# 

# GroupBy MLE Software Architecture

The GroupBy project aims to create a machine learning predictive model for customer conversion and optimizing the customer's journey through a personalized shopping experience. The team has identified 0.8+ conversions as the ideal state where the customer journey needs minimal intervention, and <0.2 conversions as a low return-on-investment category with low purchase intent. The team aims to personalize the shopping experience for customers with conversion rates ranging from 0.2 to 0.8, to increase engagement and improve conversion rates.

Here's the proposed architecture for building and deploying this model:

1. **Data Collection and Storage:** Collect customer data from various sources, including website traffic, browsing history, and transaction data, and store it in a data warehouse. Use cloud-based solutions such as Amazon S3 or Google Cloud Storage for scalable storage.
2. **Data Preprocessing and Cleaning:** Preprocess the data by performing various operations, including data normalization, feature engineering, and outlier removal. Cleaning should include identifying and correcting missing values and removing duplicates.
3. **Exploratory Data Analysis (EDA):** Analyze the preprocessed and cleaned data using statistical techniques and visualization to identify patterns, trends, and relationships among the variables.
4. **Model Training and Testing:** Train various machine learning models, such as decision trees, logistic regression, and neural networks, on the preprocessed data to predict customer behavior and personalize their shopping experience. The models should be evaluated using various metrics such as accuracy, precision, recall, and F1 score. The best performing models should be selected for deployment.
5. **Explainability and Interpretability:** Provide a mechanism to explain how the models work and how they are making decisions. This includes feature importance, SHAP plots, and LIME.
6. **Uplift Modeling:** Implement uplift modeling to predict the incremental effect of a personalized shopping experience on an individual's conversion rate. This will help identify which customers are most likely to benefit from the personalized shopping experience and prioritize their efforts accordingly.
7. **Deployment and Monitoring:** Deploy the best performing machine learning models and uplift model and monitor their performance regularly. Use containerization solutions like Streamlit, FASTAPI or Docker for easy deployment and use MLFlow for monitoring.
8. **Model Management:** Manage the machine learning models' life cycle, including versioning, archiving, and retiring the models when they become obsolete.
9. **Security**: Ensure data security, access control, and audit logging by implementing security solutions such as SSL/TLS encryption, OAuth2, and Multi-factor Authentication (MFA).
10. **Scalability:** Ensure that the software architecture is scalable to handle increasing data volumes and users. This includes using scalable cloud-based solutions, load balancing, and clustering.