

BUSINESS INTELLIGENCE

PROYECTO FINAL

Samuel Porcel Rodríguez



SAMUEL PORCEL RODRÍGUEZ

RESPONSABLE DEL DESARROLLO DEL TRABAJO FINAL DE CURSO

Fecha: 31/12/2024

ÍNDICE

1. INTRODUCCIÓN.....	2
2. EXTRACCIÓN DE DATOS.....	2
3. MODIFICACIONES EN POWER BI.....	4
4. COMBINACIONES.....	6
5. REALIZACIÓN DE MÉTRICAS.....	7
6. FILTROS.....	10



ELABORACIÓN PROYECTO FINAL:

BUSINESS INTELLIGENCE

1. INTRODUCCIÓN

En este documento sobre la elaboración del proyecto, se analizará una gran variedad de datos sobre el impacto del COVID-19 mundialmente para un medio de comunicación, abarcando desde la extracción y limpieza de datos hasta la generación de reportes y dashboards interactivos que permitan responder a preguntas clave sobre la evolución y consecuencias de la pandemia mundial. El enfoque estará en identificar patrones, tendencias y oportunidades, utilizando Power BI.

2. EXTRACCIÓN DE DATOS

La extracción de datos es el primer y uno de los pasos más cruciales en cualquier proyecto de Business Intelligence. Este proceso consiste en recopilar información relevante de diferentes fuentes.

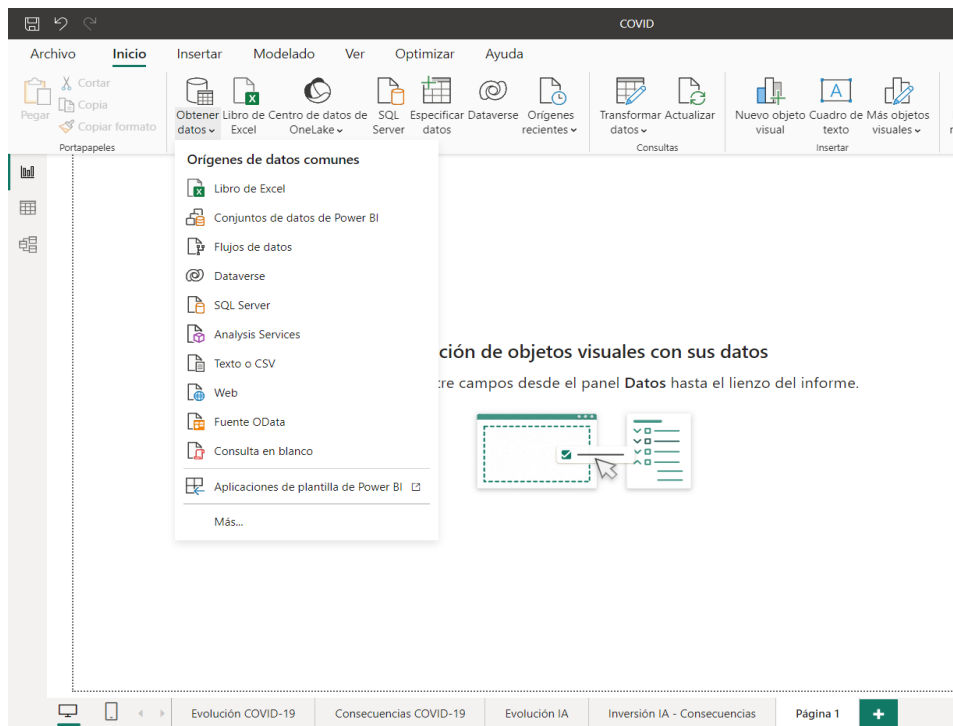
En este caso, el medio de comunicación trabaja directamente con Universidad de Oxford, por tanto, muchos datos han sido obtenidos de esta asociación. Además, hemos podido encontrar información de valor en organizaciones como el INE e Ilostat.

En total hemos obtenido 10 archivos formato CSV. A continuación, podréis ver el nombre del fichero, el link y la fecha de la última descarga del mismo:

Nombre fichero	Link	Fecha descarga
artificial-intelligence-training-computation	https://ourworldindata.org/artificial-intelligence?insight=ai-systems-perform-better-than-humans-in-language-and-image-recognition-in-some-tests#key-insights	11/12/2024
Confirmed cases	https://ourworldindata.org/coronavirus	11/12/2024
Confirmed deaths	https://ourworldindata.org/coronavirus	11/12/2024
corporate-investment-in-artificial-	https://ourworldindata.org/artificial-intelligence?insight=the-last-decades-saw-a-continuous-exponential-increase-in-the-computation-used-to-train-ai#key-insights	11/12/2024

intelligence-by-type		
Desempleo por sexo y edad	https://rshiny.ilo.org/dataexplorer37/?lang=en&id=UNE_TUNE_SEX_AGE_NB_M	11/12/2024
Empleo por sectores	https://rshiny.ilo.org/dataexplorer39/?lang=en&id=SDG_0831_SEX_ECO_RT_A	11/12/2024
owid-covid-data	covid-19-data/public/data at master · owid/covid-19-data · GitHub	11/12/2024
Quiebra por naturaleza jurídica INE	https://www.ine.es/jaxiT3/Tabla.htm?t=3181	11/12/2024
Sectores afectados IA	https://www.ilo.org/es/resource/article/como-puede-afectar-la-ia-generativa-las-distintas-ocupaciones	11/12/2024
Vaccine doses	https://ourworldindata.org/coronavirus	11/12/2024

Posteriormente, para importar estos archivos en Power BI hemos seleccionado la siguiente opción: Obtener datos > Texto o CSV.

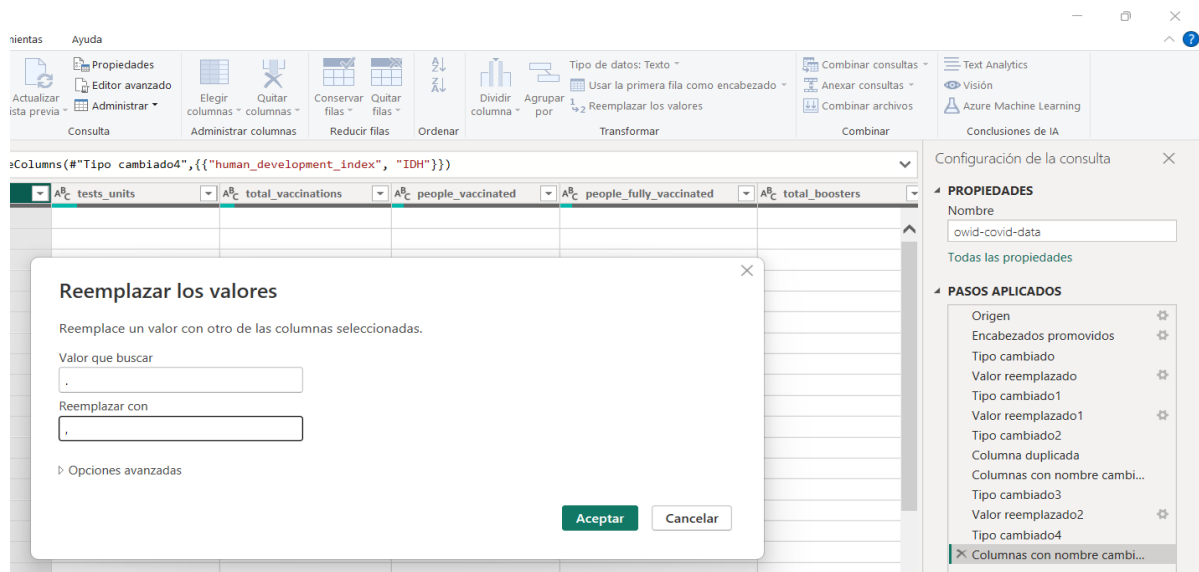


3. MODIFICACIONES EN POWER BI

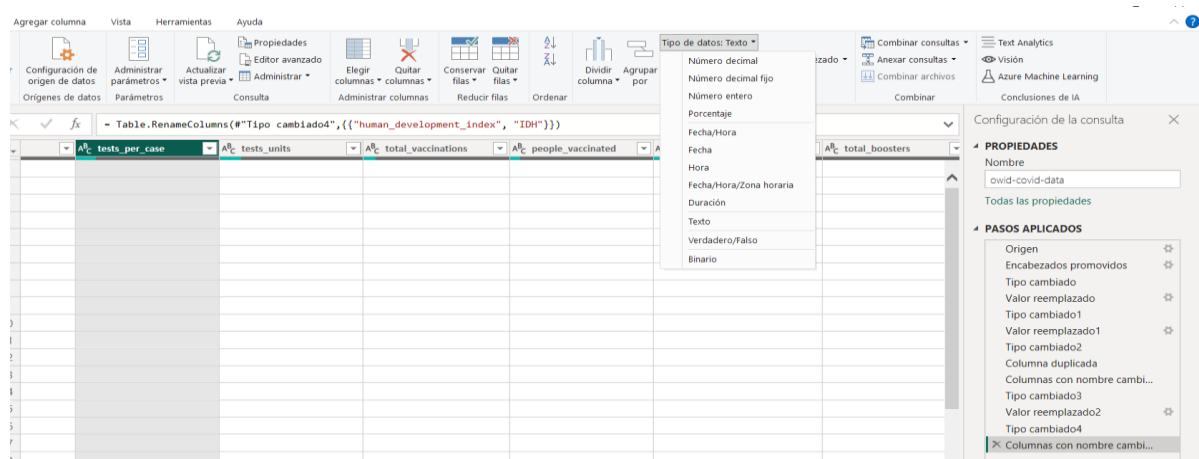
El objetivo de este apartado es mostrar cómo las modificaciones en Power BI contribuyen a transformar datos complejos en información accesible y relevante.

Las transformaciones efectuadas en Power BI en su mayoría han sido:

- Reemplazar valores > Valor que buscas > Reemplazar con > Aceptar.



- Cambiar tipo de dato > Elegir el tipo de dato deseado.





A continuación, mostraré todas las transformaciones hechas en cada tabla (CSV) y columna correspondiente:

- **Vaccine doses:**
 - COVID-19 doses: reemplazamos '.' por ',' y pasamos de texto a número decimal.
 - Day: tipo de dato de texto a fecha. Y duplicamos esta columna y la ponemos en texto para combinar consultas.
- **Owid-covid-data:**
 - aged_65_older: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - aged_70_older: tipo de dato de texto a número decimal.
 - median_age: tipo de dato de texto a número decimal.
 - date: tipo de dato de texto a fecha.
 - cardiovasc_death_rate: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - diabetes_prevalence: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - female_smokers: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - male_smokers: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - extreme_poverty: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - gdp_per_capita: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - Population: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - population_density: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - total_cases_per_million: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - total_deaths_per_million: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - total_cases: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - IDH: reemplazar '.' por ',' y pasamos de texto a número decimal.
 - total_tests: cambio de tipo de dato de texto a número entero.
 - total_tests_per_thousand: cambio de tipo de dato de texto a número entero.
- **Corporate-investment-in-artificial-intelligence-by-type:**
 - Corporate investment in AI: tipo de dato de texto a número decimal.
 - Year: tipo de dato de texto a número entero.
- **Desempleo por sexo y edad:**
 - Sex.label: reemplazar "Sex:" por "".
 - Time: reemplazar "Q1" por "/01", "Q2" por "/02", "Q3" por "/03" y "Q4" por "/04". Y pasar el tipo de dato de texto a fecha.



- **Empleo por sectores:**
 - Sex.label: reemplazar "Sex:" por " ".
 - Classif1.label: reemplazar "Economic activity" por " ". Reemplazar "(Aggregate):" por " ".
- **Quiebra por naturaleza jurídica INE:**
 - Periodo: reemplazar "T1" por "/01", "T2" por "/02", "T3" por "/03" y "T4" por "/04". Y pasar el tipo de dato de texto a fecha.
- **Sectores afectados IA:**
 - Puntuación media: reemplazar "NA" por " ". Pasar tipo de dato de texto a número decimal. Reemplazar '.' por ',' y pasamos de texto a número decimal.
 - Desviación estándar: reemplazar "NA" por " ". Pasar tipo de dato de texto a número decimal. Reemplazar '.' por ',' y pasamos de texto a número decimal.
 - Mean_exposure_level: reemplazar "NA" por " ".
- **Confirmed cases:** duplicamos la columna day, llamada Date_real y la pasamos a texto para combinar una consulta, combinando la columna COVID-19 doses de Vaccine doses en la tabla Confirmed cases.
- **icu_patients, icu_patients_per_million, hosp_patients, hosp_patients_per_million:** reemplazamos '.' por ',' y pasamos de texto a número decimal.

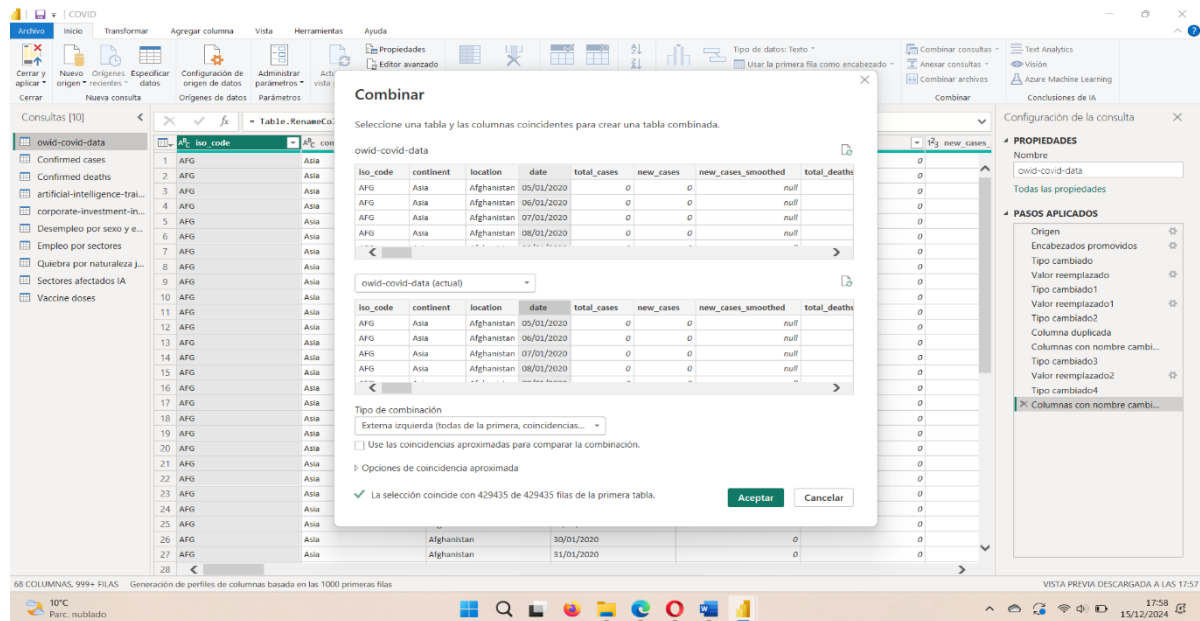
4. COMBINACIONES

La combinación de consultas es un paso esencial en el proceso de preparación de datos dentro de un proyecto de Business Intelligence, ya que permite integrar información proveniente de diferentes orígenes en un único conjunto de datos coherente y listo para el análisis. Esta etapa facilita la consolidación de datos dispersos, permitiendo a los analistas obtener una visión integral que respalde la toma de decisiones informadas.

Para combinar consultas en Power BI es necesario usar una columna que coincida entre distintas tablas, y esas columnas deben de ser de tipo texto. Haremos una combinación de consulta para cruzar los datos de IDH de la tabla "Owid-covid-data" en la tabla "Artificial-intelligence".

Para esto transformamos las columnas usadas como clave de combinación la cual es 'date' y 'day' de fecha a texto. Y para que coincidan ambas columnas se reemplazó '-' por '/' en 'day'. Se realiza la combinación y expandimos la columna 'IDH' de la tabla "Owid-covid-data".

Debido a que necesitamos la columna 'day' para las visualizaciones gráficas y no podemos cambiar el tipo de dato de texto a fecha porque si no la combinación de consulta se rompe. Por tanto, duplicamos 'day' con nuevo nombre 'Date_real' y lo pasamos de texto a fecha.

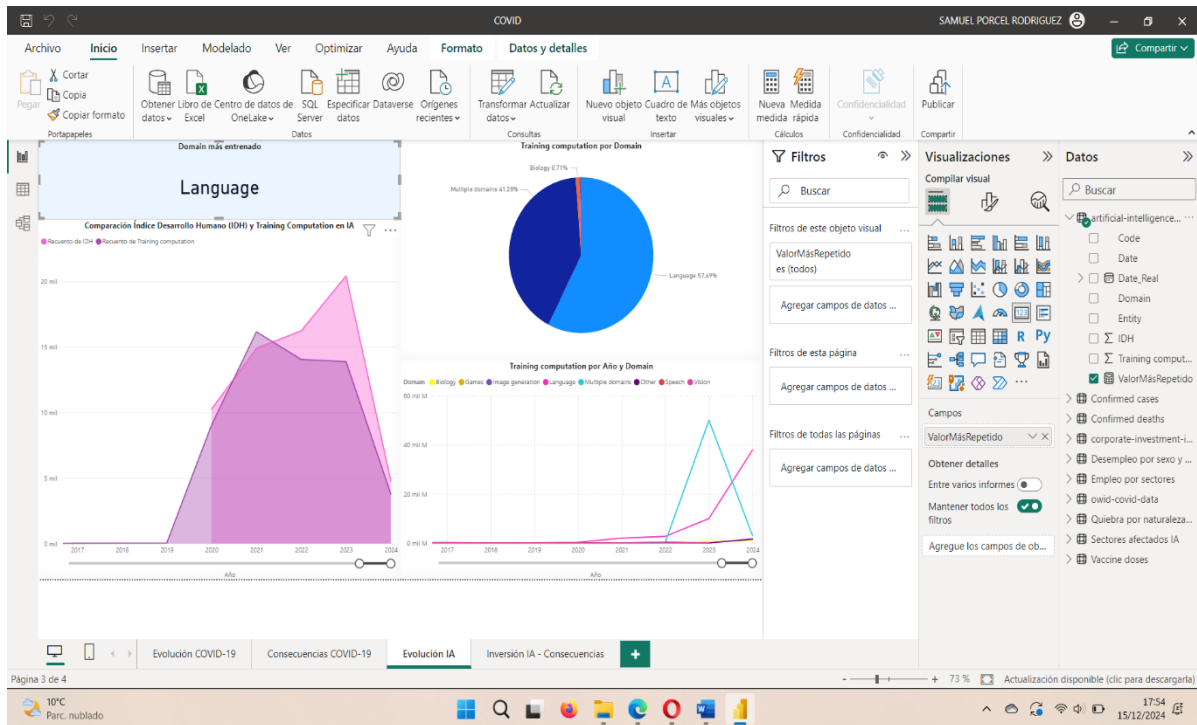


5. REALIZACIÓN DE MÉTRICAS

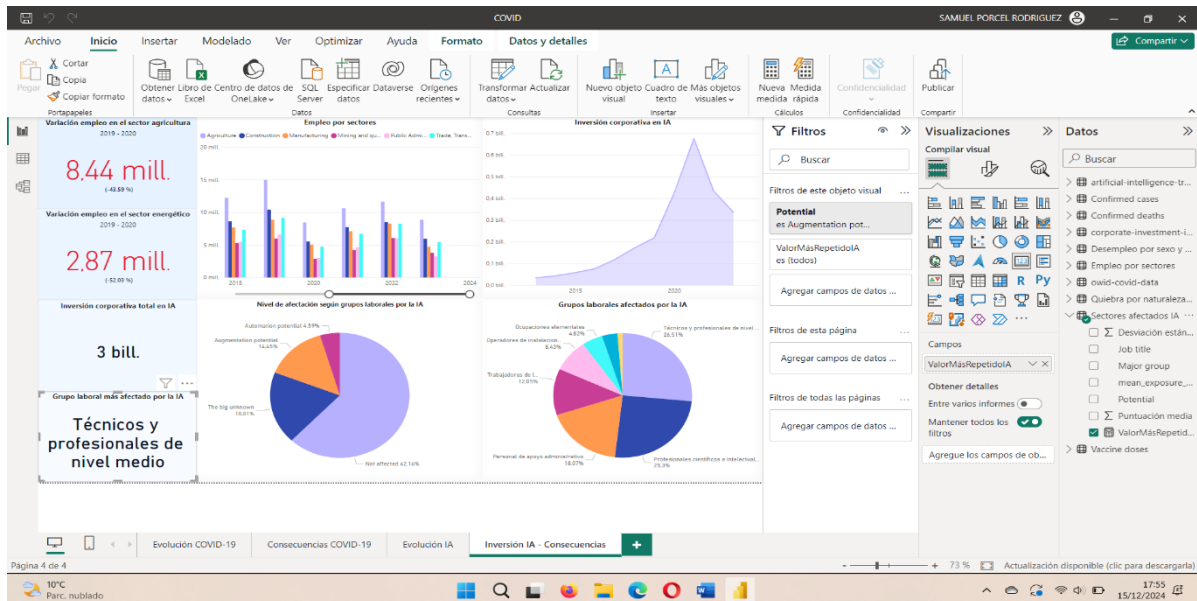
En este apartado, se describirán las adaptaciones realizadas en Power BI, incluyendo la creación y edición de métricas personalizadas mediante lenguaje DAX (Data Analysis Expressions), el ajuste de filtros y segmentaciones para focalizar el análisis.

Hemos usado el lenguaje DAX para crear distintas métricas con el objetivo de representar valores que las funciones básicas en el apartado de visualizaciones de Power BI no presenta.

- **Lenguaje:** quería mostrar el dominio más entrenado de la Inteligencia Artificial mediante una tarjeta de Power BI. Pero solo mostraba el dominio con la primera y la última fecha. Por tanto, creé la métrica ValorMásRepetido para así mostrar el dominio más repetido en el modelo.



- **Grupo laboral más afectado por la IA:** quería mostrar el grupo laboral más afectado por la IA mediante una tarjeta de Power BI. Pero solo mostraba el dominio con la primera y la última fecha. Por tanto, creé la métrica ValorMásRepetido para así mostrar el dominio más repetido en el modelo.



```

1 ValorMásRepetidoIA =
2 SELECTCOLUMNS(
3     TOPN(
4         1,
5         SUMMARIZE(
6             'Sectores afectados IA',
7             'Sectores afectados IA'[Major group],
8             "Frecuencia", COUNT('Sectores afectados IA'[Major group])
9         ),
10        [Frecuencia],
11        DESC
12    ),
13    "Resultado", 'Sectores afectados IA'[Major group])

```

- KPI: quería mostrar la evolución del empleo del año 2019 al 2020. Pero no encontraba la forma de hacerlo con las columnas que tenía en dicha tabla. Por tanto, tuve que crear dos métricas: ValorAñoActual y ValorAñoAnterior.

```

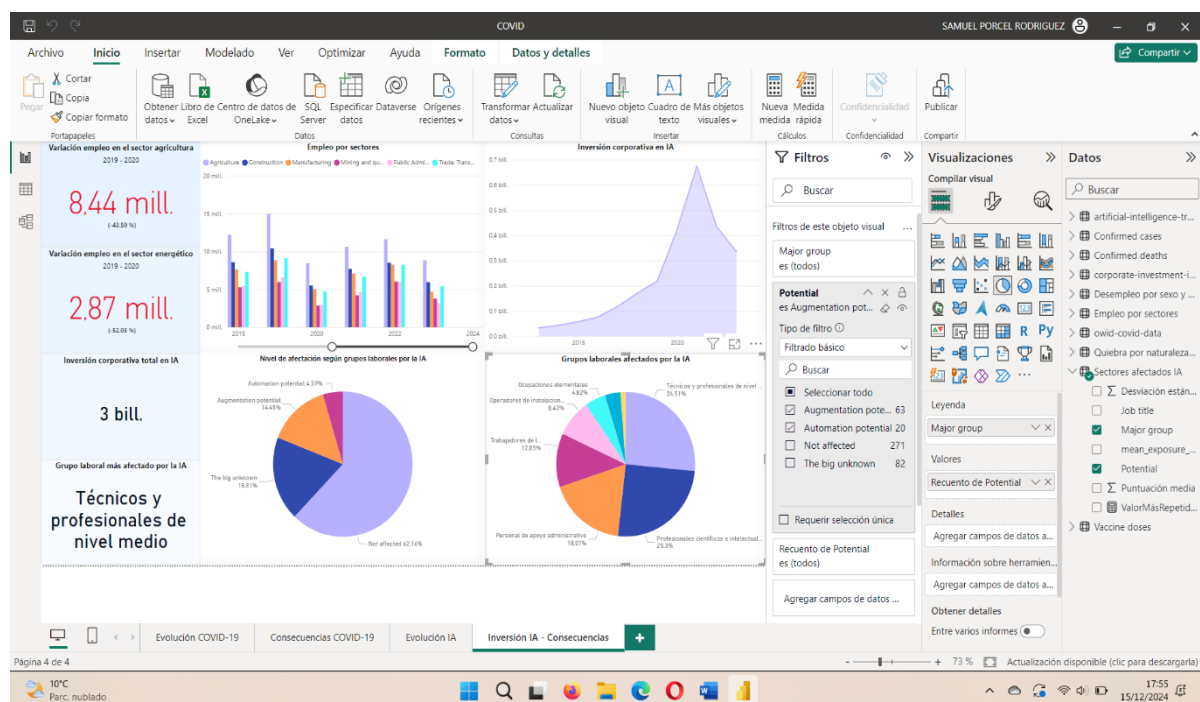
1 ValorAñoActual =
2 CALCULATE(
3     SUM('Empleo por sectores'[obs_value]),
4     'Empleo por sectores'[time] = 2020
5 )
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```

6. FILTROS

Por último, hemos hecho uso de filtros para poder visualizar en los gráficos nuestras propuestas.

En el gráfico circular “Grupos laborales afectados por la IA” la idea era mostrar solamente los grupos que puedan ser afectados por la IA, por eso filtramos dejando fuera a las opciones: “Not affected” y “The big unknown”.



Por otro lado, en los KPI la idea era mostrar la variación de los dos sectores con mayor y menor empleo. Por esta razón en el primer KPI filtramos solamente por el sector “Agriculture” y en el segundo KPI por el sector “Mining and Quarrying: Electricity, gas and water supply”.

