**Please refer the following details:**

1. **Customer Purchase Behavior Insights (from Task 1)**:

    o What were the key **customer segments** (LIFESTAGE, PREMIUM_CUSTOMER) identified?
    Answer:-
    **Analyzing the Graph: Total Sales by Customer Segments**

    **Understanding the Visual:**

    The bar graph presents a breakdown of total sales across different customer segments, categorized by both **Lifestage** and **Customer Segment**.

    ➔ **Key Observations:**
    i.      **Lifestage Impact:**
    o **Older Families** consistently exhibit the highest total sales across all customer segments.

    o **New Families** generally have the lowest sales, suggesting a potential for growth in this segment.

    o **Midage Singles/Couples** show moderate sales across all customer segments.
    ii.     **Customer Segment Performance:**
        a. **Premium** customers, regardless of lifestage, tend to generate the highest sales.
        b. **Budget** customers, especially in younger lifestages, have relatively low sales.

        c. **Mainstream** customers show moderate sales across most lifestages.
    iii.    **Segment-Specific Insights:**
        a. **Older Families:** Premium customers dominate, followed by Mainstream.

        b. **Young Singles/Couples:** Mainstream customers have the highest sales, with Premium and Budget segments showing moderate levels.

        c. **Retirees:** Premium customers lead, followed by Mainstream.

        d. **Young Families:** Mainstream customers have the highest sales, with Budget customers showing moderate levels.
        e. **Older Singles/Couples:** Premium customers dominate, followed by Mainstream.

        f. **New Families:** Mainstream customers have the highest sales, with Budget customers showing moderate levels.

    ➔ **Potential Implications:**

- **Target Market Optimization:** The data suggests that focusing on **Older Families** and **Premium** customers could yield significant revenue.

- **Segment-Specific Marketing Strategies:** Tailoring marketing efforts to the unique needs and preferences of each customer segment could enhance sales.

- **Growth Opportunities:** The relatively low sales from **New Families** and **Budget** customers might indicate potential for expansion through targeted campaigns or product offerings.

➔ **Additional Considerations:**

- **Customer Lifetime Value:** While total sales are important, analyzing customer lifetime value (CLTV) could provide a more comprehensive understanding of each segment's profitability.

- **Customer Satisfaction:** Highlighting customer satisfaction levels within each segment could help identify areas for improvement or opportunities to upsell.

- **Competitive Analysis:** Comparing these sales figures to competitors' performance could reveal market trends and potential areas for differentiation.

o Were there any notable spending patterns or preferences in terms of pack size or brand?
Answer:-

**I - Analyzing the Graph: Average Pack Size by Customer Segments**

➔ **Understanding the Visual:**
The bar graph presents a breakdown of the average pack size across different customer segments, categorized by both **Lifestage** and **Customer Segment**.

**Key Observations:**

i. **Lifestage Impact:**

o **Older Families** generally have the largest average pack size across all customer segments.

o **New Families** and **Young Singles/Couples** typically have smaller average pack sizes.

o **Midage Singles/Couples** and **Retirees** show moderate pack sizes.

ii. **Customer Segment Performance:**

o **Budget** customers, regardless of lifestage, tend to purchase the largest pack sizes.

o **Premium** customers, especially in younger lifestages, often opt for smaller pack sizes.

o **Mainstream** customers show moderate pack sizes across most lifestages.

iii. **Segment-Specific Insights:**

o **Older Families:** Budget customers have the largest average pack size, followed by Mainstream.

o **Young Singles/Couples:** Budget customers have the largest average pack size, with Mainstream and Premium customers showing moderate levels.

o **Retirees:** Budget customers have the largest average pack size, followed by Mainstream.

o **Young Families:** Budget customers have the largest average pack size, with Mainstream and Premium customers showing moderate levels.

o **Older Singles/Couples:** Budget customers have the largest average pack size, followed by Mainstream.

o **New Families:** Budget customers have the largest average pack size, with Mainstream and Premium customers showing moderate levels.

➔ **Potential Implications:**

- **Product Assortment:** The data suggests that offering larger pack sizes could be beneficial for Budget customers across all lifestages.

- **Family Size:** The larger pack sizes purchased by Older Families might indicate a higher number of family members or a preference for bulk buying.

- **Convenience Factors:** Smaller pack sizes might be preferred by younger customers or those living alone due to convenience or storage considerations.

➔ **Additional Considerations:**

- **Product Pricing:** Analyzing the relationship between pack size and price could help determine if larger pack sizes offer a cost-effective option for customers.

- **Environmental Factors:** Considering the environmental impact of packaging, businesses might explore offering refillable or reusable options to cater to eco-conscious customers.

- **Competitive Analysis:** Comparing these pack size preferences to competitors' offerings could reveal market trends and opportunities for differentiation.

II. **Analyzing the Graph: Top 5 Brands by Customer Segments**

**Understanding the Visual:**

The bar graph presents a breakdown of the top 5 brands by total sales across different customer segments, categorized by both **Lifestage** and **Customer Segment**.

➔ **Key Observations:**

i.      **Brand Popularity:**
o   **Kettle** consistently ranks as the top brand across most lifestages and customer segments.

o   **Smiths** and **Doritos** also show strong performance in multiple segments.

o   **Pringles**, **Old**, and **Thins** have more varied popularity across different groups.
ii.     **Lifestage Preferences:**
   • **Older Families** and **Midage Singles/Couples** tend to favor **Kettle** and **Smiths**.
   • **Young Singles/Couples** and **Retirees** show a broader range of brand preferences.
   • **New Families** exhibit a preference for **Kettle** and **Doritos**.

iii.    **Customer Segment Preferences:**
   • **Premium** customers often choose **Kettle** and **Smiths**.
   • **Budget** customers might be more drawn to **Pringles** and **Old**.
   • **Mainstream** customers show a wider range of brand preferences.

iv.     **Segment-Specific Insights:**
   • **Older Families:** Kettle, Smiths, and Doritos are consistently popular.
   • **Young Singles/Couples:** Kettle, Smiths, and Pringles are frequently chosen.
   • **Retirees:** Kettle, Smiths, and Doritos are popular.
   • **Young Families:** Kettle, Smiths, and Doritos are favored.
   • **Older Singles/Couples:** Kettle, Smiths, and Doritos are preferred.
   • **New Families:** Kettle, Smiths, and Doritos are popular.

➔ **Potential Implications:**
• **Brand Positioning:** The data suggests that **Kettle** and **Smiths** have a strong overall brand appeal. Businesses could focus on maintaining or enhancing this position.

• **Segment-Specific Marketing:** Tailoring marketing efforts to the specific preferences of each customer segment could improve brand awareness and sales.

• **Product Innovation:** Introducing new products or flavors that cater to the preferences of specific segments could drive growth.

➔ **Additional Considerations:**
• **Customer Satisfaction:** Analyzing customer satisfaction levels for each brand can help identify areas for improvement or opportunities to upsell.

- **Competitive Analysis:** Comparing these brand preferences to competitors' performance could reveal market trends and potential areas for differentiation.

- **Distribution Channels:** Evaluating the availability of different brands across various retail channels could impact their overall popularity.

  o Any significant trends in **total sales**, **average spend per customer**, or **purchase frequency**?
  Answer:-
  **Analyzing the Graph: Average Spend per Transaction by Customer Segments**

  ➔ **Understanding the Visual:**
  The bar graph presents a breakdown of the average spend per transaction across different customer segments, categorized by both **Lifestage** and **Customer Segment**.

  ➔ **Key Observations:**

  i.  **Lifestage Impact:**
  o **Older Families** and **Midage Singles/Couples** generally have the highest average spend per transaction across all customer segments.

  o **New Families** typically have the lowest average spend.

  o **Young Singles/Couples** and **Retirees** show moderate spending levels.

  ii.  **Customer Segment Performance:**
  - **Premium** customers, regardless of lifestage, consistently have the highest average spend per transaction.
  - **Budget** customers, especially in younger lifestages, tend to spend less per transaction.
  - **Mainstream** customers show moderate spending levels across most lifestages.

  iii.  **Segment-Specific Insights:**
  - **Older Families:** Premium customers have the highest average spend, followed by Mainstream.
  - **Young Singles/Couples:** Mainstream customers have the highest average spend, with Budget customers showing moderate levels.
  - **Retirees:** Premium customers lead, followed by Mainstream.
  - **Young Families:** Mainstream customers have the highest average spend, with Budget customers showing moderate levels.
  - **Older Singles/Couples:** Premium customers dominate, followed by Mainstream.

- **New Families:** Mainstream customers have the highest average spend, with Budget customers showing moderate levels.

➔ **Potential Implications:**

- **Pricing Strategies:** The data suggests that Premium customers are willing to pay more per transaction. Businesses could consider offering premium products or services at higher price points to target this segment.

- **Value Perception:** Analyzing the average spend per transaction can help businesses understand how different customer segments perceive value. This information can be used to optimize product offerings and promotions.

- **Customer Loyalty:** Higher average spend per transaction might indicate greater customer satisfaction and loyalty. Businesses could focus on building stronger relationships with high-spending customers.

➔ **Additional Considerations:**

- **Purchase Frequency:** While average spend per transaction is important, analyzing purchase frequency can provide a more complete picture of customer spending behavior.

- **Customer Lifetime Value:** Calculating customer lifetime value (CLTV) can help assess the long-term profitability of each customer segment.

- **Competitive Analysis:** Comparing these spending figures to competitors' performance could reveal market trends and opportunities for differentiation.

2. **Trial vs Control Store Performance (from Task 2)**:

   o How did the **trial stores** (stores 77, 86, and 88) perform compared to their respective **control stores**?
   Answer:-

I. TRIAL STORE 77 AND CONTROL STORE 233 over a period from July 2018 to June 2019.

The line graph shows a comparison of total sales between Trial Store 77 and Control Store 233 over a period from July 2018 to June 2019.

Here are some key observations:

- **Overall Trend:** Both stores experienced fluctuations in total sales throughout the period. However, Trial Store 77 generally had higher sales compared to Control Store 233.

- **Peak Periods:** Both stores saw peaks in sales during the holiday months of December 2018 and January 2019.

- **Seasonal Fluctuations:** Both stores exhibited seasonal patterns, with sales tending to increase during certain months and decrease during others.

- **Relative Performance:** While Trial Store 77 consistently outperformed Control Store 233, the difference in sales between the two stores varied over time.

II. TRIAL STORE 86 AND CONTROL STORE 155 over a period from July 2018 to June 2019.

The line graph shows a comparison of total sales between Trial Store 86 and Control Store 155 over a period from July 2018 to June 2019.

Here are some key observations:

- **Overall Trend:** Both stores experienced fluctuations in total sales throughout the period. However, Trial Store 86 generally had higher sales compared to Control Store 155.
- **Peak Periods:** Both stores saw peaks in sales during the holiday months of December 2018 and January 2019.
- **Seasonal Fluctuations:** Both stores exhibited seasonal patterns, with sales tending to increase during certain months and decrease during others.
- **Relative Performance:** While Trial Store 86 consistently outperformed Control Store 155, the difference in sales between the two stores varied over time.

III. The line graph shows a comparison of total sales between Trial Store 88 and Control Store 237 over a period from July 2018 to June 2019.
Here are some key observations:
- **Overall Trend:** Both stores experienced fluctuations in total sales throughout the period. However, Trial Store 88 generally had higher sales compared to Control Store 237.
- **Peak Periods:** Both stores saw peaks in sales during the holiday months of December 2018 and January 2019.
- **Seasonal Fluctuations:** Both stores exhibited seasonal patterns, with sales tending to increase during certain months and decrease during others.
- **Relative Performance:** While Trial Store 88 consistently outperformed Control Store 237, the difference in sales between the two stores varied over time.

o   Did you observe any significant changes in sales or customer growth?

Answer:-

I.      The total sales between a trial group and a control group.

Here are the key observations:

- **Average Sales:** The trial group and the control group have **approximately the same average total sales**. This is evident from the height of the bars, which are nearly identical.

- **Variability:** The **trial group appears to have slightly higher variability** in sales compared to the control group. This is suggested by the larger dot above the trial bar, indicating a greater spread of data points.

II.     The total sales between a trial group and a control group.

Here are the key observations:

- **Average Sales:** The trial group and the control group have **approximately the same average total sales**. This is evident from the height of the bars, which are nearly identical.
- **Variability:** The **trial group appears to have slightly higher variability** in sales compared to the control group. This is suggested by the larger dot above the trial bar, indicating a greater spread of data points.
- While the graph shows that there is no significant difference in the average total sales between the two groups, the higher variability in the trial group might warrant further investigation to understand the reasons for this variation.

o   Were there any statistical tests (e.g., t-tests) performed to validate the significance of these differences?

Answer – Yes, t-tests were held for stores 77, 86 and 88 respectively.

I.      TRIAL STORE 77:-

```
3.  # TRIAL STORE 77
4.  import matplotlib.pyplot as plt
```

```
 5.  from  scipy.stats  import  ttest_ind
 6.
 7.  # Perform t-test to check if sales during the trial are significantly different
 8.  t_stat,  p_value  =  ttest_ind(trial_metrics_77['total_sales'],  control_metrics_77['total_sales'])
 9.
10.  # Print the t-statistic and p-value
11.  print(f"T-test result: t-statistic = {t_stat}, p-value = {p_value}")
12.
13.  # Check the significance of the p-value
14.  if  p_value  <  0.05:
15.      print("The difference in total sales is statistically significant.")
16.  else:
17.      print("The difference in total sales is not statistically significant.")
18.
19.  # Create a bar plot to visualize the mean sales for each group
20.  plt.bar(['Trial', 'Control'],  [trial_metrics_77['total_sales'].mean(),  control_metrics_77['total_s
     ales'].mean()])
21.  plt.xlabel('Group')
22.  plt.ylabel('Total Sales')
23.  plt.title('Comparison of Total Sales Between Trial and Control Groups')
24.
25.  # Add error bars representing standard errors
26.  plt.errorbar(['Trial', 'Control'],  [trial_metrics_77['total_sales'].mean(),  control_metrics_77['t
     otal_sales'].mean()],
27.                   yerr=[trial_metrics_77['total_sales'].std()  /  len(trial_metrics_77),  control_
     metrics_77['total_sales'].std()  /  len(control_metrics_77)],
28.                   fmt='o')
29.
30.  # Show the plot
31.  plt.show()
```

Output:-

T-test result: t-statistic = 0.9679236368390388, p-value = 0.34360373343122186

The difference in total sales is not statistically significant.


II. TRIAL STORE 86:-

# TRIAL STORE 86

# Perform t-test to check if sales during the trial are significantly different

t_stat, p_value = ttest_ind(trial_metrics_86['total_sales'], control_metrics_86['total_sales'])


# Print the t-statistic and p-value

```python
print(f"T-test result: t-statistic = {t_stat}, p-value = {p_value}")

# Check the significance of the p-value
if  p_value  <  0.05:
        print("The difference in total sales is statistically significant.")
else:
        print("The difference in total sales is not statistically significant.")

# Create a bar plot to visualize the mean sales for each group
plt.bar(['Trial',  'Control'],  [trial_metrics_86['total_sales'].mean(),  control_metrics_86['total_sales'].mean()])
plt.xlabel('Group')
plt.ylabel('Total Sales')
plt.title('Comparison of Total Sales Between Trial and Control Groups')

# Add error bars representing standard errors
plt.errorbar(['Trial',  'Control'],  [trial_metrics_86['total_sales'].mean(),  control_metrics_86['total_sales'].mean()],
                        yerr=[trial_metrics_86['total_sales'].std()  /  len(trial_metrics_86),  control_metrics_86['total_sales'].std()  /  len(control_metrics_86)],
                        fmt='o')

# Show the plot
plt.show()
```

```
OUTPUT:-
T-test result: t-statistic = 0.019793022667677837, p-value =
0.9843869025850827
The difference in total sales is not statistically significant.
```

## IV.    TRIAL STORE 88:-

```python
# TRIAL STORE 88
# Perform t-test to check if sales during the trial are significantly different
t_stat, p_value  =  ttest_ind(trial_metrics_88['total_sales'],  control_metrics_88['total_sales'])

# Print the t-statistic and p-value
print(f"T-test result: t-statistic = {t_stat}, p-value = {p_value}")

# Check the significance of the p-value
if  p_value  <  0.05:
        print("The difference in total sales is statistically significant.")
else:
        print("The difference in total sales is not statistically significant.")

# Create a bar plot to visualize the mean sales for each group
```

```
XIX.     plt.bar(['Trial', 'Control'], [trial_metrics_88['total_sales'].mean(), control_metrics_88['total_sal
         es'].mean()])
XX.      plt.xlabel('Group')
XXI.     plt.ylabel('Total Sales')
XXII.    plt.title('Comparison of Total Sales Between Trial and Control Groups')
XXIII.
XXIV.    # Add error bars representing standard errors
XXV.     plt.errorbar(['Trial', 'Control'], [trial_metrics_88['total_sales'].mean(), control_metrics_88['tot
         al_sales'].mean()],
XXVI.                      yerr=[trial_metrics_88['total_sales'].std() / len(trial_metrics_88), control_m
         etrics_88['total_sales'].std() / len(control_metrics_88)],
XXVII.                     fmt='o')
XXVIII.
XXIX.    # Show the plot
XXX.     plt.show()
```

OUTPUT:-

T-test result: t-statistic = 1.9781599338138325, p-value = 0.06056801623928565

The difference in total sales is not statistically significant.


**3.Key Visualizations**:

     o   If you generated any graphs or charts, please describe them briefly (e.g., bar charts, line graphs comparing trial and control store sales).

Answer:-

All the graphs, charts have been described in above two points(Task 1 and Task 2)