



Informe SIES Multidimensional Proyecto Analítico de Datos

Profesor

Mario Ortiz

Alumno

Johan Órdenes

Leonel Villagra

Fecha

30/11/2020

Índice

Introducción	3
Análisis de los datos	3
Limpieza de los datos	4
Comparación de datos	6
Diagrama SIES relacional	7
Gráficos SIES relacional	8
Diagrama Estrella SIES Multidimensional	10
Gráficos SIES Multidimensional	12
Base de Datos Multidimensional	17
Tiempo de Procesos	19

1. Introducción

Una base de datos multidimensional (MDB) es un tipo de base de datos que se ha optimizado para data warehouse y aplicaciones de procesamiento analítico en línea (OLAP). Las bases de datos multidimensionales se crean con frecuencia usando entradas de las bases de datos relacionales existentes. En esta investigación se nos pide traspasar desde una base de datos relacional a una multidimensional de nuestro modelo de caso “SIES”, decidimos utilizar el esquema estrella para la elaboración de nuestro trabajo, provocando un cambio en el antiguo diagrama relacional del SIES y con una nueva base de datos.

En esta investigación se nos pide hacer el proceso de ETL, es decir exportar, transformar y cargar los datos del archivo sies.xlsx. Esto lo hicimos realizando la lectura de los datos entregados, limpieza y eliminación de los datos inservibles, y por último el almacenamiento de los datos en una base de datos en Postgres.

2. Análisis de los datos

ARANCEL ANUAL	COSTO TITULACION	AÑO, DUR. A.	DURACION CARRERA FORMAL	NIVEL CARRERA TIPO DE CARRERA	AÑO MATRICULA	TOTAL MATRICULA FEMENINO	TOTAL MATRICULA MASCULINO	TOTAL MATRICULA	MATRICULA DE 1ER AÑO FEMENINO	MATRICULA DE 1ER AÑO MASCULINO	TOTAL MATRICULA DE 1ER AÑO	MATRICULA - % DE PARTICIPAR SUBVENCIONA DO	MATRICULA - % DE PARTICIPAR SUBVENCIONA PAGADO	C. Administración Delegada	AÑO TITULADOS	TITULADOS FEMENINO	TITULADOS MASCULINO	TOTAL TITULADOS	% DE COBERTURA POS EN MATRICULA DE 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%	2016	0	0	0	-
1.550.000	161.000	2017	5	Técnica de nivel s	2017	35	86	121	15	33	49	51.7%	48.3%	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%	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	2016	0	0	0	-
3.434.000	214.700	2017	8	Profesional	2017	42	12	54	35	11	46	15.4%	46.2%	38.5%	2016	0	0	0	60% <= X < 80%
2.214.000	0	2017	10	Profesional	2017	301	52	353	50	5	55	26.7%	73.3%	0.0%	2016	29	1	30	40% <= X < 60%
2.162.000	0	2017	10	Profesional	2017	262	50	312	70	17	87	25.4%	68.3%	3.0%	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%	2016	0	0	0	60% <= X < 80%
2.214.000	0	2017	10	Profesional	2017	244	33	277	0	0	0	56.6%	34.9%	4.1%	2016	58	6	64	60% <= X < 80%
2.382.000	0	2017	10	Profesional	2017	235	32	267	76	15	91	18.4%	63.9%	11.4%	2016	16	2	18	60% <= X < 80%
3.382.000	60.000	2017	10	Profesional	2017	138	21	159	49	5	54	13.9%	76.2%	6.0%	2016	38	6	44	70% <= X < 90%
3.160.000	0	2017	10	Profesional	2017	120	71	191	48	32	80	16.5%	59.3%	22.5%	2016	21	19	40	60% <= X < 80%
2.930.000	0	2017	10	Profesional	2017	0	0	0	0	0	0	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	66.6%	25.6%	0.0%	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%	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%	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%	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%	2016	12	0	12	-
1.853.000	0	2017	8	Profesional	2017	121	86	207	24	21	45	12.5%	35.5%	52.0%	2016	8	10	18	50% <= 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%	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%	5.7%	2016	4	4	8	-
2.496.000	100.000	2017	10	Profesional	2017	299	37	336	65	7	72	49.6%	41.8%	1.8%	2016	42	2	44	70% <= X < 90%
2.389.000	325.674	2017	10	Profesional	2017	66	12	78	24	7	31	33.3%	65.3%	1.3%	2016	13	0	13	-
2.680.000	325.674	2017	10	Profesional	2017	57	0	57	12	4	16	13.8%	37.5%	10.2%	2016	17	0	17	-
2.290.000	325.674	2017	10	Profesional	2017	0	0	0	0	0	0	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%	2016	2	4	6	-
2.498.000	325.674	2017	10	Profesional	2017	41	62	103	9	21	30	60.0%	40.0%	0.0%	2016	0	0	0	-
2.690.000	325.674	2017	10	Profesional	2017	45	44	89	16	21	37	63.5%	27.5%	10.0%	2016	2	2	4	-
2.610.000	325.674	2017	10	Profesional	2017	188	80	268	25	28	53	22.5%	67.4%	6.0%	2016	17	17	34	60% <= X < 80%
4.024.000	464.000	2017	8	Profesional	2017	223	37	260	60	7	67	25.4%	63.3%	0.4%	2016	30	6	36	60% <= X < 80%
2.990.000	181.000	2017	8	Profesional	2017	0	0	0	0	0	0	NULL	NULL	NULL	2016	0	0	0	-
3.880.000	214.000	2017	8	Profesional	2017	86	41	127	22	17	39	16.6%	72.1%	11.5%	2016	8	1	9	70% <= X < 90%
1.136.000	229.000	2017	5	Técnica de nivel s	2017	176	48	224	57	22	79	41.8%	56.0%	0.0%	2016	57	8	65	-
1.120.000	229.000	2017	5	Técnica de nivel s	2017	92	11	103	44	2	46	63.5%	11.0%	5.5%	2016	0	0	0	-
1.260.000	229.000	2017	5	Técnica de nivel s	2017	148	26	174	64	15	79	66.1%	36.9%	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%	33.7%	0.0%	2016	39	7	46	-
1.500.000	50.000	2017	5	Técnica de nivel s	2017	188	30	218	68	11	79	14.6%	76.1%	0.0%	2016	40	3	43	-
1.500.000	50.000	2017	5	Técnica de nivel s	2017	161	19	180	56	5	61	62.2%	47.4%	0.0%	2016	26	12	38	-
3.740.033	418.403	2017	10	Profesional	2017	163	160	323	32	37	69	29.2%	68.0%	0.9%	2016	41	32	73	70% <= X < 90%
3.248.400	373.456	2017	10	Profesional	2017	256	207	463	60	24	84	27.9%	59.3%	0.2%	2016	31	38	69	70% <= X < 90%
3.882.500	429.625	2017	10	Profesional	2017	258	203	461	40	35	75	24.4%	71.0%	1.3%	2016	35	34	69	70% <= X < 90%
4.263.000	463.004	2017	10	Profesional	2017	111	179	290	18	21	39	23.2%	76.8%	2.0%	2016	32	30	62	70% <= X < 90%
3.246.000	86.000	2017	10	Profesional	2017	103	64	167	21	25	46	52.0%	82.2%	0.0%	2016	22	15	37	70% <= X < 90%
3.728.977	60.000	2017	10	Profesional	2017	196	222	418	38	58	96	23.4%	66.5%	9.1%	2016	15	10	25	70% <= X < 90%
3.675.197	60.000	2017	10	Profesional	2017	209	21	230	26	24	50	12.2%	69.1%	18.5%	2016	0	0	0	-
1.402.000	107.010	2017	4	Técnica de nivel s	2017	68	24	92	19	7	26	12.8%	67.4%	2.3%	2016	15	5	20	-
3.018.355	60.000	2017	8	Profesional	2017	124	368	492	17	47	64	23.4%	66.7%	7.1%	2016	28	63	91	70% <= X < 90%
2.252.750	421.000	2017	10	Profesional	2017	114	185	299	27	29	56	22.3%	77.2%	0.0%	2016	28	26	54	70% <= X < 90%
2.280.360	421.000	2017	10	Profesional	2017	104	59	163	18	12	30	17.4%	46.7%	0.0%	2016	21	5	26	70% <= X < 90%
1.550.000	161.000	2017	5	Técnica de nivel s	2017	229	1	230	110	1	111	54.1%	35.9%	0.0%	2016	0	0	0	-

Tabla de datos “sies.xlsx”

Fuente: “Escuela de Ingeniería UCN”

Para la lectura de los datos decidimos marcar de color amarillo las columnas que necesitaban limpieza, marcar de color rojo las columnas que necesitaban eliminarse y las columnas blancas sin marcar, dado que no necesitaban ningún cambio.

Las celdas vacías y erróneas contenían los siguientes N/A, NULL, s/i, -, #N/D, #VALOR!, (vacío). Lo que decidimos hacer en este caso fue:

- Pasar datos tipo texto no conocidos como “nan”
- Pasar datos tipo número no conocidos como “nan”

Para la columna “% de cobertura psu último año” decidimos dividirla y crear 2 columnas a partir de ésta, dado que contenía 2 valores diferentes dentro de la misma columna, el porcentaje mínimo y el porcentaje máximo de cobertura del último año.

3. Limpieza de los datos

También añadimos funciones que nos sirven para la limpieza luego de leer los datos de cada columna, definiendo los valores NULL como ‘nan’ que son las siguientes:

- cleanPcobertura(): Limpia los datos de la columna “% de cobertura psu último año”.
 - Transforma los datos ‘-’ a NULL
 - Reemplaza el string ‘% <= X’ < con ‘,’
 - Reemplaza el string ‘=’ con ‘
- cleanText(): Limpia los datos de las columnas tipo texto.
 - Reemplaza el string ‘s/i’ por NULL
 - Reemplaza el string ‘nan’ por NULL

- `cleanPercentages()`: Limpia los datos de tipo decimal.
 - Reemplaza el string 'nan' por NULL
- `cleanDigit()`: Limpia los datos de las columnas que contienen dígitos enteros.
 - Reemplaza el string 's/i' por NULL
 - Reemplaza el string 'nan' por NULL
- `cleanDecimal()`: Limpia los datos de las columnas que contienen dígitos flotantes.
 - Reemplaza el string 's/i' por NULL
 - Reemplaza el string 'nan' por NULL
 - Reemplaza el string ' ' por NULL
 - Reemplaza el string '-' por NULL

También hubo columnas que no se limpiaron, esto fué porque los datos están 100% funcionales y sin problemas para un posterior análisis y agregación a la base de datos, las columnas son las siguientes:

- 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

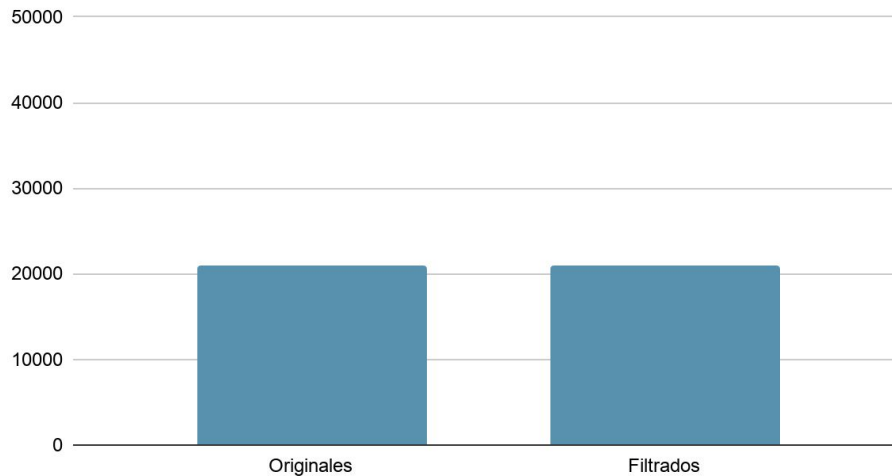
Además hubo columnas que eliminamos, dado que no eran representativas y no generaban un cambio al momento de agregarla a la base de datos, esto nos ayuda a que haya un ahorro de cantidad de datos ingresados.

- TIPO DE INSTITUCION: Fué eliminada debido a que todas las filas de esta columna poseían el mismo atributo “universidad”, en este caso estuvimos analizando datos de universidad, entonces no hubo datos de tipo “institututo” o “colegio” por ejemplo.
- AÑO_DURAC: Fué eliminada debido a que ya existía una columna con los mismos datos, solo que tenía otro nombre. En este caso, “Año duracion” contenía los mismos datos de “Año matrícula”.
- TOTAL MATRICULA, TOTAL MATRICULA 1ER AÑO, TOTAL TITULADOS: Estas 3 columnas no eran necesarias ya que por cada una de ellas existen 2, una para la suma de personas femeninas y la otra para la suma de personas masculinas, por ende el total se puede deducir de la suma entre estas 2 columnas.

4. Comparación de datos

Luego de hacer toda la limpieza de datos anterior, decidimos comparar los datos que teníamos al principio, con los datos nuevos dentro del mismo dataframe modificado y llegamos a la siguiente conclusión:

Datos Originales vs Filtrados



Cantidad de datos Originales vs Cantidad de datos filtrados

Fuente: Elaboración de Equipo

La cantidad de datos originales era de 20.944 columnas, mientras que luego de haber limpiado y filtrado todos los datos quedó la misma cantidad de filas que teníamos al principio 20.944. Esto significa que hubo un 0% de pérdida de información.

5. Diagrama SIES relacional

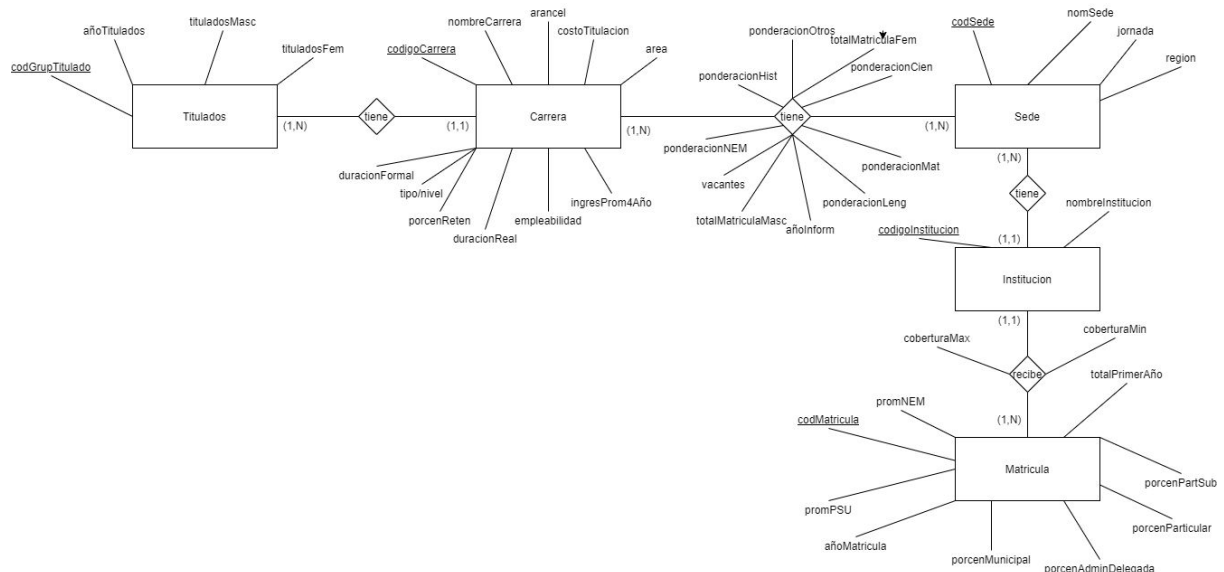


Diagrama Entidad-Relacionamiento

Fuente: Elaboración de Equipo

Como breve resumen de lo realizado en la entrega anterior: dividimos el archivo sies.xlsx en una cantidad total de 6 tablas, 5 explícitas (Titulados, Carrera, Sede, Institucion, Matricula) y 1 implícita (CarreraSede).

Con la elaboración de este trabajo pudimos encontrar diversos errores estructurales entre tablas y atributos designados, siendo corregidos en el nuevo modelo multidimensional.

6. Gráficos SIES relacional

arancel	area	nombrecarrera
7952520	Salud	ODONTOLOGIA
7947064	Salud	MEDICINA
7947064	Salud	ODONTOLOGIA
7925570	Salud	MEDICINA
7909160	Salud	MEDICINA
7909160	Salud	ODONTOLOGIA
7866790	Salud	ODONTOLOGIA
7702100	Salud	MEDICINA
7679486	Salud	ODONTOLOGIA
7457300	Salud	MEDICINA
7358392	Salud	MEDICINA
7283000	Salud	MEDICINA
7283000	Salud	ODONTOLOGIA
7214500	Salud	ODONTOLOGIA
7203000	Salud	MEDICINA
7034800	Salud	MEDICINA
7021000	Salud	ODONTOLOGIA
6985100	Salud	ODONTOLOGIA
6766300	Salud	MEDICINA
6690000	Salud	MEDICINA
6689990	Salud	ODONTOLOGIA
6631560	Salud	ODONTOLOGIA
6540000	Tecnología	INGENIERIA CIVIL DE INDUSTRIAS
6538610	Salud	MEDICINA
6418000	Salud	MEDICINA
6418000	Salud	ODONTOLOGIA
6270000	Salud	MEDICINA
6237900	Salud	ODONTOLOGIA
6234565	Administración y Comercio	INGENIERIA COMERCIAL

Ejemplo tabla modelo relacional

Fuente: Elaboración de Equipo

A través de esta Tabla designada logramos realizar un análisis utilizando la herramienta “Power Bi” que nos proyecta los gráficos que veremos a continuación

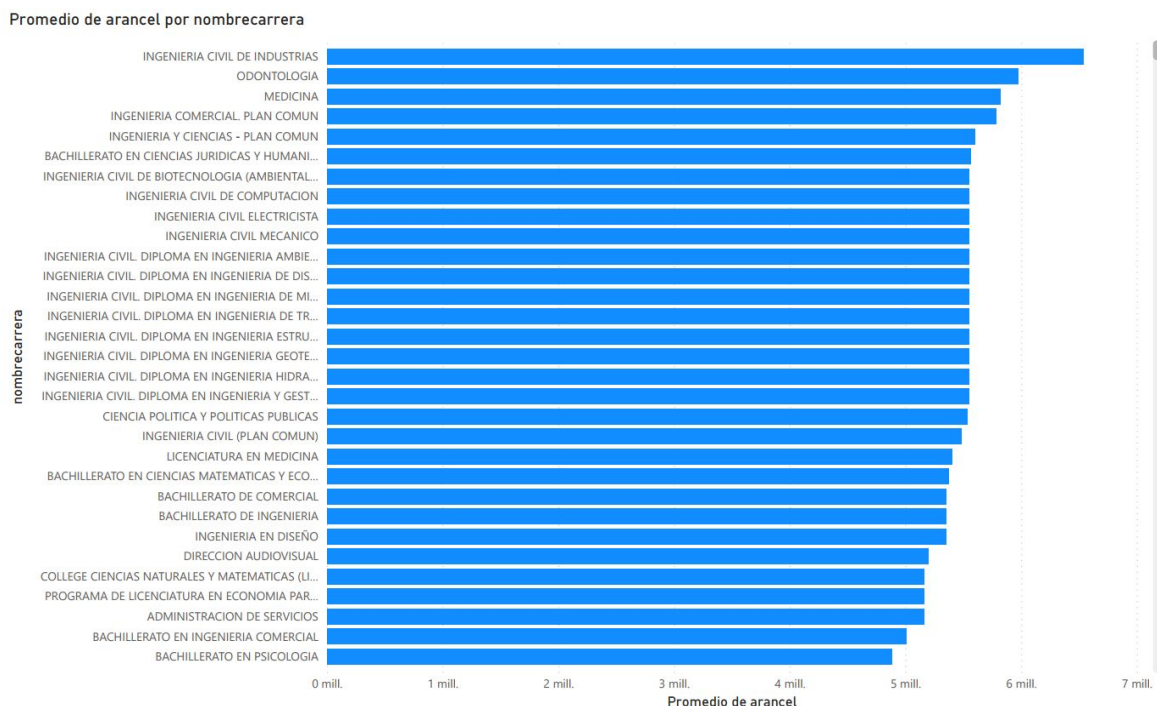


Gráfico de aranceles por carrera

Fuente: Elaboración de Equipo (Power BI)

En nuestro primer gráfico podemos analizar cada carrera con su respectivo promedio de aranceles, abarcando desde poco menos de 5 millones hasta casi los 7 millones según la carrera.

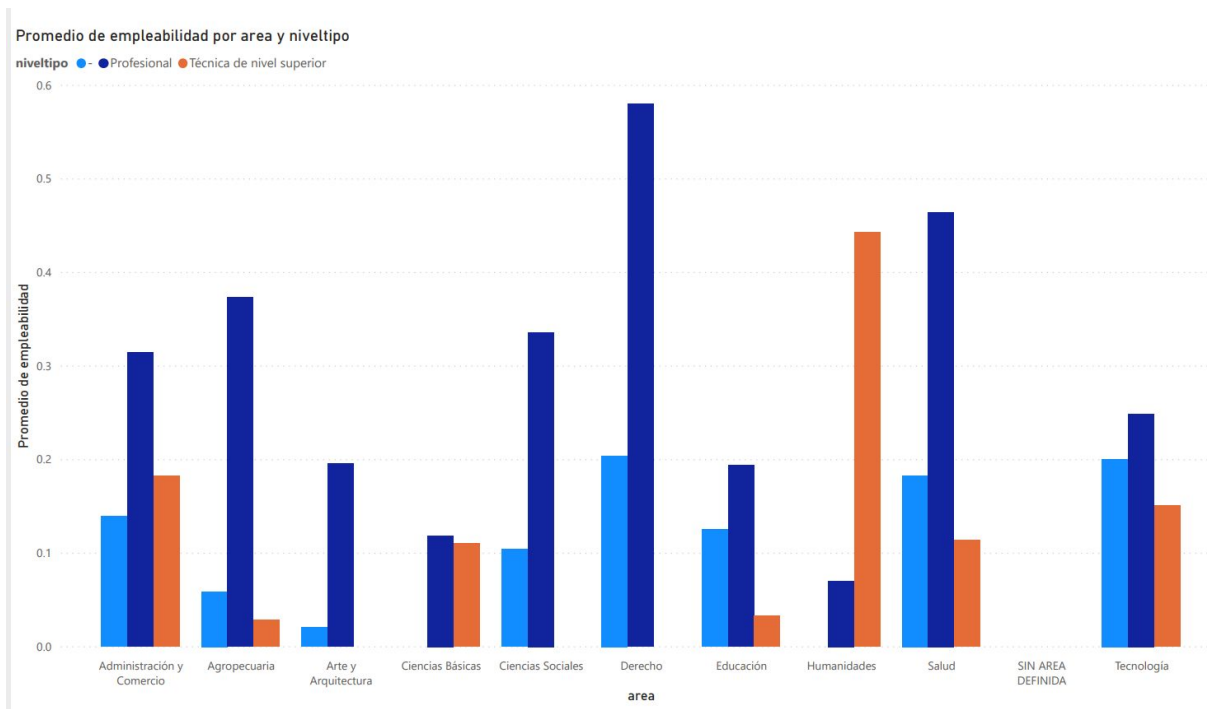


Gráfico de empleabilidad por carrera según el área y el tipo

Fuente: Elaboración de Equipo

Aquí podemos observar la cantidad de empleabilidad que hay por cada área estipulada entre las carreras.

7. Diagrama Estrella SIES Multidimensional

El esquema estrella está formado por una tabla de hechos y diversas tablas de dimensiones relacionadas a través de sus respectivas claves. Nosotros elegimos este esquema porque se caracteriza al ser: Simple de interpretar, posee mejores tiempos de respuesta, su diseño es fácil de mantener y actualizar, conteniendo las siguientes 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.

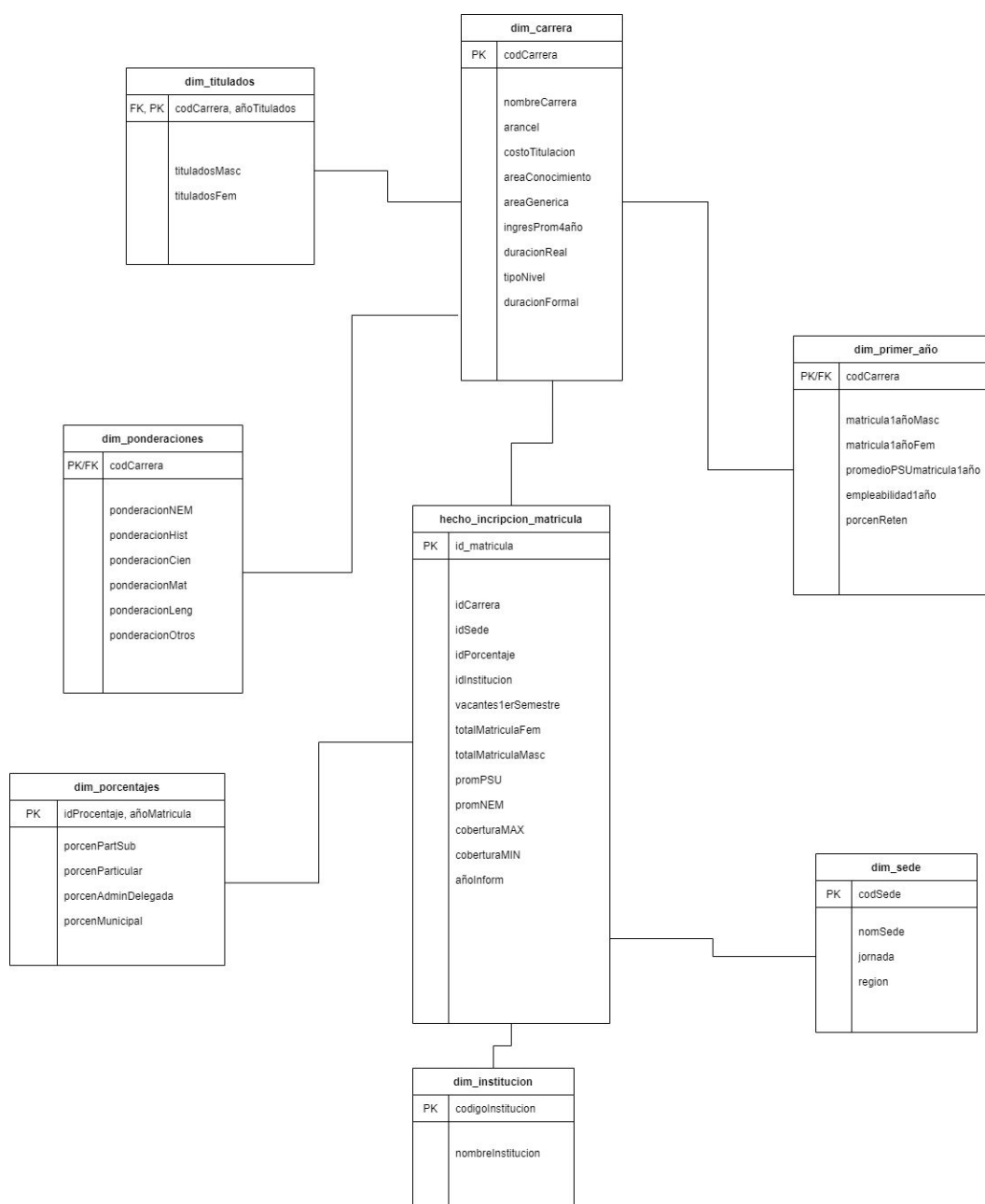


Diagrama Entidad-Relacionamiento multidimensional

Fuente: Elaboración de Equipo (Draw.io)

En nuestro caso nuestra tabla de hechos es “hecho_inscripcion_matricula” el cual deriva a las demás dimensiones expresadas en la imagen colocando como dimensión más importante la tabla “dim_carrera”.

8. Gráficos SIES Multidimensional

Ahora bien, como ya mostramos los gráficos anteriores, se nos pidió formular preguntas específicas y mostrarlas con filtros dentro de power BI.

¿Cuál es la carrera con arancel anual más alto?

ARANCEL ANUAL por Area Carrera Genérica

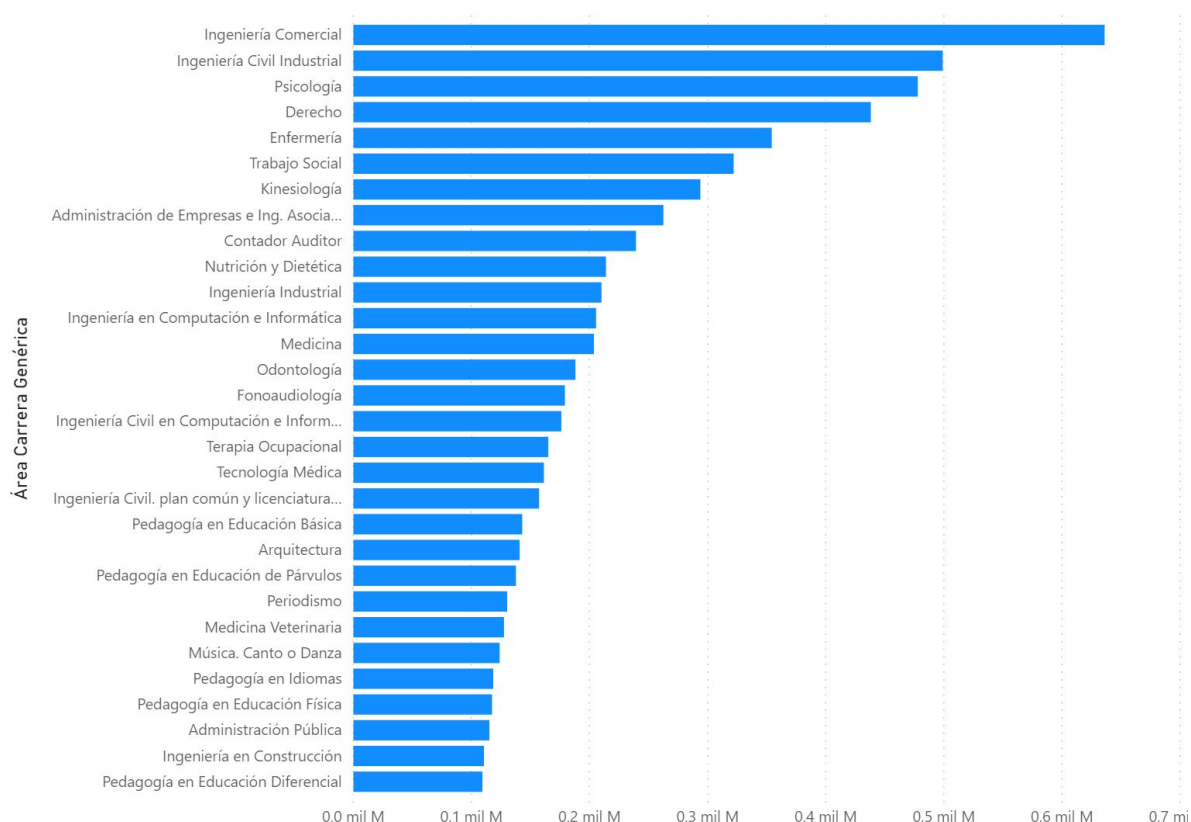


Gráfico aranceles anuales por carrera

Fuente: Elaboración de Equipo (Power BI)

En el gráfico anterior analizamos el arancel anual y el área asociada a las diversas carreras. Se puede apreciar que según el área genérica, las universidades cobran más en carreras de ingeniería, partiendo por ingeniería comercial, industrial y siguiendo por psicología.

¿Cuál es la carrera con el costo de titulación más alto?

COSTO TITULACION por AREA DE CONOCIMIENTO

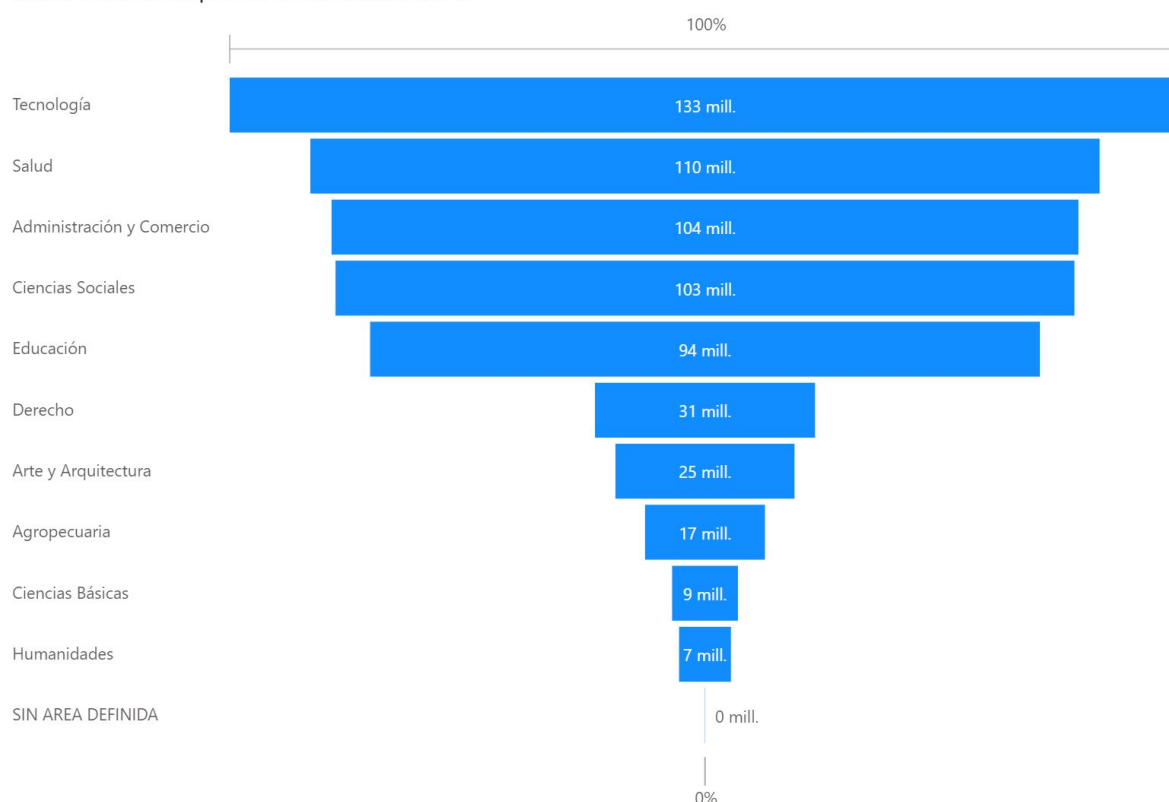


Gráfico ingresos por título

Fuente: Elaboración de Equipo (Power BI)

Aquí apreciamos la suma del costo titulación por área de conocimientos. El costo de titulación se ve más fuerte en el tema del área de la tecnología, salud, economía, sociedad y educación. Con esto podemos deducir que son las carreras que están mayoritariamente en auge y los 5 sectores más importantes del país.

¿Cuál es la cantidad de pago distribuido en las matrículas según el año?

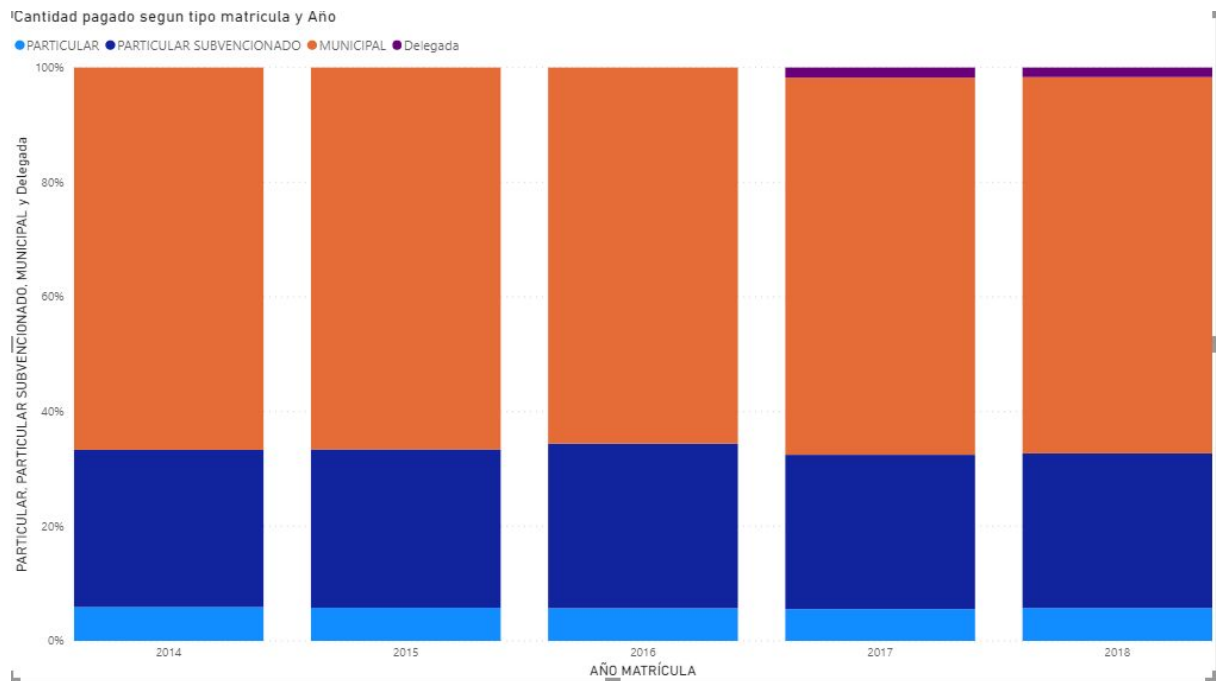


Gráfico cantidad de pagos según tipo de matrícula anual

Fuente: Elaboración de Equipo (Power BI)

Están los pagos particulares, particulares subvencionados, municipales, y el porcentaje de administración delegada de las matrículas. Se ve claramente que en los últimos 2 años 2017 y 2018 recién se aplicó el porcentaje que se va hacia la administración delegada. No varían mucho los datos con respecto a los años anteriores.

¿Cuál es la cantidad de vacantes de primer semestre en los últimos 5 años?

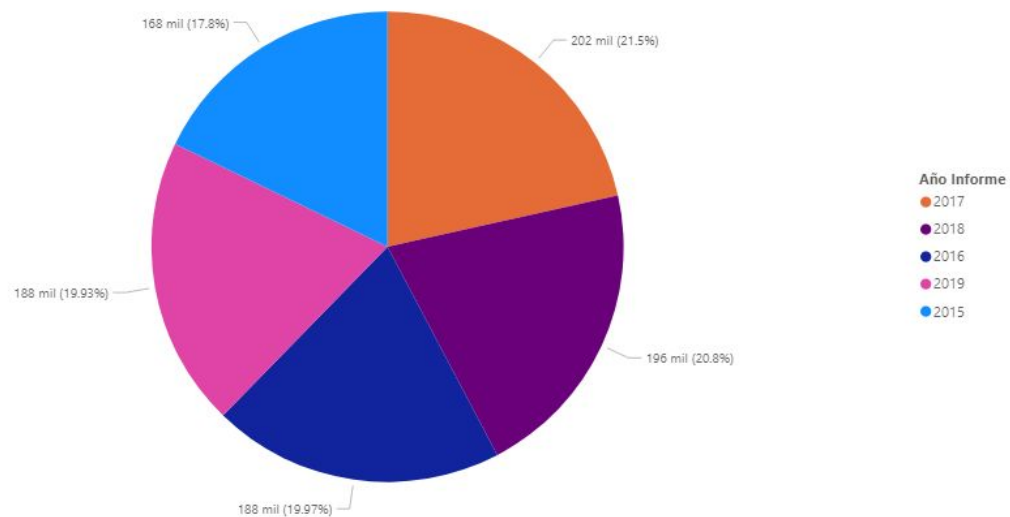


Gráfico de vacantes totales en el primer semestre por año

Fuente: Elaboración de Equipo (Power BI)

Como se puede apreciar en el gráfico, la cantidad de vacantes para alumnos en todas las universidades del país disponibles por año ha disminuido progresivamente, obteniendo 21.5% en 2017 y bajando a 17.8% en el último año. Estos datos están calculados con respecto a 5 años consecutivos.

¿Cuál es el total de personas matriculadas por sexo masculino y femenino según la región, sede, carrera e institución?

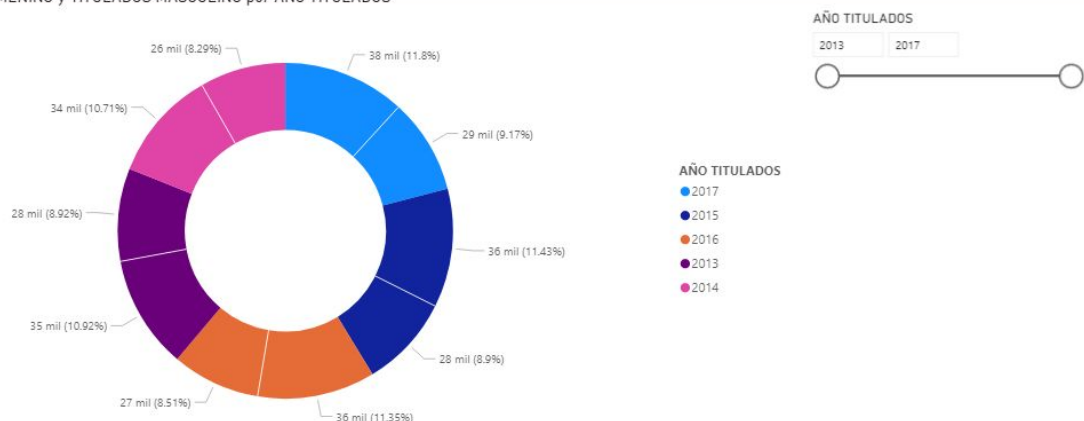


Gráfico de titulados por sexo y año

Fuente: Elaboración de Equipo (Power BI)

El total de matriculados en los años 2014 a 2017 predomina mayoritariamente el sexo masculino, aunque como vimos anteriormente, esto se ve disminuido a través de los años progresivamente.

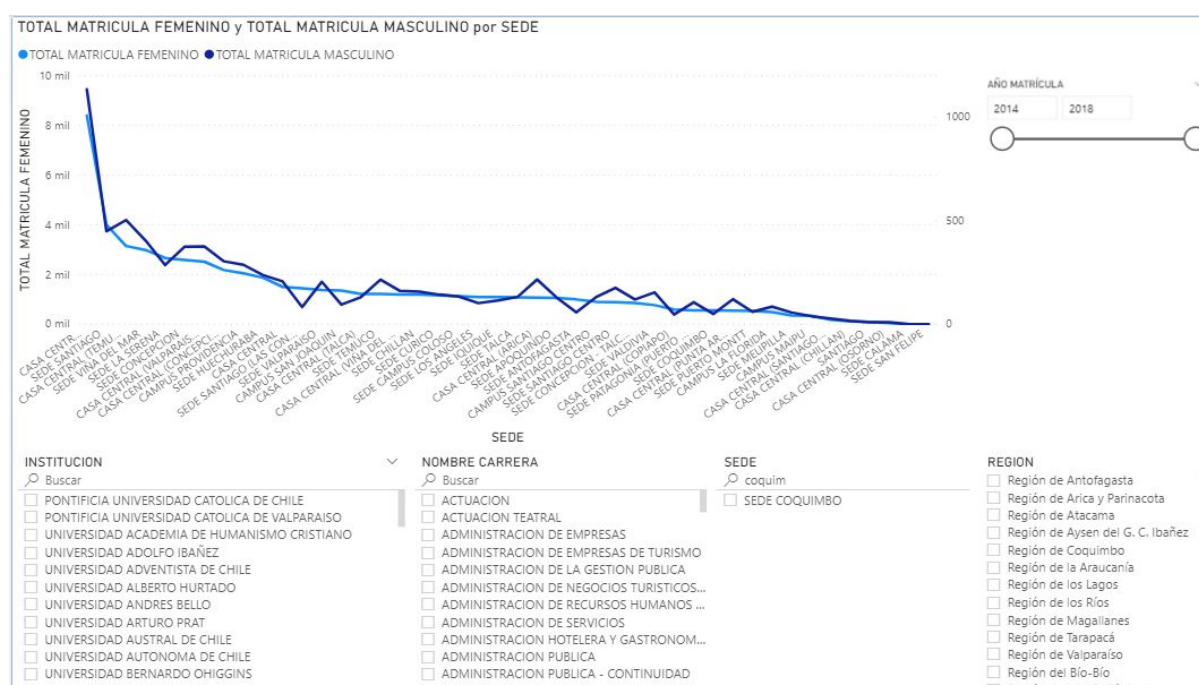


Gráfico de datos de universidades nacionales por sexo

Fuente: Elaboración de Equipo (Power BI)

El total de matrícula masculino y femenino por total podemos obtener todo tipo de valores, ya sea por institución, carrera, sede, región y año de matrícula, dándonos a conocer que la mayor cantidad de sedes y alumnos está en Santiago.

¿Cuántas son las vacantes que hay de carreras disponibles en el primer semestre por sede y región?



Gráfico de vacantes de universidades nacionales por sede
Fuente: Elaboración de Equipo (Power BI)

Las vacantes totales decidimos hacer este gráfico el cual es bastante visual para que se pueda apreciar las vacantes por sedes universitarias. Las que mayor vacantes poseen son las regiones de Santiago, Concepción y Viña del Mar.

9. Base de Datos Multidimensional

Insertamos screenshot de como queda finalmente la base de datos con el nuevo modelo multidimensional:

Query Editor Query History

```
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 FEMENINO
	text	text	bigint	double precision	bigint	bigint
1	I84S1C322JIV2	CASA CENTRAL (OSORNO)	84	5.48333333333333	45	45
2	I84S2C322JIV2	SEDE PUERTO MONTT	84	5.427659574468086	45	45
3	I84S6C322JIV2	SEDE CASTRO	84	5.370967741935483	35	35
4	I77S1C194J2V2	CASA CENTRAL (PUNTA A...	77	0	20	20
5	I3S1C126JIV2	CASA CENTRAL (SANTIA...	3	5.949999999999999	56	56
6	I54S7C69JIV1	SEDE LA SERENA	54	5.6076923076923055	50	50
7	I54S8C69JIV1	SEDE VALPARAISO	54	5.566666666666666	85	85
8	I54S12C69JIV1	SEDE CHILLAN	54	5.583333333333333	40	40
9	I54S13C69JIV1	SEDE CONCEPCION - TAL...	54	0	60	60
10	I54S21C69JIV1	SEDE APOQUINDO	54	5.426744186046509	50	50
11	I22S1C111JIV1	CASA CENTRAL (VIÑA DE...	22	5.716666666666666	80	80
12	I26S1C6JIV1	CASA CENTRAL (SANTIA...	26	5.546753246753248	80	80
13	I26S1C6J2V1	CASA CENTRAL (SANTIA...	26	0	36	0
14	I93S2C64JIV1	IT SEDE TALCAHUANO	93	5.578124999999999	80	80
15	I93S3C64JIV1	IT SEDE CAÑETE	93	5.742105263157895	40	40
16	I93S4C64J2V1	IT SEDE CAÑETE	93	5.699999999999999	90	90
17	I93S4C64JIV1	IT SEDE CHILLAN	93	5.847368421052633	40	40
18	I93S5C64JIV1	IT SEDE LOS ANGELES	93	5.846153846153845	56	56
19	I86S3C1JIV1	CAMPUS ORIENTE	86	6.353333333333333	40	40
20	I19S1C32SJIV1	CAMPUS PROVIDENCIA	19	5.344897959183673	60	60
21	I19S4C32SJIV1	CAMPUS SANTIAGO CEN...	19	5.4703125	80	80
22	I93S1C2SJIV1	CASA CENTRAL (CONCEP...	93	6.213888888888891	70	70
23	I9S2C74JIV1	SEDE LA SERENA	9	5.456666666666667	20	20
24	I9S8C74JIV1	SFHF ANTOFAGASTA	9	5.300000000000001	0	0

Tabla hechos inscripción matrícula
Fuente: Elaboración de Equipo (Postgres)

Query Editor Query History

```
1 SELECT * FROM dim_sede
```

Data Output Explain Messages Notifications

	SEDE	JORNADA	REGION
	text	text	text
1	CASA CENTRAL (OSORNO)	Diurno	Región de los Lagos
2	SEDE PUERTO MONTT	Diurno	Región de los Lagos
3	SEDE CASTRO	Diurno	Región de los Lagos
4	CASA CENTRAL (PUNTA ARENAS)	Vespertino	Región de Magallanes
5	CASA CENTRAL (SANTIAGO)	Diurno	Región Metropolitana
6	SEDE LA SERENA	Diurno	Región de Coquimbo
7	SEDE VALPARAISO	Diurno	Región de Valparaíso
8	SEDE CHILLAN	Diurno	Región del Bio-Bio
9	SEDE CONCEPCION - TALCAHUANO	Diurno	Región del Bio-Bio
10	SEDE APOQUINDO	Diurno	Región Metropolitana
11	CASA CENTRAL (VIÑA DEL MAR)	Diurno	Región de Valparaíso
12	IT SEDE TALCAHUANO	Diurno	Región del Bio-Bio
13	IT SEDE CAÑETE	Diurno	Región del Bio-Bio
14	IT SEDE CHILLAN	Diurno	Región del Bio-Bio
15	IT SEDE LOS ANGELES	Diurno	Región del Bio-Bio
16	CAMPUS ORIENTE	Diurno	Región Metropolitana
17	CAMPUS PROVIDENCIA	Diurno	Región Metropolitana
18	CAMPUS SANTIAGO CENTRO	Diurno	Región Metropolitana
19	CASA CENTRAL (CONCEPCION)	Diurno	Región del Bio-Bio
20	SEDE ANTOFAGASTA	Diurno	Región de Antofagasta
21	CASA CENTRAL (SANTIAGO - PROVIDENCIA)	Diurno	Región Metropolitana
22	CASA CENTRAL (IQUIQUE)	Diurno	Región de los Ríos
23	CASA CENTRAL (IQUIQUE)	Diurno	Región de Tarapacá
24	SEDE CALAMA	Vespertino	Región de Antofagasta

Tabla dimensión sedes
Fuente: Elaboración de Equipo (Postgres)

10. Tiempo de Procesos

Una vez iniciado el proceso de importación de archivos a la base de datos se consiguieron los siguientes resultados según cada proceso enlistado a continuación:

```
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
```

Tiempo de compilación

Fuente: Elaboración de Equipo

El código final se compiló en aproximadamente 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 completo se compiló en 39.80104732513428 segundos

El algoritmo más costoso es de orden $O(n)$, en donde n es la cantidad de ciclos. Estos algoritmos que obtuvieron dicho orden, son los encargados de filtrar los datos el cual es bastante eficiente considerando las funciones por defecto de Pandas que se ejecuta columna por columna por separado.