FINAL PROJECT DELIVERABLE #1 - PROJECT PROPOSAL

Team Members: Shilpa Sweth, Veda Sree Bojanapally, Ritvika Pillai, Samuel Jayachandran

1 - Description of the dataset

This dataset is collated using information present in the Annual Social and Economic Supplement (ASEC) which is a part of the Current Population Survey (CPS). It sources up-to-date, periodically-organized, official, national estimates of the following socioeconomic factors: poverty levels and rates of widely used measures of income. These data are obtained from the microdata files from the Current Population Survey (CPS) consisting of data concerning: families, household composition, educational attainment, income sources, poverty, health insurance coverage, geographic mobility. We have three tables: family, person and household, where person and family have been mapped to the household data with respect to their keys.

Questions highlighting the social and economic characteristics of every person who is a member of the household as of date of the interview have been asked in the survey. The questions related to income take into consideration the income that has been received in the previous year. The information acquired by measuring factors like family income and number of people in that particular household helps in indicating the poverty status.

2 - Users identification

1. **Public Affairs Research Analyst -**
    a. **Background -** A public affairs research analyst may work for government agencies, think tanks, and other organizations focused on areas such as defense, health care, energy, or environmental protection. A person in this role applies political, economic, and scientific research techniques to determine the benefits and drawbacks of current and proposed public policies. They advise legislators, government leaders, and industry executives on policy matters. They collect and analyze information to draw meaningful insights for presentation to decision-makers.
    b. **Motivations -** To provide visualization aid in the form of a year-end report to the higher officials.
    c. **Tasks -** To draw insightful conclusions from the data and provide in-depth analysis on government policies. Also, to get a comprehensive view of the pulse of the households in a particular region.

d. **Context of use -** After performing an exploratory analysis on the data, the user will then determine which snapshots of information they wish to present as a sequential data story.
e. **Challenges -** The analyst would be constrained in conveying insights from their exploratory analysis by only using a concise report.

2. **A worker being transferred to US as part of their job -**
    a. **Background -** An individual who's recently shifting to the United States and wants to compare the housing rates, taxes, safety, etc of various regions.
    b. **Motivations -** To find the best possible housing options (according to his needs) that are not only within his budget but also sound from a safety point of view.
    c. **Tasks -** To extract variations in the housing prices across different regions over the years. Also, to study the various factors that might influence the individual in buying a house in that region.
    d. **Context of use -** After studying the visualizations, the individual decides whether they would consider moving to that region. This visualization would help them in deciding without having to spend a lot of time researching ideal housing options on several websites.
    e. **Challenges -** Might be a bit difficult for the individual to initially draw understanding from the visualizations as he/they might not be very familiar with data visualizations.

3. **Real Estate Analyst**
    a. **Background -** A Real estate analyst would manage the real estate investments of organizations. A person in this role would monitor the changes in the market and decide whether to buy or sell in a particular region.
    b. **Motivations -** To maximize the profit by investing in a property with a prospect of attracting more buyers.
    c. **Tasks -** sell the properties which are showing a stagnant/ reducing prices
    d. **Context of use -** The analyst would consult visualizations to make informed decisions concerning real estate transactions with an aim to increase their profitability.
    e. **Challenges -** Forecasting always carries the risk of uncertainty, hence the analyst would benefit from not merely relying only on the visualizations.

3 - Task identification

1. **Public Affairs Research Analyst -**
    a. Gain an understanding of poverty rate trends over the years.

b. Study the correlation between income and expenditure.
c. Distribution of people across different education levels.

2. **A worker being transferred to US as part of their job -**
    a. Identify candidate areas suited to comfortably live with their current income.
    b. Is it more economic to buy a house or instead rent a house, in the long run?
    c. For each state, what is the most frequent reason for a person to move from their current locality?

3. **Real Estate Analyst -**
    a. Examine the price of properties in a region over the years, to decide whether to sell a property or not
    b. Check if there is a correlation between the income, income tax and housing prices
    c. To analyze the number of people migrating from a region so as to avoid investing much in such regions

4 - Data and task abstraction

4.1 - Data Abstraction

4.1.1 Dataset

- The datasets are of type ‘Table’

4.1.2 Attributes

Attribute	Table	Description	Value Range	Type
H_HHNUM	Household	Household number. Identifier for unique set of residents located at this sample address.	1:8	Categorical
H_IDNUM	Household	Household id number.	ID	Categorical
H_SEQ	Household	Household sequence number	00001:99999	Categorical

GEDIV	Household	Census division of current residence	0:9	Categorical
GEREG	Household	Region	1:4	Categorical
GESTFIPS	Household	State FIPS code	01:56	Categorical
GTCO	Household	Fips county code	000:810	Categorical
H_LIVQRT	Household	Type of living quarters	01:12	Categorical
HEFAMINC	Household	Family income from basic CPS income screener question.	-1:16	Categorical
H_YEAR	Household	Year of survey	1999:2999	Categorical
HHINC	Household	Total household income	0:41	Categorical
HTOTVAL	Household	Total household income	-999999:99999999	Quantitative
HEARNVAL	Household	Total household earnings	-999999:99999999	Quantitative
HPROP_VAL	Household	Estimate of current property value	-1:9999999	Quantitative
FFPOS	Family	Unique family identifier.	01:16	Categorical
FH_SEQ	Family	Household sequence number	00001:99999	Categorical
FTOT_R	Family	Total family income	0:41	Categorical
FTOTVAL	Family	Total family income	-999999:99999999	Quantitative
FEARNVAL	Family	Total family earnings	-999999:99999999	Quantitative
FAMLIS	Family	Ratio of family income to poverty threshold	-1:4	Ordinal
FHIP_VAL	Family	Total amount paid in premiums by family	0:9999999	Quantitative

FHIP_VAL2	Family	Total amount paid in premiums by family 2	0:9999999	Quantitative
FMED_VAL	Family	Total amount paid in medical expenses by family	0:9999999	Quantitative
FMOOP	Family	Family's total medical out of pocket expenditures	0:9999999	Quantitative
FMOOP2	Family	Family's total medical out of pocket expenditures with alternative measure of premiums	0:9999999	Quantitative
FOTC_VAL	Family	Total amount paid in over the counter expenses by family	0:9999999	Quantitative
P_SEQ	Person	Sequence number of person in hhld	00:16	Categorical
PERIDNUM	Person	22-digit Unique Person identifier	ID	Categorical
PF_SEQ	Person	Pointer to the sequence number of family record in household	00:16	Categorical
PH_SEQ	Person	Household seq number	00000:99999	Categorical
PPPOS	Person	Person identifier	41:79	Categorical
A_HGA	Person	Educational attainment	0:46	Categorical
A_SEX	Person	Sex	1:2	Categorical
AGE1	Person	Age - Persons 15+ years	0:17	Ordinal
PRCITSHP	Person	Citizenship group	-4:5	Categorical
A_MJIND	Person	Major industry code	-1:14	Categorical
A_MJOCC	Person	Major occupation	-1:11	Categorical

PEIOIND	Person	Industry	0:9999	Categorical
PEIOOCC	Person	Occupation	-1:9999	Categorical
A_GRSWK	Person	Earning per week at this job before deductions	0:2885	Quantitative
PTOT_R	Person	Total person income	0:41	Categorical
PTOTVAL	Person	Total persons income	-99999:99999999	Quantitative
STATETAX_A	Person	State income tax liability, after all credits	-9999:9999999	Quantitative
PERLIS	Person	Poverty level of persons	-1:4	Ordinal
MOOP	Person	Total medical out of pocket expenditures.	0:9999999	Quantitative
MOOP2	Person	Total medical out of pocket expenditures	0:9999999	Quantitative
PEMCPREM	Person	Edited Medicare premium amount	0000:99999	Quantitative
PHIP_VAL	Person	Out of pocket expenditures for comprehensive and noncomprehensive health insurance premiums	0:9999999	Quantitative
PHIP_VAL2	Person	Out of pocket expenditures for comprehensive and noncomprehensive health insurance premiums - alternative	0:9999999	Quantitative
PMED_VAL	Person	Out of pocket expenditures for non-premium medical care	0:9999999	Quantitative
POTC_VAL	Person	Out of pocket expenditures for over the counter health related spending	0:9999999	Quantitative
SPM_FedTax	Person	SPM unit's Federal tax	-999999:9999999	Quantitative

SPM_MedXpns	Person	SPM unit's Medical Out-of-Pocket (MOOP) and Medicare Part B subsidy	0:9999999	Quantitative
SPM_Poor	Person	SPM poverty status	0:1	Categorical
MIG_DIV	Person	Census division of previous year residence	0:10	Categorical
MIG_ST	Person	FIPS State code of previous residence	0:96	Categorical
NXTRES	Person	What was main reason for moving?	0:20	Categorical

4.2. Task Abstraction

User 1 : Public Affairs Research Analyst -

1. **Gain an understanding of poverty rate trends over the years.**
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, PERLIS, FAMLIS would be needed to answer this question
    c. We would not need any extra data or transformation for this in addition to the currently presented data
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Poverty Bucket' and 'Compare Trends'
2. **Study the correlation between income and expenditure**
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, HEARNVAL, HPROP_VAL, FEARNVAL, FHIP_VAL, FMOOP, FMOOP2, PTOTVAL, MOOP, MOOP2, PHIP_VAL, PHIP_VAL2, PMED_VAL, POTC_VAL, SPM_Poor
    c. The data that we have is sufficient to answer the question but it would be also useful to have all miscellaneous expenses as well
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Income and Expenditure related columns', 'Summarize Expenditure', and 'Compare Trends'
3. **Distribution of people across different education levels**
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, A_HGA
    c. We would not need any extra data or transformation for this in addition to the currently presented data
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Education Levels', 'Explore Distribution'

User 2 : A worker being transferred to US as part of their job -

1. Identify candidate areas suited to comfortably live with their current income
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, PTOTVAL, HEARNVAL, MIG_ST, SPM_Poor, PERLIS, FAMLIS, HPROP_VAL
    c. The data that we have is sufficient to answer the question but it would be also useful to have all miscellaneous expenses as well
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Income, Poverty Bucket', 'Lookup Expenditure', 'Lookup Property Rates', 'Summarize Expenditure', 'Compare Trends'
2. Is it more economic to buy a house or instead rent a house, in the long run
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, H_LIVQRT, HPROP_VAL
    c. We would not need any extra data or transformation for this in addition to the currently presented data
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Property Rates', 'Lookup Residence Types', 'Query Trends'
3. For each state, what is the most frequent reason for a person to move from their current locality
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, NXTRES
    c. We would not need any extra data or transformation for this in addition to the currently presented data
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Migration Reason', 'Identify Frequency distribution' and 'Locate Extremes'

User 3: Real Estate Analyst -

1. Examine the price of properties in a region over the years, to decide whether to sell a property or not
    a. We would need a chart to answer this question
    b. GEREG, H_YEAR, HPROP_VAL
    c. We would need no extra data or transformation for this data in addition to currently presented data
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Property Rates', 'Identify Trends'
2. Check if there is a correlation between the income, income tax and housing prices
    a. We would need a chart to answer this question
    b. GESTFIPS, H_YEAR, FEARNVAL, HEARNVAL, PTOTVAL, HPROP_VAL
    c. We would need no extra data or transformation for this data in addition to currently presented data
    d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Property Rates', 'Lookup Income', 'Lookup Income Tax', and 'Derive Dependency'
3. To analyze the number of people migrating from a region so as to avoid investing much in such regions

a. We would need a chart to answer this question
b. GESTFIPS, H_YEAR, HPROP_VAL, MIG_ST
c. We would not need any extra data or transformation for this data in addition to currently presented data
d. **Munzner's taxonomy -** This question can be mapped to 'Lookup Migration', 'Lookup Property Rates', 'Identify Trends'