

task

August 8, 2023

1 Import libraries and modules

```
[524]: # Import required libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import ttest_ind
```

Reading of data from files is done in the following way:

```
[525]: #We store data from files into variables
transaction_data = pd.read_csv('QVI_transaction_data.csv')
purchase_behaviour_data = pd.read_csv('QVI_purchase_behaviour.csv')
```

Date is not in proper format so we will format the date properly to make it accurate.

```
[526]: # Convert DATE column to datetime format
transaction_data['DATE'] = pd.to_datetime(transaction_data['DATE'],
    ↪origin='1899-12-30', unit='D')
print(transaction_data['DATE'].head())
```

```
0    2018-10-17
1    2019-05-14
2    2019-05-20
3    2018-08-17
4    2018-08-18
```

Name: DATE, dtype: datetime64[ns]

Now lets see if that worked

```
[527]: transaction_data.head(50)
```

```
[527]:
```

	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	
5	2019-05-19	4	4074	2982	57	

6	2019-05-16	4	4149	3333	16
7	2019-05-16	4	4196	3539	24
8	2018-08-20	5	5026	4525	42
9	2018-08-18	7	7150	6900	52
10	2019-05-17	7	7215	7176	16
11	2018-08-20	8	8294	8221	114
12	2019-05-18	9	9208	8634	15
13	2018-08-17	13	13213	12447	92
14	2019-05-15	19	19272	16686	44
15	2019-05-19	20	20164	17136	54
16	2018-08-18	20	20418	17413	94
17	2018-08-14	22	22411	18646	98
18	2018-08-17	22	22456	18696	93
19	2019-05-16	23	23067	19162	56
20	2019-05-19	25	25105	21815	7
21	2018-08-16	33	33081	29949	98
22	2018-08-16	36	36012	32077	31
23	2018-08-19	36	36302	33188	32
24	2018-08-15	38	38142	34181	108
25	2019-05-15	39	39144	35506	57
26	2018-08-19	39	39167	35638	111
27	2019-05-15	41	41423	38393	46
28	2018-08-15	41	41486	38472	13
29	2019-05-20	43	43110	39342	31
30	2019-05-16	43	43147	39608	99
31	2019-05-15	43	43227	40186	26
32	2019-05-20	45	45127	41122	64
33	2019-05-18	45	45220	41651	22
34	2018-08-16	51	51100	46802	48
35	2018-08-19	51	51100	46803	37
36	2018-08-18	51	51113	46828	36
37	2018-08-17	54	54226	48173	51
38	2018-08-18	54	54305	48301	44
39	2018-08-18	55	55072	48878	107
40	2019-05-14	55	55072	48883	106
41	2019-05-20	55	55073	48887	4
42	2019-05-20	55	55073	48887	113
43	2019-05-16	55	55202	49690	45
44	2018-08-18	56	56013	50090	39
45	2019-05-16	58	58324	54252	102
46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3

		PROD_NAME	PROD_QTY	TOT_SALES
0	Natural Chip	Compny SeaSalt175g	2	6.00

1	CCs Nacho Cheese	175g	3	6.30
2	Smiths Crinkle Cut Chips Chicken	170g	2	2.90
3	Smiths Chip Thinly S/Cream&Onion	175g	5	15.00
4	Kettle Tortilla ChpsHny&Jlpno Chili	150g	3	13.80
5	Old El Paso Salsa Dip Tomato Mild	300g	1	5.10
6	Smiths Crinkle Chips Salt & Vinegar	330g	1	5.70
7	Grain Waves Sweet Chilli	210g	1	3.60
8	Doritos Corn Chip Mexican Jalapeno	150g	1	3.90
9	Grain Waves Sour Cream&Chives	210G	2	7.20
10	Smiths Crinkle Chips Salt & Vinegar	330g	1	5.70
11	Kettle Sensations Siracha Lime	150g	5	23.00
12	Twisties Cheese	270g	2	9.20
13	WW Crinkle Cut Chicken	175g	1	1.70
14	Thins Chips Light& Tangy	175g	1	3.30
15	CCs Original	175g	1	2.10
16	Burger Rings	220g	4	9.20
17	NCC Sour Cream & Garden Chives	175g	1	3.00
18	Doritos Corn Chip Southern Chicken	150g	1	3.90
19	Cheezels Cheese Box	125g	1	2.10
20	Smiths Crinkle Original	330g	1	5.70
21	NCC Sour Cream & Garden Chives	175g	1	3.00
22	Infzns Crn Crnchers Tangy Gcamole	110g	1	3.80
23	Kettle Sea Salt And Vinegar	175g	1	5.40
24	Kettle Tortilla ChpsHny&Jlpno Chili	150g	2	9.20
25	Old El Paso Salsa Dip Tomato Mild	300g	1	5.10
26	Smiths Chip Thinly Cut Original	175g	2	6.00
27	Kettle Original	175g	1	5.40
28	Red Rock Deli Thai Chilli&Lime	150g	1	2.70
29	Infzns Crn Crnchers Tangy Gcamole	110g	1	3.80
30	Pringles Sthrn FriedChicken	134g	1	3.70
31	Pringles Sweet&Spcy BBQ	134g	4	14.80
32	Red Rock Deli SR Salsa & Mzzrilla	150g	2	5.40
33	Thins Chips Originl saltd	175g	1	3.30
34	Red Rock Deli Sp Salt & Truffle	150G	1	2.70
35	Smiths Thinly Swt Chli&S/Cream	175G	1	3.00
36	Kettle Chilli	175g	1	5.40
37	Doritos Mexicana	170g	1	4.40
38	Thins Chips Light& Tangy	175g	2	6.60
39	Smiths Crinkle Cut French OnionDip	150g	1	2.60
40	Natural ChipCo Hony Soy Chckn	175g	1	3.00
41	Dorito Corn Chp Supreme	380g	1	3.25
42	Twisties Chicken	270g	1	4.60
43	Smiths Thinly Cut Roast Chicken	175g	1	3.00
44	Smiths Crinkle Cut Tomato Salsa	150g	1	2.60
45	Kettle Mozzarella Basil & Pesto	175g	1	5.40
46	Infzns Crn Crnchers Tangy Gcamole	110g	1	3.80
47	Infuzions Thai SweetChili PotatoMix	110g	1	3.80

48	Smiths Crinkle	Original 330g	1	5.70
49	Kettle Sensations	Camembert & Fig 150g	1	4.60

Wonderful! Now we will go towards checking for any anomalies in the product names.

```
[528]: # Examine PROD_NAME to check for incorrect entries
productWords = pd.DataFrame(transaction_data['PROD_NAME'].str.
    ↪split(expand=True).stack(), columns=['words'])
print(productWords['words'].unique())

['Natural' 'Chip' 'Compny' 'SeaSalt175g' 'CCs' 'Nacho' 'Cheese' '175g'
'Smiths' 'Crinkle' 'Cut' 'Chips' 'Chicken' '170g' 'Thinly'
'S/Cream&Onion' 'Kettle' 'Tortilla' 'ChpsHny&Jlpno' 'Chili' '150g' 'Old'
'El' 'Paso' 'Salsa' 'Dip' 'Tomato' 'Mild' '300g' 'Salt' '&' 'Vinegar'
'330g' 'Grain' 'Waves' 'Sweet' 'Chilli' '210g' 'Doritos' 'Corn' 'Mexican'
'Jalapeno' 'Sour' 'Cream&Chives' '210G' 'Sensations' 'Siracha' 'Lime'
'Twisties' '270g' 'WW' 'Thins' 'Light&' 'Tangy' 'Original' 'Burger'
'Rings' '220g' 'NCC' 'Cream' 'Garden' 'Chives' 'Southern' 'Cheezels'
'Box' '125g' 'Infzns' 'Crn' 'Crnchers' 'Gcamole' '110g' 'Sea' 'And' 'Red'
'Rock' 'Deli' 'Thai' 'Chilli&Lime' 'Pringles' 'Sthrn' 'FriedChicken'
'134g' 'Sweet&Spcy' 'BBQ' 'SR' 'Mzzrlla' 'Originl' 'saltd' 'Sp' 'Truffle'
'150G' 'Swt' 'Chli&S/Cream175G' 'Mexicana' 'French' 'OnionDip' 'ChipCo'
'Hony' 'Soy' 'Chckn175g' 'Dorito' 'Chp' 'Supreme' '380g' 'Chicken270g'
'Roast' 'Mozzarella' 'Basil' 'Pesto' 'Infuzions' 'SweetChili' 'PotatoMix'
'Camembert' 'Fig' 'Smith' 'Mac' 'N' 'Honey' 'Seasonedchicken' 'Rib'
'Prawn' 'Crackers' 'GrnWves' 'Plus' 'Btroot' 'Jam' '180g' 'Tyrrells'
'Crisps' 'Lightly' 'Salted' '165g' 'Medium' '135g' 'Pot' 'SourCream'
'Onion' '250g' 'Chnky' 'Tom' 'Ht300g' 'Cobs' 'Popd' 'Swt/Chlli'
'&Sr/Cream' 'Woolworths' 'Co' 'Tmato' 'Hrb&Spce' 'Chs&Onion170g' 'Fries'
'Potato' 'Med' 'RRD' 'Chilli&' 'Coconut' '200g' 'Hot' 'Spicy' 'Crm'
'&Chives' 'Crnkle' 'Orgnl' 'Big' 'Bag' 'BBQ&Maple' 'D/Style' 'Crips'
'Stacked' '160g' 'CutSalt/Vinegr175g' 'Tostitos' 'Barbecue' 'Cheetos'
'Puffs' 'Splash' 'Of' 'ChpsBtroot&Ricotta' 'Tasty' '190g' 'Smoked'
'Chipotle' 'Barbeque' 'Mystery' 'Flavour' 'Ched' 'Snbts' 'Whlgrn'
'Cheddr&Mstrd' '90g' 'Chs' 'Bacon' 'Balls' 'Slt' 'Vingar'
'SourCream&Herbs' 'Veg' 'Strws' 'ChpsFeta&Garlic' 'Mango' 'Chutny'
'Papadums' '70g' 'Steak' 'Chimuchurri' 'Sunbites' 'Whlegrn' 'Frch/Onin'
'Snag&Sauce' '&OnionStacked' 'Pepper' 'Vinegr' 'Chikn&Garlic' 'Aioli'
'Slow' 'Rst' 'Pork' 'Belly' 'Pc' 'Bolognese']
```

```
[529]: transaction_data.head(50)
```

```
[529]:
```

	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
5	2019-05-19	4	4074	2982	57
6	2019-05-16	4	4149	3333	16
7	2019-05-16	4	4196	3539	24
8	2018-08-20	5	5026	4525	42
9	2018-08-18	7	7150	6900	52
10	2019-05-17	7	7215	7176	16
11	2018-08-20	8	8294	8221	114
12	2019-05-18	9	9208	8634	15
13	2018-08-17	13	13213	12447	92
14	2019-05-15	19	19272	16686	44
15	2019-05-19	20	20164	17136	54
16	2018-08-18	20	20418	17413	94
17	2018-08-14	22	22411	18646	98
18	2018-08-17	22	22456	18696	93
19	2019-05-16	23	23067	19162	56
20	2019-05-19	25	25105	21815	7
21	2018-08-16	33	33081	29949	98
22	2018-08-16	36	36012	32077	31
23	2018-08-19	36	36302	33188	32
24	2018-08-15	38	38142	34181	108
25	2019-05-15	39	39144	35506	57
26	2018-08-19	39	39167	35638	111
27	2019-05-15	41	41423	38393	46
28	2018-08-15	41	41486	38472	13
29	2019-05-20	43	43110	39342	31
30	2019-05-16	43	43147	39608	99
31	2019-05-15	43	43227	40186	26
32	2019-05-20	45	45127	41122	64
33	2019-05-18	45	45220	41651	22
34	2018-08-16	51	51100	46802	48
35	2018-08-19	51	51100	46803	37
36	2018-08-18	51	51113	46828	36
37	2018-08-17	54	54226	48173	51
38	2018-08-18	54	54305	48301	44
39	2018-08-18	55	55072	48878	107
40	2019-05-14	55	55072	48883	106
41	2019-05-20	55	55073	48887	4
42	2019-05-20	55	55073	48887	113
43	2019-05-16	55	55202	49690	45
44	2018-08-18	56	56013	50090	39
45	2019-05-16	58	58324	54252	102
46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3

	PROD_NAME	PROD_QTY	TOT_SALES
0	Natural Chip Compny SeaSalt175g	2	6.00
1	CCs Nacho Cheese 175g	3	6.30
2	Smiths Crinkle Cut Chips Chicken 170g	2	2.90
3	Smiths Chip Thinly S/Cream&Onion 175g	5	15.00
4	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.80
5	Old El Paso Salsa Dip Tomato Mild 300g	1	5.10
6	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70
7	Grain Waves Sweet Chilli 210g	1	3.60
8	Doritos Corn Chip Mexican Jalapeno 150g	1	3.90
9	Grain Waves Sour Cream&Chives 210G	2	7.20
10	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70
11	Kettle Sensations Siracha Lime 150g	5	23.00
12	Twisties Cheese 270g	2	9.20
13	WW Crinkle Cut Chicken 175g	1	1.70
14	Thins Chips Light& Tangy 175g	1	3.30
15	CCs Original 175g	1	2.10
16	Burger Rings 220g	4	9.20
17	NCC Sour Cream & Garden Chives 175g	1	3.00
18	Doritos Corn Chip Southern Chicken 150g	1	3.90
19	Cheezels Cheese Box 125g	1	2.10
20	Smiths Crinkle Original 330g	1	5.70
21	NCC Sour Cream & Garden Chives 175g	1	3.00
22	Infzns Crn Crnchers Tangy Gcamole 110g	1	3.80
23	Kettle Sea Salt And Vinegar 175g	1	5.40
24	Kettle Tortilla ChpsHny&Jlpno Chili 150g	2	9.20
25	Old El Paso Salsa Dip Tomato Mild 300g	1	5.10
26	Smiths Chip Thinly Cut Original 175g	2	6.00
27	Kettle Original 175g	1	5.40
28	Red Rock Deli Thai Chilli&Lime 150g	1	2.70
29	Infzns Crn Crnchers Tangy Gcamole 110g	1	3.80
30	Pringles Sthrn FriedChicken 134g	1	3.70
31	Pringles Sweet&Spcy BBQ 134g	4	14.80
32	Red Rock Deli SR Salsa & Mzzrlla 150g	2	5.40
33	Thins Chips Originl saltd 175g	1	3.30
34	Red Rock Deli Sp Salt & Truffle 150G	1	2.70
35	Smiths Thinly Swt Chli&S/Cream175G	1	3.00
36	Kettle Chilli 175g	1	5.40
37	Doritos Mexicana 170g	1	4.40
38	Thins Chips Light& Tangy 175g	2	6.60
39	Smiths Crinkle Cut French OnionDip 150g	1	2.60
40	Natural ChipCo Hony Soy Chckn175g	1	3.00
41	Dorito Corn Chp Supreme 380g	1	3.25
42	Twisties Chicken270g	1	4.60
43	Smiths Thinly Cut Roast Chicken 175g	1	3.00
44	Smiths Crinkle Cut Tomato Salsa 150g	1	2.60
45	Kettle Mozzarella Basil & Pesto 175g	1	5.40

46	Infzns Crn Crnchers Tangy Gcamole 110g	1	3.80
47	Infuzions Thai SweetChili PotatoMix 110g	1	3.80
48	Smiths Crinkle Original 330g	1	5.70
49	Kettle Sensations Camembert & Fig 150g	1	4.60

Just in case, we are also checking to remove any digits and special characters. These can cause error in the data.

```
[530]: # Remove digits and special characters from product words
productWords = productWords[~productWords['words'].str.contains(r'[\d&]')].
        ↪reset_index(drop=True)
print(productWords['words'].unique())
```

```
['Natural' 'Chip' 'Compny' 'CCs' 'Nacho' 'Cheese' 'Smiths' 'Crinkle' 'Cut'
 'Chips' 'Chicken' 'Thinly' 'Kettle' 'Tortilla' 'Chili' 'Old' 'El' 'Paso'
 'Salsa' 'Dip' 'Tomato' 'Mild' 'Salt' 'Vinegar' 'Grain' 'Waves' 'Sweet'
 'Chilli' 'Doritos' 'Corn' 'Mexican' 'Jalapeno' 'Sour' 'Sensations'
 'Siracha' 'Lime' 'Twisties' 'WW' 'Thins' 'Tangy' 'Original' 'Burger'
 'Rings' 'NCC' 'Cream' 'Garden' 'Chives' 'Southern' 'Cheezels' 'Box'
 'Infzns' 'Crn' 'Crnchers' 'Gcamole' 'Sea' 'And' 'Red' 'Rock' 'Deli'
 'Thai' 'Pringles' 'Sthrn' 'FriedChicken' 'BBQ' 'SR' 'Mzzrlla' 'Originl'
 'saltd' 'Sp' 'Truffle' 'Swt' 'Mexicana' 'French' 'OnionDip' 'ChipCo'
 'Hony' 'Soy' 'Dorito' 'Chp' 'Supreme' 'Roast' 'Mozzarella' 'Basil'
 'Pesto' 'Infuzions' 'SweetChili' 'PotatoMix' 'Camembert' 'Fig' 'Smith'
 'Mac' 'N' 'Honey' 'Seasonedchicken' 'Rib' 'Prawn' 'Crackers' 'GrnWves'
 'Plus' 'Btroot' 'Jam' 'Tyrrells' 'Crisps' 'Lightly' 'Salted' 'Medium'
 'Pot' 'SourCream' 'Onion' 'Chnky' 'Tom' 'Cobs' 'Popd' 'Swt/Chlli'
 'Woolworths' 'Co' 'Tmato' 'Fries' 'Potato' 'Med' 'RRD' 'Coconut' 'Hot'
 'Spicy' 'Crm' 'Crnkle' 'Orgnl' 'Big' 'Bag' 'D/Style' 'Crips' 'Stacked'
 'Tostitos' 'Barbecue' 'Cheetos' 'Puffs' 'Splash' 'Of' 'Tasty' 'Smoked'
 'Chipotle' 'Barbeque' 'Mystery' 'Flavour' 'Ched' 'Snbts' 'Whlgrn' 'Chs'
 'Bacon' 'Balls' 'Slt' 'Vingar' 'Veg' 'Strws' 'Mango' 'Chutny' 'Papadums'
 'Steak' 'Chimuchurri' 'Sunbites' 'Whlegrn' 'Frch/Onin' 'Pepper' 'Vinegr'
 'Aioli' 'Slow' 'Rst' 'Pork' 'Belly' 'Pc' 'Bolognese']
```

```
[531]: transaction_data.head(50)
```

```
[531]:
```

	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	
5	2019-05-19	4	4074	2982	57	
6	2019-05-16	4	4149	3333	16	
7	2019-05-16	4	4196	3539	24	
8	2018-08-20	5	5026	4525	42	
9	2018-08-18	7	7150	6900	52	

10	2019-05-17	7	7215	7176	16
11	2018-08-20	8	8294	8221	114
12	2019-05-18	9	9208	8634	15
13	2018-08-17	13	13213	12447	92
14	2019-05-15	19	19272	16686	44
15	2019-05-19	20	20164	17136	54
16	2018-08-18	20	20418	17413	94
17	2018-08-14	22	22411	18646	98
18	2018-08-17	22	22456	18696	93
19	2019-05-16	23	23067	19162	56
20	2019-05-19	25	25105	21815	7
21	2018-08-16	33	33081	29949	98
22	2018-08-16	36	36012	32077	31
23	2018-08-19	36	36302	33188	32
24	2018-08-15	38	38142	34181	108
25	2019-05-15	39	39144	35506	57
26	2018-08-19	39	39167	35638	111
27	2019-05-15	41	41423	38393	46
28	2018-08-15	41	41486	38472	13
29	2019-05-20	43	43110	39342	31
30	2019-05-16	43	43147	39608	99
31	2019-05-15	43	43227	40186	26
32	2019-05-20	45	45127	41122	64
33	2019-05-18	45	45220	41651	22
34	2018-08-16	51	51100	46802	48
35	2018-08-19	51	51100	46803	37
36	2018-08-18	51	51113	46828	36
37	2018-08-17	54	54226	48173	51
38	2018-08-18	54	54305	48301	44
39	2018-08-18	55	55072	48878	107
40	2019-05-14	55	55072	48883	106
41	2019-05-20	55	55073	48887	4
42	2019-05-20	55	55073	48887	113
43	2019-05-16	55	55202	49690	45
44	2018-08-18	56	56013	50090	39
45	2019-05-16	58	58324	54252	102
46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3

		PROD_NAME	PROD_QTY	TOT_SALES
0	Natural Chip	Compny SeaSalt175g	2	6.00
1		CCs Nacho Cheese 175g	3	6.30
2	Smiths Crinkle Cut	Chips Chicken 170g	2	2.90
3	Smiths Chip Thinly	S/Cream&Onion 175g	5	15.00
4	Kettle Tortilla	ChpsHny&Jlpno Chili 150g	3	13.80

5	Old El Paso Salsa	Dip Tomato Mild	300g	1	5.10
6	Smiths Crinkle Chips	Salt & Vinegar	330g	1	5.70
7	Grain Waves	Sweet Chilli	210g	1	3.60
8	Doritos Corn Chip	Mexican Jalapeno	150g	1	3.90
9	Grain Waves Sour	Cream&Chives	210G	2	7.20
10	Smiths Crinkle Chips	Salt & Vinegar	330g	1	5.70
11	Kettle Sensations	Siracha Lime	150g	5	23.00
12		Twisties Cheese	270g	2	9.20
13	WW Crinkle Cut	Chicken	175g	1	1.70
14	Thins Chips	Light& Tangy	175g	1	3.30
15		CCs Original	175g	1	2.10
16		Burger Rings	220g	4	9.20
17	NCC Sour Cream &	Garden Chives	175g	1	3.00
18	Doritos Corn Chip	Southern Chicken	150g	1	3.90
19		Cheezels Cheese Box	125g	1	2.10
20	Smiths Crinkle	Original	330g	1	5.70
21	NCC Sour Cream &	Garden Chives	175g	1	3.00
22	Infzns Crn Crnchers	Tangy Gcamole	110g	1	3.80
23	Kettle Sea Salt	And Vinegar	175g	1	5.40
24	Kettle Tortilla Chps	Hny&Jlpno Chili	150g	2	9.20
25	Old El Paso Salsa	Dip Tomato Mild	300g	1	5.10
26	Smiths Chip Thinly	Cut Original	175g	2	6.00
27		Kettle Original	175g	1	5.40
28	Red Rock Deli Thai	Chilli&Lime	150g	1	2.70
29	Infzns Crn Crnchers	Tangy Gcamole	110g	1	3.80
30	Pringles Sthrn	FriedChicken	134g	1	3.70
31	Pringles	Sweet&Spcy BBQ	134g	4	14.80
32	Red Rock Deli SR	Salsa & Mzzrlla	150g	2	5.40
33	Thins Chips	Originl saltd	175g	1	3.30
34	Red Rock Deli Sp	Salt & Truffle	150G	1	2.70
35	Smiths Thinly	Swt Chli&S/Cream	175G	1	3.00
36		Kettle Chilli	175g	1	5.40
37		Doritos Mexicana	170g	1	4.40
38	Thins Chips	Light& Tangy	175g	2	6.60
39	Smiths Crinkle Cut	French OnionDip	150g	1	2.60
40	Natural ChipCo	Hony Soy Chckn	175g	1	3.00
41	Dorito Corn Chp	Supreme	380g	1	3.25
42		Twisties Chicken	270g	1	4.60
43	Smiths Thinly Cut	Roast Chicken	175g	1	3.00
44	Smiths Crinkle Cut	Tomato Salsa	150g	1	2.60
45	Kettle Mozzarella	Basil & Pesto	175g	1	5.40
46	Infzns Crn Crnchers	Tangy Gcamole	110g	1	3.80
47	Infuzions Thai	SweetChili PotatoMix	110g	1	3.80
48	Smiths Crinkle	Original	330g	1	5.70
49	Kettle Sensations	Camembert & Fig	150g	1	4.60

Lets check the demographic and see that were is the customer focus mostly

```
[532]: # Count and sort distinct words by frequency of occurrence
word_counts = productWords['words'].value_counts().reset_index()
word_counts.columns = ['word', 'frequency']
print(word_counts)
```

	word	frequency
0	Chips	49770
1	Kettle	41288
2	Smiths	28860
3	Salt	27976
4	Cheese	27890
..
166	Frch/Onin	1432
167	Pc	1431
168	NCC	1419
169	Garden	1419
170	Fries	1418

[171 rows x 2 columns]

```
[533]: transaction_data.head(50)
```

```
[533]:
```

	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	
5	2019-05-19	4	4074	2982	57	
6	2019-05-16	4	4149	3333	16	
7	2019-05-16	4	4196	3539	24	
8	2018-08-20	5	5026	4525	42	
9	2018-08-18	7	7150	6900	52	
10	2019-05-17	7	7215	7176	16	
11	2018-08-20	8	8294	8221	114	
12	2019-05-18	9	9208	8634	15	
13	2018-08-17	13	13213	12447	92	
14	2019-05-15	19	19272	16686	44	
15	2019-05-19	20	20164	17136	54	
16	2018-08-18	20	20418	17413	94	
17	2018-08-14	22	22411	18646	98	
18	2018-08-17	22	22456	18696	93	
19	2019-05-16	23	23067	19162	56	
20	2019-05-19	25	25105	21815	7	
21	2018-08-16	33	33081	29949	98	
22	2018-08-16	36	36012	32077	31	
23	2018-08-19	36	36302	33188	32	

24	2018-08-15	38	38142	34181	108
25	2019-05-15	39	39144	35506	57
26	2018-08-19	39	39167	35638	111
27	2019-05-15	41	41423	38393	46
28	2018-08-15	41	41486	38472	13
29	2019-05-20	43	43110	39342	31
30	2019-05-16	43	43147	39608	99
31	2019-05-15	43	43227	40186	26
32	2019-05-20	45	45127	41122	64
33	2019-05-18	45	45220	41651	22
34	2018-08-16	51	51100	46802	48
35	2018-08-19	51	51100	46803	37
36	2018-08-18	51	51113	46828	36
37	2018-08-17	54	54226	48173	51
38	2018-08-18	54	54305	48301	44
39	2018-08-18	55	55072	48878	107
40	2019-05-14	55	55072	48883	106
41	2019-05-20	55	55073	48887	4
42	2019-05-20	55	55073	48887	113
43	2019-05-16	55	55202	49690	45
44	2018-08-18	56	56013	50090	39
45	2019-05-16	58	58324	54252	102
46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3

	PROD_NAME	PROD_QTY	TOT_SALES
0	Natural Chip Compny SeaSalt175g	2	6.00
1	CCs Nacho Cheese 175g	3	6.30
2	Smiths Crinkle Cut Chips Chicken 170g	2	2.90
3	Smiths Chip Thinly S/Cream&Onion 175g	5	15.00
4	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.80
5	Old El Paso Salsa Dip Tomato Mild 300g	1	5.10
6	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70
7	Grain Waves Sweet Chillli 210g	1	3.60
8	Doritos Corn Chip Mexican Jalapeno 150g	1	3.90
9	Grain Waves Sour Cream&Chives 210G	2	7.20
10	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70
11	Kettle Sensations Siracha Lime 150g	5	23.00
12	Twisties Cheese 270g	2	9.20
13	WW Crinkle Cut Chicken 175g	1	1.70
14	Thins Chips Light& Tangy 175g	1	3.30
15	CCs Original 175g	1	2.10
16	Burger Rings 220g	4	9.20
17	NCC Sour Cream & Garden Chives 175g	1	3.00
18	Doritos Corn Chip Southern Chicken 150g	1	3.90

19	Cheezels Cheese Box	125g	1	2.10
20	Smiths Crinkle	Original 330g	1	5.70
21	NCC Sour Cream &	Garden Chives 175g	1	3.00
22	Infzns Crn Crnchers	Tangy Gcamole 110g	1	3.80
23	Kettle Sea Salt	And Vinegar 175g	1	5.40
24	Kettle Tortilla ChpsHny&Jlpno	Chili 150g	2	9.20
25	Old El Paso Salsa	Dip Tomato Mild 300g	1	5.10
26	Smiths Chip Thinly	Cut Original 175g	2	6.00
27		Kettle Original 175g	1	5.40
28	Red Rock Deli Thai	Chilli&Lime 150g	1	2.70
29	Infzns Crn Crnchers	Tangy Gcamole 110g	1	3.80
30	Pringles Sthrn	FriedChicken 134g	1	3.70
31	Pringles Sweet&Spcy	BBQ 134g	4	14.80
32	Red Rock Deli SR	Salsa & Mzzrilla 150g	2	5.40
33	Thins Chips	Originl saltd 175g	1	3.30
34	Red Rock Deli Sp	Salt & Truffle 150G	1	2.70
35	Smiths Thinly	Swt Chli&S/Cream175G	1	3.00
36		Kettle Chilli 175g	1	5.40
37		Doritos Mexicana 170g	1	4.40
38	Thins Chips Light&	Tangy 175g	2	6.60
39	Smiths Crinkle Cut	French OnionDip 150g	1	2.60
40	Natural ChipCo	Hony Soy Chckn175g	1	3.00
41	Dorito Corn Chp	Supreme 380g	1	3.25
42		Twisties Chicken270g	1	4.60
43	Smiths Thinly Cut	Roast Chicken 175g	1	3.00
44	Smiths Crinkle Cut	Tomato Salsa 150g	1	2.60
45	Kettle Mozzarella	Basil & Pesto 175g	1	5.40
46	Infzns Crn Crnchers	Tangy Gcamole 110g	1	3.80
47	Infuzions Thai	SweetChili PotatoMix 110g	1	3.80
48	Smiths Crinkle	Original 330g	1	5.70
49	Kettle Sensations	Camembert & Fig 150g	1	4.60

Based on the uploaded solution, we are able to confirm that the mentioned 'salsa' are not flavoured chips but rather they are dips and hence not a part of our data. Hence, we shall remove them.

```
[534]: # Remove salsa products
transaction_data = transaction_data[~transaction_data['PROD_NAME'].str.
    .contains('salsa', case=False)]
print(transaction_data['PROD_NAME'].unique())
```

```
['Natural Chip          Compny SeaSalt175g' 'CCs Nacho Cheese    175g'
'Smiths Crinkle Cut   Chips Chicken 170g'
'Smiths Chip Thinly   S/Cream&Onion 175g'
'Kettle Tortilla ChpsHny&Jlpno Chili 150g'
'Smiths Crinkle Chips Salt & Vinegar 330g'
'Grain Waves          Sweet Chilli 210g'
'Doritos Corn Chip Mexican Jalapeno 150g'
'Grain Waves Sour     Cream&Chives 210G']
```

'Kettle Sensations Siracha Lime 150g' 'Twisties Cheese 270g'
 'WW Crinkle Cut Chicken 175g' 'Thins Chips Light& Tangy 175g'
 'CCs Original 175g' 'Burger Rings 220g'
 'NCC Sour Cream & Garden Chives 175g'
 'Doritos Corn Chip Southern Chicken 150g' 'Cheezels Cheese Box 125g'
 'Smiths Crinkle Original 330g'
 'Infzns Crn Crnchers Tangy Gcamole 110g'
 'Kettle Sea Salt And Vinegar 175g'
 'Smiths Chip Thinly Cut Original 175g' 'Kettle Original 175g'
 'Red Rock Deli Thai Chilli&Lime 150g' 'Pringles Sthrn FriedChicken 134g'
 'Pringles Sweet&Spcy BBQ 134g' 'Thins Chips Originl salted 175g'
 'Red Rock Deli Sp Salt & Truffle 150g'
 'Smiths Thinly Swt Chli&S/Cream175G' 'Kettle Chilli 175g'
 'Doritos Mexicana 170g' 'Smiths Crinkle Cut French OnionDip 150g'
 'Natural ChipCo Honey Soy Chckn175g'
 'Dorito Corn Chp Supreme 380g' 'Twisties Chicken270g'
 'Smiths Thinly Cut Roast Chicken 175g'
 'Kettle Mozzarella Basil & Pesto 175g'
 'Infuzions Thai SweetChili PotatoMix 110g'
 'Kettle Sensations Camembert & Fig 150g'
 'Smith Crinkle Cut Mac N Cheese 150g'
 'Kettle Honey Soy Chicken 175g' 'Thins Chips Seasonedchicken 175g'
 'Smiths Crinkle Cut Salt & Vinegar 170g'
 'Infuzions BBQ Rib Prawn Crackers 110g'
 'GrnWves Plus Btroot & Chilli Jam 180g'
 'Tyrrells Crisps Lightly Salted 165g'
 'Kettle Sweet Chilli And Sour Cream 175g' 'Kettle 135g Swt Pot Sea Salt'
 'Pringles SourCream Onion 134g' 'Doritos Corn Chips Original 170g'
 'Twisties Cheese Burger 250g'
 'Cobs Popd Swt/Chlli &Sr/Cream Chips 110g'
 'Natural Chip Co Tmato Hrb&Spce 175g'
 'Smiths Crinkle Cut Chips Original 170g'
 'Cobs Popd Sea Salt Chips 110g'
 'Smiths Crinkle Cut Chips Chs&Onion170g'
 'French Fries Potato Chips 175g'
 'Doritos Corn Chips Cheese Supreme 170g'
 'Pringles Original Crisps 134g' 'RRD Chilli& Coconut 150g'
 'WW Original Corn Chips 200g' 'Thins Potato Chips Hot & Spicy 175g'
 'Cobs Popd Sour Crm &Chives Chips 110g'
 'Smiths Crnkle Chip Orgnl Big Bag 380g'
 'Doritos Corn Chips Nacho Cheese 170g'
 'Kettle Sensations BBQ&Maple 150g' 'WW D/Style Chip Sea Salt 200g'
 'Pringles Chicken Salt Crisps 134g' 'WW Original Stacked Chips 160g'
 'Smiths Chip Thinly CutSalt/Vinegr175g' 'Cheezels Cheese 330g'
 'Tostitos Lightly Salted 175g' 'Thins Chips Salt & Vinegar 175g'
 'Smiths Crinkle Cut Chips Barbecue 170g' 'Cheetos Puffs 165g'
 'RRD Sweet Chilli & Sour Cream 165g' 'WW Crinkle Cut Original 175g'
 'Tostitos Splash Of Lime 175g' 'Kettle Tortilla ChpsBtroot&Ricotta 150g'

```

'CCs Tasty Cheese      175g' 'Woolworths Cheese   Rings 190g'
'Tostitos Smoked      Chipotle 175g' 'Pringles Barbeque   134g'
'WW Supreme Cheese    Corn Chips 200g' 'Pringles Mystery    Flavour 134g'
'Tyrrells Crisps      Ched & Chives 165g'
'Snbts Whlgrn Crisps  Cheddr&Mstrd 90g' 'Cheetos Chs & Bacon Balls 190g'
'Pringles Slt Vingar 134g' 'Infuzions SourCream&Herbs Veg Strws 110g'
'Kettle Tortilla ChpsFeta&Garlic 150g'
'Infuzions Mango      Chutny Papadums 70g'
'RRD Steak &          Chimuchurri 150g' 'RRD Honey Soy        Chicken 165g'
'Sunbites Whlegrn     Crisps Frch/Onin 90g' 'RRD Salt & Vinegar   165g'
'Doritos Cheese       Supreme 330g' 'Smiths Crinkle Cut   Snag&Sauce 150g'
'WW Sour Cream &OnionStacked Chips 160g' 'RRD Lime & Pepper    165g'
'Natural ChipCo Sea   Salt & Vinegr 175g'
'Red Rock Deli Chikn&Garlic Aioli 150g'
'RRD SR Slow Rst      Pork Belly 150g' 'RRD Pc Sea Salt      165g'
'Smith Crinkle Cut    Bolognese 150g']

```

Lets classify the chips packets based on their weight in grams

```

[535]: # Convert PACK_SIZE to numeric
transaction_data['PACK_SIZE'] = transaction_data['PROD_NAME'].str.
      ↪extract('(\d+)').astype(float)
print(transaction_data['PACK_SIZE'].unique())

```

```

[175. 170. 150. 330. 210. 270. 220. 125. 110. 134. 380. 180. 165. 135.
 250. 200. 160. 190.  90.  70.]

```

```

[536]: transaction_data.head(50)

```

```

[536]:
      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
6  2019-05-16         4          4149      3333        16
7  2019-05-16         4          4196      3539        24
8  2018-08-20         5          5026      4525        42
9  2018-08-18         7          7150      6900        52
10 2019-05-17         7          7215      7176        16
11 2018-08-20         8          8294      8221       114
12 2019-05-18         9          9208      8634        15
13 2018-08-17        13         13213     12447        92
14 2019-05-15        19         19272     16686        44
15 2019-05-19        20         20164     17136        54
16 2018-08-18        20         20418     17413        94
17 2018-08-14        22         22411     18646        98
18 2018-08-17        22         22456     18696        93

```

19	2019-05-16	23	23067	19162	56
20	2019-05-19	25	25105	21815	7
21	2018-08-16	33	33081	29949	98
22	2018-08-16	36	36012	32077	31
23	2018-08-19	36	36302	33188	32
24	2018-08-15	38	38142	34181	108
26	2018-08-19	39	39167	35638	111
27	2019-05-15	41	41423	38393	46
28	2018-08-15	41	41486	38472	13
29	2019-05-20	43	43110	39342	31
30	2019-05-16	43	43147	39608	99
31	2019-05-15	43	43227	40186	26
33	2019-05-18	45	45220	41651	22
34	2018-08-16	51	51100	46802	48
35	2018-08-19	51	51100	46803	37
36	2018-08-18	51	51113	46828	36
37	2018-08-17	54	54226	48173	51
38	2018-08-18	54	54305	48301	44
39	2018-08-18	55	55072	48878	107
40	2019-05-14	55	55072	48883	106
41	2019-05-20	55	55073	48887	4
42	2019-05-20	55	55073	48887	113
43	2019-05-16	55	55202	49690	45
45	2019-05-16	58	58324	54252	102
46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3
50	2019-05-20	63	63201	60709	82
51	2018-08-20	67	67053	64137	88
52	2019-05-20	71	71122	69717	40
53	2019-05-19	71	71212	70337	73

		PROD_NAME	PROD_QTY	TOT_SALES	PACK_SIZE
0	Natural Chip	Compny SeaSalt175g	2	6.00	175.0
1		CCs Nacho Cheese 175g	3	6.30	175.0
2	Smiths Crinkle Cut	Chips Chicken 170g	2	2.90	170.0
3	Smiths Chip Thinly	S/Cream&Onion 175g	5	15.00	175.0
4	Kettle Tortilla ChpsHny&Jlpno	Chili 150g	3	13.80	150.0
6	Smiths Crinkle Chips	Salt & Vinegar 330g	1	5.70	330.0
7	Grain Waves	Sweet Chilli 210g	1	3.60	210.0
8	Doritos Corn Chip Mexican	Jalapeno 150g	1	3.90	150.0
9	Grain Waves Sour	Cream&Chives 210G	2	7.20	210.0
10	Smiths Crinkle Chips	Salt & Vinegar 330g	1	5.70	330.0
11	Kettle Sensations	Siracha Lime 150g	5	23.00	150.0
12		Twisties Cheese 270g	2	9.20	270.0
13	WW Crinkle Cut	Chicken 175g	1	1.70	175.0

14	Thins Chips Light& Tangy	175g	1	3.30	175.0
15	CCs Original	175g	1	2.10	175.0
16	Burger Rings	220g	4	9.20	220.0
17	NCC Sour Cream & Garden Chives	175g	1	3.00	175.0
18	Doritos Corn Chip Southern Chicken	150g	1	3.90	150.0
19	Cheezels Cheese Box	125g	1	2.10	125.0
20	Smiths Crinkle Original	330g	1	5.70	330.0
21	NCC Sour Cream & Garden Chives	175g	1	3.00	175.0
22	Infzns Crn Crnchers Tangy Gcamole	110g	1	3.80	110.0
23	Kettle Sea Salt And Vinegar	175g	1	5.40	175.0
24	Kettle Tortilla ChpsHny&Jlpno Chili	150g	2	9.20	150.0
26	Smiths Chip Thinly Cut Original	175g	2	6.00	175.0
27	Kettle Original	175g	1	5.40	175.0
28	Red Rock Deli Thai Chilli&Lime	150g	1	2.70	150.0
29	Infzns Crn Crnchers Tangy Gcamole	110g	1	3.80	110.0
30	Pringles Sthrn FriedChicken	134g	1	3.70	134.0
31	Pringles Sweet&Spcy BBQ	134g	4	14.80	134.0
33	Thins Chips Originl salted	175g	1	3.30	175.0
34	Red Rock Deli Sp Salt & Truffle	150G	1	2.70	150.0
35	Smiths Thinly Swt Chli&S/Cream	175G	1	3.00	175.0
36	Kettle Chilli	175g	1	5.40	175.0
37	Doritos Mexicana	170g	1	4.40	170.0
38	Thins Chips Light& Tangy	175g	2	6.60	175.0
39	Smiths Crinkle Cut French OnionDip	150g	1	2.60	150.0
40	Natural ChipCo Hony Soy Chckn	175g	1	3.00	175.0
41	Dorito Corn Chp Supreme	380g	1	3.25	380.0
42	Twisties Chicken	270g	1	4.60	270.0
43	Smiths Thinly Cut Roast Chicken	175g	1	3.00	175.0
45	Kettle Mozzarella Basil & Pesto	175g	1	5.40	175.0
46	Infzns Crn Crnchers Tangy Gcamole	110g	1	3.80	110.0
47	Infuzions Thai SweetChili PotatoMix	110g	1	3.80	110.0
48	Smiths Crinkle Original	330g	1	5.70	330.0
49	Kettle Sensations Camembert & Fig	150g	1	4.60	150.0
50	Smith Crinkle Cut Mac N Cheese	150g	1	2.60	150.0
51	Kettle Honey Soy Chicken	175g	1	5.40	175.0
52	Thins Chips Seasonedchicken	175g	1	3.30	175.0
53	Smiths Crinkle Cut Salt & Vinegar	170g	2	5.80	170.0

In the data, there is random data with no identification for the products, so let's extract the brand names and see what they are.

```
[566]: # Extract brand from product name
transaction_data['BRAND'] = transaction_data['PROD_NAME'].str.split().str[0]
print(transaction_data['BRAND'].unique())
```

```
['Natural' 'CCs' 'Smiths' 'Kettle' 'Grain' 'Doritos' 'Twisties' 'WW'
 'Thins' 'Burger' 'NCC' 'Cheezels' 'Infzns' 'Red' 'Pringles' 'Dorito'
 'Infuzions' 'Smith' 'GrnWves' 'Tyrrells' 'Cobs' 'French' 'RRD' 'Tostitos']
```



```
'Cheetos' 'Woolworths' 'Snbts' 'Sunbites']
```

```
[567]: transaction_data.head(50)
```

```
[567]:
```

	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	
6	2019-05-16	4	4149	3333	16	
7	2019-05-16	4	4196	3539	24	
8	2018-08-20	5	5026	4525	42	
9	2018-08-18	7	7150	6900	52	
10	2019-05-17	7	7215	7176	16	
11	2018-08-20	8	8294	8221	114	
12	2019-05-18	9	9208	8634	15	
13	2018-08-17	13	13213	12447	92	
14	2019-05-15	19	19272	16686	44	
15	2019-05-19	20	20164	17136	54	
16	2018-08-18	20	20418	17413	94	
17	2018-08-14	22	22411	18646	98	
18	2018-08-17	22	22456	18696	93	
19	2019-05-16	23	23067	19162	56	
20	2019-05-19	25	25105	21815	7	
21	2018-08-16	33	33081	29949	98	
22	2018-08-16	36	36012	32077	31	
23	2018-08-19	36	36302	33188	32	
24	2018-08-15	38	38142	34181	108	
26	2018-08-19	39	39167	35638	111	
27	2019-05-15	41	41423	38393	46	
28	2018-08-15	41	41486	38472	13	
29	2019-05-20	43	43110	39342	31	
30	2019-05-16	43	43147	39608	99	
31	2019-05-15	43	43227	40186	26	
33	2019-05-18	45	45220	41651	22	
34	2018-08-16	51	51100	46802	48	
35	2018-08-19	51	51100	46803	37	
36	2018-08-18	51	51113	46828	36	
37	2018-08-17	54	54226	48173	51	
38	2018-08-18	54	54305	48301	44	
39	2018-08-18	55	55072	48878	107	
40	2019-05-14	55	55072	48883	106	
41	2019-05-20	55	55073	48887	4	
42	2019-05-20	55	55073	48887	113	
43	2019-05-16	55	55202	49690	45	
45	2019-05-16	58	58324	54252	102	

46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3
50	2019-05-20	63	63201	60709	82
51	2018-08-20	67	67053	64137	88
52	2019-05-20	71	71122	69717	40
53	2019-05-19	71	71212	70337	73

	PROD_NAME	PROD_QTY	TOT_SALES	PACK_SIZE	\
0	Natural Chip Compny SeaSalt175g	2	6.00	175.0	
1	CCs Nacho Cheese 175g	3	6.30	175.0	
2	Smiths Crinkle Cut Chips Chicken 170g	2	2.90	170.0	
3	Smiths Chip Thinly S/Cream&Onion 175g	5	15.00	175.0	
4	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.80	150.0	
6	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70	330.0	
7	Grain Waves Sweet Chilli 210g	1	3.60	210.0	
8	Doritos Corn Chip Mexican Jalapeno 150g	1	3.90	150.0	
9	Grain Waves Sour Cream&Chives 210G	2	7.20	210.0	
10	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70	330.0	
11	Kettle Sensations Siracha Lime 150g	5	23.00	150.0	
12	Twisties Cheese 270g	2	9.20	270.0	
13	WW Crinkle Cut Chicken 175g	1	1.70	175.0	
14	Thins Chips Light& Tangy 175g	1	3.30	175.0	
15	CCs Original 175g	1	2.10	175.0	
16	Burger Rings 220g	4	9.20	220.0	
17	NCC Sour Cream & Garden Chives 175g	1	3.00	175.0	
18	Doritos Corn Chip Southern Chicken 150g	1	3.90	150.0	
19	Cheezels Cheese Box 125g	1	2.10	125.0	
20	Smiths Crinkle Original 330g	1	5.70	330.0	
21	NCC Sour Cream & Garden Chives 175g	1	3.00	175.0	
22	Infzns Crn Crnchers Tangy Gcamole 110g	1	3.80	110.0	
23	Kettle Sea Salt And Vinegar 175g	1	5.40	175.0	
24	Kettle Tortilla ChpsHny&Jlpno Chili 150g	2	9.20	150.0	
26	Smiths Chip Thinly Cut Original 175g	2	6.00	175.0	
27	Kettle Original 175g	1	5.40	175.0	
28	Red Rock Deli Thai Chilli&Lime 150g	1	2.70	150.0	
29	Infzns Crn Crnchers Tangy Gcamole 110g	1	3.80	110.0	
30	Pringles Sthrn FriedChicken 134g	1	3.70	134.0	
31	Pringles Sweet&Spcy BBQ 134g	4	14.80	134.0	
33	Thins Chips Originl salted 175g	1	3.30	175.0	
34	Red Rock Deli Sp Salt & Truffle 150G	1	2.70	150.0	
35	Smiths Thinly Swt Chli&S/Cream175G	1	3.00	175.0	
36	Kettle Chilli 175g	1	5.40	175.0	
37	Doritos Mexicana 170g	1	4.40	170.0	
38	Thins Chips Light& Tangy 175g	2	6.60	175.0	
39	Smiths Crinkle Cut French OnionDip 150g	1	2.60	150.0	

40	Natural ChipCo	Hony Soy Chckn	175g	1	3.00	175.0
41	Dorito Corn Chp	Supreme	380g	1	3.25	380.0
42		Twisties Chicken	270g	1	4.60	270.0
43	Smiths Thinly Cut	Roast Chicken	175g	1	3.00	175.0
45	Kettle Mozzarella	Basil & Pesto	175g	1	5.40	175.0
46	Infzns Crn Crnchers	Tangy Gcamole	110g	1	3.80	110.0
47	Infuzions Thai SweetChili	PotatoMix	110g	1	3.80	110.0
48	Smiths Crinkle	Original	330g	1	5.70	330.0
49	Kettle Sensations	Camembert & Fig	150g	1	4.60	150.0
50	Smith Crinkle Cut	Mac N Cheese	150g	1	2.60	150.0
51	Kettle Honey Soy	Chicken	175g	1	5.40	175.0
52	Thins Chips	Seasonedchicken	175g	1	3.30	175.0
53	Smiths Crinkle Cut	Salt & Vinegar	170g	2	5.80	170.0

BRAND

0	Natural
1	CCs
2	Smiths
3	Smiths
4	Kettle
6	Smiths
7	Grain
8	Doritos
9	Grain
10	Smiths
11	Kettle
12	Twisties
13	WW
14	Thins
15	CCs
16	Burger
17	NCC
18	Doritos
19	Cheezels
20	Smiths
21	NCC
22	Infzns
23	Kettle
24	Kettle
26	Smiths
27	Kettle
28	Red
29	Infzns
30	Pringles
31	Pringles
33	Thins
34	Red

```

35     Smiths
36     Kettle
37     Doritos
38     Thins
39     Smiths
40     Natural
41     Dorito
42     Twisties
43     Smiths
44     Kettle
45     Infzns
46     Infuzions
47     Smiths
48     Kettle
49     Smith
50     Kettle
51     Thins
52     Smiths

```

```

[539]: # Combine similar brand names
transaction_data['BRAND'] =transaction_data['BRAND'].replace('RED', 'RRD')
print(transaction_data['BRAND'].unique())

```

```

['Natural' 'CCs' 'Smiths' 'Kettle' 'Grain' 'Doritos' 'Twisties' 'WW'
 'Thins' 'Burger' 'NCC' 'Cheezels' 'Infzns' 'Red' 'Pringles' 'Dorito'
 'Infuzions' 'Smith' 'GrnWves' 'Tyrrells' 'Cobs' 'French' 'RRD' 'Tostitos'
 'Cheetos' 'Woolworths' 'Snbts' 'Sunbites']

```

```

[540]: transaction_data.head(50)

```

```

[540]:
      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
6  2019-05-16         4           4149      3333        16
7  2019-05-16         4           4196      3539        24
8  2018-08-20         5           5026      4525        42
9  2018-08-18         7           7150      6900        52
10 2019-05-17         7           7215      7176        16
11 2018-08-20         8           8294      8221       114
12 2019-05-18         9           9208      8634        15
13 2018-08-17        13          13213     12447        92
14 2019-05-15        19          19272     16686        44
15 2019-05-19        20          20164     17136        54
16 2018-08-18        20          20418     17413        94

```

17	2018-08-14	22	22411	18646	98
18	2018-08-17	22	22456	18696	93
19	2019-05-16	23	23067	19162	56
20	2019-05-19	25	25105	21815	7
21	2018-08-16	33	33081	29949	98
22	2018-08-16	36	36012	32077	31
23	2018-08-19	36	36302	33188	32
24	2018-08-15	38	38142	34181	108
26	2018-08-19	39	39167	35638	111
27	2019-05-15	41	41423	38393	46
28	2018-08-15	41	41486	38472	13
29	2019-05-20	43	43110	39342	31
30	2019-05-16	43	43147	39608	99
31	2019-05-15	43	43227	40186	26
33	2019-05-18	45	45220	41651	22
34	2018-08-16	51	51100	46802	48
35	2018-08-19	51	51100	46803	37
36	2018-08-18	51	51113	46828	36
37	2018-08-17	54	54226	48173	51
38	2018-08-18	54	54305	48301	44
39	2018-08-18	55	55072	48878	107
40	2019-05-14	55	55072	48883	106
41	2019-05-20	55	55073	48887	4
42	2019-05-20	55	55073	48887	113
43	2019-05-16	55	55202	49690	45
45	2019-05-16	58	58324	54252	102
46	2019-05-16	59	59344	56007	31
47	2018-08-16	60	60038	56304	104
48	2019-05-15	60	60162	56825	7
49	2019-05-14	62	62177	58848	3
50	2019-05-20	63	63201	60709	82
51	2018-08-20	67	67053	64137	88
52	2019-05-20	71	71122	69717	40
53	2019-05-19	71	71212	70337	73

	PROD_NAME	PROD_QTY	TOT_SALES	PACK_SIZE	\
0	Natural Chip Compny SeaSalt175g	2	6.00	175.0	
1	CCs Nacho Cheese 175g	3	6.30	175.0	
2	Smiths Crinkle Cut Chips Chicken 170g	2	2.90	170.0	
3	Smiths Chip Thinly S/Cream&Onion 175g	5	15.00	175.0	
4	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.80	150.0	
6	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70	330.0	
7	Grain Waves Sweet Chilli 210g	1	3.60	210.0	
8	Doritos Corn Chip Mexican Jalapeno 150g	1	3.90	150.0	
9	Grain Waves Sour Cream&Chives 210G	2	7.20	210.0	
10	Smiths Crinkle Chips Salt & Vinegar 330g	1	5.70	330.0	
11	Kettle Sensations Siracha Lime 150g	5	23.00	150.0	

12	Twisties Cheese	270g	2	9.20	270.0
13	WW Crinkle Cut	Chicken 175g	1	1.70	175.0
14	Thins Chips Light&	Tangy 175g	1	3.30	175.0
15	CCs Original	175g	1	2.10	175.0
16	Burger Rings	220g	4	9.20	220.0
17	NCC Sour Cream &	Garden Chives 175g	1	3.00	175.0
18	Doritos Corn Chip	Southern Chicken 150g	1	3.90	150.0
19	Cheezels Cheese Box	125g	1	2.10	125.0
20	Smiths Crinkle	Original 330g	1	5.70	330.0
21	NCC Sour Cream &	Garden Chives 175g	1	3.00	175.0
22	Infzns Crn Crnchers	Tangy Gcamole 110g	1	3.80	110.0
23	Kettle Sea Salt	And Vinegar 175g	1	5.40	175.0
24	Kettle Tortilla ChpsHny&Jlpno	Chili 150g	2	9.20	150.0
26	Smiths Chip Thinly	Cut Original 175g	2	6.00	175.0
27	Kettle Original	175g	1	5.40	175.0
28	Red Rock Deli Thai	Chilli&Lime 150g	1	2.70	150.0
29	Infzns Crn Crnchers	Tangy Gcamole 110g	1	3.80	110.0
30	Pringles Sthrn Fried	Chicken 134g	1	3.70	134.0
31	Pringles Sweet&Spcy	BBQ 134g	4	14.80	134.0
33	Thins Chips	Originl salted 175g	1	3.30	175.0
34	Red Rock Deli Sp	Salt & Truffle 150G	1	2.70	150.0
35	Smiths Thinly	Swt Chli&S/Cream 175G	1	3.00	175.0
36	Kettle Chilli	175g	1	5.40	175.0
37	Doritos Mexicana	170g	1	4.40	170.0
38	Thins Chips Light&	Tangy 175g	2	6.60	175.0
39	Smiths Crinkle Cut	French OnionDip 150g	1	2.60	150.0
40	Natural ChipCo	Hony Soy Chckn 175g	1	3.00	175.0
41	Dorito Corn Chp	Supreme 380g	1	3.25	380.0
42	Twisties Chicken	270g	1	4.60	270.0
43	Smiths Thinly Cut	Roast Chicken 175g	1	3.00	175.0
45	Kettle Mozzarella	Basil & Pesto 175g	1	5.40	175.0
46	Infzns Crn Crnchers	Tangy Gcamole 110g	1	3.80	110.0
47	Infuzions Thai Sweet	Chili PotatoMix 110g	1	3.80	110.0
48	Smiths Crinkle	Original 330g	1	5.70	330.0
49	Kettle Sensations	Camembert & Fig 150g	1	4.60	150.0
50	Smith Crinkle Cut	Mac N Cheese 150g	1	2.60	150.0
51	Kettle Honey Soy	Chicken 175g	1	5.40	175.0
52	Thins Chips Seasoned	chicken 175g	1	3.30	175.0
53	Smiths Crinkle Cut	Salt & Vinegar 170g	2	5.80	170.0

BRAND

0	Natural
1	CCs
2	Smiths
3	Smiths
4	Kettle
6	Smiths

7	Grain
8	Doritos
9	Grain
10	Smiths
11	Kettle
12	Twisties
13	WW
14	Thins
15	CCs
16	Burger
17	NCC
18	Doritos
19	Cheezeels
20	Smiths
21	NCC
22	Infzns
23	Kettle
24	Kettle
26	Smiths
27	Kettle
28	Red
29	Infzns
30	Pringles
31	Pringles
33	Thins
34	Red
35	Smiths
36	Kettle
37	Doritos
38	Thins
39	Smiths
40	Natural
41	Dorito
42	Twisties
43	Smiths
45	Kettle
46	Infzns
47	Infuzions
48	Smiths
49	Kettle
50	Smith
51	Kettle
52	Thins
53	Smiths

Finally, we are sure that there are no further issues in transaction data so we will merge the CSV files.

```
[541]: # Merge transaction data and customer data
data = pd.merge(transaction_data, purchase_behaviour_data, how='left',
                on='LYLTY_CARD_NBR')
data.head()
```

```
[541]:
```

	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	

	PROD_NAME	PROD_QTY	TOT_SALES	PACK_SIZE	\
0	Natural Chip Compny SeaSalt175g	2	6.0	175.0	
1	CCs Nacho Cheese 175g	3	6.3	175.0	
2	Smiths Crinkle Cut Chips Chicken 170g	2	2.9	170.0	
3	Smiths Chip Thinly S/Cream&Onion 175g	5	15.0	175.0	
4	Kettle Tortilla ChpsHny&Jlpno Chili 150g	3	13.8	150.0	

	BRAND	LIFESTAGE	PREMIUM_CUSTOMER
0	Natural	YOUNG SINGLES/COUPLES	Premium
1	CCs	MIDAGE SINGLES/COUPLES	Budget
2	Smiths	MIDAGE SINGLES/COUPLES	Budget
3	Smiths	MIDAGE SINGLES/COUPLES	Budget
4	Kettle	MIDAGE SINGLES/COUPLES	Budget

1.1 Merged Data

Now we calculate the total sales and classify customers based on their age groups and their customer category.

```
[542]: # Calculate total sales by LIFESTAGE and PREMIUM_CUSTOMER
total_sales = data.groupby(['LIFESTAGE', 'PREMIUM_CUSTOMER'])['TOT_SALES'].
                sum().reset_index()
print(total_sales)
```

	LIFESTAGE	PREMIUM_CUSTOMER	TOT_SALES
0	MIDAGE SINGLES/COUPLES	Budget	33345.70
1	MIDAGE SINGLES/COUPLES	Mainstream	84734.25
2	MIDAGE SINGLES/COUPLES	Premium	54443.85
3	NEW FAMILIES	Budget	20607.45
4	NEW FAMILIES	Mainstream	15979.70
5	NEW FAMILIES	Premium	10760.80
6	OLDER FAMILIES	Budget	156863.75
7	OLDER FAMILIES	Mainstream	96413.55
8	OLDER FAMILIES	Premium	76542.60
9	OLDER SINGLES/COUPLES	Budget	127833.60

10	OLDER SINGLES/COUPLES	Mainstream	124648.50
11	OLDER SINGLES/COUPLES	Premium	123537.55
12	RETIREES	Budget	105916.30
13	RETIREES	Mainstream	145168.95
14	RETIREES	Premium	91296.65
15	YOUNG FAMILIES	Budget	129717.95
16	YOUNG FAMILIES	Mainstream	86338.25
17	YOUNG FAMILIES	Premium	78571.70
18	YOUNG SINGLES/COUPLES	Budget	57122.10
19	YOUNG SINGLES/COUPLES	Mainstream	147582.20
20	YOUNG SINGLES/COUPLES	Premium	39052.30

Customer count:

```
[543]: # Calculate number of customers by LIFESTAGE and PREMIUM_CUSTOMER
num_customers = data.groupby(['LIFESTAGE',
    ↳ 'PREMIUM_CUSTOMER'])['LYLTY_CARD_NBR'].nunique().reset_index()
print(num_customers)
```

	LIFESTAGE	PREMIUM_CUSTOMER	LYLTY_CARD_NBR
0	MIDAGE SINGLES/COUPLES	Budget	1474
1	MIDAGE SINGLES/COUPLES	Mainstream	3298
2	MIDAGE SINGLES/COUPLES	Premium	2369
3	NEW FAMILIES	Budget	1087
4	NEW FAMILIES	Mainstream	830
5	NEW FAMILIES	Premium	575
6	OLDER FAMILIES	Budget	4611
7	OLDER FAMILIES	Mainstream	2788
8	OLDER FAMILIES	Premium	2232
9	OLDER SINGLES/COUPLES	Budget	4849
10	OLDER SINGLES/COUPLES	Mainstream	4858
11	OLDER SINGLES/COUPLES	Premium	4682
12	RETIREES	Budget	4385
13	RETIREES	Mainstream	6358
14	RETIREES	Premium	3812
15	YOUNG FAMILIES	Budget	3953
16	YOUNG FAMILIES	Mainstream	2685
17	YOUNG FAMILIES	Premium	2398
18	YOUNG SINGLES/COUPLES	Budget	3647
19	YOUNG SINGLES/COUPLES	Mainstream	7917
20	YOUNG SINGLES/COUPLES	Premium	2480

Average number of units per customer:

```
[544]: # Calculate average number of units per customer by LIFESTAGE and
    ↳ PREMIUM_CUSTOMER
avg_units_per_customer = data.groupby(['LIFESTAGE',
    ↳ 'PREMIUM_CUSTOMER'])['PROD_QTY'].mean().reset_index()
print(avg_units_per_customer)
```

	LIFESTAGE	PREMIUM_CUSTOMER	PROD_QTY
0	MIDAGE	SINGLES/COUPLES	Budget 1.893626
1	MIDAGE	SINGLES/COUPLES	Mainstream 1.911942
2	MIDAGE	SINGLES/COUPLES	Premium 1.891750
3		NEW FAMILIES	Budget 1.855878
4		NEW FAMILIES	Mainstream 1.858124
5		NEW FAMILIES	Premium 1.860887
6		OLDER FAMILIES	Budget 1.945384
7		OLDER FAMILIES	Mainstream 1.948795
8		OLDER FAMILIES	Premium 1.983566
9	OLDER	SINGLES/COUPLES	Budget 1.914920
10	OLDER	SINGLES/COUPLES	Mainstream 1.911201
11	OLDER	SINGLES/COUPLES	Premium 1.913949
12		RETIREEES	Budget 1.893286
13		RETIREEES	Mainstream 1.886680
14		RETIREEES	Premium 1.901438
15		YOUNG FAMILIES	Budget 1.941226
16		YOUNG FAMILIES	Mainstream 1.941408
17		YOUNG FAMILIES	Premium 1.938149
18	YOUNG	SINGLES/COUPLES	Budget 1.808002
19	YOUNG	SINGLES/COUPLES	Mainstream 1.853510
20	YOUNG	SINGLES/COUPLES	Premium 1.807075

Average price per unit:

```
[545]: # Calculate average price per unit by LIFESTAGE and PREMIUM_CUSTOMER
avg_price_per_unit = data.groupby(['LIFESTAGE',
    ↪ 'PREMIUM_CUSTOMER'])['TOT_SALES'].mean().reset_index()
print(avg_price_per_unit) # print the result
```

	LIFESTAGE	PREMIUM_CUSTOMER	TOT_SALES
0	MIDAGE	SINGLES/COUPLES	Budget 7.108442
1	MIDAGE	SINGLES/COUPLES	Mainstream 7.637156
2	MIDAGE	SINGLES/COUPLES	Premium 7.152371
3		NEW FAMILIES	Budget 7.297256
4		NEW FAMILIES	Mainstream 7.313364
5		NEW FAMILIES	Premium 7.231720
6		OLDER FAMILIES	Budget 7.291241
7		OLDER FAMILIES	Mainstream 7.281440
8		OLDER FAMILIES	Premium 7.356329
9	OLDER	SINGLES/COUPLES	Budget 7.444305
10	OLDER	SINGLES/COUPLES	Mainstream 7.306049
11	OLDER	SINGLES/COUPLES	Premium 7.459997
12		RETIREEES	Budget 7.445786
13		RETIREEES	Mainstream 7.269352
14		RETIREEES	Premium 7.461315
15		YOUNG FAMILIES	Budget 7.302705
16		YOUNG FAMILIES	Mainstream 7.226772

17	YOUNG FAMILIES	Premium	7.285951
18	YOUNG SINGLES/COUPLES	Budget	6.663023
19	YOUNG SINGLES/COUPLES	Mainstream	7.551279
20	YOUNG SINGLES/COUPLES	Premium	6.673325

T-Test is performed on the different age categories.

```
[546]: # Perform t-test between mainstream vs premium and budget midage and young
↳singles and couples
mainstream_mask = data['PREMIUM_CUSTOMER'].isin(['Mainstream', 'Budget'])
midage_and_young_mask = data['LIFESTAGE'].isin(['MIDAGE SINGLES/COUPLES',
↳'YOUNG SINGLES/COUPLES'])
ttest_result = ttest_ind(data[mainstream_mask &
↳midage_and_young_mask]['TOT_SALES'],
                        data[~mainstream_mask &
↳midage_and_young_mask]['TOT_SALES'])
```

1.2 Graphical Visualisation of the Data

We create different variables for separate visualisations and keep our original data secure from any changes.

```
[547]: # Create a date range for the full period
start_date = transaction_data['DATE'].min()
end_date = transaction_data['DATE'].max()
date_range = pd.date_range(start_date, end_date)
```

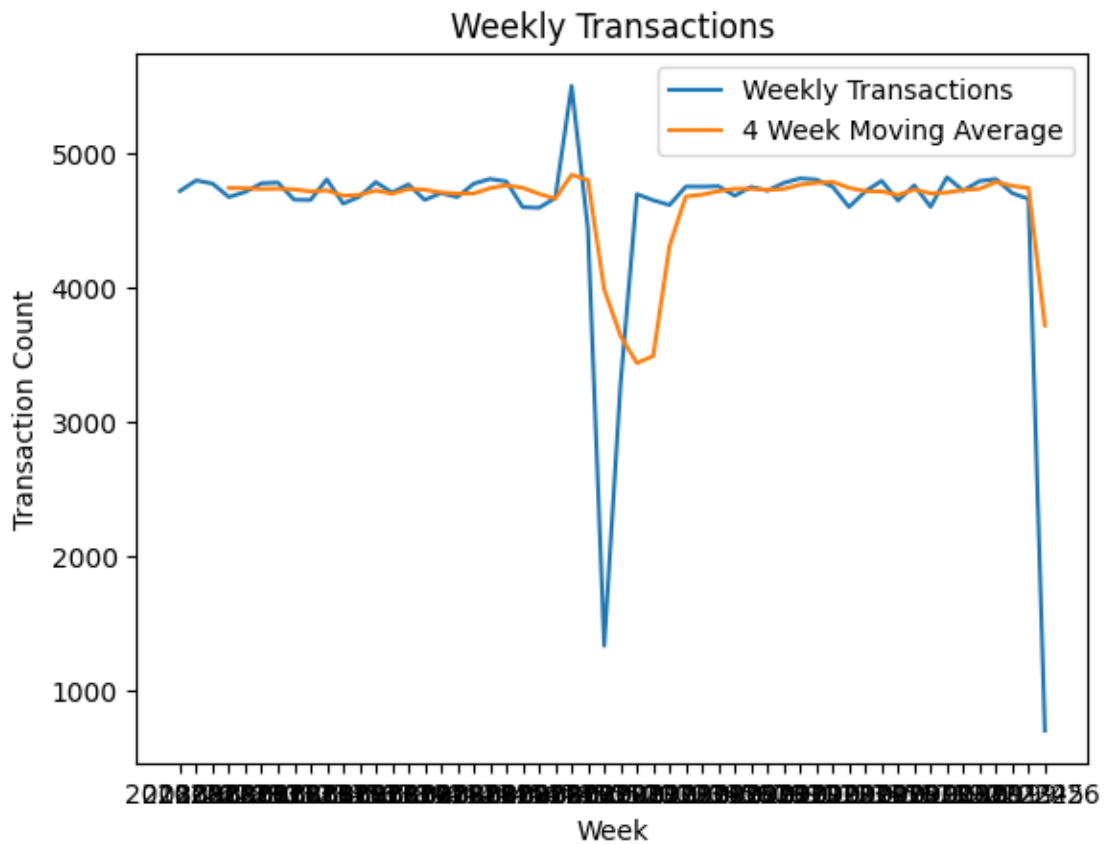
```
[548]: # Analyze transaction trends over time
transaction_data['DATE'] = pd.to_datetime(transaction_data['DATE'],
↳format='%Y-%m-%d')
transactions_by_day = transaction_data.groupby('DATE').size().
↳reset_index(name='TRANSACTION_COUNT').fillna(0)
```

```
[549]: # Resample to weekly sum and calculate 4 week moving average
transactions_by_day['WEEK'] = transactions_by_day['DATE'].dt.strftime('%Y-%U')
transactions_weekly = transactions_by_day.groupby('WEEK')['TRANSACTION_COUNT'].
↳sum().reset_index()
transactions_weekly['4WK_AVG'] = transactions_weekly['TRANSACTION_COUNT'].
↳rolling(4).mean()
```

```
[550]: # Plot weekly transaction count and 4 week moving average
plt.plot(transactions_weekly['WEEK'], transactions_weekly['TRANSACTION_COUNT'],
↳label='Weekly Transactions')
plt.plot(transactions_weekly['WEEK'], transactions_weekly['4WK_AVG'], label='4
↳Week Moving Average')
plt.xlabel('Week'); plt.ylabel('Transaction Count'); plt.title('Weekly
↳Transactions');
```

```
plt.legend()
```

[550]: <matplotlib.legend.Legend at 0x2389b6006d0>



```
[551]: basket_size = transaction_data.groupby(['DATE'])['PROD_QTY'].sum().reset_index()
dates = transaction_data[['DATE']].drop_duplicates()
basket_size = basket_size.merge(dates, how='left', on='DATE')

basket_size.head(50)
```

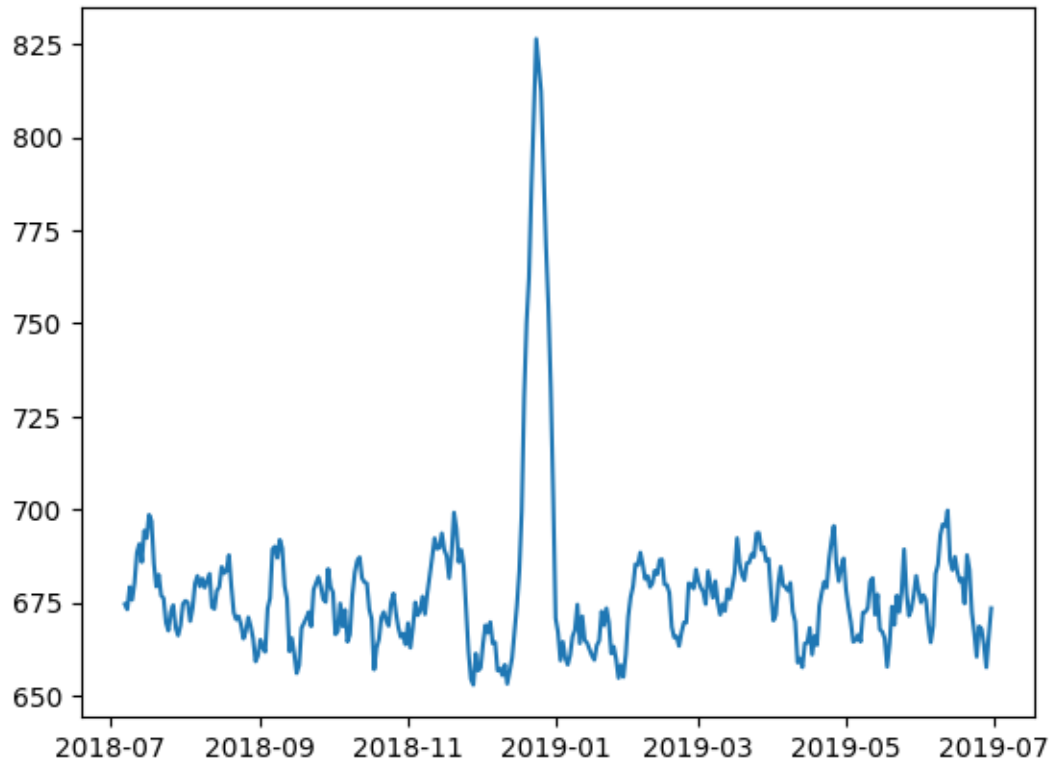
[551]:

	DATE	PROD_QTY
0	2018-07-01	1273
1	2018-07-02	1251
2	2018-07-03	1296
3	2018-07-04	1286
4	2018-07-05	1258
5	2018-07-06	1369
6	2018-07-07	1337
7	2018-07-08	1254
8	2018-07-09	1335

9	2018-07-10	1237
10	2018-07-11	1344
11	2018-07-12	1371
12	2018-07-13	1392
13	2018-07-14	1269
14	2018-07-15	1368
15	2018-07-16	1312
16	2018-07-17	1335
17	2018-07-18	1307
18	2018-07-19	1231
19	2018-07-20	1302
20	2018-07-21	1307
21	2018-07-22	1289
22	2018-07-23	1297
23	2018-07-24	1237
24	2018-07-25	1292
25	2018-07-26	1288
26	2018-07-27	1338
27	2018-07-28	1224
28	2018-07-29	1267
29	2018-07-30	1331
30	2018-07-31	1313
31	2018-08-01	1304
32	2018-08-02	1283
33	2018-08-03	1274
34	2018-08-04	1259
35	2018-08-05	1359
36	2018-08-06	1353
37	2018-08-07	1285
38	2018-08-08	1334
39	2018-08-09	1253
40	2018-08-10	1292
41	2018-08-11	1313
42	2018-08-12	1240
43	2018-08-13	1343
44	2018-08-14	1088
45	2018-08-15	1128
46	2018-08-16	1083
47	2018-08-17	1070
48	2018-08-18	1072
49	2018-08-19	1286

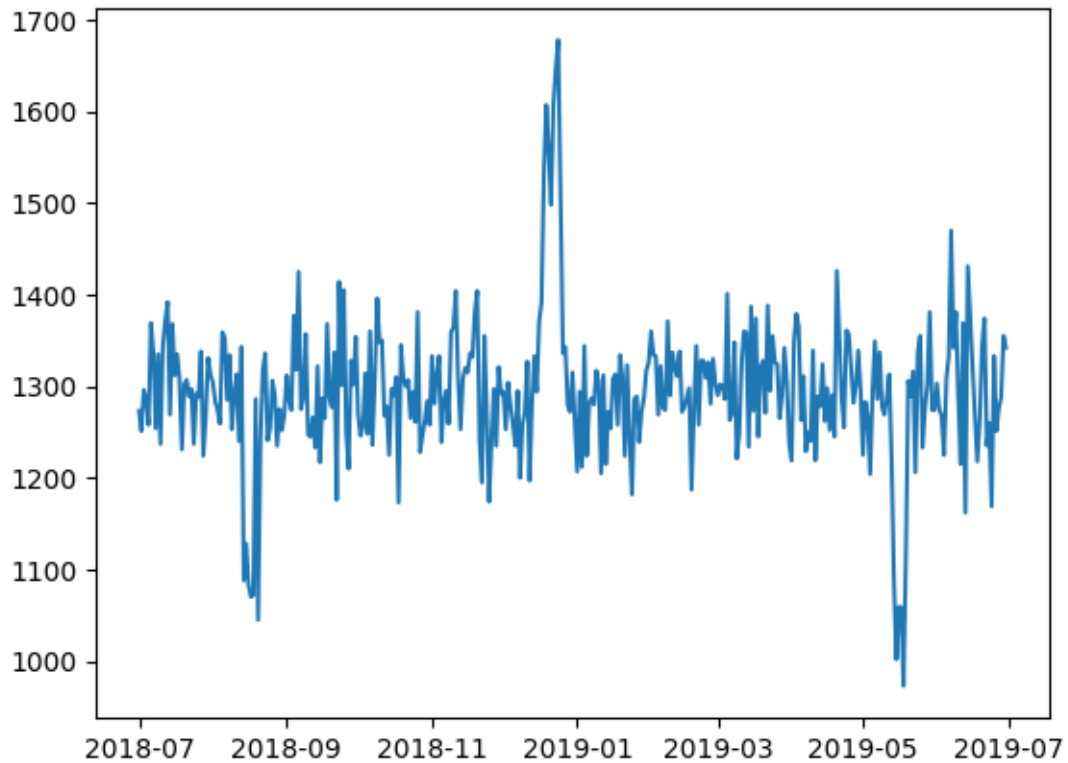
```
[552]: # Rolling averages
transactions_by_day['7D_AVG'] = transactions_by_day['TRANSACTION_COUNT'].
    ↪rolling(7).mean()
plt.plot(transactions_by_day['DATE'], transactions_by_day['7D_AVG'])
```

[552]: [<matplotlib.lines.Line2D at 0x2389b623c90>]

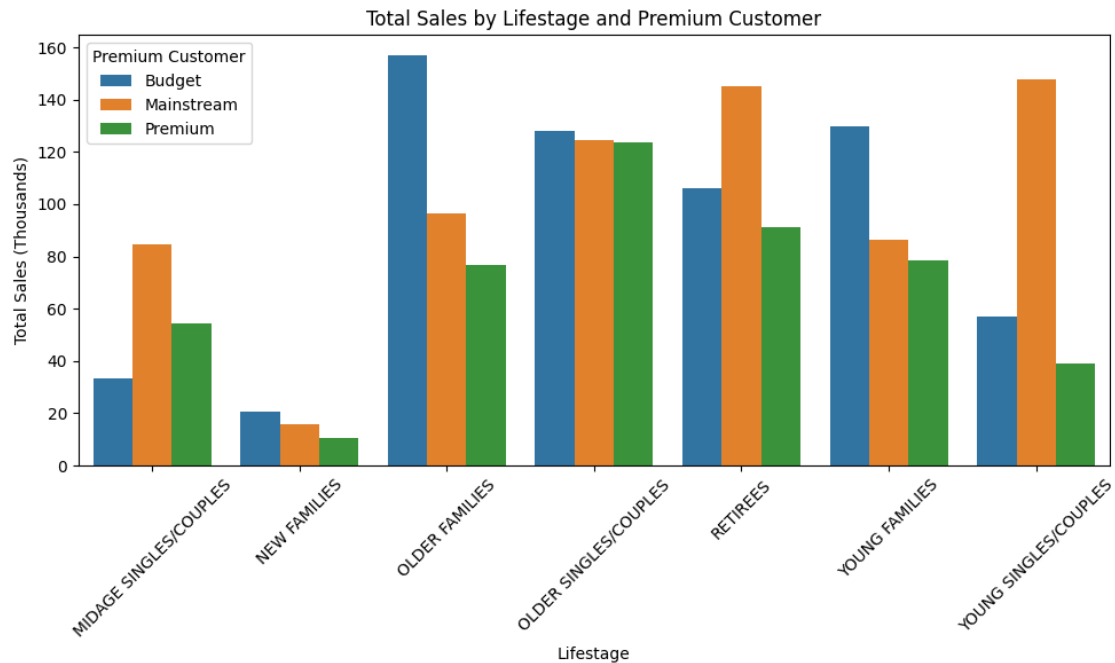


```
[553]: # Basket size
basket_size = transaction_data.groupby('DATE')['PROD_QTY'].sum().reset_index()
plt.plot(basket_size['DATE'], basket_size['PROD_QTY'])
```

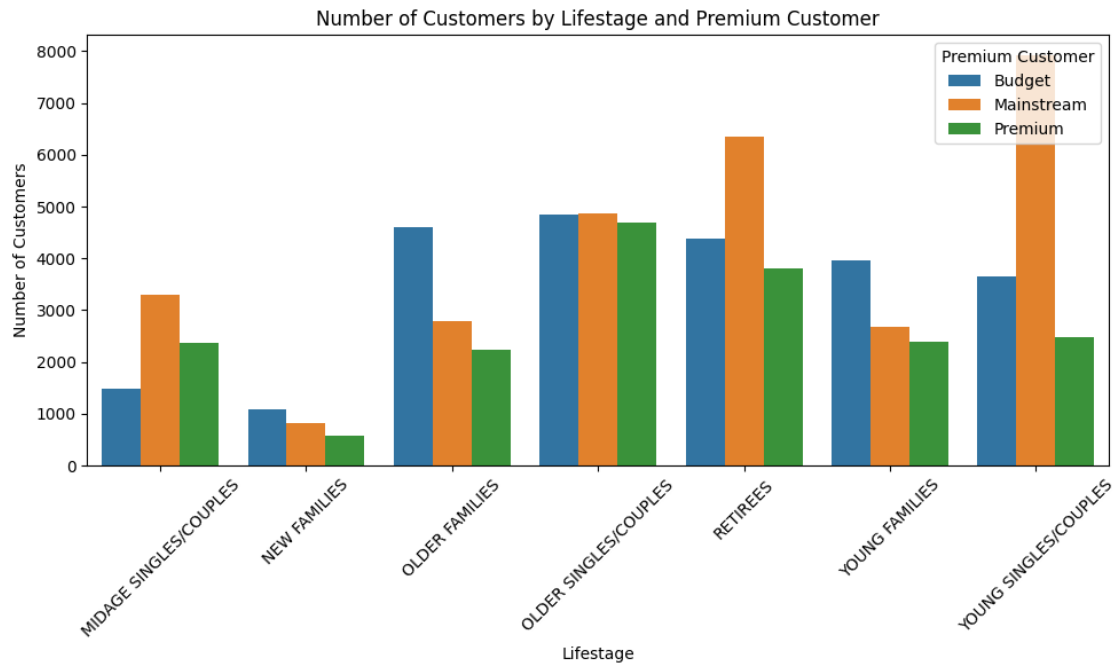
[553]: [<matplotlib.lines.Line2D at 0x2389afa5b50>]



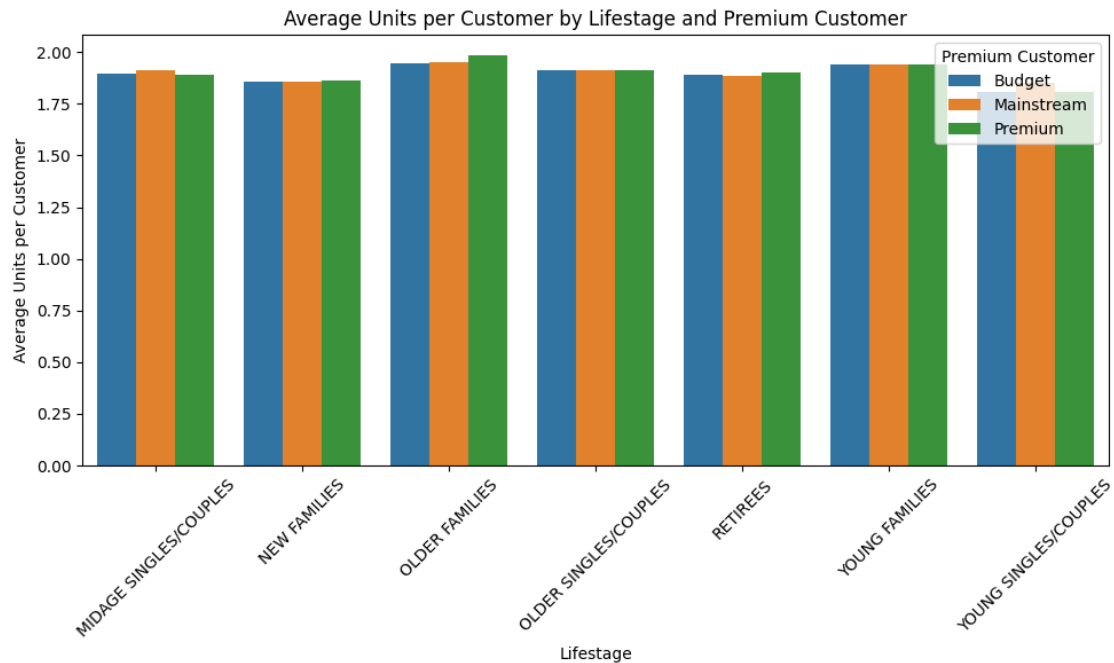
```
[554]: # Plotting total sales by LIFESTAGE and PREMIUM_CUSTOMER
plt.figure(figsize=(10, 6))
total_sales['total_sales'] = total_sales['TOT_SALES'] / 1000 # Convert to
    ↳thousands for better visualization
ax = sns.barplot(x='LIFESTAGE', y='total_sales', hue='PREMIUM_CUSTOMER',
    ↳data=total_sales)
ax.set_xlabel('Lifestage')
ax.set_ylabel('Total Sales (Thousands)')
ax.set_title('Total Sales by Lifestage and Premium Customer')
plt.xticks(rotation=45)
plt.legend(title='Premium Customer', loc='upper left')
plt.tight_layout()
plt.show()
```



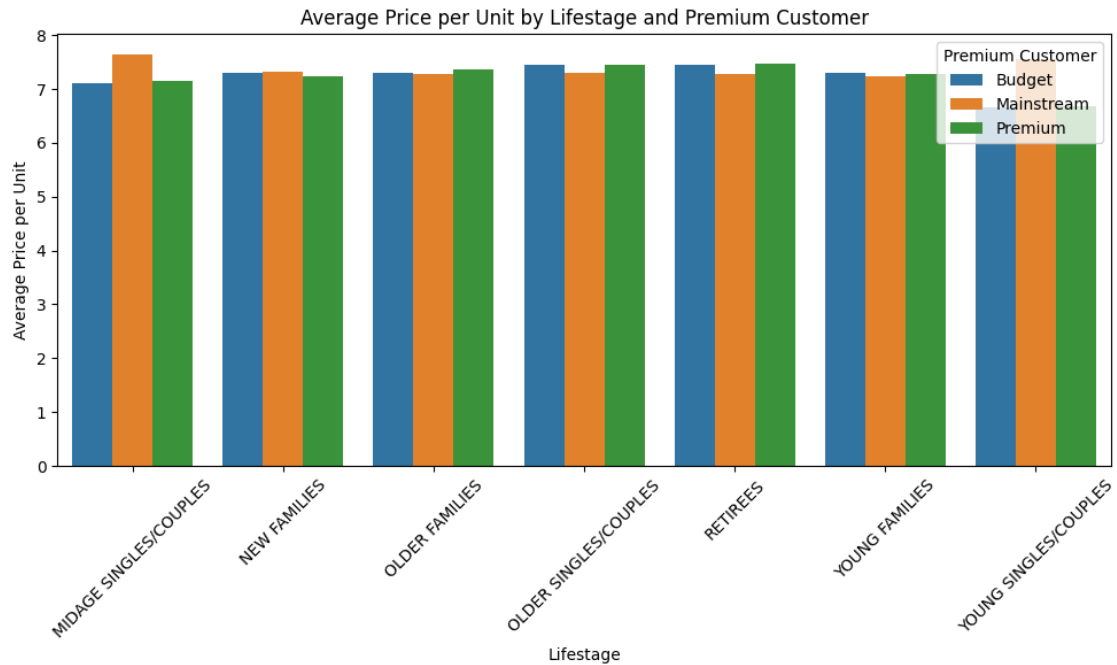
```
[555]: # Plotting number of customers by LIFESTAGE and PREMIUM_CUSTOMER
plt.figure(figsize=(10, 6))
num_customers['num_customers'] = num_customers['LYLTY_CARD_NBR']
ax = sns.barplot(x='LIFESTAGE', y='num_customers', hue='PREMIUM_CUSTOMER', data=num_customers)
ax.set_xlabel('Lifestage')
ax.set_ylabel('Number of Customers')
ax.set_title('Number of Customers by Lifestage and Premium Customer')
plt.xticks(rotation=45)
plt.legend(title='Premium Customer', loc='upper right')
plt.tight_layout()
plt.show()
```

```
[556]: # Plotting average units per customer by LIFESTAGE and PREMIUM_CUSTOMER
plt.figure(figsize=(10, 6))
ax = sns.barplot(x='LIFESTAGE', y='PROD_QTY', hue='PREMIUM_CUSTOMER',
                 data=avg_units_per_customer)
ax.set_xlabel('Lifestage')
ax.set_ylabel('Average Units per Customer')
ax.set_title('Average Units per Customer by Lifestage and Premium Customer')
plt.xticks(rotation=45)
plt.legend(title='Premium Customer', loc='upper right')
plt.tight_layout()
plt.show()
```

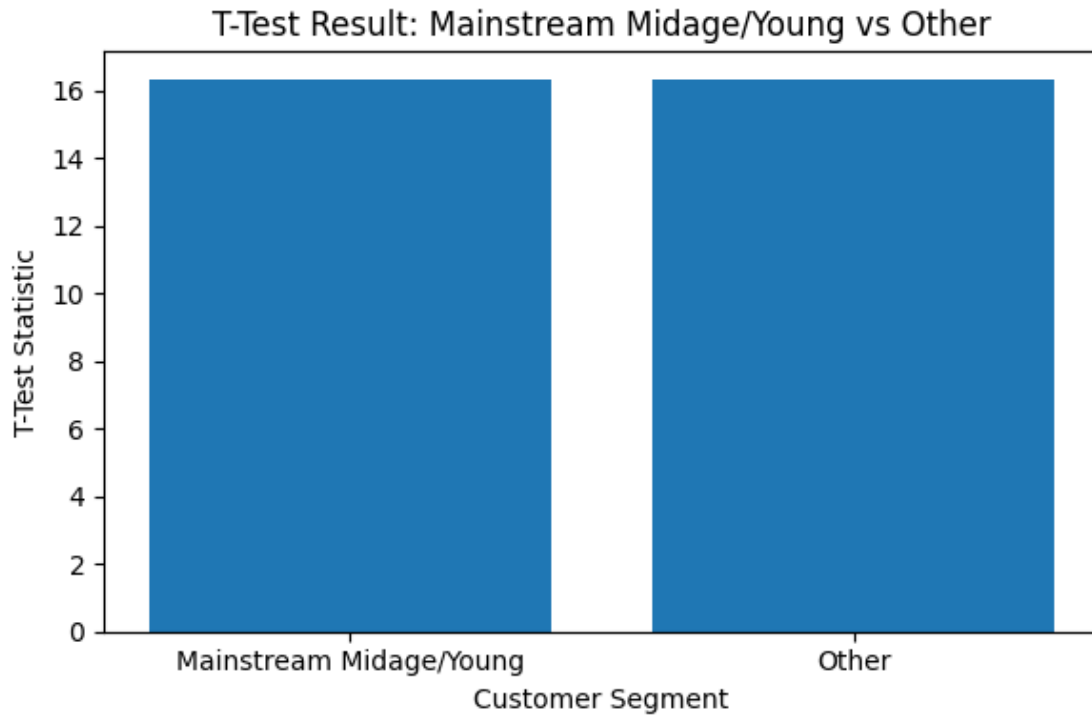


```
[557]: # Plotting average price per unit by LIFESTAGE and PREMIUM_CUSTOMER
plt.figure(figsize=(10, 6))
ax = sns.barplot(x='LIFESTAGE', y='TOT_SALES', hue='PREMIUM_CUSTOMER',
                 data=avg_price_per_unit)
ax.set_xlabel('Lifestage')
ax.set_ylabel('Average Price per Unit')
ax.set_title('Average Price per Unit by Lifestage and Premium Customer')
plt.xticks(rotation=45)
plt.legend(title='Premium Customer', loc='upper right')
plt.tight_layout()
plt.show()
```



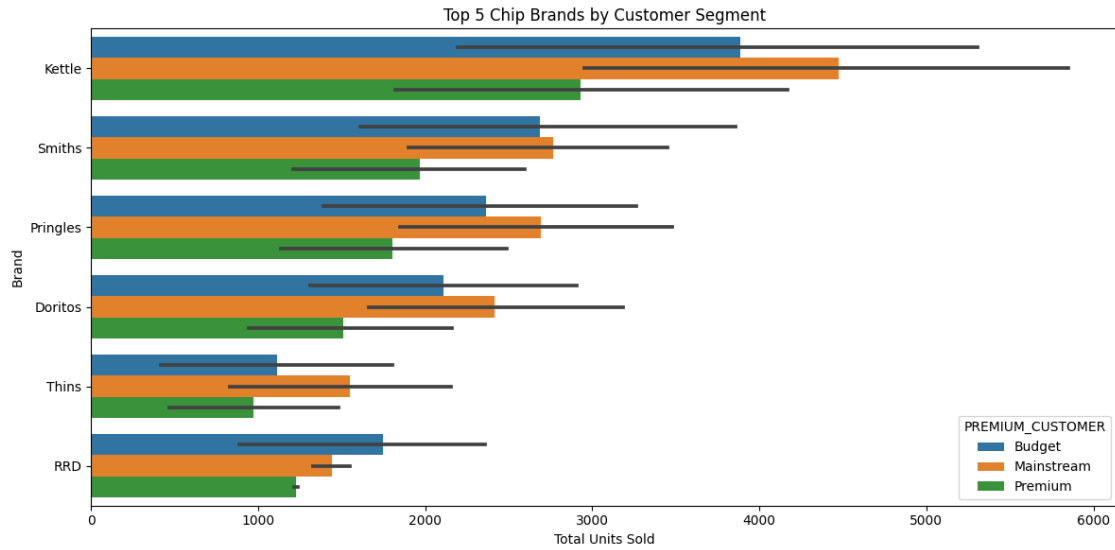
```
[558]: # Extract t-test statistic from the result
ttest_statistic = ttest_result.statistic

# Plotting t-test result
plt.figure(figsize=(6, 4))
plt.bar(['Mainstream Midage/Young', 'Other'], ttest_statistic)
plt.xlabel('Customer Segment')
plt.ylabel('T-Test Statistic')
plt.title('T-Test Result: Mainstream Midage/Young vs Other')
plt.tight_layout()
plt.show()
```

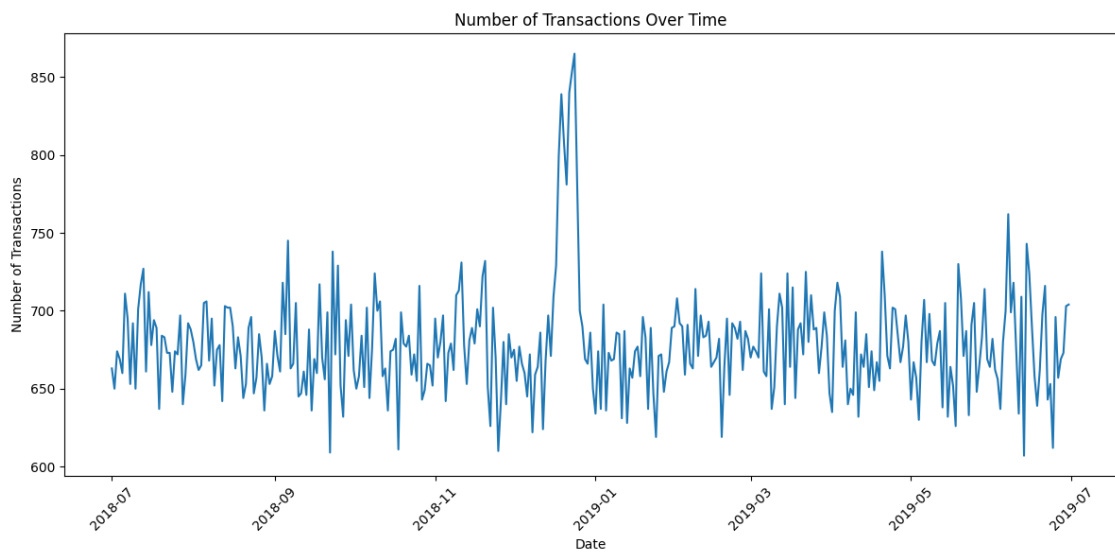


```
[559]: # Calculate the top 5 most popular chip brands by customer segment
top_brands = data.groupby(['LIFESTAGE', 'PREMIUM_CUSTOMER', '
    ↳ 'BRAND'])['PROD_QTY'].sum().reset_index()
top_brands = top_brands.sort_values(['LIFESTAGE', 'PREMIUM_CUSTOMER', '
    ↳ 'PROD_QTY'], ascending=[True, True, False])
top_brands = top_brands.groupby(['LIFESTAGE', 'PREMIUM_CUSTOMER']).head(5)
```

```
[560]: # Plotting top 5 chip brands by customer segment
plt.figure(figsize=(12, 6))
ax = sns.barplot(x='PROD_QTY', y='BRAND', hue='PREMIUM_CUSTOMER', '
    ↳ data=top_brands)
ax.set_xlabel('Total Units Sold')
ax.set_ylabel('Brand')
ax.set_title('Top 5 Chip Brands by Customer Segment')
plt.tight_layout()
plt.show()
```



```
[561]: # Plotting number of transactions over time
plt.figure(figsize=(12, 6))
ax = sns.lineplot(x='DATE', y='TRANSACTION_COUNT', data=transactions_by_day)
ax.set_xlabel('Date')
ax.set_ylabel('Number of Transactions')
ax.set_title('Number of Transactions Over Time')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



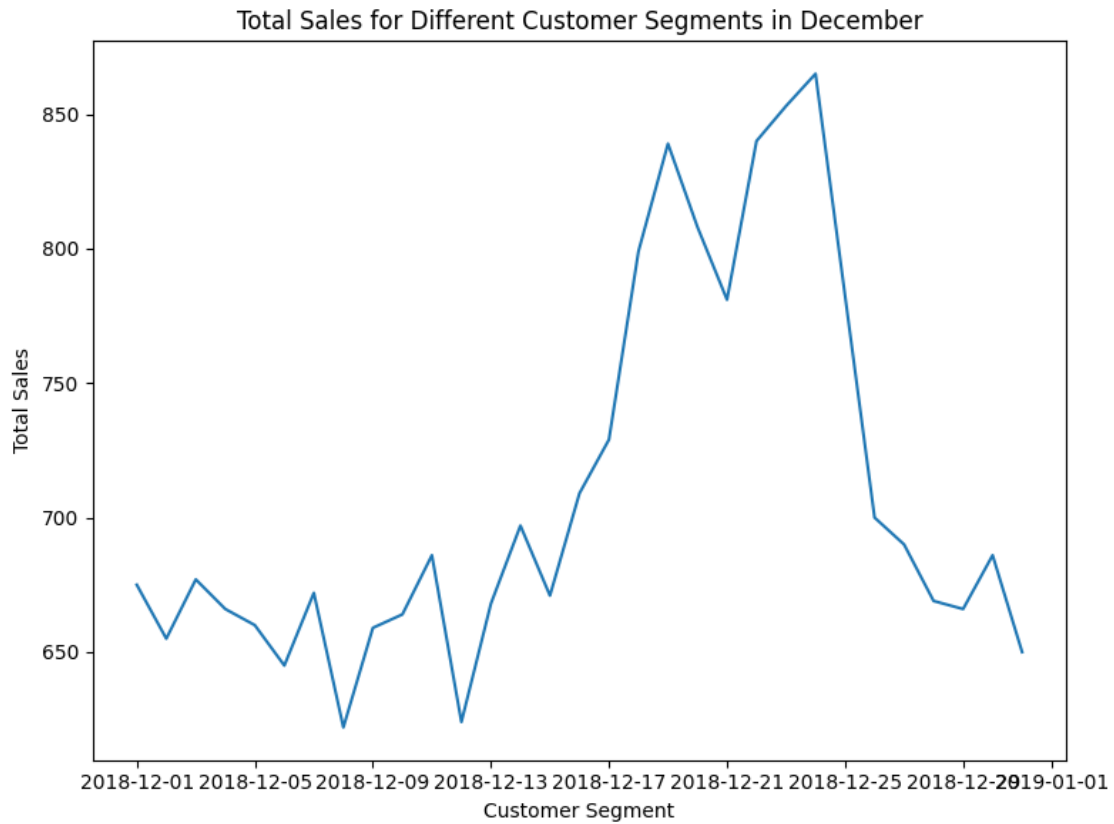
```
[562]: # Filter data for December
december_data = transactions_by_day[transactions_by_day['DATE'].dt.month == 12]
```

```
[563]: december_data.head(50)
```

```
[563]:
```

	DATE	TRANSACTION_COUNT	WEEK	7D_AVG
153	2018-12-01	675	2018-47	657.428571
154	2018-12-02	655	2018-48	663.857143
155	2018-12-03	677	2018-48	668.857143
156	2018-12-04	666	2018-48	666.857143
157	2018-12-05	660	2018-48	669.714286
158	2018-12-06	645	2018-48	664.000000
159	2018-12-07	672	2018-48	664.285714
160	2018-12-08	622	2018-48	656.714286
161	2018-12-09	659	2018-49	657.285714
162	2018-12-10	664	2018-49	655.428571
163	2018-12-11	686	2018-49	658.285714
164	2018-12-12	624	2018-49	653.142857
165	2018-12-13	668	2018-49	656.428571
166	2018-12-14	697	2018-49	660.000000
167	2018-12-15	671	2018-49	667.000000
168	2018-12-16	709	2018-50	674.142857
169	2018-12-17	729	2018-50	683.428571
170	2018-12-18	799	2018-50	699.571429
171	2018-12-19	839	2018-50	730.285714
172	2018-12-20	808	2018-50	750.285714
173	2018-12-21	781	2018-50	762.285714
174	2018-12-22	840	2018-50	786.428571
175	2018-12-23	853	2018-51	807.000000
176	2018-12-24	865	2018-51	826.428571
177	2018-12-26	700	2018-51	812.285714
178	2018-12-27	690	2018-51	791.000000
179	2018-12-28	669	2018-51	771.142857
180	2018-12-29	666	2018-51	754.714286
181	2018-12-30	686	2018-52	732.714286
182	2018-12-31	650	2018-52	703.714286

```
[564]: # Plotting the total sales for different customer segments in December
plt.figure(figsize=(8, 6))
ax = sns.lineplot(x='DATE', y='TRANSACTION_COUNT', data=december_data)
ax.set_xlabel('Customer Segment')
ax.set_ylabel('Total Sales')
ax.set_title('Total Sales for Different Customer Segments in December')
plt.tight_layout()
plt.show()
```



In the end, we have the final result and have successfully completed the task. I hope that it is easily understandable and I have tried to explain it as much as possible. Thank you for reading this notebook and I hope you have a great day!