In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv('data-export (1).csv')
df.head()
```

Out[1]:

	# ---------- ------------- ------------- ----	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed
0	Session primary channel group (Default channel...	Date + hour (YYYYMMDDHH)	Users	Sessions	Engaged sessions	Average engagement time per session	Engaged sessio per us
1	Direct	2024041623	237	300	144	47.526666666666700	0.60759493670886
2	Organic Social	2024041719	208	267	132	32.09737827715360	0.63461538461538
3	Direct	2024041723	188	233	115	39.93991416309010	0.61170212765957
4	Organic Social	2024041718	187	256	125	32.16015625	0.66844919786096

Cleaning data for retrieve important insight from this

In [2]:

```
df.columns = df.iloc[0]
df = df.drop(index=0).reset_index(drop=True)
df.columns = ['channel group', 'Date', 'Users', 'Sessions', 'Engaged sessions', 'Average
```

In [3]:

```
df.head()
```

Out[3]:

	channel group	Date	Users	Sessions	Engaged sessions	Average engagement time per session	Engaged sessions per user	Event p
0	Direct	2024041623	237	300	144	47.526666666666700	0.6075949367088610	4.67333
1	Organic Social	2024041719	208	267	132	32.09737827715360	0.6346153846153850	4.29588
2	Direct	2024041723	188	233	115	39.93991416309010	0.6117021276595740	4.58798
3	Organic Social	2024041718	187	256	125	32.16015625	0.6684491978609630	
4	Organic Social	2024041720	175	221	112	46.918552036199100	0.64	4.52941

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3182 entries, 0 to 3181
Data columns (total 10 columns):
 #   Column                               Non-Null Count  Dtype 
---  ------                               --------------  ----- 
 0   channel group                        3182 non-null   object
 1   Date                                 3182 non-null   object
 2   Users                                3182 non-null   object
 3   Sessions                             3182 non-null   object
 4   Engaged sessions                     3182 non-null   object
 5   Average engagement time per session  3182 non-null   object
 6   Engaged sessions per user            3182 non-null   object
 7   Event per session                    3182 non-null   object
 8   Engagement rate                      3182 non-null   object
 9   Event count                          3182 non-null   object
dtypes: object(10)
memory usage: 248.7+ KB
```

Change the data types of the data

In [5]:

```
df['Date'] = pd.to_datetime(df['Date'], format='%Y%m%d%H', errors='coerce')
df['Hours'] = df['Date'].dt.hour
numeric_col = df.columns.drop('channel group', 'Date')
df[numeric_col] = df[numeric_col].apply(pd.to_numeric, errors='coerce')
df['Date'] = pd.to_datetime(df['Date'])
```

In [6]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3182 entries, 0 to 3181
Data columns (total 11 columns):
 #   Column                               Non-Null Count  Dtype         
---  ------                               --------------  -----         
 0   channel group                        3182 non-null   object        
 1   Date                                 3182 non-null   datetime64[ns]
 2   Users                                3182 non-null   int64         
 3   Sessions                             3182 non-null   int64         
 4   Engaged sessions                     3182 non-null   int64         
 5   Average engagement time per session  3182 non-null   float64       
 6   Engaged sessions per user            3182 non-null   float64       
 7   Event per session                    3182 non-null   float64       
 8   Engagement rate                      3182 non-null   float64       
 9   Event count                          3182 non-null   int64         
 10  Hours                                3182 non-null   int32         
dtypes: datetime64[ns](1), float64(4), int32(1), int64(4), object(1)
memory usage: 261.2+ KB
```

In [7]:

```
df.head()
```

Out[7]:

	channel group	Date	Users	Sessions	Engaged sessions	Average engagement time per session	Engaged sessions per user	Event per session	Engagement rate	Event count
0	Direct	2024-04-16 23:00:00	237	300	144	47.526667	0.607595	4.673333	0.480000	1402
1	Organic Social	2024-04-17 19:00:00	208	267	132	32.097378	0.634615	4.295880	0.494382	1147
2	Direct	2024-04-17 23:00:00	188	233	115	39.939914	0.611702	4.587983	0.493562	1069
3	Organic Social	2024-04-17 18:00:00	187	256	125	32.160156	0.668449	4.078125	0.488281	1044
4	Organic Social	2024-04-17 20:00:00	175	221	112	46.918552	0.640000	4.529412	0.506787	1001

In [8]:

```
df.describe()
```

Out[8]:

	Date	Users	Sessions	Engaged sessions	Average engagement time per session	Engaged sessions per user	Event p sessi
count	3182	3182.000000	3182.000000	3182.000000	3182.000000	3182.000000	3182.0000
mean	2024-04-20 01:17:07.278441216	41.935889	51.192646	28.325581	66.644581	0.606450	4.6759
min	2024-04-06 00:00:00	0.000000	1.000000	0.000000	0.000000	0.000000	1.0000
25%	2024-04-13 02:15:00	20.000000	24.000000	13.000000	32.103034	0.561404	3.7500
50%	2024-04-20 02:00:00	42.000000	51.000000	27.000000	49.020202	0.666667	4.4102
75%	2024-04-26 22:00:00	60.000000	71.000000	41.000000	71.487069	0.750000	5.2176
max	2024-05-03 23:00:00	237.000000	300.000000	144.000000	4525.000000	2.000000	56.0000
std	NaN	29.582258	36.919962	20.650569	127.200659	0.264023	2.7952

In [9]:

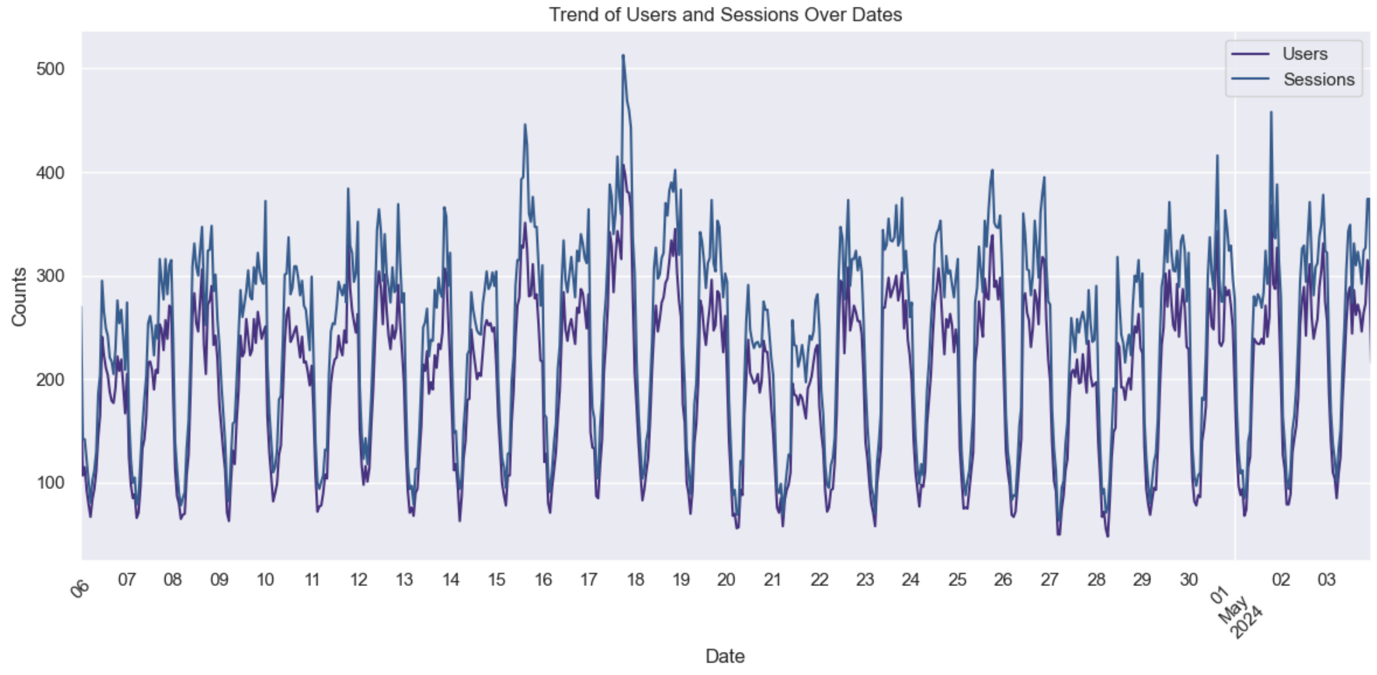
```
sns.set(style='whitegrid')
sns.set(palette='viridis')
```

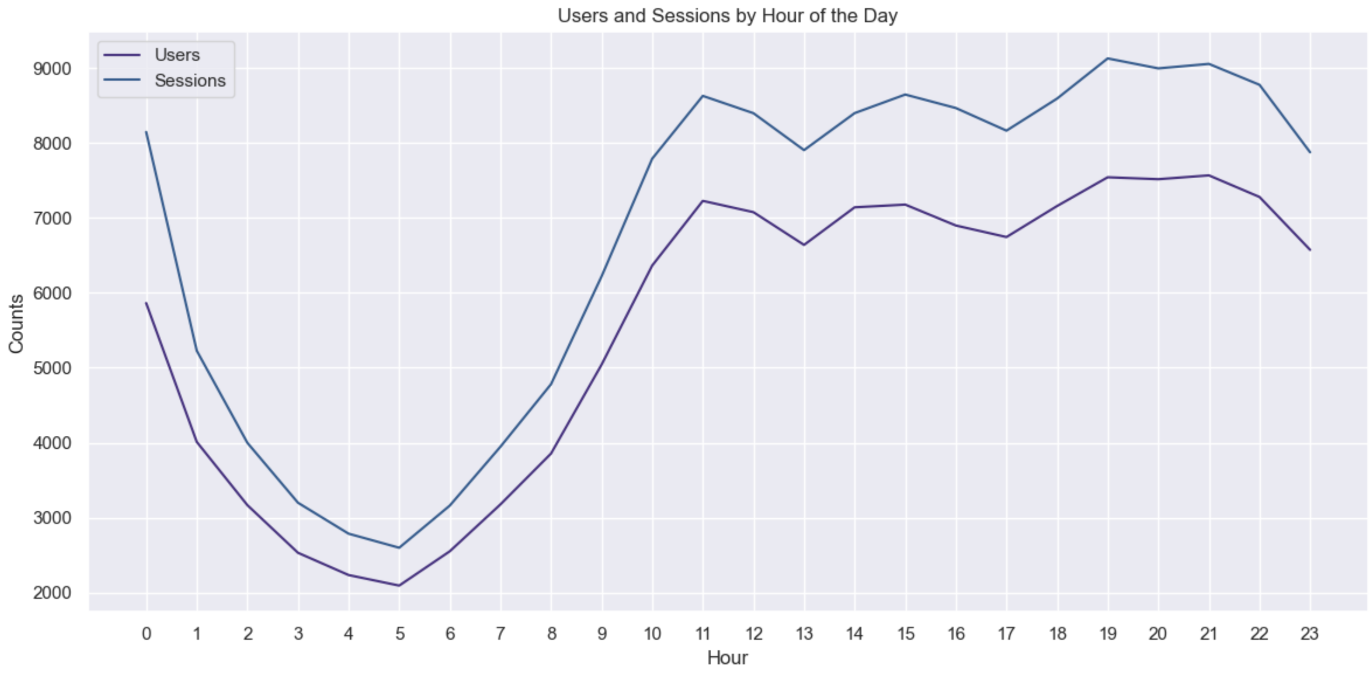
What patterns or trends can you observe in website sessions and users over time?

In [19]:

```
# pattern/Trend find on website sessions and user over time
# Plot 1: Users and Sessions by Dates
plt.figure(figsize=(12, 6))
df.groupby('Date')[['Users', 'Sessions']].sum().plot(ax=plt.gca())
plt.title('Trend of Users and Sessions Over Dates')
plt.xlabel('Date')
plt.ylabel('Counts')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

# Plot 2: Users and Sessions by Hours
plt.figure(figsize=(12, 6))
df.groupby('Hours')[['Users', 'Sessions']].sum().plot(ax=plt.gca())
plt.title('Users and Sessions by Hour of the Day')
plt.xlabel('Hour')
plt.ylabel('Counts')
plt.xticks(range(0, 24))
plt.tight_layout()
plt.show()
```

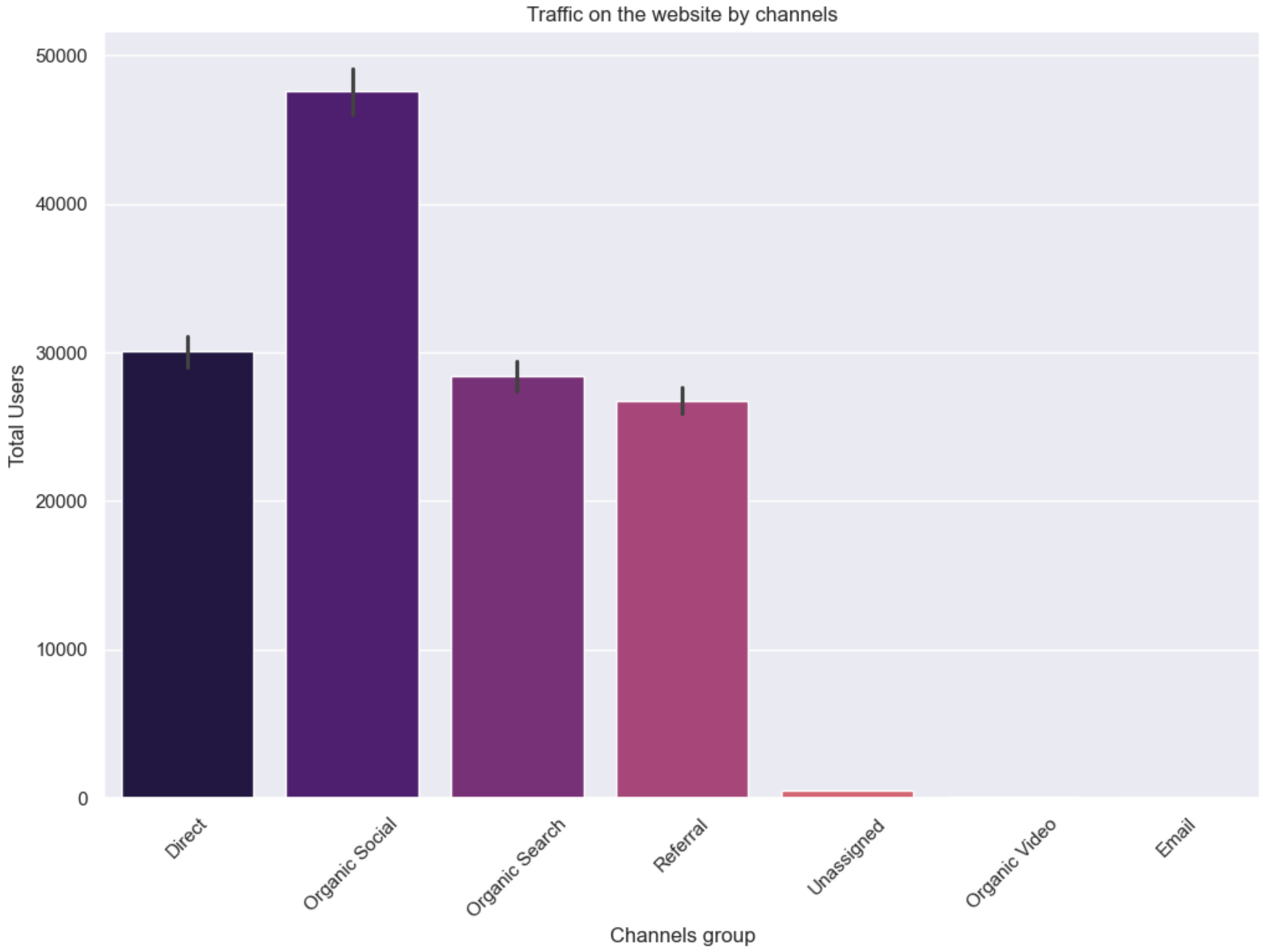




Which marketing channel brought the highest number of users to the website, and how can we use this insight to improve traffic from other sources?

In [11]:

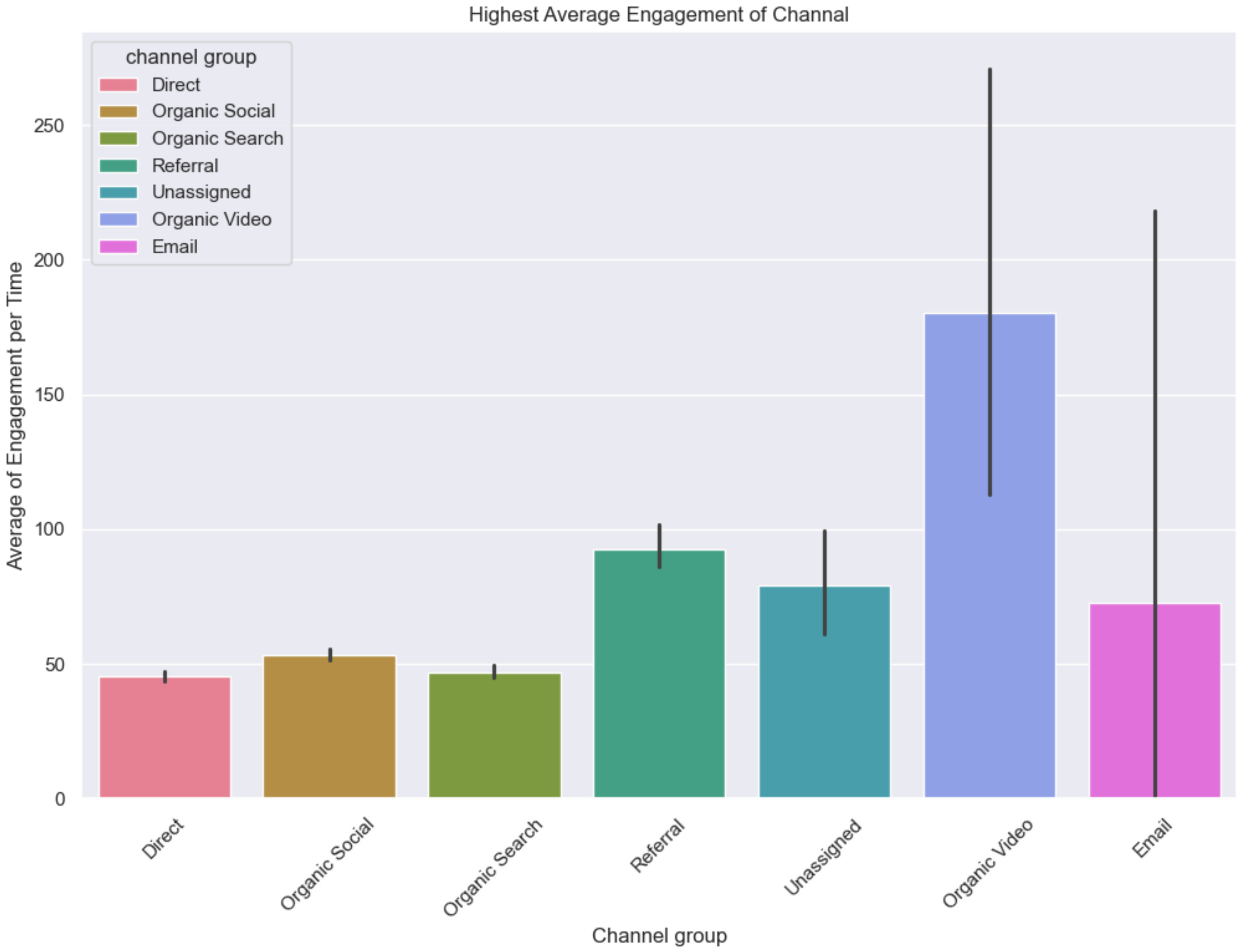
```
# Traffic of the users on the website by the channels group
plt.figure(figsize=(12,8))
sns.barplot(data=df, x='channel group', y='Users', estimator=np.sum, palette='magma', hu
plt.title('Traffic on the website by channels')
plt.xlabel('Channels group')
plt.ylabel('Total Users')
plt.xticks(rotation=45)
plt.show()
```



Which channel has the highest average engagement time, and what does that tell us about user behavior and content effectiveness?

In [12]:

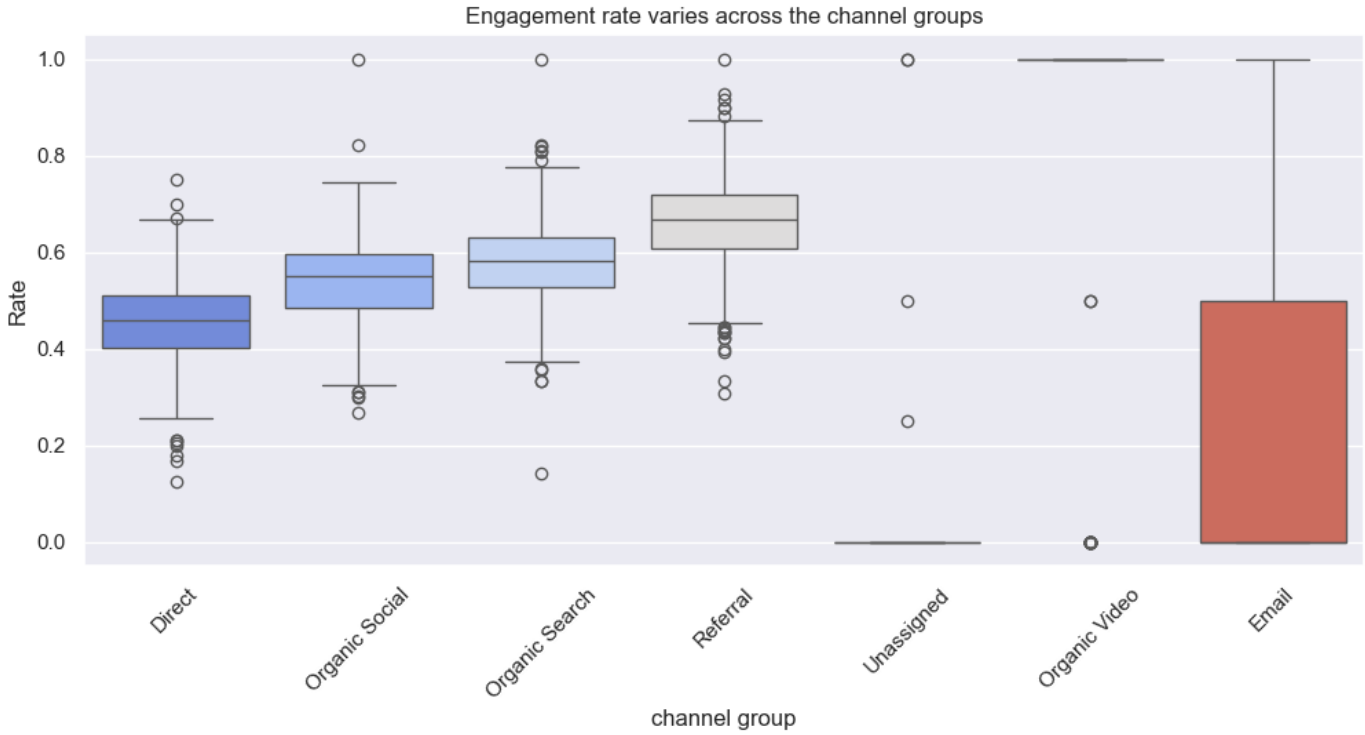
```
# Which channels have the highest average engagement
plt.figure(figsize=(12,8))
sns.barplot(data=df, x='channel group', y='Average engagement time per session', hue='ch
plt.title('Highest Average Engagement of Channal')
plt.xlabel('Channel group')
plt.ylabel('Average of Engagement per Time')
plt.xticks(rotation=45)
plt.show()
```



How does engagement rate vary across different traffic channels?

In [21]:

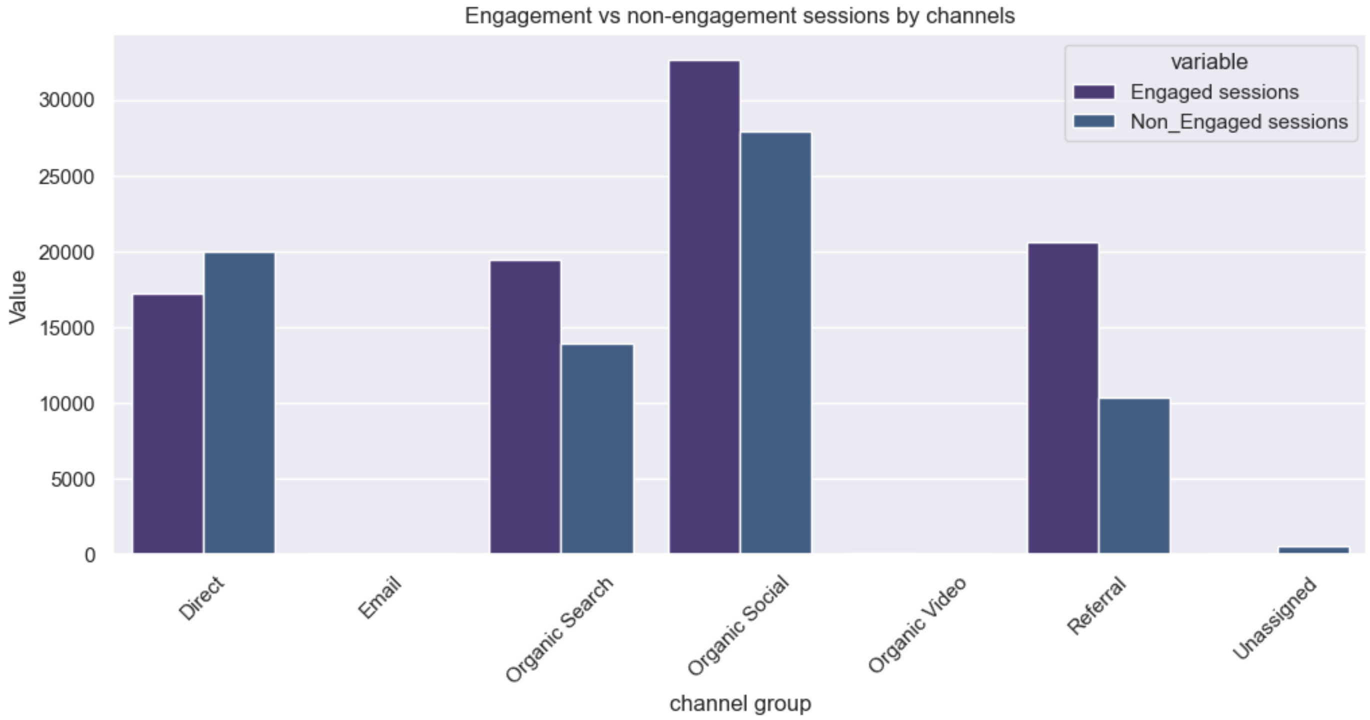
```
# Engagement rate varies across the Channel groups
plt.figure(figsize=(12,5))
sns.boxplot(data=df, x='channel group', y='Engagement rate', hue='channel group', palett
plt.title('Engagement rate varies across the channel groups')
plt.xlabel('channel group')
plt.ylabel('Rate')
plt.xticks(rotation=45)
plt.show()
```



Which channels are driving more engaged sessions compared to non-engaged ones, and what strategies can improve engagement in underperforming channels?

In [22]:

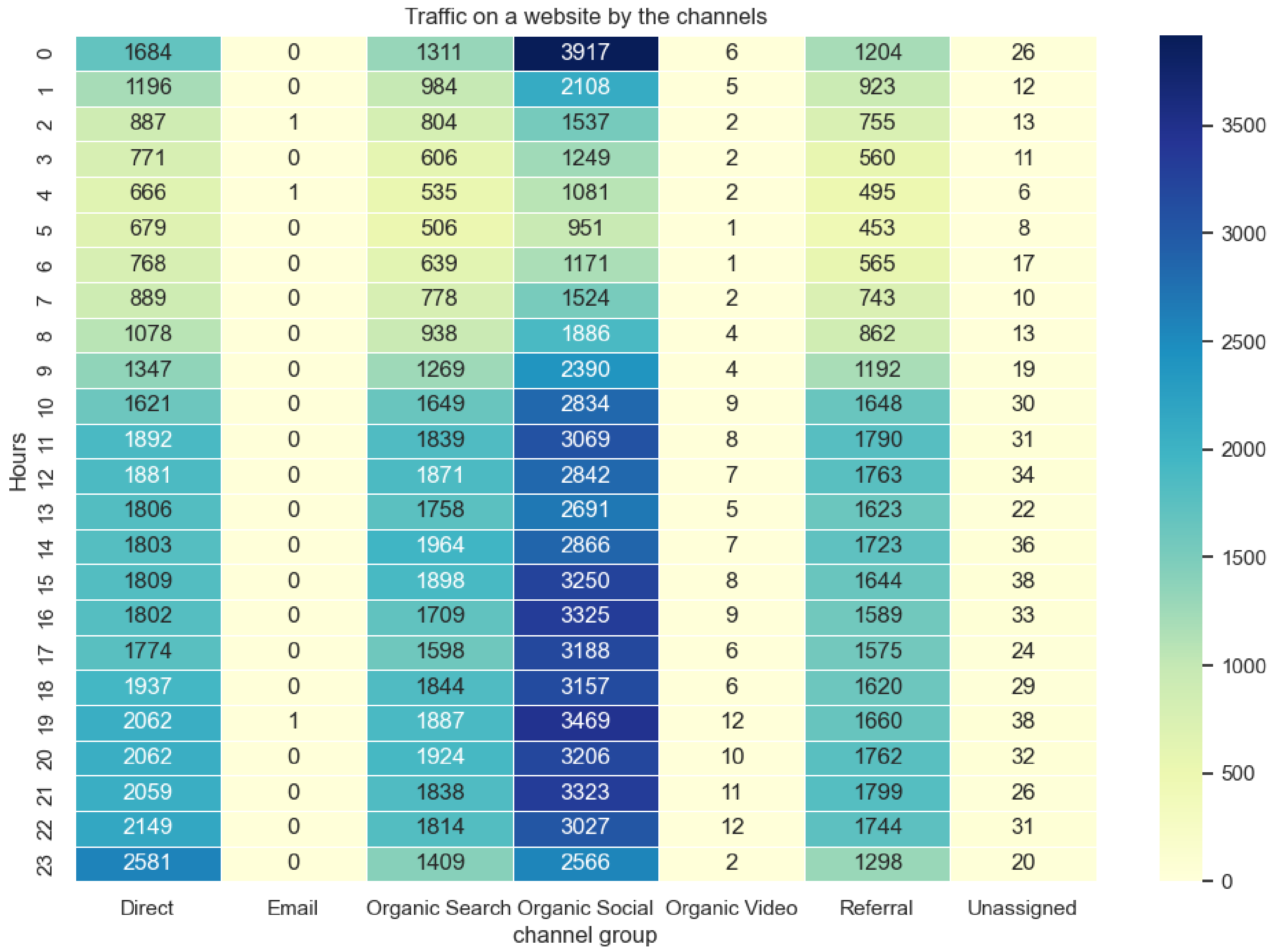
```
# Engagement vs non-engagement users come across which channel group
session_df = df.groupby('channel group')[['Sessions','Engaged sessions']].sum().reset_in
session_df['Non_Engaged sessions'] = session_df['Sessions']-session_df['Engaged sessions
sessions_df_melted = session_df.melt(id_vars=['channel group'], value_vars=['Engaged ses
plt.figure(figsize=(12,5))
sns.barplot(data=sessions_df_melted, x='channel group', y='value', hue='variable', legen
plt.title('Engagement vs non-engagement sessions by channels')
plt.xlabel('channel group')
plt.ylabel('Value')
plt.xticks(rotation=45)
plt.show()
```



At what hours of the day does each channel drive the most traffic?

In [24]:

```
Traffic_ = df.groupby(['Hours', 'channel group'])['Sessions'].sum().unstack().fillna(0)
plt.figure(figsize=(12,8))
sns.heatmap(Traffic_ , cmap="YlGnBu", linewidths = .5, annot = True, fmt = '.0f')
plt.title('Traffic on a website by the channels')
plt.show()
```



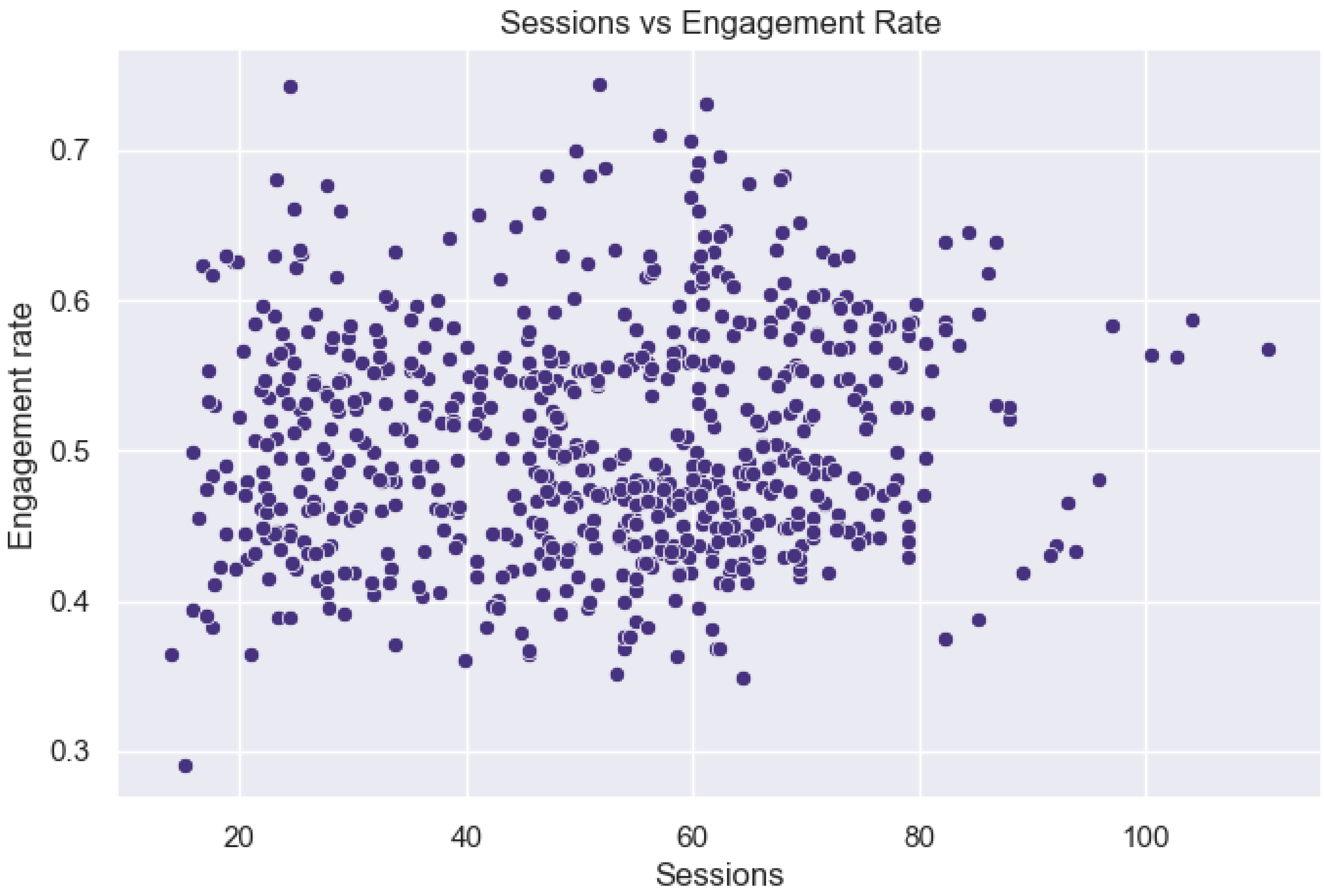
Hours	Direct	Email	Organic Search	Organic Social	Organic Video	Referral	Unassigned
0	1684	0	1311	3917	6	1204	26
1	1196	0	984	2108	5	923	12
2	887	1	804	1537	2	755	13
3	771	0	606	1249	2	560	11
4	666	1	535	1081	2	495	6
5	679	0	506	951	1	453	8
6	768	0	639	1171	1	565	17
7	889	0	778	1524	2	743	10
8	1078	0	938	1886	4	862	13
9	1347	0	1269	2390	4	1192	19
10	1621	0	1649	2834	9	1648	30
11	1892	0	1839	3069	8	1790	31
12	1881	0	1871	2842	7	1763	34
13	1806	0	1758	2691	5	1623	22
14	1803	0	1964	2866	7	1723	36
15	1809	0	1898	3250	8	1644	38
16	1802	0	1709	3325	9	1589	33
17	1774	0	1598	3188	6	1575	24
18	1937	0	1844	3157	6	1620	29
19	2062	1	1887	3469	12	1660	38
20	2062	0	1924	3206	10	1762	32
21	2059	0	1838	3323	11	1799	26
22	2149	0	1814	3027	12	1744	31
23	2581	0	1409	2566	2	1298	20

Is there any correlation between high traffic (sessions) and high engagement rate over time?

In [17]:

```
df_plot = df.groupby('Date')[['Sessions', 'Engagement rate']].mean().reset_index()
correlation = df_plot['Sessions'].corr(df_plot['Engagement rate'])
print('Correlation between Sessions and engagement Rate:',correlation)
plt.figure(figsize=(8,5))
sns.scatterplot(data=df_plot, x="Sessions", y='Engagement rate')
plt.title('Sessions vs Engagement Rate')
plt.show()
```

```
Correlation between Sessions and engagement Rate: 0.08344754002348544
```



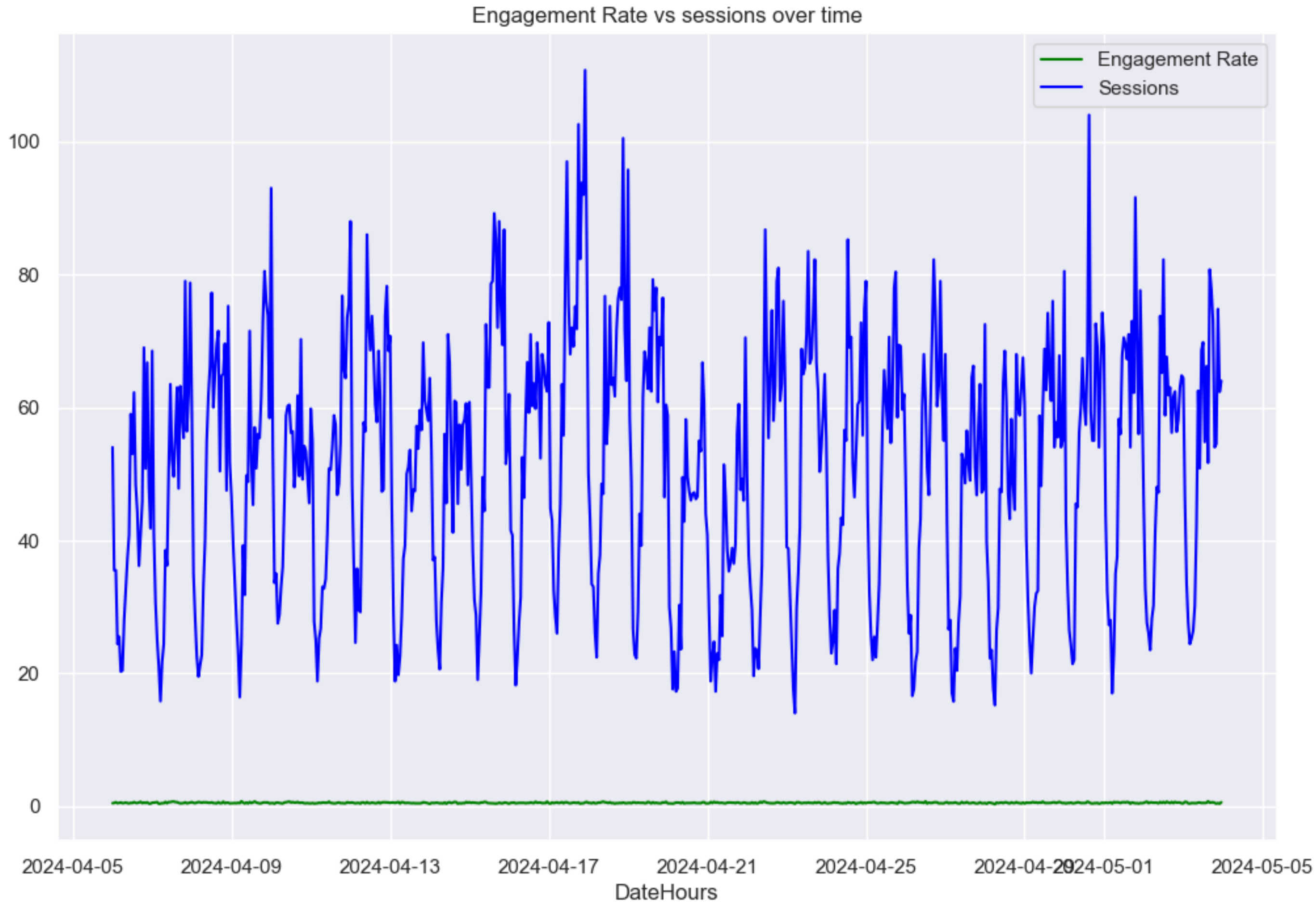
● Correlation Analysis: A correlation analysis was conducted between daily sessions and engagement rate over the observed time period. The resulting Pearson correlation coefficient was 0.08, indicating a very weak positive relationship.

● Key Insight: This suggests that there is no meaningful correlation between the volume of traffic and user engagement rate. Therefore, higher traffic does not guarantee higher engagement. User engagement appears to be influenced by other factors such as content quality, user experience, or marketing channel relevance rather than the traffic volume alone.

● Recommendation: To improve engagement, focus should be placed on optimizing content, improving UX, and targeting high-intent audiences, rather than solely increasing session volume.

In [18]:

```
plt.figure(figsize=(12,8))
plt.plot(df_plot['Date'],df_plot['Engagement rate'],  label='Engagement Rate', color='gr
plt.plot(df_plot['Date'], df_plot['Sessions'], label='Sessions', color='blue')
plt.title('Engagement Rate vs sessions over time')
plt.xlabel('DateHours')
plt.legend()
plt.show()
```



📌 Summary of Insights

1. Insights from Website Traffic Analysis
   ◆ Date-wise Trend:

- Upon analyzing the sessions and users by date, the **highest traffic was observed between the dates 15th to 19th**.
- This period consistently showed **elevated user activity**, suggesting a likely event, promotion, or regular weekly pattern.
  ◆ Hour-wise Trend:
- When analyzing by hour, the **peak sessions occurred around 19:00 (7 PM)**.
- This suggests that users are **most active during the evening hours**, possibly after work hours.
  ◆ Business Implications:
- For maximum reach and engagement, campaigns or content releases could be **timed between the 15th and 19th of each month** and scheduled for the **evening hours around 7 PM**.

2. Marketing Channel Performance Analysis
   ◆ Top Performing Channel:

- **Organic Social** generated the **highest number of users**, indicating strong performance through unpaid social media activities such as posts, shares, and community engagement.
- This shows that organic social efforts are effectively reaching and engaging the target audience.
  ◆ Underperforming Channels:

- Channels like **Email** and **Organic Video** reported **zero user traffic** during the analyzed period.
- This suggests either:
  - These channels are **inactive or not being utilized**, or
  - Current campaigns under these channels are **not reaching or engaging users**.

    ◆ Actionable Insights:
- **Replicate successful strategies** from Organic Social (e.g., regular posting, user-focused content, engagement) in other channels.
- For **Email**:
  - Ensure email campaigns are active and well-targeted.
  - Audit mailing lists and delivery performance.
- For **Organic Video**:
  - Assess if video content is being published and properly distributed (YouTube, Reels, Shorts).
  - Improve visibility through SEO and social sharing.

    ◆ Recommendation:
- Continue investing in **Organic Social** while developing a strategic plan to activate and improve the performance of **Email** and **Organic Video** as additional growth channels.

3. Engagement Time Insight

- **Organic Video** had the **highest average engagement time**, peaking over **250 seconds**, indicating strong content effectiveness and deep user interest.
- **Email** also showed a high tail (200+), but most users dropped off before 100 seconds.
- Other channels like **Direct**, **Organic Search**, and **Social** showed lower engagement (under 60 seconds), suggesting less engaging or intent-driven sessions.

  ◆ Recommendation:
- Invest more in **video content** and **optimize email follow-ups** to retain attention.
- Reassess content strategy for underperforming channels like Direct and Social.

4. Engagement Rate by Channel Group

- **Referral**, **Organic Search**, and **Organic Social** show **higher median engagement rates**.
- **Direct** has a moderate median but more spread, indicating mixed engagement.
- **Email** and **Organic Video** show **extremely wide variation**, suggesting inconsistent performance.
- **Unassigned** has the **lowest engagement rate**, often near zero.

  ◆ Recommendation:

  Focus on optimizing **Email** and **Organic Video** for consistency, and consider improving **Unassigned** traffic quality.

5. Engaged vs Non-Engaged Sessions by Channel

- **Organic Social**, **Referral**, and **Organic Search** have more **engaged sessions than non-engaged**, showing strong user interaction from these sources.
- **Direct** has more **non-engaged sessions**, suggesting low intent or weak landing experiences.
- **Email** and **Organic Video** show **no engagement**, indicating either inactive campaigns or ineffective content.

◆ Recommendation:

- Focus on enhancing content for **Direct** and activating campaigns for **Email** and **Organic Video**.
- Continue leveraging high-performing channels like **Organic Social** and **Referral** for quality traffic.

6. Peak Traffic Hours by Channel

- **Organic Social** drives the **highest traffic overall**, peaking at **3,917 sessions at 0:00 hours (midnight)**.
- Other channels like **Direct**, **Referral**, and **Organic Search** also show consistent activity between **18:00 to 22:00 hours**.
- **Email** and **Organic Video** show **no traffic across all hours**.

  ◆ Recommendation:
- Schedule key content and campaigns on **Organic Social around midnight** and **evenings (6–10 PM)** for maximum reach.
- Activate and test **Email** and **Video** campaigns to evaluate potential contribution.

In [ ]:

```

```