Exploratory Data Analysis of 2Market global

Problem Statement

2Market is a global supermarket chain offering a wide range of consumer products both online and in physical stores. They would like to identify patterns in customer purchase behaviour that will help inform marketing decisions.

How am I helping the company:

Performing an Exploratory data analysis to gain insights into their customer demographic.

Questions of interest as a Data analyst

- What is the data source? To find out how authentic the data is?
- Are there any data limitations? Any data insufficiencies or biases in data collection.
- Who are the audience? Technical or non-technical?

Questions to ask the team as a Data Analyst

- Any specific challenges that you would like deeper understanding and insights?
- How do you measure success? Customer segmentation or cost reduction?
- Are the primary audience for the presentation; technical or non-technical audience?

Questions regarding data

- Time period of the data; any significant events or campaigns that could have influenced customer behavior during that period?
- Are there any seasonal trends observed in the purchasing patterns in the online grocery market?

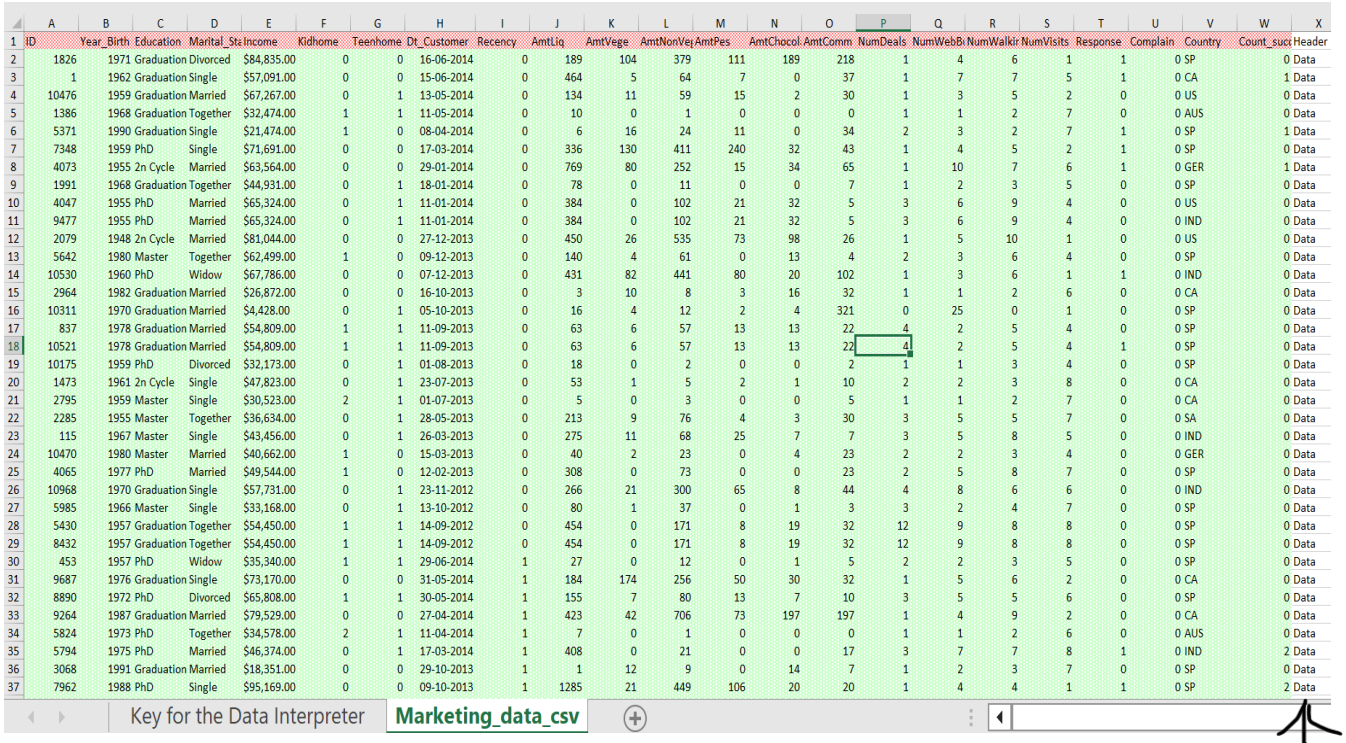
I like to perform basic statistics using Excel to get a quick glance at the aggregates. And for further Exploratory data analysis, Postgres SQL to retrieve the relevant data using Pg Admin and export to Excel files and then import these on Tableau Public to visualise the findings.

Data cleaning:

The data is provided in the raw form so my first step is to clean the data to clean it to ensure optimal performance in data analysis with reliable and refined datasets.

I used Tableau desktop version to clean the data, as it is a very time efficient tool, which transforms the data into a cleaned one in matter of seconds.

Cleaned data using Tableau



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	ID	Year_Birth	Education	Marital_Sta	Income	Kidhome	Teenhome	Dt_Customer	Recency	AmtLiq	AmtVege	AmtNonVe	AmtPes	AmtChocol	AmtComm	NumDeals	NumWebB	NumWalkI	NumVisits	Response	Complain	Country	Count_succ	Header
2	1826	1971	Graduation	Divorced	$84,835.00	0	0	16-06-2014	0	189	104	379	111	189	218	1	4	6	1	1	0	SP	0	Data
3	1	1962	Graduation	Single	$57,091.00	0	0	15-06-2014	0	464	5	64	7	0	37	1	7	7	5	1	0	CA	1	Data
4	10476	1959	Graduation	Married	$67,267.00	0	1	13-05-2014	0	134	11	59	15	2	30	1	3	5	2	0	0	US	0	Data
5	1386	1968	Graduation	Together	$32,474.00	1	1	11-05-2014	0	10	0	1	0	0	0	1	1	2	7	0	0	AUS	0	Data
6	5371	1990	Graduation	Single	$21,474.00	1	0	08-04-2014	0	6	16	24	11	0	34	2	3	2	7	1	0	SP	1	Data
7	7348	1959	PhD	Single	$71,691.00	0	0	17-03-2014	0	336	130	411	240	32	43	1	4	5	2	1	0	SP	0	Data
8	4073	1955	2n Cycle	Married	$63,564.00	0	0	29-01-2014	0	769	80	252	15	34	65	1	10	7	6	1	0	GER	1	Data
9	1991	1968	Graduation	Together	$44,931.00	0	1	18-01-2014	0	78	0	11	0	0	7	1	2	3	5	0	0	SP	0	Data
10	4047	1955	PhD	Married	$65,324.00	0	1	11-01-2014	0	384	0	102	21	32	5	3	6	9	4	0	0	US	0	Data
11	9477	1955	PhD	Married	$65,324.00	0	1	11-01-2014	0	384	0	102	21	32	5	3	6	9	4	0	0	IND	0	Data
12	2079	1948	2n Cycle	Married	$81,044.00	0	0	27-12-2013	0	450	26	535	73	98	26	1	5	10	1	0	0	US	0	Data
13	5642	1980	Master	Together	$62,499.00	1	0	09-12-2013	0	140	4	61	0	13	4	2	3	6	4	0	0	SP	0	Data
14	10530	1960	PhD	Widow	$67,786.00	0	0	07-12-2013	0	431	82	441	80	20	102	1	3	6	1	1	0	IND	0	Data
15	2964	1982	Graduation	Married	$26,872.00	0	0	16-10-2013	0	3	10	8	3	16	32	1	1	2	6	0	0	CA	0	Data
16	10311	1970	Graduation	Married	$4,428.00	0	1	05-10-2013	0	16	4	12	2	4	321	0	25	0	1	0	0	SP	0	Data
17	837	1978	Graduation	Married	$54,809.00	1	1	11-09-2013	0	63	6	57	13	13	22	4	2	5	4	0	0	SP	0	Data
18	10521	1978	Graduation	Married	$54,809.00	1	1	11-09-2013	0	63	6	57	13	13	22	4	2	5	4	1	0	SP	0	Data
19	10175	1959	PhD	Divorced	$32,173.00	0	1	01-08-2013	0	18	0	2	0	0	2	1	1	3	4	0	0	SP	0	Data
20	1473	1961	2n Cycle	Single	$47,823.00	0	1	23-07-2013	0	53	1	5	2	1	10	2	2	3	8	0	0	CA	0	Data
21	2795	1959	Master	Single	$30,523.00	2	1	01-07-2013	0	5	0	3	0	0	5	1	1	2	7	0	0	CA	0	Data
22	2285	1955	Master	Together	$36,634.00	0	1	28-05-2013	0	213	9	76	4	3	30	3	5	5	7	0	0	SA	0	Data
23	115	1967	Master	Single	$43,456.00	0	1	26-03-2013	0	275	11	68	25	7	7	3	5	8	5	0	0	IND	0	Data
24	10470	1980	Master	Married	$40,662.00	1	0	15-03-2013	0	40	2	23	0	4	23	2	2	3	4	0	0	GER	0	Data
25	4065	1977	PhD	Married	$49,544.00	1	0	12-02-2013	0	308	0	73	0	0	23	2	5	8	7	0	0	SP	0	Data
26	10968	1970	Graduation	Single	$57,731.00	0	1	23-11-2012	0	266	21	300	65	8	44	4	8	6	6	0	0	IND	0	Data
27	5985	1966	Master	Single	$33,168.00	0	1	13-10-2012	0	80	1	37	0	1	3	3	2	4	7	0	0	SP	0	Data
28	5430	1957	Graduation	Together	$54,450.00	1	1	14-09-2012	0	454	0	171	8	19	32	12	9	8	8	0	0	SP	0	Data
29	8432	1957	Graduation	Together	$54,450.00	1	1	14-09-2012	0	454	0	171	8	19	32	12	9	8	8	0	0	SP	0	Data
30	453	1957	PhD	Widow	$35,340.00	1	1	29-06-2014	1	27	0	12	0	1	5	2	2	3	5	0	0	SP	0	Data
31	9687	1976	Graduation	Single	$73,170.00	0	0	31-05-2014	1	184	174	256	50	30	32	1	5	6	2	0	0	CA	0	Data
32	8890	1972	PhD	Divorced	$65,808.00	1	1	30-05-2014	1	155	7	80	13	7	10	3	5	5	6	0	0	SP	0	Data
33	9264	1987	Graduation	Married	$79,529.00	0	0	27-04-2014	1	423	42	706	73	197	197	1	4	9	2	0	0	CA	0	Data
34	5824	1973	PhD	Together	$34,578.00	2	1	11-04-2014	1	7	0	1	0	0	0	1	1	2	6	0	0	AUS	0	Data
35	5794	1975	PhD	Married	$46,374.00	0	1	17-03-2014	1	408	0	21	0	0	17	3	7	7	8	1	0	IND	2	Data
36	3068	1991	Graduation	Married	$18,351.00	0	0	29-10-2013	1	1	12	9	0	14	7	1	2	3	7	0	0	SP	0	Data
37	7962	1988	PhD	Single	$95,169.00	0	0	09-10-2013	1	1285	21	449	106	20	20	1	4	4	1	1	0	SP	2	Data

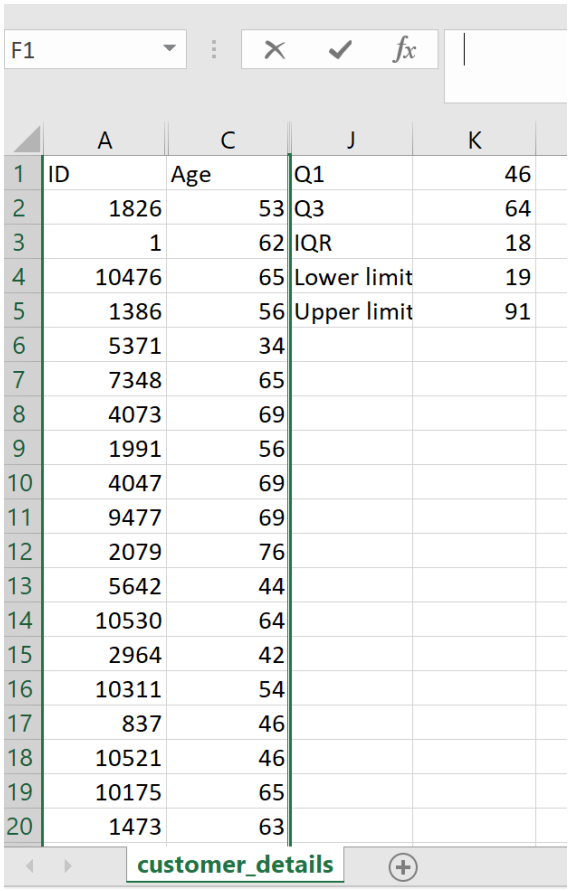
Key for the Data Interpreter | Marketing_data_csv

It adds an extra column, which I deleted as shown as X, which I deleted.

Checking for Outliers:

To start with, I have hidden the columns I don't need for this calculation for convenience.

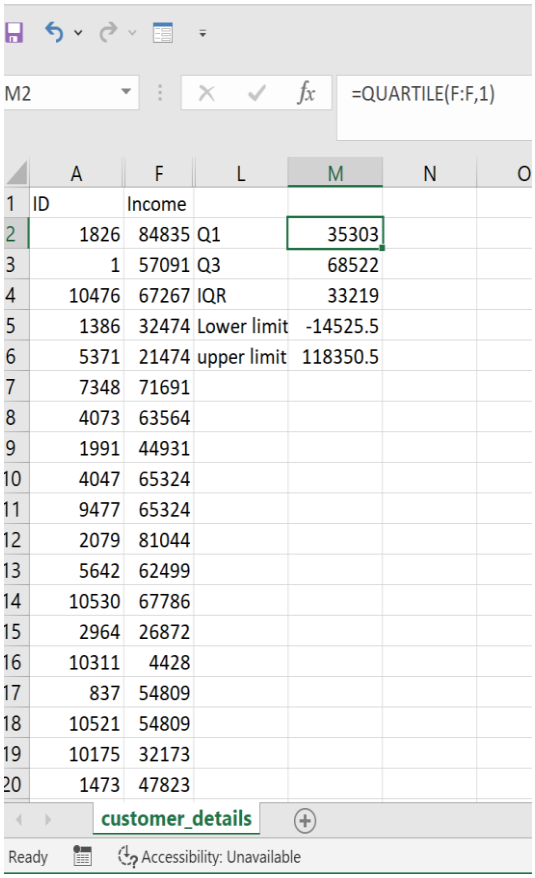
Outliers are the values below the Lower limit and above the upper limit.



F1

	A	C	J	K
1	ID	Age	Q1	46
2	1826	53	Q3	64
3	1	62	IQR	18
4	10476	65	Lower limit	19
5	1386	56	Upper limit	91
6	5371	34		
7	7348	65		
8	4073	69		
9	1991	56		
10	4047	69		
11	9477	69		
12	2079	76		
13	5642	44		
14	10530	64		
15	2964	42		
16	10311	54		
17	837	46		
18	10521	46		
19	10175	65		
20	1473	63		

customer_details



M2 =QUARTILE(F:F,1)

	A	F	L	M	N	O
1	ID	Income				
2	1826	84835	Q1	35303		
3	1	57091	Q3	68522		
4	10476	67267	IQR	33219		
5	1386	32474	Lower limit	-14525.5		
6	5371	21474	upper limit	118350.5		
7	7348	71691				
8	4073	63564				
9	1991	44931				
10	4047	65324				
11	9477	65324				
12	2079	81044				
13	5642	62499				
14	10530	67786				
15	2964	26872				
16	10311	4428				
17	837	54809				
18	10521	54809				
19	10175	32173				
20	1473	47823				

customer_details

Ready Accessibility: Unavailable

Income outlier= < 14525 >118350 the negative sign has no practical meaning for the – sign.

Age outlier = <19>91.

Q1=Quartile(C:C,1) for Income=Quartile(F:F,1)

Q3=Quartile(C:C,2)

Q3-Q1= IQR, Inter quartile range

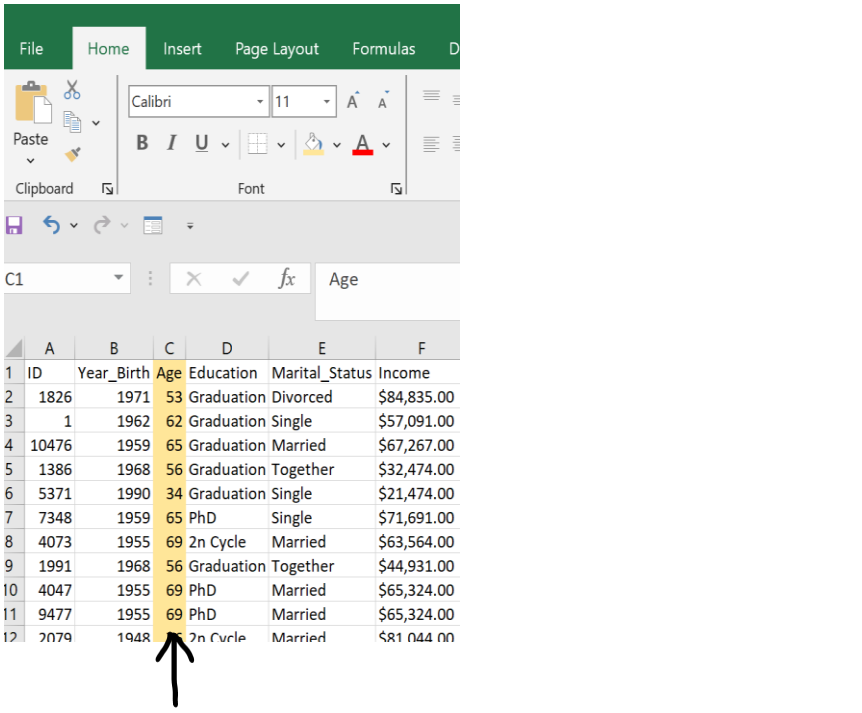
1.5 is a constant commonly used because it captures most outliers.

Questions to understand customer demographic and observe the trends to assist 2Market global in targeted marketing.

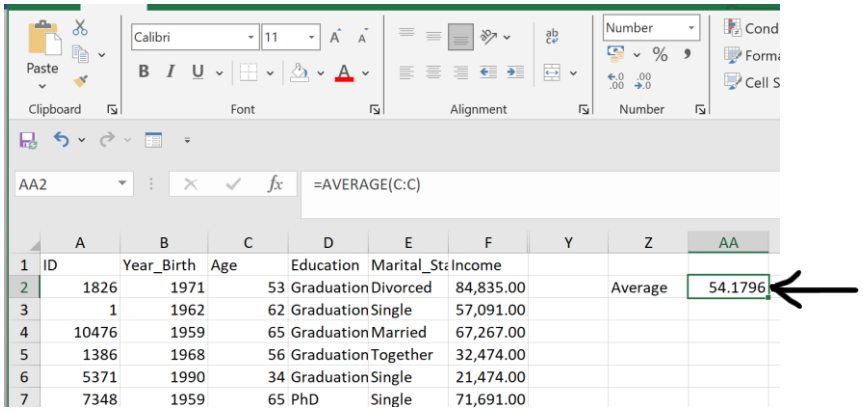
1: What is the average age of 2Market's customers?

Answer: 54

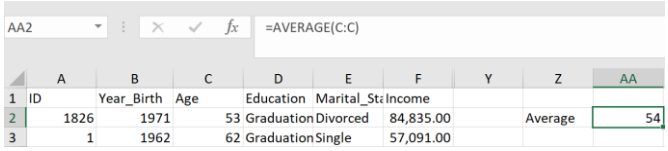
I added an extra column next to Year_Birth, used the formula = YEAR (TODAY ()) - B2, to derive the age. The result has been rounded off to 0 decimals, using the number group.



	A	B	C	D	E	F
1	ID	Year_Birth	Age	Education	Marital_Status	Income
2	1826	1971	53	Graduation	Divorced	$84,835.00
3	1	1962	62	Graduation	Single	$57,091.00
4	10476	1959	65	Graduation	Married	$67,267.00
5	1386	1968	56	Graduation	Together	$32,474.00
6	5371	1990	34	Graduation	Single	$21,474.00
7	7348	1959	65	PhD	Single	$71,691.00
8	4073	1955	69	2n Cycle	Married	$63,564.00
9	1991	1968	56	Graduation	Together	$44,931.00
10	4047	1955	69	PhD	Married	$65,324.00
11	9477	1955	69	PhD	Married	$65,324.00
12	2079	1948		2n Cycle	Married	$81,044.00



	A	B	C	D	E	F	Y	Z	AA
1	ID	Year_Birth	Age	Education	Marital_Sta	Income			
2	1826	1971	53	Graduation	Divorced	84,835.00		Average	54.1796
3	1	1962	62	Graduation	Single	57,091.00			
4	10476	1959	65	Graduation	Married	67,267.00			
5	1386	1968	56	Graduation	Together	32,474.00			
6	5371	1990	34	Graduation	Single	21,474.00			
7	7348	1959	65	PhD	Single	71,691.00			



	A	B	C	D	E	F	Y	Z	AA
1	ID	Year_Birth	Age	Education	Marital_Sta	Income			
2	1826	1971	53	Graduation	Divorced	84,835.00		Average	54
3	1	1962	62	Graduation	Single	57,091.00			

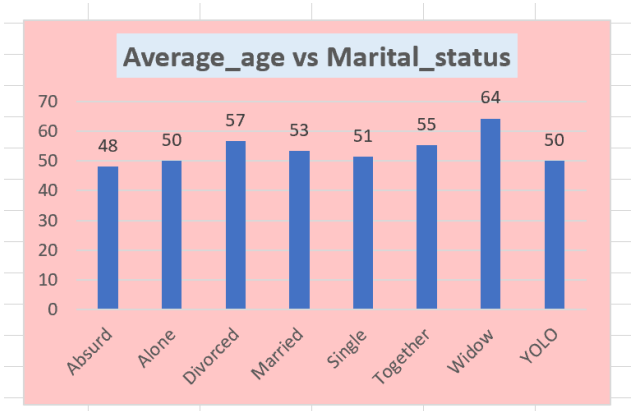
I used the formula =Average(C:C) in AA2 and adjusted it to no decimals using the Number group and removed the $sign to convert into number.

2.What is the average age of the customers belonging to each type of marital status?

Solution: I used the Pivot Table to be a very useful to represent these aggregate values. And used the Bar charts to represent.

Graphical representation

	A	B
1		
2		
3	**Row Labels**	**Average of Age**
4	Divorced	57
5	Married	53
6	Single	51
7	Together	55
8	Widow	64
9	**Grand Total**	**54**
10		



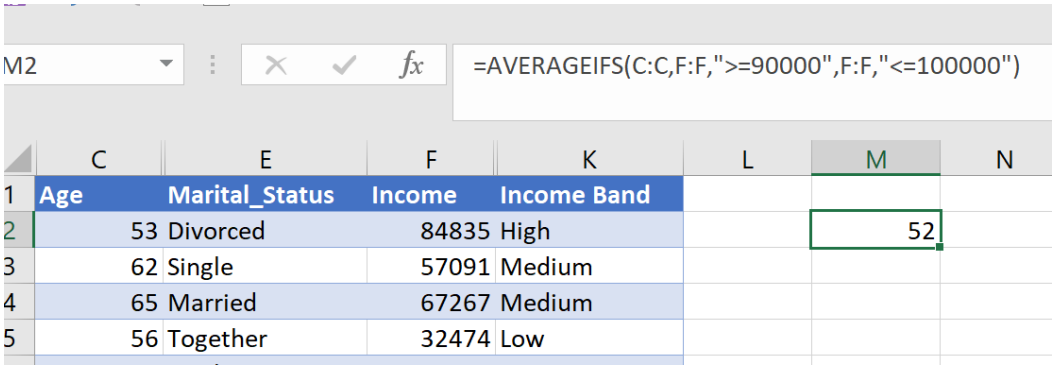
3. **What is the highest average age by marital status?**
   Answer = 64, Widow.

The Bar charts clearly show this.

4.What is the average age of customers who earn a yearly income between US$90,000 and US$100,000?

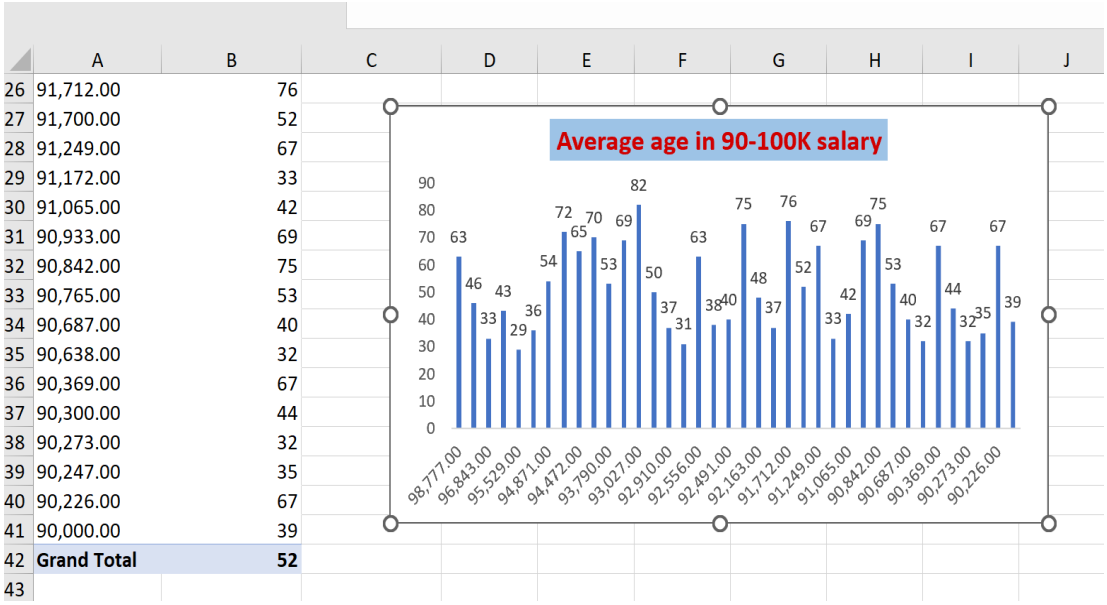
Answer= 52

Solution: I used the AverageIFS function as shown below, C:C is the Age and F:F is the income column, in a new cell with the specified range and pressed enter=rounded off using the Number group.



	C	E	F	K	L	M	N
1	Age	Marital_Status	Income	Income Band			
2	53	Divorced	84835	High		52	
3	62	Single	57091	Medium			
4	65	Married	67267	Medium			
5	56	Together	32474	Low			

5.What is the relationship between Average age and income of customers who earn between US$90000 and US$10000



	A	B
26	91,712.00	76
27	91,700.00	52
28	91,249.00	67
29	91,172.00	33
30	91,065.00	42
31	90,933.00	69
32	90,842.00	75
33	90,765.00	53
34	90,687.00	40
35	90,638.00	32
36	90,369.00	67
37	90,300.00	44
38	90,273.00	32
39	90,247.00	35
40	90,226.00	67
41	90,000.00	39
42	**Grand Total**	**52**
43		

6. **What is the average age across the various income bands?**

High = 52
Medium =56
Low =50

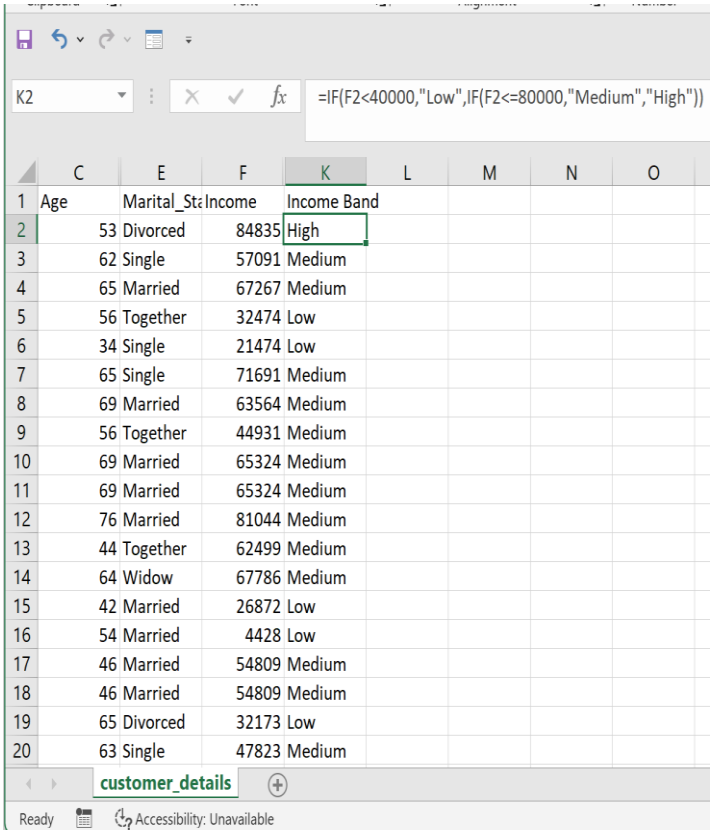
Created a new column, Income Band and used the

formula: =IF(C2<40000, "Low", IF (C2<=80000, "Medium", "High"))

Created a Pivot table, added Income Band in the 'Rows' and Age in the 'Values' field,

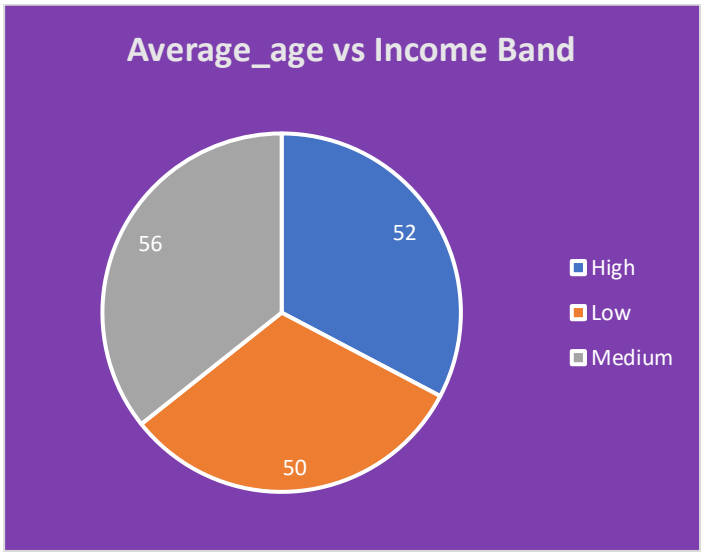
Row Labels	Average of Age
High	52
Low	50
Medium	56
Grand Total	**54**

Changed the values from default Sum to Average and rounded it off.



	C	E	F	K	L	M	N	O
1	Age	Marital_St	Income	Income Band				
2	53	Divorced	84835	High				
3	62	Single	57091	Medium				
4	65	Married	67267	Medium				
5	56	Together	32474	Low				
6	34	Single	21474	Low				
7	65	Single	71691	Medium				
8	69	Married	63564	Medium				
9	56	Together	44931	Medium				
10	69	Married	65324	Medium				
11	69	Married	65324	Medium				
12	76	Married	81044	Medium				
13	44	Together	62499	Medium				
14	64	Widow	67786	Medium				
15	42	Married	26872	Low				
16	54	Married	4428	Low				
17	46	Married	54809	Medium				
18	46	Married	54809	Medium				
19	65	Divorced	32173	Low				
20	63	Single	47823	Medium				

Graphical representation



Average age of all the 3 income bands is almost the same.

Let us answer some business questions that will help 2Market global to gain deeper insights.

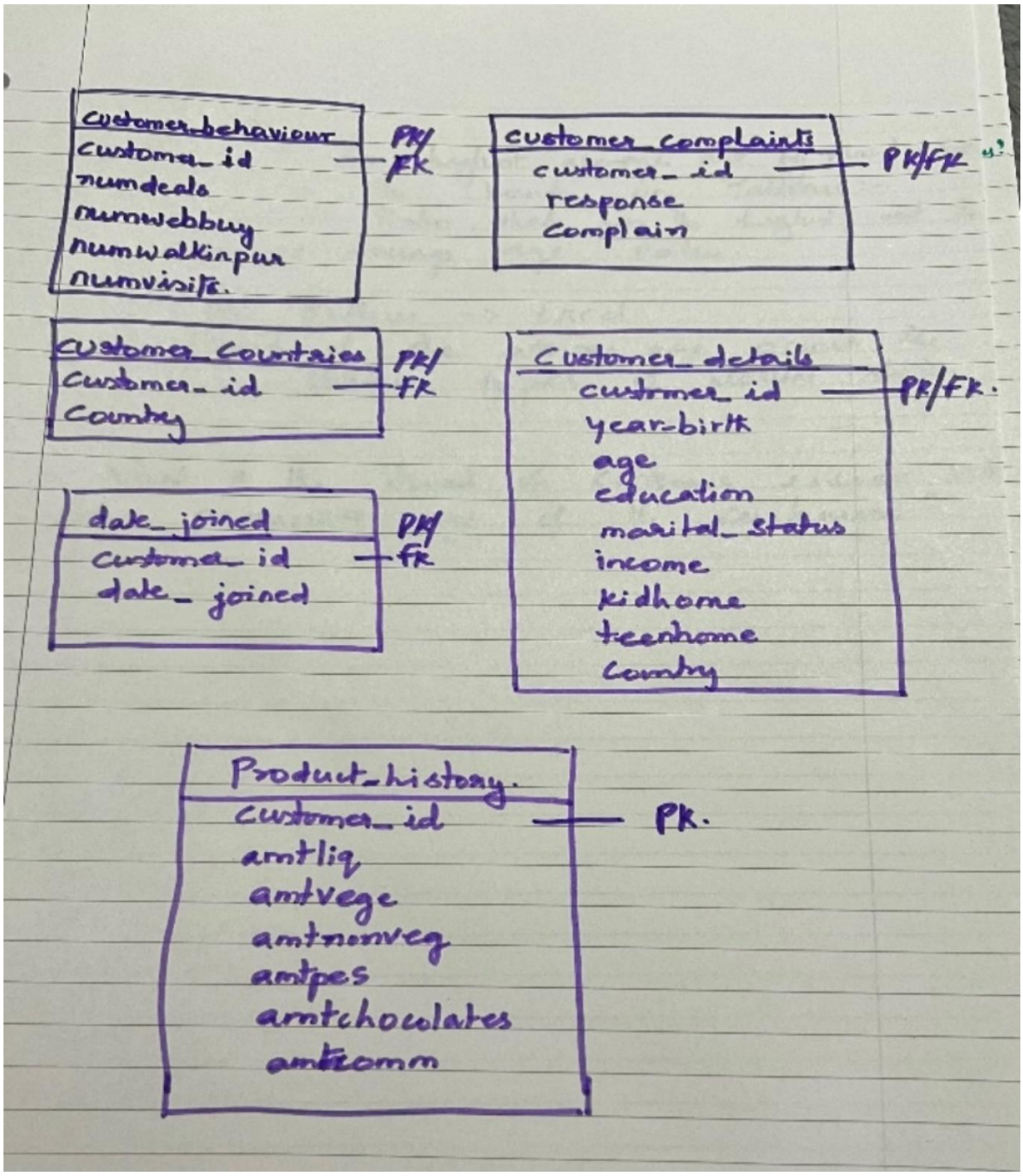
- the demographics of its customers
- which advertising channels seem to be the most effective
- which products seem to sell the best and whether that varies based on demographic.

I used SQL on Postgres, to calculate the aggregates and downloaded the output as excel files, so that I can import on Tableau to visualise and present the dashboard of the Exploratory data analysis to the stakeholders.

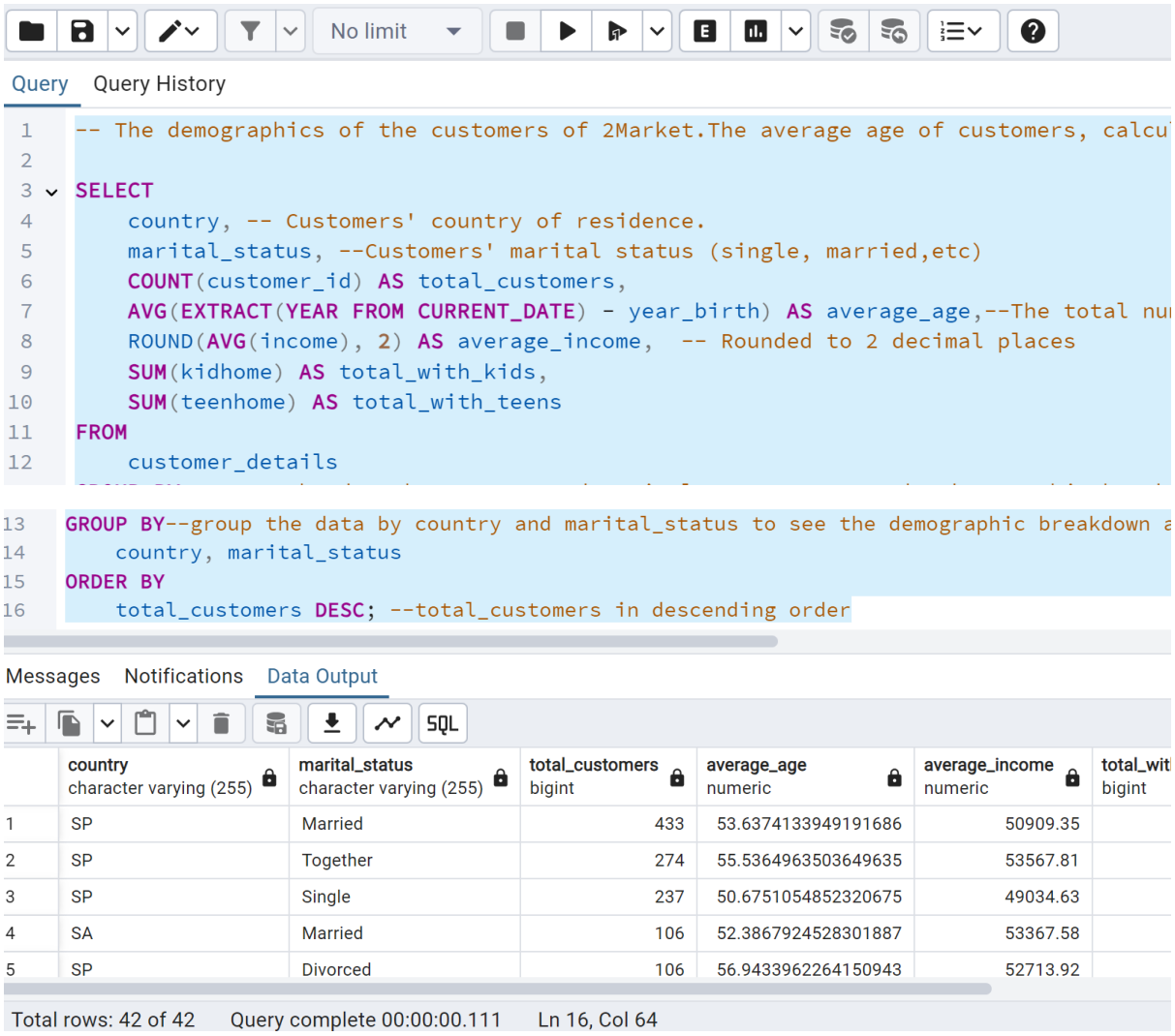
SQL can handle large datasets with high data integrity.

Visualisation:

The 2Market global file is a large data set, so I created 6 separate tables with a logical grouping and relating all the tables with the primary key, customer_id, in order to be helpful in using joins to retrieve data from multiple tables.



1. The demographics of their customers:



No limit

Query | Query History

```
1  -- The demographics of the customers of 2Market.The average age of customers, calcu
2
3  SELECT
4      country, -- Customers' country of residence.
5      marital_status, --Customers' marital status (single, married,etc)
6      COUNT(customer_id) AS total_customers,
7      AVG(EXTRACT(YEAR FROM CURRENT_DATE) - year_birth) AS average_age,--The total nu
8      ROUND(AVG(income), 2) AS average_income,  -- Rounded to 2 decimal places
9      SUM(kidhome) AS total_with_kids,
10     SUM(teenhome) AS total_with_teens
11 FROM
12     customer_details
13 GROUP BY--group the data by country and marital_status to see the demographic breakdown a
14     country, marital_status
15 ORDER BY
16     total_customers DESC; --total_customers in descending order
```

Messages | Notifications | Data Output

SQL

	country character varying (255)	marital_status character varying (255)	total_customers bigint	average_age numeric	average_income numeric	total_wit bigint
1	SP	Married	433	53.6374133949191686	50909.35	
2	SP	Together	274	55.5364963503649635	53567.81	
3	SP	Single	237	50.6751054852320675	49034.63	
4	SA	Married	106	52.3867924528301887	53367.58	
5	SP	Divorced	106	56.9433962264150943	52713.92	

Total rows: 42 of 42 Query complete 00:00:00.111 Ln 16, Col 64

2. Effectiveness of Advertising Channels:

To know this, we need to assess the relationship between advertising and customer purchases.

Query | Query History

```sql
/* Which advertising channels seem to be the most effective: advertising channels are most
let's group the data by country, marital_status, kidhome, and teenhome to see how each demo
 We can calculate the total spend on all product types (liquids, vegetables, non-vegetables
*/
SELECT
    cd.country,
    cd.marital_status,
    cd.kidhome,
    cd.teenhome,
    SUM(ph.amtliq + ph.amtvege + ph.amtnonveg + ph.amtpes + ph.amtchocolates + ph.amtcomm)
    SUM(cb.numwebbuy) AS web_conversions,
    SUM(cb.numwalkinpur) AS walkin_conversions,
    (SUM(cb.numwebbuy) + SUM(cb.numwalkinpur)) AS total_conversions --combine both
FROM
    product_history ph
JOIN
    customer_details cd ON ph.customer_id = cd.customer_id
JOIN
    customer_behaviour cb ON ph.customer_id = cb.customer_id
GROUP BY
    cd.country, cd.marital_status, cd.kidhome, cd.teenhome
ORDER BY
    total_spent DESC;
```

Messages | Notifications | Data Output

	country character varying (255)	marital_status character varying (255)	kidhome integer	teenhome integer	total_spent numeric	web_conversions bigint	walkin_conver bigint
1	SP	Married	0	0	112561.00	462	
2	SP	Together	0	0	94071.00	326	
3	SP	Married	0	1	91017.00	733	
4	SP	Single	0	0	87023.00	345	
5	SP	Together	0	1	62015.00	451	

Total rows: 193 of 193 Query complete 00:00:00.198 Ln 4, Col 3

Product demographic to understand how different customer segments interact with and purchase specific products

1. **What is the total spending per country?**

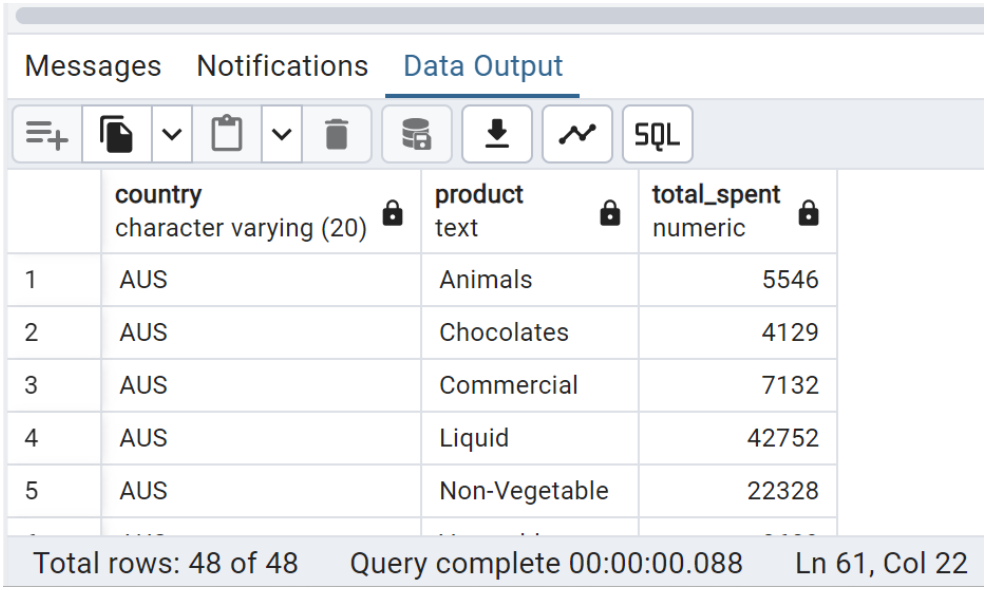
Messages Notifications Data Output

	country character varying (20)	total_spent numeric
1	SP	659557
2	SA	211071
3	CA	167403
4	AUS	85576
5	IND	77806
6	GER	73198
7	US	67546
8	ME	3122

Total rows: 8 of 8 Query complete 00:00:00.135

```sql
SELECT
    cc.country,  -- Country of the customer
    ROUND(SUM(ph.amtliq + ph.amtvege + ph.amtnonveg + ph.amtpes + ph.amtchocolates
    + ph.amtcomm))
    AS total_spent
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id  -- Join to get the country for
GROUP BY
    cc.country  -- Group by country
ORDER BY
    total_spent DESC;  -- Order by total spend in descending order
```

2. The total spend per product per country?



Messages Notifications Data Output

	country character varying (20)	product text	total_spent numeric
1	AUS	Animals	5546
2	AUS	Chocolates	4129
3	AUS	Commercial	7132
4	AUS	Liquid	42752
5	AUS	Non-Vegetable	22328

Total rows: 48 of 48 Query complete 00:00:00.088 Ln 61, Col 22

The following SQL code gave the results, which I downloaded in Excel.

```
SELECT
    cc.country,  -- Country of the customer
    'Liquid' AS product, ROUND(SUM(ph.amtliq)) AS total_spent
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id
GROUP BY
    cc.country
UNION ALL
SELECT
    cc.country,
    'Vegetable' AS product, ROUND(SUM(ph.amtvege)) AS total_spent
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id
GROUP BY
    cc.country
UNION ALL
```

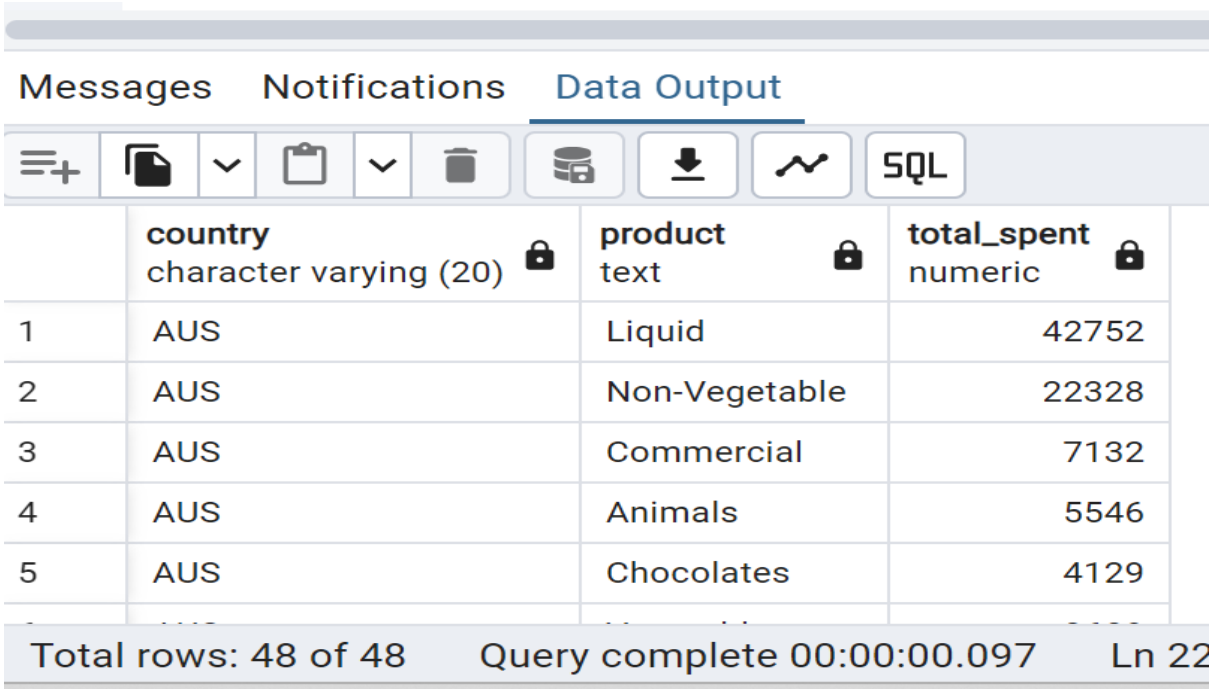
```sql
SELECT
    cc.country,
    'Non-Vegetable' AS product, ROUND(SUM(ph.amtnonveg)) AS total_spent
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id
GROUP BY
    cc.country
UNION ALL
SELECT
    cc.country,
    'Animals' AS product, ROUND(SUM(ph.amtpes)) AS total_spent --  with '
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id
GROUP BY
    cc.country
```

```sql
UNION ALL
SELECT
    cc.country,
    'Chocolates' AS product, ROUND(SUM(ph.amtchocolates)) AS total_spent
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id
GROUP BY
    cc.country
UNION ALL
SELECT
    cc.country,
    'Commercial' AS product, ROUND(SUM(ph.amtcomm)) AS total_spent
FROM
    product_history ph
JOIN
    customer_countries cc ON ph.customer_id = cc.customer_id
GROUP BY
    cc.country
```

```sql
ORDER BY
    country, product;
```

3. **Which products are the most popular in each country?**

```sql
SELECT
    cc.country,
    product,
    ROUND(SUM(amount)) AS total_spent
FROM
    customer_countries cc
JOIN
    product_history ph ON cc.customer_id = ph.customer_id
JOIN
    LATERAL ( --This is used to split the product_history tab
        VALUES
            ('Liquid', ph.amtliq),
            ('Vegetable', ph.amtvege),
            ('Non-Vegetable', ph.amtnonveg),
            ('Animals', ph.amtpes),
            ('Chocolates', ph.amtchocolates),
            ('Commercial', ph.amtcomm)
    ) AS product_spend(product, amount) ON true
GROUP BY
    cc.country, product
ORDER BY
    cc.country, total_spent DESC;
```



4. Which products are the most popular based on marital status?

```sql
SELECT
    cd.marital_status,
    product,
    ROUND(SUM(amount)) AS total_spent --This calculates the tota
FROM
    customer_details cd
JOIN
    product_history ph ON cd.customer_id = ph.customer_id
JOIN
    LATERAL ( --This creates a virtual table for each product an
        VALUES
            ('Liquid', ph.amtliq),
```

```
        VALUES
            ('Liquid', ph.amtliq),
            ('Vegetable', ph.amtvege),
            ('Non-Vegetable', ph.amtnonveg),
            ('Animals', ph.amtpes),
            ('Chocolates', ph.amtchocolates),
            ('Commercial', ph.amtcomm)
    ) AS product_spend(product, amount) ON true
GROUP BY
    cd.marital_status, product
ORDER BY -- The results are sorted first by marital status
    cd.marital_status, total_spent DESC;
```

Messages Notifications Data Output

	marital_status character varying (255)	product text	total_spent numeric
1	Absurd	Liquid	711
2	Absurd	Non-Vegetable	625
3	Absurd	Animals	411
4	Absurd	Commercial	408
5	Absurd	Vegetable	169

Total rows: 48 of 48 Query complete 00:00:00.100 Ln 22, Col 41

5.which products are the most popular based on whether or not there are children or teens in the home.

```
SELECT
    CASE --This checks if there are children (kidhome > 0) or teens (teenhome
        WHEN cd.kidhome > 0 OR cd.teenhome > 0 THEN 'With Children or Teens'
        ELSE 'No Children or Teens'
    END AS home_type,
    product,
    ROUND(SUM(amount)) AS total_spent --This calculates the total spend for e
FROM
    customer_details cd
JOIN
    product_history ph ON cd.customer_id = ph.customer_id
JOIN
    LATERAL ( -- this splits the product_history table's amount
```

Query History

```
    LATERAL ( -- this splits the product_history table's amount
        VALUES
            ('Liquid', ph.amtliq),
            ('Vegetable', ph.amtvege),
            ('Non-Vegetable', ph.amtnonveg),
            ('Animals', ph.amtpes),
            ('Chocolates', ph.amtchocolates),
            ('Commercial', ph.amtcomm)
    ) AS product_spend(product, amount) ON true
GROUP BY --This groups the results by home_type
    home_type, product
ORDER BY    --This orders the results first by the home type
    home_type, total_spent DESC;
```

Messages Notifications Data Output

	home_type text	product text	total_spent numeric
1	No Children or Teens	Liquid	308950
2	No Children or Teens	Non-Vegetable	234758
3	No Children or Teens	Animals	48500
4	No Children or Teens	Commercial	40661
5	No Children or Teens	Chocolates	33663

Total rows: 12 of 12 Query complete 00:00:00.088 Ln 25, Col 33

3.Analysis of Product popularity based on Customer Demographic:

Query Query History

```
5  SELECT
6      cd.country,
7      cd.marital_status,
8      cd.kidhome,
9      cd.teenhome,
10     'Liquid' AS product_type,
11     SUM(ph.amtliq) AS total_spent,
12     SUM(cb.numwebbuy + cb.numwalkinpur) AS total_conversions
13 FROM
14     product_history ph
15 JOIN
16     customer_details cd ON ph.customer_id = cd.customer_id
```

```
17  JOIN
18      customer_behaviour cb ON ph.customer_id = cb.customer_id
19  GROUP BY
20      cd.country, cd.marital_status, cd.kidhome, cd.teenhome
21
22  UNION ALL
23
24  SELECT
25      cd.country,
26      cd.marital_status,
27      cd.kidhome,
28      cd.teenhome,
```

```
29      'Vegetables' AS product_type,
30      SUM(ph.amtvege) AS total_spent, -- The sum of the amounts spent on each product
31      SUM(cb.numwebbuy + cb.numwalkinpur) AS total_conversions --The sum of numwebbuy and nu
32  FROM
33      product_history ph
34  JOIN
35      customer_details cd ON ph.customer_id = cd.customer_id
36  JOIN
37      customer_behaviour cb ON ph.customer_id = cb.customer_id
38  GROUP BY
39      cd.country, cd.marital_status, cd.kidhome, cd.teenhome
40
```

Query | Query History

```
40
41  UNION ALL
42
43  SELECT
44      cd.country,
45      cd.marital_status,
46      cd.kidhome,
47      cd.teenhome,
48      'Non-Vegetables' AS product_type,
49      SUM(ph.amtnonveg) AS total_spent,
50      SUM(cb.numwebbuy + cb.numwalkinpur) AS total_conversions
51  FROM
52      product_history ph
```

Query | Query History

```
    product_history ph
JOIN
    customer_details cd ON ph.customer_id = cd.customer_id
JOIN
    customer_behaviour cb ON ph.customer_id = cb.customer_id
GROUP BY
    cd.country, cd.marital_status, cd.kidhome, cd.teenhome

UNION ALL

SELECT
    cd.country,
    cd.marital_status,
```

Query | Query History

```
    cd.kidhome,
    cd.teenhome,
    'Pets' AS product_type,
    SUM(ph.amtpes) AS total_spent,
    SUM(cb.numwebbuy + cb.numwalkinpur) AS total_conversions
FROM
    product_history ph
JOIN
    customer_details cd ON ph.customer_id = cd.customer_id
JOIN
    customer_behaviour cb ON ph.customer_id = cb.customer_id
GROUP BY
    cd.country, cd.marital_status, cd.kidhome, cd.teenhome

UNION ALL

SELECT
    cd.country,
    cd.marital_status,
    cd.kidhome,
    cd.teenhome,
    'Chocolates' AS product_type,
    SUM(ph.amtchocolates) AS total_spent,
```

```
88      SUM(cb.numwebbuy + cb.numwalkinpur) AS total_conversions
89  FROM
90      product_history ph
91  JOIN
92      customer_details cd ON ph.customer_id = cd.customer_id
93  JOIN
94      customer_behaviour cb ON ph.customer_id = cb.customer_id
95  GROUP BY
96      cd.country, cd.marital_status, cd.kidhome, cd.teenhome
97
98  UNION ALL
99
```

Query | Query History

```
100 SELECT
101     cd.country,
102     cd.marital_status,
103     cd.kidhome,
104     cd.teenhome,
105     'Commercial Items' AS product_type,
106     SUM(ph.amtcomm) AS total_spent,
107     SUM(cb.numwebbuy + cb.numwalkinpur) AS total_conversions
108 FROM
109     product_history ph
110 JOIN
111     customer_details cd ON ph.customer_id = cd.customer_id
112 JOIN
113     customer_behaviour cb ON ph.customer_id = cb.customer_id
114 GROUP BY
115     cd.country, cd.marital_status, cd.kidhome, cd.teenhome -- The re
116 ORDER BY
117     total_spent DESC;
```

Messages | Notifications | Data Output

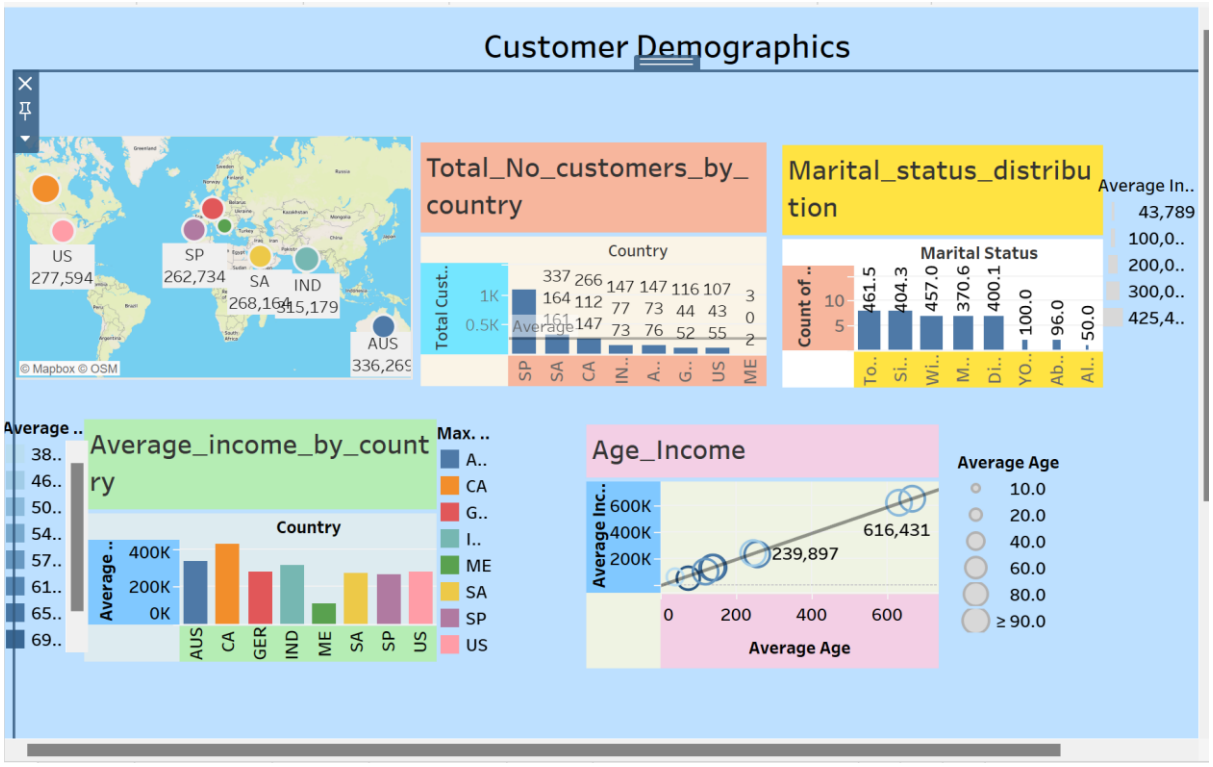
	y ter varying (255)	marital_status character varying (255)	kidhome integer	teenhome integer	product_type text	total_spent numeric	total_conversions bigint
1		Married	0	1	Liquid	52263.00	1740
2		Married	0	0	Liquid	51429.00	1228
3		Together	0	0	Liquid	41766.00	874
4		Single	0	0	Liquid	39104.00	900
5		Together	0	1	Liquid	37444.00	1083

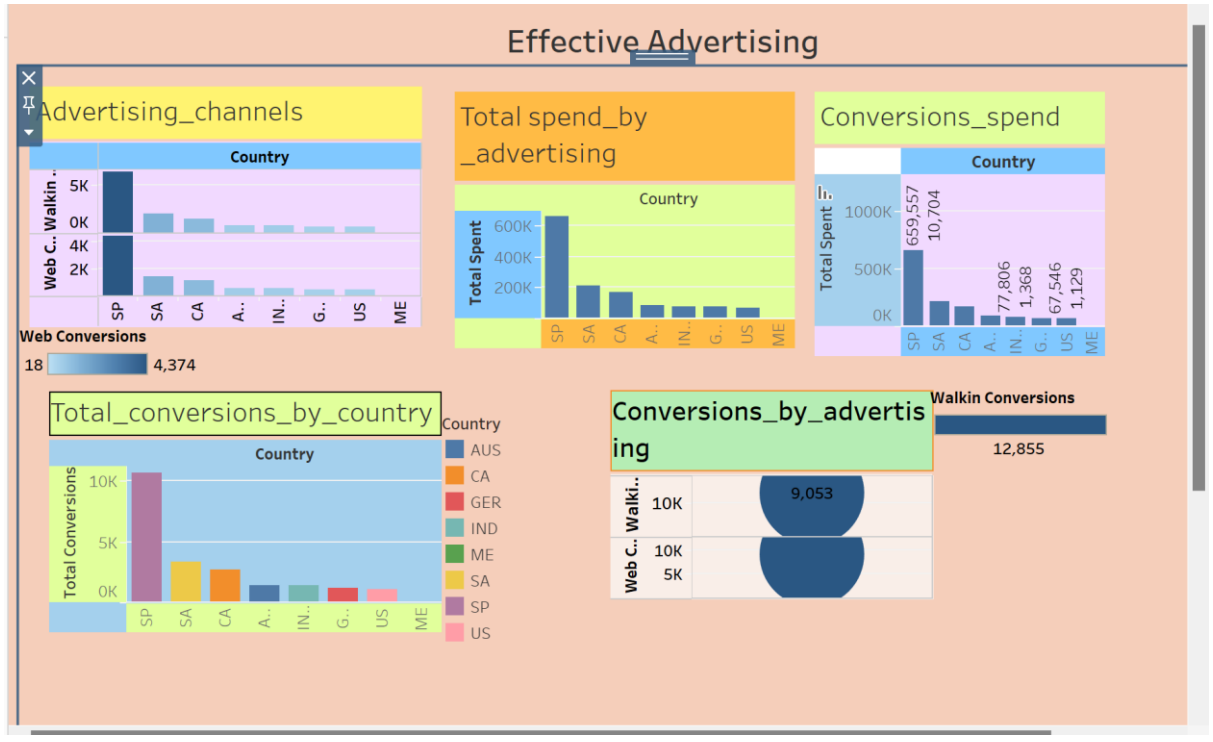
Total rows: 1000 of 1158 Query complete 00:00:00.112 Ln 115, Col 218

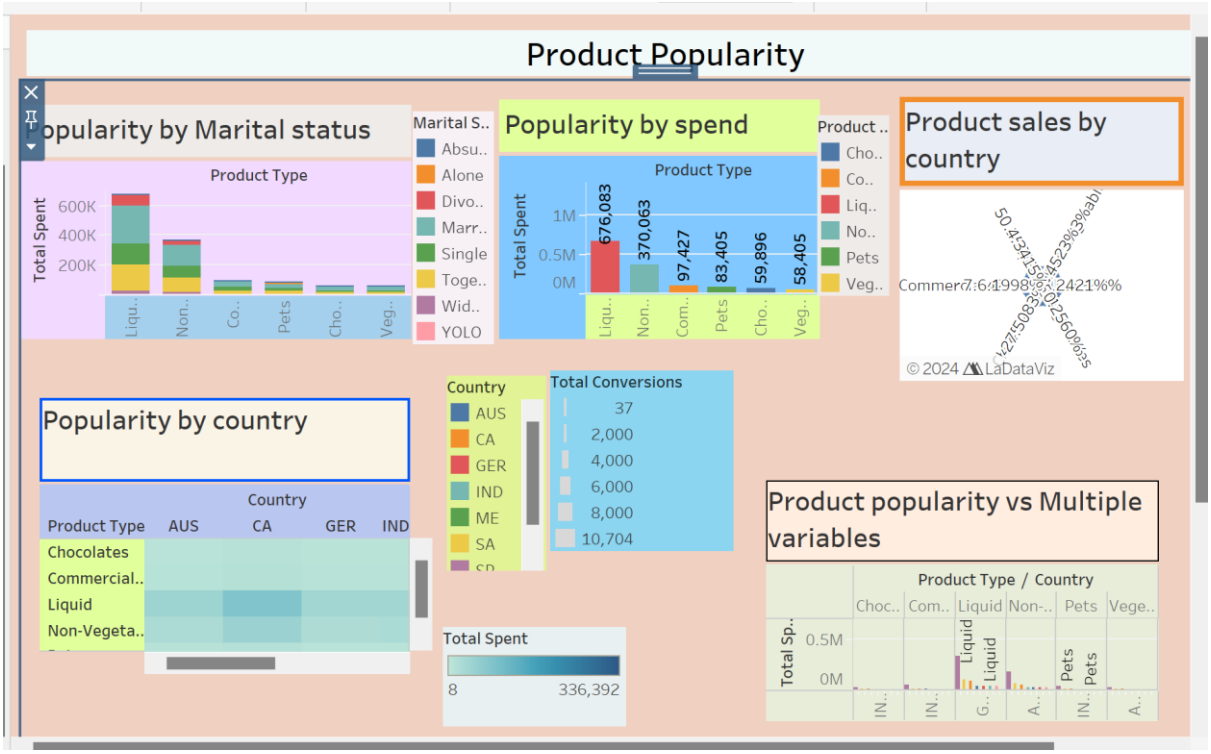
Conclusion:

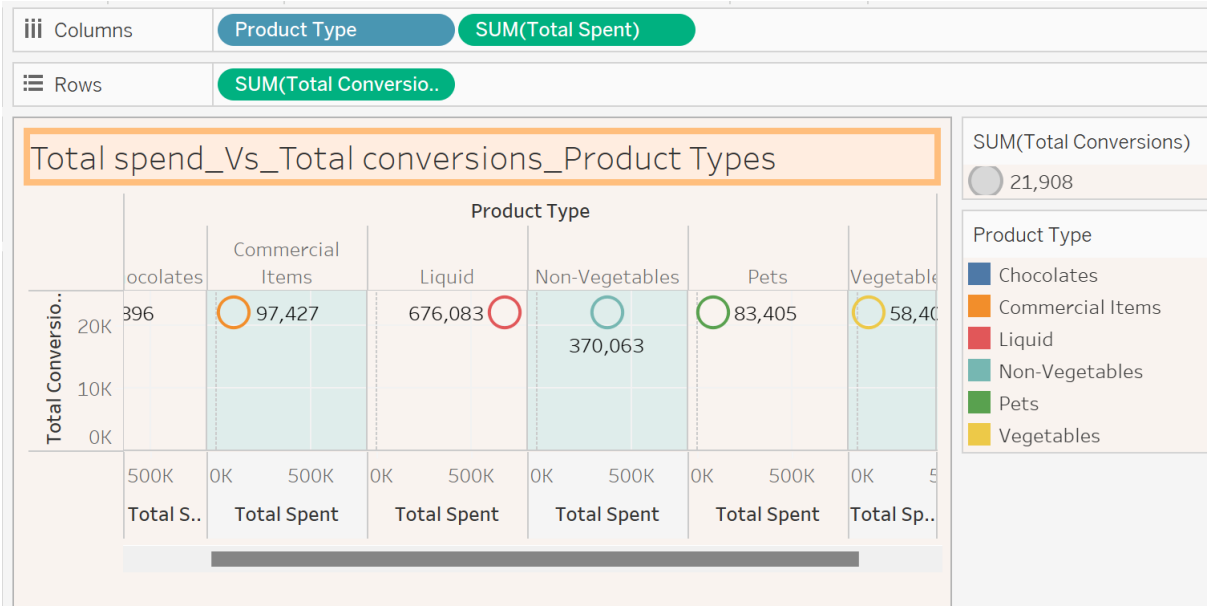
I imported all the above analysis on Tableau to visualise the findings to answer the business questions raised by our stakeholders, to gain insights into customer demographic to optimise marketing efforts.

This segmentation shows that the majority of customer fall within the average age of 50 and tend to be from Canada and highest total conversion rate is from Sao Paulo and the total spend decreased with progressive age, and our top selling products are liquids as shown below.


Customer Demographics
US
277,594
SP
262,734
SA
268,164
IND
315,179
AUS
336,269
© Mapbox © OSM
Total_No_customers_by_country
Country
Total Cust..
1K
0.5K
Average
337 266 147 147 116 107 3
164 112 77 73 44 43 0
161 147 73 76 52 55 2
SP SA CA IN.. A.. G.. US ME
Marital_status_distribution
Marital Status
Count of ..
10
5
461.5 404.3 457.0 370.6 400.1 100.0 96.0 50.0
To.. Si.. Wi.. M.. Di.. YO.. Ab.. Al..
Average In..
43,789
100,0..
200,0..
300,0..
425,4..
Average ..
38..
46..
50..
54..
57..
61..
65..
69..
Average_income_by_country
Country
Average ..
400K
200K
0K
AUS CA GER IND ME SA SP US
Max. ..
A..
CA
G..
I..
ME
SA
SP
US
Age_Income
Average Inc..
600K
400K
200K
616,431
239,897
0
200
400
600
Average Age
Average Age
10.0
20.0
40.0
60.0
80.0
≥ 90.0


Effective Advertising
Advertising_channels
Country
Walkin ..
Web C..
5K
0K
4K
2K
SP SA CA A.. IN.. G.. US ME
Web Conversions
18
4,374
Total spend_by_advertising
Country
Total Spent
600K
400K
200K
SP SA CA A.. IN.. G.. US ME
Conversions_spend
Country
Total Spent
1000K
500K
0K
659,557
10,704
77,806
1,368
67,546
1,129
SP SA CA A.. IN.. G.. US ME
Total_conversions_by_country
Country
Total Conversions
10K
5K
0K
SP SA CA A.. IN.. G.. US ME
Country
AUS
CA
GER
IND
ME
SA
SP
US
Conversions_by_advertising
Walkin Conversions
12,855
Walki..
Web C..
10K
10K
5K
9,053


Product Popularity
Popularity by Marital status
Marital S..
Absu..
Alone
Divo..
Marr..
Single
Toge..
Wid..
YOLO
Product Type
Total Spent
600K
400K
200K
Liqu..
Non..
Co..
Pets
Cho..
Veg..
Popularity by spend
Product Type
Total Spent
1M
0.5M
0M
676,083
370,063
97,427
83,405
59,896
58,405
Product ..
Cho..
Co..
Liq..
No..
Pets
Veg..
Product sales by country
© 2024 LaDataViz
Popularity by country
Country
Product Type
AUS
CA
GER
IND
Chocolates
Commercial..
Liquid
Non-Vegeta..
Country
AUS
CA
GER
IND
ME
SA
Total Conversions
37
2,000
4,000
6,000
8,000
10,704
Total Spent
8
336,392
Product popularity vs Multiple variables
Product Type / Country
Choc..
Com..
Liquid
Non-..
Pets
Vege..
Total Sp..
0.5M
0M


Columns
Product Type
SUM(Total Spent)
Rows
SUM(Total Conversio..
Total spend_Vs_Total conversions_Product Types
Product Type
ocolates
Commercial Items
Liquid
Non-Vegetables
Pets
Vegetable
Total Conversio..
20K
10K
0K
396
97,427
676,083
370,063
83,405
58,40
500K
0K
500K
Total S..
Total Spent
SUM(Total Conversions)
21,908
Product Type
Chocolates
Commercial Items
Liquid
Non-Vegetables
Pets
Vegetables