**EXAMEN 2do PARCIAL INTELIGENCIA DE NEGOCIOS**

**FECHA:** 27 julio 2022

**Hora:** 9H:00

**Duración:** 2H:00

**Profesor:** MsC. Fernando Garrido S.

**Email:** [jfgarridos@utn.edu.ec](mailto:jfgarridos@utn.edu.ec)

**Modalidad:** Presencial (Laboratorio 5)

**Herramientas a utilizar:** Python, Microsoft Power BI

**Indicaciones:**

1.Examen es personal, el estudiante puede utilizar su equipo y recursos para resolver los problemas propuestos.

2. Subir al aula virtual todo lo desarrollado en un archivo ApellidoNombre.zip

**CUESTIONARIO**

**PREGUNTA 1:** Realizar el estudio utilizando **Python** para realizar el proceso **E**xtract-**T**ransform, puede ser utilizando **Google** **Colaboratory** o **Anaconda Jupyter NoteBook** (guardar como: apellidonombre.ipynb) y **PowerBi** para el proceso **L**oad (apellidonombre.ipbx).

**ETL (Extraer, Transformar, Cargar).**

Para este caso de estudio se utilizará datos del Banco Mundial. Los datos provienen de dos fuentes (archive.zip):

* Datos de indicadores del Banco Mundial: estos datos contienen indicadores socioeconómicos de países de todo el mundo. Algunos indicadores de ejemplo incluyen la población, la tierra cultivable y la deuda del gobierno central.
* Datos de proyectos del Banco Mundial: este conjunto de datos contiene información sobre los préstamos para proyectos del Banco Mundial desde 1947.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. **Ejercicio (2 ptos)**

En Python Extraer datos de diferentes fuentes:

* archivos csv (electricity\_access\_percent.csv, projects\_data.csv, rural\_population\_percent.csv, gdp\_data.csv, mystery.csv)
* archivo json (population\_data.json)

1. **Ejercicio (5 ptos)**

Transformar datos en Python.

* **combinar datos de diferentes fuentes:** usando el método *pandas concat*. En otras palabras, encuentre la unión de los dos conjuntos de datos. (rural\_population\_percent.csv, electricity\_access\_percent.csv)
* **limpieza de datos:** cada conjunto de datos puede tener sus propios problemas, ya sea que se trate de valores faltantes, entradas duplicadas, errores de ingreso de datos, etc. Por ejemplo, los datos de proyectos (projects\_data.csv) y los datos de indicadores de población (population\_data.csv) tienen diferentes valores para los nombres de países. La columna 'Country' tiene valores NaN. Como se hizo en clase algunos datos deben estar en valores numéricos, realizar esa transformación.
* **tipos de datos:** Pandas tiene algunos métodos diferentes para convertir entre tipos de datos, y aquí hay un enlace a la documentación:
* [astype](https://pandas.pydata.org/pandas-docs/version/0.22/generated/pandas.DataFrame.astype.html#pandas.DataFrame.astype)
* [to\_datetime](https://pandas.pydata.org/pandas-docs/version/0.22/generated/pandas.to_datetime.html#pandas.to_datetime)
* [to\_numeric](https://pandas.pydata.org/pandas-docs/version/0.22/generated/pandas.to_numeric.html#pandas.to_numeric)
* [to\_timedelta](https://pandas.pydata.org/pandas-docs/version/0.22/generated/pandas.to_timedelta.html#pandas.to_timedelta)
* **análisis de fechas:** en este caso, eso significa tomar una fecha en el formato de una cadena y transformar la cadena en un tipo de fecha. Use to\_datetime.

Como ahora la fecha *boardapprovaldate*, *closingdates* están en formato de fecha y hora, cree algunas columnas nuevas en el marco de datos *df\_projects*: approvalyear

* + approvalday
  + approvalweekday
  + closingyear
  + closingday
  + closingweekday

por ejemplo, use:

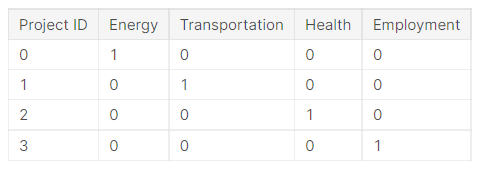
df\_projects['approvalyear'] = df\_projects['boardapprovaldate'].dt.year

* **datos perdidos:** cuando un conjunto de datos tiene valores faltantes, puede eliminar esos valores o completarlos, como se realizó en clase.
* **datos duplicados:** un conjunto de datos puede tener datos duplicados, por ejemplo, de gdp\_data, cuente el número de países que han tenido un total de proyectos (totalamt) superior a mil millones de dólares (1.000.000.000). Para obtener el recuento, deberá eliminar las filas de datos duplicados.
* **variables ficticias:** La idea es transformar datos categóricos

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

en nuevas características que se vean así:



Esto se hizo en clase.

1. **Ejercicio (3 ptos)**

Carga en PowerBI.

Visualizar los datos en Dashboards. Queda a criterio de cada estudiante crear su Dashboard con los datos tratados en Python, los data frames deben ser subidos con script en Python, se debe considerar que hay relaciones entre las tablas, estructurar antes el modelo de datos.

**¡SUERTE ¡**

**ÁREA PARA AÑADIR IMÁGENES DE LAS SOLUCIONES A LOS EJERCICIOS**