# Dollars & Data:
# Analyzing US Household Income with MySQL

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Introduction:

- Household income is one of the most important indicators of economic well-being in the United States.
- It directly impacts living standards, access to healthcare, education, housing affordability, and social mobility.
- In recent decades, trends in household income have been influenced by inflation, wage stagnation, globalization, technological change, and demographic shifts.
- Understanding income distribution patterns helps policymakers, researchers, and businesses design better economic strategies and social programs.

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Project Overview:

- The project focuses on analyzing U.S. household income data to uncover patterns, disparities, and long-term trends.
- Data sources may include U.S. Census Bureau (Current Population Survey, American Community Survey), Bureau of Labor Statistics, and Federal Reserve reports.
- The analysis covers aspects such as:

➢ Median household income (national, regional, state-level)
➢ Income inequality (Gini coefficient, top vs. bottom percentiles)
➢ Demographic breakdowns (race, gender, age, education, employment sector)
➢ Cost of living vs. income growth

- The ultimate goal is to provide data-driven insights into how income is distributed and what factors drive disparities.

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Problem Statement:

- Despite being one of the wealthiest nations, the U.S. faces growing income inequality and stagnant wages for middle- and lower-income households.
- Median household income growth has not kept pace with inflation, reducing real purchasing power.
- Wealth is increasingly concentrated in the top income brackets, while many households struggle with basic affordability (housing, healthcare, education).
- Disparities in income are highly correlated with race, education, and geographic region, reflecting deep-rooted structural inequalities.
- Without accurate data analysis, policies may fail to address the real challenges of income disparity, leaving vulnerable populations behind.

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Unstructured Data:

| row_id | id | State_Code | State_Name | State_ab | County | City | Place | Type | Primary | Zip_Code | Area_Code | ALand | AWater | Lat | Lon |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1026 | 1 | Alabama | AL | Autauga County | Elmore | Autaugaville | Track | Track | 36025 | 334 | 8020338 | 60048 | 32.4473511 | -86.4768097 |
| 2 | 10216 | 1 | Alabama | AL | Autauga County | Robertsdale | Autaugaville | Track | Track | 36567 | 251 | 737211648 | 3860933 | 30.7200267 | -87.6245437 |
| 3 | 10226 | 1 | Alabama | AL | Autauga County | Silverhill | Autaugaville | Track | Track | 36576 | 251 | 71113244 | 190587 | 30.5345606 | -87.7548736 |
| 4 | 10236 | 1 | Alabama | AL | Autauga County | Orange Beach | Autaugaville | Track | Track | 36561 | 251 | 15491986 | 20427550 | 30.3005856 | -87.5417985 |
| 5 | 10246 | 1 | Alabama | AL | Autauga County | Louisville | Autaugaville | Track | Track | 36048 | 334 | 517008078 | 983621 | 31.9222753 | -85.4498309 |
| 6 | 10256 | 1 | Alabama | AL | Autauga County | Brent | Autaugaville | Track | Track | 35034 | 205 | 691793540 | 3983687 | 32.9533249 | -87.0043466 |
| 7 | 10266 | 1 | alabama | AL | Autauga County | Remlap | Autaugaville | Track | Track | 35133 | 205 | 203716769 | 1004680 | 33.8405764 | -86.6350183 |
| 8 | 10276 | 1 | Alabama | AL | Autauga County | Georgiana | Autaugaville | Track | Track | 36033 | 334 | 352261684 | 177369 | 31.7163243 | -86.7533225 |
| 9 | 10286 | 1 | Alabama | AL | Autauga County | Anniston | Autaugaville | Track | Track | 36207 | 256 | 12116269 | 23502 | 33.6448084 | -85.796557 |
| 10 | 10296 | 1 | Alabama | AL | Autauga County | Jacksonville | Autaugaville | Track | Track | 36265 | 256 | 41129558 | 487657 | 33.7661796 | -85.7864612 |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Summary of Problems:

- Duplicate/confusing columns (row_id vs id, Type vs Primary).
- Ambiguous names (Place, Type, Primary, Area_Code).
- Inconsistent naming conventions (snake_case + PascalCase mix).
- Abbreviations that aren’t self-explanatory (ALand, AWater, Lat, Lon).
- Case-sensitivity inconsistency in values (Alabama vs alabama).

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

Q1: Show me all the unique state names from the table us_household_income.

| State_Name |
| --- |
| Alabama |
| Alaska |
| Arizona |
| Arkansas |
| California |
| Colorado |
| Connecticut |
| Delaware |
| District of Columbia |
| Florida |
| Georgia |
| georia |
| Hawaii |
| Idaho |
| Illinois |
| Indiana |
| Iowa |
| Kansas |
| Kentucky |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

Q2: Write a SQL query to display each state name along with the count of rows (places/records) for that state in the us_household_income table.

| State_Name | place_count |
|---|---|
| Alabama | 527 |
| Alaska | 87 |
| Arizona | 606 |
| Arkansas | 340 |
| California | 3268 |
| Colorado | 513 |
| Connecticut | 355 |
| Delaware | 88 |
| District of Columbia | 64 |

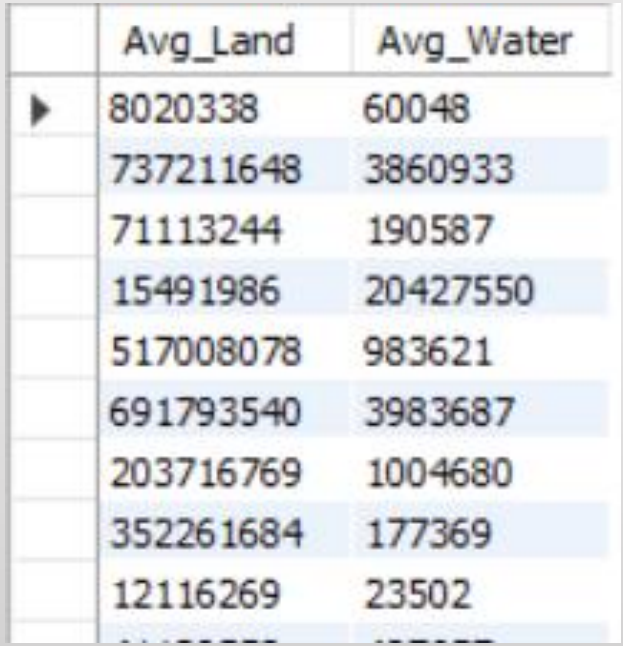
Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

Q3: Write a SQL query to list all states along with the number of places they contain, and display them in descending order of place count.

| State_Name | place_count |
|---|---|
| California | 3268 |
| Texas | 2275 |
| New York | 2161 |
| Florida | 1658 |
| Pennsylvania | 1475 |
| Illinois | 1431 |
| Ohio | 1349 |
| Michigan | 1171 |
| North Carolina | 915 |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

Q4: Write a SQL query to display the Avg_Land and Avg_Water values from the us_household_income table.



| Avg_Land | Avg_Water |
|---|---|
| 8020338 | 60048 |
| 737211648 | 3860933 |
| 71113244 | 190587 |
| 15491986 | 20427550 |
| 517008078 | 983621 |
| 691793540 | 3983687 |
| 203716769 | 1004680 |
| 352261684 | 177369 |
| 12116269 | 23502 |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

Q5: Write a SQL query to calculate the average land area and average water area for each state in the us_household_income table, rounded to two decimal places.

| State_Name | avg_land_area | avg_water_area |
|---|---|---|
| Alabama | 108850171.23 | 4587846.74 |
| Alaska | 142930048.80 | 41725737.98 |
| Arizona | 89332785.31 | 279511.36 |
| Arkansas | 171057374.72 | 3497673.13 |
| California | 27679818.89 | 1182868.28 |
| Colorado | 87163383.73 | 738978.45 |
| Connecticut | 16466399.97 | 1250218.31 |
| Delaware | 23022086.76 | 7650982.52 |
| District of Columbia | 3420132.58 | 379241.31 |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

Q6: Write a SQL query to display, for each state, the largest and smallest land area values and the largest and smallest water area values from the us_household_income table.

| State_Name | largest_land_area | smallest_land_area | largest_water_area | smallest_water_area |
|---|---|---|---|---|
| Alabama | 973585771 | 0 | 780526158 | 0 |
| Alaska | 1908739435 | 482353 | 1396368032 | 0 |
| Arizona | 1896284469 | 290861 | 39469850 | 0 |
| Arkansas | 1371775477 | 287351 | 104495490 | 0 |
| California | 2031989773 | 0 | 869682228 | 0 |
| Colorado | 2075296718 | 396151 | 39406513 | 0 |
| Connecticut | 157070494 | 0 | 93371616 | 0 |
| Delaware | 295740336 | 0 | 534631228 | 0 |
| District of Columbia | 158364992 | 204236 | 18633403 | 0 |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Structured Data:

| row_id | id | State_Code | State_Name | County_Information | County | Full_Address | Type | Primary | Area_Code | Avg_Land | Avg_Water | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1026 | 1 | Alabama | NULL | Autauga County | Elmore, Autaugaville, AL, 36025 | Track | Track | 334 | 8020338 | 60048 | 32 | -86 |
| 2 | 10216 | 1 | Alabama | NULL | Autauga County | Robertsdale, Autaugaville, AL, 36567 | Track | Track | 251 | 737211648 | 3860933 | 31 | -88 |
| 3 | 10226 | 1 | Alabama | NULL | Autauga County | Silverhill, Autaugaville, AL, 36576 | Track | Track | 251 | 71113244 | 190587 | 31 | -88 |
| 4 | 10236 | 1 | Alabama | NULL | Autauga County | Orange Beach, Autaugaville, AL, 36561 | Track | Track | 251 | 15491986 | 20427550 | 30 | -88 |
| 5 | 10246 | 1 | Alabama | NULL | Autauga County | Louisville, Autaugaville, AL, 36048 | Track | Track | 334 | 517008078 | 983621 | 32 | -85 |
| 6 | 10256 | 1 | Alabama | NULL | Autauga County | Brent, Autaugaville, AL, 35034 | Track | Track | 205 | 691793540 | 3983687 | 33 | -87 |
| 7 | 10266 | 1 | alabama | NULL | Autauga County | Remlap, Autaugaville, AL, 35133 | Track | Track | 205 | 203716769 | 1004680 | 34 | -87 |
| 8 | 10276 | 1 | Alabama | NULL | Autauga County | Georgiana, Autaugaville, AL, 36033 | Track | Track | 334 | 352261684 | 177369 | 32 | -87 |
| 9 | 10286 | 1 | Alabama | NULL | Autauga County | Anniston, Autaugaville, AL, 36207 | Track | Track | 256 | 12116269 | 23502 | 34 | -86 |
| 10 | 10296 | 1 | Alabama | NULL | Autauga County | Jacksonville, Autaugaville, AL, 36265 | Track | Track | 256 | 41129558 | 487657 | 34 | -86 |

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Summary of Structured Data:

1. row_id ek surrogate key hai jo har row ko uniquely identify karta hai.
2. id dataset ka actual unique identifier hai jo source se aaya hai.
3. State_Code ek numeric code hai jo state ko represent karta hai.
4. State_Name me state ka naam hai lekin casing inconsistent hai (Alabama vs alabama).
5. County_Information mostly NULL hai aur redundant lag raha hai.
6. County county ka naam store karta hai.
7. Full_Address composite field hai jisme city, county, state aur zip combine hain.
8. Type aur Primary dono hi "Track" repeat kar rahe hain aur extra lag rahe hain.
9. Area_Code census ka geographic area code hai, phone code nahi.
10. Avg_Land land area value hai, lekin naam misleading hai (better hota Land_Area).
11. Avg_Water water area value hai, iska bhi naam misleading hai (better hota Water_Area).
12. Latitude aur Longitude coordinates hain mapping aur GIS analysis ke liye.

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst

# Final Takeaways:

- Household income isn’t just numbers, it’s quality of life. Income trends directly shape affordability, opportunity, and long-term economic security.
- Inequality is widening. Wealth is concentrating at the top, while middle- and lower-income households face wage stagnation and rising costs of living.
- Demographics matter. Race, education, gender, and geography all play a huge role in shaping income outcomes, reflecting systemic imbalances.
- Real vs. nominal income. Even when incomes rise on paper, inflation often erodes actual purchasing power.
- Data is power. Careful analysis of household income data helps identify structural issues, forecast trends, and guide evidence-based policymaking.
- Future focus: Without addressing disparities, income inequality could threaten social mobility, economic stability, and trust in institutions.

Created by: Syed Daniyal Ahmed
Business Intelligence & Data Analyst