

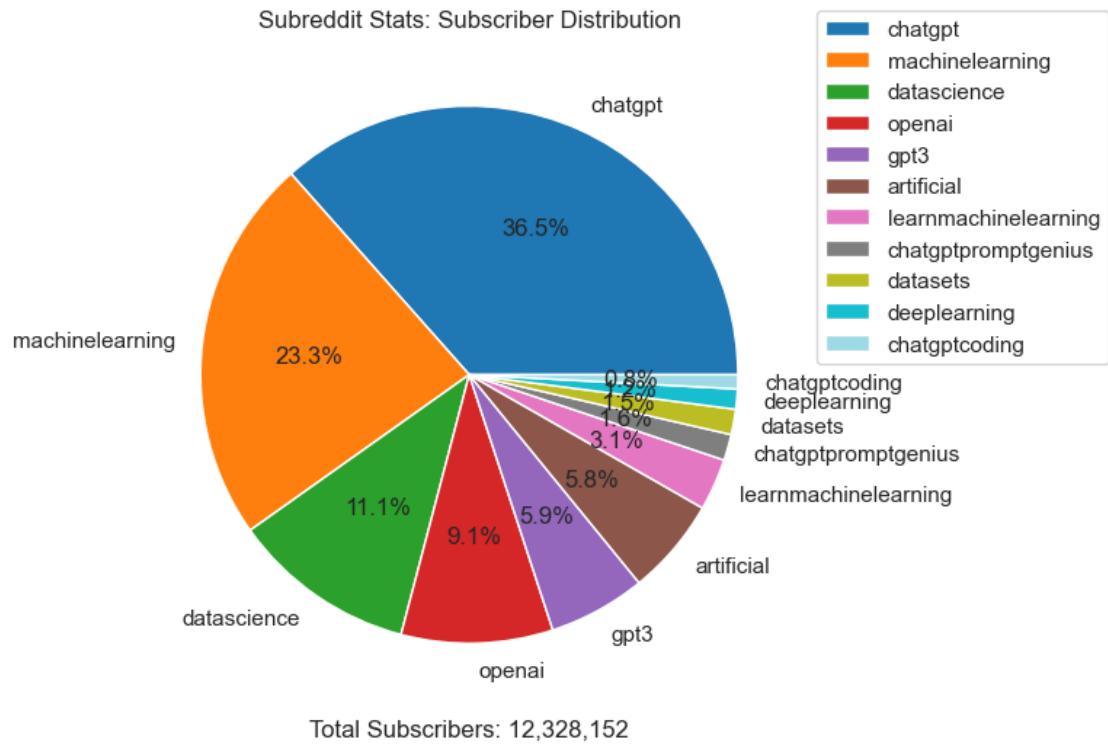
model-training

February 26, 2024

1 Reddit Subreddit API analysis

Subreddit Stats

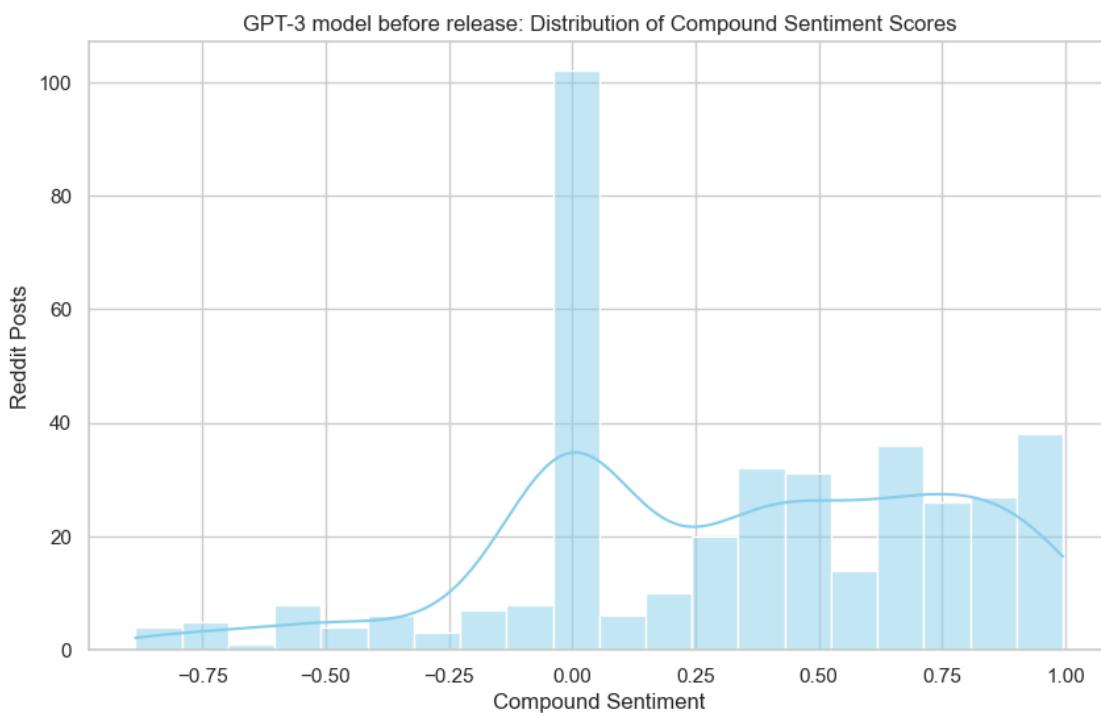
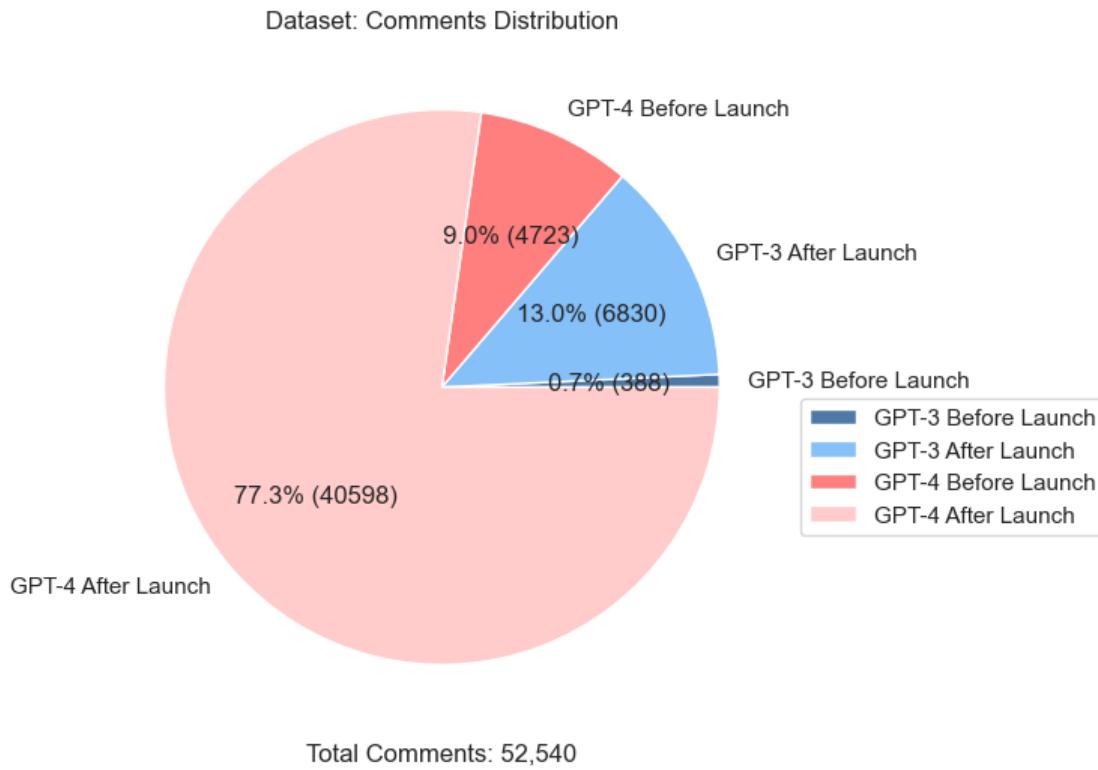
name	index	subscribers	display_name	subscribers_pct
chatgpt	1	4505500	chatgpt	36.55
machinelearning	10	2870456	MachineLearning	23.28
datascience	4	1366348	datascience	11.08
openai	11	1124730	OpenAI	9.12
gpt3	7	722076	GPT3	5.86
artificial	0	720095	artificial	5.84
learnmachinelearning	9	384514	learnmachinelearning	3.12
chatgptpromptgenius	3	193280	ChatGPTPromptGenius	1.57
datasets	5	187702	datasets	1.52
deeplearning	6	149090	deeplearning	1.21
chatgptcoding	2	104361	ChatGPTCoding	0.85



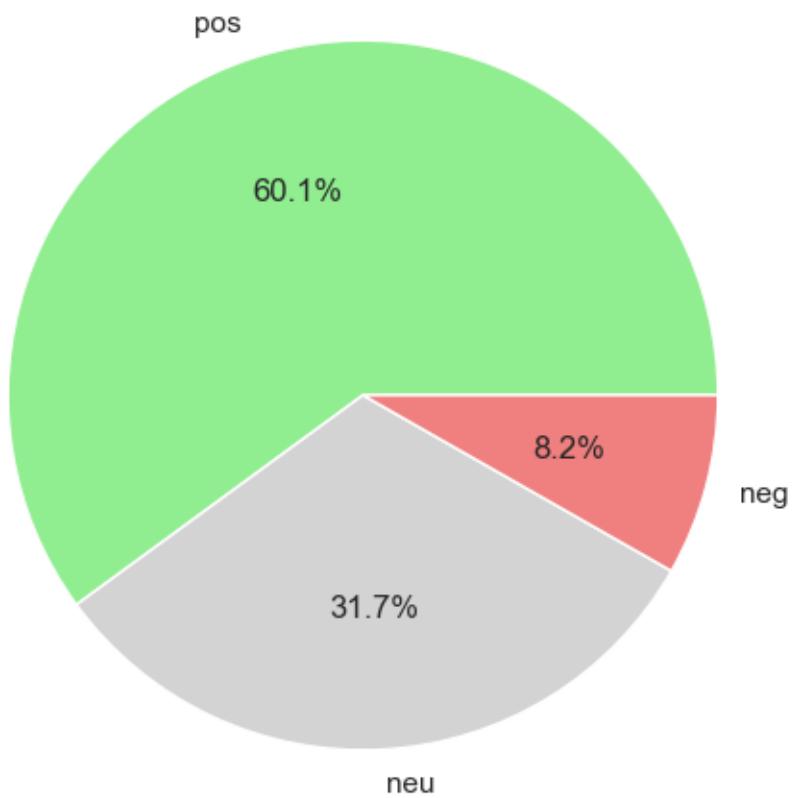
2 Reddit Search API data analysis

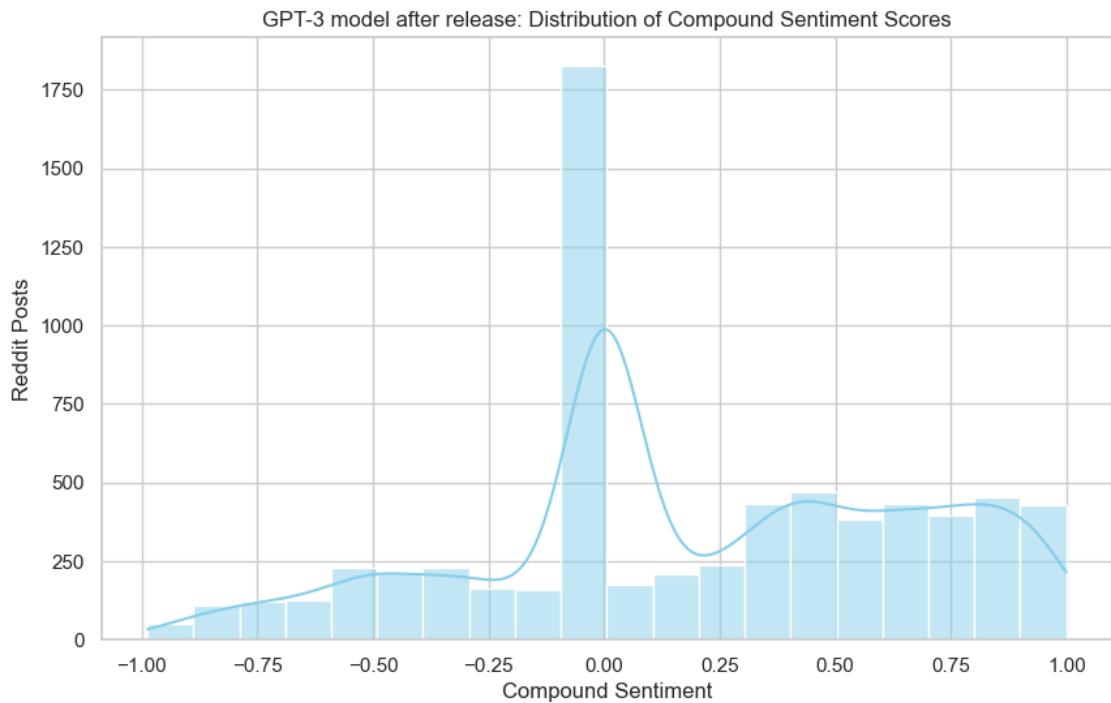
Total imported Reddit comments: 52540

3 Data split on launch dates

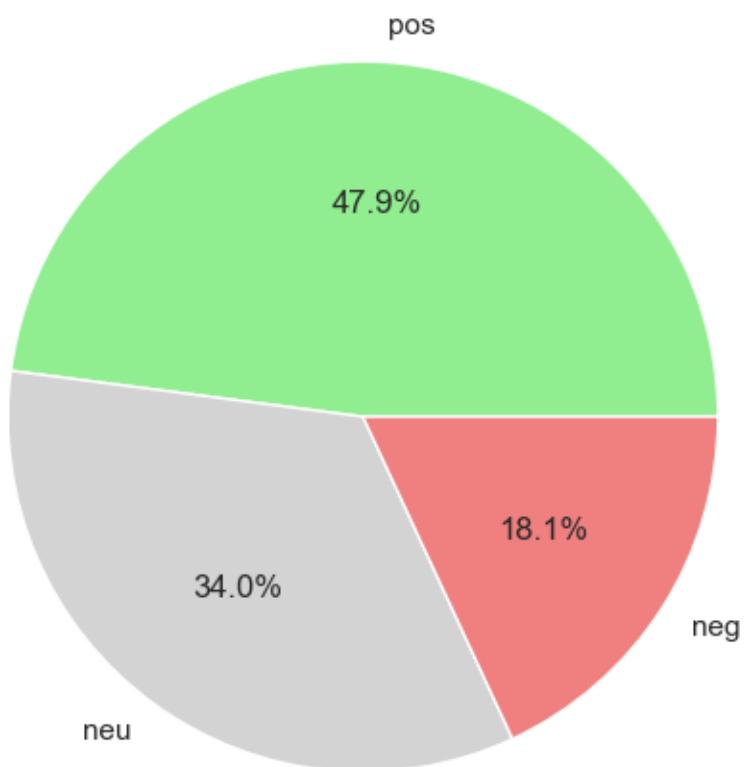


GPT-3 model before release: Sentiment Distribution

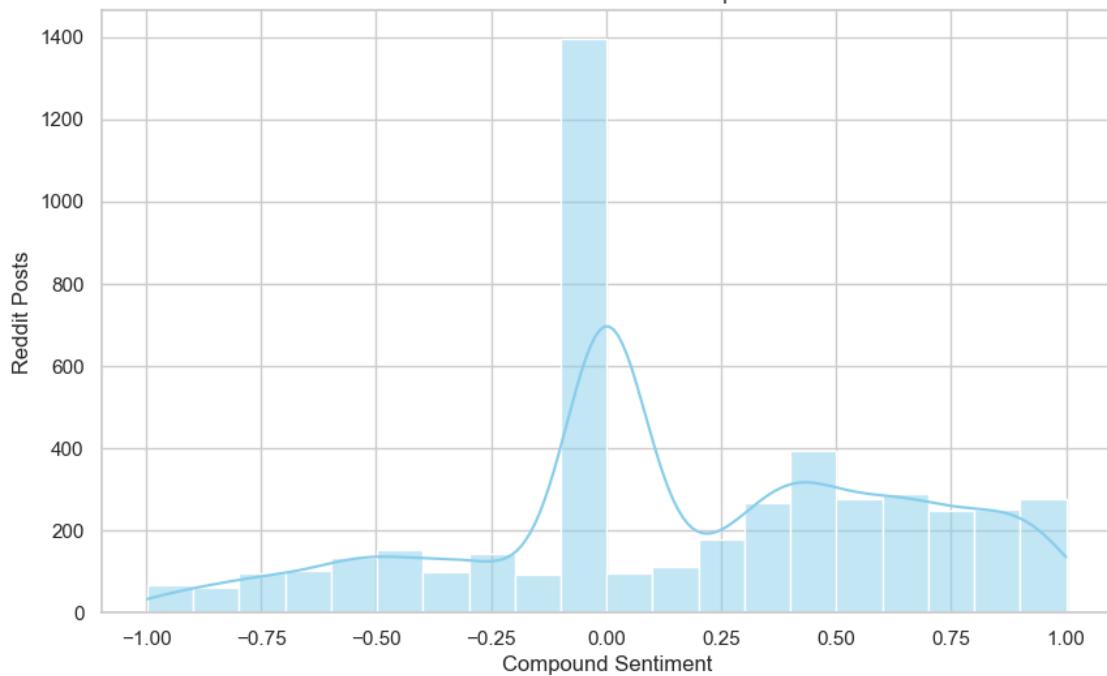




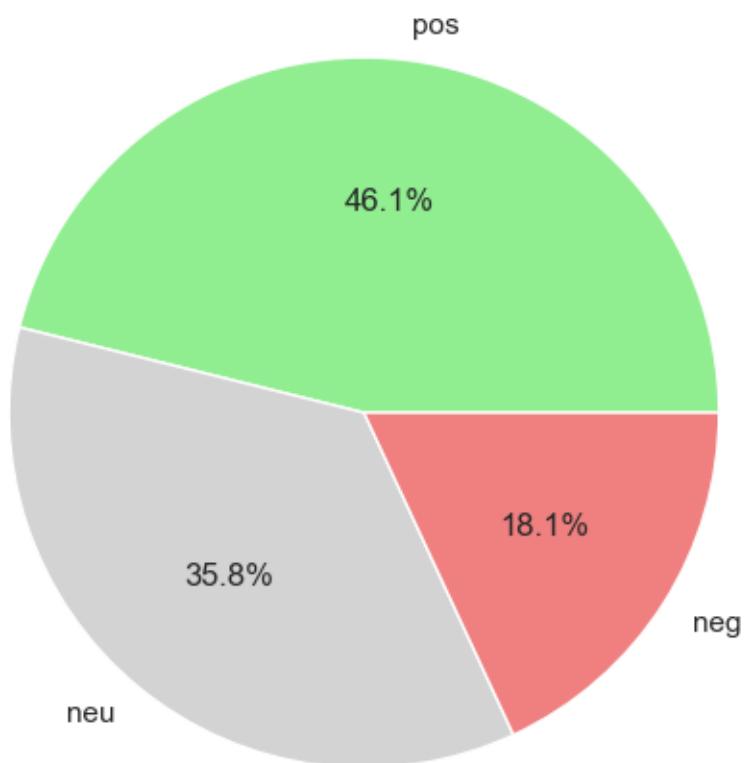
GPT-3 model after release: Sentiment Distribution

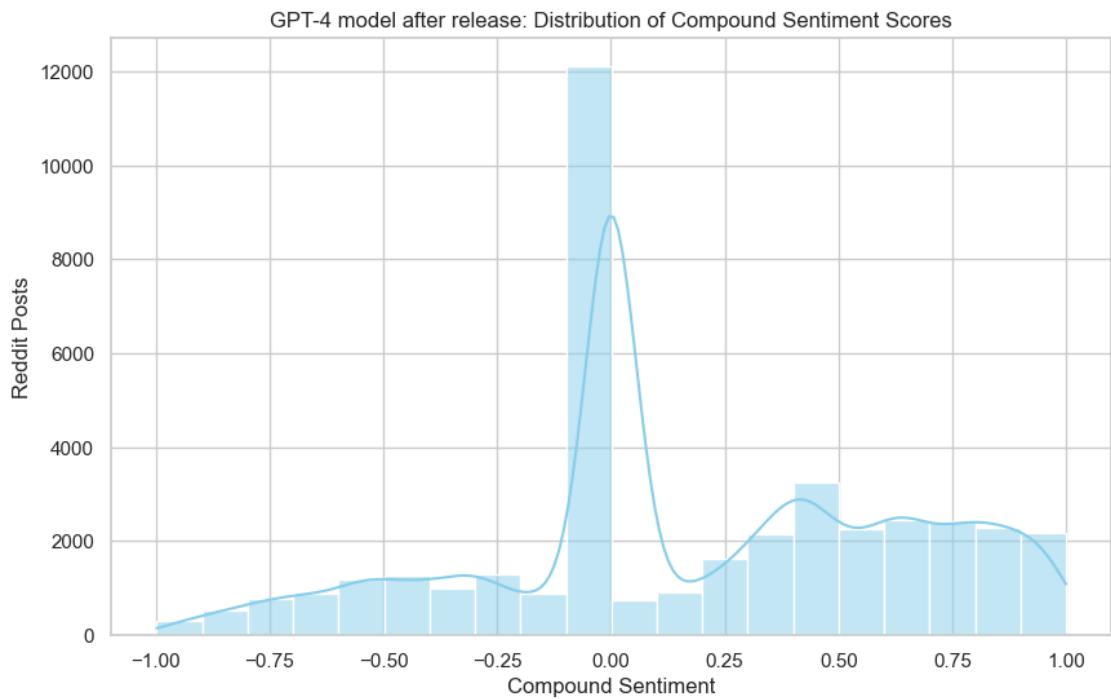


GPT-4 model before release: Distribution of Compound Sentiment Scores

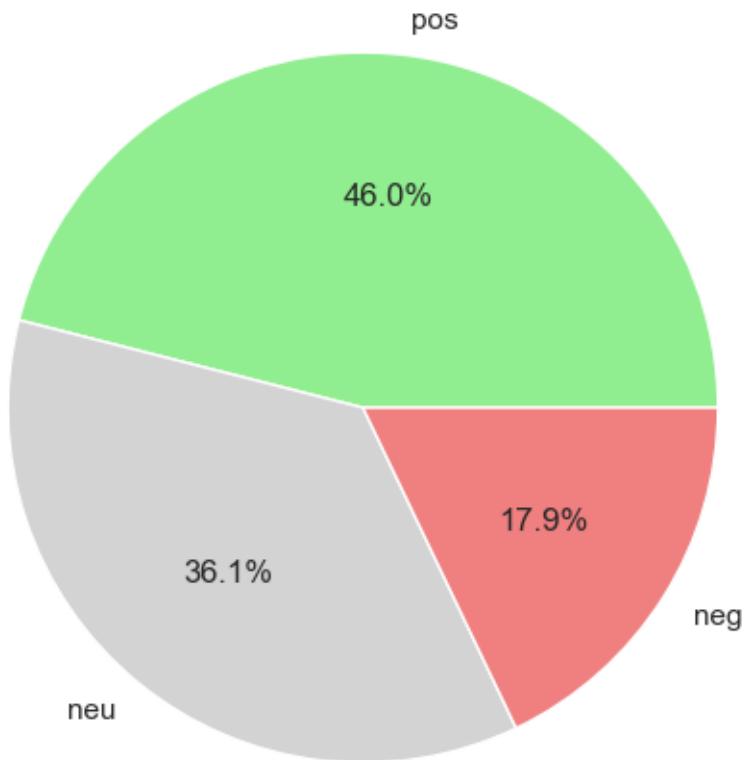


GPT-4 model before release: Sentiment Distribution



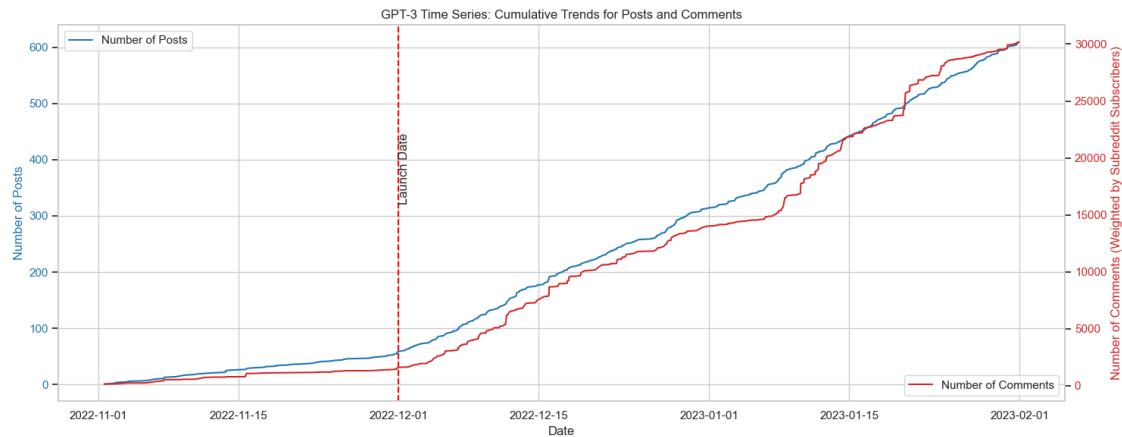


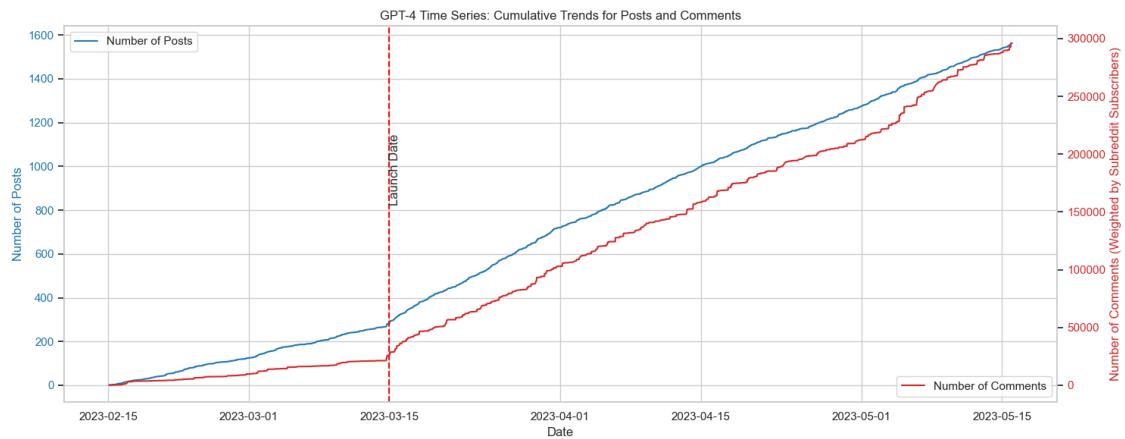
GPT-4 model after release: Sentiment Distribution



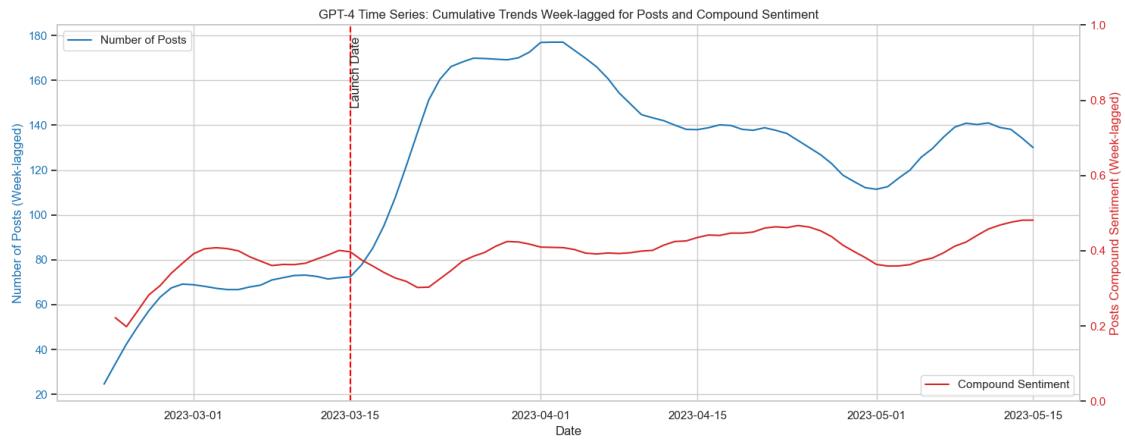
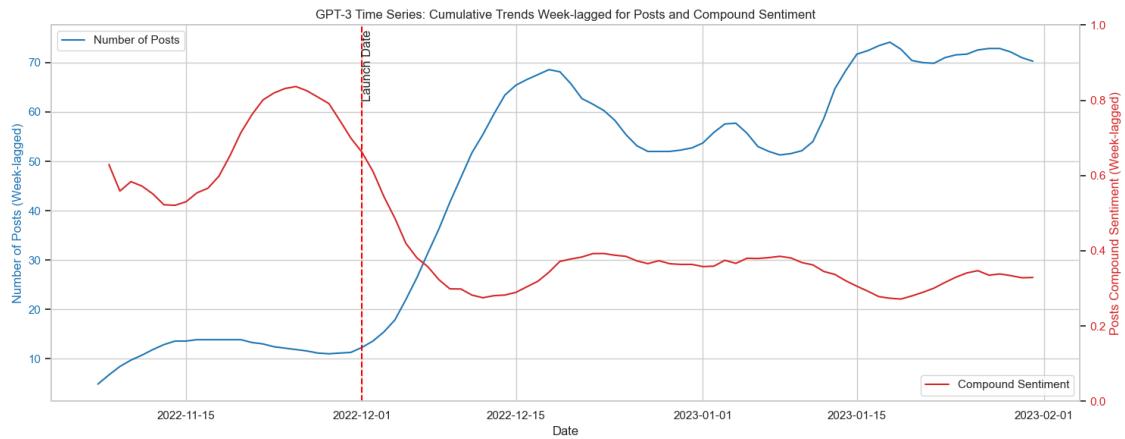
4 Time Series Analysis

4.1 Posts and Comments

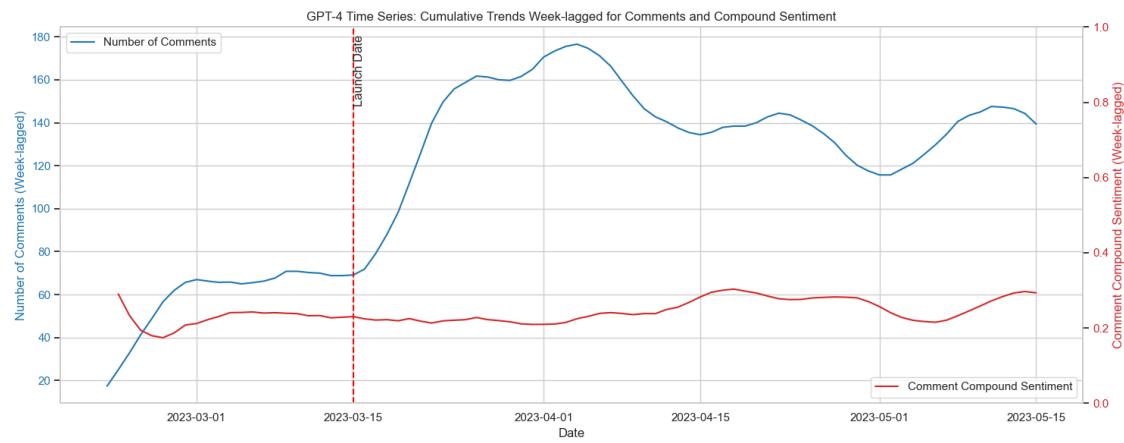
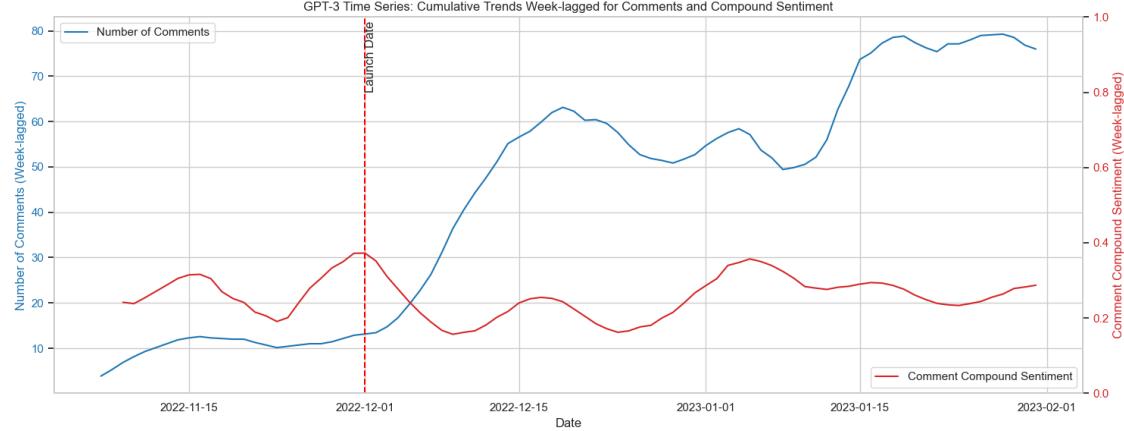




4.2 Posts and Sentiment



4.3 Comments and Sentiment



5 Preparing MLflow features and targets for Regression

```
feature_names=['launched', 'upvote_ratio', 'log_launch_distance_f',
'log_num_comments_weighted', 'log_score_weighted', 'log_comment_score_weighted',
'log_comment_launch_distance_f', 'log_comment_num_replies_weighted',
'log_count_cumulative', 'log_count_cumulative_lagged',
'log_comment_count_cumulative', 'log_comment_count_cumulative_lagged',
'compound_cumulative', 'compound_cumulative_lagged',
'comment_compound_cumulative', 'comment_compound_cumulative_lagged']
target_name=text_compound
```

GPT-3 model before release

	index	count	mean	std	
min					
401.00	index	45.00	3415.87	1626.45	
134.00	index_x	45.00	1495.56	598.34	
0.00	stickied	45.00	0.00	0.00	
0.43	upvote_ratio	45.00	0.84	0.17	
0.00	score	45.00	67.27	120.91	
1.00	num_comments	45.00	22.13	36.04	
2022-11-06	created	1970-01-01	2022-11-17	1970-01-08	
0.00	text_neg	45.00	0.04	0.07	
0.41	text_neu	45.00	0.83	0.11	
0.00	text_pos	45.00	0.13	0.10	
-0.50	text_compound	45.00	0.62	0.39	
0.00	score_weighted	45.00	79.72	147.41	
1.01	num_comments_weighted	45.00	25.91	46.18	
1751.00	index_y	45.00	35318.09	12642.02	
0.00	comment_score	45.00	4.09	6.04	
0.00	comment_num_replies	45.00	0.67	0.64	
0.00	comment_body_neg	45.00	0.03	0.07	
0.00	comment_body_neu	45.00	0.79	0.23	
0.00	comment_body_pos	45.00	0.16	0.20	
-0.55	comment_body_compound	45.00	0.31	0.40	
	comment_score_weighted	45.00	4.79	7.82	

0.00							
	comment_num_replies_weighted		45.00		0.75		0.73
0.00							
	launch_distance_f		45.00		12.59		7.40
0.10							
	comment_launch_distance_f		45.00		11.74		8.66
0.11							
	count_cumulative		45.00		12.96		1.71
8.00							
	num_comments_cumulative		45.00		935.54		340.49
281.88							
	count_cumulative_lagged		45.00		12.50		2.31
5.00							
	compound_cumulative		45.00		0.65		0.13
0.47							
	compound_cumulative_lagged		45.00		0.66		0.12
0.52							
	comment_count_cumulative		45.00		11.20		1.98
6.00							
	comment_count_cumulative_lagged		45.00		10.58		2.38
3.57							
	comment_compound_cumulative		45.00		0.28		0.07
0.11							
	comment_compound_cumulative_lagged		45.00		0.27		0.06
0.18							
	log_launch_distance_f		45.00		2.36		0.84
0.09							
	log_num_comments_weighted		45.00		2.57		1.16
0.70							
	log_score_weighted		45.00		2.81		1.89
0.00							
	log_count_cumulative		45.00		2.63		0.13
2.20							
	log_count_cumulative_lagged		45.00		2.58		0.21
1.79							
	log_comment_count_cumulative		45.00		2.49		0.18
1.95							
	log_comment_count_cumulative_lagged		45.00		2.42		0.26
1.52							
	log_comment_score_weighted		45.00		1.35		0.78
0.00							
	log_comment_num_replies_weighted		45.00		0.47		0.43
0.00							
	log_comment_launch_distance_f		45.00		2.25		0.88
0.11							
-----+-----+-----+-----+-----+-----+-----+-----							
-----+-----+-----+-----+-----+-----+-----+-----							
-----+-----+-----+-----+-----+-----+-----+-----							

	index	25%	50%	75%	
max					
6773.00	index	2085.00	3412.00	4009.00	
2313.00	index_x	1216.00	1483.00	1910.00	
0.00	stickied	0.00	0.00	0.00	
1.00	upvote_ratio	0.75	0.89	0.97	
473.00	score	3.00	12.00	50.00	
216.00	num_comments	5.00	11.00	23.00	
2022-11-29	created	2022-11-10	2022-11-16	2022-11-23	
0.45	text_neg	0.00	0.02	0.05	
1.00	text_neu	0.81	0.84	0.91	
0.59	text_pos	0.07	0.12	0.18	
1.00	text_compound	0.50	0.67	0.93	
607.43	score_weighted	3.19	12.75	53.44	
281.56	num_comments_weighted	5.31	11.73	29.74	
50346.00	index_y	34796.00	37848.00	39646.00	
30.00	comment_score	1.00	2.00	4.00	
2.00	comment_num_replies	0.00	1.00	1.00	
0.32	comment_body_neg	0.00	0.00	0.03	
1.00	comment_body_neu	0.71	0.85	0.97	
1.00	comment_body_pos	0.00	0.12	0.19	
0.98	comment_body_compound	0.00	0.34	0.64	
39.11	comment_score_weighted	1.06	2.12	4.40	
	comment_num_replies_weighted	0.00	1.06	1.10	

2.61							
		launch_distance_f		6.71	13.43	19.15	
23.79							
		comment_launch_distance_f		5.14	11.14	18.09	
43.50							
		count_cumulative		12.00	13.00	14.00	
16.00							
		num_comments_cumulative		584.76	1077.21	1195.00	
1407.37							
		count_cumulative_lagged		12.00	13.14	14.14	
14.71							
		compound_cumulative		0.55	0.61	0.77	
0.88							
		compound_cumulative_lagged		0.55	0.61	0.77	
0.85							
		comment_count_cumulative		10.00	12.00	13.00	
14.00							
		comment_count_cumulative_lagged		10.29	11.43	12.14	
13.14							
		comment_compound_cumulative		0.23	0.28	0.34	
0.39							
		comment_compound_cumulative_lagged		0.22	0.27	0.32	
0.38							
		log_launch_distance_f		2.04	2.67	3.00	
3.21							
		log_num_comments_weighted		1.84	2.54	3.43	
5.64							
		log_score_weighted		1.43	2.62	4.00	
6.41							
		log_count_cumulative		2.56	2.64	2.71	
2.83							
		log_count_cumulative_lagged		2.56	2.65	2.72	
2.75							
		log_comment_count_cumulative		2.40	2.56	2.64	
2.71							
		log_comment_count_cumulative_lagged		2.42	2.52	2.58	
2.65							
		log_comment_score_weighted		0.72	1.14	1.69	
3.69							
		log_comment_num_replies_weighted		0.00	0.72	0.74	
1.28							
		log_comment_launch_distance_f		1.82	2.50	2.95	
3.80							
-----+-----+-----+-----+-----+-----+-----+-----							
-----+-----+-----+-----+-----+-----+-----+-----							
-----+-----+-----+-----+-----+-----+-----+-----							
GPT-3 model after release							
-----+-----+-----+-----+-----+-----+-----+-----							
-----+-----+-----+-----+-----+-----+-----+-----							

		index	count	mean	std	
min						
44.00		index	543.00	3149.07	1920.01	
6.00		index_x	543.00	1382.53	790.25	
0.00		stickied	543.00	0.00	0.00	
0.12		upvote_ratio	543.00	0.82	0.20	
0.00		score	543.00	200.38	1355.50	
1.00		num_comments	543.00	42.71	81.46	
2022-11-30		created	1970-01-01	2023-01-02	1970-01-19	
0.00		text_neg	543.00	0.04	0.07	
0.00		text_neu	543.00	0.86	0.12	
0.00		text_pos	543.00	0.10	0.11	
-1.00		text_compound	543.00	0.33	0.52	
0.00		score_weighted	543.00	268.21	2127.43	
1.01		num_comments_weighted	543.00	50.94	112.54	
217.00		index_y	543.00	34920.68	15951.39	
0.00		comment_score	543.00	6.99	19.20	
0.00		comment_num_replies	543.00	0.74	1.20	
0.00		comment_body_neg	543.00	0.04	0.10	
0.00		comment_body_neu	543.00	0.81	0.22	
0.00		comment_body_pos	543.00	0.14	0.19	
-0.95		comment_body_compound	543.00	0.26	0.46	
0.00		comment_score_weighted	543.00	7.90	22.34	
0.00		comment_num_replies_weighted	543.00	0.82	1.35	

		index	25%	50%	75%	
max						
0.47		launch_distance_f	543.00	33.08	18.22	
0.56		comment_launch_distance_f	543.00	34.77	17.91	
12.00		count_cumulative	543.00	61.31	13.47	
1411.62		num_comments_cumulative	543.00	15560.33	8752.83	
12.00		count_cumulative_lagged	543.00	61.03	13.84	
0.23		compound_cumulative	543.00	0.34	0.05	
0.23		compound_cumulative_lagged	543.00	0.34	0.05	
12.00		comment_count_cumulative	543.00	60.10	16.54	
12.86		comment_count_cumulative_lagged	543.00	59.82	16.79	
0.12		comment_compound_cumulative	543.00	0.25	0.06	
0.13		comment_compound_cumulative_lagged	543.00	0.25	0.06	
0.38		log_launch_distance_f	543.00	3.30	0.80	
0.70		log_num_comments_weighted	543.00	2.98	1.35	
0.00		log_score_weighted	543.00	3.37	2.07	
2.56		log_count_cumulative	543.00	4.10	0.29	
2.56		log_count_cumulative_lagged	543.00	4.09	0.31	
2.56		log_comment_count_cumulative	543.00	4.06	0.35	
2.63		log_comment_count_cumulative_lagged	543.00	4.05	0.36	
0.00		log_comment_score_weighted	543.00	1.45	0.97	
0.00		log_comment_num_replies_weighted	543.00	0.44	0.52	
0.44		log_comment_launch_distance_f	543.00	3.38	0.73	

		index	1795.00	2965.00	4007.50	
7217.00		index_x	943.50	1374.00	1898.00	
7000.00		stickied	0.00	0.00	0.00	
0.00		upvote_ratio	0.75	0.90	0.96	
1.00		score	5.00	31.00	120.50	
26465.00		num_comments	5.50	16.00	43.00	
876.00		created	2022-12-16	2023-01-04	2023-01-17	
2023-01-31		text_neg	0.00	0.01	0.05	
0.59		text_neu	0.79	0.86	1.00	
1.00		text_pos	0.00	0.09	0.15	
1.00		text_compound	0.00	0.32	0.87	
1.00		score_weighted	5.31	32.93	130.35	
41707.66		num_comments_weighted	6.07	18.05	47.10	
1380.54		index_y	33806.00	36518.00	39635.50	
148492.00		comment_score	1.00	2.00	5.00	
294.00		comment_num_replies	0.00	0.00	1.00	
12.00		comment_body_neg	0.00	0.00	0.06	
1.00		comment_body_neu	0.74	0.86	1.00	
1.00		comment_body_pos	0.00	0.09	0.18	
1.00		comment_body_compound	0.00	0.24	0.65	
1.00		comment_score_weighted	1.06	2.12	5.31	
330.65		comment_num_replies_weighted	0.00	0.00	1.10	
13.20		launch_distance_f	16.03	35.17	48.93	
62.93						

		count	mean	std	
62.96	comment_launch_distance_f	18.25	37.88	50.60	
84.00	count_cumulative	55.00	65.00	71.00	
30173.20	num_comments_cumulative	8678.46	14418.74	22975.02	
82.71	count_cumulative_lagged	54.57	65.14	70.50	
0.52	compound_cumulative	0.31	0.34	0.37	
0.58	compound_cumulative_lagged	0.31	0.34	0.37	
87.00	comment_count_cumulative	51.00	60.00	74.00	
85.43	comment_count_cumulative_lagged	50.29	59.86	74.79	
0.43	comment_compound_cumulative	0.21	0.26	0.29	
0.42	comment_compound_cumulative_lagged	0.21	0.26	0.28	
4.16	log_launch_distance_f	2.84	3.59	3.91	
7.23	log_num_comments_weighted	1.96	2.95	3.87	
10.64	log_score_weighted	1.84	3.52	4.88	
4.44	log_count_cumulative	4.03	4.19	4.28	
4.43	log_count_cumulative_lagged	4.02	4.19	4.27	
4.48	log_comment_count_cumulative	3.95	4.11	4.32	
4.46	log_comment_count_cumulative_lagged	3.94	4.11	4.33	
5.80	log_comment_score_weighted	0.72	1.14	1.84	
2.65	log_comment_num_replies_weighted	0.00	0.00	0.74	
4.16	log_comment_launch_distance_f	2.96	3.66	3.94	
-----+-----+-----+-----+-----+					
-----+-----+-----+-----+-----+					
GPT-4 model before release					
-----+-----+-----+-----+-----+					
	index	count	mean	std	
min					
-----+-----+-----+-----+-----+					

		index	248.00	29473.59	11769.26	
218.00		index_x	248.00	1263.69	736.30	
9.00		stickied	248.00	0.00	0.00	
0.00		upvote_ratio	248.00	0.83	0.19	
0.09		score	248.00	291.03	993.41	
0.00		num_comments	248.00	60.13	130.30	
1.00		created	1970-01-01	2023-03-01	1970-01-08	
2023-02-16		text_neg	248.00	0.03	0.06	
0.00		text_neu	248.00	0.88	0.10	
0.51		text_pos	248.00	0.09	0.09	
0.00		text_compound	248.00	0.37	0.50	
-0.98		score_weighted	248.00	422.06	1565.76	
0.00		num_comments_weighted	248.00	81.59	202.59	
1.01		index_y	248.00	33109.05	14582.05	
277.00		comment_score	248.00	11.68	63.43	
0.00		comment_num_replies	248.00	0.85	1.84	
0.00		comment_body_neg	248.00	0.05	0.12	
0.00		comment_body_neu	248.00	0.77	0.26	
0.00		comment_body_pos	248.00	0.17	0.23	
0.00		comment_body_compound	248.00	0.23	0.44	
-0.97		comment_score_weighted	248.00	15.86	99.40	
0.00		comment_num_replies_weighted	248.00	1.01	2.53	
0.00		launch_distance_f	248.00	12.72	7.37	
0.22		comment_launch_distance_f	248.00	12.75	8.41	

0.08							
		count_cumulative		248.00		62.80	
8.00							
		num_comments_cumulative		248.00		11289.03	
482.96							
		count_cumulative_lagged		248.00		62.02	
5.00							
		compound_cumulative		248.00		0.34	
0.03							
		compound_cumulative_lagged		248.00		0.34	
0.06							
		comment_count_cumulative		248.00		57.72	
5.00							
		comment_count_cumulative_lagged		248.00		56.96	
3.00							
		comment_compound_cumulative		248.00		0.23	
0.12							
		comment_compound_cumulative_lagged		248.00		0.24	
0.12							
		log_launch_distance_f		248.00		2.41	
0.20							
		log_num_comments_weighted		248.00		3.12	
0.70							
		log_score_weighted		248.00		3.60	
0.00							
		log_count_cumulative		248.00		4.10	
2.20							
		log_count_cumulative_lagged		248.00		4.07	
1.79							
		log_comment_count_cumulative		248.00		3.97	
1.79							
		log_comment_count_cumulative_lagged		248.00		3.94	
1.39							
		log_comment_score_weighted		248.00		1.47	
0.00							
		log_comment_num_replies_weighted		248.00		0.45	
0.00							
		log_comment_launch_distance_f		248.00		2.38	
0.08							
	+-----+-----+-----+-----+						
	- - - - +						
	+-----+-----+-----+-----+						
	- - - - +						
		index		25%		50%	
max							
	+-----+-----+-----+-----+						
	- - - - +						
		index		30284.00		32801.50	

45171.00								
	index_x		655.75		1321.50		1793.00	
5268.00								
	stickied		0.00		0.00		0.00	
0.00								
	upvote_ratio		0.78		0.91		0.97	
1.00								
	score		5.00		35.00		134.75	
8287.00								
	num_comments		6.00		18.00		49.00	
1123.00								
	created		2023-02-22		2023-03-01		2023-03-07	
2023-03-13								
	text_neg		0.00		0.00		0.04	
0.38								
	text_neu		0.82		0.89		1.00	
1.00								
	text_pos		0.00		0.08		0.13	
0.47								
	text_compound		0.00		0.40		0.89	
1.00								
	score_weighted		5.31		39.16		152.42	
13059.94								
	num_comments_weighted		6.07		19.67		53.86	
1769.80								
	index_y		32197.00		35958.00		39330.75	
121532.00								
	comment_score		1.00		2.00		5.00	
717.00								
	comment_num_replies		0.00		0.00		1.00	
16.00								
	comment_body_neg		0.00		0.00		0.03	
1.00								
	comment_body_neu		0.70		0.83		1.00	
1.00								
	comment_body_pos		0.00		0.09		0.24	
1.00								
	comment_body_compound		0.00		0.20		0.60	
0.97								
	comment_score_weighted		1.06		2.06		5.31	
1129.96								
	comment_num_replies_weighted		0.00		0.00		1.06	
25.22								
	launch_distance_f		6.27		12.11		19.33	
25.92								
	comment_launch_distance_f		5.95		11.84		19.00	
62.02								
	count_cumulative		62.00		69.50		72.25	

79.00						
	num_comments_cumulative		5229.59		11094.70	
21296.85						
	count_cumulative_lagged		61.18		69.64	
78.29						
	compound_cumulative		0.29		0.38	
0.47						
	compound_cumulative_lagged		0.29		0.38	
0.45						
	comment_count_cumulative		55.00		66.00	
77.00						
	comment_count_cumulative_lagged		55.00		65.79	
76.14						
	comment_compound_cumulative		0.20		0.23	
0.58						
	comment_compound_cumulative_lagged		0.20		0.23	
0.58						
	log_launch_distance_f		1.98		2.57	
3.29						
	log_num_comments_weighted		1.96		3.03	
7.48						
	log_score_weighted		1.84		3.69	
9.48						
	log_count_cumulative		4.14		4.26	
4.38						
	log_count_cumulative_lagged		4.13		4.26	
4.37						
	log_comment_count_cumulative		4.03		4.20	
4.36						
	log_comment_count_cumulative_lagged		4.03		4.20	
4.35						
	log_comment_score_weighted		0.72		1.12	
7.03						
	log_comment_num_replies_weighted		0.00		0.00	
3.27						
	log_comment_launch_distance_f		1.94		2.55	
4.14						
-----+-----+-----+-----+-----+-----+-----+						
GPT-4 model after release						
-----+-----+-----+-----+-----+-----+-----+						
	index		count		mean	
min						
-----+-----+-----+-----+-----+-----+-----+						
	index		1271.00		28391.08	
61.00						
-----+-----+-----+-----+-----+-----+-----+						

		index_x	1271.00	1199.75	870.63	
0.00		stickied	1271.00	0.00	0.00	
0.00		upvote_ratio	1271.00	0.82	0.19	
0.06		score	1271.00	952.03	3383.20	
0.00		num_comments	1271.00	141.69	352.51	
1.00		created	1970-01-01	2023-04-12	1970-01-19	
2023-03-14		text_neg	1271.00	0.04	0.09	
0.00		text_neu	1271.00	0.85	0.12	
0.00		text_pos	1271.00	0.10	0.10	
0.00		text_compound	1271.00	0.42	0.53	
-1.00		score_weighted	1271.00	1456.66	5334.46	
0.00		num_comments_weighted	1271.00	207.58	555.69	
1.01		index_y	1271.00	32327.72	17169.70	
61.00		comment_score	1271.00	8.91	60.76	
0.00		comment_num_replies	1271.00	0.74	1.55	
0.00		comment_body_neg	1271.00	0.05	0.11	
0.00		comment_body_neu	1271.00	0.80	0.22	
0.00		comment_body_pos	1271.00	0.14	0.19	
0.00		comment_body_compound	1271.00	0.25	0.46	
-0.98		comment_score_weighted	1271.00	11.82	94.46	
0.00		comment_num_replies_weighted	1271.00	0.88	1.91	
0.00		launch_distance_f	1271.00	29.22	18.51	
0.31		comment_launch_distance_f	1271.00	30.35	18.53	
0.47		count_cumulative	1271.00	143.11	23.94	
72.00						

	num_comments_cumulative	1271.00	146463.43	76247.82	
21303.96					
72.86	count_cumulative_lagged	1271.00	142.99	24.10	
0.25					
0.25	compound_cumulative	1271.00	0.40	0.06	
68.00					
68.71	comment_count_cumulative	1271.00	141.74	22.78	
0.18					
0.18	comment_count_cumulative_lagged	1271.00	141.58	22.98	
0.18					
0.27	comment_compound_cumulative	1271.00	0.25	0.04	
0.70					
0.00	log_launch_distance_f	1271.00	3.12	0.89	
4.29					
4.30	log_num_comments_weighted	1271.00	3.53	1.86	
4.23					
4.24	log_score_weighted	1271.00	3.95	2.73	
0.39					
4.29	log_count_cumulative	1271.00	4.96	0.17	
4.30					
4.23	log_count_cumulative_lagged	1271.00	4.95	0.18	
4.24					
0.00	log_comment_count_cumulative	1271.00	4.95	0.18	
0.00					
0.39	log_comment_count_cumulative_lagged	1271.00	4.95	0.18	
0.39					
0.39	log_comment_score_weighted	1271.00	1.40	0.99	
0.00					
0.00	log_comment_num_replies_weighted	1271.00	0.44	0.55	
0.39					
0.39	log_comment_launch_distance_f	1271.00	3.17	0.86	
-----+-----+-----+-----+-----+-----+					
max					
45320.00	index	25%	50%	75%	
8881.00					
45320.00	index	25517.00	31847.00	35093.00	
8881.00					
8881.00	index_x	505.50	1019.00	1910.50	

	stickied	0.00	0.00	0.00	0.00	
0.00	upvote_ratio	1.00	0.75	0.90	0.96	
1.00	score	43183.00	5.00	37.00	294.50	
43183.00	num_comments	3834.00	6.00	23.00	98.00	
3834.00	created	2023-05-15	2023-03-26	2023-04-10	2023-04-28	
2023-05-15	text_neg	1.00	0.00	0.01	0.05	
1.00	text_neu	1.00	0.80	0.86	0.93	
1.00	text_pos	1.00	0.00	0.09	0.15	
1.00	text_compound	1.00	0.00	0.51	0.93	
1.00	score_weighted	68054.49	5.05	39.36	343.52	
68054.49	num_comments_weighted	6042.21	7.07	25.21	118.29	
6042.21	index_y	197410.00	27006.50	34244.00	39681.50	
197410.00	comment_score	1538.00	1.00	2.00	4.00	
1538.00	comment_num_replies	37.00	0.00	0.00	1.00	
37.00	comment_body_neg	1.00	0.00	0.00	0.06	
1.00	comment_body_neu	1.00	0.73	0.84	0.98	
1.00	comment_body_pos	1.00	0.00	0.09	0.20	
1.00	comment_body_compound	1.00	0.00	0.26	0.66	
1.00	comment_score_weighted	2423.82	1.08	2.03	4.25	
2423.82	comment_num_replies_weighted	40.71	0.00	0.00	1.10	
40.71	launch_distance_f	62.99	12.88	27.87	45.73	
62.99	comment_launch_distance_f	63.00	14.30	29.28	47.01	
63.00	count_cumulative	196.00	131.00	140.00	163.00	
196.00	num_comments_cumulative	296096.98	79459.44	142064.01	205807.96	

	count_cumulative_lagged		130.36	140.00	162.86	
193.86						
	compound_cumulative		0.37	0.41	0.45	
0.54						
	compound_cumulative_lagged		0.37	0.41	0.45	
0.53						
	comment_count_cumulative		131.00	143.00	157.00	
194.00						
	comment_count_cumulative_lagged		131.29	143.14	156.93	
192.71						
	comment_compound_cumulative		0.22	0.24	0.28	
0.33						
	comment_compound_cumulative_lagged		0.22	0.24	0.27	
0.33						
	log_launch_distance_f		2.63	3.36	3.84	
4.16						
	log_num_comments_weighted		2.09	3.27	4.78	
8.71						
	log_score_weighted		1.80	3.70	5.84	
11.13						
	log_count_cumulative		4.88	4.95	5.10	
5.28						
	log_count_cumulative_lagged		4.88	4.95	5.10	
5.27						
	log_comment_count_cumulative		4.88	4.97	5.06	
5.27						
	log_comment_count_cumulative_lagged		4.88	4.97	5.06	
5.27						
	log_comment_score_weighted		0.73	1.11	1.66	
7.79						
	log_comment_num_replies_weighted		0.00	0.00	0.74	
3.73						
	log_comment_launch_distance_f		2.73	3.41	3.87	
4.16						
-----+-----+-----+-----+						
-----+-----+-----+-----+						

6 Run MLflow Regression experiment

Test model against data after GPT-4 launch.

Coefficients of GPT-3 Regression

	Feature	Coefficient
	launched	0.04
	upvote_ratio	0.00
	log_launch_distance_f	1.13

	log_num_comments_weighted		-2.68	
	log_score_weighted		-0.91	
	log_comment_score_weighted		-1.34	
	log_comment_launch_distance_f		-0.29	
	log_comment_num_replies_weighted		0.08	
	log_count_cumulative		-0.04	
	log_count_cumulative_lagged		0.16	
	log_comment_count_cumulative		-0.02	
	log_comment_count_cumulative_lagged		-0.37	
	compound_cumulative		0.47	
	compound_cumulative_lagged		0.02	
	comment_compound_cumulative		-0.08	
	comment_compound_cumulative_lagged		-0.00	

Coefficients of GPT-4 Regression

Feature	Coefficient
launched	0.65
upvote_ratio	-0.00
log_launch_distance_f	3.69
log_num_comments_weighted	-4.45
log_score_weighted	1.17
log_comment_score_weighted	-0.77
log_comment_launch_distance_f	0.02
log_comment_num_replies_weighted	0.08
log_count_cumulative	-0.11
log_count_cumulative_lagged	-0.12
log_comment_count_cumulative	0.00
log_comment_count_cumulative_lagged	-0.36
compound_cumulative	0.59
compound_cumulative_lagged	-0.00
comment_compound_cumulative	0.04
comment_compound_cumulative_lagged	-0.07

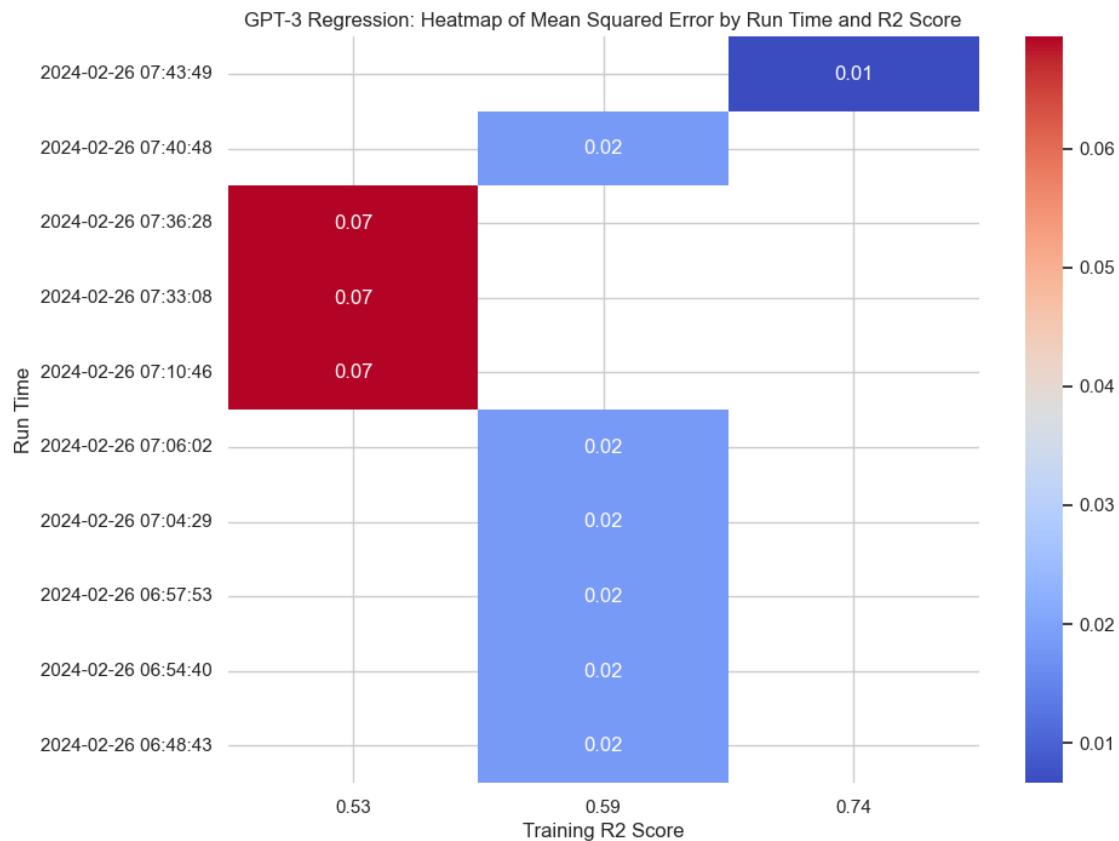
Coefficients of Before-launch Regression

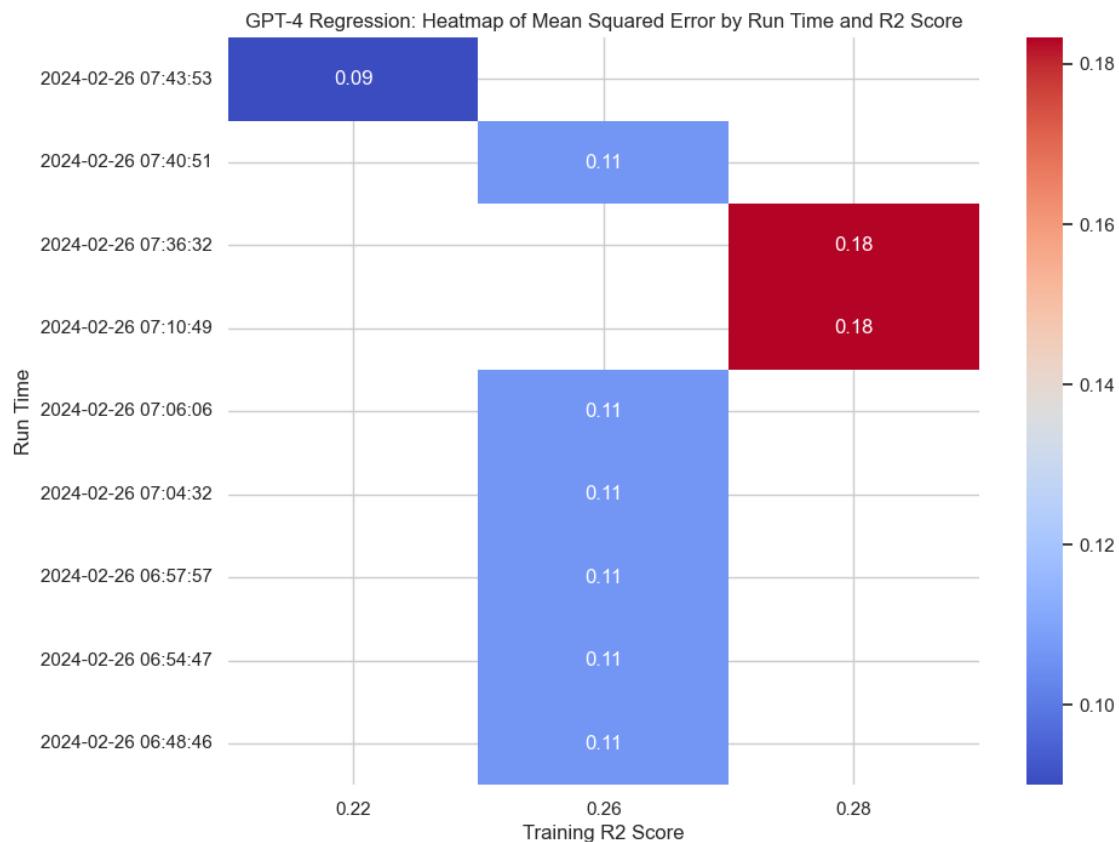
Feature	Coefficient
launched	0.04
upvote_ratio	0.00
log_launch_distance_f	1.13
log_num_comments_weighted	-2.68
log_score_weighted	-0.91
log_comment_score_weighted	-1.34
log_comment_launch_distance_f	-0.29
log_comment_num_replies_weighted	0.08
log_count_cumulative	-0.04

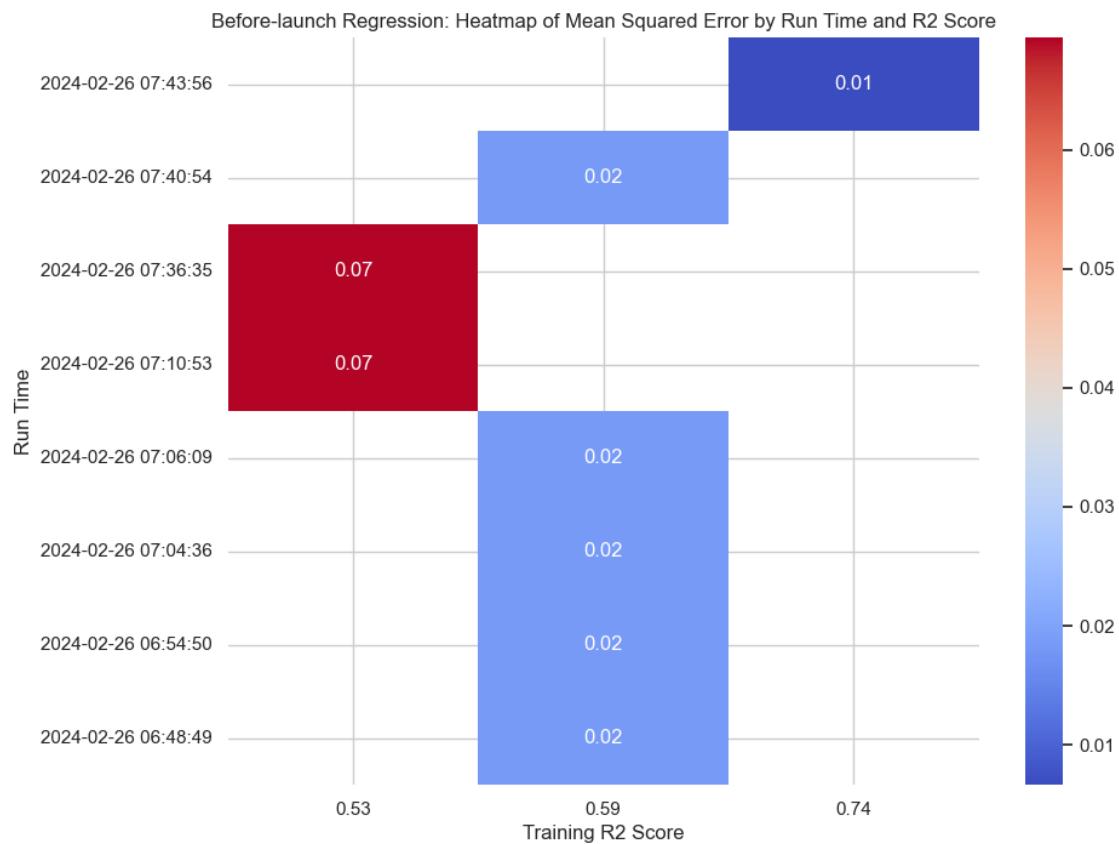
	Feature	Coefficient
	launched	0.05
	upvote_ratio	0.00
	log_launch_distance_f	6.22
	log_num_comments_weighted	-6.40
	log_score_weighted	-2.34
	log_comment_score_weighted	2.16
	log_comment_launch_distance_f	0.04
	log_comment_num_replies_weighted	0.04
	log_count_cumulative	-0.06
	log_count_cumulative_lagged	0.11
	log_comment_count_cumulative	-0.29
	log_comment_count_cumulative_lagged	0.05
	compound_cumulative	0.04
	compound_cumulative_lagged	-0.04
	comment_compound_cumulative	0.05
	comment_compound_cumulative_lagged	-0.06
+-----+-----+		
Coefficients of Comparison Regression		
	Feature	Coefficient
	launched	0.06
	upvote_ratio	-0.15
	log_launch_distance_f	4.11
	log_num_comments_weighted	-4.44
	log_score_weighted	-0.72
	log_comment_score_weighted	0.67
	log_comment_launch_distance_f	0.03
	log_comment_num_replies_weighted	0.05
	log_count_cumulative	-0.06
	log_count_cumulative_lagged	-0.19
	log_comment_count_cumulative	0.01
	log_comment_count_cumulative_lagged	-0.06
	compound_cumulative	0.14
	compound_cumulative_lagged	-0.04
	comment_compound_cumulative	0.04

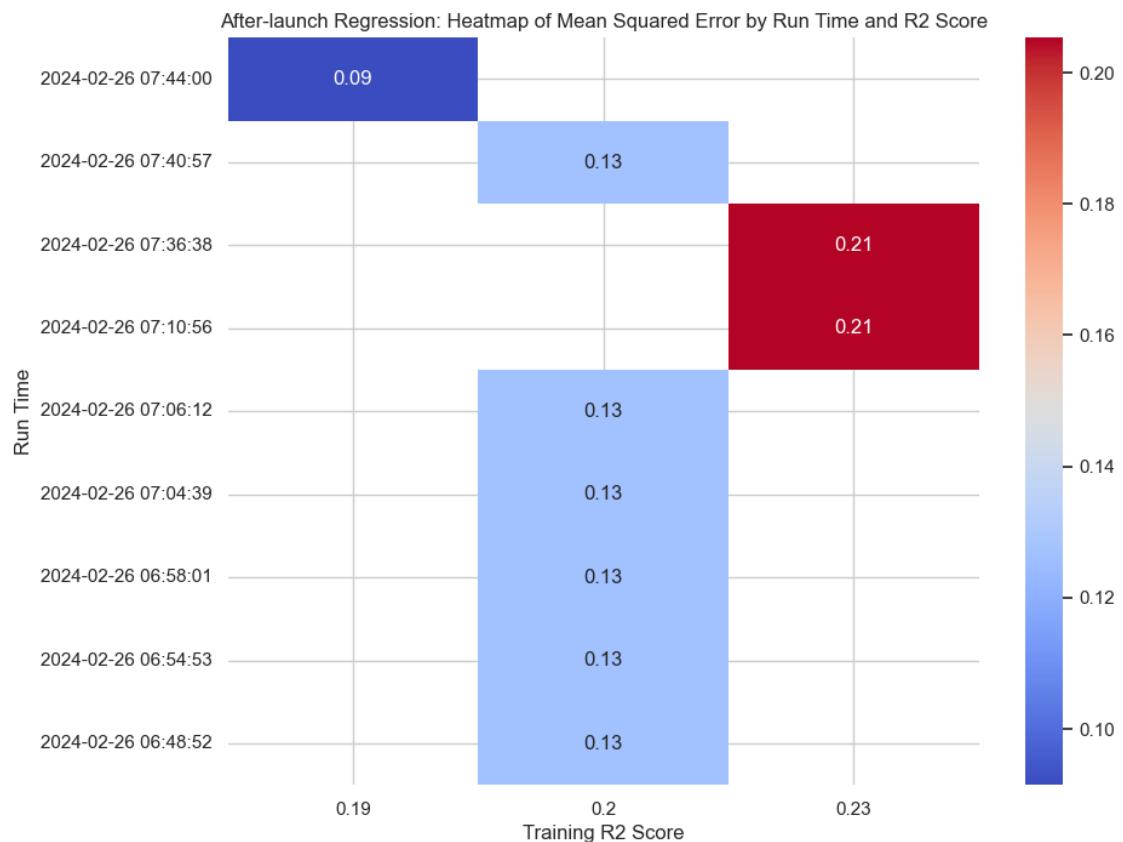
```
| comment_compound_cumulative_lagged | -0.07 |  
+-----+-----+
```

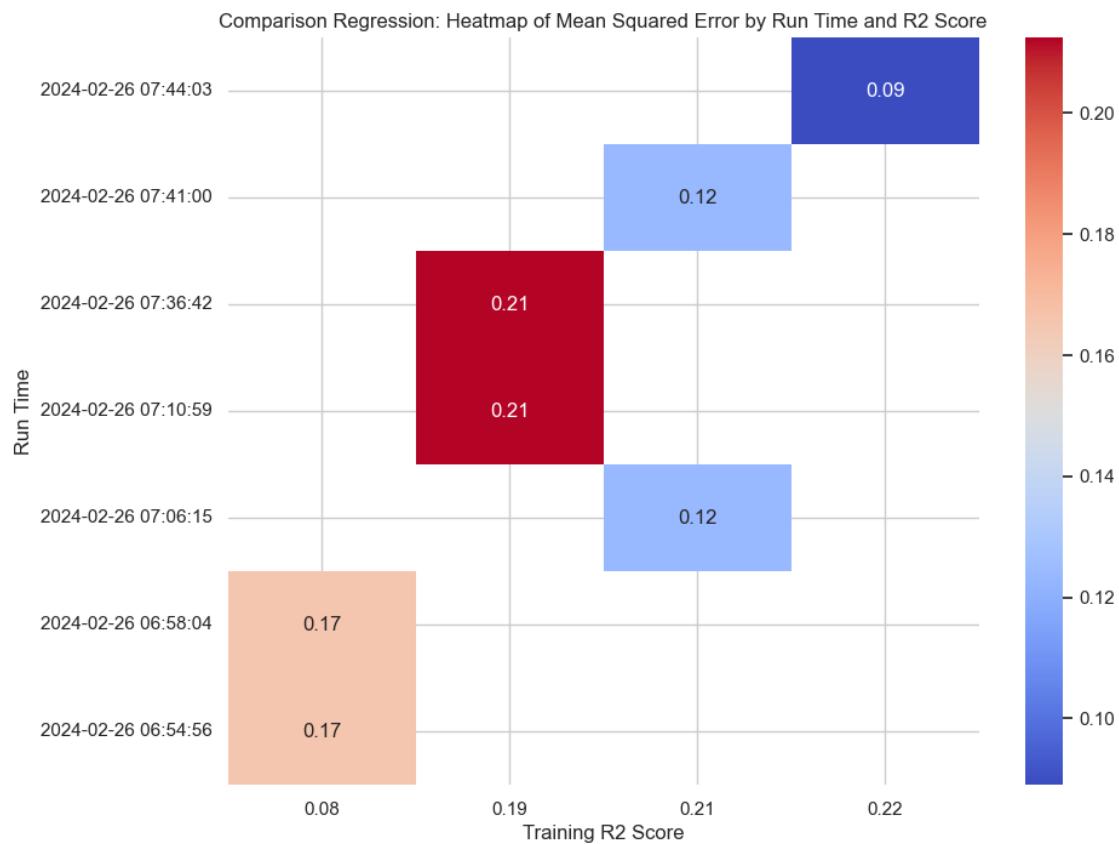
7 Heatmap of MLflow Regression runs



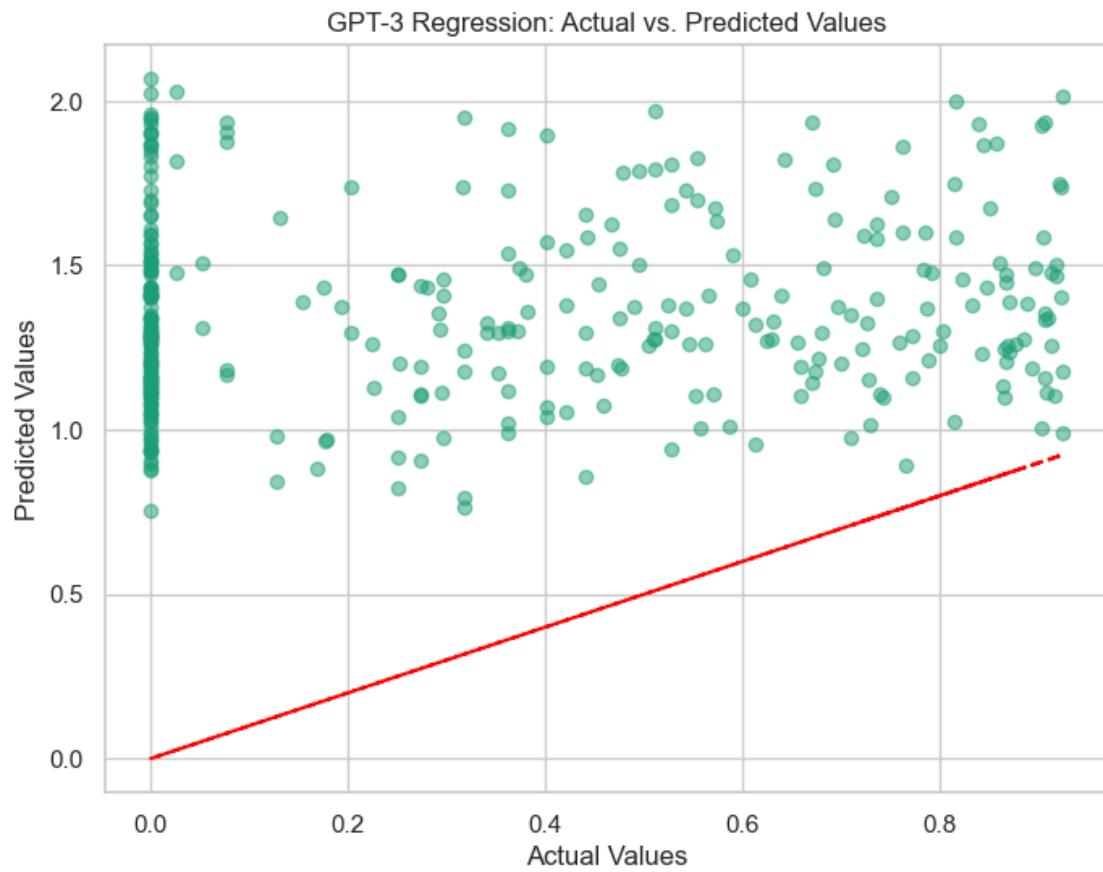


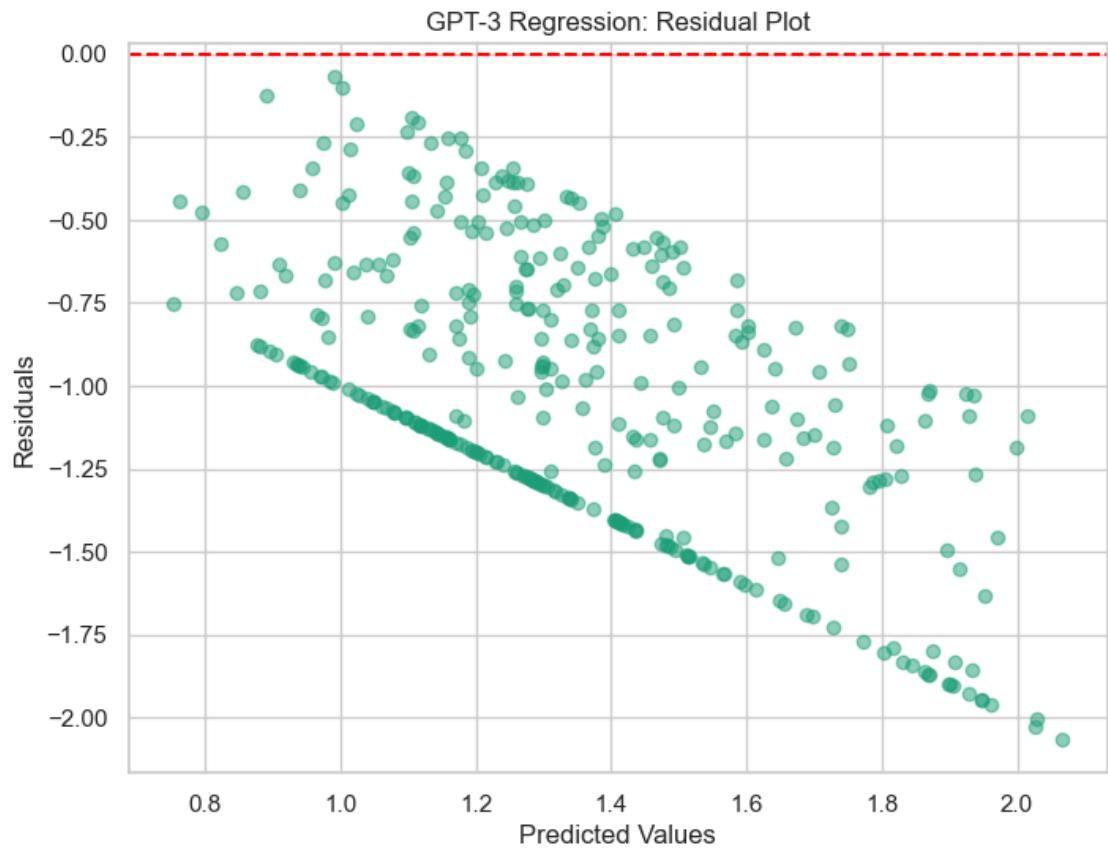




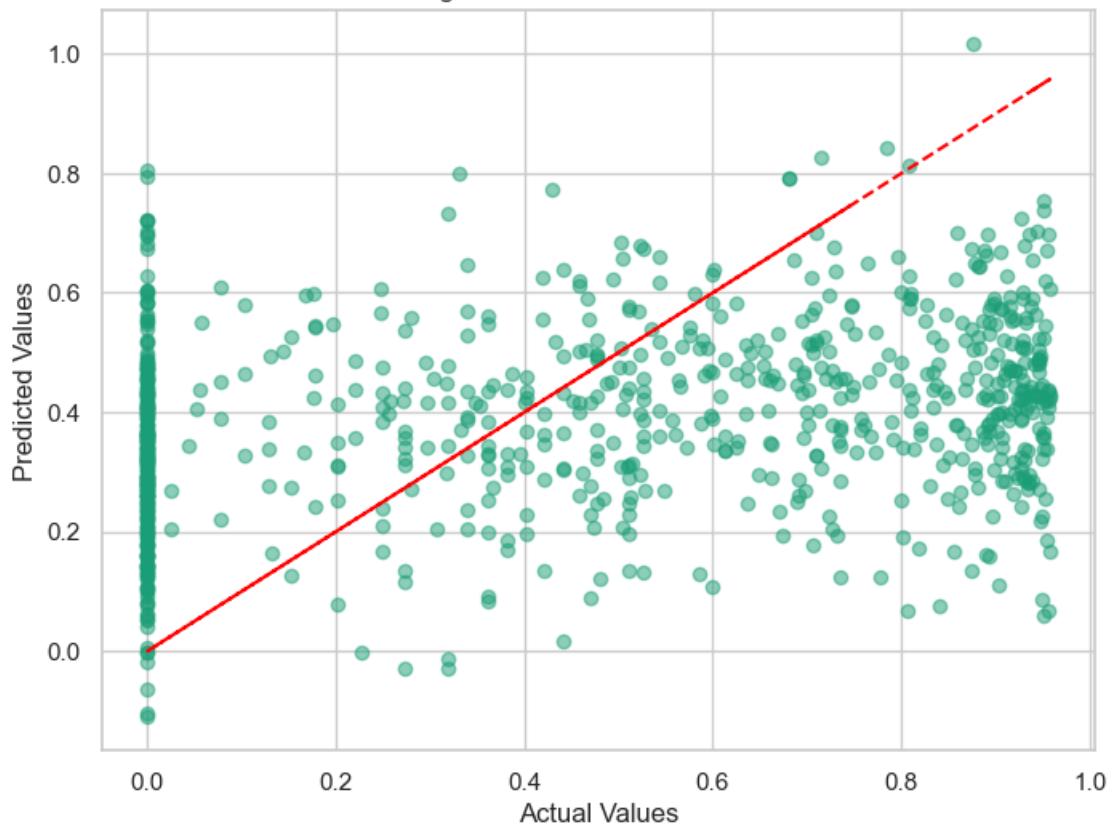


8 Feature Plots for Regression Models

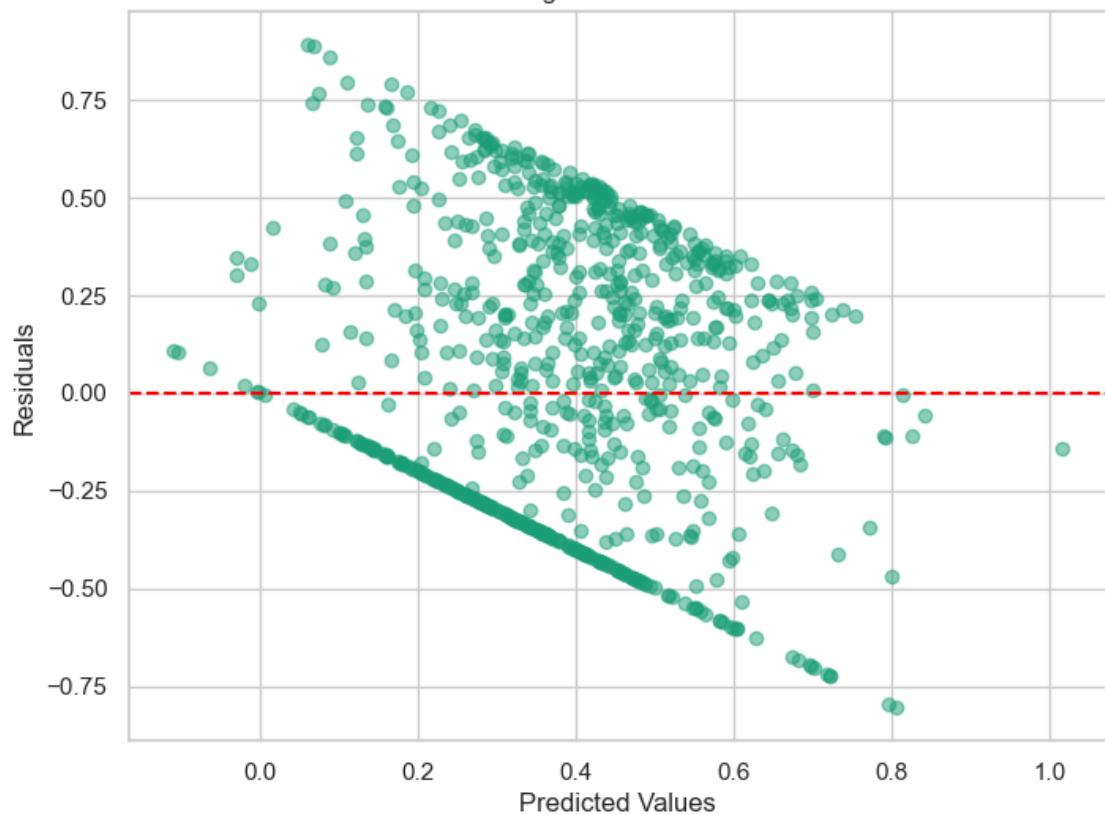




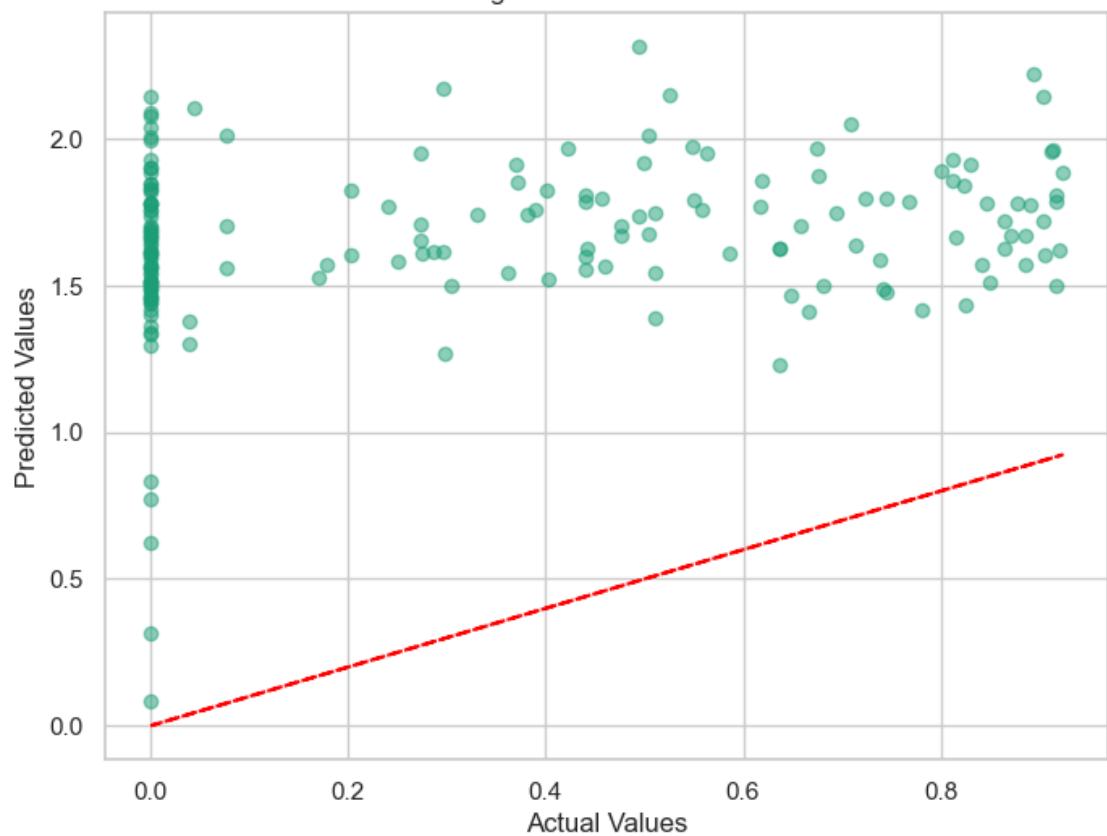
GPT-4 Regression: Actual vs. Predicted Values

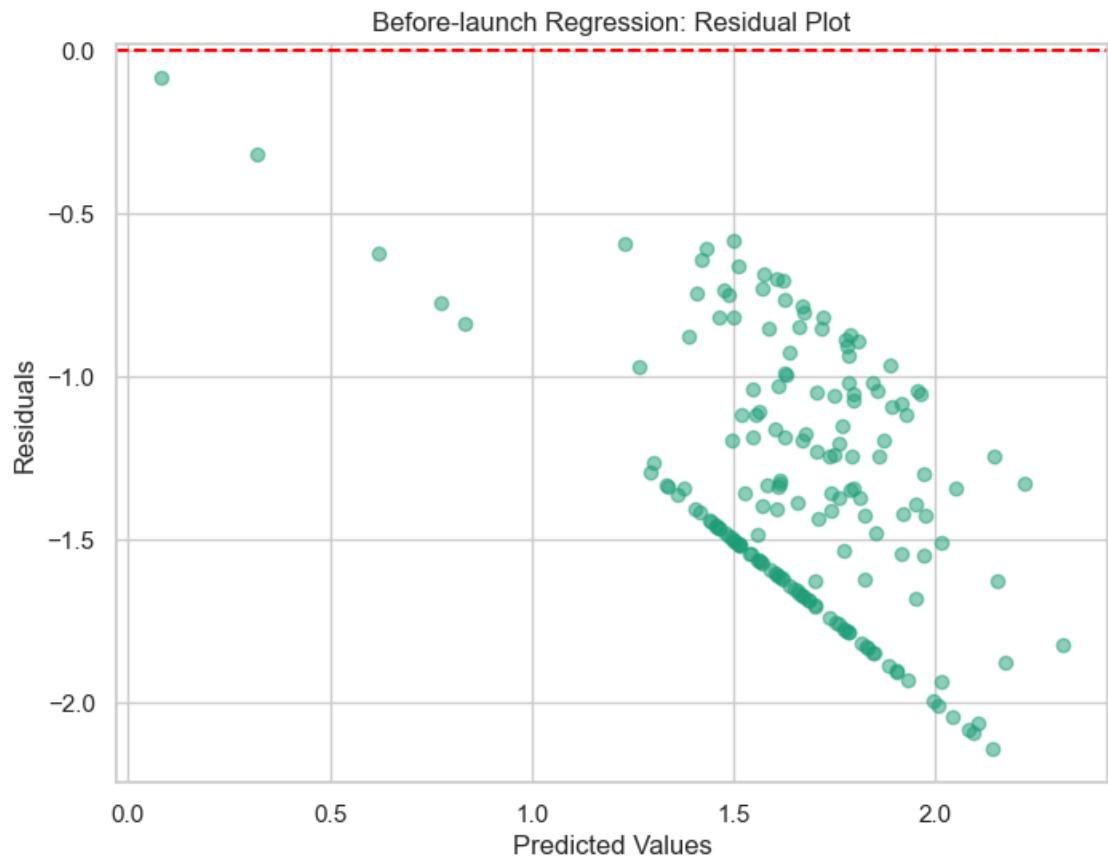


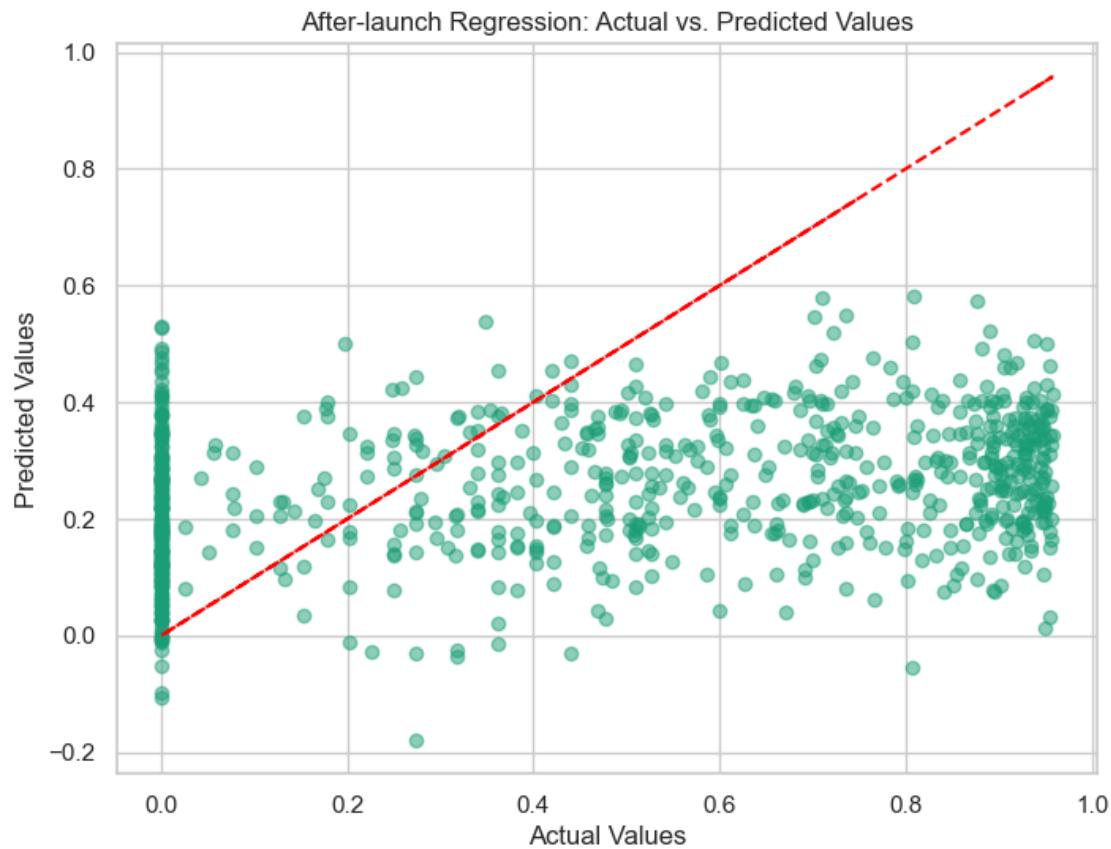
GPT-4 Regression: Residual Plot

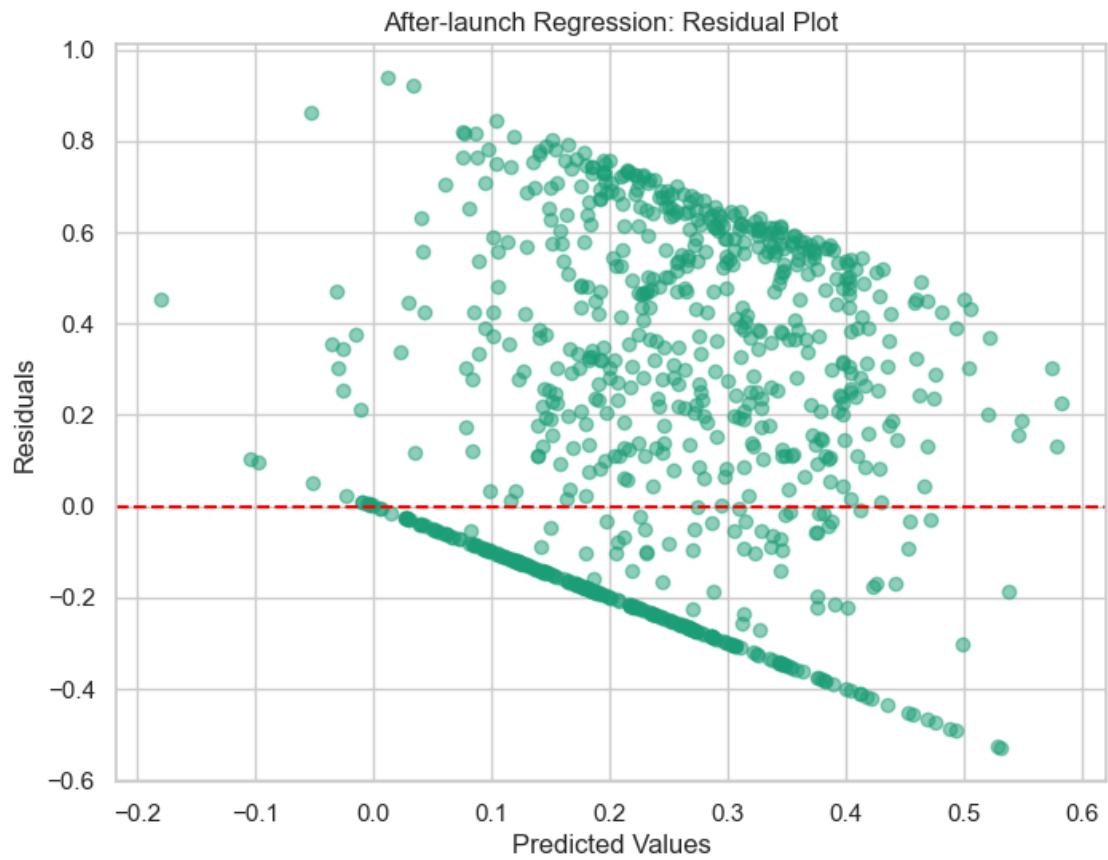


Before-launch Regression: Actual vs. Predicted Values

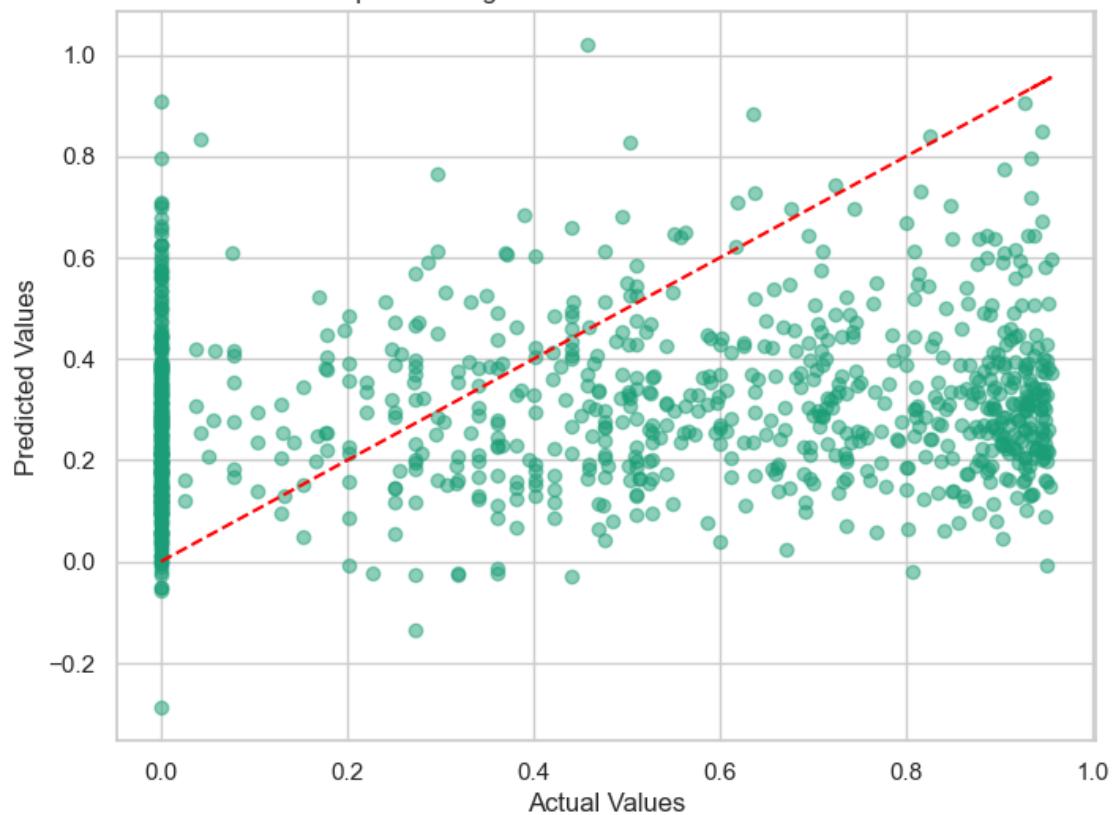




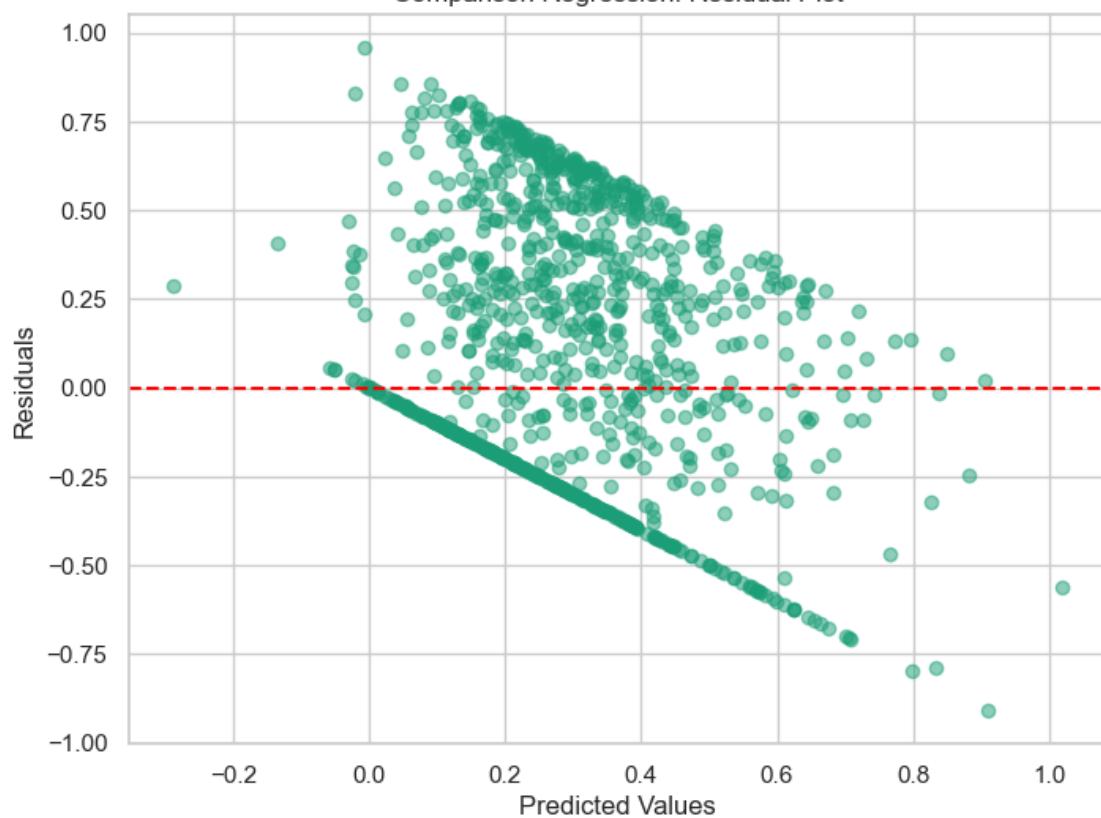


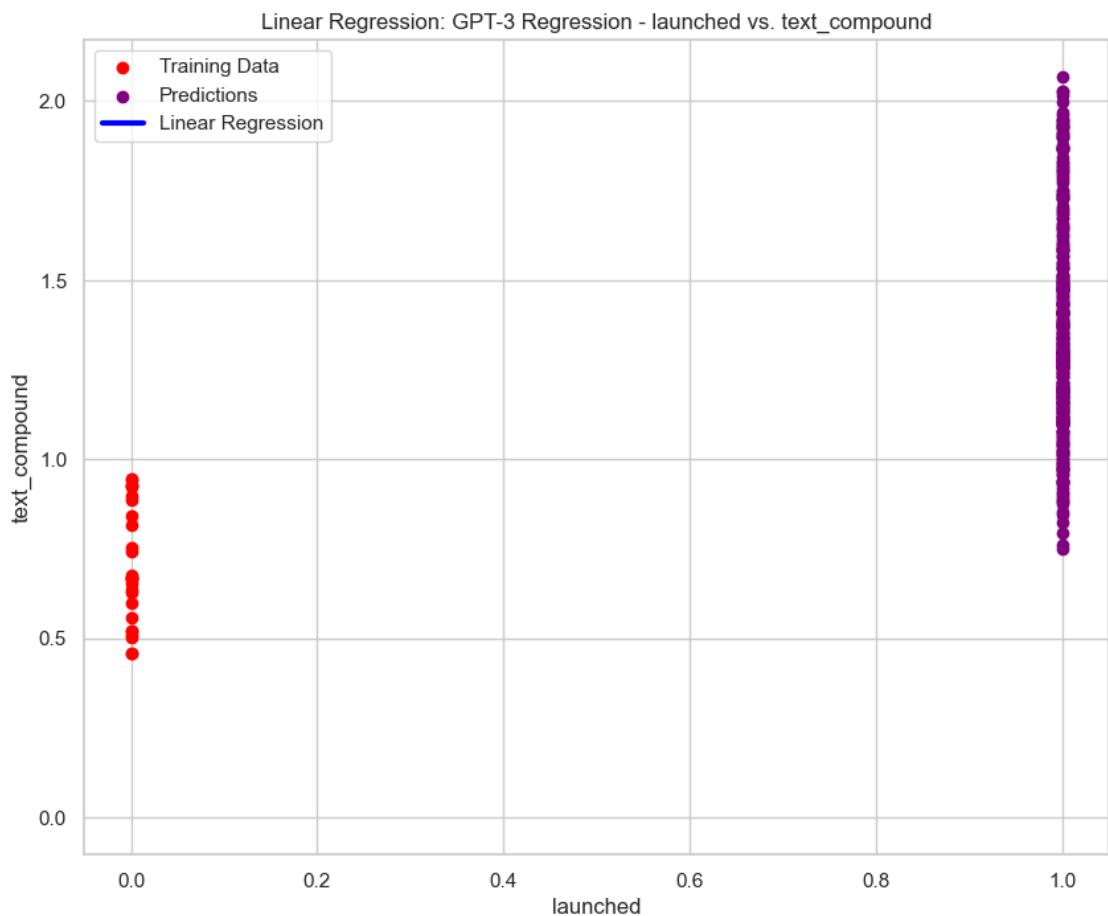


Comparison Regression: Actual vs. Predicted Values

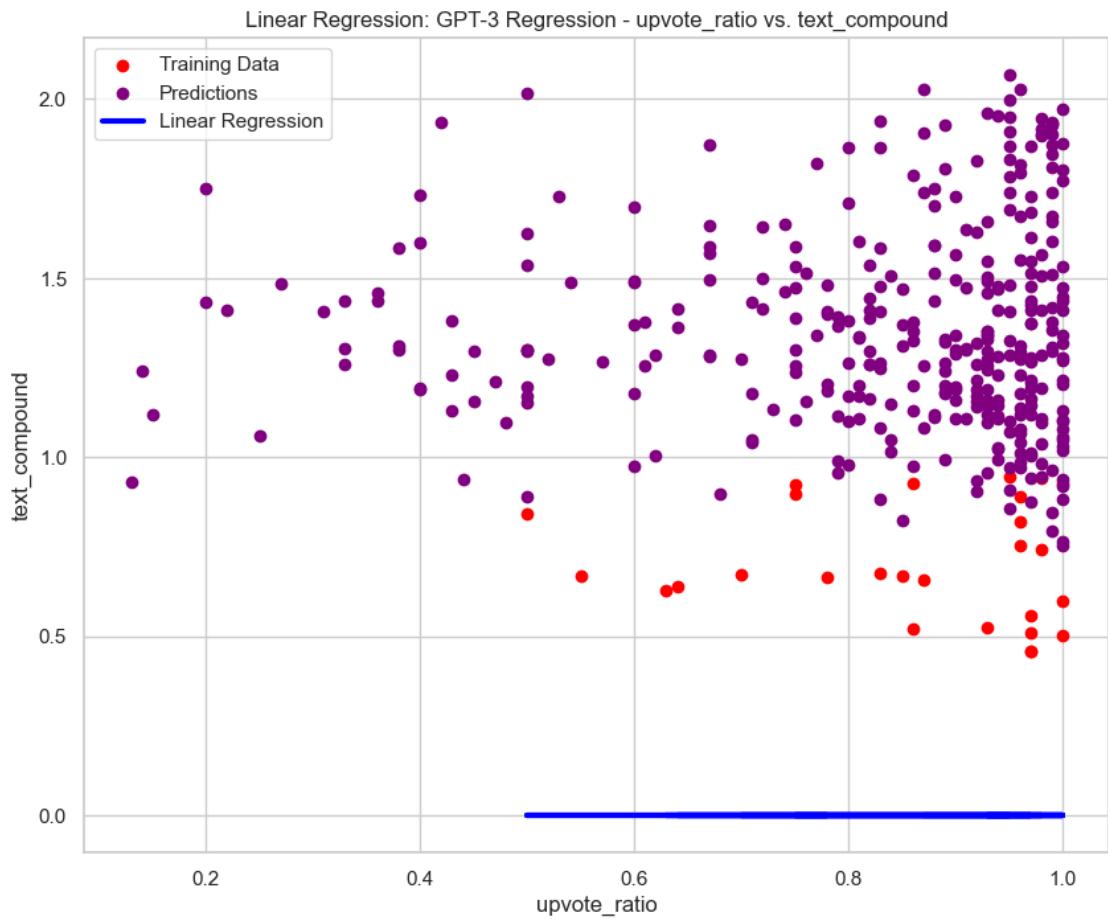


Comparison Regression: Residual Plot

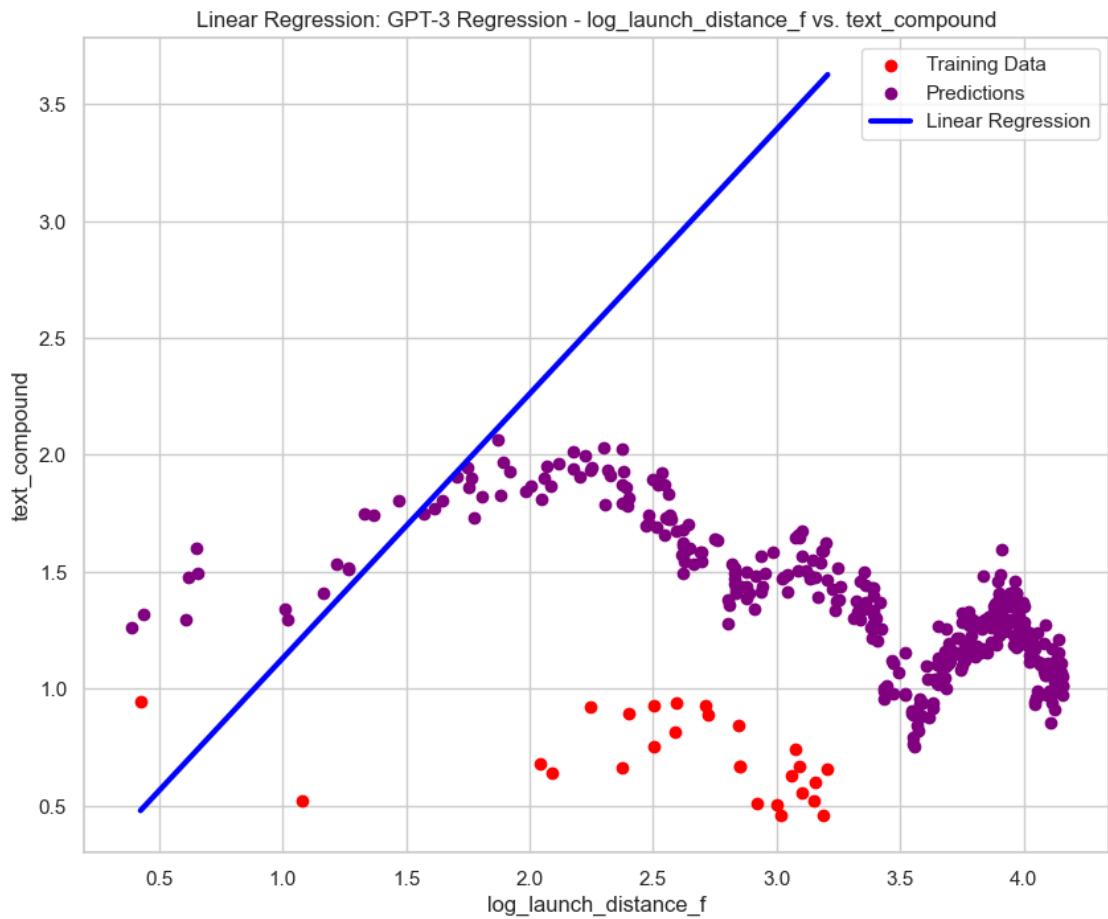




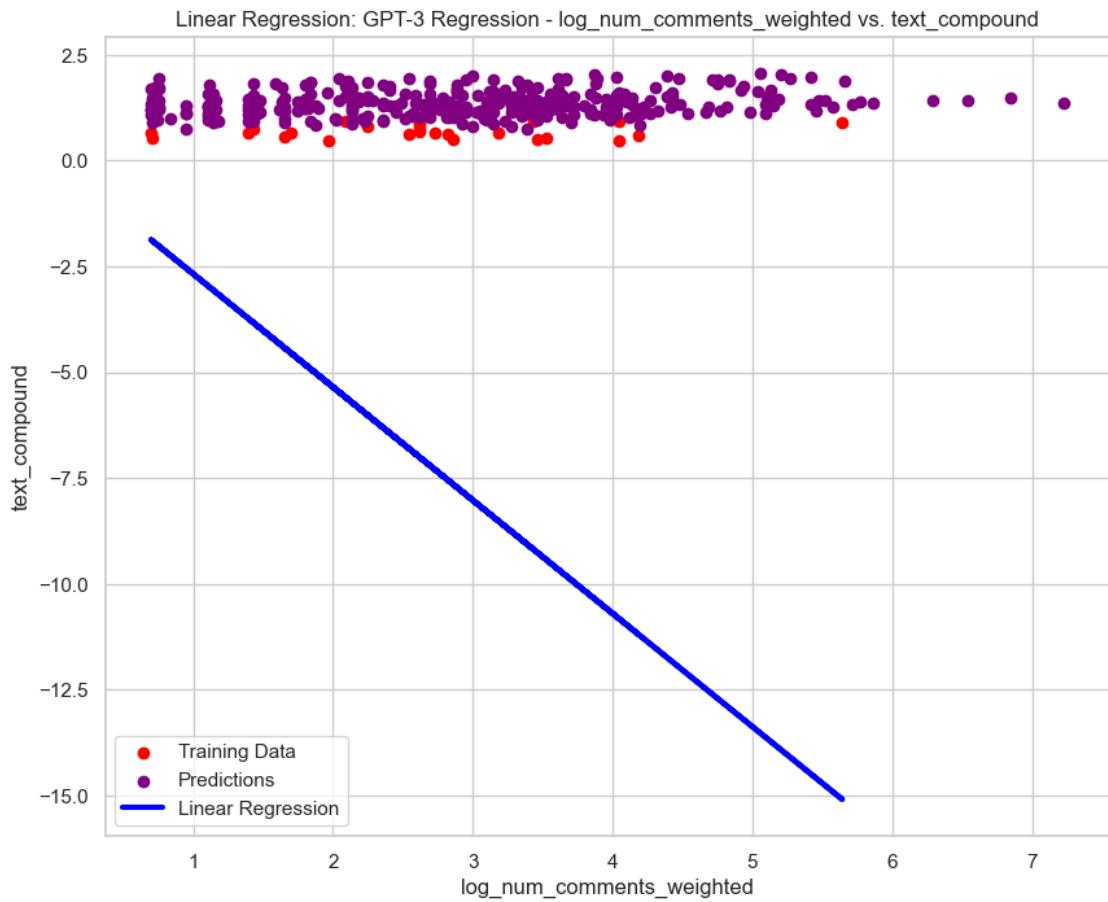
launched=0.03858679253065731



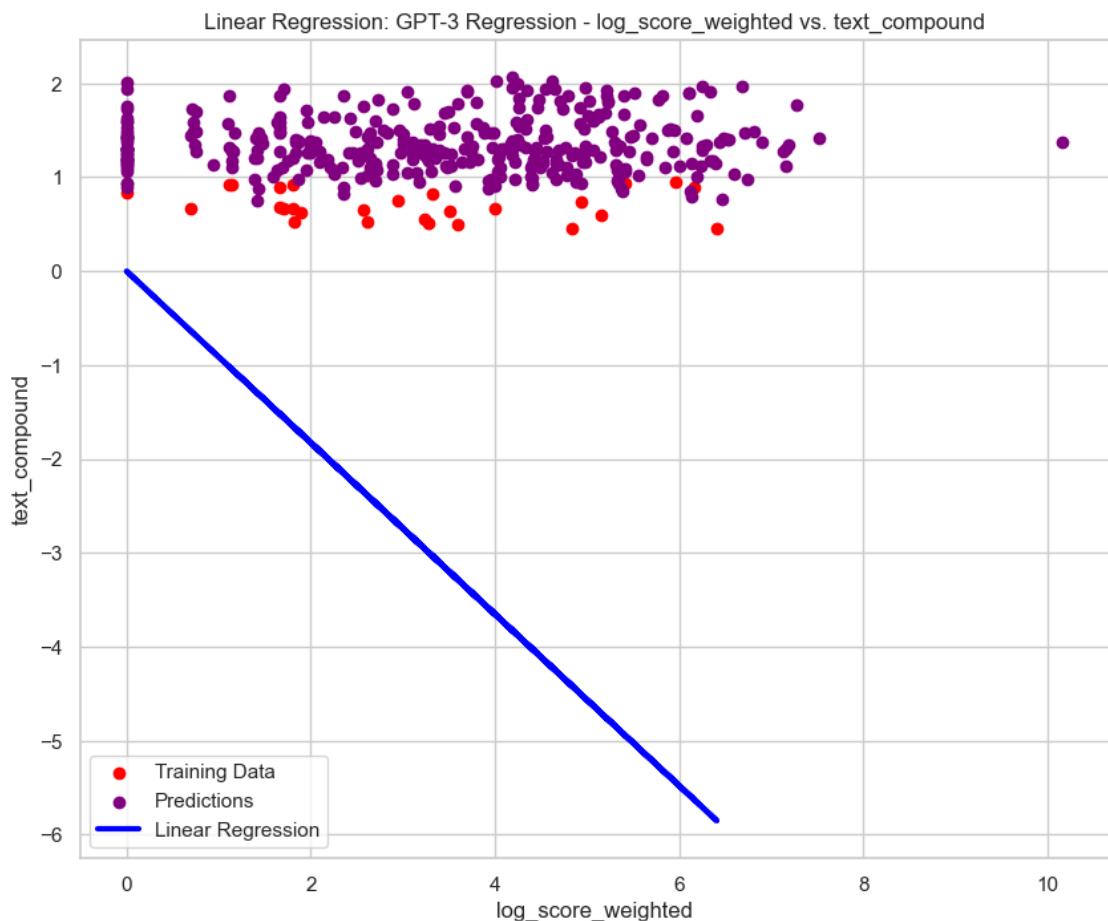
upvote_ratio=7.216449660063518e-16



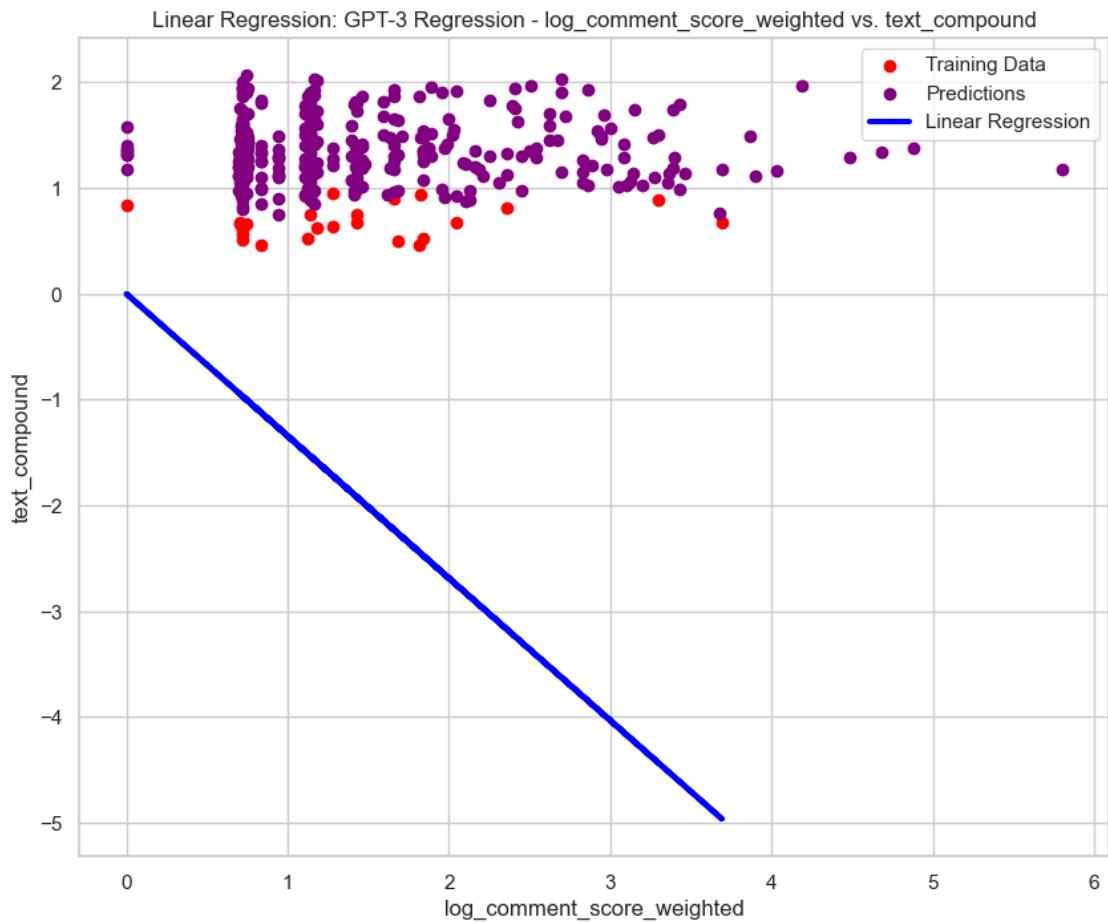
`log_launch_distance_f=1.1307324243024075`



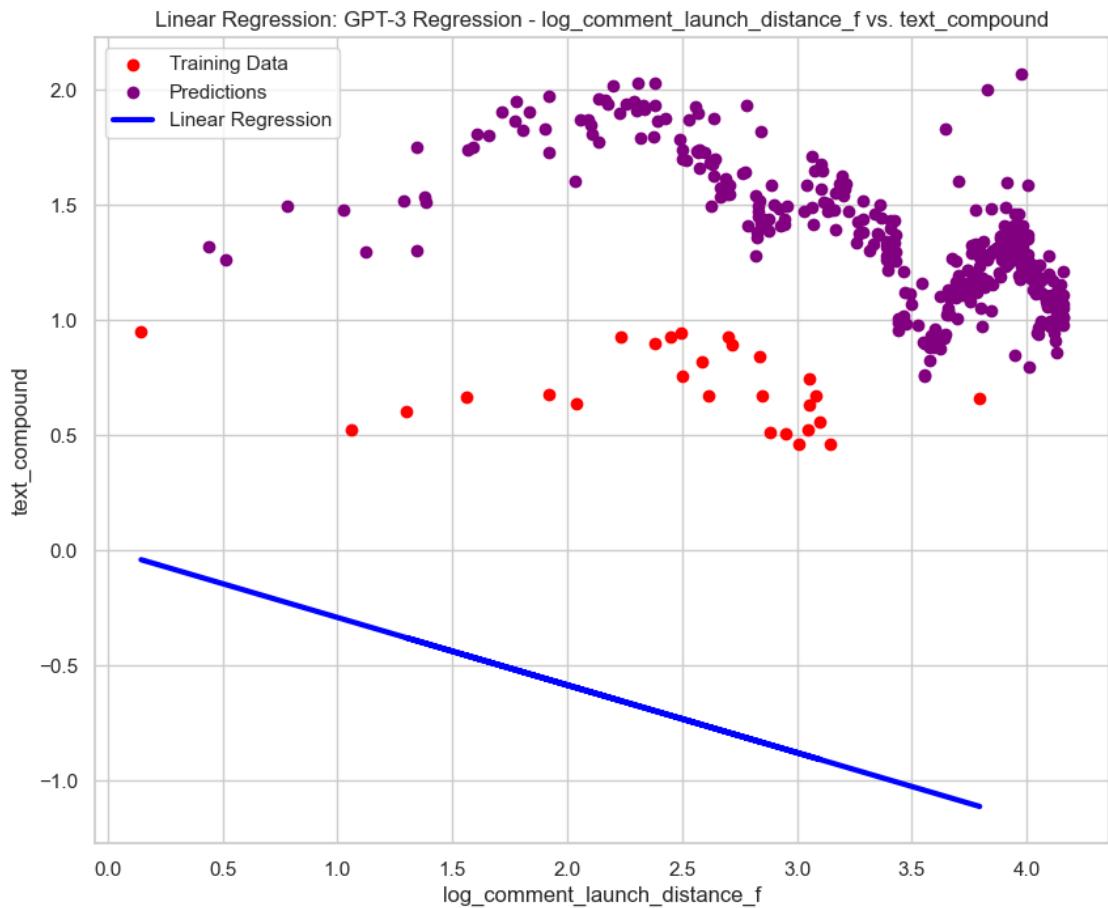
log_num_comments_weighted=-2.675335735982713



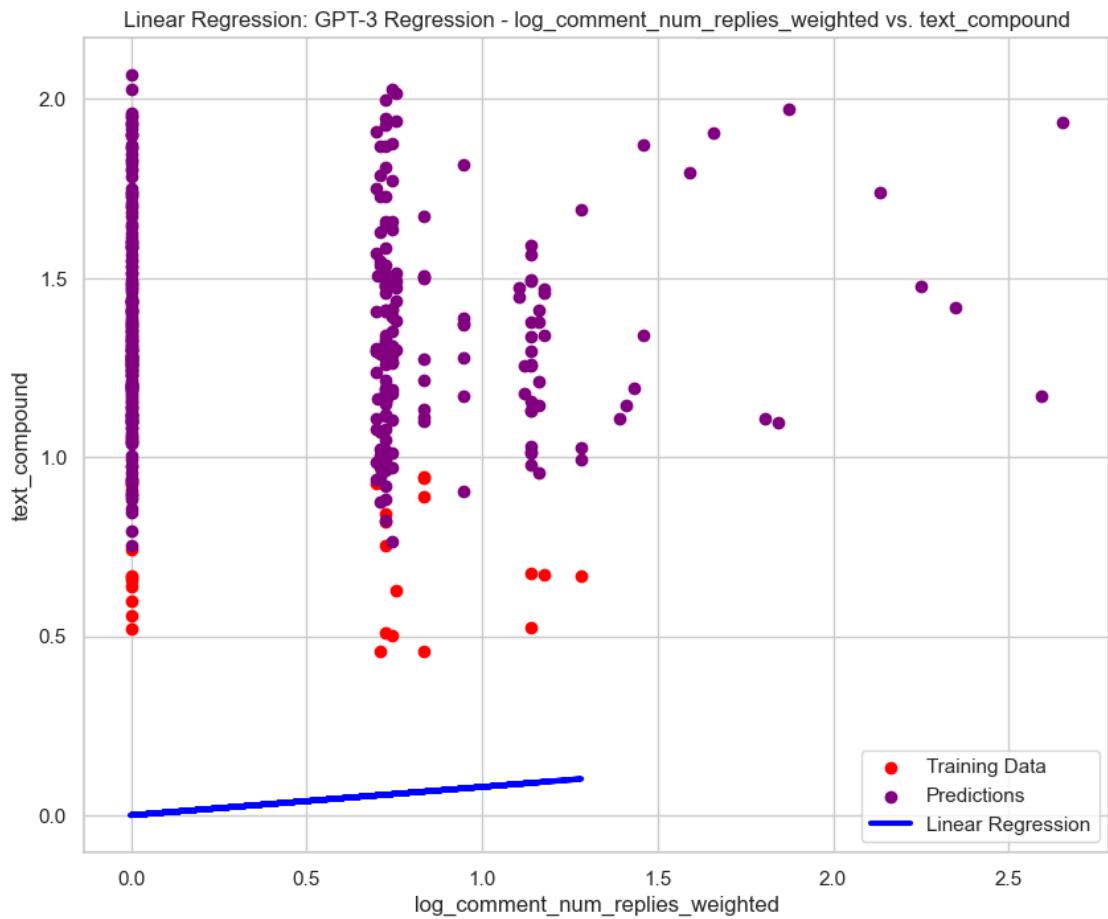
`log_score_weighted=-0.9129480729134403`



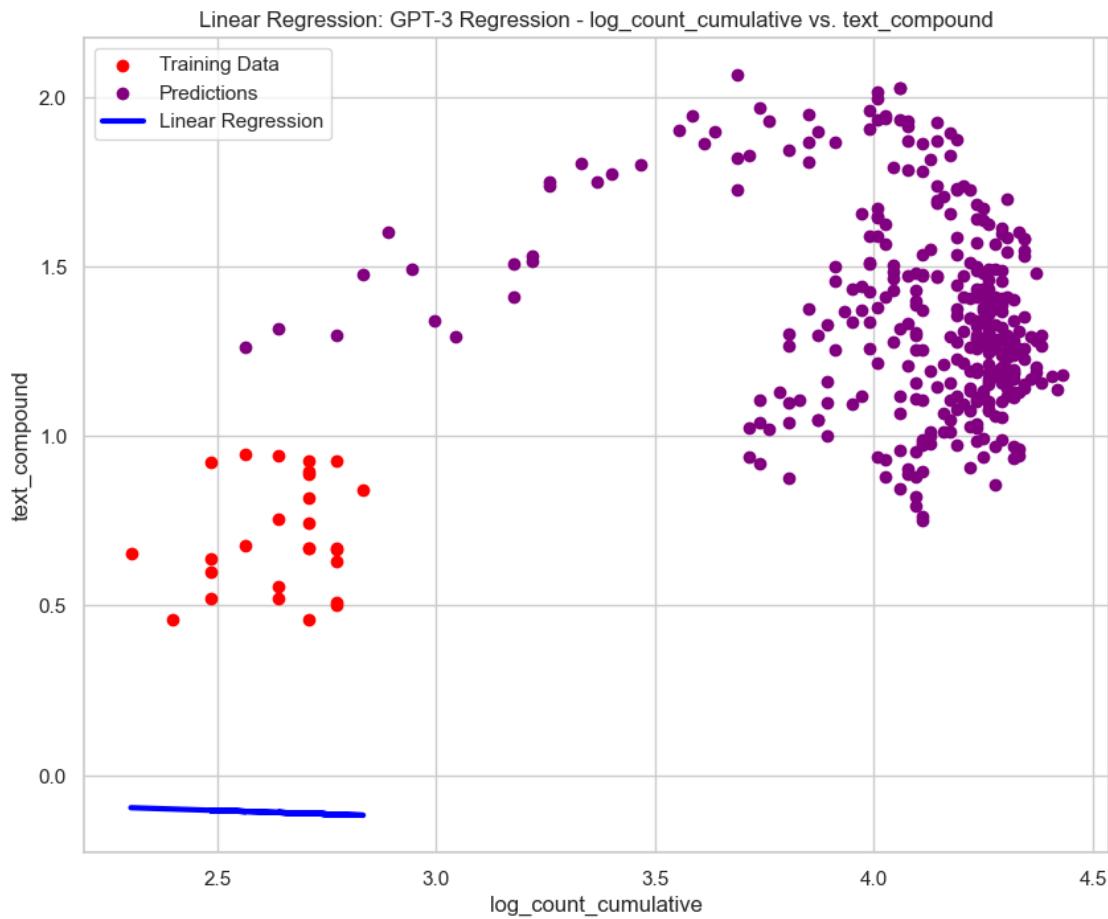
log_comment_score_weighted=-1.3436140735918558



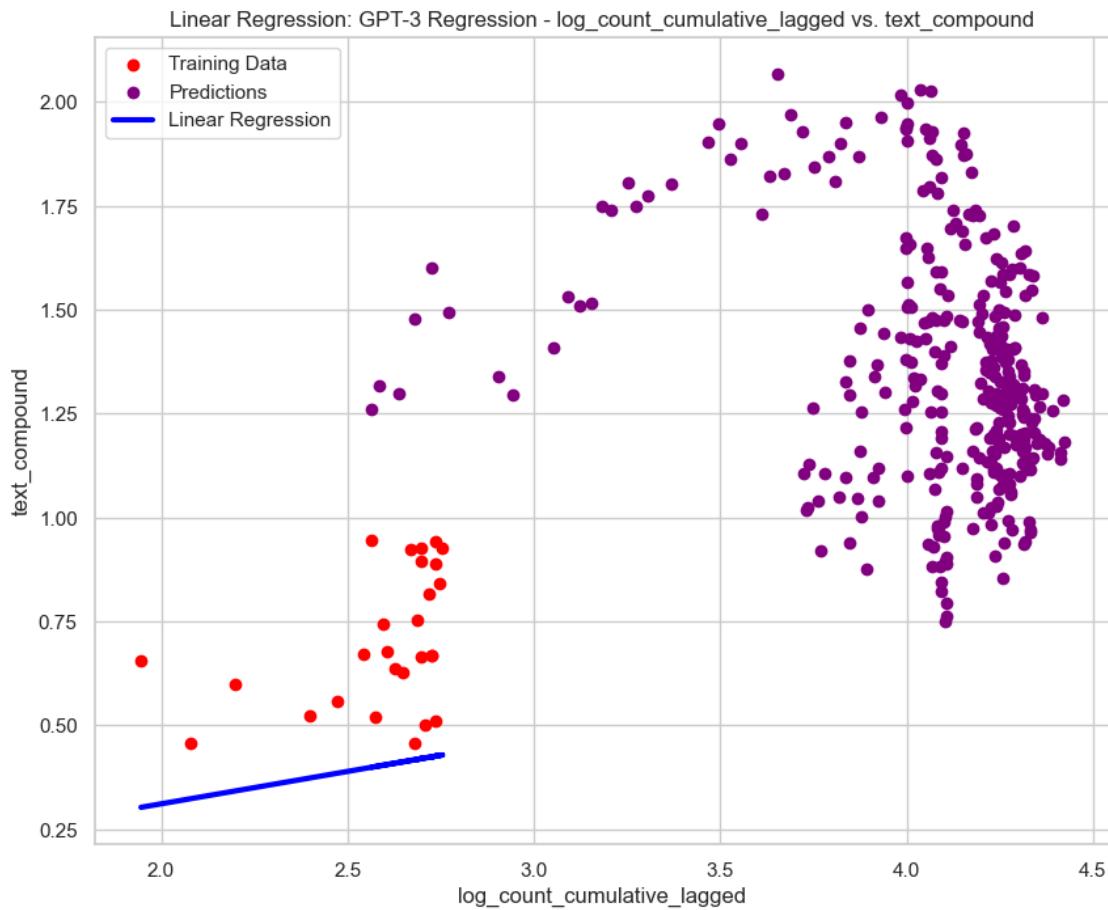
log_comment_launch_distance_f=-0.29368908950744244



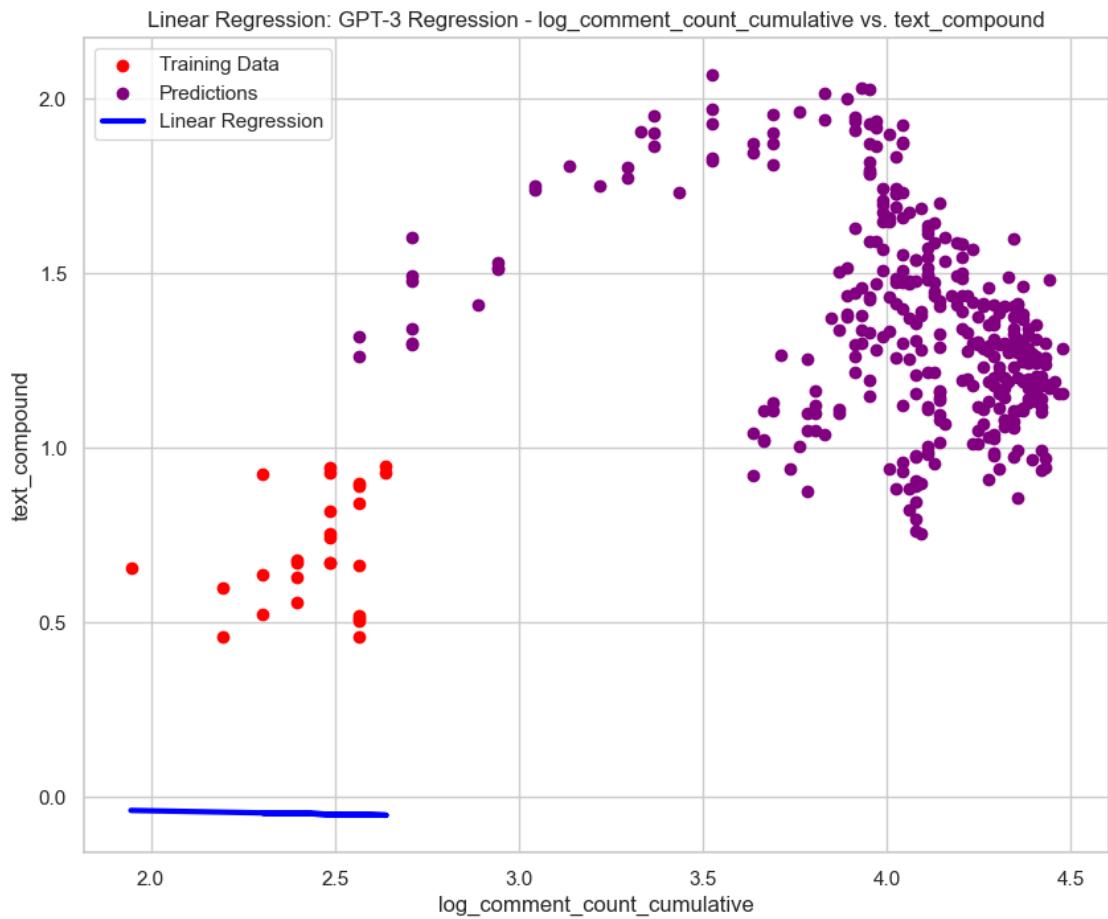
log_comment_num_replies_weighted=0.07913892483406158



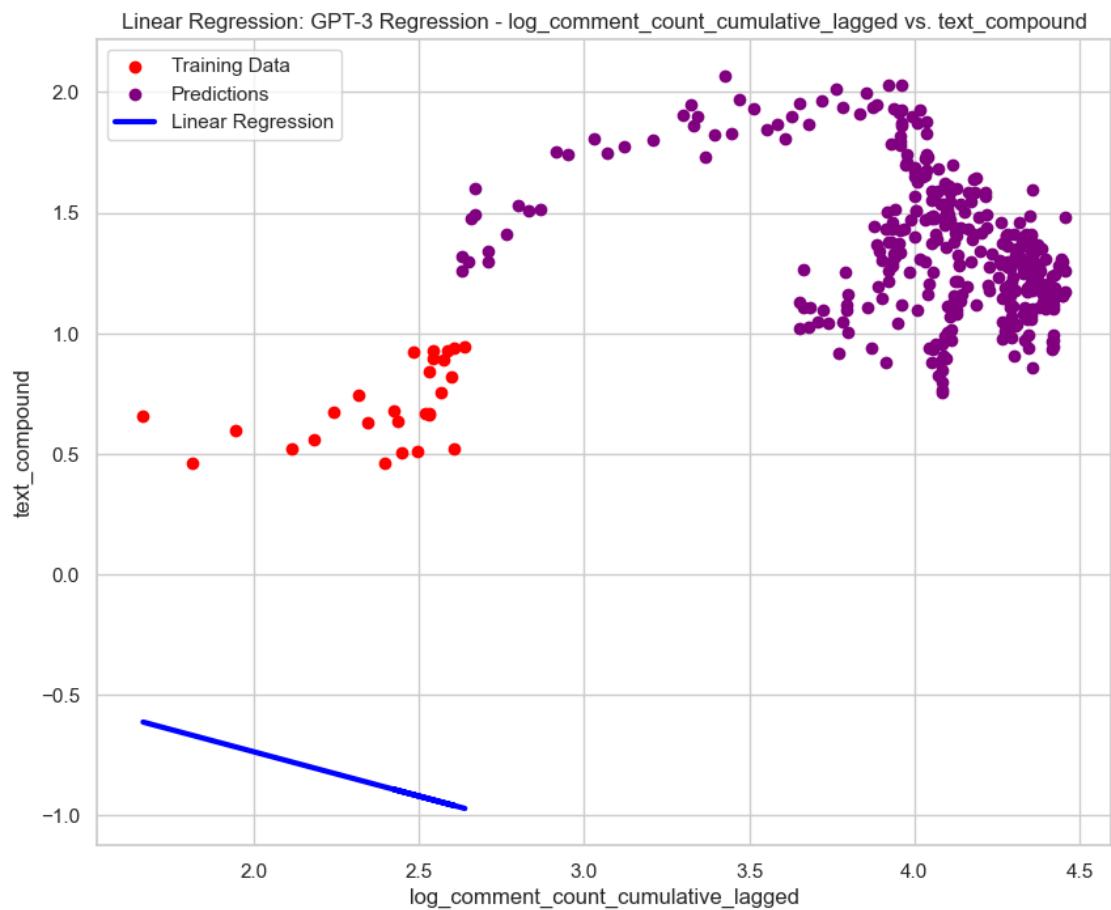
`log_count_cumulative=-0.04186038318727184`



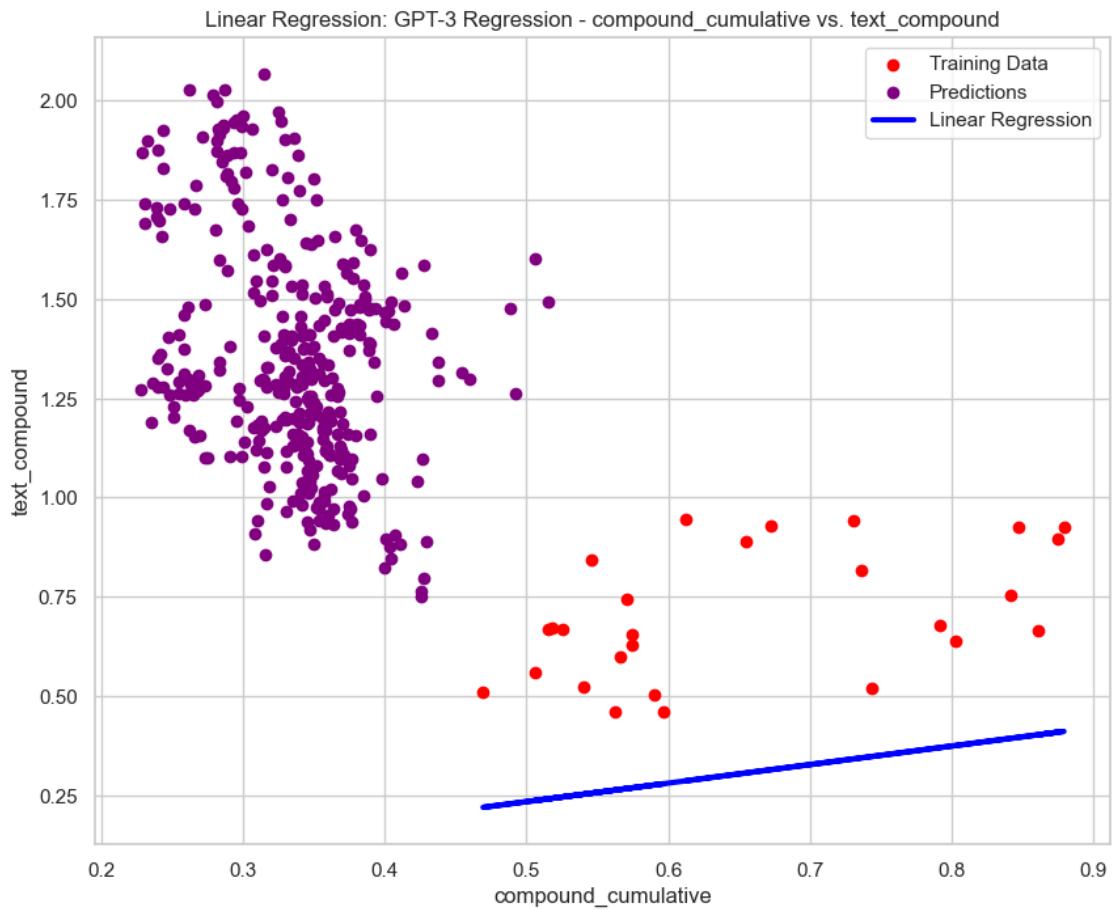
`log_count_cumulative_lagged=0.15579896373676838`



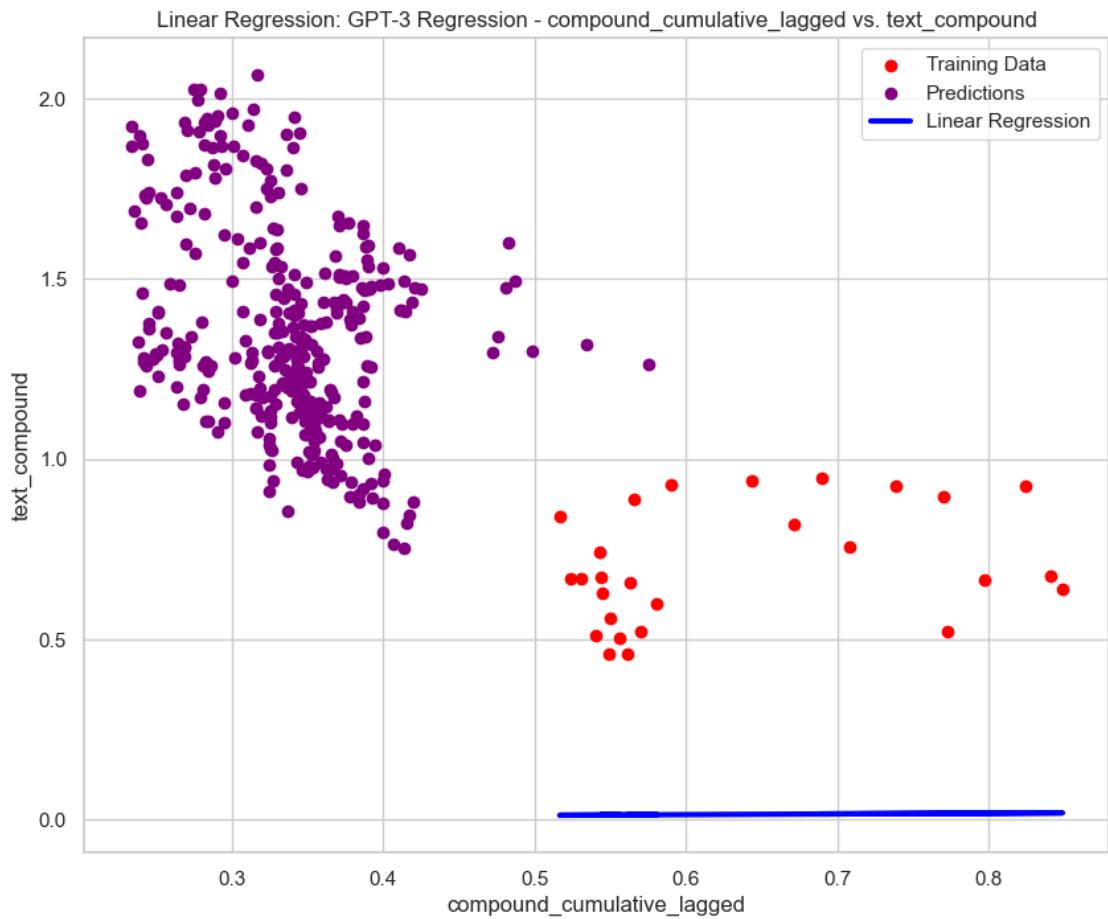
`log_comment_count_cumulative=-0.019941574615981932`



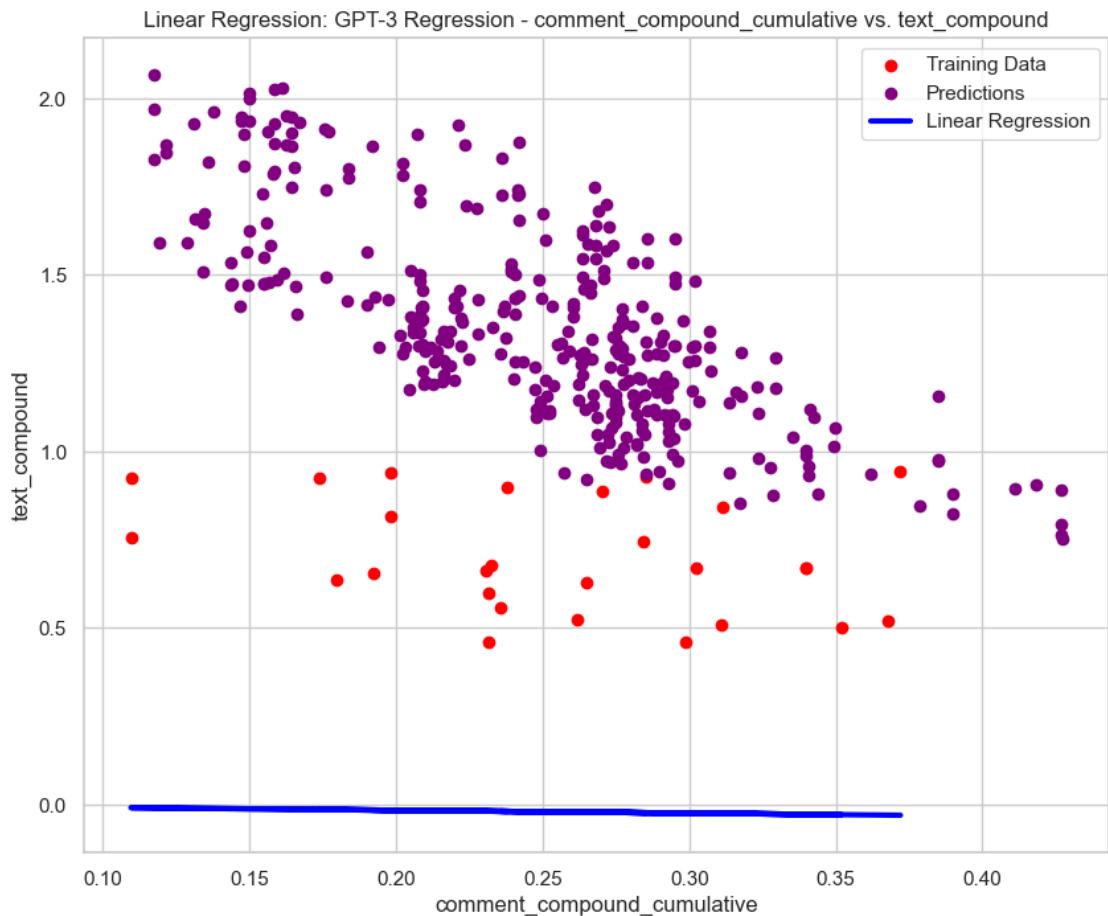
log_comment_count_cumulative_lagged=-0.3677388623725115



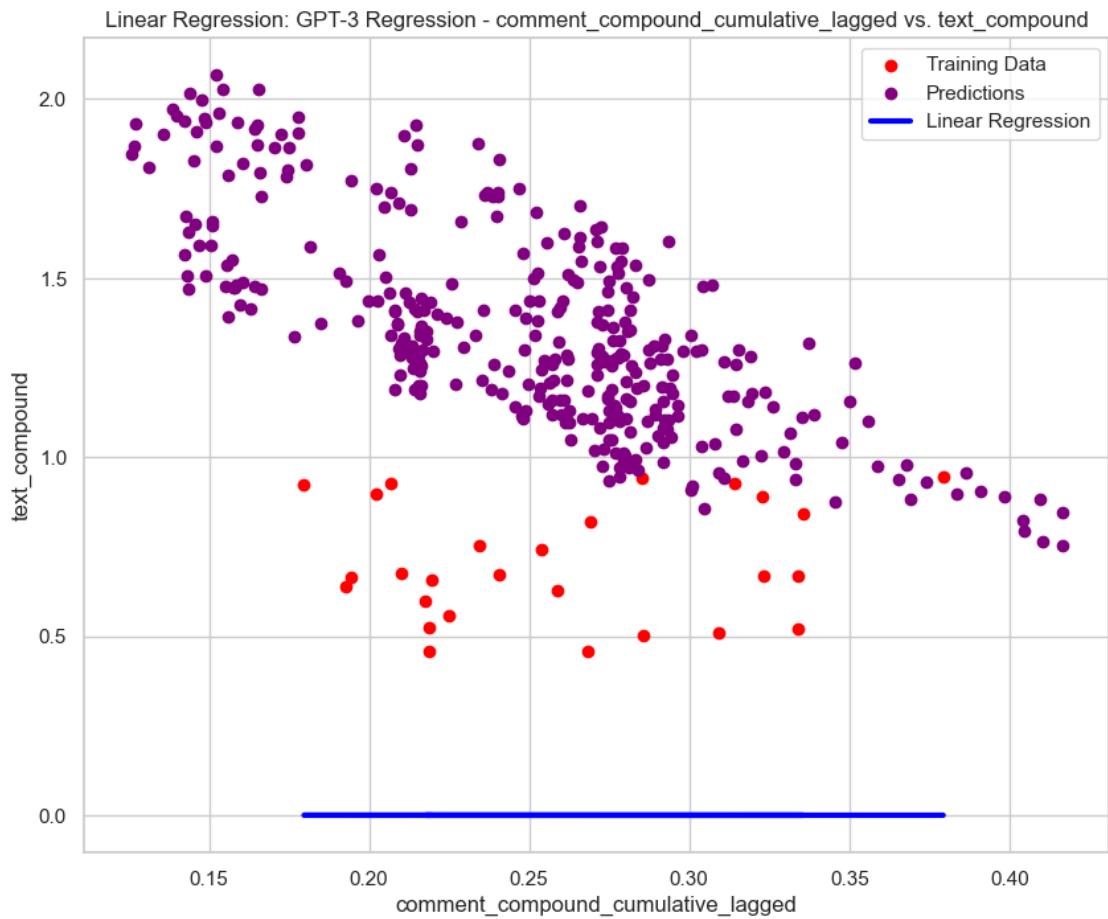
compound_cumulative=0.46712315647027425



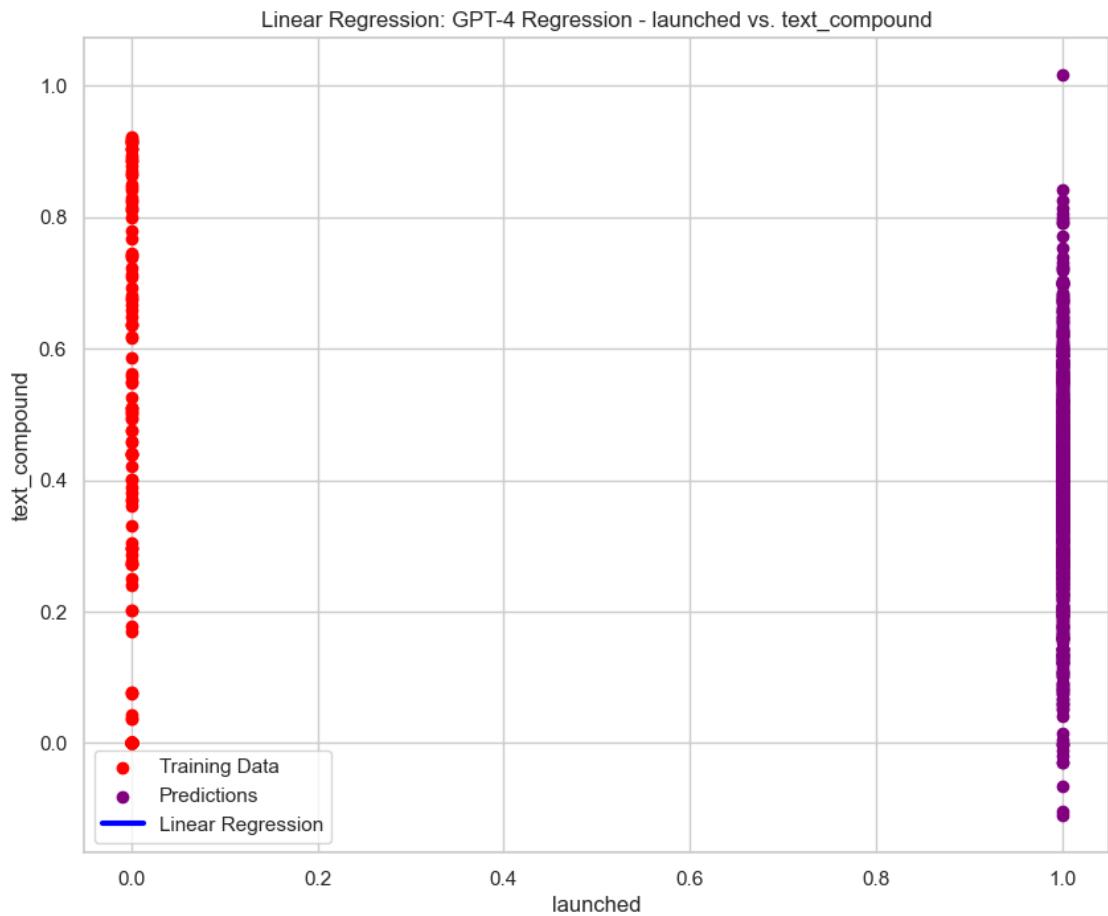
compound_cumulative_lagged=0.020234890221955737



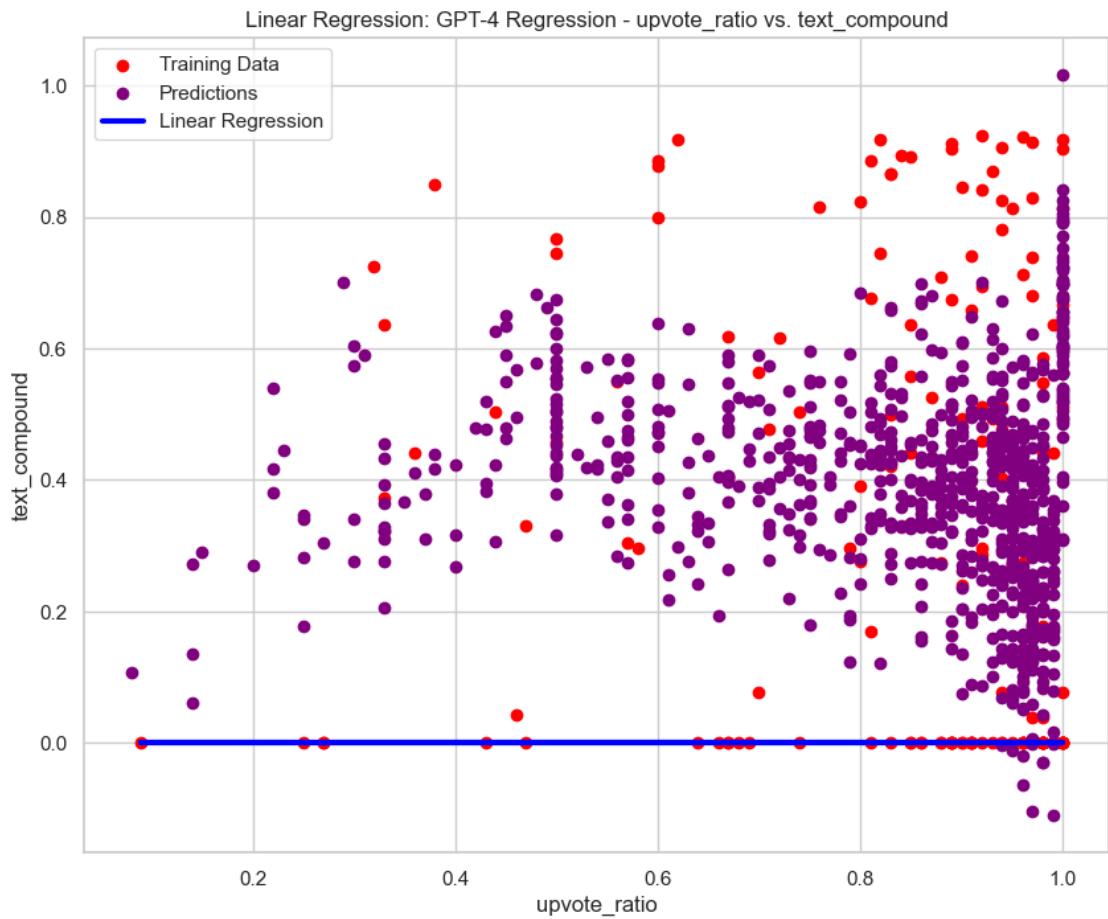
comment_compound_cumulative=-0.08186407607413103



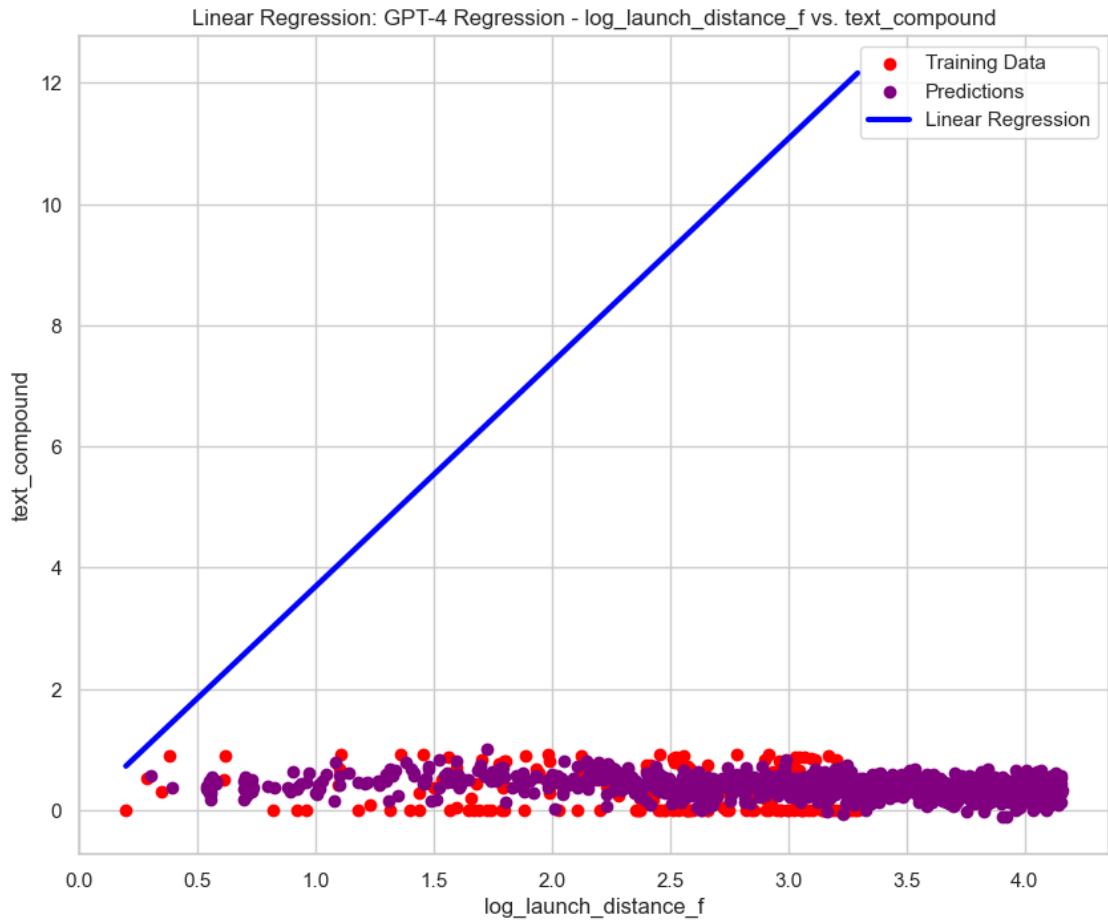
comment_compound_cumulative_lagged=-0.00021174886748573038



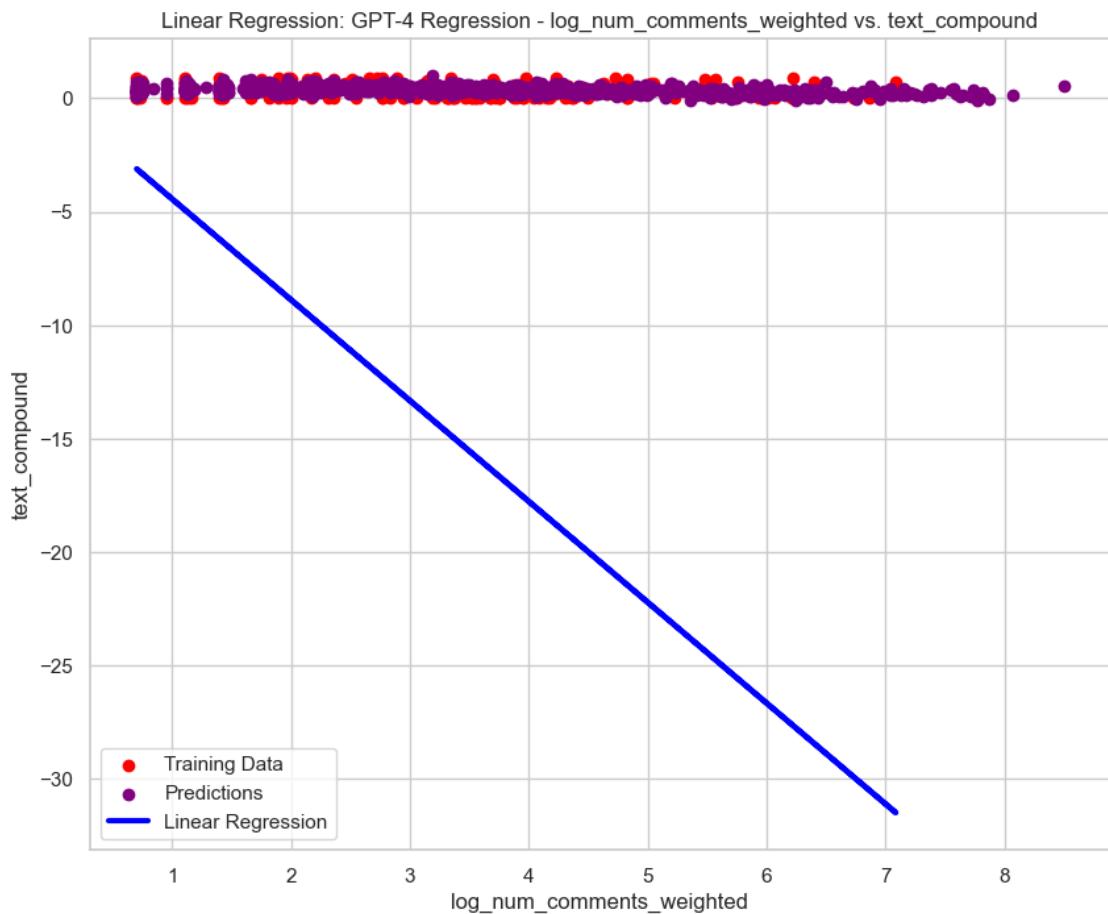
launched=0.6476356633231909



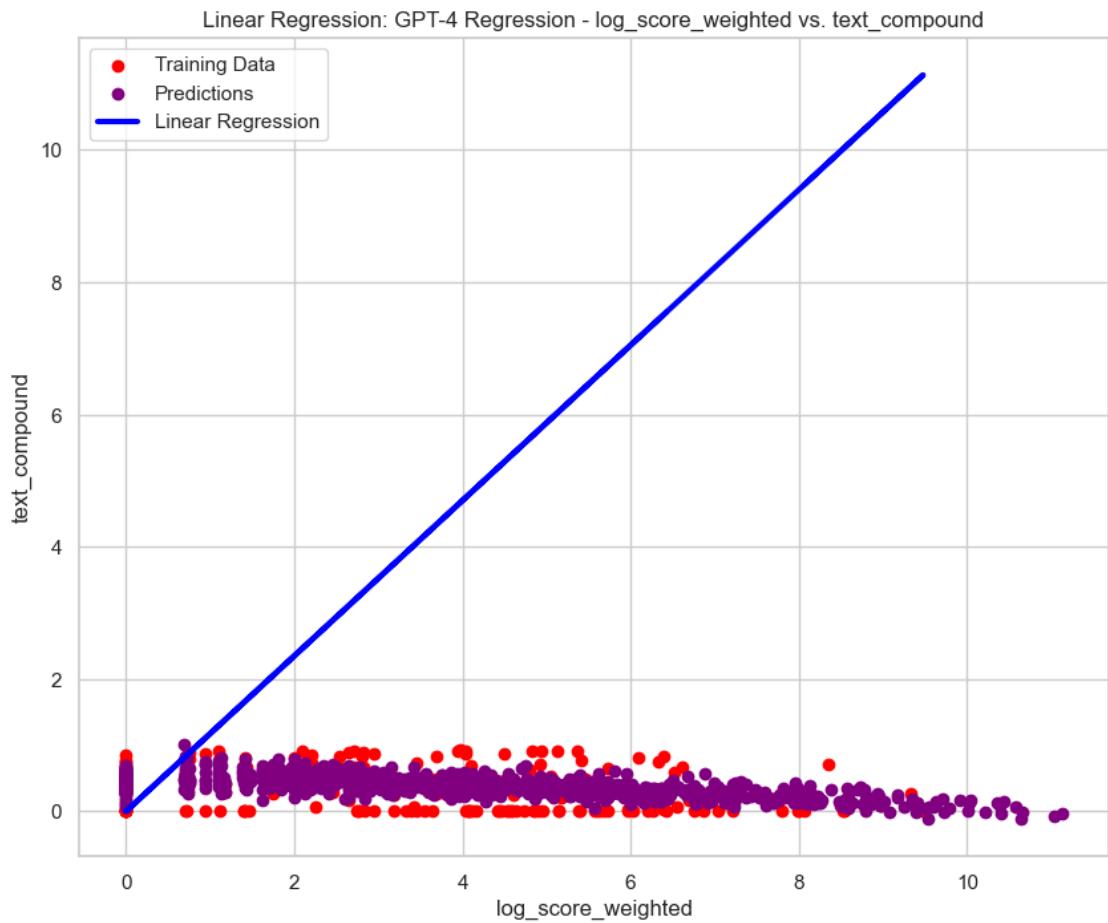
upvote_ratio=-9.492406860545088e-15



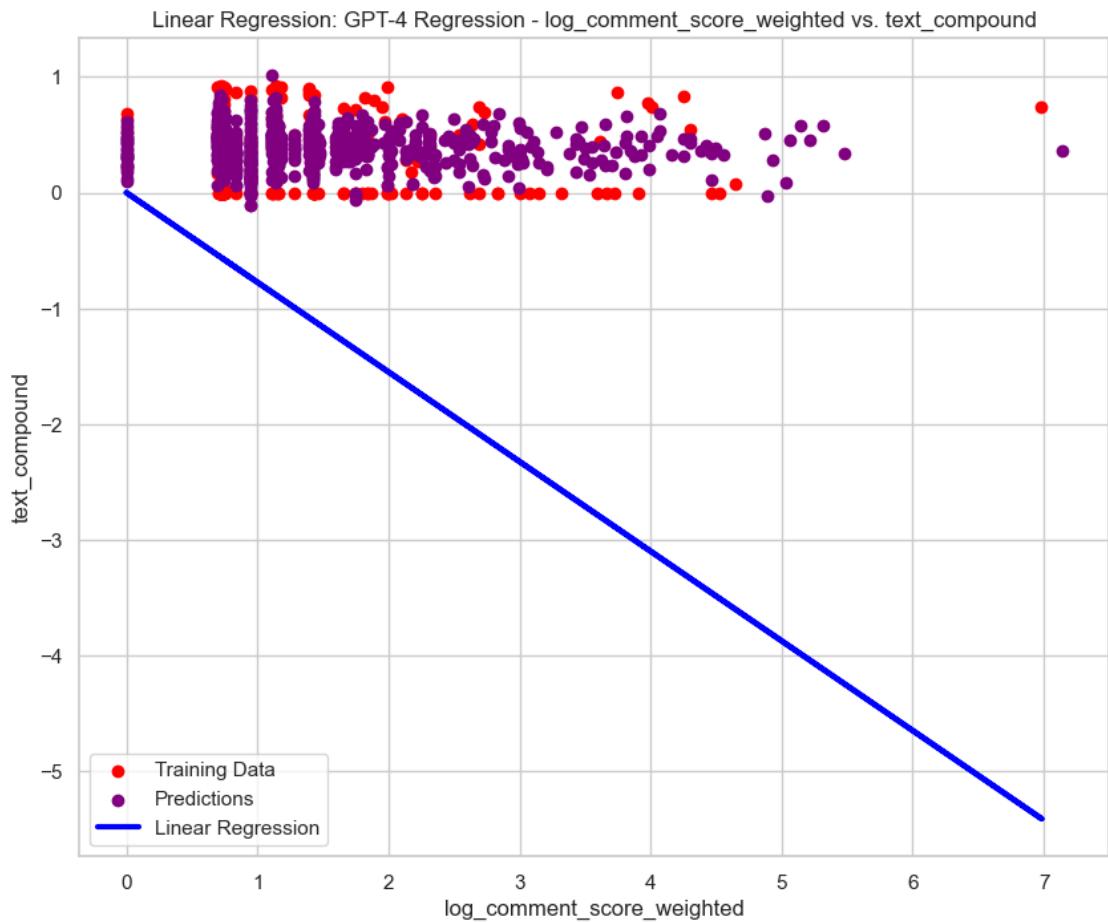
`log_launch_distance_f=3.6939971259187074`



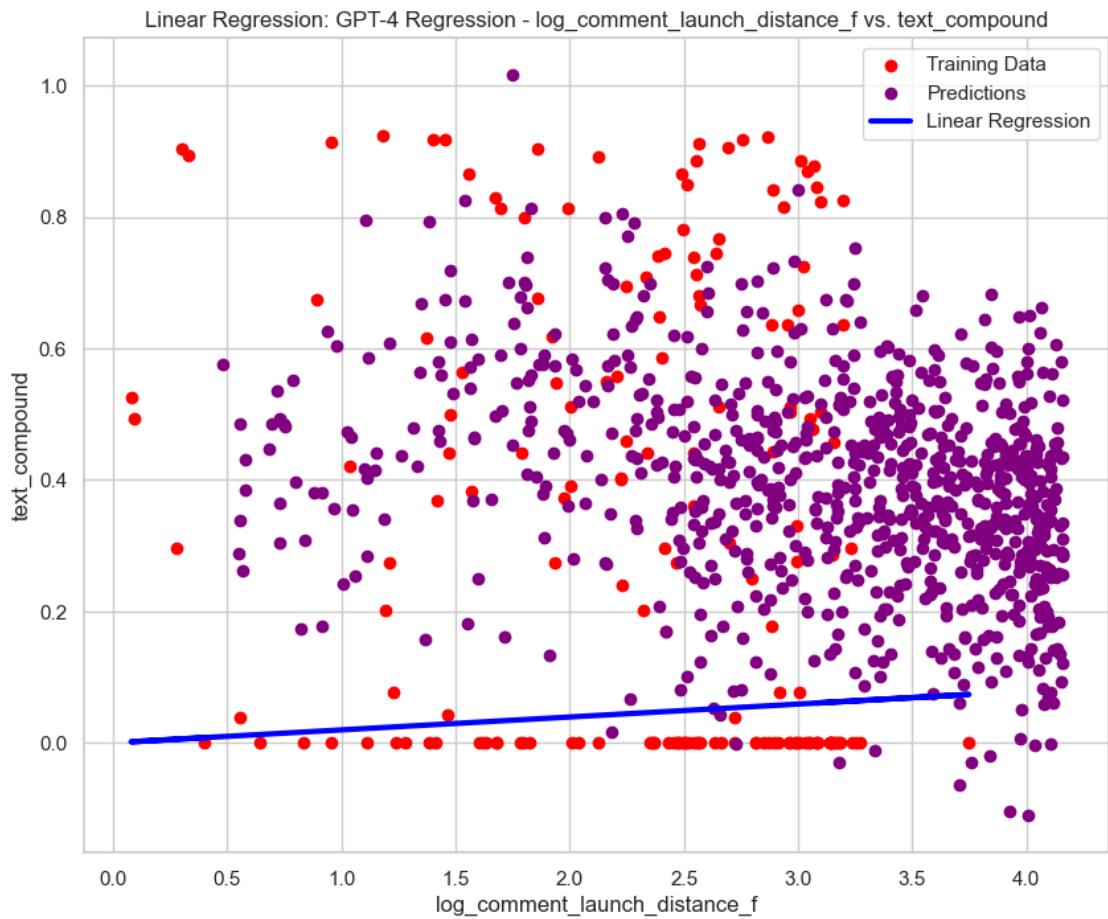
log_num_comments_weighted=-4.446818173658925



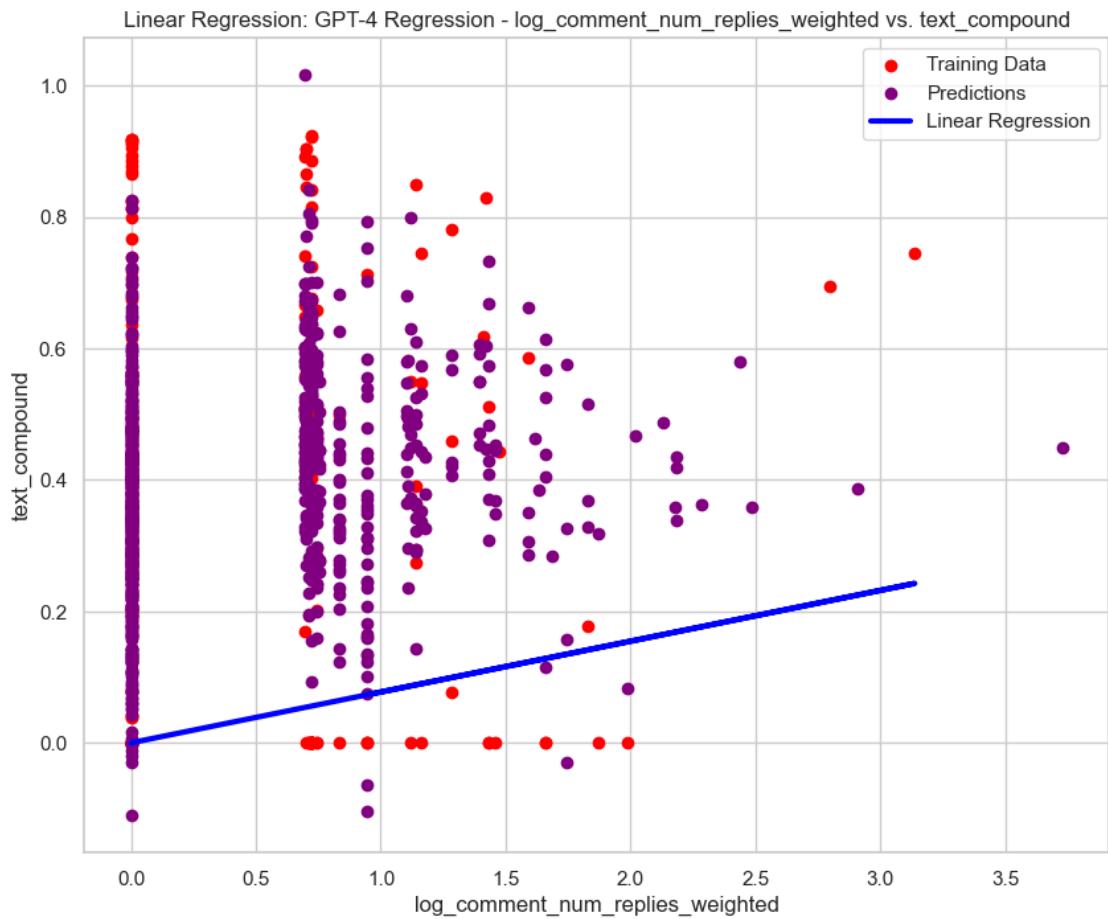
log_score_weighted=1.1744136887832775



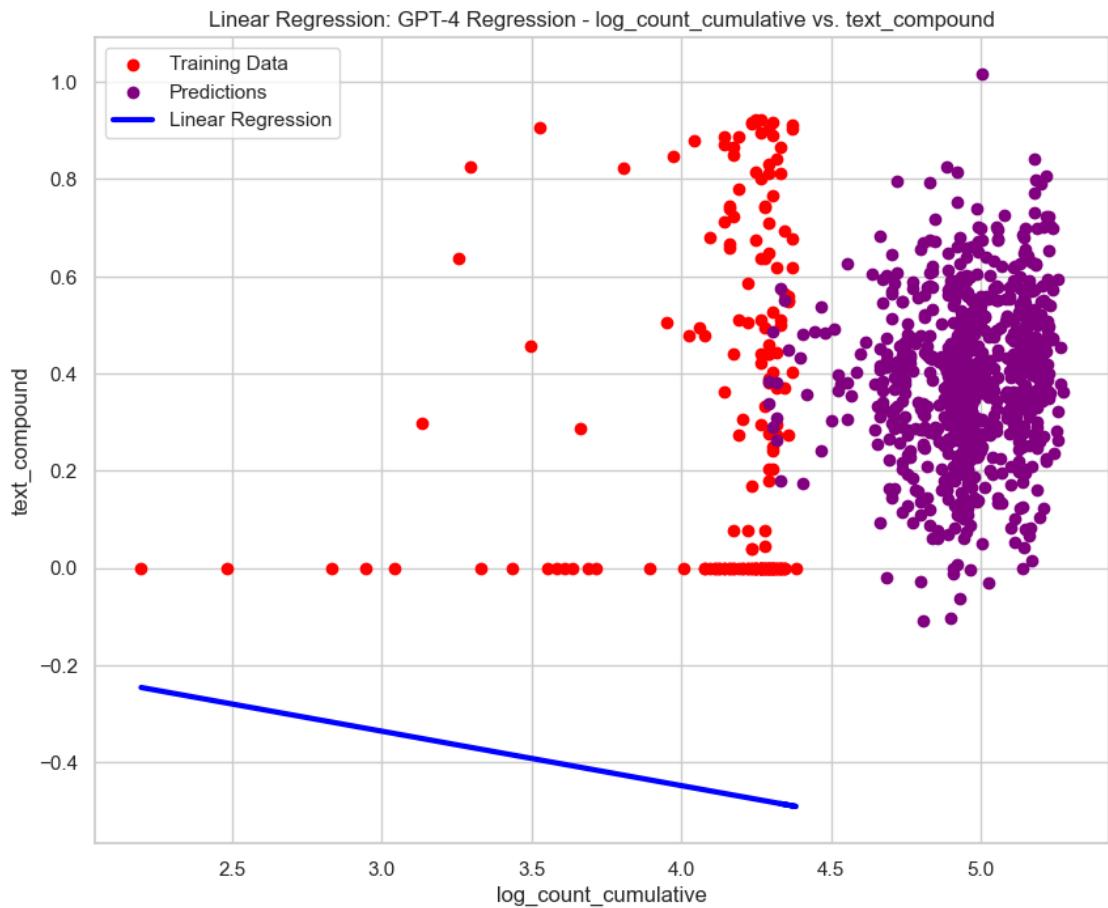
log_comment_score_weighted=-0.7747332174037064



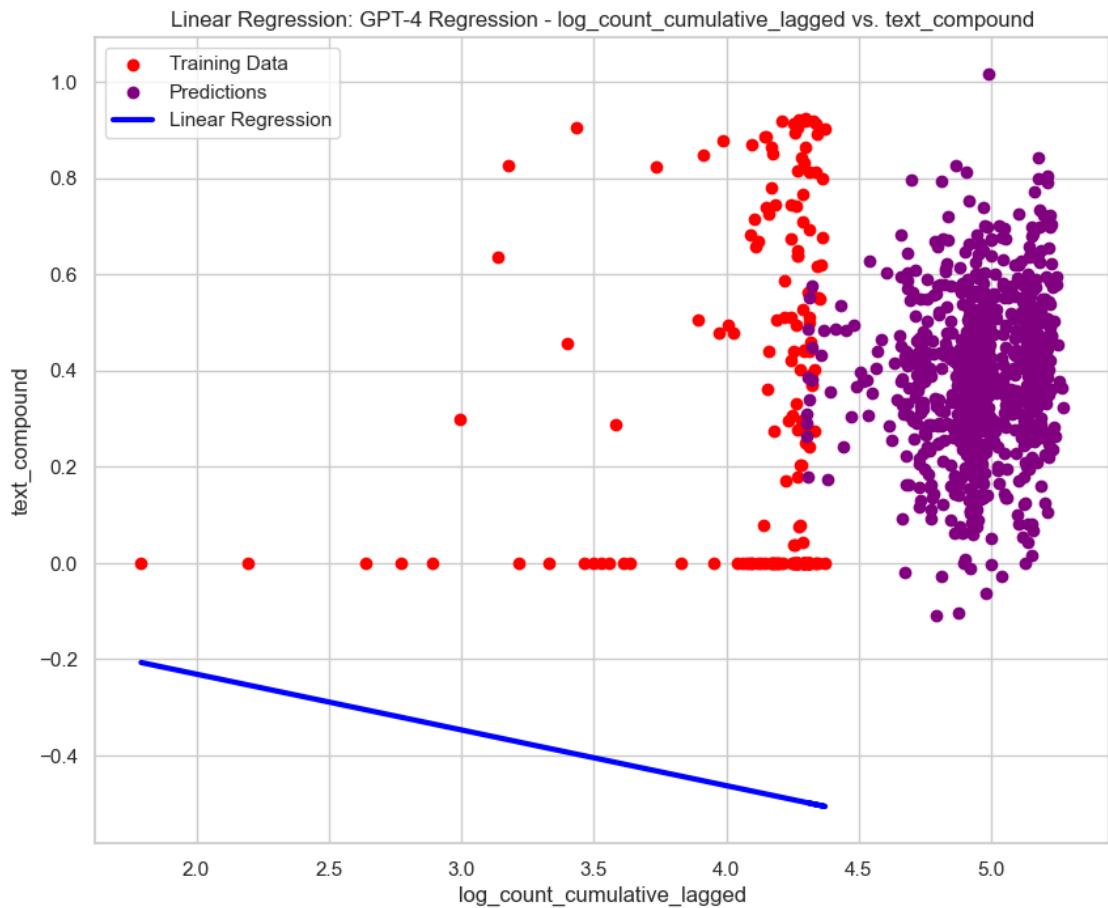
log_comment_launch_distance_f=0.019637780551439304



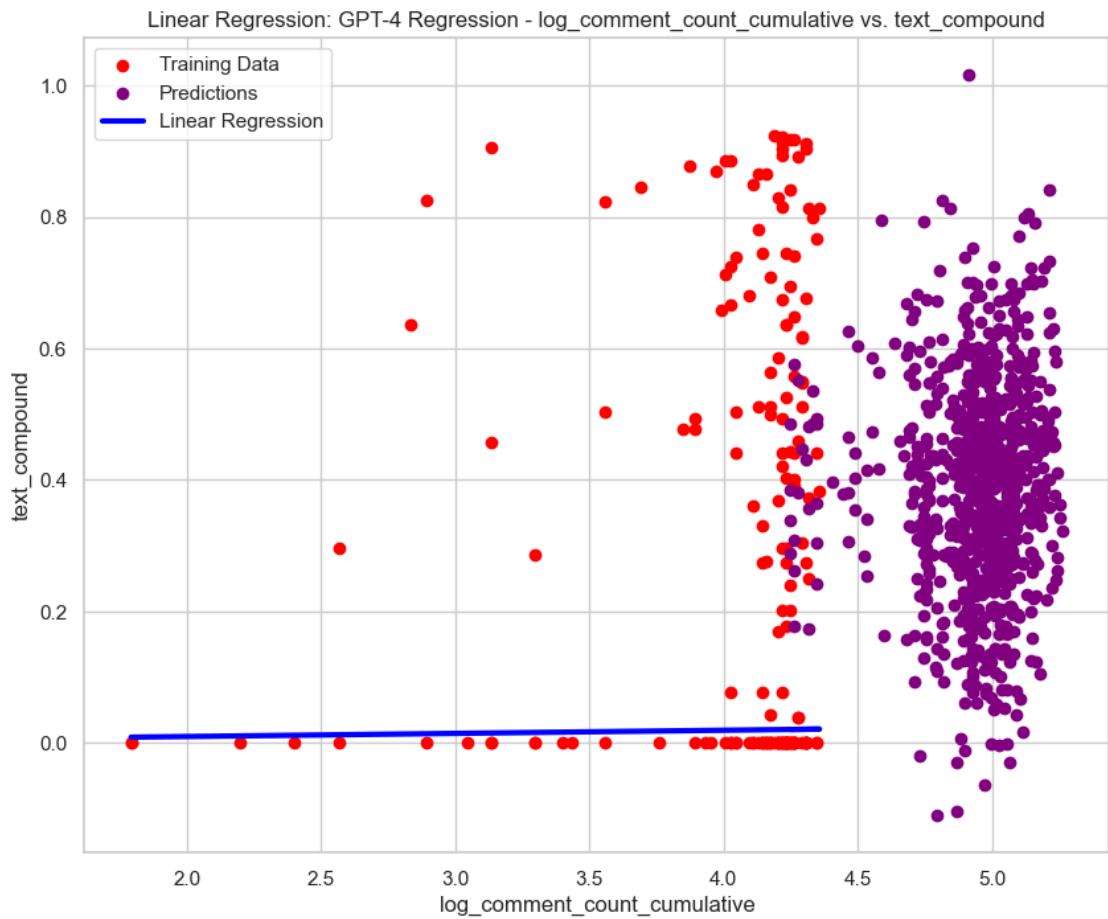
log_comment_num_replies_weighted=0.07728878833598815



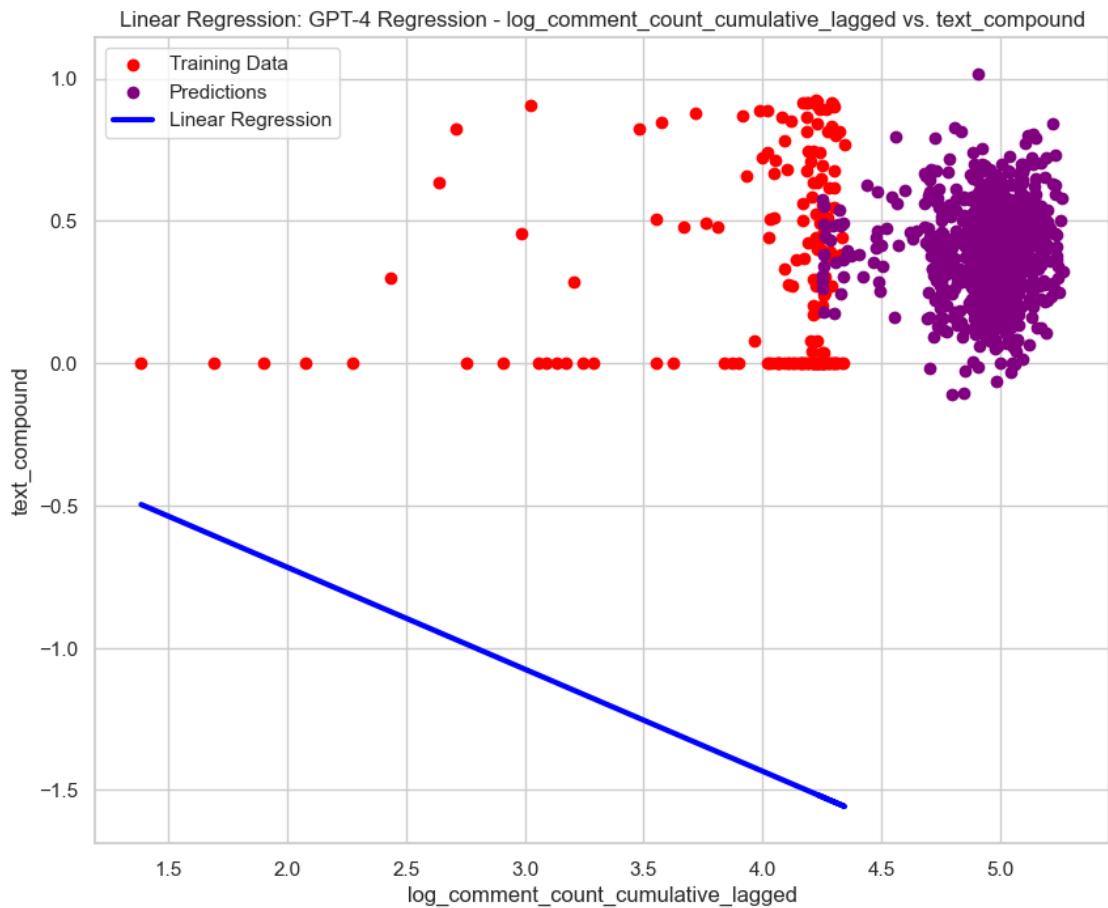
`log_count_cumulative=-0.11206030406979832`



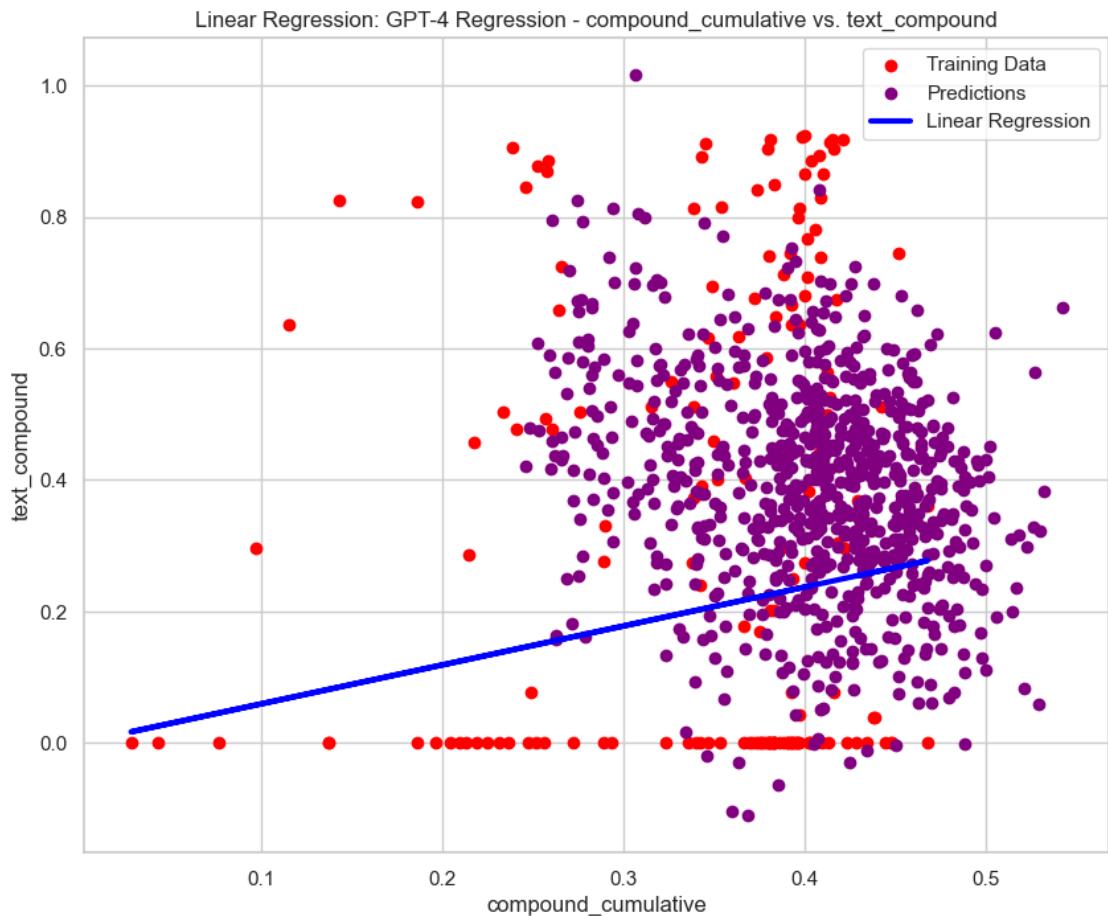
`log_count_cumulative_lagged=-0.11596675711971235`



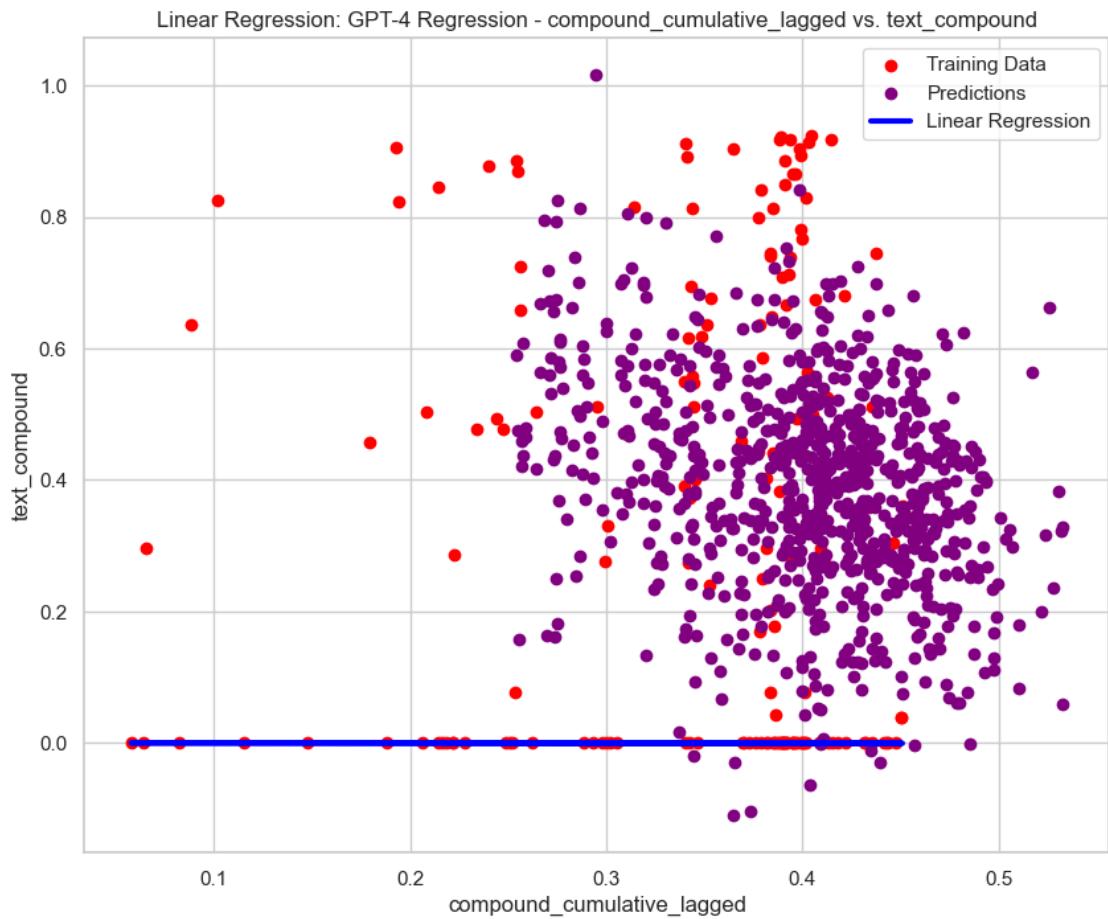
`log_comment_count_cumulative=0.004781733239506858`



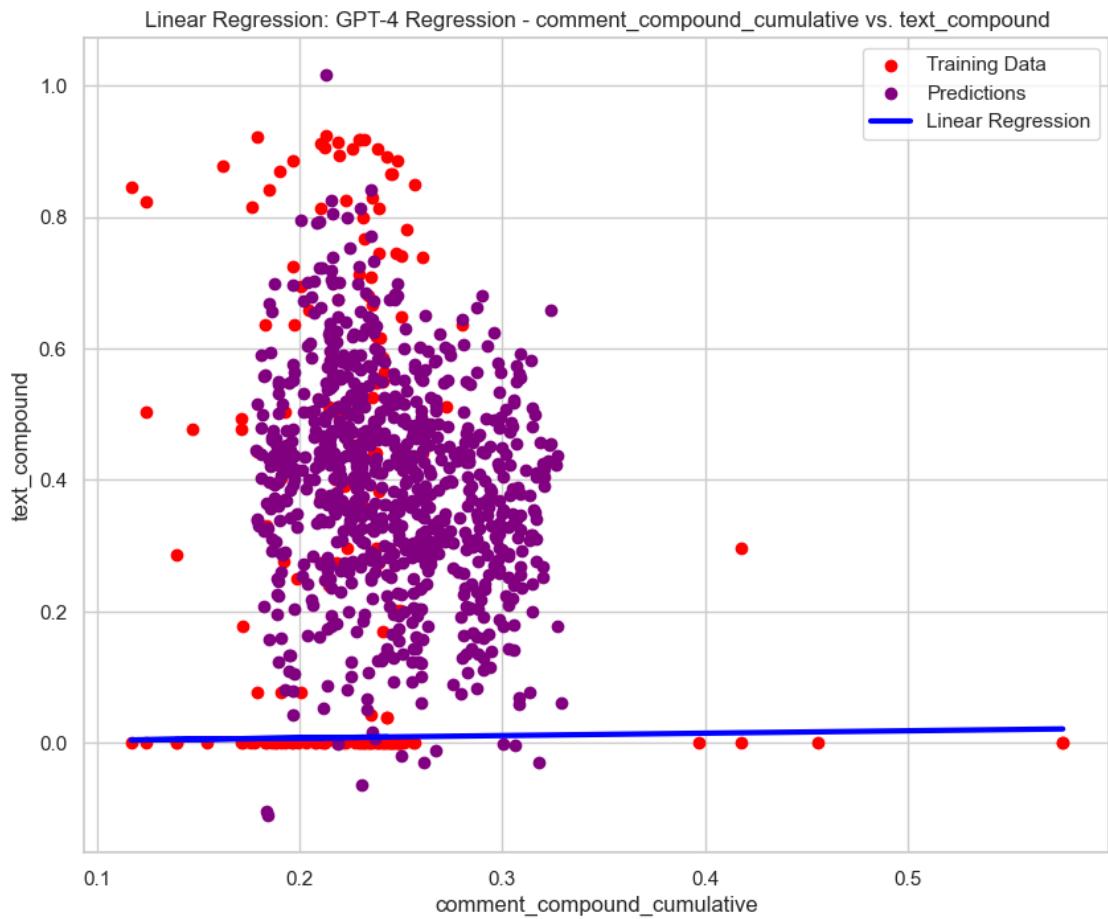
log_comment_count_cumulative_lagged=-0.358546780224581



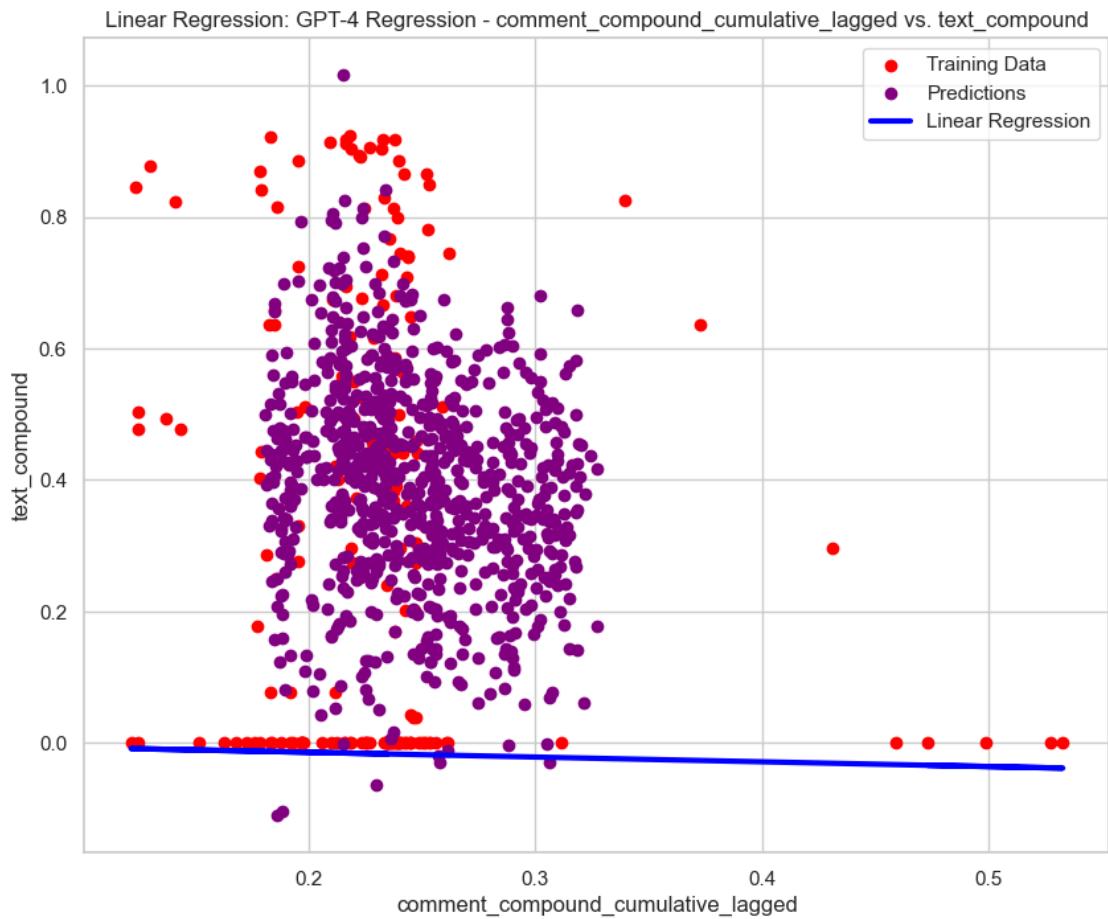
compound_cumulative=0.5925942798502565



`compound_cumulative_lagged=-0.001499556436146212`

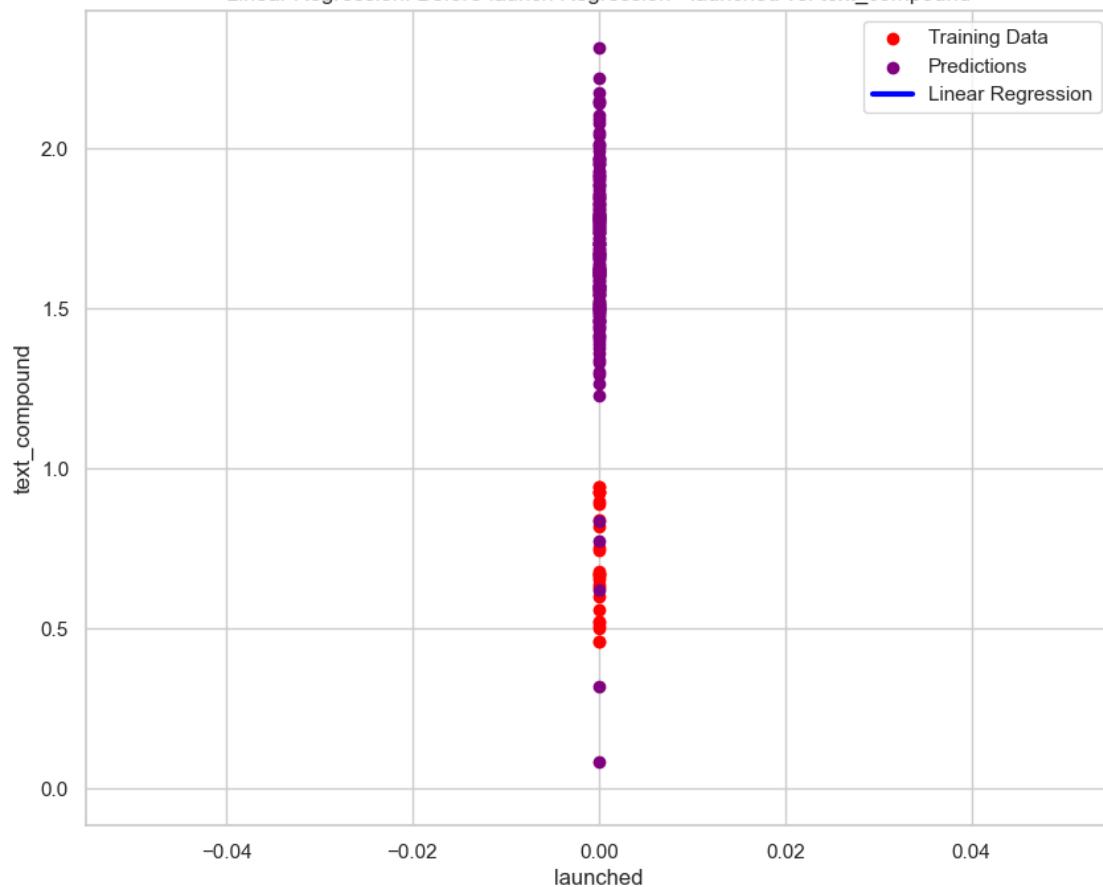


comment_compound_cumulative=0.036628461438446905

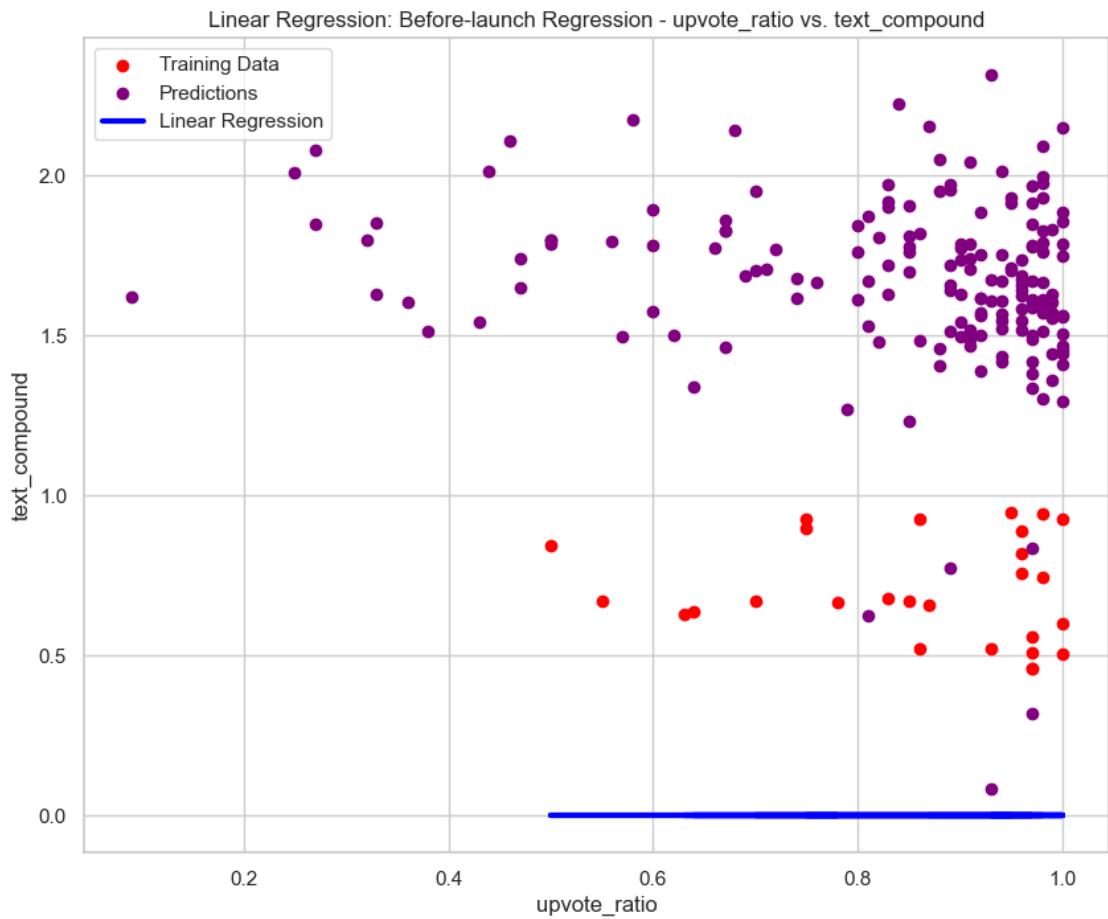


comment_compound_cumulative_lagged=-0.07214522079265877

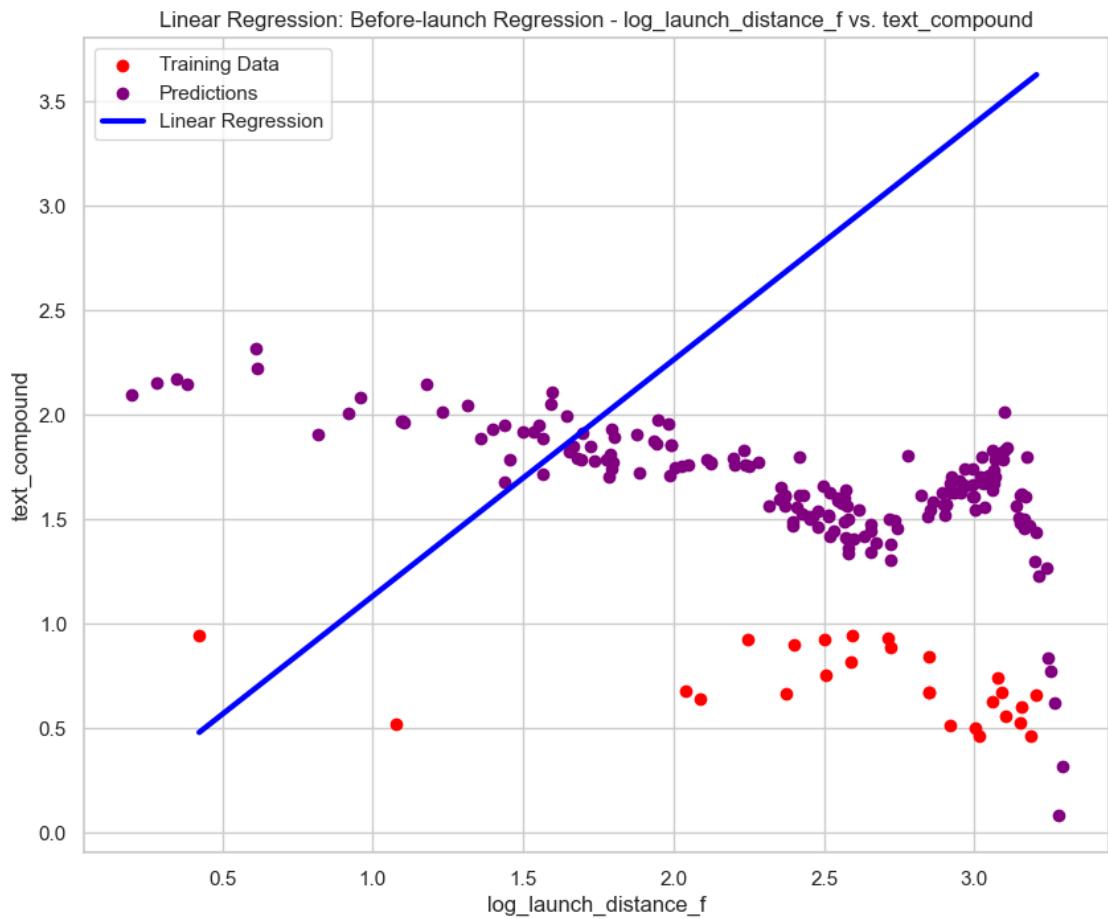
Linear Regression: Before-launch Regression - launched vs. text_compound



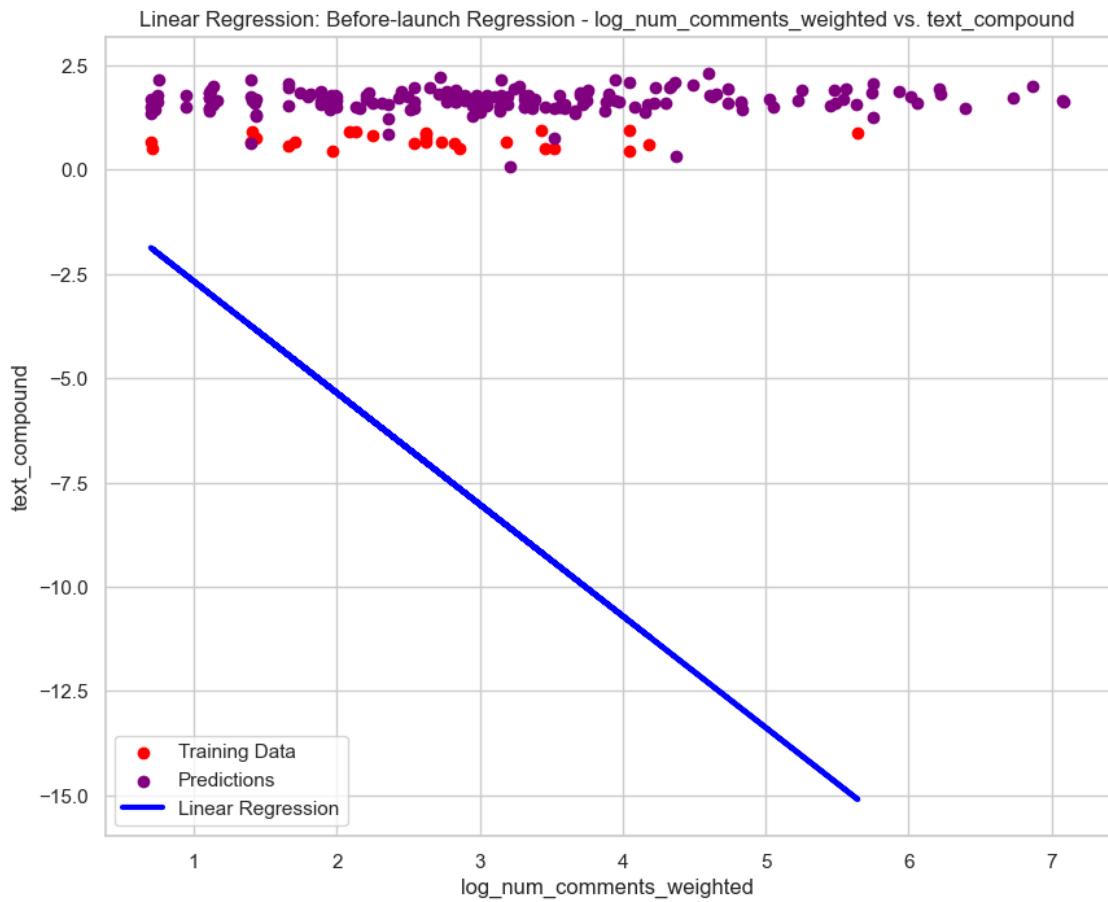
launched=0.03858679253065731



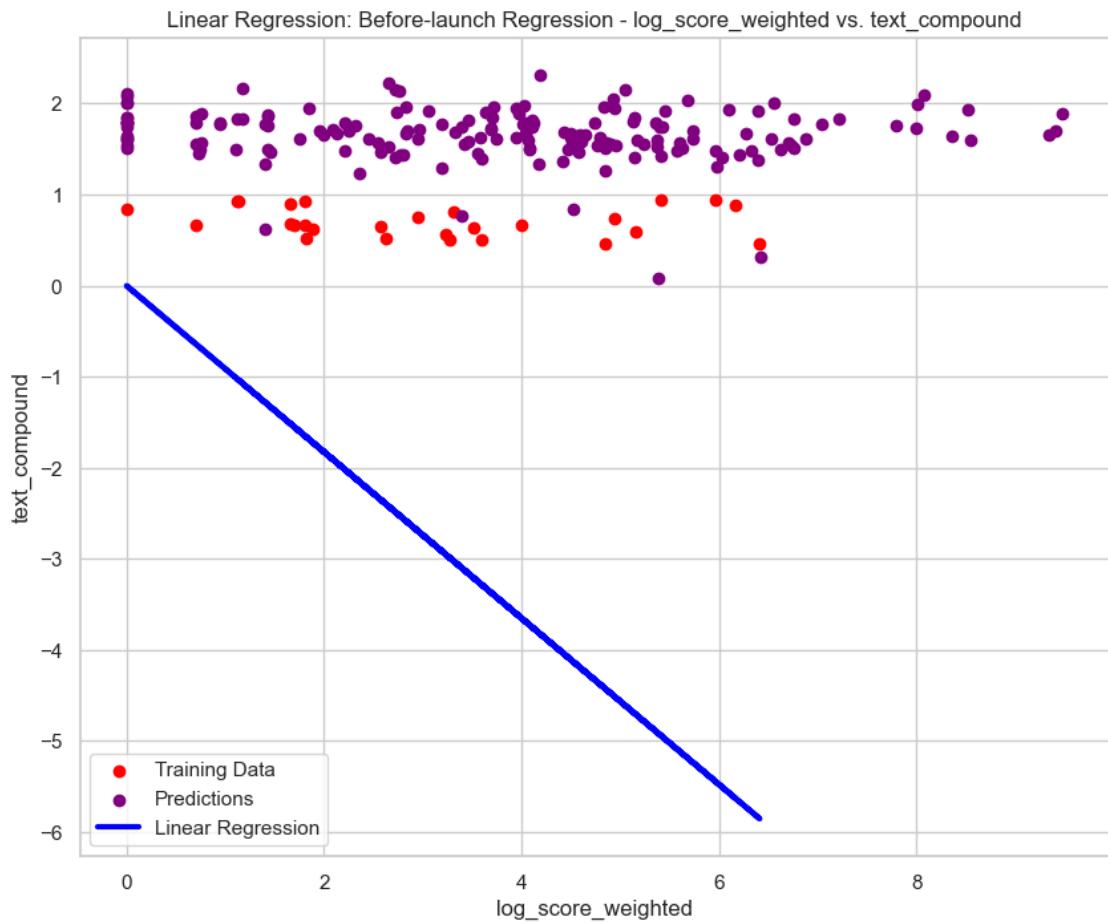
upvote_ratio=7.216449660063518e-16



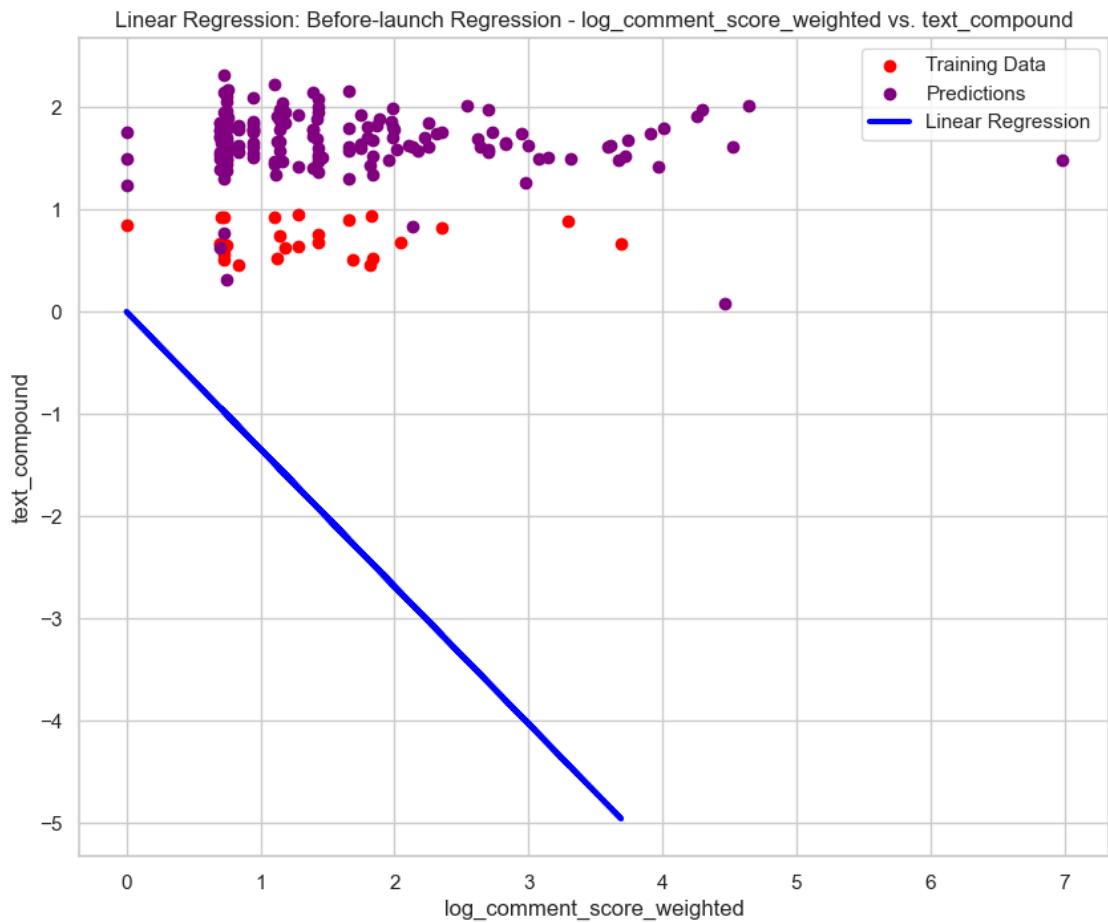
`log_launch_distance_f=1.1307324243024075`



log_num_comments_weighted=-2.675335735982713

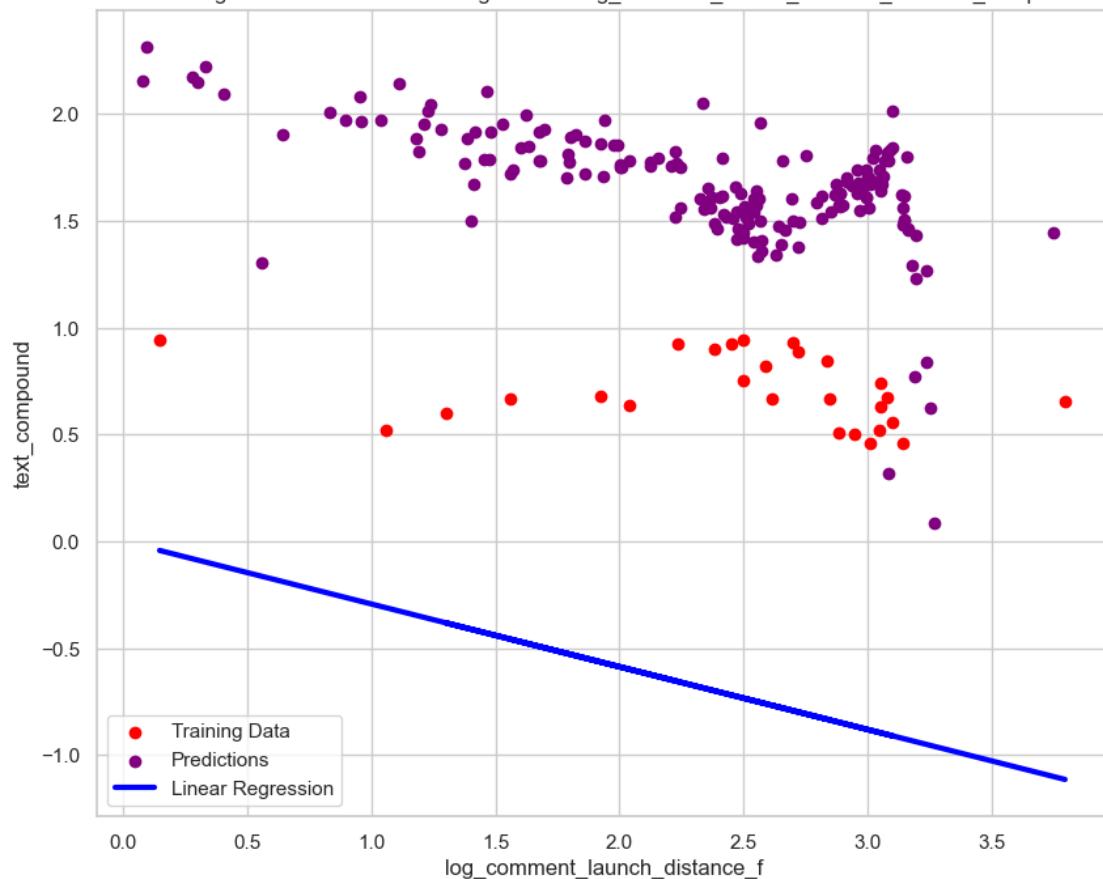


log_score_weighted=-0.9129480729134403



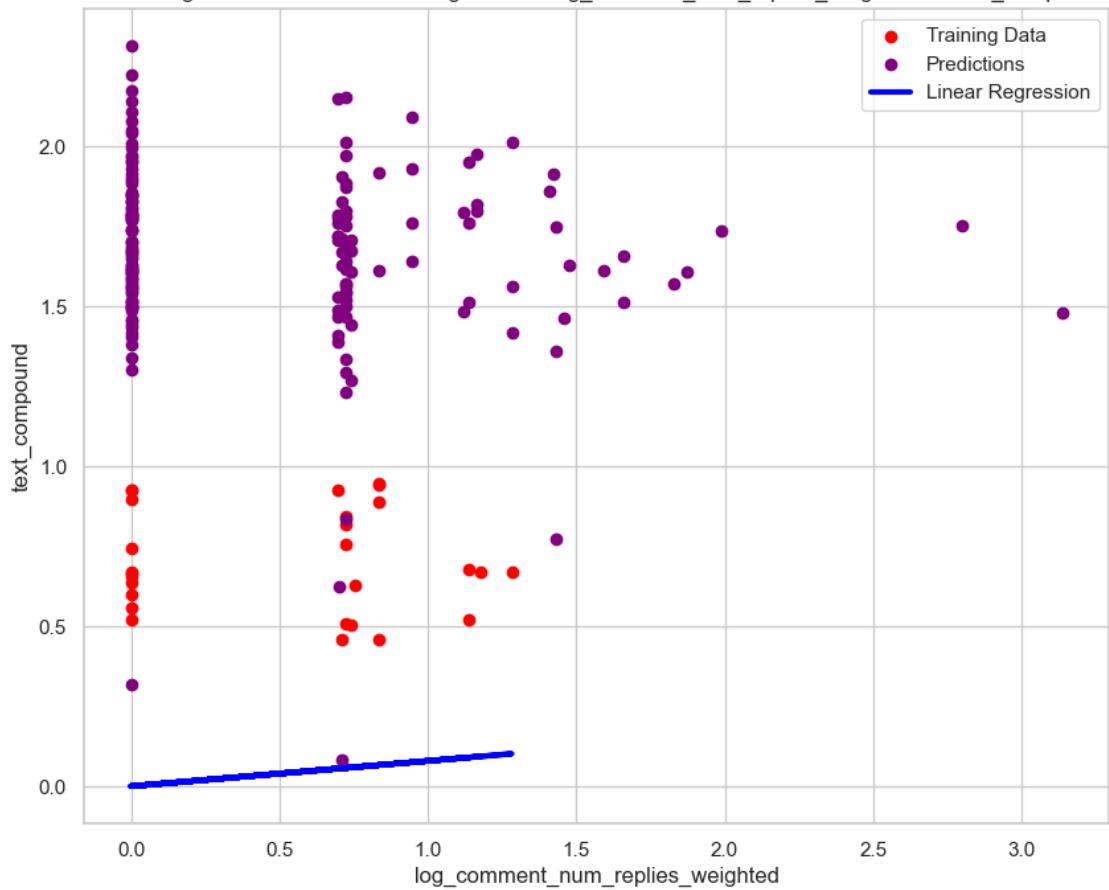
log_comment_score_weighted=-1.3436140735918558

Linear Regression: Before-launch Regression - log_comment_launch_distance_f vs. text_compound

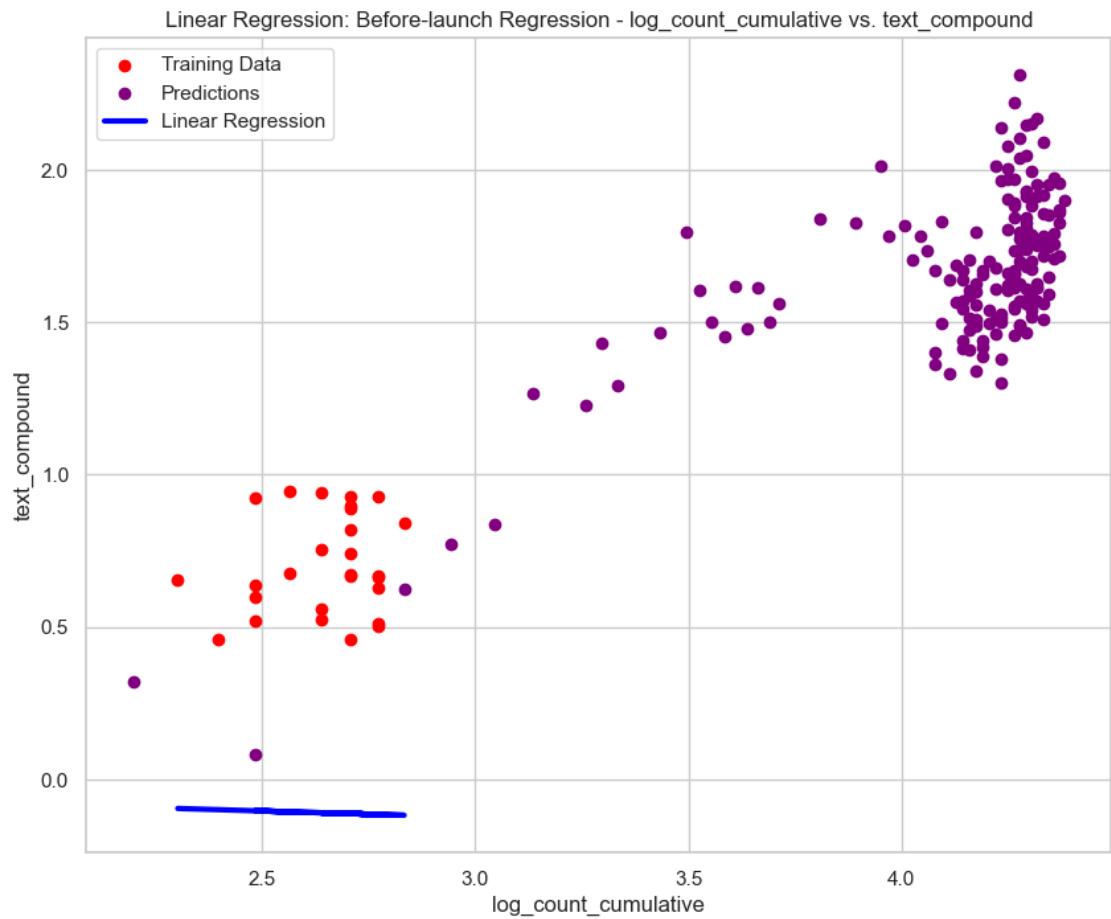


$\log_{\text{comment_launch_distance_f}} = -0.29368908950744244$

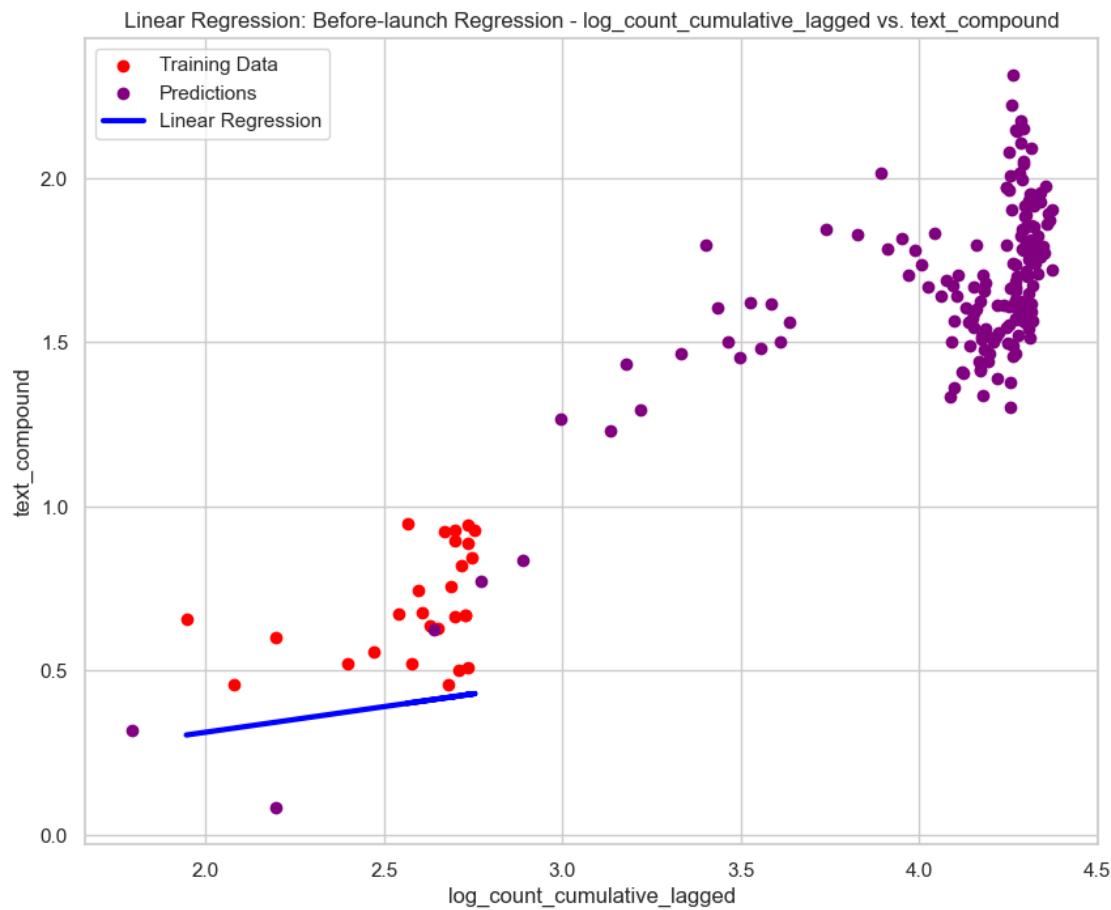
Linear Regression: Before-launch Regression - log_comment_num_replies_weighted vs. text_compound



log_comment_num_replies_weighted=0.07913892483406158

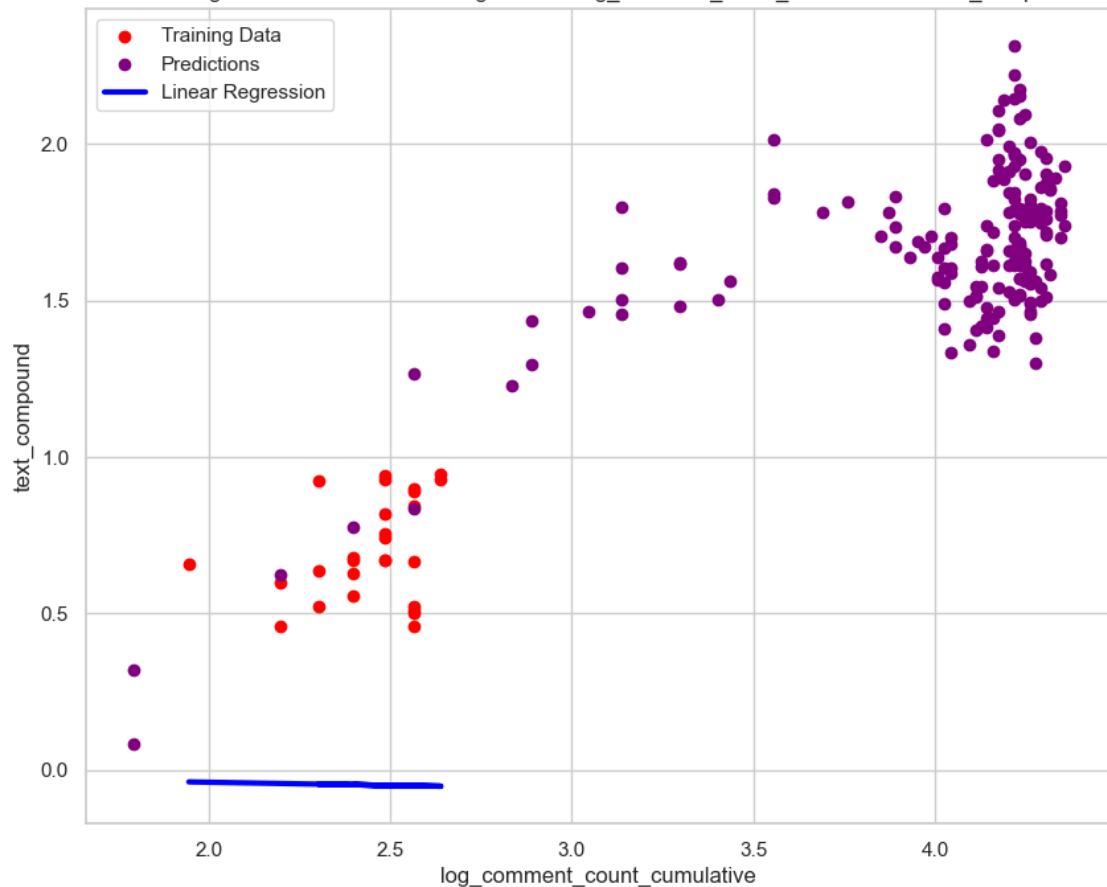


log_count_cumulative=-0.04186038318727184



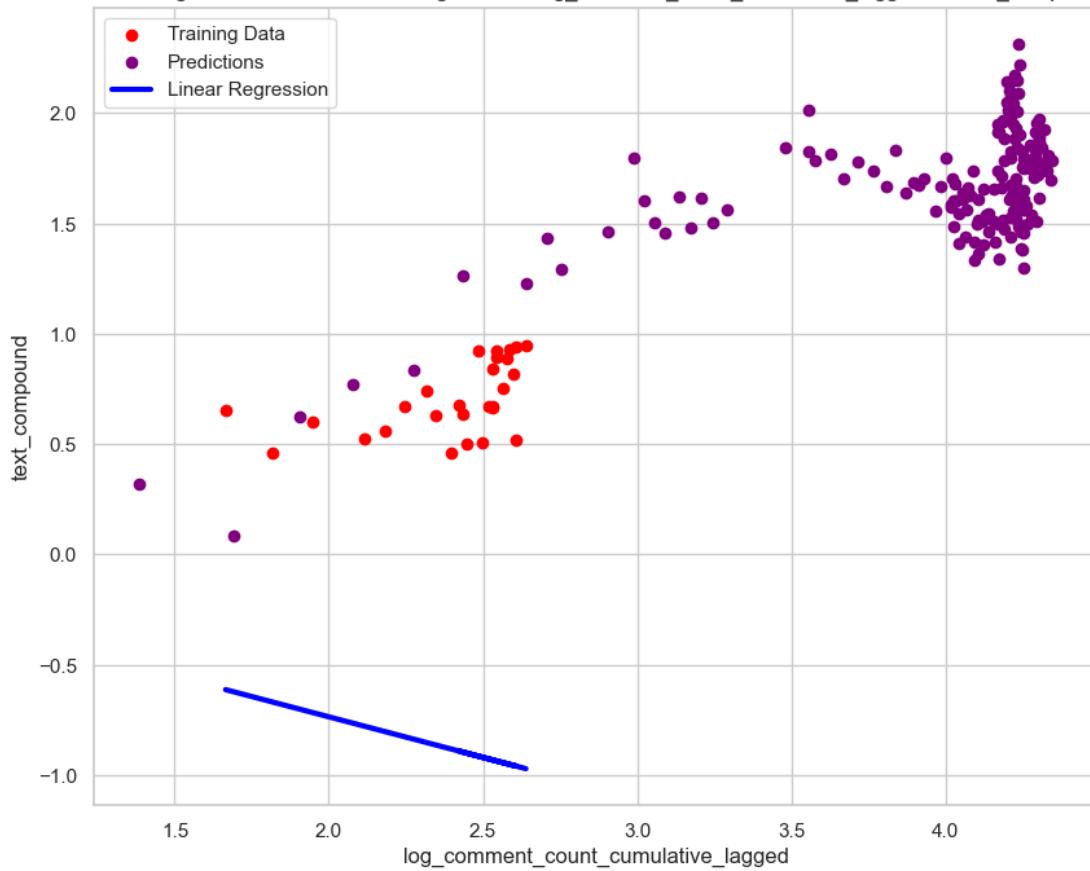
log_count_cumulative_lagged=0.15579896373676838

Linear Regression: Before-launch Regression - log_comment_count_cumulative vs. text_compound



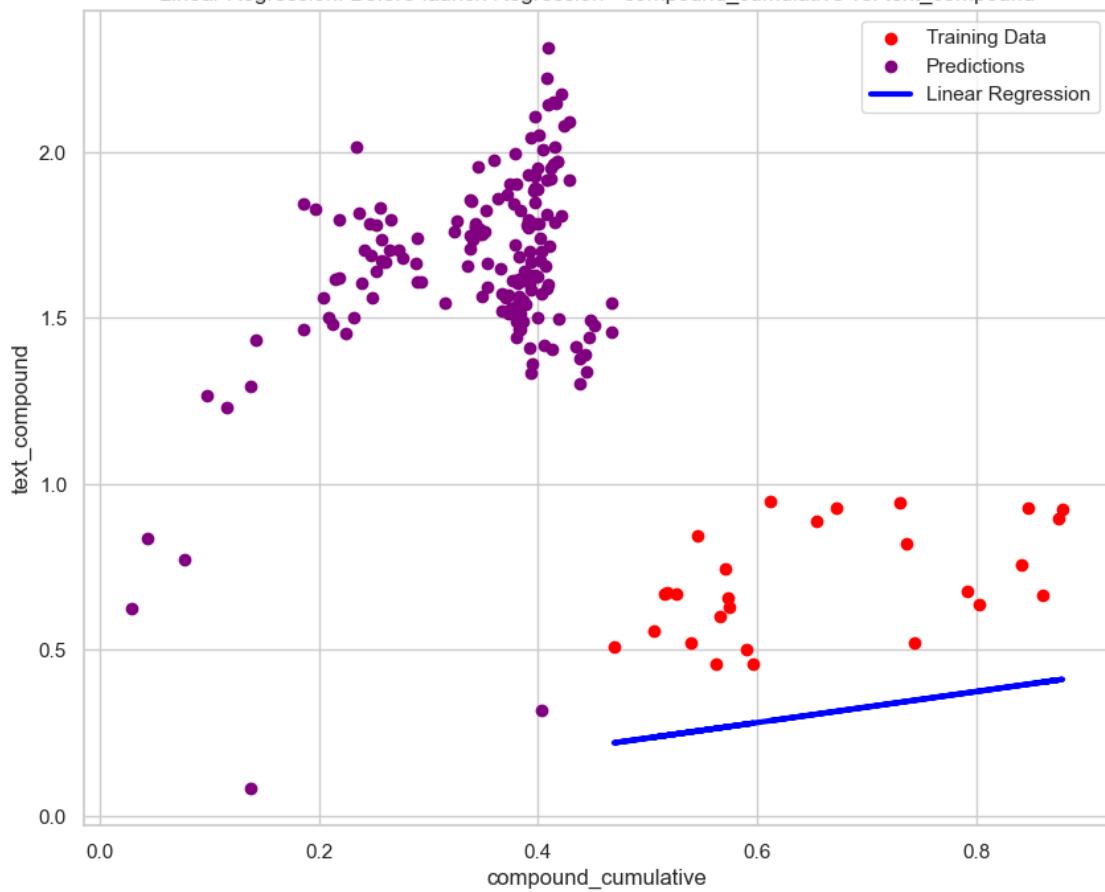
log_comment_count_cumulative=-0.019941574615981932

Linear Regression: Before-launch Regression - log_comment_count_cumulative_lagged vs. text_compound



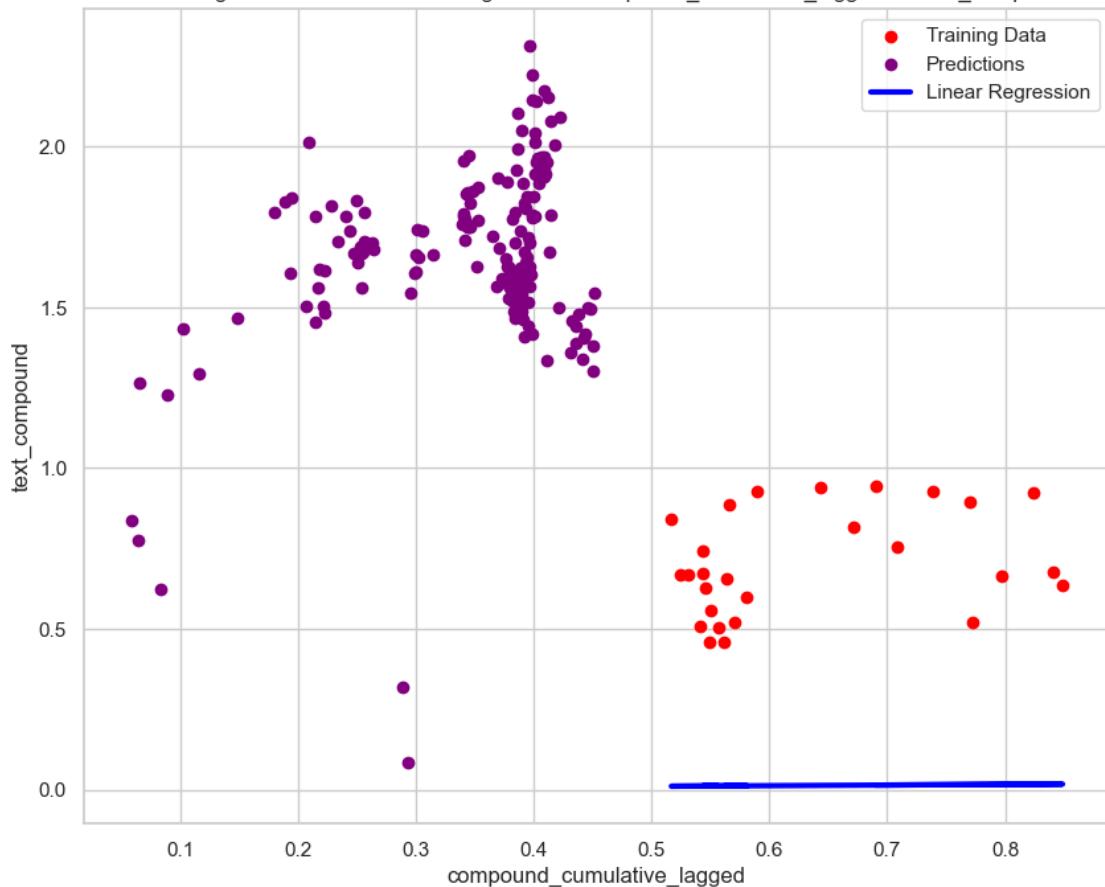
log_comment_count_cumulative_lagged=-0.3677388623725115

Linear Regression: Before-launch Regression - compound_cumulative vs. text_compound



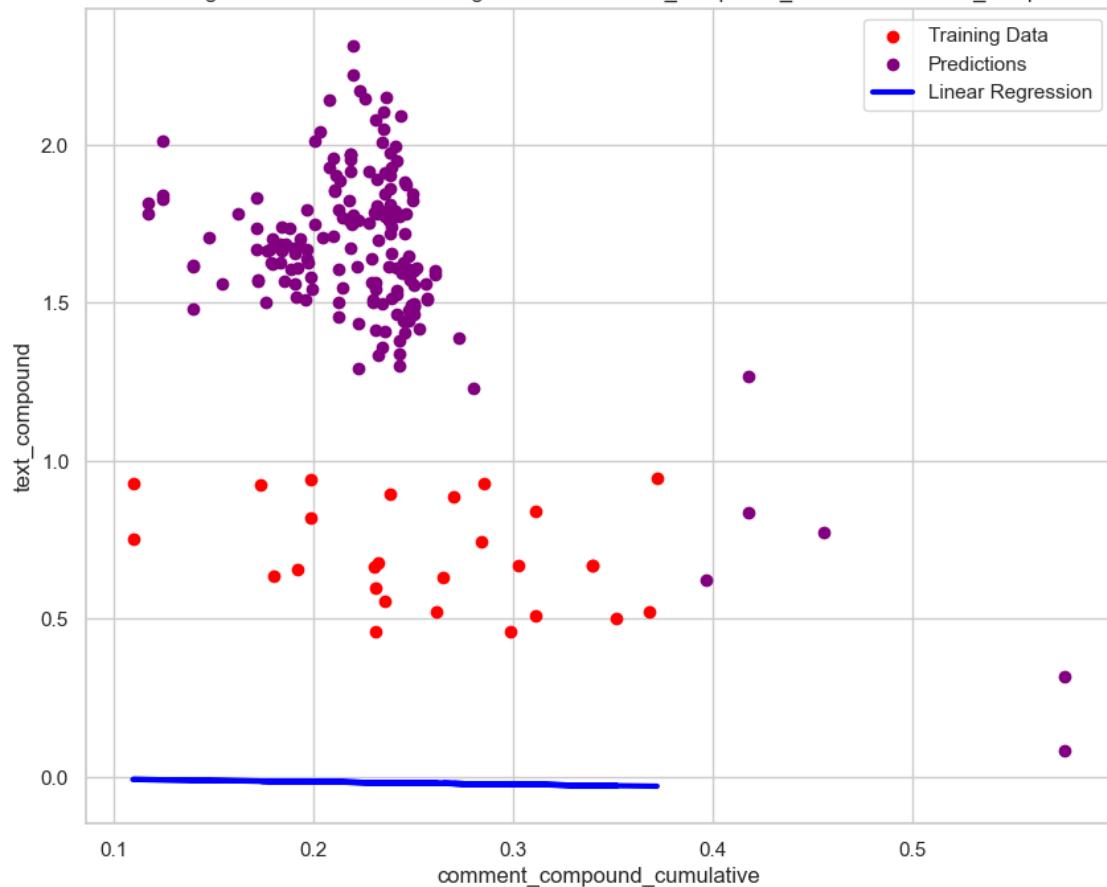
compound_cumulative=0.46712315647027425

Linear Regression: Before-launch Regression - compound_cumulative_lagged vs. text_compound



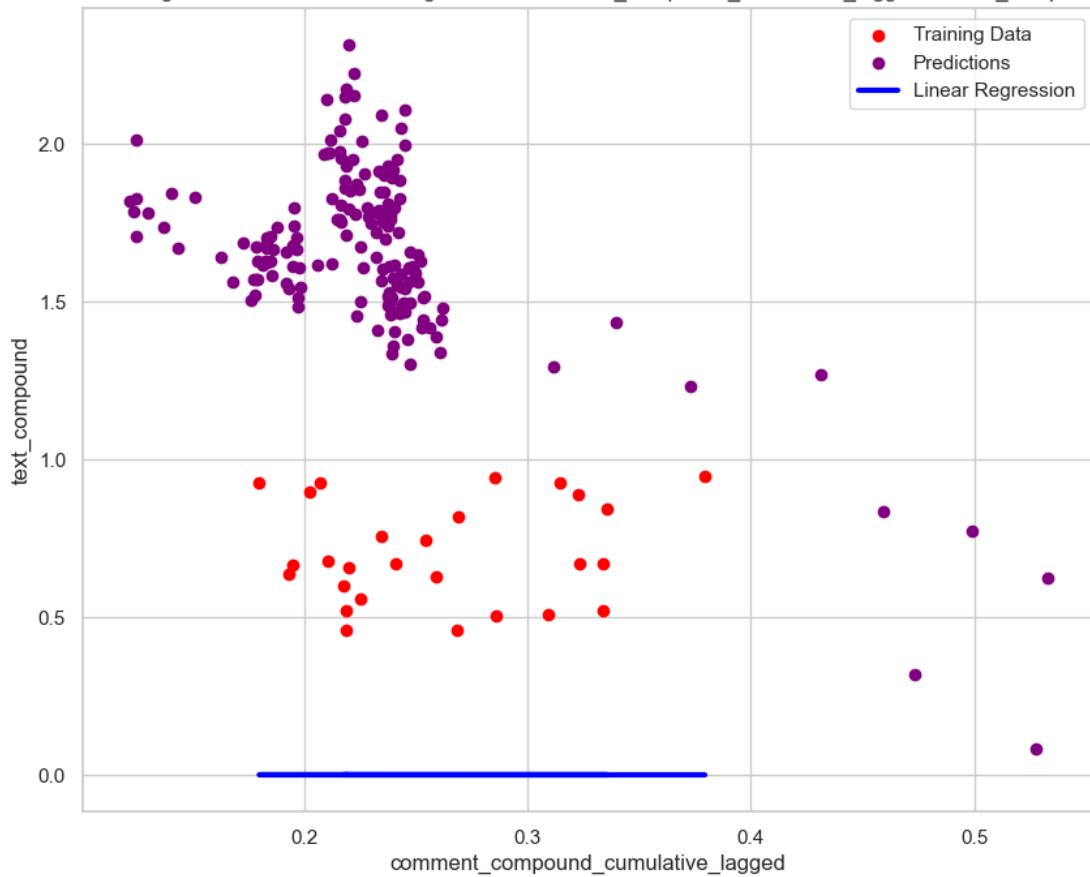
compound_cumulative_lagged=0.020234890221955737

Linear Regression: Before-launch Regression - comment_compound_cumulative vs. text_compound

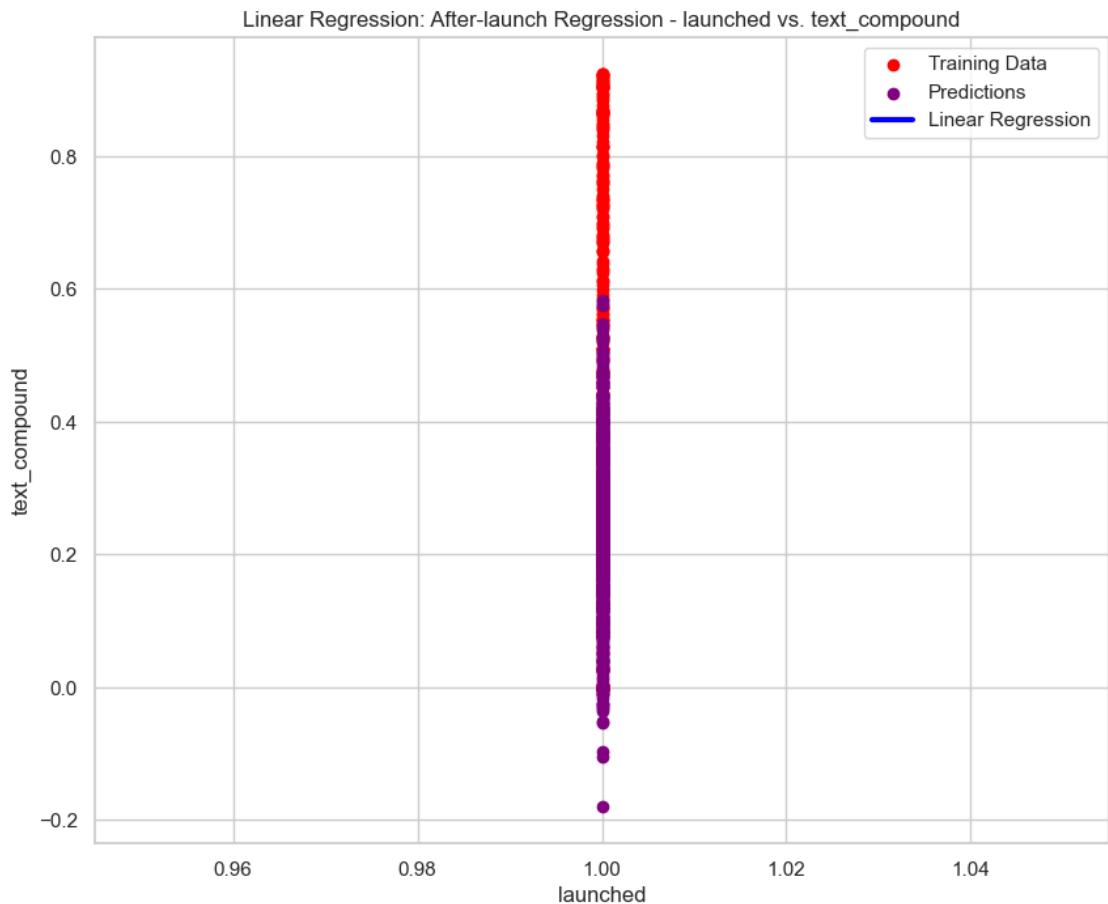


comment_compound_cumulative=-0.08186407607413103

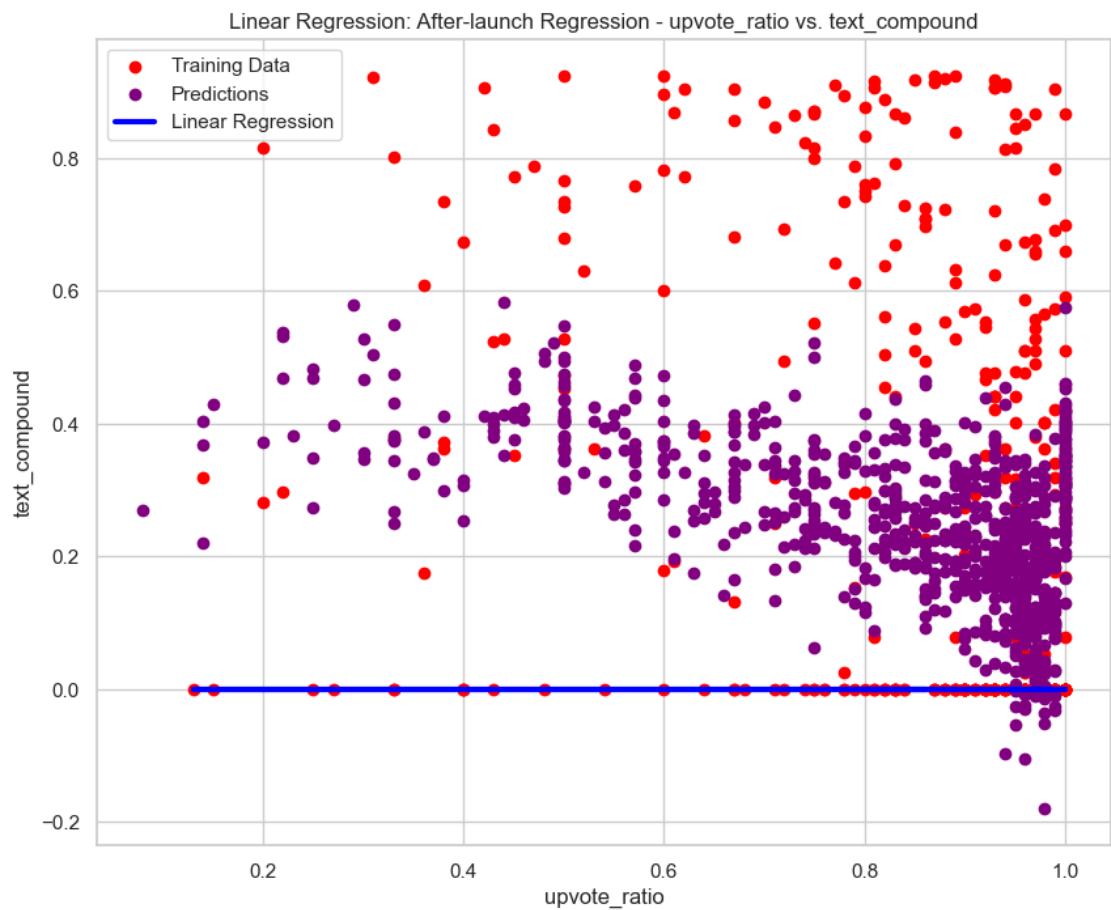
Linear Regression: Before-launch Regression - comment_compound_cumulative_lagged vs. text_compound



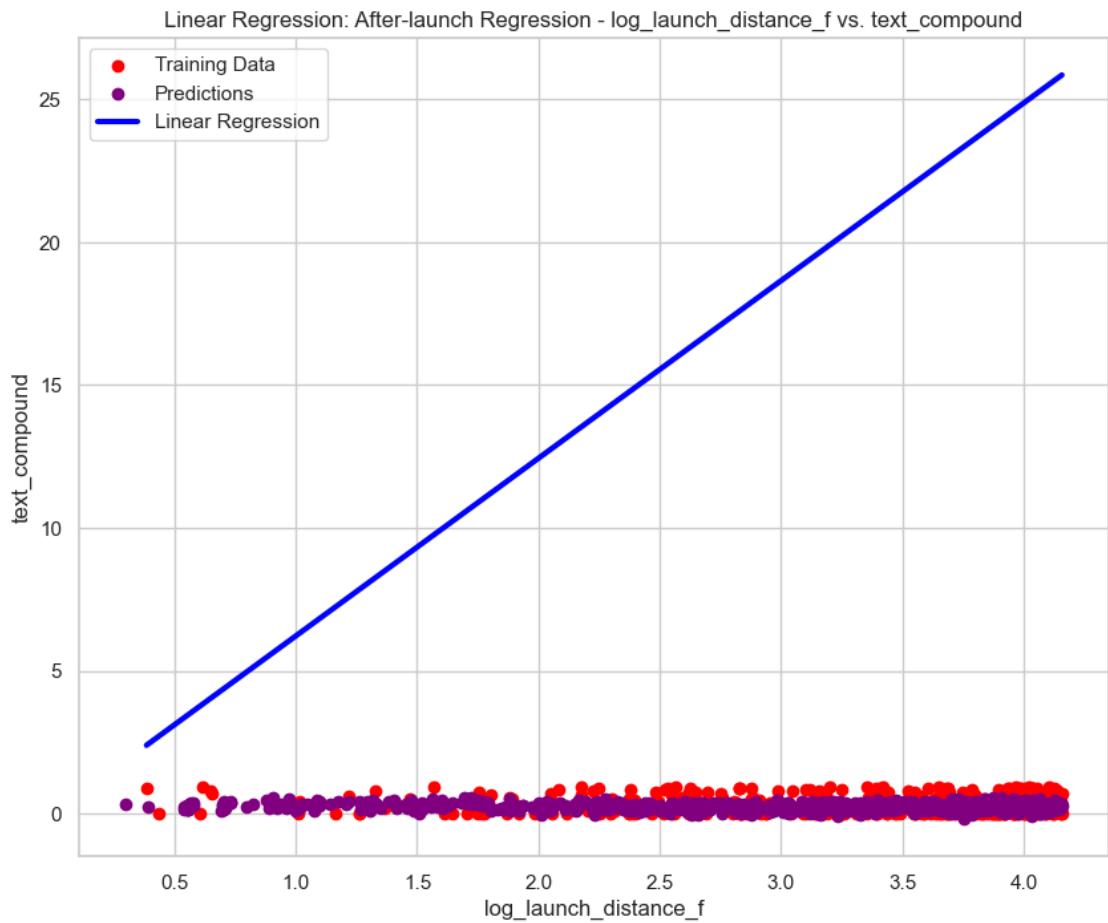
comment_compound_cumulative_lagged=-0.00021174886748573038



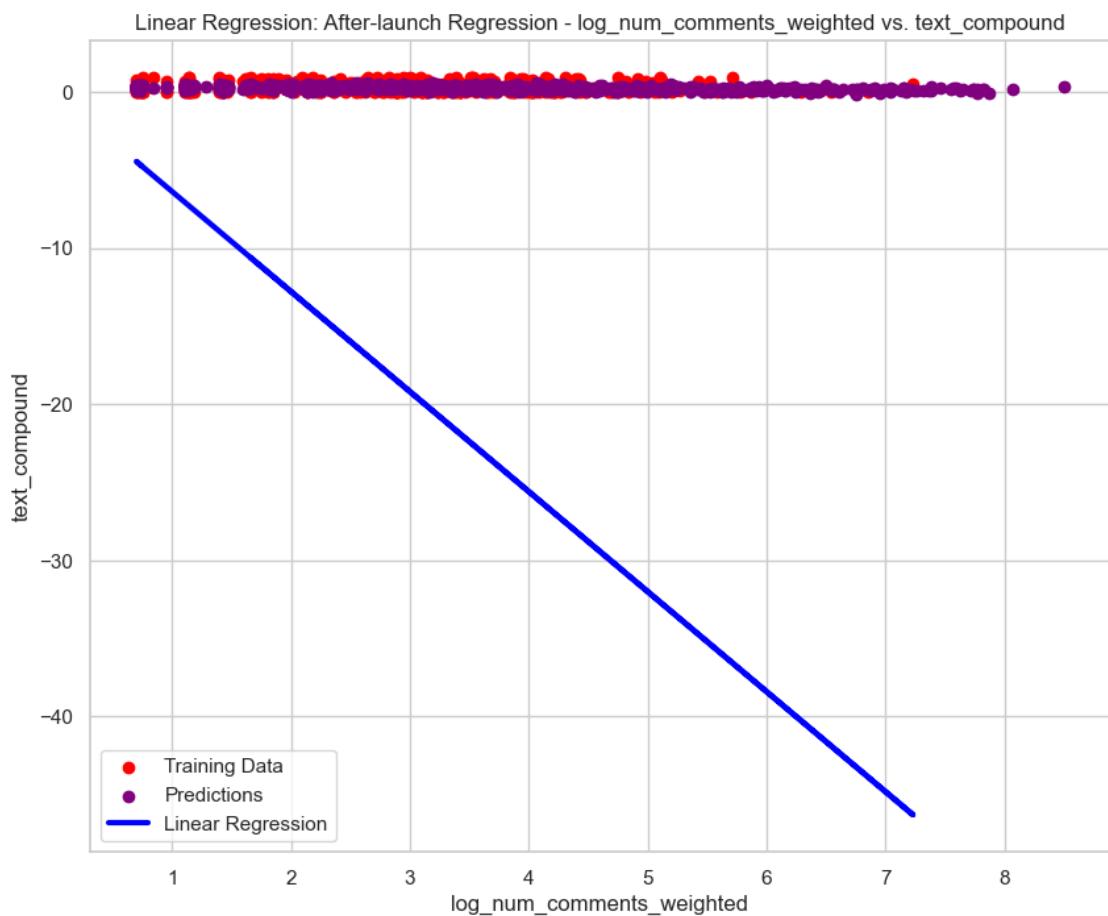
launched=0.04598565512068587



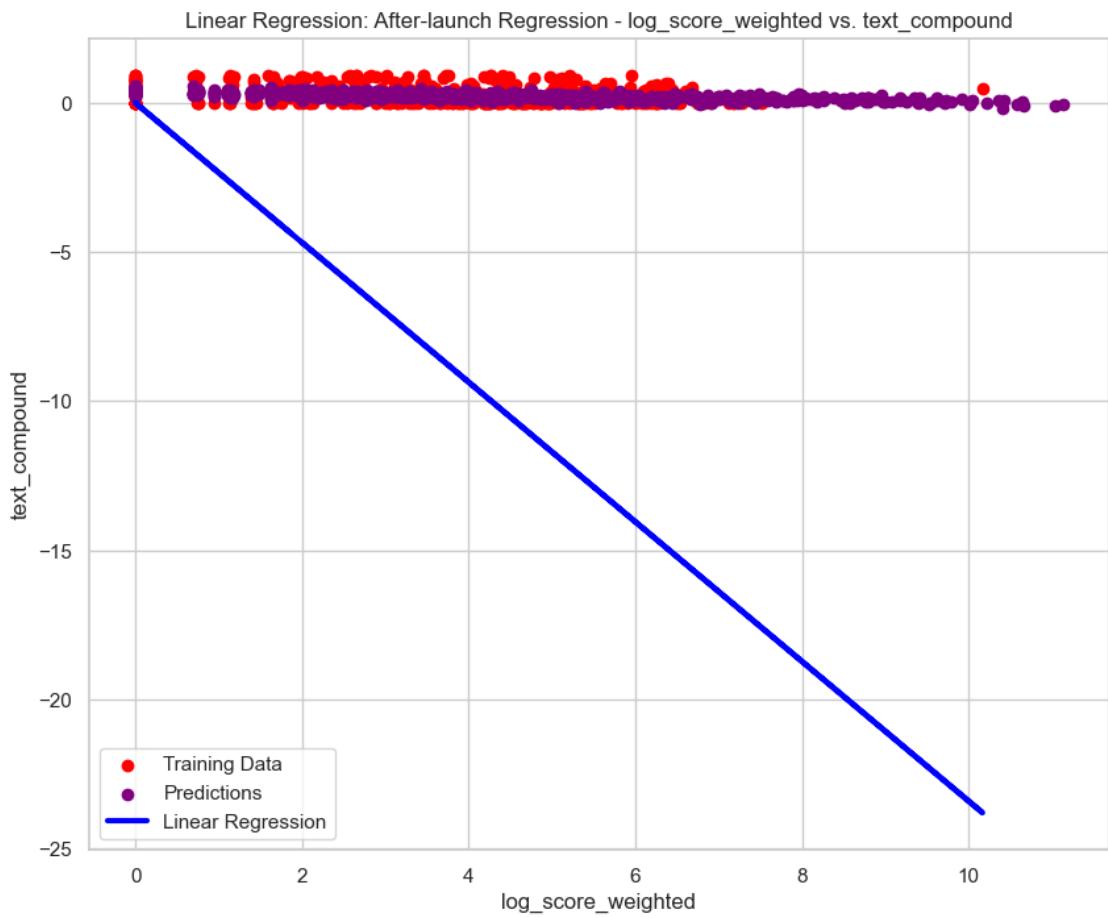
upvote_ratio=9.43689570931383e-15



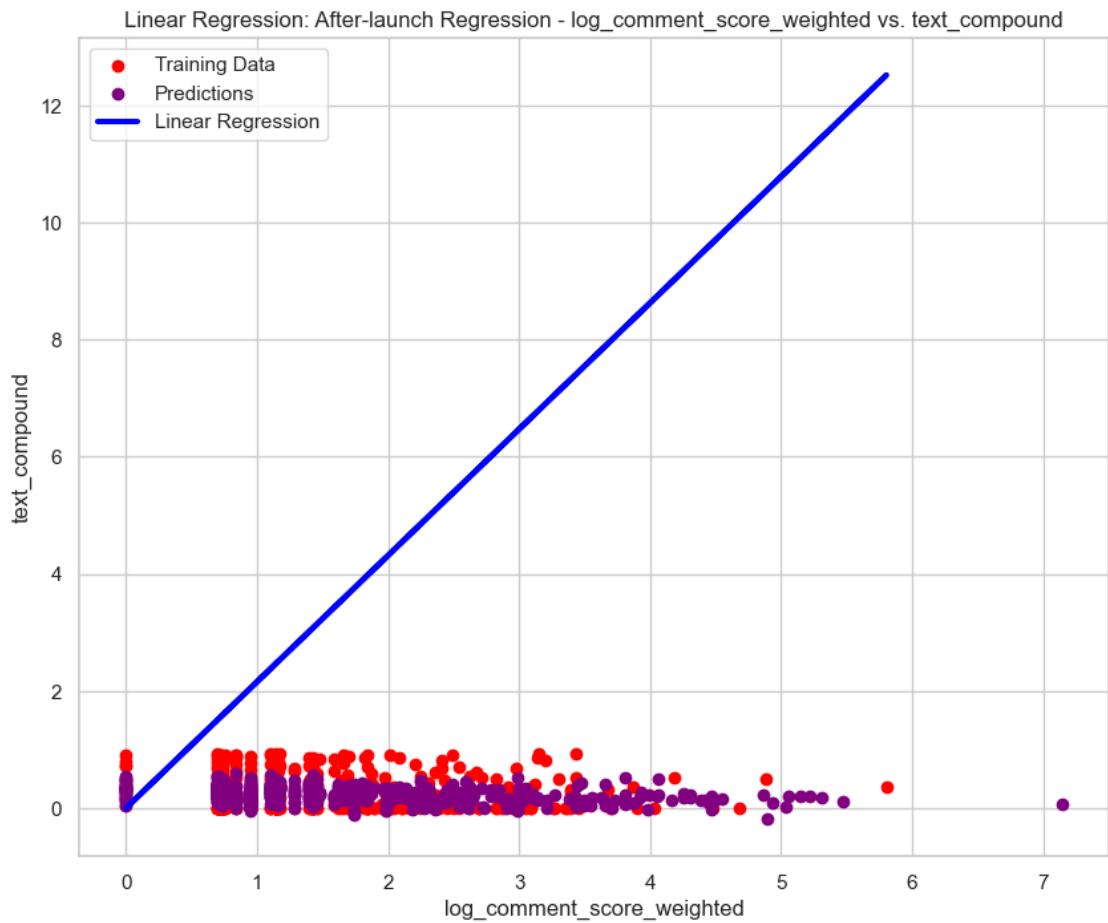
log_launch_distance_f=6.216285087278717



log_num_comments_weighted=-6.402133150871581

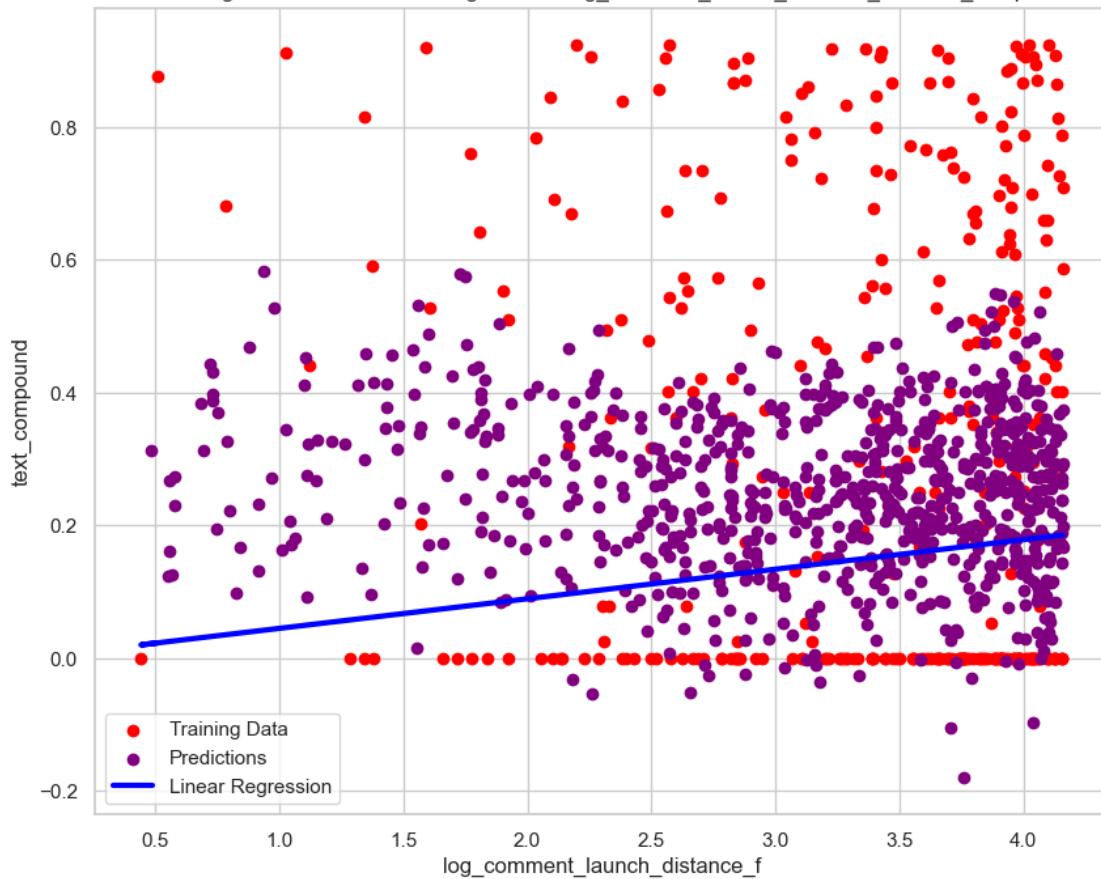


log_score_weighted=-2.340294073550759



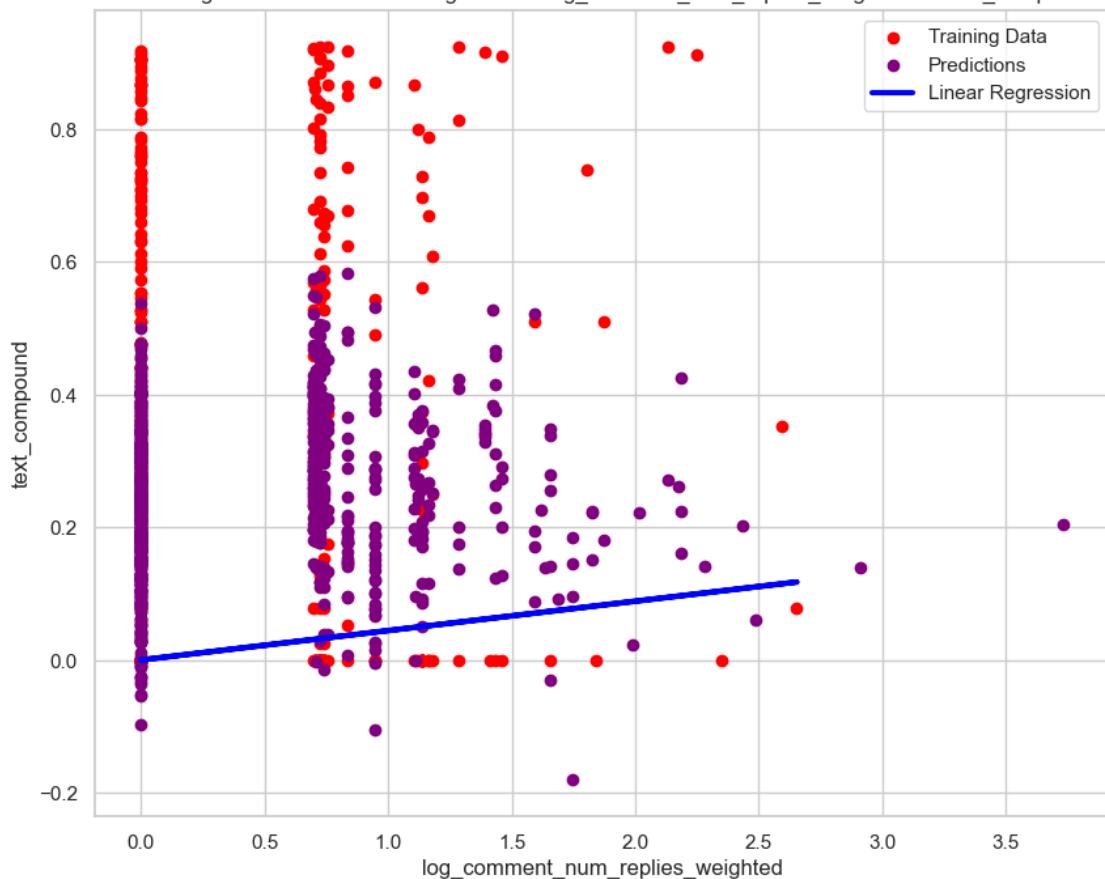
`log_comment_score_weighted=2.155760334485523`

Linear Regression: After-launch Regression - log_comment_launch_distance_f vs. text_compound

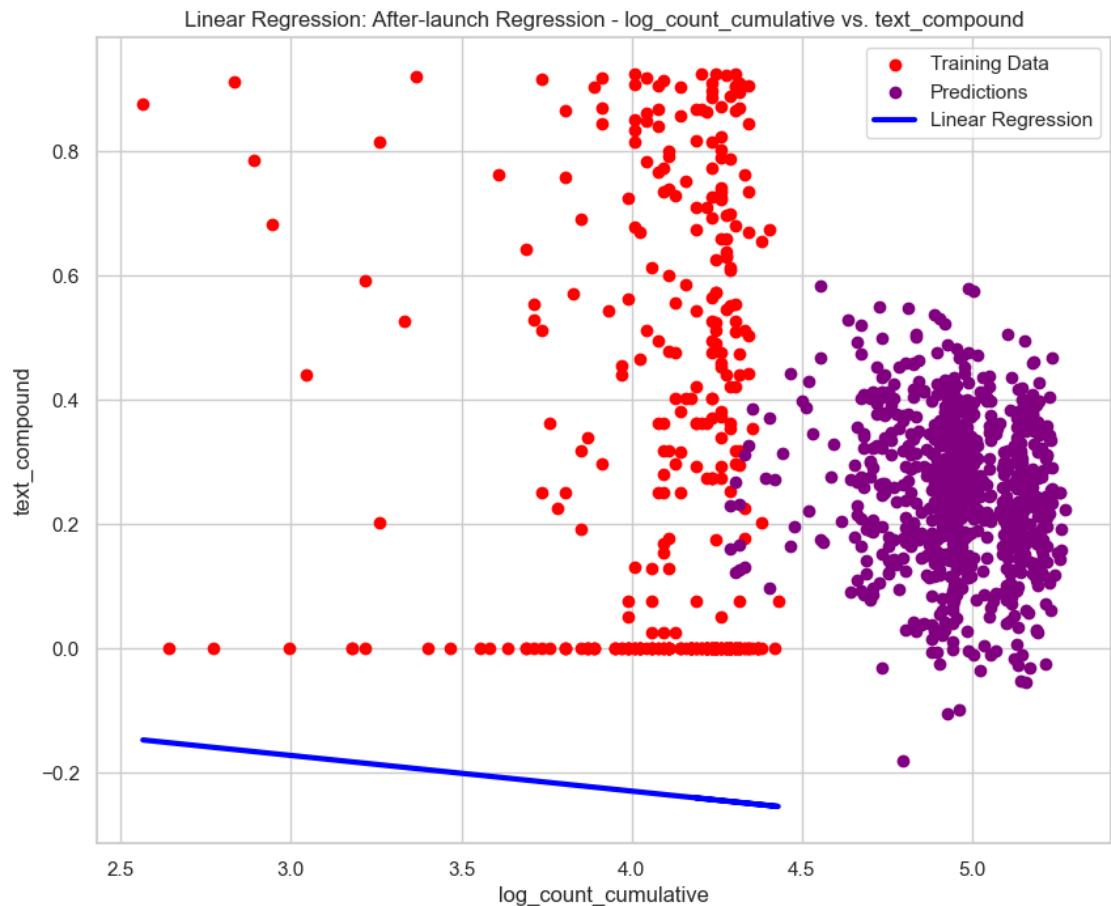


log_comment_launch_distance_f=0.04455557296563453

Linear Regression: After-launch Regression - log_comment_num_replies_weighted vs. text_compound

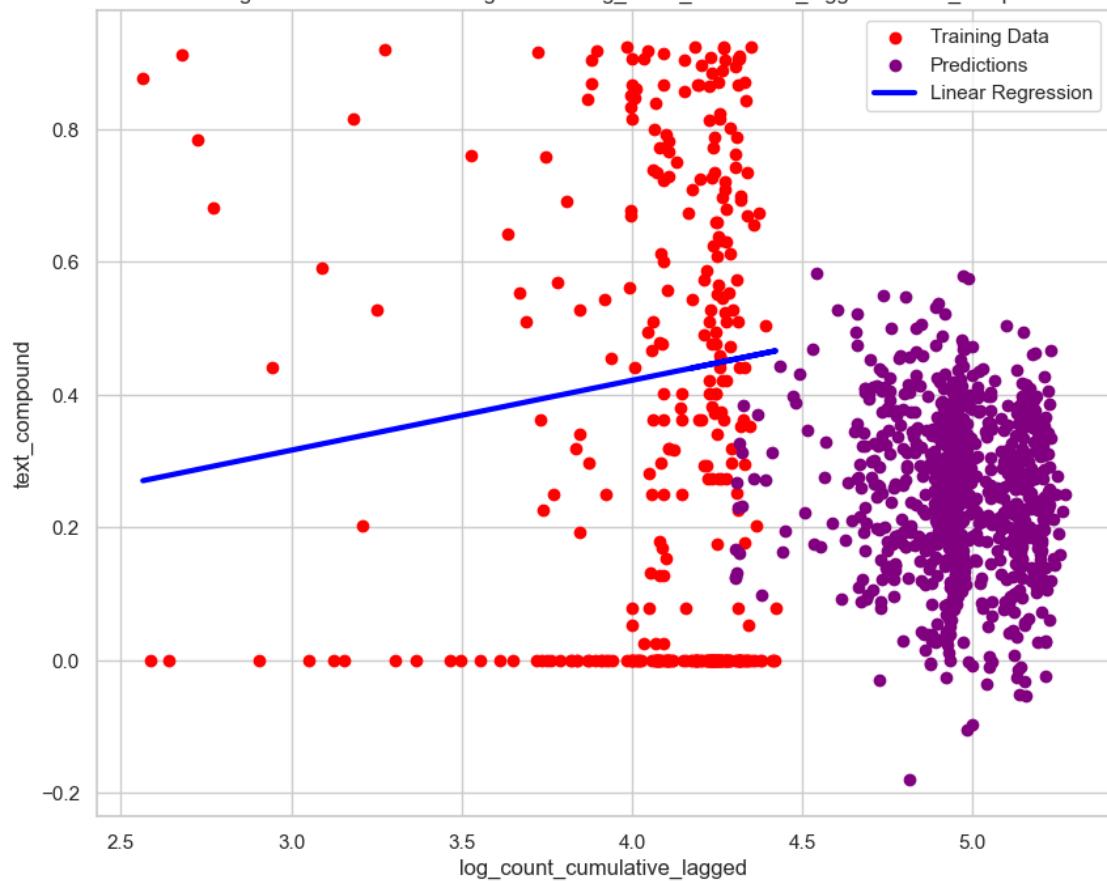


log_comment_num_replies_weighted=0.04424062264431622

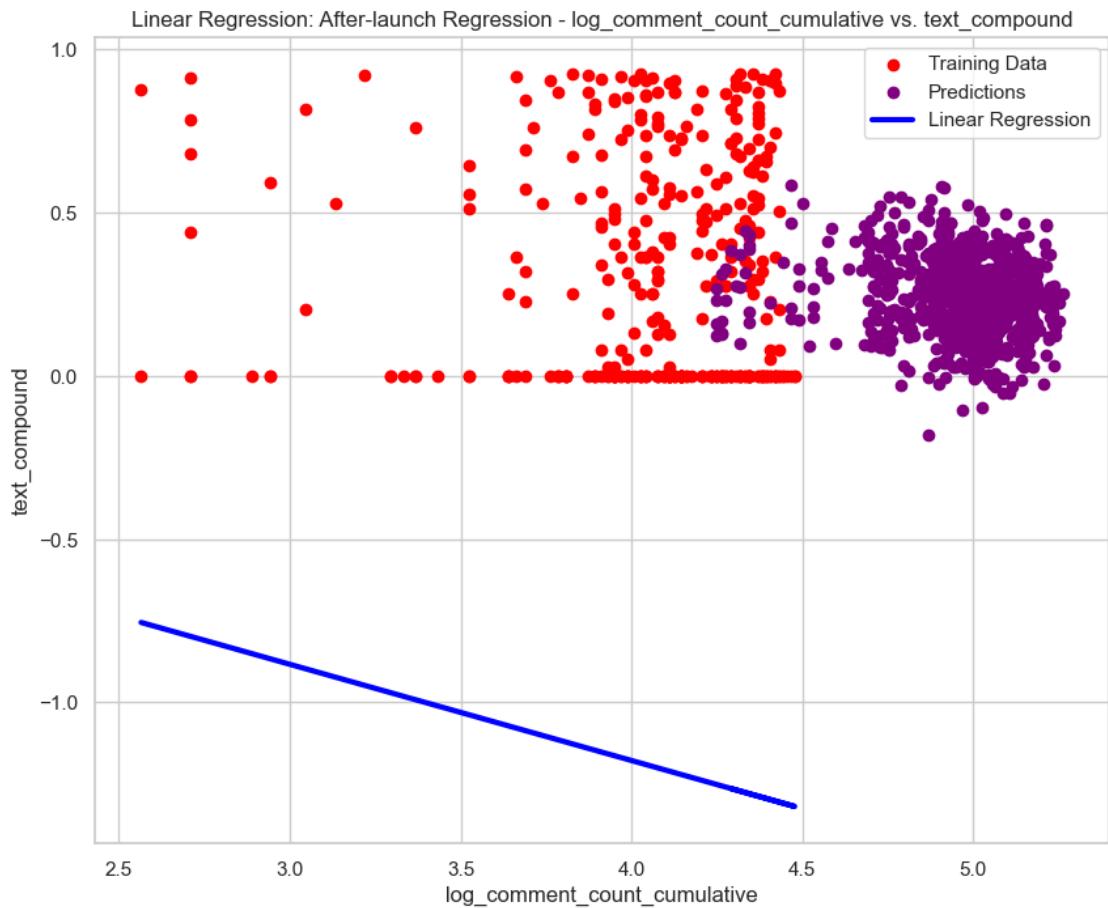


log_count_cumulative=-0.057271607131635024

Linear Regression: After-launch Regression - log_count_cumulative_lagged vs. text_compound

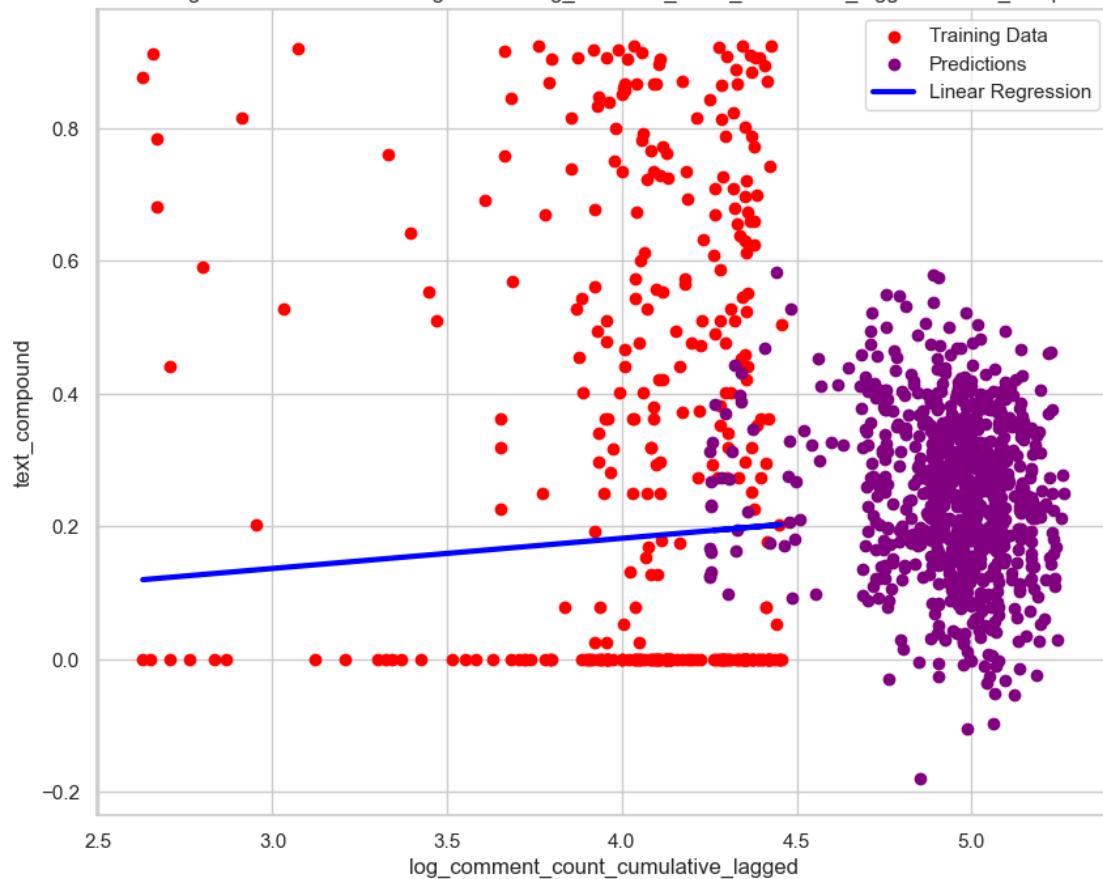


log_count_cumulative_lagged=0.1052940490925389

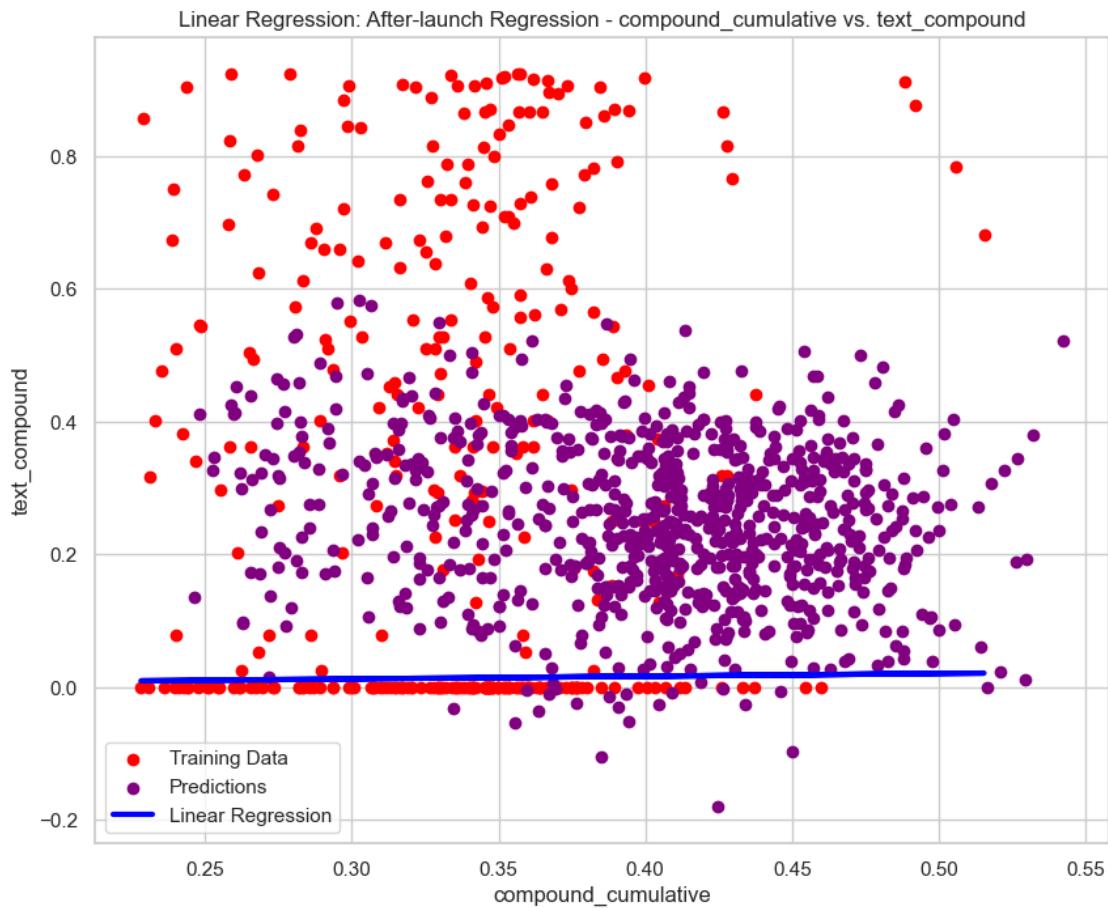


log_comment_count_cumulative=-0.2944145365185362

Linear Regression: After-launch Regression - log_comment_count_cumulative_lagged vs. text_compound

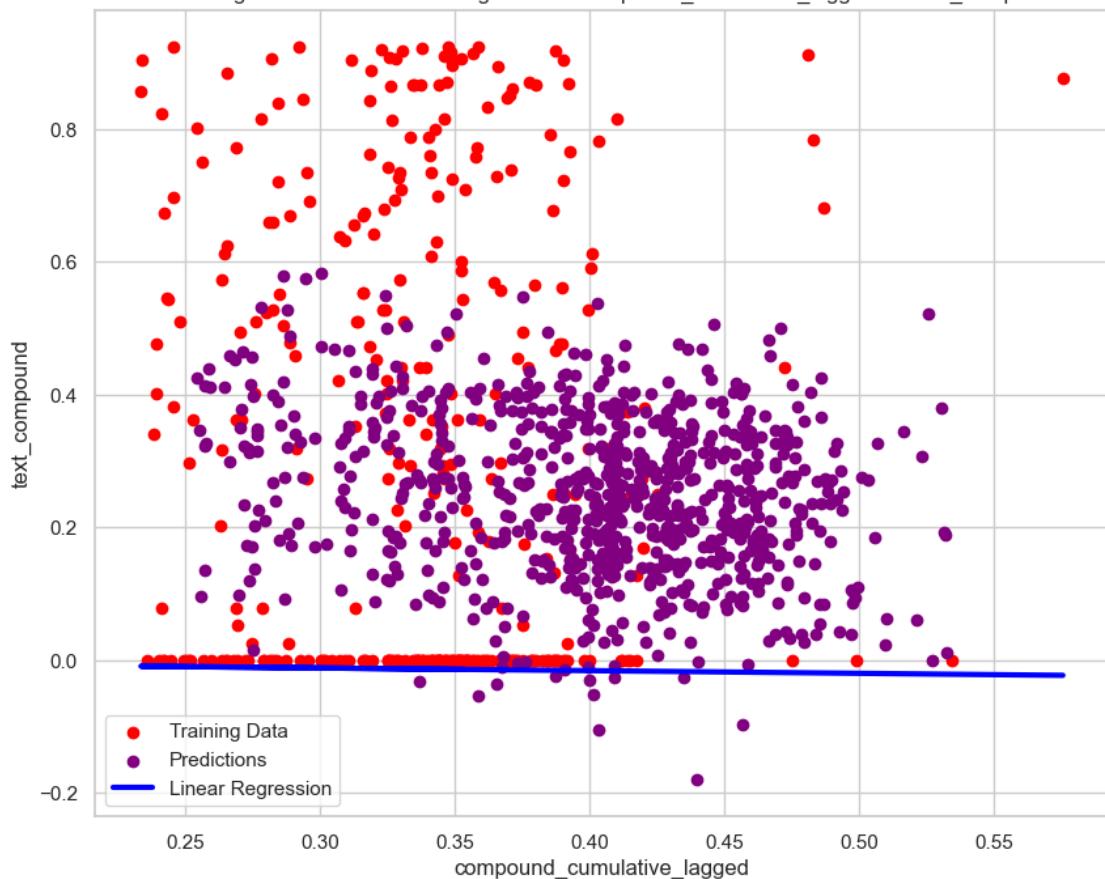


log_comment_count_cumulative_lagged=0.04542975473676008



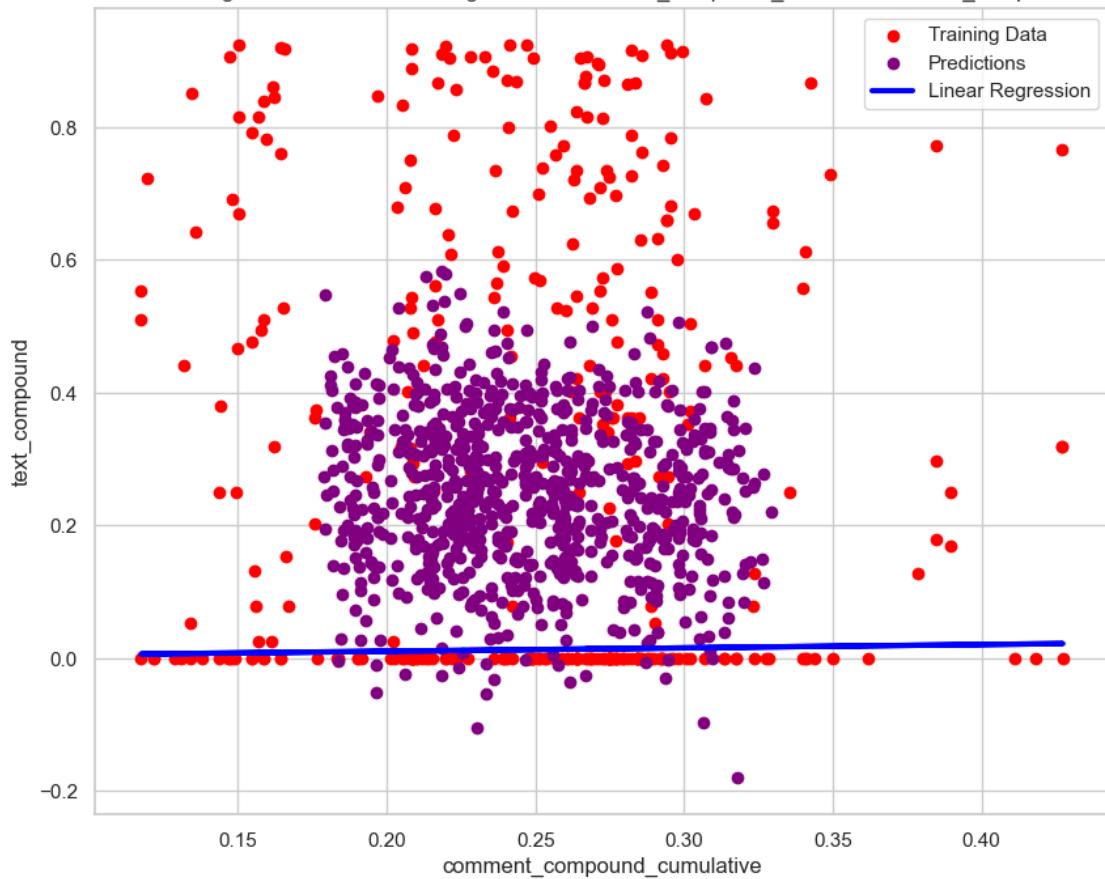
compound_cumulative=0.04047295863678872

Linear Regression: After-launch Regression - compound_cumulative_lagged vs. text_compound



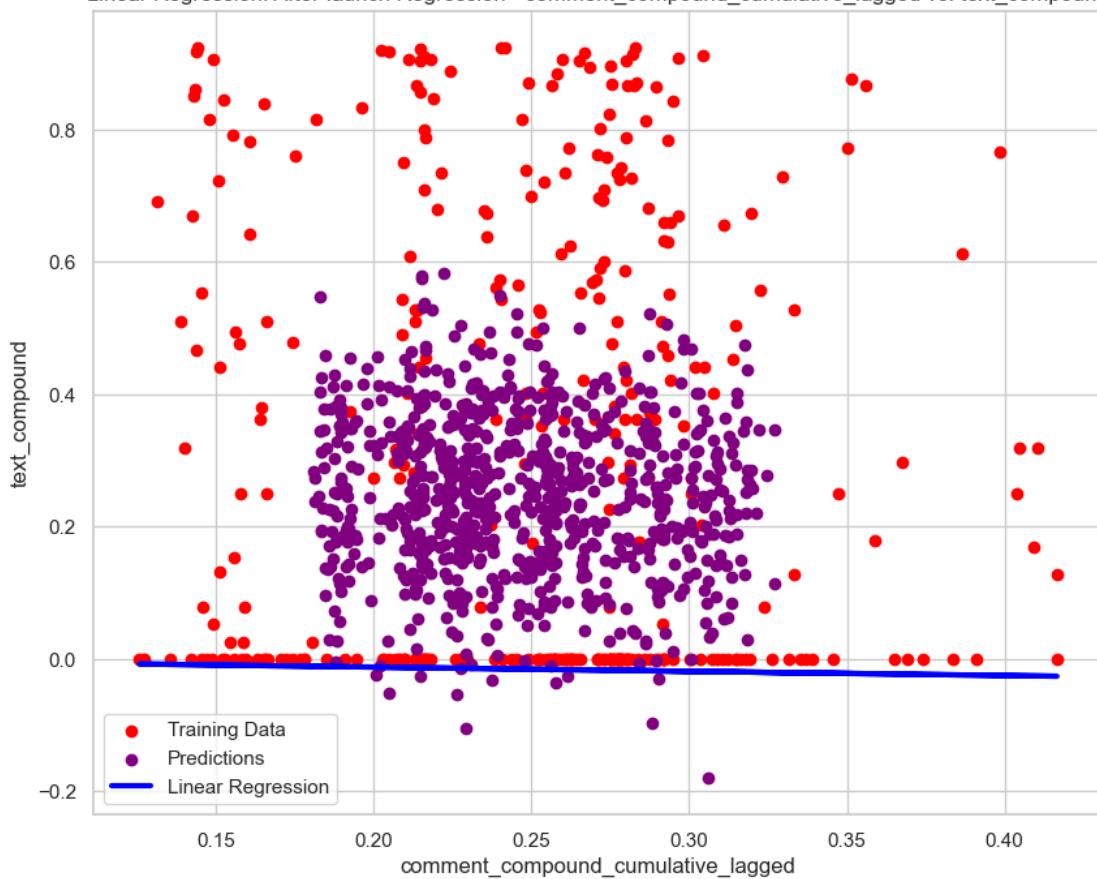
compound_cumulative_lagged=-0.04024050329275172

Linear Regression: After-launch Regression - comment_compound_cumulative vs. text_compound

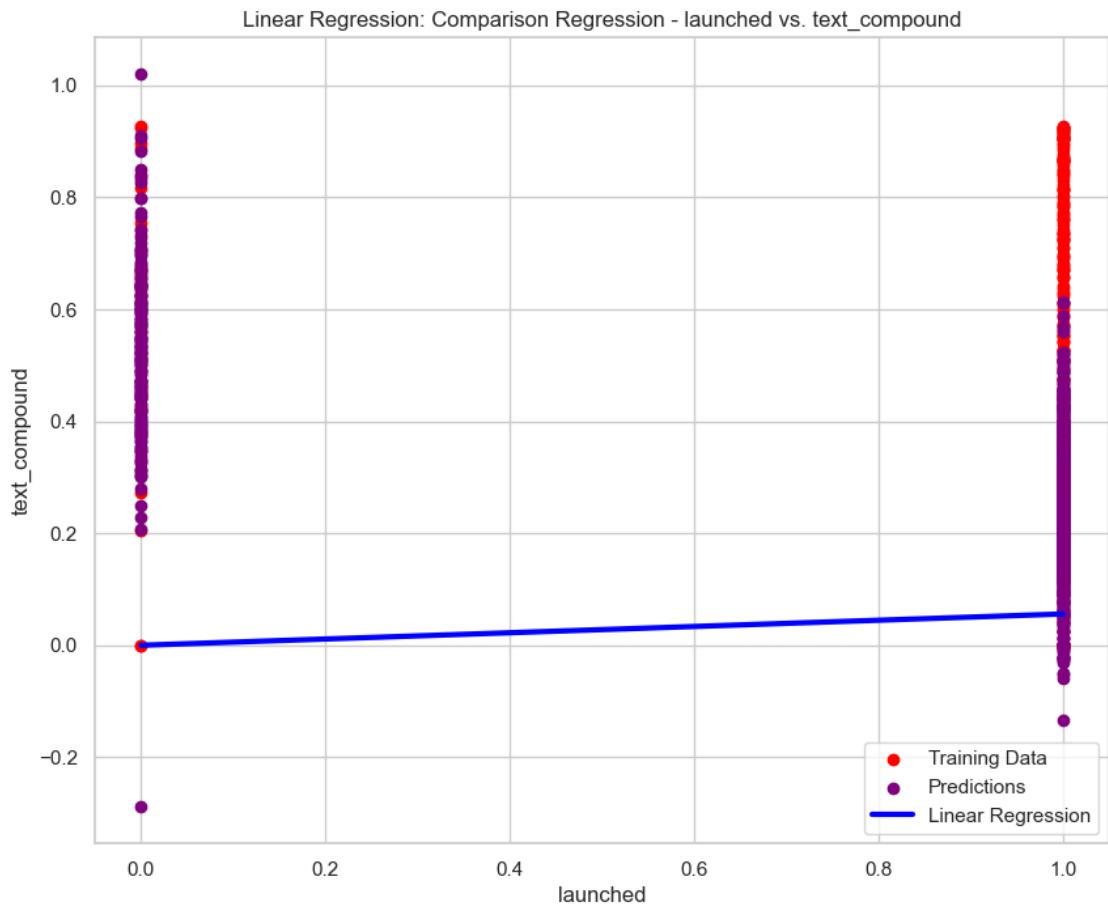


comment_compound_cumulative=0.050572671040547086

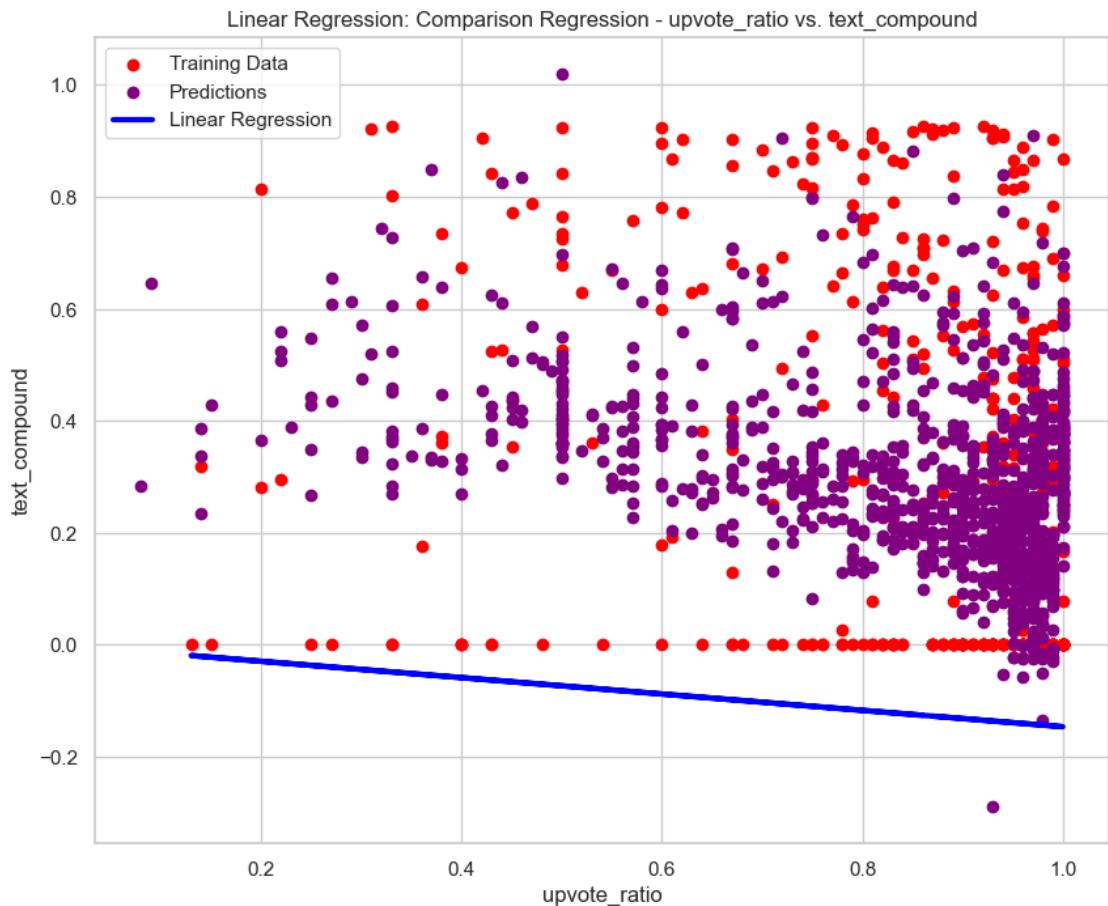
Linear Regression: After-launch Regression - comment_compound_cumulative_lagged vs. text_compound



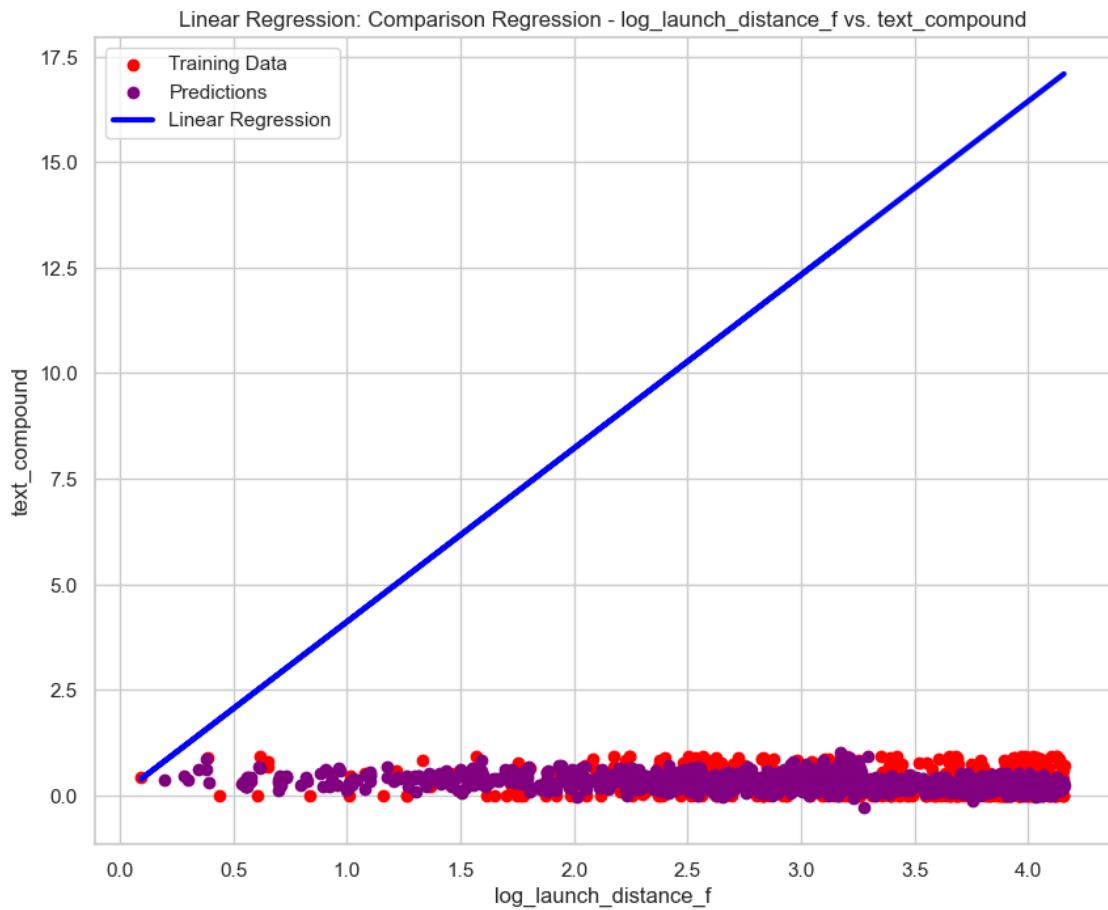
comment_compound_cumulative_lagged=-0.06349700973092488



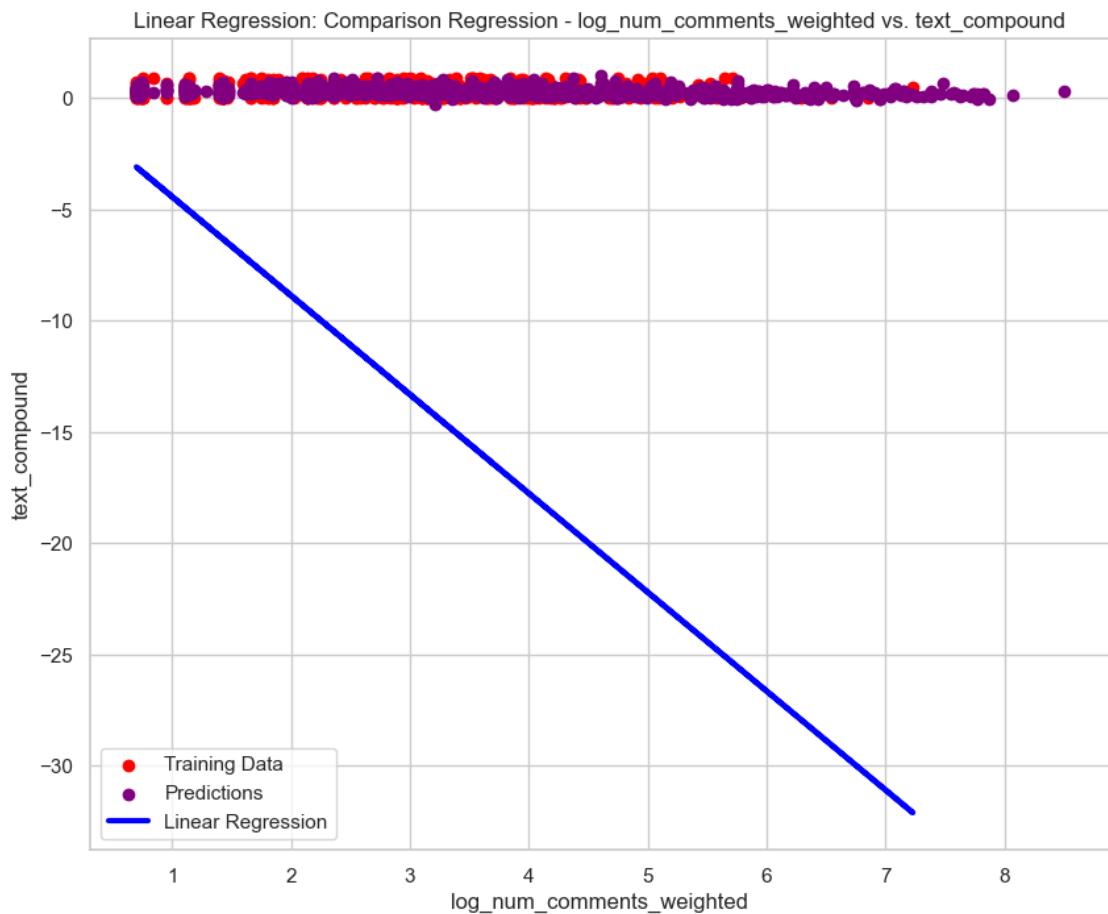
launched=0.05588844181328424



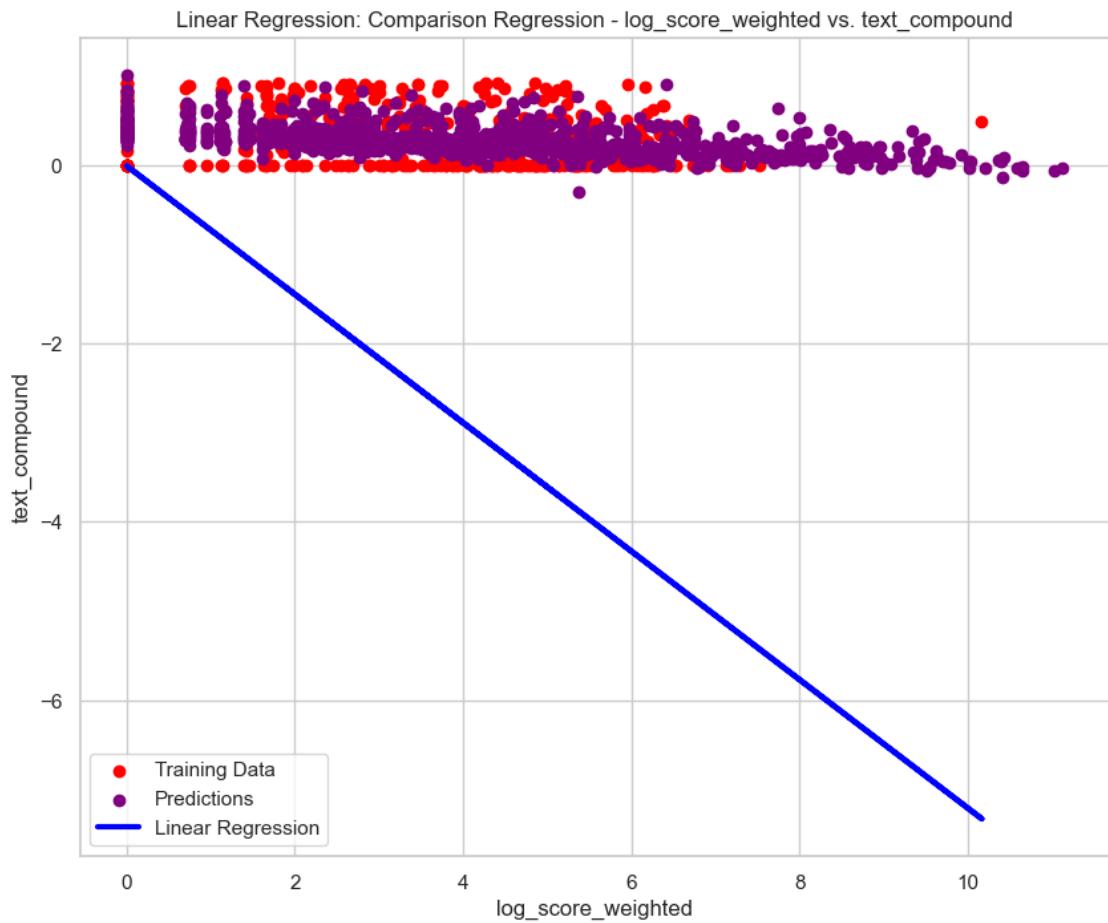
upvote_ratio=-0.14611173556891047



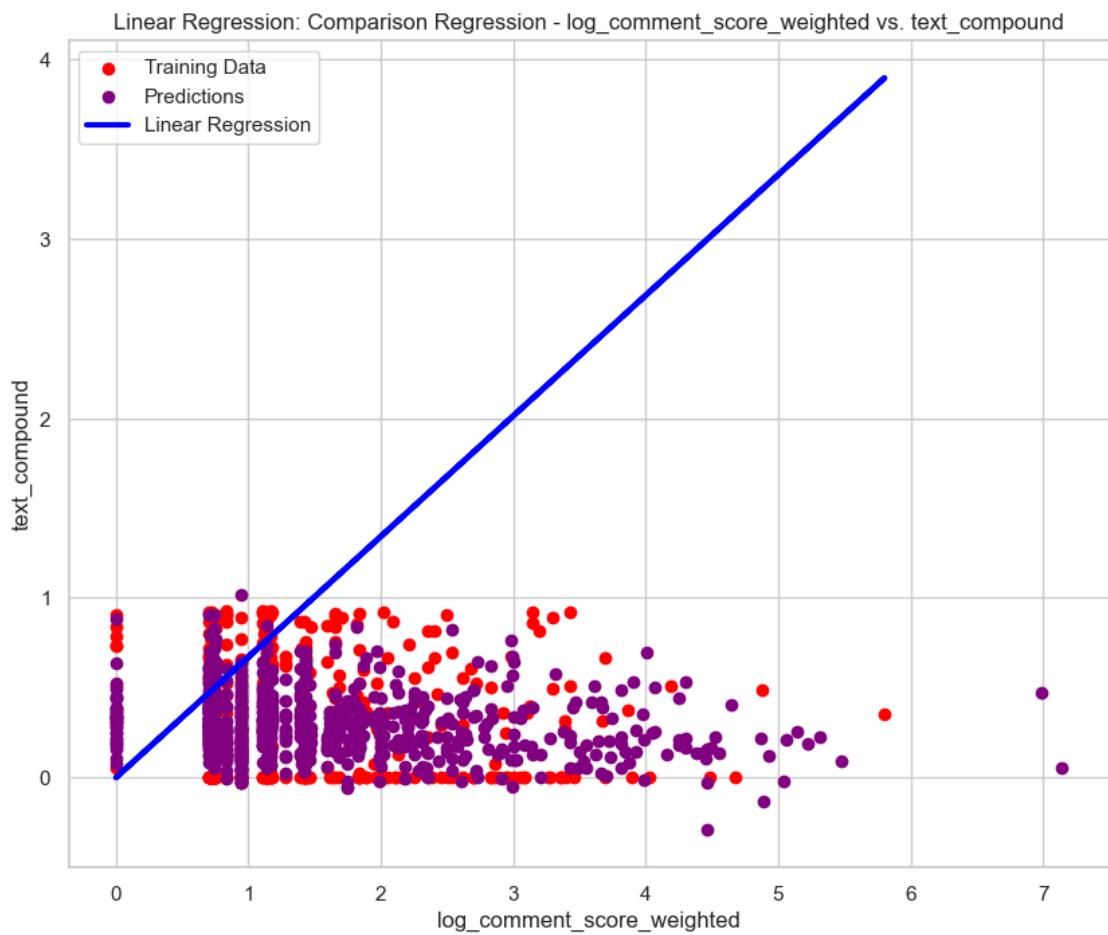
log_launch_distance_f=4.110417064926188



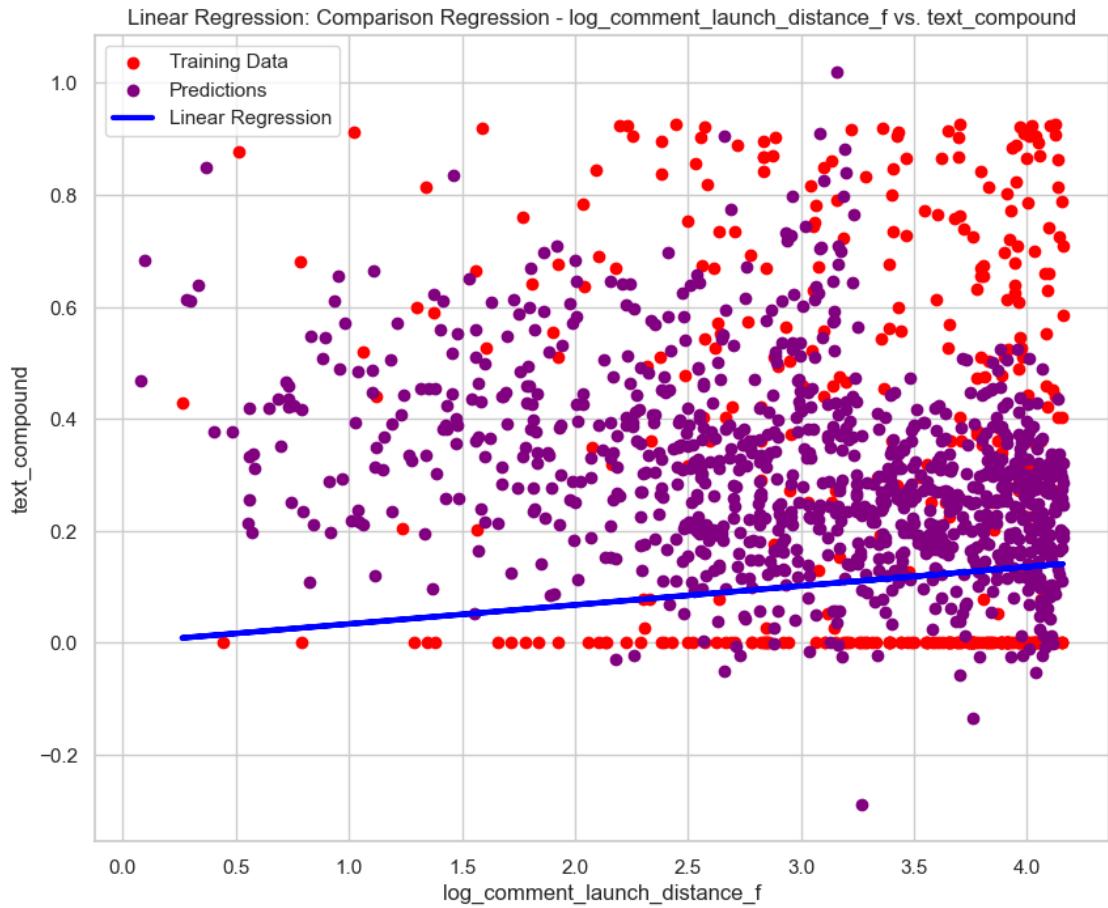
log_num_comments_weighted=-4.442379487093523



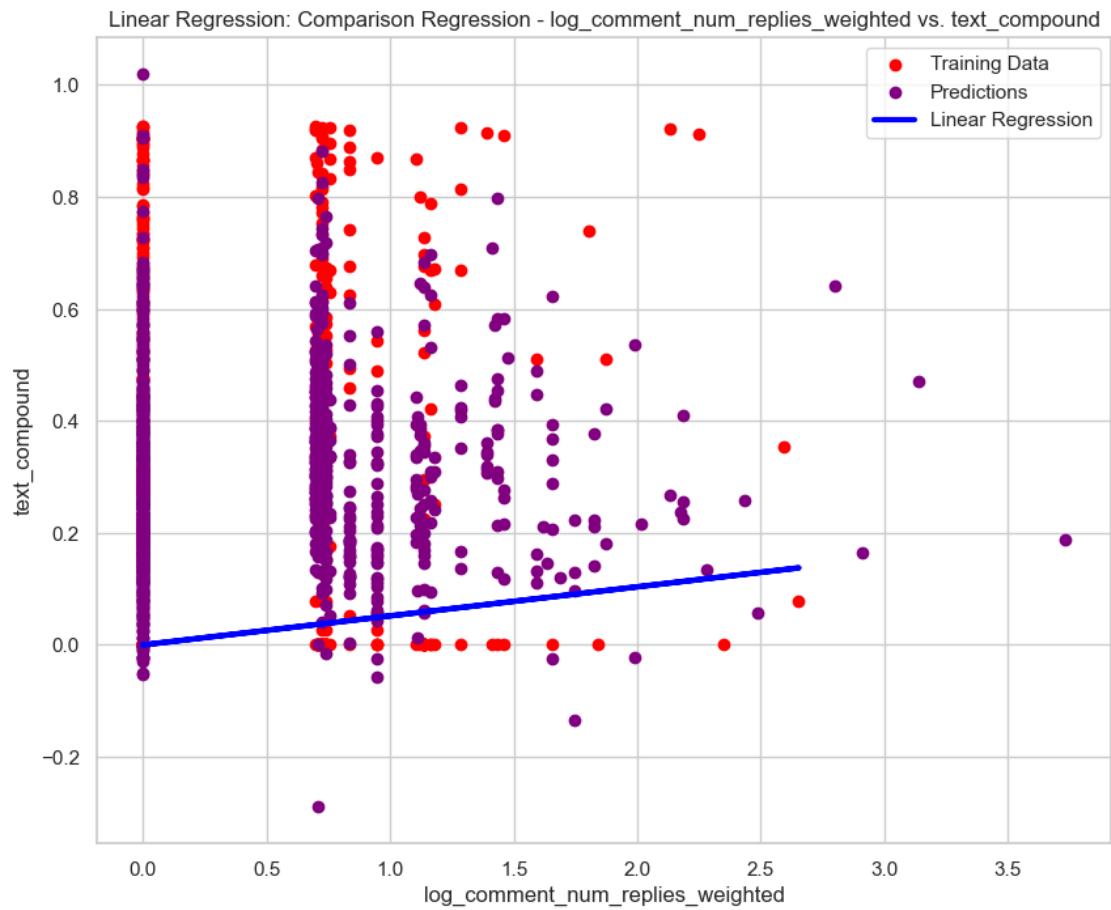
log_score_weighted=-0.7213130189659314



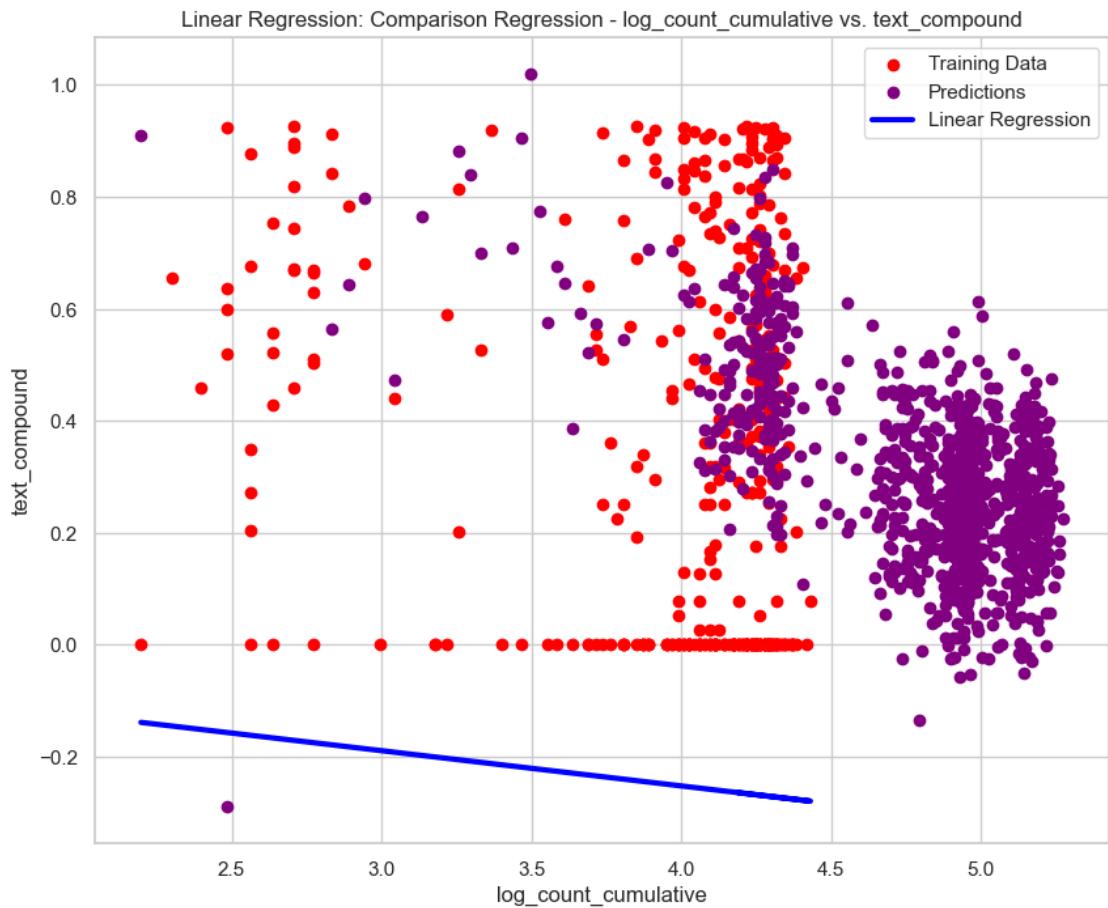
log_comment_score_weighted=0.6710407080769921



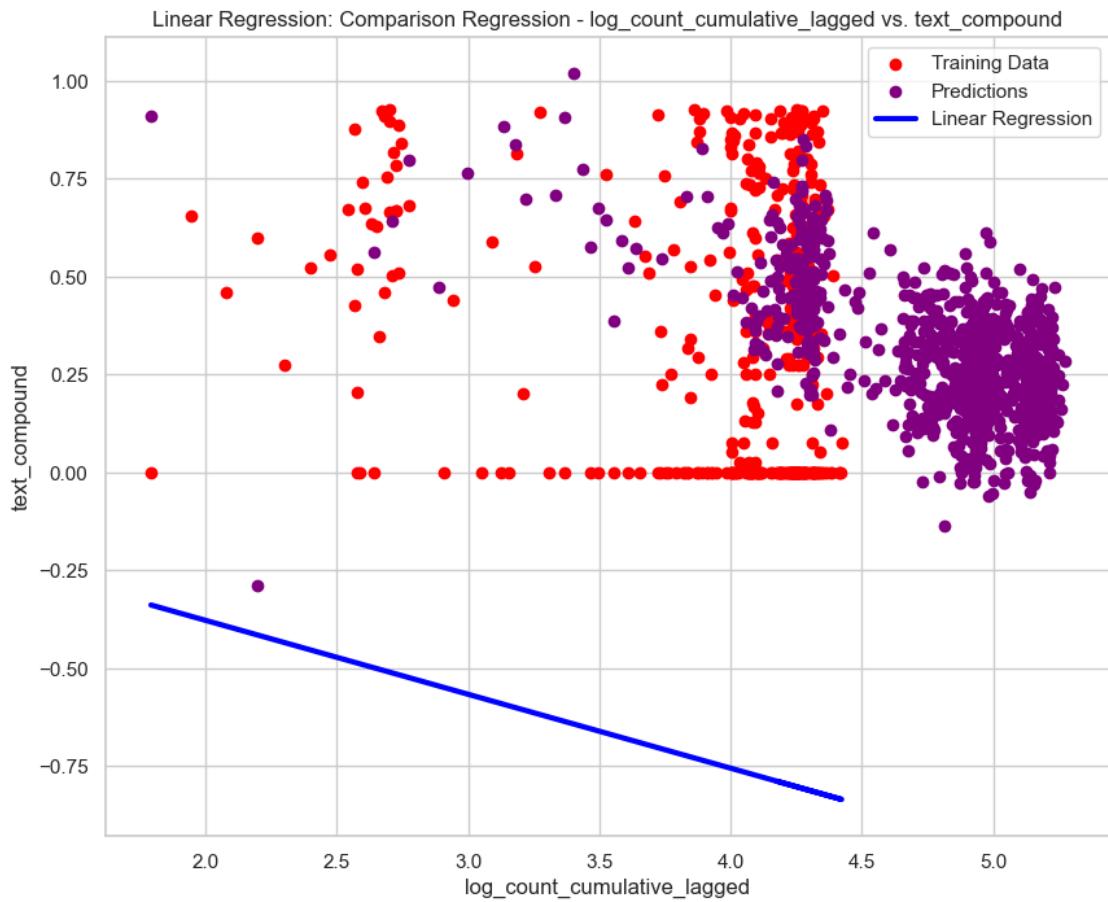
log_comment_launch_distance_f=0.03393277744007053



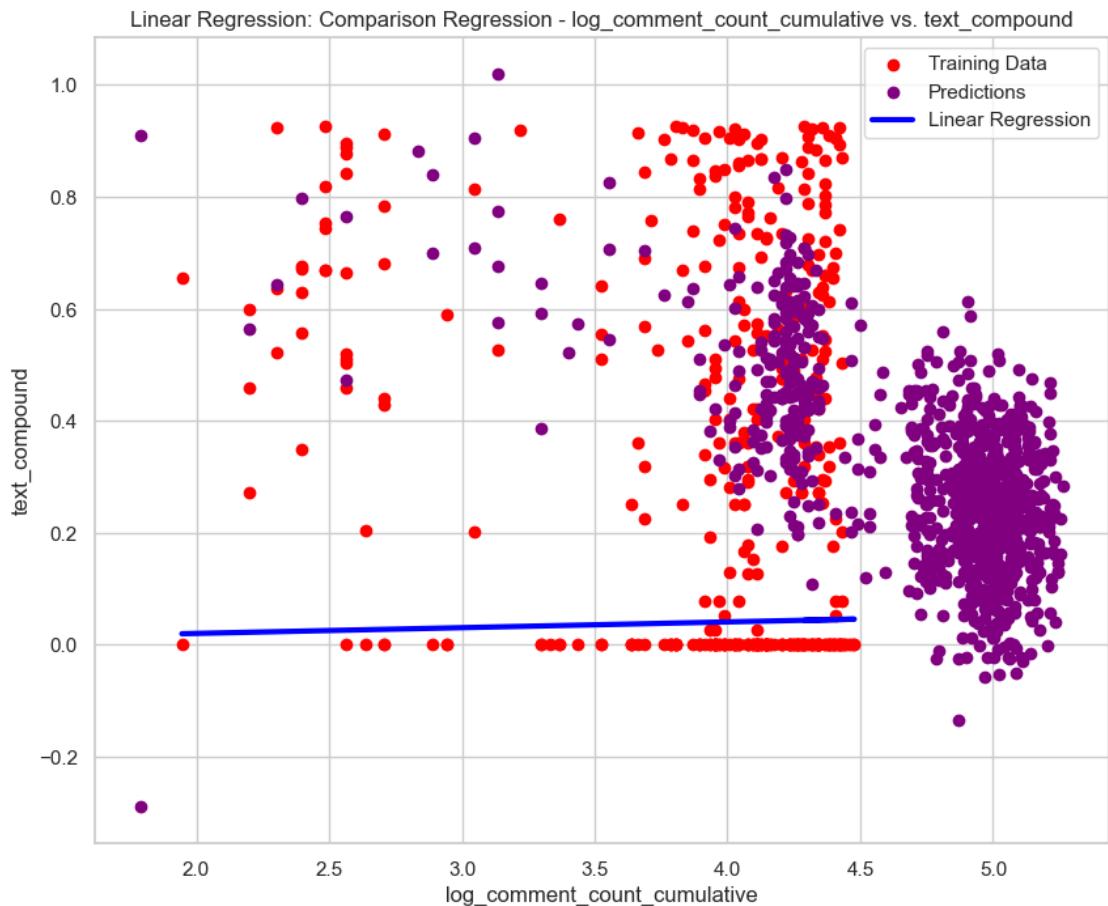
log_comment_num_replies_weighted=0.051827078897513826



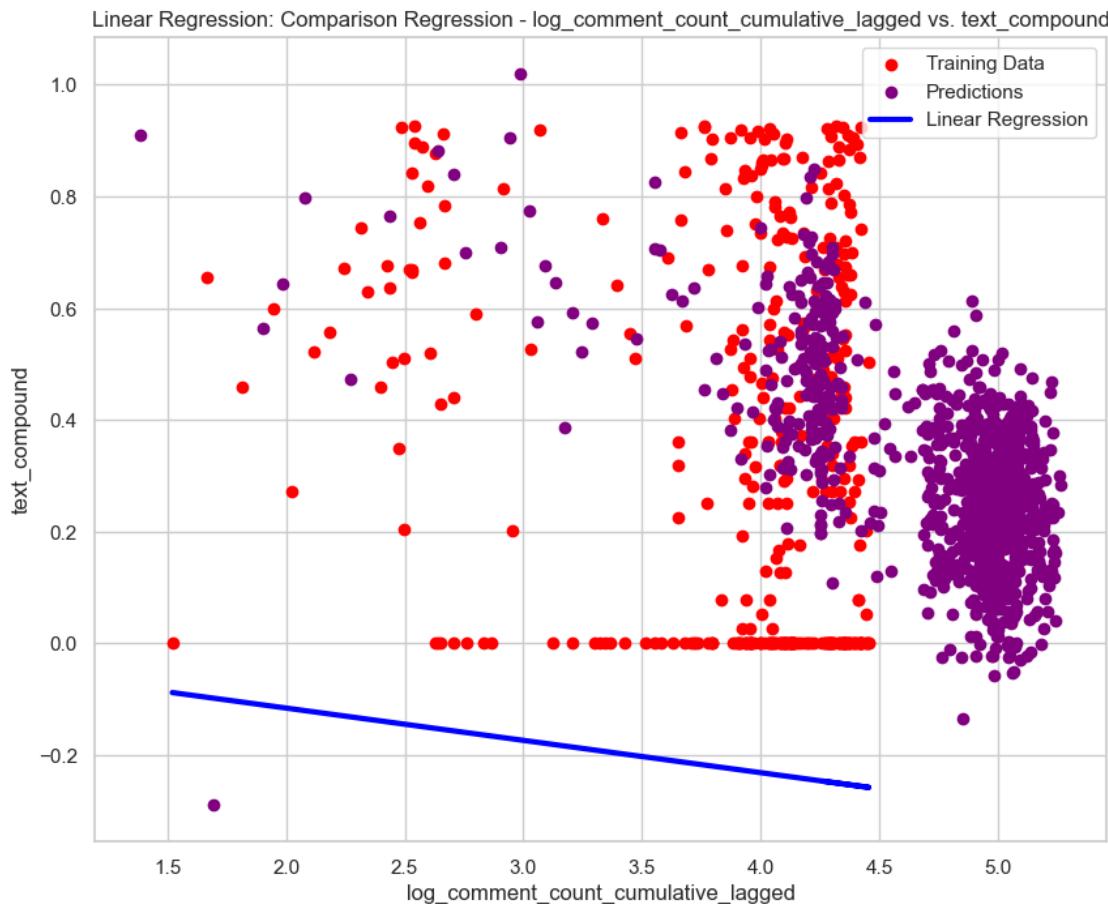
log_count_cumulative=-0.06295645876806528



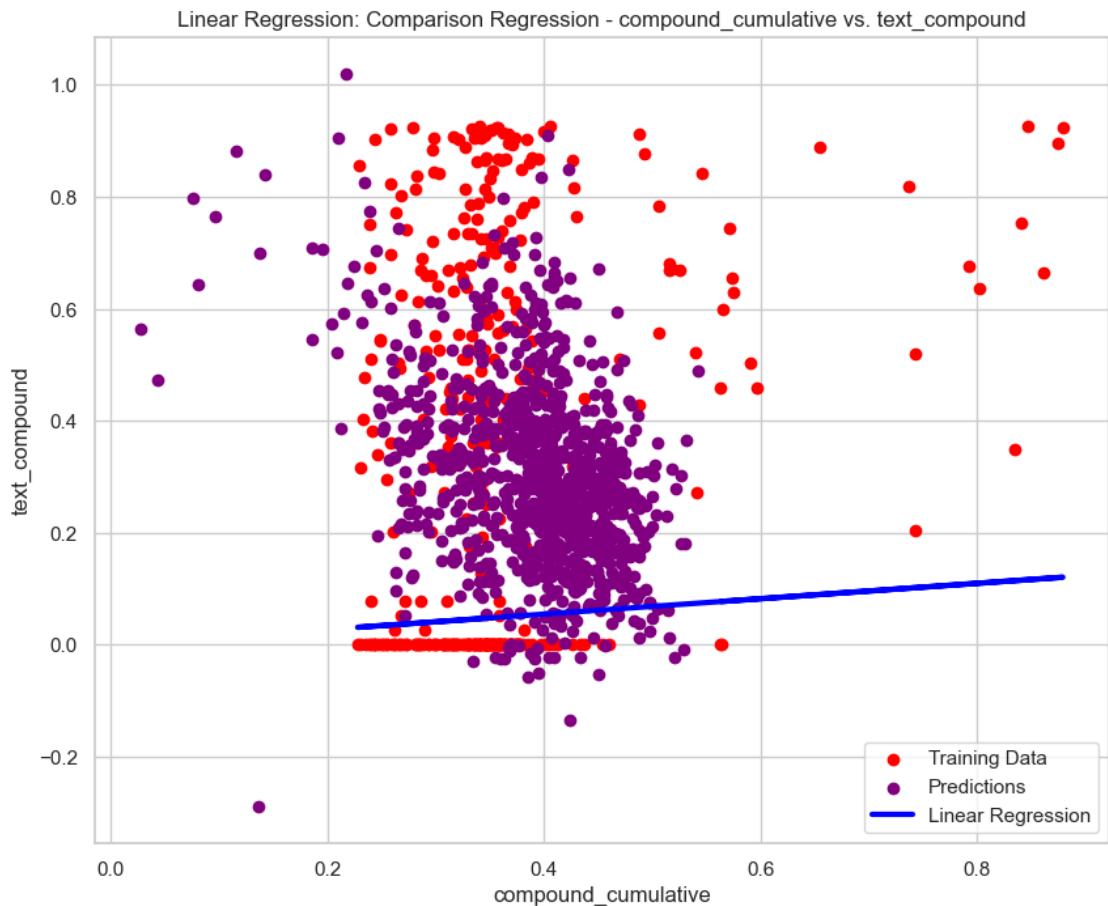
log_count_cumulative_lagged=-0.18871372467949127



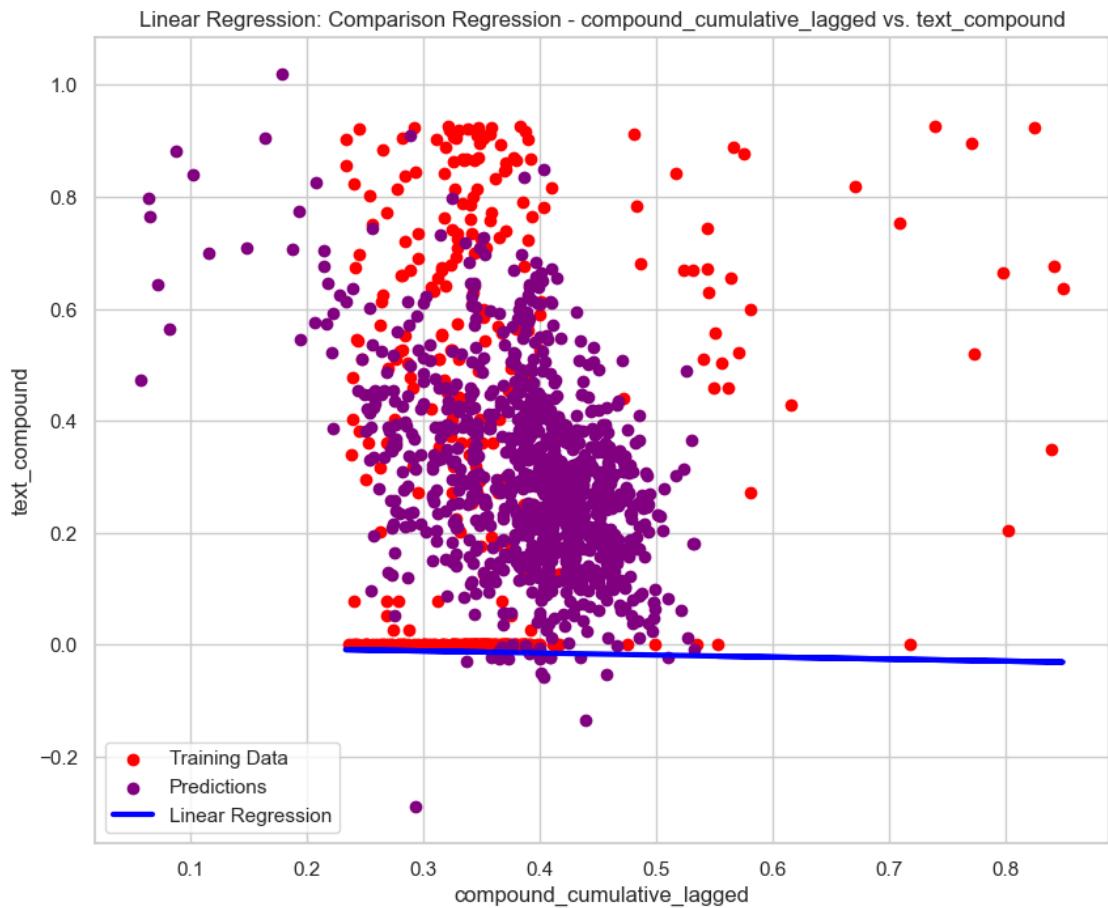
log_comment_count_cumulative=0.010256109768233193



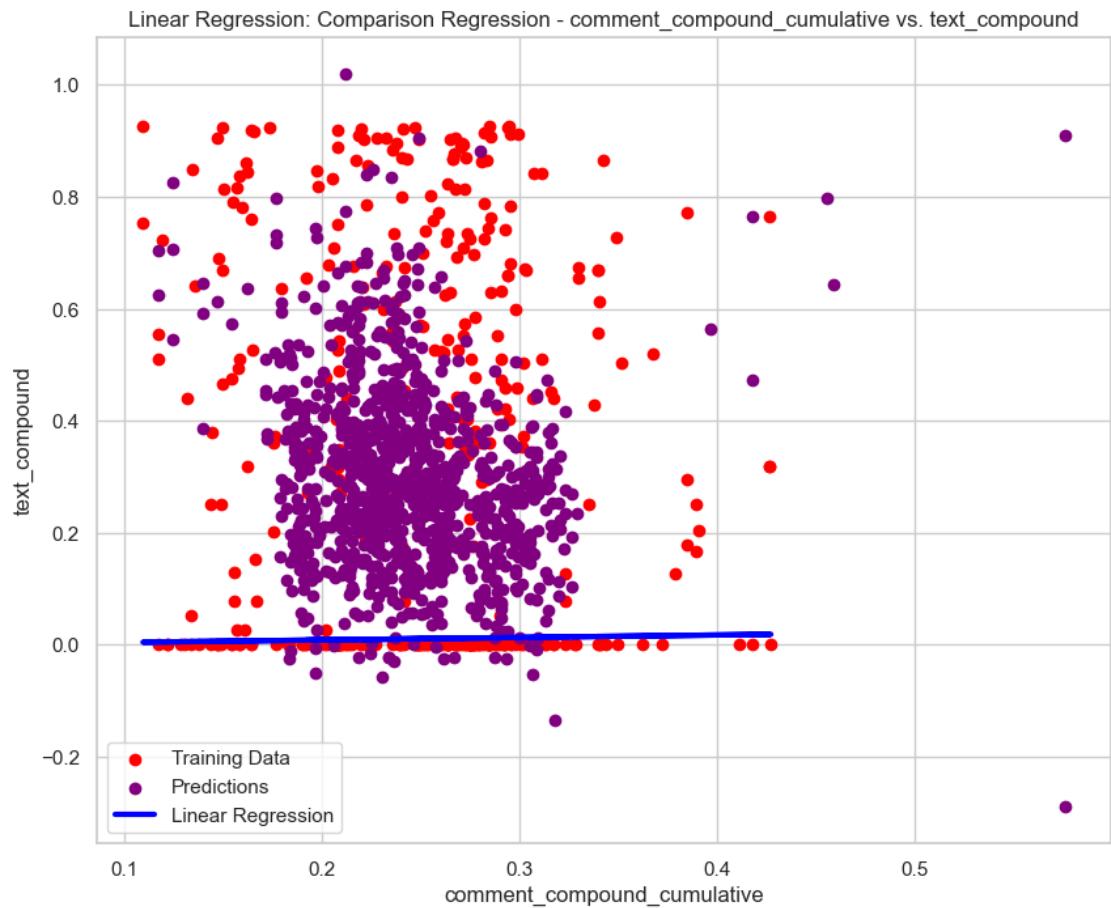
log_comment_count_cumulative_lagged=-0.057775634094383486



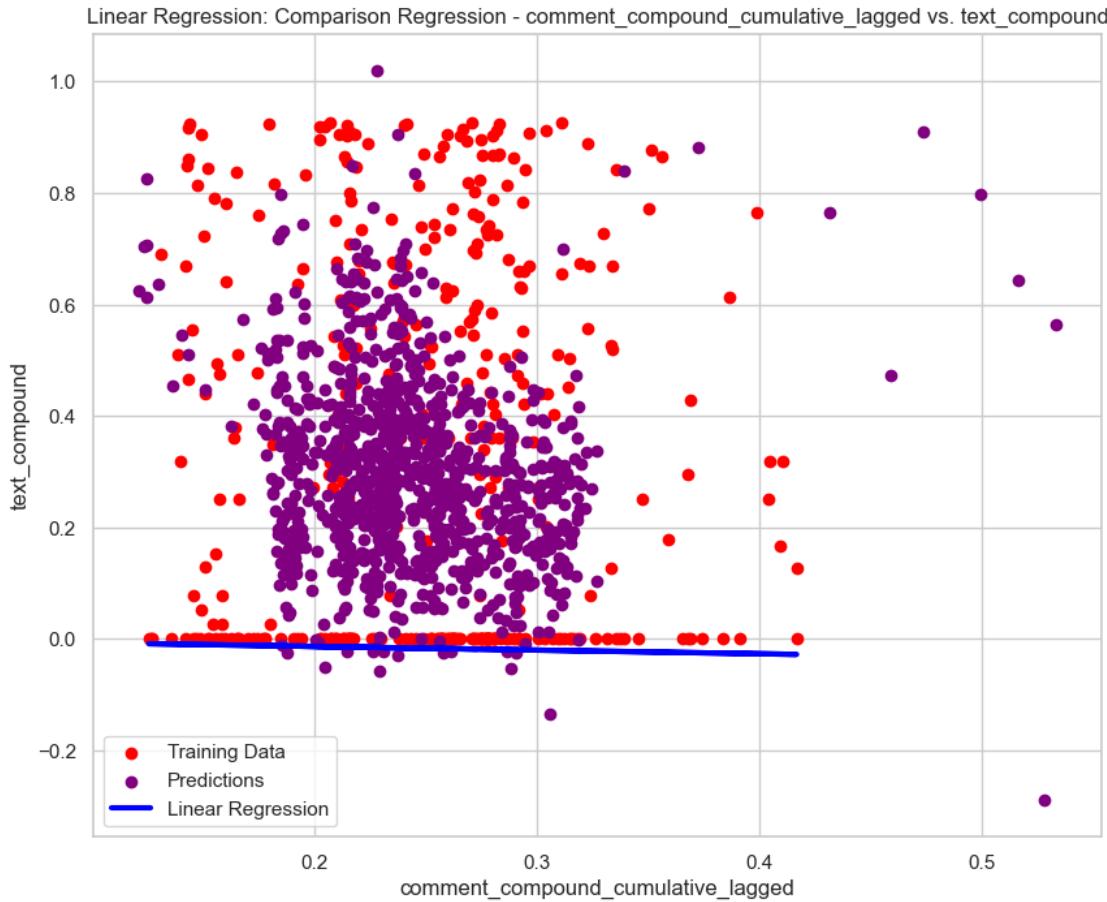
compound_cumulative=0.13747957608399317



compound_cumulative_lagged=-0.03618063260007022



comment_compound_cumulative=0.04422585608206971



```
comment_compound_cumulative_lagged=-0.06577664702367982
```

```
Index(['upvote_ratio', 'launched', 'compound_cumulative',
       'compound_cumulative_lagged', 'comment_compound_cumulative',
       'comment_compound_cumulative_lagged', 'log_launch_distance_f',
       'log_num_comments_weighted', 'log_score_weighted',
       'log_count_cumulative', 'log_count_cumulative_lagged',
       'log_comment_count_cumulative', 'log_comment_count_cumulative_lagged',
       'log_comment_score_weighted', 'log_comment_num_replies_weighted',
       'log_comment_launch_distance_f'],
      dtype='object')
```

9 Preparing MLflow features and targets for Random Forest Classifier

```
feature_names=['launched', 'upvote_ratio', 'log_launch_distance_f',
               'log_num_comments_weighted', 'log_score_weighted', 'log_comment_score_weighted',
               'log_comment_launch_distance_f', 'log_comment_num_replies_weighted',
```

```
'log_count_cumulative', 'log_count_cumulative_lagged',
'log_comment_count_cumulative', 'log_comment_count_cumulative_lagged',
'compound_cumulative', 'compound_cumulative_lagged',
'comment_compound_cumulative', 'comment_compound_cumulative_lagged']
target_name=text_sentiment
```

10 Run MLflow Random Forest experiment

Test model against data after GPT-4 launch.

```
2024/02/26 02:44:33 WARNING mlflow.sklearn.utils: roc_auc_score failed. The
metric training_roc_auc will not be recorded. Metric error: Encoders require
their input to be uniformly strings or numbers. Got ['ndarray']
2024/02/26 02:44:33 WARNING mlflow.sklearn.utils: Failed to autolog artifacts
for RandomForestClassifier. Logging error: unhashable type: 'numpy.ndarray'
```

Coefficients of GPT-3 Classifier

Feature	Importance
log_comment_count_cumulative_lagged	0.12
log_launch_distance_f	0.12
log_comment_launch_distance_f	0.12
compound_cumulative	0.11
upvote_ratio	0.10
comment_compound_cumulative_lagged	0.09
log_count_cumulative	0.06
comment_compound_cumulative	0.06
log_comment_score_weighted	0.05
log_count_cumulative_lagged	0.05
log_score_weighted	0.04
log_num_comments_weighted	0.04
log_comment_count_cumulative	0.02
compound_cumulative_lagged	0.00
log_comment_num_replies_weighted	0.00
launched	0.00

```
2024/02/26 02:44:39 WARNING mlflow.sklearn.utils: roc_auc_score failed. The
metric training_roc_auc will not be recorded. Metric error: Encoders require
their input to be uniformly strings or numbers. Got ['ndarray']
```

```
2024/02/26 02:44:39 WARNING mlflow.sklearn.utils: Failed to autolog artifacts
for RandomForestClassifier. Logging error: unhashable type: 'numpy.ndarray'
```

Coefficients of GPT-4 Classifier

Feature	Importance
log_score_weighted	0.11

log_num_comments_weighted	0.10
upvote_ratio	0.09
compound_cumulative	0.08
comment_compound_cumulative_lagged	0.07
compound_cumulative_lagged	0.07
log_comment_launch_distance_f	0.06
comment_compound_cumulative	0.06
log_comment_score_weighted	0.06
log_count_cumulative_lagged	0.06
log_comment_count_cumulative_lagged	0.05
log_launch_distance_f	0.05
log_count_cumulative	0.05
log_comment_count_cumulative	0.04
log_comment_num_replies_weighted	0.03
launched	0.00

2024/02/26 02:44:44 WARNING mlflow.sklearn.utils: roc_auc_score failed. The metric training_roc_auc will not be recorded. Metric error: Encoders require their input to be uniformly strings or numbers. Got ['ndarray']

2024/02/26 02:44:44 WARNING mlflow.sklearn.utils: Failed to autolog artifacts for RandomForestClassifier. Logging error: unhashable type: 'numpy.ndarray'

Coefficients of Before-launch Classifier

Feature	Importance
log_comment_count_cumulative_lagged	0.13
log_launch_distance_f	0.12
log_comment_launch_distance_f	0.11
upvote_ratio	0.10
compound_cumulative	0.09
comment_compound_cumulative_lagged	0.09
comment_compound_cumulative	0.07
log_count_cumulative	0.06
log_score_weighted	0.05
log_num_comments_weighted	0.05
log_comment_score_weighted	0.04
log_count_cumulative_lagged	0.04
log_comment_count_cumulative	0.03
log_comment_num_replies_weighted	0.01
compound_cumulative_lagged	0.00
launched	0.00

2024/02/26 02:44:50 WARNING mlflow.sklearn.utils: roc_auc_score failed. The metric training_roc_auc will not be recorded. Metric error: Encoders require their input to be uniformly strings or numbers. Got ['ndarray']

2024/02/26 02:44:50 WARNING mlflow.sklearn.utils: Failed to autolog artifacts

```
for RandomForestClassifier. Logging error: unhashable type: 'numpy.ndarray'
```

```
Coefficients of After-launch Classifier
```

Feature	Importance
compound_cumulative	0.10
log_score_weighted	0.09
compound_cumulative_lagged	0.09
log_num_comments_weighted	0.09
upvote_ratio	0.08
log_comment_score_weighted	0.07
comment_compound_cumulative_lagged	0.06
comment_compound_cumulative	0.06
log_comment_launch_distance_f	0.06
log_launch_distance_f	0.06
log_comment_count_cumulative_lagged	0.06
log_count_cumulative_lagged	0.05
log_comment_count_cumulative	0.05
log_count_cumulative	0.05
log_comment_num_replies_weighted	0.03
launched	0.00

```
2024/02/26 02:44:56 WARNING mlflow.sklearn.utils: roc_auc_score failed. The metric training_roc_auc will not be recorded. Metric error: Encoders require their input to be uniformly strings or numbers. Got ['ndarray']
```

```
2024/02/26 02:44:56 WARNING mlflow.sklearn.utils: Failed to autolog artifacts for RandomForestClassifier. Logging error: unhashable type: 'numpy.ndarray'
```

```
Coefficients of Comparison Classifier
```

Feature	Importance
compound_cumulative	0.10
compound_cumulative_lagged	0.09
log_score_weighted	0.09
log_num_comments_weighted	0.08
upvote_ratio	0.07
log_comment_score_weighted	0.07
log_comment_count_cumulative_lagged	0.06
log_comment_launch_distance_f	0.06
comment_compound_cumulative_lagged	0.06
comment_compound_cumulative	0.06
log_launch_distance_f	0.06
log_count_cumulative_lagged	0.06
log_comment_count_cumulative	0.05
log_count_cumulative	0.05
log_comment_num_replies_weighted	0.03

```
|           launched          |    0.00   |
+-----+-----+
```

11 Heatmap of MLflow Random Forest runs

