

MARKET TECHNICAL REPORT

Kelvin Dhondee

Data Analytics for Business Dec 2024

Table of Contents

PROBLEM STATEMENT	2
ANALYTICAL APPROACH.....	3
DATA CLEANING AND VALIDATION	3
RECENCY, FREQUENCY, MONETARY (RFM) SCORE FORMULATION	4
DASHBOARD DESIGN AND DEVELOPMENT.....	5
DESIGN PRINCIPLES.....	5
DASHBOARD DEVELOPMENT (SCREENSHOTS)	5
PATTERNS, TRENDS, INSIGHTS AND FINDINGS.....	8
ANALYSIS CONDUCTED WITH EXCEL.....	8
CUSTOMER DEMOGRAPHICS	8
EFFECTIVENESS OF ADVERTISING CHANNELS AND PROMOTION.....	12
POPULARITY OF PRODUCT CATEGORIES	13
ANALYSIS CONDUCTED WITH SQL (VIA PGADMIN)	14
CUSTOMER DEMOGRAPHICS	15
EFFECTIVENESS OF ADVERTISING CHANNELS AND PROMOTION.....	18
POPULARITY OF PRODUCT CATEGORIES	20
RFM ANALYSIS.....	22
ANALYSIS CONDUCTED WITH TABLEAU	29
DASHBOARD 1	29
DASHBOARD 2	29
DASHBOARD 3	29
DASHBOARD 4	30
RECOMMENDATIONS AND FURTHER ACTIONS	31
APPENDICES	32
APPENDIX A.....	32
DATA CLEANING AND VALIDATION CONDUCTED IN EXCEL	32
APPENDIX B	34
DASHBOARD DEVELOPMENT (DETAILED)	34
WORKS CITED.....	38

Problem Statement

Problem: 2Market wishes to better understand the demographics of its customers, the effectiveness of currently utilised advertising channels and the popularity of selected product categories.

2Market suspects that its current advertising strategy is currently based on a scattergun approach without a clear understanding of the market especially given the fact that it has a customer base in eight countries.

Current state: The 5W1H Analysis Framework (Visual Paradigm Online, 2024) adapted for a customer-centric approach is used to establish the current state of relevant factors/variables. Recommendations will be offered based on data-driven insights, patterns and trends identified for some of the most compelling factors and variables.



Desired outcome with business context: Enable 2Market to formulate a more targeted advertising strategy with the aim boosting revenue by 15% within the next two years. This is in line with established strategic goals of profitability and expanding market share in all locations.

Analytical Approach

Data cleaning and validation

In terms of data cleaning and validation, checks were conducted to identify missing values, incorrect spelling, duplicates, data type inconsistencies/anomalies and outliers. For a detailed account of all the steps undertaken (see [Appendix A](#)).

Four outliers were removed: **Year_Birth** of 1894, 1900, 1901 and ID 9432. The removal of these four records represented 0.18% of initial total. Summary statistics for age and income suggest that removal had minimal impact on average values.

Used 2014 as the reference year for when analysis is being undertaken.

Find & Replace was used to rationalise **Marital_Status**:

Changed Absurd (x2) & YOLO (x2) to Unknown.
Grouped Alone(x3) with Single.
Grouped Married (x857) with Together(x573) and renamed to Relationship.
Kept Divorced and Widow as these may provide insights for these distinct groups, particularly given that the Divorced status comprises of 232 entries.

Dealt with inconsistent date formats in **Dt_Customer** as follows:

Dt_Customer	MM	DD	YY	YYYY	DATE_FN	IF_ISNUMBER
05/04/2013	41369			#N/A	#N/A	05/04/2013
6/15/14	6	15	14	2014	15/06/2014	15/06/2014
8/18/12	8	18	12	2012	18/08/2012	18/08/2012
7/22/13	7	22	13	2013	22/07/2013	22/07/2013
09/06/2013	41434			#N/A	#N/A	09/06/2013
09/01/2013	41283			#N/A	#N/A	09/01/2013
12/03/2012	40980			#N/A	#N/A	12/03/2012
2/15/14	2	15	14	2014	15/02/2014	15/02/2014
11/16/12	11	16	12	2012	16/11/2012	16/11/2012
09/08/2012	41130			#N/A	#N/A	09/08/2012
4/24/14	4	24	14	2014	24/04/2014	24/04/2014
05/11/2014	41948			#N/A	#N/A	05/11/2014
12/07/2012	41102			#N/A	#N/A	12/07/2012
9/18/13	9	18	13	2013	18/09/2013	18/09/2013
5/26/13	5	26	13	2013	26/05/2013	26/05/2013

Recency, Frequency, Monetary (RFM) Score Formulation

RFM scores (Murphy, 2024) were allocated to each customer based on three quantitative factors:

1. Recency: How recently a customer made a purchase.
2. Frequency: How often a customer makes a purchase.
3. Monetary: How much a customer spends on purchases.

The following were used for this purpose:

recency – to calculate Recency Score.

numwebuy, numwalkinpur – to calculate Frequency Score

amtliq, amtvege, amtnonveg, amtpes, amtchocolates, amtcomm – to calculate Monetary Score.

Scores were allocated to quartile bands:

Recency (Days since last purchase)	
0-24	4
25-49	3
50-74	2
75-99	1

Frequency (Number of purchases in store + online)	
0-5	1
6-10	2
11-14	3
15-27	4

Monetary (Value of purchases)	
0-69	1
70-397	2
398-1048	3
1049-2525	4

The average of the three scores was calculated to provide an overall RFM Score.

Dashboard Design and Development

Four dashboards were created on Tableau to address the three dimensions of the problem statement.

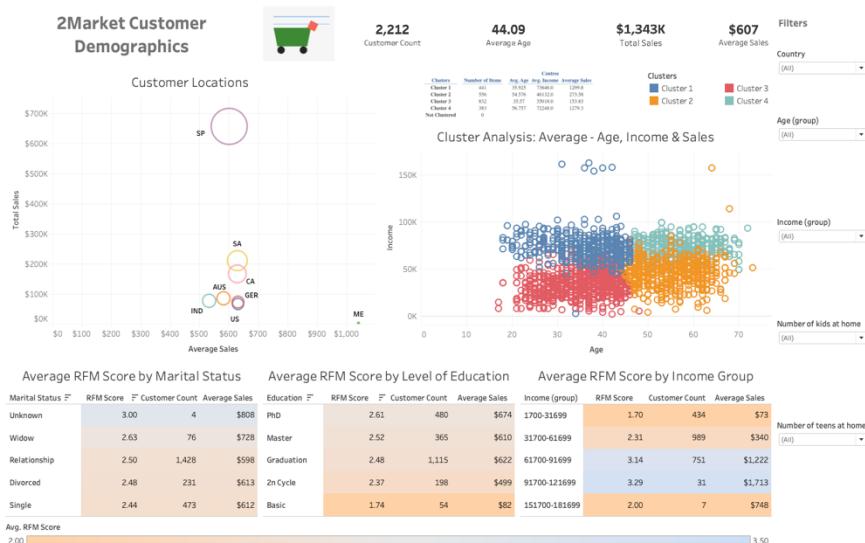
Design Principles

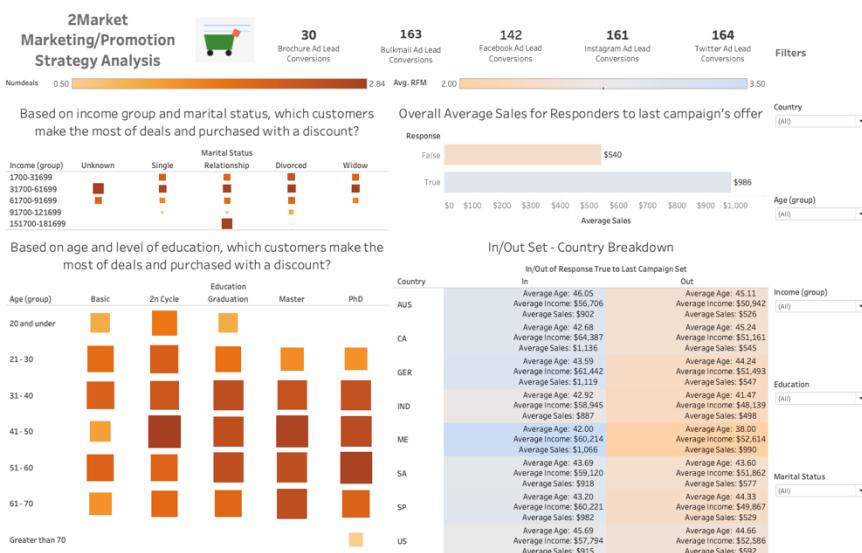
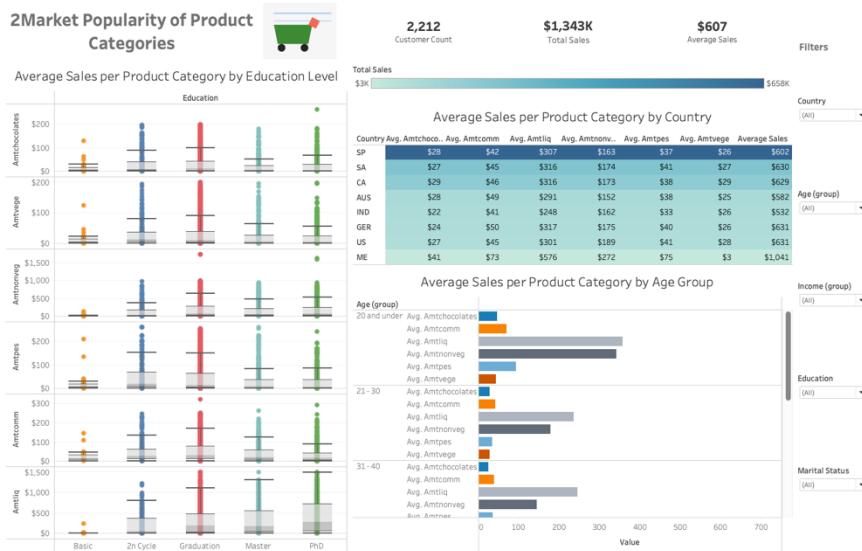
Design principles were as follows:

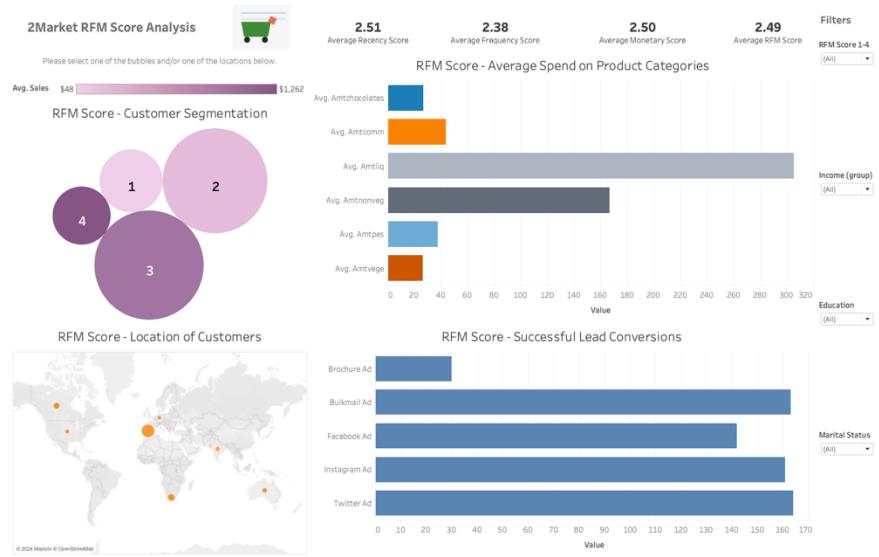
- (i) A variety of suitable visualisations were employed to maintain interest. These were complemented with interactive elements such as filters to the side or selectable elements on the visualisation themselves.
- (ii) Ensured labels, headings and text were legible.
- (iii) Colour palette selected with accessibility considerations in mind.
- (iv) Clutter was minimised – gridlines removed; number of visualisations limited to five per dashboard.
- (v) Dashboards sectioned in a way to aid navigability.

Dashboard Development (Screenshots)

See [Appendix B](#) for more details regarding rationale behind choice of visualisations.







Patterns, Trends, Insights and Findings

Analysis conducted with Excel

Customer demographics

age					
Average	44				
Median	43				
Mode	37				
Min	17				
Max	73				
Q1	36				
Q2	43				
Q3	54				
IQR	18				
Income					
Total	114932889				
Average	51959				
Median	51369				
Min	1730				
Max	162397				
Q1	35234				
Q2	51369				
Q3	68481				
IQR	33247				
Customers per country					
SA	335				
CA	266				
SP	1092				
GER	116				
IND	146				
US	107				
AUS	147				
ME	3				
	<u><u>2212</u></u>				

Aggregation results.

customer_id	age	education	marital_status	income	kidhome	teenhome	dt	customer	recency	amtltq	amtvisge	amtnonveg	amtipes	amtchocolates	amtcomm	numdeals	numwebbuy	numwalkinpur	numvisits	response	complain	country	count_success
1503	37	PhD	Relationship	162,397.00	1	1	06/03/2013	31	85	1	16	1622	17	2	1	2	0	0	1	1	0	SP	0
1501	42	PhD	Relationship	160,803.00	0	0	06/04/2013	21	55	1	16	1622	17	3	4	15	0	0	1	0	US	0	
5326	42	Master	Relationship	160,300.00	1	0	06/04/2013	37	39	1	6	2	0	8	0	0	1	1	1	0	SP	0	
8475	40	PhD	Relationship	157,243.00	0	1	03/01/2014	98	20	2	2	1582	1	2	1	15	0	0	1	0	IND	0	
4931	36	Graduation	Relationship	157,146.00	0	0	29/04/2013	13	1	0	1725	2	1	1	0	0	0	0	1	0	SA	0	
11181	64	PhD	Relationship	156,924.00	0	0	29/08/2013	85	2	1	2	1	1	1	0	0	0	0	0	0	CA	0	
5555	38	Graduation	Divorced	155,924.00	0	0	02/07/2014	81	1	1	1	1	1	1	0	0	0	0	0	0	SP	0	
4611	43	Graduation	Relationship	155,800.00	0	0	06/04/2013	9	8	2	1	262	3	0	27	0	1	0	0	0	SP	0	
4611	43	Graduation	Relationship	105,471.00	0	0	21/04/2013	36	109	181	104	202	207	0	9	0	0	13	3	1	0	SP	0
10099	39	Graduation	Divorced	102,970.00	0	0	04/05/2013	5	168	188	444	32	177	149	1	6	13	3	1	0	S&H	0	

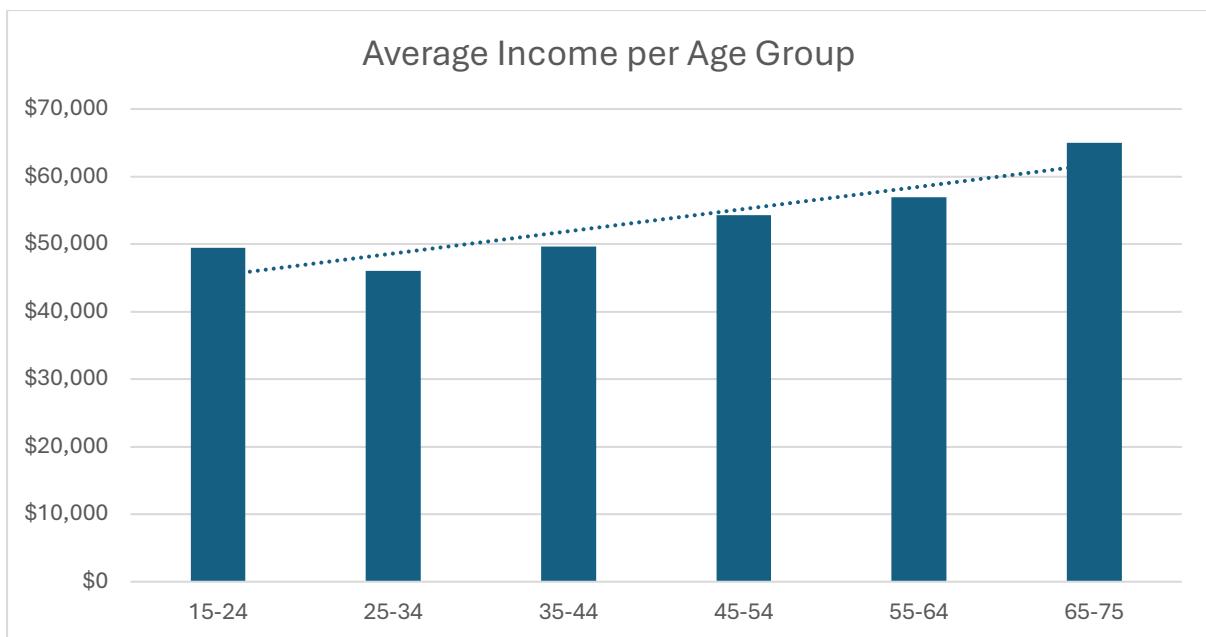
Sorting the income column in descending order revealed the characteristics of top earners being highly educated and predominantly in a relationship.

customer_id	age	education	marital_status	income	lidhome	teenhome	dt_customer	recency	amtbig	amtnorm	amtopen	amtchocolates	amtcoffee	total_spent	numdeals	numwebbuy	numwalkup	numvisits	response	complain	country	count_success
5500	22	Master	Single	90,630.00	0	0	13/02/2014	29	1156	120	915	94	144	2628	1	3	10	1	0	SA	1	
5755	23	Master	Single	90,630.00	0	0	13/02/2014	29	1156	120	915	94	144	95	2525	1	3	10	1	0	SP	1
1763	25	Graduation	Relationship	87,679.00	0	0	27/07/2013	62	1259	172	815	97	148	357	2524	1	7	10	4	1	CA	3
4586	44	Graduation	Relationship	75,769.00	0	0	25/10/2013	46	1394	22	708	89	91	187	2486	1	9	9	5	1	SA	3
4476	64	PhD	Relationship	69,986.00	0	0	16/02/2013	82	1315	22	780	145	0	178	2440	1	7	9	5	0	SP	0
5453	57	Master	Relationship	99,226.00	0	0	26/09/2012	26	1085	108	649	153	151	108	2352	1	4	2	0	0	SA	1
1833	41	Master	Single	90,630.00	0	0	13/02/2014	18	1302	86	731	88	114	249	0	6	12	2	1	CA	9	
4414	41	Master	Relationship	93,151.00	0	0	23/10/2013	98	986	147	842	137	44	2346	1	5	10	2	1	SA	3	
5386	60	Graduation	Relationship	94,384.00	0	0	03/04/2013	62	1111	790	160	45	172	2302	0	5	5	2	1	SA	3	
6024	60	Graduation	Relationship	94,384.00	0	0	09/04/2013	62	1111	24	790	160	45	172	2302	0	5	5	2	1	SP	3

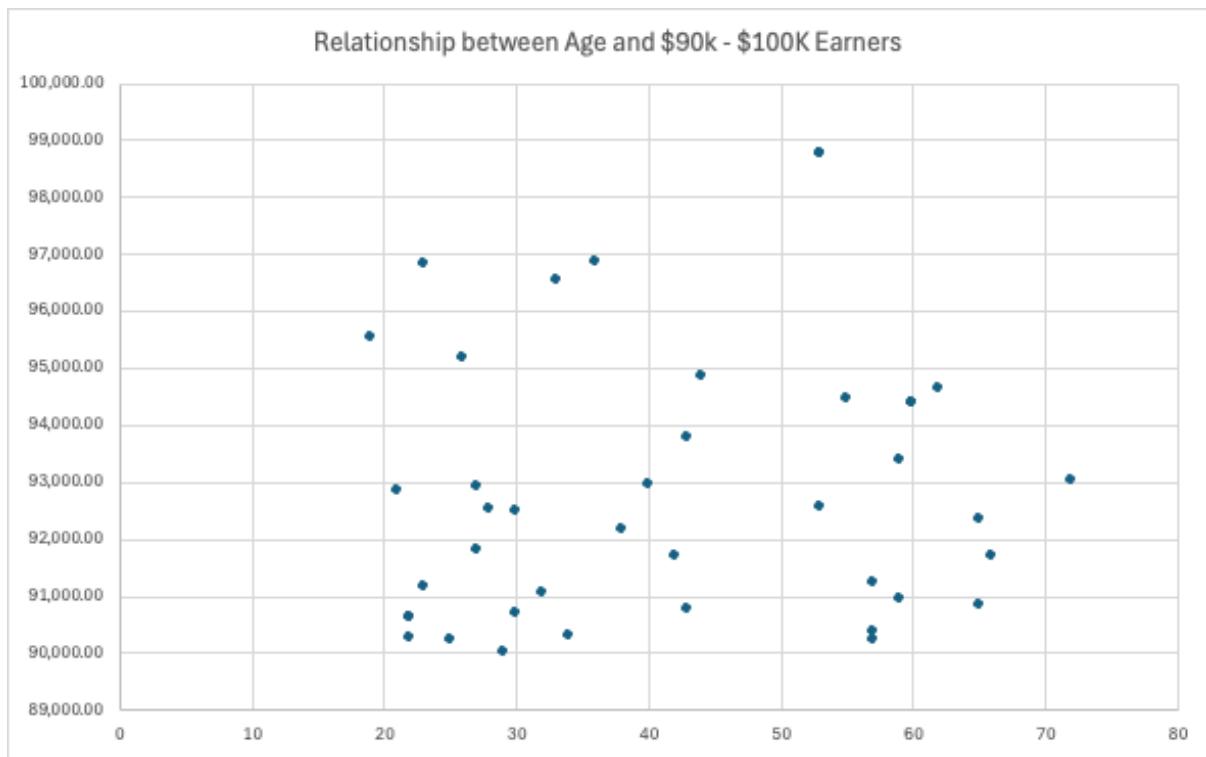
Top spenders share similar characteristics with top earners but show higher count.

Both groups did not have a lot of teens or kids at home.

% breakdown of customer population by marital status	
Row Labels	% by marital status
Divorced	10.44%
Relationship	64.56%
Single	21.38%
Unknown	0.18%
Widow	3.44%
Grand Total	100.00%



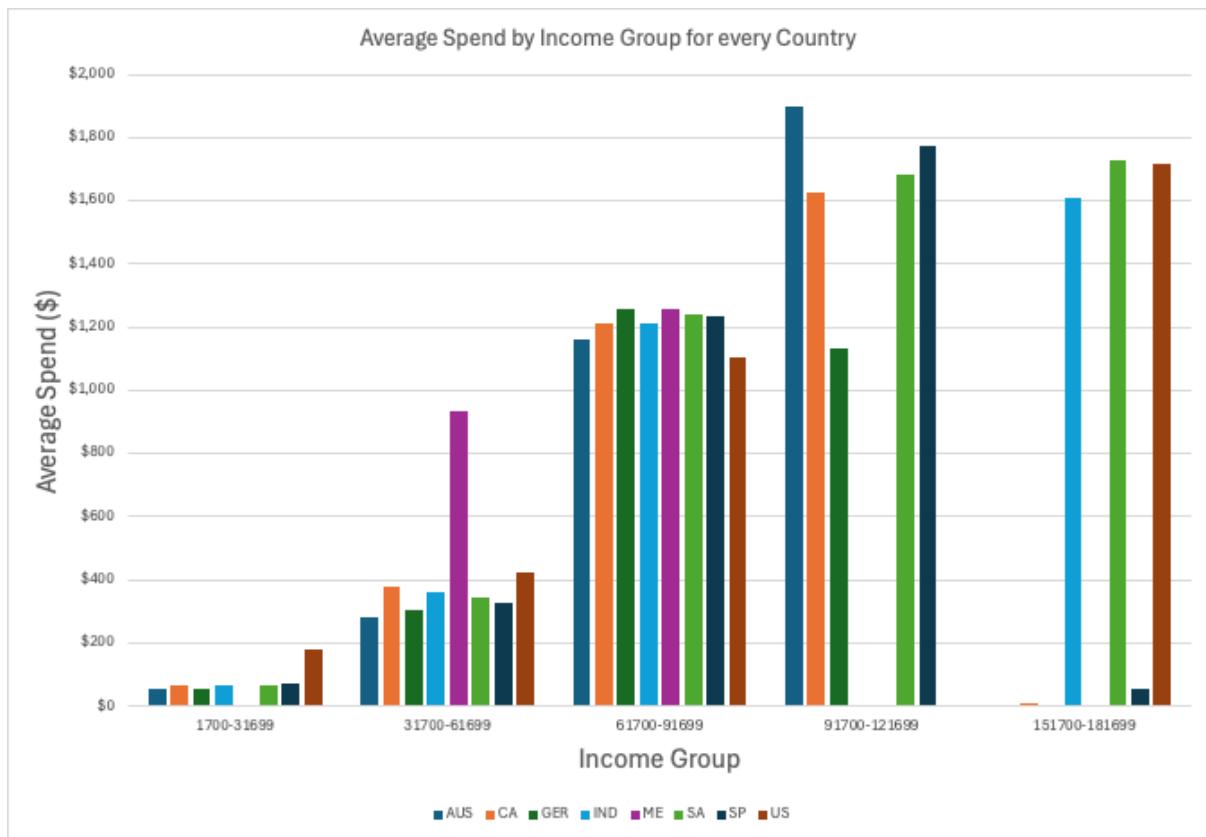
Income generally increases with age with some individual exceptions in the 15-24 range.



No discernible relationship. There are high earners across that age range for that income range.

Number of customers and market share		
Row Labels	Count of customer_id	Market Share
AUS	147	6.37%
CA	266	12.46%
GER	116	5.45%
IND	146	5.79%
ME	3	0.23%
SA	335	15.71%
SP	1092	48.96%
US	107	5.03%
Grand Total	2212	100.00%

Significantly more customers based in Spain.



91700-121699 highest average spend at \$1,713, followed by 61700-91699 at \$1,222.

Effectiveness of advertising channels and promotion

Number of customers who accepted last ad campaign's offer		
Row Labels	Count of response	
No		1879
Yes		333
Grand Total	2212	

Successful lead conversions		
Lead conversion value	Count per lead conversion value	Sum per lead conversion value
0	1754	0
1	322	322
2	81	162
3	44	132
4	11	44
Grand Total	2212	660

333 customers accepted last ad campaign's offer (15.05%)

660 successful lead conversions.

Popularity of product categories

Average Spend of every country by product category									
Values	Column Labels								
	AUS	CA	GER	IND	ME	SA	SP	US	Grand Total
Average of amtvege	\$25	\$29	\$26	\$26	\$3	\$27	\$26	\$28	\$26
Average of amtchocolates	\$28	\$29	\$24	\$22	\$41	\$27	\$28	\$27	\$27
Average of amtpes	\$38	\$38	\$40	\$33	\$75	\$41	\$37	\$41	\$38
Average of amtcomm	\$49	\$46	\$50	\$41	\$73	\$45	\$42	\$45	\$44
Average of amtnonveg	\$152	\$173	\$175	\$162	\$272	\$174	\$163	\$189	\$167
Average of amtliq	\$291	\$316	\$317	\$248	\$576	\$316	\$307	\$301	\$305

Considerable spending on liquor. Bestselling product category across all countries.

Average Spend every income group by each product category						
Values	Column Labels					Grand Total
	1700-31699	31700-61699	61700-91699	91700-121699	151700-181699	
Average of amtvege	\$5	\$13	\$55	\$62	\$3	\$26
Average of amtchocolates	\$6	\$13	\$56	\$84	\$1	\$27
Average of amtpes	\$8	\$18	\$78	\$97	\$4	\$38
Average of amtcomm	\$16	\$36	\$70	\$64	\$3	\$44
Average of amtnonveg	\$21	\$67	\$360	\$595	\$708	\$167
Average of amtliq	\$16	\$193	\$602	\$810	\$29	\$305

Increasing income generally shows higher spend for liquor and non-veg categories.

Exception in 151700-181699 income range.

How does level of education affect average spending for each product category						
Values	Column Labels					
	Basic	Graduation	2n Cycle	Master	PhD	Grand Total
Average of amtvege	\$11	\$31	\$30	\$21	\$20	\$26
Average of amtchocolates	\$12	\$31	\$35	\$21	\$20	\$27
Average of amtpes	\$17	\$43	\$48	\$31	\$27	\$38
Average of amtcomm	\$23	\$51	\$47	\$40	\$32	\$44
Average of amtnonveg	\$11	\$181	\$136	\$163	\$169	\$167
Average of amtliq	\$7	\$285	\$203	\$333	\$406	\$305

Other than customers with basic level of education, spend on liquor and non-veg higher within groups based on same education level.

Decomposition of Total Sales by Product Category	
Values	Sum
Sum of amtvege	\$58,241
Sum of amtchocolates	\$59,827
Sum of amtpes	\$83,279
Sum of amtcomm	\$97,164
Sum of amtnonveg	\$369,470
Sum of amtliq	\$675,296
Sum of total_sales	\$1,343,277

Analysis conducted with SQL (via pgAdmin)

Screenshots of SQL queries and results encompassing all aspects of the business problem.

Selected commentary:

- (i) In terms of recency days, customers who complain did not stay away too long (51 days) compared to those who did not complain (49 days).
- (ii) Investigated the activity from Montenegro (country with highest average sales). This pertained to three highly-educated customers only.
- (iii) Even though Spain has more customers, five countries rank higher in terms of average sales.
- (iv) Customers in a relationship were responsible for higher numbers of successful lead conversions.
- (v) Alcohol most popular product category across all countries and for all marital statuses.
- (vi) Extensive RFM analysis inspired the RFM Score Analysis dashboard on Tableau.

Customer demographics

-- Of those customers who have children, what is the breakdown of kids/teens based on marital status?

```
SELECT marital_status,
       SUM(kidhome) AS kids_at_home,
       SUM(teenhome) AS teens_at_home
  FROM marketing_data
 GROUP BY marital_status;
```

	marital_status character varying (20)	kids_at_home bigint	teens_at_home bigint
1	Unknown	0	2
2	Divorced	95	137
3	Widow	18	49
4	Single	218	190
5	Relationship	646	741

Total rows: 5 of 5 Query complete 00:00:00.437 Ln 7, Col 25

--What is the profile for customers who have complained?

```
SELECT
    customer_id, age, marital_status, income, teenhome, kidhome, recency, complain
  FROM
    marketing_data
 WHERE
    complain='true'
 GROUP BY
    marital_status, customer_id
 ORDER BY
    income DESC;
```

	customer_id [PK] integer	age integer	marital_status character varying (20)	income numeric	teenhome integer	kidhome integer	recency integer	complain boolean
1	4427	18	Single	83257	0	0	56	true
2	7966	54	Single	80982	1	1	48	true
3	2875	35	Relationship	67023	0	0	93	true
4	7192	58	Relationship	65748	1	0	58	true
5	6142	70	Relationship	65073	0	0	65	true
6	9500	54	Relationship	58113	1	0	66	true
7	9483	54	Relationship	57957	1	0	24	true
8	359	63	Relationship	48070	1	0	33	true
9	10637	49	Relationship	40800	2	1	77	true
10	3710	29	Single	39684	0	1	41	true

Total rows: 20 of 20 Query complete 00:00:00.212 Ln 20, Col 14

```

-- What is the average recency days value of customers who have complained?

SELECT ROUND(AVG(recency)) AS recency_days_of_complainers
from (SELECT
    customer_id, age, marital_status, income, teenhome, kidhome, recency, complain
FROM
    marketing_data
WHERE
    complain='true'
GROUP BY
    marital_status, customer_id
ORDER BY
    income DESC);

```

The screenshot shows a SQL query execution interface. At the top, there is a toolbar with various icons. Below the toolbar, the query is displayed. The result set consists of one row with the column name 'recency_days_of_complainers' and its value '51'. At the bottom of the interface, a status bar displays 'Total rows: 1 of 1' and 'Query complete 00:00:00.081 Ln 34, Col 15'.

	recency_days_of_complainers
1	51

```

36 -- What is the average recency days value of customers who have not complained?
37
38 v SELECT ROUND(AVG(recency)) AS recency_days_of_complainers
39   from (SELECT
40     customer_id, age, marital_status, income, teenhome, kidhome, recency, complain
41   FROM
42     marketing_data
43 WHERE
44     complain='false'
45 GROUP BY
46     marital_status, customer_id
47 ORDER BY
48     income DESC);
49

```

The screenshot shows a SQL query execution interface. At the top, there is a toolbar with various icons. Below the toolbar, the query is displayed. The result set consists of one row with the column name 'recency_days_of_complainers' and its value '49'. At the bottom of the interface, a status bar displays 'Total rows: 1 of 1' and 'Query complete 00:00:00.107 Ln 48, Col 15'.

	recency_days_of_complainers
1	49

```

i1 -- What is the profile for customers from ME?
i2
i3 ✓ SELECT *
i4   FROM marketing_data
i5   WHERE country = 'ME';
i6
i7 -- Total amount spent by ME customers for each product category.
i8
i9 ✓ SELECT
i10    SUM(amtliq) AS me_liq,
i11    SUM(amtvege) AS me_veg,
i12    SUM(amtnonveg) AS me_nonveg,
i13    SUM(amtpes) AS me_pes,
i14    SUM(amtchocolates) AS me_choc,
i15    SUM(amtcomm) AS me_comm
i16   FROM marketing_data
i17   WHERE country = 'ME';
i18
i19 ✓ */
i20 -- Average spend per country
i21
i22 SELECT country, SUM(amtliq+amtvege+amtnonveg+amtpes+amtchocolates+amtcomm)/COUNT(customer_id) AS average_spend
i23   from marketing_data
i24   GROUP BY country
i25   ORDER BY average_spend DESC;

```

Data Output Messages Notifications

	customer_id	age	education	marital_status	income	kidhome	teenhome	dt_customer	recency	amtliq	amt
	[PK] integer	integer	character varying (20)	character varying (20)	numeric	integer	integer	date	integer	numeric	num
1	2920	38	PhD	Single	52614	0	1	2012-01-12	63	789	
2	5080	20	Graduation	Single	70515	0	0	2013-10-21	12	420	
3	9323	64	Master	Relationship	49912	0	1	2012-07-09	5	520	

Total rows: 3 of 3 Query complete 00:00:00.104 Ln 55, Col 22

≡+ | 📁 | ⏮ | 🗂 | ⏴ | 🗃 | 🗁 | 🗂 | ⏵ | 🗃 | ⏴ | SQL

	me_liq	me_veg	me_nonveg	me_pes	me_choc	me_comm
	numeric	numeric	numeric	numeric	numeric	numeric
1	1729	8	817	226	122	220

Total rows: 1 of 1 Query complete 00:00:00.120 Ln 67, Col 22

≡+ | 📁 | ⏮ | 🗂 | ⏴ | 🗃 | ⏵ | 🗂 | ⏴ | SQL

	country	average_spend
	character varying (20)	numeric
1	ME	1040.6666666666666667
2	US	631.2710280373831776
3	GER	631.0172413793103448
4	SA	629.8119402985074627
5	CA	629.3345864661654135
6	SP	602.2930402930402930
7	AUS	582.1496598639455782
8	IND	532.4726027397260274

Total rows: 8 of 8 Query complete 00:00:00.110 Ln 75, Col 29

Effectiveness of advertising channels and promotion

```

2 -- Which is the most effective method of advertising in each country?
3 -- (In this case, consider the total number of lead conversions as a measure of effectiveness).
4
5   SELECT
6     country,
7       SUM(bulkmail_ad) AS bulkmail,
8       SUM(twitter_ad) AS twitter,
9       SUM(instagram_ad) AS instagram,
10      SUM/facebook_ad AS facebook,
11      SUM(brochure_ad) AS brochure,
12      SUM(bulkmail_ad+brochure_ad) AS total_printed_lead_conversions,
13      SUM(twitter_ad+instagram_ad+facebook_ad) AS total_digital_lead_conversion,
14      SUM(bulkmail_ad+twitter_ad+instagram_ad+facebook_ad+ brochure_ad) AS total_lead_conversions
15  FROM
16    marketing_data
17  JOIN
18    ad_data
19  ON
20    marketing_data.customer_id = ad_data.customer_id
21  GROUP BY
22    country
23  ORDER BY
24    total_lead_conversions DESC;
25
26
27
28

```

Data Output Messages Notifications

	country character varying (20)	bulkmail bigint	twitter bigint	instagram bigint	facebook bigint	brochure bigint	total_printed_lead_conversions bigint	total_digital_lead_conversion bigint	total_lead_conversions bigint
1	SP	83	87	88	76	16	99	251	350
2	CA	18	24	21	18	6	24	63	87
3	SA	21	20	21	20	4	25	61	86
4	GER	10	11	8	7	2	12	26	38
5	IND	13	10	6	7	2	15	23	38
6	AUS	9	6	12	7	0	9	25	34
7	US	8	6	5	7	0	8	18	26
8	ME	1	0	0	0	0	1	0	1

Total rows: 8 of 8 Query complete 00:00:00.095 Ln 24, Col 30

```

29 -- Which is the most effective method of advertising based on marital status?
30 -- (In this case, consider the total number of lead conversions as a measure of effectiveness).
31
32   SELECT
33     marital_status,
34       SUM(bulkmail_ad) AS bulkmail,
35       SUM(twitter_ad) AS twitter,
36       SUM(instagram_ad) AS instagram,
37       SUM/facebook_ad AS facebook,
38       SUM(brochure_ad) AS brochure,
39       SUM(bulkmail_ad+brochure_ad) AS total_printed_lead_conversions,
40       SUM(twitter_ad+instagram_ad+facebook_ad) AS total_digital_lead_conversion,
41       SUM(bulkmail_ad+twitter_ad+instagram_ad+facebook_ad+ brochure_ad) AS total_lead_conversions
42  FROM
43    marketing_data
44  JOIN
45    ad_data
46  ON
47    marketing_data.customer_id = ad_data.customer_id
48  GROUP BY
49    marital_status
50  ORDER BY
51    total_lead_conversions DESC;
52

```

Data Output Messages Notifications

	marital_status character varying (20)	bulkmail bigint	twitter bigint	instagram bigint	facebook bigint	brochure bigint	total_printed_lead_conversions bigint	total_digital_lead_conversion bigint	total_lead_conversions bigint
1	Relationship	100	104	109	94	19	119	307	426
2	Single	39	32	31	30	5	44	93	137
3	Divorced	20	18	13	12	5	25	43	68
4	Widow	4	10	7	5	1	5	22	27
5	Unknown	0	0	1	1	0	0	2	2

Total rows: 5 of 5 Query complete 00:00:00.120 Ln 51, Col 30

```

55 -- Assignment Activity 5(b)(iii)
56 -- Which social media platform(s) seem to be the most effective per country?
57 -- (In this case, assume that purchases were in some way influenced by lead conversions from a campaign).
58
59 ✓ SELECT
60     m.country, m.response,
61     SUM(a.twitter_ad) AS twitter,
62     SUM(a.instagram_ad) AS instagram,
63     SUM(a.facebook_ad) AS facebook,
64     SUM(m.amtliq) AS alcohol_spend,
65     SUM(m.amtvege) AS veg_spend,
66     SUM(m.amtnonveg) AS non_veg_spend,
67     SUM(m.amtpes) AS fish_spend,
68     SUM(m.amtchocolates) AS chocolates_spend,
69     SUM(m.amtcomm) AS commodity_spend,
70     SUM(m.amtliq+m.amtvege+m.amtnonveg+m.amtpes+m.amtchocolates+m.amtcomm+a.twitter_ad+a.instagram_ad+a.facebook_ad) AS tot
71 FROM
72     marketing_data m
73 JOIN
74     ad_data a
75 ON
76     m.customer_id = a.customer_id
77 WHERE
78     response = True
79 GROUP BY
80     country, response
81 ORDER BY total_sales DESC;

```

Data Output Messages Notifications

	country	response	twitter	instagram	facebook	alcohol_spend	veg_spend	non_veg_spend	fish_spend	chocolates_spend	commodity_spend	total_sales
1	SP	true	37	48	42	88256	6979	51761	8713	6881	10300	173017
2	SA	true	5	14	12	23664	1897	14034	2407	2076	3650	47759
3	CA	true	10	13	11	22788	1438	13083	2051	1437	2379	43210
4	AUS	true	3	7	5	9419	730	5588	1624	856	1629	19861
5	GER	true	2	3	4	10275	691	5954	825	478	805	19037
6	US	true	2	3	2	6387	459	3127	908	504	512	11904
7	IND	true	3	3	3	5642	432	3602	479	433	940	11537
8	ME	true	0	0	0	940	8	675	214	113	182	2132

Total rows: 8 of 8 Query complete 00:00:00.110 Ln 81, Col 27

Popularity of product categories

```

1 -- What is the total spend per country?
2 -- What is the total spend per product category per country?
3 -- Which products are the most popular in each country?
4
5 v SELECT
6   country,
7   SUM(amtliq) AS spend_alcohol,
8   SUM(amtvege) AS spend_vegetables,
9   SUM(amtnonveg) AS spend_meat,
10  SUM(amtpes) AS spend_fish,
11  SUM(amtchocolates) AS spend_chocolates,
12  SUM(amtcomm) AS spend_commodities,
13  SUM(amtliq + amtvege + amtnonveg + amtpes + amtchocolates + amtcomm) AS total_spend,
14  GREATEST(
15    SUM(amtliq),
16    SUM(amtvege),
17    SUM(amtnonveg),
18    SUM(amtpes),
19    SUM(amtchocolates),
20    SUM(amtcomm)
21  ) AS highest_category_total
22 FROM
23   marketing_data
24 GROUP BY
25   country
26 ORDER BY
27   highest_category_total DESC;
28
29 .. */

```

Data Output Messages Notifications

	country character varying (20)	spend_alcohol numeric	spend_vegetables numeric	spend_meat numeric	spend_fish numeric	spend_chocolates numeric	spend_commodities numeric	total_spend numeric	highest_category_total numeric
1	SP	335637	28144	177847	40049	30070	45957	657704	335637
2	SA	105901	8923	58375	13655	9018	15115	210987	105901
3	CA	84066	7681	45925	9980	7607	12144	167403	84066
4	AUS	42752	3689	22328	5546	4129	7132	85576	42752
5	GER	36776	2980	20272	4601	2801	5768	73198	36776
6	IND	36221	3782	23721	4811	3217	5989	77741	36221
7	US	32214	3034	20185	4411	2863	4839	67546	32214
8	ME	1729	8	817	226	122	220	3122	1729

Total rows: 8 of 8 Query complete 00:00:00.091 Ln 27, Col 30

-- Which products are the most popular based on marital status?

```

31
32
33 SELECT
34   marital_status,
35   SUM(amtliq) AS spend_alcohol,
36   SUM(amtvege) AS spend_vegetables,
37   SUM(amtnonveg) AS spend_meat,
38   SUM(amtpes) AS spend_fish,
39   SUM(amtchocolates) AS spend_chocolates,
40   SUM(amtcomm) AS spend_commodities,
41   SUM(amtliq + amtvege + amtnonveg + amtpes + amtchocolates + amtcomm) AS total_spend,
42   GREATEST(
43     SUM(amtliq),
44     SUM(amtvege),
45     SUM(amtnonveg),
46     SUM(amtpes),
47     SUM(amtchocolates),
48     SUM(amtcomm)
49  ) AS highest_category_total
50 FROM
51   marketing_data
52 GROUP BY
53   marital_status
54 ORDER BY
55   highest_category_total DESC;
56
57 .. */

```

Data Output Messages Notifications

	marital_status character varying (20)	spend_alcohol numeric	spend_vegetables numeric	spend_meat numeric	spend_fish numeric	spend_chocolates numeric	spend_commodities numeric	total_spend numeric	highest_category_total numeric
1	Relationship	432927	36435	232682	52666	37892	61237	853839	432927
2	Single	137763	12852	87138	18278	12772	20476	289279	137763
3	Divorced	75349	6357	34840	8123	6218	10714	141601	75349
4	Widow	27902	2422	14085	3793	2878	4245	55325	27902
5	Unknown	1355	175	725	419	67	492	3233	1355

Total rows: 5 of 5 Query complete 00:00:00.133 Ln 55, Col 30

```

-- Which products are the most popular based on whether or not there are teens in the home?
59
60
61 SELECT
62     CASE
63         WHEN teenhome=0 THEN 'no teens'
64         WHEN teenhome=1 THEN 'one teen'
65         ELSE 'two teenagers'
66     END AS teens,
67     SUM(amtlia) AS spend_alcohol,
68     SUM(amtvege) AS spend_vegetables,
69     SUM(amtnonveg) AS spend_meat,
70     SUM(amtpes) AS spend_fish,
71     SUM(amtchocolates) AS spend_chocolates,
72     SUM(amtcomm) AS spend_commodities,
73     SUM(amtlia + amtvege + amtnonveg + amtpes + amtchocolates + amtcomm) AS total_spend
74
FROM
    marketing_data
75
GROUP BY
76     teens
77
ORDER BY
78     total_spend DESC;
79
80
81
82
83
84
85

```

Data Output Messages Notifications

	teens	spend_alcohol	spend_vegetables	spend_meat	spend_fish	spend_chocolates	spend_commodities	total_spend
	text	numeric	numeric	numeric	numeric	numeric	numeric	numeric
1	no teens	350161	38082	259348	55868	38413	51484	793356
2	one teen	307052	19268	103806	26220	20628	43363	520337
3	two teenagers	18083	891	6316	1191	786	2317	29584

Total rows: 3 of 3 Query complete 00:00:00.073 Ln 79, Col 19

```

-- Which products are the most popular based on whether or not there are kids in the home?
88
89
90 SELECT
91     CASE
92         WHEN kidhome=0 THEN 'no kids'
93         WHEN kidhome=1 THEN 'one kid'
94         ELSE 'two kids'
95     END AS kids,
96     SUM(amtlia) AS spend_alcohol,
97     SUM(amtvege) AS spend_vegetables,
98     SUM(amtnonveg) AS spend_meat,
99     SUM(amtpes) AS spend_fish,
100    SUM(amtchocolates) AS spend_chocolates,
101    SUM(amtcomm) AS spend_commodities,
102    SUM(amtlia + amtvege + amtnonveg + amtpes + amtchocolates + amtcomm) AS total_spend
103
FROM
    marketing_data
104
GROUP BY
105     kids
106
ORDER BY
107     total_spend DESC;
108
109 */
110
111
112
113

```

Data Output Messages Notifications

	kids	spend_alcohol	spend_vegetables	spend_meat	spend_fish	spend_chocolates	spend_commodities	total_spend
	text	numeric	numeric	numeric	numeric	numeric	numeric	numeric
1	no kids	579458	50407	324552	72084	52014	76812	1155327
2	one kid	92526	7556	43534	10879	7634	19576	181705
3	two kids	3312	278	1384	316	179	776	6245

Total rows: 3 of 3 Query complete 00:00:00.079 Ln 108, Col 19

RFM Analysis

```
1 -- RFM analysis
2
3 -- Which country has the highest average RFM score?
4
5 SELECT
6     m.country,
7     ROUND(AVG(r.recenty_score), 2) AS avg_recency_score,
8     ROUND(AVG(r.frequency_score), 2) AS avg_frequency_score,
9     ROUND(AVG(r.monetary_score), 2) AS avg_monetary_score,
10    ROUND(AVG(r.rfm_score), 2) AS avg_rfm_score
11
12 FROM
13     marketing_data m
14 JOIN
15     rfm_score r ON m.customer_id = r.customer_id
16 GROUP BY
17     m.country
18 ORDER BY
19     avg_rfm_score DESC;
20
21
22
23
24
25
26
27
28
29
```

Data Output Messages Notifications

The screenshot shows a database query interface with the following details:

- Toolbar:** Includes icons for New, Open, Save, Print, Copy, Paste, Delete, Import, Export, Refresh, and SQL.
- Query Editor:** Displays the SQL code for the RFM analysis.
- Results:** A table showing the average RFM scores by country. The table has columns: country, avg_recency_score, avg_frequency_score, avg_monetary_score, and avg_rfm_score.

	country character varying (20)	avg_recency_score numeric	avg_frequency_score numeric	avg_monetary_score numeric	avg_rfm_score numeric
1	ME	3.33	2.67	3.33	3.00
2	CA	2.62	2.43	2.54	2.56
3	US	2.44	2.50	2.62	2.55
4	AUS	2.67	2.34	2.46	2.54
5	SA	2.50	2.42	2.54	2.51
6	SP	2.49	2.37	2.47	2.47
7	GER	2.46	2.46	2.51	2.46
8	IND	2.43	2.23	2.42	2.34

Total rows: 8 of 8 Query complete 00:00:00.106 Ln 28, Col 2

```

30 -- Which marital_status has the highest average RFM score?
31
32 SELECT
33     m.marital_status,
34     ROUND(AVG(r.reency_score), 2) AS avg_recency_score,
35     ROUND(AVG(r.frequency_score), 2) AS avg_frequency_score,
36     ROUND(AVG(r.monetary_score), 2) AS avg_monetary_score,
37     ROUND(AVG(r.rfm_score), 2) AS avg_rfm_score
38
39 FROM
40     marketing_data m
41 JOIN
42     rfm_score r ON m.customer_id = r.customer_id
43 GROUP BY
44     m.marital_status
45 ORDER BY
46     avg_rfm_score DESC;
47
48
49
50
51
52
53
54
55
56

```

Data Output Messages Notifications

	marital_status character varying (20)	avg_recency_score numeric	avg_frequency_score numeric	avg_monetary_score numeric	avg_rfm_score numeric
1	Unknown	3.25	2.75	3.50	3.00
2	Widow	2.51	2.58	2.80	2.63
3	Relationship	2.52	2.39	2.48	2.50
4	Divorced	2.47	2.42	2.48	2.48
5	Single	2.49	2.32	2.51	2.44

Total rows: 5 of 5 Query complete 00:00:00.085 Ln 45, Col 21

```

56
57 -- Which level of education has the highest average RFM score?
58
59 SELECT
60   m.education,
61   ROUND(AVG(r.reency_score), 2) AS avg_recency_score,
62   ROUND(AVG(r.frequency_score), 2) AS avg_frequency_score,
63   ROUND(AVG(r.monetary_score), 2) AS avg_monetary_score,
64   ROUND(AVG(r.rfm_score), 2) AS avg_rfm_score
65
FROM
66   marketing_data m
67
JOIN
68   rfm_score r ON m.customer_id = r.customer_id
69
GROUP BY
70   m.education
71
ORDER BY
72   avg_rfm_score DESC;
73
74

```

Data Output Messages Notifications

	education character varying (20)	avg_recency_score numeric	avg_frequency_score numeric	avg_monetary_score numeric	avg_rfm_score numeric
1	PhD	2.56	2.51	2.65	2.61
2	Master	2.58	2.41	2.47	2.52
3	Graduation	2.47	2.40	2.52	2.48
4	2n Cycle	2.50	2.27	2.32	2.37
5	Basic	2.52	1.20	1.39	1.74

Total rows: 5 of 5 Query complete 00:00:00.096 Ln 72, Col 21

```

86 -- Average RFM score for customers with teens at home.
87
88 SELECT
89   m.teenhome,
90   ROUND(AVG(r.reency_score), 2) AS avg_recency_score,
91   ROUND(AVG(r.frequency_score), 2) AS avg_frequency_score,
92   ROUND(AVG(r.monetary_score), 2) AS avg_monetary_score,
93   ROUND(AVG(r.rfm_score), 2) AS avg_rfm_score
94
FROM
95   marketing_data m
96
JOIN
97   rfm_score r ON m.customer_id = r.customer_id
98
GROUP BY
99   m.teenhome
100
ORDER BY
101   avg_rfm_score DESC;
102

```

Data Output Messages Notifications

	teenhome integer	avg_recency_score numeric	avg_frequency_score numeric	avg_monetary_score numeric	avg_rfm_score numeric
1	1	2.52	2.50	2.43	2.52
2	2	2.12	2.65	2.57	2.49
3	0	2.52	2.27	2.55	2.46

Total rows: 3 of 3 Query complete 00:00:00.100 Ln 101, Col 21

```

103
104    -- Average RFM score for customers with kids at home.
105
106    SELECT
107        m.kidhome,
108        ROUND(AVG(r.reency_score), 2) AS avg_reency_score,
109        ROUND(AVG(r.frequency_score), 2) AS avg_frequency_score,
110        ROUND(AVG(r.monetary_score), 2) AS avg_monetary_score,
111        ROUND(AVG(r.rfm_score), 2) AS avg_rfm_score
112    FROM
113        marketing_data m
114    JOIN
115        rfm_score r ON m.customer_id = r.customer_id
116    GROUP BY
117        m.kidhome
118    ORDER BY
119        avg_rfm_score DESC;
120
121
122

```

Data Output Messages Notifications

	kidhome integer	avg_reency_score numeric	avg_frequency_score numeric	avg_monetary_score numeric	avg_rfm_score numeric
1	0	2.52	2.87	3.07	2.85
2	1	2.51	1.72	1.71	1.99
3	2	2.28	1.61	1.61	1.96

Total rows: 3 of 3 Query complete 00:00:00.082 Ln 119, Col 21

```

127    -- Who are the customers with an RFM score of 4?
128
129    SELECT
130        m.customer_id,
131        m.age,
132        m.income,
133        m.marital_status,
134        m.education,
135        m.country,
136        m.teenhome,
137        m.kidhome,
138        r.rfm_score
139    FROM
140        marketing_data m
141    JOIN
142        rfm_score r ON m.customer_id = r.customer_id
143    GROUP BY
144        m.customer_id, r.rfm_score
145    HAVING
146        rfm_score = 4;
147
148
149
150
151
152
153

```

Data Output Messages Notifications

	customer_id integer	age integer	income numeric	marital_status character varying (20)	education character varying (20)	country character varying (20)	teenhome integer	kidhome integer	rfm_score numeric
1	9597	44	73448	Relationship	Graduation	CA	0	0	4
2	6521	59	77972	Relationship	Graduation	SA	0	0	4
3	2021	38	61456	Relationship	Graduation	SP	1	0	4
4	9723	53	67716	Relationship	Graduation	CA	1	0	4
5	5303	39	79632	Relationship	PhD	SA	0	0	4
6	2807	44	56796	Relationship	Graduation	SP	1	0	4
7	4785	43	77622	Relationship	PhD	SA	2	0	4
8	2429	59	72071	Divorced	Graduation	CA	1	0	4
9	8318	34	90300	Relationship	Graduation	SP	0	0	4
10	2407	31	69109	Relationship	Graduation	US	0	0	4

Total rows: 247 of 247 Query complete 00:00:00.081 Ln 146, Col 19

```

156 -- What is the average age of the customers with an RFM score of 4?
157
158 SELECT
159     ROUND(AVG(m.age),2) AS avg_age_of_high_rfmm_customers
160 FROM
161     marketing_data m
162 JOIN
163     rfm_score r ON m.customer_id = r.customer_id
164 WHERE
165     rfm_score = 4;
166

```

Data Output Messages Notifications

	avg_age_of_high_rfmm_customers	numeric
1		45.72

Total rows: 1 of 1 Query complete 00:00:00.076 Ln 166 Col 10

```

184
185 -- What is the average income of the customers with an RFM score of 4?
186
187 SELECT
188     ROUND(AVG(m.income),2) AS avg_income_of_high_rfmm_customers
189 FROM
190     marketing_data m
191 JOIN
192     rfm_score r ON m.customer_id = r.customer_id
193 WHERE
194     rfm_score = 4;
195

```

Data Output Messages Notifications

	avg_income_of_high_rfmm_customers	numeric
1		71533.90

Total rows: 1 of 1 Query complete 00:00:00.069 Ln 194, Col 19

```

202 -- Average rfm score for customers with successful lead conversion via brochure_ad
203
204 < SELECT
205     ROUND(AVG(r.rfm_score), 2) AS avg_rfm
206     FROM
207         rfm_score r
208     JOIN
209         ad_data a
210     ON
211         r.customer_id = a.customer_id
212     WHERE
213         brochure_ad = 1;
214
215

```

Data Output Messages Notifications

avg_rfm	
numeric	lock
1	3.10

Total rows: 1 of 1 Query complete 00:00:00.083 Ln 213, Col 18

```

216 -- Average RFM score by every country based on lead conversion success/failure per ad.
217
218 < SELECT
219     m.country, a.brochure_ad,
220     ROUND(AVG(r.rfm_score), 2) AS avg_rfmscore
221     FROM
222         marketing_data m
223     JOIN
224         rfm_score r ON m.customer_id = r.customer_id
225     JOIN
226         ad_data a ON m.customer_id = a.customer_id
227     WHERE
228         brochure_ad = 1
229     GROUP BY
230         m.country, a.brochure_ad
231     ORDER BY
232         avg_rfmscore DESC;

```

Data Output Messages Notifications

	country	brochure_ad	avg_rfmscore
	character varying (20)	integer	numeric
1	SP	1	3.25
2	GER	1	3.00
3	IND	1	3.00
4	SA	1	3.00
5	CA	1	2.83

Total rows: 5 of 5 Query complete 00:00:00.079 Ln 232, Col 21

```

1 ✓ SELECT
2     r.rfm_score,
3     AVG(m.amtliq) AS avg_spend_alcohol,
4     AVG(m.amtvege) AS avg_spend_vegetables,
5     AVG(m.amtnonveg) AS avg_spend_meat,
6     AVG(m.amtpes) AS avg_spend_fish,
7     AVG(m.amtchocolates) AS avg_spend_chocolates,
8     AVG(m.amtcomm) AS avg_spend_commodities,
9     AVG(m.amtliq + m.amtvege + m.amtnonveg + m.amtpes + m.amtchocolates + m.amtcomm) AS avg_total_spend,
L0    GREATEST(
L1        AVG(m.amtliq),
L2        AVG(m.amtvege),
L3        AVG(m.amtnonveg),
L4        AVG(m.amtpes),
L5        AVG(m.amtchocolates),
L6        AVG(m.amtcomm)
L7    ) AS highest_category_avg
L8 FROM
L9     marketing_data m
L10 JOIN
L11     rfm_score r
L12 ON
L13     m.customer_id = r.customer_id
L14 GROUP BY
L15     r.rfm_score;

```

Data Output Messages Notifications

	rfm_score	avg_spend_alcohol	avg_spend_vegetables	avg_spend_meat	avg_spend_fish	avg_spend_chocolates	avg_spend_commodities	avg_total_spend	highest_category_avg
	numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric
1	1	16.1758620689655172	2.9379310344827586	11.5758620689655172	4.2931034482758621	3.1344827586206897	9.5931034482758621	47.7103448275862069	16.1758620689655172
2	2	106.9202988792029888	10.7559153175591532	60.8468244084682441	15.333748443374844	11.4246575342465753	25.8231631382316314	231.1046077210460772	106.9202988792029888
3	4	665.6275303643724696	52.4979757085020243	346.5546558704453441	71.5991902834008097	54.4988785425101215	71.6356275303643725	1262.4048582995951417	665.6275303643724696
4	3	482.0389908256880734	41.0378440366972477	265.6582568807339450	59.6743119266055046	41.6112385321100917	64.1651376146788991	954.1857798165137615	482.0389908256880734

Total rows: 4 of 4 Query complete 00:00:00.097 Ln 25, Col 17

```

1 ✓ create or replace view tableau.customer_rfm_4 as
2 SELECT
3     m.customer_id,
4     m.age,
5     m.income,
6     m.marital_status,
7     m.education,
8     m.country,
9     m.teenhome,
10    m.kidhome,
11    m.amtliq,
12    m.amtvege,
13    m.amtnonveg,
14    m.amtpes,
15    m.amtchocolates,
16    m.amtcomm,
17    m.response,
18    m.count_success,
19    r.rfm_score
20 FROM
21     marketing_data m
22 JOIN
23     rfm_score r ON m.customer_id = r.customer_id
24 GROUP BY
25     m.customer_id, r.rfm_score
26 HAVING
27     r.rfm_score = 4;
28

```

Data Output Messages Notifications

	customer_id	age	income	marital_status	education	country	teenhome	kidhome	amtliq	amtvege	amtnonveg	amtpes	amtchocolates	amtcomm	response
	integer	integer	numeric	character varying (20)	character varying (20)	character varying (20)	integer	integer	numeric	numeric	numeric	numeric	numeric	numeric	boolean
1	9597	44	73448	Relationship	Graduation	CA	0	0	236	106	189	23	41	130	false
2	6521	59	77972	Relationship	Graduation	SA	0	0	613	22	319	33	102	12	false
3	2021	38	61456	Relationship	Graduation	SP	1	0	563	76	384	84	192	89	false
4	9723	53	67716	Relationship	Graduation	CA	1	0	530	142	217	62	9	56	false
5	5303	39	79632	Relationship	PhD	SA	0	0	471	0	510	99	21	12	false
6	2807	44	56796	Relationship	Graduation	SP	1	0	656	38	161	62	47	37	false
7	4785	43	77622	Relationship	PhD	SA	2	0	520	7	154	19	0	14	false
8	2429	59	72071	Divorced	Graduation	CA	1	0	531	69	300	150	138	150	false
9	8318	34	90300	Relationship	Graduation	SP	0	0	594	134	786	33	134	57	false
10	7417	21	49109	Relationship	Graduation	US	n	n	873	75	450	174	70	65	true

Total rows: 247 of 247 Query complete 00:00:00.090 Ln 30, Col 29

Analysis conducted with Tableau

In relation to the clustering analysis on the Customer Demographics dashboard, those were the Analysis of Variance statistics (all three statistically significant).

Analysis of Variance:

Variable	F-statistic	p-value	Model		Error	
			Sum of Squares	DF	Sum of Squares	DF
Average Sales	566.2	0.0	97.23	3	126.4	2208
Avg. Age	516.3	0.0	67.72	3	96.54	2208
Avg. Income	447.3	0.0	24.12	3	39.69	2208

Clusters	Number of Items	Centres		
		Avg. Age	Avg. Income	Average Sales
Cluster 1	441	35.925	73646.0	1299.8
Cluster 2	556	54.576	46132.0	273.58
Cluster 3	832	35.57	35018.0	153.83
Cluster 4	383	56.757	72248.0	1279.3
Not Clustered	0			

Dashboard 1

Marital statuses of Unknown and Widows show higher average sales. Although at a combined count of 80 customers, they are a minority.

Although those in a relationship rank last in terms of average sales, they have the highest RFM score (2.50) other than Unknowns and Widows.

PhD holders with highest average sales and RFM score.

Income ranges 61700 – 121699 (782 customers) offer exceptional readings for both average sales and RFM score

Clustering algorithm has established four distinct clusters. Selecting a data point shows the dashboard values for that customer.

Dashboard 2

Spend on alcohol generally rises with age and education level (and by extension income).

Dashboard 3

Purchase with deals tends to increase with age and education level.

Particularly good at making the most of deals of those in the age range of 31700-61699.

An examination of the uptake for the last ad campaign shows that the 333 customers who did respond positively to the campaign tend to have higher average RFM score than those who did not across all countries.

Strikingly, they were on average higher income earners and generated more sales on average for 2Market.

Dashboard 4

247customers with an RFM score of 4. More successful lead conversions via Instagram and Facebook.

872 with an RFM score of 3. More successful lead conversions via Twitter and Instagram.

Recommendations and Further Actions

More information as to how 2Market caters to the specific national and cultural needs of its transnational customer base in terms of timing and seasonality of promotions and product offerings e.g. Diwali in India, Thanksgiving in the US.

Competitor benchmarking would also have been a worthwhile exercise.

In terms of the Tableau clustering algorithm, the clusters may be added as a group (like a dimension), which may in turn be used as a forecasting tool and to visualise its association with other variables.

Number of stores per country, profit data for product categories and trackable sales information to specific occurrence of a successful lead conversion, would have enabled more precise analysis. The metadata did not provide this information. It was not possible to assume, particularly in instances whereby the count success for a customer was more than one and from different advertising channels.

Appendices

Appendix A

Data cleaning and validation conducted in Excel

- COUNTBLANK() used to determine missing values. Result was NIL.
- Spell Check performed – opted against any corrections.
- Identified **ID** column as primary key. Checked for duplicates for that column using Conditional Formatting. None were identified.
- A substantial number of records only had subtle differences. This is a recurring pattern in the data. These were not removed but would ordinarily warrant sense-checking with the business user.
- Used filters on every column to check for anomalies, data type consistency.
- Four outliers were removed:

Year_Birth 1894, 1900 and 1901 (pertaining to records 11004, 1150 and 7829 respectively). This is unlikely given that 11004 and 7829 have stated 1 for **Teenhome** and 1 for **Kidhome** respectively.

ID 9432 had an annual income of \$666,666. This is significantly higher than the next highest income of \$162,397. It does not look like a natural variation in the population.

Removal of the four records represented 0.18% of initial total. Summary statistics for age and income suggest that removal had minimal impact.

- Fixed **Income** column values with Find & Replace to remove \$ sign and convert values to number format.
- Given that date of most recent **Dt_Customer** was 06/12/2014, used 2014 as the reference year for when analysis is being undertaken. Replaced **Year_Birth** column with **Age**.
- Converted column headers in lower case in anticipation of analysis on SQL.
- Find & Replace was used to rationalise **marital_status**:

Changed Absurd (x2) & YOLO (x2) to Unknown.
Grouped Alone(x3) with Single.
Grouped Married (x857) with Together(x573) and renamed to Relationship.
Kept Divorced and Widow as these may provide insights for these distinct groups, particularly given that the Divorced status comprises of 232 entries.

- Renamed **id** column as **customer_id** for the sake of clarity.
- Assigned the values of FALSE for 0 and TRUE for 1 in respect of **response** and **complain**.
- Dealt with inconsistent date formats in **dt_customer** as follows:

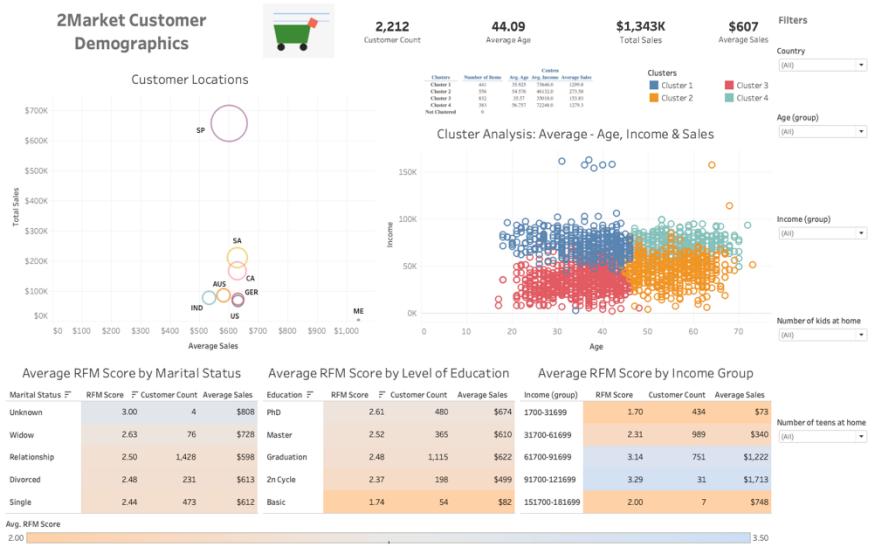
Dt_Customer	MM	DD	YY	YYYY	DATE_FN	IF_ISNUMBER
05/04/2013	41369			#N/A	#N/A	05/04/2013
6/15/14	6	15	14	2014	15/06/2014	15/06/2014
8/18/12	8	18	12	2012	18/08/2012	18/08/2012
7/22/13	7	22	13	2013	22/07/2013	22/07/2013
09/06/2013	41434			#N/A	#N/A	09/06/2013
09/01/2013	41283			#N/A	#N/A	09/01/2013
12/03/2012	40980			#N/A	#N/A	12/03/2012
2/15/14	2	15	14	2014	15/02/2014	15/02/2014
11/16/12	11	16	12	2012	16/11/2012	16/11/2012
09/08/2012	41130			#N/A	#N/A	09/08/2012
4/24/14	4	24	14	2014	24/04/2014	24/04/2014
05/11/2014	41948			#N/A	#N/A	05/11/2014
12/07/2012	41102			#N/A	#N/A	12/07/2012
9/18/13	9	18	13	2013	18/09/2013	18/09/2013
5/26/13	5	26	13	2013	26/05/2013	26/05/2013

- In [marketing_data_cleaned_sql.csv](#), for the purpose of facilitating import of the csv into pgAdmin for the SQL element of the analysis **dt_customer** was changed to format YYYY-MM-DD.
- A primary key, **ad_customer_id**, was included for the [ad_data_cleaned.csv](#). Although this key was not utilised in the joining process on either Tableau or SQL.
- In the version of [marketing_data.csv](#) used for Tableau, [marketing_data_cleaned_tableau.csv](#), the foreign key **rfm_customer_id** was included. Although this was not required at any point during the analysis.
- The **rfm_customer_id** was set as the primary key in [rfm_score.csv](#). This file also had **customer_id** as a foreign key.

Appendix B

Dashboard Development (detailed)

Dashboard 1: 2Market Customer Demographics

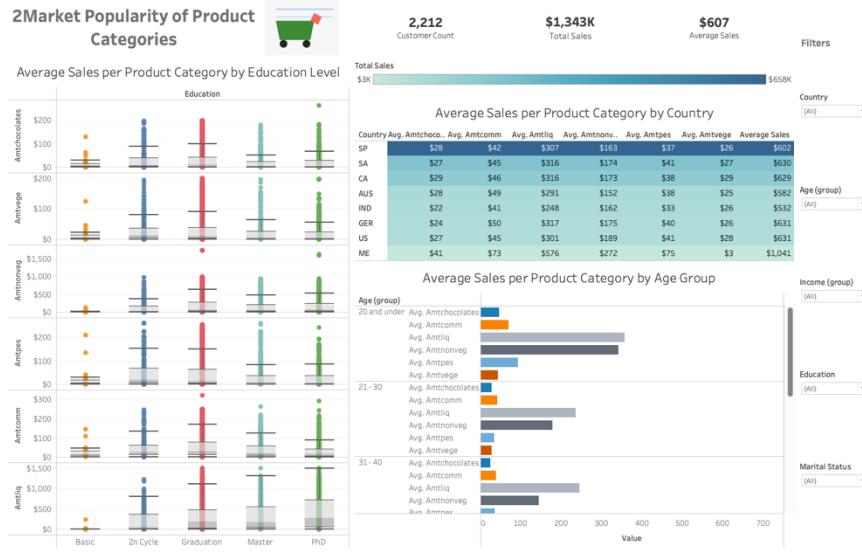


Visualisation 1: Customer Locations depicted with a scatter plot. Chosen for compact form. Data points may be used as filter.

Tables 1-3: Pertinent categorical variables. Count and average sales for each group.

Visualisation 2: Cluster Analysis using Tableau's clustering algorithm (Tableau, 2024). This creates k-number of clusters based on the data points in the data space that are closest to the cluster centres as determined by the algorithm. Chosen to depict an alternative customer segmentation technique to RFM analysis. Data points may be used as filter.

Dashboard 2: 2Market Popularity of Product Categories

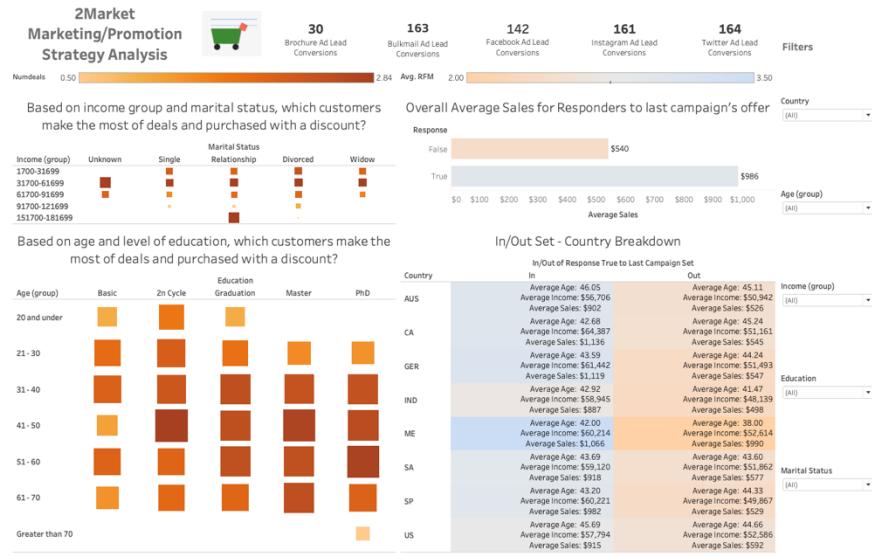


Visualisation 1: Box plot for education categories showing spend per product category. Chosen to better identify trends.

Table: Average sales by country for each product category and coloured to show intensity of total sales.

Visualisation 2: Bar chart showing spend by each age group for each product category. Chosen for ease of interpretation.

Dashboard 3: 2Market Marketing/Promotion Strategy Analysis

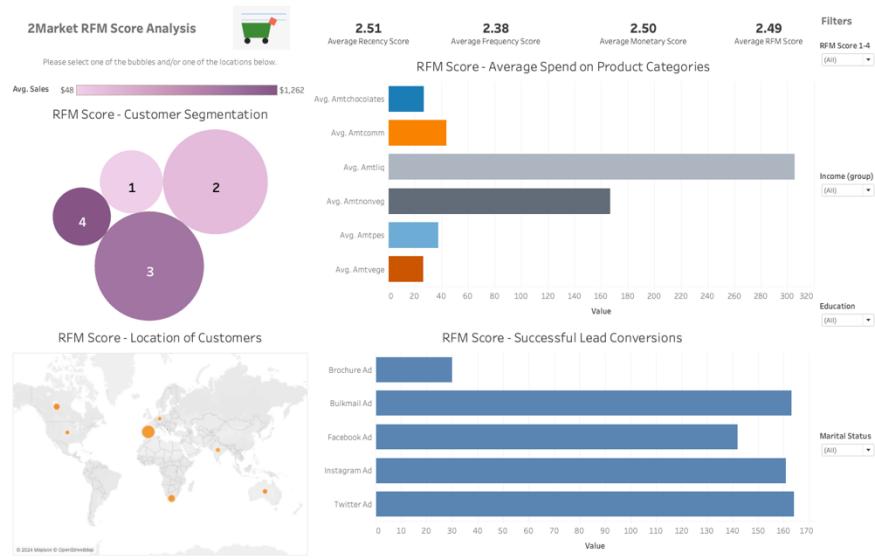


Visualisation 1-2: Heat maps to explore uptake of discount offerings. Easy to draw insights.

Visualisation 3 and Table: Analysis of last advertising campaign and set created to further analyse responders' characteristics.

Cards: Advertising Lead Conversions.

Dashboard 4: 2Market RFM Score Analysis



Visualisation 1: Bubble chart with size depicting count and label showing RFM Score. May be used as filter.

Visualisation 2: Map showing location with further information in tooltip. May be used as filter.

Visualisation 3: Average Spend on Product Categories to identify what customers of a certain RFM score tend to buy.

Visualisation 4: Bar chart showing successful lead conversions per advertising channel. Use to work out which advertising channel most efficiently reaches customers of a certain RFM score.

Cards: Recency, Frequency, Monetary and RFM average scores.

Works Cited

- Murphy, C. (2024, August 14). *What is Recency, Frequency, Monetary Value (RFM) in Marketing?* Retrieved from Investopedia: <https://www.investopedia.com/terms/r/rfm-recency-frequency-monetary-value.asp>
- Tableau. (2024). *Find Clusters in Data*. Retrieved from Tableau Desktop and Web Authoring Help: <https://help.tableau.com/current/pro/desktop/en-us/clustering.htm>
- Visual Paradigm Online. (2024). *How to Research a Topic with 5W1H Questioning Technique*. Retrieved from Visual Paradigm Online: <https://online.visual-paradigm.com/knowledge/5w1h/how-to-research-topic-with-5w1h/>