



XYZ COMPANY SALES MODEL REPORT

HELLO!

This is a report that analyzes and interprets XYZ company's sales and product category data; drawing meaningful insights using the Multiplicative linear regression model in R-programming language and providing strategic directions to the company.



BENEFITS OF REPORT

Strategic Insights

Provided areas for improvement

ORIGINAL XYZ'S DATASET

```
## # tibble [34,867 x 15] (S3: tbl_df/tbl/data.frame)
## # $ Date : POSIXct[1:34867], format: "2016-02-19" "2016-02-20" ...
## # $ Year : num [1:34867] 2016 2016 2016 2016 2016 ...
## # $ Month : chr [1:34867] "February" "February" "February" "March" ...
## # $ Customer Age : num [1:34867] 29 29 29 29 29 29 29 29 29 ...
## # $ Customer Gender : chr [1:34867] "F" "F" "F" "F" ...
## # $ Country : chr [1:34867] "United States" "United States" "United States" "United States" ...
## # $ State : chr [1:34867] "Washington" "Washington" "Washington" "Washington" ...
## # $ Product Category: chr [1:34867] "Accessories" "Clothing" "Accessories" "Accessories" ...
## # $ Sub Category : chr [1:34867] "Tires and Tubes" "Gloves" "Tires and Tubes" "Tires and Tubes" ...
## # $ Quantity : num [1:34867] 1 2 3 2 3 1 2 1 2 2 ...
## # $ Unit Cost : num [1:34867] 80 24.5 3.67 87.5 35 66 52 60 8 2.5 ...
## # $ Unit Price : num [1:34867] 109 28.5 5 116.5 41.7 ...
## # $ Cost : num [1:34867] 80 49 11 175 105 66 104 60 16 5 ...
## # $ Revenue : num [1:34867] 109 57 15 233 125 78 120 68 20 6 ...
## # $ Column1 : num [1:34867] NA NA NA NA NA NA NA NA NA ...
```

OBSERVATION OF THE DATASET

N/A values present

No competitor's
values included

Exclusion of
environmental
factors; like
seasonality and
weekends

CLEANED DATA SET

```
data.frame': 34866 obs. of 12 variables:
$ Revenue          : num 109 57 15 233 125 78 120 68 20 6 ...
$ Easter            : num 0 0 0 0 0 1 1 1 0 0 ...
$ Summer            : num 0 0 0 0 0 0 0 0 0 0 ...
$ Zoo_Price         : num 10900 2850 500 11650 4167 ...
$ December          : num 0 0 0 0 0 0 0 0 0 0 ...
$ Country           : chr "United States" "United States" "United States" ...
$ Customer_Gender   : chr "F" "F" "F" "F" ...
$ Product_Category  : chr "Accessories" "Clothing" "Accessories" "Accessories" ...
$ Unit_Cost          : num 80 24.5 3.67 87.5 35 66 52 60 8 2.5 ...
$ Month              : chr "February" "February" "February" "March" ...
$ January             : num 0 0 0 0 0 0 0 0 0 0 ...
$ Unit_Price         : num 109 28.5 5 116.5 41.7 ...
- attr(*, "na.action")= 'omit' Named int 34867
.. - attr(*, "names")= chr "34867"
```

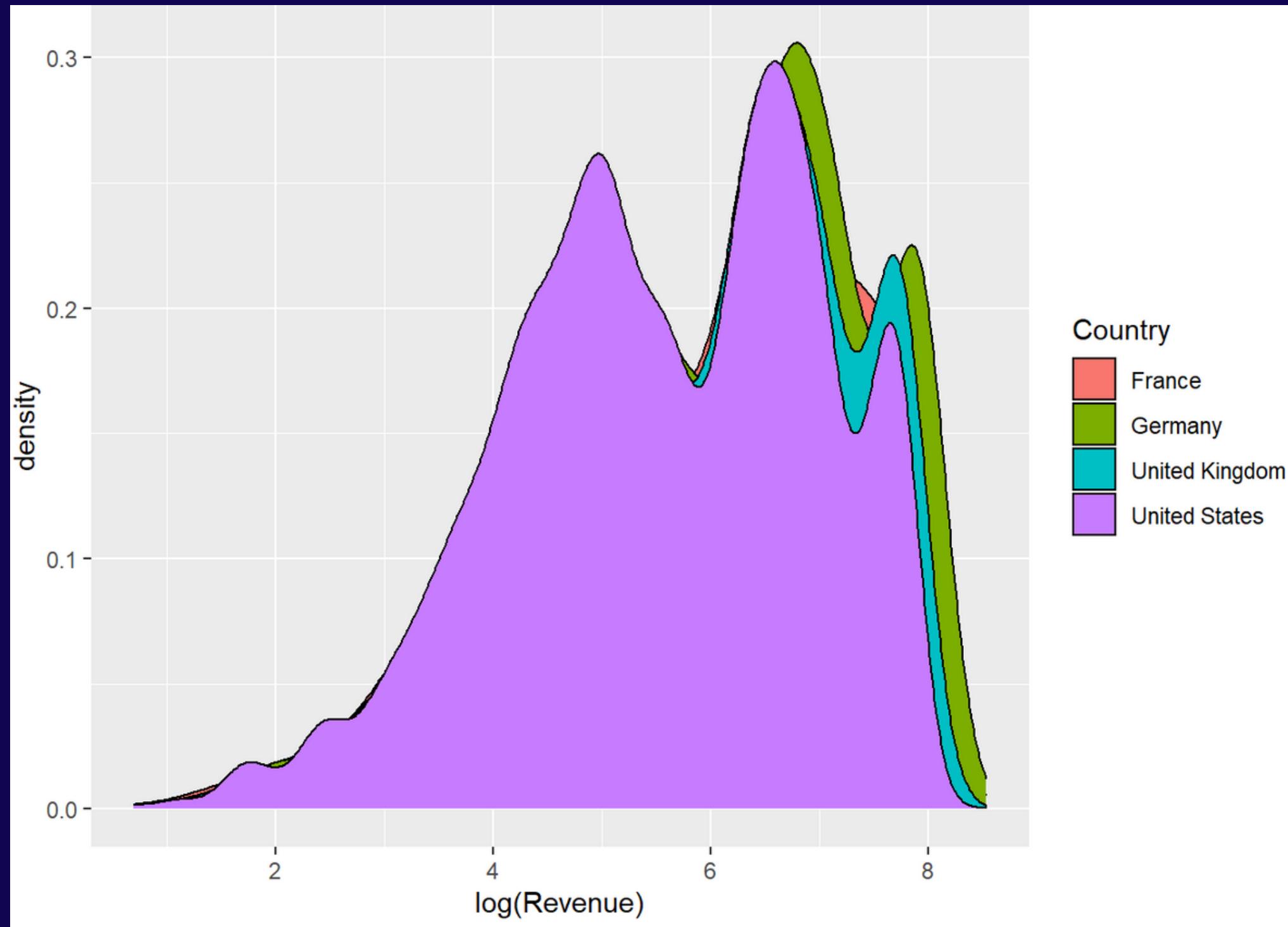


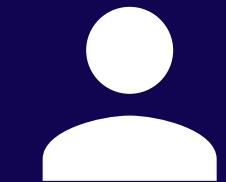
Observation of the cleaned dataset

No NAs anylonger

Dummy variables to account for environmental factors; inclusion a competitors price called ZOO price

Data Visualizations by Country for Insights

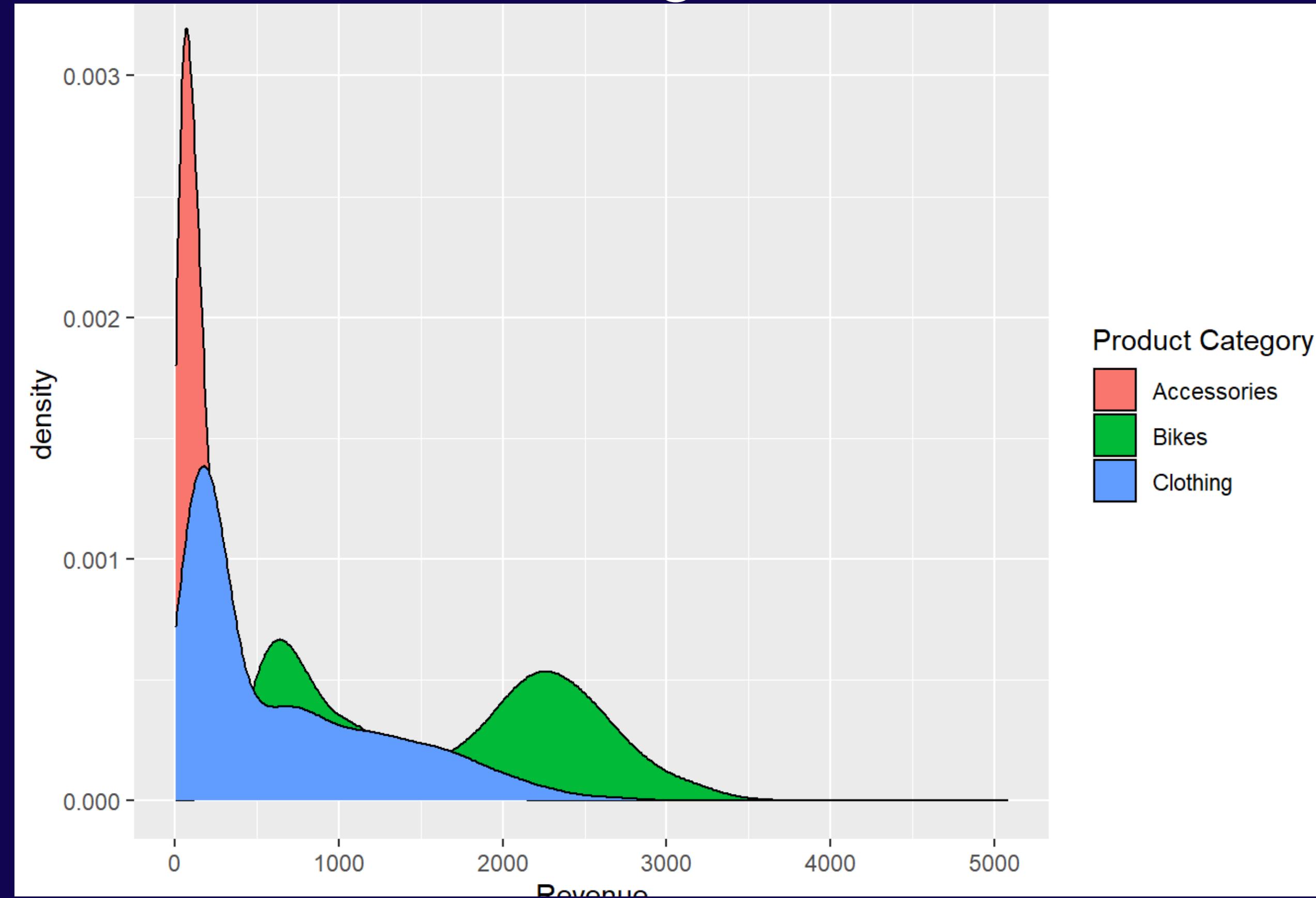




Observation of the plot

The US provided most revenue followed by Germany.

Data Visualizations by XYZ's Product Category for Insights

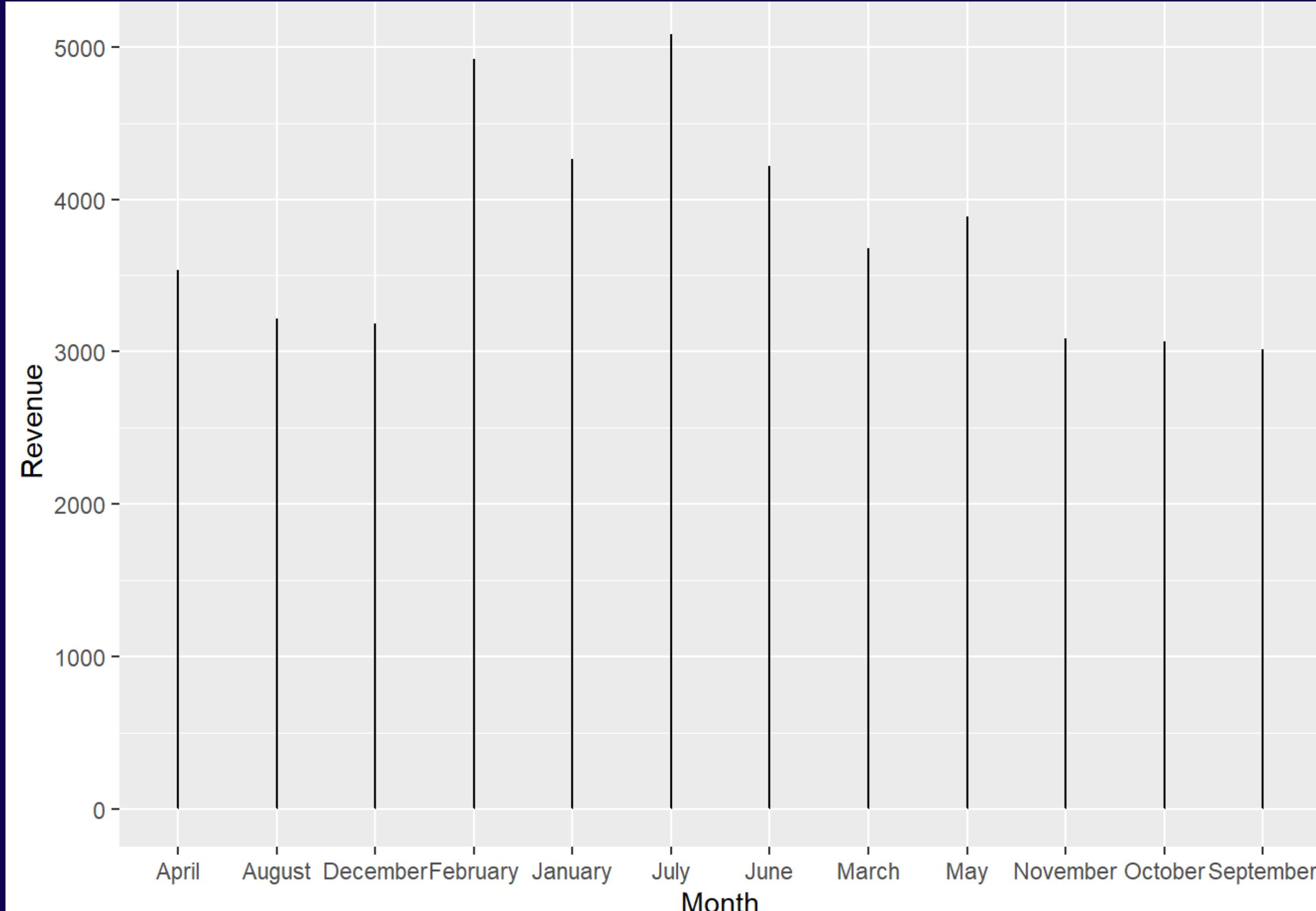




Observation of the plot

Accessories contributed to the most revenue
followed by Bikes

DATA VISUALIZATIONS BY MONTHS FOR INSIGHTS





OBSERVATION OF THE PLOT

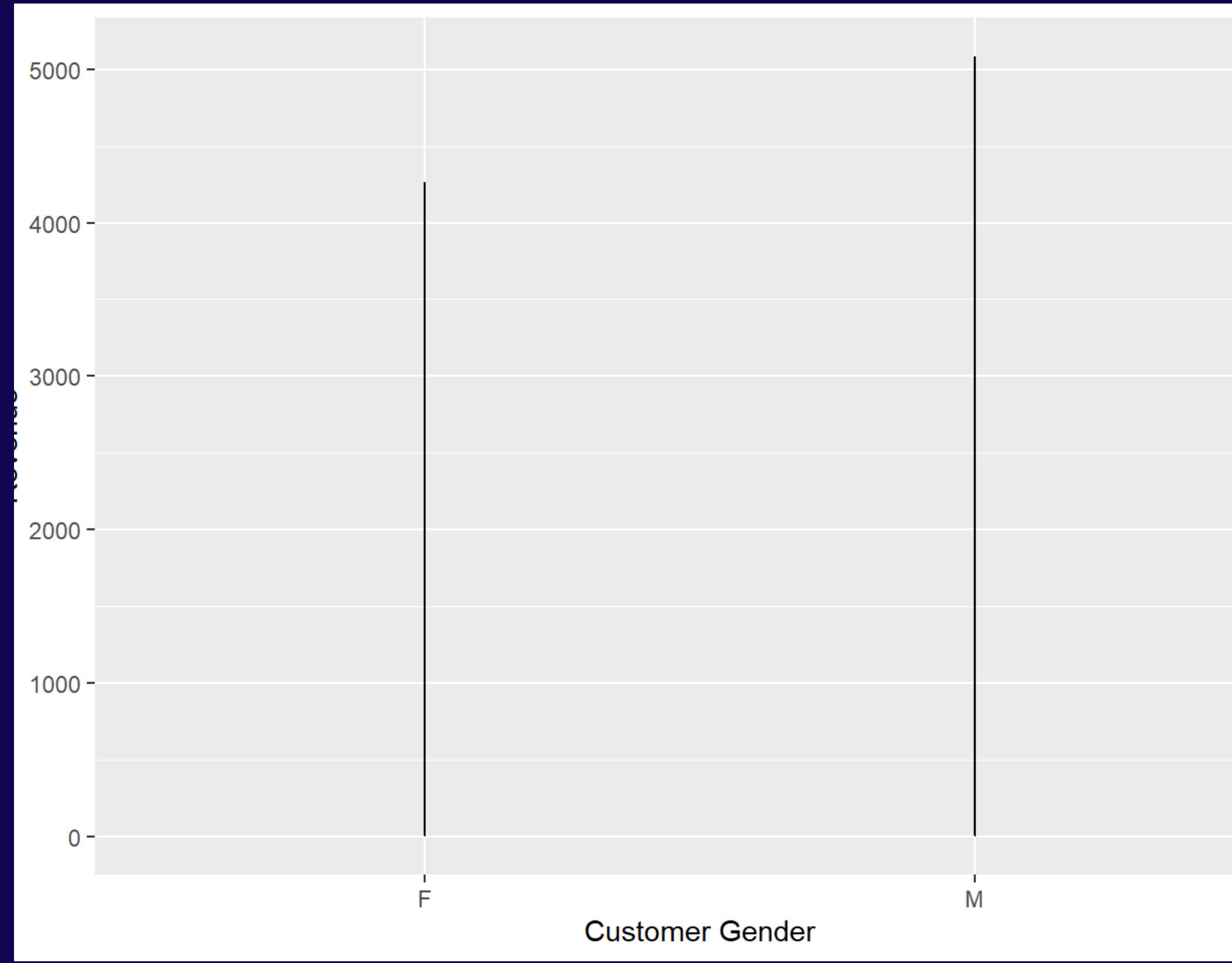
Month of July has
the highest
average revenue

Month of February
has the 2nd highest
average revenue

Month of September
has the lowest
average revenue

Data Visualizations by Gender for Insights

EE



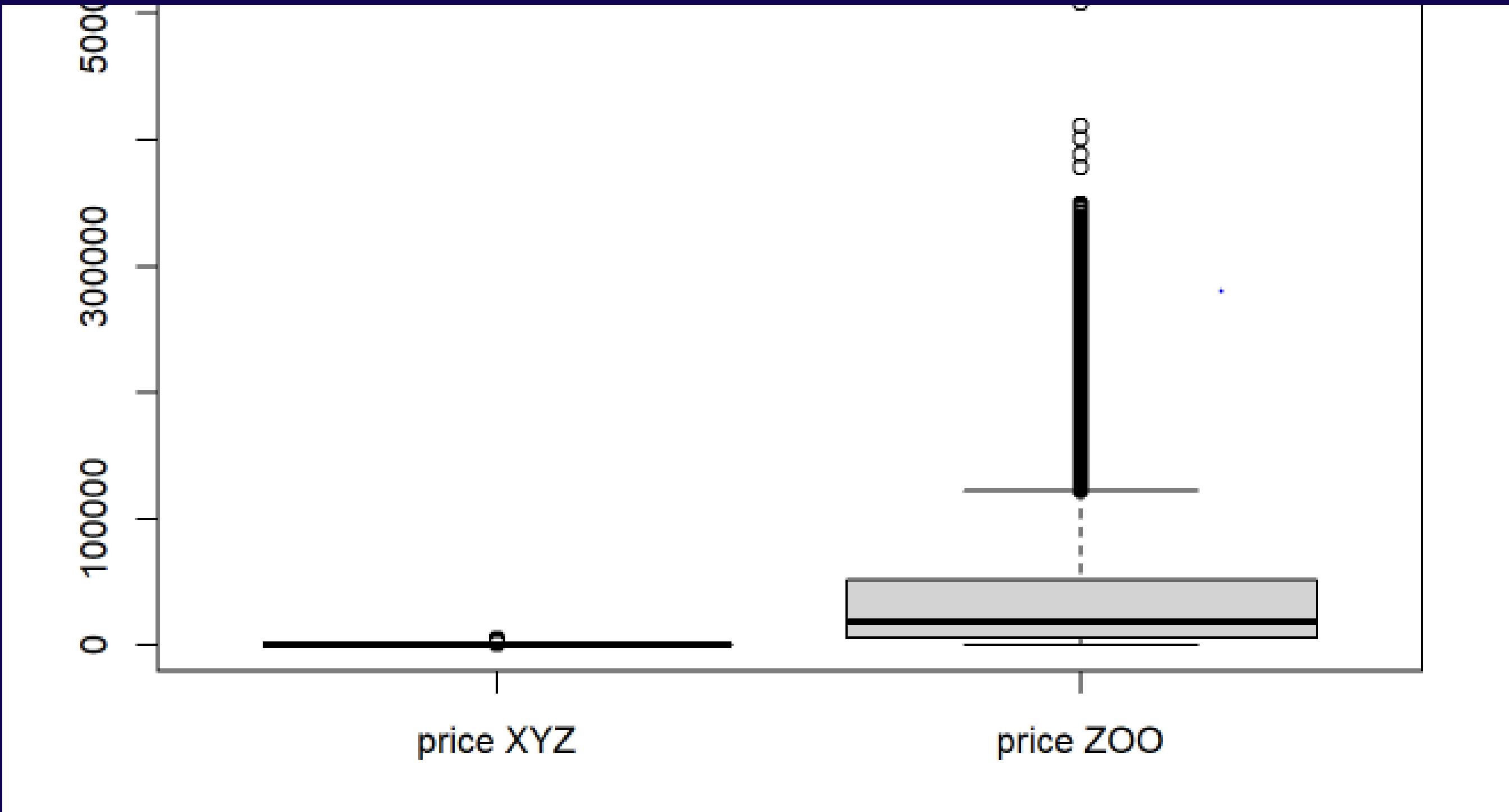
“



OBSERVATION OF THE PLOT

Male contributed more to the revenue of XYZ company compared to females

Data Visualizations by Competitor's Price for Insights



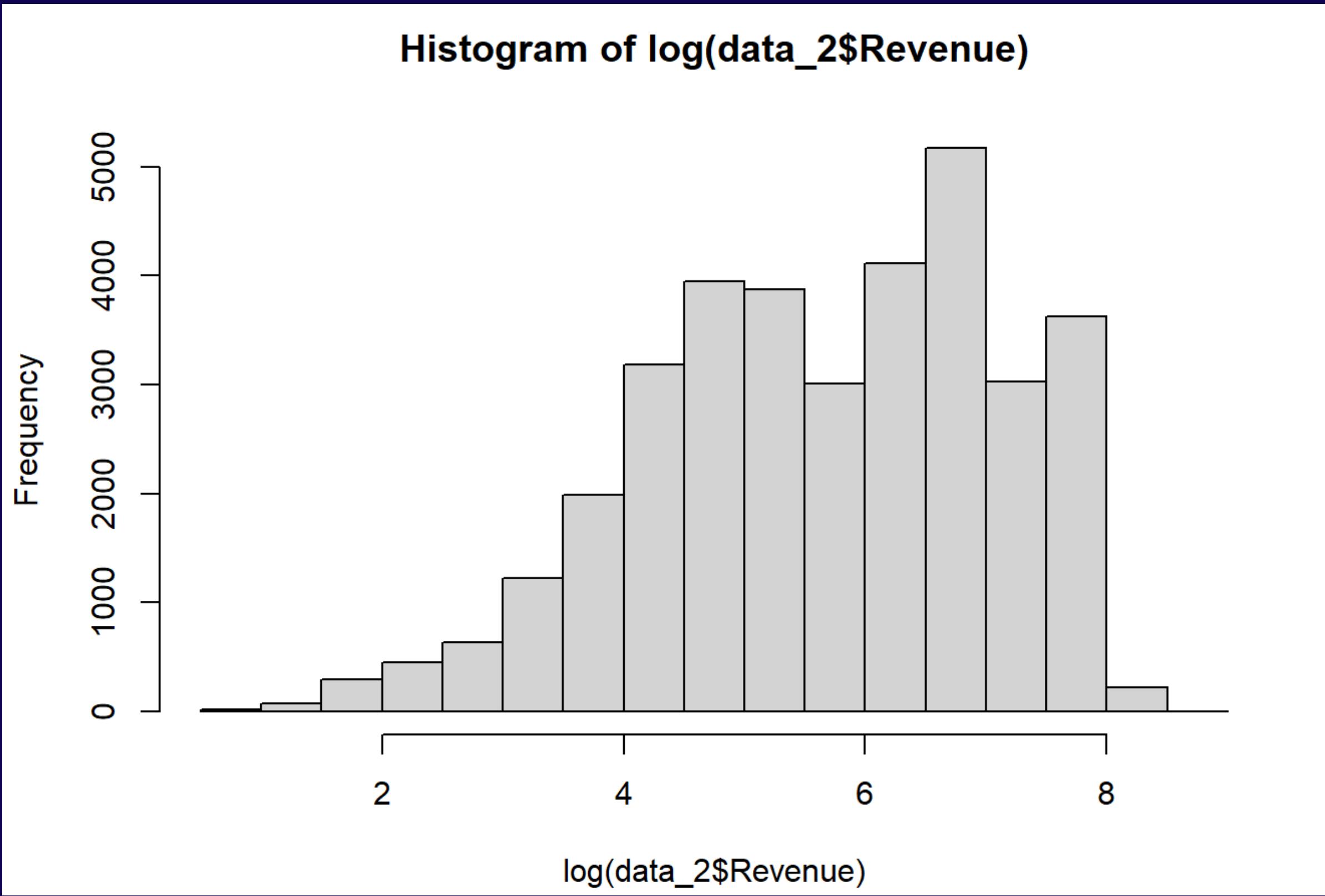


XYZ company's price was lower
comapred to the ZOO price.

Data Visualization of XYZ Company's Revenue

“

Histogram of $\log(\text{data_2\$Revenue})$



”

LOG-LOG LINEAR REGRESSION MODEL OF XYZ SALES MODEL

```
## lm(formula = log(Revenue) ~ Easter + Summer + log(Zoo_Price) +
##     December + `Customer Gender` + `Product Category` + log(`Unit Cost`) +
##     January + Country, data = data_2)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -1.0785 -0.4166  0.0762  0.3435  0.8663 
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)                 -3.59911   0.09685 -37.16 < 0.00000000000002 *** 
## Easter                      -0.00537   0.00794  -0.68     0.50    
## Summer                      -0.00533   0.00940  -0.57     0.57    
## log(Zoo_Price)                1.03341   0.02017  51.23 < 0.00000000000002 *** 
## December                     -0.00207   0.00785  -0.26     0.79    
## `Customer Gender`M          -0.00586   0.00453  -1.30     0.20    
## `Product Category`Bikes     0.31049   0.00799  38.85 < 0.00000000000002 *** 
## `Product Category`Clothing  0.11017   0.00667  16.51 < 0.00000000000002 *** 
## log(`Unit Cost`)            -0.16272   0.02023  -8.04  0.00000000000091 *** 
## January                      -0.01103   0.00837  -1.32     0.19    
## CountryGermany              -0.01440   0.00919  -1.57     0.12    
## CountryUnited Kingdom        -0.00916   0.00790  -1.16     0.25    
## CountryUnited States         -0.00542   0.00667  -0.81     0.42    
## ---
## Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.422 on 34853 degrees of freedom
## Multiple R-squared:  0.915,  Adjusted R-squared:  0.915
```



XYZ has a lower price

CONCLUSIONS AND FINDINGS

The revenue of XYZ responds

strongly to own unit cost changes



XYZ company revenue benefited greatly from Zoo's price



THANK YOU!