ClickPredict - Enhancing Online Advertising Effectiveness

Introduction:

In the realm of online advertising, optimizing click-through rates (CTRs) is paramount for advertisers to maximize their return on investment (ROI). Understanding the factors that influence user clicks can enable more targeted and effective advertising campaigns.

The project "Click Predict - Enhancing Online Advertising Effectiveness" aims to optimize the impact of online advertisements by analyzing user interaction data. With the rise of digital marketing, understanding how different demographics engage with ads is crucial for businesses to improve their targeting strategies and maximize return on investment. This project utilizes a combination of exploratory data analysis (EDA) and predictive analytics to uncover key insights about user behavior and ad performance.

Objective:

Click-Through Rate (CTR) Analysis: Calculate the click-through rates for various ad categories, positions, and user demographics to determine which factors significantly impact user engagement.

Identify key factors: Determine the most significant factors that influence user clicks, such as user demographics, ad characteristics, and contextual information.

Optimize advertising campaigns: Provide insights and recommendations to advertisers for improving the effectiveness of their campaigns by targeting the right users with the right ads.

Enhance ROI: Help advertisers maximize their return on investment by optimizing their advertising spend and improving click-through rates.

Gain a deeper understanding: Explore the underlying factors that drive user behavior and ad performance.

Contribute to the field: Advance the understanding of online advertising by providing a data-driven approach to improving click-through rates.

Dataset Overview:

The dataset titled **“ad_click_data”** consists of 10 columns, which are as follows:

- id
- user_id
- ad_id
- ad_category
- ad_position
- click
- impression_date
- user_age
- user_gender
- user_location

Connecting to MySQL Database and Fetching Data:

To begin with our analysis, we first connect our MySQL database to using a MySQL Connector. After successfully connecting to our database we then fetch the dataset **“ad_click_data”** from the database.

Exploratory Data Analysis:

Age Distribution Analysis

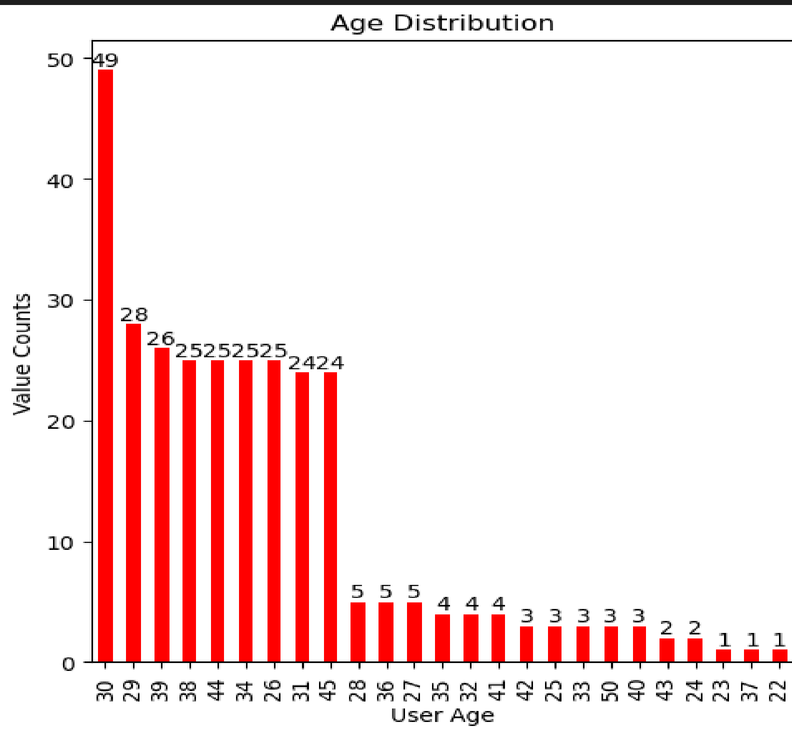
Goal: The aim of this analysis is to explore the distribution of users across different age groups, which can provide insights into which age segments are most represented in the user base.

In this analysis, a bar plot was generated to show the **age distribution** of users. Each bar represents the number of users within a specific age group, with the height of the bars indicating the count of users for each age. This visualization helps identify which age groups have a higher user count, offering insights into the age segments that are more active or engaged. Understanding these peaks can assist in focusing advertising efforts on the most prevalent age groups, thus optimizing campaign reach and effectiveness. The distribution also helps identify potential gaps in user engagement across different age ranges, which could inform strategies for expanding reach.

Code Snippet:

```
plt.subplot(1,2,1)
```

```
age_counts=df['user_age'].value_counts().plot.bar(xlabel='User Age',
ylabel='Value Counts',color='red', title='Age Distribution')

age_counts.bar_label(age_counts.containers[0])
```



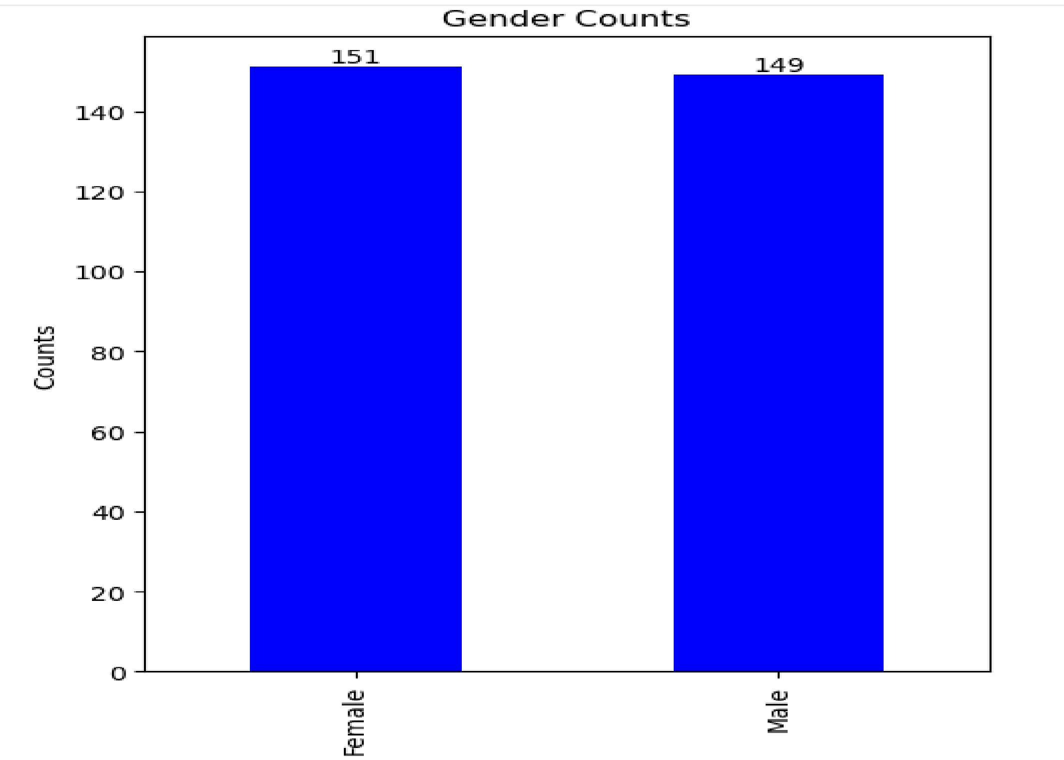
Gender Distribution Analysis

Goal: The goal of this analysis is to understand the distribution of users based on gender, helping to identify which gender groups are more represented among the users. This insight aids in designing targeted ad campaigns that resonate more effectively with the dominant user group.

The bar chart visualizes the count of users by gender.The gender distribution plot reveals that there is no significant difference in proportion of male and female users within the dataset. Understanding this balance allows for better targeting of ad content to suit the preferences of each gender.

Code Snippet:

```
plt.subplot(1,2,2)
gender_counts=df['user_gender'].value_counts().plot.bar(xlabel='Gender',yl
abel='Counts', title='Gender Counts',color='blue')
gender_counts.bar_label(gender_counts.containers[0])
```



CTR by Ad Category

Goal: The purpose of this analysis is to examine the performance of different ad categories based on the total number of clicks they receive.

The bar chart displays the total number of clicks for each ad category, highlighting which categories attract the most user engagement. Categories like Travel, Electronics and Toys show highest CTR.By analyzing the click distribution across various categories, advertisers can determine which ad types resonate more with users, allowing them to allocate their resources to the more successful categories. Higher click counts indicate that users find certain categories more relevant or appealing, while lower counts may suggest areas that need improvement or re-evaluation. This analysis enables advertisers to optimize their ad strategy by focusing on categories with proven user interest, thereby improving the overall effectiveness of ad campaigns.

Code Snippet:

```
lt.subplot(1,2,1)
CTR_category=df.groupby('ad_category')['click'].sum().plot.bar(xlabel='Cat
egories',ylabel='Clicks',color='darkblue')
CTR_category.bar_label(CTR_category.containers[0])
```

Analysis of Click-Through Rates by Gender and Ad Category

Goal: The goal of this analysis is to explore how different ad categories perform across male and female users by analyzing the number of clicks each gender generates for different ad types.

This analysis utilizes a bar plot to illustrate the aggregated click data for each advertisement category, segmented by user gender. The visualization reveals that categories such as Travel, Electronics, and Toys experience significantly higher engagement from female users, indicating a strong interest and suggesting potential for targeted marketing campaigns in these areas.In contrast, male users showed engagement based on categories.

Code Snippet:

```
plt.subplot(1,2,2)
CTR_gender_category=df.groupby(['ad_category','user_gender'])['click'].sum
().reset_index()
sns.barplot(data=CTR_gender_category,      x='ad_category',      y='click',
hue='user_gender')
plt.xticks(rotation=90)
```

Analysis of Click-Through Rates by Ad Position

Goal: The goal of this analysis is to evaluate the impact of advertisement positioning on click-through rates (CTR).

This analysis employs a bar plot to depict the total number of clicks received by advertisements according to their position on the page. The visualization reveals that ads located at the side consistently receive the highest click counts, indicating that users are more likely to engage with prominently displayed content. Whereas, ads situated in top and bottom positions experienced fewer clicks, but the difference is not drastic.

Code Snippet:

```
lt.subplot(1,2,2)

CTR_ad_position_categories=df.groupby(['ad_category','ad_position'])['clic
k'].sum().reset_index()

sns.barplot(data=CTR_ad_position_categories,    x='ad_category',y='click',
hue='ad_position')

plt.xlabel('Categories')
```

```
plt.ylabel('Click Count')

plt.title('Click for each category with respect to each position')
```

Analysis of Click-Through Rates by Ad Category and Position

Goal: The goal of this analysis is to explore the interaction between advertisement categories and their positions on user engagement, as measured by click-through rates (CTR).

This analysis utilizes a bar plot to visualize the total clicks received by each advertisement category, segmented by their respective positions on the page. The visualization reveals that Travels category shows highest and equal CTR for all the three Ad positions indicating a high preference for Travel Ads.The graph facilitates a comparison across categories, showing that while some categories consistently perform well in specific positions, others may require adjustments to their placement strategy to improve click rates. This analysis underscores the importance of optimizing both category selection and ad positioning to maximize user engagement and overall advertising effectiveness.

Code Snippet:

```
plt.subplot(1,2,1)
CTR_ad_positions=df.groupby('ad_position')['click'].sum().plot.bar(xlabel=
'AD Position',ylabel='Click Count', title='Click per position',
color='darkblue')
CTR_ad_positions.bar_label(CTR_ad_positions.containers[0])
```

Analysis of Click-Through Rates by User Location

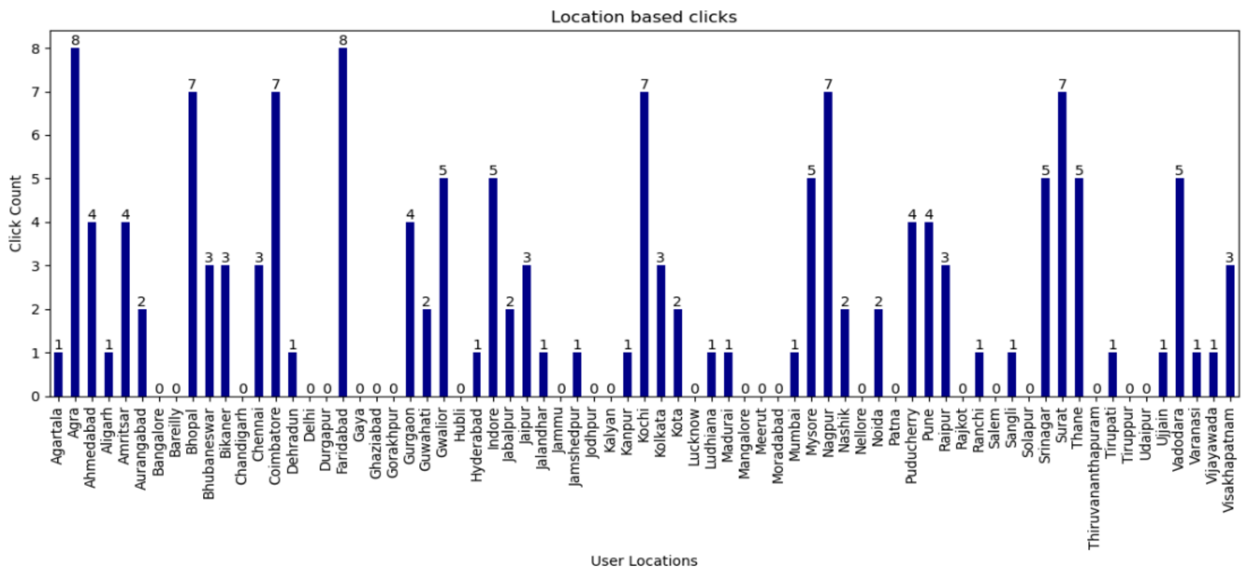
Goal: The goal of this analysis is to assess how user location influences click-through rates (CTR) for advertisements. Understanding geographic engagement patterns can help advertisers tailor their campaigns to specific audiences more effectively.

This analysis utilizes a bar plot to illustrate the total number of clicks generated by users based on their geographic locations. The visualization reveals significant variations in user engagement across different regions, with locations like Agartala and Faridabad showing highest CTR. A diverse range of variations can be seen across the visualization and emphasizes the importance of considering geographic factors in advertising strategies to maximize reach and engagement.

Code Snippet:

```
plt.figure(figsize=(12,6))
```

```
CTR_location=df.groupby('user_location')['click'].sum().plot.bar(xlabel='U
ser Locations',ylabel='Click Count', title='Location based clicks',
color='darkblue')
CTR_location.bar_label(CTR_location.containers[0])
plt.tight_layout()
plt.show()
```



Analysis of Click-Through Rates by Ad Category

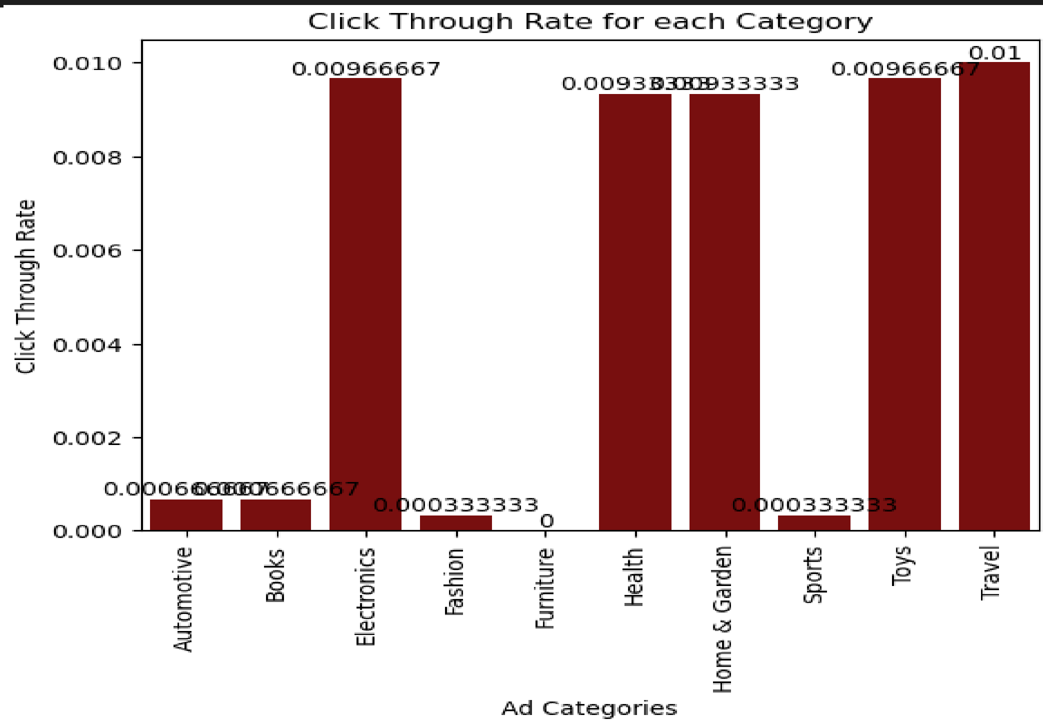
Goal: The goal of this analysis is to calculate the click-through rates (CTR) for different advertisement categories by examining both the total number of clicks and the total number of impressions.

This analysis utilizes a bar plot to display the calculated CTR for each advertisement category. The visualization reveals distinct performance levels across categories, highlighting which ads are most effective in converting impressions into clicks.For instance, categories with a high CTR indicate that users are engaging with the ads more frequently, suggesting strong alignment between the ad content and user interests. Conversely, categories with lower CTR values may signify areas where advertising strategies need refinement, such as improved targeting or more compelling messaging.

Code Snippet:

```
CTR_category_plot=sns.barplot(data=ctr_category, x='ad_category',y='CTR',
color='darkred')
CTR_category_plot.bar_label(CTR_category_plot.containers[0])
```

```
plt.xlabel('Ad Categories')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Category")
plt.xticks(rotation=90)
plt.show()
```



Analysis of Click-Through Rates by User Location

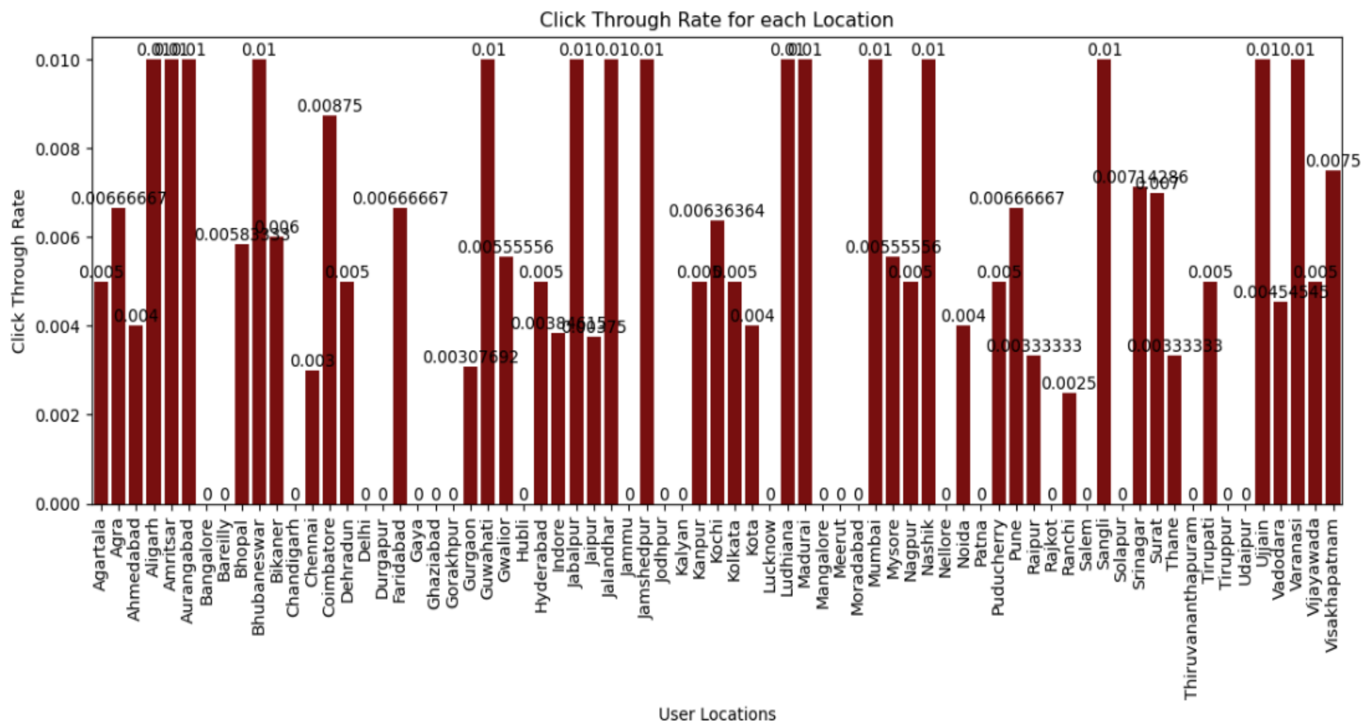
Goal: The goal of this analysis is to calculate the click-through rates (CTR) based on user locations, providing insights into how geographic factors influence user engagement with advertisements.

This analysis employs a bar plot to illustrate the CTR for each user location, providing a visual comparison of engagement levels across different regions. The visualization highlights significant differences in user interactions with ads, revealing that some locations demonstrate much higher CTR values than others.

Code Snippet:

```
plt.figure(figsize=(12,6))
CTR_location_plot=sns.barplot(data=ctr_location,
x='user_location',y='CTR', color='darkred')
CTR_location_plot.bar_label(CTR_location_plot.containers[0])
```

```
plt.xlabel('User Locations')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Location")
plt.xticks(rotation=90)
plt.tight_layout()
plt.show()
```



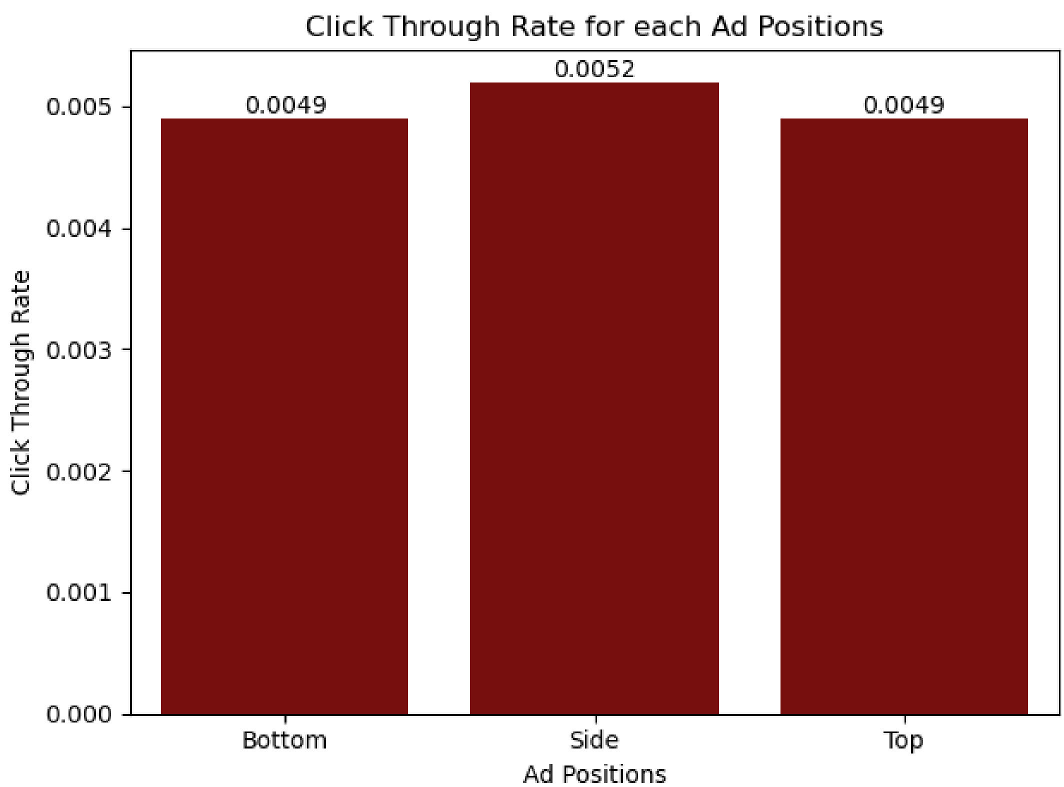
Analysis of Click-Through Rates by Ad Position

Goal: The goal of this analysis is to visualize the click-through rates (CTR) associated with various advertisement positions on the page. This visualization aims to provide insights into how ad placement impacts user engagement.

This analysis employs a bar plot to depict the total number of clicks received by advertisements according to their position on the page. The visualization reveals that ads located at the side consistently receive the highest click-through rate, indicating that users are more likely to engage with prominently displayed content.This analysis underscores the critical importance of strategic ad positioning in maximizing user engagement and enhancing overall advertising effectiveness.

Code Snippet:

```
CTR_position_plot=sns.barplot(data=ctr_position,  x='ad_position',y='CTR', color='darkred')

CTR_position_plot.bar_label(CTR_position_plot.containers[0])

plt.xlabel('Ad Positions')

plt.ylabel('Click Through Rate')

plt.title("Click Through Rate for each Ad Positions")

plt.tight_layout()

plt.show()
```



Analysis of Click-Through Rates by User Age

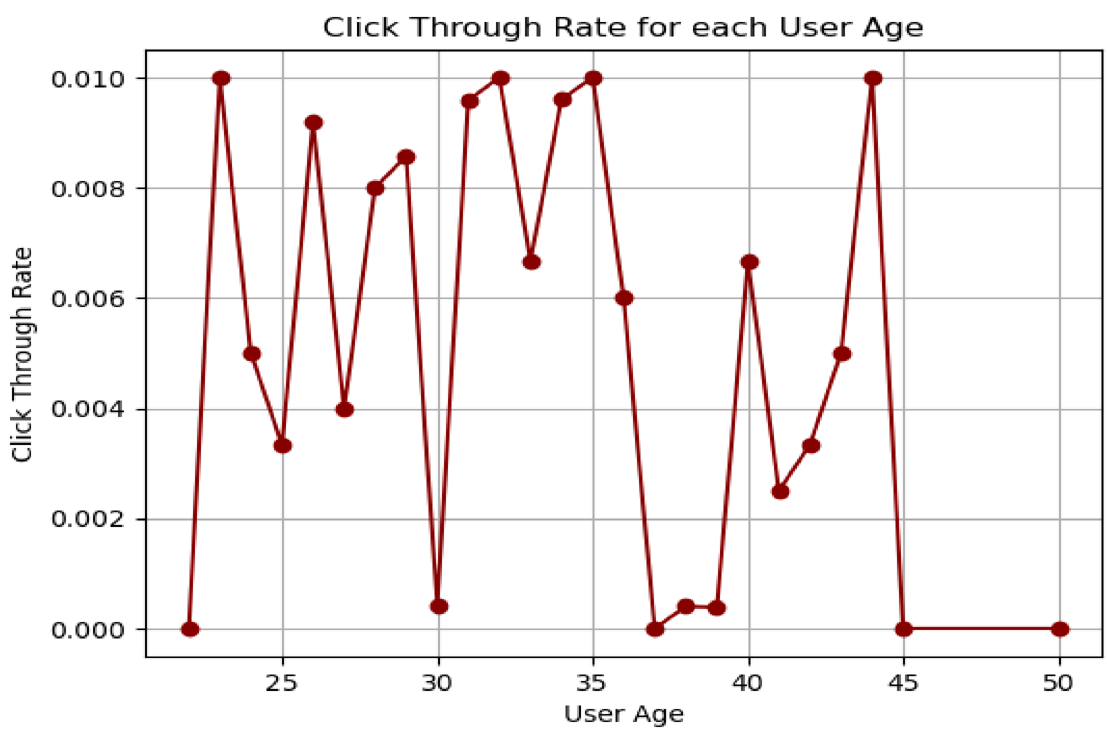
Goal: The goal of this analysis is to visualize the click-through rates (CTR) for different user age groups.

Description

This analysis utilizes a line plot to illustrate the CTR for each user age group, enabling a clear visualization of how engagement varies with age. The plot highlights distinct trends, showing that certain age groups demonstrate significantly higher CTRs, indicating stronger engagement with advertisements.The analysis of click-through rates (CTRs) across different user age groups reveals that there is no consistent pattern. While some age groups, such as 25 and 45, exhibit significantly higher CTRs, others, like 30 and 40, demonstrate lower rates. This suggests that advertisers may benefit from tailoring their campaigns to specific age demographics to maximize engagement and improve overall performance.

Code Snippet:

```
CTR_age_plot=plt.plot(ctr_age['user_age'],ctr_age['CTR'], marker='o',
color='darkred')
plt.xlabel('User Age')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each User Age")
plt.grid(True)
plt.show()
```



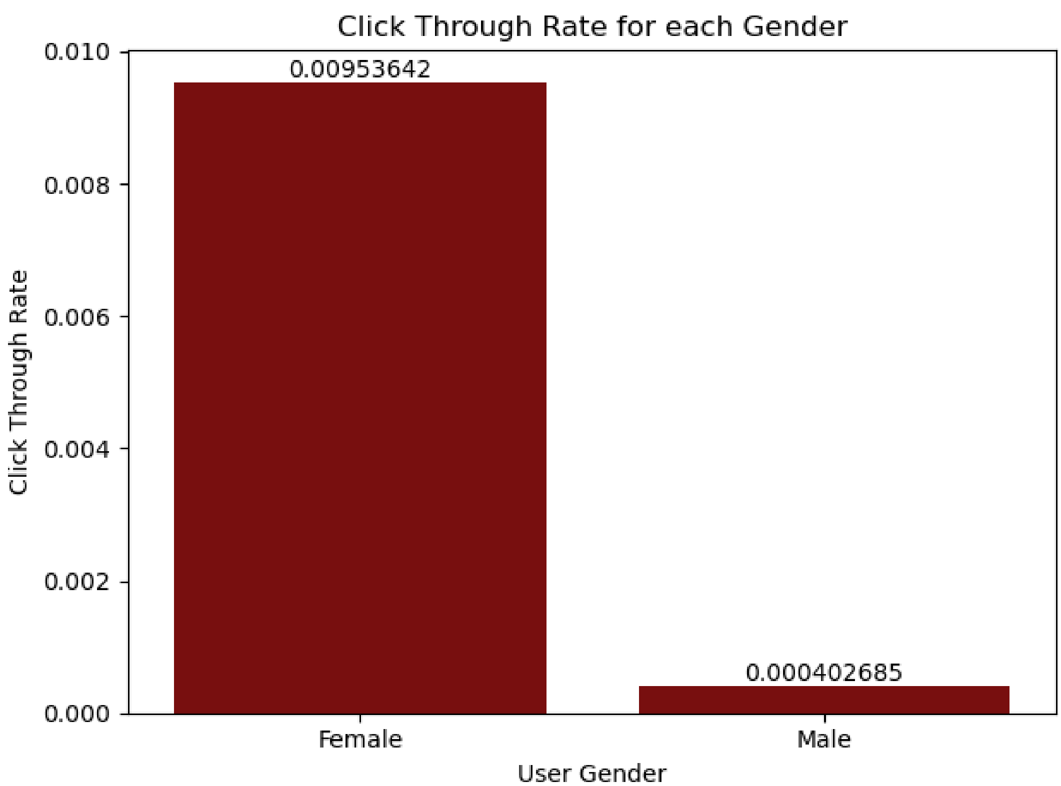
Analysis of Click-Through Rates by User Gender

Goal: The goal of this analysis is to visualize the click-through rates (CTR) based on user gender.

The graph demonstrates a significant disparity in click-through rates (CTRs) between genders. Females exhibit a markedly higher CTR of approximately 0.0095 compared to males, whose CTR is only 0.0004. This substantial gender gap suggests that targeting females with advertisements could be a more effective strategy for generating clicks and improving campaign performance. This analysis emphasizes the importance of gender-specific insights in advertising strategies to enhance engagement and improve overall campaign effectiveness.

Code Snippet:

```
CTR_gender_plot=sns.barplot(data=ctr_gender,      x='user_gender',y='CTR',
color='darkred')
CTR_gender_plot.bar_label(CTR_gender_plot.containers[0])
plt.xlabel('User Gender')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Gender")
plt.tight_layout()
plt.show()
```



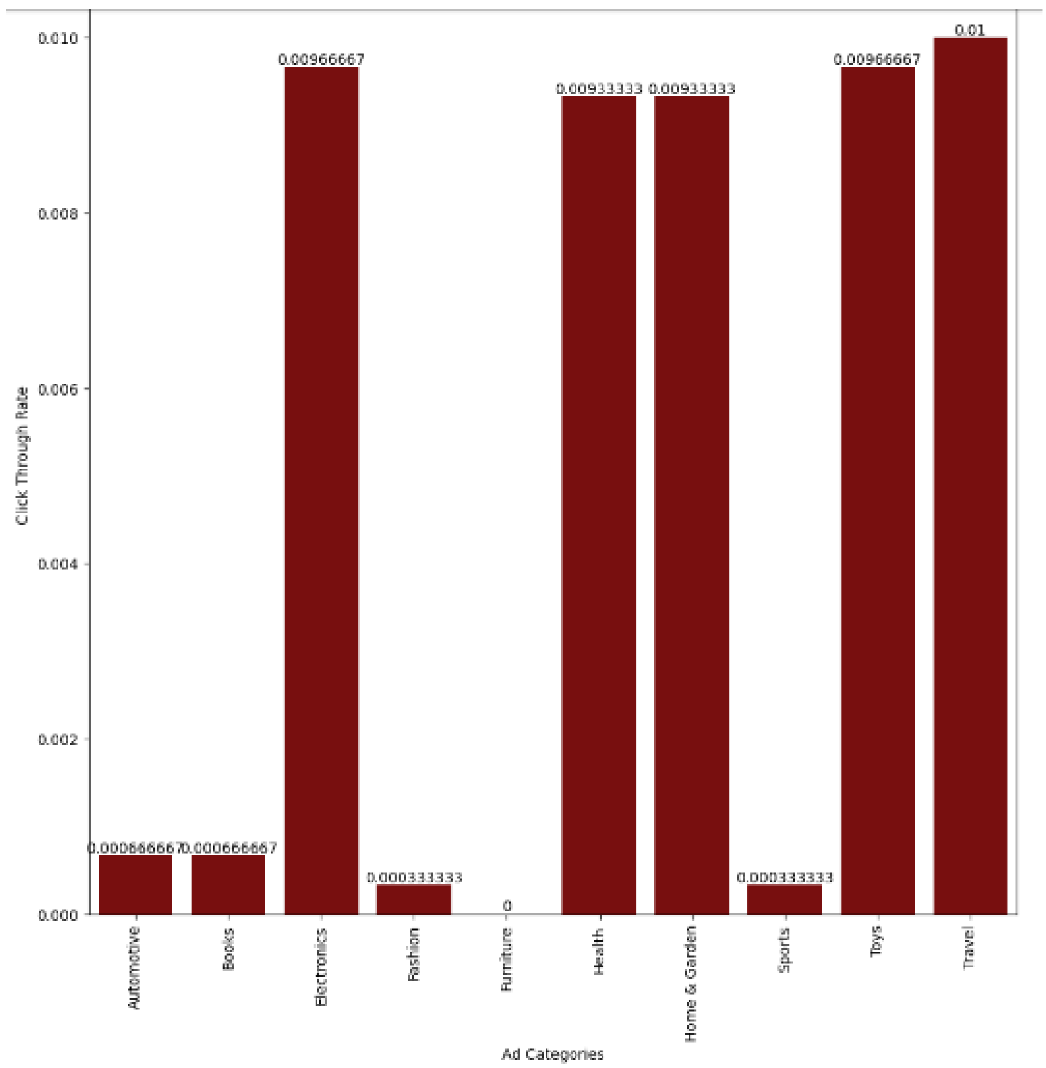
Analysis of Click-Through Rates by Ad Category

Goal: The goal of this analysis is to visualize the click-through rates (CTR) across different advertisement categories.

This analysis uses a bar plot to illustrate the CTR for each advertisement category, making it easy to compare the performance of different ad types. The graph demonstrates a significant disparity in click-through rates (CTRs) across different ad categories. While categories such as "Toys" and "Travel" exhibit notably higher CTRs, others, like "Furniture" and "Automotive," display considerably lower rates. This suggests that advertisers may benefit from tailoring their campaigns to specific ad categories to maximize engagement and improve overall performance. However, further investigation into the factors influencing CTR variations across different categories is necessary to gain a deeper understanding and develop more effective targeted advertising strategies.

Code Snippet:

```
#CTR by category
plt.subplot(2,2,1)
CTR_category_plot=sns.barplot(data=ctr_category, x='ad_category',y='CTR',
color='darkred')
CTR_category_plot.bar_label(CTR_category_plot.containers[0])
plt.xlabel('Ad Categories')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Category")
plt.xticks(rotation=90)
```



Analysis of Click-Through Rates by Ad Positions

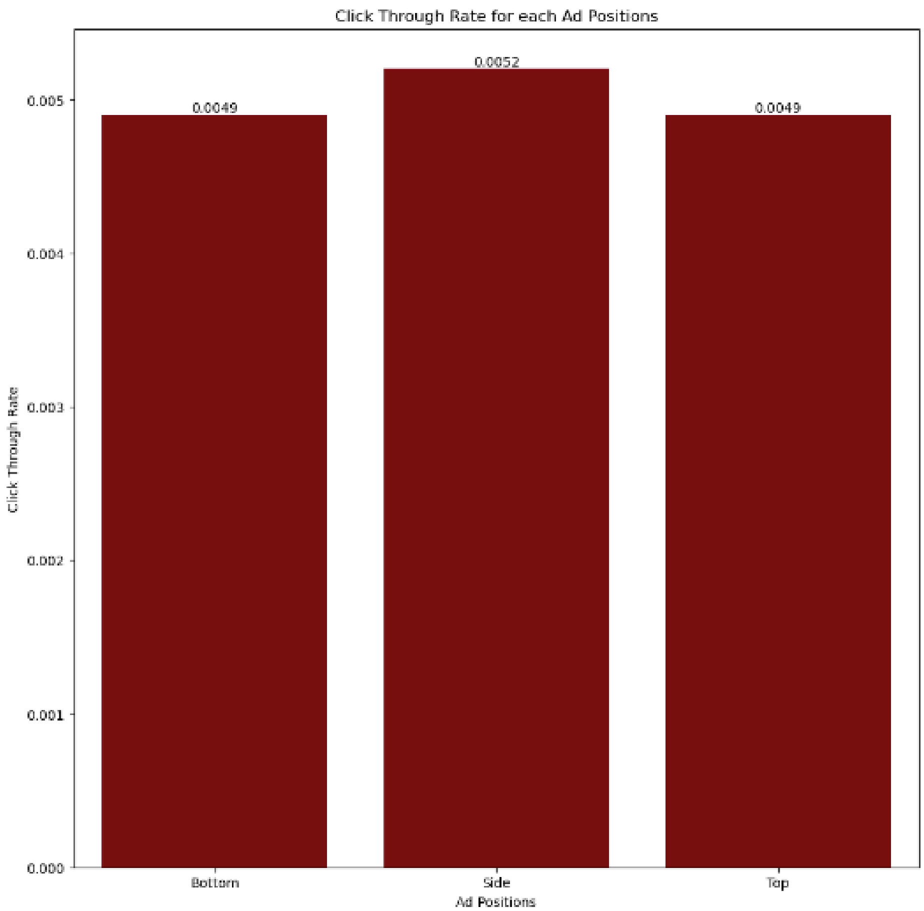
Goal: The goal of this analysis is to visualize the click-through rates (CTR) associated with different advertisement positions.

This analysis uses a bar plot to present the CTR for each ad position, making it easy to compare the effectiveness of various placements.The graph demonstrates a slight disparity in click-through rates (CTRs) across different ad positions. While the "Side" ad position exhibits a marginally higher CTR than the "Bottom" and "Top" positions, the overall variations are relatively minimal. This suggests that advertisers may be able to effectively utilize all ad positions without significant impact on campaign performance.

Code Snippet:

```
#CTR by Ad Positions
```

```
plt.subplot(2,2,2)
CTR_position_plot=sns.barplot(data=ctr_position, x='ad_position',y='CTR',
color='darkred')
CTR_position_plot.bar_label(CTR_position_plot.containers[0])
plt.xlabel('Ad Positions')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Ad Positions")
```



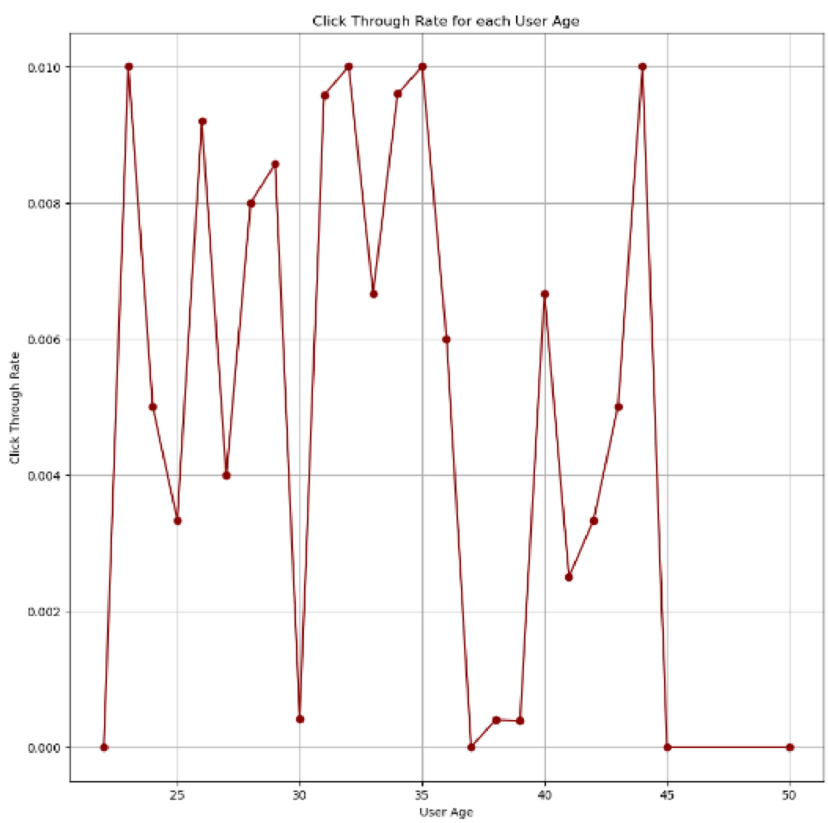
Analysis of Click-Through Rates by User Age

Goal: The goal of this analysis is to visualize the click-through rates (CTR) across different user age groups.

The graph demonstrates a lack of a consistent pattern in click-through rates (CTRs) across different user age groups. While some age groups, such as 25 and 45, exhibit significantly higher CTRs, others, like 30 and 40, demonstrate lower rates. This suggests that advertisers may benefit from tailoring their campaigns to specific age demographics to maximize engagement and improve overall performance.

Code Snippet:

```
#CTR by age
plt.subplot(2,2,3)
CTR_age_plot=plt.plot(ctr_age['user_age'],ctr_age['CTR'],       marker='o',
color='darkred')
plt.xlabel('User Age')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each User Age")
plt.grid(True)
```



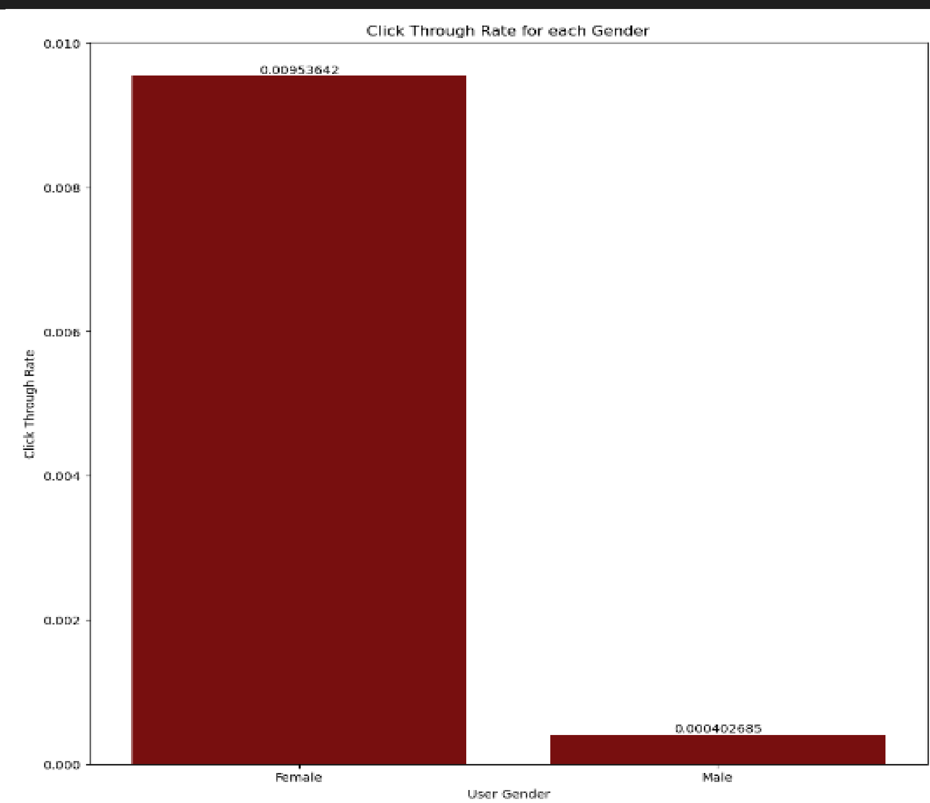
Analysis of Click-Through Rates by Gender

Goal: The goal of this analysis is to visualize the click-through rates (CTR) for different user genders.

The graph demonstrates a significant disparity in click-through rates (CTRs) between genders. Females exhibit a markedly higher CTR of approximately 0.0095 compared to males, whose CTR is only 0.0004. This substantial gender gap suggests that targeting females with advertisements could be a more effective strategy for generating clicks and improving campaign performance.

Code Snippet:

```
#CTR By gender
plt.subplot(2,2,4)
CTR_gender_plot=sns.barplot(data=ctr_gender, x='user_gender',y='CTR',
color='darkred')
CTR_gender_plot.bar_label(CTR_gender_plot.containers[0])
plt.xlabel('User Gender')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Gender")
```



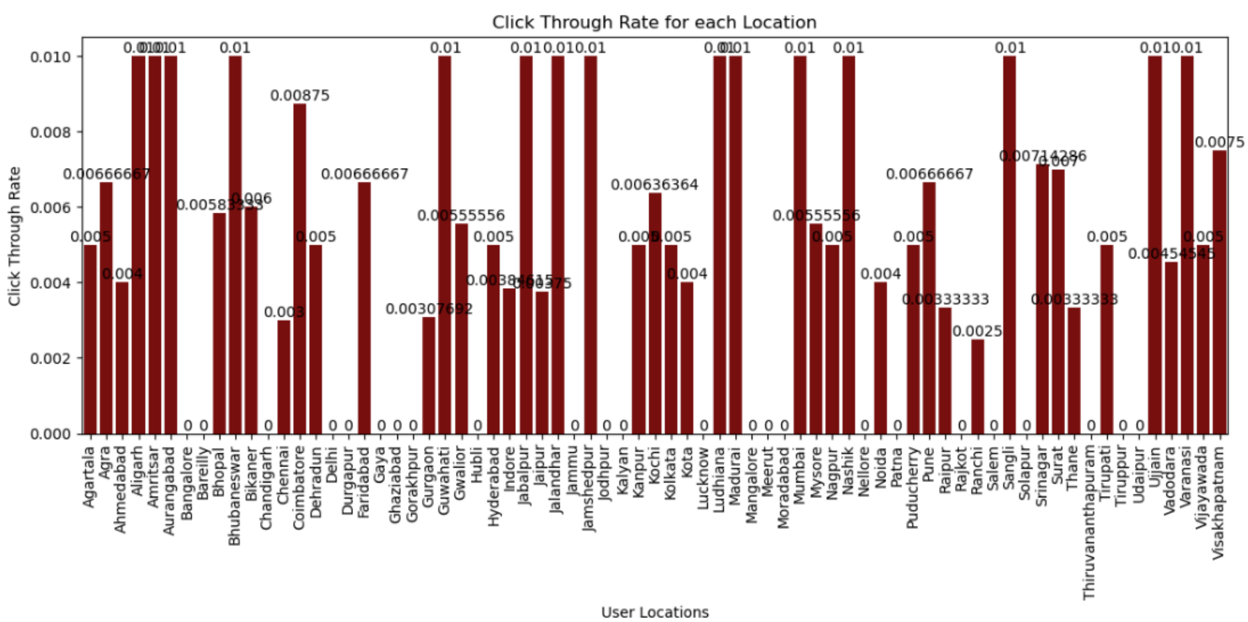
Analysis of Click-Through Rates by User Location

Goal: The goal of this analysis is to visualize the click-through rates (CTR) across different user locations.

This analysis uses a bar plot to represent the CTR for users across various locations. Each bar shows the CTR for a particular geographic region, offering a direct comparison of how different areas respond to advertisements. The plot reveals variations in engagement, where certain locations exhibit higher CTR values, indicating that users in those areas are more likely to click on ads. High-CTR locations could be prioritized for more focused ad campaigns due to their potential for higher conversion rates. Conversely, regions with lower CTRs might suggest the need for adapting the messaging or format of ads to better appeal to users in those areas.

Code Snippet:

```
#CTR by location

plt.figure(figsize=(12,6))
CTR_location_plot=sns.barplot(data=ctr_location,
x='user_location',y='CTR', color='darkred')
CTR_location_plot.bar_label(CTR_location_plot.containers[0])
plt.xlabel('User Locations')
plt.ylabel('Click Through Rate')
plt.title("Click Through Rate for each Location")
plt.xticks(rotation=90)
```



Model Implementation:

Regression:

Regression is a type of analysis used to understand the relationship between variables. In simple terms, it helps predict a certain value (like sales, temperature, or costs) based on other related data. It comes under **Supervised Machine Learning.**

For example, if you want to predict the cost of a car based on its age and mileage, regression would allow you to find a pattern and make accurate predictions.

In essence, regression helps us estimate or forecast a continuous outcome (like a number) based on other known data points.

Logistic Regression:

Logistic Regression is a statistical method used for binary classification problems, where the target variable has two classes (e.g., 0 and 1). It predicts the probability that a given input belongs to a particular class. Instead of directly predicting the outcome, logistic regression predicts the probability of the outcome, which is then mapped to a binary result using a threshold (usually 0.5).

The model uses the logistic function (also known as the sigmoid function) to convert the linear combination of input features into a probability.

In our project the logistic regression focuses on Ad Suggestion.

```
df=df.merge(ctr_category[['ad_category','CTR']],          how='left',
on='ad_category', suffixes=('','_category'))

df=df.merge(ctr_location[['user_location','CTR']],         how='left',
on='user_location', suffixes=('','_location'))

df=df.merge(ctr_position[['ad_position','CTR']],          how='left',
on='ad_position', suffixes=('','_position'))

df=df.merge(ctr_age[['user_age','CTR']],    how='left',    on='user_age',
suffixes=('','_age'))

df=df.merge(ctr_gender[['user_gender','CTR']],
how='left',on='user_gender', suffixes=('','_gender'))
```

In the above given code snippet we have merged the Click-Through Rate to our main dataframe.

```
df_encoded=pd.get_dummies(df[['user_gender','ad_category','user_location',
'ad_position']],drop_first=True)
```

Over here we have made use of one-hot encoding technique to assign it with binary values.

```
[118]: accuracy = accuracy_score(y_test, y_pred)
       print(f'Accuracy: {accuracy * 100:.2f}%')

       Accuracy: 95.56%

[120]: print(classification_report(y_test, y_pred))

                     precision    recall  f1-score   support

                  0       0.95      0.95      0.95        44
                  1       0.96      0.96      0.96        46

           accuracy                           0.96        90
          macro avg       0.96      0.96      0.96        90
       weighted avg       0.96      0.96      0.96        90
```

Based on the performance metrics of our model, we observe strong results across the board. The precision, recall, and F1-score for both classes (0 and 1) are consistently high, around 95-96%, indicating that the model is effective in correctly identifying both clicks and non-clicks. This balance between precision and recall suggests the model is not only making accurate predictions but also capturing the true positive cases without introducing many false positives or negatives. Additionally, the overall accuracy of 96% is well-supported by the macro and weighted averages, further confirming the robustness of the model.

In conclusion, the model shows strong predictive capabilities, with balanced performance across key metrics, making it a reliable tool for ad click-through rate prediction.