CASE STUDY 1 – DEEP LEARNING

O2R2 Telecom: Expanding with Data-driven Marketing

Cours

Instru

sub

Student Names	Contribution
Abhi Rajeshbhai Patel	Performed basic Exploratory Data Analysis, Handled Data Imbalance. Worked on Data Preprocessing and Data Imbalancing Slides.
Ajit Gurung	Performed Model Tuning
Chinmay Hiteshkumar Parmar	Built Logistic Regression and Deep Neural Network
Jaivin Jacob	Data Preprocessing
Jasleen Kaur	Built Logistic Regression Model
Kanchan Laxmi Rai	Feature Engineering – Created Features, Performed Feature Binning and Trained Logistic Regression model on that data. Worked on Presentation Model Comparison Slides.
Kritika	Derived Insights from plotted graphs and Worked on Overview, Key Findings, Introduction slides of the presentation and Business Proposals slides
Kuldip Hareshbhai Mangrola	Model Comparison, Plotted AUC,ROC and Execution time graphs
N W Mudiyanselage Rajantha Tikiri Bandara Samarakoon	Built Logistic Regression Model

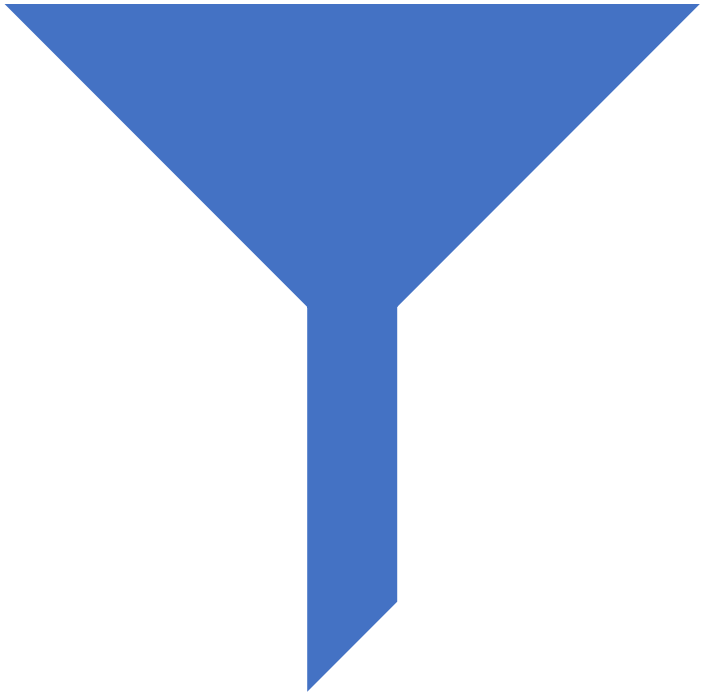
INTRODUCTION

- O2R2 Mobile is a leading telecom company with a focus on expanding its customer base through a new wellness plan.

- The product team aims to introduce a new offering, the wellness plan and wants to leverage data to make market more efficient. They want to predict potential customers for the wellness plan through predictive analysis for wellness plan

OVERVIEW

- The primary goal is to predict which customers are more likely to purchase the newly introduced Wellness Plan.
- It is to enhance customer lifestyle and adapting to changes and ensuring growth of the telecom sector.
- Currently, O2R2 Mobile offers 5 distinct Plans: Basic, Standard, Deluxe, Super Deluxe and King. Observations from previous year’s data reveal’s that approximately 18% of customers opted for these existing plans.

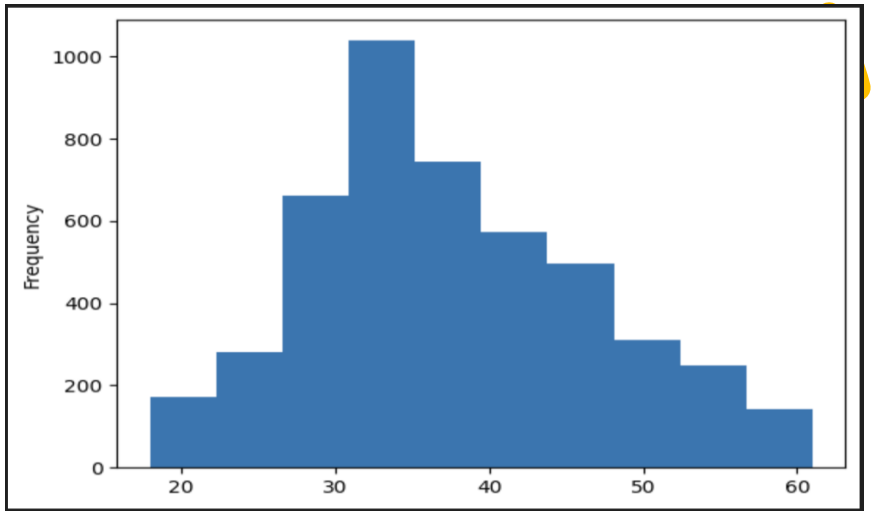


LIMITATIONS

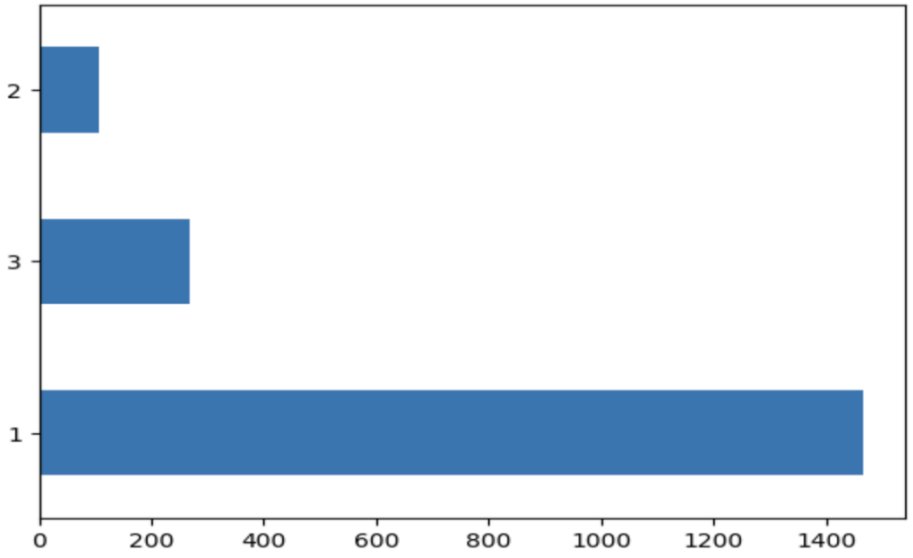
The previous campaign had limitations such as random customer contact. This approach is inefficient and has low conversion rates, highlighting the need for a more targeted and data driven market strategy.

KEY FINDINGS AND INSIGHTS OF DATA

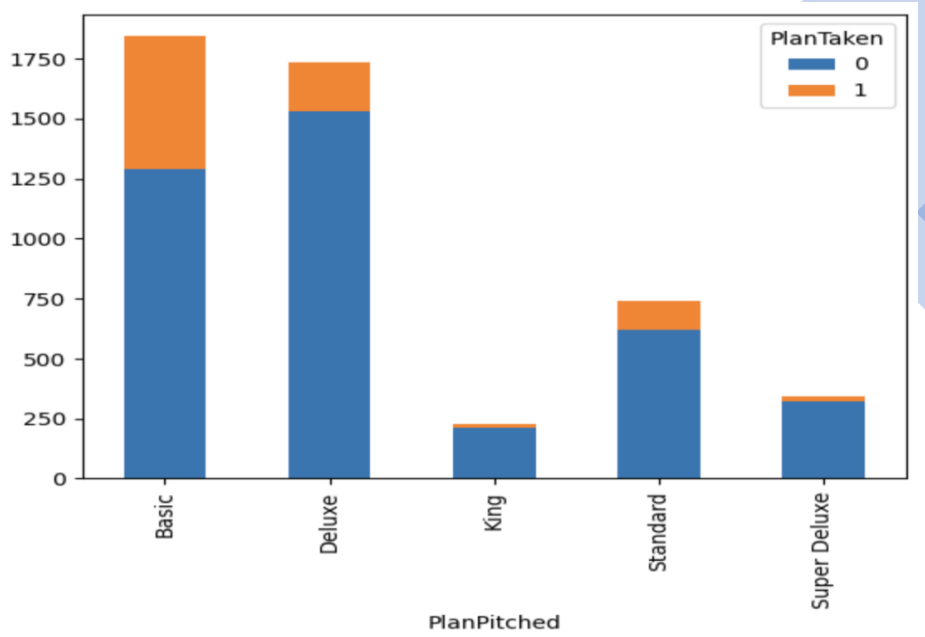
Age distribution – The analysis shows that we have more customers between age of mid-30s to mid-40s. With this information we can target this age group and make marketing strategies according to it.



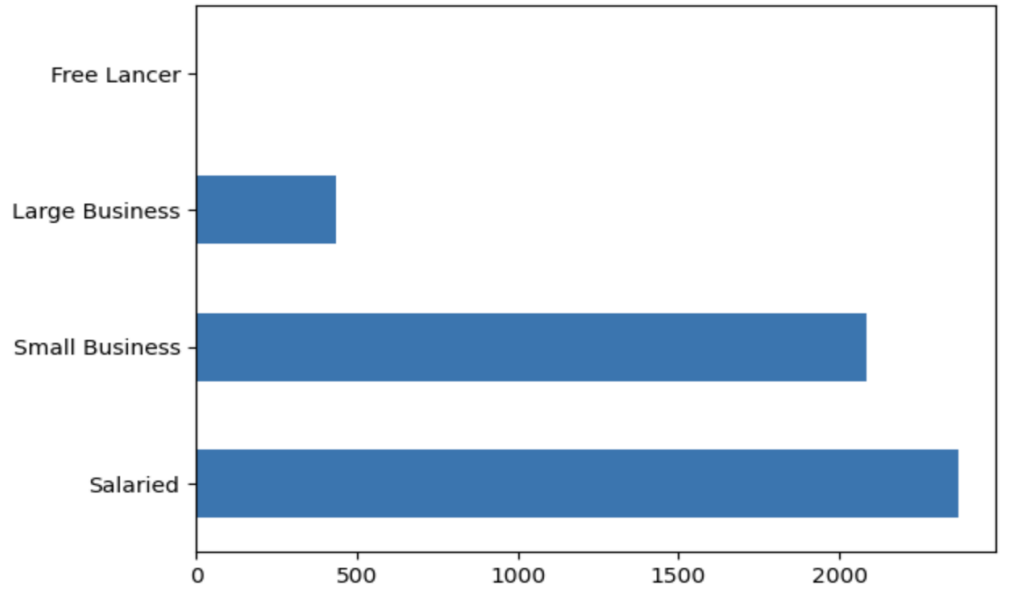
City Tier – The analysis shows that we have 3 city tier which are Tier1, Tier 2 and Tier 3. With this we can understand customer preference in all 3 tier, and we can customize plan for the targeted customers.



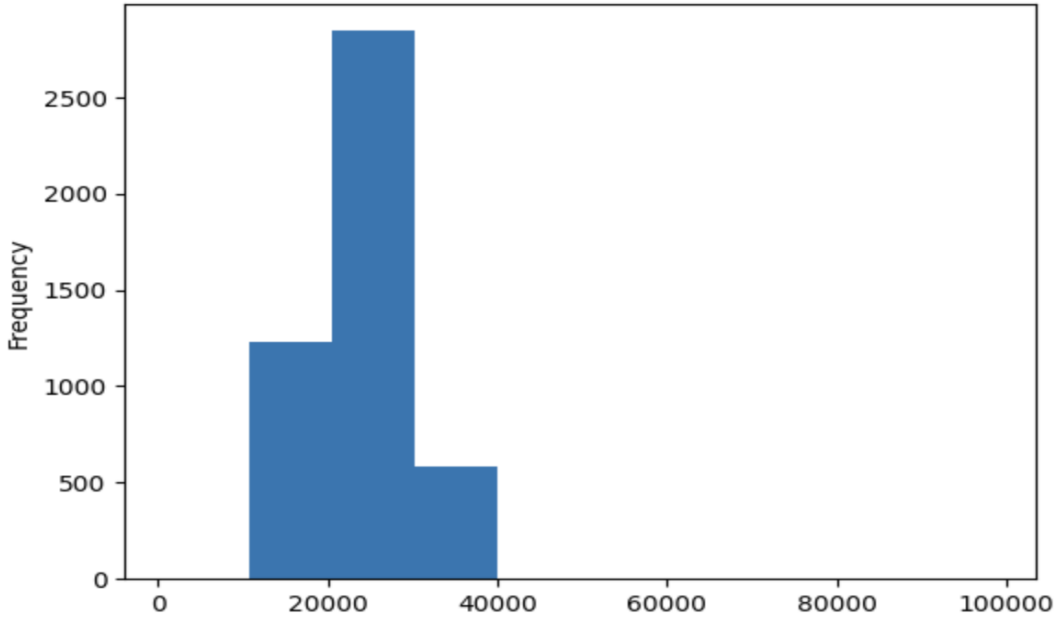
Plan Pitched – With this we can see positive correlations observed between sales pitch satisfaction scores and plan purchase. Here orange color shows how many plans were purchased after the pitch.



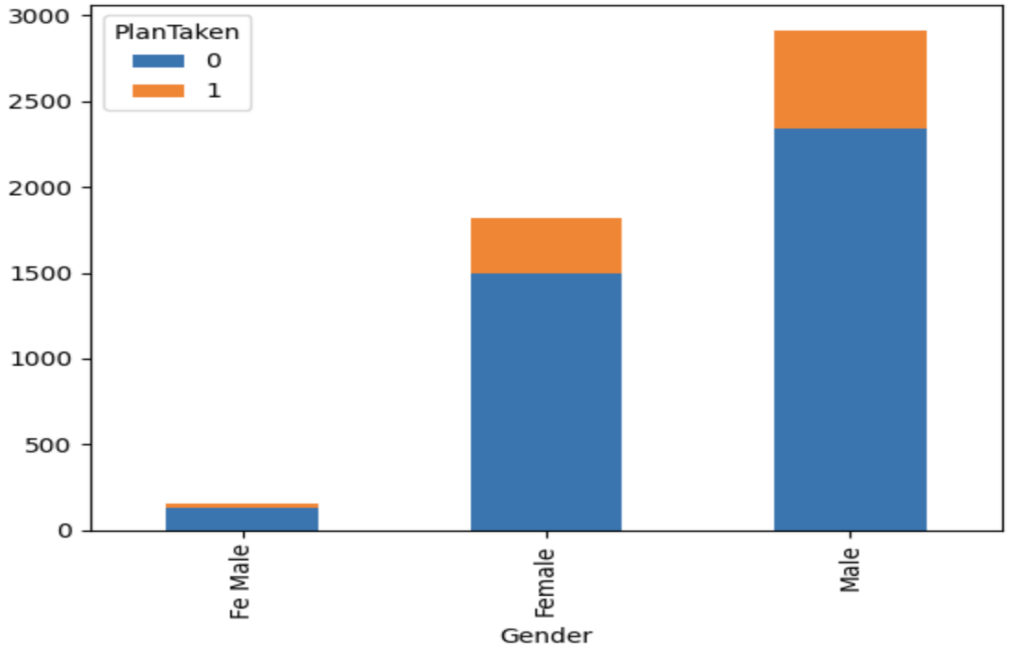
Occupation – Different occupation have different preferences for the plans so if we identify patterns among these occupations, we can tell the specific offerings according to occupation.



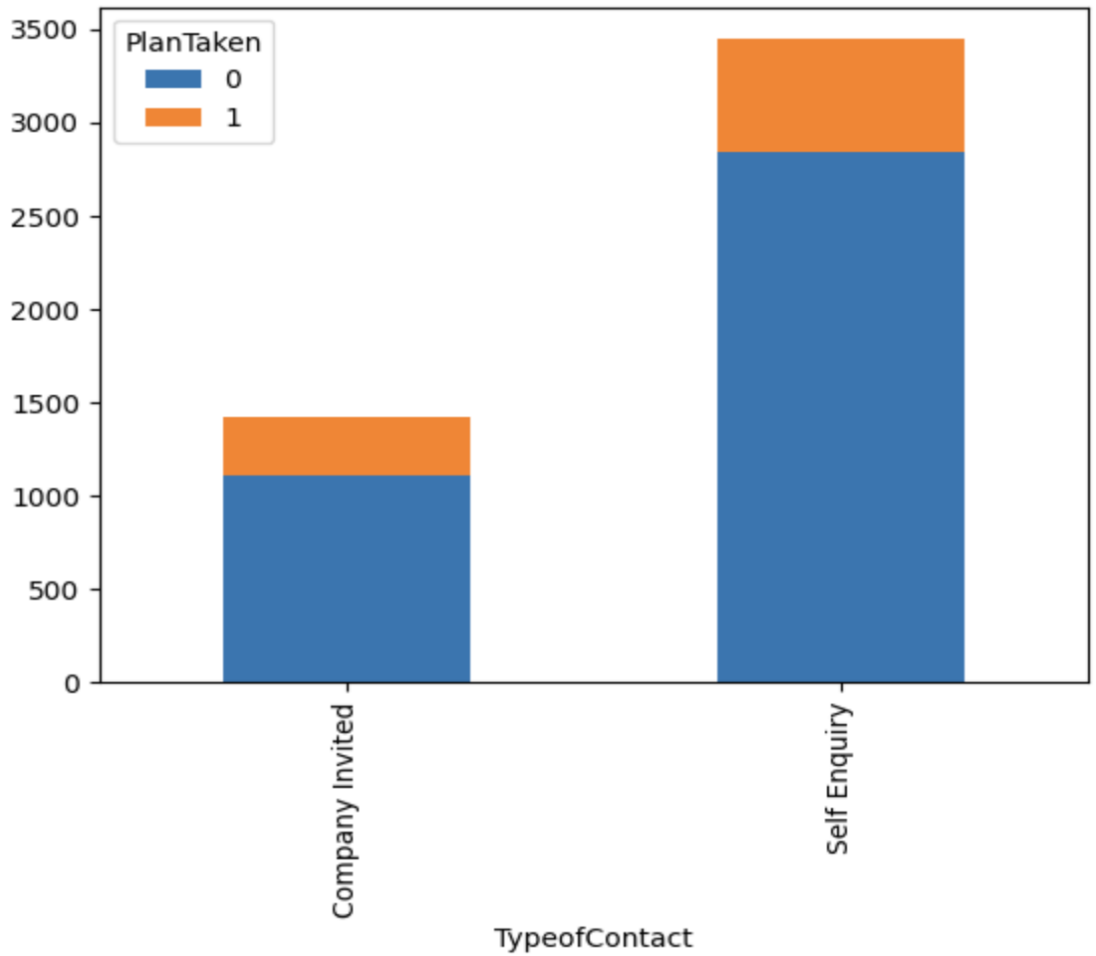
Monthly Income – By observing monthly income we can see that the customers choose plans according to their monthly income so if we want to introduce the new plan, we can have discounts so that most of the customers can afford it.



Gender – Based on the gender we may reveal insights that which customer is opting for more plans. This can be used to target the specific gender more and to develop gender specific plans.

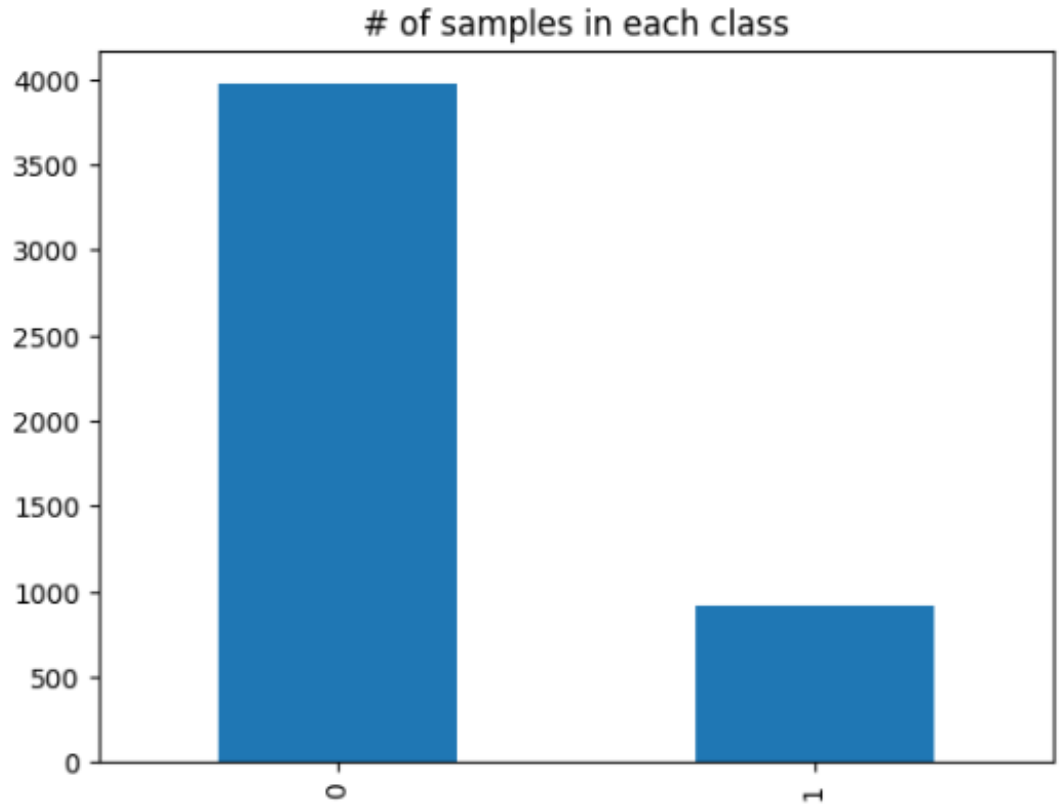


Type of Contact – With this we can analyze that customers reached through Self Inquired tend to have different preferences compared to those contacted via Company Invited. We can also how many plans were taken after the contact. Here orange color is used plan taken.



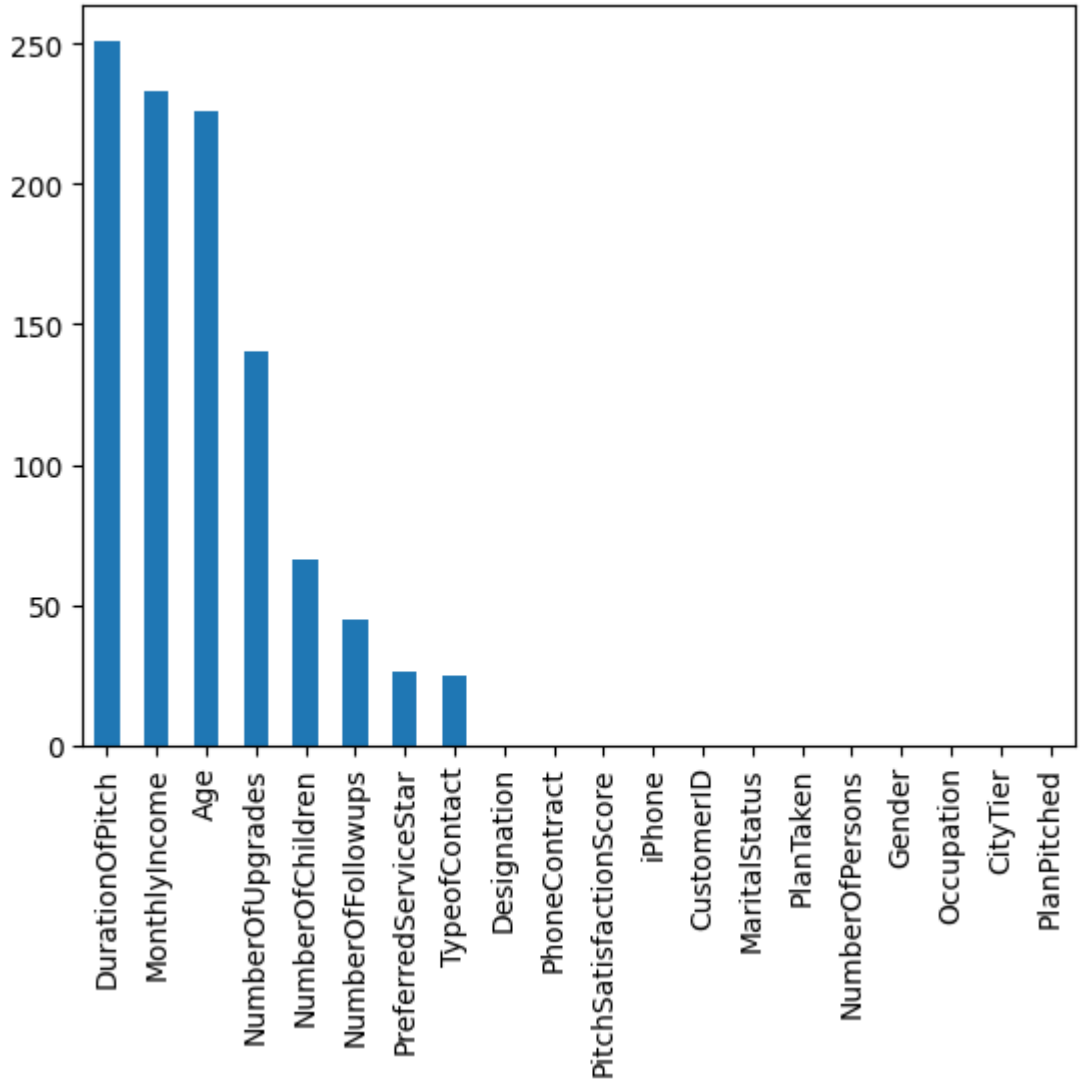
DATA PREPROCESSING – HANDLING IMBALANCES

- The target variable “PlanTaken” is highly **imbalanced**. The “Not Plan taken” class has **4 times** more sample than “Plan Taken” class.
- To solve the imbalances, we performed oversampling on the “Bought” class.



DATA PREPROCESSING – MISSING VALUES

- Many of the features such as DurationOfPitch, MonthlyIncome, Age, etc. have more than 100 missing values.
- For features, having normal distribution we used **mean** to fill the missing values.
- For features, having skewed distribution, we used **median** to fill the missing values.
- And, we used **mode** to fill the missing values of the categorical columns.



MODEL – LOGISTIC REGRESSION

Parameters	All	Features Selected(Age (Binned), MonthlyIncome, PlanPitched, Occupation, Designation)
F1 Score	“Bought”: 0.40 “Not Bought”: 0.9	“Bought”:0.42 “Not Bought”: 0.69
Recall	“Bought”: 0.29 “Not Bought”: 0.96	“Bought”: 0.62 “Not Bought”: 0.58
Training Accuracy	0.83	0.61
Testing Accuracy	0.83	0.59
Oversampled?	No	Yes

MODEL – DEEP NEURAL NETWORK

Parameters	All	Features Selected(Age (Binned), MonthlyIncome, PlanPitched, Occupation, Designation)
F1 Score	“Bought”: 0.57 “Not Bought”: 0.84	“Bought”:0.44 “Not Bought”: 0.74
Recall	“Bought”: 0.67 “Not Bought”: 0.80	“Bought”: 0.60 “Not Bought”: 0.65
Training Accuracy	0.77	0.64
Testing Accuracy	0.77	0.59
Oversampled?	Yes	Yes
Network Architecture	8->16->1	8->16->1

BUSINESS RECOMMENDATION

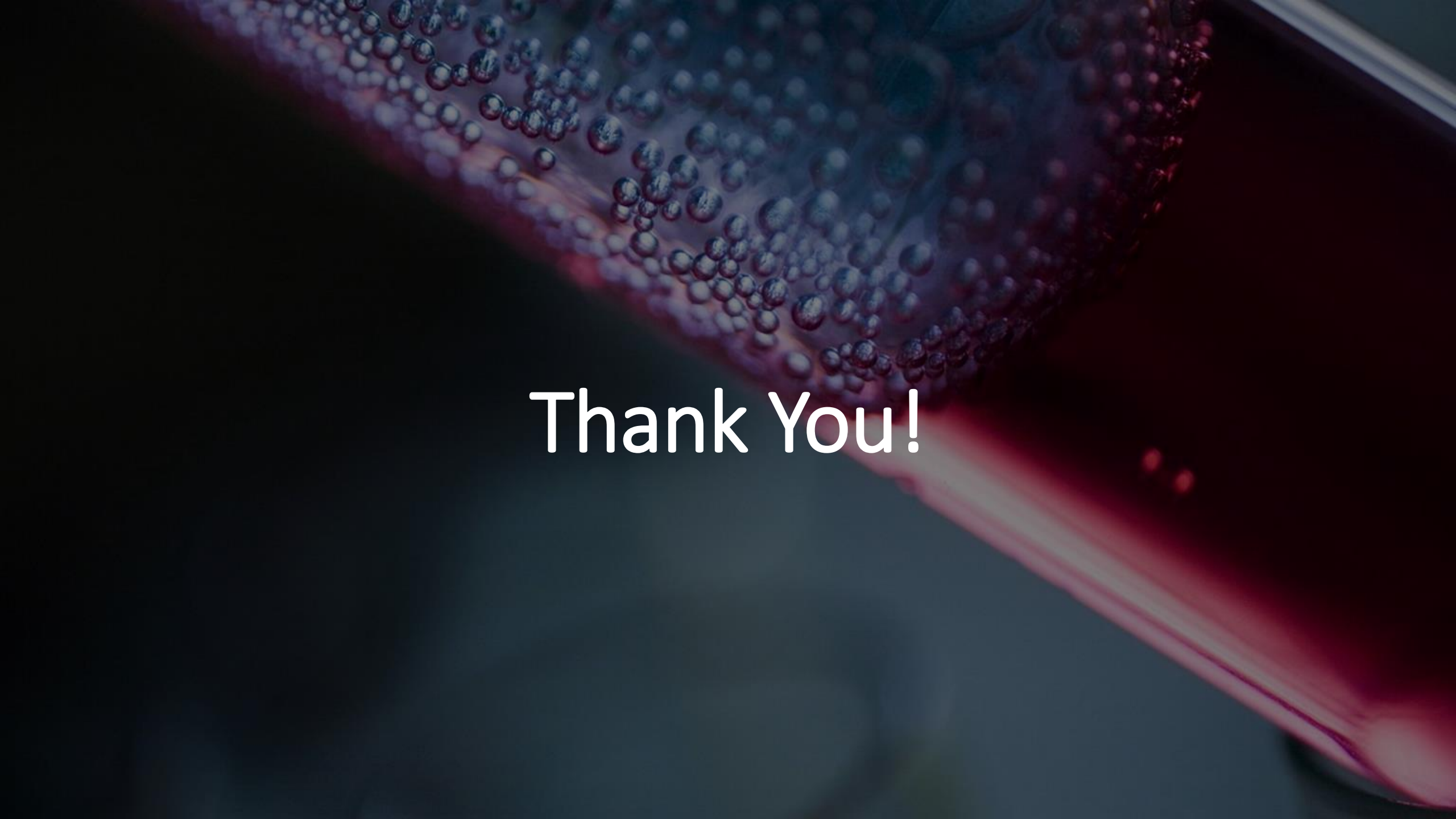
1. Enhance customer experience by identifying preferences using the model. Giving Discounts or exclusive offer to customer so that more customers can afford it and more customer can take the plan.

2. Collect feedback from customer who have opted for Wellness Plan. Identify successful patterns in sales pitches to make a more effective pitch for the future.

3. Make a marketing plan with which it will reach most of the targeted customers. Such as Married People having Kids are more likely to buy.

CONCLUSION

- Deep learning model has performed better than logistic regression


Thank You!