Internship Report - Task 1: Data Cleaning

Overview

This report summarizes Task 1 completed as part of the Data Analysis Internship at RD Infro Technology. The goal of the task was to clean and prepare the Telco Customer Churn dataset for analysis.

Dataset Used

The dataset used in this task is the 'Telco Customer Churn' dataset. It contains information about telecommunication customers and whether or not they have churned.

Steps Performed

1. **Imported the Dataset**
   - Used Google Colab to upload and view the dataset using pandas.

2. **Checked for Missing Values**
   - Used `df.isnull().sum()` to detect missing values in the dataset.

3. **Data Type Conversion**
   - Converted the 'TotalCharges' column to numeric using `pd.to_numeric()`.
   - Handled coercion errors and dropped rows with missing 'TotalCharges' values using `dropna()`.

4. **Encoding Categorical Variables**
   - Applied one-hot encoding using `pd.get_dummies()` with `drop_first=True` to convert categorical columns.

5. **Final Dataset**
   - The cleaned dataset (`df_encoded`) was prepared, now ready for visualization and modeling.

Internship Report - Task 1: Data Cleaning

Python Code Used

```
# Step 1: Import the dataset
from google.colab import files
uploaded = files.upload()

import pandas as pd
df = pd.read_csv("Telco_Customer_Churn.csv")
df.head()

# Step 2: Check for missing values
print("Missing values in each column:")
print(df.isnull().sum())

# Step 3: Convert 'TotalCharges' to numeric and drop missing values
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'], errors='coerce')
df.dropna(inplace=True)

# Step 4: Encode categorical variables
df_encoded = pd.get_dummies(df, drop_first=True)

# Step 5: Display cleaned data
df_encoded.head()

# Output: Missing values in each column
# gender              0
# SeniorCitizen       0
# Partner             0
# Dependents          0
# tenure              0
# PhoneService        0
# ...
# TotalCharges        0
# Churn               0
```

Internship Report - Task 1: Data Cleaning

Additional Notes and Skills Practiced

Note:

This task is part of a larger internship project involving data analysis of customer churn. Future tasks will include data visualization, feature engineering, and predictive modeling using machine learning algorithms.

Skills Practiced:

- Python Programming
- Data Cleaning with Pandas
- Data Preprocessing
- Handling Missing Data
- Encoding Categorical Variables

The insights gained during this project have been instrumental in understanding the workflow of data science projects. Special thanks to RD Infro Technology for the opportunity to learn and grow in a professional setting.

Conclusion

This task provided practical experience with real-world data cleaning tasks using Python. Techniques like handling missing values, data type conversion, and categorical encoding were practiced. The cleaned data will serve as the foundation for further analysis in upcoming tasks.