

THE IMPACT OF CHANGES IN THE FREQUENCY OF SMALL SHOP'S ANALYSIS OF
CUSTOMER DATA ON THEIR PROFITABILITY

A Research Proposal

Presented to

The Graduate and Continuing Studies

Department of School of Technology and Workforce Learning

Pittsburg State University

Pittsburg, Kansas

In Partial Fulfillment

of the Requirements for the Degree

Master of Science and Technology: Emphasis - Information Technology

by

Shoinbek Shoinbekov

May 23

© Shoinbek Shoinbekov 2023 [NOTE: Do NOT include copyright information on proposals]

Keywords: Data analysis, Frequency of customer data analysis, Transactional tracking,
Profitability, and Change in the frequency of customer data analysis

ABSTRACT

Throughout the whole U.S, small retail shops face numerous challenges in competing against one another and are in constant search for finding the correct approach to outcompete their rivals. One such approach to address this challenge is the analysis of customer data, which can provide valuable insight into customer preferences and behavior. However, the frequency with which small firms analyze this data can significantly impact their capacity to make a profit.

This study attempts to investigate the connection between small shops' use of consumer data analysis and their financial success in the American retail sector. For this study, a group of small business owners in major metropolitan areas like Kansas City and Saint Louis will be targeted using a cluster-sampling strategy. The study will try to equally divide a sample group of 70 small business owners by gender, with a focus on those between the ages of 29 and 45. Since the shops are the ones collecting the data, the shop owners will be in charge of overseeing the data collection process. For the purpose of this study, the researcher will contact the owners of these shops using a variety of means, including emails, phone calls, LinkedIn, and other social media platforms, to request their consent to analyze their customer data in order to address the study's central topic. To ascertain whether there is a relationship between the independent variable, the length of time it takes to collect and analyze data, and the dependent variable, profit, the data will be analyzed using a variety of statistical techniques, including Pearson Correlational Coefficient, Inferential Statistics, Descriptive Statistics, and Regression through various software such as RStudio, Python, Excel and Tableau. The relationship and direction of the x and y variables will also be shown through graphic representations such bar charts, scatterplots, and maps.

DEFINITION OF TERMS

Small shops: A store that carries a diverse range of products, usually within a limited physical space. These shops typically sell a variety of goods, including groceries, household items, toiletries, electronics, clothing, and other consumer goods.

Frequency of customer data analysis: The regularity with which a business reviews and evaluates the information gathered from its clients. This information might range from demographics to feedback to shopping habits to preferences.

Predictive Analysis: Method of analyzing past data and making predictions about future occurrences or behaviors using statistical algorithms and machine learning techniques. Finding patterns and links in the data that can be used to predict future outcomes entails using data mining, statistical modeling, and other analytical tools.

Descriptive Analysis: practice of examining past data to gain knowledge and comprehend what has transpired. In order to effectively summarize and describe data, it uses statistical and graphical approaches.

Data Visualization: Data visualization is the practice of presenting data and information in a visual manner, such as maps, charts, and graphs. Making complex data sets more approachable, intelligible, and actionable is the aim of data visualization.

Regression: Regression is a statistical technique used to represent the correlation between one or more independent variables and a dependent variable.

Pearson Correlational Coefficient: The degree and direction of the linear link between two continuous variables are assessed statistically using the Pearson correlation coefficient. It is represented by the letter "r" and has a value between -1 and +1.

Inferential Statistics: Making inferences and predictions about a population based on a sample

of data falls under the category of inferential statistics. It entails analyzing and interpreting data using statistical techniques, as well as inferring information about a broader population or group of people from a smaller subset of data.

Python: High-level, general-purpose programming language which is used in a variety of industries including artificial intelligence, machine learning, data science, web development and more.

RStudio: Integrated development environment (IDE) which offers various tools for debugging, data visualization, and data analysis, as well as a user-friendly interface for developing and running R code.

Tableau: A user-friendly data visualization software for users to connect, analyze, and visualize data. Businesses, academics, and analysts use it to swiftly produce interactive dashboards, reports, and visualizations from large and complicated datasets.

Bar Charts: A form of data visualization chart in which data is shown as rectangular bars. The length or height of each bar, which represents a category, reflects the value or frequency of the data for that category.

Scatterplots: A typical style of chart used in data visualization that shows the correlation between two numerical variables is a scatterplot. With one variable displayed on the horizontal axis and the other variable shown on the vertical axis, each point on the chart represents a single observation or data point.

Data Analysis: A procedure for gathering and evaluating customer data in order to understand customer behavior.

TABLE OF CONTENTS

ABSTRACT.....	ii
DEFINITION OF TERMS.....	
LIST OF FIGURES.....	v
CHAPTER 1: INTRODUCTION.....	8
1.1 Purpose of the Study.....	8, 9
1.2. Significance of the Study.....	9,10
1.3. Research Questions/Hypothesis.....	10,11,12
1.4. Assumptions.....	12
1.5. Limitations of the Study.....	13
CHAPTER 2: LITERATURE REVIEW.....	14
2.1 Literature Search Methodology.....	14
2.2. What is data and data analysis.....	15
2.3. Customer Data Analysis and Consumer Decision Making (CDM).....	16,17
2.4. Frequent Pattern Mining (FPM)	17,18
CHAPTER 3: METHODOLOGY.....	19
3.1. Research Design.....	19,20
3.2. Population.....	20
3.3. Sample.....	20,21
3.4. Data Collection Methods: Quantitative, Transactional Tracking, Surveys.....	21,22,23
3.5. Data Analysis Procedures: Pearson Correlational, Regression, Descriptive statistics, Data visualization, Data Interpretation.....	23,24
3.6. Role of the Researcher.....	24,25

3.7. Validity and Reliability.....	25,26
SUMMARY.....	26
REFERENCES.....	27
APPENDIX A: SURVEY QUESTIONNARE.....	28,29,30

LIST OF FIGURES

Figure 1. Graphical Representation of Alternative Hypothesis.....	10
Figure 2. Graphical Representaion of Null Hypothesis.	11
Figure 3. Data Collection Methodology Design.....	22
<i>Figure 4.</i> Bar charts, Scatterplot and Pearson Correlation Coefficient formula.....	24

CHAPTER 1

INTRODUCTION

If we take a moment and dive deep into the principles of economy and how it works, we will come to an understanding that small businesses play a crucial role in the world of economy. This paper is mainly concerned with small shops in metropolitan cities which are also known as mom-and-pop shops or corner stores. Small shops in metropolitan cities are just as important to the economy as they are in smaller towns or rural areas. In fact, they may be even more crucial in metropolitan areas due to the concentration of population and business activity. Small shops in urban areas contribute to the local economy's diversification. They frequently offer distinctive services that bigger businesses might not, which can draw clients and encourage economic development. Small shops are frequently at the forefront of innovation and fresh concepts in major cities. Because of their flexibility and agility, small shops are able to adapt swiftly to shifting market conditions and create new services to satisfy their clients' needs. One of their main benefits is to contribute to the growth and development of local economies. This is due to their purchase of goods and services from local suppliers, which support other businesses in the area.

1.1: Purpose of Study

The goal of this study is to determine how small retail shops' ability to turn a profit in the American retail sector is impacted by changes in how frequently they analyze customer data. Small retail shops have increased their data collection and analysis efforts in recent years in an effort to better understand customer behavior and preferences. As a result, there is now more competition in the retail sector as small shops try to exploit this information to their advantage.

There has been a study on the advantages of data analysis for larger businesses, but there is less evidence on how the change in frequency of small shops' analysis of customer data affects their capacity to make a profit in the U.S retail industry.

The study will use a quantitative research approach, collecting customer data from a sample of small retail shops in the U.S retail industry in big metropolitan cities. The goal of the research's correlation design is to ascertain whether there is a correlation between the frequency with which these shops analyze their customer data and the profit they make during a certain period of time. For the purpose of this study a cluster sampling approach will be used to select the appropriate 70 candidates of small shop owners equally divided into males and females whose age ranges between 29 and 45 years old. In case of not finding the appropriate number of participants, the scope of the research will have to be enlarged to come up with the desired number of participants.

1.2: Significance of Study

The significance of this study rests in its potential to provide light on the connection between small shops profitability and consumer data analysis in the U.S. retail sector. Small shops are increasingly depending on customer data analysis to get insights into consumer behavior and preferences as a result of the expanding availability of data and developments in data analytics. The correlation between the frequency of data analysis and small shop profitability, however, is not well understood.

This study is significant because it can inform small shops about the optimal frequency of customer data analysis necessary to achieve profitability. By determining the optimal frequency small retail shops can better understand the preferences of their customers, optimize inventory management, target their marketing efforts more effectively and improve customer satisfaction

and loyalty. Furthermore, the finding of this study can help small shops to identify behavioral patterns of their customers and develop theories. The main significance of this research is to become a roadmap for not only small businesses but all businesses to determine an optimal frequency of customer data analysis necessary to achieve the best result for the business as well as for the customers.

1.3: Research Question/Hypothesis

A Research question is a crucial element in a research study. It aids in defining the research's scope and directs the researcher in selecting the right methodologies, data, and supporting materials to gather. The research question for this particular study is “How does the change in frequency of small shops’ analysis of customer data affect their capacity to make a profit in the U.S retail industry?”. In order to determine a correlation between the two variables the following hypotheses are necessary to be tested:

- H0: There is no significant relationship between the change in frequency of small shops’ analysis of customer data and their capacity to make profit in the U.S retail industry.

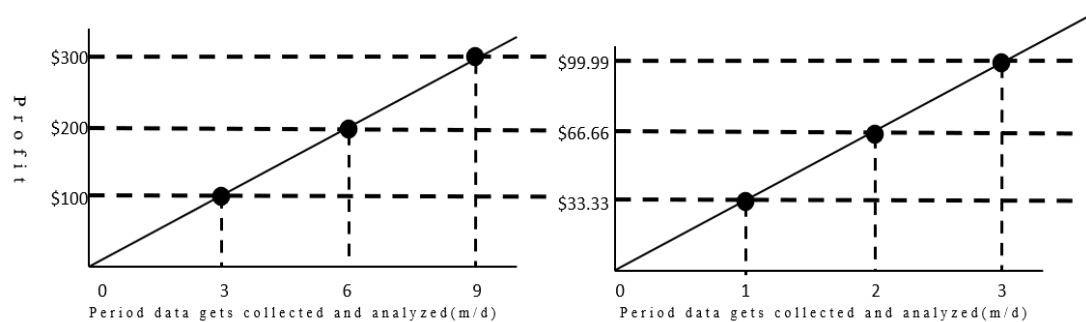


Figure 1. Graphical representation of Null Hypothesis

To understand what the graph above for the null hypothesis means, we have to come up with a hypothetical retail shop which collects customer data for a period of up to three months

and then analyzes it. Once they analyze the data and implement the findings to their business, they generate \$100 dollars. If they continue to do the same thing every other three months they may generate \$100 dollars every time, but what if they don't have to wait for three months to do the customer data analysis? What if they do the analysis every month and by doing so, they notice that they only generate \$33.33 dollars. This would mean that the change in frequency as it shows on the graph ($\$100/3\text{months} = \$33.33/1\text{month}$) did not affect profit at all, thus meaning the relationship between the change in frequency of customer data analysis and profit does not exist.

- HA: There is a significant relationship between the change in the frequency of small firms' analysis of customer data and their capacity to make profit.

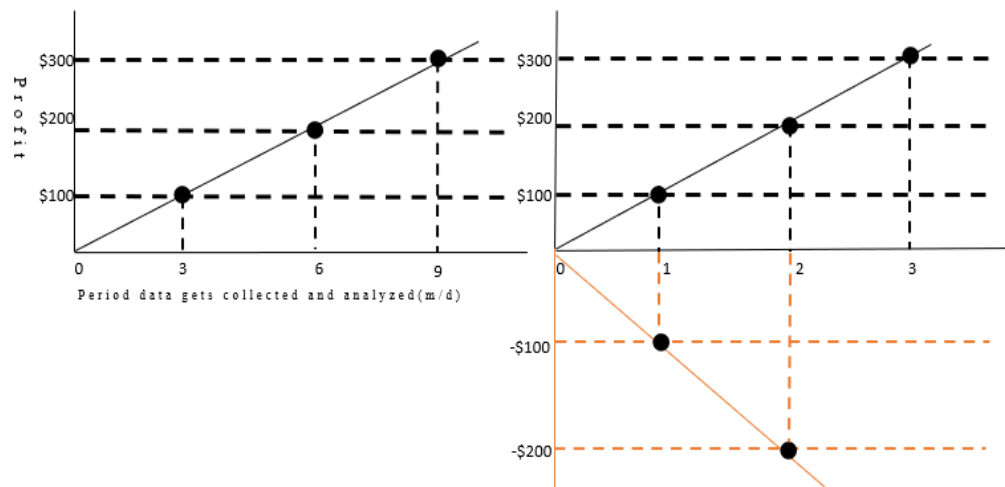


Figure 2. Graphical Representation of Alternative Hypothesis

Alternatively, to the hypothetical scenario above, we can see that in this case there is a relationship between the change in the frequency of customer data analysis and profit. The effect of the change in frequency can either result in a profit for the small retail shop or a loss. As we can see from the graph above the hypothetical retail shop has the potential to make \$100 in just one month if it were to change its customer data analysis frequency from 3 months to just 1

month. In three months, the shop has a potential to make \$300 dollars which supports the idea of positive correlational alternative hypothesis. It could also end up as a negative correlational hypothesis which could result in a loss of profits for the shop.

1.4: Assumptions

When conducting research, it's critical to recognize the underlying assumptions underpinning the research question because they may have an impact on the design of the study and how the findings are interpreted. Some of the assumptions that are relevant for this study are:

1. Assumption of causality: There is a causal assumption that there is a link between the frequency of customer data analysis and small shops' ability to turn a profit in the US retail sector.
2. Assumption of relevance: This type of assumption is based on the idea that customer data analysis offers insightful data on consumer behavior and preferences, and that the data can be used to boost profitability.
3. Assumption of validity of data: The research topic makes the assumption that the information gathered on the analysis of customer data and profitability by small businesses is valid and credible. This is based on the supposition that the procedures used to gather the data are suitable, that the data is measured and reported accurately, and that any biases in the data are found and corrected.
4. Assumption of normality: There is an assumption that the variables under consideration are regularly distributed. This is predicated on the notion that the variable distribution is symmetrical, and that the majority of observations are centered around the mean.

1.5: Limitations of Study

1. The study is limited to participants of small retail shop owners in big metropolitan cities whose age ranges between 29 to 45.
2. The study has a limited sample size of 70 individuals.
3. The study is limited to self-reported data by the shops which could introduce bias or inaccuracies in the data.
4. Small shops' ability to turn a profit could be impacted by external factors, such as changes in the economy or adjustments in consumer tastes, regardless of how frequently they analyze customer data. As a result, it may be difficult to ascribe any increases in profitability to data analysis alone.
5. Small businesses may not assess consumer data with consistent or reliable quality, which could restrict the value of the data in guiding profitable decisions.
6. Due to time constraints, the study may not be as thorough or comprehensive in its data analysis as it may be.

CHAPTER 2

LITEATURE REVIEW

2.1: Literature Search Methodology

The methodology used for this study's literature search entails performing a thorough search of all existing research studies, articles, and publications that address the research question, "How does the change in frequency of small shops' analysis of customer data affect their capacity to make a profit in the U.S. retail industry"? Starting with the study question, the goal will be to search for pertinent terms and phrases such "customer data analysis," "small firms," "profit," and "retail industry." Then the aim will be to look for pertinent studies and papers using academic resources like Google Scholar, JSTOR, and ProQuest. To choose which papers are pertinent to the study, inclusion and exclusion criteria will be established. A publication's date range, format, language, and applicability to the research question are just a few examples of the criteria that may be used. Even though the search for academic articles did not turn up as many results as hoped, there were a handful that were at least somewhat relevant to the study's core subject. This only opens up a possibility for further investigation and the writing of more research papers on such untouched topics.

2.2: What is data and data analysis.

Data is regarded as one of the most precious resources for corporations, organizations, and governments in the modern world. The significance of data and data analysis has increased significantly as a result of the quick development of technology and the expanding availability of data. Any information that can be gathered, saved, and analyzed to offer perceptions and aid in decision-making is referred to as data. A variety of sources, including social media, website traffic, consumer interactions, financial transactions, and more, can provide data. Businesses can enhance their operations and reach their objectives by using this data to analyze consumer behavior, market trends, and other vital information. According to the article “Big Data: What is it and Why Should You Care” written by Richard L. Villars, Mathew Eastwood and Carl W. Olofson, just in 2010 over 1ZB (Zettabyte) of data was generated, and they predicted that by 2014 the world will generate 7ZB of data each year. Nowadays that number is much larger than what it was back in 2010.

Data analysis, which is the process of methodically studying and interpreting data in order to get valuable insights and make inferences, has a promising future. Data analysis is becoming even more common and available to businesses of all kinds thanks to the ongoing growth of data, technological advancements, and more readily available technologies. Businesses can derive even more important insights from their data as data analysis becomes more integrated with other technologies like artificial intelligence and the Internet of Things (IoT).

John W. Tukey in his article “The future of data analysis” gives much praise for the advancement of data analysis to statistics and believes that in the future it can “contribute much more” (Tukey, 1961).

2.3: Customer Data Analysis and Consumer Decision Making (CDM)

For small retail stores trying to gain a competitive edge and boost profitability, customer data analysis is a crucial tool. These retail stores can acquire insights into the behavior, tastes, and demands of their customers by studying customer data. This knowledge will enable such stores to modify their goods and services to better suit customer requirements, which may boost client satisfaction and loyalty.

A wonderful study has been conducted by Scarlett Wesley Melody LeHew, and Arch G. Woodside to understand “how consumers' decision-making styles relate to their shopping mall behavior” (Wesley, LeHew, Woodside, 2005) which is highly relevant to this study. In the study the researchers used “data analysis including the use of the comparative method” (Wesley, LeHew, Woodside, 2005) to recognize the effects of customer decision-making patterns. They used face-to-face data collection from consumers in the shopping malls whose number reached about 527 individuals. The researchers were mainly concerned with the two-super regional “(West Edmonton Mall and Mall of America in Bloomington) and two regional (Pier 39 in San Francisco and Forum Shops in Las Vegas)” malls.

According to the study's findings, customers' buying behaviors fluctuate significantly depending on how they make decisions. The study specifically identified four different decision-making styles: reliant, logical, intuitive, and spontaneous. It was discovered that each of these styles was linked to certain shopping habits and preferences.

For instance, it was discovered that customers with a rational decision-making style were more likely to shop alone, stay longer in the mall, and use coupons or discounts. On the other side, it was discovered that consumers with intuitive decision-making styles were more likely to shop with friends or family, spend less time in the mall, and make impulsive purchases.

According to the study, consumers' purchasing habits varied significantly depending on their age, gender, and financial level. For instance, it was discovered that younger shoppers were more prone to make impulsive purchases, but elderly shoppers were more likely to shop alone and take advantage of discounts or coupons.

The study by Scarlett Wesley Melody LeHew and Arch G. Woodside is significant to this study since it emphasizes how crucial it is to analyze customer data in order to comprehend consumer behavior and preferences. The results of the study by LeHew and Woodside show that there are substantial variances in customer decision-making patterns based on purchasing activity, which can be exploited to create more effective marketing strategies. Similarly, the goal of this study is to investigate how variations in the frequency of customer data analysis can impact the financial performance of small shops in the retail sector.

2.4 Frequent Pattern Mining (FPM)

Frequent pattern mining is an approach used to identify patterns or correlations that regularly co-occur in a dataset. It is also known as market basket analysis or association rule mining. A research “Frequent pattern mining: current status and future directions” conducted by Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan dive deep into the world of frequent pattern mining and believe that “frequent pattern mining research has substantially broadened the scope of data analysis” (Han, Cheng, Xin, Yan, 2006).

An example of frequent pattern that is being demonstrated in the research is the frequent appearance of a “set of items, such as milk and bread” together. Identifying such types of frequencies is the bread and milk of data analysis itself. Data analysis relies on these types of patterns, and it is because of them small retail shops can identify customer’s behavior pattern and enhance their marketing strategies in order to generate more profit.

The very first-time frequent pattern mining (FPM) was proposed by Agrawal et al. (1993) to be used for “basket analysis in the form of association rule mining” (Han, Cheng, Xin, Yan, 2006). He used it for the purpose of identifying the relationship between various goods that customers place in their “shopping baskets”. For example, if a customer was to buy milk at a store, how likely is he or she is going to buy cereal as well (and what would be the type of cereal they may purchase)? This type of information can assist small retail shops with targeted marketing and shelf space organization which can result in improved sales.

Overall, the research conducted by Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan provides an overview of where FPM research stood at that time and where it was heading. The research can also assist with understanding the numerous frequent pattern mining algorithms and methodologies as well as the existing and future uses of FPM in a variety of industries, including the retail sector. The study by Jiawei Han, Hong Cheng, Dong Xin, and Xifeng Yan on "Frequent pattern mining: current status and future directions" is directly related to data analysis and offers helpful insights for investigation into the effects of change in the frequency of customer data analysis on small shops' profitability in the US retail industry.

CHAPTER 3

METHODOLOGY

3.1. Research Design

For the purpose of this study a correlational approach will be employed to understand the relationship between the frequent use of customer data analysis conducted by small retail shops and its impact on their profitability. The use of correlation approach is an ideal design for this study simply because it is heavily reliant on data and data analysis. What also makes correlational approach an ideal design for this study is the fact that the data is not fixed but continuously changing.

The study would start by selecting a sample of small shops through cluster sampling in big metropolitan cities such as Kansas City and Saint Louis in the U.S. retail sector and determining how frequently they analyzed consumer data during a predetermined time frame, like a month. The quantity of times consumer data was examined throughout that time period could be used to gauge the frequency of analysis.

Small shops' profitability would be assessed using financial information like sales, profit margin, and return on investment (ROI). Financial reports or company records may also contain this information. A correlation study would be carried out once data for both variables had been gathered to ascertain the degree and direction of the association between the frequency of customer data analysis and the profitability of small shops in the U.S. retail industry. Software for statistical analysis, such as Statistical Package for Social Sciences (SPSS) could be used for this analysis.

If a statistically significant positive association is discovered, while conducting the Pearson Correlation Coefficient it could mean that small shops in the U.S. retail sector will be more profitable if consumer data analysis is conducted more frequently. However, to prove a causal relationship between these factors, further study would be required, and it is crucial to remember that correlational studies do not establish causality.

3.2. Population

The population for this study consists of 85 small retail shop owners who are divided between men and women. The population as a whole is aged between 29 and 45 and has been in business for roughly 6 to 7 years. According to the data provided by Robert V. Krejcie and Daryle W. Morgan, for a population of 85 a searcher only needs a sample size of 70 individuals which will be equally divided into 35 males and 35 females for the purpose of this study.

A representative sample of the population will be chosen using cluster sampling. In order to do this, the population will first be segmented into smaller groups or clusters based on location or other relevant variables, and then a sample will be randomly chosen from each cluster.

After the sample has been chosen, data will be gathered from the owners of small retail shops using a survey technique to provide them with an in-depth purpose of the study and ask them for their customer data to be used for the research study. The information gathered should relate to the company's business practices, marketing tactics, financial performance, customer data, frequency of consumer data analysis, and other pertinent factors.

3.3. Sampling

As previously stated, a cluster sampling strategy will be used for the purposes of this study. Cluster sampling can be an effective approach for this type of study because the population of small retail shop owners is likely to be widely dispersed across different

geographic areas or regions. Cluster sampling involves dividing the population into smaller groups or clusters based on certain criteria, such as geographic location, and then randomly selecting a sample from each cluster. The use of cluster sampling allows researchers to gather information quickly and cost-effectively from a representative sample of small shop owners from various locales or regions without having to visit each one separately. Cluster sampling can also aid in ensuring that the sample is diverse and representative of the larger population.

3.4. Data Collection Methods

The data collection approach is one of the most crucial factors in producing an appropriate response to the research question for a highly data-driven research project like this one. In order to collect the accurate consumer data required for the research study, the researcher is heavily dependent on the owners of these small retail stores and their computer system. The following procedure will be used to reach out to the study participants and ask them for the data the researcher needs.

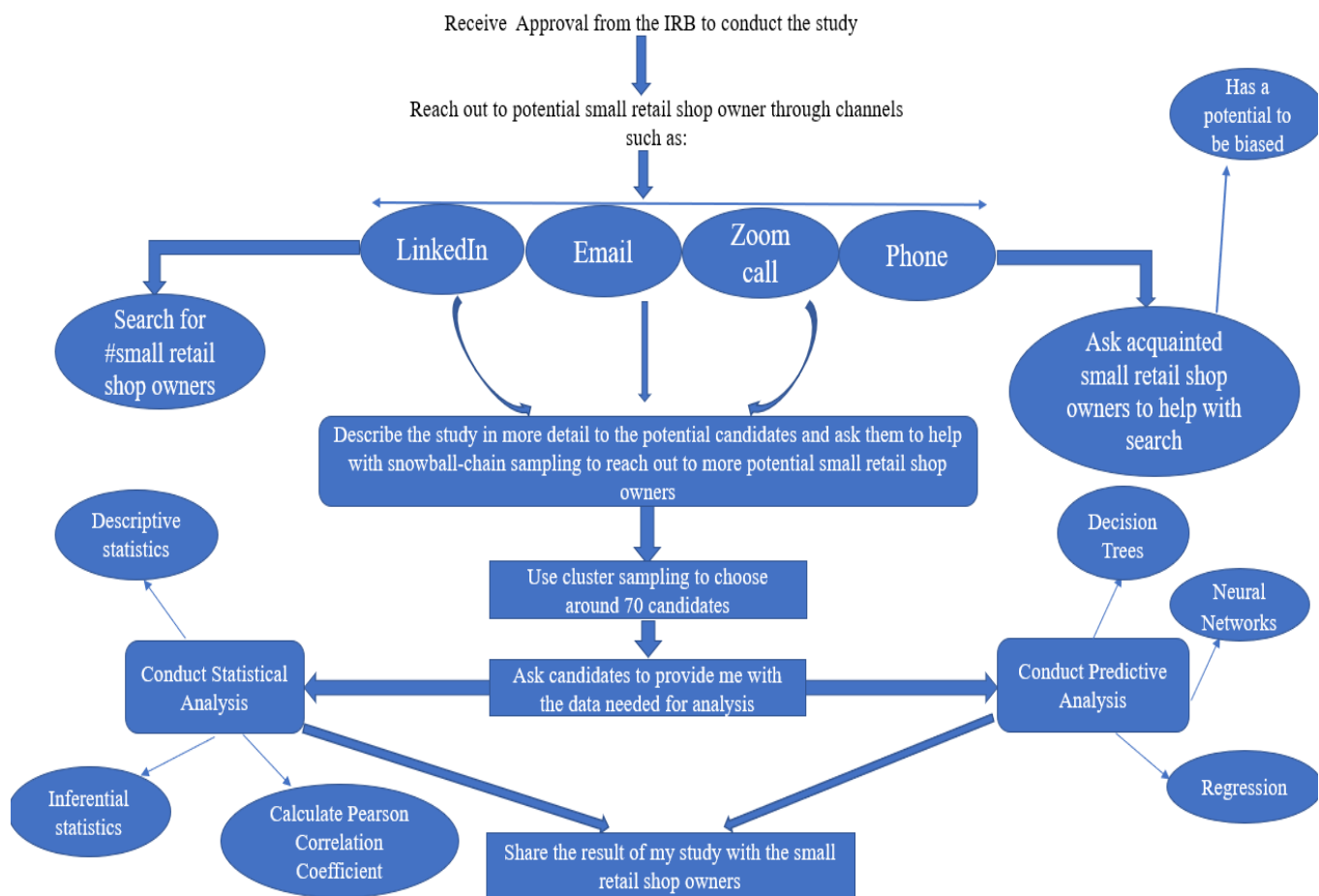


Figure 3. Research design. A flowchart of the study's process

Asking potential candidates for referrals though snowball-chain sampling as noted in the figure above may also result in bias.

To get the best quality data needed for the research the candidates will be asked to utilize transactional tracking. Through transactional tracking data such as time of the transaction, place of the transaction, price of the item bought, payment method and quantities and qualities associated with the transaction will be recorded. This type of data is ideal for identifying patterns and trends in customer behavior, such as which products are popular and when they are most

frequently purchased. Small retail shops can manage their inventories, develop pricing plans, and plan marketing efforts by evaluating this data. Transaction tracking also enables small retail shops to spot and address any problems that might be influencing their sales, such issues with a specific product or a fall in client pleasure.

3.5. Data Analysis Procedures

Data analysis is the process of transforming raw data into meaningful insights or conclusions. It is a critical component of any research study, as it allows the researcher to make sense of the data collected and draw conclusions about the research question. Once the data is collected the very first step will be taken toward data cleaning and preparation. This includes examining the data for errors or omissions, eliminating duplicates, and converting it into an analysis-ready format. After the data has been cleaned and is ready for analysis the next step will be to perform descriptive statistics on the data. This will entail computing statistics for the various variables in the data collection, including mean, median, standard deviation, and range. This will offer a broad picture of the data and make it easier to see any outliers or odd patterns.

Once the descriptive statistics approach is over the next move will be to conduct correlation analysis. This is the primary analysis for this study which will involve running a link between the change in the frequency of small shops analyzes of customer data and those shops' ability to turn a profit in the U.S. retail sector. The main correlational technique that will be used during this phase is Pearson Correlation Coefficient which is a measurement based on statistics that establishes the magnitude and direction of linear relationship between two continuous variables.

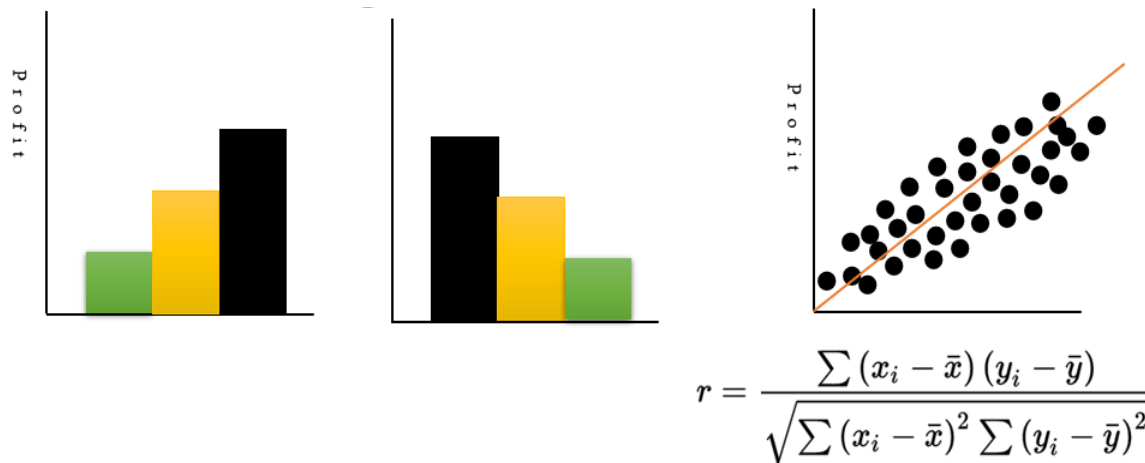


Figure 4. Bar charts, Scatterplot and Pearson Correlation Coefficient formula

The targeted Pearson Correlation Coefficient which the research is hoping to achieve is around ($0.7 < R < 0.9$) for the dependent and independent variables to be considered highly correlated. However, a Pearson Correlation Coefficient of ($0.5 < R < 0.7$) will also be considered for the purpose of this study.

After this step another statistical analysis which is called regression analysis will have to be conducted. In addition to correlation analysis, regression analysis may also be used to investigate the relationship between the frequency of customer data analysis and profit in the retail industry, while accounting for other factors like the size, age, and location of the business. The next two steps after regression analysis will be to visualize the data using spreadsheets, Tableau, RStudio and interpret the result of the study.

3.6. Role of the Researcher

The role of the researcher for this study will be essential in ensuring that the research is carried out in a morally and professionally appropriate manner and that the data gathered is accurate and trustworthy. The researcher will be in charge of deciding on the best data gathering techniques, like transaction tracking, and making sure that they are properly implemented.

Additionally, the researcher must ensure that the study's participants are properly aware of the study's goal, their part in it, and the protection of their privacy and their customer's data. The researcher will need to properly arrange and handle the data once it has been acquired. In order to do this, it may be necessary to build databases, clean and prepare the data for analysis, and look for any flaws or discrepancies.

Additionally, the researcher will be responsible for analyzing the data and using appropriate statistical methods to answer the research question. It is also the responsibility of the researcher to draw conclusions from the findings and document every procedure in a clear and detailed manner. In order to guarantee the study's transparency and reproducibility, this documentation will be essential. Additionally, the researcher must be open to suggestions for improvement and be prepared to modify their approach in light of new information or advice from peers and colleagues.

3.7: Validity and Reliability

Any quantitative research being conducted must be valid in order for the findings to fairly represent the topic being investigated. For this research, however validity is crucial for determining the reliability and accuracy of the findings. For the purposes of this study, internal validity will be the main focus because it ensures that no other factors can account for the observed association between the frequency of customer data analysis and profit. For instance, geography, competition, and economic conditions could all have an impact on how profitable small retail businesses are. The study may need the use of statistical techniques like regression analysis to account for these variables, which can isolate the impact of the independent variable on the dependent variable. It is essential for this study to be internally valid and in order for it to be internally valid extraneous variables such as market trends, competition or sampling bias such

as selection of particular participants or history of the shop will either need to be eliminated or minimized to the point where they will have no effect on the dependent and interdependent variables.

3.8: Summary

The goal of this study is to determine whether there is a connection between the regularity with which small shops analyze consumer data and their ability to turn a profit in the very competitive US retail sector. The main goal of the study is to determine how frequently analyzing consumer data affects the profitability of small retail stores in big metropolitan cities such as Kansas City and Saint Lois. A correlational research design will be used in the study to accomplish this goal, and data from a sample of 70 small retail shop owners who have been in operation for about 6-7 years will be collected.

To obtain accurate and trustworthy information about the frequency of customer data analysis and the profitability of the small retail stores, the data gathering technique will combine transaction tracking and surveys for the small retail shop owners. The results of the study will have substantial ramifications for small retail shop owners in the United States, giving them crucial insights into how the analysis of consumer data may be used to increase the profitability of their companies.

REFERENCES

1. Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607-610. <https://doi.org/10.1177/001316447003000308>
2. Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White paper, IDC, 14*, 1-14. https://www.admin-magazine.com/HPC/content/download/5604/49345/file/IDC_BigData_whitepaper_final.pdf
3. Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1), 1-67. <https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-33/issue-1/The-Future-of-Data-Analysis/10.1214/aoms/1177704711.pdf>
4. Wesley, S., LeHew, M., & Woodside, A. G. (2006). Consumer decision-making styles and mall shopping behavior: Building theory using exploratory data analysis and the comparative method. *Journal of Business Research*, 59(5), 535-548. <https://www.sciencedirect.com/science/article/abs/pii/S0148296306000051>
5. Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery*, 15(1), 55-86. <https://doi.org/10.1007/s10618-006-0059-1>

APPENDIX A: SURVEY QUESTIONNAIRE

1. What is your age range?

- ☐ 18-25
- ☐ 26-35
- ☐ 36-45
- ☐ 46-55
- ☐ 56 and above

2. How long have you been running your small retail shop?

- ☐ Less than 1 year
- ☐ 1-2 years
- ☐ 3-4 years
- ☐ 5-6 years
- ☐ More than 6 years

3. Do you analyze customer data to make business decisions?

- ☐ Yes
- ☐ No

4. How often do you analyze customer data?

- ☐ Weekly
- ☐ Monthly
- ☐ Once every 2 months
- ☐ Once every 5 months
- ☐ Once in a year

- Never
5. How do you store and manage your customer data?
- Spreadsheet software (e.g., Microsoft Excel, Google Sheets)
 - Customer relationship management (CRM) software
 - Other (please specify)
6. Have you noticed a change in your shop's profits since analyzing customer data?
- Yes, profits have increased.
 - Yes, profits have decreased.
 - No, there has been no significant change in profits.
 - I have not analyzed customer data.
7. In your opinion, what is the most effective way to obtain customer data?
- Customer surveys
 - Transaction tracking
 - social media monitoring
 - Website analytics
 - Other (please specify)
8. Would you be willing to share your customer data for the purpose of this study?
- Yes
 - No
9. How important do you believe customer data analysis is to the success of your small retail shop?

- Very important
- Somewhat important
- Neutral
- Somewhat unimportant
- Not important at all

10. In your opinion does the change in the frequency of customer data analysis effects small retail stores' ability to generate income?

- Yes
- No