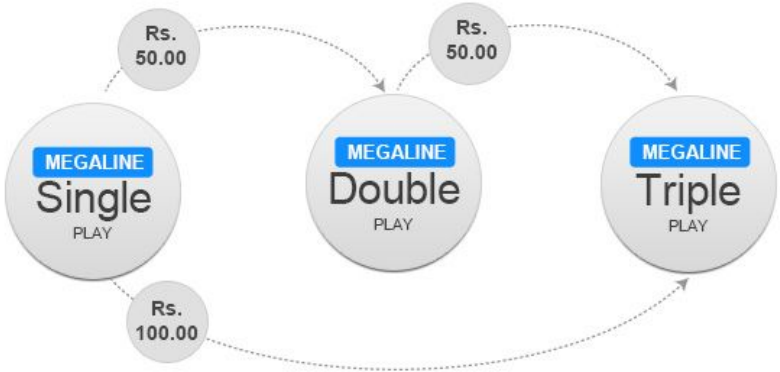
Telecommunication Data Analysis Workflow

Chiran Hewawitharana IM/2016/046

Supervised By: Dr. Chathura Rajapaksha, Mr. Dinesh Asanka

Usage Analysis (VOICE)



VOICE calls usage analysis

- Categories: ONNET, OFFNET, IDD (INCOMING & OUTGOING)
- Separate analysis for each dataset.
- User Identifier : ***event_source_hash***
- Steps:
    - Handle null values
    - Drop unnecessary columns
    - Pivot by **month-year** usage
    - Generating Usage rating by **Call-count & Duration**
    - Bin by User Locations
    - Outlier Analysis
    - Extract Usage and ratings (Categorical & Scaled)
    - Provide visualizations

VOICE calls usage: Pivot table

	call_count										...	duration						
year-month	201908	201909	201910	201911	201912	202001	202002	202003	202004	202005	...	201911	201912	202001	202002	202003	202004	202005
event_source.hash																		
00221451f705ebe26051158bb14f567a	6	1	2	1	2	4	6	2	0	1	...	14	74	80	368	70	0	60
002fb47f60400713c854b0f69ff78c0a	1	1	3	9	2	6	5	5	8	7	...	2042	500	1744	1560	3140	1993	1920
00567a4a3c474aca1ce5cd6570648932	0	0	2	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
006be321aad541ff1d6c32a43dd7cdde	10	5	16	7	23	11	6	14	6	6	...	5760	10558	3718	4008	8465	4720	2834
00724ae2470df1ed0fe919800d02517d	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
ffb4fa6778bc5b3d9c858f5f4fa1e0c2	4	11	9	32	8	14	6	2	0	1	...	17526	6477	14409	6407	2969	0	60
ffb63ff99cf1354dff1da862f379da80	46	70	48	71	44	54	24	21	8	14	...	6314	3259	3476	1451	1360	945	1045
ffd697d66cc01c8897ce3e37d693673a	3	7	3	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
ffd96f173d6cd683bf3384bdc2f99713	25	18	27	22	27	22	16	64	53	31	...	3969	6588	4015	1833	10585	9217	4991
ffee845c0caa320b55e31bb156827d16	9	6	2	4	1	17	18	37	54	7	...	1891	60	1200	1080	4287	8349	475

VOICE calls usage: Descriptive analysis

- Call Count
  - 0-26 within 75% (Q3) of the distribution
  - 26-6890 within remaining 25% of the distribution

- Call Duration
  - 0-4500 seconds within 75% (Q3) of the distribution
  - 4500-70042 seconds within remaining 25% of the distribution

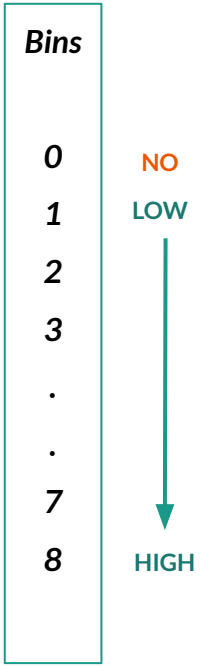
	call_count	duration
count	32427.00000	32427.00000
mean	31.43439	4689.88938
std	144.07306	16137.82189
min	1.00000	1.00000
25%	4.00000	388.00000
50%	10.00000	1542.00000
75%	26.00000	4500.00000
max	6890.00000	700042.00000

VOICE ONNET INCOMING DATA

VOICE calls usage: Usage Ranking

- Bined Call Counts - 8 Bins
  - Bin ranges selected considering skewness of data distribution
  - **Freedman-Diaconis's Rule based approach (Using Interquartile range)**
  - Ex: For ONNET INCOMING Data : [0, 6,15, 30, 100,1200,2400,4800,6890]

- Bined Call Duration - 8 Bins
  - Bin ranges selected considering skewness of data distribution
  - **Freedman-Diaconis's Rule based approach (Using Interquartile range)**
  - Considered Business meaning of Durations (As given in seconds, rounded to the nearest 60s)
  - Ex: For ONNET INCOMING Data : [0, 300, 1500, 4200,16800,33600,66000,240000,700042]



Reference:
https://en.wikipedia.org/wiki/Freedman%E2%80%93Diaconis_rule#:~:text=For%20a%20set%20of%20empirical,of%20the%20theoretical%20probability%20distribution.

https://stats.stackexchange.com/questions/143438/optimal-number-of-bins-in-histogram-by-the-freedman-diaconis-rule-difference-be

VOICE calls usage: Usage Ranking

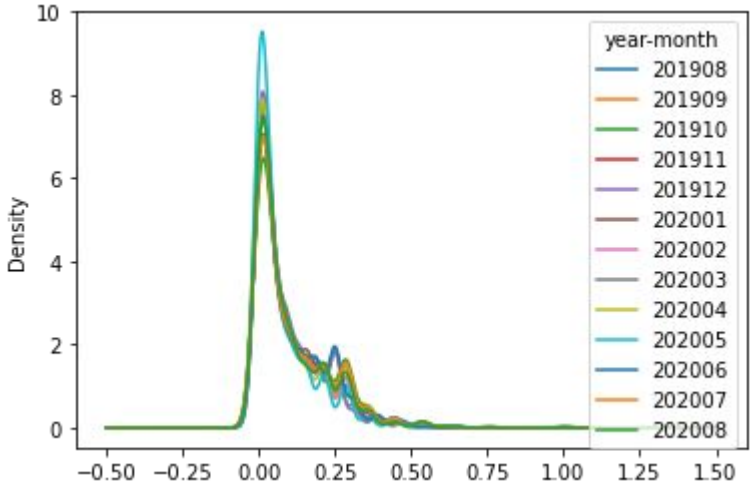
- Overall Usage ranking
  - **Binned Call_Count x Binned Duration**
  - *Value between 0-64*
- *ToDo: More advanced technique to combine variables?*
- Scaled values using Min-Max scalar **(0-1)**

Reference:

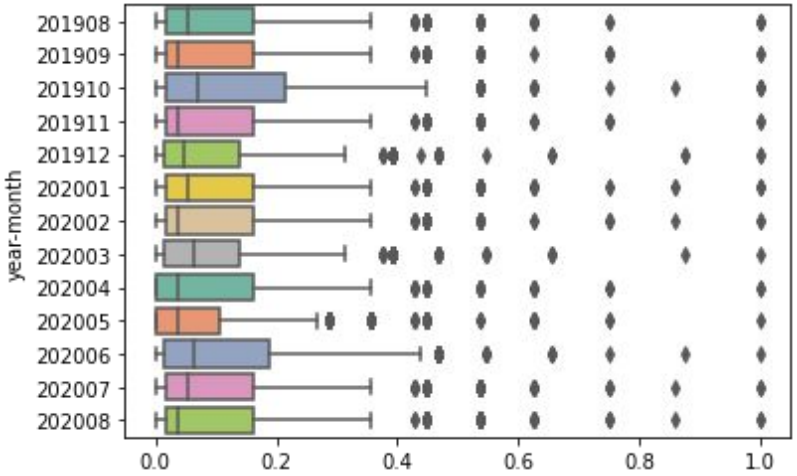
https://doi.org/10.1109/ICCE48956.2021.9352111, 2021 (Pham, C. D., Anh Chu, T., Pham, H. H., Linh Dao, M., Pham, T. S., Hung Trinh, V., & Nguyen, D. H. A recommendation system for offers in telecommunications)

http://www.utstat.toronto.edu/~brunner/DataAnalysisText/Interactions.pdf

VOICE calls usage: Outlier analysis



VOICE ONNET INCOMING DATA : Density and Skewness



VOICE ONNET INCOMING DATA : Outliers

VOICE calls usage: Outlier analysis

- Quantitative statistical methods to detect outliers
  - Tukey's box plot method
  - Internally studentized residuals (AKA z-score method)
  - Median Absolute Deviation method

- Tukey's box plot method
  - Tukey distinguishes between possible and probable outliers. A possible outlier is located between the inner and the outer fence, whereas a probable outlier is located outside the outer fence.
  - **IQR =Q3 - Q1**, (whereas q3 := 75th quartile and q1 := 25th quartile)
  - **Inner fence = [Q1 - 1.5 IQR , Q3 + 1.5 IQR]**
  - **Outer fence = [Q1 – 3 IQR , Q3 + 3 IQR]**

VOICE calls usage: Outlier analysis

- Outlier thresholds identified using Tukey's box plot method:

{201908: {'threshold': 0.42857142857142855, 'count': 82}},
{201909: {'threshold': 0.42857142857142855, 'count': 66}},
{201910: {'threshold': 0.5357142857142857, 'count': 44}},
{201911: {'threshold': 0.42857142857142855, 'count': 66}},
{201912: {'threshold': 0.375, 'count': 83}},
{202001: {'threshold': 0.42857142857142855, 'count': 80}},
{202002: {'threshold': 0.42857142857142855, 'count': 69}},
{202003: {'threshold': 0.375, 'count': 73}},
{202004: {'threshold': 0.42857142857142855, 'count': 48}},
{202005: {'threshold': 0.26785714285714285, 'count': 252}},
{202006: {'threshold': 0.46875, 'count': 44}},
{202007: {'threshold': 0.42857142857142855, 'count': 84}},
{202008: {'threshold': 0.42857142857142855, 'count': 68}}

VOICE calls usage: Outlier analysis

- Median Absolute Deviation method
  - This method is highly limited as the distributions mean and standard deviation are sensitive to outliers. This means that finding one outlier is dependent on other outliers as every observation directly affects the mean.

$$\text{MAD} = \text{median}(\left|X_i - \overline{X}\right|)$$

VOICE calls usage: Outlier analysis

- Outlier thresholds identified using Median Absolute Deviation method:

{201908: {'threshold': 0.3571428571428571, 'count': 166}},

{201909: {'threshold': 0.3571428571428571, 'count': 129}},

{201910: {'threshold': 0.3571428571428571, 'count': 201}},

{201911: {'threshold': 0.3571428571428571, 'count': 133}},

{201912: {'threshold': 0.3125, 'count': 152}},

{202001: {'threshold': 0.3571428571428571, 'count': 167}},

{202002: {'threshold': 0.3571428571428571, 'count': 142}},

{202003: {'threshold': 0.3125, 'count': 162}},

{202004: {'threshold': 0.3571428571428571, 'count': 142}},

{202005: {'threshold': 0.3571428571428571, 'count': 83}},

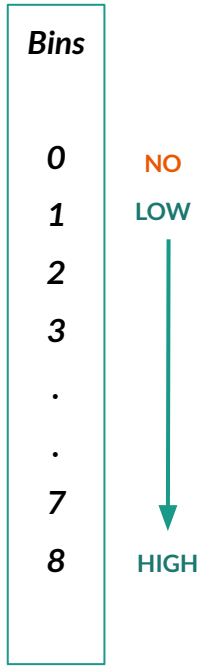
{202006: {'threshold': 0.3125, 'count': 221}},

{202007: {'threshold': 0.3571428571428571, 'count': 177}},

{202008: {'threshold': 0.3571428571428571, 'count': 147}}

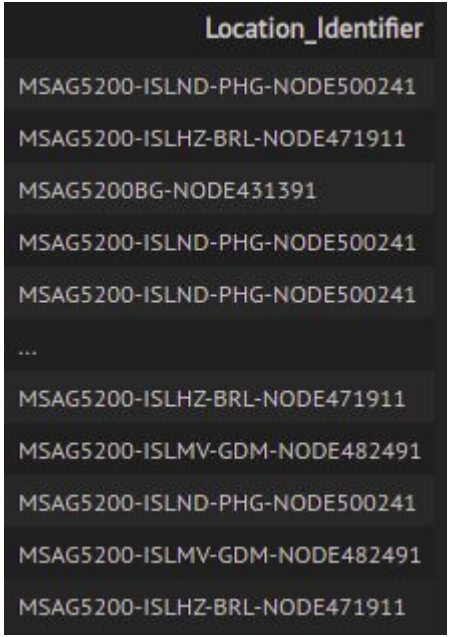
VOICE calls usage: Usage Ranking Categories

- Usage Rankings Categorized into four categories
  - Categories: LOW/ MEDIUM/ HIGH/ NO
  - **NO - No usage at all**
  - **Bin sizes defined using the same approach: considering data distribution (IQR)**
  - Ex: Bins For ONNET INCOMING Data : [0, 0.08, 0.3,1]



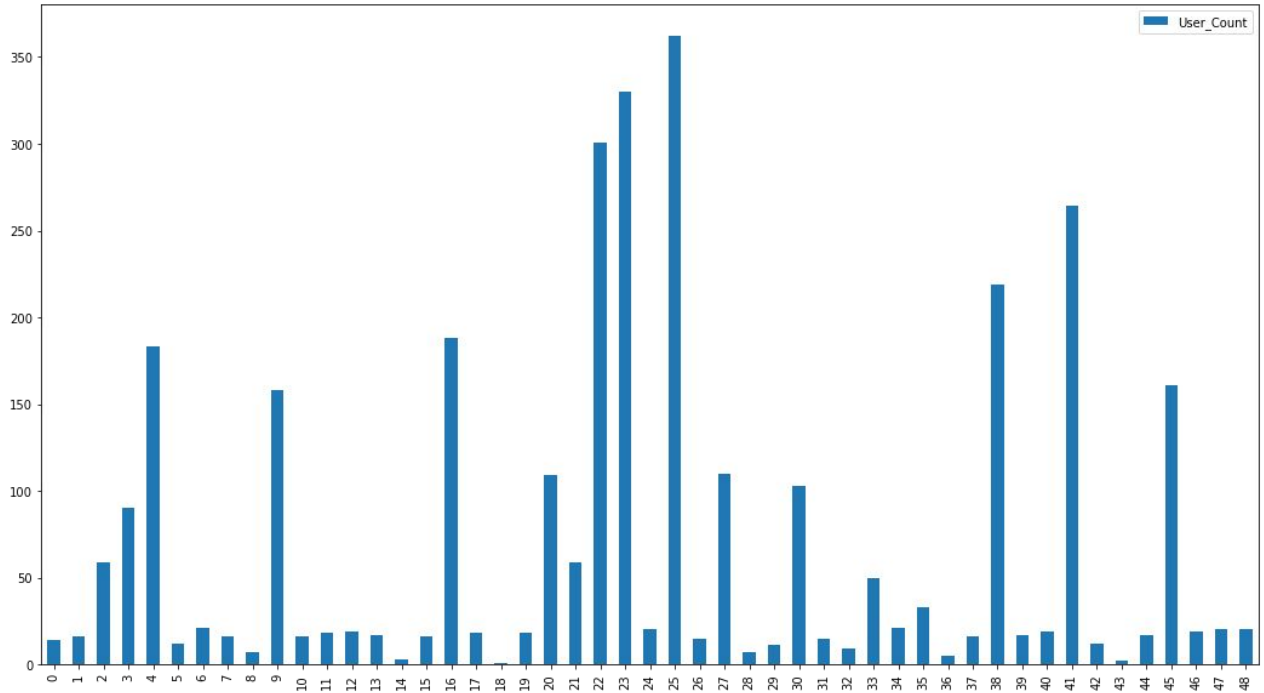
VOICE subscribers: Group by Location

- Combined four columns together. To identify unique location of a subscriber
    - **MSAN**
    - **Location_Code**
    - **EQUP_ID**
    - **EQUP_Index**
- Added new feature to User Profile : **Location_Identifier**



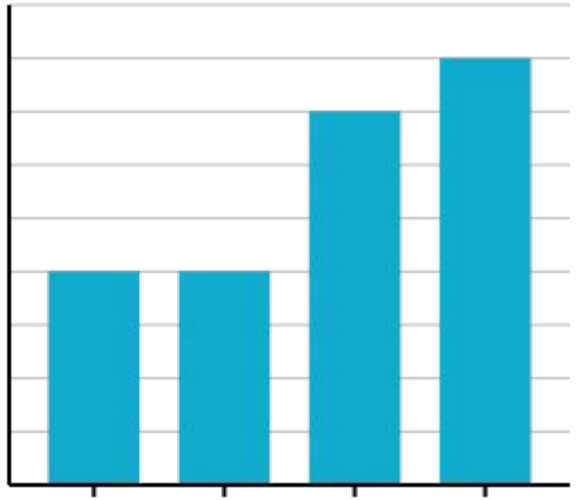
VOICE ONNET INCOMING DATA

VOICE subscribers: Group by Location



0	C300MHE-DMD-NODE15221561
1	C300MHT-NE-NODE8152461
2	C300MIDH-IDJ-NODE11065541
3	C300MKI-NODE10668164
4	C300MNW-MEE-NODE18887751
5	C300MPDT-NODE13379431
6	C300MTP-NODE8700911
7	C300MVH-NODE9874041
8	MA5603TAD-SRV-NODE10224991
9	MSAG5200-ISLBZ-NODE435881
10	MSAG5200-ISLGE-NODE5577981
11	MSAG5200-ISLHC-NAP-NODE478721
12	MSAG5200-ISLHK-OCC-NODE506561
13	MSAG5200-ISLHNT-NODE423851
14	MSAG5200-ISLHO-NODE524692
15	MSAG5200-ISLHZ-BRL-NODE471911
16	MSAG5200-ISLIDH-KGW-NODE6194491
17	MSAG5200-ISLJA-PKV-NODE460421
18	MSAG5200-ISLKI-KRB-NODE479411
19	MSAG5200-ISLMB-PSL-NODE482261
20	MSAG5200-ISLMH-POL-NODE491451
21	MSAG5200-ISLMLT-VMD-NODE4145091
22	MSAG5200-ISLMT-AVR-NODE467161
23	MSAG5200-ISLMV-GDM-NODE482491
24	MSAG5200-ISLMX-NODE5391821
25	MSAG5200-ISLND-PHG-NODE500241
26	MSAG5200-ISLNL-HGW-NODE491281
27	MSAG5200-ISLPC-AGT-NODE480031

Usage Analysis (BroadBand)



BroadBand usage analysis

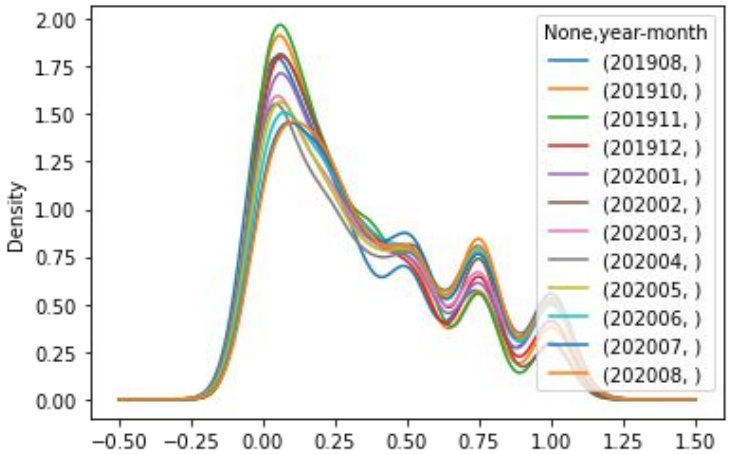
- User Identifier : ***PSTN_hash***
- Steps:
  - Handle null values
  - Drop unnecessary columns
  - Pivot by **month-year** usage
  - Generating Usage rating by Usage (**Upload + download) & Duration**
  - Bin by User Locations
  - Extract Usage and ratings (Categorical & Scaled)
  - Provide visualizations

BroadBand usage: Pivot table

	Duration										...	Usage				
year-month	201908	201910	201911	201912	202001	202002	202003	202004	202005	202006	...	201911	201912	202001	202002	202003
PSTN.hash																
00567a4a3c474aca1ce5cd6570648932	2277937	1256652	795222	1417310	2176277	2212790	2030474	1725246	2133617	1777555	...	11764247072	27618734644	33279185555	36946597062	2886039
00724ae2470df1ed0fe919800d02517d	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
00a8c3e3baed0557c31ab9b68a086ed2	699711	1153572	374914	511928	1194567	493510	431098	552825	421133	1049049	...	13435042644	10508792092	17492553338	12583099729	1504337
00cc5a820586ce291c0fa4c2a99a4816	0	0	0	0	157239	548360	0	0	1119319	966783	...	0	0	5498326	466553446	0
00d1f690ec4af27702febbd0db6942c2	713219	663171	558688	377493	570854	545675	697625	520233	350390	222968	...	7795652629	7632107196	8961346005	8920192587	9256717
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
feca6cdb74a4a95dec4fe88f3fb2f938	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
ff1093d5ef61bf46f8c88e9b7614e311	681240	791191	872112	938131	666153	693108	1223891	1615199	1541898	813031	...	10917038375	15030006772	10187461833	15969021804	3610257
ff45fdfb4b6dac191aba17a281181d39	2292467	2332487	2276132	2523725	2548082	2422885	2645146	2558089	2545852	2542633	...	21145367276	24093109565	21815269999	62009852907	1531017
ffb2ec238eba276f65157693ed329c7f	0	0	0	140358	2056665	1820231	2413915	2272529	2597092	2555086	...	0	1329073213	57067166125	30448045515	3710563
ffd697d66cc01c8897ce3e37d693673a	1405744	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

BroadBand usage: Descriptive analysis

- Usage Columns
  - Usage (Upload + Download) - in bytes
  - Used Duration - in seconds



	Duration	Usage
count	13043.00000	13043.00000
mean	1521226.50564	22678527038.04807
std	870908.53424	28481239994.13892
min	120.00000	0.00000
25%	739867.00000	6834388399.50000
50%	1518326.00000	13811012030.00000
75%	2427392.00000	27456020350.00000
max	2709612.00000	537118249202.00000

BB USAGE DATA

BroadBand usage: Usage Ranking

- Bined Data Usage (Uploads+Downloads) - 4 Bins
  - Bin ranges selected considering skewness of data distribution
  - Bins: [0, 7057561283.25, 14188526279.5, 28226707481.75, 537118249202.0]

- Bined Usage Duration - 4 Bins
  - Bin ranges selected considering skewness of data distribution
  - Considered Business meaning of Durations (As given in seconds, rounded to the nearest 60s)
  - Bins : [0, 745381.25, 1522064.0, 2438034.0, 2709612.0]

BroadBand usage: Usage Ranking

- Overall Usage ranking
  - **Binned Usage x Binned Duration**
  - *Value between 0-16*
- *ToDo: More advanced technique to combine variables?*
- Scaled values using Min-Max scalar **(0-1)**

Reference:
https://doi.org/10.1109/ICCE48956.2021.9352111, 2021 (Pham, C. D., Anh Chu, T., Pham, H. H., Linh Dao, M., Pham, T. S., Hung Trinh, V., & Nguyen, D. H. A recommendation system for offers in telecommunications)

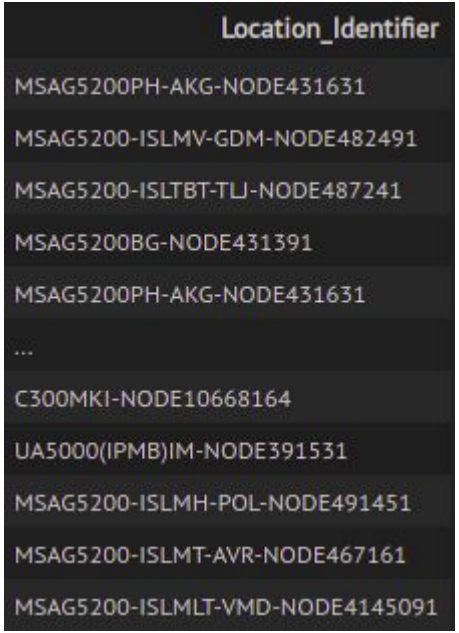
http://www.utstat.toronto.edu/~brunner/DataAnalysisText/Interactions.pdf

BroadBand usage: Usage Ranking Categories

- Usage Rankings Categorized into four categories
  - Categories: LOW/ MEDIUM/ HIGH/ NO
  - **NO - No usage at all**
  - **Bin sizes (equal size strata)**
  - Bins: [0, 0.33, 0.66, 1]

BroadBand subscribers: Group by Location

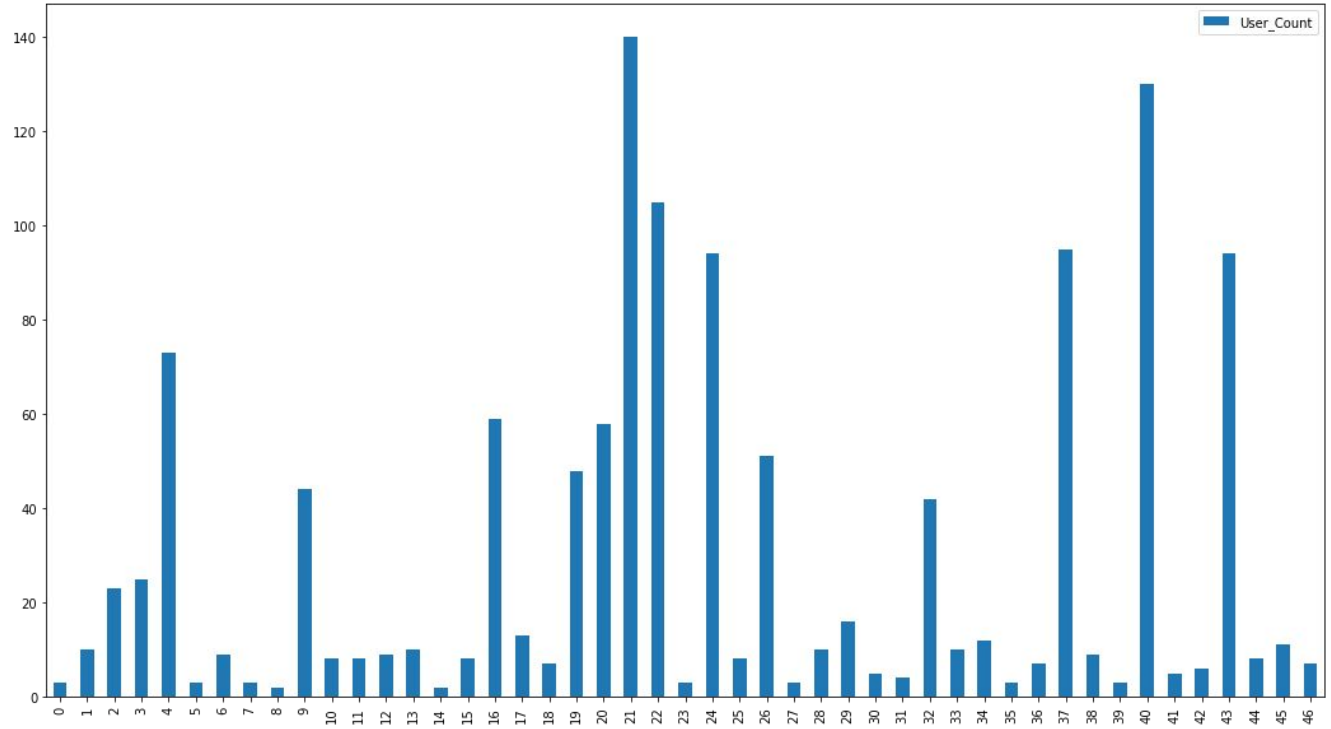
- Combined four columns together. To identify unique location of a subscriber
    - **MSAN**
    - **Location_Code**
    - **EQUP_ID**
    - **EQUP_Index**
- Added new feature to User Profile : **Location_Identifier**



Location_Identifier
MSAG5200PH-AKG-NODE431631
MSAG5200-ISLMV-GDM-NODE482491
MSAG5200-ISLTBT-TLJ-NODE487241
MSAG5200BG-NODE431391
MSAG5200PH-AKG-NODE431631
...
C300MKI-NODE10668164
UA5000(IPMB)IM-NODE391531
MSAG5200-ISLMH-POL-NODE491451
MSAG5200-ISLMT-AVR-NODE467161
MSAG5200-ISLMLT-VMD-NODE4145091

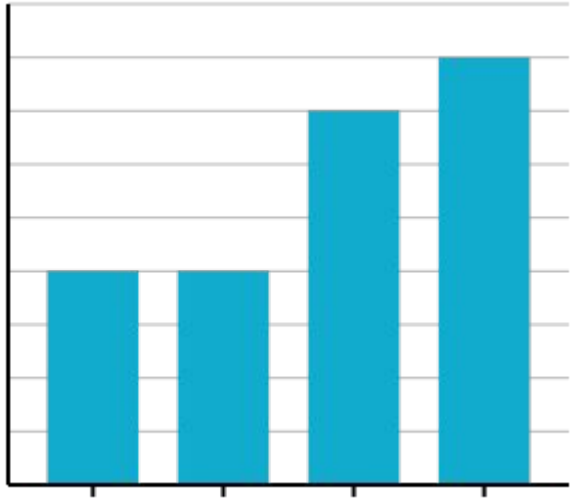
BB DATA

BroadBand subscribers: Group by Location



0	C300MHE-DMD-NODE15221561
1	C300MHT-NE-NODE8152461
2	C300MIDH-IDJ-NODE11065541
3	C300MKI-NODE10668164
4	C300MNW-MEE-NODE18887751
5	C300MPDT-NODE13379431
6	C300MTP-NODE8700911
7	C300MVH-NODE9874041
8	MA5603TAD-SRV-NODE10224991
9	MSAG5200-ISLBZ-NODE435881
10	MSAG5200-ISLGE-NODE5577981
11	MSAG5200-ISLHC-NAP-NODE478721
12	MSAG5200-ISLHK-OCC-NODE506561
13	MSAG5200-ISLHNT-NODE423851
14	MSAG5200-ISLHO-NODE524692
15	MSAG5200-ISLHZ-BRL-NODE471911
16	MSAG5200-ISLIDH-KGW-NODE6194491
17	MSAG5200-ISLJA-PKV-NODE460421
18	MSAG5200-ISLMB-PSL-NODE482261
19	MSAG5200-ISLMH-POL-NODE491451
20	MSAG5200-ISLMLT-VMD-NODE4145091
21	MSAG5200-ISLMT-AVR-NODE467161
22	MSAG5200-ISLMV-GDM-NODE482491
23	MSAG5200-ISLMX-NODE5391821
24	MSAG5200-ISLND-PHG-NODE500241
25	MSAG5200-ISLNL-HGW-NODE491281
26	MSAG5200-ISLPC-AGT-NODE480031
27	MSAG5200-ISLPK-PNU-NODE6136021
28	MSAG5200-ISLPRN-NODE427901
29	MSAG5200-ISLSI-DWS-NODE520721

Usage Analysis (PeoTV)



PeoTV usage analysis

- User Identifier : ***Phone_Number_hash***
- Steps:
  - Handle null values
  - Drop unnecessary columns
  - Pivot by **month-year** usage
  - Transform Usage duration into measurable units (hours)
  - Generating Usage rating by Usage **WatchTime (Duration)**
  - Bin by User Locations
  - Identify PeoTV Packages and bin users by packages
  - Extract Usage and ratings (Categorical & Scaled)
  - Provide visualizations

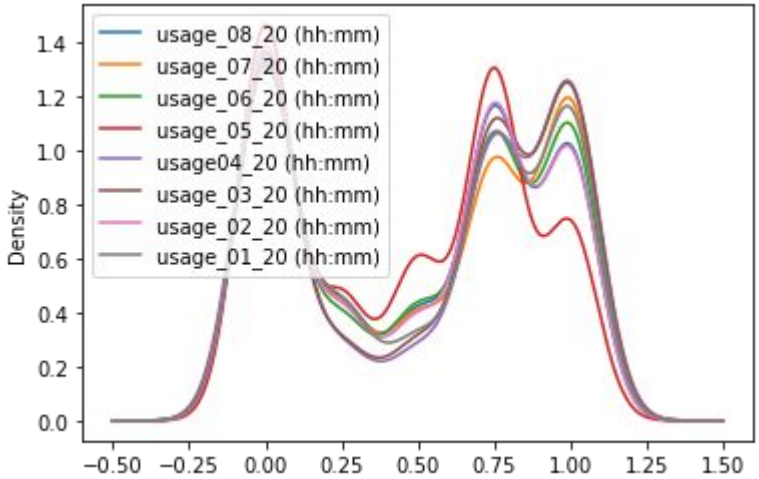
PeoTV usage: Transformed WatchTime

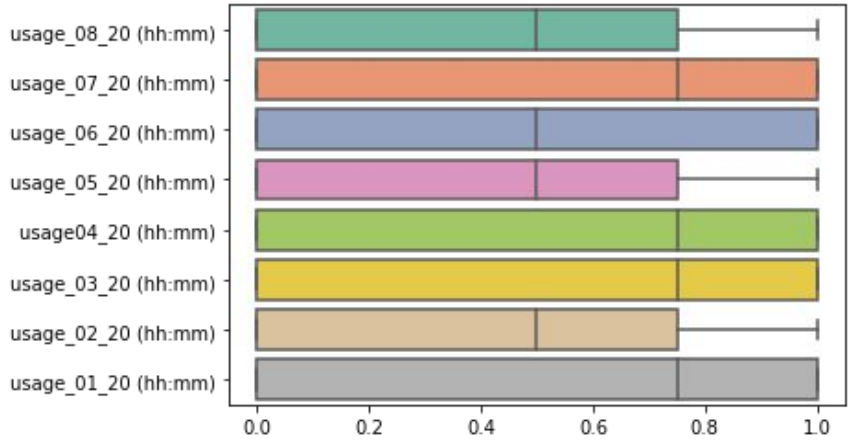
Phone_Number.hash	Peo_TV_Package	usage_08_20 (hh:mm)	usage_07_20 (hh:mm)	usage_06_20 (hh:mm)	usage_05_20 (hh:mm)	usage04_20 (hh:mm)	usage_03_20 (hh:mm)	usage_02_20 (hh:mm)	usage_01_20 (hh:mm)
46960b77dad950ddae76dd53bd7848cd	PEO_UTHAYAM	145.150000	159.250000	183.650000	212.116667	311.700000	251.400000	132.233333	225.800000
71d08664b609dcaddf02d6ffc2532aa7	PEO_SILVER	15.216667	11.516667	17.433333	32.166667	79.183333	92.666667	16.966667	21.266667
f98bd7d6c96317fc2656bc43441df8bb	PEO_SILVER	87.333333	70.850000	101.816667	72.516667	120.066667	115.716667	67.866667	93.750000
a41d8c2af7a4d478417f2368ee852296	PEO_UTHAYAM	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
09e7342066a13ca589d4a58ea0c685ae	PEO_SILVER	136.500000	135.416667	175.100000	113.433333	172.733333	197.633333	238.700000	211.083333
...	...	...	...	...	...	...	...	...	...
3ec6441eb5df6ef9ba2bdd63f919ce8b	PEO_SILVER	192.850000	244.200000	202.566667	165.933333	208.216667	221.033333	198.550000	214.983333
013736757d2ddad83cdfaa418921605c	PEO_SILVER	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
5fc30724801d125e0048b486655723f6	PEO_SILVER	94.500000	103.316667	68.516667	77.633333	106.333333	91.033333	78.000000	67.566667
e034fe9b7cb567273a07edbad9946fde	PEO_SILVER_PLUS	72.816667	92.083333	86.350000	104.100000	151.900000	113.950000	74.233333	92.300000
2ab8436cc34b4f16de3c452300f975c3	PEO_SILVER	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	41.316667	83.000000

PeoTV usage: Descriptive analysis

	usage_08_20 (hh:mm)	usage_07_20 (hh:mm)	usage_06_20 (hh:mm)	usage_05_20 (hh:mm)	usage04_20 (hh:mm)	usage_03_20 (hh:mm)	usage_02_20 (hh:mm)	usage_01_20 (hh:mm)
count	1428.00000	1428.00000	1428.00000	1428.00000	1428.00000	1428.00000	1428.00000	1428.00000
mean	94.34700	104.30008	97.35943	73.10051	106.76998	106.35979	93.02690	101.47883
std	118.65254	127.48594	119.67962	86.37740	124.25824	123.20679	114.96505	123.34116
min	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
25%	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
50%	59.08333	61.35000	59.07500	49.12500	74.85833	76.80000	59.80000	66.05000
75%	141.67917	162.76250	151.57500	117.85000	170.72500	167.20833	140.56667	162.55000
max	743.71667	734.26667	695.78333	470.40000	664.70000	678.68333	669.75000	679.35000

PeoTV usage: Descriptive analysis





PeoTV usage: Usage Ranking

- Bined WatchTime (Hours) - 4 Bins
  - Bin ranges selected considering skewness of data distribution
  - **Freedman-Diaconis's Rule based approach (Using Interquartile range)**
  - Bins: [0, 30, 60, 150, 744]

PeoTV usage: Usage Ranking

- Overall Usage ranking
  - **Binned Usage x Binned Duration**
  - *Value between 0-16*
- *ToDo: More advanced technique to combine variables?*
- Scaled values using Min-Max scalar **(0-1)**

Reference:
https://doi.org/10.1109/ICCE48956.2021.9352111, 2021 (Pham, C. D., Anh Chu, T., Pham, H. H., Linh Dao, M., Pham, T. S., Hung Trinh, V., & Nguyen, D. H. A recommendation system for offers in telecommunications)

http://www.utstat.toronto.edu/~brunner/DataAnalysisText/Interactions.pdf

PeoTV usage: Usage Ranking Categories

- Usage Rankings Categorized into four categories
  - Categories: LOW/ MEDIUM/ HIGH/ NO
  - **NO - No usage at all**
  - **Bin sizes (equal size strata)**
  - Bins: [0, 0.33, 0.66, 1]

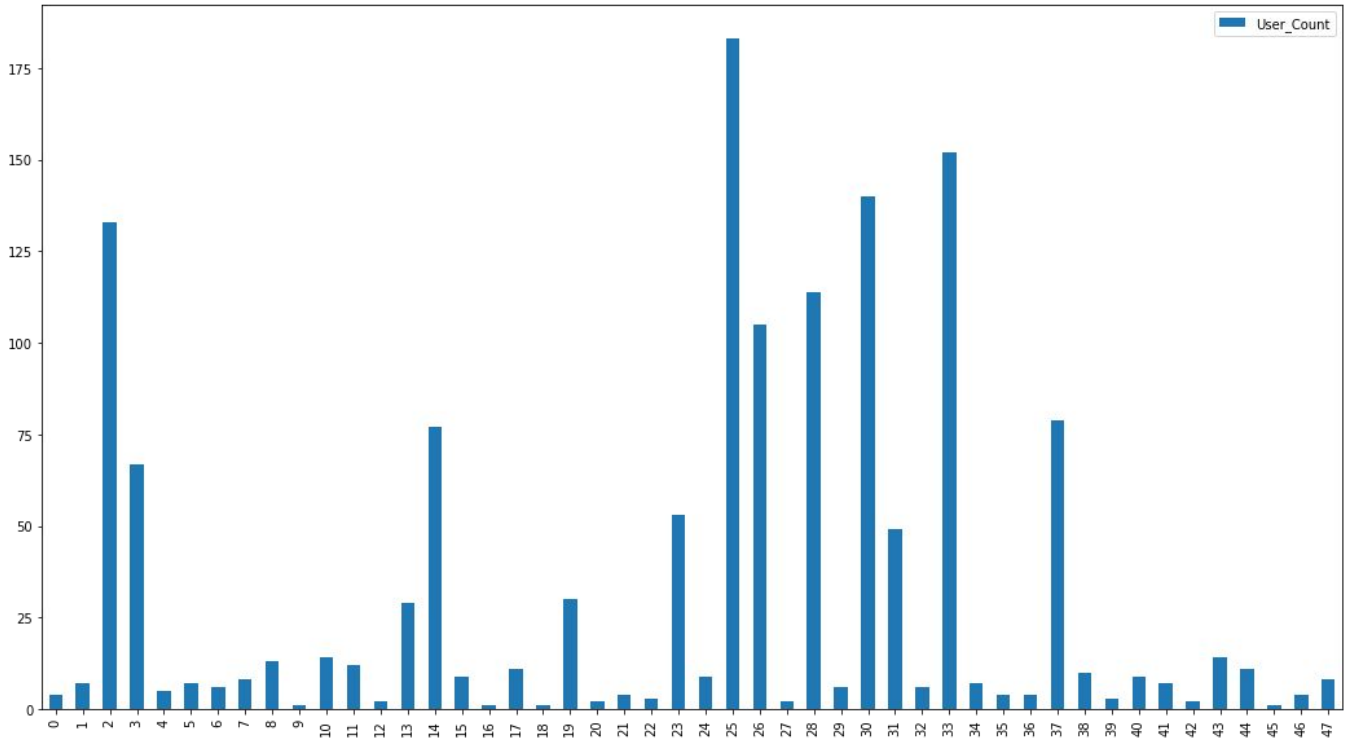
PeoTV subscribers: Group by Location

- Combined four columns together. To identify unique location of a subscriber
    - **MSAN**
    - **Location_Code**
    - **EQUP_ID**
    - **EQUP_Index**
- Added new feature to User Profile : **Location_Identifier**

	Location_Identifier	User_Count
0	AD-SRV-NODEMA5603T10224991	4
1	AG-PNP-NODEZXDSL9806H-ISL529761	7
2	BG-NODEMSAG5200431391	133
3	BZ-NODEMSAG5200-ISL435881	67
4	GE-NODEMSAG5200-ISL5577981	5
5	HC-NAP-NODEMSAG5200-ISL478721	7
6	HE-DMD-NODEC300M15221561	6
7	HK-OCC-NODEMSAG5200-ISL506561	8
8	HNT-NODEMSAG5200-ISL423851	13
9	HO-NODEMSAG5200-ISL524692	1
10	HPG-NODEMSAG5200432141	14
11	HT-NE-NODEC300M8152461	12
12	HZ-BRL-NODEMSAG5200-ISL471911	2
13	IDH-IDJ-NODEC300M11065541	29
14	IDH-KGW-NODEMSAG5200-ISL6194491	77
15	IM-NODEUA5000(IPMB)391531	9

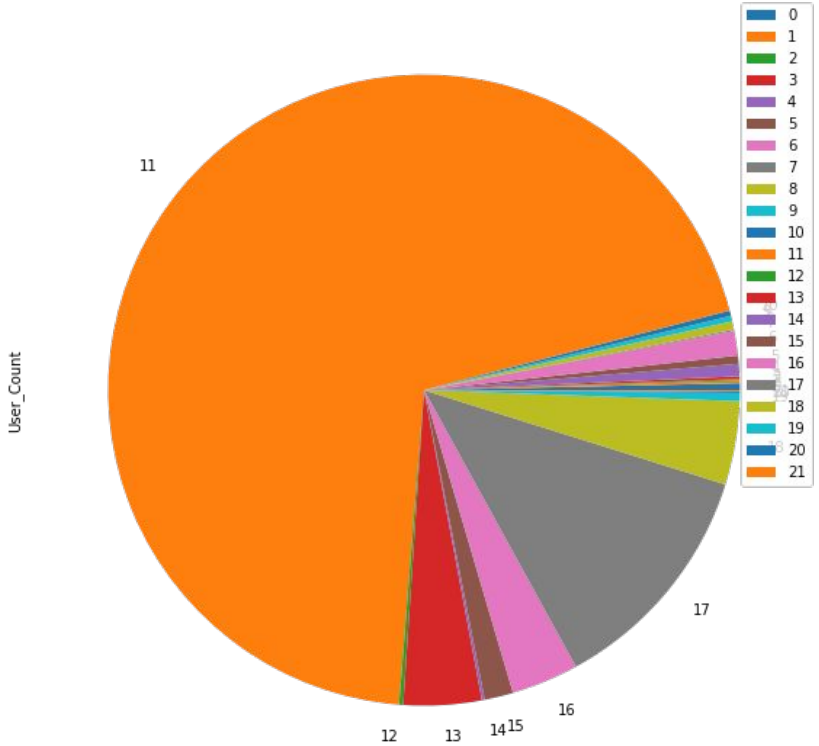
PEOTV DATA

PeoTV subscribers: Group by Location



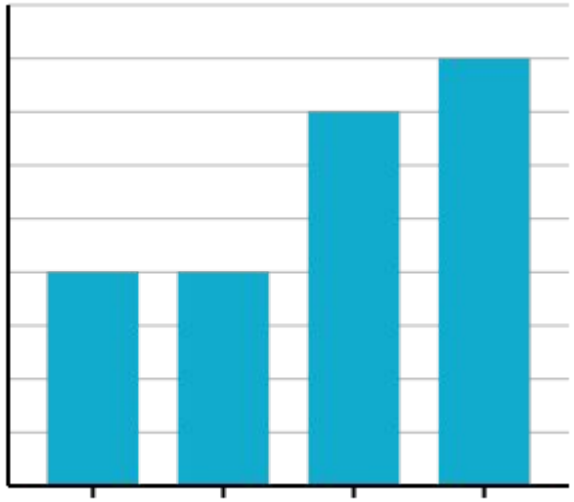
0	AD-SRV-NODEMA5603T10224991
1	AG-PNP-NODEZXDSL9806H-ISL529761
2	BG-NODEMSAG5200431391
3	BZ-NODEMSAG5200-ISL435881
4	GE-NODEMSAG5200-ISL5577981
5	HC-NAP-NODEMSAG5200-ISL478721
6	HE-DMD-NODEC300M15221561
7	HK-OCC-NODEMSAG5200-ISL506561
8	HNT-NODEMSAG5200-ISL423851
9	HO-NODEMSAG5200-ISL524692
10	HPG-NODEMSAG5200432141
11	HT-NE-NODEC300M8152461
12	HZ-BRL-NODEMSAG5200-ISL471911
13	IDH-IDJ-NODEC300M11065541
14	IDH-KGW-NODEMSAG5200-ISL6194491
15	IM-NODEUA5000(IPMB)391531
16	JA-PKV-NODEMSAG5200-ISL460421
17	KE-NRP-NODEZXDSL9806H-ISL495731
18	KI-KRB-NODEMSAG5200-ISL479411
19	KI-NODEC300M10668164
20	KL-PRC-NODEZXDSL9806H-ISL482361
21	MB-PSL-NODEMSAG5200-ISL482261
22	MGE-NODEMSAG5200428381
23	MH-POL-NODEMSAG5200-ISL491451
24	MLT-VMD-NODEMSAG5200-ISL4145091
25	MT-AVR-NODEMSAG5200-ISL467161
26	MV-GDM-NODEMSAG5200-ISL482491
27	MX-NODEMSAG5200-ISL5391821
28	ND-PHG-NODEMSAG5200-ISL500241

PeoTV subscribers: Group by PeoTV Packages



0	BOC
1	DOUBLEPROMO
2	LEISURE_SPORTS
3	LOTUS
4	NEW_SLT_STAFF_PACKAGE
5	NEW_YEAR_TREAT
6	PEO_ENTERTAINMENT
7	PEO_FAMILY
8	PEO_GOLD
9	PEO_PLATINUM
10	PEO_RELIGIOUS_PACKAGE
11	PEO_SILVER
12	PEO_SILVER_FTTH
13	PEO_SILVER_PLUS
14	PEO_SILVER_PLUS_FTTH
15	PEO_TITANIUM
16	PEO_UNNATHAM
17	PEO_UTHAYAM
18	PRANAMA
19	SLTStaff_PEO_SILVER
20	SLT_STAFF_PEO_UNNATHAM_FTTH
21	SLT_STAFF_TRIPLE_PLAY

Product State Changes Analysis



Product State Changes analysis

- User Identifier: ***Phone_Number.hash***
- Steps:
  - Identify Products List associated with users
  - Group phone numbers associated with each account
  - Group products purchased by each account/ phone number
  - Products count per each account/ phone number
  - Group users by products
  - Export generated data and maps for future use
  - Provide visualizations

Phone numbers and Products associated

ACCOUNT_NUM.hash	Phone_Number.hash	PRODUCT_NAME
0009e7e4d940c2a539b89342af07e7f1	[5e411c13c17e851e8cdf6fbdcc10537a]	[AB_Service Vacation]
008bca99f0cccb5d07d4f03744709cf8	[ba3e4a690c1d811b4192005e480bcd55]	[V_Installment]
01cbc9b050fa000d1de651103cd30fc3	[70d7089a840256ad28912710e9cdc608, 70d7089a840...	[E_PeoTV Initiation, E_Video on Demand, E_TSTV]
02059e815d999bda708149ae263946d3	[a554cc1065efdea68627e133958bca32, a554cc1065e...	[V_Call Forwarding - Immediate, V_Call Forward...
026368f4bccda09644665c13cbe13079	[b0247cfa960b7d693a0e3e7d8e1eebd7, b0247cfa960...	[E_PeoTV Initiation, E_Video on Demand, E_TSTV...
...	...	...
fd3b30118c8c6faa2ba4dc83e1f35bab	[76d2be98265496481fa8ed9700e256fe, 76d2be98265...	[V_Call back on busy, V_Call holding, V_Call F...
fdf274d421e09f7677f3b579d45aa60b	[d9f57e2dfadee74f1bc968513e2ad476, d9f57e2dfad...	[V_Call holding, V_Incoming Call Memory, V_Inc...
fea406eb917dcb3f3e93fec553e5d897	[7961eca52aa61ed953d0dff1a2a5a390]	[AB_Relocation]
ff0f03a90ae06a091ac177504385e69a	[ab7fe1473cfcfcc995de024aa4ce9c94, ab7fe1473cf...	[AB_Copper Access Bearer, AB_Megaline]
ff98ca51d1c83549f22f957b7de68815	[c035b93b6ba248056606ecf6ec725c35, c035b93b6ba...	[V_Call holding, V_Outgoing Call Memory, V_Abs...

Maps: Products purchased by Phone numbers

- Examples:

{'b635fc31617ffc9c1e6b5015fb32e108': 'V_Single VAS Bundle'},

{'ef74b1171331411a4d2e9fada9158215': 'E_Referral offer Discount'},

{'31110bf10f07e3d5e23da9680b1dd287': 'V_Absentee service'},

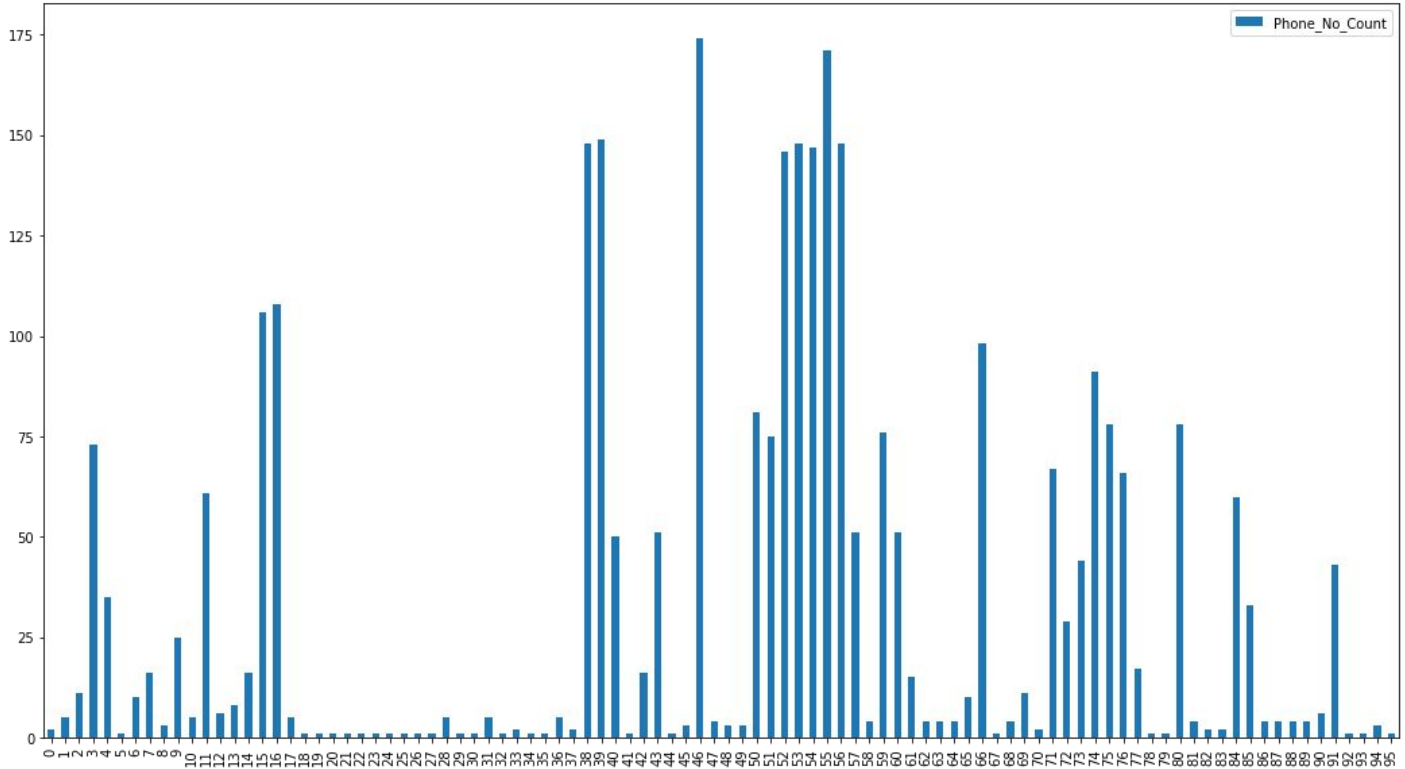
{'b6404a24d0ae0f5dbb11c78c783bd1fc': 'V_Hotline Service'},

{'b635fc31617ffc9c1e6b5015fb32e108': 'V_Caller Line Identification'},

{'b6404a24d0ae0f5dbb11c78c783bd1fc': 'V_Absentee service'},

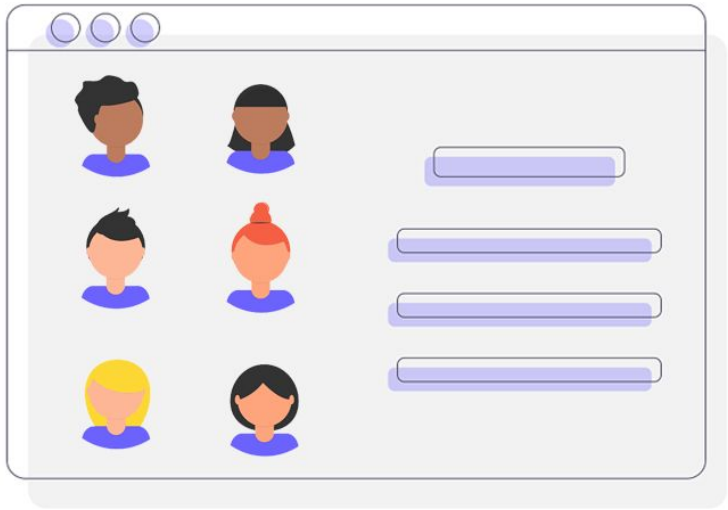
{'b635fc31617ffc9c1e6b5015fb32e108': 'V_Call Forwarding Offline'},

Users grouped by Products



	Products
0	V_Single VAS Bundle
1	E_Referral offer Discount
2	V_Absentee service
3	V_Hotline Service
4	V_Caller Line Identification
...	...
91	P_Call Forwarding by time
92	V_Basic Phone
93	P_Incoming Call Memory
94	P_Anonymous call barring
95	P_Call park

Extended User Profiles



User Profiles

- User Profile Includes:
  - **VOICE Usage Rankings** (Scaled/ Categorical)
  - **BroadBand Usage Rankings** (Scaled/ Categorical)
  - **PeoTV Usage Rankings** (Scaled/ Categorical)
  - Identified **PeoTV packages** for each user
  - Identified **VOICE packages** for each user
  - Identified **OTHER (not categorized) packages** per each user
  - Unique **Location Identifier**
  - Subscription **Type and Usage**
  - Is **Outlier**? (YES/NO)
  - Derived **socio-economic features**
    - Household types (Kids/ Working Professional or Student/ Senior citizen/ Family member abroad)
    - Having Insurance? (Tele life/Tele health) (YES/NO)

User Profile cont'd.

- VOICE Usage:
  - ONNET INCOMING Usage ranks
  - ONNET OUTGOING Usage ranks
  - OFFNET INCOMING Usage ranks
  - OFFNET OUTGOING Usage ranks
  - IDD INCOMING Usage ranks
  - IDD OUTGOING Usage ranks
  - **Overall ONNET Usage ranks**
  - **Overall OFFNET Usage ranks**
  - **Overall IDD Usage ranks**
  - **Overall VOICE Usage ranks**

- Play Types:
  - SINGLE PLAY
  - DOUBLE PLAY
  - TRIPLE PLAY

- Play Usage:
  - BB (HGH/LOW/MEDIUM/NO)
  - VOICE (HGH/LOW/MEDIUM/NO)
  - PEOTV (HGH/LOW/MEDIUM/NO)

User Profile: Usage ranking methods

- VOICE Overall Usage binning method:
  - 2 out 0f 3 Usage Rankings are **HIGH**, then **HIGH**
  - 2 out 0f 3 Usage Rankings are **LOW**, then **LOW**
  - 2 out 0f 3 Usage Rankings are **MEDIUM**, then **MEDIUM**
  - 2 out 0f 3 Usage Rankings are **NO**, then **LOW**
  - If **all** Usage Rankings are **NO**, then **NO**
  - If **1 LOW, 1 HIGH, 1 MEDIUM** Then **MEDIUM**

User Profile cont'd.

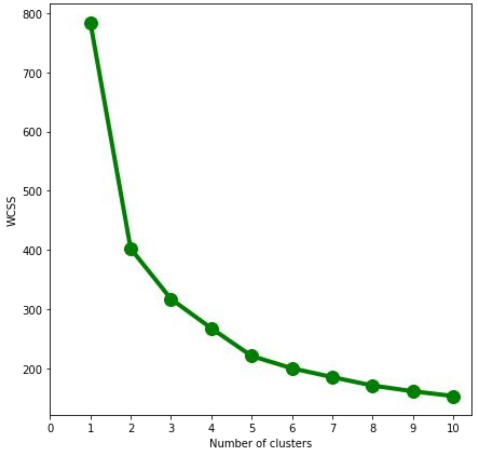
	event_source.hash	OFFNET_Overall_Categorical	IDD_Overall_Categorical	VOICE_Categorical	Play_Type	Play_Usage	Voice_Packages	Other_Packages	Household_Types	Having_Insurance
0	00221451f705ebe26051158bb14f567a	LOW	NO	LOW	SINGLE_PLAY	[VOICE_LOW, BB_NO, TV_NO]	NO	NO	NO	NO
1	002fb47f60400713c854b0f69ff78c0a	LOW	LOW	LOW	SINGLE_PLAY	[VOICE_LOW, BB_NO, TV_NO]	NO	NO	NO	NO
2	00567a4a3c474aca1ce5cd6570648932	LOW	NO	LOW	TRIPLE_PLAY	[VOICE_LOW, BB_MEDIUM, TV_MEDIUM]	NO	NO	NO	NO
3	00724ae2470df1ed0fe919800d02517d	LOW	NO	LOW	DOUBLE_PLAY	[VOICE_LOW, BB_LOW, TV_NO]	NO	NO	NO	NO
4	007400db8afa996e788d23986739f3db	LOW	LOW	LOW	SINGLE_PLAY	[VOICE_LOW, BB_NO, TV_NO]	NO	NO	NO	NO
...	...	...	...	...	...	...	...	...	...	...
3302	ffb4fa6778bc5b3d9c858f5f4fa1e0c2	LOW	LOW	LOW	SINGLE_PLAY	[VOICE_LOW, BB_NO, TV_NO]	NO	NO	NO	NO
3303	ffb63ff99cf1354dff1da862f379da80	HIGH	NO	MEDIUM	SINGLE_PLAY	[VOICE_MEDIUM, BB_NO, TV_NO]	NO	NO	NO	NO
3304	ffd697d66cc01c8897ce3e37d693673a	LOW	NO	LOW	DOUBLE_PLAY	[VOICE_LOW, BB_LOW, TV_NO]	NO	NO	NO	NO
3305	ffd96f173d6cd683bf3384bdc2f99713	MEDIUM	MEDIUM	MEDIUM	SINGLE_PLAY	[VOICE_MEDIUM, BB_NO, TV_NO]	NO	NO	NO	NO
3306	ffee845c0caa320b55e31bb156827d16	LOW	NO	LOW	DOUBLE_PLAY	[VOICE_LOW, BB_NO, TV_HIGH]	NO	NO	[Kids]	NO

User Profile : Clustering Users

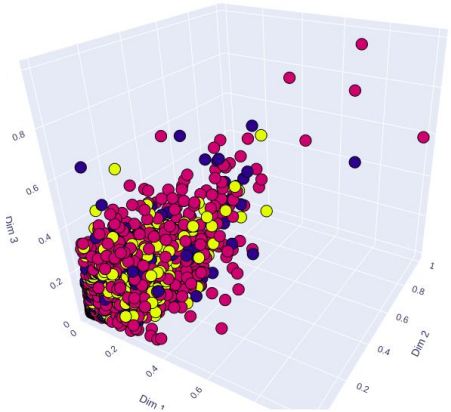
- K-Means Clustering based on Scaled user features (Usage Ratings)
  - Best no of clusters (k) was selected using Elbow method
  - Three clusters were identified
    - Cluster 1: 914 users,
    - Cluster 2: 2145 users,
    - Cluster 3: 248 users

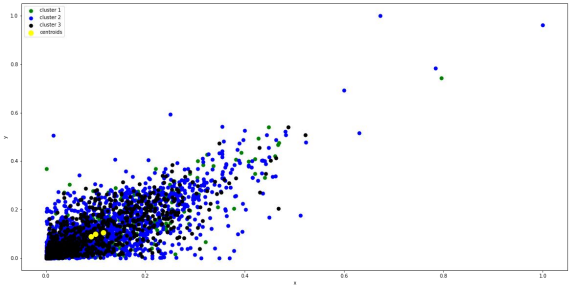
- Hierarchical (Agglomerative) Clustering based on Scaled user features (Usage Ratings)
  - Best no of clusters were identified as 3.
  - Considered euclidean distances

User Profile : Clustering Users



KMeans clustering - ELBOW

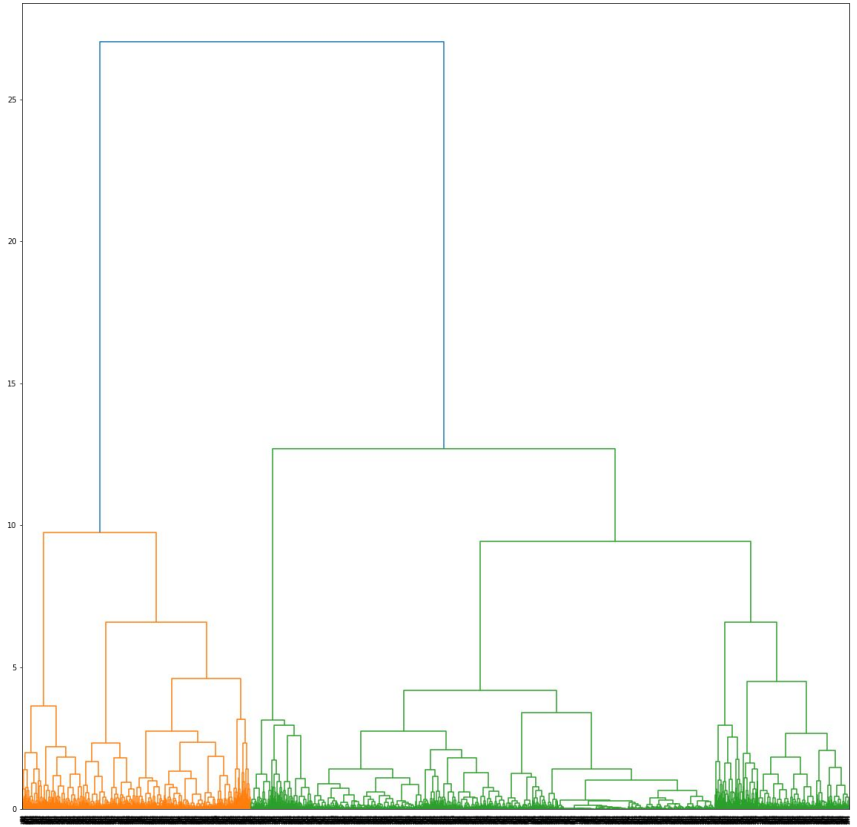




Three clusters identified

User Profile : Clustering Users

Clustering: Dendrogram



Extended Product Profiles

Product Profiles

- Product Profile Includes:
  - **Product/package details**
  - **Subscription Plans**
  - **Price (Downpayment)**
  - **Monthly Installments**
  - **Data plans (for BroadBand packages) (Time-based/Anytime/Unlimited)**
  - **Call Charges (VOICE packages)**
  - **TV Channels and monthly charges (PeoTV Packages)**
  - **Value Added services details (VAS)**
  - **Conditions**
  - *** Dependent Products**
  - *** Available locations**

Data Extracted From:
SLT Web site, Annual report 2020

Product Profiles: Available features

- Common features:
  - Product_ID - product code
  - Base_Type (BB/VOICE/PEOTV)
  - Pricing_Type (PAID/FREE)
  - Package_Type (ADSL/Fibre/4G/Telephone/PeoTV)
  - VAS (YES/NO) - value added service
  - Title and Description
  - Included_Packages - packages shipped with a product
  - Price (Rs.) - totl cost, downpayment or first installment
  - Monthly_Rental (Rs.)
  - Subscription_Type (SINGLE_PLAY/DOUBLE_PLAY/TRIPLE_PLAY)
  - Minimum_Subscription_Period (years)
  - Recidence_Type (Home/Office)
  - Tax_Status (INCLUDED/EXCLUDED)
  - Conditions - list of conditions for package
  - Available_Regions - list of available regions: MSAN or related level
  - Dependent_Packages (other products it depends on)

Product Profiles: Available features

- BroadBand specific features:
  - BB_Data_standard (GB) - Standard data for a Time-based package
  - BB_Data_Free (GB) - Free data for a Time-based package
  - BB_Data_Anytime (GB) - data for anytime package
  - BB_Data_Unlimited (GB) - data for unlimited package
  - BB_Connection_Type (Time-based/Anytime/Unlimited)
  - BB_Connection_Speed (Download Speed/ Upload Speed)

- PeoTV specific features:
  - PEOTV_No_of_Channels - No of channels in a PEO TV package

- VOICE specific features:
  - VOICE_Home_SLT_Instrument_Rental (Rs.) - Home Telephone rental (with SLT provided telephone)
  - VOICE_Home_Customer_Instrument_Rental (Rs.) - Home Telephone rental (with Customer provided telephone)
  - VOICE_Charge_Active_Hours (SLT-STL, SLT-Other) (Rs.) - Voice calls charges for Active hours
  - VOICE_Charge_Leisure_Hours (SLT-STL, SLT-Other) (Rs.) - Voice calls charges for Leisure hours
  - VOICE_Free_Minutes - Free voice call minutes given per package
  - VOICE_Telehelth_Insurance_Benefits (Rs.) - Awarded Benefit at fulfillment for SLT Telehealth Insurance packages
  - VOICE_Tele_Life_Insurance_Benefits (Rs.) - Awarded Benefit at fulfillment for SLT Tele Life Insurance packages

Product Profiles

	Product_ID	Base_Type	Pricing_Type	Package_Type	VAS	Title	Description	BB_Data_Standard	BB_Data_Free	BB_Data_Anytime	...	VOICE_Telehelth_Insurance_Benefits
0	BB_Higher_Education	BB	PAID	ADSL	NO	HIGHER EDUCATION	NaN	4.0	6	NaN	...	NaN
1	BB_Web_Lite	BB	PAID	ADSL	NO	WEB LITE	NaN	6.0	9	NaN	...	NaN
2	BB_Entree	BB	PAID	ADSL	NO	ENTREE	NaN	NaN	NaN	7.0	...	NaN
3	BB_Web_Starter_4G	BB	PAID	4G	NO	WEB STARTER 4G	NaN	11.0	17	NaN	...	NaN
4	BB_Web_Starter_ADSL	BB	PAID	ADSL	NO	WEB STARTER ADSL	NaN	11.0	17	NaN	...	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...
240	V_Referral offer Discount	VOICE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
241	V_Miscellaneous Service	VOICE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
242	E_PeoTV_Package Transfer	VOICE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
243	V_One-Time Detailed Bill	VOICE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN
244	V_Basic Phone	VOICE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN

Demonstration

Thank You