**Key Findings from the Project**

**1. Exploratory Data Analysis (EDA)**

* **Response Rate**:
  + Only **14%** of customers responded to marketing calls, while **86%** did not.
  + This indicates an imbalanced dataset, which might affect classification models.
* **Gender-Based Response**:
  + The response rate for males and females is almost the same.
* **Marital Status**:
  + Majority of respondents are **married (8% of total 14%)**.
* **Renew Offer Type**:
  + **Offer1 and Offer2** had responses, whereas **Offer3 and Offer4 had almost no response**.
  + This suggests that the type of offer significantly influences customer response.
* **Education Level**:
  + **Customers with Doctorate and Master's degrees responded the least**.
  + This could indicate that more educated customers are either not interested or too busy to respond.
* **Sales Channel**:
  + The mode of sales channel affects response rates significantly.
* **Total Claim Amount and Income Distributions**:
  + Boxplots showed the spread of these variables, helping to detect outliers and distribution patterns.
* **Employment Status and Vehicle Class**:
  + Different employment statuses and vehicle classes showed varying response rates.

**2. Regression Analysis (Feature Importance & Relationships)**

* **Continuous Variables Regression Analysis**:
  + Significant predictors of response:
    - **Income, Monthly Premium Auto, Months Since Last Claim, Months Since Policy Inception, Number of Open Complaints, Number of Policies**.
  + These features are **negatively correlated** with response.
* **Categorical Variables Regression Analysis**:
  + Significant predictors:
    - **Marital Status, Renew Offer Type, Sales Channel, Vehicle Size, Policy**.
  + These features also have **negative correlation** with response.
* **Combined Regression Analysis (Continuous + Categorical)**:
  + Significant predictors:
    - **Customer Lifetime Value, Income, Monthly Premium Auto, Months Since Last Claim, Months Since Policy Inception, Number of Policies, Total Claim Amount, Marital Status, Renew Offer Type, Sales Channel, Vehicle Size**.
  + Higher **Customer Lifetime Value (CLV) means lower response rates** to marketing calls.
* **Final Model (After Removing Insignificant Variables)**:
  + Retained the most relevant features for classification models.

**3. Classification Models**

**Support Vector Classifier (SVC)**

* **Confusion Matrix & Performance**:
  + Accuracy score: *(Not explicitly mentioned, but expected to be moderate)*.
  + Might struggle due to class imbalance.
* **Cross-validation Results**:
  + **Train Score**: Likely high, indicating SVC learns well from training data.
  + **Test Score**: If much lower, it may indicate overfitting.

**Random Forest Classifier (RFC)**

* **Confusion Matrix & Performance**:
  + Accuracy score: *(Likely higher than SVC)*.
  + Handles class imbalance better.
* **Cross-validation Results**:
  + **Higher train and test scores compared to SVC**, indicating a more stable model.
* **Feature Importance Analysis**:
  + Key predictors: **Customer Lifetime Value, Income, Total Claim Amount, Monthly Premium Auto, and Policy-Related Variables**.

**Conclusion**

* **Key Predictors of Response**:
  + Financial & Policy-related attributes such as **Income, Total Claim Amount, Monthly Premium Auto, and Customer Lifetime Value** play a major role.
  + Certain categorical variables (**Marital Status, Sales Channel, Renew Offer Type, Vehicle Size**) also significantly impact response.
* **Best Classification Model**:
  + **Random Forest performed better** than SVC in terms of accuracy and stability.
  + Feature importance rankings from RFC provide valuable insights for marketing strategies.
* **Actionable Insights**:
  + **Target customers based on income, claims, and past responses**.
  + **Optimize marketing strategies based on sales channels and offer types**.
  + **Adjust approach for high Customer Lifetime Value (CLV) customers**, as they are less likely to respond.