Data Analyst - Product Intelligence

Case Study

By Oluwatobi Ojo

Task 1

Session Investigation

Challenge 1 - Descriptive Analysis

Trivago operates as a meta search engine company for accommodations around the world and offering accurate comparisons and recommendations for its users making profit from charging platforms onboarded from clicks, listings, advertisements, subscriptions etc.

This to me means as much as possible we need to keep users engaged on the platform and determine what is key to track, which features drives adoption, are the users willing to come back, and how much is done by a user during sessions.

Let’s dive in, I’ll put up some of the basic things from the data to explore and explain my findings.

Nb: Note that this report does not prioritize the analysis of Bookings, as Trivago's revenue model does not rely on this Key Performance Indicator (KPI).

- Data structure (data cleaning around the country_name column)

- Statistical measures of numerical columns in the dataset (clickouts, bookings, session_duration, total_ctp)

- Distribution of sessions across various categorical columns such as device used, platform, traffic type etc.

- How does specific categorical columns influence clickouts, retention rate (repeat visitor rate)

- Time series analysis of sessions

- Relationships/patterns between specific variables in the data

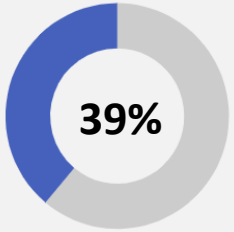
Statistical Measures

	clickouts	bookings	session_duration	total_ctp
count	900000.000000	900000.000000	900000.000000	900000.000000
mean	0.843981	0.009877	392.361784	6.486588
std	1.813060	0.107203	989.495531	30.995589
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	12.000000	0.000000
50%	0.000000	0.000000	65.000000	0.000000
75%	1.000000	0.000000	287.000000	1.000000
max	86.000000	8.000000	83335.000000	3662.000000

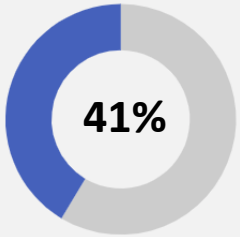
Nb: To be certain the large values are not outliers, and they are distributed across a number of sessions, I also plotted a boxplot to view the distributions, which showed they are not isolated outliers.

This shows that on an average, most sessions did not make clickouts, spent averagely just about a minute on the site and viewed less than a content when they visited the site. These are worrisome findings, and I would suggest exploring more on the reasons that could be behind these behaviors. Asking questions like;

- %of sessions that ended up making at least a click vs ones that didn't.
- Does being a repeat user affect usage of the product positively compared to those who aren't repeat users.
- Does the device used, platform of usage, traffic/marketing channel in which the user originate from matter
- Has the app usage changed over time positively, could be because of newer features added or product team making modifications



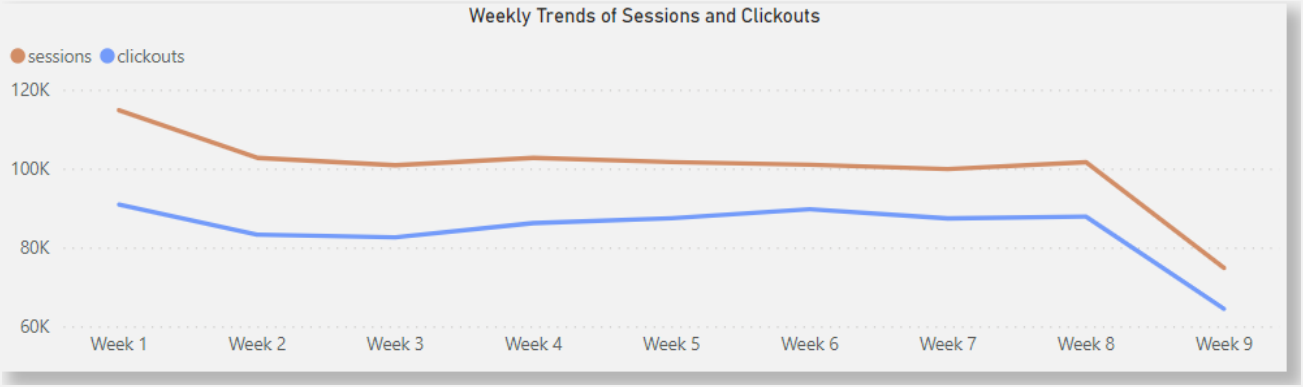
Only 39% of sessions led to the user making at least a clickout



Checking just repeat users

Only 41% of sessions by repeat users led to the user making at least a clickout

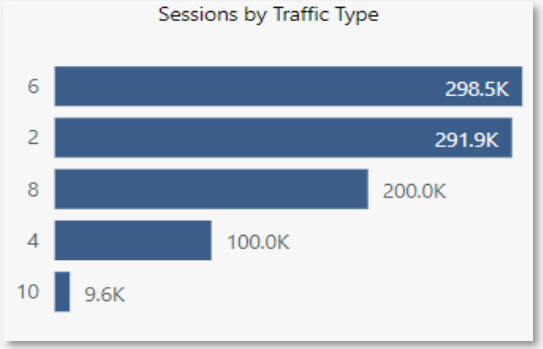
This shows that being a repeat user has little or no effect in determining if a user gets to make a clickout during a session

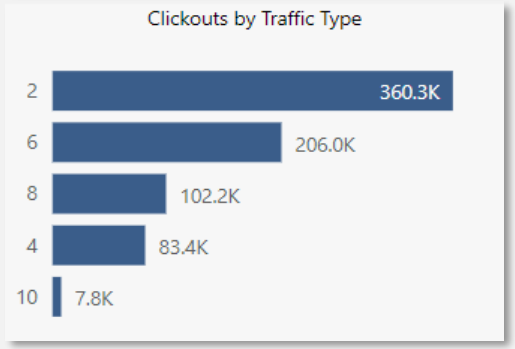


This shows there has been a slight decline in the number of sessions and clickouts from the launch till date. This is a worry as these trends should tend to increase as modifications are made on the product.

Categorical Variables Distribution

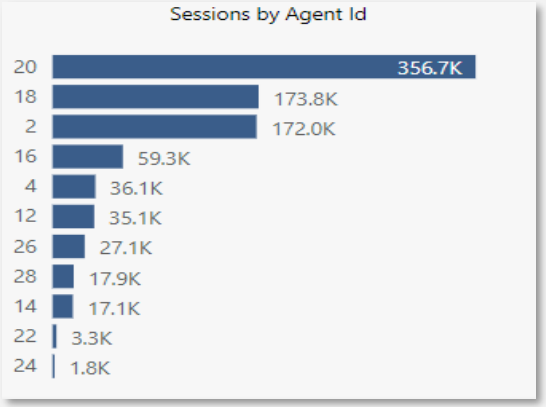
Traffic Type (marketing channel)

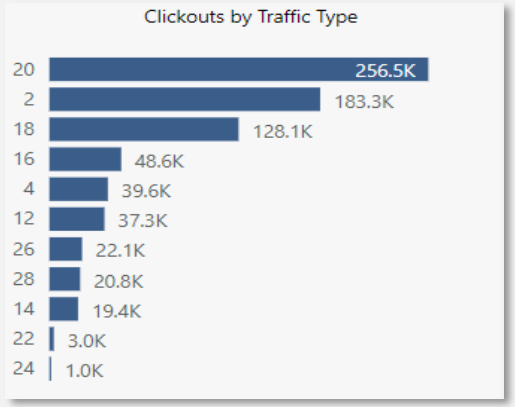




The charts highlight traffic sources' performance in driving user engagement and clickouts. Traffic source 6 exhibits low effectiveness in prompting user clickouts, whereas Traffic source 2 significantly excels in encouraging user engagement through clickouts.

Agent Id (device type)





The charts highlight device used performance in driving user engagement and clickouts. Device 18 exhibits low effectiveness in prompting user clickouts, whereas device 2 significantly excels in encouraging user engagement through clickouts.

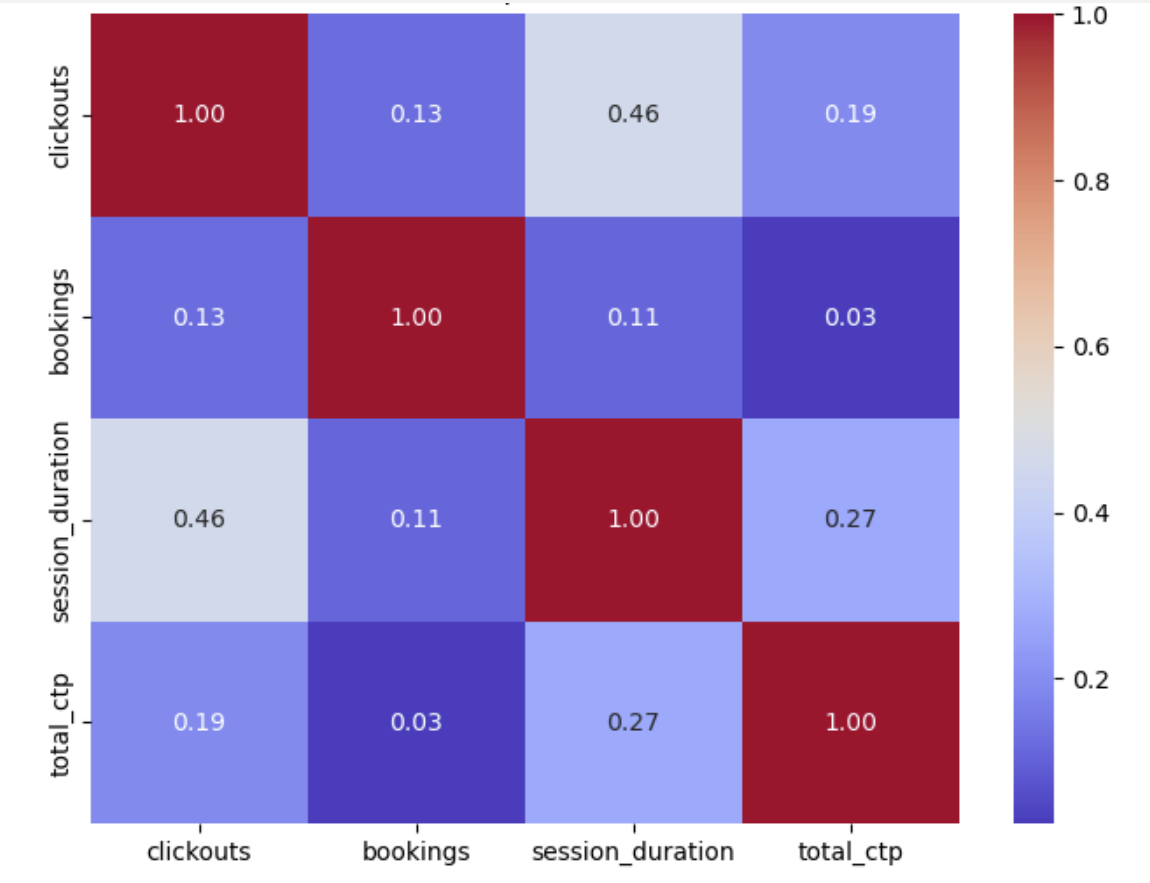
Platform (country code) – top 20 by sessions

platform	sessions	total clickouts	clickouttosession
US	117589	87240	-25.81%
UK	55945	47428	-15.22%
IN	54670	34625	-36.67%
BR	54087	42804	-20.86%
TR	48254	59227	22.74%
JP	45800	49963	9.09%
DE	44240	41711	-5.72%
IT	41296	41494	0.48%
ES	39397	38613	-1.99%
AU	32894	28528	-13.27%
MX	31322	23410	-25.26%
FR	27579	21421	-22.33%
CA	22121	19119	-13.57%
RU	19151	17934	-6.35%
AR	17114	12613	-26.30%
GR	13380	11398	-14.81%
PT	13065	11085	-15.15%
NL	12749	11210	-12.07%
AA	12132	6605	-45.56%
MY	11690	10014	-14.34%

Clickouttosession = (Total clickouts – sessions) *100
sessions

The analysis displays the top 20 countries based on session counts, total clickouts, and clickouts per session. Despite lower session counts in country codes TR and JP, their clickout rates are significantly high. Further investigation would be warranted to understand the comparatively lower clickout rates in other countries.

Relationships between Numerical Variables



The heatmap shows the correlation matrix between specific numerical variables in the dataset. It reveals varying degrees of correlation among the variables. Specifically:

- Clickouts and Session Duration display a moderate positive correlation of **0.46**, indicating a tendency for longer sessions to have more clickouts.
- Session Duration and Total_ctp exhibit a moderate positive correlation of **0.27**, suggesting a not so strong connection between session length and the total content viewed.
- Clickouts and Total_ctp demonstrate a moderate positive correlation of **0.19**, hinting at a relationship between the number of clickouts and the total content viewed.
- Bookings have a relatively weaker correlation with the other variables, showing low correlations ranging from **0.03** to **0.13** with Clickouts, Session Duration, and Total_ctp.

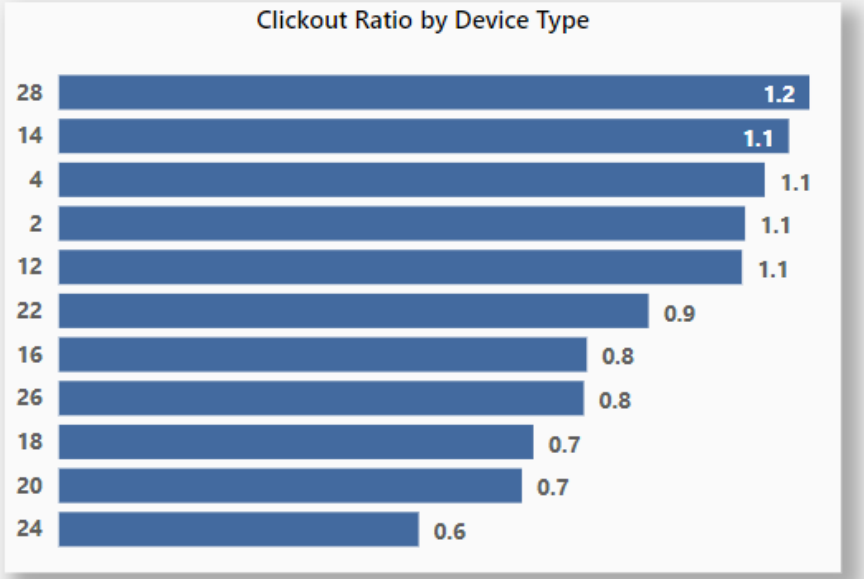
Observations

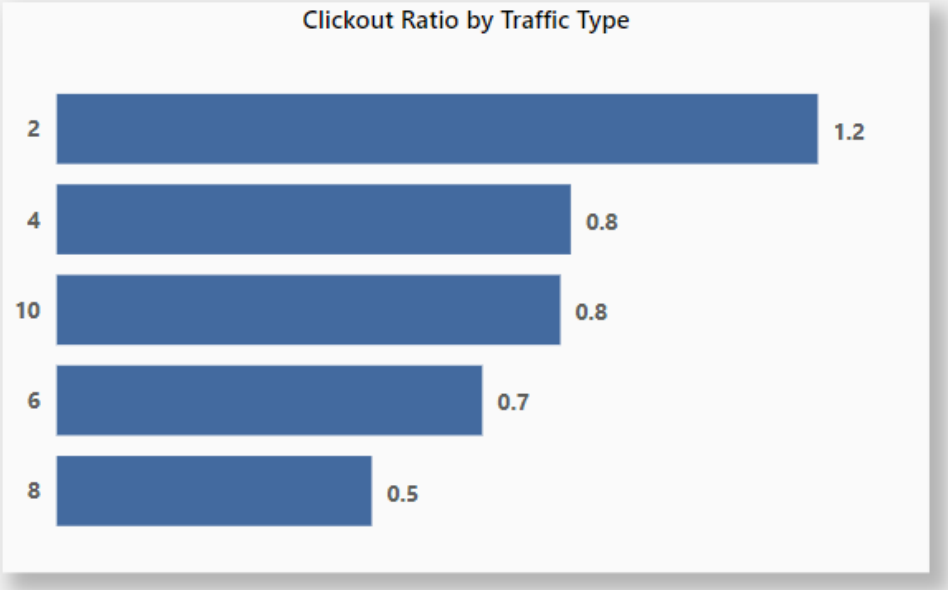
- The data had formatting issues(csv structure in a .xlsx file), particularly in the 'country_name' column with multiple commas, necessitating extensive cleaning to ensure data accuracy and integrity across all columns.
- Consider prioritizing the development of a mobile app (seeing as no session was performed on a mobile app) to improve user engagement on a more user-friendly platform. Additionally, if a mobile app already exists, encourage users to utilize it for a better experience.
- With under **40%** of sessions leading to at least a clickout and a minute average session durations, users may struggle to find available accommodations. Considering an AI chat feature for personalized recommendations or boosting listings on platforms with higher sessions but fewer clickouts could be a viable strategy.
- Certain traffic/marketing channels exhibit higher efficiency in directing users more prone to making clickouts compared to others, despite having similar or lower session counts. Allocating increased advertising revenues to these high-performing traffic channels is recommended.
- Investigating the gradual decline in sessions since the product's launch on May 1st is essential. Strategies to encourage user revisits and focus on effective marketing channels directing users to the product may be beneficial.
- Enhancing user engagement through interactive features could boost clickouts, considering the moderate positive correlation between clickouts and session duration.

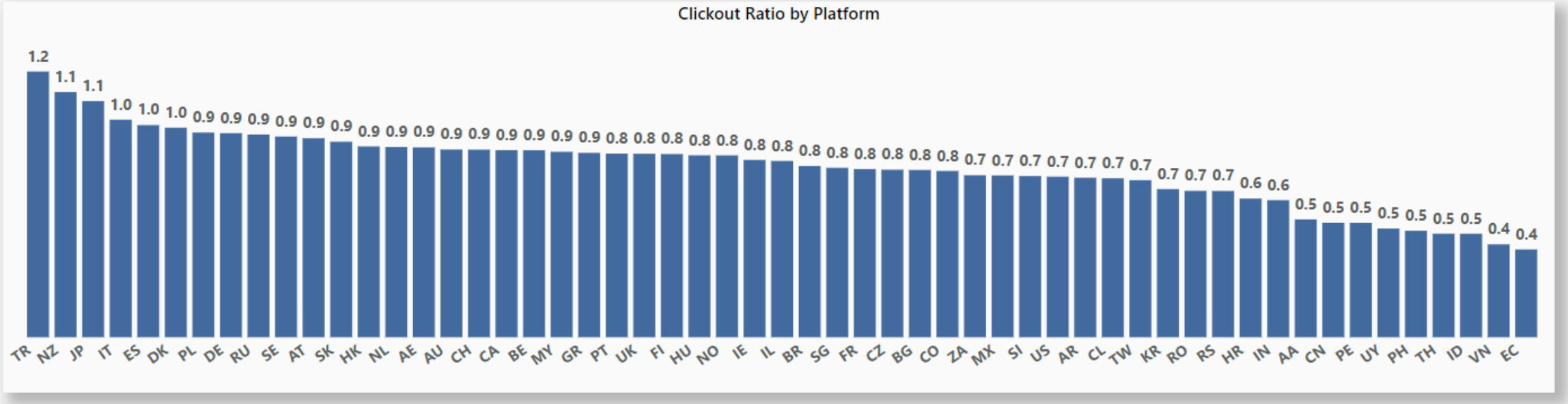
Important Metrics

- Clickout Ratio
- Repeat Visitor Rate
- Time series analysis of engagement metrics such as average contents viewed per session with time
- Traffic source performance for the various metrics above
- Platform performance for the various metrics above
- Agent id (device type) performance for the various metrics above
- Page entry and exit rates (for user actions, to see which pages/actions are prone to drop offs)

Challenge 2 – Clickout Ratio (COR)







Observations

- Platform with the highest COR – **Platform TR**

- Device with the lowest COR – **Device 24**

- Are there differences by traffic type – Yes , there are significant differences, with traffic type **2** having about **50%** more COR over any other traffic type.

- Upon examining session counts in page 4 and clickout ratios (COR) among different traffic types, it's evident that although traffic type **6** generates the highest number of sessions, it exhibits a significantly lower COR of **0.7**. Conversely, despite having the fewest sessions, traffic type **10** demonstrates commendable COR performance. I recommend directing marketing efforts and allocating ad revenues toward traffic channels such as **2**, **10**, and **4**, which yield sessions resulting in higher COR, fostering better performance and generating higher revenue.

Part 2 – Additional KPIs

1. **Repeat visitor rate** – This metric measures user satisfaction by tracking how often previous users return to the site. It provides insights into user enjoyment and helps differentiate actions taken by returning users from those who don't revisit.

   Repeat Visitor Rate = (Number of Repeat Visitors / Total Number of Visitors) * 100

   Where:

   - Number of Repeat Visitors: count of unique visitors who have visited more than once.
   - Total Number of Visitors: The total count of unique visitors.

2. **Average Session Duration** – As identified during descriptive analysis, this correlates with clickouts, which is a vital metric for us. This metric signifies the average session length on the platform. It can also be used to note if various platforms don’t find listings suitable to their regions and they tend to drop off early.

   Average Session Duration = Total Duration of Sessions / Number of Sessions

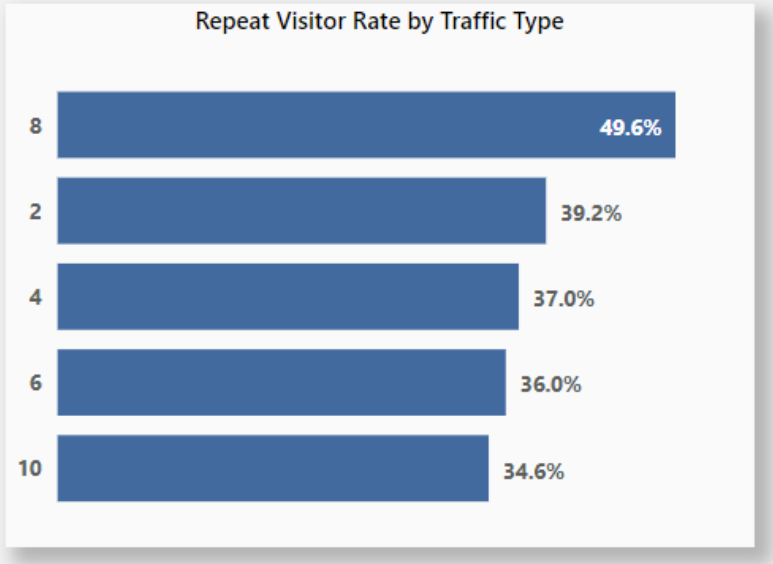
Repeat Visitor Rate & Avg Session Durations

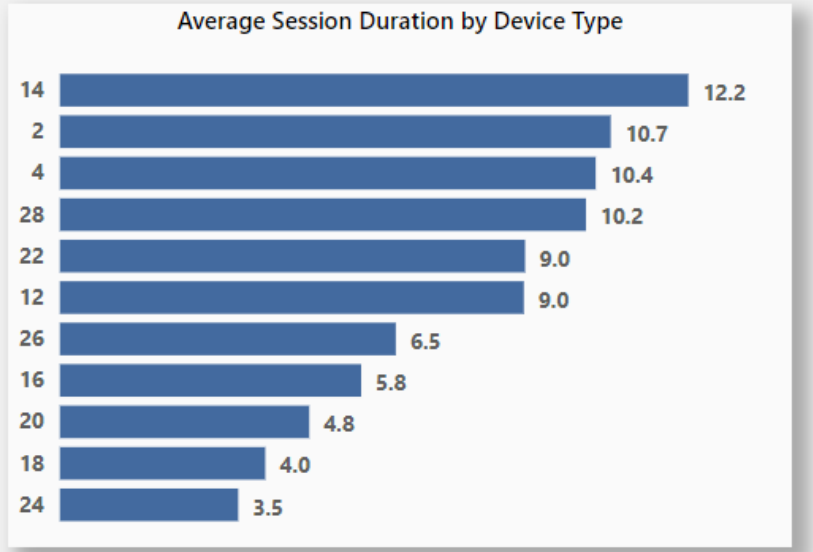
Repeat Visitor Rate

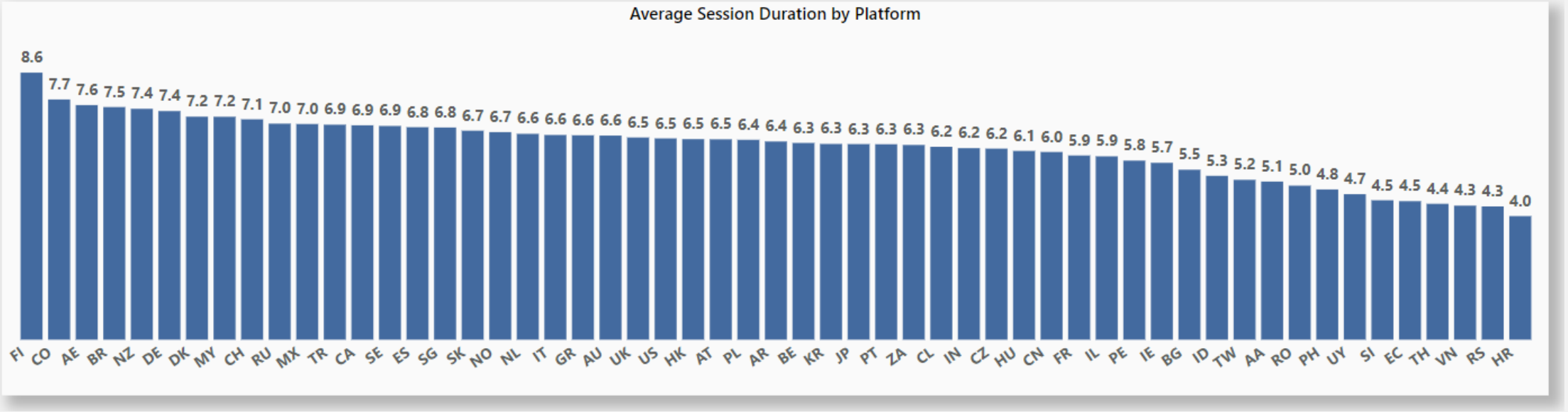
40.1%

Average Session Duration (mins)

6.5







Observations

- Only about 40% of our users revisit our site after initial visit.
- Our user retention rate has shown a gradual decline since our inception.
- These observations suggest a need for the team to prioritize user retention strategies, including enhancing user experience, personalizing content for users, and potentially introducing loyalty programs.
- Upon analyzing the traffic channels directing users to our site, it's evident that Channel **8** consistently attracts repeat users even with a bad clickout ratio . But Channels **2** and **4** not only drive users with a high inclination for clickouts but also tend to exhibit stronger user retention, they show viable strong traffic channel options.
- Users utilizing devices **18** and **24** tend to have a poor average session duration compared to devices like **14**, **2** and **4**. These devices need to be looked at and investigated further.
- Platforms such as FL, CO, and AE exhibit longer average session durations among users, contrasting with platforms like VN, RS, and HR. This trend highlights the varying user engagement across different platforms.
- The comparison between the platform-based average session duration chart and the initial analysis, where only around half of our sessions exceed a minute, indicates that within each platform, there are frequent extremely lengthy sessions, causing the average values to be higher. Investigating the actions of users in this category can offer insights into why they spend extended periods on the site. Understanding these behaviors could aid in enhancing the overall user experience to encourage similar lengthy sessions for other users.

Task 2

User Actions

Challenge 2 – User Actions (Part 1)

I tried exploring a typical user session on the Trivago website and observe the interactions of a typical user journey on the platform. Some of the ways I can see a session starting are highlighted below:

Acceptance of Cookies Prompt: This step involves consenting to the 'all cookies' pop-up, enabling Trivago to uniquely identify the user via a tracking ID. This action also provides crucial metadata, including device information and user location, enhancing the session experience.

"Sign Up/Log In" Functionality: This feature serves the purpose of user identification across multiple devices. This pivotal action aids Trivago in maintaining a seamless and personalized experience for the user, ensuring continuity, ease of access.

Engagement with the "ChatGPT AI" Interface: This AI-driven conversational feature serves as a valuable resource, providing users with personalized guidance, supplementing traditional search methodologies.

Utilization of the "Search Feature": The feature reflects users' intent to directly explore and discover optimal accommodation options. This feature streamlines the user journey by providing arrays of filter options to better aid the users' search.

These steps typically represent how I would expect sessions to start for users on the trivago website. Delving into what these actions could be from the data can be seen on the right.

Action ID	Unique Session Starts	%Unique Session Starts
2100	25615	71.71%
2113	3289	9.21%
2116	2933	8.21%
2111	1122	3.14%
2115	537	1.50%
2114	470	1.32%
2227	400	1.12%
2142	328	0.92%
2160	166	0.46%
2365	95	0.27%
2306	67	0.19%
2358	67	0.19%
2374	67	0.19%
2455	67	0.19%
2257	59	0.17%
2156	51	0.14%
8001	51	0.14%
2375	46	0.13%
2345	39	0.11%
Total	35722	100.00%

This analysis shows the predominant actions that initiate sessions on the Trivago website based on the user action data. Notably, action 2100 stands out significantly, constituting approximately 72% of all session starts, suggesting its prominence among the user interactions observed. It is presumed that these top actions align with the earlier mentioned start features integrated into the platform from my exploration.

Delving into the actions diverging from this prevalent pattern gives the product team a valuable opportunity to gain insights into user sessions that deviate from the norm. This could provide a pathway for the team to enhance and optimize user experiences by better understanding the nuances and needs of users engaging in different session initiation behaviors.

Challenge 2 – User Actions (Part 2)

Drop Rate for Action 1 = (Number of sessions where Action 1 is the last action / Total number of sessions where Action 1 occurs) * 100

S/N	Action ID	Drop Rate (%)
1	2506	92.31%
2	8101	80.00%
3	2374	79.17%
4	2120	78.57%
5	2149	66.67%
6	2371	52.91%
7	8001	51.70%
8	2395	50.00%
9	2472	33.33%
10	8006	28.57%
11	2286	27.97%
12	2186	27.78%
13	2476	27.27%
14	2116	25.51%
15	8091	25.00%
16	2364	24.17%
17	8002	23.97%
18	2155	23.85%
19	2385	22.58%
20	2227	22.42%
21	2142	20.43%
22	2113	20.21%
23	8020	19.05%
24	2459	18.18%
25	2465	18.00%
26	2111	17.83%
27	2375	16.84%
28	2114	16.65%
29	2464	16.33%
30	2262	14.78%
31	2319	14.29%
32	2380	14.29%
33	2458	14.29%
34	2296	13.73%

S/N	Action ID	Drop Rate (%)
35	2115	13.45%
36	2100	13.00%
37	2255	12.65%
38	2470	12.50%
39	2313	12.20%
40	2128	12.20%
41	2324	11.67%
42	2156	11.13%
43	2455	10.73%
44	2146	10.49%
45	2108	10.17%
46	2188	10.14%
47	2131	10.00%
48	2318	10.00%
49	2445	10.00%
50	2501	9.42%
51	2145	9.23%
52	2119	8.64%
53	2452	8.54%
54	2257	8.49%
55	2284	8.42%
56	2345	8.13%
57	2216	7.91%
58	2440	7.78%
59	2502	7.32%
60	2132	7.14%
61	2353	7.08%
62	2446	6.98%
63	2474	6.90%
64	2302	6.76%
65	2136	6.42%
66	2133	6.19%
67	2160	6.09%
68	2143	5.86%

S/N	Action ID	Drop Rate (%)
69	2451	5.81%
70	0	5.56%
71	2168	5.26%
72	2121	5.00%
73	2351	4.38%
74	2175	4.32%
75	2358	4.16%
76	2350	3.98%
77	2205	3.60%
78	2200	3.34%
79	2443	3.33%
80	2356	3.02%
81	2442	2.97%
82	2399	2.77%
83	2123	2.63%
84	2279	2.60%
85	2124	2.58%
86	2125	2.44%
87	2235	2.38%
88	2126	2.03%
89	2122	1.97%
90	2135	1.92%
91	2369	1.70%
92	2130	1.66%
93	2306	0.33%
94	2269	0.00%
95	2352	0.00%
96	2365	0.00%
97	2479	0.00%
98	2307	0.00%
99	2134	0.00%
100	2463	0.00%
101	2475	0.00%
102	2408	0.00%

Challenge 2 – User Actions (Part 3)

Considering potential actions that might be viewed as conversions while I was navigating the Trivago website, I've compiled the following list;

- **Click Throughs to partner sites** – Users clicking through to external partner sites to view deals, explore accommodations, or complete bookings. This action indicates a significant step towards potential hotel reservations.

- **User Account Sign Up** – Users registration or sign-up for personalized experiences on the website. Account creation fosters user engagement, offering features like saved preferences and easy access to previous searches.

- **Return Visits to the Site** – Users revisiting the Trivago website after their initial visit. Returning visits often signify sustained interest or ongoing engagement with accommodation options.

- **Voluntary Subscription to Notifications** – Opting-in for notifications such as travel recommendations, exclusive offers, or newsletters. This action demonstrates user interest in receiving tailored information for future travel plans.

- **Saving or Sharing Accommodation Options** – Users saving preferred accommodations to favorites or sharing them. This action indicates a higher level of user engagement and intent to consider specific options for potential bookings.

- **Engagement with Trivago's AI Travel Assistant or Customer Support** – Interacting with AI-driven travel assistance or customer support features. Engagements with these services signify users seeking guidance or assistance in their accommodation search, potentially guiding them towards booking decisions.

Challenge 2 – User Actions (Part 3)

Upon reviewing the available data, certain fields within the sessions dataset appear to align with identifying potential conversions as mentioned earlier:

- **Clickouts Column in the Sessions Dataset** – This field helps to find sessions that resulted in click-throughs redirecting users to partner sites. It aids in understanding user behaviors that potentially lead to external bookings.

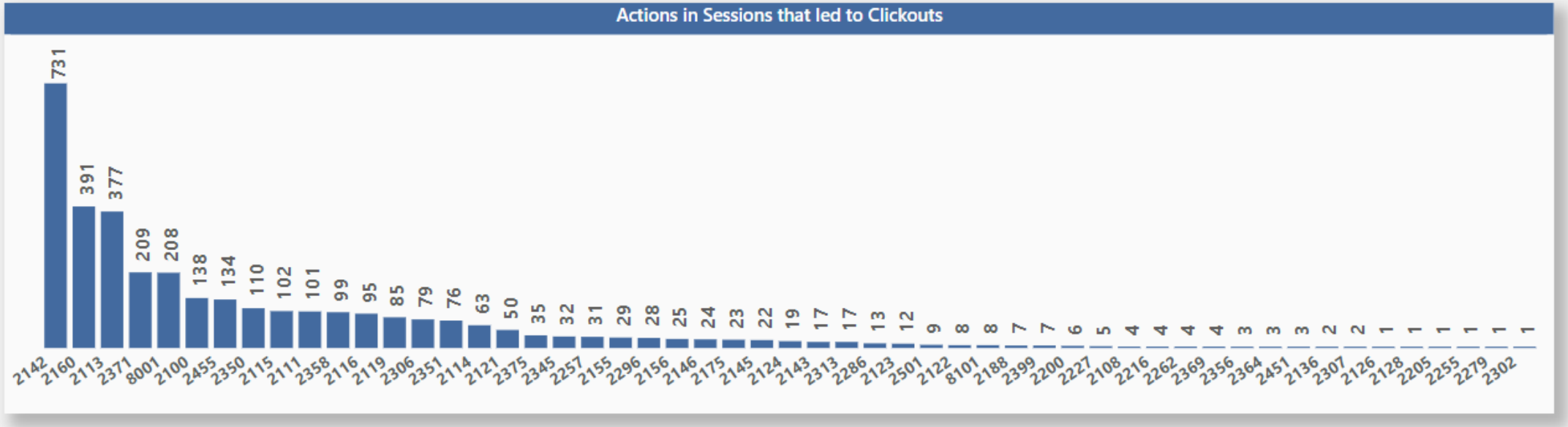
- **Is_Repeater Column in the Sessions Dataset** – This field assists in identifying users who revisit the Trivago site, indicating sustained engagement or interest over multiple sessions.

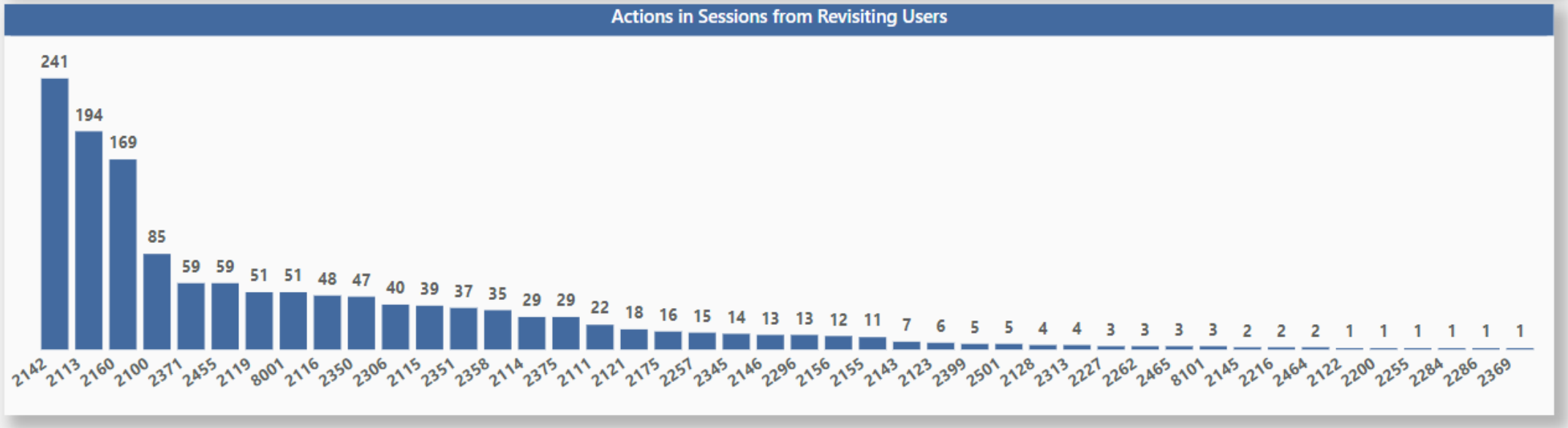
Unfortunately, the datasets either encodes or do not contain sufficient information to capture the other potential conversions mentioned earlier, thereby limiting the ability to assess those specific conversion actions based solely on the available data.

- Steps used to determine the actions to higher numbers of conversions for the cases above;

S/N	Clickouts Column in the Sessions Dataset	Is_Repeater Column in the Sessions Dataset
1	Find all sessions that led to at least a clickout in the sessions data and filter the dataset down to just those sessions	Find all sessions that had the is_repeater flag as 1 in the sessions data and filter the dataset down to just those sessions
2	Join the filtered dataset to the actions dataset to include only actions in the sessions from the filtered above.	Join the filtered dataset to the actions dataset to include only actions in the sessions from the filtered above.
3	Count the actions across the sessions to see which actions are prominent to sessions that eventually lead to clickouts	Count the actions across the sessions to see which actions are prominent to sessions that eventually lead to a user revisiting the site

Challenge 2 – User Actions (Part 3)





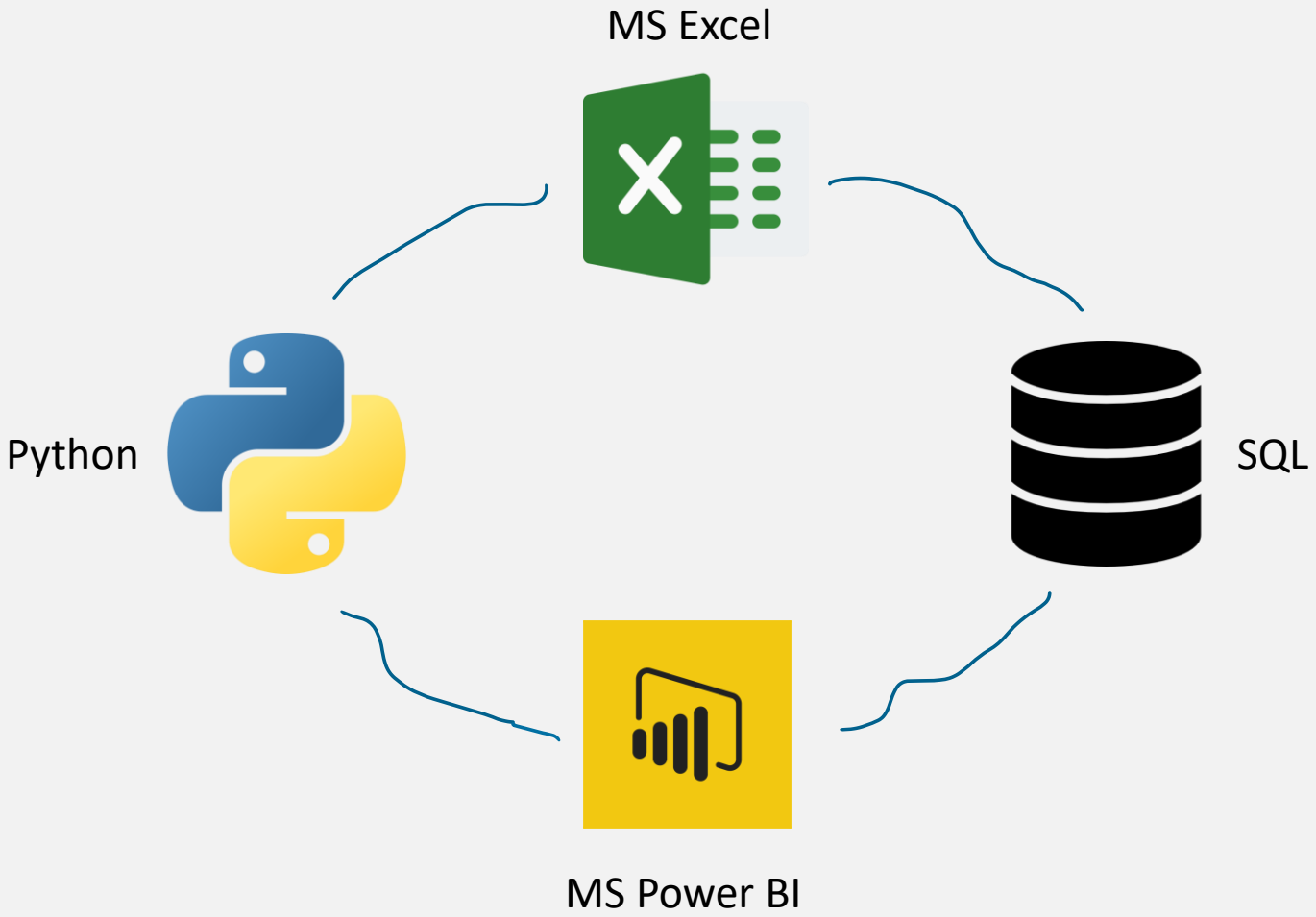
Challenge 2 – User Actions (Part 3)

Upon close examination of both charts that shows a depiction of action distributions among users in sessions resulting in clickouts and those revisiting the site, it becomes evident that notable similarities exist. The actions correlating with higher numbers of conversions exhibit striking similarities across both scenarios and prominently stand out in the charts. The top six actions identified are as follows:

- **2142**
- **2160**
- **2113**
- **2371**
- **8001**
- **2001**

These actions demonstrate consistency in their occurrence among sessions leading to clickouts and users revisiting the site, indicating their potential significance in contributing to conversions in both contexts.

Tools Used while working on this Case Study



trivago

THANK YOU

By Oluwatobi Ojo