Forage

quantium

Chip Analysis

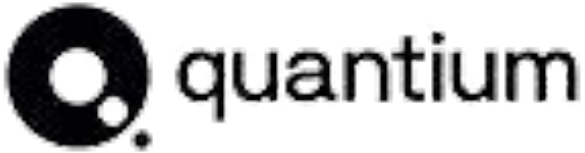
Understand customer purchasing behaviour on Chips Sales

Imam Ahmad Qusyairi - Retail Analytics Team

Virtual Internship Program

About virtual internship program

- Currently i'm participating in **Quantum Data Analytics** Virtual Experience Program by *theForage.com*
- I've given task that designed to replicate life in the **Retail Analytics and Strategy team** at Quantium



Background

Story behind this project

- As **Category Manager** , Our client wants to better **understand** the types of customers who **purchase Chips** and their purchasing behaviour within the region.
- The insights from your analysis will feed into the **supermarket's strategic plan** for the chip category in the next half year.

OUR TASKS

Job we need to get done !

1. Clean Data
2. Create & Interpret Data
3. Exploratory Data Analysis
4. Provide Clear Recommendation

Table of Contents

Contents of this project

01

Data Check

Check data quality

Data Overview

Things We Analyze

We have 2 dataset files used in this project :

Purchasing Behaviour Data

Customer Personal information including membership, lifestage, etc.

Transaction Data

Detailed record of transaction data including product name, quantity, etc.

Quality Check

Make sure the data was ready

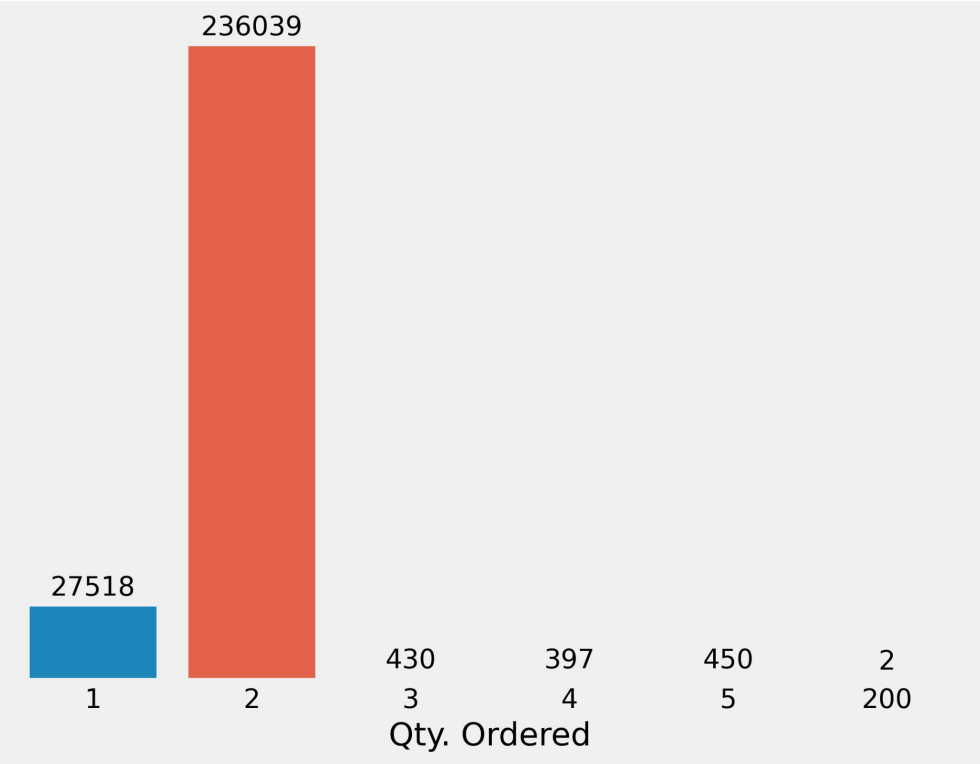
	Purchasing Data	Transaction Data
Missing Values	No	No
Duplicated Values	No	No
Outliers	No	2 Instances

The data was clean except **2 outliers** from **Trans. data**

The Outliers

Verify & Clean the Instances

- There are 2 customer that order **200 chips !**
- It's possible that they were **distributor**, not a customer
- We will **exclude** them from analysis



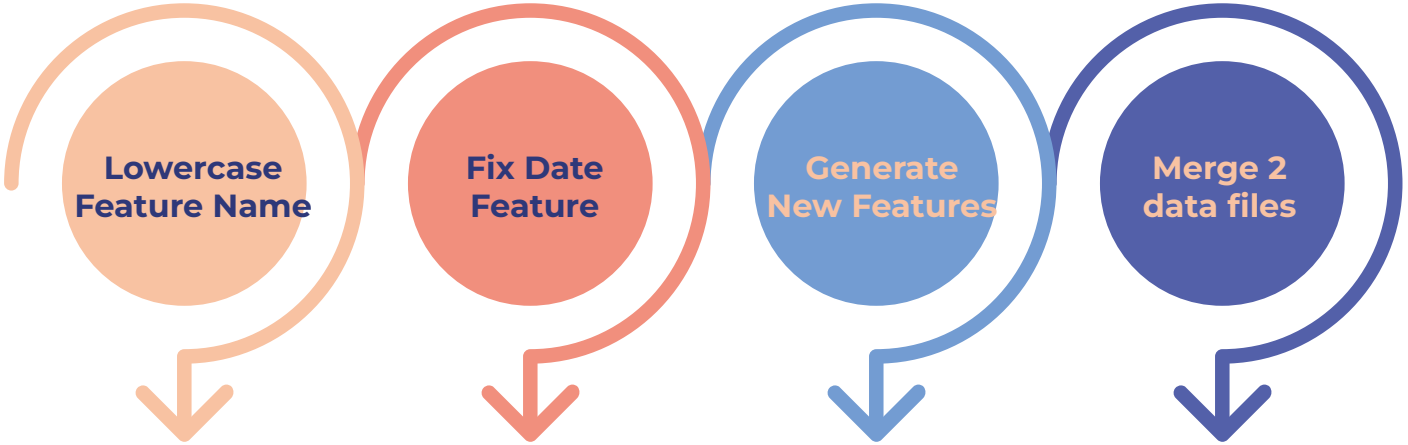
02

Feature Engineering

Modify existing & create new features

Feature Engineering

What We Will Do :



Lowercase all feature names so it's easier to type

Convert them into Y-M-D format

Generate chips brand & chips size features

Inner join the 2 data

Feature Engineering

Lowercase Feature Name

Make feature names easier to typed

Before

	DATE	STORE_NBR	LYLTY_CARD_NBR	TXN_ID	PROD_NBR
0	43390	1	1000	1	5
1	43599	1	1307	348	66
2	43605	1	1343	383	61
3	43329	2	2373	974	69
4	43330	2	2426	1038	108

After

	date	store_nbr	lylty_card_nbr	txn_id	prod_nbr
0	2018-10-17	1	1000	1	5
1	2019-05-14	1	1307	348	66
2	2019-05-20	1	1343	383	61
3	2018-08-17	2	2373	974	69
4	2018-08-18	2	2426	1038	108

After lowercase, the feature names will be easier to typed

Feature Engineering

Fix Date Feature

Revise date into correct format date

Before

```
0    43390
1    43599
2    43605
3    43329
4    43330
5    43604
6    43601
7    43601
8    43332
9    43330
Name: date, dtype: int64
```

After

```
0    2018-10-17
1    2019-05-14
2    2019-05-20
3    2018-08-17
4    2018-08-18
5    2019-05-19
6    2019-05-16
7    2019-05-16
8    2018-08-20
9    2018-08-18
Name: date, dtype: object
```

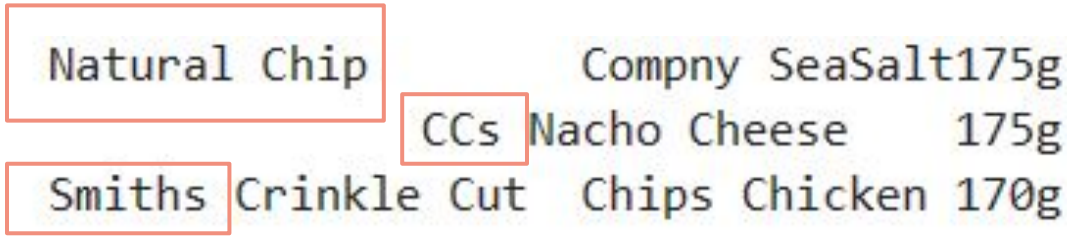
- Previously, date feature was integer value of day after **'date origins'**
- We revise them into correct date by **counting days** from the integer value
- **Date Origins** : 30-12-1899

Feature Engineering

Generate New Feature : Brand

Create New Feature from Existing Data

Product Name Feature Examples



- We found that for **every first word** from each instances of **product name feature**, indicate a **brand name**
- We will **extract** brand name form product name feature by **splitting string** values by **whitespace** and take the **first splitting** into **brand name**
- We also **fix** several **mispelled** brand name so it shows only **one distinct** brand

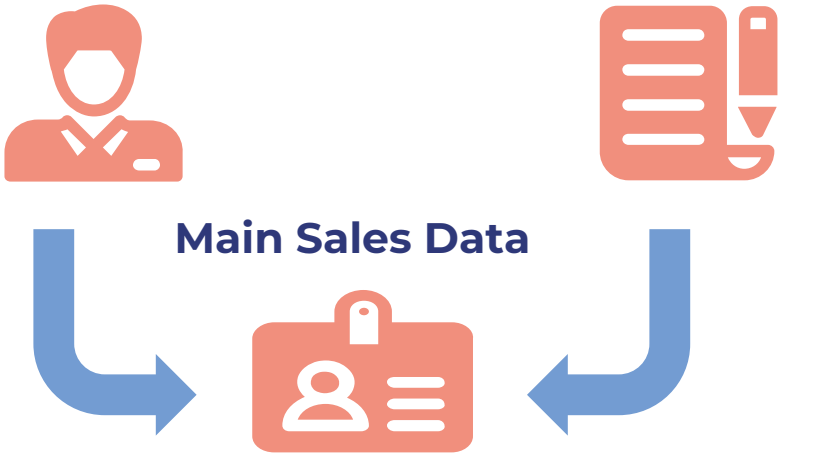
Feature Engineering

Merge 2 Files Data

Unite 2 different files into 1 main dataset

Purchasing Behaviour Data

Transaction Data



03

Exploratory Data Analysis

Deep-dive analysis toward sales data

Exploratory Data Analysis

Sales Performance

Descriptively Analyze Sales Performance Data

	Total	Mean	Median	Min	Max
Product (Qty)	504,724	1.905	2.0	1.0	5.0
Revenue ($)	1,933,114	7.299	7.4	1.5	29.5

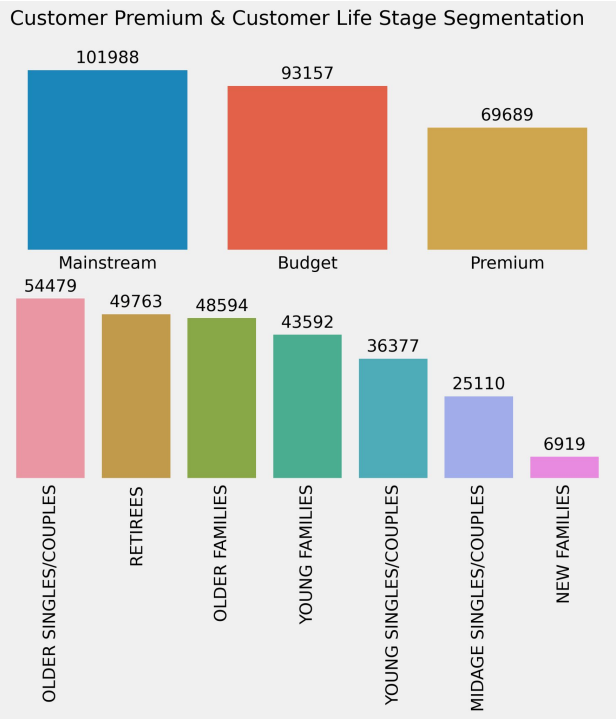
- Sales performance in 1 Year (July 2018 - July 2019) reach total product **sales 504,724 units** and **total revenue $1,933,114**
- Product sales has **mean values of 1.905**. Meaning for every chips purchasing, customer buy at least **2 chips**
- Revenue has **mean values of $7.299**. Meaning for every chips purchase, **$7.3** dollars collected

Exploratory Data Analysis

Customer Segmentation

Customer Group based on Membership & Life Stage

- **Most customer** who buy chips was **not a member** / regular customer. Where Customer who buy the **premium** only **reach ~30%** of total customer who buy chips
- **Older singles/couples** founded as a most people **who buy chips**. Young customer founded **relatively low** compared to elderly for buying chips

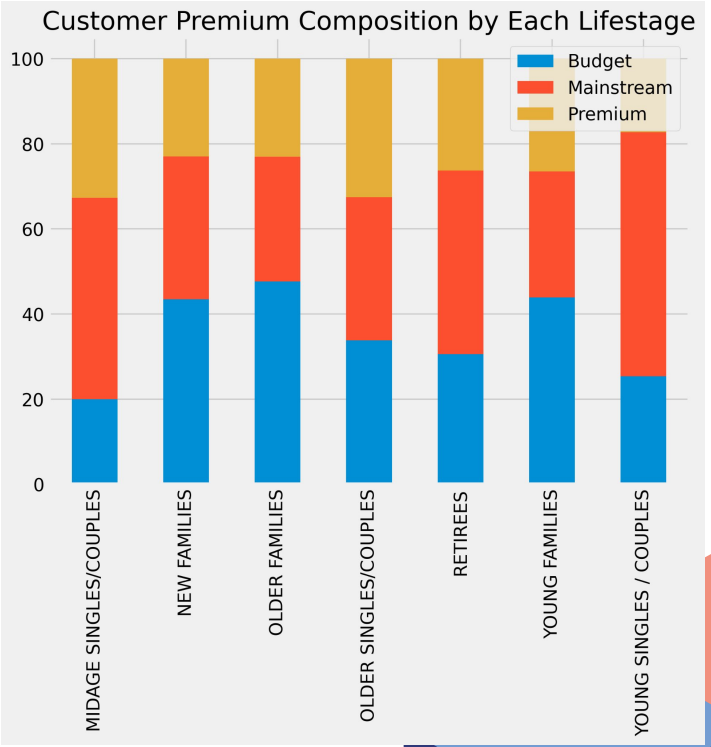


Exploratory Data Analysis

Membership Composition

Membership Proportion from Customer Life Stage

- **Middle age** customer tend to have **higher membership propotion** compared to **younger** and **elder** age customer
- **Young singles / couples** have **lowest membership** proportion. Should be the main concern if we want to target the younger customer market

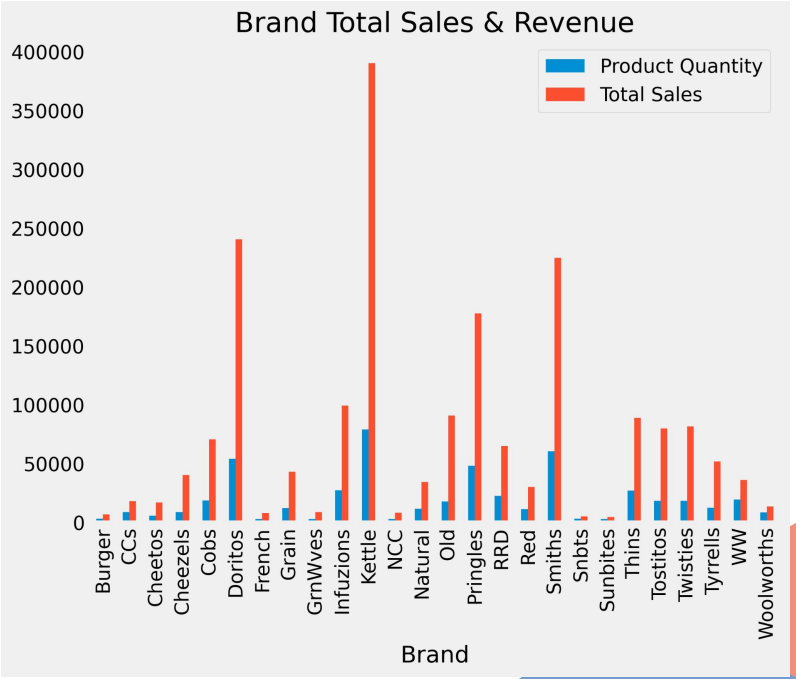


Exploratory Data Analysis

Brand Sales Performance

Quantity Product Sold and Revenue from Chip Brands

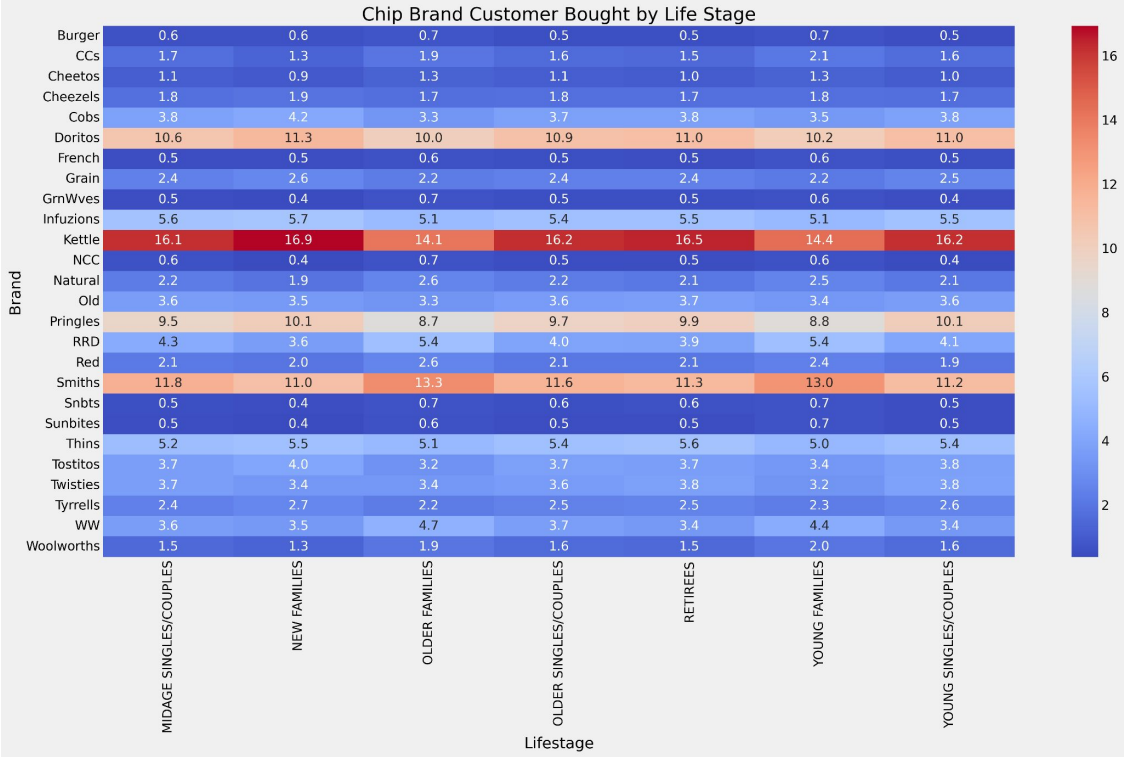
- We found **Kettle** as the most **success brand** in. They have highest quantity of product sold and highest revenue
- **Doritos** have higher revenue compared to **Smiths** although Doritos have **smaller** quantity of **product** saled to Smiths



Exploratory Data Analysis

Customer Preferred Chips Brands (i)

Chip Brand Bought by Customer per Life Stage

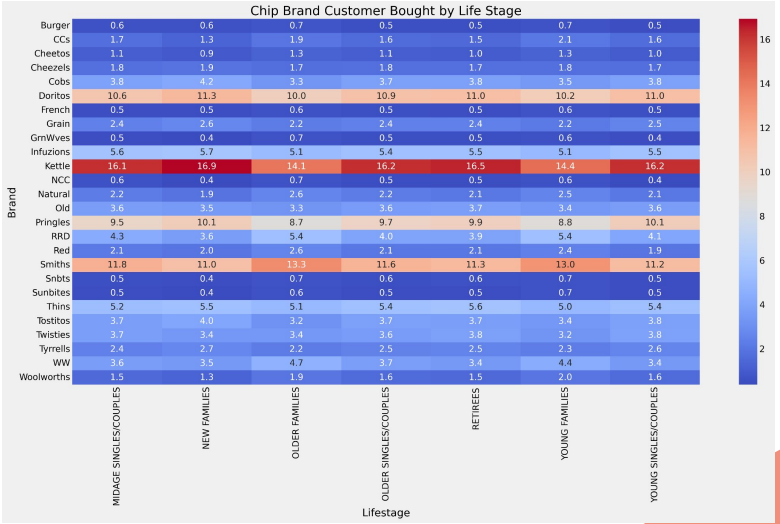


Exploratory Data Analysis

Customer Preferred Chips Brands (ii)

Chip Brand Bought by Customer per Life Stage

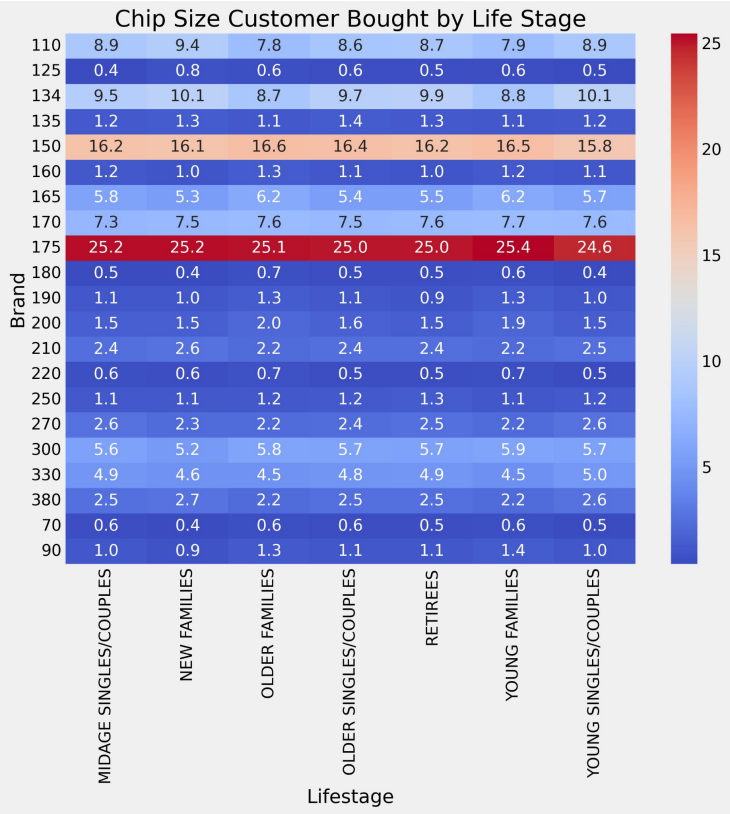
- **Consistent** with other customer **life stage**, **Kettle** became the most purchased brand compared
- We can **assume** that **between** different **customer lifestage**, they have relatively **similar** preferred brands



Exploratory Data Analysis

Customer Preferred Chips Size (i)

Chip Packaging Size Bought by Customer per Life Stage

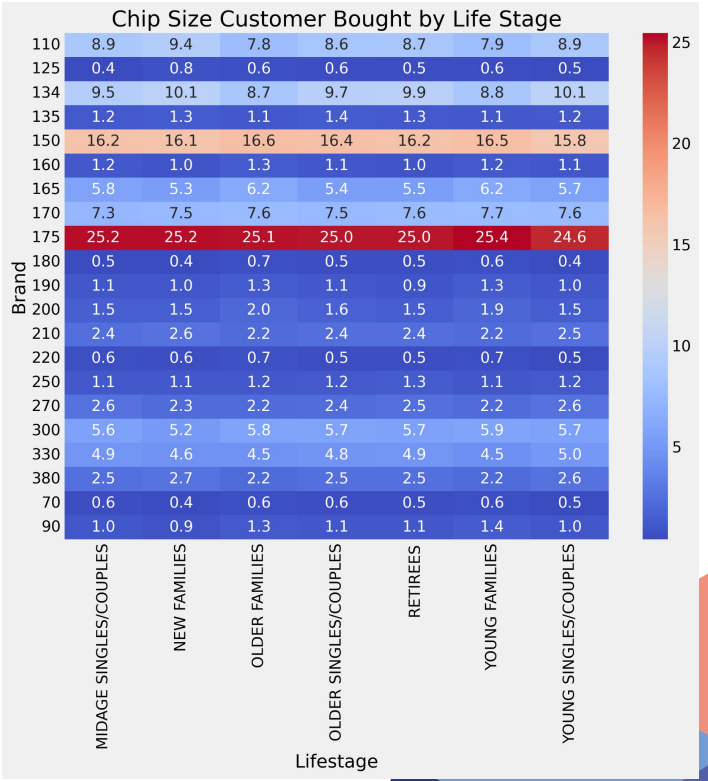


Exploratory Data Analysis

Customer Preferred Chips Size (ii)

Chip Packaging Size Bought by Customer per Life Stage

- **Chips** packed in **175 g** founded **most purchased** chips pack size between different customer lifestage
- Similar with previous findings, there is relatively **similar preferences** for **chips** packaging **size** between different customer lifestage

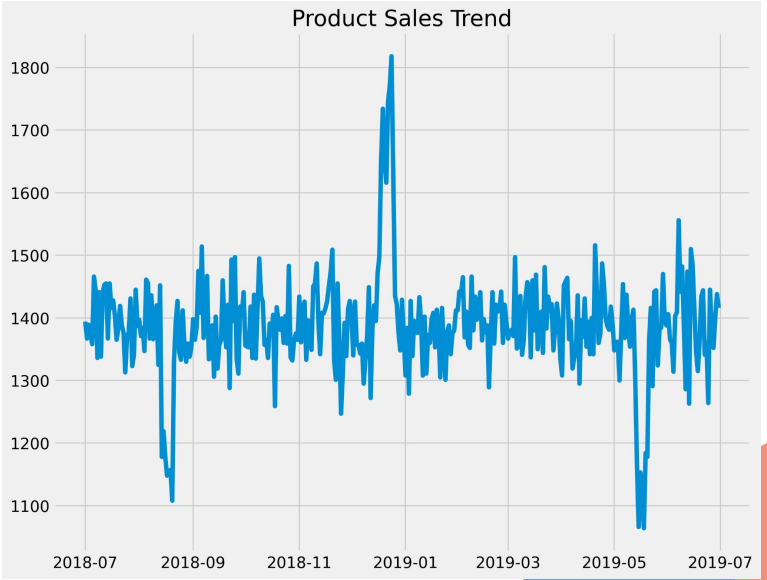


Exploratory Data Analysis

Product Sales Trend

Chip Product Sales Trend Accross the Year

- There is **drastically increase** of in **December 2018** followed then in January 2018.
- Maybe caused by the **increasing demand** of chips for **Christmas**
- There is **drastically decrease** of quantity product **saled** in **June 2019**. We need to inspect why this happen in the future work



04

Conclusion & Recommendation

What This Project Tells Us

Conclusion & Recommendation

Conclusion

What we found from this project

- **Elderly** customer have **higher** number of **buying** chips compared to **younger** customer
- **Young** customer has **lower number** of buying **premium** membership compared to **elder customer**
- **Most sucess** chips **brands** in terms of quantity chips saled and revenue collected was **Kettle** brand
- There is **no significant difference** of chip **brands** and **chips** package size **between customer lifestage**

Conclusion & Recommendation

Recommendation

What We Could Improve Based on Data

- **Set marketing** srategy targetting on **Younger** customer. Since younger customer have huge potential and we should **exploit** them
- Perform **Collaboration** with Kettle brand to **attract existing** customer and **new** customer
- **Perform Clustering analysis** using unsupervised learning in the future

Thanks

Have Any Questions ? Contact me !

imam.aqusairi@gmail.com

Or

Click This Link To See the Code

Github Link

Tools I Used :

Jupyter Notebook

Virtual Environment for Python

Google Slides

Create & Modify Presentation Slides

Obsidian Notes

Note Taking System