ETL Exercise

The ETL Process will be used to carry out the following exercise in Data Analysis.

We have a set of customer sales data that we want to extract information from.

For each customer, we want to know:

- How much each customer spent?
- What was their average spend?
- How many of each item did a customer buy?

Write a python script that executes the below steps.

Extract

1. Extract all the data from the sales_data.csv. The columns for the csv are `customer_id`, `purchase_date`, `purchase_amount` and `product_id`.
2. Clean that data (minimum requirement is to remove any rows that contain null cells).

Transform

1. Transform that data, calculating the following results: a. Each customer's total spend b. Each customer's average spend c. How many times each customer has purchased a specific item
2. Calculate this data for the period 1 December 2020 - 5 December 2020

Load

1. Load the results into a database. You can choose how to structure the table(s) you load the data into.

Analyse

1. What does the data in the tables tell you about the different customers purchasing habits?