

Segundo proyecto - Entrega 2

Inteligencia de Negocios

1a.

El requerimiento analítico a trabajar corresponde a: "Identificar características deseables de salones"
Este requerimiento es apoyado por los siguientes objetivos de negocio:

- Agilizar etapa de negociación
- Aumentar comodidad de estudiantes y profesores

Roles de las personas en la institución que se benefician:

- Estudiantes
- Profesores
- Ingenieros de salones

b.

El periodo de tiempo a analizar es el primer semestre de 2018. En este periodo de tiempo es que se generan y resuelven las solicitudes de salones y también es cuando se llevan a cabo los eventos para los que se solicitan los salones.

Antes de proceder, es importante mencionar que solo trabajaremos con 2 archivos. [solicitudes_enero_2018.xlsx](#) (252,695 KB) y el inventario de salones provisto por la gerencia del campus¹. Todos los demás archivos fueron descartados para el análisis dado que no contenían datos interesantes para el mismo. En particular, el único dato que tenía alguna relación con los salones era la disponibilidad de la sección. Además, todos los demás archivos contenían información relacionada con salones de secciones de asignaturas dictadas en la universidad, pero esto no es de nuestro interés porque no nos interesa el salón que fue asignado sino el salón que se quería (y sus características). Por otro lado, notemos que en el diagrama aparecen 2 tablas de hechos. Esta era la intención al comienzo pero dada la calidad de los datos nos vimos forzados a trabajar en la implementación solo con la tabla de hechos Solicitud Salon. Por esta razón el punto 2 solo considerará esa tabla.

2.

Dada la naturaleza de los datos, no se puede implementar la tabla de hechos SolicitudCambioSalon (véase el diagrama abajo). Sin embargo, si se puede implementar SolicitudSalon como se muestra a continuación:

Tabla SolicitudSalon

Especificar la granularidad: Una fila en esta tabla corresponde a cualquier solicitud creada en enero de 2018 de un salón para algún evento que se llevará a cabo durante el semestre. De hecho, hay una correspondencia biyectiva entre las entradas de esta fact table y los registros en el archivo [solicitudes_enero_2018.xlsx](#).

Las medidas de esta tabla son: Estado, Capacidad deseada, Permanente.

¹ <https://campusinfo.uniandes.edu.co/es/recursos/espaciosacademicos/salones>

Estado y permanente corresponden a medidas no aditivas (no son numéricas). Aun así, sobre ellas podemos ejecutar operaciones que son de nuestro interés, por ejemplo, podemos agrupar y contar para saber la cantidad de solicitudes de un estado en particular, o saber cuántas solicitudes son permanentes.

La capacidad deseada es una medida semi-aditiva. Sumar las capacidades de todas las solicitudes no tienen ningún tipo de sentido ni valor en el análisis. Aun así, es de nuestro interés conocer máximos, mínimos y promedios de capacidades solicitadas.

bi.

DimFechaEvento:

Mes: este atributo es de interés porque queremos usarlo para filtrar. En el proceso de asignación de espacios puede haber un momento del semestre en el que haya bastantes solicitudes y la disponibilidad de los espacios sea reducida.

Semana: Este atributo es de interés por razones similares al atributo mes.

Día: A parte de las razones expuestas anteriormente (para semana y mes) este atributo es particularmente importante para solicitudes de espacios permanentes, pues el espacio que se asigne será utilizado ese día en particular durante todo el semestre.

DimSalonPreferencia

idSalon (PK) es importante para identificar un salon en particular.

Edificio: A veces queremos filtrar por bloques, es de interés identificar los bloques más solicitados, o identificar características deseables que estén asociadas a un edificio en particular.

Numero Salon: Este atributo es importante para poder utilizar el dato “salon de preferencia” de los archivos que vamos a usar. Con el numero del salon y el edificio y los datos de gerencia del campus podemos tener información completa de ese salon.

DimCaracteristicaSalon

idCaracteristicas (PK): Este atributo es importante para poder identificar un conjunto particular de características de un salon

tipoSalon: Al ser una característica del salon, es importante dado que nuestro proceso de negocio es identificar características.

Acceso a internet: Al ser una característica del salon, es importante dado que nuestro proceso de negocio es identificar características.

DimActividad:

tipoActividad (PK): Este atributo es importante porque las características solicitadas pueden depender de la actividad para la que se solicita el espacio. Por ejemplo, para una conferencia se suelen necesitar espacios de gran capacidad.

DimDependencia:

Facultad-dependencia: Este atributo es importante porque queremos filtrar las solicitudes de los salones por dependencia de la Universidad. Por ejemplo, es más probable que la Facultad de Ingeniería solicite laboratorios con tecnologías particulares a que lo haga la Facultad de Derecho.

Departamento: Este atributo es importante por lo mismo que el anterior. Aca aparece una jerarquía, por ejemplo el departamento es mas especifico que la facultad

DimHora:

Los atributos de esta dimensión son importantes por razones similares a la fecha. Queremos saber y filtrar por el momento en el que sucede el evento. Es una dimensión aparte para no tener una cardinalidad tan grande en una sola dimensión.

ii. Slowly Changing Dimension:

La mayoría de atributos no cambian con el tiempo, por lo que tenemos tipos 0. En particular, todos los identificadores son de este tipo

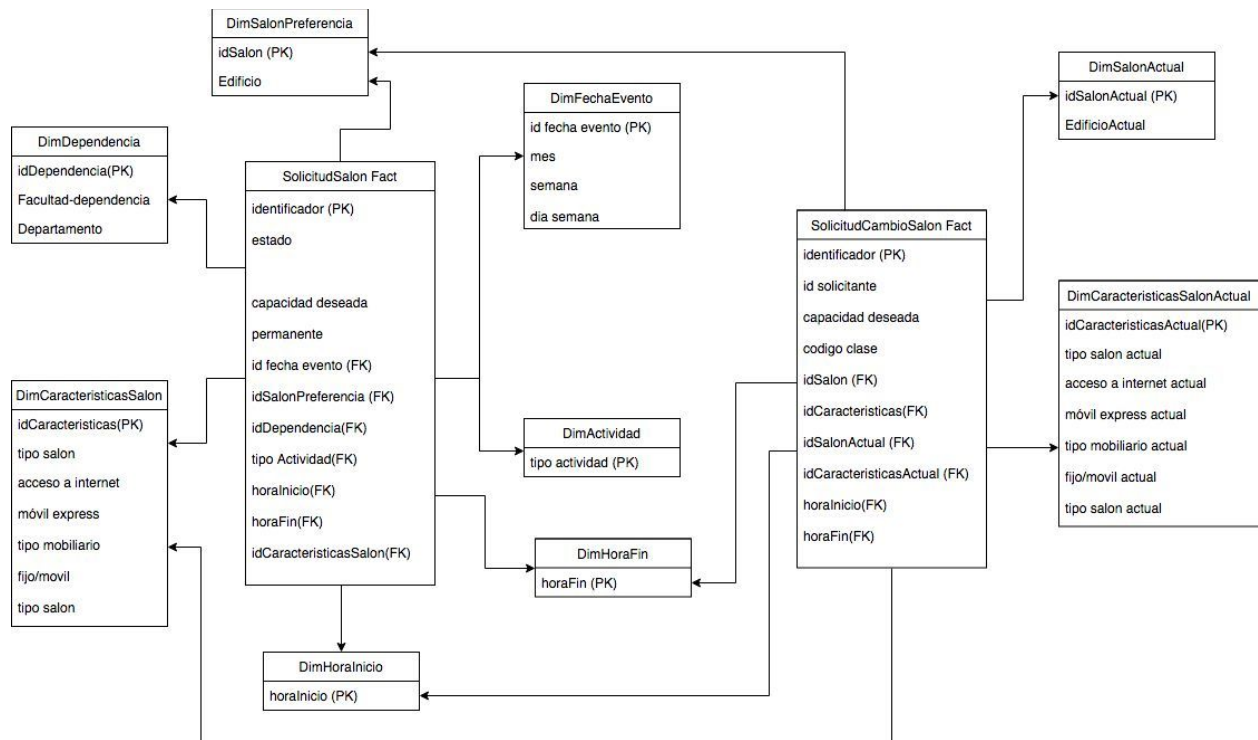
Por otro lado, los siguientes son los campos que pueden cambiar con el tiempo:

Estado: Puede cambiar de PENDIENTE a ANULADA, RECHAZADA o LISTO.

La política para actualizar este campo en caso de ser necesario será sobre-escribir la entrada de la tabla. Es decir, acá tenemos un tipo 1.

En la fact table que no se implementó la dimensión que cambia es el salón actual. De haberla podido implementar (si los datos lo hubieran permitido), nuestra política para actualizar esta dimensión sería agregar una nueva fila, es decir, tipo 2. Esta decisión teniendo en cuenta que quisiéramos llevar una historia del proceso de reasignación del salon.

c.



d.

El siguiente es un ejemplo de la fact table Solicitud Salon

| Id | estado | cap | M.exprss | id.f.e | preferenc. | Id dep | tipo act | hora in | horafin | id car |
|--------|--------|-----|----------|--------|------------|--------|----------|---------|---------|--------|
| 532307 | LISTO | 10 | NO | 1 | ML_110 | 4 | Clase | 8 | 9 | 15 |

El siguiente es un ejemplo de la dimensión fecha evento

| Id fecha evento | Mes | Semana | Día semana |
|-----------------|-----|--------|------------|
| 1 | 2 | 3 | Jueves |

3a.

Como se mencionó anteriormente, solo se usaron dos archivos para esta entrega. A continuación exponemos los resultados del perfilamiento para cada uno de ellos.

Archivo 1: [solicitudes_enero_2018.xlsx](#)

El objetivo de esta fuente es exponer información acerca de solicitudes hechas por miembros de la comunidad estudiantil de salones para actividades diferentes de clases (monitorias, exámenes, foros, entre otras). Contiene 1513 filas. Además, tiene información del dónde y el cuándo del evento en detalle. En particular, los siguientes campos son considerados importantes:

lk01identificador: Identificador único de la solicitud. Al realizar el perfilamiento de los datos encontramos que, en efecto, este campo es único para cada registro.

Cn01nombreactividad: Tipo de la actividad para la que se solicita el espacio (Clase, Monitoria, Sustentación) acompañado de una pequeña descripción. Nos interesa solo el tipo de la actividad, las descripciones aumentan considerablemente la cardinalidad de este campo. Antes de esta corrección se tienen 834 valores distintos.

Dn01fechaeventoinicial: Fecha en la que inicia el evento. Durante todo el semestre se solicitaron salones para 165 días diferentes. El día que mas se solicitó fue el 2 de febrero.

Dn01fechaeventofinal: Fecha en la que finaliza el evento, este campo y "Dn01fechaeventoinicial" son exactamente iguales, por lo que uno de ellos será eliminado.

Cn01estado: Estado de la solicitud, toma valores del conjunto {"LISTO", "ANULADA", "RECHAZADA", "PENDIENTE"}. Hay 1242 LISTO, 114 ANULADA, 95 RECHAZADA, 62 PENDIENTE.

Cn01horainicial: Hora en la que inicia el evento. Adopta 29 posibles valores, la hora de inicio más frecuente es 8:00AM

Cn01horafinal: Hora en la que finaliza el evento. Adopta 29 posibles valores, la hora de finalización más frecuente es 2:00PM

Cn01preferencia: Salon de preferencia de quien solicita. A veces no se especifica salón sino edificio. Solo nos interesa el lugar, además las descripciones aumentan considerablemente la cardinalidad de este campo. Antes de transformar los datos de este campo hay 865 valores

Cn01tiposalon: Tipo del salon que se necesita para el evento. Sólo hay 5 valores diferente para este campo. Siendo el más frecuente "Ninguno en especial" (904). Hay 3 registros sin datos en este campo.

Cn01capacidaddeseada: La capacidad requerida para poder llevar a cabo el evento. Hay 67 diferentes valores para este campo. El máximo es 250, el mínimo 3 y la media 38.626. Hay 3 registros que no tienen datos en este campo.

Cn01accesointernet: Indica si el salon requerido para el evento debe tener conexión a internet o no. Solo puede adoptar valores en el conjunto {SI, NO}. Hay 1217 SI, 293 NO. Hay 3 registros que no tienen datos en este campo.

Cn01permanente: Este campo indica si la solicitud es para uso permanente del salon durante el semestre. Hay 1232 NO, 278 SI y 3 faltantes

Cn01dependencia: La dependencia de la universidad a la que está asociada la solicitud. Hay 179 valores diferentes. Hay tres registros que no tienen información de este campo, a estos les asignaremos "FACULTAD DE INGENIERÍA" que es el valor más común. Es importante notar que la calidad de los datos en este campo no es tan buena como quisiéramos. En particular, hay entradas diferentes que significan lo mismo, por ejemplo "Dpto de industrial", "iind", "Departamento de Ingeniería Industrial". Corregir esto es un problema difícil entonces no lo haremos. Aun así, estos datos se pueden usar porque no todas las entradas que quieren significar lo mismo son diferentes y en esa medida este campo puede darnos una noción acertada acerca de la dependencia de la universidad que está relacionada con la solicitud. Por ejemplo, este campo nos sirve para saber que una de las dependencias que más solicita salones es la facultad de Ingeniería con mínimo 301 de las 1513 solicitudes.

Cn01diasemana: Día de la semana en la que se llevará a cabo el evento. El día de la semana que más tiene eventos es el jueves con 354 solicitudes.

No se consideran importantes los demás campos del archivo en la medida en que contienen información que no afecta de manera considerable el proceso de negocio que estamos trabajando. Por ejemplo, los datos de la persona que solicita el salon, y de la persona que procesa la solicitud no son de interés (Además se asume que la asignación de salones es independiente del usuario)

Archivo 2: Inventario de salones Gerencia del Campus

Bloque: El edificio en el que está ubicado el salón. Hay 20 distintos valores para este campo. Hay 2 filas que no tienen este dato.

Capacidad: Capacidad del salón, el valor máximo es 9284, el mínimo 10 y la media 94.312

Movil Express: Indica si el salón es compatible con movil express. Hay que considerar que hay entradas con "Si", "si" y "SI" (entradas diferentes que quieren significar lo mismo). Hay 121 SI (Suma de "Si", "si" y "SI") y 77 NO.

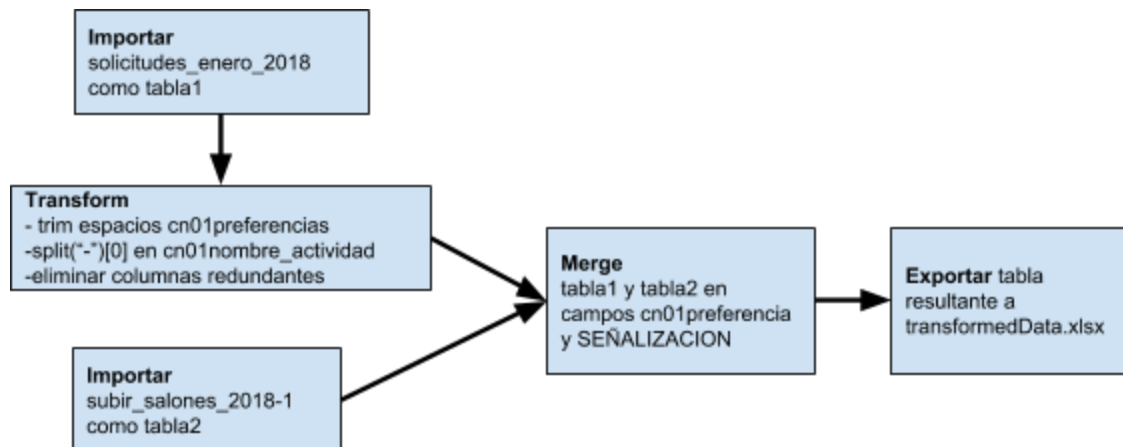
Señalización: nombre del salón, para nosotros será el identificador (es único)

Tipo mobiliario: Indica el tipo de muebles con los que cuenta el salón, puede tomar 7 distintos valores, el más común de ellos es "Silla móvil con mesa fija" con 87 apariciones.

Fijo/móvil: Indica si los muebles son fijos o móviles. Hay 102 fijos, 96 móviles.

Descartamos el campo área útil porque está fuertemente correlacionada con la capacidad. También descartamos la extensión porque no es de nuestro interés.

b.



c.

Las estadísticas detalladas de la tabla resultante se muestran en el archivo finalProfiling.html.

4a.

Dados nuestros objetivos de negocio:

- Agilizar etapa de negociación
- Aumentar comodidad de estudiantes y profesores

Dados los datos que nos proveyeron, podemos hacer dos análisis OLAP:

1. Ver las características deseables de un salón por la comunidad uniandina y qué bloques tienen estas características
2. Ver los bloques preferidos de la comunidad uniandina

El número 1, puede agilizar la etapa de negociación al proponer salones que se saben tienen características deseables para la comunidad uniandina. Esto probablemente disminuiría la cantidad de salones propuestos antes de que los coordinadores académicos accedan a tomar ese salón.

Además, el número 1 puede aumentar la comodidad de los profesores y estudiantes pues con esta información, se pueden reservar salones cómodos y agradables para que no sean asignados para las clases y así estén disponibles para que la comunidad los pueda reservar a lo largo del semestre.

El número 2, cumple con los mismos objetivos de los dos párrafos anteriores.

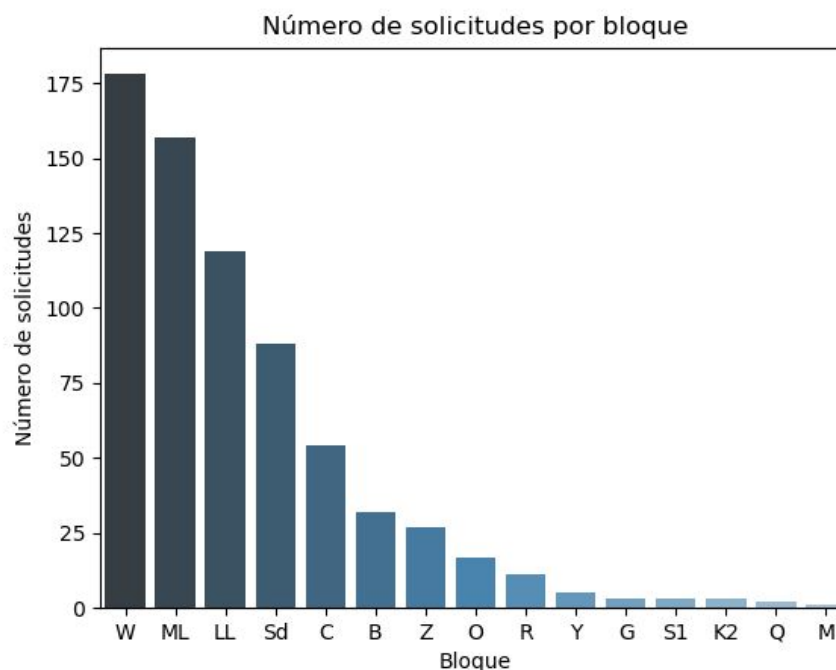
b.

Con esta información, un analista de negocio podría facilitar el proceso de negociación significativamente. Proveer a los coordinadores académicos salones que se sabe de antemano son preferidos por la comunidad uniandina aumenta las probabilidades de que acepten el salón en un menor número de propuestas.

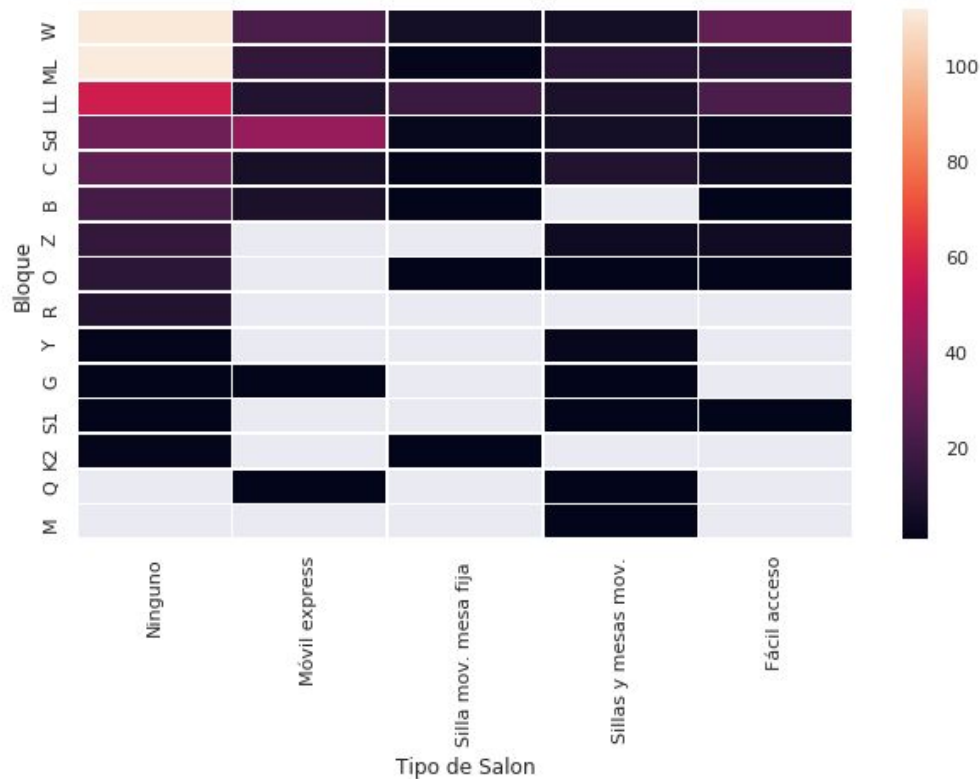
Así mismo, un analista en el negocio podría reservar un número pequeño de salones “favoritos” y facilitar que los estudiantes y los profesores tengan la oportunidad de reservarlos a lo largo del semestre. Así pues, aumentaría la comodidad de estos dos stakeholders.

c.

La visualización de los resultados se hizo utilizando la librería seaborn en python. El archivo donde fue implementado en análisis se llama cubito.xlsx y se encuentra adjunto en los documentos de entrega. Los resultados son los siguientes:



El bloque más deseado es el W seguido del ML, LL, etc. Por las razones descritas anteriormente, por las razones descritas en los dos puntos anteriores, esto es de altísima utilidad.



Aquí se evidencia que la características más deseadas (aparte de ninguna), son Móvil Express y Fácil Acceso. Es decir, para la comunidad uniandina, lo más deseable es que los salones sean de fácil acceso y tengan servicio de móviles express. Además, se evidencia que los bloques donde se encuentran estas características deseables son principalmente el Sd, el LL y el W. De nuevo, por las razones descritas en los dos puntos anteriores, esto es de altísima utilidad.

d.

La presentación se encuentra adjunta en el folder de la entrega.

El repositorio de respaldo se encuentra en: <https://github.com/mneira10/BIEntrega2.git>