

# Data Preparation and Customer Analytics

For Chips category

Nhi (Chelsea) Le

Retail Analytics Team  
Quantium

May 2020

# Let's have a look at the clean data after tidying, wrangling and joining two data sets (Transaction and Purchase Behaviour) together

```
## Rows: 264,839
## Columns: 14
## $ loyalty_card_number <dbl> 1000, 1002, 1003, 1003, 1004, 1005, 1007, 1007,...
## $ full_date           <date> 2018-10-17, 2018-09-16, 2019-03-07, 2019-03-08...
## $ store_number        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ transaction_id      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, ...
## $ product_number      <dbl> 5, 58, 52, 106, 96, 86, 49, 10, 20, 51, 59, 84,...
## $ product_quantity    <dbl> 2, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 2, 1, 1, 1,...
## $ total_sales          <dbl> 6.0, 2.7, 3.6, 3.0, 1.9, 2.8, 3.8, 2.7, 5.7, 8....
## $ premium_customer    <chr> "Premium", "Mainstream", "Budget", "Budget", "M...
## $ brand_name           <chr> "Natural Chip          Compny SeaSalt", "Red Rock...
## $ lifestage            <chr> "young singles/couples", "young singles/couples...
## $ brand                <chr> "natural", "rrd", "grnwves", "natural", "woolwo...
## $ date                 <dbl> 43390, 43359, 43531, 43532, 43406, 43462, 43438...
## $ product_name         <chr> "Natural Chip          Compny SeaSalt", "Red Rock...
## $ package_size         <dbl> 175, 150, 210, 175, 160, 165, 110, 150, 330, 17...
```

# Check for missing values in the data

- There is a **very small, almost negligible, number of missing values in almost all of the variables** in our data
- Removing these missing values, therefore, would not impact our analysis to a significant extent.
- No imputation of missing values, hence, is needed.

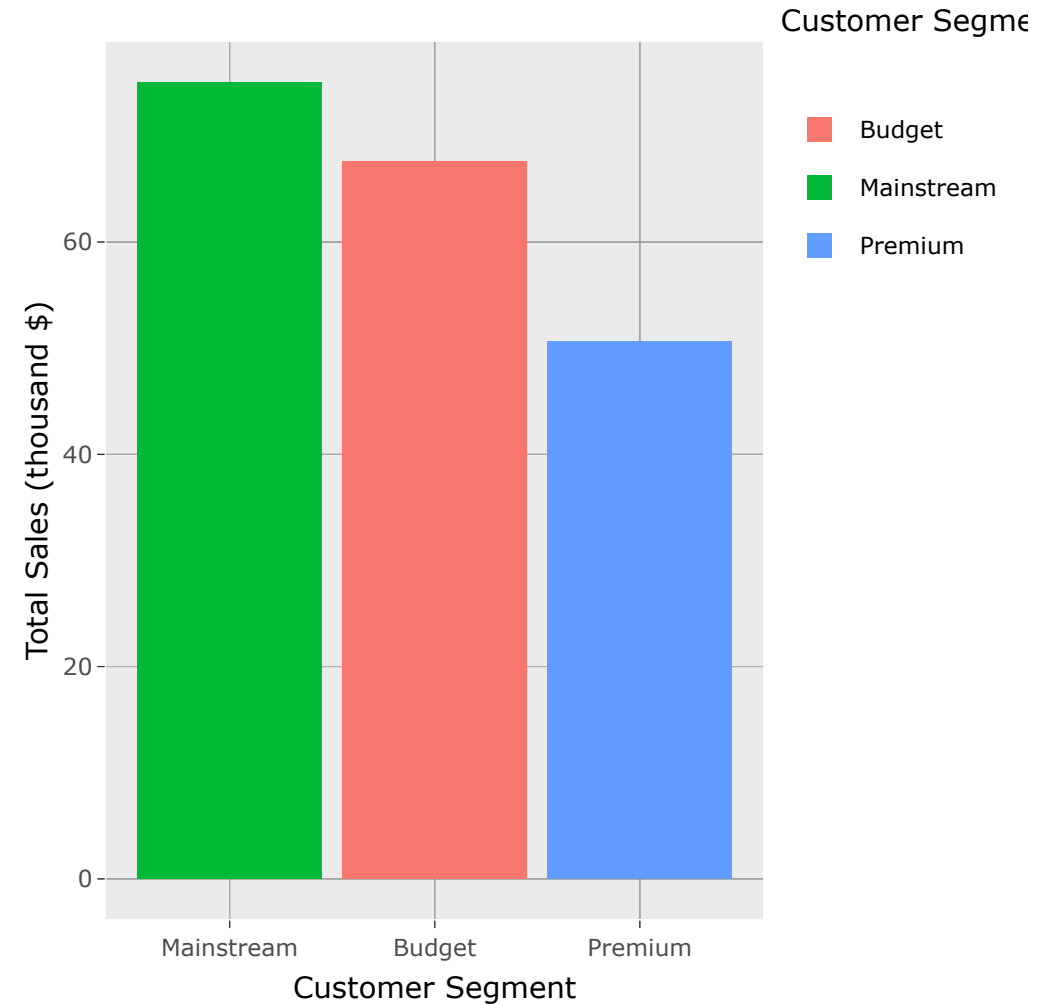
variable	n_miss	pct_miss
full_date	3	0.00113276367 906539
brand_name	3	0.00113276367 906539
brand	3	0.00113276367 906539
package_size	3	0.00113276367 906539
premium_customer	2	0.00075517578 6043596
lifestage	2	0.00075517578 6043596
store_number	1	0.00037758789 3021798
transaction_id	1	0.00037758789 3021798

# The metrics in the data:

- Loyalty card number
- Date of transaction
- Store number
- Transaction ID
- Product number
- Product quantity
- Total sales (per transaction)
- Brand name and product name
- Lifestage of buyers
- Package size

# Total sales by customer segments

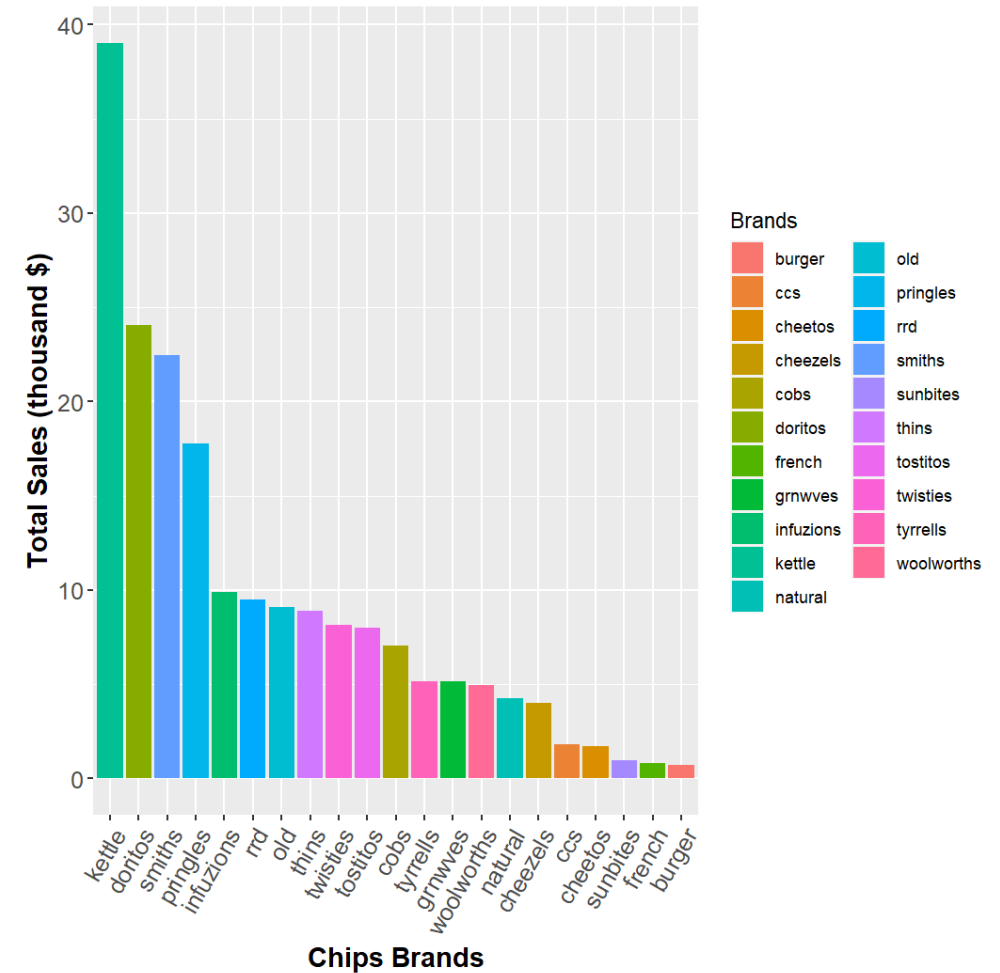
- **Mainstream customers** are the highest sales-generating buyers for the chips category, generating more than \$75,000 in sales for the whole period.



# Total sales by brand

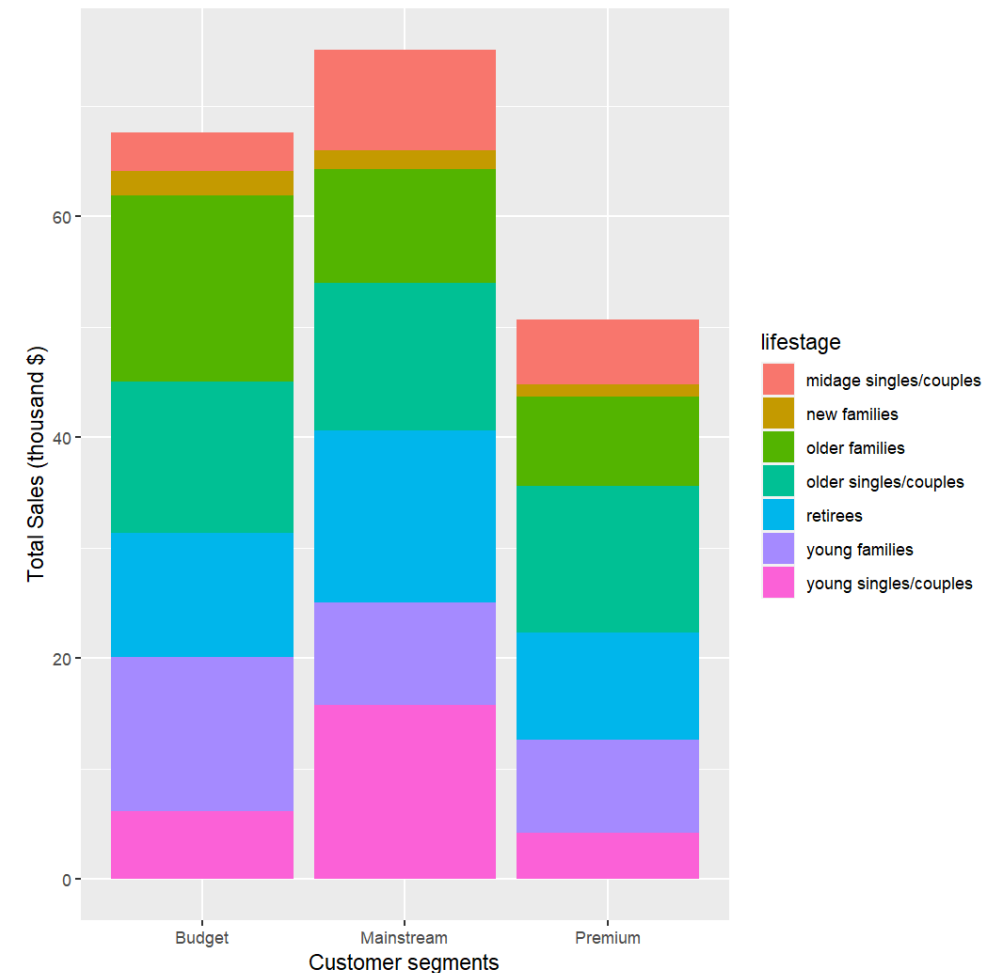
## Top 8 most popular chip brands are:

1. Kettle
2. Doritos
3. Smiths
4. Pringles
5. Infuzions
6. RRD SR
7. Old El Paso
8. Thins Chips



# Total sales by customer segments

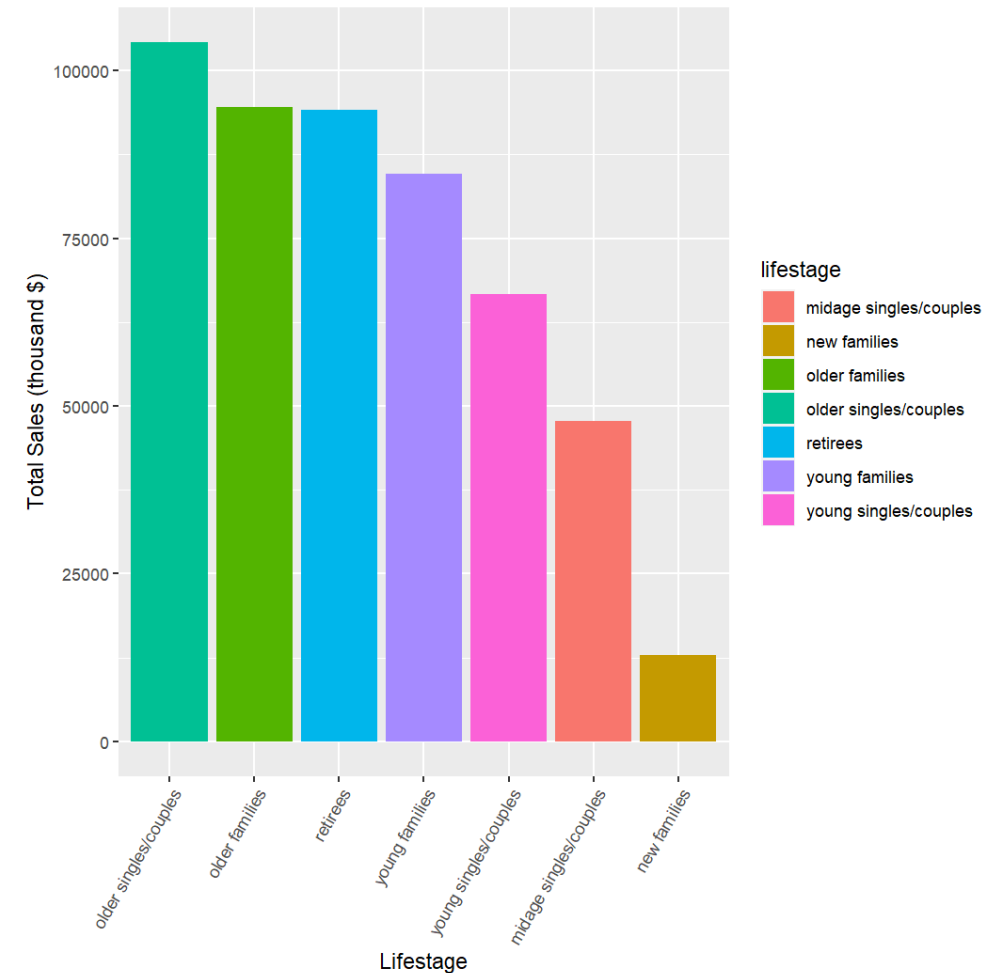
- **Mainstream buyers** contribute the most to sales across all chips products
- Second most are **budget buyers**
- Among mainstream buyers, the majority of buyers are **retirees** and **young singles/couples**
- **Young families, older families, older singles/couples** make up the majority of budget buyers
- Most premium buyers are those from **older families**



# Total sales by lifestage

## Top 4 types of customers for chips:

1. older singles/couples
2. older families
3. retirees
4. young families





# Total sales by package size:

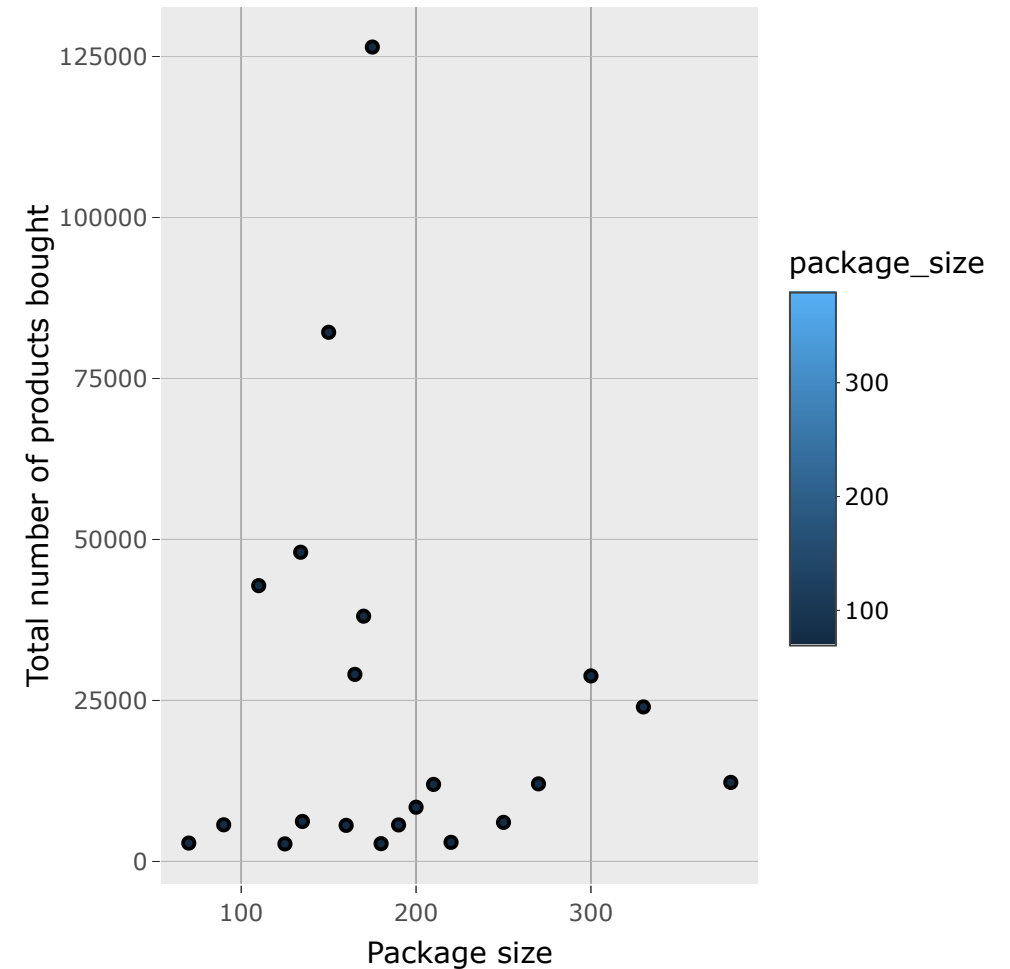
Package sizes of

- 175g
- 150g

generate the most sales amongst all chips products

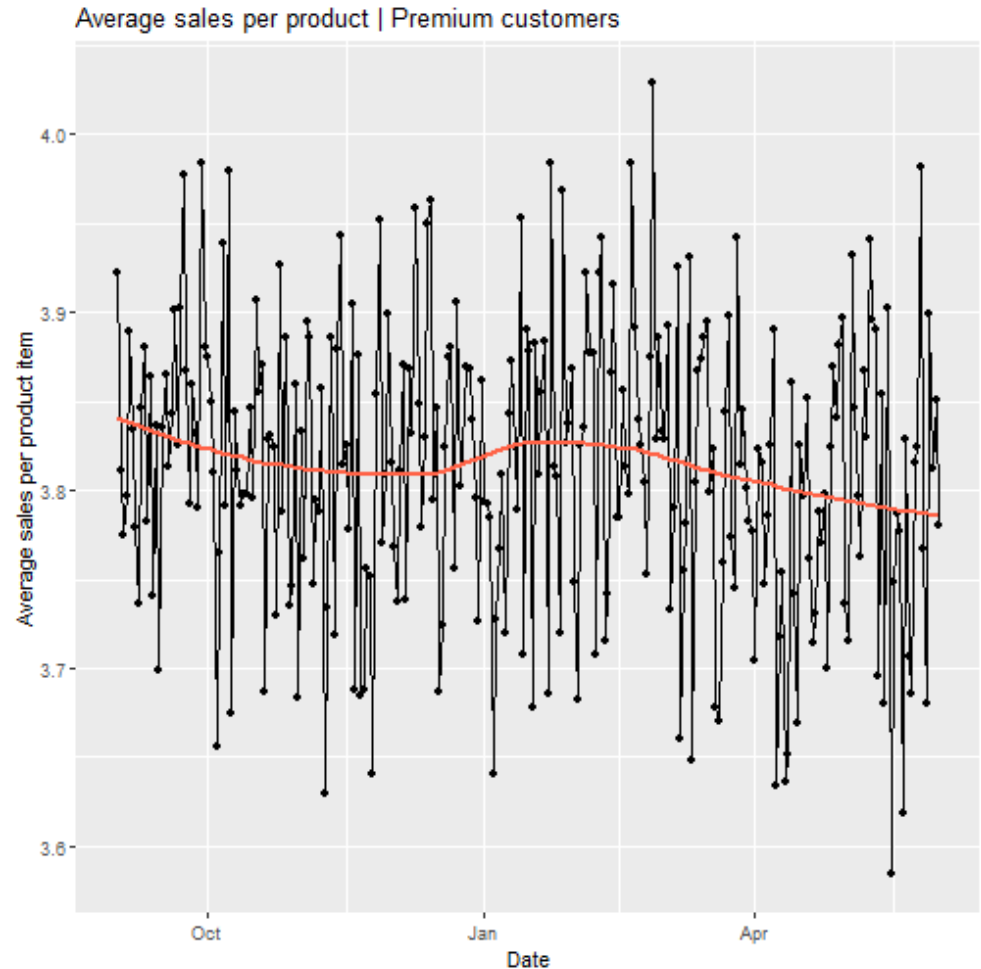
Other high sales-generating options are:

- 110g
- 134g
- 165g
- 170g



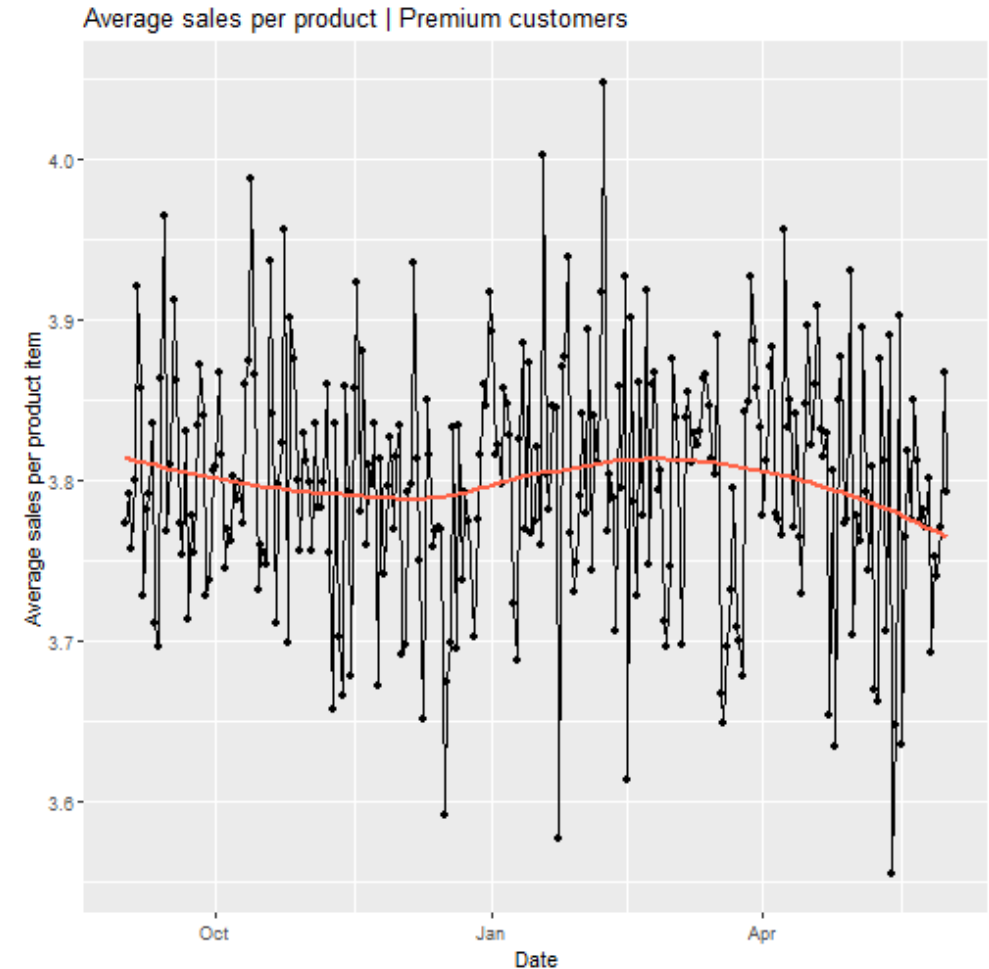
# Looking at Premium Customers:

- Despite experiencing lots of fluctuations, average sales per product item for premium customers remained quite stable over the last 9 months.
- This suggests that there hasn't been a significant increase in sales contributed by premium customers.



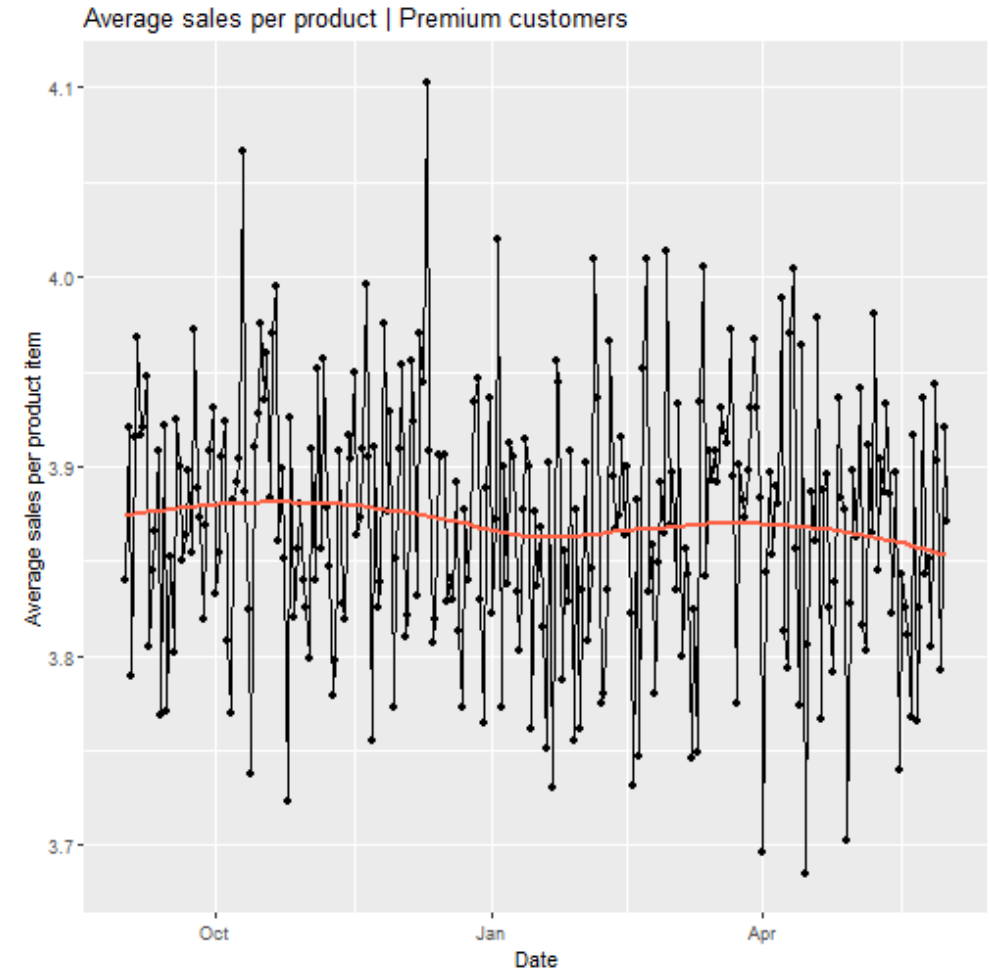
# Looking at Budget Customers:

- Average sales per product item for budget customers decreased slightly towards the end of the period.
- This suggests that maybe demand for this class of customers is likely to further decrease in upcoming months.



# Looking at Mainstream Customers:

- Average sales per product item for budget customers remained quite constant throughout the last 9 months.
- This suggests that maybe demand for this class of customers is likely to remain at this level for upcoming months.



# Multi-pack buying behaviours

- On average, a customer tends to buy 2 packs of chips, regardless of what segment they come from.
- This suggests that maybe a customer's shopping behaviour isn't really related to the number of chips products they purchase.

premium_cus	quantity	customers	s_per_person
Budget	177898	93157	1.90965788936956
Mainstream	193965	101988	1.90184139310507
Premium	132865	69691	1.90648720781737

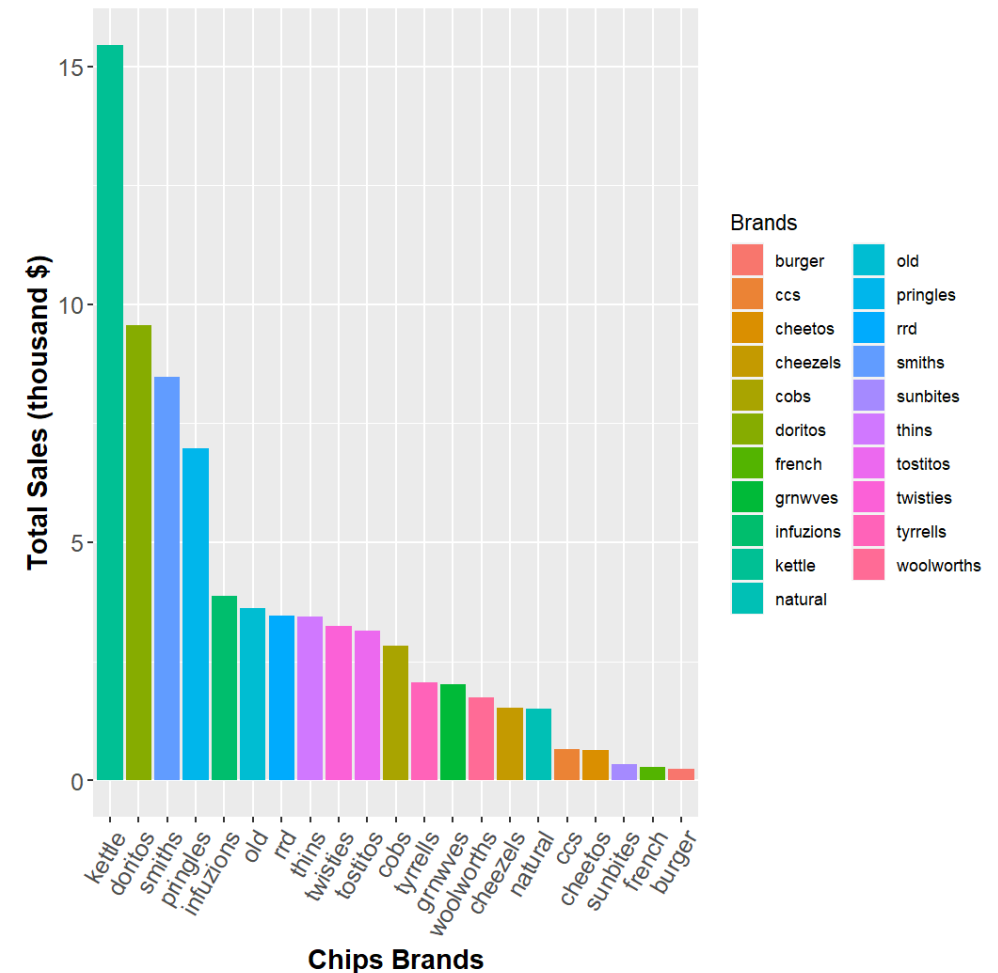
# Most popular products among Mainstream Customers:

- Since mainstream customers are the highest sales-generating customer class, looking at their common chips options might be worthwhile.

## Top 4 most popular chips brands amongst mainstream customers are:

- Kettle
- Doritos
- Smiths
- Pringles

which are also the most popular brand names amongst all customers.



# Recommendations:

- Target at Mainstream customers because they generate the most sales for chips.
- Budget customers are also a high sales-generating class of buyers.
- Ensure stocks are adequate for popular brands of chips like Kettle, Doritos, Smiths, Pringles, Infuzions.
- Stock more of 175g and 150g packages than of smaller or larger sizes.
- Quantity of products to stock should be twice that of customers since most are multi-pack buyers.