

Feature name	Source	Value type	Scale	Data type	Unique count	Min	Max	asure	unversion	fac	Outlier count	Outlier treatment	Null	Notes
location	DB	Categorical	Nominal	Varchar	2						0			location is associated with sale_labor_day, item rating rank by category and all prices (uk prices are expressed in \$ with avg conversion rate)
season	DB	Categorical	Nominal	Varchar	4						0			variable season show association with image count
description_word_cnt	DB	Numeric	Ratio	Integer	4363	1	15151				86913	Transformation		correlation with ebay_price,amazon price, profit, returns count
item_rating_rank_by_category	DB	Numeric	Interval	Integer	1532	1	1617				0			correlated with reviews count and revised star rating
category_size_rank	DB	Numeric	Ratio	Integer	42	1	42				15557	Categorize		correlation to item_rating_rank_by_category
creation_time	DB	Date	Time-series	Date	1242									
image_cnt	DB	Numeric	Ratio	Integer	15	0	27				0			41% value 1, 2.4% 0
review_count	DB	Numeric	Interval	Decimal	11837	1	904570				306697	Categorize	NA	56% NA,correlated with description word count, with ebay_price and profit
star_rating	DB	Numeric	Interval	Decimal	41	1	5				290289	Categorize	NA	56% NA,Normal distribution (skewed left),star rating is correlated to ebay price, amazon price, competitor price
revised_star_rating	DB	Numeric	Ratio	Integer	1629	0	28				0		NA	0 = NA, 2 means review count low (under 135 reviews),When analyzing the data without these values we get a normal distribution. correlated with rating rank by category, review count, popular brand and revies cnt category
returns_cnt	DB	Numeric	Ratio	Decimal	20	0	45	\$			39232	Drop	NA	92% with 0 returns. Variable returns count is correlated with ebay_price, amazon price, category size ranking and to star_rating
ebay_price	DB	Numeric	Ratio	Decimal	22111	0	10202	\$			48326	Transformation		Distribution skewed right, correlated with amazon price and profit
competitor_price	DB	Numeric	Ratio	Decimal	20138	0	920000	\$			47990	Transformation		Distribution skewed right, correlated with loction,condition,ebay price,amazon price and profit
amazon_price	DB	Numeric	Ratio	Decimal	22201	0	2520	\$			47899	Transformation		
price_diff_competitor	DB	Numeric	Ratio	Decimal	12185	-999	99.9	\$			57646	Transformation	NA	values -999 are meant to be NA in the original data, variable price_diff_competitor is associated with sale days
profit	DB	Numeric	Ratio	Decimal	49086	-138	2239				51194	Transformation		variable profit is correlated with ebay price, competitor price and amazon price, condition, location, sale days: black friday,labor day,4th of july, item_sell_rank by category
profit_pct	DB	Numeric	Ratio	Integer	81897	-1599	100				34052	Transformation	NA	association with sale days, location and condition
category	DB	Categorical	Nominal	Text	42									3 highest categories hold 45% of all items
brand	DB	Categorical	Nominal	Text	107861									5% with no brand, 28% of all items are from a small brand (5 items or less)
brand size	DB	Categorical	Ordinal	Varchar	3	Small	Large						NA	brand size is associated with image count
popular brand	DB	Categorical	Nominal	Boolean	2									Variable reviews cnt category is correlated with
sales_cyber_monday	DB	Categorical	Nominal	Integer	2									item rating rank by category, revised star rating
sales_black_friday	DB	Categorical	Nominal	Integer	2									
sales_colombus_day	DB	Categorical	Nominal	Integer	2									
sales_labor_day	DB	Categorical	Nominal	Integer	2									
sales_presidents_day	DB	Categorical	Nominal	Integer	2									
sales_4th_july	DB	Categorical	Nominal	Integer	2									
sales_newyear	DB	Categorical	Nominal	Integer	2									
sales_xmas	DB	Categorical	Nominal	Integer	2									
reviews_cnt_category	DB	Categorical	Ordinal	Varchar	3	Low	High							Variable reviews cnt category is correlated with
success	DB	Categorical	Nominal	Boolean	2									item rating rank by category, revised star rating