2Market: Exploratory Data Analysis and Predictive Modeling of Customer Spend

Student: Monica Urquiza Ribeiro Baracho

Table of Contents

1. Overview

This report outlines key insights from an exploratory data analysis conducted for **2Market**, a global supermarket with both online and in-store. The analysis aimed to provide insights that will inform 2Market's marketing strategies by examining customer demographics, purchasing patterns, and the effectiveness of various advertising channels

To achieve this, I analysed two key datasets—**customer profiles** and **advertising conversions**—using **Excel**, **SQL**, **Tableau and R for a regression analysis**. The analysis focused on answering the following core business questions:

- Who are our customers? (age, income, education, marital status)
- Which advertising channels generate the most conversions?
- Which product categories are most popular among different customer segments?

This report outlines the full methodology, key findings, and business recommendations, supported by interactive dashboards and a multiple regression model to predict customer spending.

2. Tools and Methodology

Data Sources

- **marketing_data.csv**: Contains customer demographics, spending behaviours, income, education, and marital status.
- **ad_data.csv**: Records customer responses to various marketing channels (e.g., Instagram, Facebook, brochures).

Tools

- **Excel**: For initial data cleaning, handling missing values, calculating age, and generating basic visualisations.
- **SQL (PostgreSQL)**: Used to join datasets, calculate spending by segment, and summarize category spend by education level.
- **Tableau**: Built two interactive dashboards to present key insights to stakeholders.
- **RStudio**: regression analysis

Data Preparation

- **Cleaning**: Handled missing values, removed duplicates, and standardized data types (e.g., formatted income fields).
- **Category Grouping**: Combined "YOLO", "Alone", and "Single" into a single **"Single"** category to reflect similar behavioural patterns in spending and campaign response (see Appendix A1 – Average Age by Marital Status).
- **Invalid Values**: Replaced "Absurd" entries with #N/A to prevent skewed analysis.
- **Feature Engineering**: Calculated **age** from Year_Birth using 2025 as the reference year, and created total spend variables by category.
- **Data Join**: Merged marketing_data and ad_data using the ID field.
- **Exploration**: Ran SQL queries to identify trends across customer segments.

SQL queries[1] were used extensively to support data cleaning, aggregation, and exploration. For instance, queries were designed to calculate total product spend by marital status, identify top-spending customer segments, and group data by family type.

Regression Analysis Summary

A multiple linear regression model was developed to identify which customer attributes and ad channels best predict spending behaviour. The dependent variable, total spend, was log-transformed for normalization.

Independent variables included demographics (age, marital status, country, education) and ad exposures (Instagram, Facebook, Twitter, Bulk mail).

The model showed strong performance (Adjusted $R^2 = 0.582$), and confirmed **income and Instagram ads** as the strongest positive predictors of customer spend. **Twitter, Facebook**, and even **Bulk mail ads** also showed significant positive effects, while education level (Basic only) and country (US) were minor but significant contributors. (see **Appendix C – Regression Analysis in R**)

3. Key Findings

Three interactive Tableau dashboards were developed to present key customer insights clearly and engagingly.

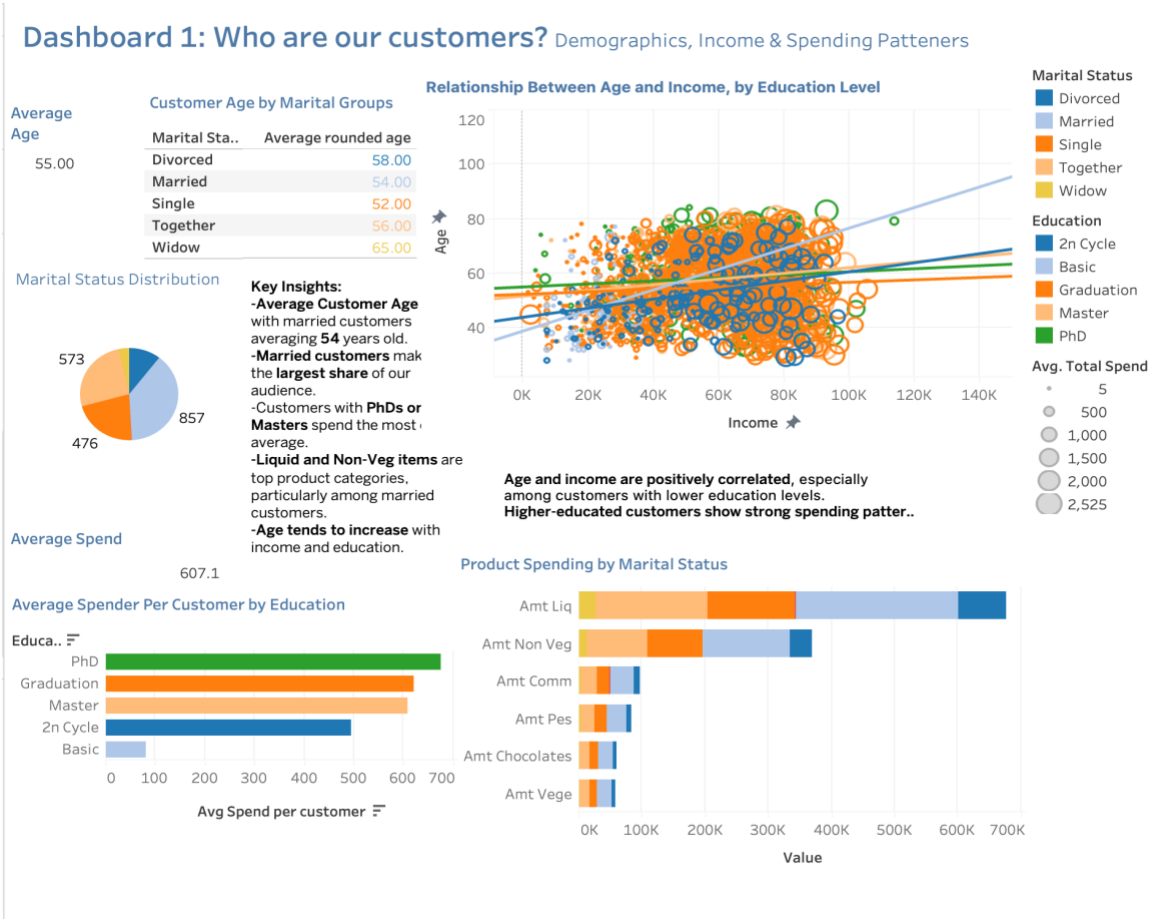
The Demographics & Spending Patterns dashboard highlights the following:

- The average customer age is approximately 55 years.
- Married individuals represent the largest customer segment, with around 857 customers.
- There is a positive correlation between age, income, and education level.
- Customers with master's or PhD degrees demonstrate the highest average spending across all segments.
- Liquor and non-vegetables are the most purchased product categories, particularly among higher-educated and married customers.

This scatterplot illustrates the relationship between **income and age**, segmented by **education level**. Each point represents an individual customer, with circle size indicating their **average total spend**.
The trend lines reveal a **positive correlation** between age and income, particularly for customers with **Basic** Nd **2n Cycle** education levels. Higher education groups (e.g., **master's** and **PhD**) show a **more stable income pattern across age**, suggesting earlier earning potential. Notably, customers with **higher spending power** tend to be **older and better educated**.

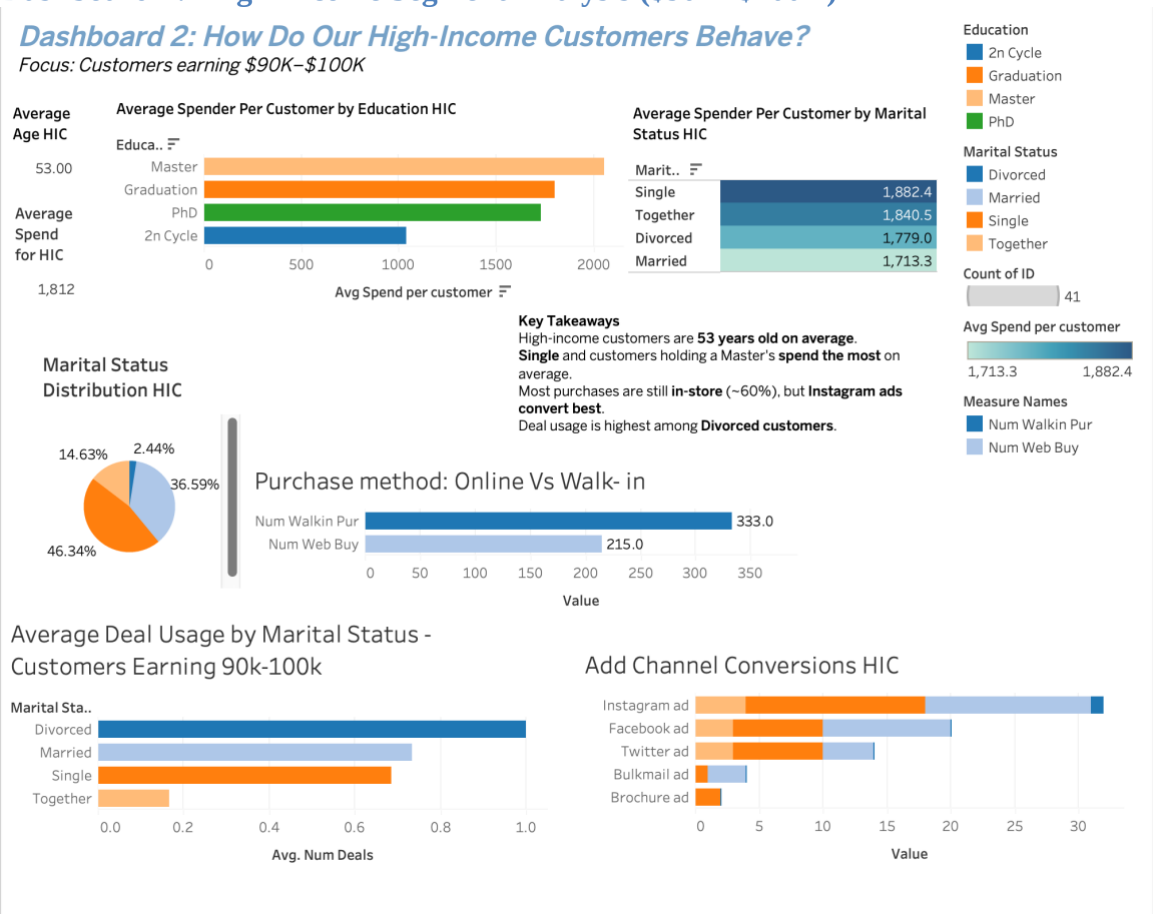
[1] The full SQL queries can be found in Appendix B – SQL Queries

Dashboard 1: Demographics & Spending



The average customer is 53 years old, with the highest spenders being single or divorced individuals. Although they earn more, they remain deal-conscious — especially the divorced group. Instagram stands out as the top ad channel, suggesting that visual, aspirational content resonates best with this segment. Surprisingly, walk-in purchases still dominate, highlighting the continued importance of in-store experiences, even for premium customers.

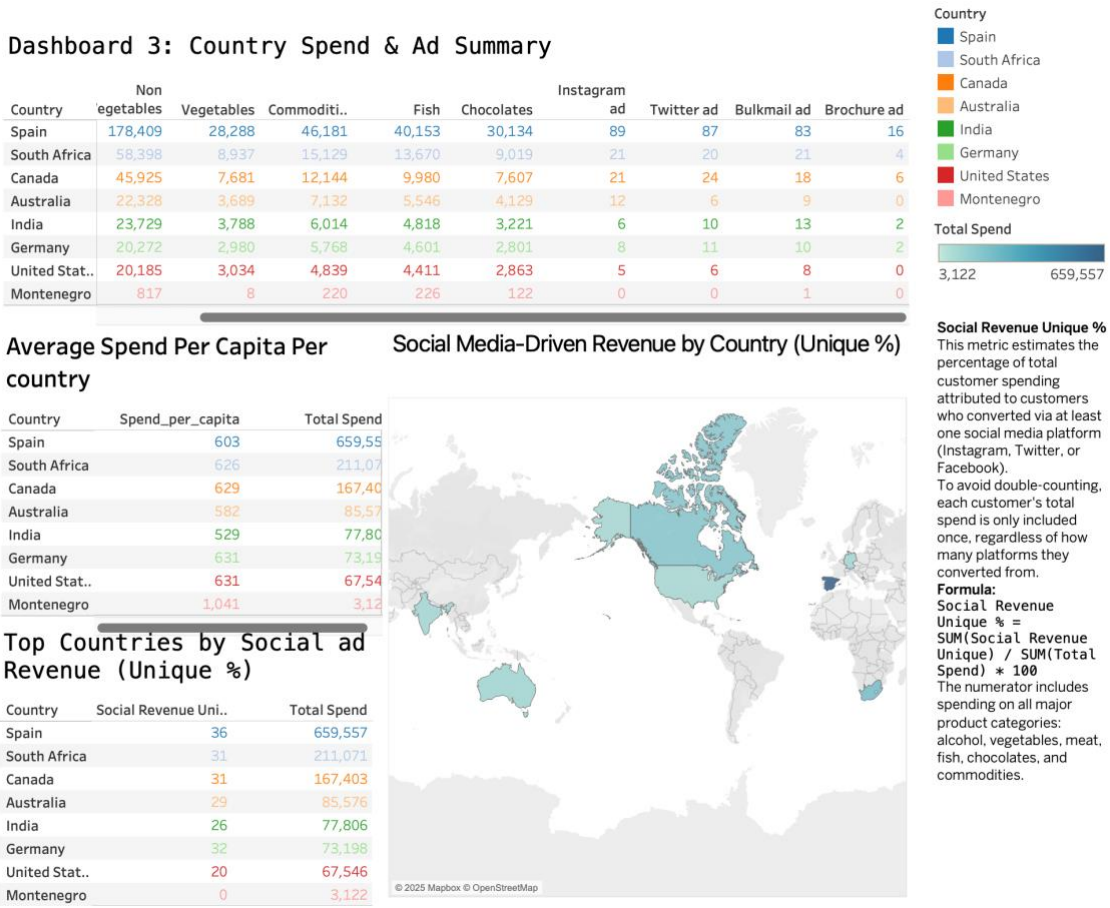
Dashboard 2: High-Income Segment Analysis ($90K–$100K)



Dashboard 3 below shows country-level customer spend, response to social media ads, and **Social Revenue Unique %** — the share of spend from customers who converted via at least one social platform.

- **Spain** had the highest total spend ($659,557) and **Social Revenue Unique %** (36%).
- **Canada** and **South Africa** followed, each with 31%.
- **Montenegro** had no revenue from social media conversions. This insight helps identify where social ads are most effective and where to focus future campaigns.

Dashboard 3: Country Spend & Ad Summary



Country	Non egetables	Vegetables	Commoditi..	Fish	Chocolates	Instagram ad	Twitter ad	Bulkmail ad	Brochure ad
Spain	178,409	28,288	46,181	40,153	30,134	89	87	83	16
South Africa	58,398	8,937	15,129	13,670	9,019	21	20	21	4
Canada	45,925	7,681	12,144	9,980	7,607	21	24	18	6
Australia	22,328	3,689	7,132	5,546	4,129	12	6	9	0
India	23,729	3,788	6,014	4,818	3,221	6	10	13	2
Germany	20,272	2,980	5,768	4,601	2,801	8	11	10	2
United Stat..	20,185	3,034	4,839	4,411	2,863	5	6	8	0
Montenegro	817	8	220	226	122	0	0	1	0

Country	Spend_per_capita	Total Spend
Spain	603	659,55
South Africa	626	211,07
Canada	629	167,40
Australia	582	85,57
India	529	77,80
Germany	631	73,19
United Stat..	631	67,54
Montenegro	1,041	3,12

Country	Social Revenue Uni..	Total Spend
Spain	36	659,557
South Africa	31	211,071
Canada	31	167,403
Australia	29	85,576
India	26	77,806
Germany	32	73,198
United Stat..	20	67,546
Montenegro	0	3,122

4. Recommendations

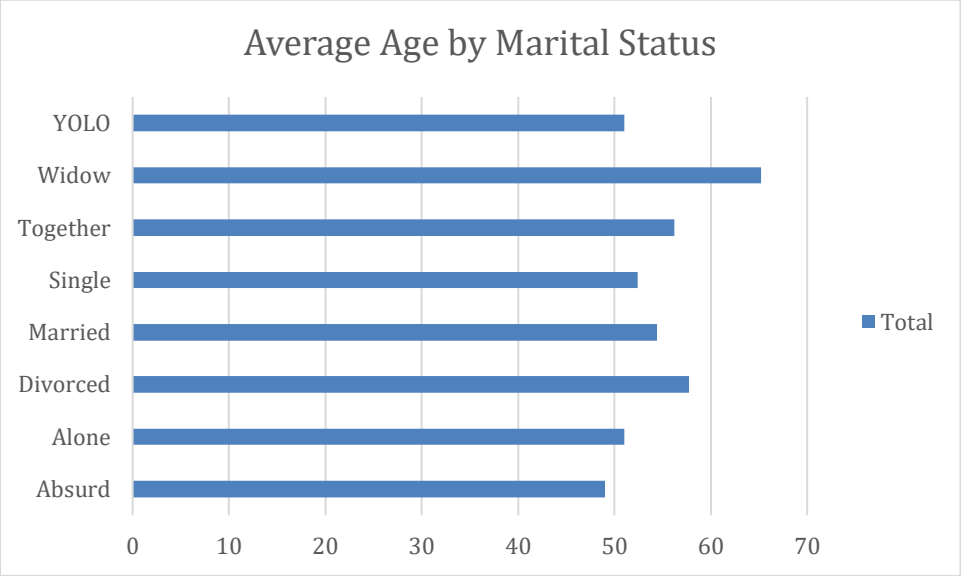
Strategy Area	Recommendation
Customer Segmentation & Targeting	Focus on married, educated, and older customers, who represent the most profitable and sizable segment.
Customer Segmentation & Targeting	Target high-income customers ($90K–$100K), particularly Single and Divorced individuals, with premium product offerings.
Advertising Strategy	Invest more heavily in Instagram and Twitter campaigns, as they drive the highest returns.
Advertising Strategy	Reallocate budget away from underperforming channels such as brochures, bulk mail, and Facebook ads.
Sales Channel Optimization	Maintain a strong in-store presence, as a majority of purchases still happen offline.
Sales Channel Optimization	Reinforce online sales with retargeting strategies for digitally active customers.

Geographic Strategy	Prioritize Spain, South Africa, and Canada for future campaigns based on spend and ad conversion performance.
Geographic Strategy	Re-evaluate strategy in the United States, where digital engagement is low despite relatively high per-capita spend.
Data Strategy Enhancements	Track product quantity and price to better understand the drivers of spend.
Data Strategy Enhancements	Incorporate behavioural variables like deal usage, product preference, and campaign history to refine segmentation models.

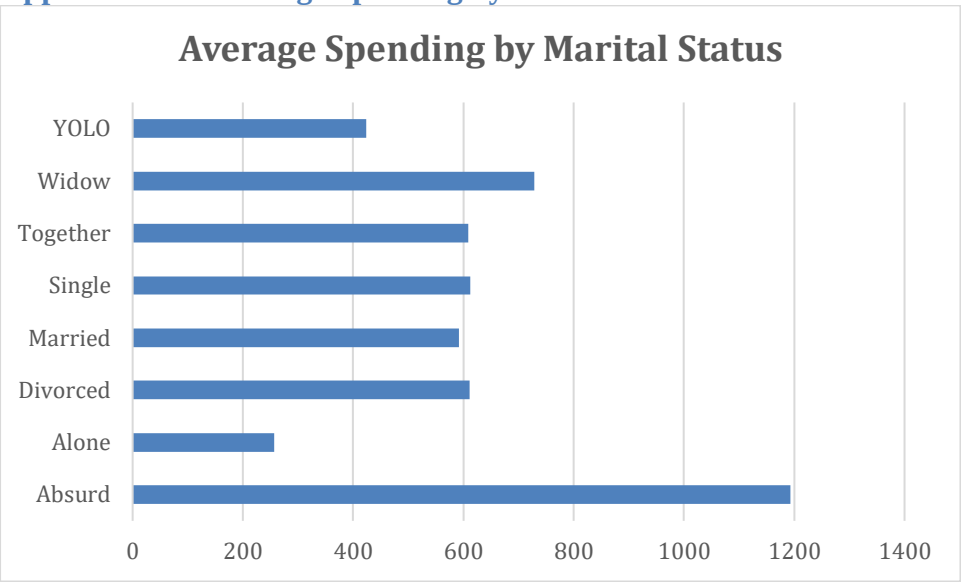
Appendices

Appendix A: Excel Visualizations

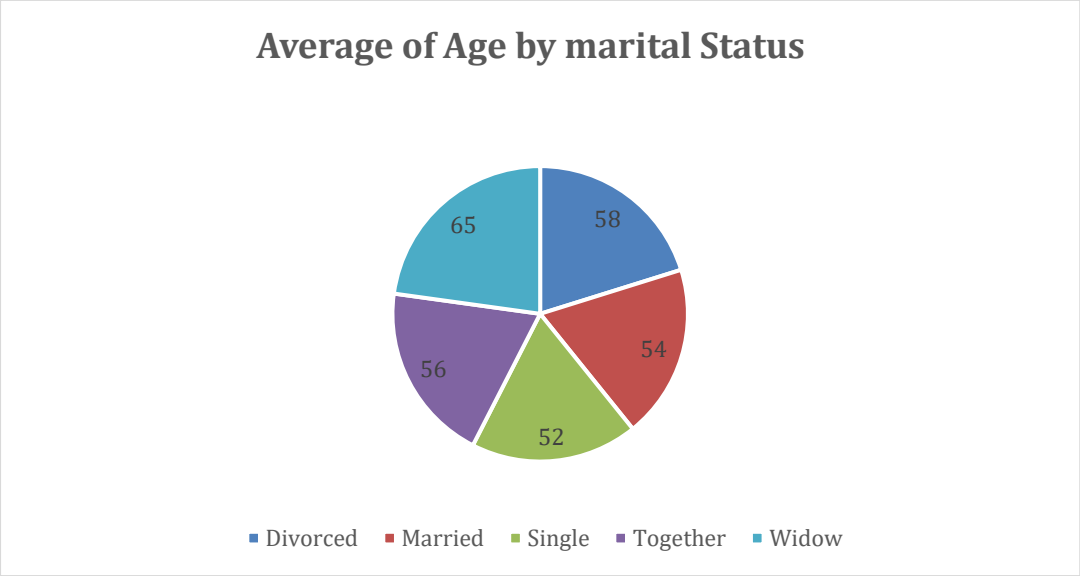
Appendix A1 – Average Age by Marital Status



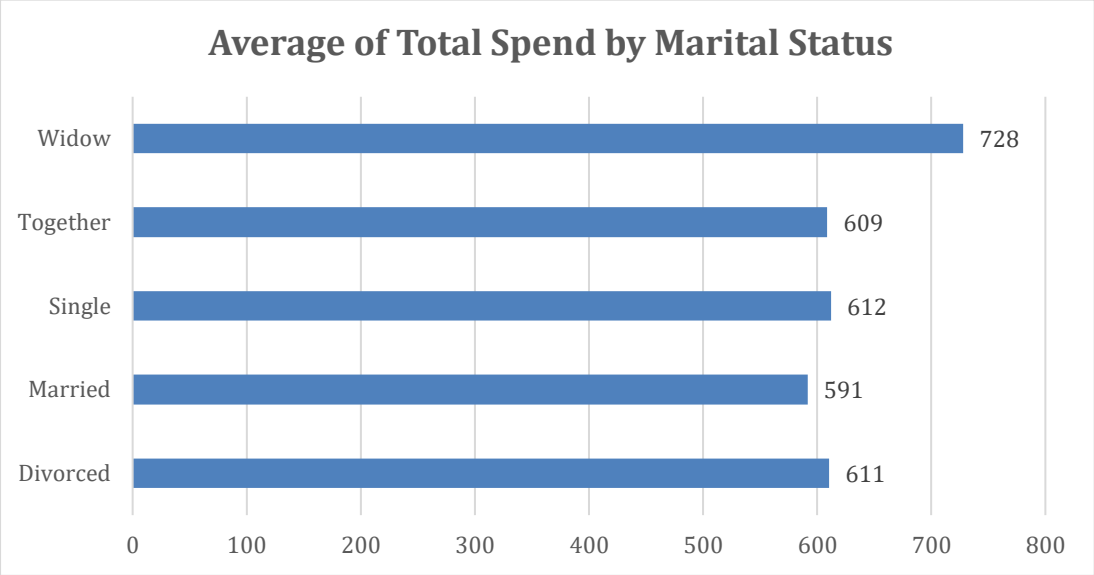
Appendix A2 – Average Spending by Marital Status



Appendix A.3 – Average of Age by Marital Status after joining "YOLO, Single and Alone" into a single category



Appendix A.4 – Average Spend by Marital Status after joining "YOLO, Single, and Alone" into a single category



Appendix B – SQL Queries

Appendix B.1– Country-Level Total Spend and Product Category Breakdown with Ranking

Query | Query History

```
58  WITH country_spend AS (
59   SELECT
60       Country,
61       SUM(AmtLiq) AS Liquor,
62       SUM(AmtVege) AS Vegetables,
63       SUM(AmtNonVeg) AS Meat,
64       SUM(AmtPes) AS Fish,
65       SUM(AmtChocolates) AS Chocolates,
66       SUM(AmtComm) AS Commodities,
67       SUM(AmtLiq + AmtVege + AmtNonVeg + AmtPes + AmtChocolates + AmtComm) AS Total_Spend
68   FROM marketing_data
69   GROUP BY Country
70  )
71  SELECT *,
72        RANK() OVER (ORDER BY Total_Spend DESC) AS Spend_Rank
73  FROM country_spend;
74
75
```

Data Output | Messages | Graph Visualiser | Notifications

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

	country text	liquor bigint	vegetables bigint	meat bigint	fish bigint	chocolates bigint	commodities bigint	total_spend bigint	spend_rank bigint
1	SP	336392	28288	178409	40153	30134	46181	659557	1
2	SA	105918	8937	58398	13670	9019	15129	211071	2
3	CA	84066	7681	45925	9980	7607	12144	167403	3
4	AUS	42752	3689	22328	5546	4129	7132	85576	4
5	IND	36236	3788	23729	4818	3221	6014	77806	5
6	GER	36776	2980	20272	4601	2801	5768	73198	6
7	US	32214	3034	20185	4411	2863	4839	67546	7
8	ME	1729	8	817	226	122	220	3122	8

Total rows: 8 | Query complete 00:00:00.093 | LF | Ln 73, Col 20

Appendix B.2 - Marital Status-Level Spend and Product Category Analysis with Rank

```sql
WITH marital_spend AS (
    SELECT
        Marital_Status,
        SUM(AmtLiq) AS Liquor,
        SUM(AmtVege) AS Vegetables,
        SUM(AmtNonVeg) AS Meat,
        SUM(AmtPes) AS Fish,
        SUM(AmtChocolates) AS Chocolates,
        SUM(AmtComm) AS Commodities,
        SUM(AmtLiq + AmtVege + AmtNonVeg + AmtPes + AmtChocolates + AmtComm) AS Total_Spend
    FROM marketing_data
    GROUP BY Marital_Status
)
SELECT *,
       RANK() OVER (ORDER BY Total_Spend DESC) AS Spend_Rank
FROM marital_spend;
```

Data Output | Messages | Graph Visualiser | Notifications

Showing rows: 1 to 5 | Page No: 1 of 1

	marital_status text	liquor bigint	vegetables bigint	meat bigint	fish bigint	chocolates bigint	commodities bigint	total_spend bigint	spend_rank bigint
1	Married	256976	21981	137888	30395	22926	36719	506885	1
2	Together	176715	14612	95374	22383	15031	24754	348869	2
3	Single	139126	13027	87868	18704	12839	20970	292534	3
4	Divorced	75364	6363	34848	8130	6222	10739	141666	4
5	Widow	27902	2422	14085	3793	2878	4245	55325	5

Total rows: 5 | Query complete 00:00:00.094 | LF | Ln 130, Col 1

Appendix B.3 - Family Type and Most Popular Product Category Analysis

Query | Query History

```
147 WITH family_product_spend AS (
148     SELECT
149         CASE WHEN Kidhome + Teenhome = 0 THEN 'No Children/Teens' ELSE 'Has Children/Teens' END AS Family_T
150         'Liquor' AS Product,
151         SUM(AmtLiq) AS Total_Spend
152     FROM marketing_data
153     GROUP BY Family_Type
154
155     UNION ALL
156     SELECT
157         CASE WHEN Kidhome + Teenhome = 0 THEN 'No Children/Teens' ELSE 'Has Children/Teens' END AS Family_T
158         'Vegetables',
159         SUM(AmtVege)
160     FROM marketing_data
161     GROUP BY Family_Type
162
163     UNION ALL
```

Data Output | Messages | Graph Visualiser | Notifications

Showing rows: 1 to 2 | Page No: 1 of 1

	family_type text	most_popular_product text	total_spend bigint
1	Has Children/Teens	Liquor	367133
2	No Children/Teens	Liquor	308950

Total rows: 2 | Query complete 00:00:00.074 | LF | Ln 206, Col 1

Appendix B.4 - Country-Level Advertising Channel Conversion Analysis

```
209
210  -- Step 2: Total Lead Conversions by Social Media Platform per Country
211  SELECT
212      m.Country,
213      SUM(CAST(a.Twitter_ad AS INT)) AS Twitter_Leads,
214      SUM(CAST(a.Instagram_ad AS INT)) AS Instagram_Leads,
215      SUM(CAST(a.Facebook_ad AS INT)) AS Facebook_Leads,
216      SUM(CAST(a.Bulkmail_ad AS INT)) AS Bulkmail_Leads,
217      SUM(CAST(a.Brochure_ad AS INT)) AS Brochure_Leads
218  FROM marketing_data m
219  JOIN ad_data a ON m.ID = a.ID
220  GROUP BY m.Country
221  ORDER BY m.Country;
222  -- Step 3 (updated): Total Lead Conversions by Ad Channel per Marital Status
223  SELECT
224      m.Marital_Status,
225      SUM(CAST(a.Twitter_ad AS INT)) AS Twitter_Leads,
226      SUM(CAST(a.Instagram_ad AS INT)) AS Instagram_Leads,
```

Data Output | Messages | Graph Visualiser | Notifications

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

	country text	twitter_leads bigint	instagram_leads bigint	facebook_leads bigint	bulkmail_leads bigint	brochure_leads bigint
1	AUS	6	12	7	9	0
2	CA	24	21	18	18	6
3	GER	11	8	7	10	2
4	IND	10	6	7	13	2
5	ME	0	0	0	1	0
6	SA	20	21	20	21	4
7	SP	87	89	76	83	16
8	US	6	5	7	8	0

Total rows: 8 | Query complete 00:00:00.087 | LF | Ln 221, Col 20

Appendix B.5 - Ad Channel Conversion Analysis by Marital Status

```
221  ORDER BY m.Country;
222  -- Step 3 (updated): Total Lead Conversions by Ad Channel per Marital Status
223  SELECT
224      m.Marital_Status,
225      SUM(CAST(a.Twitter_ad AS INT)) AS Twitter_Leads,
226      SUM(CAST(a.Instagram_ad AS INT)) AS Instagram_Leads,
227      SUM(CAST(a.Facebook_ad AS INT)) AS Facebook_Leads,
228      SUM(CAST(a.Bulkmail_ad AS INT)) AS Bulkmail_Leads,
229      SUM(CAST(a.Brochure_ad AS INT)) AS Brochure_Leads
230  FROM marketing_data m
231  JOIN ad_data a ON m.ID = a.ID
232  GROUP BY m.Marital_Status
233  ORDER BY m.Marital_Status;
234
235  -- Step 4 (updated): Product Spend + All Ad Channels per Country
236  SELECT
237      m.Country,
```

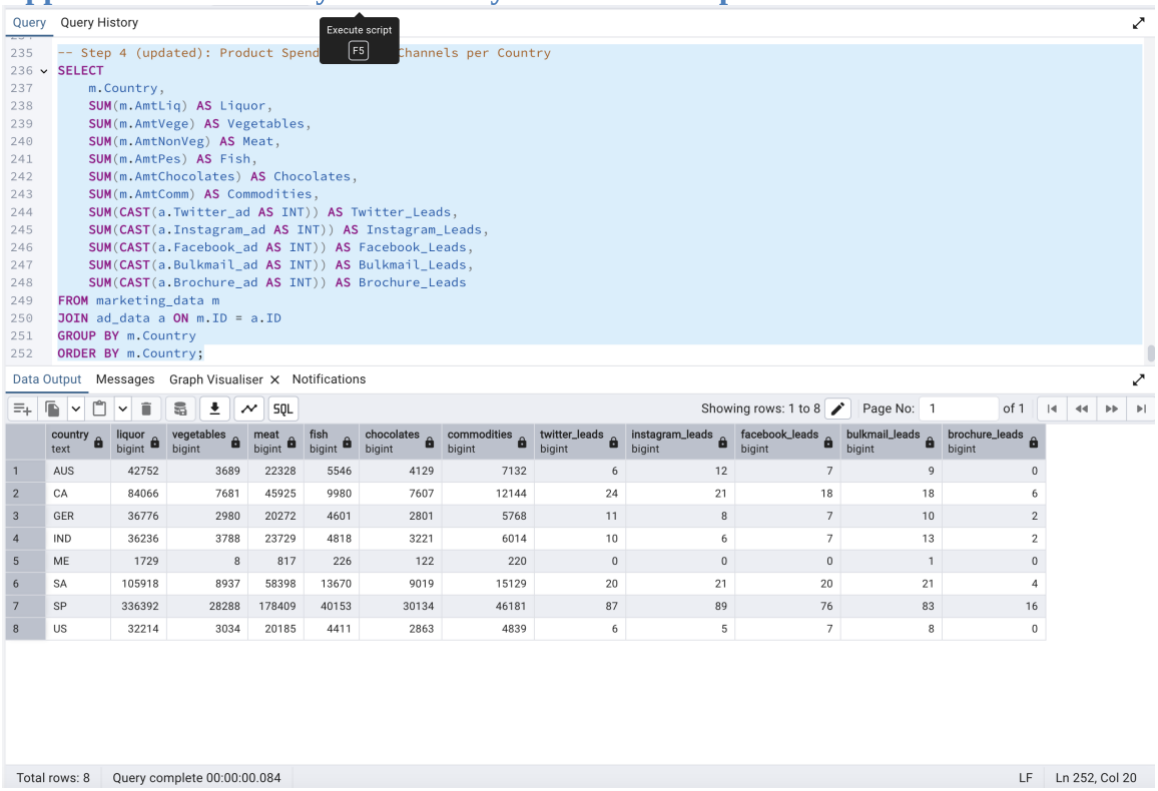
Data Output | Messages | Graph Visualiser | Notifications

Showing rows: 1 to 5 | Page No: 1 of 1

	marital_status text	twitter_leads bigint	instagram_leads bigint	facebook_leads bigint	bulkmail_leads bigint	brochure_leads bigint
1	Divorced	18	13	12	20	5
2	Married	62	66	62	63	7
3	Single	32	32	31	39	5
4	Together	42	44	32	37	12
5	Widow	10	7	5	4	1

Total rows: 5 | Query complete 00:00:00.090 | LF | Ln 233, Col 27

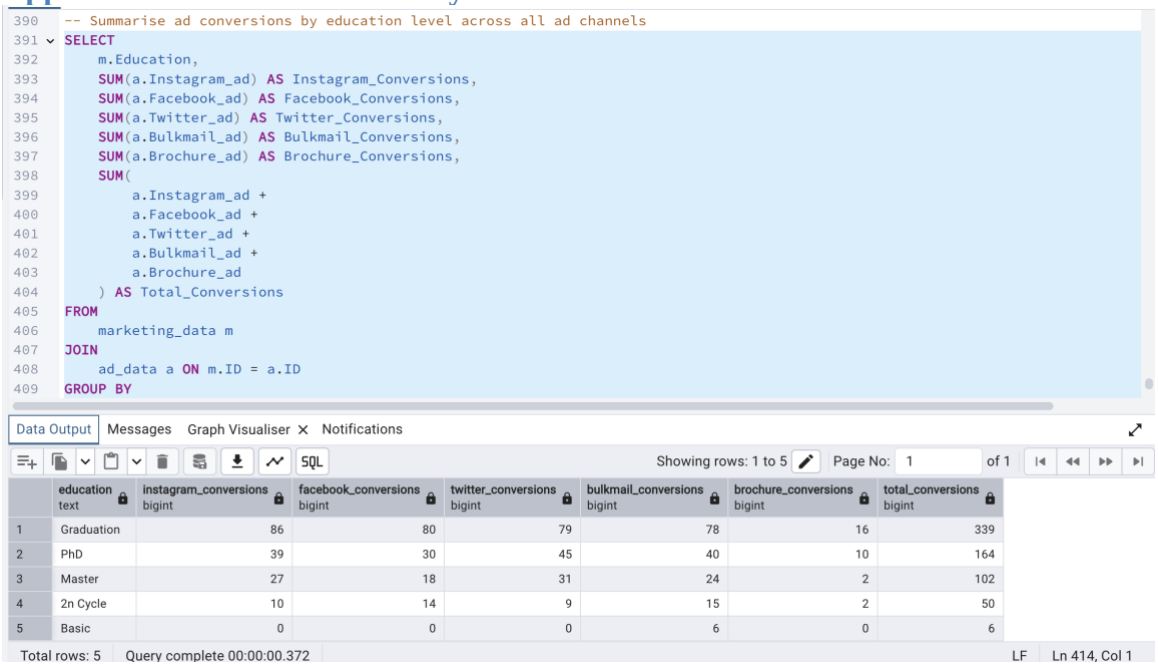
Appendix B.6 - Country-Level Analysis of Product Spend and Ad Channel Conversions



```sql
-- Step 4 (updated): Product Spend & Channels per Country
SELECT
    m.Country,
    SUM(m.AmtLiq) AS Liquor,
    SUM(m.AmtVege) AS Vegetables,
    SUM(m.AmtNonVeg) AS Meat,
    SUM(m.AmtPes) AS Fish,
    SUM(m.AmtChocolates) AS Chocolates,
    SUM(m.AmtComm) AS Commodities,
    SUM(CAST(a.Twitter_ad AS INT)) AS Twitter_Leads,
    SUM(CAST(a.Instagram_ad AS INT)) AS Instagram_Leads,
    SUM(CAST(a.Facebook_ad AS INT)) AS Facebook_Leads,
    SUM(CAST(a.Bulkmail_ad AS INT)) AS Bulkmail_Leads,
    SUM(CAST(a.Brochure_ad AS INT)) AS Brochure_Leads
FROM marketing_data m
JOIN ad_data a ON m.ID = a.ID
GROUP BY m.Country
ORDER BY m.Country;
```

	country text	liquor bigint	vegetables bigint	meat bigint	fish bigint	chocolates bigint	commodities bigint	twitter_leads bigint	instagram_leads bigint	facebook_leads bigint	bulkmail_leads bigint	brochure_leads bigint
1	AUS	42752	3689	22328	5546	4129	7132	6	12	7	9	0
2	CA	84066	7681	45925	9980	7607	12144	24	21	18	18	6
3	GER	36776	2980	20272	4601	2801	5768	11	8	7	10	2
4	IND	36236	3788	23729	4818	3221	6014	10	6	7	13	2
5	ME	1729	8	817	226	122	220	0	0	0	1	0
6	SA	105918	8937	58398	13670	9019	15129	20	21	20	21	4
7	SP	336392	28288	178409	40153	30134	46181	87	89	76	83	16
8	US	32214	3034	20185	4411	2863	4839	6	5	7	8	0

Appendix B.8 – ad conversions by education level across all ad channels



```sql
-- Summarise ad conversions by education level across all ad channels
SELECT
    m.Education,
    SUM(a.Instagram_ad) AS Instagram_Conversions,
    SUM(a.Facebook_ad) AS Facebook_Conversions,
    SUM(a.Twitter_ad) AS Twitter_Conversions,
    SUM(a.Bulkmail_ad) AS Bulkmail_Conversions,
    SUM(a.Brochure_ad) AS Brochure_Conversions,
    SUM(
        a.Instagram_ad +
        a.Facebook_ad +
        a.Twitter_ad +
        a.Bulkmail_ad +
        a.Brochure_ad
    ) AS Total_Conversions
FROM
    marketing_data m
JOIN
    ad_data a ON m.ID = a.ID
GROUP BY
```

	education text	instagram_conversions bigint	facebook_conversions bigint	twitter_conversions bigint	bulkmail_conversions bigint	brochure_conversions bigint	total_conversions bigint
1	Graduation	86	80	79	78	16	339
2	PhD	39	30	45	40	10	164
3	Master	27	18	31	24	2	102
4	2n Cycle	10	14	9	15	2	50
5	Basic	0	0	0	6	0	6

Appendix B.10 - Top Advertising Channel by Country (Based on Leads)

Query | Query History

```
    CASE
        WHEN Country = 'AUS' THEN 'Australia'
        WHEN Country = 'CA' THEN 'Canada'
        WHEN Country = 'GER' THEN 'Germany'
        WHEN Country = 'IND' THEN 'India'
        WHEN Country = 'ME' THEN 'Montenegro'
        WHEN Country = 'SA' THEN 'South Africa'
        WHEN Country = 'SP' THEN 'Spain'
        WHEN Country = 'US' THEN 'United States'
        ELSE Country
    END AS Country_Name,
    Channel AS Top_Channel,
    Leads
FROM ranked_channels
WHERE channel_rank = 1
ORDER BY Country_Name;
```

Data Output | Messages | Graph Visualiser | Notifications

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

	country_name text	top_channel text	leads bigint
1	Australia	Instagram	12
2	Canada	Twitter	24
3	Germany	Twitter	11
4	India	Bulkmail	13
5	Montenegro	Bulkmail	1
6	South Africa	Bulkmail	21
7	Spain	Instagram	89
8	United States	Bulkmail	8

Total rows: 8 | Query complete 00:00:00.081 | LF | Ln 390, Col 1

Appendix B.11 - average amount spent on each product category by households, segmented by: Number of kids or Teens

```
743
744 SELECT
745     Kidhome,
746     Teenhome,
747     CASE
748         WHEN Kidhome = 0 AND Teenhome = 0 THEN 'No Kids or Teens'
749         WHEN Kidhome > 0 AND Teenhome = 0 THEN 'Kids Only'
750         WHEN Kidhome = 0 AND Teenhome > 0 THEN 'Teens Only'
751         WHEN Kidhome > 0 AND Teenhome > 0 THEN 'Kids and Teens'
752     END AS family_type,
753     ROUND(AVG(AmtLiq), 1) AS avg_liquor,
754     ROUND(AVG(AmtVege), 1) AS avg_vegetables,
755     ROUND(AVG(AmtChocolates), 1) AS avg_chocolates,
756     ROUND(AVG(AmtComm), 1) AS avg_commercial,
757     ROUND(AVG(AmtPes), 1) AS avg_fish,
758     ROUND(AVG(AmtNonVeg), 1) AS avg_meat
759 FROM marketing_data
760 GROUP BY Kidhome, Teenhome
761 ORDER BY Kidhome, Teenhome;
762
```

Data Output | Messages | Graph Visualiser | Notifications

	kidhome integer	teenhome integer	family_type text	avg_liquor numeric	avg_vegetables numeric	avg_chocolates numeric	avg_commercial numeric	avg_fish numeric	avg_meat numeric
1	0	0	No Kids or Teens	488.1	52.3	53.2	64.2	76.6	370.9
2	0	1	Teens Only	417.7	27.2	28.8	55.9	36.6	139.3
3	0	2	Teens Only	409.6	20.7	19.1	57.0	33.7	133.5
4	1	0	Kids Only	82.4	9.9	9.4	21.3	14.6	49.2
5	1	1	Kids and Teens	124.1	6.5	7.4	22.8	9.4	45.5
6	1	2	Kids and Teens	276.0	12.9	10.1	28.9	8.6	110.1
7	2	0	Kids Only	61.2	14.6	8.1	28.0	13.4	42.1
8	2	1	Kids and Teens	78.3	1.0	1.4	10.3	3.0	23.1

Appendix C – Regression Analysis in R

RStudio

```
+     geom_text(aes(label = sig), hjust = -0.4, size = 3, color = "black") +
+     geom_hline(yintercept = 0, linetype = "dashed", color = "gray50") +
+     coord_flip() +
+     labs(
+         title = "Significant Predictors of Customer Spend",
+         x = "Predictors",
+         y = "Estimated Effect on log(Total Spend)"
+     ) +
+     theme_minimal(base_size = 12)
>
> summary(model_extended)

Call:
lm(formula = log_TotalSpend ~ log_Income + Age + Marital_Status +
    Country + Instagram_ad + Facebook_ad + Twitter_ad + Bulkmail_ad +
    Education, data = merged_data)

Residuals:
    Min      1Q  Median      3Q     Max
-6.8844 -0.5111  0.1746  0.6173  7.7939

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)         -16.926303   0.491583 -34.432  < 2e-16 ***
log_Income            2.075932   0.045884  45.243  < 2e-16 ***
Age                   0.001305   0.001799   0.725  0.46840
Marital_StatusMarried 0.042550   0.070989   0.599  0.54898
Marital_StatusSingle  0.068375   0.077024   0.888  0.37480
Marital_StatusTogether 0.005160  0.074295   0.069  0.94464
Marital_StatusWidow   0.132797   0.126671   1.048  0.29459
CountryCA             0.037886   0.098026   0.386  0.69917
CountryGER           -0.033888   0.118637  -0.286  0.77517
CountryIND            0.073932   0.111551   0.663  0.50755
CountryME             0.993883   0.556022   1.787  0.07400 .
CountrySA             0.039270   0.094365   0.416  0.67734
CountrySP             0.032422   0.083861   0.387  0.69907
CountryUS             0.279282   0.121236   2.304  0.02134 *
Instagram_ad          0.382508   0.090192   4.241 2.32e-05 ***
Facebook_ad           0.305324   0.092802   3.290  0.00102 **
Twitter_ad            0.354257   0.083169   4.259 2.14e-05 ***
Bulkmail_ad           0.207063   0.078727   2.630  0.00859 **
EducationBasic        0.313455   0.149946   2.090  0.03669 *
EducationGraduation   0.009794   0.073542   0.133  0.89407
EducationMaster      -0.059605   0.084647  -0.704  0.48141
EducationPhD         -0.092434   0.081508  -1.134  0.25690
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9507 on 2194 degrees of freedom
Multiple R-squared:  0.5859,    Adjusted R-squared:  0.582
F-statistic: 147.8 on 21 and 2194 DF,  p-value: < 2.2e-16
```

Environment		History		Connections		Tutorial
Import Dataset	12 MiB	List				
R	Global Environment					

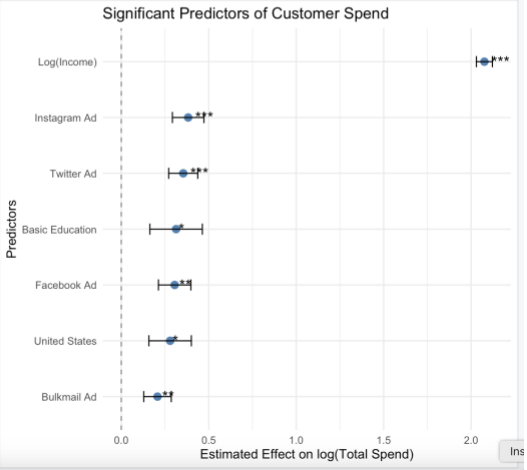
Data

ad_data	2216 obs. of 6 variables
ad_long	661 obs. of 26 variables
avg_revenue	34 obs. of 4 variables
marketing_data_clean…	2216 obs. of 23 variables
merged_data	2216 obs. of 33 variables
model	Large lm (13 elements, 797 kB)
model_campaign	Large lm (12 elements, 754.8 kB)
model_extended	Large lm (13 elements, 1.1 MB)
model_interact	Large lm (12 elements, 773.4 kB)
model_log	Large lm (12 elements, 763.6 kB)
model_log_campaign	Large lm (12 elements, 793.2 kB)
model_responders	List of 12

Large lm (12 elements, 754.8 kB)

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish



RStudio

```
+     ) +
+     theme_minimal()
> library(broom)
> write_csv(tidy(model), "regression_model_results.csv")
Error in write_csv(tidy(model), "regression_model_results.csv") :
  could not find function "write_csv"
> merged_data$predicted <- predict(model)
>
> ggplot(merged_data, aes(x = log_TotalSpend, y = predicted)) +
+     geom_point(alpha = 0.3, color = "dodgerblue") +
+     geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
+     labs(title = "Predicted vs Actual (log) Spend",
+          x = "Actual log Spend",
+          y = "Predicted log Spend") +
+     theme_minimal()
>
> library(readr)
> write_csv(tidy(model), "regression_model_results.csv")
>
> merged_data <- merged_data %>%
+     mutate(predicted = predict(model)) %>%
+     filter(predicted < quantile(predicted, 0.99))  # Keep only bottom 99% for visual
>
> ggplot(merged_data, aes(x = log_TotalSpend, y = predicted)) +
+     geom_point(alpha = 0.3, color = "dodgerblue") +
+     geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
+     labs(
+         title = "Predicted vs Actual (log) Spend",
+         x = "Actual log Spend",
+         y = "Predicted log Spend"
+     ) +
+     theme_minimal()
>
> model_tidy <- tidy(model) %>%
+     mutate(sig = ifelse(p.value < 0.001, "***",
+                         ifelse(p.value < 0.01, "**",
+                                ifelse(p.value < 0.05, "*", ""))))
>
> ggplot(merged_data, aes(x = log_TotalSpend, y = predicted)) +
+     geom_point(alpha = 0.3, color = "dodgerblue") +
+     geom_smooth(method = "lm", se = FALSE, color = "blue", linetype = "solid") +  # THIS IS YOUR TREND LINE
+     geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
+     labs(
+         title = "Predicted vs Actual (log) Spend",
+         x = "Actual log Spend",
+         y = "Predicted log Spend"
+     ) +
+     theme_minimal()
`geom_smooth()` using formula = 'y ~ x'
>
>
```

Environment		History		Connections		Tutorial
Import Dataset	15 MiB	List				
R	Global Environment					

Data

ad_data	2216 obs. of 6 variables
ad_long	661 obs. of 26 variables
avg_revenue	34 obs. of 4 variables
marketing_data_clea…	2216 obs. of 23 variables
merged_data	2193 obs. of 34 variables
model	Large lm (13 elements, 797 kB)
model_campaign	Large lm (12 elements, 754.8 kB)
model_extended	Large lm (13 elements, 1.1 MB)
model_interact	Large lm (12 elements, 773.4 kB)
model_log	Large lm (12 elements, 763.6 kB)
model_log_campaign	Large lm (12 elements, 793.2 kB)
model_responders	List of 12

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish