**DATA CLEANING AND EXPLORATION**

**Potential bias and it’s affects :**

**1 . Gender**

* **Description:** The column representing the gender of customers (e.g., Male, Female).
* **Potential Bias:** If there is an unequal distribution of churn rates between genders, the model may inadvertently learn biased patterns. For example, if men are more likely to churn than women, this could skew the model's predictions**.**

**2. Age**

* **Description:** The age of the customer.
* **Potential Bias:** Age could introduce bias if certain age groups (e.g., older customers) are more likely to churn, or if there is an imbalance in the distribution of ages in the dataset.

**3. Tenure**

* **Description**: The number of months or years the customer has been with the telecom company.
* **Potential Bias**: New customers or those with short tenure may have higher churn rates than long-term customers. If this pattern is not carefully handled, it could introduce a bias where the model focuses too much on tenure as a factor for churn.

**4. Contract Type**

* **Description**: The type of contract the customer has (e.g., Month-to-Month, One Year, Two Year).
* **Potential Bias**: Customers on month-to-month contracts may be more likely to churn than those with long-term contracts. If contract type is not treated carefully, it can lead to bias where the model overemphasizes contract type in predicting churn.

**5. Payment Method**

* **Description**: The method by which the customer pays for services (e.g., Electronic Check, Credit Card, Mailed Check, Bank Transfer).
* **Potential Bias**: Certain payment methods might correlate with higher or lower churn, potentially leading to biased churn predictions.

Expected output :

|  |  |  |  |
| --- | --- | --- | --- |
| **Column** | **Description** | **Potential Bias** | **Impact** |
| Gender | The gender of the customer (e.g., Male, Female). | Unequal churn rates between genders may lead to biased predictions if one gender is overrepresented. | The model may be biased towards predicting churn for one gender over the other, leading to skewed results. |
| Age | The age of the customer**.** | Certain age groups may have different churn behaviors, and an imbalance in age distribution can lead to biased results. | The model may unfairly predict churn for certain age groups, leading to inaccurate predictions**.** |
| Payment Method | The method used by the customer to pay for services (e.g., Credit Card, Bank Transfer). | Certain payment methods may be associated with a higher or lower likelihood of churn, introducing bias. | The model may over- or under-predict churn based on payment method, leading to inaccuracies. |
| Tenure | The number of months the customer has been with the company. | Short-tenure customers may be more likely to churn, creating bias if the dataset is not balanced by tenure length. | The model may overemphasize tenure as a factor in predicting churn, affecting long-term customer predictions. |
| Contract Type | The type of contract the customer has (e.g., Month-to-Month, One Year, Two Year). | Month-to-month contracts may have a higher churn rate, leading to bias in churn predictions if not properly balanced. | The model may wrongly predict churn based on contract type, potentially ignoring other relevant factors. |
| Dependents | Whether the customer has dependents (e.g., Yes/No). | Churn patterns may differ between customers with or without dependents. If not represented properly, this could bias predictions. | The model may make incorrect churn predictions based on the presence or absence of dependents. |