To: michaelsean@me.com

From: Mabel Alamu, Arthur An, Dainty Cuevas, Sonny Maz

Hi Sean,

Here is a quick summary of the ETL Project you requested. As of Wednesday, Jan. 27, our team finished the project within the timeframe we originally planned, even though the allocated time spent on each step varied from what we estimated. Please see below for further details.

**Purpose:**

The purpose of this project was to demonstrate the process of extracting and transforming data from different sources with varying data formats (csv, json, etc.). Then, loading transformed data to a target database system, in this case, Postgres.

**Process:**

We sourced our alcohol substance abuse data from Kaggle, then scraped per capita GDP data from Wikipedia.

Although we finished the project in time, we spent more time cleaning/transforming data than we had planned.

* First, we found that the population data contained more information than we needed and decided to focus only on one year (2018) and one data point (aggregated alcohol abuse for both genders).
* Second, we then tried to group the country and average all the GDP per capita for each country. However, we found that there were non-numeric values in the GDP per capita column, so we had to drop those non-numeric values using their indexes.
* Finally, we joined the two data frames on the 'country' column.

When we tried to load the data into the target databases, we got a system error. We found that the data type assigned in pandas were different from the data types in Postgres. We proceeded to change the data types in pandas, so we could successfully load the data into Postgres.

**Results**

Overall, we successfully extracted, transformed, and loaded as planned.

You can find the project results through the following link:

<https://github.com/daintycuevas/ETL-project.git>.

Thank you!

The Worldly Company

worldlycompany.org