Practical 2

Find the below data set and perform the following operations: Dataset name: - Churn_DataDescription

1. Find the no. of duplicate records in the churn dataframe based on the cutomerID column.
2. In the churn dataframe, what are the total no. of missing values for the variable TotalCharges?
3. From the churn dataframe, what is the average monthly charge paid by a customer for the services he/she has signed up for?
4. In the churn dataframe, under the variable Dependents how many records have "1@#" ?
5. Find the data type of the variable tenure from the churn dataframe.

1. Number of duplicate records based on CustomerID column:

```
import pandas as pd

# Load the dataset

churn = pd.read_csv('Churn_DataDescription.csv')

# Find duplicate records based on CustomerID column

duplicate_records = churn.duplicated(subset=['CustomerID']).sum()


print(f"Number of duplicate records based on CustomerID: {duplicate_records}")
```

2. Total number of missing values for the variable TotalCharges:

```
# Find total number of missing values for TotalCharges

missing_total_charges = churn['TotalCharges'].isnull().sum()

print(f"Total number of missing values for TotalCharges: {missing_total_charges}")
```

3. Average monthly charge paid by a customer:

```
Assuming the column representing monthly charges is named MonthlyCharges:
# Calculate average monthly charge
average_monthly_charge = churn['MonthlyCharges'].mean()
print(f"Average monthly charge paid by a customer: {average_monthly_charge}")
```

4. Number of records in the Dependents column that have "1@#":

```
# Count records in Dependents column that have "1@#"
dependents_count = (churn['Dependents'] == '1@#').sum()
print(f"Number of records in Dependents column with '1@#': {dependents_count}")
```

5. Find the data type of the variable tenure from the churn dataframe.

```
# Check the data type of the variable 'tenure'
tenure_dtype = churn['tenure'].dtype print(f"Data type of the variable 'tenure': {tenure_dtype}")
```