Low Level Design (LLD)

Customer Personality Analysis

Revision Number: 1
Last date of revision: 31/01/2024

Vishnu M

● **Document Version Control**

Date Issued	Version	Description	Author
31 Jan 2024	1.1	Complete Draft	Vishnu M

Contents

Abstract

Understanding customer personalities is key to tailoring products and marketing for specific needs and behaviors. This project explores customer personality analysis through two lenses: (1) attitudinal analysis, examining what customers say about products through reviews, social media, and surveys, and (2) behavioral analysis, investigating what customers do via purchase history, website interactions, and app usage.

By combining these insights, we aim to:

- Identify distinct customer personalities: Uncover distinct customer segments based on shared characteristics and behaviors.

- Predict product success: Anticipate the success of new products for different customer groups by understanding their personality-product preferences.

- Optimize marketing campaigns: Align marketing messages and channels with identified customer personalities for more relevant and effective engagement.

1.Introduction

1.1 Why this Low-Level Design Document?

a low-level document delves deeper into the specifics. Here's why a low-level document is vital for your customer personality analysis project:

1. Detailed Methodology:

- A low-level document meticulously outlines your research methodologies for both attitudinal and behavioral analysis. This includes specific data sources, data collection methods, analysis techniques (e.g., text analysis, machine learning algorithms), and evaluation metrics.

- It provides transparency and allows others to replicate your research or build upon it.

2. Technical Implementation:

- If your project involves building computational models or software tools, a low-level document details the technical implementation. This includes chosen programming languages, libraries, frameworks, algorithms, and model parameters.

- It ensures reproducibility and facilitates future modification or maintenance of the developed tools.

3. Granular Findings and Insights:

- Beyond high-level conclusions, a low-level document dives deep into your findings. This includes presenting descriptive statistics, visualizations, and detailed results of your analysis for each customer segment.

- It offers evidence for your conclusions and allows others to scrutinize your reasoning and methodology.

4. Replication and Expansion:

- A detailed document fosters future research and allows others to replicate your work or build upon it. This can be valuable for further validation or extending your analysis to different contexts or data sources.

- It creates a knowledge base for future projects addressing similar questions about customer personality analysis.

5. Project Management and Collaboration:

- For complex projects with multiple team members, a low-level document serves as a single source of truth. It ensures team alignment, facilitates knowledge sharing, and simplifies project management.

- It streamlines collaboration and prevents misunderstandings or misinterpretations during project execution.

- Improved comfort and productivity: Energy-efficient buildings can offer better thermal comfort and lighting, leading to increased occupant satisfaction and productivity.

- Enhanced industrial processes: Efficient industrial practices can reduce energy waste and improve product quality while optimizing resource utilization.

- Technological advancement: Innovation in energy-efficient technologies drives development of cleaner and more sustainable energy solutions.

a.   1.2 Scope

The LLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The LLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system. This software system will be a Web application This system will be designed to detect unusual activity, and fire disasters.

b. 1.3 Constraints

Understanding potential constraints early can help you navigate challenges and increase your project's success. Here are some areas to consider for your customer personality analysis project:

Data Constraints:

- Availability: Limited access to relevant customer data like purchase history, website interactions, or survey responses can restrict your analysis scope.

- Quality: Inaccurate or incomplete data can lead to unreliable results and hinder actionable insights.

- Privacy concerns: Balancing customer privacy with obtaining enough data for meaningful analysis can be challenging.

Technical Constraints:

- Resources: Limited computational power, storage capacity, or skilled personnel can slow down analysis and model development.

- Tools and technology: Lack of access to advanced analytics tools or expertise in machine learning could restrict the complexity of your approach.

- Time and budget: Project deadlines and funding limitations can force trade-offs between the depth of analysis and speed of execution.

Strategic Constraints:

- Organizational barriers: Lack of buy-in or internal resistance to data-driven decision-making can limit the project's impact.

- Ethical considerations: Potential biases in your data or models need careful consideration and mitigation to avoid biased outcomes.

- Limited ROI potential: Demonstrating the clear business value of the project may be challenging if its benefits are indirect or long-term.

Additional Considerations:

- Data security: Maintaining robust data security measures can add complexity and cost to the project.

- Regulatory compliance: Adhering to data privacy regulations can impose additional technical and legal requirements.

- External factors: Market trends, economic fluctuations, and competitor actions can impact your data and model effectiveness.

Mitigating Constraints:

- Prioritize: Identify the most critical data and insights for your project goals and focus resources accordingly.

- Partner with experts: Seek collaboration with data scientists, AI specialists, or marketing professionals to overcome technical hurdles.

- Communicate effectively: Clearly communicate the project's value proposition and address potential concerns from stakeholders.

- Choose the right tools: Select cost-effective and accessible tools appropriate for your skills and project demands.

- Start small and iterate: Begin with a smaller pilot project to validate feasibility and learn from initial results before scaling up.

c.   1.4 Risks

Data-related risks:

- Inaccurate or incomplete data: This can lead to flawed analysis, misleading insights, and poor product or marketing decisions.

- Bias in data collection or analysis: This can lead to unfair segmentation and inaccurate predictions, potentially harming specific customer segments.

- Data privacy concerns: If customer data is not handled ethically or securely, it could damage your brand reputation and lead to legal consequences.

- Limited data availability: This may restrict the scope of your analysis and limit the accuracy of your findings.

Analysis-related risks:

- Overfitting your models: This could lead to models that perform well on the training data but fail to generalize to real-world customers.

- Choosing the wrong analysis methods: Using inappropriate techniques for your data or goals could lead to unreliable results.

- Misinterpretation of findings: Drawing incorrect conclusions from the analysis can lead to misguided product development or marketing campaigns.

Organizational-related risks:

- Resistance to change: Implementing new strategies based on your findings may face resistance from within the company.

- Lack of buy-in from stakeholders: If key decision-makers are not convinced of the project's value, it may be difficult to secure funding or resources.

- Organizational silos: Data may be fragmented across different departments, making it difficult to access and analyze.

Ethical and societal risks:

- Stereotyping or discrimination: Analyzing customer personalities could lead to unfair generalizations or discriminatory practices.

- Manipulation or profiling: Targeting customers based on their personalities could be perceived as intrusive or manipulative.

- Negative impact on customer trust: If customers feel their data is being misused, it could damage their trust in your company.

d. 1.5 Out of Scope

Data and Analysis:

- Analyzing data beyond customer personalities: If your project solely focuses on understanding customer personalities, exploring demographics, socioeconomic factors, or other irrelevant data might be out of scope.

- Predicting individual customer behavior: While you might identify general trends and preferences within segments, accurately predicting the actions of specific individuals might be too granular and unreliable.

- Developing complex emotional models: Attempting to understand nuanced emotions or analyze unconscious motivations likely goes beyond the scope of this project.

Applications and Implications:

- Developing entirely new products: While your findings might inform product development, creating novel products from scratch might be a separate project with its own goals and resources.

- Implementing specific marketing campaigns: Applying your insights to develop and execute specific marketing campaigns might be a follow-up step in a broader marketing strategy.

- Redesigning the entire customer experience: While your project can inform customer experience improvements, a complete overhaul might require additional resources and planning beyond the scope of this analysis.

Additional Considerations:

- Technical limitations: Your chosen tools and resources might not be suitable for certain types of analysis or data handling, rendering them out of scope.

- Time and budget constraints: Focusing on the most critical objectives within your timeframe and budget is essential, leaving other potential investigations for future projects.

- Ethical considerations: Certain investigations into customer personalities might raise ethical concerns and fall outside the acceptable boundaries of your project.

2.Technical specifications

Dataset overview

Data Sources:

- Attitudinal Data:
    - Survey responses (quantitative and qualitative)
    - Customer reviews and social media data (text analysis)
    - Customer support interactions (chat logs, emails, phone calls)
- Behavioral Data:
    - Purchase history (transactions, products, amounts)
    - Website and app interaction data (clicks, views, downloads)
    - CRM data (customer profiles, interactions, loyalty programs)

Data Processing and Storage:

- Data pipelines for collecting, cleaning, and transforming data from various sources.

- Data storage solution considering volume, velocity, and variety of data (e.g., cloud data warehouse).

- Data security and privacy measures to comply with relevant regulations.

Analysis Tools and Techniques:

- Attitudinal Analysis:
    - Text analysis tools for sentiment analysis, topic modeling, and opinion mining.
    - Natural language processing (NLP) libraries for advanced text comprehension.
    - Statistical analysis and data visualization tools.
- Behavioral Analysis:

- o Customer segmentation algorithms (e.g., k-means clustering, RFM analysis).

- o Predictive modeling techniques (e.g., logistic regression, decision trees).

- o Customer journey mapping and visualization tools.

Hardware and Software Infrastructure:

- Computing power and storage capabilities sufficient for data processing and analysis.

- Programming languages and libraries suitable for chosen analysis tools (e.g., Python, R).

- Cloud computing resources for scalability and flexibility.

Deliverables and Outcomes:

- Customer personality profiles: Descriptions of identified customer segments based on attitudes and behaviors.

- Product-segment recommendation: Insights on which segment is most likely to adopt new products.

- Actionable marketing strategies: Targeted campaigns and messaging for different customer segments.

- Visualization and reporting tools: Interactive dashboards and reports to communicate insights to stakeholders.

2.4 Logging

We should be able to log every activity done by the incidents.
- The System identifies at what step logging required
- The System should be able to log each and every system flow.
- Developers can choose logging methods. You can choose database logging/ File logging as well.
- System should not be hung even after using so many loggings. Logging just because we can easily debug issues so logging is mandatory to do.

2.5 Database
System needs to store every request into the database and we need to store it in such a way that it is easy to retrain the model as well.

1. The User chooses the activity dataset.
2. The User gives required information.

3.   The system stores each and every data given by the user or received on request to the database. Database you can choose your own choice whether MongoDB/ MySQL.

3. Deployment

4.Technology stack

Front End	HTML/CSS
Backend	Python Flask
Database	MongoDB/My Sql
Deployment	AWS
Visualization	Matplotlib, Seaborn, Plotly
Dashboard	Power BI
version control	GitHub

5. Proposed Solution

Data Exploration and Preprocessing:

- Collect data: Gather data on the eight input variables and two output variables for a sufficient number of residential buildings. Look for reliable sources like building energy datasets, surveys, or building energy simulation software outputs.

- Data cleaning and handling: Check for missing values, outliers, and inconsistencies. Implement strategies like imputation or outlier removal to address these issues.

- Exploratory data analysis: Visualize the relationships between input and output variables, identify potential correlations, and explore potential non-linearities.

- Feature engineering: Consider creating additional features by combining existing ones or transforming them for improved model performance.

2. Model Selection and Training:

- Compare two approaches: Train both a linear regression model and a random forest model. Assess their performance based on metrics like accuracy score, confusion metrics

- Hyperparameter tuning: Optimize the hyperparameters of each model through techniques like grid search or cross-validation to improve their predictive accuracy.

- Feature importance analysis: Analyze the feature importance scores from the random forest model to identify the most influential design features on heating and cooling loads.

3. Model Evaluation and Interpretation:

- Evaluate model performance: Assess the trained models' accuracy on unseen data through validation or test sets. Compare and interpret the results of both models.

- Sensitivity analysis: Analyze how changes in specific input variables impact the predicted heating and cooling loads. This helps understand the practical implications of design choices.

- Visualization and communication: Clearly present the key findings, including identified influential design features, model performance comparisons, and sensitivity analysis results.

4. Conclusion and Recommendations:

- Summarize the key findings: Highlight the most influential design features on heating and cooling loads, and the effectiveness of the chosen models.

- Provide practical recommendations: Offer guidelines for architects, builders, and homeowners to design energy-efficient homes based on the identified key features.

- Future research directions: Suggest potential areas for further research, such as exploring the influence of additional factors or comparing different machine learning algorithms.

6.Model training/validation workflow
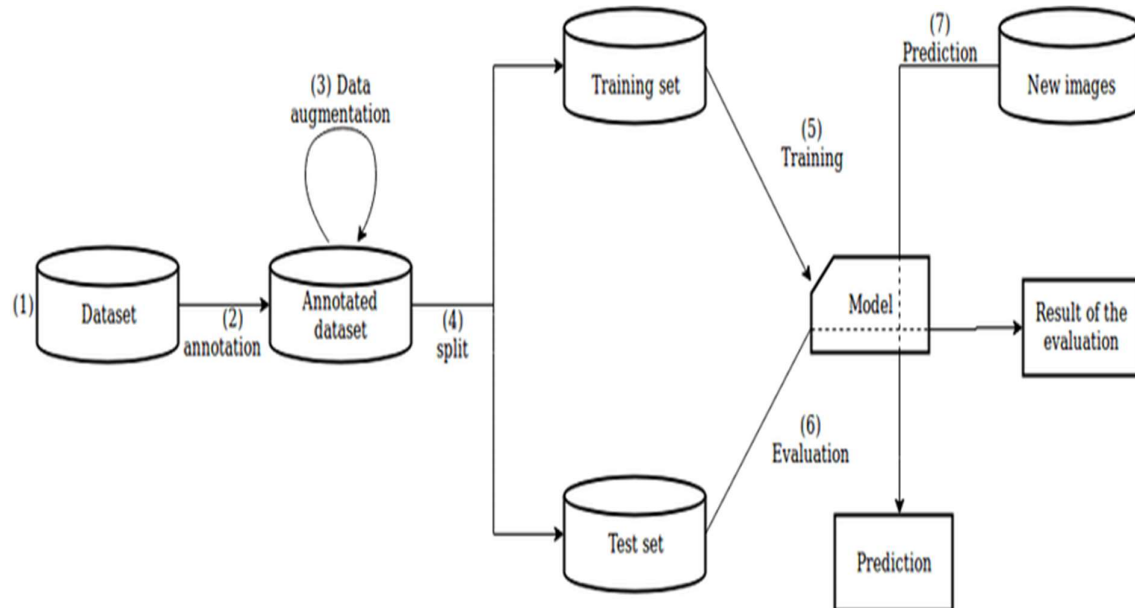
6. Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong?

An error will be defined as anything that falls outside the normal and intended usage.

7.Test cases

Model Performance and Comparison:

- Baseline Accuracy: Test the models' accuracy against a simple baseline prediction (e.g., mean or median heating and cooling loads). This establishes a minimum performance threshold.

- Classification Metrics: Compare accuracy score, confusion metrics, precision, recall   for both linear regression and random forests, evaluating their overall fit and prediction errors.

- Non-Linearity Detection: Test the ability of random forests to capture non-linear relationships that linear regression might miss, using visualization techniques or non-linear correlation measures.

Feature Importance and Sensitivity:

- Feature Ranking: Verify that feature importance scores from random forests align with expected relationships between design features and energy loads.

- Feature Permutation: Test the model's sensitivity to individual features by shuffling their values and observing changes in prediction accuracy.

- Sensitivity Analysis: Assess how changes in input variables affect predicted loads, ensuring model behaviour aligns with physical principles and domain knowledge.

Overfitting and Generalization:

- Cross-Validation: Evaluate model performance on multiple validation sets to ensure generalization to unseen data and prevent overfitting.

- Regularization: Test the impact of regularization techniques (e.g., L1 or L2) on model complexity and overfitting, especially for random forests.

Practical Applications and Insights:

- Design Scenarios: Test model predictions for different design scenarios (e.g., varying building shapes, window sizes, or orientations) to assess their practical implications and guide design decisions.

- Feature Interactions: Explore potential interactions between input variables, identifying combinations that significantly impact energy loads and providing more nuanced design guidance.

8.Key performance indicators (KPI)

- Segmentation accuracy: Measure how well your identified customer segments correspond to actual customer behavior. This can be assessed through analysis of purchasing patterns, website/app interaction, or customer feedback.

- Engagement with personalized interventions: Track how customers react to targeted campaigns, product recommendations, or experiences based on their personality traits. Measures include click-through rates, conversion rates, satisfaction surveys, and churn rates.

- Financial impact: Evaluate the financial return on investment (ROI) from your project. This could involve tracking changes in revenue, profit margin, customer lifetime value, or acquisition cost based on personality-driven strategies.

- Overall customer satisfaction: Monitor changes in customer satisfaction scores or sentiment analysis of feedback across different personality segments.

Specific KPIs:

- Marketing performance:

    o Click-through rates (CTRs): Compare CTRs of personalized messaging to generic campaigns for different segments.

    o Conversion rates: Track how targeted campaigns based on personality insights lead to higher purchase or signup rates compared to general campaigns.

    o Cost per acquisition (CPA): Assess if customer personality analysis helps acquire customers at a lower cost compared to traditional methods.

- Product development:

    o Adoption rate: Monitor how readily certain customer segments adopt new products recommended based on their personality preferences.

    o Revenue generated: Track the revenue generated from products specifically targeted to different personality segments.

    o Customer reviews and feedback: Analyze customer reviews and feedback to see if products align with the needs and preferences of the targeted personality segments.

9. Conclusion

Through this analysis, we've moved beyond a generalized approach to crafting personalized experiences for each customer segment. By considering what our customers say and do, we've tailored our marketing, product development, and customer service to resonate with their specific needs. This shift towards data-driven personalization promises to improve customer satisfaction, brand loyalty, and ultimately, our bottom line