
Exploratory Analysis of WOWAH Dataset

— Presented Yuming Fang —

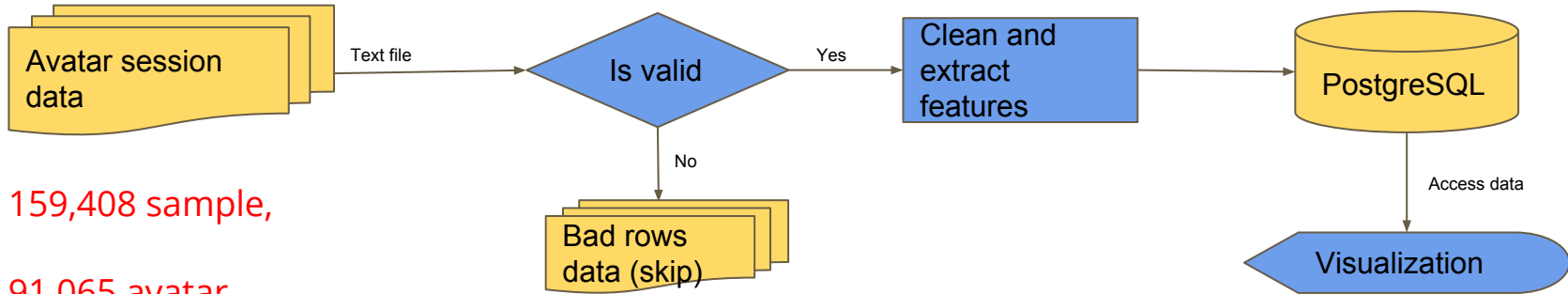
Motivation of the Analysis

- For game companies, players' behavior is one of the most important factor when designing a game
- Exploring players' gaming time and playing patterns are crucial to increase business revenue
- Being able to predict how long a player will stay in the game will make impact for business growth



query_time	avatar_id	guild_id	user_level	user_race	user_class	game_zone
2005-12-31 23:59:46	1	-1	9	Orc	Shaman	Durotar
2005-12-31 23:59:52	4	-1	14	Orc	Shaman	Durotar
2005-12-31 23:59:52	2	-1	13	Orc	Shaman	Durotar
2005-12-31 23:59:52	6	-1	18	Orc	Warlock	The Barrens
2005-12-31 23:59:52	5	-1	16	Orc	Hunter	The Barrens

Data Processing



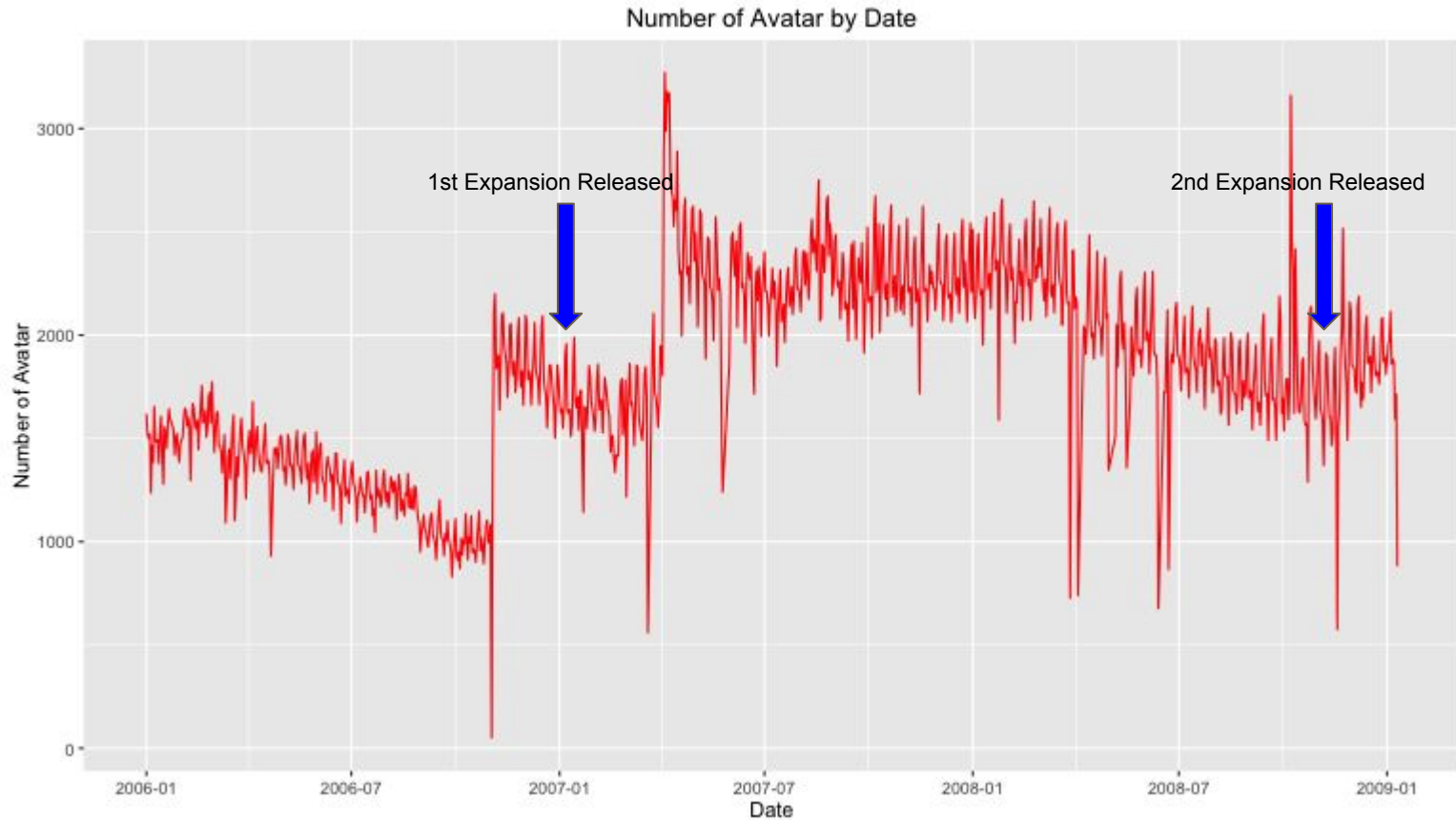
159,408 sample,

91,065 avatar,

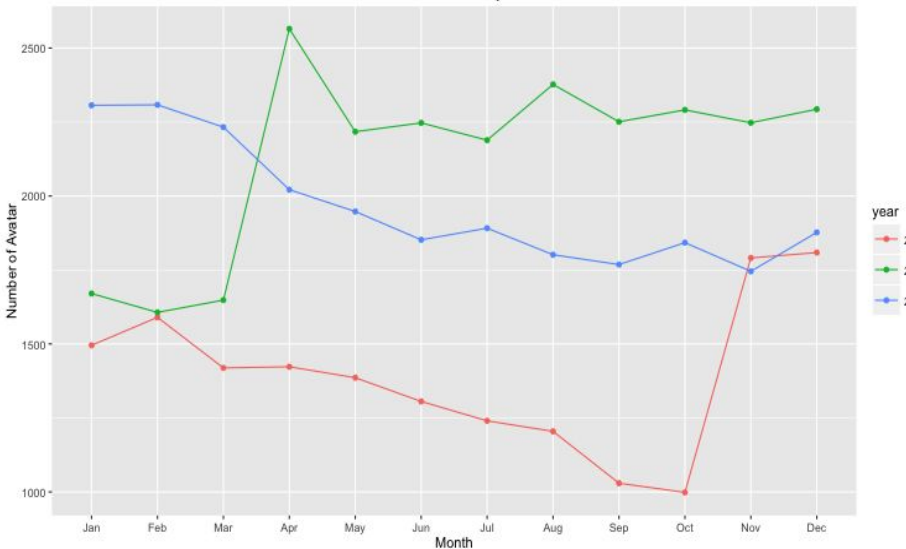
1,107 days

3.7 G

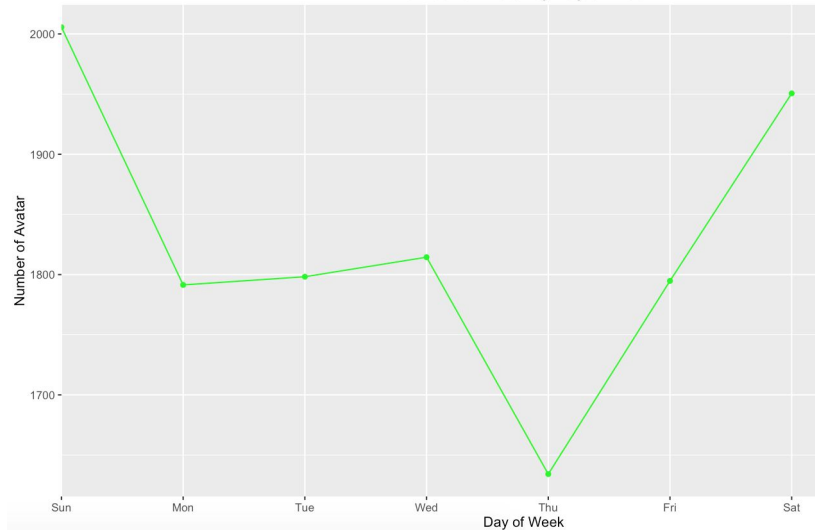
Initial Data Analysis



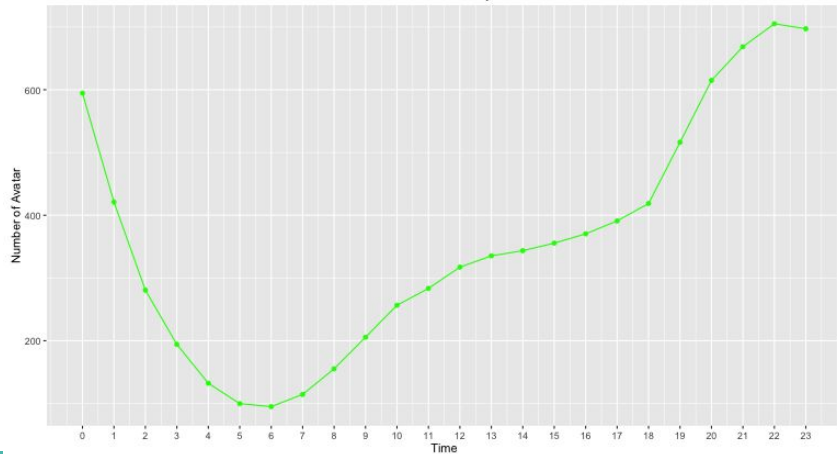
Number of Avatar by Month



Number of Avatar by Day of Week



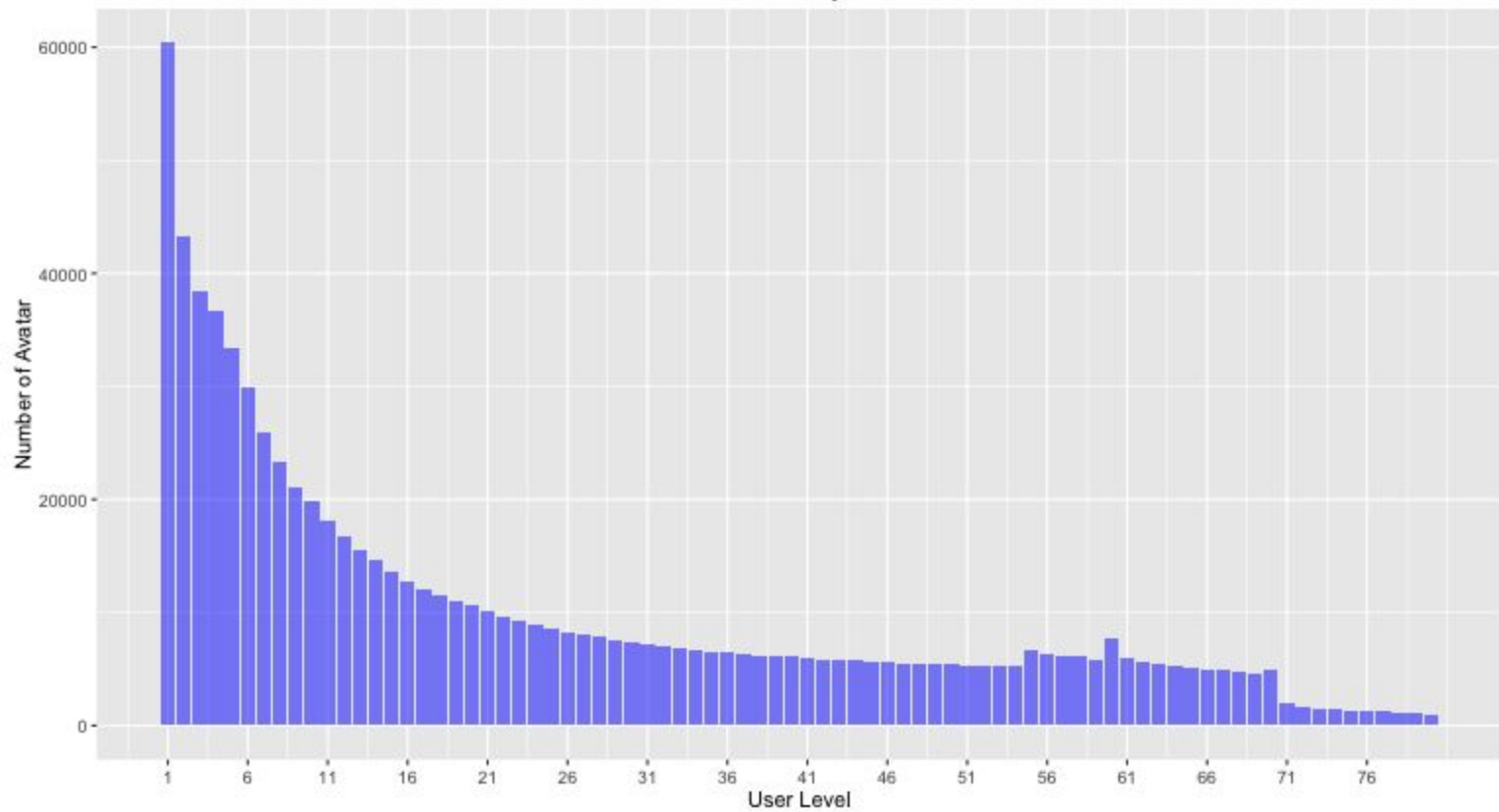
Number of Avatar by Hour



Insights

- It shows three different moving range: 2006-01 to 2006-11; 2006-11 to 2007-04; 2007-04 to 2009-01.
- For each moving range, the number of avatars increased sharply then slowly decreasing.
- The release of expansion does not seem to cause the sharply increase in the number of avatars.
- The business is doing really well in the year 2007 and starts to going down after March, 2008
- Players are spending most of their time playing the game on weekends.
- The game operator schedules weekly maintenance down time on Thursday morning
- Peak traffic happens between 8pm to midnight; the lowest traffic happens between 3am to 9am.

Number of Avatar by Game Level



user_class	count
Warrior	21935
Hunter	11557
Rogue	11167
Mage	11128
Warlock	9817
Paladin	8302
Shaman	7987
Priest	6705
Druid	4162
Death Knight	1776

user_race	count
Blood Elf	29091
Orc	18191
Undead	16920
Troll	14755
Tauren	14517

game_zone	num
Durotar	42996
Eversong Woods	30273
Orgrimmar	27136
Tirisfal Glades	26809
Mulgore	22153



I am a Blood
Elf Warrior!

Insights

- As game level increases, less avatars were observed.
- The number of avatar drops quickly before level 20 and stays stable between level 30 and 70.
- A big drop was observed between level 70 and 71.
- Zone Durotar has the most number of avatars
- A lot of users like class Warrior.
- They also like the race Blood elf.

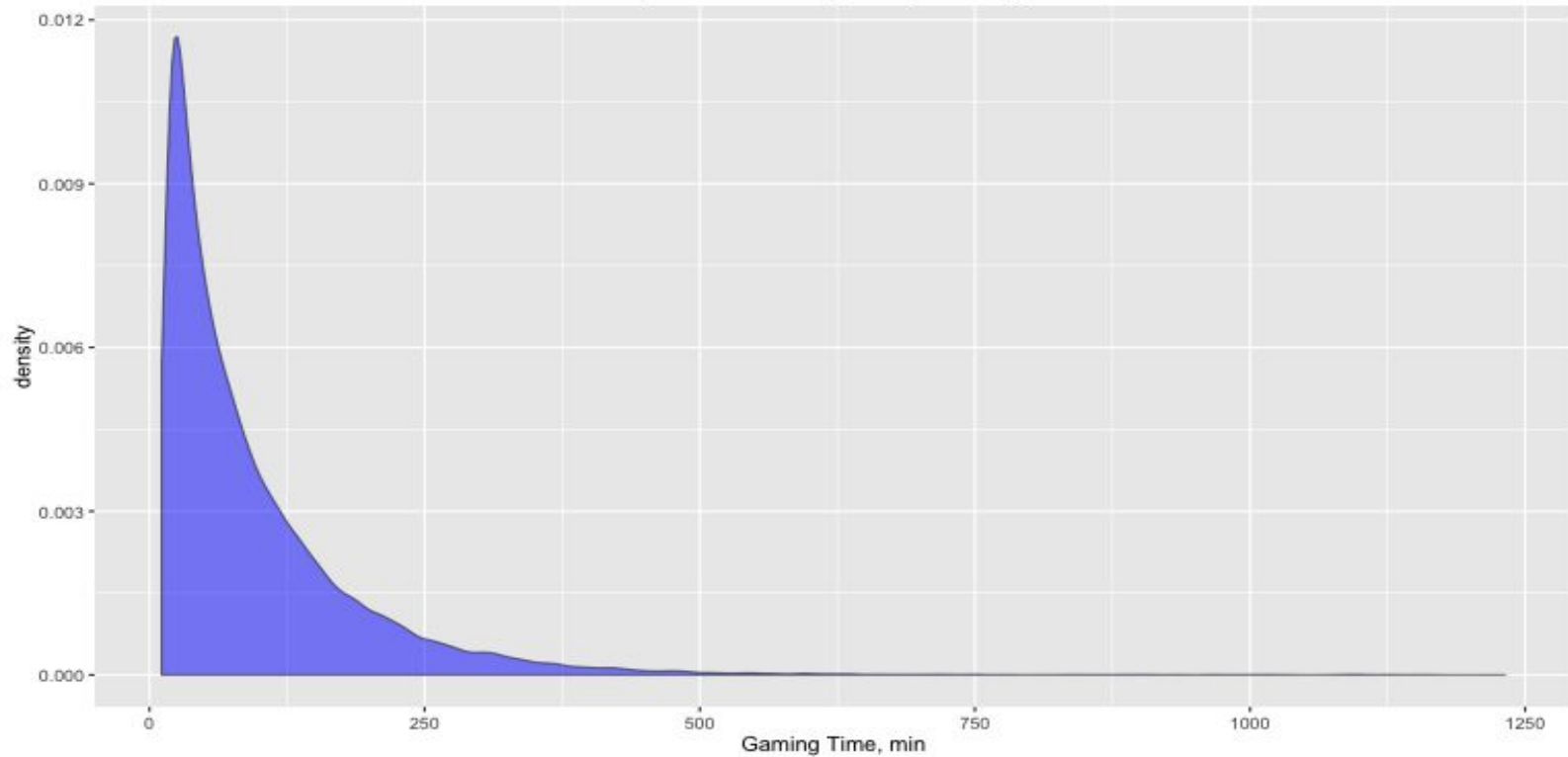
How long does a user play WOW per day?

How to calculate the daily gaming time?

- If the same avatar shows up in each consecutive sample by about 10 mins interval, it indicates the user is playing the game during this period.
- Using the total amount of the in session time over by the number of days that avatar was observed
- Avatars that have daily gaming time less than 10mins were excluded

Statistical summary shows that more than half of the active users spending more than 64 mins per day playing the game

Density Plot of Average Daily Gaming Time

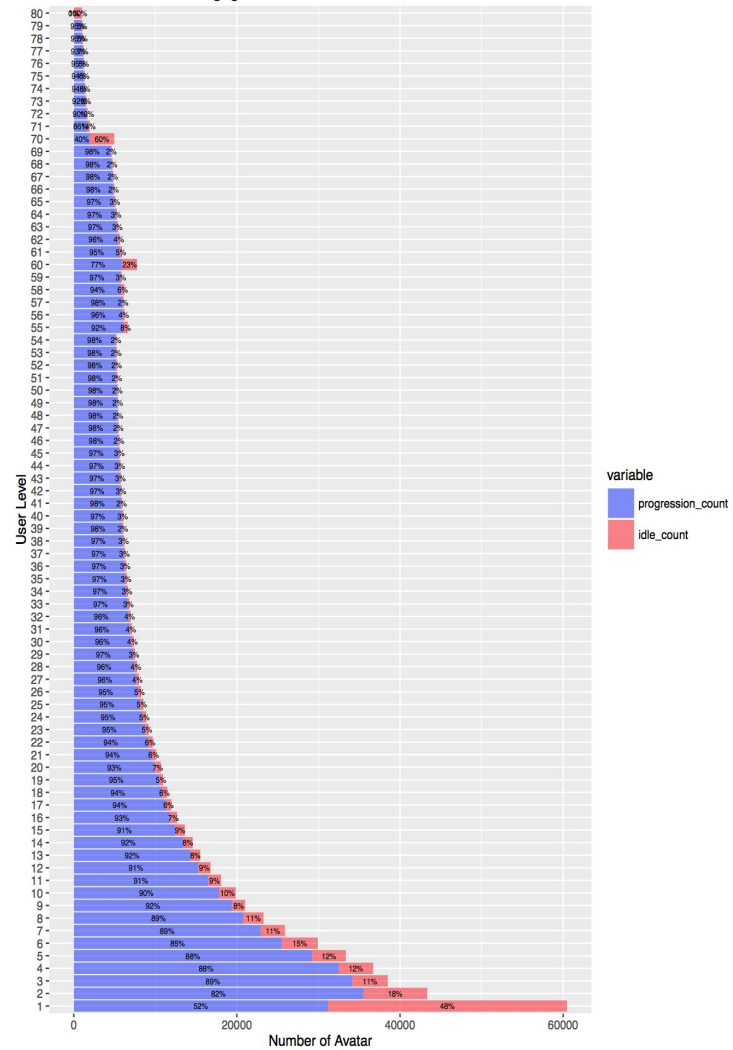


Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
11.00	30.00	64.00	97.07	127.00	1232.00

Engagement rate at different game level

- At level one, almost half of avatars does not make any progression.
- At most of the levels, the user engagement rate is over 80%.
- Level 80 is the highest level so no one can make any more progression.
- Level 60 and 70 used to be the highest level until expansion package was released. So there is an decrease in engagement rate due to some users might abandon the game when they reach the highest game level.

User Engagement Rate at Different Game level

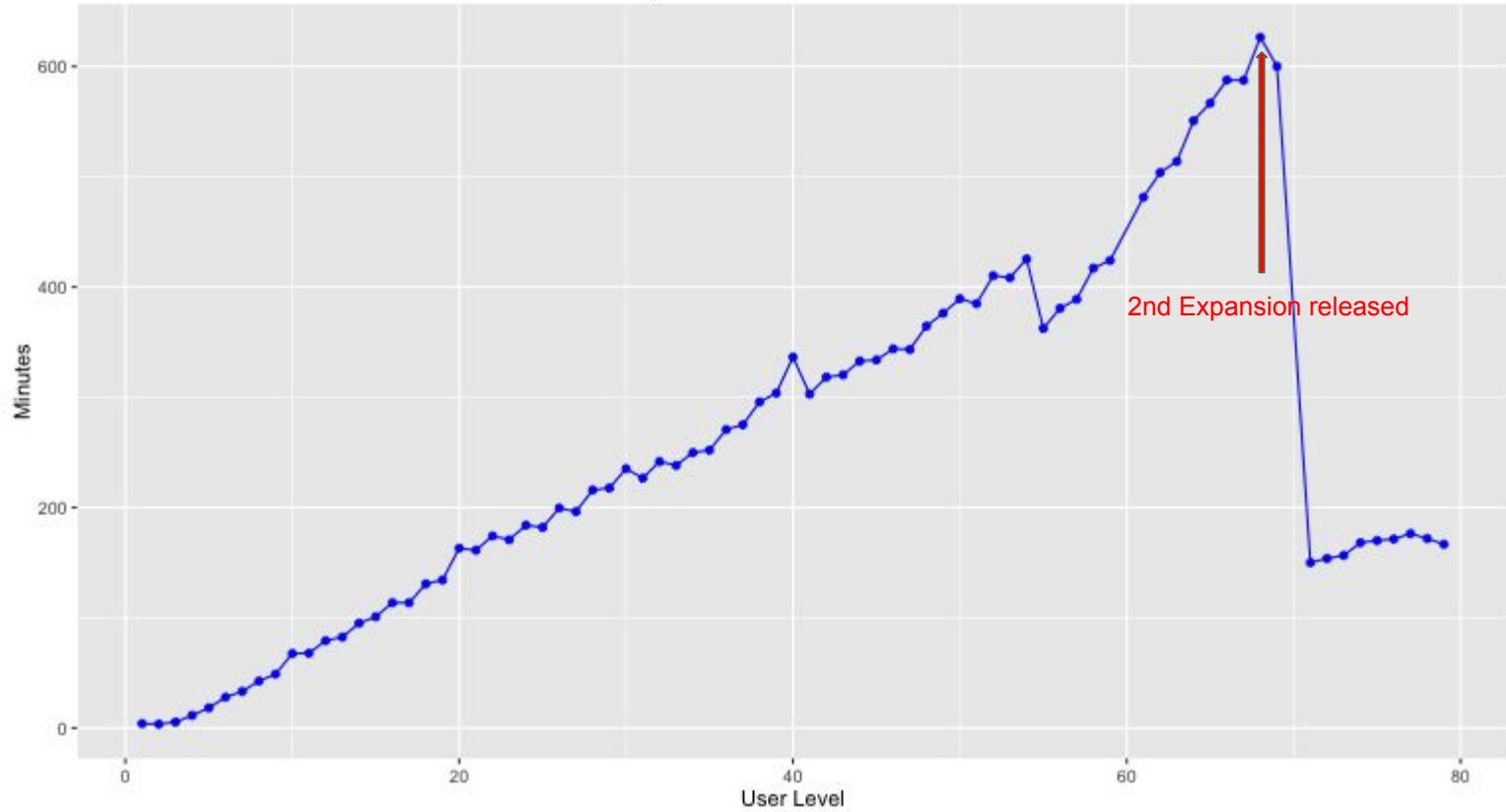


How quickly a user's avatar advanced to a new level?

- Level 60 and 70 were excluded due to the waiting period of expansion package
- The graph shows a linear increase of progression time when user level increases from 1 to 70.
- When the second expansion package was released, users make progression very quickly.

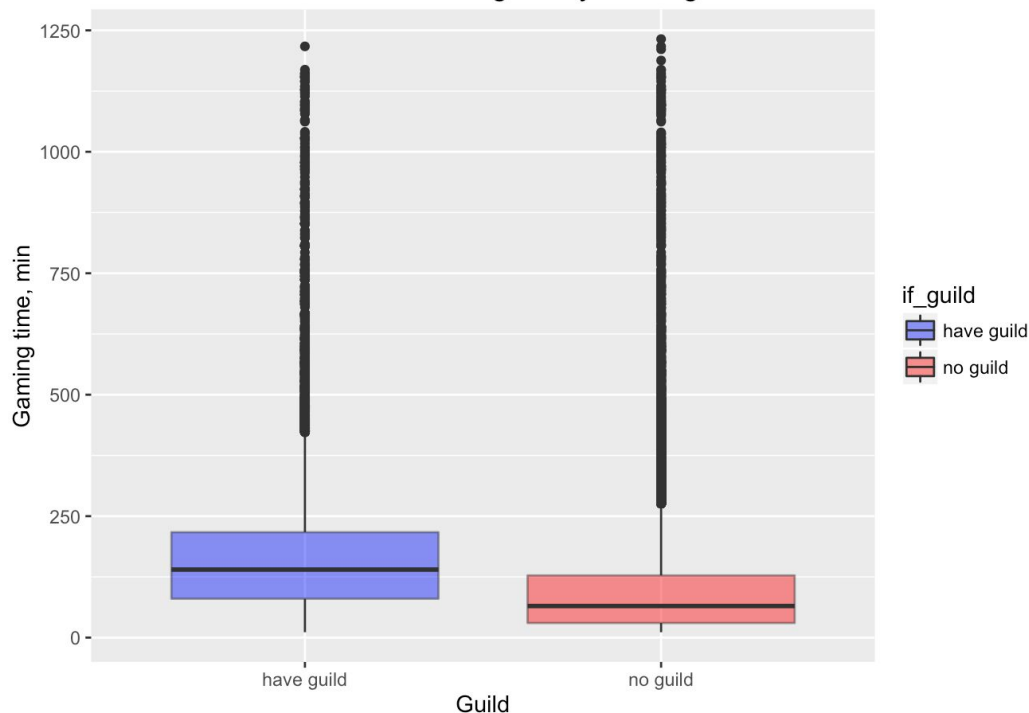


Porgression Time for Each Level



Does having a guild affect engagement?

Effect of Guild on Average Daily Gaming Time



If fact, it does!

- Two sample T test shows there is significant difference between the means of two groups
- Users have no guild have an average of daily gaming time about 98 minutes, while users do have guild have an average of 164 minutes.

User segmentation



Hard Core
Player

VS

Casual
Player



Why user segmentation:

- 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

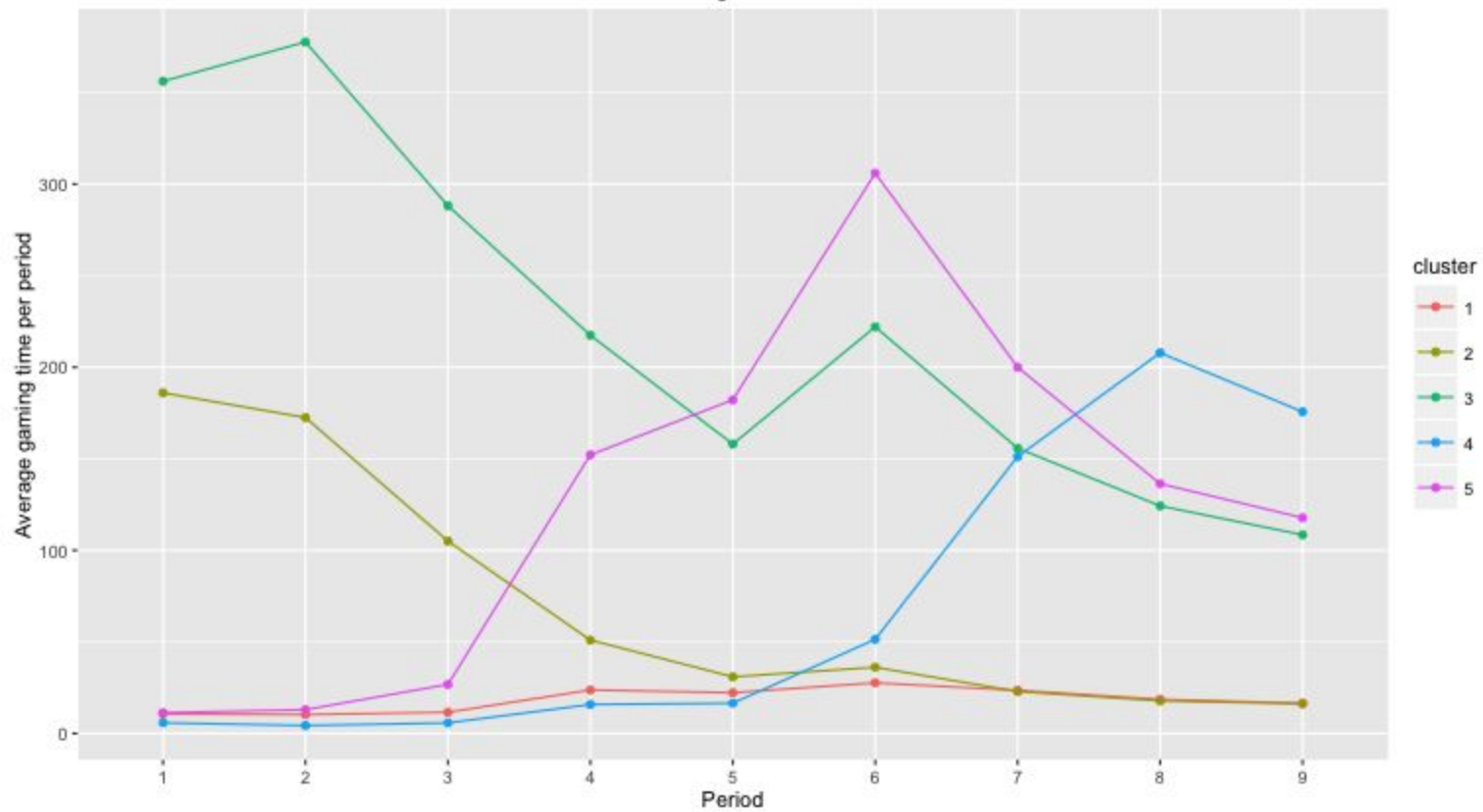
Feature generation:

- 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

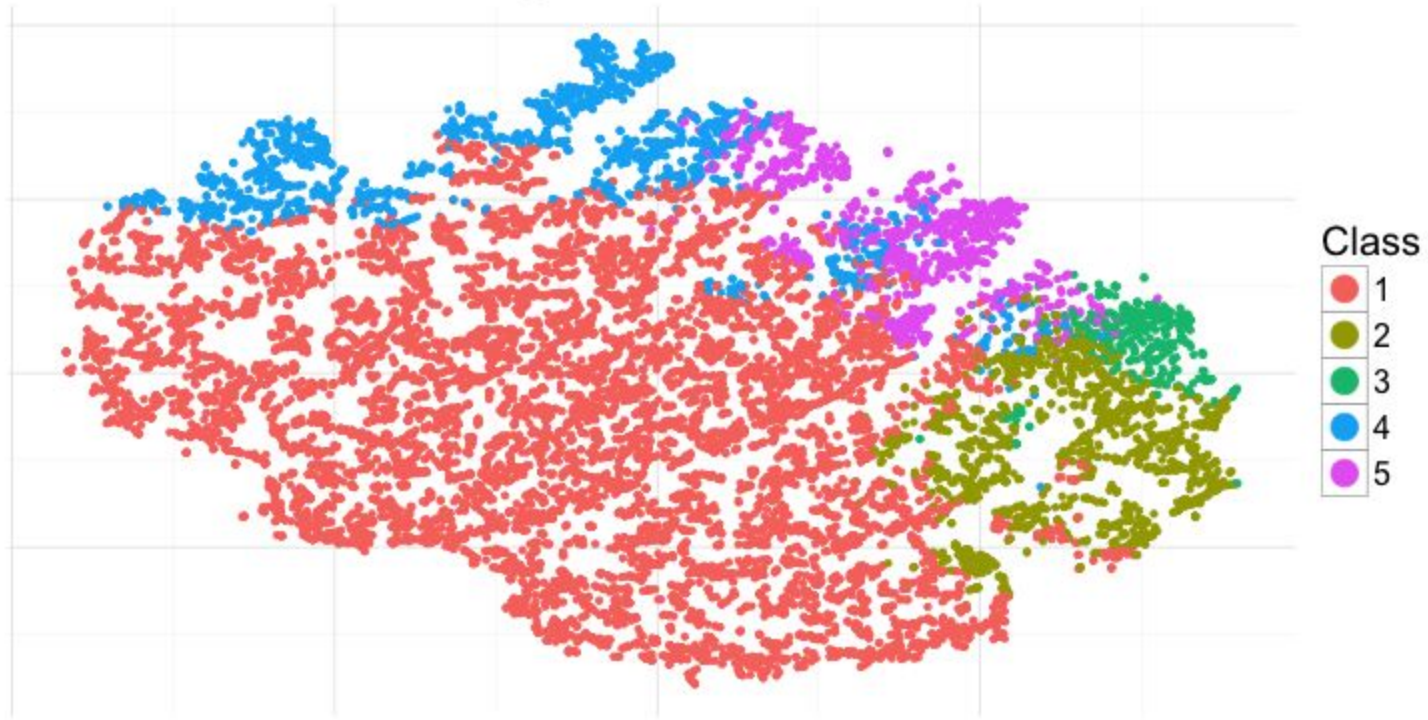
User segmentation

- Use k-means to group users into 5 clusters
- Applied t-SNE to reduce the user space into 2 dimensional to visualize it
 - 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)*

User Segmentation



t-SNE 2D Segmentation of WoW User



Predicting how long a user will stay in the game

Motivation: 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

The most important feature is
the most recent playing
density