

IMPORTS

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [2]: dataset=pd.read_csv("ad_click_dataset.csv")
dataset
```

```
Out[2]:
```

	id	full_name	age	gender	device_type	ad_position	browsing_history	time_of_day	click
0	670	User670	22.0	NaN	Desktop	Top	Shopping	Afternoon	1
1	3044	User3044	NaN	Male	Desktop	Top	NaN	NaN	1
2	5912	User5912	41.0	Non-Binary	NaN	Side	Education	Night	1
3	5418	User5418	34.0	Male	NaN	NaN	Entertainment	Evening	1
4	9452	User9452	39.0	Non-Binary	NaN	NaN	Social Media	Morning	0
...
9995	8510	User8510	NaN	NaN	Mobile	Top	Education	NaN	0
9996	7843	User7843	NaN	Female	Desktop	Bottom	Entertainment	NaN	0
9997	3914	User3914	NaN	Male	Mobile	Side	NaN	Morning	0
9998	7924	User7924	NaN	NaN	Desktop	NaN	Shopping	Morning	1
9999	3056	User3056	44.0	Male	Tablet	Top	Social Media	Morning	0

10000 rows × 9 columns

Data Visualization

```
In [3]: dataset.head()
```

```
Out[3]:
```

	id	full_name	age	gender	device_type	ad_position	browsing_history	time_of_day	click
0	670	User670	22.0	NaN	Desktop	Top	Shopping	Afternoon	1
1	3044	User3044	NaN	Male	Desktop	Top	NaN	NaN	1
2	5912	User5912	41.0	Non-Binary	NaN	Side	Education	Night	1
3	5418	User5418	34.0	Male	NaN	NaN	Entertainment	Evening	1
4	9452	User9452	39.0	Non-Binary	NaN	NaN	Social Media	Morning	0

Data Analytics

```
In [4]: dataset.describe()
```

```
Out[4]:
```

	id	age	click
count	10000.000000	5234.000000	10000.000000
mean	5060.211400	40.197363	0.650000
std	2861.758265	13.126420	0.476993
min	5.000000	18.000000	0.000000
25%	2529.000000	29.000000	0.000000
50%	5218.000000	39.500000	1.000000
75%	7466.000000	52.000000	1.000000
max	10000.000000	64.000000	1.000000

```
In [5]: dataset.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  --
 0   id            10000 non-null  int64  
 1   full_name     10000 non-null  object  
 2   age           5234 non-null   float64 
 3   gender        5307 non-null   object  
 4   device_type   8000 non-null   object  
 5   ad_position   8000 non-null   object  
 6   browsing_history 5218 non-null   object  
 7   time_of_day   8000 non-null   object  
 8   click         10000 non-null  int64  
dtypes: float64(1), int64(2), object(6)
memory usage: 703.3+ KB
```

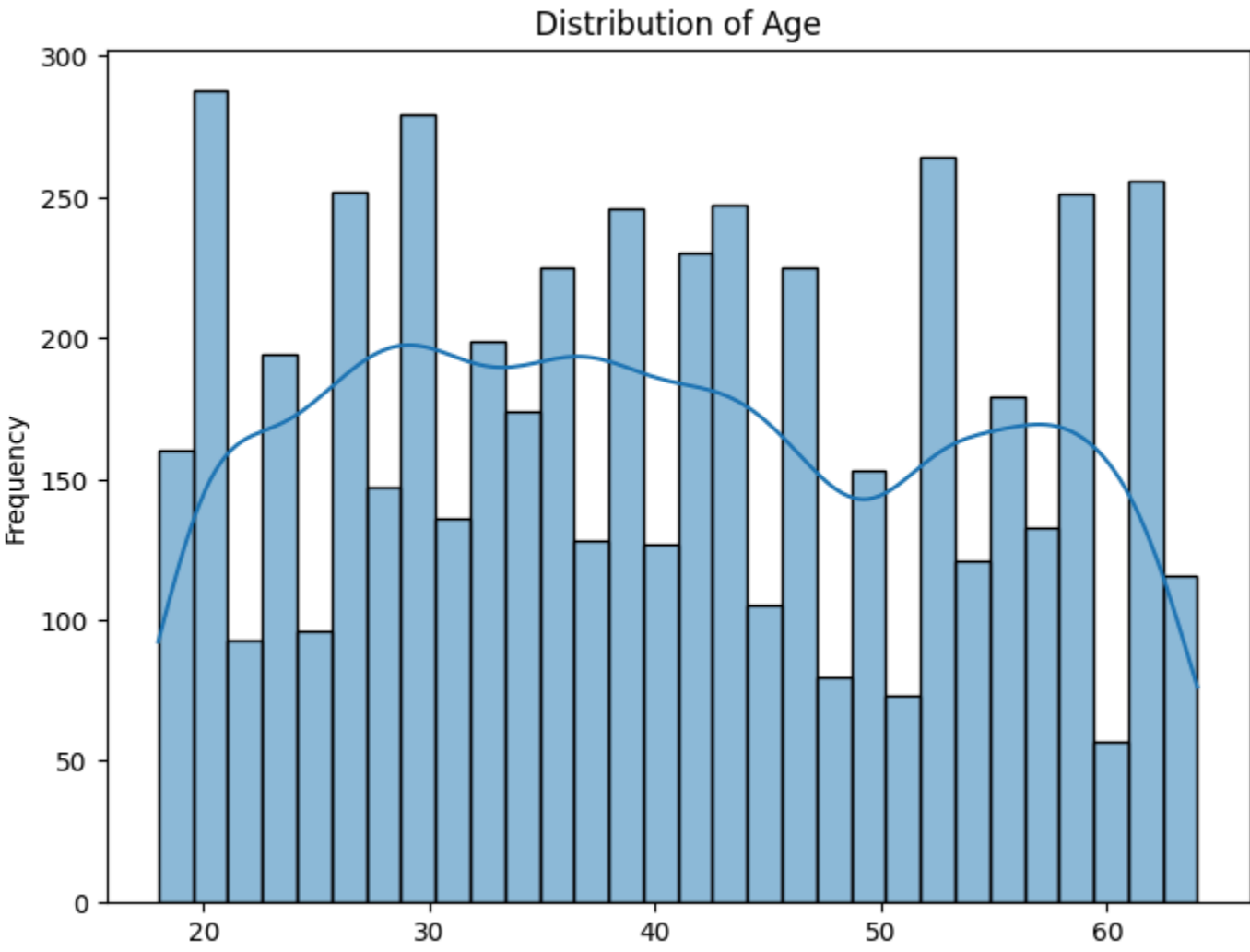
```
In [6]: dataset['id'].value_counts()
dataset['age'].value_counts()
dataset['gender'].value_counts()
dataset['ad_position'].value_counts()
dataset['browsing_history'].value_counts()
dataset['time_of_day'].value_counts()
```

```
Out[6]: time_of_day
Morning    2126
Afternoon  2016
Evening    1958
Night      1900
Name: count, dtype: int64
```

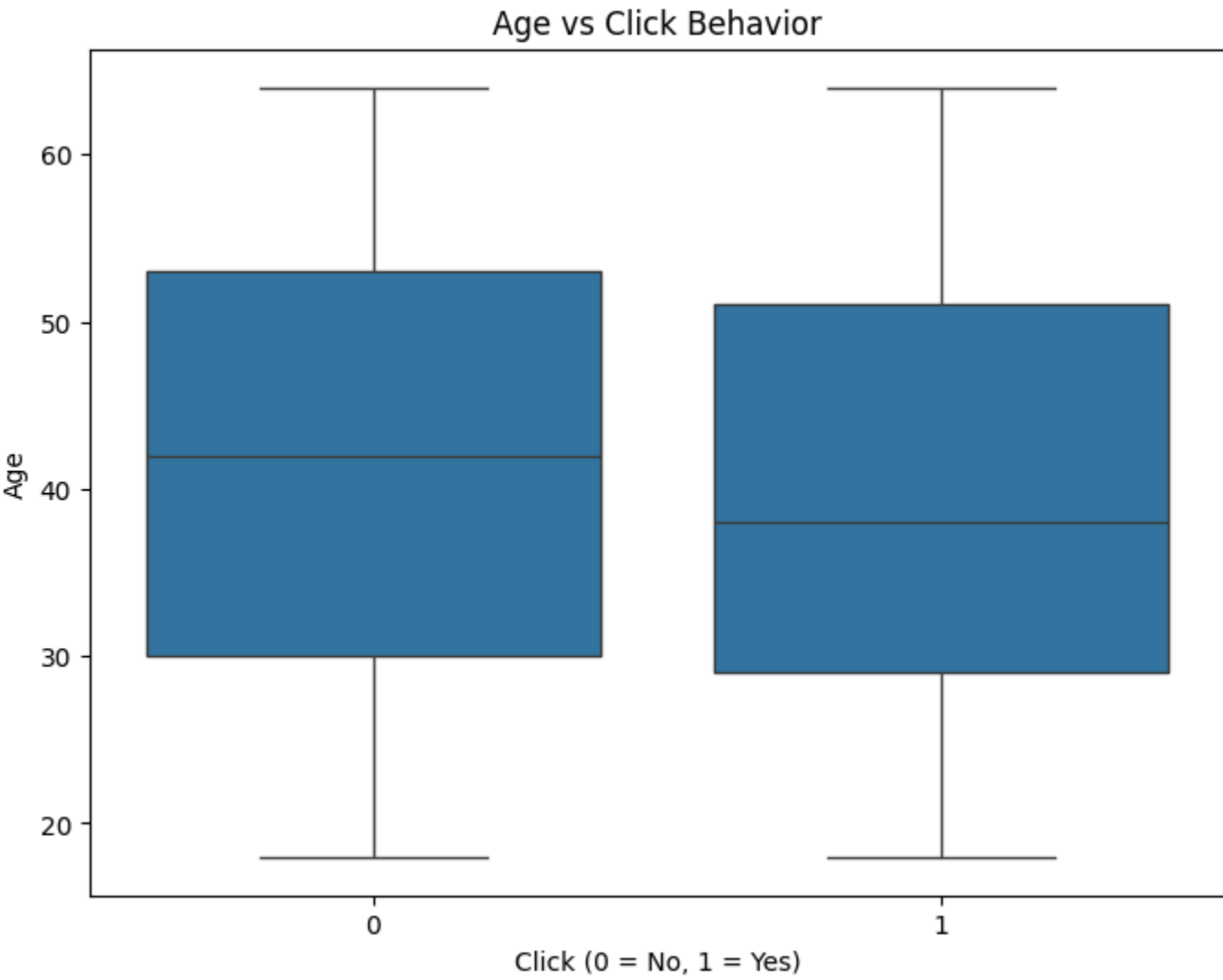
```
In [7]: dataset['click'].value_counts(normalize=True)
```

```
Out[7]: click
1    0.65
0    0.35
Name: proportion, dtype: float64
```

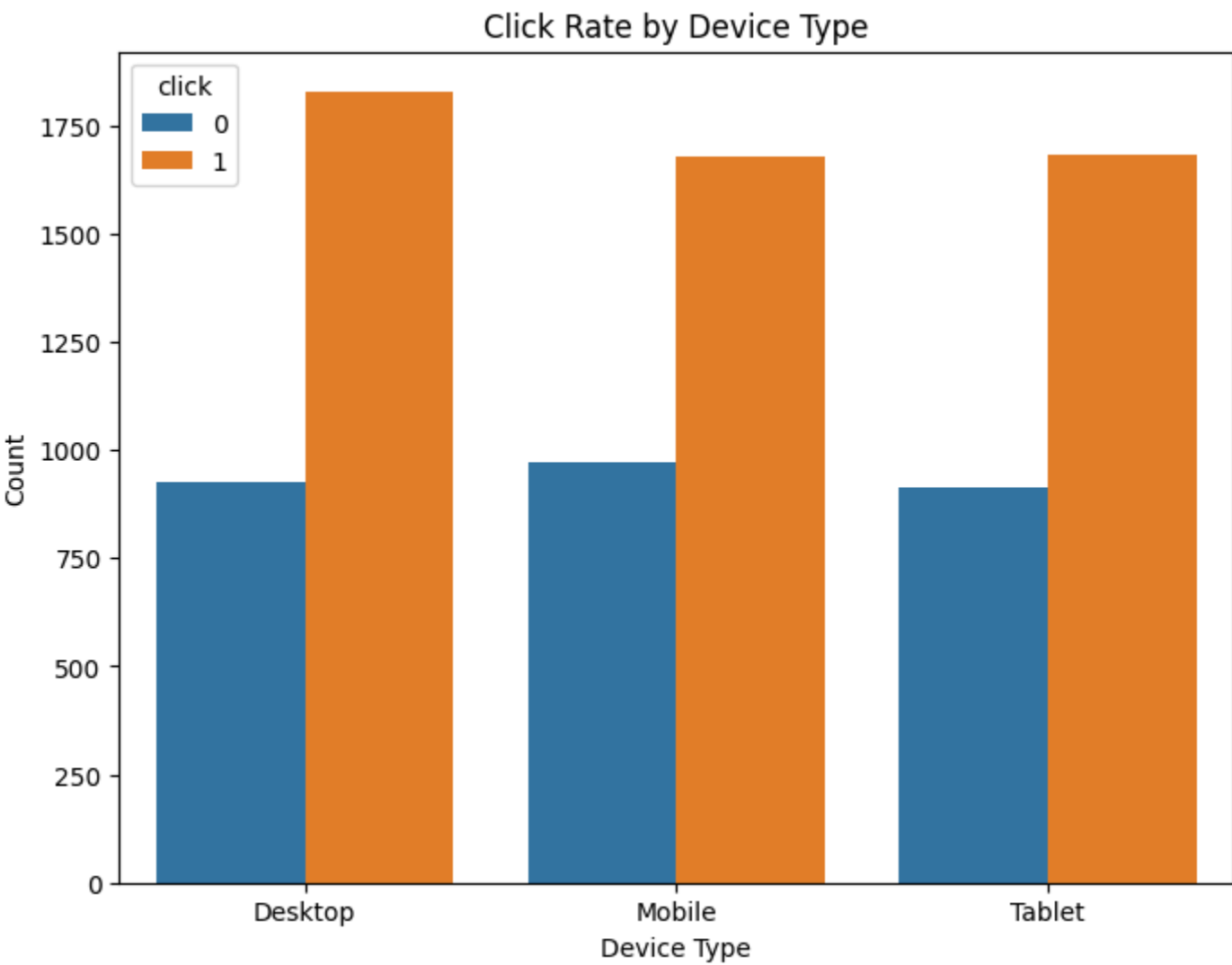
```
In [8]: plt.figure(figsize=(8,6))
sns.histplot(dataset['age'], bins=30, kde=True)
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



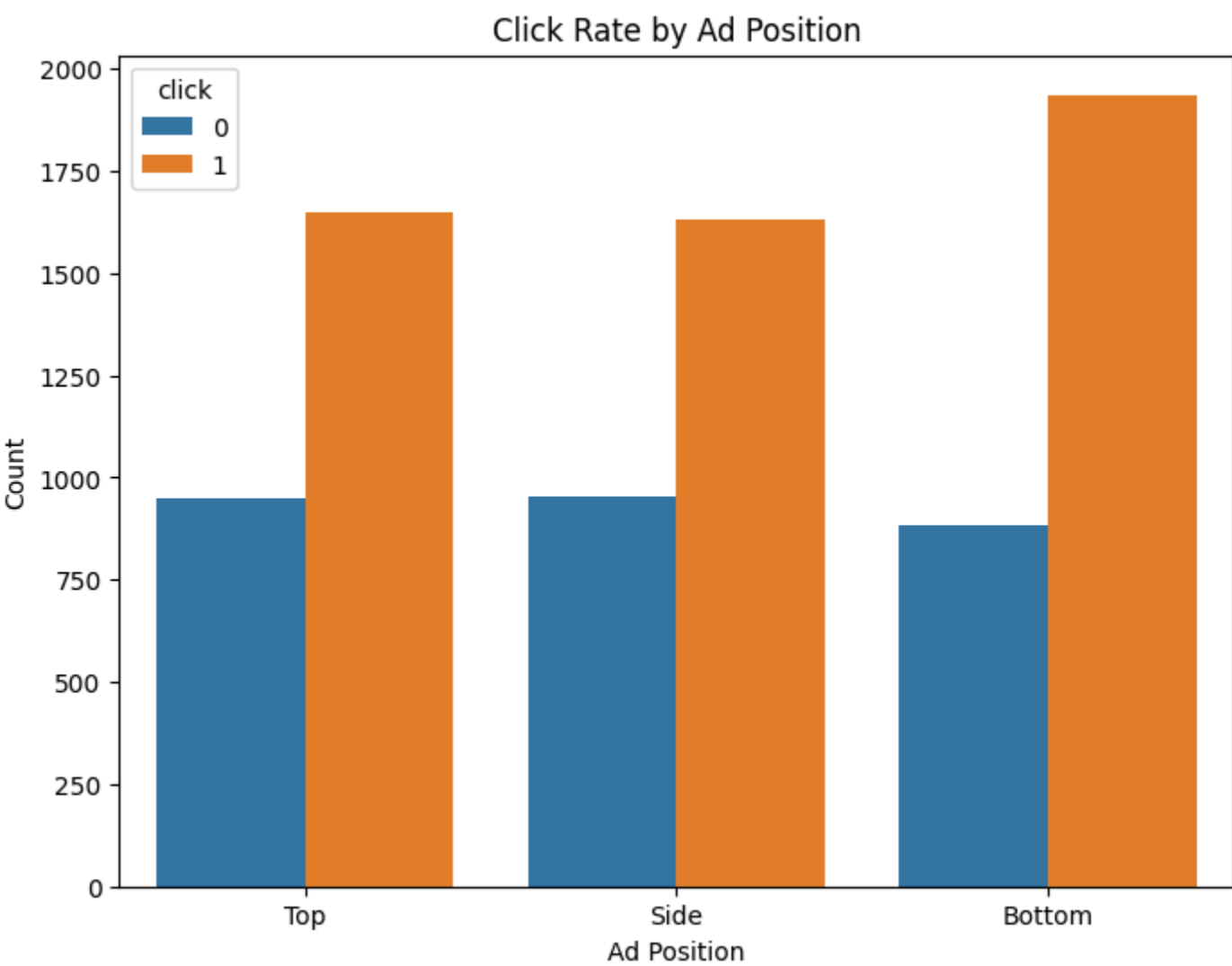
```
In [9]: plt.figure(figsize=(8,6))
sns.boxplot(x='click', y='age', data= dataset)
plt.title('Age vs Click Behavior')
plt.xlabel('Click (0 = No, 1 = Yes)')
plt.ylabel('Age')
plt.show()
```



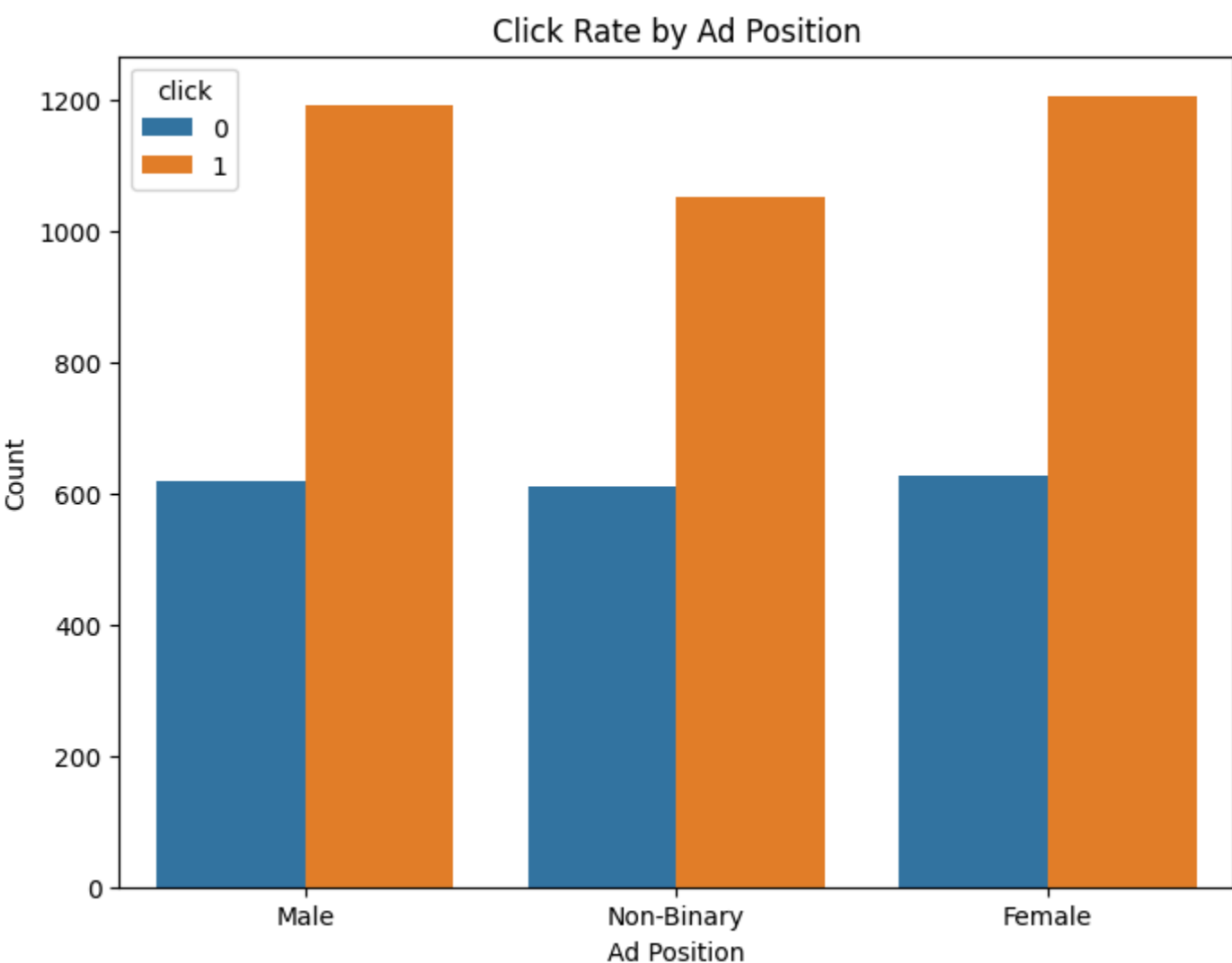
```
In [10]: plt.figure(figsize=(8,6))
sns.countplot(x='device_type', hue='click', data=dataset)
plt.title('Click Rate by Device Type')
plt.xlabel('Device Type')
plt.ylabel('Count')
plt.show()
```



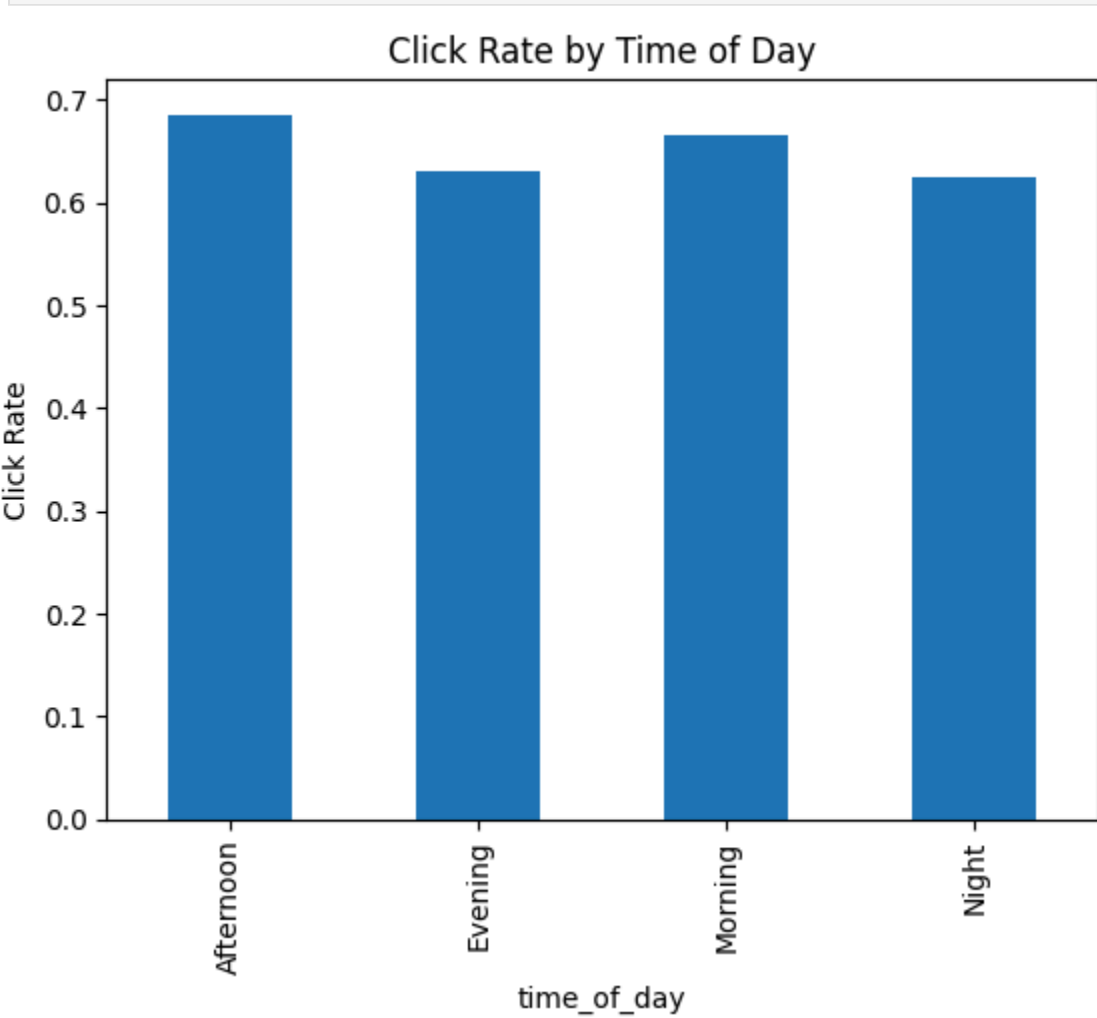
```
In [11]: plt.figure(figsize=(8,6))
sns.countplot(x='ad_position', hue='click', data=dataset)
plt.title('Click Rate by Ad Position')
plt.xlabel('Ad Position')
plt.ylabel('Count')
plt.show()
```



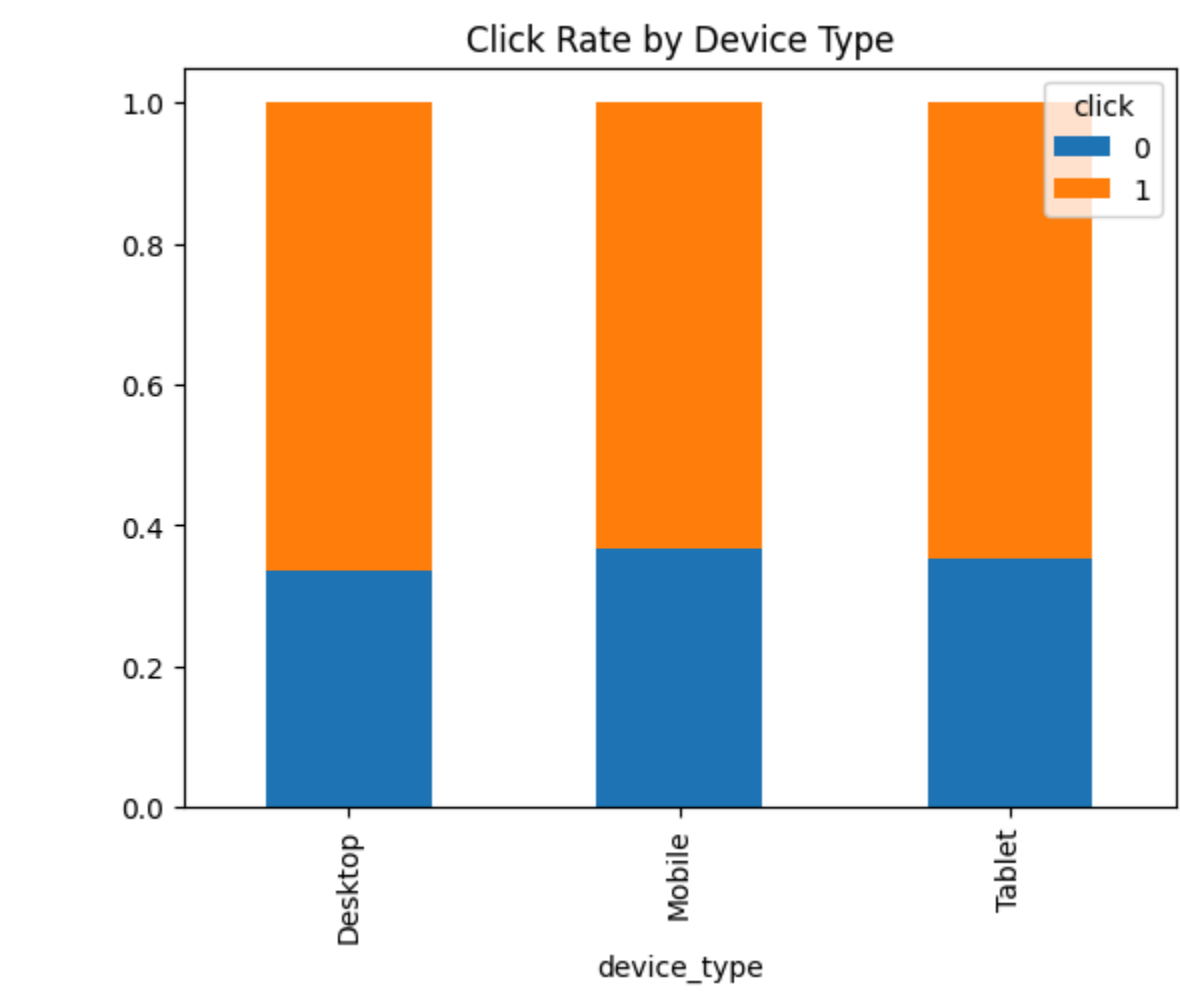
```
In [12]: plt.figure(figsize=(8,6))
sns.countplot(x='gender', hue='click', data=dataset)
plt.title('Click Rate by Ad Position')
plt.xlabel('Ad Position')
plt.ylabel('Count')
plt.show()
```



```
In [14]: dataset.groupby('time_of_day')['click'].mean().plot(kind='bar')
plt.title('Click Rate by Time of Day')
plt.ylabel('Click Rate')
plt.show()
```



```
In [13]: pd.crosstab(dataset['device_type'], dataset['click'], normalize='index').plot(kind='bar', stacked=True)
plt.title('Click Rate by Device Type')
plt.show()
```

in []: