

BUSINESS APPLICATIONS OF DATA SCIENCE

PROJECT 2

ARUNIMA MANDWARIYA | BHART LAL | NEHA S K | SIDRA SHAIKH

QUESTION 1

Which channel performs better in generating sales? Q

Social Media

Social Media is the channel that performs better than others in generating sales.

ANOVA

METHODOLOGY

Through Analysis of Variance, we concluded that there exists statistically significant difference in the means of Email, SMS and Social Media.

Subsequently, upon comparing mean sales and total sales for all channels, we find **Social Media** to be the most effective in generating sales.

ANOVA Table:

	sum_sq	df	F	PR(>F)
C(channel)	1.387097e+08	2.0	449.843944	1.033532e-153
Residual	2.308005e+08	1497.0	NaN	NaN

Average sales per channel:

channel

Social Media 2510.079635

Email 1982.010739

SMS 1783.564473

Name: sales, dtype: float64

QUESTION2

Do male and female customers respond differently in terms of engagement?



NO

Gender of the customers does not have a statistically significant effect on the engagement.

Hypothesis Testing

Shapiro-Wilk test

Levene's Test

t-Test

METHODOLOGY

Shapiro Wilks test proves that the data is normally distributed for both the mentioned genders, Male and Female.

Moreover, Levene's test proves that the variances across the groups of Male and Female is similar enough to run a pooled variance t-test.

```
Shapiro Male: W=0.9971, p=0.1802  
Shapiro Female: W=0.9988, p=0.9061  
Levene's test: stat=0.0202, p=0.8870  
  
Independent t-test (equal variances): stat=0.4509, p=0.6521
```

Finally, the t-test proves that there is **no statistically significant difference** in engagement between male and female customers.

QUESTION 3

Can we predict engagement score using advertising budget and click-through rate?



YES

Advertising budget has the highest impact on engagement followed by Click through rate (CTR). However, there's a significant gap between the impact of Advertising budget and CTR.

Multiple Linear Regression

Levene's Test

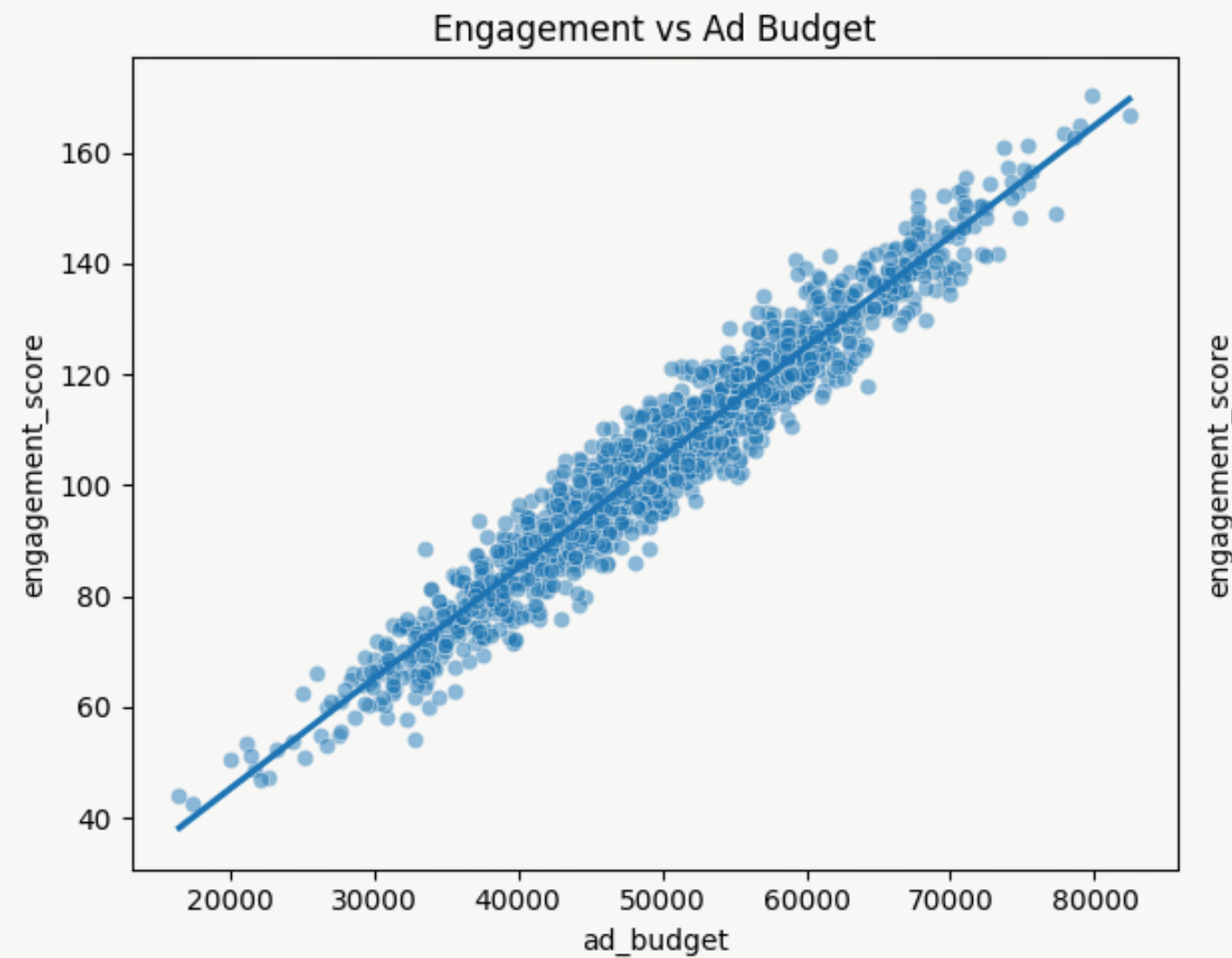
Shapiro-Wilk Test

ANOVA

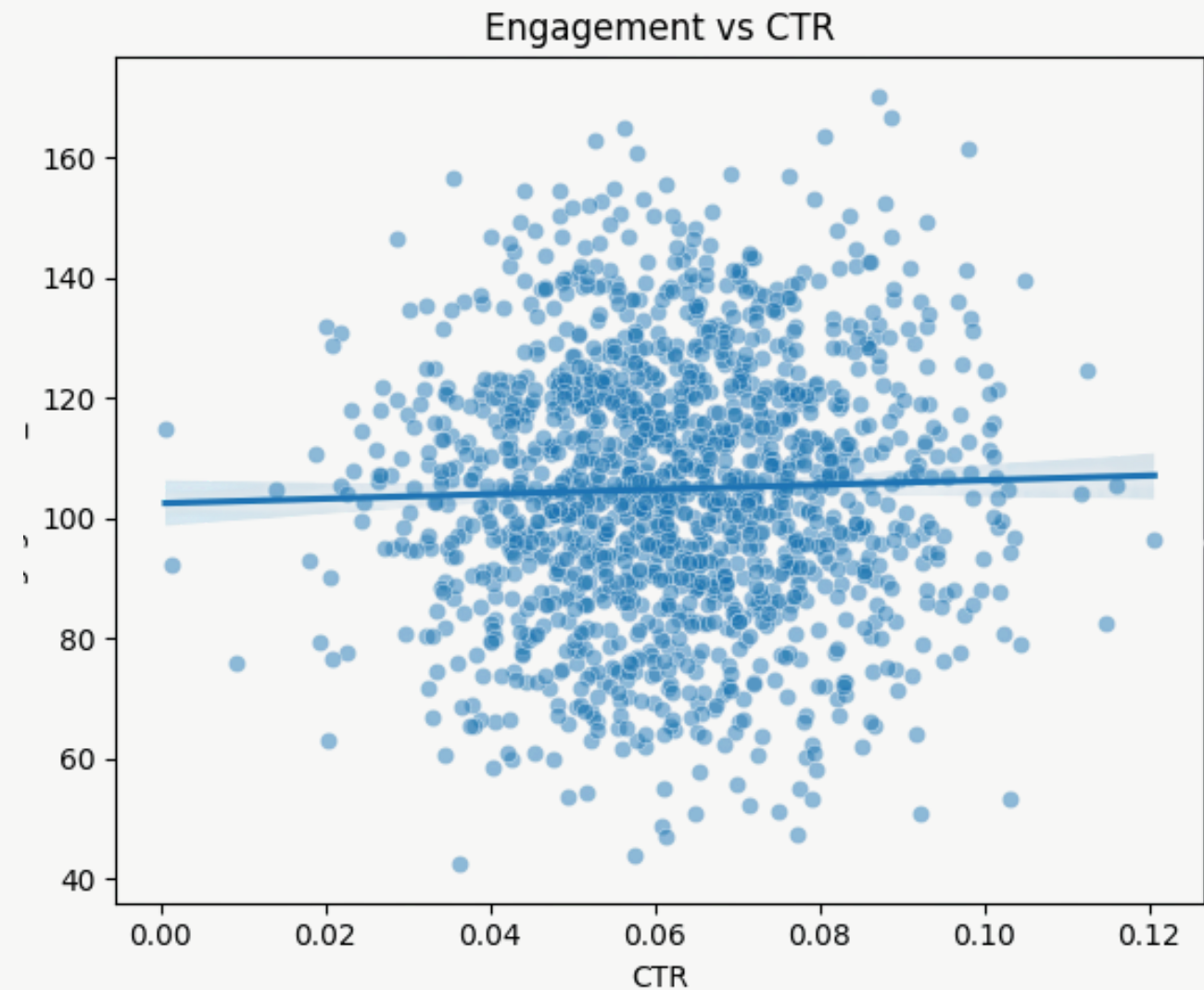
Breusch-Pagan Test

METHODOLOGY

1. Linearity check (scatter plots)



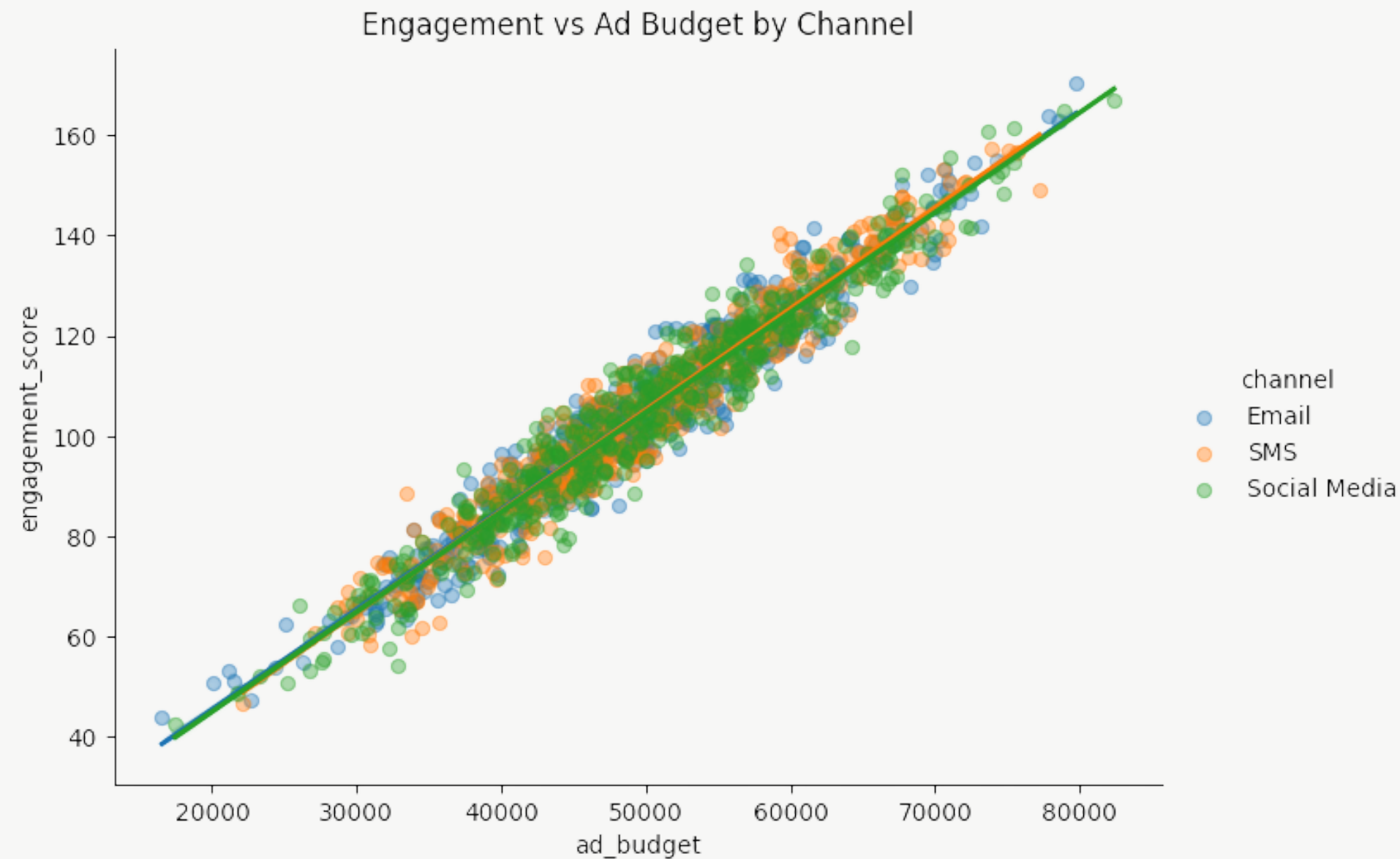
Higher ad budgets strongly predict higher engagement i.e there is a strong positive linear relationship



CTR may not be a strong predictor because there is no linear relationship

METHODOLOGY

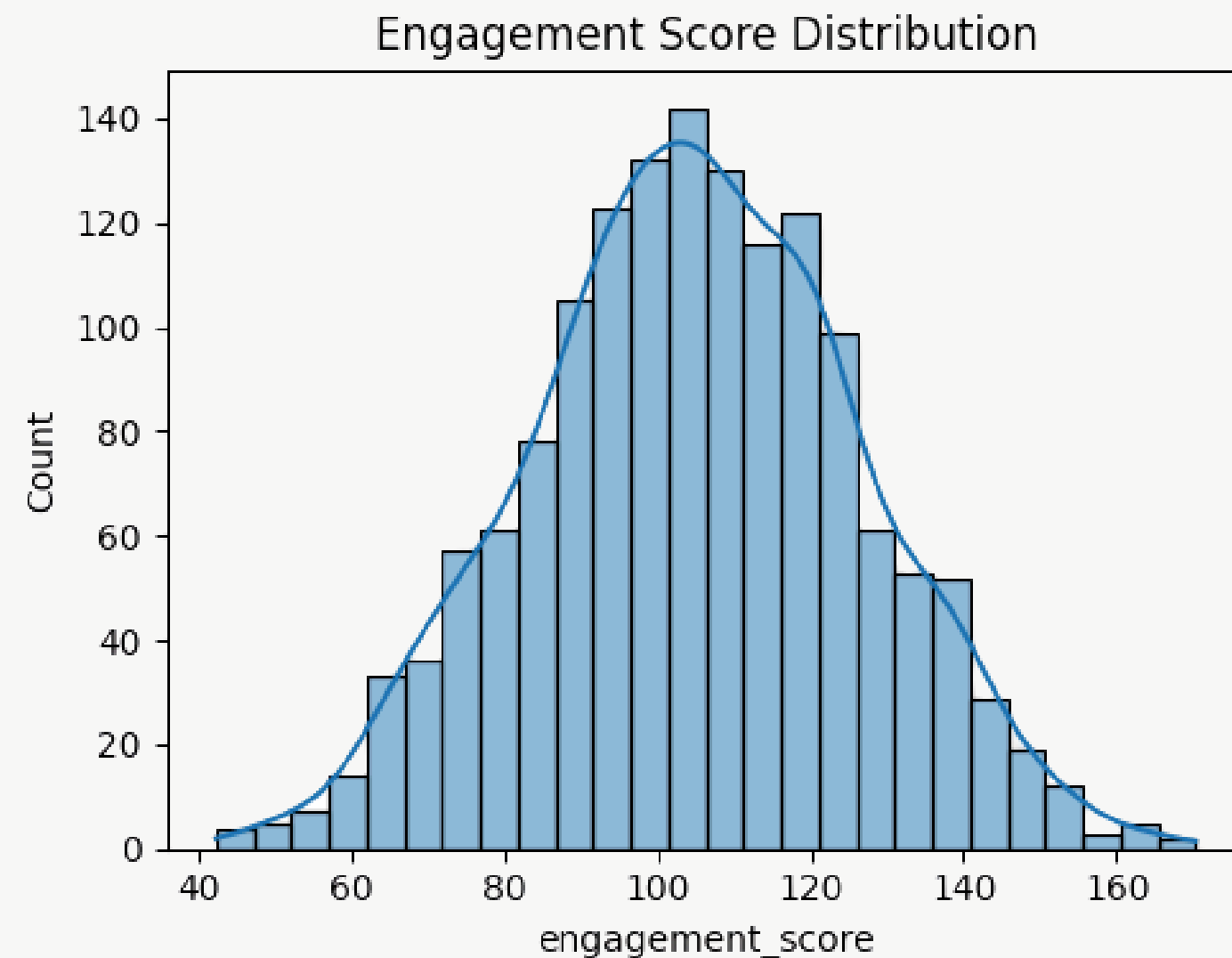
1. Linearity check (scatter plots)



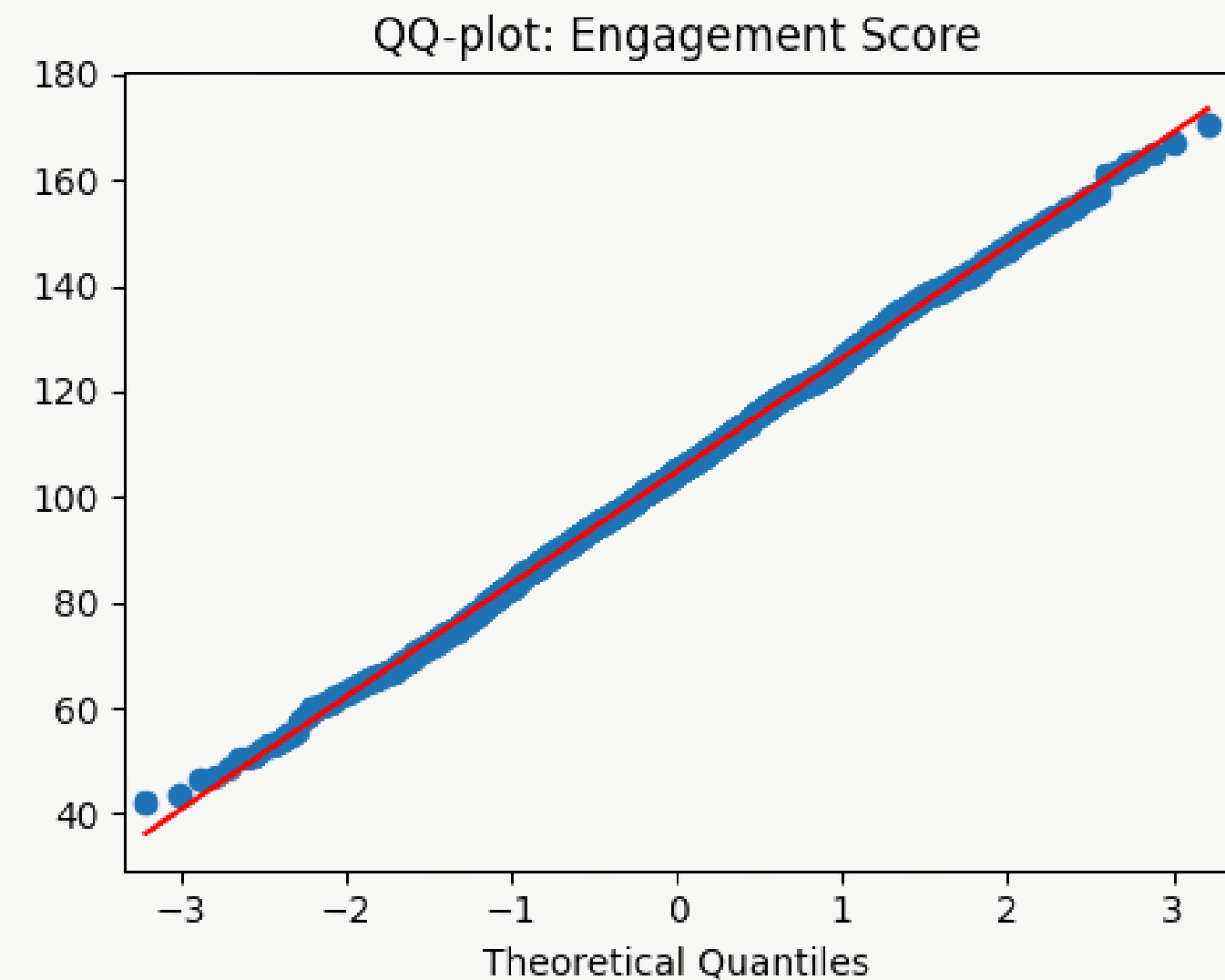
Relationship between
ad budget and
engagement is strong
and similar across
channels → minimal
interaction effect

METHODOLOGY

2. Checking Normality of the data using (Shapiro-Wilk Test)



All three variables have p-values well above 0.05, meaning there is no evidence to suggest they deviate from normality



Q-Q plot confirm the normality of the data as points closely follow the expected straight line

METHODOLOGY

3. Homogeneity of Variances using Levene's test and Group Differences Using ANOVA

```
ANOVA: engagement_score ~ channel
              sum_sq      df      F      PR(>F)
C(channel)    269.521471    2.0    0.294784  0.744735
Residual    684353.984965 1497.0      NaN      NaN

ANOVA: engagement_score ~ gender
              sum_sq      df      F      PR(>F)
C(gender)     92.924119    1.0    0.203352  0.652094
Residual    684530.582318 1498.0      NaN      NaN
```

By Channel: $p = 0.70 \rightarrow$ No significant difference in variances across Email, SMS, and Social Media

By Gender: $p = 0.89 \rightarrow$ No significant difference in variances across genders

METHODOLOGY

4. Multiple linear regressions (4 model variants)

Formulas that will be fitted:

all_vars: engagement_score ~ ad_budget + CTR + sales + C(channel) + C(gender) Adj. R-squared: 0.941

no_ad_budget: engagement_score ~ CTR + sales + C(channel) + C(gender) Adj. R-squared: -0.002

no_ctr: engagement_score ~ ad_budget + sales + C(channel) + C(gender) Adj. R-squared: 0.938

no_ad_no_ctr: engagement_score ~ sales + C(channel) + C(gender) Adj. R-squared: -0.002

Ad Budget is the strongest predictor of engagement score ($p < 0.001$) which means that higher budgets consistently lead to higher engagement.

CTR also has a significant positive effect, but smaller in magnitude than ad budget

METHODOLOGY

4. Multiple linear regressions (4 model variants)

Full model explains 94% of engagement score variation ($R^2 = 0.941$)

Removing ad budget causes R^2 to drop from 94% to almost 0%, confirming that ad budget has a dominant role

Removing CTR has a minimal effect on model fit (R^2 only drops slightly from 94% to 93.8%)

Ad budget allocation is likely to have the greatest impact on engagement

METHODOLOGY

6. Normal distribution of Residuals (Shapiro-Wilk Test)

```
Residual diagnostics for model: all_vars  
Residuals: mean = -0.000000, std = 5.169480, n = 1500  
residuals (all_vars): W-stat=0.9994, p-value=0.9364  
residuals (all_vars) looks normally distributed.
```

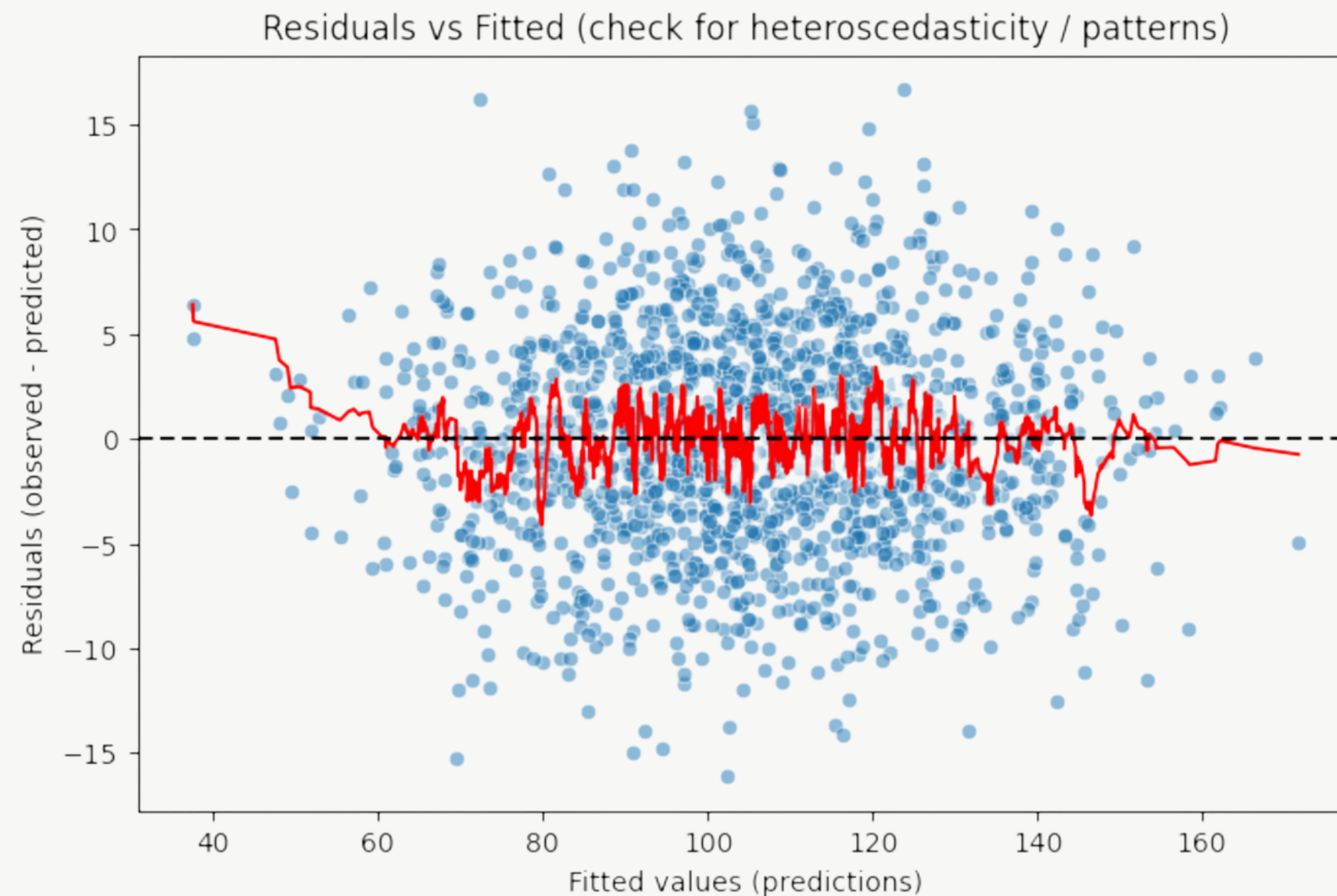
Residuals have a mean 0 i.e. there is no systematic over/under prediction

Standard deviation is ~5.17, consistent with the model's expected error range

Statistical test ($p = 0.94$) confirms residuals are normally distributed meaning that the prediction is reliable and unbiased

METHODOLOGY

6. Breusch-Pagan test

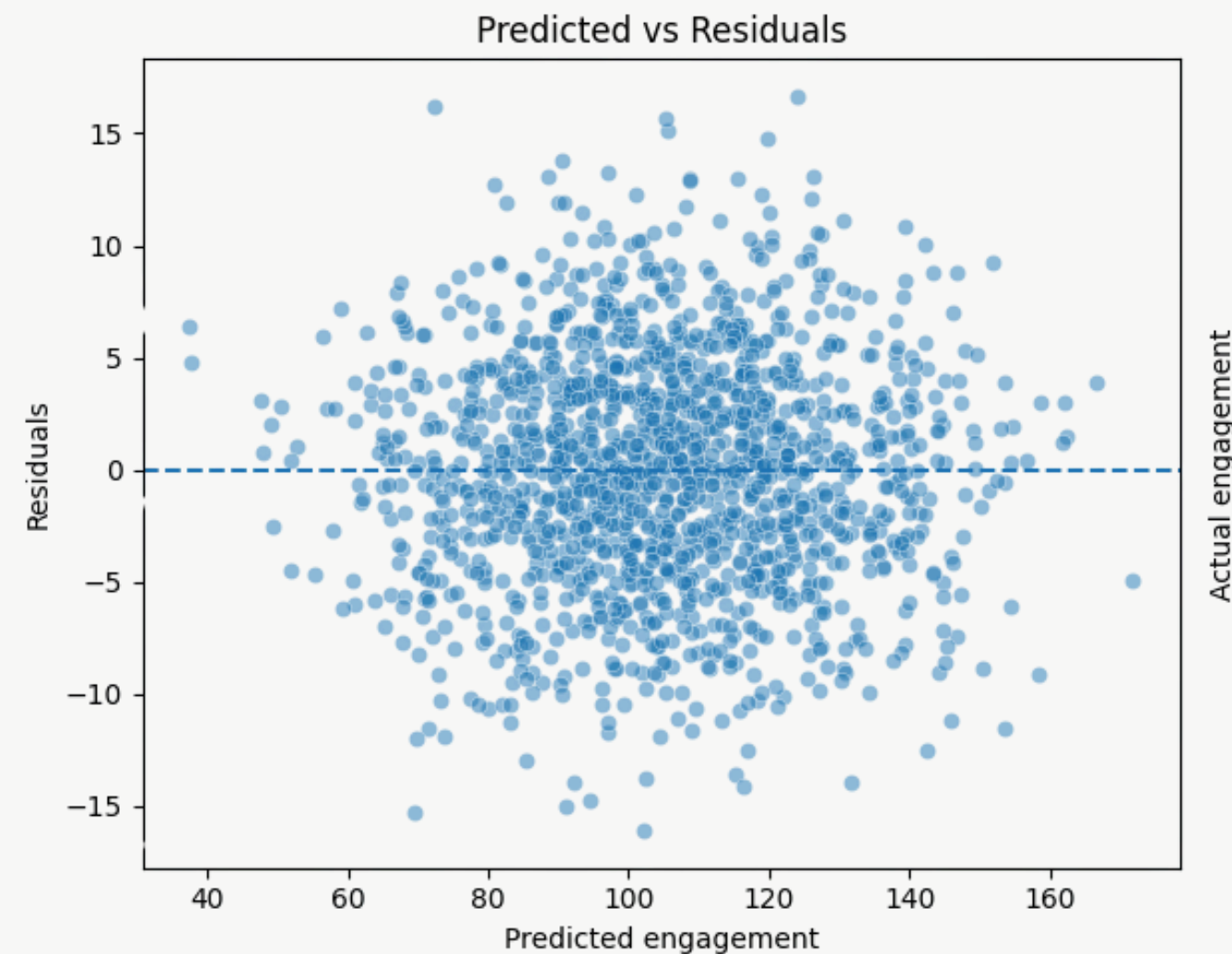


```
Lagrange multiplier stat: 1.48903  
p-value: 0.96022  
f-value: 0.247259  
f p-value: 0.960498
```

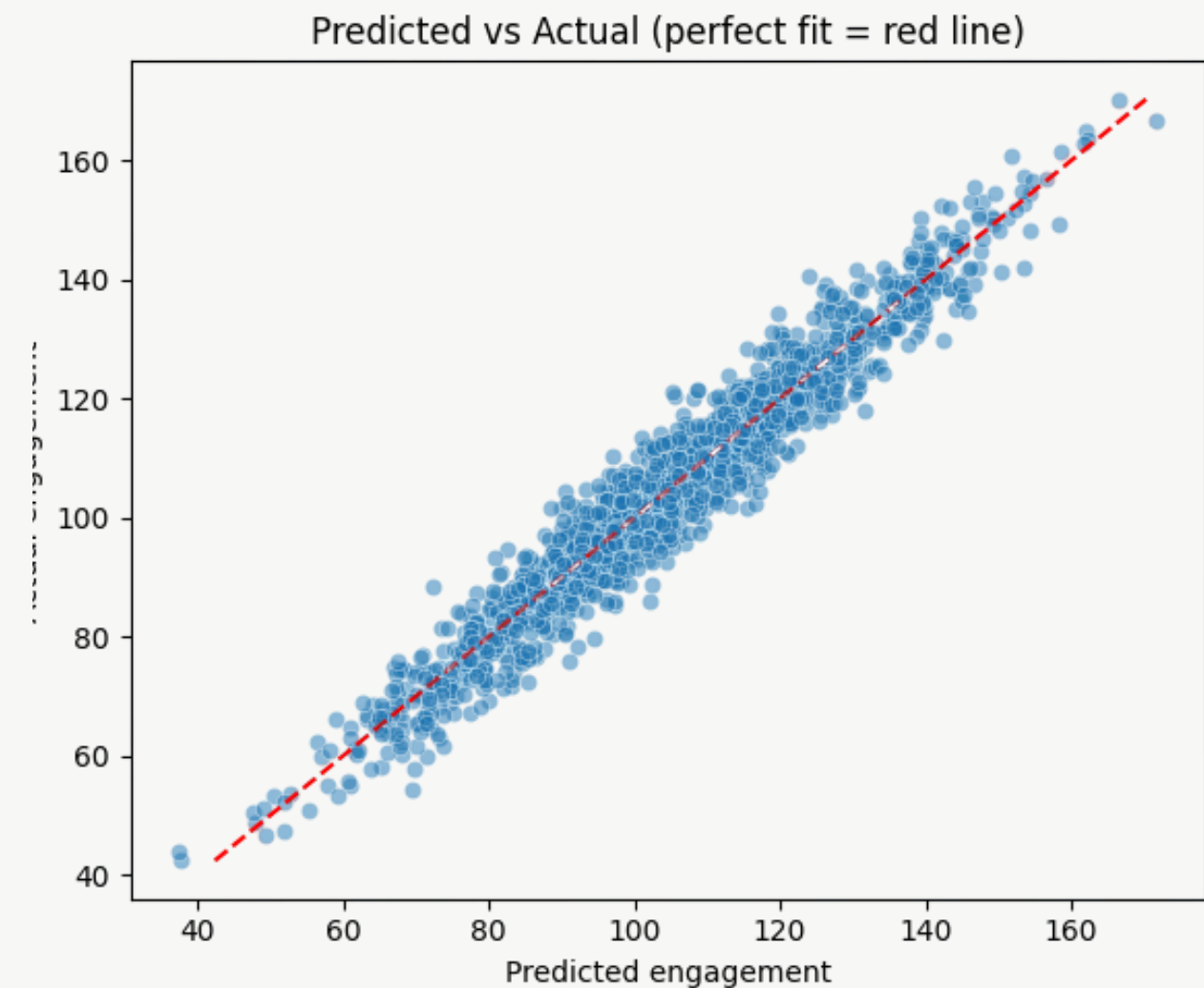
p-value = 0.96 means there is no evidence of heteroscedasticity. Therefore, the model's estimates are statistically valid and efficient

METHODOLOGY

6. Predictions vs errors scatterplot



implies that the predictions are consistent and unbiased and the assumptions are met



confirms that the model's predictions are highly accurate and closely match the actual results

METHODOLOGY

7. Goodness-of-fit comparison (R-squared, adj R-squared)

Best Model: The model using all variables (all_vars) is the best fit, explaining 94.1% of the variation in engagement score.

Key Drivers: The most statistically significant factors driving engagement are Advertising Budget and CTR (Click-Through Rate).

BUSINESS APPLICATIONS OF DATA SCIENCE

THANK YOU

ARUNIMA MANDWARIYA | BHART LAL | NEHA S K | SIDRA SHAIKH