2Market Data Analysis - Technical Documentation

Sujith Kumaar K C (Sept 9 2024)

1. Problem Statement:

2Market seeks insights into customer purchase behaviour across demographics, advertising channels, and product preferences

Stakeholders: 2Market's C-Suite, Global and Regional Sales and Marketing teams, Product Management teams, Technology partners

2. Business Objective Analysis:

While the 2Market team has outlined WHAT data they need, it's crucial to understand WHY. A more comprehensive view of potential business objectives obtained through 5-WHY technique could be:

- Increasing profitability by boosting revenue and reducing marketing, procurement costs
- Improving market positioning in specific demographics, regions
- Enhancing customer experience through marketing, optimized online/in-store experiences.

3. Data Analysis Framework:

Data assumptions defined based on metadata and data exploration to avoid ambiguity:

Data Assumptions:

1. "Count_success" in marketing_data.csv indicates total successful lead conversions across all channels.
2. All customers receive marketing, as no opt-out data is available.
3. Customer registration data spans July 2012 to June 2014, representing a snapshot as of June 30, 2014.
4. Due to the short timeframe, seasonality analysis is omitted to avoid incorrect trend identification.
5. Averages are used for metric comparisons to mitigate population skews.
6. Deal purchases (Num_Deals) are utilized across in-store and online shopping as count of deals is less than or equal to total purchases.
7. Social media success is attributed to online purchases, hence Bulkmail and Brochure success is attributed to in-store purchases.
8. The "Response" column indicates successful conversion for the latest campaign, while other marketing data relates to previous campaigns within the period.

Data analysis was approached using a **combination of SMART and 5W1H** frameworks categorizing analysis into 3 broad categories in line with the problem statement. Identified data insights will help inform the business insights to achieve stated business objectives.

3.1. Customer Demographics (Who, What, Where):

Specific Goal	Analyse demographic characteristics of customers.
Measurable Data	Age, Education, Marital Status, Income, Family Composition (Kids and Teens), Country
Approach	Who and where are the customers? - Segment customers based on demographic factors. What do they purchase? - Identify patterns and trends within segments. - Cross-reference with purchasing behaviour.

3.2 Advertising Channel Effectiveness (How, Where):

Specific Goal	Determine the most successful advertising methods for lead conversions.
Measurable Data	Bulkmail_ad, Twitter_ad, Instagram_ad, Facebook_ad, Brochure_ad, Count_success
Approach	How effective is each marketing channel? - Calculate conversion rates for each channel Where are each marketing channel effective? - Analyse effectiveness across different demographic segments - Consider technological aspects (online vs. traditional channels)

3.3 Product Sales Analysis (What, Why):

Specific Goal	Identify best-selling products and correlations with demographics.
Measurable Data	Revenue from -Liquor, Vegetables, Meat, Fish, Chocolates, Commodities
Approach	What are the product categories driving the revenue? - Rank products by sales volume Why are they performing well and others not performing well? - Analyse sales patterns across demographic segments

4. Implementation:

4. 1. Data Preparation and Cleaning – MS Excel :

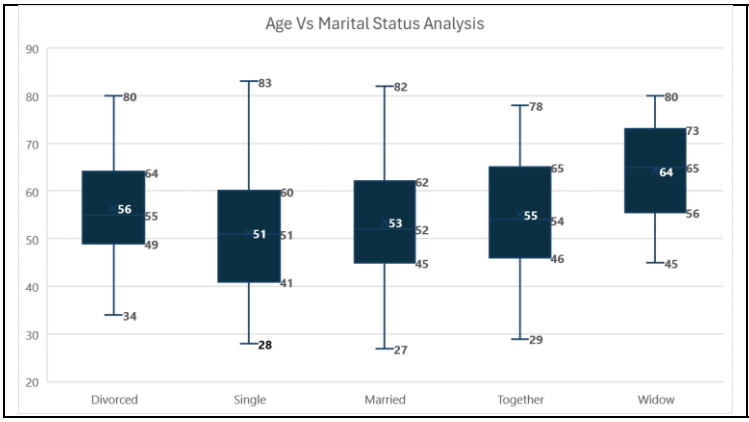
- Imported to excel as a table "marketing_data_raw"
- Cleanup performed on copy
- Blanks, Spelling errors corrected using SORT, FILTER, FIND functions.
- Data types checked and corrected as necessary using formulae and data type changes into cleaned columns, retaining original data for reference.
    - Income converted from text to currency using `=VALUE(RIGHT(G2,LEN(G2)-1))` and Data type select.
    - Retained boolean columns "Response" and "Complain" as numeric to get aggregates.
    - Calculated "Age" using `=YEAR(NOW())-YEAR_BIRTH`
    - Identified and removed Age outliers with AVERAGE, MAX and MIN functions – **3 rows of customers with age>115 were removed.**
    - Cleaned "Marital Status" using a lookup array and VLOOKUP function:
      `=IFNA(VLOOKUP([@[Marital_Status]],MSLookupArray,1,FALSE),"UNKNOWN")`
    - Converting "Dt_Customer" text values to uniform date format using TEXTSPLIT, IF, and DATE functions:
      `=DATE(IF(LEN(X2)=2,CONCATENATE("20",X2),X2),V2,W2)`
- 47 duplicates identified when Customer ID column excluded. None were removed as Customer ID is unique.

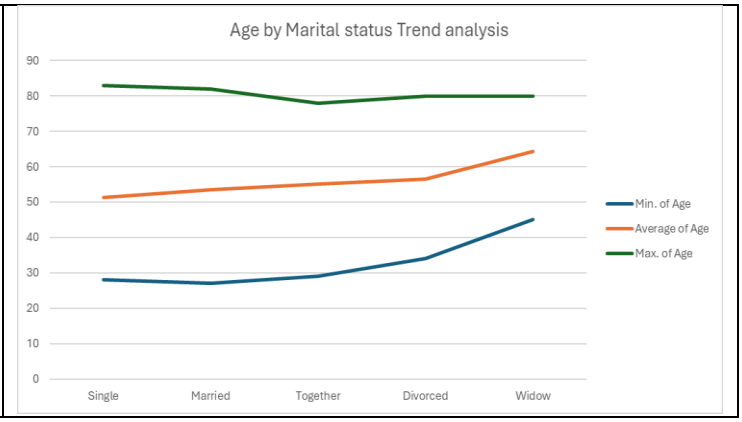
ID	Year_Birth	Age	Education	Marital_Status	Marital_Status_Clean	Income	IncomeClean
492	1974	=YEAR(NOW())-marketing_data_clean!$B105	PhD	YOLO	=IFNA(VLOOKUP(marketing_data_clean!$E105,MSLookupArray,1,FALSE),"UNKNOWN")	$48,432.00	=VALUE(RIGHT(G105,LEN(G105)-1))
11133	1974	50	PhD	YOLO	UNKNOWN	$48,432.00	$48,432.00
4491	1950	74	Master	Married	Married	$62,845.00	$62,845.00
873	1950	74	Master	Married	Married	$62,845.00	$62,845.00
1631	1966	58	PhD	Together	Together	$65,220.00	$65,220.00
8278	1991	33	PhD	Married	Married	$74,214.00	$74,214.00
6815	1981	43	2n Cycle	Married	Married	$96,547.00	$96,547.00

4.2 Exploratory Data Analysis – MS Excel:

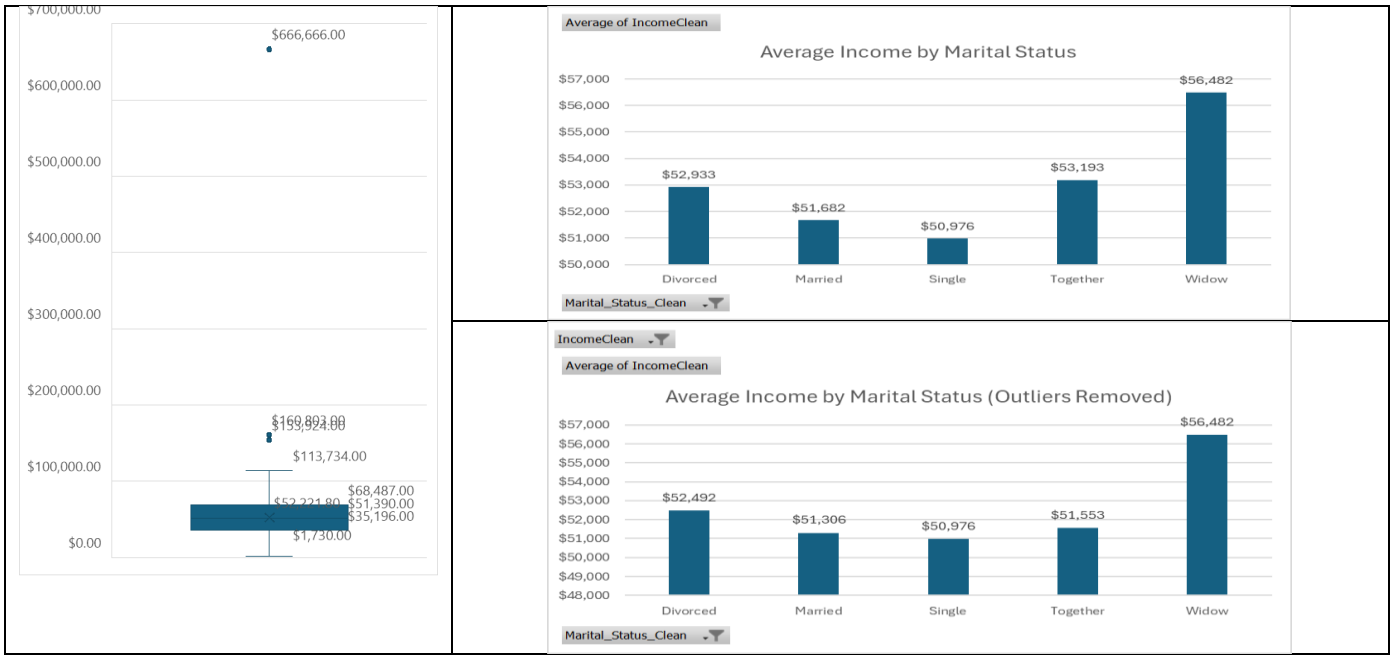
Visualized demographic distributions – created box plots charts to check for Age outliers & line graph to check trends.

Marital Status	Min. of Age	Average of Age	Max. of Age
Divorced	34	56	80
Married	27	53	82
Single	28	51	83
Together	29	55	78
Widow	45	64	80
Grand Total	**27**	**54**	**83**





Identified **income outliers** above $113k and an extreme outlier at $666k. Decided to keep these outliers but filter them out as needed for analysis – performed in Tableau as Data Filters for $666k. Compared charts with and without outliers and found minimal impact, confirming the decision to retain them.



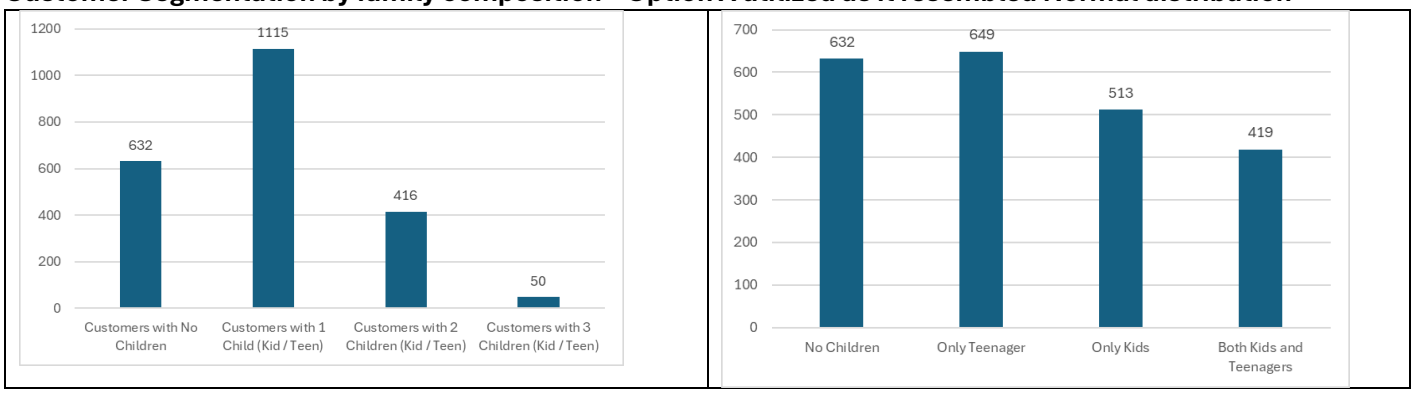
Are NumDeals already included in WebBuy/InStore Purchases?

Analysis via Excel (Nested IFs & PIVOT Table) revealed NumDeals is a subset. Therefore, it was excluded from the total purchase calculation.

```
=IF(D3=0,"No Purchase",IF(A3<D3,"Subset",IF(D3=A3,"Full Deal","Superset?"))).
```

Row Labels	Count of Classification
Subset	2197
No Purchase	10
Full Deal	5
Superset?	1
Grand Total	**2213**

Customer Segmentation by family composition – Option A utilized as it resembled Normal distribution



4.3 SQL and Excel Data Analysis:

SQL queries were written to join tables and obtain best selling product by country, best media channel by country. The logic used to determine these results were broken down from SQL into workable components and implemented in Tableau using Calculated fields to be able to analyse and compare results quickly – and develop the dashboard in parallel. SQL results were also exported to Excel for quick validations and comparisons.

SQL Query Screenshot – Most Popular Product in each country:

```sql
/* 3. Which products are the most popular in each country */
SELECT "Country",
	CASE
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtLiq") THEN 'Liquor'
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtVege") THEN 'Vegetables'
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtNonVeg") THEN 'Meat'
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtPes") THEN 'Fish'
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtChocolates") THEN 'Chocolates'
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtComm") THEN 'Commodities'
	END AS Most_Popular_Product,
	CASE
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtLiq") THEN sum("AmtLiq")
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtVege") THEN sum("AmtVege")
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtNonVeg") THEN sum("AmtNonVeg")
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtPes") THEN sum("AmtPes")
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtChocolates") THEN sum("AmtChocolates")
		WHEN GREATEST(sum("AmtLiq"), sum("AmtVege"), sum("AmtNonVeg"), sum("AmtPes"), sum("AmtChocolates"), sum("AmtComm")) = sum("AmtComm") THEN sum("AmtComm")
	END AS Product_Sale_Amount
FROM marketing_data
GROUP BY "Country";
```

Tableau calculated field utilizing the same logic – Best Selling Product for any filtered context of a worksheet:

Formula

```
IF SUM([Amt Liq]) >= SUM([Amt Vege]) AND SUM([Amt Liq]) >= SUM([Amt Non Veg]) AND SUM([Amt Liq]) >= SUM([Amt
Pes]) AND SUM([Amt Liq]) >= SUM([Amt Chocolates]) AND SUM([Amt Liq]) >= SUM([Amt Comm]) THEN 'Liquor'
ELSEIF SUM([Amt Vege]) >= SUM([Amt Liq]) AND SUM([Amt Vege]) >= SUM([Amt Non Veg]) AND SUM([Amt Vege]) >=
SUM([Amt Pes]) AND SUM([Amt Vege]) >= SUM([Amt Chocolates]) AND SUM([Amt Vege]) >= SUM([Amt Comm]) THEN
'Vegetables'
ELSEIF SUM([Amt Non Veg]) >= SUM([Amt Liq]) AND SUM([Amt Non Veg]) >= SUM([Amt Vege]) AND SUM([Amt Non Veg]) >=
SUM([Amt Pes]) AND SUM([Amt Non Veg]) >= SUM([Amt Chocolates]) AND SUM([Amt Non Veg]) >= SUM([Amt Comm]) THEN
'Meat'
ELSEIF SUM([Amt Pes]) >= SUM([Amt Liq]) AND SUM([Amt Pes]) >= SUM([Amt Vege]) AND SUM([Amt Pes]) >= SUM([Amt Non
Veg]) AND SUM([Amt Pes]) >= SUM([Amt Chocolates]) AND SUM([Amt Pes]) >= SUM([Amt Comm]) THEN 'Fish'
ELSEIF SUM([Amt Chocolates]) >= SUM([Amt Liq]) AND SUM([Amt Chocolates]) >= SUM([Amt Vege]) AND SUM([Amt
Chocolates]) >= SUM([Amt Non Veg]) AND SUM([Amt Chocolates]) >= SUM([Amt Pes]) AND SUM([Amt Chocolates]) >=
SUM([Amt Comm]) THEN 'Chocolates'
ELSEIF SUM([Amt Comm]) >= SUM([Amt Liq]) AND SUM([Amt Comm]) >= SUM([Amt Vege]) AND SUM([Amt Comm]) >= SUM([Amt
Non Veg]) AND SUM([Amt Comm]) >= SUM([Amt Pes]) AND SUM([Amt Comm]) >= SUM([Amt Chocolates]) THEN 'Commodities'
END
```

SQL JOIN Query & Result:

```sql
/*Which social media platform (Twitter, Instagram, or Facebook) is the most effective method of advertising in each country?*/
SELECT
country,
SUM(count_success) AS Total_LeadConv,
sum(twitter_ad) AS Total_TwitterConv,
sum(insta_ad) AS Total_InstaConv,
sum(fb_ad) AS Total_FbConv,
CASE
    WHEN GREATEST(sum(twitter_ad),sum(insta_ad),sum(fb_ad))=sum(fb_ad)THEN 'Facebook'
    WHEN GREATEST(sum(twitter_ad),sum(insta_ad),sum(fb_ad))=sum(insta_ad)THEN 'Instagram'
    WHEN GREATEST(sum(twitter_ad),sum(insta_ad),sum(fb_ad))=sum(twitter_ad)THEN 'Twitter'
END AS Most_Effective_Method
FROM marketing_data mk
INNER JOIN ad_data ad
USING (cust_id)
GROUP BY country
ORDER BY sum(count_success)desc;
```

Data Output | Messages | Notifications

	country character varying (4)	total_leadconv bigint	total_twitterconv bigint	total_instaconv bigint	total_fbconv bigint	most_effective_method text
1	SP	351	87	89	76	Instagram
2	CA	87	24	21	18	Twitter
3	SA	86	20	21	20	Instagram
4	GER	38	11	8	7	Twitter
5	IND	38	10	6	7	Twitter
6	AUS	34	6	12	7	Instagram
7	US	26	6	5	7	Facebook
8	ME	1	0	0	0	Facebook

A scenario where Excel, SQL and Tableau were utilized together to arrive at calculating marketing revenue:

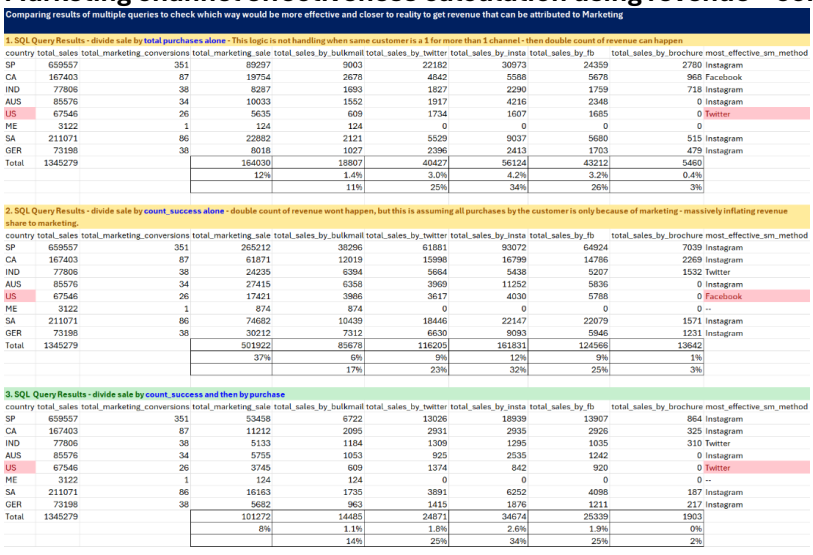
SQL Query using a subquery – to obtain social media effectiveness by revenue (using purchase and sales):

```
/*3. Which social media channel seem to be the most effective per country in terms of sales - Using Purchases and Revenue*/
SELECT
country,sum(total_sale)AS total_sales,
sum(num_web_buy) AS online_purchases,
sum(num_visits) AS online_visits,
sum(num_walkin_pur) AS instore_purchases,
sum(num_deals) AS total_deals,
sum(count_success) AS total_marketing_conversions,
(sum(bulkmail_sales)+sum(twitter_sales)+sum(insta_sales)+sum(facebook_sales)+sum(brochure_sales)) AS total_marketing_sale,
sum(bulkmail_sales) AS total_sales_by_bulkmail,
sum(twitter_sales) AS total_sales_by_twitter,
sum(insta_sales) AS total_sales_by_insta,
sum(facebook_sales) AS total_sales_by_fb,
sum(brochure_sales) AS total_sales_by_brochure,
CASE
    WHEN GREATEST(sum(twitter_sales),sum(insta_sales),sum(facebook_sales))=sum(facebook_sales)THEN 'Facebook'
    WHEN GREATEST(sum(twitter_sales),sum(insta_sales),sum(facebook_sales))=sum(insta_sales)THEN 'Instagram'
    WHEN GREATEST(sum(twitter_sales),sum(insta_sales),sum(facebook_sales))=sum(twitter_sales)THEN 'Twitter'
END AS Most_Effective_SM_Method
FROM
(SELECT cust_id,country,total_sale,count_success,num_walkin_pur,num_web_buy,num_visits,num_deals,
CASE WHEN count_success > 0 AND num_walkin_pur > 0 THEN
    ((total_sale/count_success) * bulkmail_ad)/num_walkin_pur
ELSE 0 END AS bulkmail_sales,
CASE WHEN count_success > 0 AND num_web_buy > 0 THEN
     ((total_sale/count_success) * twitter_ad)/num_web_buy
ELSE 0 END AS twitter_sales,
CASE WHEN count_success > 0 AND num_web_buy > 0 THEN
    ((total_sale/count_success) * insta_ad)/num_web_buy
ELSE 0 END AS insta_sales,
CASE WHEN count_success > 0 AND num_web_buy > 0 THEN
    ((total_sale/count_success) * fb_ad)/num_web_buy
ELSE 0 END AS facebook_sales,
CASE WHEN count_success > 0 AND num_walkin_pur > 0 THEN
    ((total_sale/count_success) * brochure_ad)/num_walkin_pur
ELSE 0 END AS brochure_sales
FROM marketing_data join ad_data USING (cust_id)
)GROUP BY country;
```

	country character var	total_sales bigint	online_purchases bigint	online_visits bigint	instore_purch bigint	total_deals bigint	total_marketing_conversions bigint	total_marketing_sale bigint	total_sales_by_bulkmail bigint	total_sales_by_twitter bigint	total_sales_by_insta bigint	total_sales_by_fb bigint	total_sales_by_brochure bigint	most_effective_sm_method text
1	SP	659557	4374	5810	6330	2462	351	53458	6722	13026	18939	13907	864	Instagram
2	CA	167403	1142	1407	1554	640	87	11212	2095	2931	2935	2926	325	Instagram
3	IND	77806	584	814	784	365	38	5133	1184	1309	1295	1035	310	Twitter
4	AUS	85576	595	762	830	334	34	5755	1053	925	2535	1242	0	Instagram
5	US	67546	479	598	650	285	26	3745	609	1374	842	920	0	Twitter
6	ME	3122	18	18	19	7	1	124	124	0	0	0	0	Facebook
7	SA	211071	1397	1782	1988	815	86	16163	1735	3891	6252	4098	187	Instagram
8	GER	73198	464	596	700	241	38	5682	963	1415	1876	1211	217	Instagram

Excel analysis on SQL queried results:

Marketing channel effectiveness calculation using revenue – comparing results of different approaches:



Comparing results of multiple queries to check which way would be more effective and closer to reality to get revenue that can be attributed to Marketing

1. SQL Query Results - divide sale by total purchases alone - This logic is not handling when same customer is a 1 for more than 1 channel - then double count of revenue can happen

country	total_sales	total_marketing_conversions	total_marketing_sale	total_sales_by_bulkmail	total_sales_by_twitter	total_sales_by_insta	total_sales_by_fb	total_sales_by_brochure	most_effective_sm_method
SP	659557	351	89297	9003	22182	30973	24359	2780	Instagram
CA	167403	87	19754	2678	4842	5588	5678	968	Facebook
IND	77806	38	8287	1693	1827	2290	1759	718	Instagram
AUS	85576	34	10033	1552	1917	4216	2348	0	Instagram
US	67546	26	5635	609	1734	1607	1685	0	Twitter
ME	3122	1	124	124	0	0	0	0	
SA	211071	86	22882	2121	5529	9037	5680	515	Instagram
GER	73198	38	8018	1027	2396	2413	1703	479	Instagram
Total	1345279		164030	18807	40427	56124	43212	5460	
			12%	1.4%	3.0%	4.2%	3.2%	0.4%	
				11%	25%	34%	26%	3%	

2. SQL Query Results - divide sale by count_success alone - double count of revenue wont happen, but this is assuming all purchases by the customer is only because of marketing - massively inflating revenue share to marketing.

country	total_sales	total_marketing_conversions	total_marketing_sale	total_sales_by_bulkmail	total_sales_by_twitter	total_sales_by_insta	total_sales_by_fb	total_sales_by_brochure	most_effective_sm_method
SP	659557	351	265212	38296	61881	93072	64924	7039	Instagram
CA	167403	87	61871	12019	15998	16799	14786	2269	Instagram
IND	77806	38	24235	6394	5664	5438	5207	1532	Twitter
AUS	85576	34	27415	6358	3969	11252	5836	0	Instagram
US	67546	26	17421	3986	3617	4030	5788	0	Facebook
ME	3122	1	874	874	0	0	0	0	--
SA	211071	86	74682	10439	18446	22147	22079	1571	Instagram
GER	73198	38	30212	7312	6630	9093	5946	1231	Instagram
Total	1345279		501922	85678	116205	161831	124566	13642	
			37%	6%	9%	12%	9%	1%	
				17%	23%	32%	25%	3%	

3. SQL Query Results - divide sale by count_success and then by purchase

country	total_sales	total_marketing_conversions	total_marketing_sale	total_sales_by_bulkmail	total_sales_by_twitter	total_sales_by_insta	total_sales_by_fb	total_sales_by_brochure	most_effective_sm_method
SP	659557	351	53458	6722	13026	18939	13907	864	Instagram
CA	167403	87	11212	2095	2931	2935	2926	325	Instagram
IND	77806	38	5133	1184	1309	1295	1035	310	Twitter
AUS	85576	34	5755	1053	925	2535	1242	0	Instagram
US	67546	26	3745	609	1374	842	920	0	Twitter
ME	3122	1	124	124	0	0	0	0	--
SA	211071	86	16163	1735	3891	6252	4098	187	Instagram
GER	73198	38	5682	963	1415	1876	1211	217	Instagram
Total	1345279		101272	14485	24871	34674	25339	1903	
			8%	1.1%	1.8%	2.6%	1.9%	0%	
				14%	25%	34%	25%	2%	

Tableau calculated fields were created based on the logic used in Query 3 (Count_succes and then by purchase):

brochure_rev

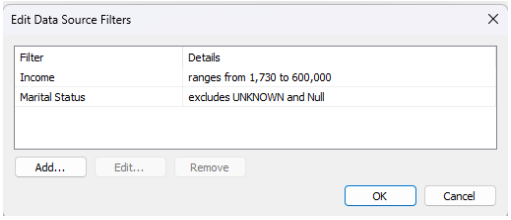
```
IF (([Count success]>0) AND ([Num Walkin Pur]>0)) THEN
(([Total Revenue]/[Count success])*[Brochure ad])/[Num Walkin Pur]
ELSE 0
END

/* Logic derived from SQL Query (Assignment Module 5):
To measure Media effectiveness using means other than conversion, Revenue and Purch
```

4.4 Dashboard development in Tableau:

Iterative Approach:

- Began with exploratory analysis results from Excel, guiding the creation of initial views for customer demographics and product purchases.
- Data source filters were applied to exclude the Income and Marital Status outliers.



- Continuous refinement of visualizations based on new insights and business requirements.

Visualization Techniques:

- Moved from a geographical map to a highlight table for country-wise analysis, enhancing comparison of revenue and customer base.
- Developed **41 calculated fields** to develop visualizations that could summarize the findings effectively, ensuring accurate representation and allow for proper exploration across segments. [Full list available in appendix]
- Integrated filters and KPIs to allow dynamic exploration and highlight key performance indicators for all 3 dashboards.
- Implemented Dynamic Zone Visibility in Marketing Dashboard to allow for detailed and dynamic exploration of each channel.

Business Objective Focus:

- Dashboards were tailored to meet the needs of 2Market's global and regional management teams, ensuring that insights could be drawn at global, continental, or country levels.
- Customized metrics such as Best-Selling Products were dynamically updated based on filter conditions, aligning with the objective to optimize sales strategies.

Inclusivity and Ethics:

- Care was taken to ensure that no demographic group was excluded, transitioning from segmenting customers by marital status to a broader approach focusing on key factors like age, income, and family size.
- Visualization designs adhered to ethical guidelines, ensuring transparency and accessibility across all customer segments.

Tableau – Calculated fields examples:

LifeStage

```
IF
[Age]<25 THEN "Youth"
ELSEIF
[Age]<=34 THEN "Young Adults"
ELSEIF
[Age]<=44 THEN "Established Adults"
ELSEIF
[Age]<=54 THEN "Mid-Life"
ELSEIF
[Age]<=64 THEN "Pre-Retirees"
ELSEIF
[Age]<=74 THEN "Young-Retirees"
ELSEIF
[Age]<=90 THEN "Senior-Retirees"
END
```

brochure_rev

```
IF (([Count success]>0) AND ([Num Walkin Pur]>0)) THEN
(([Total Revenue]/[Count success])*[Brochure ad])/[Num Walkin Pur]
ELSE 0
END

/* Logic derived from SQL Query (Assignment Module 5):
To measure Media effectiveness using means other than conversion, Revenue and Purch
If any media conversion was successful (count_success>0),
the individual customer's total revenue generated amount will be divided by count_s
This divided amount will then be multiplied by Brochure ad's data.

If Brochure was unsuccessful for the customer - will be multiplied by 0 and no reve
If Brochure was successful, will be multiplied by 1 and the divided revenue amount
This value will then be divided by Number of Walkin Purchases to obtain a Per Purch
- to avoid inflating the value (As the Revenue amount is across multiple purchases)
```

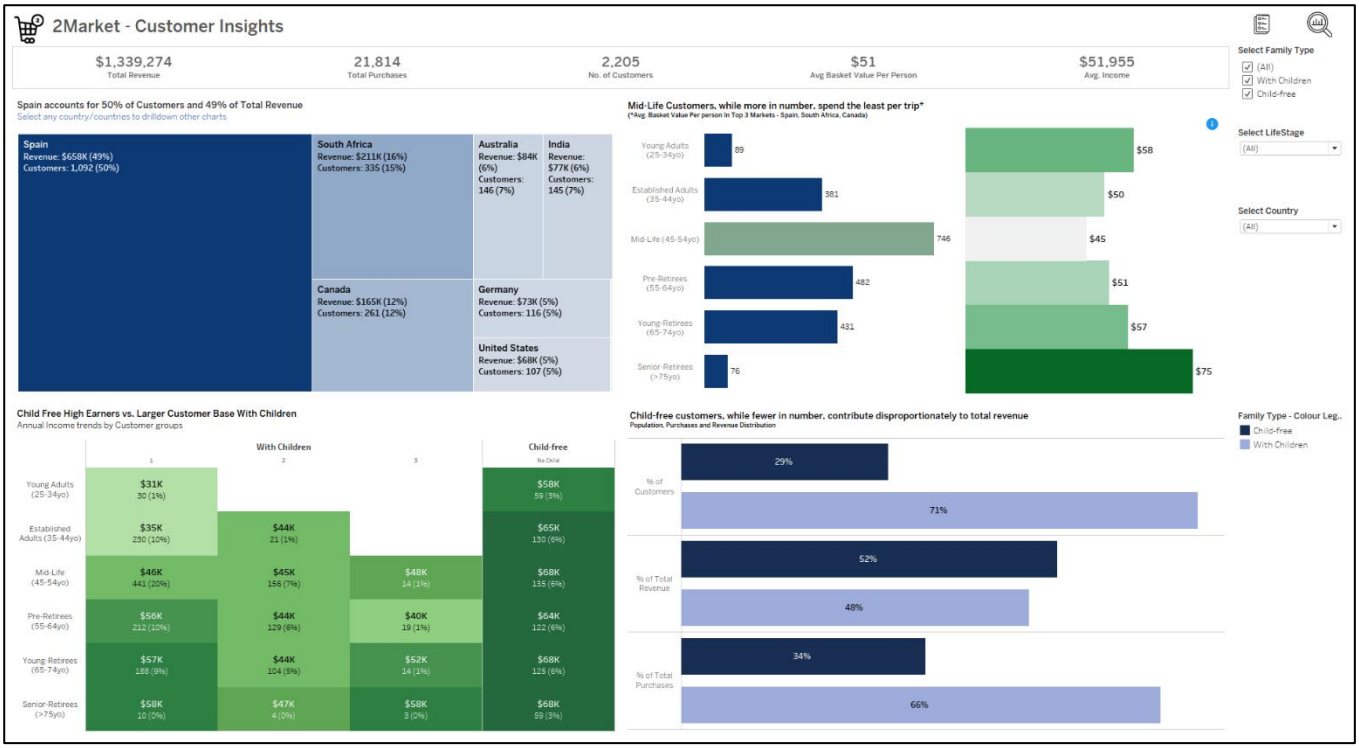
AvgBasketValue

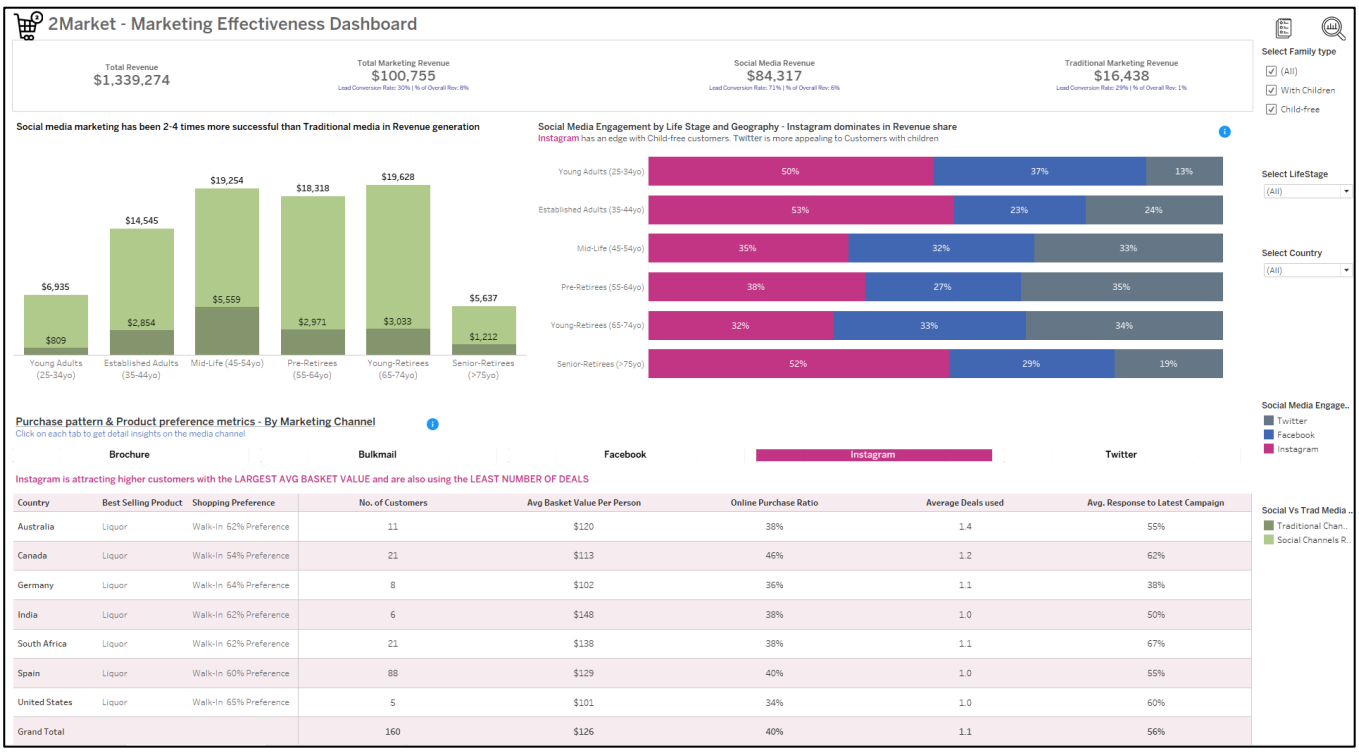
```
[Total Revenue] / [TotalPurchases]
```

Marketing Revenue Share

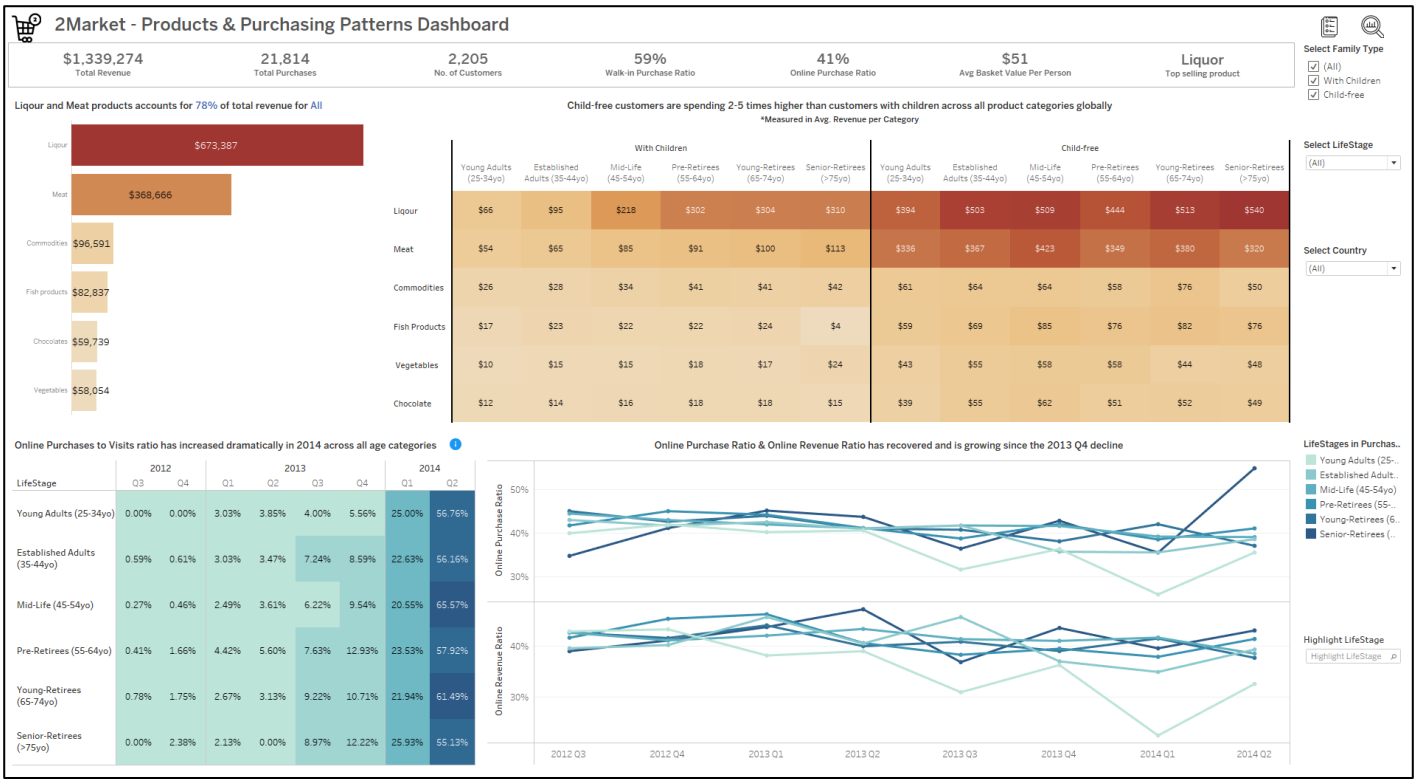
```
[bulkmail_rev]+[fb_rev]+[brochure_rev]+[insta_rev]+[twitter_rev]
```

Dashboard Screenshots:





Country	Best Selling Product	Shopping Preference	No. of Customers	Avg Basket Value Per Person	Online Purchase Ratio	Average Deals used	Avg. Response to Latest Campaign
Australia	Liquor	Walk-In 62% Preference	11	$120	38%	1.4	55%
Canada	Liquor	Walk-In 54% Preference	21	$113	46%	1.2	62%
Germany	Liquor	Walk-In 64% Preference	8	$102	36%	1.1	38%
India	Liquor	Walk-In 62% Preference	6	$148	38%	1.0	50%
South Africa	Liquor	Walk-In 62% Preference	21	$138	38%	1.1	67%
Spain	Liquor	Walk-In 60% Preference	88	$129	40%	1.0	55%
United States	Liquor	Walk-In 65% Preference	5	$101	34%	1.0	60%
Grand Total			160	$126	40%	1.1	56%



5. Data Insights derived:

Key Insights derived:

- Spain, South Africa, Canada are the top 3 markets.
- Revenue and customer count show strong positive correlation.
- The MidLife segment (45-54 years) constitutes the largest customer group across all countries and they have the lowest Average Basket Value Per Person in top 3 markets (Spain, South Africa, Canada).
- Child-free customers are earning higher income than those with children across all age groups. They also generate higher revenue - $695,384 with only 630 customers against $643,890 with 1,575 customers.
- Marketing efforts can be attributed for 8% of overall revenue from Q3 2012 to Q2 2014.
- Social media channels (Facebook, Twitter, Instagram) were 4x more effective.
- Brochure: Least effective, attracting only 30 customers.
- Bulk mail: More effective than brochures but attracts lower-value customers (average basket value: $61, average deals used: 2.2).
- Social media channel performance:
    - Instagram: Most popular among ages 25-44 and 75+. Highest average basket value ($126) and lowest deal usage (1.1).
    - Twitter: Most customers, but lowest average basket value ($88). Highest deal usage (2.4).
    - Facebook: Moderate performance (average deal usage: 1.4).
- Recent campaign success rates:
    - Instagram: 56%, Facebook: 55%, Twitter: 38%
- Liquor and Meat products dominate sales, accounting for 75% of overall revenue across all countries and age categories. Their sales show strong correlation.
- Child-free customers consistently outspend those with children by 2-5x across all product categories and life stages.
- Walk-in shopping is preferred globally.
- Monthly Online Purchase to Web visits ratio calculated has improved dramatically since Q1 2014. Online Purchase Ratio and Online Revenue Ratio as well have improved in Q1 2014.

6. Technical recommendation and Conclusion:

The analysis provided valuable insights into customer behaviour, which will help 2Market optimize marketing strategies and improve sales across key demographics.

The analysis can be enhanced to gain deeper insights by incorporating more granular data—such as product subcategories, invoice-level details of customer purchases, and marketing data at the campaign level, including views, clicks, and partial or full customer journeys.

Appendix – List of calculated fields and definitions for the tableau dashboard:

#	Calculated Field Name	Logic Used	Purpose / Reasoning for creation
Demography Metrics - 3			
1	FamilySize	`[Kidhome]+[Teenhome]`	Gets a single value of size of the family, reducing complexity
2	LifeStage	`IF [Age]<25 THEN "Youth" ELSEIF [Age]<=34 THEN "Young Adults" ELSEIF [Age]<=44 THEN "Established Adults" ELSEIF [Age]<=54 THEN "Mid-Life" ELSEIF [Age]<=64 THEN "Pre-Retirees" ELSEIF [Age]<=74 THEN "Young-Retirees" ELSEIF [Age]<=90 THEN "Senior-Retirees" END`	Customized Age buckets to reduce the number of buckets and also to avoid ambiguity in understanding the age range in charts
3	Set - With Kids	`Yes for : LifeStage = ALL and FamilySize > 0` `No for : LifeStage = ALL and FamilySize = 0`	To classify customers in charts as With Kids / Without Kids
Product Metrics - 2			
4	Total Revenue (New Column during Excel Cleanup)	`=AmtLiq+AmtNonVeg+AmtVege+AmtPes+AmtChoc+AmtComm`	New column added in Excel with all products summed at row level
5	Best Selling Product	`IF SUM([Amt Liq]) >= SUM([Amt Vege]) AND SUM([Amt Liq]) >= SUM([Amt Non Veg]) AND SUM([Amt Liq]) >= SUM([Amt Pes]) AND SUM([Amt Liq]) >= SUM([Amt Chocolates]) AND SUM([Amt Liq]) >= SUM([Amt Comm]) THEN 'Liquor' ELSEIF [VegeCheck] THEN 'Vegetables' ELSEIF [MeatCheck] THEN 'Meat' ELSEIF [PesCheck] THEN 'Fish' ELSEIF [ChocolatesCheck] THEN 'Chocolates' ELSEIF [CommoditiesCheck] THEN 'Commodities' END`	Nested IF statements to identify the best selling product within a context. Used in Products Dashboard and Marketing Dashboard to highlight product preference for any specific customer demography
Ad-Marketing Conversion Metrics – 8			
6	Lead Conversion Rate	`sum([Count success])/COUNT([ID])`	Overall conversion rate of all Marketing channels to get a metric on success ratio of marketing. ***Assumption made that all customers will receive some form of marketing.***
7	FB Conversion Rate	`sum([Facebook ad])/sum([Count success])`	Conversion Rate for Facebook within Successful Media conversions
8	Insta Conversion Rate	`sum([Instagram ad])/sum([Count success])`	Conversion Rate for Instagram within Successful Media conversions

9	Twitter Conversion Rate	`sum([Twitter ad])/sum([Count success])`	Conversion Rate for Twitter within Successful Media conversions
10	Bulkmail Conversion Rate	`sum([Brochure ad])/sum([Count success])`	Conversion Rate for Brochure within Successful Media conversions
11	Brochure Conversion Rate	`sum([Bulkmail ad])/sum([Count success])`	Conversion Rate for Bulkmail within Successful Media conversions
12	Traditional Conversion Rate	`[Bulkmail Conversion Rate]+[Brochure Conversion Rate]`	Conversion rate of Traditional media channels within Successful Media conversions.
13	Social Conversion Rate	`[FB Conversion Rate]+[Twitter Conversion Rate]+[Insta Conversion Rate]`	Conversion rate of Social media channels within Successful Media conversions.
Ad-Marketing Revenue Metrics - 15			
14	Brochure Revenue (brochure_rev)	`IF (([Count success]>0) AND ([Num Walkin Pur]>0)) THEN (([Total Revenue]/[Count success])*[Brochure ad])/[Num Walkin Pur] ELSE 0 END`	Logic derived from SQL Query (Assignment Module 5): To measure Media effectiveness using means other than conversion, Revenue and Purchase was considered. If any media conversion was successful (count_success>0), the individual customer's total revenue generated amount will be divided by count_success value first to avoid double counting when multiple channels are successful. This divided amount will then be multiplied by Brochure ad's data. If Brochure was unsuccessful for the customer - will be multiplied by 0 and no revenue amount for the customer, if Brochure was successful, will be multiplied by 1 and the divided revenue amount of the cusotmer will get allocated as Brochure Revenue. This value will then be divided by ***Number of Walkin Purchases*** to obtain a Per Purchase value of the Revenue - to avoid inflating the value (As the Revenue amount is across multiple purchases) and Successful conversion in Brochure is considered as completing purchase instore.
15	Bulkmail Revenue (bulkmail_rev)	`IF (([Count success]>0) AND ([Num Walkin Pur]>0)) THEN (([Total Revenue]/[Count success])*[Bulkmail ad])/[Num Walkin Pur] ELSE 0 END`	Same logic as Brochure Revenue - Both are considered Traditional media and will be associated with Walkin purchases only. Assumption made based on metadata definition of successful conversion of these channels.
16	Facebook Revenue (fb_rev)	`IF (([Count success]>0) AND ([Num Web Buy]>0)) THEN (([Total Revenue]/[Count success])*[Facebook ad])/[Num Web Buy] ELSE 0 END`	Logic derived from SQL Query (Assignment Module 5): To measure Media effectiveness using means other than conversion, Revenue and Purchase was considered. If any media conversion was successful (count_success>0), the individual customer's total revenue generated amount will be divided by count_success value first to avoid double counting when multiple channels are successful. This divided amount will then be multiplied by Facebook ad's data. If Facebook was unsuccessful for the customer - will be multiplied by 0 and no revenue amount for the customer, if Facebook was successful, will be multiplied by 1 and the divided revenue amount of the customer will get allocated as Facebook Revenue. This value will then be divided by Number of Web Purchases to obtain a Per Purchase value of the Revenue - to avoid inflating the value (As the Revenue amount is across multiple purchases) and Successful conversion in FB is considered as FB Ad to Website and completing purchase as defined in the metadata.
17	Instagram Revenue (insta_rev)	`IF (([Count success]>0) AND ([Num Web Buy]>0)) THEN (([Total Revenue]/[Count success])*[Instagram ad])/[Num Web Buy] ELSE 0 END`	Same logic as Facebook Revenue - Considered as Social media and will be associated with Web purchases only. Assumption made based on metadata definition of successful conversion of these channels.
18	Twitter Revenue (twitter_rev)	`IF (([Count success]>0) AND ([Num Web Buy]>0)) THEN (([Total Revenue]/[Count success])*[Twitter ad])/[Num Web Buy] ELSE 0 END`	Same logic as Facebook Revenue - Considered as Social media and will be associated with Web purchases only. Assumption made based on metadata definition of successful conversion of these channels.

19	Marketing Revenue Share	`[bulkmail_rev]+[fb_rev]+[brochure_rev]+[insta_rev]+[twitter_rev]`	To get a dollar revenue value of how much of the total revenue can be attributed to marketing channels / marketing efforts in general.
20	Social Channels Rev	`[fb_rev]+[insta_rev]+[twitter_rev]`	To get a dollar revenue value of how much revenue *Social marketing channels* created
21	Brochure Revenue Ratio	`sum([brochure_rev])/(sum([bulkmail_rev])+sum([brochure_rev]))`	To get the ratio of Brochure's Revenue share against Traditional Marketing Revenue
22	Bulkmail Revenue Ratio	`sum([bulkmail_rev])/(sum([bulkmail_rev])+sum([brochure_rev]))`	To get the ratio of Bulkmail's Revenue share against Traditional Marketing Revenue
23	Facebook Revenue Ratio	`sum([fb_rev])/sum([Social Channels Rev])`	To get the ratio of Facebook's Revenue share against Social media Marketing Revenue
24	Instagram Revenue Ratio	`sum([insta_rev])/sum([Social Channels Rev])`	To get the ratio of Instagram's Revenue share against Social media Marketing Revenue
25	Twitter Revenue Ratio	`sum([twitter_rev])/sum([Social Channels Rev])`	To get the ratio of Twitter's Revenue share against Social media Marketing Revenue
26	Marketing Revenue Ratio	`sum([Marketing Revenue Share])/sum([Total Revenue])`	To get the ratio of Marketing Revenue share against Total Revenue
27	SocialRev to TotalRev Ratio	`sum([fb_rev]+[insta_rev]+[twitter_rev])/sum([Total Revenue])`	To get a ratio of social media revenue to Total revenue to measure social media's impact on the company's success.
28	TraditionalRev To TotalRev Ratio	`sum([brochure_rev]+[bulkmail_rev])/sum([Total Revenue])`	To get a ratio of Traditional media revenue to Total revenue to measure Traditional media's impact on the company's success.
Purchase Metrics - 13			
29	TotalPurchases	`[Num Web Buy] + [Num Walkin Pur]`	To obtain a single value for number of purchases per customer.
30	Average Basket Value	`[Total Revenue] / [TotalPurchases]`	To obtain an individual customer's total revenue per purchase or basket value per purchase event.
31	Window Min Basket Value	`WINDOW_MIN(AVG([AvgBasketValue]))`	To identify the minimum of average basket value in any context. Used in Customer Demographics Dashboard
32	OnlinePurchase Ratio	`sum([Num Web Buy])/sum([TotalPurchases])`	To get a ratio of how much of total purchases were online
33	Walk-in Purchase Ratio	`sum([Num Walkin Pur])/sum([TotalPurchases])`	To get a ratio of how much of total purchases were in-store
34	Estimated Walk-in Revenue	`([Total Revenue] / ([Num Web Buy] + [Num Walkin Pur])) * [Num Walkin Pur]`	To obtain the in-store shopping share of supermarket's total revenue
35	Estimated Web Revenue	`[Total Revenue]-[estimated_walkin_rev]`	To obtain the online shopping share of supermarket's total revenue
36	Walk-in rev percent	`sum([estimated_walkin_rev])/sum([Total Revenue])`	Percentage of Instore revenue to Total Revenue - for comparitive representation
37	Web rev percent	`1-[walkin_rev_percent]`	Percentage of Website Revenue to Total Revenue - for comparitive representation (Used 1-Walkin Rev Percent to avoid rounding issues during Aggergation).
38	MonthsSinceCust	`DATEDIFF('month',[Dt Customer],date("2014-07-01"))`	To get the number of months the customer has been with the supermarket. *Assumption made that the data has been cut as of 07-01-2014* as the data ends on 06/29/2014. ***Will be required to calculate the avg web buy per month which can then be used along with web visit rate (num visits column in the available data that is defined as number of visits per month)***
39	Avg Web Buy Per Month	`ROUND([Num Web Buy]/[MonthsSinceCust])`	To get the average number of online purchases made per month by the customer.
40	Web Buy to Visit Ratio	`AVG([Avg Web Buy Per Month])/AVG([Num Web Buy])`	To get a ratio of how many times customers are going through with an online purchase against how many times they visit the website on a monthly basis.
41	Shopping Preference	`IF [Online Purchase Ratio]>=[Walk-in Purchase Ratio] THEN 'Online'+' '+STR(int([Online Purchase Ratio]*100))+'% Preference' ELSE 'Walk-In'+' '+STR(int([Walk-in Purchase Ratio]*100))+'% Preference' END`	To get a preference of the customer segment for Online or In-store shopping based on the purchase ratios (Online Purchase Ratio is Web Buy / Total Purchases, Walk-In Ratio is Walk-in Purchases / Total Purchases) ***Since the ratios are very close, displayed the ratio as preference in % terms.***