

UNIVERSIDAD NACIONAL DE LA MATANZA

DEPARTAMENTO DE INGENIERÍA
E INVESTIGACIONES TECNOLÓGICAS

INGENIERIA EN INFORMATICA

BASE DE DATOS

Evaluación Grupal



2024

I. Objetivo

La Evaluación Grupal tiene como objetivo poder diseñar una base de datos, aplicando todos los conceptos aprendidos a lo largo de la materia. De esta forma, se diseñará un modelo de datos real, fomentando los debates e interacción grupal.

Los grupos elegirán el modelo de datos de su interés, según los casos propuestos.

II. Modelo de Datos propuesto

→ 1. Starbucks:

- **Fuente:** <https://www.kaggle.com/code/juniorbueno/starbucks-maps-location/input>
- **URL:** <https://www.starbucks.com.ar/>
- **Descripción del dataset:** El dataset consiste en las locaciones de todos los locales de la marca Starbucks.
- **Modelo de datos:** El modelo debe permitir registrar la venta online de productos de Starbucks, tal como se requiere en su página web, registrando los productos vendidos, los clientes que lo solicitan, desde qué sucursal se enviará, el destino del envío y todo lo que considere necesario para cumplir con el negocio a modelar.

→ 2. Películas de Netflix:

- **Fuente:** <https://www.kaggle.com/code/shivamb/netflix-shows-and-movies-exploratory-analysis>
- **URL:** <https://www.netflix.com/ar/>
- **Descripción del dataset:** El dataset contiene todas las películas de Netflix disponibles para ser elegidas.
- **Modelo de datos:** El modelo debe permitir almacenar la disponibilidad de películas que ofrece Netflix según sus distintas clasificaciones. Se registrarán sus abonados con sus tipos de suscripciones, las películas que hayan visualizado y todo lo que considere necesario para cumplir con el negocio a modelar.

→ 3. Casas en Venta o Alquiler:

- **Fuente:** <https://datos.gob.ar/dataset/ign-unidades-territoriales>
- **URL:** <https://www.argenprop.com/>
- **Descripción del dataset:** El dataset contiene los datos de las provincias, partidos y localidades de argentina.
- **Modelo de datos:** El modelo debe permitir poder registrar las propiedades para su venta, tal como lo realiza la página web. En este caso, las ventas no se concretan a través de la página, pero sólo los usuarios registrados pueden obtener el contacto de quien publica la propiedad. Se registrarán todos los datos necesarios que permitan cargar los datos y operar en la página web indicada, tal como las propiedades, los usuarios interesados y todo lo que considere necesario para cumplir con el negocio a modelar.

→ 4. Tienda de Anime:

- **Fuente:** <https://www.kaggle.com/code/fatihbilgin/story-of-anime/input>
- **URL:** <https://tiendaanime.com.ar/>
- **Descripción del dataset:** El dataset contiene los datos de los anime disponibles registrados para la tienda.
- **Modelo de datos:** El modelo debe permitir poder vender los anime ya sean nuevos o usados. Se deben almacenar todos los anime y luego sólo estarán a la venta aquellos que haya disponibilidad. Registrar los datos según la URL indicada y todo lo que considere necesario para cumplir con el negocio a modelar.

→ 5. Alquiler en Airbnb:

- **Fuente:** <https://www.kaggle.com/code/benyehuda/buenos-aires-airbnb/input>
- **URL:** <https://www.airbnb.com.ar/>
- **Descripción del dataset:** El dataset contiene los alojamientos cargados en Airbnb.
- **Modelo de datos:** El modelo debe permitir poder registrar las propiedades disponibles para alquiler, tal como realiza la página mencionada. Se registran las fechas disponibles de cada una de ellas, los precios según las fechas y las reservas ya efectuadas por los distintos usuarios. Se deben considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

→ 6. Denuncias en Defensa al consumidor:

- **Fuente:** <https://www.kaggle.com/datasets/agustinyanzn/denuncias-defensa-del-consumidor-20192020/data>
- **URL:** <https://www.argentina.gob.ar/servicio/iniciar-un-reclamo-ante-defensa-del-consumidor>
- **Descripción del dataset:** El dataset contiene algunas denuncias realizadas.
- **Modelo de datos:** El modelo debe permitir poder almacenar las denuncias que un consumidor quiera registrar. Además se realizará un seguimiento de las distintas citaciones que se mediará entre el consumidor y la entidad demandada. Se debe considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

→ 7. Venta de autos:

- **Fuente:** <https://www.kaggle.com/datasets/ahmettalhabektas/argentina-car-prices/data>
- **URL:** <https://www.kavak.com/ar>
- **Descripción del dataset:** El dataset contiene algunos vehículos ofrecidos para la venta.
- **Modelo de datos:** El modelo debe permitir poder registrar los vehículos que se quieran vender. En este caso, las ventas no se concretan a través de la página, pero sólo los usuarios registrados pueden obtener el contacto de quien publica el vehículo. Se debe considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

→ 8. Campeonato local de fútbol:

- **Fuente:** https://www.kaggle.com/datasets/mcarujo/south-america-football-season-2023?select=matches_primera-division-de-argentina-23.csv

- **URL:** <https://onefootball.com/en/competition/primer-division-de-argentina-183>
- **Descripción del dataset:** El dataset contiene algunos partidos disputados en el campeonato nacional de fútbol de primera división.
- **Modelo de datos:** El modelo debe permitir poder registrar los distintos partidos de fútbol de primera división en los distintos torneos locales. También se debe almacenar los distintos resultados y puntajes obtenidos en cada uno de los partidos disputados. Se gestionará un sistema de alertas personalizado, para los usuarios que decidan tener novedades de los resultados de su interés. Se debe considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

→ 9. Juegos Olímpicos:

- **Fuente:** <https://www.kaggle.com/datasets/umerhaddii/paris-olympic-medalists-2024>
<https://www.kaggle.com/datasets/mohamedamgad2002/2024-olympics-in-paris>
<https://www.kaggle.com/datasets/munsif2mu/olympics-2024-medals-table>
- **URL:** <https://olympics.com/es/paris-2024>
- **Descripción del dataset:** El dataset contiene el resultado de los juegos olímpicos 2024.
- **Modelo de datos:** El modelo debe permitir poder registrar las distintas disciplinas de los juegos olímpicos, los países concursantes, los equipos y jugadores y los resultados de cada certamen. Se gestionará un sistema de alertas personalizado, para los usuarios que decidan tener novedades de las distintas disciplinas, su pre-calificaciones/calificaciones que sean de su interés. Se debe considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

→ 10. Cine:

- **Fuente:** <https://www.kaggle.com/datasets/ashutoshdevpura/imdb-top-10000-movies-updated-august-2023>
- **URL:** <https://www.cinemarkhoyts.com.ar>
- **Descripción del dataset:** El dataset contiene un listado de películas.
- **Modelo de datos:** El modelo debe permitir poder registrar la venta de entradas para el cine a los usuarios, según la cartelera disponible. También permitirá visualizar los datos de cada una de las películas, tales como tiempo de duración, género, artistas y todos los datos que permitan cumplir con el modelo propuesto, tal como se realiza en la URL señalada.

→ 11. Reserva de hoteles:

- **Fuente:** <https://www.kaggle.com/datasets/mojtaba142/hotel-booking>
- **URL:** <https://www.booking.com>
- **Descripción del dataset:** El dataset contiene el resultado de reservas realizadas en hoteles.
- **Modelo de datos:** El modelo debe permitir poder registrar los alojamientos disponibles para su reserva, almacenando todos los datos necesarios para la registración de los alquileres y disponibilidad de los alojamientos, según la URL indicada. También se realiza una escala de puntos de los usuarios para ofrecerles bonificaciones. Se debe considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

→ 12. Clima:

- **Fuente:** <https://www.kaggle.com/datasets/nelgiriyeewithana/global-weather-repository>
- **URL:** <https://www.smn.gob.ar/>
- **Descripción del dataset:** El dataset contiene el resultado de las temperaturas observadas.

- **Modelo de datos:** El modelo debe permitir poder registrar las temperaturas pasadas, las actuales y aquellas que son predecidas de cada una de las regiones de Argentina. Verificar los datos necesarios, según la URL indicada. Se gestionará un sistema de alertas personalizado, para los usuarios que decidan tener novedades que sean de su interés. Se debe considerar el almacenamiento de todos los datos que se requieran para poder cumplir con el negocio a modelar.

III. Consignas Generales

- La evaluación se realizará en grupos de hasta **5** integrantes.
- La entrega debe contener un pdf con su carátula, enunciado y resolución de cada uno de los puntos.
- Los scripts utilizados en cada uno de los pasos deben estar adjuntos en la entrega.
- La entrega se realizará en un único archivo llamado *2024_Evaluacion_GrupoXX.zip*. Este archivo contendrá el pdf y todos los scripts desarrollados.
- Se aconseja realizar la evaluación en Microsoft SQL Server versión 2022.
- Cada uno de los scripts deben contemplar comentarios relevantes que se quieran detallar de las acciones realizadas. (Comentarios: /* */ ó --)
- Cada código debe encabezarse con un comentario que haga referencia al punto desarrollado

IV. Cronograma

Fecha	Actividad
25-October-2024	Entrega por Miel de la Evaluación Grupal.
28-October-2024	Exposición oral I de Evaluación por cada grupo, según horario pactado.
4-Noviembre-2024	Exposición oral II de Evaluación por cada grupo, según horario pactado.

V. Condiciones de aprobación

- La evaluación tendrá una nota grupal, pero sólo será efectiva por cada integrante si pudiera defenderla en la fecha de exposición oral.
- En el caso que algún integrante no sea consistente con sus respuestas, impidiendo demostrar su trabajo participativo en la evaluación, deberá realizar un recuperatorio convencional en forma escrita, en la fecha indicada en el cronograma general.

VI. Enunciado

Diseño

1. Elegir uno de los casos señalados en la Sección I. Tener en cuenta que sólo podrá elegirse un caso por grupo, sin poder repetirse.
2. Verificar el link mencionado con el set de datos propuesto. Cada caso tiene un dataset que se debe tener en cuenta para la carga de datos en el modelo diseñado. Será obligatorio en cada trabajo que ese dataset sea cargado, por lo cual se deben tomar como requerimiento inicial de los datos.
3. Verificar el link del modelo de negocio que se requiere diseñar. Cada caso se ejemplifica a través de una web comercial, para que puedan contemplar todos los actores que intervienen en cada modelo. El ejercicio desarrollado debe poder almacenar los datos para que funcione esa página web.
4. Escribir el Requerimiento inicial que se va a Modelar en modo descriptivo.
5. Diseñar un diagrama de Entidad Relación para el modelo de negocios elegido. Se debe contemplar el desarrollo gráfico de un DER. En este caso, se podrá elegir alguna aplicación que permita la diagramación del modelo.
 - a. Debe contener no menos de 12 entidades
 - b. Al menos una entidad debe ser débil
 - c. Al menos debe existir alguna relación de jerarquía
 - d. Al menos un tipo de atributo de los explicados en clase
 - e. alguna de las relaciones debe contemplar un almacenamiento histórico de datos
 - f. Se deben detallar todos los atributos en cada entidad
 - g. Las relaciones deben estar definidas con conjugaciones verbales
 - h. Los nombres de entidades deben ser definidos en singular
 - i. Tanto las relaciones como entidades deben contener nombres únicos y de aspecto descriptivo de su contenido
 - j. La lectura del DER será de izquierda a derecha y de arriba hacia abajo
 - k. Se debe evitar el entrecruzamiento de líneas, si fuera posible
6. Generar el Modelo Relacional del DER propuesto en el punto 5.
7. Verificar que el Modelo se encuentre al menos en 3FN. Si no lo está se deberá efectuar el proceso de normalización.

Desarrollo

8. **p_CrearDB(borrar_si_existe?):** Desarrollar un *procedimiento almacenado* que permita crear la base de datos con sus tablas, claves primarias, claves foráneas y todo lo relacionado al esquema de la base de datos. Según el parámetro enviado, se debe o no borrar la estructura si existía previamente.
9. **p_LimpiarDatos():** Desarrollar un *procedimiento almacenado* que permita limpiar todos los datos almacenados en la base de datos, sin borrar estructuras.

10. **p_CagarDataset**(archivo): Desarrollar un *procedimiento almacenado* que permita cargar el dataset propuesto en cada caso. Se podrá contemplar la carga en una tabla temporal para luego distribuir los datos en las distintas tablas. En el caso que el dataset tuviera datos incorrectos, no consistentes o faltantes, se podría implementar la estrategia de completarlo, corregirlo o suprimirlo. La carga se debe realizar desde el path del archivo enviado por parámetro.
11. **p_CargaAleatoria**(): Desarrollar un *procedimiento almacenado* que permita cargar el juego de datos que se utilizarán para la generación de filas en cada una de las entidades diseñadas en el modelo. Se deben controlar los errores de carga.
12. **tg_auditar**(): Elegir alguna de las tablas que contenga datos sensibles dentro de cada modelo. Sobre dicha tabla realizar un *trigger* de auditoría que permita controlar cada una de las acciones realizadas, permitiendo conocer la acción, el momento de la acción, el usuario que lo realizó, el valor cambiado y todo lo que considere relevante.
13. **f_funcion1**(valor): Realizar una función que permita darle formato frecuente a un campo determinado. Se deberá elegir la función, según lo que resulte útil para el modelo diseñado.
14. **f_funcion2**(contraseña): Realizar una función que permita controlar los caracteres permitidos y reglas para el seteo de contraseñas en la aplicación. Devolverá 1 si es válida ó 0 si no cumple las condiciones. Contemplar al menos 5 condiciones de diferente característica, por ejemplo, la longitud.
15. **v_vista1**() ... **v_vista3**: Diseñar 3 vistas que permitan listar 3 requerimientos funcionales que puedan presentarse en cada modelo de negocio. Las mismas deben:
 - tener complejidad media/alta, con el uso de joins, agrupaciones, subconsultas, etc.
 - Al menos una de ellas debe mostrar un ranking de los 10 “mejores”/“peores”/“mayores”/“menores”/etc, según el requerimiento elegido
 - Al menos una de ellas debe tener recursividad con jerarquías
 - Al menos una de ellas debe contener un análisis anual de algún requerimiento oportuno al modelo
16. **p_reporte1**() ... **p_reporte5**(): Diseñar 5 procedimientos almacenados que permitan listar 5 requerimientos funcionales que puedan presentarse en cada modelo de negocio. Los mismos deben contemplar las siguientes estructuras técnicas:
 - a. Al menos 2 de ellos deben recibir parámetros de entrada
 - b. Al menos 2 de ellos deben retornar datos de salida
 - c. Se deben validar que los parámetros de entrada contengan valores esperados
 - d. Se deben contemplar el uso de todos los operadores vistos en clase:
 - Joins
 - Agrupaciones y funciones de agrupación
 - Union / Union all
 - All / any / some
 - Subconsultas
 - exists / in
 - with
17. **p_proceso1**(datos_del_usuario): Diseñar un procedimiento almacenado que permita realizar el alta/baja/modificación de un nuevo usuario en nuestro sistema, contemplando diferentes características, tales como que el usuario no exista/exista, que cumpla con las

condiciones de contraseñas, que haya completado todos los datos mínimos necesarios, las relaciones que se afectarían, etc.

VII. Formato de Entrega

Se entregará un único archivo llamado *2024_Evaluacion_GrupoXX.zip* con el siguiente contenido:

- Archivo *2024_Evaluacion_GrupoXX.pdf* con todo el contenido del TP. Este archivo no debe contener el código de los scripts, solo su mención y organización de los puntos resueltos.
- Archivo *2024_Evaluacion_GrupoXX_DDL.sql* con cada uno de los scripts utilizados para el diseño (puntos 8, 9, 10, 11)
- Archivo *2024_Evaluacion_GrupoXX_DML.sql* con cada uno de los scripts utilizados para el desarrollo (puntos 12, 13, 14, 15, 16, 17)