**STELLA IBOR**

**APRIL COHORT**

## **Analysis of Census Data**

**Question 1 -** Which state has the most countries in it? future questions too...)

**Answer:**

# Group the data by state and count the number of unique counties in each state

state\_counties = Census.groupby('STNAME')['CTYNAME'].nunique()

# Find the state with the maximum number of counties

state\_with\_most\_counties = state\_counties.idxmax()

print(state\_with\_most\_counties)

**Texas, Georgia, Virginia came top of the chat while District of Columbia, Delware were the least counties.**

**Question 2 -**Only looking at the three most populous countries for each state, what are the three most populous

states (in order of highest population to lowest population)? Use CENSUS2010POP

**Answer:**

# Group the data by state and county and sum the population for each countycounty\_populations = Census.groupby(['STNAME', 'CTYNAME'])['CENSUS2010POP'].sum()

# Group the data by state and sort the counties by population in descending orderstate\_populations = county\_populations.groupby('STNAME').apply(lambda x: x.sort\_values(ascending=False)[:3])

# Calculate the total population for each state and sort in descending order

state\_total\_populations = state\_populations.groupby('STNAME').sum().sort\_values(ascending=False)

# Get the top three most populous states

top\_3\_states = state\_total\_populations.head(3)

# Display the three most populous states

print(top\_3\_states)

**We have California, Texas and New York as the three most populus countries.**

**Question 3 -**Which city has the most countries in it?

**Answer:**

# Group the data by city and county and count the number of counties in each city

city\_counties = Census.groupby(['CTYNAME', 'STNAME'])['COUNTY'].count()

# Reset the index of the grouped data

city\_counties = city\_counties.reset\_index()

# Find the city with the maximum number of counties

city\_with\_most\_counties = city\_counties.loc[city\_counties['COUNTY'].idxmax()]

# Display the city with the most counties

print("City with the most counties:", city\_with\_most\_counties['CTYNAME'])

print("State:", city\_with\_most\_counties['STNAME'])

print("Number of counties:", city\_with\_most\_counties['COUNTY'])

**City and State with the most counties is District of Columbia having 2 counties.**

**Question 4 -**Which region has the most division in it?

**Answer:**

# Group the data by region and division and count the number of divisions in each region

region\_divisions = Census.groupby(['REGION', 'DIVISION'])['DIVISION'].count()

# Reset the index of the grouped data and rename the count column

region\_divisions = region\_divisions.reset\_index(name='DIVISION\_COUNT')

# Find the region with the most divisions

top\_region = region\_divisions.sort\_values('DIVISION\_COUNT', ascending=False).iloc[0]

# Print the result

print(f"The region with the most divisions is Region {top\_region['REGION']} (Division {top\_region['DIVISION']}) with {top\_region['DIVISION\_COUNT']} divisions.")

**Region 2 (Division 4) has the highest number of divisions**