



Proyecto No. 1 **Desarrollo y Consulta de Base de Datos** **(Parejas)**

Modalidad y fecha de entrega

- El proyecto se hará en forma individual y ser enviado antes de la fecha límite de entrega:
- No se permitirá la entrega o envío de proyecto más allá de la fecha límite.

Descripción general del proyecto

El proyecto contempla el uso de tecnología de Base de Datos para la creación, modelación y carga de modelos datos, con el objetivo de utilizar lenguaje SQL para investigación, desarrollo y presentación de resultados sobre preguntas de negocio para apoyo en toma de decisiones.

El conjunto de datos a utilizar serán los archivos formato csv que pueden descargarse de la dirección web: https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp. Estos datos contemplan información histórica de vuelos comerciales que datan del año 1987 al 2018. Para fines del proyecto, utilizaremos datos históricos del año 2017 a 2018 (los más recientes).

Los pasos sugeridos para el inicio correcto del proyecto son:

- **Descarga de archivos csv a repositorio local (elaborado por TA)**
- **Instalación de servicio Postgres o SQLite**
- Entendimiento de Modelo de Datos (tablas, atributos, relaciones) diseñada sobre los datos históricos descargados
- **Creación de Modelo de Datos y carga de los datos crudos en manejador de Base de Datos.**

Una vez creado el modelo de datos y cargado con la data necesaria, se recomienda la interacción por medio de interfaces SQL hacia la base de datos creada con el objetivo de familiarizarse con la información de la misma.

Las preguntas que se deben de contestar son todas aquellas que puedan apoyar a la toma la siguiente decisión: **¿qué aerolínea se desempeñó mejor durante el 2017-2018?**

Para contestar esta pregunta, la tarea es, basado en el entendimiento del modelo de datos, cuál información de índices o métricas podría ser base para justificar una toma de decisión de este tipo. Algunas preguntas que pueden plantearse para ayudar a la toma de decisión son:

- ¿Qué aerolíneas ejecutan más?
- ¿Cuál aerolínea presentan más tiempo de retraso?

- ¿Qué aerolíneas tienen una menor proporción de tiempo de retraso con respecto a la cantidad de sus vuelos?
- ¿Qué aerolíneas recorren mayores distancias?
- ¿Qué aerolíneas han aparecido o desaparecido durante el 2017-2018?
- ¿Qué aerolíneas están creciendo en operaciones o decreciendo en operaciones?
- ¿Cuál es la causa principal de atrasos para los vuelos entre las aerolíneas?
- etc.

Especificando las posibles preguntas a contestar para apoyar a la toma de decisión, se deben de traducir estas preguntas a sentencias SQL que nos brinden el resultado deseado. Con estos resultados (que generalmente se denominan KPIs) se debe de justificar la inversión en una aerolínea u en otra. Todas sus conclusiones deben de estar basados en un resultado de alguna sentencia SQL.

Requerimientos mínimos a completar:

- Se debe presentar el resultado de al menos 10 queries
- Deben presentarse al menos tres queries con agrupación (GROUP BY)
- Deben presentarse al menos tres queries con JOINS entre dos o más tablas
- Debe presentarse al menos una consulta que haga uso de subqueries

Especificación de tecnología

- Sistema manejador de base de datos: Postgres o SQLite
- Interfaz de Interacción con BD: Jupyter Notebook, DBeaver o similar
- Solución: PDF

Temas a reforzar

- Lenguaje SQL: DDL / DML
- Postgres o SQLite
- Aplicación de Base de Datos hacia el negocio.

Documentos a entregar

- Archivo comprimido con:
 - Documentación de modelo de datos correspondiente a la Base de Datos utilizada.
 - Se debe generar un diagrama E/R de la base de datos utilizada
 - Jupyter notebook o similar con las consultas SQL realizadas para la solución a la pregunta de negocio planteada.
 - Documento PDF con la respuesta a la pregunta de negocio planteada, incluyendo toda la base que justifica la solución.



VI. Evaluación

- Diseño de preguntas ejecutivas: **20 puntos**
- Diseño de consultas e interacción con base de datos: **60 puntos**
- Análisis de resultados: **10 puntos**
- Presentación de solución a pregunta de negocio **10 puntos**

Total: 100 puntos

Puntos extras:

- Presentación de resultados web: 10 puntos
- Presentación de gráficas de análisis: 10 puntos