Case Study

1. Write an ETL job in python:

Create an ETL job that pulls the temperature from your city (using a public weather data api) and loads it to a daily temperature table.

Your script needs to be idempotent and easy to maintain. Your target table (to which you are loading) should always be available, and in case of load failure, the target table should still hold the old data.

The SQL part can be mock code as long as it's well explained.

2. Using http://sqlfiddle.com/, please tables and SQL code to answer the following questions using the attached csv file.

    1. How many % of salespeople sold at least one item?
    2. Out of all the people that sold something, how high is the average discount that a sales person gives?
    3. Out of all orders, how high is the average discount given?
    4. What other questions do you think would make sense to answer with this data and how should the reporting table for this look like?
    5. Create a table/view for Data Analysts that contains the following information:

- sales_person_id
- sales_person_name
- first_sales_date
- last_sales_date
- avg_count_order_per_month

avg_order_value

avg_discount

6. which other information do you think would make sense to add into this table?