

Eve – Online Final Pitch

Prof. Aron Lindberg

Harsh Bhotika

Chirantan Soni

Swapnil Junjare

Shruti Agrawal

Index

- ▶ Problem Statement
- ▶ Prototype Model
- ▶ Correlation with R
- ▶ Graphs with normalized P value
- ▶ Inverse Graph with Negative P values
- ▶ Predicted Independent variables
- ▶ Multivariable linear regression
- ▶ Regression Model
- ▶ Summary
- ▶ Inferences

Problem Statement

- To predict positive sentiment
- What is Positive Sentiment?
- ▶ Sentiment is what the author wants to convey to the reader from the text. It basically refers to the contextual polarity of the statement.
- ▶ The sentences that have a positive polarity attached to them have positive sentiment. Sentiment analysis is very useful in Social media and Marketing since it helps in forecasting and gathering public opinion

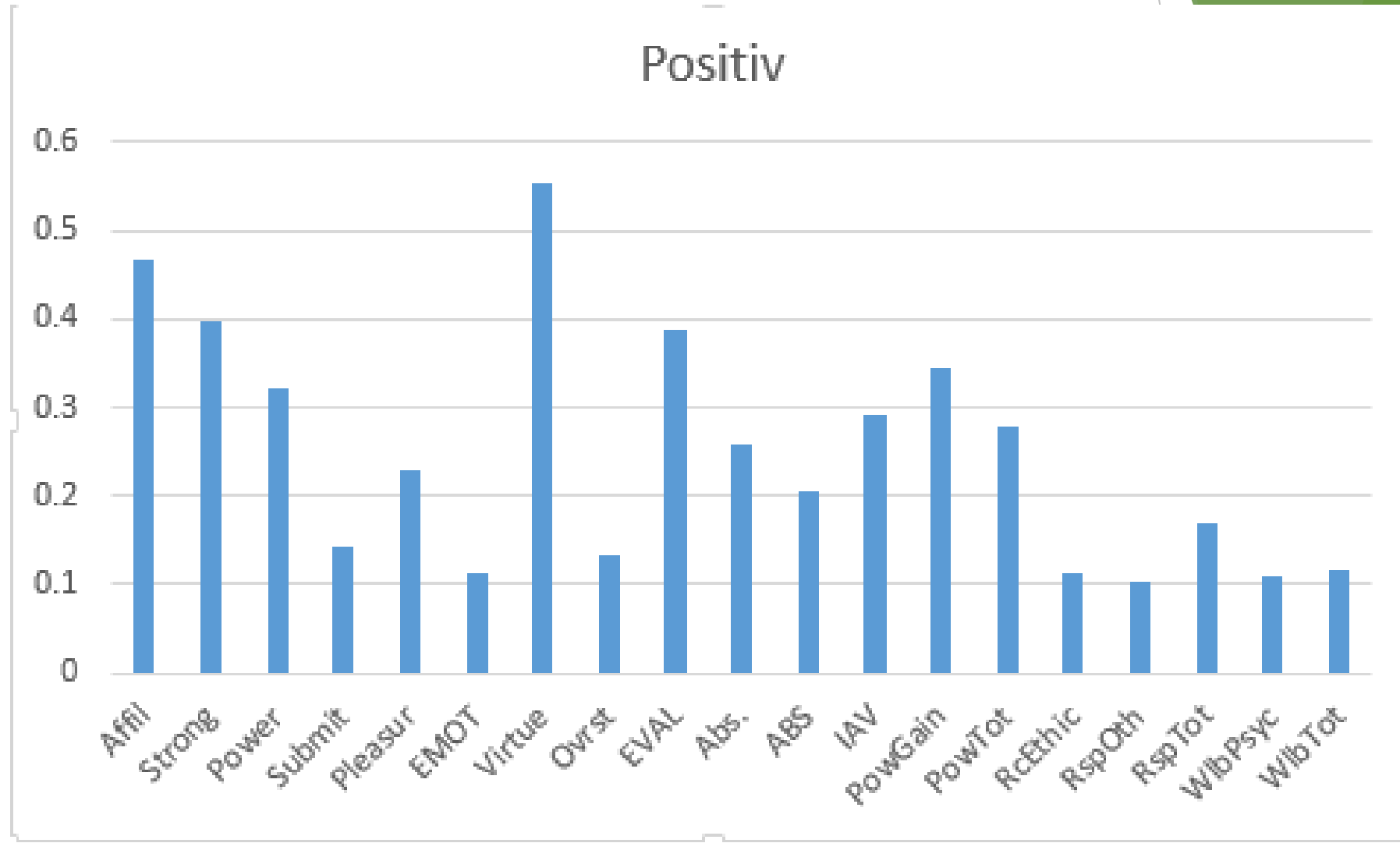
Final model

- ▶ Dataset containing 15000 comments
- ▶ 'Positiv' from dataset decided as a dependent Variable
- ▶ Derived a Correlation matrix using R

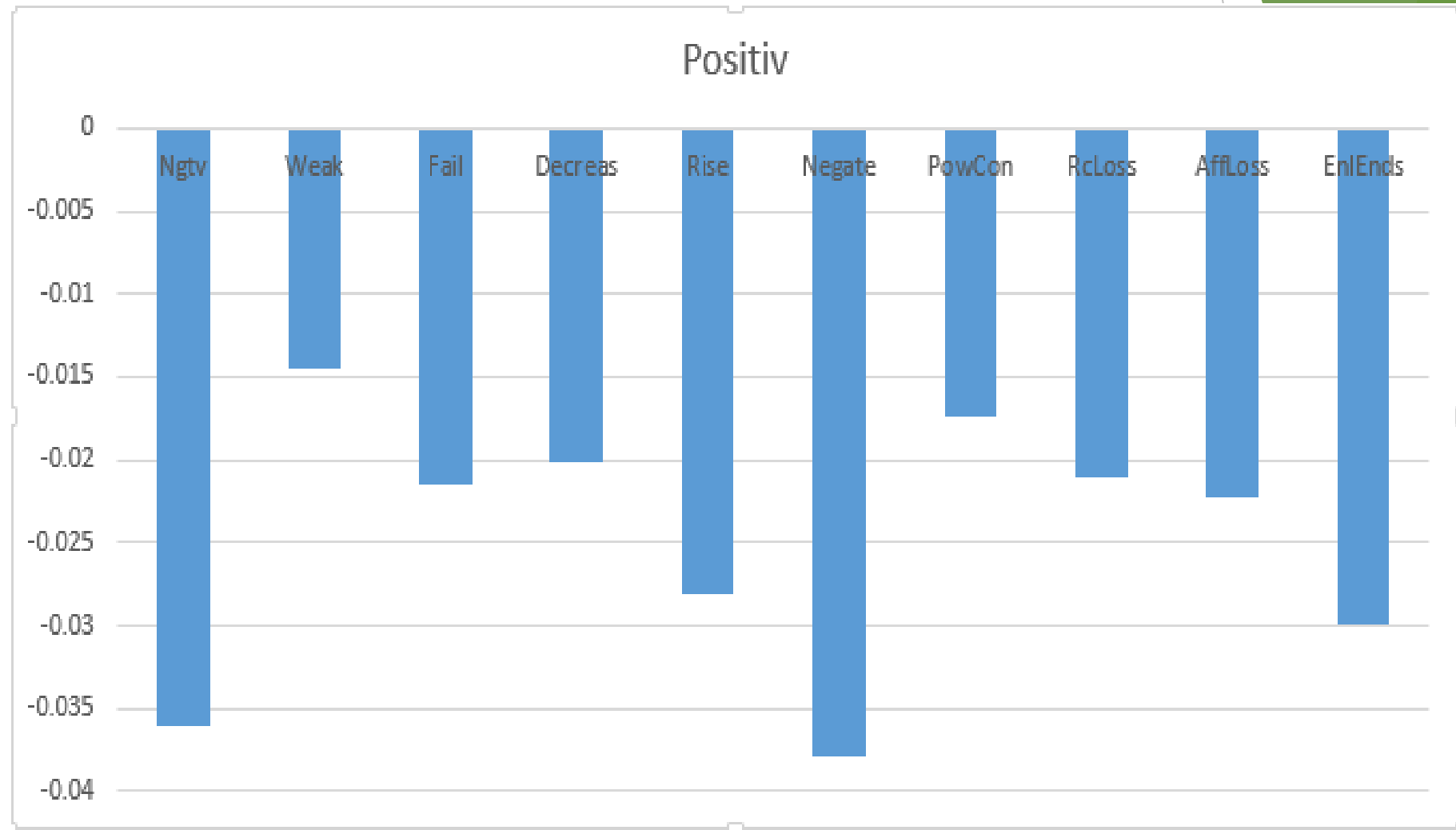
Correlation with R

- ▶ To predict positive sentiment , we have:
 - ▶ Dependent variable - 'Positiv'
 - ▶ Independent variables - Determined using correlation matrix with correlation < 0.05 and logically have effect on the 'Positiv'

Corelation Graph of Normalized data with Low P values



Inverse Graphs with Negative Value



Predicted Independent Variables

1. **Pstv** - positive words
2. **Negativ** - words of negative outlook
3. **Ngtev** - negative words
4. **Affil** - for words indicating affiliation or supportiveness.
5. **Weak** - words implying weakness
6. **Strong** - words implying strength
7. **Active** - words implying an active orientation.
8. **Virtue** - words indicating an assessment of moral approval or good fortune
9. **Passive** - words indicating a passive orientation
10. **Ovrst** - words indicating Overstated
11. **EVAL** - words which imply judgment and evaluation, whether positive or negative, including means-ends judgments.
12. **ABS** - words reflecting tendency to use abstract vocabulary
13. **PowTot** - words for the whole domain
14. **SureLw** - words indicating "a feeling of sureness, certainty and firmness."
15. **IAV** - verbs giving an interpretative explanation of an action
16. **SV** - state verbs describing mental or emotional states

Multivariable Linear Regression Model

Null Hypothesis, H_0 : All the independent Variables have equal effect on Dependent Variable('Positiv')

Alternative Hypothesis, H_1 : Not H_0

Multivariable Linear Regression Model

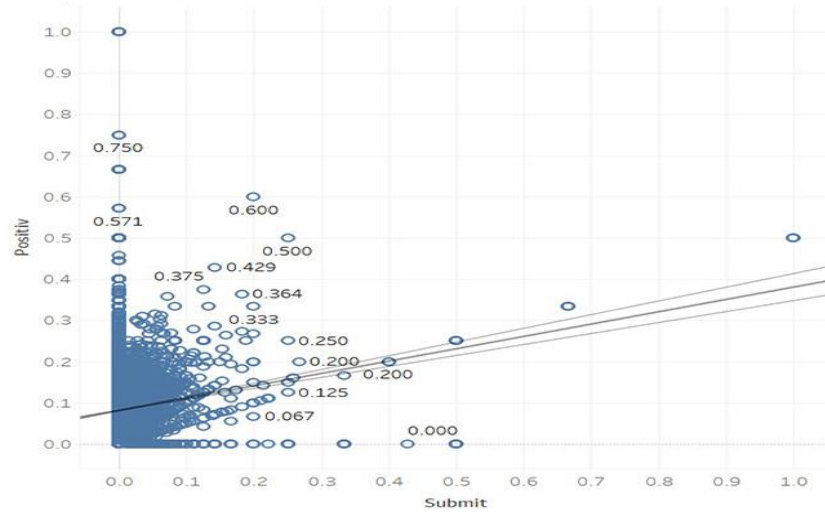
- ▶ Correlation Coefficient, R : Higher the value stronger the relation
- ▶ R Square: 0.985663147
- ▶ Adjusted R Square: 0.985508468
- ▶ F-Ratio = $MS_{\text{Regression}}/MS_{\text{Residual}}$
- ▶ F-Ratio = 6372.294782, Higher the F value show that model is the good fit for the data

Regression Model

P-value: < 0.0001

Equation: $\text{Positiv} = 0.299146 \cdot \text{Submit} + 0.0810802$

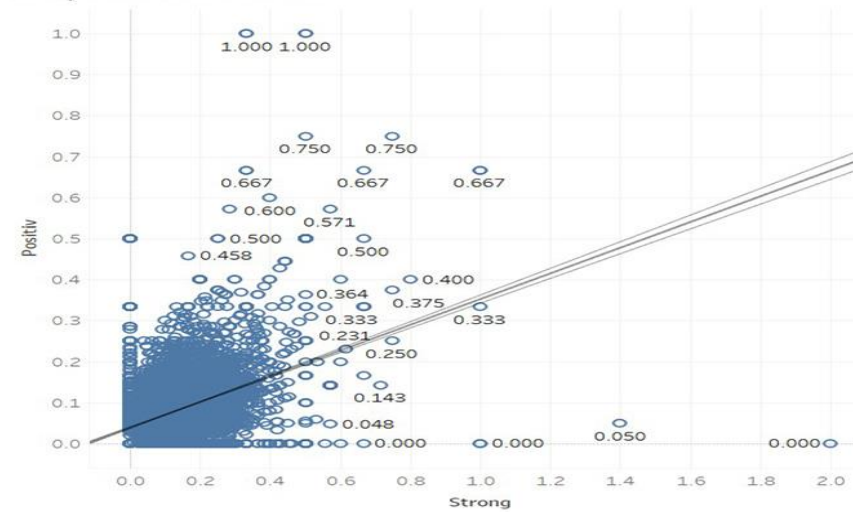
R-Squared: 0.0203466



P-value: < 0.0001

Equation: $\text{Positiv} = 0.313357 \cdot \text{Strong} + 0.0386234$

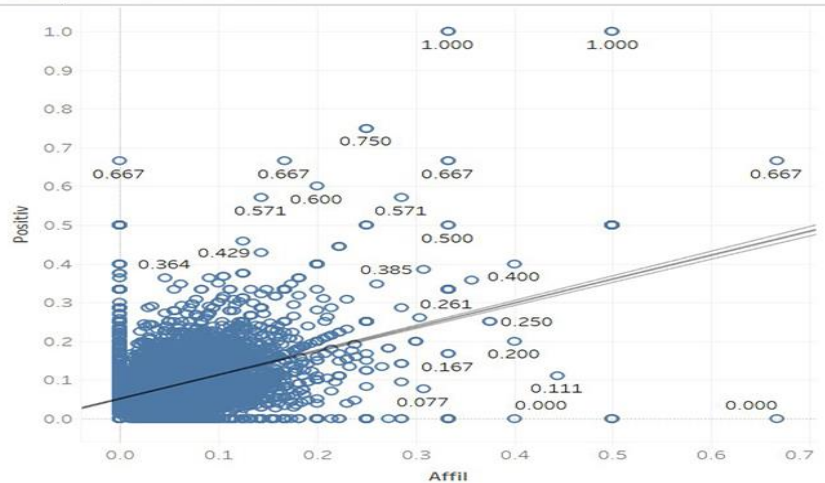
R-squared: 0.158354



P-value: < 0.0001

Equation: $\text{Positiv} = 0.618023 \cdot \text{Affil} + 0.0509096$

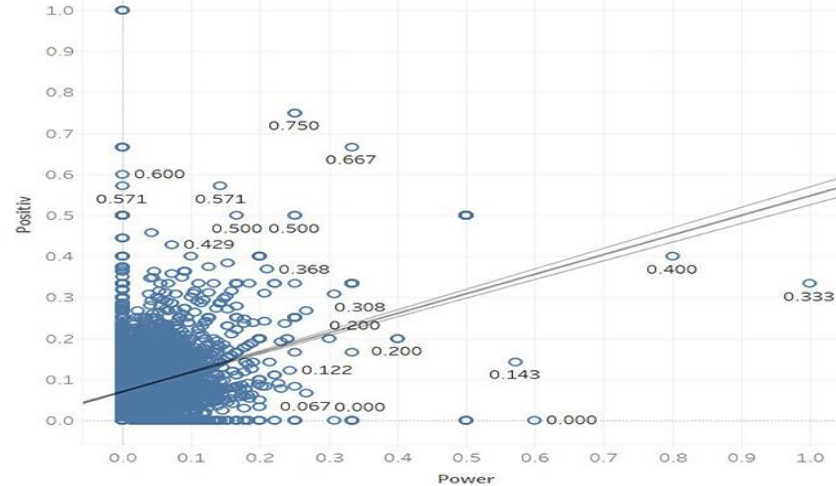
R-Square: 0.217659



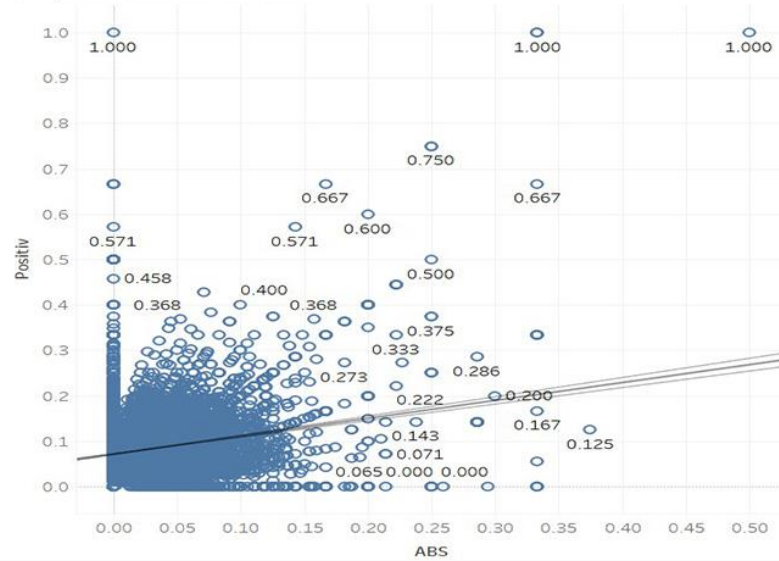
P-value: < 0.0001

Equation: $\text{Positiv} = 0.477248 \cdot \text{Power} + 0.0696898$

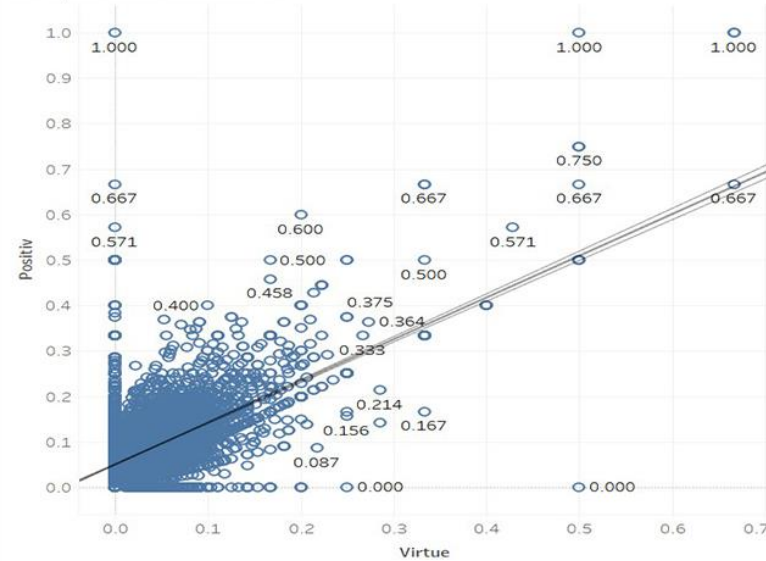
R-Squared: 0.10269



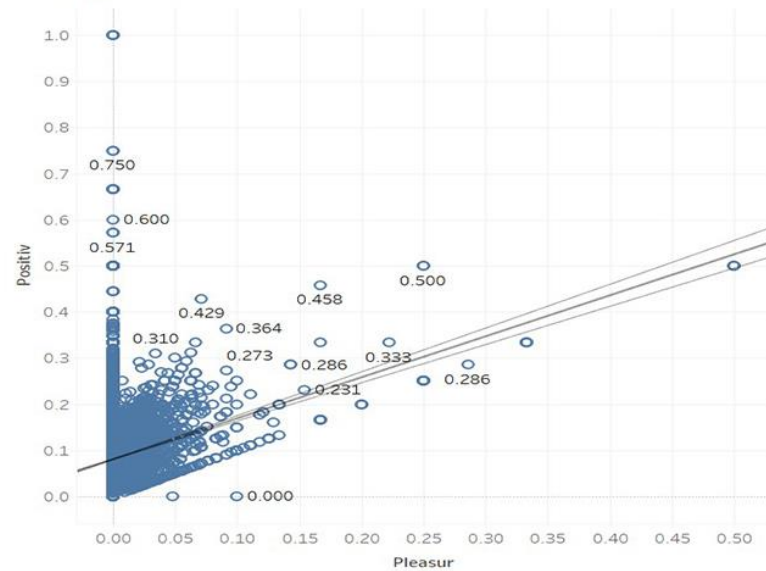
P-value: < 0.0001
Equation: $\text{Positiv} = 0.393836 \cdot \text{ABS} + 0.0713038$
R-squared : 0.0422061



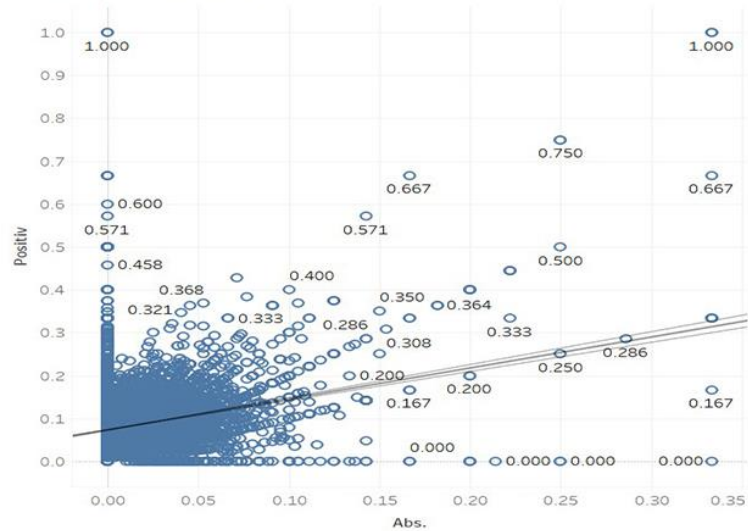
P-value: < 0.0001
Equation: $\text{Positiv} = 0.918346 \cdot \text{Virtue} + 0.0499119$
R-Squared: 0.30728



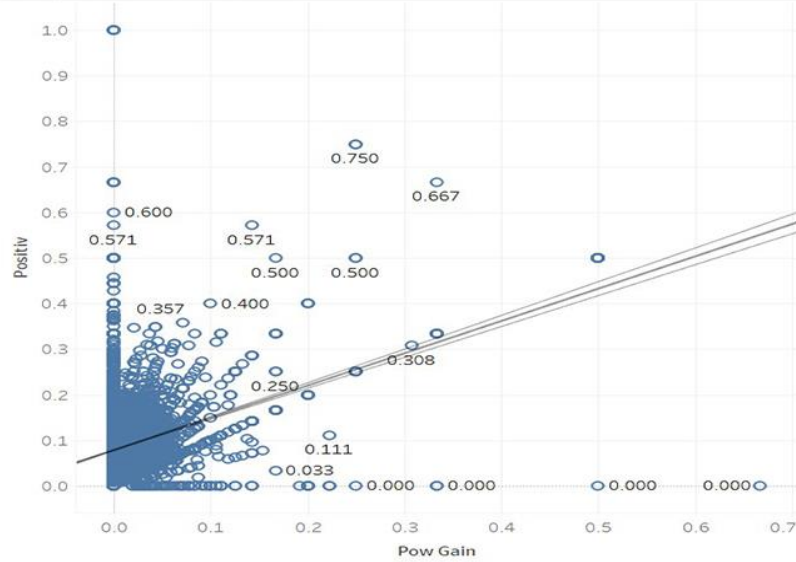
P-value: < 0.0001
Equation: $\text{Positiv} = 0.888391 \cdot \text{Pleasur} + 0.0802791$
R-squared: 0.05255



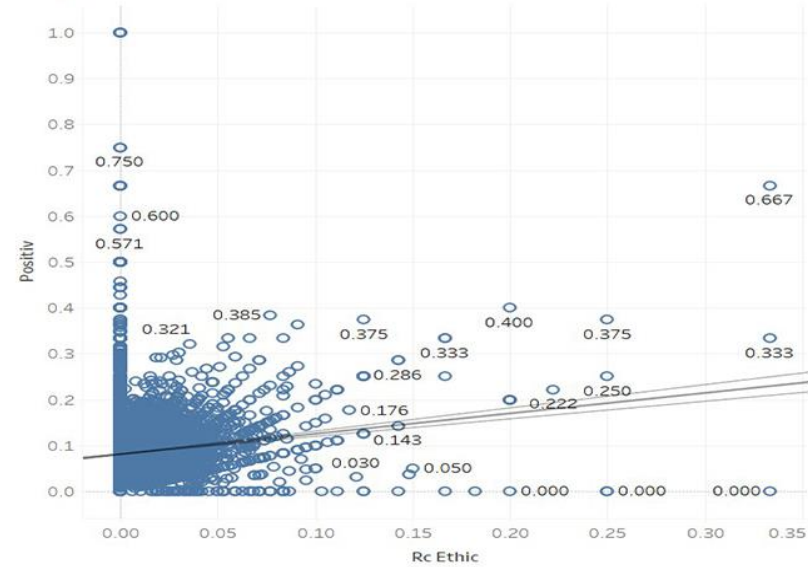
P-value: < 0.0001
Equation: $\text{Positiv} = 0.721967 \cdot \text{Abs.} + 0.0731213$
R-Squared: 0.0662299



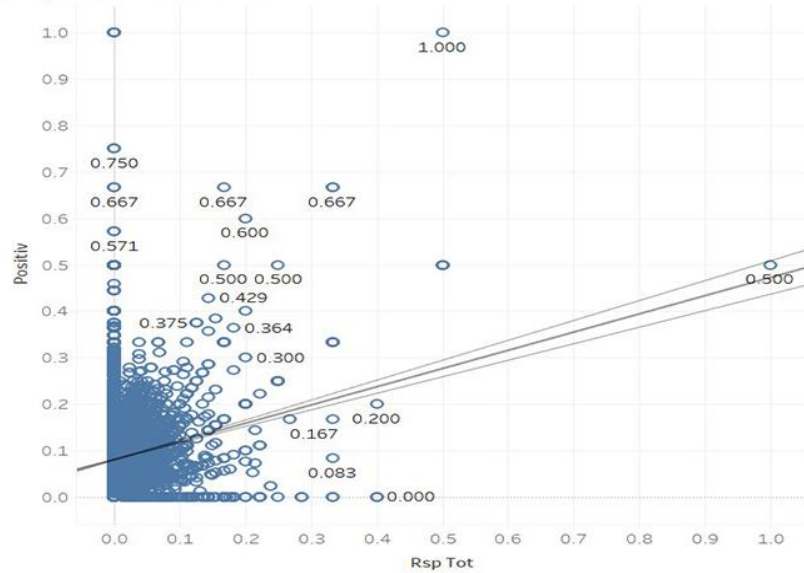
P-value: < 0.0001
Equation: Positiv = 0.709018*Pow Gain + 0.0780609
R-squared: 0.118628



P-value: < 0.0001
Equation: Positiv = 0.445106*Rc Ethic + 0.0806857
R-Squared : 0.0129478



P-value: < 0.0001
Equation: Positiv = 0.709018*Pow Gain + 0.0780609
R-squared : 0.0289119



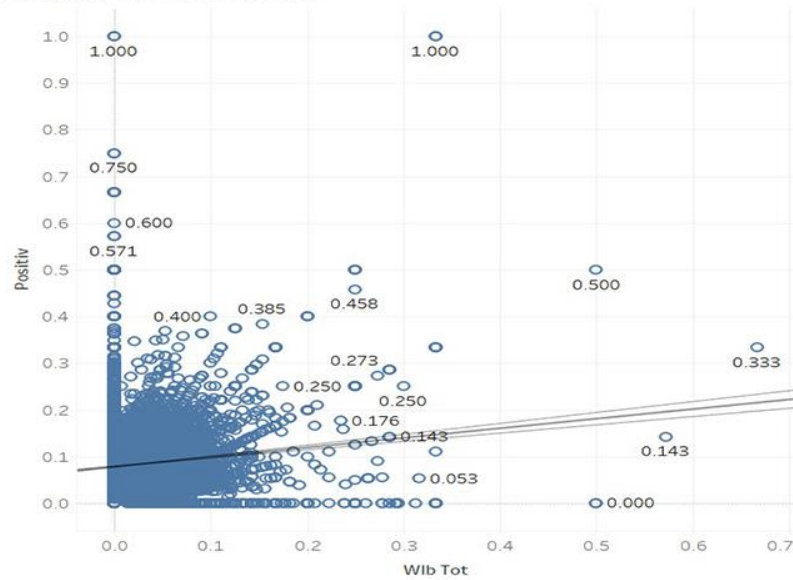
P-value: < 0.0001
Equation: Positiv = 0.286322*IAV + 0.0516713
R-Squared : 0.0848169



P-value: < 0.0001

Equation: Positiv = 0.205509*Wlb Tot + 0.078118

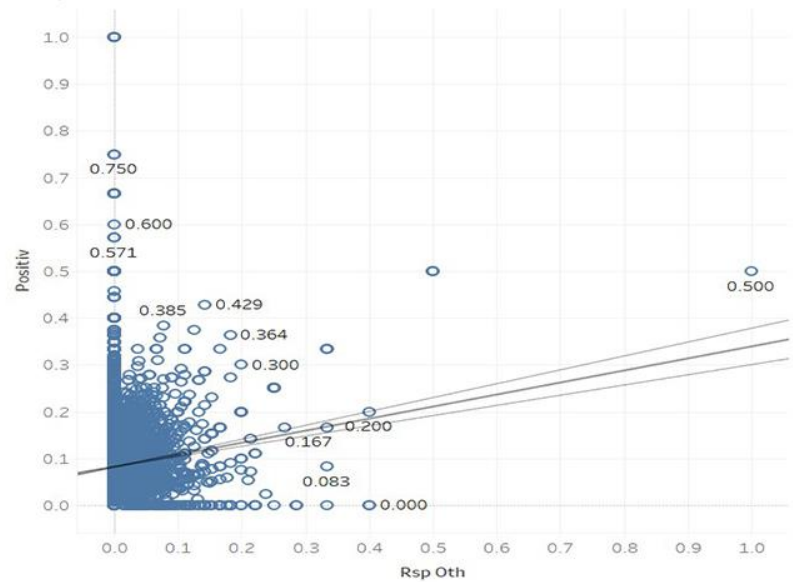
R-Squared: 0.0134986



P-value: < 0.0001

Equation: Positiv = 0.25714*Rsp Oth + 0.0818323

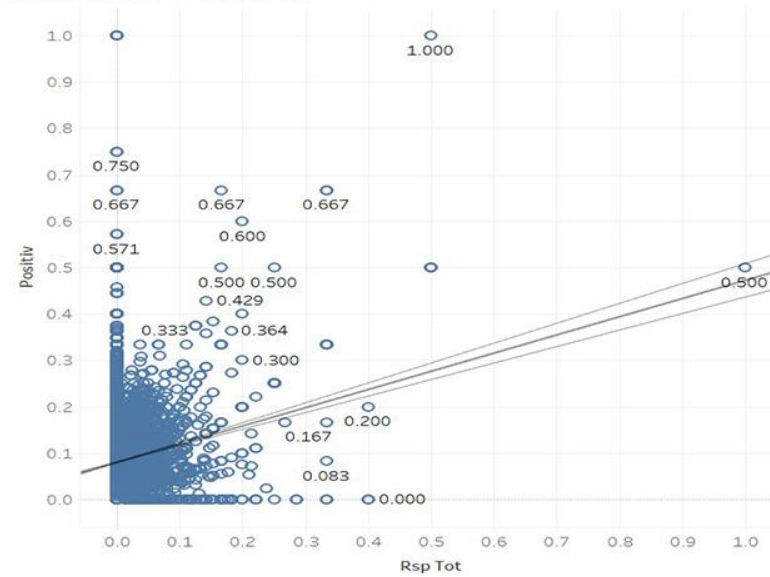
R-squared: 0.0108855



P-value: < 0.0001

Equation: Positiv = 0.392132*Rsp Tot + 0.079848

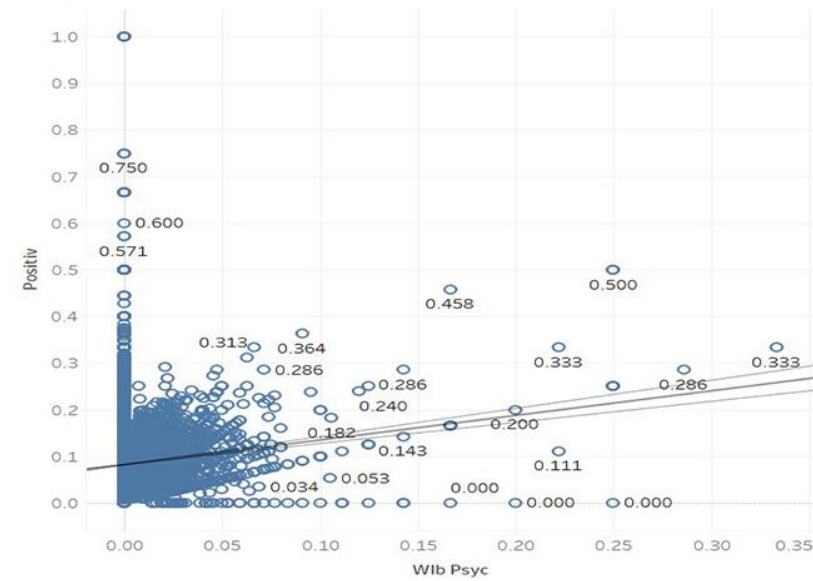
R-Squared : 0.0289119



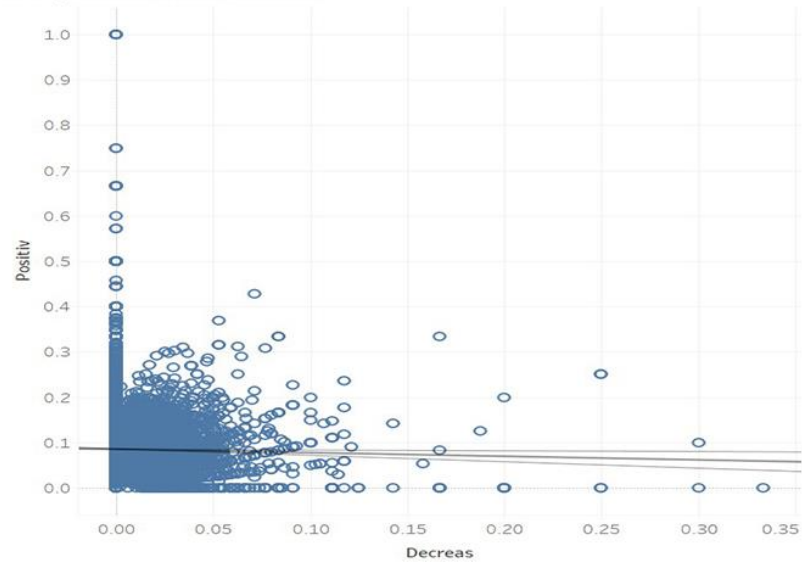
P-value: < 0.0001

Equation: Positiv = 0.527733*Wlb Psyc + 0.0819904

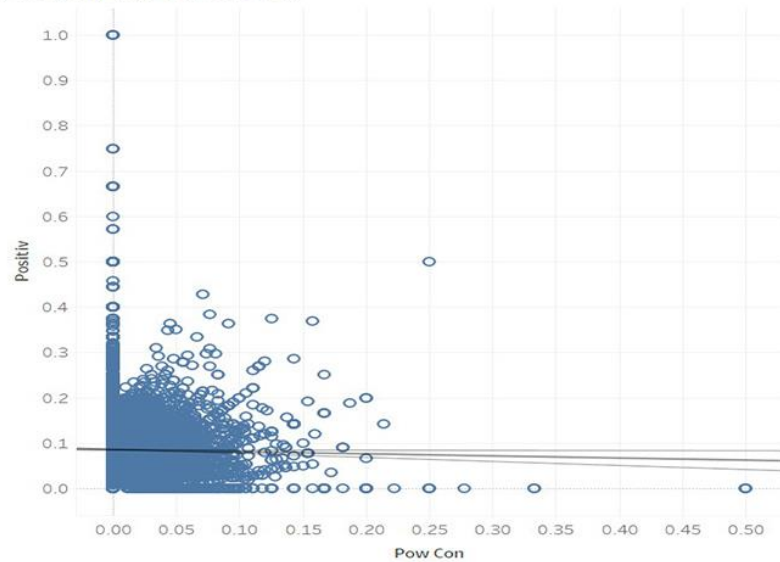
R-squared: 0.0120847



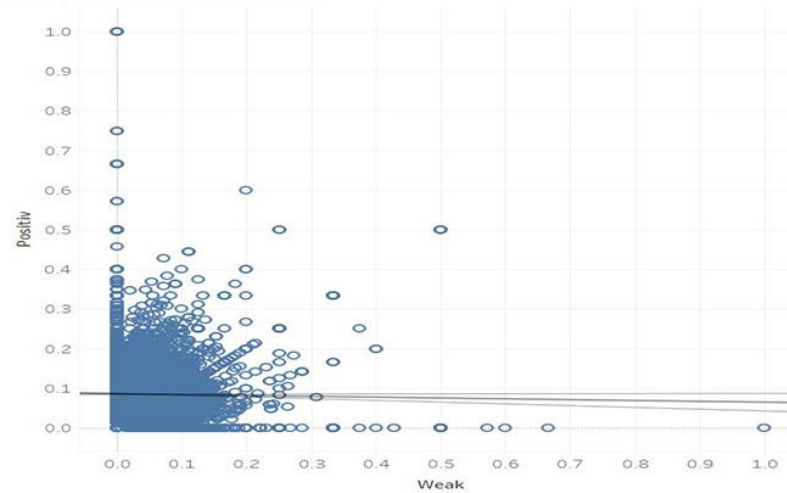
P-value: 0.0134956
Equation: Positiv = $-0.0792657 \cdot \text{Decreas} + 0.0853285$
R-squared: 0.0004068



P-value: 0.0336815
Equation: Positiv = $-0.0463354 \cdot \text{Pow Con} + 0.0852847$
R-Squared: 0.0003007



P-value: 0.0751414
Equation: Positiv = $-0.0210129 \cdot \text{Weak} + 0.0856023$
R-Squared: 0.0002111



Variables	Meaning	R ²	R
Positiv	words of positive outlook		
Affil	for words indicating affiliation or supportiveness	0.217659	0.46654
Strong	words implying strength	0.158354	0.39794
Power	indicating a concern with power, control or authority.	0.10269	0.32045
Weak	words implying weakness	0.0002111	-0.0145
Submit	connoting submission to authority or power, dependence on others, vulnerability to others, or withdrawal.	0.0203466	0.14264
Pleasur	words indicating the enjoyment of a feeling, including words indicating confidence, interest and commitment.	0.05255	0.22925
Virtue	words indicating an assessment of moral approval or good fortune	0.30728	0.55433
Decreas	words for processes found in nature, birth to death	0.0004068	-0.0202
Abs.	words reflecting tendency to use abstract vocabulary	0.0662299	0.25735
ABS	words reflecting tendency to use abstract vocabulary	0.0422061	0.20544
IAV	verbs giving an interpretative explanation of an action	0.0848169	0.29123
PowGain	words about power increasing	0.118628	0.34442
PowCon	words for ways of conflicting.	0.0003007	-0.0173
RcEthic	words of values concerning the social order.	0.0129478	0.11379
RspTot		0.0289119	0.10433
RspOth	words regarding respect that are neither gain nor loss	0.0108855	0.17004
WlbPsys	words connoting the psychological aspects of well being, including its absence.	0.0120847	0.10993
WlbTot	words in well-being domain	0.0134986	0.11618

Multivariable Linear Regression Model

- ▶ **P-value:** Is the calculated Probability, if $P\text{-value} < 0.05$ we can reject the null hypothesis.
- ▶ Observing P-value for each Independent variables we found that some of the variables have $p\text{-value} > 0.05$, so they do not add much in the predictive power.
- ▶ So accepting only those variables who have $P\text{-value} < 0.05$

- Positiv
- Affil
- Strong
- Power
- Weak
- Submit
- Pleasur
- Virtue

- Decreas
- Abs.
- ABS
- IAV
- PowGain
- PowCon
- RcEthic
- RspTot
- RspOth
- WlbPsync
- WlbTot

Summary

Residual standard error: 0.0222 on 14969 degrees of freedom

Multiple R-squared: 0.8723, Adjusted R-squared: 0.872 F-statistic: 3408 on 30 and 14969 DF, p-value: $< 2.2e-16$

In the regression model the coefficient of determination states that regression line fits 87% in the data.

So the Independent variables given help in determining positive sentiment on the forums.

Also the highlighted words are the ones that have inverse correlation can help in determining how their use in the discuss can cause negative effect on forums

PowGain: addition of new swarms, new ships and redesigning also addition of (Kronos Release - 2014 Q3 & Q4), addition of set of characters (Goonswarm - 2012 Q3), Intense battles (Battle of Asakai 2013 Q1), In game Update releases (Bloodbath of B-R5RB 2014 Q1, Inacarna Release 2011 Q3)

PowCon: In game conflicts (summer of rage - Inacarna Release 2011 Q3 & Q4), In game battles that lead to conflict (Goonswarm - 2012 Q3, Bloodbath of B-R5RB 2014 Q1, The battle of M-OEE8 2016 Q2)

Emotive Markers: Virtue and pleasure are related to changes made to In game features (Citadel Introduced new player owned stations, force auxiliaries, super weapons etc)

- ▶ Thus, the regression model helped in analysing the areas of in game features for the Eve Online game developers
- ▶ Based on the updates rolled out for the game and based on user feedbacks in relevance to the roll outs analysis can be done and areas of further game development can be determined
- ▶ Which crowd of people like what and what geographical areas like which aspect of the game can be predicted
- ▶ Better the predictions better the revenues for the business.

Thank You!
Questions?