



Manual Técnico

Indicadores Académicos

A continuación se detalla el sistema de análisis de datos de indicadores académicos desarrollado, con el fin de que este sirva para que la facultad de ingeniería pueda analizar la información relacionada a las estadísticas de las pruebas de admisión que realizan año con año los candidatos que desean estudiar alguna carrera de Ingeniería.

Paquete SSIS

Carga de Archivos de Excel a Modelo Relacional (ETL)

Este paquete consta de tareas que permite extraer los datos de dos fuentes de archivos de excel y la carga a una base de datos relacional. El paquete consta de los siguiente procesos:

- La tarea “Clear”, se utiliza para limpiar las tablas del modelo que no forman parte de catálogos. De forma que tablas como municipios, departamentos o países funcionen como un lugar al que únicamente se añaden nuevos valores y evitar así la pérdida de los identificadores de los mismos.
- Luego existe una tarea para cada una de las siguientes tablas:
 - País
 - Departamento
 - Municipios
 - Colegio
 - Diversificado

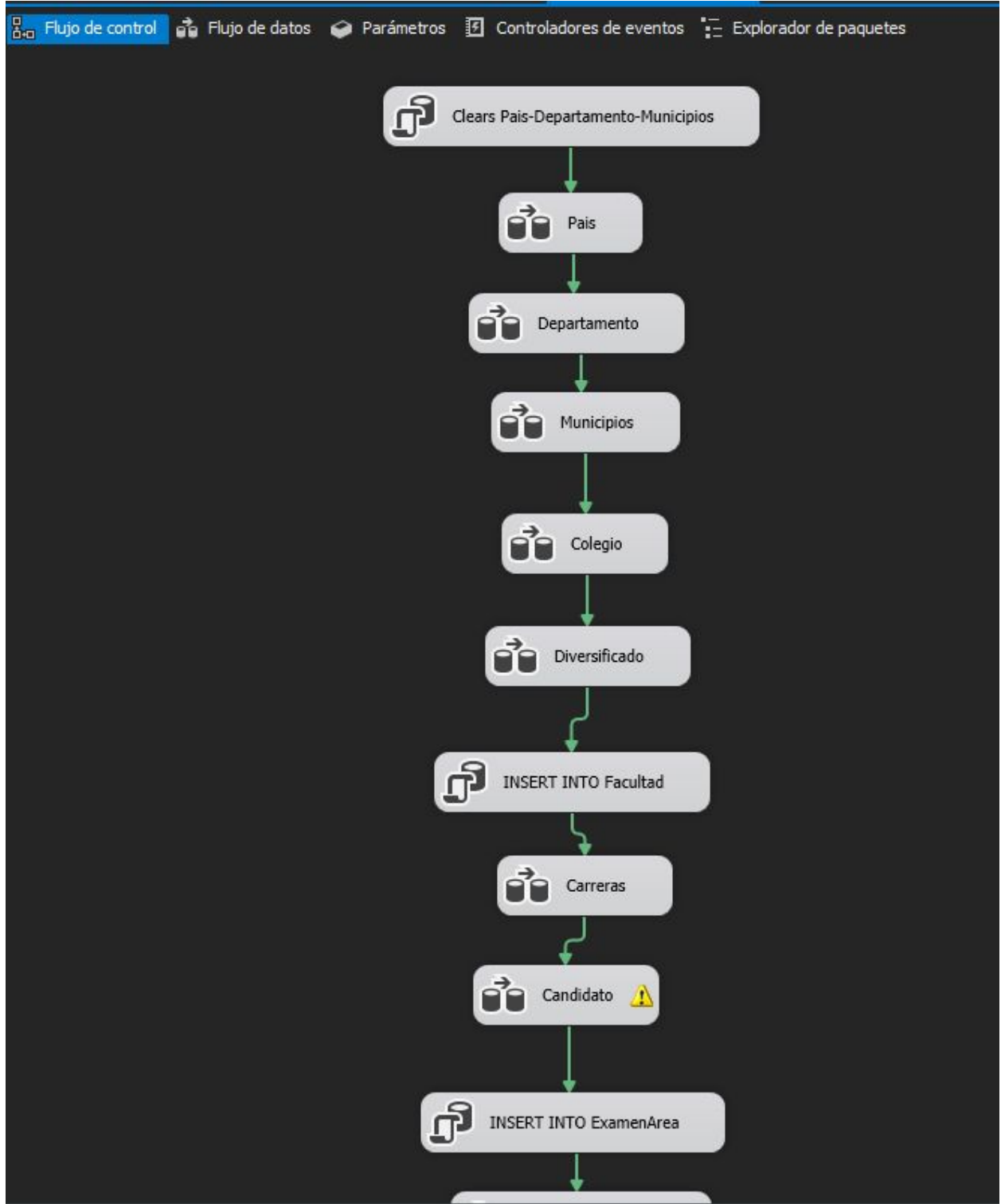
Estas tablas conforman los catálogos de la base de datos. Existen relaciones entre las tablas de País, Departamento y Municipios debido a que un país se conforma de varios departamentos y un departamento de varios municipios. El ETL se encarga de hacer este emparejamiento basado en la columna de “Departamento” del archivo de excel.

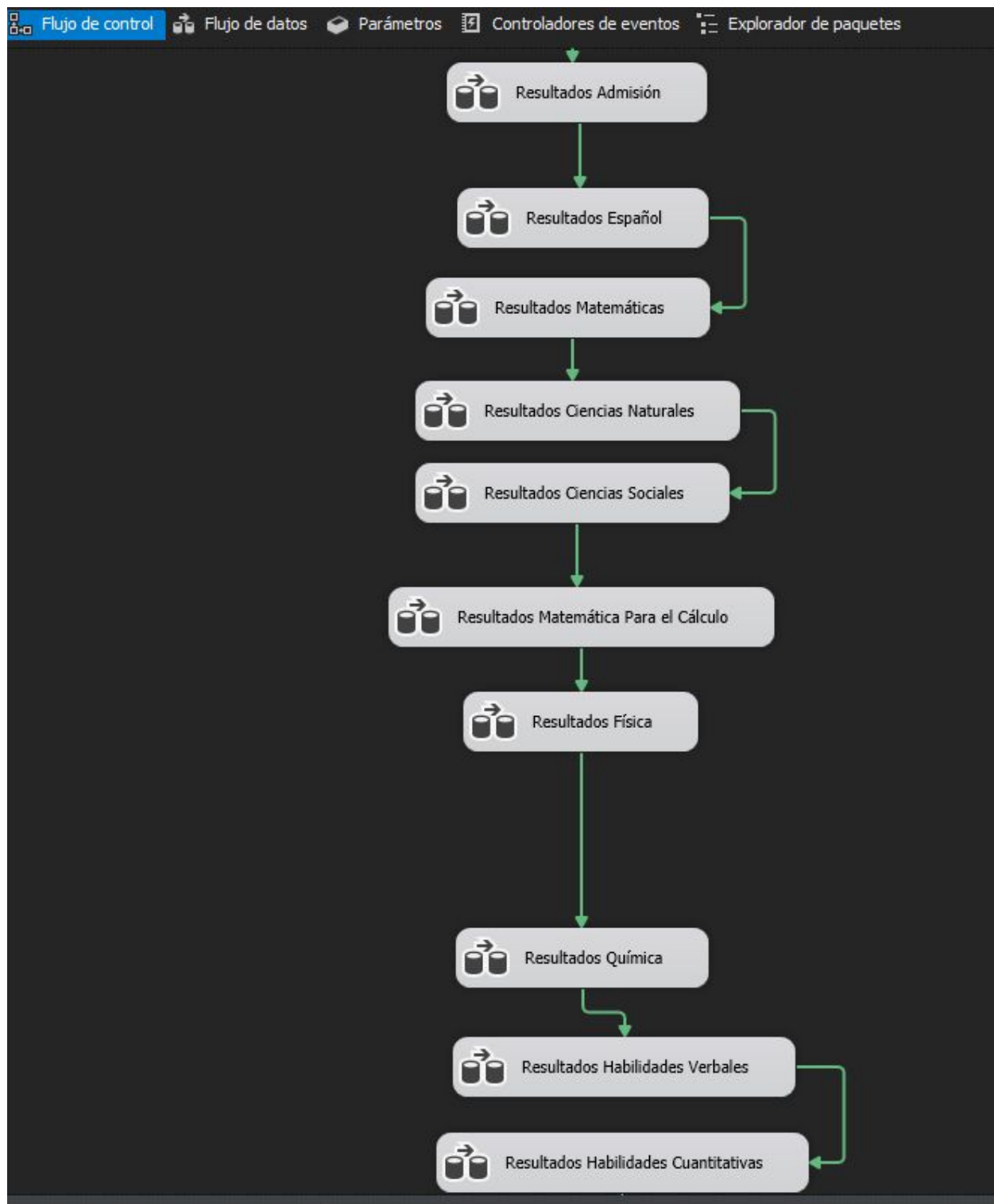
Los procesos de carga de Colegios y Diversificados únicamente cargan los datos de las columnas “Colegio” y “Diversificado” del archivo de resultados de excel, sin realizar algún tipo de asociación dentro de la base de datos.

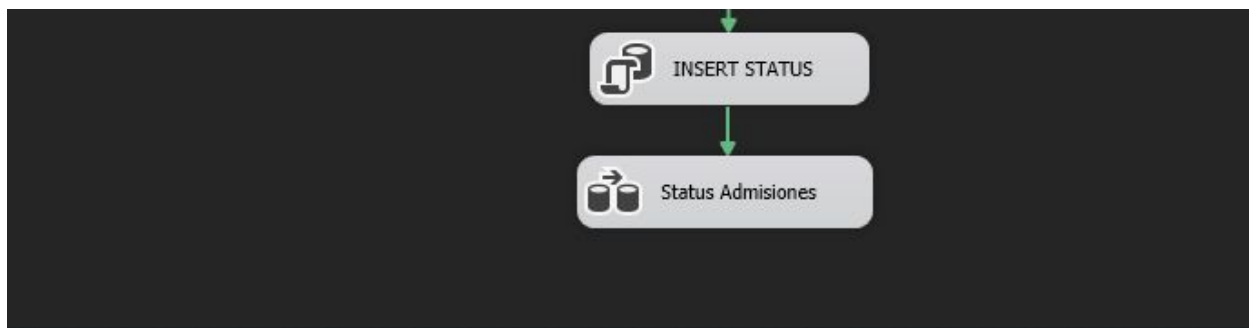
- La tarea “INSERT INTO FACULTAD”, inserta en la tabla de Facultad el registro “Ingeniería” en caso de que este aún no haya sido creado.
- La tarea “Carreras”, obtiene del archivo de excel los valores de esa columna y comprueba que los valores a insertar no existan basándose en el nombre de este.
- La tarea “Candidato”, se encarga de realizar la carga de nombres de candidatos y sus identificadores basados en la columna “Número” del archivo de excel. Esta tarea realiza

también la asociación de las columnas de municipio, colegio, diversificado. Propiedades que conforman a un candidato.

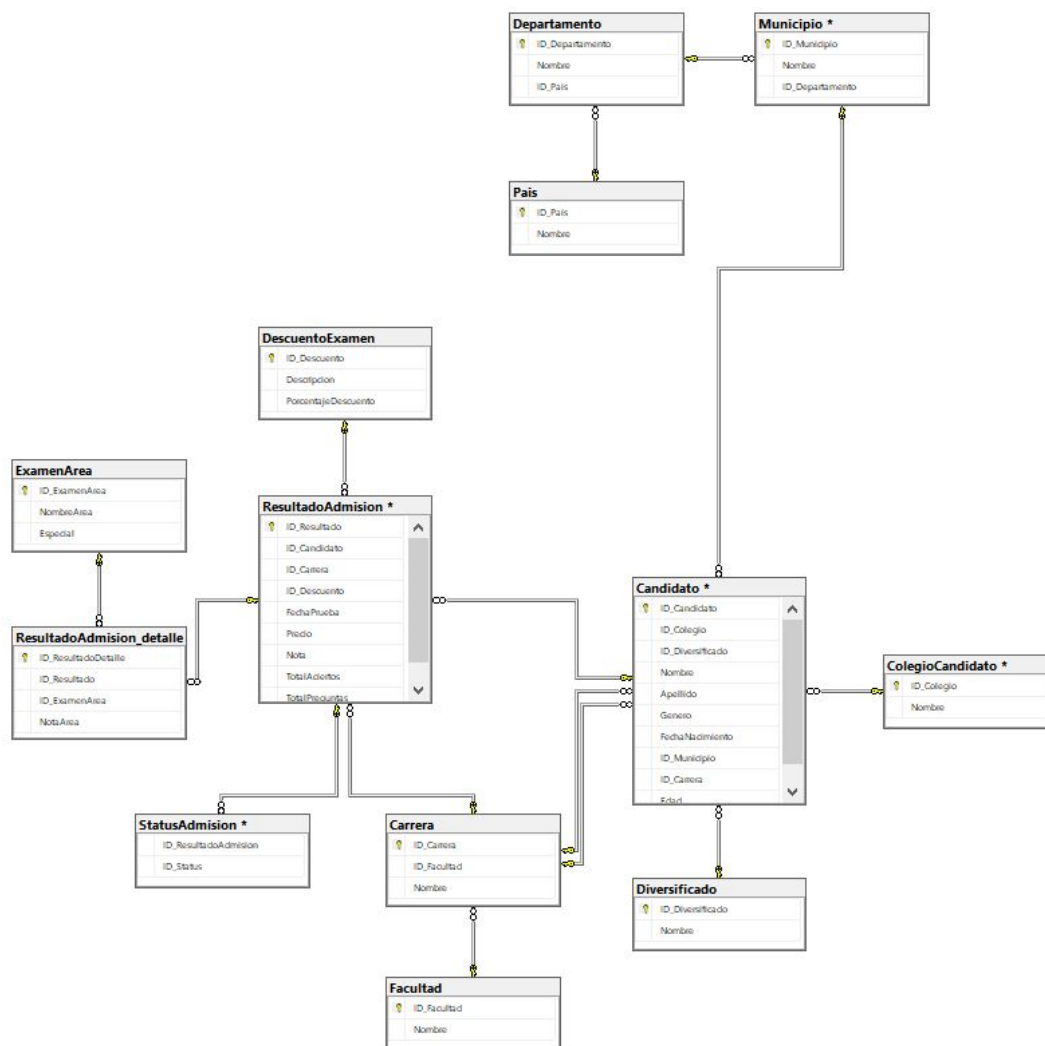
- Luego se encuentran las tareas que realizan la inserción de los resultados obtenidos en las distintas áreas. Estos resultados se asocian a un “resultado general” que pertenece a un estudiante en una fecha específica.
- Finalmente, se realiza la carga del archivo que contiene información del status de los estudiantes que tienen clasificación B. Esta tarea toma los registros del archivo de Status y los asigna a un resultado de un estudiante igual al valor de la columna “Número”.







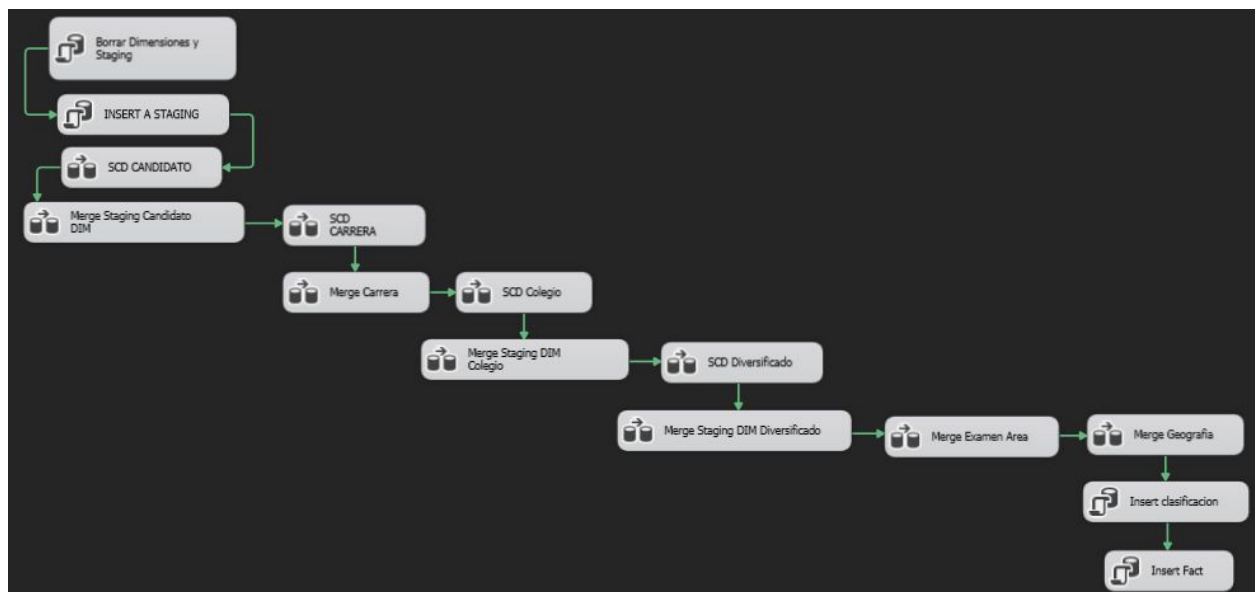
A continuación se muestra el diseño de la base de datos relacional.



Carga de Modelo Relacional a Data Ware House (ELT)

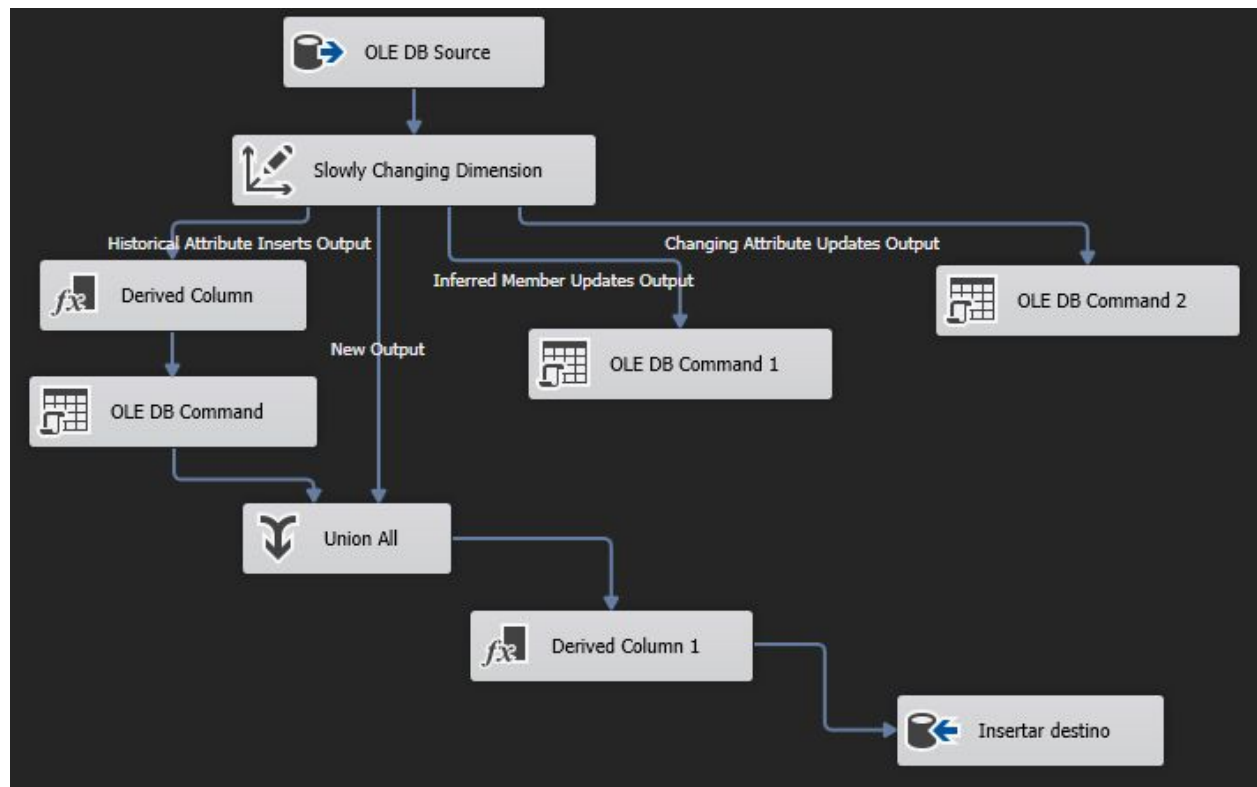
Este paquete se utiliza para realizar la carga de los datos del Staging a las dimensiones correspondientes a continuación se explican los procesos:

- La tarea llamada “Borrar Dimensiones y Staging” se utiliza para realizar un truncate de las tablas de Staging, tabla de Fact y dimensiones de fecha y clasificación.
- La tarea llamada “Insert a Staging” obtiene los datos de la base de datos relacional y hace un insert a las tablas del insert.
- Luego a cada una de las columnas que pueden llegar a variar en el tiempo se les realiza un “Slowly Changing Dimension” la tarea se describe cómo SCD, para poder guardar una copia de los datos que llegasen a cambiar.
- Y como paso final por columna se realiza un merge de lo que está en “Staging” y “Dimension” verificando que columnas han cambiado en el “Staging” y las que han cambiado se trasladan a la tabla de su dimensión correspondiente.



Estructura Slowly Changing Dimension (SCD)

La estructura de SCD se realiza con una Base de Datos fuente la cual nos indica que tabla se está trabajando conjunto a su información, la tarea de SSIS de SCD y el destino de la tabla que se está trabajando.



A continuación se describen los valores utilizados para las columnas de SCD los cuales son:

- **Atributos Históricos:** Este tipo de valor se le coloca a todas las variable que puedan cambiar en el tiempo, es decir que tengan una repercusión en el historial del estudiante a la hora de realizar la evaluación además se utilizan para llevar un record de los cambios que pueda tener el estudiante en aspectos que generen valor a la hora de realizar un reporte.
- **Atributos Variables:** Estos valores son colocados en todas las variables que no llegan a influir a la hora de querer conocer el historial de un estudiante. Cómo por ejemplo la

edad es un valor que tiende a cambiar año con año y no genera algún valor a la hora de realizar la reportería necesaria.

En las columnas de Slowly Changing Dimension se utilizan valores para guardar un historial para conocer de donde provienen los cambios o cuando se realizaron a continuación estas se llegan a aplicar en todas las columnas que requieran guardar datos históricos. A continuación se listan los atributos generales

- FechaCreacion
- FechaModificacion
- ID_Batch
- ID_SourceSystem
- UsuarioCreacion
- UsuarioModificacion

Las columnas que llevan SCD se listan a continuación

- **Candidato**

Columnas de dimensión	Cambiar tipo
ApellidoCandidato	Atributo histórico
Edad	Atributo variable
FechaCreacion	Atributo variable
FechaModificacion	Atributo variable
FechaNacimiento	Atributo histórico
Genero	Atributo histórico
ID_Batch	Atributo variable
ID_SourceSystem	Atributo variable
Municipio	Atributo histórico
NombreCandidato	Atributo histórico
NombreColegio	Atributo variable
NombreDiversificado	Atributo variable
UsuarioCreacion	Atributo variable
UsuarioModificacion	Atributo variable

- Carrera

Columnas de dimensión	Cambiar tipo
FechaCreacion	Atributo variable
FechaModificacion	Atributo variable
ID_Batch	Atributo variable
ID_SourceSystem	Atributo variable
NombreCarrera	Atributo histórico
NombreFacultad	Atributo histórico
UsuarioCreacion	Atributo variable
UsuarioModificacion	Atributo variable

- Colegio

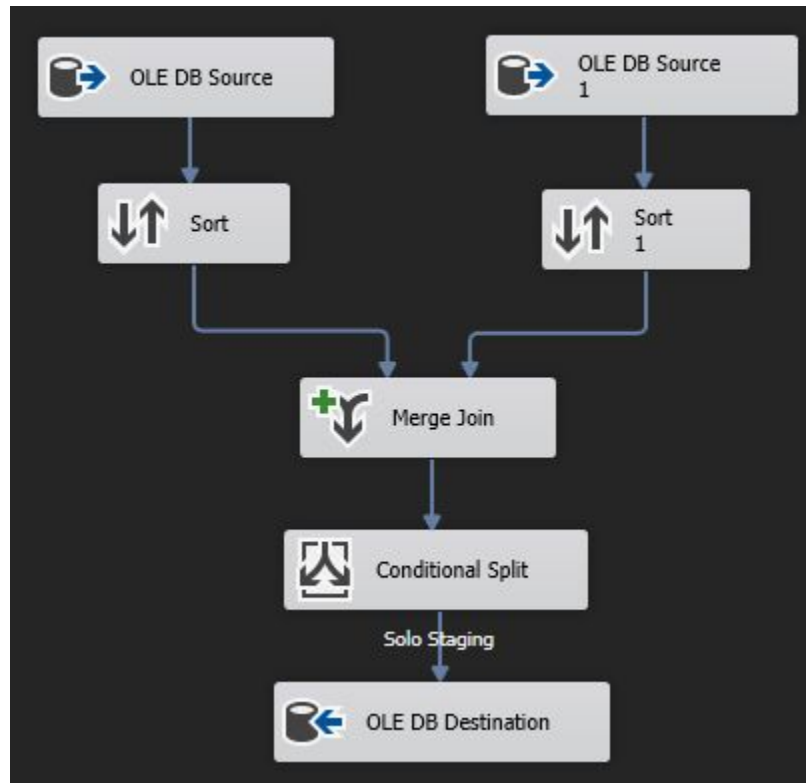
Columnas de dimensión	Cambiar tipo
FechaCreacion	Atributo variable
FechaModificacion	Atributo variable
ID_Batch	Atributo variable
ID_SourceSystem	Atributo variable
Nombre	Atributo histórico
UsuarioCreacion	Atributo variable
UsuarioModificacion	Atributo variable

- Diversificado

Columnas de dimensión	Cambiar tipo
FechaCreacion	Atributo variable
FechaModificacion	Atributo variable
ID_Batch	Atributo variable
ID_SourceSystem	Atributo variable
Nombre	Atributo histórico
UsuarioCreacion	Atributo variable
UsuarioModificacion	Atributo variable

Estructura Merge

A continuación se describe la estructura utilizada para realizar el merge de las tablas de “Staging” y “Dimension” correspondientes.

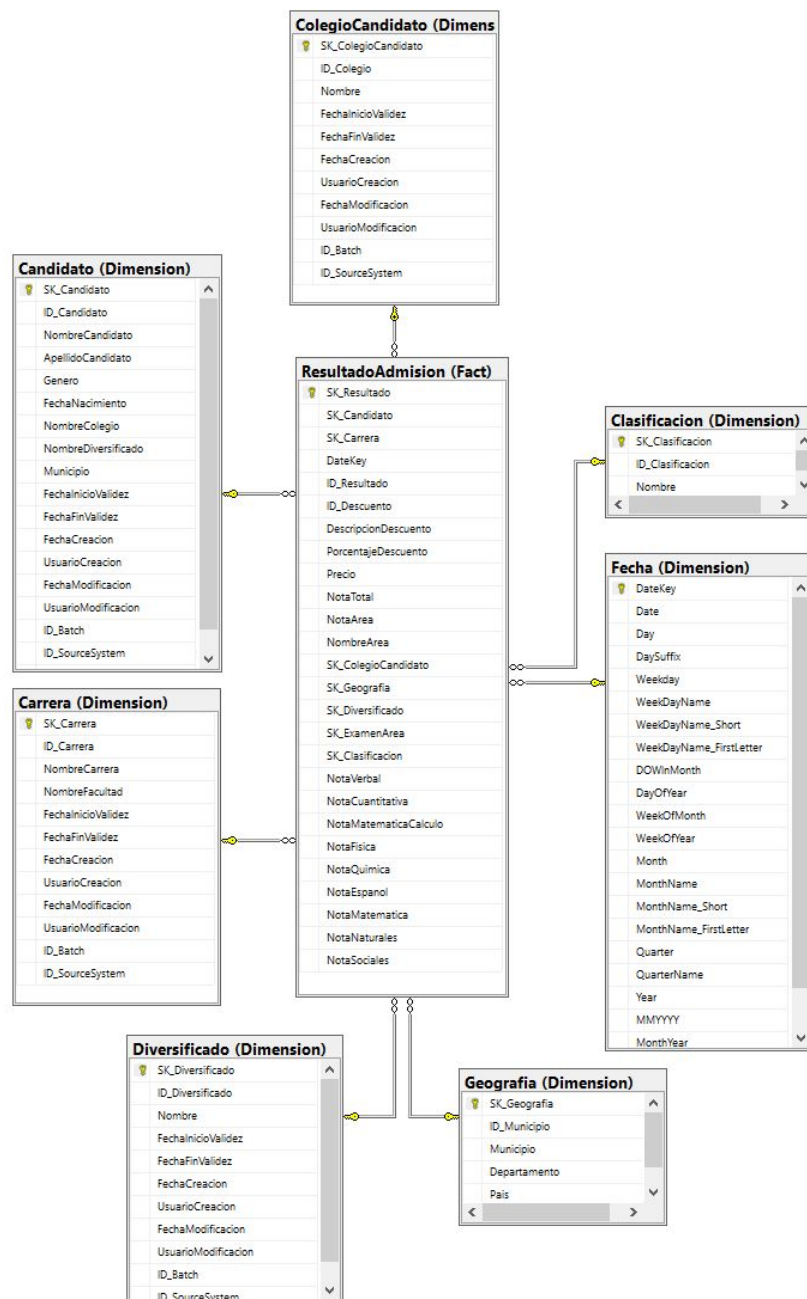


Se utilizaron dos fuentes de bases de datos una que tiene la conexión a la tabla de “Staging” y otra con la conexión a la tabla de “Dimension”, a cada una de ellas se les ordenó mediante su identificador el cual empieza con “ID_” ordenándolos ascendentemente.

Luego se realiza un JOIN LEFT para obtener todos los datos que estén en “Staging” y los que se tengan en común con “Dimension”. Sin embargo como se mencionó anteriormente también se traen los datos que se tienen en común estas dos tablas por lo cual se debe de realizar un “Conditional Split” lo cual se realiza la discriminación de todos los datos que no están en “Dimension” y únicamente están en “Staging”.

Luego de pasar por el “Conditional Split” esté entrega como salida todos los valores que no están únicamente en “Staging” y se hace la inserción en la tabla “Dimensión”.

Estructura DWH



El esquema utilizado en el DWH de este proyecto fue el esquema de estrella. La razón principal de esto es debido a que los esquemas en estrella están desnormalizados, lo que significa que las reglas normales de normalización aplicadas a las bases de datos relacionales transaccionales se relajan durante el diseño y la implementación del esquema en estrella. Los beneficios de la desnormalización del esquema estelar son:

- Consultas más simples: la lógica de combinación de esquema en estrella es generalmente más simple que la lógica de combinación requerida para recuperar datos de un esquema transaccional altamente normalizado.
- Aumento del rendimiento de las consultas: los esquemas en estrella pueden proporcionar mejoras de rendimiento para las aplicaciones de informes de solo lectura en comparación con los esquemas altamente normalizados.
- Agregaciones rápidas: las consultas más simples contra un esquema en estrella pueden mejorar el rendimiento de las operaciones de agregación.
- Cubos de alimentación: el esquema en estrella es ampliamente utilizado por todos los sistemas OLAP para diseñar cubos OLAP de manera eficiente. De hecho, los principales sistemas OLAP ofrecen un modo de operación ROLAP que puede usar un esquema en estrella como fuente sin diseñar una estructura de cubo.
- Business Intelligence: El esquema de estrella es el más utilizado en las implementaciones de almacenamiento de datos, por lo cual es ampliamente respaldado por una gran cantidad de herramientas de inteligencia empresarial.

Jobs

Junto con el deployment se creó un job el cual ejecuta periódicamente los paquetes de Integration Services los cuales refrescan la información, este se llama Grupo8_Job_ETL_ELТ y ejecuta los paquetes ETL.dtsx y ELТ.dtsx, esto lo ejecuta cada seis horas.

Proyecto SSAS

El proyecto SSAS consiste en la creación del cubo tomando como base el modelo de Data Ware House descrito anteriormente. Para su elaboración se utilizó el wizard que provee Visual Studio.

Se utilizó un modelo OLAP para la construcción del cubo porque permite realizar consultas y análisis fácilmente basándose en las distintas dimensiones. También facilita el proceso de generar reportería.

Las tabla que se utilizó como medida fue la de Resultados de Admisión. Esta tabla tiene asociaciones de llave foránea a las dimensiones que componen el cubo. Estas dimensiones son el área a la que pertenecen los resultados de las evaluaciones, la clasificación dada al estudiante basada en el resultado de su prueba, la carrera del estudiante al que corresponde el resultado; la geografía, que consiste en una jerarquía compuesta por país, municipio y departamento; la dimensión de fecha consta de una jerarquía que permite realizar consultas por año, trimestres, meses, semanas y días. La dimensión de candidato relaciona cada uno de los resultados con la tabla de candidatos, permitiendo así obtener el nombre, edad, género y datos geográficos del estudiante. Finalmente se cuenta con la dimensión de Diversificado que contiene únicamente el nombre del diversificado al que pertenece el estudiante de los resultados.

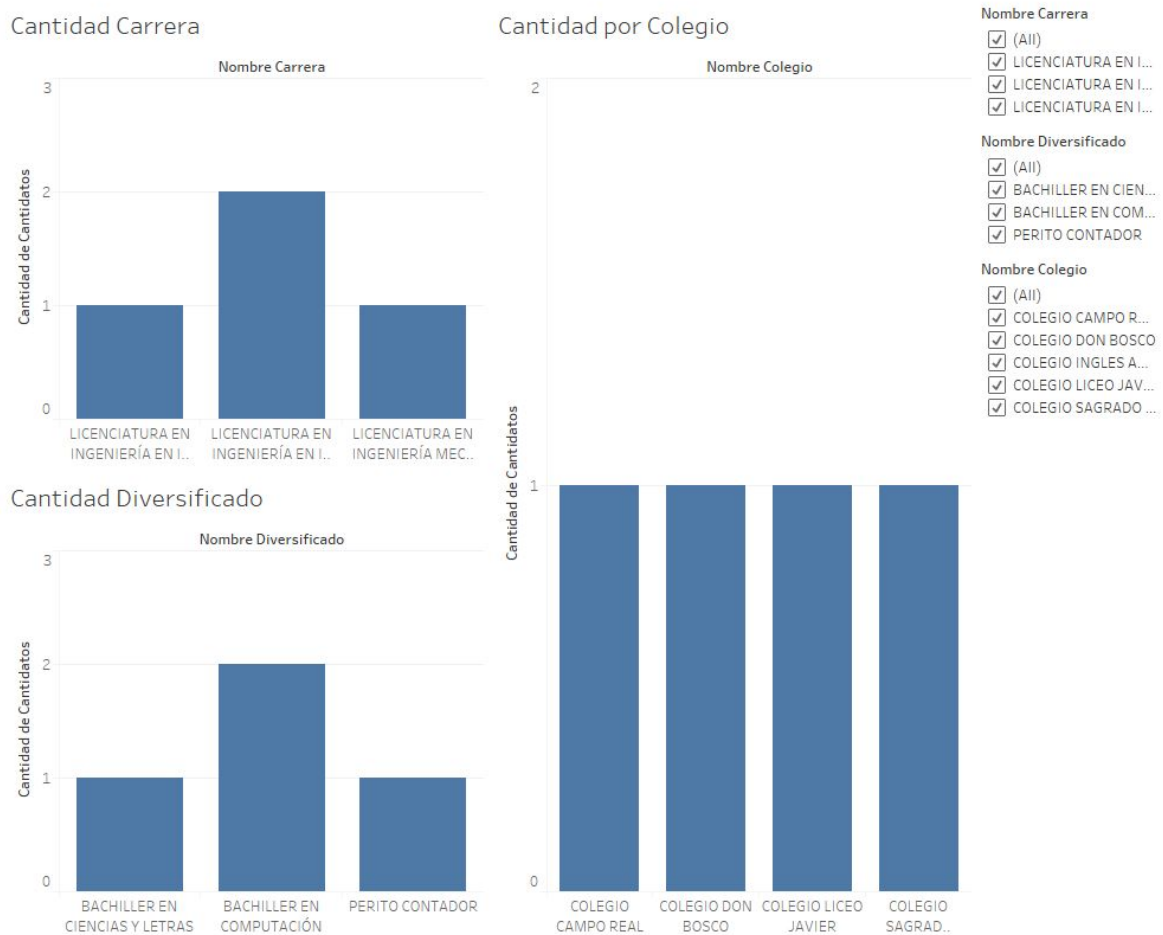
A continuación se muestra el diagrama del cubo. Puede observarse que es un modelo de estrella en la primera forma normal.



Reportes Tableau

En lo que respecta a la reportería del sistema de análisis de datos de indicadores académicos, se crearon reportes bajo tres categorías: Cantidades de Candidatos, Mapas de Candidatos, y Notas de Evaluaciones de Admisión.

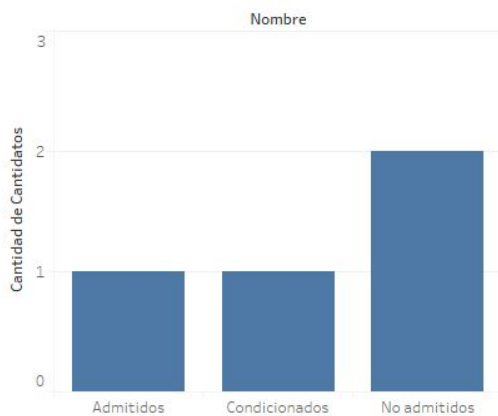
Gráficas de Cantidad #1



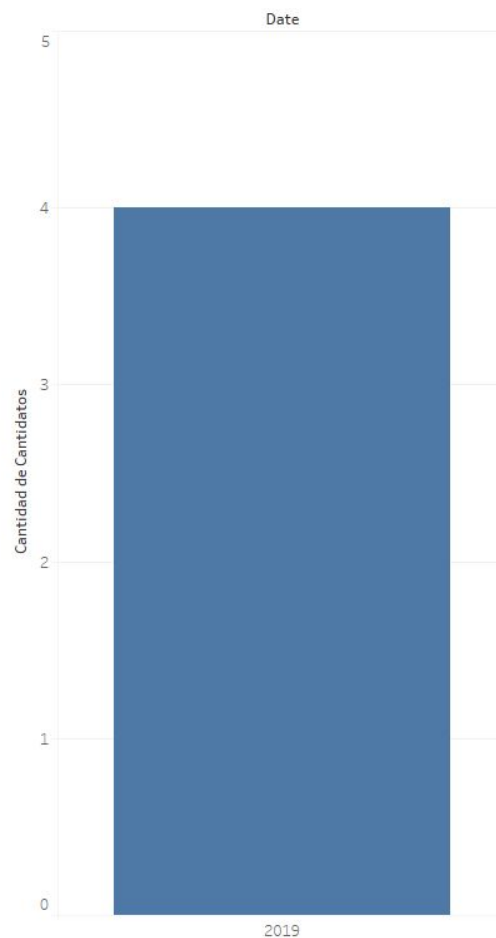
En este se incluyen reportes de Cantidad de Candidatos por Carrera a la que se aplica, por Colegio proveniente y por el Diversificado del Candidato.

Gráficas de Cantidad #2

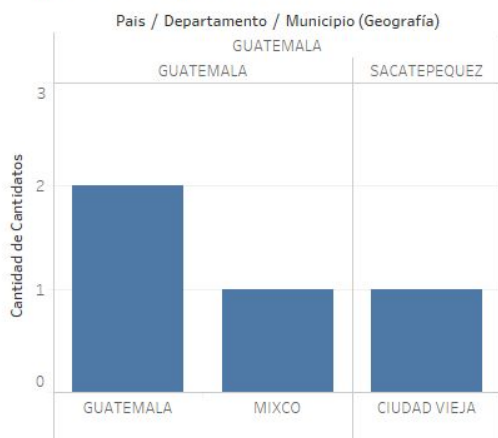
Cantidad Clasificación



Cantidad Fecha



Cantidad Geografía

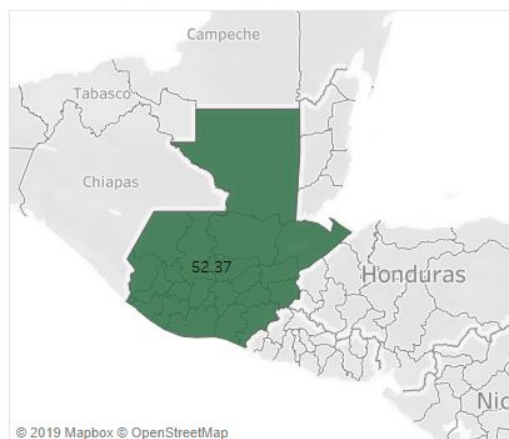


- Nombre
- ☒ (All)
 - ☒ Admitidos
 - ☒ Condicionados
 - ☒ No admitidos
- Date
- 1/1/2016 12/30/2020
- Municipio (Geografía)
- ☒ (All)
 - ☒ CIUDAD VIEJA
 - ☒ GUATEMALA
 - ☒ MIXCO
- Departamento
- ☒ (All)
 - ☒ GUATEMALA
 - ☒ SACATEPEQUEZ

En este se incluyen reportes de Cantidad de Candidatos por Clasificación obtenida, por rango de fechas y por el área geográfica de la cual proviene el estudiante..

Reportería Geográfica

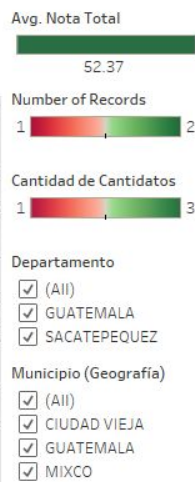
Promedio de Notas por País



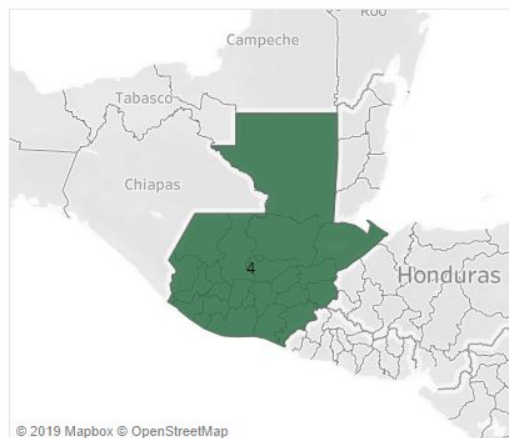
Número de Candidatos por Municipio



Promedio de Notas por Municipio



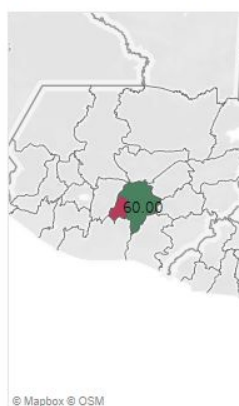
Número de Candidatos por País



Número de Candidatos por Departamento



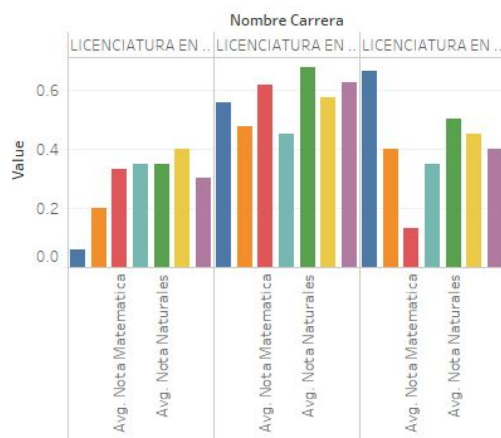
Promedio de Notas por Departamento



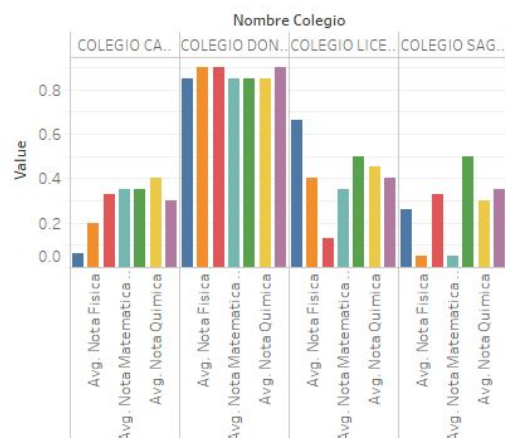
En este se incluyen reportes geográficos los cuales muestran el número de candidatos y promedio de notas por País, Departamento y Municipio.

Reporteria de Notas #1

Promedio de notas por área de Carrera



Promedio de notas por área de colegio



Measure Names

- Avg. Nota Español
- Avg. Nota Física
- Avg. Nota Matemática
- Avg. Nota Matemática
- Avg. Nota Naturales
- Avg. Nota Química
- Avg. Nota Sociales

Nombre Diversificado

- BACHILLER EN CIENCIAS Y LETRAS
- BACHILLER EN COMPUTACIÓN
- PERITO CONTADOR

Avg. Nota Total

149.74

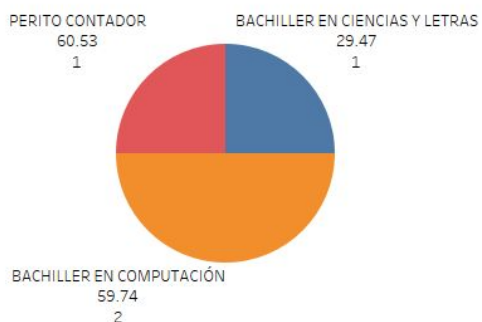
Nombre

- Admitidos
- Condicionados
- No admitidos

Nombre Diversificado

- ☒ (All)
- ☒ BACHILLER EN CIENCIAS Y LETRAS
- ☒ BACHILLER EN COMPUTACIÓN
- ☒ PERITO CONTADOR

Nota promedio por diversificado



Candidatos por diversificados admitidos y no admitidos

Nombre Diversificado

BACHILLER EN CIENCIAS Y LETRAS

1
No admitidos

BACHILLER EN COMPUTACIÓN

1
Admitidos

PERITO CONTADOR

1
Condicionados

Nombre

- ☒ (All)
- ☒ Admitidos
- ☒ Condicionados
- ☒ No admitidos

En este se incluyen reportes de notas de los candidatos. Se poseen los siguientes reportes:

- Reporte de promedio de notas por área por carrera a la que se aplica.
- Reporte de promedio de notas por área por colegio proveniente.
- Reporte de promedio de notas por diversificado proveniente.
- Reporte de candidatos admitidos y no admitidos por diversificado proveniente.

Reporteria de Notas #2

Promedio por Diversificados Admitidos y no Admitidos

Nombre Diversificado

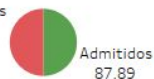
BACHILLER EN CIENCIAS Y LETRAS

No admitidos
29.47

BACHILLER EN COMPUTACIÓN

No admitidos

31.58



PERITO CONTADOR

Condicionados
60.53

Nombre

- ☒ (All)
- ☒ Admitidos
- ☒ Condicionados
- ☒ No admitidos

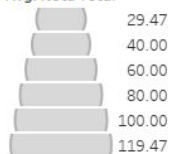
Nombre Diversificado

- ☒ (All)
- ☒ BACHILLER EN CIEN...
- ☒ BACHILLER EN COM...
- ☒ PERITO CONTADOR

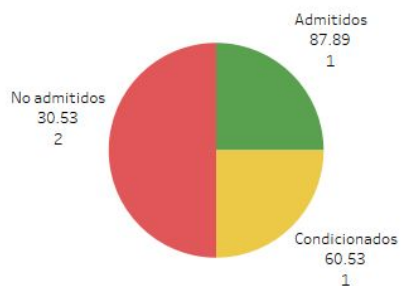
Nombre

- ☒ Admitidos
- ☒ Condicionados
- ☒ No admitidos

Avg. Nota Total



Admitidos vrs Condicionados vrs No admitidos



En este se incluyen reportes de notas de los candidatos. Se poseen los siguientes reportes:

- Reporte de promedio de notas de candidatos admitidos, no admitidos y condicionados; por diversificado proveniente.
- Reporte de promedio de notas de candidatos Admitidos, No Admitidos y Condicionados.