DAT 522 Cameron Farrugia

**Final Project**

Introduction

Video games have always been a large part of my life, so I found it fitting to do my final project based on the economy of the biggest video game of all time. World of Warcraft has what is called in the game - the auction house. The auction house is a simpler version of Ebay. You are allowed to sell items you find or that you make in the game. There are two ways to sell items, just like on Ebay. You can set a bid price and a buyout price just like the popular website. The goal of this project was to collect a lot of data over the course of multiple weeks and try to find patterns and answer some questions. One of the major questions that I wanted to ask was as follows: “When is the best day to buy and sell goods?” In order to answer this question, I needed to collect data everyday and during all points of the day.

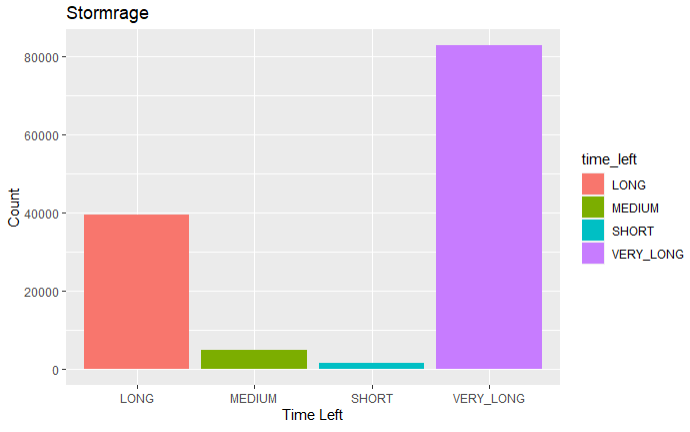
Data Gathering and Initial Analyses

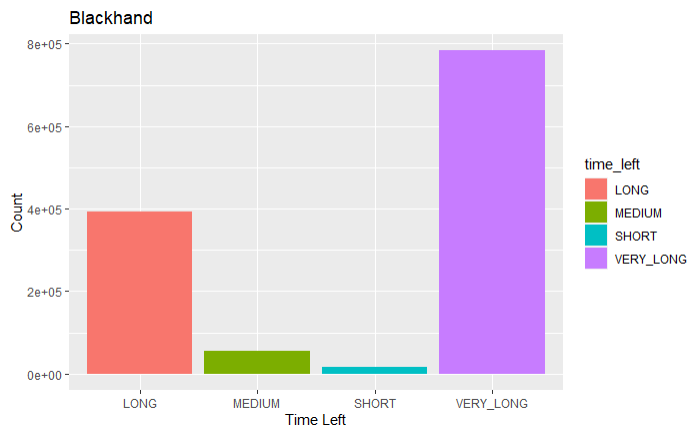
My first task was to gather the data. I looked into how to get auction house data on World of Warcraft. Luckily, Blizzard, the company that created the game, allows users to interact with their API. After applying for permission to access the API, I was able to write a python code that collected some data on the server I play on. (World of Warcraft has about 100 servers in North America alone that players can play on). The data wasn’t everything I was hoping for, It had some good information, but was missing a lot of detail. For instance, it didn’t have the item names in the data, but it did have an item number. The new problem became figuring out how to enrich the data so that I understood the items I was working with. Unfortunately, Blizzard didn’t have a place where I could easily get this information.

There is a website that a lot of players use for any and all kinds of information on World of Warcraft called WowHead. This website has a plethora of information on it including an item repository. I then used the information from the WowHead website to enrich the data. However the WowHead data was in the form of json which is not something I worked with in the past. It took some time to navigate json in order to get the information I needed to continue working with the Blizzard data. In addition to the names of items, I found more information about many of the items. The item repository had a lot of information about each item that I could have used if I wanted to go in more depth, but I wanted to start off by simply getting the item names and categories. I knew that if I wanted more information I could go back and easily pull more information if needed.

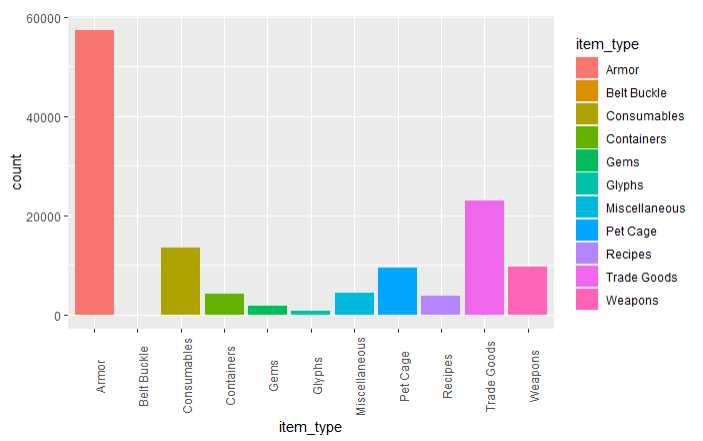
The game breaks items down into categories and some of those categories were useful and some were just numbers. This obviously was a major problem, so I went into excel to see if I could make sense of where some of those items belonged. One of the items was pretty obvious because there was exact correlation between one of the unknown categories and the amount of “Pet Cages” (Pets in World of Warcraft are like pokemon, there are a lot you can collect and you are able to do pet battles with them). The next two unknown categories were a bit harder. I came to the conclusion that the smallest category were belt buckles. I can see why this category stands out by itself because it is not technically armor and therefore doesn’t have a place. The final unknown category I could only describe as miscellaneous because the items in this category did not fit any clear pattern. In this sample set of data, I then changed the names of these categories in order to more easily work with the data. I could have gotten more information on the items from the WowHead json files, but for what I wanted to do with the data the name and category was sufficient. What is interesting and challenging about this data is that there is no metric for the auctions in order to know if these auctions actually sold or if they ended without a sale. Another challenge I noticed was that the time and date was a timestamp that I did not recognize. It took some navigating, but I learned that epoch time is used by Blizzard. I then converted epoch time into Eastern Standard time and date.

Once I had a data set that had relevant information in it, I looked for patterns and other relevant details. One of the things I focused on was the fact that the game allows you to post your auctions for different amounts of time. You can post in increments of 12 hours up to a maximum of 48 hours. I wanted to see what was the most common time increment for posted items and if there was a difference based on the server population. In order to do this I had to get more information. I then got information from two other servers. The initial server (stormrage) is considered a full server meaning the population of this server is at max capacity, the second server is considered high population (Norgannon) and the third server is considered low population (Blackhand). Having data from three servers was useful to determine if there were differences between the servers and if so by how much. It turned out that even though there was a lot less data per hour on the other two servers, the graphs and the post times looked almost identical. I decided to only compare the two extremes, the full population server and the low population server. As you can see below the graphs look pretty much identical.



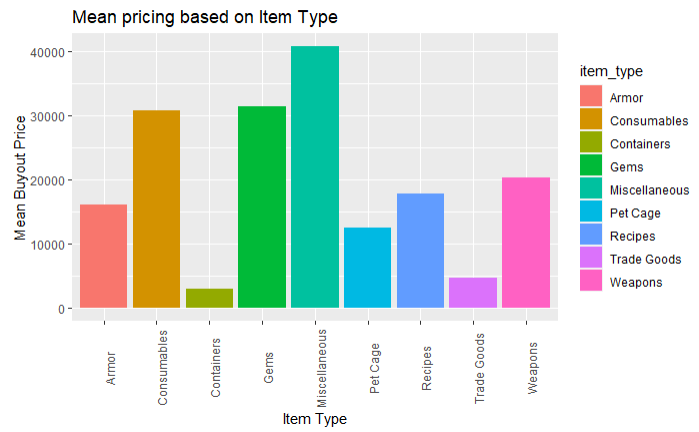


Next I wanted to figure out which item category was the most popular. I decided to only look at the full server because it had the most data per hour. That was pretty simple and it was clear that armor was the number one category. Based on my knowledge of the game I would have guessed that trade goods would have been the number one category by a large margin because the trade goods are sold in bulk. What I didn’t take into account was that this data is of individual auctions posted and you are able to post large stacks of items within one auction. For example, you can sell one piece of dragonscale armor or you could sell 200 of one herb and they would both be considered one auction. There was a quantity column in the data that could have been used in conjunction with some other columns to discover which category had the most items on sale at one time, but that was something that was beyond the focus of this project. Below is a graph of each category with the most amount of unique items being sold.



Just like eBay, users are allowed to set their own prices and sometimes they ask for much more for some items that aren’t worth the asking price. This can skew this data because the data is on the auctions posted, not on the ones that result in a sale or not. It is very likely the auctions that are posted for considerably higher prices than most auctions are never bought. Due to the ambiguity around whether an auction was successful or not, I looked into the minimum, mean and maximum prices on the second most popular item. Both the mean and maximum were very high compared to the minimum. In reality most buyers are paying closer to the minimum prices rather than the other two because a lot of auctions clearly go unsold. This gave me an additional idea to explore which I will go over later in this paper.

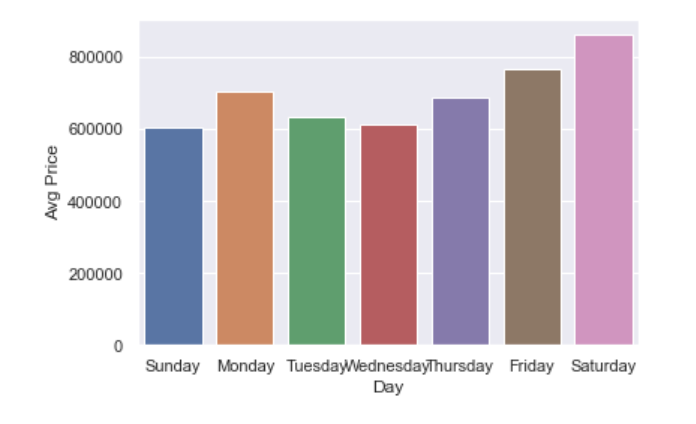
At this point I had figured out that armor was the most popular category, but I also wanted to know what was the most expensive category. I did this by finding the mean buyout price for every category and comparing these numbers. It turned out that even though armor was the most popular item, it wasn’t the most expensive by a large margin.



Answering the Initial Question and Additional Analyses

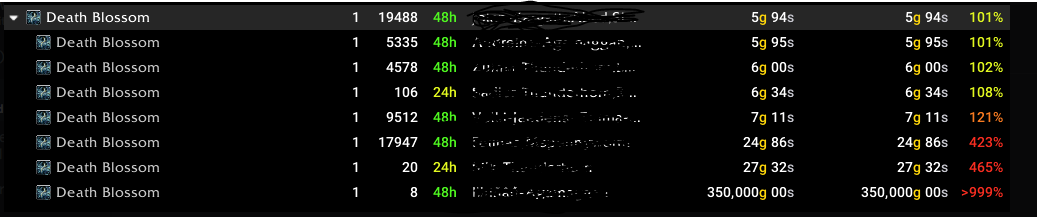
The information and observations made thus far, were from a small sample limited to one hour worth of data. The next step in the project was to do some calculations on a much larger scale. When I set up my data collection I had it automatically record data every hour. With the help of a friend's server I was able to keep it going for three months. Since I was working with so much data I decided to work with it in python and learn one of its packages called pandas. It turns out pandas is a very intuitive package that allowed me to answer my original question of what is the best day to buy or sell.

Since there are so many items on the auction house and item prices can fluctuate based on numerous factors I decided to focus on a product that I knew was always in high demand. This item is called Death Blossom. Death Blossom is a herb in the game that is used for a variety of purposes and by many different “professions”. The code created for this process allowed me to pick any item from the auction house to do this calculation. For the three months of data there were 150 thousand individual auctions of the trade good - Death Blossom. This provided a good deal of data on the average price of Death Blossom for each day of the week. Below is the result of the average prices by day of the week:

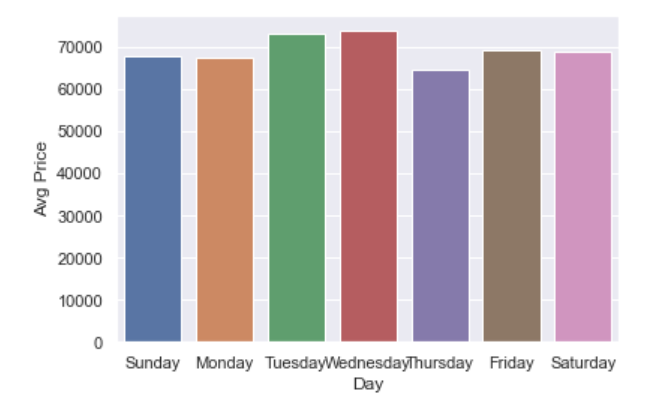


It was noteworthy that there was a significant daily price difference. Further, it was clear that Sunday and Wednesday were the best days to buy and Saturday was the best day to sell. This observation makes sense because there are likely more players playing the game over weekends. More players on, should equate to more demand and the rising prices.

However I noticed there was likely a flaw with this logic. Since I play this game, I know a few important factors that an outsider might not recognize in a cursory review of the data. Tuesday is known as the “server reset day”. This is the day that a lot of the weekly content that the game provides, resets and allows players to redo some aspects of the game for more rewards. Even though logically the weekends might be more busy, I know that Tuesdays are actually one of the busiest days of the week for the game. Based on that knowledge, I decided to refine my process to be more accurate. This brought me back to the idea that not all of these auctions sell, in fact most of them probably don’t. This is a screenshot of the auction house showing all of the Death Blossom that were on sale at the time of the screenshot.



As you can see, almost 20 thousand Death Blossoms were for sale at 400 percent over normal market price. Those 20 thousand Death Blossoms have almost no likelihood of selling, therefore I needed to figure out a way to ignore those auctions and focus on the ones that would actually sell. I decided to look at the lowest 10% of the prices because those auctions are most likely to sell. Similar to the previous analysis, the minimum prices for the Death Blossom were much lower than the mean and maximum. Even though there are a lot of auctions, only the cheapest ones will likely sell. Keeping this in mind I did the same calculation again and got a totally different result.



After taking into account “the server reset day”, this calculation made more sense. Tuesday and Wednesday were the best days to sell and Thursday was the best day to buy. Overall the days are a lot closer together than the original calculation, but this is probably a more accurate representation of the best day to buy because this analysis uses the auctions that are most likely to have resulted in a sale. Demand is high on Tuesday and Wednesday because that is essentially the beginning of the week for the game content. I was a bit surprised that Thursday is the best day to buy because Thursday is still early in the game week. There are influxes on the reset days and then again on the weekend when people are more free to play. Additional analysis could be explored by looking at different items to see if patterns changed at all. It is likely that different patterns could emerge due to the way items are used with different aspects of the game content. For example, the armor category isn’t all about enhancing the strength of a character. There is a whole market for fashion in the game as well. Some armor is bought and sold on the auction house purely based on its looks. That exploration would require more enrichment of the data to figure out some key details to see if the armor is used for fashion or for power. Pet cages are another category that would be interesting to explore because in the auction house they all have the same name, but what is contained in the cage is where the real value lies. Some pets are more rare than others and would sell for higher prices. Looking into either of these items in more detail would probably yield different results than the Death Blossom because of the different uses of these items. This allows for a lot more potential exploration tailored to specific items and content which is an exciting prospect.

Relation to Data Analytics Program

The learning goals from the data analytics program that were demonstrated in this project include “Multi-disciplinary analytic capabilities” and “Ethical Data Stewardship”. This project was based on the economics of a video game market that acts a lot like real life markets. In order to complete the project I worked with python to collect and store the data. I worked in R to do some basic data cleaning and exploration. I was able to adapt and learn how to deal with json files and python pandas. I was able to read big data and answer my main question with these new techniques.

My experience from the data analytics classes prepared me for this project because I had an essential understanding of a lot of the tools I used in the project. I used R, python, and jupyter notebooks which were all tools I learned throughout the program. Looking back at the project I could have set up a SQL database in the beginning to collect and store the data more efficiently. I am happy with the range of tools the College provided allowing me to effectively use them to complete this project. Throughout this whole process I was able to remain ethical by not using data that would harm individuals or companies. There were parts of the data that could be used to trace back to individuals. I made sure not to include that data in my calculations. It is important to be careful with information. We need to determine and be mindful of what could cause harm.

Overall this project was fun to do because I was able to work with data that is very relevant to me. I know I only scratched the surface of what could be done with the data, but I am happy with what I found and I will be able to use this knowledge in the game. If I continued working with the data from this project, I would try to figure out what time of day was the best time to buy and sell goods. I would also look into some item niche’s to see if their sell patterns were any different from Death Blossom. There are thousands of items in World of Warcraft and I only looked at one that has a high turnover rate. It would be interesting to look into items that are used much less frequently. This project gave me a new appreciation for economics and being able to find trends in data. I can easily see myself working with similar data in the future and can only hope that my job will be as interesting as this project.