Eve Online is a Massive Multiplayer Online Role Playing Game. Players are set in the far future and tasked with piloting spaceships to make money (ISK) and do whatever they desire. Players are able to come together to form corporations, then corporations of like minded players form alliances. Eve online has a completely player driver market, that means everything is created and destroyed within the game with very few external injections. The space in Eve online is divided into two main categories, High security space which is safe with an ingame police force and Null security space where outlaws roam free. Each of these groups are then further broken down into regions then constellations, then individual solar systems.

The data analyzed was all market data for the year of 2021 with the following attributes:

Type\_id: The specific code associated to each unique item

Region\_id: The specific code to region of space

Volume: The amount that a specific item was traded in one region

Order\_count: How many individual buy and sell requests were placed on the market combined

Highest: The highest price an item sold for

Lowest: The lowest price an item sold for

Average: The average price an item was sold for

I ended up adding the following attributes for later use.

Name: The ingame name of each region correlating to the Region\_id

Type\_name: The ingame name of each item correlating to the type\_id

Highlow\_difference: The difference between the highest and item sold for and the lowest

In the dataset there were 49,032 values each with the 10 attributes.

Initial plan for Data Exploration:

Initially I had three main goals. Find a way to look at all items in one specific region of space, find the data for one item across the entire universe, and find the greatest highlow\_difference for an item.

Actions taken for data cleaning and feature engineering:

Thankfully the data I pulled was clean of any human errors since it was pulled directly from the game itself. This was the point that I had added the names for regions and items as well as the highlow\_difference attribute. When it came to outliers I had my first big stop.

The data is broad. In my volume attribute some items varied from 1 to 2.4 billion, average varied from .01 to 30 billion ISK. Then just because something was an outlier in one attribute didn’t mean it was in another. I determined finally that I would not remove any values from my data, and that as I analyzed my specific hypothesis I would check for outliers there.

When it came to feature engineering again I did very little, I simply converted the High/Null values for the type of space to 1’s and 0’s.

Key findings:

My first question was just how much ‘volume’ was moving through the universe and where it was going:

Total volume: 11,64,051,792

High security space volume: 10,004,443,763 (70%)

Null security space volume: 1,639,608,763 (30%)

Of that high security space volume 6,174,394,724 (53%) occurred in the major trade hub of the game, and 689,707,957 (5%) occurred in the null security space where the current the current largest player alliance of over thirty-one thousand players call home.

Once I knew a distribution of where stuff was moving I was then able to further index into the dataset and find out some really interesting stuff such as what type of fuel the outlaw player groups were using, which types of ships the high security players were using, and what items yielded the greatest returns for trade haulers traveling from region to region.

Formulating 3 hypothesis about the data:

My big hurdle here was my data was over the course of a year, so I wasn’t able to test my initial pre-hypothesis of tracking market trends so I had to shift to something that my data could actually tell me.

1. More trade volume moves through high than null security space.
2. Player alliances in null security space sell cheaper critical items to their alliance than high security trade hubs?
3. Items with a lower trade volume have a larger high/low difference?

The first one went pretty smoothly. I pulled the total volume of trades in high and null security and quickly saw that 70% did occur in high security space. This makes sense as it is safer to move items around and that all major trade hubs are in higher security space.

The second one was more interesting with sixteen thousand unique items. I had to bring my sample sizes down to test this. I reached out to two ingame players and got their opinions. I then had three picked items that were important for outlaw alliances and three random items to compare.

Of my three recommended items, they appeared in null security space 37% more often and 32% cheaper than in high security space. While the three random items appeared in high security space 73% of the time and were on average 91% more expensive in null security space. I took this hypothesis as plausible, I think future testing would show that items that provided infrastructure would prove yes but that is only a small sample of all sixteen thousand items in the game.

The third and final hypothesis was where I conducted my formal significance test.

Formal significance test:

Going into this test my research hypothesis was “Items with lower volume with have a higher high/low difference. With the Null hypothesis being “There is no relationship between trade volume and high/low difference. In order to determine “High” and “Low volume I picked the outer 15% of all volume. I decided to use a one tailed t-test for my statistics.

| High/Low Difference | Low volume | High Volume |
| --- | --- | --- |
| Number of observations | 7,335 | 7,335 |
| Mean | 6,259,450 | 1,223,207 |
| Standard Deviation | 32,189,369 | 8,554,7776 |
| Variance | 1,036,172,890,788,937 | 73,184,194,834,541 |

t= 12.949, df=14,668, p<.05

Since my t value had to exceed 1.96 we accept the research hypothesis as true and reject the null hypothesis.

Next steps for analyzing the data:

The next step for analyzing this specific set of data would just be finding out the right question, you could pull tens of thousands of different observations from the data that is there. After this project I plan to go back to talk to some high level players and see what information would be relevant for their causes.

I would like to in the future take some smaller chunks of the data by specific region and on a months time scale instead of a whole year to try and actually make some predictions about future prices. This would also allow me to see the impact of game updates and player wars on the market.

Final notes:

As far as the quality of the data set goes, I don’t think it could be improved just broken up into smaller chunks based off of months instead of a whole year. Also you could group the items into groups, like ships, weapons or minerals for quicker analysis of specific groups. For someone who didn’t know a thing about statistics before this class the final step was difficult and I learned a lot on the way. Overall this took me about 27 hours of work, I would say 75% of that time was researching how to do one specific task, whether that was making plots, merging CSV files, or the statistics and it was time well spent.

References:

-<https://everef.net/data>

-<https://evemaps.dotlan.net/>

-Experts:

1.NCC 1701E- A player who has been playing eve for 17 consecutive years and currently holds the all time rank of 445, a veteran null security Fleet Commander who is currently the CEO of the corporation Criterion in the alliance Pandemic Legion.

2. Blazingbunny- A player who has been playing eve for 13 consecutive years, a veteran null security Fleet Commander and tactician. He has led player fleets of all classes of ships into battles, scaling from 50 players to over 2,000.

3. Typherian- A player who has been playing eve for over 14 years, he currently holds the title of head logistican for the null security corporation Criterion.

4.Ishbuanium-A player who has been playing Eve for 16 years, and currently holds the title of Capital ship acquisition for the Null security corporation Criterion.