

QVI Chips Analysis Report

Geraldine Nyika

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE) library(tidyverse) library(readxl)  
library(janitor) library(ggplot2) library(stringr)
```

Executive Summary

Using Quantum's transaction and customer datasets, this report offers a data-driven analysis of chip buying behavior. The objective is to identify high-value customer segments, understand purchasing patterns, and recommend strategic actions for the upcoming category review.

Key Findings

- Older Singles/Couples and Retirees generate the highest total sales.
- Mainstream shoppers contribute the most revenue compared with Budget or Premium customers.
- Pack sizes between **150–175g** dominate sales across all segments.
- Top-performing brands include **Kettle, Doritos, and Smith's**.

Strategic Recommendations

- Prioritise older mainstream households for targeted promotions.
- Focus stocking and promotions on **mid-size pack formats (150–175g)**.
- Maintain strong visibility for Kettle, Doritos, and Smith's.
- Develop targeted offers for Premium shoppers who under-index in chip purchases.

1. Data Importing and Cleaning

```
## {r} transactions <- read_excel("Data/QVI_transaction_data.xlsx") %>% clean_names()  
customers <- read_csv("Data/QVI_purchase_behaviour.csv") %>% clean_names()
```

```
## Clean Transaction Data
```

```
```{r}
transactions <- transactions %>%
 mutate(
 date = as.Date(date, origin = "1899-12-30"),
 prod_name = str_squish(prod_name),
 pack_size = as.numeric(str_extract(prod_name, "\d+")),
 brand = word(prod_name, 1)
)

```

## Standardise Brand Names

```
{r} transactions <- transactions %>% mutate(brand_clean = case_when(
 %in% c("Smiths", "Smith") ~ "Smith's",
 brand %in% c("Doritos", "Dorito") ~ "Doritos",
 brand == "Kettle" ~ "Kettle",
 brand == "Pringles" ~ "Pringles",
 brand == "Thins" ~ "Thins",
 brand == "Cobs" ~ "Cobs",
 brand == "Twisties" ~ "Twisties",
 brand == "Cheezels" ~ "Cheezels",
 brand == "Cheetos" ~ "Cheetos",
 brand == "CCs" ~ "CC's",
 brand %in% c("Natural", "NCC") ~ "Natural Chip Co",
 brand %in% c("Grain", "GrnWves") ~ "Grain Waves",
 brand %in% c("Sunbites", "Snbts") ~ "Sunbites",
 brand %in% c("Red", "RRD") ~ "Red Rock Deli",
 brand == "Old" ~ "Old El Paso",
 brand %in% c("WW", "Woolworths") ~ "Woolworths",
 brand == "French" ~ "French Fries",
 brand == "Tostitos" ~ "Tostitos",
 TRUE ~ brand
))
```

## Filter Outliers

---

```
{r} transactions <- transactions %>% filter(pack_size > 20 & pack_size < 500,
 prod_qty > 0 & prod_qty < 10)
```

---

## 2. Merge Data and Create Features

---

```
{r} merged <- transactions %>% inner_join(customers, by = "lylty_card_nbr") %>% mutate(unit_price =
 tot_sales / prod_qty)
```

---

## 3. Exploratory Analysis

### Sales by Lifestage

```
```{r} sales_lifestage <- merged %>% group_by(lifestage) %>% summarise(total_sales = sum(tot_sales),
  .groups = "drop")
sales_lifestage
```

```

### Plot - Lifestage Sales

```{r}
ggplot(sales_lifestage, aes(x = reorder(lifestage, total_sales), y = total_sales)) +
 geom_col(fill = "steelblue") +
 coord_flip() +
 labs(title = "Total Sales by Lifestage", x = "Lifestage", y = "Sales") +
 theme_minimal()

```

## Sales by Premium Customer Segment

```

```{r} sales_premium <- merged %>% group_by(premium_customer) %>% summarise(total_sales =
sum(tot_sales), .groups = "drop")
sales_premium

```

Plot - Premium Segment Sales

```

```{r}
ggplot(sales_premium, aes(x = premium_customer, y = total_sales)) +
 geom_col(fill = "darkseagreen3") +
 theme_minimal() +
 labs(title = "Total Sales by Premium Customer Type", x = "Customer Type", y = "Sales")

```

---

## 4. Segment Interaction (Lifestage × Premium)

```

```{r} segment_summary <- merged %>% group_by(lifestage, premium_customer) %>% summarise( total_sales = sum(tot_sales), avg_price = mean(unit_price), qty = sum(prod_qty), transactions = n(), .groups = "drop" )
segment_summary

```

Heatmap - Segment Performance

```

```{r}
ggplot(segment_summary, aes(x = premium_customer, y = lifestage, fill = total_sales)) +
 geom_tile(color = "white") +
 scale_fill_gradient(low = "lightblue", high = "darkblue") +
 labs(title = "Heatmap: Total Sales by Customer Segment",
 x = "Premium Customer Type", y = "Lifestage") +
 theme_minimal()

```

---

## 5. Pack Size Analysis

```
``{r} packsize_summary <- merged %>% group_by(pack_size) %>% summarise(total_sales = sum(tot_sales), avg_price = mean(unit_price), qty = sum(prod_qty), transactions = n(), .groups = "drop")
```

```
packsize_summary
```

```
Plot - Pack Size
```

```
```{r}
ggplot(packsize_summary, aes(x = factor(pack_size), y = total_sales)) +
  geom_col(fill = "darkolivegreen3") +
  coord_flip() +
  theme_minimal() +
  labs(title = "Total Sales by Pack Size", x = "Pack Size (g)", y = "Total Sales")
```

6. Brand Performance

```
``{r} brand_summary <- merged %>% group_by(brand_clean) %>% summarise( total_sales = sum(tot_sales), avg_price = mean(unit_price), qty = sum(prod_qty), transactions = n(), .groups = "drop" )
```

```
brand_summary
```

```
### Plot - Brand Sales
```

```
```{r}
ggplot(brand_summary, aes(x = reorder(brand_clean, total_sales), y = total_sales)) +
 geom_col(fill = "steelblue") +
 theme_minimal() +
 labs(title = "Total Sales by Brand", x = "Brand", y = "Total Sales") +
 theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

---

## 7. Export Outputs

```
{r} write_csv(sales_lifestage, "outputs/summary_tables/sales_lifestage.csv") write_csv(sales_premium,
"outputs/summary_tables/sales_premium.csv") write_csv(segment_summary, "outputs/summary_tables/segment",
write_csv(packsize_summary, "outputs/summary_tables/packsize_summary.csv") write_csv(brand_summary,
"outputs/summary_tables/brand_summary.csv")
```

---

## **Conclusion**

Pack size preferences, brand performance, and segment behaviors are all clearly understood thanks to this analysis. The insights support targeted strategic opportunities for retail optimisation and customer engagement within the chips category.