

# Crime Prediction from Hotel Reviews

Robroo Somkiadcharoen, Lin Bofan  
รอบรู้ สมเกียรติเจริญ, 林博凡



The background is a dark, blurred image of a police vehicle at night. The top of the vehicle is visible, showing red and blue emergency lights that are out of focus, creating a bokeh effect. The overall color palette is dominated by dark blues and blacks, with hints of red and blue from the lights.

# Crime Sensing: Hotel Review to predict Crime Hotspots

# Hotel Dataset (2003-2016)

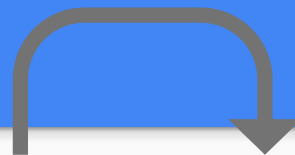
B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Hotel Name	Hotel Review Stars	Hotel Address	City	ZIP	Review Title	Review Date	Review Content	Review Stars	Reviewer Name	Reviewer Location	Reviewer Profile	Reviewer Total Reviews	Reviewer Hotel Reviews	Reviewer Helpful Votes
Travelodge	3.5	37 London	Bromley	BR	A Great stay in Br	###		5	Graham E	Port Elizabeth, South		1 review	36 hotel re	31
Travelodge	3.5	37 London	Bromley	BR	A regular stay at	###		4	John L	Folkestone, United Ki		12 review	3 hotel rev	7
Travelodge	3.5	37 London	Bromley	BR	As expected	###		3	496niccib	Bristol, Ur	<a href="https://www">https://www</a>	8 reviews	4 hotel rev	1
Travelodge	3.5	37 London	Bromley	BR	Avoid at all costs	###		1	P683YHda	Penzance,	<a href="https://www">https://www</a>	2 reviews	5 hotel rev	10
Travelodge	3.5	37 London	Bromley	BR	Awful	###		1	Catherine	St. Albans, United Kir		1 review	3 hotel rev	7
Travelodge	3.5	37 London	Bromley	BR	Awful room	###		2	325ellah	London, U	<a href="https://www">https://www</a>	1 review	7 hotel rev	18
Travelodge	3.5	37 London	Bromley	BR	Basic but good	###		3	Joh2008	Barcelona	<a href="https://www">https://www</a>	100 review	11 hotel re	32
Travelodge	3.5	37 London	Bromley	BR	Bed is like a brick	###		3	Hayley S	Liverpool, United Kin		1 review	6 hotel rev	22
Travelodge	3.5	37 London	Bromley	BR	Bromley Travelodge	###		3	C6258HA	london	<a href="https://www">https://www</a>	1 review	36 hotel re	31
Travelodge	3.5	37 London	Bromley	BR	Central position a	###		4	Fliss55	Greater M	<a href="https://www">https://www</a>	15 review	3 hotel rev	7

# Crime Dataset (3/2017 - 8/2017)

	A	B	C	D	E	F	G	H	I	J	K
1	Crim	Month	Reported by	Falls with	Longitude	Latitude	Location	LSO	LSOA name	Crime type	Last outcome category
2	6c2	2018-08	Metropolitan	Metropol	-0.805749	51.798149	On or near Lower Road	E010	Aylesbury Vale 0	Violence and sexual offences	Under investigation
3		2018-08	Metropolitan	Metropol	0.140035	51.589112	On or near Beansland Grove	E010	Barking and Dag	Anti-social behaviour	
4		2018-08	Metropolitan	Metropol	0.136387	51.589215	On or near Billet Road	E010	Barking and Dag	Anti-social behaviour	
5		2018-08	Metropolitan	Metropol	0.136513	51.588214	On or near Kingston Close	E010	Barking and Dag	Anti-social behaviour	
6		2018-08	Metropolitan	Metropol	0.134947	51.588063	On or near Mead Grove	E010	Barking and Dag	Anti-social behaviour	
7		2018-08	Metropolitan	Metropol	0.140192	51.582311	On or near Hatch Grove	E010	Barking and Dag	Anti-social behaviour	
8		2018-08	Metropolitan	Metropol	0.136387	51.589215	On or near Billet Road	E010	Barking and Dag	Anti-social behaviour	
9	2ba	2018-08	Metropolitan	Metropol	0.133181	51.586719	On or near Rosehatch Avenue	E010	Barking and Dag	Bicycle theft	Under investigation
10	7be	2018-08	Metropolitan	Metropol	0.140192	51.582311	On or near Hatch Grove	E010	Barking and Dag	Burglary	Under investigation
11	710	2018-08	Metropolitan	Metropol	0.140035	51.589112	On or near Beansland Grove	E010	Barking and Dag	Criminal damage and arson	Under investigation
12	7cb	2018-08	Metropolitan	Metropol	0.135866	51.587336	On or near Gibbfield Close	E010	Barking and Dag	Criminal damage and arson	Under investigation
13	9c5b	2018-08	Metropolitan	Metropol	0.140634	51.583427	On or near Rams Grove	E010	Barking and Dag	Other theft	Under investigation
14	ad6	2018-08	Metropolitan	Metropol	0.135866	51.587336	On or near Gibbfield Close	E010	Barking and Dag	Public order	Under investigation
15	106	2018-08	Metropolitan	Metropol	0.135554	51.584985	On or near Rose Lane	E010	Barking and Dag	Shoplifting	Under investigation
16	98b	2018-08	Metropolitan	Metropol	0.148434	51.595164	On or near Park/Open Space	E010	Barking and Dag	Vehicle crime	Under investigation
17	d55	2018-08	Metropolitan	Metropol	0.134947	51.588063	On or near Mead Grove	E010	Barking and Dag	Vehicle crime	Under investigation



# Feature Extraction



Index	Hotel Name	Reviewer Name	Review Stars	Review Title	Sentiment	Effective Star Rating
28	Travelodge London Bromley	Jean T	1	Nightmare	Negative	1
29	Travelodge London Bromley	P1475E0simo...	1	No Air conditioning - too hot, couldn't s...	Negative	1
30	Travelodge London Bromley	joyceh337	3	Noisy plumbing	Neutral	3
31	Travelodge London Bromley	izabelac800	5	One of the best if not the best hotel in ...	Positive	5
32	Travelodge London Bromley	true_angel2...	4	Overall a pleasant stay	Positive	4
33	Travelodge London Bromley	colin p	4	Pleasant stay	Positive	4

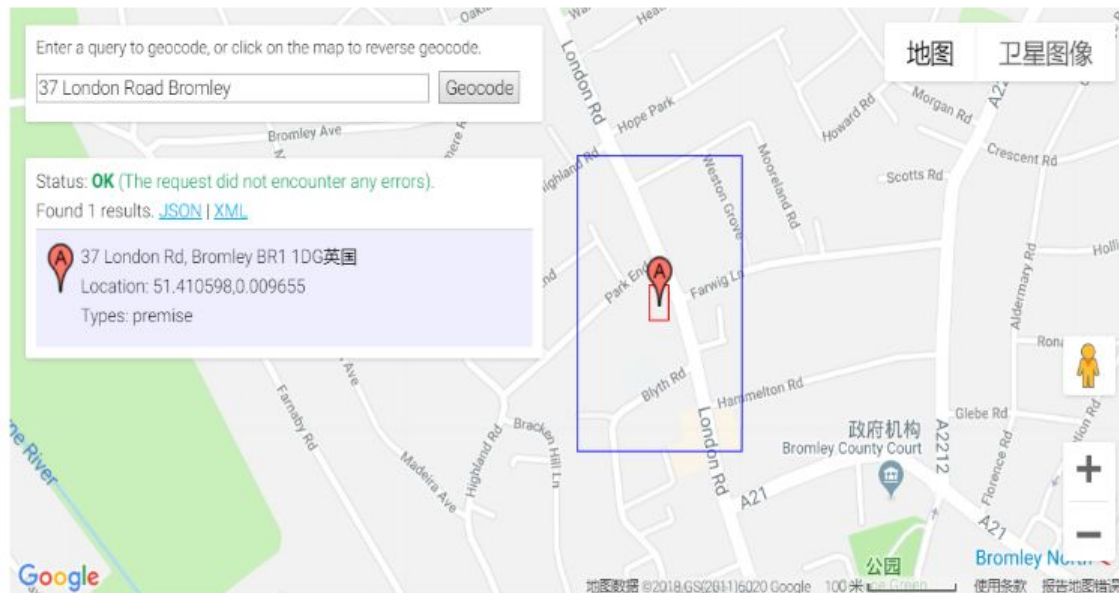


34.4k/130k contrast Review Text&Star



Travelodge South Croydon	bradwebb2002	3	rude desk clerk, and was told I had no room in spit...	Negative	nan
Travelodge South Croydon	totty01	3	Terrible smell	Negative	nan

# Google Geocoding



Using the Hotel address and City For example, '37 London Road Bromley' converts to Latitude '51.410598' Longitude '0.009655'

# Data Preprocessing

- In the original data, there are some missing value of 'Hotel address'. For those hotels, we convert the 'Hotel name' and 'City' to 'Latitude' and 'Longitude'.

Hotel Name	Hotel Review Stars	Hotel Address	City	ZIP	Review Title	Review Date
Tamasha Hotel	2.5	131 Widmore Road	Bromley	BR1 3AX	YUCK! YUCK! YUCK! AVOID	2012/7/15
Sundridge Park Manor	3		Bromley	BR1 3FZ	2nd visit...	2014/6/17

# Feature Extraction

Avg\_crime\_n = Average of crime occurred in a month  
(during 6 months time frame)

Hotel Name	Hotel Address	Latitude_x	Longitude_x
1 Lexham Gardens	Lexham Gardens	51.4951	-0.190451
11 Cadogan Gardens	Cadogan Gardens	51.4935210	-0.1595370
130 Queen's Gate	Queen's Gate	51.4945545	-0.1784508
196 Bishopsgate	196 Bishopsgate	51.5179486	-0.0796542
37 Collingham Place Lo...	Collingham Place	51.4937606	-0.1900999
44 Curzon Street by Ma...	Curzon Street	51.5067241	-0.1466451
45 Park Lane - Dorches...	Park Lane	51.5063709	-0.1515359
47 Park Street - Grand...	Park Street	51.5109406	-0.1550038
47 Prince's Gardens Gu...	Prince's Gardens	51.4994810	-0.1721698
51 Kensington Court Li...	Kensington Court	51.5019	-0.188234

Latitude	Longitude	location	avg_crime_n
51.476836	0.129202	Abbott'S Walk	1.50
51.504915	0.118104	Abbotts Close	2.00
51.614641	0.005168	Abbotts Crescent	1.33
51.568199	-0.314136	Abbotts Drive	1.00
51.57209	-0.005465	Abbotts Park Road	2.17

Map location using text mining  
Compute Geo-distance between Hotel & Crime Scene

Hotel Name	Hotel Address	Latitude_x	Longitude_x	Latitude_y	Longitude_y	location	avg_crime_n	distance
1 Lexham Gardens	Lexham Gardens	51.4951	-0.190451	51.4959	-0.192421	Lexham Gard...	2.25	0.161633
11 Cadogan Gardens	Cadogan Gardens	51.4935210	-0.1595370	51.4929	-0.159782	Cadogan Gar...	1.33333	0.0763515
130 Queen's Gate	Queen's Gate	51.4945545	-0.1784508	nan	nan	nan	0	nan
196 Bishopsgate	196 Bishopsgate	51.5179486	-0.0796542	nan	nan	nan	0	nan
37 Collingham Place Lo...	Collingham Place	51.4937606	-0.1900999	51.4939	-0.189544	Collingham ...	2	0.0432589
44 Curzon Street by Ma...	Curzon Street	51.5067241	-0.1466451	51.5067	-0.148028	Curzon Stre...	1.4	0.0960698
45 Park Lane - Dorches...	Park Lane	51.5063709	-0.1515359	51.3584	-0.218407	Park Lane	1	17.1088
47 Park Street - Grand...	Park Street	51.5109406	-0.1550038	51.3735	-0.098394	Park Street	9.66667	15.7923
47 Prince's Gardens Gu...	Prince's Gardens	51.4994810	-0.1721698	nan	nan	nan	0	nan
51 Kensington Court Li...	Kensington Court	51.5019	-0.188234	51.5013	-0.187856	Kensington ...	1.5	0.0723952



# Final Table

		X1	X2	X3	Y
Hotel Name	Hotel Address	Effective Star Rating	ratio_negreviews	distance	avg_crime_n
1 Lexham Gardens	Lexham Gardens	3.11905	0.152174	0.161633	2.25
11 Cadogan Gardens	Cadogan Gardens	3.95495	0.157143	0.0763515	1.33333
130 Queen's Gate	Queen's Gate	4	0.122807	nan	0
196 Bishopsgate	196 Bishopsgate	3.88889	0.192857	nan	0
37 Collingham Place Lo...	Collingham Place	2.42	0.428571	0.0432589	2
44 Curzon Street by Ma...	Curzon Street	3.71739	0.102804	0.0960698	1.4
45 Park Lane - Dorches...	Park Lane	4.04412	0.166667	17.1088	1
47 Park Street - Grand...	Park Street	4.66667	0	15.7923	9.66667
47 Prince's Gardens Gu...	Prince's Gardens	3.91304	0.289474	nan	0
51 Kensington Court Li...	Kensington Court	3.68571	0.24	0.0723952	1.5

Effective Star Rating = Avg(Eff. Per hotel)

Ratio Negative Reviews = #neg\_reviews/#all\_reviews

# Preparation for Analysis

- **Fraud Detection:** Trying to eliminate users who gave negative\_reviews for all\_reviews
  - Not high enough to be fraud
- **Missing values:** are replaced by mean of their columns

```
In [10]: Counter(fraud_users["all_reviews"])  
Out[10]: Counter({1: 15186, 2: 372, 3: 10, 4: 2})
```

# Correlation Analysis & OLS Regression

Y

Index	Effective Star Rating	ratio_negreviews	avg_crime_n	distance
Effective Star Rating	1	-0.577021	0.0163366	0.069958
ratio_negreviews	-0.577021	1	-0.0433527	-0.047953
avg_crime_n	0.0163366	-0.0433527	1	-0.00897664
distance	0.069958	-0.047953	-0.00897664	1

OLS Regression Results

Dep. Variable:

avg\_crime\_n

R-squared:

0.002

Model:

OLS

Adj. R-squared:

-0.000

Method:

Least Squares

F-statistic:

0.8265

Date:

Mon, 10 Dec 2018

Prob (F-statistic):

0.479

Time:

11:41:23

Log-Likelihood:

-2806.7

No. Observations:

1196

AIC:

5621.

Df Residuals:

1192

BIC:

5642.

Df Model:

3

Covariance Type:

nonrobust

P-value < 0.05?

	coef	std err	t	P> t	[0.025	0.975]
const	2.6279	0.758	3.467	0.001	1.141	4.115
Effective Star Rating	-0.0611	0.178	-0.343	0.732	-0.411	0.289
ratio_negreviews	-1.4480	1.009	-1.436	0.151	-3.427	0.531
distance	-2.261e-05	6.94e-05	-0.326	0.745	-0.000	0.000
Omnibus:	964.458	Durbin-Watson:	1.920			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	22842.646			
Skew:	3.622	Prob(JB):	0.00			
Kurtosis:	23.147	Cond. No.	1.75e+04			

P-value < 0.05?

Next Step:

- Visualization on Map?
- Select new features?

$Y = \text{const} + (w1 * \text{EFFrating}) + (w2 * \text{NegRatio}) + (w3 * \text{Distance}) + \text{error}$   
(No variable is significant at 5% of significance level except constant)