Executive Summary: Digital Marketing Performance

Business Problem

The company needed to determine whether weekly advertising spend significantly affects weekly conversions, and to find the most accurate model to forecast conversions. In simpler terms, management wanted to know:

> *"Does spending more on advertising actually increase conversions?"*

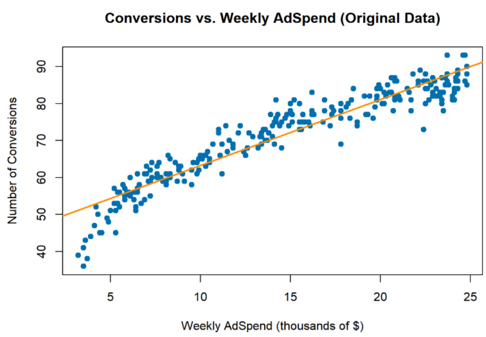
The claims that need to be evaluated also include if 1% increase in ad spend would increase ad conversions by at least 5. The goal was to test this relationship statistically and recommend a predictive model for future planning.

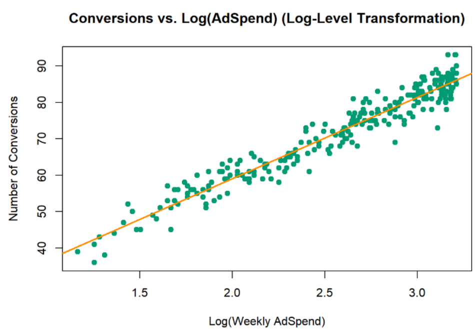
Key Findings

The relationship between advertising spend and conversions is strong and positive (slope coefficient was statistically significant), demonstrating higher ad spend leads to higher number of conversions. However, there was a critical finding that while advertising is effective, its marginal impact is not linear.

Two linear models were tested:

- Linear model (Conversions ~ AdSpend), representing absolute change in ad spending





- Linear-log model (Conversions ~ log(AdSpend)), representing a percentage change in ad spending

The linear-log model performed better, accurately demonstrating an effective linear relationship, explaining roughly 93.7% of the variation in conversions (compared to 89% for the linear model) and producing smaller forecasting errors.

The central finding from this model is that a 1% increase in advertising spend currently generates approximately 0.22 additional conversions. This performance level is substantially below the stated financial target.

Recommendations

We recommend that the marketing budget not be approved with the current performance expectation. The analysis provides conclusive evidence that the existing marketing strategy is incapable of delivering the return on investment (ROI) required to meet the benchmark of five conversions per 1% spend increase. We advise a two-pronged approach: either re-evaluate and adjust the performance targets to align with a data-driven reality or invest in fundamentally new marketing strategies designed to improve efficiency and bridge the significant performance gap.

The linear-log model should be adopted for weekly forecasting and budgeting decisions to guide more efficient resource allocation. By using this model, the marketing team can identify the optimal spending range, thereby maximizing campaign efficiency. Performance should be monitored continuously, with the model updated as new campaign data becomes available to ensure accuracy over time.

Supporting Evidence

- **Model Insights:** Both linear and linear-log models confirm a strong and reliable relationship between ad spending and conversions. However, the curved (linear-log) model provides a more accurate fit, capturing how returns gradually level off as spending rises.
- **Predictable Diminishing Returns:** The positive impact of ad spending weakens at higher levels. This indicates a natural marginal return on increased ad spending.
- **High Confidence in Findings:** Our conclusions are based on a highly precise model. We are 95% confident that the true return for a 1% spend increase is between 0.217 and 0.231 additional conversions, which is far below the target of five.
- **Right Interval for Specific Business Problems**: We noted it is important to use the right interval (across prediction interval and confidence interval) to answer specific business questions. The prediction interval showed a broader range for short-term model prediction, while confidence interval is useful for long-term planning.

Limitations

- This simple regression utilizes only *one* predictor variable (AdSpend); external factors (such as seasonality, competitor pricing, and website user experience changes) that influence the remaining conversion variability are not accounted for, suggesting the need for a more comprehensive, multivariate analytical approach in subsequent studies.
- This analysis is based on historical data and assumes a linear-log relationship between ad spending and conversions for future. The results may not hold if there are structural changes or if ad spend moves outside the range observed in the data.

Appendix

Data Loading & Setup

```
# Load the data file
data <- read.csv("Case2_Marketing.csv")

# Verify the data samples
print(head(data))
```

```
##   Week AdSpend Conversions
## 1    1     4.8          49
## 2    2     6.0          56
## 3    3    24.1          81
## 4    4    14.1          75
## 5    5     9.6          64
## 6    6    20.8          86
```

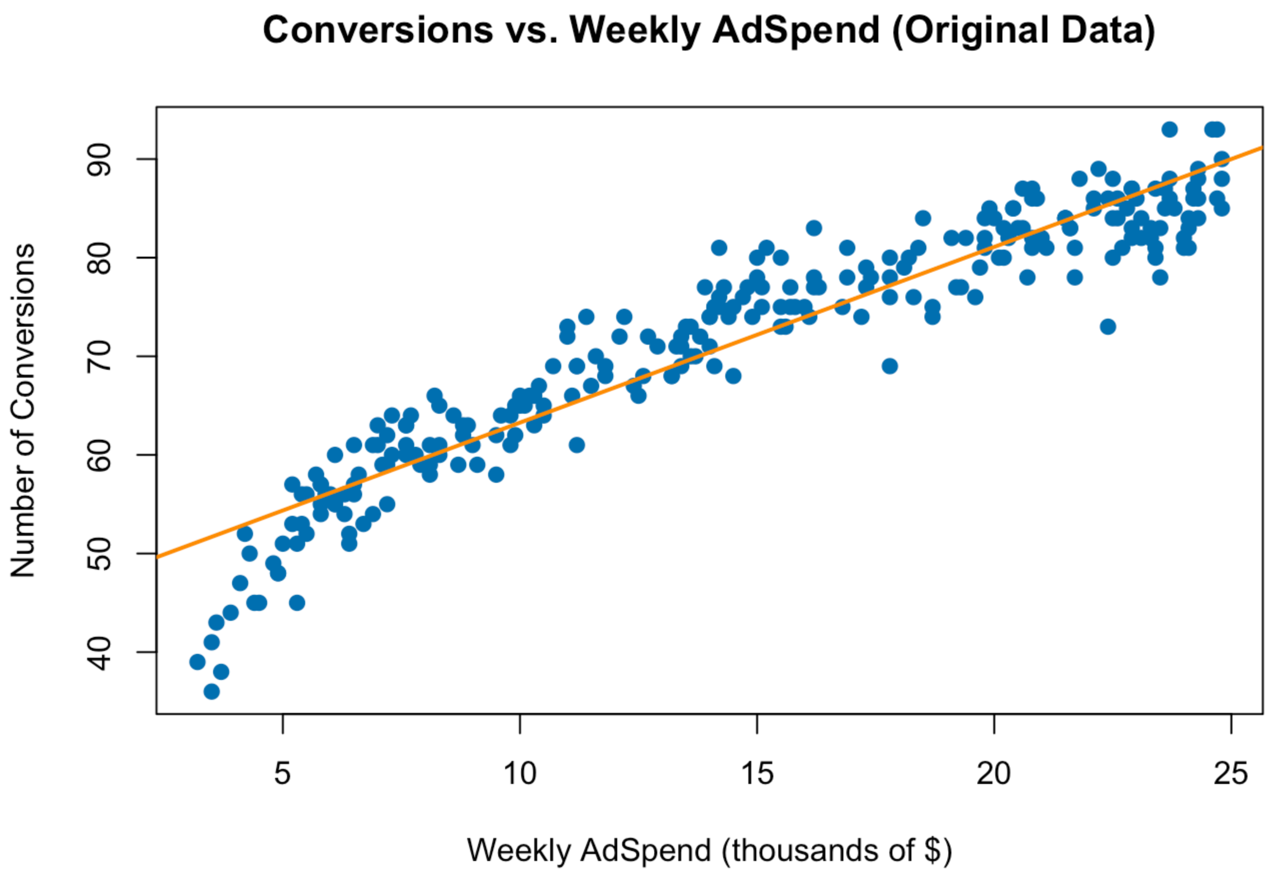
```
print(paste("Sample Size (n):", nrow(data)))
```

```
## [1] "Sample Size (n): 260"
```

Task 1: Scatterplot and Linearity Check

```
plot(data$AdSpend, data$Conversions,
     main = "Conversions vs. Weekly AdSpend (Original Data)",
     xlab = "Weekly AdSpend (thousands of $)",
     ylab = "Number of Conversions",
     pch = 19,
     col = "#0072B2") # Blue color for data points

abline(lm(Conversions ~ AdSpend, data = data), col = "darkorange", lwd = 2)
```



Observation: The scatterplot shows a positive, but clearly non-linear, concave-down relationship (diminishing marginal returns). Even though the relationship is positive, the relationship between the two variables is not linear as observed from the scatterplot graph plotted with the superimposed regression line.

Task 2: Linear Regression Model (No transformation)

```
model1 <- lm(Conversions ~ AdSpend, data = data)

summary(model1)
```

```
##
## Call:
## lm(formula = Conversions ~ AdSpend, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.6941  -2.0609   0.3282   2.7302  10.2378
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.45691    0.61685   73.69   <2e-16 ***
## AdSpend      1.78206    0.03864   46.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.082 on 258 degrees of freedom
## Multiple R-squared:  0.8918, Adjusted R-squared:  0.8914
## F-statistic:  2127 on 1 and 258 DF,  p-value: < 2.2e-16
```

Interpretation: The estimated regression model equation is: `Conversions_hat = 45.46 + 1.78 * AdSpend`.

- **Intercept (β_0 = 45.46):** This is the predicted number of weekly conversions when the advertising spend is $0. The model suggests that if the client spent nothing on ads, they would still expect to get approximately 45 conversions per week. This may or may not be correct due to extrapolation error for extreme values.
- **AdSpend Coefficient (β_1 = 1.78):** This is the slope of the regression line. It indicates that for each additional $1,000 increase in weekly ad spend, the number of website conversions is predicted to increase by approximately 1.78.

Task 3: Assess Model Fit (No transformation)

```
r_squared_model1 <- summary(model1)$r.squared
rmse_model1 <- summary(model1)$sigma

print(paste("Linear-Linear Model R-squared:", round(r_squared_model1, 3)))
```

```
## [1] "Linear-Linear Model R-squared: 0.892"
```

```
print(paste("Linear-Linear Model Standard Error:",
            round(rmse_model1, 2), "Conversions"))
```

```
## [1] "Linear-Linear Model Standard Error: 4.08 Conversions"
```

Interpretation: The R-squared of 0.892 indicates that 89.2% of the variability in conversions is explained by ad spend. The standard error of 4.08 suggests that, on average, the model's predictions are off by about four conversions. The t-test for the slope is highly significant ($t = 46.12,\ p < 0.001$), confirming that ad spend is a statistically significant predictor of conversions.

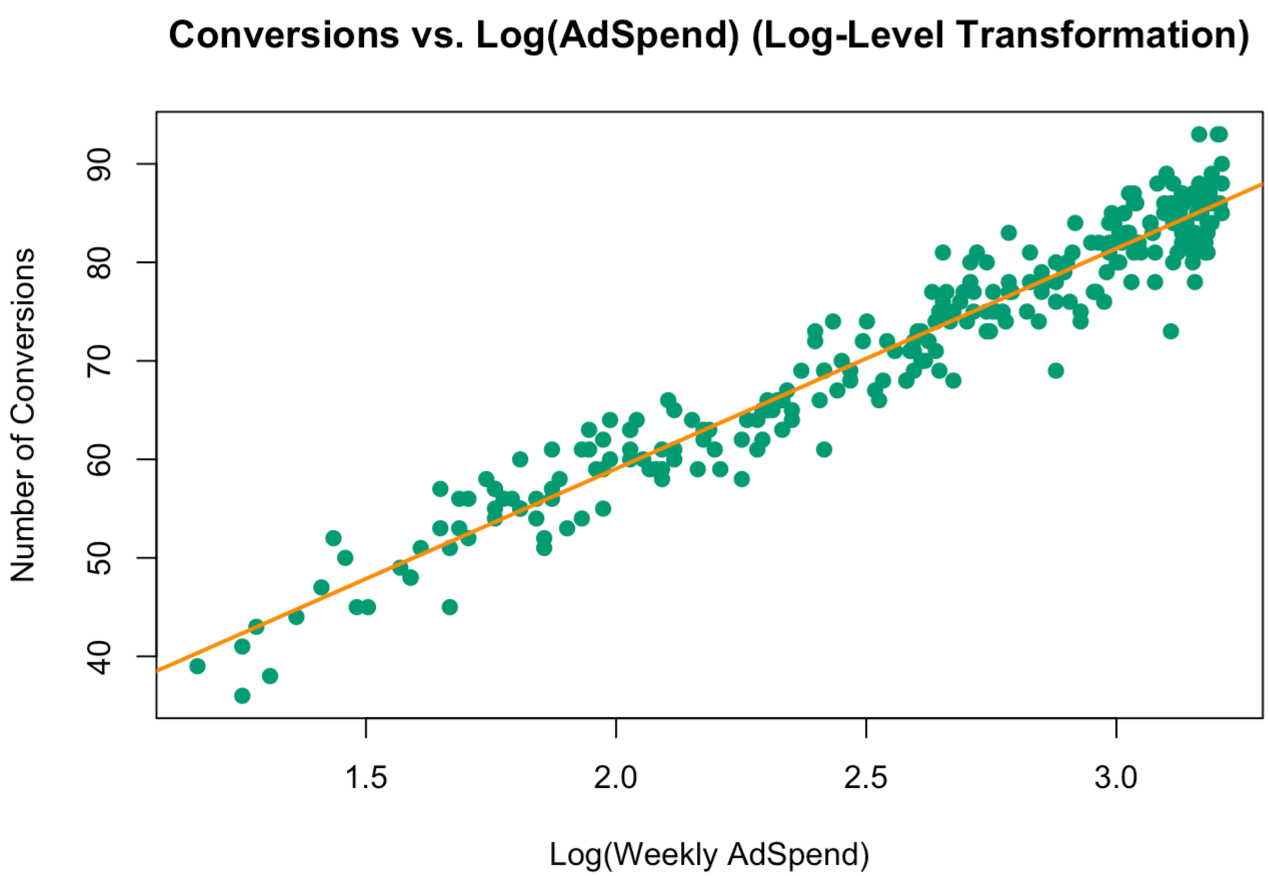
Task 4: Tukey's Bulging Rule and Variable Transformation

The concave-down shape (diminishing returns) and the need to test a % change in X (AdSpend) against an absolute change in Y (Conversions) suggest the Linear-Log model.

```
data$Log_AdSpend <- log(data$AdSpend)

# Plot the transformed data to check for linearity improvement
plot(data$Log_AdSpend, data$Conversions,
     main = "Conversions vs. Log(AdSpend) (Log-Level Transformation)",
     xlab = "Log(Weekly AdSpend)",
     ylab = "Number of Conversions",
     pch = 19,
     col = "#009E73") # Green color for data points

# Add the new linear trend line
ll_fit <- lm(Conversions ~ Log_AdSpend, data = data)
abline(ll_fit, col = "darkorange", lwd = 2)
```



Observation: The scatterplot shows a strong, positive linear relationship between AdSpend and Conversions, after Log(AdSpend) transformation.

Task 5: Linear Regression Model (With transformation)

```
model2 <- lm(Conversions ~ Log_AdSpend, data = data)

summary(model2)
```

```
##
## Call:
## lm(formula = Conversions ~ Log_AdSpend, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.8556  -1.9563   0.0238   1.8802   7.8829
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.3335     0.9405   15.24   <2e-16 ***
## Log_AdSpend  22.3611     0.3607   61.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.113 on 258 degrees of freedom
## Multiple R-squared:  0.9371, Adjusted R-squared:  0.9368
## F-statistic:  3843 on 1 and 258 DF,  p-value: < 2.2e-16
```

Interpretation: The estimated regression model equation is:
`Conversions_hat = 14.33 + 22.36 * log(AdSpend)`.

- **Intercept (β_0 = 14.33):** The model predicts approximately 14 conversions per week when the advertising spend is $1 (since log(1) = 0).
- **Log(AdSpend) Coefficient (β_1 = 22.36):** This is the slope of the log-transformed regression line. It indicates that for every 1% increase in advertising spend, the number of website conversions is predicted to increase by approximately 0.22 (22.36 / 100).

Task 6: Assess Model Fit (With transformation)

```
# Assess Model Fit
rmse_model2 <- summary(model2)$sigma
r_squared_model2 <- summary(model2)$r.squared

print(paste("Linear-Log Model R-squared:", round(r_squared_model2, 3)))
```

```
## [1] "Linear-Log Model R-squared: 0.937"
```

```
print(paste("Linear-Log Model Standard Error:", round(rmse_model2, 2), "Conversions"))
```

```
## [1] "Linear-Log Model Standard Error: 3.11 Conversions"
```

Interpretation: The model demonstrates an excellent fit to the data. The R-squared of 0.937 indicates that 93.7% of the variability in conversions is explained by ad spend, signifying strong explanatory power. The standard error of 3.11 suggests that, on average, the model's predictions are off by about three conversions. The t-test for the slope is highly significant ($t = 61.99$, $p < 0.001$), confirming that ad spend is a statistically significant predictor of conversions.

Task 7: Compare the Two Models

```
print(paste("Linear Model -- R-squared:",
            round(r_squared_model1, 3), "| RMSE:", round(rmse_model1, 2)))
```

```
## [1] "Linear Model -- R-squared: 0.892 | RMSE: 4.08"
```

```
print(paste("Linear-Log Model -- R-squared:",
            round(r_squared_model2, 3), "| RMSE:", round(rmse_model2, 2)))
```

```
## [1] "Linear-Log Model -- R-squared: 0.937 | RMSE: 3.11"
```

Interpretation: The transformed linear-log model is the more suitable choice for this analysis. Model 2 is suitable due to higher R-squared (0.937 vs. 0.892), lower RMSE (3.11 vs. 4.08), and appropriate functional form for interpreting percentage changes in AdSpend as desired by the CFO.

Task 8: Confidence Interval for the Slope (Transformed Model)

```
conf_interval_slope <- confint(model2, level = 0.95)
print(conf_interval_slope)
```

```
##                2.5 %   97.5 %
## (Intercept) 12.48154 16.18554
## Log_AdSpend 21.65084 23.07142
```

Interpretation: We are 95% confident that the true population value of the slope coefficient (β_1) lies between 21.65 and 23.07. Since the slope of the transformed model represents the change in conversions for a 1% increase in spending ($\beta_1/100$), this means we are 95% confident that the true return for a 1% increase in AdSpend falls between 0.2165 and 0.2307 additional conversions.

Task 9: Test the CFO’s Claim

```
# Extract slope estimate (b1) and its standard error (SE_b1)
s <- summary(model2)$coefficients
b1 <- s["Log_AdSpend", "Estimate"]
SE_b1 <- s["Log_AdSpend", "Std. Error"]
df <- model2$df.residual

b1_threshold <- 500  # because 0.01*b1 >= 5  <=>  b1 >= 500

# One-sided t-test statistic for H0: b1 >= 500 vs H1: b1 < 500
t_stat <- (b1 - b1_threshold) / SE_b1
print(paste("t-test statistic:", round(t_stat, 2)))
```

```
## [1] "t-test statistic: -1324.2"
```

```
# One-sided p-value (left tail, since H1: b1 < 500)
p_val <- pt(t_stat, df = df, lower.tail = TRUE)
print(paste("p-value:", round(p_val, 2)))
```

```
## [1] "p-value: 0"
```

Interpretation: To evaluate the claim, a one-sided hypothesis test was conducted.

- **Null Hypothesis (H_0):** The campaign meets or exceeds the target (a 1% increase in ad spend yields ≥ 5 conversions).
- **Alternative Hypothesis (H_1):** The campaign does not meet the target.

The test yielded a highly negative t-statistic of -1324.2 and a p-value of essentially zero. Since the p-value is far below significance level ($\alpha = 0.05$), we strongly reject the null hypothesis. This provides strong statistical evidence that the campaign’s performance is significantly below the claimed target.

Task 10: Intervals for $15000 AdSpend

```
new_spend <- 15
new_data <- data.frame(AdSpend = new_spend)
new_data$Log_AdSpend <- log(new_spend)

# Predict the Conversion: Point estimate
point_estimate <- predict(model2, new_data)
print(paste("Point Forecast (AdSpend = $15k):",
            round(point_estimate, 2), "Conversions"))
```

```
## [1] "Point Forecast (AdSpend = $15k): 74.89 Conversions"
```

```
prediction_interval <- predict(model2,
                               new_data, interval = "prediction", level = 0.95)
print("95% Prediction Interval for next week's conversions:")
```

```
## [1] "95% Prediction Interval for next week's conversions:"
```

```
print(round(prediction_interval, 2))
```

```
##     fit   lwr   upr
## 1 74.89 68.75 81.03
```

```
confidence_interval <- predict(model2,
                               new_data, interval = "confidence", level = 0.95)
print("95% Confidence Interval for mean conversions:")
```

```
## [1] "95% Confidence Interval for mean conversions:"
```

```
print(round(confidence_interval, 2))
```

```
##     fit   lwr   upr
## 1 74.89 74.49 75.28
```

Observation: For a planned Ad Spend of $15,000 next week, the model provides the following forecasts:

- **95% Prediction Interval:** We are 95% confident that the conversions for a single, specific week with a $15k spend will fall between 68.75 and 81.03.
- **95% Confidence Interval:** We are 95% confident that the average number of conversions for a $15k spend falls between 74.49 and 75.28.

Task 11: Compare and Explain Both Intervals

```
pi_width <- prediction_interval[1, "upr"] - prediction_interval[1, "lwr"]
print(paste("Prediction Interval Width:", round(pi_width, 2)))
```

```
## [1] "Prediction Interval Width: 12.29"
```

```
ci_width <- confidence_interval[1, "upr"] - confidence_interval[1, "lwr"]
print(paste("Confidence Interval Width:", round(ci_width, 2)))
```

```
## [1] "Confidence Interval Width: 0.79"
```

Interpretation: The prediction interval is significantly wider because it must account for two sources of uncertainty: the uncertainty in the model's estimated relationship plus the random variability of an individual week's performance. The confidence interval only accounts for the former.

The specific business scenarios where each would be relevant is as follows:

- Use the prediction interval for tactical, short-term decisions, like ensuring enough inventory or staff is available for next week's expected outcomes.
- Use the confidence interval for strategic planning, like setting a long-term quarterly performance benchmark (e.g., "our average weekly conversions should be greater than 70").

Task 12: Recommendation

Based on this analysis, the client's CFO **should not approve** the marketing budget with the expectation of meeting the current performance goal. The primary evidence is the direct refutation of the CFO's claim. The analysis in Task 9 conclusively rejects the hypothesis that a 1% increase in ad spend yields at least 5 additional conversions (p-value ~ 0).

The more accurate transformed model (Task 5) predicts that a 1% increase in spending generates only about 0.22 additional conversions. This is significantly below the CFO's required threshold. While ad spending is a statistically significant driver of conversions and positively correlated, it does not deliver the return on investment required to justify the budget under the specified criteria.

Disclaimer: I have consulted the Gemini 2.5 Pro model for helping me with some of the R code and syntax.