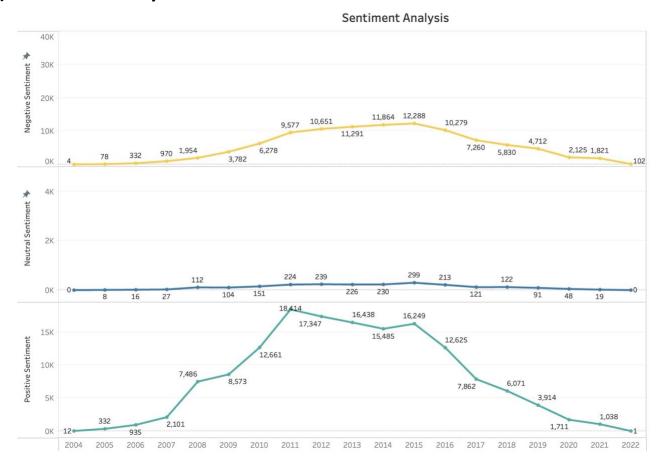
DATA 228 – Big Data Technologies Customer Segmentation and Analysis on Yelp Dataset

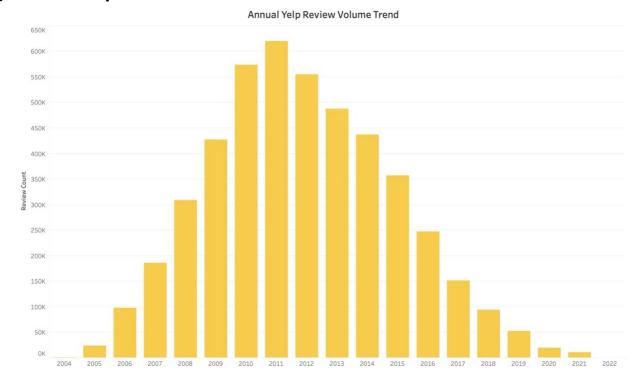
Data Visualizations on Tableau

1) Sentiment Analysis:



This visualization represents the volume of sentiments (negative, neutral, and positive) across different years, highlighting the changes in sentiment distribution over time. We can see that the number of positive insights during the years 2007 to 2020 are significantly higher than that of neutral and negative compliments.

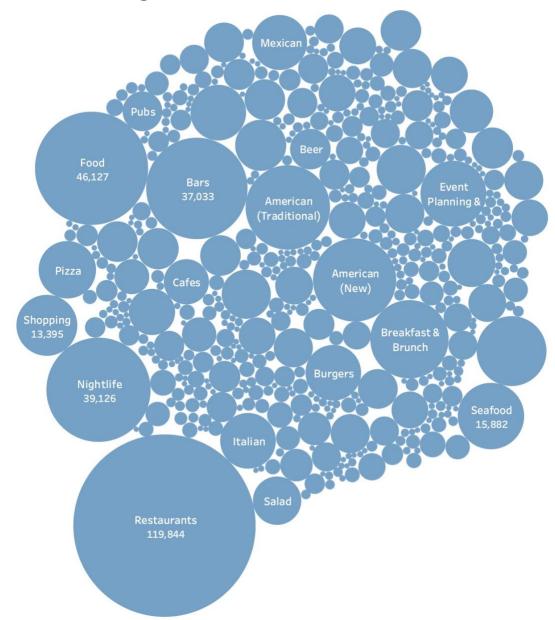
2) Annual Yelp Review Volume Trend



This shows the trend of Yelp review volumes year-over-year, with the height of the bars indicating the number of reviews, which can signal the platform's growth or decline in user activity.

3) World Cloud

World Cloud Of Categories



A visual representation of the popularity of categories, with the size of each term indicating the frequency or importance, useful for identifying the most prominent categories in a glance.

4) Frequency Distribution of review Counts by User Activity

Frequency Distribution of Review Counts by User Activity

							Review Cou	nt						
No. of Days	0	50	100	150	200	250	300	350	400	450	500	550	600	650
100	1,255	70	30	5	13	7	2	2	3		2		1	
200	967	48	27	11	3	5	2		3		1			
300	894	58	24	14	5	7	3	2	1	1		1	2	
400	794	65	27	11	7	2	1	3	3			2		
500	709	39	23	11	4	8	5	2	2	2	2			
600	647	47	22	11	6	2		2	4	2	1	4		
700	663	44	25	15	7	4	2	6					1	
800	596	40	12	12	4	7	3	5	2	1		1		
900	518	40	16	6	5	4	2	2	1		1		2	
1000	512	42	17	5	5	5	1		2	1	1		1	
1100	470	38	17	15	2	4	2	2	1	3	2			
1200	475	28	17	9	13	10	2	3	3	4	1	3	1	
1300	412	44	19	11	7	1	1	4			2	1	2	
1400	368	27	17	9	6	1	5	3	1	1	1	2	1	
1500	319	39	14	7	5	2	1	4	1	2		2		
1600	306	18	10	8	3	2		2	1			1		
1700	312	37	16	8	5	4	2	3	1	2	1	1	1	
1800	263	26	9	7	8		3	1	1	1	1	2		
1900	257	34	13	8	4	5	2	3		2			1	
2000	203	17	10	8	4	1	1		1					
2100	209	23	12	6	2	2	2	1	1		2	1	1	
2200	194	23	10	6	3	6		1	1	1			1	
2300	170	22	7	5	3	1		1	2		1		1	
2400	130	26	4	7	4	1				2		1	1	
2500	128	16	7	6	3	1		1	2			1		
2600	94	15	10	2	2			5	2	1	1	1		
2700	89	19	1	1	1	5	3	3	1	1	1		2	
2800	87	11	3	6	1	2	2	1			1			
2900	85	13	2	6	1	2	2	1				1		
3000	54	12	6	5	1	3							1	
3100	57	10	3	1	1		1		2	3				

This is a heatmap that visualizes the concentration of reviews over different time spans, with darker shades typically representing higher frequencies.

5) Distribution of Sentiment analysis



This visual represents the percentage of sentiments (negative, neutral, and positive) with respect to each other. This gives us an overview of the comparative trend of each sentiment.