Proyecto Final

Línea corta

[Integrantes](#_4lqp25cx7kth) 2

[Temática](#_q5dfpefed0q) 2

[Objetivo](#_8tpw7lofsrgc) 2

[Alcance](#_kym8boudco7l) 2

[Usuario Final y nivel de aplicación y análisis](#_73wb4lfx0lkg) 2

[**Fuente de Datos**](#_dcl1ebvetl4x) **3**

[Modelo Relacional](#_zhblwp1yg7kd) 3

[Base de Datos](#_ovawxjoi07ia) 3

[Cómo se relacionan las tablas](#_lzr2si3oshbv) 5

[Tipos de datos en cada tabla](#_tvquuf8hh7tc) 5

[**Transformación de Datos**](#_lqb2lwvzjqlu) **9**

[**Análisis funcional del tablero**](#_39o9jy3ssvo3) **14**

[Dashboard: here](#_3jmxz66jkecw) 14

[**Medidas Calculadas**](#_76c7i25iy7fv) **18**

## Integrantes

María Laura Fantoni  
María Emilia Derimais  
Martín Ossés

## Temática

En nuestro trabajo analizaremos el comportamiento y las variables de ciertas posiciones de trabajo (IT) ofrecidas durante la pandemia en el año 2020.

Las posiciones bajo estudio son las de Data Analyst, Business Analyst y Data Science.

Como consecuencia de la pandemia, en 2020 el rubro IT tuvo una demanda destacada sobre los demás. El objetivo es analizar qué posiciones de IT fueron las más solicitadas durante el año 2020 y, en función de este análisis, evaluar las proyecciones de dichas posiciones para los próximos años.

## Objetivo

El objetivo del tablero es poder analizar cómo se comportó la demanda de ciertas posiciones de IT en EE.UU en el año 2020. El comportamiento que se analizará serán las ciudades con mayor demanda, las industrias que más requirieron posiciones de IT, los sueldos promedios y las skills requeridas.

Teniendo esta información, el usuario será capaz de estimar cómo será la demanda de los próximos años, dónde debe focalizarse el training, las industrias que mejor se adaptaron al nuevo normal post-pandemia y el rango de salarios de cada posición.

## Alcance

Aunque la información y el análisis es del mercado de EE.UU en el año 2020, al ser Estados Unidos el líder del mercado de IT (y quien marca tendencia), puede extenderse el alcance a los principales países desarrollados.

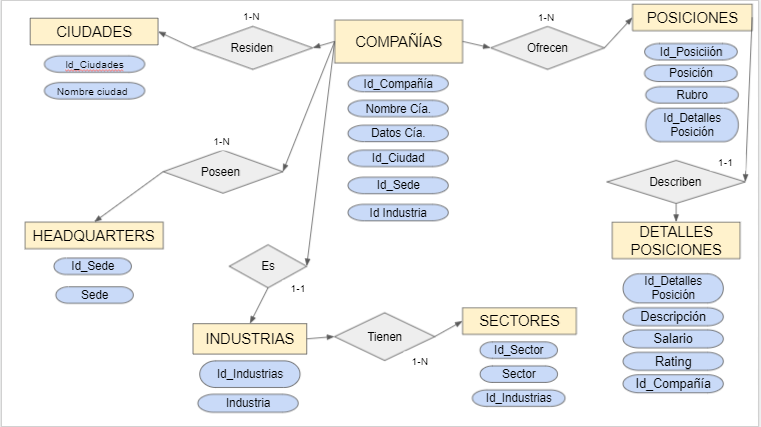
## Usuario Final y nivel de aplicación y análisis

Los usuarios finales serían los profesionales de IT y aquellas personas interesadas en esta industria: pueden ser free-lancers, asistentes a cursos de Data, etc. También identificamos como usuario a los departamentos de HR de las empresas de Tecnología.

## 

# Fuente de Datos

## Modelo Relacional



## Base de Datos

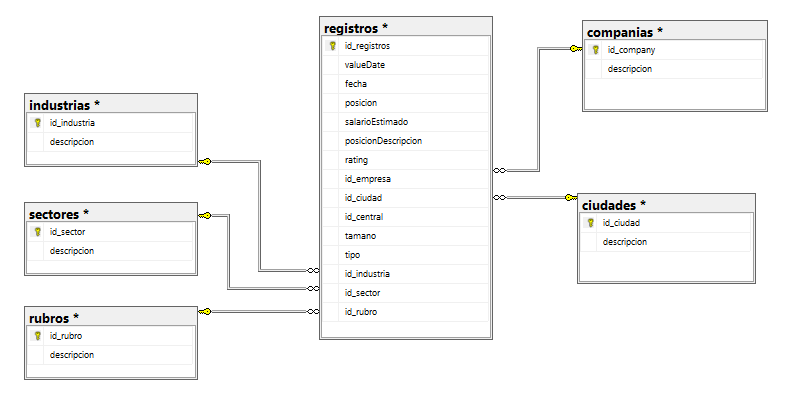
La base de datos se compuso unificando 3 datasets encontrados, los cuales eran un extracto de posiciones laborales ofrecidas durante el año 2020 para puestos de Business Analyst, Data Analyst y Data Science principalmente en Estados Unidos.

Para poder implementar un modelo de base de datos relacional, tuvimos que distribuir los datos en tablas auxiliares y de esta manera relacionar las tablas para poder elaborar consultas que nos provean de información precisa ante una necesidad. Además, hubo un trabajo de normalización en la calidad de los datos ya que los csv crudos tenían columnas fuera de lugar o campos vacíos. Al mismo tiempo, las compañías comenzaban a repetirse ya que empresas que buscaban Business Analysts también buscaban Data Analysts, por lo tanto se tuvo que realizar un trabajo en la limpieza de los datos con el fin de registrar una sola vez la compañía y evitar la duplicación de registros.

Link de la base de datos: [aquí](https://docs.google.com/spreadsheets/d/1FD1rYyp3KrWr7IFm0LmsRpo1fxy0ZAUl4zscPmIyGk8/edit?usp=sharing)

## Cómo se relacionan las tablas

Como fue mencionado previamente, la base de datos está compuesta por una tabla principal y varias tablas auxiliares con las cuales se completan los registros. La tabla registros es la que concentra todas las posiciones de trabajo, con el detalle de las mismas, y las tablas auxiliares son las que complementan con información relacionada a la posición, a la industria o a la compañía.



En el archivo de la base de datos, se puede ver que cada registro tiene información propia de la posición y campos con IDs de referencia.

## Tipos de datos en cada tabla

**Tabla**: Ciudades  
**Descripción**: esta tabla contiene 1019 ciudades en donde la mayoría son de Estados Unidos. Hay ciudades de otras partes del mundo debido a que la ciudad es para identificar tanto las oficinas de la empresa, como la sede central.

| **Columna** | **Tipo** | **Descripción** | **Columna del índice** | **Null** | **Descripción** |
| --- | --- | --- | --- | --- | --- |
| id\_ciudad | int | ID de la ciudad | Índice único | No | Es el identificador único para cada ciudad. Se utiliza para referenciar tanto al campo ciudad como al campo sede. |
| descripcion | nvarchar(500) | Nombre de la ciudad |  | Si | Contiene el nombre propio de cada ciudad |

**Tabla**: Industrias  
**Descripción**: la tabla contiene 107 registros diferentes que tienen relación con el sector en donde se desempeñan actividades las empresas que ofrecen los puestos de trabajo.

| **Columna** | **Tipo** | **Descripción** | **Columna del índice** | **Null** | **Descripción** |
| --- | --- | --- | --- | --- | --- |
| id\_industria | int | ID de la industria | Índice único | No | Es el identificador único para cada industria. |
| descripcion | nvarchar(500) | Nombre de la industria |  | Si | Contiene el nombre que identifica al campo de trabajo de la compañía. |

**Tabla**: Companías  
**Descripción**: si bien la tabla está compuesta por 4144 registros. Son las empresas que ofrecen las posiciones de trabajo.

| **Columna** | **Tipo** | **Descripción** | **Columna del índice** | **Null** | **Descripción** |
| --- | --- | --- | --- | --- | --- |
| id\_compania | int | ID de la compañía | Índice único | No | Es el identificador único para cada compañía. |
| descripcion | nvarchar(500) | Nombre de la compañía |  | Si | Contiene el nombre de la compañía |

**Tabla**: Rubros  
**Descripción**: En esta tabla tenemos cargados los tipos de posición. Los tipos de posición son Business Analyst, Data Analyst y Data Scientist. De esta manera podemos identificar en la tabla de registros cuales son las ofertas laborales realizadas por rubro o tipo de posición.

| **Columna** | **Tipo** | **Descripción** | **Columna del índice** | **Null** | **Descripción** |
| --- | --- | --- | --- | --- | --- |
| id\_rubro | int | ID del rubro | Índice único | No | Es el identificador único para cada rubro. |
| descripcion | nvarchar(50) | Nombre del rubro |  | Si | Contiene el nombre del rubro o el tipo de posición. |

**Tabla**: Sectores  
**Descripción**: es otra tabla auxiliar en donde almacenamos los diferentes tipos de sectores que existen y se agrupan industrias.

| **Columna** | **Tipo** | **Descripción** | **Columna del índice** | **Null** | **Descripción** |
| --- | --- | --- | --- | --- | --- |
| id\_sector | int | ID del sector | Índice único | No | Es el identificador único para cada sector. |
| descripcion | nvarchar(500) | Nombre del sector |  | Si | Contiene el nombre del sector. |

**Tabla**: Registros  
**Descripción**: es la tabla principal. Básicamente es la tabla que contiene las diferentes ofertas laborales con su descripción. En esta tabla agrupamos todas las ofertas para los tres tipos de posiciones o rubros. Además, contiene datos importantes como la oficina de trabajo, la sede central de la compañía, la industria y el sector al que pertenece, una calificación correspondiente a la posición, el salario estimado y más.

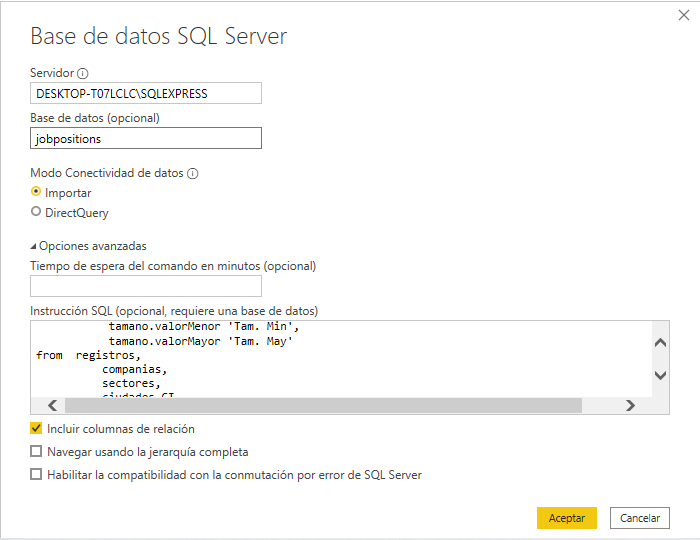
| **Columna** | **Tipo** | **Descripción** | **Columna del índice** | **Null** | **Descripción** |
| --- | --- | --- | --- | --- | --- |
| id\_registro | int | ID del registro | Índice único | No | Es el identificador único para cada propuesta laboral |
| fecha | date | fecha |  | No | Es la fecha en la que se publicó la búsqueda laboral |
| posicion | nvarchar(1000) | Encabezado |  | Si | Es el anuncio sobre la posición |
| salarioEstimado | nvarchar(1000) | Salario mínimo y máximo |  | Si | Indica cuanto es lo mínimo y máximo que se puede pagar por esa posición |
| posicionDescripcion | nvarchar(4000) | Detalle |  | Si | Es el detalle de la posición, las funciones a realizar, lo que se espera del empleado |
| rating | float | Calificación |  | Si | Es una calificación que se habrá asignado desde el portal o la agencia que publicaba estas ofertas de trabajo |
| id\_empresa | int | ID de la empresa | FK | No | Es el número de empresa que publicó la posición en cuestión |
| id\_ciudad | int | ID de la ciudad | FK | No | Corresponde a la ciudad en donde está ubicada la oficina |
| id\_central | int | ID de la oficina central | FK | No | Corresponde a la oficina central de la compañía |
| tamano | nvarchar(200) | Tamaño aproximado |  | Si | Indica la cantidad de empleados aproximados que tiene la compañía |
| id\_industria | int | ID de la industria | FK | No | Corresponde a la industria que pertenece la compañía |
| id\_sector | int | ID del sector | FK | No | Es la referencia del sector al cual pertenece la industria |
| id\_rubro | int | ID del rubro | FK | No | Es el tipo de posición al cual pertenece cada registro |

# 

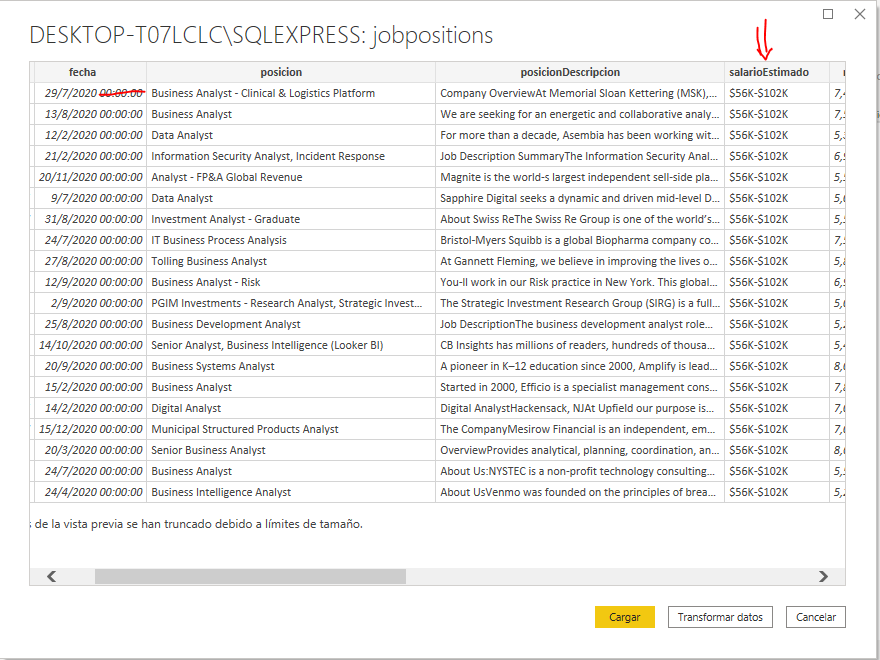
# Transformación de Datos

En nuestro proyecto, tomamos como fuente de datos la base de datos que venimos manteniendo y al momento de conectarla con Power BI, le pasamos una consulta que nos trae todos los datos como buscamos mostrarlos en los tableros.

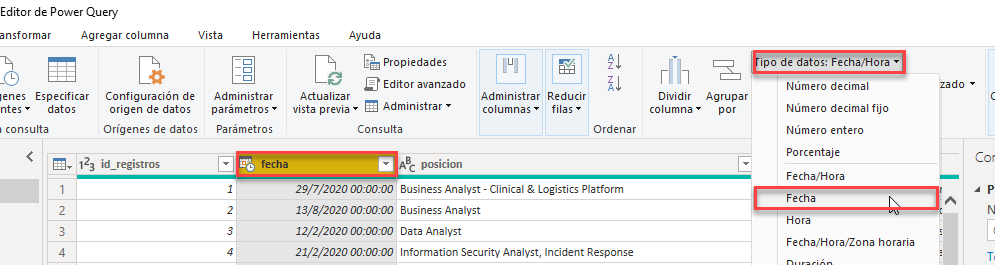
Base de Datos: [Here](https://drive.google.com/file/d/1vQKpsvQXDnGEkrKitBoDmSZ00w2-Jl2c/view?usp=sharing)  
Query: [here](https://drive.google.com/file/d/1uVN9AinVOyqq_afGWSU_YESqeD8pqF67/view?usp=sharing)



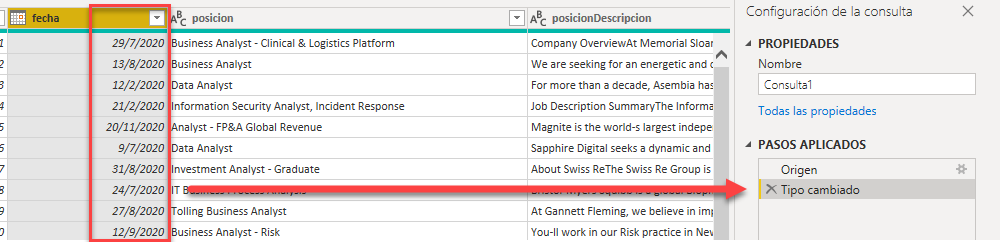
Hasta el momento no tuvimos la necesidad de crear tablas auxiliares o puentes porque la consulta con la cual generamos los datos desde el SQL, nos provee de la mayoría de la información que esperamos usar, ya que está relacionada con las tablas que contiene los datos que necesitamos, pero una vez que conectamos con Power Bi, encontramos que podemos transformar ciertos campos que nos permitirían mostrar los indicadores de una manera más simple y clara, e incluso la posibilidad de presentar un indicador más. En caso de que en el futuro, a medida que desarrollamos los indicadores planificados, vemos la necesidad de generar tablas puentes, vamos a anexar en este documento el modelo relacional entre estas tablas y la de la consulta que tenemos.



Por un lado, el campo que necesitamos transformar es el campo fecha, que viene como un datetime, y queremos manejarlo como un date. Para eso seleccionamos la columna de la fecha, hacemos click en tipo de datos, y seleccionamos el tipo de dato al que queremos cambiar.

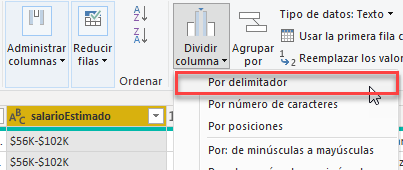
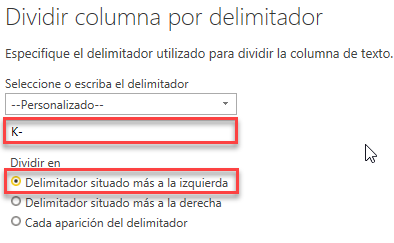


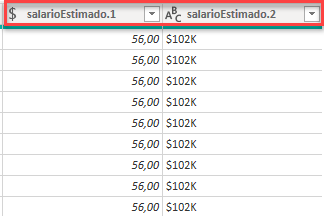
Podemos ver como se remueven los valores de horas, minutos y segundos, dejando sólo la fecha como valor y si vemos el panel de pasos aplicados, podemos ver nuestra primera transformación.



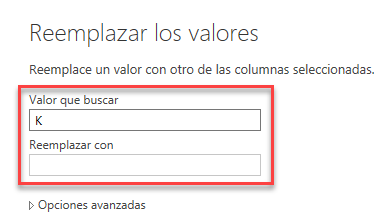
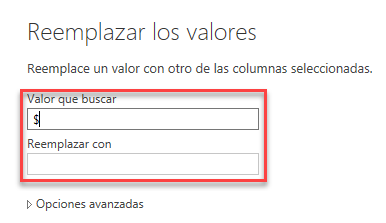
Para continuar, encontramos que uno de nuestros datos proveniente de la base de datos nos sería muy útil para desarrollar un indicador que es importante para la exposición del análisis. El campo salario. Estimado, es un campo nvarchar en nuestra base de datos y nos trae el salario mínimo y salario máximo que una empresa ofrece para cierta posición y si nosotros tenemos interés en mostrar posiciones que presenten un salario entre X y Z salario, no podríamos hacerlo sin transformar estos datos.

Las transformaciones que aplicamos son:

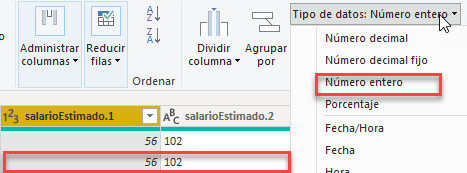
* Primero dividimos la columna utilizando como separador los caracteres “K-” y marcando la opción “delimitador situado más a la izquierda” logrando así obtener el salario mínimo en una columna y el salario máximo en otra  
   



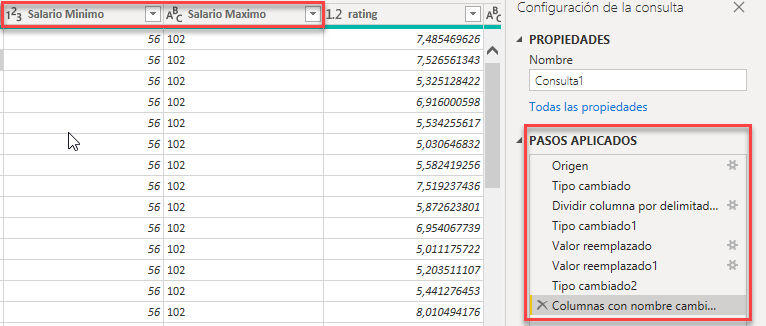
* Para la columna salario.estimado2 vamos a quitar el resto de caracteres que no necesitamos, tales como el $ y la K, y lo hacemos reemplazando estos caracteres por “vacío”.

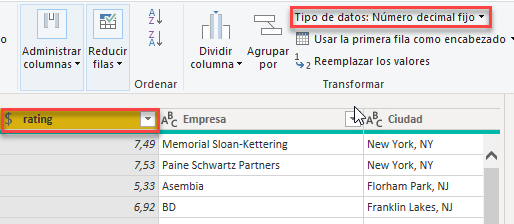
* A la columna salario.estimado1 le cambiamos el tipo de dato a número entero logrando de esta manera obtener en ambas columnas, valores que pueden ser filtrados por “mayor a”, “menor a”, “valor entre y entre” y es justamente lo que estábamos buscando.



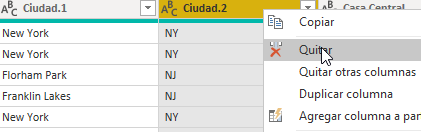
* Finalmente, cambiamos los nombres de cada columna, a donde las vamos a nombrar Salario Mínimo y Salario Máximo.



* A la columna de rating, le cambiamos el tipo de dato de formato decimal, a formato decimal fijo para que nos quede un número con dos dígitos.



* Las columnas ciudad y central, quitamos la abreviatura de cada ciudad para facilitar las tareas sobre los filtros. Para lograrlo, lo hacemos en dos pasos. Primero dividimos por columnas, usando la coma como separador, y luego quitamos la columna generada con la abreviatura.



Para finalizar estas transformaciones, damos click en cerrar y aplicar.

# Análisis funcional del tablero

## Dashboard: [here](https://drive.google.com/file/d/1fTjx68q4fzIX8cHP0Agbf2-jWNCrfBZV/view?usp=sharing)

Al momento de realizar la entrega correspondiente, el tablero cuenta con 6 páginas, en donde la primera página corresponde a la presentación del grupo y la temática.

**La segunda hoja, Resumen inicial** es un resumen del estudio realizado en base a los datos del año 2020. Tiene 6 indicadores:

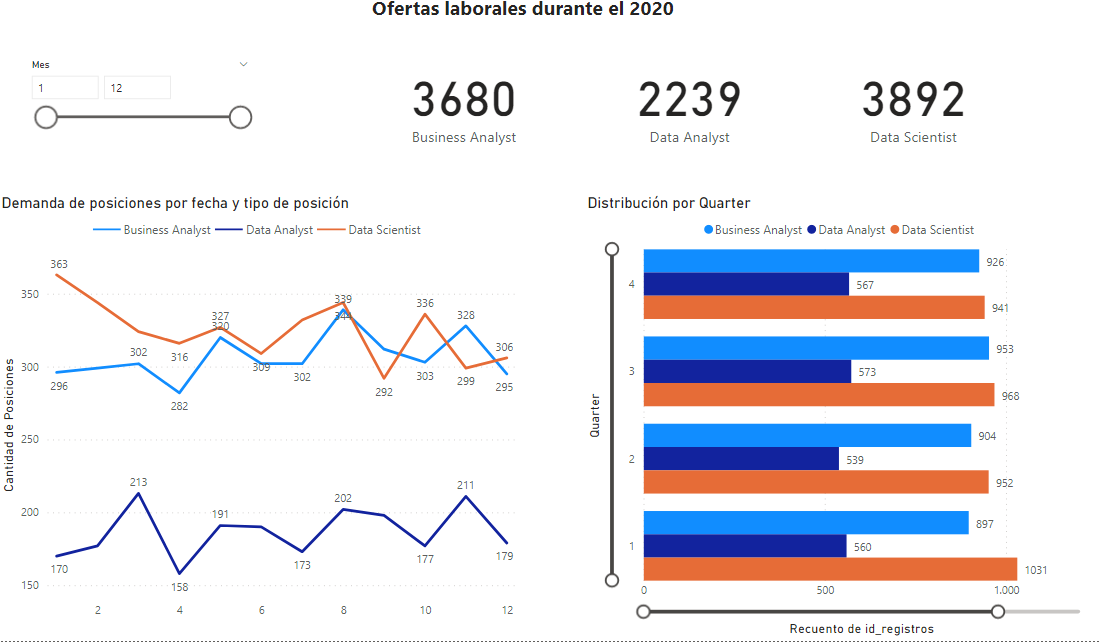
* Cantidad de posiciones analizadas
* Promedio de salario mínimo
* Promedio de salario máximo
* El rubro del mercado más demandado durante el 2020
* Cantidad de industrias que buscaron contratar los profesionales bajo estudio
* Cantidad de empresas que publicaron solicitaron al menos un profesional de los rubros estudiados.



**En la tercera hoja, Posiciones durante el tiempo,** se busca reflejar cómo fue la distribución de la demanda de posiciones durante el año 2020. Esta página cuenta con:

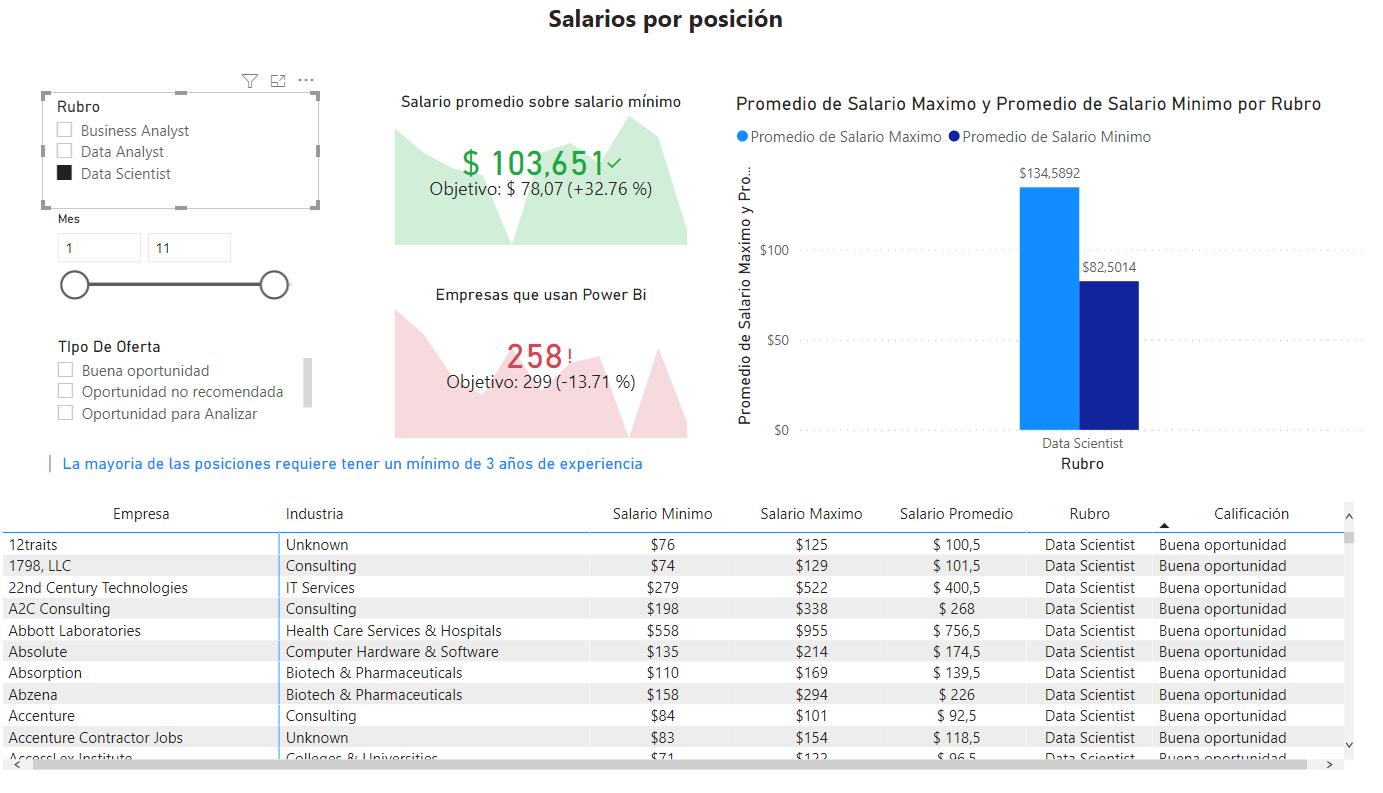
* Filtro para segmentar resultados entre el mes 1 y el 12.
* Cantidad de posiciones por rubro (BA, DA, DS)
* Distribución lineal por mes que muestra la cantidad de posiciones demandadas por rubro durante el año o durante el segmento seleccionado del filtro.
* Distribución por cuatrimestre en relación a la cantidad de posiciones ofrecidas.

Haciendo click en cualquiera de los gráficos, se aplican los filtros en todos los indicadores.



En **la cuarta hoja, Salarios por posición,** se hace un análisis sobre la relación salarios-posición, en donde tenemos:

* Filtro por rubro (selección única)
* Filtro por mes
* Filtro por tipo de oportunidad
* KPI entre el salario promedio sobre el salario mínimo.
* KPI sobre Empresas que usan Power Bi
* Gráfico de barras en donde indica el promedio de salario mínimo y máximo para un rubro y durante el tiempo segmentado.
* Tabla con las empresas que ofrecen dichas posiciones.

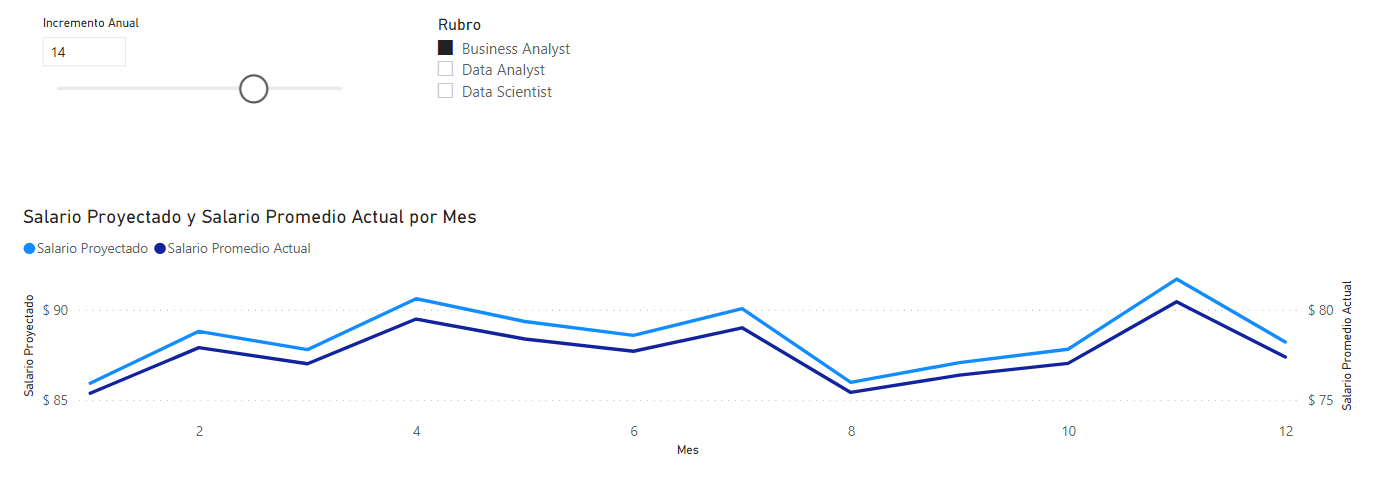


En la **quinta hoja, posiciones que requieren Power Bi y/o SQL,** tiene información que puede ser de utilidad al momento de querer indagar sobre las posiciones, salarios y los rubros en general. Hasta el momento están agrupadas en una hoja llamada resumen, pero posiblemente estos indicadores se distribuyen en las demás hojas una vez que el tablero esté más cerca de su finalización.

* Filtro por rubro
* Filtro si la posición requiere de Power Bi o no
* Filtro si la posición requiere de SQL o no
* Filtro por industria
* Filtro por mes
* Tarjetas con resultados basados en los filtros:
  + Cantidad de posiciones
  + Promedio de salario mínimo
  + Promedio de salario máximo
  + Cantidad de empresas que ofrecen posiciones similares
  + Cantidad de industrias a las que pertenecen estas empresas
  + Cantidad de ciudades en donde están localizadas las empresas
* Gráfico de barras que indica que ciudad ofrece más posiciones según los filtros aplicados.
* Posiciones y detalles de la job description.



Finalmente, en la última hoja, **Proyección Salarial 2021,** buscamos reflejar como los salarios para las posiciones bajo estudio se incrementarían en un rango del 1 al 20%.



# Medidas Calculadas

## Listado de medidas calculadas

* **Cantidad de Ciudadas**
  + Con esta medida buscamos el total de ciudades en donde radican las empresas para identificar geográficamente las ubicaciones más convenientes por rubro.
* **Cantidad de Empresas**
  + Esta medida nos permite tener el total de empresas ofreciendo puestos de trabajo relacionados a los rubros bajo estudio y la aplicamos en diversos indicadores (desarrollados y por desarrollar)
* **Cantidad de Industrias**
  + Esta medida es simplemente para obtener el total de industrias implicadas en el estudio y cuál es la industria que más demanda en el mercado.
* **Promedio salario mínimo**
  + Esta medida nos permite obtener el promedio de salario mínimo y lo aplicamos en los indicadores relacionados con los salarios.
* **Promedio salario máximo**
  + Esta medida nos permite obtener el promedio de salario máximo y lo aplicamos en los indicadores relacionados con los salarios.
* **Salario promedio**
  + Esta medida nos permite obtener el promedio de salario entre el mínimo y el máximo para presentar de manera más general lo que un profesional puede ganar por su trabajo y determinar cuál de los rubros es el que mejor pagó durante el 2020.
* **Total salario máximo y salario mínimo**
  + Estas medidas las usamos para sacar KPI entre sueldo promedio y mínimo-máximo
* **Años de experiencia**
  + Esta medida la creamos para saber el promedio de años de experiencia que se piden en las posiciones. Para hacer la medida de años promedios, primero necesitábamos el dato, por lo cual le dimos origen a esta tabla obteniendo la información desde el campo posicionDescripcion.
* **Tipo de Oferta**
  + Con esta medida buscamos catalogar las posiciones. Aplicamos 2 filtros en 3 rangos:
    - Oportunidad no recomendada
      * Salario menor a 85k
      * Calificación de la empresa menor o igual a 6
    - Oportunidad para analizar
      * Salario menor a 85k
      * Calificación de la empresa mayor a 6
    - Buena oportunidad
      * Salario mayor a 85k
      * Calificación de la empresa mayor o igual a 7
* **Mínimo de Experiencia**
  + Esta medida la calculamos para poder lograr cumplir con el pedido de entrega. Si bien no teníamos el dato, obtuvimos algunos datos como para poder hacer un cálculo. Para lograrlo, se declararon 2 variables en donde se almacena la sumatoria de los años requeridos entre todas las posiciones, y en otra variable se almacena la cantidad de posiciones ofrecidas. En base a estas dos variables se obtiene el promedio de años de experiencia que requieren la mayoría de las posiciones. Por lo tanto, con un condicional obtenemos el mensaje.

| Mínimo de Experiencia =  var anos = CALCULATE(sum(Consulta1[Años de experiencia])) var cantidad = CALCULATE(count(Consulta1[id\_registros])) var prom = anos / cantidad return if(prom < 3,"La mayoria de las posiciones requiere tener un mínimo de 3 años de experiencia","Requiere 3 años o más de experiencia") |
| --- |

* **SalarioPromIncremento**:
  + Esta medida fue creada para aplicar el uso de parámetros. Lo que hacemos en esta medida es proyectar cuánto se incrementan los salarios promedios estimando que exista un aumento de 20% anual. El indicador está diferenciado de los demás en la hoja “con parametro” ya que aun no está definido si será parte del entregable final o no.
* **Incremento salario promedio**
  + Incremento de Salario Promedio = GENERATESERIES(1, 30, 1)

## Tablas generadas

* **Calendario**
  + Es una tabla de fechas, en donde buscamos poder manejar el filtro de fechas por día, mes, quarter y semana.
* **Índice**
  + Tabla generada para identificar la cantidad de registros únicos.
* **Salarios**
  + Tabla en donde trabajamos los salarios con relación a las empresas y a las industrias.