

Project X2: Visualization of big messy data

Here, the Dillard's department store data was analyzed to compare sales in various US states in 2016. After the preprocessing of data, the sales results were visualized in Fig. 1, and the top 10 departments with largest sales were found as shown in Fig. 2. The departments were named with numbers in this database. The SQL code to retrieve data are available in a separate file in this project.

- Data preprocessing showed missing data.
- The store big data were located in Teradata (see Ref. [1] for data source)
- Tableau were used for data visualization.
- The Dillards' database schema is shown in Fig.3

Sales by State

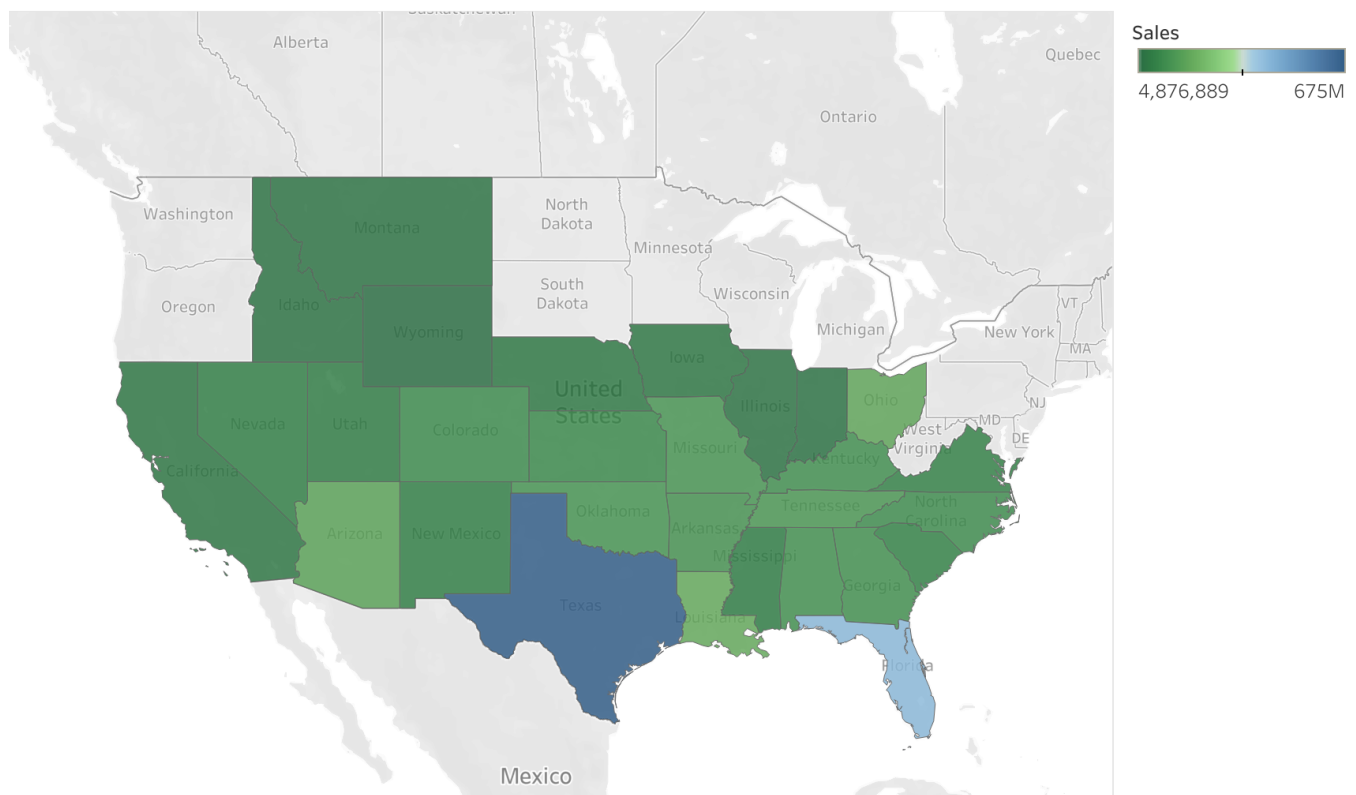


Fig.1: Comparison of total sales by state for Dillards' department store (Sales results in some states were not available in the database.)

Sales by Department

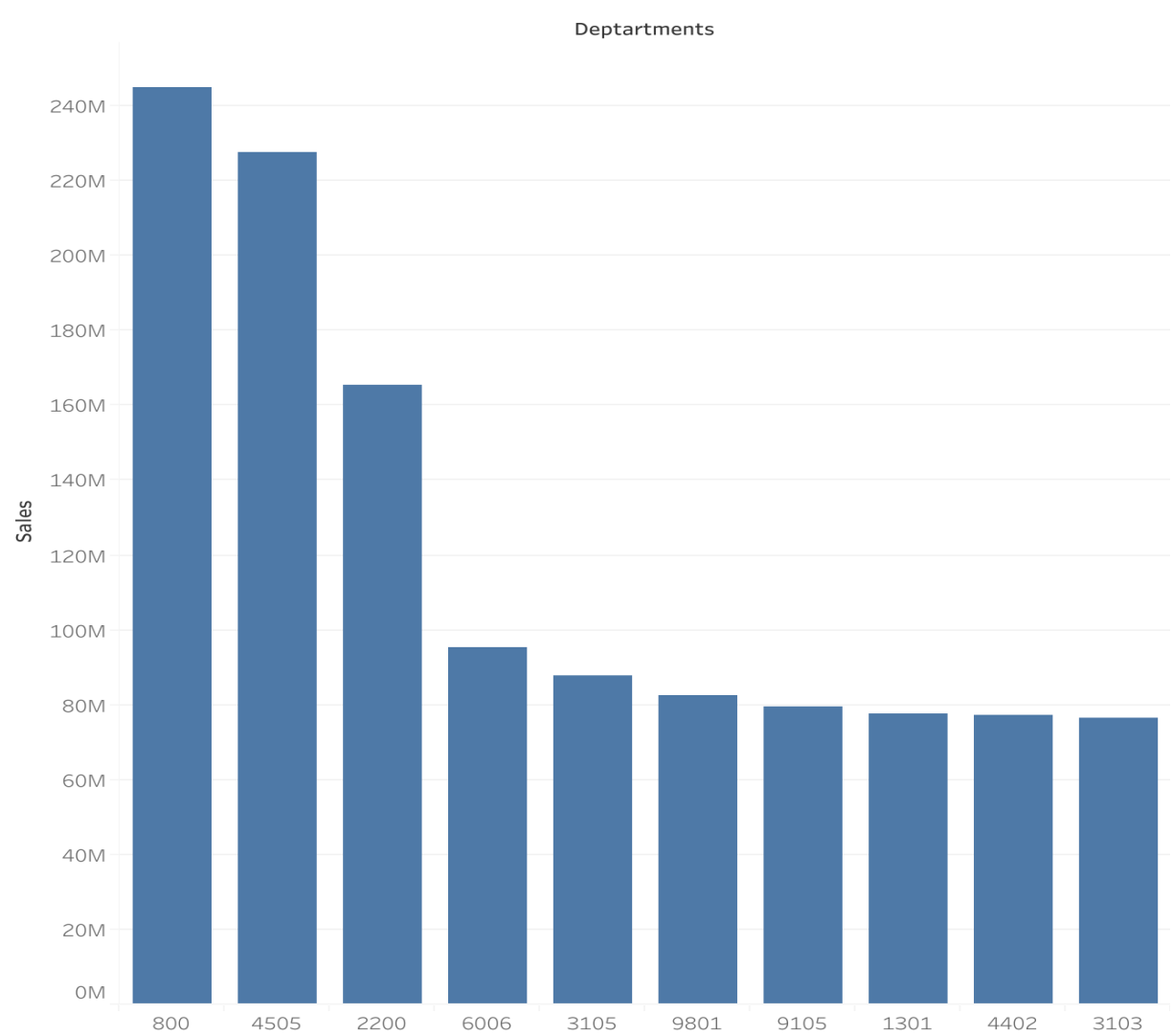


Fig. 2: Top 10 departments with maximum sales

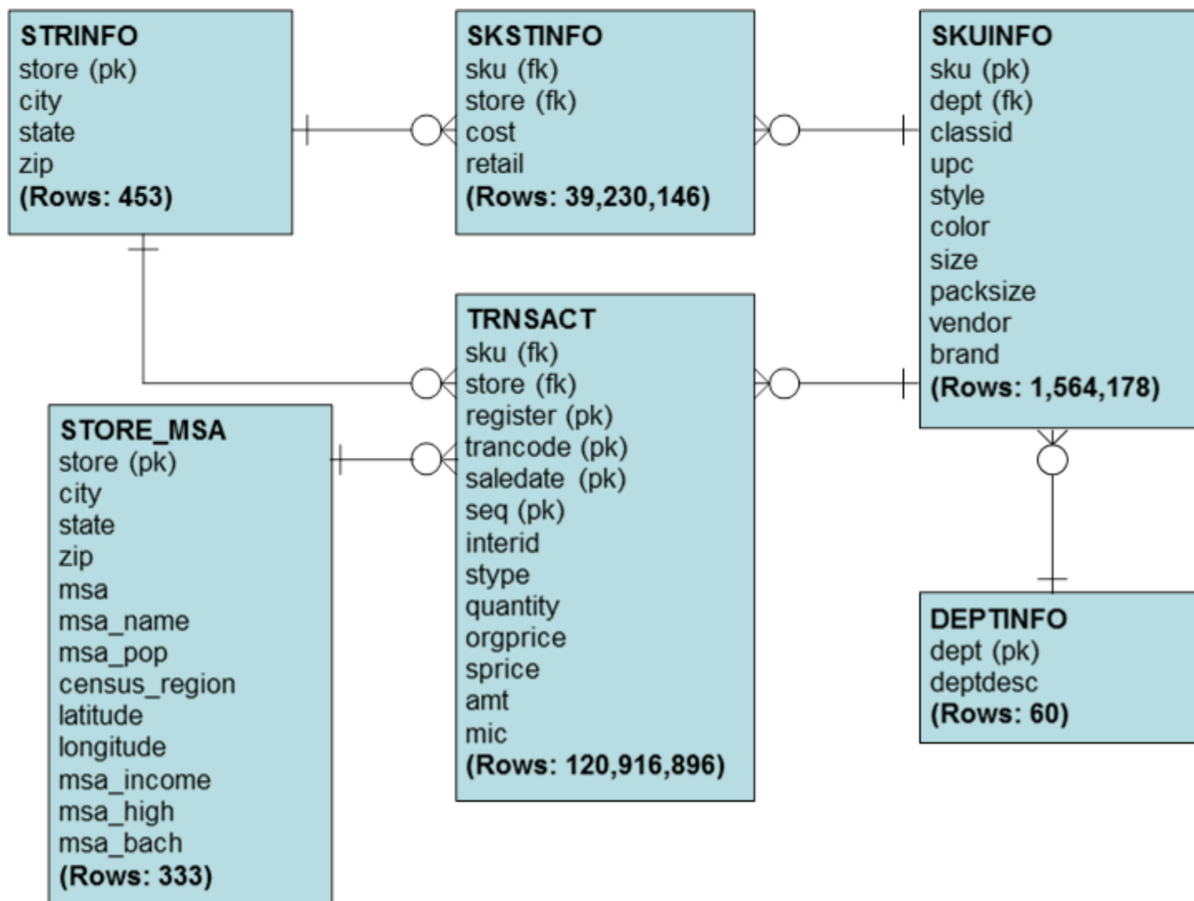


Fig 3: Dillards' database relational schema

Ref [1]: Managing big data with MYSQL, Duke University course @ Coursera.