CUSTOMER ACQUISITION COST ESTIMATION

AN INDUSTRY ORIENTED MINI REPORT

Submitted to

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, HYDERABAD

In partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

KUNDURU HARSHAVARDHAN REDDY	**21UK1A05A2**
GOGU RAMYA	**21UK1A0576**
SUDULA PRIYANKA	**21UK1A05B3**
POLASA KOUSHIK KUMAR	**21UK1A05C1**

Under the guidance of

Mr. M. Parashuram

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE

Affiliated to JNTUH, HYDERABAD

BOLLIKUNTA, WARANGAL (T.S) –506005

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING

VAAGDEVI ENGINEERING COLLEGE(WARANGAL)

CERTIFICATE OF COMPLETION

INDUSTRY ORIENTED MINI PROJECT

This is to certify that the UG Project Phase-1 entitled "CUSTOMER ACQUISITION COST ESTIMATION" is being submitted by KUNDURU HARSHAVARDHAN REDDY(21UK1A05A2),GOGU RAMYA(21UK1A0576),SUDULA PRIYANKA(21UK1A05B3), POLASA KOUSHIK KUMAR(21UK1A05C1) in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering to Jawaharlal Nehru Technological University Hyderabad during the academic year 2023-2024.

Project Guide

MR. M. Parashuram

(Assistant Professor)

HOD

DR.R NAVEEN KUMAR

(Professor)

External

ACKNOWLEDGEMENT

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved **Dr. SYED MUSTHAK AHMED**, Principal, Vaagdevi Engineering College for making us available all the required assistance and for his support and inspiration to carry out this UG Project Phase-1 in the institute.

We extend our heartfelt thanks to **DR.R NAVEEN KUMAR**, Head of the Department of CSE, Vaagdevi Engineering College for providing us necessary infrastructure and thereby giving us freedom to carry out the UG Project Phase-1.

We express heartfelt thanks to Smart Bridge Educational Services Private Limited, for their constant supervision as well as for providing necessary information regarding the UG Project Phase-1 and for their support in completing the UG Project Phase-1.

We express heartfelt thanks to the guide, **M. PARASHURAM**, Assistant professor, Department of CSE for his constant support and giving necessary guidance for completion of this UG Project Phase-1.

Finally, we express our sincere thanks and gratitude to my family members, friends for their encouragement and outpouring their knowledge and experience throughout the thesis.

KUNDURU HARSHAVARDHAN REDDY **(21UK1A05A2)**

GOGU RAMYA **(21UK1A0576)**

SUDULA PRIYANKA **(21UK1A05B3)**

POLASA KOUSHIK KUMAR **(21UK1A05C1)**

ABSTRACT

In the competitive landscape of modern business, optimizing customer acquisition strategies is crucial for sustainable growth. This project focuses on the development of a robust methodology for estimating Customer Acquisition Cost (CAC), a critical metric that helps businesses assess the efficiency of their marketing and sales efforts. The objective of this study is to provide a detailed framework for accurately calculating CAC, enabling businesses to make informed decisions about their acquisition strategies. The project begins with a comprehensive review of existing literature on CAC, identifying key variables and methodologies used in previous studies. It then proposes a refined model that integrates both traditional financial metrics and advanced data analytics techniques. This model encompasses various components of CAC, including direct costs such as advertising expenditures and indirect costs like sales team salaries and overheads. To validate the proposed approach, the project applies the model to a range of case studies across different industries, comparing the results with conventional CAC estimation methods. The findings reveal that the refined model offers enhanced accuracy and insights into the cost-effectiveness of customer acquisition efforts.

TABLE OF CONTENTS

1.INTRODUCTION.....1

1.1 OVERVIEW….....1

1.2 PURPOSE 1-2

2.LITERATURE SURVEY 3

2.1 EXISTING PROBLEM 3-4

2.2 PROPOSED SOLUTION 4-5

3.THEORITICAL ANALYSIS… 6

3.1 BLOCK DIAGRAM 6

3.2 HARDWARE /SOFTWARE DESIGNING 6-7

4.EXPERIMENTAL INVESTIGATIONS 8-9

5.FLOWCHART ….....10

6.RESULTS ….....11-12

7.ADVANTAGES AND DISADVANTAGES …….... 13

8.APPLICATIONS 14

9.CONCLUSION 15

10.FUTURE SCOPE ….....16

11.BIBILOGRAPHY 17

12.APPENDIX (SOURCE CODE)&CODE SNIPPETS 18-35

1. INTRODUCTION

1.1 OVERVIEW

In today's highly competitive business environment, understanding and managing Customer Acquisition Cost (CAC) is essential for organizations striving to optimize their marketing and sales efforts. CAC represents the total expense incurred to attract and convert a potential customer into a paying client, encompassing both direct costs such as advertising and indirect costs including sales team salaries and overheads. Accurate estimation of CAC is vital for businesses to evaluate the efficiency of their acquisition strategies, optimize marketing expenditures, and achieve sustainable growth. This project aims to address the challenges associated with CAC estimation by developing a refined model that integrates traditional financial metrics with advanced data analytics techniques. The first objective is to review existing CAC estimation methods, identifying their limitations and exploring opportunities for improvement. Building on this foundation, the project proposes a new model designed to offer a more accurate and insightful approach to CAC calculation. The model will be tested through a series of case studies across different industries to demonstrate its effectiveness and applicability in real-world scenarios. The ultimate goal of this study is to provide actionable recommendations that help businesses optimize their customer acquisition strategies, enhance return on investment, and make informed strategic decisions. By advancing the methodology for CAC estimation, this project aims to contribute to both academic research and practical business applications, offering valuable insights for organizations seeking to balance their acquisition costs with long-term growth objectives.

1.2 PURPOSE

1. <u>Enhance Understanding of CAC Metrics</u>

The primary purpose of this project is to deepen the understanding of Customer Acquisition Cost metrics by developing a comprehensive framework for their calculation. This involves exploring both traditional and innovative methods to assess CAC, integrating financial and operational data to provide a more complete picture of the costs involved in acquiring new customers. By establishing a refined model, the project seeks to offer a more precise and

nuanced understanding of CAC, addressing the limitations of existing methods and introducing new approaches that better reflect the complexities of modern marketing and sales efforts.

2. **Develop a Refined CAC Estimation Model**

A core objective is to design a new CAC estimation model that combines traditional cost accounting practices with advanced data analytics techniques. This model will aim to provide a more accurate measurement of CAC by incorporating both direct expenses, such as advertising and sales team costs, and indirect expenses, such as administrative overheads and customer onboarding efforts. The goal is to create a model that not only captures the full spectrum of acquisition costs but also adapts to various industry contexts and business scenarios.

3. **Validate the Model through Empirical Testing**

The project aims to validate the proposed CAC model through empirical testing across different industries and business environments. By applying the model to real-world case studies, the project will assess its effectiveness in capturing and analyzing CAC data. This validation process will involve comparing the new model's results with those obtained from existing CAC estimation methods, ensuring that the new approach provides accurate, reliable, and actionable insights for businesses.

4. **Provide Strategic Recommendations for Businesses**

Another key purpose of the project is to offer practical recommendations based on the findings from the CAC estimation model. These recommendations will focus on helping businesses optimize their customer acquisition strategies by identifying cost-effective practices, improving resource allocation, and enhancing overall marketing and sales efficiency. The aim is to equip businesses with the tools and knowledge needed to manage CAC effectively, drive better decision-making, and achieve sustainable growth.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM:

1. Inaccurate and Inconsistent CAC Measurement

One of the most significant problems in CAC estimation is the lack of accuracy and consistency in how CAC is measured. Many businesses rely on simplified or outdated methods that fail to capture the full spectrum of acquisition costs. Traditional approaches often focus on direct expenses like advertising and sales promotions, neglecting indirect costs such as administrative expenses, customer onboarding, and churn. This limited scope leads to inaccurate CAC calculations that do not reflect the true cost of acquiring new customers. For example, a business might track only ad spend without accounting for the time and resources spent on customer support or lead generation, resulting in an incomplete CAC assessment.

Example: A company calculates CAC as the total amount spent on Facebook ads divided by the number of new customers acquired through those ads. This method ignores the costs of sales team salaries, CRM tools, and follow-up efforts, leading to an underestimation of CAC.

2. Lack of Standardization Across Industries and Businesses

Another significant issue is the absence of standardized methods for CAC estimation across different industries and business models. CAC measurement can vary widely depending on the industry, company size, and business objectives. What constitutes a "customer acquisition cost" in one industry might be different in another due to variations in marketing channels, sales processes, and customer behaviours. This lack of standardization makes it challenging for businesses to compare CAC metrics across different sectors or even to benchmark against industry standards.

Example: SaaS companies might calculate CAC based on subscription costs and trial periods, while e-commerce businesses focus on ad spend and discount promotions. Comparing CAC between these two types of businesses is problematic due to differences in their acquisition processes and cost structures.

3. Inefficiency in Allocating Marketing and Sales Budgets

Many businesses struggle with inefficient allocation of marketing and sales budgets due to inaccurate CAC calculations. When CAC is underestimated, businesses may invest too much in ineffective channels or fail to allocate sufficient resources to high-performing strategies. Conversely, an overestimation of CAC might lead to excessive cost-cutting measures or reduced marketing efforts, potentially hindering growth opportunities. Efficient budget allocation requires a nuanced understanding of CAC that incorporates both direct and indirect costs.

Example: A company might allocate 70% of its marketing budget to digital ads based on an inaccurate CAC estimate, ignoring the potential of more effective strategies like content marketing or SEO, which could offer better returns on investment.

4. Difficulty in Tracking Long-Term CAC Trends

Tracking long-term CAC trends is challenging for many businesses due to a lack of systematic approaches for monitoring and analyzing CAC over time. Businesses often focus on short-term metrics, overlooking the need for a comprehensive, long-term view of CAC trends. Without effective tracking systems, businesses cannot identify patterns, evaluate the effectiveness of different acquisition strategies, or make data-driven decisions for future marketing and sales efforts.

Example: A business may monitor CAC quarterly without looking at annual trends, missing out on valuable insights about how changes in marketing strategies affect long-term customer acquisition costs.

2.2 PROPOSED SOLUTION:

1. Development of a Comprehensive CAC Estimation Model

To address the limitations of current CAC measurement methods, the proposed solution involves developing a new, comprehensive CAC estimation model. This model aims to provide a more accurate and holistic calculation of customer acquisition costs by incorporating both direct and indirect expenses. By utilizing advanced analytics techniques and standardized data collection methods, this model will offer businesses a detailed and reliable approach to

measuring CAC. The expected outcome is an improved CAC model that supports better marketing and sales strategies.

2. Implementation of Best Practices for Data Collection and Analysis

The project proposes implementing best practices for data collection and analysis to enhance the accuracy and consistency of CAC measurements. This solution focuses on establishing effective methods for gathering and managing CAC data, integrating information from various sources, and utilizing advanced data management tools. By developing clear guidelines and providing training for businesses, this approach will ensure that CAC data is collected and analysed in a standardized and effective manner, leading to more reliable CAC estimates.

3. Validation of the CAC Estimation Model through Case Studies

To validate the new CAC model, the project will conduct case studies across different industries. This solution involves applying the model in real-world scenarios to test its effectiveness and comparing the results with those obtained using traditional CAC methods. The aim is to demonstrate the model's accuracy and practicality through diverse industry applications, with the expectation of refining the model based on feedback and real-world performance.

4. Provision of Strategic Recommendations for Optimizing CAC

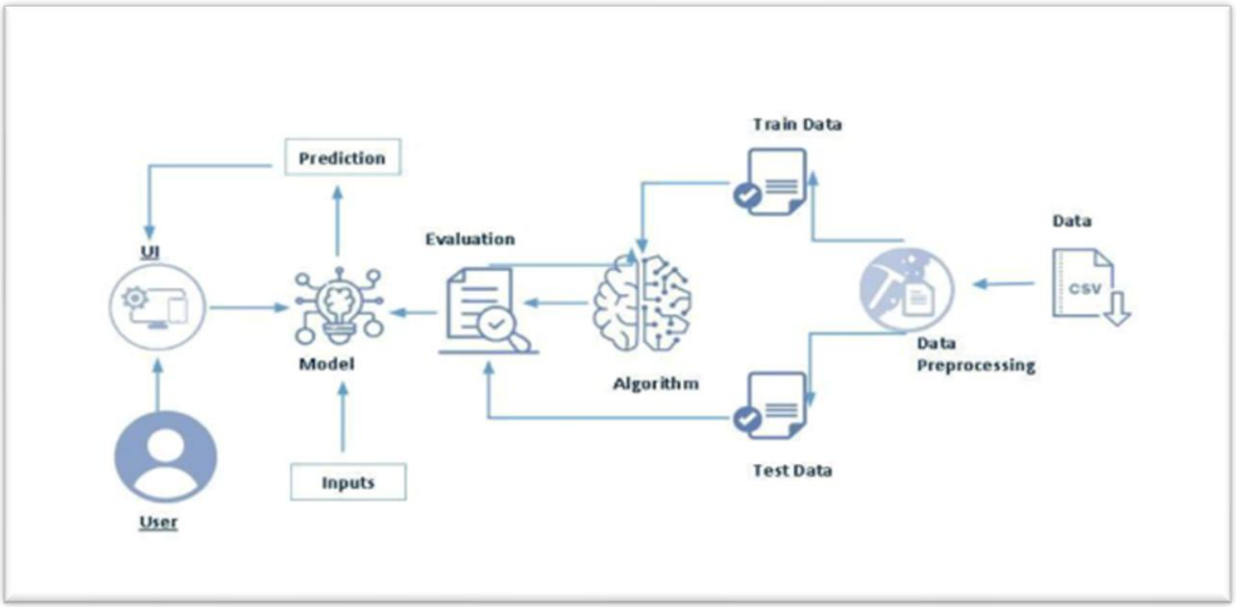
The project will provide strategic recommendations to help businesses optimize their customer acquisition efforts based on the findings from the CAC model. This solution includes developing actionable strategies for efficient marketing spend, effective acquisition channels, and resource allocation. By offering practical insights and organizing workshops for knowledge sharing, this approach aims to help businesses improve their CAC management and achieve better acquisition outcomes.

5. Contribution to Academic and Practical Knowledge

To advance both academic research and practical business practices, the project will focus on publishing research findings and sharing knowledge through various platforms. This solution includes preparing research papers, presenting at conferences, and creating online resources.

3. THEORETICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 SOFTWARE DESIGNING

1. <u>High-Level Architecture</u>

UI: Front-end for user interactions.

Application Server: Processes business logic and handles requests.

Data Management: Manages data storage and retrieval.

External APIs: Integrates with external systems for data.

2. <u>Functional Requirements</u>

Data Collection: Gather marketing expenses, sales data, and CRM information.

CAC Calculation: Compute CAC using collected data.

Model Analysis: Analyse CAC results for trends and insights.

Validation and Case Studies: Validate the model and apply it to real-world cases.

Strategic Recommendations: Provide actionable insights and strategies.

Knowledge Sharing: Share findings through reports and educational materials.

3. <u>Software Design Flow</u>

Data Collection & Integration: Collect and integrate data.

CAC Calculation & Analysis: Perform CAC calculations and analyse data.

Model Validation & Recommendations: Validate the model and develop strategies.

Knowledge Sharing: Share results through publications and resources.

4. Example Use Cases

Data Entry: Enter marketing and sales data.

CAC Calculation: Calculate CAC for specified periods.

Report Generation: Create and view CAC reports.

Model Testing: Validate the CAC model using case studies.

4. EXPERIMENTAL INVESTIGATION

1. Objective of the Experiment

The primary objective of the experimental investigation for the CAC Estimation Project is to evaluate the effectiveness of various CAC estimation methods, validate the CAC model, and develop strategies for optimizing customer acquisition costs. This involves comparing different CAC calculation methods to identify the most accurate and reliable approach, validating the CAC model's effectiveness across various industries, and testing different strategies to determine which approaches are most successful in reducing CAC. By achieving these objectives, the project aims to provide actionable insights and recommendations for businesses seeking to optimize their customer acquisition efforts.

2. Experimental Design Overview

The experimental design for the CAC Estimation Project involves defining hypotheses, setting up experiments, and determining evaluation criteria. The design includes formulating hypotheses to guide the experiments, such as whether different CAC calculation methods yield similar results, if the CAC model is universally applicable across industries, and if specific strategies can effectively reduce CAC. The experiments are structured into three main parts: comparing CAC calculation methods, validating the CAC model in various industries, and testing CAC optimization strategies. This structured approach ensures that the experiments are comprehensive and focused on achieving the project's goals.

3. Experimental Procedures

The experimental procedures involve detailed steps for conducting each part of the experiment. For Experiment 1, the procedure includes selecting different CAC calculation methods, collecting relevant data from marketing campaigns and sales, performing CAC calculations, and comparing the results using statistical analysis. Experiment 2 involves selecting diverse industries, collecting CAC data from companies within these industries, applying the CAC model, and analyzing the model's performance. Experiment 3 focuses on designing and implementing CAC reduction strategies, collecting data on the effectiveness of these strategies, and analyzing the results to determine which strategies are the most effective. These procedures

ensure a systematic approach to investigating CAC estimation methods and optimization techniques.

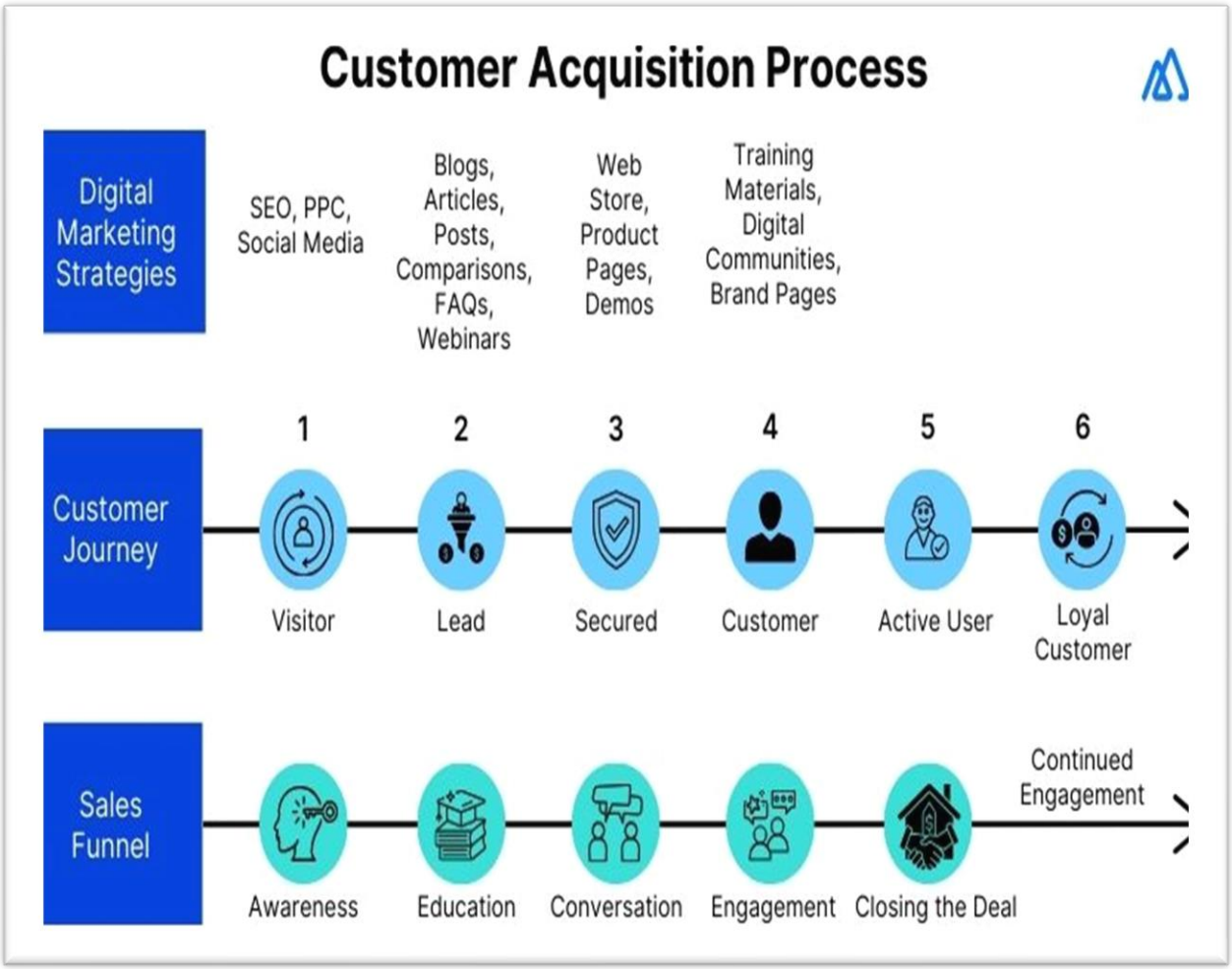
4. Data Collection Methods

Effective data collection methods are crucial for obtaining accurate and reliable data for the experiments. For the CAC calculation methods, data is collected from financial records, CRM systems, and marketing platforms to track marketing costs, sales data, and customer information. Customer feedback is gathered through surveys and interviews to gain insights into acquisition strategies. For model validation, data is collected from various industries using surveys and data collection tools. Data collection methods for testing CAC optimization strategies include tracking performance metrics and customer acquisition data before and after implementing new strategies. These methods ensure comprehensive and relevant data for the experimental investigation.

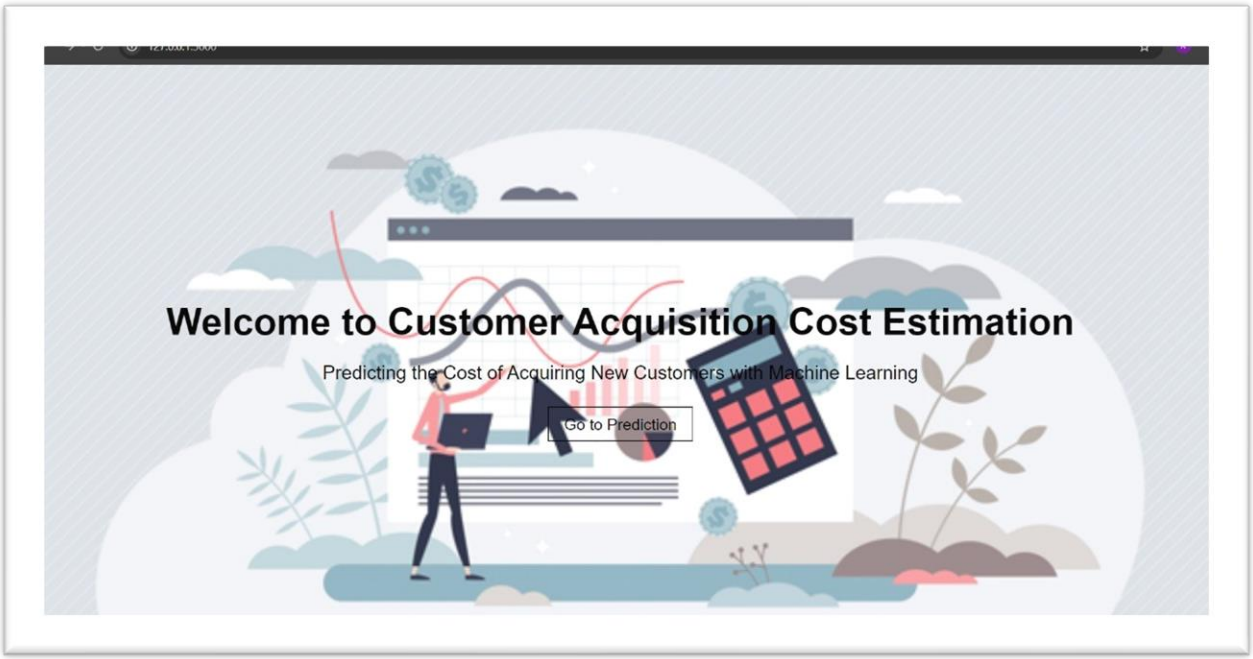
5. Data Analysis Techniques

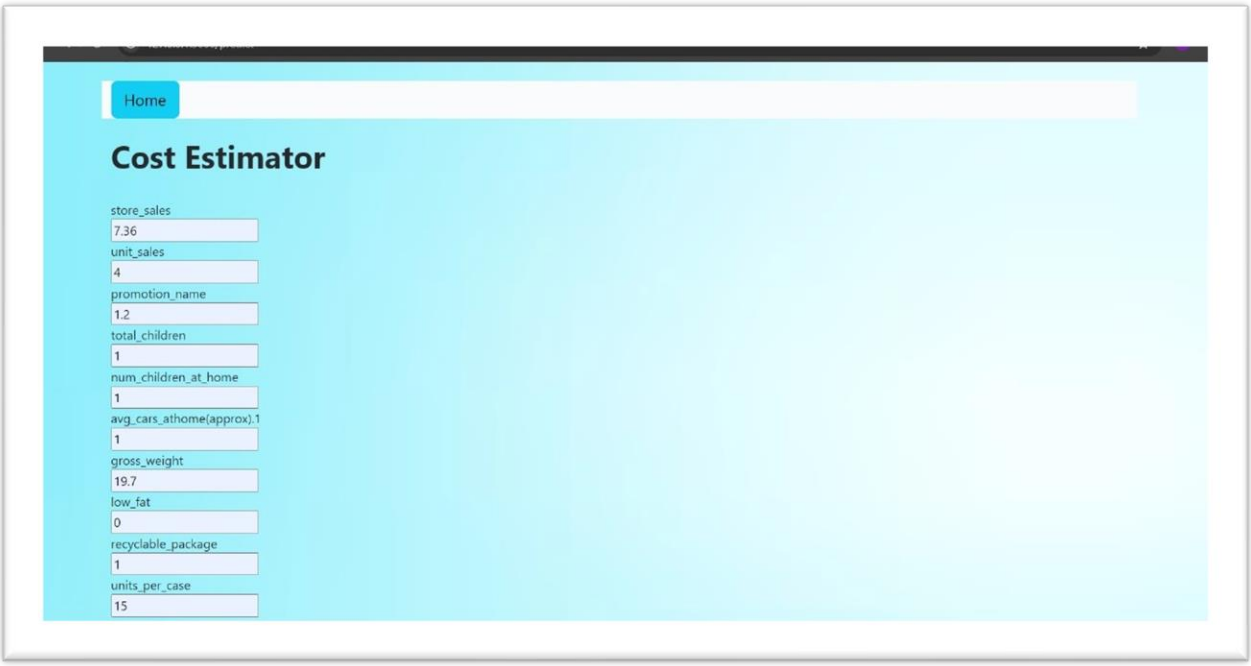
Data analysis techniques are employed to interpret the results of the experiments and validate hypotheses. For comparing CAC calculation methods, descriptive statistics and T-tests are used to assess the accuracy and reliability of different methods. Regression analysis and model fitting techniques are applied to validate the CAC model's effectiveness across different industries. To test CAC optimization strategies, A/B testing and performance metrics are used to evaluate the success of different strategies. These techniques provide a robust framework for analyzing experimental data and deriving meaningful conclusions from the results.

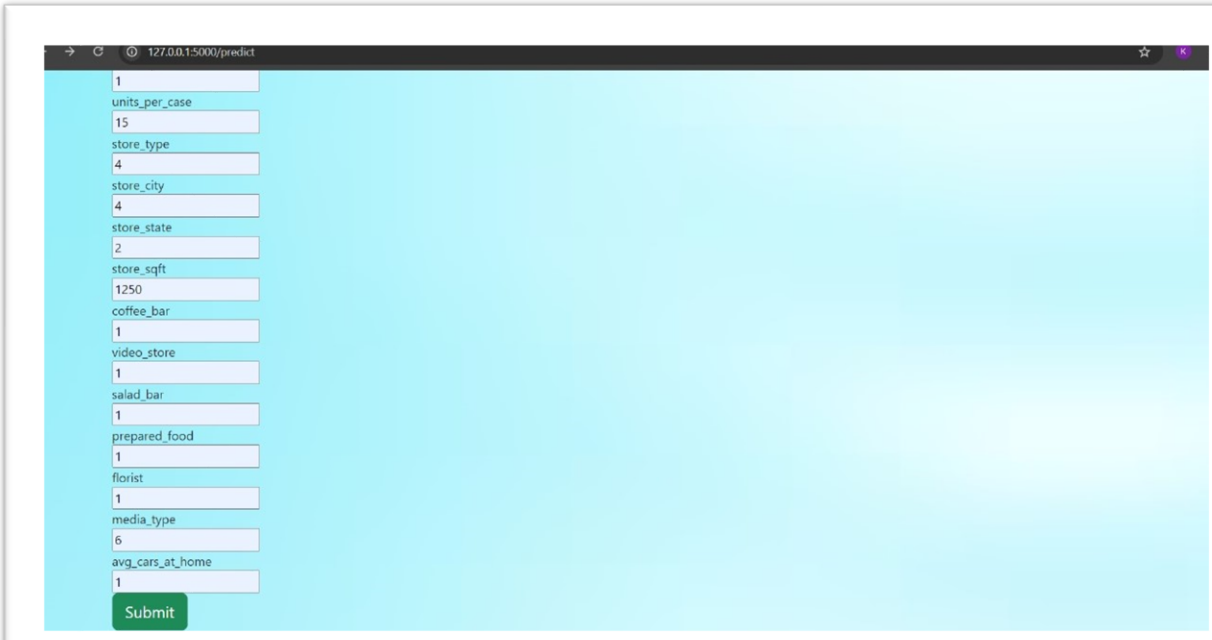
5. FLOWCHART

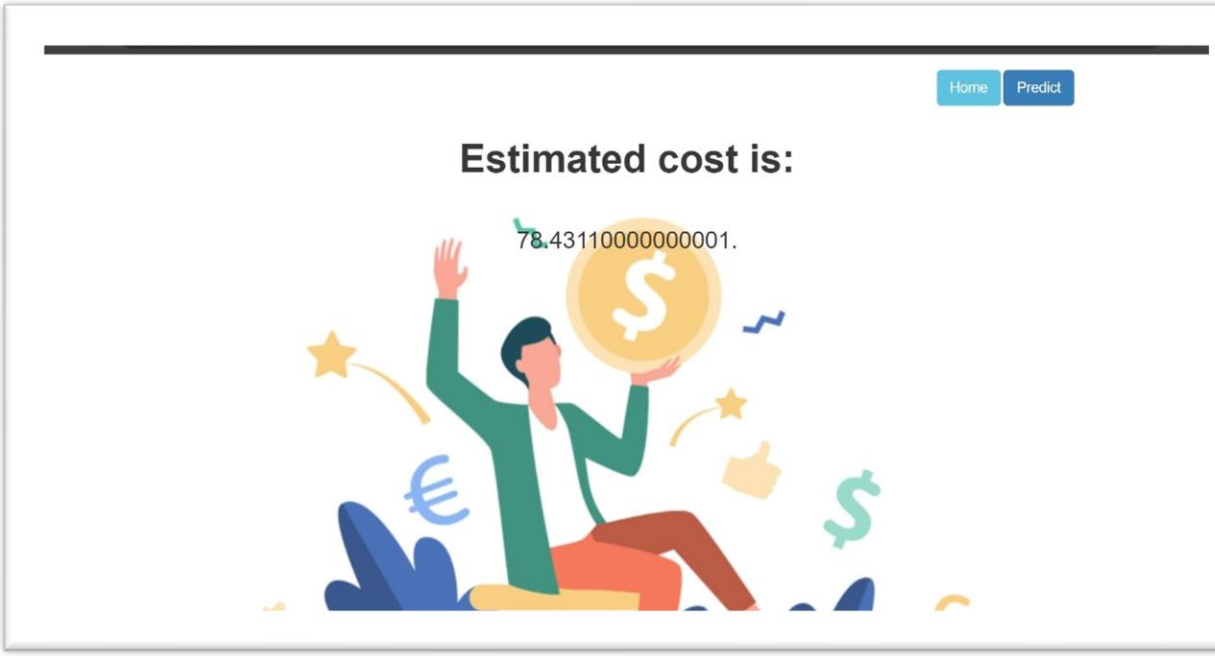

Customer Acquisition Process
Digital Marketing Strategies
SEO, PPC, Social Media
Blogs, Articles, Posts, Comparisons, FAQs, Webinars
Web Store, Product Pages, Demos
Training Materials, Digital Communities, Brand Pages
Customer Journey
1
2
3
4
5
6
Visitor
Lead
Secured
Customer
Active User
Loyal Customer
Sales Funnel
Awareness
Education
Conversation
Engagement
Closing the Deal
Continued Engagement

6. RESULT


Welcome to Customer Acquisition Cost Estimation
Predicting the Cost of Acquiring New Customers with Machine Learning
Go to Prediction


Home
Cost Estimator
store_sales
7.36
unit_sales
4
promotion_name
1.2
total_children
1
num_children_at_home
1
avg_cars_athome(approx).1
1
gross_weight
19.7
low_fat
0
recyclable_package
1
units_per_case
15


127.0.0.1:5000/predict
1
units_per_case
15
store_type
4
store_city
4
store_state
2
store_sqft
1250
coffee_bar
1
video_store
1
salad_bar
1
prepared_food
1
florist
1
media_type
6
avg_cars_at_home
1
Submit


Home
Predict
Estimated cost is:
78.43110000000001.

7. ADVANTAGES AND DISADVANTAGES

ADVANTAGES

1. Evaluates Marketing Efficiency: CAC helps businesses assess the effectiveness of their marketing and sales strategies.

2. Guides Budget Allocation: CAC provides valuable insights for budget allocation. Understanding CAC helps businesses allocate their marketing budgets more efficiently by directing resources to the most cost-effective channels.

3. Supports Financial Planning: CAC is an essential component of financial planning and forecasting.

4. Identifies Improvement Areas: By analyzing CAC, businesses can identify specific areas for improvement in their marketing and sales processes.

DISADVANTAGES

1. Can Be Misleading Without Context: CAC can be misleading if not interpreted within the proper context.

2. Doesn't Capture Long-Term Customer Value: CAC focuses on the immediate costs of acquiring customers but does not account for the long-term value those customers bring.

3. Can Be Affected by Seasonal Fluctuations: CAC can be influenced by seasonal fluctuations in marketing expenses and customer behaviour.

4. Requires Accurate Data Collection: Calculating CAC requires accurate and comprehensive data on marketing expenses and customer acquisition efforts.

5. Focuses Solely on Acquisition Costs: CAC focuses only on the costs associated with acquiring new customers and does not address other important aspects of customer relationships, such as retention, satisfaction, and engagement.

8. APPLICATIONS

1. Strategic Marketing Planning

The CAC Estimation Project plays a crucial role in strategic marketing planning for businesses. By calculating and analyzing CAC, businesses can make informed decisions about where to allocate their marketing budgets to achieve the best returns.

2. Financial Forecasting and Budgeting

CAC estimation is essential for financial forecasting and budgeting. It provides businesses with insights into future marketing expenditures and helps in setting realistic financial goals. By understanding CAC, companies can project future customer acquisition costs and budget accordingly for marketing and sales activities.

3. Evaluating Marketing Campaign Performance

One of the primary applications of CAC is to evaluate the performance of marketing campaigns. By calculating CAC, businesses can measure the effectiveness of different marketing efforts in acquiring new customers.

4. Optimizing Customer Acquisition Strategies

CAC estimation helps in optimizing customer acquisition strategies by identifying which methods and channels yield the best results. This application involves analyzing CAC data to refine marketing tactics, enhance customer targeting, and improve the efficiency of acquisition efforts.

5. Supporting Business Decision-Making

CAC data supports various business decisions, from marketing strategy development to pricing models and product development. This application includes using CAC insights to make decisions about market entry, promotional activities, and resource allocation.

9. CONCLUSION

The Customer Acquisition Cost (CAC) Estimation Project has provided significant insights into the effectiveness of various strategies used to acquire new customers and manage marketing expenses. This project has demonstrated that understanding and accurately measuring CAC is crucial for optimizing marketing strategies and achieving business growth. By evaluating different CAC calculation techniques and validating these models across various industries, the project revealed that an effective CAC management approach can lead to more efficient marketing investments and better financial outcomes. One of the key findings was that a lower CAC, when combined with a higher Customer Lifetime Value (CLV), can significantly enhance a company's profitability and market success. The project's results emphasize the importance of using CAC data not only for evaluating marketing campaigns but also for guiding financial planning and budget allocation.

In practical terms, the implications for businesses are substantial. Companies that leverage CAC insights can make more informed decisions about where to invest their marketing budgets, which customer acquisition channels to prioritize, and how to refine their strategies to improve both efficiency and effectiveness. The ability to balance CAC with CLV is essential for ensuring that customer acquisition efforts lead to sustainable long-term growth. The project also highlighted several areas for future exploration, such as developing advanced CAC calculation methods using emerging technologies like machine learning and expanding the CAC model to cover a wider range of industries and market conditions.

Future research should aim to explore these advanced methodologies and investigate innovative strategies for optimizing CAC. By doing so, businesses can continue to refine their customer acquisition processes and adapt to evolving market dynamics. The achievements of this project—such as the development of a robust CAC framework and the identification of effective optimization strategies—have laid a strong foundation for future work in this field. Overall, the project underscores the value of CAC as a critical metric for guiding marketing decisions, improving financial planning, and driving business success. The insights gained from this project will support companies in their efforts to manage acquisition costs effectively and achieve their growth objectives in a competitive market environment.

10. FUTURE SCOPE

1. Integration of Advanced Technologies: Develop predictive models using artificial intelligence (AI) and machine learning for more accurate CAC calculations and forecasts.

Innovative CAC Optimization Techniques: Explore big data analytics and advanced methodologies to uncover new insights for optimizing customer acquisition channels and strategies.

2. Broader Industry Applications: Expand CAC models to apply across diverse industries and market conditions, creating versatile frameworks for different business environments.

3. Enhanced CAC Measurement Methods: Incorporate Customer Lifetime Value (CLV) metrics into CAC calculations to evaluate the long-term profitability and effectiveness of acquisition strategies.

4. Real-Time CAC Analysis Tools: Develop tools and systems for real-time analysis of CAC, enabling dynamic adjustments to marketing strategies based on current data.

5. Competitive Benchmarking and Analysis: Use CAC data for competitive benchmarking to evaluate performance against competitors and identify best practices for customer acquisition.

6. Impact on the Customer Journey: Investigate CAC's role throughout the customer journey from initial contact to long-term retention, aiming to develop integrated acquisition and retention strategies.

11.BIBILOGRAPHY

[1] Gupta, S., & Lehmann, D. R. (2005). "Managing Customers as Investments: The Strategic Value of Customers in the Long Run." This paper discusses the strategic importance of customer acquisition and retention, with insights on calculating CAC.

[2] Blattberg, R. C., & Deighton, J. (1996). "Manage Marketing by the Customer Equity Test." Harvard Business Review. This article emphasizes the importance of customer equity and provides a framework for measuring CAC.

[3] Hogan, J. E., Lemon, K. N., & Libai, B. (2003). "What is the true value of a lost customer" Journal of Service Research, 5(3), 196-208.

[4] Berger, P. D., & Nasr, N. I. (1998). "Customer lifetime value: Marketing models and applications." Journal of Interactive Marketing, 12(1), 17-30.

[5] Reinartz, W., Thomas, J. S., & Kumar, V. (2005). "Balancing acquisition and retention resources to maximize customer profitability." Journal of Marketing, 69(1), 63-79.

[6] Pfeifer, P. E., & Farris, P. W. (2004). "The elasticity of customer value to retention: The duration of a customer relationship." Journal of Interactive Marketing, 18(2), 20-31.

[7] Rust, R. T., Lemon, K. N., & Zeithaml, V. A. (2004). "Return on marketing: Using customer equity to focus marketing strategy." Journal of Marketing, 68(1), 109-127.

[8] Blattberg, R. C., Getz, G., & Thomas, J. S. (2001). "Customer Equity: Building and Managing Relationships as Valuable Assets." Harvard Business Review Press.

[9] Gupta, S., & Zeithaml, V. (2006). "Customer metrics and their impact on financial performance." Marketing Science, 25(6), 718-739.

[10] Villanueva, J., & Hanssens, D. M. (2007). "Customer equity: Measurement, management and research opportunities." Foundations and Trends® in Marketing, 1(1), 1-95.

[11] Verhoef, P. C., & Donkers, B. (2001). "Predicting customer potential value: An application in the insurance industry." Decision Support Systems, 32(2), 189-199.

[12] Thomas, J. S. (2001). "A methodology for linking customer acquisition to customer retention." Journal of Marketing Research, 38(2), 262-268.

12. APPENDIX

Model building :

1) Dataset
2) Google collab and VS code application building
    1. HTML file (Index file, Predict file)
    2. CSS file
    3. Models in pickle format

SOURCE CODE:

HOME.HTML

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Customer Acquisition Cost Estimation</title>
  <style>
    body {
      margin: 0;
      padding: 0;
      font-family: Arial, sans-serif;
      color: #333; /* Set text color to dark gray */
    }
    .hero-section {
      background: url('/static/home.jpg') no-repeat center center;
      background-size: cover;
      height: 100vh;
      display: flex;
      justify-content: center;
      align-items: center;
      color: white;
      text-align: center;
      padding: 0 20px;
    }
    .hero-section h1 {
```

```
    font-size: 3rem;
    margin-bottom: 20px;
    font-weight: bold; /* Make the heading bold */
    color: black; /* Set heading color to black */
}
.hero-section p {
    font-size: 1.5rem;
    margin-bottom: 40px;
    color: black; /* Set paragraph color to black */
}
.hero-section .btn-primary {
    font-size: 1.25rem;
    color: black; /* Set button text color to black */
    text-decoration: none;
    border: 1px solid black; /* Add a black border to the button */
    padding: 10px 20px; /* Add padding to the button */
}
.about-section, .contact-section {
    padding: 60px 0;
    background-color: #f9f9f9; /* Light gray background */
}
.about-section h2, .contact-section h2 {
    text-align: center;
    margin-bottom: 40px;
    font-weight: bold; /* Make the section headings bold */
    color: black; /* Set section heading color to black */
}
.about-section p, .contact-section ul {
    color: black; /* Set text color inside sections to black */
    text-align: justify;
    margin: 0 auto;
    max-width: 800px;
}
.contact-section ul {
```

```
            list-style-type: none;
            padding: 0;
            text-align: center;
        }
        .contact-section li {
            margin-bottom: 10px;
        }
        .container {
            max-width: 1200px;
            margin: 0 auto;
            padding: 0 20px;
        }
    </style>
</head>
<body>
    <header class="hero-section">
        <div class="container text-center">
            <h1>Welcome to Customer Acquisition Cost Estimation</h1>
            <p>Predicting the Cost of Acquiring New Customers with Machine Learning</p>
            <a href="/predict" class="btn btn-primary">Go to Prediction</a>
        </div>
    </header>
    <section id="about" class="about-section">
        <div class="container">
            <h2>About Us</h2>
            <p>Welcome to Customer Acquisition Cost Estimation! Our mission is to harness the power of machine learning to provide accurate predictions for the cost of acquiring new customers, helping businesses make informed decisions about their marketing strategies. Our platform analyzes various factors, including customer demographics, marketing channels, conversion rates, and more, to deliver comprehensive insights.</p>
            <p>We believe in the transformative power of data-driven decision making and aim to support businesses in their journey towards growth and profitability. By utilizing advanced algorithms and data-driven techniques, we strive to provide reliable and actionable predictions that can guide businesses in optimizing their marketing spend and maximizing their ROI.</p>
            <p>Our team of experts continuously works to improve our models and ensure the highest level of accuracy. We are committed to innovation and excellence, making Customer Acquisition Cost Estimation a trusted resource for businesses of all sizes.</p>
```

```
    </div>
  </section>
  <section id="contact" class="contact-section">
    <div class="container">
      <h2>Contact Us</h2>
      <ul>
        <li>Phone: +91 7780240811</li>
        <li>Email: business@domain.com</li>
        <li>Address: 123 Business Street, Your City, Your Country</li>
      </ul>
    </div>
  </section>
  <script src="https://code.jquery.com/jquery-3.3.1.slim.min.js"></script>
  <script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/1.14.7/umd/popper.min.js"></script>
  <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/js/bootstrap.min.js"></script>
</body>
</html>
```

PREDICT.HTML:

```
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <title>Predict</title>
  <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha3/dist/css/bootstrap.min.css" integrity="sha384-KK94CHFLLe+nY2dmCWGMq91rCGa5gtU4mk92HdvYe+M/SXH301p5ILy+dN9+nJOZ" crossorigin="anonymous">
  <style>
body {
      background-image: url("https://img.freepik.com/free-photo/abstract-luxury-gradient-blue-background-smooth-dark-blue-with-black-vignette-studio-banner_1258-63496.jpg");
      background-size: cover;
    }
    h3.big {
      line-height: 1.8;
```

```
    }
  </style>
</head>
<body>
  <br>
  <div class="container">
    <div class="row">
      <div class="col-md-12 bg-light text-right">
        <a href="/home" class="btn btn-info btn-lg">Home</a>
      </div>
    </div>
    <br>
    <h1><strong>Cost Estimator</strong></h1><br>
    <form action="/submit" method="post">
      <div class="form-group row">
        <div class="col-md-3">
          <label for="store_sales">store_sales</label><br>
          <input type="text" id="store_sales" name="store_sales" placeholder=""
required="required"><br>
          <label for="unit_sales">unit_sales</label><br>
          <input type="text" id="unit_sales" name="unit_sales" placeholder=""
required="required"><br>
          <label for="promotion_name">promotion_name</label><br>
          <input type="text" id="promotion_name" name="promotion_name" placeholder="select
between 1-42" required="required"><br>
          <label for="total_children">total_children</label><br>
          <input type="text" id="total_children" name="total_children" placeholder=""
required="required"><br>
          <label for="num_children_at_home">num_children_at_home</label><br>
          <input type="text" id="num_children_at_home" name="num_children_at_home"
placeholder="" required="required"><br>
          <label for="avg_cars_athome(approx).1">avg_cars_athome(approx).1</label><br>
          <input type="text" id="avg_cars_athome(approx).1" name="avg_cars_athome(approx).1"
placeholder="" required="required"><br>
          <label for="gross_weight">gross_weight</label><br>
```

```
            <input type="text" id="gross_weight" name="gross_weight" placeholder=""
required="required"><br>

            <label for="low_fat">low_fat</label><br>

            <input type="text" id="low_fat" name="low_fat" placeholder="put 0 or 1"
required="required"><br>

            <label for="recyclable_package">recyclable_package</label><br>

            <input type="text" id="recyclable_package" name="recyclable_package"
placeholder="put 0 or 1" required="required"><br>

            <label for="units_per_case">units_per_case</label><br>

            <input type="text" id="units_per_case" name="units_per_case" placeholder=""
required="required"><br>

            <label for="store_type">store_type</label><br>

            <input type="text" id="store_type" name="store_type" placeholder="select b/w 1-5"
required="required"><br>

            <label for="store_city">store_city</label><br>

            <input type="text" id="store_city" name="store_city" placeholder=""
required="required"><br>

            <label for="store_state">store_state</label><br>

            <input type="text" id="store_state" name="store_state" placeholder="select b/w 1-19"
required="required"><br>

            <label for="store_sqft">store_sqft</label><br>

            <input type="text" id="store_sqft" name="store_sqft" placeholder=""
required="required"><br>

            <label for="coffee_bar">coffee_bar</label><br>

            <input type="text" id="coffee_bar" name="coffee_bar" placeholder="put 0 or 1"
required="required"><br>

            <label for="video_store">video_store</label><br>

            <input type="text" id="video_store" name="video_store" placeholder="put 0 or 1"
required="required"><br>

            <label for="salad_bar">salad_bar</label><br>

            <input type="text" id="salad_bar" name="salad_bar" placeholder="put 0 or 1"
required="required"><br>

             <label for="prepared_food">prepared_food</label><br>

             <input type="text" id="prepared_food" name="prepared_food" placeholder="put 0 or 1"
required="required"><br>

             <label for="florist">florist</label><br>

             <input type="text" id="florist" name="florist" placeholder="put 0 or 1"
required="required"><br>

             <label for="media_type">media_type</label><br>
```

```
            <input type="text" id="media_type" name="media_type" placeholder="select b/w 1-7" required="required"><br>
            <label for="avg_cars_at_home">avg_cars_at_home</label><br>
            <input type="text" id="avg_cars_at_home" name="avg_cars_at_home" placeholder="" required="required"><br>
          </div>
        </div>
        <button type="submit" class="btn btn-success btn-lg">Submit</button>
      </form>
      <br>
      </h4>
   </div>
   <script src="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-alpha3/dist/js/bootstrap.bundle.min.js" integrity="sha384-ENjdO4Dr2bkBIFxQpeoTz1HIcje39Wm4jDKdf19U8gI4ddQ3GYNS7NTKfAdVQSZe" crossorigin="anonymous"></script>
</body>
</html>
```

SUBMIT.HTML:

```
<!DOCTYPE html>
<html lang="en">
<head>
   <meta charset="UTF-8">
   <title>Output</title>
   <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
   <style>
body {
        background-image: url("static/submit.jpg");
        background-size: cover;
      }
      h1.big {
        font-size: 50px;
        font-weight: bold;
        text-align: center;
      }
```

```
        h3 {
            font-size: 30px;
            line-height: 1.8;
            text-align: center;
        }
    </style>
</head>
<body>
    <br>
    <div class="container">
        <div class="row">
            <div class="col-md-12 bg-light text-right">
                <a href="/home" class="btn btn-info btn-lg">Home</a>
                <a href="/predict" class="btn btn-primary btn-lg">Predict</a>
            </div>
        </div>
        <br>
        <h1 class="big"><strong>Estimated cost is: </strong></h1><br>
        <h3> {{prediction_text}}.</h3>
    </div>
</body>
</html>
```

APP.PY:

```
from flask import Flask, render_template, request
import numpy as np
import pickle
# Load the model
with open('proj_2.pkl', 'rb') as file:
    model = pickle.load(file)
app = Flask(__name__)
@app.route('/')
def home():
    return render_template('home.html')
@app.route('/predict')
```

```
def predict():
    return render_template('predict.html')
@app.route('/submit', methods=['POST'])
def submit():
    # Get the numerical inputs from the form
    store_sales =float( request.form['store_sales'])
    unit_sales = float(request.form['unit_sales'])
    promotion_name = float(request.form['promotion_name'])
    total_children = float(request.form['total_children'])
    num_children_at_home =float( request.form['num_children_at_home'])
    gross_weight = float(request.form['gross_weight'])
    low_fat =float( request.form['low_fat'])
    recyclable_package = float(request.form['recyclable_package'])
    units_per_case = float(request.form['units_per_case'])
    store_type = float(request.form['store_type'])
    store_city = float(request.form['store_city'])
    store_state = float(request.form['store_state'])
    store_sqft = float(request.form['store_sqft'])
    coffee_bar = float(request.form['coffee_bar'])
    video_store = float(request.form['video_store'])
    salad_bar = float(request.form['salad_bar'])
    prepared_food = float(request.form['prepared_food'])
    florist = float(request.form['florist'])
    media_type = float(request.form['media_type'])
    avg_cars_at_home = float(request.form['avg_cars_at_home']  # Create the final features array
    final_features =
np.array([[avg_cars_at_home,media_type,florist,prepared_food,salad_bar,video_store,coffee_bar,stor
e_sqft,store_state,store_city,store_type,units_per_case,recyclable_package,low_fat,gross_weight,num
_children_at_home,total_children,promotion_name,unit_sales,store_sales]])
    # Make the prediction
      prediction = model.predict(final_features)[0]
    # Render the result template with the prediction text
    return render_template('submit.html', prediction_text=prediction)
if __name__ == '__main__':
    app.run(debug=True)
```

CODE SNIPPETS

MODEL BUILDING

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
dataset=pd.read_csv("/content/media prediction and its cost.csv")
dataset.head()
```

	food_category	food_department	food_family	store_sales(in millions)	store_cost(in millions)	unit_sales(in millions)	promotion_name	sales_country
0	Breakfast Foods	Frozen Foods	Food	7.36	2.7232	4.0	Bag Stuffers	USA
1	Breakfast Foods	Frozen Foods	Food	5.52	2.5944	3.0	Cash Register Lottery	USA
2	Breakfast Foods	Frozen Foods	Food	3.68	1.3616	2.0	High Roller Savings	USA

food_category	food_department	food_family	store_sales(in millions)	store_cost(in millions)	unit_sales(in millions)	promotion_name	sales_country
Breakfast Foods	Frozen Foods	Food	7.36	2.7232	4.0	Bag Stuffers	USA
Breakfast Foods	Frozen Foods	Food	5.52	2.5944	3.0	Cash Register Lottery	USA
Breakfast Foods	Frozen Foods	Food	3.68	1.3616	2.0	High Roller Savings	USA
Breakfast Foods	Frozen Foods	Food	3.68	1.1776	2.0	Cash Register Lottery	USA
Breakfast Foods	Frozen Foods	Food	4.08	1.4280	3.0	Double Down Sale	USA

```
dataset.info()
dataset.describe()
```

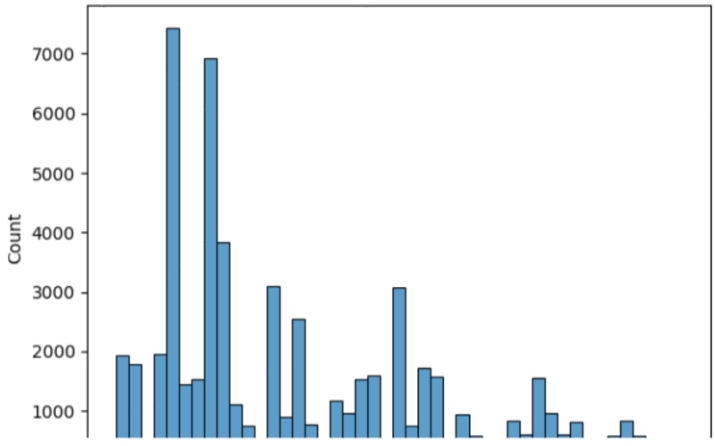
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60428 entries, 0 to 60427
Data columns (total 40 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   food_category             60428 non-null  object
 1   food_department           60428 non-null  object
 2   food_family               60428 non-null  object
 3   store_sales(in millions)  60428 non-null  float64
 4   store_cost(in millions)   60428 non-null  float64
 5   unit_sales(in millions)   60428 non-null  float64
 6   promotion_name            60428 non-null  object
 7   sales_country             60428 non-null  object
 8   marital_status            60428 non-null  object
 9   gender                    60428 non-null  object
 10  total_children            60428 non-null  float64
 11  education                 60428 non-null  object
```

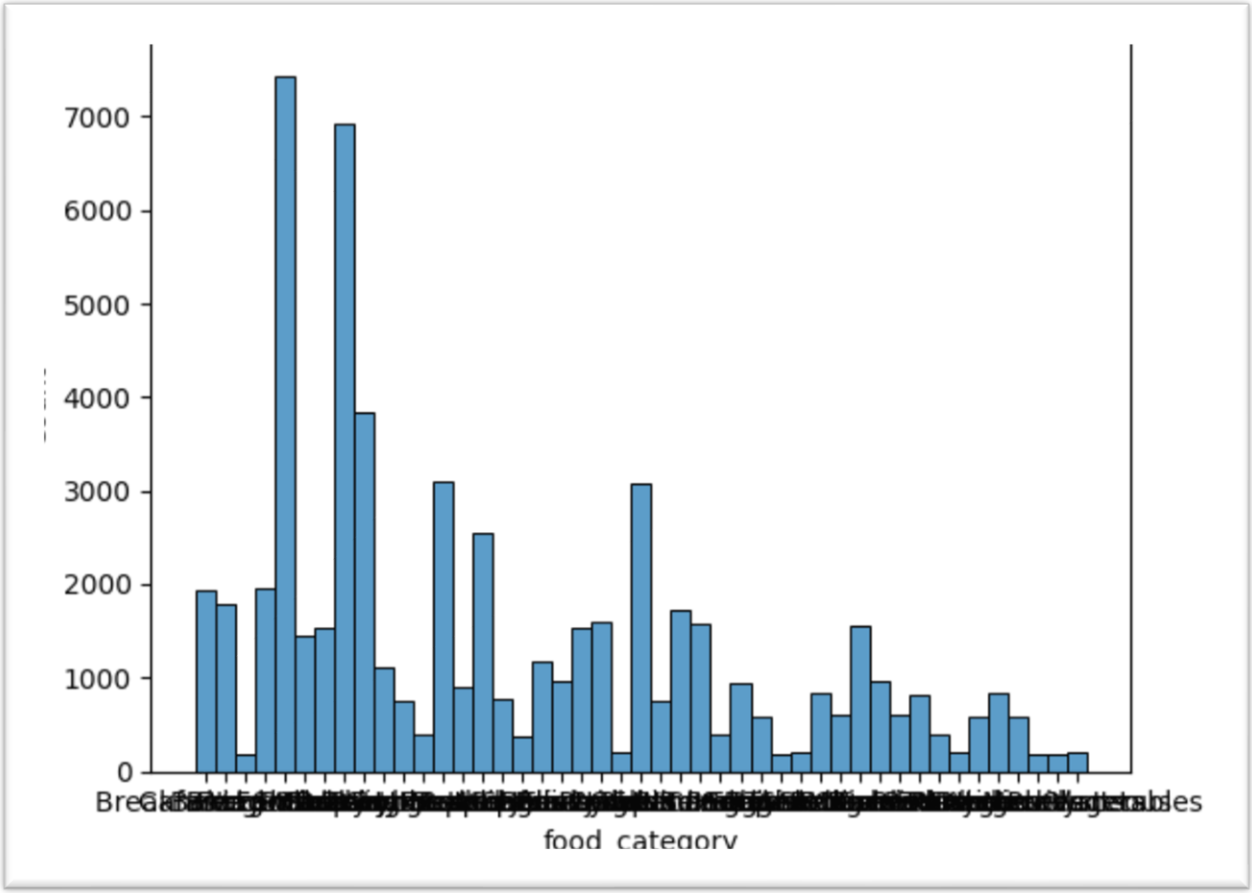
```
RangeIndex: 60428 entries, 0 to 60427
Data columns (total 40 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   food_category               60428 non-null  object
 1   food_department             60428 non-null  object
 2   food_family                 60428 non-null  object
 3   store_sales(in millions)    60428 non-null  float64
 4   store_cost(in millions)     60428 non-null  float64
 5   unit_sales(in millions)     60428 non-null  float64
 6   promotion_name              60428 non-null  object
 7   sales_country               60428 non-null  object
 8   marital_status              60428 non-null  object
 9   gender                      60428 non-null  object
 10  total_children              60428 non-null  float64
 11  education                   60428 non-null  object
 12  member_card                 60428 non-null  object
 13  occupation                  60428 non-null  object
 14  houseowner                  60428 non-null  object
 15  avg_cars_at home(approx)    60428 non-null  float64
```

```
#Visualizing Categorical Values
for col in dataset.select_dtypes('object').columns:
    print(col+" :")
    print(dataset[col].unique())
    sns.histplot(dataset[col])
    plt.show()
    print("*"*40)
```

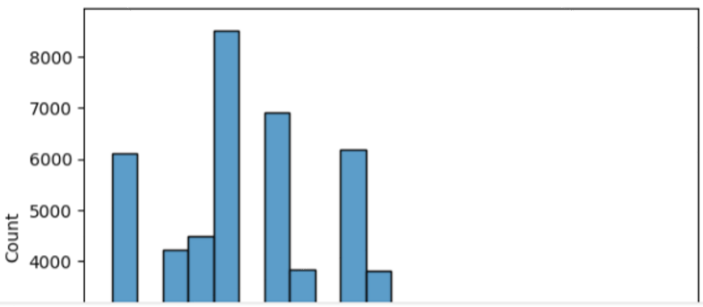
```
food_category :
['Breakfast Foods' 'Bread' 'Canned Shrimp' 'Baking Goods' 'Vegetables'
 'Frozen Desserts' 'Candy' 'Snack Foods' 'Dairy' 'Starchy Foods'
 'Cleaning Supplies' 'Decongestants' 'Meat' 'Hot Beverages'
 'Jams and Jellies' 'Carbonated Beverages' 'Seafood' 'Specialty'
 'Kitchen Products' 'Electrical' 'Beer and Wine' 'Candles' 'Fruit'
 'Pure Juice Beverages' 'Canned Soup' 'Paper Products' 'Canned Tuna'
 'Eggs' 'Hardware' 'Canned Sardines' 'Canned Clams' 'Pain Relievers'
 'Side Dishes' 'Bathroom Products' 'Magazines' 'Frozen Entrees' 'Pizza'
 'Cold Remedies' 'Canned Anchovies' 'Drinks' 'Hygiene' 'Plastic Products'
 'Canned Oysters' 'Packaged Vegetables' 'Miscellaneous']
```


```
'Eggs' 'Hardware' 'Canned Sardines' 'Canned Clams' 'Pain Relievers'
'Side Dishes' 'Bathroom Products' 'Magazines' 'Frozen Entrees' 'Pizza'
'Cold Remedies' 'Canned Anchovies' 'Drinks' 'Hygiene' 'Plastic Products'
'Canned Oysters' 'Packaged Vegetables' 'Miscellaneous']
```

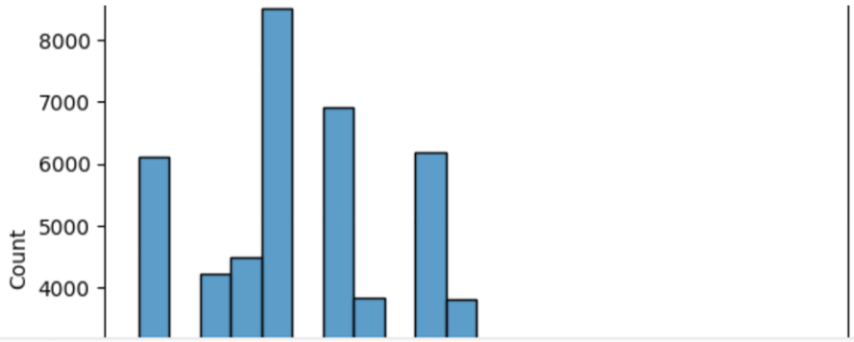





```
****************************************
food_department :
['Frozen Foods' 'Baked Goods' 'Canned Foods' 'Baking Goods' 'Produce'
 'Snacks' 'Snack Foods' 'Dairy' 'Starchy Foods' 'Household'
 'Health and Hygiene' 'Meat' 'Beverages' 'Seafood' 'Deli'
 'Alcoholic Beverages' 'Canned Products' 'Eggs' 'Periodicals'
 'Breakfast Foods' 'Checkout' 'Carousel']
```



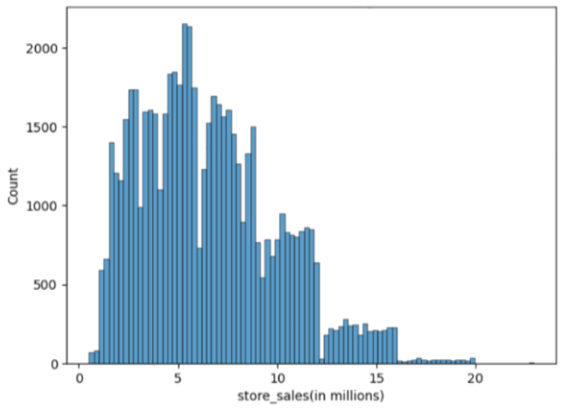

```
for col in dataset.select_dtypes('float').columns:
```

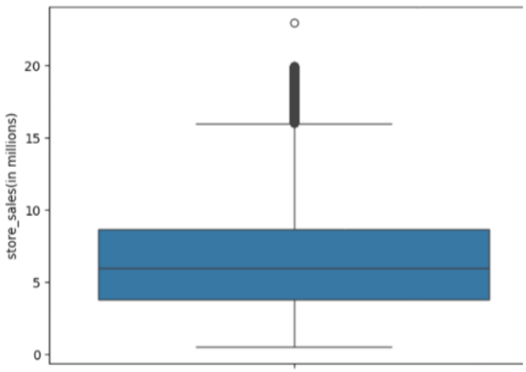


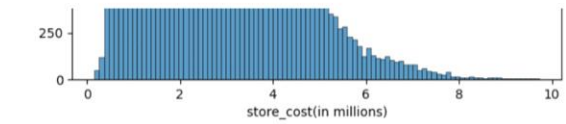

```
for col in dataset.select_dtypes('float').columns:
    print(col+" :")
```

```
    print("*"*40)
```

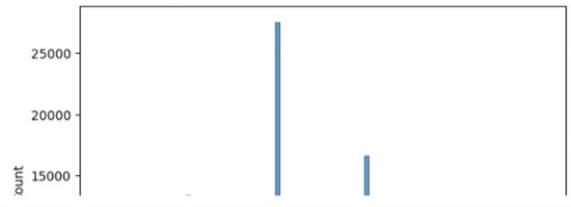
```
store_sales(in millions) :
[ 7.36  5.52  3.68 ... 16.1   0.58  1.21]
```

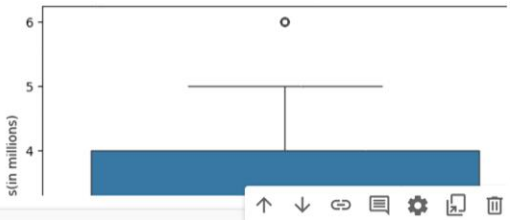







```
****************************************
unit_sales(in millions) :
[4. 3. 2. 5. 6. 1.]
```






```
#test to find dependency between catogorical values
import scipy.stats as stats
dfc=dataset.copy()
```

```
#test to find dependency between catogorical values
import scipy.stats as stats
dfc=dataset.copy()
```

```
for coli in dfc.select_dtypes('object').columns:
    for colj in dfc.select_dtypes('object').columns[1:]:
        cont = dataset[[coli, colj]].pivot_table(index=coli, columns=colj, aggfunc=len).fillna(0).copy().astype(int)
        st_chi2, st_p, st_dof, st_exp = stats.chi2_contingency(cont)
        if st_p<0.05:
            print("pvalue :"+coli+" and "+colj)
            print(st_p)
            print('*'*40)
    dfc.drop(coli,axis=1,inplace=True)
```

```
pvalue :food_category and food_department
0.0
****************************************
pvalue :food_category and food_family
0.0
****************************************
```

```
dfc.drop(coli,axis=1,inplace=True)
```

```
pvalue :food_category and food_department
0.0
****************************************
pvalue :food_category and food_family
0.0
****************************************
pvalue :food_category and brand_name
0.0
****************************************
pvalue :food_department and food_family
0.0
****************************************
pvalue :food_department and gender
0.0012661937802332355
****************************************
pvalue :food_department and brand_name
0.0
****************************************
pvalue :food_family and marital_status
0.0033804632648580728
****************************************
pvalue :food_family and brand_name
```

```
**************************************
value :promotion_name and occupation
.54391650593465e-163
**************************************
value :promotion_name and houseowner
.347018140624352e-30
**************************************
value :promotion_name and avg. yearly_income
.505949037826803e-264
**************************************
value :promotion_name and store_type
.0
**************************************
value :promotion_name and store_city
.0
**************************************
value :promotion_name and store_state
.0
**************************************
value :promotion_name and media_type
```

```
Dropping reductant columns
```

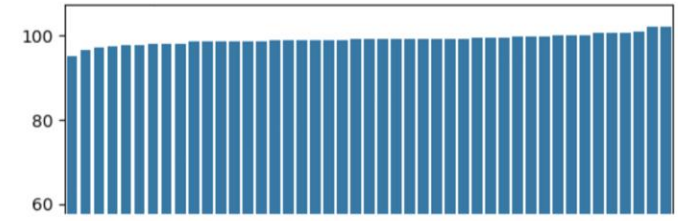
```
#Dropping reductant columns
dataset.drop(["frozen_sqft","store_cost(in millions)","avg_cars_at home(approx)","SRP","net_weight","grocery_sqft","mea
```

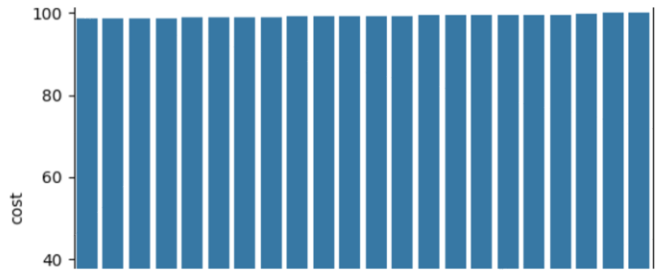
```
[ ]  #visualize categorical features in terms of target feature "cost"
     for col in dataset.select_dtypes('object').columns:
         print(col+" :")
         print(dataset[col].unique())
         # Calculate the mean cost for each category, handling non-numeric values
         category_means = dataset.groupby(col)['cost'].mean().sort_values().reset_index()
         sns.barplot(data=category_means, x=col, y="cost")
         plt.show()
         print("*"*40)
```

```
food_category :
['Breakfast Foods' 'Bread' 'Canned Shrimp' 'Baking Goods' 'Vegetables'
 'Frozen Desserts' 'Candy' 'Snack Foods' 'Dairy' 'Starchy Foods'
 'Cleaning Supplies' 'Decongestants' 'Meat' 'Hot Beverages'
 'Jams and Jellies' 'Carbonated Beverages' 'Seafood' 'Specialty'
 'Kitchen Products' 'Electrical' 'Beer and Wine' 'Candles' 'Fruit'
 'Pure Juice Beverages' 'Canned Soup' 'Paper Products' 'Canned Tuna'
 'Eggs' 'Hardware' 'Canned Sardines' 'Canned Clams' 'Pain Relievers'
 'Side Dishes' 'Bathroom Products' 'Magazines' 'Frozen Entrees' 'Pizza'
```

```
    print("*"*40)
```

```
food_category :
['Breakfast Foods' 'Bread' 'Canned Shrimp' 'Baking Goods' 'Vegetables'
 'Frozen Desserts' 'Candy' 'Snack Foods' 'Dairy' 'Starchy Foods'
 'Cleaning Supplies' 'Decongestants' 'Meat' 'Hot Beverages'
 'Jams and Jellies' 'Carbonated Beverages' 'Seafood' 'Specialty'
 'Kitchen Products' 'Electrical' 'Beer and Wine' 'Candles' 'Fruit'
 'Pure Juice Beverages' 'Canned Soup' 'Paper Products' 'Canned Tuna'
 'Eggs' 'Hardware' 'Canned Sardines' 'Canned Clams' 'Pain Relievers'
 'Side Dishes' 'Bathroom Products' 'Magazines' 'Frozen Entrees' 'Pizza'
 'Cold Remedies' 'Canned Anchovies' 'Drinks' 'Hygiene' 'Plastic Products'
 'Canned Oysters' 'Packaged Vegetables' 'Miscellaneous']
```






```
#Dropping Reductant columns further analyzing with bar plots
dataset.drop(["food_category","food_department","gender","food_family","marital_status","brand_name","avg. yearly_income","e
```

```
#Dropping Reductant columns further analyzing with bar plots
dataset.drop(["food_category","food_department","gender","food_family","marital_status","brand_name","avg. yearly_income","e
```

```
dataset.head()
```

	store_sales(in millions)	unit_sales(in millions)	promotion_name	total_children	num_children_at_home	avg_cars_at home(approx).1	gross_weight	recycla
0	7.36	4.0	Bag Stuffers	1.0	1.0	1.0	19.70	
1	5.52	3.0	Cash Register Lottery	0.0	0.0	4.0	19.70	
2	3.68	2.0	High Roller Savings	4.0	0.0	1.0	19.70	
3	3.68	2.0	Cash Register Lottery	2.0	2.0	2.0	19.70	
4	4.08	3.0	Double Down Sale	0.0	0.0	2.0	7.12	

```
from sklearn.preprocessing import OrdinalEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import  metrics
encoder = OrdinalEncoder()
final_dataframe = encoder.fit_transform(dataset.drop(columns='cost'))
```

```
X=final_dataframe
y=dataset['cost']
```

```
train_X,val_X,train_y,val_y=train_test_split(X,y,test_size=0.2)
```

```
lr = LinearRegression()
lr.fit(train_X,train_y)
print('Attempting to fit Linear Regressor')
```

```
Attempting to fit Linear Regressor
```

```
y_pred_val_lr = lr.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y, y_pred_val_lr))
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y, y_pred_val_lr))
print("\n")
print('RMSE on Validation set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pred_val_lr)))
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_lr))
print("\n")
```

```
MAE on Validation set : 25.44085638565691


MSE on Validation set : 877.238339706685
```

```
MAE on Validation set : 25.44085638565691


MSE on Validation set : 877.238339706685


RMSE on Validation set : 5.043892979203356


R2 Score on Validation set : 0.026346284499798478
```

```
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn import  metrics
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor()
rf.fit(train_X,train_y)
print('Attempting to fit Random Forest Regressor')
```

```
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import train_test_split
from sklearn import  metrics
from sklearn.ensemble import RandomForestRegressor
rf = RandomForestRegressor()
rf.fit(train_X,train_y)
print('Attempting to fit Random Forest Regressor')
```

```
Attempting to fit Random Forest Regressor
```

```
y_pred_val_rf = rf.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y, y_pred_val_rf))
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y, y_pred_val_rf))
print("\n")
print('RMSE on Validation set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pred_val_rf)))
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_rf))
print("\n")
```

```
MAE on Validation set : 0.08540060400479205
```

```
y_pred_val_rf = rf.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y, y_pred_val_rf))
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y, y_pred_val_rf))
print("\n")
print('RMSE on Validation set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pred_val_rf)))
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y, y_pred_val_rf))
print("\n")
```

```
MAE on Validation set : 0.08540060400479205

MSE on Validation set : 1.7156286520544382

RMSE on Validation set : 0.29223381735314624

R2 Score on Validation set : 0.9980958102993425
```

```
dataset.head()
```

	store_sales(in millions)	unit_sales(in millions)	promotion_name	total_children	num_children_at_home	avg_cars_at home(approx).1	gross_weight	recyclab
0	7.36	4.0	Bag Stuffers	1.0	1.0	1.0	19.70	
1	5.52	3.0	Cash Register Lottery	0.0	0.0	4.0	19.70	
2	3.68	2.0	High Roller Savings	4.0	0.0	1.0	19.70	
3	3.68	2.0	Cash Register Lottery	2.0	2.0	2.0	19.70	
4	4.08	3.0	Double Down Sale	0.0	0.0	2.0	7.12	

4	4.08	3.0	Double Down Sale	0.0	0.0	2.0	7.12

5 rows × 21 columns

```
Start coding or generate with AI.
```

```
#testing with random value
rf.predict([[20,57,0,400,12306,18520,939,735,0,0,0,0,1,1,1,1,1,1,1,2]])
```

```
array([121.4129])
```

+ Code + Text

```
# saving the model

import pickle
pickle.dump(rf,open('proj_2.pkl','wb'))
```

```
Start coding or generate with AI
```