

Generando el Data Set Final para el análisis de la encuesta

A partir de la base de datos proporcionada con las respuestas de la encuesta sobre tu educación durante el tiempo de pandemia, desarrollar una aplicación ETL en Python que realice lo siguiente.

Extracción

1. Conectar con la base de datos ya sea de manera local o remota.
2. Consultar todas las respuestas, permitiendo manejar los siguientes filtros.
 - a. Todas las carreras
 - b. Ingeniería en Desarrollo y Gestión de Software
 - c. Ingeniería en Redes y Ciberseguridad
 - d. Ingeniería en Entornos Virtuales y Negocios Digitales

Transformación

A partir de los datos obtenidos en la etapa anterior realizar las transformaciones necesarias para obtener un data set con las siguientes columnas.

ID: Deberá ser una clave única para cada respuesta.

Pregunta1: Nombre del estado con nombre no id.

Pregunta2: Nombre del municipio con nombre no id.

Pregunta3: Verificar que la fecha esté en el formato correcto y que sea una fecha válida, que al menos tenga 18 años.

Edad: Calculado a partir de la fecha de nacimiento a la fecha actual.

Pregunta4: Sexo con nombre o carácter F o M

Pregunta5: Carrera: Con nombre

Pregunta6: Número de cuatrimestre

Pregunta7: Número de cuatrimestre

Pregunta8:

Si

No

Compartido

Pregunta9:

Si

No

Pregunta10:

50 a 100

101 a 300

301 a 500

501 o más

Pregunta11:

Computadora

Tablet

Celular

Televisión

Pregunta12:

Si

No

Pregunta13:

Correo electrónico,

Plataformas como classroom, teams, etc.

Grupos de WhatsApp

Ninguno

Pregunta14:

Mucho

Regular

Poco

Muy poco

Pregunta15:

Examen oral

Examen en línea

Prueba escrita abierta

Elaboración de trabajos, proyectos y casos

Pregunta16:

Examen oral

Examen en línea

Prueba escrita abierta

Elaboración de trabajos, proyectos y casos

Pregunta17:

Examen oral

Examen en línea

Prueba escrita abierta

Elaboración de trabajos, proyectos y casos

Pregunta18:

Problemas de conexión

Dificultad para preguntar dudas durante el examen

Nervios adicionales por la incertidumbre ante la evaluación online

Excesiva carga de trabajo

Exámenes escritos a resolver en tiempo reducido

Coincidencia de entregar de trabajos de distintas asignaturas

Pregunta19:

Mucho

Regular

Poco

Muy poco

Pregunta20:

Mucho

Regular

Poco

Muy poco

Pregunta21:

Mucho

Regular

Poco

Muy poco

Pregunta22:

De 3 a 4 horas

De 5 a 6 horas

De 6 a 7 horas

Más de 7 horas

Pregunta23:

Más que si hubiera ido a clase

Igual que si hubiera ido a clase

Menos que si hubiera ido a clase

Pregunta24:

Mucho

Regular

Poco

Muy poco.

Pregunta25:

Mucho

Regular

Poco

Muy poco

Pregunta26:

Mucho

Regular

Poco

Muy poco

Pregunta27:

Mucho

Regular

Poco

Muy poco

Nada

Pregunta28:

Estudio

Trabajo

Apoyo en casa

Cuidado de menores/mayores

Otros

Pregunta29:

Si

No

Pregunta30:

Texto

Tokens:

Palabras principales (Tokens) encontradas en la pregunta 30 eliminando los stopwords para español

En cada columna revisar que no existan datos faltantes y en dado caso reemplazar por el valor mas adecuado en cada caso.

Carga

Almacenar el nuevo data set en un formato adecuado para su análisis posterior, CSV, JSON, EXCEL, etc.