

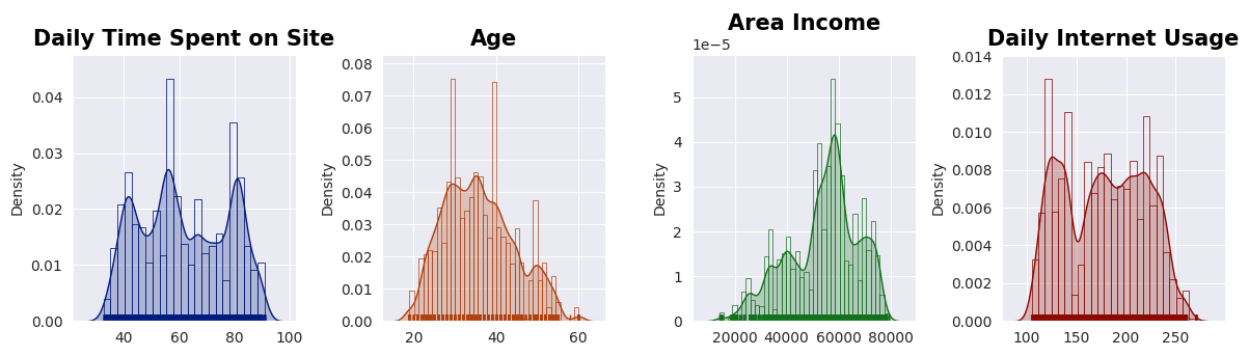
# Introduction: Predicting Click-Through Rates (CTR) in Online Advertising

Click-Through Rate (CTR) plays a pivotal role in the success of online advertising campaigns. As users interact with ads, understanding their likelihood to click becomes crucial for platforms and marketers alike. In this report, we delve into user interaction data to predict whether a visitor will click on an online advertisement.

## Dataset Overview

The dataset at hand contains 10 columns, with 9 features representing data instances. Our target variable is the “Clicked on Ad” column, indicating whether a user clicked the ad. Let’s explore this data to uncover insights and build an effective predictive model.

## EDA ANALISYYS



### 1. Daily Time Spent on Site:

- The histogram shows two overlapping distributions (in blue and orange). Users' time spent on the site appears to follow a bimodal pattern.
- Consider investigating further to understand the reasons behind these two peaks. Are there specific user segments (e.g., mobile vs. desktop users) driving this behavior?

### 2. Age Distribution:

- Again, we see two overlapping distributions (in blue and orange). The age groups seem to cluster around certain values.
- Explore whether age correlates with ad engagement. Are certain age ranges more likely to click on ads?

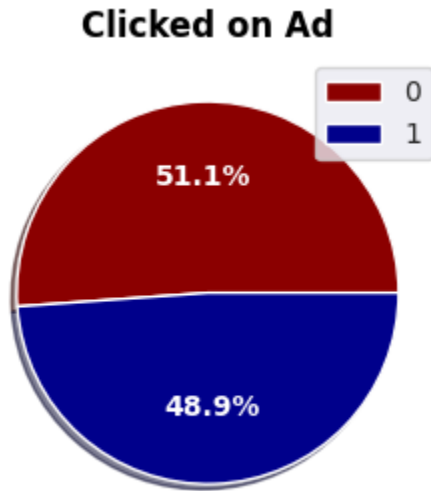
### 3. Area Income:

- The green and red distributions represent income levels across an area.
- Consider segmenting users by income brackets and analyzing their ad interaction patterns. Do higher-income individuals engage differently with ads?

### 4. Daily Internet Usage:

- The green and red densities show internet usage time.

- Investigate whether users who spend more time online exhibit different ad-click behavior.



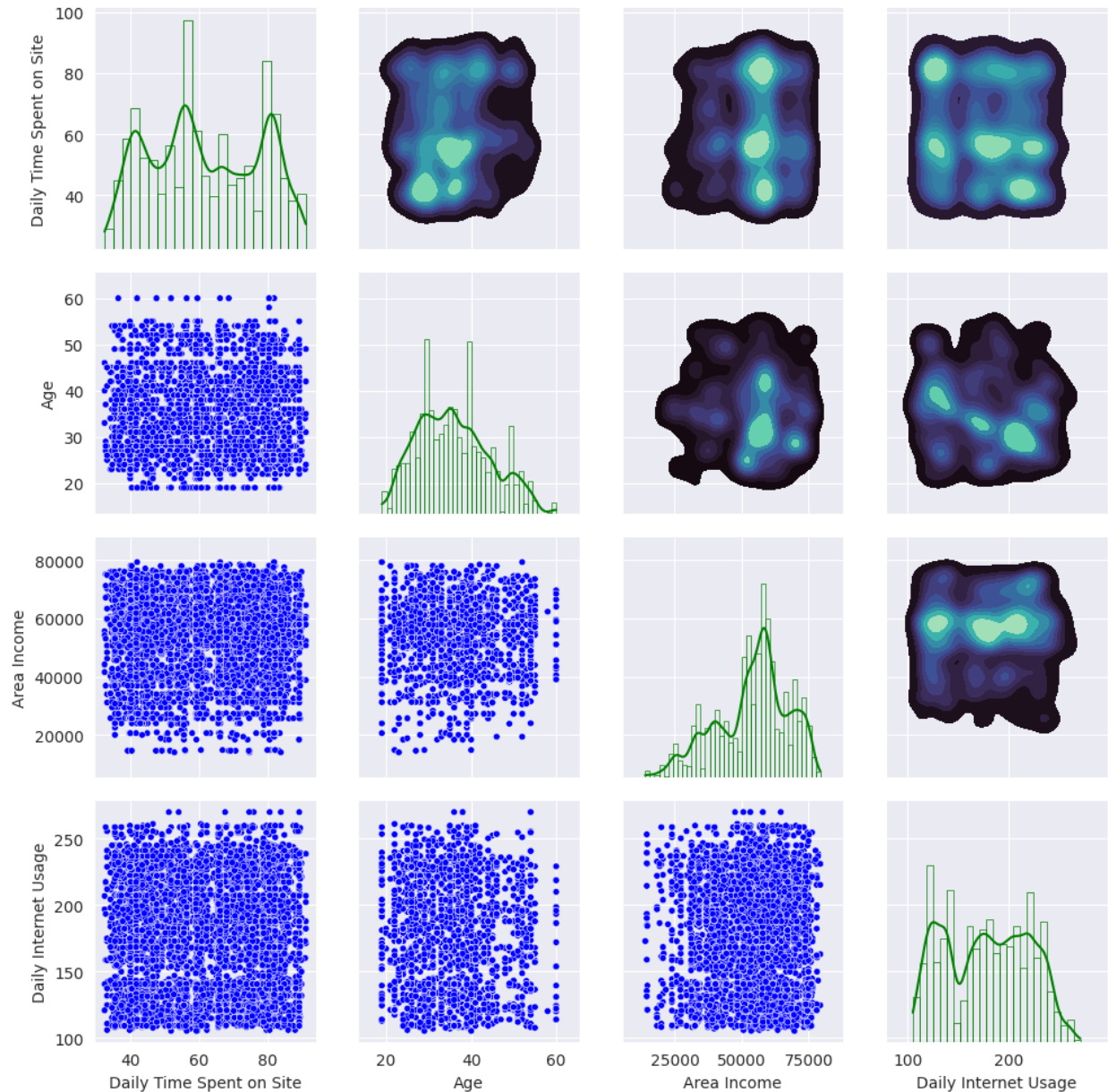
## Clicked on Ad Distribution

The pie chart illustrates the distribution of user interactions with ads. Here are the key points:

- **51.1%** of users clicked on the ad (represented by the red section).
- **48.9%** did not click on the ad (represented by the blue section).

## Interpretation

- The majority of users engaged with the ad, but a significant portion did not.
- Consider exploring factors that influence ad engagement, such as demographics, time of day, or ad content.



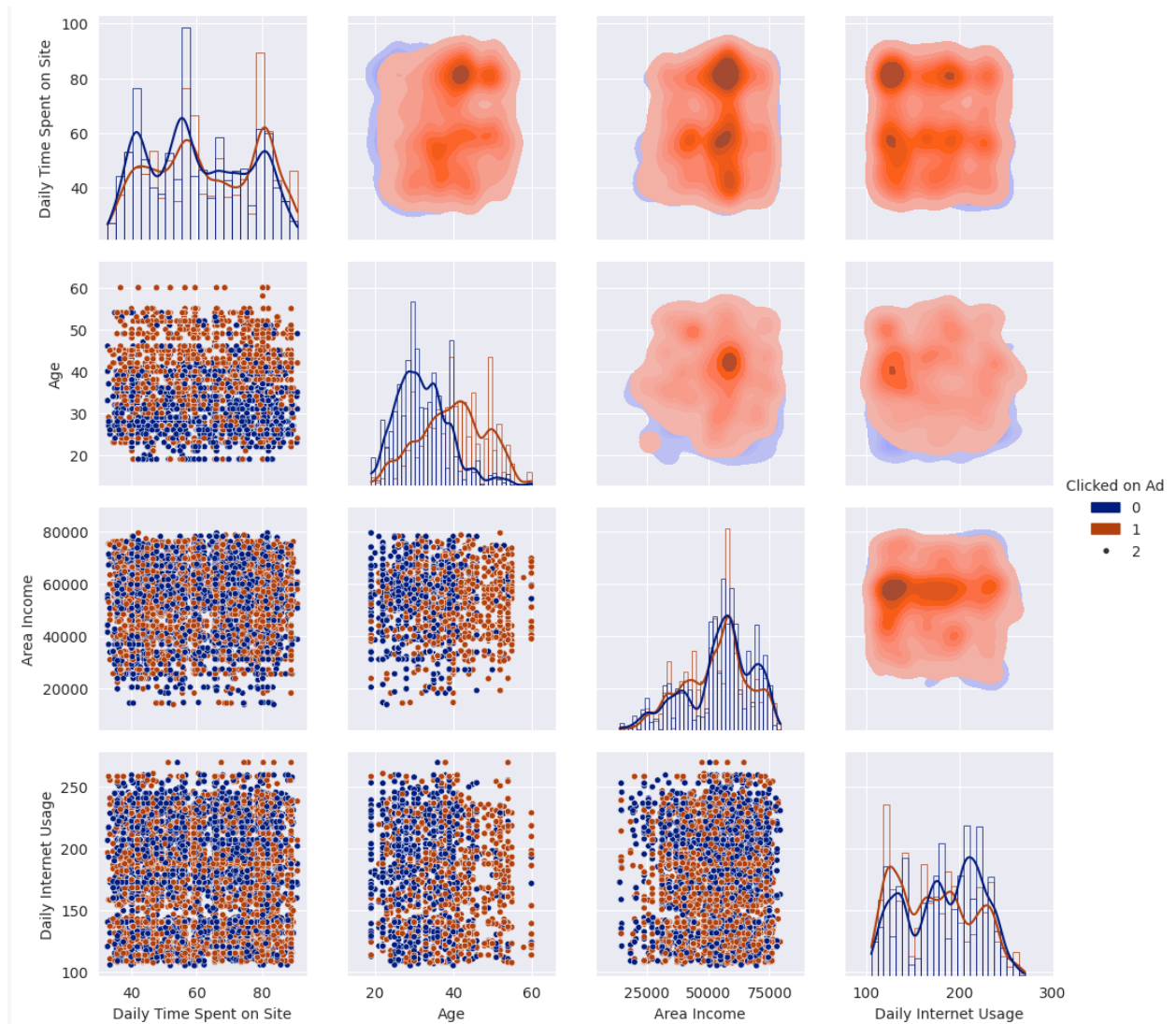
## Multivariate Analysis: Relationships Between Variables

The matrix of plots provides a comprehensive view of how different variables interact within the dataset. Here are the key observations:

### 1. Daily Time Spent on Site vs. Age:

- The scatter plot shows a negative correlation between age and time spent on the site. Younger users tend to spend more time on the site.

- Consider exploring age-specific ad engagement patterns.
- 2. **Daily Time Spent on Site vs. Area Income:**
  - No clear linear relationship is evident. Users across different income levels exhibit varying time spent on the site.
  - Investigate whether other factors (e.g., content relevance, ad placement) influence this behavior.
- 3. **Daily Time Spent on Site vs. Daily Internet Usage:**
  - A positive correlation exists between time spent on the site and daily internet usage.
  - Users who spend more time online tend to engage more with ads.
- 4. **Age vs. Area Income:**
  - No strong linear relationship: income distribution varies across age groups.
  - Consider segmenting users by income brackets and analyzing ad interactions.
- 5. **Age vs. Daily Internet Usage:**
  - No clear pattern: users of different ages exhibit diverse internet usage habits.
  - Explore whether specific age groups respond differently to ad content.
- 6. **Area Income vs. Daily Internet Usage:**
  - No strong correlation: income and internet usage are not directly related.
  - Investigate other factors influencing ad engagement.



## Multivariate Analysis: Relationships Between Variables

The matrix of scatter plots and density plots provides valuable insights into the relationships between different variables. Here are the key observations:

1. **Daily Time Spent on Site vs. Age:**
  - The scatter plot reveals a negative correlation between age and time spent on the site. Younger users tend to spend more time browsing.
  - Consider exploring age-specific ad engagement patterns further.
2. **Daily Time Spent on Site vs. Area Income:**
  - No clear linear relationship is evident. Users across different income levels exhibit varying time spent on the site.
  - Investigate whether other factors (content relevance, ad placement) influence this behavior.
3. **Daily Time Spent on Site vs. Daily Internet Usage:**

- A positive correlation exists between time spent on the site and daily internet usage.
- Users who spend more time online tend to engage more with ads.
- 4. **Age vs. Area Income:**
  - Income distribution varies across age groups, but there is no strong linear relationship.
  - Consider segmenting users by income brackets and analyzing ad interactions.
- 5. **Age vs. Daily Internet Usage:**
  - No clear pattern: users of different ages exhibit diverse internet usage habits.
  - Explore whether specific age groups respond differently to ad content.
- 6. **Area Income vs. Daily Internet Usage:**
  - No strong correlation: income and internet usage are not directly related.
  - Investigate other factors influencing ad engagement.