

# USER ENGAGEMENT ANALYSIS

-Manikandan Lalitprasad

## **I. Background:**

- **Showwcase** is an online platform where members of the tech community connect, share their knowledge, work experiences, projects built, people they have worked with - and find new opportunities.
- Showwcase users can like, follow, comment, and share posts on the platform. The company's goal is to build a platform that encourages users to share, exchange ideas, connect, and learn from other users.
- Every time a user logs into their Showwcase account, a new session begins. In each session, we track how and where a user is engaged on the platform.

## **II. Analysis Goals:**

- To identify user behavior patterns from the website and identify specific users that are more active than others – These customers can be strategically encouraged to upgrade to the PRO version of Showwcase, in turn increasing revenue.
- To categorize the different types of user interaction with the website and check for correlations between the actions – This knowledge can be used to create more product feature offerings that are engaging and increase product usage.
- To identify users that are not active users and with a higher risk of churn – This information would allow Showwcase to address their needs specifically, while additionally adopting general mechanisms that improve customer experience and potentially prevent attrition.

## **III. Assumptions:**

- The data is up-to-date, and the time of analysis is November 1<sup>st</sup>, 2019.
- "session\_duration" is the active time spent on the website, while "inactive\_duration" is a mutually exclusive time spent on the website where the user doesn't perform any actions.

## **IV. Data Preparation:**

- Added a new column "Total\_session\_time" = "Session Duration + Inactive Duration.
- Calculated a new column "Active\_ratio" = "Session Duration / Total Session Time.
- Grouped the data on customer level to view all the fields particular to each customer.
- Extracted an additional feature "day\_of\_month" from "login\_date" to identify patterns in specific days where customers are active.

- Extracted an additional feature “day\_of\_week” from “login\_date” to identify the specific day in which customers are active.
- Converted the session duration to minutes and added as a new column “Session\_minutes”.

## V. Exploratory Data Analysis:

At the outset, we inferred the following:

- **48** unique customers visited the website in the month of October in 2019.
- The total number of projects added is **486**, the number of likes given is **3127**, the number of comments given is **722**. From the given data, this quantifies the measurable activity on the website for October 2019.
- The total session duration is **5933.82** minutes and the total inactive duration / idle time is **3664.67** mins for all customers
- The total bugs the customers faced is **370**. We have no indication of the criticality of the bugs.

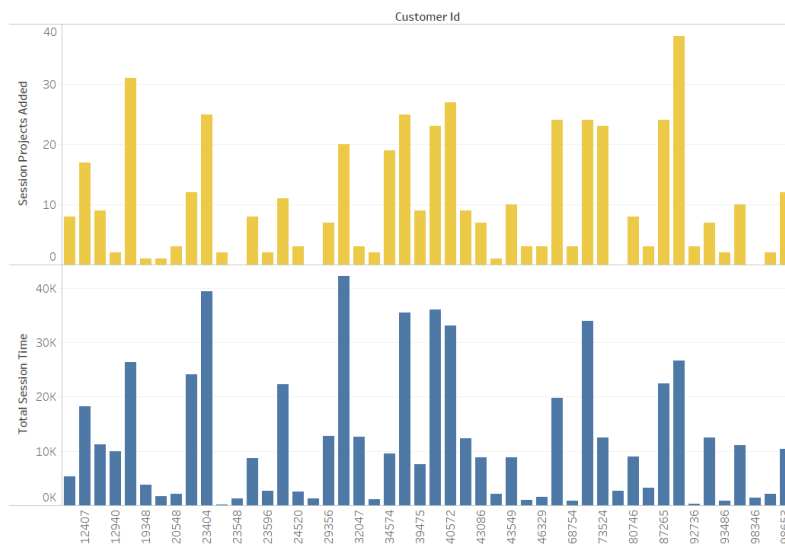
The average session duration spent per user is **123.61** minutes in which the average inactive time is **76.34** minutes. Hence the customers actively spend an average of **62%** actively on the website. This indicates the vitality of the website. We should therefore focus on increasing the active session time which in turn decides the overall engagement on the website. Since a customer encounters an average of **7** bugs per session, it could be a potential reason for high inactivity.

The date **October 26, 2019** experienced the highest web traffic of **22** unique customers, where **31** projects were added. The highest session duration on this day was **457.31 minutes**, which also had the maximum bugs of **43**, and an inactive duration of **335.61 minutes**. Moreover, 247 likes were given.

The maximum time spent by a unique customer is **505.33 minutes**. However, this customer did not create too many projects (not even top 10), but instead was very active socially with respect to likes (189 in number, ranking 2<sup>nd</sup> highest) and comments (44 in number, ranking 3<sup>rd</sup> highest). On the other hand, the least time spent by a customer is only **1.43** minutes, who gave **5** likes.

It can therefore be inferred that the customers may access the website for various reasons. Some may want to primarily learn from other user's projects and would have therefore given more likes and comments in appreciation, while others are more interested in contributing and showcasing many projects. However, overall, we do see a significant correlation between projects and social activity, with customers who are socially active also tend to publish projects later.

Projects and Session Duration



The adjoining graph plots session time with projects contributed during that session. We see a high correlation between the two, implying that customers spend more time during session where they are also publishing a project, as opposed to the sessions where they are not publishing any project.  
X-Axis: Customer\_ID  
Y-Axis: Session Projects Added & Total Session Time

Login Frequency Days vs Bugs

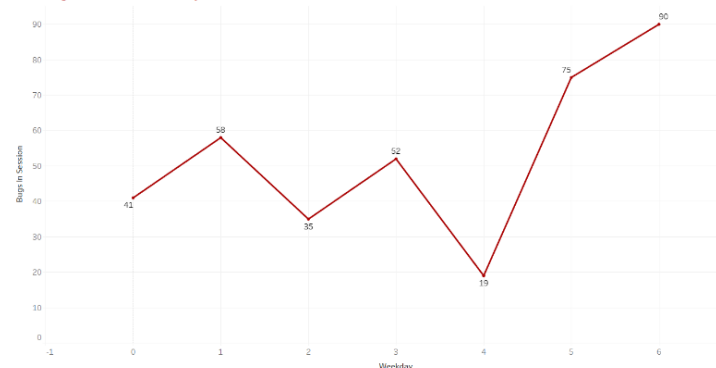


From the adjoining visualization, we find that the customers faced more bugs when they had login multiple times.

It is therefore important to provide bug-free user Interface to ensure satisfactory customer participation.

X-Axis: Customer ID  
Y-Axis: Login Frequency Days & Bugs in Session

Total Bugs occurred in each day



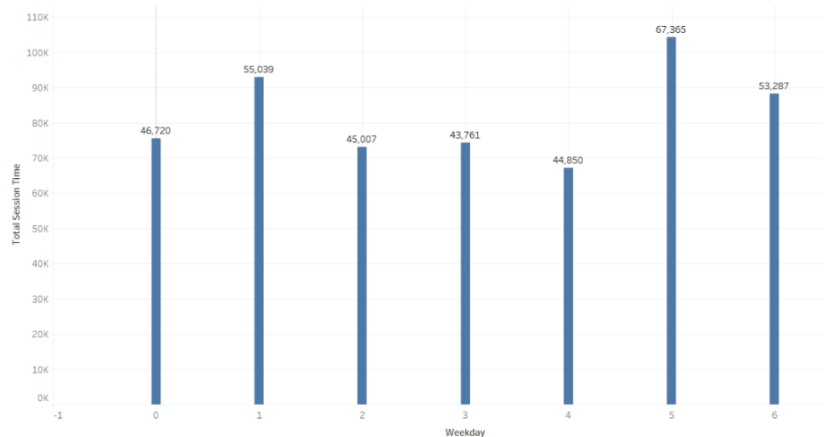
The adjacent figure shows that twice the number of bugs is encountered during weekends (20 bugs on average) when compared to weekdays (10 bugs on average)

X-Axis: Day of the week  
Y-Axis: Total Bugs

The adjacent graph shows that on an average, sessions last for 251 minutes on weekends, when compared to the 196 average minutes during weekdays. Hence, we see a higher margin of customers accessing the websites during the weekends than during the weekdays.

X- Axis: Days in a week (0-6: Monday-Sunday)  
Y- Axis: Total Session Time

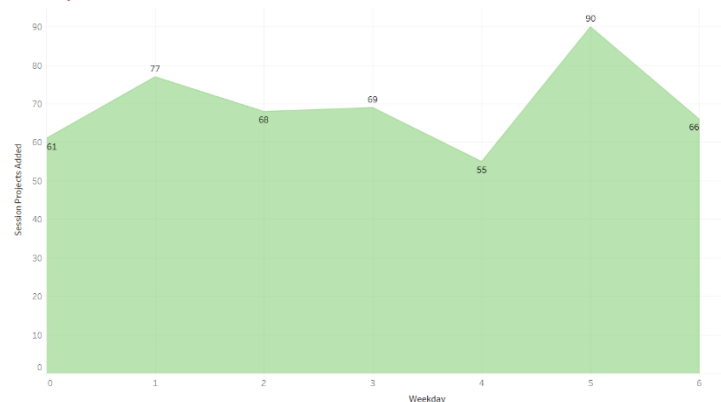
Total Session Duration in a week



The average number of projects added during weekends is 20 which is higher than the average number of projects added during weekdays (16 in number)

X-Axis: Days in a week  
Y-Axis: Total Projects Added

Total Projects Added in a Week



## VI. Inferential Analysis:

### (a) Customer Value Score (Modified Approach of RFM):

Recency, frequency, monetary (RFM) value is a marketing analysis tool used to identify an organization's best customers. The RFM model is calculated as the product of three quantitative factors:

Recency: How recently a customer has made a purchase

Frequency: How often a customer makes a purchase

Monetary Value: How much money a customer spends on purchases

RFM analysis numerically ranks a customer in each of these three categories, on a scale. (the higher the number, the better the result). The "best" customer would receive a top score in every category.

For this analysis, we have modified and applied this technique to identify our best customers and accordingly make recommendations.

**(b) Algorithm:**

- Net Value of Customer (RFM Score) = Projects(scaled) + Likes(scaled) + Comments(scaled) + Login Frequency(scaled) + Recent Login Date(scaled) + Session Minutes(scaled)
- Data is scaled by dividing each of the data points of a given field with the maximum value in that field
- The “most recent login scaled” column is assigned 0.5 if the user's final visit for the month is in the first ten days, and 1.0 if the user visits anytime after that. This particular field is an indicator for the propensity to churn, because only loyal and engaged customers continue to frequent the website.

**VII. Insights:**

- From the analysis, we have identified that the number of project contributions and frequency of usage are the biggest contributors to what makes a customer high-value, and differentiable from the others. The remaining factors such that the ratio of active time-spent and social activity are multi-collinearly related to these primary features.
- The recency column, while highly useful in detecting risk of churn when the dataset is large (spanning over many months and years), in this case can only be used as an indicator of churn since the dataset is only spans for one month. Therefore, the recency factor has been intentionally designed to not overly impact the customer value score.
- Inactive time is not a big indicator of a less valuable customer, as even
- There is also immense scope for bug fixing and product feature enhancements, as the number of bugs only increase with prolonged usage.

**VIII. Recommendations:**

- The top 20 users that have been ranked in the analysis can be offered the PRO version of Showwcase, as they really like the website and have demonstrated high interaction in a multitude of ways.
- To address the risk of churning by the bottom 10 ranked users, Showwcase can offer more automated support in terms of how to use the platform better by surfing more example projects, creating “Help” buttons that can solve any confusion they face while creating projects
- The identified customers with 0.5 score on Latest Login Date can be sent promotional emails to attract them into logging onto the website again.

## **XI. Future Scope:**

- With more data access for a larger time frame, a more accurate and precise analysis can be performed.
- With information on the type of bugs on the website, the criticality of it and its root cause can also be analyzed.
- Given the social activity of which customer liked and commented on which other user's projects, graphical mappings can be performed and a social network can be built. Based on this, better project recommendations can be sent to users.
- Using Google Analytics, advertising campaigns can be made better, by knowing the source from where users joined the website.