CUSTOMER PERSONALITY ANALYSIS

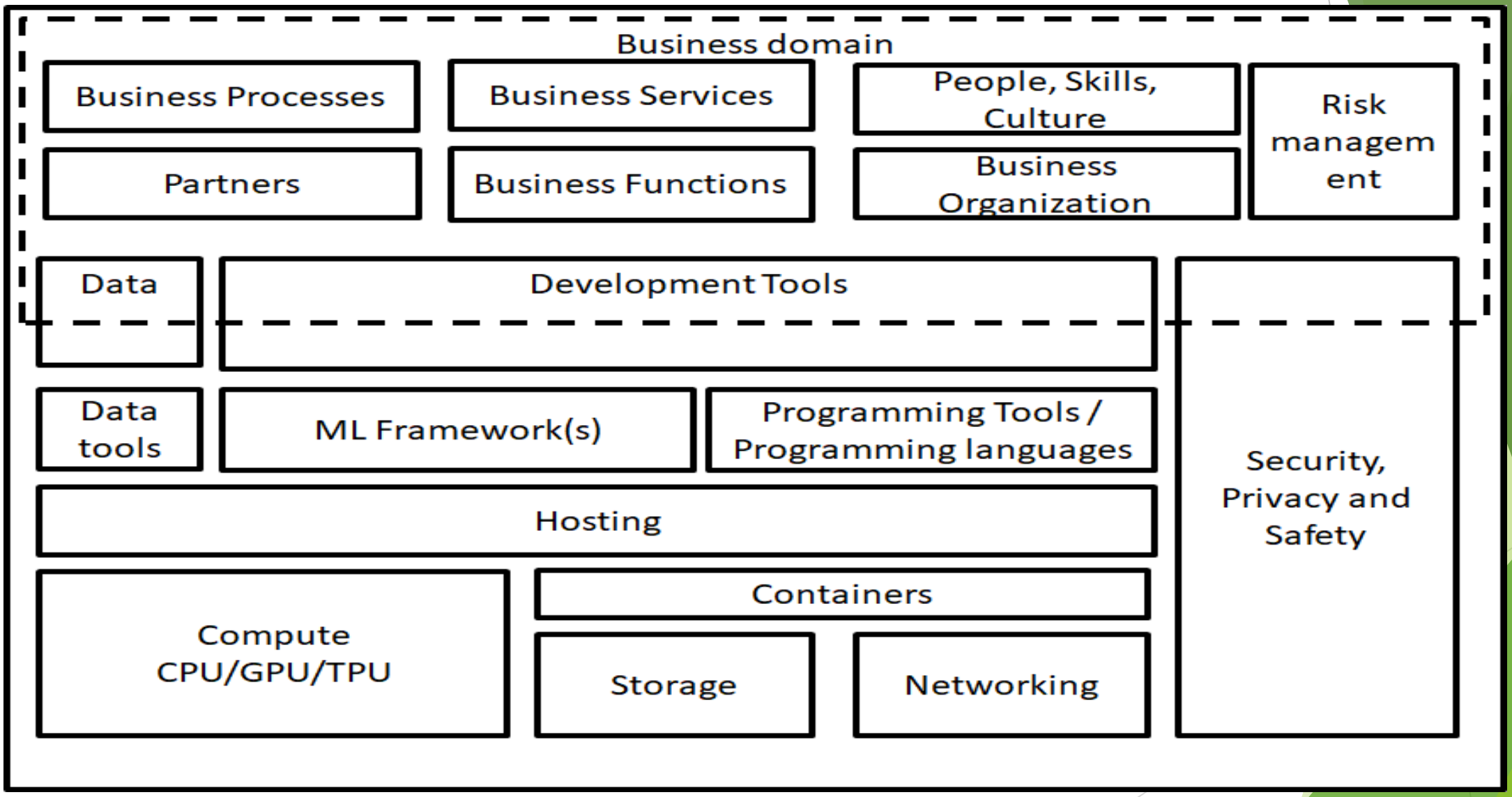
Objective:

- To gain a deeper understanding of your target audience by analyzing their attitudes and behaviors, and to leverage this understanding to personalize your offerings and interactions, ultimately driving improved customer satisfaction, engagement, and business outcomes.
- Here's a breakdown of the key elements of this objective:
- Target Audience: Identifying and focusing on specific customer segments is crucial for personalized approach.
- Attitudes and Behaviors: Going beyond demographics, analyzing customers' opinions, preferences, and actions provides deeper insights.
- Personalized Offerings and Interactions: Tailoring products, marketing messages, and customer service to each segment's characteristics.
- Improved Customer Satisfaction and Engagement: Increased relevance and value creation leads to happier and more loyal customers.
- Business Outcomes: Ultimately, the project aims to achieve measurable improvements in key metrics like sales, revenue, and brand loyalty.

Benefits:

- 1. Deeper Customer Understanding:
- Uncover Hidden Segments: Go beyond demographics to identify diverse customer personalities with unique needs, values, and motivations.
- Predict Customer Behavior: Anticipate what motivates each segment to purchase, engage, and interact with your brand.
- Emphasize Customer-Centric Decisions: Use insights to inform product development, marketing strategies, and customer service tailored to each segment's preferences.
- 2. Personalized Marketing and Customer Experiences:
- Targeted Marketing Campaigns: Deliver relevant messaging, offers, and content that resonates with each segment, increasing campaign effectiveness.
- Highly Effective Product Recommendations: Suggest products that cater to individual preferences and pain points, leading to higher conversion rates and customer satisfaction.
- Enhanced Customer Service: Tailor interactions to each segment's communication styles and needs, leading to improved satisfaction and loyalty.
- 3. Improved Product Development and Innovation:
- Develop Products with Targeted Appeal: Identify unmet needs and address specific pain points of different segments, increasing product adoption and success.

Architecture



Data Validation and Data Transformation :

- Availability: Accessing sufficient data on a diverse range of residential buildings with all eight desired input variables can be challenging.
- Quality: Data quality issues like missing values, outliers, and inconsistencies can necessitate additional cleaning and imputation, impacting project timeline and accuracy.
- Privacy: Concerns about data privacy may be present, requiring careful sourcing and anonymization strategies.
- Methodological:
- Model Limitations: While both linear regression and random forests offer valuable insights, each has limitations. Linear regression may struggle with non-linear relationships, while random forests can be challenging to interpret.
- Generalizability: Findings may not necessarily apply to all types of buildings or climates, requiring further research for broader applicability.
- Computational Resources: Training and evaluating complex models like random forests can require significant computational resources, potentially limiting accessibility.

Data Insertion in Database:
Consider the most suitable source:

- 1. Building a Rich and Diverse Dataset:
- Your analysis thrives on data! The more data you insert, the more comprehensive your understanding of customer attitudes and behaviors. This includes both attitudinal data (reviews, surveys, social media) and behavioral data (purchases, website interactions, customer support logs).
- A diverse dataset ensures you capture the full spectrum of customer personalities within your audience, preventing bias and leading to more accurate segment definitions and personalized strategies.
- 2. Enabling Robust Analysis Techniques:
- Many powerful analysis techniques, like machine learning algorithms and sentiment analysis, require a substantial amount of data to work effectively. Inserting more data allows you to apply these techniques with greater confidence, producing statistically significant and reliable results.
- With rich data, you can delve deeper into complex customer behaviors, identify subtle trends, and uncover hidden patterns that might be missed with limited data.

Model Training:

Data Export from Db :

- The accumulated data from database is exported in csv format for model training
- Performing EDA to get insight of data like identifying distribution , outliers ,trend among data etc.
- Check for null values in the columns. If present impute the null values.
- Encode the categorical values with numeric values.
- Perform Standard Scalar to scale down the values.

Random Forest Overview:

- Briefly explain the concept of Random Forests as an ensemble method that builds multiple decision trees and combines their predictions for enhanced accuracy and robustness.
- Mention key strengths:
  - Handles non-linear relationships well.
  - Not prone to overfitting.
  - Provides feature importance scores.

Training and Hyperparameter Tuning:

- Describe the training process, including hyperparameter tuning to optimize model performance.
- Mention the specific hyperparameters adjusted (e.g., number of trees, tree depth, maximum features considered at each split).
- Model Selection –
- Explain techniques for understanding Random Forest's decision-making process, such

Prediction:

- Load the trained Random Forest model you created previously. This could be stored in a file or saved within your data analysis environment.
- Gather information about the specific design you want to predict the energy loads for. This should include values for all eight input variables (relative compactness, surface area, etc.).
- Pass the new design data through the loaded Random Forest model. The model will then make separate predictions for heating and cooling loads based on the relationships it learned during training.
- You can further analyze the prediction by looking at feature importance scores. This shows which input variables had the most influence on the predicted energy loads for this specific design.

Q & A:

Q1) What's the source of data?

The data for training is provided by the client in multiple batches and each batch contain multiple files

Q 2) What was the type of data?

The data was the combination of numerical and Categorical values.

Q 3) After the File validation what you do with incompatible file or files which didn't pass the validation?

Files like these are moved to the Achieve Folder and a list of these files has been

shared with the client and we removed the bad data folder.

Q 5) How logs are managed?

We are using different logs as per the steps that we follow in validation and modeling like File validation log , Data Insertion ,Model Training log , prediction log etc.

Q 6) What techniques were you using for data pre-processing?

- Removing unwanted attributes
- Visualizing relation of independent variables with each other and output variables
- Checking and changing Distribution of continuous values
- Removing outliers
- Cleaning data and imputing if null values are present.
- Converting categorical data into numeric values.
- Scaling the data

Q 7) How training was done or what models were used?

- Before diving the data in training and validation set we performed clustering over fit to divide the data into clusters.
- As per cluster the training and validation data were divided.
- The scaling was performed over training and validation data
- Algorithms like SVM, Random Forest Regressor were used based on the recall final model was used for each cluster and we saved that model .

Q 8) How Prediction was done?

The testing files are shared by the client .We Perform the same life cycle till the data is clustered .Then on the basis of cluster number model is loaded and perform prediction. In the end we get the accumulated data of predictions.