User Engagement Analysis - Data Analyst @ Showwcase

So the task here is to better understand "user engagement" on the platform in the October month to define engagement, find insights, and question any other metrics we can use to understand our approach better.

Tool used: Microsoft Excel – The reason to use this tool is because of its functionalities which help in giving deep insights of the data.

Additional Columns added:

1. Active_duration(column O) which is session_duration(N)-Inactive_duration(L)
2. P,Q,R,S,T are the columns which are the binary values for Columns D,E,F,G,H because values given are TRUE and False which are difficult to be find insights. So the values are converted to binary 1 for TRUE and 0 for FALSE.
3. Column U is the day part of the date
4. V is sessionId rank , by calculating using RANK.AVG(val1,val2) . This gives the rank within the column.

1)

Goal:

To find the relationship between session duration and the bug_occured.

Data Column Used:Excel1

Session_duration(N), Valuebug(P)

Method Used:

The variable types given are all different.
So the binary values can be considered categorical duration may be considered continuous.
Normal Correlation formula cannot be used for these variables.
So point biserial correlation is used - it is simply the correlation between one dichotmous variable and one continuous variable. It turns out that this is a special case of the Pearson correlation.
The following function is provided in the Real Statistics Resource Pack.

Results:

The result was p value with 0.65 with point biserial value as -0.03, which means there is large concentration around centre that is 0.5. The tests are more reliable when they are around the centre.
The results showed more the number of duration more will be the bugs.

Suggestions:

The product team has to get a note as to why the bugs are getting generated. The team can get the feedback from the customers and know after how long the bugs are generated.

2)

Goal:

To find any trend between sessionid and the day

Data Column Used:Excel1

day(U), sessionid rank(V)

Method Used:

The variable types given are continuous
So spearman method is used where in ranks of the columns are calculated and the code is applied.

Results: The result was value with 0.076 value which defines some close relationship between sessions and the day

3)

Goal:

To find trend with customer and the bug

Data Column Used:Excel2

AA to AF columns , the ones where duplicates are being removed from the columns A to F

Method Used:

Chart1

Results and suggestions:

There are 49 customers in the month of October with different session id

In the Chart 1 plotted, following trends are observed:

1 It shows number of bugs, projects, likes and comments combined for each session and customer.
2 Its seen that the blue line shows there were many bugs detected for almost all sessions. The product team has to look into it. The metrics that must be added to rectify this issue is to know if the bugs were reported by the customer and rectified by the product team. Otherwise the bugs may come up more often.
3 The customerId 29375 has accessed 22 times with at least 18 bugs .
4 The trend seems to be the more number of times website accessed, more bugs detected.
5 Next observation is when a customer is active atleast 1 like will be given by that customer.
6 Same is observed for project.
7 There may be customers who may be working on the similar projects. There should be an addition of another variable called share, which will help them to share with friends or someone working on similar lines.

4)

Goal:

To find trend with activities in the October month

Data Column Used:Excel2

AK to AR columns, the ones where duplicates are being removed from the columns P to W

Method Used:

Chart2, showing the 30 day trend

Results and suggestions:

In the Chart2 plotted, following trends are observed

1. Shows the day, session_duration, inactive_duration, session_projects, session_comments, session_likes,bugs_in_ session, active.
2. Here it is observed that though the session duration is high, the inactive status is also proportional. The product team needs to get feedback on why the active duration is less. Have the users just logged in and doing other work or other issues.
3. On 26th October there was a sudden increase in session_duration , and the next day 27th October there was very less traffic on the website. The product team need to analyse what made the sudden change

Code used:
Point Biserial Correlation- BCORREL(R1, R2) = the biserial correlation coefficient corresponding to the data in column ranges R1 and R2, where R1 is assumed to contain only 0's and 1's

*Can be written in Python pandas too