

Exploratory Analysis of 2Market's Customer Purchasing Behaviour

How can we account for the varying purchasing behaviours of 2Market's customers?

This question concerns a few key areas, which are: What are the different demographics of the customers? Which products are the most popular? Which marketing channels are the most popular?

Questions that we want to ask about the data:

- Which countries sell the most?
- What is the average age of 2Market's customers?
- Which age groups have the highest number of purchases?
- Which products are the most popular?
- Which marketing channels are the most popular?
- What patterns can we see based on marital status?
- What patterns can we see based on the number of kids and teens at home?

Data cleaning and analysis approach in Excel and PostgreSQL

There are a few different types of errors in the marketing_data dataset, which include: duplicate values, misspelled text or incorrect text, incorrect formatting and gaps in the data.

Firstly, the ID column should have unique values, and when we use 'remove duplicates' it returns none were found. However, if we use conditional formatting on the income values, we can see a pattern of duplicate data entries.

ID	Income	Age	Gender	Marital Status	Number of Kids	Number of Teens	Number of Pets	Number of Cars	Number of TVs	Number of Computers	Number of Cell Phones	Number of Tablets	Number of Smart Home Devices	Number of Social Media Accounts	Number of Email Subscriptions	Number of Website Visits	Number of Purchases
1	\$44,931.00	0	1	1/18/14	0	78	0	11	0	0	7	1	2	3	5	0	0
2	\$65,324.00	0	1	11/1/14	0	384	0	102	21	32	5	3	6	9	4	0	0
3	\$65,324.00	0	1	11/1/14	0	384	0	102	21	32	5	3	6	9	4	0	0
4	\$81,044.00	0	0	12/27/13	0	450	26	535	73	98	26	1	5	10	1	0	0
5	\$62,499.00	1	0	9/12/13	0	140	4	61	0	13	4	2	3	6	4	0	0
6	\$67,786.00	0	0	7/12/13	0	431	82	441	80	20	102	1	3	6	1	1	0
7	\$26,872.00	0	0	10/16/13	0	3	10	8	3	16	32	1	1	2	6	0	0
8	\$4,428.00	0	1	5/10/13	0	16	4	12	2	4	321	0	25	0	1	0	0
9	\$54,809.00	1	1	11/9/13	0	63	6	57	13	13	22	4	2	5	4	0	0
10	\$54,809.00	1	1	11/9/13	0	63	6	57	13	13	22	4	2	5	4	1	0
11	\$32,173.00	0	0	1/8/13	0	18	0	2	0	0	2	1	1	3	4	0	0
12	\$47,823.00	0	1	7/23/13	0	53	1	5	2	1	10	2	2	3	8	0	0
13	\$30,523.00	2	1	1/7/13	0	5	0	3	0	0	5	1	1	2	7	0	0
14	\$36,634.00	0	1	5/28/13	0	213	9	76	4	3	30	3	5	5	7	0	0
15	\$43,456.00	0	1	3/26/13	0	275	11	68	25	7	7	3	5	8	5	0	0
16	\$40,662.00	1	0	3/15/13	0	40	2	23	0	4	23	2	2	3	4	0	0
17	\$49,544.00	1	0	12/2/13	0	308	0	73	0	0	23	2	5	8	7	0	0
18	\$57,731.00	0	1	11/23/12	0	266	21	300	65	8	44	4	8	6	6	0	0
19	\$33,168.00	0	1	10/13/12	0	80	1	37	0	1	3	3	2	4	7	0	0
20	\$54,450.00	1	1	9/14/12	0	454	0	171	8	19	32	12	9	8	8	0	0
21	\$54,450.00	1	1	9/14/12	0	454	0	171	8	19	32	12	9	8	8	0	0
22	\$35,340.00	1	1	6/29/14	1	27	0	12	0	1	5	2	2	3	5	0	0
23	\$73,170.00	0	0	5/31/14	1	184	174	256	50	30	32	1	5	6	2	0	0

Figure.1

We need to delete these values as some of the surrounding information will skew the results and could lead to spurious relationships. To make the marketing_data consistent with the

ad_data, we need to delete the matching rows in the ad_data.

G	H	I	J	K	L	
nhc	Sum_Ki	Dt_Custom	Recenc	AmtLiq	AmtVe	Amt
0	0	6/16/14	0	189	104	
0	0	6/15/14	0	464	5	
1	1	5/13/14	0	134	11	
1	2	11/5/14	0	10	0	
0	1	8/4/14	0	6	16	
0	0	3/17/14	0	336	130	
0	0	1/29/14	0	769	80	
1	1	1/18/14	0	78	0	
0	0	12/27/13	0	450	26	
0	1	9/12/13	0	140	4	
0	0	7/12/13	0	431	82	
0	0	10/16/13	0	3	10	
1	1	5/10/13	0	16	4	
1	1	1/8/13	0	18	0	
1	1	7/23/13	0	53	1	
1	3	1/7/13	0	5	0	
1	1	5/28/13	0	213	9	
1	1	3/26/13	0	275	11	
0	1	3/15/13	0	40	2	

Figure.2

Some values are registered as text, whereas other values are registered as numbers. In the Income column, we can use 'find and replace' to replace the \$ symbol with a blank. We can then set the column to US currency values. Additionally, we format the date to US MM/DD/YY to match the text values.

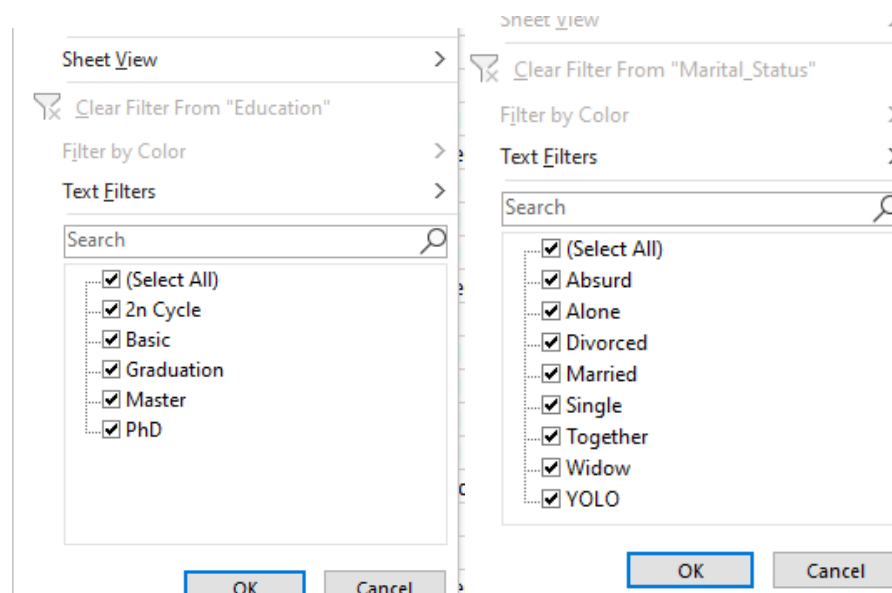


Figure.3

Spelling errors or incorrect text in Marital_Status include, 'Absurd', 'Alone' and 'YOLO'. We should delete these values and replace with '#N/A.'

As for Education there are: 'Basic', 'Graduation' and '2n Cycle'. In this case, we have removed the column based on the question: Is education indicative of purchasing behaviours?

The columns Age, Sum_Kid_Teen and Sum_Purchases have been created. While this has been done in Excel, we could also achieve the same result in PostgreSQL by aggregating the values. The age value represents the customer's current age. Since the max value of the recency of purchases (Figure 5) is 99, the current age of the customers is still indicative of their purchasing behaviour.

	Age		Sum_Kid_Teen		Sum_Purchases	
1	Mean	53.26101	Mean	0.950991	Mean	607.4791
2	Standard Error	0.281088	Standard Error	0.017406	Standard Error	14.14746
3	Median	52	Median	1	Median	396
4	Mode	46	Mode	1	Mode	55
5	Standard Deviation	11.97845	Standard Deviation	0.741764	Standard Deviation	602.8878
6	Sample Variance	143.4833	Sample Variance	0.550214	Sample Variance	363473.8
7	Kurtosis	1.058491	Kurtosis	-0.21665	Kurtosis	-0.36449
8	Skewness	0.415971	Skewness	0.40279	Skewness	0.859089
9	Range	103	Range	3	Range	2519
10	Minimum	26	Minimum	0	Minimum	5
11	Maximum	129	Maximum	3	Maximum	2524
12	Sum	96722	Sum	1727	Sum	1103182
13	Count	1816	Count	1816	Count	1816
14						
15						

Figure.4

Z	AA	AB	AC	AD
Recency Max	99			

Figure.5

The max age in the dataset is 129 years old. We can omit anomalies such as this, as it does not represent the customer population. We can see that the range of kids and teens for the customer population is between 0 and 3 and the most common number of kids and/or teens is 1. Sum_Purchases has wide range (2519 purchases) between customers. This is an indication that popularity of 2Market varies dramatically depending on the demographic.

In PostgreSQL, we need to create the tables for the cleaned marketing and advert data:

```
1  -- Create the tables
2  --1st table: marketing data
3
4  create table marketing_data_cleaned (
5  ID integer primary key,
6  Year_Birth integer,
7  Age integer,
8  Marital_Status varchar(50),
9  Income integer,
10 Kidhome integer,
11 Teenhome integer,
12 Sum_Kid_Teen integer,
13 Recency integer,
14 AmtLiq integer,
15 AmtVege integer,
16 AmtMeat integer,
17 AmtFish integer,
18 AmtChocolates integer,
19 AmtComm integer,
20 Sum_Purchases integer,
21 NumDeals integer,
22 NumWebBuy integer,
23 NumWalkinPur integer,
24 NumVisits integer,
25 Response boolean,
26 Complain boolean,
27 Country varchar(50),
28 Count_Success integer);
29
```

Figure.6

```
9
10 --2nd table: ad data
11 create table ad_data_cleaned (
12 ID integer primary key,
13 Bulkmail_ad boolean,
14 Twitter_ad boolean,
15 Instagram_ad boolean,
16 Facebook_ad boolean,
17 Brochure_ad boolean);
18 */
```

Figure.7

Next, we need to import the cleaned Excel CSV files into the tables, checking the columns and their respective data types to ensure we don't receive an error when importing.

We have updated the country column so that the tables display the country names rather than country code, to make the results more readable.

```
39
40 --Update country names
41 update "marketing_data_cleaned"
42 set "country" = replace("country", 'AUS', 'Australia');
43
44 update "marketing_data_cleaned"
45 set "country" = replace("country", 'CA', 'Canada');
46
47 update "marketing_data_cleaned"
48 set "country" = replace("country", 'GER', 'Germany');
49
50 update "marketing_data_cleaned"
51 set "country" = replace("country", 'IND', 'India');
52
53 update "marketing_data_cleaned"
54 set "country" = replace("country", 'ME', 'Montenegro');
55
56 update "marketing_data_cleaned"
57 set "country" = replace("country", 'SA', 'South Africa');
58
59 update "marketing_data_cleaned"
60 set "country" = replace("country", 'SP', 'Spain');
61
62 update "marketing_data_cleaned"
63 set "country" = replace("country", 'US', 'USA');
64 */
```

Figure.8

Another step taken to clean the data is to remove the remaining #N/A values left in the Marital_Status column.

```
93 --Remove #N/A from marketing data table
94 delete from "marketing_data_cleaned"
95 where "marital_status" = '#N/A';
96 */
```

Figure.9

Before we begin interrogating the data, it is useful to carry out sense-checks.

```

121 select "country",
122 conversion,
123 sum(bulkmail + twitter + instagram + facebook + brochure),
124 bulkmail,
125 twitter,
126 instagram,
127 facebook,
128 brochure
129 from (
130     select mar."country",
131     sum(mar."sum_purchases") as country_spend,
132     sum((mar."count_success")::int) as conversion,
133     sum((ad."bulkmail_ad")::int) as bulkmail,
134     sum((ad."twitter_ad")::int) as twitter,
135     sum((ad."instagram_ad")::int) as instagram,
136     sum((ad."facebook_ad")::int) as facebook,
137     sum((ad."brochure_ad")::int) as brochure
138     from public."marketing_data_cleaned" mar
139     inner join public."ad_data_cleaned" ad using ("id")
140     group by mar."country"
141 ) as a
142 group by a."country",
143 a.conversion,
144 a.bulkmail,
145 a.twitter,
146 a.instagram,
147 a.facebook,
148 a.brochure,
149 a.country_spend
150 order by country_spend desc;

```

Figure. 10

	country character varying (50)	conversion bigint	sum numeric	bulkmail bigint	twitter bigint	instagram bigint	facebook bigint	brochure bigint
1	Spain	301	259	69	65	59	55	11
2	South Africa	67	78	23	20	17	17	1
3	Canada	71	78	19	13	23	21	2
4	Australia	23	37	6	9	11	8	3
5	India	31	35	6	12	7	6	4
6	Germany	32	31	8	5	9	8	1
7	USA	21	30	6	9	7	5	3
8	Montenegro	0	0	0	0	0	0	0

Figure. 11

From the table above, we can see that the total number of successful conversions is inconsistent with the sum of the lead conversions of the different marketing channels. Therefore, we shouldn't use the Count_Success column as a comparison.

To analyse the data in SQL, the following queries are responses to the questions that we asked earlier:

```
65  --How much does each country spend?
66
67  select "country",
68  country_spend,
69  to_char(country_spend*100/sum(country_spend) over(), 'fm00D00') as percent
70  from (
71      select "country",
72      sum("sum_purchases") as country_spend
73      from public."marketing_data_cleaned"
74      group by "country"
75      order by country_spend desc
76  ) as c;
```

Figure.12

```
78  --How much does each country spend on each product?
79
80  select "country",
81  sum("amtliq") as liq,
82  sum("amtvege") as veg,
83  sum("amtmeat") as meat,
84  sum("amtfish") as fish,
85  sum("amtchocolates") as choc,
86  sum("amtcomm") as comm
87  from public."marketing_data_cleaned"
88  group by "country"
89  order by sum("sum_purchases") desc;
90  */
```

Figure.13

```
106  --Which products are the most popular based on marital status?
107  select "marital_status",
108         sum("amtliq") as liq,
109         sum("amtvege") as veg,
110         sum("amtmeat") as meat,
111         sum("amtfish") as fish,
112         sum("amtchocolates") as choc,
113         sum("amtcomm") as comm
114  from public."marketing_data_cleaned"
115  group by "marital_status";
116  */
```

Figure.14

```
107  --Which products are the most popular depending on how many kids are at home?
108
109  select "sum_kid_teen",
110         sum("amtliq") as liq,
111         sum("amtvege") as veg,
112         sum("amtmeat") as meat,
113         sum("amtfish") as fish,
114         sum("amtchocolates") as choc,
115         sum("amtcomm") as comm
116  from public."marketing_data_cleaned"
117  group by "sum_kid_teen";
118  */
```

Figure.15

```
152  --Which social media platform is the most effective per country?
153
154  select "country",
155         twitter,
156         instagram,
157         facebook
158  from (
159         select mar."country",
160                sum(mar."sum_purchases") as country_spend,
161                sum((ad."twitter_ad")::int) as twitter,
162                sum((ad."instagram_ad")::int) as instagram,
163                sum((ad."facebook_ad")::int) as facebook
164         from public."marketing_data_cleaned" mar
165         inner join public."ad_data_cleaned" ad using ("id")
166         group by mar."country") as a
167  order by country_spend desc;
168  */
```

Figure.16


```
170 --Which social media platforms are associated with the highest sales per country?
171 select mar."country",
172 sum(mar."sum_purchases") as country_spend,
173 sum((ad."twitter_ad")::int) as twitter,
174 sum((ad."instagram_ad")::int) as instagram,
175 sum((ad."facebook_ad")::int) as facebook
176 from public."marketing_data_cleaned" mar
177 left join public."ad_data_cleaned" ad using ("id")
178 group by mar."country"
179 order by country_spend desc;
180 */
```

Figure.17

```
181 --Which marketing channel is the most effective based on marital status?
182
183 select mar."marital_status",
184 sum((ad."twitter_ad")::int) as twitter,
185 sum((ad."instagram_ad")::int) as instagram,
186 sum((ad."facebook_ad")::int) as facebook
187 from public."marketing_data_cleaned" mar
188 left join public."ad_data_cleaned" ad using ("id")
189 group by mar."marital_status";
190 */
```

Figure.18

Dashboard

Metrics used in the visualisations:

- The total purchases of each country.
- The total purchases of different age groups.
- Average age of the customer population.
- Total purchases of each product type.
- Number of lead conversions for each marketing channels.

The rationale behind the design of the dashboard is that we can filter the data, such as the popularity of the products and the popularity of marketing channels, depending on which country we select. The country symbols are displayed at the top, acting as tabs.

In the middle of the dashboard is the total purchases of each age group, represented as a bar chart. We can also filter the data depending on the age range. To do this we can select an age group or highlight a range of age groups. Additionally, we can combine the age filter with the country filter to display the popularity of certain products based on both country and age. The average age is displayed in the top right corner, which changes based on the demographic selected.

For the purpose of functionality over visualisation, the decision was made to keep the size of the country tabs consistent, while displaying their respective total purchases. This is because we are focusing on the volume of purchases for each country and the values of the larger markets, such as Spain, would visually overwhelm the values of smaller markets, such as Montenegro.

The colour scheme is set up to differentiate the countries and average age from the rest of the screen. Countries are set to the automatic colour scheme, while the bar charts have the colour-blind setting applied to it. This is to emphasise the product types and marketing channels, to make them more identifiable.

Insights

	country character varying (50)	country_spend bigint	percent text
1	Spain	547652	49.75
2	South Africa	169656	15.41
3	Canada	140362	12.75
4	Australia	65619	05.96
5	India	61017	05.54
6	Germany	58113	05.28
7	USA	57044	05.18
8	Montenegro	1258	00.11

Figure.19

	country character varying (50)	liq bigint	veg bigint	meat bigint	fish bigint	choc bigint	comm bigint
1	Spain	278541	23361	149134	32912	25378	38326
2	South Africa	84811	7213	47206	11064	7279	12083
3	Canada	71625	6473	38146	7899	6557	9662
4	Australia	31892	2952	17239	4553	3292	5691
5	India	29590	2759	18440	3422	2386	4420
6	Germany	29746	2485	16002	3694	2095	4091
7	USA	25983	2714	18025	3725	2520	4077
8	Montenegro	420	0	452	182	64	140

Figure.20

From the tables above, we can see that Spain has by far the highest number of total purchases of the customer population. The Spanish market accounts for 49.75% of the total purchases. As for product purchases, alcohol and meat are consistently within the top 2 most popular products by quite large margins.

	marital_status character varying (50)	liq bigint	veg bigint	meat bigint	fish bigint	choc bigint	comm bigint
1	Together	144412	11955	79400	17138	12151	19195
2	Married	212105	19074	121007	28343	19567	29551
3	Divorced	56488	5066	30745	6475	5029	9053
4	Single	120009	10328	62575	13256	10770	18061
5	Widow	19594	1534	10917	2239	2054	2630

Figure.21

Figure 21, shows the total purchases of each product based on marital status. We can see that married couples have the highest number of purchases among the different groups. Furthermore, within each group we can see a similar pattern reoccurring with the total purchases of alcohol and meat.

Data Output Messages Notifications							
	sum_kid_teen integer	liq bigint	veg bigint	meat bigint	fish bigint	choc bigint	comm bigint
1	0	253244	26896	191653	38222	27331	32018
2	1	245226	18023	93062	25015	19032	36962
3	3	6421	287	2561	220	246	764
4	2	47717	2751	17368	3994	2962	8746

Figure.22

Figure 22, above, show the purchases of different products depending on the total number of kids and teens at home. We can see that a smaller family size tends to purchase more.

	country character varying (50)	country_spend bigint	twitter bigint	instagram bigint	facebook bigint
1	Spain	547652	65	59	55
2	South Africa	169656	20	17	17
3	Canada	140362	13	23	21
4	Australia	65619	9	11	8
5	India	61017	12	7	6
6	Germany	58113	5	9	8
7	USA	57044	9	7	5
8	Montenegro	1258	0	0	0

Figure. 23

	marital_status character varying (50)	twitter bigint	instagram bigint	facebook bigint
1	Together	31	32	29
2	Married	57	56	52
3	Divorced	14	8	9
4	Single	27	32	26
5	Widow	4	5	4

Figure. 24

Figure 23 shows the lead conversions from different social media compared to the total purchases in different countries. We can see Twitter has been the most effective in Spain, South Africa, India and USA. Additionally, social media has not affected the market in Montenegro.

In Figure 24, the different social media channels have approximately the same influence as each other.

Recommendations

An area of great interest to 2Market is Spain, as they have nearly half the total purchases of the entire customer population. We know that alcohol and meat are consistently among the top 2 total purchases, which may be an indication that the customers prefer these products over their competitors alternatives. We also know that Twitter is the most effective form of social media. The marketing team can use this to tailor more Twitter ads to the Spanish market. Moreover, the same principle can be used to try and gain outreach in other markets, such as Montenegro. Alcohol and meat could feature more often in the advertising to entice more

Angus Ratcliffe

customers to register with 2Market.

Family packages could also be considered in the marketing. We also know that 2Market is less popular for bigger families. To incentivise more people to register and buy more products, 2Market can sell package deals or 'family size' versions of their products. For example, deals on meat hampers.

Appendix

er	\$44,931.00	0	11/18/14	0	78	0	11	0	0	7	1	2	3	5	0	0	S
d	\$65,324.00	0	11/11/14	0	384	0	102	21	32	5	3	6	9	4	0	0	L
d	\$65,324.00	0	11/11/14	0	384	0	102	21	32	5	3	6	9	4	0	0	L
d	\$81,044.00	0	012/27/13	0	450	26	535	73	98	26	1	5	10	1	0	0	L
er	\$62,499.00	1	09/12/13	0	140	4	61	0	13	4	2	3	6	4	0	0	S
v	\$67,786.00	0	07/12/13	0	431	82	441	80	20	102	1	3	6	1	1	0	L
d	\$26,872.00	0	010/16/13	0	3	10	8	3	16	32	1	1	2	6	0	0	C
d	\$4,428.00	0	15/10/13	0	16	4	12	2	4	321	0	25	0	1	0	0	S
d	\$54,809.00	1	11/9/13	0	63	6	57	13	13	22	4	2	5	4	0	0	S
d	\$54,809.00	1	11/9/13	0	63	6	57	13	13	22	4	2	5	4	1	0	S
ed	\$32,173.00	0	11/8/13	0	18	0	2	0	0	2	1	1	3	4	0	0	S
	\$47,823.00	0	17/23/13	0	53	1	5	2	1	10	2	2	3	8	0	0	C
	\$30,523.00	2	11/7/13	0	5	0	3	0	0	5	1	1	2	7	0	0	C
er	\$36,634.00	0	15/28/13	0	213	9	76	4	3	30	3	5	5	7	0	0	S
	\$43,456.00	0	13/26/13	0	275	11	68	25	7	7	3	5	8	5	0	0	L
d	\$40,662.00	1	03/15/13	0	40	2	23	0	4	23	2	2	3	4	0	0	C
d	\$49,544.00	1	012/2/13	0	308	0	73	0	0	23	2	5	8	7	0	0	S
	\$57,731.00	0	11/23/12	0	266	21	300	65	8	44	4	8	6	6	0	0	L
	\$33,168.00	0	110/13/12	0	80	1	37	0	1	3	3	2	4	7	0	0	S
er	\$54,450.00	1	19/14/12	0	454	0	171	8	19	32	12	9	8	8	0	0	S
er	\$54,450.00	1	19/14/12	0	454	0	171	8	19	32	12	9	8	8	0	0	S
v	\$35,340.00	1	16/29/14	1	27	0	12	0	1	5	2	2	3	5	0	0	S
	\$73,170.00	0	05/31/14	1	184	174	256	50	30	32	1	5	6	2	0	0	C

Figure.1 – Conditional Formatting: Duplicate Values

G	H	I	J	K	L	
nhc	Sum_Ki	Dt_Custom	Recenc	AmtLiq	AmtVe	Amt
0	0	6/16/14	0	189	104	
0	0	6/15/14	0	464	5	
1	1	5/13/14	0	134	11	
1	2	11/5/14	0	10	0	
0	1	8/4/14	0	6	16	
0	0	3/17/14	0	336	130	
0	0	1/29/14	0	769	80	
1	1	1/18/14	0	78	0	
0	0	12/27/13	0	450	26	
0	1	9/12/13	0	140	4	
0	0	7/12/13	0	431	82	
0	0	10/16/13	0	3	10	
1	1	5/10/13	0	16	4	
1	1	1/8/13	0	18	0	
1	1	7/23/13	0	53	1	
1	3	1/7/13	0	5	0	
1	1	5/28/13	0	213	9	
1	1	3/26/13	0	275	11	
0	1	3/15/13	0	40	2	

Figure.2 – Correcting Data Types: Date

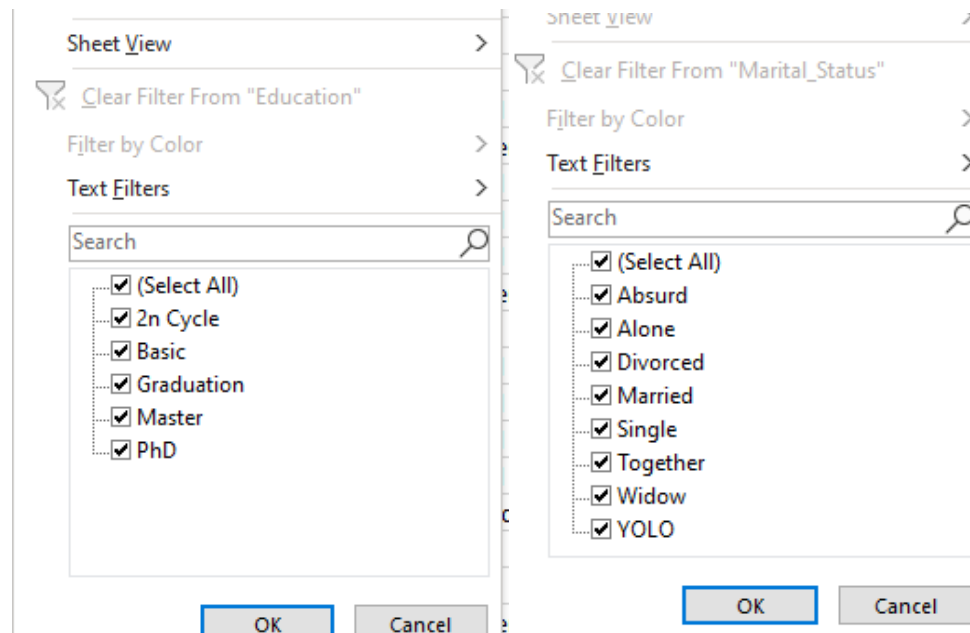


Figure.3

1	Age		Sum_Kid_Teen		Sum_Purchases	
2	Mean	53.26101	Mean	0.950991	Mean	607.4791
3	Standard Error	0.281088	Standard Error	0.017406	Standard Error	14.14746
4	Median	52	Median	1	Median	396
5	Mode	46	Mode	1	Mode	55
6	Standard Deviation	11.97845	Standard Deviation	0.741764	Standard Deviation	602.8878
7	Sample Variance	143.4833	Sample Variance	0.550214	Sample Variance	363473.8
8	Kurtosis	1.058491	Kurtosis	-0.21665	Kurtosis	-0.36449
9	Skewness	0.415971	Skewness	0.40279	Skewness	0.859089
10	Range	103	Range	3	Range	2519
11	Minimum	26	Minimum	0	Minimum	5
12	Maximum	129	Maximum	3	Maximum	2524
13	Sum	96722	Sum	1727	Sum	1103182
14	Count	1816	Count	1816	Count	1816
15						

Figure.4

Z	AA	AB	AC	A
cess	Recency Max			
	99			

Figure.5

```
1  -- Create the tables
2  --1st table: marketing data
3
4  create table marketing_data_cleaned (
5  ID integer primary key,
6  Year_Birth integer,
7  Age integer,
8  Marital_Status varchar(50),
9  Income integer,
10 Kidhome integer,
11 Teenhome integer,
12 Sum_Kid_Teen integer,
13 Recency integer,
14 AmtLiq integer,
15 AmtVege integer,
16 AmtMeat integer,
17 AmtFish integer,
18 AmtChocolates integer,
19 AmtComm integer,
20 Sum_Purchases integer,
21 NumDeals integer,
22 NumWebBuy integer,
23 NumWalkinPur integer,
24 NumVisits integer,
25 Response boolean,
26 Complain boolean,
27 Country varchar(50),
28 Count_Success integer);
29
```

Figure.6

```
30
31 --2nd table: ad data
32 create table ad_data_cleaned (
33 ID integer primary key,
34 Bulkmail_ad boolean,
35 Twitter_ad boolean,
36 Instagram_ad boolean,
37 Facebook_ad boolean,
38 Brochure_ad boolean);
39 */
```

Figure.7


```
39
40 --Update country names
41 update "marketing_data_cleaned"
42 set "country" = replace("country", 'AUS', 'Australia');
43
44 update "marketing_data_cleaned"
45 set "country" = replace("country", 'CA', 'Canada');
46
47 update "marketing_data_cleaned"
48 set "country" = replace("country", 'GER', 'Germany');
49
50 update "marketing_data_cleaned"
51 set "country" = replace("country", 'IND', 'India');
52
53 update "marketing_data_cleaned"
54 set "country" = replace("country", 'ME', 'Montenegro');
55
56 update "marketing_data_cleaned"
57 set "country" = replace("country", 'SA', 'South Africa');
58
59 update "marketing_data_cleaned"
60 set "country" = replace("country", 'SP', 'Spain');
61
62 update "marketing_data_cleaned"
63 set "country" = replace("country", 'US', 'USA');
64 */
```

Figure.8

```
93 --Remove #N/A from marketing data table
94 delete from "marketing_data_cleaned"
95 where "marital_status" = '#N/A';
96 */
```

Figure.9

```

121 select "country",
122 conversion,
123 sum(bulkmail + twitter + instagram + facebook + brochure),
124 bulkmail,
125 twitter,
126 instagram,
127 facebook,
128 brochure
129 from (
130     select mar."country",
131     sum(mar."sum_purchases") as country_spend,
132     sum((mar."count_success")::int) as conversion,
133     sum((ad."bulkmail_ad")::int) as bulkmail,
134     sum((ad."twitter_ad")::int) as twitter,
135     sum((ad."instagram_ad")::int) as instagram,
136     sum((ad."facebook_ad")::int) as facebook,
137     sum((ad."brochure_ad")::int) as brochure
138     from public."marketing_data_cleaned" mar
139     inner join public."ad_data_cleaned" ad using ("id")
140     group by mar."country"
141 ) as a
142 group by a."country",
143 a.conversion,
144 a.bulkmail,
145 a.twitter,
146 a.instagram,
147 a.facebook,
148 a.brochure,
149 a.country_spend
150 order by country_spend desc;

```

Figure. 10

	country character varying (50)	conversion bigint	sum numeric	bulkmail bigint	twitter bigint	instagram bigint	facebook bigint	brochure bigint
1	Spain	301	259	69	65	59	55	11
2	South Africa	67	78	23	20	17	17	1
3	Canada	71	78	19	13	23	21	2
4	Australia	23	37	6	9	11	8	3
5	India	31	35	6	12	7	6	4
6	Germany	32	31	8	5	9	8	1
7	USA	21	30	6	9	7	5	3
8	Montenegro	0	0	0	0	0	0	0

Figure. 11

```
65  --How much does each country spend?
66
67  select "country",
68  country_spend,
69  to_char(country_spend*100/sum(country_spend) over(), 'fm00D00') as percent
70  from (
71      select "country",
72      sum("sum_purchases") as country_spend
73      from public."marketing_data_cleaned"
74      group by "country"
75      order by country_spend desc
76  ) as c;
```

Figure.12

```
78  --How much does each country spend on each product?
79
80  select "country",
81  sum("amtliq") as liq,
82  sum("amtvege") as veg,
83  sum("amtmeat") as meat,
84  sum("amtfish") as fish,
85  sum("amtchocolates") as choc,
86  sum("amtcomm") as comm
87  from public."marketing_data_cleaned"
88  group by "country"
89  order by sum("sum_purchases") desc;
90  */
```

Figure.13

```
--Which products are the most popular based on marital status?
select "marital_status",
sum("amtliq") as liq,
sum("amtvege") as veg,
sum("amtmeat") as meat,
sum("amtfish") as fish,
sum("amtchocolates") as choc,
sum("amtcomm") as comm
from public."marketing_data_cleaned"
group by "marital_status";
*/
```

Figure.14

```

106  */
107  --Which products are the most popular depending on how many kids are at home?
108
109  select "sum_kid_teen",
110  sum("amtliq") as liq,
111  sum("amtvege") as veg,
112  sum("amtmeat") as meat,
113  sum("amtfish") as fish,
114  sum("amtchocolates") as choc,
115  sum("amtcomm") as comm
116  from public."marketing_data_cleaned"
117  group by "sum_kid_teen";
118  */

```

Figure.15

```

151  */
152  --Which social media platform is the most effective per country?
153
154  select "country",
155  twitter,
156  instagram,
157  facebook
158  from (
159      select mar."country",
160      sum(mar."sum_purchases") as country_spend,
161      sum((ad."twitter_ad")::int) as twitter,
162      sum((ad."instagram_ad")::int) as instagram,
163      sum((ad."facebook_ad")::int) as facebook
164      from public."marketing_data_cleaned" mar
165      inner join public."ad_data_cleaned" ad using ("id")
166      group by mar."country") as a
167  order by country_spend desc;
168  */

```

Figure.16

```

170  --Which social media platforms are associated with the highest sales per country?
171  select mar."country",
172  sum(mar."sum_purchases") as country_spend,
173  sum((ad."twitter_ad")::int) as twitter,
174  sum((ad."instagram_ad")::int) as instagram,
175  sum((ad."facebook_ad")::int) as facebook
176  from public."marketing_data_cleaned" mar
177  left join public."ad_data_cleaned" ad using ("id")
178  group by mar."country"
179  order by country_spend desc;
180  */

```

Figure.17

```

181  --Which marketing channel is the most effective based on marital status?
182
183  select mar."marital_status",
184  sum((ad."twitter_ad")::int) as twitter,
185  sum((ad."instagram_ad")::int) as instagram,
186  sum((ad."facebook_ad")::int) as facebook
187  from public."marketing_data_cleaned" mar
188  left join public."ad_data_cleaned" ad using ("id")
189  group by mar."marital_status";
190  */

```

Figure. 18

	country character varying (50)	country_spend bigint	percent text
1	Spain	547652	49.75
2	South Africa	169656	15.41
3	Canada	140362	12.75
4	Australia	65619	05.96
5	India	61017	05.54
6	Germany	58113	05.28
7	USA	57044	05.18
8	Montenegro	1258	00.11

Figure. 19

	country character varying (50)	liq bigint	veg bigint	meat bigint	fish bigint	choc bigint	comm bigint
1	Spain	278541	23361	149134	32912	25378	38326
2	South Africa	84811	7213	47206	11064	7279	12083
3	Canada	71625	6473	38146	7899	6557	9662
4	Australia	31892	2952	17239	4553	3292	5691
5	India	29590	2759	18440	3422	2386	4420
6	Germany	29746	2485	16002	3694	2095	4091
7	USA	25983	2714	18025	3725	2520	4077
8	Montenegro	420	0	452	182	64	140

Figure. 20

	marital_status character varying (50)	liq bigint	veg bigint	meat bigint	fish bigint	choc bigint	comm bigint
1	Together	144412	11955	79400	17138	12151	19195
2	Married	212105	19074	121007	28343	19567	29551
3	Divorced	56488	5066	30745	6475	5029	9053
4	Single	120009	10328	62575	13256	10770	18061
5	Widow	19594	1534	10917	2239	2054	2630

Figure.21

Data Output Messages Notifications							
	sum_kid_teen integer	liq bigint	veg bigint	meat bigint	fish bigint	choc bigint	comm bigint
1	0	253244	26896	191653	38222	27331	32018
2	1	245226	18023	93062	25015	19032	36962
3	3	6421	287	2561	220	246	764
4	2	47717	2751	17368	3994	2962	8746

Figure.22

	country character varying (50)	country_spend bigint	twitter bigint	instagram bigint	facebook bigint
1	Spain	547652	65	59	55
2	South Africa	169656	20	17	17
3	Canada	140362	13	23	21
4	Australia	65619	9	11	8
5	India	61017	12	7	6
6	Germany	58113	5	9	8
7	USA	57044	9	7	5
8	Montenegro	1258	0	0	0

Figure.23

	marital_status character varying (50)	twitter bigint	instagram bigint	facebook bigint
1	Together	31	32	29
2	Married	57	56	52
3	Divorced	14	8	9
4	Single	27	32	26
5	Widow	4	5	4

Figure.24