Data Analysis Assignment- Team 3

1. **What did you learn about data mining? Identify three concepts.**

* Data mining is identifying valid, novel, potentially useful, and ultimately understandable patterns in data.
* Searching for relationships, patterns, and trends not known to exist or not visible.
* Providing answers to questions decision maker not thought to ask.

**Concepts:**

From the data mining exercise with the coffee shop data, three key concepts learned are:

1. **Clustering**: This involves segmenting the data into groups where members of a group are more like each other than to those in other groups. For instance, customers could be clustered based on their coffee preferences and visit times, revealing patterns like professionals preferring venti sizes in the morning.
2. **Association Rule Discovery:** This technique helps in finding relationships between different variables. For example, an association rule might reveal that customers who buy pastries also tend to order grande size coffees, indicating a market-basket pattern.
3. **Characterization:** This is about summarizing and generalizing the data. In our exercise, characterization could involve summarizing the average income level of customers who stay longer in the store, or the most common hair colour among students who visit in the afternoon.
4. **Identify two other things you would need to know about either the data or the business if this were a real-world application.**

In a real-world application, beyond the data provided, it would be essential to know:

1. **Customer Feedback:** Understanding customer satisfaction levels, preferences, and feedback on the coffee, pastries, and overall store experience. This qualitative data can complement the quantitative data we have, providing insights into areas for improvement or expansion.
2. **Competitor Information:** Knowledge about the offerings, pricing, and customer base of competing coffee shops in the vicinity. This can help us to understand the competitive landscape and identify unique selling propositions (USPs) or areas where our coffee shop can differentiate itself.
3. **What, if anything, would you do differently to perform this data mining task? Note: This answer should relate to the in-class exercise, not generic data mining.**

**Incorporate Time-Series Analysis:** Given the "Time of store visit" data, a time-series analysis could provide insights into peak hours, customer flow patterns, and seasonal variations in coffee preferences or pastry sales. This would enable more targeted marketing, staffing, and inventory management strategies.