Title:

Social and economic worldwide indicators.

Question Ideas:

- Which is the most populated country?
- Is x country growing economically?
- Are these facts related to the population's physical conditions?

Team Members:

Nicholas, Robbie, Liam, Carlos

Datasets to be used:

- https://www.kaggle.com/datasets/sudalairajkumar/undata-country-profiles?resource=download
- https://www.kaggle.com/datasets/majyhain/height-of-male-and-female-by-country-2022

Project Description:
The datasets cover ten economic and social indicators from 200 countries worldwide. The primary purpose of this project is to study the relationship between social and economic indicators and if they affect to population's physical conditions.

Breakdown of tasks:
Person 1: Citing and extracting the data from the sources.
Person 2: Transforming the data.
Person 3: Loading data into a database.
Person 4: Creating a report with specific documentation.

Document Key:

Titles	Title
Intructions	Grey text is from the Project 2 Detail Page on the Dashboard
Content	Compile standard text together to make up the reports content
Highlight	Report Requirements (For the Report) | Supporting Materials

Report Requirements

Once you've identified your datasets, you'll perform the ETL process and create the documentation, which must include:

The datasets that you used and their sources.

Data Source Citation

Srk (2017) Download (43 kB) Country Statistics UNData | Overview. This dataset contains key statistical indicators covering General Information Economic Indicators Social Indicators Environmental & Infrastructure Indicators. www.kaggle.com/datasets/sudalairajkumar/undata-country-profiles

Majyhain (2022) Download (4 kB) Height of Male and Female by Country | Overview This dataset contains tallest to smallest average heights of males and females by country covering rank, country name, male height and female height. www.kaggle.com/datasets/majyhain/height-of-male-and-female-by-country-2022

WHO (2014). Download (6.7 MB) Global status report on noncommunicable diseases | Overview The report provides data on the current situation, identifying bottlenecks as well as opportunities and priority actions for attaining global targets. www.who.int/publications-detail-redirect/9789241564854

The types of data wrangling that you performed (such as cleaning, joining, filtering, and aggregating).

Data Wrangling

basic cleaning was done in excel including removing columns that contained irrelevant data.
Transforming the data (such as cleaning, joining, filtering, and aggregating it)

The schemata that you used in the final production database.

Entity Relationship Diagram (ERD)

Project Specialisation Requirement

You should focus your efforts on a specific industry. To help you choose, the following subsections provide examples of how ETL gets used in various possible specialisations.

Here's an example of using ETL in the healthcare industry:
An analyst working at a major hospital is tasked with reviewing their policies regarding the upcoming flu season. The analyst keeps the following questions in mind: How many patients does the hospital expect this year? How severe will the flu season be this year? Will regional differences or similarities occur??

Project Specialisation Brief

In UWA's anthropology department, the lead analyst wants to collect data from different sources in order to compile a database they can use to analyse and explore the social and economic relationships between economies, average heights and the body mass index's of populations around the world.

The dataset produced covers ten economic and social indicators from 200 countries worldwide. The primary purpose of the dataset is to enable study into the relationship between economic and population health metrics.

Internal Milestones

Project ideation	Testing
Data fetching/API integration	Creating documentation
Data analysis	Creating the presentation

Project Proposal

Before you start writing any code, your group should outline the scope and purpose of your project. This will help provide direction and safeguard against scope creep.

The proposal is essentially a brief summary of your interests and intent. Be sure to include the following details:

> The kind of data you'd like to work with and the field you're interested in (finance, healthcare surveys, etc.)

ETL - Global Health and Economic Metrics

> The questions you'll ask of the data

- Which is the most populated country?
- Is x country growing economically?
- Are these facts related to the population's physical conditions?

> Possible source for the data

www.kaggle.com/datasets

Aim / Scope

Use the following example for guidance:
The aim of our project is to uncover patterns in credit card fraud. We'll examine relationships between transaction types and location, purchase prices and times of day, purchase trends over the course of a year, and other related relationships derived from the data.

Important to Note

Whenever you use a dataset or create a new dataset based on other sources (such as existing datasets or information scraped from websites), make sure to use the following guidelines:

Check for copyright protections, and make sure that the way you plan to use this dataset is within the bounds of fair use.

Document how you intend to use this dataset now and in the future. Find any licences or terms of use associated with the dataset, and review them to confirm that your intended use is in compliance.

Investigate how the dataset was collected. Identify any indicators that the data was obtained from a source that the compilers were not authorised to access.

You'll likely have to adjust your project plan as you explore the available data. That's okay! This is all part of the process. Just make sure that everyone in the group is aligned on the project's goals as you make changes.

Make sure that your datasets are not too large for your personal computer. Big datasets are difficult to manage locally, so consider using data subsets or different datasets altogether.

All Instructions

For Project 2, you'll work with your team to complete the ETL process on a dataset.

Data Cleanup and Analysis Requirements
Your team will be responsible for:

Citing the data sources.

Extracting the data from those sources.

Transforming the data (such as cleaning, joining, filtering, and aggregating it)

Loading the data into a database (either relational or nonrelational)

Report Requirements
Once you've identified your datasets, you'll perform the ETL process and create the documentation, which must include:

The datasets that you used and their sources.

The types of data wrangling that you performed (such as cleaning, joining, filtering, and aggregating).

The schemata that you used in the final production database.

Project Specialisation Requirement
You should focus your efforts on a specific industry. To help you choose, the following subsections provide examples of how ETL gets used in various possible specialisations.

ETL and Finance
Here’s an example of using ETL in the financial industry:

Current treasury benchmarks are at an all-time low. So, a financial analyst is tasked with studying the last 30 years' worth of rates.

After pulling the historical data, the analyst cleans and explores the data to perform an analysis, with the intent of predicting future benchmark trends.

Once the analyst has extracted, transformed, and loaded the historical data into a database, they turn their attention to the present-day data. Using an API, they pull the most up-to-date information so that it can be added to the company’s established database.

The analyst has now extracted the new data but needs to ensure that it’s correctly formatted before loading it into the existing database. Then, once the data is transformed, the analyst can load it into the database and continue with the analysis.

ETL and Healthcare
Here’s an example of using ETL in the healthcare industry:

An analyst working at a major hospital is tasked with reviewing their policies regarding the upcoming flu season. The analyst keeps the following questions in mind: How many patients does the hospital expect this year? How severe will the flu season be this year? Will regional differences or similarities occur??

The analyst wants to collect and analyse data from different sources so that they can make predictions about the upcoming flu season.

Before combining the hospital’s own data with the externally acquired regional data, the analyst needs to extract the new data, transform it to match the existing data, and then load it into the database.

ETL and Other Industries
Here are examples of using ETL in the other industries:

In the marketing department, an analyst gets data from the company's competitors to find out how their products compare. They need to extract, transform, and load multiple data sources into a common database before doing an analysis.

An analyst working for a large retail chain is in charge of moving a legacy database into a cloud-based data warehouse.

An entrepreneur has a grand business idea but wants to get a feel for how it will be received. So, they use web scraping and APIs to pull data from various social media platforms with the intent of analysing consumer reactions.

Rubric
Project 2 Rubric (Links to an external site.)

Submission
To submit your project, click Submit, and then provide the URL of your GitHub repository for grading.

NOTE
Projects are requirements for graduation. While you are allowed to miss up to two Challenge assignments and still earn your certificate, projects cannot be missed.

Working with Your Group
When working on an online group project, it's crucial to meet with your group and communicate regularly. Plan for significant collaboration time outside of class. The following tips can help you make the most of your time:

Decide how you're going to communicate with your group members when you begin. Create a Slack channel, exchange phone numbers, and ensure that the group knows each group member's available working hours.

Set up an agile project by using GitHub Projects (Links to an external site.) so that your group can track tasks.

Create internal milestones to ensure that your group is on track. Set due dates for these milestones so that you have a timeline for completing the project. Some of these milestones might include:

Project ideation

Data fetching/API integration

Data analysis

Testing

Creating documentation

Creating the presentation

Because this is a one-week project, make sure that you've completed at least half of the project by the middle of the week to stay on track.

Although you will divide the work among the group members, it's essential to collaborate and communicate while working on different parts of the project. Be sure to check in with your teammates regularly and offer support.

Support and Resources
Your instructional team will provide support during classes and office hours. You will also have access to learning assistants and tutors to help you with topics as needed. Make sure to take advantage of these resources as you collaborate with your group on this project.

Project Guidelines
The following project guidelines focus on teamwork, your project proposal, data sources, and data cleanup and analysis.

Collaborating with Your Team
Remember that projects are a group effort. Working closely with your teammates will benefit the outcome of your project AND help you in your future careers. You'll learn collaborative workflows that will enable you to approach and solve complex problems. In other words, working in groups allows you to work smart and dream big. Take advantage!

Project Proposal
Before you start writing any code, your group should outline the scope and purpose of your project. This will help provide direction and safeguard against scope creep.

The proposal is essentially a brief summary of your interests and intent. Be sure to include the following details:

The kind of data you'd like to work with and the field you're interested in (finance, healthcare surveys, etc.)

The questions you'll ask of the data

Possible source for the data

Use the following example for guidance:

The aim of our project is to uncover patterns in credit card fraud. We'll examine relationships between transaction types and location, purchase prices and times of day, purchase trends over the course of a year, and other related relationships derived from the data.

Finding Data
Once your group has written a proposal, it's time to start searching for data. We recommend the following curated sources of high-quality data:

IMPORTANT
You must have at least two data sources.

data.world
Kaggle
Data.gov.au
Awesome Public Datasets
Public-APIs
Awesome API

IMPORTANT
Whenever you use a dataset or create a new dataset based on other sources (such as existing datasets or information scraped from websites), make sure to use the following guidelines:

Check for copyright protections, and make sure that the way you plan to use this dataset is within the bounds of fair use.

Document how you intend to use this dataset now and in the future. Find any licences or terms of use associated with the dataset, and review them to confirm that your intended use is in compliance.

Investigate how the dataset was collected. Identify any indicators that the data was obtained from a source that the compilers were not authorised to access.

You'll likely have to adjust your project plan as you explore the available data. That's okay! This is all part of the process. Just make sure that everyone in the group is aligned on the project's goals as you make changes.

Make sure that your datasets are not too large for your personal computer. Big datasets are difficult to manage locally, so consider using data subsets or different datasets altogether.