7-2 Project Two

DAT 300 Valid Data: Getting Data Right

Professor G. Dean Snyder, MBA, MSCS SNHU

Andrea Plunkett Jackson

10/19/2024

**Step-by-Step Data Mining Plan**

**Step 1: Data Cleaning and Preparation**

* **Task**: Clean the sample survey data, handle missing values, and standardize formats.
* **Purpose**: Ensure data quality and consistency for accurate analysis.

**Step 2: Exploratory Data Analysis (EDA)**

* **Task**: Perform initial exploration of the survey data to identify patterns, distributions, and anomalies.
* **Purpose**: Gain an understanding of the data and formulate hypotheses for further analysis.

**Step 3: Cluster Analysis**

* **Task**: Conduct cluster analysis on survey responses to identify natural customer segments.
* **Purpose**: Determine how customers group based on their satisfaction levels and purchasing behaviors.
* **Method**: Use k-means or hierarchical clustering algorithms.
* **Output**: Visualization of customer clusters with descriptions of each segment.

**Step 4: Linear Regression**

* **Task**: Use linear regression to model the relationship between customer satisfaction scores and various factors.
* **Purpose**: Identify key drivers of customer satisfaction.
* **Method**: Fit a regression model using factors like purchase frequency, average spend, and channel used as predictors.
* **Output**: Coefficients indicating the strength and direction of relationships.

**Step 5: Logistic Regression**

* **Task**: Apply logistic regression to predict the probability of customer repurchase based on satisfaction scores.
* **Purpose**: Predict customer loyalty and identify at-risk customers.
* **Method**: Fit a logistic regression model using satisfaction scores and demographic data.
* **Output**: Probabilities of repurchase for each customer.

**Step 6: Result Interpretation and Visualization**

* **Task**: Interpret the results from cluster, linear, and logistic regression analyses.
* **Purpose**: Translate data insights into actionable business strategies.
* **Output**: Visualizations and reports summarizing key findings and recommendations.

**Step 7: Addressing Limitations and Errors**

* **Task**: Identify and resolve any sources of error or inaccuracies in the data.
* **Purpose**: Ensure the robustness and reliability of the analysis.
* **Method**: Conduct validation checks and handle outliers or noise.

**Potential Errors and Fixes**

1. **Missing Data**: Use imputation methods or remove incomplete records.
2. **Inconsistent Data**: Standardize formats and validate entries.
3. **Outliers**: Detect and handle outliers using statistical methods or domain knowledge.

**Meaningful Patterns**

* **Observation**: Identify distinct customer segments with specific satisfaction drivers.
* **Implication**: Tailor marketing and service strategies to each customer segment.

**Additional Research Questions**

* What are the demographic characteristics of each customer segment?
* How do purchasing behaviors differ between customer segments?

**Alternative Analytic Methods**

* **Decision Trees**: For understanding decision-making processes.
* **Neural Networks**: For capturing complex relationships.