Table of Contents

[PART 1: Data Section 2](#_Toc437126836)

[1. Summary 2](#_Toc437126837)

[2. Source description and statistics 2](#_Toc437126838)

[3. Metadata 2](#_Toc437126839)

[4. Data processing 4](#_Toc437126840)

[5. SAS code 4](#_Toc437126842)

[6. Excel functions used 4](#_Toc437126843)

[PART 2: Analytics Section 5](#_Toc437126844)

[1. Plan of analysis 5](#_Toc437126845)

[2. Cluster Analysis 5](#_Toc437126846)

[2.1 Cluster Summary 8](#_Toc437126847)

[3. Profiling 8](#_Toc437126848)

[4. SAS code 9](#_Toc437126850)

[5. Excel functions used 11](#_Toc437126851)

[6. Other findings and analysis 12](#_Toc437126852)

[7. Summary 14](#_Toc437126853)

[APPENDIX 15](#_Toc437126854)

# PART 1: Data Section

## Summary

The Public Use Microdata Sample (PUMS) contains a sample of actual responses to the American Community Survey (ACS). PUMS files for an individual year, such as 2013, contain records of data from approximately one percent of the United States population. The PUMS dataset includes variables for nearly every question on the survey. Each record in the file represents single housing unit. The 2013 ACS PUMS files rely on PUMA boundaries that were drawn by state governments using data from the 2010 Census.

This PUMS file is downloaded from below link -

[**http://www.census.gov/programs-surveys/acs/data/pums.html**](http://www.census.gov/programs-surveys/acs/data/pums.html)

psam\_h17.sas7bdat, SAS format file is downloaded and copied in SAS ‘Final\_Project’ folder to make it available for use. In SAS, ‘census’ library is created to access this data from '/folders/myfolders/Final\_Project' path. This dataset is further used to generate clusters to form groups of data having similar characteristics.

## Source description and statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Item** | **Source**  **(e.g., website)** | **Description** | **Statistics**  **(e.g., # of rows)** | **Comments** |
| psam\_h17.sas7bdat | [Illinois Housing Unit Records](http://www2.census.gov/acs2013_1yr/pums/unix_hil.zip) | It contains the information of housing units of Illinois state for year 2013. | No. of rows : 58206  No. of columns : 231  Total tax groups: 68 | Housing data is considered for only Illinois state (year 2013). |

## Metadata

Below variables are selected for analysis from source dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Item** | **Description** | **Valid Range Values** | **Purpose for selection** |
| HINCP | Household income (past 12 months) | All positive values. Negative or NULL values are ignored. | HINCP is useful to get the idea of current household income as it plays major role in deciding tax increase depending on it’s high to low brackets. |
| VALP | Property value | All positive values. Missing values are ignored. | VALP is useful to understand the asset value on which householders are paying taxes. It helped in clustering to group same ranged households together. |
| TAXP | Property taxes (yearly amount) | TAX range starts from 0 i.e no tax to $10K +. | TAXP is useful to group households into high to low classes based on their tax amount which further can be used in deciding tax increase schemes according to tax value ranges. |
| MV | When moved into this house or apartment | Starting from less than 12 months to greater than 30 years. | It helps to understand since when owner is paying property tax so it can be deciding factor while doing tax increase. |
| WIF | Workers in family during the past 12 months | No workers  1 worker  2 workers  3 or more workers in family | WIF is useful to find out the total workers in family which can directly impact tax increase so single worker may face minimum tax increase. |
| WORKSTAT | Work status of householder or spouse in family households | 15 different combinations of working status of family. E.g Neither husband nor wife in labor force | WORKSTAT gives the status of workers in family which can also contribute in deciding tax increment. E.g if household income is below average and single worker is present in family then tax increase applied will be less or minimum. |
| SMX | Second or junior mortgage or home equity loan status | 1 .Yes, second mortgage  2 .Yes, home equity loan  3 .No  4 .Both second mortgage and home equity loan | SMX is useful to understand the what kind of second mortgages householders have recently which depicts the stability of households to decide tax increase. |

More details about all variables are available on below Link:

<http://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMSDataDict13.pdf>

## Data processing

## Outliers, missing values and negative values are ignored as there is no impact on analysis. No data trimming is performed as data was good enough to perform analysis.

|  |  |  |
| --- | --- | --- |
| **Variable Item** | **Description** | **Comments** |
| VALP | Very few records have extremely high property value i.e above $2M. | These are considered as outliers and grouped into clusters of it's own. |
| HINCP | Very few records have extremely high household income value i.e above $1M. | These are considered as outliers and grouped into clusters of it's own. |
| HINCP | Negative and missing values for household income | If cluster has 70% to 80% negative or missing values then that cluster is ignored while doing analysis. |
| VALP, TAXP, SMX and MV | Missing values | Missing values are ignored as there is no impact on analysis. |

## SAS code

|  |  |  |
| --- | --- | --- |
| **SAS Program Name** | **Copy/Paste Code** | **Comments** |
| Setup.sas | %let path=/folders/myfolders/Final\_Project;  libname census "&path"; | It is used to initialize a library to access data from the path. Every time after starting up SAS, it has to be executed before accessing the data in SAS. |
| content.sas | proc contents data=census.psam\_h17;  run; | This program is used to analyze details about dataset like statistics, data types etc. |

## Excel functions used

N/A

# PART 2: Analytics Section

## Plan of analysis

Objective of this analysis is to find out the property tax increase strategy for Illinois state using housing data set for year 2013. Below is the approach which is followed to decide tax strategy.

* + **Cluster Generation**

Clusters are generated using FASTCLUS method of SAS based on housing data set and, considering property value, household income variables. Scatter plot of clusters is used to depict graphical view of clusters. These clusters are created to group the data sets having common factors or similarities so it's easy to compare clusters with each other to decide tax strategy for each cluster.

* + **Characterization of variables of each cluster**

Each cluster is then characterized using above selected variables from source data set to understand the behavior of cluster. Characterization involved analysis of histogram of each variable's distribution in that cluster, frequency, minimum and maximum values etc. This further helped in ranking clusters using high to low range based on selected variables.

* + **Cluster Analysis and Profiling**

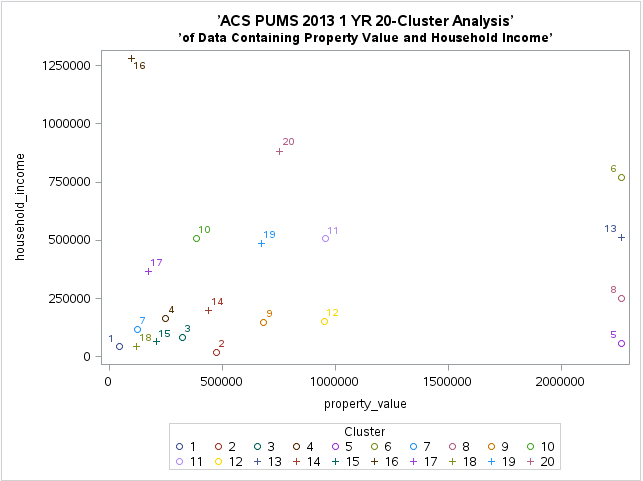
Cluster profiling is done by analyzing characteristics of HINCP, VALP, SMX, MV, TAXP and WORKSTAT variables and clusters are then ranked and grouped on the basis of their similar characteristics for deciding tax increase schemes. So it’s easier to apply tax increase strategy to whole group instead of working on all clusters which hold similar properties. High to low ranking for all clusters is given by deciding the ranges for TAXP, HINCP and VALP variables.

* + **Tax Strategy for profiled groups**

Various schemes for tax increase are decided and applied to all groups considering group behavior and ranking allotted to TAXP, HINCP and VALP from high to low ranges.

## Cluster Analysis

* + 20 Clusters are generated to analyze dataset. Previously, less than 20 clusters selected but it didn't form separate cluster for outliers and so it required additional sub-grouping. Using 20 clusters, outliers formed their own groups and so it helped in profiling to treat them separately.
  + E.g 16, 5, 8, 3 and 6 clusters are outliers.



*1.1 Graphical representation of cluster*

* + In SAS, cluster output of FASTCLUS procedure is stored in subset1\_20clusters table and further 20 tables are created to store data for each cluster separately. These clusters are characterized using below steps in SAS-

1. Go to 'Tasks' in navigator and click on 'Characterize data'
2. Select table from 'Data' tab and then choose variables from dropdown for which characterization is required.
3. Click on Run option; it displays histogram of each variable and frequency, percentage, minimum, maximum values etc.
   * Each cluster is then analyzed and profiled using above characteristics of variables. Numeric variables like TAXP, VALP and HINCP are ranked according to below ranges and these ranges are decided looking at minimum, average and maximum values from characterization.

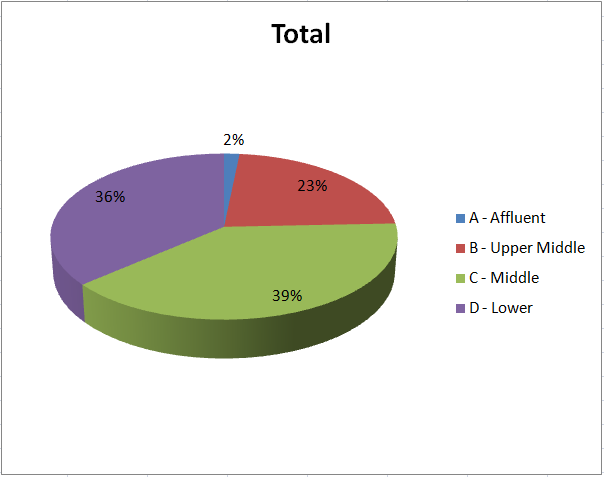
|  |  |  |
| --- | --- | --- |
| **\*\*Income** |  |  |
| **Amount in $** | **Desc** | **Rank** |
| 150k > | high | 1 |
| 91K to 150K | mid-high | 2 |
| 35K to 90K | mid | 3 |
| 20K to 34K | mid low | 4 |
| 20K < | low | 5 |
| **\*\*Property** | | |
| **Amount in $** | **Desc** | **Rank** |
| >2M | high | 1 |
| 700K to 2M | mid-high | 2 |
| 500K to 700K | mid | 3 |
| 100K to 500K | mid-low | 4 |
| <100K | low | 5 |
| **\*\*Tax Band** | | |
| **Band** | **Amount in $** | **Rank** |
| 68 | 10K + | A |
| 62 to 67 | 5K to 9.9K | B |
| 37 to 61 | 2.5K to 4.9K | C |
| 37< | 2.5K < | D |

* + Clusters are then grouped on the basis of their similar characteristics and divided it into 4 groups as follows –
  + **Group A –Affluent Class (Cluster : 20,8,6,11,12,13)**

1. Mid-High or high property value
2. High income
3. 2 workers in family
4. Highest tax payer i.e 10K $+
5. Both husband and wife are employed
   * **Group B –Upper-middle Class (Cluster : 4,7,9,10,14,17,19)**
6. Mid or mid-low property value
7. Mid-high or High income
8. 2 workers in family
9. Pays tax more than 5K $
10. Owning house for more than 10 years
11. Both husband and wife are employed
12. Home equity loans as second mortgage
    * **Group C –Middle Class (Cluster : 3,15,18)**
13. Mid-low property value
14. Average or medium income
15. 2 workers in family and also includes cases where only husband is working
16. Pays tax in between $2.5K to $4.9K
    * **Group D –Lower Class (Cluster : 1,2)**
17. Low or mid-low property value
18. Medium or below average income
19. No tax or less than 3K $
20. Non-working, female household

### 2.1 Cluster Summary

* 7989 records are omitted from analysis due to missing values.
* 50217 records are divided into 20 clusters.
* Out of 50217 , 1092 records i.e clusters 5 and 16 are ignored due to high missing values of VAPL and HINCP variables.
* Maximum households are from middle class(C group) who pays taxes in between $2.5K to $9.5K. So tax strategy will be decided considering this fact.



*1.2 Pie chart represents percentage of population for each class*

## Profiling

## Below tax increase schemes are applied to all 4 groups based on above cluster analysis and ranking given to clusters. Households can fit into multiple schemes but it will start hierarchically from scheme 1 i.e highest tax bracket and so whichever is the highest tax bracket scheme fits to household among all is applicable to that household.

* + **Scheme 1**

30% tax increment for households who are currently paying taxes more than $10K, having property value above $2M and yearly income is more than $150K.

* + **Scheme 2**

25% tax increment for households who are currently paying taxes more than $5K, having property value above $700K and yearly income is more than $150K.

* + **Scheme 3**

20% tax increment for households who are currently paying taxes more than $5K, having property value less than $700K and yearly income is more than $150K.

* + **Scheme 4**

15% tax increment for households who are currently paying taxes in between $2.5K to 5K, having yearly income more than $90K and workers in family are more than 1 except which falls in non-working categories (Workstat : 15,9,12).

* + **Scheme 5**

10% tax increment for households who are currently paying taxes in between $1K to 2.499K, having yearly income more than $70K and workers in family are more than 1 except which falls in non-working categories (Workstat : 15,9,12).

* + ***Rest all are exempted from tax increase.***
  + **Steps followed for tax calculation -**
  + In data dictionary, tax minimum and maximum ranges are given for all 68 tax groups so first of all averages of all groups are calculated in excel using formula '=AVERAGE(B4,C4)' by copying min values in column B and max values in C. This sheet is imported in SAS using IMPORT utility and created census.tax\_lkp table which is further used to calculate tax increase.
  + For e.g if household's tax group is 62 which has min value $5K and max value $5499 then it is paying average tax $5249.50 and if it's falling under 20% tax increase scheme then incremented tax for that household is 5249.50\*0.20 = $1049.90.
  + By following above analysis for all groups, total tax increment around $8.5M can be anticipated.

## SAS code

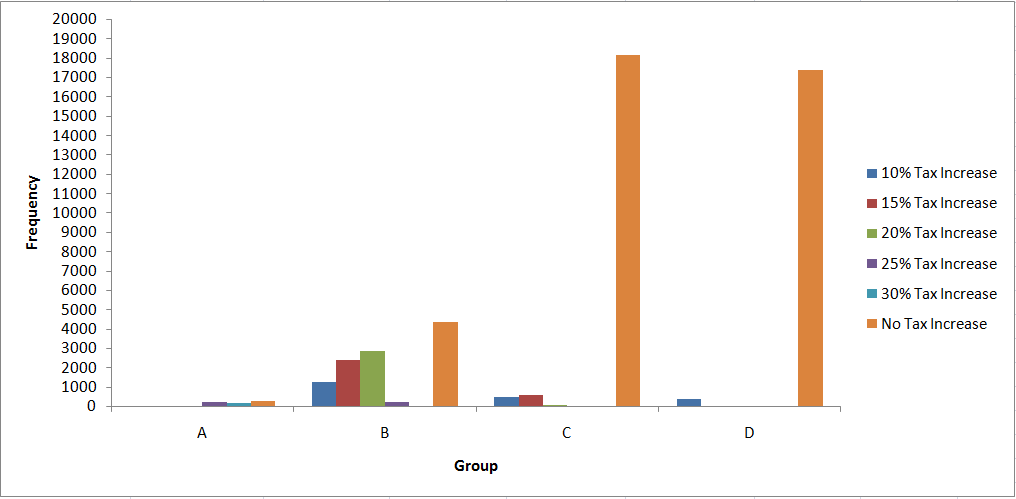
|  |  |  |
| --- | --- | --- |
| **SAS Program Name** | **Copy/Paste Code** | **Comments** |
| Cluster\_20.sas | /\*20 Cluster run of FASTCLUS\*/  Proc fastclus data=census.psam\_h17\_subset1  radius=0 replace=full  converge=0 maxiter=200  maxclusters=20  OUTSTAT=census.psam\_h17\_subset1\_20clusters\_stat  OUT=census.psam\_h17\_subset1\_20clusters  distance;  id SERIALNO;  var VALP HINCP;  run;  proc sgplot;  scatter y=HINCP x=VALP / group=cluster;  title ’ACS PUMS 2013 1 YR 20-Cluster Analysis’;  title2 ’of Data Containing Property Value and Household Income’;  run; | This program is used to generate 20 clusters from source dataset. |
| Agg\_variables.sas | /\*Join additional variables from psam\_h17 table for profiling\*/  procsql;  title '20 Cluster aggregated information for profiling';  create table census.psam\_h17\_subset1\_20clusters\_agg as  select a.SERIALNO, a.CLUSTER, a.DISTANCE, a.HINCP, a.VALP, a.TAXP,  b.MV, b.SMX, b.WORKSTAT, b.WIF  from census.psam\_h17\_subset1\_20clusters as a left join census.psam\_h17 as b  on a.SERIALNO=b.SERIALNO;  run; | This program is used to merge other variables with cluster dataset |
| Clust\_dataset.sas | %macro sqlloop;  procsql;  /\*Loop through 20 times to create a table for each cluster\*/  %DO k=1 %TO 20;  create table census.psam\_h17\_subset1\_cluster\_agg&k. as  select \*  from census.psam\_h17\_subset1\_20clusters\_agg  where CLUSTER=&k.;  %END;  QUIT;  %mend;  %sqlloop; | This program is used to insert each cluster data into separate tables. |
| Clust\_group.sas | proc sql;  create table census.psam\_h17\_subset1\_20clusters\_grp as  select 'A' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg20  union all  select 'A' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg8  union all  select 'A' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg6  union all  select 'A' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg11  union all  select 'A' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg12  union all  select 'A' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg13  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg4  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg7  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg9  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg10  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg14  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg17  union all  select 'B' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg19  union all  select 'C' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg3  union all  select 'C' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg15  union all  select 'C' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg18  union all  select 'D' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg1  union all  select 'D' as GROUP,\* from census.psam\_h17\_subset1\_cluster\_agg2;  run; | This program is used to re-aggregate all clusters group wise. Groups are formed after cluster analysis. |
| Tax\_calc.sas | proc sql;  create table census.psam\_h17\_subset1\_cluster\_tax as  select a.\*, b.AVG,case when (b.No\_=68 and a.VALP>2000000 and a.HINCP>150000 and VALP is not null) then '30% Tax Increase'  when (b.No\_>=62 and a.VALP>=700000 and a.HINCP>150000 and VALP is not null) then '25% Tax Increase'  when (b.No\_>=62 and a.VALP<=700000 and a.HINCP>150000 and VALP is not null and VALP > 0) then '20% Tax Increase'  when (b.No\_>=37 and b.No\_<62 and a.HINCP>90000 and a.WIF<>'1' and a.WORKSTAT not in ('15','9','12')  and VALP is not null and VALP > 0) then '15% Tax Increase'  when (b.No\_>=22 and b.No\_<37 and a.HINCP>70000 and a.WIF<>'1' and a.WORKSTAT not in ('15','9','12')  and VALP is not null and VALP > 0) then '10% Tax Increase'  else 'No Tax Increase' end as SCHEME,  case when (b.No\_=68 and a.VALP>2000000 and a.HINCP>150000 and VALP is not null) then (b.AVG\*0.3)  when (b.No\_>=62 and a.VALP>=700000 and a.HINCP>150000 and VALP is not null) then (b.AVG\*0.25)  when (b.No\_>=62 and a.VALP<=700000 and a.HINCP>150000 and VALP is not null and VALP > 0) then (b.AVG\*0.20)  when (b.No\_>=37 and b.No\_<62 and a.HINCP>90000 and a.WIF<>'1' and a.WORKSTAT not in ('15','9','12')  and VALP is not null and VALP > 0) then (b.AVG\*0.15)  when (b.No\_>=22 and b.No\_<37 and a.HINCP>70000 and a.WIF<>'1' and a.WORKSTAT not in ('15','9','12')  and VALP is not null and VALP > 0) then (b.AVG\*0.10)  else 0 end as TAXP\_INCREASE  from census.psam\_h17\_subset1\_20clusters\_grp as a left join  census.tax\_lkp as b  on input(a.taxp,18.) =b.No\_;  run; | This program is used to calculate tax for each cluster applying new schemes. |
| Summ\_tax.sas | proc sql;  select group,cluster,scheme,count(\*) as frequency,sum(TAXP\_INCREASE) as total from census.psam\_h17\_subset1\_cluster\_tax  group by group,cluster,scheme;  run; | It's used to sum up tax amount by group, cluster and scheme wise. |

## Excel functions used

|  |  |  |
| --- | --- | --- |
| **Excel Function** | **Description or Copy/Paste Code** | **Comments** |
| Pivot Table | First pivot table is used to display the count of households per group falling under various tax increase schemes. Second table is used to display tax increase amount per group and per scheme in tabular format. | Manually exported the data of census.psam\_h17\_subset1\_cluster\_tax final table from SAS into excel and then created pivot table using below steps:  Go to Data ribbon and click on pivot table option to generate table for selected data.  Group the data on the basis of schemes and group/class and sum up frequency, amount. |
| Pivot Chart | First pivot chart is graphical representation of pivot table 1 to understand maximum and minimum frequency. Second pivot chart is graphical representation of pivot table 2 to understand maximum revenue getting generated from tax increase. | Steps:  Go to Data ribbon and click on pivot chart option to generate chart for selected pivot table. |

## Other findings and analysis

Below chart shows that maximum tax increase benefit can be obtained from group B i.e upper middle; 6779 households out of 11164 from group B are falling under new tax increase schemes where as group D i.e lower class has lowest tax increase; 17440 households out of 17841 are exempted from tax increase. Group A i.e affluent class mostly falls in 30% or 25% tax increment scheme.

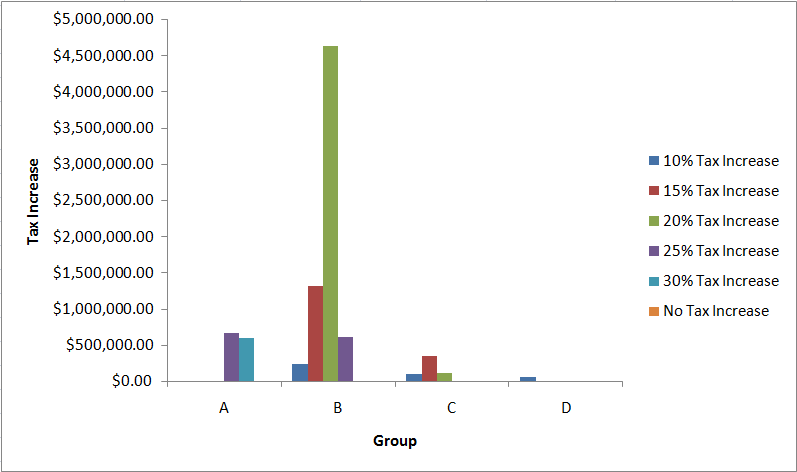


*1.3 Group versus Frequency distribution*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sum of FREQUENCY** | **Scheme** |  |  |  |  |  |  |
| **Group** | **10% Tax Increase** | **15% Tax Increase** | **20% Tax Increase** | **25% Tax Increase** | **30% Tax Increase** | **No Tax Increase** | **Grand Total** |
| A | 7 | 8 | 9 | 274 | 202 | 294 | 794 |
| B | 1265 | 2401 | 2857 | 256 |  | 4385 | 11164 |
| C | 499 | 584 | 74 |  |  | 18169 | 19326 |
| D | 384 | 17 |  |  |  | 17440 | 17841 |
| **Grand Total** | **2155** | **3010** | **2940** | **530** | **202** | **40288** | **49125** |

*1.4 Frequency per group per scheme*

Below chart shows that Group B i.e upper middle class has majority of households falling in10% to 25% tax increment schemes and around $6M can be obtained from it whereas very less i.e around $67K can be obtained from Group D. Only group A i.e affluent class falls in 30% tax increment scheme.



*1.3 Group versus Amount distribution*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sum of TAX-INCREASE** | **Scheme** |  |  |  |  |  |  |
| **Group** | **10% Tax Increase** | **15% Tax Increase** | **20% Tax Increase** | **25% Tax Increase** | **30% Tax Increase** | **No Tax Increase** | **Grand Total** |
| A | $1,324.65 | $3,794.40 | $17,199.80 | $673,996.75 | $606,000.00 | $0.00 | $1,302,315.60 |
| B | $234,381.75 | $1,321,462.43 | $4,640,552.10 | $624,933.50 |  | $0.00 | $6,821,329.78 |
| C | $92,480.05 | $346,021.20 | $122,094.10 |  |  | $0.00 | $560,595.35 |
| D | $59,550.80 | $8,286.23 |  |  |  | $0.00 | $67,837.03 |
| **Grand Total** | **$387,737.25** | **$1,679,564.25** | **$4,779,846.00** | **$1,298,930.25** | **$606,000.00** | **$0.00** | **$8,752,077.75** |

*1.4 Tax increase amount per group per scheme*

## Summary

* Total anticipated tax increase is $8,752,077.75.
* Highest tax payers and very high income householders are in group A so they fall in maximum i.e 30% tax increment scheme . 476 households out of 794 from group A i.e affluent class falls in 25% and 30% tax increment schemes and total benefit of group A can go up to $1,302,315.60.
* Low tax payers having low household incomes are exempted from tax increase and these are majorly from middle and lower classes. 40288 households out of 49125 are exempted from tax increase.
* For Group C, maximum returns are anticipated from 15% tax increase scheme; total anticipated returns from Group C is $560,595.35.
* 98% of households from Group D are exempted from tax increase.
* Most stable group is B as maximum anticipated returns for upper middle class is $6,821,329.78.
* Around 82% of households cannot go under tax increase scheme so it is the next focus area to look after on their existing tax strategy and how it's working out currently.

## APPENDIX

Detailed description for each cluster after characterization of data.

* + **Cluster 1 :**

1. High frequency band – 10460 i.e lot of people fall in this criteria.
2. Average income and low property value.
3. Pays no tax or very low i.e around 1200$.
4. 17% are non-workers.
   * **Cluster 2 :**
5. High frequency band – 7381 i.e lot of people fall in this criteria.
6. Poor income (~18K$) and mid-low property value.
7. Female householders.
8. Mostly one worker in family.
   * **Cluster 3 :**
9. Medium frequency band – 4138
10. Above average income and low property value.
11. Pays tax more than 6000$.
12. In 17% cases only husband is working.
13. Moved in house approx. 10 years back.
    * **Cluster 4:**
14. Good income and mid-low property value.
15. Pays tax more than 6000$.
16. Husband and wife both employed
    * **Cluster 5 :**
17. Average income and high property value i.e over 2 Millions.
18. 90% property values are missing so cluster is ignored.
    * **Cluster 6 :**
19. Low frequency band – 45 i.e only few people fall in this criteria.
20. Very high income i.e around 770K$ and high property value i.e over 2M.
21. Falls in highest tax band i.e 10K $+
22. 2 workers in family
    * **Cluster 7 :**
23. High income but mid-low property value.
24. Pays tax more than 3K $
25. Both husband and wife employed
    * **Cluster 8 :**
26. Low frequency band – 175 i.e few people fall in this criteria.
27. High income and very high property value.
28. Falls in highest tax band i.e 10K $+
29. 2 workers in family
    * **Cluster 9 :**
30. High income and average property value.
31. Falls in highest tax band i.e 10K $+
32. 2 workers in family
33. 30% people have home equity loans
    * **Cluster 10 :**
34. Low frequency band – 284 i.e few people fall in this criteria.
35. Very high income and low property value.
36. Pays tax more than 6000$.
37. 17% are non-workers.
    * **Cluster 11 :**
38. Low frequency band – 101 i.e very few people fall in this criteria.
39. Very high income i.e around 500K $ and mid-high property value.
40. Falls in highest tax band i.e 10K $+
41. Both husband and wife employed.
    * **Cluster 12 :**
42. Low frequency band – 294 i.e few people fall in this criteria.
43. Good income and mid-high property value.
44. Falls in highest tax band i.e 10K $+
45. Mostly both husband and wife are employed.
    * **Cluster 13 :**
46. Low frequency band – 116 i.e few people fall in this criteria.
47. Very high income (>500K $) and very high property value i.e over 2M $.
48. Falls in highest tax band i.e 10K $+
    * **Cluster 14 :**
49. Good income and low property value.
50. Falls in highest tax band i.e 10K $+
51. 22% of people have home equity loans
52. Both husband and wife are employed.
    * **Cluster 15 :**
53. Average income and low property value.
54. Pays tax more than 5000$.
55. Moved in house approx 10 years back.
    * **Cluster 16 :**
56. Low frequency band – 99 i.e very few people fall in this criteria.
57. Lot of income values are missing so cluster is ignored
    * **Cluster 17 :**
58. High income and low property value.
59. Pays tax more than 5000$.
60. Moved in house approx 10 years back.
    * **Cluster 18 :**
61. High frequency band – 8454 i.e lot of people fall in this criteria.
62. Average income and low property value.
63. Pays tax more than3K $.
64. In 21% cases, both husband and wife are not working
    * **Cluster 19 :**
65. High income and average property value.
66. Falls in highest tax band i.e 10K $+
67. Mostly both husband and wife are employed.
    * **Cluster 20 :**
68. Very low frequency band i.e only 33 people fall in this criteria.
69. Very high income i.e around 800K $ and mid-high property value.
70. Falls in highest tax band i.e 10K $+
71. Mostly both husband and wife are employed.