

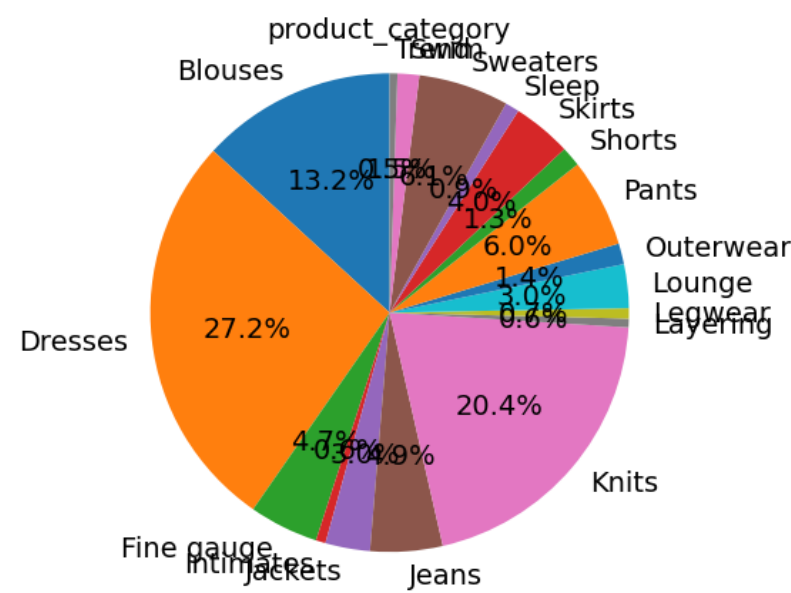
# Analysis Report

We report the following SageMaker analysis.

## Pre-training Bias Metrics

We computed the bias metrics for the label `sentiment` using label value(s)/threshold `1` for the below facets.

- product\_category**  
The groups are represented in the dataset with the following proportions.



Value(s)/Threshold: Blouses

| name | description  | value    |
|------|--|----------|
| CI   | Class Imbalance (CI)                               | 0.736321 |
| DPL  | Difference in Positive Proportions in Labels (DPL) | 0.016356 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000186 |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.000737 |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.016356 |
| LP   | L-p Norm (LP)                                      | 0.023131 |
| TVD  | Total Variation Distance (TVD)                     | 0.016356 |

Value(s)/Threshold: Dresses

| name | description  | value    |
|------|--|----------|
| CI   | Class Imbalance (CI)                               | 0.45682  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | 0.022482 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000352 |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.001392 |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.022482 |
| LP   | L-p Norm (LP)                                      | 0.031795 |
| TVD  | Total Variation Distance (TVD)                     | 0.022482 |

Value(s)/Threshold: Pants

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.880668  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.026661 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000522  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.002119  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.026661  |
| LP   | L-p Norm (LP)                                      | 0.037704  |
| TVD  | Total Variation Distance (TVD)                     | 0.026661  |

Value(s)/Threshold: Knits

| name | description  | value    |
|------|--|----------|
| CI   | Class Imbalance (CI)                               | 0.59109  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | 0.011213 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000088 |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.00035  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.011213 |
| LP   | L-p Norm (LP)                                      | 0.015857 |
| TVD  | Total Variation Distance (TVD)                     | 0.011213 |

Value(s)/Threshold: Intimates

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.987006  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.025599 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000483  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.001959  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.025599  |
| LP   | L-p Norm (LP)                                      | 0.036203  |
| TVD  | Total Variation Distance (TVD)                     | 0.025599  |

Value(s)/Threshold: Outerwear

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.971802  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.026121 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000503  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.00204   |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.026121  |
| LP   | L-p Norm (LP)                                      | 0.036941  |
| TVD  | Total Variation Distance (TVD)                     | 0.026121  |

Value(s)/Threshold: Lounge

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.940864  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.045509 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.001573  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.006474  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.045509  |
| LP   | L-p Norm (LP)                                      | 0.06436   |
| TVD  | Total Variation Distance (TVD)                     | 0.045509  |

Value(s)/Threshold: Sweaters

| name | description  | value    |
|------|--|----------|
| CI   | Class Imbalance (CI)                               | 0.878016 |
| DPL  | Difference in Positive Proportions in Labels (DPL) | 0.021044 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000305 |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.001207 |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.021044 |
| LP   | L-p Norm (LP)                                      | 0.029761 |
| TVD  | Total Variation Distance (TVD)                     | 0.021044 |

Value(s)/Threshold: Skirts

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.92018   |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.021053 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000323  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.001308  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.021053  |
| LP   | L-p Norm (LP)                                      | 0.029773  |
| TVD  | Total Variation Distance (TVD)                     | 0.021053  |

Value(s)/Threshold: Fine gauge

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.906391  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.020859 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000317  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.001283  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.020859  |
| LP   | L-p Norm (LP)                                      | 0.0295    |
| TVD  | Total Variation Distance (TVD)                     | 0.020859  |

Value(s)/Threshold: Sleep

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.981084  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.047723 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.001743  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.007185  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.047723  |
| LP   | L-p Norm (LP)                                      | 0.067491  |
| TVD  | Total Variation Distance (TVD)                     | 0.047723  |

Value(s)/Threshold: Jackets

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.939627  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.035868 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000961  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.003928  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.035868  |
| LP   | L-p Norm (LP)                                      | 0.050725  |
| TVD  | Total Variation Distance (TVD)                     | 0.035868  |

Value(s)/Threshold: Swim

| name | description  | value    |
|------|--|----------|
| CI   | Class Imbalance (CI)                               | 0.970653 |
| DPL  | Difference in Positive Proportions in Labels (DPL) | 0.01162  |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000094 |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.000373 |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.01162  |
| LP   | L-p Norm (LP)                                      | 0.016433 |
| TVD  | Total Variation Distance (TVD)                     | 0.01162  |

Value(s)/Threshold: Trend

| name | description  | value    |
|------|--|----------|
| CI   | Class Imbalance (CI)                               | 0.98957  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | 0.110042 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.00748  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.028876 |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.110042 |
| LP   | L-p Norm (LP)                                      | 0.155623 |
| TVD  | Total Variation Distance (TVD)                     | 0.110042 |

Value(s)/Threshold: Jeans

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.902413  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.055597 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.002382  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.009875  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.055597  |
| LP   | L-p Norm (LP)                                      | 0.078626  |
| TVD  | Total Variation Distance (TVD)                     | 0.055597  |

Value(s)/Threshold: Legwear

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.986034  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.027173 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.000545  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.002215  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.027173  |
| LP   | L-p Norm (LP)                                      | 0.038428  |
| TVD  | Total Variation Distance (TVD)                     | 0.027173  |

Value(s)/Threshold: Shorts

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.973128  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.019247 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.00027   |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.001091  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.019247  |
| LP   | L-p Norm (LP)                                      | 0.027219  |
| TVD  | Total Variation Distance (TVD)                     | 0.019247  |

Value(s)/Threshold: Layering

| name | description  | value     |
|------|--|-----------|
| CI   | Class Imbalance (CI)                               | 0.988332  |
| DPL  | Difference in Positive Proportions in Labels (DPL) | -0.086077 |
| JS   | Jensen-Shannon Divergence (JS)                     | 0.006138  |
| KL   | Kullback-Liebler Divergence (KL)                   | 0.026226  |
| KS   | Kolmogorov-Smirnov Distance (KS)                   | 0.086077  |
| LP   | L-p Norm (LP)                                      | 0.121732  |
| TVD  | Total Variation Distance (TVD)                     | 0.086077  |

## Analysis Configuration Parameters

```
{
  "dataset_type": "text/csv",
  "headers": [
    "sentiment",
    "review_body",
    "product_category"
  ],
  "label": "sentiment",
  "label_values_or_threshold": [
    1
  ],
  "facet": [
    {
      "name_or_index": "product_category"
    }
  ],
  "methods": {
    "pre_training_bias": {
      "methods": [
        "CI",
        "DPL",
        "KL",
        "JS",
        "LP",
        "TVD",
        "KS"
      ]
    },
    "report": {
      "name": "report",
      "title": "Analysis Report"
    }
  }
}
```