|  |
| --- |
| Top 10 rules by lift:  1. BSA\_Category\_Large -> Smoke\_ES  Support: 0.1010  Confidence: 0.5263  Lift: 2.4812  Conviction: 1.6633  Leverage: 0.0603  Support Count: 0  2. BSA\_Category\_Small -> Smoke\_NS, Sex\_M  Support: 0.1111  Confidence: 0.6111  Lift: 2.0862  Conviction: 1.8182  Leverage: 0.0579  Support Count: 0  3. occupation\_category\_Housewife, Sex\_F, BMI\_Category\_Normal -> Smoke\_NS, BSA\_Category\_Medium  Support: 0.1010  Confidence: 0.9091  Lift: 1.8000  Conviction: 5.4444  Leverage: 0.0449  Support Count: 0  4. occupation\_category\_Housewife, Sex\_F -> Smoke\_NS, BSA\_Category\_Medium, BMI\_Category\_Normal  Support: 0.1010  Confidence: 0.6250  Lift: 1.7187  Conviction: 1.6970  Leverage: 0.0422  Support Count: 0  5. Smoke\_NS, BSA\_Category\_Small -> Sex\_M  Support: 0.1111  Confidence: 0.6471  Lift: 1.6858  Conviction: 1.7458  Leverage: 0.0452  Support Count: 0  6. occupation\_category\_Housewife, Age\_Category\_Senior, Sex\_F -> Smoke\_NS, BSA\_Category\_Medium  Support: 0.1010  Confidence: 0.8333  Lift: 1.6500  Conviction: 2.9697  Leverage: 0.0398  Support Count: 0  7. occupation\_category\_Housewife, Sex\_F -> Smoke\_NS, Age\_Category\_Senior, BSA\_Category\_Medium  Support: 0.1010  Confidence: 0.6250  Lift: 1.6283  Conviction: 1.6431  Leverage: 0.0390  Support Count: 0  8. occupation\_category\_Housewife, Sex\_F -> Smoke\_NS, BSA\_Category\_Medium  Support: 0.1313  Confidence: 0.8125  Lift: 1.6087  Conviction: 2.6397  Leverage: 0.0497  Support Count: 0  9. Smoke\_NS, occupation\_category\_Housewife, Sex\_F, BMI\_Category\_Normal -> BSA\_Category\_Medium  Support: 0.1010  Confidence: 1.0000  Lift: 1.5968  Conviction: inf  Leverage: 0.0378  Support Count: 0  10. occupation\_category\_Housewife, Sex\_F, BMI\_Category\_Normal -> BSA\_Category\_Medium  Support: 0.1111  Confidence: 1.0000  Lift: 1.5968  Conviction: inf  Leverage: 0.0415  Support Count: 0  Rules related to large BSA:  Top 10 rules by lift:  1. BSA\_Category\_Large -> Smoke\_ES  Support: 0.1010  Confidence: 0.5263  Lift: 2.4812  Conviction: 1.6633  Leverage: 0.0603  Support Count: 0  2. BSA\_Category\_Large -> Age\_Category\_Senior  Support: 0.1414  Confidence: 0.7368  Lift: 1.0274  Conviction: 1.0747  Leverage: 0.0038  Support Count: 0  Detailed Rule Information:  Rule 1:  Antecedents: BSA\_Category\_Large  Consequents: Smoke\_ES  Support: 0.1010  Confidence: 0.5263  Lift: 2.4812  Rule 2:  Antecedents: BSA\_Category\_Small  Consequents: Smoke\_NS, Sex\_M  Support: 0.1111  Confidence: 0.6111  Lift: 2.0862  Rule 3:  Antecedents: occupation\_category\_Housewife, Sex\_F, BMI\_Category\_Normal  Consequents: Smoke\_NS, BSA\_Category\_Medium  Support: 0.1010  Confidence: 0.9091  Lift: 1.8000  Rule 4:  Antecedents: occupation\_category\_Housewife, Sex\_F  Consequents: Smoke\_NS, BSA\_Category\_Medium, BMI\_Category\_Normal  Support: 0.1010  Confidence: 0.6250  Lift: 1.7187  Rule 5:  Antecedents: Smoke\_NS, BSA\_Category\_Small  Consequents: Sex\_M  Support: 0.1111  Confidence: 0.6471  Lift: 1.6858  Rule 6:  Antecedents: occupation\_category\_Housewife, Age\_Category\_Senior, Sex\_F  Consequents: Smoke\_NS, BSA\_Category\_Medium  Support: 0.1010  Confidence: 0.8333  Lift: 1.6500  Rule 7:  Antecedents: occupation\_category\_Housewife, Sex\_F  Consequents: Smoke\_NS, Age\_Category\_Senior, BSA\_Category\_Medium  Support: 0.1010  Confidence: 0.6250  Lift: 1.6283  Rule 8:  Antecedents: occupation\_category\_Housewife, Sex\_F  Consequents: Smoke\_NS, BSA\_Category\_Medium  Support: 0.1313  Confidence: 0.8125  Lift: 1.6087  Rule 9:  Antecedents: Smoke\_NS, occupation\_category\_Housewife, Sex\_F, BMI\_Category\_Normal  Consequents: BSA\_Category\_Medium  Support: 0.1010  Confidence: 1.0000  Lift: 1.5968  Rule 10:  Antecedents: occupation\_category\_Housewife, Sex\_F, BMI\_Category\_Normal  Consequents: BSA\_Category\_Medium  Support: 0.1111  Confidence: 1.0000  Lift: 1.5968  Summary Statistics:  Total number of rules: 289  Average lift: 1.2006  Average confidence: 0.7019  Average support: 0.2022 |

|  |
| --- |
| Analysis by Claude on the above results and dataset as context: |
| 1. BSA and Smoking:     The strongest association (highest lift of 2.4812) is between large Body Surface Area (BSA) and being an ex-smoker (ES). This suggests that patients with larger BSA are more likely to be ex-smokers. This could indicate that larger individuals might have been more prone to smoking in the past, possibly due to lifestyle factors or stress-related behaviors.  2. BSA, Sex, and Smoking:     There's a strong association (lift 2.0862) between small BSA, being male, and being a non-smoker (NS). This suggests that among asthma patients, men with smaller body surface areas are more likely to be non-smokers.  3. Female Housewives:     Several rules involve female housewives, indicating that this group has some distinct characteristics:     - They are likely to have medium BSA, be non-smokers, and have normal BMI (lift 1.8000).     - Many of them are seniors (lift 1.6283).     - All female housewives with normal BMI have medium BSA (lift 1.5968, confidence 1.0000).  4. Age and BSA:     There's a weak association between large BSA and being a senior (lift 1.0274). This suggests that older patients tend to have larger body surface areas, which is consistent with general aging patterns.  5. Smoking Patterns:     Ex-smokers are more associated with large BSA, while non-smokers are more associated with small BSA and being male. This could suggest that quitting smoking might be related to weight gain or overall body size increase in this patient group.  6. BMI and BSA:     Normal BMI is associated with medium BSA, particularly among female housewives. This indicates a consistent relationship between these body measurements in this specific demographic group.  Key Insights:  1. Smoking History: There appears to be a relationship between body size (BSA) and smoking history. Larger individuals are more likely to be ex-smokers, while smaller males are more likely to be non-smokers.  2. Female Housewives: This group shows consistent characteristics - they tend to have medium BSA, normal BMI, and are often non-smokers. Many are also seniors.  3. Body Measurements: There's a consistent relationship between BMI and BSA, particularly among female housewives, suggesting that these measurements are reliable indicators of body composition in this group.  4. Age Factors: Seniors tend to have larger BSA, which could be related to age-related changes in body composition.  These insights can help in developing more targeted asthma management strategies and identifying potential risk factors based on patient characteristics. However, it's important to note that association rules don't imply causation, and further clinical studies would be needed to confirm these relationships and their implications for asthma management. |

|  |
| --- |
| Cluster Characteristics:  Age BMI BSA Sex Smoke occupation\_category  Cluster  -1 64.111111 23.904463 1.730312 F ES Manufacturing  0 62.000000 22.432574 1.503493 M NS Housewife  1 62.666667 24.890745 1.646673 F NS Housewife  2 64.800000 23.864662 1.615990 F NS Unemployed  3 80.000000 23.467741 1.686694 M NS Unemployed  4 57.666667 22.222177 1.589440 F NS Office  5 65.363636 22.807041 1.631797 F NS Other  6 62.800000 23.678243 1.723075 F NS Service Industry  7 66.750000 23.237277 1.600286 M NS Education  8 67.750000 23.836232 1.799895 F NS Manufacturing  9 59.250000 21.883293 1.733283 M ES Construction  Cluster Sizes:  Cluster  -1 18  0 10  1 15  2 5  3 5  4 9  5 11  6 10  7 8  8 4  9 4  Name: count, dtype: int64  Number of outliers: 18  Clustered data saved to 'clustered\_asthma\_patients.csv' |

Loadings for the first three principal components:

|  | **PC1** | **PC2** | **PC3** |
| --- | --- | --- | --- |
| Age | 0.086194 | -0.154646 | -0.211222 |
| BMI | -0.217689 | 0.158694 | 0.555338 |
| BSA | -0.459094 | 0.076532 | 0.235661 |
| Sex\_F | -0.055228 | 0.624453 | -0.137128 |
| Sex\_M | 0.055228 | -0.624453 | 0.137128 |
| Smoke\_ES | -0.500584 | -0.139796 | -0.154008 |
| Smoke\_NS | 0.526442 | 0.070552 | 0.128197 |
| Smoke\_SM | -0.122288 | 0.157012 | 0.046791 |
| occupation\_category\_Agriculture | 0.086596 | 0.107759 | -0.183092 |
| occupation\_category\_Construction | -0.175020 | -0.065636 | -0.050880 |
| occupation\_category\_Education | 0.076440 | -0.089395 | -0.063216 |
| occupation\_category\_Finance | -0.118779 | -0.166869 | -0.023983 |
| occupation\_category\_Healthcare | 0.094354 | 0.076331 | -0.293776 |
| occupation\_category\_Housewife | 0.234761 | 0.058001 | 0.521700 |
| occupation\_category\_Manufacturing | -0.203958 | 0.036949 | 0.001383 |
| occupation\_category\_Office | 0.001402 | 0.118244 | -0.235986 |
| occupation\_category\_Other | -0.040559 | -0.108850 | -0.233056 |
| occupation\_category\_Service Industry | -0.082522 | 0.058417 | -0.000152 |
| occupation\_category\_Transportation | -0.115529 | -0.152471 | 0.016618 |
| occupation\_category\_Unemployed | 0.011593 | -0.034415 | 0.045362 |

Explained variance ratio: PC1: 14.43% PC2: 11.35% PC3: 7.96%

|  |
| --- |
| Analysis by Claude on the above results and dataset as context: |
| The output provides valuable insights into the structure of our data and how different features contribute to the principal components. Let's analyze this information:  1. Principal Component 1 (PC1) - 14.43% of variance explained:     - Strongest positive contributors: Smoke\_NS (0.526442), occupation\_category\_Housewife (0.234761)     - Strongest negative contributors: Smoke\_ES (-0.500584), BSA (-0.459094), BMI (-0.217689)     Interpretation: PC1 seems to primarily differentiate between non-smokers (particularly housewives) and smokers with larger body sizes. It might represent a lifestyle/health factor.  2. Principal Component 2 (PC2) - 11.35% of variance explained:     - Strongest positive contributors: Sex\_F (0.624453), BMI (0.158694)     - Strongest negative contributors: Sex\_M (-0.624453), Age (-0.154646)     Interpretation: PC2 appears to mainly distinguish between females and males, with some influence from BMI and age. It could represent a gender-related factor, possibly including some age and body composition aspects.  3. Principal Component 3 (PC3) - 7.96% of variance explained:     - Strongest positive contributors: BMI (0.555338), occupation\_category\_Housewife (0.521700)     - Strongest negative contributors: occupation\_category\_Healthcare (-0.293776), occupation\_category\_Office (-0.235986)     Interpretation: PC3 seems to differentiate between housewives with higher BMI and healthcare/office workers. This could represent a combination of occupation and body composition factors.  Key Insights:  1. Smoking status, body size (BSA and BMI), gender, and certain occupations (particularly housewife) are the most influential factors in distinguishing between patients in this dataset.  2. The first three principal components together explain only 33.74% of the total variance (14.43% + 11.35% + 7.96%). This suggests that the data is quite complex and cannot be easily reduced to just a few dimensions without losing a significant amount of information.  3. There appears to be a strong correlation between being a non-smoker and being a housewife in this dataset.  4. Gender (Sex\_F and Sex\_M) has a very strong influence on PC2, indicating that there are significant differences between male and female patients in this dataset.  5. BMI appears in all three top PCs, suggesting that body composition is a key factor in differentiating between patients across multiple dimensions.  6. Some occupations (like Healthcare, Office, and Housewife) seem to have more influence on the principal components than others, which could indicate lifestyle or environmental factors associated with these occupations that affect asthma patients.  These insights can guide further analysis:  1. We might want to investigate the relationship between smoking status, occupation (especially housewives), and asthma symptoms or severity.  2. The strong gender component suggests we should look at how asthma manifestation or treatment efficacy differs between males and females.  3. The importance of BMI across all components indicates that weight management could be a key factor in asthma management for this patient group.  Remember that while PCA is useful for dimensionality reduction and identifying key factors, it's always important to consider these mathematical insights in the context of clinical knowledge about asthma and patient care. |