DOCUMENTACIÓN TÉCNICA

2023

DATALYST

Tabla de contenido

1. Introducción al proyecto 2
2. Objetivo principal 2
3. Fuentes de datos 2
   a. Detalle de las Fuentes de datos 2
4. Procesamiento de datos (ETL) 6
   a. Indicadores/Métricas 7
   b. Modelo de datos 9
   c. Procesos 10
5. Aplicación 10
   a. Visión General 10
   b. Selectores y Filtros 11
6. Otros 12
   a. Seguridad y Acceso 12
   b. Comentarios adicionales 12

1. Introducción al proyecto

Este es un Proyecto en el cual una institución educativa nos contacta para que realicemos diferentes tableros sobre la información de sus alumnos, correctores, materias, calificaciones, exámenes, tiempos de publicación, etc.

Este permite explorar y comprender los datos de manera intuitiva, facilitando la toma de decisiones basada en información precisa.

Primero nos centramos en la integración y transformación de datos, después, se diseñará el modelo de datos, a través de la confección de diferentes tablas. Más tarde nos adentraremos en el diseño y la creación de los gráficos, para finalmente terminar con los detalles y la validación de Datos.

2. Objetivo principal

El objetivo de este proyecto no es más que brindar la información a los/as directivos de la institución educativa, para que, a través de nuestra aplicación y nuestros tableros, puedan tomar las mejores decisiones que conlleven a mejorar sus procesos, eficiencia, identificar cuales son sus materias o exámenes mas duros, cual es el rendimiento promedio de los estudiantes, todo esto con el fin de mejorar su calidad educativa y como institución.

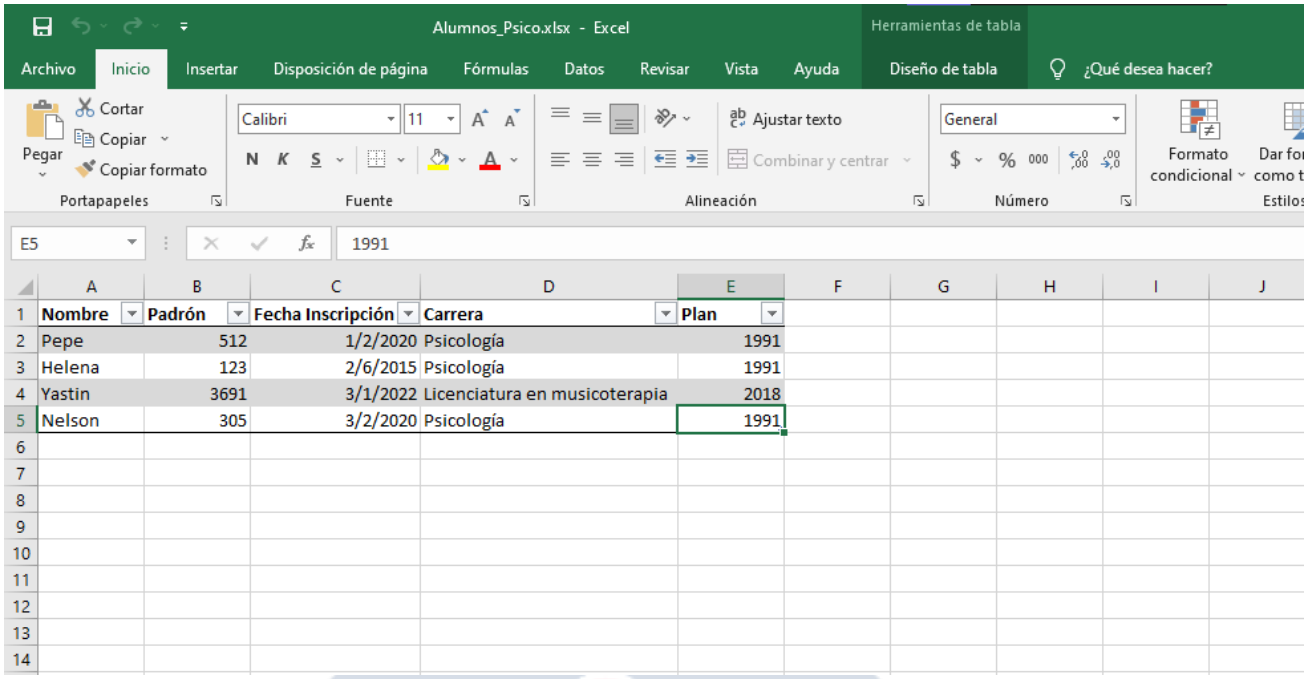
3. Fuentes de datos

Los datos están almacenados en diferentes locaciones. Pese a ser todos de el mismo tipo de Archivo (.xlsx), se encuentran divididos en muchas “tablas”. Por lo que tendremos que analizar cuáles deben ser llevadas a nuestro modelo, con cuales campos y con que tipos de formato.

a. Detalle de las Fuentes de datos

Existen 6 Archivos diferentes de Fuentes de Datos, los cuales son:

Alumnos_Psico.xlsx

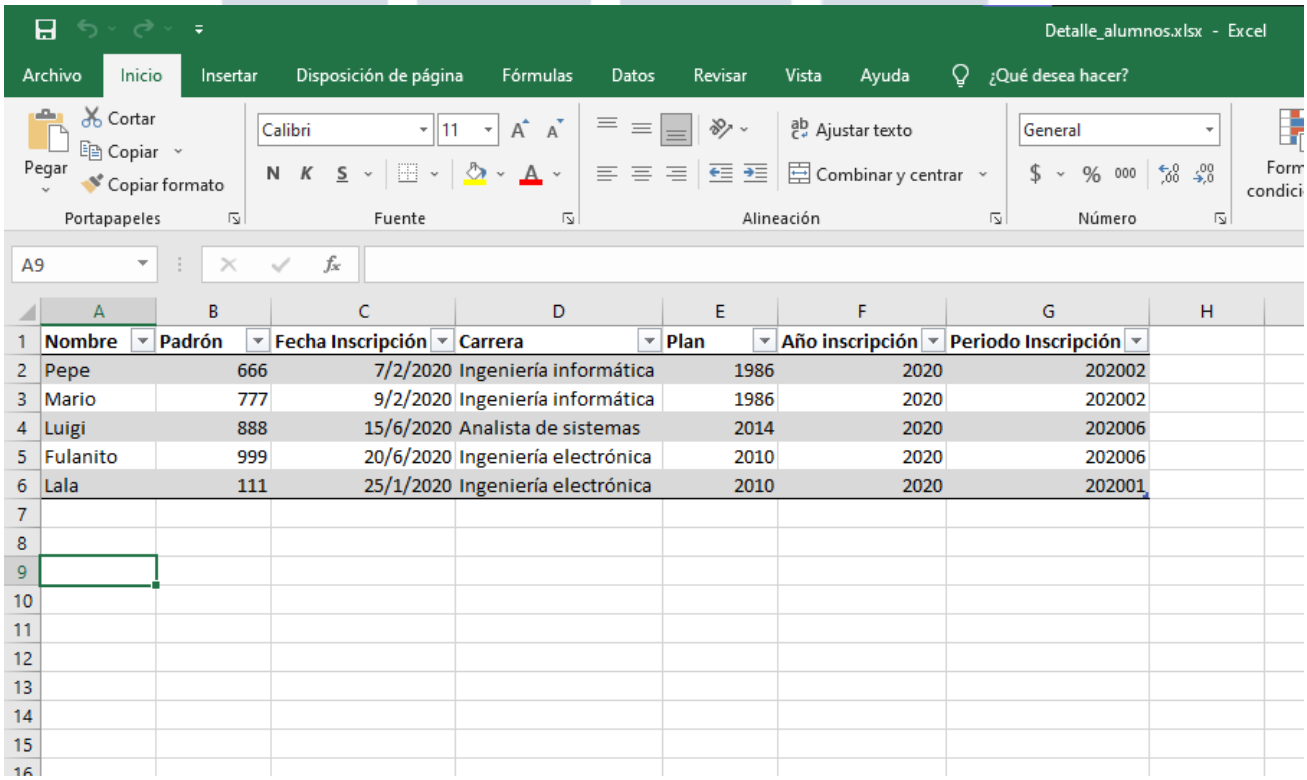


Nombre	Padrón	Fecha Inscripción	Carrera	Plan
Pepe	512	1/2/2020	Psicología	1991
Helena	123	2/6/2015	Psicología	1991
Yastin	3691	3/1/2022	Licenciatura en musicoterapia	2018
Nelson	305	3/2/2020	Psicología	1991

Esta fuente de datos, contiene información sobre los alumnos que asisten a las carreras de Psicología o Afines, contiene campos dimensionales tales como Nombre, Padrón, Fecha Inscripción, Carrera y Plan.

El campo que se relacionará con las otras tablas es el de Padrón.

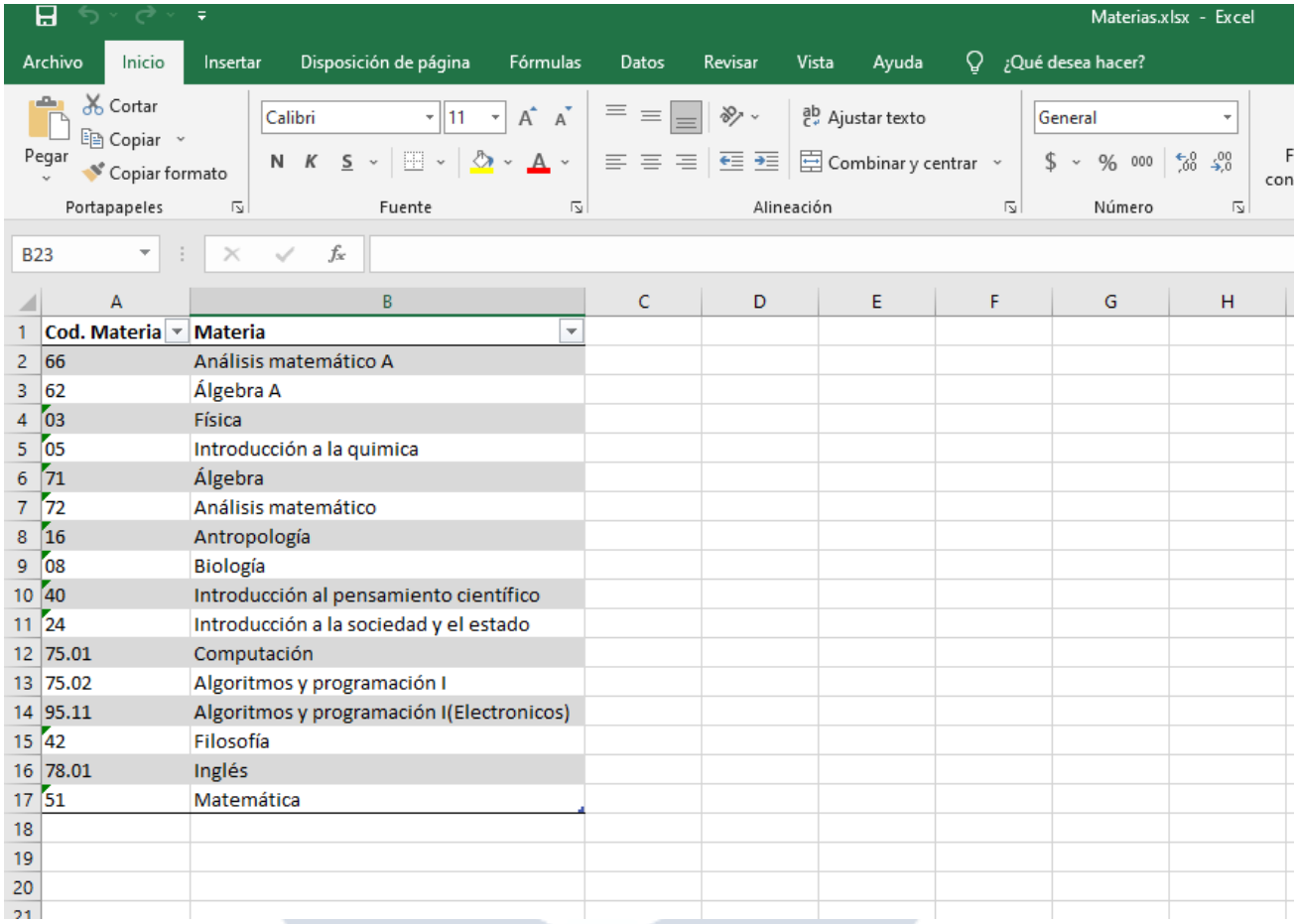
Detalle_alumnos.xls



Nombre	Padrón	Fecha Inscripción	Carrera	Plan	Año inscripción	Periodo Inscripción
Pepe	666	7/2/2020	Ingeniería informática	1986	2020	202002
Mario	777	9/2/2020	Ingeniería informática	1986	2020	202002
Luigi	888	15/6/2020	Analista de sistemas	2014	2020	202006
Fulanito	999	20/6/2020	Ingeniería electrónica	2010	2020	202006
Lala	111	25/1/2020	Ingeniería electrónica	2010	2020	202001

Esta tabla es muy similar a la anterior, solo contiene campos dimensionales, de los cuáles, comparte la Mayoría con Alumnos_Psico.xlsx

Materias.xlsx:



Cod. Materia	Materia
66	Análisis matemático A
62	Álgebra A
03	Física
05	Introducción a la quimica
71	Álgebra
72	Análisis matemático
16	Antropología
08	Biología
40	Introducción al pensamiento científico
24	Introducción a la sociedad y el estado
75.01	Computación
75.02	Algoritmos y programación I
95.11	Algoritmos y programación I(Electronicos)
42	Filosofía
78.01	Inglés
51	Matemática

Esta tabla solamente tiene 2 campos y sirve para relacionar los códigos de materia con el nombre de esta.

Hay que tener en cuenta que hay algunos números que están guardados como texto (seguramente fue un error 😉), por lo que tendremos que procesar o integrar estos datos para después poder utilizarlos.

Notas_Alumnos.xlsx

Código Mat	Plan_Carrera	Fecha Final	Condición	Hora Inicio	Hora Fin	Hora Entrega	Padrón_A	Alumno	Corrector	Fecha Publicación	Nota	Llamado
66	1986	1/6/2020	Regular	09:00:00	12:00:00	10:01:00	666	Pepe	Acero	1/6/2020	2	1
66	1986	1/6/2020	Regular	09:00:00	12:00:00	10:59:00	777	Mario	Maulhart	4/6/2020	1	1
66	2010	8/6/2020	Regular	09:00:00	12:00:00	11:35:00	111	Lala	Sirne	15/6/2020	4	2
03	2010	11/2/2021	Regular	13:00:00	17:00:00	16:20:00	999	Fulanito	Fontana	15/2/2021	2	1
03	2010	18/2/2021	Regular	13:00:00	17:00:00	15:59:00	999	Fulanito	Conde	25/2/2021	4	2
75.02	1986	21/6/2020	Regular	10:00:00	14:00:00	12:40:00	666	Pepe	Essaya	23/6/2021	8	3
75.01	2014	9/2/2021	Libre	07:00:00	11:00:00	10:22:00	888	Luigi	Flores	10/2/2021	2	1
03	2010	18/2/2021	Regular	13:00:00	17:00:00	13:01:00	111	Lala	Acosta	20/2/2021	1	2
75.02	1986	14/6/2020	Libre	10:00:00	14:00:00	13:55:00	777	Mario	Azcurra	16/6/2020	4	2
62	2014	7/12/2020	Regular	09:00:00	12:00:00	11:55:00	888	Luigi	Grynberg	8/12/2020	7	1
95.11	2010	12/12/2020	Libre	14:00:00	16:00:00	15:29:00	999	Fulanito	Essaya	12/12/2020	2	2
66	1986	8/6/2020	Regular	09:00:00	12:00:00	11:22:00	666	Pepe	Unger	12/6/2020	4	2
03	2014	25/2/2021	Regular	13:00:00	17:00:00	16:40:00	888	Luigi	Fontana	29/2/2021	8	3
62	1986	14/12/2020	Regular	09:00:00	12:00:00	11:35:00	777	Mario	Vargas	14/12/2020	5	2
95.11	2010	5/6/2020	Regular	14:00:00	17:00:00	16:35:00	111	Lala	Azcurra	10/6/2020	9	1
03	1986	7/3/2021	Regular	13:00:00	17:00:00	16:22:00	666	Pepe	Fontana	9/3/2021	4	5
62	2010	21/12/2020	Regular	09:00:00	12:00:00	11:17:00	111	Lala	Vargas	22/12/2020	2	3
03	1986	2/3/2021	Regular	13:00:00	17:00:00	15:00:00	777	Mario	Conde	7/3/2021	1	4
66	2010	15/12/2020	Regular	09:00:00	12:00:00	11:45:00	999	Fulanito	Acero	15/12/2020	4	2
66	2014	22/12/2020	Regular	09:00:00	12:00:00	11:00:00	888	Luigi	Unger	23/12/2020	1	3

Esta es una tabla de hechos, en la cual se registran todos los Finales que rindió cada alumno (puede ser más de uno por materia), su nota, condición, hora, fecha, entre otros. Campo de Codigo Mat con un formato erróneo.

Notas_Alumnos_Psico.xlsx

Código M	Plan_Carr	Fecha Fin	Condición	Hora Inici	Hora Fin	Hora Entr	Padrón_Alumno	Corrector	Fecha_Pu	Sede	Nota	Certificad	Llamado
42	1991	29/8/2017	LIBRE	09:00:00	11:00:00	09:21:00	0000000000000123	Dodorex	29/8/2017	The Center	10	Si	1
42	1991	1/6/2021	REGULAR	09:00:00	11:00:00	-	0000000000000512	Gutiérrez	4/6/2021	Martínez	AUS	-	1
51	1991	8/12/2020	LIBRE	09:00:00	11:00:00	12:02:00	0000000000000305	Martínez	15/6/2020	Martínez	4	No	2
51	2018	10/2/2022	LIBRE	09:00:00	11:00:00	11:14:00	0000000000003691	Ayana	10/2/2022	Las Heras	5	Si	2
42	1991	11/2/2021	REGULAR	09:00:00	11:00:00	-	0000000000000512	Gutiérrez	15/2/2021	Ciudad universitaria	AUS	-	1
42	2010	18/2/2021	REGULAR	09:00:00	11:00:00	-	0000000000000512	Gutiérrez	25/2/2021	Ciudad universitaria	AUS	-	2
42	1986	21/6/2020	REGULAR	09:00:00	11:00:00	10:40:00	0000000000000305	Gutiérrez	23/6/2021	Ciudad universitaria	4	No	3
40	2014	9/2/2021	REGULAR	07:00:00	09:00:00	08:22:00	0000000000000305	Bianchi	10/2/2021	Ciudad universitaria	4	Si	1
24	2010	18/2/2021	REGULAR	07:00:00	08:30:00	08:17:00	0000000000000305	Fonsi	20/2/2021	Ciudad universitaria	8	Si	2
42	1986	14/6/2020	REGULAR	09:00:00	11:00:00	13:55:00	0000000000000512	Campi	16/6/2020	Ciudad universitaria	1	No	2

Esta tabla también es de hechos, almacena los exámenes finales de Psicología y , además de que comparte gran cantidad de campos con la tabla anterior, sus registros se refieren a lo mismo, por lo que pueden ser concatenadas.

Campos de código materia y Padrón Alumno con formato incorrecto.

Notas_Cursada.xlsx

Alumno	Padrón	Código Mat	Txt_Nota	Nota Cursada	Nro_Parcial	Fecha Publi
Pepe, Le Pew	666	66	Seis	6	1	viernes, 3 de abril de 2020
Pepe, Le Pew	666	66	Siete	7	2	martes, 5 de mayo de 2020
Pepe, Le Pew	666	75.02	Siete	7	1	domingo, 4 de abril de 2021
Pepe, Le Pew	666	75.02	Cinco	5	2	miércoles, 12 de mayo de 2021
Pepe, Le Pew	666	3	Cuatro	4	1	miércoles, 12 de agosto de 2020
Pepe, Le Pew	666	3	Ocho	8	2	sábado, 17 de octubre de 2020
Mario	777	66	Nueve	9	1	viernes, 3 de abril de 2020
Mario	777	66	Cinco	5	2	domingo, 10 de mayo de 2020
Mario	777	75.02	Ocho	8	1	sábado, 4 de abril de 2020
Mario	777	75.02	Siete	7	2	lunes, 27 de abril de 2020
Mario	777	62	Diez	10	1	viernes, 7 de agosto de 2020
Mario	777	62	Cuatro	4	2	lunes, 12 de octubre de 2020
Mario	777	3	Cinco	5	1	lunes, 7 de septiembre de 2020
Mario	777	3	Seis	6	2	domingo, 15 de noviembre de 2020
Luigi	888	75.01	Seis	6	1	jueves, 13 de agosto de 2020
Luigi	888	75.01	Ocho	8	2	jueves, 17 de septiembre de 2020
Luigi	888	62	Seis	6	1	sábado, 12 de septiembre de 2020
Luigi	888	62	Nueve	9	2	martes, 17 de noviembre de 2020
Luigi	888	3	Seis	6	1	44125
Luigi	888	3	Cuatro	4	2	domingo, 15 de noviembre de 2020
Luigi	888	66	Ocho	8	1	jueves, 1 de octubre de 2020
Luigi	888	66	Diez	10	2	martes, 3 de noviembre de 2020
Billy	4001	16	Cuatro	4	1	lunes, 17 de mayo de 2021
Billy	4001	16	Cinco	5	2	domingo, 20 de junio de 2021
Fulanito	999	3	Siete	7	2	domingo, 15 de noviembre de 2020
Fulanito	999	3	Seis	6	1	domingo, 13 de septiembre de 2020
Fulanito	999	95.11	Nueve	9	1	viernes, 7 de agosto de 2020
Fulanito	999	95.11	Nueve	9	2	domingo, 27 de septiembre de 2020

Esta es otra tabla de eventos, pero a diferencia de las dos anteriores, esta se refiere a otra información.

En esta se guardan los registros de los parciales que ocurrieron durante la cursada de la materia, por lo que bajo mi punto de vista, no sería conveniente concatenarla con las otras tablas(notas alumnos y notas alumnos psico). Para esto se utilizará una link table.

Campos innecesarios, como txt nota, fechas con otro formato y algunos alumnos que si rindieron parciales pero no finales.

4. Procesamiento de datos (ETL)

Notas_cursada Notas_Alumnos, notas_alumnos_psico:
Diferentes funciones que utilicé para la carga de datos:

```
NUM(
REPLACE(
LTRIM(
REPLACE(Padrón_Alumno,'0',' ')),' ','0' ))      AS      PADRON_ALUMNO,
```

En algunos casos, el num por si solo no funcionaba, ya que por ejemplo el campo padrón alumno Superaba los 14 dígitos que qlik permite para un número, por lo cual, devolvía nulo

El num solo no anda, por que supera el máximo de dígitos de qlik (14) por lo cual devuelve nulo

El replace, reemplaza en el campo que se le asigne, cuando encuentre la primer condición, la cambia por la segunda. En el primer uso reemplaza todos los 0 por espacios.

El Ltrim borra todos los espacios a la izquierda

El Segundo Replace, reemplaza los espacios (ceros dentro del número) por ceros, para que por ejemplo

Un 101 no se interprete como un 11.

Desconozco si esta es la mejor solución, tampoco tengo el conocimiento sobre cuan eficientes son estas funciones, pero al ver el muy mal formato que traían, fue la única manera que encontré para poder utilizarlos

```
TIME([Hora Fin])        AS    HORA_FIN,
```

Interpreta este campo como un tiempo.

```
DATE([Fecha Publicación])   AS     FECHA_PUBLICACION,
```

Interpreta este campo como una fecha.

Notas_Alumnos:

En este archivo, había registros vacíos y Alumnos que estuvieron, ausentes, los cuales no se iban a tener en cuenta en este proyecto, por esa razón fueron eliminados a tarves de estas líneas.

```
filters(
Remove(Row, RowCnd(CellValue, 12, StrCnd(contain, 'AUS'))),
Remove(Row, Pos(Top, 9)),
….
```

a. Indicadores/Métricas

Algunos son:

NOTAS_FINAL, provienen desde Notas_Alumnos y Notas_Alumno_psico y como su nombre lo indica, es el valor que los alumnos obtuvieron en la resolución de su examen final.

NOTAS, proviene de Notas_cursada y es el valor que los alumnos obtuvieron en la resolución de los parciales.

TIEMPO EJECUCIÓN: se obtiene de la resta entre los valores de Hora Entrega y Hora inicio, nos indica cuanto se tardo esa persona en realizar el examen. Se utiliza para el punto 5

V_FUERA_TIEMPO: Se obtiene si la resta de Hora Fin y Hora inicio es menor que la resta entre Hora entrega y Hora inicio. Actúa como un Flag.

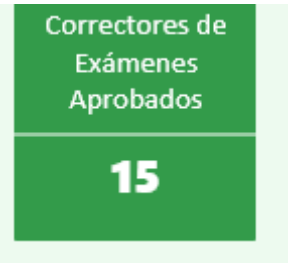
NOTA_P: Se refiere a el conjunto de notas de NOTAS_FINAL y NOTAS (parciales y finales), se utiliza para el punto 19 en el cual se pide un promedio de estas.

INTENTOS: Se refiere a la cantidad de veces que un alumno rindió un final en una materia hasta aprobarla.

AVG(NOTA_P)
Es el promedio entre las 3 notas de una materia por alumno (2 parciales y 1 final)

Aprobados /(Aprobados+Desaprobados)
Es el porcentaje de alumnos aprobados por materia

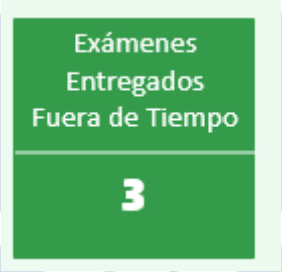
Indicadores:



=SUM(CONTADOR_CORRECTORES_APROBADOS)

Este dato se obtuvo del script, en donde se creo una flag sobre los correctores que habían corregido exámenes que estuviesen aprobados. Luego se los conto (sin repetirlos) y se los almaceno en una tabla llamada CONTADOR_CORRECTORES

Es la cantidad de Correctores diferentes que dieron un examen por aprobado



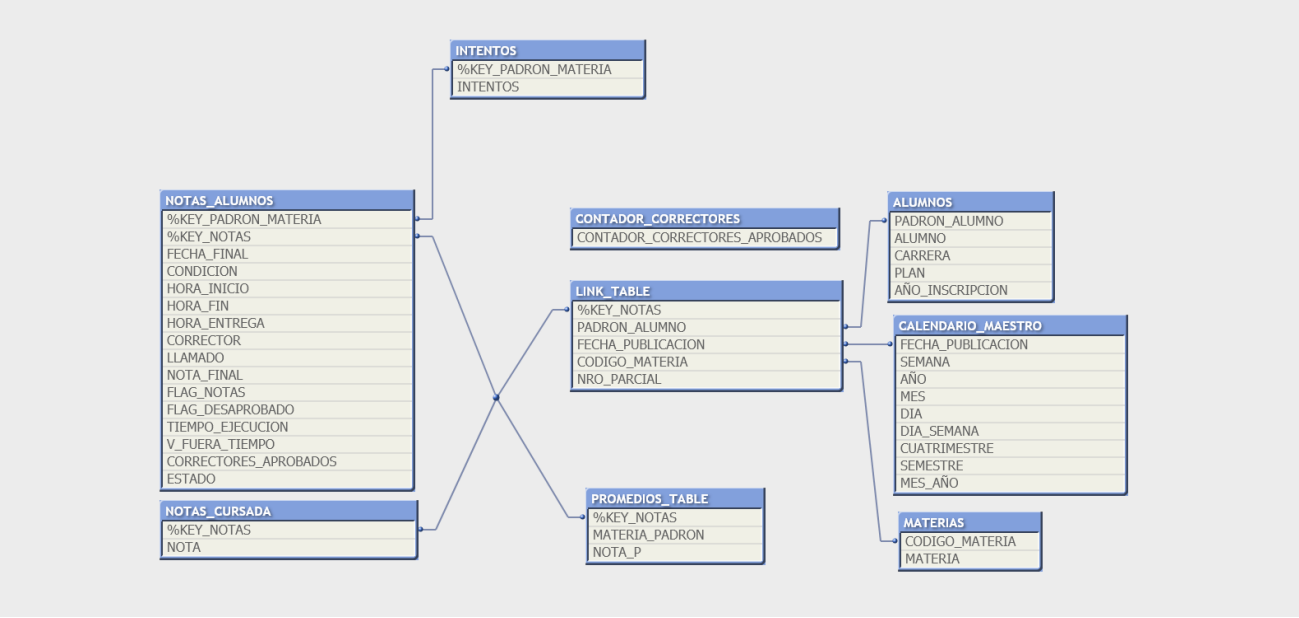
=SUM(V_FUERA_TIEMPO)

Es la cantidad de exámenes que tomaron más de el tiempo establecido. Viene de la resta de Hora Fin y Hora inicio, menor que la resta entre Hora entrega y Hora inicio.

Es el promedio de las notas de los exámenes finales por año, además, muestra información, como la cantidad de los mismos.

b. Modelo de datos

Para describir mi Modelo, primero lo mostraré.



Principalmente tenemos la tabla Notas_Alumnos, la cual se concatenó con Notas_Alumnos_psico, que almacena las notas de los finales de estas dos fuentes.

Luego está la tabla Notas_Cursada que contiene las notas de los finales.

Para poder conectar estas 2 tablas (necesario en el punto 19) usé una Link table, que se relaciona con las otras tablas a través de una key.

Luego tenemos otras tablas tales como Alumnos y Materias que nos sirven para obtener mas dimensiones sobre nuestras tablas de hechos.

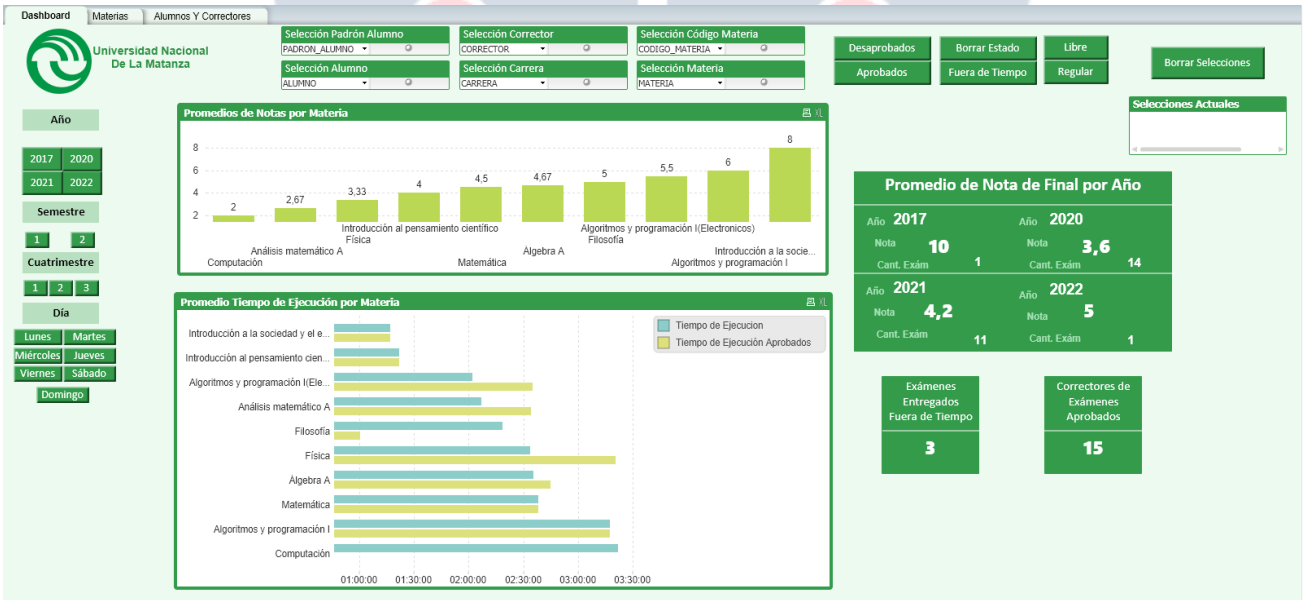
Tambien tenemos un Calendario Maestro creado a partir de la fecha publicación, como se nos fue indicado y otras tablas que son necesarias para resolver otros puntos tales como intentos (punto 7), Promedios_Table (punto 19) y Contador_correctores (punto 9)

c. Procesos

5. Aplicación

a. Visión General

Primeramente nos encontramos con 3 Hojas, en el primero visualizamos el Dashboard compuesto por 2 gráficos y 3 KPI, uno de ellos principal, y 2 secundarios. El objetivo es brindar una información rápida y un "resumen" del archivo.



En la segundo Hoja, Tenemos las tablas de Promedios por Materia , Por alumno y Materia

Y la tabla de aprobados de Materias, las cuales nos dan una visión general de los datos

Promedios por Materia

MATERIA	ALUMNO	FECHA_FINAL	NOTA FINAL	TIEMPO EJECUCION
Álgebra A	Lala	21/12/2020	2	02:17:00
	Luigi	7/12/2020	7	02:55:00
	Mario	14/12/2020	5	02:35:00
	Total		**4,67**	**02:35:40**
Algoritmos y programación I	Mario	14/6/2020	4	03:55:00
	Pepe	21/6/2020	8	02:40:00
	Total		**6**	**03:17:30**
Algoritmos y programación I (Electronicos)	Fulanito	12/12/2020	2	01:29:00
	Lala	5/6/2020	9	02:35:00
	Total		**5,5**	**02:02:00**
Análisis matemático A	Fulanito	15/12/2020	4	02:45:00
	Lala	8/6/2020	4	02:35:00
	Luigi	22/12/2020	1	02:00:00
	Mario	1/6/2020	1	01:59:00
	Pepe	1/6/2020	2	01:01:00
		8/6/2020	4	02:22:00
	Total		**2,67**	**02:07:00**
Computación	Luigi	9/2/2021	2	03:22:00
	Total		**2**	**03:22:00**
Filosofía			5	02:18:40
Física	Fulanito	11/2/2021	2	03:20:00
		18/2/2021	4	02:59:00
	Lala	18/2/2021	1	00:01:00
	Luigi	25/2/2021	8	03:40:00
	Mario	2/3/2021	1	02:00:00
	Pepe	7/3/2021	4	03:22:00
	Total		**3,33**	**02:33:40**
Introducción a la sociedad y el estado	Nelson	18/2/2021	8	01:17:00
	Total		**8**	**01:17:00**
Introducción al pensamiento cientifico	Nelson	9/2/2021	4	01:22:00
	Total		**4**	**01:22:00**

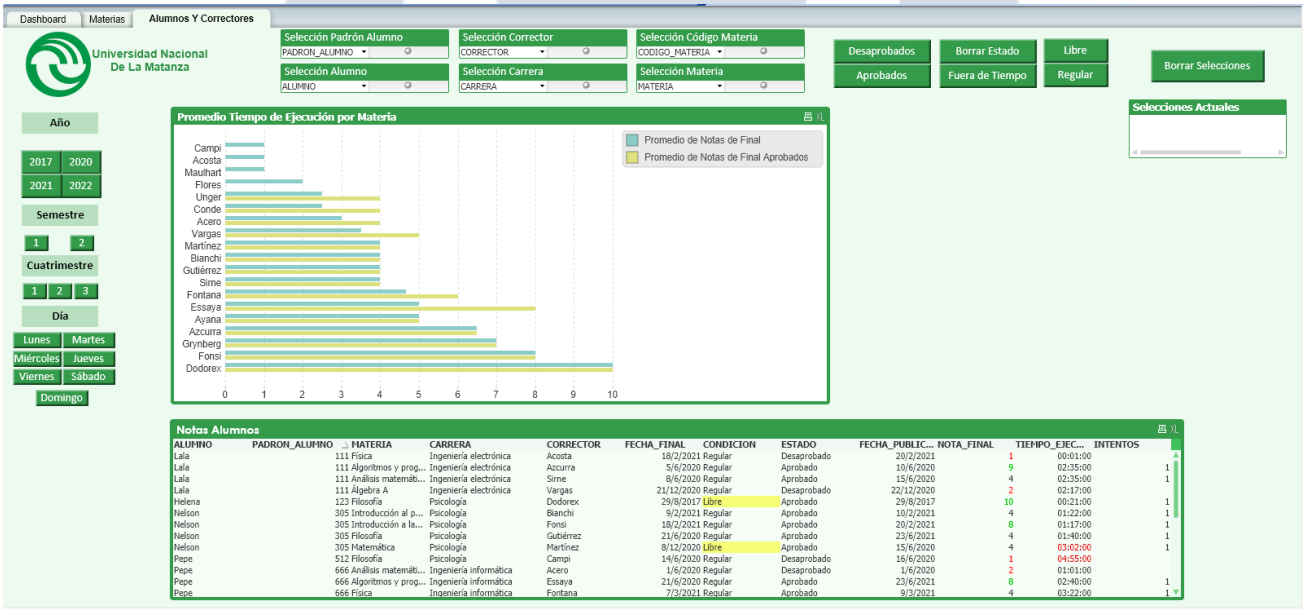
Promedios Alumno Y Materia

ALUMNO	MATERIA	NRO_PARCIAL	NOTA
Fulanito	Análisis matemático A	1	4
		2	4
		3	4
		Total	**4**
	Física	1	6
		2	7
		3	4
		Total	**5,67**
Helena	Filosofía	1	10
		2	10
		3	10
		Total	**10**
Lala	Algoritmos y programación I (Electronicos)	1	5
		2	7
		3	9
		Total	**7**
	Análisis matemático A	1	7
		2	8
		3	4
		Total	**6,33**
Luigi	Álgebra A	1	6
		2	9
		3	7
		Total	**7,33**

Materias

MATERIA	Aprobados	Desaprobados	Porcentaje
Álgebra A	2	1	67%
Algoritmos y prog...	2	0	100%
Algoritmos y prog...	1	1	50%
Análisis matemáti...	3	3	50%
Computación	0	1	0%
Filosofía	2	1	67%
Física	3	3	50%
Introducción a la...	1	0	100%
Introducción al p...	1	0	100%
Matemática	2	0	100%

En el tercero encontramos data mucho mas especifica, con muchos registros, tabla de notas alumnos y gráfico sobre los correctores

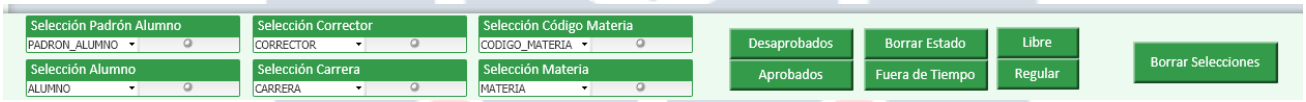


b. Selectores y Filtros

Los selectores y filtros están separados en 2 secciones

Los de Fechas

Y los de selecciones



6. Otros

a. Seguridad y Acceso

b. Comentarios adicionales

Correcciones realizadas:

Ejercicios:

6) Creación de botón para eliminar solamente la selección de Aprobado / Desaprobado. Asigne 2 variables a los respectivos botones para que cambien de color cuando estos sean activados. También asigne que cuando se pulsen los botones de Borrar estado o Borrar Selecciones, regresen a su color original.

7) Renombre el campo PADRON_MATERIA a %KEY_PADRON_MATERIA para darle un formato correcto y que no hayan confusiones.

9) Eliminé la sentencia Count Distinct en el front, y lo reemplace con una tabla auxiliar, por lo que ahora en el front solamente hay un Sum, lo cual es más eficiente.

11) Eliminé los Count del front, y los reemplacé por Sum de flags en el script.

Diseño:

Fijar los objetos en la hoja y cambiar algunos nombres de los botones.

Modelo:

Cambiar las keys por %key y renombrar la key Padrón_Materia

Solucionar el problema de la carga estática a través de la sentencia WHERE