

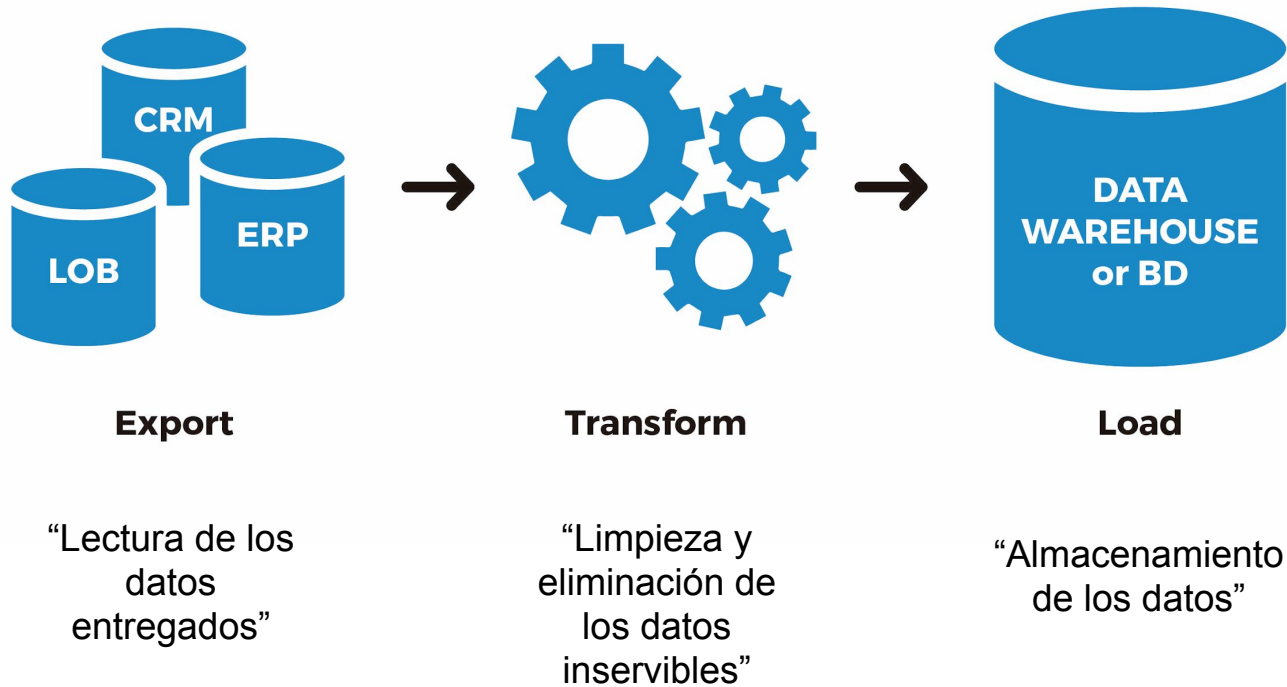


Universidad
Católica del Norte

Proyecto Analítico de Datos “SIES”

Johan Ordenes
Leonel Villagra

SIES



Lectura de archivo .xlsx

ARANCEL ANUAL	COSTO TITULACION	AÑO DUR AC	DURACION CARRERA FORMAL	NIVEL CARRERA O TIPO DE CARRER	AÑO MATRÍCULA	TOTAL MATRÍCULA FEMENINO	TOTAL MATRÍCULA MASCULINO	TOTAL MATRÍCULA	MATRÍCULA DE 1ER AÑO FEMENINO	MATRÍCULA 1ER AÑO MASCULINO	TOTAL MATRÍCULA 1ER AÑO	MATRÍCULA - % DE MUNICIPAL	MATRÍCULA - % DE PARTICULAR SUBVENCIONADO	MATRÍCULA - % DE PARTICULAR PAGADO	C. Administración Delegada	AÑO TITULADOS	TITULADOS FEMENINO	TITULADOS MASCULINO	TOTAL TITULADOS	% DE COBERTURA PSU EN MATRÍCULA 1ER AÑO
1.550.000	161.000	2017	5	Técnica de nivel s	2017	29	79	108	15	35	50	51,4%	47,6%	0,0%	1,0%	2016	0	0	0	-
1.550.000	161.000	2017	5	Técnica de nivel s	2017	35	86	121	16	33	49	51,7%	48,3%	0,0%	0,0%	2016	0	0	0	-
1.550.000	161.000	2017	5	Técnica de nivel s	2017	19	48	67	7	24	31	62,1%	37,9%	0,0%	0,0%	2016	0	0	0	-
1.390.000	36.000	2017	4	Técnica de nivel s	2017	0	0	0	0	0	0	NULL	NULL	NULL	NULL	2016	0	0	0	-
3.434.000	214.200	2017	8	Profesional	2017	42	12	54	35	11	46	15,4%	46,2%	38,5%	0,0%	2016	0	0	0	60% <= X < 80%
2.214.000	0	2017	10	Profesional	2017	301	32	333	50	5	55	28,3%	71,1%	0,6%	0,0%	2016	29	1	30	40% <= X < 60%
2.162.000	0	2017	10	Profesional	2017	292	50	342	50	17	67	25,4%	68,3%	3,0%	3,3%	2016	27	1	28	60% <= X < 80%
2.214.000	0	2017	10	Profesional	2017	50	13	63	44	13	57	16,9%	79,7%	0,0%	3,4%	2016	0	0	0	60% <= X < 80%
2.214.000	0	2017	10	Profesional	2017	244	33	277	0	0	0	56,6%	34,5%	4,1%	4,9%	2016	58	6	64	-
2.382.000	0	2017	10	Profesional	2017	235	32	267	76	15	91	18,4%	63,9%	11,4%	6,3%	2016	16	2	18	40% <= X < 60%
3.382.924	60.000	2017	10	Profesional	2017	238	33	271	49	5	54	13,9%	78,2%	6,0%	1,9%	2016	38	6	44	0% <= X <= 10%
3.160.000	0	2017	10	Profesional	2017	120	71	191	48	32	80	16,5%	59,3%	22,5%	1,6%	2016	21	19	40	60% <= X < 80%
2.900.000	0	2017	10	Profesional	2017	0	0	0	0	0	0	NULL	NULL	NULL	NULL	2016	0	0	0	-
1.240.000	50.000	2017	4	Técnica de nivel s	2017	136	0	136	73	0	73	69,4%	25,6%	0,0%	5,0%	2016	18	0	18	-
1.240.000	50.000	2017	4	Técnica de nivel s	2017	93	0	93	39	0	39	68,2%	31,8%	0,0%	0,0%	2016	30	0	30	-
1.240.000	50.000	2017	4	Técnica de nivel s	2017	63	0	63	39	0	39	53,3%	46,7%	0,0%	0,0%	2016	27	0	27	-
1.240.000	50.000	2017	4	Técnica de nivel s	2017	88	0	88	59	0	59	50,0%	44,2%	0,0%	5,8%	2016	25	0	25	-
1.240.000	50.000	2017	4	Técnica de nivel s	2017	66	0	66	31	0	31	72,4%	25,9%	0,0%	1,7%	2016	12	0	12	-
1.883.000	0	2017	8	Profesional	2017	121	80	201	24	21	45	12,5%	30,5%	57,0%	0,0%	2016	8	10	18	60% <= X < 80%
1.833.300	26.472	2017	5	Técnica de nivel s	2017	64	28	92	38	15	53	13,4%	59,8%	22,0%	4,9%	2016	8	1	9	-
1.833.300	26.472	2017	5	Técnica de nivel s	2017	71	44	115	40	28	68	22,0%	64,2%	3,7%	10,1%	2016	4	4	8	-
2.496.000	100.000	2017	10	Profesional	2017	299	37	336	65	7	72	49,6%	41,5%	1,8%	7,2%	2016	42	2	44	0% <= X <= 10%
2.269.000	325.674	2017	10	Profesional	2017	66	12	78	24	7	31	33,3%	65,3%	1,3%	0,0%	2016	13	0	13	-
2.690.000	325.674	2017	10	Profesional	2017	57	8	65	12	4	16	51,8%	37,5%	10,7%	0,0%	2016	17	0	17	-
2.290.000	325.674	2017	10	Profesional	2017	0	0	0	0	0	0	NULL	NULL	NULL	NULL	2016	0	0	0	-
2.798.000	325.674	2017	10	Profesional	2017	26	32	58	7	5	12	21,8%	67,3%	7,3%	3,6%	2016	2	4	6	-
2.498.000	325.674	2017	10	Profesional	2017	41	63	103	9	21	30	34,0%	60,0%	6,0%	0,0%	2016	0	0	0	-
2.690.000	325.674	2017	10	Profesional	2017	45	44	89	16	21	37	62,5%	27,5%	10,0%	0,0%	2016	2	2	4	-
2.910.000	325.674	2017	10	Profesional	2017	108	80	188	25	26	51	22,5%	67,4%	0,6%	9,6%	2016	17	17	34	60% <= X < 80%
4.024.000	464.000	2017	8	Profesional	2017	223	37	260	60	7	67	6,9%	29,4%	63,3%	0,4%	2016	30	6	36	60% <= X < 80%
2.993.000	183.000	2017	8	Profesional	2017	0	0	0	0	0	0	NULL	NULL	NULL	NULL	2016	0	0	0	-
3.888.000	214.000	2017	8	Profesional	2017	86	41	127	22	17	39	15,6%	72,1%	11,5%	0,8%	2016	8	1	9	0% <= X <= 10%
1.336.000	229.000	2017	5	Técnica de nivel s	2017	176	48	224	57	22	79	41,8%	56,0%	0,0%	2,2%	2016	57	8	65	-
1.320.000	229.000	2017	5	Técnica de nivel s	2017	92	11	103	44	2	46	83,5%	11,0%	5,5%	0,0%	2016	0	0	0	-
1.200.000	229.000	2017	5	Técnica de nivel s	2017	148	26	174	64	15	79	69,1%	30,9%	0,0%	0,0%	2016	0	0	0	-
1.500.000	50.000	2017	5	Técnica de nivel s	2017	173	36	209	62	11	73	58,2%	35,7%	0,0%	8,2%	2016	39	7	46	-
1.500.000	50.000	2017	5	Técnica de nivel s	2017	188	30	218	68	11	79	61,6%	38,4%	0,0%	0,0%	2016	40	3	43	-
1.500.000	50.000	2017	5	Técnica de nivel s	2017	161	19	180	56	5	61	42,2%	47,4%	0,0%	10,4%	2016	78	12	90	-
3.740.033	418.603	2017	10	Profesional	2017	163	160	323	32	37	69	29,2%	68,0%	0,9%	1,9%	2016	41	32	73	0% <= X <= 10%
3.238.400	373.456	2017	10	Profesional	2017	256	207	463	60	24	84	37,9%	59,3%	0,7%	2,2%	2016	31	38	69	0% <= X <= 10%
3.862.500	429.625	2017	10	Profesional	2017	258	203	461	40	35	75	24,4%	71,0%	1,3%	3,3%	2016	35	34	69	0% <= X <= 10%
4.243.600	463.924	2017	10	Profesional	2017	211	179	390	53	21	74	22,7%	70,8%	3,7%	2,9%	2016	32	30	62	0% <= X <= 10%
3.246.000	86.000	2017	10	Profesional	2017	103	82	185	21	20	41	36,3%	55,5%	8,2%	0,0%	2016	22	15	37	0% <= X <= 10%
3.778.977	60.000	2017	10	Profesional	2017	196	222	418	38	58	96	21,4%	66,5%	9,1%	3,0%	2016	15	10	25	0% <= X <= 10%
3.675.197	60.000	2017	10	Profesional	2017	209	73	282	78	24	102	12,2%	69,1%	18,3%	0,4%	2016	0	0	0	60% <= X < 80%
1.420.000	107.010	2017	4	Técnica de nivel s	2017	68	24	92	19	7	26	12,8%	67,4%	2,3%	17,4%	2016	15	5	20	-
3.018.555	60.000	2017	8	Profesional	2017	124	8	132	40	17	57	23,4%	66,7%	7,1%	2,7%	2016	28	63	91	0% <= X <= 10%
2.252.750	421.000	2017	10	Profesional	2017	114	105	219	27	29	56	22,3%	77,2%	0,5%	0,0%	2016	28	0	28	0% <= X <= 10%
2.280.560	421.000	2017	10	Profesional	2017	104	59	163	18	12	30	57,4%	40,7%	0,0%	1,9%	2016	21	5	26	0% <= X <= 10%
1.550.000	161.000	2017	5	Técnica de nivel s	2017	229	1	230	110	1	111	44,1%	55,9%	0,0%	0,0%	2016	0	0	0	-

Datos erróneos

ARANCEL ANUAL

Ordenar

z↓ Ascendente A↓ Descendente

Por color: Ninguno

Filtro

Por color: Ninguno

Elige uno

Buscar

- ✓ 7.970.500
- ✓ 7.990.000
- ✓ 8.210.237
- ✓ 8.237.788
- ✓ 8.403.095
- ✓ N/A
- ✓ NULL
- ✓ s/i

Borrar filtro

COSTO TITULACION

Ordenar

z↓ Ascendente A↓ Descendente

Por color: Ninguno

Filtro

Por color: Ninguno

Elige uno

Buscar

- ✓ (Seleccionar todo)
- ✓ -
- ✓ 0
- ✓ 5.500
- ✓ 17.000
- ✓ 18.500
- ✓ 18.860
- ✓ 19.100

Borrar filtro

NIVEL CARRERA O TIPO DE CARRERA

Ordenar

z↓ Ascendente A↓ Descendente

Por color: Ninguno

Filtro

Por color: Ninguno

Elige uno

Buscar

- ✓ (Seleccionar todo)
- ✓ Profesional
- ✓ Técnica de nivel superior
- ✓ #N/D

Borrar filtro

Limpieza de los datos

Para celdas vacías o erróneas:

- Datos tipo texto no conocidos como “nan”
- Datos tipo número no conocidos como “nan”



Funciones:

- `cleanPcobertura()`: Limpia los datos de la columna “% de cobertura psu ultimo año”
- `cleanText()`: Limpia los datos de las columnas tipo texto
- `cleanPercentages()`: Limpia los datos de tipo decimal
- `cleanDigit()`: Limpia los datos de las columnas que contienen dígitos enteros
- `cleanDecimal()`: Limpia los datos de las columnas que contienen dígitos flotantes

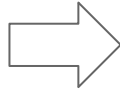
Transformaciones:

- Cambiamos el nombre a algunas tablas para poder subirlas directamente a la base de datos y evitar errores.

Limpieza con funciones

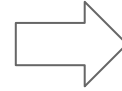
cleanPcobertura():

- Elimina '-' a NULL
- Reemplaza '% <= X' <' con ','
- Reemplaza '=' con ''
- Corta final '%'
- Divide ','



60% <= X <80%

ENTRADA



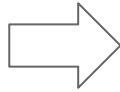
60

80

SALIDA

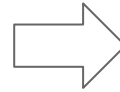
cleanText():

- Reemplaza s/i
- Reemplaza NULL



s/i

ENTRADA



nan

SALIDA

cleanPercentage():

- Limpia NULL

cleanDigit():

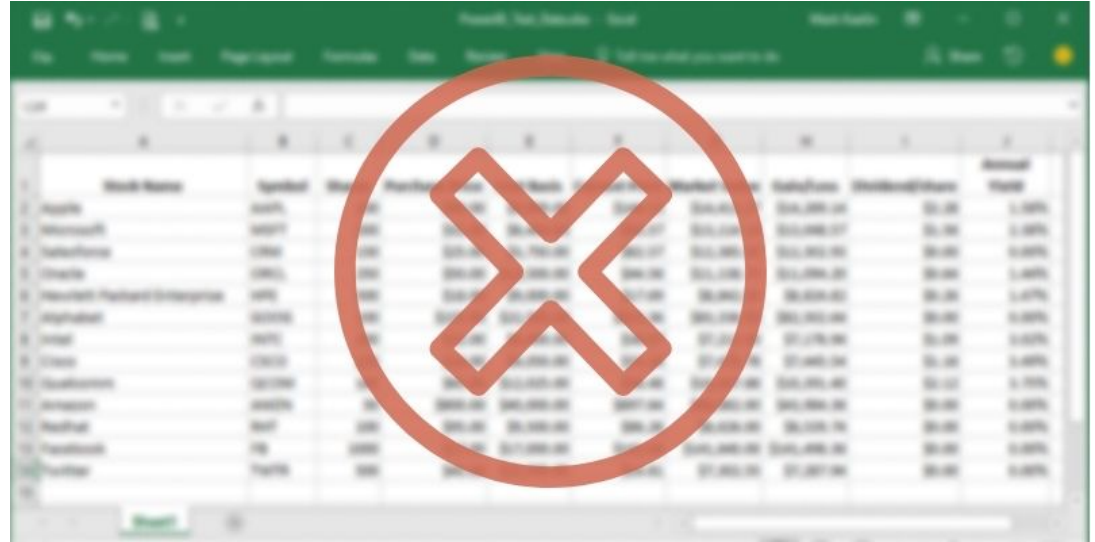
- Limpia '-'
- Limpia 's/i'

cleanDecimal():

- Limpia '-'
- Limpia 's/i'

Columnas sin limpiar

- CODIGO UNICO DE CARRERA
- AÑO_INFORM
- CODIGO DE INSTITUCIÓN
- AREA DE CONOCIMIENTO
- INSTITUCION
- NOMBRE CARRERA
- REGION
- JORNADA
- SEDE
- DURACION CARRERA FORMAL
- AÑO TITULADOS
- PSU PONDERACION NOTAS EM
- PSU PONDERACION RANKING
- PSU PONDERACION LENGUAJE
- PSU PONDERACION MATEMATICAS
- PSU PONDERACION HISTORIA
- PSU PONDERACION CIENCIAS
- PSU PONDERACION OTROS



Columnas eliminadas

Innecesarias:

- TIPO DE INSTITUCION (Todas son universitarias)

Columnas iguales:

- AÑO_DURAC (Año duracion = Año matrícula)

Representan la suma de masculino + femenino:

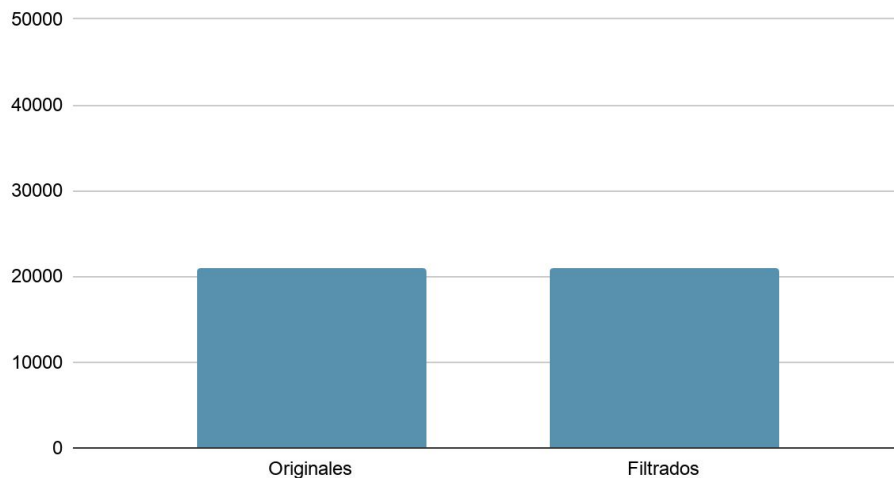
- TOTAL MATRICULA
- TOTAL MATRICULA 1ER AÑO
- TOTAL TITULADOS

Análisis de datos

	CODIGO UNICO DE CARRERA	AÑO_INFORM	CODIGO DE INSTITUCIÓN	AREA DE CONOCIMIENTO	TIPO DE INSTITUCIÓN	INSTITUCION	NOMBRE CARRERA
201884TECNICO DEPORTIVO UNIVERSITARIO	I84S1C322J1V2	2018	84	Educación	Universidades	UNIVERSIDAD DE LOS LAGOS	TECNICO DEPORTIVO UNIVERSITARIO
201884TECNICO DEPORTIVO UNIVERSITARIO	I84S2C322J1V2	2018	84	Educación	Universidades	UNIVERSIDAD DE LOS LAGOS	TECNICO DEPORTIVO UNIVERSITARIO
201884TECNICO DEPORTIVO UNIVERSITARIO	I84S6C322J1V2	2018	84	Educación	Universidades	UNIVERSIDAD DE LOS LAGOS	TECNICO DEPORTIVO UNIVERSITARIO
201877TECNICO DE NIVEL SUPERIOR EN EDUCACION PARVULARIA	I77S1C194J2V2	2018	77	Educación	Universidades	UNIVERSIDAD DE MAGALLANES	TECNICO DE NIVEL SUPERIOR EN EDUCACION PARVULARIA
20183ARTES VISUALES	I3S1C126J1V2	2018	3	Arte y Arquitectura	Universidades	UNIVERSIDAD DIEGO PORTALES	ARTES VISUALES
201854NUTRICION Y DIETETICA	I54S7C69J1V1	2018	54	Salud	Universidades	UNIVERSIDAD TECNOLOGICA DE CHILE INACAP	NUTRICION Y DIETETICA
201854NUTRICION Y DIETETICA	I54S8C69J1V1	2018	54	Salud	Universidades	UNIVERSIDAD TECNOLOGICA DE CHILE INACAP	NUTRICION Y DIETETICA
201854NUTRICION Y DIETETICA	I54S12C69J1V1	2018	54	Salud	Universidades	UNIVERSIDAD TECNOLOGICA DE CHILE INACAP	NUTRICION Y DIETETICA
201854NUTRICION Y DIETETICA	I54S13C69J1V1	2018	54	Salud	Universidades	UNIVERSIDAD TECNOLOGICA DE CHILE INACAP	NUTRICION Y DIETETICA
201854NUTRICION Y DIETETICA	I54S21C69J1V1	2018	54	Salud	Universidades	UNIVERSIDAD TECNOLOGICA DE CHILE INACAP	NUTRICION Y DIETETICA

Reducción de datos

Datos Originales vs Filtrados



Cantidad de datos

- Originales: 20944
- Filtrados: 20944

Sin pérdida de información

Tiempo de compilación

El código se compiló en promedio 40 segundos.

- La declaración de variables y funciones se tardó 6.7602715492248535 segundos
- La limpieza se tardó 7.39152979850769 segundos
- La exportación de datos en HTML tardó 19.50882053375244 segundos
- La subida a la base de datos demoró 39.80104732513428 segundos
- El código se compiló en 39.80104732513428 segundos

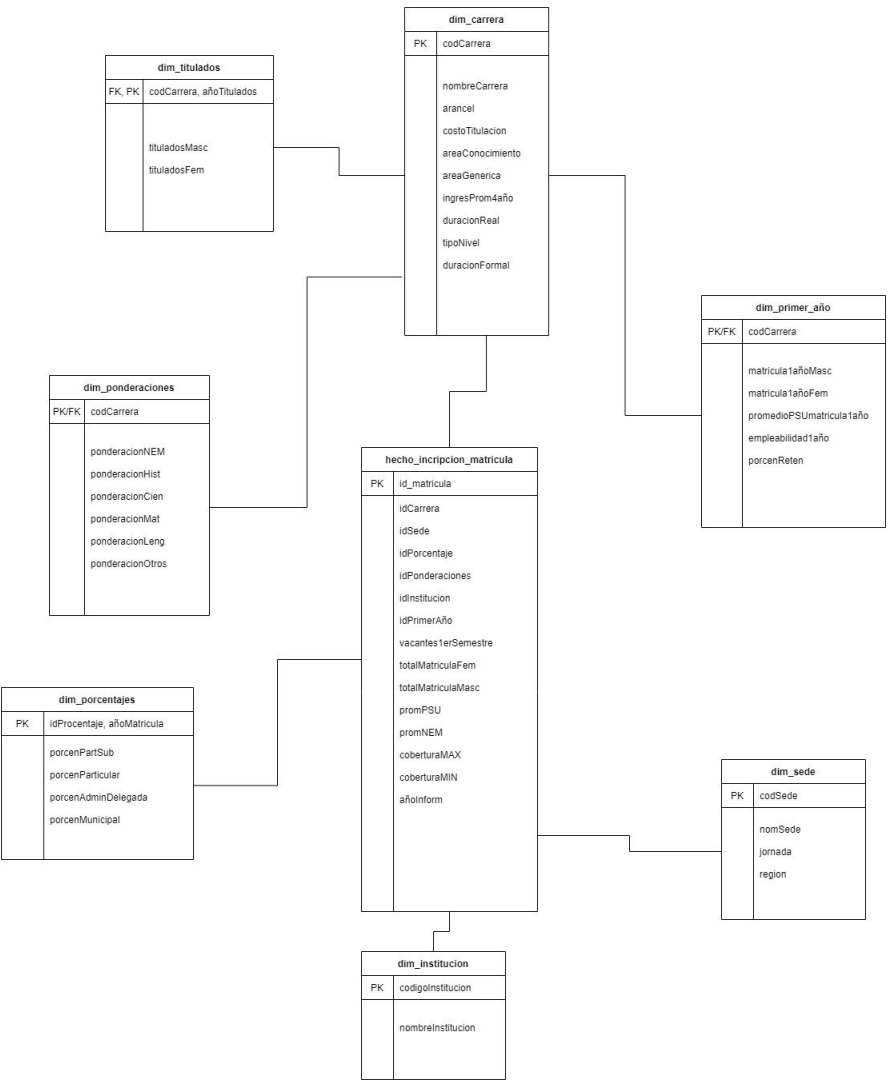
El algoritmo más costoso es de $O(n)$, en donde n es la cantidad de ciclos, los cuales se utilizaron para filtrar los datos y guardar .sies para subir a la base de datos.

Elección de tablas

- dim_titulados: Contiene los datos de los titulados.
- dim_carrera: Contiene los datos de todas las carreras.
- dim_sede: Contiene los datos de las sedes.
- dim_primer_ano: Contiene los datos de primer año de carrera.
- dim_ponderaciones: Contiene las ponderaciones para ser aceptado.
- dim_porcentajes: Contiene los porcentajes de rendimiento de una matrícula.
- dim_institucion: Contiene los datos de las instituciones.
- hecho_inscripcion_matricula: Contiene los datos de las inscripciones de matrícula

Total: 8 tablas.

Diagrama SIES RELACIONAL



Exportación de tablas en HTML

- Exportamos los datos directamente a postgres y también como archivos HTML



df_1_AÑO



df_CARRERA



df_INSTITUCION



df_MATRICULA



df_PONDERACIONES



df_PORCENTAJES



df_SEDE



df_TITULADOS

BASE DE DATOS

Admin File Object Tools Help

Dashboard Properties SQL Statistics Dependencies Dependents SIES_MULTI/postgres@PostgreSQL 13 *

Browser Servers (1) PostgreSQL 13 Databases (3) SIES SIES_MULTI Casts Catalogs Event Triggers Extensions Foreign Data Wrappers Languages Schemas (1) public Collations Domains FTS Configurations FTS Dictionaries FTS Parsers FTS Templates Foreign Tables Functions Materialized Views Procedures Sequences Tables (8) dim_1_ano dim_carrera dim_institucion dim ponderaciones dim_porcentajes dim_sede dim_titulados hecho_inscripcion_matricula Trigger Functions Types Views postgres Login/Group Roles Tablespaces

Query Editor Query History Scratch Pad

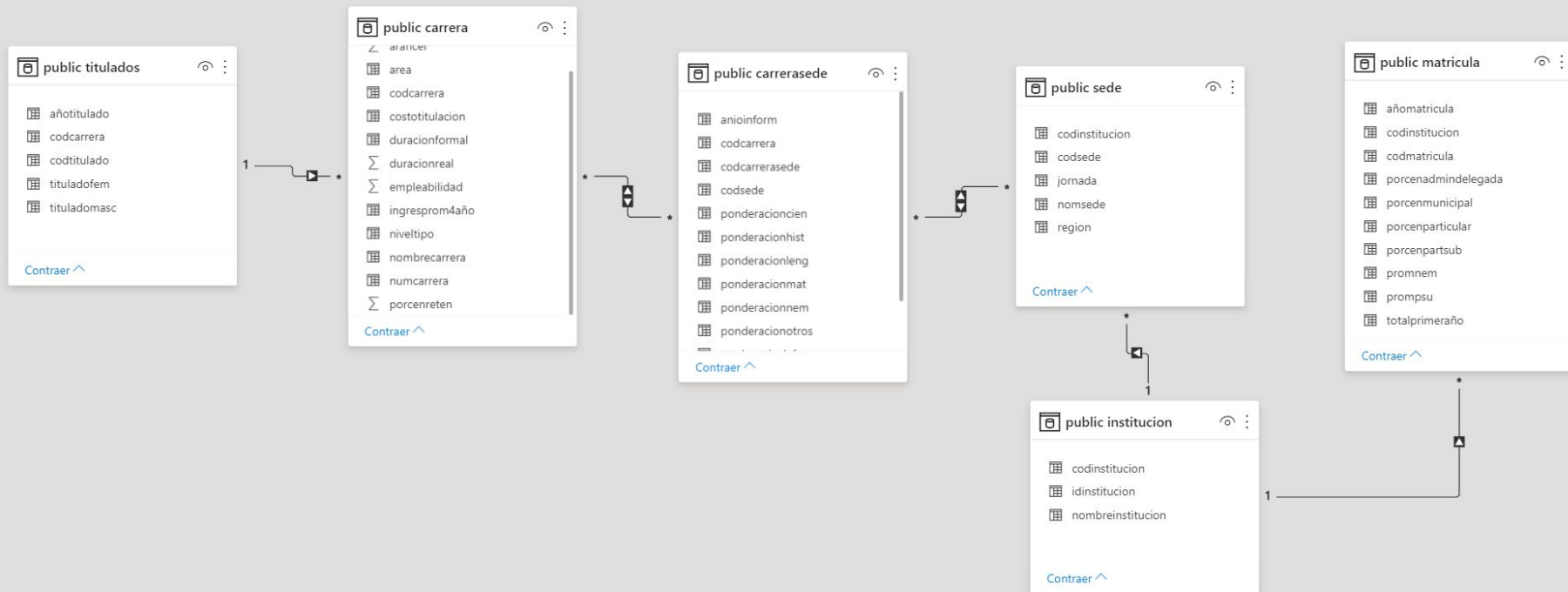
```
1 SELECT * FROM hecho_inscripcion_matricula
```

Data Output Explain Messages Notifications

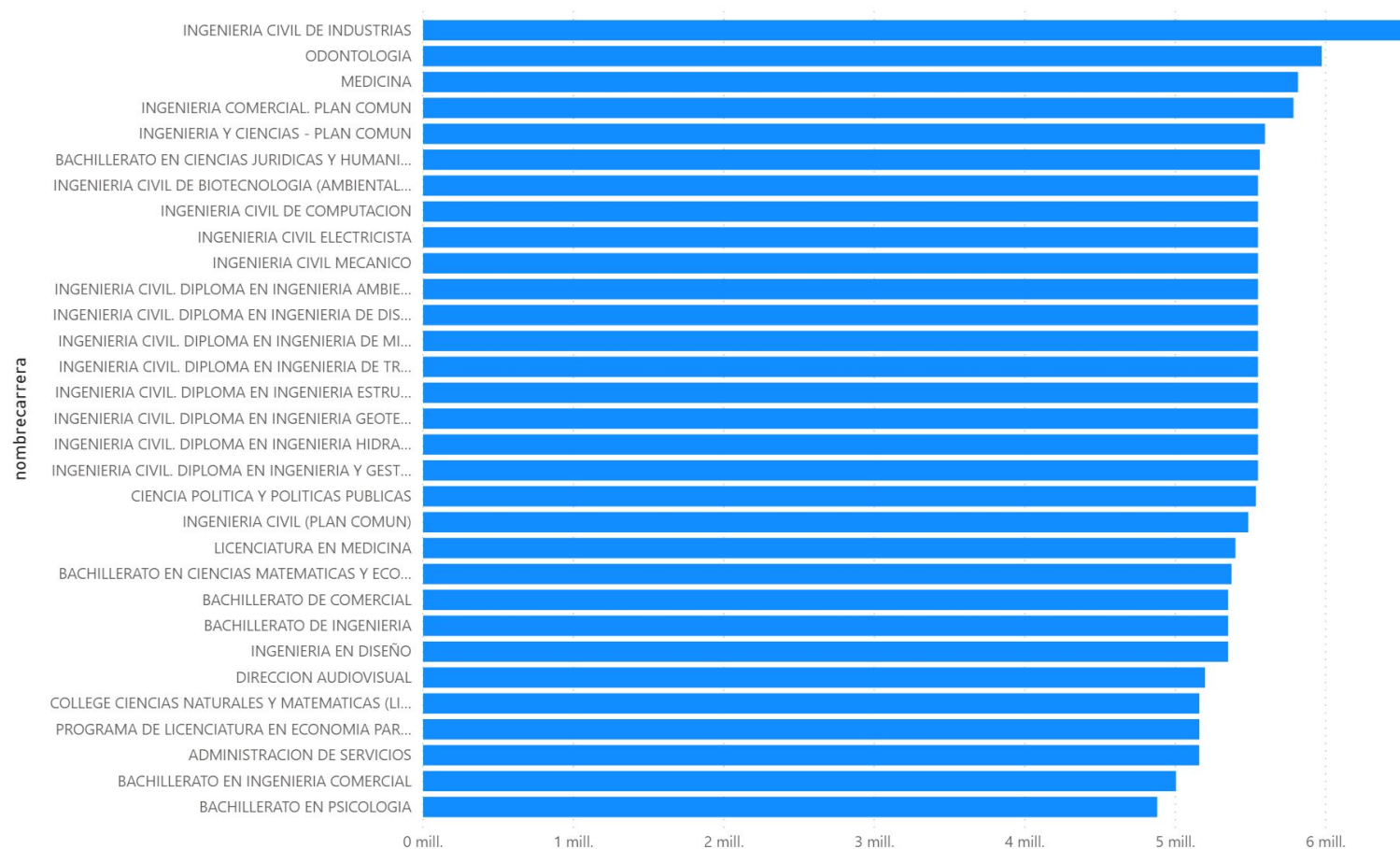
	CODIGO UNICO DE CARRERA	SEDE	CODIGO DE INSTITUCIÓN	PROMEDIO NEM EN MATRICULA	VACANTES 1ER SEMESTRE	TOTAL MATRICULA FEM
	text	text	bigint	double precision	bigint	bigint
1	I84S1C322J1V2	CASA CENTRAL (OSORNO)		84	5.48333333333333	45
2	I84S2C322J1V2	SEDE PUERTO MONTT		84	5.427659574468086	45
3	I84S6C322J1V2	SEDE CASTRO		84	5.370967741935483	35
4	I77S1C194J2V2	CASA CENTRAL (PUNTA AR...		77	0	20
5	I3S1C126J1V2	CASA CENTRAL (SANTIAGO)		3	5.949999999999999	56
6	I54S7C69J1V1	SEDE LA SERENA		54	5.6076923076923055	50
7	I54S8C69J1V1	SEDE VALPARAISO		54	5.566666666666666	85
8	I54S12C69J1V1	SEDE CHILLAN		54	5.583333333333333	40
9	I54S13C69J1V1	SEDE CONCEPCION - TALCA...		54	0	60
10	I54S21C69J1V1	SEDE APOQUINDO		54	5.426744186046509	50
11	I22S1C111J1V1	CASA CENTRAL (VINA DEL ...		22	5.716666666666666	80
12	I26S1C6J1V1	CASA CENTRAL (SANTIAGO)		26	5.546753246753248	80
13	I26S1C6J2V1	CASA CENTRAL (SANTIAGO)		26	0	36
14	I93S2C64J1V1	IT SEDE TALCAHUANO		93	5.578124999999999	80
15	I93S3C64J1V1	IT SEDE CAÑETE		93	5.742105263157895	40
16	I93S3C64J2V1	IT SEDE CAÑETE		93	5.699999999999999	40
17	I93S4C64J1V1	IT SEDE CHILLAN		93	5.847368421052633	90
18	I93S5C64J1V1	IT SEDE LOS ANGELES		93	5.846153846153845	56
19	I86S3C1J1V1	CAMPUS ORIENTE		86	6.353333333333333	40
20	I19S1C32SJ1V1	CAMPUS PROVIDENCIA		19	5.344897959183673	60
21	I19S4C32SJ1V1	CAMPUS SANTIAGO CENTRO		19	5.4703125	80
22	I93S1C2SJ1V1	CASA CENTRAL (CONCEPCI...		93	6.213888888888891	70
23	I9S2C74J1V1	SEDE LA SERENA		9	5.456666666666667	20
24	I9S8C74J1V1	SEDE ANTOFAGASTA		9	5.300000000000001	20
25	I9S9C74J1V1	SEDE CHILLAN		9	0	20
26	I84S2C322J1V2	CASA CENTRAL (SANTIAGO)		84	5.999999999999999	20

Activar Windows
Ve a Configuración para activar Windows.

POWER BI RELACIONAL

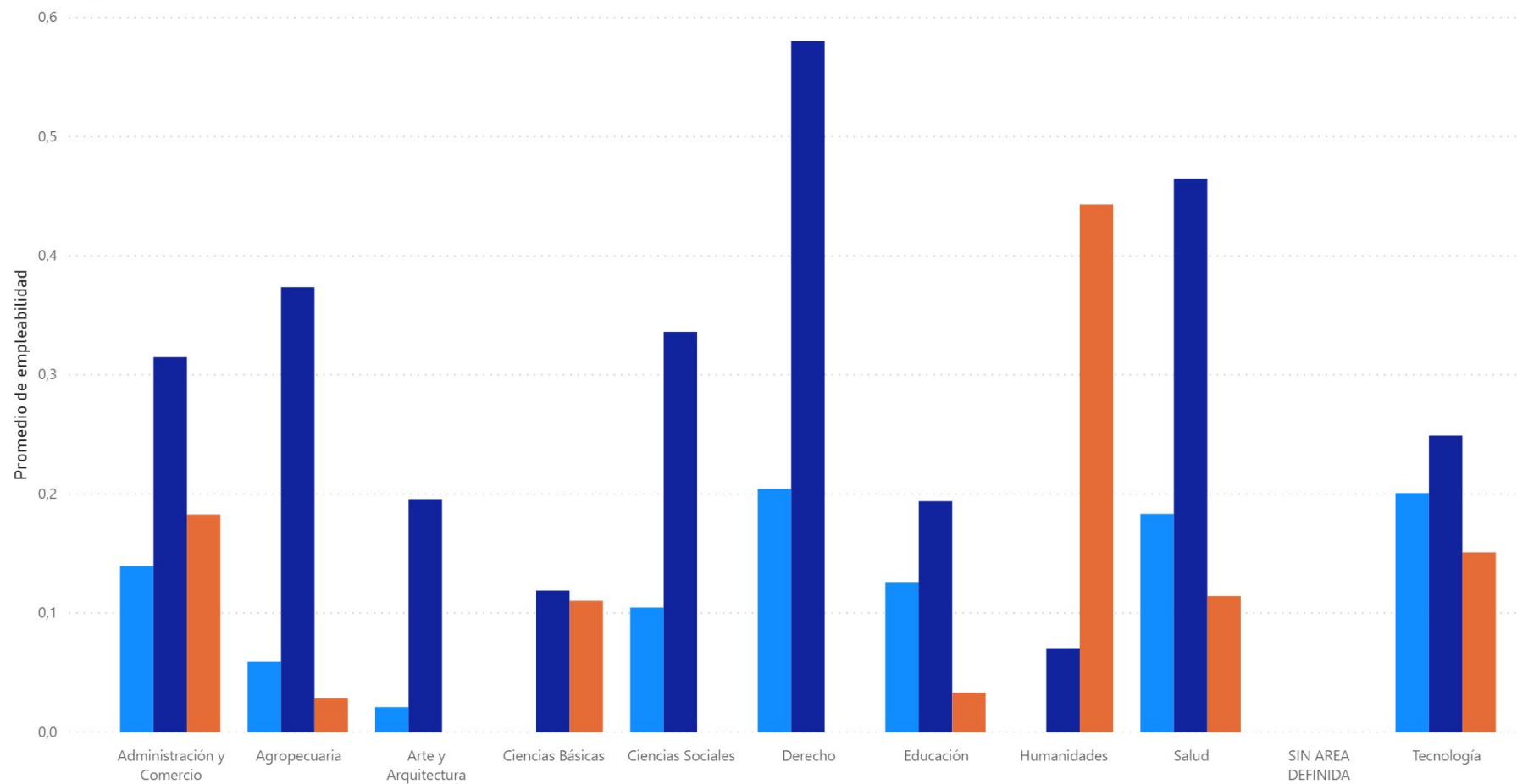


Promedio de arancel por nombrecarrera

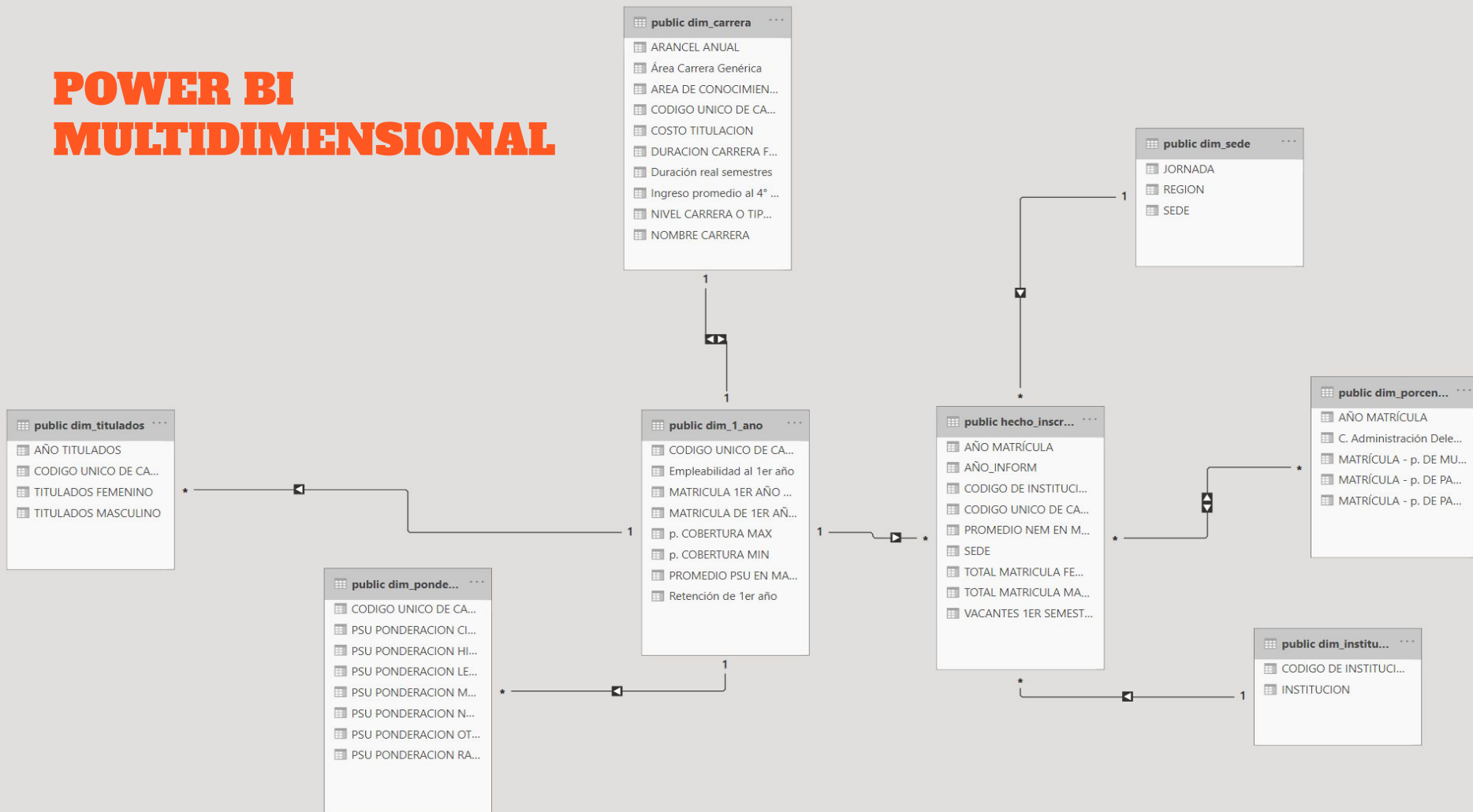


Promedio de empleabilidad por area y niveltipo

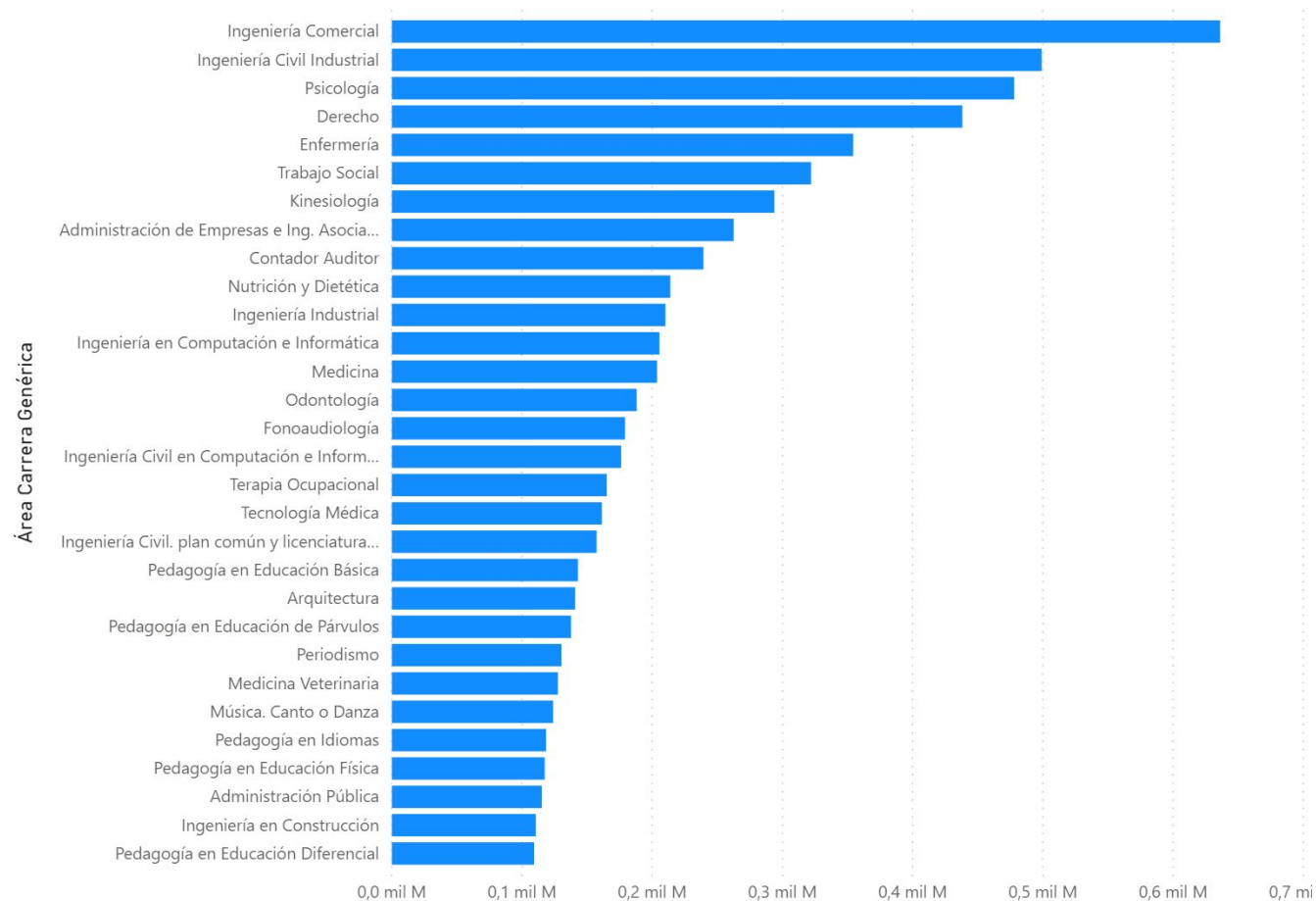
niveltipo ● - Profesional ● Técnica de nivel superior



POWER BI MULTIDIMENSIONAL



ARANCEL ANUAL por Area Carrera Genérica



COSTO TITULACION por AREA DE CONOCIMIENTO

