149946101...]	[[0.41167757497305...], 1.0]
[[60.0, -0.49858550...]	0	[[15.0088235667885...]	[[0.75044117833942...], 0.0]

[[48.0,-0.55260144...	1 [13.6578560132873... [0.68289280066436...	0.0
[[49.0,0.182015440...	0 [14.6542632135253... [0.73271316067626...	0.0
[[55.0,-0.40135679...	0 [14.6529874581943... [0.73264937290971...	0.0
[[56.0,-0.48778231...	0 [14.7493579809483... [0.73746789904741...	0.0
[[58.0,-0.55260144...	0 [13.4391167887229... [0.67195583943614...	0.0
[[59.0,-0.42296317...	0 [14.3220839654410... [0.71610419827205...	0.0
[[60.0,-0.59581420...	1 [8.23355149946101... [0.41167757497305...	1.0
[[60.0,-0.49858550...	0 [15.0088235667885... [0.75044117833942...	0.0
[[60.0,-0.00163878...	0 [15.3452686915104... [0.76726343457552...	0.0
[[62.0,-0.59581420...	0 [9.47806727571065... [0.47390336378553...	1.0
[[62.0,-0.48778231...	0 [14.7973498008325... [0.73986749004162...	0.0
[[62.0,-0.22850576...	0 [14.7920726493068... [0.73960363246534...	0.0

only showing top 25 rows

22/04/26 23:01:00 WARN scheduler.TaskSetManager: Stage 42 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.

22/04/26 23:01:01 WARN scheduler.TaskSetManager: Stage 44 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.

22/04/26 23:01:01 WARN scheduler.TaskSetManager: Stage 46 contains a task of very large size (126 KB). The maximum recommended task size is 100 KB.

Accuracy = 0.6331292282347907

Test Error = 0.3668707717652093

host_neighbourhood

host_neighbourhood	price \					
	count	mean	std	min	25%	
Alameda	2.0	109.500000	70.003571	60.0	84.75	
Alamo Square	59.0	241.711864	290.425493	59.0	100.00	
Alphabet City	1.0	300.000000	NaN	300.0	300.00	
Alta Mesa Community Association	2.0	192.000000	66.468037	145.0	168.50	
Amed	1.0	85.000000	NaN	85.0	85.00	
...	
Western Addition/NOPA	188.0	227.569149	223.669950	30.0	110.00	
Westlake	1.0	248.000000	NaN	248.0	248.00	
Westside	3.0	444.666667	314.937031	200.0	267.00	
Yerba Buena	5.0	142.000000	36.844267	100.0	125.00	
Zephyr Cove	1.0	80.000000	NaN	80.0	80.00	

host_neighbourhood	number_of_reviews \			
	50%	75%	max	count
Alameda	109.5	134.25	159.0	2.0
Alamo Square	130.0	235.50	1795.0	59.0
Alphabet City	300.0	300.00	300.0	1.0
Alta Mesa Community Association	192.0	215.50	239.0	2.0
Amed	85.0	85.00	85.0	1.0
...
Western Addition/NOPA	169.5	249.25	1995.0	188.0
Westlake	248.0	248.00	248.0	1.0
Westside	334.0	567.00	800.0	3.0
Yerba Buena	140.0	145.00	200.0	5.0
Zephyr Cove	80.0	80.00	80.0	1.0

host_neighbourhood	\					
	mean	std	min	25%	50%	
Alameda	1.500000	2.121320	0.0	0.75	1.5	
Alamo Square	42.644068	56.963014	0.0	4.00	25.0	
Alphabet City	0.000000	NaN	0.0	0.00	0.0	
Alta Mesa Community Association	0.000000	0.000000	0.0	0.00	0.0	
Amed	6.000000	NaN	6.0	6.00	6.0	
...	
Western Addition/NOPA	48.180851	81.827386	0.0	4.00	16.0	
Westlake	30.000000	NaN	30.0	30.00	30.0	
Westside	9.000000	14.730920	0.0	0.50	1.0	

	mean	std	min	25%	50%
host_neighbourhood					
Alameda	1.500000	2.121320	0.0	0.75	1.5
Alamo Square	42.644068	56.963014	0.0	4.00	25.0
Alphabet City	0.000000	NaN	0.0	0.00	0.0
Alta Mesa Community Association	0.000000	0.000000	0.0	0.00	0.0
Amed	6.000000	NaN	6.0	6.00	6.0
...
Western Addition/NOPA	48.180851	81.827386	0.0	4.00	16.0
Westlake	30.000000	NaN	30.0	30.00	30.0
Westside	9.000000	14.730920	0.0	0.50	1.0
Yerba Buena	0.200000	0.447214	0.0	0.00	0.0
Zephyr Cove	13.000000	NaN	13.0	13.00	13.0

	review_scores_rating			\
	75%	max	count	mean
host_neighbourhood				
Alameda	2.25	3.0		1.0 5.000000
Alamo Square	58.50	261.0		51.0 4.787255
Alphabet City	0.00	0.0		0.0 NaN
Alta Mesa Community Association	0.00	0.0		0.0 NaN
Amed	6.00	6.0		1.0 5.000000
...
Western Addition/NOPA	51.50	497.0		170.0 4.777176
Westlake	30.00	30.0		1.0 4.970000
Westside	13.50	26.0		2.0 4.980000
Yerba Buena	0.00	1.0		1.0 5.000000
Zephyr Cove	13.00	13.0		1.0 4.690000

	std	min	25%	50%	75%	max
host_neighbourhood						
Alameda	NaN	5.00	5.00	5.00	5.00	5.00
Alamo Square	0.387283	2.50	4.73	4.90	5.00	5.00
Alphabet City	NaN	NaN	NaN	NaN	NaN	NaN
Alta Mesa Community Association	NaN	NaN	NaN	NaN	NaN	NaN
Amed	NaN	5.00	5.00	5.00	5.00	5.00
...
Western Addition/NOPA	0.567456	0.00	4.78	4.89	5.00	5.00
Westlake	NaN	4.97	4.97	4.97	4.97	4.97
Westside	0.028284	4.96	4.97	4.98	4.99	5.00
Yerba Buena	NaN	5.00	5.00	5.00	5.00	5.00
Zephyr Cove	NaN	4.69	4.69	4.69	4.69	4.69

	reviews_per_month			\
	count	mean	std	min
host_neighbourhood				
Alameda	1.0	0.710000	NaN	0.71
Alamo Square	51.0	1.459608	4.050064	0.03
Alphabet City	0.0	NaN	NaN	NaN
Alta Mesa Community Association	0.0	NaN	NaN	NaN
Amed	1.0	0.150000	NaN	0.15
...
Western Addition/NOPA	170.0	1.273706	1.952391	0.02
Westlake	1.0	4.710000	NaN	4.71
Westside	2.0	0.780000	0.311127	0.56
Yerba Buena	1.0	0.310000	NaN	0.31
Zephyr Cove	1.0	0.510000	NaN	0.51

	25%	50%	75%	max	crime_counts \ count
host_neighbourhood					
Alameda	0.710	0.71	0.710	0.71	2.0

Yerba Buena 1.0 0.310000 NaN 0.31
Zephyr Cove 1.0 0.510000 NaN 0.51

host_neighbourhood crime_counts \\\nAlameda 0.710 0.71 0.710 0.71 2.0
Alamo Square 0.245 0.65 1.300 29.00 59.0
Alphabet City NaN NaN NaN NaN 1.0
Alta Mesa Community Association NaN NaN NaN NaN 2.0
Amed 0.150 0.15 0.150 0.15 1.0
...
Western Addition/NOPA 0.300 0.60 1.415 18.33 188.0
Westlake 4.710 4.71 4.710 4.71 1.0
Westside 0.670 0.78 0.890 1.00 3.0
Yerba Buena 0.310 0.31 0.310 0.31 5.0
Zephyr Cove 0.510 0.51 0.510 0.51 1.0

host_neighbourhood \\\nmean std min 25%
Alameda 3826.500000 645.588491 3370.0 3598.25
Alamo Square 4895.067797 238.112974 3306.0 4916.50
Alphabet City 4106.000000 NaN 4106.0 4106.00
Alta Mesa Community Association 4632.000000 31.112698 4610.0 4621.00
Amed 1667.000000 NaN 1667.0 1667.00
...
Western Addition/NOPA 4685.228723 203.195819 3920.0 4566.00
Westlake 1102.000000 NaN 1102.0 1102.00
Westside 4591.666667 315.271840 4244.0 4458.00
Yerba Buena 4459.800000 83.016263 4375.0 4418.00
Zephyr Cove 4478.000000 NaN 4478.0 4478.00

host_neighbourhood distance \\\n50% 75% max count mean
Alameda 3826.5 4054.75 4283.0 2.0 0.000533
Alamo Square 4943.0 4987.00 5045.0 59.0 0.000692
Alphabet City 4106.0 4106.00 4106.0 1.0 0.000308
Alta Mesa Community Association 4632.0 4643.00 4654.0 2.0 0.001359
Amed 1667.0 1667.00 1667.0 1.0 0.000770
...
Western Addition/NOPA 4712.0 4842.00 5115.0 188.0 0.000621
Westlake 1102.0 1102.00 1102.0 1.0 0.001409
Westside 4672.0 4765.50 4859.0 3.0 0.000460
Yerba Buena 4421.0 4501.00 4584.0 5.0 0.000381
Zephyr Cove 4478.0 4478.00 4478.0 1.0 0.000347

host_neighbourhood \\\nstd min 25% 50%
Alameda 0.000277 0.000337 0.000435 0.000533
Alamo Square 0.000394 0.000208 0.000396 0.000558
Alphabet City NaN 0.000308 0.000308 0.000308
Alta Mesa Community Association 0.000030 0.001338 0.001349 0.001359
Amed NaN 0.000770 0.000770 0.000770
...
Western Addition/NOPA 0.000322 0.000038 0.000395 0.000564
Westlake NaN 0.001409 0.001409 0.001409
Westside 0.000329 0.000161 0.000284 0.000407
Yerba Buena 0.000253 0.000122 0.000246 0.000297
Zephyr Cove NaN 0.000347 0.000347 0.000347

Amed		NaN	0.000770	0.000770	0.000770
...	
Western Addition/NOPA		0.000322	0.000038	0.000395	0.000564
Westlake		NaN	0.001409	0.001409	0.001409
Westside		0.000329	0.000161	0.000284	0.000407
Yerba Buena		0.000253	0.000122	0.000246	0.000297
Zephyr Cove		NaN	0.000347	0.000347	0.000347

75% max

host_neighbourhood					
Alameda		0.000631	0.000729		
Alamo Square		0.000851	0.001607		
Alphabet City		0.000308	0.000308		
Alta Mesa Community Association		0.001370	0.001381		
Amed		0.000770	0.000770		
...			
Western Addition/NOPA		0.000836	0.001488		
Westlake		0.001409	0.001409		
Westside		0.000610	0.000813		
Yerba Buena		0.000463	0.000777		
Zephyr Cove		0.000347	0.000347		

[163 rows x 48 columns]

neighbourhood_cleansed

neighbourhood_cleansed	price	count	mean	std	min	25%	50%	\
Bayview	167.0	111.916168	72.237678	37.0	56.00	90.0		
Bernal Heights	295.0	186.166102	133.511533	35.0	109.50	150.0		
Castro/Upper Market	288.0	267.854167	286.321083	10.0	126.00	185.0		
Chinatown	123.0	373.967480	1252.258058	31.0	95.00	195.0		
Crocker Amazon	33.0	91.333333	43.483234	38.0	50.00	86.0		
Diamond Heights	12.0	236.500000	177.243797	78.0	116.00	179.5		
Downtown/Civic Center	661.0	262.313162	707.415891	28.0	89.00	154.0		
Excelsior	117.0	117.102564	84.629964	36.0	55.00	93.0		
Financial District	167.0	754.455090	3392.235956	0.0	90.00	179.0		
Glen Park	58.0	245.982759	196.196598	28.0	121.00	192.5		
Golden Gate Park	4.0	196.750000	69.610224	155.0	155.00	166.0		
Haight Ashbury	297.0	186.020202	180.211211	28.0	90.00	132.0		
Inner Richmond	172.0	192.191860	170.159654	24.0	88.75	140.5		
Inner Sunset	146.0	265.349315	506.743989	30.0	85.00	121.0		
Lakeshore	46.0	124.021739	130.215119	40.0	55.25	80.0		
Marina	183.0	296.655738	353.765221	37.0	125.00	186.0		
Mission	556.0	178.052158	133.517846	27.0	91.75	135.0		
Nob Hill	328.0	202.786585	328.298966	0.0	80.00	120.0		
Noe Valley	232.0	223.806034	193.083997	36.0	119.00	172.0		
North Beach	189.0	266.243386	371.906294	25.0	100.00	179.0		
Ocean View	86.0	107.418605	82.810267	33.0	58.00	79.5		
Outer Mission	108.0	160.564815	148.540140	34.0	75.75	117.5		
Outer Richmond	151.0	186.827815	157.566526	28.0	92.50	132.0		
Outer Sunset	265.0	141.867925	122.634904	38.0	75.00	111.0		
Pacific Heights	146.0	291.095890	474.674921	39.0	119.00	172.0		
Parkside	110.0	158.081818	100.481513	35.0	85.25	133.5		
Potrero Hill	150.0	246.253333	218.705041	35.0	126.00	195.0		
Presidio	9.0	300.222222	179.455968	113.0	140.00	271.0		
Presidio Heights	25.0	210.560000	173.075147	60.0	109.00	150.0		
Russian Hill	168.0	325.119048	408.774374	26.0	125.00	187.5		
Seacliff	12.0	167.250000	164.216113	49.0	63.50	118.5		
South of Market	465.0	204.569892	469.045252	33.0	112.00	164.0		
Twin Peaks	50.0	264.360000	202.718641	62.0	122.25	188.5		
Visitacion Valley	47.0	142.680851	110.406976	50.0	68.50	96.0		
West of Twin Peaks	85.0	173.952941	125.835012	40.0	87.00	134.0		

South of Market	465.0	204.569892	469.045252	33.0	112.00	164.0
Twin Peaks	50.0	264.360000	202.718641	62.0	122.25	188.5
Visitacion Valley	47.0	142.680851	110.406976	50.0	68.50	96.0
West of Twin Peaks	85.0	173.952941	125.835012	40.0	87.00	134.0
Western Addition	462.0	210.116883	213.957918	30.0	95.00	140.0

neighbourhood_cleansed	number_of_reviews				\	
	75%	max	count	mean		
Bayview	149.50	415.0	167.0	42.652695		
Bernal Heights	225.00	1179.0	295.0	60.708475		
Castro/Upper Market	303.25	3000.0	288.0	63.545139		
Chinatown	285.00	9999.0	123.0	18.382114		
Crocker Amazon	111.00	224.0	33.0	71.484848		
Diamond Heights	295.00	650.0	12.0	106.416667		
Downtown/Civic Center	209.00	10000.0	661.0	15.835098		
Excelsior	138.00	449.0	117.0	62.025641		
Financial District	299.50	25000.0	167.0	46.952096		
Glen Park	318.75	1100.0	58.0	45.672414		
Golden Gate Park	207.75	300.0	4.0	173.000000		
Haight Ashbury	220.00	1536.0	297.0	53.282828		
Inner Richmond	222.00	895.0	172.0	57.889535		
Inner Sunset	199.00	3800.0	146.0	43.075342		
Lakeshore	135.75	690.0	46.0	54.956522		
Marina	300.00	2596.0	183.0	37.196721		
Mission	218.25	1080.0	556.0	46.500000		
Nob Hill	219.00	4607.0	328.0	13.807927		
Noe Valley	249.25	1250.0	232.0	51.823276		
North Beach	300.00	3990.0	189.0	30.714286		
Ocean View	132.25	625.0	86.0	44.430233		
Outer Mission	184.25	1000.0	108.0	53.370370		
Outer Richmond	225.00	1179.0	151.0	77.245033		
Outer Sunset	160.00	1264.0	265.0	84.384906		
Pacific Heights	262.00	5000.0	146.0	44.993151		
Parkside	186.00	542.0	110.0	69.327273		
Potrero Hill	292.50	1750.0	150.0	68.773333		
Presidio	350.00	649.0	9.0	89.000000		
Presidio Heights	250.00	900.0	25.0	48.560000		
Russian Hill	325.00	3250.0	168.0	29.940476		
Seacliff	163.00	535.0	12.0	112.833333		
South of Market	217.00	10000.0	465.0	17.565591		
Twin Peaks	325.75	1100.0	50.0	56.520000		
Visitacion Valley	163.00	507.0	47.0	88.723404		
West of Twin Peaks	225.00	716.0	85.0	86.282353		
Western Addition	250.00	1995.0	462.0	40.049784		

neighbourhood_cleansed	std	min	25%	50%	75%	max
	73.414890	0.0	1.00	7.0	50.00	325.0
Bayview	93.724828	0.0	4.00	19.0	79.50	568.0
Bernal Heights	104.838509	0.0	4.00	18.0	70.25	639.0
Castro/Upper Market	48.369398	0.0	0.00	2.0	7.00	285.0
Crocker Amazon	123.308587	0.0	7.00	24.0	84.00	659.0
Diamond Heights	159.806393	0.0	7.75	22.5	132.25	478.0
Downtown/Civic Center	44.394427	0.0	0.00	3.0	15.00	497.0
Excelsior	81.639606	0.0	4.00	26.0	95.00	423.0
Financial District	98.404305	0.0	0.00	2.0	18.50	489.0
Glen Park	72.985792	0.0	1.25	8.0	60.50	326.0
Golden Gate Park	119.124585	2.0	149.00	206.0	230.00	278.0
Haight Ashbury	100.326296	0.0	1.00	12.0	58.00	558.0
Inner Richmond	101.109196	0.0	1.00	9.5	69.75	577.0
Inner Sunset	77.642513	0.0	1.00	6.0	44.00	372.0

Mission	675.0	0.000564	0.000321	0.000044
Mission Bay	39.0	0.001051	0.000695	0.000246
Nob Hill	226.0	0.000487	0.000253	0.000027
Noe Valley	136.0	0.001001	0.000543	0.000039
North Beach	123.0	0.000726	0.000414	0.000037
Oceanview/Merced/Ingleside	60.0	0.001246	0.000758	0.000102
Outer Mission	106.0	0.001216	0.000615	0.000106
Outer Richmond	181.0	0.001218	0.000723	0.000047
Pacific Heights	112.0	0.000650	0.000289	0.000036
Portola	25.0	0.001154	0.000632	0.000254
Potrero Hill	122.0	0.000803	0.000440	0.000043
Presidio	1.0	0.002480	NaN	0.002480
Presidio Heights	71.0	0.000880	0.000507	0.000038
Russian Hill	167.0	0.000561	0.000257	0.000100
South of Market	213.0	0.000493	0.000271	0.000014
Sunset/Parkside	363.0	0.001387	0.000799	0.000099
Tenderloin	248.0	0.000359	0.000164	0.000024
Twin Peaks	55.0	0.001059	0.000548	0.000025
Visitacion Valley	38.0	0.000934	0.000500	0.000131
West of Twin Peaks	130.0	0.001488	0.000854	0.000030
Western Addition	166.0	0.000459	0.000195	0.000084

	25%	50%	75%	max
Neighborhoods - Analysis Boundaries				
Bayview Hunters Point	0.000508	0.000824	0.001305	0.006581
Bernal Heights	0.000495	0.000792	0.001109	0.002354
Castro/Upper Market	0.000465	0.000810	0.001261	0.004205
Chinatown	0.000243	0.000370	0.000497	0.002018
Excelsior	0.000742	0.001257	0.001741	0.003129
Financial District/South Beach	0.000294	0.000436	0.000650	0.001705
Glen Park	0.000552	0.000796	0.001176	0.002330
Golden Gate Park	0.001262	0.001577	0.001892	0.002207
Haight Ashbury	0.000156	0.000504	0.000916	0.002983
Hayes Valley	0.000351	0.000525	0.000759	0.001607
Inner Richmond	0.000616	0.000933	0.001439	0.002287
Inner Sunset	0.000492	0.000843	0.001442	0.004379
Japantown	0.000384	0.000673	0.000866	0.001260
Lakeshore	0.000931	0.001370	0.001756	0.004945
Lone Mountain/USF	0.000475	0.000751	0.001071	0.002237
Marina	0.000389	0.000671	0.001052	0.001824
Mission	0.000312	0.000508	0.000734	0.002120
Mission Bay	0.000470	0.000838	0.001668	0.002963
Nob Hill	0.000292	0.000408	0.000624	0.001242
Noe Valley	0.000548	0.000978	0.001398	0.002472
North Beach	0.000371	0.000652	0.000988	0.001959
Oceanview/Merced/Ingleside	0.000767	0.001091	0.001529	0.003148
Outer Mission	0.000811	0.001203	0.001586	0.003093
Outer Richmond	0.000660	0.001117	0.001623	0.004606
Pacific Heights	0.000437	0.000640	0.000845	0.001807
Portola	0.000653	0.001055	0.001488	0.002336
Potrero Hill	0.000463	0.000747	0.001115	0.002312
Presidio	0.002480	0.002480	0.002480	0.002480
Presidio Heights	0.000460	0.000848	0.001232	0.002326
Russian Hill	0.000339	0.000569	0.000729	0.001238
South of Market	0.000289	0.000452	0.000658	0.001176
Sunset/Parkside	0.000805	0.001241	0.001844	0.003905
Tenderloin	0.000251	0.000348	0.000447	0.000855
Twin Peaks	0.000700	0.001088	0.001460	0.002656
Visitacion Valley	0.000527	0.000865	0.001302	0.001922
West of Twin Peaks	0.000835	0.001352	0.002118	0.004414
Western Addition	0.000300	0.000433	0.000567	0.000987
Parking Tax				

Presidio Heights 0.000460 0.000848 0.001232 0.002326
Russian Hill 0.000339 0.000569 0.000729 0.001238
South of Market 0.000289 0.000452 0.000658 0.001176
Sunset/Parkside 0.000805 0.001241 0.001844 0.003905
Tenderloin 0.000251 0.000348 0.000447 0.000855
Twin Peaks 0.000700 0.001088 0.001460 0.002656
Visitacion Valley 0.000527 0.000865 0.001302 0.001922
West of Twin Peaks 0.000835 0.001352 0.002118 0.004414
Western Addition 0.000300 0.000433 0.000567 0.000987

Parking Tax
 price \\
 count mean std min 25% 50% 75%
Parking Tax
False 6364.0 217.126336 406.124886 0.0 92.00 145.0 232.0
True 48.0 1752.166667 6072.364615 40.0 103.75 158.0 209.0

number_of_reviews \\
 max count mean std min 25% 50%
Parking Tax
False 10000.0 6364.0 44.585324 85.528637 0.0 1.0 7.0
True 25000.0 48.0 26.770833 75.251137 0.0 0.0 2.0

review_scores_rating \\
 75% max count mean std min 25%
Parking Tax
False 45.0 877.0 5046.0 4.736581 0.544254 0.0 4.710
True 9.5 435.0 30.0 4.458333 0.863090 1.0 4.425

reviews_per_month \\
 50% 75% max count mean std min
Parking Tax
False 4.89 5.0 5.0 5046.0 2.051974 5.727341 0.01
True 4.78 5.0 5.0 30.0 2.697000 8.394827 0.07

crime_counts \\
 25% 50% 75% max count mean
Parking Tax
False 0.2100 0.67 2.1300 162.00 6364.0 3867.508171
True 0.2325 0.39 1.4725 46.11 48.0 4622.541667

distance \\
 std min 25% 50% 75% max count
Parking Tax
False 1396.238645 414.0 3622.00 4486.0 4765.0 5325.0 6364.0
True 365.721901 4014.0 4178.75 4810.0 4852.0 5249.0 48.0

\\
 mean std min 25% 50% 75%
Parking Tax
False 0.000776 0.000593 0.000008 0.000363 0.000612 0.001019
True 0.000419 0.000171 0.000139 0.000288 0.000397 0.000504

max
Parking Tax
False 0.006581
True 0.001091
Transient Occupancy Tax
 price \\
 count mean std min 25% 50%
Transient Occupancy Tax
False 6168.0 227.065824 680.219719 0.0 91.0 145.0
True 244.0 267.844262 410.562327 28.0 100.0 154.5

```

max
Parking Tax
False      0.006581
True       0.001091
Transient Occupancy Tax
    price
    count      mean       std   min   25%   50%   \
Transient Occupancy Tax
False      6168.0  227.065824  680.219719  0.0   91.0  145.0
True       244.0   267.844262  410.562327  28.0  100.0  154.5

    number_of_reviews
    75%      max      count      mean   \
Transient Occupancy Tax
False      230.0  25000.0      6168.0  44.814851
True       250.0  4607.0      244.0   35.278689

    std   min   25%   50%   75%      max   \
Transient Occupancy Tax
False      86.358876  0.0   1.0   7.0   45.00  877.0
True       57.863649  0.0   1.0   9.0   45.25  506.0

    review_scores_rating
    count      mean       std   min   25%   \
Transient Occupancy Tax
False      4881.0  4.732772  0.552340  0.0   4.710
True       195.0   4.789128  0.386093  1.0   4.735

    reviews_per_month
    50%      75%      max      count      mean       std   \
Transient Occupancy Tax
False      4.89    5.0     5.0      4881.0  2.066107  5.735646
True       4.89    5.0     5.0      195.0   1.797436  6.002735

    crime_counts
    min      25%      50%      75%      max      count   \
Transient Occupancy Tax
False      0.01    0.21    0.67    2.16   162.00      6168.0
True       0.01    0.27    0.71    1.70    78.95      244.0

    mean       std   min   25%   50%   \
Transient Occupancy Tax
False      3854.163262 1400.420784  414.0  3582.75  4472.0
True       4353.381148 1082.823336  580.0  4533.75  4608.0

    distance
    75%      max      count      mean       std   \
Transient Occupancy Tax
False      4763.00  5325.0     6168.0  0.000776  0.000596
True       4819.75  5023.0     244.0   0.000701  0.000473

    min      25%      50%      75%      max
Transient Occupancy Tax
False      0.000008  0.000361  0.000609  0.001018  0.006581
True       0.000027  0.000365  0.000564  0.000957  0.003516

Full model performance:
accu      0.634843
recall    0.645963
precision 0.609375
f1        0.627136

```

```
True      0.01  0.27  0.71  1.70  78.95      244.0
```

```
          mean       std     min    25%   50%
```

Transient Occupancy Tax	mean	std	min	25%	50%
False	3854.163262	1400.420784	414.0	3582.75	4472.0
True	4353.381148	1082.823336	580.0	4533.75	4608.0

```
          distance  
          75%      max   count      mean      std
```

Transient Occupancy Tax	75%	max	count	mean	std
False	4763.00	5325.0	6168.0	0.000776	0.000596
True	4819.75	5023.0	244.0	0.000701	0.000473

```
          min      25%      50%      75%      max
```

Transient Occupancy Tax	min	25%	50%	75%	max
False	0.000008	0.000361	0.000609	0.001018	0.006581
True	0.000027	0.000365	0.000564	0.000957	0.003516

Full model performance:

```
accu      0.634843  
recall    0.645963  
precision 0.609375  
f1        0.627136  
auc       0.690845
```

dtype: float64

Random Forest Graph:

```
[1.99 0.99 0.98 0.97 0.96 0.94 0.92 0.91 0.9  0.89 0.88 0.87 0.86 0.85  
 0.84 0.83 0.82 0.81 0.8  0.79 0.78 0.77 0.76 0.75 0.74 0.73 0.72 0.71  
 0.7  0.69 0.68 0.67 0.66 0.65 0.64 0.63 0.62 0.61 0.6  0.59 0.58 0.57  
 0.56 0.55 0.54 0.53 0.52 0.51 0.5  0.49 0.48 0.47 0.46 0.45 0.44 0.43  
 0.42 0.41 0.4  0.39 0.38 0.37 0.36 0.35 0.34 0.33 0.32 0.31 0.3  0.29  
 0.28 0.27 0.26 0.25 0.24 0.23 0.22 0.21 0.2  0.19 0.18 0.17 0.16 0.15  
 0.14 0.13 0.12 0.11 0.1  0.09 0.08 0.07 0.06 0.05 0.04 0.03 0.01 0.  ]
```

```
price      1.000000  
price_rating 0.000840  
crime_counts 0.000139  
number_of_reviews -0.000126  
reviews_per_month -0.000070
```

dtype: float64

```
crime_counts -0.616795  
distance     0.551331  
number_of_reviews 0.360542  
price_rating  -0.312962  
reviews_per_month 0.295926
```

dtype: float64

```
[[ 4.35835111e+02  5.55959578e-04 -1.17111925e-04  7.62559783e-04  
  3.50445495e-06  6.41085396e-05 -6.95790071e-07 -2.06739462e-06]  
[-5.47691558e-02  4.54436061e-01  6.33636157e-01  4.77214656e-02  
 -6.22022775e-01 -5.96908325e-04  6.66196075e-04  3.14047903e-05]  
[-3.07199519e-02  3.72992599e-01  7.14248667e-01  5.10833966e-03  
  5.89835318e-01  4.52424310e-02 -8.43200724e-05 -1.01079167e-04]  
[ 6.07298759e-02 -7.77425036e-01  3.03388059e-01  1.43221631e-01  
 -6.13018511e-02  5.25163408e-01 -2.81390524e-03 -2.22549039e-04]  
[-2.04422842e-02  6.94911985e-01 -3.30451620e-01 -4.11060940e-01  
  2.98244556e-04  4.88576367e-01 -3.54222856e-04  1.99094305e-04]  
[ 1.01965468e-02 -1.64750377e-03  1.26469220e-03  6.39903171e-03  
  1.04614953e-03  9.27129996e-04 -1.80814728e-03  7.56434982e-02]  
[ 2.97488343e-03 -1.18487528e-02  2.13068119e-03  5.54722155e-04  
  1.50513754e-03  8.55000000e-03  1.91618360e-01  7.11016944e-04]  
[ 3.66289748e-01 -3.94465718e-01  2.25198091e-01 -9.46648090e-01  
 -3.75700067e-02 -1.32474305e-01 -1.38720791e-04  3.92655514e-04]]
```

>>> █