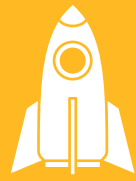# Customer Conversion Prediction

Digital Marketing Campaigns Analysis

Project 2: Machine Learning
Ironhack Data Analytics Course

Cristina Marco Tirot

Key Goals

Identify key factors that drive customer conversion and build a model to predict the likelihood of conversion based on these insights.

1 Analyze Customer and Engagement Patterns

Understand key customer demographics, behaviors, and engagement metrics to identify trends related to conversion.

2 Identify Key Predictors of Conversion

Determine which factors most strongly influence customer conversion through feature exploration and correlation analysis.

3 Develop and Evaluate a Predictive Model

Build and assess a model to predict conversion likelihood, focusing on actionable insights to improve targeted marketing efforts.
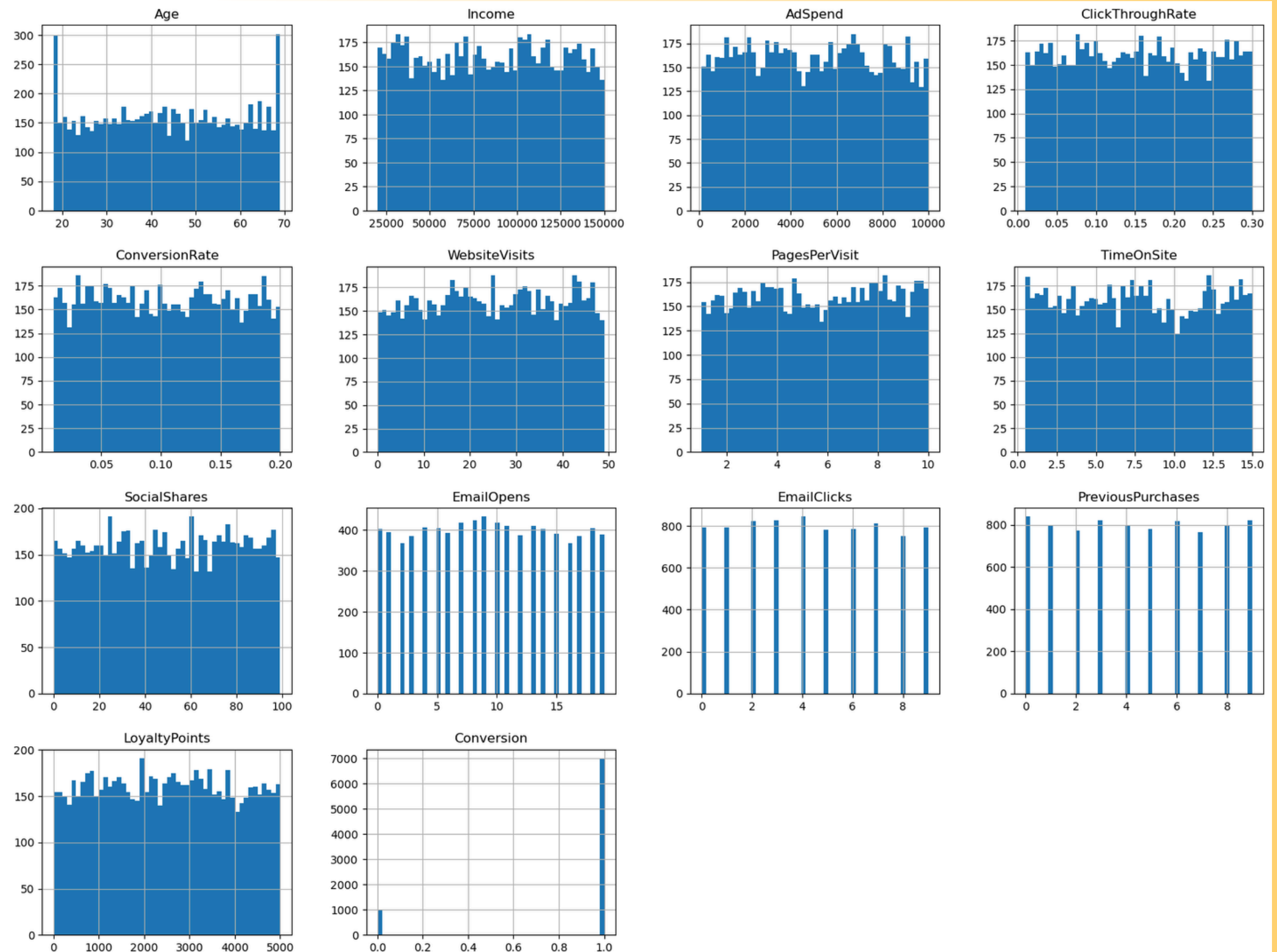
Analyze Customer and Engagement Patterns

Features

- Demographic information
- Marketing-specific Variables
- Customer Engagement Variables
- Historical Data

Target

- Conversion: Binary variable indicating whether the customer converted (1) or not (0).

Shape: 8,000 rows with 20 features

Identify Key Predictors of Conversion

Key Takeaways on Conversion Correlations

- Positive indicators: Longer site visits, email engagement, ad spend and past purchases.
- Minimal impact: Demographics like age, income, and gender.
- Complex Drivers: No single factor has a strong correlation with conversion, indicating that multiple variables likely work together to drive customer actions.

These findings suggest a need for a multi-factor approach in analyzing and predicting conversions.

Correlation Heatmap

Conversion	1.00
TimeOnSite	0.13
EmailClicks	0.13
EmailOpens	0.12
AdSpend	0.12
ClickThroughRate	0.12
PreviousPurchases	0.11
PagesPerVisit	0.10
CampaignType_Conversion	0.10
LoyaltyPoints	0.10
ConversionRate	0.09
WebsiteVisits	0.08
Income	0.01
CampaignChannel_Referral	0.01
CampaignChannel_PPC	0.01
Age	0.00
Gender_Male	0.00
CampaignChannel_SEO	0.00
Gender_Female	-0.00
CampaignChannel_Email	-0.01
SocialShares	-0.01
CampaignChannel_Social Media	-0.01
CampaignType_Retention	-0.03
CampaignType_Awareness	-0.04
CampaignType_Consideration	-0.04
AdvertisingPlatform_IsConfid	
AdvertisingTool_ToolConfid	

Develop and Evaluate a Predictive Model

Supervised learning: classification

- Logistic Regression achieved the highest accuracy (89.9%), excelling in overall correct classifications.
- Random Forest showed strong precision and low false negatives, making it a close second
- Decision Tree with Class Weight improved recall over the standard Decision Tree, balancing predictions better across classes.

Recommendation: For overall accuracy, Logistic Regression is best.

Comparison of Classification Metrics

Confusion Matrix Comparison
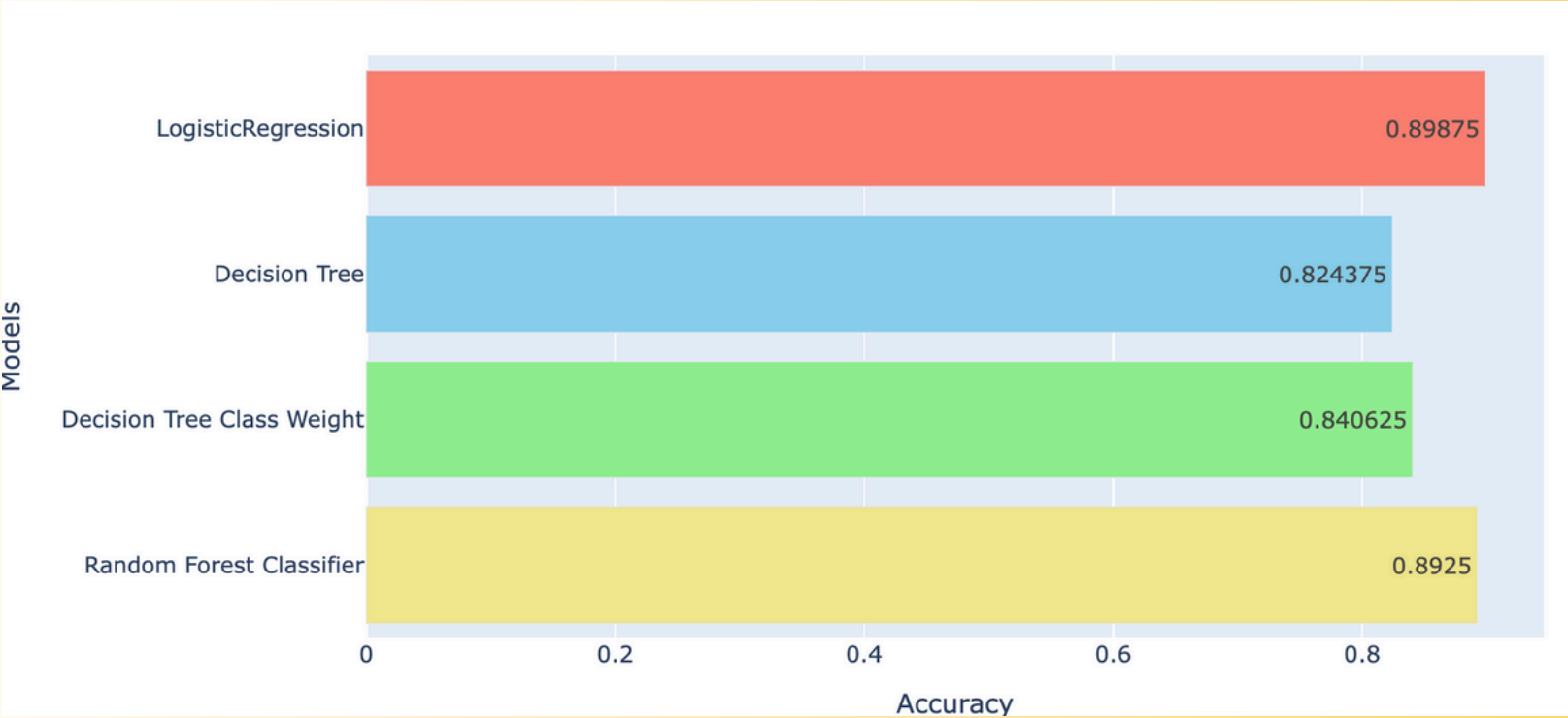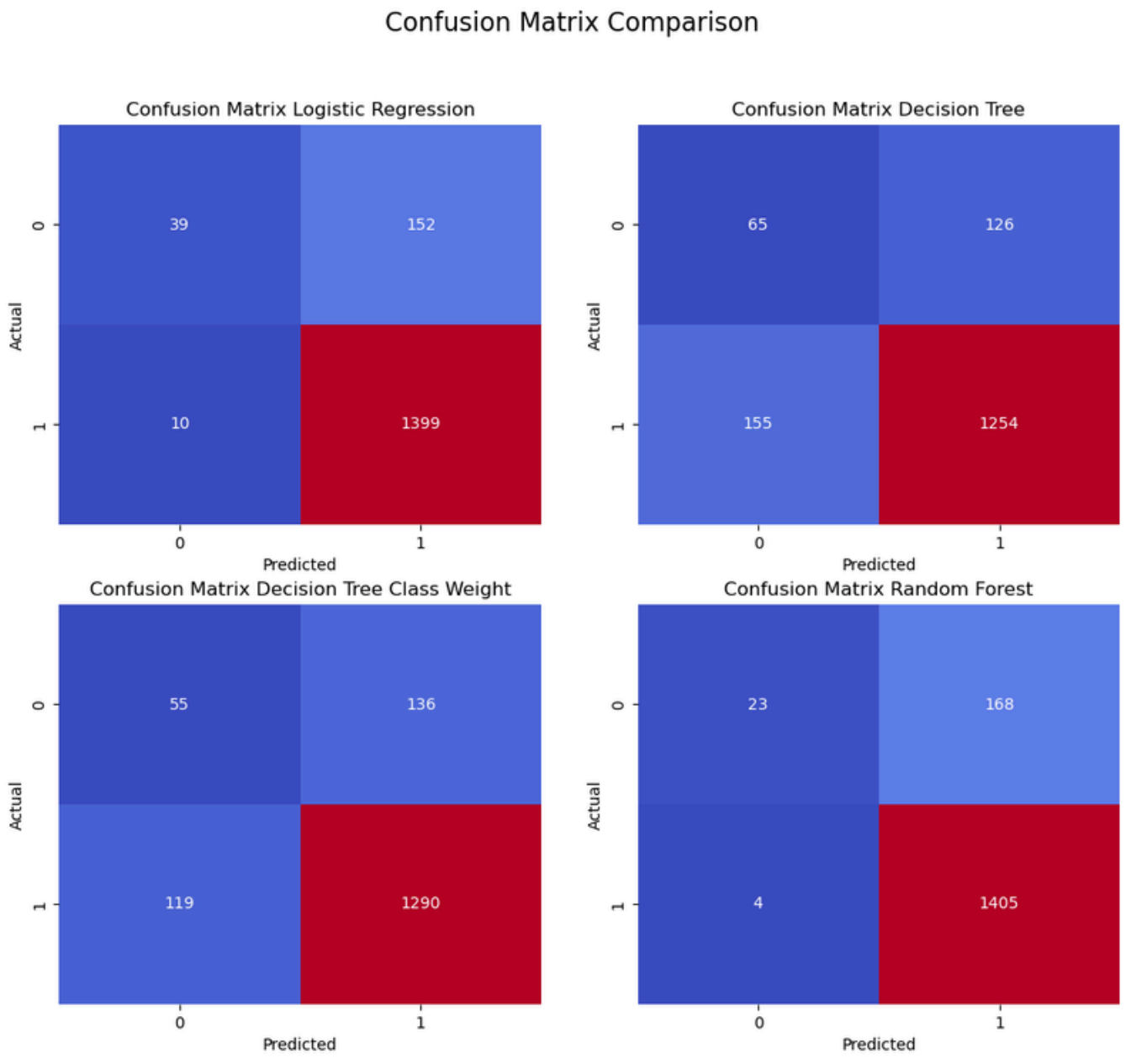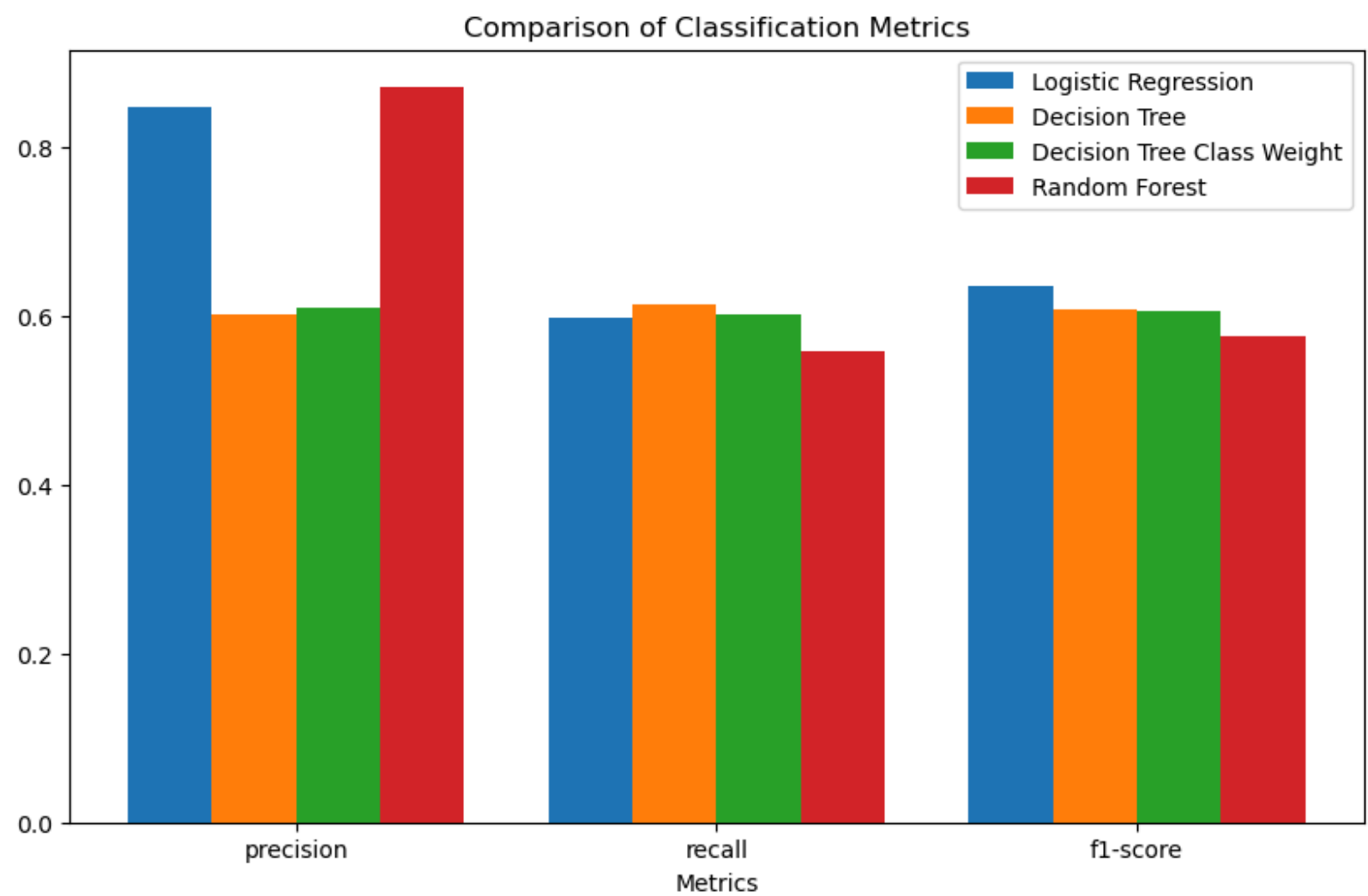
Identify Key Predictors of Conversion

Key Takeaways from Logistic Regression Coefficients:

- Positive Impact: AdSpend, Email Engagement and CTR boost conversions.
- Negative Impact: Consideration, Retention and Awareness Campaigns reduce conversions.
- Irrelevant: Some features like AdvertisingPlatform_IsConfid don't affect the model.

Conclusion: focusing on improving the Email Engagement and CTR can increase conversion likelihood, while campaigns targeting Consideration, Awareness and Retention may need to be reevaluated for better performance.

Logistic Regression Coefficients
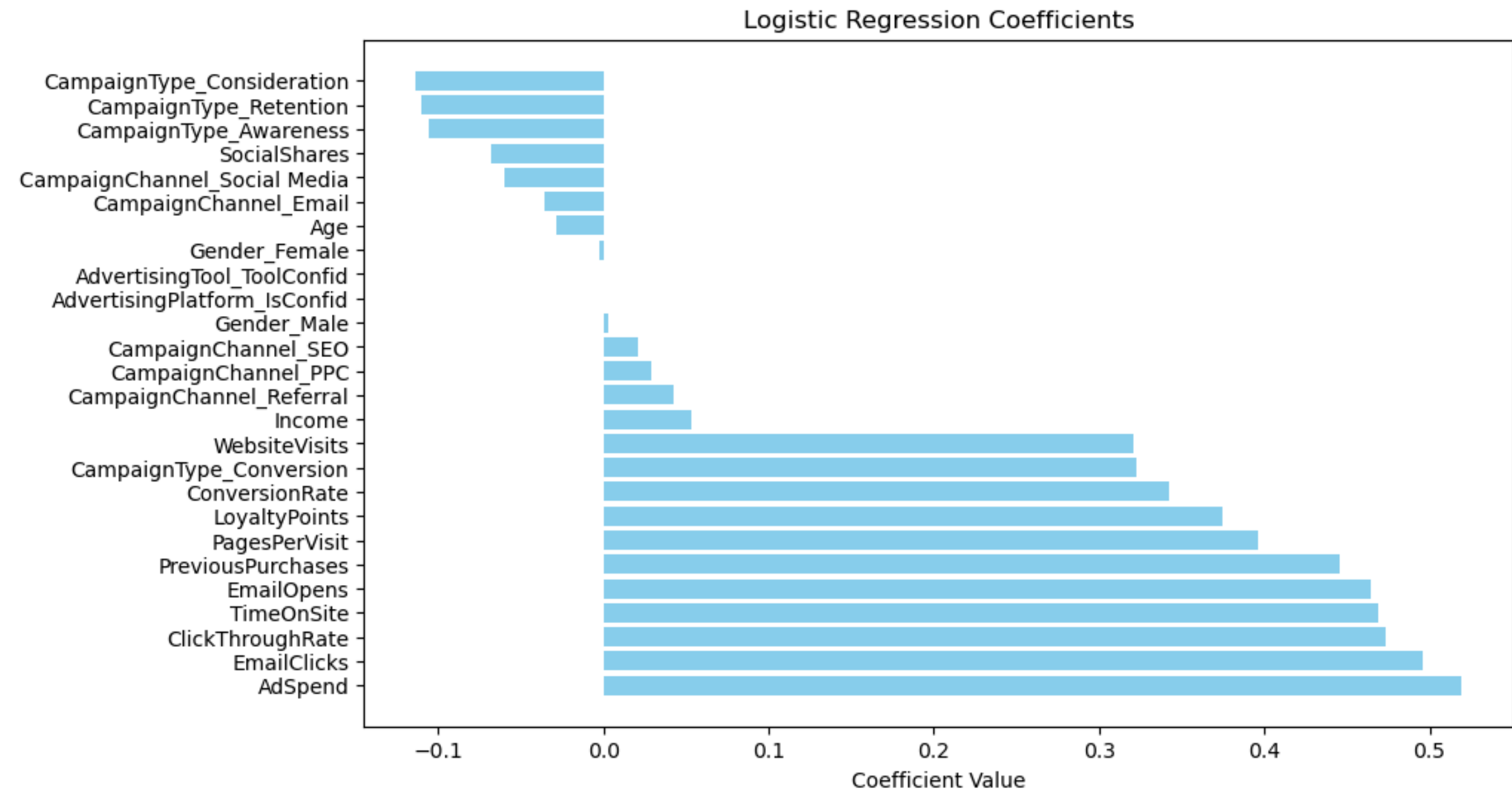
Thank you!

Feel free to contact me if you have any questions