Eric Dishmon

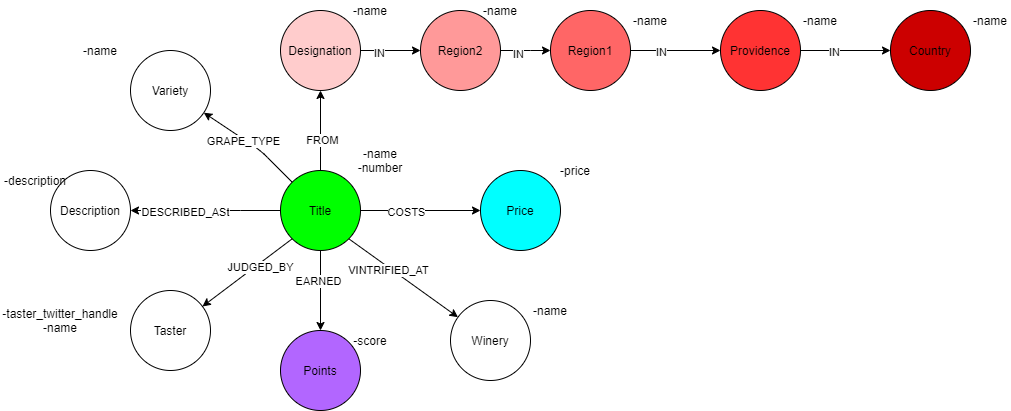
7/08/2021

**Business Case:**

As the owner of a restaurant, providing the best experience for my patrons is on the forefront of my mind when making decisions. Considering different factors from the ambience, to the food, to the overall price of a meal, customer satisfaction is very important. That is why I want to perform graph analytics on the Wine Reviews dataset to be able to provide my patrons the best wines in terms of highest ratings and lowest price. It will also be interesting to see where these wines are grown to be on the lookout for other wines from this area or to try to establish a relationship with the specific wineries. Increasing customer satisfaction will translate into more brand awareness, customer loyalty, and higher sales.

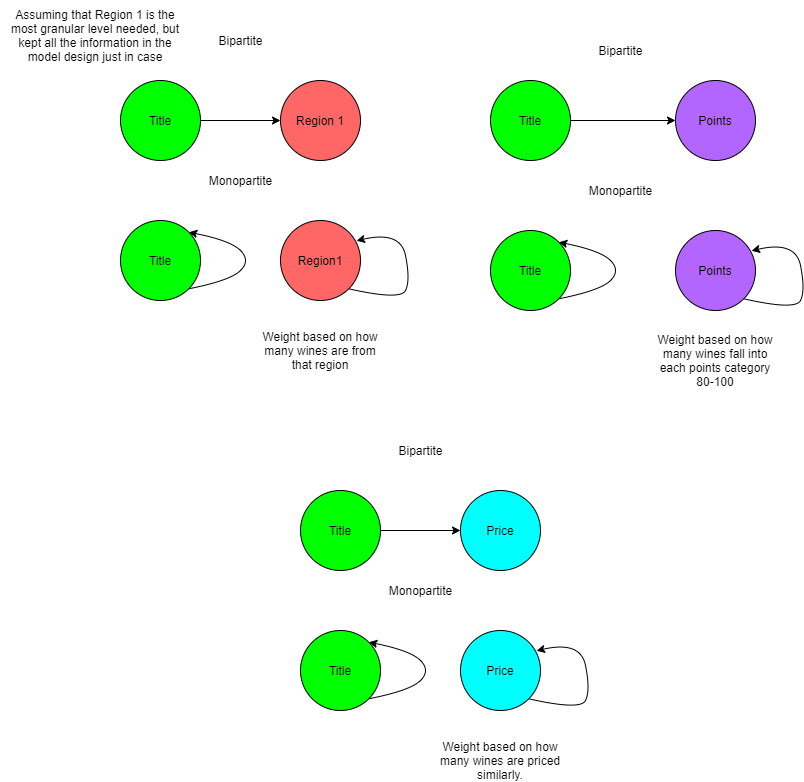
**Data Model:**

The data model for the Wine Reviews dataset keeps most attributes only two hops away from each other. The long chain of location based nodes can be rolled up into one node based on how granular the location needs to be. This will help keep the scope of the location flexible depending on future needs. I have also color coded the nodes of importance while leaving the other nodes monotone so that it is easier to identify key nodes when viewing the data.



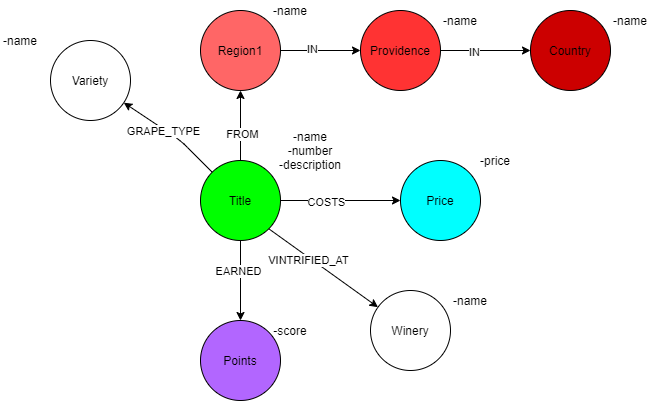
**Projections:**

For the projections, I have included three monopartite examples from the data model covering the three nodes of importance. I chose the Region 1 node to focus on because it allows for a more granular look at where the grapes are grown without being hyper specific to locality. The Points and Price nodes are also projected as monopartite with the Title node because it could be important to analyze the weights of how similarly priced and scored wines are to each other to determine the best valued wine. The Price node may need to be changed into larger buckets because of the large amount of variation in prices, but that will be determined once modeling begins.



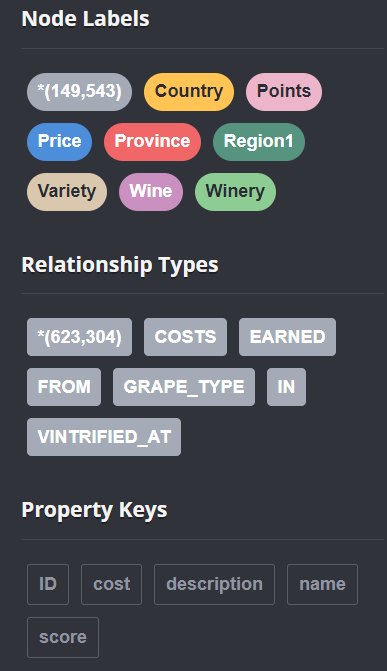
**Model Updates:**

I have made the below changes to the data model. This updated model allows for easier querying and algorithms by reducing unnecessary nodes, such as Taster, and connecting nodes with less null values, such as Region1, to create a more connected model.



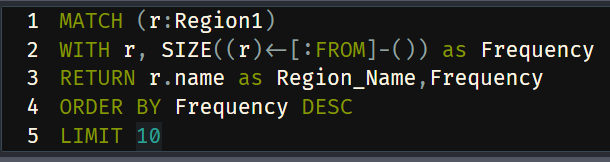
**Database Setup:**

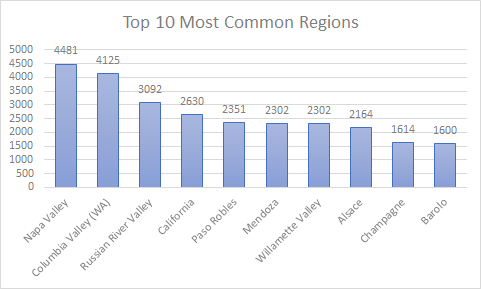
Below is a screenshot of the database setup. This model has 8 different node types and almost 150 thousand nodes in total.



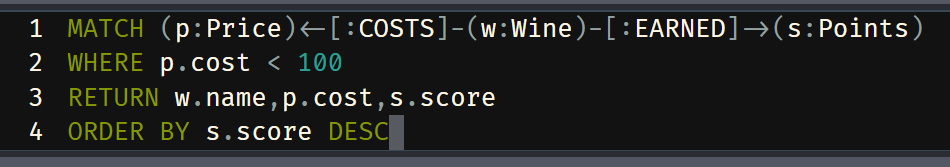
**Cypher Queries:**

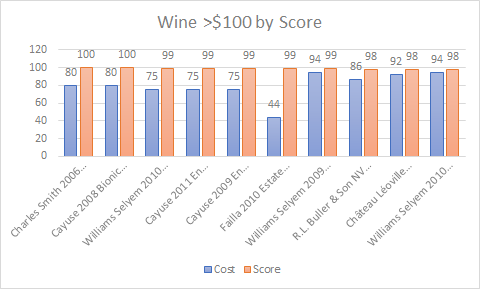
Query 1: The first query shows the top 10 most common regions from where wine comes from. This is useful to determine where most of the wines with a rating of 80 or higher are coming from to be able to get a general idea of where the “wine hotspots” are located geographically. Potentially opening a store near these regions could create an easy symbiotic relationship between the wineries and the restaurant.





Query 2: This query shows the wines that are under $100, a price limit for the restaurant to keep costs low and wine prices low as well, and ranks them by their score. This full list will show all of the options for the restaurant to choose from, narrowing down the results to about half. Shown below are the top 10 results.





Query 3: The final query was found first by determining the median price of the wines, $25, and then ranking the scores by for the wines that are the median. This helps the restaurant get a sense of where an average priced wine falls on the points scale to help determine what it should expect for quality when making its selections.

