UNDERSTANDING AUDIENCE: A MACHINE LEARNING APPROACH TO CUSTOMER SEGMENTATION

1. INTRODUCTION

1.1 OVERVIEW
These days, you can personalize everything. There's no one-size-fits-all approach. But, for business, this is actually a great thing. It creates a lot of space for healthy competition and opportunities for companies to get creative about how they acquire and retain customers.

One of the fundamental steps towards better personalization is customer segmentation. This is where personalization starts, and proper segmentation will help you make decisions regarding new features, new products, pricing, marketing strategies, even things like in-app recommendations.

But, doing segmentation manually can be exhausting. Why not employ machine learning to do it for us? In this article, I'll tell you how to do just that.

1.2 Purpose
The goal of this project is to decide how to relate to customers in each segment in order to predict the financial status of customers. We can identify the most active users/customers, and optimize your application/offer towards their needs.

1. LITERATURE SURVEY

2.1 EXISTING SYSTEM
The problem of existing system is that it is not able to predict the financial status of a customer. And firm will not be able to develop or innovate product according to customer desire. So that firm will not be able to target right customer with right product.

2.2 REFERENCES
https://www.daitm.org.in/wp-content/uploads/2019/04/15499016029_Abhijit-Bag.pdf

2.3 PROBLEM STATEMENT DEFINITION
The problem statement aims to explore and implement a machine learning approach for customer segmentation with a focus on understanding audience behaviour, ultimately enhancing targeted marketing strategies and improving overall customer engagement.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas
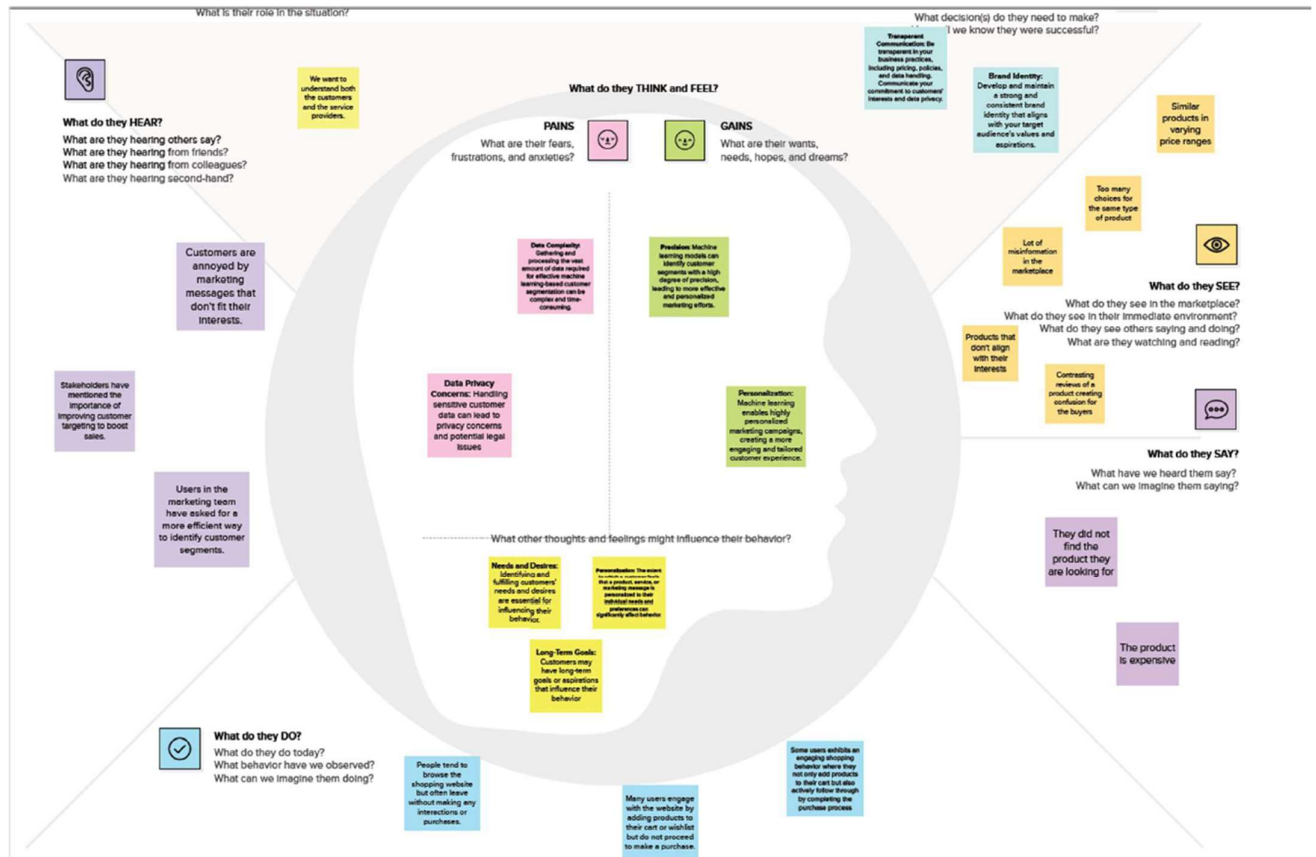
3.2 Ideation & Brainstorming

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- 10 minutes to prepare
- 1 hour to collaborate
- 2-8 people recommended

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

⏱ 10 minutes

A. Team gathering
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

B. Set the goal
Think about the problem you'll be focusing on solving in the brainstorming session.

C. Learn how to use the facilitation tools
Use the Facilitation Superpowers to run a happy and productive session.

Open article →

1 Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

⏱ 5 minutes

> **PROBLEM**
> How might we [your problem statement]?

Key rules of brainstorming
To run an smooth and productive session

- Stay in topic.
- Defer judgment.
- Go for volume.
- Encourage wild ideas.
- Listen to others.
- If possible, be visual.

Adwaith M Nambiar

Customer Lifetime Value (CLV) Prediction: Segment customers based on their predicted lifetime value. This can help businesses identify and prioritize high-value customers for special treatment.

Retail Customer Segmentation: Implement customer segmentation for a retail business. Use purchase history, frequency of visits, and spending patterns to categorize customers into segments. This can help the business target different customer groups more effectively.

Subscription Services: For subscription-based businesses, segment customers to understand churn and retention patterns. This information can be used to implement strategies to reduce churn and improve customer satisfaction.

Manisha Rout

Recommendation Systems: Build recommendation engines that use machine learning to suggest products or content to customers based on their behavior and preferences, further refining customer segmentation.

Predictive Customer Segmentation: Develop machine learning models that use historical data to predict how customers will behave in the future. This can help businesses tailor their marketing strategies to each segment's specific needs.

Real-time Audience Segmentation: Implement a machine learning system that can segment customers in real-time as they interact with a website or app. This can lead to personalized content.

Shrey Gupta

Attribute segmentation Implement customer segmentation based on attributes like Demography (age, gender, location), Behavior (purchase history, website interaction) and Psychography (lifestyle, values, interests).

Prduct based profiling: Use data such as number of views, clicks, searches and purchases to identify which product is in demand, during which time period, and correlate it with the customer profile to have a better understanding of the market.

NLP for text analysis: Analyze customer reviews, feedback or social media mentions using NLP techniques to understand sentiment and extract insights that can be used for segmentation and product development

Aditya Ghosh

E-commerce Personalization: Use K-means clustering to group e-commerce customers based on their behavior, enabling tailored recommendations and marketing.

Financial Services Customer Profiling: Use machine learning algorithms to profile financial customers for personalized product and service offerings.

Telecom Churn Prediction: Apply machine learning algorithms to identify customer segments for predicting and reducing telecom churn

---

Customer Lifetime Value (CLV) Prediction: Segment customers based on their predicted lifetime value. This can help businesses identify and prioritize high-value customers for special treatment.
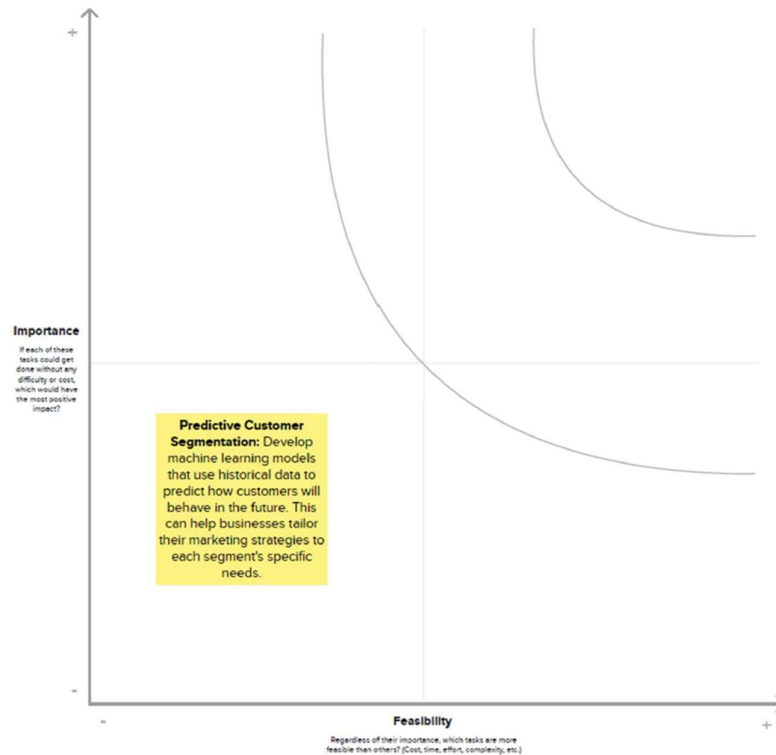
Predictive Customer Segmentation: Develop machine learning models that use historical data to predict how customers will behave in the future. This can help businesses tailor their marketing strategies to each segment's specific needs.

E-commerce Personalization: Use K-means clustering to group e-commerce customers based on their behavior, enabling tailored recommendations and marketing.

Attribute segmentation Implement customer segmentation based on attributes like Demography (age, gender, location), Behaviour (purchase history, website interaction) and Psychography (lifestyle, values, interests).

Importance

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

Predictive Customer Segmentation: Develop machine learning models that use historical data to predict how customers will behave in the future. This can help businesses tailor their marketing strategies to each segment's specific needs.

Feasibility

Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

4. REQUIREMENT ANALYSIS

4.1   Functional requirement

1.   Data Collection: Gather diverse and relevant customer data, including demographics, purchase history, and online behaviour.

2.   Feature Engineering: Identify key features for machine learning models, considering factors such as customer preferences, geographic location, and interaction frequency.

3.   Model Selection: Choose appropriate machine learning algorithms for segmentation, such as clustering techniques (k-means, hierarchical clustering) to group customers based on similarities.

4. Training and Validation: Train the model on historical data, and validate its performance using a separate dataset to ensure accuracy and reliability.

5.   Performance Monitoring: Establish mechanisms for ongoing monitoring of the model's performance, with the ability to retrain or update the model as needed.

4.2   Non-Functional requirements

1. Performance: The system should provide efficient and timely customer segmentation results, with minimal latency, even as the dataset grows.

2. Scalability: The solution should be scalable to handle an increasing number of customers and data points without a significant decrease in performance.

3. Reliability: Ensure high availability and reliability of the system, minimizing downtime and disruptions to customer segmentation processes.

4. Security: Implement robust security measures to safeguard customer data, including encryption, access controls, and compliance with data protection regulations.

5. Compatibility: Ensure compatibility with various data sources and formats to facilitate seamless integration with existing systems and tools.

6. PROJECT DESIGN

6.1   Data Flow Diagrams & User Stories
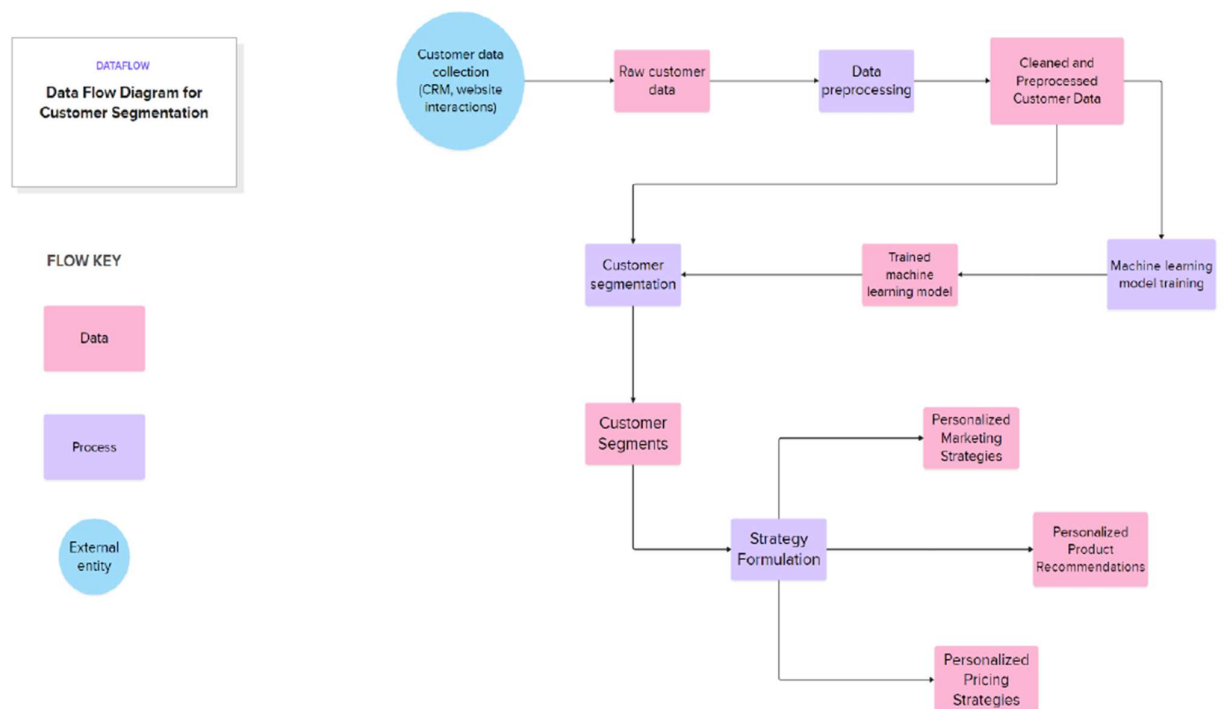
6.2 Solution Architecture

1. Data Ingestion Layer:
   - Responsibility: Ingests data from various sources like surveys, transactions, and social media.

2. Data Storage and Processing Layer:
   - Responsibility: Stores and processes raw and pre-processed data for analysis.
   - Technologies: BigQuery, Apache Spark for data processing.

3. Feature Engineering and ML Model Layer:
   - Responsibility: Extracts meaningful features and applies machine learning models for segmentation.
   - Technologies: Pandas for feature engineering, scikit-learn for ML models.
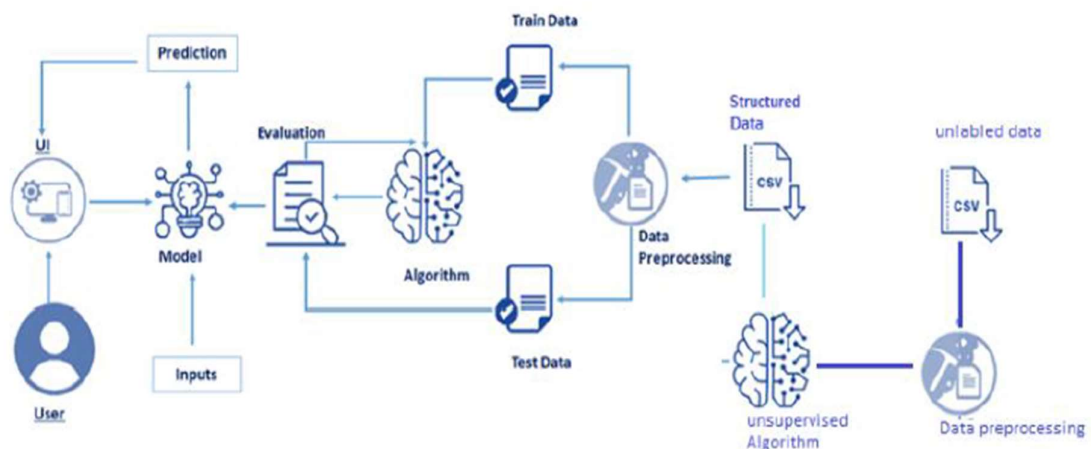
4. Real-time Processing Layer:
   - Responsibility: Enables real-time adaptation to customer behaviour changes.

5. User Interface Layer:
   - Responsibility: Provides a user-friendly interface for business users to interact with segmented data.
   - Technologies: React for front-end development.

7. PROJECT PLANNING & SCHEDULING

7.1 Technical Architecture

6.2 Sprint Planning & Estimation

Sprint 1: Data Preparation
- Define data sources and collection methods (1hr).
- Develop data cleaning and preprocessing scripts (3hrs).

Sprint 2: Feature Engineering
- Identify key features for customer segmentation (2hrs).
- Implement feature engineering techniques to enhance model input (4hrs).

Sprint 3: Model Selection and Training
- Research and choose appropriate machine learning algorithms for customer segmentation (2hrs).
- Implement the selected algorithms and train the model on historical data (5hrs).

Sprint 4: Real-time Processing
- Integrate real-time processing capabilities for dynamic customer behaviour adaptation (4hrs).
- Test and validate real-time processing functionality (2hrs).

Sprint 5: User Interface
- Design and implement a user-friendly interface for business users (5hrs).
- Conduct usability testing and gather feedback for improvements (3hrs).

Sprint 6: Scalability and Performance Optimization
- Implement scalability measures and optimize system performance (4hrs).
- Conduct performance testing and make necessary adjustments (3hrs).

Sprint 7: Security and Compliance
- Implement security measures and ensure compliance with data protection regulations (5hrs).
- Conduct security audits and address any vulnerabilities (2hrs).

Sprint 8: Monitoring and Maintenance
- Set up monitoring mechanisms for model performance (2hrs).
- Develop processes for ongoing maintenance and updates (4hrs).

Estimation:

Based on a standard 1-week sprint:

- Total number of sprints: 8
- Estimated duration: 8 weeks (8 sprints * 1 weeks/sprint)

6.3 Sprint Delivery Schedule

Sprint 1 (Week 1): Data Preparation
- Deliverable: Data collection methods, initial database structure, and data cleaning scripts.

Sprint 2 (Week 2): Feature Engineering
- Deliverable: Identified key features and implemented feature engineering techniques.

Sprint 3 (Week 3): Model Selection and Training
- Deliverable: Chosen ML algorithms, implemented models, and initial training on historical data.

Sprint 4 (Week 4): Real-time Processing
- Deliverable: Integrated real-time processing capabilities and validated functionality.

Sprint 5 (Week 5): User Interface
- Deliverable: User-friendly interface design and initial implementation.

Sprint 6 (Week 6): Scalability and Performance Optimization
- Deliverable: Implemented scalability measures and optimized system performance.

Sprint 7 (Week 7): Security and Compliance
- Deliverable: Implemented security measures, ensured compliance, and addressed vulnerabilities.

Sprint 8 (Week 8): Monitoring and Maintenance
- Deliverable: Set up monitoring mechanisms, established maintenance processes, and conducted final testing.

7. CODING & SOLUTIONING

7.1   Feature 1

Data Handling and Preprocessing:

Feature Engineering: Develop code to extract relevant features from the raw data, potentially using domain knowledge to enhance the dataset.

Data Cleaning: Implement routines to handle missing or inconsistent data, ensuring the dataset is suitable for training.

7.2   Feature 2

Machine Learning Algorithms:

Algorithm Implementation: Code the machine learning algorithms used for customer segmentation, such as k-means clustering or neural networks.

Model Training: Develop functions to train the models on the prepared dataset.

8. PERFORMANCE TESTING

8.1 Performance Metrics

MAE:

```python
from sklearn.metrics import mean_absolute_error

print ('MAE =',mean_absolute_error(y_test, y_pred))

MAE = 0.0
```

MSE:

```python
from sklearn.metrics import mean_squared_error

print ('MSE =',mean_squared_error(y_test,y_pred ))

MSE = 0.0
```

R2 SCORE:

```
from sklearn.metrics import r2_score

print ('R Squared =',r2_score(y_test,y_pred ))

R Squared = 1.0
```

RMSE:

```
from sklearn.metrics import mean_squared_error
import math

RMSE = math.sqrt(MSE)
print("Root Mean Square Error:\n")
print(RMSE)

Root Mean Square Error:

0.0
```
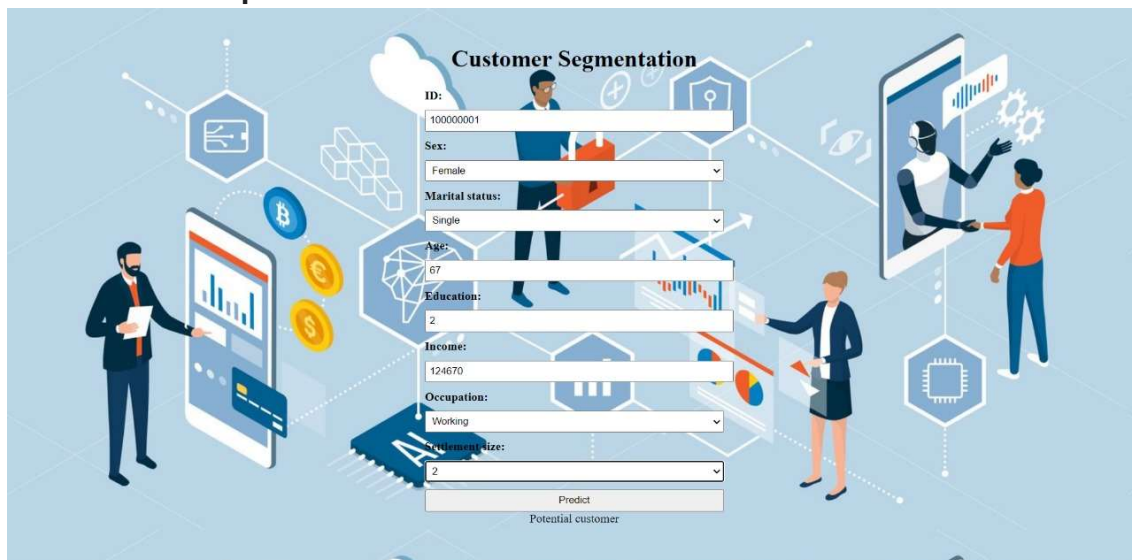
9. RESULTS

9.1 Output Screenshots

10. ADVANTAGES & DISADVANTAGES

Advantages:

1. Enhanced Customer Understanding:
   - The model provides a deeper understanding of customer behaviour, preferences, and trends, enabling more targeted and effective marketing strategies.

2. Real-time Adaptation:
   - Real-time processing allows the system to adapt dynamically to changes in customer behaviour, ensuring the segmentation remains relevant and up-to-date.

3. Scalability:
   - The solution is designed with scalability in mind, capable of handling growing datasets and increased computational demands as the customer base expands.

4. User-Friendly Interface:
   - The user interface facilitates easy interaction for business users, allowing them to explore customer segments intuitively and make informed decisions.

5. Automated Monitoring and Maintenance:
   - Automated scripts monitor key performance indicators and handle routine maintenance tasks, reducing the need for manual intervention and ensuring system reliability.

Disadvantages:

1. Complex Implementation:
   - The model's complexity may require a skilled team for development and maintenance, potentially posing challenges for smaller organizations with limited resources.

2. Data Quality Dependencies:
   - The effectiveness of the model relies heavily on the quality of input data. Inaccurate or incomplete data may lead to skewed segmentation results.

3. Initial Setup Time:
   - Setting up the entire solution, including data pipelines, feature engineering, and model training, may require a significant initial time investment before delivering tangible results.

4. Cost Considerations:
   - Utilizing advanced technologies like Apache Flink and BigQuery may involve higher operational costs, which could be a concern for budget-conscious organizations.

5. Dependency on External Tools:
   - Integration with external tools such as Apache Kafka and marketing platforms introduces dependencies that require careful management and may impact overall system stability.

11. CONCLUSION

While this guide provides a step-by-step process for identifying, prioritizing, and targeting your best current customer segments, simply following it does not guarantee success. To be effective, you must prepare and plan for the various challenges and hurdles that each step may present, and always make sure to adapt your process to any new information or feedback that might change its output. Additionally, you cannot force feed this process on your business. If the key stakeholders that will be impacted by the best current customers segmentation process do not fully buy-in, then the outputs produced from it will be relatively meaningless.

12. FUTURE SCOPE

If you properly manage the best current customer segmentation process, however, the impact it can have on every part of your organization — sales, marketing, product development, customer service, etc. — is immense. Your business will possess stronger customer focus and market clarity, allowing it to scale in a far more predictable and efficient manner. Ultimately, that means no longer needing to take on every customer that is willing to pay for your product or service, which will allow you to instead hone in on a specific subset of customers that present the most profitable opportunities and efficient use of resources. That is critical for every business, of course, but at the expansion stage, it can often be the difference between incredible success and certain failure.