The background of the page is a high-angle aerial photograph of a tropical island. The island is covered in dense green forests and rocky terrain. Several small boats are scattered across the clear, turquoise-blue water. In the distance, more islands and a range of mountains are visible under a bright sky.

Introduction

Project Overview

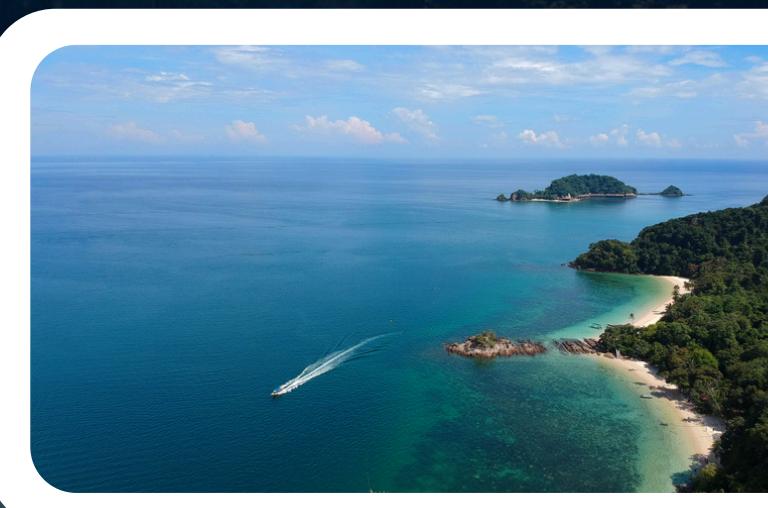
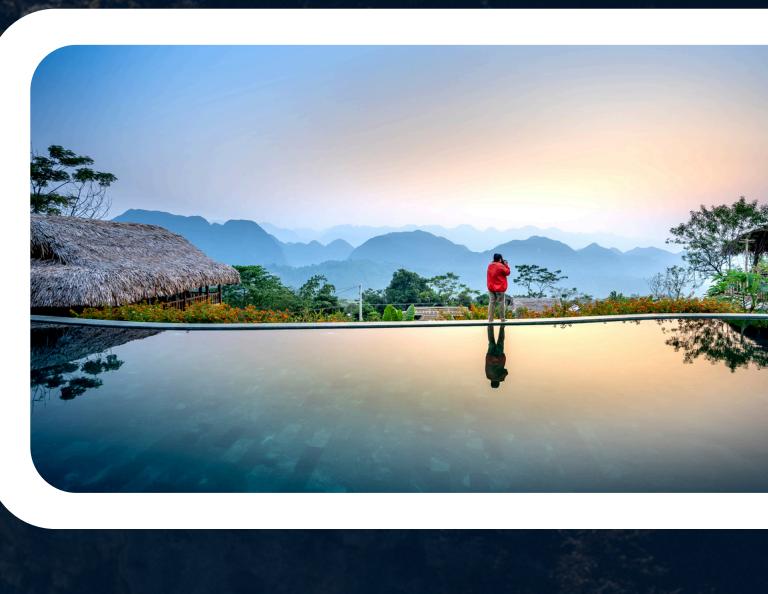
Project Details

Contact



# TRAVEL PACKAGE PREDICTION

Project Link: [here](#) ➔



# 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 and Age plays a role in differentiating customer preferences, highlighting the need to tailor marketing efforts based on demographic.
- Number of Follow-ups suggests that regular engagement with potential customers enhances the likelihood of a purchase.





# Business Understanding

Travel packages are gaining popularity due to their convenience and cost savings, but understanding what drives customers to purchase them is key for travel businesses. This project analyzes a travel package dataset to predict customer behavior, allowing agencies to personalize offerings, target marketing, and ultimately improve customer satisfaction and revenue. The insights gained will benefit both businesses and travelers, creating more tailored and value-driven experiences.



**Goal:** Determine customer profile and purchasing behaviour

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



**4888** ROWS,  
**20** COLUMNS

# Data Information

**19** FEATURES

## Customer Profile

- Customer ID
- Age
- City Tier (*1: less urban → 3: highly urbanized*)
- 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

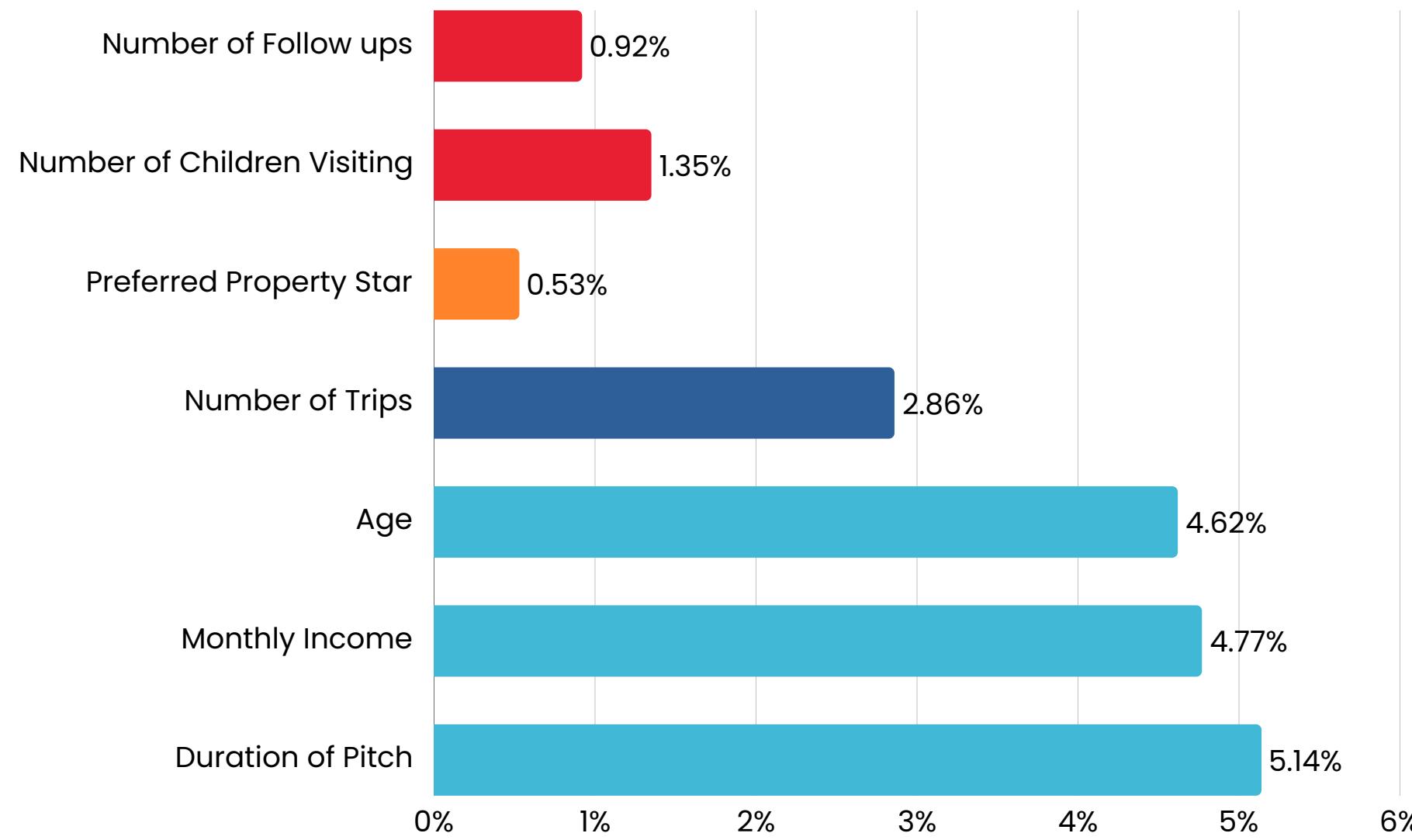
**1** TARGET

## ProdTaken

(whether the product was purchased or not)

# Data Quality Overview

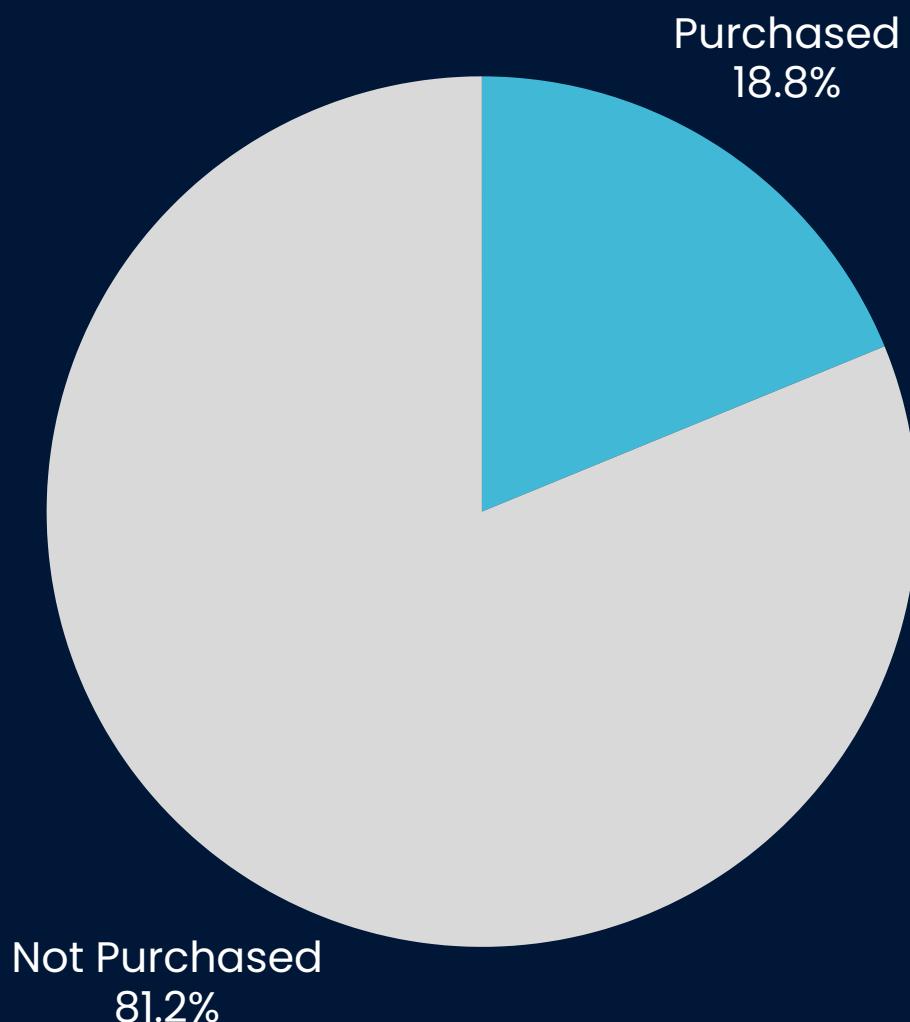
**Percentage of Missing Data**



 **No Data Duplicates**

# Exploratory Data Analysis

## Proportion of Purchased Travel Packages

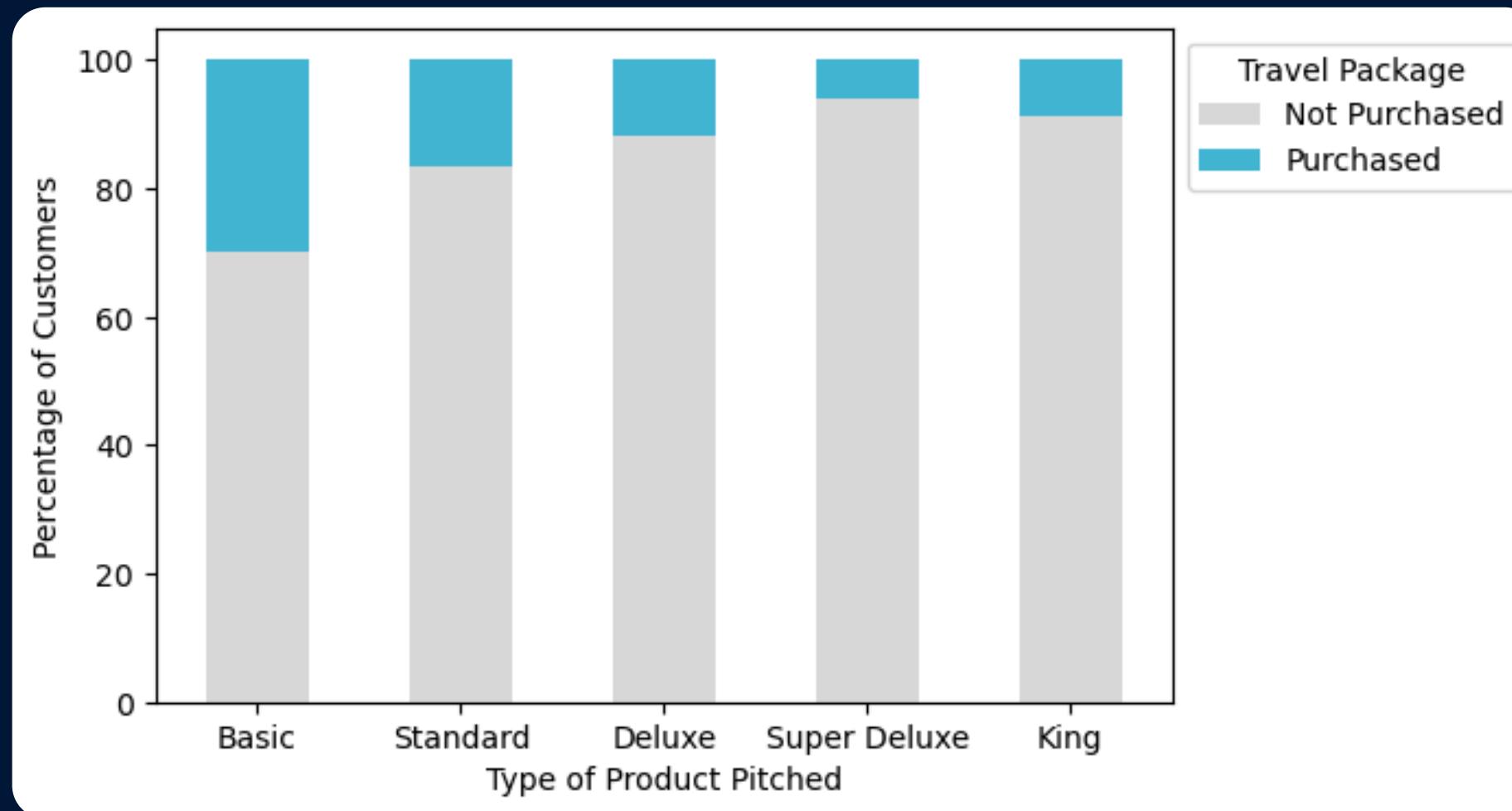


### Business Questions:

- What are the key demographic, socioeconomic, and behavioral differences between customers who purchase the travel package and those who do not?
- What are the primary drivers and barriers influencing the purchase decision regarding the travel package?

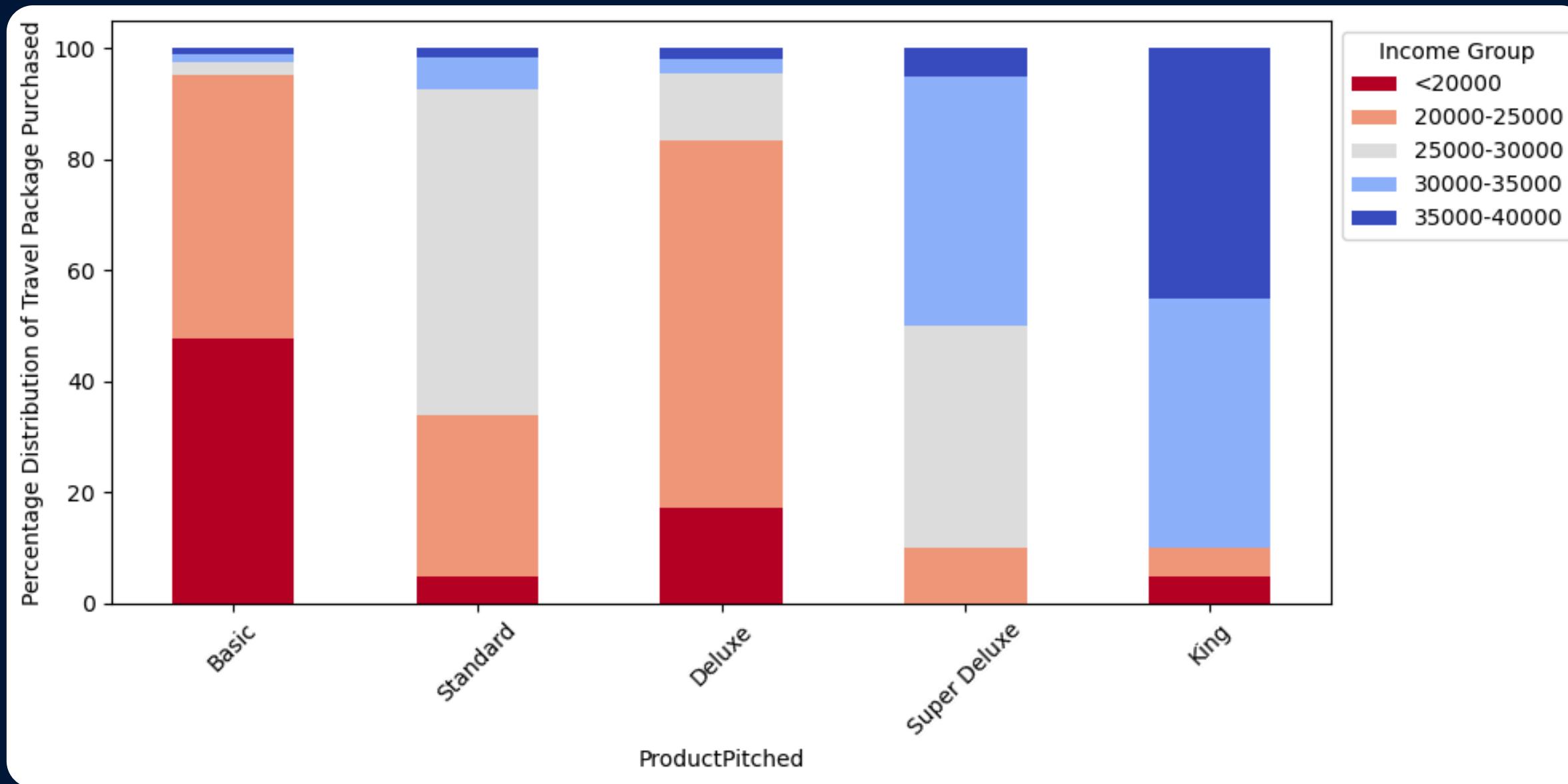
# Exploratory Data Analysis

## Percentage Distribution of Travel Package Purchased by Product Pitched



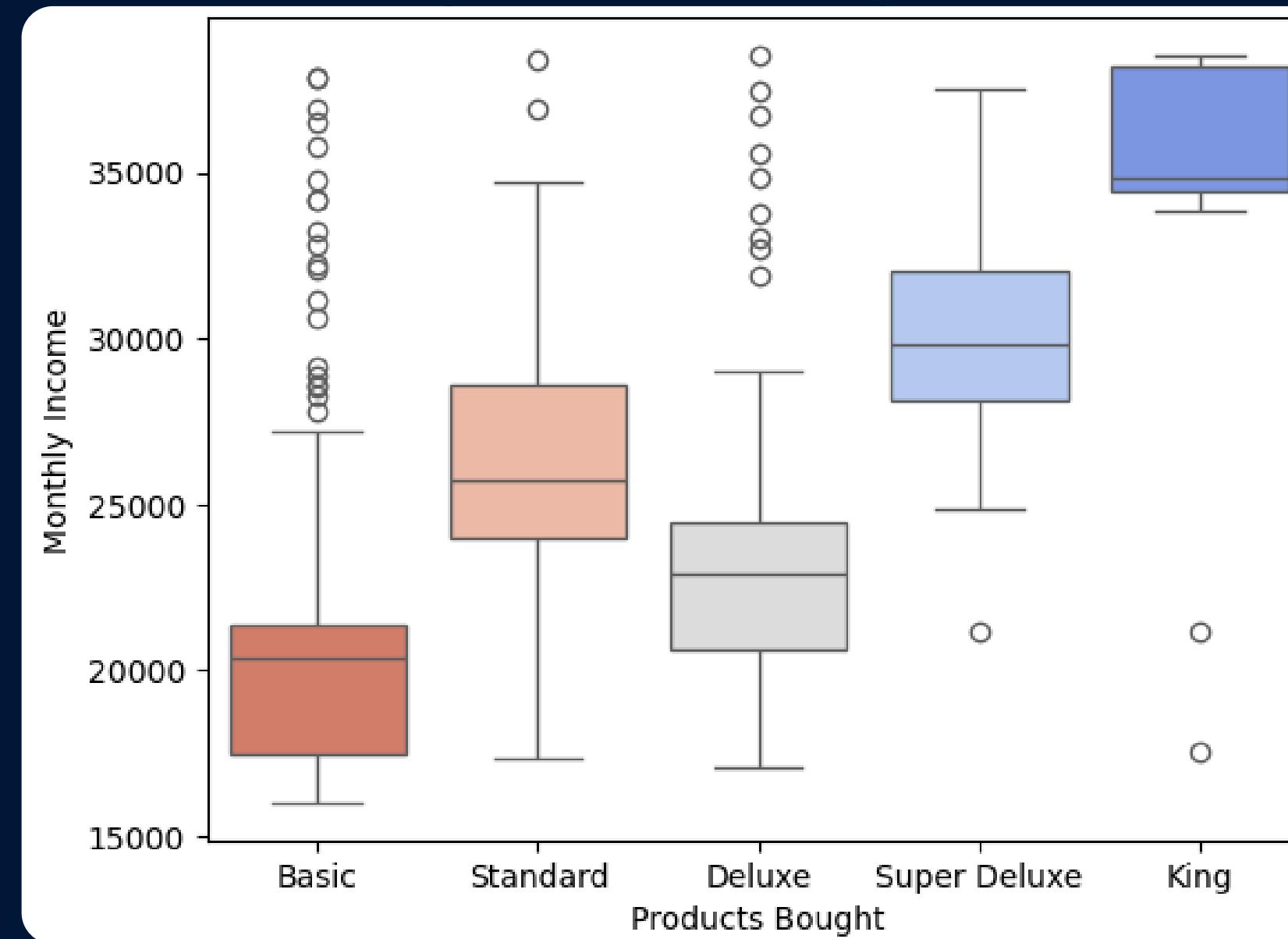
- More customers chose the Basic travel package over Super Deluxe and King options.
- The Basic package appeals to most customers, likely due to its affordability and broader suitability.

# Exploratory Data Analysis | Customer Age Insights



- Basic package buyers demonstrate strong interest in affordable travel among early-career individuals.
- Older customers prefer Premium packages indicating a trend toward luxury with age.

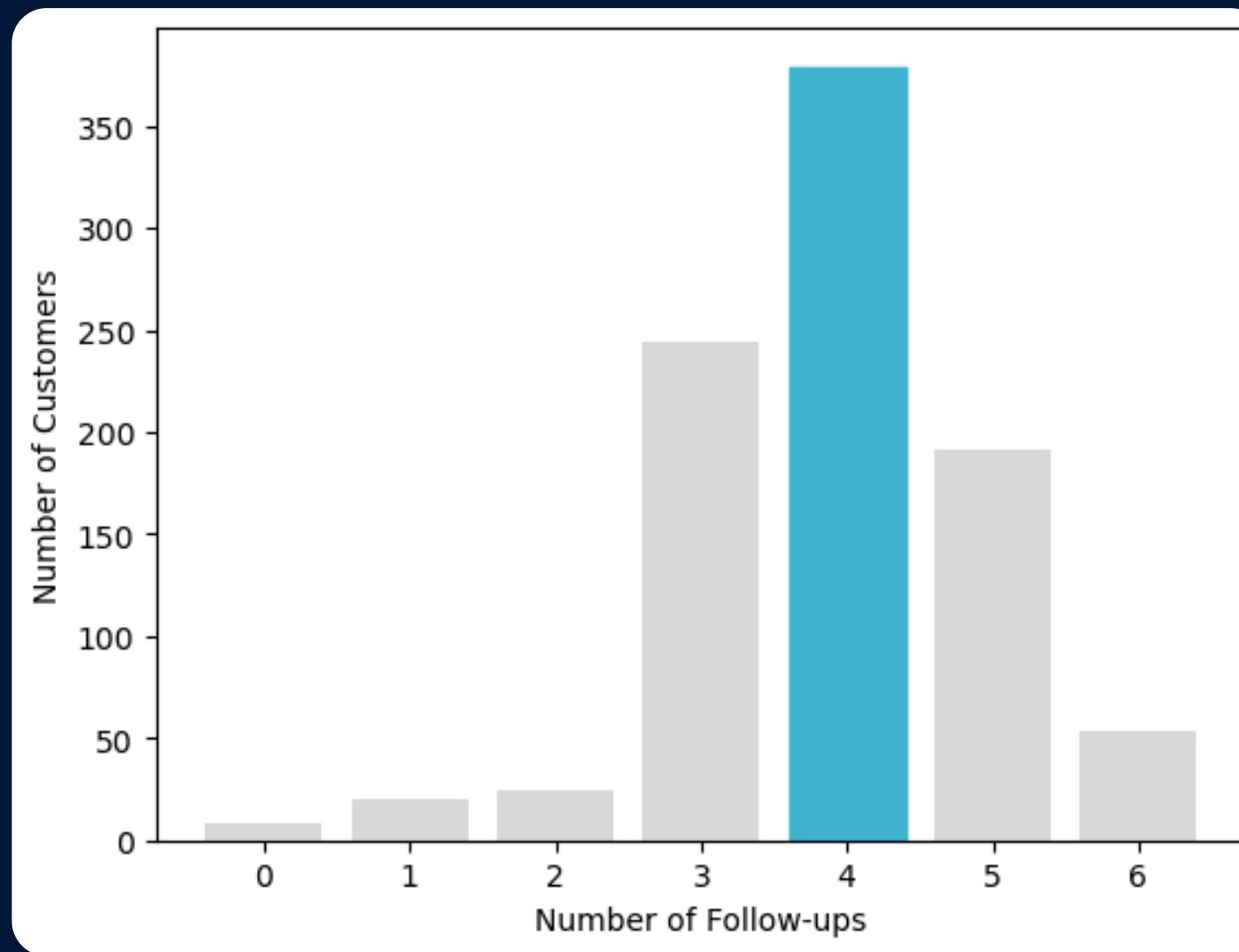
# Exploratory Data Analysis | Customer Income Insights



- Average monthly income increases with package tier, showing a strong link between income and travel package choice.
- Premium package buyers have the highest median incomes, confirming that luxury packages attract high-income segments.

# Exploratory Data Analysis

## Distribution of Number of Follow-ups by Customers who Purchased



- Frequent follow-ups increase the likelihood of purchasing a travel package.
- Consistent sales team engagement positively influences customer decisions, suggesting regular follow-ups are key in converting leads into buyers.

# Customer Profile

## Basic



<20;  
20-30



<20000;  
20000-25000



Executives



City Tier 1/2

## Standard



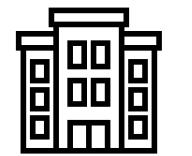
20-30;  
31-40



25000-30000



Managers



City Tier 1/2

## Deluxe



31-40



20000-25000



Senior  
Managers



City Tier 3

## Super Deluxe



31-50



30000-35000



AVPs



City Tier 3

## King



40+



30000-35000;  
35000-40000



VPs



City Tier 3

# 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	
		F1-Score	ROC-AUC
Logistic Regression	Train	39.00%	80.28%
	Test	32.98%	80.59%
Decision Tree	Train	100.00%	100.00%
	Test	74.25%	83.26%
Random Forest	Train	100.00%	100.00%
	Test	74.19%	96.27%
XGBoost	Train	99.93%	100.00%
	Test	79.19%	94.65%
LightGBM	Train	94.13%	99.85%
	Test	69.65%	94.34%

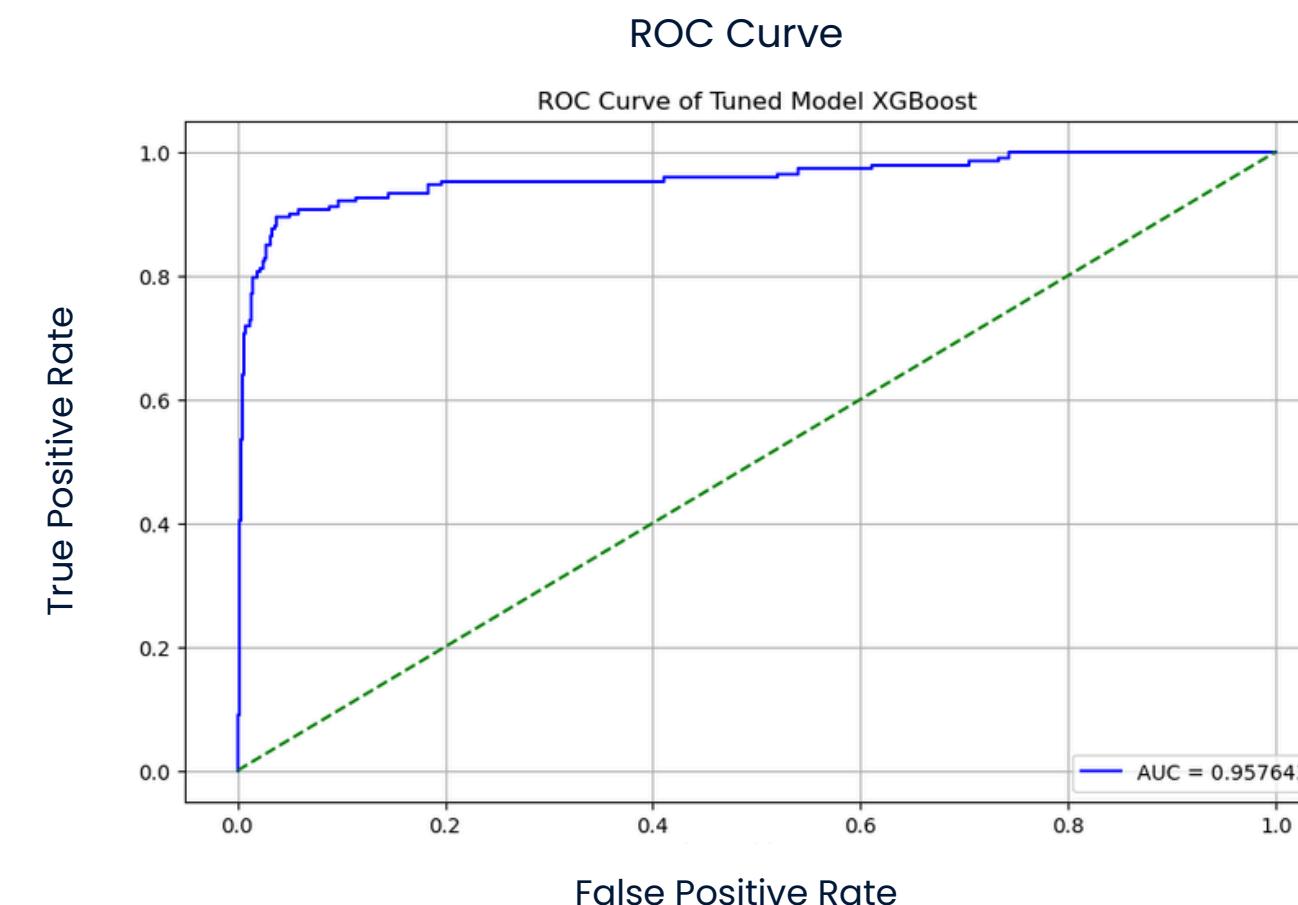
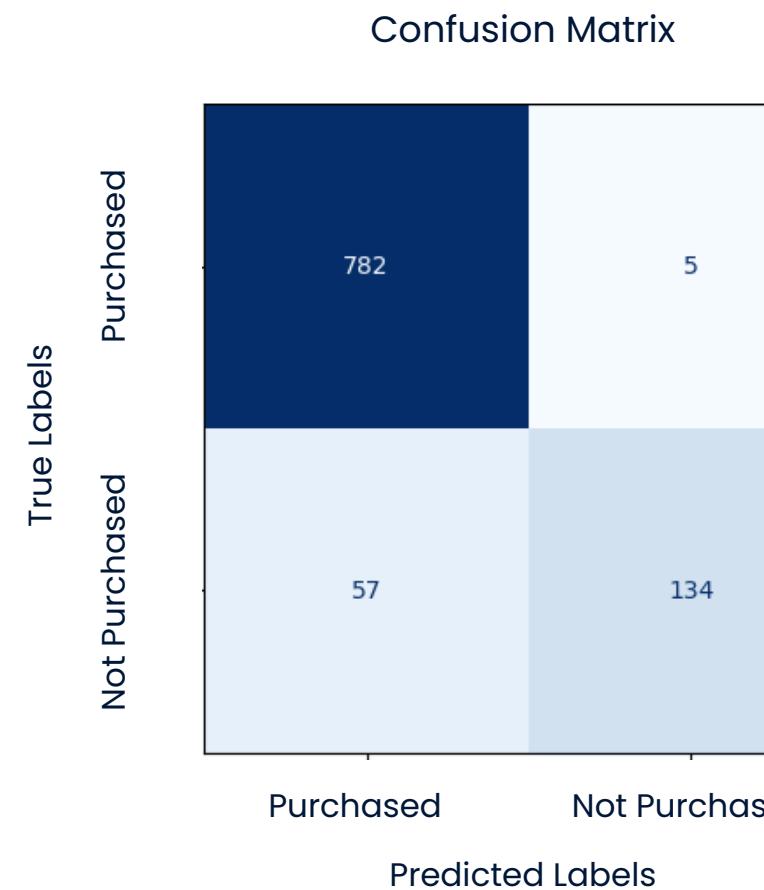
XGBoost demonstrates strong generalization with a **high F1-score** of 79.2% while maintaining a **high ROC AUC** of 94.7% on the test set, capturing complex patterns in the data without compromising its ability to perform on unseen examples.

# 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	Data	Metrics	
Base XGBoost	Train	F1-score	ROC-AUC
<b>Tuned XGBoost (without SMOTE)</b>	Train	99.93%	100%
	Test	79.19%	94.65%
<b>Tuned XGBoost with SMOTE</b>	Train	100%	100%
	Test	81.21%	95.76%

# Model Analysis

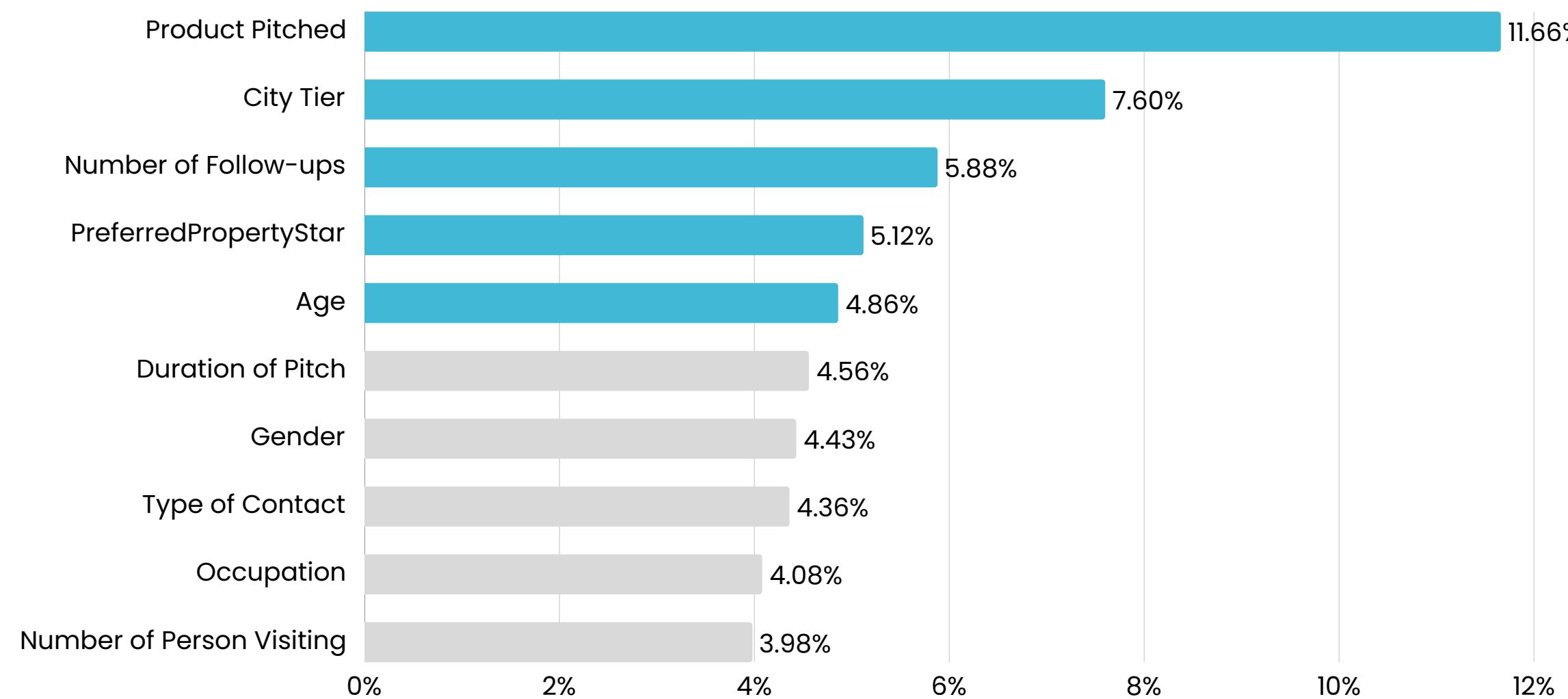


- **True Positive (TP):** 740 customers were correctly identified
- **False Positives (FP):** 60 customers were predicted to buy but did not
- **True Negatives (TN):** 137 customers were predicted to buy and actually did not
- **False Negatives (FN):** 13 customers were correctly predicted not to buy

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



Key insights based on the model include:

- **Package type** significantly influences customer purchasing decisions.
- Demographic location (**City Tier**) and **Age** affects customer preferences.
- Consistent engagement through **Number of Follow-ups** boosts purchase likelihood.

# Key Recommendations

## Lead with Best-Performing Products

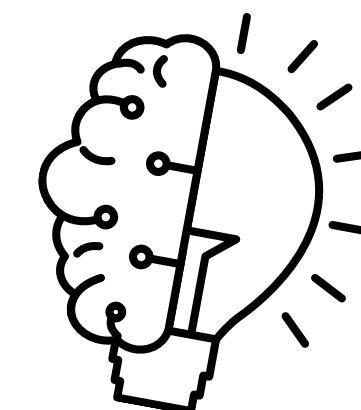
- Start pitches with the **Basic** or **Standard** package before introducing or upselling premium options.

## Target High-Potential Segments

- Focus marketing efforts on **City Tier 3** customers and age group most likely to convert (**25-40 years old**).
- Tailor messaging and offers to maintain engagement with other segments.

## Strengthen Follow-up & Personalization

- Set a minimum of **3-4 follow-ups** per lead for higher chance of conversion.
- Personalize outreach and **highlight relevant property options** based on customer preferences.



# THANK YOU!



**WatsApp**

(+62) 812 1927 5737



**LinkedIn**

[www.linkedin.com/in/mkhairunnisa/](https://www.linkedin.com/in/mkhairunnisa/)



**Email**

[kh.manda@gmail.com](mailto:kh.manda@gmail.com)



**GitHub Portfolio**

[www.github.com/mrndkhrnnsa](https://www.github.com/mrndkhrnnsa)