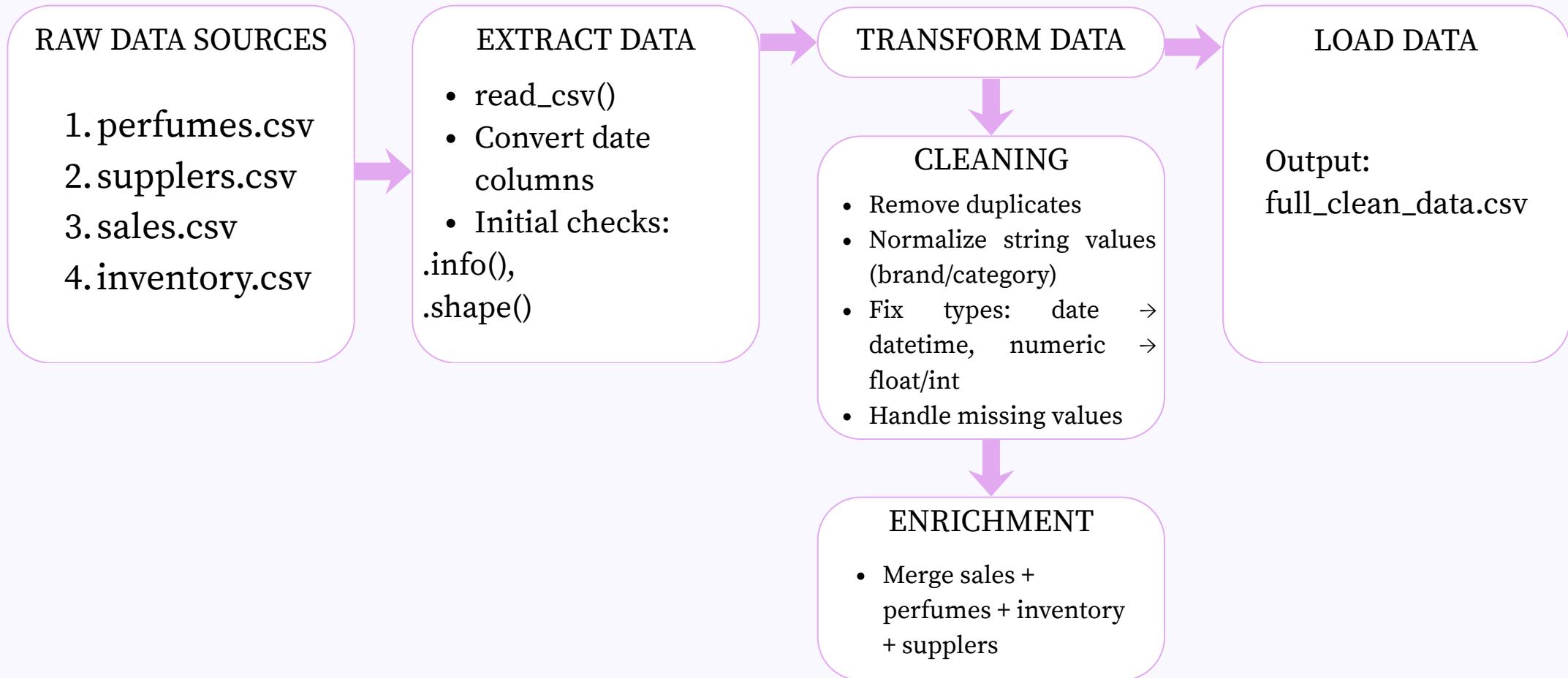


# PYTHON ETL PIPELINE (pandas)



# PYTHON EDA PIPELINE

## IMPORT EDA TOOLS AND DATA

- Pandas
- Matplotlib
- Seaborn
- NumPy
- full\_clean\_data.csv

## DATA OVERVIEW

- .info()
- .head()
- .isna().sum()
- .describe()

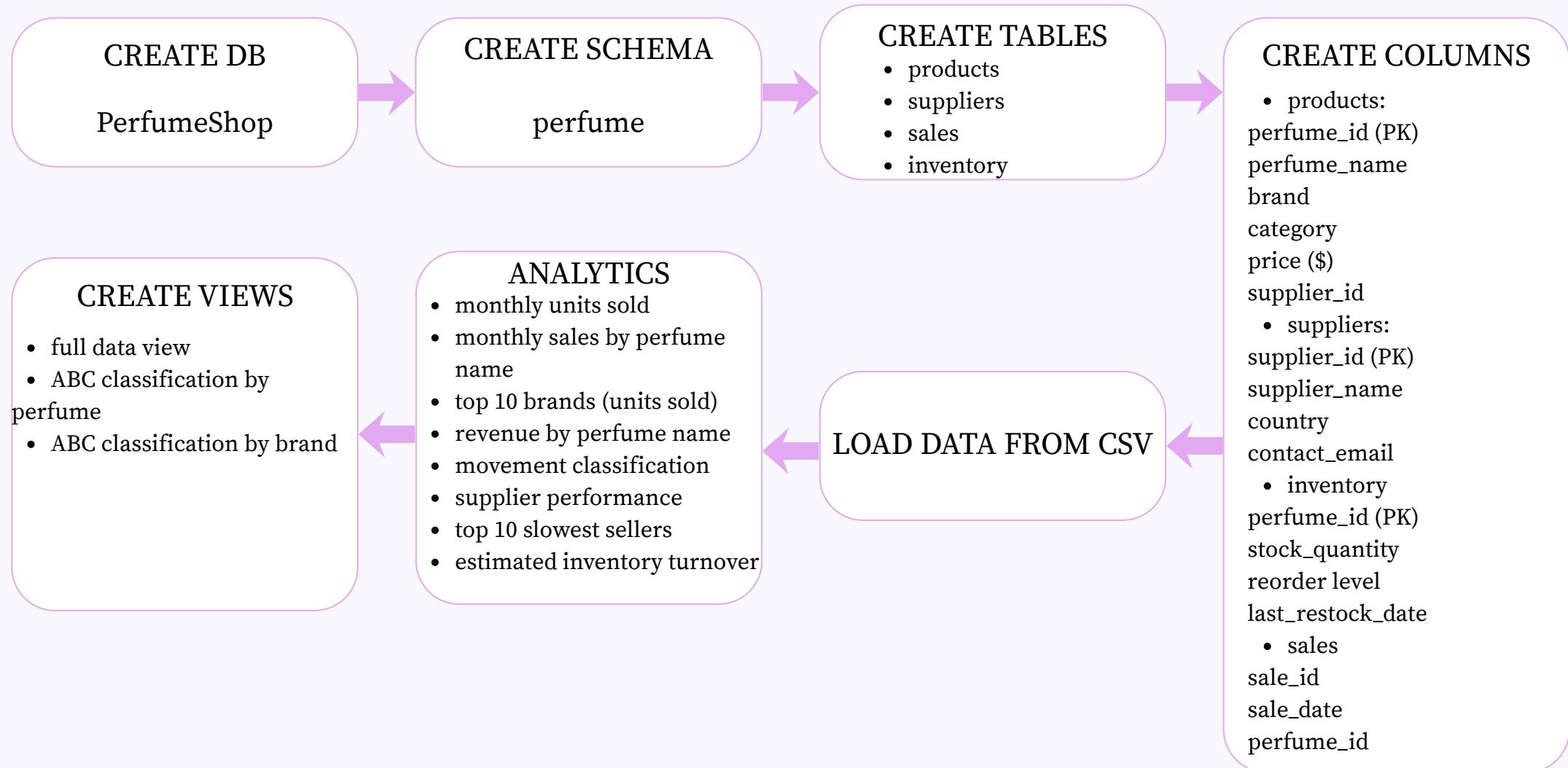
## DATA TYPE NORMALIZING AND ADDING 'MONTH' COLUMN (PERIOD)

- ## VISUALS
- number of perfumes by brand
  - price distribution
  - sales count per day
  - monthly sales volume
  - revenue by brand
  - top 10 perfumes by revenue
  - revenue by supplier
  - daily revenue
  - revenue share by category
  - stock quantity by brand

## KEY INSIGHTS

- Brands **Dior** and **Yves Saint Laurent** have the largest number of perfumes available in the shop
- The largest sales count was in mid-February (>25), and smallest sales count was in mid-May (<5)
- Brands with highest revenue: **Diptyque, Dior, Jo Malone, Tom Ford, Yves Saint Laurent** (>\$25k), lowest revenue: **Guerlain, Dunhill, Ralph Lauren, Estee Lauder** (<\$3.5k)
- Perfume with highest revenue: **Neroli Portofino** (>\$20k)
- Supplier with highest revenue: **L'Oreal Luxe** (>\$30k), lowest revenue: **Inter Parfums UK Ltd.** (<\$5k)

# SQL (POSTGRESQL) PIPELINE



# POWER BI PIPELINE

