



Final Project – ISM2410C

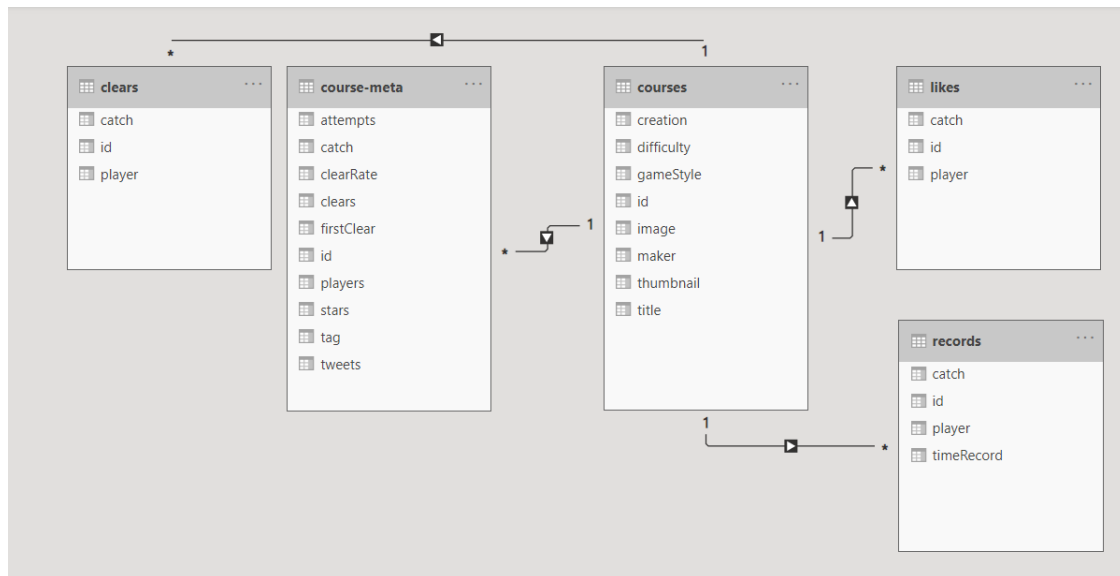
Choosing and Connecting a Data Source

I chose to use data from a video game I am very familiar with called Super Mario Maker. The data was pulled from [Kaggle.com](https://www.kaggle.com) and was a zip that contained multiple CSVs. On the Power BI desktop version, I uploaded each CSV and checked for data consistency, such as date & time formats, so that I would not have a skew in pulling groupings. I liked that there were multiple sheets that linked many-to-many columns, it made for a broader range of aspects to dig into.

I started by testing a few charts to look for trends and outliers, this helped me look for what I wanted to finalize and conclude.

Using and Transforming the Dataset

I wanted to be able to visualize the statistics on clears (or 'wins') on top levels, player (or "Maker" in this community) popularity rates, and overall time averages. Since this data came in very clean from the video game system, there was nothing really to transform. I did have to check on my relationships between CSVs, however, which were lacking some links:



I then created two new columns under the 'clears' sheet so that I may link the player data to their regions. I first split the 'Player ID' to just have the first 4-digits (using LEFT function), then used these to identify their regions/continents (I had to make up fictional regions for this, but it worked out). This assisted in showing the map viz of where the levels were created. I was then able to add a slicer for difficulty level and another for game type.

I wanted to see the ratio of stars awarded to number of players so I could then compare against difficulty and time to beat. Because of this, I added a calculated column for "Stars Awarded" (# of stars on level / # of unique players). I then used this in the visual along with the difficulty. Because the difficulty column was originally a text string, I created an additional calculated column that transformed the levels to number rankings. Example Easy = 1, Super Expert = 4.

I ended up analyzing the Average Clears by month in which I then linked a number card to explain the highest clear month. This also worked with the slicers for more of an in-depth analysis.

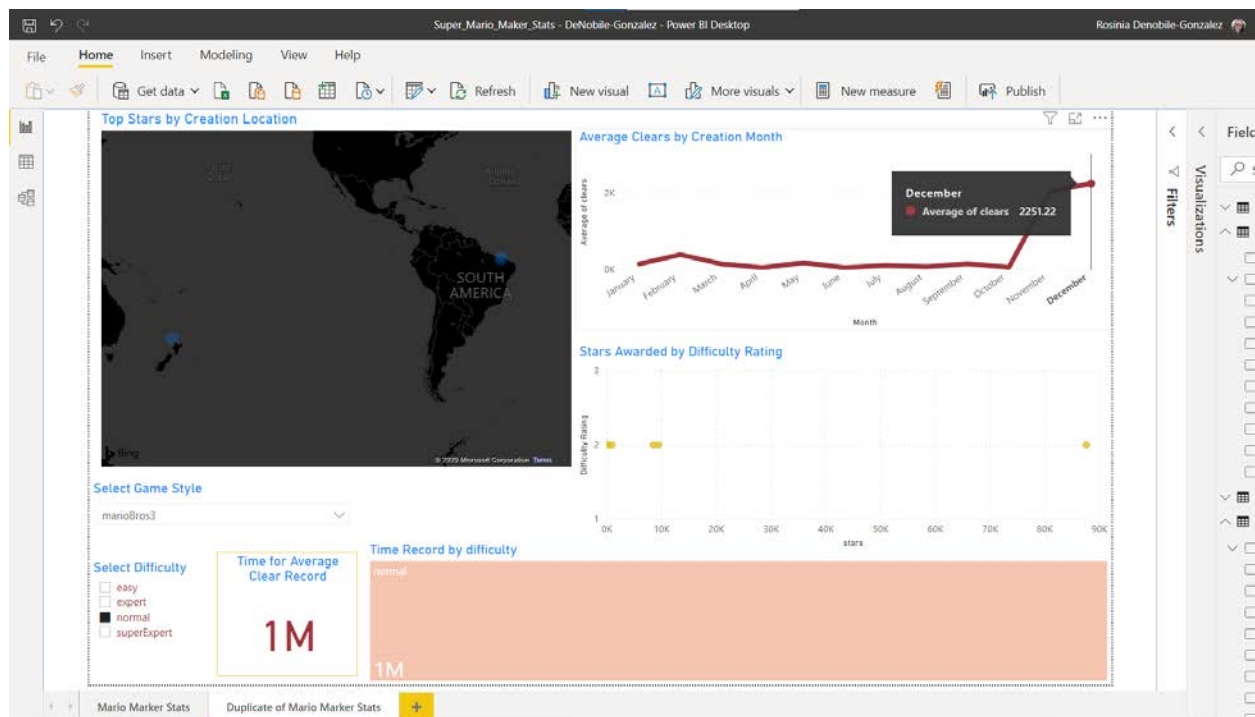
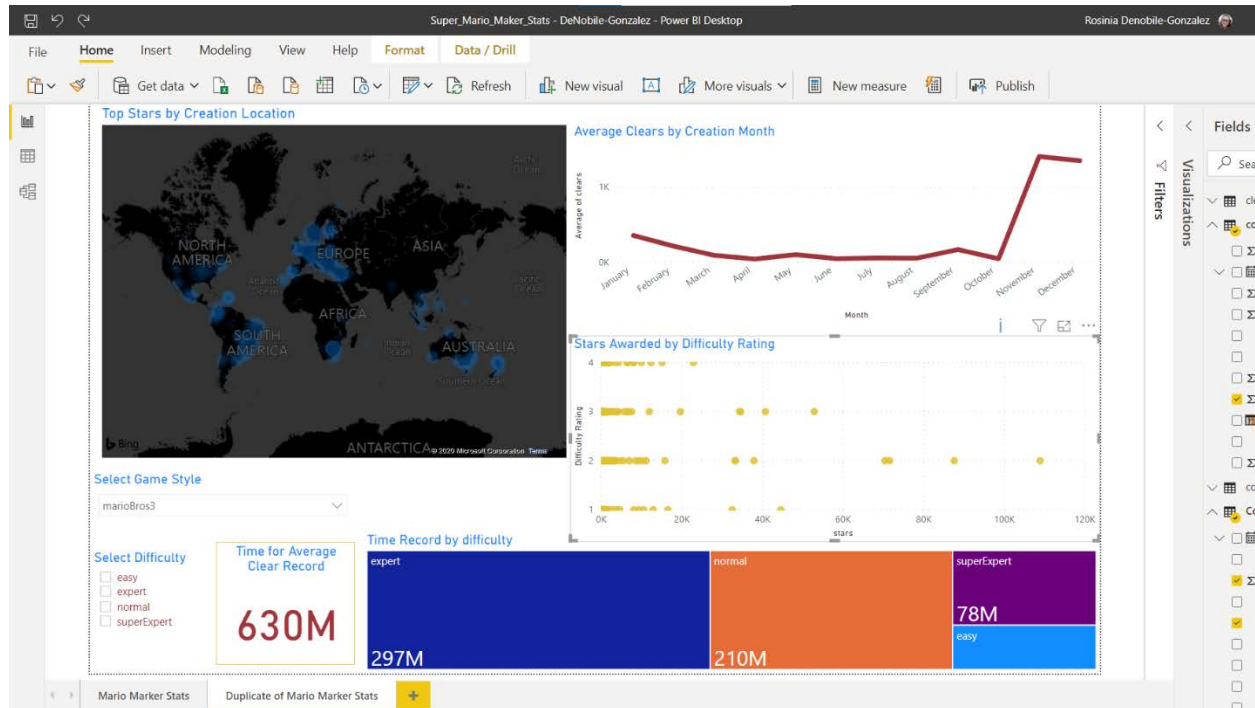
Using the Dataset

After pulling all the data together and playing around with the filters and vizs, I came to a couple of conclusions. One conclusion was that the average time to beat and difficulty had a direct relation to the amount of stars awarded to the level.

The other was the difficulty level and regions directly correlated. Most of the Expert and Super Expert levels originated from South America and Europe. North America ranged from Normal to Expert but took longer to beat. This makes me wonder if the individual sub-communities in these countries play a factor in the type of levels they create.

As a video game developer, you could create specific downloadable content and online competitive features geared towards these types of level creators. Outside of this game, Nintendo can base new regional games off of the level styles that are most abundant in these regions.

Here is my dashboard in action!



Thank you for an amazing semester, Professor Lee! I had a blast doing this project!

