

Quantitative Data Analysis Report

Customer Distance, Weekly Spend and Behavioral Evidence in a Neighbourhood Shop

Student name: Guy Kendal Richards

Student number: 2541860

Module: BMPF7330D

Word count: 3257

Contents

1. Introduction
2. Data, Method & Measurement Levels
3. Descriptive Analysis & Interpretation
 - 3.1 Overall customer profile
 - 3.2 Distance and weekly spend
 - 3.2.1 Robustness of the distance–spend relationship
 - 3.3 Gender and spending behaviour
 - 3.4 Newspaper preference and customer segmentation
 - 3.5 Opinion of the local borough council
 - 3.6 Alternative relationships and analytical limitations
4. Why Dataset Analysis Matters
5. Missing Variables & Improved Data Collection
6. Practical Recommendations for Management
7. Conclusion
8. Appendix A. Provided Dataset

1. Introduction

This report analyses a quantitative dataset relating to 20 customers of a local neighbourhood shop or newsagent. The dataset records distance from the shop in meters, average weekly spend in UK pounds, gender, regular newspaper purchased and opinion of the local borough council. It therefore provides a concise view of who uses the shop, how much customers spend and whether spending differs according to location or customer characteristics. The initial pattern is that weekly spend generally falls as distance from the shop increases. Smaller differences also appear across gender and newspaper groups, while council opinion shows a much weaker relationship with spending. The purpose of the report is to interpret these patterns for management rather than simply present calculations.

The data was most likely produced through a structured, cross-sectional customer questionnaire or short interview, possibly supported by observation or shop records. Data appears to have been collected through a structured customer survey or short observational exercise. A structured method would explain why every customer has the same set of coded responses while the inclusion of average weekly spend and council opinion suggests that at least some information may have been self-reported. Distance may have been estimated from a home address or postcode although the method is not stated. Because each customer appears only once, the dataset provides a snapshot at one point in time rather than evidence of change over time. The sampling method is also unknown: if customers were selected because they were available in the shop, the sample may be convenient rather than representative of the full customer base.

The principal finding is a strong decline in weekly spend as distance increases indicating that the shop depends heavily on its immediate catchment area. However, the dataset does not record travel mode, carrying capacity, weather, competitors or basket contents or weekly spend was recalled or taken from till records. These omissions limit the conclusions that can be drawn and mean that the analysis can identify associations but cannot establish why they occur.

To give the analysis a clear direction, the report addresses three related questions. First, how strongly is customer distance associated with average weekly spend? Second, do gender, newspaper preference or council opinion provide credible alternative explanations for differences in spending? Third, what additional evidence would management need before acting on the results? These questions connect the statistical work to practical decisions and prevent the analysis from becoming a list of disconnected calculations. They also ensure that each finding is considered in terms of its usefulness, its limitations and the strength of the evidence supporting it.

2. Data, Method & Measurement Levels

The dataset contains scale, nominal and ordinal data. Distance and average weekly spend are scale variables because they are numeric, measured in defined units and have meaningful differences between values. Distance is measured in meters and weekly spend in pounds sterling. Gender and newspapers purchased are nominal variables because they place customers into categories without a natural order. Council opinion is ordinal because it is ranked from 1 to 4, where 1 means very discontent and 4 means very content. Correctly identifying measurement level is important because it determines which summaries are appropriate. Means, medians, standard deviations and correlations can be used for the scale variable, while counts and percentages are more suitable for nominal categories. Ordinal responses can be summarised using frequencies, medians and cautions comparisons of average scores, provided the ranking is not treated as a perfectly equal scale.

The spreadsheet was prepared by removing the descriptive notes and retaining the 20 customer records. Distance, spend and council opinion were treated as numeric variables while gender and newspaper were treated as categories. Four distance bands were created: 0-50 meters, 51-200 meters, 201-500 meters and over 500 meters. This makes the results easier to interpret for a management because a shop manager may find distance bands more meaningful than 20 separate distances. Clear coding, consistent units and accurate labels are important because errors introduced during preparation can affect every later calculation and conclusion. Data preparation is not a minor technical step; it affects the quality of the conclusion.

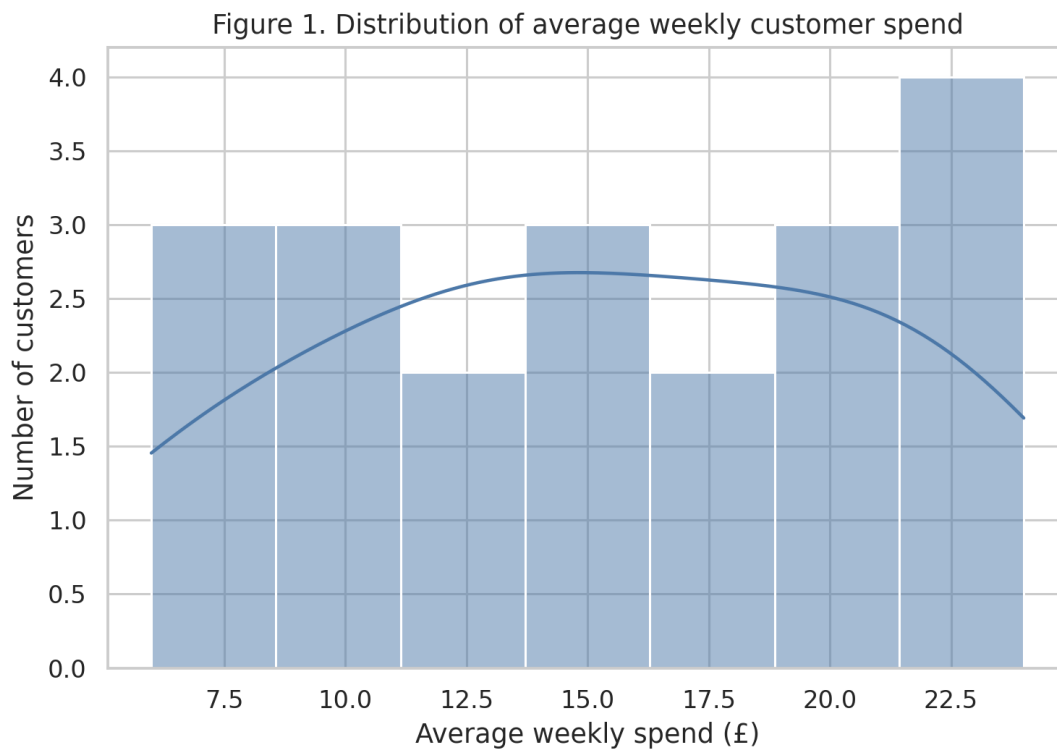
Table 1. Descriptive Statistics for Key Variables

Index	Distance M	Spend £	Council Opinion Score
count	20.0	20.0	20.0
mean	259.5	15.25	2.4
median	160.0	15.0	2.5
min	5.0	6.0	1.0
max	1100.0	24.0	4.0
std	323.43	5.83	1.14

3. Descriptive Analysis & Interpretation

3.1 Overall Customer Profile

The sample contains 20 customers. The average distance from the shop is 259.50 meters while the median is 160 meters. The mean is higher than the median because a few customers live much farther away including customers at 700, 1,000 and 1,100 meters. This shows that the shop's customer base is concentrated locally but includes a small number of distant customers. Average weekly spend is £15.25 and the median is £15.00 suggesting that spending is less skewed than distance. The lowest weekly spend is £6 and the highest is £24 providing a useful range for comparing lower and higher spending customers.



Distribution of average weekly customer spend

Figure 1 shows that spending is clustered around the middle and upper part of the observed range with fewer customers at the lowest level. This is useful because it suggests that many customers have some regular weekly value to the shop. However, the chart does not explain why one customer spends £24 and another spends £6. For that explanation, the dataset would need more variables about the purpose of the visit, frequency of visits, basket contents, transport and customer circumstances. The distribution is therefore a starting point for interpretation rather than a complete explanation.

3.2 Distance and Weekly Spend

The strongest relationship in the dataset is between distance and average weekly spend. The Pearson correlation between distance and spend is -0.875, which indicates a very strong negative association. In practical terms, customers who live farther from the shop tend to spend less each week. A simple trend line suggests that each additional meter from the shop is associated with about £0.016 less weekly spend, which is approximately £1.58 less for each additional 100 meters. This is important for management, but it remains an association within a small cross-sectional sample and should not be presented as proof that distance directly causes lower spending.

3.2.1 Robustness of the Distance–Spend Relationship

The main result was tested in several ways to determine whether it was stable rather than being created by a particular calculation or a small number of unusual customers. The original Pearson correlation is -0.875 ($p < .001$), and distance alone explains approximately 76.6% of the observed variation in weekly spend. The Spearman rank correlation is even stronger at -0.997 ($p < .001$). Spearman correlation considers the order of the values rather than assuming the relationship must follow a perfectly straight line. The result means that customers can be ranked with near consistency: as distance increases, weekly spend almost always decreases.

The analysis was also repeated after removing the three customers living beyond 500 meters. Among the remaining 17 customers, the correlation remains very strong at -0.950 ($p < .001$). The relationship is therefore not created solely by the observations at 700, 1,000 and 1,100 meters. A small, adjusted model including distance, gender and council opinion explains 78.3% of the variation in spending, with distance remaining statistically significant ($p < .001$). These checks increase confidence that the observed relationship is stable within this sample. Nevertheless, 20 cases are insufficient for dependable forecasting or for a complex model with several predictors, so the result should be tested using a larger and more representative sample.

Table 2. Robustness Checks for the Distance–Spend Relationship

Robustness check	Result	Meaning
Original Pearson correlation	$r = -0.875, p < .001$	Very strong negative linear association
Spearman rank correlation	$\rho = -0.997, p < .001$	Spending falls almost consistently as distance ranks increase
Customers within 500 meters	$r = -0.950, p < .001; n = 17$	Finding remains after removing the three most distant customers
Distance-only linear model	$R^2 = 0.766$	Distance explains 76.6% of observed spending variation
Adjusted model	$R^2 = 0.783; \text{distance } p < .001$	Distance remains important after including gender and opinion

Figure 2. Distance and spend by customer segment

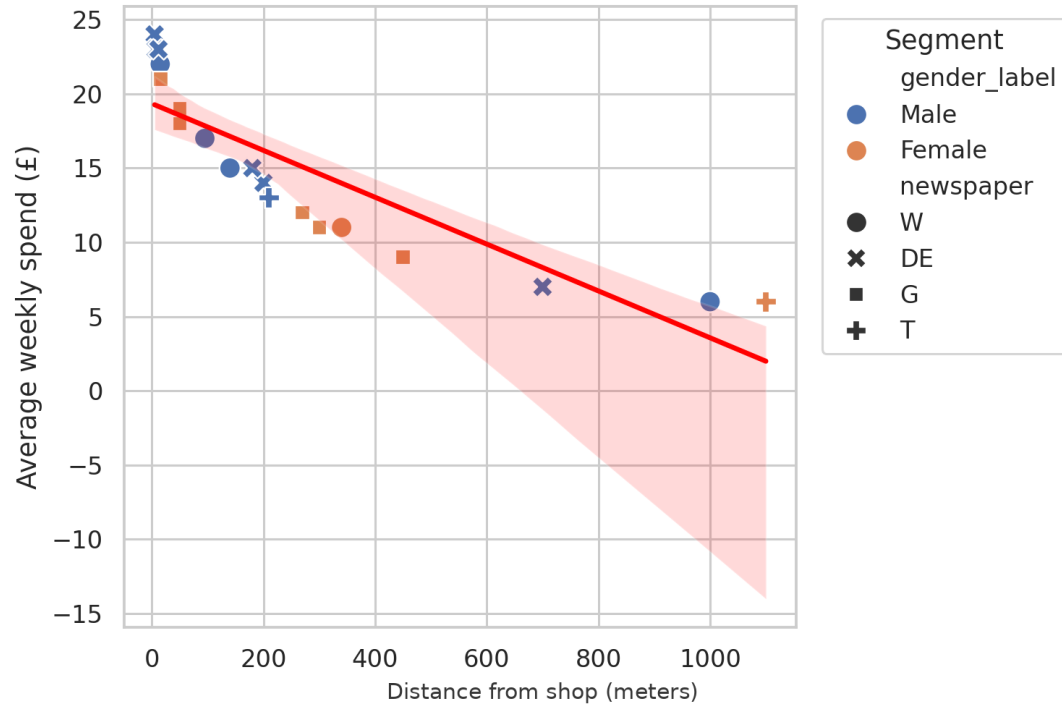


Figure 3. Average weekly spend by distance band

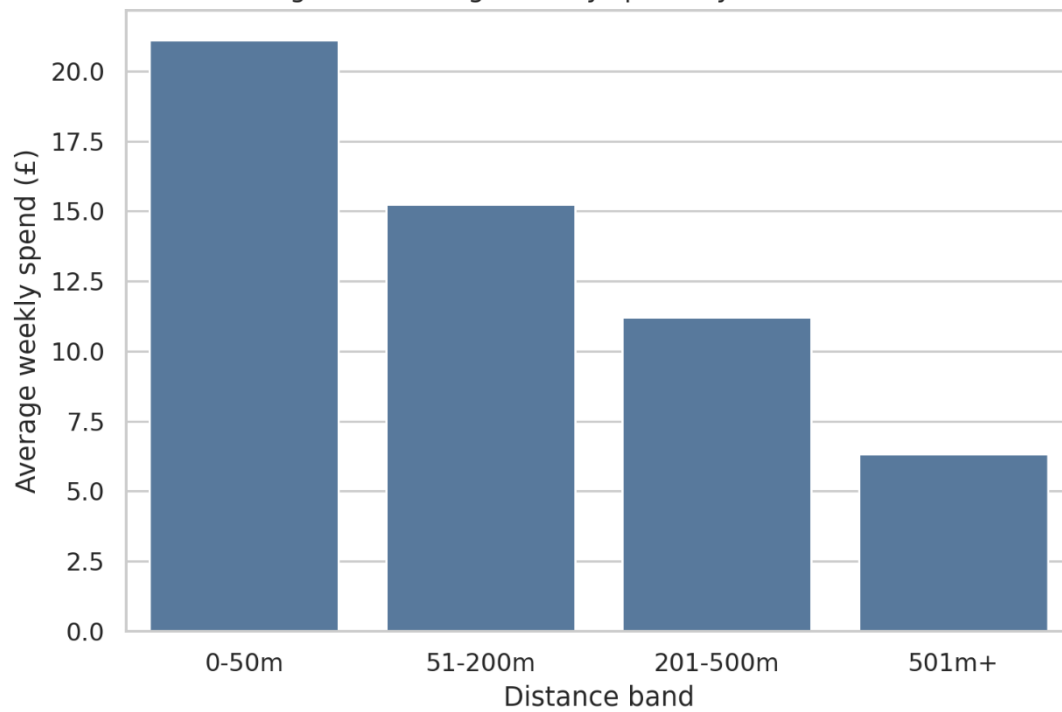


Table 3. Spend and Opinion by Distance Band

Distance Band	Customers	Average Spend £	Median Spend £	Average Opinion Score
0-50m	8	21.12	21.5	2.0
51-200m	4	15.25	15.0	2.5
201-500m	5	11.2	11.0	2.8
500m+	3	6.33	6.0	2.67

The distance-band analysis makes the pattern especially clear. Customers within 50 meters spend an average of £21.12 per week. This falls to £15.25 for customers 51-200 meters away, £11.20 for customers 201-500 meters away, and £6.33 for customers more than 500 meters away. The steady decline suggests that convenience and proximity are likely to be important. For management, this result indicates that revenue may depend heavily on nearby residents who use the shop as part of their regular routines. Protecting this local customer base through reliability, availability and good service should therefore be a priority.

Distance may represent more than geography. It can reflect journey time, physical effort, travel cost, weather exposure, carrying capacity or access to competing shops. For example, a driver and a pedestrian travelling 700 meters face different constraints. Distance is therefore a strong indicator of spending in this sample but not a complete explanation of customer behaviour.

3.3 Gender & Spending Behaviour

Table 4. Customer Profile by Gender

Gender	Customers	Average Distance M	Average Spend £	Average Opinion Score
Female	9	291.78	14.0	2.33
Male	11	233.09	16.27	2.45

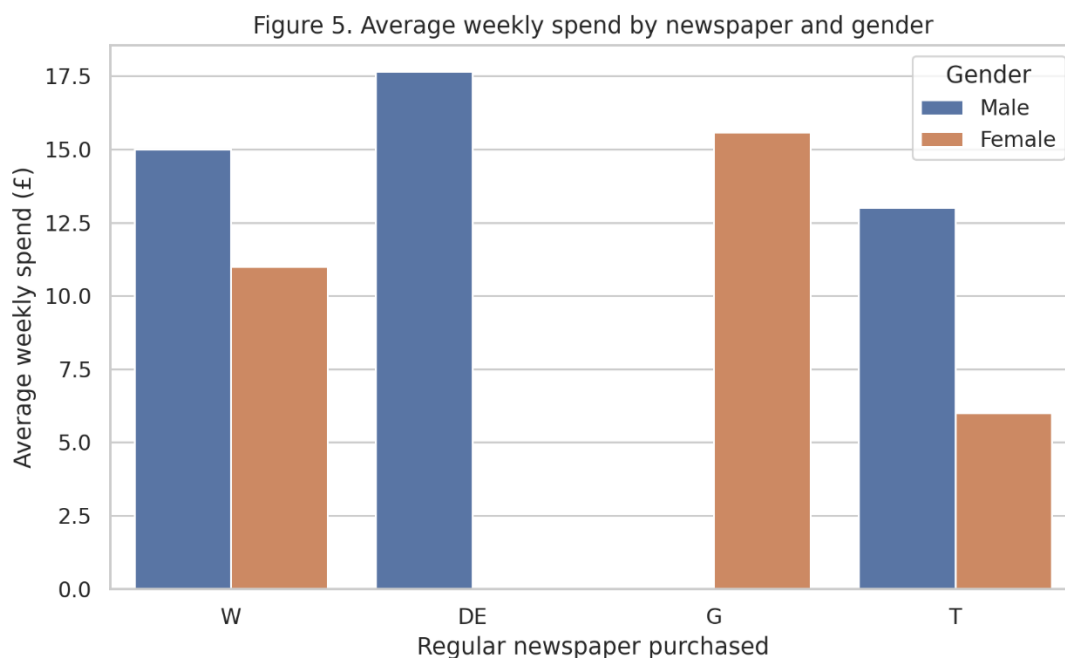
The sample includes 11 male and 9 female customers. Male customers spend £16.27 per week on average while female customers spend £14.00. This appears to suggest that male customers spend more. However, male customers also live closer to the shop on average at 233.09 meters compared with 291.78 meters for female customers. The distance in spending may therefore reflect distance, at least partly, rather than gender itself. A careful analysis should not treat the difference between two small group averages as evidence of a direct gender effect.

With a larger sample, the shop could compare customers of different genders at similar distances and could also consider travel mode, visit frequency and basket type. In this dataset, gender provides useful descriptive context but does not offer a reliable independent explanation of spending differences.

3.4 Newspaper Preference and Customer Segmentation

Table 5. Customer profile by newspaper

Newspaper	Customers	Average Distance M	Average Spend £	Average Opinion Score
DE	6	184.0	17.67	2.17
G	7	169.43	15.57	1.86
T	2	655.0	9.5	4.0
W	5	318.0	14.2	2.8



Newspaper preference may help the shop understand customer routines. In the sample above, G is purchased by 7 customers, DE by 6, W by 5 and T by 2. Customers purchasing DE have the highest average spend at £17.67, followed by G at £15.57, W at £14.20 and T at £9.50. However, the T group contains only two customers, who live much farther away on average at 655 meters. It would therefore be weak to conclude that the newspaper itself explains their lower spending. A better interpretation is that newspaper preference may identify customer segments but the sample is too small to separate newspaper effects from distance and other characteristics.

All DE customers in the sample are male and all G customers are female although this may be a sampling effect rather than a stable local pattern. The more interesting and useful question for management is whether regular newspaper purchases generate additional sales. Basket-level data could show whether newspapers drive purchases of groceries, drinks or snacks or mainly represent single-item transactions.

3.5 Opinion of the Local Borough Council

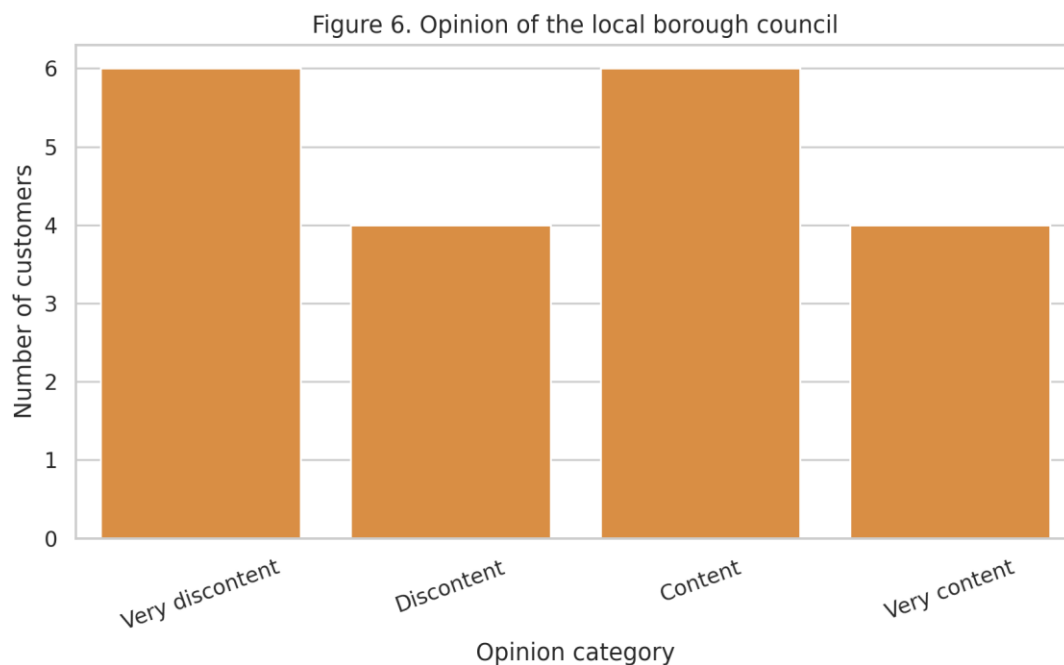


Table 6. Opinion of the Local Borough Council

Opinion Category	Customers
Very discontent	6
Discontent $p < .001$	4
Content	6
Very content	4

Council opinion is included as an ordinal variable and is spread across all four categories. Six customers are very discontent, four are discontent, six are content and four are very content. The relationship between council opinion and spend is weak, with a Pearson correlation of -0.269. This suggests that council opinion is not a major direct predictor of weekly spend in this dataset. However, the variable may still be useful if future research asks more specific questions about local conditions. *Parking restrictions, roadworks, cycling facilities, street lighting, pavements and public safety may all influence whether customers choose this shop.*

Figure 4. Correlation heatmap for scale and ordinal variables

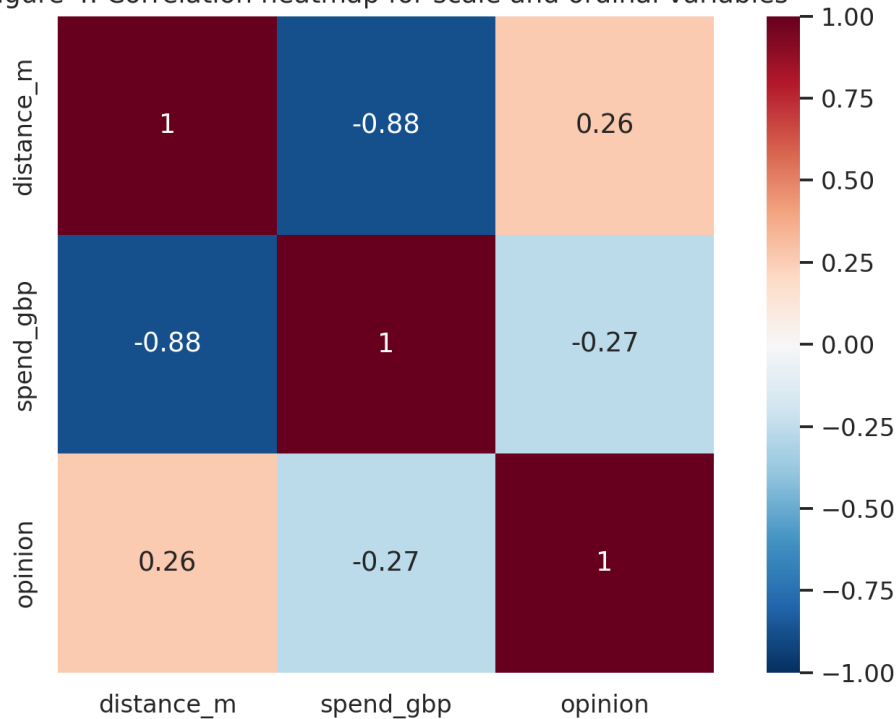


Figure 7. Correlation heatmap

Table 7. Summary of Key Relationships

Relationship	Pearson R	P-Value
Distance and spend	-0.88	$p < .001$
Council opinion and spend	-0.27	0.25
Distance and council opinion	0.26	0.27

The relationship summary confirms that distance and spend is the dominant pattern, while council opinion has much weaker associations with spend and distance. The evidence therefore supports a geographically structured revenue pattern although the available variables cannot explain it fully.

3.6 Alternative Relationships and Analytical Limitations

Several secondary patterns reveal why the results require careful interpretation. Newspaper choice and gender overlap substantially: all six DE customers are male while all seven G customers are female. This makes it difficult to decide whether differences in average spend relate to gender, newspaper preference, distance or a combination of all three. The pattern may reflect genuine customer segments, but it may also result from the small sample. It should therefore be used to develop questions for further research rather than to stereotype customers or design gender-based promotions.

Council opinion provides a useful comparison because spending does not rise or fall consistently across the four opinion categories. It appears less commercially relevant than accessibility and customer routine. Convivence, competition and ease of access may influence store choice, but the importance of each factor can only be established through evidence from the local setting. The analysis also needs a transparent and reproducible process: variables should be clearly coded, calculations should be checkable and each graph should answer a defined question. These controls strengthen credibility even when the dataset remains limited.

4. Why Dataset Analysis Matters

Dataset analysis turns separate observations into usable evidence for decision-making. Instead of relying on assumptions such as 'nearby customers are loyal' or 'newspaper buyers spend more', management can test these claims against recorded results. Descriptive statistics show what is typical, charts make patterns visible and correlations identify relationships that deserve further investigation. This supports business decision based on observable evidence rather than intuition alone.

The same evidence can guide different users. The shop owner may consider local promotions, collection or delivery, while the council may review access and infrastructure. However, the right response depends on why distant customers spend less. Good analysis must therefore separate confirmed findings from possible explanations and unanswered questions.

This distinction is critical. The dataset confirms that spending falls as distance increases but it cannot show whether the cause is carrying limits, weather, competitors or single-item newspaper visits. Identifying this evidence gap directs better data collection. Clear visuals and plain language then help decision-makers understand the priority finding without overstating what the data proves.

5. Missing Variables & Improved Data Collection

The dataset is useful for the assessment because it allows several forms of descriptive analysis but it is limited for professional decision-making. The most important weakness is that it records where customers live but not how they experience the journey to the shop. Distance in meters is not the same as accessibility. A 500-meter journey may be easy for a cyclist with panniers, manageable for a driver, difficult for an older pedestrian and unattractive during heavy rain. To make the dataset more credible, future research should include variables that explain travel, capacity, timing, purpose and context.

- Mode of travel, including walking, bicycle, car, bus, taxi or other transport.
- Carrying capacity, including backpack, shopping bags, bicycle basket, pram, trolley or vehicle boot.
- Weather and daylight conditions, including rain, temperature, wind, season and darkness.
- Visit frequency, because weekly spend may come from one large visit or several small visits.
- Basket composition, including newspapers, groceries, drinks, snacks, tobacco, lottery and household goods.
- Purpose of visit, such as planned newspaper purchase, emergency top-up, commuting stop or social routine.
- Access conditions, including parking, bicycle racks, pavement quality, roadworks and perceived safety.
- Competing shops, including whether customers pass closer alternatives on the way.
- Customer context, such as age band, household size, employment status and price sensitivity.
- Customer loyalty, including years using the shop, satisfaction with service and loyalty-scheme status.

These variables would allow the shop to test explanations and select appropriate responses. If distant customers spend less because they walk and cannot carry much, pre-ordering, delivery or collection bundles may help. If they pass a competitor, pricing, range or service differentiation may be more relevant. Further data should therefore be complete enough to address the decision, relevant to the local context and recorded accurately.

A stronger follow-up study should combine a short customer questionnaire with till-record data. The questionnaire should ask about travel mode, journey purpose, carrying capacity, visit frequency, weather conditions and access barriers. Till data should record actual basket value and product categories because recalled weekly spend may be inaccurate. The sample should also be larger and collected across different weekdays, weekends, mornings, evenings and weather conditions. This would improve reliability and reduce the risk that the current findings reflect an unusual collection period. Customers should also be informed of the purpose of the research, responses should be anonymous and sensitive demographic questions should be optional.

The final report should retain analytical precision while using accessible language. Statistical terms should be applied correctly and explained in relation to the shop. Each visual then serves a distinct purpose: the distribution shows customer value, the scatterplot displays the distance–spend relationship, the banded chart supports management interpretation and the heatmap compares the relative strength of relationships.

6. Practical Recommendations for Management

The analysis supports three practical recommendations: First, the shop should protect its nearby customer base because customers within 50 meters are currently the highest weekly spenders. This could be done through reliable availability of everyday essentials, accurate newspaper ordering, friendly service and a simple loyalty offer for repeat local customers. These actions fit the evidence because the strongest spending group is already local so the easiest growth may come from retaining and slightly increasing the value of customers who already use the shop regularly.

Second, the shop should test ways to increase spending among customers who live farther away. The data show that these customers spend less but the reason is unknown. A small trial could ask distant customers whether delivery, reserved newspaper collection, click-and-collect, bicycle parking or multi-buy bundles would increase their spending. This would turn the analysis into a useful management decision-making tool. It would also help management avoid guessing whether the barrier is distance, transport, carrying capacity, competing shops or product choice.

Third, the shop should improve its evidence base before making major changes. The current dataset is useful, but it is too small for expensive decisions such as changing the whole product range or investing heavily in delivery. A sensible next step would be a two-week data collection exercise combining till receipts with a short customer questionnaire. This would allow management to compare actual basket value with travel mode, visit purpose, weather and frequency. In this way, the business could move from descriptive analysis to targeted action based on stronger evidence.

7. Conclusion

This exercise demonstrated how a small dataset can support a neighbourhood shop when organised through tables, charts and comparisons. An important finding was that customer value is strongly associated with proximity. Distance had a very strong negative correlation with average weekly spend and accounts for approximately 76.6% of the variation but did not directly prove that distance itself causes lower spending. Nearby customers were therefore the highest-value group, supporting a higher focus on customer retention, product availability, accurate newspaper ordering and consistent service.

The analysis was equally valuable because it exposed what the data cannot explain. Distance may represent journey effort, transport, carrying capacity, weather, local safety, public transport access or competition. Other factors could materially change the conclusion including relative prices, promotions, product range and freshness, cleanliness, local reputation to include historical criminal activity, ease of communication with staff and the wider environment. A follow-up survey should prioritise travel mode, visit purpose and frequency, basket contents, competitor use, price perceptions, promotions, product availability, personal safety and shop security plus overall customer satisfaction. Survey variables may also include if the shop is used simply for convenience, trust or access to certain products. Combining survey responses with till records would provide a more complete basis for any future decisions. The current dataset cannot distinguish between these explanations.

Actionable management decisions require rigorous data integrity, not speculative assumptions drawn from limited samples. Accurate records, spreadsheets, tables and charts turn customer information into management evidence. Management should not make costly changes based on small samples alone. Customer data from surveys can show management where to investigate, *but conclusions are only as strong as the variables collected.*

8. Appendix A. Dataset Provided

Table A1. Dataset Used for Analysis

Customer	Distance M	Spend £	Gender	Newspaper	Opinion Score	Opinion Label	Distance Band
a	95	17	Male	W	1	Very discontent	51-200m
b	5	24	Male	DE	1	Very discontent	0-50m
c	7	23	Male	DE	3	Content	0-50m
d	200	14	Male	DE	4	Very content	51-200m
e	270	12	Female	G	2	Discontent	201-500m
f	15	22	Male	W	3	Content	0-50m
g	340	11	Female	W	4	Very content	201-500m
h	450	9	Female	G	1	Very discontent	201-500m
i	16	21	Female	G	1	Very discontent	0-50m
j	700	7	Male	DE	1	Very discontent	501m+
k	50	19	Female	G	3	Content	0-50m
l	1000	6	Male	W	3	Content	501m+
m	50	19	Female	G	2	Discontent	0-50m
n	140	15	Male	W	3	Content	51-200m
o	300	11	Female	G	3	Content	201-500m
p	210	13	Male	T	4	Very content	201-500m
q	12	23	Male	DE	2	Discontent	0-50m
r	180	15	Male	DE	2	Discontent	51-200m
s	50	18	Female	G	1	Very discontent	0-50m
t	1100	6	Female	T	4	Very content	500m+