

# Desafío 7: Primera Entrega del Proyecto Final

Nombre del proyecto: **Data Science Jobs Salaries**

Nombres integrantes: Daniel Ramirez Daniel Arbelaez Daniela Cano

## PRIMERA ENTREGA DEL PROYECTO FINAL

**Formato:** La base de datos debe ser un archivo plano de tipo xls, csv o txt. La documentación debe ser formato pdf.

**Sugerencia:** activar comentarios en el archivo, y también la opción compartir.

Proyecto  
Final



### >>Objetivos Generales:

1. Reconocer el modelo relacional.
2. Documentar la iniciativa de análisis de datos.

### >>Objetivos Específicos:

1. Reconocer las tablas
2. Diseñar el modelo relacional.
3. Identificar los tipos de datos de las tablas.
4. Reconocer cómo se relaciona las tablas.

## PRIMERA ENTREGA DEL PROYECTO FINAL

**Formato:** La base de datos debe ser un archivo plano de tipo xls, csv o txt. La documentación debe ser formato pdf.

**Sugerencia:** activar comentarios en el archivo, y también la opción compartir.

Proyecto  
Final



### >>Se debe entregar:

Cada documento debe llevar por título el nombre del proyecto y los nombres de los integrantes. En el caso de la base de datos, agregarlos en un documento txt adicional o en la primera solapa.

### En el archivo pdf:

1. Descripción de la temática de los datos.
2. Diagrama entidad-relación de las tablas seleccionadas.
3. Listado de tablas, con definición de clave primaria y/o clave foránea, según corresponda.
4. Listado de columnas por tablas, con definiciones de tipos de datos.

### En el archivo plano:

1. Cada solapa/archivo debe llevar un nombre claro que permita identificar la temática.

## Tabla de contenido

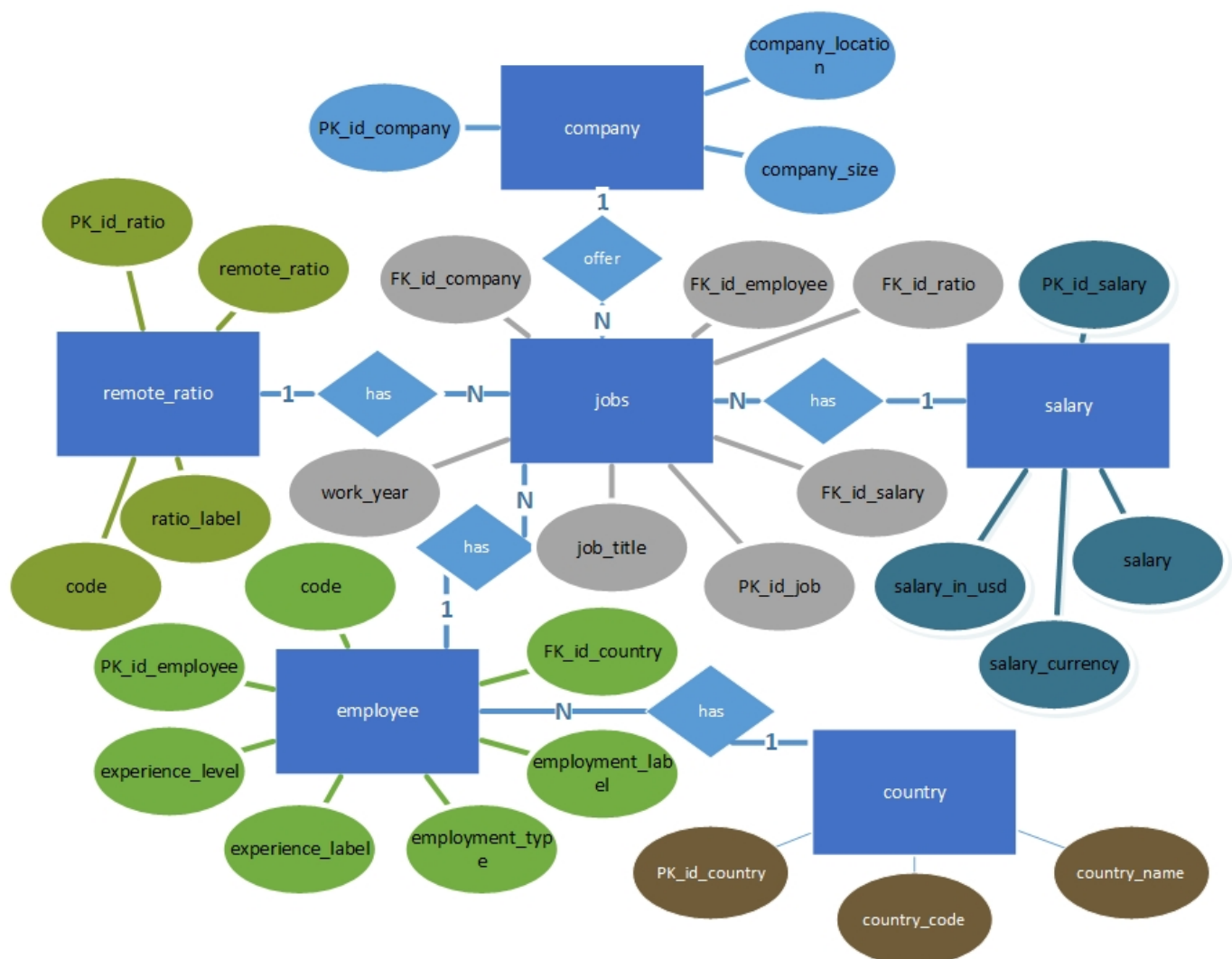
- [Desafío 7: Primera Entrega del Proyecto Final](#)
  - [Tabla de contenido](#)
  - [Descripción de la temática de los datos](#)
  - [Diagrama Entidad - Relación](#)
  - [Listado de tablas con PK y FK](#)
  - [Listado de columnas por tablas con tipo de datos](#)

## Descripción de la temática de los datos

En el presente trabajo se realizó inicialmente una búsqueda de datasets que comprendían datos sobre salarios de los empleos en áreas de data science; seguido se realizó inicialmente el análisis de los datos que comprendían el dataset elegido, en este caso, "Data Science Jobs Salaries", el cual fue encontrado en el repositorio *Kaglee* al cual se puede acceder a través del siguiente [link](#)

## Diagrama Entidad - Relación

Para el diseño del presente diagrama se tuvo en cuenta que la tabla jobs se consideró como la tabla hechos por tratarse del tema principal en cuanto a los salarios de empleos en áreas de Data Science, las demás fueron clasificadas como tablas de dimensiones. El diagrama se realizó en el programa de [Microsoft Visio](#) y se exportó en formato `.jpg`:



## Listado de tablas con PK y FK

- Tabla "job":

PK_id_job	job_title	work_year	FK_id_ratio	FK_id_company	FK_id_salary	FK_id_employee
1	Data Science Consultant	2021	2	1	1	1
2	Data Scientist	2020	3	2	2	2
3	Head of Data Science	2021	1	3	3	3
4	Head of Data	2021	2	4	4	4
5	Machine Learning Engineer	2021	3	5	5	5
6	Data Analytics Manager	2021	3	6	6	6
7	Research Scientist	2020	1	7	7	7
8	Data Analyst	2020	2	8	8	8
9	Data Engineer	2020	2	9	9	9
10	Data Science Engineer	2021	2	10	10	10

- Tabla "remote\_ratio":

PK_id_rat	code	remote_ratio	ratio_label
1	2	50	Partially remote
2	3	100	Fully remote
3	1	0	No remote work
4	2	50	Partially remote
5	3	100	Fully remote
6	3	100	Fully remote
7	1	0	No remote work
8	2	50	Partially remote

- Tabla "company"

PK_id_company	company_location	company_size
1	DE	L
2	US	L
3	RU	M
4	RU	L
5	US	S
6	US	M
7	US	M
8	FR	L

- Tabla "salary"

PK_id_salary	salary	salary_currency	salary_in_usd	tipo_caml
1	54000	EUR	64369	1,19201852
2	60000	EUR	68428	1,14046667
3	85000	USD	85000	1
4	230000	USD	230000	1
5	125000	USD	125000	1
6	120000	USD	120000	1
7	450000	USD	450000	1
8	41000	EUR	46759	1,14046341
9	65000	EUR	74130	1,14046154
10	158500	CAD	127512	0,79864262

- Tabla "employee"

PK_id_em	code	experience_level	experience_label	employment_type	employment_label	FK_id_cou
1		1 EN	Entry-level / Junior	FT	Full-time	84
2		3 SE	Senior-level / Expert	FT	Full-time	87
3		4 EX	Executive-level / Director	FT	Full-time	184
4		4 EX	Executive-level / Director	FT	Full-time	184
5		1 EN	Entry-level / Junior	FT	Full-time	237
6		3 SE	Senior-level / Expert	FT	Full-time	237
7		2 MI	Mid-level / Intermediate	FT	Full-time	237
8		2 MI	Mid-level / Intermediate	FT	Full-time	77

Listado de columnas por tablas con tipo de datos

Tabla	Columna	Tipo de dato
job	PK_id_job	Int
job	job_title	Varchar(40)
job	work_year	Varchar(5)
job	FK_id_ratio	Int
job	FK_id_company	Int
job	FK_id_salary	Int
job	FK_id_employee	Int
-	-	-
ratio	PK_id_ratio	Int
ratio	code	Int
ratio	remote_ratio	Int
ratio	ratio_label	Varchar(20)
-	-	-
company	PK_id_company	Int
company	company_location	Varchar(5)
company	company_size	Varchar(5)
-	-	-
salary	PK_id_salary	Int
salary	salary	Int
salary	salary_currency	Varchar(5)
salary	salary_in_usd	Int
salary	tipo_cambio	Int
-	-	-
employee	PK_id_employee	Int

Tabla	Columna	Tipo de dato
employee	code	Varchar(5)
employee	experience_level	Varchar(5)
employee	experience_label	Varchar(40)
employee	employment_type	Varchar(5)
employee	employment_label	Varchar(20)
employee	FK_id_country_code	Varchar(5)
-	-	-
country_code	PK_id_country	Int
country_code	code	Varchar(5)
country_code	country_name	Varchar(40)