Background/context 2Market is a global supermarket that operates both online and in-store, selling a wide range of products to a diverse customer base. The company seeks to gain insights into customer demographics, assess the effectiveness of its advertising channels, and identify top-selling products, particularly in relation to different customer segments.
This data analytics project should help 2Market to optimize marketing strategies, consider specifics of online and in-store parts of the business, tailor current product offerings, perhaps introduce new products, also improve customer targeting, ultimately boosting sales and enhancing business growth.
Key focus areas and questions in ***Appendix 1***

Analytical approach: Since the course began with Excel and the dataset was relatively small, most of the data cleaning and validation was conducted in Excel (***Appendix 2***). This approach allowed for straightforward manipulation, and I exported the cleaned tables to CSV format for further processing in PostgreSQL (pgAdmin). The same cleaning steps could have been performed directly in PostgreSQL (***Appendix 3***) and using PostgreSQL would have allowed for a more structured approach, working with progressively refined schemas such as ‘raw’ (for initially imported data), ‘staging’ (for cleaning), and ‘reporting’. However, given the dataset’s size, Excel was sufficient for handling all tasks.
During the cleaning process, several key decisions were made. Given the context of the dataset and income in USD, I assumed the date format to be American (mm/dd/yyyy). Certain values were standardized to ensure consistency across records: the '2nd Cycle' in the Education field was combined with 'Master', while 'Absurd' and 'YOLO' in the Marital Status field were merged under 'Unknown'. Similarly, the 'Alone' value was consolidated with 'Single'. When calculating customers' ages, I used 2024 as the reference year and replaced three outlier values with nulls, but retained the rest of the records for analysis. No records were removed, as all customer IDs were unique and in order from a primary key’s perspective.
PostgreSQL allowed for a more granular analysis of the data, enabling in detail exploration of patterns and trends, particularly regarding product preferences and marketing channel effectiveness. (***Appendix 5***). I utilized metrics such as Total_Spent and aggregated average spent to examine customer distributions across various demographics.
The primary data source was the Marketing_data file, which provided the most detailed information, while the Ad_data file was left-joined into it using the primary key column, id.

The design and development of all three **dashboards** aimed to prioritize clarity and simplicity for the user. Each dashboard was focused on answering a specific business question, and only the most relevant information was included. A consistent colour scheme and uniform fonts were used to tie the elements together and enhance the visual coherence across the dashboards. The dashboards were developed for a fixed desktop browser size (1400x800) using containers for layout control and element structuring. Spacing was applied to ensure better alignment between objects. All filters

were set to apply to specific worksheets only, ensuring a smooth and focused user experience.

Customer Dashboard

A Packed Bubbles chart was chosen to visualize the distribution of customer demographics, as it effectively compares group sizes and highlights prominent categories. The Country filter was applied to all views, with a static colour range bar used to indicate the size of each group. Most views on the dashboard used the Count of Id measure combined with a demographic dimension (e.g., Age_Group, Children) (***Appendix 6***). Whereas Income Spent Brackets view gives joint group consisting of specific income group in relation to customers spending group (***Appendix 7***).

Advertising Dashboard

Given that the data was not extensive enough to fully assess the effectiveness of advertising channels, I opted to compare the average Total_Sales of customers who encountered ads versus those who did not. Initial exploration was conducted in pgAdmin (***Appendix 8***), then moved to Tableau, where two new columns—Total_Sales_with_ad and Total_Sales_no_ad—were created for further analysis. The table view clearly outlines low counts of successful ad conversions, suggesting that the data may not yet be sufficient for definitive conclusions.

Different angle on advertising and possible customers' lack of awareness of business' channels as well as opportunity for advertising is reflected on double horizonal bars chart with measure of Total_Sales and average Total_Sales for number and percentage share of customers on labels.

Product Dashboard

The focus of this dashboard was on Total_Sales, percentage share of total sales for each product, and the average spend on each product. Simple cards were used to display key metrics for each product, while a horizontal bar chart showes the multichannel distribution of products and average customer spend. The dashboard includes more filters to allow users to explore product distribution across various dimensions, giving them the flexibility to explore the data.

Patterns, Trends, and Insights: The dataset was heavily skewed towards Spain, which accounted for 49% of all customers, with South Africa and Canada contributing 15% and 12%, respectively. This could be due to factors such as data collection methods, business strategy, or market characteristics, and warrants further investigation.

Key customer patterns reveal that the majority are in the 44-58 age group, married or in a relationship, with children, and hold a degree or higher. Most prefer in-store shopping, earn approximately $50K annually, and spend on average under $300. (***Appendix 4***). The latter may derive from pattern where it was found that people with children tend to be more cautious with spending. These insights are consistent across countries.

Regarding advertising effectiveness, the data is insufficient to definitively measure the impact of specific marketing channels also duration of campaigns is unknown. That may have had an impact too. However, the analysis suggests that customers who

were not targeted by ads tend to spend more on average, whether shopping online or in-store.

On the product front, limiting factor was the absence of unit sales and pricing data. Nonetheless, alcoholic beverages emerged as the top-selling category based on total and average sales, followed by meat. Other product categories contributed significantly less. This trend held across most demographics and countries. Notably, dairy products were excluded from the dataset, which may have skewed product distribution analysis.

Appendix 1

Key Focus Areas:

1. Customer demographic analysis:
   - What are the key demographic attributes (e.g., age, education, marital status, children, income, location) of the customer base?
   - Are there any noticeable patterns or preferences among different demographic groups?
   - How do customer demographics differ between online and in-store shopping?

2. Advertising channel effectiveness:
   - Which advertising channels are most effective at driving sales both online and in-store?
   - Which advertising channels are most effective at driving sales per country?
3. Top-selling products and customer segmentation
   - What are the top-selling products across different customer segments?
   - Are there any products that perform better with certain demographics or in certain geographic areas?
   - How do customer preferences differ across product categories?
   - Are there any under-performing products or categories that may need rebranding or discontinuation?
   - Are there any missing product groups (e.g. dairy)

4. Considering the dual nature of 2Market's operations, understanding the split between online and in-store business performance is essential. Insights from this analysis can drive decisions related to resource allocation, supply chain optimization, and inventory management.
   - What is the revenue split between online and in-store sales?
   - Are there different product preferences between online shoppers and in-store shoppers?
   - How does customer behaviour (average spend etc) differ across these two channels?

Appendix 2

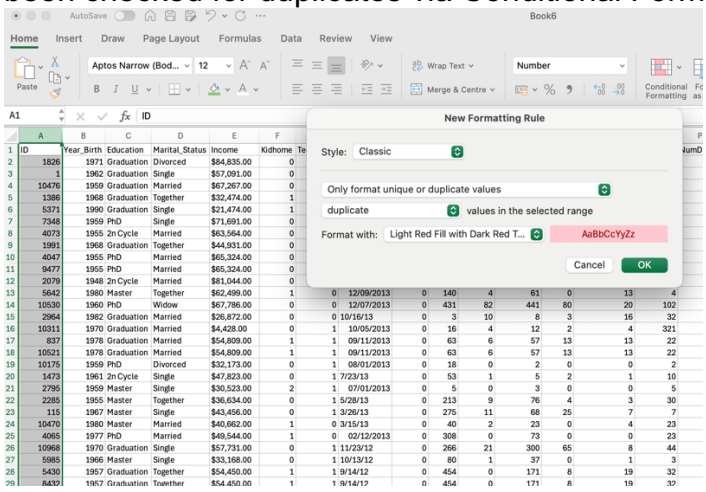
Data cleaning and validating in Excel.

For convenience of using dataset in the future, converted column headers into lower cases via TRIM (A1:W1) and LOWER(A1:W1) afterwards

As per provided metadata_2Market.txt:

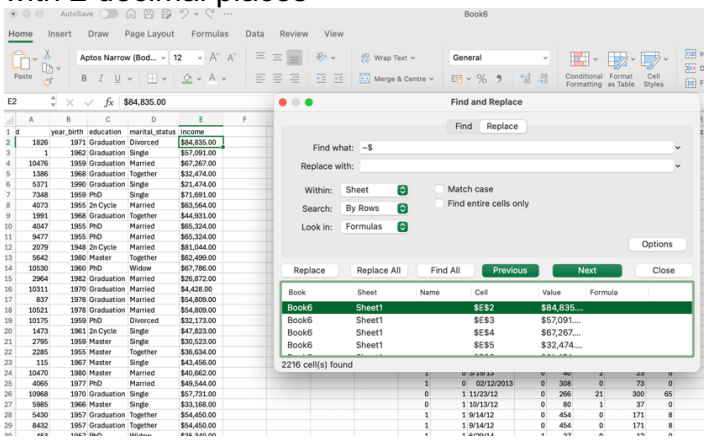
1. marketing_data file

'id' column = unique customer identifier, hence data type = whole number and it has been checked for duplicates via Conditional Formatting & Filtering by Cell colour.

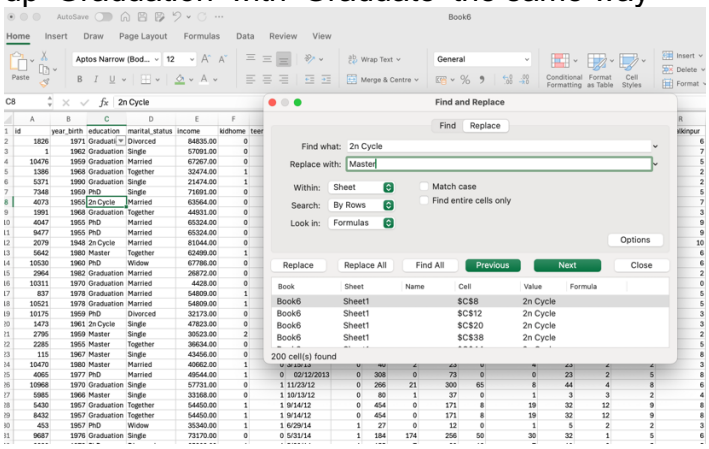


Columns 'year_birth', 'kidhome', 'teenhome', 'recency', 'amtliq', 'amtveg', 'amtnonveg', 'amtpes', 'amtchocolates', 'amtcomm', 'numdeals', 'numwebbuy', 'numwalkinpur', 'numvisits', 'response', 'complain' & 'count_success' are also of whole number data type, hence columns formatted to number with no decimals.

'income' column – Find ~$ & remove all. Afterwards format the column as number with 2 decimal places

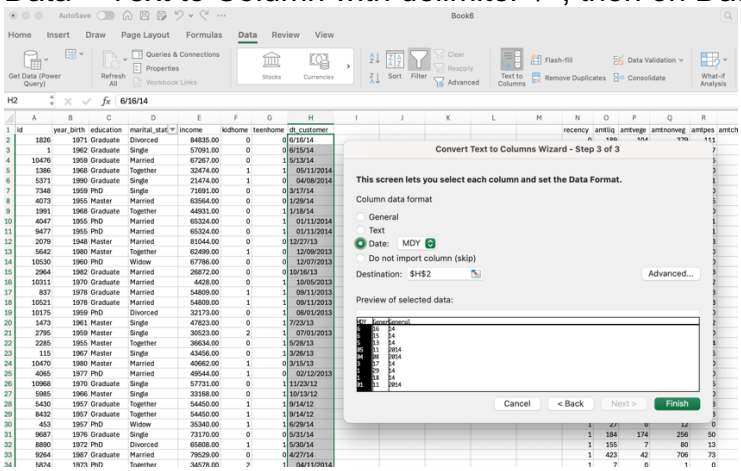


‘education column’ – of text type, that has ‘2nd Cycle’ as one of values, what as per google is on par with ‘Master’, hence Find ‘2n Cycle’ & Replace with ‘Master’, also tidy up ‘Graduation’ with ‘Graduate’ the same way



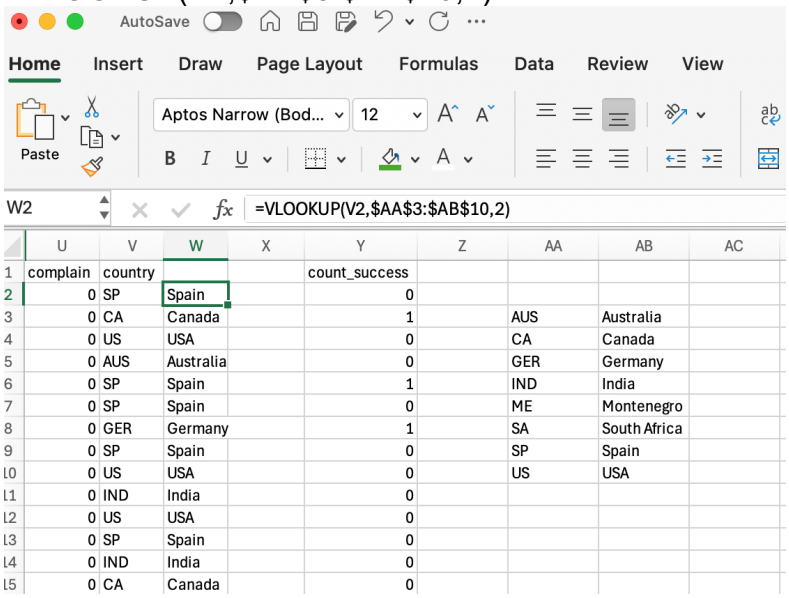
‘marital status’ – text type column, Find ‘Absurd’ and ‘YOLO ‘& Replace with ‘Unknown’, combine ‘Alone’ with ‘Single’ & Replace ‘Together’ with ‘In Relationship’

‘dt_customer’ column – given income column had $, therefore assumed dates were in mm/dd/yy or mm/dd/yyyy i.e. American way too with some of it appearing as text and some as number type clinging to either right or the left side of the cell respectively. Data > Text to Column with delimiter ‘/’ , then on Date> chose MDY format.



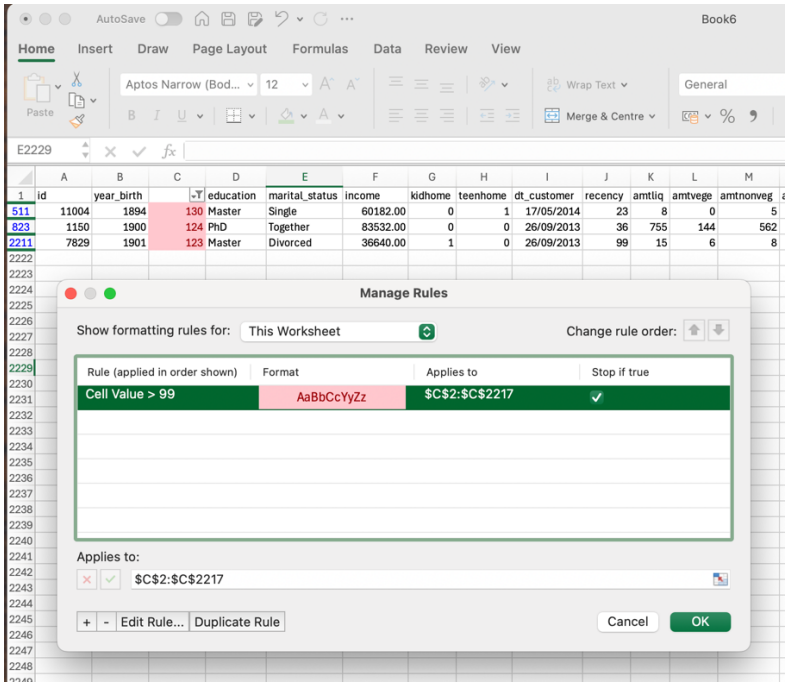
Sense-check that month column does not have values above 12, day <31 and on new column with year values use IF(J2=12,2012,IF(J2=13,2013,IF(J2=14,2014,J2))) to make it all in YYYY format. Then use DATE(J2,H2,I2) to turn values in separate columns back into one date of dd/mm/yyyy format.

In **'country'** column replace country abbreviations via =VLOOKUP(V2,$AA$3:$AB$10,2)



	U	V	W	X	Y	Z	AA	AB	AC
1	complain	country			count_success				
2	0	SP	Spain		0				
3	0	CA	Canada		1		AUS	Australia	
4	0	US	USA		0		CA	Canada	
5	0	AUS	Australia		0		GER	Germany	
6	0	SP	Spain		1		IND	India	
7	0	SP	Spain		0		ME	Montenegro	
8	0	GER	Germany		1		SA	South Africa	
9	0	SP	Spain		0		SP	Spain	
10	0	US	USA		0		US	USA	
11	0	IND	India		0				
12	0	US	USA		0				
13	0	SP	Spain		0				
14	0	IND	India		0				
15	0	CA	Canada		0				

Add **'age'** column. Note, I used 2024 as reference year to find customers' age. Sense-checked it for anything >99 and replaced 3 values with '#N/A' to not lose other values from these records.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	year_birth		education	marital_status	income	kidhome	teenhome	dt_customer	recency	amtliq	amtvege	amtnonveg
511	11004	1894	130	Master	Single	60182.00	0	1	17/05/2014	23	8	0	5
823	1150	1900	124	PhD	Together	83532.00	0	0	26/09/2013	36	755	144	562
2211	7829	1901	123	Master	Divorced	36640.00	1	0	26/09/2013	99	15	6	8

2. ad_data file

The data cleaning and validation process for the file involved several straightforward steps:

- column name standardization where converted all column names to lowercase for consistency and ease of use.
- checked for duplicates within the ‘id’ column to ensure uniqueness.
- verified that the values in numerical columns were whole numbers and fell within the valid range specified in the metadata file.
- ensured there were no irregularities, such as incorrect data types, or out-of-range values, in any columns.

Appendix 3

Data cleaning and validating in PostgreSQL.

Create three Schemas in 2Market database to be used for storing initial data in 'raw', all cleaning done in 'staging' and ready tables/views to be used for reporting in 'reporting'

Created two tables - 'marketing_data' & 'ad_data' - as per cvs provided in raw schema just to import initial data. To avoid any issues with importing, other than 'ID' column in both tables data type = bigserial primary key, everything else of data type varchar(255).

Check for duplicates of primary key in each table (there was none) & move on to 'staging' Schema to carry out cleaning and validating

```
create schema raw;
create schema staging;
create schema reporting;

create table raw.marketing_data (
"ID" bigserial primary key,
"Year_Birth" varchar(225),
"Education" varchar(225),
"Marital_Status"varchar(225),
"Income"varchar(225),
"Kidhome"varchar(225),
"Teenhome" varchar(225),
"Dt_Customer" varchar(225),
"Recency" varchar(225),
"AmtLiq" varchar(225),
"AmtVege" varchar(225),
"AmtNonVeg" varchar(225),
"AmtPes" varchar(225),
"AmtChocolates" varchar(225),
"AmtComm" varchar(225),
"NumDeals" varchar(225),
"NumWebBuy"varchar(225),
"NumWalkinPur" varchar(225),
"NumVisits"varchar(225),
"Response" varchar(225),
"Complain"varchar(225),
"Country" varchar(225),
"Count_success" varchar(225)
);

create table raw.ad_data(
"ID" bigserial primary key,
"Bulkmail_ad"varchar(225),
"Twitter_ad" varchar(225),
"Instagram_ad" varchar(225),
"Facebook_ad" varchar(225),
"Brochure_ad" varchar(225)
);

select "ID", count(*)
from raw.ad_data
group by "ID"
having count(*)>1
;
```

Create a table of both tables in ‘staging ‘ Schema for further interaction

Change column names in both tables to be in lower case for convenience & to avoid using quotation marks

```
create table staging.ad_data as
select * from raw.ad_data;

create table staging.marketing_data as
select * from raw.marketing_data;

--- change column names in both tables to be lower case for conve
alter table  staging.ad_data
rename column "ID" to id;
alter table  staging.ad_data
rename column "Bulkmail_ad" to bulkmail_ad;
alter table  staging.ad_data
rename column "Twitter_ad" to twitter_ad;
alter table  staging.ad_data
rename column "Instagram_ad" to instagram_ad;
alter table  staging.ad_data
rename column "Facebook_ad" to facebook_ad;
alter table  staging.ad_data
rename column "Brochure_ad" to brochure_ad
;

alter table staging.marketing_data
rename column "ID" to id;
alter table staging.marketing_data
rename column "Year_Birth" to year_birth;
alter table staging.marketing_data
rename column "Education" to education;
alter table staging.marketing_data
rename column"Marital_Status" to marital_status;
alter table staging.marketing_data
rename column"Income" to income;
alter table staging.marketing_data
rename column"Kidhome"to kidhome;
alter table staging.marketing_data
rename column"Teenhome" to teenhome;
alter table staging.marketing_data
rename column"Dt_Customer" to dt_customer;
alter table staging.marketing_data
rename column"Recency" to recency;
alter table staging.marketing_data
rename column"AmtLiq" to amtliq;
alter table staging.marketing_data
rename column"AmtVege" to amtveg;
alter table staging.marketing_data
rename column"AmtNonVeg" to amtnonveg;
alter table staging.marketing_data
rename column"AmtPes" to amtpes;
alter table staging.marketing_data
rename column"AmtChocolates" to amtchocolates;
alter table staging.marketing_data
rename column"AmtComm" to amtcomm;
alter table staging.marketing_data
rename column"NumDeals" to numdeals;
alter table staging.marketing_data
rename column"NumWebBuy"to numwebbuy;
alter table staging.marketing_data
rename column"NumWalkinPur" to numwalkinpur;
alter table staging.marketing_data
rename column"NumVisits"to numvisits;
alter table staging.marketing_data
rename column"Response" to response;
alter table staging.marketing_data
rename column"Complain"to complain;
alter table staging.marketing_data
rename column"Country" to country;
alter table staging.marketing_data
rename column"Count_success" to count_success;
```

Change data type to appropriate for columns in both tables:

- 'id' columns in both tables have been imported as bigint primary key;
- for the remaining columns, before data type can be changed to the appropriate, to check:

---in ad_data table:

a) whether integer appearing values have accidental spaces (trim)
b) whether values have any other symbols ('^\d+$')

```
129    alter table staging.ad_data
130    alter column brochure_ad set data type integer
131    using case
132    when trim(brochure_ad)~'^\d+$' then trim(brochure_ad)::integer
133    else 0
134    end;
```

Cleaning, validation for ad-data table is DONE

---in marketing_data table:

columns **'year_birth', 'kidhome', 'teenhome', 'recency', 'amtliq', 'amtveg', 'amtnonveg', 'amtpes', 'amtchocolates', 'amtcomm', 'numdeals', 'numwebbuy', 'numwalkinpur', 'numvisits', 'response', 'complain' & 'count_**success' check for:

a) whether integer appearing values have accidental spaces (trim)
b) whether values have any other symbols ('^\d+$')

```
144    alter table staging.marketing_data
145    alter column count_success set data type integer
146    using case
147    when trim(count_success)~'^\d+$' then trim(count_success)::integer
148    else 0
149    end;
```

column **'income'** to be changed to numeric (12,2)
remove $ & thousands delimiter ','

```
153    alter table staging.marketing_data
154    alter column income set data type numeric (12,2)
155    using replace(replace (income,'$',''),',','')::numeric(12,2);
```

column **'dt_customer** ' data type to be changed to date type
given income is in $, assumption was made that all date values are in mm/dd/yy format
used to_date() command to explicitly convert the mm/dd/yy format to date type.

```
161    alter table staging.marketing_data
162    alter column dt_customer set data type date
163    using to_date(dt_customer,'mm/dd/yy');
```

update **'country'** column values to full country name

```
166     update staging.marketing_data
167     set country = case
168     when country = 'AUS' then   'Australia'
169     when country =  'CA' then   'Canada'
170     when country =  'GER' then  'Germany'
171     when country = 'IND' then 'India'
172     when country =  'ME' then 'Montenegro'
173     when country = 'SA' then 'South Africa'
174     when country = 'SP' then 'Spain'
175     when country = 'US' then 'USA'
176     end;
```

check what's in the data and tidy up '**marital_status'** by using 'case when then' to combine irregular values

```
183     select distinct marital_status, count(marital_status)
184         from staging.marketing_data
185         group by marital_status;
186
187         update staging.marketing_data
188         set marital_status = case
189         when marital_status = 'Alone' then 'Single'
190         when marital_status = 'Divorced' then 'Divorced'
191         when marital_status = 'Married' then 'Married'
192         when marital_status = 'Single' then 'Single'
193         when marital_status = 'Together' then 'In Relationship'
194         when marital_status = 'Widow' then 'Widow'
195         else 'Unknow'
196         end;
```

check what's in the data and tidy up '**education' column** by running replace '2nd Cycle' with 'Master'

```
200     select distinct education, count(education)
201         from staging.marketing_data
202         group by education;
203
204     update staging.marketing_data
205         set education = replace (education, '2nd Cycle', 'Master')
206         where education = '2n Cycle';
```

add new column for '**age'** of customers (reference point 2024)
sense check age i.e. not <15 and not >100 & replace any of it with null to preserve the rest of the record for analysis

```
206    alter table staging.marketing_data
207    add column age integer;
208    update staging.marketing_data
209    set age=2024-year_birth
210    where year_birth is not null;
211
212    select id, age
213    from staging.marketing_data
214    where age < 15 or age>100;
215
216    update staging.marketing_data
217    set age  = case
218    when age >100 then null
219    else age
220    end;
221
222    select *
223    from staging.marketing_data
224    where age is null;
225
```

check for null in this column brings the same three records as in excel

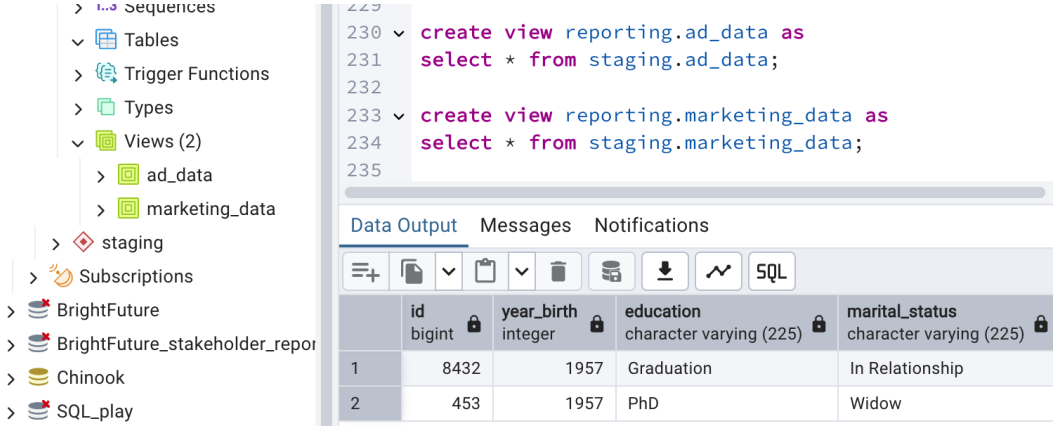
```
221  select *
222  from staging.marketing_data
223  where age is null;
224
```

Data Output | Messages | Notifications

	id bigint	year_birth integer	education character varying (225)	marital_status character varying (225)	in nu
1	1150	1900	PhD	In Relationship	
2	11004	1894	2n Cycle	Single	
3	7829	1901	2n Cycle	Divorced	

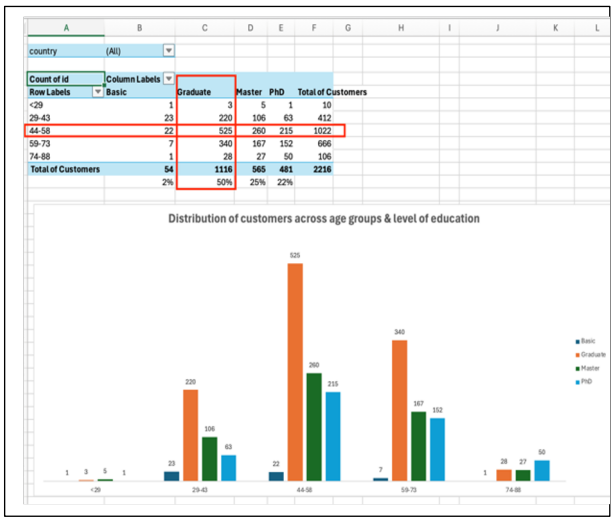
Cleaning, validation for marketing_data table is DONE

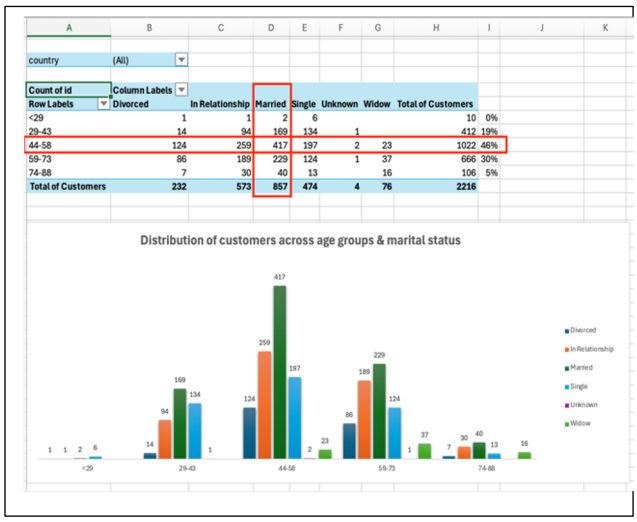
In 2Market case not much of a difference whether we'll create tables or views in 'reporting' Schema to get on with further interaction with the data.



Appendix 4

Initial data exploration in Excel





Children per home by age group

Row Labels	Sum of children per home	Average of children per home
<29	2.00	0.20
29-43	251.00	0.61
44-58	1116.00	1.09
59-73	678.00	1.02
74-88	52.00	0.49
Grand Total	**2099.00**	**0.95**

Row Labels	No of Customer
Spain	1093.00
South Africa	337.00
Canada	266.00
Australia	147.00
India	147.00
Germany	116.00
USA	107.00
Montenegro	3.00
Grand Total	**2216.00**

Averadge income increasing with age, but average spent dips for the group with most customers

Row Labels	Count of id	Average of income($)
<29	10	50,696.00
29-43	412	46,879.00
44-58	1022	50,599.85
59-73	666	56,244.78
74-88	106	64,025.73
Grand Total	**2216**	**52,247.25**

Row Labels	Count of id	Average of total_spent
<29	10	588.40
29-43	412	573.66
44-58	1022	539.28
59-73	666	681.03
74-88	106	927.70
Grand Total	**2216**	**607.08**

In store shopping preferred over online across all demographics

education	(All)
marital_status	(All)
income($)	(All)
children_combined	(All)
total_spent	(All)
country	(All)

Row Labels	Sum of numwebbuy	Sum of numwalkinpur	Count of id
<29	25	49	10
29-43	1424	2177	412
44-58	4111	5673	1022
59-73	2938	4199	666
74-88	555	757	106
Grand Total	**9053**	**12855**	**2216**

Appendix 5

Further data exploration in PostgreSQL

```
--Q: What is the total spend per product per country?
--OBSERVATION:
---apart from amount spent on certain category no further data, e.g. unit price or quantities provided,
---hence it's not possible to establish whether product seems to be popular
---because customer base in particular country is larger,
---or perhaps it is more expensive and overall less people need to buy it to drive total sales.
select country,
sum(amtliq)as total_spent_alcohol,
sum(amtvege)as total_spent_vegies,
sum(amtnonveg)as total_spent_meat,
sum(amtpes)as total_spent_fish,
sum(amtchocolates)as total_spent_chocolates,
sum(amtcomm)as total_spent_commodities
from marketing_data
group by country;
```

Data Output | Messages | Notifications

Showing rows: 1 to 8 | Page No: 1 of 1

	country character varying (255)	total_spent_alcohol numeric	total_spent_vegies numeric	total_spent_meat numeric	total_spent_fish numeric	total_spent_chocolates numeric	total_spent_commodities numeric
1	Spain	336392.00	28288.00	178409.00	40153.00	30134.00	46181.00
2	South Africa	105918.00	8937.00	58398.00	13670.00	9019.00	15129.00
3	Montenegro	1729.00	8.00	817.00	226.00	122.00	220.00
4	Australia	42752.00	3689.00	22328.00	5546.00	4129.00	7132.00
5	Germany	36776.00	2980.00	20272.00	4601.00	2801.00	5768.00
6	Canada	84066.00	7681.00	45925.00	9980.00	7607.00	12144.00
7	India	36236.00	3788.00	23729.00	4818.00	3221.00	6014.00
8	USA	32214.00	3034.00	20185.00	4411.00	2863.00	4839.00

```
-------------
-- data exploration
-- Q: Avg spent per coustomer per country
-- OBSERVATION:
----most customer are in Spain (49%) & Montenegro data set only has 3 customers,
----hence can be disregarded for purpose of this query
select
country,count(id) as customers,
        round((count(id) * 100.0 / (select count(*) from marketing_data))) as percent_of_total_customers,
        round(avg(total_spent),2) as avg_spent_per_customer
from marketing_data
group by country
order by avg_spent_per_customer desc;
```

Data Output | Messages | Notifications

Showing rows: 1 to 8 | Page No: 1

	country character varying (255)	customers bigint	percent_of_total_customers numeric	avg_spent_per_customer numeric
1	Montenegro	3	0	1040.67
2	USA	107	5	631.27
3	Germany	116	5	631.02
4	Canada	266	12	629.33
5	South Africa	337	15	626.32
6	Spain	1093	49	603.44
7	Australia	147	7	582.15
8	India	147	7	529.29

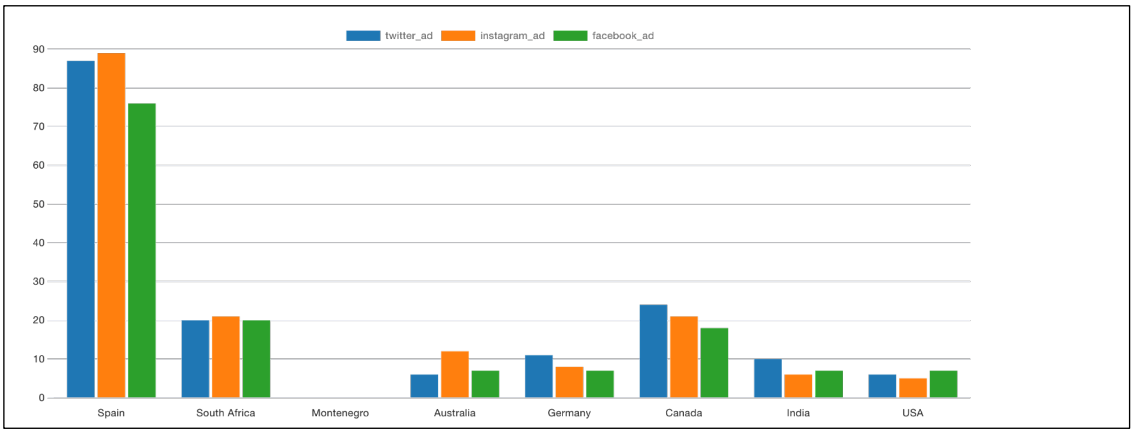
Q: Which social media platform(s) seem to be the most effective per country?

```sql
-- Q: Which social media platform(s) seem to be the most effective per country?
---(In this case, assume that purchases were in some way influenced by lead conversions from a campaign)


select m.country,                       --- the country column from the marketing_data table aliased as m. It represents the country of the customer.
    sum(a.twitter_ad) as twitter_ad,    --- sums ea media platform's values to output in 3 separate columns
    sum(a.instagram_ad) as instagram_ad,
    sum(a.facebook_ad) as facebook_ad,
    case when                           --- will evaluate sums of conversons for ea platform for ea country
            sum(a.twitter_ad) > sum(a.instagram_ad) and --when sum of twitter conversions is greater then instagram
            sum(a.twitter_ad) > sum(a.facebook_ad)  --- and facebook
            then 'Twitter'                          --- the last column most_effective_platform will return value 'Twitter'
        when
            sum(a.instagram_ad) > sum(a.twitter_ad) and
            sum(a.instagram_ad) > sum(a.facebook_ad)
            then 'Instagram'
        when
            sum(a.facebook_ad) > sum(a.twitter_ad) and
            sum(a.facebook_ad) > sum(a.instagram_ad)
            then 'Facebook'
        when
            sum(a.twitter_ad) = sum(a.instagram_ad) and --- checks for when values of conversions are the same
            sum(a.twitter_ad) = sum(a.facebook_ad)
            then 'Multiple Platforms'
    end as most_effective_platform          --- ends the case with the column most_effective_platform
from marketing_data m                       --- specifies marketing_data table aliased as m is primary
left join ad_data a                         --- ad_data table aliased as a is left joined to marketing_data table to ensure all
                                            ----records from the marketing_data table will be included, even if there is no matching record in the ad_data table
using (id)                                  --- join's done using primary key id present in both tables
group by m.country;                         --- this will group the output by country & query will end. Aggregating will not work without grouping.

```

Data Output | Messages | Notifications

Showing rows: 1 to 8

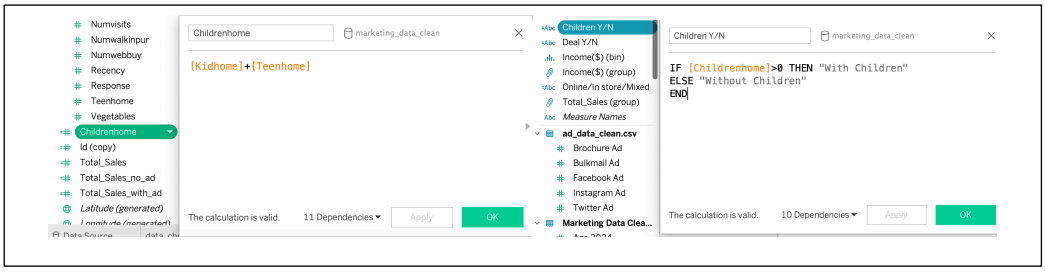
	country character varying (255)	twitter_ad bigint	instagram_ad bigint	facebook_ad bigint	most_effective_platform text
1	Spain	87	89	76	Instagram
2	South Africa	20	21	20	Instagram
3	Montenegro	0	0	0	Multiple Platforms
4	Australia	6	12	7	Instagram
5	Germany	11	8	7	Twitter
6	Canada	24	21	18	Twitter
7	India	10	6	7	Twitter
8	USA	6	5	7	Facebook



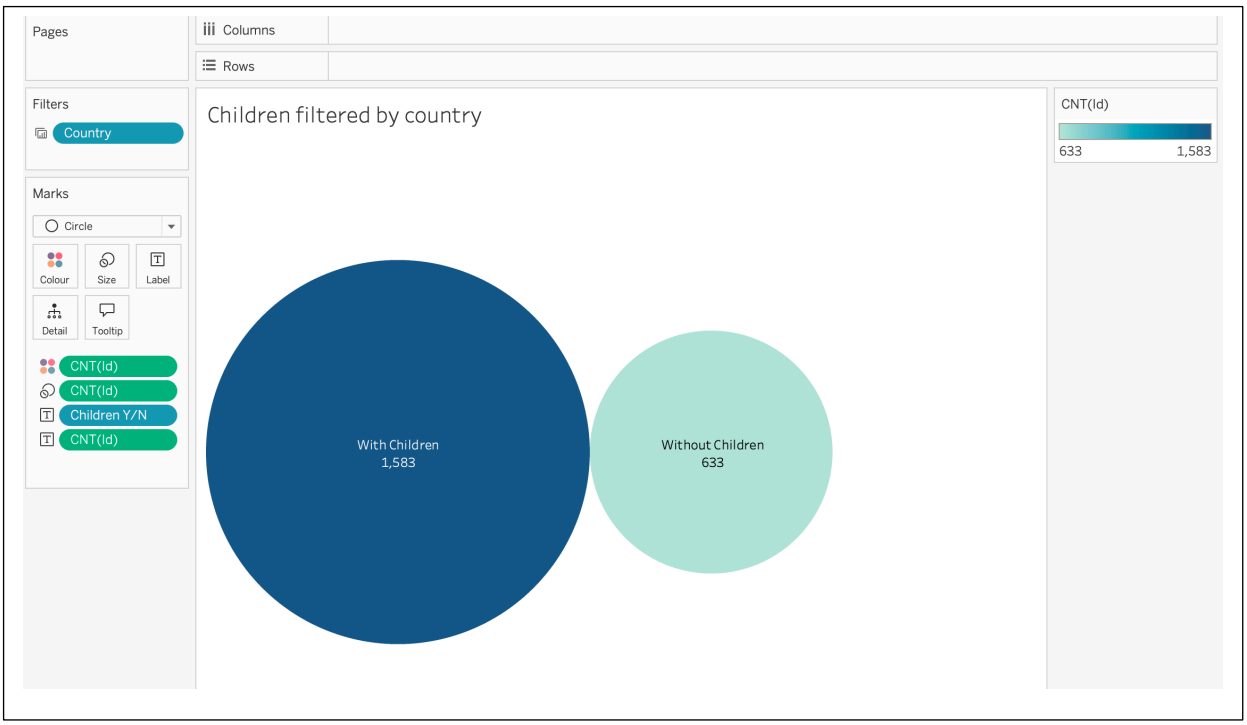
Appendix 6

Children View for Customer Dashboard

To create the **Children** view, I first combined **Kidhome** and **Teenhome** into a new measure, **Childrenhome**, using a calculated field. Then, the second calculated field created **Children Y/N** dimension based on the **Childrenhome** measure.



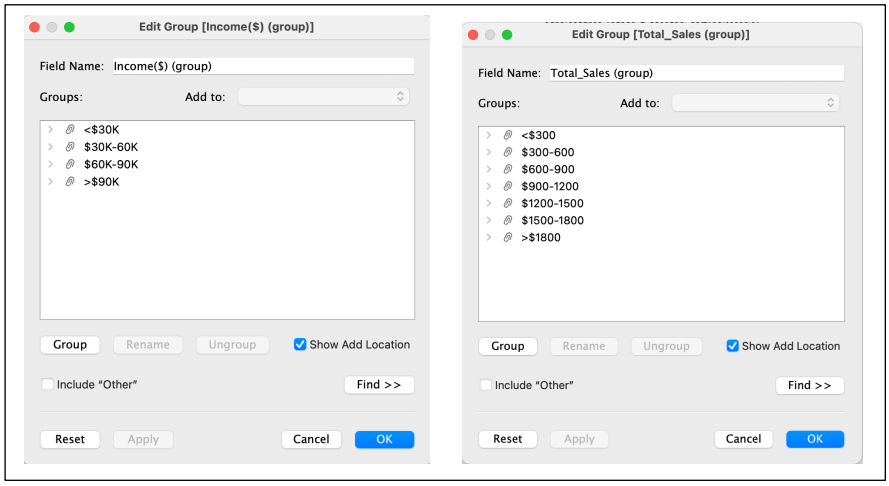
In the view, I placed the **Children Y/N** dimension on the **Rows** shelf and **Count of Id** on the **Columns** shelf. After selecting the **Packed Bubbles** chart type, I added **Count of Id** to **Colour**, **Size**, and **Label** on the **Marks card**, while placing the **Children Y/N** dimension on **Label** to ensure consistent colour and labelling across the dashboard.

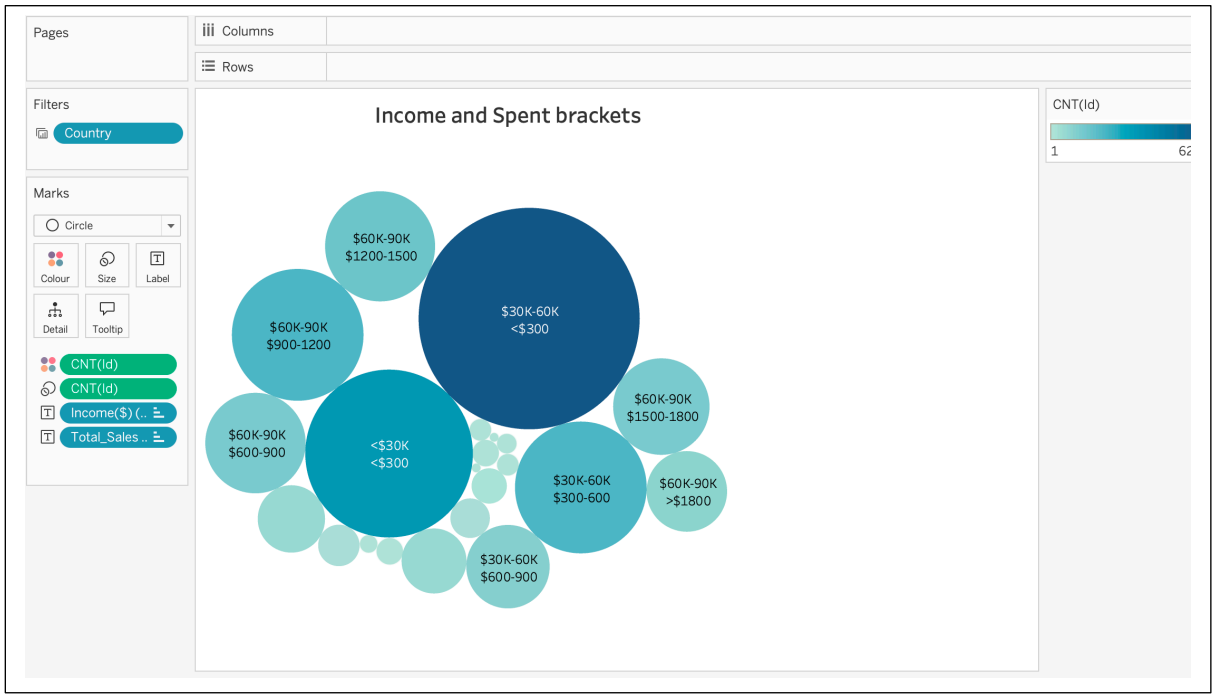


Appendix 7

Income_Spent view for Customer Dashboard

To create the **Income_Spent** view, I created custom **Income Group** and **Total_Spent Group** dimensions (renaming the **Total_Sales** column). I opted for custom groups instead of automated bins. The process to build the **Packed Bubbles** chart was the same as before. I chose not to allow labels to overlap smaller, less significant groups, relying on tooltips when hovering over the bubbles instead.





Appendix 8

Business Q: Which advertising channels seem to be the most effective bringing the most sales (per country)?

Logic & limitations:

- we do not have revenue/total_spent derived from ad channels, therefore we cannot apportion sales
- 'count_success' is sum of all (Bulkmail+FB+Instagram,etc) successful leads for any given customer, so we'll assume the purchase happened
- and will compare sales from those whose count_success>0 and those whose count_success = 0, aka no approach from any marketing campaign
- note 'response' is not the same as 'count-success' & connection to ads file unclear

```
---step1: counts & lists in the decending order total ad campaign_effect per country
select m.country,
    sum (m.response)as accepted_offers,
    sum (m.count_success) as campaign_effect,
    sum (a.bulkmail_ad)as bulkmail_ad,
    sum (a.twitter_ad)as twitter_ad,
    sum (a.instagram_ad)as instagram_ad,
    sum (a.facebook_ad)as facebook_ad,
    sum (a.brochure_ad)as brochure_ad
from marketing_data m
left join ad_data a
using (id)
group by country
order by campaign_effect desc;
```

	country character varying (255)	accepted_offers bigint	campaign_effect bigint	bulkmail_ad bigint	twitter_ad bigint	instagram_ad bigint	facebook_ad bigint	brochure_ad bigint
1	Spain	176	351	83	87	89	76	16
2	Canada	38	87	18	24	21	18	6
3	South Africa	52	86	21	20	21	20	4
4	Germany	17	38	10	11	8	7	2
5	India	13	38	13	10	6	7	2
6	Australia	22	34	9	6	12	7	0
7	USA	13	26	8	6	5	7	0
8	Montenegro	2	1	1	0	0	0	0

```
---step2 show countries in order
---if count_success > 0 sum total_spent as total_spent_with_ad else total_spent_no_ad
select m.country,
sum (case when m.count_success>0 then m.total_spent else 0 end) as total_spent_with_ad,
sum (case when m.count_success=0 then m.total_spent else 0 end) as total_spent_no_ad
from marketing_data m
```

Data Output | Messages | Graph Visualiser | Notifications

Showing rows: 1 to 8

	country character varying (255)	total_spent_with_ad numeric	total_spent_no_ad numeric
1	Spain	265271.00	394286.00
2	South Africa	74692.00	136379.00
3	Montenegro	874.00	2248.00
4	Australia	27422.00	58154.00
5	Germany	30218.00	42980.00
6	Canada	61882.00	105521.00
7	India	24239.00	53567.00
8	USA	17424.00	50122.00