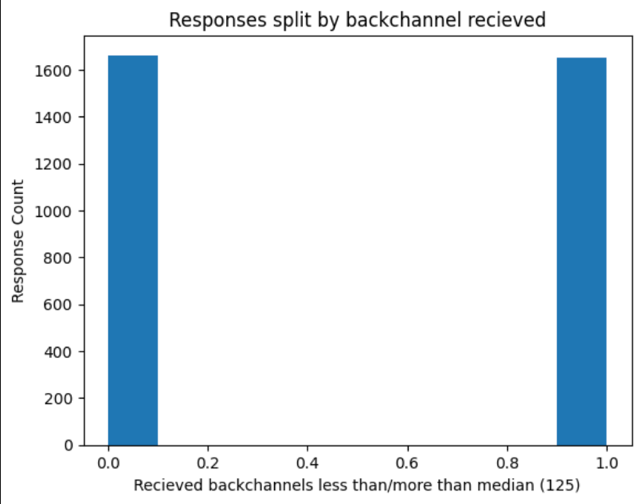
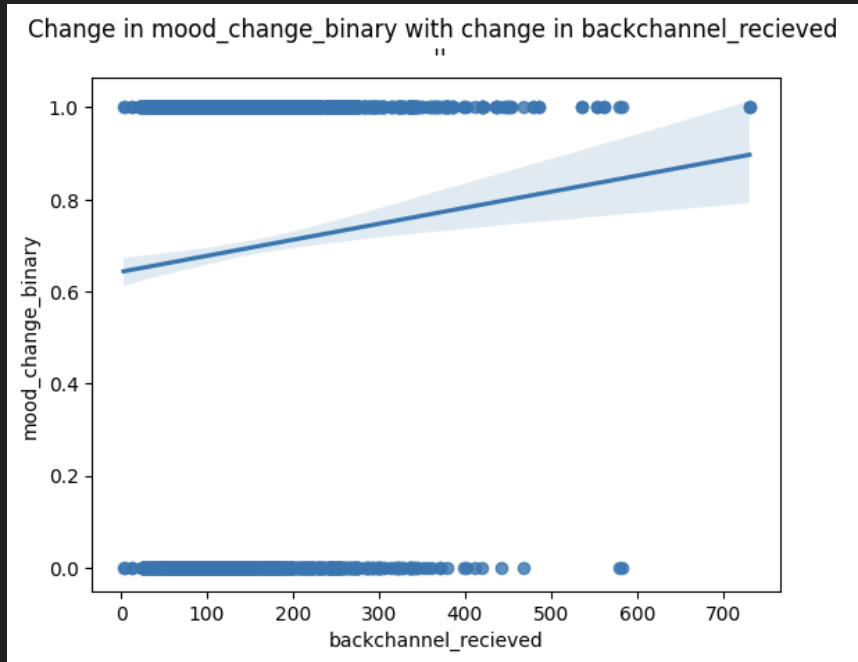
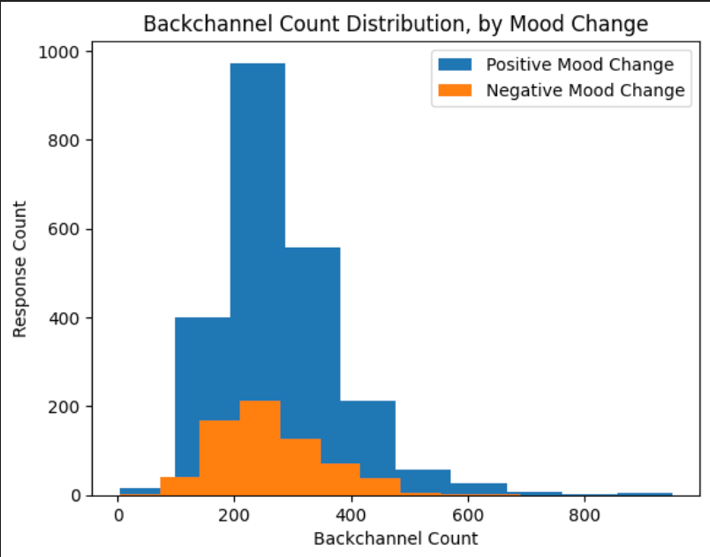
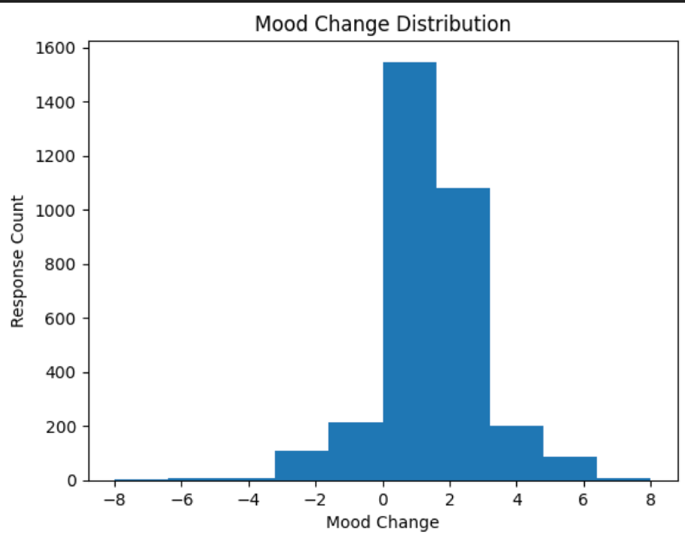
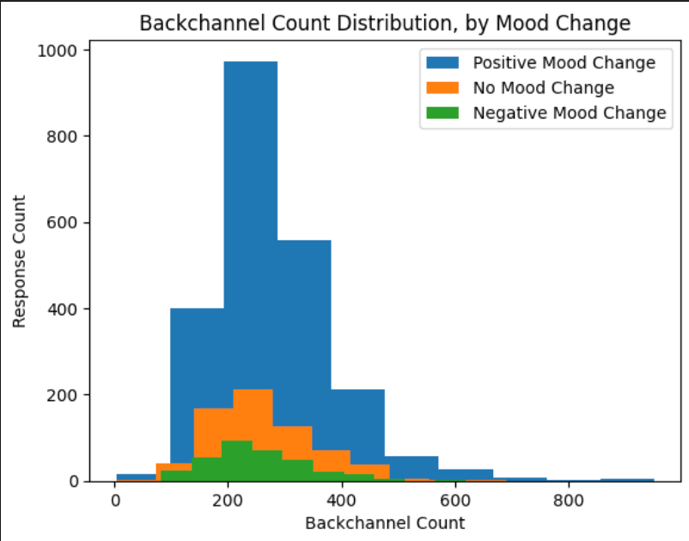
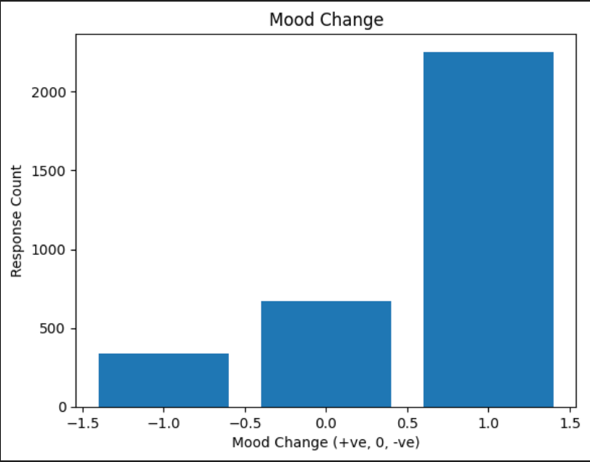
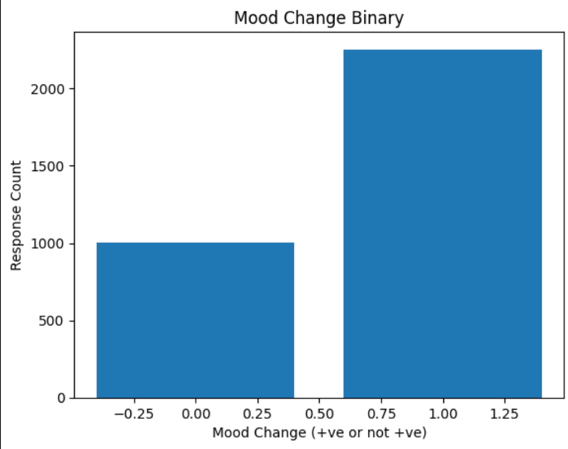
|  |  |  |  |
| --- | --- | --- | --- |
| Item | Todo | Assigned | Progress |
|  | Does **received backchannel count/total turn length** ratio have a clearer difference in outcome?   * Aim: maybe it has more obvious effect than just backchannel count by itself | jessica | While average is around 0.0972, 65.96% consist of 0.0 ratio |
|  | For *each person*, is there correlation between average backchannel given and average ratings by other users?  Rephrase: is there correlation between average backchannel received and rating of other user?   * ratings:'you\_like\_me','you\_feel\_isolated','you\_feel\_left\_out','you\_think\_i\_am\_intelligent','you\_think\_i\_am\_quickwitted','you\_think\_i\_am\_competent','you\_think\_i\_am\_kind','you\_think\_i\_am\_friendly','you\_think\_i\_am\_warm','you\_think\_my\_status' (these are subjective) * ratings: get rows where | jessica |  |
|  | Binarize variable (ie. backchannel count >= mean) ⇒ do hypothesis test (is there a difference in engagement between groups with vs without backchannels? Between groups with high vs low (above/below average) backchannels?) |  |  |
|  | Read papers → what qualitative ways are there to analyze textual/survey data with regards to the effect of one variable? |  | Progress |
|  | Paper draft: describing dataset |  | Not started |
|  | Paper draft:  Methods section for data analysis (what are we doing right now) |  | Not started |
|  | *Backchannel received* vs ***mood change*** (+/-) (end\_effect - pre\_affect)  Groupby positive/negative mood change  Figure (up): affect change >0 vs affect change < 0 (dropped no affect change)  Figure (down): affect change >0 vs affect change <=0.  Is there a difference in backchannel count between people with positive/negative mood changes? |  |  |
|  | Split conversationalists into “highly ranked” vs “low ranked” conversationalists: k-nearest neighbours |  | Not started |
|  | Split conversations into “highly ranked conversations” vs “low ranked conversations”: look for variables of interest with k-nearest neighbors   * Q: What are “good” vs “bad” conversations? “Effective” vs “less effective” conversations? * Aim: finding other variables of interest that distinguish good conversations besides backchannel count |  | Not started |

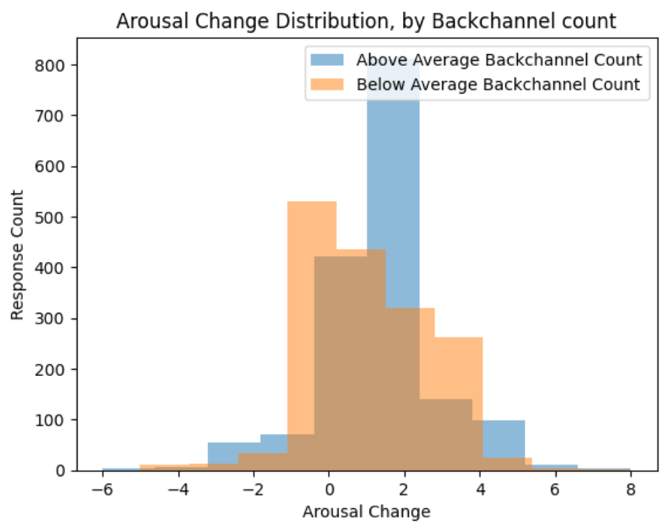
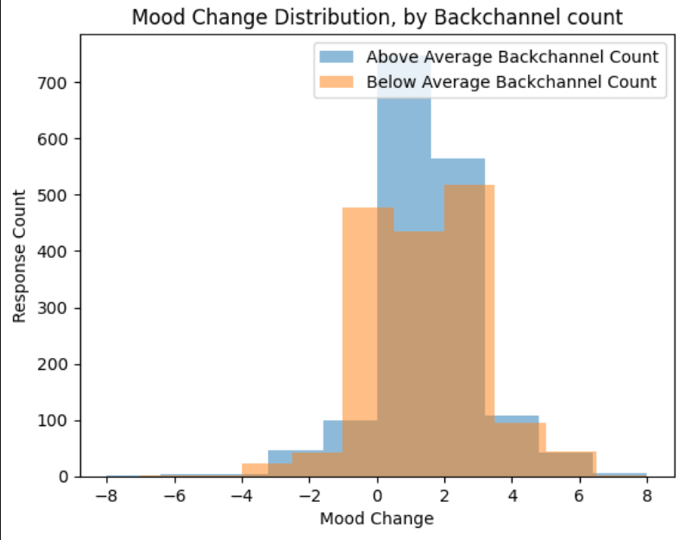


Binarize backchannel count

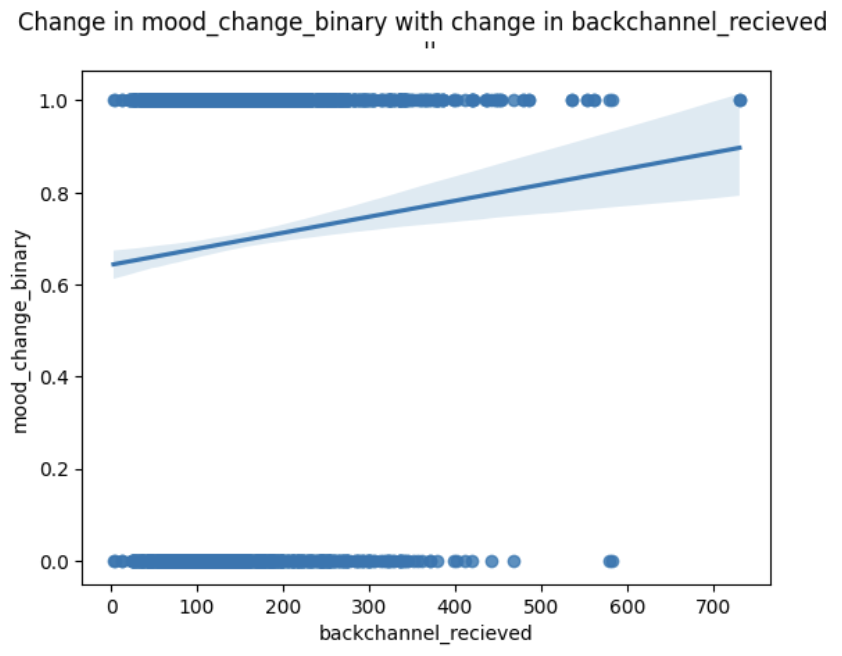
**Binarize “mood” outcome:** positive vs no positive mood change, +/0/-ve mood change. Is there a difference in backchannel count distribution between them?



**Binarize backchannel count:** above/below average backchannel count. Is there a difference between mean “mood” outcome between them?



**Continuous backchannel 🡪 metrics**



* Intercept 0.6425 (when no backchannel, that is mood change)
* backchannel\_recieved 0.00039 (each extra backchannel increases positive mood change by this much)

Correlations

*Measures the strength and direction of correlation between two variables.*

* Above 0 = positive correlation, Below 0 = negative correlation
* Absolute value above 0.5 = strong correlation, below 0.5 = weak correlation
* Summary: backchannel received has a weak positive correlation with the below variables

Pearman's coefficient: linear relationship (shown here)

Spearman's coefficient: non-linear relationship (not shown, but similar results)

|  |  |
| --- | --- |
| *Notes* | *Correlation with “Backchannel Received”* |
| Measure of “engagement” | how\_long\_other 0.153920 (2056 missing responses)  how\_long\_you 0.120976 (2056 missing responses)  your\_mind\_wander -0.090263  my\_mind\_wander -0.069034 |
| Measure of  “mood” | affect 0.102438  best\_affect 0.085162  affect\_change  affect\_change\_is\_positive |
| Measure of “conversation quality” | how\_enjoyable 0.100970  my\_partner\_was\_clear\_and\_coherent 0.095538  conversationalist 0.080799  *measure of “trust”?*  i\_felt\_close\_to\_my\_partner 0.100229  you\_are\_kind 0.082094  you\_are\_giving 0.081692  you\_are\_friendly 0.080599  i\_like\_you 0.078568 |

Linear Regression

*Measures if there is a linear relationship between var x and var y*

|  |  |
| --- | --- |
| Variable | Plot |
| Your mind wander  My mind wander | A picture containing text, screenshot, font, line  Description automatically generated |
| Affect  Best affect  Affect change  Affect change is positive |  |
| How enjoyable  my\_partner\_was\_clear\_and\_coherent  conversationalist |  |
| i\_felt\_close\_to\_my\_partner  you\_are\_kind  you\_are\_giving  you\_are\_friendly  i\_like\_you |  |

**Binarize Backchannel Count (Continuous 🡪 Discrete), Hypothesis Test**

Partition data based on backchannel count into 2 equal partitions. See if there is a difference in the distributions.

* Cons = losing information, somewhat arbitrary separation
* Pros = conduct hypothesis test

*Null Hypothesis*: There is no difference in mean amount of positive affect change between the group with below average backchannel count and the group with above average backchannel count.

*Alternate Hypothesis:* There is a difference

A screen shot of a graph

Description automatically generated with low confidence

Mean “Affect Change is Positive”: 0.6831, 0.6391 (Difference of ***0.044***)

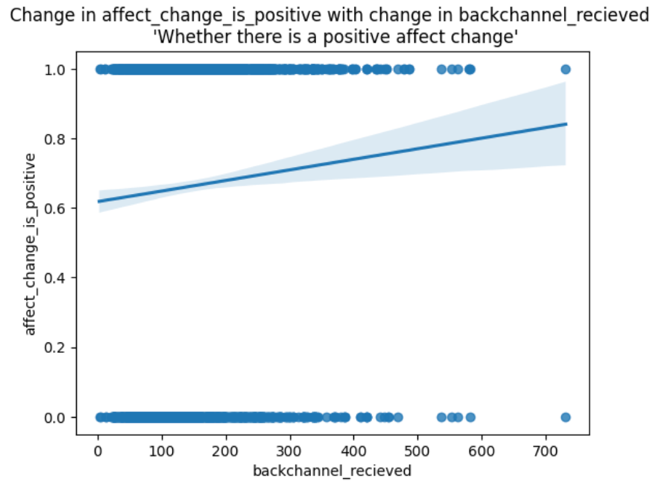
T-statistic=2.652, p value=***0.00803*** (ie. the difference between mean positive affect change change is significant)

* Interpret as 0.04% more likely to have a positive mood change for those with higher backchannels?
* How to approximate no backchannels vs have backchannels? Maybe try lower quantile vs higher quantile as well here.
* Note: here, variable approximately normally distribution shifted to the right. Saw that with high sample size may not matter. Double check

**Binarize Outcome Measure (Categorical 🡪 Binary), Regression with *binary* *outcomes***

* Split outcome into binary groups (ie. proxies for “good” and “bad” conversations, mood change, engagement)
* Keep backchannel count as continuous

Example: *Has Positive Affect Change*

**

R-squared: 0.003

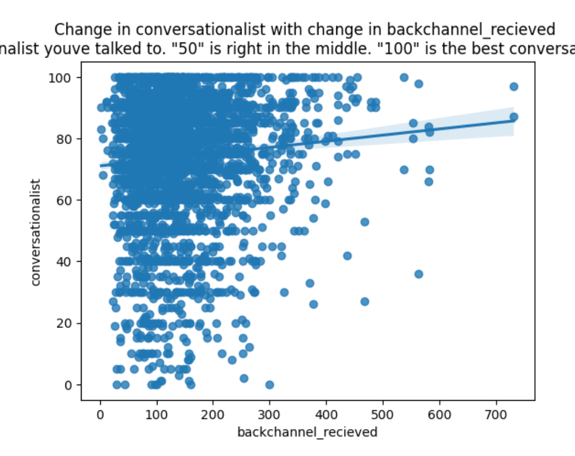
coef std err t P>|t| [0.025 0.975]

Intercept **0.6177** 0.017 36.890 0.000 0.585 0.651

backchannel\_recieved **0.0003** 0.000 2.979 0.003 0.000 0.001

* No backchannels, there is a 0.62 affect change
* Per extra backchannel, there is a 0.0003 affect change (ie. every extra 1000 backchannels increases affect change by 0.3)

*Conversationalist*

A picture containing text, font, screenshot

Description automatically generated

* Recall: this was linear regression on continuous conversationalist
* Now, split into two groups: conversationalist above 75 and ≦75 (“great” and “okay”)

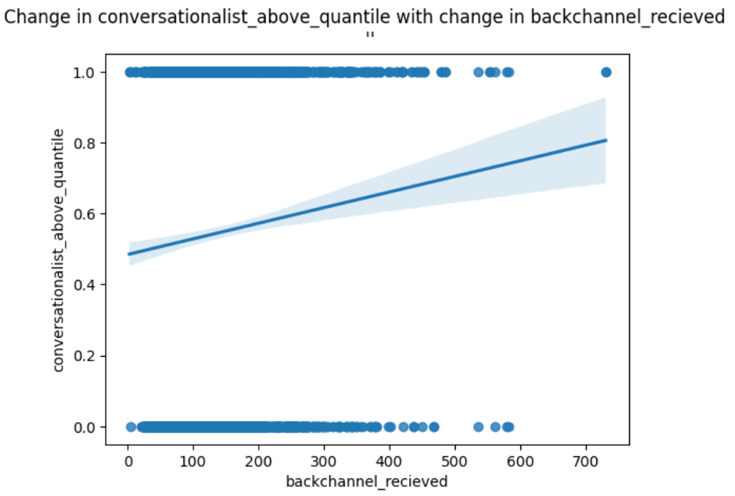
*“Best” vs other conversationalist*

A picture containing screenshot, display, rectangle, text

Description automatically generated

* Number of responses that rated partner conversationalist ≦75 vs >75

A picture containing text, screenshot, diagram, plot

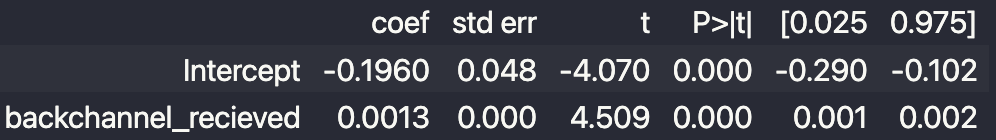
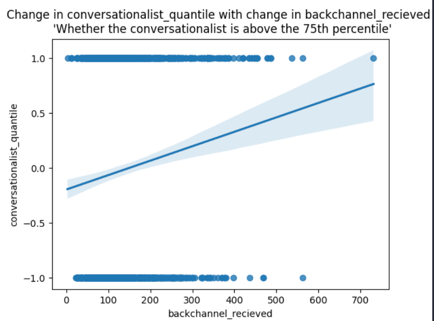
Description automatically generated



* There is 0.0004 change in conversationalist ranking with every extra backchannel (ie. 1000 more backchannels = ranking goes up by 4)

“Bad” vs “Good” conversationalists

* *25% quantile = 65, 75% quantile = 90* (how they split it in the CANDOR paper)



Notes

* Personality test before chatbot
* Backchannel opportunity