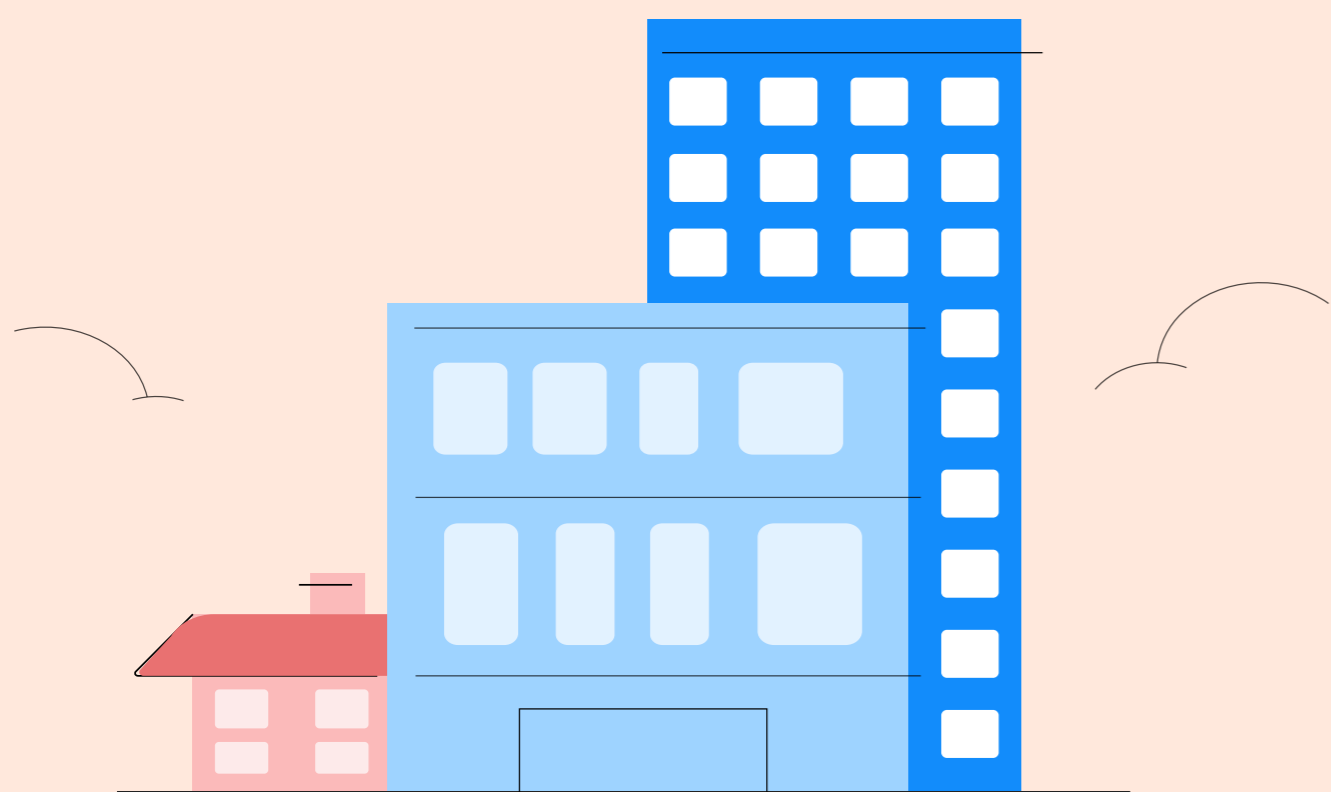


Deconstructing the ETL process



Let us deconstruct how ETL process helps an organization analyse its sales data through a simple process of bringing data from two of its stores - one in the US and the other in India - and preparing it for downstream analytics reporting.

Phase 1 - Extract

In this phase, we extract sales data from New York and Mumbai stores for the month of February 2023.

Transaction ID	Date	Product Type	Currency	Final Amount incl. Tax
NEW-001	2024-02-01	Accessory	USD	105.55
NEW-002	2024-02-02	Accessory	USD	115.54
NEW-003	2024-02-03	Smartphone	USD	2262.97
NEW-004	2024-02-04	Smartphone	USD	711.12
NEW-005	2024-02-05	Accessory	USD	160.74
NEW-006	2024-02-06	Accessory	USD	210.05

Transaction ID	Date	Product Type	Currency	Final Amount incl. Tax
MUM-001	01-02-2024	Accessory	INR	10446.96
MUM-002	02-02-2024	Smartphone	INR	71927.7
MUM-003	03-02-2024	Accessory	INR	15038.58
MUM-004	04-02-2024	Accessory	INR	4453.57
MUM-005	05-02-2024	Smartphone	INR	141279.78
MUM-006	06-02-2024	Smartphone	INR	64522.69

Key differences in data

- ◆ Different date formats - YYYY-MM-DD format in the US store and DD-MM-YYYY format in the Indian store.
- ◆ Financial data is captured in US dollars in New York store, and in Indian Rupees in the Mumbai store.

Phase 2 - Transform

Steps in the 'Transform' phase

- ◆ Join datasets by bringing together appropriate data in columns
- ◆ Standardize date form to YYYY-MM-DD
- ◆ Standardize financial details to reflect US dollars.

Transaction ID	Date	Product Type	Currency	Final Amount incl. Tax
NEW-001	2024-02-01	Accessory	USD	105.55
NEW-002	2024-02-02	Accessory	USD	115.54
NEW-003	2024-02-03	Smartphone	USD	2262.97
MUM-004	2024-02-04	Accessory	USD	55.67
MUM-005	2024-02-05	Smartphone	USD	1766
MUM-006	2024-02-06	Smartphone	USD	806.53

Phase 3 - Load

The combined dataset is then loaded to a database. This data is then loaded to an analytic platform to generate reports.

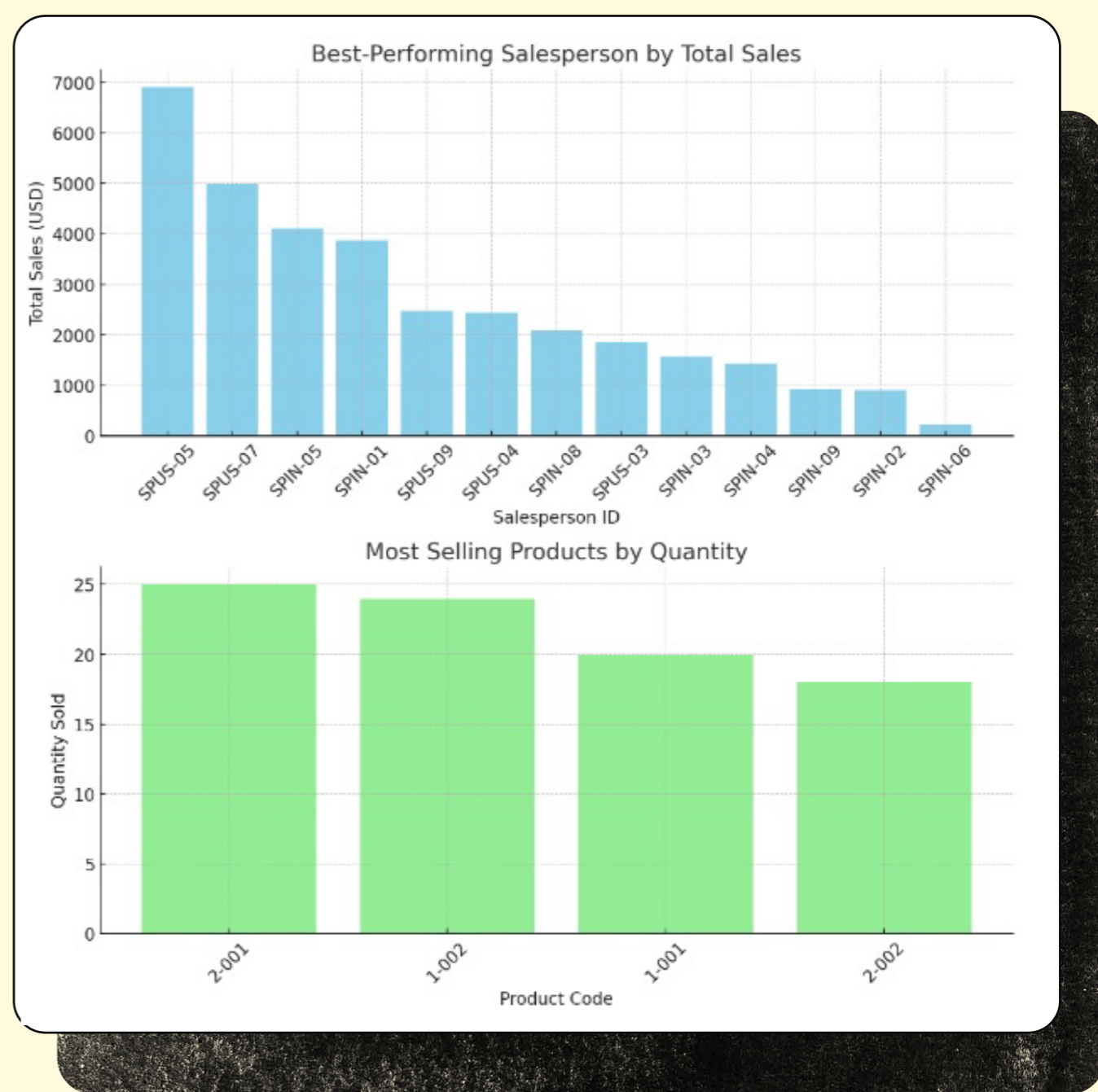
- ◆ Best performing sales person across stores
- ◆ Best selling product of the month



Best performing sales person



Best Selling product



Extract

Transform

Load

ETL processes can be engineered for other reports.



Sales Trends Over Time



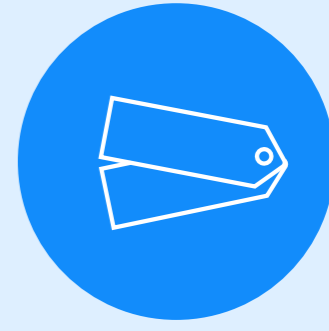
Average Transaction Value



Discount Impact Analysis



Sales Distribution by Location



Product Category Performance