Data 115 Final Project by Jaehong Lee

Jaehong Lee

2023-05-05

library(tidyverse) library(rmarkdown)

superbowl_ads_dataframe <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/superbowl-ads.csv", header = T) View(superbowl_ads_dataframe) head(superbowl_ads_dataframe) summary(superbowl_ads_dataframe)

predictor1 <- superbowl_ads_dataframe[,5] predictor2 <- superbowl_ads_dataframe[,6] predictor3 <- superbowl_ads_dataframe[,7] predictor4 <- superbowl_ads_dataframe[,8] predictor5 <- superbowl_ads_dataframe[,10] predictor7 <- superbowl_ads_dataframe[,11]

#model <- glm(predictor1 ~ predictor2 + predictor3 + predictor4 + predictor5 + predictor6 + predictor7, data = superbowl_ads_dataframe, family = binomial) # Extract Specific columns. Alt + - will print out an arrow. #data.frame is a function for extracting specific columns. extract1 <-

 $\label{lem:data-frame} data.frame(superbowl_ads_dataframe) year, superbowl_ads_dataframe) print(extract1) view(extract1)$

budlight_sales_dataframe <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/monthly Budlight sales.csv") budweiser_sales_dataframe <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/montly sales for Budweiser.csv") hyundai_sales_dataframe <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/monthly sales for Hyundai cars.csv") kia_sales_dataframe <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/monthly sales for Kia cars.csv") toyota_sales_dataframe <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/monthly sales for Toyota cars.csv")

view(budlight_sales_dataframe) view(budweiser_sales_dataframe)
view(hyundai sales dataframe) view(kia sales dataframe) view(toyota sales dataframe)

ggplot(hyundai_sales_dataframe) ggplot(hyundai_sales_dataframe, mapping = $aes(x=month, y = sales, color = factor(year))) + geom_point() + labs(title = "Changes of Hyundai sales in the US from 2015 to 2019")$

ggplot(budlight_sales_dataframe) ggplot(budlight_sales_dataframe, mapping = aes(x= month, y = sales, color = factor(year))) + geom_point() + labs(title = "Changes of Budlight sales in the US from 2015 to 2019")

ggplot(budweiser_sales_dataframe) ggplot(budweiser_sales_dataframe, mapping = aes(x = month, y = sales, color = factor(y = month)) + geom_point() + labs(y = month) = "Changes of Budweiser sales in the US from 2015 to 2019")

budlight_data2016 <- budlight_sales_dataframe[12:23,] ggplot(budlight_data2016) ggplot(budlight_data2016, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Budlight sales after 1 times of ads during the Superbowl 2016")

budlight_data2017 <- budlight_sales_dataframe[24:35,] ggplot(budlight_data2017) ggplot(budlight_data2017, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Budlight sales after 0 times of ads during the Superbowl 2017")

budlight_data2018 <- budlight_sales_dataframe[36:47,] ggplot(budlight_data2018) ggplot(budlight_data2018, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Budlight sales after 2 times of ads during the Superbowl 2018")

budlight_data2019 <- budlight_sales_dataframe[48:59,] ggplot(budlight_data2019) ggplot(budlight_data2019, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Budlight sales after 4 times of ads during the Superbowl 2019")

budweiser_data2016 <- budweiser_sales_dataframe[12:23,] ggplot(budweiser_data2016) ggplot(budweiser_data2016, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.",

"May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(y="sales in millions", title = "Budweiser sales after 2 times of ads during the Superbowl 2016")

budweiser_data2017 <- budweiser_sales_dataframe[24:35,] ggplot(budweiser_data2017) ggplot(budweiser_data2017, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(y="sales in millions", title = "Budweiser sales after 1 times of ads during the Superbowl 2017")

budweiser_data2018 <- budweiser_sales_dataframe[36:47,] ggplot(budweiser_data2018) ggplot(budweiser_data2018, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(y="sales in millions", title = "Budweiser sales after 1 times of ads during the Superbowl 2018")

budweiser_data2019 <- budweiser_sales_dataframe[48:59,] ggplot(budweiser_data2019) ggplot(budweiser_data2019, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(y="sales in millions", title = "Budweiser sales after 1 times of ads during the Superbowl 2019")

hyundai_data2016 <- hyundai_sales_dataframe[12:23,] ggplot(hyundai_data2016) ggplot(hyundai_data2016, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Hyundai sales after 2 times of ads during the Superbowl 2016")

hyundai_data2017 <- hyundai_sales_dataframe[24:35,] ggplot(hyundai_data2017) ggplot(hyundai_data2017, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Hyundai sales after 1 times of ads during the Superbowl 2017")

hyundai_data2018 <- hyundai_sales_dataframe[36:47,] ggplot(hyundai_data2018) ggplot(hyundai_data2018, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Hyundai sales after 0 times of ads during the Superbowl 2018")

hyundai_data2019 <- hyundai_sales_dataframe[48:59,] ggplot(hyundai_data2019) ggplot(hyundai_data2019, mapping = aes(x= month, y = sales, group = 1, color =

factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Hyundai sales after 1 times of ads during the Superbowl 2019")

hyundai_data2016 <- hyundai_sales_dataframe[12:23,] ggplot(hyundai_data2016) ggplot(hyundai_data2016, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Kia sales after 1 times of ads during the Superbowl 2016")

hyundai_data2017 <- hyundai_sales_dataframe[24:35,] ggplot(hyundai_data2017) ggplot(hyundai_data2017, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Kia sales after 1 times of ads during the Superbowl 2017")

hyundai_data2018 <- hyundai_sales_dataframe[36:47,] ggplot(hyundai_data2018) ggplot(hyundai_data2018, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Kia sales after 1 times of ads during the Superbowl 2018")

hyundai_data2019 <- hyundai_sales_dataframe[48:59,] ggplot(hyundai_data2019) ggplot(hyundai_data2019, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Kia sales after 2 times of ads during the Superbowl 2019")

```
toyota_data2016 <- toyota_sales_dataframe[12:23,] ggplot(toyota_data2016) ggplot(toyota_data2016, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Toyota sales after 1 times of ads during the Superbowl 2016")
```

toyota_data2017 <- toyota_sales_dataframe[24:35,] ggplot(toyota_data2017) ggplot(toyota_data2017, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Toyota sales after 0 times of ads during the Superbowl 2017")

toyota_data2018 <- toyota_sales_dataframe[36:47,] ggplot(toyota_data2018) ggplot(toyota_data2018, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Toyota sales after 1 times of ads during the Superbowl 2018")

toyota_data2019 <- toyota_sales_dataframe[48:59,] ggplot(toyota_data2019) ggplot(toyota_data2019, mapping = aes(x= month, y = sales, group = 1, color = factor(year))) + geom_point() + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec.")) + geom_line() + labs(title = "Toyota sales after 2 times of ads during the Superbowl 2019")

influencers_in_ads_survey <- read.csv("C:/Users/USER/Desktop/2023 Spring Semester CPTS 121/Csv data to use for a data final project/Whatsgoodly - Thought Catalog Influencers (1).csv") view(influencers_in_ads_survey)

ads_survey_result <- influencers_in_ads_survey[,c(1, 4, 5)] view(ads_survey_result) summary(ads_survey_result) survey_responses_when <- influencers_in_ads_survey[c(4004:4774),c(1, 4, 5)] survey_responses_influencer <- influencers_in_ads_survey[c(1:2915),c(1, 4, 5)] survey_responses_platform <- influencers_in_ads_survey[c(609:1863),c(1, 4, 5)]

view(survey_responses_when) view(survey_responses_influencer)
view(survey_responses_platform)

survey_responses_platform <- data.frame(Count = 1:251) # facebook 31626 sum(survey_responses_platform\$Count) survey_responses_platform <- data.frame(Count = 252:502) # instagram 94627 sum(survey_responses_platform\$Count) survey_responses_platform <- data.frame(Count = 503:753) # none 157628 sum(survey_responses_platform\$Count) survey_responses_platform <- data.frame(Count = 754:1004) # snapchat 220629 sum(survey_responses_platform\$Count) survey_responses_platform <- data.frame(Count = 1005:1255) # twitter 283630 sum(survey_responses_platform\$Count) survey_responses_when <- data.frame(Count = 1:247) # Yes very much so. 30628 sum(survey_responses_when\$Count) survey_responses_when <- data.frame(Count = 248:494) # somewhat 91637 sum(survey_responses_when\$Count) survey_responses_when <- data.frame(Count = 510:756) # No, its whatever. 156351 sum(survey_responses_when\$Count)

survey responses influencer <- data.frame(Count = 1:251) # twitter 31626 sum(survey responses influencer\$Count) survey responses influencer <data.frame(Count = 252:402) # This is lame 49377 sum(survey_responses_influencer\$Count) survey_responses_influencer <-</pre> data.frame(Count = 1444:1706) # snapchat 414225 sum(survey_responses_influencer\$Count) survey_responses_influencer <-</pre> data.frame(Count = 1707:1858) # other (comment) when an influencer 270940 sum(survey responses influencer\$Count) survey responses influencer <data.frame(Count = 1859:2109) # None (social media) 497984 sum(survey responses influencer\$Count) survey responses influencer <data.frame(Count = 2110:2261) # Is the product cool? 332196 sum(survey responses influencer\$Count) survey responses influencer <data.frame(Count = 2262:2512) # instagram 599137 sum(survey responses influencer\$Count) survey responses influencer <data.frame(Count = 2513:2664) # Get that money! 393452 sum(survey responses influencer\$Count) survey responses influencer <data.frame(Count = 2665:2915) # facebook 700290 sum(survey responses influencer\$Count)

which_influencer_responses <- c(49377, 393452, 332196, 270940) options(scipen = 999) influencer_responses_labels <- c("This is lame", "Get that money!", "Is the product cool?", "other (comments)") barplot(which_influencer_responses) barplot(which_influencer_responses, names.arg = influencer_responses_labels, xlab = "Types of responses", ylab = "Number of responses", main = "What do you think when an influencer is obviously selling a product?", col = 'green') #(A total of 1045965 responses)

##Create a new dataframe budlight_linear_regression_data <- data.frame(how_many_times_aired = c(0:4), average_annual_revenue = c(50024, 42282, 43188, 41680, 52329), average_first_quarter_revenue = c(11688, 9202, 10303, 10264, 12224), stringsAsFactors = FALSE)

view(budlight_linear_regression_data)

plot(budlight_linear_regression_data how_m any $_t$ imes $_a$ ired, $budlight_linear_r$ egression $_d$ ataa verage_annual_revenue, main = "Effects on how many ads aired during Super Bowlon Budlight company's revenue", xlab = "Number of times ads aired during Super Bowl", ylab = "Average revenue from that year (Millions of US \$)", pch = 19, col = 'brown', cex = 3)

ggplot(budlight_linear_regression_data, mapping = aes(x=how_many_times_aired, y = average_annual_revenue)) + geom_point(size = 10, color = 'brown') + geom_smooth(method = "lm", se = FALSE) + labs(x = "Number of times ads aired during Super Bowl", y = "Average revenue from that year (Millions of US \$)", title = "Effects on how many ads aired during Super Bowlon Budlight company's annual revenue") lm_result_budlight_annual <- lm(formula = how_many_times_aired ~ average_annual_revenue, data = budlight_linear_regression_data) summary(lm_result_budlight_annual)

$$\begin{split} & ggplot(budlight_linear_regression_data, mapping = aes(x=how_many_times_aired, y = average_first_quarter_revenue)) + geom_point(size = 10, color = 'yellow') + geom_smooth(method = "lm", se = FALSE) + labs(x = "Number of times ads aired during Super Bowl", y = "Average first quaterly revenue from that year (Millions of US $)", title = "Effects on how many ads aired during Super Bowlon Budlight company's first quaterly revenue") lm_result_budlight_quaterly <- lm(formula = how_many_times_aired ~ average_first_quarter_revenue, data = budlight_linear_regression_data) summary(lm_result_budlight_quaterly) \\ \end{aligned}$$

#dt <- survey_responses_when(x = c("No, it's whatever", "Somewhat", "","C", "C", "B"))
#ggplot(data = hyundai sales dataframe)</pre>

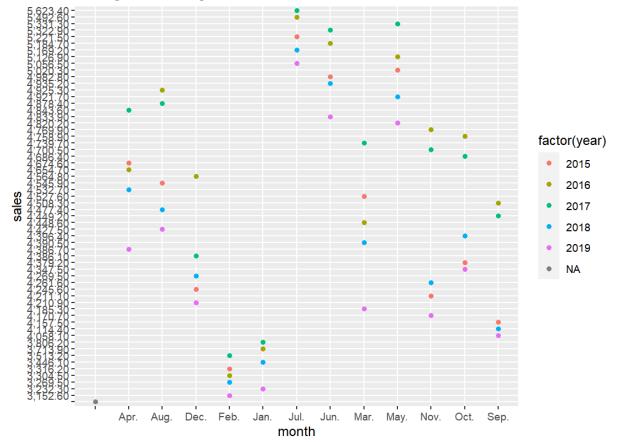
#ggplot(data = hyundai_sales_dataframe) + #geom_point(mapping = aes(x = month, y =
sales, color = year)) + scale_x_discrete(limits = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jun.",
"Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec."))

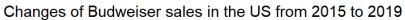
#hyundai_sales_dataframe $month < -factor(hyundai_sales_dataframe$ month, levels = c("Jan.", "Feb.", "Mar.", "Apr.", "May.", "Jul.", "Aug.", "Sep.", "Oct.", "Nov.", "Dec."))

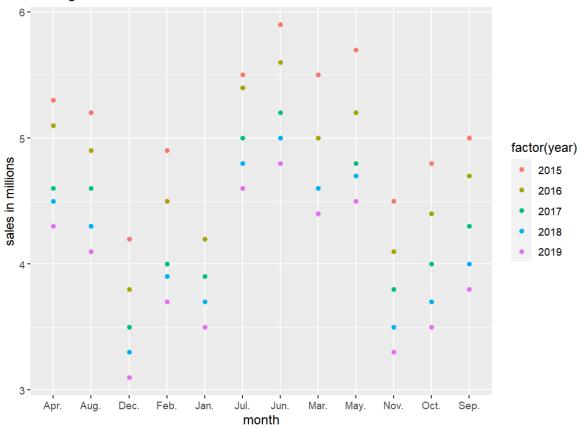
#ggplot(data = hyundai_sales_dataframe) + # geom_point(mapping = aes(x = month, y =
sales, color = year)) + # scale_color_manual(values = c("2015" = "green", "2016" = "yellow",
"2017" = "orange", "2018" = "blue", "2019" = "brown"))

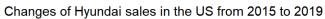
Including Plots

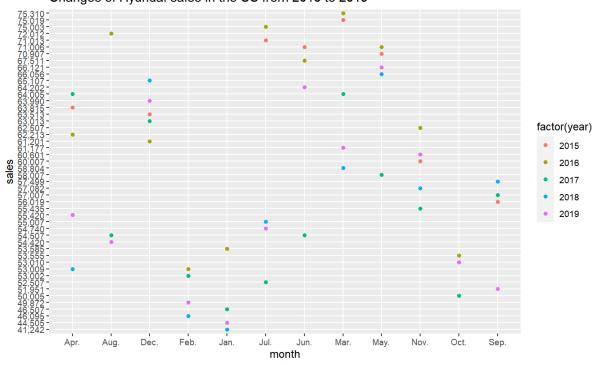
Changes of Budlight sales in the US from 2015 to 2019

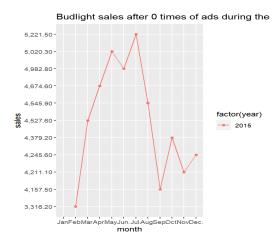


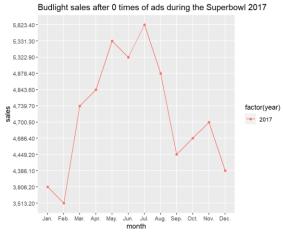


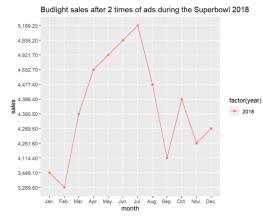


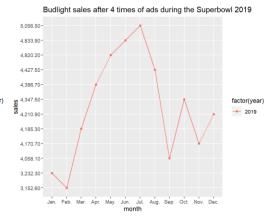




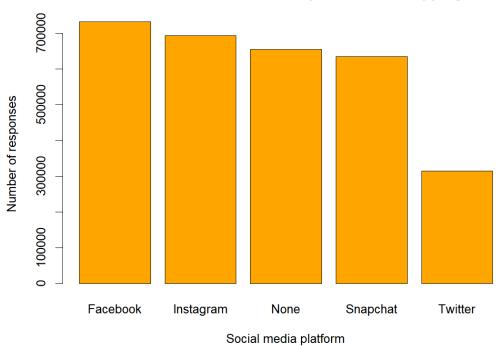




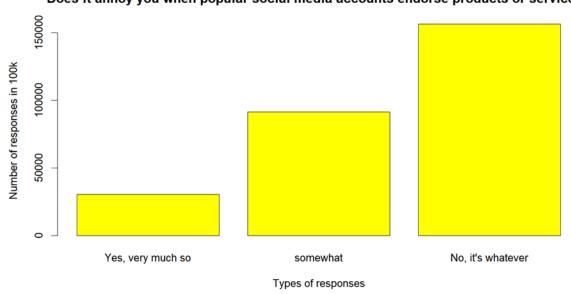




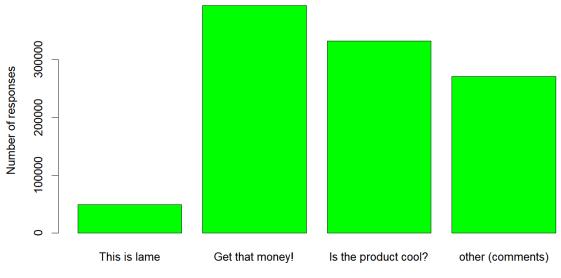
What social media has influenced your online shopping most?



Does it annoy you when popular social media accounts endorse products or services?

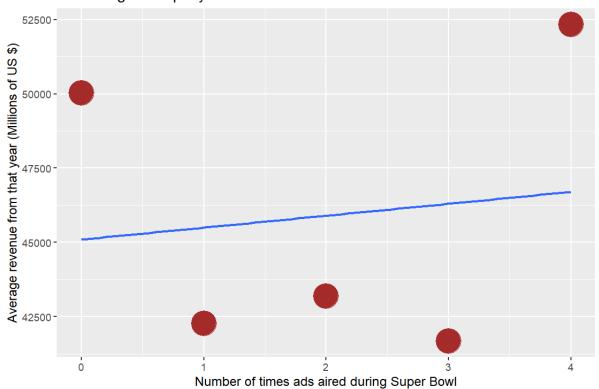


What do you think when an influencer is obviously selling a product?



Types of responses

Effects on how many ads aired during Super Bowl on Budlight company's annual revenue



Effects on how many ads aired during Super Bowl on Budlight company's first quaterly revenue

