QUANTUM WOLF
DATA INTELLIGENCE RESEARCH LAB
DATE: 14/02/2025 – 16/02/2025
NAME: SURIYA E

Real-Time Customer Segmentation & Personalized Offers

1. Problem Statement

The retail industry is increasingly competitive, with e-commerce platforms setting high standards for personalized customer experiences. Traditional customer segmentation methods are often static and fail to capture real-time customer behavior, leading to missed opportunities for targeted marketing and customer engagement. Retailers need a solution that can:

- Analyze customer behavior in real-time.
- Segment customers dynamically based on their purchasing patterns and preferences.
- Deliver personalized discounts and offers to improve customer retention, conversion rates, and revenue.

Key Challenges:

- Lack of real-time insights into customer behavior.
- Inability to deliver hyper-personalized offers.
- Difficulty in competing with e-commerce giants.

2. Solution

To address these challenges, we propose a Real-Time Customer Segmentation & Personalized Offers solution using KNIME. This solution leverages:

- **Real-Time Data Integration:** Streaming POS transaction data, CRM data, and IoT data (e.g., in-store foot traffic sensors).

- **Advanced Analytics:** Clustering algorithms (k-means, DBSCAN) for customer segmentation.
- **Machine Learning:** Predictive modeling to forecast discount usage.
- **Automation:** Dynamic discount triggering via email or APIs.
- **Real-Time Dashboards:** KNIME WebPortal for monitoring customer segments and campaign performance.

Key Benefits:

- Real-time customer insights.
- Hyper-personalized offers.
- Improved customer retention and revenue.

3. Working Process of Real-Time Customer Segmentation & Personalized Offers Using KNIME

The workflow is divided into the following steps:

1. Data Ingestion
2. Data Preprocessing
3. Feature Engineering
4. Customer Segmentation (Clustering)
5. Predictive Modeling (Discount Usage Prediction)
6. Personalized Discount Assignment
7. Data Visualization
8. Email Automation for Personalized Offers

Step 1: Data Ingestion

- **Node Used**: CSV Reader
- **Input File**: customer_transactions.csv

- **Configuration**:
  - Detect delimiter (comma ,)
  - Ensure correct data types (e.g., PurchaseAmount as Double, TransactionDate as Date/Time)
- **Output**: A structured dataset containing customer transactions ready for preprocessing.

File Table - 3:11 - CSV Reader

File Edit Hilite Navigation View

Table "default" - Rows: 200 | Spec - Columns: 10 | Properties | Flow Variables

Row ID	S Custom...	I Age	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	I StoreVi...	I LastPur...	S Discoun...
Row0	C001	29	Female	Silver	55.12	Sports	01-01-2024	4	3	No
Row1	C002	56	Male	Bronze	475.64	Fashion	02-01-2024	1	14	Yes
Row2	C003	19	Female	Gold	207.99	Sports	03-01-2024	1	24	No
Row3	C004	55	Female	Silver	272.22	Sports	04-01-2024	9	25	No
Row4	C005	19	Male	Silver	389.56	Fashion	05-01-2024	7	26	Yes
Row5	C006	43	Male	Bronze	308.43	Grocery	06-01-2024	6	4	No
Row6	C007	31	Female	Bronze	471.2	Home & Kitc...	07-01-2024	3	22	Yes
Row7	C008	42	Male	Silver	441.63	Fashion	08-01-2024	1	1	No
Row8	C009	53	Male	Platinum	290.32	Home & Kitc...	09-01-2024	10	15	No
Row9	C010	36	Female	Silver	28.45	Sports	10-01-2024	7	24	Yes
Row10	C011	40	Male	Platinum	150.91	Electronics	11-01-2024	7	14	Yes
Row11	C012	58	Female	Gold	96.06	Electronics	12-01-2024	8	14	Yes
Row12	C013	65	Male	Silver	20.32	Home & Kitc...	13-01-2024	5	26	Yes
Row13	C014	24	Female	Bronze	311.51	Sports	14-01-2024	1	10	No
Row14	C015	51	Female	Platinum	445.46	Grocery	15-01-2024	3	5	No
Row15	C016	35	Female	Bronze	455.94	Home & Kitc...	16-01-2024	5	23	Yes

Step 2: Data Preprocessing

- **Nodes Used**:
  - **Missing Value Node**: Handles missing data.
    - Numeric values replaced with median/mean.
    - Categorical values replaced with the most frequent entry.
  - **Normalizer Node**: Scales numerical features (PurchaseAmount, StoreVisits) to ensure uniformity.
- **Output**: A cleaned and normalized dataset ready for feature engineering.

Normalized table - 3:6 - Normalizer

File Edit Hilite Navigation View

Table "default" - Rows: 200 | Spec - Columns: 10 | Properties | Flow Variables

Row ID	S Custom...	D Age	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...
Row0	C001	29	Female	Silver	0.088	Sports	01-01-2024	0.333	3	No
Row1	C002	56	Male	Bronze	0.953	Fashion	02-01-2024	0	14	Yes
Row2	C003	19	Female	Gold	0.403	Sports	03-01-2024	0	24	No
Row3	C004	55	Female	Silver	0.535	Sports	04-01-2024	0.889	25	No
Row4	C005	19	Male	Silver	0.776	Fashion	05-01-2024	0.667	26	Yes
Row5	C006	43	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No
Row6	C007	31	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes
Row7	C008	42	Male	Silver	0.883	Fashion	08-01-2024	0	1	No
Row8	C009	53	Male	Platinum	0.572	Home & Kitc...	09-01-2024	1	15	No
Row9	C010	36	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes
Row10	C011	40	Male	Platinum	0.285	Electronics	11-01-2024	0.667	14	Yes
Row11	C012	58	Female	Gold	0.172	Electronics	12-01-2024	0.778	14	Yes
Row12	C013	65	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes
Row13	C014	24	Female	Bronze	0.616	Sports	14-01-2024	0	10	No
Row14	C015	51	Female	Platinum	0.891	Grocery	15-01-2024	0.222	5	No
Row15	C016	35	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes

Step 3: Feature Engineering

- **Nodes Used**:
  - **Math Formula Node**: Computes AvgSpendPerVisit = PurchaseAmount / StoreVisits.
  - **Date&Time Difference Node**: Computes Recency (days since the last purchase).
- **Output**: An enriched dataset with newly derived features (AvgSpendPerVisit, Recency).

Output table - 3:18 - Missing Value

File Edit Hilite Navigation View

Table "default" - Rows: 200 | Spec - Columns: 13 | Properties | Flow Variables

Row ID	S Custom...	D Age	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...	D Gender...	D Loyalty...	D AvgSpe...
Row0	C001	29	Female	Silver	0.088	Sports	01-01-2024	0.333	3	No	0	0	0.264
Row1	C002	56	Male	Bronze	0.953	Fashion	02-01-2024	0	14	Yes	1	1	0.854
Row2	C003	19	Female	Gold	0.403	Sports	03-01-2024	0	24	No	0	2	0.854
Row3	C004	55	Female	Silver	0.535	Sports	04-01-2024	0.889	25	No	0	0	0.602
Row4	C005	19	Male	Silver	0.776	Fashion	05-01-2024	0.667	26	Yes	1	0	1.164
Row5	C006	43	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No	1	1	1.097
Row6	C007	31	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes	0	1	4.248
Row7	C008	42	Male	Silver	0.883	Fashion	08-01-2024	0	1	No	1	0	0.854
Row8	C009	53	Male	Platinum	0.572	Home & Kitc...	09-01-2024	1	15	No	1	3	0.572
Row9	C010	36	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes	0	0	0.05
Row10	C011	40	Male	Platinum	0.285	Electronics	11-01-2024	0.667	14	Yes	1	3	0.428
Row11	C012	58	Female	Gold	0.172	Electronics	12-01-2024	0.778	14	Yes	0	2	0.221
Row12	C013	65	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes	1	0	0.037
Row13	C014	24	Female	Bronze	0.616	Sports	14-01-2024	0	10	No	0	1	0.854
Row14	C015	51	Female	Platinum	0.891	Grocery	15-01-2024	0.222	5	No	0	3	4.01
Row15	C016	35	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes	0	1	2.053

Step 4: Customer Segmentation (Clustering)

- **Nodes Used**:
  - **k-Means Clustering**: Identifies distinct customer groups by analyzing behavioral features.
  - **DBSCAN**: Used for density-based clustering to detect noise and anomalies.
- **Features Used**:
  - PurchaseAmount
  - StoreVisits
  - AvgSpendPerVisit
  - Recency

- **Configuration**:
  - k-Means: Optimal k determined using the Elbow Method.
  - DBSCAN: Defined epsilon and min points for density estimation.
- **Output**: Segmented customer groups such as High-Value Customers, Frequent Buyers, At-Risk Customers, and One-Time Buyers.

Labeled input - 3:12 - k-Means

File Edit Hilite Navigation View

Table "default" - Rows: 200 | Spec - Columns: 14 | Properties | Flow Variables

Row ID	S Custom...	D Age	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...	D Gender...	D Loyalty...	D AvgSpe...	S Cluster
Row0	C001	29	Female	Silver	0.088	Sports	01-01-2024	0.333	3	No	0	0	0.264	cluster_1
Row1	C002	56	Male	Bronze	0.953	Fashion	02-01-2024	0	14	Yes	1	1	0.854	cluster_2
Row2	C003	19	Female	Gold	0.403	Sports	03-01-2024	0	24	No	0	2	0.854	cluster_1
Row3	C004	55	Female	Silver	0.535	Sports	04-01-2024	0.889	25	No	0	0	0.602	cluster_0
Row4	C005	19	Male	Silver	0.776	Fashion	05-01-2024	0.667	26	Yes	1	0	1.164	cluster_0
Row5	C006	43	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No	1	1	1.097	cluster_0
Row6	C007	31	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes	0	1	4.248	cluster_2
Row7	C008	42	Male	Silver	0.883	Fashion	08-01-2024	0	1	No	1	0	0.854	cluster_2
Row8	C009	53	Male	Platinum	0.572	Home & Kitc...	09-01-2024	1	15	No	1	3	0.572	cluster_0
Row9	C010	36	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes	0	0	0.05	cluster_1
Row10	C011	40	Male	Platinum	0.285	Electronics	11-01-2024	0.667	14	Yes	1	3	0.428	cluster_0
Row11	C012	58	Female	Gold	0.172	Electronics	12-01-2024	0.778	14	Yes	0	2	0.221	cluster_0
Row12	C013	65	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes	1	0	0.037	cluster_1
Row13	C014	24	Female	Bronze	0.616	Sports	14-01-2024	0	10	No	0	1	0.854	cluster_2
Row14	C015	51	Female	Platinum	0.891	Grocery	15-01-2024	0.222	5	No	0	3	4.01	cluster_2
Row15	C016	35	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes	0	1	2.053	cluster_2

Step 5: Predictive Modeling (Discount Usage Prediction)

- **Nodes Used**:
  - **Partitioning Node**: Splits data (80% training, 20% testing).
  - **Random Forest Learner / Decision Tree Learner**: Trains models on past customer behaviors.
  - **Predictor Node**: Applies trained models to predict discount usage likelihood.
  - **Scorer Node**: Evaluates model accuracy using metrics like precision and recall.
- **Target Variable**: DiscountUsed (Yes/No)
- **Features Used**:
  - Purchase behavior
  - Customer demographics
  - Cluster label from segmentation step
- **Output**: Predicted probability of discount redemption for each customer.

Classified Data - 3:23 - Decision Tree Predictor

File Edit Hilite Navigation View

Table "default" - Rows: 60 | Spec - Columns: 15 | Properties | Flow Variables

Row ID	S Custom...	D Age	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...	D Gender...	D Loyalty...	D AvgSpe...	S Cluster	S Predicti...
Row2	C003	19	Female	Gold	0.403	Sports	03-01-2024	0	24	No	0	2	0.854	cluster_1	cluster_1
Row5	C006	43	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No	1	1	1.097	cluster_0	cluster_0
Row6	C007	31	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes	0	1	4.248	cluster_2	cluster_2
Row9	C010	36	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes	0	0	0.05	cluster_1	cluster_0
Row12	C013	65	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes	1	0	0.037	cluster_1	cluster_1
Row15	C016	35	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes	0	1	2.053	cluster_2	cluster_2
Row16	C017	34	Male	Bronze	0.782	Fashion	17-01-2024	0.333	29	Yes	1	1	2.346	cluster_2	cluster_2
Row23	C024	60	Female	Silver	0.483	Home & Kitc...	24-01-2024	0.111	25	Yes	0	0	4.35	cluster_2	cluster_1
Row29	C030	64	Male	Bronze	0.039	Sports	30-01-2024	0.444	22	Yes	1	1	0.087	cluster_1	cluster_1
Row32	C033	36	Male	Platinum	0.689	Fashion	02-02-2024	0.889	17	No	1	3	0.776	cluster_0	cluster_0
Row36	C037	40	Female	Gold	0.445	Fashion	06-02-2024	0.111	11	Yes	0	2	4.002	cluster_1	cluster_1
Row41	C042	30	Male	Silver	0.391	Sports	11-02-2024	0.556	3	No	1	0	0.704	cluster_0	cluster_1
Row45	C046	58	Female	Silver	0.152	Fashion	15-02-2024	1	27	Yes	0	0	0.152	cluster_0	cluster_0
Row46	C047	31	Male	Platinum	0.04	Electronics	16-02-2024	0.778	16	No	1	3	0.051	cluster_0	cluster_0
Row49	C050	20	Male	Bronze	0.498	Home & Kitc...	19-02-2024	0.778	28	No	1	1	0.64	cluster_0	cluster_0
Row51	C052	56	Male	Gold	0.447	Sports	21-02-2024	0.222	16	Yes	1	2	2.011	cluster_1	cluster_1

Step 6: Personalized Discount Assignment

- **Nodes Used**:
  - **Rule Engine Node**: Assigns discount categories based on customer segmentation and predicted discount usage.
- **Example Rules**:
  - $Prediction(Cluster)$ = "cluster_0" => "High-Value"
  - $Prediction(Cluster)$ = "cluster_1" => "Frequent Buyer"
  - $Prediction(Cluster)$ = "cluster_2" => "At-Risk"
- **Discount Rules**:
  - $Customer Segment$ = "High-Value" => "10% Off"
  - $Customer Segment$ = "Frequent Buyer" => "Special Offer"
  - $Customer Segment$ = "At-Risk" => "Retention Deal"
  - $Customer Segment$ = "One-Time Buyer" => "Welcome Back"
- **Output**: Discount offers tailored to individual customer profiles.

Classified values - 3:28 - Rule Engine

File Edit Hilite Navigation View

Table "default" - Rows: 60 | Spec - Columns: 17 | Properties | Flow Variables

Row ID	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...	D Gender...	D Loyalty...	D AvgSpe...	S Cluster	S Predicti...	S Custome...	S Discoun...
Row2	Female	Gold	0.403	Sports	03-01-2024	0	24	No	0	2	0.854	cluster_1	cluster_1	Frequent Buyer	20% offer
Row5	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No	1	1	1.097	cluster_0	cluster_0	High-Value	30% offer
Row6	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes	0	1	4.248	cluster_2	cluster_2	At-Risk	Special offer
Row9	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes	0	0	0.05	cluster_1	cluster_0	High-Value	30% offer
Row12	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes	1	0	0.037	cluster_1	cluster_1	Frequent Buyer	20% offer
Row15	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes	0	1	2.053	cluster_2	cluster_2	At-Risk	Special offer
Row16	Male	Bronze	0.782	Fashion	17-01-2024	0.333	29	Yes	1	1	2.346	cluster_2	cluster_2	At-Risk	Special offer
Row23	Female	Silver	0.483	Home & Kitc...	24-01-2024	0.111	25	Yes	0	0	4.35	cluster_2	cluster_1	Frequent Buyer	20% offer
Row29	Male	Bronze	0.039	Sports	30-01-2024	0.444	22	Yes	1	1	0.087	cluster_1	cluster_1	Frequent Buyer	20% offer
Row32	Male	Platinum	0.689	Fashion	02-02-2024	0.889	17	No	1	3	0.776	cluster_0	cluster_0	High-Value	30% offer
Row36	Female	Gold	0.445	Fashion	06-02-2024	0.111	11	Yes	0	2	4.002	cluster_1	cluster_1	Frequent Buyer	20% offer
Row41	Male	Silver	0.391	Sports	11-02-2024	0.556	3	No	1	0	0.704	cluster_0	cluster_1	Frequent Buyer	20% offer
Row45	Female	Silver	0.152	Fashion	15-02-2024	1	27	Yes	0	0	0.152	cluster_0	cluster_0	High-Value	30% offer
Row46	Male	Platinum	0.04	Electronics	16-02-2024	0.778	16	No	1	3	0.051	cluster_0	cluster_0	High-Value	30% offer
Row49	Male	Bronze	0.498	Home & Kitc...	19-02-2024	0.778	28	No	1	1	0.64	cluster_0	cluster_0	High-Value	30% offer
Row51	Male	Gold	0.447	Sports	21-02-2024	0.222	16	Yes	1	2	2.011	cluster_1	cluster_1	Frequent Buyer	20% offer
Row55	Female	Gold	0.801	Fashion	25-02-2024	0.889	29	Yes	0	2	0.902	cluster_0	cluster_0	High-Value	30% offer
Row56	Female	Platinum	0.361	Home & Kitc...	26-02-2024	0.333	2	No	0	3	1.083	cluster_1	cluster_1	Frequent Buyer	20% offer
Row57	Female	Bronze	0.955	Home & Kitc...	27-02-2024	0.222	11	Yes	0	1	4.3	cluster_2	cluster_2	At-Risk	Special offer

Step 7: Data Visualization

- **Nodes Used**:
    - **Bar Chart / Pie Chart**: Displays customer segment distribution.
    - **Stacked Bar Chart**: Shows discount redemption rates across segments.
    - **Box Plot**: Illustrates spending behavior variations.
    - **Heatmap**: Visualizes revenue contributions by customer segment.
- **Insights**:
    - Identify high-value segments.
    - Evaluate discount effectiveness.
    - Analyze customer spending trends.

Step 8: Email Automation for Personalized Offers

- **Nodes Used**:
    - **Row Filter Node**: Selects customers based on criteria (e.g., email ID, purchase history).
    - **Rule Engine Node**: Assigns personalized discount codes.
    - **Send Email Node**: Sends tailored emails via SMTP configuration.
- **Example Email Template**:
- Subject: "Exclusive Discount Just for You!"
- Message:

    Dear Customer,

    We noticed you love shopping with us! Here's your special offer:

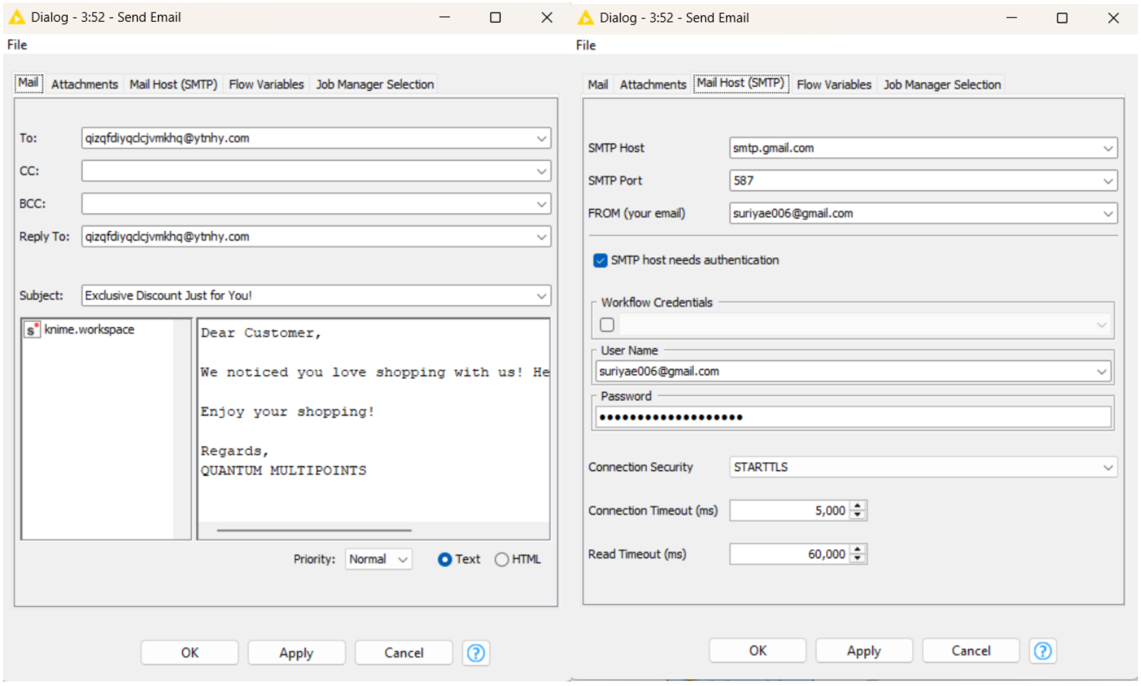
    $Discount_Message$

    Enjoy your shopping!

Regards,

QUANTUM MULTIPOINTS

- **Output**: Automated email campaigns delivering personalized promotions.



4. Final Workflow & Output

The final output of the project includes:

1. **Real-Time Customer Segments:**

   - High-Value Customers → Spend frequently, get a 10% discount.
   - Frequent Buyers → Visit often but spend moderately, get a special offer.
   - At-Risk Customers → Haven't purchased recently, get a retention discount.
   - One-Time Buyers → Low engagement, receive a welcome back offer.

Output:

Classified values - 3:28 - Rule Engine

File Edit Hilite Navigation View

Table "default" - Rows: 60 | Spec - Columns: 17 | Properties | Flow Variables

Row ID	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...	D Gender...	D Loyalty...	D AvgSpe...	S Cluster	S Predicti...	S Custome...	S Discoun...
Row2	Female	Gold	0.403	Sports	03-01-2024	0	24	No	0	2	0.854	cluster_1	cluster_1	Frequent Buyer	20% offer
Row5	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No	1	1	1.097	cluster_0	cluster_0	High-Value	30% offer
Row6	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes	0	1	4.248	cluster_2	cluster_2	At-Risk	Special offer
Row9	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes	0	0	0.05	cluster_1	cluster_0	High-Value	30% offer
Row12	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes	1	0	0.037	cluster_1	cluster_1	Frequent Buyer	20% offer
Row15	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes	0	1	2.053	cluster_2	cluster_2	At-Risk	Special offer
Row16	Male	Bronze	0.782	Fashion	17-01-2024	0.333	29	Yes	1	1	2.346	cluster_2	cluster_2	At-Risk	Special offer
Row23	Female	Silver	0.483	Home & Kitc...	24-01-2024	0.111	25	Yes	0	0	4.35	cluster_2	cluster_1	Frequent Buyer	20% offer
Row29	Male	Bronze	0.039	Sports	30-01-2024	0.444	22	Yes	1	1	0.087	cluster_1	cluster_1	Frequent Buyer	20% offer
Row32	Male	Platinum	0.689	Fashion	02-02-2024	0.889	17	No	1	3	0.776	cluster_0	cluster_0	High-Value	30% offer
Row36	Female	Gold	0.445	Fashion	06-02-2024	0.111	11	Yes	0	2	4.002	cluster_1	cluster_1	Frequent Buyer	20% offer
Row41	Male	Silver	0.391	Sports	11-02-2024	0.556	3	No	1	0	0.704	cluster_0	cluster_1	Frequent Buyer	20% offer
Row45	Female	Silver	0.152	Fashion	15-02-2024	1	27	Yes	0	0	0.152	cluster_0	cluster_0	High-Value	30% offer
Row46	Male	Platinum	0.04	Electronics	16-02-2024	0.778	16	No	1	3	0.051	cluster_0	cluster_0	High-Value	30% offer
Row49	Male	Bronze	0.498	Home & Kitc...	19-02-2024	0.778	28	No	1	1	0.64	cluster_0	cluster_0	High-Value	30% offer
Row51	Male	Gold	0.447	Sports	21-02-2024	0.222	16	Yes	1	2	2.011	cluster_1	cluster_1	Frequent Buyer	20% offer
Row55	Female	Gold	0.801	Fashion	25-02-2024	0.889	29	Yes	0	2	0.902	cluster_0	cluster_0	High-Value	30% offer
Row56	Female	Platinum	0.361	Home & Kitc...	26-02-2024	0.333	2	No	0	3	1.083	cluster_1	cluster_1	Frequent Buyer	20% offer
Row57	Female	Bronze	0.955	Home & Kitc...	27-02-2024	0.222	11	Yes	0	1	4.3	cluster_2	cluster_2	At-Risk	Special offer
Row65	Female	Platinum	0.993	Sports	07-03-2024	0.889	28	No	0	3	1.117	cluster_0	cluster_0	High-Value	30% offer
Row66	Male	Silver	0.818	Grocery	08-03-2024	0.778	4	No	1	0	1.051	cluster_0	cluster_0	High-Value	30% offer

2. **Personalized Discount Offers:**

Predictions for Discount Usage (Based on Decision Tree/Random Forest Model)

- Customers likely to use discounts → More aggressive promotions.
- Customers unlikely to use discounts → Adjust strategy to increase engagement.

Output:

Classified Data - 3:35 - Decision Tree Predictor

File Edit Hilite Navigation View

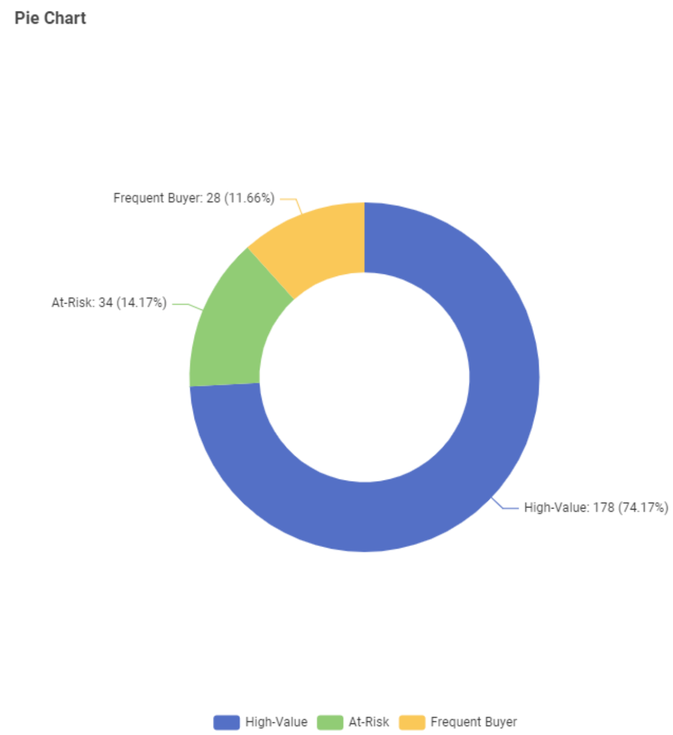
Table "default" - Rows: 60 | Spec - Columns: 17 | Properties | Flow Variables

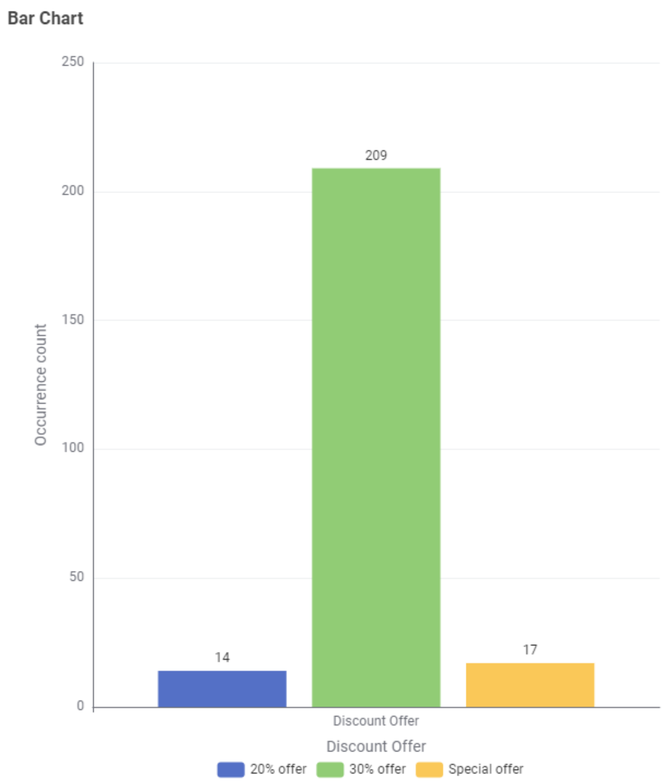
Row ID	S Gender	S Loyalty...	D Purcha...	S Category	S Transa...	D StoreVi...	D LastPur...	S Discoun...	D Gender...	D Loyalty...	D AvgSpe...	S Cluster	S Predicti...	S Custome...	S Predict ...
Row2	Female	Gold	0.403	Sports	03-01-2024	0	24	No	0	2	0.854	cluster_1	cluster_1	Frequent Buyer	No
Row5	Male	Bronze	0.609	Grocery	06-01-2024	0.556	4	No	1	1	1.097	cluster_0	cluster_0	High-Value	No
Row6	Female	Bronze	0.944	Home & Kitc...	07-01-2024	0.222	22	Yes	0	1	4.248	cluster_2	cluster_2	At-Risk	Yes
Row9	Female	Silver	0.033	Sports	10-01-2024	0.667	24	Yes	0	0	0.05	cluster_1	cluster_0	High-Value	Yes
Row12	Male	Silver	0.016	Home & Kitc...	13-01-2024	0.444	26	Yes	1	0	0.037	cluster_1	cluster_1	Frequent Buyer	Yes
Row15	Female	Bronze	0.913	Home & Kitc...	16-01-2024	0.444	23	Yes	0	1	2.053	cluster_2	cluster_2	At-Risk	Yes
Row16	Male	Bronze	0.782	Fashion	17-01-2024	0.333	29	Yes	1	1	2.346	cluster_2	cluster_2	At-Risk	Yes
Row23	Female	Silver	0.483	Home & Kitc...	24-01-2024	0.111	25	Yes	0	0	4.35	cluster_2	cluster_1	Frequent Buyer	Yes
Row29	Male	Bronze	0.039	Sports	30-01-2024	0.444	22	Yes	1	1	0.087	cluster_1	cluster_1	Frequent Buyer	Yes
Row32	Male	Platinum	0.689	Fashion	02-02-2024	0.889	17	No	1	3	0.776	cluster_0	cluster_0	High-Value	No
Row36	Female	Gold	0.445	Fashion	06-02-2024	0.111	11	Yes	0	2	4.002	cluster_1	cluster_1	Frequent Buyer	No
Row41	Male	Silver	0.391	Sports	11-02-2024	0.556	3	No	1	0	0.704	cluster_0	cluster_1	Frequent Buyer	No
Row45	Female	Silver	0.152	Fashion	15-02-2024	1	27	Yes	0	0	0.152	cluster_0	cluster_0	High-Value	Yes
Row46	Male	Platinum	0.04	Electronics	16-02-2024	0.778	16	No	1	3	0.051	cluster_0	cluster_0	High-Value	No
Row49	Male	Bronze	0.498	Home & Kitc...	19-02-2024	0.778	28	No	1	1	0.64	cluster_0	cluster_0	High-Value	No
Row51	Male	Gold	0.447	Sports	21-02-2024	0.222	16	Yes	1	2	2.011	cluster_1	cluster_1	Frequent Buyer	Yes
Row55	Female	Gold	0.801	Fashion	25-02-2024	0.889	29	Yes	0	2	0.902	cluster_0	cluster_0	High-Value	Yes
Row56	Female	Platinum	0.361	Home & Kitc...	26-02-2024	0.333	2	No	0	3	1.083	cluster_1	cluster_1	Frequent Buyer	No
Row57	Female	Bronze	0.955	Home & Kitc...	27-02-2024	0.222	11	Yes	0	1	4.3	cluster_2	cluster_2	At-Risk	No
Row65	Female	Platinum	0.993	Sports	07-03-2024	0.889	28	No	0	3	1.117	cluster_0	cluster_0	High-Value	Yes
Row66	Male	Silver	0.818	Grocery	08-03-2024	0.778	4	No	1	0	1.051	cluster_0	cluster_0	High-Value	No
Row73	Male	Platinum	0.943	Fashion	15-03-2024	0.111	4	Yes	1	3	8.486	cluster_2	cluster_2	At-Risk	Yes
Row74	Male	Platinum	0.315	Grocery	16-03-2024	0.778	12	Yes	1	3	0.405	cluster_0	cluster_0	High-Value	No

3. **Real-Time Dashboards:**

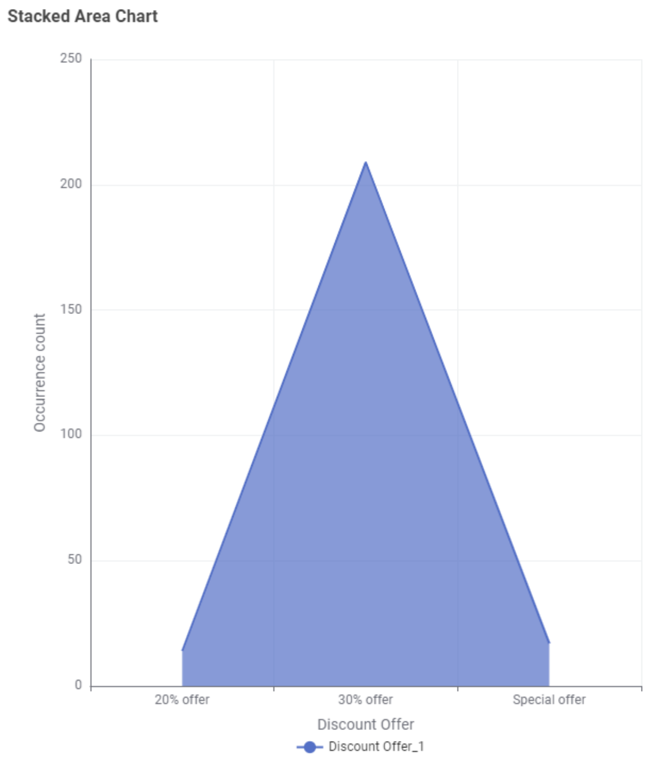
- Customer segment distribution.
- Discount redemption rates.
- Revenue impact of personalized campaigns.

- **Bar Chart / Pie Chart: Displays customer segment distribution.**

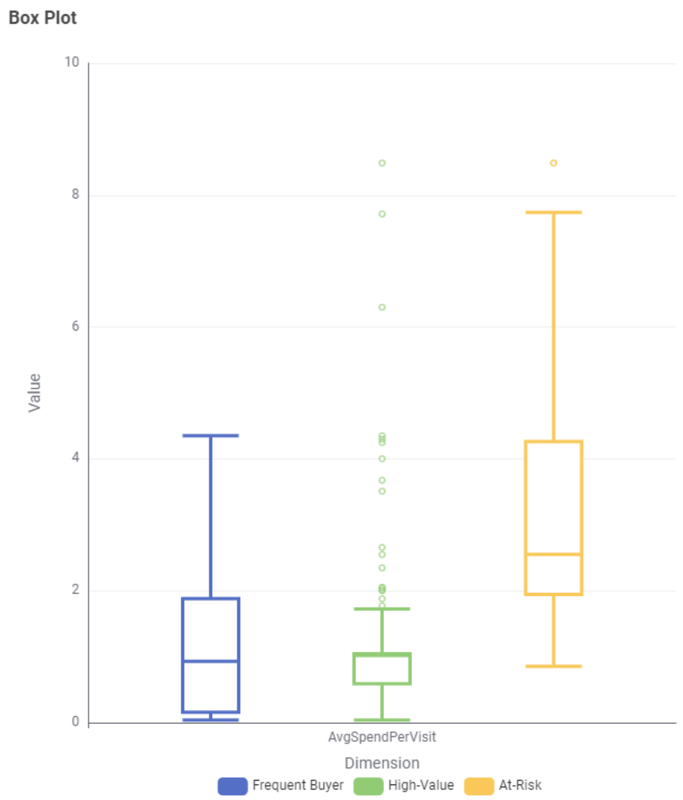




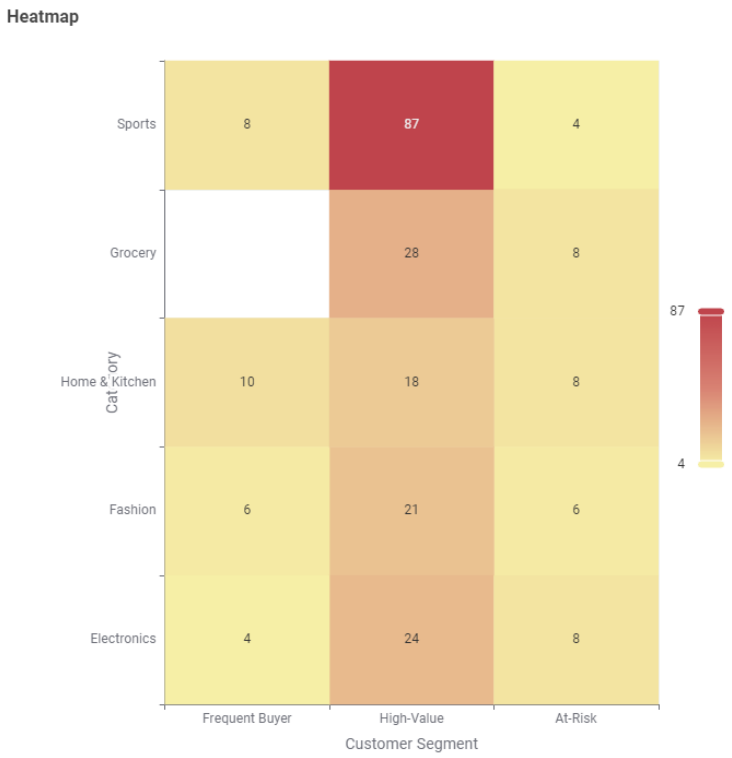
- **Stacked Bar Chart: Shows discount redemption rates across segments.**



- **Box Plot: Illustrates spending behavior variations.**



- **Heatmap: Visualizes Category by customer segment.**



- **Table View: Live updates of customer transactions and segmentation.**

Dialog - 3:39 - Table View

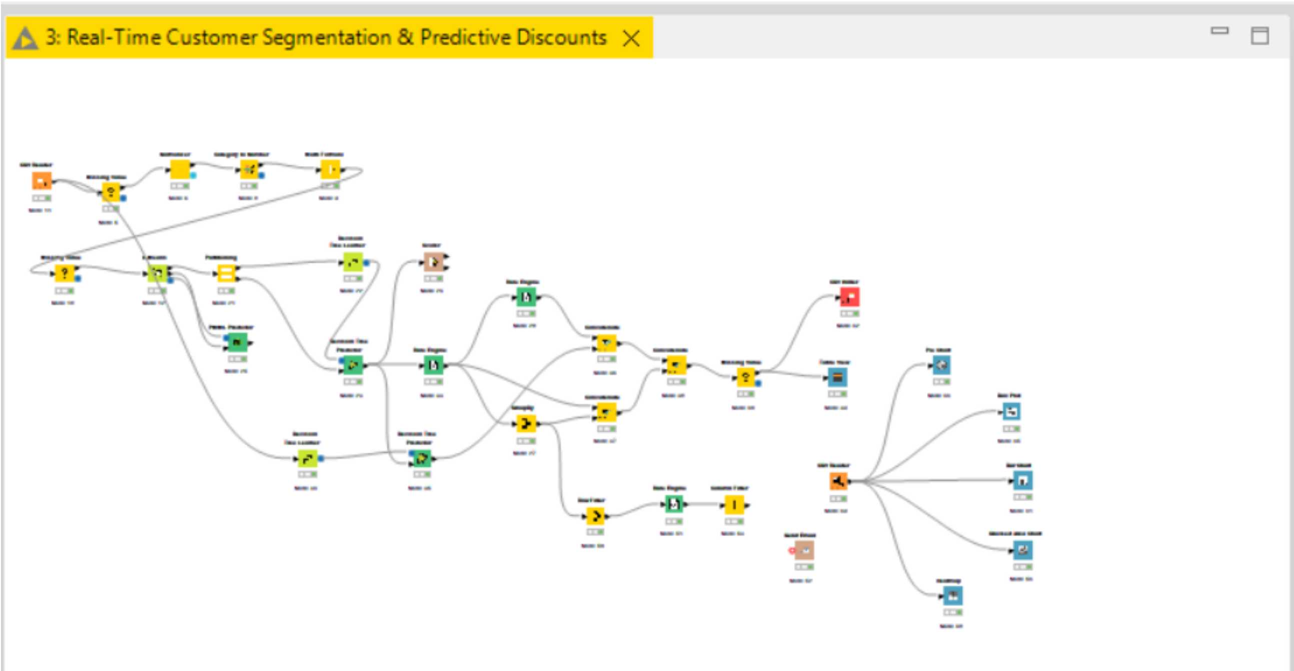
Table View

Rows: 240 | Columns: 5

RowID	CustomerID String	AvgSpendPerVisit Number (double)	Customer Segment String	Discount Offer String	Predict Probability String
Row0	C003	0.854	Frequent Buyer	20% offer	No
Row1	C006	1.097	High-Value	30% offer	No
Row2	C007	4.248	At-Risk	Special offer	No
Row3	C010	0.05	High-Value	30% offer	No
Row4	C013	0.037	Frequent Buyer	20% offer	No
Row5	C016	2.053	At-Risk	Special offer	No
Row6	C017	2.346	At-Risk	Special offer	No
Row7	C024	4.35	Frequent Buyer	20% offer	No
Row8	C030	0.087	Frequent Buyer	20% offer	No
Row9	C033	0.776	High-Value	30% offer	No
Row10	C037	4.002	Frequent Buyer	20% offer	No

4. **Business Impact:**
    - Increased customer retention and loyalty.
    - Higher conversion rates due to targeted offers.
    - Improved revenue from upselling and cross-selling.

Final Workflow:



5. Problems Faced During the Project

5.1. Data Integration Challenges

- Integrating real-time POS data with CRM and IoT data required careful handling of data formats and synchronization.
- **Issue:** Mismatched data formats caused delays in the initial stages.
- **Solution:** Used KNIME's data transformation nodes to standardize formats.

5.2. Math Formula Node Configuration

- **Issue:** During Step 3 (Feature Engineering), the Math Formula Node was not configured properly, leading to errors in calculating AvgSpendPerVisit.

- **Impact:** This issue delayed the workflow by almost 30 minutes.
- **Solution:** After troubleshooting, the correct formula (PurchaseAmount / StoreVisits) was applied, and the node was successfully configured.

5.3. Clustering Algorithm Issues

- **Issue:** The k-Means Clustering Node did not work as expected. It selected incorrect columns for clustering instead of the intended features (PurchaseAmount, StoreVisits, AvgSpendPerVisit, Recency).
- **Impact:** This caused a delay of 20 minutes in the clustering process.
- **Solution:** Reconfigured the node and ensured the correct columns were selected for clustering.

5.4. Predictive Modeling Challenges

- **Issue 1:** The Partitioning Node and Decision Tree Learner Node faced configuration issues during Step 5 (Predictive Modeling).
- **Issue 2:** The Predictor Node was not initially available in the KNIME platform.
- **Impact:** This caused a delay of almost 3 hours in the predictive modeling step.
- **Solution:** After seeking help, the Decision Tree Predictor Node was identified as an alternative and successfully implemented.

5.5. Email Node Connectivity Issues

- **Issue:** The Send Email Node failed to connect with the previous node in the workflow, preventing the automation of personalized emails.
- **Impact:** Despite trying multiple methods and alternate solutions, the issue could not be resolved.
- **Workaround:** Instead of automating the email process, a normal file was uploaded and used to send emails manually.

5.6. General Platform Challenges

- **Issue:** KNIME's platform occasionally experienced performance lags, especially when handling large datasets or complex workflows.
- **Impact:** This slowed down the overall workflow execution and required frequent troubleshooting.
- **Solution:** Workflow optimization and breaking down complex processes into smaller, manageable steps.

6. Conclusion

The **Real-Time Customer Segmentation & Personalized Offers** solution using KNIME successfully addresses the challenges faced by retailers in delivering hyper-personalized experiences. By leveraging real-time data integration, advanced analytics, and automation, retailers can:

- Gain real-time insights into customer behavior.
- Deliver targeted discounts to improve customer retention and conversion rates.
- Compete effectively with e-commerce giants.

Key Outcomes:

- A **15-20% increase in sales** from personalized campaigns.
- Improved customer loyalty and satisfaction.
- Enhanced revenue through upselling and cross-selling.