Solution 1. The marketing team can potentially be interested in whether income levels affect the likelihood of customers accepting the final marketing campaign offer (measured by the Response variable). In such a case, we have the following hypotheses:

- **Null Hypothesis (H0):** There is no significant difference in the average income of customers who accepted the offer (Response = 1) compared to those who did not (Response = 0).
- **Alternative Hypothesis (H1):** There is a significant difference in the average income of customers who accepted the offer compared to those who did not.

The reason this hypothesis is interesting to marketeers is because it seeks to understand whether income influences the acceptance rate of marketing campaigns. In turn, it provides strategies for future campaigns - High-income customers may respond better to premium offers or exclusive experiences. Low-income customers might prefer discounts or budget-friendly promotions. This insight can guide segmentation and offer personalization.

````
```{r}
data <- read.csv("marketing_campaignhw3.csv", sep = ';')
```
````

````
```{r}
# Check for rows with null values
null_rows <- data %>% filter_all(any_vars(is.na(.)))

# Print number of null rows
cat("Number of rows with null values:", nrow(null_rows), "\n")
```
````

```
Number of rows with null values: 24
```

````
```{r}
# drop these null rows
data <- data %>% filter_all(all_vars(!is.na(.)))
```
````

```{r}
library(dplyr)
summary_result <- data %>%
  group_by(Response) %>%
  summarise(mean = mean(Income))

summary_result
```

A tibble: 2 × 2

Response <fctr>	mean <dbl>
0	50839.13
1	60209.68

2 rows

```{r}
data$Response <- as.factor(data$Response)
data$Income <- as.numeric(data$Income)

t_test_result <- t.test(Income ~ Response, data = data) #two tailed t test #direction not specified
t_test_result

# Null is no significant difference in the average income of customers who accepted the offer (Response = 1)
compared to those who did not (Response = 0)

# Reject the null - There is a significant difference in the average income of customers who accepted the
offer compared to those who did not.
```

```
	Welch Two Sample t-test

data:  Income by Response
t = -6.7032, df = 482.05, p-value = 5.713e-11
alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
95 percent confidence interval:
 -12117.309  -6623.777
sample estimates:
mean in group 0 mean in group 1 
       50839.13        60209.68 
```

Since the p-value is less than 0.05, we reject the null hypothesis and say that there is a significant difference in average income between customers who accepted and those who did not.

Implications - The marketing team could analyse income segments further to design differentiated campaigns. For example, luxury products might target high-income customers, while discount offers might focus on low-income customers.

```{r}
t_test_result <- t.test(Income ~ Response, data = data, alternative = 'less') # one tail # interested in directionality
print(t_test_result)

# Null is the average income of customers who do not accept the offer (Response = 0) is greater than or equal to those who accept the offer (Response = 1)
# Alternate is the average income of customers who do not accept the offer is less than those who accept the offer
```

```
	Welch Two Sample t-test

data:  Income by Response
t = -6.7032, df = 482.05, p-value = 2.857e-11
alternative hypothesis: true difference in means between group 0 and group 1 is less than 0
95 percent confidence interval:
      -Inf -7066.744
sample estimates:
mean in group 0 mean in group 1
       50839.13        60209.68
```

We also perform a one tailed t test and still end up rejecting the null hypothesis (p value (2.857e-11) is very low). In turn, we can conclude there is evidence in support of the alternate hypothesis or that the average income of customers who do not accept the offer is less than that of those who accept the offer.

Solution 2. The marketing team is likely interested in exploring the relationship between a customer's income and the amount spent on premium products such as gold products (MntGoldProds).

Hypothesis:

- **Null Hypothesis (H_0):** There is no significant correlation between income and spending on premium products (e.g. gold).
- **Alternative Hypothesis (H_1):** There is a significant positive correlation between income and spending on premium products.

The reason this hypothesis is interesting because it explores whether income is strongly correlated with spending on premium products. If so, marketing campaigns for these products can focus on higher-income customers. Also, marketeers would want to tap into cross-selling opportunities. Insights from this analysis can help marketers identify customers who are more likely to purchase premium products and design offers targeting them.

```{r}
data$Income <- as.numeric(data$Income)
data$MntGoldProds <- as.numeric(data$MntGoldProds)
data <- data %>% filter(!is.na(Income) & !is.na(MntGoldProds))

cor_income_gold <- cor(data$Income, data$MntGoldProds)
cor_income_gold

cat("Correlation between Income and Spending on Gold Products:", cor_income_gold, "\n")
```

```
[1] 0.3259164
Correlation between Income and Spending on Gold Products: 0.3259164
```

A correlation value of 0.326 suggests a weak correlation or no meaningful correlation. We fail to reject the null hypothesis. It suggests that income may not be the primary driver of spending on premium products. The marketing team should investigate other factors, such as purchase behavior, family demographics, or lifestyle. We can also conclude from our analysis that while high-income customers may spend slightly more on premium products, a weak correlation means not all high-income customers prioritize premium products. Marketers should adopt a multi-faceted approach, combining income with other behavioural and demographic factors for better segmentation and targeting.

Solution 3. Let's say we are interested in the association of two categorical variables - education level and complaints**.** We want to determine if customers with different education levels are more or less likely to lodge complaints.

Hypothesis:

- **Null Hypothesis (H_0):** There is no association between education level and the likelihood of lodging complaints.
- **Alternative Hypothesis (H_1):** There is a significant association between education level and the likelihood of lodging complaints.

If certain education levels are associated with higher complaint rates, it can indicate varying expectations or communication gaps. Understanding the association can help the marketing team develop personalized communication for different education segments. If some groups are more likely to complain, marketers can proactively target these customers to improve their experience.

````
```{r}
# Ensure variables are factors
data$Education <- as.factor(data$Education)
data$Complain <- as.factor(data$Complain)

# Create a contingency table
education_complain_table <- table(data$Education, data$Complain)

# Perform the chi-square test
chi_square_result <- chisq.test(education_complain_table)

# Print the contingency table and chi-square test results
cat("Contingency Table:\n")
print(education_complain_table)
cat("\nChi-Square Test Result:\n")
print(chi_square_result)
```
````

```
                0    1
  2n Cycle    196    4
  Basic        54    0
  Graduation 1102   14
  Master      363    2
  PhD         480    1

Chi-Square Test Result:

        Pearson's Chi-squared test

data:  education_complain_table
X-squared = 7.4209, df = 4, p-value = 0.1152
```

Since the p-value is greater than 0.05, we fail to reject the null hypothesis, indicating no significant association. Based on the above test, there is no statistically significant association between education level and the likelihood of lodging complaints. In other words, complaints are distributed similarly across all education levels, and education is not a factor influencing whether customers lodge complaints. It suggests that marketers should investigate other factors like income, spending habits, or product preferences.

Additionally, marketers could design standardized complaint-handling strategies that apply to all education levels. If education is not a factor, complaints may be linked to universal issues like product quality, service inefficiencies, or delivery delays.

Solution 4. In the t test, we rejected the null hypothesis, indicating a significant difference in average income between the two groups (response = 0 and 1). However, we cannot make a causal statement with this result. A t-test identifies differences between two groups but does not establish causation. Our test result only indicated that the average incomes differ significantly between customers who accepted the campaign and those who did not. Causation would require evidence that income causes the likelihood of accepting the campaign, which cannot be determined without a controlled experiment or further analysis.

In correlation test, we fail to reject the null hypothesis. The correlation value of 0.326 indicates a weak positive correlation. Again, we cannot make a causal statement with this result. The weak correlation suggests that income is not a strong predictor of spending on premium products, but even with a stronger correlation, causation would require additional evidence. In our case, we found that the relationship between income and spending is somewhat associative, not causal.

Lastly, in the chi-squared test, we failed to reject the null hypothesis, indicating no significant association between education level and complaints. We cannot make a causal statement with this result. The chi-square test only assesses whether an association exists between two categorical variables. It does not indicate causation or explain the nature of the relationship. Even if the test showed a significant association, causation could not be inferred because other variables (e.g., income, customer behavior) could influence the likelihood of complaints.

None of our above tests can establish causation**.** To infer causality, marketers would need to conduct experiments (e.g. randomized controlled trials) or use advanced statistical techniques.