**Trabajo final - Entrega 1**

**Introducción a la analítica de negocios**

Jhonatan Valencia Ocampo. Andrés Penna Hernández. Steven Zapata Zuleta. Departamento de Ingeniería Industrial, Universidad de Antioquia.

Febrero 2022

**Pregunta:** ¿Cuál es la factibilidad de una persona para emplearse en Canadá? **1. Contexto de la pregunta**

**a.¿Qué situación se va analizar?**

Se analizará el comportamiento del empleo en Canadá enfocados en su principal provincia, Ontario, esta posee el mayor desarrollo industrial respecto al total de los demás estados, además se encuentra la mayor parte de población y la ciudad capital del país, Ottawa. Identificando la factibilidad de la población para laborar a partir de diversos factores, como lo son nivel educativo, tipo de sector, indicadores económicos y población.

**b. ¿Por qué le llamó la atención?**

El tema de empleabilidad nos llama la atención, porque, en el entorno actual (Colombia) no nos permite visualizar un comportamiento favorable para los desempleados. Por lo que se quiere conocer cómo es el comportamiento en un país como Canadá, que es uno de los países desarrollados del continente, con grandes recursos, educación y economía pero en contraste presenta un decrecimiento poblacional que puede incidir en la empleabilidad.

**c. ¿Para qué puede servir el análisis de esta situación?**

El análisis de la empleabilidad en Canadá y su principal provincia (Ontario), permite hallar riesgos y oportunidades en el mercado laboral, ayuda en la elaboración de políticas y toma de decisiones por parte del gobierno, correlación de situaciones sociales que se producen como efectos colaterales.

**2. Recolección y extracción:**

**Describa estas fuentes de datos:**

**i. ¿De dónde fueron tomados los datos? - link y bases de datos**

La información fue tomada de la base de datos Open Government Canada en: https://open.canada.ca/en

Las bases de datos elegidas fueron:

1. Duration of unemployment.

https://open.canada.ca/data/en/dataset/4edcd678-6f24-4064-970b-b7254fc6671e

2. Unemployment rates of 25- to 29-year-olds, by educational attainment, Canada and provinces

https://open.canada.ca/data/en/dataset/0f68117f-a98c-49c9-956f-2fa44baa18a4

3. Employment by class of worker

https://open.canada.ca/data/en/dataset/112e6dee-b54b-4869-bf45-511a49b74f74

4. Real gross domestic product at market prices, annual percent change, Canada and Alberta

https://open.canada.ca/data/en/dataset/3f1ed900-66f4-4dfa-9ea5-e7711c0f2fd7

La información fue tomada de data commons en:

https://datacommons.org/?hl=es

5. Población en Ontario

https://datacommons.org/place/wikidataId/Q1904?utm\_medium=explore&mprop=cou nt&popt=Person&hl=es#

**ii. ¿De qué período es la información?**

| **Base de datos Periodo** |
| --- |
| Duration of unemployment. 2001-2020 |
| Unemployment rates of 25- to  29-year-olds, by educational attainment,  1995- 2020  Canada and provinces |
| Employment by class of worker 2001- 2015 |

| Real gross domestic product at market  1981-2020  prices, annual percent change, Canada  and Alberta |
| --- |
| Población en Ontario 1971-2019 |

Para efectos prácticos se trabajará con un periodo homogéneo entre las tablas que comprende desde el año 2001 hasta el año 2015.

**iii. ¿Qué información contiene?: variables, tipos, significado de cada variable. Descripción base de datos**

**1.**

| Nombre: **Duration of unemployment. (Tiempo)** | | |
| --- | --- | --- |
| **Variables** | **Tipo** |  |
| MONTH | **String** | Es el registro de la fecha a la cual corresponde la información suministrada, dicho registro se hace trimestralmente. |
| GEOGRAPHY | **String** | Comprende las zonas geográficas de Ontario de donde fue tomada la información. |
| DURATION  AGE\_GROUP | **String**  **String** | Tiempo que la persona permanece desempleada, está dado en semanas.  Edad de la persona desempleada durante un periodo de tiempo, se registraba varios rangos de edad que van desde los 15 años hasta los 64. |
| Both\_sexes | **Float** | Número de personas que están desempleadas. |
| Female | **Float** | Número de mujeres que están desempleadas. |

| Male | **Float** | Número de hombres que están desempleadas. |
| --- | --- | --- |

**2.**

| Nombre: **Unemployment rates of 25- to 29-year-olds, by educational attainment, Canada and provinces (Formación).** | | |
| --- | --- | --- |
| **Variables** | **Tipo** |  |
| \_REF\_DATE\_ | Integer | Indica el año en el que fue registrada la información. |
| GEO | String | Zona geográfica de Canadá de donde se toma la información, comprende las provincias que conforman al país. |
| Educational\_attaiment | String | Niveles educativos de las personas registradas. |

**3.**

| Nombre: **Employment by class of worker (Clase).** | | |
| --- | --- | --- |
| **Variables** | **Tipo** |  |
| Geography | **String** | Comprende las zonas geográficas de Ontario de donde fue tomada la información. |
| Class\_of\_worker | **String** | Sector económico del cual hace parte la persona. |
| Year | **Integer** | Es el registro de la fecha a la cual corresponde la información suministrada |
| Value | **Float** | Valor de los registros (población) de acuerdo con las segmentación de variables. |

**4.**

| Nombre: **Real gross domestic product at market prices, anual percent change, Canada and Alberta (PIB).** | | |
| --- | --- | --- |
| **Variables** | **Tipo** |  |
| REF\_DATE | **Integer** | Es el registro de la fecha a la cual corresponde la información suministrada |
| GEO | **String** | Zona geográfica de Canadá de donde se toma la información. |
| Prices | **String** | Registra en concepto del valor. (Ajuste del dólar para el año correspondiente) |
| Estimates | **String** | Producto interno bruto a precios de mercado. |
| UOM | **String** | Unidades en las que fue registrada la información. (Dólar) |
| SCALAR\_FACTOR | **String** | Unidades en las que están registrados los datos. |
| VALUE | **String** | Hace referencia al valor de interés (PIB). |
| GDP\_Percentage\_Change\_ | **String** | Indica el porcentaje de cambio respecto al período inmediatamente anterior. |

**5.**

| Nombre: **Población en Ontario** | | |
| --- | --- | --- |
| **Variables** | **Tipo** |  |
| Fecha | **Integer** | Es el registro de la fecha a la cual corresponde la información suministrada |
| Poblacion | **Integer** | Número de personas en Ontario |

**3. Carga y almacenamiento:**

**Describa estas fuentes de datos:**

**a)**

**1. Base de datos:** Duration of unemployment

**Sentencia:** CREATE TABLE Tiempo( MONTH varchar(100) not null, GEOGRAPHY varchar(100) not null, DURATION varchar(100) not null, AGE\_GROUP varchar(100) not null, Both\_sexes FLOAT default "0", Male FLOAT default "0", Female FLOAT default "0");

**2. Base de datos:** Unemployment rates of 25- to 29-year-olds, by educational attainment, Canada and provinces

**Sentencia:** CREATE TABLE Formacion( \_REF\_DATE\_ int, GEO varchar(100) not null, DGUID varchar(100) not null, Educational\_attainment varchar(100) not null, UOM varchar(100) not null, UOM\_ID INTEGER not null, SCALAR\_FACTOR varchar(100) not null, SCALAR\_ID INTEGER not null, VECTOR varchar(100) not null, COORDINATE FLOAT not null, VALUE FLOAT not null, STATUS varchar(100) default "x", SYMBOL varchar(100), TERMINATED varchar(100), DECIMALS INTEGER not null, primary key ( \_REF\_DATE\_ , VECTOR ));

**3. Base de datos:** Employment by class of worker

**Sentencia:** CREATE TABLE Clase(Geography varchar(100) not null, Class\_of\_worker varchar(100) not null, Year INTEGER not null CHECK(Year>2000), Value float, primary key ( Geography, Class\_of\_worker, Year));

**4. Base de datos:** Real gross domestic product at market prices, anual percent change, Canada and Alberta

**Sentencia:** CREATE TABLE PIB(REF\_DATE int, GEO varchar(100) not null, Prices varchar(100), Estimates varchar(100) not null, UOM varchar(100) default “Dollars” not null, SCALAR\_FACTOR varchar(100) default “millions” not null, VALUE int, GDP\_Percentage\_Change\_ float, primary key (REF\_DATE,GEO));

**5. Base de datos:** Canadá-Población

**Sentencia:** CREATE TABLE Poblacion(Fecha int UNIQUE, Poblacion int);

**Preguntas**:

1-¿Qué porcentaje de personas se encuentran desempleadas a final de año en Ontario con respecto a la población total para el año 2015?

**Código:** SELECT T2.Fecha, (T1.Personas/T2.Poblacion)\*100 as Porcentaje\_personas\_desempleadas FROM (SELECT \*,CAST(Fechaid AS INT64) AS Fecha2 FROM(SELECT GEOGRAPHY,AGE\_GROUP,REPLACE(MONTH,'Dec2015','2015') as Fechaid, SUM(Both\_sexes) as Personas FROM

`trabajo-final-1.canada.Tiempo` WHERE GEOGRAPHY="Total, Ontario regions" AND AGE\_GROUP="Total, 15 years and over" GROUP BY AGE\_GROUP,GEOGRAPHY,MONTH) WHERE Fechaid="2015") T1 INNER JOIN `trabajo-final-1.canada.Poblacion` T2 ON T1.Fecha2=T2.Fecha

**Resultado:**

****

2-¿Cual es el promedio de desempleo por sexo en Ontario?

**Código:** SELECT AVG(Male) AS promedio\_hombres, AVG(Female) AS promedio\_mujeres FROM `trabajo-final-1.canada.Tiempo`

**Resultado:**

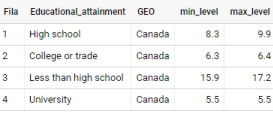
****

3- ¿Cuál fue el porcentaje poblacional máximo y mínimo, obtenido por cada nivel académico entre todo el período comprendido entre los años 2001 y 2015?

**Código**: SELECT Educational\_attainment, GEO, MIN(VALUE)as min\_level, Max(VALUE) as max\_level FROM

`trabajo-final-1.canadatrabajo-final-1.Formacion` WHERE GEO = 'Canada' AND Educational\_attainment in ('ALL levels of education', 'Less than high school', 'High school', 'College or trade', 'University') AND \_REF\_DATE\_ in (2001, 2015) Group by Educational\_attainment, GEO

**Resultado:**

****

4-¿Qué rango de edad en la población de Ontario tiene mayor dificultad para emplearse?

**Código:** SELECT AGE\_GROUP as grupo\_edad, SUM(Both\_sexes) as can\_personas FROM `trabajo-final-1.canada.Tiempo` WHERE AGE\_GROUP IN (' 20-24',' 25-44',' 45-54',' 55-64') group by AGE\_GROUP ORDER BY can\_personas DESC LIMIT 1

**Resultado:**

****

5-¿Qué región representa un mayor grado de informalidad con respecto a la población de Ontario?

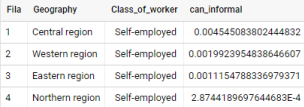
**Codigo:** SELECT T1.Geography, T1.Class\_of\_worker,

sum(Value)/T2.Poblacion\*100 as can\_informal from

`trabajo-final-1.canada.Clase` T1 inner join

`trabajo-final-1.canada.Poblacion` T2 on T1.Year = T2.Fecha WHERE Class\_of\_worker = ' Self-employed' AND Geography IN ('Eastern region','Central region','Western region','Northern region') AND Year = 2015 Group by Geography, Class\_of\_worker, Value, Poblacion ORDER BY can\_informal DESC

**Resultado:**

****

6-¿Cuál es el nivel educativo que tiene mayor probabilidad de estar desempleado?

**Código:** SELECT Educational\_attainment

,(sum(VALUE)/count(Educational\_attainment)) as prob\_desempleados FROM `trabajo-final-1.canada.Formacion` group by Educational\_attainment order by prob\_desempleados desc limit 1

**Resultado:**

****

7-¿Hacia dónde se dirige la región de Ontario en la conformación de sectores formal e informal comparando los años 2001 y 2015?

**Código:**SELECT Year,Tipo\_trabajador, (Value\*100)/SUM(Value)OVER() as Porcentaje FROM(SELECT Year,UPPER(Class\_of\_worker) as Tipo\_trabajador, Value FROM `trabajo-final-1.canada.Clase` WHERE (Class\_of\_worker=" Employees" OR Class\_of\_worker=" Self-employed") AND Year=2001 AND Geography="Total, Ontario regions")UNION ALL SELECT Year, Tipo\_trabajador, (Value\*100)/SUM(Value)OVER() as Porcentaje FROM(SELECT Year,UPPER(Class\_of\_worker) as Tipo\_trabajador, Value FROM `trabajo-final-1.canada.Clase` WHERE (Class\_of\_worker=" Employees" OR Class\_of\_worker=" Self-employed") AND Year=2015 AND Geography="Total, Ontario regions")

**Resultado:**

****

8-¿El indicador económico PIB está directamente relacionado con el nivel de población?

**Código:** SELECT CORR(PIB,Poblacion) as Correlacion FROM (SELECT T1.Fecha as Date, T2.VALUE as PIB, T1.Poblacion as Poblacion FROM `trabajo-final-1.canada.Poblacion` T1 inner join `trabajo-final-1.canada.Canada.PIB` T2 on T1.Fecha = T2.REF\_DATE WHERE GEO ='Canada' Group by Fecha, T2.VALUE, T1.Poblacion)

**Resultado:**

****

9-¿Qué año tuvo mayor variación del PIB con respecto a su año anterior?

**código:**SELECT REF\_DATE as Date, MAX(GDP\_Percentage\_Change\_\_\_\_) as Variacion\_PIB FROM `trabajo-final-1.canada.PIB` WHERE REF\_DATE

BETWEEN 2001 AND 2015 GROUP BY REF\_DATE order by Variacion\_PIB desc limit 1

**Resultado:**

****

10-¿En qué región de Ontario las personas permanecen más tiempo desempleadas?

**Código:** SELECT GEOGRAPHY, DURATION as T\_desempleado, SUM(Both\_sexes) as can\_personas FROM `trabajo-final-1.canada.Tiempo` WHERE DURATION = ' 53 weeks or more' and GEOGRAPHY IN ('Central region','Eastern region','Western region','Northern region') group by GEOGRAPHY,DURATION ORDER BY can\_personas DESC limit 1

**Resultado:**

****

11-Entre los periodos 2001 a 2015, ¿Cuál de los sectores (público y privado) ha generado mayor impacto en la empleabilidad de Ontario?

**Código:** SELECT Class\_of\_worker ,sum(Value) as cant\_empleados FROM `trabajo-final-1.canada.Clase` WHERE Class\_of\_worker IN (' Public sector employees', ' Private sector employees') AND Year BETWEEN 2001 and 2015 group by Class\_of\_worker order by cant\_empleados desc

**Resultado:**

****

12-¿Qué porcentaje ha crecido el empleo en 2015 con respecto al 2001 en Ontario?

**Código:**SELECT((MAX(cant\_empleados)-MIN(cant\_empleados))/MAX(cant \_empleados))\*100 AS Variacion\_empleo FROM (SELECT Year, sum(Value) as cant\_empleados FROM `trabajo-final-1.canada.Clase` WHERE Year IN (2001,2015) group by Year)

**Resultado:**

****

13-¿Qué porcentaje promedio de la población desempleada no terminó el colegio entre los períodos 2015 respecto el 2001?

**Código**: SELECT Educational\_attainment ,AVG(VALUE) as porcentaje\_noschool FROM `trabajo-final-1.canada.Formacion` WHERE Educational\_attainment IN ('Less than high school') AND \_REF\_DATE\_ BETWEEN 2001 and 2015 group by Educational\_attainment order by porcentaje\_noschool desc

**Resultado:**

****

14-¿El aumento del PIB aumenta las probabilidades de conseguir empleo?

**Código:** SELECT CORR(PIB,Empleos) as Correlacion FROM (SELECT T1.Year as Date, T2.VALUE as PIB, SUM(T1.Value) as Empleos FROM `trabajo-final-1.canada.Clase` T1

inner join `trabajo-final-1.canada.PIB` T2 on T1.Year = T2.REF\_DATE WHERE (GEO ='Canada') AND (Year BETWEEN 2001 AND 2015) Group by Year, T2.VALUE)

**Resultado:**

****

15-¿Qué porcentaje de mujeres y hombres permanecen menos tiempo desempleados en Ontario?

**Código:** SELECT DURATION as T\_desempleado, (SUM(Male)/SUM(Both\_sexes)) as porc\_male, (SUM(Female)/SUM(Both\_sexes)) as porc\_female FROM `trabajo-final-1.canada.Tiempo` WHERE DURATION = ' 1 - 4 weeks' group by DURATION

**Resultado:**

****