Project Proposal

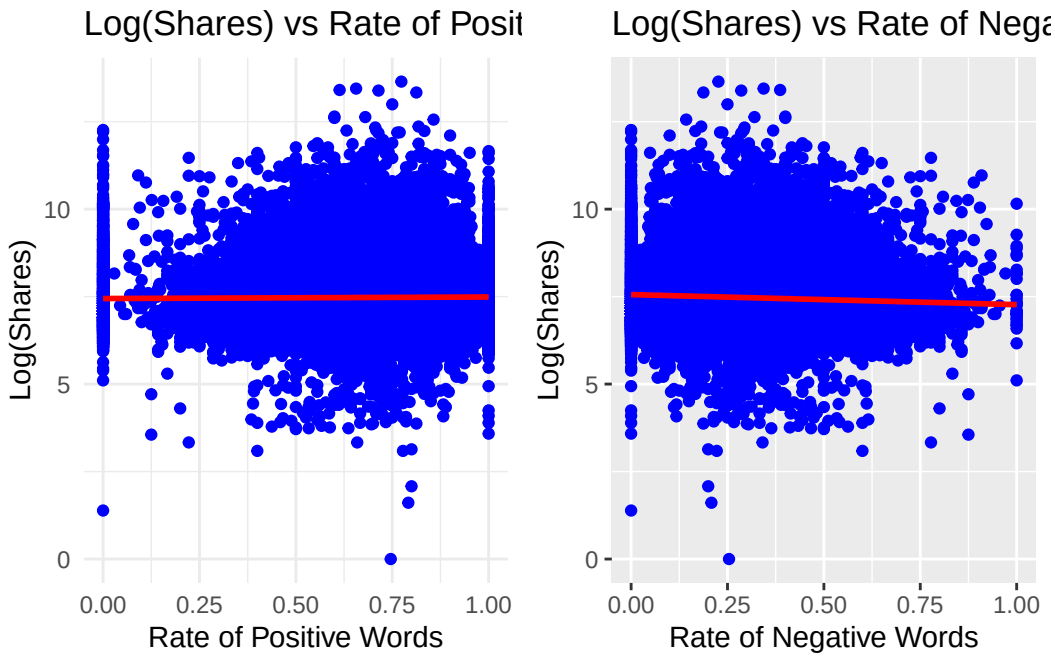
The BEST Fit - Olivia Encarcion, Leo Yang, Philip Lin, Allison Yang

```
library(tidyverse)
library(tidymodels)
library(ggplot2)
library(dplyr)
library(gridExtra)
library(knitr)
if (!requireNamespace("car", quietly = TRUE)) {
  install.packages("car")
}
library(car)

newsdf <- read_csv("data/OnlineNewsPopularity.csv")
```

```
p1 <- ggplot(newsdf, aes(x = rate_positive_words, y = log(shares))) +
  geom_point( color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  labs(title = "Log(Shares) vs Rate of Positive Words",
       x = "Rate of Positive Words",
       y = "Log(Shares)") +
  theme_minimal()

p2 <- ggplot(newsdf, aes(x = rate_negative_words, y = log(shares))) +
  geom_point(color = "blue") +
  geom_smooth(method = "lm", color = "red") +
  labs(title = "Log(Shares) vs Rate of Negative Words",
       x = "Rate of Negative Words",
       y = "Log(Shares)")
grid.arrange(p1, p2, ncol = 2)
```




```
model <- lm(log(shares) ~ rate_negative_words + rate_positive_words, data =newsdf )
  tidy(model)|>
  kable(digits=3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	7.655	0.027	283.712	0
rate_negative_words	-0.360	0.035	-10.206	0
rate_positive_words	-0.113	0.029	-3.894	0

```
vif_values <- vif(model)
print("Variance Inflation Factors:")
```

```
[1] "Variance Inflation Factors:"
```

```
print(vif_values)
```

```
rate_negative_words rate_positive_words
           1.392604            1.392604
```