BAMA 517 Data Driven Marketing

Capstone Project: Quantium

Group 7
Angga Adi Saputra | 51010908 | angga23@student.ubc.ca
Namith Mohan | 45130770 | namith22@student.ubc.ca
Tony Yang | 68638030 | ty070101@student.ubc.ca
Yihan Yang | 91179846 | yihany@student.ubc.ca
Yitong Wang | 55240584 | ywan3672@student.ubc.ca

Table of Content

Project Background

The team is tasked with delivering a strategic recommendation to Julia, supported by data that she can utilize for the upcoming category review. To achieve this, an analysis of the current purchasing trends and behaviors is underway. The client is especially interested in understanding customer segments and their behavior regarding chip purchases. Various metrics are being considered to effectively describe the customers' purchasing behavior. The client is keen on receiving a data-driven recommendation based on the analysis, aiding in the decision of whether to extend the trial layout to all stores.

Quantium's retail analytics team has been approached by their client, the Category Manager for Chips. The Category Manager has tasked the team with testing the impact of the new trial layouts and providing a data-driven recommendation on whether or not the trial layout should be rolled out to all the stores.

Data Cleaning

The dataset underwent several preprocessing steps to ensure its integrity and relevance for subsequent analysis. Initially, a review of the DATE column revealed its data type as integers, prompting conversion to actual date values for accurate temporal analysis.

```
# Create a Corpus
corpus <- Corpus(VectorSource(transaction_data$PROD_NA
ME))

# Preprocess the Text
corpus <- tm_map(corpus, content_transformer(tolower))
corpus <- tm_map(corpus, removePunctuation)
corpus <- tm_map(corpus, removeNumbers)
corpus <- tm_map(corpus, removeWords, stopwords("en"))

# Remove punctuation, numbers and common English
stopwords

# Create a Document-Term Matrix (DTM)
dtm <- DocumentTermMatrix(corpus)

# Convert DTM to a Data Frame
dtm_df <- as.data.frame(as.matrix(dtm))

# Summarize Word Frequencies
word_frequencies <- colSums(dtm_df)
sorted_word_frequencies <- sort(word_frequencies,
decreasing = TRUE)

# Display 20 Most Common Words
head(sorted_word_frequencies, 20)
```

```
chips   kettle   smiths     salt   cheese pringles  doritos  crinkle     corn original
 49770    41288    28860    27976    27890    25102    24962    23960    22063    21560
   cut     chip    salsa  chicken      sea    thins   chilli     sour   crisps  vinegar
 20754    18645    18094    15407    14145    14075    13895    13882    12607    12402
```

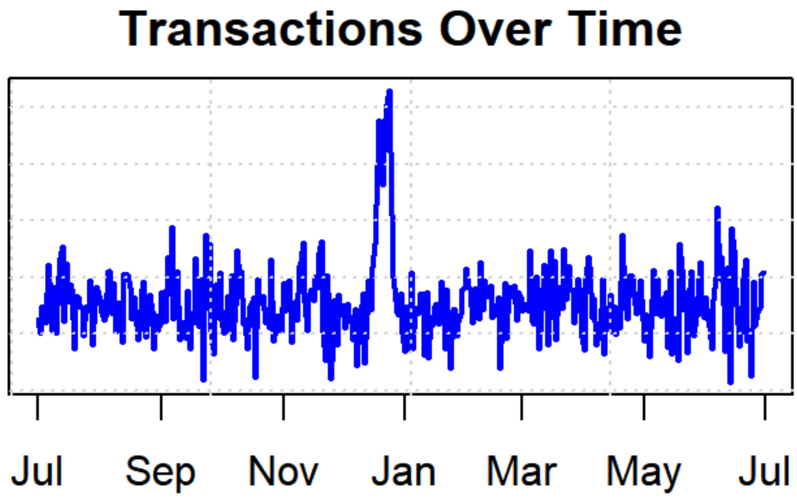
Subsequently, a text analysis was conducted to validate the product category. Insignificant elements such as punctuation, numbers, and common English stop words were removed, and the 20 most frequent words were identified. Notably, 'salsa' emerged as the 13th most common word, leading to its exclusion from the dataset.

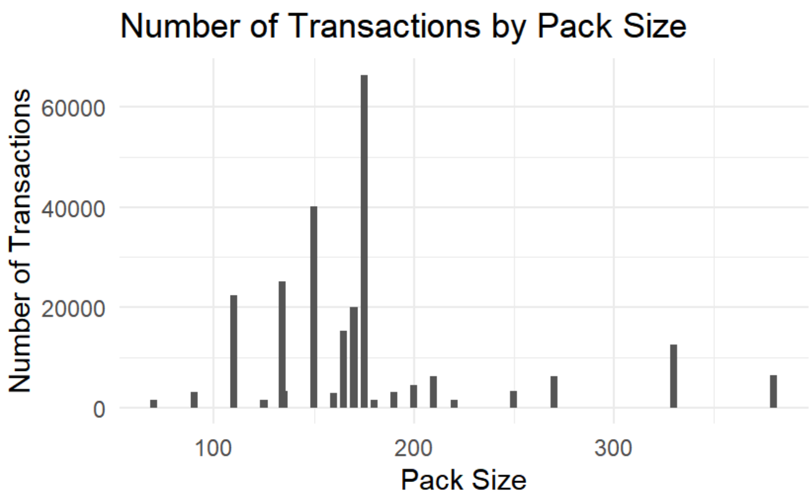
```
   STORE_NBR     LYLTY_CARD_NBR        TXN_ID           PROD_NBR
 Min.   :  1.0   Min.   :   1000   Min.   :      1   Min.   :  1.00
 1st Qu.: 70.0   1st Qu.:  70015   1st Qu.:  67569   1st Qu.: 26.00
 Median :130.0   Median : 130367   Median : 135183   Median : 53.00
 Mean   :135.1   Mean   : 135531   Mean   : 135131   Mean   : 56.35
 3rd Qu.:203.0   3rd Qu.: 203084   3rd Qu.: 202654   3rd Qu.: 87.00
 Max.   :272.0   Max.   :2373711   Max.   :2415841   Max.   :114.00
    PROD_QTY         TOT_SALES
Min.   :  1.000   Min.   :  1.700
1st Qu.:  2.000   1st Qu.:  5.800
Median :  2.000   Median :  7.400
Mean   :  1.908   Mean   :  7.321
3rd Qu.:  2.000   3rd Qu.:  8.800
Max.   :200.000   Max.   :650.000
```

DATE <date>	STORE_NBR <int>	LYLTY_CARD_NBR <int>	TXN_ID <int>	PROD_NBR <int>	PROD_NAME <chr>	PROD_QTY <int>	TOT_SALES <dbl>
2018-08-19	226	226000	226201	4	Dorito Corn Chp Supreme 380g	200	650
2019-05-20	226	226000	226210	4	Dorito Corn Chp Supreme 380g	200	650

LYLTY_CARD_NBR <int>	LIFESTAGE <chr>	PREMIUM_CUSTOMER <chr>
226000	OLDER FAMILIES	Premium

No null values were found, but upon summarizing the dataset, a significant disparity between the max and mean of total sales was observed. Data points where total sales exceeded $650 were filtered out. From the above results, there are only two transactions of the same product and the same quantity which are made by the same customer. With a closer look, this customer is from the older family and has a premium status. Finally, these outliers are removed.





Exploratory data analysis involved mapping transaction amounts over time, revealing a surge in sales in late December and early January. This might be potentially due to the Christmas and winter break. Extraction of package size and brand name from the Product Name column facilitated the creation of new columns for subsequent analysis. From the graph, it seems customers tend to purchase the 175g package size of the chip instead of larger ones. Ultimately, the cleaned transaction dataset was merged with the customer dataset to enable a comprehensive analysis of the combined data.

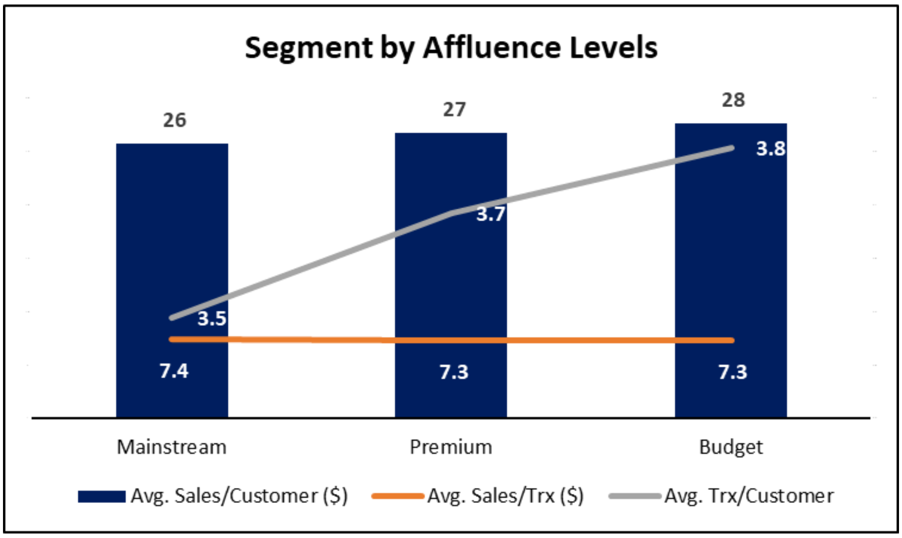
Business Overview

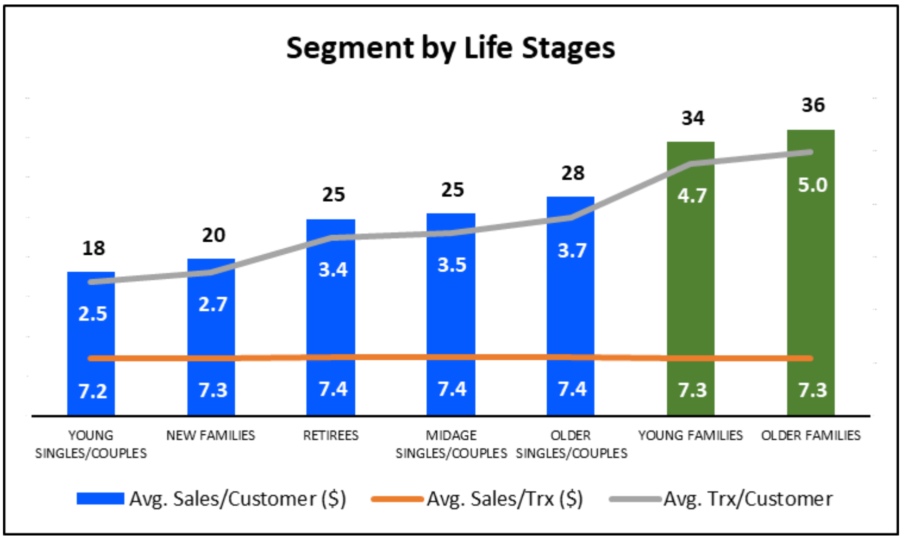
Client Overview

Based on the cleaned dataset, topline analysis was first conducted to gain a good grasp of the client's current performance. The dataset shows that our client has 114 different types of chips (21 different brands and 21 different pack sizes ranging from 70 gr - 380 gr) distributed in 272 stores with more or less 73 thousand customers. They generate around $1.9 million revenue/year (Jul 2018 - Jun 2019), with over 260 thousand transactions and 500 thousand products sold.

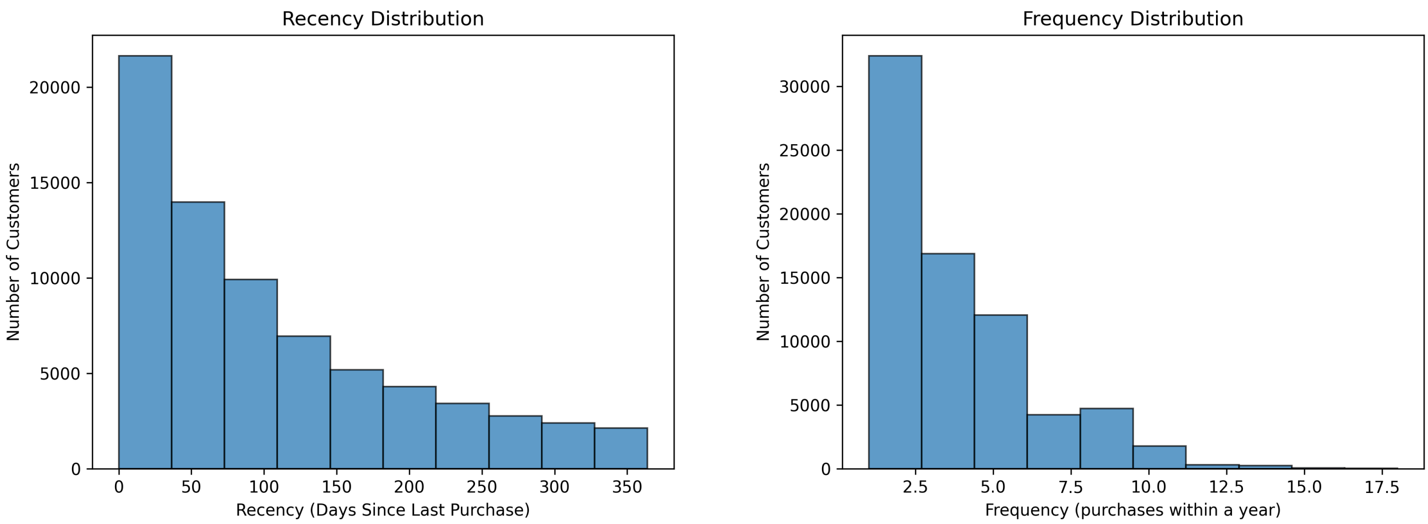
Customer Overview

The client possesses a total of 21 sub-segments comprising 3 distinct affluence levels (Budget, Mainstream, and Premium) and 7 life stages (Midage singles/couples, New families, Older families, Older singles/couples, Retirees, Young families, Young singles/couples.





After analyzing the purchasing patterns, it becomes evident that the average spent/customer on chips falls within the range of \$18 to \$36. Notably, the average sales per transaction remain consistent across the diverse segments (\$7.2 - \$7.4). Consequently, it implies that the variability in the average sales/customer is influenced by disparities in the number of transactions/customers rather than the amount spent/transaction. An examination of the average number of transactions/customers reveals a strong positive correlation between average expenditure and the number of transactions/customers. Based on this finding, deeper analysis was conducted to further define customer purchase behavior in terms of recency and frequency.



Recency:

The graph represents the recency of the customer purchases which spans over the period of one year on the X-axis and the number of customers within each recency bucket on the Y-axis. The recency score of the customers are divided into three buckets which are from July - October, November - February, and March - June. The more recent the visits, the higher the Recency Score. From the depicted graphical trend, it can be inferred that approximately only 50% of customers make purchases within the last 50 days. Businesses can leverage this graph to identify crucial customer segments and formulate targeted marketing strategies aimed at enhancing sales by encouraging the recency of transactions among customers.

Frequency:

The graph illustrates the distribution of customer frequency, representing the number of times customers made purchases within a one-year timeframe. The frequency ranges from 1 to 18 purchases. Notably, the majority of customers make an average of 2.5 purchases per year. Businesses can extract valuable insights from this distribution by analyzing seasonal trends, devising customer retention strategies, and implementing marketing initiatives. Ultimately, these efforts aim to optimize their approach and increase the frequency of purchases within their customer base.

Recency Codes:

Recency code

In [6]:
```python
df['DATE'] = pd.to_datetime(df['DATE'])

df_recency = df.groupby(by='LYLTY_CARD_NBR', as_index=False)['DATE'].max()
df_recency.columns = ['LYLTY_CARD_NBR', 'LastPurchaseDate']
recent_date = df_recency['LastPurchaseDate'].max()
df_recency['Recency'] = df_recency['LastPurchaseDate'].apply(
    lambda x: (recent_date - x).days)

# Reverse the recency score so that a higher score corresponds to more recent purchases
df_recency['Recency'] = df_recency['Recency'].max() - df_recency['Recency']

df_recency
```

Out[6]:

	LYLTY_CARD_NBR	LastPurchaseDate	Recency
0	1000	2018-10-17	108
1	1002	2018-09-16	77
2	1003	2019-03-08	250
3	1004	2018-11-02	124
4	1005	2018-12-28	180
...	...	...	...
72631	2370651	2018-08-03	33
72632	2370701	2018-12-08	160
72633	2370751	2018-10-01	92
72634	2370961	2018-10-27	118
72635	2373711	2018-12-14	166

72636 rows × 3 columns

Recency Score

In [7]:
```python
# Assuming 'DATE' is the column name containing dates in string format
df['DATE'] = pd.to_datetime(df['DATE'])

# Calculate recency
df_recency = df.groupby(by='LYLTY_CARD_NBR', as_index=False)['DATE'].max()
df_recency.columns = ['LYLTY_CARD_NBR', 'LastPurchaseDate']
recent_date = df_recency['LastPurchaseDate'].max()
df_recency['Recency'] = (recent_date - df_recency['LastPurchaseDate']).dt.days

# Define recency score ranges
bins = [0, 120, 240, 365]
labels = [3, 2, 1]  # Assign scores in reverse order

# Assign recency scores
df_recency['RecencyScore'] = pd.cut(df_recency['Recency'], bins=bins, labels=labels, right=True)

df_recency
```

Out[7]:

	LYLTY_CARD_NBR	LastPurchaseDate	Recency	RecencyScore
0	1000	2018-10-17	256	1
1	1002	2018-09-16	287	1
2	1003	2019-03-08	114	3
3	1004	2018-11-02	240	2
4	1005	2018-12-28	184	2
...	...	...	...	...
72631	2370651	2018-08-03	331	1
72632	2370701	2018-12-08	204	2
72633	2370751	2018-10-01	272	1
72634	2370961	2018-10-27	246	1
72635	2373711	2018-12-14	198	2

Frequency Codes:

FREQUENCY

In [14]:
```
frequency_df = df.drop_duplicates().groupby(
    by=['LYLTY_CARD_NBR'], as_index=False)['TXN_ID'].count()
frequency_df.columns = ['LYLTY_CARD_NBR', 'Frequency']
frequency_df
```

Out[14]:

	LYLTY_CARD_NBR	Frequency
0	1000	1
1	1002	1
2	1003	2
3	1004	1
4	1005	1
...	...	...
72631	2370651	1
72632	2370701	1
72633	2370751	1
72634	2370961	2
72635	2373711	1

Frequency score

1-6 vists - score 1

6-12 visits - score 2

12-18 visits - score 3

In [15]:
```
import pandas as pd

# Assuming 'DATE' is the column name containing dates in string format
df['DATE'] = pd.to_datetime(df['DATE'])

# Calculate frequency
frequency_df = df.drop_duplicates().groupby(
    by=['LYLTY_CARD_NBR'], as_index=False)['DATE'].count()
frequency_df.columns = ['LYLTY_CARD_NBR', 'Frequency']

# Define frequency score ranges
bins = [1, 6, 12, 18, float('inf')]
labels = [1, 2, 3, 4]

# Assign frequency scores
frequency_df['FrequencyScore'] = pd.cut(frequency_df['Frequency'], bins=bins, labels=labels, right=False)

frequency_df
```

Out[15]:

	LYLTY_CARD_NBR	Frequency	FrequencyScore
0	1000	1	1
1	1002	1	1
2	1003	2	1
3	1004	1	1
4	1005	1	1
...	...	...	...
72631	2370651	1	1
72632	2370701	1	1

Problem Statement

Based on further findings above, it has been apparent that there has been a significant challenge related to low customer purchase frequency, impacting overall revenue and sales performance of the business. With reference to Americans' average annual chips consumption of 6.6 lbs/person - ±3,000 gr/person (Tensley, 2002), there is ample opportunity to drive current purchase frequency (2.5 times/year) based on customer loyalty and preference.

Customer Segmentation

Lifestage Group (Greater Sales Variability)
This group has a larger variance in sales. Imagine a group of customers with diverse spending habits - some make frequent small purchases, others make occasional large purchases, and the rest fall somewhere in between. This diversity in purchasing behavior leads to a wider range of sales figures within the group. However, despite this variability, the group as a whole still shows a consistent pattern or trend that passes the ANOVA test, indicating that the differences in sales among these customers are statistically significant.

Premium Group (Smaller Sales Variability)
The premium group, on the other hand, exhibits less variability in sales. Here, imagine customers with more uniform purchasing behaviors - perhaps most of them make regular, high-value purchases. The sales figures within this group are more consistent and less spread out compared to the Lifestage group. Like the Lifestage group, the premium group's sales pattern also passes the ANOVA test, suggesting that any differences in sales are not due to random chance but are still statistically significant.

The key takeaway is that while both groups show statistically significant patterns in sales (as evidenced by passing the ANOVA test), the Lifestage group exhibits a broader range of sales behaviors compared to the more uniform sales patterns in the premium group.

	user	profit_per_purchase	profit_per_user
LIFESTAGE			
OLDER SINGLES/COUPLES	14609	7.386823	27.546495
RETIREES	14805	7.364325	24.753185
OLDER FAMILIES	9779	7.253307	36.043276
YOUNG FAMILIES	9178	7.252709	34.447603
YOUNG SINGLES/COUPLES	14441	7.158515	18.032359
MIDAGE SINGLES/COUPLES	7275	7.357678	25.395368
NEW FAMILIES	2549	7.289124	19.785583

	user	profit_per_purchase	profit_per_user
PREMIUM_CUSTOMER			
Mainstream	29245	7.361106	25.670867
Budget	24470	7.258838	27.634309
Premium	18921	7.263111	26.751173

ANOVA tests for both groups:

```
anova - premium group
p value is 3.8067935691157215e-28 < 0.05, null hypothesis is rejected
anova - lifestage group
p value is 3.8067935691157215e-28 < 0.05, null hypothesis is rejected
```

Lifestage contribution group (group 5 is the highest contribution group)

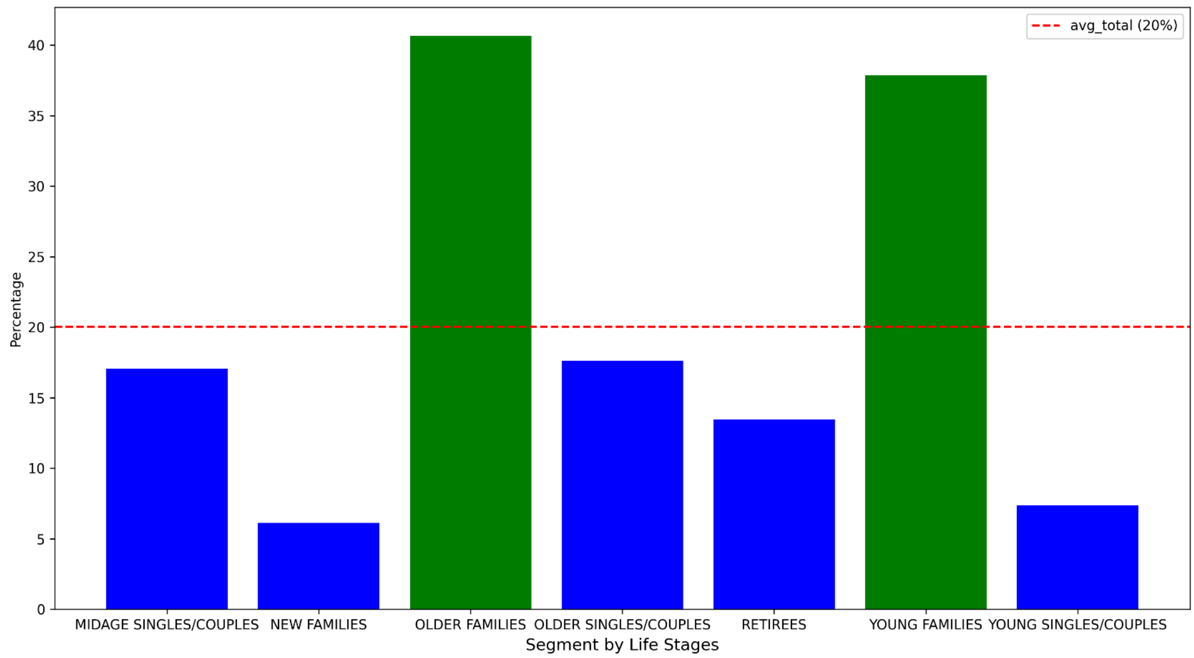
Further investigation is done by dividing the customers into five distinct groups based on their sales contributions (5 is the highest). The objective is to identify which lifestage segment is more likely to be high-contributing consumers. Chi-square test showed significant differences among groups. The bar chart illustrates the percentage of individuals within various life stage segments that also fall into the highest contribution group. The red dashed line represents the average percentage across all segments set at 20% The chart indicates strategies should focus on Families. Older and Young Families represent key target markets for products and services. Tailored marketing campaigns, family-sized packaging, and value offers could be effective. While New Families currently fall below the average percentage in terms of their representation in the highest contribution group, they are a critical demographic that cannot be ignored. As these families transition into the 'Older Families' or 'Young Families' segments over time, their purchasing power and priorities are likely to change, potentially increasing their contribution to the market.

Customers is divided into 5 groups based on their contribution to sales

sales_group LIFESTAGE	Group 1	Group 2	Group 3	Group 4	Group 5
MIDAGE SINGLES/COUPLES	20.426117	20.316151	20.920962	21.264605	17.072165
NEW FAMILIES	24.440957	25.068654	23.656336	20.714005	6.120047
OLDER FAMILIES	15.676450	14.735658	15.298088	13.641477	40.648328
OLDER SINGLES/COUPLES	17.386543	16.832090	19.152577	28.995824	17.632966
RETIREES	20.148598	19.000338	20.412023	26.984127	13.454914
YOUNG FAMILIES	16.877315	15.929396	15.014164	14.295053	37.884071
YOUNG SINGLES/COUPLES	29.007686	28.349837	24.568936	10.698705	7.374836

```
Chi-squared Statistic: 90.7470264582343
P-value: 1.0840048935199542e-09
Degrees of Freedom: 24
p value is 1.0840048935199542e-09 < 0.05, null hypothesis is rejected
```

40% of customers of family group are also in the highest contribution group as shown below:

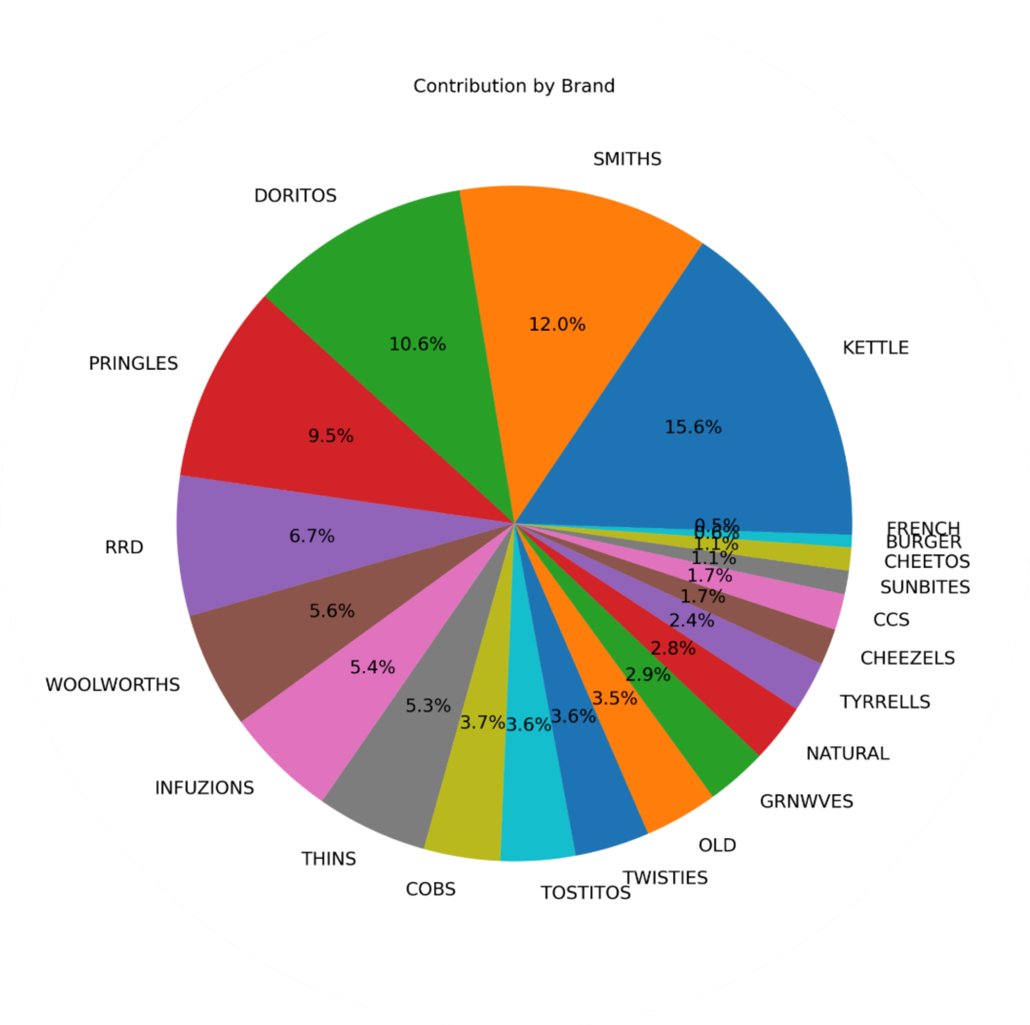


Market Dynamics

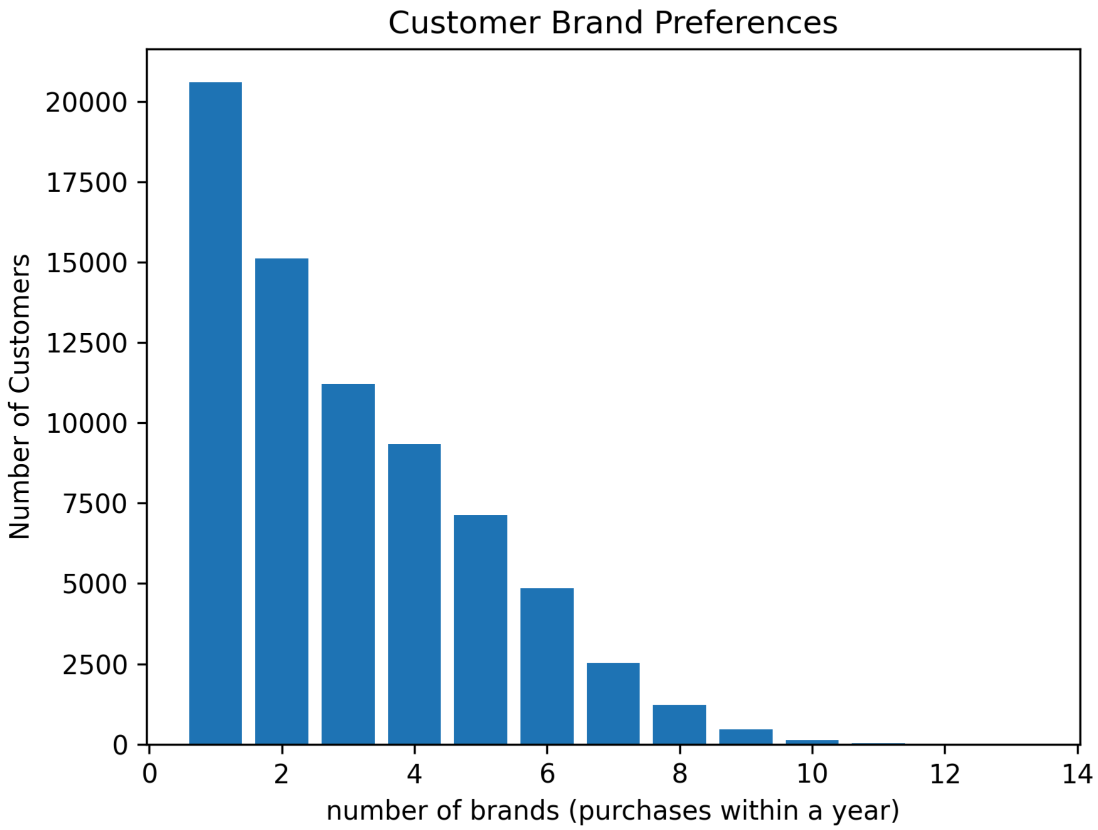
The chip market exhibits a competitive landscape with a variety of players, each holding a portion of the market share. Kettle leads with a significant share of 15.6%, indicating a strong consumer preference or wide market reach. Smiths follows closely at 12.0%, showcasing its prominence in the market. Doritos and Pringles also hold substantial portions, with 10.6% and 9.5% respectively, suggesting their strong market presence. These top four brands collectively account for almost half of the market share, highlighting their dominance in the snack industry. RRD, Woolworths, and Infuzions are mid-level competitors with shares ranging from 5.4% to 6.7%, which points to a moderate but stable market position. The remaining brands, including Thins, Cobs, Tostitos, Twisties, Natural, GRNWVES, and Old, each have between 2.8% to 5.3% of the market share. This indicates a more fragmented market segment with a mix of niche players and smaller brands that cater to specific consumer preferences or demographics. Lastly, a group of brands including Bench, Cheetos, Sunbites, CCS, Cheezels, and Tyrrells have the smallest individual shares, ranging from 1.1% to 2.9%. These brands may either be emerging entities, niche products, or ones with limited market penetration. Overall, the market dynamics suggest a competitive environment where a few brands have a strong hold on the market, while a number of smaller brands are vying for consumer attention.

```python
Analysis of Sales Revenue and Volume by Brand
# Grouping data by 'BRAND' to understand sales revenue and volume
brand_sales = sales_data.groupby(['BRAND']).agg(
    total_sales_revenue=pd.NamedAgg(column='TOT_SALES', aggfunc='sum'),
    total_sales_volume=pd.NamedAgg(column='PROD_QTY', aggfunc='sum'),
    total_sales_times=pd.NamedAgg(column='PROD_QTY', aggfunc='count')
).sort_values(by=['total_sales_revenue'], ascending=False)
brand_sales['profit_per_unit'] = brand_sales['total_sales_revenue'] / brand_sales['total_sales_volume']
brand_sales['profit_per_chase'] = brand_sales['total_sales_revenue'] / brand_sales['total_sales_times']
brand_sales
```

	total_sales_revenue	total_sales_volume	total_sales_times	profit_per_unit	profit_per_chase
BRAND					
KETTLE	390239.8	79051	41288	4.936557	9.451652
DORITOS	240590.9	53816	28145	4.470620	8.548264
SMITHS	224660.2	60339	31823	3.723300	7.059680
PRINGLES	177655.5	48019	25102	3.699692	7.077344
INFUZIONS	99047.6	27119	14201	3.652332	6.974692



New Segmentation



The bar chart depicting customer brand preferences in terms of the number of different brands they purchase within a year suggests that a significant number of customers are relatively brand-agnostic, buying from multiple brands, while a smaller segment displays higher brand loyalty, purchasing from fewer brands. To craft effective market strategies from this data, here are some suggestions:

Loyalists:

For customers showing high brand loyalty (purchasing from 1-2 brands), personalized marketing campaigns can reinforce their loyalty. This could include loyalty programs, personalized discounts, and rewards for repeat purchases.

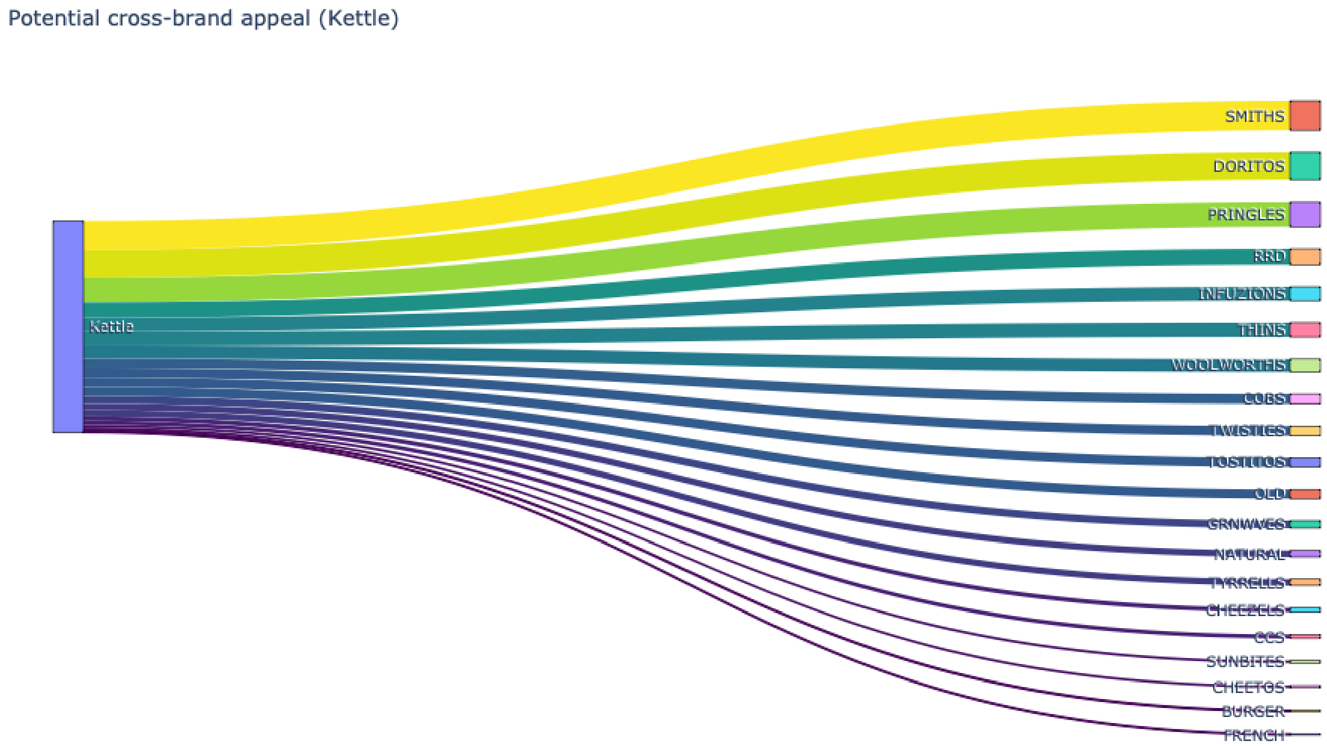
Explorer:

For customers purchasing more than 3 brands, strategies could focus on diversification, such as offering bundle deals that include a variety of products or flavors. the deals could be based on the preference of consumer who also have bought products of the same brand, as below :

```python
sales_brand = sales_data[['BRAND','TOT_SALES']].groupby(by = 'BRAND').sum()
sales_brand = sales_brand.sort_values(by = 'TOT_SALES', ascending = False)
# Calculate cumulative sum of sales
sales_brand['cumulative_sales'] = sales_brand['TOT_SALES'].cumsum()

# Calculate total sales
total_sales = sales_brand['TOT_SALES'].sum()

# Calculate cumulative percentage
sales_brand['cumulative_percentage'] = (sales_brand['cumulative_sales'] / total_sales) * 100
sales_brand
```

```
Perference of people who has bought Keetle
               count
BRAND
SMITHS      0.685362
DORITOS     0.652370
PRINGLES    0.584979
RRD         0.367716
INFUZIONS   0.327641
THINS       0.327447
WOOLWORTHS  0.305711
COBS        0.230459
TWISTIES    0.222600
TOSTITOS    0.220513
OLD         0.217214
GRNWVES     0.177090
NATURAL     0.152928
TYRRELLS    0.150357
CHEEZELS    0.103585
CCS         0.093591
SUNBITES    0.060599
CHEETOS     0.059289
BURGER      0.034351
FRENCH      0.029838
```

It also can be illustrated as below :



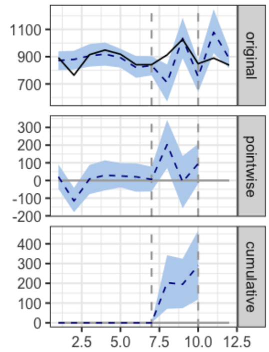
The widths of the bands in the Sankey Diagram are proportional to the quantity of the flow. Each flow typically has a specific value associated with it, and the varying thickness of the bands helps visualize the relative magnitudes of these values. For example, based on the purchase data of Kettle, it might be a profitable option to put products of Smiths, Doritos, and Pringles into a bundle

Causal Analysis

The CausalImpact package in R was utilized to analyze the impact of the store layout trial. The graphs and tables below demonstrate the causal impact of the trial on the monthly total sales of store 86 and store 88. To account for any potential confounding factors, the sales data of stores that had a correlation higher than 0.8 with the trial stores during the pre-trial period were used as control variables.

The results show that the new store layout has a positive impact on the monthly total sales in both stores, with relative effects of 12% and 6.3% respectively. The probability of obtaining this effect by chance is very low (Bayesian one-sided tail-area probability p = 0.001). Therefore, the causal effect can be considered statistically significant.

- Sales of Store 86- controlling store 204, store 161

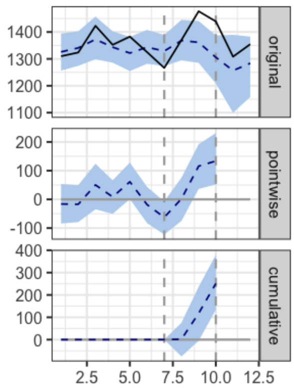


Posterior inference {CausalImpact}

	Average	Cumulative
Actual	929	2788
Prediction (s.d.)	833 (29)	2498 (88)
95% CI	[774, 890]	[2321, 2670]
Absolute effect (s.d.)	97 (29)	290 (88)
95% CI	[39, 156]	[118, 467]
Relative effect (s.d.)	12% (4%)	12% (4%)
95% CI	[4.4%, 20%]	[4.4%, 20%]

Posterior tail-area probability p: 0.001
Posterior prob. of a causal effect: 99.8996%

- Sales of store 88, controlling store 2, store 95, store 224



Posterior inference {CausalImpact}

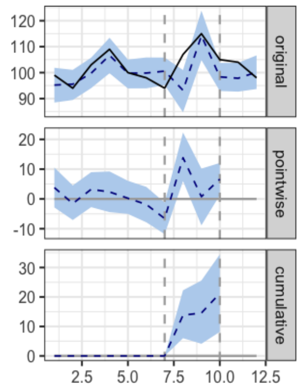
	Average	Cumulative
Actual	1429	4287
Prediction (s.d.)	1345 (21)	4036 (62)
95% CI	[1303, 1384]	[3908, 4153]
Absolute effect (s.d.)	84 (21)	251 (62)
95% CI	[45, 126]	[134, 379]
Relative effect (s.d.)	6.3% (1.6%)	6.3% (1.6%)
95% CI	[3.2%, 9.7%]	[3.2%, 9.7%]

Posterior tail-area probability p: 0.001
Posterior prob. of a causal effect: 99.9%

The graphs and tables below demonstrate the causal impact of the trial on the monthly number of customers at store 86 and store 88. To account for any potential confounding factors, the number of customers data from stores that had a correlation higher than 0.8 with the trial stores during the pre-trial period were used as control variables.

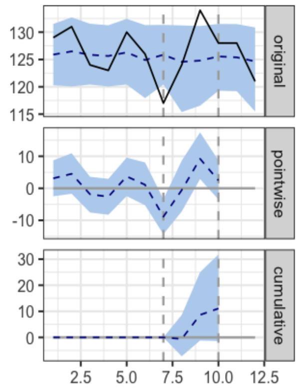
The results indicate that the new store layout has a positive impact on the monthly total sales at both stores, with relative effects of 7% and 3% respectively. The probability of obtaining this effect by chance is very low (Bayesian one-sided tail-area probability p = 0.001 and 0.049). Therefore, the causal effect can be considered statistically significant.

- Numbers of customers of store 86, controlling store 125, store 161




```
Posterior inference {CausalImpact}

                         Average        Cumulative
Actual                   109            327
Prediction (s.d.)        102 (2.2)      306 (6.6)
95% CI                   [98, 106]      [293, 319]

Absolute effect (s.d.)   7.1 (2.2)      21.2 (6.6)
95% CI                   [2.7, 11]      [8.2, 34]

Relative effect (s.d.)   7% (2.3%)      7% (2.3%)
95% CI                   [2.6%, 12%]    [2.6%, 12%]

Posterior tail-area probability p:   0.001
Posterior prob. of a causal effect:  99.8996%
```

- Numbers of customers of store 88, controlling store 200, store 182, store 185, store 142.




```
Posterior inference {CausalImpact}

                         Average        Cumulative
Actual                   129            386
Prediction (s.d.)        125 (2.6)      375 (7.9)
95% CI                   [118, 129]     [354, 387]

Absolute effect (s.d.)   3.7 (2.6)      11.1 (7.9)
95% CI                   [-0.46, 11]    [-1.39, 32]

Relative effect (s.d.)   3% (2.2%)      3% (2.2%)
95% CI                   [-0.36%, 9%]   [-0.36%, 9%]

Posterior tail-area probability p:   0.049
Posterior prob. of a causal effect:  95.1%
```

However, for trial store 77, the new layout does not have a statistically significant impact on the monthly total sales and number of customers. It may be necessary to examine the store more closely to determine the reasons.

Based on the results obtained from store 86 and store 88, the new store layout has proven to be effective in boosting sales and attracting new customers. The clients should consider implementing the new layout in other similar stores. However, it is important to note that the impact of the new store layout ranges from 3% to 12%, which is relatively moderate. Therefore, the clients should also consider the cost of the trial to determine if it is worth the investment.

Recommendation

Based on the findings above, four strategic recommendations are being put forward to drive sales performance of the chips manufacturing company;

1. **Focus on Improving Frequency**

As identified in the Customer Overview section, the fluctuation in the average sales/customer is significantly impacted by differences in the number of transactions/customers rather than the monetary value/transaction. This might be caused by chips being a highly saturated product category with a lot of variety of options available. This makes pricing within this category to be limited in variation, resulting in a narrow range of prices. Hence, incentivizing more frequent purchases is advisable to motivate customers in increasing repeat purchase for the same brand which could drive overall revenue and promote brand loyalty at the same time.

2. **Focus on Family**

As highlighted in the Market Dynamics section, it was identified that families tend to consume more chips due to their larger size. The team has also taken into account the potential influx of new and young families within the customer demographic as the existing ones age. To address this consideration, the team has proposed implementing family-themed campaigns, drawing inspiration from successful examples such as Disney's "Share Your Ears" Campaign, Coca-Cola's "Share a Coke" Campaign, and IKEA's "Where Life Happens" Campaign.

3. **Consumer-centric Promo**

Based on the findings on New Segmentation analysis, there are 2 different marketing initiatives tailored to each of the newly identified segment based on their preference and behavior:

- Loyalists

With consideration to Recommendation 1 & 2, it is advisable for the client to initiate a family-themed campaign that motivates family segments to make more frequent purchases. This could take form in limited lucky pack promotion that gives customers the opportunity to win trips to Disneyland, which highly appeals to family segments and incentivizes higher consumption and leads to increased number of transactions.

- Explorers

With reference to the Sankey diagram under the New Segmentation analysis, it is recommended for the client to introduce a mix-pack bundle. There is a big portion of chip customers that like to explore different flavors in small pack size options. Hence, having a bundled option would give them greater convenience, and using customer historical purchase data to define the flavor mix would ensure relevance to customer preference.

4. **Implement new store layout changes**

With reference to the findings from the Causal Analysis section, the new store layout has revealed a positive and statistically significant impact on monthly total sales, with relative effects of 12% and 6.3% in store 86 and store 88 respectively. Therefore, it is recommended that the clients proceed with the implementation of the new store layout in similar stores, considering the proven positive impact and potential for enhanced business outcomes.

References

Tensley, B. (2022, January 1). How the potato chip took over America. Smithsonian.com. https://www.smithsonianmag.com/arts-culture/curious-history-potato-chip-180979232/#:~:text=Americans%20consume%20about%201.85%20billion,around%206.6%20pounds%20per%20person.