10/15/2024

**SQL Project: Analysis of Canadian and Foreign-Controlled Enterprises: Distribution, Size, and Financial Performance (1999-2022)**

**Kina Gajipara (8850437)**

**Business Analytics, Conestoga college**

**SQL and Data Analysis**

**Professor: Jey Kumaresan**

**Table Of Content**

1. Goal of Project ……………………………………………………………………… 2
2. Data source and cleaning Process …………………………………………………... 3
3. ERD diagram and Explanation ……………………………………………………… 3
4. Query Output and insights …………………………………………………………... 6
5. Conclusion …………………………………………………………………………... 9

**Goal of the Project:**

The goal of this project is to analyse the characteristics of the Canadian enterprises during the years 1999 to 2022. The focus is to understand the distribution of these enterprises based on the main factors: Country od Control (whether it is controlled by Canada or foreign), their size (small, medium or large enterprises), and how well they performed financially, specifically measured by average operating income. This examination seeks to offer understanding on how these factors impacted the configuration and functioning of Canadian businesses in this timeframe.

Using SQL queries and a well-structured database, this project provides a detailed analysis of business control, size, industry type, and financial data. The goal is to learn how these factors affect the Canadian economy, whether the businesses are Canadian-owned or foreign-controlled. By breaking down the data by year and different business characteristics, this project offers valuable insights for business leaders and policymakers. It helps reveal trends and gives a clearer picture of the role that different types of businesses play in Canada's economy.

In conclusion, this project uses SQL and database tools to analyze Canadian and foreign-owned businesses in detail. The findings can help guide business decisions and policies by highlighting important trends in business size, ownership, and financial performance. Understanding these patterns will help stakeholders assess how different businesses contribute to Canada's economic growth.

**Data Source and Cleaning Process:**

The primary data source for this project is the csv file named “Foreign-controlled enterprises in Canada, counts by operating revenue size groups”. This file Contain information about enterprises and about characteristics. The data is sourced from Statistics Canada, which can be accessed at <https://www.statcan.gc.ca>.

In data warehousing, the first step is to create a database, which we named **EnterprisesDB**, to organize and store the data effectively. it is essential to create dimension and fact tables to effectively organize and structure data for analysis. At first, a temporary table is generated to act as a staging area for importing raw data from a source, like a CSV file. This provisional table enables initial data cleaning and transformation before inserting data into permanent tables. After the temporary table is filled, distinct values are selected to fill dimension tables that contain descriptive characteristics linked to the data. In our project dimension tables contain details such as industry categorizations, company attributes, controlling country, enterprise size, measurement units, and scaling factors. By adding unique values to these dimension tables, duplication is reduced, and referential integrity is maintained.

After setting up a dimension table, the next step involves a creating a fact table, which contain measurable data – mostly numerical values that can be aggregated for analysis. The fact table references the dimension table through the foreign key, which links each quantitative measure to a qualitative attributed. Suppose we have a fact table called Fact\_Enterprises that includes a measurable value for VALUE (e.g., operating revenue (VALUE)). It also includes COUNTRY\_OF\_CONTROL\_ID, which links to a dimension table called Dim\_COUNTRY\_CONTROL. The Dim\_Country\_Control contains descriptive attributes like the country controlled by Canada or foreign. By linking these tables, you can easily query data to find, for example, the total revenue generated by Canadian or foreign controlled in the manufacturing sector.

**ERD Diagram and Explanation:**

The entity-relationship diagram (ERD) describes the structure of a database designed to store and manage enterprise-related data. This schema includes one central fact table (fact\_enterprises) surrounded by several dimension tables (e.g., dim\_enterprise\_characteristics, dim\_size\_of\_enterprise, dim\_country\_control, etc.). The fact table captures transactional or measurable data, while the dimension tables store descriptive information. This model, often referred to as a star schema, is widely used for analytical purposes and reporting.

* **Schema Overview:**

This Schema contains total 6 tables: One is the fact table and other 5 are the dimension table.

* Fact Table: fact\_enterprises
* Dimension Tables: dim\_enterprise\_characteristics

dim\_size\_of\_enterprise  
 dim\_country\_control

dim\_uom

dim\_scalar

**Primary Keys and Foreign keys:**

**Primary Key:** A primary key is the unique identifier for records in table. It ensures that each record in within table can be uniquely identifiable.

**Foreign Key:** A foreign key is used to create a link between two tables. In this schema, foreign keys from the dimension tables are referenced in the fact table to maintain relationships.

**Dimension and Fact table Description:**

* **dim\_enterprise\_characteristics**: Stores qualitative descriptions of enterprise characteristics, such as business type and size.
* **dim\_size\_of\_enterprise**: Contains data about the size of enterprises (e.g., small, medium, large).
* **dim\_country\_control**: Holds information about the controlling country of the enterprise.
* **dim\_uom (Unit of Measure)**: Stores units of measure (e.g., kilograms, liters) used for quantifying enterprise attributes.
* **dim\_scalar**: Contains scalar factors used to scale enterprise data for analysis, such as adjusting values by a factor to express in different units (e.g., millions).
* **fact\_enterprises**
* **Primary Key**: ref\_date
* **Foreign Keys**:
  + enterprise\_characteristics\_id (from dim\_enterprise\_characteristics)
  + size\_of\_enterprise\_id (from dim\_size\_of\_enterprise)
  + country\_of\_control\_id (from dim\_country\_control)
  + uom\_id (from dim\_uom)
  + scalar\_id (from dim\_scalar)

In this database, the relationships between the **fact** and **dimension** tables are important for organizing the data and making it easy to analyze, create reports, and generate insights.

The ERD below shows the layout of the database schema, focusing on how the fact and dimension tables are connected. This diagram helps you understand how the different parts work together, making it easier to see how the data is organized and how to analyze it effectively.

A screenshot of a computer screen

Description automatically generated

* **Fact Table**: The **fact\_enterprises** table is the central table where the main business data (like enterprise metrics) is stored. It connects to several dimension tables through foreign keys.
* **Dimension Tables**: These tables store additional details that describe the data in the fact table. For example, the **enterprise characteristics**, **size**, and **country control** of each enterprise are stored in their respective dimension tables, providing context for the metrics in the fact table.
* **Star Schema**: This structure is known as a **star schema**. The fact table is at the center, and all the dimension tables connect to it, forming a shape like a star. This setup makes querying the data faster and easier by reducing the number of complex joins between tables.

**Query, Output and Insights:**

* **Analyzing Enterprises by Country of Control and Enterprise Size**

This query aims to examine how enterprises in Canada are distributed in 2022, with a specific focus on the size of enterprises (small, medium, large) and whether they are Canadian or foreign controlled by creating view in SQL. The assessment will categorize businesses based on these factors and present important measurements like total number of businesses, total business value, and average value per business. This can assist in recognizing trends in the sizes of businesses managed by various countries and their economic contributions.

A screenshot of a computer code

Description automatically generated

This query creates a view to summarize data about enterprises, grouping them by **country of control** (e.g., domestic or foreign) and **enterprise size** (e.g., small, medium, large). First, it retrieves key information such as the **year** of the data, **geographic region** (Canada), **country controlling the enterprise**, and the **size of the enterprise**.

Query calculates: **Total Enterprises**: The total sum of enterprises' values, **Enterprise Count**: The number of enterprises in each group, **Average Enterprise Value**: The average value per enterprise. Finally, it groups the results by year, geography, country of control, enterprise size.

**Output**: The view helps to quickly analyze and understand how the control of enterprises (domestic or foreign) and their size are distributed and contribute to the economy. A second query then retrieves data specifically for **Canada in the year 2022**, showing how many enterprises of different sizes are controlled by foreign or domestic entities.

**Insights**:

* Canadian-controlled enterprises significantly outnumber foreign-controlled enterprises. For example, the total number of enterprises under Canadian control is **1,970,982**, while under foreign control, it is only **146,444**.
* Small enterprises dominate the market in both Canadian and foreign control. Canadian-controlled small enterprises account for **1,953,034** total enterprises, while foreign-controlled small enterprises account for only **11,295**. This indicates that small enterprises play a crucial role in both canadian and foreign-controlled businesses.
* Foreign-controlled businesses primarily consist of larger enterprises, while Canadian businesses maintain a strong presence across all sizes, particularly in the small and medium segments.
* **Analyzing Enterprise Characteristics by Geography and Characteristics**

The query aims to analyze enterprises by geography (e.g., countries) and specific characteristics (such as industry, size, etc.) to provide insights into how enterprises are distributed across different regions and what their defining features are. By creating a view, the query allows for a quick and efficient assessment of enterprise data, enabling comparisons across different geographies.

A screenshot of a computer code

Description automatically generated

**Output**:

* The view presents summarized data that allows easy analysis of enterprise distribution by various characteristics and geography.
* A subsequent query retrieves data specifically for the year 2022, providing a snapshot of how enterprises are distributed across different regions and their characteristics during that period.

**Insights:**

* The analysis is centered on enterprises in **Canada for the year 2022**, providing a focused view of key enterprise metrics. Total mean operating revenues amount to **1,353,676 dollars**, across 8 enterprises.
* The **average revenue per enterprise** is **169,209.5 dollars**, indicating a substantial revenue generation capacity for businesses in Canada.
* Canadian businesses are prominent in the market, generating significant revenue and featuring a wide variety of companies that contribute to the economy.
* **Analyzing total enterprises with control category**

The aim of this query is to **analyze the total number of enterprises in Canada in 2022**, categorizing them by control levels based on their size. The "control" is determined by the number of enterprises in each category, and it also retrieves additional details such as the enterprise size, count, and average value.

A computer screen shot of text

Description automatically generated

**Output**:

* The view summarizes enterprise data, making it easy to analyze how enterprises are distributed by control type, size, and value.
* The query specifically retrieves data for 2022, showing how Canadian and foreign-controlled enterprises are distributed across different sizes and control levels for that year.

**Insights**:

* Canadian-controlled enterprises are significantly higher in number, especially in small enterprises with "High Control."
* Foreign-controlled enterprises are mostly in the **Low Control** category, with notable numbers in medium and small enterprise sectors.
* Canadian-controlled enterprises, especially small ones, show much higher average values compared to foreign-controlled counterparts, indicating more valuable enterprises in the domestic sector.

**Conclusion:**

This project provided analysis of Canadian enterprises from 1999 to 2022, focusing on the country of control, size, and financial performance. Using SQL queries and a structured database, we explored how Canadian and foreign-controlled businesses differ in terms of size, number, and economic contributions.

The results show that **Canadian-controlled businesses** are far more common, especially in the small and medium-sized sectors, and they tend to be more valuable compared to foreign-controlled ones. **Foreign-controlled businesses** are fewer but are more concentrated in the large enterprise category.

By breaking down the data by year and other factors, this analysis provides useful insights for business leaders and policymakers. It highlights the key role that small Canadian-owned businesses play in the economy and shows important trends in foreign investment.

In summary, this SQL-based analysis shows the central role Canadian-owned businesses play in economic growth, while also offering insights into foreign investment trends. The findings can help guide decisions in business policy, investment, and economic strategy for the future.