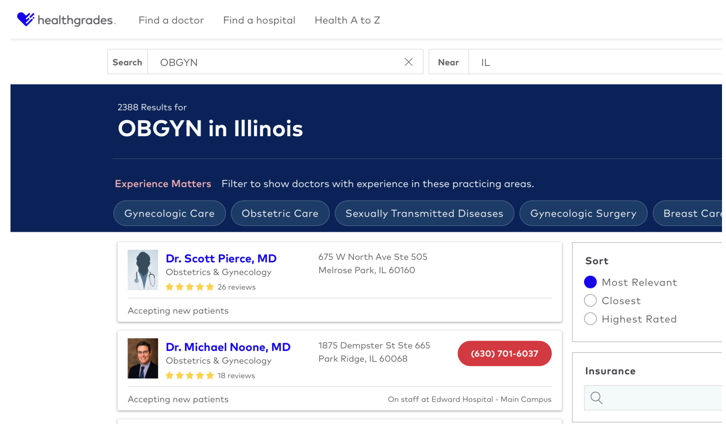
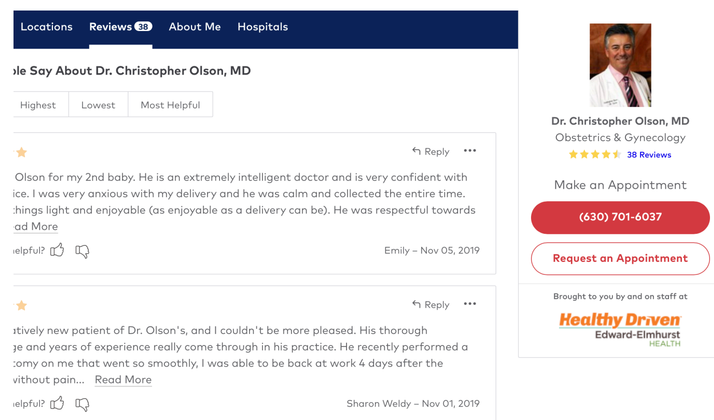
ONLINE PHYSICIAN REVIEWS- TEXT MINING

Varun Maheshwari

Web Scraping – Healthgrades

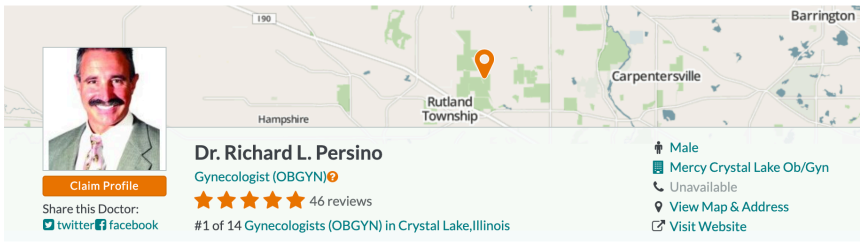
- Selenium API used along with Chrome driver in Python
- Reviews scraped for OBGYN doctors in state of Illinois for 2388 doctors
- Doctors having less than 2 reviews were removed
- Final – 841 doctors with over 6000 reviews
- Demographics scraped – Age, Gender, Rating, Affiliated Hospitals, University, Experience in Years





Web Scraping – RateMDs

- Used Selenium API in Python
- Removed all the Doctors with less than 2 reviews
- Reviews for 891 doctors were scraped
- Doctor Names, Gender, Doctor Reviews, Doctor Ratings, Hospital Affiliations, Experience Years were scraped.



Doctor_names	Doctor_Genc	Doctor_reviews	Doctor_ratin	Hospital Affiliations
Dr. Jennifer M. Ozan	Female	["I love this office and I love Jennifer Ozan! It's always clean, and they are qui	4.97	Jennifer M. Ozan Clinic Evanston
Dr. Carlos Sandoval-Herrera	Male	['Dr. Sandoval is the best. He also has a great staff. They were all very helpful.	4.94	MOUNT SINAI
Dr. Richard L. Persino	Male	['Dr Persino is my favorite doctor! He is always so happy and nice and a great	4.78	Mercy Crystal Lake Ob/Gyn
Dr. Gail D. Miller	Female	['I came for an ultrasound and my experience was excellent! Doc is always ver	4.66	Miller Gail D MD
Dr. Lori C. Leipold	Female	['The nicest, most caring, honest, compassionate Ob/Gyn.', '5 stars plus!!!!! F	4.73	Leipold Lori MD
Dr. Thomas M. Kazmierczak	Male	['Dr Kaz is one of the best doctors I have met. As a nurse, I notice particular th	5	Thomas M. Kazmierczak Clinic Minooka
Dr. Mary Jane A. Nowak	Female	['I am so lucky to have found a dr as nice as her. Shes extremely kind in explai	5	Mary Jane A. Nowak Clinic Oak Lawn

Buckets – Themes Explained

Buckets	Positive	Negative
Bedside Manners	Friendly/Postive Behaviour,Bonding, General Information sharing(listening/explaining), Substantial Visit time, Good Pyshcological Support	Rude/Uncaring attitude,bonding, behaviour,poor explainations, little time spent with patients, Does not address stress/anxiety
Waiting Time	Short time spent waiting to see doctor at Waiting room	Excessive waiting time spent to see doctor at waiting room, Delays
Ease of Appointment	Flexible and easy scheduling of appointment	Hard/Inconvinient scheduling of appointment
Office Environment	Clean clinic, good parking/other facilities available, Hygiene, Location	Unclean enviroment, insufficient patient facilities, Reachability
Office Staff	Supportive and good mannerism, Reachability of staff	Rude/Unsupportive behaviour, Unavailability of staff
Medical Expertise	Effective Treatment, Correct Diagnosis, Best use of tests/surgery, Clinical decision-making, Treatment plan	Ineffective treatment, Misdiagnosis, Unnecesary tests/failed surgeries, Unorganised treatment plan
Costs/Expenses	Inexpensive, Hassel-free Billing, Reimbursements	Expensive, Complex billing, overhead charges, high copay

Snapshot of words in each Bucket

Bedside Manners	Waiting Time	Ease of Appointment	Office Environment	Staff	Medical Expertise	Costs/Expenses
manner	delay	voicemail	administrative	assistant	surgical	amount
accomodating	early	appointment	water	manager	delivery	bill
interaction	fast	urgent	atmosphere	attending	abrasive	business
care	hour	appt	washroom	counter	fibroids	cash
advocate	hours	available	building	nurse	aggressive	charge
personality	late	booked	center	coworkers	misdiagnosed	claim
condescending	long	busy	city	customer	knowledgeable	copay
rushes	minute	called	clean	reception	approach	costly
professional	minutes	calls	rooms	receptionist	biopsy	costs
comforting	overbook	cancel	clinic	desk	csection	coverage
answers	overbooked	cancelled	department	employee	surgery	dollars
trust	quick	confirmation	dirty	front	examination	expensive
approachable	room	ease	dog	secretary	hysterectomy	fees
guidance	short	forms	drive	ladies	judgement	greedy
assurance	timely	phone	experiences	lady	skilled	insurance
attention	visit	vacation	facility	technician	cesarean	money
clarify	visited	reschedule	hospital	service	treatment	overcash
demeanor	visits	rescheduling	location	staff	laparoscopic	overcharged
unethical	wait	returned	office	paperwork	diagnosis	payment
encouraging	waited	schedule	parking	personnel	expertise	reimburse

Splitting of Reviews

Reviews were separated into smaller phrases based on “.”, “,” “!”, “?”, “and” & “but” using a split function.

For Example: “He is a little older, so he is old-school on some things, but in all makes sure that you understand what is going on.”, “This doctor is very rude and has a terrible bedside manner. He pushes his appoints back sometimes as much as 2 hours!”, “Dr. Ruby is very knowledgeable about gyne issues.

Phrase 1: He is a little older	**Phrase 2:** so he is old-school on some things	**Phrase 3:** in all makes sure that you understand what is going on	**Phrase 4:** This doctor is very rude	**Phrase 5:** has a terrible bedside manner	**Phrase 6:** He pushes his appoints back sometimes as much as 2 hours	**Phrase 7:** Dr. Ruby is very knowledgeable about gyne issues

Identifying sentiment of each phrase

- Using the open source sentiment analysis algorithm called TextBlob, sentiment of each phrase was determined.
- A compound score of <0 will be negative
- > 0 will be positive
- 0 will be Neutral
- All the Neutral sentences will be removed

Phrase	Sentiment
He is a little older	Negative
so he is old-school on some things on	Negative
in all makes sure that you understand what is going	Positive
This doctor is very rude	Negative
has a terrible bedside manner	Negative
He pushes his appoints back sometimes as much as 2 hours	Negative
Dr. Ruby is very knowledgeable about gyne issues	Positive

Data Pre-processing before Tokenization

- Removed all the special characters
- Removed Stop words using inbuilt NLTK stop words list
- Removed all the numbers
- Converted all the words to lower case
- Removed Doctor’s names and frequently occurring neutral words such as Dr., Doctor, etc.
- Created Unigram and bigrams from the reviews

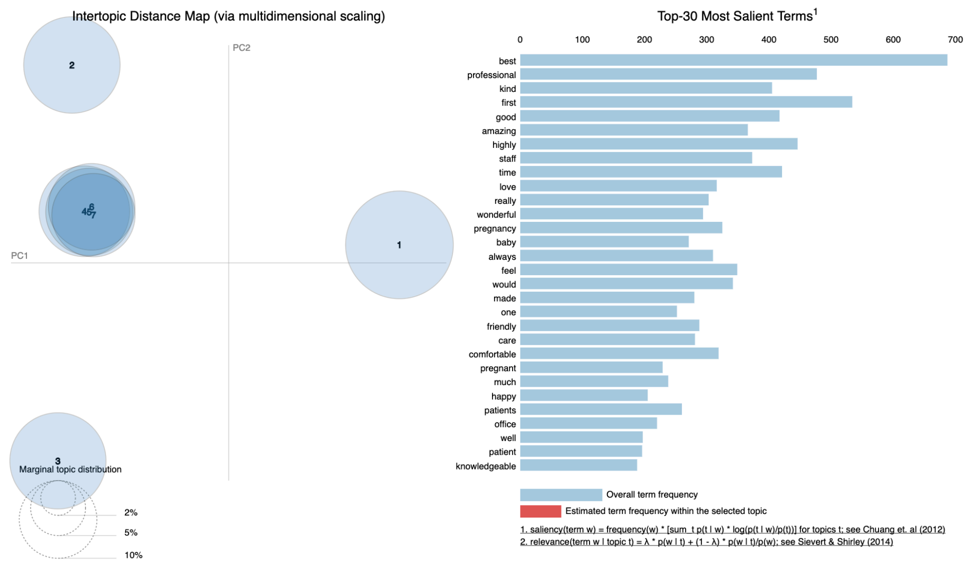
Tokenization & Lemmatization

- Each pre-processed phrase was converted into tokens using NLTK Tokenization function.
- Each of these tokens were then lemmatized using using Lemmatizer under NLTK.

Pre-processed phrase	Upon Tokenization	Upon Lemmatization
little old	[“little”, “old”]	[“little”, “old”]
oldschool some things	[“oldschool”, “some”, “things”]	[“oldschool”, “some”, “things”]
makes sure understand going	[“makes”, “sure”, “understand”, “going”]	“makes”, “sure”, “understand”, “going”]
rude	[“rude”]	[“rude”]
bedside manner	[“bedside”, “manner”]	[“bedside”, “manner”]
pushes appoints back sometimes hours	[“pushes”, ”appoints”, “back”, “sometimes”, “hours”]	[“push”, “appoint”, “back”, ”sometime”, ”hour”]
knowledgeable gyne issues	[“knowledgeable” “gyne”, “issues”]	[“knowledge”, “gyne”, “issue”]

Topic Modelling

- Performed Unsupervised LDA using the reviews
- Found that few of the topics are overlapping i.e., there are few words that are common for the topics 4, 5, 6 & 7
- Therefore, we understood that we should move to a semi-supervised LDA model.



Topic Modelling – Anchored LDA

- By using, Anchored latent dirichlet allocation method in Corex, each of the phrases separated was classified into different topics.
- The bag of words for each topic was used as the anchors.

```
Topic #1: feel, comfortable, feel comfortable, sure, makes, questions, made, make, made feel, make sure, makes feel,
answer, easy talk, makes sure, things, answer questions, made sure, feel like, always, talk
Topic #2: time, pregnant, get, appointment, wait, first time, weeks, felt, took, see, back, minutes, went, much, told,
takes, months, call, day, times
Topic #3: ever, best, would highly, highly, best ever, would, one best, highly recommended, recommended, anyone,
bed_side, one, best ob_gyn, highly anyone, best doctors, ob_gyn, bed_side manner, far best, obgyn, ever seen
Topic #4: first, baby, pregnancy, delivered, child, years, new, many, first pregnancy, first child, healthy, many years,
first visit, deliver, high_risk, delivered first, son, first baby, ob, delivering
Topic #5: staff, office staff, office, friendly, staff friendly, humor, sense, nice, sense humor, staff wonderful, staff
always, staff also, also, staff nice, staff amazing, good experience, always friendly, really nice, nursing, nursing
staff
Topic #6: like, better, could, really, know, going, another, said, ask, surgery, right, find, want, wish, doctors,
available, someone
Topic #7: patients, cares, care, really cares, cares patients, health, best care, listens, genuinely, care patients,
really listens, many patients, interest, blood_pressure, seems, level, best gynecologist, trust, really care, time
patients
```

Before:

```
Topic #1: knowledgeable, practice, child, delivered, surgery, health, job, issues, medical, birth, life, section, labor,
deliver, exam, check, options, test, ultrasound, pregnancies, decision, treatment, complications, expertise
Topic #2: feel, care, comfortable, patient, questions, bedside_manner, kind, compassionate, talk, willing, pleasant,
help, answer, gentle, attentive, unprofessional, concerns, manner, confident, listen, answers, explain, bed_side,
respectful, answered, informative, special
Topic #3: staff, nurse, service, nursing, team, front, dealing, assistant, staff friendly, staff wonderful, staff
always, staff also, staff nice, staff amazing, nursing staff, staff awesome, staff helpful, wonderful staff, friendly
staff, staff kind, staff well, also friendly, staff best, always friendly
Topic #4: insurance, pay, rate, paid, amount, bill, charge, billing, money, higher, lost, self mistake
Topic #5: office, far, hospital, area, clean, clinic, drive, city, offices, center, new office, far away, love office,
doctors office, moved, doctors, contact, close
Topic #6: appointment, busy, appointments, called, schedule, ease, phone, appt, calls, receptionist, scheduled, routine,
first appointment, get appointment, get, make appointment, day, call, scheduling, right_away, morning, past
Topic #7: wait, early, quick, minutes, available, room, fast, long, visits, hour, waited, minute, short, late, punctual,
worth wait, always available, wait time, wait minutes, waiting_room, hours, exam room, wait hour
```

After:

Topic Modelling – Anchored LDA

- The output of the model classified each of the phrases into the buckets scoring each topic as 1 and 0 if the phrase does not have the topic.
- Using this topic allocation, the proportion of of each topic for each doctor was calculated

Doc_names	Sentiment	texts	med_experti	bedside_mai	office_staff_	clinic_pos	clinic_envt_	ease_schedu	waiting_tim	med_experti	bedside_mai	office_staff_	costs_neg	clinic_envt_	ease_schedu	waiting_time_neg
Dr. Howard I	Negative	wait, times, arent, bad, max, ive, waited, mi	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Dr. Howard I	Negative	arof, well, every, seen, knowledgeable	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dr. Humbert	Negative	first, initial, appointment, impossible, get, h	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Dr. Humbert	Negative	even, call, times, get, results, hvg, exam	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Dr. Humbert	Negative	chloride, months, get, call, anyone, check	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Dr. Humbert	Negative	number, like, reviewers, said	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dr. Humbert	Negative	staff, horrible	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Dr. Humbert	Negative	scoccia, office, waste, time	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Dr. Abraham	Positive	best, exuiptment	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	staff, supportative	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	absolute, best, staff, wor	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	made, easy, see	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	appointment, times, flexi	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	easy, parking, illinois, ma	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	ultrasounds, right, office,	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	technology, fabulous	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Dr. Abraham	Positive	send, script, pharmacy, el	0	0	0	0	0	1	0	0	0	0	0	0	0	0

Data Creation

- As a score for each bucket, the proportion of total no. of phrases falling under each bucket was calculated.

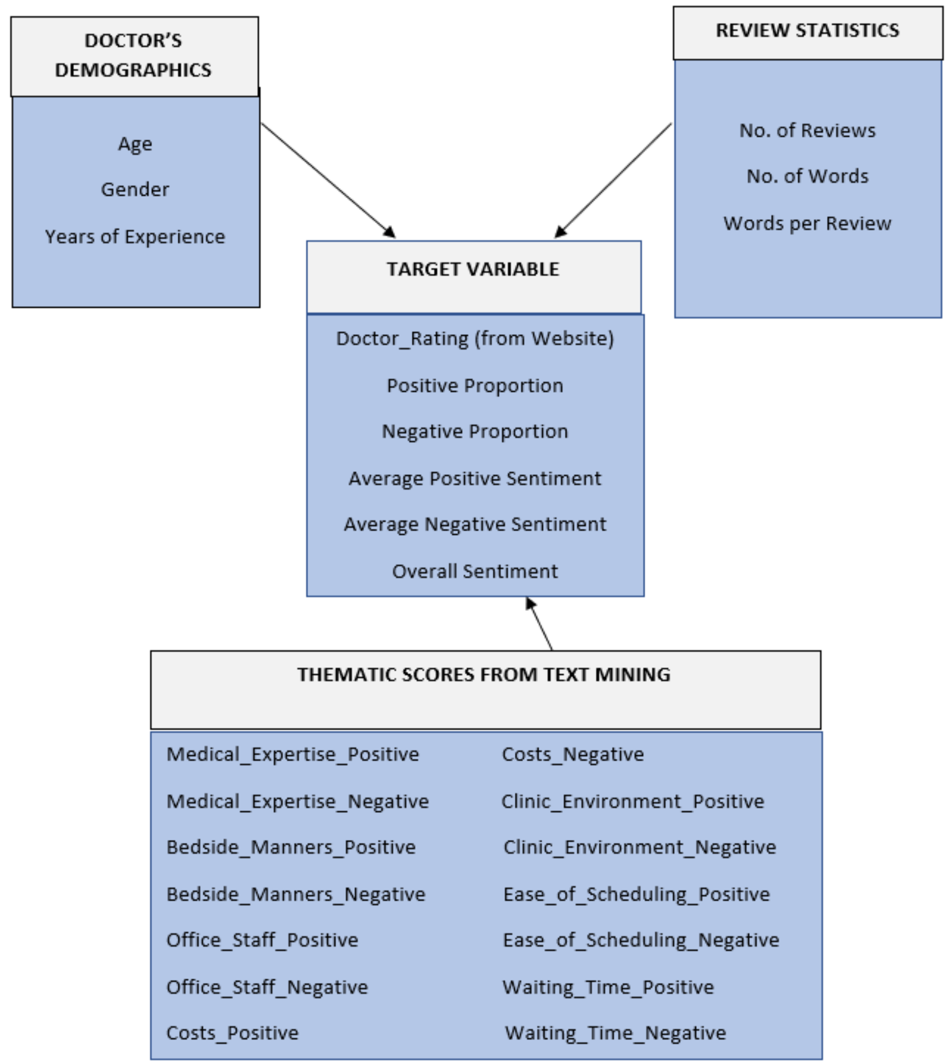
$$\frac{Sum\ of\ topic\ score\ for\ each\ buckets}{Total\ no.of\ phrases}$$

- Then, the sentiment score for each reviews was calculated and the depended variable Average Positive Sentiment score and Average Negative Sentiment score.
- Similarly, the proportion of positive and negative reviews were calculated.

DoctorName	Age	Affiliated H	Doctor_Gen	Education	Experience	Rating	Number_of_	Number_of_	Words_per_	Reviews
Dr. Aarathi	42	Edward Hos	Female	Northeast C	21	4.5	148	3	49.333333	['Dr cholkeri
Dr. Abbie Ro	51	Northweste	Female	Rush Medica	33	4	41	2	20.5	['Dr. Roth is
Dr. Abrahar	51	Advocate Ill	Male	University c	29	4.7	680	6	113.33333	["It is with g
Dr. Ada Kag	49	Blessing Hos	Female	Johns Hopki	22	4.2	40	2	20	['Dr. Kaguml
Dr. Adam Co	47	Northweste	Male	Rush Medica	40	4.6	228	5	45.6	['Amazing D
Dr. Adam Ga	44	Evanston Ho	Male	Yale Univers	25	3.8	371	7	53	["I have bee
Dr. Adam Ra	65	Adventist H	Male	Ain Shams U	41	4.4	43	2	21.5	['Would high
Dr. Adeeb A	49	Community	Male	Aleppo Mec	33	5	46	2	23	['Dr. Alshahı
Dr. Adel Har	66	Decatur Me	Male	Cairo Univer	42	3.7	125	3	41.666667	['Ok', 'Dr ha
Dr. Adriena	48	Memorial H	Female	Nova South	27	3.7	254	8	31.75	['My care w
Dr. Akemi N	48	Advocate Co	Female	Medical Coll	18	4.4	199	4	49.75	["Dr Nakanu
Dr. Akua Afi	46	Loyola Univ	Female	University C	18	5	152	3	50.666667	['Dr Afraiyie
Dr. Alan Joh	69	Alexian Brot	Male	Loyola Univ	24	4.7	428	7	61.142857	["Dr. Johnso

DoctorName	med_expert	bedside_ma	office_staff_	costs_pos	clinic_envt_	ease_schedu	waiting_tim	med_expert	bedside_ma	office_staff_	costs_neg	clinic_envt_	ease_schedu	waiting_tim	Postive_Pro	Negative_P	avg_pos_se	avg_sent_n	avg_sentneg2
Dr. Aarathi	0.31	0.54	0	0.08	0	0	0	0	0.08	0	0	0	0	0	0.4	0.2	0.6605	-0.0777778	0.0777778
Dr. Abbie Ro	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.4625	0	0
Dr. Abrahar	0.15	0.53	0.03	0.06	0	0	0.03	0.18	0.03	0	0	0	0	0	0.5121951	0.097561	0.2955224	-0.299375	0.299375
Dr. Ada Kag	0	0.75	0	0	0	0	0	0.25	0	0	0	0	0	0	1	0	0.48125	0	0
Dr. Adam Co	0.21	0.29	0.14	0.07	0	0	0	0.07	0.07	0.07	0	0	0	0.07	0.375	0.125	0.4452778	-0.275	0.275
Dr. Adam Ga	0.25	0.5	0.03	0.03	0	0	0.03	0.09	0.06	0	0	0	0	0	0.625	0.0833333	0.3538924	-0.6	0.6
Dr. Adam Ra	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.4866667	0	0
Dr. Adeeb A	0	0.33	0.33	0	0.33	0	0	0	0	0	0	0	0	0	1	0	0.709697	0	0
Dr. Adel Har	0	0.5	0	0	0	0	0	0.25	0.13	0	0	0	0	0.13	0.875	0	0.3714881	0	0
Dr. Adriena	0.2	0.4	0.13	0	0	0	0.07	0.13	0.07	0	0	0	0	0	0.6428571	0.1428571	0.5147222	-0.1667063	0.1667063
Dr. Akemi N	0	0.5	0.08	0.08	0	0	0	0	0.17	0	0	0	0.08	0.08	0.8	0.2	0.579881	-0.11625	0.11625
Dr. Akua Afi	0.14	0.43	0	0	0	0	0.14	0.14	0.14	0	0	0	0	0	0.6666667	0	0.2580588	0	0
Dr. Alan Joh	0.18	0.45	0.05	0.09	0	0	0.09	0.14	0	0	0	0	0	0	0.7142857	0	0.5477778	0	0

Conceptual Analytical Model



Exploratory Data Analysis – Univariate (Healthgrades)

- Mean ratings in Healthgrades is 3.924, median is 4 and the range is 1 through 5.
- On an average, people have used around 53 words per review, with 182 words per review as maximum
- Apart from Medical Expertise and Bedside Manners, rest all the buckets have a normal distribution
- Median of sentiment scores for all the buckets is 0 except for Medical Expertise and Bedside Manners
- Mean of Whole Sentiment score is 0.232 and median is 0.242, which means it's a normal distribution with minimum score as -0.753 and maximum as 0.875

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Rating	3.924	0.771	1.000	3.400	4.000	4.500	5.000
Number_of_Words	279.032	287.921	12	121	202	340	2,738
Number_of_Reviews	5.260	5.526	1	2	4	6	73
Words_per_review	53.210	19.318	4.000	40.272	51.500	65.333	182.000
med_expertise_pos	0.157	0.145	0.000	0.060	0.140	0.220	1.000
bedside_manners_pos	0.386	0.228	0.000	0.238	0.365	0.500	1.000
office_staff_pos	0.040	0.065	0	0	0	0.1	0
costs_pos	0.052	0.074	0.000	0.000	0.000	0.090	0.670
clinic_envt_pos	0.032	0.060	0	0	0	0.05	0
ease_schedule_pos	0.024	0.055	0	0	0	0.03	1
waiting_time_pos	0.026	0.053	0	0	0	0.04	0
med_expertise_neg	0.086	0.124	0	0	0.05	0.1	1
bedside_manners_neg	0.122	0.144	0.000	0.000	0.080	0.200	1.000
office_staff_neg	0.017	0.046	0	0	0	0	0
costs_neg	0.006	0.025	0	0	0	0	0
clinic_envt_neg	0.015	0.047	0	0	0	0	1
ease_schedule_neg	0.013	0.036	0	0	0	0	0
waiting_time_neg	0.025	0.054	0	0	0	0.03	0
Postive_Proportion	0.620	0.229	0.000	0.500	0.625	0.750	1.000
Negative_Proportion	0.166	0.180	0.000	0.000	0.125	0.250	1.000
avg_pos_sent_score	0.387	0.141	0.000	0.303	0.385	0.465	1.000
avg_sent_neg1	-0.183	0.197	-1.000	-0.282	-0.135	0.000	0.000
avg_sentneg2	0.183	0.197	0.000	0.000	0.135	0.282	1.000
Whole_Sentiment	0.232	0.175	-0.753	0.129	0.242	0.349	0.875

Exploratory Data Analysis – Univariate (RateMds)

- Mean ratings in Ratemds is 3.752, median being 3.885
- Here, maximum words used per review is 634, with minimum being 12 and average being 70
- Here, apart from the medical expertise and Bedside Manners, even Office Staff have a normal distribution
- Even here, on average doctors are reviewed positive, median being 0.202 and mean is 0.197. So, even this is a normal distribution with minimum sentiment score being -0.474 and maximum being 0.691

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Doctor_ratings	3.752	0.936	1.000	3.140	3.885	4.508	5.000
No..of.Words	517.065	623.376	33	145	293	607.8	5,036
No..of.Reviews	7.464	8.304	1	3	5	9	90
Words_per_review	70.629	43.007	12.330	43.410	64.000	86.020	634.000
med_expertise_pos	0.160	0.156	0.000	0.050	0.135	0.230	1.000
bedside_manners_pos	0.303	0.222	0.000	0.150	0.265	0.428	1.000
office_staff_pos	0.055	0.094	0	0	0	0.1	1
costs_pos	0.008	0.032	0	0	0	0	0
clinic_envt_pos	0.024	0.055	0	0	0	0.03	1
ease_schedule_pos	0.018	0.045	0	0	0	0.02	0
waiting_time_pos	0.026	0.054	0.000	0.000	0.000	0.040	0.500
med_expertise_neg	0.133	0.171	0.000	0.000	0.090	0.210	1.000
bedside_manners_neg	0.066	0.116	0.000	0.000	0.000	0.100	1.000
office_staff_neg	0.057	0.107	0.000	0.000	0.000	0.090	1.000
costs_neg	0.025	0.079	0	0	0	0	1
clinic_envt_neg	0.026	0.075	0	0	0	0	1
ease_schedule_neg	0.043	0.098	0	0	0	0.05	1
waiting_time_neg	0.050	0.110	0	0	0	0.05	1
Positive.Proportion	0.763	0.255	0.000	0.636	0.810	1.000	1.000
Negative.Proportion	0.205	0.245	0.000	0.000	0.143	0.333	1.000
Average_Positive_Sentiment	0.321	0.151	0.000	0.231	0.323	0.410	1.000
Avg_sent_neg1	-0.118	0.160	-1.000	-0.187	-0.058	0.000	0.000
Avg_sent_neg2	0.118	0.160	0.000	0.000	0.058	0.187	1.000
Whole_Sentiment	0.197	0.165	-0.474	0.090	0.202	0.302	0.691

Exploratory Data Analysis - Univariate

- Healthgrades have 20% more female doctors than males, whereas RateMds have almost same proportion of Male and Female doctors
- Only 10% of doctors are below age 40 in Healthgrades with rest of the doctors being almost the same proportion
- In Healthgrades, there are 6% doctors having 0-10 years of experience, whereas in RateMds there are 11%
- Number of doctors with 40+ years of experience are 45% more in Healthgrades than RateMds

Healthgrades					
Gender	**Male**	**Female**			
	331	509			
Age	**<40**	**41-50**	**51-60**	**60+**	
	72	228	272	227	
Experience	**0-10**	**11-20**	**21-30**	**31-40**	**40+**
	52	206	280	203	99

Ratemds					
Gender	**Male**	**Female**			
	414	476			
Experience	**0-10**	**11-20**	**21-30**	**31-40**	**40+**
	85	270	234	97	37

P-values Healthgrades

Healthgrades p-values					
	Doctor Ratings	**Avg Positive score**	**Avg Negative score**	**Positive proportion**	**Negative proportion**
Gender	0.00000000000146	0.0354	0.148	0.0182	0.0000909
Experience	0.806	0.664	0.695	0.715	0.667
Number of words	0.245	0.059	0.00000168	0.023	0.759
Number of Reviews	0.111	0.476	0.0000574	0.948	0.056
Words Per Review	0.0000000000000472	<0.0000000000000002	0.0737	<0.0000000000000002	0.0000000552
Med_Expertise_pos	0.000129	0.0844	0.000698	0.000000000756	0.000000000052
Med_Expertise_neg	0.000000000698	0.00000364	<0.0000000000000002	<0.0000000000000002	<0.0000000000000002
Bedside_manners_pos	<0.0000000000000002	0.0000000000000049	<0.0000000000000002	<0.0000000000000002	<0.0000000000000002
Bedside_manners_neg	<0.0000000000000002	0.00000000000963	<0.0000000000000002	<0.0000000000000002	<0.0000000000000002
Office_staff_pos	0.0146	0.00439	0.495	0.0123	0.00303
Office_staff_neg	0.0000000000994	0.0208	0.00000000577	0.0000000000205	0.000000000000636
Costs_pos	0.87	0.55	0.34	0.859	0.0539
Costs_neg	0.0334	0.00767	0.0213	0.00702	0.00199
Clinical_envt_pos	0.219	0.012	0.0502	0.113	0.0000441
Clinical_envt_neg	0.000000000163	0.0172	0.000146	0.000000000988	0.000000000344
Ease_scheduling_pos	0.79	0.071	0.782	0.579	0.578
Ease_scheduling_neg	0.00000374	0.0114	0.006	0.0000000275	0.0000000000286
Waiting_time_pos	0.99	0.00958	0.209	0.649	0.193
Waiting_time_neg	0.0000000344	0.00243	0.000255	0.000000000359	0.00000000297
Age	0.0536	0.105	0.251	0.0182	0.0895

- Medical Expertise, Bedside Manners and Office Staff have an impact on all the Dependent Variables for both positive and negative sentiments
- Costs Negative and Clinical environment negative sentiment is important across all the dependent variables
- Costs positive and clinical environment pos is important for determining the proportion of negative sentiments

P-values (Ratemds)

RateMds p-values					
	Doctor Ratings	Avg Positive score	Avg Negative score	Positive proportion	Negative proportion
Gender	0.00000000959	0.0135	0.00263	0.000131	0.000024
Experience	0.776	0.382	0.5	0.638	0.661
Number of words	0.00000165	0.0125	0.000000767	0.154	0.0627
Number of Reviews	0.481	0.0516	0.000000249	0.704	0.611
Words Per Review	<0.0000000000000002	<0.0000000000000002	0.00359	0.00145	0.00000125
Med_Expertise_pos	0.000000475	0.017	0.0244	0.000000000231	0.00000000000787
Med_Expertise_neg	0.000000000063	<0.0000000000000002	0.0141	<0.0000000000000002	<0.0000000000000002
Bedside_manners_pos	<0.0000000000000002	<0.0000000000000002	8.33E-11	<0.0000000000000002	<0.0000000000000002
Bedside_manners_neg	0.00000000000000109	0.00171	0.0000000000833	0.00000000000248	0.0000000000000738
Office_staff_pos	0.000793	0.000000212	0.00478	0.00211	0.0000155
Office_staff_neg	0.0000000000209	0.000821	0.00174	0.00000373	0.0000000374
Costs_pos	0.03	0.428	0.621	0.752	0.707
Costs_neg	0.00000000000011	0.000132	0.00477	0.000000000000376	0.0000000000000787
Clinical_envt_pos	0.000123	0.00882	0.0472	0.0164	0.0409
Clinical_envt_neg	0.00000000000383	0.00568	0.0000000405	0.0000000000000178	<0.0000000000000002
Ease_scheduling_pos	0.499	0.377	0.0618	0.0232	0.0284
Ease_scheduling_neg	0.000198	0.0000000253	0.842	0.00268	0.0147
Waiting_time_pos	0.434	0.591	0.725	0.084	0.0473
Waiting_time_neg	0.0000000369	0.000000661	0.587	0.0315	0.00454

- Almost all the sentiments are contributing to the prediction of dependent variables
- Gender and Words per review are important in both the websites
- Even here strangely, positive bucket of waiting time is only contributing to the negative proportion of reviews

Correlation Coefficients (HealthGrades)

Healthgrades Correlation Coefficients						
	Doctor Ratings	Avg Positive Sentiment	Avg Negative Sentiment	Positive Proportion	Negative Proportion	Overall Sentiment
Age	-0.090	0.050	0.050	-0.100	0.020	-0.030
Experience	-0.030	0.070	0.010	-0.030	-0.030	0.030
Number of words	-0.040	-0.070	0.160	-0.080	-0.010	-0.060
Number of Reviews	0.060	0.020	0.140	0.000	-0.070	0.030
Words Per Review	-0.260	-0.320	0.060	-0.280	0.190	-0.310
Med_Expertise_pos	0.130	0.060	-0.120	0.210	-0.220	0.210
bedside_manners_pos	0.420	0.270	-0.320	0.510	-0.510	0.620
office_staff_pos	0.080	0.100	-0.020	0.090	-0.100	0.140
costs_pos	0.010	-0.020	-0.030	0.010	-0.070	0.030
clinic_envt_pos	0.040	0.090	-0.070	0.050	-0.140	0.130
ease_schedule_pos	-0.010	-0.060	-0.010	-0.020	-0.020	-0.020
waiting_time_pos	0.000	-0.090	-0.040	0.020	-0.050	0.010
med_expertise_neg	-0.210	-0.160	0.190	-0.350	0.370	-0.420
bedside_manners_neg	-0.380	-0.230	0.360	-0.480	0.570	-0.630
office_staff_neg	-0.220	-0.080	0.200	-0.230	0.240	-0.260
costs_neg	-0.070	-0.090	0.080	-0.090	0.110	-0.160
clinic_envt_neg	-0.220	-0.080	0.130	-0.210	0.210	-0.230
ease_schedule_neg	-0.160	-0.090	0.090	-0.190	0.240	-0.220
waiting_time_neg	-0.190	-0.100	0.130	-0.230	0.200	-0.240

Correlation Coefficients (RateMDs)

RateMds Correlation Coefficients						
	Doctor Ratings	Avg Positive Sentiment	Avg Negative Sentiment	Positive Proportion	Negative Proportion	Overall Sentiment
Experience	0.001	-0.034	-0.031	-0.009	-0.002	-0.012
Number of words	-0.160	-0.084	0.165	-0.048	0.062	-0.134
Number of Reviews	-0.024	0.065	0.172	0.013	-0.017	0.007
Words Per Review	-0.332	-0.456	-0.098	-0.107	0.162	-0.338
Med_Expertise_pos	0.168	0.080	-0.075	0.210	-0.227	0.203
bedside_manners_pos	0.534	0.390	-0.215	0.420	-0.426	0.568
office_staff_pos	0.112	0.173	-0.094	0.103	-0.144	0.224
costs_pos	-0.073	-0.027	0.017	0.011	-0.013	-0.036
clinic_envt_pos	0.128	0.088	-0.067	0.080	-0.069	0.096
ease_schedule_pos	0.023	0.030	-0.063	0.076	-0.073	0.044
waiting_time_pos	-0.026	0.018	-0.012	0.058	-0.067	0.049
med_expertise_neg	-0.217	-0.272	0.082	-0.287	0.279	-0.358
bedside_manners_neg	-0.264	-0.105	0.234	-0.232	0.247	-0.288
office_staff_neg	-0.222	-0.112	0.105	-0.154	0.183	-0.240
costs_neg	-0.246	-0.128	0.095	-0.240	0.247	-0.242
clinic_envt_neg	-0.230	-0.064	0.183	-0.262	0.288	-0.248
ease_schedule_neg	-0.124	-0.185	0.007	-0.101	0.082	-0.178
waiting_time_neg	-0.183	-0.166	0.018	-0.072	0.095	-0.143

Exploratory Data Analysis - Healthgrades

- **Age v/s Rating**
- Rating is significantly more for age group ‘less than 40’ as compared to ‘50-60’
- Rating is significantly less for age group ‘more than 60’ as compared to ‘less than 40’

- **Age v/s Bedside Manners**
- Doctors with age below 40 have more positive bed side manners as compared to doctors with age more than 50 {50-60, 60+}

- **Gender v/s Rating**
- In the context of gynecology: Male doctors have a better rating than female doctors
- Male doctors have an average rating of 4.14 and females have an average rating of 3.77

- **Gender v/s Negative bedside manners**
- Females are more negatively reviewed as compared to males in the context of bedside manners
- Female doctors have an average negative bedside rating of 0.132 as compared to male’s average rating of 0.104

- **Gender v/s Positive medical expertise**
- Males are more positively reviewed as compared to females in the context of medical expertise
- Male doctors have an average positive medical expertise as 0.174 as compared to female’s expertise of 0.145

Bivariate Data Analysis - RateMDs

- **Gender v/s Experience**:

Male physicians show significantly higher Experience Years compared to Female doctors

- **Gender v/s Ratings**:

Male physicians show significantly higher Star ratings.

- **Gender v/s Bedside Manners**:

Male doctors are more positively reviewed in the context of bedside manners.

- **Ratings v/s Experience**:

Star Ratings Mean lowest for Doctors with Experience of above 40 years

- **Words/Review v/s Experience**:

Star Ratings significantly decrease with increase in Words/Review with p value of <2e-16

Gender	Experience Mean	Star Ratings Mean	Bedside_Manner_Pos
Female	18.41176	3.585903	0.2761765
Male	25.18715	3.943647	0.33343
p-value	2.20E-16	7.68E-09	0.0001305

Linear Regression – Ratemds

- We can see that Star rating will increase by a unit of 0.197 if a physician is Male.
- Similarly, average no. of words per review also contributes strongly to few of the below mentioned models. More no. of words is used in the case of negative reviews when compared to positive reviews.
- Our data also confirms that the negative representation of various themes in the review greatly affects most of the target variables than positive representation of themes.

	RateMds					
	Avg +ve score	Avg -ve score	Prop +ve reviews	Prop -ve reviews	Overall-Sentiment	Star rating
	(1)	(2)	(3)	(4)	(5)	(6)
Doctor_GenderMale	0.014	-0.018	0.028*	-0.031**	0.014*	0.197***
	(0.009)	(0.012)	(0.016)	(0.014)	(0.008)	(0.054)
Experience21 - 30	0.006	0.029	0.014	-0.005	0.012	0.040
	(0.015)	(0.019)	(0.024)	(0.022)	(0.013)	(0.084)
Experience31 - 40	0.003	0.023	-0.007	0.018	-0.002	-0.036
	(0.015)	(0.019)	(0.025)	(0.023)	(0.013)	(0.087)
Experience41	above	-0.028*	0.013	-0.015	0.010	-0.006
	(0.017)	(0.021)	(0.028)	(0.025)	(0.015)	(0.095)
Words_per_review	-0.001***	-0.001***	0.0003	0.0001	-0.001***	-0.004***
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)	(0.001)
med_expertise_pos	0.103*	-0.020	0.032	-0.072	0.092	0.377
	(0.063)	(0.081)	(0.105)	(0.096)	(0.056)	(0.362)
bedside_manners_pos	0.152**	-0.079	0.113	-0.113	0.168***	1.049***
	(0.060)	(0.077)	(0.100)	(0.092)	(0.053)	(0.344)
office_staff_pos	0.244***	-0.082	-0.058	-0.093	0.251***	0.467
	(0.075)	(0.097)	(0.126)	(0.115)	(0.067)	(0.434)
costs_pos	-0.032	0.069	-0.009	-0.074	-0.179	-0.468
	(0.164)	(0.210)	(0.274)	(0.251)	(0.146)	(0.944)
clinic_envt_pos	0.244***	-0.089	0.064	-0.004	0.112	1.349**
	(0.092)	(0.119)	(0.155)	(0.142)	(0.083)	(0.533)
ease_schedule_pos	0.131	-0.109	0.011	-0.082	-0.022	0.460
	(0.120)	(0.154)	(0.200)	(0.184)	(0.107)	(0.691)
waiting_time_pos	0.084	0.023	0.212	-0.289*	0.142	0.168
	(0.100)	(0.129)	(0.167)	(0.154)	(0.089)	(0.577)
med_expertise_neg	-0.073	0.134*	-0.486***	0.412***	-0.280***	-0.826**
	(0.062)	(0.080)	(0.104)	(0.095)	(0.056)	(0.358)
bedside_manners_neg	0.008	0.316***	-0.550***	0.513***	-0.311***	-1.523***
	(0.067)	(0.087)	(0.113)	(0.104)	(0.060)	(0.389)
office_staff_neg	-0.050	0.066	-0.245**	0.263**	-0.208***	-1.037**
	(0.071)	(0.091)	(0.118)	(0.109)	(0.063)	(0.408)
costs_neg	-0.040	0.220**	-0.859***	0.775***	-0.391***	-2.163***
	(0.077)	(0.099)	(0.129)	(0.118)	(0.069)	(0.445)
clinic_envt_neg	0.009	0.312***	-0.876***	0.911***	-0.442***	-2.105***
	(0.080)	(0.103)	(0.134)	(0.123)	(0.072)	(0.463)
ease_schedule_neg	-0.136*	-0.014	-0.248**	0.121	-0.175***	-0.577
	(0.071)	(0.091)	(0.119)	(0.109)	(0.063)	(0.409)
waiting_time_neg	-0.121*	0.015	-0.242**	0.265**	-0.177***	-1.162***
	(0.069)	(0.089)	(0.116)	(0.106)	(0.062)	(0.399)
Constant	0.343***	0.132*	0.874***	0.093	0.255***	3.996***
	(0.059)	(0.076)	(0.099)	(0.090)	(0.053)	(0.340)
Observations	732	732	732	732	732	732
R2	0.382	0.162	0.389	0.432	0.591	0.487
Adjusted R2	0.365	0.139	0.373	0.417	0.580	0.474
Residual Std. Error (df = 712)	0.116	0.149	0.194	0.178	0.104	0.668
F Statistic (df = 19; 712)	23.153***	7.233***	23.880***	28.493***	54.235***	35.625***
Note:						*p<0.1; **p<0.05; ***p<0.01

Linear Regression – Healthgrades

- We can see that Star rating will increase by a unit of 0.278 if a physician is Male.
- Like Ratemds, average no. of words per review also contributes strongly to few of the below mentioned models. More no. of words is used in the case of negative reviews when compared to positive reviews.
- Unlike Ratemds, positive themes contribute to the star rating more than negative themes.
- Although, for other target variables, negative themes are contributing more than the positive themes.

	Healthgrades					
	Avg +ve score	Avg -ve score	Prop +ve reviews	Prop -ve reviews	Overall-Sentiment	Star rating
	(1)	(2)	(3)	(4)	(5)	(6)
Doctor_GenderMale	0.005	0.004	0.008	-0.007	0.012	0.278***
	(0.010)	(0.013)	(0.013)	(0.009)	(0.008)	(0.048)
Experience11 - 20	0.003	0.016	0.007	-0.021	0.021	-0.113
	(0.020)	(0.028)	(0.027)	(0.019)	(0.016)	(0.101)
Experience21 - 30	0.001	0.031	-0.004	-0.025	0.012	-0.131
	(0.020)	(0.027)	(0.027)	(0.019)	(0.016)	(0.099)
Experience31 - 40	0.004	0.004	0.004	-0.035*	0.029*	-0.175*
	(0.021)	(0.029)	(0.028)	(0.020)	(0.017)	(0.104)
Experience41	above	0.035	0.0001	-0.011	-0.042*	0.034**
	(0.022)	(0.030)	(0.029)	(0.021)	(0.017)	(0.110)
Words_per_review	-0.002***	-0.001	-0.002***	0.0001	-0.001***	-0.005***
	(0.0002)	(0.0003)	(0.0003)	(0.0002)	(0.0002)	(0.001)
med_expertise_pos	-0.032	0.141	0.457***	-0.047	-0.084	1.707***
	(0.127)	(0.176)	(0.170)	(0.123)	(0.101)	(0.634)
bedside_manners_pos	-0.012	0.084	0.474***	-0.078	-0.003	1.966***
	(0.125)	(0.173)	(0.167)	(0.120)	(0.099)	(0.622)
office_staff_pos	0.123	0.248	0.456**	-0.021	0.024	2.108***
	(0.142)	(0.197)	(0.191)	(0.137)	(0.113)	(0.710)
costs_pos	-0.048	0.173	0.301	-0.030	-0.135	1.531**
	(0.137)	(0.189)	(0.184)	(0.132)	(0.108)	(0.682)
clinic_envt_pos	0.074	0.087	0.290	-0.160	-0.020	1.378*
	(0.144)	(0.199)	(0.192)	(0.139)	(0.113)	(0.715)
ease_schedule_pos	-0.196	0.255	0.151	0.093	-0.327***	1.283*
	(0.148)	(0.204)	(0.198)	(0.142)	(0.117)	(0.735)
waiting_time_pos	-0.293**	0.078	0.320	-0.001	-0.213*	1.347*
	(0.149)	(0.206)	(0.200)	(0.144)	(0.118)	(0.742)
med_expertise_neg	-0.144	0.390**	-0.108	0.422***	-0.543***	0.861
	(0.129)	(0.178)	(0.172)	(0.124)	(0.102)	(0.641)
bedside_manners_neg	-0.187	0.541***	-0.183	0.552***	-0.686***	0.263
	(0.128)	(0.177)	(0.171)	(0.123)	(0.101)	(0.637)
office_staff_neg	-0.084	0.724***	-0.119	0.391**	-0.522***	-0.086
	(0.160)	(0.222)	(0.215)	(0.155)	(0.127)	(0.799)
costs_neg	-0.261	0.404	0.308	0.143	-0.523***	1.807*
	(0.215)	(0.298)	(0.289)	(0.208)	(0.170)	(1.073)
clinic_envt_neg	-0.118	0.411*	-0.145	0.435***	-0.491***	-0.367
	(0.158)	(0.218)	(0.211)	(0.152)	(0.125)	(0.786)
ease_schedule_neg	-0.170	0.218	-0.089	0.626***	-0.514***	0.610
	(0.178)	(0.246)	(0.238)	(0.171)	(0.140)	(0.884)
waiting_time_neg	-0.133	0.359*	-0.055	0.242*	-0.412***	0.625
	(0.150)	(0.207)	(0.200)	(0.144)	(0.118)	(0.745)
Constant	0.540***	-0.022	0.425**	0.101	0.457***	2.801***
	(0.126)	(0.175)	(0.169)	(0.122)	(0.100)	(0.629)
Observations	840	840	840	840	840	840
R2	0.188	0.202	0.445	0.532	0.669	0.322
Adjusted R2	0.169	0.182	0.431	0.521	0.661	0.306
Residual Std. Error (df = 819)	0.129	0.178	0.173	0.124	0.102	0.642
F Statistic (df = 20; 819)	9.501***	10.360***	32.837***	46.600***	82.899***	19.481***

Note: *p<0.1; **p<0.05; ***p<0.01

Linear Regression – Combined

- The data from both the websites were then combined to check if the hypothesis holds strong.
- In this case, even experience of more than 30 years is positively affecting the proportion of positive reviews and overall sentiment
- Combined data also shows that overall star rating and other dependent variables except proportion of negative reviews increases for male physicians when compared to female physicians.
- Negative themes are contributing more than the positive themes which is also reflecting the results of each websites.

	Both Websites					
	Avg +ve score (1)	Avg -ve score (2)	Prop +ve reviews (3)	Prop -ve reviews (4)	Overall-Sentiment (5)	Star rating (6)
Doctor_GenderMale	0.004 (0.007)	-0.011 (0.009)	$\boxed{0.030*** (0.011)}$	$\boxed{-0.017** (0.008)}$	$\boxed{0.012** (0.006)}$	$\boxed{0.246*** (0.036)}$
Experience11 - 20	0.001 (0.019)	-0.003 (0.025)	0.050 (0.030)	-0.029 (0.023)	0.025 (0.016)	-0.063 (0.100)
Experience21 - 30	0.004 (0.018)	0.011 (0.025)	$\boxed{0.072** (0.030)}$	-0.036 (0.023)	$\boxed{0.027* (0.015)}$	-0.043 (0.097)
Experience31 - 40	0.005 (0.019)	-0.010 (0.025)	$\boxed{0.071** (0.030)}$	-0.030 (0.023)	$\boxed{0.030* (0.016)}$	-0.107 (0.100)
Experience41	above (0.019)	0.004 (0.026)	-0.015 (0.031)	$\boxed{0.058* (0.024)}$	-0.039 (0.016)	$\boxed{0.028* (0.103)}$
Words_per_review	$\boxed{-0.001*** (0.0001)}$	$\boxed{-0.001*** (0.0001)}$	$\boxed{0.0004** (0.0002)}$	0.0001 (0.0001)	$\boxed{-0.001*** (0.0001)}$	$\boxed{-0.005*** (0.001)}$
med_expertise_pos	0.079 (0.058)	0.021 (0.078)	0.091 (0.094)	-0.059 (0.071)	0.058 (0.048)	$\boxed{0.672** (0.308)}$
bedside_manners_pos	$\boxed{0.119** (0.056)}$	-0.032 (0.075)	0.110 (0.090)	-0.092 (0.069)	$\boxed{0.146*** (0.047)}$	$\boxed{1.055*** (0.296)}$
office_staff_pos	$\boxed{0.211*** (0.068)}$	0.014 (0.092)	0.087 (0.110)	-0.071 (0.084)	$\boxed{0.189*** (0.057)}$	$\boxed{0.880** (0.362)}$
costs_pos	0.100 (0.074)	$\boxed{0.206** (0.100)}$	$\boxed{-0.564*** (0.120)}$	-0.004 (0.091)	-0.031 (0.062)	0.159 (0.393)
clinic_envt_pos	$\boxed{0.212*** (0.075)}$	-0.011 (0.102)	-0.052 (0.122)	-0.093 (0.093)	0.104 (0.063)	$\boxed{0.947** (0.401)}$
ease_schedule_pos	-0.013 (0.083)	0.095 (0.113)	$\boxed{-0.231* (0.135)}$	0.039 (0.103)	$\boxed{-0.148** (0.070)}$	0.337 (0.444)
waiting_time_pos	-0.073 (0.080)	0.017 (0.109)	0.052 (0.130)	-0.132 (0.099)	0.021 (0.067)	0.325 (0.427)
med_expertise_neg	-0.078 (0.058)	$\boxed{0.206*** (0.079)}$	$\boxed{-0.417*** (0.094)}$	$\boxed{0.415*** (0.072)}$	$\boxed{-0.351*** (0.049)}$	-0.380 (0.310)
bedside_manners_neg	-0.025 (0.059)	$\boxed{0.460*** (0.080)}$	$\boxed{-0.677*** (0.095)}$	$\boxed{0.534*** (0.073)}$	$\boxed{-0.485*** (0.049)}$	$\boxed{-0.933*** (0.313)}$
office_staff_neg	-0.079 (0.067)	0.145 (0.091)	-0.071 (0.109)	$\boxed{0.295*** (0.083)}$	$\boxed{-0.264*** (0.057)}$	$\boxed{-0.758** (0.359)}$
costs_neg	-0.099 (0.075)	$\boxed{0.241** (0.102)}$	$\boxed{-0.604*** (0.122)}$	$\boxed{0.721*** (0.093)}$	$\boxed{-0.431*** (0.063)}$	$\boxed{-1.600*** (0.401)}$
clinic_envt_neg	-0.013 (0.074)	$\boxed{0.350*** (0.100)}$	$\boxed{-0.714*** (0.120)}$	$\boxed{0.768*** (0.092)}$	$\boxed{-0.450*** (0.062)}$	$\boxed{-1.630*** (0.395)}$
ease_schedule_neg	$\boxed{-0.163** (0.069)}$	0.016 (0.093)	-0.110 (0.111)	$\boxed{0.187** (0.085)}$	$\boxed{-0.237*** (0.058)}$	-0.231 (0.365)
waiting_time_neg	$\boxed{-0.126* (0.065)}$	0.086 (0.088)	-0.158 (0.106)	$\boxed{0.256*** (0.081)}$	$\boxed{-0.233*** (0.055)}$	$\boxed{-0.724** (0.348)}$
Constant	0.392*** (0.058)	0.123 (0.078)	0.700*** (0.094)	0.117* (0.071)	0.277*** (0.049)	3.781*** (0.308)
Observations	1,572	1,572	1,572	1,572	1,572	1,572
R2	0.297	0.178	0.350	0.460	0.626	0.401
Adjusted R2	0.288	0.167	0.342	0.453	0.621	0.394
Residual Std. Error (df = 1551)	0.124	0.168	0.201	0.153	0.104	0.660
F Statistic (df = 20; 1551)	32.779***	16.803***	41.744***	65.941***	129.681***	52.006***
Note:						*p<0.1; **p<0.05; ***p<0.01

Conclusions

- When compared to female physicians, start rating of male physicians increase in Ratemds and Healthgrades by a unit of 0.197 and 0.278 respectively.
- Although Experience of the physician does not make an effect in Ratemds, it proves significant in determining proportion of negative sentiment, overall sentiment score and star rating in Healthgrades.
- The relationship between average words per review and dependent variables is almost similar in Ratemds and Healthgrades. We can also say that, on average, negative review has more no. of words per review that positive review.
- For both the websites, the negative themes contribute to each of the dependent variables than positive themes.
- In both the websites, among the themes, “bedside manners” affect the star rate the most whereas “ease of scheduling” affects the least.
- In Ratemds, the negative themes seem to contribute more towards the star rating than positive themes. Whereas, in Healthgrades, positive themes contribute towards star rating than negative themes.
- Although, in Ratemds, gender plays a role in determining proportion of positive reviews, proportion of negative reviews, overall sentiment and star rating, in Healthgrades, gender plays a role in determining only star rating.