

Positive vs Neutral vs Negative Star Ratings

Added a new categorical column, review_rating, based on the number of stars. (1-2 stars = Negative, 3 = Neutral, 4-5 = Positive)

```
alter table sukanya_final add column review_rating varchar(30) default
"Positive";
update sukanya_final set review_rating= "Neutral" where stars=3;
update sukanya_final set review_rating= "Negative" where stars=2;
update sukanya_final set review_rating= "Negative" where stars=1;
```

This command converts the categorical variables into values using one hot encoding.

```
./dlatkInterface.py -d scg_covid_pa -t sukanya_final -c message_id --
correlate --csv --outcome_table sukanya_final --outcomes review_rating
--categorical review_rating --outcome_with_outcome_only --output ~/
correlations
```

Ngrams:

```
./dlatkInterface.py -d scg_covid_pa -t sukanya_final -c message_id \
-f 'feat$1to3grams$sukanya_final$message_id$16to16$0_01$pmi3_0' \
--outcome_table sukanya_final --group_freq_thresh 1 \
--outcomes review_rating --output_name word_clouds1to3_review_rating \
--tagcloud --make_wordclouds --categorical review_rating --csv --
logistic
```

Topics:

```
./dlatkInterface.py -d scg_covid_pa -t sukanya_final -c message_id \
-f 'feat$cat_reviews_cov_app25_cp_w$sukanya_final$message_id$1gra' \
--outcome_table sukanya_final --topic_tagcloud --make_topic_wordcloud
--topic_lexicon reviews_cov_app25_freq_t5011 --group_freq_thresh 1 --
tagcloud_colorscheme bluered --csv --outcomes review_rating --
output_name topics_review_rating --rmatrix --logistic --categorical
review_rating
```