```
import pandas as pd
import numpy as np
import time


from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
df = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/adult.csv")


df
```

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	capit
0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	
1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	
2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	
3	53	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	
4	28	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	
...	...	...	...	...	...	...	...	...	...	...	...	
32556	27	Private	257302	Assoc-acdm	12	Married-civ-spouse	Tech-support	Wife	White	Female	0	
32557	40	Private	154374	HS-grad	9	Married-civ-spouse	Machine-op-inspct	Husband	White	Male	0	

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
def check_recursive_cl_diversity(df_group, c, l):
    counts = df_group["income"].value_counts()

    if len(counts) < l:
        return False

    sorted_counts = counts.sort_values(ascending=False)
    f1 = sorted_counts.iloc[0]

    sum_rest = sorted_counts.iloc[l-1:].sum()
    return f1 < c * sum_rest


def recursive_partition_cl(df_group, c, l, min_size):

    if len(df_group) < min_size:
        return []

    if check_recursive_cl_diversity(df_group, c, l):
        return [df_group]

    df_group = df_group.sort_values("age")
    mid_index = len(df_group) // 2
    group1 = df_group.iloc[:mid_index]
    group2 = df_group.iloc[mid_index:]


    groups1 = recursive_partition_cl(group1, c, l, min_size)
    groups2 = recursive_partition_cl(group2, c, l, min_size)

    return groups1 + groups2
```

```
c_value = 2.0
l_threshold = 2
min_group_size = 50

start_time = time.time()
final_groups = recursive_partition_cl(df, c_value, l_threshold, min_group_size)
end_time = time.time()


if final_groups:
    df_anonymized = pd.concat(final_groups)
else:
    df_anonymized = pd.DataFrame(columns=df.columns)

suppressed_count = len(df) - len(df_anonymized)


df_anonymized
```

	age	workclass	fnlwgt	education	education_num	marital_status	occupation	relationship	race	sex	capital_gain	cap
8413	31	Private	226443	Some-college	10	Married-civ-spouse	Craft-repair	Husband	White	Male	0	
10107	31	Private	341672	Bachelors	13	Married-civ-spouse	Adm-clerical	Husband	Asian-Pac-Islander	Male	0	
24918	31	Federal-gov	113688	HS-grad	9	Never-married	Machine-op-inspct	Unmarried	White	Female	0	
9389	31	Local-gov	144949	HS-grad	9	Divorced	Craft-repair	Unmarried	White	Male	0	
11147	31	Private	185480	HS-grad	9	Married-civ-spouse	Adm-clerical	Wife	White	Female	0	
...	...	...	...	...	...	...	...	...	...	...	...	
24043	90	Self-emp-not-inc	82628	HS-grad	9	Never-married	Exec-managerial	Not-in-family	White	Male	2964	
32277	90	Private	313749	HS-grad	9	Widowed	Adm-clerical	Unmarried	White	Female	0	

Next steps: Generate code with df_anonymized | View recommended plots | New interactive sheet

```
group_sizes = [len(g) for g in final_groups]
num_groups = len(final_groups)

print("=== Recursive (c, l)-Diversity Performance ===")
print(f"Original Dataset Size: {len(df)}")
print(f"Anonymized Dataset Size (Non-suppressed): {len(df_anonymized)}")
print(f"Suppressed Records: {suppressed_count}")
print(f"Number of Final Equivalence Classes: {num_groups}")
if group_sizes:
    print("Equivalence Class Size Distribution:")
    print(pd.Series(group_sizes).describe())
print(f"Total Processing Time: {end_time - start_time:.4f} seconds")


df_anonymized.to_csv("adult_recursive_cl_diverse.csv", index=False)
print("\nFinal anonymized dataset saved as 'adult_recursive_cl_diverse.csv'")
```

```
=== Recursive (c, l)-Diversity Performance ===
Original Dataset Size: 32561
Anonymized Dataset Size (Non-suppressed): 16981
Suppressed Records: 15580
Number of Final Equivalence Classes: 5
Equivalence Class Size Distribution:
count        5.000000
mean      3396.200000
std       7205.403021
min         63.000000
25%         64.000000
50%         64.000000
75%        509.000000
max      16281.000000
dtype: float64
Total Processing Time: 0.2581 seconds

Final anonymized dataset saved as 'adult_recursive_cl_diverse.csv'
```