

1. Build an OLS regression model that predicts the drivers of click-through rate

```
6 . reg logCTR loglagCTR logadrank lognumberofwords retailer brandname logadQuality
```

Source	SS	df	MS	Number of obs	=	891
Model	26.5781272	6	4.42968787	F(6, 884)	=	1504.33
Residual	2.60305656	884	.002944634	Prob > F	=	0.0000
				R-squared	=	0.9108
				Adj R-squared	=	0.9102
Total	29.1811838	890	.032787847	Root MSE	=	.05426

logCTR	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
loglagCTR	.802425	.0168634	47.58	0.000	.769328 .8355219
logadrank	-.0035653	.0020926	-1.70	0.089	-.0076724 .0005418
lognumberofwords	-.0164653	.0042289	-3.89	0.000	-.0247652 -.0081654
retailer	.0361856	.0051061	7.09	0.000	.0261642 .0462071
brandname	.0067389	.0058118	1.16	0.247	-.0046677 .0181455
logadQuality	.0199	.0039687	5.01	0.000	.0121109 .0276891
_cons	.0040629	.007615	0.53	0.594	-.0108827 .0190084

Summary

The above ordinary least-squares regression model analyzed the key drivers of clickthrough rate (CTR) using the log transformed variables loglagCTR, logadrank, lognumberofwords, logadQuality, in addition to the categorical predictors for retailer and brandname. The model achieved a high adjusted R-squared of 0.9102, indicating it explains over 91% of the variation in logCTR. Furthermore, a low root MSE of 0.05426 that suggest the model has high predictive accuracy. The most impactful predictor was loglagCTR with a coefficient of 0.8024, suggesting that past CTR strongly predicts current CTR. Other significant predictors included logadrank, lognumberofwords, retailer, and logadQuality. Only brandname was found to be statistically insignificant to the model.

Drivers

<i>Driver</i>	<i>Coefficient</i>	<i>Significance</i>	<i>Description</i>
<i>loglagCTR</i>	0.8024	Highly	A 1% increase in previous CTR increases current CTR by 0.802%
<i>logadrank</i>	-0.0036	Yes	Higher ad rank (i.e. worse positions) decreases CTR
<i>lognumberofwords</i>	-0.0165	Yes	More words decreases CTR
<i>retailer</i>	0.0362	Yes	Including the retailer's name increases CTR
<i>logadQuality</i>	0.0199	Yes	Higher ad quality increases CTR

<i>brandname</i>	0.0067	No	Including the brand name does not meaningfully impact CTR
------------------	--------	----	---

2. search engine ad ranking prediction model to estimate the drivers of the "ad rank"

To predict Ad Rank, we generated two Ordinary Least Squares (OLS) regression models using different sets of explanatory variables.

- Model 1:** The first model included key factors that are expected to influence Ad Rank based on paid search auction dynamics
 - logCTR*
 - logbidprice*
 - logadquality*
 - loglandquality*
 - lognumberofwords*
 - brandname*
 - retailer*
- Model 2:** The second model built upon the first by introducing date fixed effects to account for daily fluctuations in ranking.

Model 1:

```
. reg logadrank logCTR logbidprice logadQuality loglandQuality lognumberofwords brandname retailer
```

Source	SS	df	MS	Number of obs	=	1,010
Model	293.287225	7	41.898175	F(7, 1002)	=	55.51
Residual	756.325593	1,002	.754815961	Prob > F	=	0.0000
				R-squared	=	0.2794
				Adj R-squared	=	0.2744
Total	1049.61282	1,009	1.04025056	Root MSE	=	.8688

logadrank	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
logCTR	-1.401817	.2573365	-5.45	0.000	-1.906797	-.8968365
logbidprice	-.946042	.2627559	-3.60	0.000	-1.461657	-.4304271
logadQuality	-.0804169	.0634161	-1.27	0.205	-.2048605	.0440268
loglandQuality	-.0528444	.0636374	-0.83	0.407	-.1777222	.0720334
lognumberofwords	.0571885	.0688146	0.83	0.406	-.0778487	.1922257
brandname	-.8758437	.0763268	-11.47	0.000	-1.025622	-.726065
retailer	.2930679	.0811505	3.61	0.000	.1338235	.4523124
_cons	2.179997	.1138458	19.15	0.000	1.956594	2.403401

- Observation:** The low R-squared (27.94%) suggests that there are other important factors influencing Ad Rank that are not captured in this model.

Interpreting the Coefficients

Variable	Coeff.	Interpretation	Statistically Significant?
logCTR	-1.4018	A 1% increase in CTR decreases Ad Rank by 1.40% (higher CTR improves ranking).	✓ (p = 0.000)
logBidPrice	-0.9460	A 1% increase in bid price decreases Ad Rank by 0.95% (higher bids improve ranking).	✓ (p = 0.000)
logAdQuality	-0.0840	A 1% increase in Ad Quality Score decreases Ad Rank by 0.08% (higher quality ads rank better).	✗ (p = 0.205, not significant)
logLandingQuality	-0.0528	A 1% increase in Landing Page Quality decreases Ad Rank by 0.05%.	✗ (p = 0.407, not significant)
logNumberOfWords	0.0518	A 1% increase in the number of words slightly increases Ad Rank (longer keywords slightly lower ranking).	✗ (p = 0.406, not significant)
Brandname (Dummy: 1 if present)	-0.8758	If the brand name is present, Ad Rank improves significantly.	✓ (p = 0.000, highly significant)
Retailer (Dummy: 1 if present)	-0.2931	If a retailer name is present, Ad Rank improves significantly.	✓ (p = 0.000, highly significant)

Model 2:

- Recognizing that Ad Rank can vary by date due to fluctuations in competition, search trends, and external factors, we introduced Date effects in the second model.
 - day_id** – A unique numerical ID assigned to each date (capturing daily variation in Ad Rank).

```
. egen day_id = group( datestring )

. reg logadrank logCTR logbidprice logadQuality loglandQuality lognumberofwords retailer brandname day_id
```

Source	SS	df	MS	Number of obs	=	1,010
Model	372.089862	8	46.5112328	F(8, 1001)	=	68.72
Residual	677.522956	1,001	.67684611	Prob > F	=	0.0000
				R-squared	=	0.3545
				Adj R-squared	=	0.3493
Total	1049.61282	1,009	1.04025056	Root MSE	=	.82271

logadrank	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
logCTR	-.6267304	.2540504	-2.47	0.014	-1.125263 - .128198
logbidprice	-1.128889	.2493916	-4.53	0.000	-1.618279 -.6394989
logadQuality	-.1084303	.0601076	-1.80	0.072	-.2263818 .0095211
loglandQuality	-.0145482	.0603655	-0.24	0.810	-.1330055 .1039092
lognumberofwords	-.0138464	.0654953	-0.21	0.833	-.1423701 .1146774
retailer	.3808751	.0772747	4.93	0.000	.229236 .5325141
brandname	-.7942642	.0726716	-10.93	0.000	-.9368703 -.6516581
day_id	-.0799418	.0074088	-10.79	0.000	-.0944804 -.0654032
_cons	2.701867	.1181579	22.87	0.000	2.470001 2.933732

Comparing Model 1 and Model 2 Performance

Metric	Without Date Fixed Effects	With Date Fixed Effects (day_id)	Change
R-squared	0.2794	0.3545	+7.5%
Adj R-squared	0.2744	0.3493	+7.5%
Root MSE	0.8688	0.8227	Improved

R-squared increased from 27.94% to 35.45%, meaning more variation in Ad Rank is explained by the model when Date Fixed Effects are included. Root MSE decreased, meaning the model's prediction error has reduced.

Conclusion: Date significantly impacts Ad Rank, and including day_id improves model accuracy.

Interpreting the Coefficients

Variable	Coeff.	Interpretation	Statistically Significant?
logCTR	-0.6276	A 1% increase in CTR decreases Ad Rank by 0.63% (higher CTR improves ranking).	(p = 0.014)
logBidPrice	-1.1288	A 1% increase in bid price decreases Ad Rank by 1.13% (higher bids improve ranking).	(p = 0.000)
logAdQuality	-0.1084	A 1% increase in Ad Quality decreases Ad Rank by 0.11%.	✗ (p = 0.072, not significant)
logLandingQuality	-0.0155	A 1% increase in Landing Page Quality decreases Ad Rank by 0.02%.	✗ (p = 0.810, not significant)
logNumberOfWords	0.0572	A 1% increase in keyword length slightly increases Ad Rank.	✗ (p = 0.833, not significant)
Retailer (Dummy: 1 if present)	-0.3809	Including a retailer name significantly improves Ad Rank.	(p = 0.000)
Brandname (Dummy: 1 if present)	-0.7946	Including a brand name significantly improves Ad Rank.	(p = 0.000)
Day Fixed Effects (day_id)	-0.0799	Ad Rank varies significantly across different days.	(p = 0.000)

Differences between Model 1 and Model 2:

1. R-squared increased from 27.94% to 35.45% → This means daily variations explain an extra 7.5% of Ad Rank changes.
2. logCTR is still significant but has a weaker effect (-0.63 instead of -1.40)
3. logBidPrice remains highly significant (-1.13) → Bid price still plays a major role in ranking.
4. Retailer and Brandname remain highly significant → These factors strongly influence Ad Rank.
5. Ad Quality and Landing Page Quality are still NOT significant → These factors don't seem to impact Ad Rank much in this dataset.
6. Day Fixed Effects (day_id) is significant → Ad Rank changes based on the day.

Key Takeaways

Final Conclusion: Ad Rank depends heavily on Bid Price, CTR, and whether the ad contains a brand or retailer name. It also fluctuates daily, making Date an important factor.