

# Data Description Report: SSENSE Dataset

Dylan Hong

2024-02-16

## Introduction

This report provides an in-depth description of the SSENSE dataset, focusing on the quantity and quality of the data, including an analysis of product listings comprising brand, description, price, and type. The purpose is to support data mining projects by detailing the dataset's characteristics.

## Data Quantity

The dataset is structured in a CSV format and consists of 19,919 rows and 4 columns. These dimensions indicate a substantial amount of data for analysis, covering a wide range of products.

## Format of the Data

The data is tabular, stored in a CSV file, suitable for analysis in R or any other data processing framework.

## Data Capture Method

The dataset appears to be compiled from an e-commerce platform, specifically targeting luxury fashion items.

## Size of the Database

The database contains 19,919 records (rows) and 4 attributes (columns), encompassing a diverse selection of fashion items.

## Data Quality

### Relevance to Business Question

Attributes included are:

- **brand**: Categorical, indicating the manufacturer or designer of the item.
- **description**: Textual, providing details about the item.
- **price\_usd**: Numerical, listing the item's price in US dollars.
- **type**: Categorical, indicating the fashion category (e.g., menswear, womenswear).

These attributes are critical for analyzing market trends, pricing strategies, and inventory management.

## Data Types Present

- **brand** and **type**: Categorical (String)
- **description**: Textual (String)
- **price\_usd**: Numerical (Integer)

## Key Attributes Basic Statistics

The key numerical attribute, `price_usd`, has the following statistics:

- Count: 19,919
- Mean: \$559.41
- Standard Deviation: \$682.36
- Min: \$20
- 25th Percentile: \$185
- Median: \$350
- 75th Percentile: \$648.50
- Max: \$13,000

These statistics reveal a wide range in the pricing of items, from affordable to high-end luxury products.

## Observations and Insights

The dataset demonstrates a significant diversity in product offerings, with prices ranging broadly from \$20 to \$13,000. This variability suggests the presence of both accessible and high-end luxury items. The standard deviation in price points to a considerable dispersion in item prices, which could be indicative of varied target demographics or product exclusivity.

## Conclusion

This Data Description Report has outlined the SSENSE dataset's fundamental characteristics, emphasizing its potential utility in business analyses related to pricing strategies, market positioning, and inventory categorization. The dataset's rich assortment of fashion items, combined with detailed attributes, provides a robust foundation for in-depth market analysis and strategic business decision-making.