

HOTELS REVIEW REVIEWED

Sentiment Analysis for Hotel Reviews

Group 9

WHAT WE OFFER ?

WHO WE OFFER TO?

- Common Mass, trying to book tickets for best-fit hotels according to filters and preferences.
- Currently Existing Booking Platforms, trying to increase their conversion rate by showing reviews and ratings to users.
- Hotel Managers, trying to get to know users' views about their hotel.

WHO WE OFFER TO?

- Common Mass, trying to book tickets for best-fit hotels according to filters and preferences.
- **Currently Existing Booking Platforms, trying to increase their conversion rate by showing reviews and ratings to users.**
- Hotel Managers, trying to get to know users' views about their hotel.

WHO WE OFFER TO: BOOKING PLATFORMS

- Total rating – 7.6
- Rank in “Andheri East” – 16
- 102 reviews
- 30 mentions about food, food rating – 9.1
- Very positive food reviews – [“Loved the buffet”, The food and service was awesome” ...

WHO WE OFFER TO?

- Common Mass, trying to book tickets for best-fit hotels according to filters and preferences.
- Currently Existing Booking Platforms, trying to increase their conversion rate by showing reviews and ratings to users.
- **Hotel Managers, trying to get to know users' views about their hotel.**

HOW DID WE?

STEPS

- Data Acquisition System
- Sentiment Analysis
- Score Methodology
- Feature-Aspect Based Analysis
- Score Normalisation

DATA ACQUISITION SYSTEM

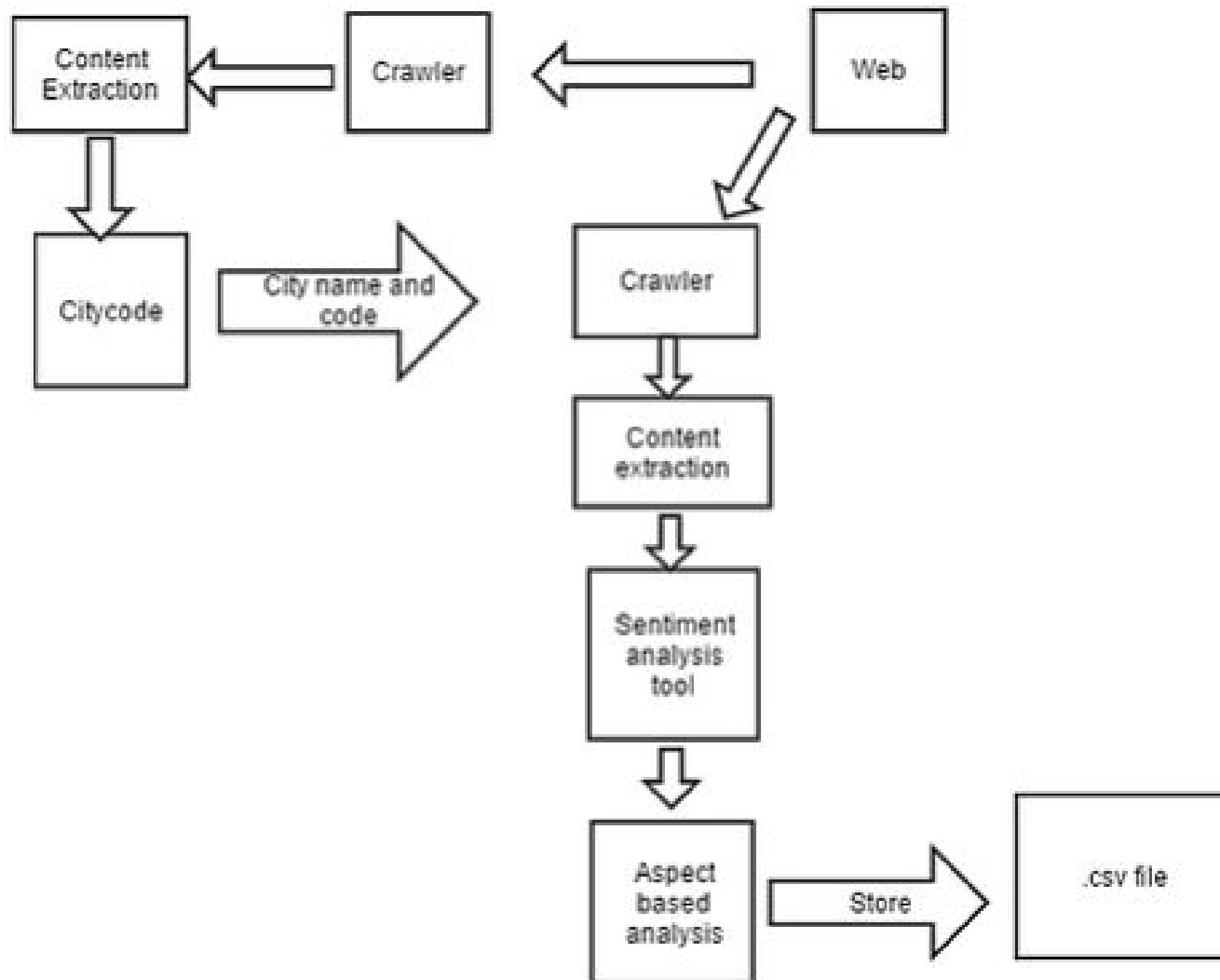
- Software used: SCRAPY
- Websites crawled: expedia.co.in, tripadvisor.in, holidayiq.com

STEPS

- Data Acquisition System
- **Sentiment Analysis**
- Score Methodology
- Feature-Aspect Based Analysis
- Score Normalisation

SENTIMENT ANALYSIS

- Software used: Stanford's coreNLP
- Algorithm used: Recursive Neural Tensor Network



STEPS

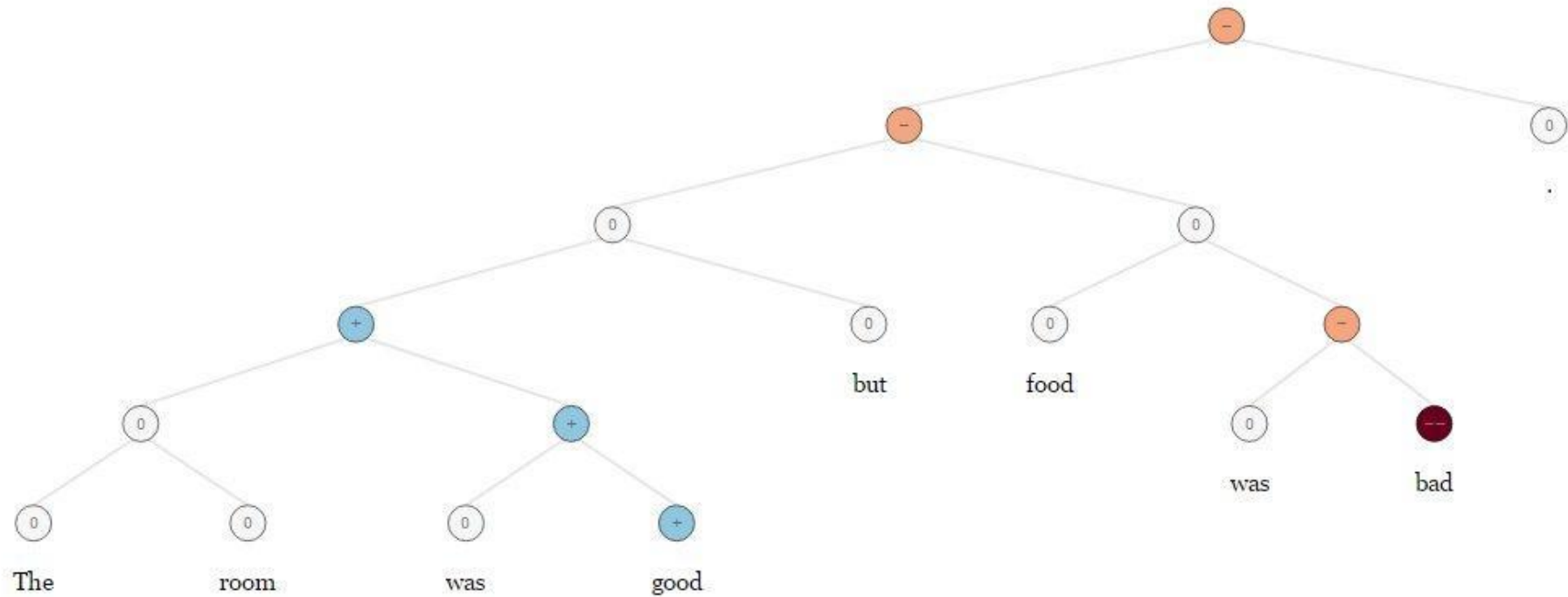
- Data Acquisition System
- Sentiment Analysis
- **Score Methodology**
- Feature-Aspect Based Analysis
- Score Normalisation

SENTENCES, NOT REVIEWS

SCORE METHODOLOGY

```
(0 (1 (2 (3 (4 (5 The) (6 room)) (7 (8 was) (9 good))) (10 but)) (11 (12 food) (13 (14 was) (15 bad)))) (16 .))
0: 0.1767 0.5951 0.1881 0.0267 0.0134
1: 0.1362 0.6056 0.1999 0.0441 0.0143
2: 0.0251 0.2705 0.4991 0.1894 0.0160
3: 0.0450 0.2424 0.2397 0.4222 0.0508
4: 0.0024 0.0368 0.9377 0.0220 0.0011
5: 0.0004 0.0029 0.9933 0.0029 0.0006
6: 0.0007 0.0091 0.9872 0.0023 0.0006
7: 0.0088 0.0421 0.0711 0.8142 0.0639
8: 0.0003 0.0022 0.9960 0.0009 0.0004
9: 0.0007 0.0049 0.0020 0.9874 0.0050
10: 0.0018 0.0052 0.9897 0.0016 0.0018
11: 0.0938 0.4257 0.4451 0.0276 0.0079
12: 0.0006 0.0032 0.9932 0.0016 0.0014
13: 0.2330 0.4251 0.3272 0.0071 0.0076
14: 0.0003 0.0022 0.9960 0.0009 0.0004
15: 0.8057 0.1500 0.0070 0.0025 0.0348
16: 0.0004 0.0012 0.9965 0.0013 0.0005
```

SCORE METHODOLOGY



SCORE METHODOLOGY

```
(0 (1 (2 (3 (4 (5 The) (6 room)) (7 (8 was) (9 good))) (10 but)) (11 (12 food) (13 (14 was) (15 bad)))) (16 .))
0: 0.1767 0.5951 0.1881 0.0267 0.0134
1: 0.1362 0.6056 0.1999 0.0441 0.0143
2: 0.0251 0.2705 0.4991 0.1894 0.0160
3: 0.0450 0.2424 0.2397 0.4222 0.0508
4: 0.0024 0.0368 0.9377 0.0220 0.0011
5: 0.0004 0.0029 0.9933 0.0029 0.0006
6: 0.0007 0.0091 0.9872 0.0023 0.0006
7: 0.0088 0.0421 0.0711 0.8142 0.0639
8: 0.0003 0.0022 0.9960 0.0009 0.0004
9: 0.0007 0.0049 0.0020 0.9874 0.0050
10: 0.0018 0.0052 0.9897 0.0016 0.0018
11: 0.0938 0.4257 0.4451 0.0276 0.0079
12: 0.0006 0.0032 0.9932 0.0016 0.0014
13: 0.2330 0.4251 0.3272 0.0071 0.0076
14: 0.0003 0.0022 0.9960 0.0009 0.0004
15: 0.8057 0.1500 0.0070 0.0025 0.0348
16: 0.0004 0.0012 0.9965 0.0013 0.0005
```

FEATURE-ASPECT BASED ANALYSIS

FEATURE-ASPECT BASED ANALYSIS

- General rating
- Food rating
- Room rating
- Service rating
- Location rating
- Cleanliness rating

FEATURE-ASPECT BASED ANALYSIS

Seed array for location: ["location", "position", "place", "property", "situation", "site", "locality", "local", "spot", "whereabouts", "whereabout", "scene", "setting", "area", "environment", "venue", "address", "view", "sight", "beach"]

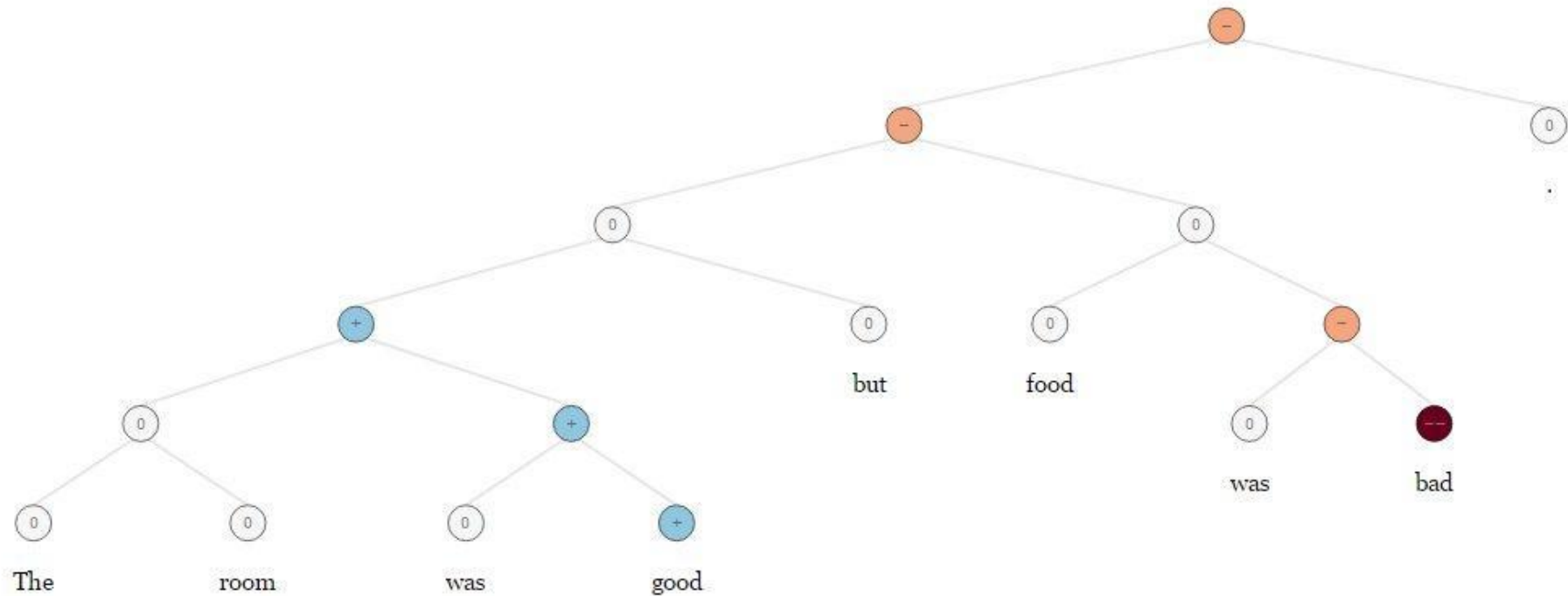
FEATURE-ASPECT BASED ANALYSIS

“I had a wonderful meal at Hotel Oberoi, I
loved the puddings.”

FEATURE-ASPECT BASED ANALYSIS

“The room was good but the food was very bad.”

FEATURE-ASPECT BASED ANALYSIS



FEATURE-ASPECT BASED ANALYSIS

```
(0 (1 (2 (3 (4 (5 The) (6 room)) (7 (8 was) (9 good))) (10 but)) (11 (12 food) (13 (14 was) (15 bad)))) (16 .))
0: 0.1767 0.5951 0.1881 0.0267 0.0134
1: 0.1362 0.6056 0.1999 0.0441 0.0143
2: 0.0251 0.2705 0.4991 0.1894 0.0160
3: 0.0450 0.2424 0.2397 0.4222 0.0508
4: 0.0024 0.0368 0.9377 0.0220 0.0011
5: 0.0004 0.0029 0.9933 0.0029 0.0006
6: 0.0007 0.0091 0.9872 0.0023 0.0006
7: 0.0088 0.0421 0.0711 0.8142 0.0639
8: 0.0003 0.0022 0.9960 0.0009 0.0004
9: 0.0007 0.0049 0.0020 0.9874 0.0050
10: 0.0018 0.0052 0.9897 0.0016 0.0018
11: 0.0938 0.4257 0.4451 0.0276 0.0079
12: 0.0006 0.0032 0.9932 0.0016 0.0014
13: 0.2330 0.4251 0.3272 0.0071 0.0076
14: 0.0003 0.0022 0.9960 0.0009 0.0004
15: 0.8057 0.1500 0.0070 0.0025 0.0348
16: 0.0004 0.0012 0.9965 0.0013 0.0005
```


STEPS

- Data Acquisition System
- Sentiment Analysis
- Score Methodology
- Feature-Aspect Based Analysis
- **Score Normalisation**

SCORE NORMALISATION: THE NEED

Mean	Median	Standard Deviation
5.73	5.75	0.49

SCORE NORMALISATION: THE SOLUTION

- Simply don't consider objective reviews.
- We defined the range of 5.0 to 6.0
- If a review lies in this range, don't consider the review for overall rating.
- Applied only for general rating

SCORE NORMALISATION: THE OUTPUT

Mean	Median	Standard Deviation
6.08	6.30	0.91

OKAY, HOW TO USE IT ?

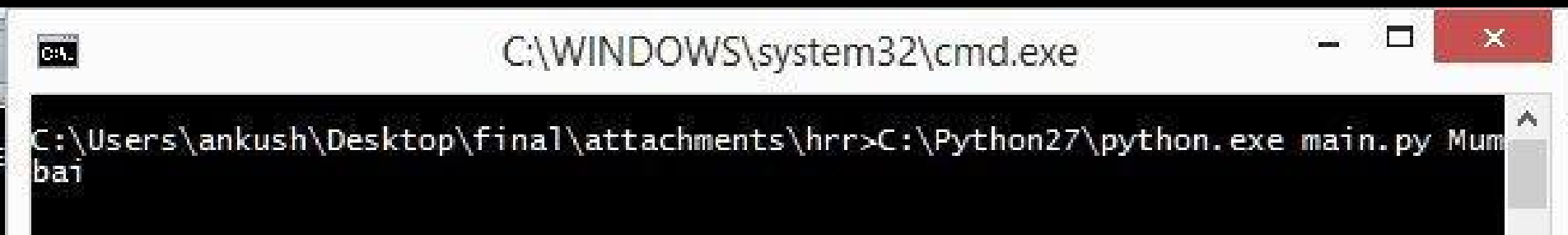
OKAY, HOW DO WE USE IT ?

- Command Line Interface
- Web-based simple interface
- Web-based advanced interface

COMMAND LINE INTERFACE

- Code written in PYTHON
- Works in windows
- Single command to roll

COMMAND LINE INTERFACE



A screenshot of a Windows Command Prompt window. The title bar shows the path `C:\WINDOWS\system32\cmd.exe`. The command prompt shows the current directory as `C:\Users\ankush\Desktop\final\attachments\hrr`. The command being executed is `C:\Python27\python.exe main.py Mumbai`. The output of the command is `bai`.

```
C:\WINDOWS\system32\cmd.exe
C:\Users\ankush\Desktop\final\attachments\hrr>C:\Python27\python.exe main.py Mumbai
bai
```


COMMAND LINE INTERFACE

- Files generated in \$city_name directory
 - review.json
 - rate.csv
 - text.txt

OKAY, HOW DO WE USE IT ?

- Command Line Interface
- **Web-based minimal interface**
- Web-based advanced interface

Hotels Reviews Reviewed

Select city name and format

I want rating for mumbai
hotels in csv format.



DOWNLOAD

Hotels Reviews Reviewed

Select city name and

bangalore

delhi

goa



kolkata

shimla

I want rating for
hotels in csv for



DOWNLOAD

Hotels Reviews Reviewed

Select city name and format

I want rating for goa

hotels in format

CSV ✓

json



DOWNLOAD

WEB-BASED MINIMAL INTERFACE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	title	location	region	stars	amenities	start price	end price	total norm	total ratin	total coun	total very	total posit	total nega	total very	total very	total posit	total nega	total very
2	Hotel Airp	Twin Arca	Military R	NA		2468	4010	3.34585	5.32326	10	0	4	6	0 []		['recieved	['stayed th	[]
3	Pearl Resi	14, Usha K	NA	2	Free Parki	1419	2036	3.407623	4.491049	23	0	5	13	5 []		['and the i	['they doe	['most hor
4	iHOTEL	90 Feet Rc	Andheri E	2	Room Ser	1357	2098	3.451417	3.865683	3	0	0	2	1 []		[]	['but appa	['the dusb
5	Orange In	Sai Baba E	Makhwan	2.5	Room Ser	1481	2221	3.781988	4.575845	25	0	5	20	0 []		['in the en	['we went	[]
6	Skyway In	Premises	Andheri /	2.5	Budget, Fr	987	2591	4.106135	4.711902	16	1	3	10	2	['food was	['and of cc	['most of t	['this was
7	Ashwin	Marol-Ma	Andheri (NA	Restauran	2776	5244	4.313463	4.9319	28	0	9	18	1 []		['lots of tr	['plaster c	['heres a b
8	Marol Res	Opp Seve	Andheri	2.5	Room Ser	1234	1604	4.665568	5.038583	15	0	5	8	2 []		['positive!	['the awfu	['it was su
9	Hotel Sate	213/214, C	Vile Parle	2.5	Room Ser	2098	2961	4.7149	5.016467	18	0	4	14	0 []		['pm', 'the	['not so fri	[]
10	Hotel Arm	Rani Arcac	Sakinaka	1.5	Room Ser	1851	3146	4.72275	5.161535	18	1	5	12	0	['for coupl	['room ser	['the area	[]
11	Hotel Cosi	Seepz Poc	Andheri E	2.5	Room Ser	1172	3640	4.756035	5.06341	20	0	9	10	1 []		['very sma	['seriously	['worst ex
12	Hotel Arm	1st Floor,	Bandra (E	1	Room Ser	1851	2838	5.015393	5.350264	23	0	11	12	0 []		['entrance	['no eleva	[]
13	Hotel Plar	Plot No 12	Near Garv	3.5	Room Ser	3887	6601	5.133436	5.286093	30	2	10	16	2	['the front	['on th', 'tl	['first imp	['the wors
14	Hotel Om	Bhosle pre	Andheri E	2	Room Ser	1913	3455	5.257871	5.338051	36	2	16	15	3	['very nice	['there is',	['the bed :	['im all for
15	The LaLiT	Sahar Airp	Andheri E	5	Luxury, Pc	7588	10426	5.280645	5.255484	38	2	12	23	1	['the lobb	['take care	['this mak	['really tha
16	KC Reside	Maharash	Marol	2		1542	2036	5.301325	5.160131	24	1	8	15	0	['this hote	['close to	['it made r	[]
17	Hotel Cre:	Near Asha	Andheri C	2.5	Room Ser	1851	3085	5.391619	5.47605	22	2	9	9	2	['room ser	['once', 'h	['the stan	['waste of
18	Eden Gue:	2- A Marol	Opposite	NA		1110	1789	5.460609	5.377027	26	3	7	14	2	['really en	['its very c	['youll prc	['very bad
19	Hotel Ori	Near Dom	Andheri (3.5	Free Breal	3948	4936	5.511328	5.535865	24	1	10	13	0	['around c	['the room	['i stayed	[]
20	Savoy Suit	Saki Vihar	NA	3	Restauran	2591	3702	5.541039	5.506932	27	1	9	16	1	['room ser	['will defi	['the chec	['it was th
21	Hotel Soli	New Sanje	Opp. APT	3	Restauran	4750	5676	5.557923	5.493025	57	4	25	26	2	['rooms sp	['i stayed	['the hote	['mee and
22	Aifa Resid	Andheri A	Near Soci	2.5	Budget, Fr	1542	1974	5.57338	5.526557	22	1	11	10	0	['so far , s	['rooms w	['the trou	[]
23	The Regal	Plot No 31	Plot No 31	4	Room Ser	4257	5861	5.687204	5.720849	18	0	10	8	0 []		['keep it u	['been sta	[]
24	Hotel JK R	30, Old Na	Near Sai S	2.5	Room Ser	2221	4134	5.687874	5.501193	36	4	12	20	0	['manager	['also pret	['the staff	[]
25	Hotel Roy.	Patel Com	Sakinaka	2.5	Budget, Fr	2468	4010	5.698211	5.521563	29	0	13	15	1 []		['it provid	['however	['worst ex

WEB-BASED MINIMAL INTERFACE

S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
location rating	location cour	location v	location p	location n	location v	location v	location p	location n	location v	room ratir	room cour	room very	room posi	room neg	room very	room very	room posi
NA	0	0	0	0	0	0	0	0	0	5.514075	2	0	1	1	0	0	['the hote
NA	0	0	0	0	0	0	0	0	0	NA	0	0	0	0	0	0	0
NA	0	0	0	0	0	0	0	0	0	4.586838	2	0	1	1	0	0	['actually i
5.678083333	3	0	3	0	0	0	['the locat	0	0	5.4837	1	0	0	1	0	0	0
NA	0	0	0	0	0	0	0	0	0	4.423269	4	0	0	3	1	0	0
5.04685	1	0	0	1	0	0	0	['hotel ask	0	5.302633	3	0	1	2	0	0	['the room
5.526325	1	0	1	0	0	0	['the locat	0	0	NA	0	0	0	0	0	0	0
4.966975	4	0	2	2	0	0	['location	['the hote	0	6.170233	3	0	3	0	0	0	['dirty sink
5.538825	1	0	1	0	0	0	['the area	0	0	4.764108	3	0	0	3	0	0	0
4.2381375	2	0	1	0	1	0	['the locat	0	['worst ex	5.4837	1	0	0	1	0	0	0
NA	0	0	0	0	0	0	0	0	0	5.341025	1	0	0	1	0	0	0
NA	0	0	0	0	0	0	0	0	0	4.880288	4	0	1	3	0	0	['the front
NA	0	0	0	0	0	0	0	0	0	5.904485	5	1	2	2	0	['staff was	['each staf
7.554925	1	0	1	0	0	0	['the food	0	0	5.564381	8	0	4	4	0	0	['rooms ar
5.586525	1	0	1	0	0	0	['the staff	0	0	5.536563	2	0	1	1	0	0	['the hote
6.227841667	3	0	2	1	0	0	['hotel wa	['hotel is r	0	4.34705	2	0	0	1	1	0	0
NA	0	0	0	0	0	0	0	0	0	5.50935	2	0	2	0	0	0	['rooms w
5.565025	1	0	1	0	0	0	['location	0	0	4.925769	4	0	3	0	1	0	['the room
5.543975	1	0	1	0	0	0	['location	0	0	5.496525	4	0	2	2	0	0	['rooms ar
6.78805	3	1	2	0	0	['hotel is l	['location	0	0	5.192432	7	0	5	2	0	0	['com con
5.1525875	2	0	1	1	0	0	['location	['no view	0	5.40518	5	0	3	2	0	0	['rooms w
NA	0	0	0	0	0	0	0	0	0	5.820125	2	0	1	1	0	0	['service v
5.51405	1	0	1	0	0	0	['the view	0	0	5.268828	10	0	6	3	1	0	['the room
5.543975	1	0	1	0	0	0	['location	0	0	5.908181	4	0	3	1	0	0	['the hote

OKAY, HOW DO WE USE IT ?

- Command Line Interface
- Web-based minimal interface
- **Web-based advanced interface**

WEB-BASED ADVANCED INTERFACE

Show 100 entries

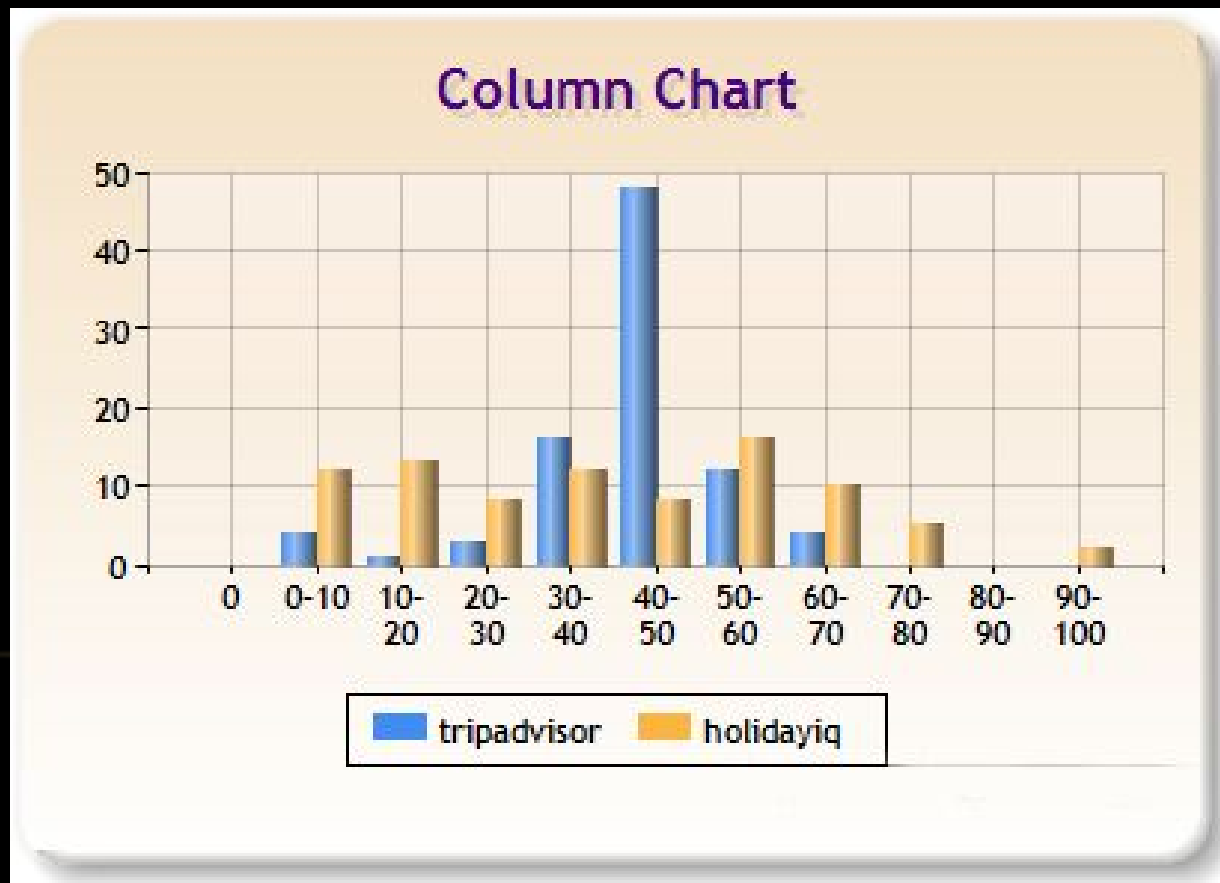
Search: Andheri

Hotel	location	region	stars	amenities	start price	end price	rating	location rating	room rating	food rating	service rating	cleanliness rating
New Shahana	Shop No-5, Marol Maroshi road, Andheri, 400059, India	Andheri	2.5	Restaurant, Bar/Lounge, Room Service, Free High Speed Internet (WiFi)	1049	1913	6.028352272727273	6.44055	5.5239625	NA	7.0603125	NA
Hotel Cosmo	Seepz Pocket Cross Road, Near Vijay Nagar Flyover Bridge, Marol Maroshi, Andheri East, 400059, India	Andheri East	2.5	Room Service, Free Parking, Restaurant, Free High Speed Internet (WiFi)	1172	3640	4.873966666666666	4.238137500000001	5.4837	NA	NA	7.697225
Marol Residency	Opp Seven Hills Hospital near Marol Maroshi Bus Stop, Andheri, 400059, India	Andheri	2.5	Room Service, Free Parking, Free High Speed Internet (WiFi)	1234	1604	4.922647916666667	5.526325000000001	NA	NA	NA	NA
iHOTEL	90 Feet Road, Near Saki Naka Junction, Andheri East, 400072, India	Andheri East	2	Room Service, Free Parking, Free Breakfast, Free High Speed Internet (WiFi)	1357	2098	3.8159375000000004	NA	4.5868375	NA	NA	5.18815

WEB-BASED ADVANCED INTERFACE

Hotel	location	region	stars	amenities	start price	end price	rating	location rating	room rating	food rating	service rating	cleanliness rating
Ascot Inn	38 Garden Road,, Colaba, 400039, India	Colaba	3	Room Service, Restaurant, Free High Speed Internet (WiFi)	8205	9378	7.304580434782609	8.242525	6.8199125	7.016275	5.978350000000001	7.370383333333334
The Residency	At The Acres Club, Hemu Kalani Marg, Chembur, 400 071, India	Chembur	4	Room Service, Restaurant, Pool, Free Internet, Free Parking, Bar/Lounge, Free Breakfast, Fitness Centre with Gym / Workout Room	5923	19187	7.095364583333333	6.7002625	6.009633333333333	5.6371	5.446225000000001	7.536575
Hotel Airlink	75, Off. Nehru Road, 400099, India	Off. Nehru Road	2.5	Room Service, Free Parking, Restaurant, Free Internet, Bar/Lounge, Free Breakfast	3455	4750	7.052763068181818	5.99905	5.800825000000001	5.71075	6.50995	6.692233333333334
				Vile Parle East, Luxury								

VISUALISATION



Percentage of hotels for each amount of reviews interval.

Website	Total reviews	Positive reviews	Negative reviews	Neutral reviews
holidayiq	3334	1583	698	1053
Tripadvisor	3788	1171	1195	1422

The segments were classified with respect to their polarity, including the neutral polarity.

WHAT'S IN THE FUTURE ?

WHAT'S IN THE FUTURE ?

- Social information gathering- from twitter/facebook
- Sarcasm

WHAT'S IN THE FUTURE ?

- Our immediate plan would be to scale for more data.
- For hotel managers, we would be adding visualised results with charts and trend-lines displaying how the hotel is doing with time.

WHAT'S IN THE FUTURE ?

- To get more accurate results, we will train the sentiment tree bank with specific hotel reviews.
- We would add our solution to read smiles/emoticons, images and different form of texts present in reviews.

THIS IS IT.

Thank You