17LIVE Data Analyst Assignment

Objective:

- Identify optimal ad placement times
- Evaluate the ad campaign performance

```
In [1]: # Step 1: Import Libraries and Load Data
    import pandas as pd
    import numpy as np
    import matplotlib.pyplot as plt
    import seaborn as sns
    from datetime import datetime
    from scipy.stats import ttest_ind, mannwhitneyu

sns.set(style="whitegrid")
    plt.rcParams['figure.figsize'] = (12, 6)

df = pd.read_csv("all_stream_info.csv")
    df.head()
```

Out[1]:		liveStreamID	beginTime	endTime	duration	closeBy	maxLiveViewerCount	maxLive\
	0	109437538	2020-06- 22 11:55:21 UTC	22	16918	normalEnd	363	16:2
	1	109441785	2020-06- 22 14:55:26 UTC	2020-06- 22 21:31:19 UTC	23753	normalEnd	100	19:07
	2	109438205	2020-06- 22 12:20:34 UTC	2020-06- 22 16:02:46 UTC	13332	disconnect	471	14:53
	3	109438917	2020-06- 22 12:54:21 UTC	2020-06- 22 14:47:27 UTC	6786	normalEnd	44	14:29
	4	109442185	2020-06- 22 15:18:20 UTC	2020-06- 22 15:48:02 UTC	1782	normalEnd	52	15:42

5 rows × 29 columns

```
In [2]: # Step 2: Initial Data Exploration
    df.info()
    print(df.isnull().sum())
    df.describe()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3148 entries, 0 to 3147
Data columns (total 29 columns):

		1113).		
#	Column	Non-N	Null Count	Dtype
0	liveStreamID	3148	non-null	int64
1	beginTime	3148	non-null	object
2	endTime	3148	non-null	object
3	duration	3148	non-null	int64
4	closeBy	3051	non-null	object
5	maxLiveViewerCount	3148	non-null	int64
6	maxLiveViewerTime	1838	non-null	object
7	privateLiveStream	3148	non-null	int64
8	receivedLikeCount	3148	non-null	int64
9	streamerType		non-null	int64
10	isShow		non-null	bool
11	cultureGroup		n-null	float64
12	userID		non-null	object
13	registerTime		non-null	object
	registerCountry		non-null	object
14 15	•			•
15	isContracted		n-null	float64
16	uniqueViewerCount		non-null	int64
17	ios		non-null	int64
18	android		non-null	int64
19	durationGTE5sec		non-null	int64
20	durationGTE2min		non-null	int64
21	durationGTE10min	3148	non-null	int64
22	totalViewerDuration	1689	non-null	float64
23	avgViewerDuration	1689	non-null	float64
24	avgStreamJoinDuration	1689	non-null	float64
25	count	3148	non-null	int64
26	followIncreaseEstimate	d 3148	non-null	int64
27	receivePointEstimated	3148	non-null	int64
28	dau	0 nor	n-null	float64
	es: bool(1), float64(6)		(15), object	(7)
	ry usage: 691.8+ KB		. ,, ,	` '
	-	0		
liveS	itreamID	0 0		
liveS begir	treamID Time	0		
liveS begin endTi	treamID Time .me	0 0		
liveS begir endTi durat	treamID Time .me :ion	0 0 0		
liveS begir endTi durat close	treamID Time me ion By	0 0 0 97		
liveS begin endTi durat close maxLi	treamID Time me tion BBy veViewerCount	0 0 0 97 0		
liveS begir endTi durat close maxLi maxLi	treamID Time .me tion By veViewerCount veViewerTime	0 0 0 97 0 1310		
liveS begin endTi durat close maxLi maxLi priva	treamID Time me tion By veViewerCount veViewerTime tteLiveStream	0 0 97 0 1310		
liveS begir endTi durat close maxLi maxLi priva recei	treamID Time me tion By veViewerCount veViewerTime teLiveStream vedLikeCount	0 0 97 0 1310 0		
liveS begir endTi durat close maxLi maxLi priva recei strea	streamID Time me tion By veViewerCount veViewerTime teLiveStream vedLikeCount merType	0 0 97 0 1310 0 0		
liveS begir endTi durat close maxLi maxLi priva recei strea isSho	streamID Time me ion By veViewerCount veViewerTime ateLiveStream vedLikeCount merType	0 0 97 0 1310 0 0		
liveS begir endTi durat close maxLi maxLi priva recei strea isSho	streamID Time me ion By veViewerCount veViewerTime ateLiveStream vedLikeCount merType	0 0 97 0 1310 0 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI	streamID Time me cion By veViewerCount veViewerTime cteLiveStream vedLikeCount merType ow ireGroup	0 0 97 0 1310 0 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI	streamID ITime .me .ion By .veViewerCount .veViewerTime .teLiveStream .vedLikeCount .merType .we	0 0 97 0 1310 0 0 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis	streamID Time me cion By veViewerCount veViewerTime cteLiveStream vedLikeCount merType ow ireGroup	0 0 97 0 1310 0 0 0 3148		
lives begir endTi durat close maxLi priva recei strea isSho cultu userI regis regis	treamID Time me tion By veViewerCount veViewerTime teLiveStream vedLikeCount merType bw preGroup ID terTime terCountry	0 0 97 0 1310 0 0 0 3148 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor	streamID ITime Ime Ime Ition IBy IveViewerCount IveViewerTime IteLiveStream IvedLikeCount ImerType ImeGroup ID IterTime IterCountry	0 0 97 0 1310 0 0 0 3148 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor	treamID Time .me .ion By .veViewerCount .veViewerTime .teLiveStream .vedLikeCount .merType .w .ireGroup .b .terTime .terCountry .tracted	0 0 97 0 1310 0 0 0 3148 0 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor uniqu	streamID ITime Ime Incompany Incompany	0 0 97 0 1310 0 0 0 3148 0 0 3148		
lives begir endTi durat close maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro	streamID ITime Ime Incompany Incompany	0 0 97 0 1310 0 0 0 3148 0 0 3148		
lives begir endTi durat close maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro durat	streamID ITime Ime Ime Ition IBy IveViewerCount IveViewerTime IveLiveStream IvedLikeCount ImerType ID IterTime IterCountry Itracted IeViewerCount	0 0 97 0 1310 0 0 0 3148 0 0 3148 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro durat durat	streamID ITime Ime Ime Ition By IveViewerCount IveViewerTime IveLiveStream IvedLikeCount ImerType IveGroup ID IterTime IterCountry Itracted IveViewerCount Ind Id Id Id Id Id Id Id Id I	0 0 97 0 1310 0 0 0 3148 0 0 0 3148		
lives begir endTi durat close maxLi priva recei strea isSho cultu userI regis isCor uniqu ios andro durat durat	streamID Time .me .ion By .veViewerCount .veViewerTime .teLiveStream .vedLikeCount .merType .me .terTime .terTime .terCountry .tracted .eViewerCount .ionGTE5sec .ionGTE5min .ionGTE10min	0 0 97 0 1310 0 0 0 3148 0 0 3148 0 0 0		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis isCor uniqu ios andro durat durat total	streamID Time .me .ion By .veViewerCount .veViewerTime .teLiveStream .vedLikeCount .merType .meGroup .D .terTime .terCountry .tracted .eViewerCount .ionGTE5sec .ionGTE2min .ionGTE10min .ViewerDuration	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 0 1459		
lives begir endTi durat close maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro durat durat total avgVi	streamID Time .me .ion By .veViewerCount .veViewerTime .vedLikeCount .merType .w .reGroup .D .terTime .terCountry .tracted .eViewerCount .id .iionGTE5sec .ionGTE2min .ionGTE10min .ViewerDuration .ewerDuration	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 1459 1459		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro durat durat total avgVi avgSt	streamID ITime Ime Ime Ion By IveViewerCount IveViewerTime IveLiveStream IvedLikeCount ImerType Ive IneGroup ID IterTime IterCountry Itracted IeViewerCount Ind IionGTE5sec IionGTE2min IionGTE10min IviewerDuration IveerDuration IreamJoinDuration	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 1459 1459		
lives begin endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis isCor uniqu ios andro durat durat total avgVi avgSt count	streamID ITime Ime Ime Ition By IveViewerCount IveViewerTime IteLiveStream IvedLikeCount ImerType ImerGroup ID IterTime IterCountry Itracted IeViewerCount Id ItionGTE5sec IonGTE2min IonGTE10min ViewerDuration IvewerDuration IreamJoinDuration Iteramy Iteramy Incomplete Incomplete Incomplete Incomplete Incomplete	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 1459 1459 1459		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis isCor uniqu ios andro durat durat total avgVi avgSt count follo	streamID Time .me .ion By .veViewerCount .veViewerTime .vedLikeCount .merType .me .reGroup .D .terTime .terCountry .tracted .eviewerCount .eviewerDuration .reamJoinDuration .reamJoinDuration .cowIncreaseEstimated	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 0 1459 1459 1459		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro durat durat total avgVi avgSt count follo	streamID Time .me .ion By .veViewerCount .veViewerTime .vedLikeCount .merType .w .reGroup .D .terTime .terCountry .tracted .eViewerCount .id .iionGTE5sec .ionGTE2min .ionGTE10min .ViewerDuration .ewerDuration .reamJoinDuration .reamJoinDuration .cowIncreaseEstimated .vePointEstimated	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 1459 1459 1459 1459		
lives begir endTi durat close maxLi maxLi priva recei strea isSho cultu userI regis regis isCor uniqu ios andro durat durat total avgVi avgSt count follo recei dau	streamID Time .me .ion By .veViewerCount .veViewerTime .vedLikeCount .merType .w .reGroup .D .terTime .terCountry .tracted .eViewerCount .id .iionGTE5sec .ionGTE2min .ionGTE10min .ViewerDuration .ewerDuration .reamJoinDuration .reamJoinDuration .cowIncreaseEstimated .vePointEstimated	0 0 97 0 1310 0 0 0 3148 0 0 0 3148 0 0 0 0 1459 1459 1459		

receivedLikeC	privateLiveStream	maxLiveViewerCount	duration	liveStreamID	
3148.00	3148.0	3148.000000	3148.000000	3.148000e+03	count
358.42	0.0	15.416773	3495.864041	1.093797e+08	mean
2145.57	0.0	45.505183	6651.916426	3.381310e+04	std
0.00	0.0	0.000000	0.000000	1.093211e+08	min
0.00	0.0	0.000000	136.750000	1.093518e+08	25%
0.00	0.0	2.000000	1032.500000	1.093835e+08	50%
62.00	0.0	8.000000	4801.500000	1.094082e+08	75%
65831.00	0.0	493.000000	121258.000000	1.094427e+08	max

8 rows × 21 columns

Out[2]:

In [4]: print(df.columns.tolist())

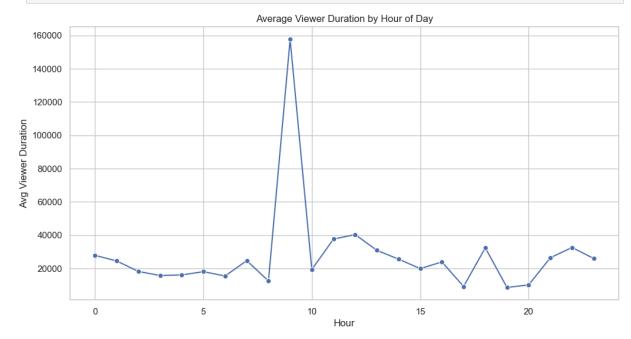
['liveStreamID', 'beginTime', 'endTime', 'duration', 'closeBy', 'maxLiveViewerCoun t', 'maxLiveViewerTime', 'privateLiveStream', 'receivedLikeCount', 'streamerType', 'isShow', 'cultureGroup', 'userID', 'registerTime', 'registerCountry', 'isContracte d', 'uniqueViewerCount', 'ios', 'android', 'durationGTE5sec', 'durationGTE2min', 'du rationGTE10min', 'totalViewerDuration', 'avgViewerDuration', 'avgStreamJoinDuratio n', 'count', 'followIncreaseEstimated', 'receivePointEstimated', 'dau']

```
In [5]: # Step 3: Time Features using 'beginTime'
df['beginTime'] = pd.to_datetime(df['beginTime'])

# Extract hour and day of week from beginTime
df['hour'] = df['beginTime'].dt.hour
df['dayofweek'] = df['beginTime'].dt.day_name()
```

```
In [6]: # Step 4: Optimal Ad Placement Time
hourly = df.groupby('hour').agg({
        'totalViewerDuration': 'mean',
        'userID': 'count'
}).rename(columns={'userID': 'streamCount'})

sns.lineplot(data=hourly, x=hourly.index, y='totalViewerDuration', marker='o')
plt.title('Average Viewer Duration by Hour of Day')
plt.xlabel('Hour')
plt.ylabel('Avg Viewer Duration')
plt.grid(True)
plt.show()
```



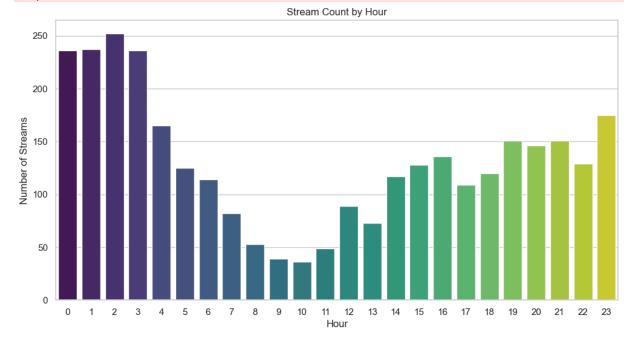
```
In [8]: sns.barplot(data=hourly.reset_index(), x='hour', y='streamCount', palette='viridis'
plt.title('Stream Count by Hour')
plt.xlabel('Hour')
```

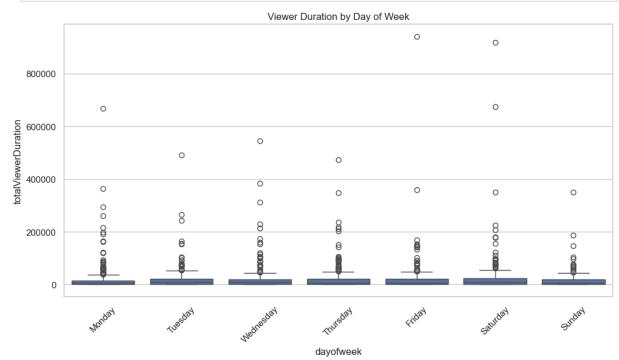
```
plt.ylabel('Number of Streams')
plt.show()
```

C:\Users\jeeve\AppData\Local\Temp\ipykernel_17368\775848521.py:1: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.1 4.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

sns.barplot(data=hourly.reset_index(), x='hour', y='streamCount', palette='viridi
s')



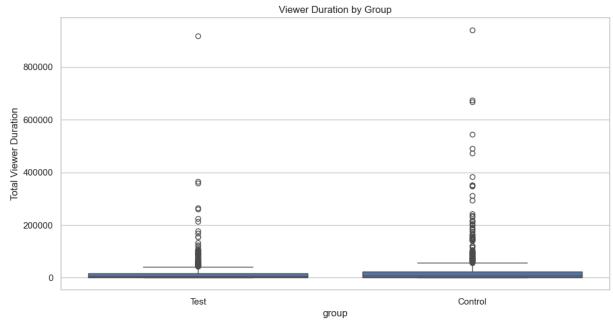


```
In [10]: # Step 5: Marketing Campaign Evaluation
    control_suffix = ['d', 'f', '1', '2', '3', '4', '5', '8']
    test_suffix = ['a', 'c', 'b', 'e', '0', '6', '7', '9']

def assign_group(user_id):
    if isinstance(user_id, str) and len(user_id) > 0:
        suffix = user_id[-1]
        return 'Control' if suffix in control_suffix else 'Test'
    return 'Unknown'

df['group'] = df['userID'].apply(assign_group)
    df = df[df['group'].isin(['Control', 'Test'])]
```

```
In [11]: sns.boxplot(x='group', y='totalViewerDuration', data=df)
  plt.title('Viewer Duration by Group')
  plt.ylabel('Total Viewer Duration')
  plt.show()
```



```
In [12]: control_dur = df[df['group'] == 'Control']['totalViewerDuration']
    test_dur = df[df['group'] == 'Test']['totalViewerDuration']

stat, p_value = mannwhitneyu(test_dur, control_dur)
    print(f"Mann-Whitney U Test: U={stat:.2f}, p-value={p_value:.4f}")
```

Mann-Whitney U Test: U=nan, p-value=nan

Final Report

- 1. Optimal Ad Time Slots:
- Use the hours with peak avg. viewer duration and high stream count.
- 2. Campaign Evaluation:
- p-value = nan
- Not significant difference observed.

```
In [ ]:
```