

Introduction

Project Overview

Project Details

Contact



TRAVEL PACKAGE PREDICTION

MIRANDA KHAIRUNNISA

Learn More →



Executive Summary



Background:

Travel packages are popular for their convenience, cost savings, and curated experiences, yet still not universally appealing. Travel agencies often struggle to identify likely buyers, leading to inefficient marketing and missed opportunities. This project leverages a travel package prediction dataset to analyze customer behavior, enabling travel agencies to create personalized recommendations, target campaigns, effectively, and enhance customer satisfaction. The insights aim to maximize revenue, improve resource allocation, and foster loyalty.

Problem Statement:

Inefficient marketing can lead to wasted budgets and missed opportunities. Thus, understanding purchase drivers is crucial for travel agencies to optimize their offerings and marketing strategies.

Business Understanding

Data Collection & Preparation

Data Analysis & Exploration

Modeling

Recommendations

Goals & Objectives:

- Identify factors affecting customers purchasing behaviour.
- Predict customer likelihood of purchasing travel packages.



Key Insights:

- Product Pitched is the most influential feature. The type of travel package offered strongly influences customer purchasing choices.
- City Tier plays a role in differentiating customer preferences, highlighting the need to tailor marketing efforts based on their demographic location.
- Number of Follow-ups suggests that regular engagement with potential customers enhances the likelihood of a purchase.



Data Science Workflow Overview



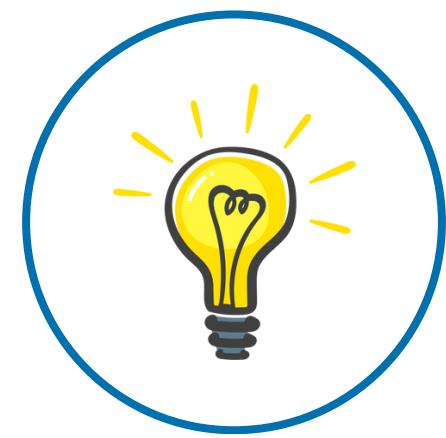
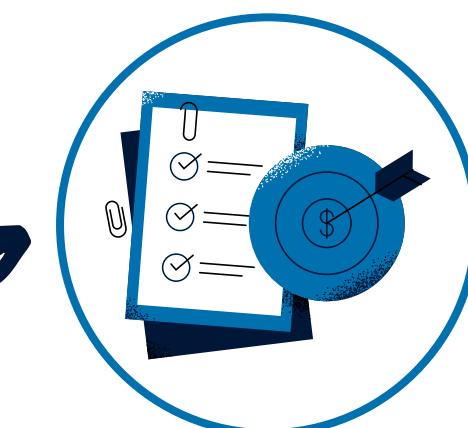


Business Understanding

Goal: Determine customer profile and purchasing behaviour



Problem Statement: Travel agencies struggle to identify potential customers, leading to wasted marketing efforts and poor customer experiences



Objective: Develop a prediction model to predict the probability of travel package purchase based on customer profile



Performance Metrics: precision, recall, F1-score, ROC-AUC

Learn More →

Introduction

Project Overview

Project Details

Contact



4888, 20

ROWS

COLUMNS

Data Information

19 FEATURES

Customer Profile

- Customer ID
- Age
- City Tier
- Occupation
- Gender
- Number of Person Visiting
- Preferred Property Star
- Marital Status
- Number of Trips
- Passport Ownership
- Car Ownership
- Number of Children Visiting
- Customer Designation
- Monthly Income

Customer Interaction

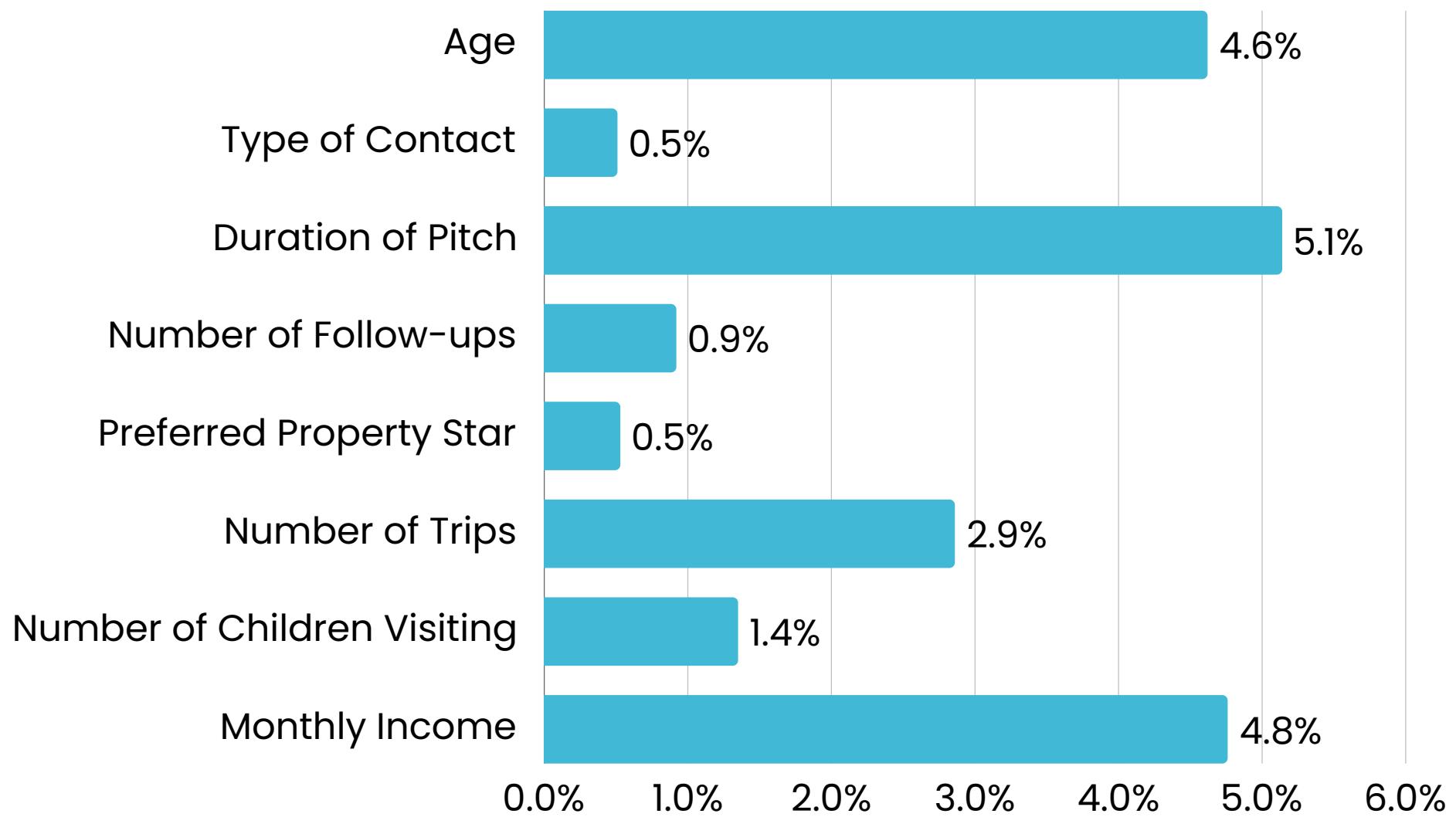
- Type of Contact with customer
- Duration of Pitch
- Number of Follow-ups
- Product Pitched
- Pitch Satisfaction Score

1 TARGET

ProdTaken
(whether the product was purchased or not)

Data Quality Overview

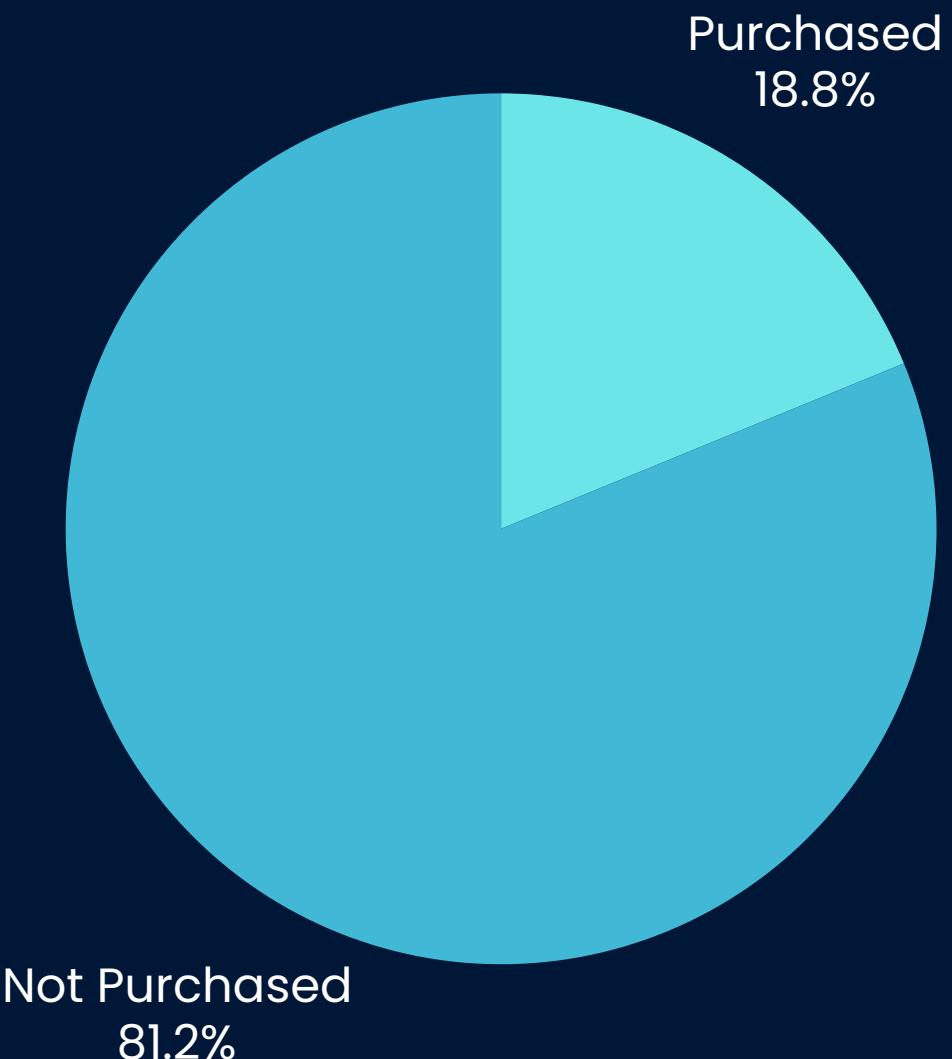
Percentage of Missing Data



141 (2.8%) Data Duplicates

Exploratory Data Analysis

Proportion of Purchased Travel Packages



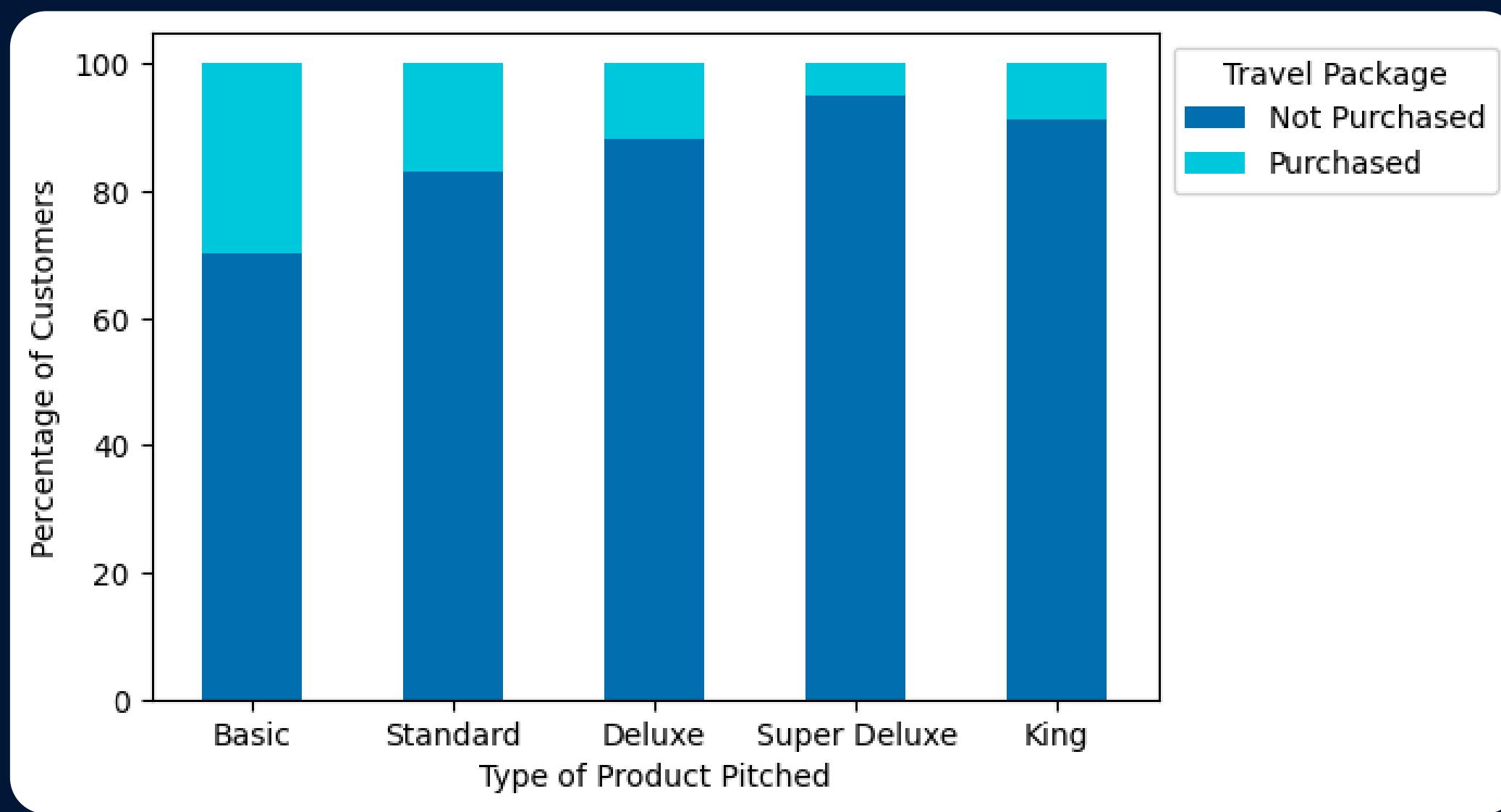
There are less people who purchase the travel package than those who do not. Therefore, key factors such as:

- customer profiles
- demographics
- background
- customer engagement

driving purchase behaviour need to be identified.

Exploratory Data Analysis

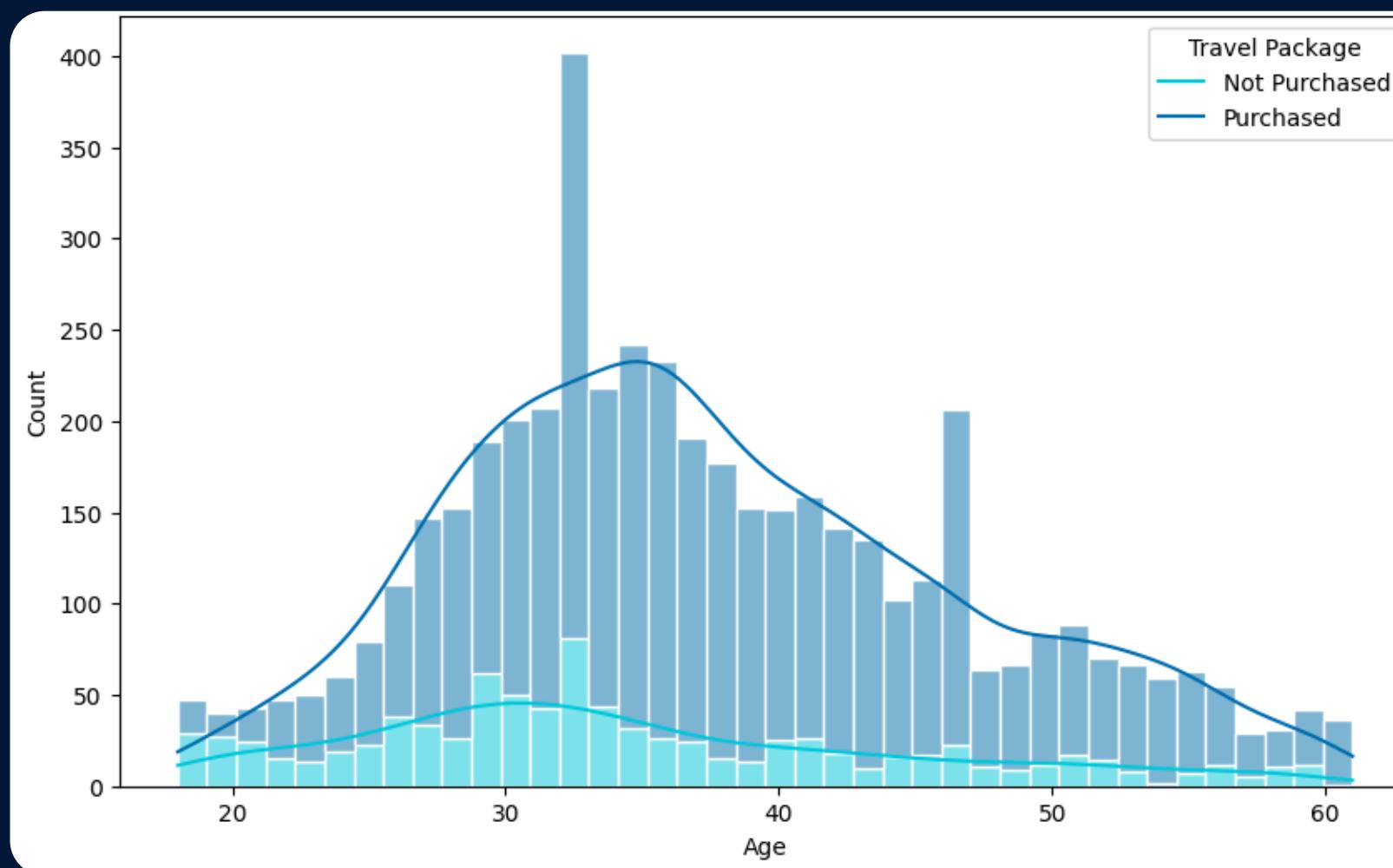
Percentage Distribution of Travel Package Purchased by Product Pitched



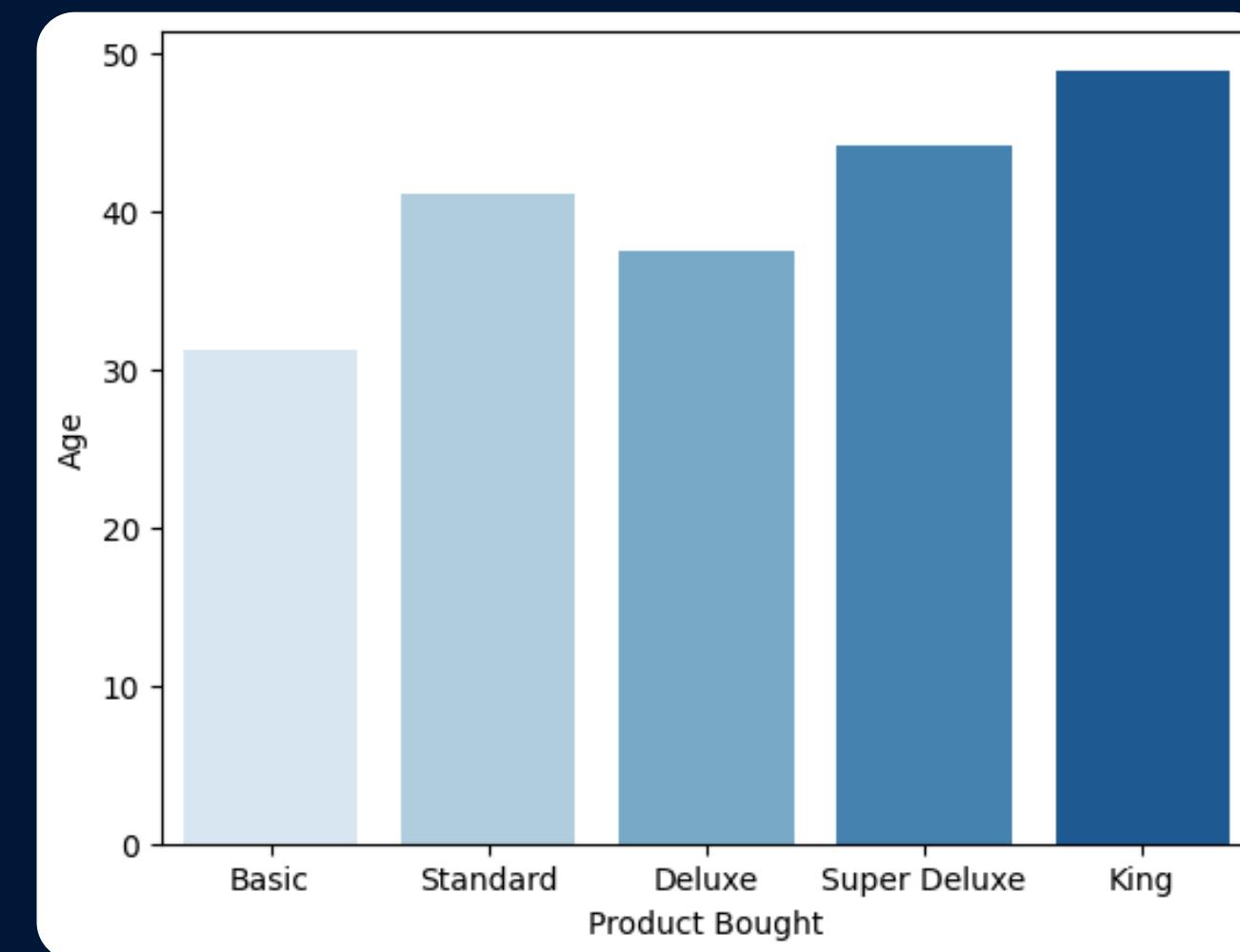
- More customers purchased the travel package when pitched the basic product, compared to the Super Deluxe and King options
- The basic package resonates more with the majority of customers, potentially due to its perceived affordability and suitability for a wider audience

Exploratory Data Analysis | Customer Age Insights

Age Distribution



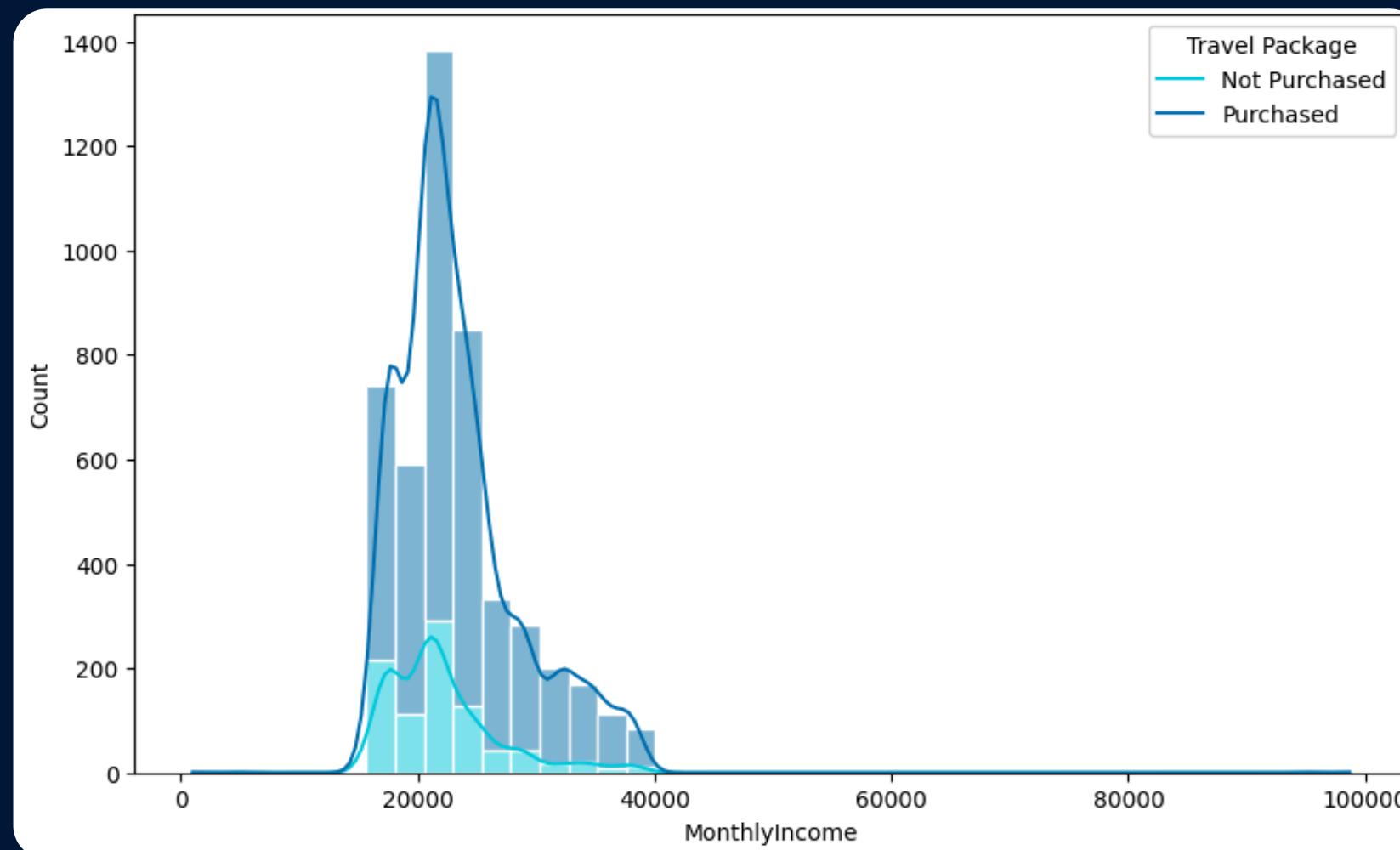
Distribution of Mean Age by Product Type



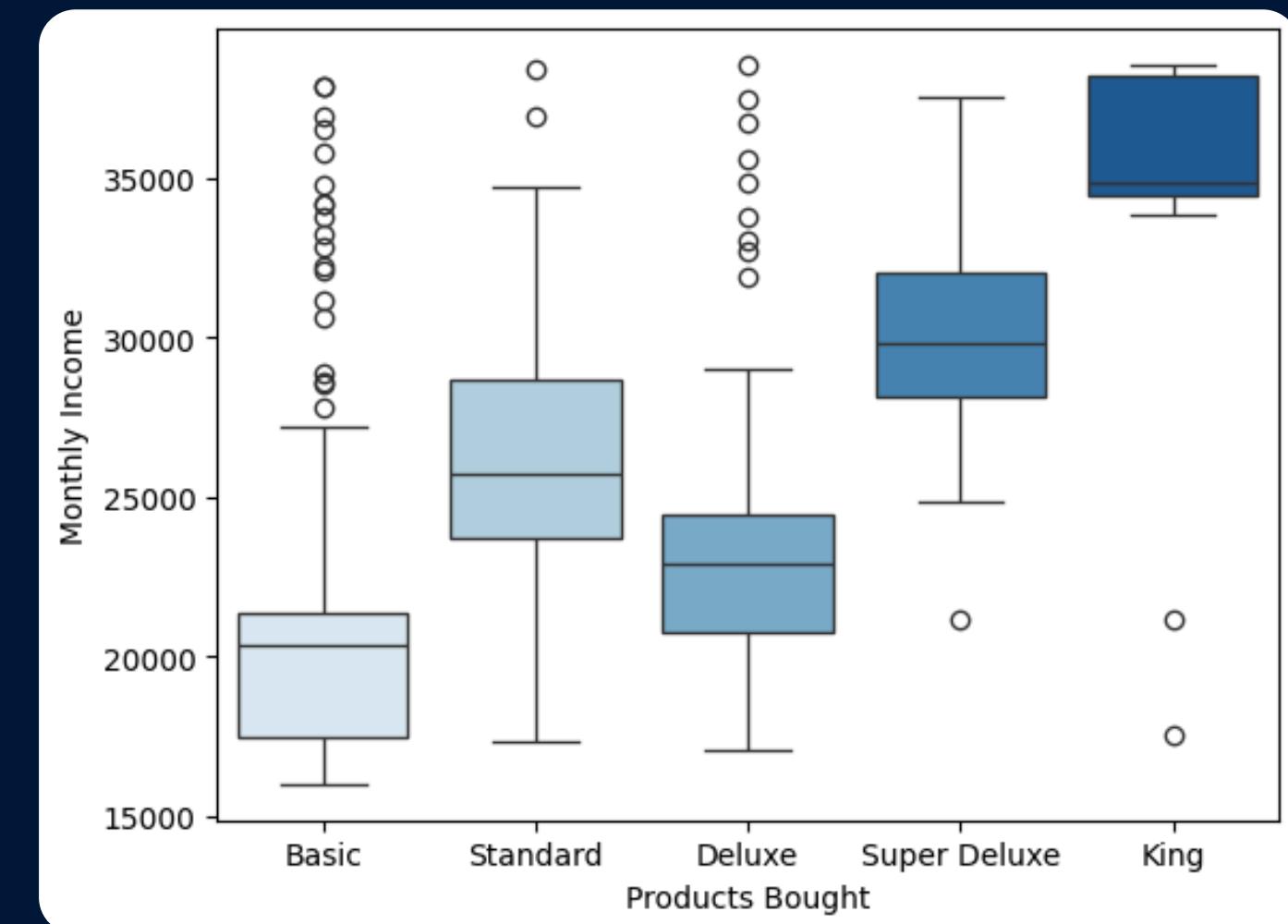
- Most purchases occur between 25–40 years, peaking around 30 years.
- Product preference shifts with age. Higher-tier packages appeal to older customers, while younger customers favor budget-friendly options.

Exploratory Data Analysis | Customer Income Insights

Monthly Income Distribution



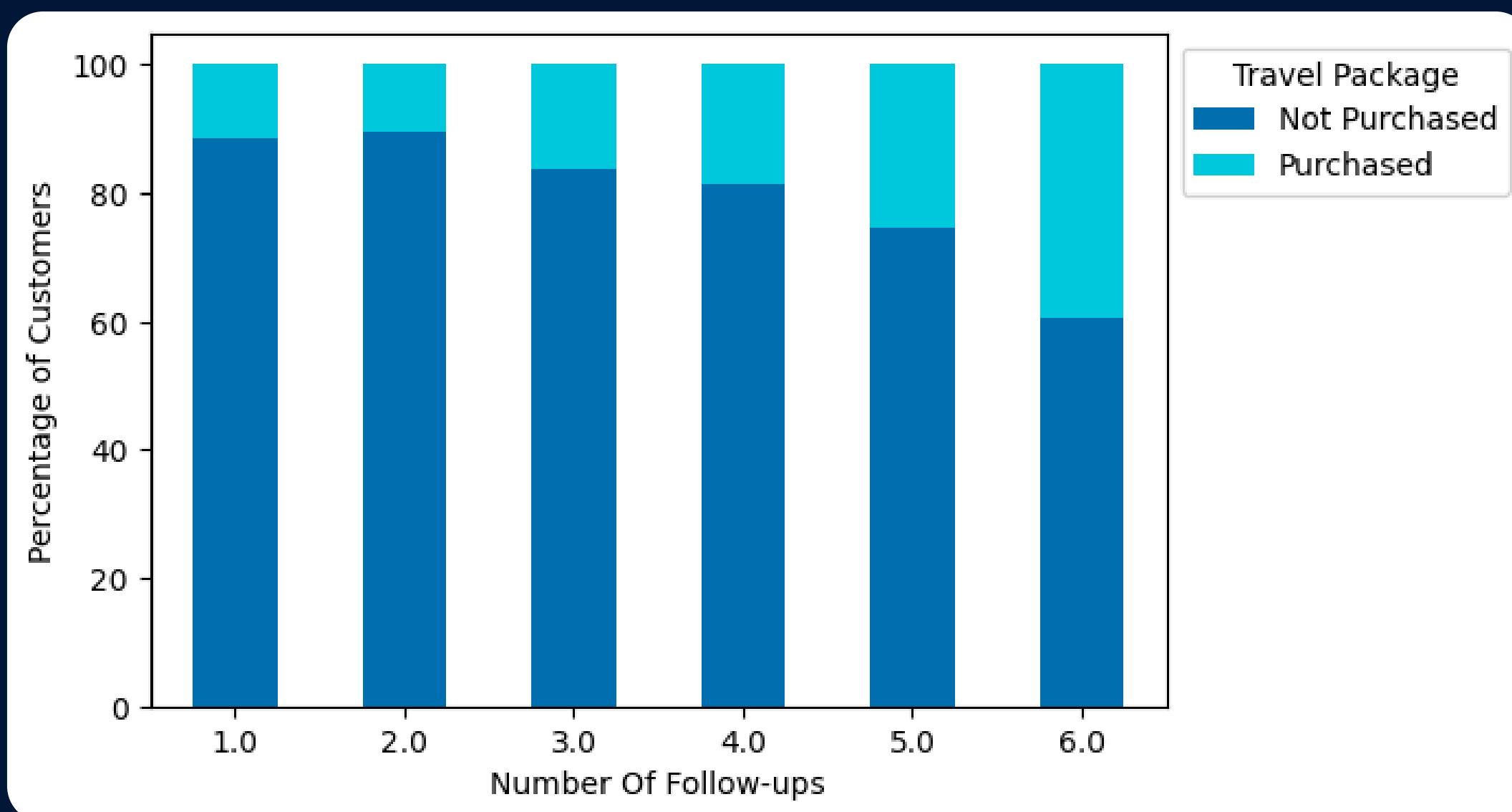
Boxplot of Monthly Income by Product Type



- Income distribution is concentrated between \$15,000–30,000, peaking at \$20,000–25,000.
- Higher-tier products align with higher income. Product preferences correlate with income, supporting segmentation by purchasing power.

Exploratory Data Analysis

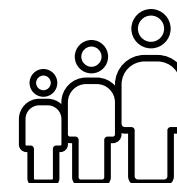
Percentage Distribution of Travel Package Purchased by Number of Follow-ups



- Customers with a higher number of follow-ups were more likely to purchase a travel package.
- Increased engagement from the sales team positively impacts customer decisions, indicating that persistent follow-up efforts may play a significant role in converting potential leads into actual buyers.

Customer Profile

Basic



20-30

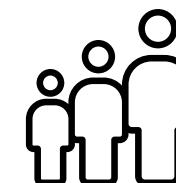


<20000

Free
Lancers,
Salaried,
Executives

City Tier 3

Standard



31-40

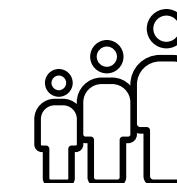


25000-30000

Small
Business
Owners,
Managers

City Tier 3

Deluxe



31-40

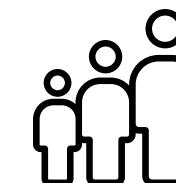


20000-25000

Small
Business
Owners,
Senior
Manager

City Tier 3

Super Deluxe



41-50



30000-35000

Salaried,
AVPs

City Tier 3

King



51+



35000-40000

Small
Business
Owners,
VP

City Tier 1

Data Pre-Processing



Data Cleaning

*handling duplicated
and missing data*

Feature Selection & Engineering

encoding categorical variables & checking multicollinearity



Train-Test Split

80% Train: 20% Test



Outlier Handling

IQR Method

Model Selection

Model	Data	Metrics			
		Precision	Recall	F1-Score	ROC-AUC
Decision Tree	Train	100.00%	100.00%	100.00%	100.00%
	Test	71.36%	72.08%	71.72%	82.26%
Random Forest	Train	100.00%	100.00%	100.00%	100.00%
	Test	89.31%	59.39%	71.34%	94.79%
XGBoost	Train	100.00%	99.69%	99.84%	100.00%
	Test	87.97%	70.56%	78.31%	93.98%
LightGBM	Train	99.49%	92.04%	95.62%	99.89%
	Test	88.32%	61.42%	72.45%	91.60%

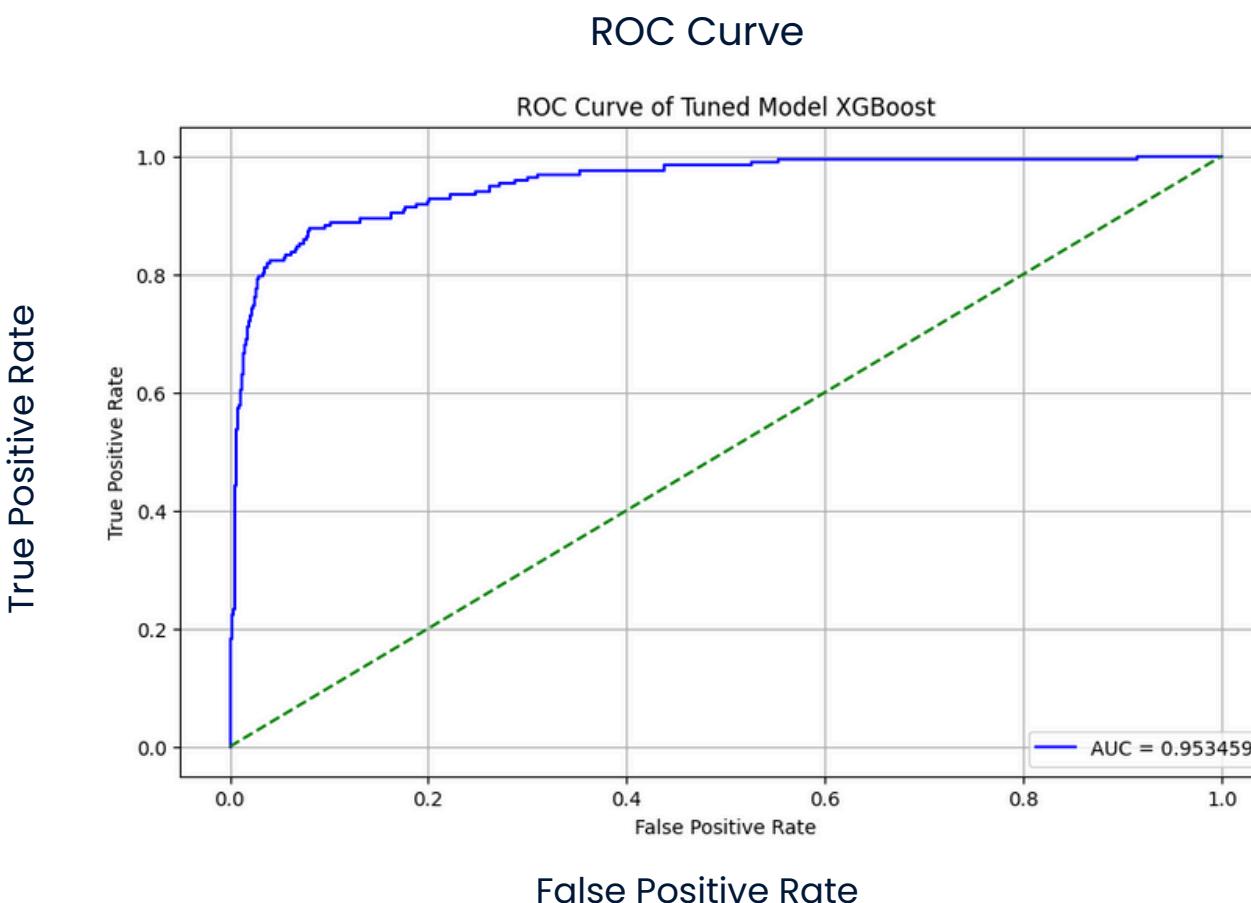
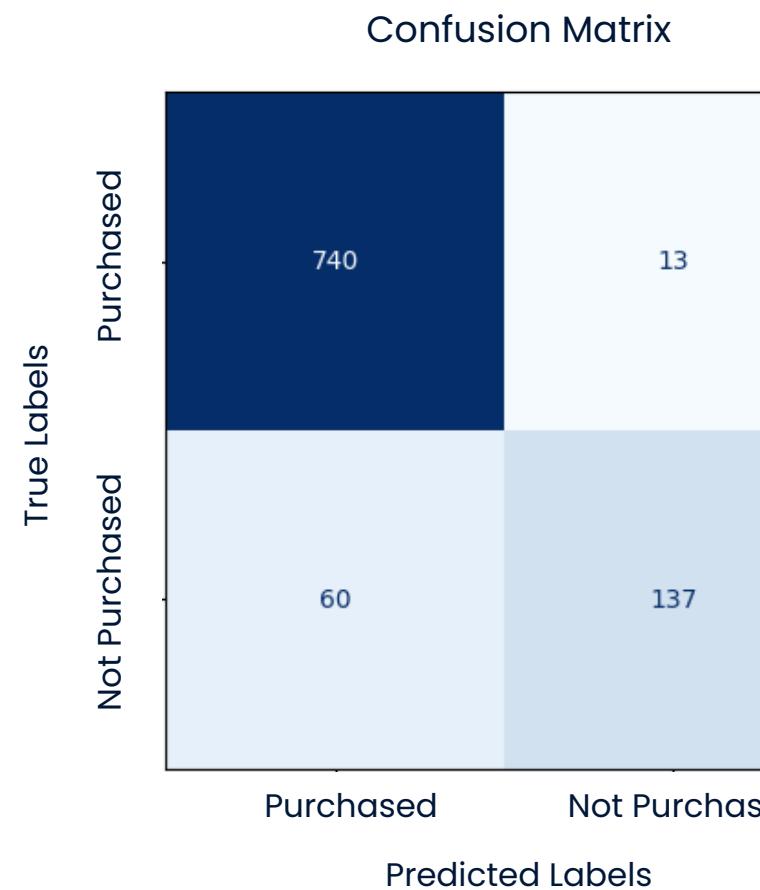
XGBoost offers strong generalization and high F1-score on the test data (78.31%) and a fairly decent ROC AUC (93.98%), which indicates a good balance between precision and recall and a good ability to distinguish between true positives and negatives.

Optimizing XGBoost Parameters

- **Improved Metrics:** Tuned XGBoost outperforms the base model across all metrics
- **SMOTE Impact:** SMOTE likely introduces noise from synthetic samples
- **Best Model:** The Tuned XGBoost Model (without SMOTE) delivers the highest and most balanced performance

Model	Metrics			
	Precision	Recall	F1-score	ROC-AUC
Base XGBoost	87.97%	70.56%	78.31%	93.98%
Tuned XGBoost (without SMOTE)	91.33%	69.54%	78.96%	95.35%
Tuned XGBoost with SMOTE	86.21%	63.45%	73.10%	93.30%

Model Analysis

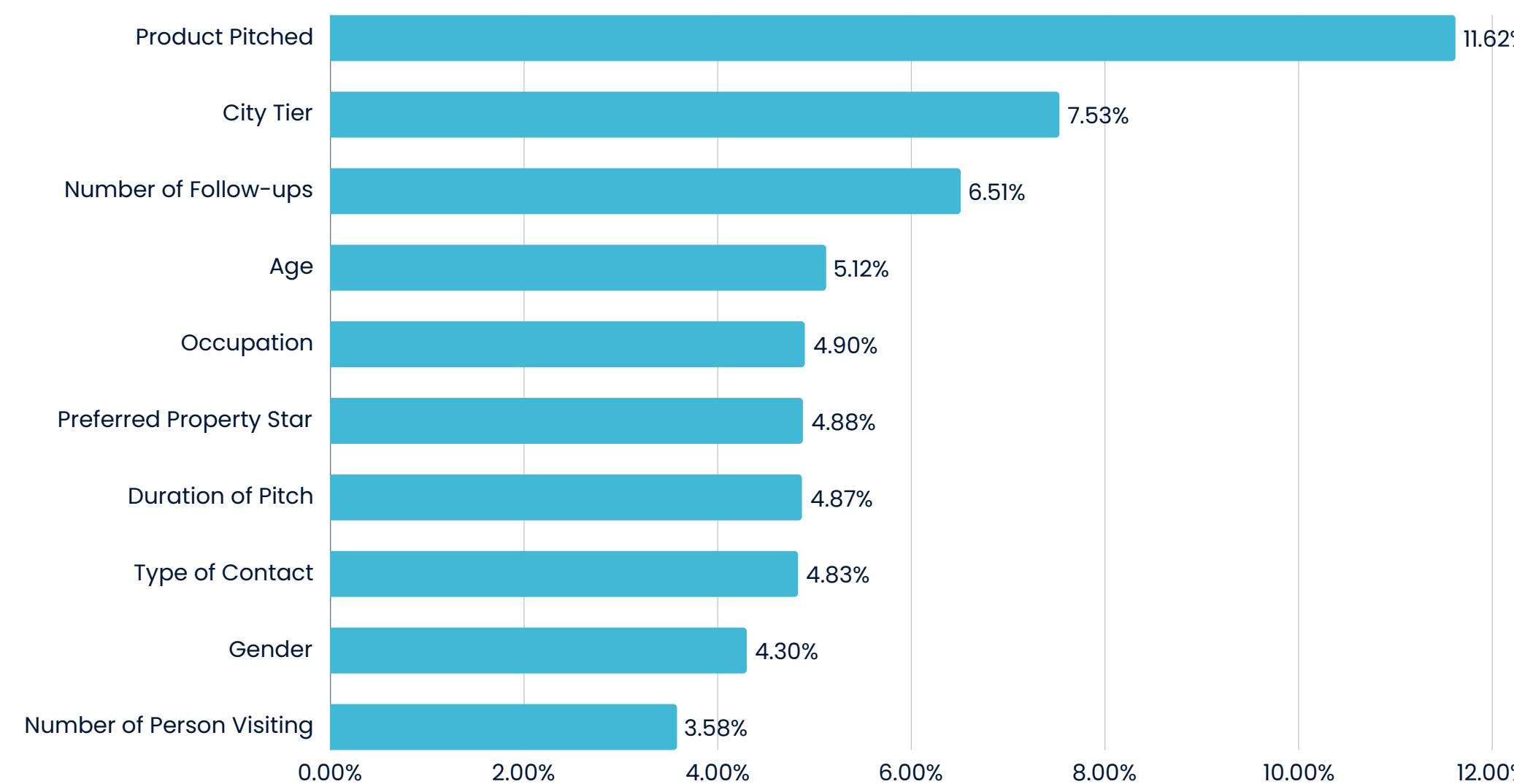


- **True Positive (TP):** 740 customers were correctly identified
- **False Positives (FP):** 60 customers were predicted to buy but did not
- **True Negatives (TN):** 137 customers were predicted to buy and actually did not
- **False Negatives (FN):** 13 customers were correctly predicted not to buy

The model has an AUC of 95.35% demonstrating that the model effectively distinguishes between customers likely to purchase and those who aren't.

Model Analysis

Top 10 Most Important Features



Based on the model:

- The type of package offered significantly impacts customer purchasing decisions.
- Demographic location affects customer preferences.
- Consistent engagement increases the likelihood of purchase.
- Occupation provides insight into customer's financial capacity and interests, impacting their likelihood to purchase specific packages. Since this was not prominent in EDA, it can be a step for improvement in further research.

Recommendations

Personalize Product Pitches by Customer Profiles

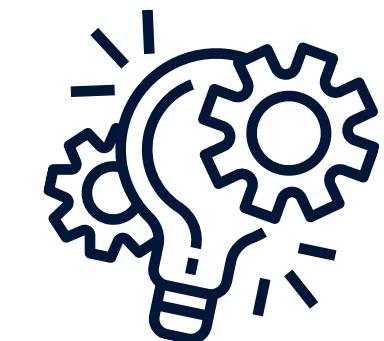
- Strategize and focus marketing specifically to customers in their 30s with a Monthly Income of around \$20,000 - \$40,000
- Pitch products according to each customer profile

Customize communication strategies for each City Tier demographic

- Prioritize mid-tier packages such as *Standard* and *Deluxe* for City Tier 3
- Focus promoting higher-end packages like *King* in City Tier 1

Increase Follow-up Efforts Strategically

- Create targeted follow-up campaigns for leads showing interest but have not yet converted
- Ensure at least 4 - 5 follow ups to increase success rates



LET'S GET IN TOUCH!



LinkedIn

www.linkedin.com/in/mkhairunnisa/



Email

kh.manda@gmail.com



GitHub Portfolio

www.github.com/mrndkhrnnsa