In [1]:
```
#Using Python to clean the data
#Importing pandas
import pandas as pd
```

In [2]:
```
#Importing the Census's 5 Year 2020 Income Data CSV file.
#The link for the CSV file is below.
#https://data.census.gov/cedsci/table?q=income&tid=ACSST1Y2021.S1901&moe=false&tp=true
Income2020 = pd.read_csv('Income2020.csv')
Income2020
```

Out[2]:

	GEO_ID	NAME	S1901_C01_001E	S1901_C01_001EA	S1901_C01_001M	S1901_C01_001MA	S
0	Geography	Geographic Area Name	Estimate!!Households!!Total	Annotation of Estimate!!Households!!Total	Margin of Error!!Households!!Total	Annotation of Margin of Error!!Households!!Total	Estimate!!Househ
1	0100000US	United States	122354219	NaN	211970	NaN	

2 rows × 259 columns

In [3]:
```
#Dropping columsn with NaN values
Income2020 = Income2020.dropna(axis = 1)
#Setting the columns names to the first row's values
Income2020.columns=Income2020.iloc[0]
#Dropping columns with names that include the string 'Margin of Error'
Income2020 = Income2020[Income2020.columns.drop(list(Income2020.filter(regex='Margin of Error')))]
#Dropping columns that have (X) values
Income2020 = Income2020.loc[:, ~(Income2020 == '(X)').any()]
#Updating the Income2020 dataframe to only include the 2nd row and the 3rd column and beyond.
Income2020 = Income2020.iloc[1:, 2:15]
Income2020
```

Out[3]:

	Estimate!!Households!!Total	Estimate!!Households!!Total!!Less than $10,000	Estimate!!Households!!Total!!$10,000 to 14,999	Estimate!!Households!!Total!!$15,000 to 24,999
1	122354219	5.8	4.1	8.5

In [4]:
```
Income2020 = Income2020.rename(
columns = {Income2020.columns[0]:'Total Households',
Income2020.columns[1]:'Household Income less than $10k',
Income2020.columns[2]:'Household Income $10k to $14.999k',
Income2020.columns[3]:'Household Income $15k to $24,999k',
Income2020.columns[4]:'Household Income $25k to $34.999k',
Income2020.columns[5]:'Household Income $35k to $49.999k',
Income2020.columns[6]:'Household Income $50k to $74.999k',
Income2020.columns[7]:'Household Income $75k to $99.999k',
Income2020.columns[8]:'Household Income $100k to $149.999k',
Income2020.columns[9]:'Household Income $150k to $199.999k',
Income2020.columns[10]:'Household Income $200k or more',
Income2020.columns[11]:'Household Median Income',
Income2020.columns[12]:'Household Mean Income',
})
Income2020
```

Out[4]:

	Total Households	Household Income less than $10k	Household Income $10kto 14.999k	Household Income $15kto 24,999k	Household Income $25kto 34.999k	Household Income $35kto 49.999k	Household Income $50kto 74.999k	Household Income $75kto 99.999k	Household Income $100kto 149.999k	Household Income $150kto 199.999k	Household Income $200k or more	Hou M I
1	122354219	5.8	4.1	8.5	8.6	12.0	17.2	12.8	15.6	7.1	8.3	

In [5]:
```
#Exporting the Income2020 dataframe to a CSV
Income2020.to_csv('CleanIncome2020.csv')
```

In [ ]:
```
%%sql
--Renaming the CleanIncome2020 table columns
EXEC sp_rename 'CleanIncome2020.column2', 'Household Income less than $10k','Column'
EXEC sp_rename 'CleanIncome2020.column3', 'Household Income $10k to $14.999k','Column'
EXEC sp_rename 'CleanIncome2020.column4', 'Household Income $15k to $24,999k','Column'
EXEC sp_rename 'CleanIncome2020.column5', 'Household Income $25k to $34.999k','Column'
EXEC sp_rename 'CleanIncome2020.column6', 'Household Income $35k to $49.999k','Column'
EXEC sp_rename 'CleanIncome2020.column7', 'Household Income $50k to $74.999k','Column'
EXEC sp_rename 'CleanIncome2020.column8', 'Household Income $75k to $99.999k','Column'
EXEC sp_rename 'CleanIncome2020.column9', 'Household Income $100k to $149.999k','Column'
EXEC sp_rename 'CleanIncome2020.column10', 'Household Income $150k to $199.999k','Column'
EXEC sp_rename 'CleanIncome2020.column11', 'Household Income $200k or more','Column'
EXEC sp_rename 'CleanIncome2020.column12', 'Household Median Income','Column'
EXEC sp_rename 'CleanIncome2020.column13', 'Household Mean Income','Column'
```

In [ ]:
```
%%sql
--Showing all of the rows and columns in the CleanIncome2020 table
SELECT *
FROM CleanIncome2020;
```

In [ ]:
```
%%sql
--Creating the Mean-Median Differential by subtracting the Household Median Income from the Household Mean Income
SELECT CleanIncome2020.[Household Mean Income] - CleanIncome2020.[Household Median Income] as [Mean-Median Differential
FROM CleanIncome2020;
```

In [ ]:
```
%%sql
/*2022 POVERTY GUIDELINES FOR THE 48 CONTIGUOUS STATES AND THE DISTRICT OF COLUMBIA
$27,750 for a household with 4 family members
Source: https://aspe.hhs.gov/topics/poverty-economic-mobility/poverty-guidelines
Using this dollar number to see the total percentage of people in brackets that are lower than or contain $27,750*/

/*Selecting the Rough Percentages for the Household Income Brackets that are
1.below or near the Poverty Line
2.greater than the Poverty Line*/
SELECT CONCAT(ROUND(
CleanIncome2020.[Household Income less than $10k] +
CleanIncome2020.[Household Income $10k to $14.999k] +
CleanIncome2020.[Household Income $15k to $24,999k] +
CleanIncome2020.[Household Income $25k to $34.999k]
,2),'%') AS [Rough Percentage of Household Income Brackets that are inside or near the Poverty Line],
CONCAT(100 -
ROUND(
CleanIncome2020.[Household Income less than $10k] +
CleanIncome2020.[Household Income $10k to $14.999k] +
CleanIncome2020.[Household Income $15k to $24,999k] +
CleanIncome2020.[Household Income $25k to $34.999k]
,2),'%') AS [Rough Percentage of Household Income Brackets that are NOT inside or near the Poverty Line]
FROM CleanIncome2020;
```

In [ ]:
```
%%sql
--Ensuring that the percentages of Household Incomes equate to 100%
SELECT CONCAT(
                ROUND(
CleanIncome2020.[Household Income less than $10k] +
CleanIncome2020.[Household Income $10k to $14.999k]+
CleanIncome2020.[Household Income $15k to $24,999k]+
CleanIncome2020.[Household Income $25k to $34.999k]+
CleanIncome2020.[Household Income $35k to $49.999k]+
```

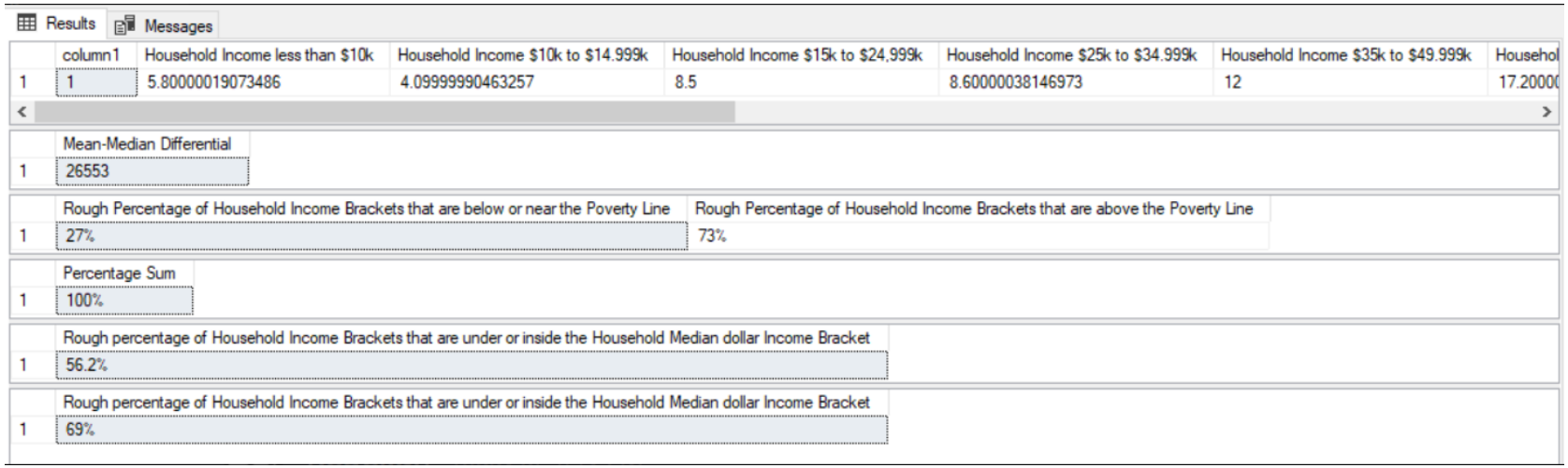
```
CleanIncome2020.[Household Income $50k to $74.999k]+
CleanIncome2020.[Household Income $75k to $99.999k]+
CleanIncome2020.[Household Income $100k to $149.999k]+
CleanIncome2020.[Household Income $150k to $199.999k]+
CleanIncome2020.[Household Income $200k or more], 2), '%')
AS [Percentage Sum]
FROM CleanIncome2020;
```

In [ ]:

```
%%sql
--Rough Percentage of Household Income Brackets that are under or inside the Household Median Income ($64,994) Bracket
SELECT CONCAT(
                ROUND(
CleanIncome2020.[Household Income less than $10k] +
CleanIncome2020.[Household Income $10k to $14.999k]+
CleanIncome2020.[Household Income $15k to $24,999k]+
CleanIncome2020.[Household Income $25k to $34.999k]+
CleanIncome2020.[Household Income $35k to $49.999k]+
CleanIncome2020.[Household Income $50k to $74.999k],2), '%')
AS [Rough percentage of Household Income Brackets that are under or inside the Household Median dollar Income Bracket]
FROM CleanIncome2020;
```

In [ ]:

```
%%sql
--Rough Percentage of Household Income Brackets that are under or inside the Household Mean Income ($91,547) Bracket
SELECT CONCAT(
                ROUND(
CleanIncome2020.[Household Income less than $10k] +
CleanIncome2020.[Household Income $10k to $14.999k]+
CleanIncome2020.[Household Income $15k to $24,999k]+
CleanIncome2020.[Household Income $25k to $34.999k]+
CleanIncome2020.[Household Income $35k to $49.999k]+
CleanIncome2020.[Household Income $50k to $74.999k]+
CleanIncome2020.[Household Income $75k to $99.999k],2), '%')
AS [Rough percentage of Household Income Brackets that are under or inside the Household Median dollar Income Bracket]
FROM CleanIncome2020;
```



Results | Messages

	column1	Household Income less than $10k	Household Income $10k to $14.999k	Household Income $15k to $24.999k	Household Income $25k to $34.999k	Household Income $35k to $49.999k	Household
1	1	5.80000019073486	4.09999990463257	8.5	8.60000038146973	12	17.2000

	Mean-Median Differential
1	26553

	Rough Percentage of Household Income Brackets that are below or near the Poverty Line	Rough Percentage of Household Income Brackets that are above the Poverty Line
1	27%	73%

	Percentage Sum
1	100%

	Rough percentage of Household Income Brackets that are under or inside the Household Median dollar Income Bracket
1	56.2%

	Rough percentage of Household Income Brackets that are under or inside the Household Median dollar Income Bracket
1	69%