



DIGITAL MARKET CAMPAIGN PREDICTION PROJECT

Presented by: Sahil Anand



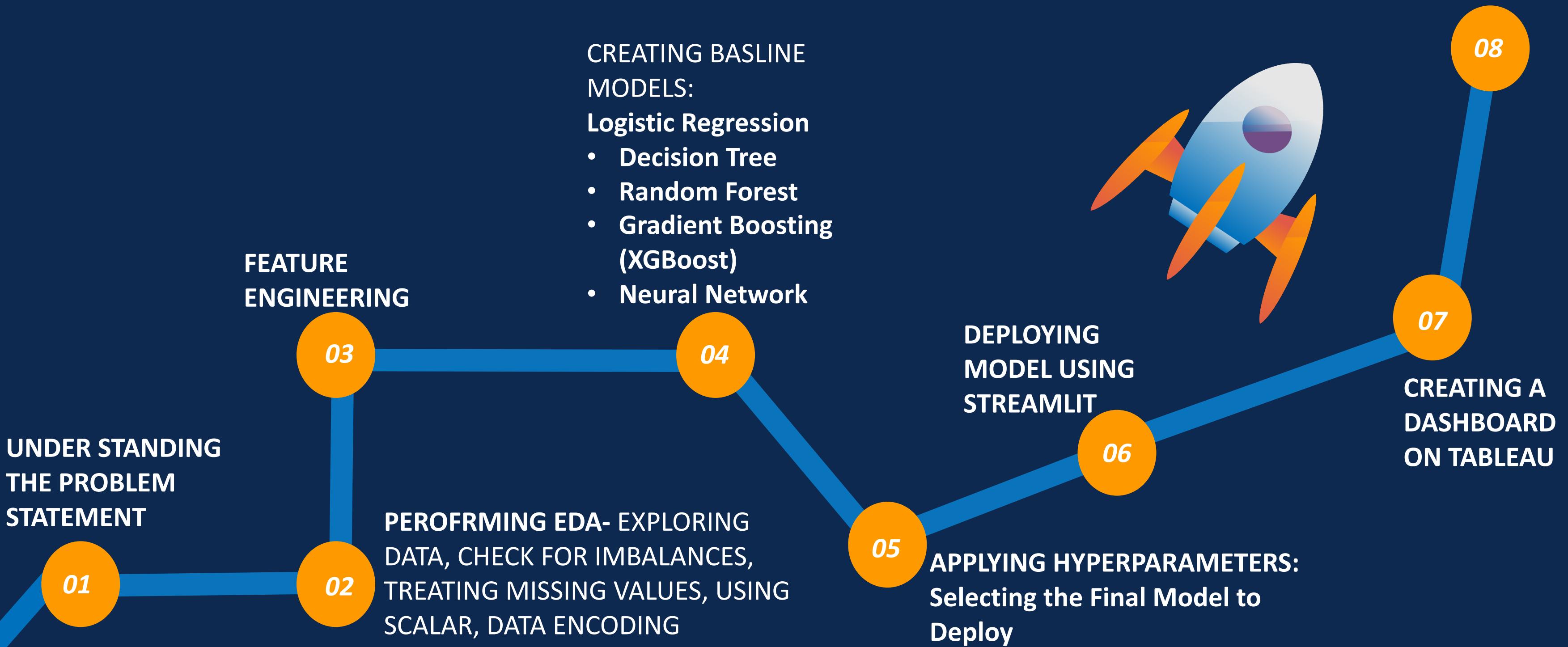


INDEX

ROADMAP	3
OBAJECTIVE	4
EXPLORATORY DATA ANALYSIS	5
FEATURE ENGINEERING	13
MODEL ELECTION	14
BEST MODEL	15
MODEL DEPLOYMENT	18
THE DASHBOARD	21
CONCLUSION	26



ROADMAP



OBJECTIVE

This project aims to enhance the effectiveness of digital marketing campaigns by accurately predicting customer conversions. By leveraging machine learning, the project seeks to identify potential converters and optimize marketing strategies. The objective is to develop a robust model that predicts customer conversions based on demographic and engagement data, enabling:

- Improve Campaign Targeting: Identify potential converters, allowing for more precise and efficient marketing efforts.
- Increase Conversion Rates: Enhance the effectiveness of campaigns by focusing on the most promising leads.
- Maximize Return on Advertising Spend (ROAS): Optimize resource allocation to reduce wasted ad spend and improve overall campaign performance.

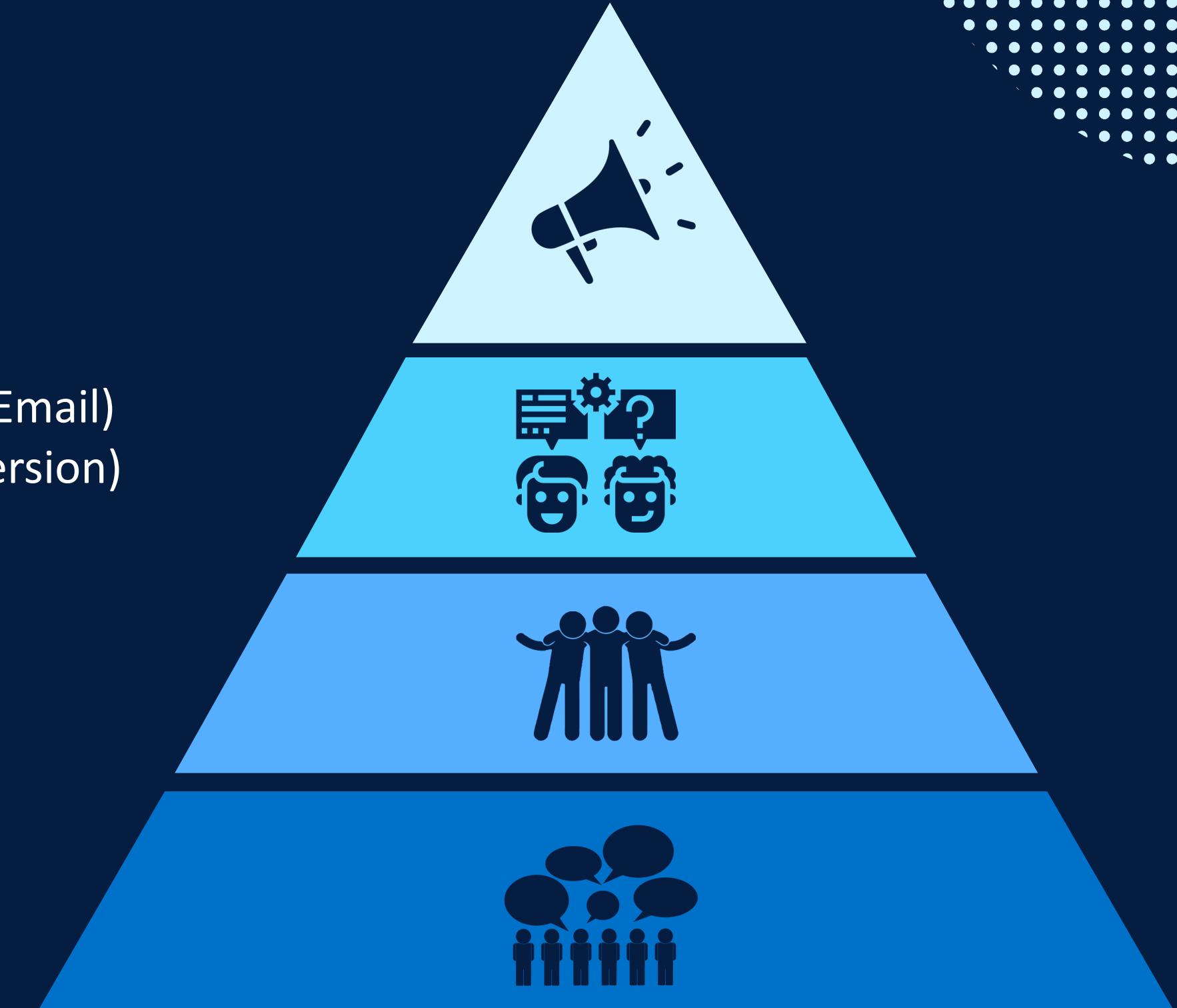
DATASET OVERVIEW

The dataset used for this project includes a comprehensive set of features that capture both demographic and engagement metrics of customers. Key features include:

01 Demographics: Age, Gender, Income

02 Engagement Metrics:

- ❖ Campaign Channel (e.g., Social Media, Email)
- ❖ Campaign Type (e.g., Awareness, Conversion)
- ❖ Ad Spend
- ❖ Click-Through Rate (CTR)
- ❖ Conversion Rate
- ❖ Website Visits
- ❖ Pages Per Visit
- ❖ Time on Site
- ❖ Social Shares
- ❖ Email Opens and Clicks





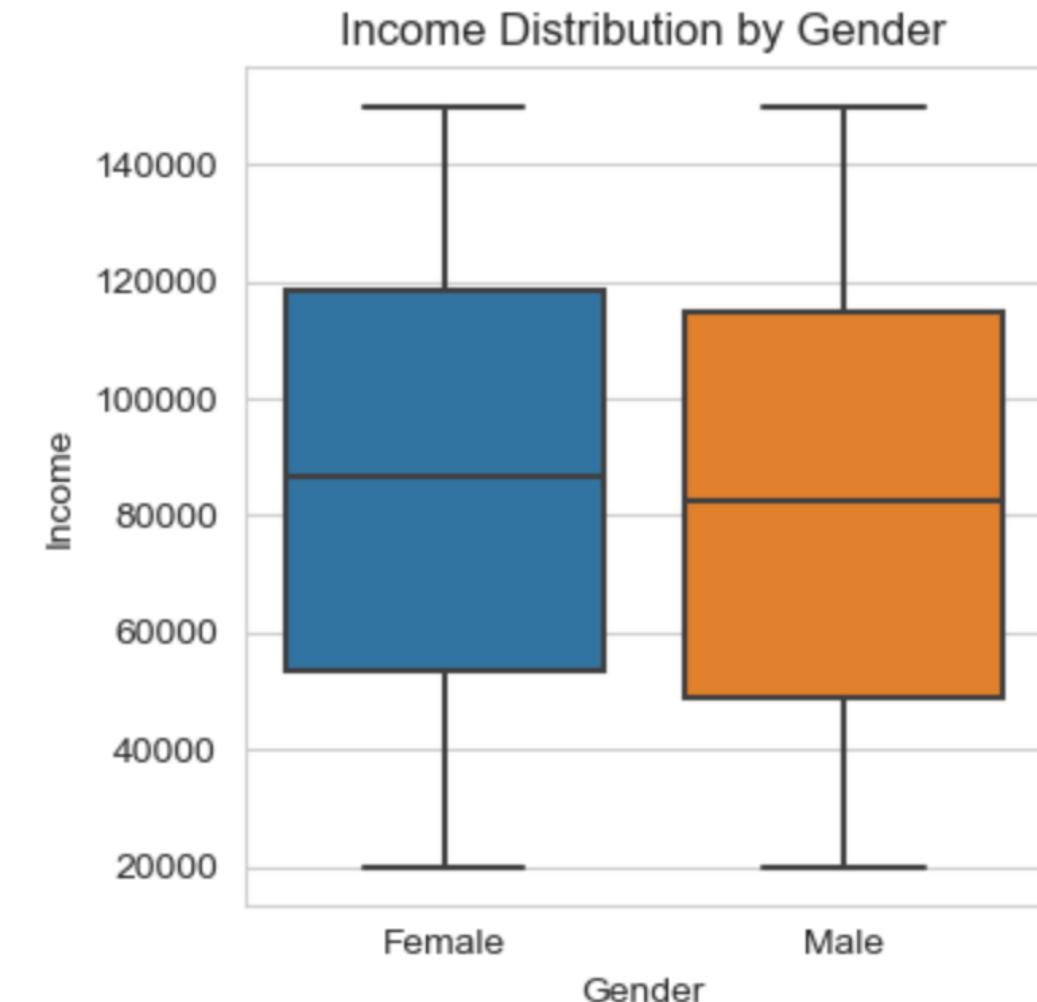
EXPLORATORY DATA ANALYSIS (EDA) INSIGHTS



Income Distribution by Gender

It presents the income distribution for males and females:

- Females: The median income is slightly higher than that of males. The interquartile range (IQR) is also larger, **indicating more variability in female incomes**. The whiskers suggest that the income range is broader for females.
- Males: The median income is lower compared to females, with a smaller IQR, **indicating less variability**. The whiskers are shorter, suggesting a narrower income range.



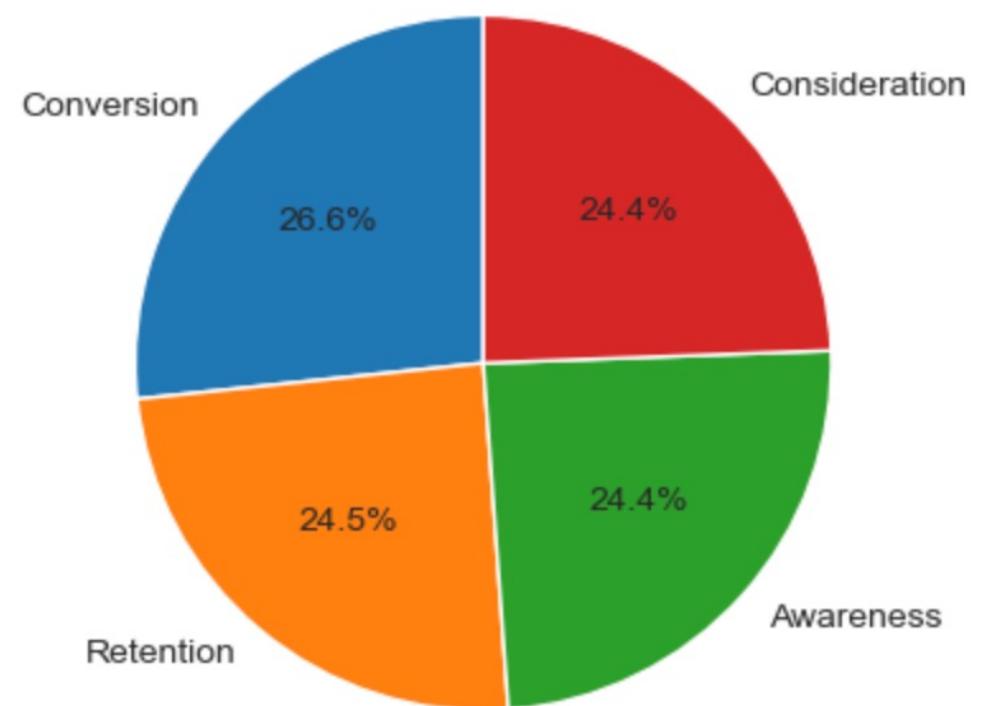
Conversion Rate by Campaign Type

It shows the distribution of conversion rates across four types of marketing campaigns:

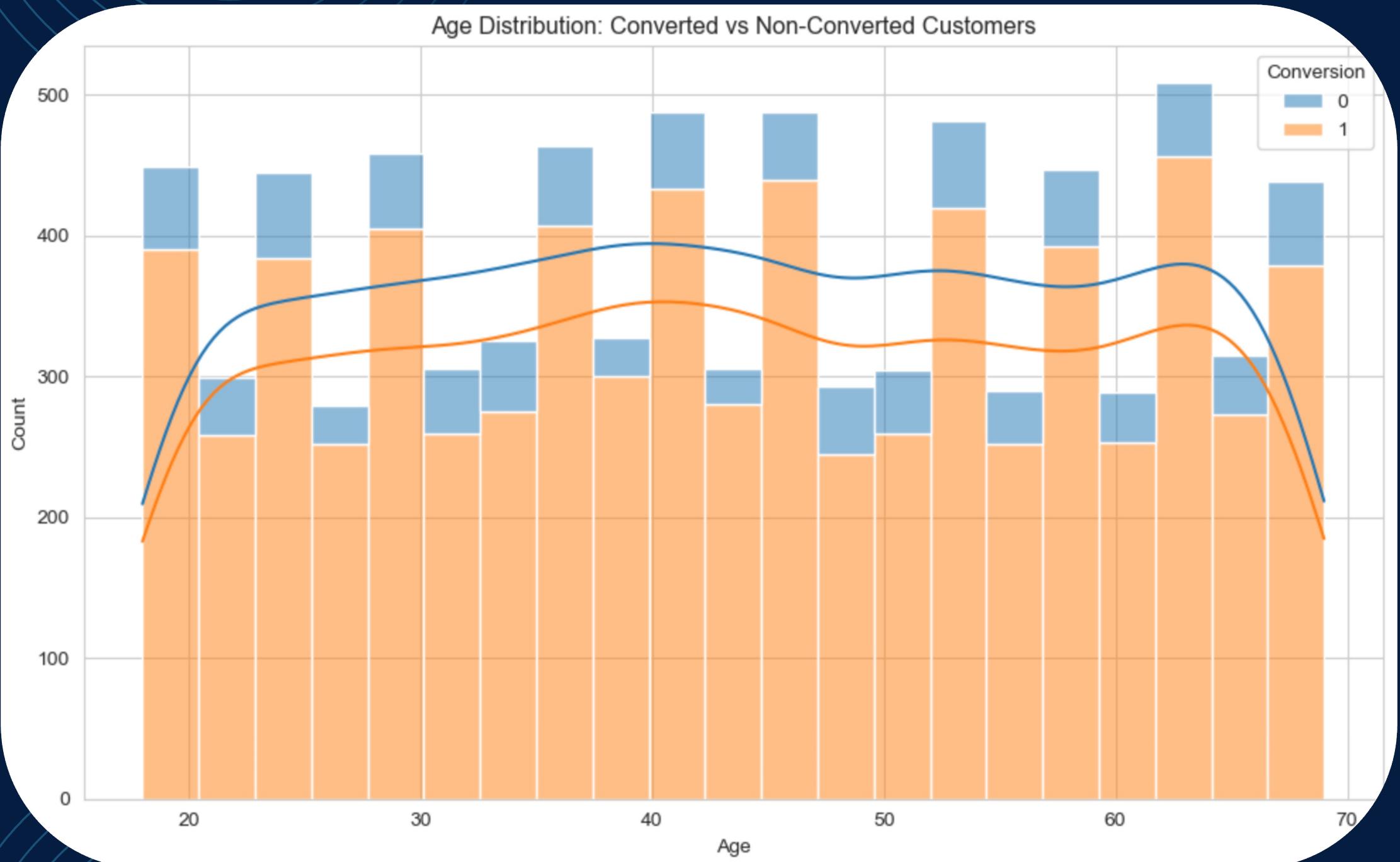
- Conversion: 26.6%
- Consideration: 24.4%
- Awareness: 24.4%
- Retention: 24.5%

This chart indicates that **the "Conversion" campaign type has the highest conversion rate**, slightly more than the other three types, which are nearly equal.

Conversion Rate by Campaign Type



Stacked Bar Chart: AGE DISTRIBUTION

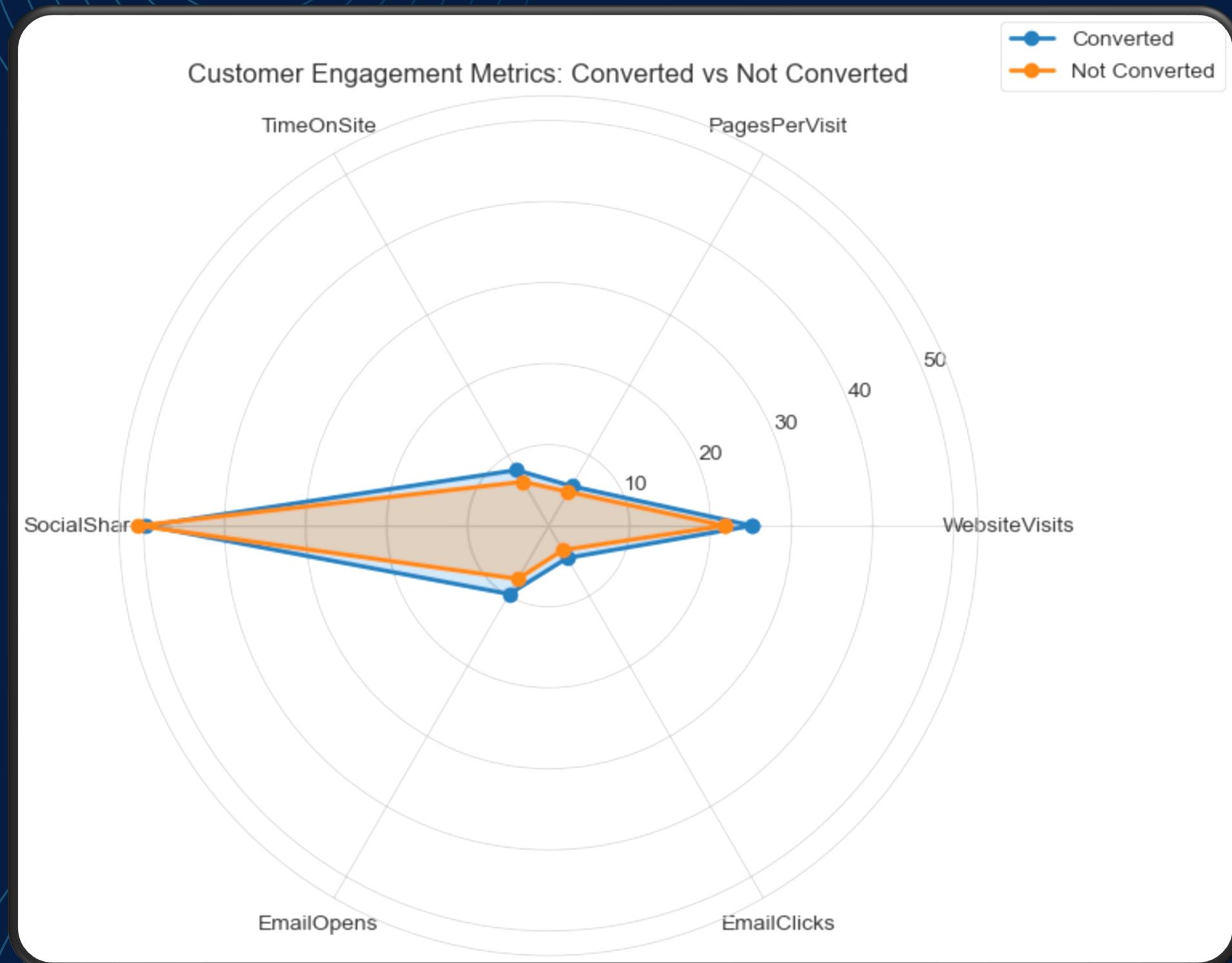


This stacked bar chart illustrates the age distribution of converted (orange) and non-converted (blue) customers.

Each bar represents a different age group, with the height indicating the count of customers. The chart shows:

- Converted Customers: The orange sections indicate the number of customers who converted.
- Non-Converted Customers: The blue sections represent those who did not convert.
- Trends: The lines show trends for both converted and non-converted customers across age groups, with a noticeable peak around ages 30-50.

Radar Chart: Customer Engagement Metrics: Converted vs Not Converted



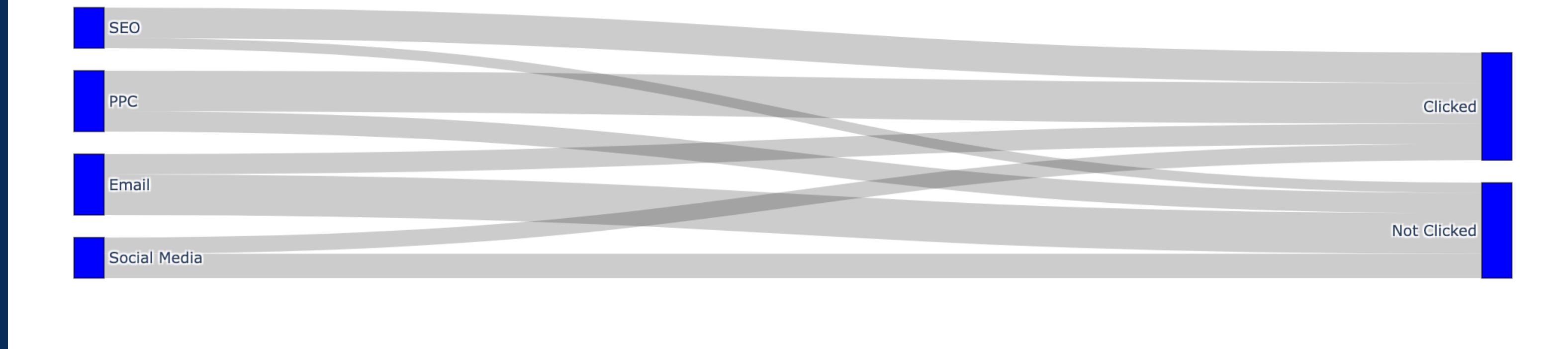
Insights

- Time On Site: The converted group spends slightly more time on the site compared to the not converted group.
- Pages Per Visit: Converted users tend to view more pages per visit.
- Website Visits: Both groups have similar website visit numbers, with a slight edge for the converted group.
- Email Opens and Email Clicks: Converted users show higher engagement in both opening and clicking emails.
- Social Share: The not converted group has a marginally higher social sharing activity.

Overall Observations

- **Engagement:** Converted users generally exhibit higher engagement across most metrics, particularly in email interactions and site activity.
- **Potential Focus Areas:** Enhancing email marketing strategies and encouraging more page interactions could potentially increase conversions.

Simplified Flow: Customer flow through Marketing Channels



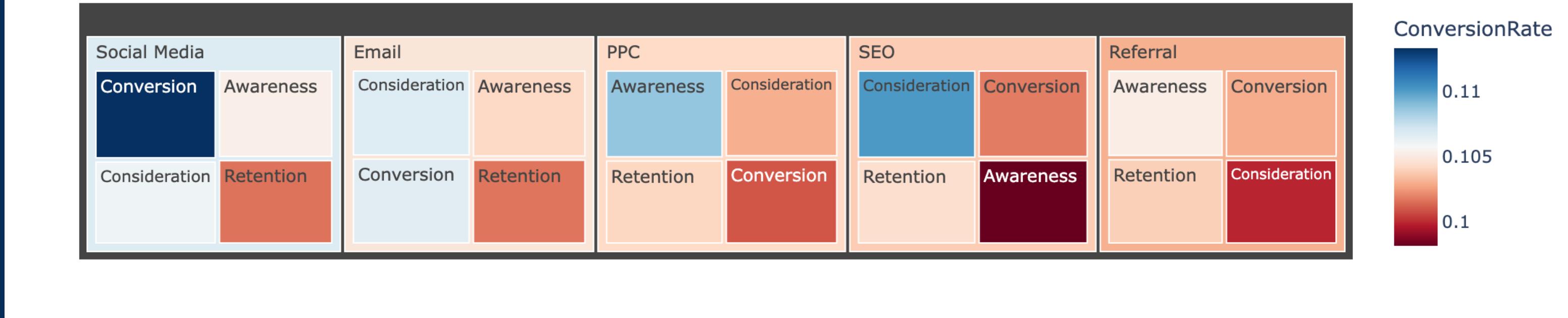
Insights from the Sankey Diagram

- Flow of Customers: A significant proportion of users from SEO end up clicking through.
- PPC (Pay Per Click): Shows a balanced distribution between users who click and those who do not.
- Email: A higher number of users from this channel tend to click through, indicating effective email campaigns.
- Social Media: A **larger proportion of users do not click through** compared to other channels.

Overall Observations

- Effectiveness: **Email and SEO channels appear to be more effective in driving clicks** compared to PPC and Social Media.
- Potential Improvement: Strategies to increase engagement and click-through rates in PPC and Social Media could be beneficial.
- User Behaviour: Understanding the reasons behind the higher click-through rates in Email and SEO could provide insights to improve other channels.

Campaign Performance Treemap



Overall structure

- The tree map visualizes marketing performance across five channels: Social Media, Email, PPC, SEO, and Referral.
- Each channel is divided into stages of the marketing funnel: Awareness, Consideration, Conversion, and Retention.

Color Coding

- Colors range from blue (high conversion rates) to red (low conversion rates).
- Conversion rates are depicted on a scale from 0.1 to 0.11.

Strategic Insights

- Focus on enhancing Retention in Social Media.
- Leverage strong Conversion performance in PPC and SEO for further optimization.
- Consider strategies to improve Awareness in Referral channels.

INSIGHTS FROM THE CORRELATION HEATMAP

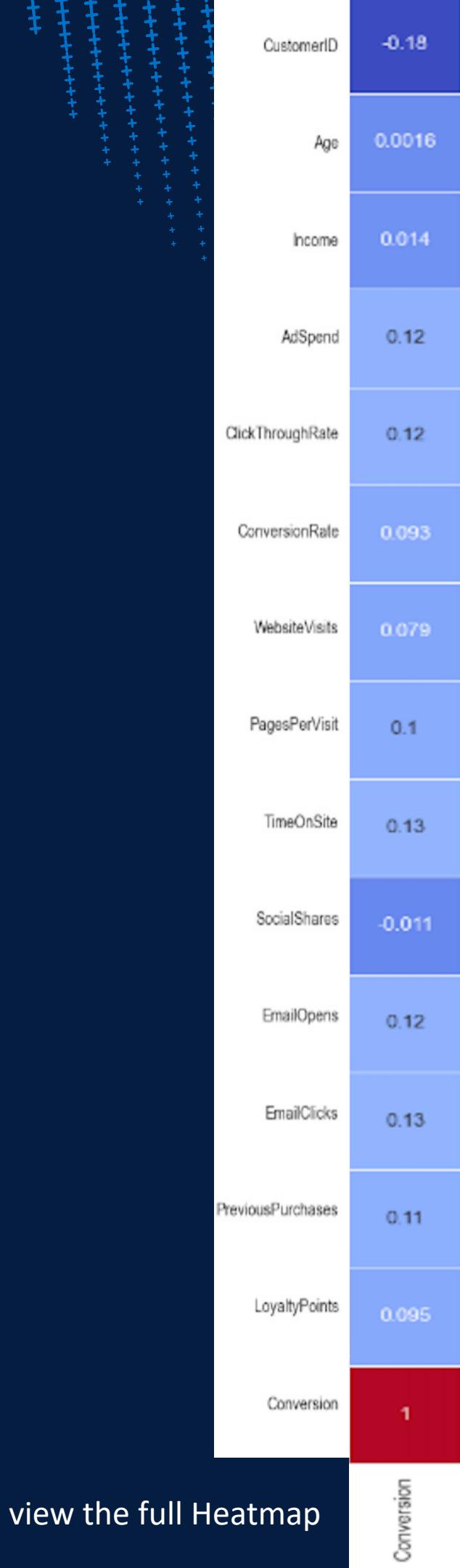
The correlation heatmap provides insights into the relationships between various campaign variables. Here are the key observations:

Conversion

- Click Through Rate: Shows a moderate positive correlation with conversion, suggesting that higher click-through rates are associated with more conversions.
- Pages Per Visit and Time On Site: Both have moderate positive correlations with conversion, indicating that more page views and longer site visits are linked to higher conversion rates.
- Ad Spend: Also positively correlated, suggesting that increased spending can lead to more conversions.

Other Variables

- Email Opens and Email Clicks: Show positive correlations with conversion, highlighting the importance of effective email campaigns.
- Social Shares: Has a slight positive correlation, indicating some impact on conversion.



*Click [here](#) to view the full Heatmap

FEATURE ENGINEERING

Techniques Used

Encoding Categorical Variables:

- One-hot encoding was applied to the 'Gender' and 'Campaign Channel' columns.
- Label encoding was used for the 'Campaign Type' column.

Feature Scaling:

- Standard Scaler was employed to normalize numerical features, ensuring all features are on a similar scale.

Feature Creation:

Several new features were engineered from the existing data:

- Click Through Rate: Calculated as a ratio of clicks to impressions.
- Conversion Rate: Computed as the ratio of conversions to clicks.
- Engagement Score: A composite score based on Web site Visits, Pages Per Visit, and Time On Site.
- Customer Life time Value: Estimated based on Previous Purchases and Loyalty Points.
- Campaign Efficiency: Calculated as Conversion divided by Ad Spend.



MODEL SELECTION

ALGORITHMS EVALUATED

- 1) **Logistic Regression**
- 2) **Decision Tree**
- 3) **Random Forest**
- 4) **Gradient Boosting
(XGBoost)**
- 5) **Neural Network**



MODEL SELECTION PROCESS

The model selection process involved the following steps:

Data Pre processing:

- Handling missing values
- Encoding categorical variables
- Feature scaling

Train-Test Split:

- Data was split into 80% training and 20% testing sets.
- Cross-Validation: 5-fold cross-validation was used to assess model performance.



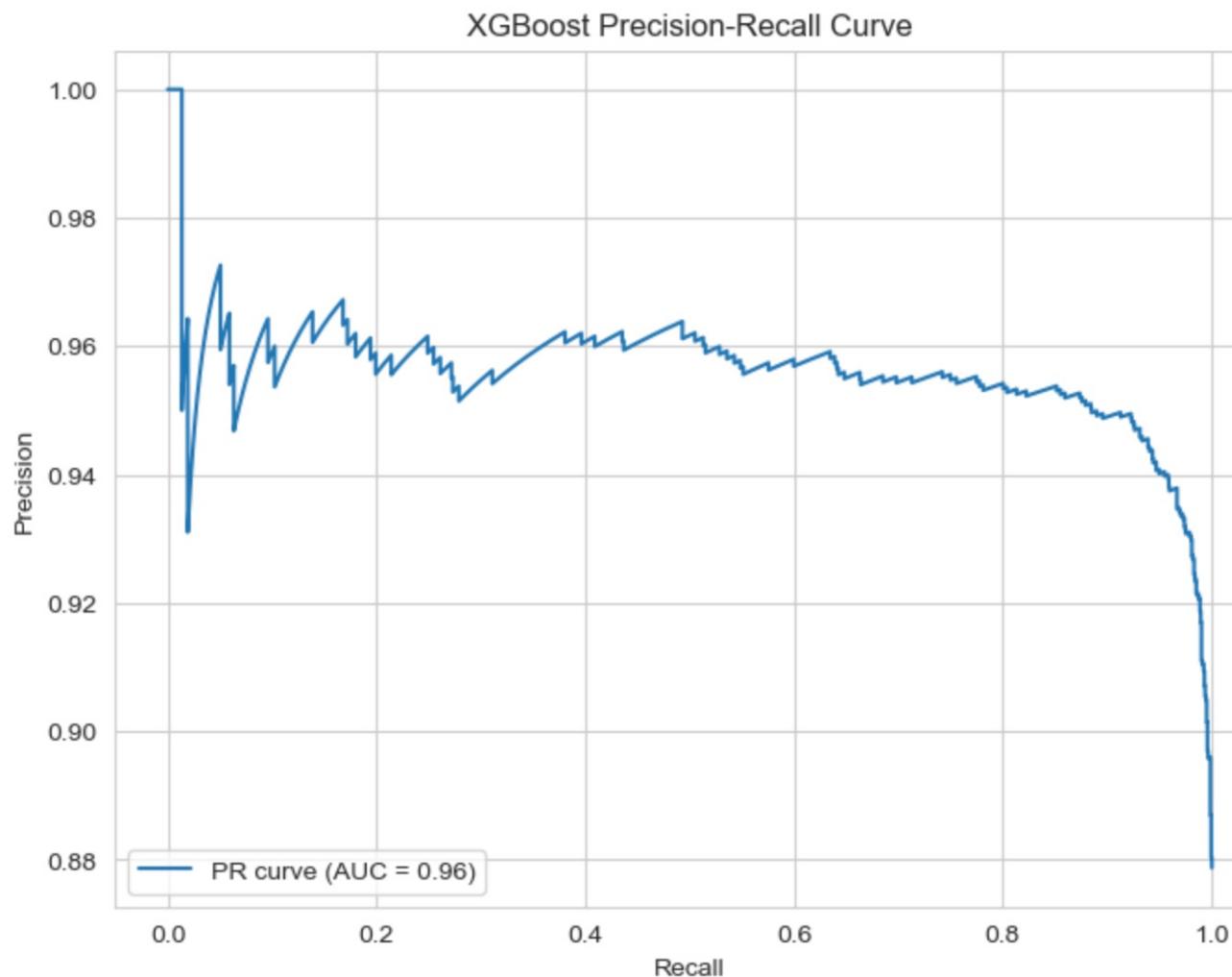
Performance Metrics:

- Models were evaluated based on accuracy, precision, recall, F1-score, and AUC-ROC.
- Hyperparameter Tuning: Grid search with cross-validation was performed for each model.

BEST MODEL: XGBOOST

The XGBoost model emerged as the best performer, both with and without hyperparameter tuning. It provided superior predictive performance compared to other models.

Before tuning the model:



	Precision	Recall	F1 Score	support
0	0.77	0.45	0.57	194
1	0.93	0.98	0.95	1406
Accuracy			0.92	1600
Macro	0.85	0.72	0.76	1600
Avg				
Weighted Avg	0.91	0.92	0.91	1600

Confusion Matrix:

88 - True Negatives	106 - False Negatives:
27 - False Positives	1379 - True Positives

Areas for Improvement:

Enhance recall for class 0 to reduce false negatives, possibly by adjusting the decision threshold or using techniques like cost-sensitive learning.

Precision-Recall Curve

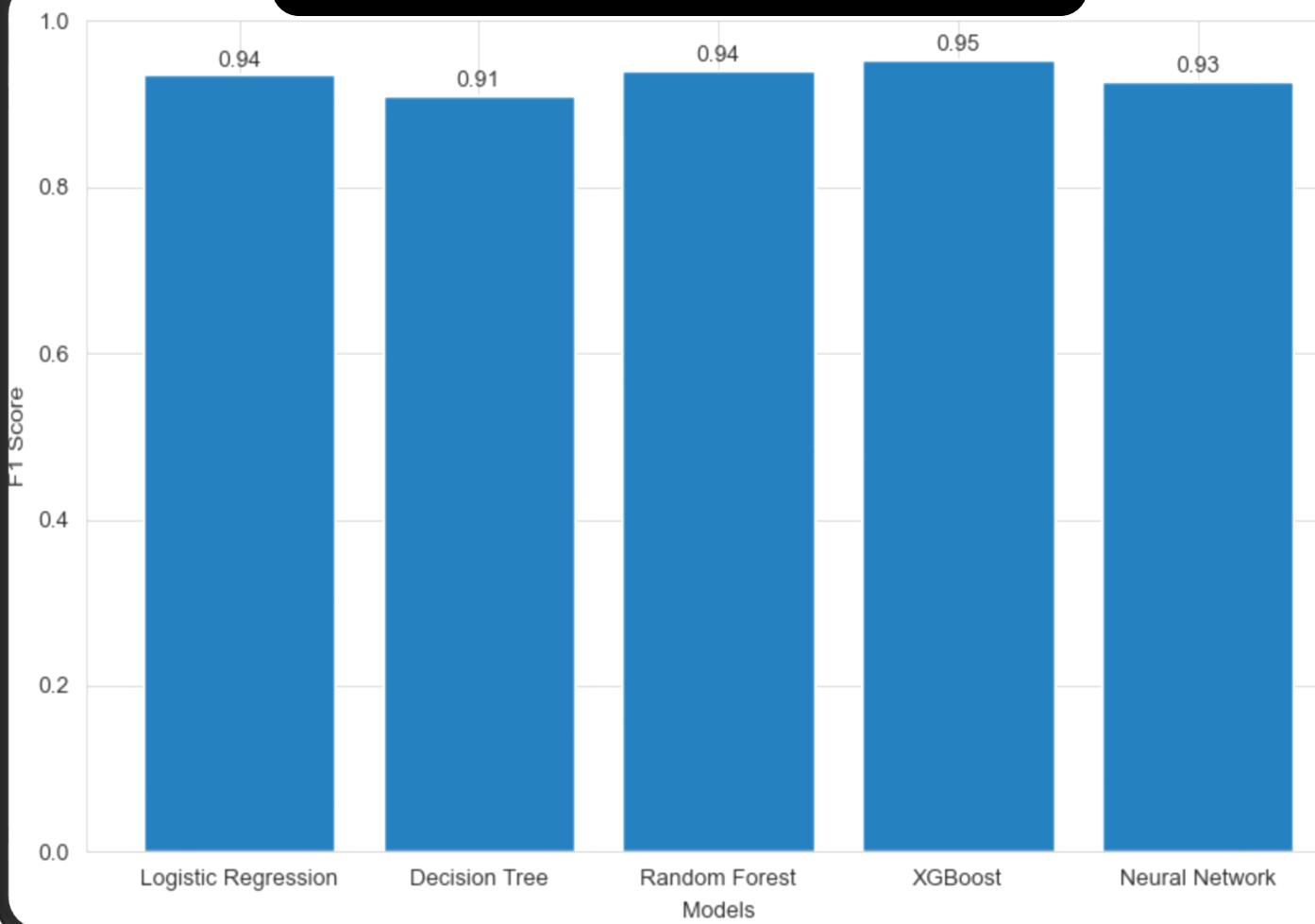
- AUC (Area Under Curve): 0.96 Indicates excellent model performance with high precision across various recall levels.
- The curve shows the model maintains high precision even at high recall values, suggesting robustness.

Key Insights

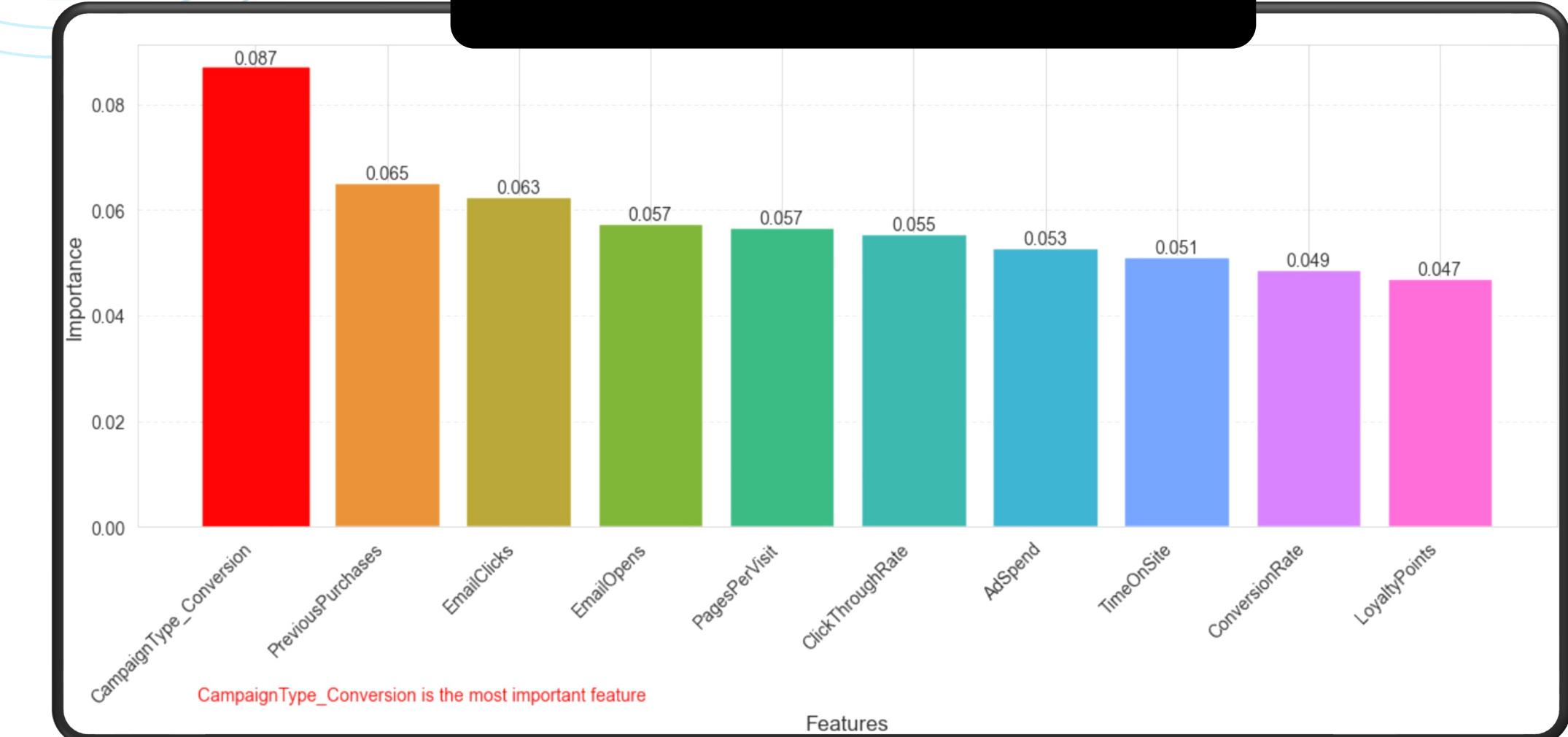
- Strengths:
 - The model performs exceptionally well in predicting the positive class (class 1), which is often the focus in many applications.
 - High AUC-PR value indicates reliable predictions.

BEST MODEL: XGBOOST Comparison_Best_Features

F1 SCORE COMPARISON



TOP 10 FEATURE IMPORTANCE FOR XGBOOST



A visual comparison of F1 scores across different models can be presented in a bar chart, highlighting the superior performance of XGBoost.

XGBoost allows for easy extraction of feature importance, which provides insights into the most influential features in predicting conversions.

BEST MODEL: XGBOOST with Hyperparameters

Best parameters for XGBoost:

subsample	1.0
n_estimators	400
max_depth	3
learning_rate	0.05
gamma	0.4
colsample_bytree	0.6

*These parameters optimize model performance by controlling complexity and learning rate, enhancing accuracy and generalization.



Best F1 score for XGBoost: 0.9600033957834031

Accuracy: 0.93

NEXT IS:

MODEL DEPLOYMENT

Platform: Streamlit

The Digital Marketing Campaign Conversion Predictor is deployed on Streamlit, providing an interactive web application for real-time predictions. Users can input customer information through a user-friendly form, and the application displays model accuracy and prediction results.

Digital Marketing Campaign Conversion Predictor

Model Accuracy: 0.93

Enter Customer Information

Age	Website Visits
29	- + 100 - +
Gender	Pages Per Visit
Female	2.00 - +
Annual Income	Time on Site
49997	- + 5.00 - +
Campaign Channel	Social Shares
Email	13 - +
Campaign Type	Email Opens
Awareness	50 - +
Ad Spend	Email Clicks
999.98	- + 17 - +
Click Through Rate	Previous Purchases
0.14	- + 1 - +
Conversion Rate	Loyalty Points
0.10	- + 104 - +

Predict

Prediction:

The customer is likely to convert!

Technical Implementations :



Data Processing:

- Caching for Efficient Data Loading: Utilizes `@st.cache_data` to speed up data loading and prevent reloading on each interaction.

Pre-processing Steps:

- Feature encoding and scaling are performed to prepare the data.
- Categorical variables like gender and campaign channels are converted to numerical values using mapping and one-hot encoding.
- Unnecessary columns are dropped to streamline the dataset.

Model Training:

- XGBoost Classifier: The model is trained using an XGBoost Classifier.
- RandomizedSearchCV: Hyperparameter tuning is done with RandomizedSearchCV to optimize model performance.
- StandardScaler: Applied for feature normalization, ensuring that all features contribute equally to the model.

Key Features:

- Interactive Web Application: Allows users to interact with the model in real-time.
- User Input Form: Collects customer information such as age, gender, income, and campaign details.
- Model Accuracy Display: Shows the model's accuracy, which is 0.93.
- Prediction Results: Provides immediate feedback on the likelihood of customer conversion.

Technical Implementations :



Prediction Pipeline:

- Real-Time Prediction: Based on user input, predictions are made instantly.
- Error Handling and Logging: Robust error handling and logging are implemented to ensure smooth operation and easy debugging.

Code Implementation:

- The code is structured to load and pre-process data, train the model, and make predictions. Key components include:
- Data Loading: Handled by `load_data()`, which reads the dataset and logs success or failure.
- Data Preprocessing: `preprocess_data()` prepares the data by encoding categorical variables and splitting features and targets.
- Model Training: `train_model()` trains the XGBoost model with hyperparameter tuning and returns the trained model, scaler, and accuracy.
- Prediction: `predict()` scales input data and predicts conversion likelihood.

Streamlit Application

- The main Streamlit application (`main()`) sets up the user interface, handles data input, and displays predictions. It includes:
 - Input Form: Collects customer data through interactive widgets.
 - Prediction Button: Triggers the prediction process.
 - Result Display: Shows whether the customer is likely to convert based on the model's prediction.

THE DASHBOARD

PRESENTING YOU:

Click [here](#) to Access this Dashboard

DIGITAL MARKETING DASHBOARD

Total Ad Spent
40,007,559

Total Website Visits
1,98,013

Avg Time on Site
7.728

Avg Click Rate
0.1548

Avg Conversion Rate
0.1044

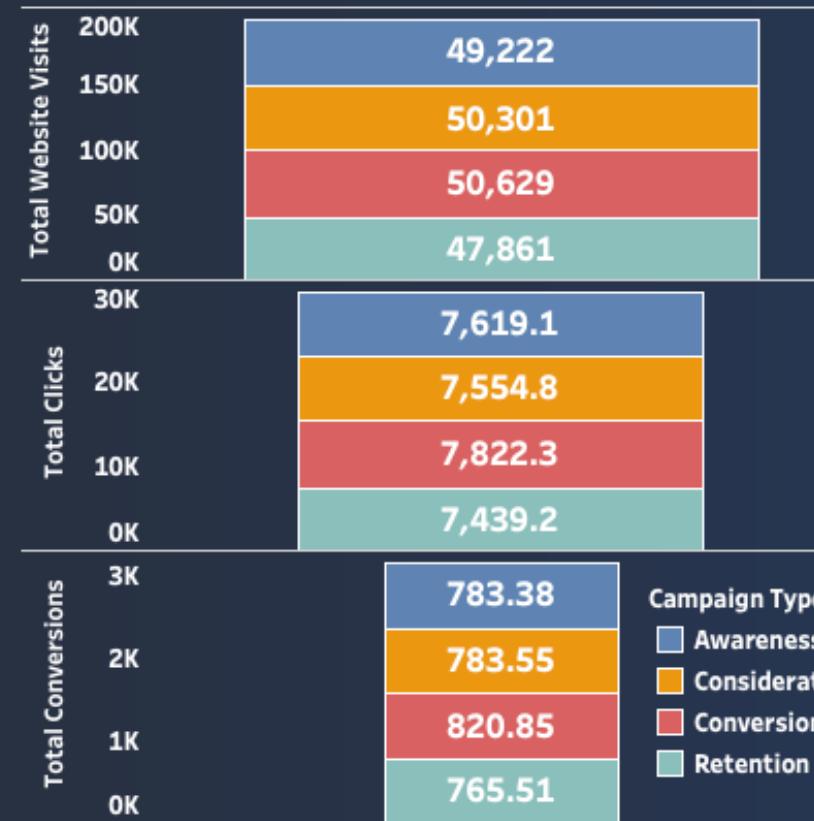
Customer Segmentation

Gender	Avg. Ad Spend	Avg. Age	Avg. Income
Female	5,005	44	85,923
Male	4,995	43	82,738

Averages

Avg. Time On Site	7.73
Avg. Website Visits	24.75
Avg. Social Shares	49.80
Avg. Pages Per Visit	5.55

Funel- Analysis



Gender: (All) Campaign Channel: (All) Campaign Type: (All)



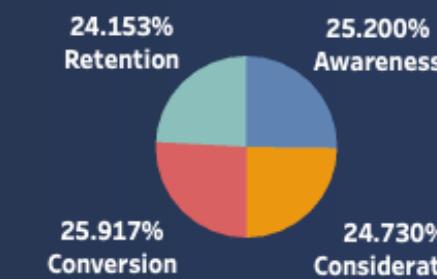
Customer Loyalty Heatmap



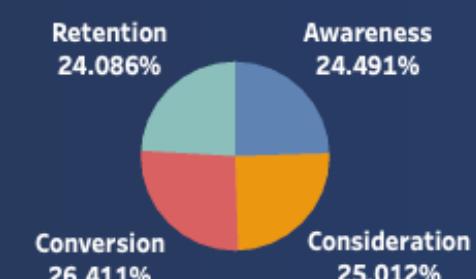
Total- Campaign Type



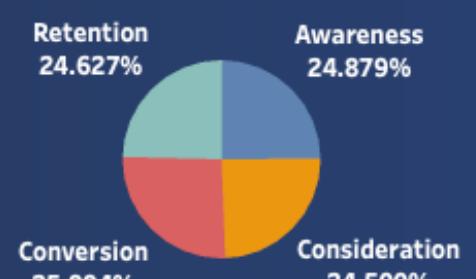
Social Engagement



Email Clicks



Email Opens





The digital marketing dashboard provides an overview of key metrics:

- Total Ad Spend: 40,007,559
- Total Website Visits: 198,013
- Average Time on Site: 7.728 minutes
- Average Click Rate: 0.1548
- Average Conversion Rate: 0.1044

Customer Segmentation: Insights from customer segmentation include:

Gender:

- Female: Average ad spend is 5,005, average age is 44, and average income is 85,923.
- Male: Average ad spend is 4,995, average age is 43, and average income is 82,738.

Customer Loyalty Heatmap:

- Shows varying loyalty levels across income and age bins, with higher loyalty in certain income brackets like 100K and age groups like 40.

Campaign Performance: Performance analysis of different campaign types and channels:

- Campaign Types:
- Social Media: Ad spend of 7,542,323
- SEO: Ad spend of 7,740,904
- Email: Ad spend of 7,871,576
- PPC: Highest ad spend at 8,199,237
- Referral: Ad spend of 8,653,519

Funnel Analysis:

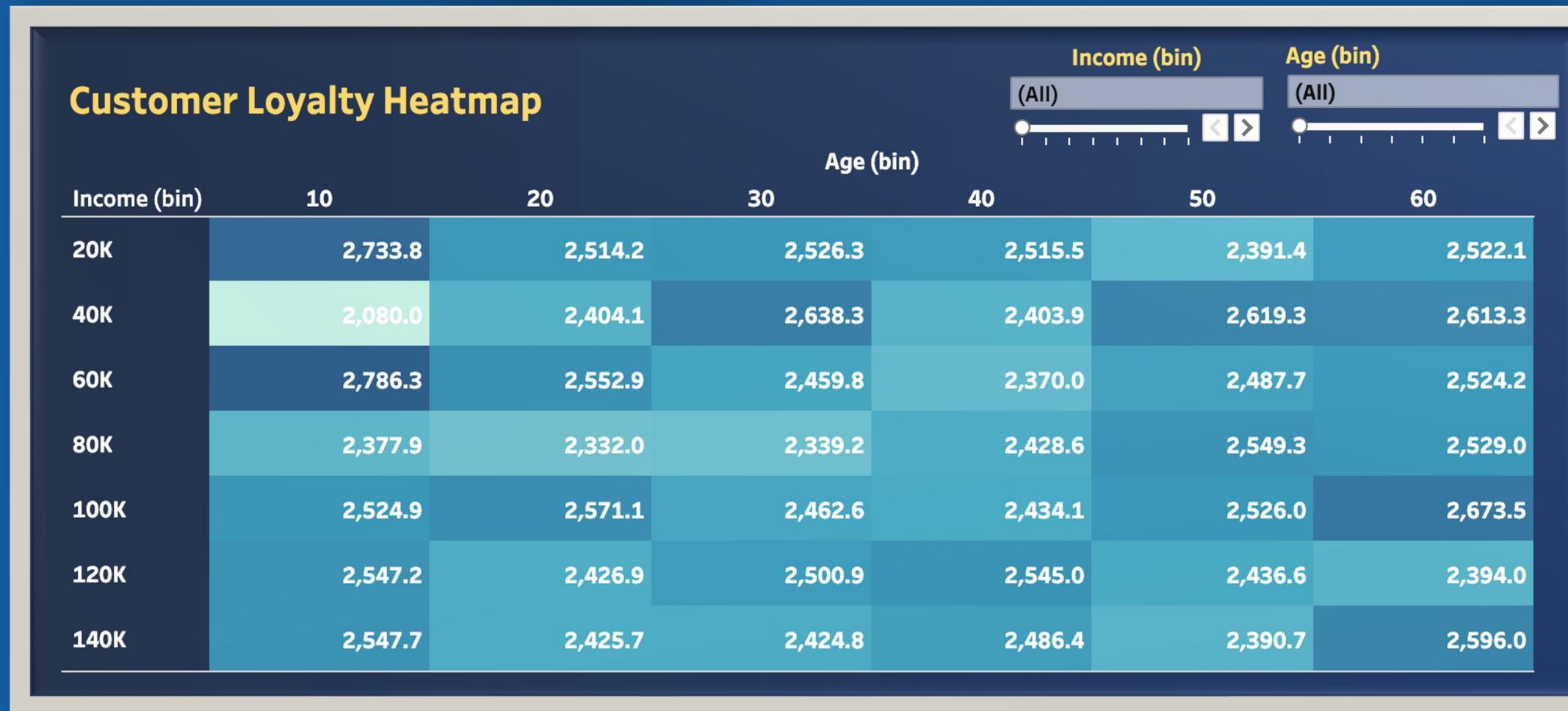
- Highlights the number of website visits, clicks, and conversions at each stage, indicating areas for improvement.

These insights can help optimize marketing strategies and allocate resources effectively.



Customer Loyalty Heatmap

The customer loyalty heatmap provides insights into how different income and age groups correlate with loyalty metrics. Key Observations include:



Combination of Age and Income:

- The highest loyalty is observed in the 50-year age group with 140K income, scoring 2,596.0.
- The 30-year age group with 100K income also shows significant loyalty, with scores around 2,571.1.

These insights suggest that targeting marketing efforts towards higher-income individuals in their 40s and 50s could enhance customer loyalty.

Income Bins:

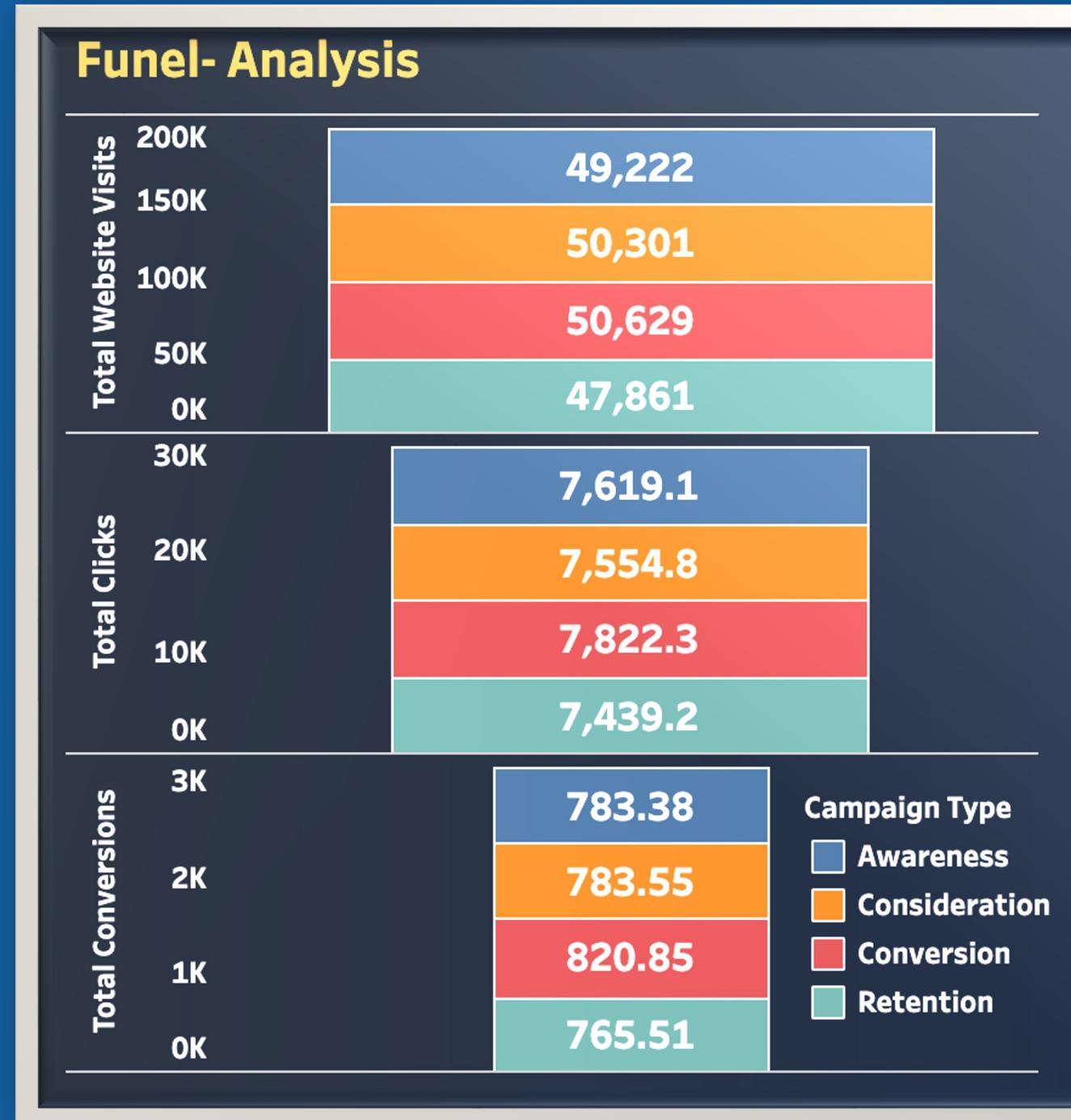
- Customers with an income of 100K show higher loyalty, with scores around 2,524.9 to 2,673.5.
- Loyalty tends to be lower in the 20K income bracket, with scores ranging from 2,080.0 to 2,514.2.

Age Bins:

- The 40-year age group shows consistent loyalty across various income levels, with scores between 2,428.6 and 2,571.1.
- Customers aged 50 and 60 also exhibit strong loyalty, particularly in higher income brackets.



Funnel Analysis



The funnel analysis illustrates the progression from website visits to conversions:

- Total Website Visits: 198,013
- Total Clicks: 40,222
- Total Conversions: 3,000

Insights

- Drop-off Points: A significant drop-off occurs between website visits and clicks, with only about 20% of visitors clicking through.
- Another notable drop-off is from clicks to conversions, where only about 7.5% of clicks result in conversions.

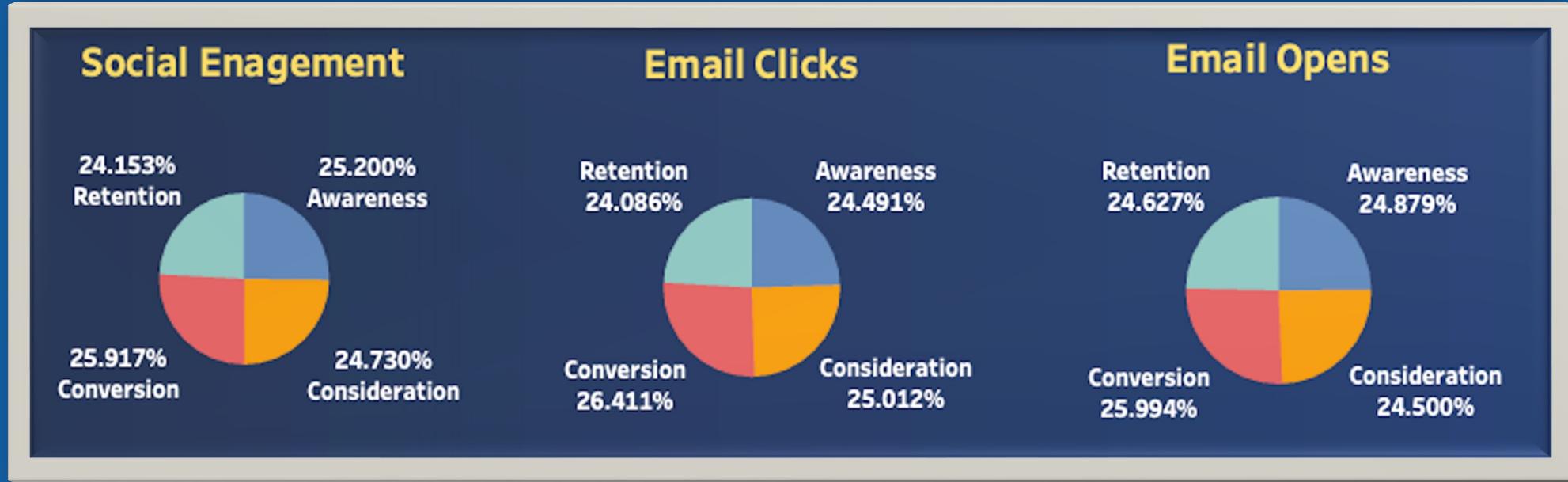
Areas for Improvement:

- Increase Click-Through Rate: Strategies like improving ad creatives and targeting could help reduce the initial drop-off.
- Enhance Conversion Rate: Optimizing landing pages and streamlining the checkout process could improve conversion rates.

Using different filters, such as campaign type or channel, can provide further insights into specific areas needing attention.



Social Engagement and Email Metrics



Correlation with Conversion Rates

- The correlation heatmap provides additional insights: Email clicks have a moderate positive correlation (0.13) with conversions, indicating their effectiveness in driving conversions.
- Social shares show a weak negative correlation (-0.011) with conversions, suggesting that while they may increase awareness, they might not directly lead to conversions.
- Time on site has a moderate positive correlation (0.13) with conversions, indicating that engaging content that keeps users on the site longer may lead to more conversions.

Engagement Overview

- Social engagement metrics show a significant impact on conversions:
- The highest percentage in social engagement is for conversion (25.917%), indicating that social media efforts are effectively driving conversions

Email Metrics

- Email marketing shows strong performance in both opens and clicks
- Email clicks have the highest conversion rate at 26.411%, slightly outperforming email opens (25.994%) and social engagement (25.917%) in driving conversions

These insights suggest that while social engagement is important for awareness and consideration, email marketing, particularly email clicks, is more effective in driving actual conversions. Strategies should focus on optimizing email campaigns and creating engaging on-site content to improve conversion rates.



CONCLUSION



Summary

The digital marketing dashboard provides a comprehensive overview of the campaign's performance, highlighting key metrics such as total ad spend, website visits, and conversion rates. The machine learning model, implemented using an XGBoost classifier, has effectively improved marketing strategies by optimizing ad spend and targeting specific customer segments. Key findings include:

- Total Ad Spend: \$40,007,559 with a focus on gender-based segmentation showing similar spending across genders.
- Website Performance: 198,013 total visits with an average time on site of 7.73 minutes and a click-through rate of 0.1548.
- Conversion Rate: Achieved an average conversion rate of 0.1044, indicating effective campaign strategies.
- Campaign Effectiveness: Email and PPC channels showed the highest conversion rates, suggesting a focus on these channels for future campaigns.
- Customer Loyalty: Heatmap analysis reveals higher loyalty among customers with incomes between 100K and 140K, suggesting targeted retention strategies.

Recommendations

To further enhance marketing strategies, consider the following areas for improvement and research:

- Advanced Segmentation: Implement more granular customer segmentation using additional demographic and behavioural data to tailor marketing efforts more precisely.
- A/B Testing: Conduct A/B tests on different campaign types and channels to identify the most effective strategies and optimize conversion rates.
- Predictive Analytics: Develop predictive models to forecast customer behaviour and optimize marketing spend accordingly.
- Cross-Channel Integration: Explore the integration of social media and email campaigns to create a cohesive marketing strategy that maximizes engagement and conversion.
- Real-Time Analytics: Implement real-time analytics to monitor campaign performance and make data-driven adjustments promptly.

These initiatives will help refine marketing strategies, enhance customer engagement, and improve overall campaign effectiveness.



THANK YOU

Connect with me.



+91 9891285304



sahilanand1416@gmail.com



[LinkedIn](#)



[GitHub](#)



[Drive Link to Project Repository](#)



[Click here to Access this Dashboard](#)

