Test assignment for Data engineer intern

Part 1: Aggregated Monthly Payments (SQL)

The code for the task can be found here:

https://app.snowflake.com/europe-west4.gcp/bp79014/w1POMS8BCgaF#query

Part 2: Weekend Data Processing (Python)

I used Google colab for the task; you can download and run it locally as well

https://colab.research.google.com/drive/1IorR-32wbEeLUCXkdbextiagZ2p9IZFa?usp=sharing

Part 3: Questions

What are the major differences between ETL and ELT approaches to data transformation?

The primary distnction between ETL and ELT is the data transformation order. In ETL, data is transformed before being loaded into the target system; while in ELT, data is loaded into the target system first and then transformed. In other words, ETL transforms data on a separate processing server, while ELT transformations occur within the data warehouse.

- Loading data from multiple sources process is faster on ELT
- With ELT, data might be transformed while loading
- ELT offers more opportunities for raw data queries
- ETL typically works with structured data; ELT might contain unstructured
- ETL is good for small datasets with complex transformations
- ELT is good for bigger datasets that need to be processed faster

What are the major differences between data analytics and data engineering?

Data engineers aim to build data infrastructure and a functional environment useful for data analysts to extract patterns and meanings from data. Data analytics and data engineering have different objectives and roles in a project. For a business, data engineers design, develop, and maintain the data architecture. The architecture is utilized by data analysts when working with data; they make calculations, visualisations, and statistical analysis.

Is there any specific software/tool/technology you are really interested in working with?

In general, I have a lot of curiosity about discovering new software and learning new tools. For instance, I learned SQL using BigQuery, however answering the first question, I switched to Snowflake and had a great time getting to know this new tool. Since I am just starting my career in data engineering, I would adore to work with anything that the Inbank team uses.