# Market Entry Analysis Report

## Analytical Report for ABG Motors' Entry into the Indian Market

Harshit Jain

## Introduction

ABG Motors, a Japanese car manufacturer, is exploring the potential of entering the Indian market. Their primary goal is to assess whether market conditions in India align with their experiences in Japan and to estimate the feasibility of achieving a sales target of 10,000 units in the first year.

The analysis was conducted using two datasets: one representing the Japanese market (including purchase information) and another for the Indian market (without purchase data). By building a predictive model using logistic regression on the Japanese dataset, we could estimate sales potential in the Indian market.

Key highlights include:

- A logistic regression model was developed based on Japanese market data, with multiple independent variables influencing customer purchase decisions.
- The model was applied to Indian market data to predict the number of purchases.
- Several classification thresholds were tested, with a cutoff of 0.2 yielding the highest accuracy.
- The model predicts 29,735 sales in the Indian market within the first year.
- Additional steps were taken to segment and normalize key variables for more accurate analysis.

## Data

The datasets used for analysis contained the following variables:

- ID: Unique identifier for each customer.
- Age: Age of the customer.
- Gender: Male or Female.
- Annual Income: Income of the customer in local currency (yen for Japan, rupees for India).
- Previous Car Age: The age of the customer's previous car, if applicable.
- Purchase: Purchase outcome (only available in the Japanese dataset).

**Example Values:**

| ID | CURR_AGE | GENDER | ANN_INCOME(INR) | AGE _CAR (segmented) | PURCHASE |
|---|---|---|---|---|---|
| 00001Q15YJ | 50 | 1 | 0.49 | 3 | 0 |
| 00003I71CQ | 35 | 1 | -1.44 | 2 | 0 |
| 00003N47FS | 59 | 0 | 0.82 | 3 | 1 |

**Segmentations and Adjustments:**

- Previous Car Age Binning: For easier analysis, the previous car age variable was segmented into four bins, represented as categories: 1, 2, 3, and 4. This allowed us to capture general trends more effectively without getting lost in granular data.

- Income Normalization: The annual income in the Japanese dataset was normalized to match the Indian dataset, ensuring consistent analysis between the two datasets. This was crucial to accurately predict purchasing power and consumer behavior in India

**Abnormalities and Steps Taken:**

- Missing Values: Some customer data (e.g., income or previous car age) was missing. These were either imputed based on median values or removed.
- Outliers: Extreme values in income and previous car age were capped based on interquartile range (IQR) analysis.
- Data Wrangling: Microsoft Excel 2021 was used to clean and organize the datasets.

# Analysis

**Methods Used:**

1. Exploratory Data Analysis (EDA):
    - Conducted using pandas and Matplotlib in Python to understand customer demographics, income distribution, and purchasing patterns in the Japanese market.
2. Logistic Regression:
    - A logistic regression model was developed on the Japanese dataset with the dependent variable being the Purchase (Yes/No). Independent variables included age, gender, annual income, and previous car age.
    - Different cutoff values were tested (0.5, 0.4, 0.3, 0.2). The cutoff of 0.2 yielded the highest classification accuracy, meaning it was the most effective in predicting whether a customer would purchase a car.
3. Multiple Linear Regression:
    - The following formula was used to predict the number of purchases in the Indian market:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \ldots + \beta nXn$$

Where Y represents the dependent variable (purchases), and X1, X2, ... Xn represent independent variables namely CURR_AGE, GENDER, ANN_INCOME, and AGE_CAR (segmented). This model was trained on the Japanese data and then applied to the Indian dataset to predict sales.

$\beta_0$ is the intercept
$\beta_1, \beta_2, \beta_3, \beta_4$ are the intercept coefficients of the independent variables

$$Purchase = \beta_0 + \beta_1 Curr\_Age + \beta_2 Gender + \beta_3 Ann\_ Income + \beta_4 Age\_car(segmented)$$

**Significant Findings:**

- The logistic regression model showed that annual income and age were the most significant factors influencing purchasing behavior in the Japanese market.
- Gender had a lesser but still notable impact on purchasing decisions.
- The predicted number of sales in the Indian market was 29,735, significantly higher than the target of 10,000.

# Result and Conclusion

*The analysis predicts 29,735 car sales in the Indian market within the first year, well above ABG Motors' initial target of 10,000 units. This suggests strong potential for success in India, particularly if the company can replicate the conditions that led to purchasing decisions in Japan.*

**Summary and Recommendations for Action:**

1. Aggressive Entry Strategy: Given the predicted demand, ABG Motors should pursue an aggressive market entry strategy, with a focus on high-income urban areas where car ownership is more prevalent.
2. Targeted Marketing: Tailor marketing efforts to customers aged 30–45 with higher annual incomes, as these demographics were identified as the most likely to make a purchase.
3. Product Offering: Consider pricing and features that appeal to the Indian middle class, while maintaining brand value.
4. Further Testing: It would be advisable to further validate this model with more detailed Indian data, as additional factors such as cultural differences may impact purchasing behavior.

**This analysis was performed using:**

- Microsoft Excel 2021 (v16.0) for data wrangling,
- pandas 2.2.2 and Matplotlib 3.9.2 in Python 3.10.6 for exploratory data analysis (EDA),
- Tableau 2024.1.3 for data visualization.