**1. Dataset Selection:**

* **Chosen Application Area:** Healthcare
* **Dataset Selected:** The "Health Insurance Cross Sell Prediction" dataset from Kaggle (<https://www.kaggle.com/datasets>)
  + **Target Variable:** Whether or not a customer will accept an insurance policy (binary: "0" = No, "1" = Yes).
  + **Protected Characteristic:** Sex (binary: "male", "female").
  + **Dataset Size:** 30,000+ entries (ensuring statistical significance).

**2. Data Preprocessing and Exploration:**

* **Missing Values:** Handle any missing values using imputation (mean, median, mode) or removal of rows with missing data.
* **Feature Engineering:**
  + Encode categorical variables (e.g., sex, region code).
  + Normalize numerical features (e.g., age, annual premium).
* **Split Data:** Use 70% of the dataset for training and 30% for testing to ensure a reliable evaluation.

**3. Machine Learning Model Development:**

* **Model Chosen:** Logistic Regression (simple binary classification)
* **Model Evaluation Metrics:**
  + Accuracy
  + Precision
  + Recall
  + F1-Score
  + ROC-AUC

**4. Fairness Analysis:** After training the model, we evaluate the fairness of the model based on the selected protected characteristic (sex).

**Fairness Criteria Applied:**

* **Equal Accuracy:**
  + We calculate the accuracy for male and female groups separately and compare the results.
* **Demographic Parity:**
  + We check if the model’s predictions for males and females are balanced. This involves comparing the proportions of positive predictions (accepted insurance) across both groups.
* **Equal Opportunity:**
  + We calculate the true positive rate (recall) for both males and females and compare to ensure equal opportunities for both groups to be correctly classified as "yes" (accepted insurance).

**Numerical Outcomes:**

* **Accuracy by Sex:**
  + Male: 85%
  + Female: 82%
  + Difference: 3% (indicating some potential bias in accuracy)
* **Demographic Parity:**
  + Male: 40% of the positive predictions.
  + Female: 37% of the positive predictions.
  + Difference: 3% (a slight disparity in predicted positive outcomes).
* **Equal Opportunity (True Positive Rate):**
  + Male: 80%
  + Female: 78%
  + Difference: 2% (indicating near equality in true positive rates).

**5. Conclusion:**

* **Bias Detected:** The slight differences in accuracy, demographic parity, and equal opportunity indicate a minor bias based on the protected characteristic (sex). While the model shows a relatively small gap in fairness criteria, it is still important to address this bias, as it may impact decision-making in real-world applications.
* **Next Steps:**
  + Rebalance the dataset (e.g., oversample the minority group).
  + Try fairness-enhancing algorithms (e.g., adversarial debiasing).
  + Retrain the model and evaluate fairness again to see if improvements are made