Findings and Report

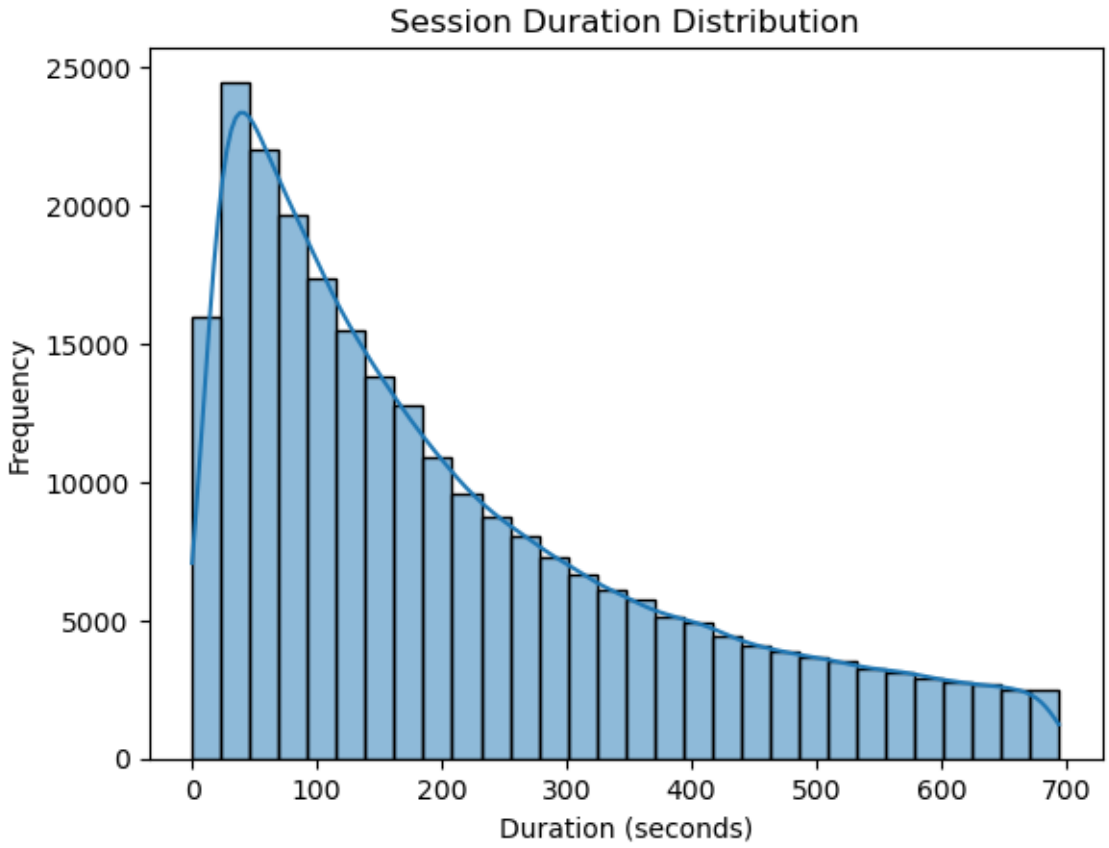
Introduction

This report looks at user sessions to understand how people interact with the platform. It focuses on key numbers like Clickout Ratio (COR), Platform, Traffic Type, Booking Rate, and Engagement Ratio to find out what works well and what needs improvement. The report also gives easy-to-follow suggestions to make the platform better for users, attract the right traffic, and increase bookings.

Tool used for the analysis: **Python and Excel**

1. Descriptive Analysis

- **Session Data**:
    - Total sessions: 254,177.
    - Average session duration: ~128 seconds, with a max of 694 seconds.
    - Most sessions have fewer than 200 seconds, indicating short browsing activity.



- **Clickouts and Bookings**:
    - Average clickouts per session: ~1.58.

- Bookings per session are very low (mean = 0.016), indicating most sessions don't result in bookings.

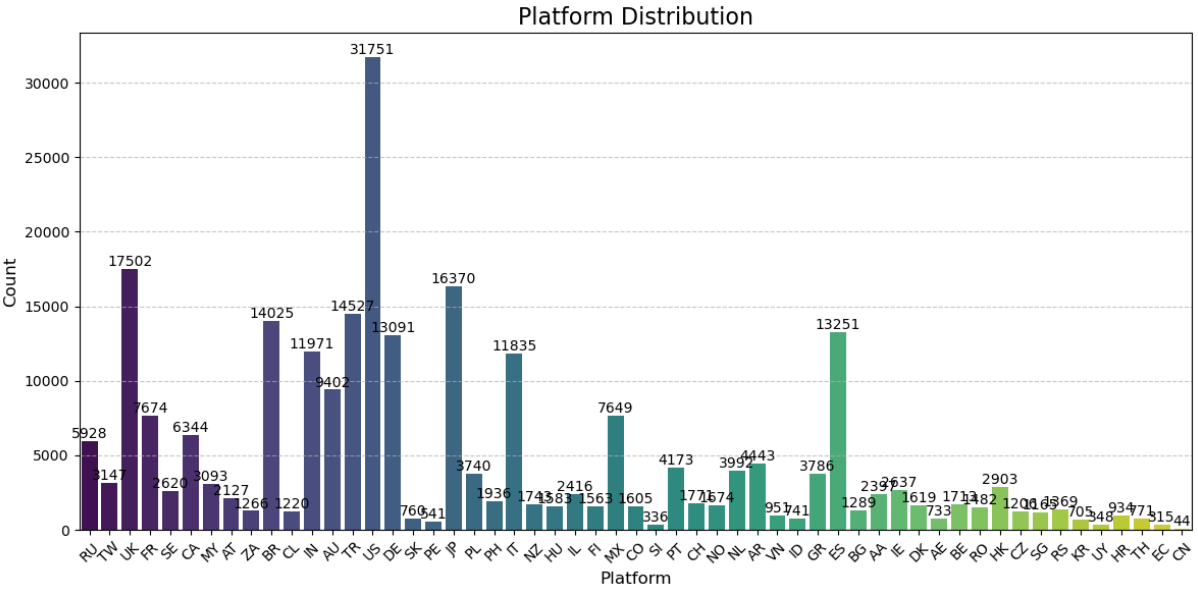
```
In [12]: # Descriptive statistics
         print(df_split.describe())  # Summary of numerical columns
         print(df_split['platform'].value_counts())  # Frequency of platforms
         print(df_split['traffic_type'].value_counts())  # Frequency of traffic types
```

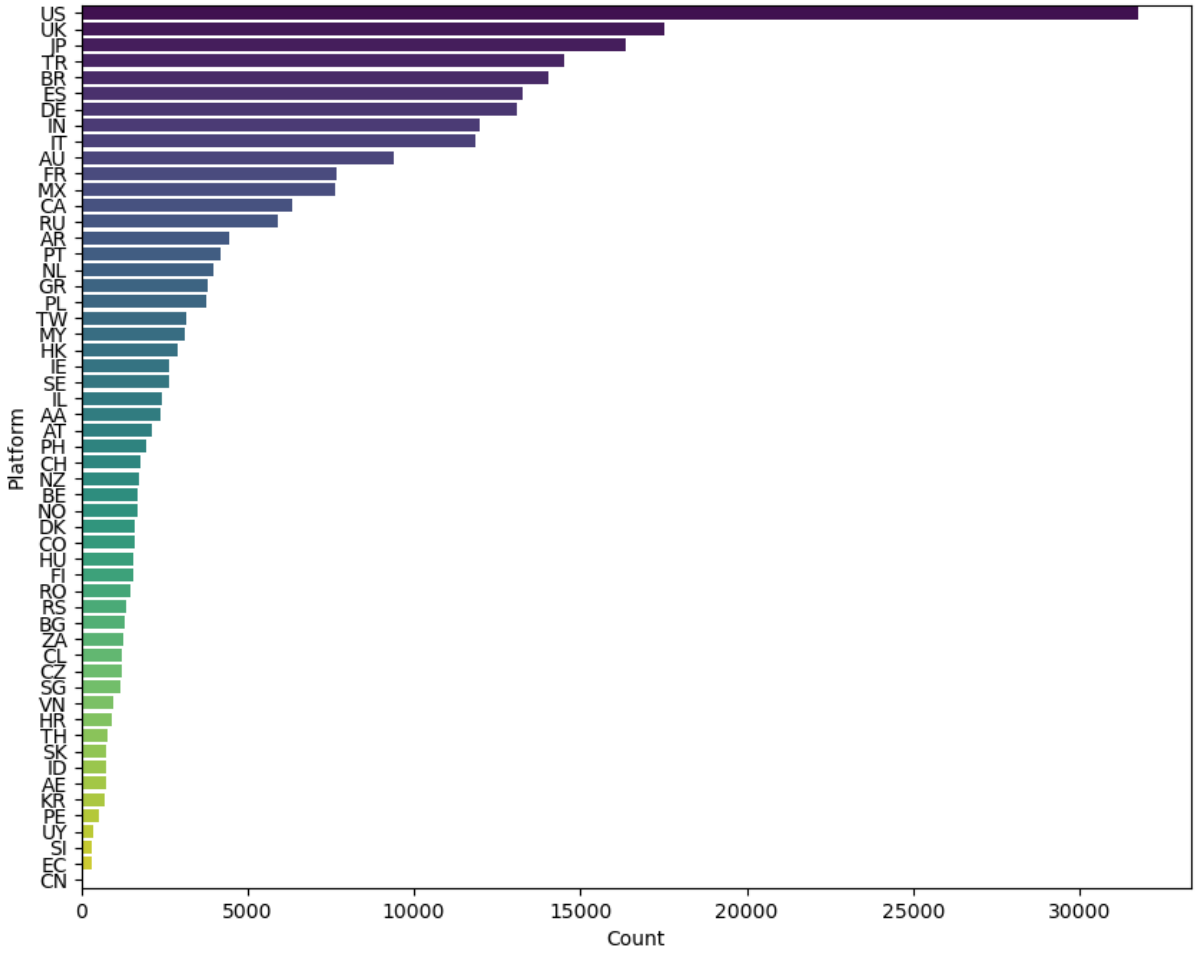
```
                                ymd    session_id    is_app    is_repeater  \
count                        254177  2.541770e+05  254177.0  254177.000000
mean   2022-05-30 19:26:51.407798528  2.022056e+18       0.0       0.431648
min              2022-05-01 00:00:00  2.022050e+18       0.0       0.000000
25%              2022-05-15 00:00:00  2.022052e+18       0.0       0.000000
50%              2022-05-31 00:00:00  2.022053e+18       0.0       0.000000
75%              2022-06-15 00:00:00  2.022062e+18       0.0       1.000000
max              2022-06-30 00:00:00  2.022063e+18       0.0       1.000000
std                              NaN  5.062950e+12       0.0       0.495307

        traffic_type       agent_id      clickouts       bookings  \
count  254177.000000  254177.000000  254177.000000  254177.000000
mean        4.730562      15.212974       1.581587       0.016150
min         2.000000       2.000000       1.000000       0.000000
25%         2.000000      12.000000       1.000000       0.000000
50%         4.000000      18.000000       1.000000       0.000000
75%         6.000000      20.000000       2.000000       0.000000
max        10.000000      28.000000      41.000000       8.000000
std         2.437011       7.454489       1.092401       0.130259

       session_duration     entry_page      total_ctp  \
count     254177.000000  254177.000000  254177.000000
mean         212.479268    2114.565106       5.342694
min            0.000000    2100.000000       0.000000
25%           71.000000    2100.000000       0.000000
50%          158.000000    2113.000000       0.000000
75%          315.000000    2116.000000       2.000000
max          694.000000    2502.000000     621.000000
std          175.382184      33.691789      16.353974
```

- **Platforms**:
  - Most traffic comes from the US (31,751 sessions), followed by the UK and Japan.
  - Significant variation across countries in platform usage.





- **Traffic Types**:
    - Traffic types are coded (e.g., 2, 6, 8), with 2 being the most frequent (94,331 sessions).

```
Name: count, dtype: int64
traffic_type
2     94331
6     68958
8     53321
4     33708
10     3859
Name: count, dtype: int64
```

Calculate the clickout ratio per platform and device type - what platform has the highest COR? What device has the lowest COR?

```
In [19]: # Overall COR
         overall_cor = df_split['clickouts'].sum() / df_split['session_id'].nunique()
         print("Overall Clickout Ratio:", overall_cor)

         Overall Clickout Ratio: 1.581586846960976
```

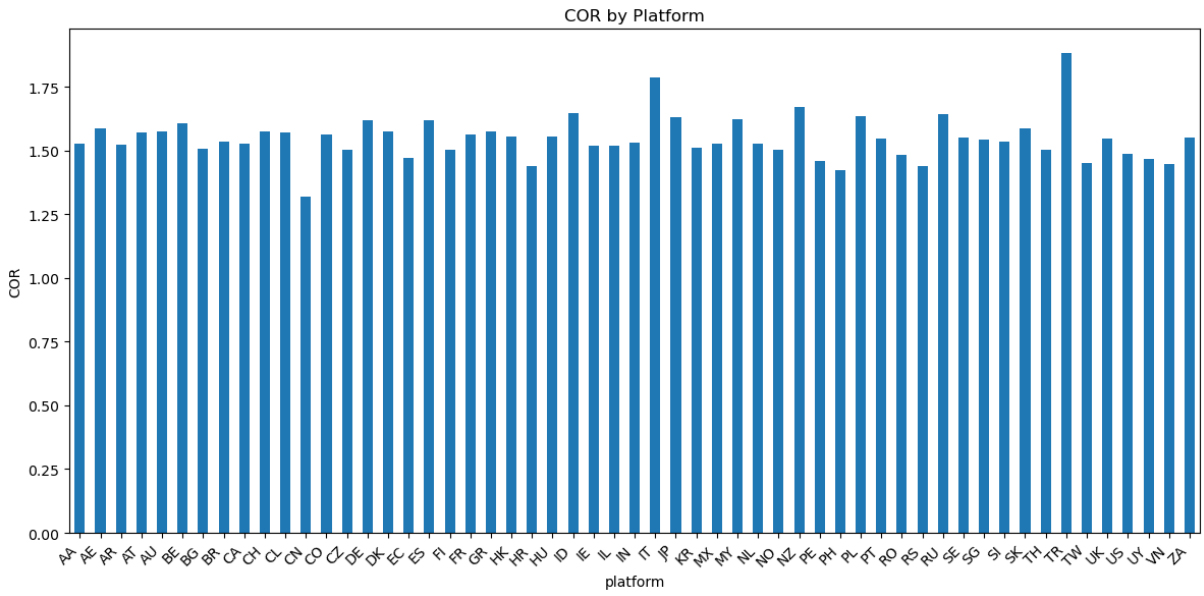
Clickout Ratio (COR) by Platform

- **Highest COR**: The platform **TR** (Turkey) has the highest COR at **1.882**, indicating strong user engagement and interaction on this platform.

```
SI              516         336  1.535714
SK             1205         760  1.585526
TH             1159         771  1.503243
TR            27347       14527  1.882495
TW             4562        3147  1.449635
UK            27089       17502  1.547766
US            47184       31751  1.486063
```

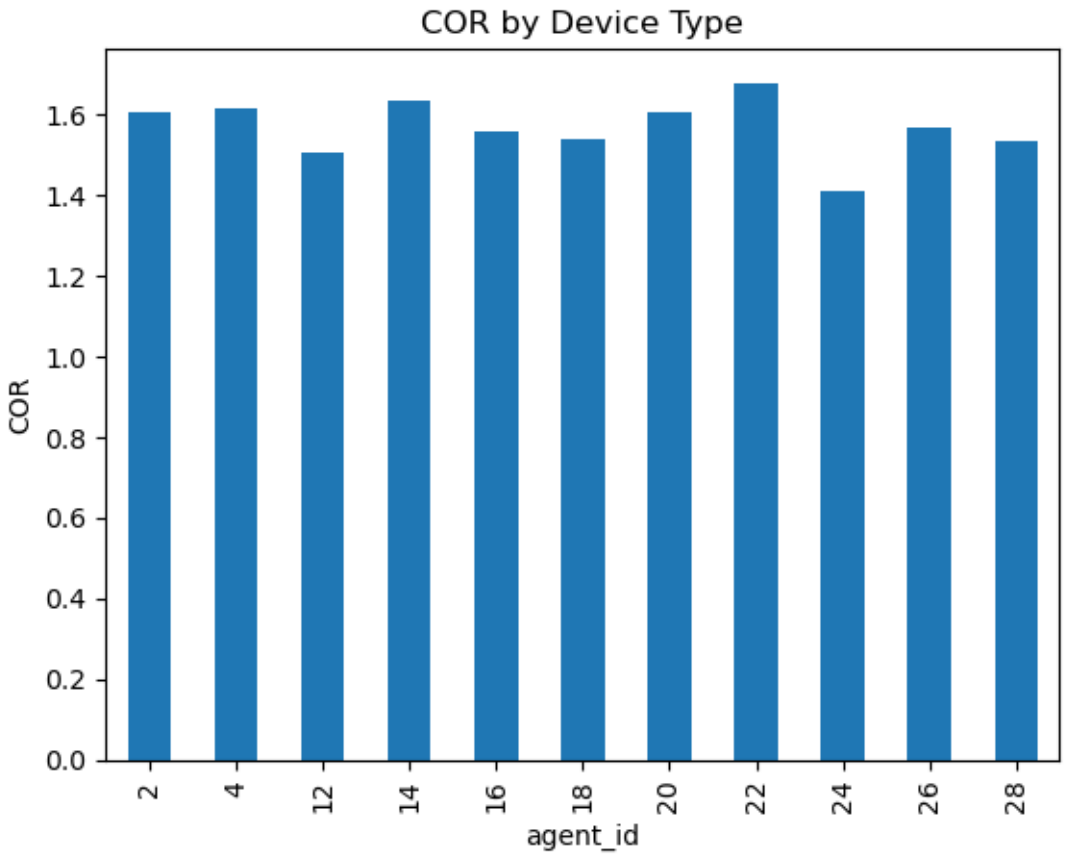
- **Lowest COR**: The platform **CN** (China) has the lowest COR at **1.381**, suggesting minimal user engagement compared to other platforms.

```
CA             9696        6344  1.528373
CH             2790        1771  1.575381
CL             1916        1220  1.570492
CN               58          44  1.318182
CO             2506        1605  1.561371
CZ             1811        1206  1.501658
DE            21180       13091  1.617905
```



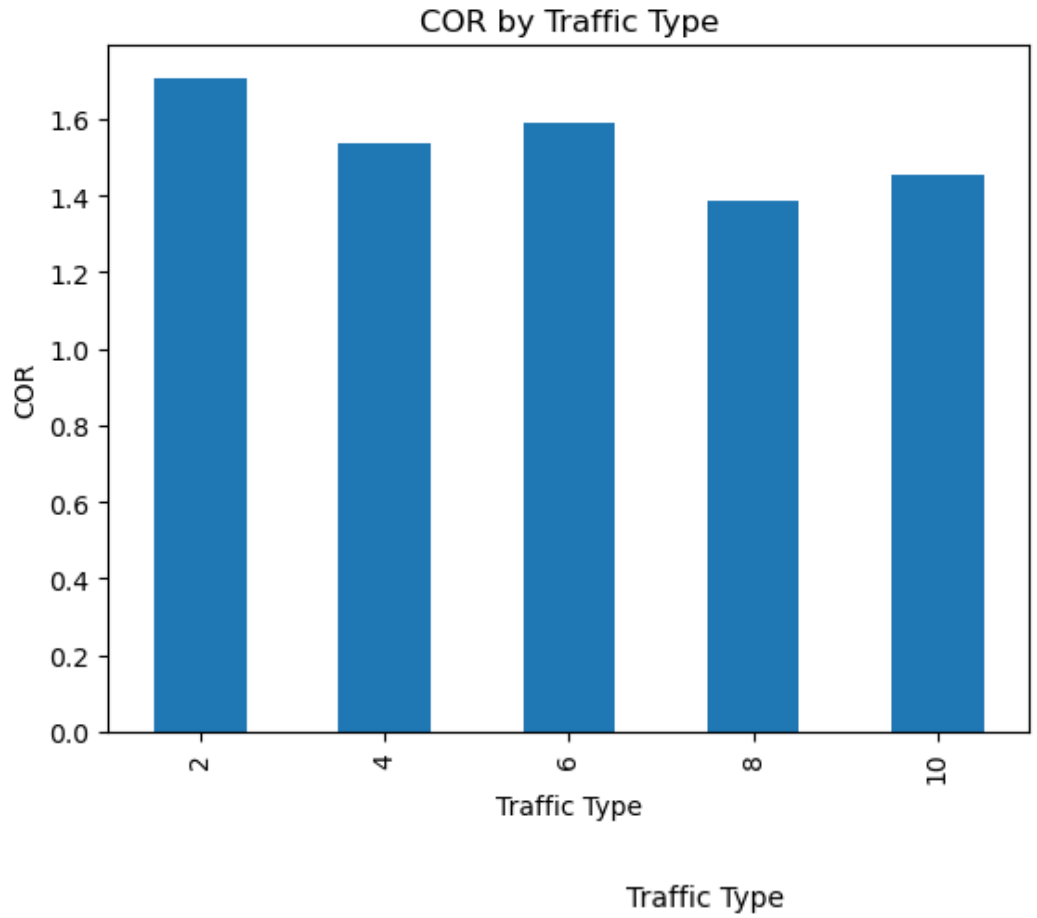
Clickout Ratio (COR) by Device Type

- **Lowest COR**: Device ID **24** has the lowest COR at **1.410**, suggesting room for improvement in user interaction on this device type.



Are there differences by traffic type? Can you draw any conclusions from the ratios about the coded values for traffic type?

traffic_type	clickouts	session_id	COR
2	160959	94331	1.706321
4	51775	33708	1.535986
6	109782	68958	1.592013
8	73874	53321	1.385458
10	5613	3859	1.454522



Traffic Type

Traffic Types with High COR:

traffic_type	clickouts	session_id	COR
2	160959	94331	1.706321
4	51775	33708	1.535986
6	109782	68958	1.592013

Traffic Types with Low COR:

traffic_type	clickouts	session_id	COR
8	73874	53321	1.385458
10	5613	3859	1.454522

Yes, the analysis shows significant differences in **Clickout Ratio (COR)** across the traffic types:

- **Traffic Type 2**: Has the highest COR of **1.706**, indicating strong user engagement from this traffic source.
- **Traffic Type 8**: Has the lowest COR of **1.385**, suggesting this traffic source contributes less to user engagement.
- Other traffic types like **4** and **6** have moderate COR values, falling between the extremes.

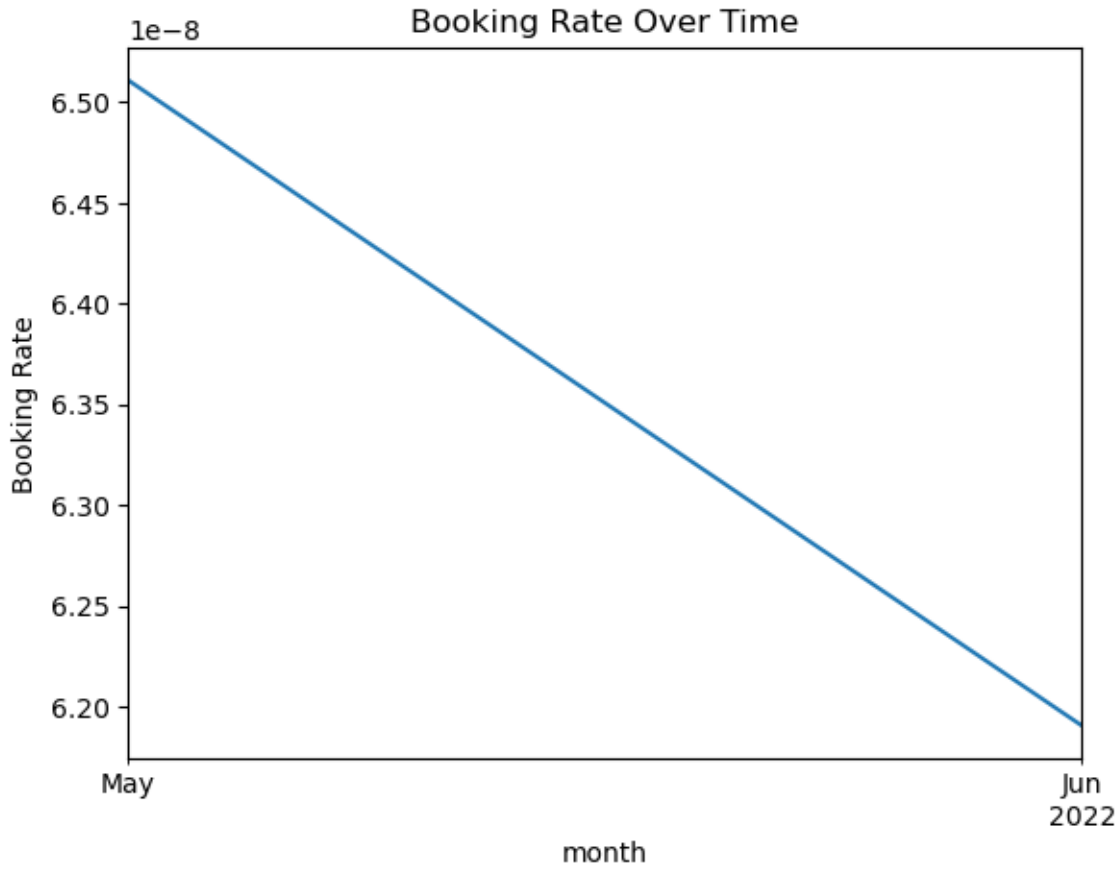
Conclusions from the Ratios

1. **High COR Traffic Types**:
    - **Traffic Type 2** likely represents targeted traffic sources (e.g., paid ads or direct marketing campaigns) that are optimized for engagement.
    - These sources are likely high-quality traffic, driving users who are more inclined to interact with the platform.
2. **Low COR Traffic Types**:
    - **Traffic Type 8** may represent less-targeted or irrelevant sources (e.g., organic search or broad referral links), which result in lower engagement.
    - **Traffic Type 10**, while less frequent, also shows lower COR, which could indicate untapped potential or a need for optimization.

Can you think of an additional two KPIs? Please provide KPI definitions and charts or tables showing the development over time along with descriptions and critical analysis of what you see.

Additional KPIs

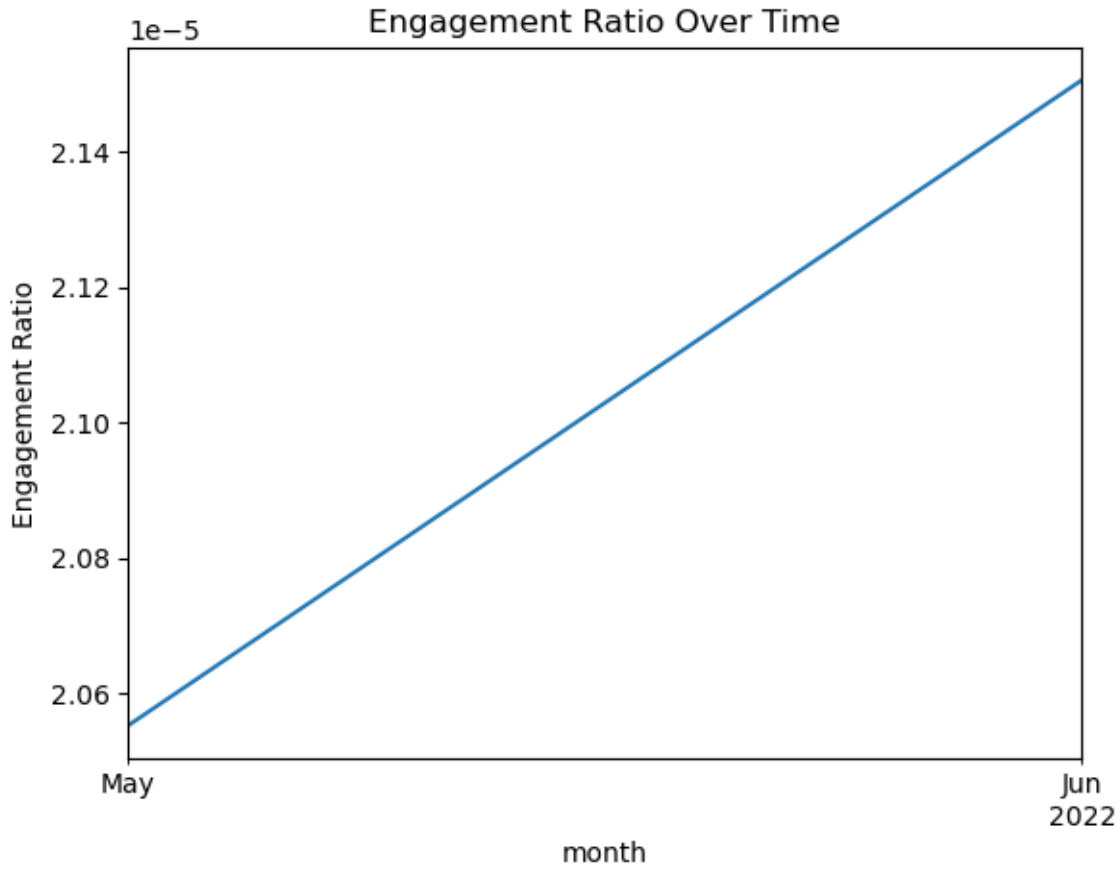
1. **Booking Rate**:
    - Definition: Measures the proportion of sessions that result in bookings:
    - **Trend**: The chart shows a **decline in Booking Rate** over time. This suggests a decreasing ability to convert sessions into bookings.



Critical Analysis:

- A declining booking rate may indicate issues with the booking process (e.g., usability problems or ineffective marketing).

2. **Engagement Ratio**:
    - Definition: Measures how much content users view during a session:
    - **Trend**: The chart shows a **steady increase in Engagement Ratio** over time. This indicates users are interacting more with the platform's content.



Critical Analysis:

- While engagement is increasing, this does not translate into bookings. There may be a disconnect between content engagement and actions like booking.

Findings and Recommendations Table

Category	Findings	Reason	Recommendations
Session Metrics	Avg. session is 128 seconds, COR is 1.58 (moderate). Booking Rate is very low (0.016). Engagement is increasing.	Users like content but are not booking due to barriers or process issues.	Make bookings easier and give users incentives to book.
Traffic Types	Type 2 has the best COR (1.706), Type 8 has the worst (1.385).	Type 2 is from good traffic sources, Type 8 has less relevant users.	Focus on Type 2 traffic and improve engagement for Type 8.
Platforms	TR has the best COR (1.882), CN has the worst (1.381).	TR may have better design or targeting, CN may need improvements.	Learn from TR and fix issues on CN for better results.

Category	Findings	Reason	Recommendations
Devices	Device 22 has the best COR (1.677), Device 24 has the worst (1.410).	Device 22 works well; Device 24 might have issues with performance or usability.	Optimize for Device 24 to match the performance of Device 22.
Conversion Metrics	10-15% of sessions convert. CN and Traffic Types 8 and 10 have high bounce rates.	Bounce rates are high because of poor landing pages or slow performance.	Fix landing pages and make the site faster to reduce bounce rates.