Eric Dishmon

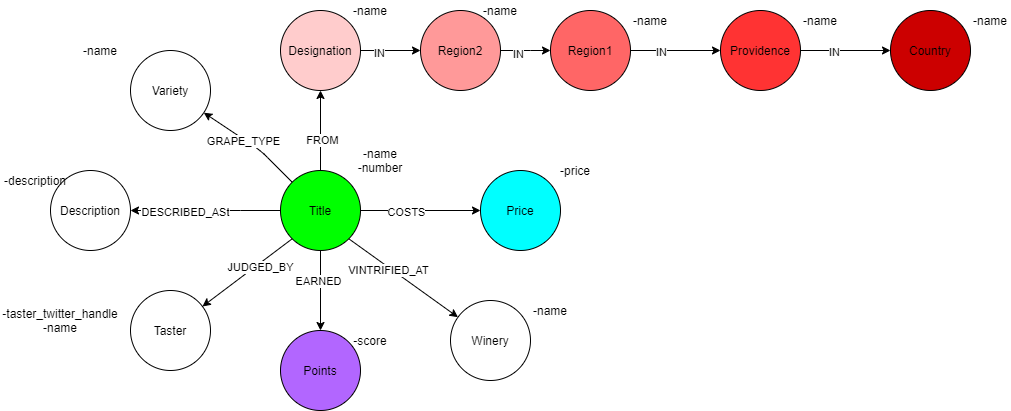
7/25/2021

**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.

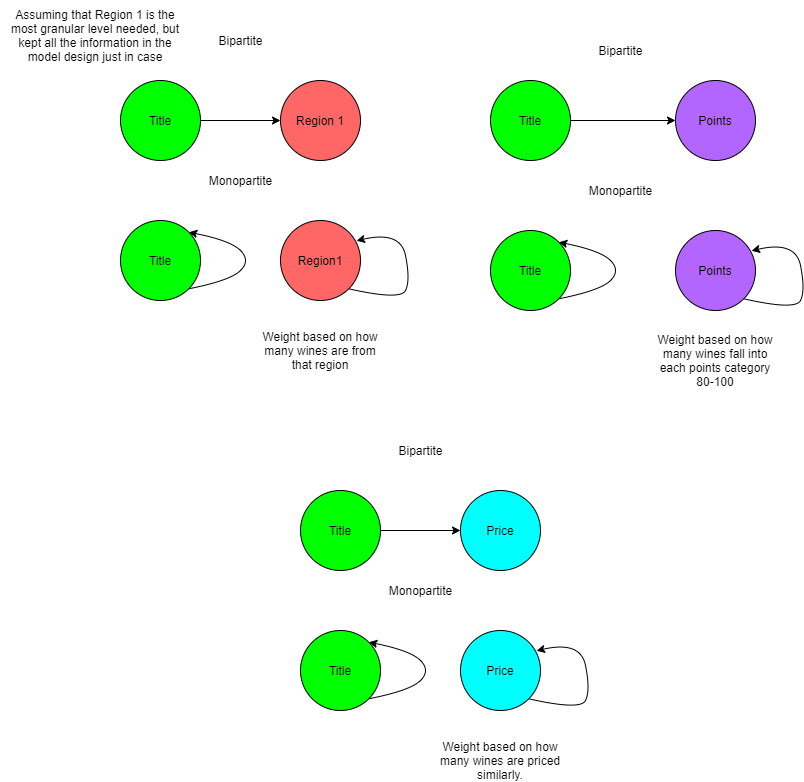
**Data Model:**

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.



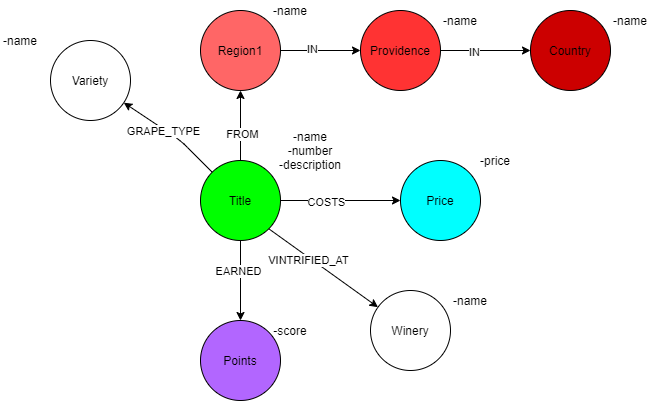
**Projections:**

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.



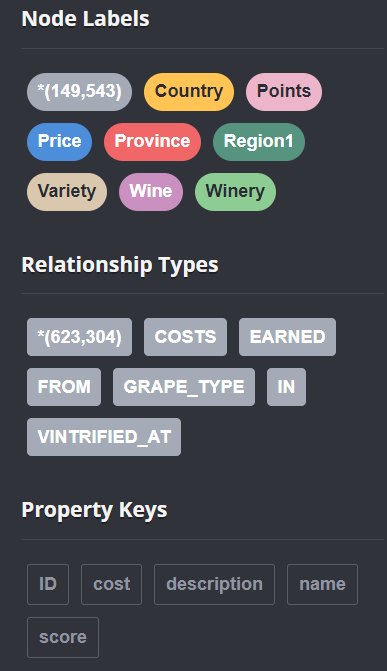
**Model Updates Project Deliverable 2:**

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.



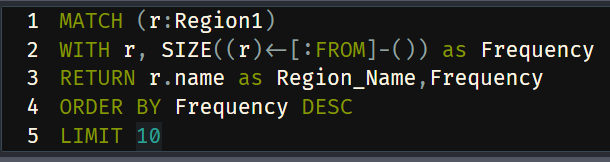
**Database Setup:**

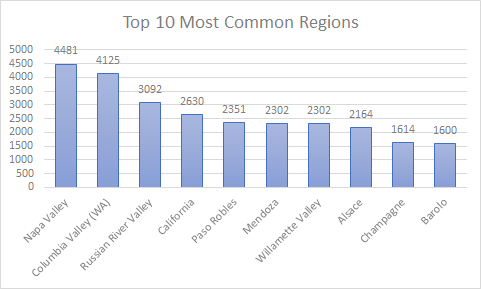
Below is a screenshot of the database setup. This model has 8 different node types and almost 150 thousand nodes in total.



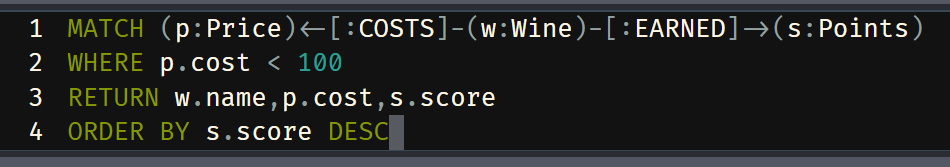
**Cypher Queries:**

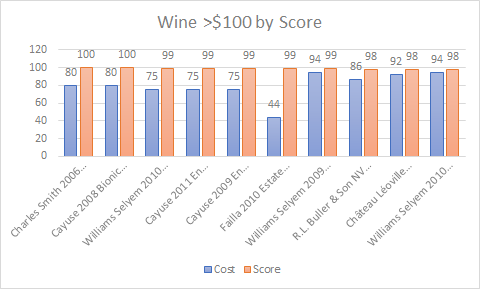
Query 1: The first query shows the top 10 most common regions from where wine comes from. This is useful to determine where most of the wines with a rating of 80 or higher are coming from to be able to get a general idea of where the “wine hotspots” are located geographically. Potentially opening a store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



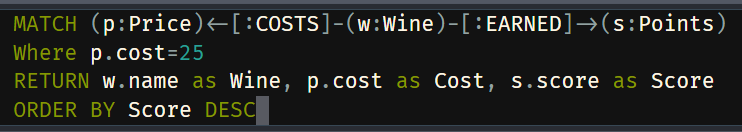


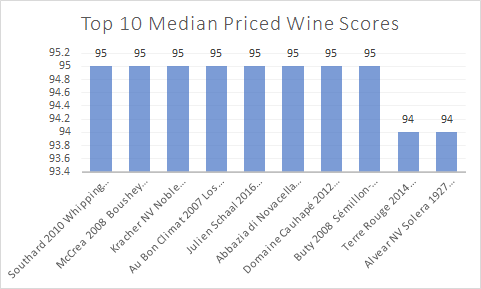
Query 2: This query shows the wines that are under $100, a price limit for the restaurant to keep costs low and wine prices low as well, and ranks them by their score. This full list will show all of the options for the restaurant to choose from, narrowing down the results to about half. Shown below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





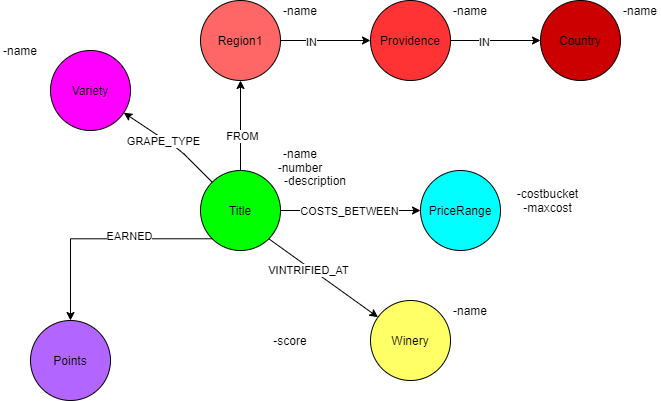
Query 3: The final query was found first by determining the median price of the wines, $25, and then ranking the scores by for the wines that are the median. This helps the restaurant get a sense of where an average priced wine falls on the points scale to help determine what it should expect for quality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



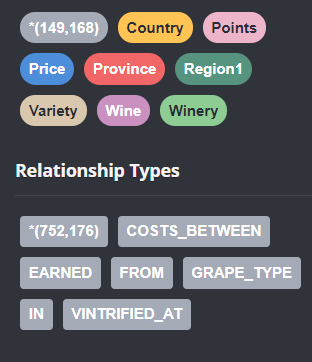


**Model Updates Project Deliverable 3:**

The price node in the data model was changed to include price buckets, with maximum price as a property to create less nodes so that the algorithms and queries can be run more efficiently. I chose the price buckets to be from 0-10,11-20,21-30 dollars ect. until 131-140 dollars and finally a bucket that was greater than 140 dollars. I chose this as the cutoff point after analyzing the distribution of prices of the wines and above 140 dollars there was a significant dropoff in volume. I chose buckets of 10 dollars just for ease of dividing.



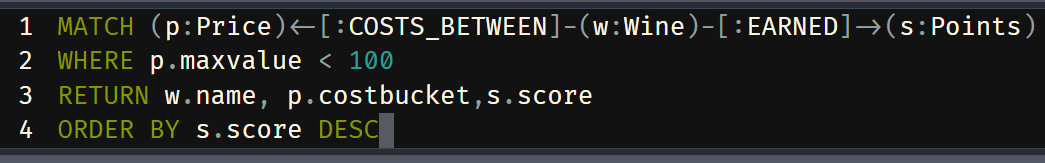
**Database Setup Update:**

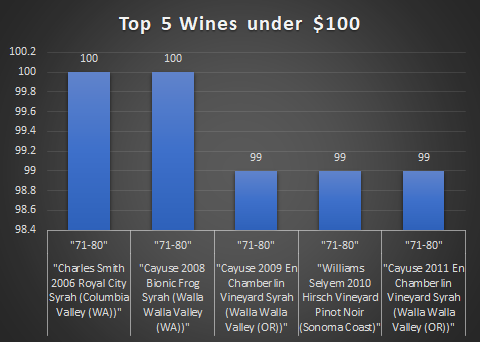
****

**Query Updates:**

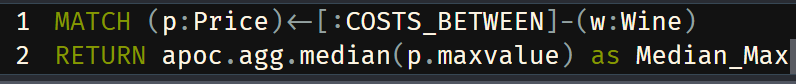
Queries 2 and 3 were updated to reflect this change in the model.

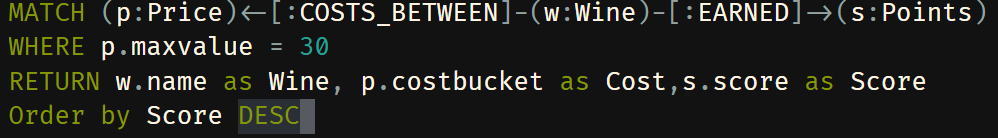
Query 2:

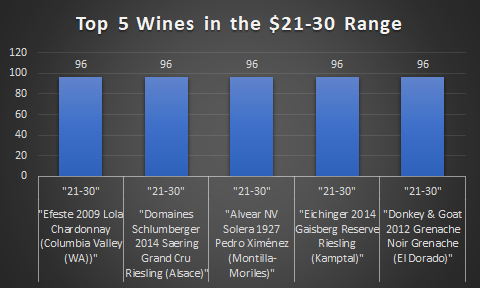




Query 3:



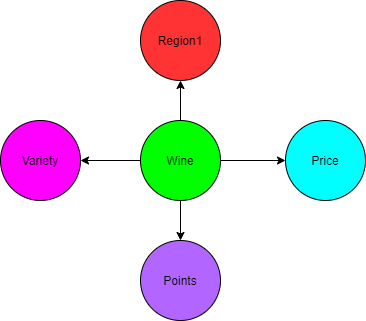




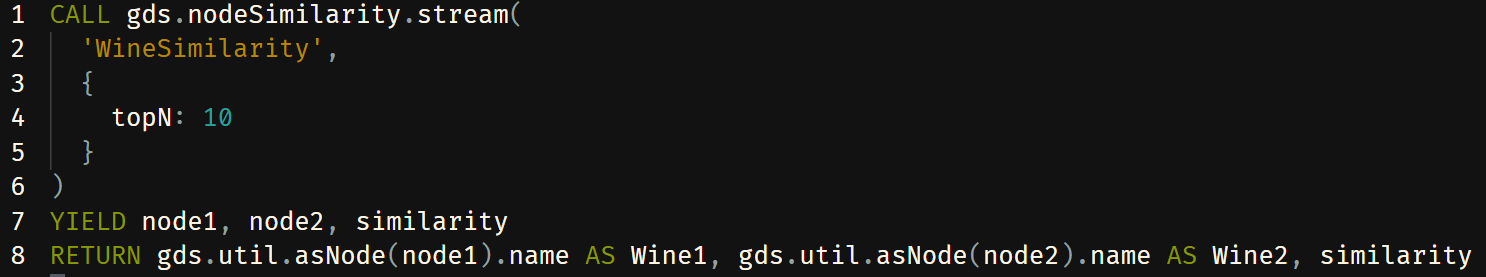
**Algorithms:**

**Algorithm 1:**

Projection:



Code:

****

Results:

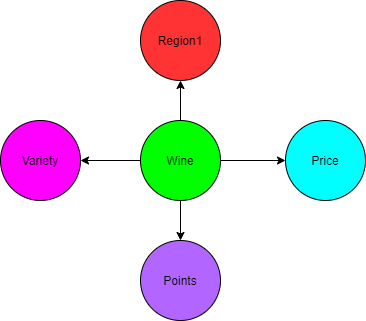


Takeaways:

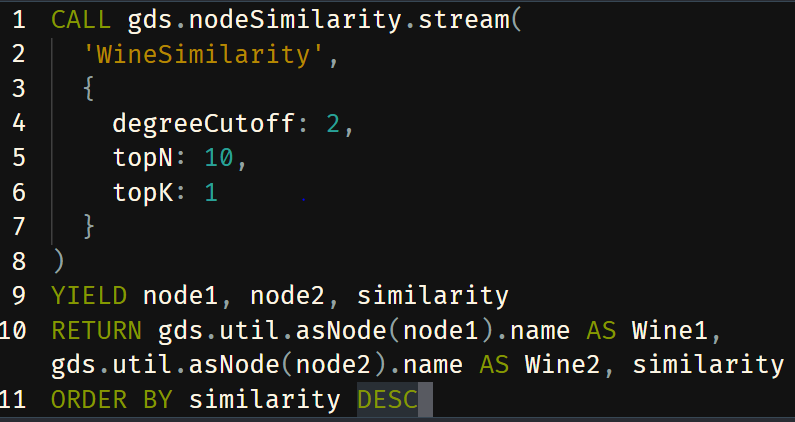
This algorithm is useful for finding wines that are similar to each other based on the location, variety, price, and score. This is very useful for when you find a specific wine that I like and want to find wines that are comparable. I can use this to find wines that may be slightly cheaper, but come from a similar area, type and score. This algorithm was limited to the top 10 for processing power limitations, but would ideally be all inclusive.

**Algorithm 2:**

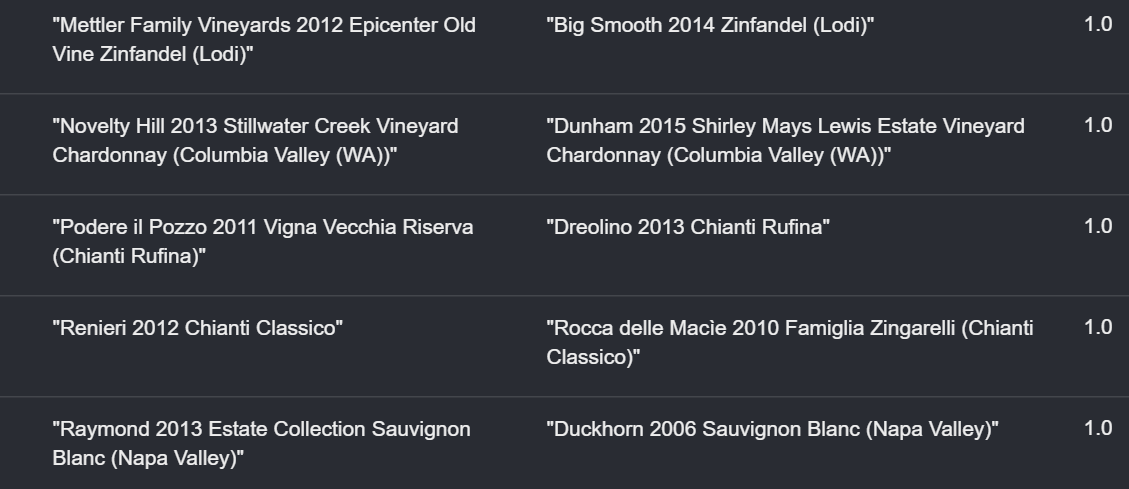
Projection:



Code:



Results:

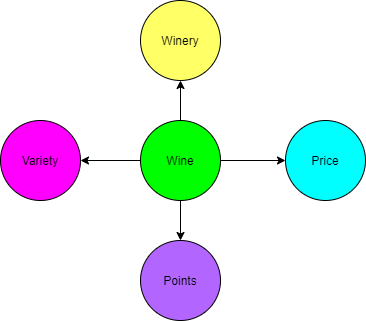


Takeaways:

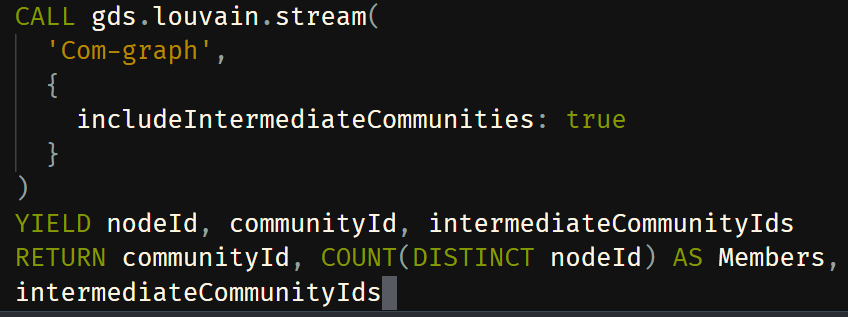
Like the algorithm above, this algorithm finds the similarity between two wines based on location,price,score, and variety, but this algorithm finds the most similar wine instead of creating a list of every wine and how it compares to each other. This is useful to find the most similar wine for every wine in the dataset. Again, it only shows the top 10 because of processing limitations.

**Algorithm 3:**

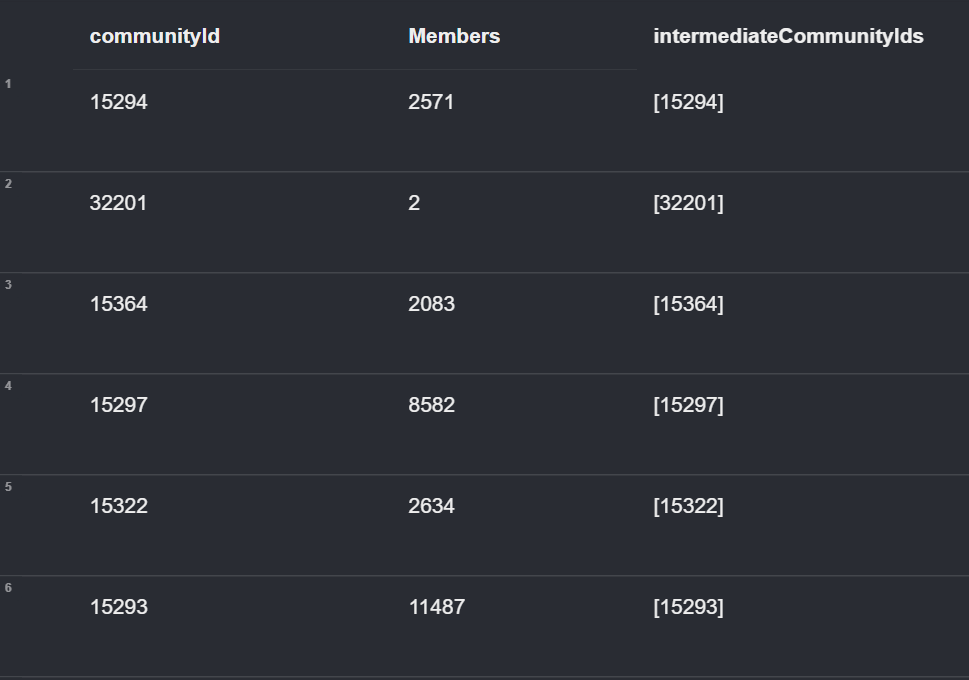
Projection:



Code:



Results:



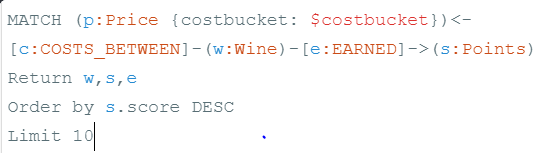
Takeaways: This algorithm assigns communities within my database. This is great for identifying wines that are similar to each other based on the same criteria as the first algorithm. This is different because it does not compare each wine to another wine, but rather just gathers the most similar wines together. I can use this to find more wines than the similarity algorithm that are similar, but the similarity algorithm will show me the closest wine to the one being compared. In short, this is better for identifying groups of wines and similarity is good for identifying the closest wine to a wine.

**Cypher Search Phrases:**

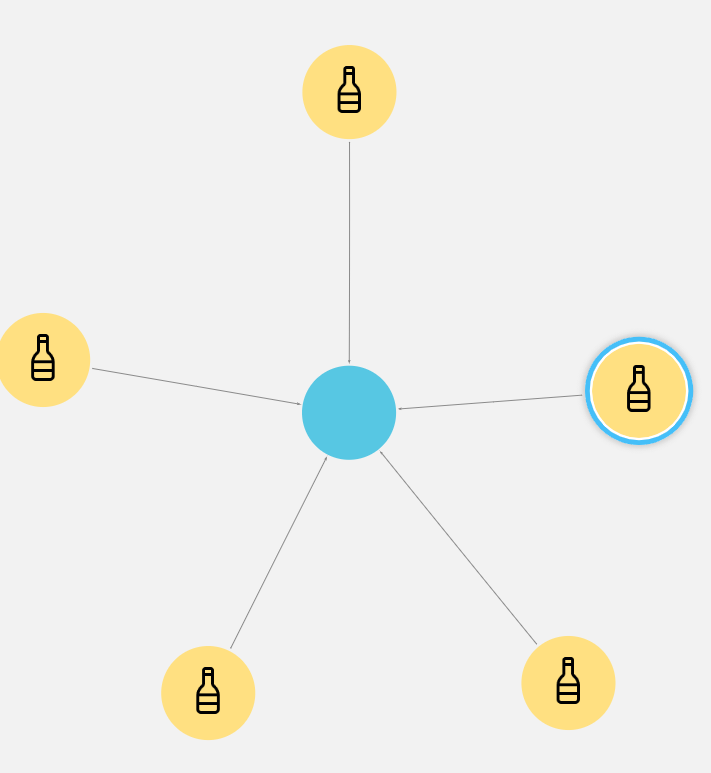
Phrase 1)

Search phrase: Best Wines in $costbucket Price Range

Description: Returns the top ten highest rated wines in a certain price range and their score



Example: This phrase is very useful to be able to quickly search for the best wines in a certain range. I will use this when deciding on which wines to purchase for the restaurant in different ranges to offer a variety of choices for the customer. After finding the wines, I can also explore more details about the wines to get a better understanding of the product.

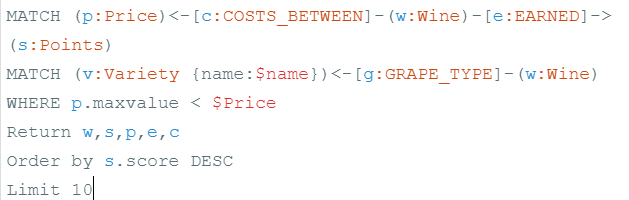


Phrase 2)

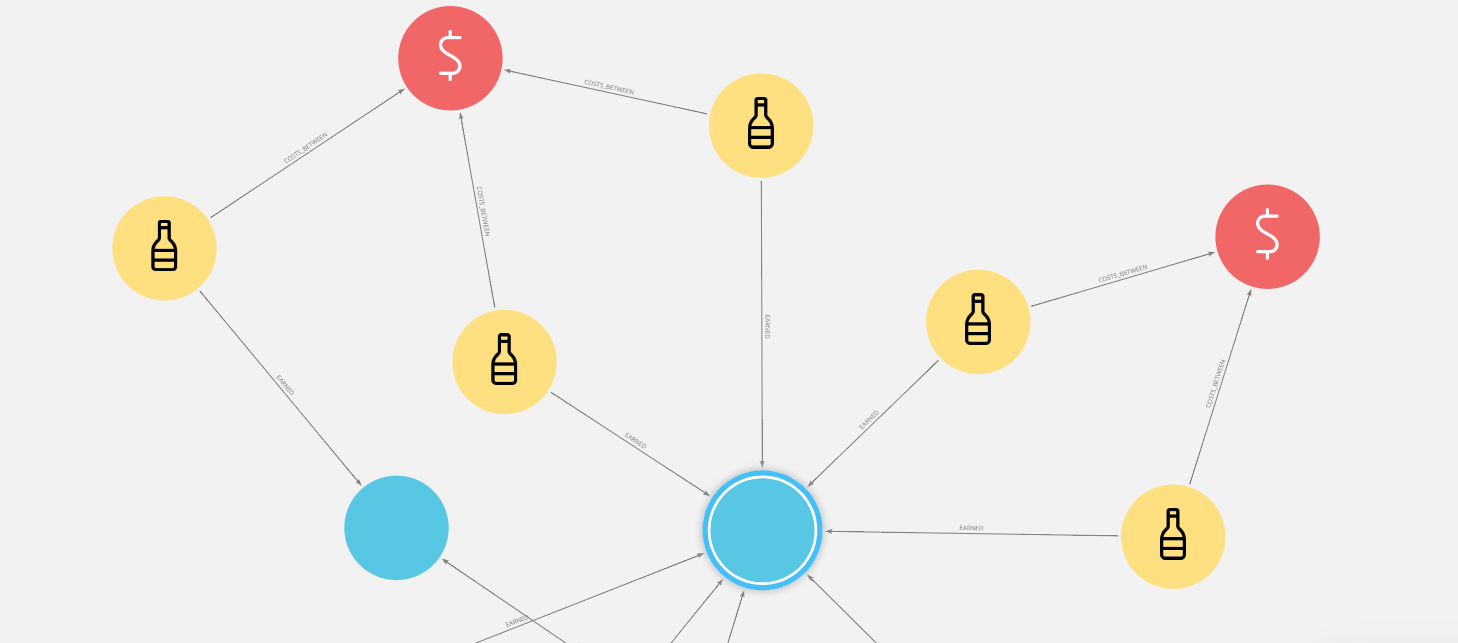
Search phrase: Find the Best Wines in the $name Variety under $Price Dollars

Description: Finds the top ten highest rated wines per variety that are under a certain dollar amount based on maxvalue for pricebuckets

Code:



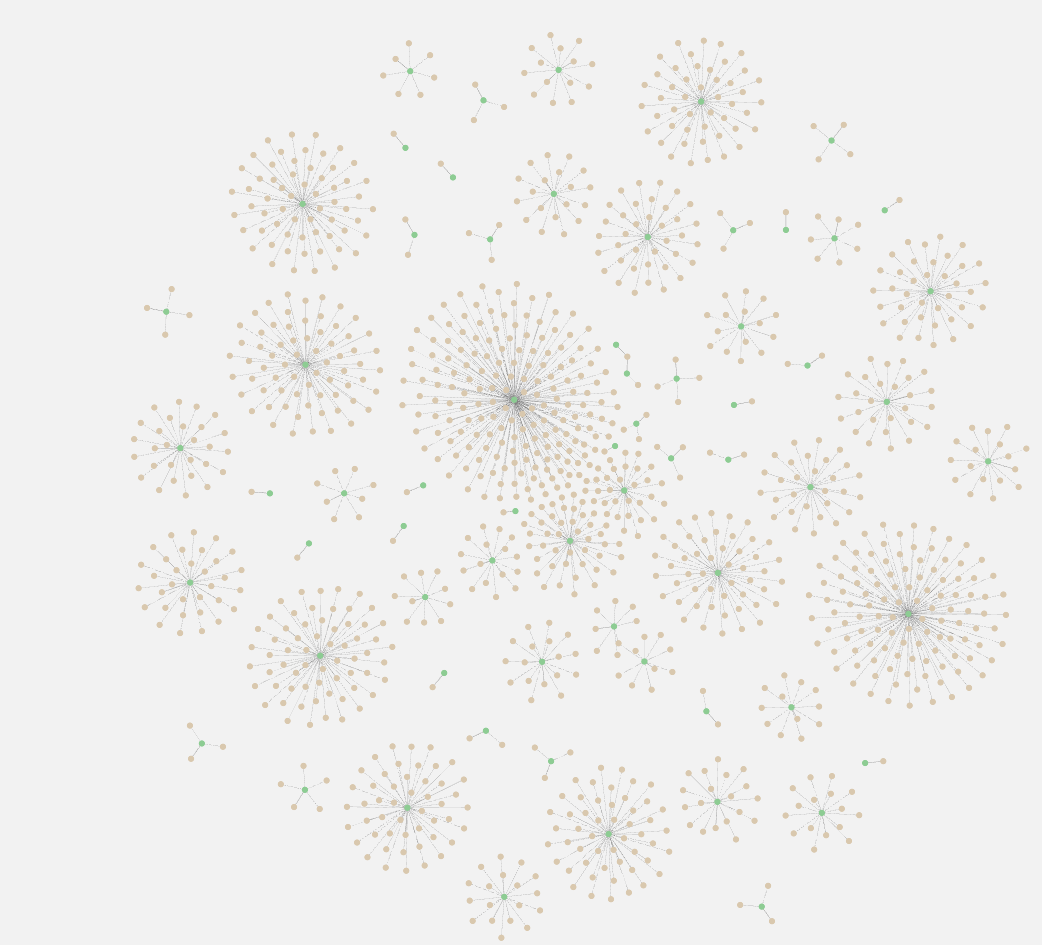
Example: This search phrase is useful because it allows me to quickly find the best wines under a certain price for a specific variety. If I need something that pairs well with a white meat fish, but I don’t want to spend more than $55 per bottle, I simply search and find the top wines that fit that description.



**Visualizations:**

Visualization 1)

This is a visualization of the different regions within each province. From afar, it makes it easier to see which provinces have the most regions and the least regions. I can use this information to focus on provinces that have the most regions because although this dataset varies in scores and prices, all of these wines are above an 80 on a scale of 100, so they are all good wines. I can use my time more efficiently by visiting and exploring wineries in these different provinces/regions to create a relationship with them while not having to travel all over the world.



Visualization 2)

This smaller visualization looks specifically at two wines. These two wines are the highest rated wines, with a score of 100, in the lowest price range. These are the wines that I will purchase for my restaurant for opening nights where I want to impress all of the guests. These two wines will not only leave a lasting impression, but will also not cost as much as other wines with ratings of 100. I can easily contact the wineries to order a few cases for an upcoming opening.

