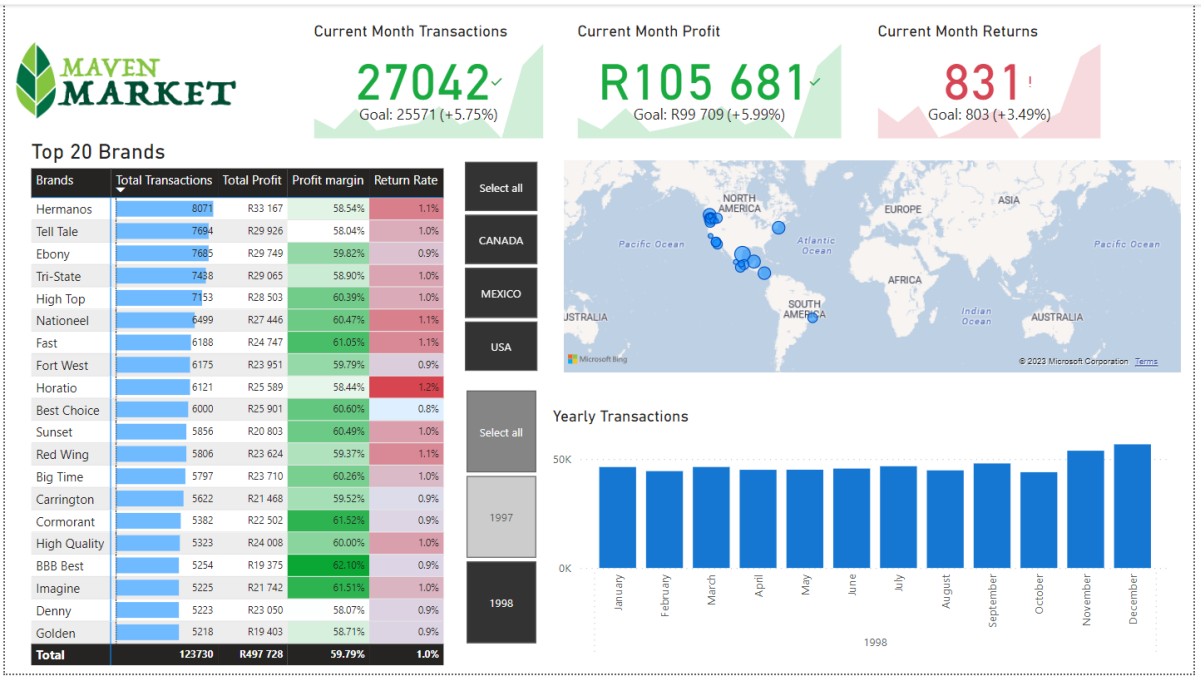
About

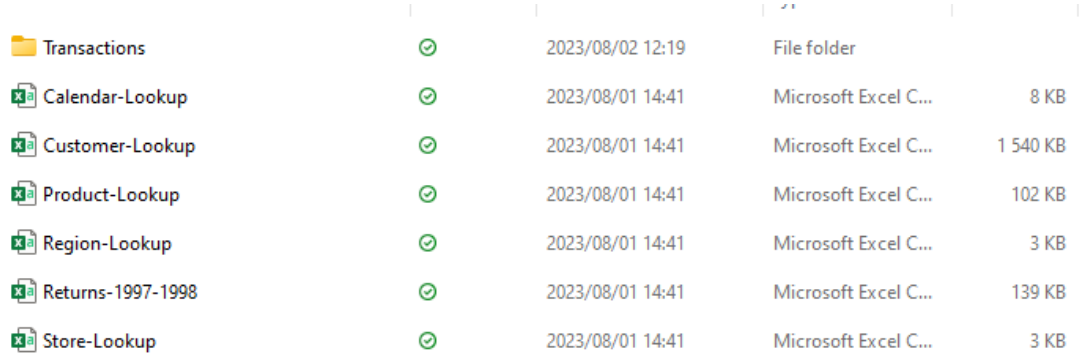
By: Zaahid Allie

This process report serves to provide an overview of the steps undertaken during this Power BI and Power Query project. From data acquisition and transformation to advanced analysis and visualization, each stage is highlighted in a concise manner showcasing the techniques employed to transform raw data into actionable insights. This report serves as a roadmap, offering a clear insight into my adeptness at using Power BI and Power Query to derive meaningful conclusions from complex datasets.



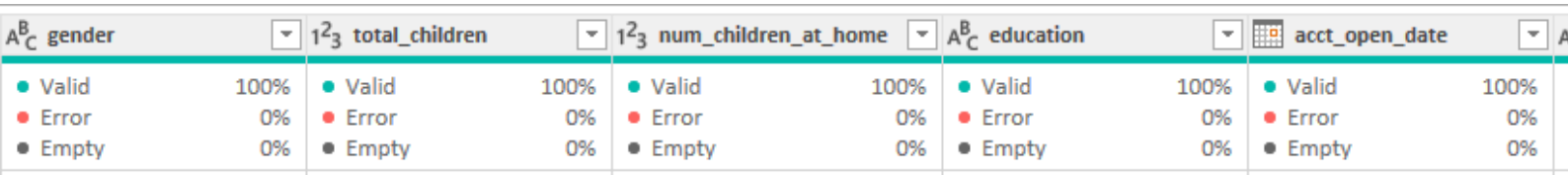
Shaping & Transforming the Data

- The dataset was obtained from MavenAnalytics, in the form of CSV files.
- The first step was to go through the data using Power Query and Power Query Editor to see which of the tables contained useful information for analysis.



Transactions	✓	2023/08/02 12:19	File folder	
Calendar-Lookup	✓	2023/08/01 14:41	Microsoft Excel C...	8 KB
Customer-Lookup	✓	2023/08/01 14:41	Microsoft Excel C...	1 540 KB
Product-Lookup	✓	2023/08/01 14:41	Microsoft Excel C...	102 KB
Region-Lookup	✓	2023/08/01 14:41	Microsoft Excel C...	3 KB
Returns-1997-1998	✓	2023/08/01 14:41	Microsoft Excel C...	139 KB
Store-Lookup	✓	2023/08/01 14:41	Microsoft Excel C...	3 KB

- This involved ensuring that that there were no errors in the tables. Any errors would be appropriately rectified.



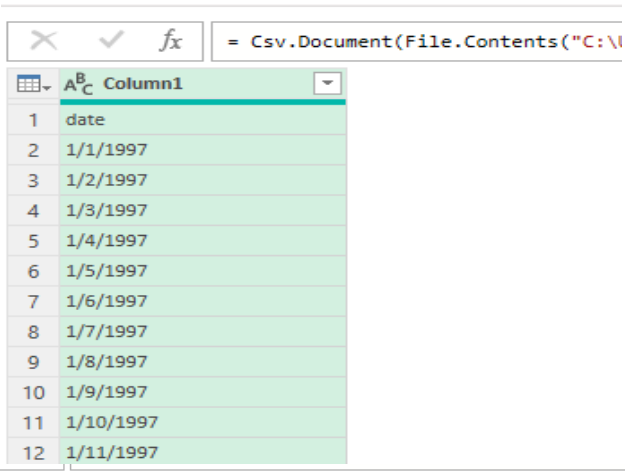
gender		total_children		num_children_at_home		education		acct_open_date	
Valid	100%	Valid	100%	Valid	100%	Valid	100%	Valid	100%
Error	0%	Error	0%	Error	0%	Error	0%	Error	0%
Empty	0%	Empty	0%	Empty	0%	Empty	0%	Empty	0%

- After error checking, the next step was to ensure that the data in columns were of the correct type.
    - In the example case of the image below, the columns “product cost” and “product weight” were being treated as text.
- Rectifying this involved converting them to their appropriate types as a currency and a decimal number.

product_cost	product_weight
0.94	8.39
0.26	7.42
0.4	13.1
1.64	10.6
0.77	6.66
0.37	15.8
0.91	18
0.8	8.97
0.77	7.14

product_cost	product_weight
R0,94	8,39
R0,26	7,42
R0,40	13,1
R1,64	10,6
R0,77	6,66
R0,37	15,8
R0,91	18
R0,80	8,97
R0,77	7,14
R0,50	8,13

- Once the data had been appropriately validated, new columns were pre-emptively added to the tables that would be useful for later analysis.
    - For example, deconstructing dates from the calendar lookup table into their quarters, months, week of the year etc so that a more granular filter could be later applied.



	date	Start of Year	Start of Month	Month Number	Start of Week
1	1997/01/01	1997/01/01	1997/01/01	1	1996/12/29
2	1997/01/02	1997/01/01	1997/01/01	1	1996/12/29
3	1997/01/03	1997/01/01	1997/01/01	1	1996/12/29
4	1997/01/04	1997/01/01	1997/01/01	1	1996/12/29
5	1997/01/05	1997/01/01	1997/01/01	1	1997/01/05
6	1997/01/06	1997/01/01	1997/01/01	1	1997/01/05
7	1997/01/07	1997/01/01	1997/01/01	1	1997/01/05
8	1997/01/08	1997/01/01	1997/01/01	1	1997/01/05
9	1997/01/09	1997/01/01	1997/01/01	1	1997/01/05
10	1997/01/10	1997/01/01	1997/01/01	1	1997/01/05
11	1997/01/11	1997/01/01	1997/01/01	1	1997/01/05

- The process of adding new columns for analysis continued either deconstructing columns for more granular filters or grouping information into smaller more defined groups.
- For example, adding a new column to the customer lookup table, in which I condensed the customers education into 2 groups “graduates” and “non-graduates”.

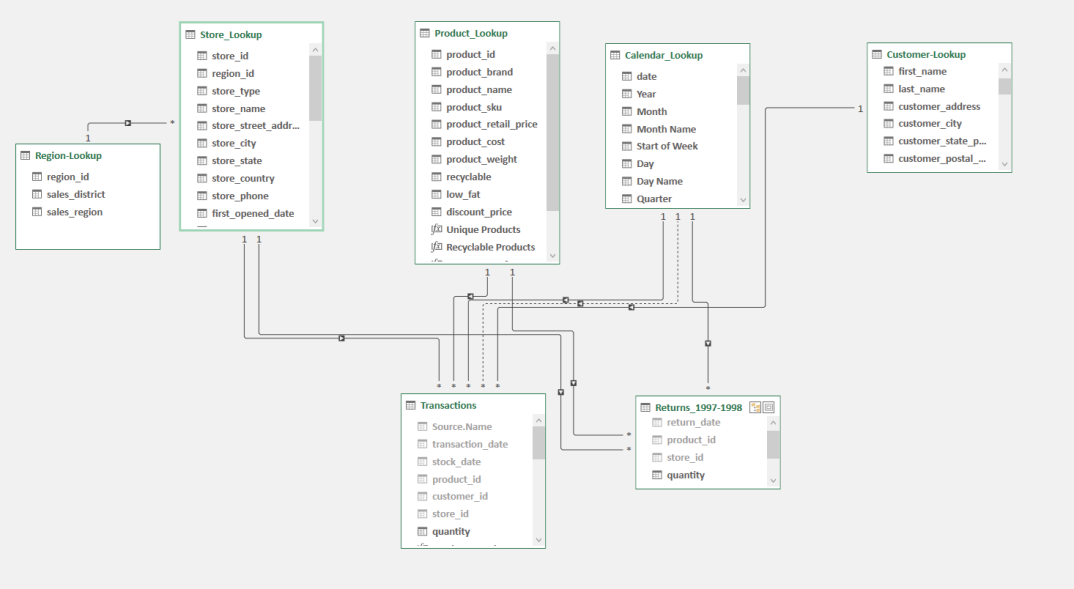
```
=IF(OR('Customer-Lookup'[education]="Bachelors Degree",'Customer-Lookup'[education]="Graduate Degree"),"Graduate","
```

education	education_level
Partial High School	Non-Graduate
Partial High School	Non-Graduate
Bachelors Degree	Graduate
Partial High School	Non-Graduate
Partial College	Non-Graduate
Bachelors Degree	Graduate
Partial High School	Non-Graduate
Bachelors Degree	Graduate
Partial High School	Non-Graduate
Bachelors Degree	Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate
Bachelors Degree	Graduate
Graduate Degree	Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate
Partial College	Non-Graduate
High School Degree	Non-Graduate
Partial High School	Non-Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate
High School Degree	Non-Graduate

- Once all the CSV files had been added to the data model and the various initial columns were added, it was time to create connections between the tables.

Creating the data model

- Utilising Excels data model diagram, connections were made between the tables.
- This creates efficiency in the model, by preventing multiple tables having recurring data, or from using a single table with redundant data.



- Additionally, to prevent incorrectly filtering on the wrong metrics (foreign keys) in Pivot tables or Power Bi, I hid all foreign keys from outside of the model, so that the filter upstream in the lookup tables were the only filters being applied.

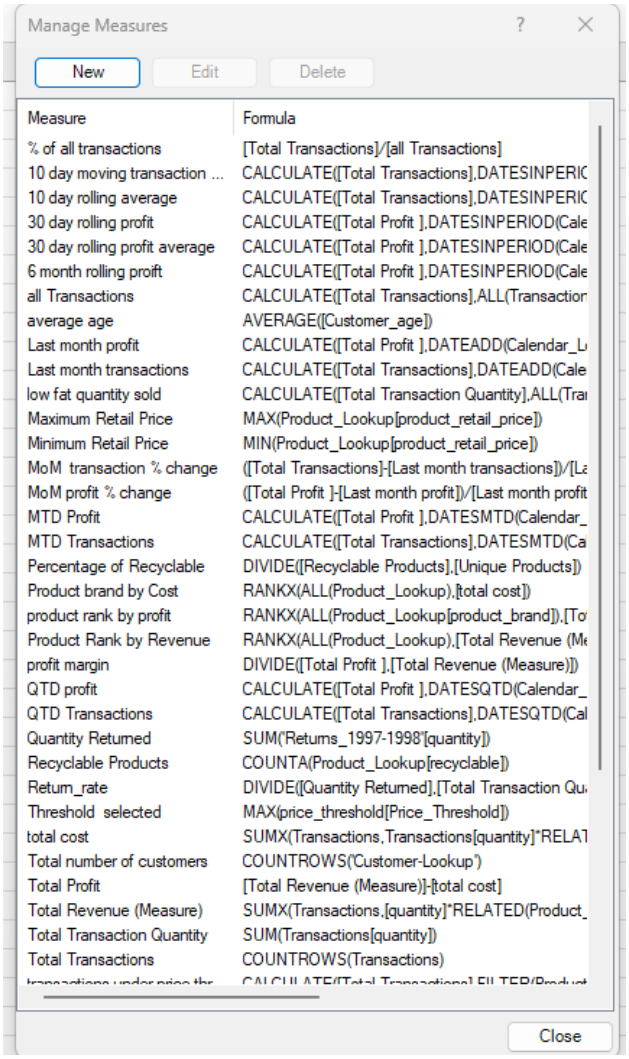
- The filterable options that remained in the data tables were the quantities sold and returned.

- This ensures that the outputs do not display misleading values from incorrectly filtering on the wrong “key”.

Creating Measures for analysis

- By leveraging the data available in the provided tables, I utilised Data Analysis Expressions (DAX), the formula language native to Power Query and Power BI to create an array of measures that efficiently dissected the data, delivering valuable insights.

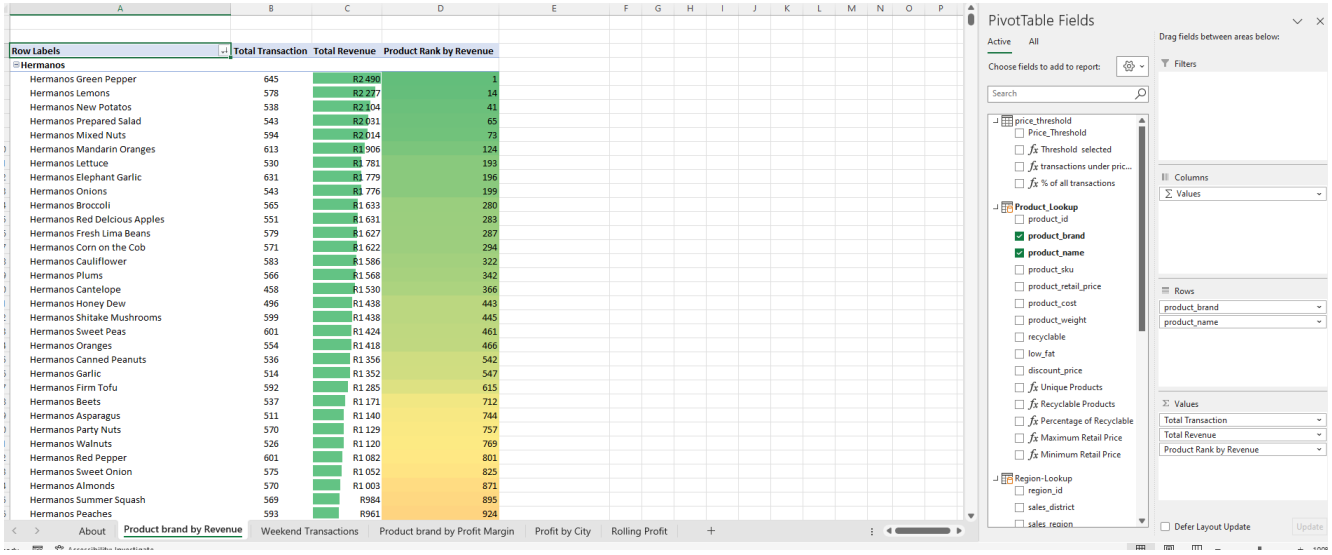
- This entailed creating a range of measures to compute fundamental metrics such as profit, loss, revenue, and total cost, alongside more intricate measurements like breaking down the percentage of recyclable products or the proportion of transactions occurring on weekends or the return rate of specific products etc.



Analysing the Data with Power Pivot

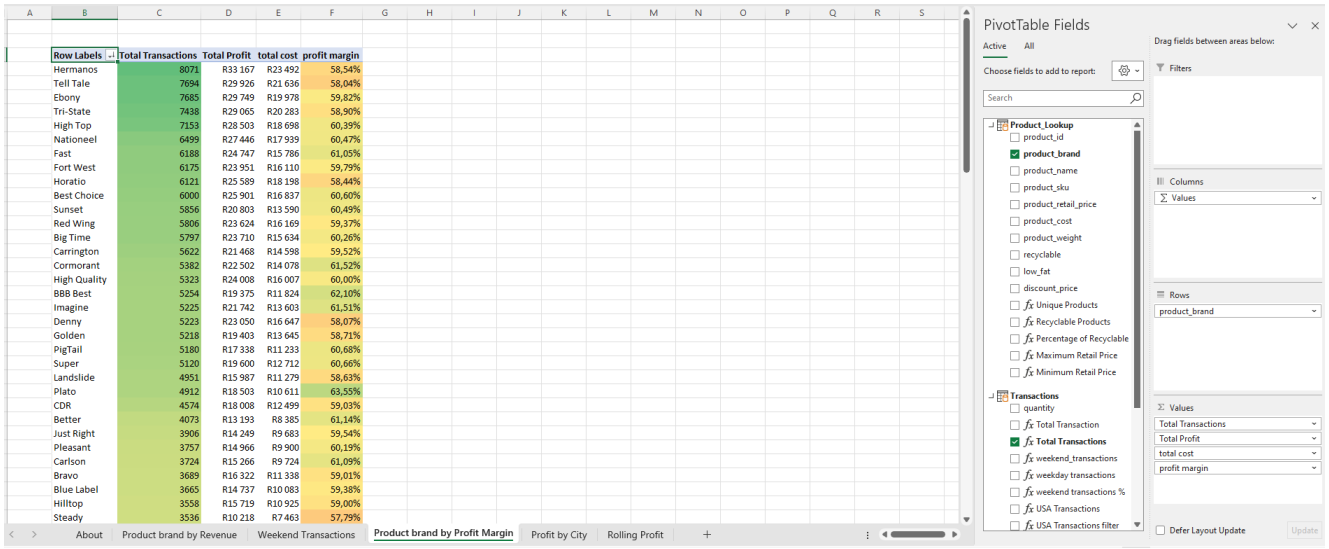
- After establishing the various measures using DAX, Power Pivot was utilised to better understand the data and start delivering useable insights.

Insights into the revenue generated from most transacted items.



Row Labels	Total Transaction	Total Revenue	Product Rank by Revenue
Hermanos			
Hermanos Green Pepper	645	R2 490	1
Hermanos Lemons	578	R2 227	14
Hermanos New Potatos	538	R2 104	41
Hermanos Prepared Salad	543	R2 031	65
Hermanos Mixed Nuts	594	R2 014	73
Hermanos Mandarin Oranges	613	R1 906	124
Hermanos Lettuce	530	R1 781	193
Hermanos Elephant Garlic	631	R1 779	196
Hermanos Onions	543	R1 776	199
Hermanos Broccoli	565	R1 633	280
Hermanos Red Delcious Apples	551	R1 631	283
Hermanos Fresh Lima Beans	579	R1 627	287
Hermanos Corn on the Cob	571	R1 622	294
Hermanos Cauliflower	583	R1 586	322
Hermanos Plums	566	R1 568	342
Hermanos Cantelope	458	R1 530	366
Hermanos Honey Dew	496	R1 438	443
Hermanos Shitake Mushrooms	599	R1 438	445
Hermanos Sweet Peas	601	R1 424	461
Hermanos Oranges	554	R1 418	466
Hermanos Canned Peanuts	536	R1 356	542
Hermanos Garlic	514	R1 352	547
Hermanos Firm Tofu	592	R1 285	615
Hermanos Beets	537	R1 171	712
Hermanos Asparagus	511	R1 140	744
Hermanos Party Nuts	570	R1 129	757
Hermanos Walnuts	526	R1 120	769
Hermanos Red Pepper	601	R1 082	801
Hermanos Sweet Onion	575	R1 052	825
Hermanos Almonds	570	R1 003	871
Hermanos Summer Squash	569	R984	895
Hermanos Peaches	593	R961	924

Understanding the profit margins from items sold

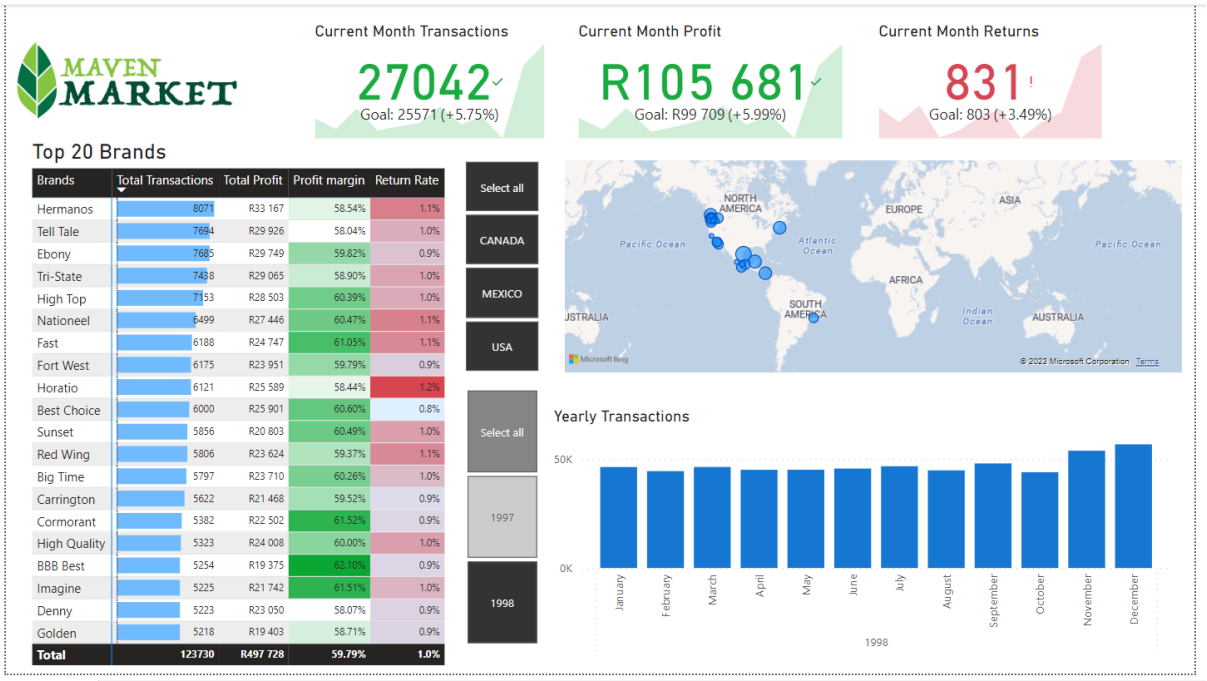


Row Labels	Total Transactions	Total Profit	total cost	profit margin
Hermanos	8071	R33 167	R23 492	58,54%
Tell Tale	7694	R29 926	R21 636	58,04%
Ebony	7685	R29 749	R19 978	59,82%
Tri-State	7438	R29 065	R20 283	58,90%
High Top	7153	R28 503	R18 698	60,39%
Nationeel	6499	R27 446	R17 939	60,47%
Fast	6188	R24 747	R15 786	61,05%
Fort West	6175	R23 951	R16 110	59,79%
Horatio	6121	R25 589	R18 198	58,44%
Best Choice	6000	R25 901	R16 837	60,60%
Sunset	5856	R20 803	R13 590	60,49%
Red Wing	5806	R23 624	R16 169	59,37%
Big Time	5797	R23 710	R15 634	60,26%
Carrington	5622	R21 468	R14 598	59,52%
Cormorant	5382	R22 502	R14 078	61,52%
High Quality	5323	R24 008	R16 007	60,00%
BBB Best	5254	R19 375	R11 824	62,10%
Imagine	5225	R21 742	R13 603	61,51%
Denny	5223	R23 050	R16 647	58,07%
Golden	5218	R19 403	R13 645	58,71%
PigTail	5180	R17 338	R11 233	60,68%
Super	5120	R19 600	R12 712	60,66%
Landslide	4951	R15 987	R11 279	58,63%
Plato	4912	R18 503	R10 611	63,55%
CDR	4574	R18 008	R12 499	59,03%
Better	4073	R13 193	R8 385	61,14%
Just Right	3906	R14 249	R9 683	59,54%
Pleasant	3757	R14 966	R9 900	60,19%
Carlson	3724	R15 266	R9 724	61,09%
Bravo	3689	R16 322	R11 338	59,01%
Blue Label	3665	R14 737	R10 083	59,38%
Hilltop	3558	R15 719	R10 925	59,00%
Steady	3536	R10 218	R7 463	57,79%

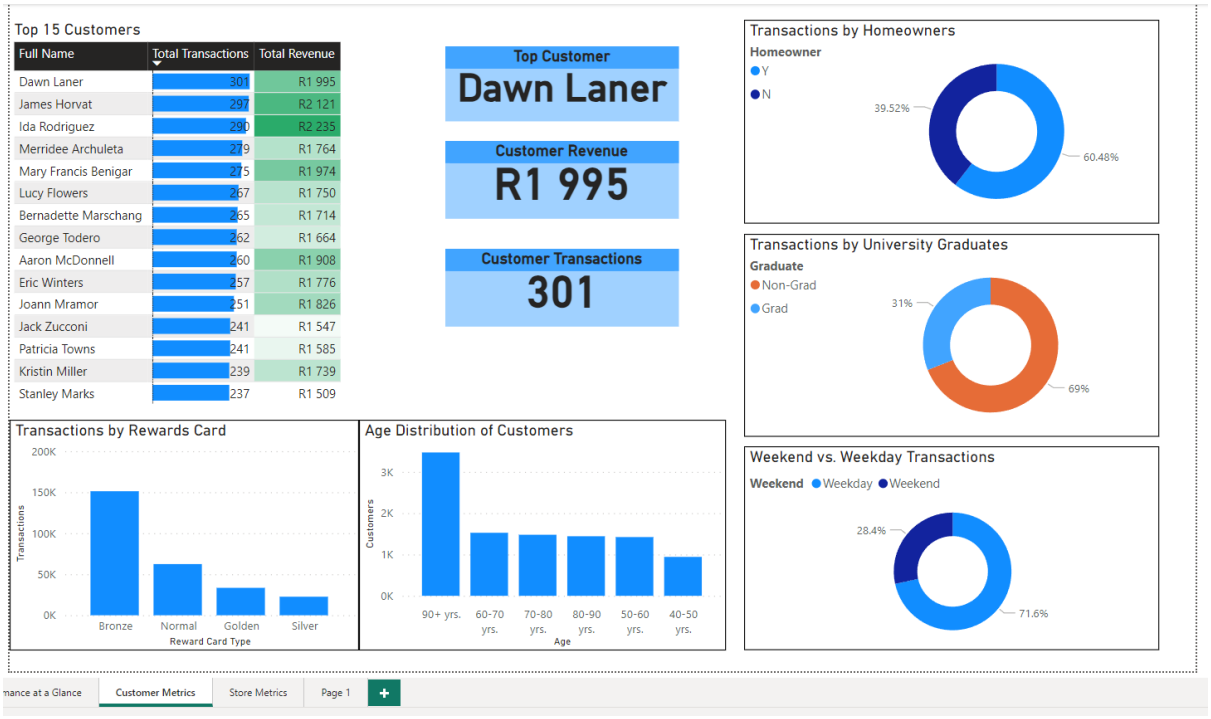
Visualising the Data

- While Excel natively supports and provides a comprehensive suite of tools to visualise data, I used this opportunity to showcase my ability to use Power Bi.
- Utilising the model and measures created in Excel, it was imported into Power Bi.
- 3 distinct and interactive pages (pictured below) were created:
    - Overview of the business for the past month
    - Customer Metrics over the last year
    - Store Metrics over the last year

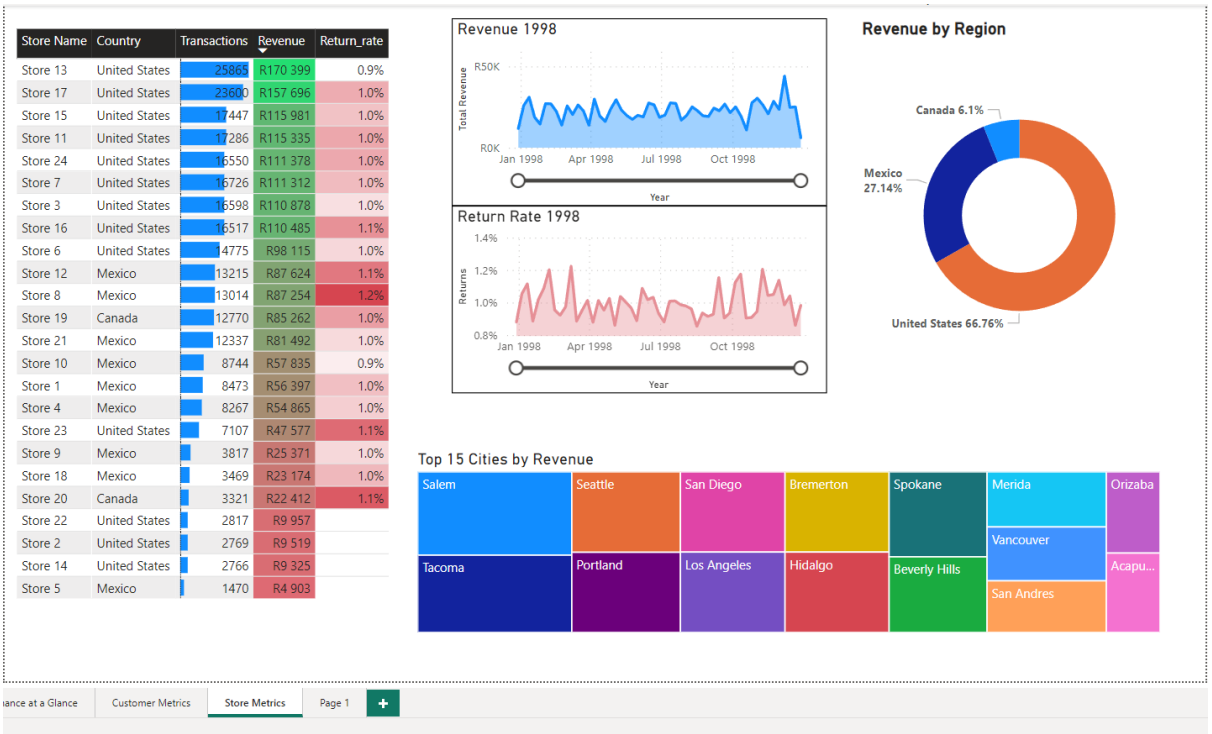
Overview of business



Customer Metrics



Store Metrics



Conclusions

Thank you for reviewing this process report. I am confident that it showcases my adeptness in harnessing Microsoft's array of business intelligence tools. Furthermore, it underscores my capacity to extract valuable analysis and insights from data. Another project of mine, titled "CEO Report," serves as a testament to my proficiency in researching, visualizing data, and composing comprehensive reports on subjects that are new to me.

To view that report as well as other projects of mine they can be found at:

https://github.com/ItsZeed/Data-Analyst-Portfolio