Exploratory Analysis of WOWAH Dataset

Yuming Fang

Agenda

1. Why this analysis?
2. Methodology
3. Summary of the data
4. User segmentation
5. Predicting how long a player will stay in the game

Motivation of the Analysis

- Predicting players' behavior pattern is crucial for game design and revenue growth
- User segmentation is important for improving user experience

Methodology

Avatar session data

Text file

Is valid

Yes → Clean data → PostgreSQL

No → Bad rows data (skip)
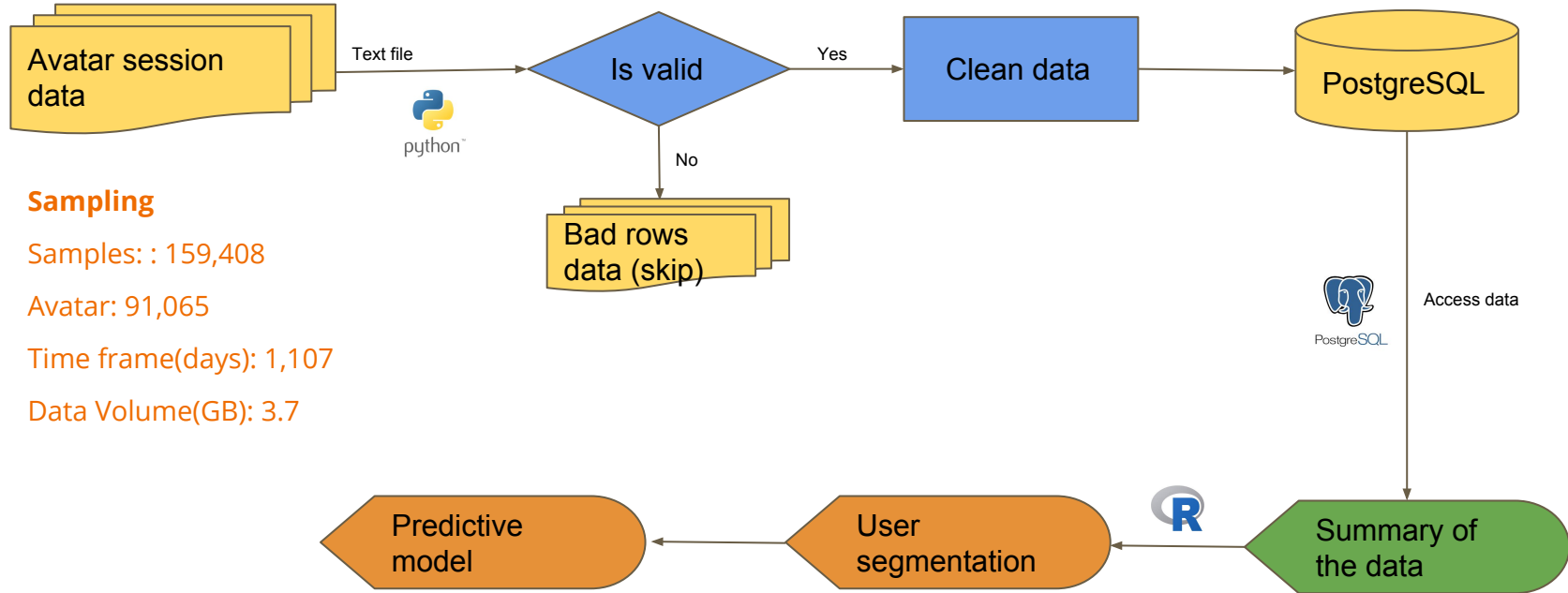
Sampling

Samples: : 159,408

Avatar: 91,065

Time frame(days): 1,107

Data Volume(GB): 3.7

Access data

Summary of the data ← User segmentation ← Predictive model

Summary of the Data

- The second expansion have an immediate impact on user growth
- Assumed causes of spikes: winter break, summer break,etc.
- Recommendation:
  - When to release expansion
  - What to do for spikes-decrease/Increase

Number of Avatar by Date

1st Expansion Released(2007-1)

2nd Expansion Released (2008-11)
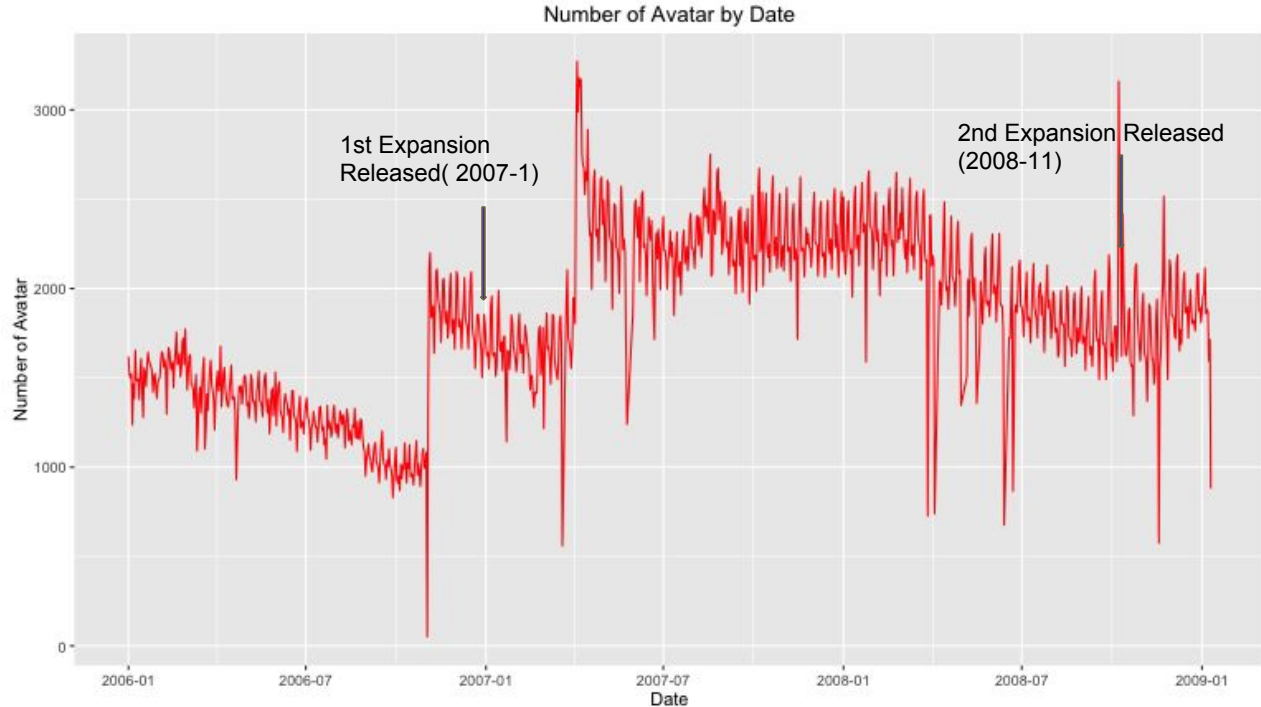
Analysis by Number of Avatar and Game Level

Findings:
Level 1-6: 2000-6000
Level 6-21: 1000-2000
Level 21-71:500-1000
Level 71: Approx. <250

Recommendation:
- Sharp decrease from Level 1 to Level 6, therefore needs improvement on user acquisition - more fun, easier for progression.
- Another sharp decrease from level 70 to 71, therefore needs to release expansion package sooner before users lost interest to the game
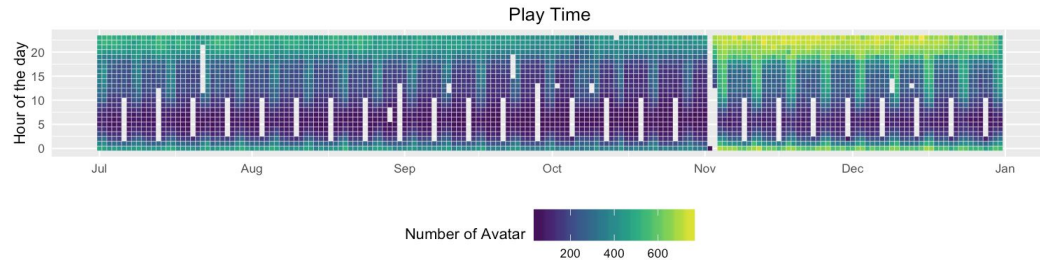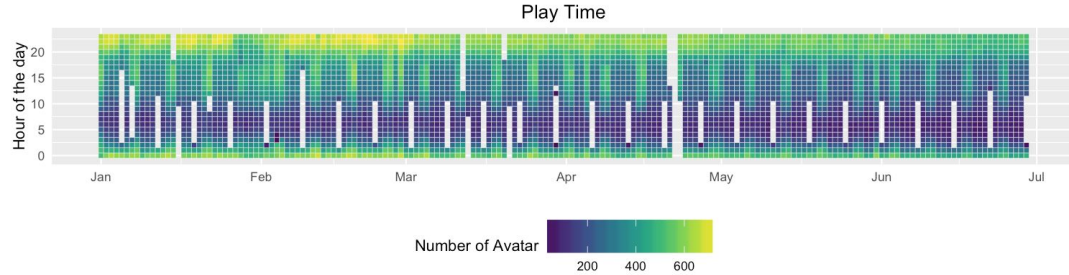

Number of Avatar by Game Level
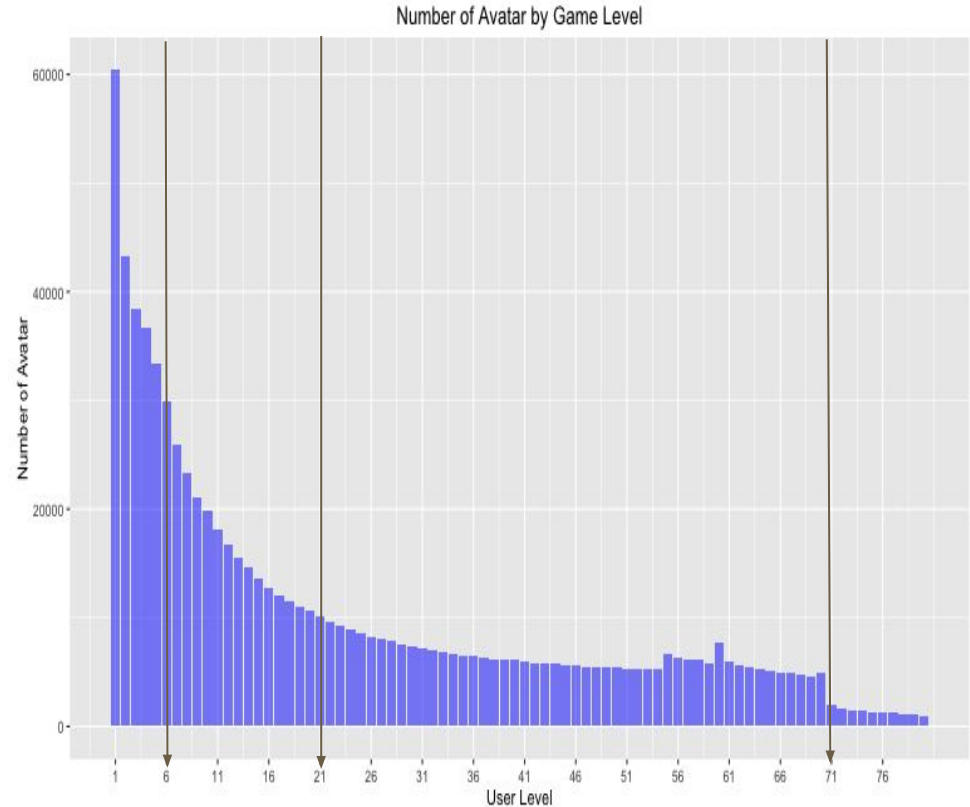
Analysis of Retention Rate

- There are 45% of users remain playing the game in day 1 since they registered.
- After one year, there are 10% of the users still playing the game

Retention Curve
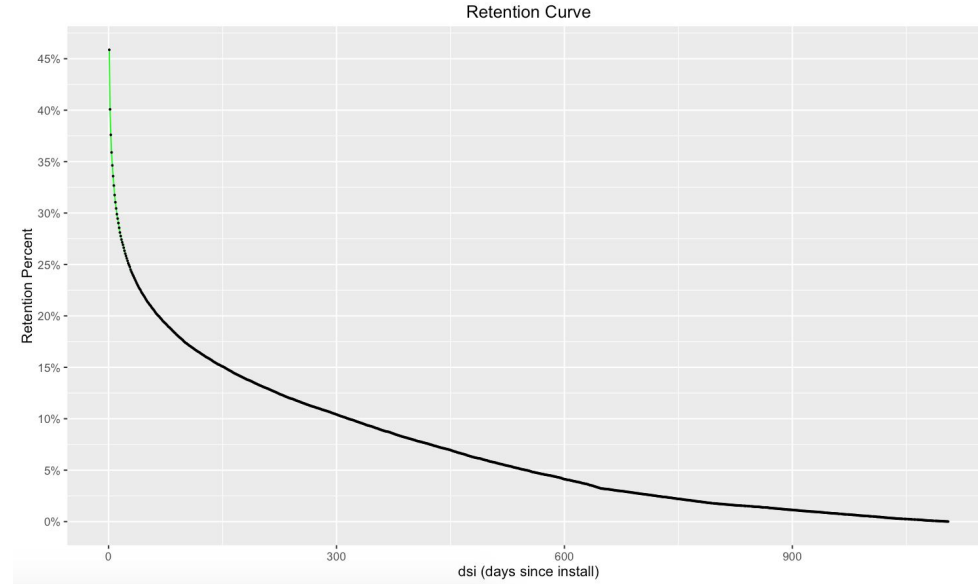
Analysis of Play Time

- There are more users playing the game at night time
- More users are observed between November and March

Additional Findings

Class Warrior is the most popular character

Blood elf is most popular race

Zone Durotar has the most number of avatars

User Daily Gaming Time Analysis

- Sample Qualification: The same avatar(user) shows up in each consecutive sample by 5-10 mins interval indicating that the user is playing the game during this period qualifies to be a sample. Avatars that have daily gaming time less than 10mins were excluded.
- Method of Calculation: Using the total amount of the in session time over by the number of days that avatar was observed.
- Recommendation: Conduct user segmentation by identifying hardcore players and casual players.

Density Plot of Average Daily Gaming Time
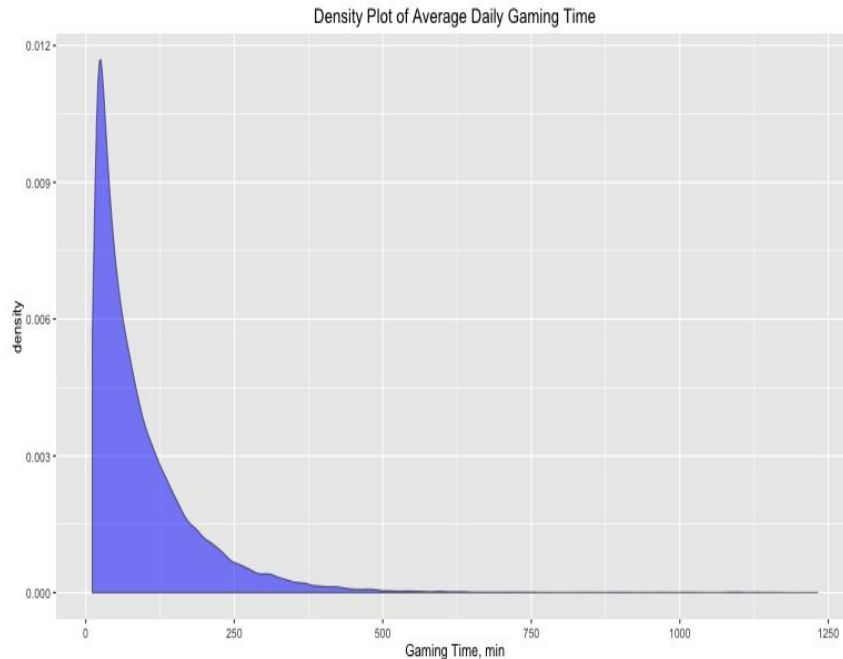
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
11.00	30.00	64.00	97.07	127.00	1232.00

Why User Segmentation ?

Hard Core Player **VS** Casual Player

Separate the users and feed them customized recommendations based on their needs can significantly improve their gaming experience which in turns, **increase users' engagement and business revenue.**

User Segmentation Process

Step 1: Feature generation:
- Use avatars detected between 2006-01 to 2008-03, resulting in 28,444 unique avatars
- Divide each avatar's history into 9 periods, each period has 3 months of data
- Calculate average daily play time and playing density in each period
- Use these features to group users into clusters

Step 2: Use k-means to group users into 5 clusters

Step 3:Applied t-SNE to reduce the user space into 2 dimensional to visualize it

User Segmentation Process

- Cluster 1: Not interested in the game(10,704)*
- Cluster 2: Starts moderate high gaming time, lose interest over time and not interested in the game anymore(1,576)*
- Cluster 3: Hard core player at the beginning, lose interest over time but still keep moderate high gaming time(425)*
- Cluster 4: starts with low gaming time, getting more interested over periods and having a moderate high gaming time (1,741)*
- Cluster 5: starts with low gaming time, getting more interested over periods and being hard core player for short period of time, then lose interest over time but still keep moderate high gaming time(1,501)*
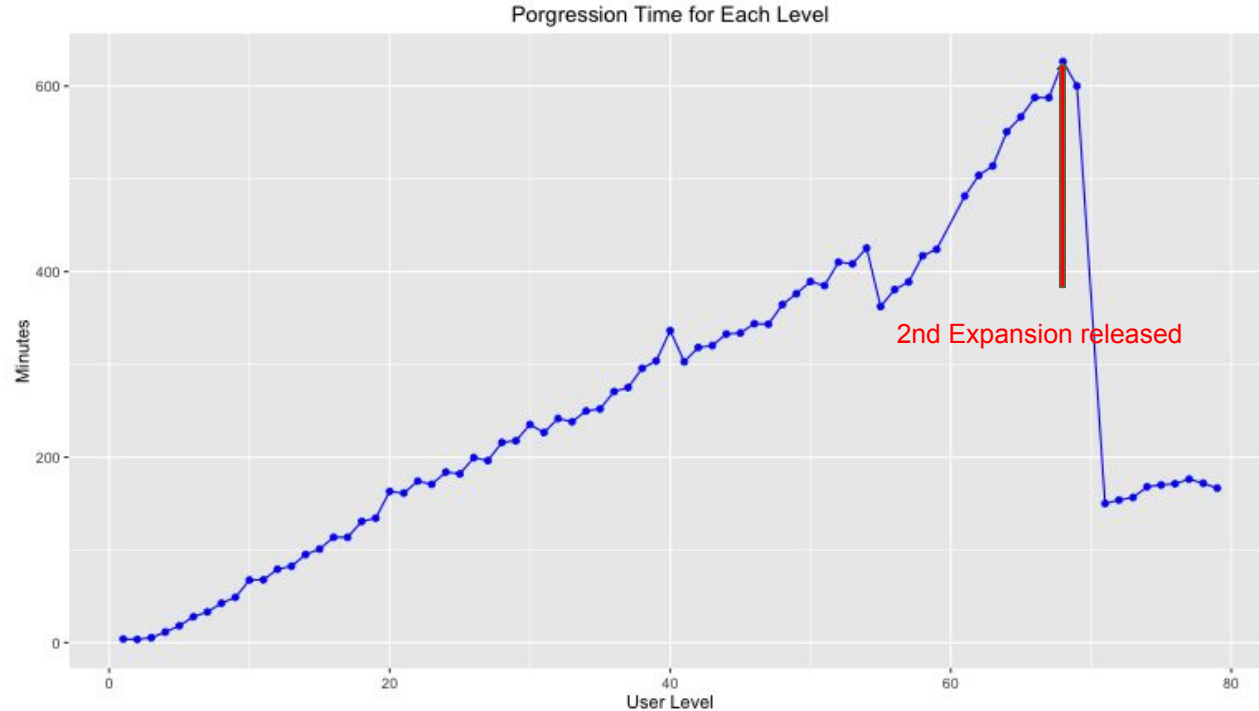
*Number of avatars observed
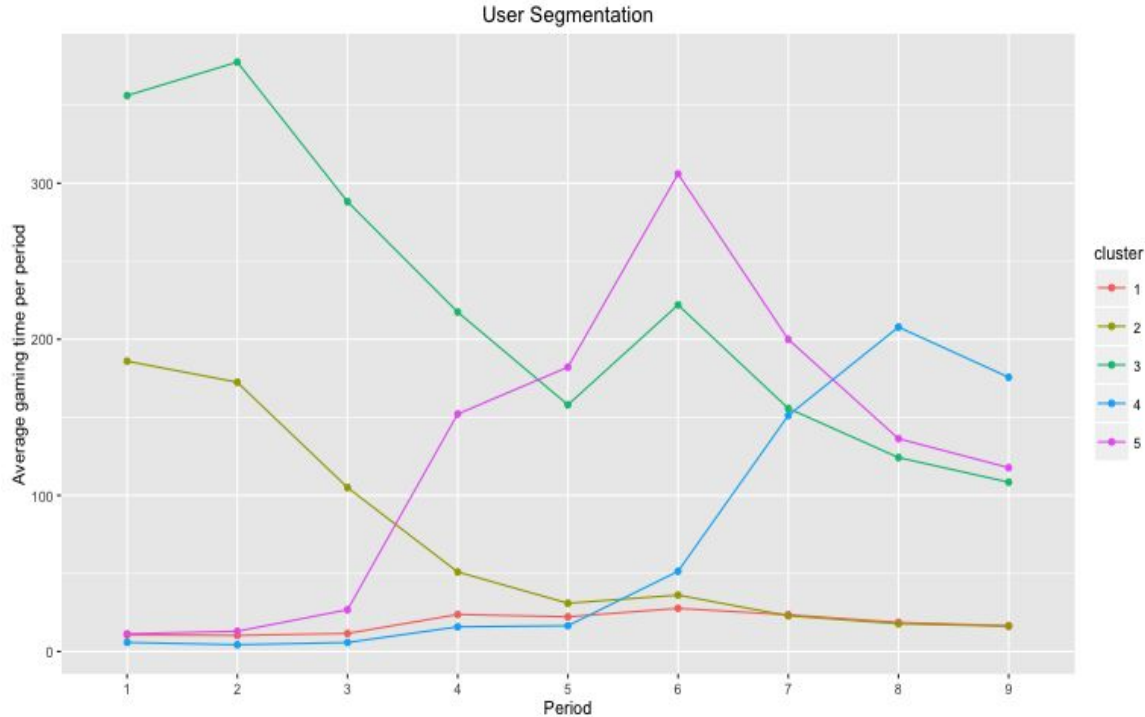
t-SNE 2D Segmentation of WoW User
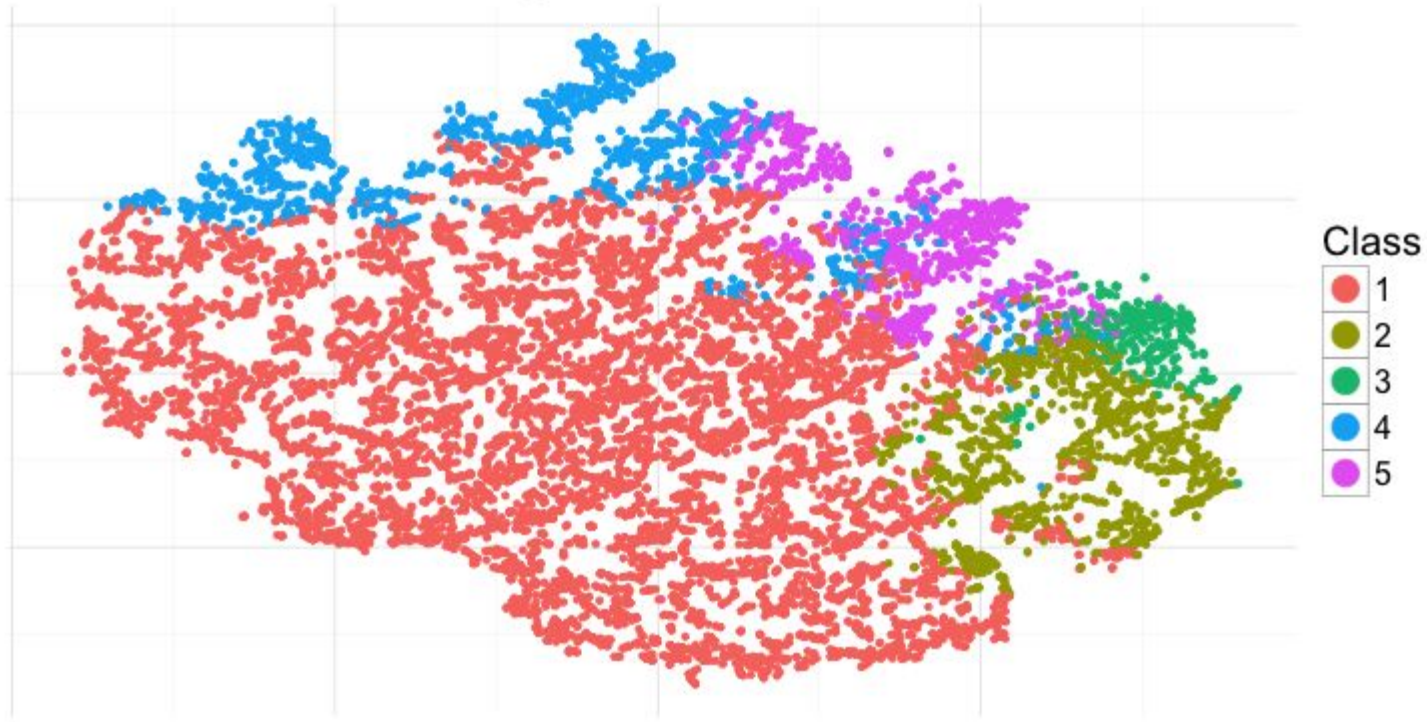
User progression speed analysis

- The difficulty of levels are positively linear correlated with the amount of time spent
- Level 70-80 has unusual low progression speed

Porgression Time for Each Level

2nd Expansion released

Guild affects engagement

- Guild: group interaction
- Two sample T test shows there is significant difference between the means of users <u>with</u> guild and users <u>without</u> guild
- Users with no guild have an average of daily gaming time about 98 minutes; users do have guild have an average of 164 minutes.
- Recommendation: Encourage users using guild

Effect of Guild on Average Daily Gaming Time
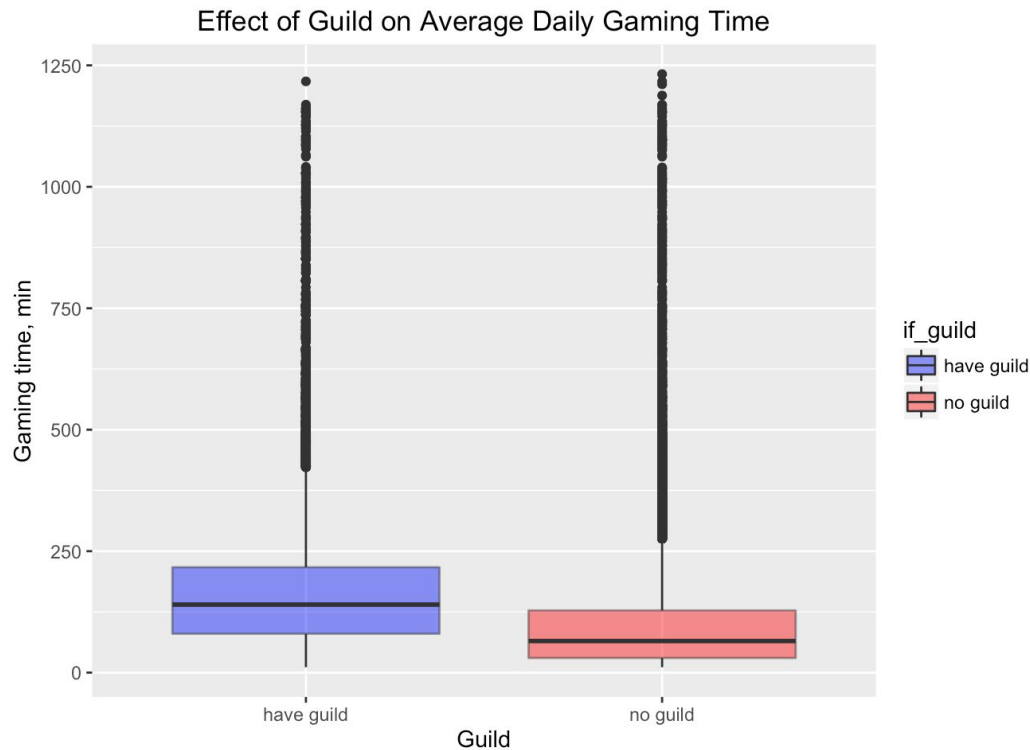
Beyond the Analysis: Predicting how long a user will stay in the game

If we can predict a user is leaving the game before he actually quits the game, the game operator can take actions to prevent it from happening.

- Using average daily play time and playing density in each period as features
- The future time was binned in one week interval for 7 weeks. The prediction is to find out whether or not an avatar will be seen in the game within a defined time frame.
- For each time bin, gradient boosting machine(GBM) classifier was applied to predict if an avatar will be seen in the game
- AUC was used to evaluate the model

Predicting how long a user will stay in the game

Avatar	1 week	2 week	3 week	4 week	5 week	6 week	7 week	8 week	9 week
x	YES	YES	YES	NO	NO	YES	NO	YES	NO
y	YES	YES	YES	YES	YES	NO	NO	NO	YES

Action taken

Action taken

Key takeaways:
- Real time playing density plays a key role
- Actions can be taken to re-engage a user if prediction indicates a high probability of quitting.