

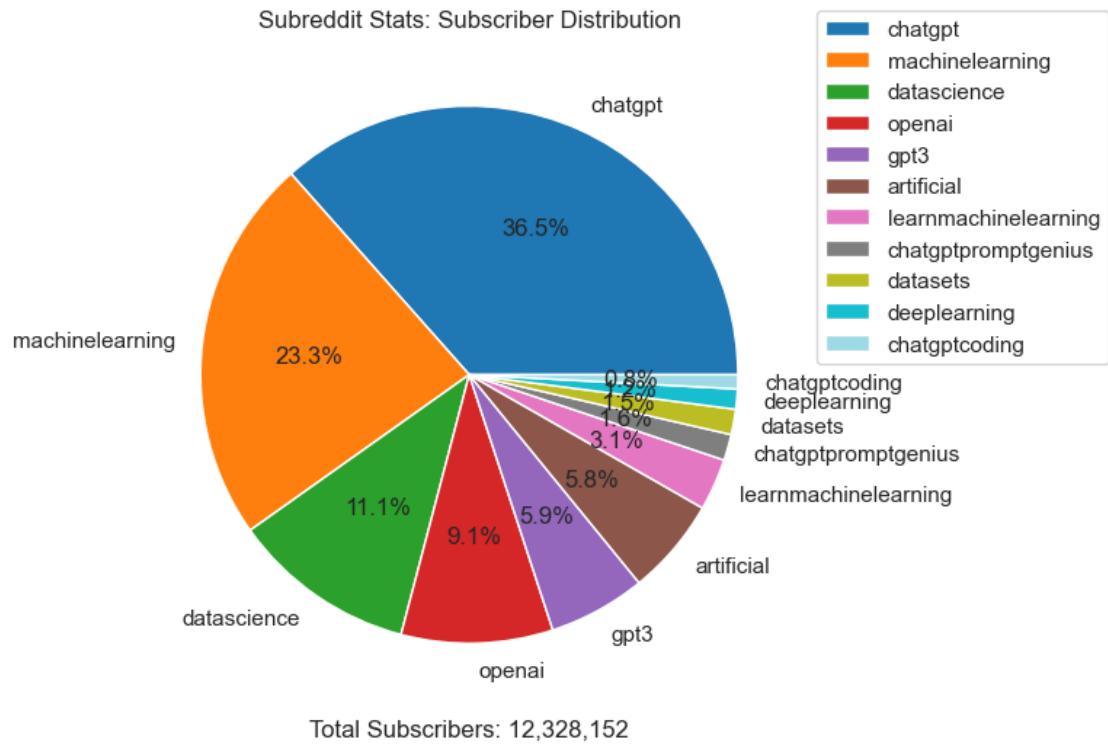
model-training

February 28, 2024

1 Reddit Subreddit API analysis

Subreddit Stats

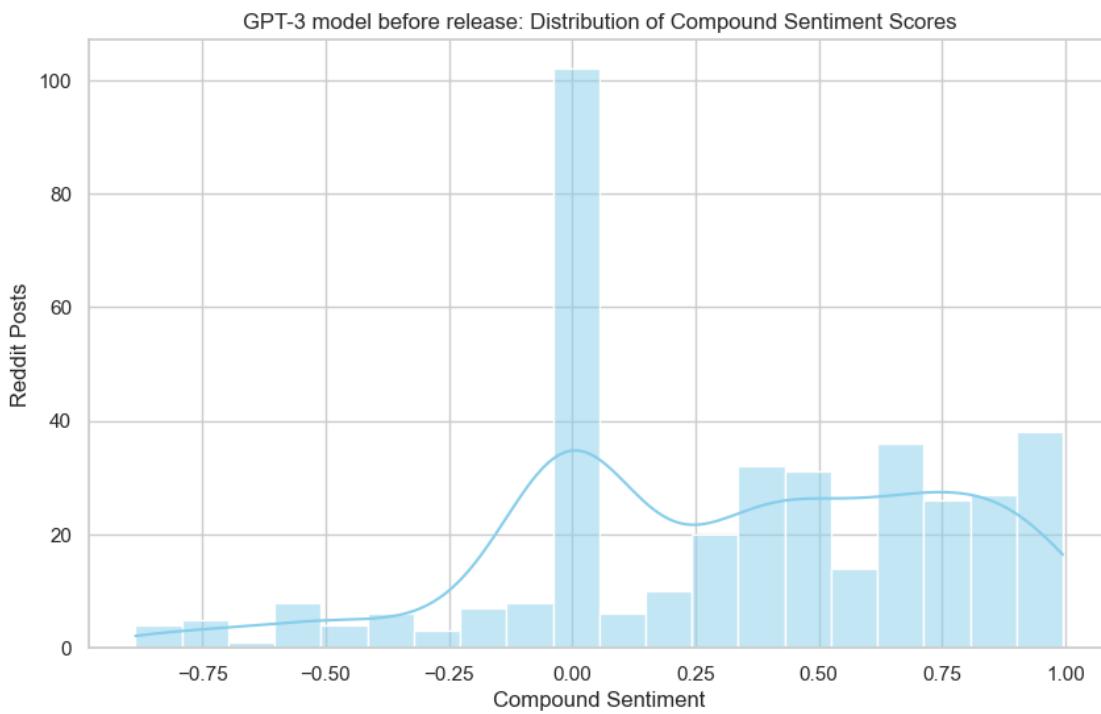
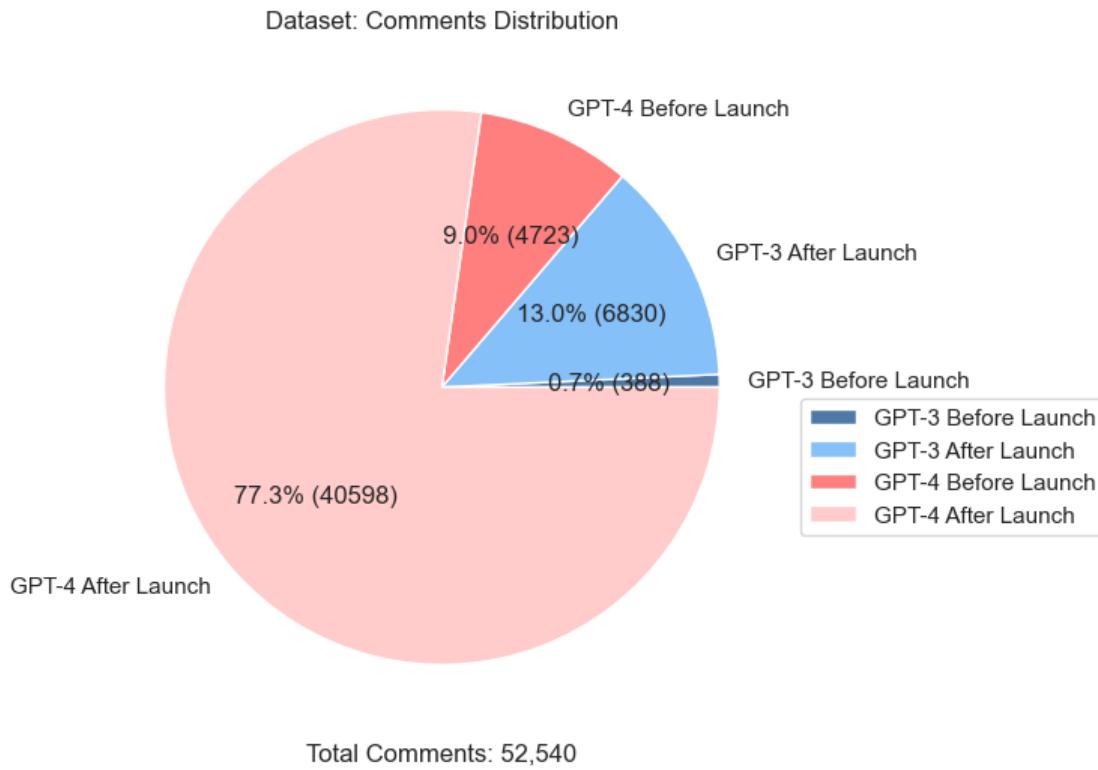
name	subscribers	subscribers_pct
chatgpt	4505500	36.55
machinelearning	2870456	23.28
datascience	1366348	11.08
openai	1124730	9.12
gpt3	722076	5.86
artificial	720095	5.84
learnmachinelearning	384514	3.12
chatgptpromptgenius	193280	1.57
datasets	187702	1.52
deeplearning	149090	1.21
chatgptcoding	104361	0.85



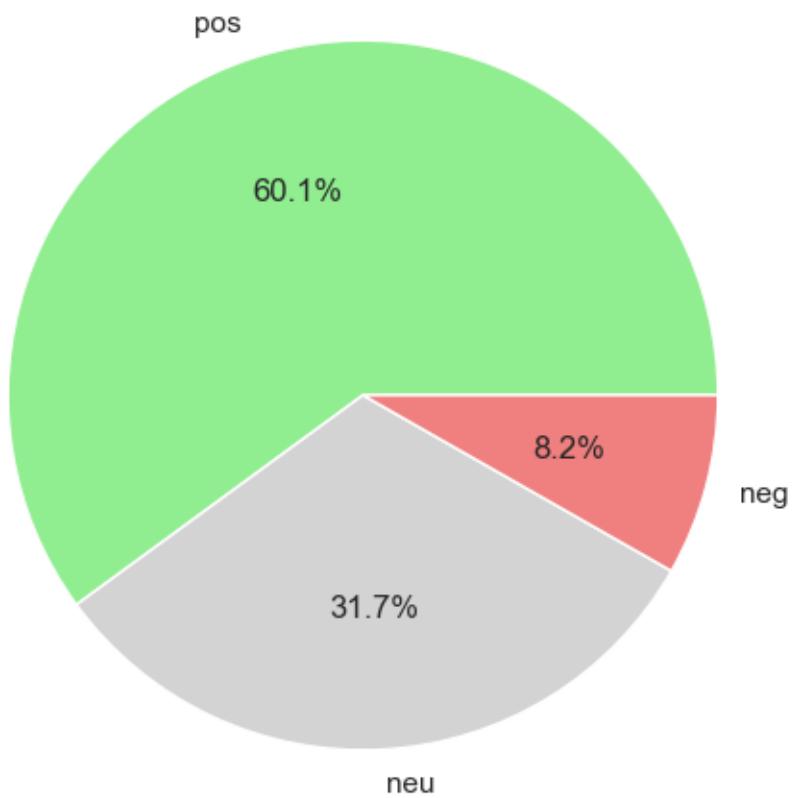
2 Reddit Search API data analysis

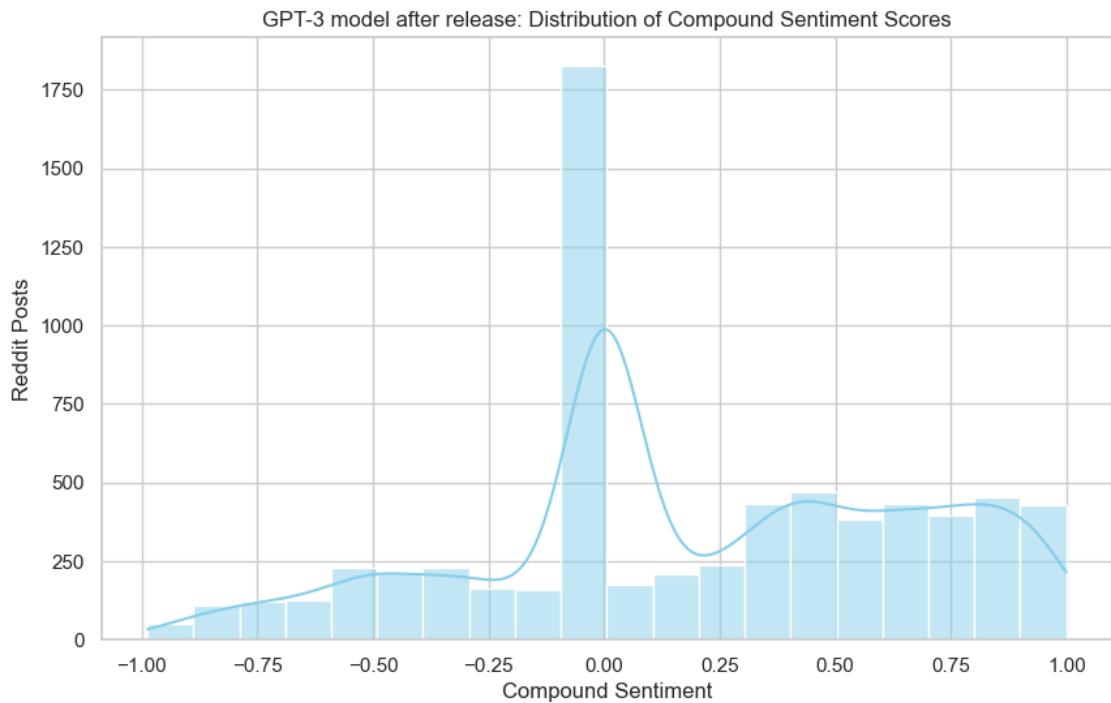
Total imported Reddit comments: 52540

3 Data split on launch dates

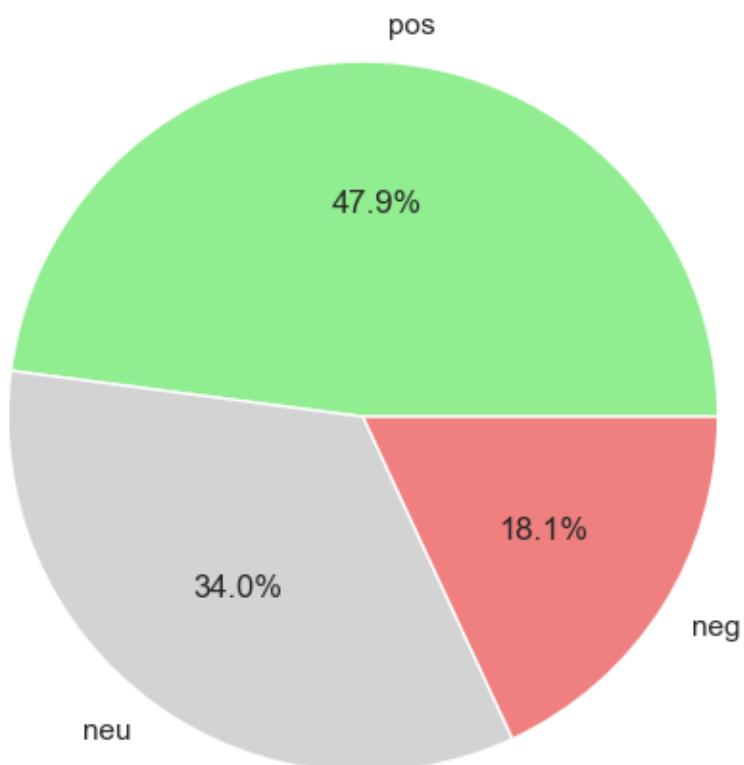


GPT-3 model before release: Sentiment Distribution

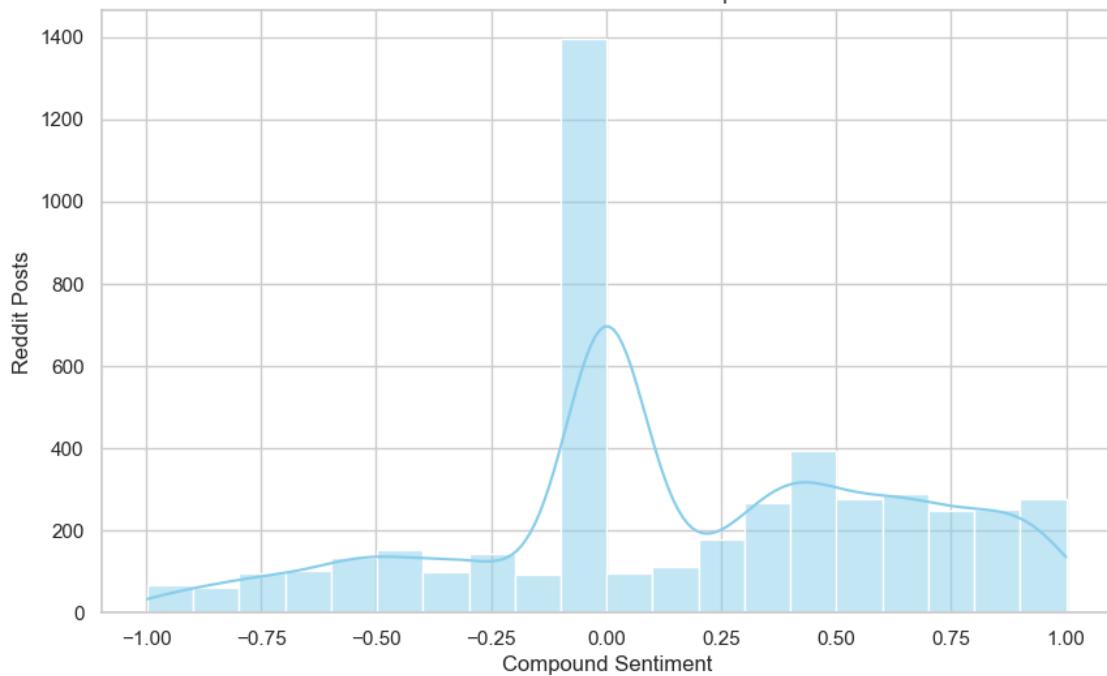




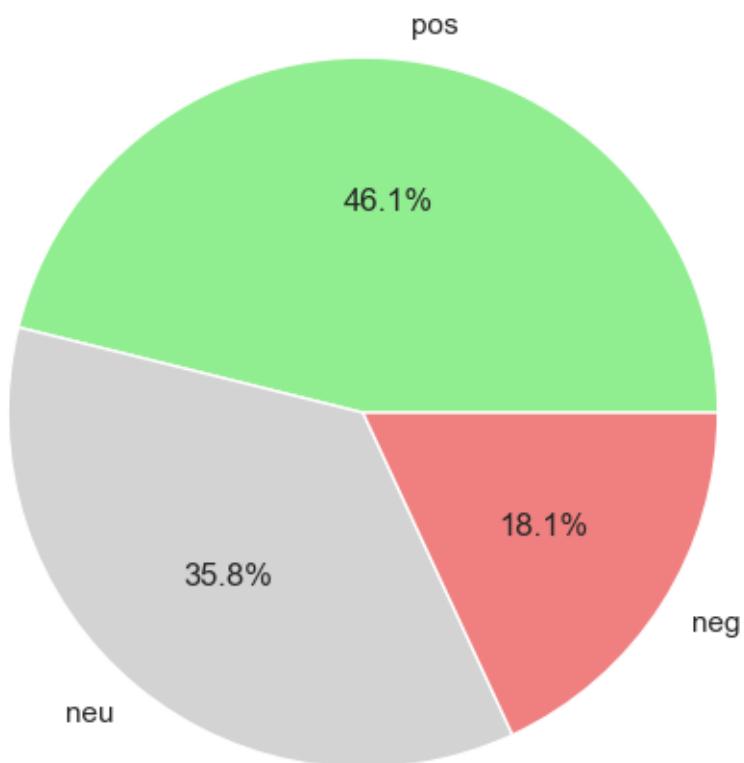
GPT-3 model after release: Sentiment Distribution

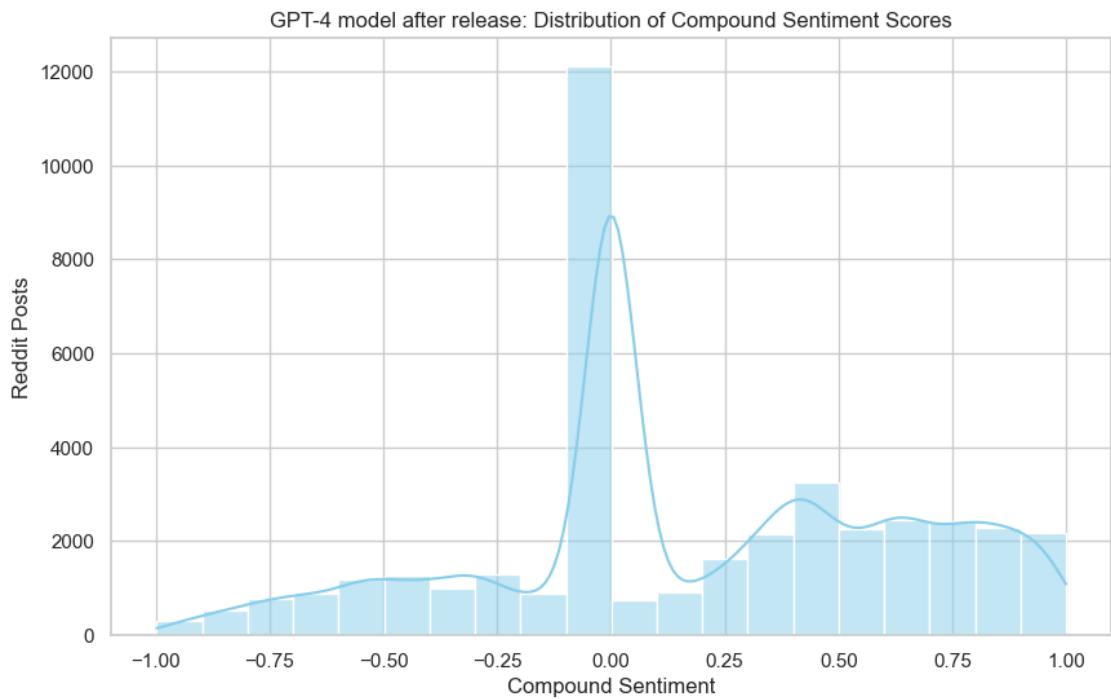


GPT-4 model before release: Distribution of Compound Sentiment Scores

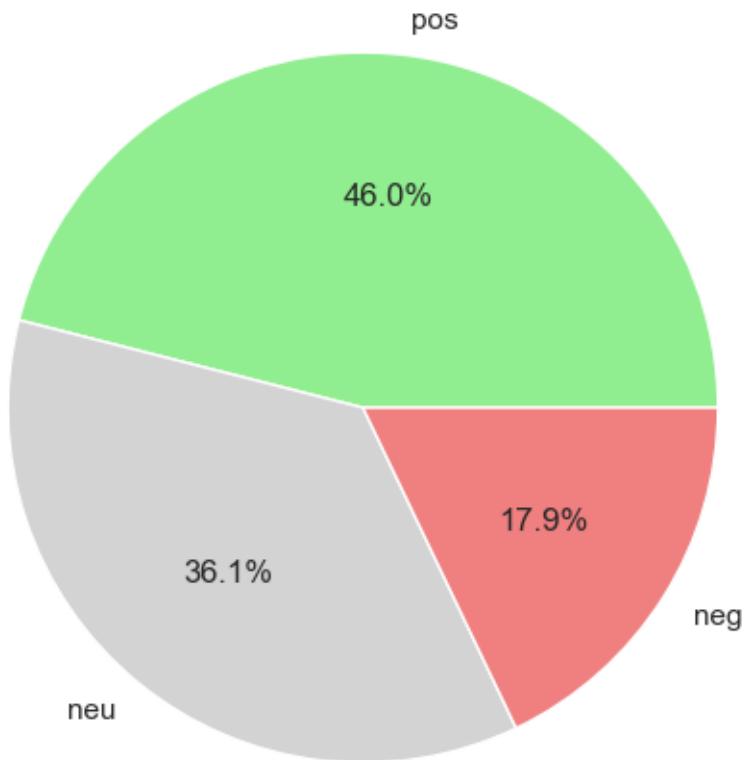


GPT-4 model before release: Sentiment Distribution



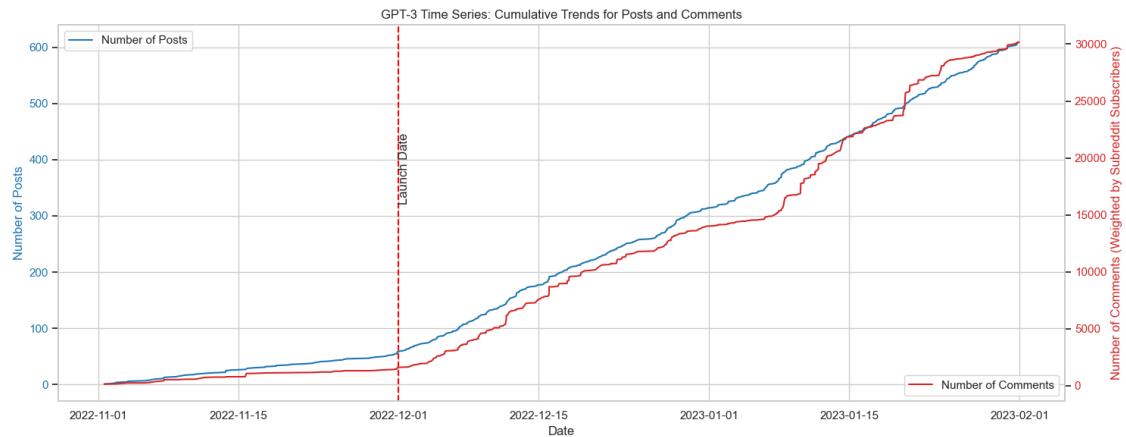


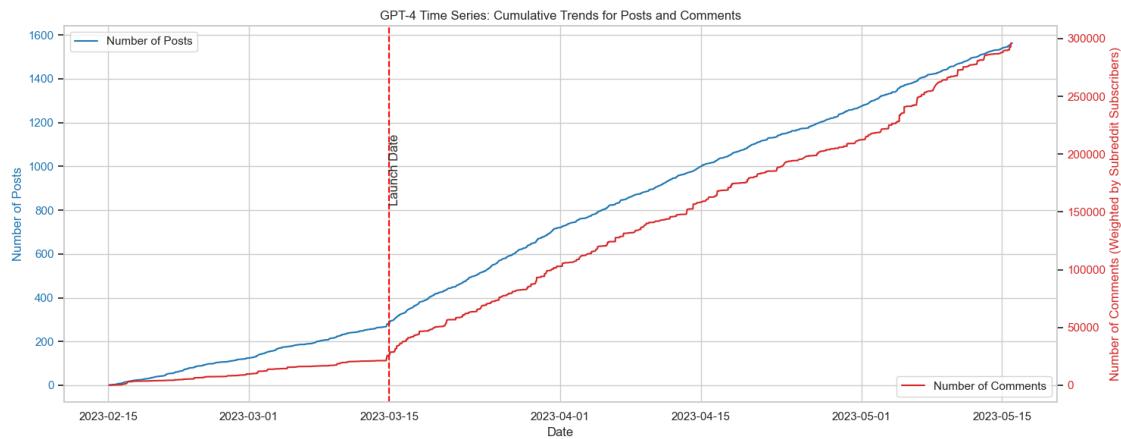
GPT-4 model after release: Sentiment Distribution



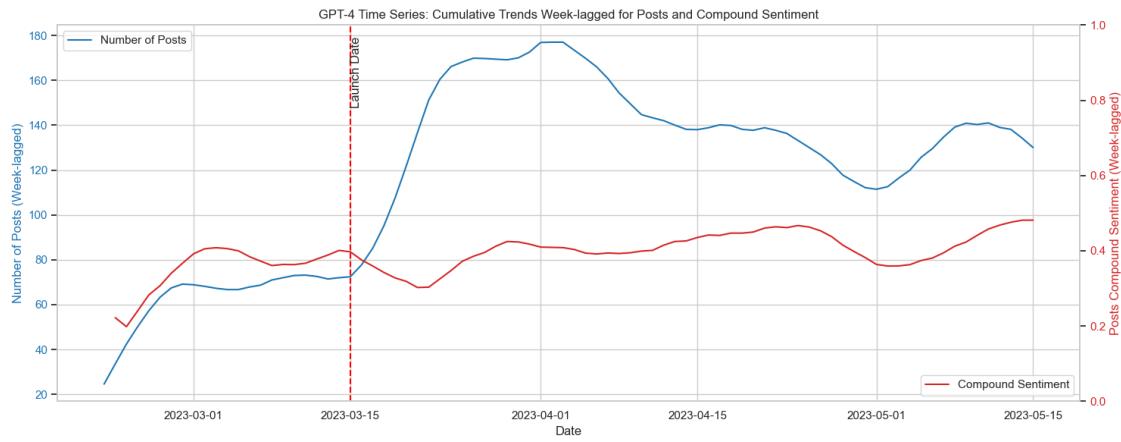
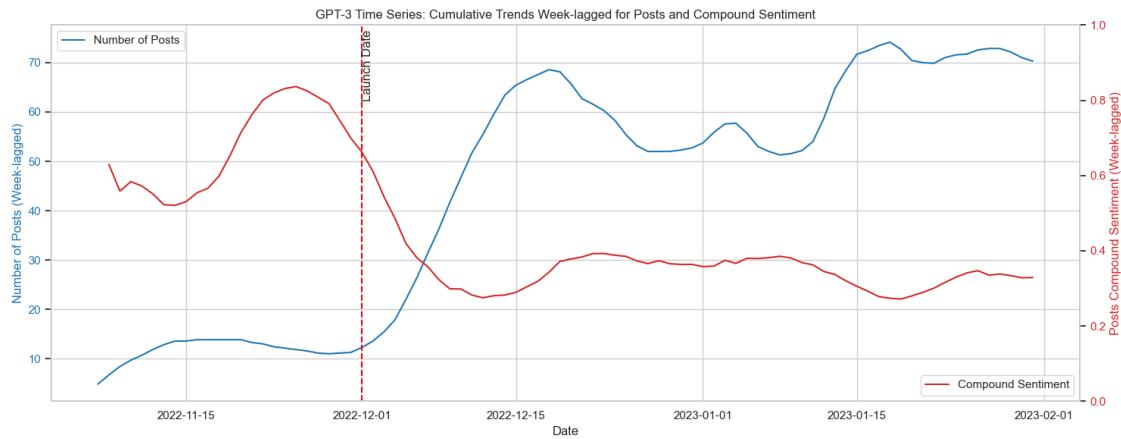
4 Time Series Analysis

4.1 Posts and Comments

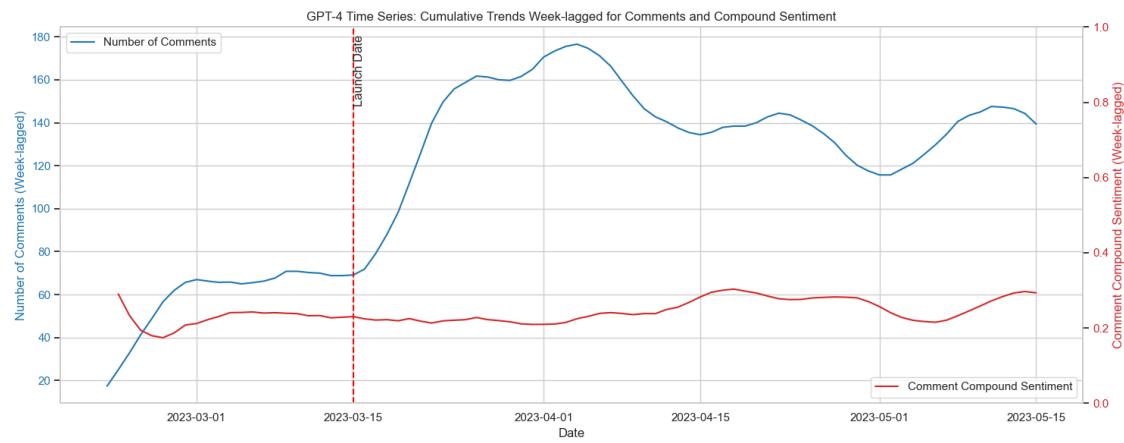
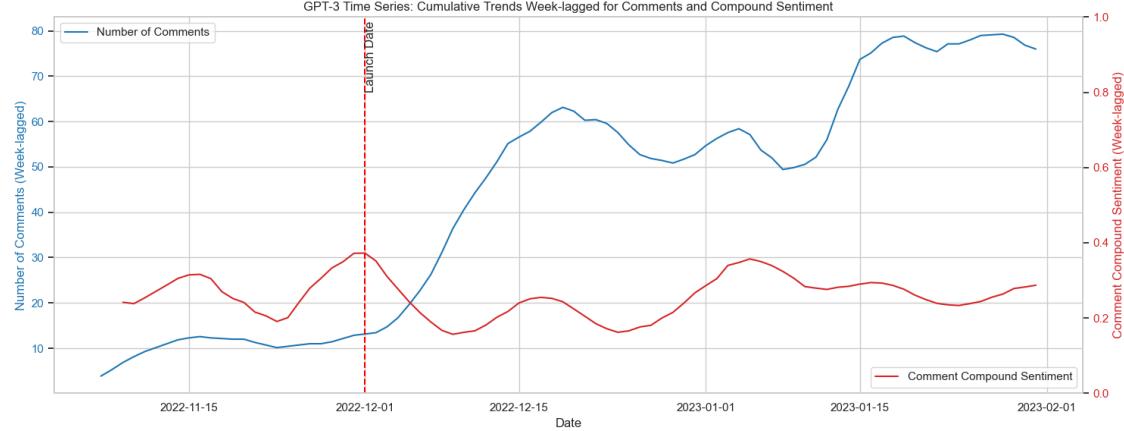




4.2 Posts and Sentiment



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
index	45.00	3415.87	1626.45	401.00
index_x	45.00	1495.56	598.34	134.00
stickied	45.00	0.00	0.00	0.00
upvote_ratio	45.00	0.84	0.17	0.43
score	45.00	67.27	120.91	0.00
num_comments	45.00	22.13	36.04	1.00
created	1970-01-01	2022-11-17	1970-01-08	2022-11-06
text_neg	45.00	0.04	0.07	0.00
text_neu	45.00	0.83	0.11	0.41
text_pos	45.00	0.13	0.10	0.00
text_compound	45.00	0.62	0.39	-0.50
score_weighted	45.00	79.72	147.41	0.00
num_comments_- weighted	45.00	25.91	46.18	1.01
index_y	45.00	35318.09	12642.02	1751.00
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	0.00
comment_body_- compound	45.00	0.31	0.40	-0.55
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	45.00	10.58	2.38	3.57
cumulative_lagged				
comment_compound_- cumulative	45.00	0.28	0.07	0.11
comment_compound_-	45.00	0.27	0.06	0.18

cumulative_lagged						
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		
<hr/>						
index	25%	50%	75%	max		
index	2085.00	3412.00	4009.00	6773.00		
index_x	1216.00	1483.00	1910.00	2313.00		
stickied	0.00	0.00	0.00	0.00		
upvote_ratio	0.75	0.89	0.97	1.00		
score	3.00	12.00	50.00	473.00		
num_comments	5.00	11.00	23.00	216.00		
created	2022-11-10	2022-11-16	2022-11-23	2022-11-29		
text_neg	0.00	0.02	0.05	0.45		
text_neu	0.81	0.84	0.91	1.00		
text_pos	0.07	0.12	0.18	0.59		
text_compound	0.50	0.67	0.93	1.00		
score_weighted	3.19	12.75	53.44	607.43		
num_comments_weighted	5.31	11.73	29.74	281.56		
index_y	34796.00	37848.00	39646.00	50346.00		
comment_score	1.00	2.00	4.00	30.00		
comment_num_replies	0.00	1.00	1.00	2.00		
comment_body_neg	0.00	0.00	0.03	0.32		
comment_body_neu	0.71	0.85	0.97	1.00		
comment_body_pos	0.00	0.12	0.19	1.00		
comment_body_compound	0.00	0.34	0.64	0.98		
comment_score_weighted	1.06	2.12	4.40	39.11		
comment_num_replies_	0.00	1.06	1.10	2.61		

	weighted					
launch_distance_f	6.71	13.43	19.15	23.79		
comment_launch_	5.14	11.14	18.09	43.50		
distance_f						
count_cumulative	12.00	13.00	14.00	16.00		
num_comments_	584.76	1077.21	1195.00	1407.37		
cumulative						
count_cumulative_	12.00	13.14	14.14	14.71		
lagged						
compound_cumulative	0.55	0.61	0.77	0.88		
compound_cumulative_	0.55	0.61	0.77	0.85		
lagged						
comment_count_	10.00	12.00	13.00	14.00		
cumulative						
comment_count_	10.29	11.43	12.14	13.14		
cumulative_lagged						
comment_compound_	0.23	0.28	0.34	0.39		
cumulative						
comment_compound_	0.22	0.27	0.32	0.38		
cumulative_lagged						
log_launch_distance_	2.04	2.67	3.00	3.21		
f						
log_num_comments_	1.84	2.54	3.43	5.64		
weighted						
log_score_weighted	1.43	2.62	4.00	6.41		
log_count_cumulative	2.56	2.64	2.71	2.83		
log_count_	2.56	2.65	2.72	2.75		
cumulative_lagged						
log_comment_count_	2.40	2.56	2.64	2.71		
cumulative						
log_comment_count_	2.42	2.52	2.58	2.65		
cumulative_lagged						
log_comment_score_	0.72	1.14	1.69	3.69		
weighted						
log_comment_num_	0.00	0.72	0.74	1.28		
replies_weighted						
log_comment_launch_	1.82	2.50	2.95	3.80		
distance_f						

GPT-3 model after release

	index	count	mean	std	min	
index	543.00	3149.07	1920.01	44.00		
index_x	543.00	1382.53	790.25	6.00		
stickied	543.00	0.00	0.00	0.00		
upvote_ratio	543.00	0.82	0.20	0.12		
score	543.00	200.38	1355.50	0.00		

	num_comments	543.00	42.71	81.46	1.00	
	created	1970-01-01	2023-01-02	1970-01-19	2022-11-30	
	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	0.00	
	text_compound	543.00	0.33	0.52	-1.00	
	score_weighted	543.00	268.21	2127.43	0.00	
	num_comments_	543.00	50.94	112.54	1.01	
	weighted					
	index_y	543.00	34920.68	15951.39	217.00	
	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	0.00	
	comment_body_	543.00	0.26	0.46	-0.95	
	compound					
	comment_score_	543.00	7.90	22.34	0.00	
	weighted					
	comment_num_replies_	543.00	0.82	1.35	0.00	
	weighted					
	launch_distance_f	543.00	33.08	18.22	0.47	
	comment_launch_	543.00	34.77	17.91	0.56	
	distance_f					
	count_cumulative	543.00	61.31	13.47	12.00	
	num_comments_	543.00	15560.33	8752.83	1411.62	
	cumulative					
	count_cumulative_	543.00	61.03	13.84	12.00	
	lagged					
	compound_cumulative	543.00	0.34	0.05	0.23	
	compound_cumulative_	543.00	0.34	0.05	0.23	
	lagged					
	comment_count_	543.00	60.10	16.54	12.00	
	cumulative					
	comment_count_	543.00	59.82	16.79	12.86	
	cumulative_lagged					
	comment_compound_	543.00	0.25	0.06	0.12	
	cumulative					
	comment_compound_	543.00	0.25	0.06	0.13	
	cumulative_lagged					
	log_launch_distance_	543.00	3.30	0.80	0.38	
	f					
	log_num_comments_	543.00	2.98	1.35	0.70	
	weighted					
	log_score_weighted	543.00	3.37	2.07	0.00	
	log_count_cumulative	543.00	4.10	0.29	2.56	
	log_count_	543.00	4.09	0.31	2.56	
	cumulative_lagged					

log_comment_count_cumulative	543.00	4.06	0.35	2.56
log_comment_count_cumulative_lagged	543.00	4.05	0.36	2.63
log_comment_score_weighted	543.00	1.45	0.97	0.00
log_comment_num_replies_weighted	543.00	0.44	0.52	0.00
log_comment_launch_distance_f	543.00	3.38	0.73	0.44
index	25%	50%	75%	max
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
comment_launch_distance_f	18.25	37.88	50.60	62.96
count_cumulative	55.00	65.00	71.00	84.00
num_comments_cumulative	8678.46	14418.74	22975.02	30173.20
count_cumulative_lagged	54.57	65.14	70.50	82.71

compound_cumulative	0.31	0.34	0.37	0.52
compound_cumulative_lagged	0.31	0.34	0.37	0.58
comment_count_cumulative	51.00	60.00	74.00	87.00
comment_count_cumulative_lagged	50.29	59.86	74.79	85.43
comment_compound_cumulative	0.21	0.26	0.29	0.43
comment_compound_cumulative_lagged	0.21	0.26	0.28	0.42
log_launch_distance_f	2.84	3.59	3.91	4.16
log_num_comments_weighted	1.96	2.95	3.87	7.23
log_score_weighted	1.84	3.52	4.88	10.64
log_count_cumulative	4.03	4.19	4.28	4.44
log_count_cumulative_lagged	4.02	4.19	4.27	4.43
log_comment_count_cumulative	3.95	4.11	4.32	4.48
log_comment_count_cumulative_lagged	3.94	4.11	4.33	4.46
log_comment_score_weighted	0.72	1.14	1.84	5.80
log_comment_num_replies_weighted	0.00	0.00	0.74	2.65
log_comment_launch_distance_f	2.96	3.66	3.94	4.16

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_-		248.00		0.23		0.44		-0.97	
	compound									
	comment_score_-		248.00		15.86		99.40		0.00	
	weighted									
	comment_num_replies_-		248.00		1.01		2.53		0.00	
	weighted									
	launch_distance_f		248.00		12.72		7.37		0.22	
	comment_launch_-		248.00		12.75		8.41		0.08	
	distance_f									
	count_cumulative		248.00		62.80		16.47		8.00	
	num_comments_-		248.00		11289.03		6152.70		482.96	
	cumulative									
	count_cumulative_-		248.00		62.02		17.45		5.00	
	lagged									
	compound_cumulative		248.00		0.34		0.10		0.03	
	compound_cumulative_-		248.00		0.34		0.09		0.06	
	lagged									
	comment_count_-		248.00		57.72		18.64		5.00	
	cumulative									
	comment_count_-		248.00		56.96		19.40		3.00	
	cumulative_lagged									
	comment_compound_-		248.00		0.23		0.07		0.12	
	cumulative									
	comment_compound_-		248.00		0.24		0.08		0.12	
	cumulative_lagged									
	log_launch_distance_-		248.00		2.41		0.74		0.20	
	f									
	log_num_comments_-		248.00		3.12		1.52		0.70	
	weighted									
	log_score_weighted		248.00		3.60		2.24		0.00	
	log_count_cumulative		248.00		4.10		0.40		2.20	
	log_count_-		248.00		4.07		0.46		1.79	
	cumulative_lagged									
	log_comment_count_-		248.00		3.97		0.56		1.79	
	cumulative									
	log_comment_count_-		248.00		3.94		0.61		1.39	
	cumulative_lagged									
	log_comment_score_-		248.00		1.47		1.06		0.00	
	weighted									
	log_comment_num_replies_weighted		248.00		0.45		0.58		0.00	
	log_comment_launch_-		248.00		2.38		0.79		0.08	

distance_f						
index	25%	50%	75%	max		
index	30284.00	32801.50	34846.25	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_	6.07	19.67	53.86	1769.80		
weighted						
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_-	0.00	0.20	0.60	0.97		
compound						
comment_score_-	1.06	2.06	5.31	1129.96		
weighted						
comment_num_replies_-	0.00	0.00	1.06	25.22		
weighted						
launch_distance_f	6.27	12.11	19.33	25.92		
comment_launch_	5.95	11.84	19.00	62.02		
distance_f						
count_cumulative	62.00	69.50	72.25	79.00		
num_comments_-	5229.59	11094.70	16617.05	21296.85		
cumulative						
count_cumulative_-	61.18	69.64	72.46	78.29		
lagged						
compound_cumulative	0.29	0.38	0.40	0.47		
compound_cumulative_-	0.29	0.38	0.40	0.45		
lagged						
comment_count_-	55.00	66.00	69.00	77.00		
cumulative						
comment_count_-	55.00	65.79	69.18	76.14		
cumulative_lagged						
comment_compound_-	0.20	0.23	0.24	0.58		
cumulative						

comment_compound_	0.20	0.23	0.24	0.58
cumulative_lagged				
log_launch_distance_	1.98	2.57	3.01	3.29
f				
log_num_comments_	1.96	3.03	4.00	7.48
weighted				
log_score_weighted	1.84	3.69	5.03	9.48
log_count_cumulative	4.14	4.26	4.29	4.38
log_count_	4.13	4.26	4.30	4.37
cumulative_lagged				
log_comment_count_	4.03	4.20	4.25	4.36
cumulative				
log_comment_count_	4.03	4.20	4.25	4.35
cumulative_lagged				
log_comment_score_	0.72	1.12	1.84	7.03
weighted				
log_comment_num_	0.00	0.00	0.72	3.27
replies_weighted				
log_comment_launch_	1.94	2.55	3.00	4.14
distance_f				

GPT-4 model after release

index	count	mean	std	min
index	1271.00	28391.08	12259.98	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_	1271.00	207.58	555.69	1.01
weighted				
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_	1271.00	0.25	0.46	-0.98
compound				
comment_score_	1271.00	11.82	94.46	0.00

	weighted					
comment_num_replies_	1271.00	0.88	1.91	0.00		
weighted						
launch_distance_f	1271.00	29.22	18.51	0.31		
comment_launch_	1271.00	30.35	18.53	0.47		
distance_f						
count_cumulative	1271.00	143.11	23.94	72.00		
num_comments_	1271.00	146463.43	76247.82	21303.96		
cumulative						
count_cumulative_	1271.00	142.99	24.10	72.86		
lagged						
compound_cumulative	1271.00	0.40	0.06	0.25		
compound_cumulative_	1271.00	0.40	0.06	0.25		
lagged						
comment_count_	1271.00	141.74	22.78	68.00		
cumulative						
comment_count_	1271.00	141.58	22.98	68.71		
cumulative_lagged						
comment_compound_	1271.00	0.25	0.04	0.18		
cumulative						
comment_compound_	1271.00	0.25	0.04	0.18		
cumulative_lagged						
log_launch_distance_	1271.00	3.12	0.89	0.27		
f						
log_num_comments_	1271.00	3.53	1.86	0.70		
weighted						
log_score_weighted	1271.00	3.95	2.73	0.00		
log_count_cumulative	1271.00	4.96	0.17	4.29		
log_count_	1271.00	4.95	0.18	4.30		
cumulative_lagged						
log_comment_count_	1271.00	4.95	0.18	4.23		
cumulative						
log_comment_count_	1271.00	4.95	0.18	4.24		
cumulative_lagged						
log_comment_score_	1271.00	1.40	0.99	0.00		
weighted						
log_comment_num_	1271.00	0.44	0.55	0.00		
replies_weighted						
log_comment_launch_	1271.00	3.17	0.86	0.39		
distance_f						
<hr/>						
	index	25%	50%	75%	max	
<hr/>						
	index	25517.00	31847.00	35093.00	45320.00	
	index_x	505.50	1019.00	1910.50	8881.00	
	stickied	0.00	0.00	0.00	0.00	
	upvote_ratio	0.75	0.90	0.96	1.00	

score	5.00	37.00	294.50	43183.00	
num_comments	6.00	23.00	98.00	3834.00	
created	2023-03-26	2023-04-10	2023-04-28	2023-05-15	
text_neg	0.00	0.01	0.05	1.00	
text_neu	0.80	0.86	0.93	1.00	
text_pos	0.00	0.09	0.15	1.00	
text_compound	0.00	0.51	0.93	1.00	
score_weighted	5.05	39.36	343.52	68054.49	
num_comments_	7.07	25.21	118.29	6042.21	
weighted					
index_y	27006.50	34244.00	39681.50	197410.00	
comment_score	1.00	2.00	4.00	1538.00	
comment_num_replies	0.00	0.00	1.00	37.00	
comment_body_neg	0.00	0.00	0.06	1.00	
comment_body_neu	0.73	0.84	0.98	1.00	
comment_body_pos	0.00	0.09	0.20	1.00	
comment_body_	0.00	0.26	0.66	1.00	
compound					
comment_score_	1.08	2.03	4.25	2423.82	
weighted					
comment_num_replies_	0.00	0.00	1.10	40.71	
weighted					
launch_distance_f	12.88	27.87	45.73	62.99	
comment_launch_	14.30	29.28	47.01	63.00	
distance_f					
count_cumulative	131.00	140.00	163.00	196.00	
num_comments_	79459.44	142064.01	205807.96	296096.98	
cumulative					
count_cumulative_	130.36	140.00	162.86	193.86	
lagged					
compound_cumulative	0.37	0.41	0.45	0.54	
compound_cumulative_	0.37	0.41	0.45	0.53	
lagged					
comment_count_	131.00	143.00	157.00	194.00	
cumulative					
comment_count_	131.29	143.14	156.93	192.71	
cumulative_lagged					
comment_compound_	0.22	0.24	0.28	0.33	
cumulative					
comment_compound_	0.22	0.24	0.27	0.33	
cumulative_lagged					
log_launch_distance_	2.63	3.36	3.84	4.16	
f					
log_num_comments_	2.09	3.27	4.78	8.71	
weighted					
log_score_weighted	1.80	3.70	5.84	11.13	
log_count_cumulative	4.88	4.95	5.10	5.28	
log_count_	4.88	4.95	5.10	5.27	

cumulative_lagged						
log_comment_count_	4.88	4.97	5.06	5.27		
cumulative						
log_comment_count_	4.88	4.97	5.06	5.27		
cumulative_lagged						
log_comment_score_	0.73	1.11	1.66	7.79		
weighted						
log_comment_num_	0.00	0.00	0.74	3.73		
replies_weighted						
log_comment_launch_	2.73	3.41	3.87	4.16		
distance_f						

6 Run MLflow Regression experiment

Test model against data after GPT-4 launch.

Coefficients of GPT-3 Regression

Feature	Coefficient
launched	0.44
upvote_ratio	0.00
log_launch_distance_f	2.59
log_num_comments_weighted	0.04
log_score_weighted	-1.26
log_comment_score_weighted	6.05
log_comment_launch_distance_f	-0.15
log_comment_num_replies_weighted	0.02
log_count_cumulative	-0.02
log_count_cumulative_lagged	0.78
log_comment_count_cumulative	2.50
log_comment_count_cumulative_lagged	0.20
compound_cumulative	-3.15
compound_cumulative_lagged	0.01
comment_compound_cumulative	0.02
comment_compound_cumulative_lagged	-0.03

Coefficients of GPT-4 Regression

Feature	Coefficient
launched	0.89
upvote_ratio	-0.00
log_launch_distance_f	5.37
log_num_comments_weighted	-6.25
log_score_weighted	1.48
log_comment_score_weighted	-0.86

log_comment_launch_distance_f	-0.04	
log_comment_num_replies_weighted	0.09	
log_count_cumulative	-0.13	
log_count_cumulative_lagged	-1.03	
log_comment_count_cumulative	0.82	
log_comment_count_cumulative_lagged	-1.08	
compound_cumulative	1.22	
compound_cumulative_lagged	-0.02	
comment_compound_cumulative	0.00	
comment_compound_cumulative_lagged	-0.02	

Coefficients of Before-launch Regression

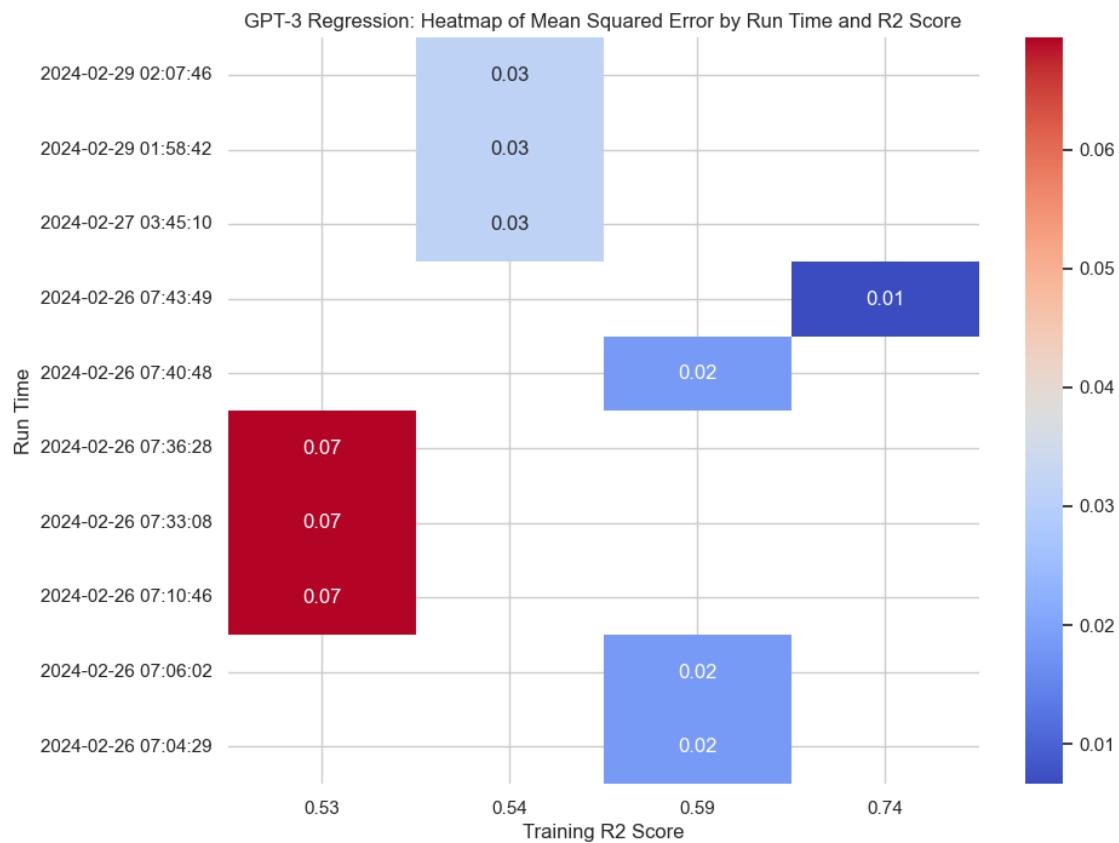
Feature	Coefficient
launched	0.44
upvote_ratio	0.00
log_launch_distance_f	2.59
log_num_comments_weighted	0.04
log_score_weighted	-1.26
log_comment_score_weighted	6.05
log_comment_launch_distance_f	-0.15
log_comment_num_replies_weighted	0.02
log_count_cumulative	-0.02
log_count_cumulative_lagged	0.78
log_comment_count_cumulative	2.50
log_comment_count_cumulative_lagged	0.20
compound_cumulative	-3.15
compound_cumulative_lagged	0.01
comment_compound_cumulative	0.02
comment_compound_cumulative_lagged	-0.03

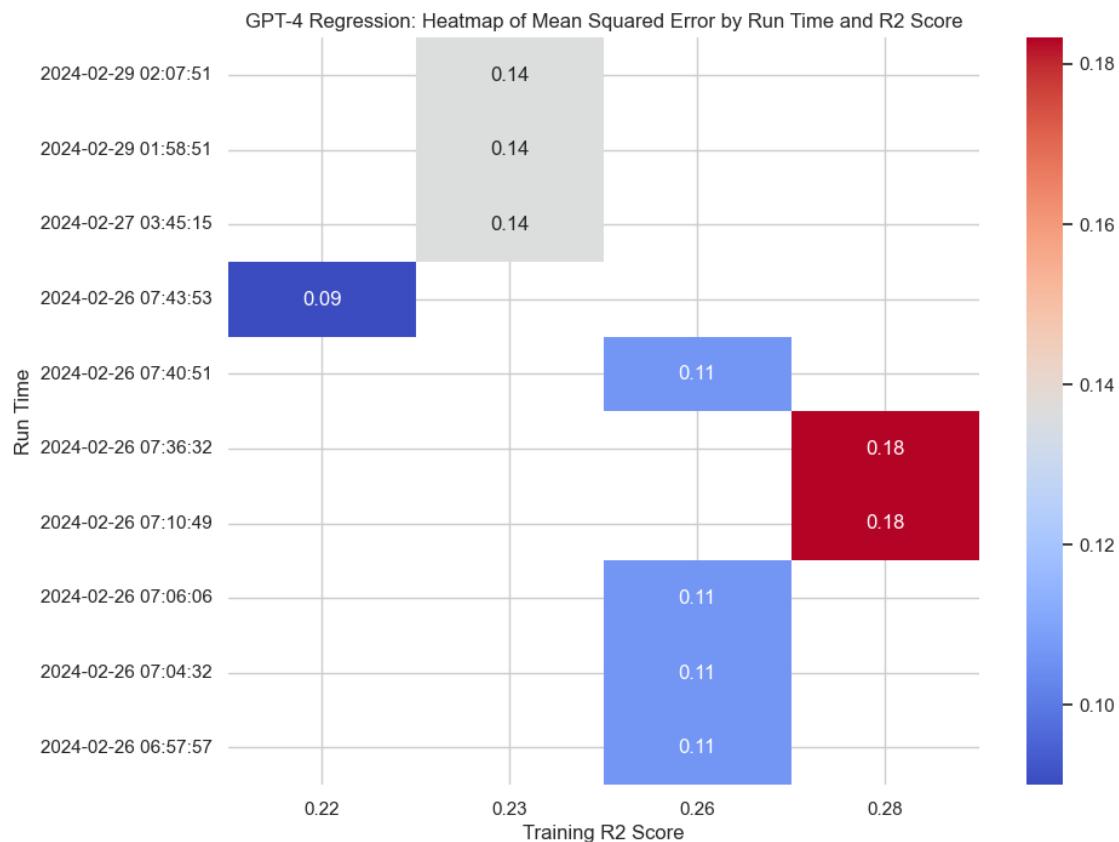
Coefficients of After-launch Regression

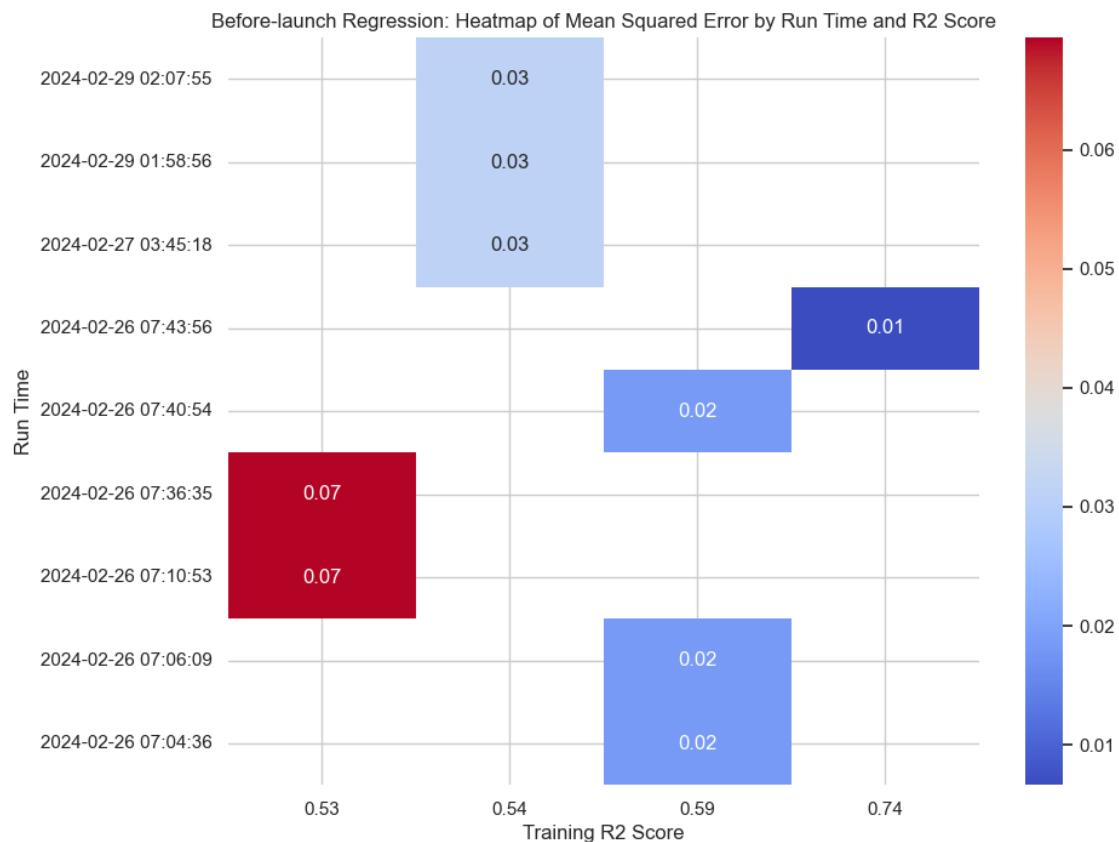
Feature	Coefficient
launched	0.15
upvote_ratio	0.00
log_launch_distance_f	10.76
log_num_comments_weighted	-10.61
log_score_weighted	-1.24
log_comment_score_weighted	0.83
log_comment_launch_distance_f	0.07
log_comment_num_replies_weighted	0.05
log_count_cumulative	-0.06
log_count_cumulative_lagged	-1.08
log_comment_count_cumulative	0.62
log_comment_count_cumulative_lagged	0.27

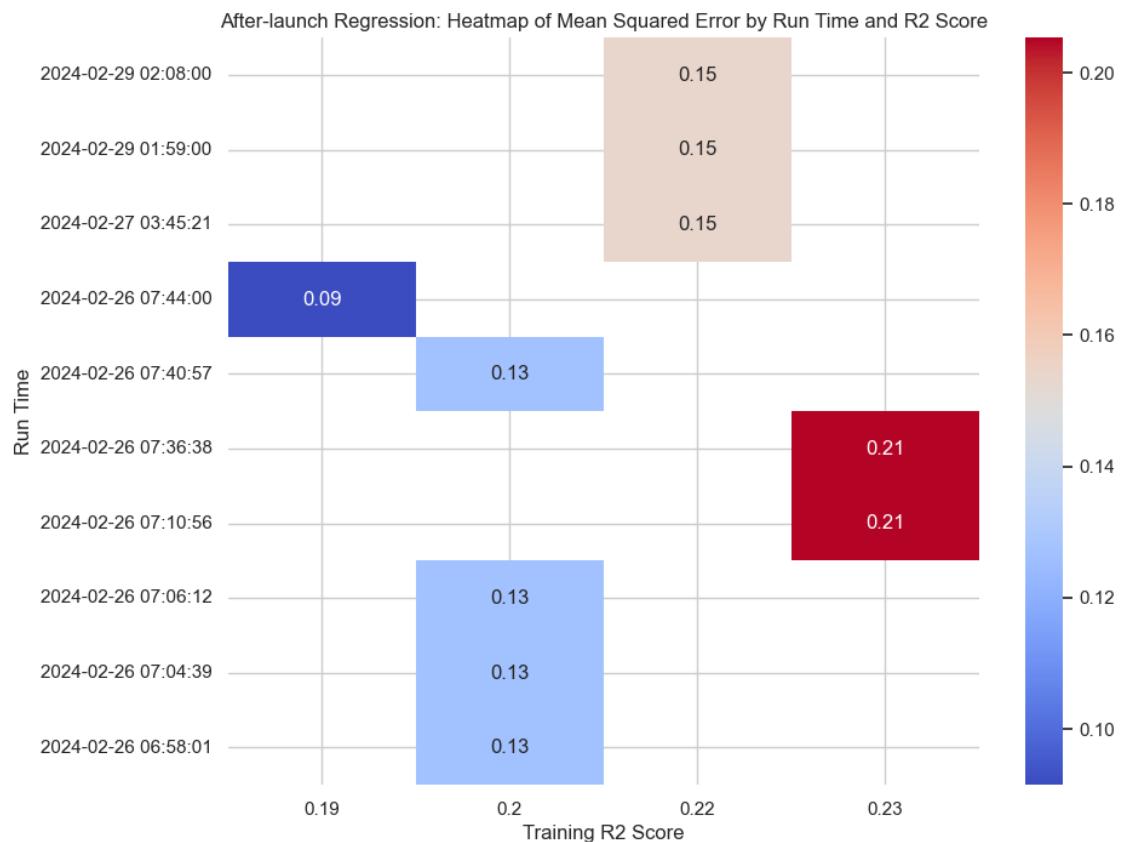
	compound_cumulative		-0.05	
	compound_cumulative_lagged		-0.02	
	comment_compound_cumulative		0.02	
	comment_compound_cumulative_lagged		-0.09	
+-----+				
Coefficients of Comparison Regression				
+-----+-----+				
	Feature		Coefficient	
+-----+-----+				
	launched		0.12	
	upvote_ratio		0.06	
	log_launch_distance_f		6.73	
	log_num_comments_weighted		-6.63	
	log_score_weighted		0.76	
	log_comment_score_weighted		-1.03	
	log_comment_launch_distance_f		0.04	
	log_comment_num_replies_weighted		0.06	
	log_count_cumulative		-0.07	
	log_count_cumulative_lagged		-0.72	
	log_comment_count_cumulative		0.14	
	log_comment_count_cumulative_lagged		0.43	
	compound_cumulative		-0.00	
	compound_cumulative_lagged		-0.02	
	comment_compound_cumulative		0.03	
	comment_compound_cumulative_lagged		-0.09	
+-----+-----+				

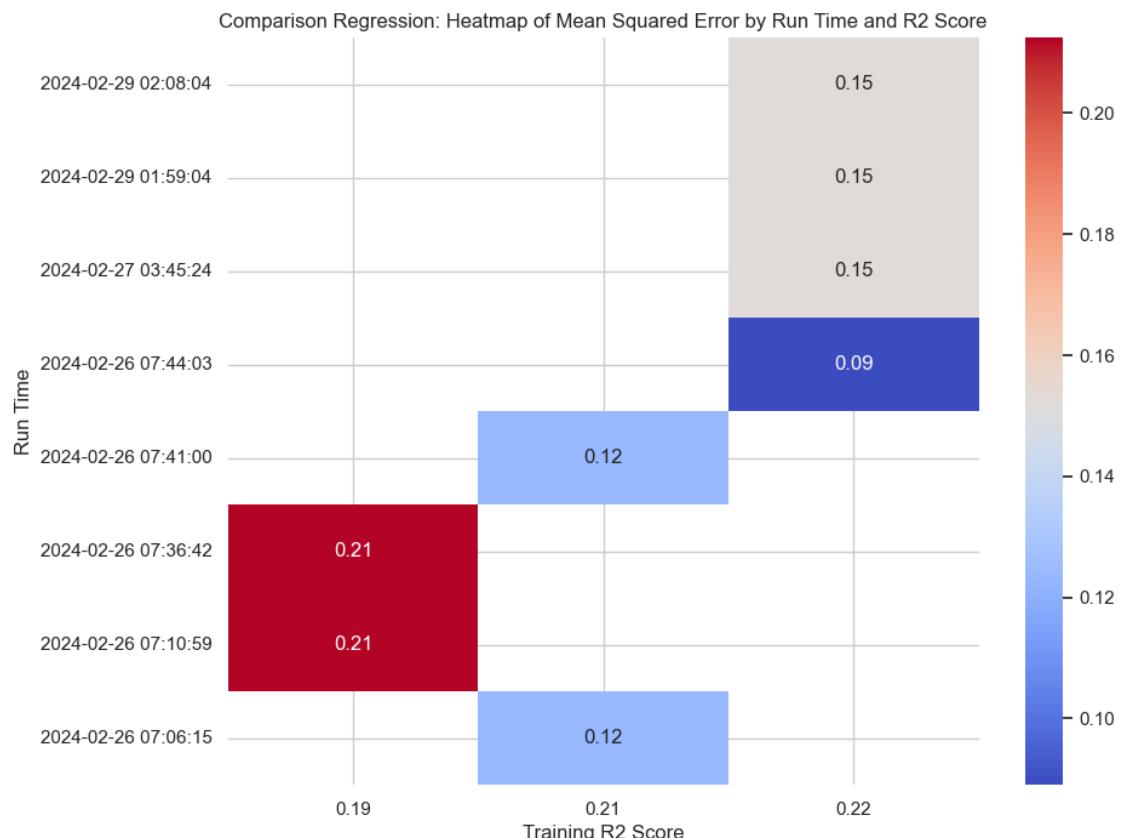
7 Heatmap of MLflow Regression runs







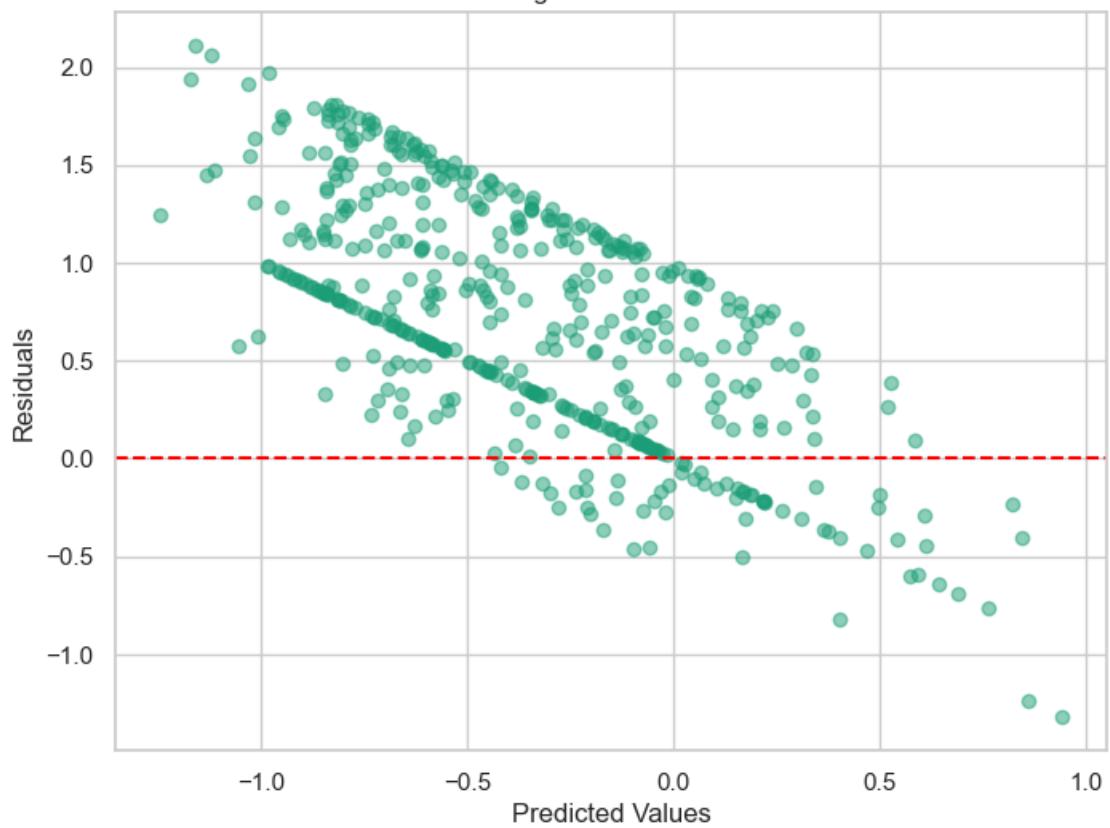




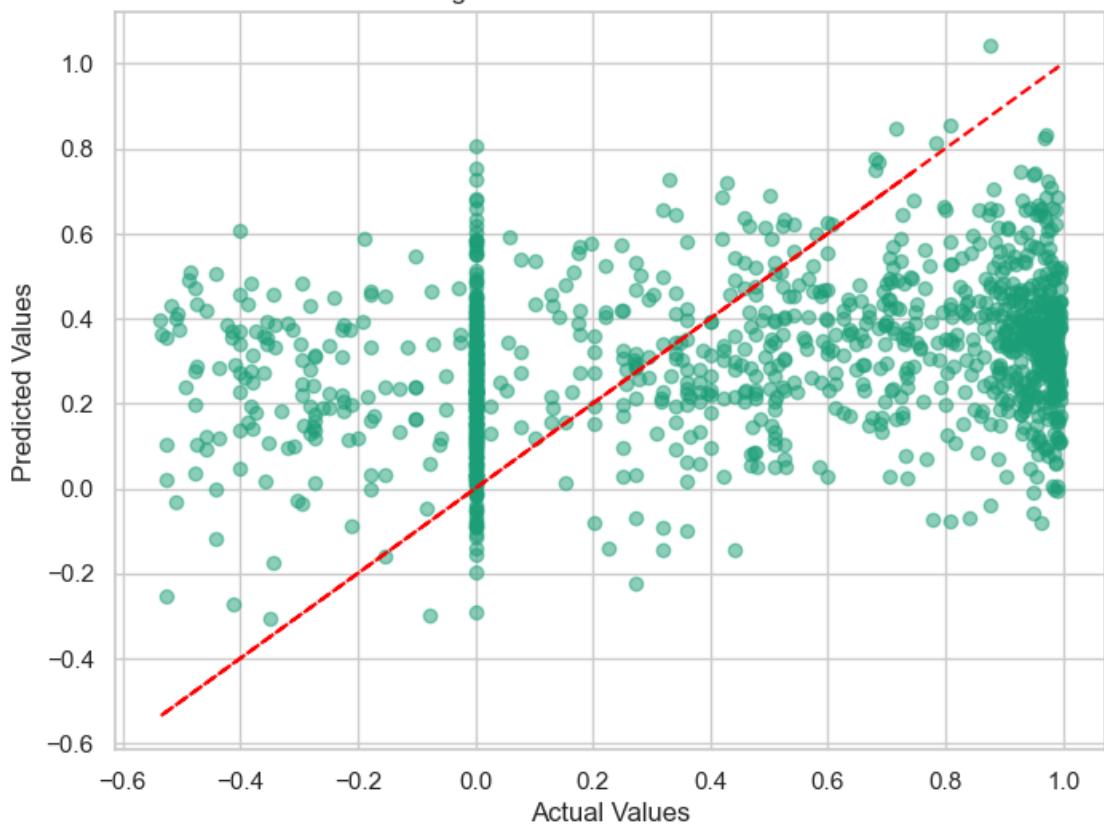
8 Feature Plots for Regression Models



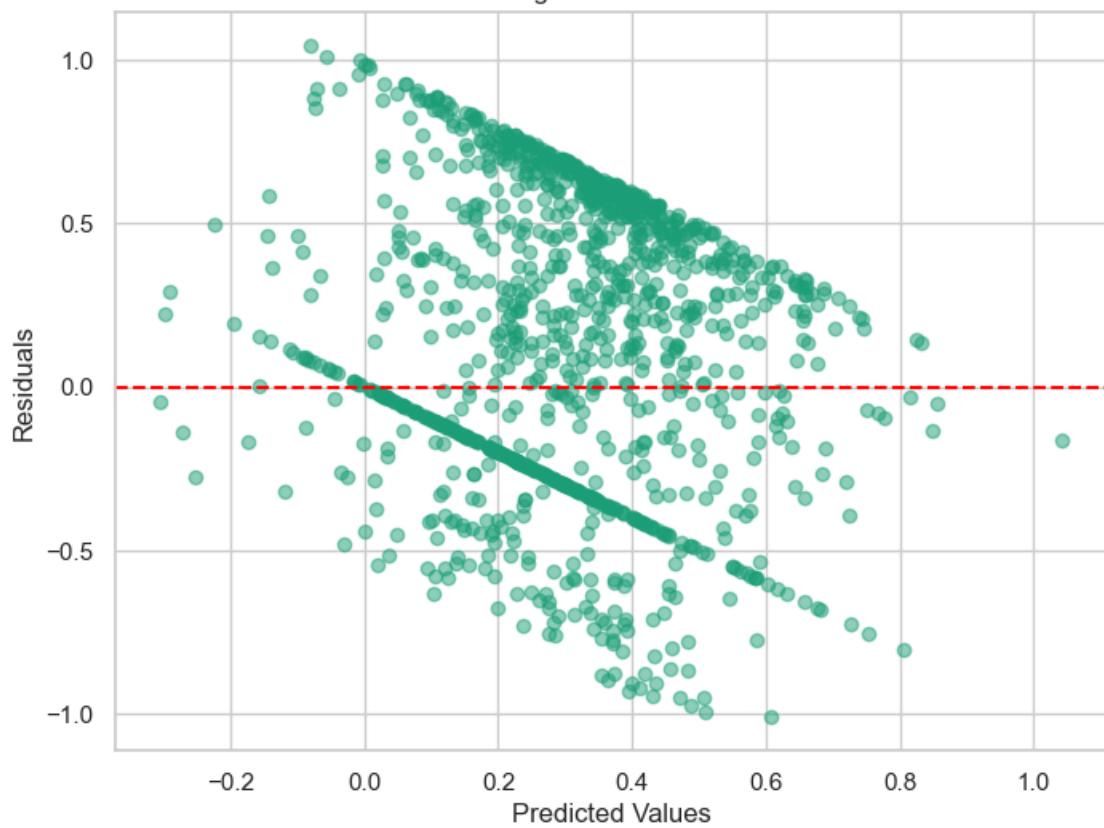
GPT-3 Regression: Residual Plot



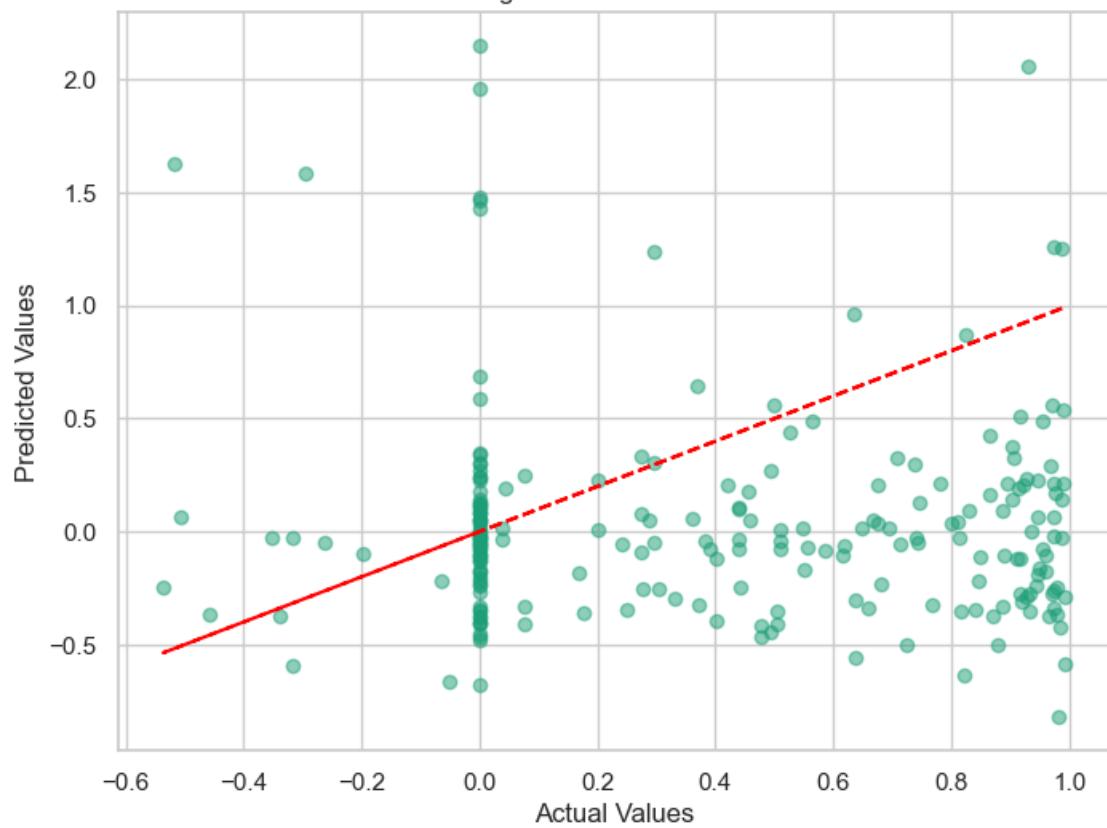
GPT-4 Regression: Actual vs. Predicted Values

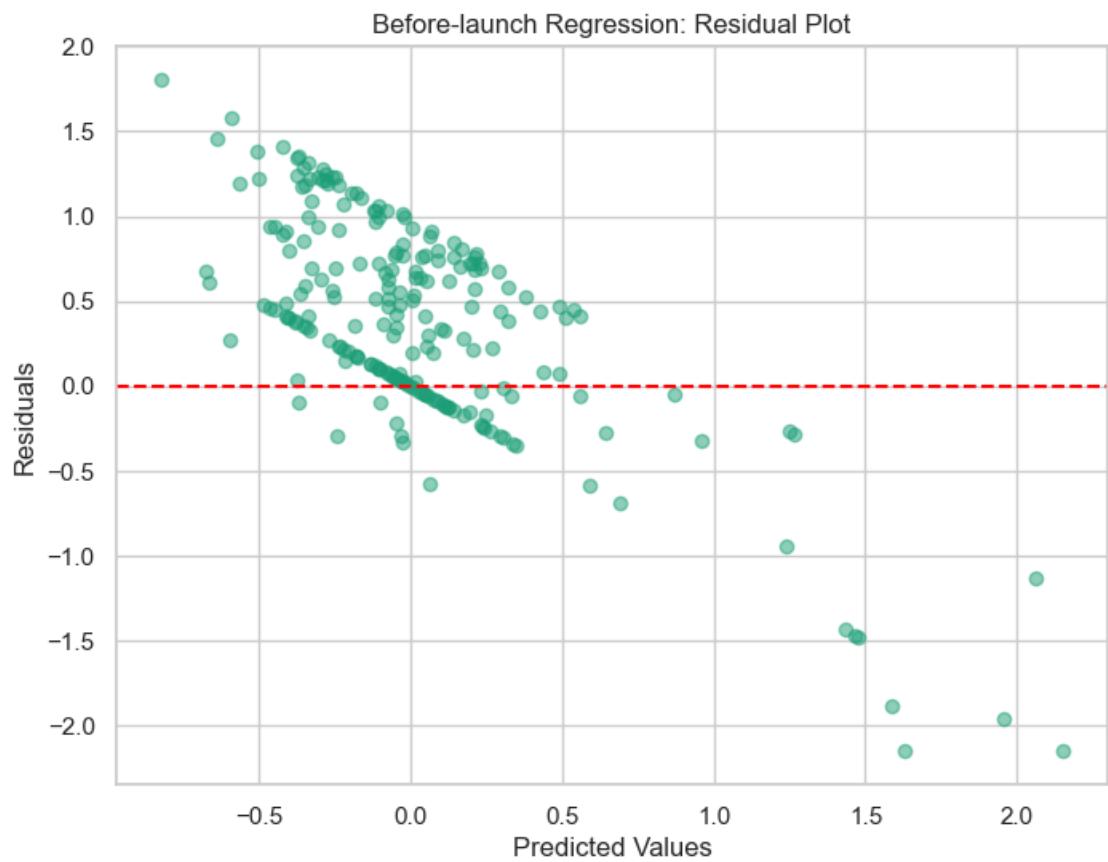


GPT-4 Regression: Residual Plot

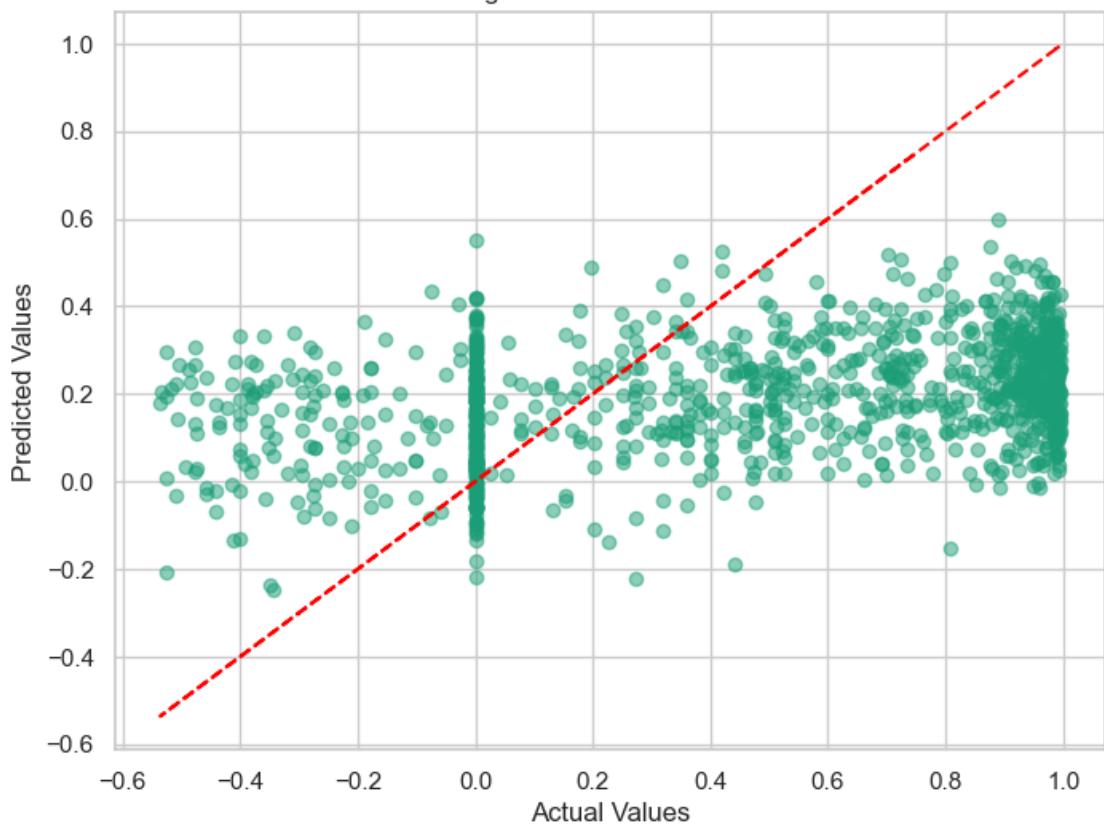


Before-launch Regression: Actual vs. Predicted Values

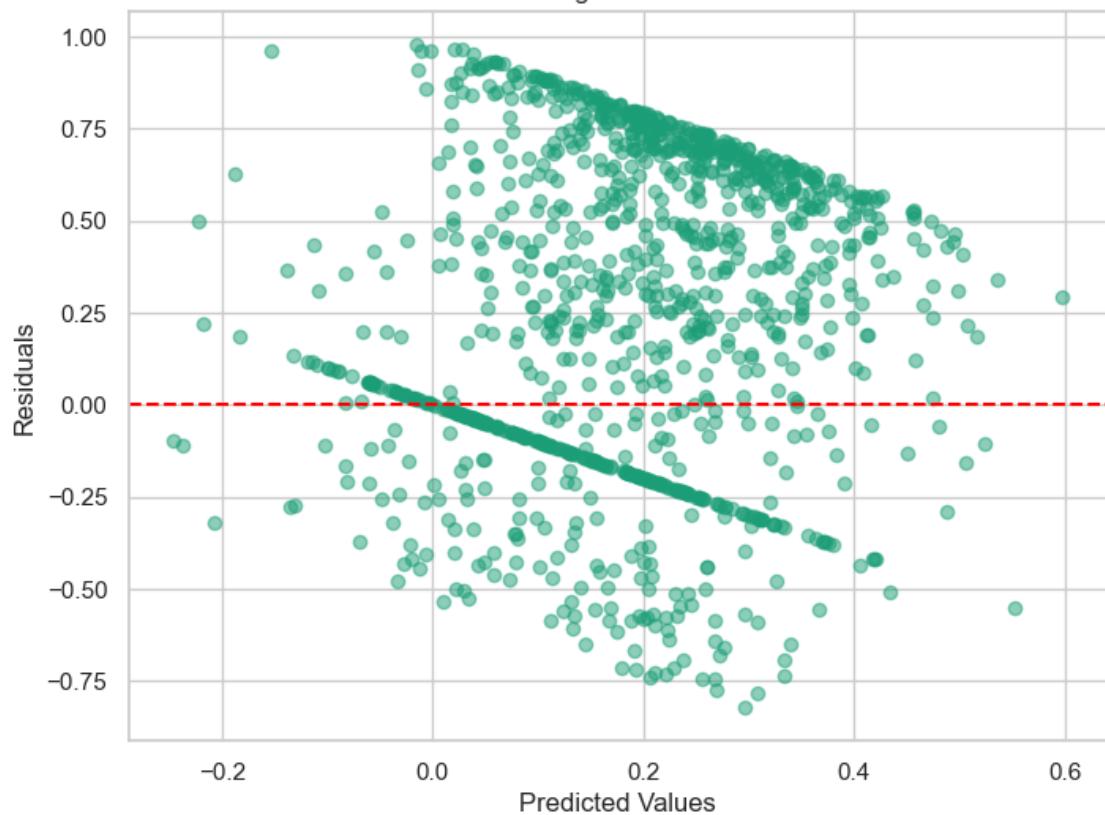


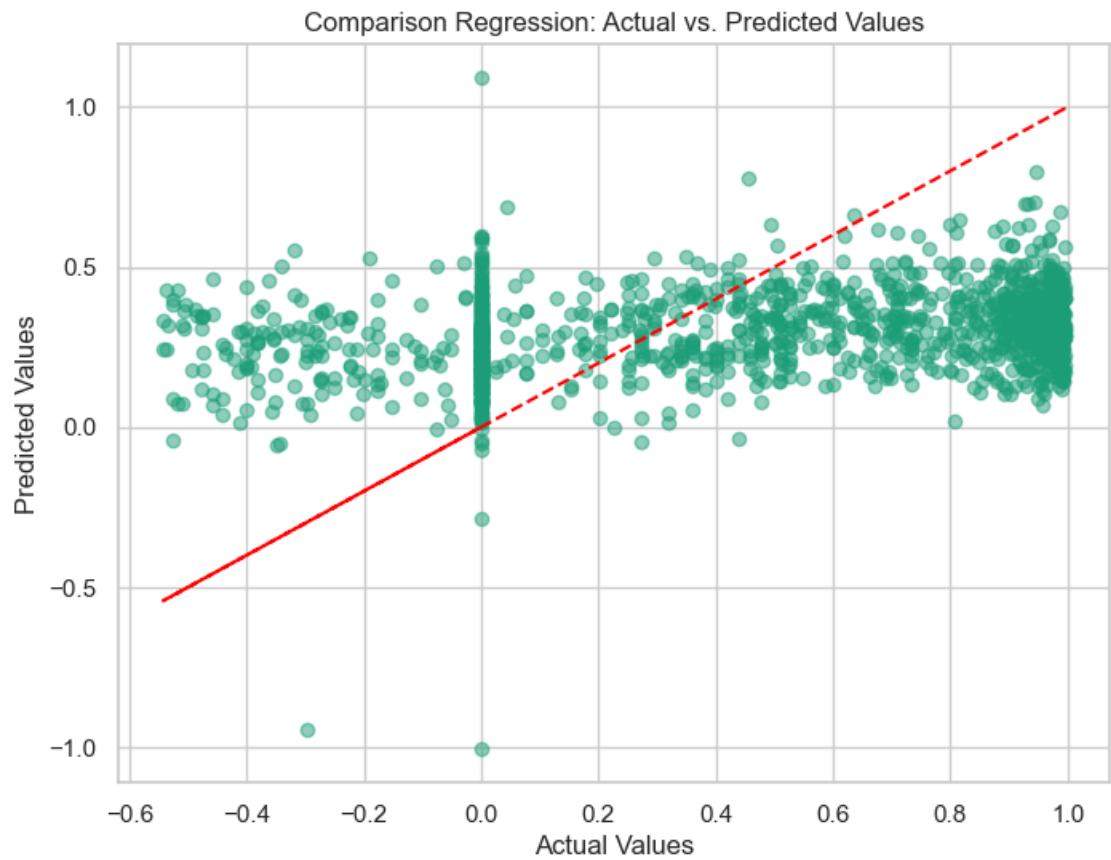


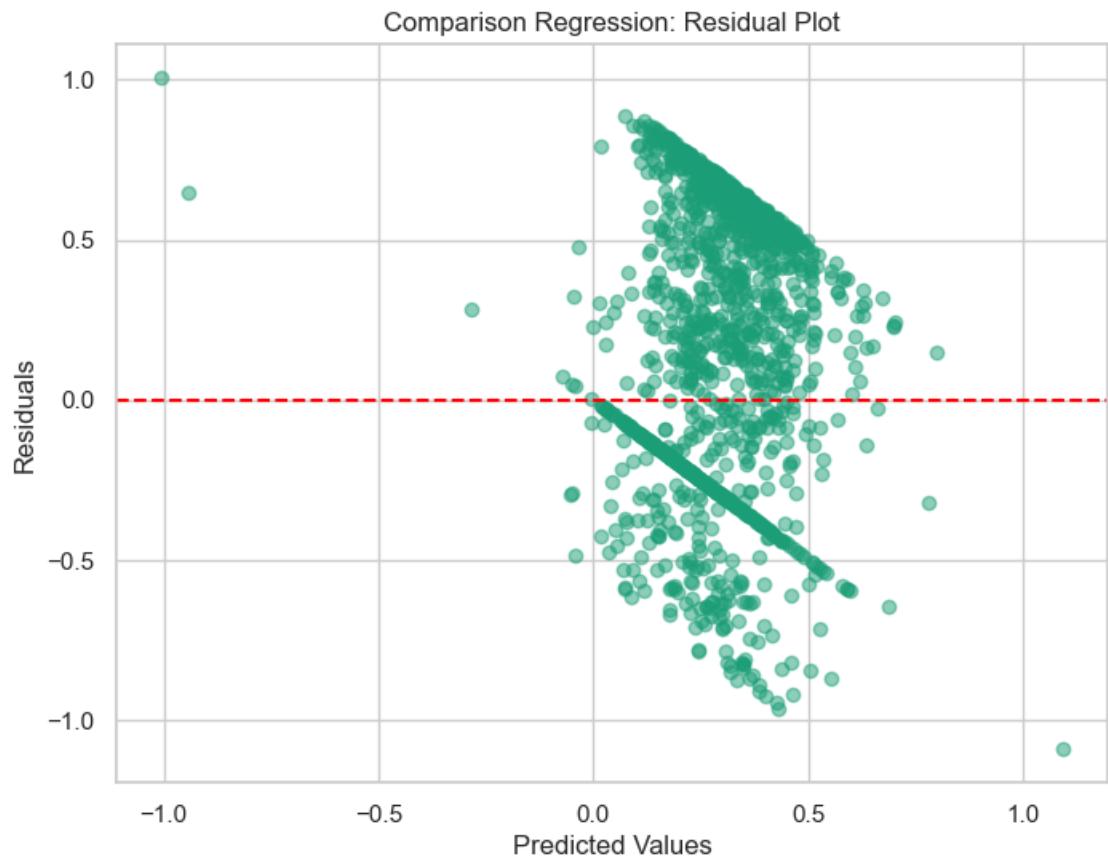
After-launch Regression: Actual vs. Predicted Values

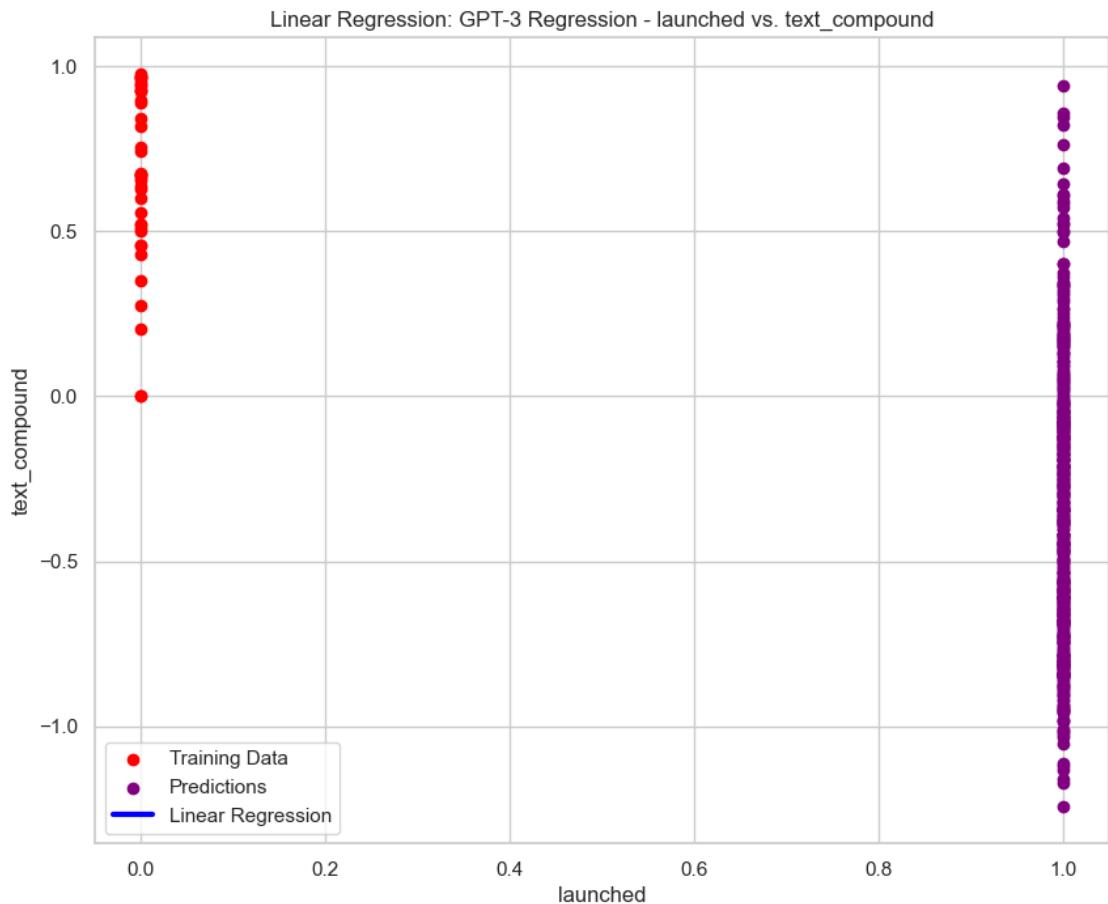


After-launch Regression: Residual Plot

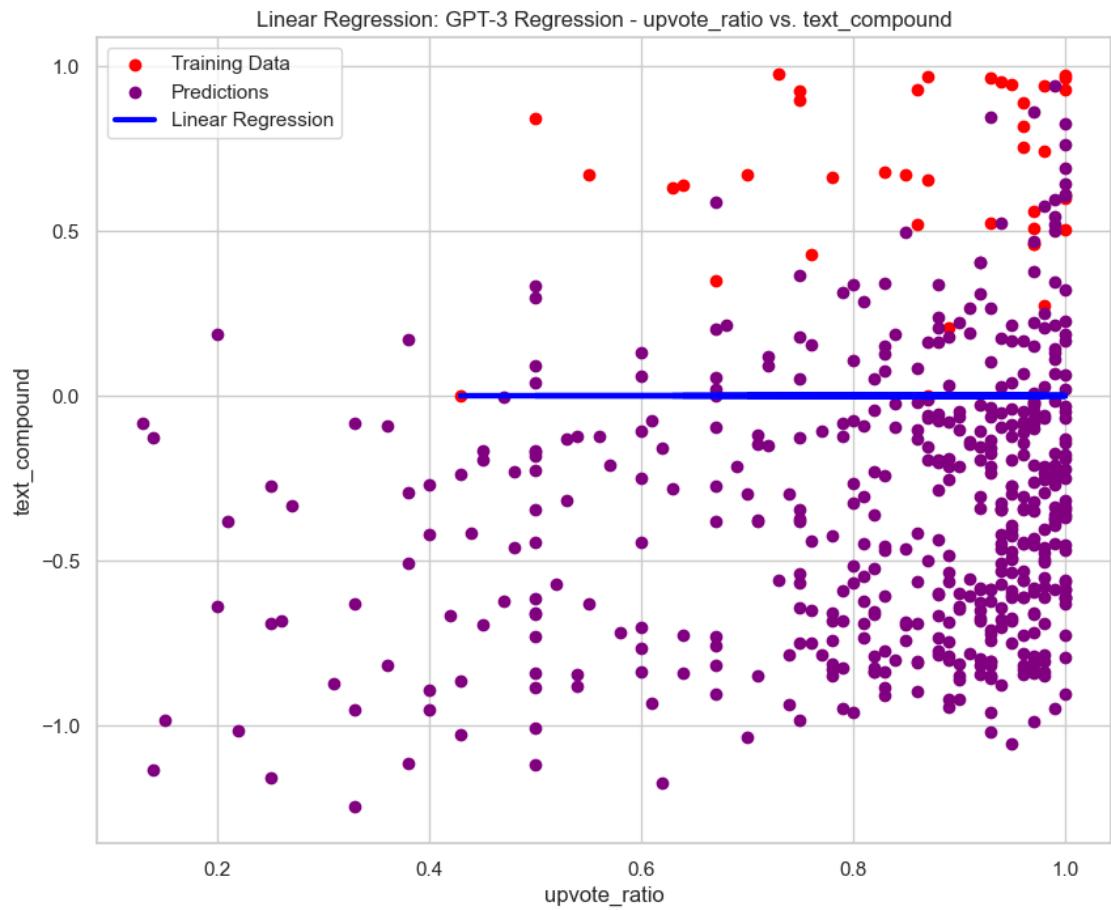




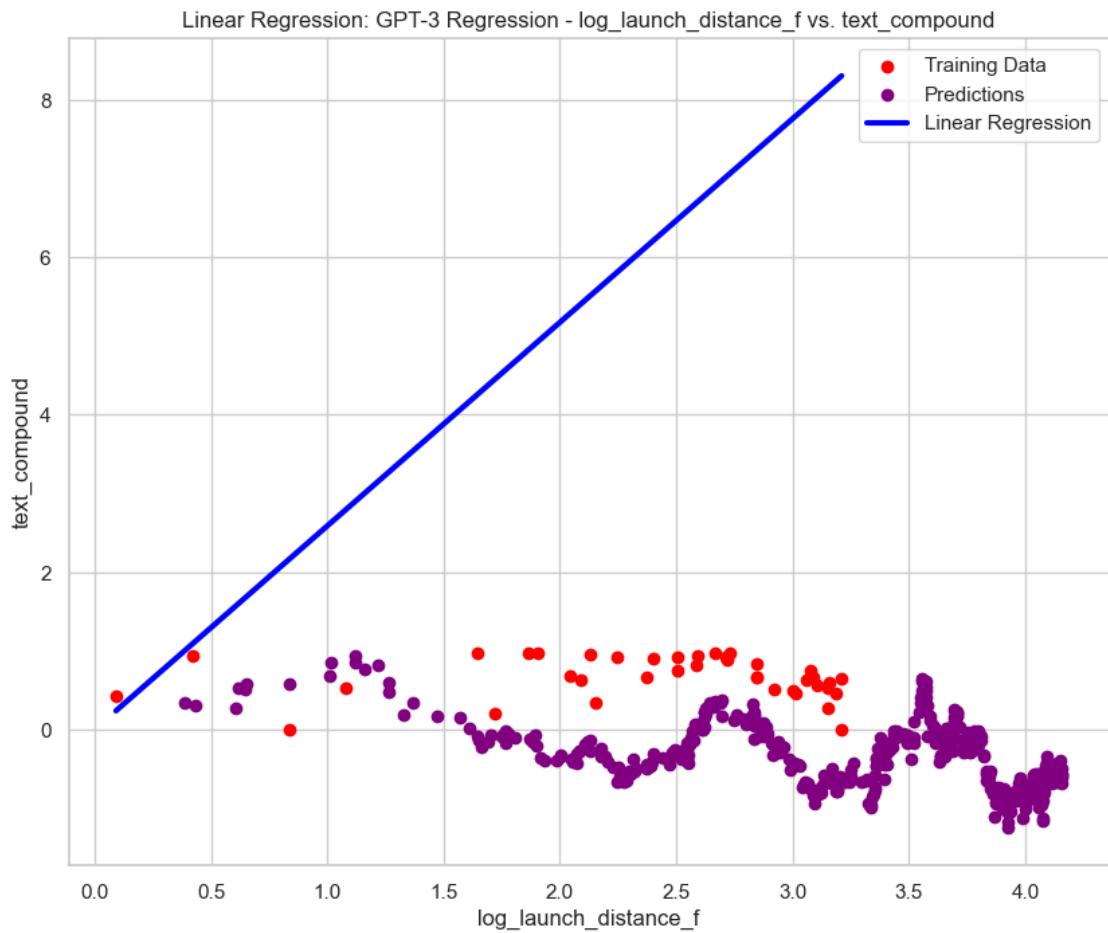




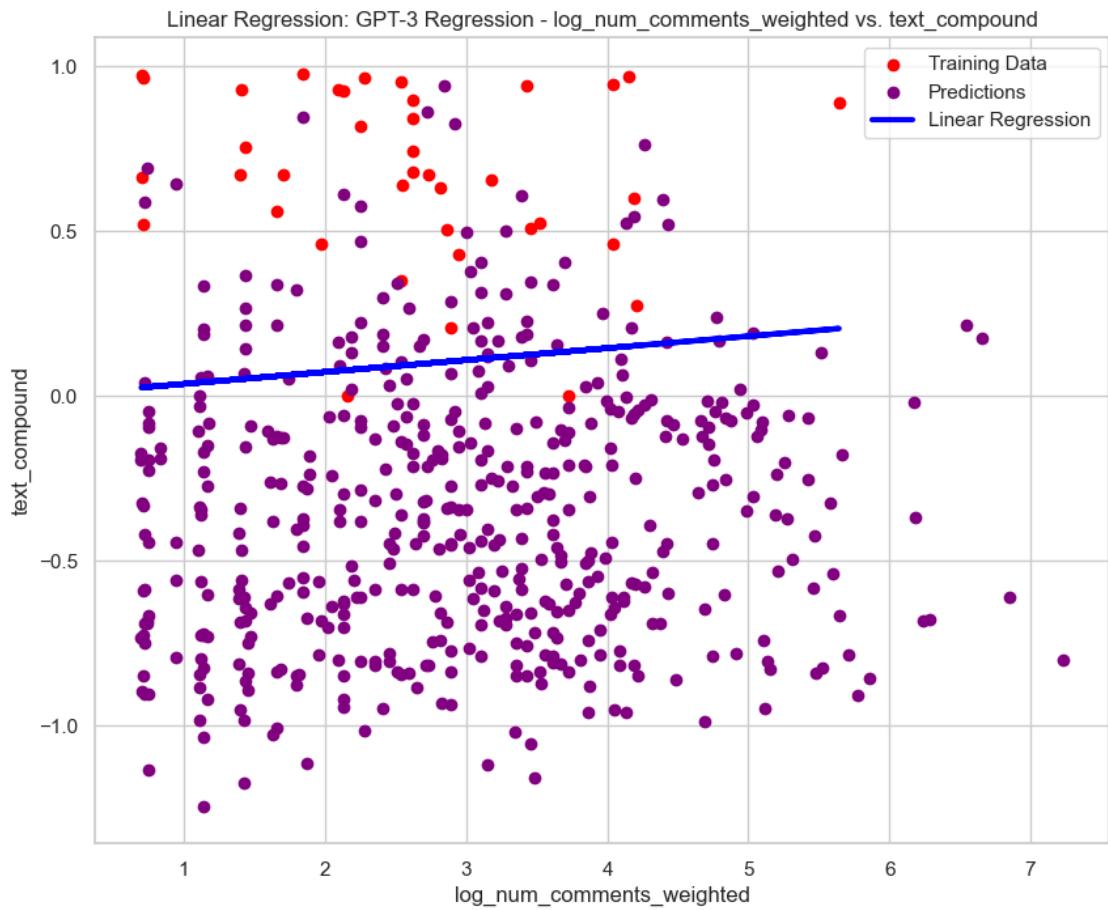
launched=0.4397653973506887



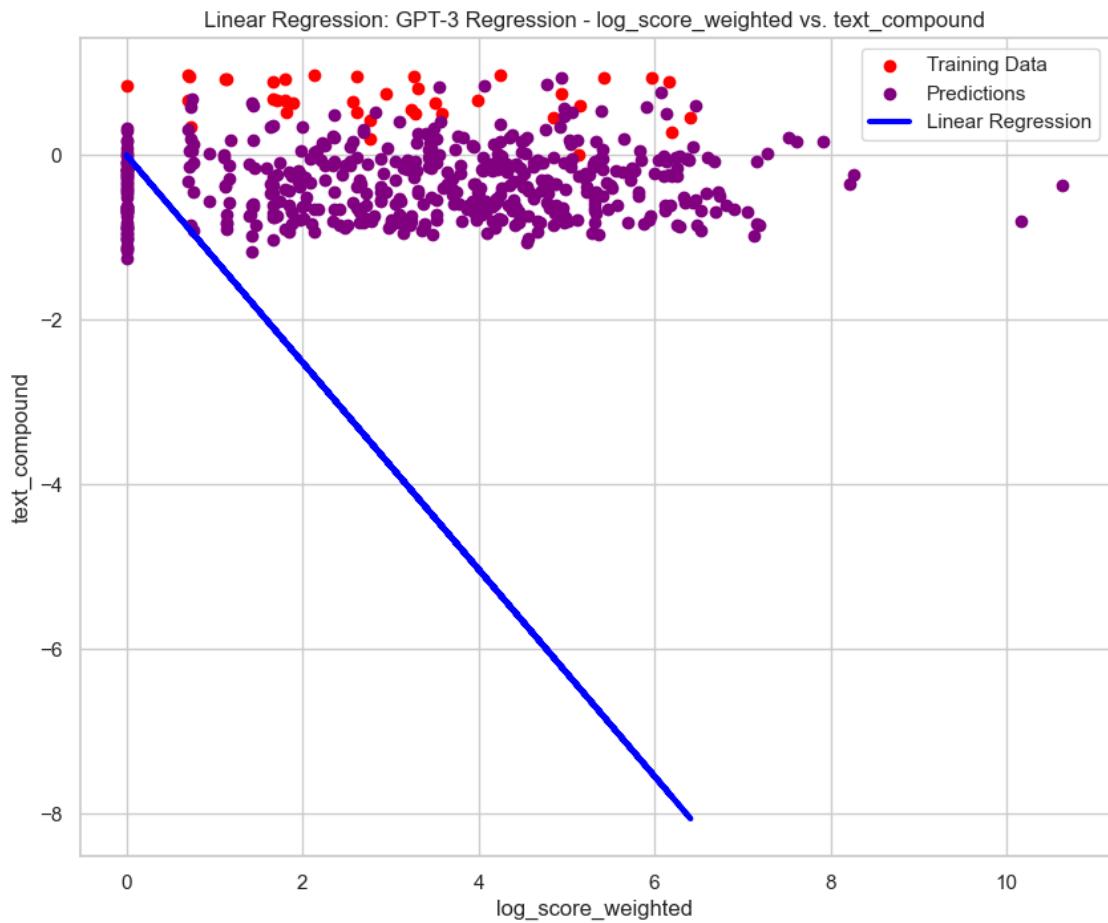
upvote_ratio=1.9626661407201595e-14



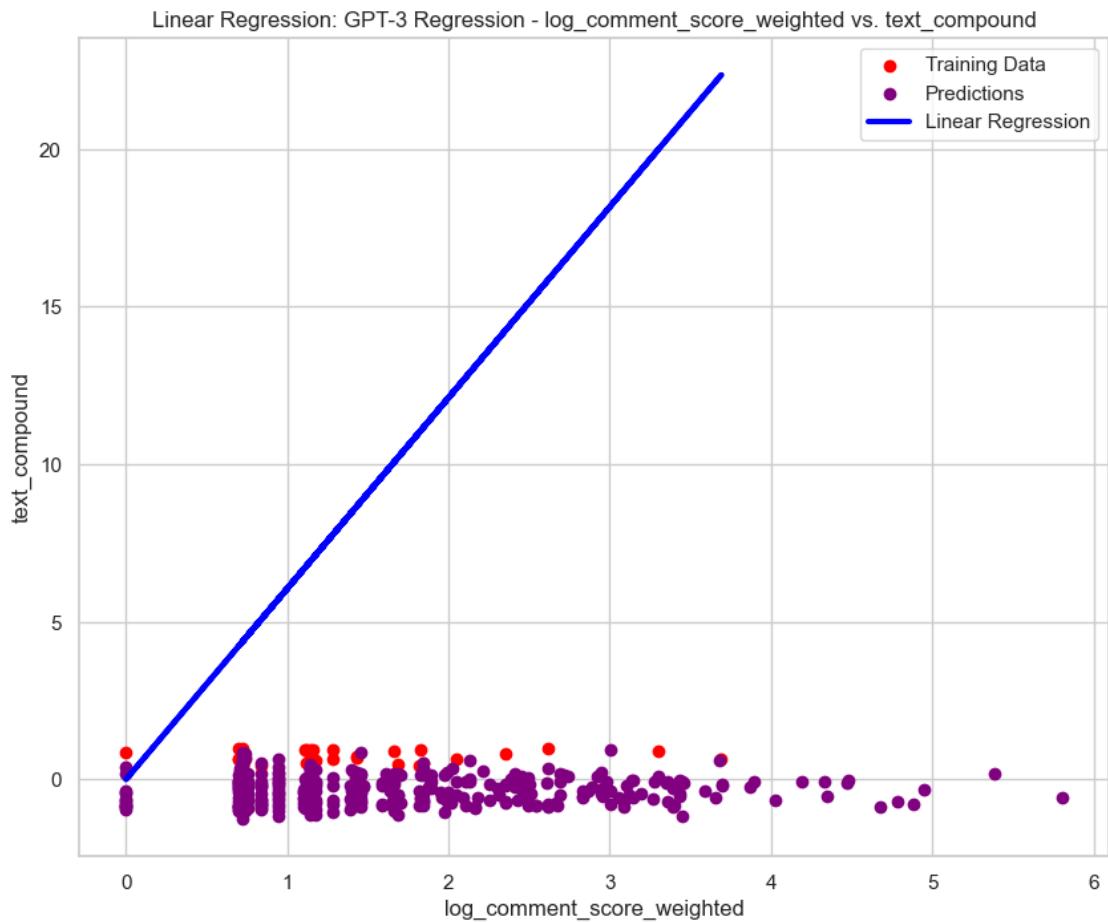
log_launch_distance_f=2.5856984477923586



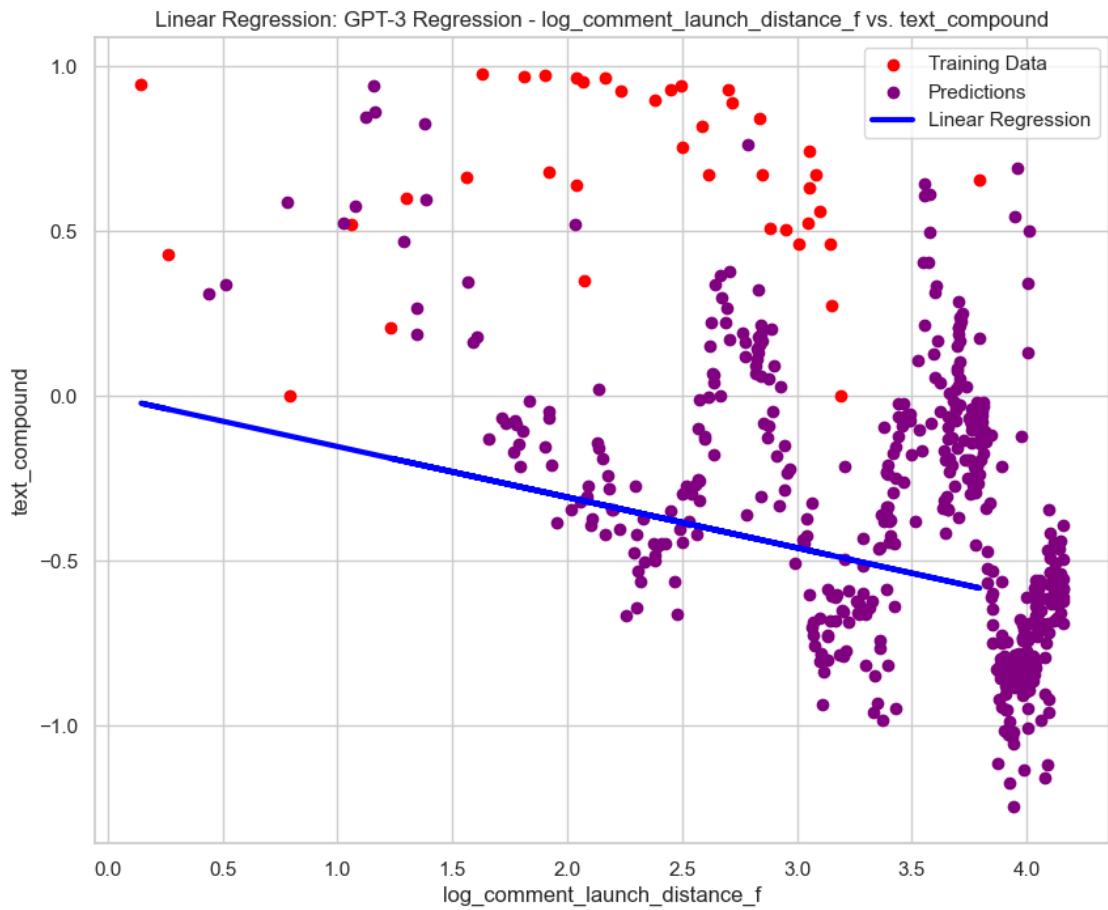
log_num_comments_weighted=0.03612687699553232



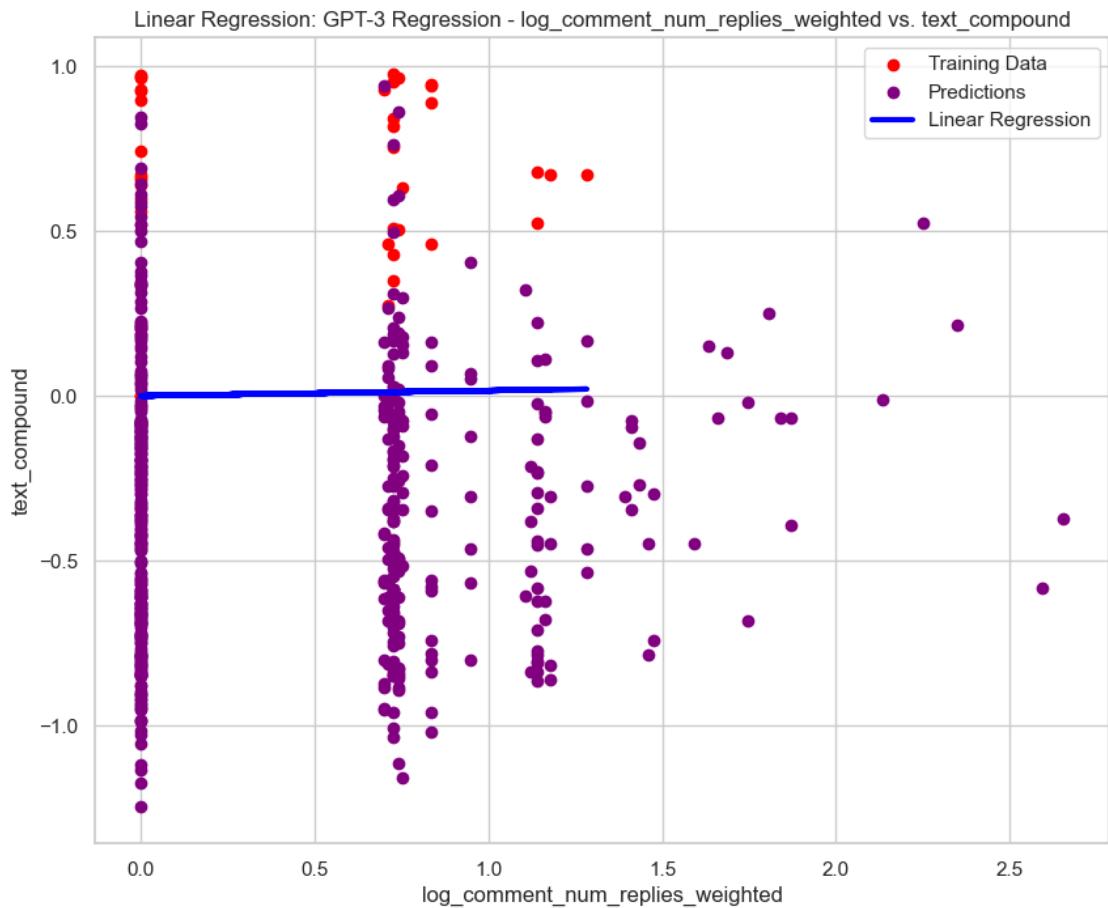
log_score_weighted=-1.257073338244043



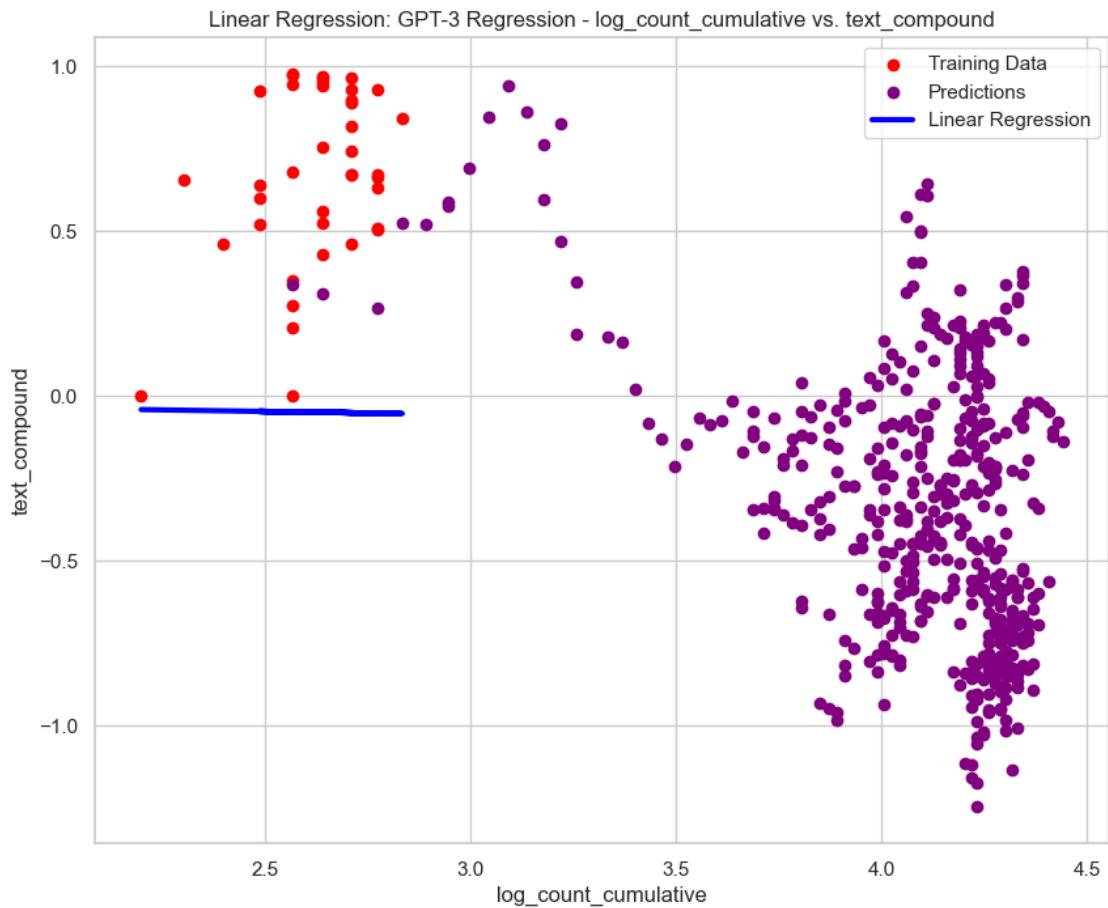
log_comment_score_weighted=6.054710137043305



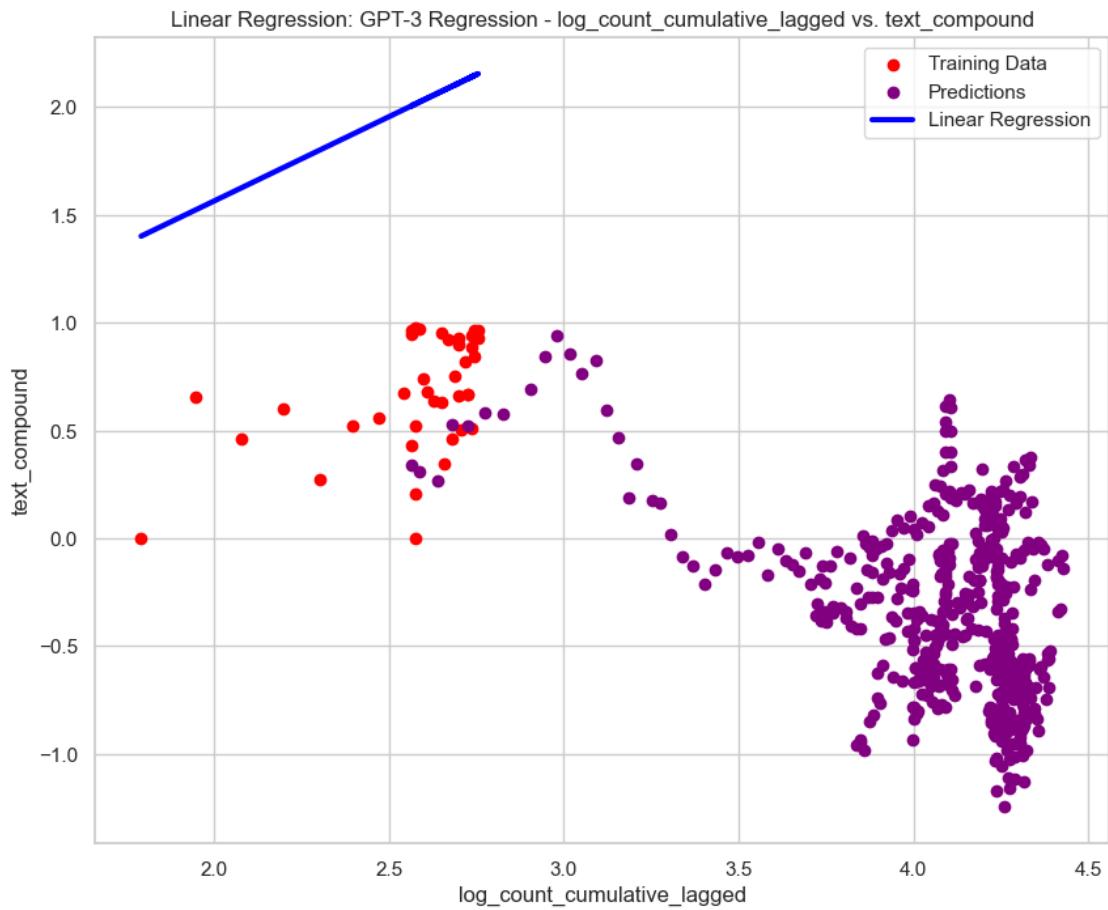
log_comment_launch_distance_f=-0.15343994928527524



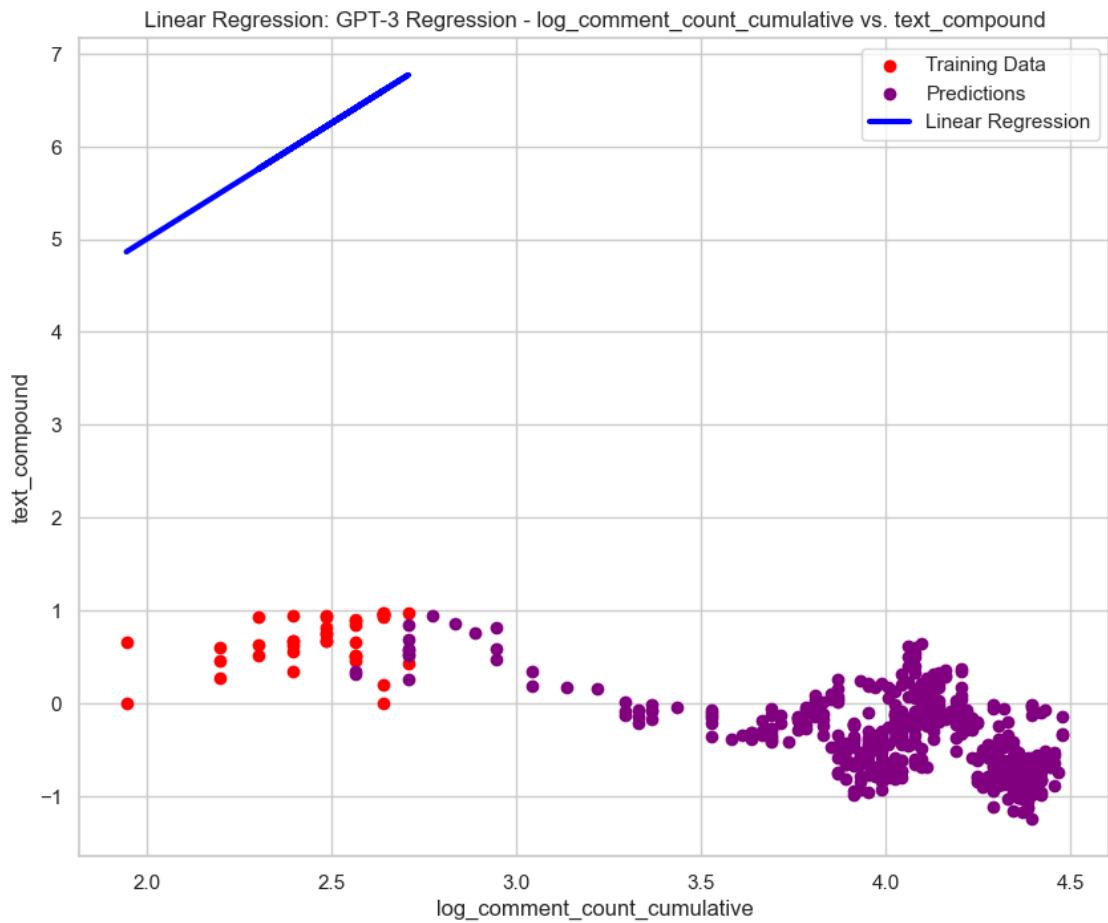
log_comment_num_replies_weighted=0.015980590335701712



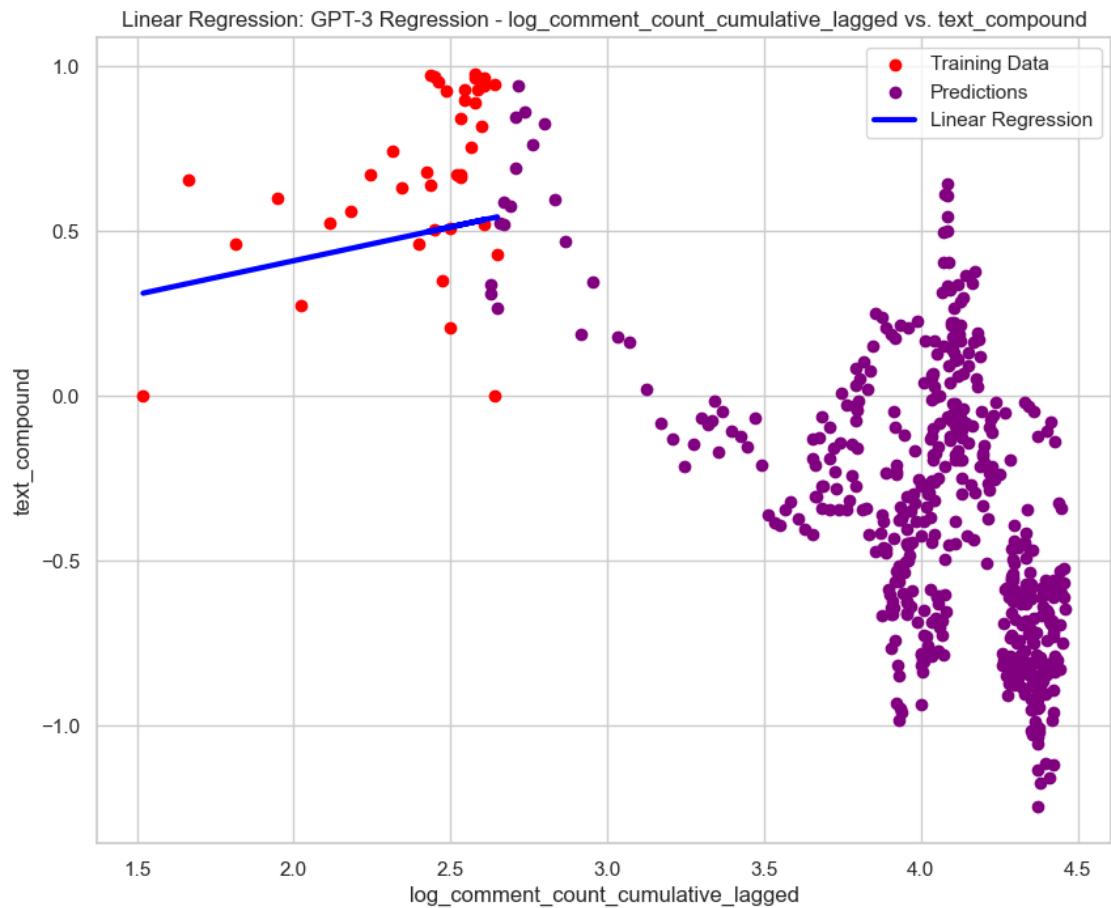
log_count_cumulative=-0.018934099486996116



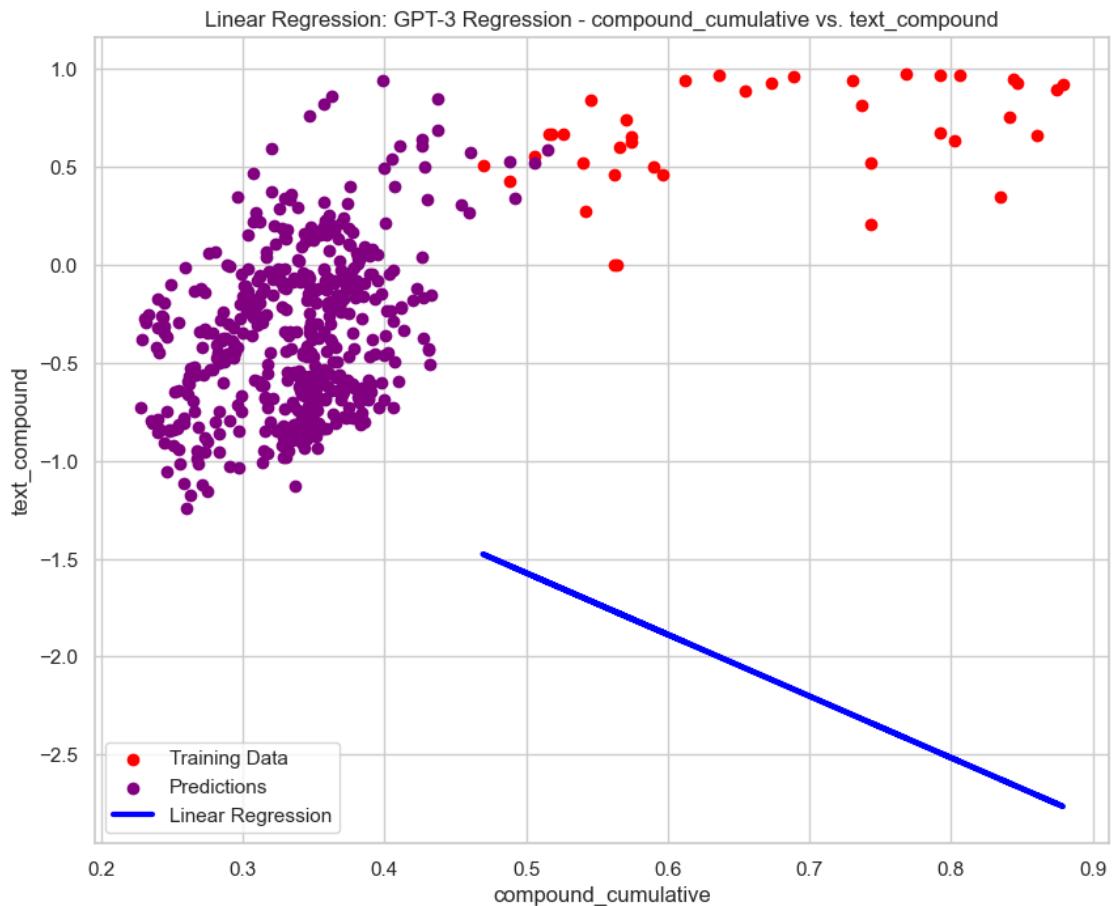
log_count_cumulative_lagged=0.7821090882097406



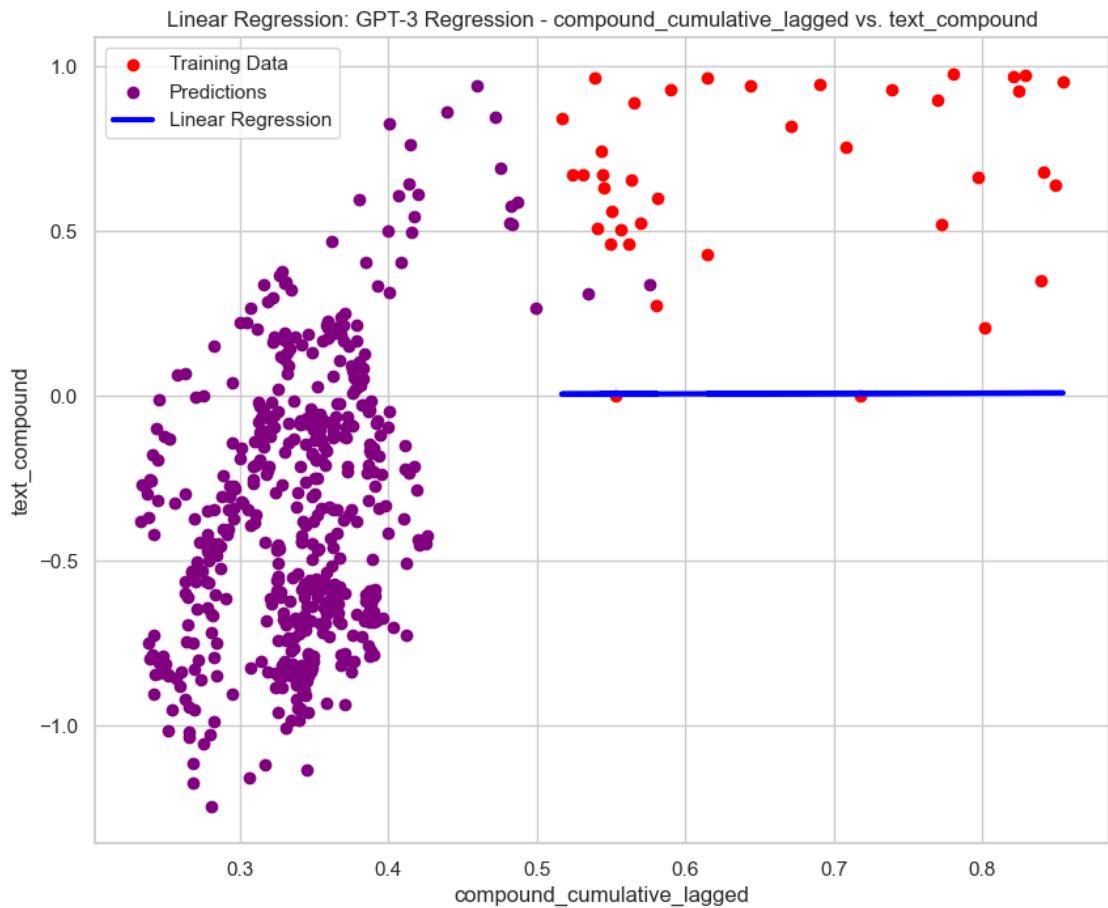
log_comment_count_cumulative=2.4989714440362287



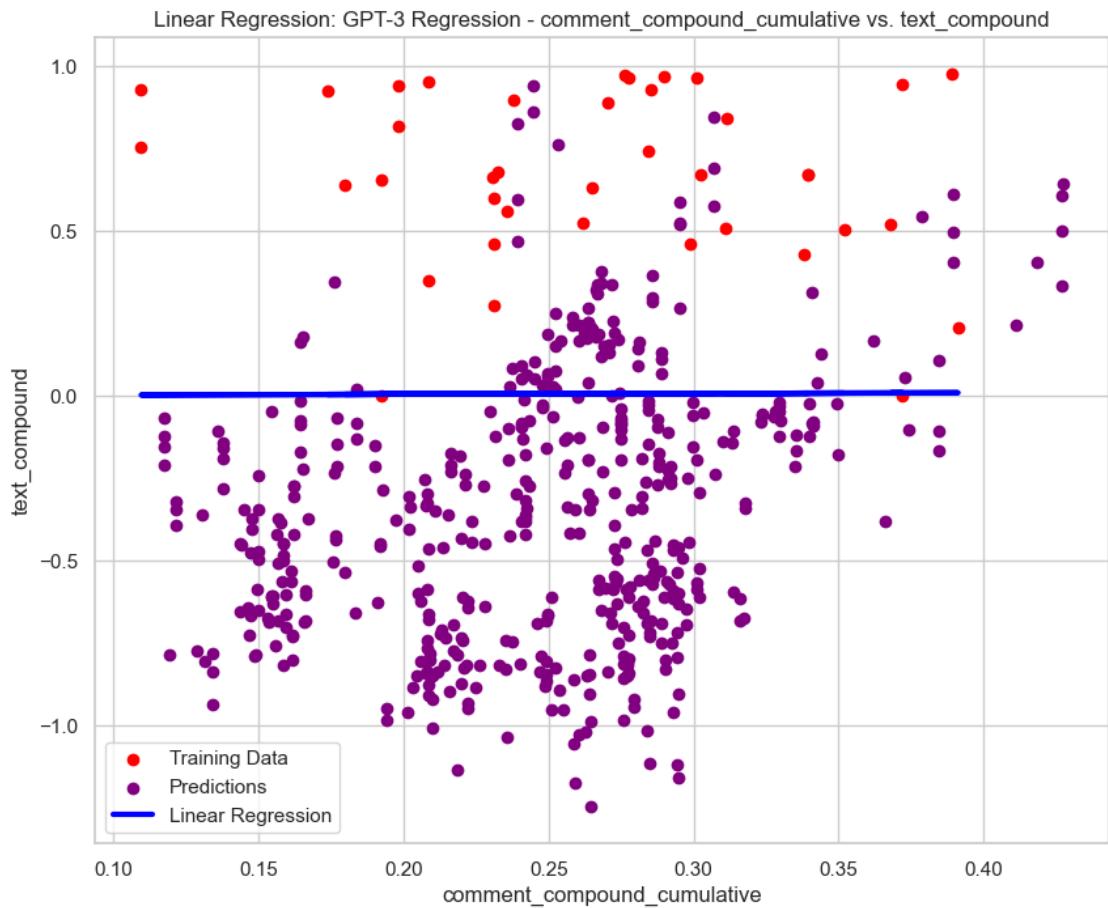
log_comment_count_cumulative_lagged=0.2046216036151977



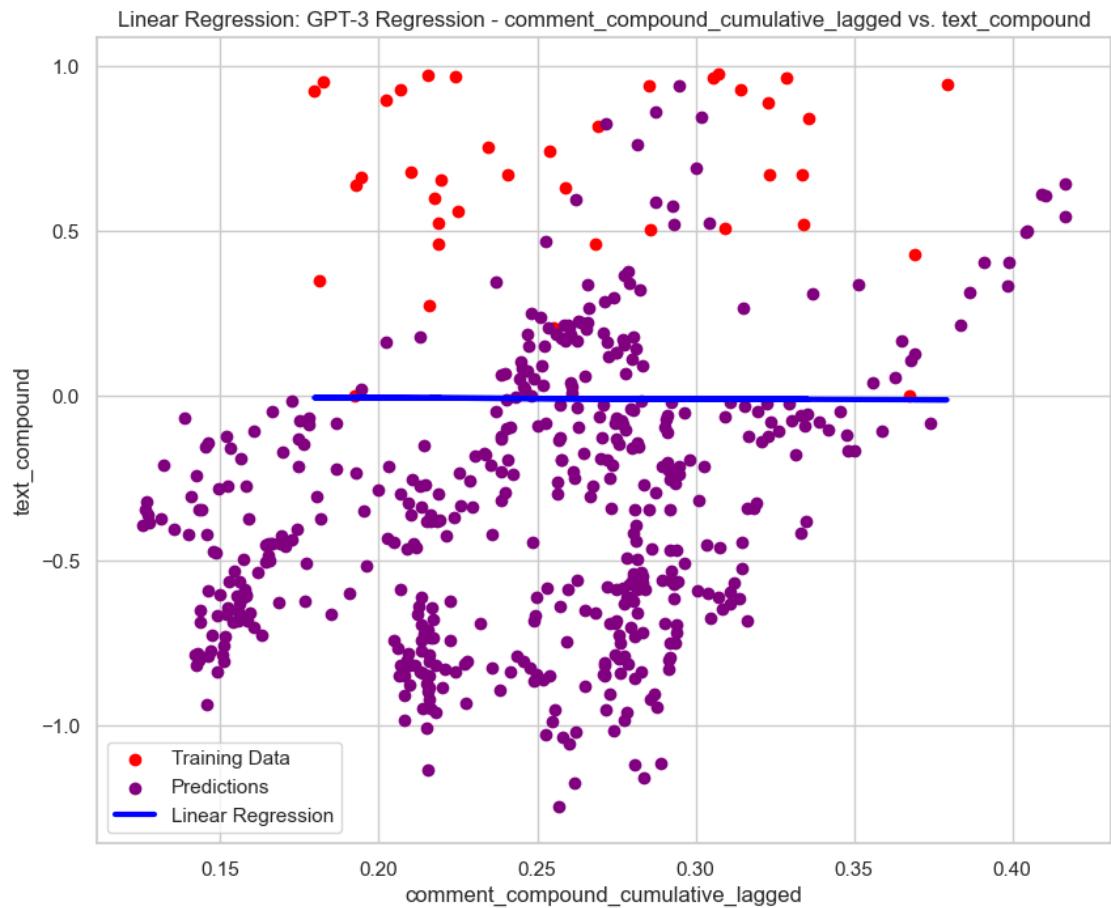
compound_cumulative=-3.1452136748772848



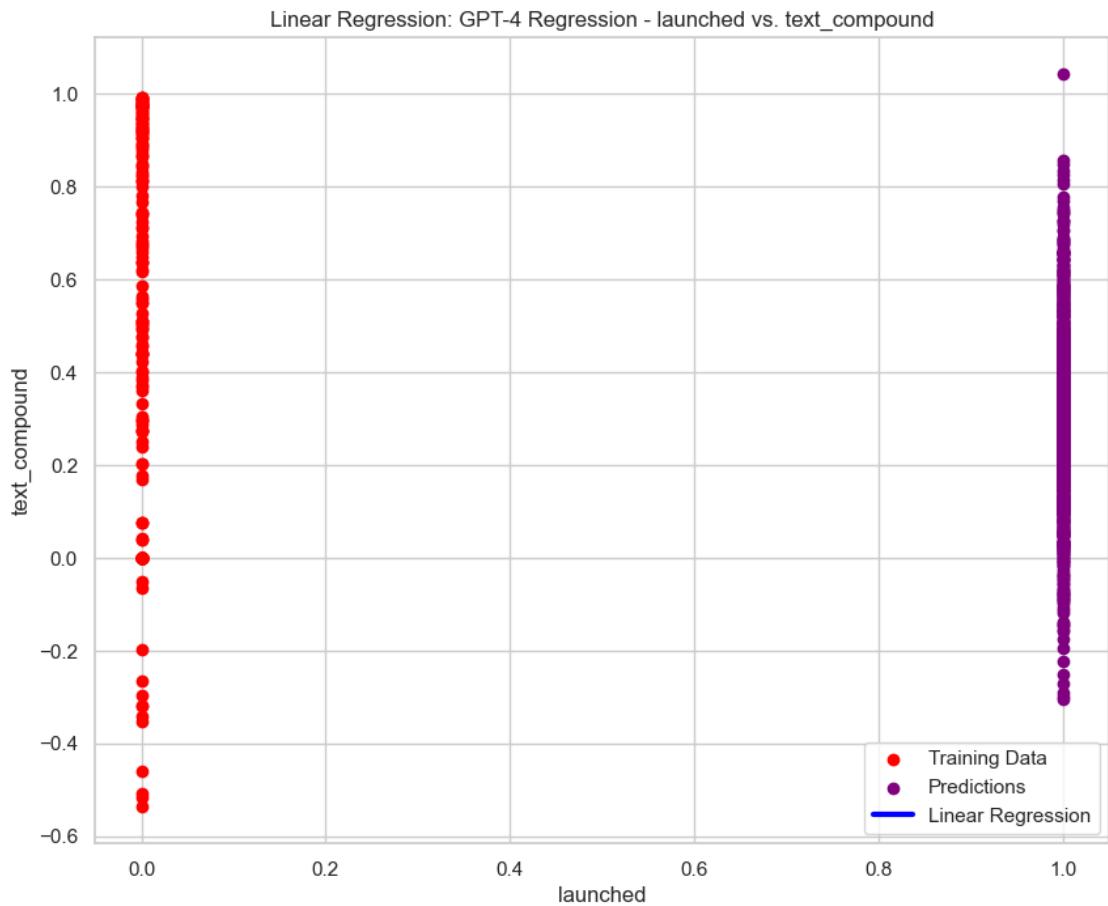
compound_cumulative_lagged=0.010201292425556978



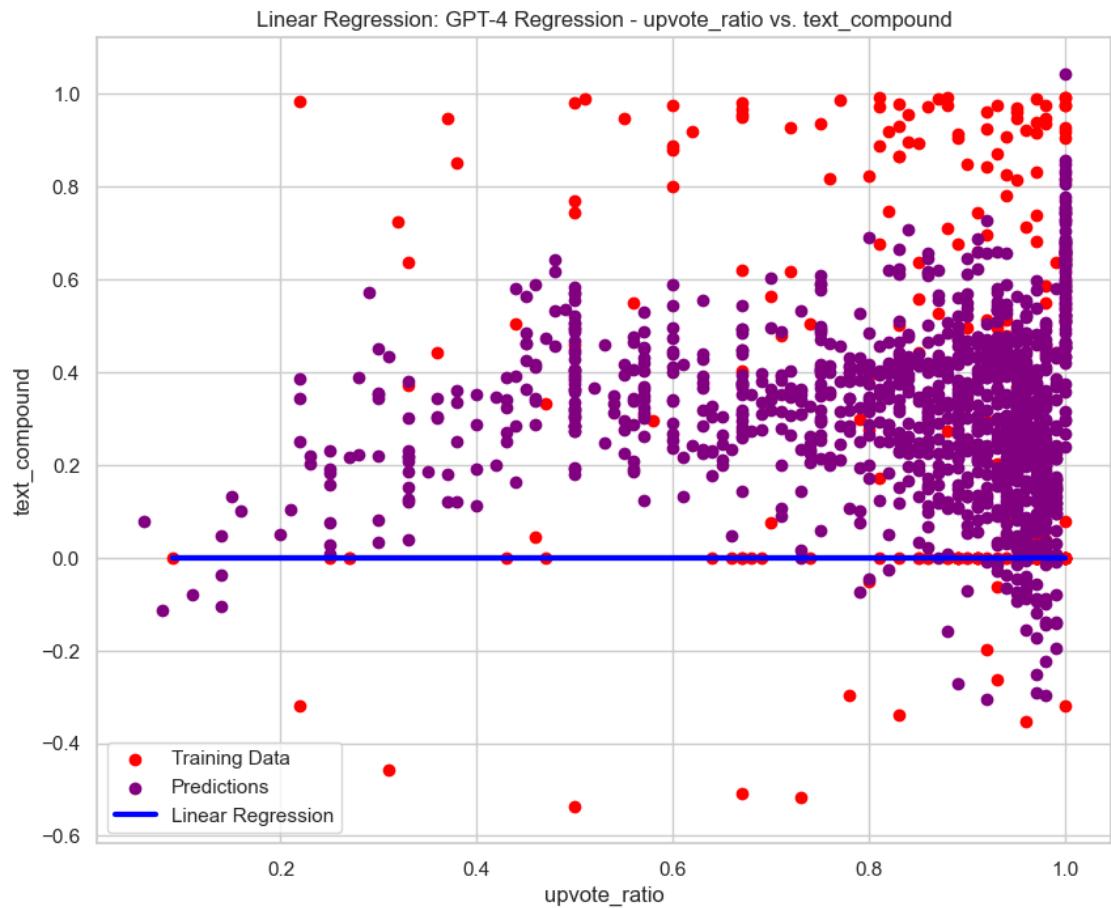
comment_compound_cumulative=0.02395088648333308



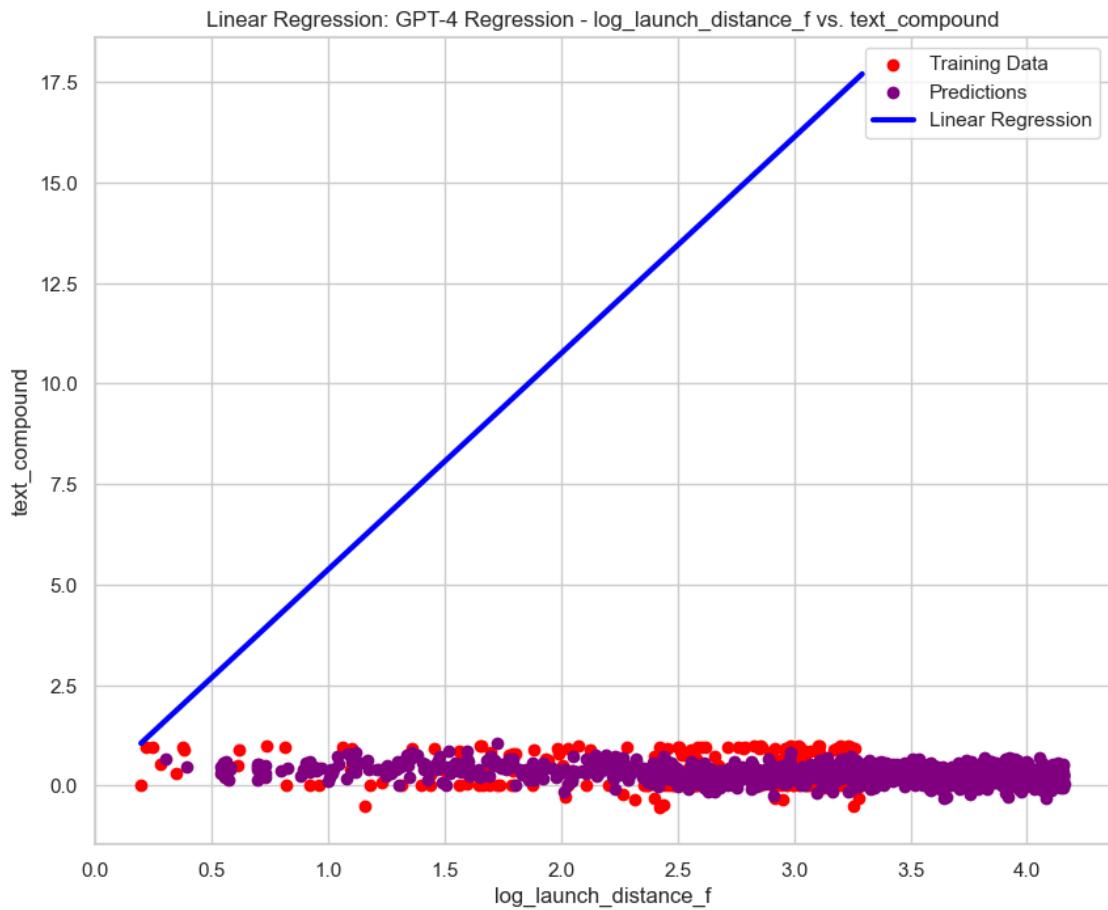
comment_compound_cumulative_lagged=-0.031650710097004604



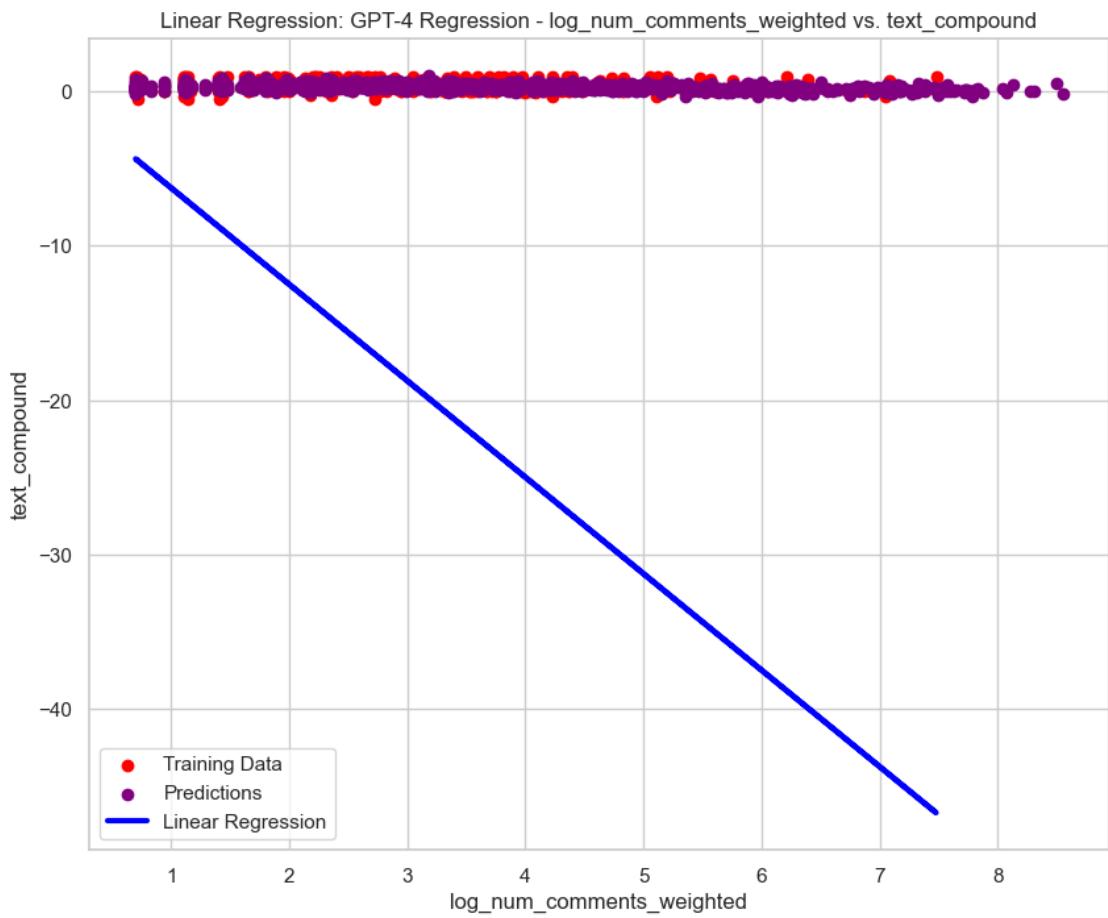
launched=0.8888992276148665



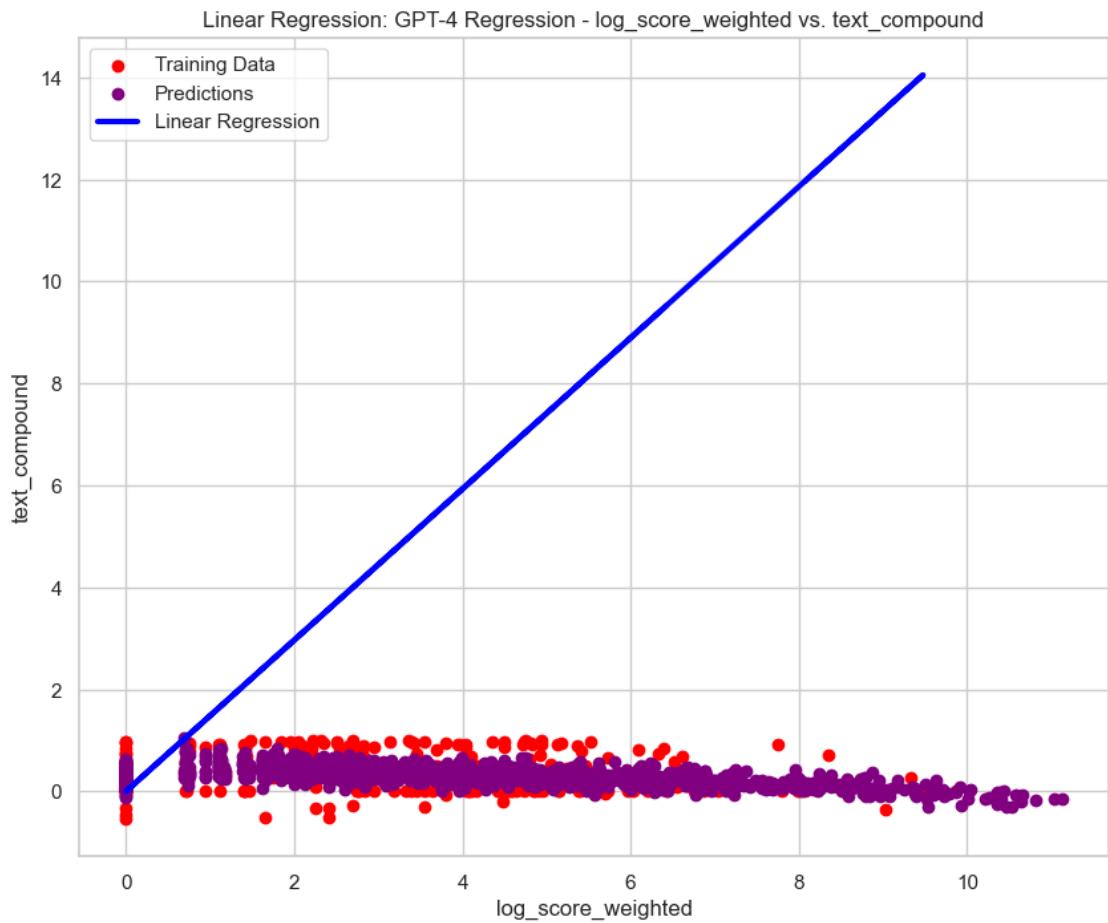
upvote_ratio=-1.1227130336521896e-14



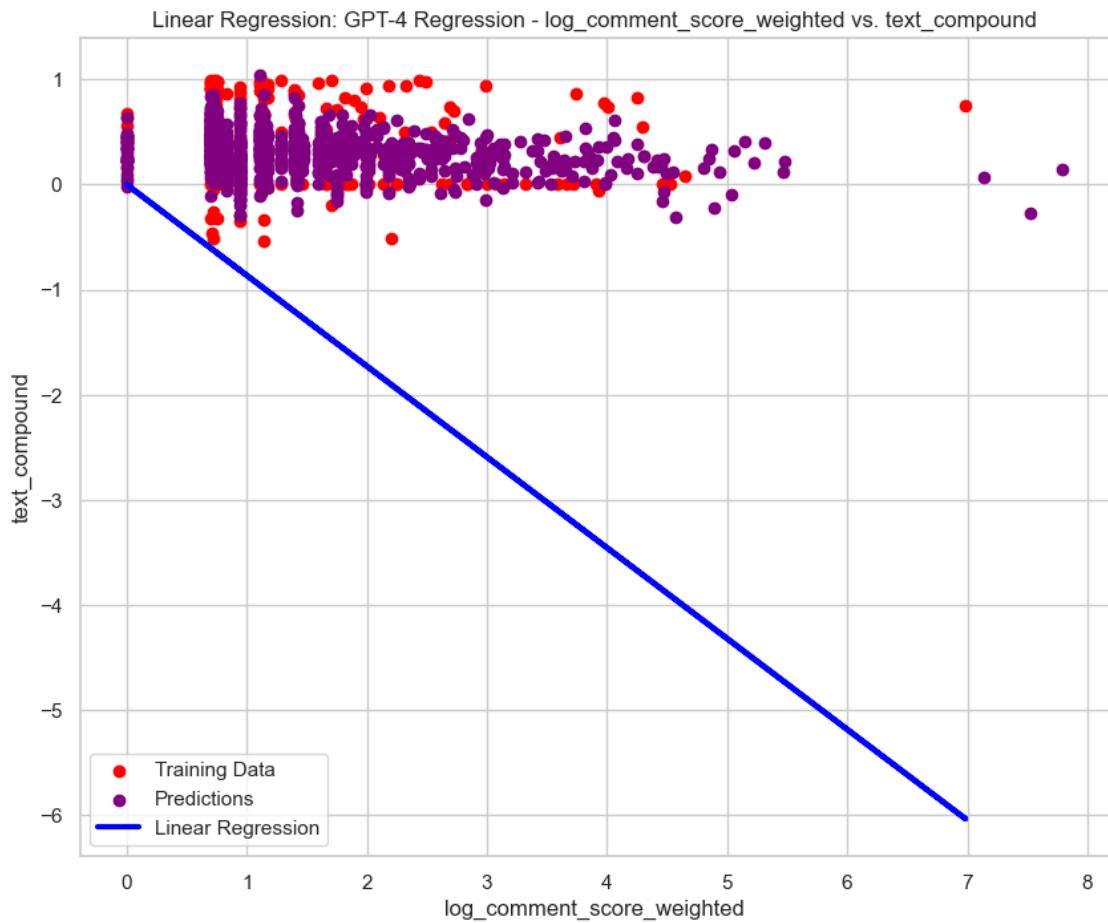
`log_launch_distance_f=5.374317438103284`



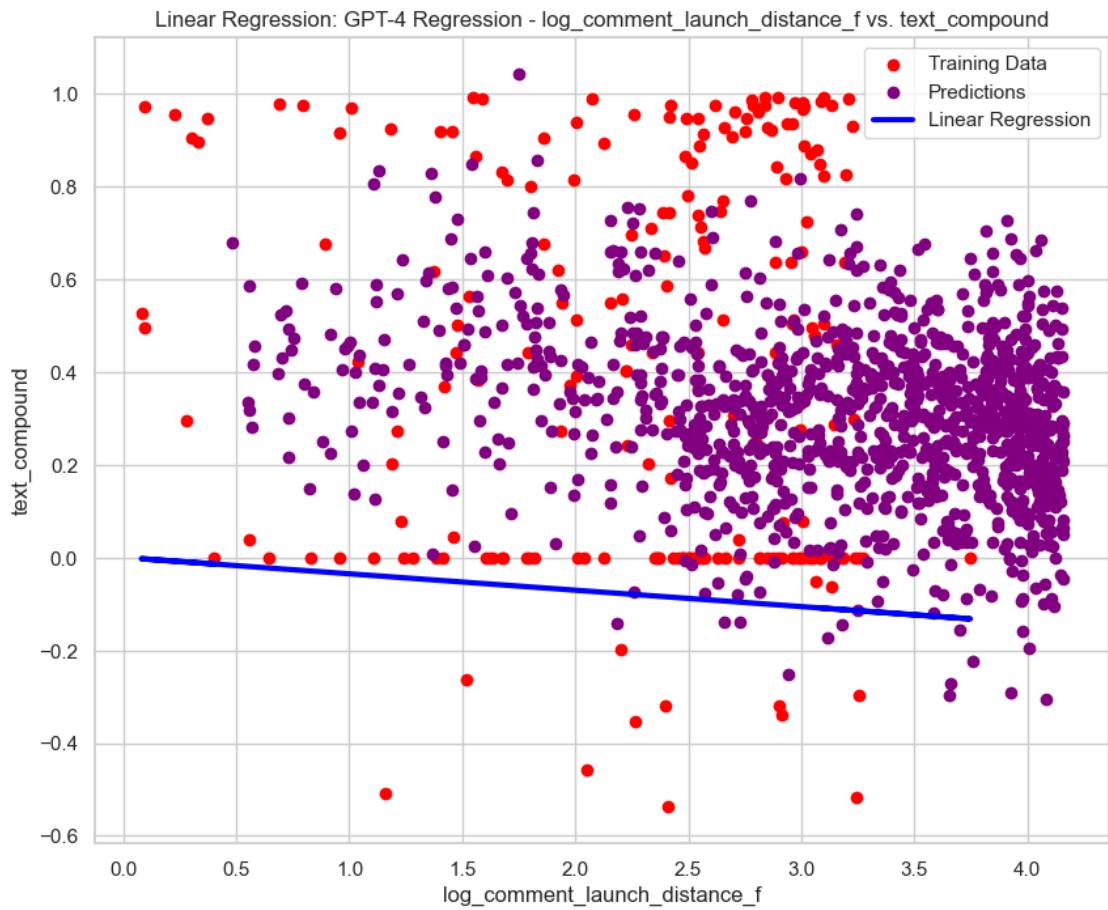
log_num_comments_weighted=-6.247890604479563



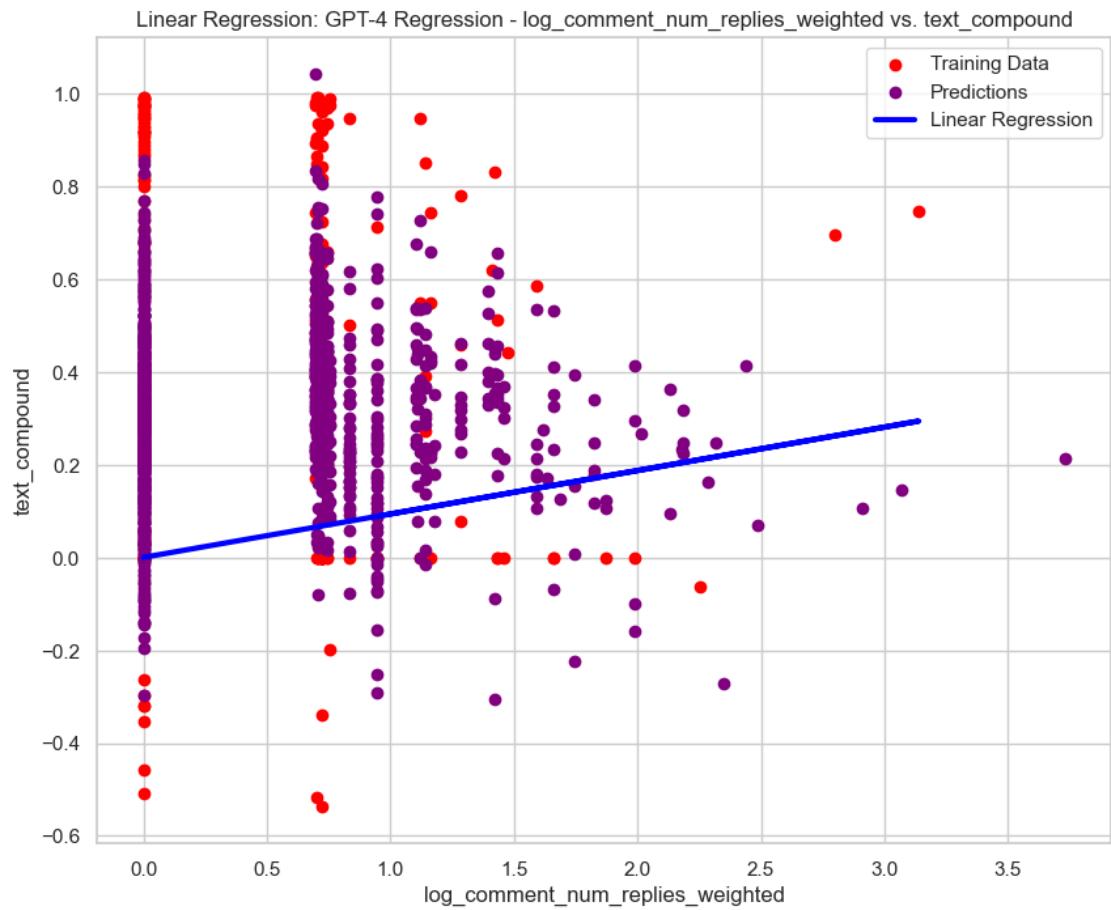
log_score_weighted=1.482504538917876



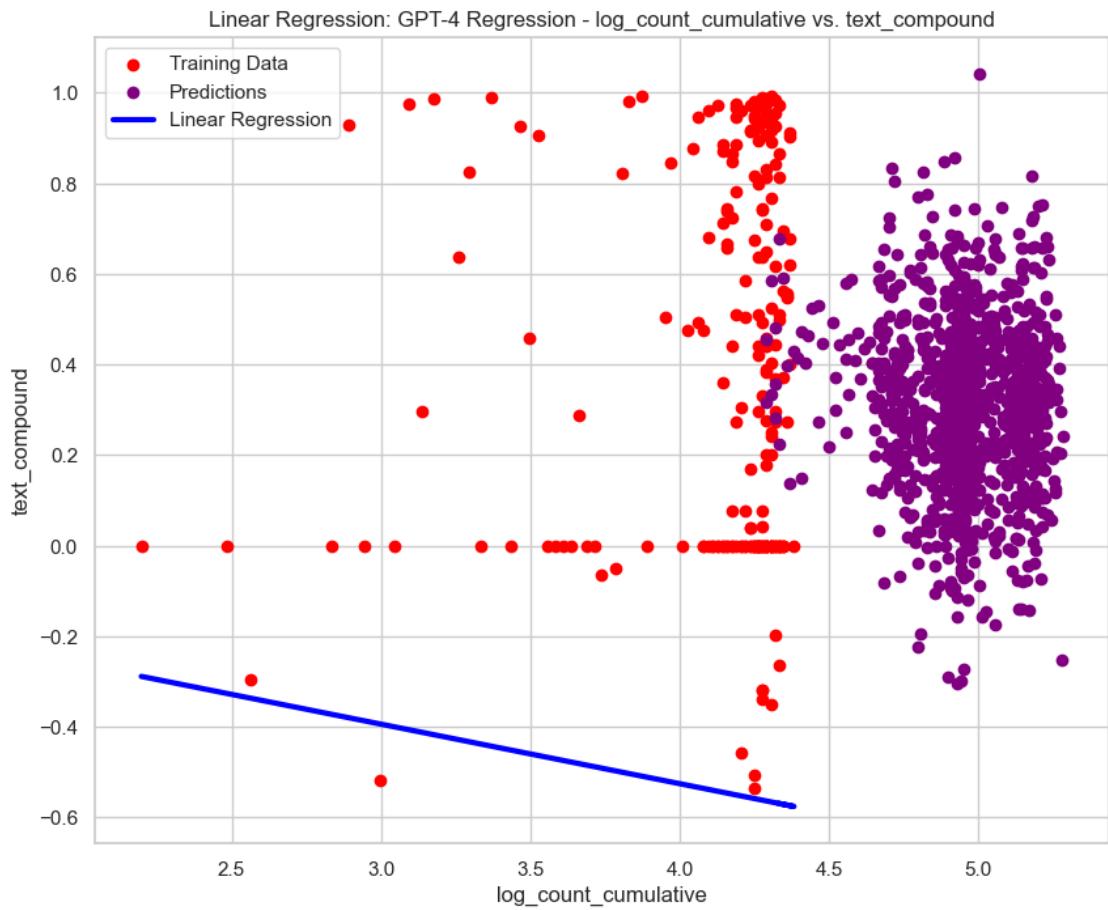
log_comment_score_weighted=-0.8642924263616882



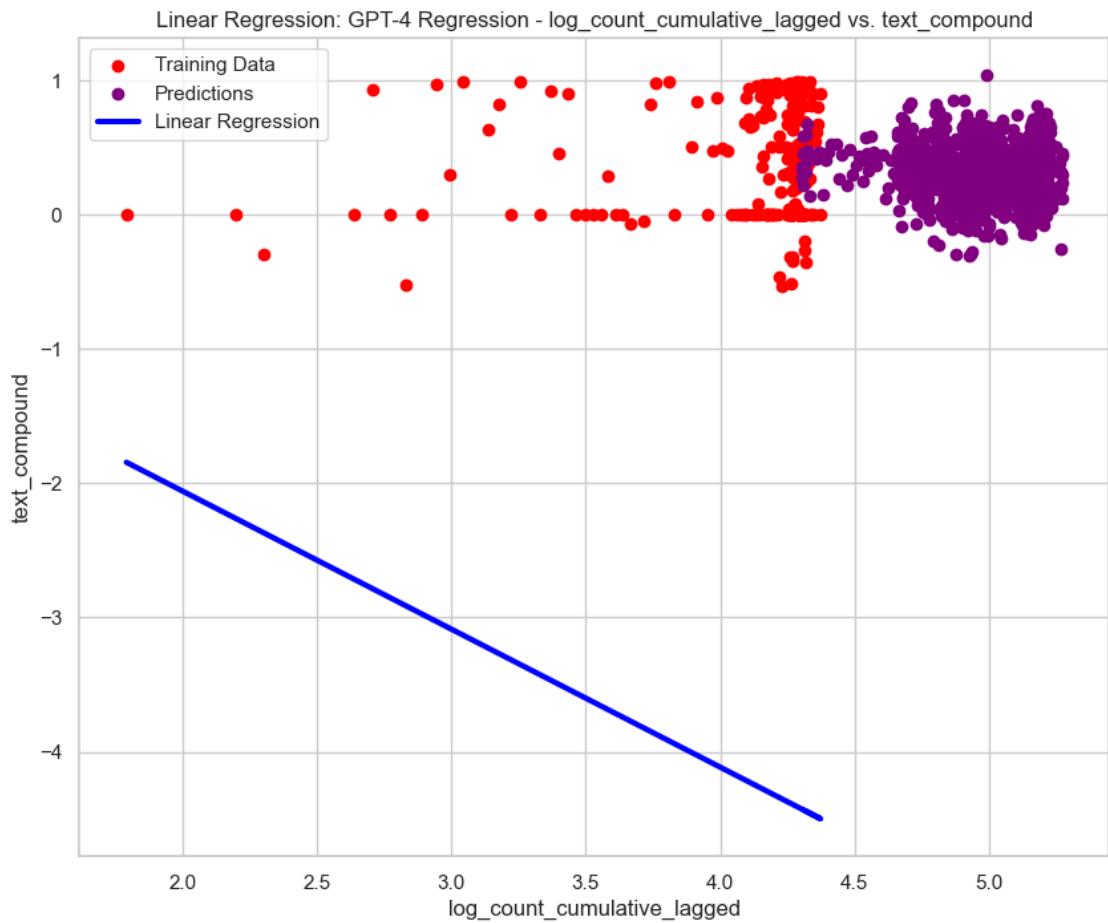
`log_comment_launch_distance_f=-0.03531010513615149`



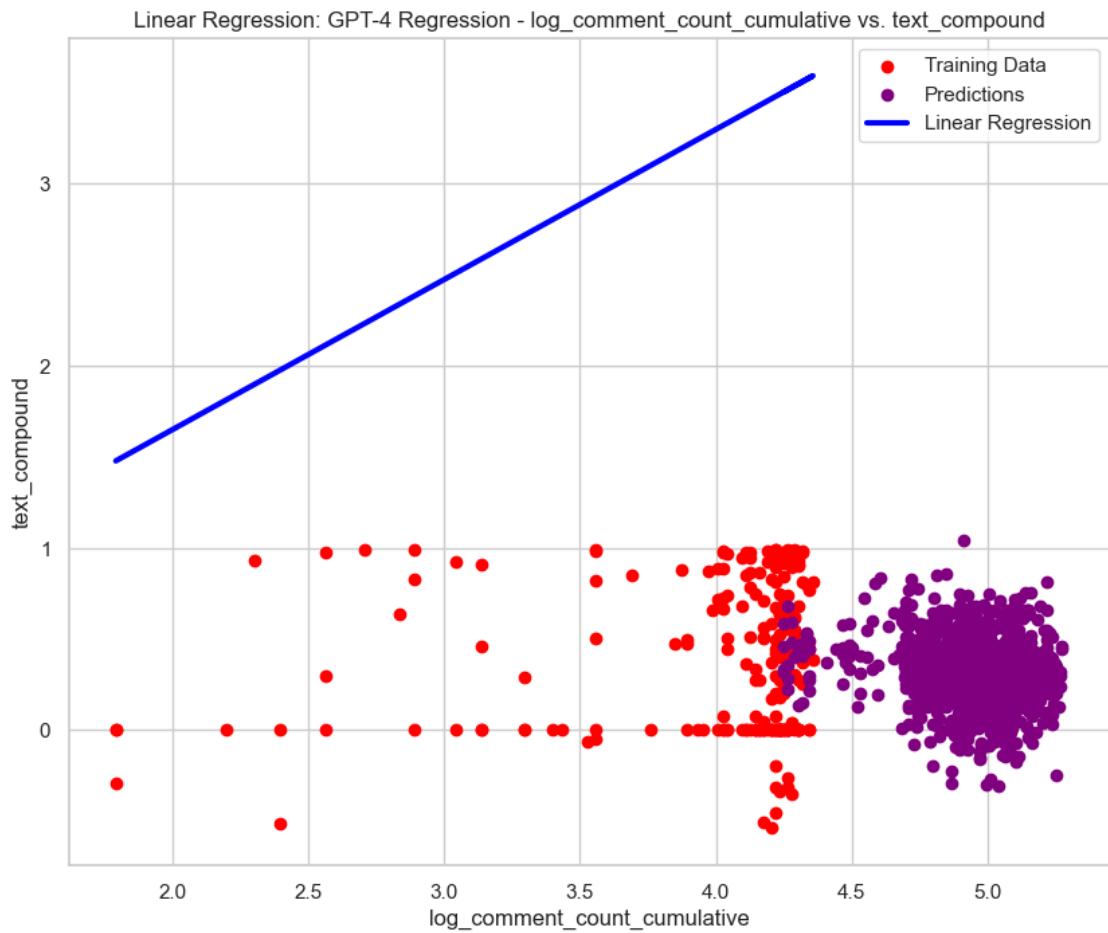
log_comment_num_replies_weighted=0.09359049921661934



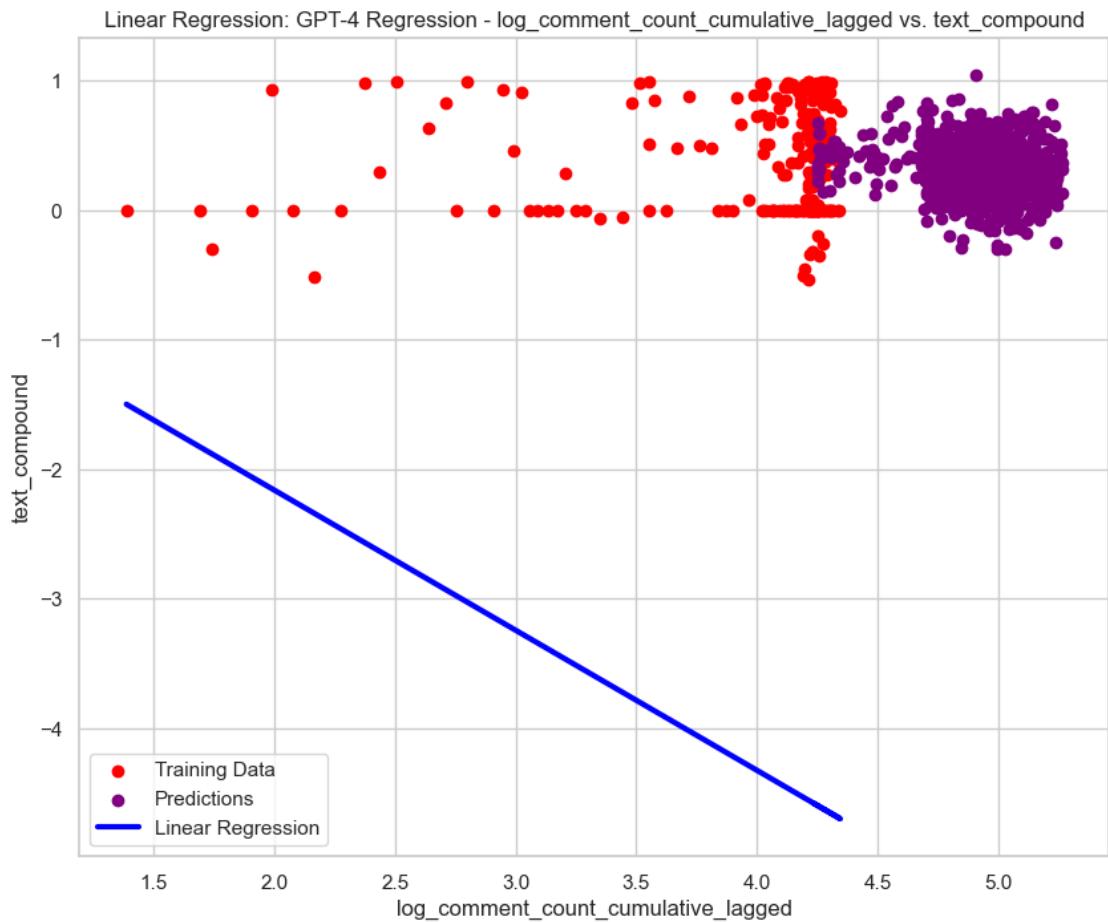
`log_count_cumulative=-0.13146685607372943`



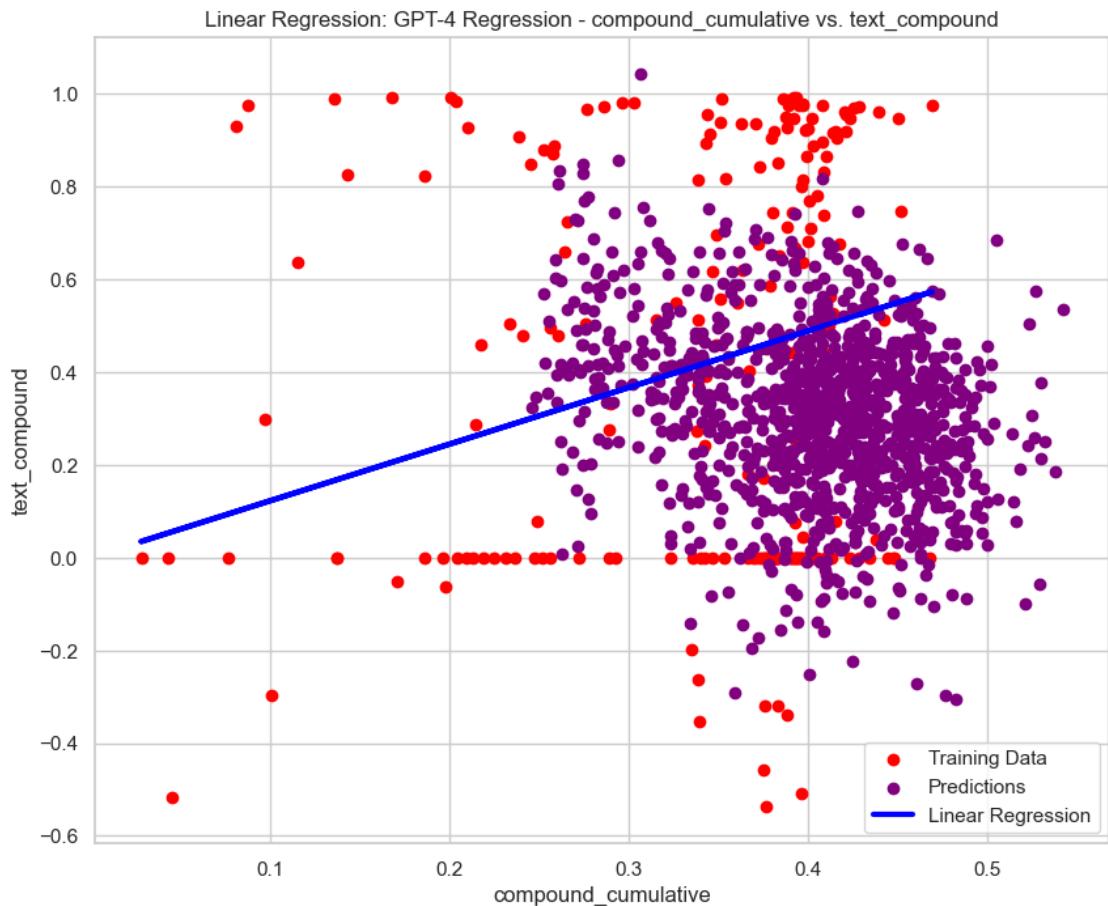
log_count_cumulative_lagged=-1.0287992087352718



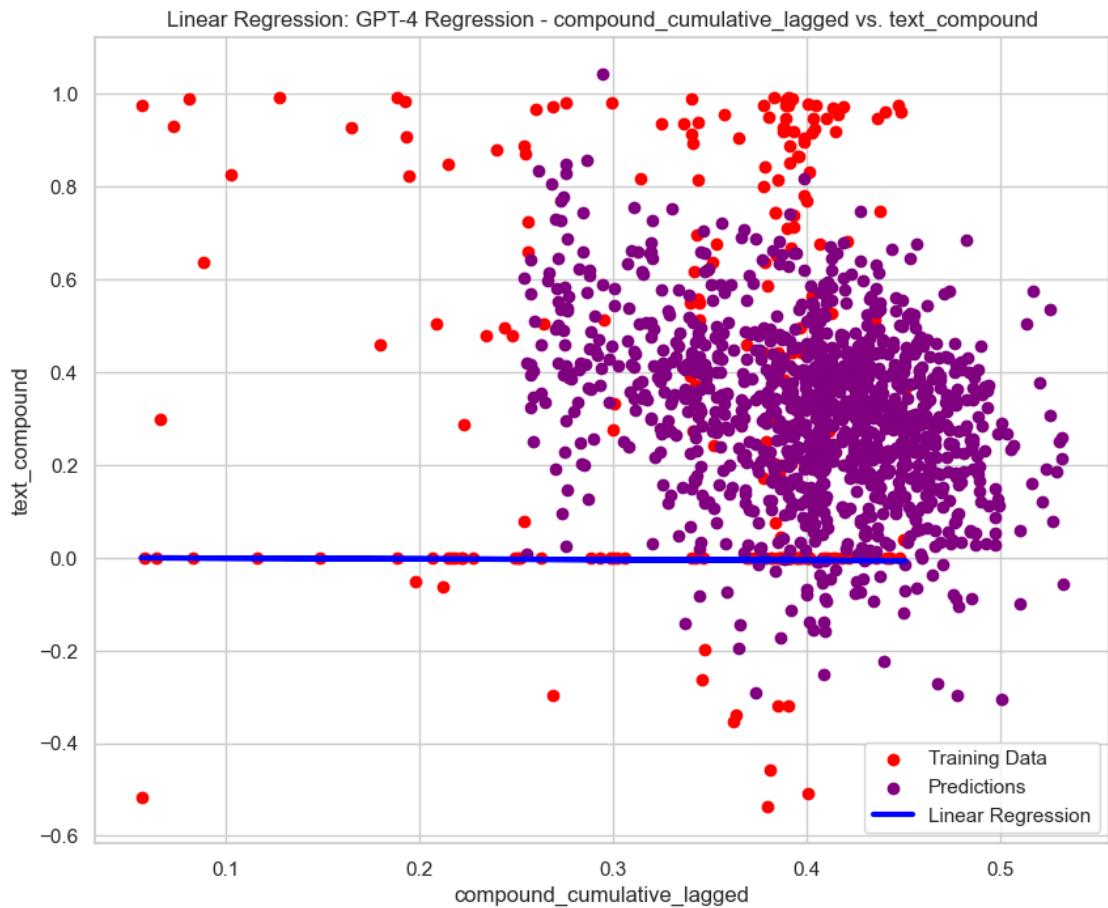
log_comment_count_cumulative=0.8247090035290257



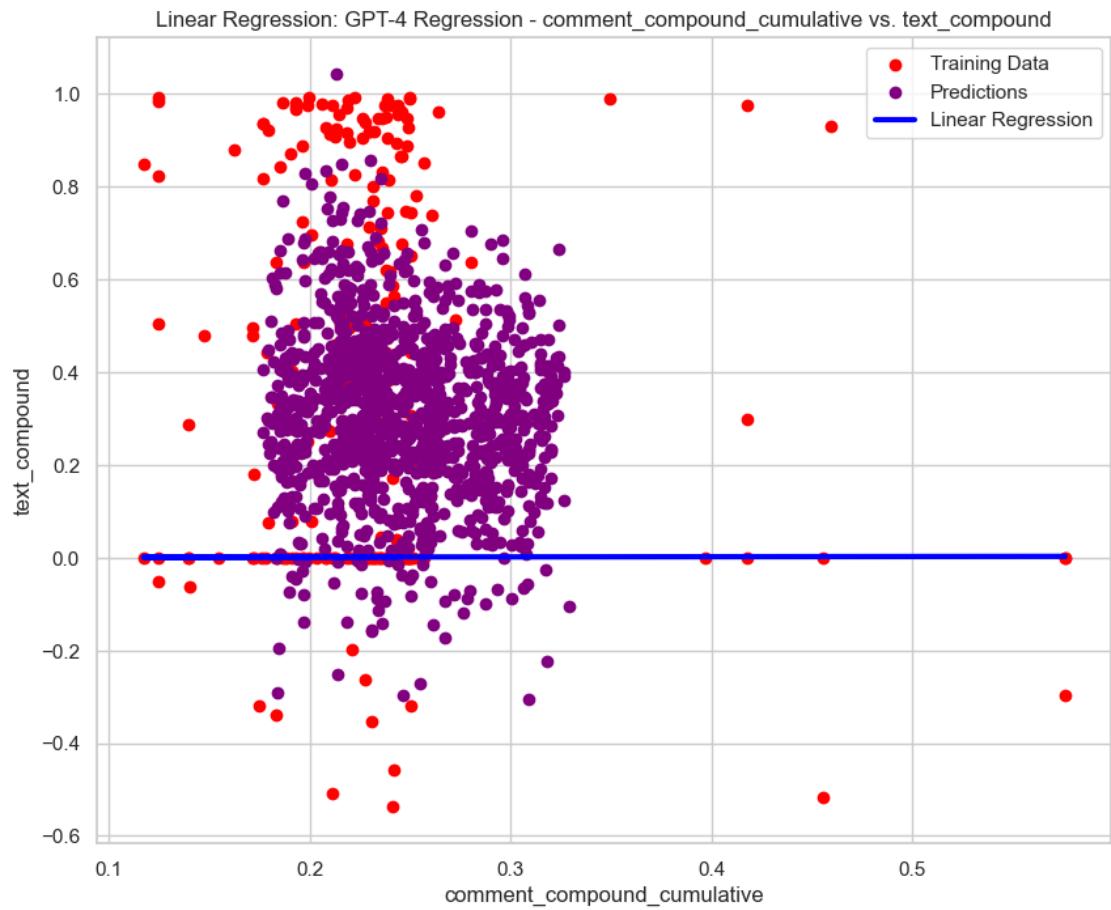
log_comment_count_cumulative_lagged=-1.0804842597811075



compound_cumulative=1.2202651759864822

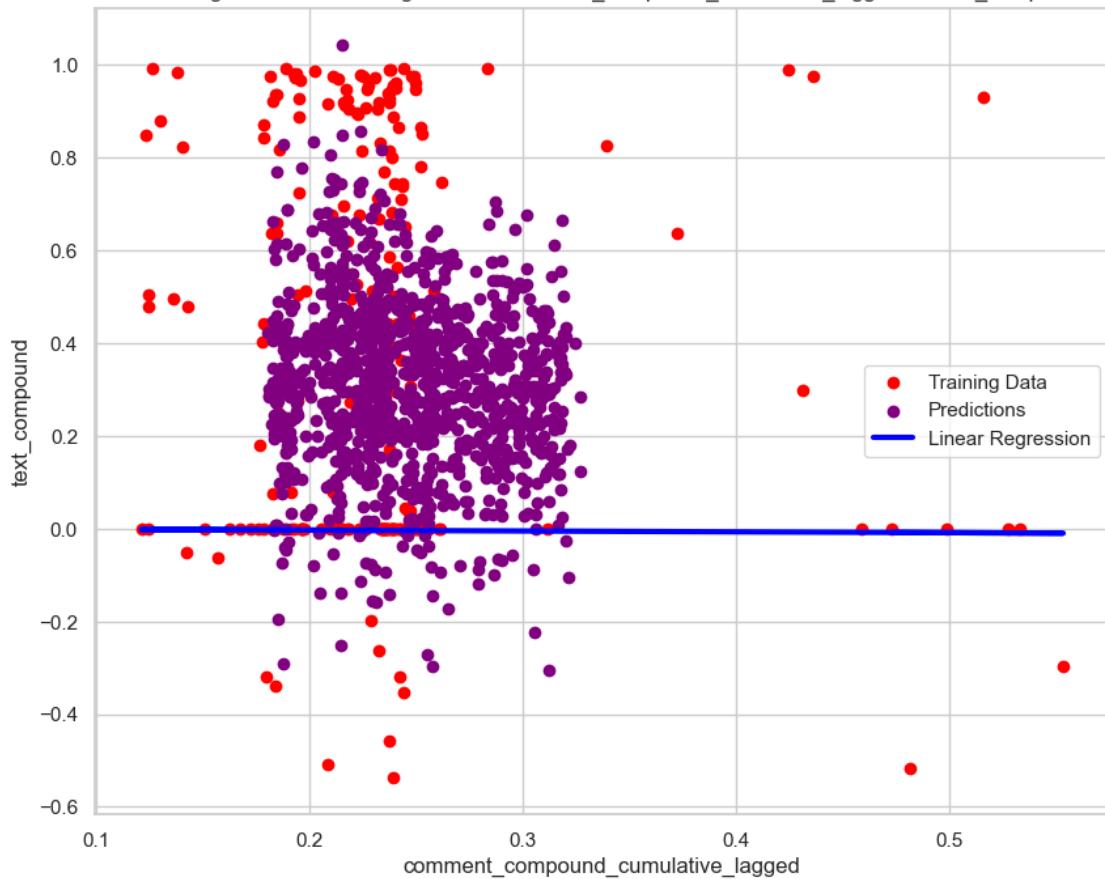


compound_cumulative_lagged=-0.015823577929702112

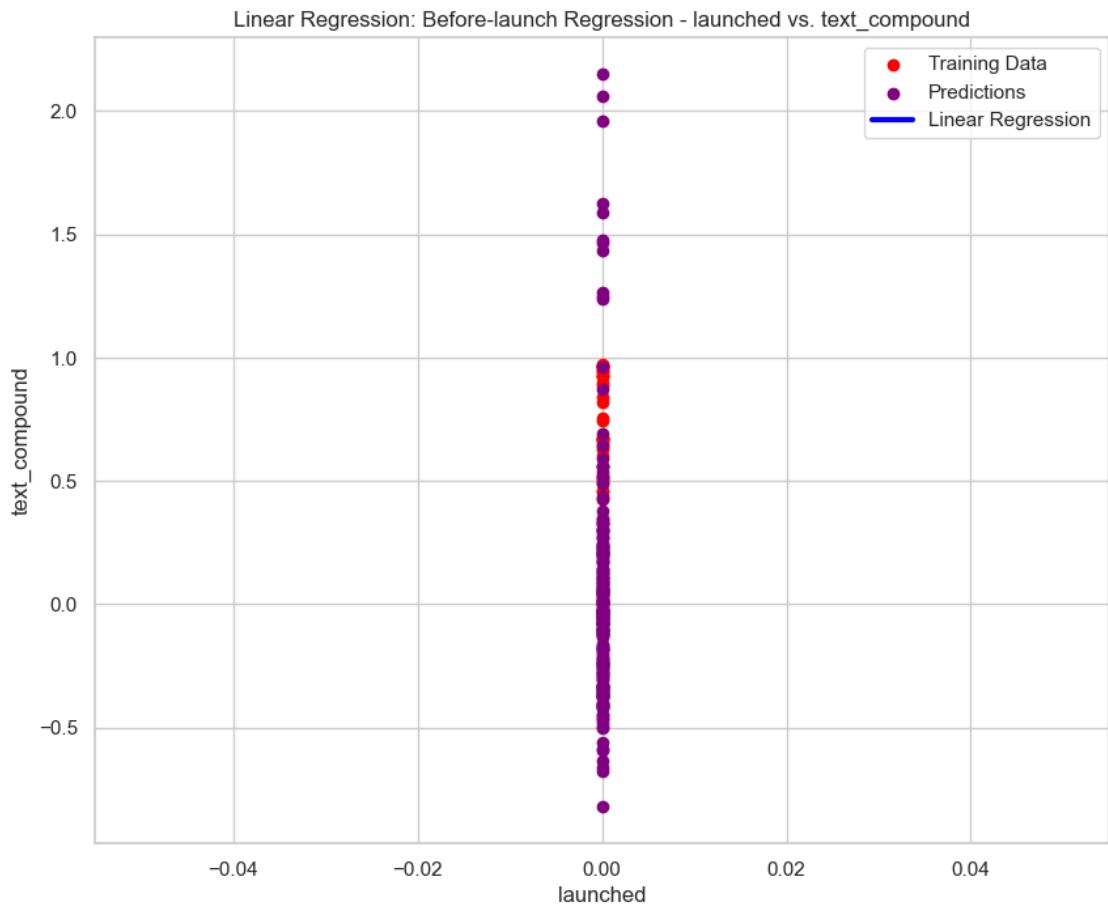


comment_compound_cumulative=0.003933989457936083

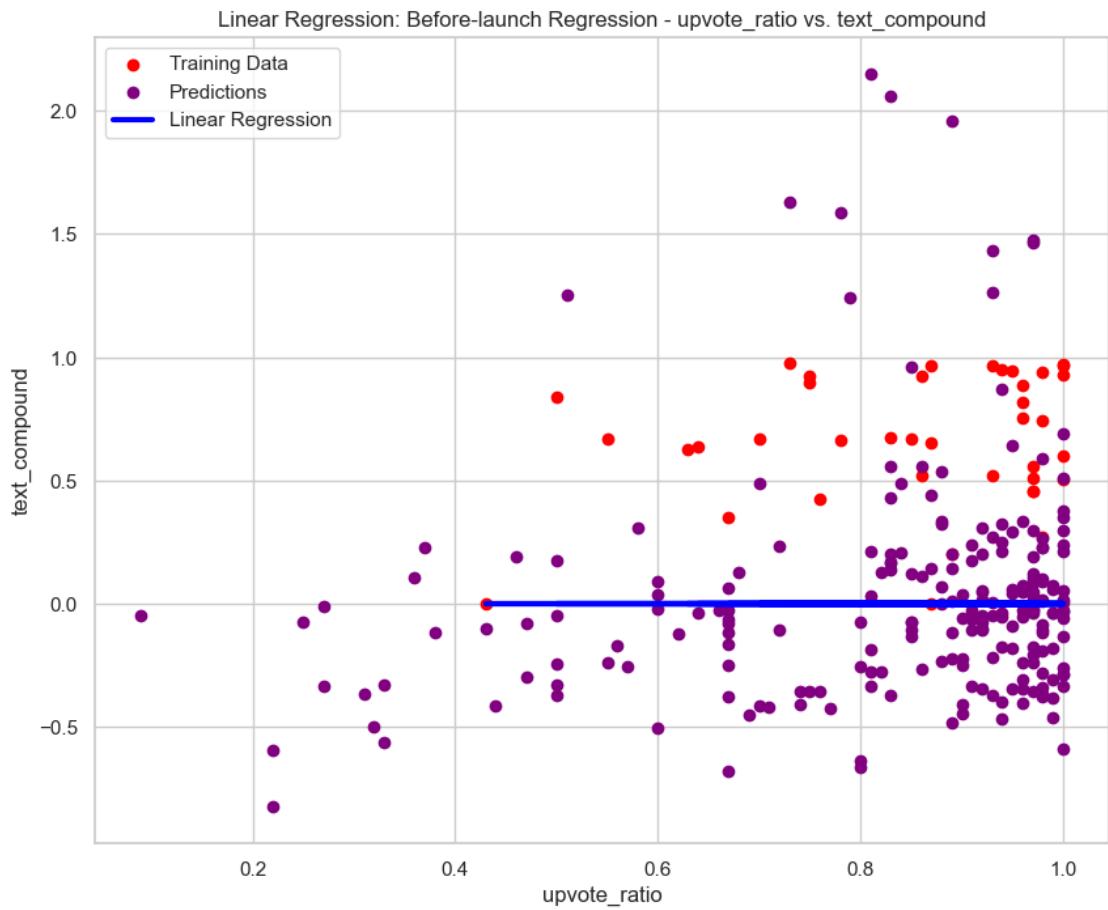
Linear Regression: GPT-4 Regression - comment_compound_cumulative_lagged vs. text_compound



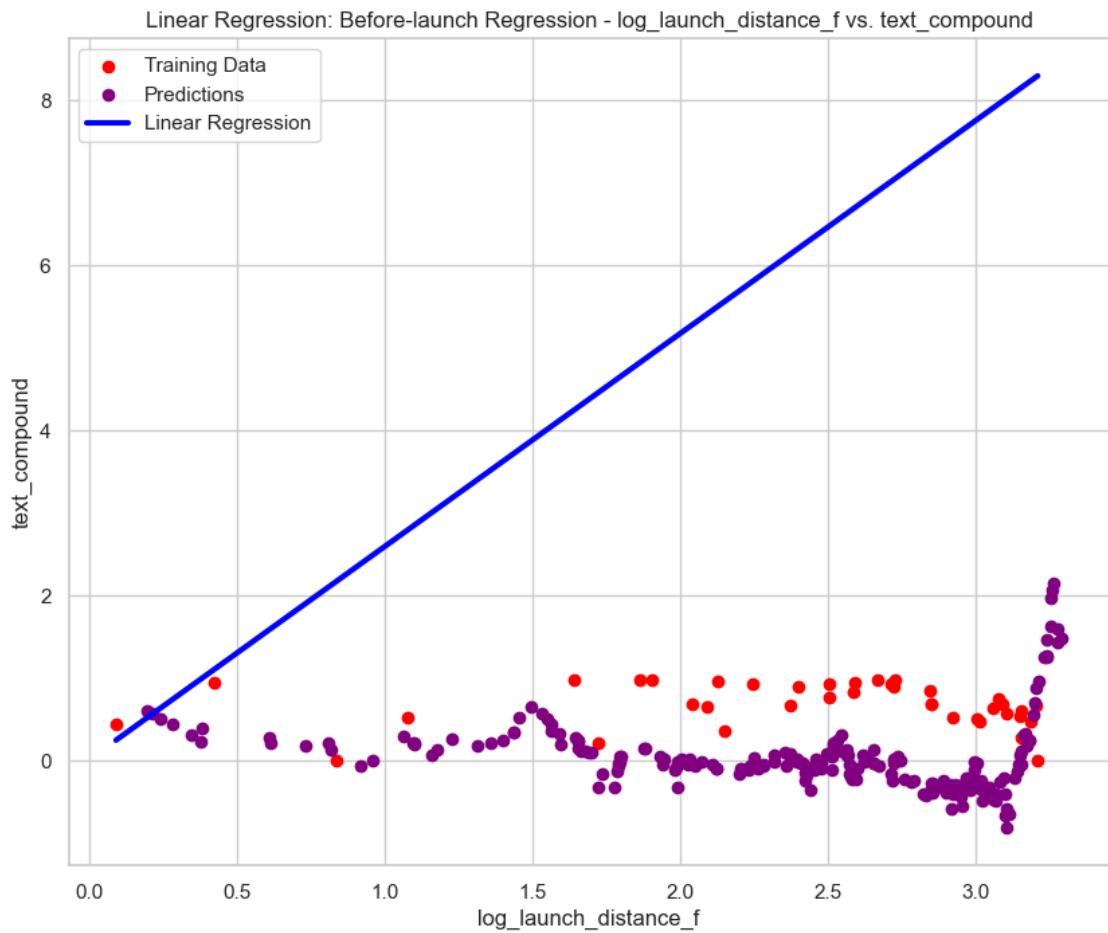
comment_compound_cumulative_lagged=-0.018431934782265146



launched=0.4397653973506887

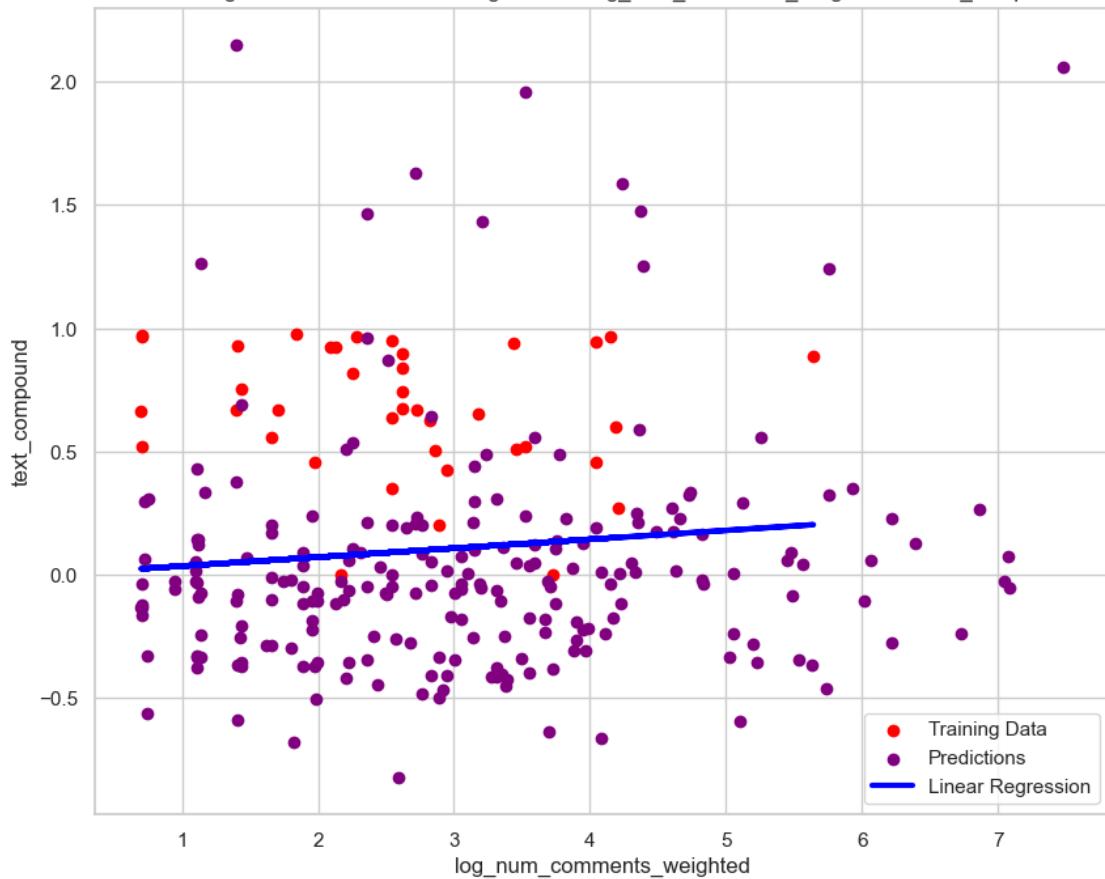


upvote_ratio=1.9626661407201595e-14

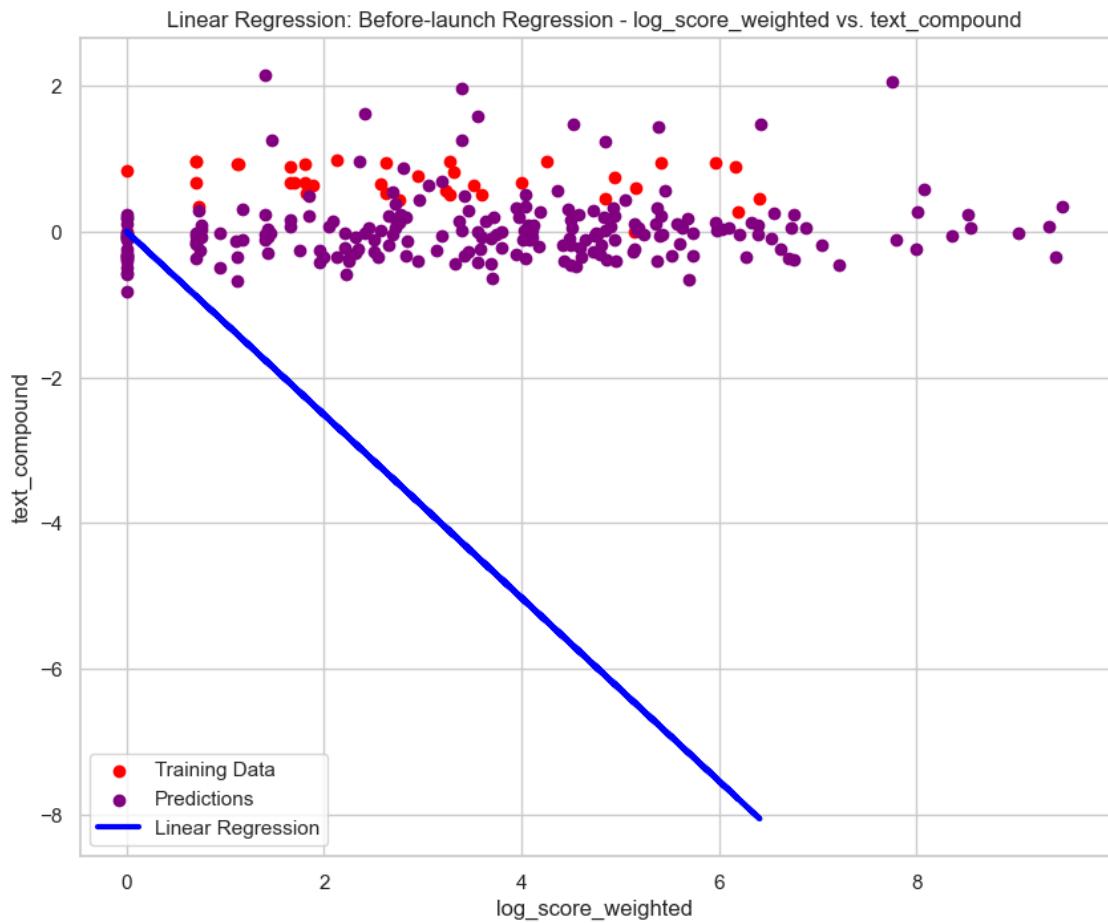


`log_launch_distance_f=2.5856984477923586`

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

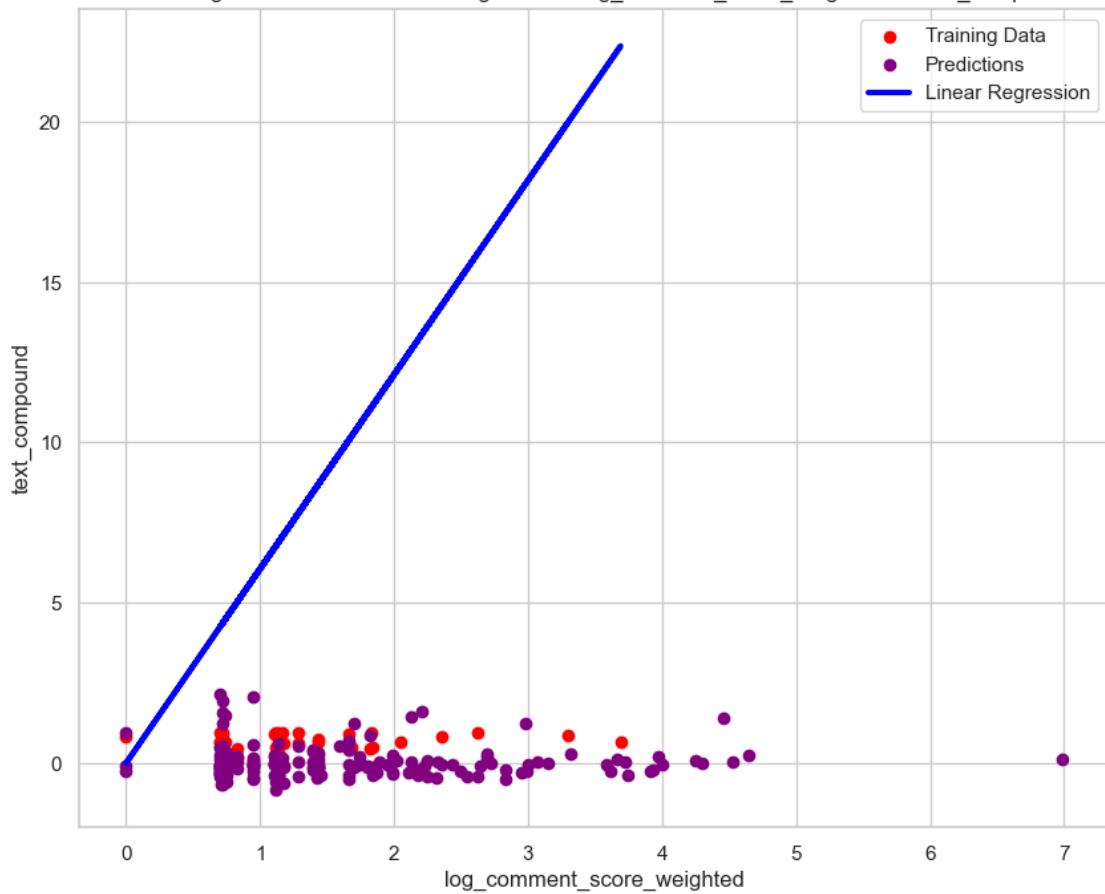


log_num_comments_weighted=0.03612687699553232

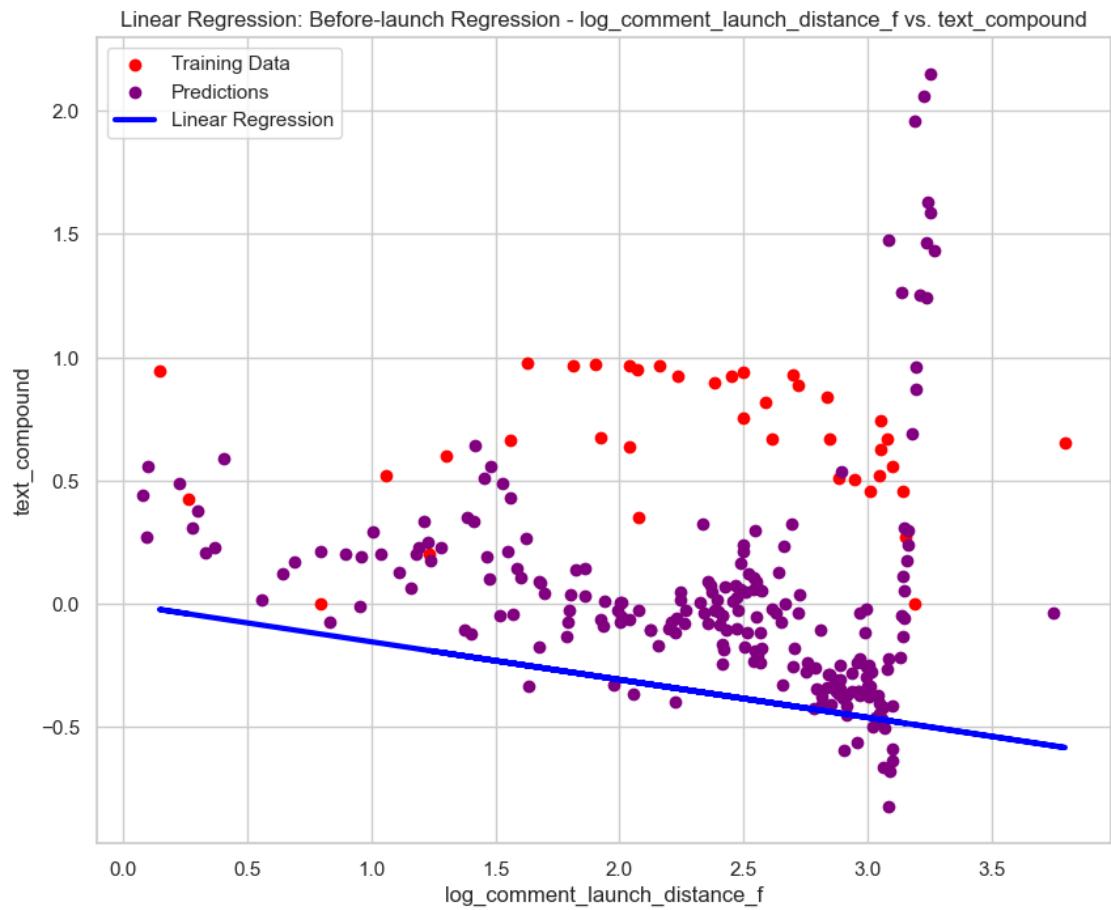


log_score_weighted=-1.257073338244043

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

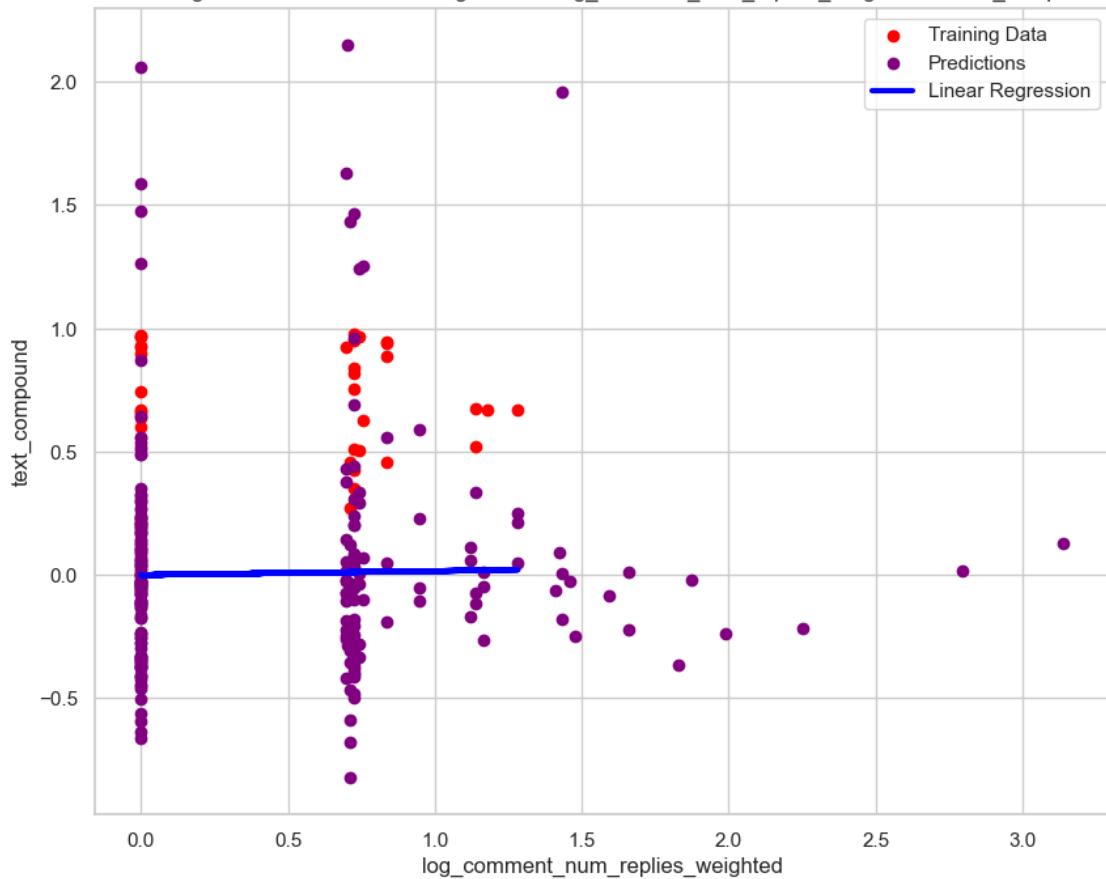


log_comment_score_weighted=6.054710137043305

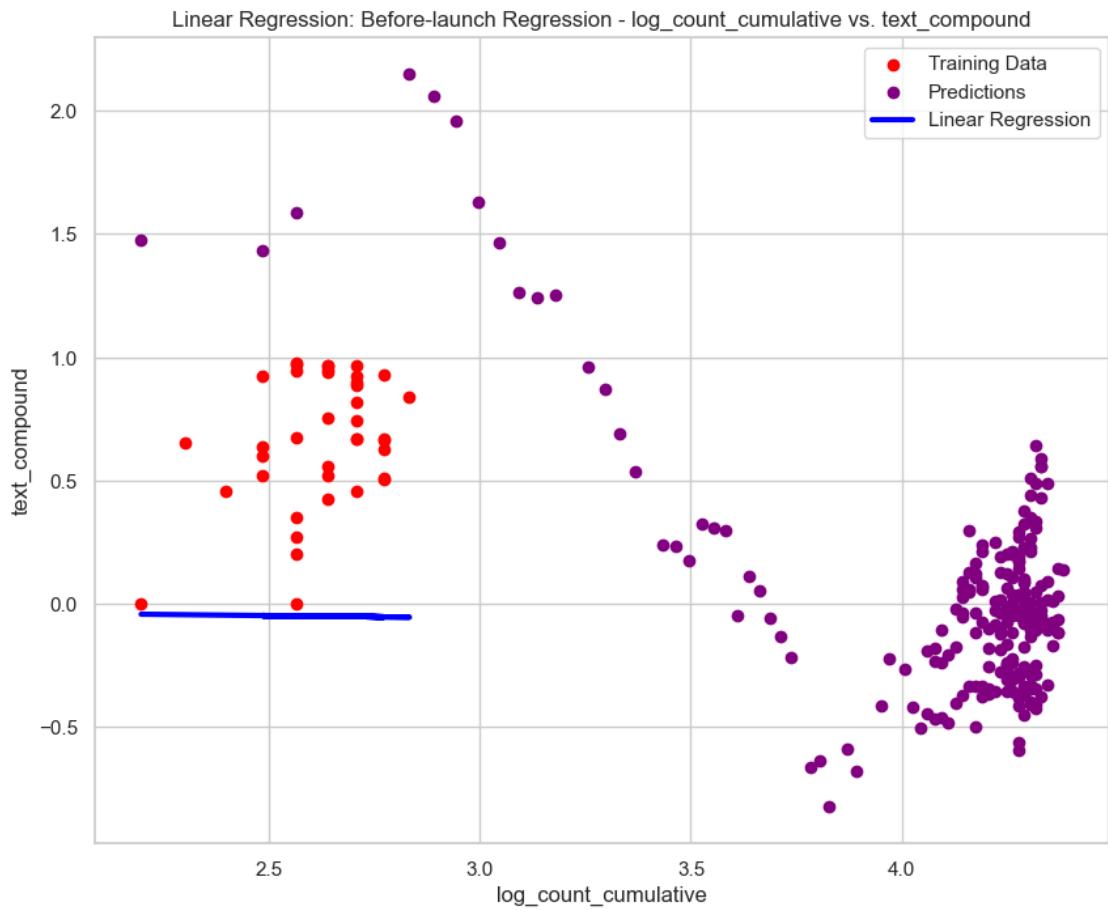


`log_comment_launch_distance_f=-0.15343994928527524`

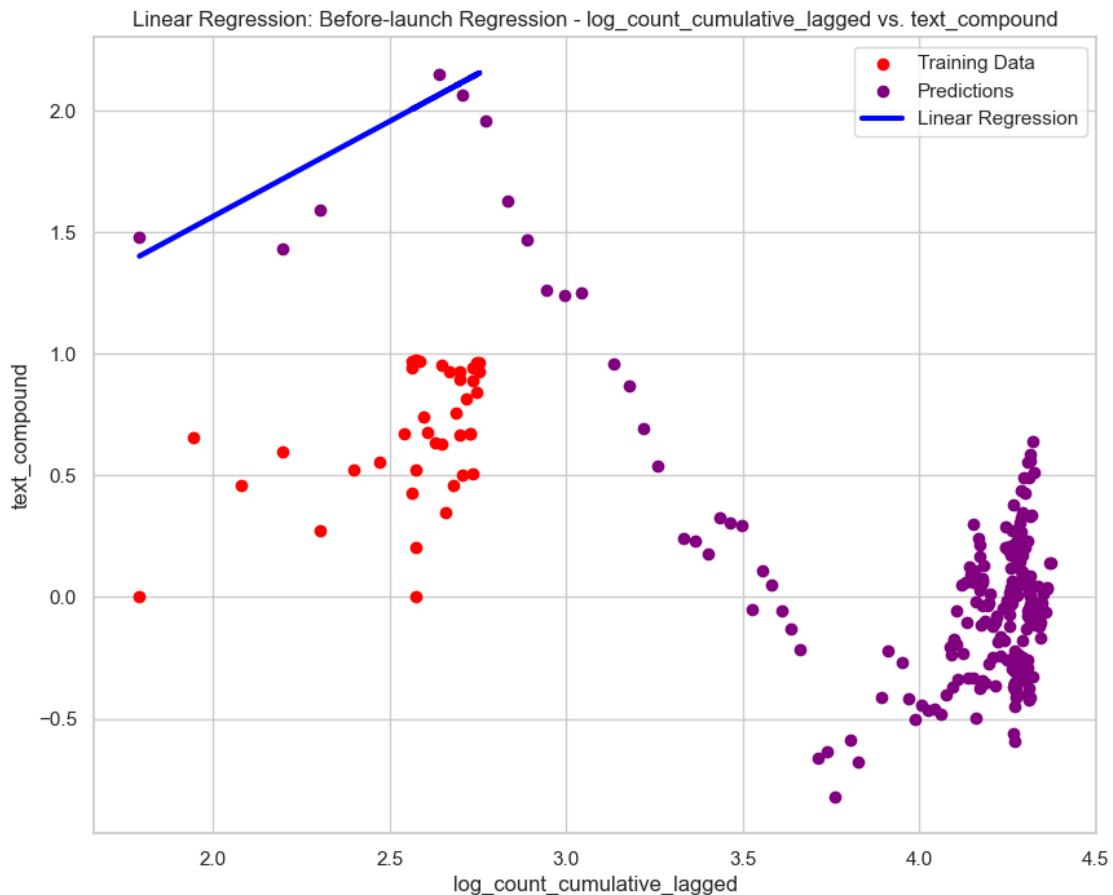
Linear Regression: Before-launch Regression - log_comment_num_replies_weighted vs. text_compound



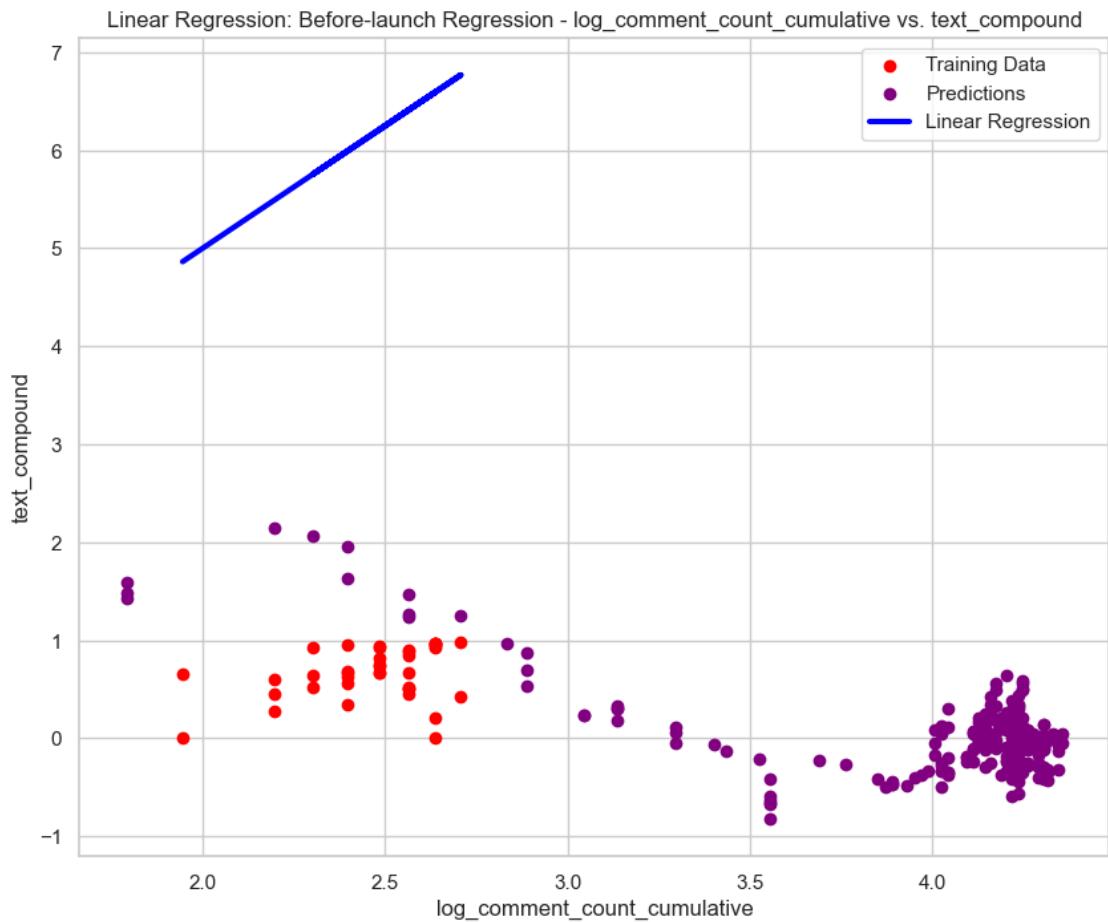
log_comment_num_replies_weighted=0.015980590335701712



`log_count_cumulative=-0.018934099486996116`

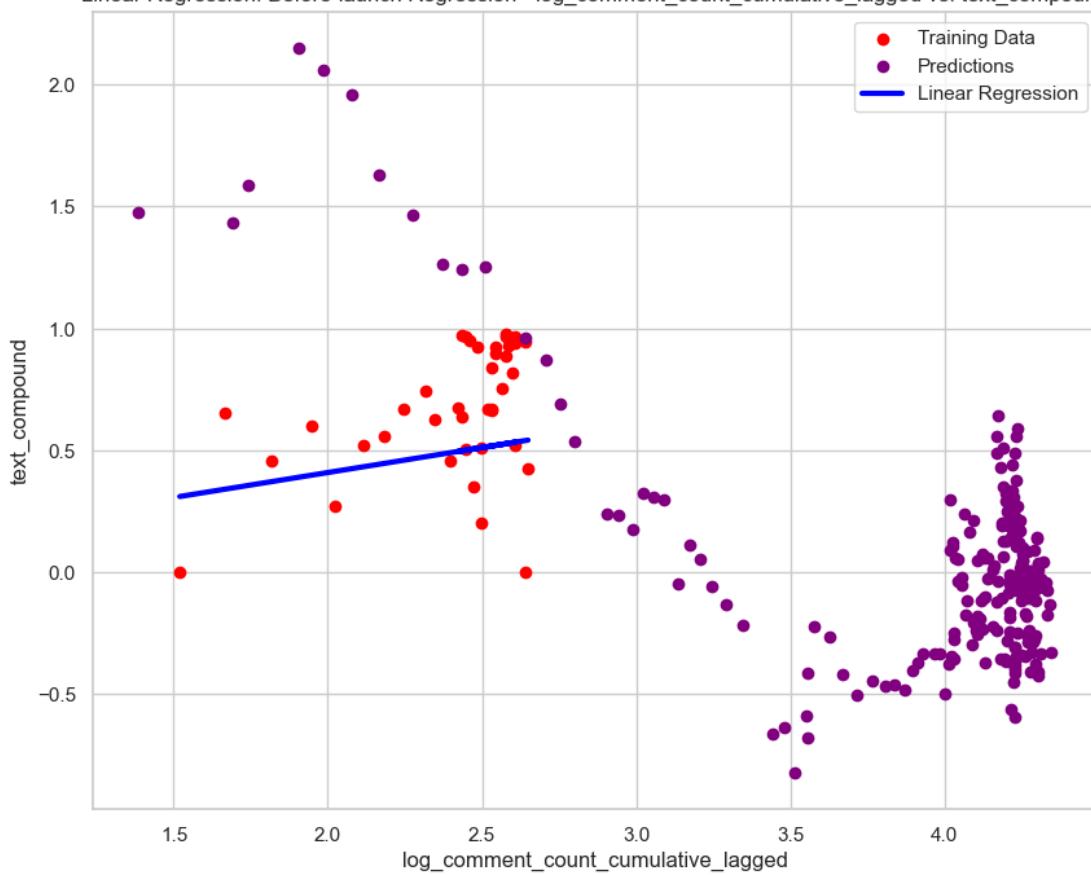


log_count_cumulative_lagged=0.7821090882097406

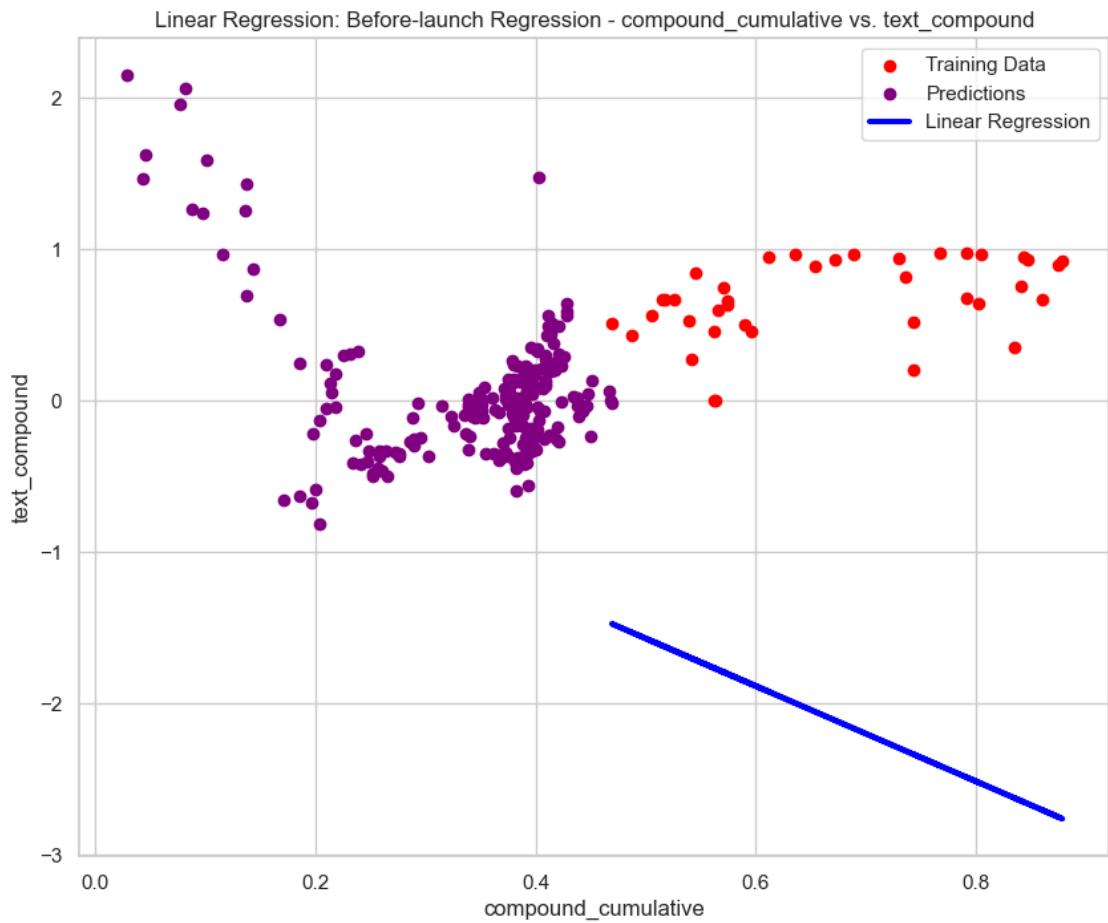


log_comment_count_cumulative=2.4989714440362287

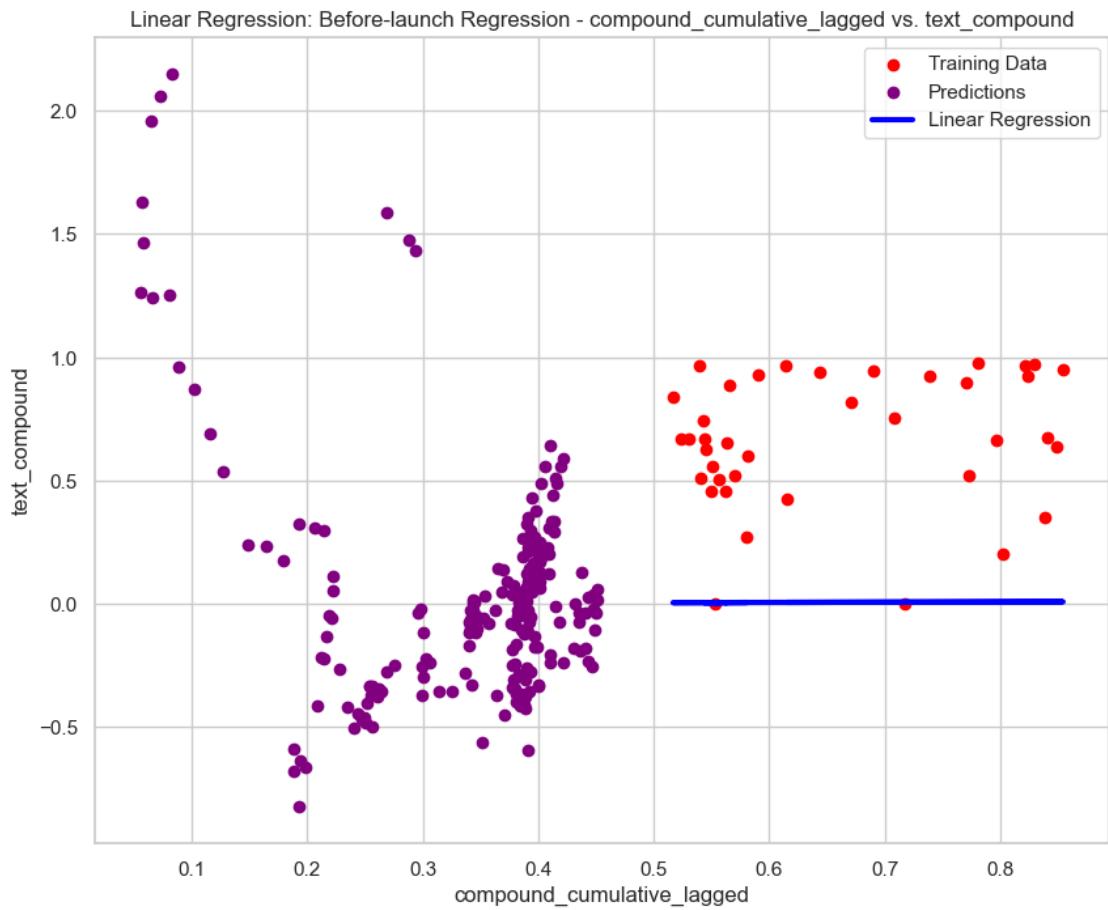
Linear Regression: Before-launch Regression - log_comment_count_cumulative_lagged vs. text_compound



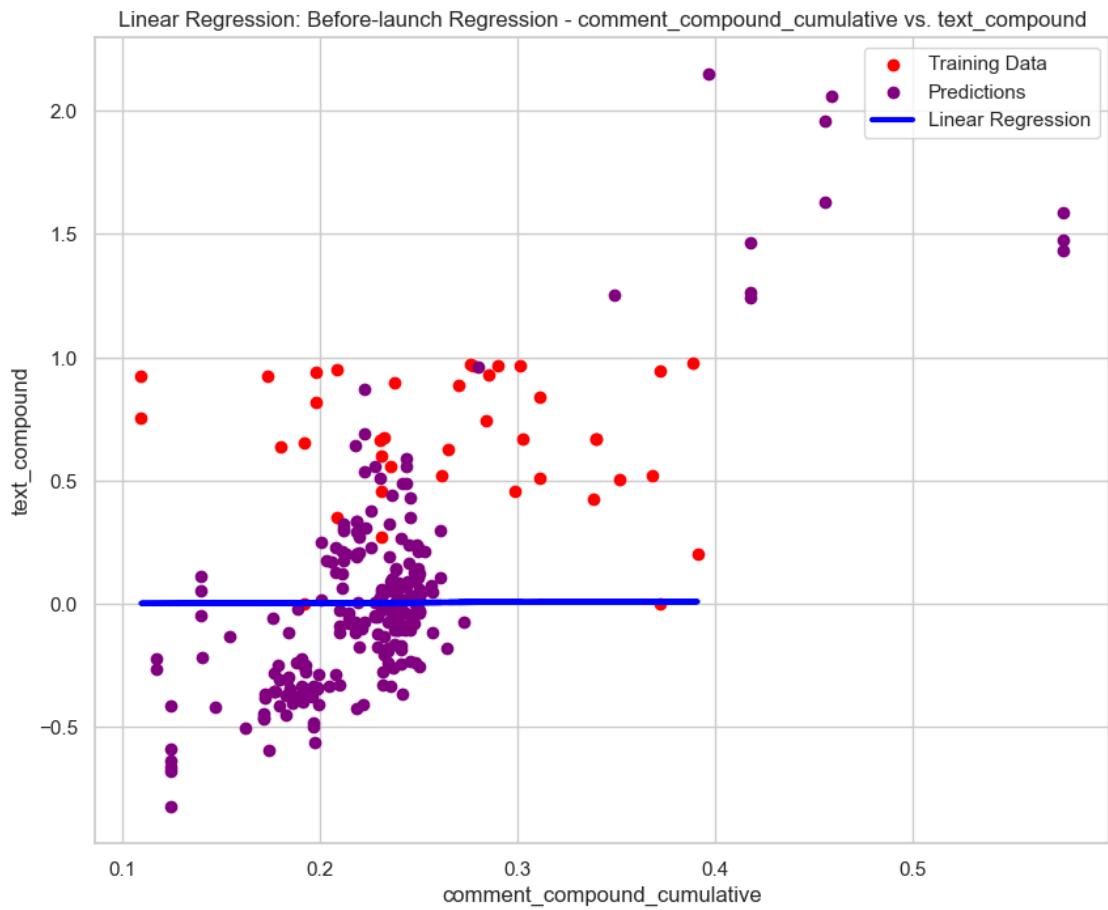
log_comment_count_cumulative_lagged=0.2046216036151977



compound_cumulative=-3.1452136748772848

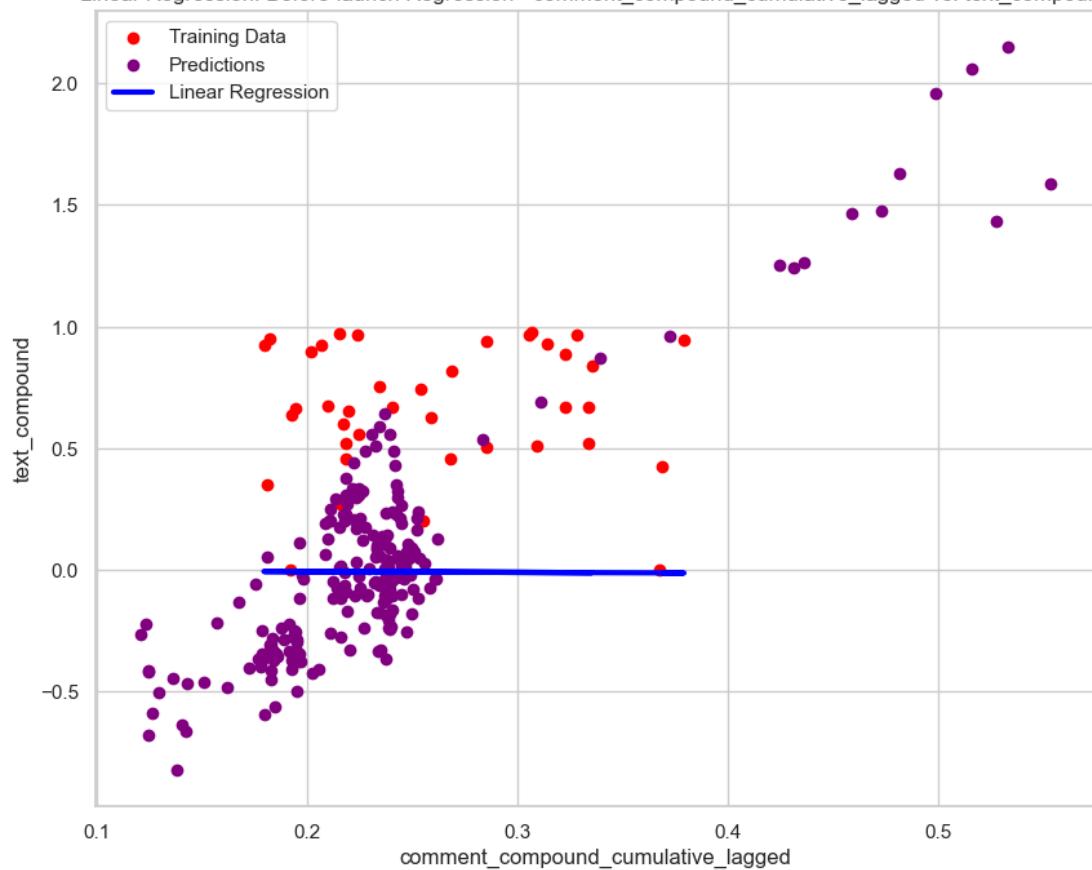


compound_cumulative_lagged=0.010201292425556978

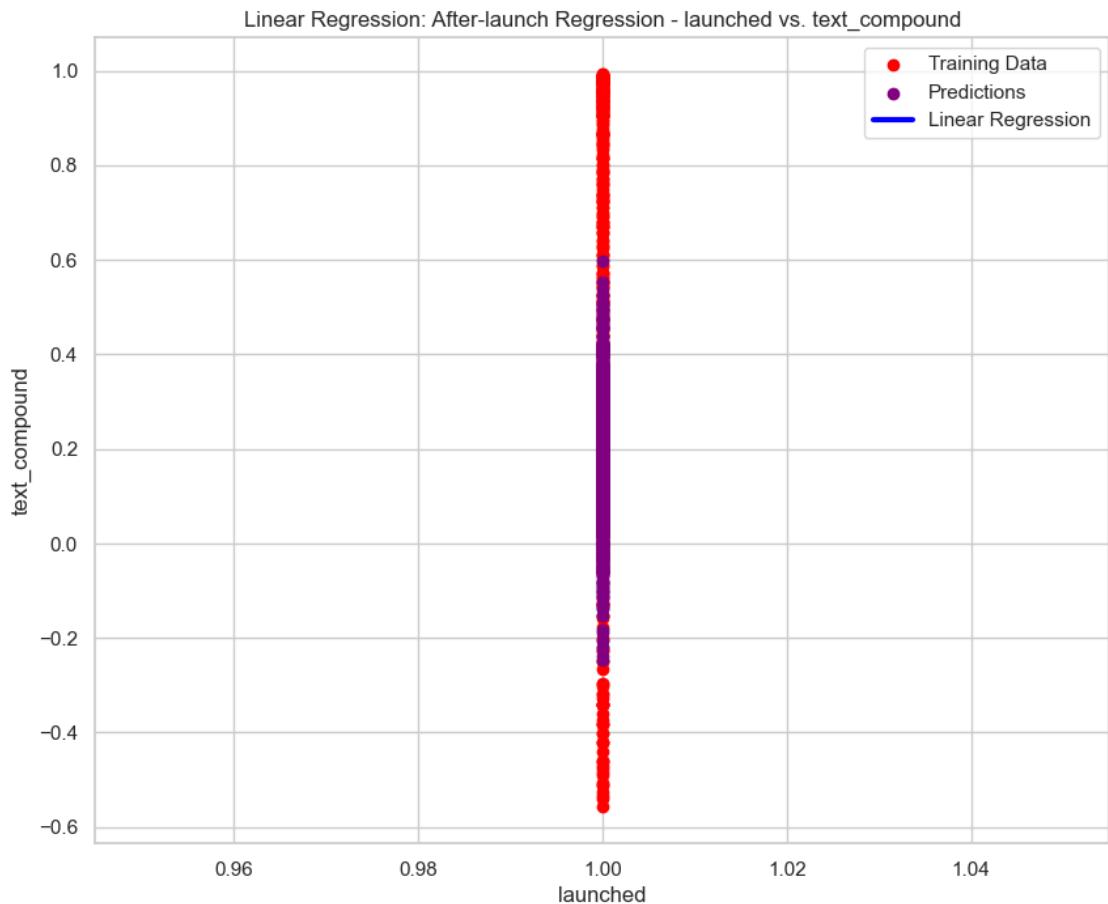


comment_compound_cumulative=0.02395088648333308

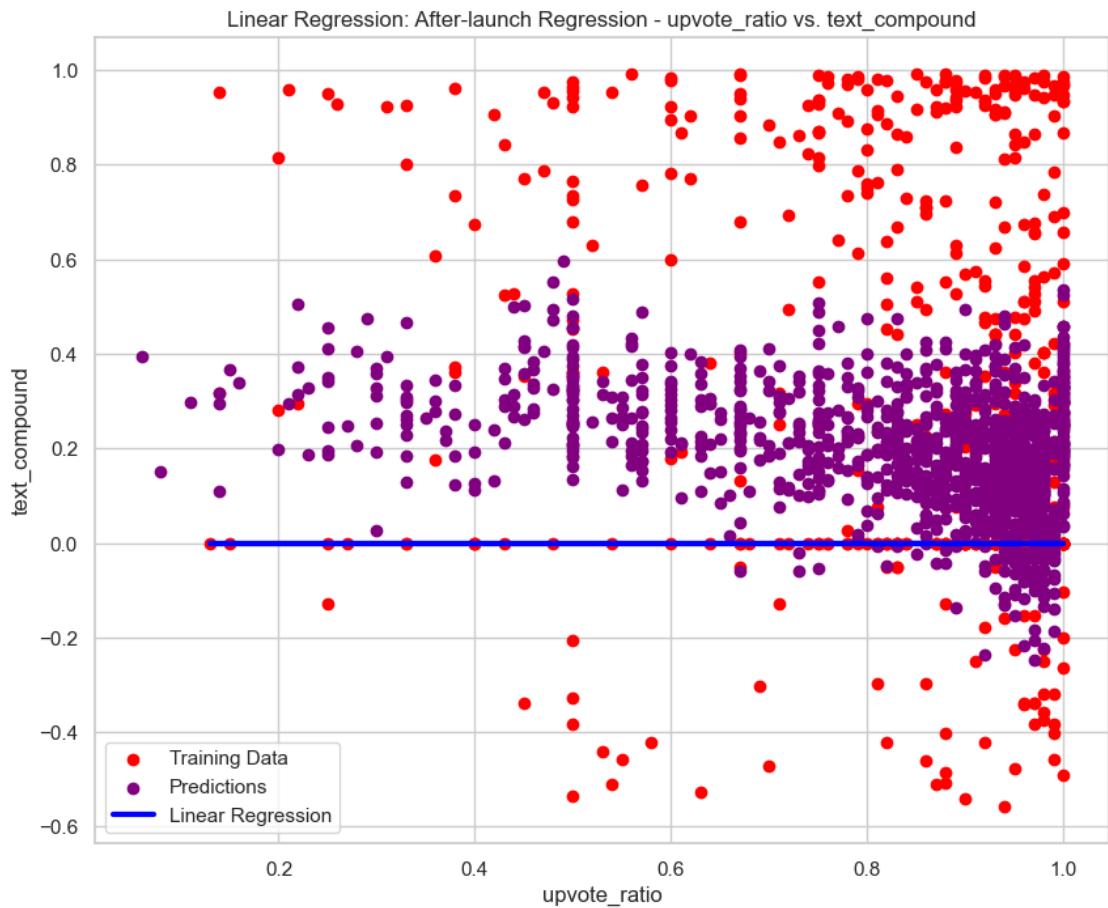
Linear Regression: Before-launch Regression - comment_compound_cumulative_lagged vs. text_compound



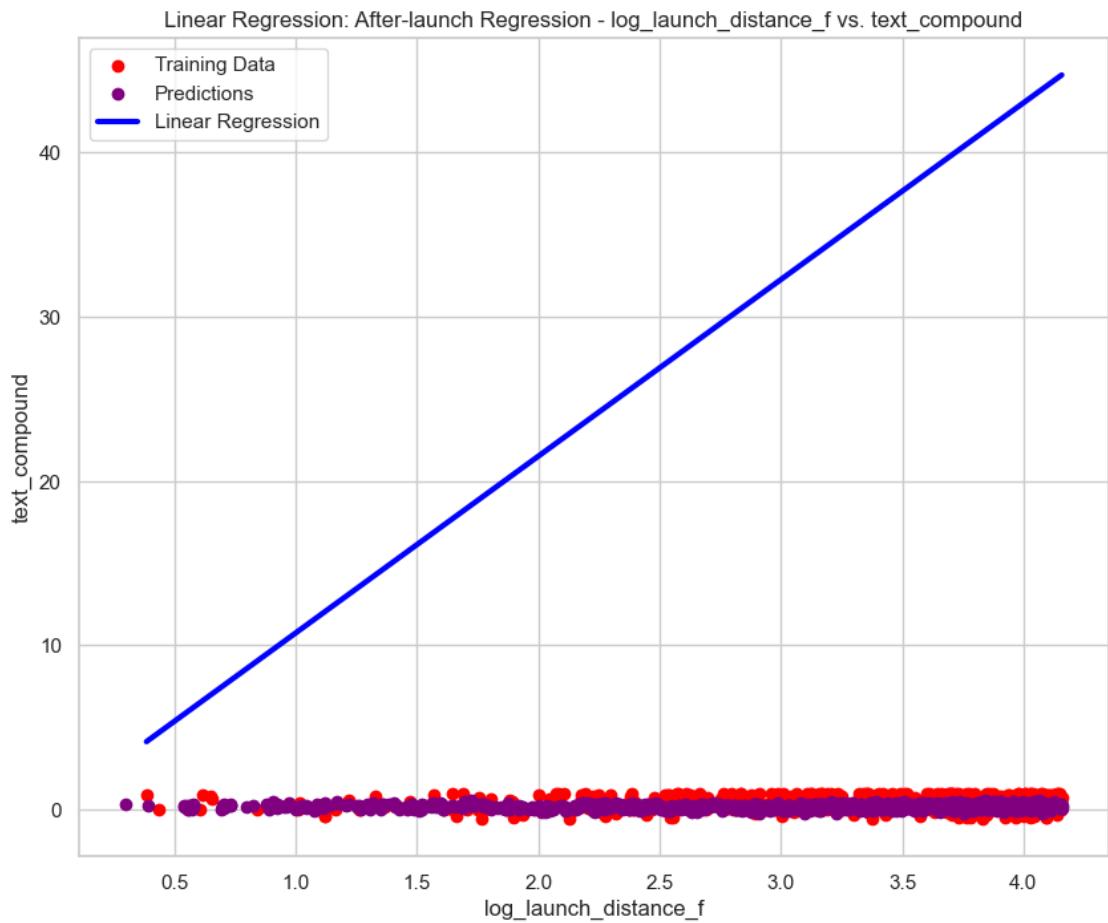
comment_compound_cumulative_lagged=-0.031650710097004604



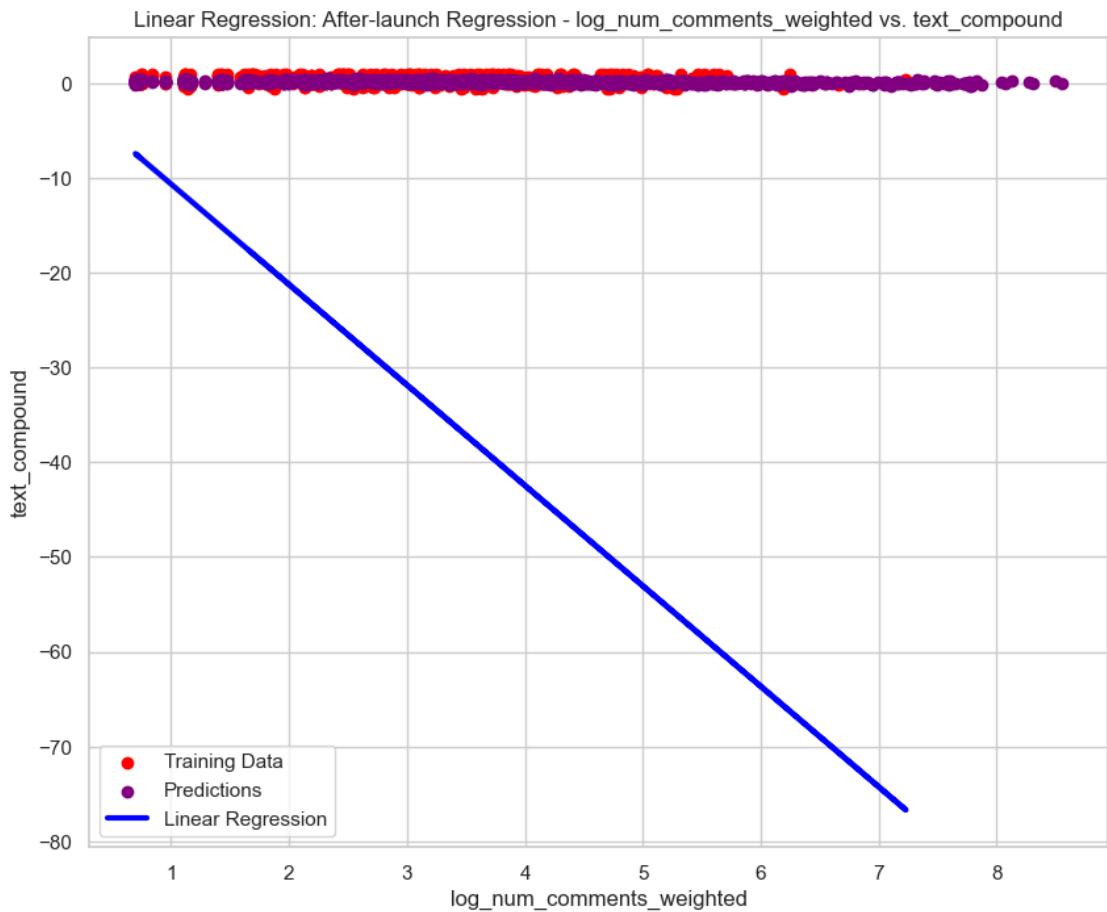
launched=0.14915135209219882



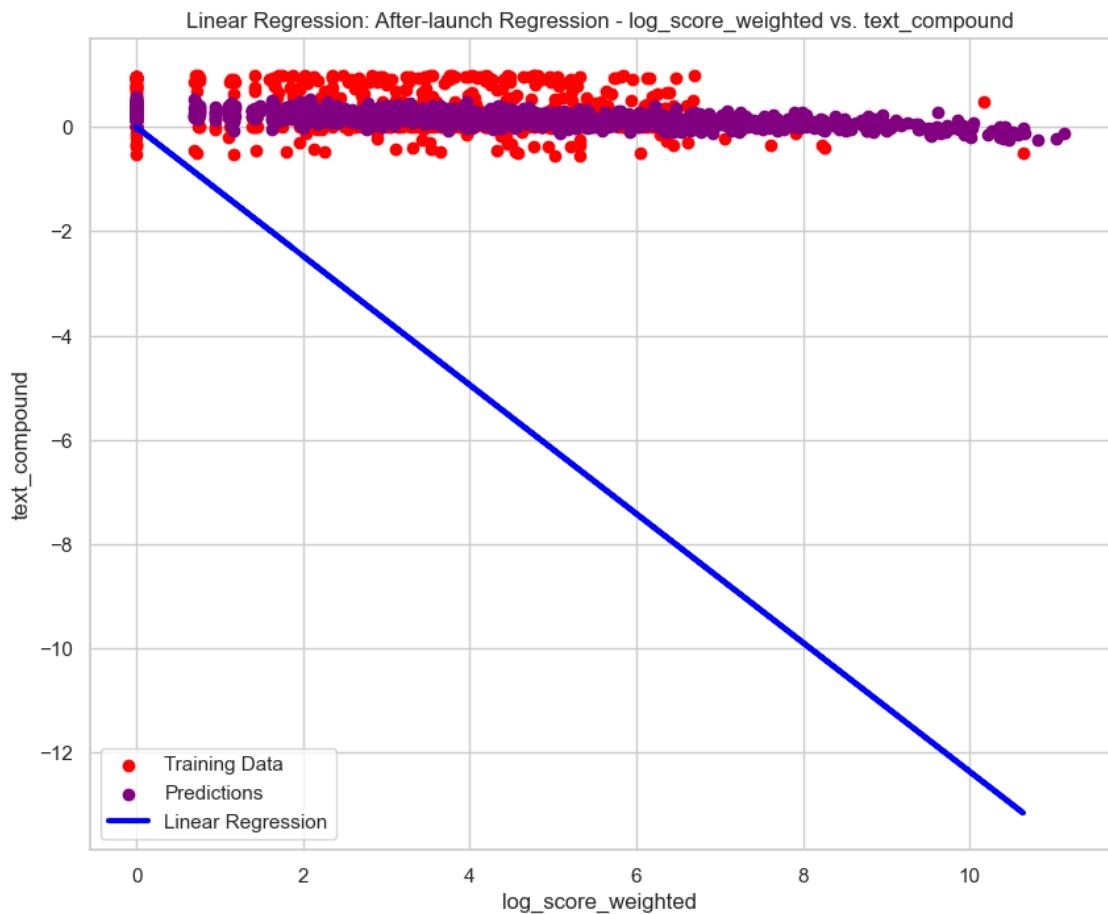
upvote_ratio=6.723788192886104e-15



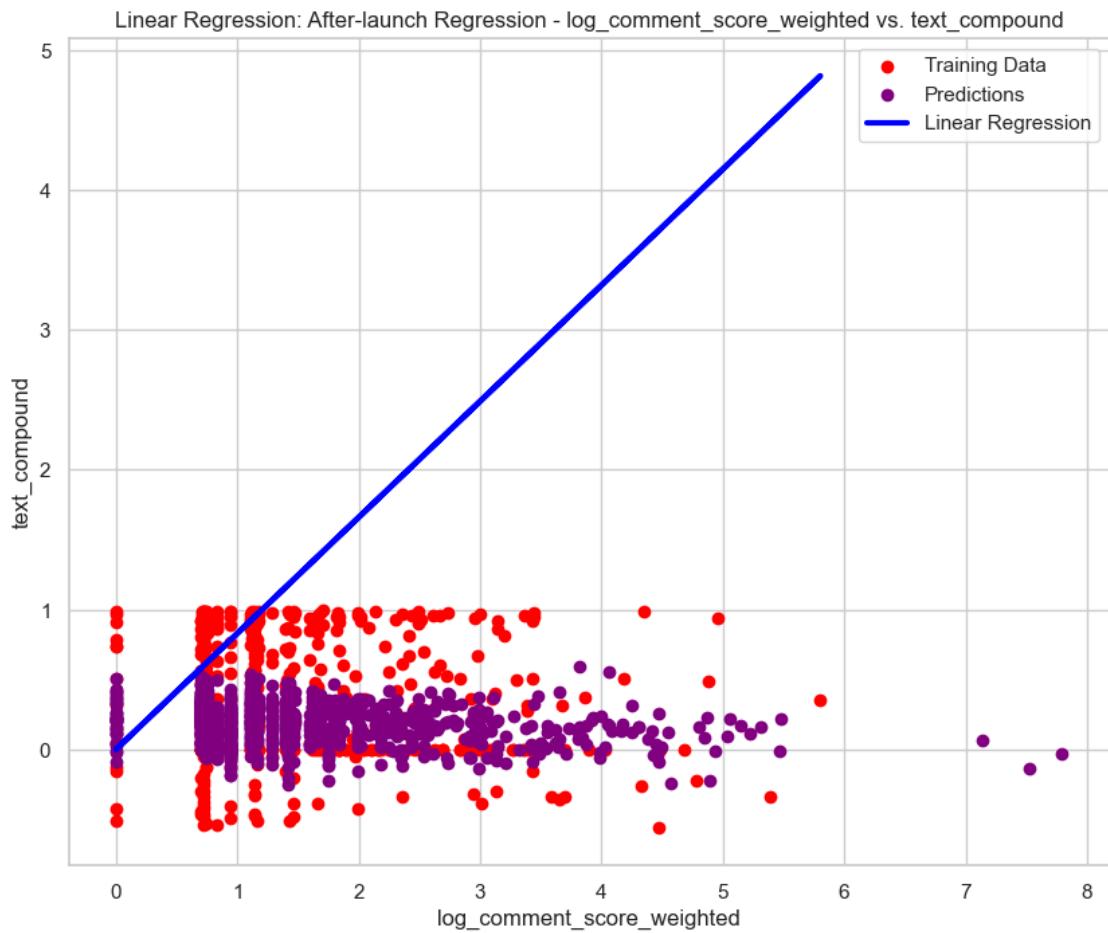
log_launch_distance_f=10.758769556375263



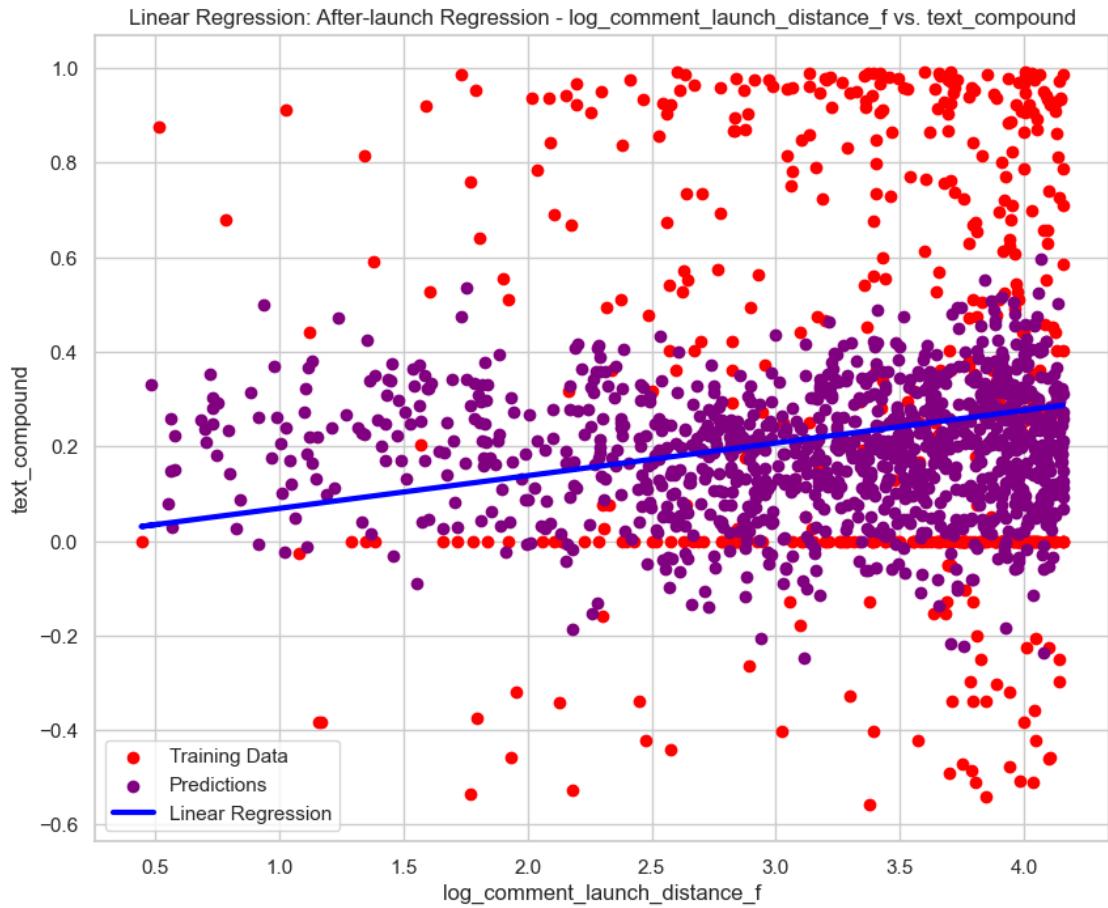
log_num_comments_weighted=-10.609254032298887



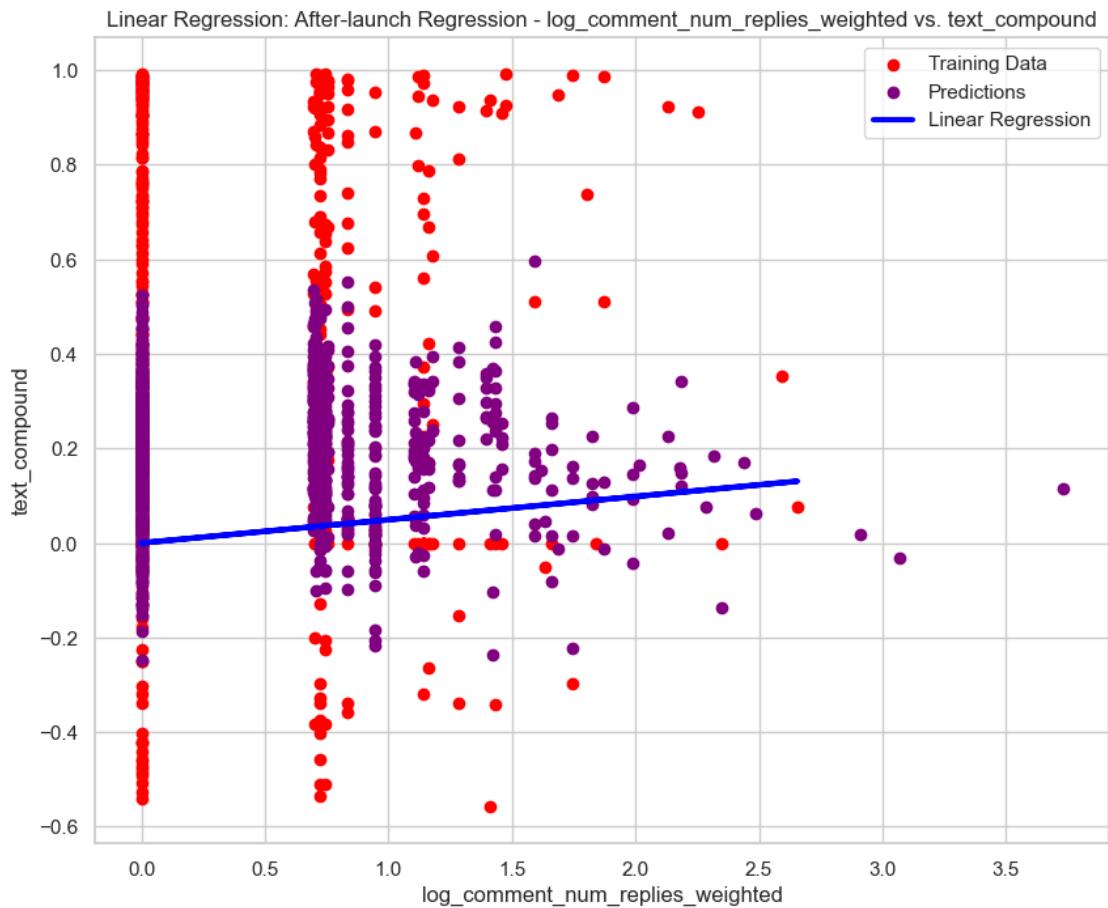
log_score_weighted=-1.2366378798318873



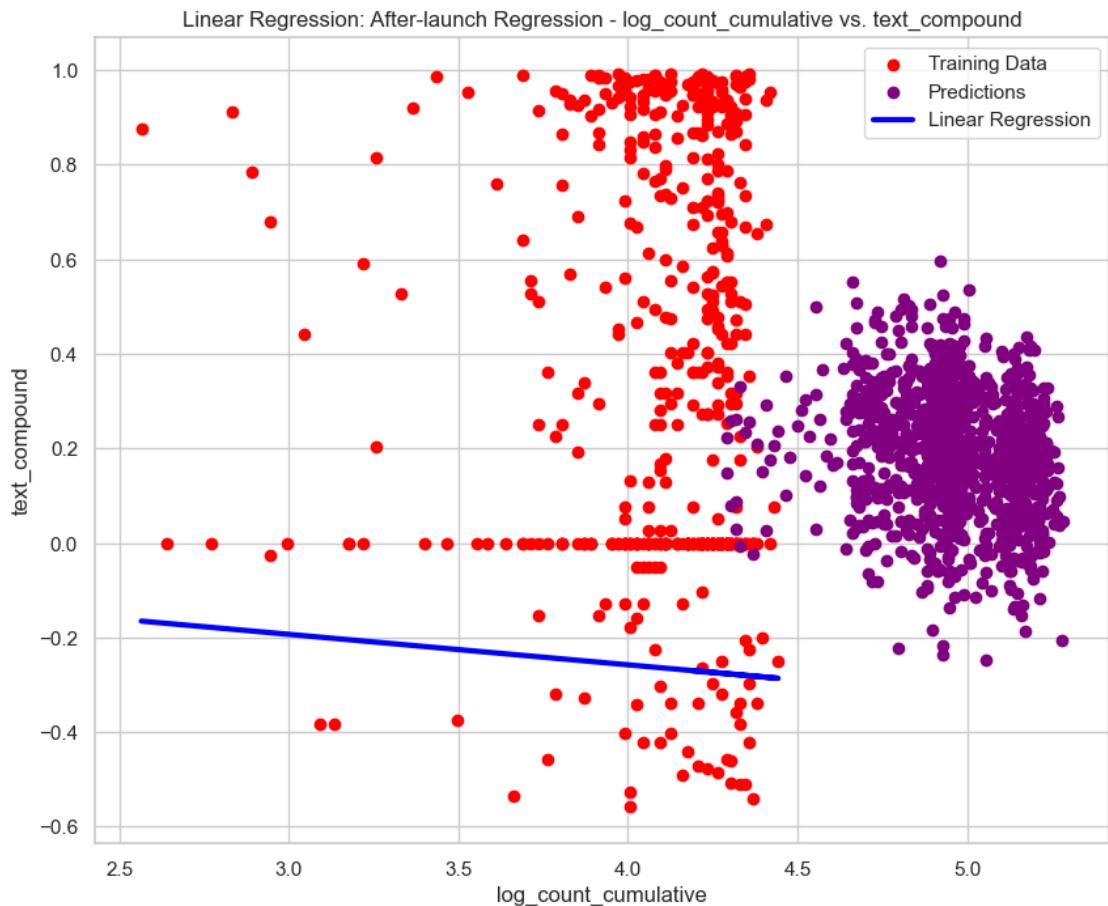
log_comment_score_weighted=0.8298221717035702



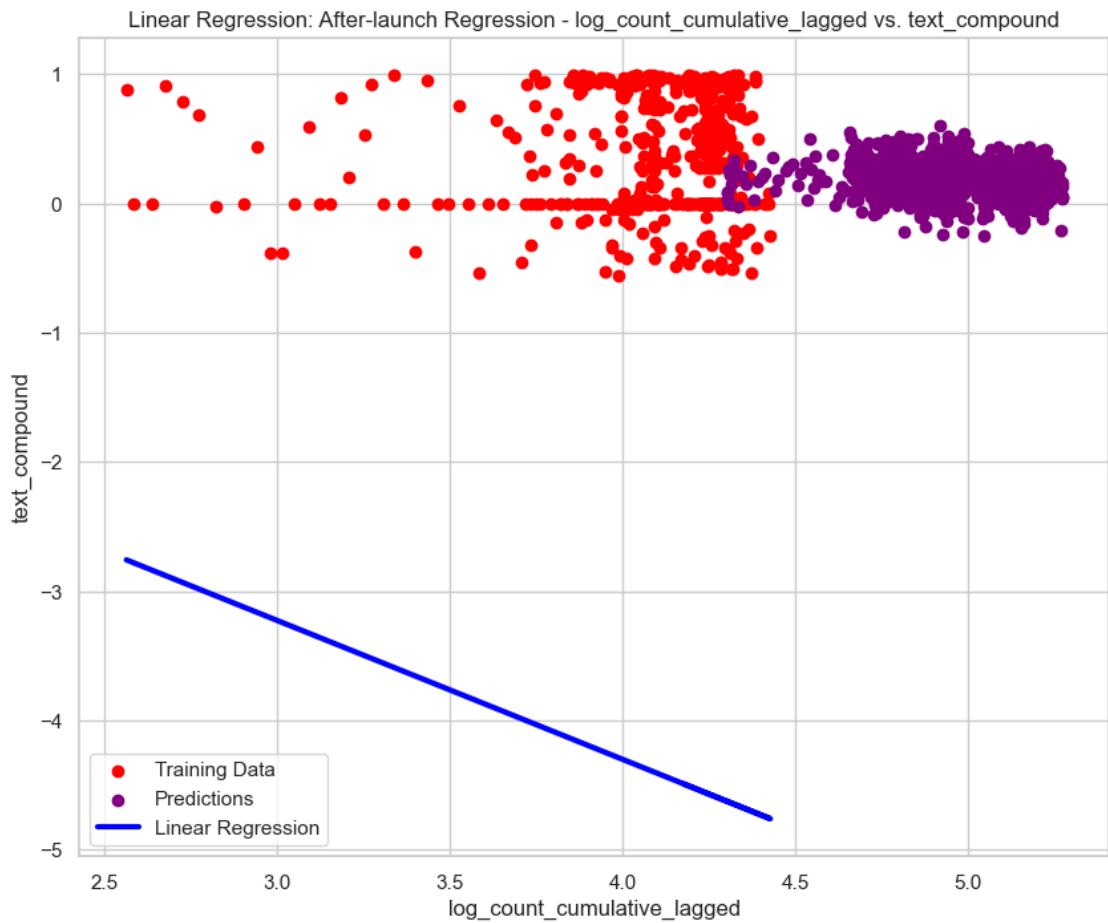
log_comment_launch_distance_f=0.0690578734854035



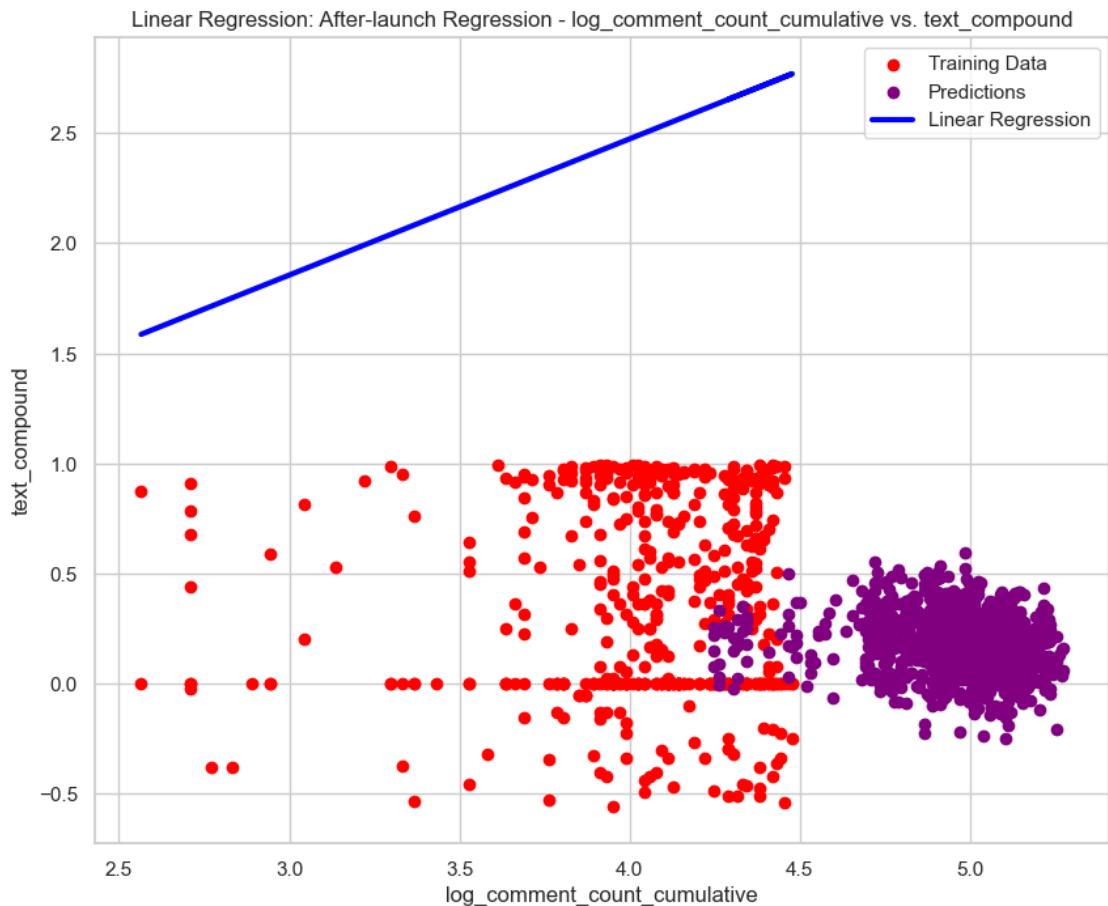
log_comment_num_replies_weighted=0.04917109878904852



log_count_cumulative=-0.06444035711810703



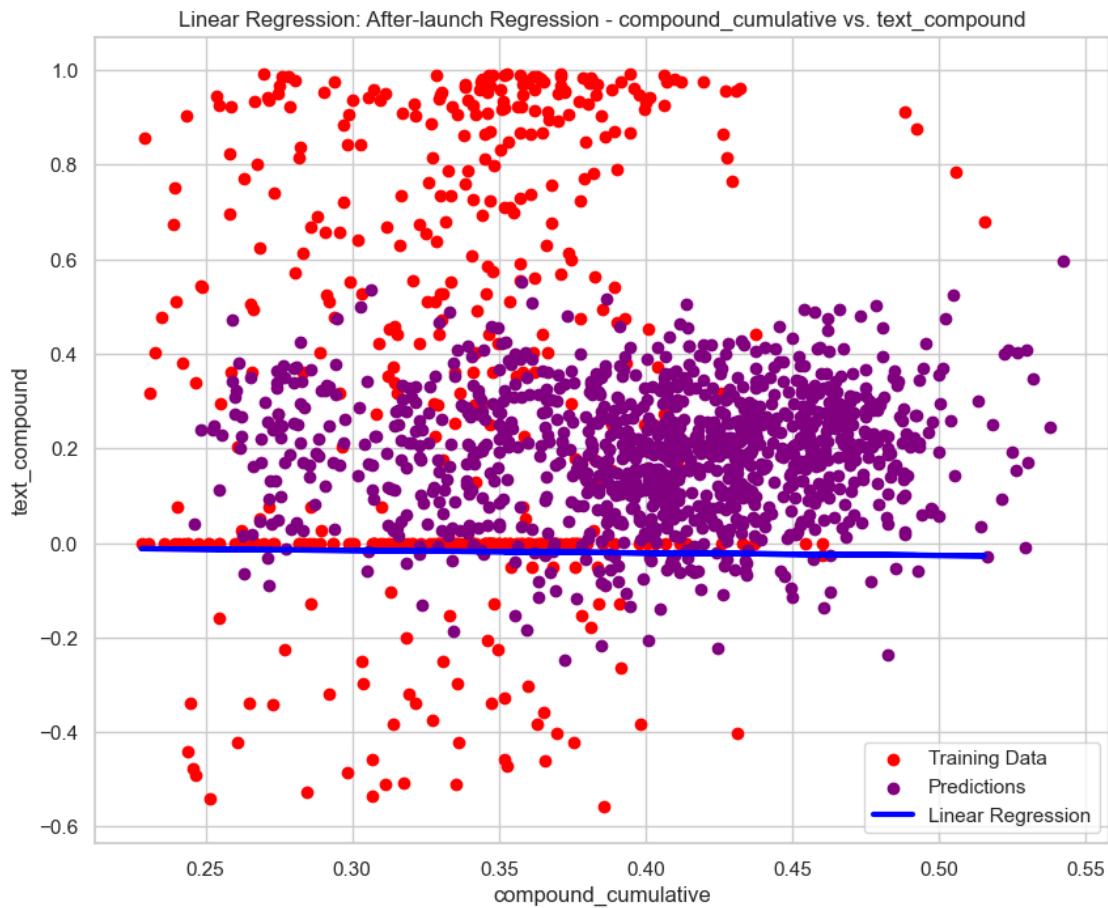
log_count_cumulative_lagged=-1.0751365259488168



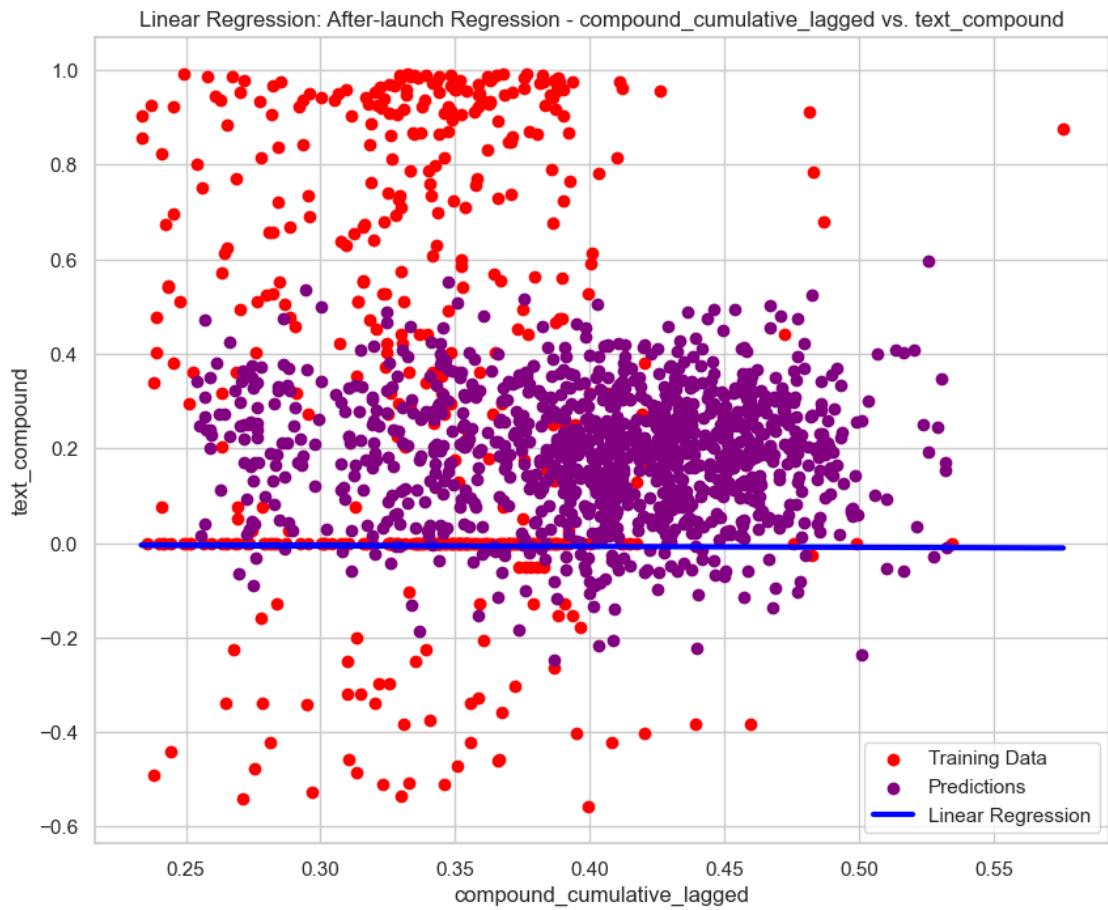
log_comment_count_cumulative=0.6184931524144314



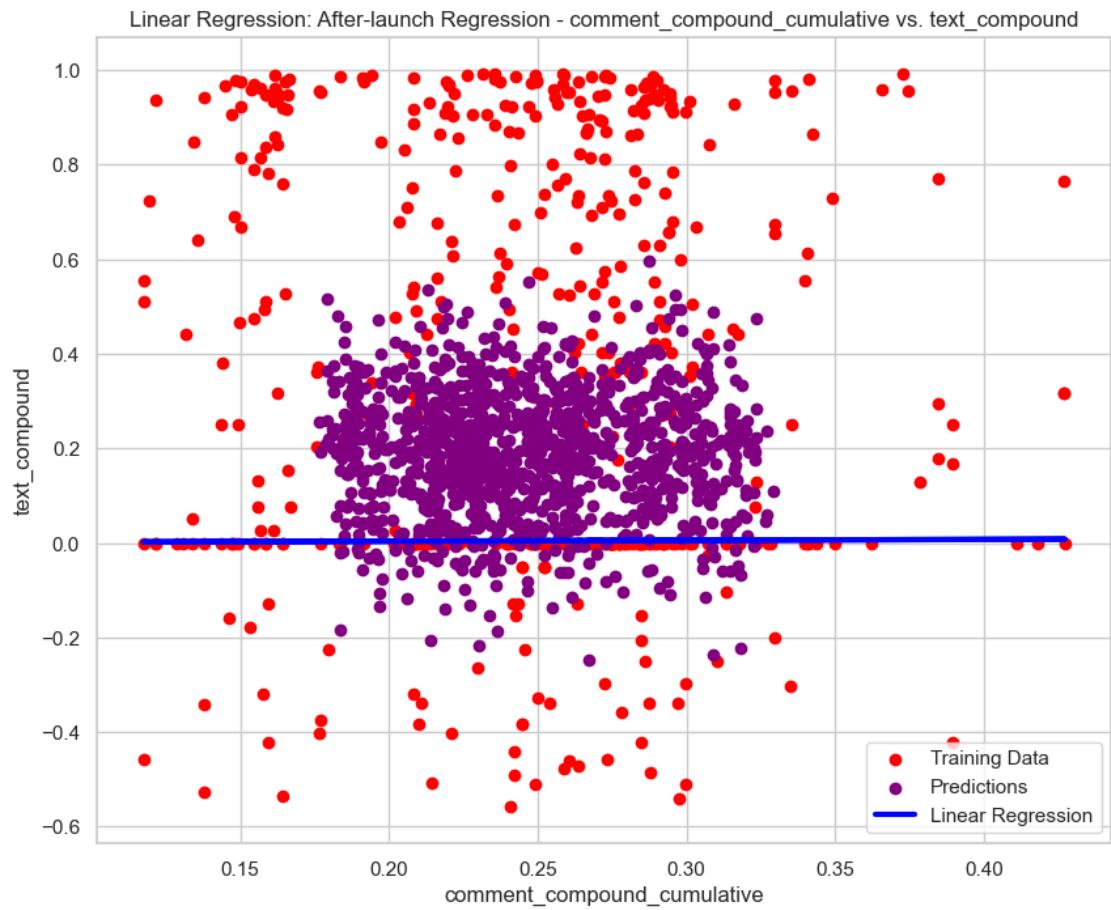
log_comment_count_cumulative_lagged=0.26993273336885415



compound_cumulative=-0.05273122559988909

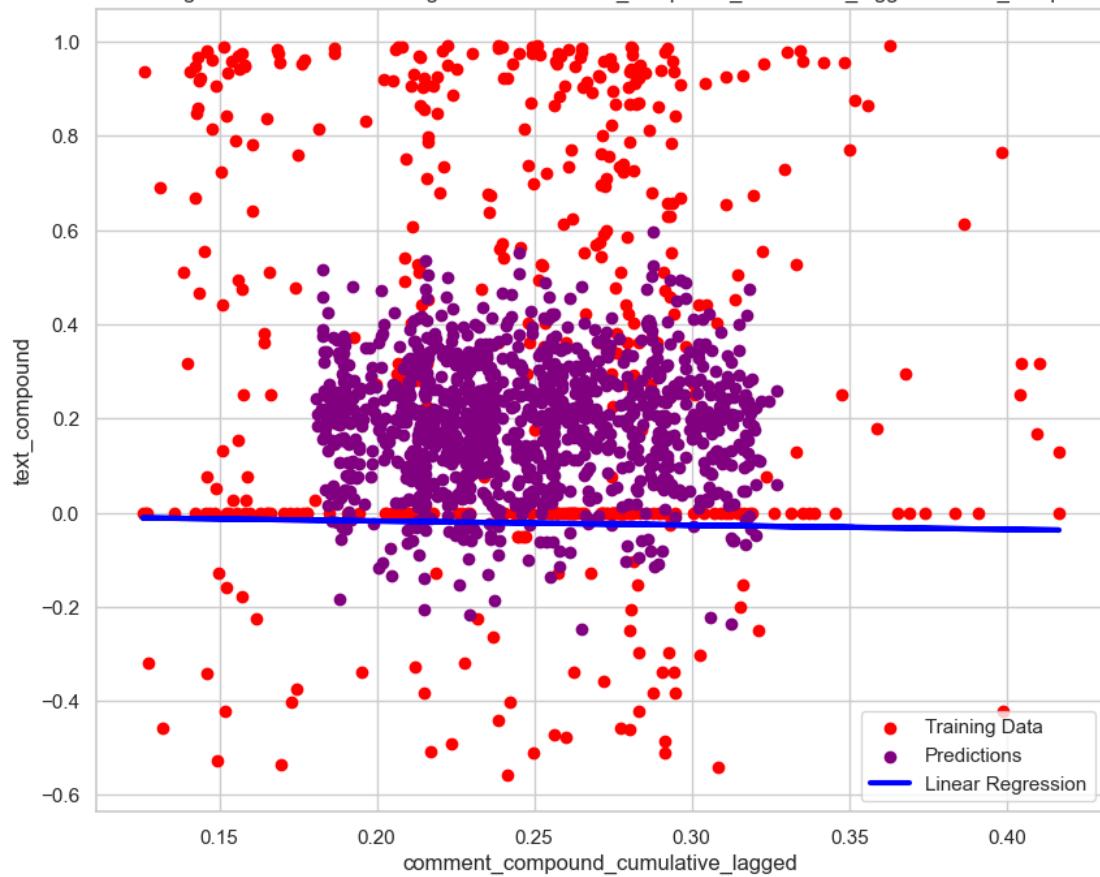


compound_cumulative_lagged=-0.01843616096827612

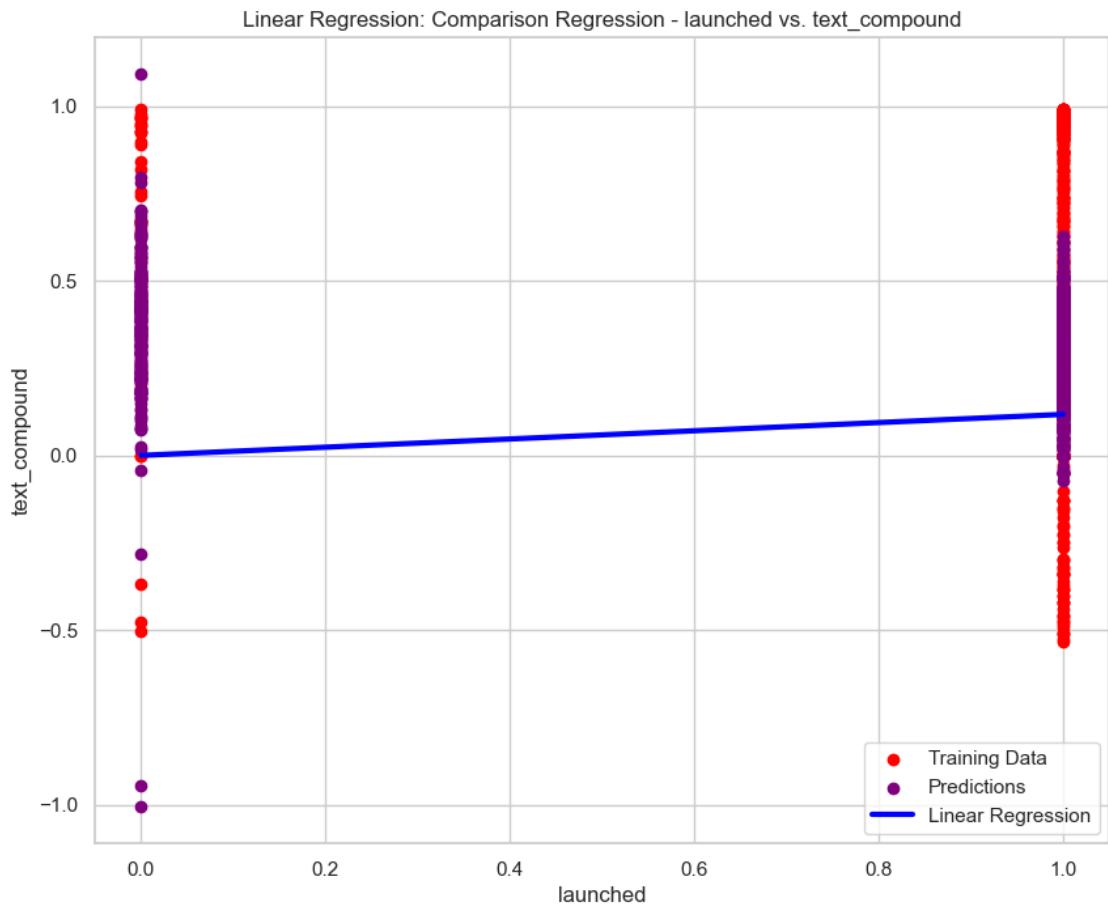


comment_compound_cumulative=0.018902675290534396

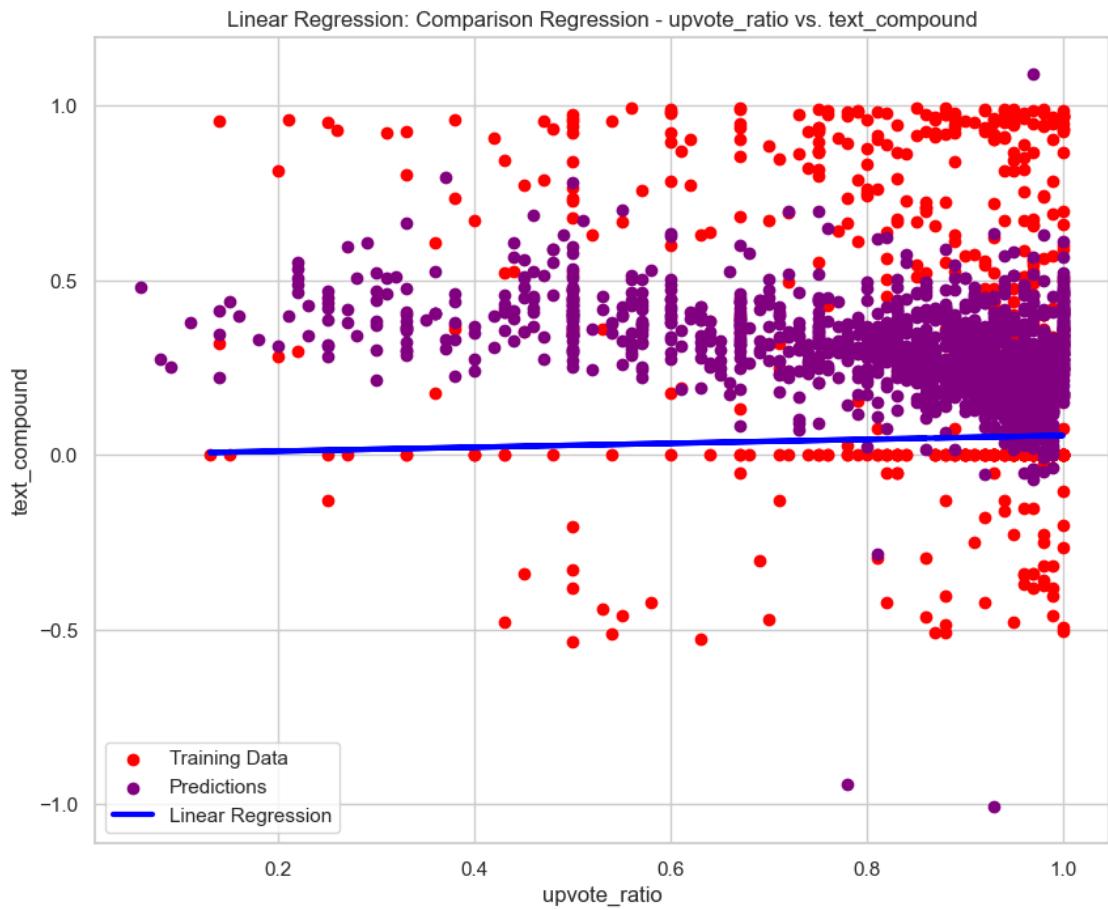
Linear Regression: After-launch Regression - comment_compound_cumulative_lagged vs. text_compound



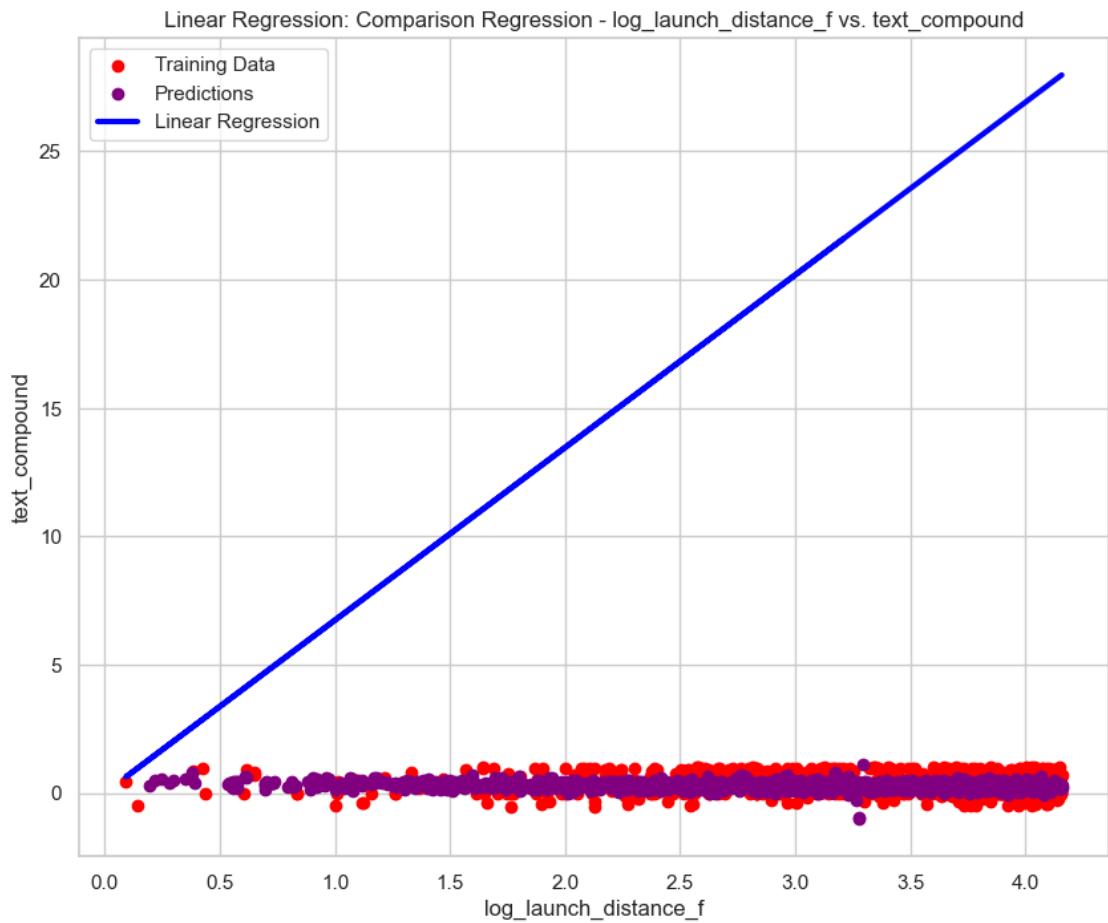
comment_compound_cumulative_lagged=-0.08909466758717874



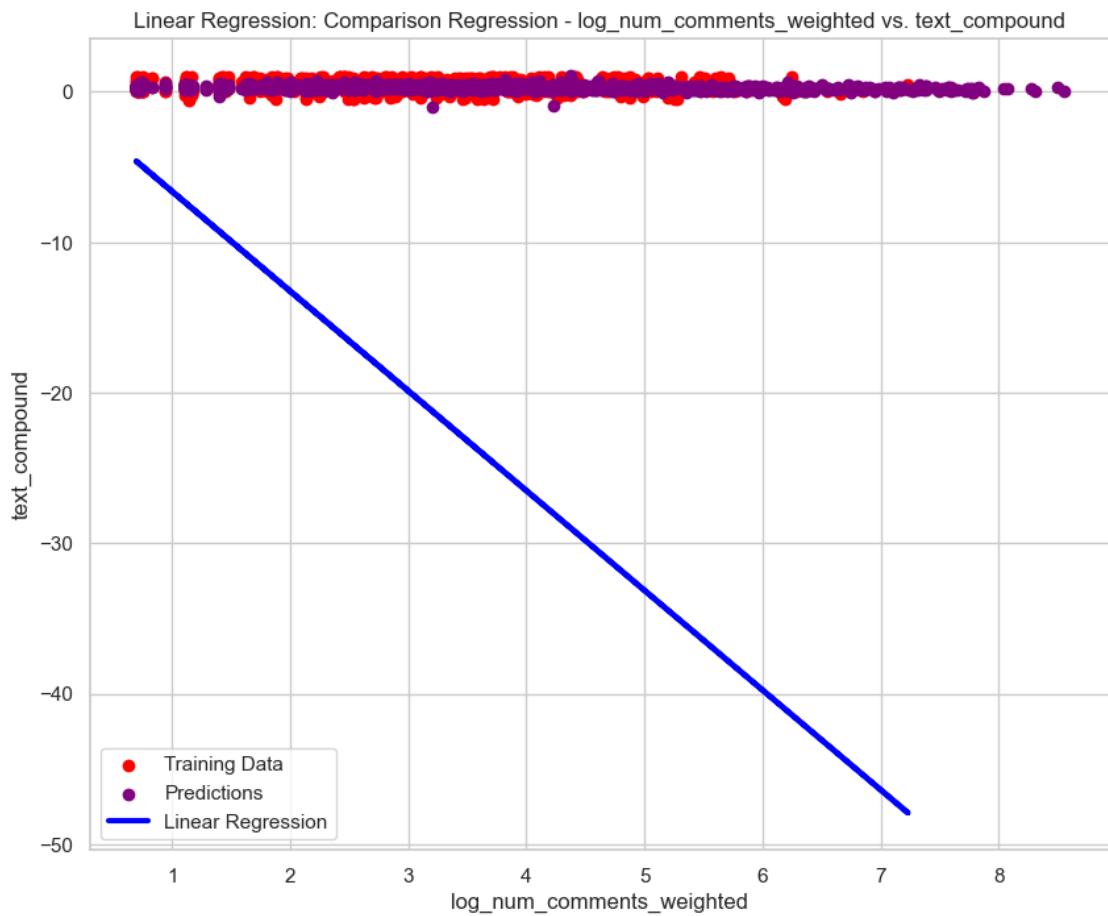
launched=0.11738017265940516



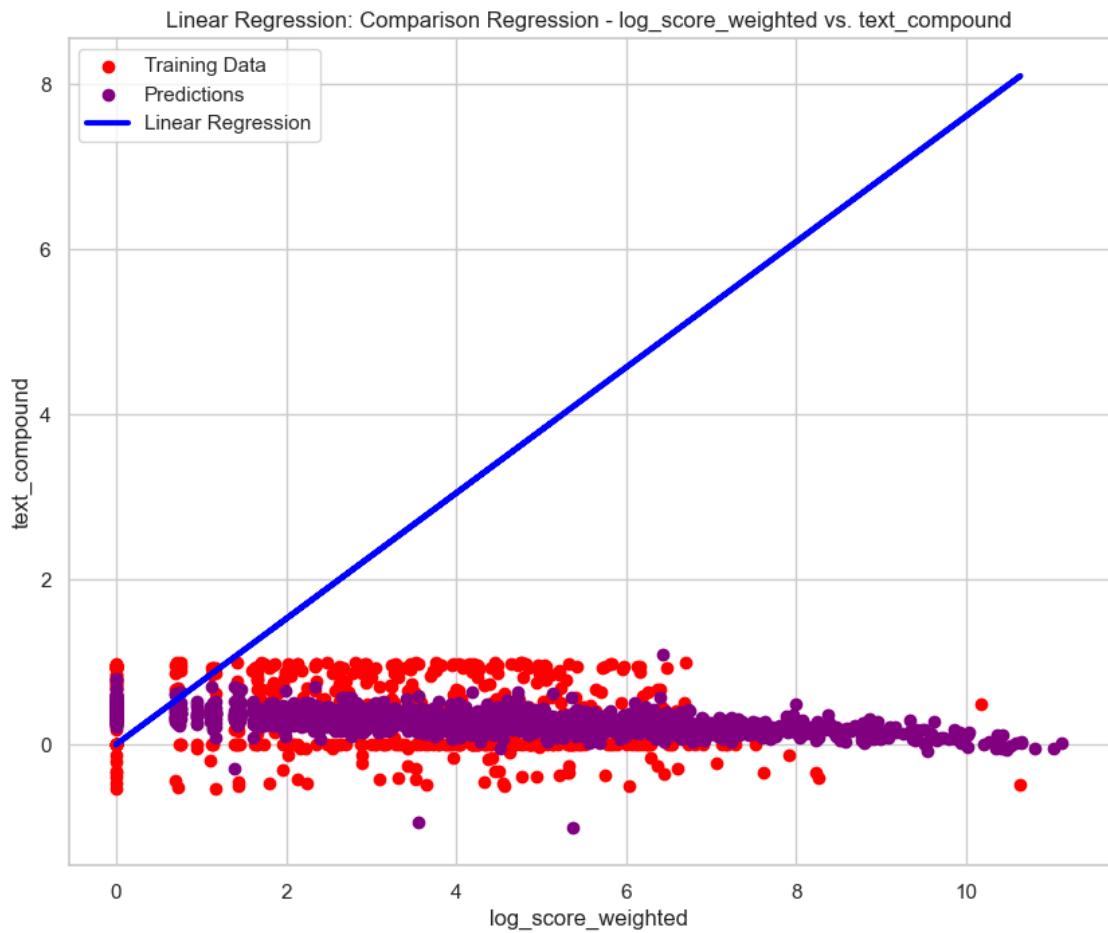
upvote_ratio=0.05597966289250515



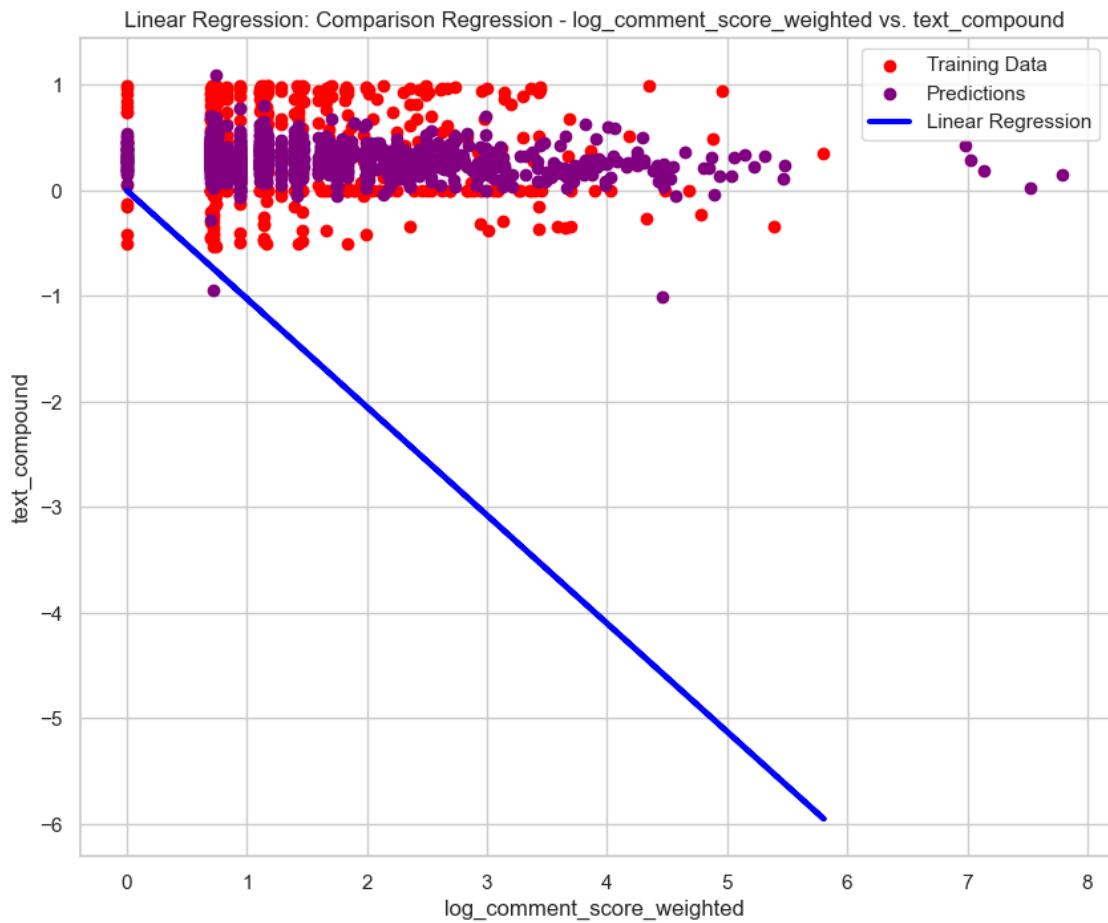
log_launch_distance_f=6.728804371800689



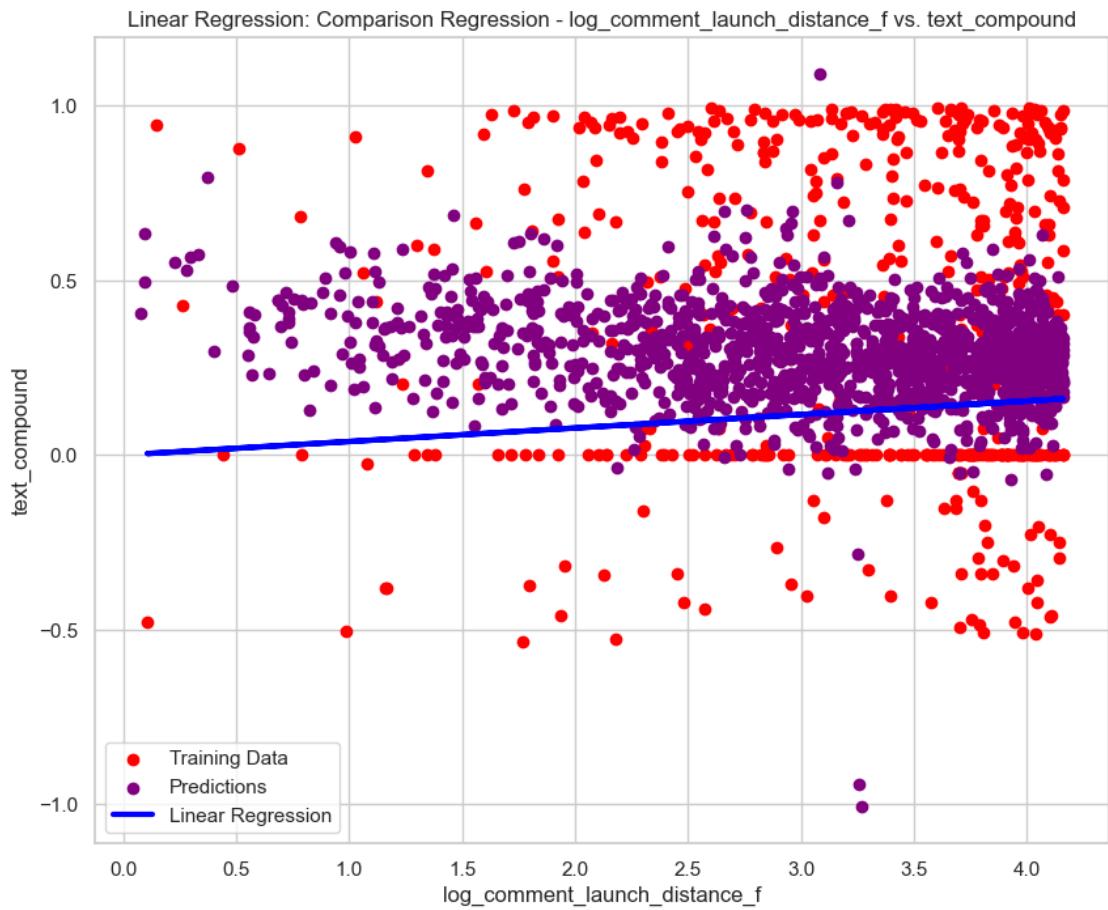
log_num_comments_weighted=-6.625961696064441



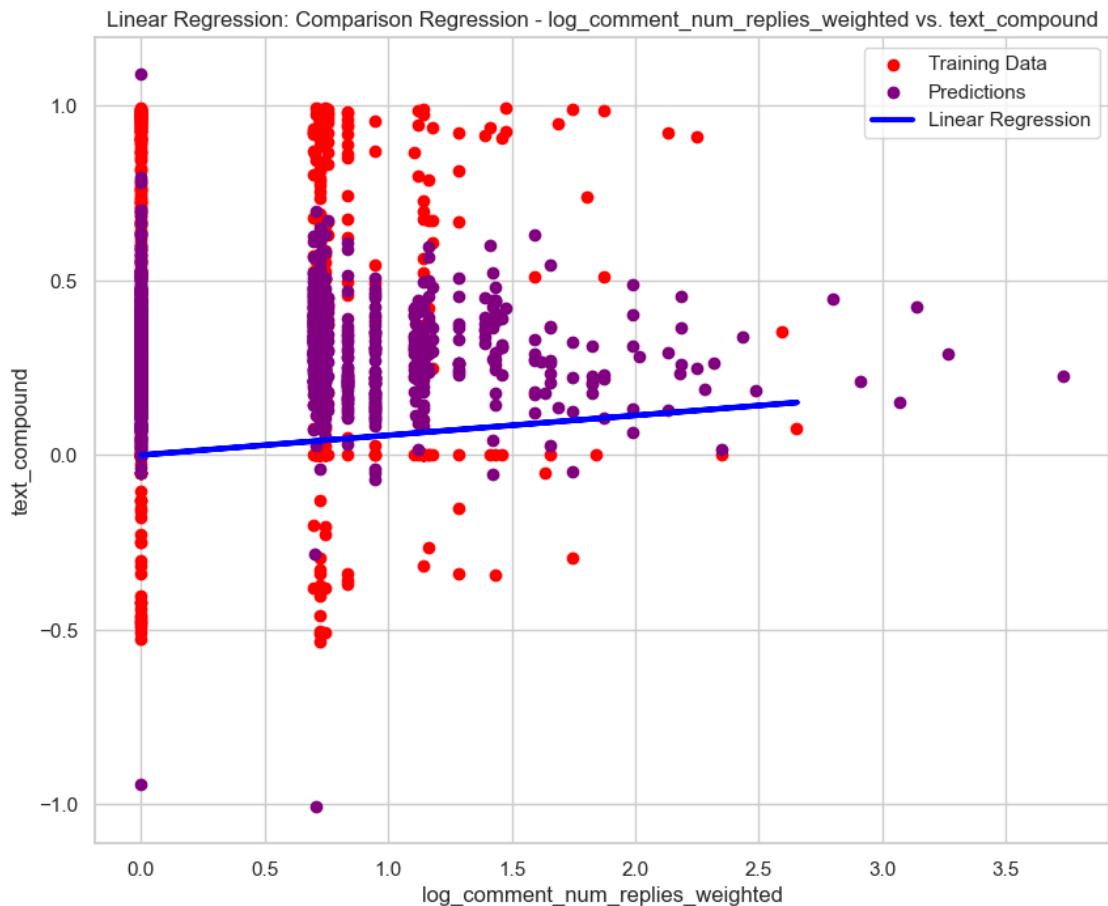
log_score_weighted=0.7619478539391425



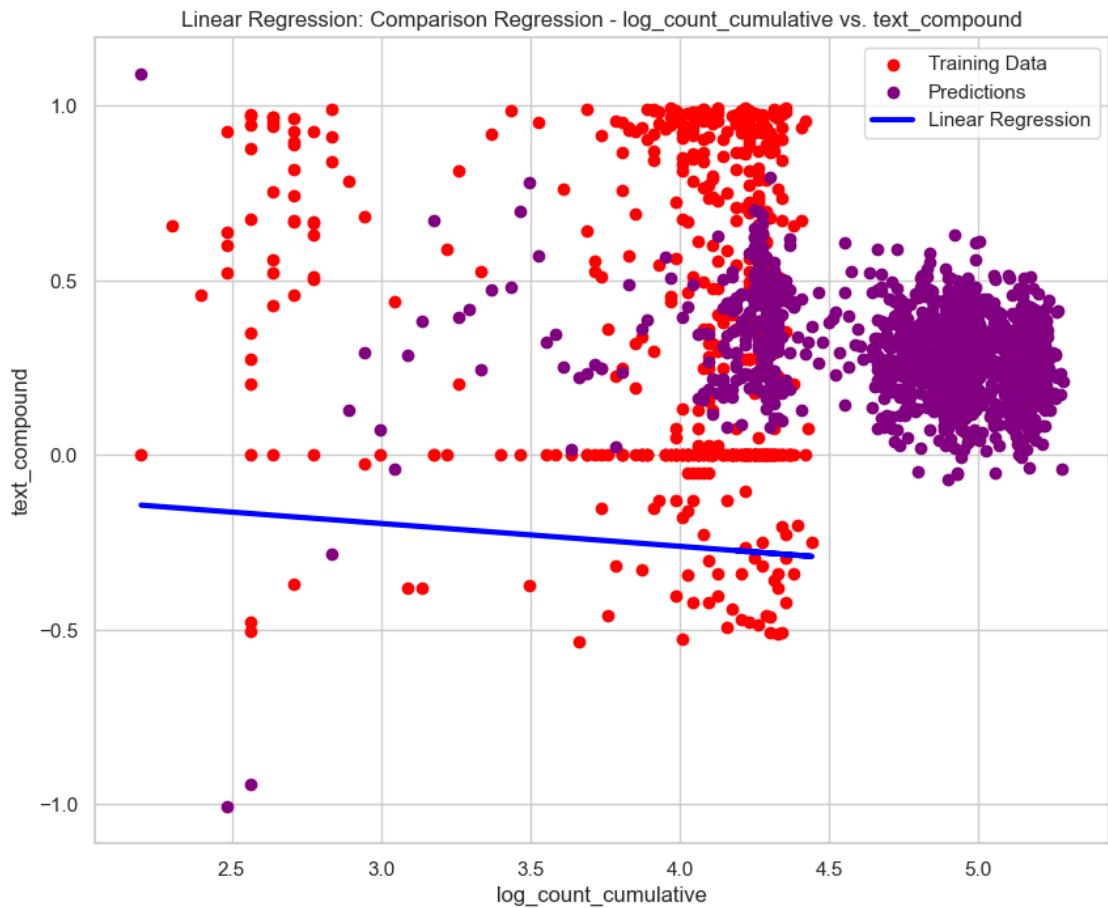
log_comment_score_weighted=-1.0251475221073796



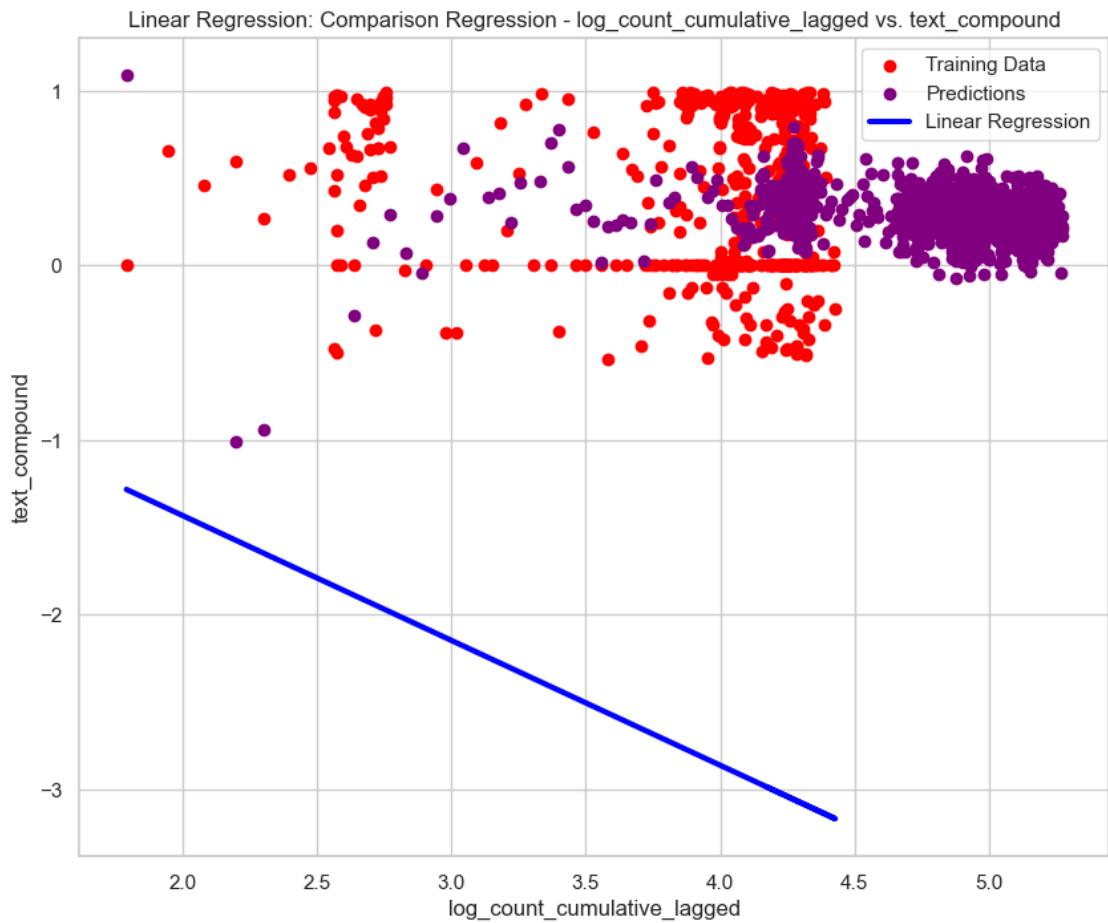
log_comment_launch_distance_f=0.03867064906493635



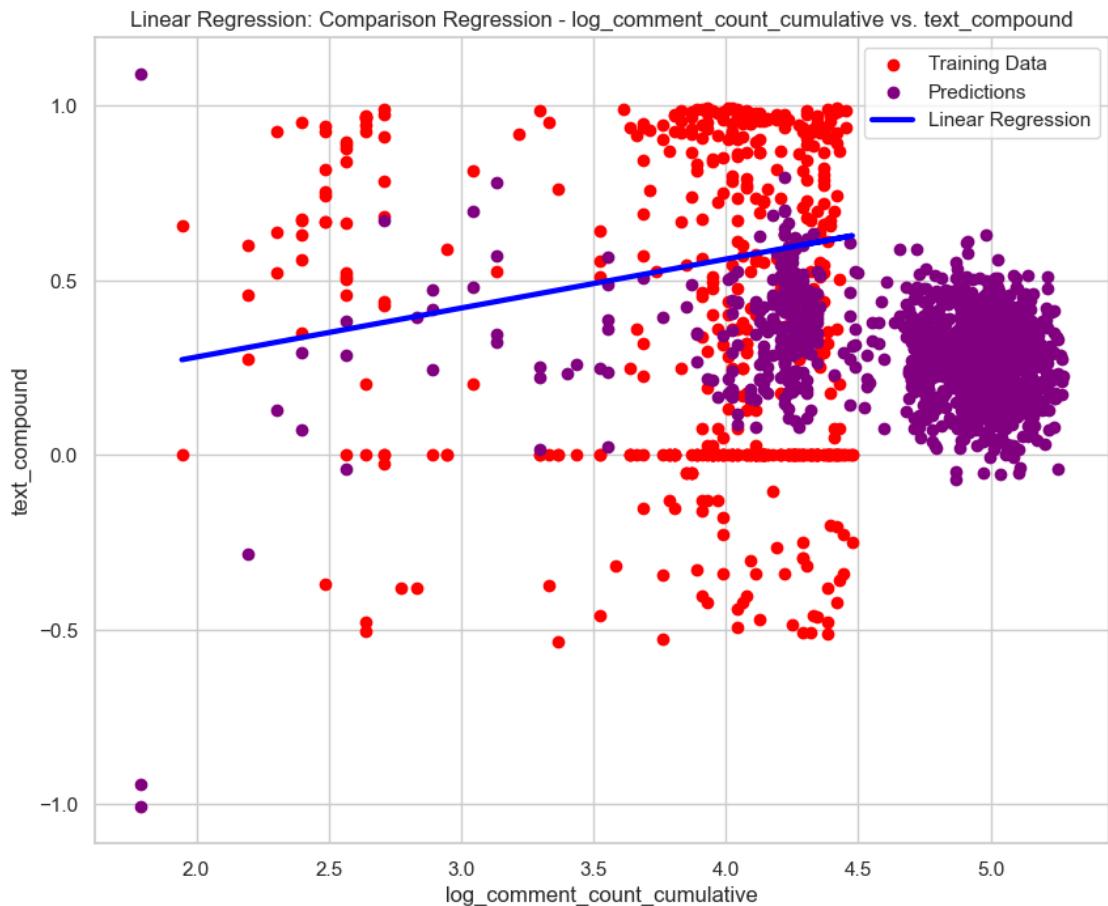
log_comment_num_replies_weighted=0.056655353114418515



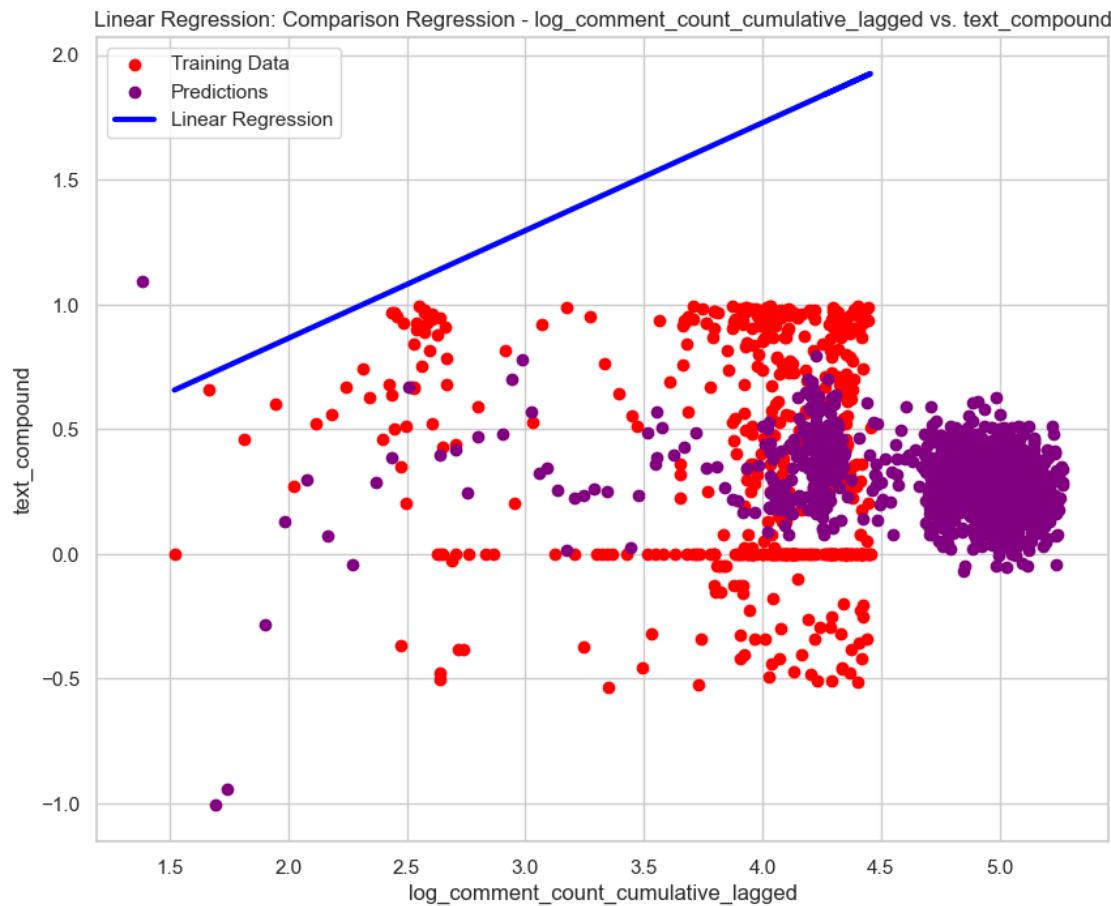
log_count_cumulative=-0.06527589476904148



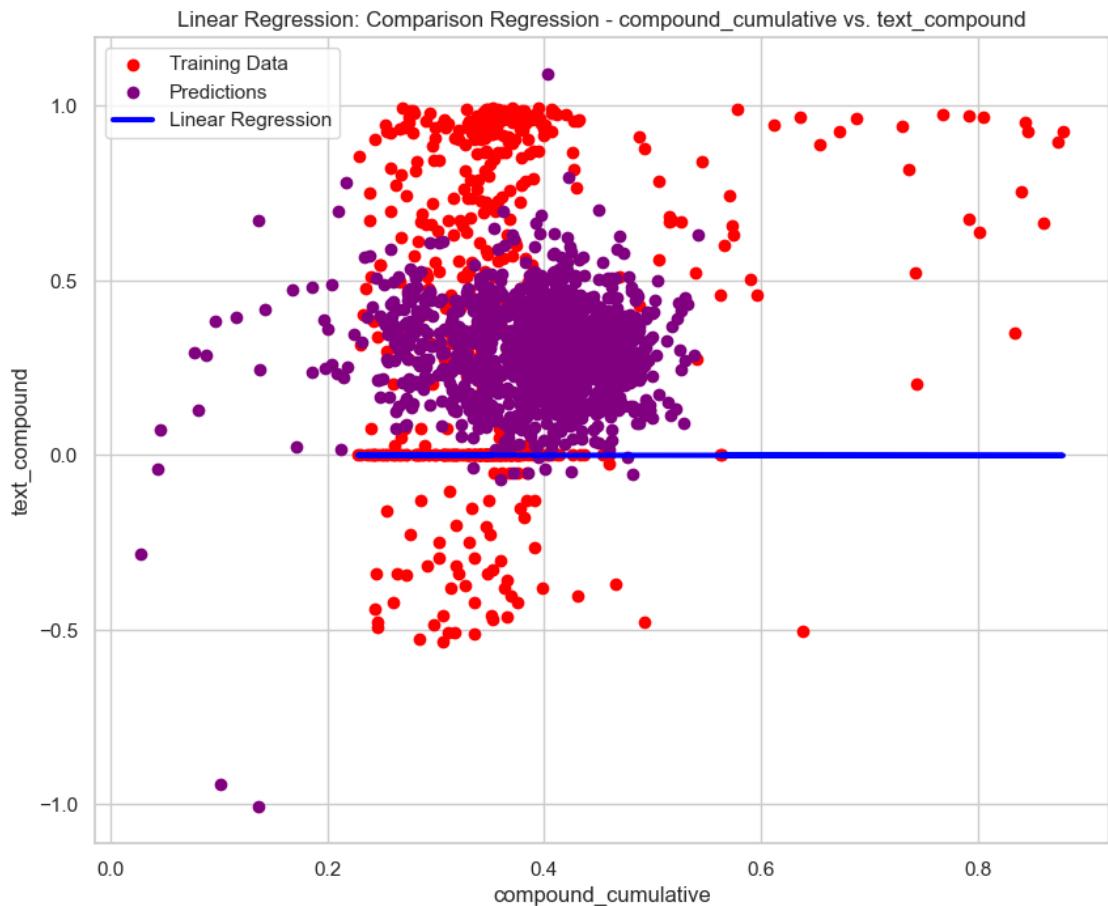
log_count_cumulative_lagged=-0.7153097602936294



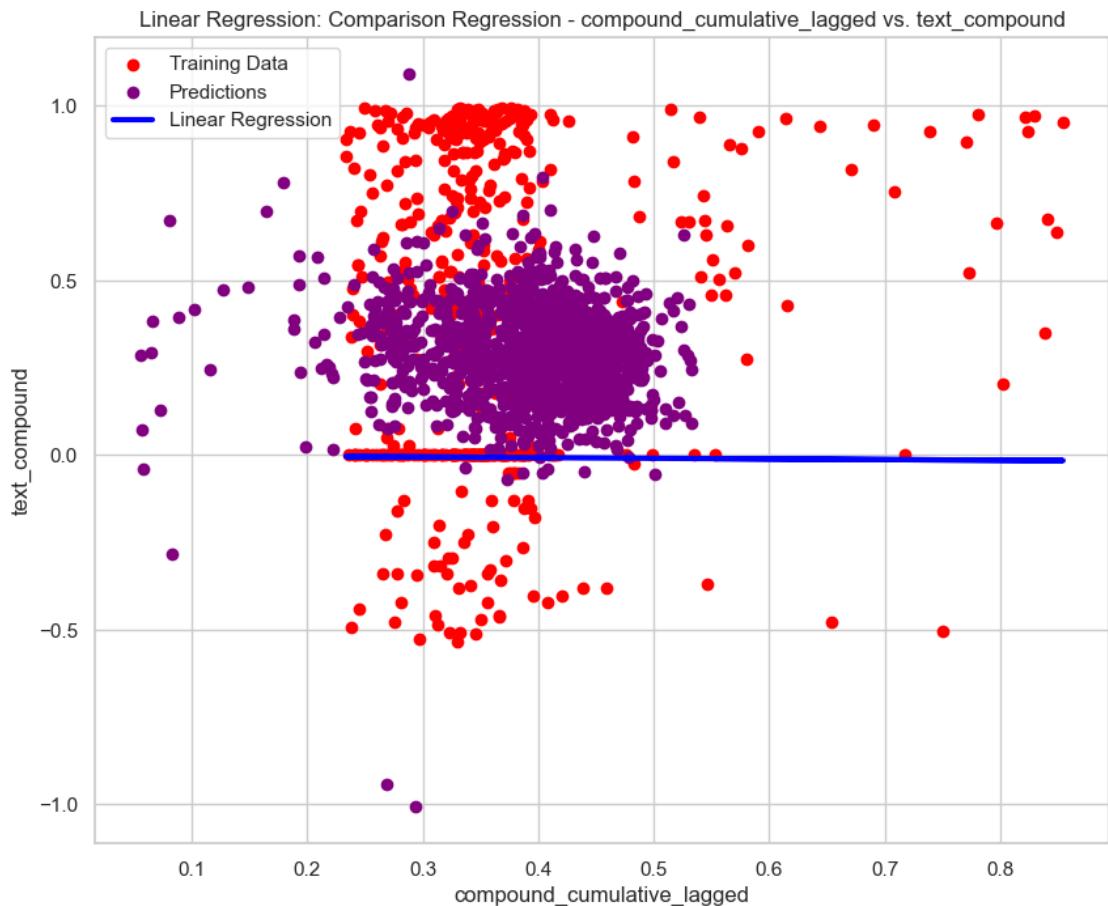
log_comment_count_cumulative=0.14018682192835294



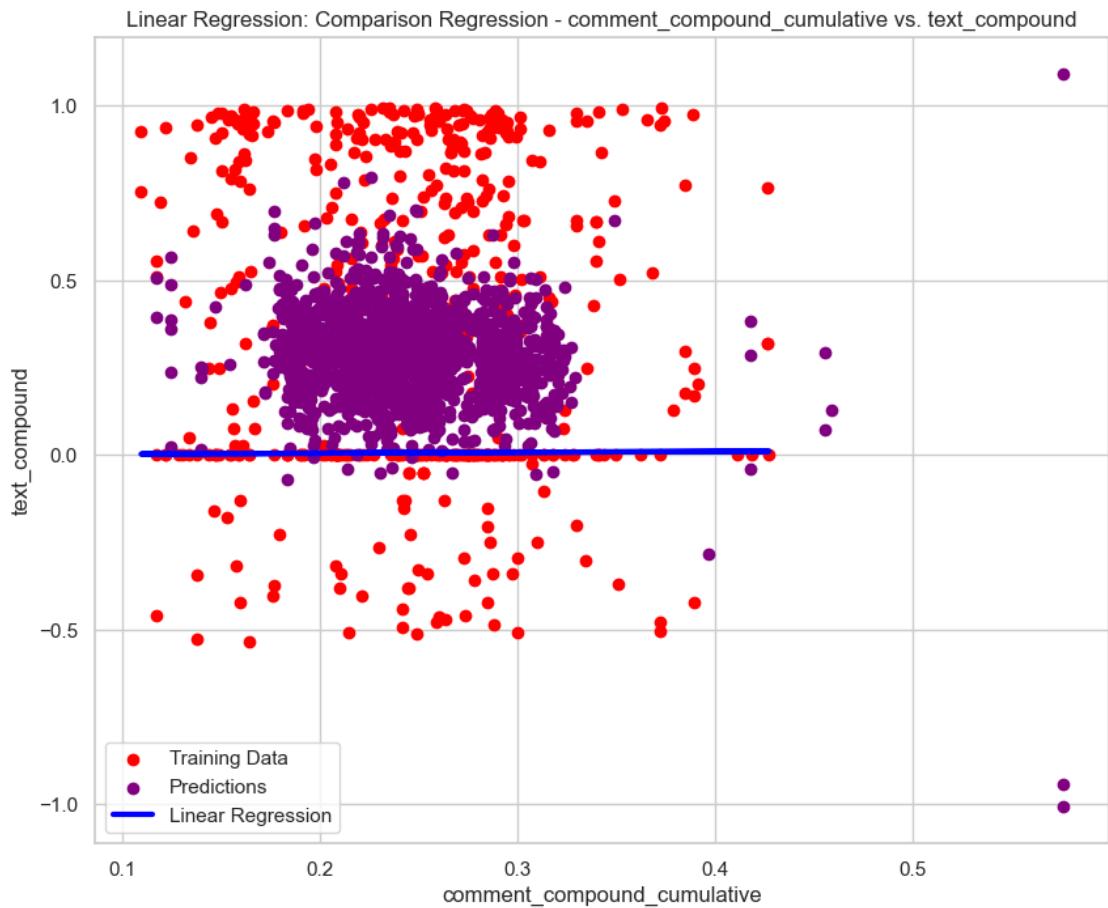
`log_comment_count_cumulative_lagged=0.4318884265142763`



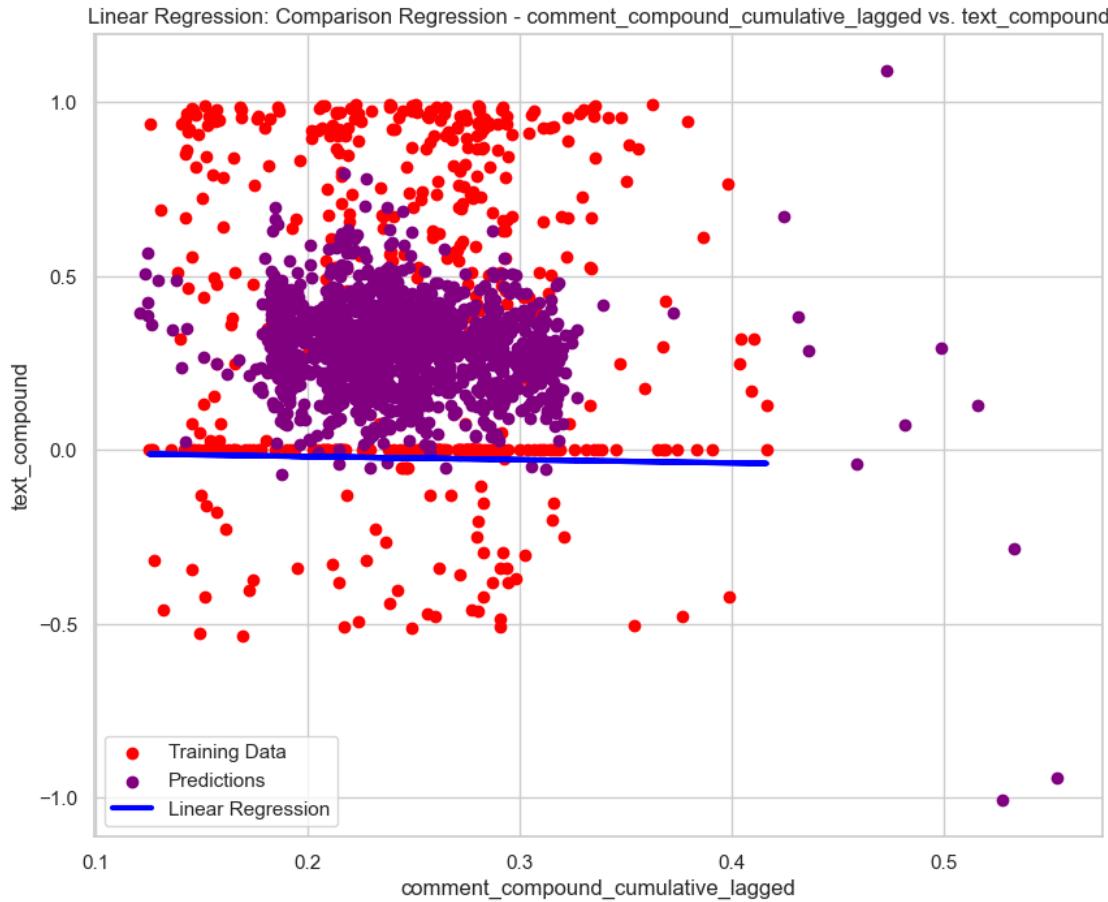
compound_cumulative=-0.0014654141986383635



compound_cumulative_lagged=-0.01858295544797081



comment_compound_cumulative=0.02573363314967698



```
comment_compound_cumulative_lagged=-0.09242568601378859
```

```
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.

Coefficients of GPT-3 Classifier

Feature	Importance
log_comment_launch_distance_f	0.16
comment_compound_cumulative	0.15
log_launch_distance_f	0.11
log_comment_count_cumulative_lagged	0.10
compound_cumulative	0.09
upvote_ratio	0.08
comment_compound_cumulative_lagged	0.06
log_count_cumulative	0.05
log_num_comments_weighted	0.05
log_count_cumulative_lagged	0.05
log_comment_score_weighted	0.04
log_score_weighted	0.03
log_comment_count_cumulative	0.03
compound_cumulative_lagged	0.01
log_comment_num_replies_weighted	0.00
launched	0.00

Coefficients of GPT-4 Classifier

Feature	Importance
log_score_weighted	0.12
log_num_comments_weighted	0.10
compound_cumulative	0.10
upvote_ratio	0.09
comment_compound_cumulative_lagged	0.07
compound_cumulative_lagged	0.07
log_comment_launch_distance_f	0.06
comment_compound_cumulative	0.06
log_count_cumulative_lagged	0.06
log_comment_score_weighted	0.06
log_comment_count_cumulative_lagged	0.05
log_launch_distance_f	0.05
log_count_cumulative	0.04

log_comment_num_replies_weighted 0.04
log_comment_count_cumulative 0.04
launched 0.00

Coefficients of Before-launch Classifier

Feature	Importance
comment_compound_cumulative	0.15
log_comment_launch_distance_f	0.14
log_launch_distance_f	0.12
log_comment_count_cumulative_lagged	0.10
compound_cumulative	0.09
upvote_ratio	0.08
comment_compound_cumulative_lagged	0.06
log_count_cumulative_lagged	0.06
log_count_cumulative	0.05
log_comment_score_weighted	0.04
log_num_comments_weighted	0.04
log_score_weighted	0.04
log_comment_count_cumulative	0.03
compound_cumulative_lagged	0.01
log_comment_num_replies_weighted	0.00
launched	0.00

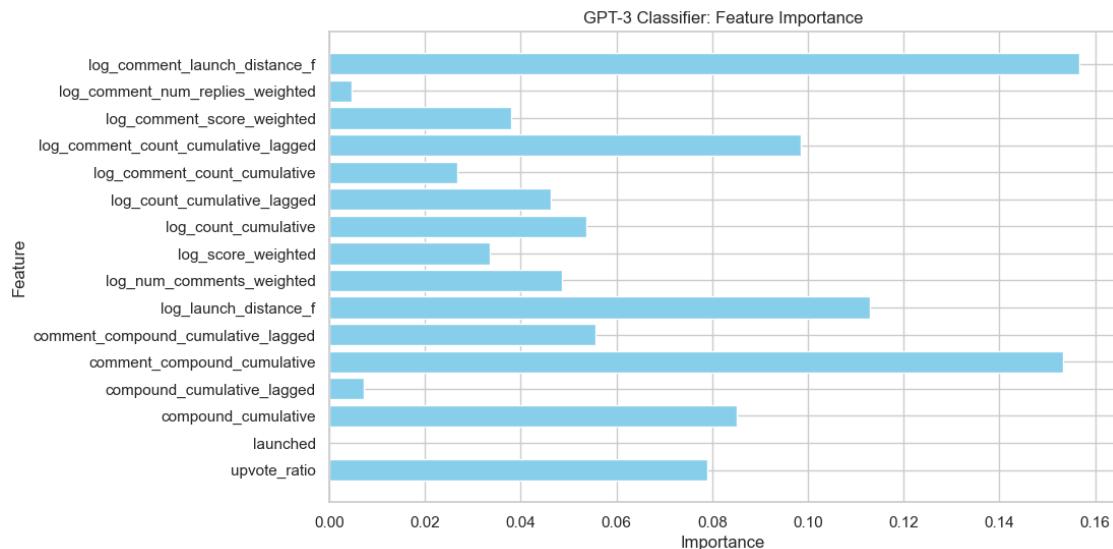
Coefficients of After-launch Classifier

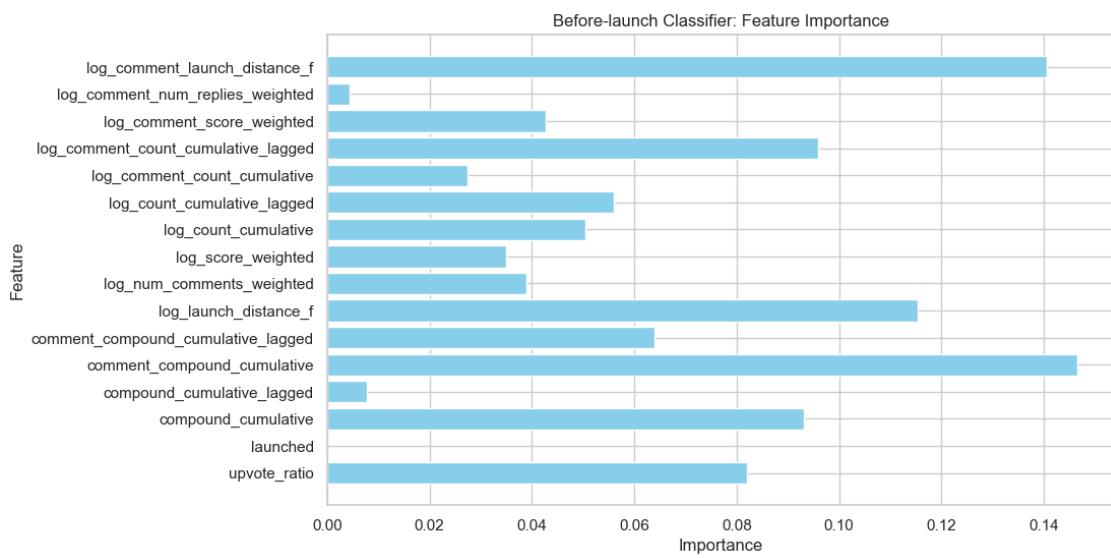
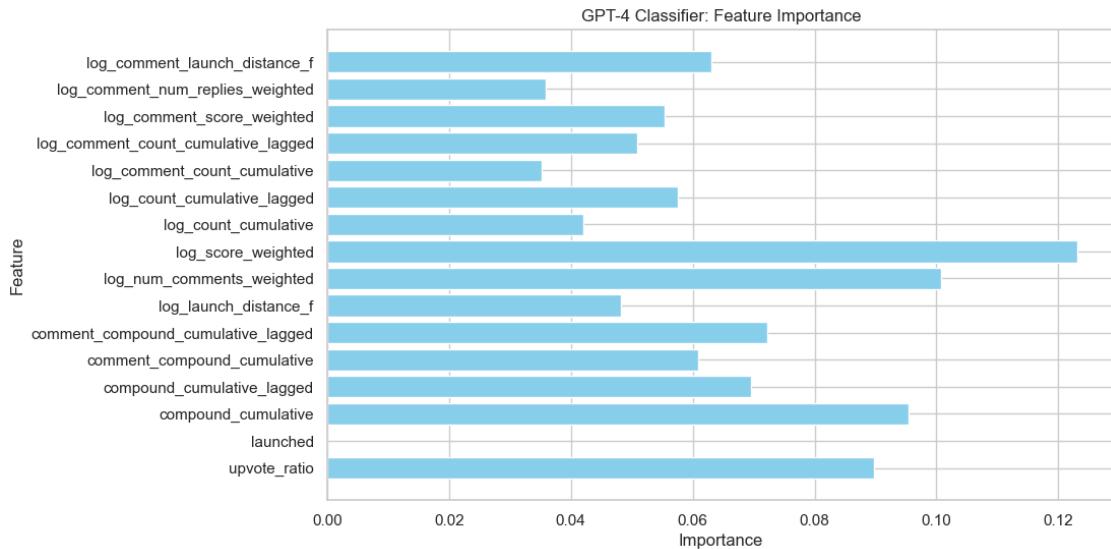
Feature	Importance
compound_cumulative	0.10
log_score_weighted	0.09
compound_cumulative_lagged	0.09
log_num_comments_weighted	0.09
upvote_ratio	0.08
log_comment_score_weighted	0.07
log_comment_launch_distance_f	0.06
comment_compound_cumulative_lagged	0.06
log_comment_count_cumulative_lagged	0.06
log_count_cumulative_lagged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

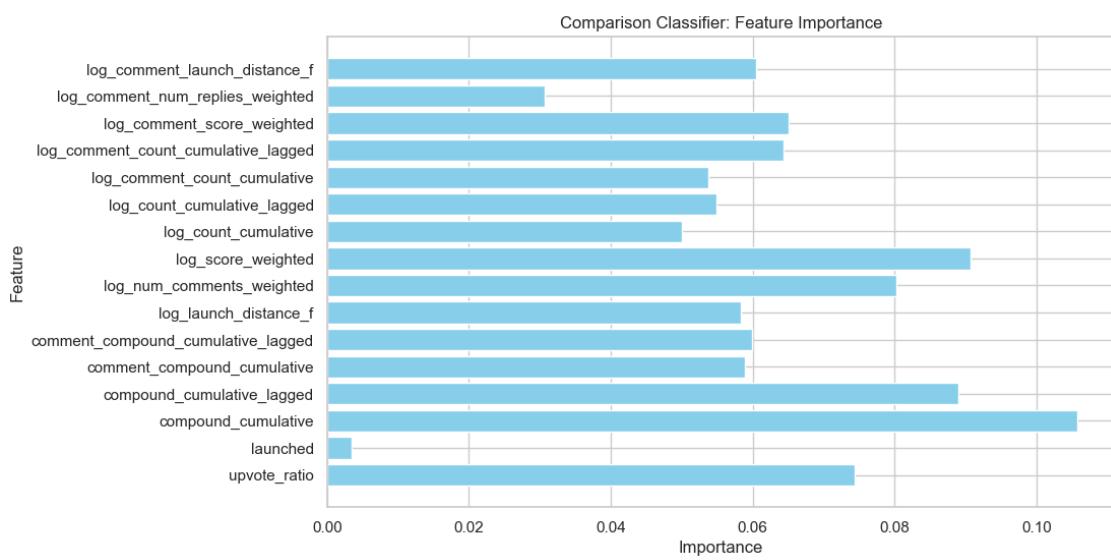
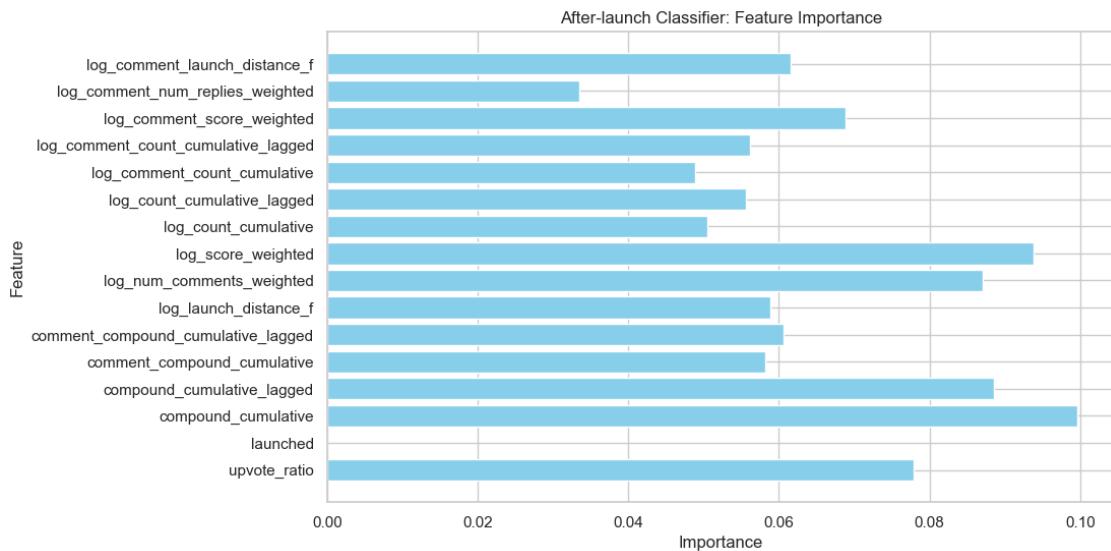
Coefficients of Comparison Classifier

Feature	Importance
compound_cumulative	0.11
log_score_weighted	0.09
compound_cumulative_lagged	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	0.05
log_comment_count_cumulative	0.05
log_count_cumulative	0.05
log_comment_num_replies_weighted	0.03
launched	0.00

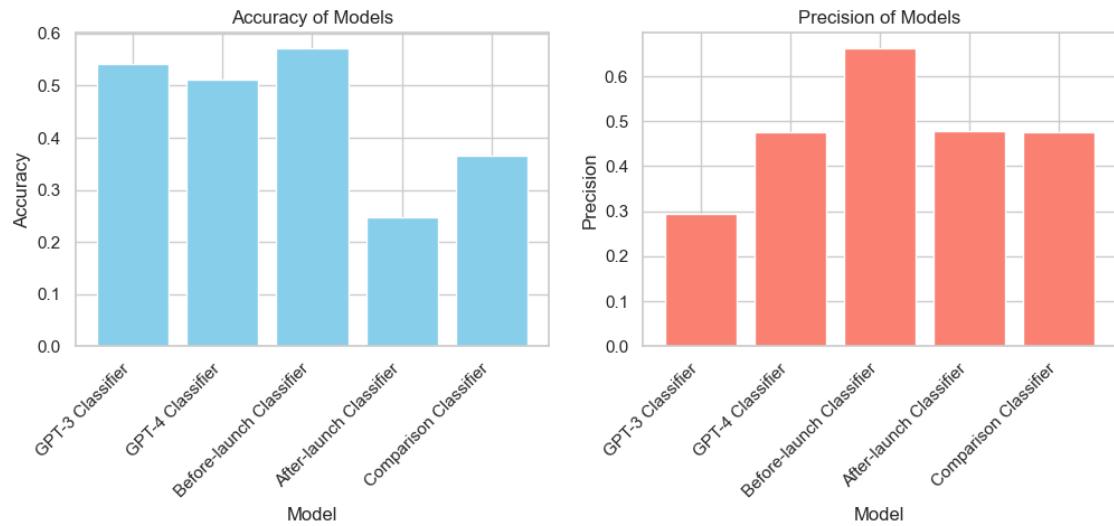
11 Bar charts of MLflow Random Forest runs ranking importance

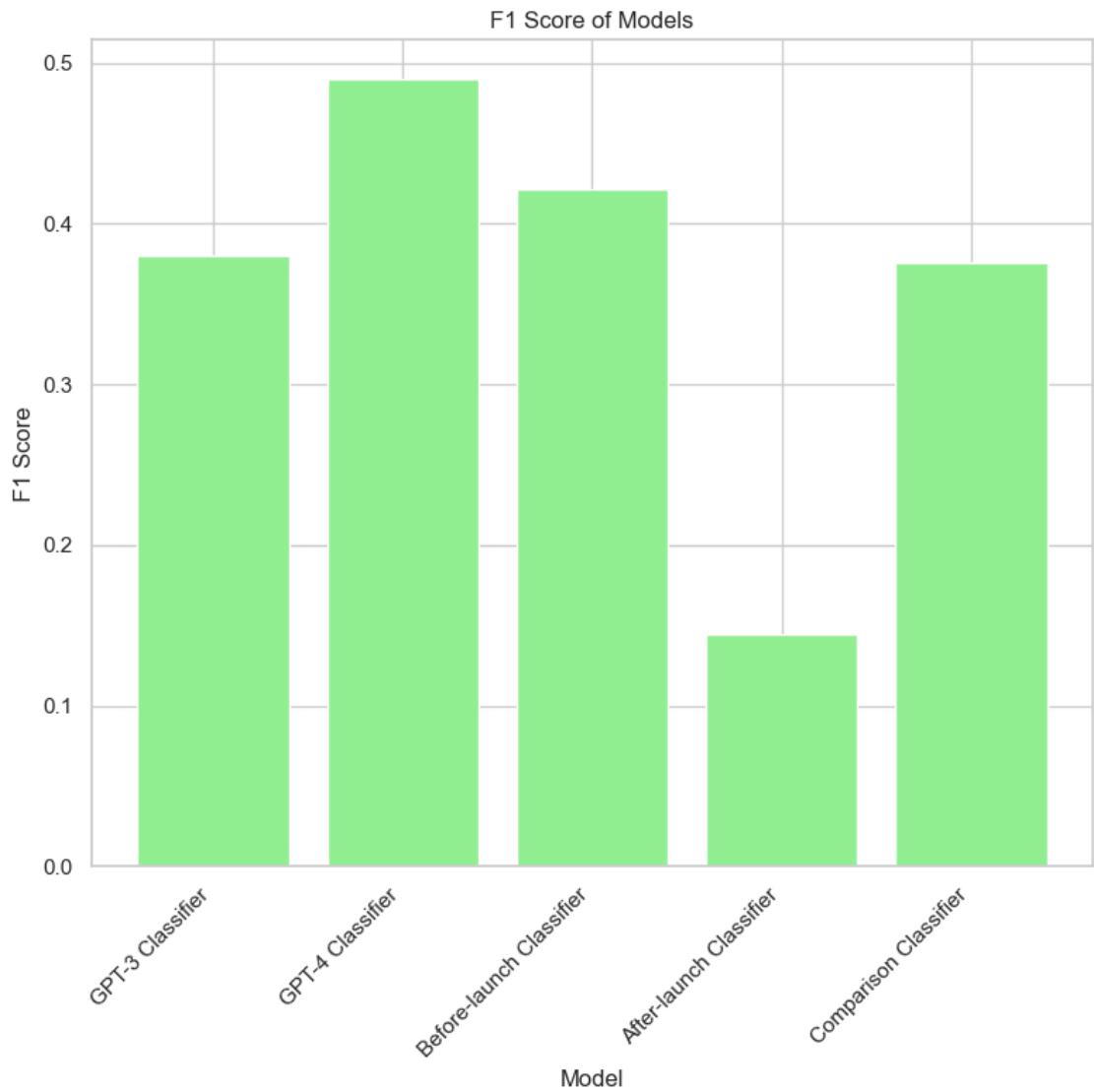






12 Scoring of MLflow Random Forest runs





13 Heatmap of MLflow Random Forest runs

