

Mentor:

Dr. Narayana D

Interim Report

Team members:

* Sulekha Aloorravi
* Deepa Venugopal
* Niharika Thanavarapu
* Lakshmipriya M

**Behavioral pattern recognition of multiplayer online role-playing game players using Big Data Analytics and Machine Learning**

Contents

[I. Domain and Context 2](#_Toc502580586)

[1. Domain 2](#_Toc502580587)

[2. Industry worth 2](#_Toc502580588)

[3. Context of this Project 2](#_Toc502580589)

[4. Objective 3](#_Toc502580590)

[II. Problem Statement 3](#_Toc502580591)

[1. Perform Exploratory analysis 3](#_Toc502580592)

[2. Perform Predictive Analytics by applying Machine Learning Models 3](#_Toc502580593)

[III. Proposed Solution 3](#_Toc502580594)

[IV. Evaluation Metrics 4](#_Toc502580595)

[V. Exploratory Data Analysis 4](#_Toc502580596)

[Step 1: Parse Wow Logs 5](#_Toc502580597)

[Step 2: Clean up incorrect records 6](#_Toc502580598)

[VI. Exploratory Visualization 8](#_Toc502580599)

[Tableau Visualizations 8](#_Toc502580600)

[Insights from Exploratory Analysis 13](#_Toc502580601)

[VII. Summary of Initial Findings 13](#_Toc502580602)

[Models attempted 13](#_Toc502580603)

[VIII. Challenges 24](#_Toc502580604)

[IX. Next Steps 24](#_Toc502580605)

[X. References 25](#_Toc502580606)

# Domain and Context

## Domain

A massively multiplayer online game (more commonly, MMO) is an online game which is capable of supporting large numbers of players, typically from hundreds to thousands, simultaneously from around the world.

These games can be found for most network-capable platforms, including the personal computer, video game console, or smartphones and other mobile devices. MMOs can enable players to cooperate and compete with each other on a large scale, and sometimes to interact meaningfully with people around the world.

## Industry worth

The UK MMO-market is worth £195 million in 2009 compared to the £165 million and £145 million spent by German and French online gamers. The US gamers spend more, however, spending about $3.8 billion overall on MMO games. $1.8 billion of that money is spent on monthly subscription fees. The money spent averages out to $15.10 between both subscription and free-to-play MMO gamers. The study published by “Today’s Gamers MMO Focus Report” also found that 46% of 46 million players in the US pay real money to play MMO games.

## Context of this Project

It is challenging to develop the database engines that are needed to run a successful MMOG with millions of players. Understanding the behavior of players using their activity data is more important for these game developers to come up with better strategies in game development.

The variety, volume, velocity, value and veracity (Big Data 5Vs) of data that is involved in these Gaming environments exceed the limits of analysis and manipulation of conventional tools, therefore, Big Data platforms are required to handle and interpret this data.

Great volumes of data are generated all the time in these environments. Each interaction made by a player creates data that are transferred and stored, and if properly analyzed, can contain valuable information. This information can be vital for the continuity and improvement of a game. Patterns can be detected from these data and even predictive analysis can be made to foresee the actions and intentions of the players inside the game.

## Objective

Objective of this Project is to perform analytics on one such Big Data Gaming Environment and the results would help game developers in:

* Optimizing user experience
* Improving revenue
* Raise the level of control over the environment

# Problem Statement

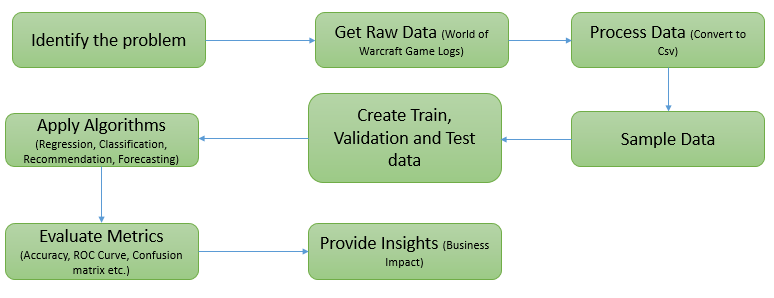
## Perform Exploratory analysis

* 1. To cluster players into different groups based on features in dataset
  2. To analyze and visualize timeline patterns of players by different groups and parameters
  3. To create heat map based on the gaming zones
  4. To visualize patterns based on Guilds they belong to

## Perform Predictive Analytics by applying Machine Learning Models

* 1. Forecast the number of players expected in future time point
  2. Predict player churning
  3. Recommend guilds to players for effective gaming

# Proposed Solution

Following workflow will be followed to solve the identified problems in this Project:

# Evaluation Metrics



# Exploratory Data Analysis

We have chosen an online game named “World of Warcraft” which is most suitable for this Project.

A large and scalable dataset with 3 years of player logs are released by Blizzard Entertainment for research purposes. We are using this dataset of our Project.

|  |  |
| --- | --- |
| **Data set Summary** | |
| **Attribute** | **Value** |
| Data duration (in days) | 1107 |
| Sampling Rate per day | 124 |
| No. of Samples | 138084 |
| No. of Records (rows) | 36,513,647 |
| No. of Values (Data points) | 438,163,764 |
| Size of data (in GB) | 3.4 |
| Dataset Type | Logs |
| Format | Text Files |
| No. of Folders | 1095 |

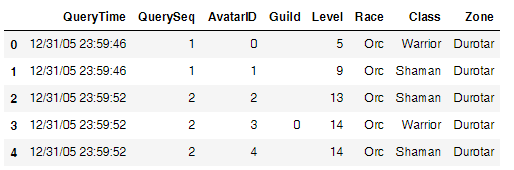
|  |  |  |
| --- | --- | --- |
| **Field Description** | | |
| **Field** | **Description** | **Data Type** |
| Query Time | Date and time when logs were generated | integer |
| Query Seq. # | Sequence of queries | integer |
| Avatar ID | Unique id for each user | integer |
| Guild | Group id of the player | integer |
| Level | Game level of the player | integer |
| Race | Blood Elf, Orc, Tauren, Troll, Undead | String |
| Class | Death Knight, Druid, Hunter, Mage, Paladin, Priest, Rogue, Shaman, Warlock, Warrior | String |
| Zone | One of the 229 Zones in World of WarCraft game | String |

## Step 1: Parse Wow Logs

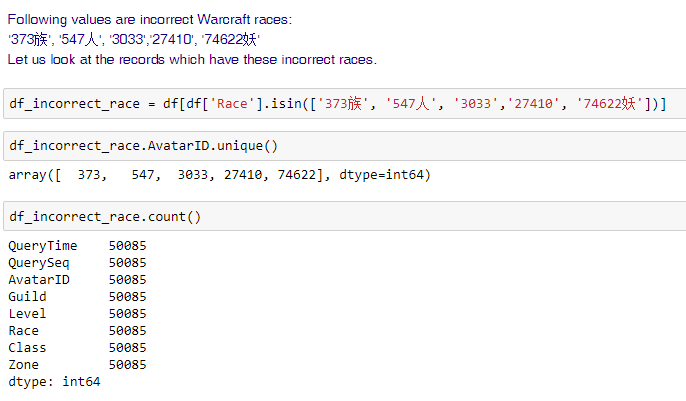


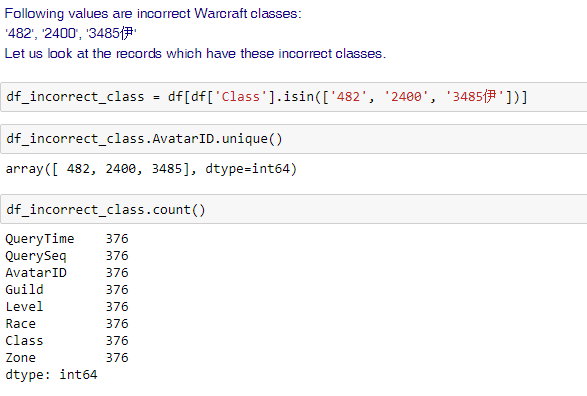


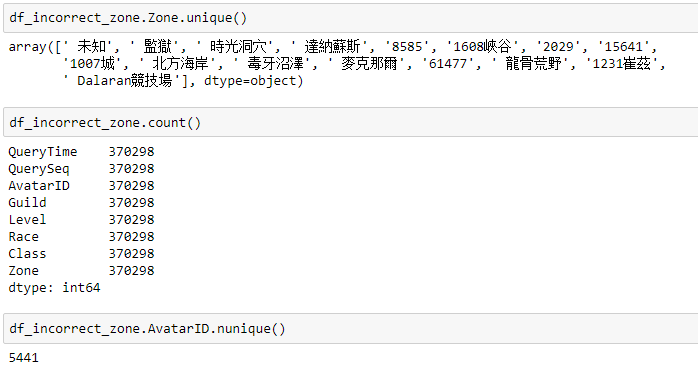


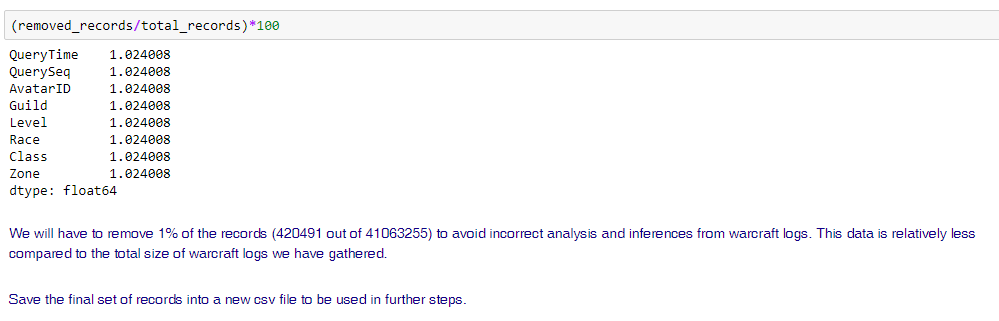


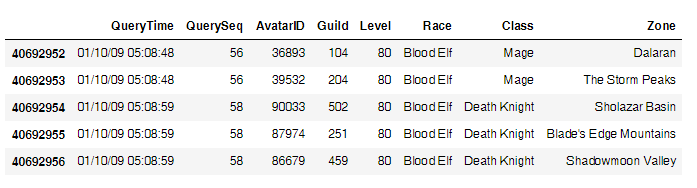
## Step 2: Clean up incorrect records







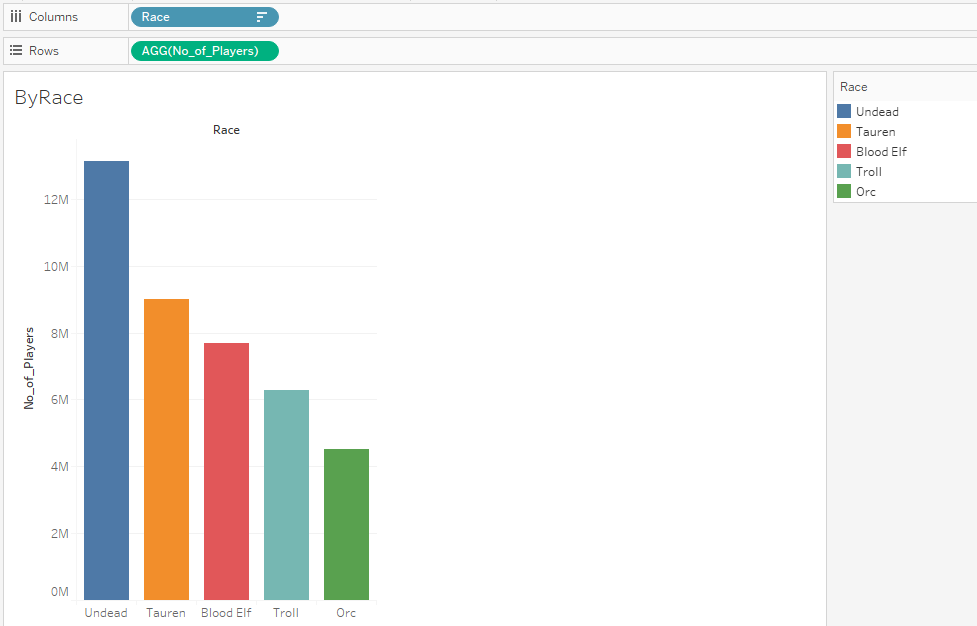


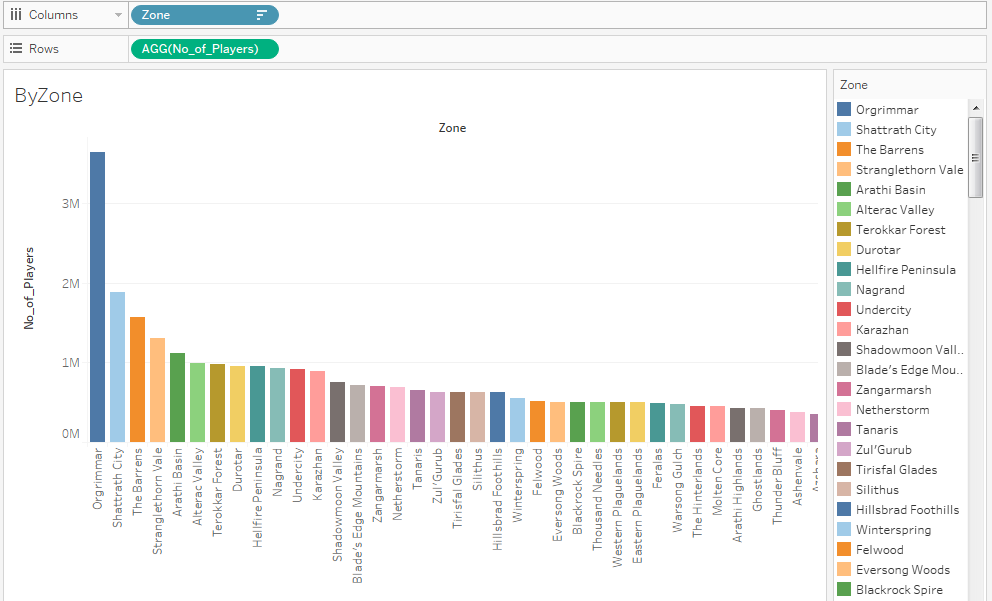


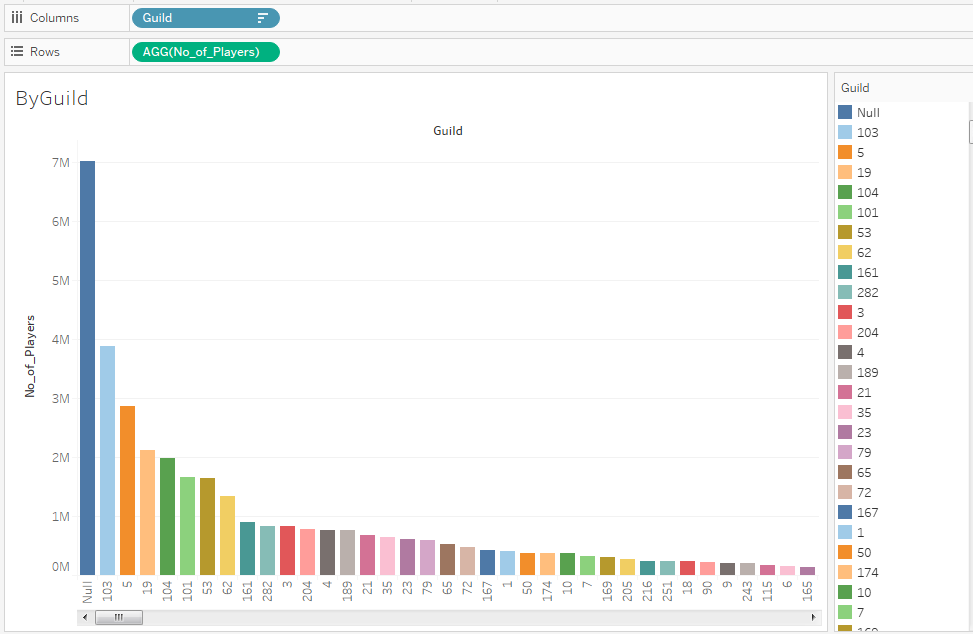
# Exploratory Visualization

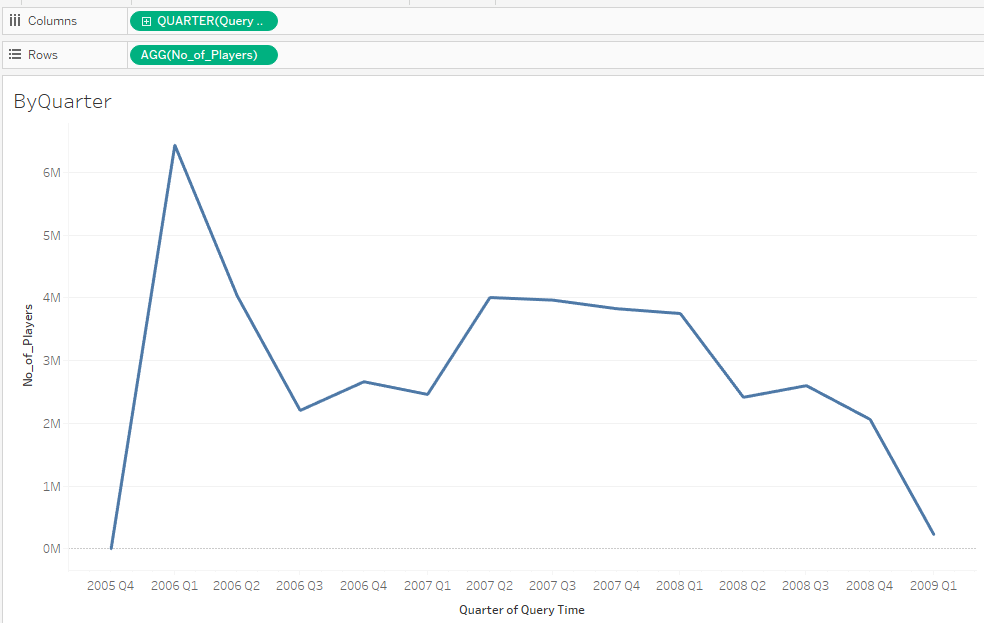
## Tableau Visualizations

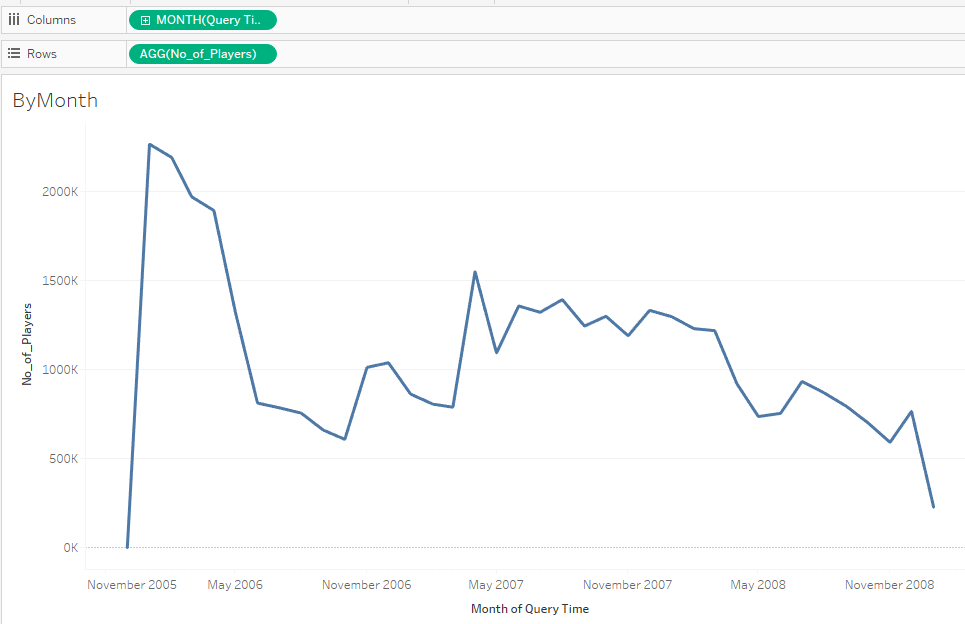


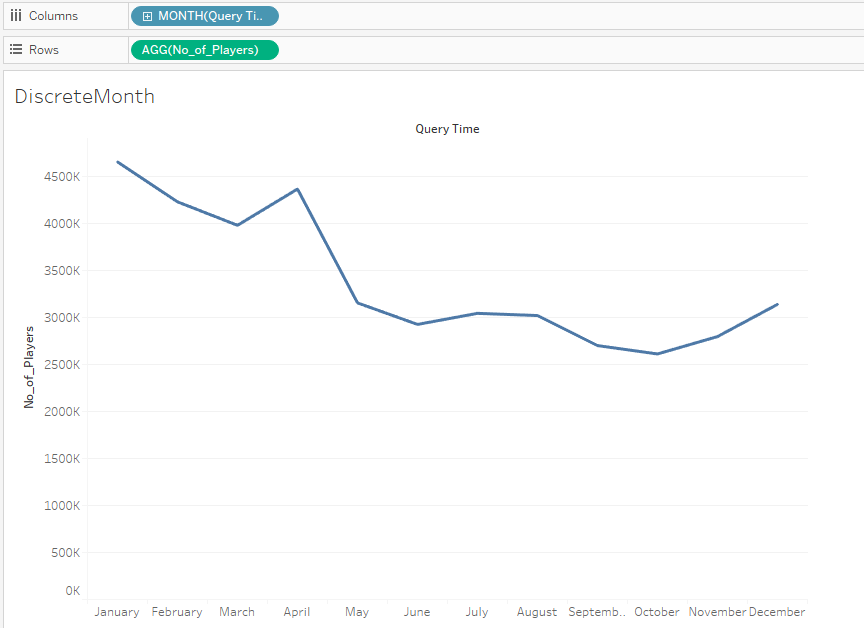


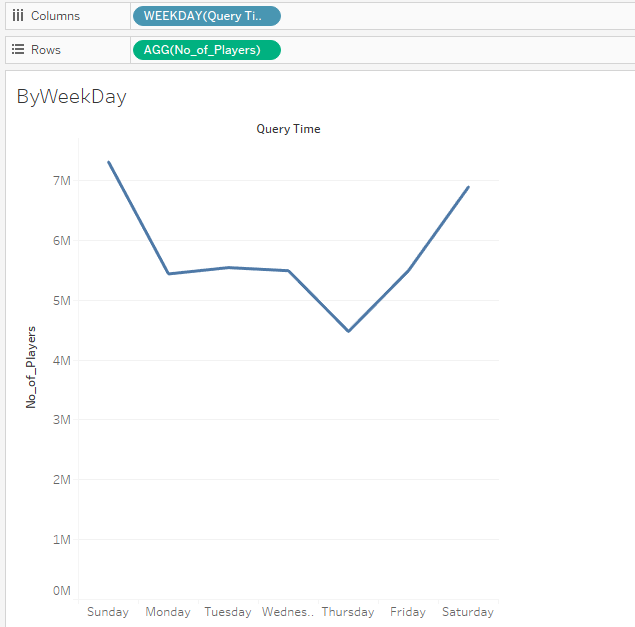


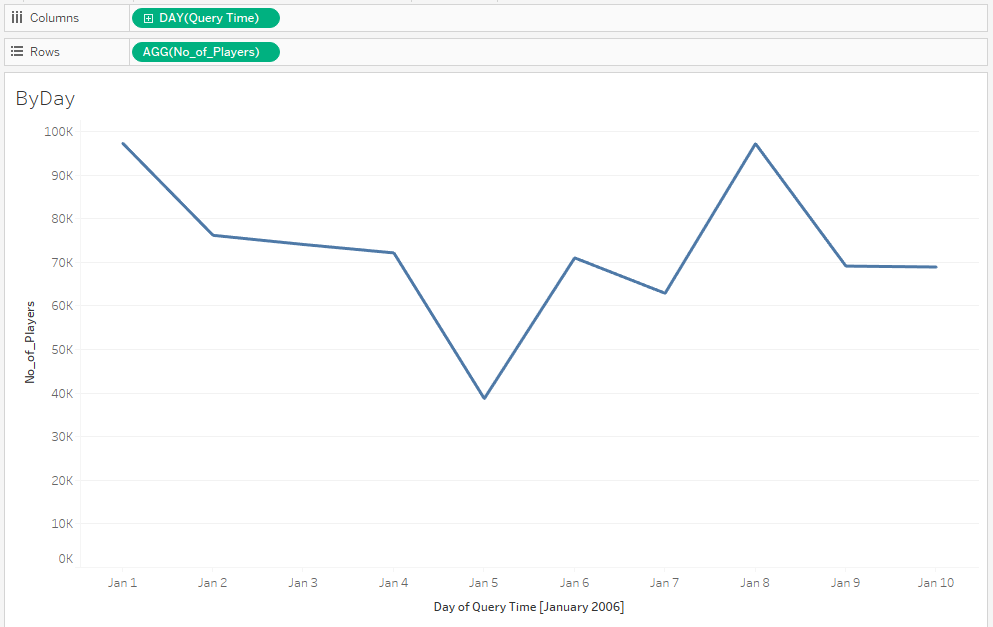


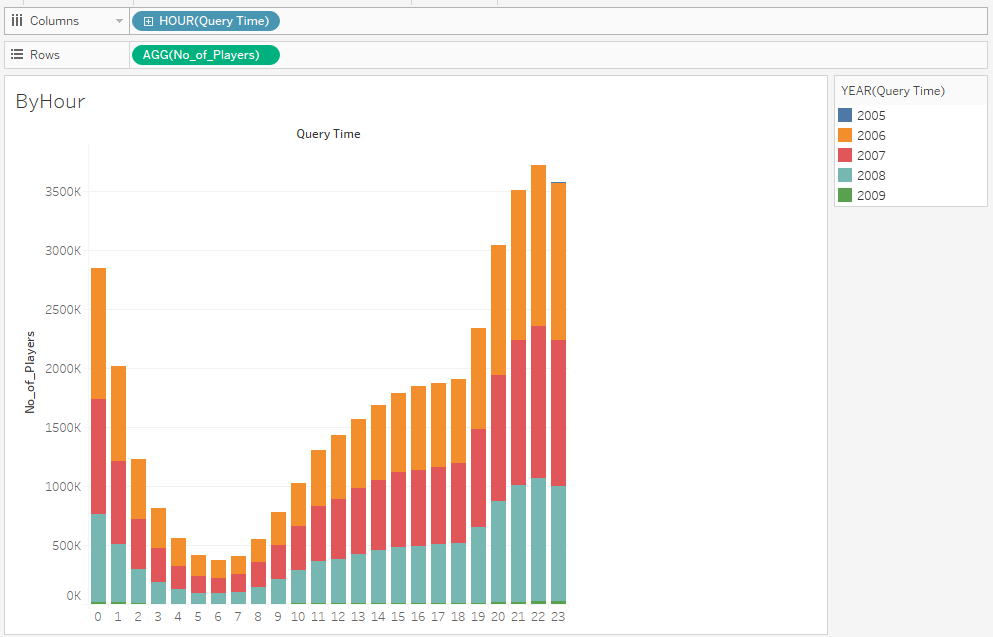












## Insights from Exploratory Analysis

1. Hunter is the class chosen by most of the players
2. Death Knight is chosen by least number of players
3. Undead is the Race chosen by most of the players
4. Orc is chosen by least number of players
5. Orgrimmar is the Zone chosen by most of the players
6. 7,014,160 players are playing without joining any guilds
7. Within the Data collection period, maximum number of players played world of Warcraft during Q1 2006
8. 2005 and 2009 data cannot be considered for exploratory insights since it covers data of less than a month
9. January is the month with most players and October is with least players every year
10. Sunday and Saturday (Weekends) are the days with more players and Thursday is the day with least players every week
11. 10:00 PM is the most played hour in a day and 6:00 AM is the least played

# Summary of Initial Findings

## Models attempted

#### 1. Neural networks for pattern recognition - Exploratory

In the process of solving problem statements, we first came up with identifying the game play patterns of players from the dataset.

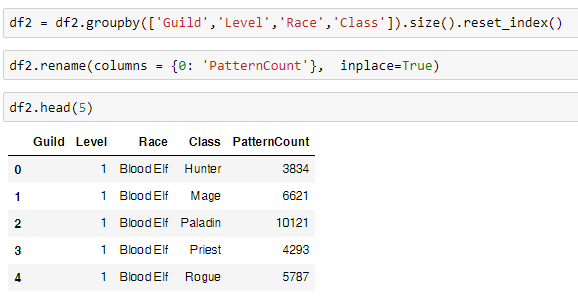
For this process, we have converted the dataset into numerical data that can be provided as input to Pattern recognizing neural networks.

This model would still be part of exploratory data analysis.

Step 1: Read data into a dataframe



Step 2: Group records by Guild, Level, Race and Class and count the number of records following each unique combination



Step 3: Convert string values on the above columns into their encoded numeric value

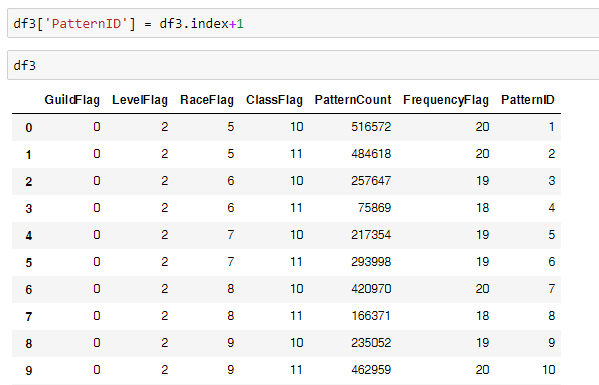




Step 4: Create Frequency Flag by creating a rule for each frequency



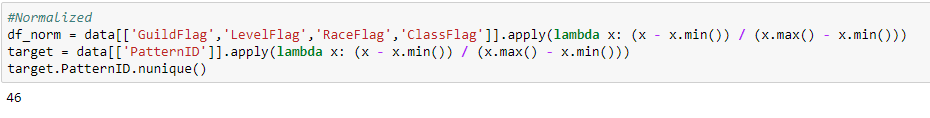
Step 5: Create Pattern ID for each unique pattern

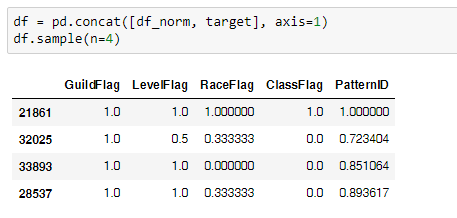


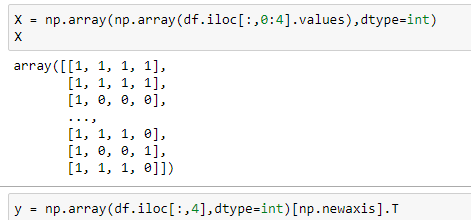
Step 6: Identify the number of unique patterns in the dataset



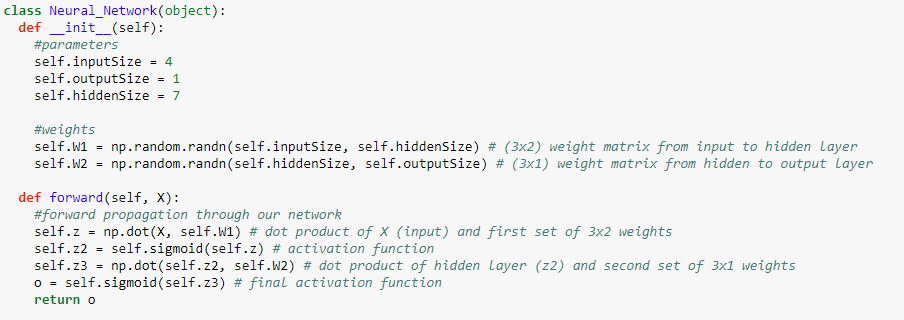
Step 7: Normalize the data and create Train & Test data



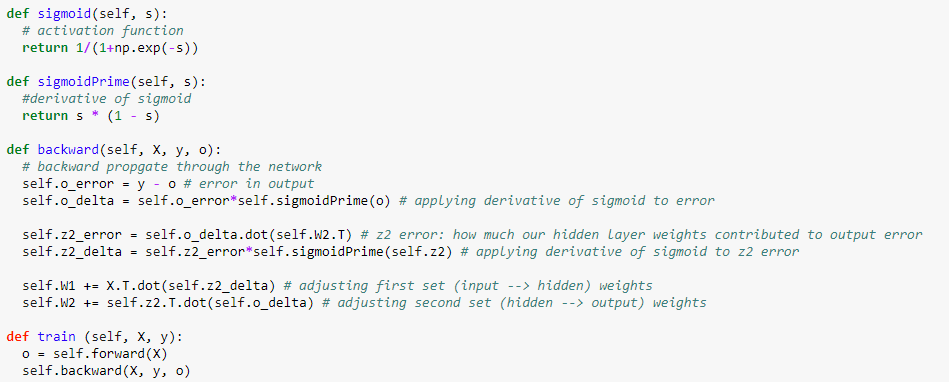




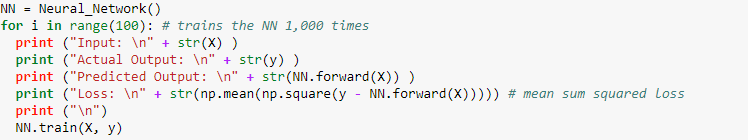
Step 8: Provide inputs to Multiple Back Propagation neural networks and predict patterns (Input Layer = 4, Hidden Layer = 7 and Output Layer = 1)



Step 9: Use sigmoid function to activate the neurons



Step 10: Predict Pattern id and Loss on Test Data



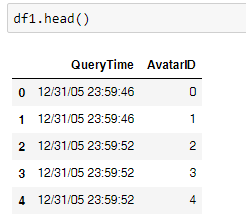


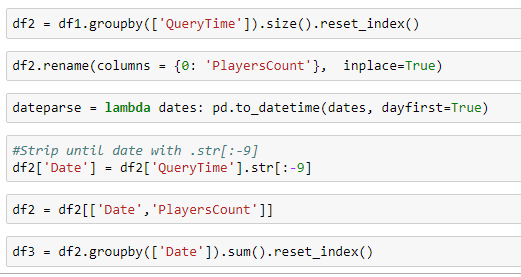
The results of the tests with the neural networks shows how it can be trained to recognize groups of players by their characteristics and learn the unique patterns of Guild, Level, Race, Class combinations of these groups in the game.

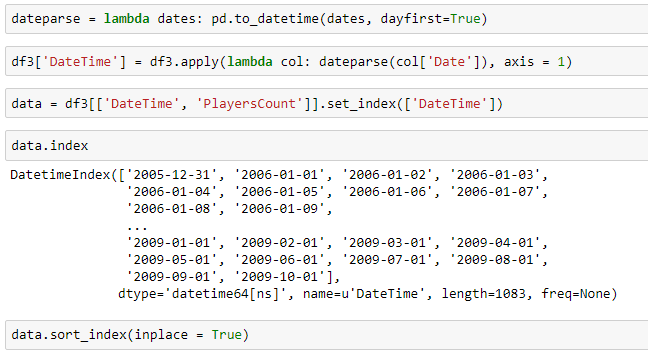
#### 2. Time series forecasting to predict number of players

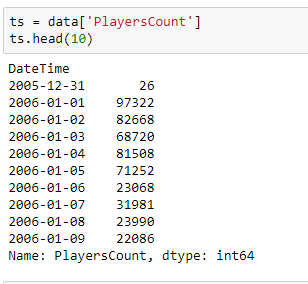
Another algorithm applied so far on this dataset is ARIMA model **Auto-Regressive Integrated Moving Averages** to predict number of players expected on a future time series

Step1: Load and handle Time series data in Python

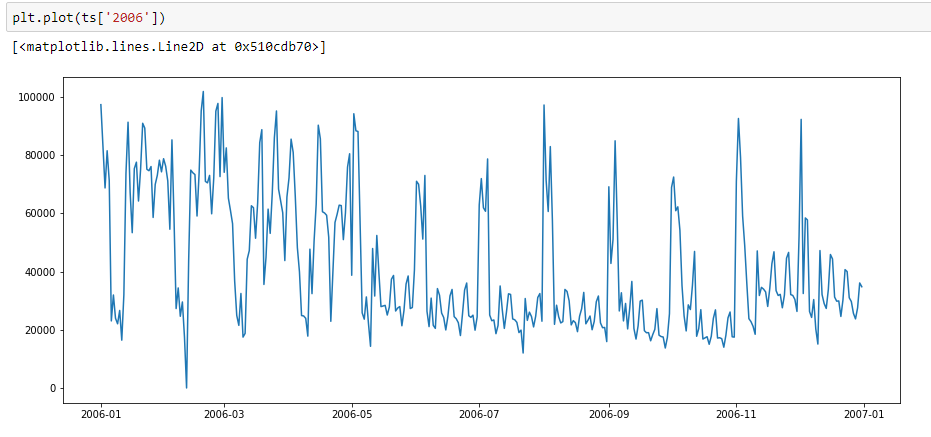




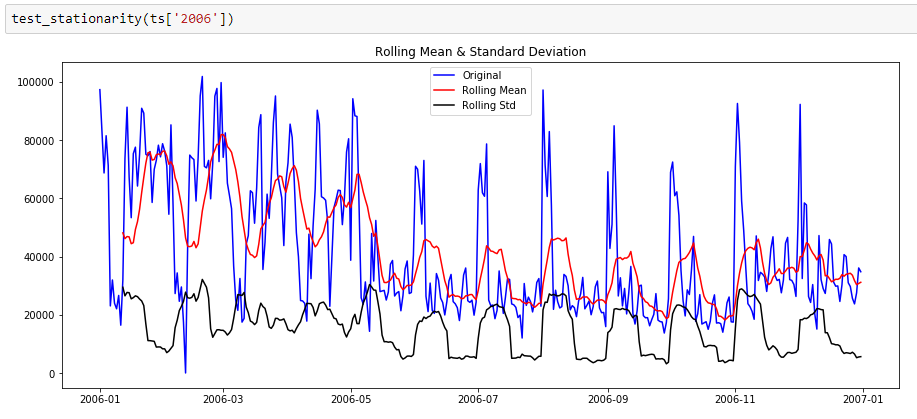


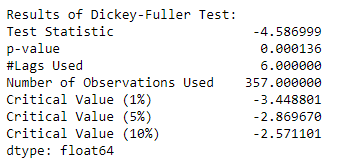


Step 2: Check stationarity of a Time series

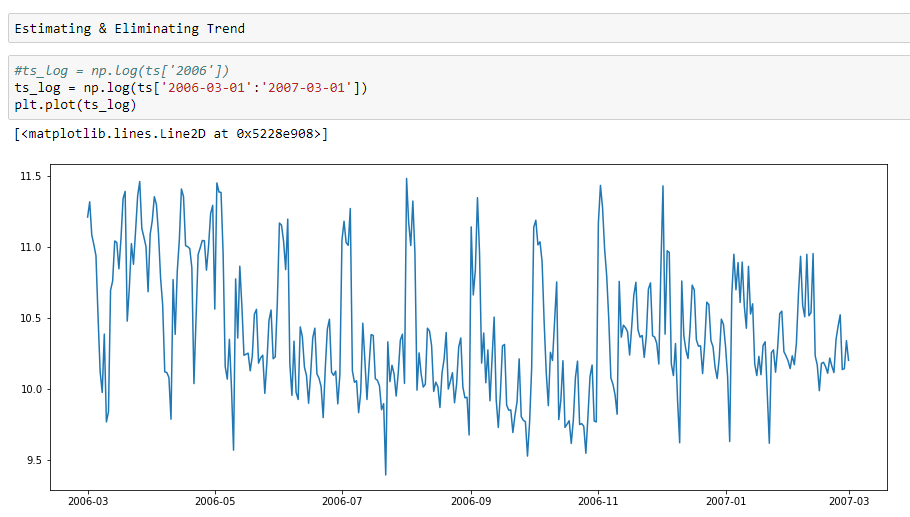


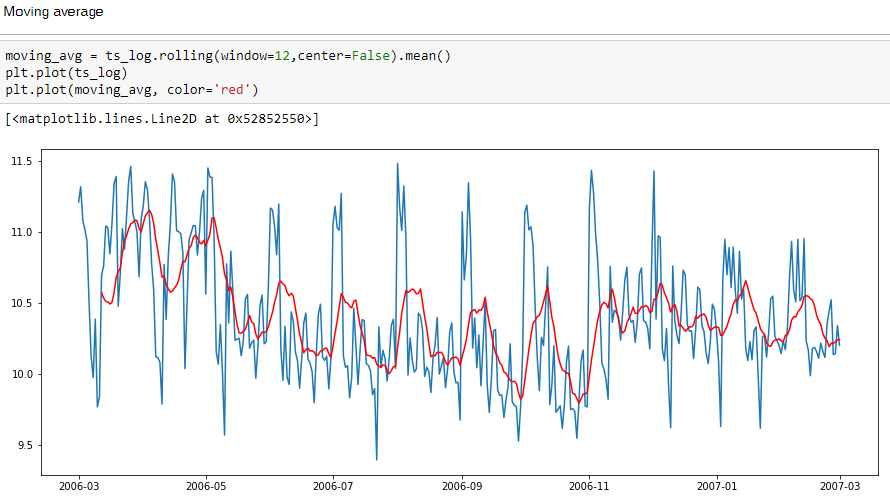


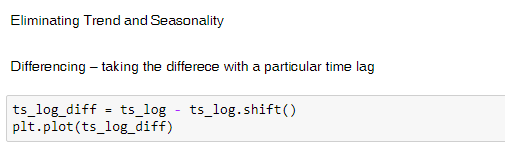


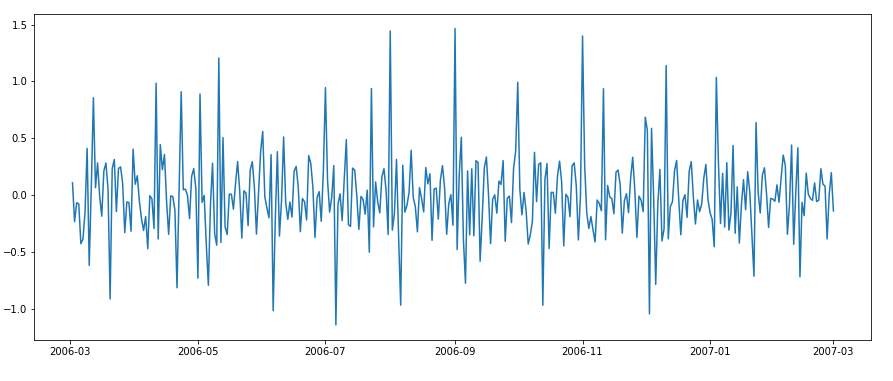


Step 3: Make time series stationary

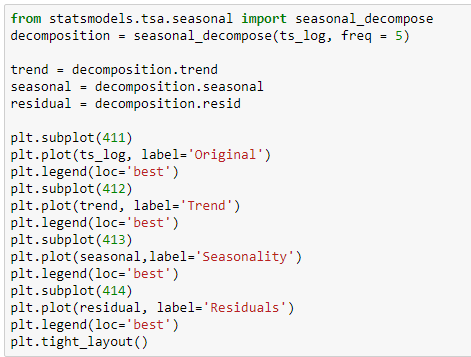


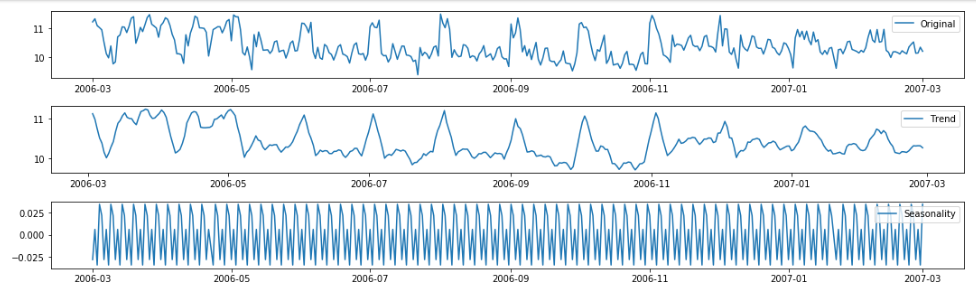


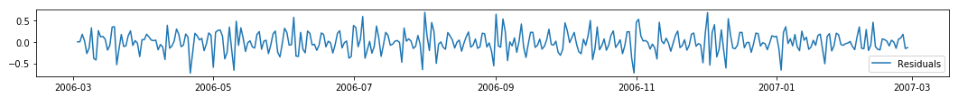






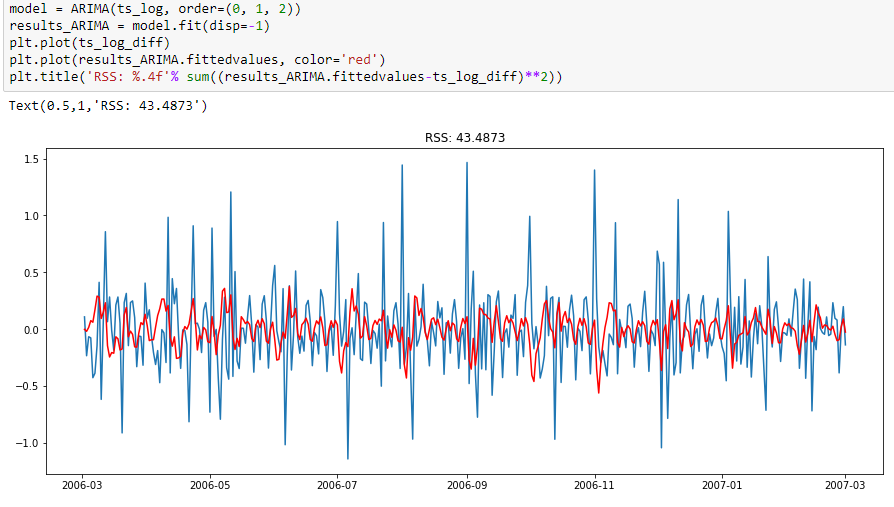


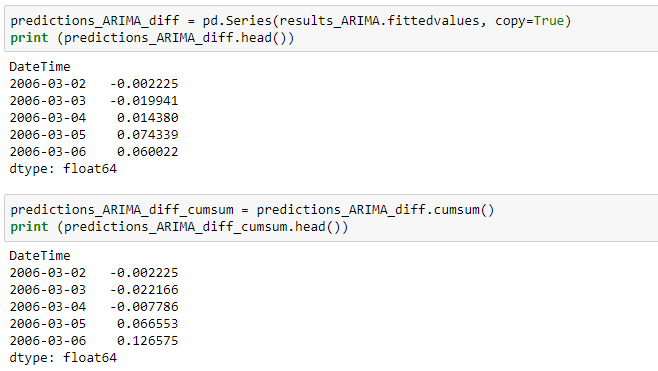


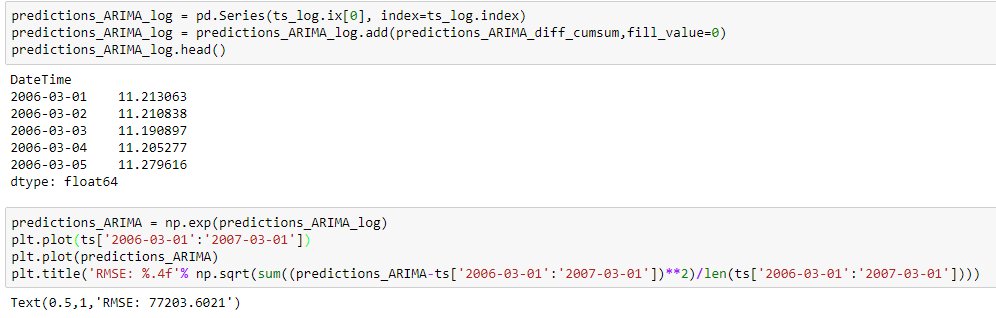


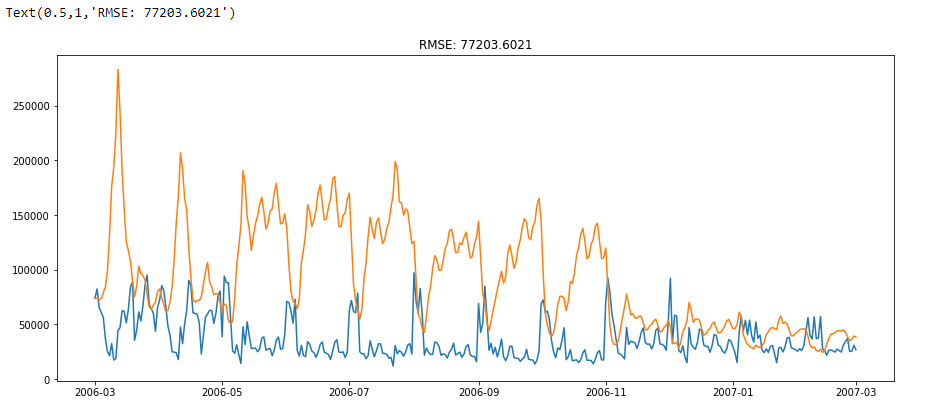
Step 4: Forecast on Time series

Applying ARIMA model









# Challenges

* Data size of 438,163,764 data points is too time consuming to process in Python
* Data set had to be sampled into smaller chunks to apply models
* Handling data in Tableau for exploratory data analysis is also more time consuming but complete set of data needs to be used for Exploratory analysis and sampling will not provide appropriate insights
* Neural networks or Matrix Factorization is unable to handle all data points
* In Time series forecasting – ARIMA Model: RMSE value is very high and the model needs to be revisited.

# Next Steps



## References

1. **Big Data Analytics Using Neural networks**

***Author****: Chetan Sharma*

***Master’s Thesis:*** *San Jose State University*

1. **Game Analytics - Maximizing the Value of Player Data**

***Authors:*** *Magy Seif El-Nasr, Anders Drachen and Alessandro Canossa*

***Publisher:*** *Springer*

1. **Big Data Analytics in Cloud Gaming**

***Authors:*** *Victor Perazzolo Barros and Pollyana Notargiacomo*

**Paper:** 2016 IEEE International Conference on Big Data

1. **Setting Players’ Behaviors in World of Warcraft through Semi-Supervised Learning*****Authors:*** *Marcelo Souza Nery, Victor do Nascimento Silva, Roque Anderson S. Teixeira and Adriano Alonso Veloso*