Design Document: Customer Journey Analysis Using Clustering and Dimensionality Reduction

# 1. Project Overview

This project focuses on analyzing customer journeys using unsupervised learning techniques such as clustering and dimensionality reduction. The entire implementation is performed in Google Colab using Python and relevant machine learning libraries.

# 2. System Architecture

The system architecture comprises the following components:  
- Data Input Layer: Loads customer journey data (CSV, JSON, etc.).  
- Preprocessing Module: Cleans, normalizes, and prepares data.  
- Dimensionality Reduction Layer: Applies PCA or t-SNE.  
- Clustering Engine: Applies K-Means or other clustering algorithms.  
- Visualization Layer: Displays clusters using matplotlib or seaborn.  
- Insight Generation: Interprets and documents clustering results.

# 3. Workflow Design

1. Import libraries and load data.  
2. Handle missing values and normalize features.  
3. Apply PCA and/or t-SNE to reduce dimensions.  
4. Determine optimal number of clusters (e.g., using Elbow Method).  
5. Apply clustering (e.g., K-Means).  
6. Visualize and interpret results.  
7. Save results and generate actionable insights.

# 4. Data Flow Description

1. Raw data is ingested from external sources.  
2. Preprocessed using standardization and encoding.  
3. Dimensionality reduction compresses data for better visualization.  
4. Clustering groups customers into meaningful segments.  
5. Visualizations are generated to explore and explain results.

# 5. Tools and Libraries

- Google Colab  
- Python 3  
- pandas, numpy for data handling  
- scikit-learn for ML algorithms  
- matplotlib, seaborn for visualization  
- PCA, t-SNE for dimensionality reduction  
- KMeans, DBSCAN for clustering

# 6. Implementation Environment

Google Colab provides a cloud-based Jupyter Notebook environment with free access to GPU/TPU support. It is ideal for rapid prototyping, sharing, and collaboration. Python packages are pre-installed or can be added easily using pip.