2MARKET ANALYSIS

EXCEL INITIAL INSIGHTS:

I have cleaned the data using excel. Data was already relatively clean.
No duplicates or missing values.
Also checked for spelling errors and correct data type formats.

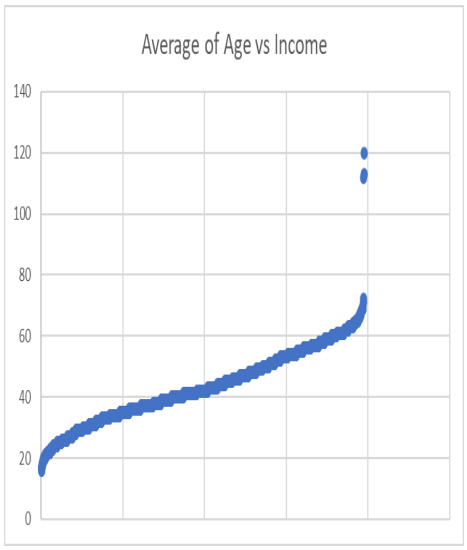
Most notable change was removing the dollar sign from income in order to perform calculations.
Absurd and YOLO in marital status are anomalies and thus ignored.
The question I am investigating is as the number of web purchases are low compared to the number of walk in purchases.
How do we increase web purchases in order to shift to a more online service which reduces employment and overhead costs of the business?

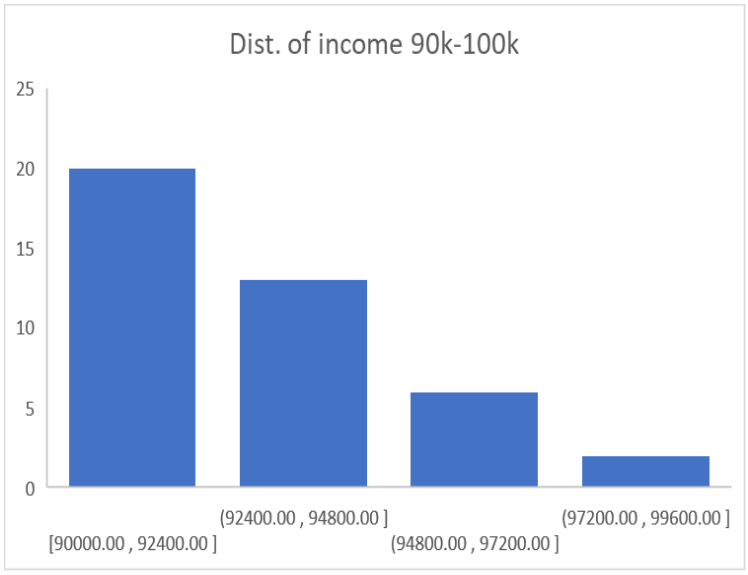
Age was calculated by the dt customer column and yr birth column.
(Using YEAR function and subtracting).
We see quite clearly that the bin of ages 36-43 is the most common age.

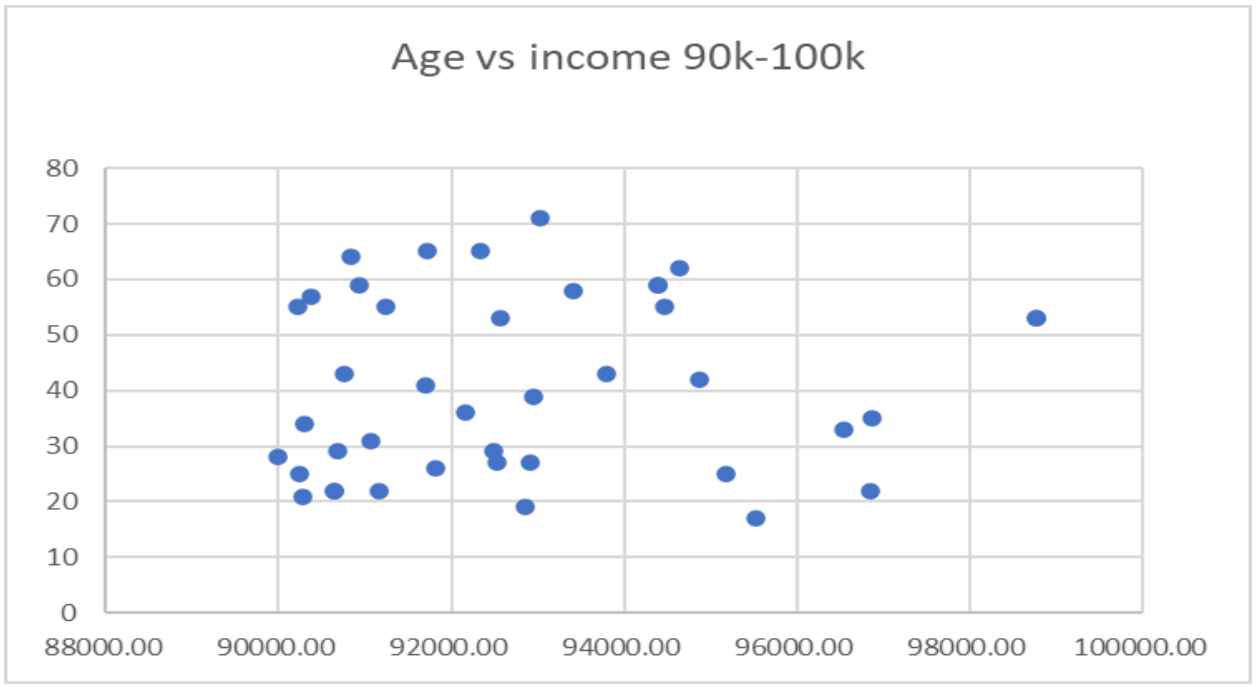
Pivot Tables were used for the visualizations.

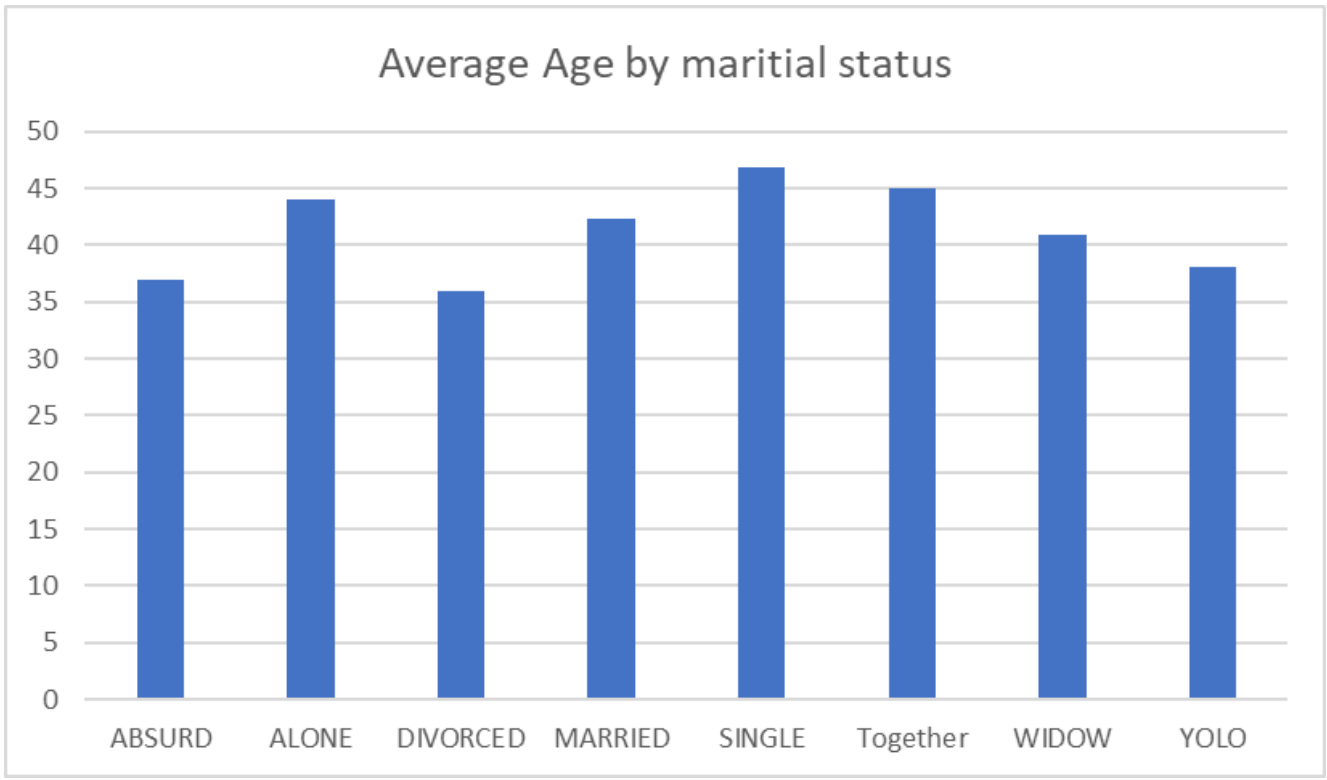
Age is fairly normally distributed as shown by the histogram.
Single has the highest average age and divorce has the lowest.
Income is highly correlated with the number of web purchases.
Average age has a positive correlation with income.
We see the bracket 90-92.4k has the highest frequency. Income between 90-100k is right skewed.
Spain has the highest number of webuy and majority is coming from liquor.
Customers are more likely to purchase commodities , chocolates and liquor online so we target this. Graduation has the highest number of web purchases. It has an average age of 43.
From this initial analysis we see that to increase web buys we have to target the age group 36-43
in specific graduates (who are usually around 42 years old) in Spain and increase spending and advertising of liquor.
We also should target higher incomes people in the range 90-92.4k going above that in the 90-100k range would lead to reduced efficiency
due to right-skewness of data.
We should target Divorced and married people as their age falls in the
age bracket.
Also families with 0 kids have a significantly higher number of web purchases.
Ages 71 and above are outliers.
Average age was calculated using AVERAGE function..

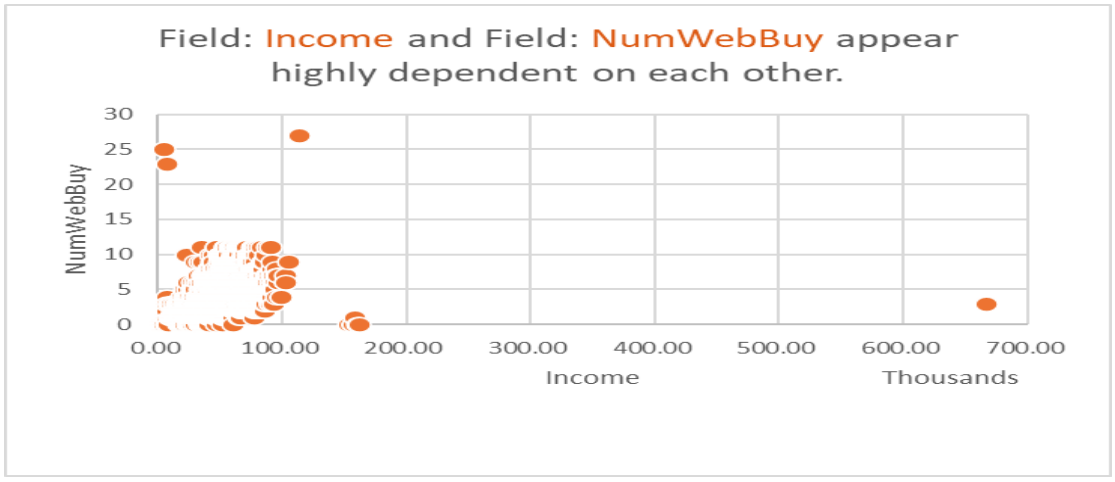
Income was sorted to find average age of 90k-100k income range.

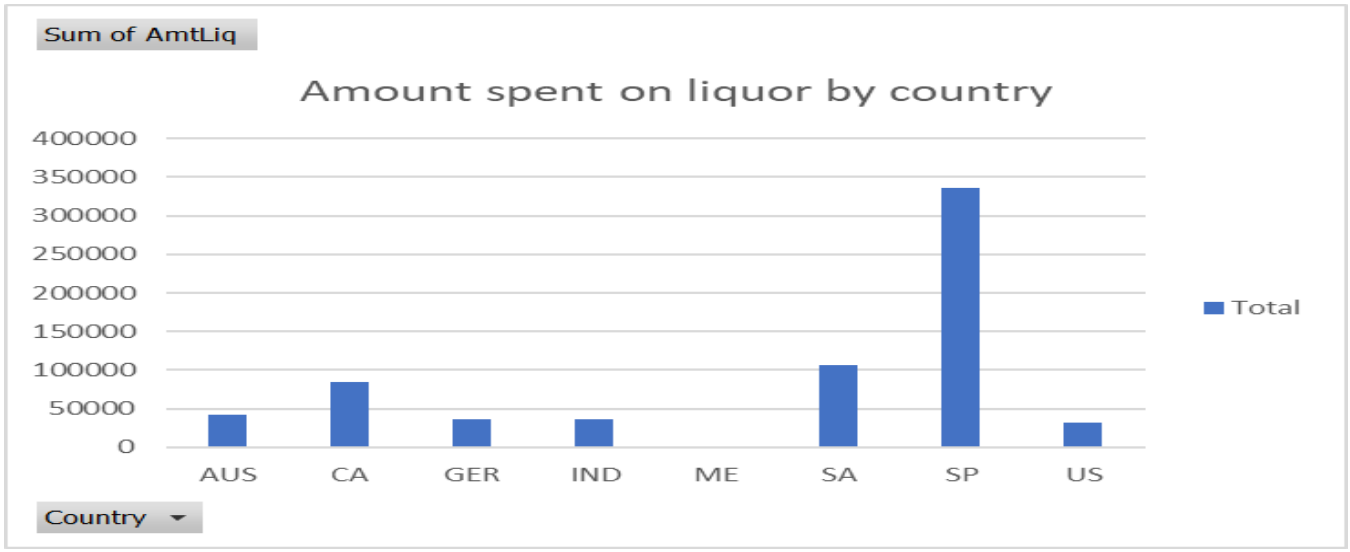


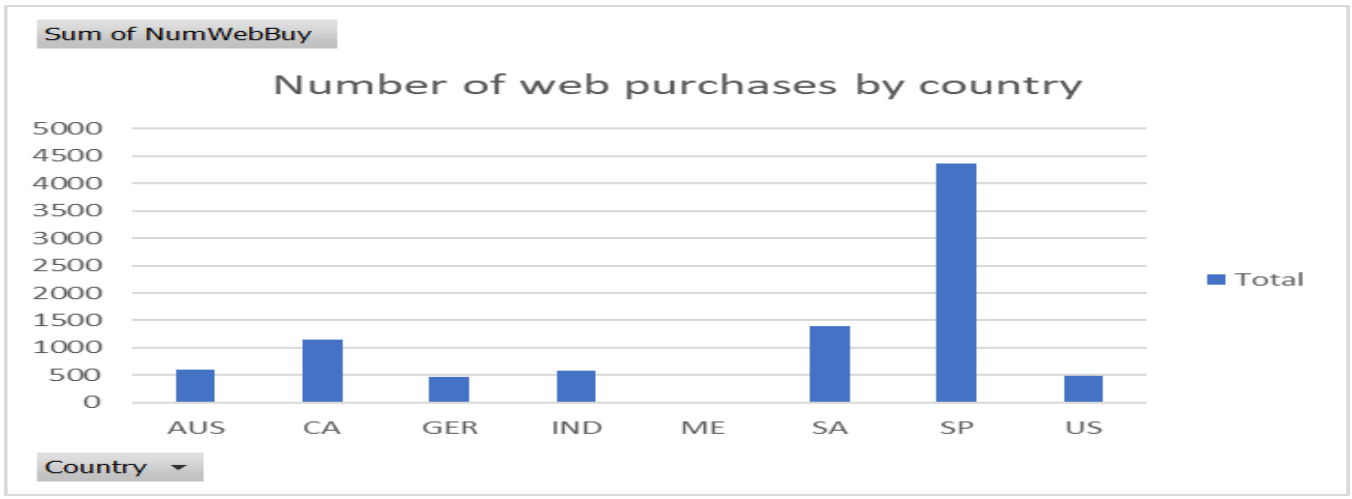


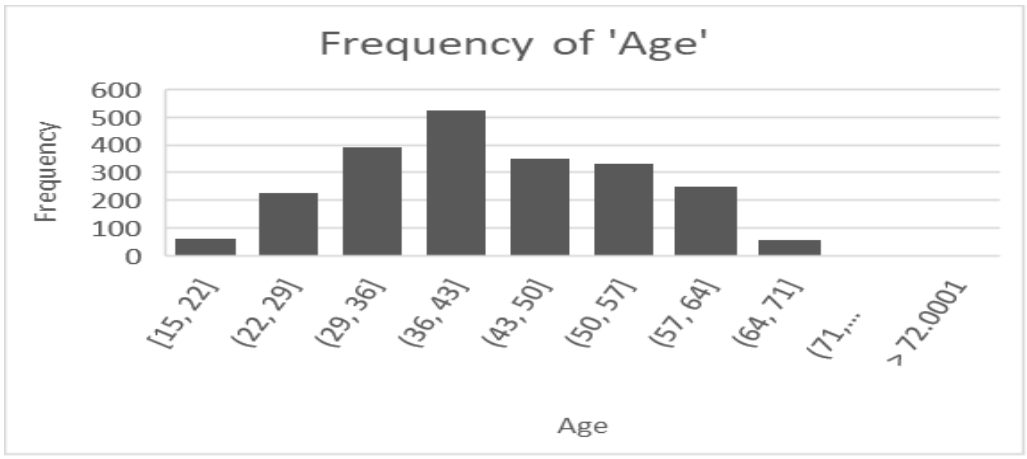



Average Age by maritial status
50
45
40
35
30
25
20
15
10
5
0
ABSURD
ALONE
DIVORCED
MARRIED
SINGLE
Together
WIDOW
YOLO


Field: Income and Field: NumWebBuy appear highly dependent on each other.
NumWebBuy
30
25
20
15
10
5
0
0.00
100.00
200.00
300.00
400.00
500.00
600.00
700.00
Income
Thousands


Sum of AmtLiq
Amount spent on liquor by country
400000
350000
300000
250000
200000
150000
100000
50000
0
AUS
CA
GER
IND
ME
SA
SP
US
Total
Country


Sum of NumWebBuy
Number of web purchases by country
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
AUS
CA
GER
IND
ME
SA
SP
US
Total
Country


Frequency of 'Age'
Frequency
600
500
400
300
200
100
0
[15, 22]
(22, 29]
(29, 36]
(36, 43]
(43, 50]
(50, 57]
(57, 64]
(64, 71]
(71,...
> 72.0001
Age

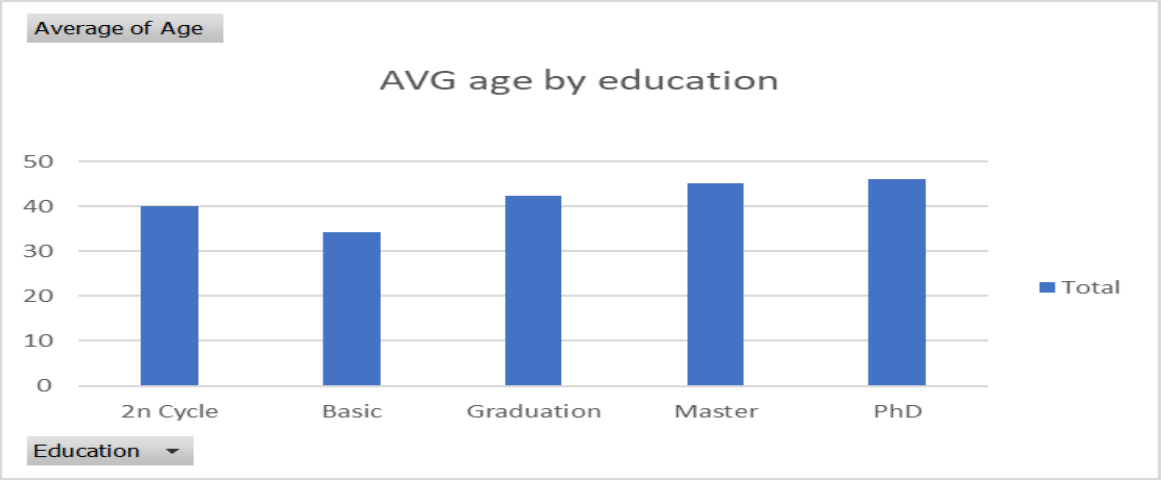


TABLEAU:

Total sales is lowest in July at 50k and max in August. Other than the major dip in July the trend is quite consistent. We optimize sales by increasing marketing and supply in August.
Alcohol is clearly the best seller and chocolates and vegetables are lowest.
We can advertise chocolates, alcohol and commodities more on the website.

Max Sales occur in 40k-60k income range.

Spain has the most online and in person sales.
In person sales are higher, 6k vs 4k.
We spend less money on particularly the middle east especially due to alcohol restriction.

Looking at the sales breakdown if we can make commodities, chocolates and alcohol online only and assuming they are purchased at the same rate we can shift the majority of sales from in person to the website.

In spain Instagram and twitter is most successful for online purchases.
Again nothing in the middle east. We should definitely reduce money spent on the middle east.
Canada the same as spain.
SA roughly all the same.

The dashboard was made with the Total sales by month metric, Total sales and breakdown by product metric, distribution of age by the marketing channel to show which marketing channel Might be the best option, and the comparison of in person and web purchases by country.
The dashboard has been made accessible for mobile devices by fitting one visual on each page. The total sales per month per product visual and the total sales breakdown
By product visual share the same colour key. Shown by the common color key.

Off/online sales have labels to account for potential color-blindness. The total sales and breakdown by product and monthly time series have the same colour key as labels would cause congestion. The breakdown of age by marketing channel is all the same colour to again address colorblindness.

SQL CODE AND INSIGHTS:
I used the CREATE function to create the marketing_data and ad_data tables (and the corresponding columns).
I then imported the data from the cleaned excel data sets.

```
SELECT COALESCE( kidhome,teenhome) AS "kid/teen_home", sum(AmtLiq) as
total_alc,count(AmtVege) as total_veg,
sum(AmtNonVeg)as total_meat,sum(AmtPes) as "total_fish`",sum(AmtChocolates) as
"total_choc"
,sum(AmtComm) AS "total_comm"

FROM public."marketing_data"

WHERE COALESCE( kidhome,teenhome) >= 1
 GROUP BY "kid/teen_home"
```

The COALESCE function was used to combine the kidhome and teen home columns.
The sum function was used to get the total sales of each product.

From the table we see when 1 kid or teen is at home alcohol and overall spend is a lot more than if 2 were at home.

```
SELECT maritial_status, sum(AmtLiq) as total_alc,count(AmtVege) as total_veg,
sum(AmtNonVeg)as total_meat,sum(AmtPes) as "total_fish`",sum(AmtChocolates) as
"total_choc"
,sum(AmtComm) AS "total_comm"
```

```
FROM public."marketing_data"
Group by "maritial_status"
```

Couples who are together and married spend the most on alcohol.
Single is 3rd.

```
SELECT country, sum(AmtLiq) as total_alc,sum(AmtVege) as total_veg,
sum(AmtNonVeg)as total_meat,sum(AmtPes) as "total_fish",sum(AmtChocolates) as
"total_choc"
,sum(AmtComm) AS "total_comm"


FROM public."marketing_data"

 GROUP BY "country"
```

Spain spends the most on alcohol.

```
CREATE table combined AS
SELECT md.year_birth,md.education,md.income,md.kidhome,
md.teenhome,md.dt_customer,md.maritial_status,
md.recency,md.amtliq,md.amtvege,md.amtnonveg,md.amtpes,
md.amtchocolates,md.amtcomm,md.numdeals,md.numwebbuy,
md.numwalkinpur,md.numvisits,md.response,md.complain,md.country,md.count_success,
md.age, ad.bulkmail_ad,ad.twitter_ad,ad.instagram_ad,ad.facebook_ad,
ad.brochure_ad

FROM public.marketing_data  md
INNER JOIN public.ad_data ad
on md.id=ad.id

SELECT * FROM public.combined;
```

I performed an inner join on marketing_data and ad_data using the common id(primary key) column.
I then am able to query the new combined table.

```
SELECT country,sum(facebook_ad) as totfacebook,
sum(instagram_ad) as totinsta , sum(twitter_ad) as twitter
```

FROM public.combined

GROUP BY country

Twitter and Instagram seem to be more successful than facebook.

SELECT maritial_status,sum(facebook_ad) as totfacebook,
sum(instagram_ad) as totinsta , sum(twitter_ad) as twitter

FROM public.combined

GROUP BY maritial_status

From the table showing successful lead conversions on different social media.
Widow's use twitter the most. Again as couples and married couples and single people make up the majority of the sample they have the most lead conversions. YOLO and Absurd are outliers.
We thus ignore them. Married couples have most on instagram,
couples together have most on instagram but twitter is also high.
Twitter and Instagram seems most effective.
This shows that as graduates have the most purchases that facebook is less common for young people.
Thus spend less on facebook marketing.

SELECT sum(amtliq) as totalalcohol, sum(amtvege)as totalveg, sum(amtnonveg) as totalmeat,
sum(amtpes) as totalfish, sum(amtchocolates)as totalchoc,
sum(amtcomm) as totalcomm, country, sum(facebook_ad)as facebook,
sum(instagram_ad)as insta, sum(twitter_ad) as twitter

FROM public.combined
GROUP BY country

We see alcohol is the best seller especially in spain. Advertising should occur through twitter and instagram.

SELECT country, SUM(COALESCE(twitter_ad,instagram_ad,facebook_ad)) as totalsocialmedia,
sum(bulkmail_ad) as totalbulkmail ,sum(brochure_ad) as totalbrochure

FROM public.combined

GROUP by country
ORDER by totalsocialmedia desc

Table shows brochure is least effective.
Bulkmail or social media most effective.
(nested coalesce function used)

SELECT AVG(income) as avgincomeusd, country

from public.combined

GROUP by country
ORDER by AVG(income) desc

Middle east has highest average income while India has the lowest.

SELECT sum(numwebbuy)/sum(numvisits) as convrate,
country

from public.combined
GROUP by country

Canada has highest conversion rate. India has lowest.
We thus focus our social media and bulkmail marketing to Canada and America and less focus to india.
Middle east is an outlier due to very small sample size.

CONCLUSION AND SUGGESTIONS:

I have used an inner join between the two tables in tableau with the intersecting ID column.

Looking at the web purchases compared to walk in purchases, we see walk in is more.
Spain is the highest in web purchases by a large margin. We thus should focus our efforts on spain to maximize web purchases. The age group with the highest total sales and thus web purchases is 32-48.
We see from the breakdown that Alcohol, commodities and chocolates, which are all products that can easily made online only
make up most of the total sales, this once again supports a increase in expenditure in increasing online purchases.
Looking at the total sales by product we see a maximum in August and a minimum in July, Alcohol makes up most of the sales.
also note non veg items are well selling but people are most likely to buy this in person for freshness.
We focus on alcohol sales. The distribution of Age by marketing channel shows social media is most normally distributed while
bulkmail is right skewed and brochure is left skewed.
We focus on social media and Bulkmail where majority of lead conversions come from.
Social media seems more effective than bulkmail due to it being more normally distributed in specific, Instagram and Twitter from our
previous insights. Families with 1 or less kids/teens at home spend more.
Overall with initial insights for maximum efficiency we should target sales of alcohol
increasing spend into advertising on twitter and instagram during august in Spain.
We should also target graduates and people with 1 or less kids or teens at home.
Canada is also somewhere we should increase ad spend as it has the highest conversion rate.
Our campaigns are less successful in india and should be minimised.
By creating a calculated field I calculated a social media measure.
The distribution of age is most normally distributed in social media,while bulkmail is
right skewed and Brochures are left skewed.
We should thus use the social media marketing channel as it targets a wider range of ages.
In particular Twitter and Instagram.

To Summarize:

- Increase production of Alcohol.
- Invest more in online marketing in Spain and North america.
- Invest less in marketing in India and the Middle East.
- Address major dip in Sales in July.
- Focus marketing through Twitter and Instagram.
- Target Graduates, married and unmarried couples in the age range of 32-48.
- Families with less children spend more.
- Make Alcohol, Chocolates and commodities online to shift majority of sales to online.
- Continue selling meat, fish and vegetables in store.