

# Technical Report: 2Market Analysis

**By:** Lotti Monika Knobel

**Date:** 15/12/2024

## Background / Business Context

2Market seeks to better understand their customer demographics to inform future business strategies. This analysis focuses on three key questions:

1. Who are the customers, and what are their life circumstances?
2. Who spends how much on 2Market, and on which items?

Through this analysis, actionable insights were derived to better target key customer segments, optimize advertising strategies, and identify the most popular product categories. The findings aim to inform future business decisions and unlock growth opportunities.

Challenges arose due to small customer bases in most countries, particularly outside Spain, making it difficult to analyze trends effectively in other regions.

## Analytical Approach

### Data Cleaning in Excel

Two datasets were provided by 2Market. These datasets were cleaned and formatted to ensure consistency and usability:

- **Spell Check:** Corrected any spelling errors.
- **Formatting:** Standardized date, currency, and numerical formats.
- **Duplicates:** No duplicate entries were found.
- **Empty Cells:** No incomplete data was identified.

This cleaning process was conducted in Excel to ensure that the source data was ready for use across different tools and to minimize rework for potential collaborators.

### Additional Data Preparation

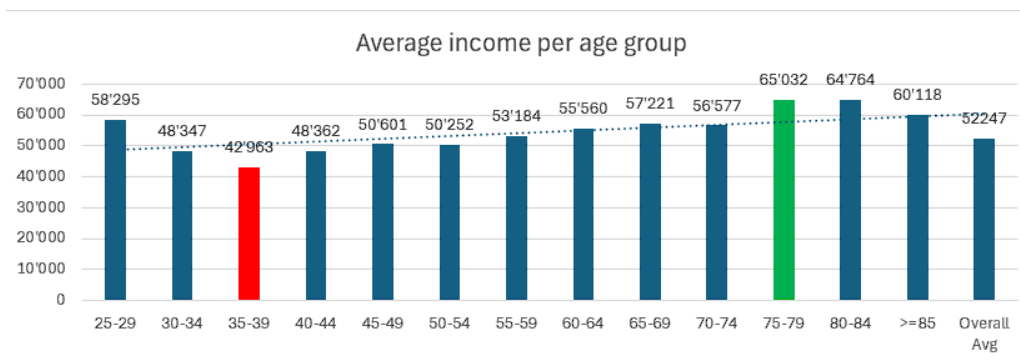
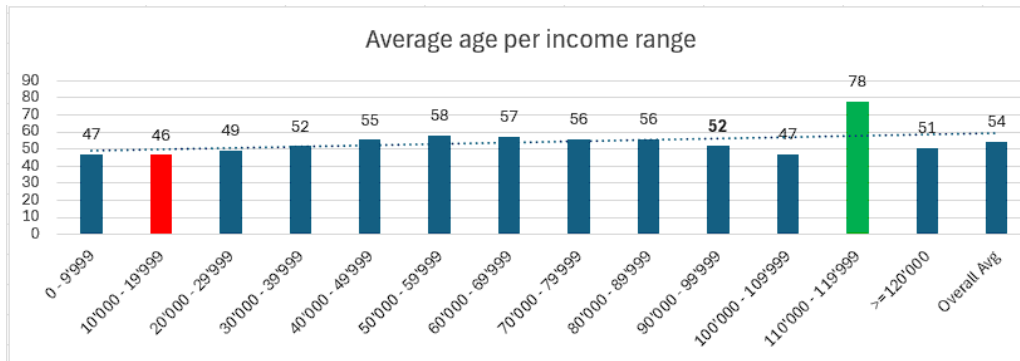
To enhance analysis, an additional column was created in Excel to calculate customer age based on their year of birth. This ensured that age could be used as a key variable in subsequent analyses.

## Initial Analysis

In **Excel**, early trends emerged:

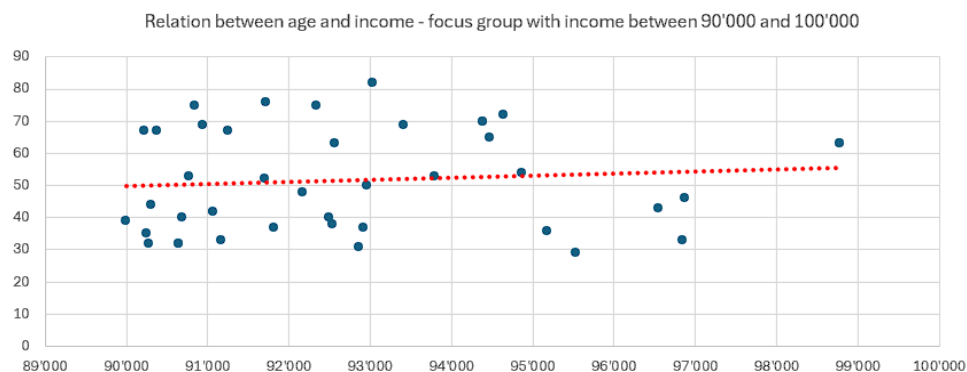
- **Age and Income Relationship:**

The initial analysis revealed a positive correlation between age and income: "the higher the age, the higher the income." This insight, visualized below, suggests that targeting older demographics might yield better financial returns. However, further validation is necessary in Tableau.



- **Specific Income Range Analysis:**

At 2Market's request, a deeper analysis was conducted on customers within a specific income range. The trend of higher income with higher age persisted within this group. However, the dataset contained relatively few customers within this range, which may limit the robustness of this finding.



For these analyses formulas with the functions FILTER and AVERAGEIFS were used to create graphs for visual representation.

## Advanced Analysis in PostgreSQL

PostgreSQL was used for advanced data integration and deeper analysis. The datasets were joined using a full outer join, simplifying further analysis. Key insights from this step included:

- **Advertising Effectiveness:** Online ads had significantly higher lead conversion rates compared to brochures (Appendices 4.1 and 4.2).

```
-- most effective advertising method per country

select
  country,
  case
    when sum(bulkmail_ad::int) > sum(twitter_ad::int) and sum(bulkmail_ad::int) > sum(instagram_ad::int) and sum(bulkmail_ad::int) > sum(facebook_ad::int) and sum(bulkmail_ad::int) > sum(brochure_ad::int) then 'bulkmail'
    when sum(twitter_ad::int) > sum(instagram_ad::int) and sum(twitter_ad::int) > sum(facebook_ad::int) and sum(twitter_ad::int) > sum(brochure_ad::int) then 'twitter'
    when sum(instagram_ad::int) > sum(twitter_ad::int) and sum(instagram_ad::int) > sum(facebook_ad::int) and sum(instagram_ad::int) > sum(brochure_ad::int) then 'instagram'
    when sum(facebook_ad::int) > sum(twitter_ad::int) and sum(facebook_ad::int) > sum(instagram_ad::int) and sum(facebook_ad::int) > sum(brochure_ad::int) then 'facebook'
    when sum(brochure_ad::int) > sum(twitter_ad::int) and sum(brochure_ad::int) > sum(instagram_ad::int) and sum(brochure_ad::int) > sum(facebook_ad::int) then 'brochure'
    else 'several'
  end as most_effective_platform,
  sum(bulkmail_ad::int) as bulkmail_success,
  sum(twitter_ad::int) as twitter_success,
  sum(instagram_ad::int) as instagram_success,
  sum(facebook_ad::int) as facebook_success,
  sum(brochure_ad::int) as brochure_success
from twomarket_data
group by country
order by country;
```

	country character (30)	most_effective_platform text	bulkmail_success bigint	twitter_success bigint	instagram_success bigint	facebook_success bigint	brochure_success bigint
1	Australia	instagram	9	6	12	7	0
2	Canada	twitter	18	24	21	18	6
3	Germany	twitter	10	11	8	7	2
4	India	bulkmail	13	10	6	7	2
5	Montenegro	bulkmail	1	0	0	0	0
6	South Africa	several	21	20	21	20	4
7	Spain	instagram	83	87	89	76	16
8	United States of America	bulkmail	8	6	5	7	0

- **Item Popularity:** Alcohol and meat emerged as the most popular product categories, irrespective of geography, marital status, or household type (Appendices 5.1 to 7.2).

```
1 -- discovering the most popular item per country
2
3 select m.country,
4 case
5   when max(m.amtliq) > max(m.amtvege) and max(m.amtliq) > max(m.amtnonveg) and max(m.amtliq) > max(m.amtpes) and max(m.amtliq) > max(m.amtchocolates) and max(m.amtliq) > max(m.amtcomm) then 'alcohol'
6   when max(m.amtvege) > max(m.amtnonveg) and max(m.amtvege) > max(m.amtpes) and max(m.amtvege) > max(m.amtchocolates) and max(m.amtvege) > max(m.amtcomm) then 'meat'
7   when max(m.amtnonveg) > max(m.amtpes) and max(m.amtnonveg) > max(m.amtchocolates) and max(m.amtnonveg) > max(m.amtcomm) then 'vegetables'
8   when max(m.amtpes) > max(m.amtchocolates) and max(m.amtpes) > max(m.amtcomm) then 'fish'
9   when max(m.amtchocolates) > max(m.amtcomm) then 'chocolates'
10  else 'commodities'
11  end as most_popular_item
12 from marketing_data m
13 group by m.country;
```

	country character (30) 🔒	most_popular_item text 🔒
1	United States of America	meat
2	South Africa	meat
3	Spain	liquor
4	Montenegro	liquor
5	Australia	liquor
6	Germany	liquor
7	Canada	liquor
8	India	meat

```

1  -- discovering the most popular item by marital status
2
3  ▼ select
4      m.marital_status,
5      case
6          when max(m.amtliq) > max(m.amtvege) and max(m.amtliq) > max(m.amtnonveg) and max(m.amtliq) > m
7          when max(m.amtvege) > max(m.amtnonveg) and max(m.amtvege) > max(m.amtpes) and max(m.amtvege) >
8          when max(m.amtnonveg) > max(m.amtpes) and max(m.amtnonveg) > max(m.amtchocolates) and max(m.am
9          when max(m.amtpes) > max(m.amtchocolates) and max(m.amtpes) > max(m.amtcomm) then 'fish'
10         when max(m.amtchocolates) > max(m.amtcomm) then 'chocolates'
11         else 'commodities'
12     end as most_popular_item
13 from marketing_data m
14 group by
15     m.marital_status;
16

```

	marital_status character varying (50) 🔒	most_popular_item text 🔒
1	Widow	liquor
2	YOLO	liquor
3	Together	meat
4	Alone	liquor
5	Absurd	meat
6	Married	meat
7	Divorced	liquor
8	Single	liquor

```

1  -- discovering the most popular item with households with and/or without kids & teenagers
2
3  ▼ select
4      case
5          when kidhome = 0 then '0 children'
6          when kidhome = 1 then '1 child'
7          when kidhome = 2 then '2 children'
8          else 'more than 2 children'
9          end as child_count,
10     case
11         when teenhome = 0 then '0 teenagers'
12         when teenhome = 1 then '1 teenager'
13         when teenhome = 2 then '2 teenagers'
14         else 'more than 2 teenagers'
15         end as teenager_count,
16     case
17         when max(amtliq) > max(amtvege) and max(amtliq) > max(amtnonveg) and max(amtliq) > max(amtpes) and
18         when max(amtvege) > max(amtnonveg) and max(amtvege) > max(amtpes) and max(amtvege) > max(amtchocol
19         when max(amtnonveg) > max(amtpes) and max(amtnonveg) > max(amtchocolates) and max(amtnonveg) > max
20         when max(amtpes) > max(amtchocolates) and max(amtpes) > max(amtcomm) then 'fish'
21         when max(amtchocolates) > max(amtcomm) then 'chocolates'
22         else 'commodities'
23     end as most_popular_item
24 from marketing_data
25 group by kidhome, teenhome
26 order by kidhome, teenhome;

```

	child_count text	teenager_count text	most_popular_item text
1	0 children	0 teenagers	meat
2	0 children	1 teenager	meat
3	0 children	2 teenagers	liquor
4	1 child	0 teenagers	meat
5	1 child	1 teenager	liquor
6	1 child	2 teenagers	liquor
7	2 children	0 teenagers	liquor
8	2 children	1 teenager	liquor

- **Regional Spending:** Spending patterns by product category and country provided a granular view of regional preferences (Appendices 8.1 and 8.2).

```

1  -- calculating the spend per country for each product group to identify the most successful product gro
2
3  ▼ select m.country,
4         sum(m.amtliq) as liquor_spend,
5         sum(m.amtvege) as vegetable_spend,
6         sum(m.amtnonveg) as meat_spend,
7         sum(m.amtpes) as fish_spend,
8         sum(m.amtchocolates) as chocolate_spend,
9         sum(m.amtcomm) as commodities_spend
10 from marketing_data m
11 group by m.country
12 order by sum(m.amtliq + m.amtvege + m.amtnonveg + m.amtpes + m.amtchocolates + m.amtcomm)desc;

```

	country character (30)	liquor_spend numeric	vegetable_spend numeric	meat_spend numeric	fish_spend numeric	chocolate_spend numeric	commodities_spend 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	Canada	84066.00	7681.00	45925.00	9980.00	7607.00	12144.00
4	Australia	42752.00	3689.00	22328.00	5546.00	4129.00	7132.00
5	India	36236.00	3788.00	23729.00	4818.00	3221.00	6014.00
6	Germany	36776.00	2980.00	20272.00	4601.00	2801.00	5768.00
7	United States of Ameri...	32214.00	3034.00	20185.00	4411.00	2863.00	4839.00
8	Montenegro	1729.00	8.00	817.00	226.00	122.00	220.00

This approach leveraged PostgreSQL's flexibility to combine datasets and extract actionable insights beyond Excel's capabilities.

## Dashboard Design and Development

To present the findings effectively, **Tableau** dashboards were developed with the following considerations:

### Demographic Dashboard

#### 1. Visualizations:

- A geographic map allows stakeholders to hover over countries and view total customer spend per country.
- A bar chart shows the largest customer base by distinct count, enabling quick identification of high-priority markets.

- Two pie charts display education levels and marital statuses, with proportional percentages to highlight distribution across demographics.

## 2. Rationale:

- A mix of chart types provided variety and clarity for different types of data.
- Interactive elements (e.g., hover feature) were incorporated for deeper stakeholder exploration.

## Spending Behavior Dashboard

### 1. Visualizations:

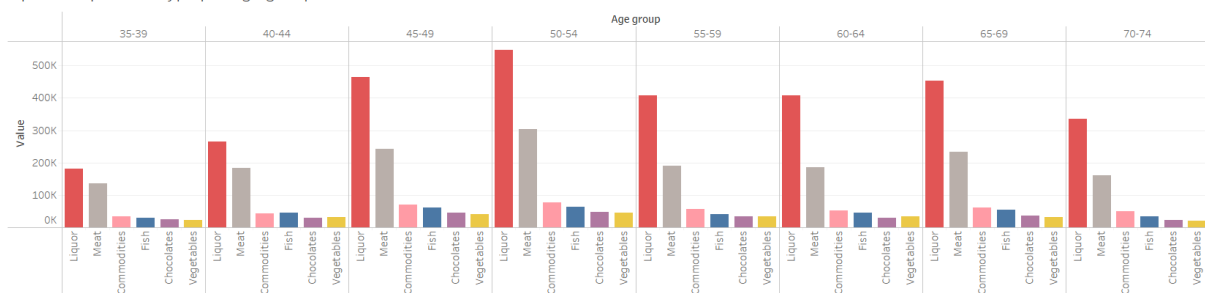
- A table summarizes total and average spend per country to give both macro and per-customer perspectives.

### Spend per age group

Age group	Total spend	Total spend percen..	Customer count	Avg spend per cust..
35-39	426.885	6%	167	2.556
40-44	597.490	9%	215	2.779
45-49	923.790	14%	369	2.503
50-54	1.083.870	16%	381	2.845
55-59	762.535	11%	260	2.933
60-64	754.335	11%	222	3.398
65-69	872.335	13%	245	3.561
70-74	623.975	9%	188	3.319

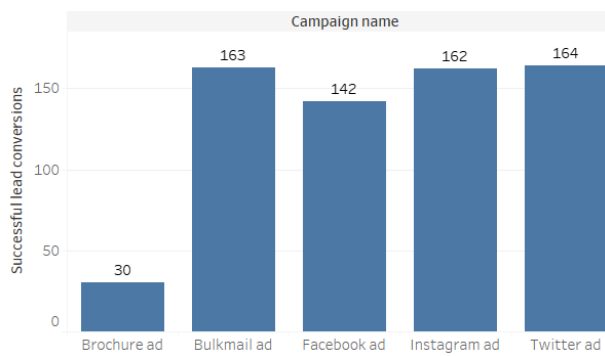
- A side-by-side bar chart depicts spending by age group and product category, filtered for distinct customer counts (default: 100).

Spend on product type per age group

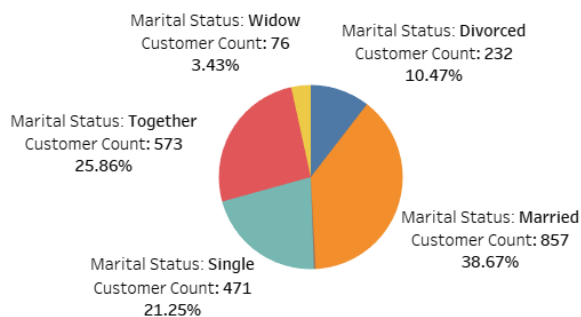
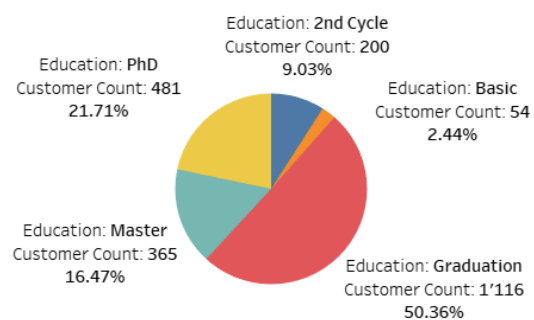


- A bar chart visualizes lead conversions by advertising channel with a country-specific filter, enabling localized campaign assessments.

## Advertising campaigns



- Two pie charts visualising the distribution of education and marital status.



- A worldmap to show the distribution of the customer base.

## Geographical demographics based on total spend



## 2. **Rationale:**

- Filtering options ensured flexibility for stakeholders to tailor insights to their needs.
- Side-by-side comparisons helped illustrate relationships between demographics and spending.

## 3. **Design Principles**

- **Interactivity:** Filters and hover functions enable stakeholders to explore specific metrics in greater depth.
- **Clarity:** Labels, legends, and proportional representations ensured insights were intuitive and actionable.
- **Accessibility:** A colorblind-friendly palette was chosen, avoiding the use of green to ensure all stakeholders can interact with the dashboards effectively.

# Patterns, Trends, and Insights

- The largest customer base falls within the age range of 50 to 55.
- Customers aged 60 to 70 have the highest average spend per person, making them a priority target demographic.
- Alcohol and meat are the most popular items across all demographics, reflecting consistent customer preferences.
- Online advertising campaigns significantly outperform brochures in terms of lead conversion rates. Brochures, as the least effective channel, warrant discontinuation.
- Spending behavior varies by region, but alcohol and meat remain dominant categories across geographies.

# Recommendations

## 1. **Target Older Demographics:**

Focus marketing efforts on customers aged 50 and above, particularly those aged 60 to 70, as they represent the highest average spend.

## 2. **Optimize Advertising Strategies:**

Shift resources away from brochure campaigns to more effective online advertising channels.

## 3. **Promote Popular Items:**

- Continue promoting alcohol and meat as key product categories.
- Consider campaigns for health-focused items to cater to potential shifts in customer preferences.

## 4. **Explore Regional Strategies:**

Use insights from spending patterns by region to develop localized marketing campaigns tailored to regional preferences.

Future analyses could investigate more specific preferences within key demographics or test alternative advertising channels to optimize performance.