Carrera: Ingeniería en Informática y Sistemas

Curso: Análisis de Datos

Catedrático: Stanly Bolaños



Manual Técnico

Manuel Andoni Zamora Hernández

Joseph Abraham Soto Gutiérrez

Pedro Pablo Pineda Izquierdo

Tabla de contenido

[Modelo de la Base de Datos 3](#_Toc23696535)

[Tablas 3](#_Toc23696536)

[**Modelo Data Warehouse** 5](#_Toc23696537)

[**Paquete de Integración SSIS** 6](#_Toc23696538)

[**Carga de Datos** 6](#_Toc23696539)

[**Paquete de analisis SSAS** 9](#_Toc23696540)

[**Reportes** 10](#_Toc23696541)

# **Modelo de la Base de Datos**

****

## **Tablas**

* **Candidato la tabla de candidato contiene la información respectiva al candidato como lo es el nombre, fecha de nacimiento (calculado mediante la edad), Genero y sus respectivos datos al momento de realizar la prueba.**
* **Carrera la tabla carrera contiene las carrearas ofrecidas por las distintas facultades de la universidad, son identificadas por un nuevo número al momento de crear un nuevo registro en la base de datos**
* **ColegioCandidato la tabla de colegioCandidato contiene el nombre de los distintos colegios donde los candidatos se encuentran estudiando o se graduaron de ellos.**
* **Condicion la tabla condición ayuda a identificar el estado en el que se encuentre el estudiante si en caso de haber perdido las pruebas se le concede de forma condicionada continuar con la solicitud. Se identifica por medio de las siglas:**
  + PED - Pendiente Entrevista Director
  + PTD - Pendiente Tarea Director
  + AP - Asignado Propedéutico
  + PG - Propedéutico Ganado
* **Departamento permite almacenar los distintos departamentos que tenga un país**
* **DescuentoExamen tabla que registra los distintos descuentos otorgados a los estudiantes para poder ser aplicados al momento de pagar las cuotas mensuales. Actualmente no es utilizada, pero puede implementarse en un futuro.**
* **Diversificado registra las diferentes carreras de las que pueden estar cursando el candidato actualmente**
* **ExamenArea en esta tabla se registran las diferentes áreas de exámenes existentes siendo actualmente, pero con la capacidad de actualizarse las siguientes:**
  + **Area Basica**
    - Habilidades Verbales
    - Habilidades Cuantitativas
    - Español
    - Matemáticas
    - Ciencias naturales
    - Ciencias sociales
  + **Area Especifica**
    - Matemática para el cálculo
    - Física
    - Química
* **Facultad esta tabla almacena las distintas facultades que la universidad puede tener para cubrir las áreas de estudio que tengan actualmente.**
* **Municipio la tabla de municipio almacena identifica a un nivel mas exacto el lugar en el que vive**
* **Pais identifica el país del que proviene el candidato**
* **ResultadoAdmision está tabla contiene la información del estado de la admisión del usuario así como el resultado que obtuvo en general y permite tener toda la información accesible mediante las relaciones.**
* **ResultadoAdmision\_detalle esta tabla contiene la información respecto al detalle de notas en cada examen que realizo el estudiante**

# **Modelo Data Warehouse**

****

**Para el modelo data warehouse se escogió un esquema de tipo estrella debido a que la información que se tiene actualmente puede representarse en dimensiones separadas de forma completa, este modelo cuenta con las dimensiones Carrera, Región, Fecha, Candidato y Examen. El modelo estrella presenta la ventaja de ser más simple por lo que ofrece un mayor rendimiento en comparación al modelo copo de nieve.**

**La desnormalización fue efectuada alrededor de la tabla de Hechos, Resultado de admisión. Esto es debido a que las notas totales nos permiten hacer varios análisis que le dan valor a la universidad. Las dimensiones nos permitirán dar contexto a las notas para poder entender varios aspectos de como se comportan estos datos.**

## **Tablas**

Dimensiones

Las siguientes dimensiones permiten hacer filtrado de la data fácilmente e identificar los datos para su fácil análisis.

* Candidato
* Carrera
* Fecha
* Examen
* Region

Hechos

La tabla de hechos contiene las medidas aditivas y semi aditivas necesarias para comparar los resultados y poder realizar análisis.

* ResultadoAdmision

Staging

Estas tablas apoyan al esquema para poder realizar la carga de datos

* Candidato
* Carrera
* Examen
* Fecha
* Region

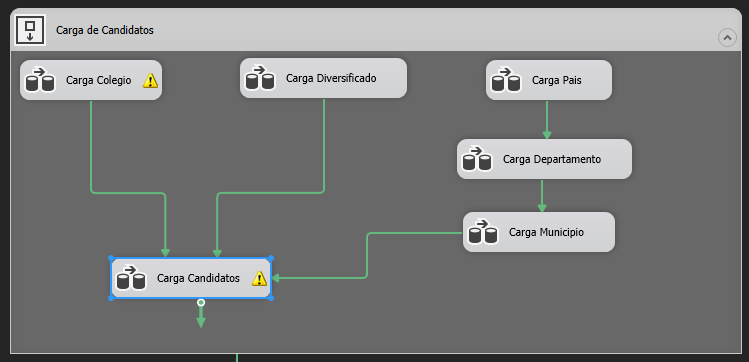
****

# **Paquete de Integración SSIS**

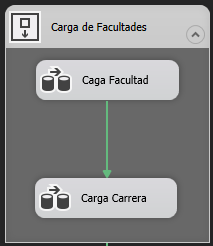
## **Carga de Datos**

Para la carga de datos se escogió utilizar el proceso de ETL debido a que nos ofrece una gran habilidad de transformar la data y manipularla antes de que llegue al data warehouse. Al mismo tiempo permite separar varios aspectos de la lógica y debido a que la carga de datos se hace en fechas específicas del año puede ser manejable.

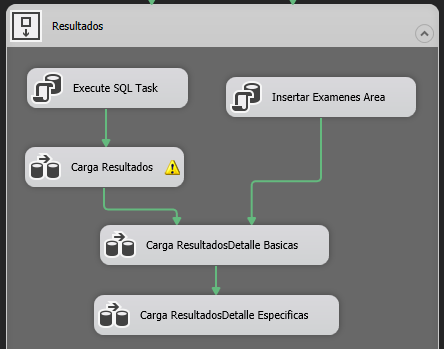
Primero se realiza la carga de datos de candidatos, siendo las tablas que no tienen dependencias hacia otra las primeras por ser cargadas.



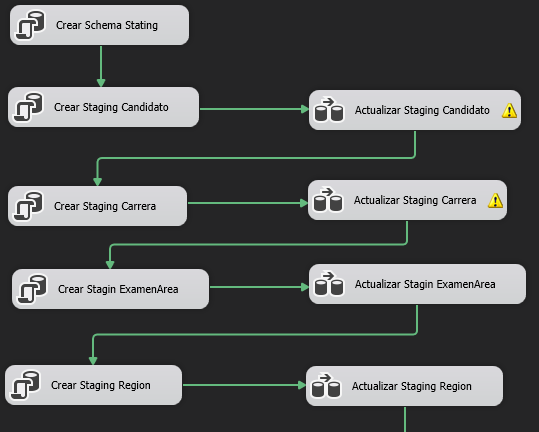
**Se extraen la data de facultades y la carga de carreras que aspiran los candidatos**



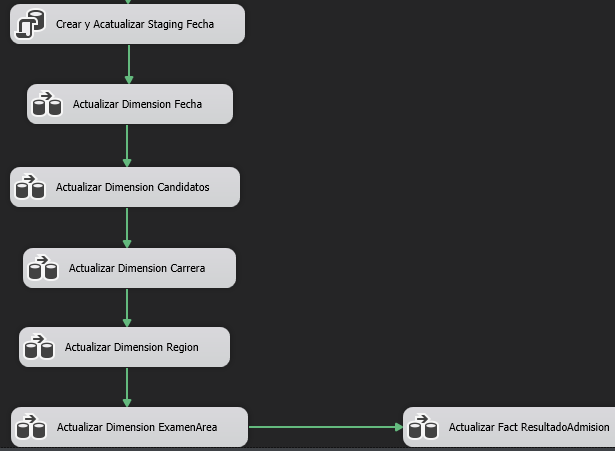
**Por ultimo se cargan los resultados de los estudiantes en su respectivo ID por examen.**



**Luego se pasa a la carga de datos al modelo data wawrehouse, para esto se crean las tablas de Staging y procede a cargar los datos en cada una de ellas**

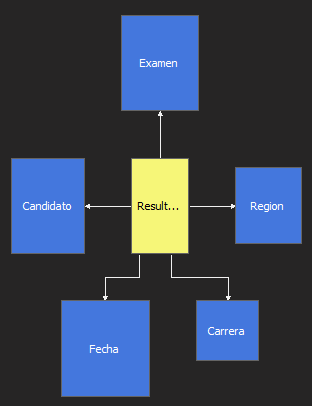


**Por último se actualizan las fechas para poder obtener información adicional y se procede a la carga en de cada una de las dimensiones.**



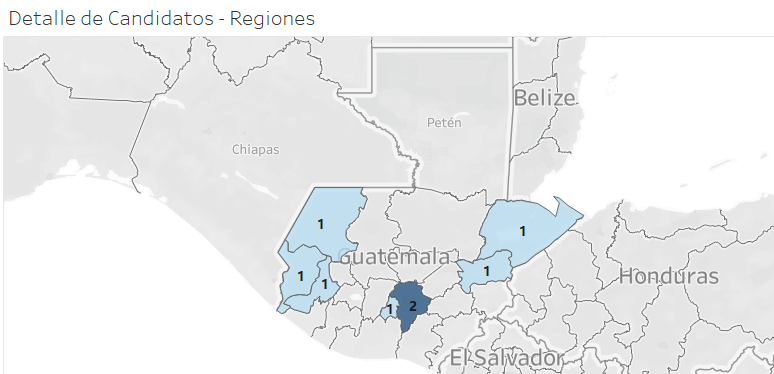
# **Paquete de análisis SSAS**

Se utilizo un cubo MOLAP para tener análisis de múltiples dimensiones. Debido a que se crea una copia esto nos permite tomar la data que se desee analizar y trabajar con ella sin tener que recurrir a una búsqueda en las particiones de la base de datos. Al mismo tiempo nos permite tener mayor eficiencia al momento de realizar consultas grandes y que requieran de computo.

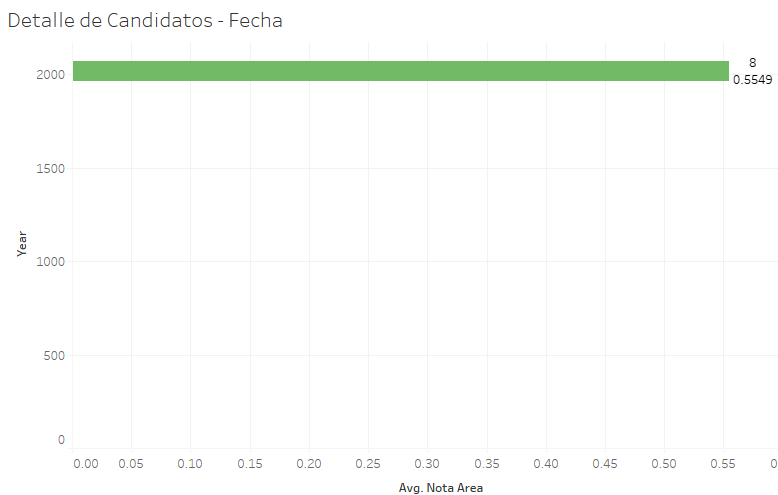


# **Reportes**

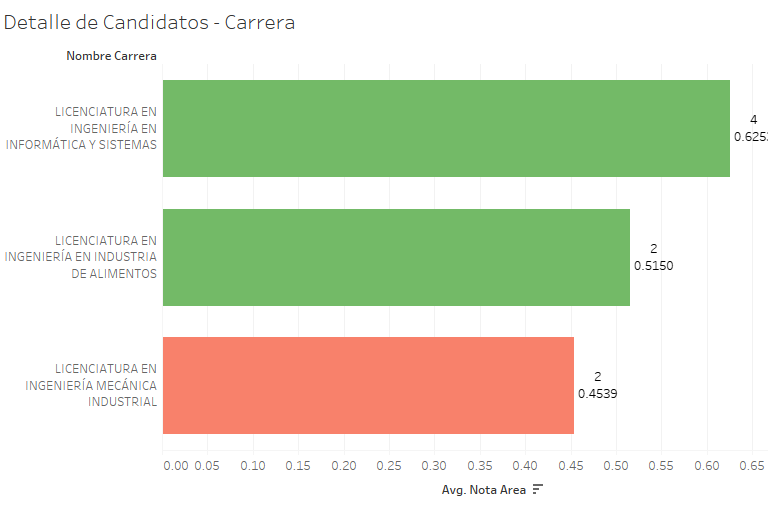
* Gráficas de cantidades de candidatos por
  + Año
  + Carrera
  + Diversificado
* Gráficas de mapas de candidatos por
  + País
  + Departamento
  + Municipio
* Gráficas de notas de evaluaciones de admisión por
  + Admitidos vs Condicionados vs No Admitidos
  + Departamento (preferible mapa con escala de colores)
  + Diversificado
  + Áreas de evaluación



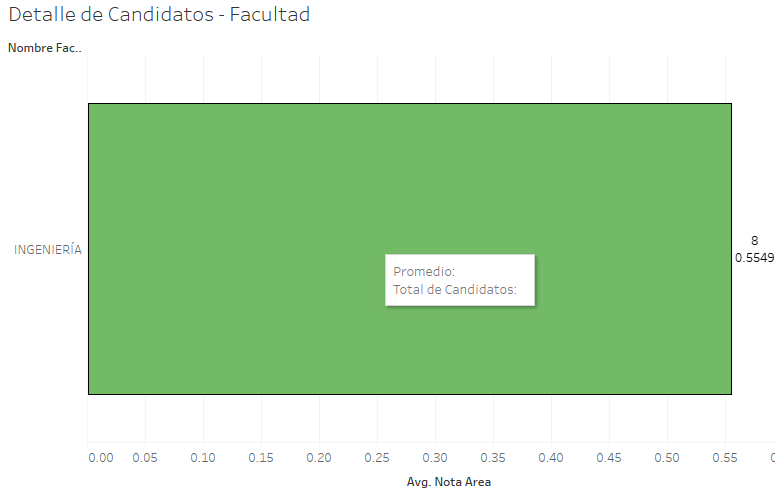
Muestra la cantidad de alumnos por región



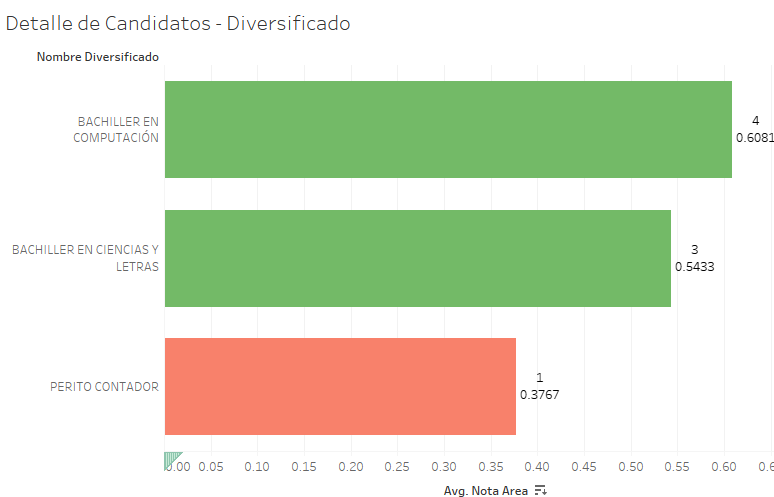
Muestra el promedio de nota por fecha



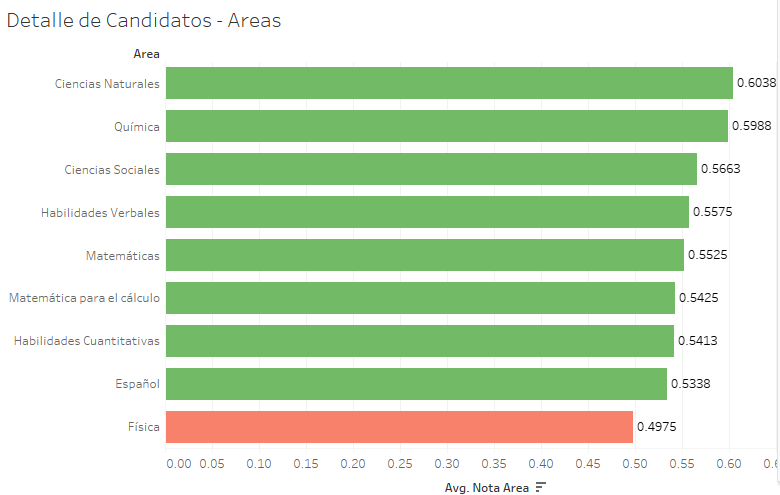
Muestra el promedio de notas por carrera



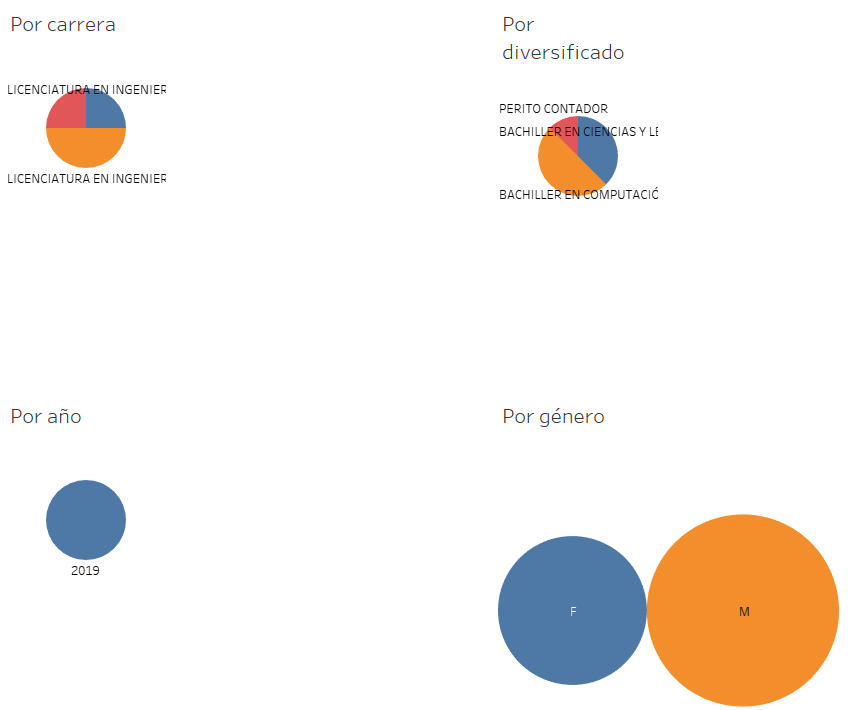
Muestra el promedio de notas por facultad



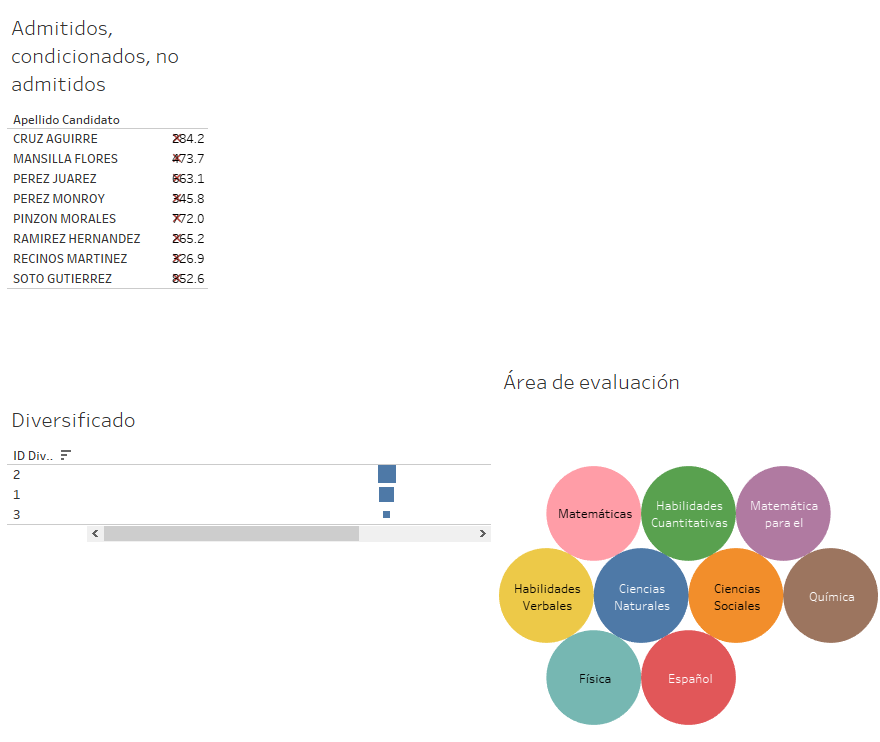
Muestra el promedio de nota por diversificado



Muestra el promedio de nota por Área



Metra el total de personas por los criterios año, carrera, diversificado y genenro



Muestra el total de condicionados y las áreas que se evaluaron