Phase 1

Domain:

The domain explores the distribution of females in management positions, compared to their male counterparts.

Investigative Questions:

1. What is the relationship between provinces and territories' GDP and the ratio of males to females that occupy management positions in those provinces and territories?
2. What is the relationship between the ratio of females occupying middle management and senior management positions and the number of married couples by province/territory?
3. What is the relationship between females occupying middle management and senior management positions and the number of females that enroll in postsecondary institutions vs the number of males that enroll in postsecondary education?

Location for Data Retrieval:

1) Proportion of female and men employed in management positions, annual:
   https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410033503

2) Gross domestic product (GDP) at basic prices, by industry, provinces and territories (x 1,000,000):

   https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3610040201

3) Marital status and opposite- and same-sex status by sex for persons aged 15 and over living in private households for both sexes, total, presence and age of children, 2016 counts, Canada, provinces and territories, 2016 Census – 100% Data

   https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/fam/Table.cfm?Lang=E&T=11&Geo=00

4) Postsecondary enrolments, by registration status, institution type, status of student in Canada and gender
   https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3710001801

Choices in scraping data

Relevant vs Irrelevant Data: We will aim to look at 2016, as our year of relevance, and we will discard all other years to focus our data. We also have data available for 2016 across all our data sets.

For the third data set, we will ignore the categories for same sex relationships, as we are only concerned with the number of couples that are married in general, to be able to make an association about how locations with high marriage rates affect the number of females that are in the management work force.

For all these data, we will look at by-province/territory results, so that we can understand how concerns of particular areas affect its population.

Cleaning up: By categorizing the information into relations, we are able to extract only the data that we need. See "Relation" section below for the information extracted.

For the data set that denotes the number of married couples, we will have to extract a ratio, from the data by dividing the number couples over the total number of people that are legally able to get married. We will perform this process for the number of single individuals and common law partners so that we can see ratios for comparison purposes.

Learning: To interpret this data, we will primarily be working with queries to extract relevant information. As far as reading the data after the extraction goes, perhaps we might need to figure out how to organize the data, such that we can compare each province across all the spectrums, both in terms of the number of females in particular management divisions, as well as the relevant factors that might result in that number of females partaking in those divisions. We would have to find a way to organize the data, such that we can then compare those results across the provinces/territories to then be able to draw conclusions.

Perhaps a good topic to look into would be organization strategies.

Relations:

1) mmSpecialized (<u>location</u>, ratioMale, ratioFemale)
   A tuple in this relation represents the ratio of males vs the ratio of females that occupy "specialized middle management positions" by province/territory. *location* denotes the province/territory, and ratioMale and ratioFemale denotes the ratio of males and ratio of females occupying those specialized middle management positions in their respective *location*.

2) mmRetailCustomerService(<u>location</u>, ratioMale, ratioFemale)
   A tuple in this relation represents the ratio of males and ratio of females that occupy "middle management occupations in retail and wholesale trade and customer services".

location denotes the province/territory, and ratioMale and ratioFemale denotes the ratio of males and ratio of females occupying middle management occupations in retail and wholesale trade and customer services in their respective *location*.

3) mmTradesProduction(<u>location</u>, ratioMale, ratioFemale)
A tuple in this relation represents the ratio of males and the ratio of females that occupy "middle management occupations in trades, transportation, production and utilities". *location* denotes the province/territory, and ratioMale and ratioFemale denotes the ratio of males and ratio of females occupying occupations in trades, transportation, production and utilities.

4) mmSenior(<u>location</u>, ratioMale, ratioFemale)
A tuple in this relation represents the ratio of males and the ratio of females that occupy "senior management positions". *location* denotes the province/territory, and ratioMale and ratioFemale denotes the ratio of males and ratio of females occupying senior management positions.

5) GDP(<u>location</u>, GDP)
A tuple in this relation represents the Gross Domestic Product (GDP) by province/territory. *location* denotes the province/territory and GDP denotes the Gross Domestic Product in that respective province/territory.

6) relationshipStatus(<u>location</u>, ratioCouplesMarried, ratioCouplesCommonLaw, ratioSingle)

A tuple in this relation represents the relationship status of the population in its respective province/territory. *ratioCouplesMarried* represents the number of couples that are married, *ratioCouplesCommonLaw* represents number of couples in common law relationships, and *ratioSingle* represents number of people that are single.

7) postsecondaryInstitutions(<u>location</u>, enrollmentFemale, enrollmentMale)

A tuple in this relation represents enrollment into postsecondary institutions in different provinces/territories by sex. *location* denotes the province/territory, and *enrollmentFemale* represents enrollment of females in postsecondary institutions and *enrollmentMale* represents enrollment of males in postsecondary institutions in their respective province/territory.

Data Dictionary

Attribute	Description	Type	Required	Default
Relations 1-4				
location	The location of the province/territory for which the data was recorded	TEXT	Yes	
ratioMale	The ratio of males in the management position described	DOUBLE	Yes	
ratioFemale	The ratio of females in the management position described	DOUBLE	Yes	
Relation 5				
location	The location of the province/territory for which the data was recorded	TEXT	Yes	
GDP	The gross domestic product of the area where the data is recorded	DOUBLE	Yes	
Relation 6				
location	The location of the province/territory for which the data was recorded	TEXT	Yes	
ratioCouplesMarried	The ratio of the married couple in the respective province/territory	INT	Yes	
ratioCouplesCommonLaw	The ratio of common-law couples in the respective province/territory	INT	Yes	
ratioSingle	The ratio of single individuals in the respective province/territory	INT	Yes	
Relation 7				
location	The location of the province/territory for which the data was recorded	TEXT	Yes	
enrollmentFemale	The number of females who enrolled in postsecondary education	INT	Yes	
enrollmentMale	The number of males who enrolled in postsecondary education	INT	Yes	

Constraints

mmSpecialized[location] ⊆ mmRetailCustomerServices[location] ⊆ mmTradesProductions[location] - mmSenior[location] ⊆ GDP[location] ⊆ relationshipStatus[location] ⊆ postsecondaryInstitutions[location] ⊆ {"Newfoundland and Labrador", "Prince Edward Island", "Nova Scotia", "New Brunswick", "Quebec", "Ontario", "Manitoba", "Saskatchewan", "Alberta", "British Columbia", "Yukon", "Northwest Territories", "Nunavut"}

Justification of Relations

These relations are created such that they aim to focus on the relevant questions that we want to answer.

Relations 1-4 in the list above aim to showcase the distribution of male and female in those particular management positions. This division will allow us to visualize which management positions have higher densities of male or females in relative comparison to the other province/territories. Such a division will allow us to ask questions like, "it seems most province/territories have a reduced amount of female in this field, but this particular province/territory has a larger amount of female— perhaps this says something about how the needs and the sociopolitical/ economic states of the province/territory affects the female workforce."

Relation 5 aims to showcase the variety of GDP values based on location. This relation will allow us to analyze how GDP might reflect the distribution of males and females in the workforce. By analyzing GDP, perhaps we will be able to create associations as to why certain areas with certain GDP values have more/less females in management positions (derived from previous relations).

Relation 6 aims to highlight how the ratio of married couples might affect the distribution of males and females in the workforce. It will allow us to perhaps create association as to how the ratio of couples in a particular location versus the number of single individuals affects female partaking in time consuming management positions.

As management positions usually require graduation, it is beneficial to look at the graduation rates in each province. This will allow us to perhaps associate the number of female graduates, to the ratio of female enrolled in postsecondary education.

Fundamentally, relations 5-7 focus on particular/definitive categories that might allow us to create associations between the distribution of females to males in management positions in different provinces/territories.