TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[Division Name] - [Engagement Manager], [Senior Consultant], [Junior Consultant]

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Introduction

Sprocket Central Pty Ltd specializes in high-quality bikes and cycling accessories. They have shared a list of 1000 potential customers lacking transaction history.

- The analysis aims to optimize resource allocation for targeted marketing.
- The goal of this project is to identify high-value customers among the new potential customers.
- The marketing team aims to boost business by analyzing customer trends and behavior.
- The project is scoped for 3 weeks with three phases: Data Exploration, Model Development, and Interpretation.
- Each phase is allocated one week.

Introduction

New Customers Analysis

Analyze the following factors for new customers:

1. Age distribution criteria
2. Assess whether there are gender-related trends in value.
3. Distribution of customers across wealth segments.
4. Distribution of customers across different job titles.
5. Check if higher education correlates with high value.
6. Identify regions with potential high-value customers.
7. Check if customers from certain industries tend to be high value.
8. Visualize the distribution of bike-related purchases.
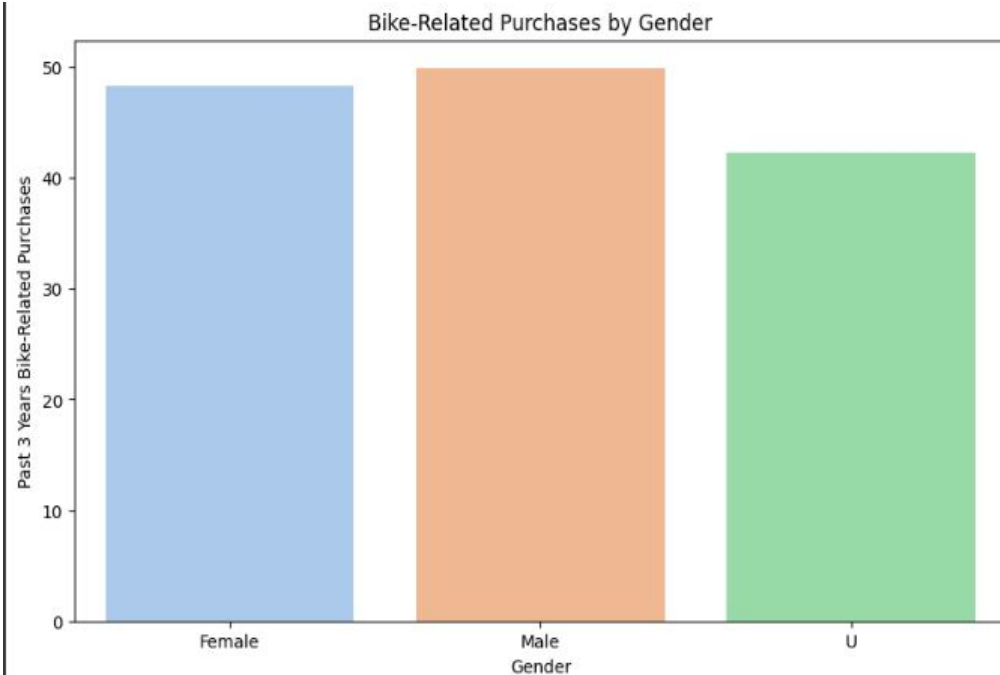
Data Exploration

Age Distribution

- The age group "46-60" represents the largest proportion of high-value customers among the specified age categories.
- Consider tailoring marketing strategies specifically for customers aged 46-60.
- This age group might respond well to promotions, campaigns, or product features that align with their preferences and needs.

Distribution of High-Value Customers Across Age Groups by Gender
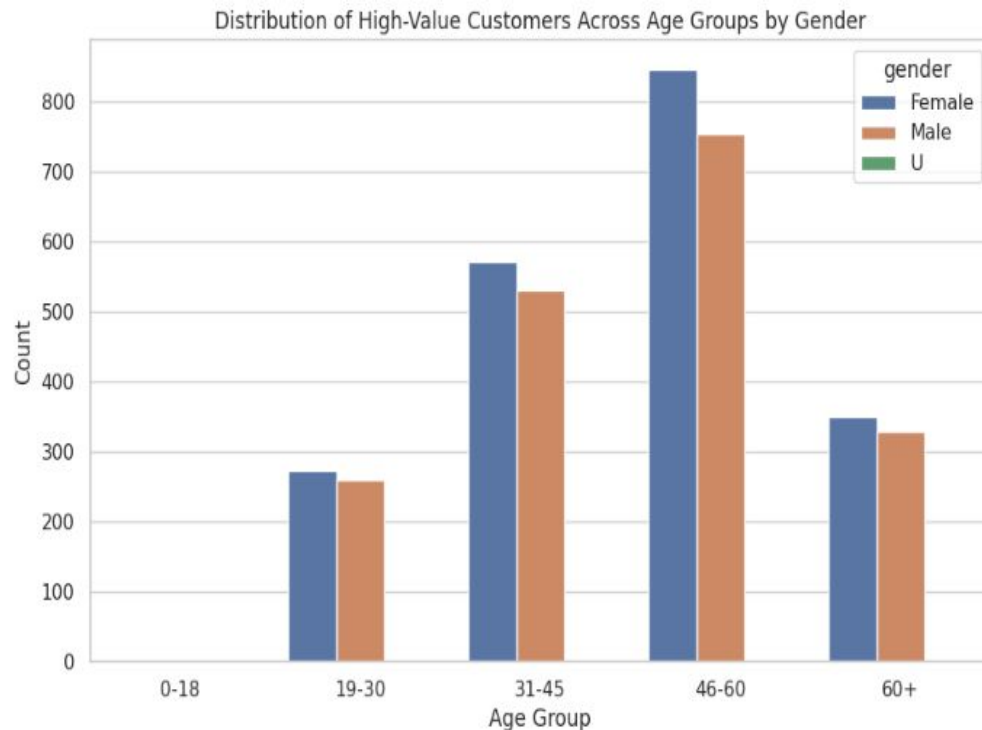
Data Exploration

Purchase History of Bikes (last 3 years)

- Around 51% (25,212 bikes) males purchased a bike within the last 3 years and the female purchase sums up to 47% (23,765 bikes).
- The marketing team might develop campaigns that specifically target male customers, leveraging their higher engagement with biking-related products.

Bike-Related Purchases by Gender

Data Exploration

Wealth Segment

- "Mass Customer" especially females indicates that a large number of customers belong to this wealth segment.
- While having a large number of mass customers is positive for broad market reach, it's essential to ensure that your marketing strategies are also reaching and appealing to other wealth segments, especially if they are more profitable.

Distribution of Customers Across Wealth Segments by Gender

Data Exploration

Job Industry

- "Mass Customer" especially females indicates that a large number of customers belong to this wealth segment.
- While having a large number of mass customers is positive for broad market reach, it's essential to ensure that your marketing strategies are also reaching and appealing to other wealth segments, especially if they are more profitable.

Model Development

Modeling Strategy

- Our objective is to predict high-value customers among the new potential customers.
- We have considered several classification models for this task, including:
  1. Logistic Regression
  2. Decision Trees
  3. Random Forests
  4. Gradient Boosting
- Our choice of models is based on factors such as interpretability, ability to capture complex relationships, and robustness in handling potential imbalances in the data

Model Development

Modeling Strategy

- Logistic Regression is a suitable choice for binary classification problems like ours. It provides a probabilistic interpretation of the results, making it easier to interpret the likelihood of a customer being high-value.
- Decision Trees are advantageous for capturing non-linear relationships and interactions among features. They can identify key decision points that contribute to a customer being classified as high-value.
- Random Forests offer an ensemble of decision trees, improving predictive accuracy and reducing overfitting. They are robust in handling various types of data and can handle a large number of features
- Gradient Boosting is effective for boosting the predictive power of weak models. It builds trees sequentially, focusing on correcting errors from the previous trees. This can lead to superior predictive performance

Model Development

Training & Testing

- We will split our dataset into training (80%) and testing (20%) sets to train and evaluate our models effectively
- Given the nature of our problem, where high-value customers may be a minority, we will employ techniques like oversampling or undersampling to handle imbalanced classes during model training.