Task 2: Data Cleaning & Exploration

**Load the Telco Customer Churn dataset.**

**Identify columns that may introduce bias (e.g., gender, senior citizen status, payment method).**

**Document any potential bias and how it may affect analysis.**

**📌 Expected Output: A list of columns that could introduce bias in the dataset.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **Column Name** | | **Potential Bias**   |  | | --- | |  | | **Why It Might Be Problematic**   |  | | --- | |  |      |  | | --- | |  | |
| gender | Gender bias | Certain genders may be overrepresented or underrepresented in churn patterns. |
| PaymentMethod | Financial bias | Married customers or those with dependents may have different service retention behavior. |
| Partner & Dependents | AI models are often trained on Western-centric data | AI might assume Christmas is universally celebrated but ignore other cultural holidays. |
| Tenure | |  | | --- | |  |  |  | | --- | | Experience bias | | Long-term customers may be treated differently than new customers. |
| Monthly Charges & Total Charges | Socioeconomic bias | High-income customers may afford expensive plans and have lower churn. |

**How Bias May Affect Analysis**

* **Gender Bias** → If one gender shows higher churn, models might assume the same for all customers.
* **Age Bias (Senior Citizen)** → Older customers might churn more due to **digital literacy** or **service needs**, but a biased model may unfairly penalize them.
* **Financial Bias (Payment Method & Monthly Charges)** → Certain payment methods or **higher bills** might correlate with churn, unfairly affecting low-income users.