**Curso Ing. de Datos – Rocking Data – Santiago Sosa**

Primero que nada, quiero agradecerles por haberme elegido para participar de este curso, aprendí mucho y me dejó con ganas de seguir incorporando conocimientos, también les agradezco por la amabilidad y la predisposición para responder preguntas con mucha paciencia.

Para la resolución del proyecto, seguí los siguientes pasos:

1. Creé un jupyter notebook llamado “**Proyecto Final - Ing Datos - Santiago Sosa.ipynb**”. Cargué el dataset, realicé un EDA y transformé la información para que se adaptara al nuevo esquema.
2. Creé un esquema estrella utilizando la herramienta diagram, el resultado se puede ver en la imagen “**Esquema.png**”. Elegí crear 4 tablas para el datawarehouse, comprendiendo que eran 3 las dimensiones necesarias para la base de datos (podrían ser más, pero decidí no crear tablas que no fueran 100% necesarias)

Diagrama

Descripción generada automáticamente

1. Creé una query SQL “**db\_creation.sql**” donde armé en SQL los comandos para crear la estructura planteada en el esquema “Esquema.png”.
2. Nuevamente sobre el jupyter notebook creado en el punto 1, cargué las tablas de cities\_states y soc con la información necesaria previamente transformada para adaptarse al esquema. Para la tabla cities\_states utilicé un dataset de un censo de estados unidos, para enriquecer la base con información de las ciudades y a que estados corresponden.
3. Creé un nuevo notebook para el ETL, utilizando el código de ejemplo visto en el curso, logré crear la función extract, pero no comprendí como continuar con las funciones transform y load. Esto es porque no sabía donde tomar la información Raw: en el ejemplo se toma de la base de datos con el esquema ya cargado y en este caso la información raw está en un dataset y no separado por tablas. Me parece que no logré entender esta parte y me gustaría, en lo posible, saber cómo terminarlo.