# **Analytics Engineering Data Challenge**

You have 3 different files, each containing different information:

* user information,
* transactional information,
* package delivery information.

Based on these files, do the following:

1. Explore the three files and, based on your findings, write a data dictionary for each file and a UML diagram.

**Hint: Think of all the areas that can help others better understand the data**

1. Create a merged file that consolidates the information. Make sure that:
   * Date formats are standardized to YYYY-MM-DD,
   * Timestamp fields have the Mexico City time zone,
   * Numeric fields have two decimal places,
   * Text strings do not exceed 256 characters.

**Please write comments in the script explaining each code statement.**

1. Design a DAG for the production deployment of your merged table creation. Based on that, answer the following questions:
   * Would you choose ETL or ELT? Why?
   * Which part of the process would you emphasize most? Why?
2. Write SQL queries to answer the following:
   * How many users are there in total?
   * How many users have transacted?
   * What is the card delivery rate?
   * Which is the most efficient package courier?
   * Which are the top 10 categories with the most transactions?

**Hints:**

1. **Check data quality**
2. **There are different ways to measure efficiency with the information provided, think of one that can best measure impact.**
3. **Categories refers to transaction\_type**