

Kendall Stimart

DS 2002 Documentation

October 27, 2025

In SQL, I created a new database called `adventureworks_dw`, which I used as my data warehouse for the project. In this database, I built four tables: `fact_sales`, `dim_product`, `dim_customer`, and `dim_date`. I populated these tables using data from the original AdventureWorks database, following a similar process to what we practiced in class and on the homework. The `fact_sales` table contained all of the business data, while the remaining three tables described each of the sales. The `dim_product` stored product names, product numbers, and list prices. The `dim_customer` contained territory ID, account number, and customer type. Lastly, the `dim_date` organized order dates into year, month, and day.

After setting up the data warehouse in SQL, I used Python to connect to my MySQL database. I used the connection setup from the top half of Lab 4 to ensure that the connection worked properly. Using Pandas, I extracted each of the four tables into separate dataframes and then merged them together. First, I merged `fact_sales` with `dim_product`, then joined the `dim_customer` and `dim_date`. To incorporate an additional data source, I created a CSV file in Excel named `dim_salesperson.csv`. This file contained a list of 5 salespeople, along with their salesperson ID, first name, and last name. I imported this file into Python and merged it with my existing combined dataset. Once I joined the CSV file with the combined dataset, I renamed the table `fact_sales_final`. Then, I ran `fact_sales_final` into SQL. Finally, I ran 3 different queries to ensure the data ran correctly.