Analysis Report

Global dataset report

This report is the output of the Amazon SageMaker Clarify analysis. The report is split into following parts:

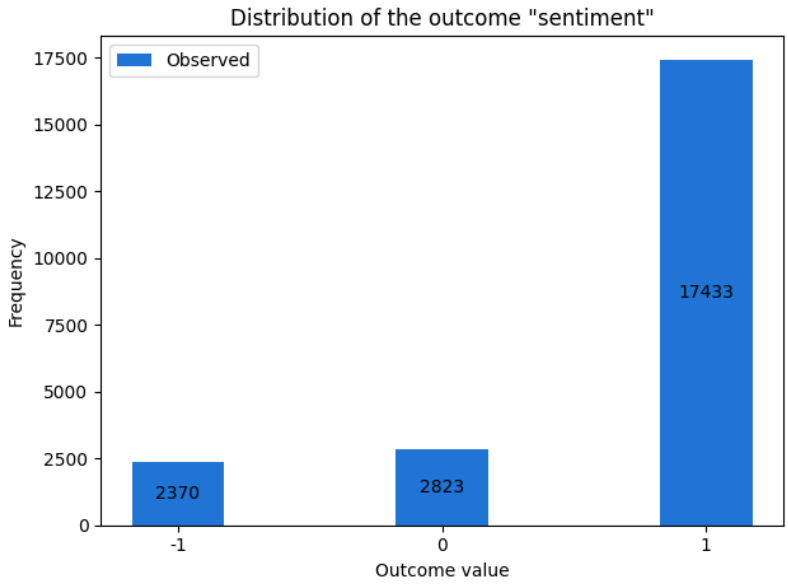
1. Analysis configuration
2. Pretraining bias metrics

Analysis Configuration

Bias analysis requires you to configure the outcome label column, the facet and optionally a group variable. Generating explanations requires you to configure the outcome label. You configured the analysis with the following variables. The complete analysis configuration is appended at the end.

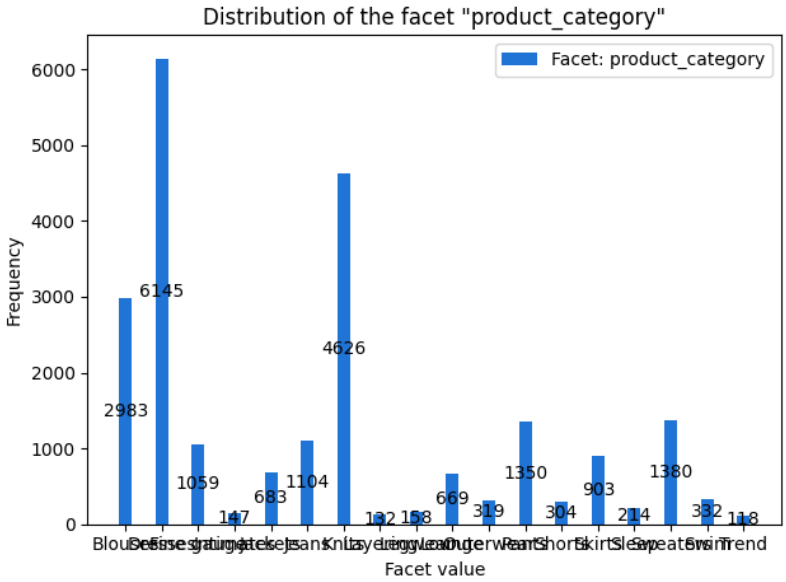
Outcome label: You chose the column `sentiment` in the input data as the outcome label. Bias metric computation requires designating the positive outcome. You chose `sentiment = 1` as the positive outcome. `sentiment` consisted of values `[-1, 0, 1]`.

The figure below shows the distribution of values of `sentiment`.



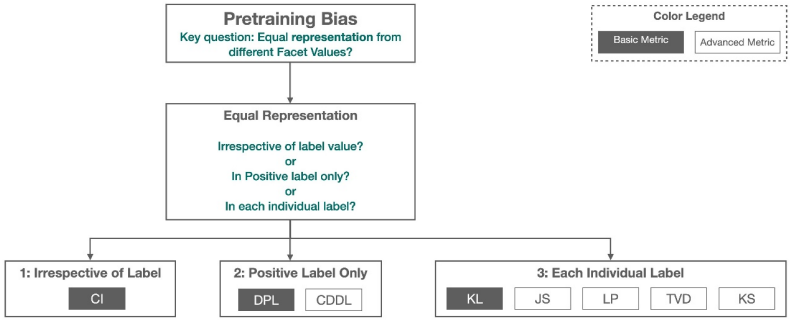
Facet: You chose the column `product_category` in the input data as the facet. `product_category` consisted of values `['Blouses', 'Dresses', 'Fine gauge', 'Intimates', 'Jackets', 'Jeans', 'Knits', 'Layering', 'Legwear', 'Lounge', 'Outerwear', 'Pants', 'Shorts', 'Skirts', 'Sleep', 'Sweaters', 'Swim', 'Trend']` . Bias metrics were computed by comparing the inputs `product_category = Blouses` with all other inputs, then by comparing inputs `product_category = Dresses` with all other inputs, then by comparing inputs `product_category = Pants` with all other inputs, then by comparing inputs `product_category = Knits` with all other inputs, then by comparing inputs `product_category = Intimates` with all other inputs, then by comparing inputs `product_category = Outerwear` with all other inputs, then by comparing inputs `product_category = Lounge` with all other inputs, then by comparing inputs `product_category = Sweaters` with all other inputs, then by comparing inputs `product_category = Skirts` with all other inputs, then by comparing inputs `product_category = Fine gauge` with all other inputs, then by comparing inputs `product_category = Sleep` with all other inputs, then by comparing inputs `product_category = Jackets` with all other inputs, then by comparing inputs `product_category = Swim` with all other inputs, then by comparing inputs `product_category = Trend` with all other inputs, then by comparing inputs `product_category = Jeans` with all other inputs, then by comparing inputs `product_category = Legwear` with all other inputs, then by comparing inputs `product_category = Shorts` with all other inputs, then by comparing inputs `product_category = Layering` with all other inputs.

The figure below shows the distribution of values of `product_category`.



Pre-training Bias Metrics

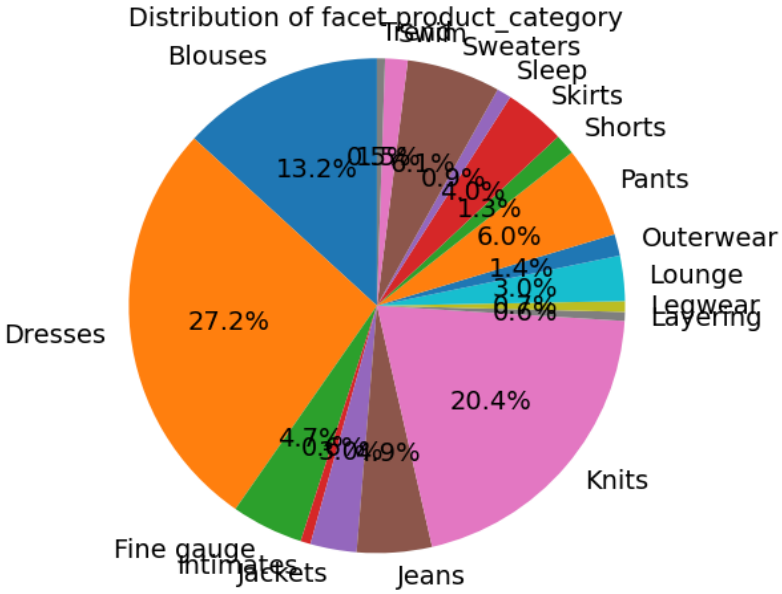
Pretraining bias metrics measure imbalances in facet value representation in the training data. Imbalances can be measured across different dimensions. For instance, you could focus imbalances within the inputs with positive observed label only. The figure below shows how different pretraining bias metrics focus on different dimensions. For a detailed description of these dimensions, see Learn How Amazon SageMaker Clarify Helps Detect Bias.



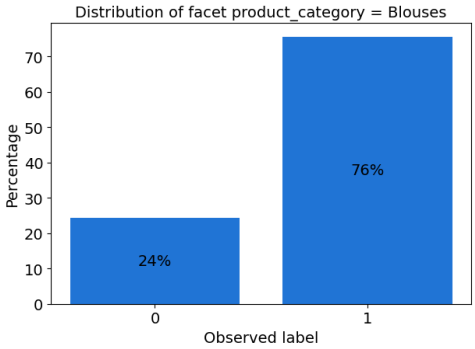
The metric values along with an informal description of what they mean are shown below. For mathematical formulas and examples, see the Measure Pretraining Bias section of the AWS documentation.

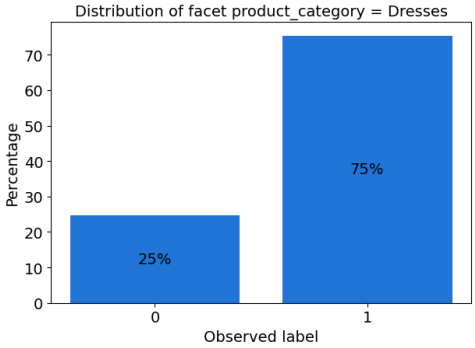
We computed the bias metrics for the label `sentiment` using label value(s)/threshold `sentiment = 1` for the following facets:

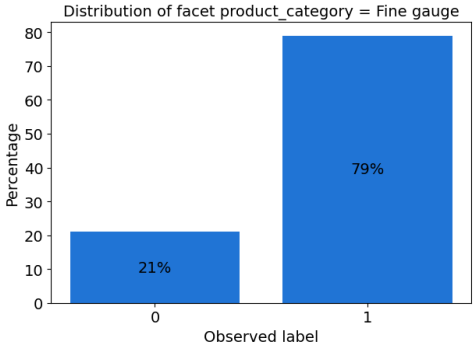
- Facet column: **product_category**
  The pie chart shows the distribution of facet column `product_category` in your data.

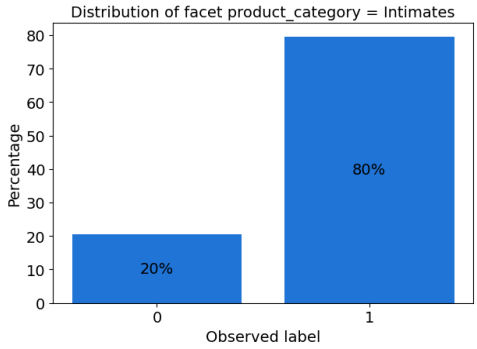


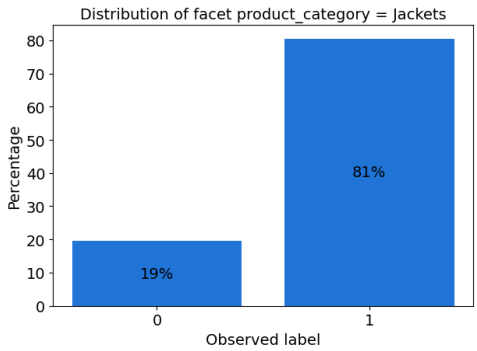
The bar plot(s) below show the distribution of facet column `product_category` in your data.

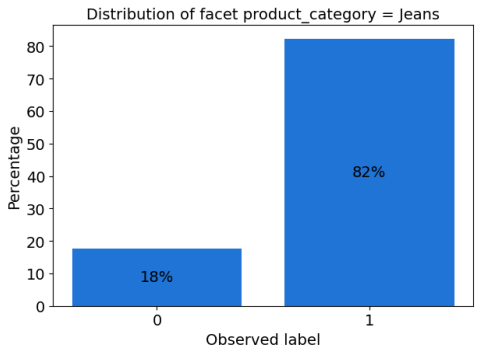


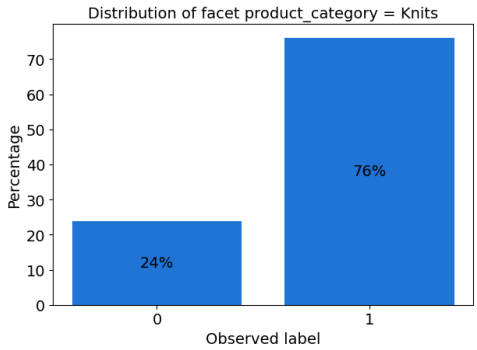


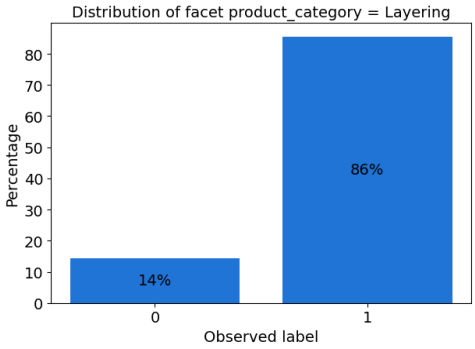


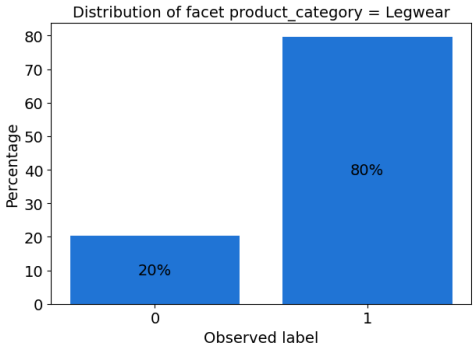

Distribution of facet product_category = Intimates
Percentage
Observed label
20%
80%
0
1

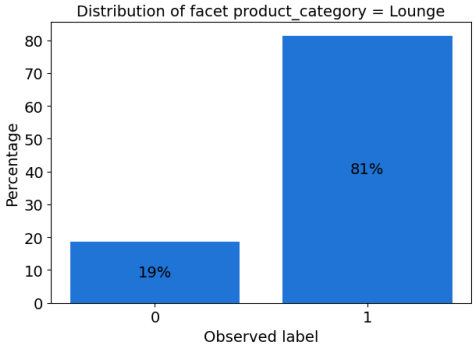

Distribution of facet product_category = Jackets
Percentage
Observed label
19%
81%
0
1

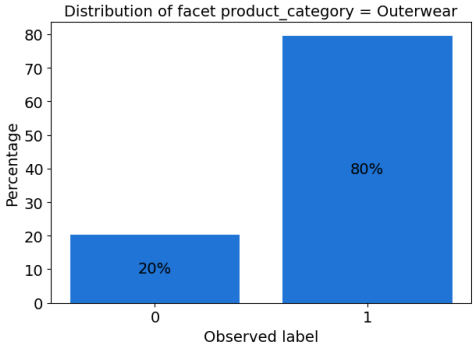

Distribution of facet product_category = Jeans
Percentage
Observed label
18%
82%
0
1

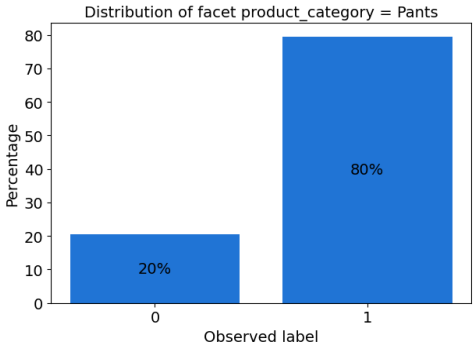

Distribution of facet product_category = Knits
Percentage
Observed label
24%
76%
0
1

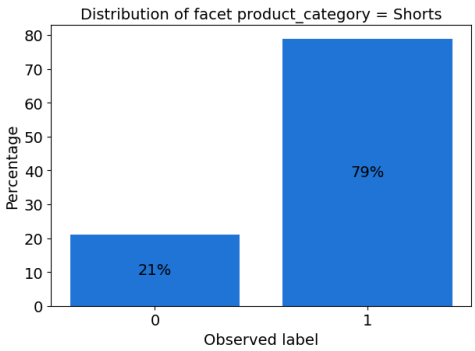

Distribution of facet product_category = Layering
Percentage
Observed label
14%
86%
0
1

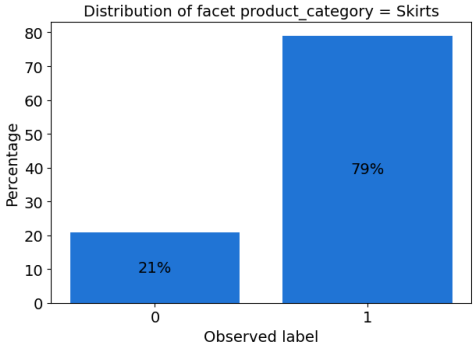

Distribution of facet product_category = Legwear
Percentage
Observed label
20%
80%
0
1

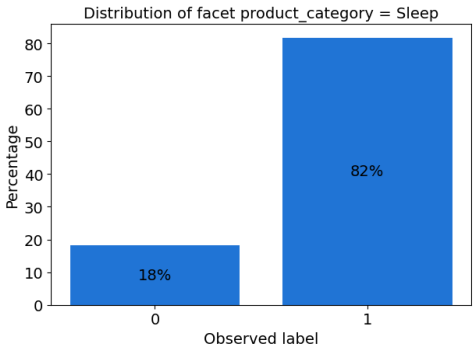

Distribution of facet product_category = Lounge
Percentage
Observed label
19%
81%
0
1

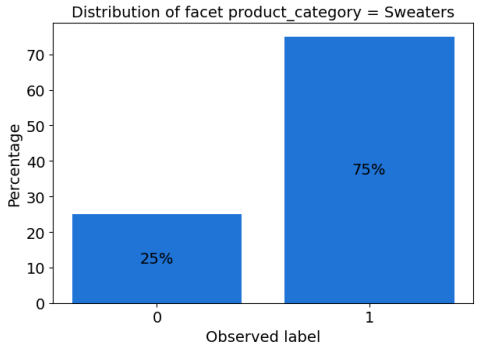

Distribution of facet product_category = Outerwear
Percentage
Observed label
20%
80%
0
1

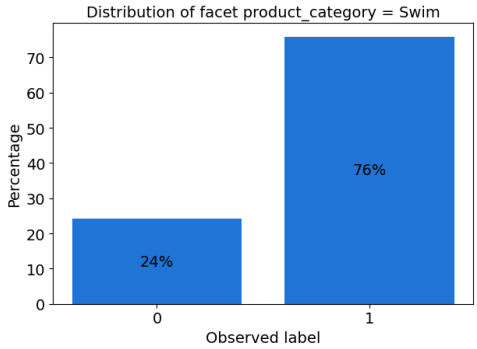

Distribution of facet product_category = Pants
Percentage
Observed label
20%
80%
0
1

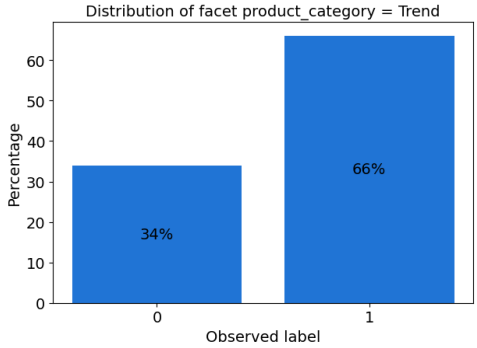

Distribution of facet product_category = Shorts
Percentage
Observed label
21%
79%
0
1


Distribution of facet product_category = Skirts
Percentage
Observed label
21%
79%
0
1


Distribution of facet product_category = Sleep
Percentage
Observed label
18%
82%
0
1







Facet Value(s)/Threshold: `product_category = Blouses`

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `product_category = Blouses` and rest of the inputs.	0.736
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `product_category = Blouses` and rest of the inputs.	0.016
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `product_category = Blouses` and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `product_category = Blouses` and rest of the inputs diverge from each other entropically.	0.001
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `product_category = Blouses` and rest of the inputs in the dataset.	0.016
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `product_category = Blouses` rest of the inputs in the dataset.	0.023
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `product_category = Blouses` and rest of the inputs in the dataset.	0.016

Facet Value(s)/Threshold: `product_category = Dresses`

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Dresses and rest of the inputs.	0.457
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Dresses and rest of the inputs.	0.022
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Dresses and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Dresses and rest of the inputs diverge from each other entropically.	0.001
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Dresses and rest of the inputs in the dataset.	0.022
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Dresses rest of the inputs in the dataset.	0.032
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Dresses and rest of the inputs in the dataset.	0.022

Facet Value(s)/Threshold: product_category = Pants

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Pants and rest of the inputs.	0.881
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Pants and rest of the inputs.	-0.027
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Pants and rest of the inputs diverge from each other entropically.	0.001
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Pants and rest of the inputs diverge from each other entropically.	0.002
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Pants and rest of the inputs in the dataset.	0.027
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Pants rest of the inputs in the dataset.	0.038
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Pants and rest of the inputs in the dataset.	0.027

Facet Value(s)/Threshold: product_category = Knits

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Knits and rest of the inputs.	0.591
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Knits and rest of the inputs.	0.011
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Knits and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Knits and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Knits and rest of the inputs in the dataset.	0.011
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Knits rest of the inputs in the dataset.	0.016
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Knits and rest of the inputs in the dataset.	0.011

Facet Value(s)/Threshold: product_category = Intimates

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Intimates and rest of the inputs.	0.987
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Intimates and rest of the inputs.	-0.026
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Intimates and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Intimates and rest of the inputs diverge from each other entropically.	0.002
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Intimates and rest of the inputs in the dataset.	0.026
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Intimates rest of the inputs in the dataset.	0.036
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Intimates and rest of the inputs in the dataset.	0.026

Facet Value(s)/Threshold: product_category = Outerwear

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Outerwear and rest of the inputs.	0.972
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Outerwear and rest of the inputs.	-0.026
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Outerwear and rest of the inputs diverge from each other entropically.	0.001
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Outerwear and rest of the inputs diverge from each other entropically.	0.002
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Outerwear and rest of the inputs in the dataset.	0.026
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Outerwear rest of the inputs in the dataset.	0.037
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Outerwear and rest of the inputs in the dataset.	0.026

Facet Value(s)/Threshold: product_category = Lounge

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Lounge and rest of the inputs.	0.941
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Lounge and rest of the inputs.	-0.046
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Lounge and rest of the inputs diverge from each other entropically.	0.002
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Lounge and rest of the inputs diverge from each other entropically.	0.006
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Lounge and rest of the inputs in the dataset.	0.046
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Lounge rest of the inputs in the dataset.	0.064
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Lounge and rest of the inputs in the dataset.	0.046

Facet Value(s)/Threshold: product_category = Sweaters

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Sweaters and rest of the inputs.	0.878
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Sweaters and rest of the inputs.	0.021
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Sweaters and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Sweaters and rest of the inputs diverge from each other entropically.	0.001
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Sweaters and rest of the inputs in the dataset.	0.021
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Sweaters rest of the inputs in the dataset.	0.030
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Sweaters and rest of the inputs in the dataset.	0.021

Facet Value(s)/Threshold: product_category = Skirts

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Skirts and rest of the inputs.	0.920
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Skirts and rest of the inputs.	-0.021
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Skirts and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Skirts and rest of the inputs diverge from each other entropically.	0.001
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Skirts and rest of the inputs in the dataset.	0.021
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Skirts rest of the inputs in the dataset.	0.030
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Skirts and rest of the inputs in the dataset.	0.021

Facet Value(s)/Threshold: product_category = Fine gauge

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Fine gauge and rest of the inputs.	0.906
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Fine gauge and rest of the inputs.	-0.021
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Fine gauge and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Fine gauge and rest of the inputs diverge from each other entropically.	0.001
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Fine gauge and rest of the inputs in the dataset.	0.021
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Fine gauge rest of the inputs in the dataset.	0.029
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Fine gauge and rest of the inputs in the dataset.	0.021

Facet Value(s)/Threshold: product_category = Sleep

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `product_category = Sleep` and rest of the inputs.	0.981
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `product_category = Sleep` and rest of the inputs.	-0.048
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `product_category = Sleep` and rest of the inputs diverge from each other entropically.	0.002
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `product_category = Sleep` and rest of the inputs diverge from each other entropically.	0.007
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `product_category = Sleep` and rest of the inputs in the dataset.	0.048
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `product_category = Sleep` rest of the inputs in the dataset.	0.067
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `product_category = Sleep` and rest of the inputs in the dataset.	0.048

Facet Value(s)/Threshold: `product_category = Jackets`

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `product_category = Jackets` and rest of the inputs.	0.940
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `product_category = Jackets` and rest of the inputs.	-0.036
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `product_category = Jackets` and rest of the inputs diverge from each other entropically.	0.001
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `product_category = Jackets` and rest of the inputs diverge from each other entropically.	0.004
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `product_category = Jackets` and rest of the inputs in the dataset.	0.036
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `product_category = Jackets` rest of the inputs in the dataset.	0.051
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `product_category = Jackets` and rest of the inputs in the dataset.	0.036

Facet Value(s)/Threshold: `product_category = Swim`

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `product_category = Swim` and rest of the inputs.	0.971
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `product_category = Swim` and rest of the inputs.	0.012
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `product_category = Swim` and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `product_category = Swim` and rest of the inputs diverge from each other entropically.	0.000
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `product_category = Swim` and rest of the inputs in the dataset.	0.012
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `product_category = Swim` rest of the inputs in the dataset.	0.016
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `product_category = Swim` and rest of the inputs in the dataset.	0.012

Facet Value(s)/Threshold: `product_category = Trend`

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Trend and rest of the inputs.	0.990
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Trend and rest of the inputs.	0.110
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Trend and rest of the inputs diverge from each other entropically.	0.007
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Trend and rest of the inputs diverge from each other entropically.	0.029
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Trend and rest of the inputs in the dataset.	0.110
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Trend rest of the inputs in the dataset.	0.156
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Trend and rest of the inputs in the dataset.	0.110

Facet Value(s)/Threshold: product_category = Jeans

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Jeans and rest of the inputs.	0.902
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Jeans and rest of the inputs.	-0.056
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Jeans and rest of the inputs diverge from each other entropically.	0.002
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Jeans and rest of the inputs diverge from each other entropically.	0.010
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Jeans and rest of the inputs in the dataset.	0.056
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Jeans rest of the inputs in the dataset.	0.079
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Jeans and rest of the inputs in the dataset.	0.056

Facet Value(s)/Threshold: product_category = Legwear

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values product_category = Legwear and rest of the inputs.	0.986
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values product_category = Legwear and rest of the inputs.	-0.027
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically.	0.001
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values product_category = Legwear and rest of the inputs diverge from each other entropically.	0.002
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values product_category = Legwear and rest of the inputs in the dataset.	0.027
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values product_category = Legwear rest of the inputs in the dataset.	0.038
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values product_category = Legwear and rest of the inputs in the dataset.	0.027

Facet Value(s)/Threshold: product_category = Shorts

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `product_category = Shorts` and rest of the inputs.	0.973
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `product_category = Shorts` and rest of the inputs.	-0.019
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `product_category = Shorts` and rest of the inputs diverge from each other entropically.	0.000
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `product_category = Shorts` and rest of the inputs diverge from each other entropically.	0.001
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `product_category = Shorts` and rest of the inputs in the dataset.	0.019
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `product_category = Shorts` rest of the inputs in the dataset.	0.027
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `product_category = Shorts` and rest of the inputs in the dataset.	0.019

Facet Value(s)/Threshold: `product_category = Layering`

Metric	Description	Value
Class Imbalance (CI)	Measures the imbalance in the number of inputs with facet values `product_category = Layering` and rest of the inputs.	0.988
Difference in Proportions of Labels (DPL)	Measures the imbalance of positive observed labels between facet values `product_category = Layering` and rest of the inputs.	-0.086
Jensen-Shannon Divergence (JS)	Measures how much the observed label distributions of facet values `product_category = Layering` and rest of the inputs diverge from each other entropically.	0.006
Kullback-Leibler Divergence (KL)	Measures how much the observed label distributions of facet values `product_category = Layering` and rest of the inputs diverge from each other entropically.	0.026
Kolmogorov-Smirnov (KS)	Measures maximum divergence between the observed label distributions for facet values `product_category = Layering` and rest of the inputs in the dataset.	0.086
Lp-norm (LP)	Measures a p-norm difference between the observed label distributions associated with facet values `product_category = Layering` rest of the inputs in the dataset.	0.122
Total Variation Distance (TVD)	Measures half of the L1-norm difference between the observed label distributions associated with facet values `product_category = Layering` and rest of the inputs in the dataset.	0.086

Appendix: 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"
      }
   }
}
```