

UJIAN TENGAH SEMESTER GENAP
TAHUN AKADEMIK 2021/2022

Nama : LAILA SIRRI HAYATI

NIM : 19611050

Pengantar Data Mining - C

Jawab:

1. Dipunyai data pelanggaran lalu lintas ("traffic violations.csv") yang dapat di muat turun dari website <https://data.amerigeoss.org/dataset/traffic-violations-56dda>
 - a. Buatlah analisis deskriptif dari data tersebut yang memudahkan bagi para pengambil kebijakan

Jawab :

```
> summary(trans)
transactions as itemMatrix in sparse format with
906990 rows (elements/itemsets/transactions) and
14467 columns (items) and a density of 0.0001217975

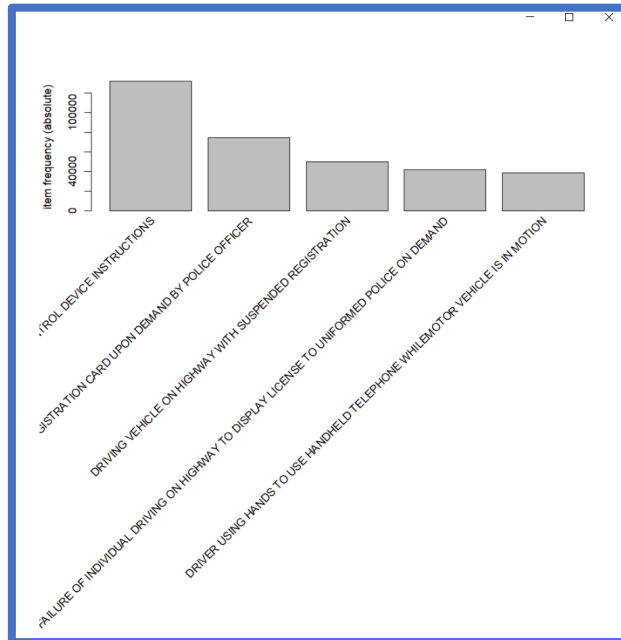
most frequent items:
      DRIVER FAILURE TO OBEY PROPERLY PLACED TRAFFIC CONTROL DEVICE INSTRUCTIONS
                                                    131981
      FAILURE TO DISPLAY REGISTRATION CARD UPON DEMAND BY POLICE OFFICER
                                                    74437
      DRIVING VEHICLE ON HIGHWAY WITH SUSPENDED REGISTRATION
                                                    50160
FAILURE OF INDIVIDUAL DRIVING ON HIGHWAY TO DISPLAY LICENSE TO UNIFORMED POLICE ON DEMAND
                                                    41895
      DRIVER USING HANDS TO USE HANDHELD TELEPHONE WHILEMOTOR VEHICLE IS IN MOTION
                                                    38941
      (other)
                                                    1260743
```

Berdasarkan *output* di atas dapat dilihat bahwa data banyak pelanggaran lalu lintas 906990 dari 14467 jenis pelanggaran lalu lintas. Adapun 5 teratas dari kejadian yang sering terjadi di lalu lintas sebagai berikut.

- Driver failure to obey properly placed traffic control device instructions sebanyak 131981 data.
- Failure to display registration card upon demand by police officer sebanyak 74437 data
- Driving vehicle on highway with suspended registration sebanyak 50160 data
- Failure of individual driving on highway to display license to uniformed police on demand sebanyak 41895 data
- Driver using hands to use handheld telephone whilemotor vehicle is in motion sebanyak 38941 data

b. Tentukan faktor-faktor yang memungkinkan terjadinya pelanggaran lalu lintas

Jawab:



Berdasarkan *Output* diatas dapat dilihat bahwa 5 faktor teratas pada gambar diatas merupakan yang memungkinkan terjadinya pelanggaran lalu lintas .

c. Buatlah association rules dari beberapa faktor yang bisa didapatkan hubungan antar factor tersebut dengan menentukan nilai support, confidence dan lift rasionya.

Jawab:

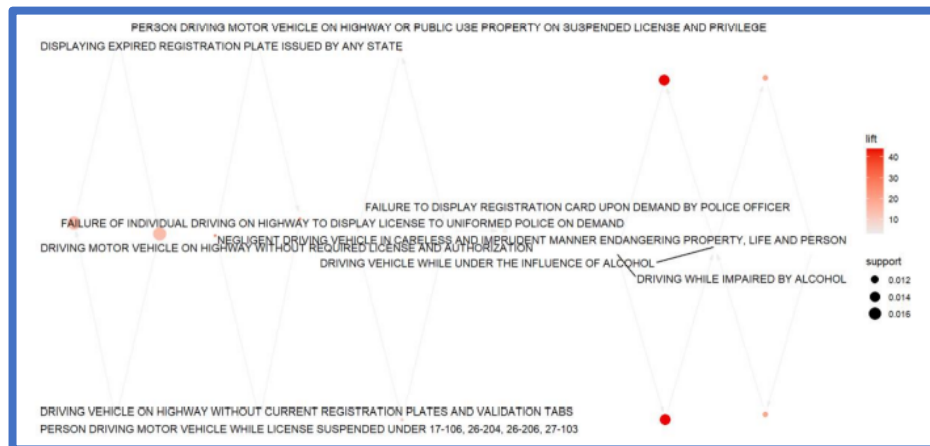
lhs	rhs
support confidence coverage lift count	
[1] [DRIVING VEHICLE ON HIGHWAY WITHOUT CURRENT REGISTRATION PLATES AND VALIDATION TABS]	=> [DISPLAYING EXPIRED REGISTRATION PLATE ISSUED BY ANY STATE]
0.01741585 0.6016836 0.02894519 14.458101 15796	=> [DRIVING VEHICLE ON HIGHWAY WITHOUT CURRENT REGISTRATION PLATES AND VALIDATION TABS]
[2] [DISPLAYING EXPIRED REGISTRATION PLATE ISSUED BY ANY STATE]	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
0.01741585 0.4184925 0.04161567 14.458101 15796	=> [DRIVING WHILE IMPAIRED BY ALCOHOL]
[3] [DRIVING WHILE IMPAIRED BY ALCOHOL]	=> [DRIVING WHILE IMPAIRED BY ALCOHOL]
0.01624163 0.9762094 0.01663745 44.758476 14731	
[4] [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]	=> [PERSON DRIVING MOTOR VEHICLE ON HIGHWAY OR PUBLIC USE PROPERTY ON SUSPENDED LICENSE AND PRIVILEGE]
0.01681510 0.9707713 0.02053521 28.351645 10595	=> [PERSON DRIVING MOTOR VEHICLE WHILE LICENSE SUSPENDED UNDER 17-106, 26-204, 26-206, 27-103]
[5] [PERSON DRIVING MOTOR VEHICLE WHILE LICENSE SUSPENDED UNDER 17-106, 26-204, 26-206, 27-103]	
0.01681510 0.9707713 0.02053521 28.351645 10595	
[6] [PERSON DRIVING MOTOR VEHICLE ON HIGHWAY OR PUBLIC USE PROPERTY ON SUSPENDED LICENSE AND PRIVILEGE]	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
0.01681510 0.3388990 0.03468086 28.351645 10595	=> [DRIVING WHILE IMPAIRED BY ALCOHOL]
[7] [DRIVING UNDER THE INFLUENCE OF ALCOHOL PER SE]	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
0.01020879 0.9960376 0.01109494 45.575884 10003	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
[8] [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
0.01020879 0.9960376 0.01109494 45.575884 10003	
[9] [DRIVING UNDER THE INFLUENCE OF ALCOHOL PER SE]	=> [DRIVING WHILE IMPAIRED BY ALCOHOL]
0.01020217 0.9934421 0.01109494 59.713156 9997	=> [DRIVING WHILE IMPAIRED BY ALCOHOL]
[10] [DRIVING WHILE IMPAIRED BY ALCOHOL]	=> [DRIVING UNDER THE INFLUENCE OF ALCOHOL PER SE]
0.01020217 0.6624927 0.01663745 59.713156 9997	
[11] [DRIVING UNDER THE INFLUENCE OF ALCOHOL PER SE]	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
0.01099571 0.9979993 0.01102217 45.738186 9973	
[12] [DRIVING UNDER THE INFLUENCE OF ALCOHOL PER SE]	=> [DRIVING WHILE IMPAIRED BY ALCOHOL]
0.01099571 0.9970009 0.01102879 59.925106 9973	
[13] [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]	
0.01099571 0.6770077 0.01624163 61.019496 9973	
[14] [FAILURE OF INDIVIDUAL DRIVING ON HIGHWAY TO DISPLAY LICENSE TO UNIFORMED POLICE ON DEMAND]	=> [FAILURE TO DISPLAY REGISTRATION CARD UPON DEMAND BY POLICE OFFICER]
0.01075205 0.2327724 0.04619125 2.836254 9752	=> [FAILURE OF INDIVIDUAL DRIVING ON HIGHWAY TO DISPLAY LICENSE TO UNIFORMED POLICE ON DEMAND]
[15] [DRIVING MOTOR VEHICLE ON HIGHWAY WITHOUT REQUIRED LICENSE AND AUTHORIZATION]	=> [DRIVING MOTOR VEHICLE ON HIGHWAY WITHOUT REQUIRED LICENSE AND AUTHORIZATION]
0.01054687 0.4474484 0.02357137 9.686866 9566	
[16] [FAILURE OF INDIVIDUAL DRIVING ON HIGHWAY TO DISPLAY LICENSE TO UNIFORMED POLICE ON DEMAND]	=> [DRIVING MOTOR VEHICLE ON HIGHWAY WITHOUT REQUIRED LICENSE AND AUTHORIZATION]
0.01054687 0.2283327 0.04619125 9.686866 9566	
[17] [DRIVING MOTOR VEHICLE ON HIGHWAY WITHOUT REQUIRED LICENSE AND AUTHORIZATION]	=> [PERSON DRIVING MOTOR VEHICLE ON HIGHWAY OR PUBLIC USE PROPERTY ON SUSPENDED LICENSE AND PRIVILEGE]
0.01033087 0.4382806 0.02357137 12.715225 9370	=> [DRIVING MOTOR VEHICLE ON HIGHWAY WITHOUT REQUIRED LICENSE AND AUTHORIZATION]
[18] [PERSON DRIVING MOTOR VEHICLE ON HIGHWAY OR PUBLIC USE PROPERTY ON SUSPENDED LICENSE AND PRIVILEGE]	=> [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]
0.01033087 0.2997153 0.04619125 12.715225 9370	
[19] [NEGLIGENT DRIVING VEHICLE IN CARELESS AND IMPRUDENT MANNER ENDANGERING PROPERTY, LIFE AND PERSON]	=> [NEGLIGENT DRIVING VEHICLE IN CARELESS AND IMPRUDENT MANNER ENDANGERING PROPERTY, LIFE AND PERSON]
0.01023936 0.4232968 0.02438990 15.407568 9287	
[20] [DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL]	
0.01023936 0.4694671 0.02181060 15.407568 9287	

Setelah di uji didapatkan 20 rules dengan menggunakan nilai *supportnya* minimal 0.01 dan *confidencenya* 0.2. Dapat dilihat pada hasil nomor 1 apabila terjadi

pelanggaran *DRIVING VEHICLE ON HIGHWAY WITHOUT CURRENT REGISTRATION PLATES AND VALIDATION TABS* maka *DISPLAYING EXPIRED REGISTRATION PLATE ISSUED BY ANY STATE* juga akan terjadi dengan nilai *support*, *confidence*, *lift*, dan *countnya* masing-masing sebesar 0.01741585, 0.6016836, 14.458101 dan 15796, begitu juga seterusnya sampai nomor 20 dimana pelanggaran *DRIVING VEHICLE WHILE UNDER THE INFLUENCE OF ALCOHOL* maka *NEGLIGENT DRIVING VEHICLE IN CARELESS AND IMPRUDENT MANNER ENDANGERING PROPERTY, LIFE AND PERSON* juga akan terjadi dengan nilai *support*, *confidence*, *lift* dan *countnya* masing-masing sebesar 0.01023936, 0.4694672, 19.407568, dan 9287.

- d. Apa insight yang didapatkan dari association rules yang anda dapatkan sebagai saran bagi pengambil kebijakan

Jawab:



Berdasarkan *output* di atas dapat dilihat bahwa pada garis terjadi pelanggaran *Negligent driving vehicle in careless and imprudent manner endangering property, life, and person* maka akan terjadi pelanggaran *driving while impaired by alcohol* dimana artinya bahwa seseorang membawa kendaraan tidak hati – hati sampai membahayakan properti dan nyawa seseorang karena orang berkendara tersebut di bawah pengaruh alkohol. Oleh sebab itu, perlu adanya kebijakan terkait keamanan demi memperketat keamanan di sekitar wilayah tersebut guna untuk menghindari kejadian seseorang mengendarai dalam kondisi pengaruh alkohol.

2. Dipunyai data pengguna internet dunia yang ada pada <https://www.internetworldstats.com/stats.htm> Berilah analisa terhadap beberapa pertanyaan berikut,

a. Lakukan analisis deskriptif untuk data di benua Asia tersebut.

Jawab:

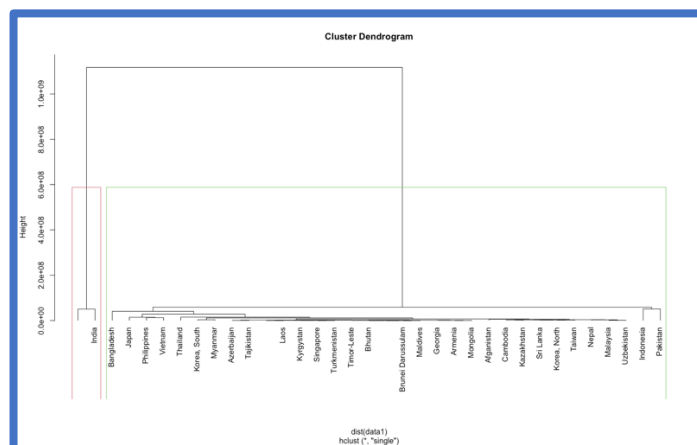
```
> summary(data)
      ASIA      Population..2021.Est.. Internet.Users..Year.2000. Internet.Users.30.JUNE.2021
Length:35      Min.   :4.415e+05      Min.   :      0      Min.   : 20000
Class :character 1st Qu.:6.007e+06      1st Qu.: 6000      1st Qu.: 2876628
Mode :character  Median :2.150e+07      Median : 51600      Median : 7991630
                  Mean  :1.236e+08      Mean  : 3265829      Mean  : 79200375
                  3rd Qu.:6.238e+07      3rd Qu.: 2000000      3rd Qu.: 53960542
                  Max.   :1.444e+09      Max.   :47080000      Max.   :989080566

Penetration.Population Users.Asia Facebook.30.JUNE.2021
Min.   :0.0010      Min.   :0.00000      Min.   : 14000
1st Qu.:0.5135      1st Qu.:0.00100      1st Qu.: 2301500
Median :0.7380      Median :0.00300      Median : 4996000
Mean   :0.6684      Mean   :0.02854      Mean   : 33139063
3rd Qu.:0.8295      3rd Qu.:0.01950      3rd Qu.: 27120000
Max.   :1.0450      Max.   :0.35700      Max.   :416600000
```

Berdasarkan *Output summary* di atas dapat dilihat bahwa data tersebut merupakan data pada benua Asia, dapat dilihat pada tahun 2021 bahwa Benua Asia memiliki populasi terkecil 4.415×10^5 dan nilai populasi terbesar 1.444×10^9 dengan rata-rata populasinya sebesar 1.236×10^8 . Selanjutnya pada tahun 2000 pengguna internet dengan jumlah paling sedikit adalah 0 dan jumlah pengguna internet paling banyak adalah 47080000 pengguna, rata-rata pengguna internet di Benua Asia yaitu 3265829. Kemudian 2021, kini rata-rata pengguna internet meningkat sebesar 79200375 pengguna, dengan paling sedikit adalah 20000 pengguna dan paling banyak adalah 989080566 pengguna.

b. Buatlah kluster untuk negara-negara di benua Asia menurut Populasinya

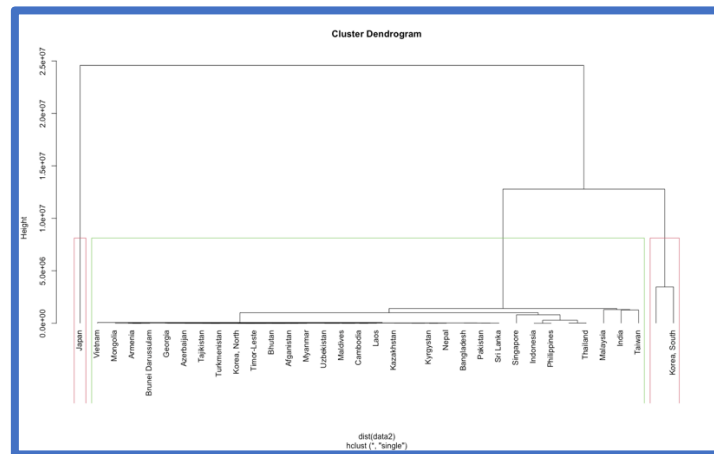
Jawab:



Berdasarkan *output* di atas dapat dilihat bahwa cluster untuk benua Asia berdasarkan populasinya terbagi menjadi 2 cluster, dimana cluster 1 beranggotakan negara Bangladesh sampai dengan negara Pakistan, sedangkan cluster 2 beranggotakan China dan India.

- c. Buatlah kluster untuk negara-negara di benua Asia menurut Pengguna internet pada tahun 2000.

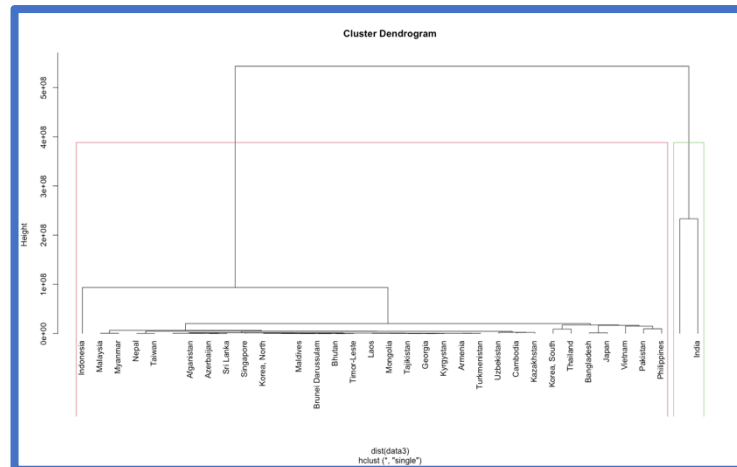
Jawab:



Berdasarkan *output* di atas dapat dilihat bahwa cluster untuk benua Asia menurut pengguna internet tahun 2000 yang terbentuk sebanyak 3 cluster. Cluster 1 beranggotakan negara Vietnam sampai negara Taiwan, cluster 2 beranggotakan China dan Korea Selatan, dan cluster 3 beranggotakan Jepang.

- d. Buatlah kluster untuk negara-negara di benua Asia menurut Pengguna internet pada tahun 2021 (sampai 30 Juni)

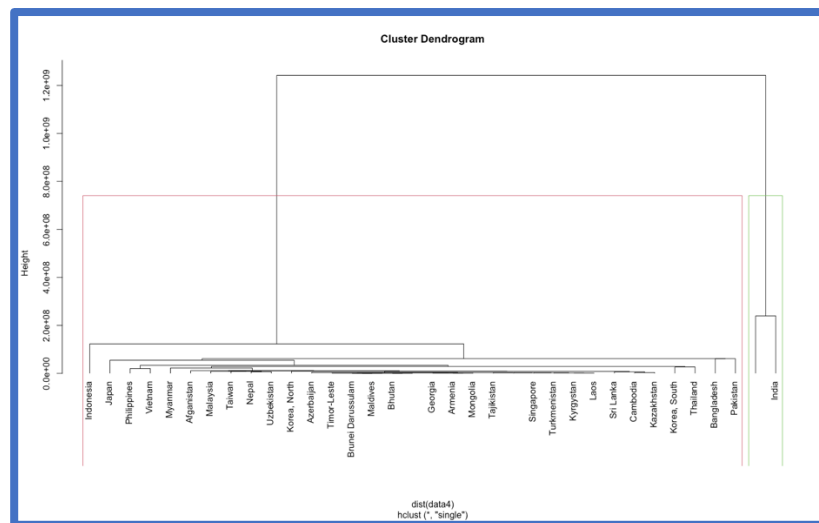
Jawab:



Berdasarkan *output* di atas dapat dilihat bahwa cluster untuk benua Asia berdasarkan pengguna internet tahun 2021 (sampai 30 Juni) yang terbentuk sebanyak 2 cluster. Untuk cluster 1 beranggotakan negara Indonesia sampai negara Philippines sedangkan cluster 2 memiliki 2 anggota yaitu China dan India.

- e. Buatlah kluster untuk negara-negara di benua Asia menurut Populasi, Pengguna internet pada tahun 2000 dan pengguna internet pada tahun 2021 (sampai 30 Juni).

Jawab:

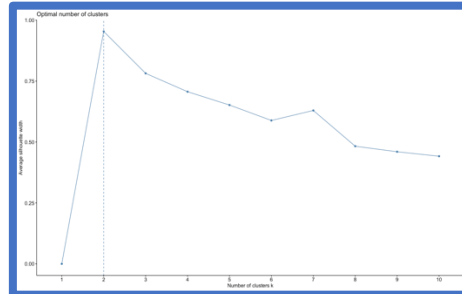


Berdasarkan populasinya, pengguna internet pada tahun 2000 dan pengguna internet pada tahun 2021 di benua Asia dikelompokkan menjadi 2 cluster. Untuk cluster 1 beranggotakan negara Indonesia sampai Pakistan, sedangkan cluster 2 terdapat 2 negara yaitu China dan India.

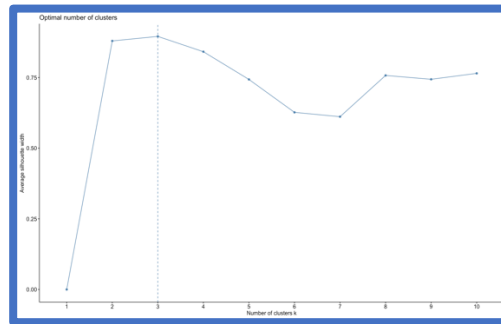
f. Untuk pertanyaan dari b sampai e berapakah nilai optimal jumlah klusternya?

Jawab:

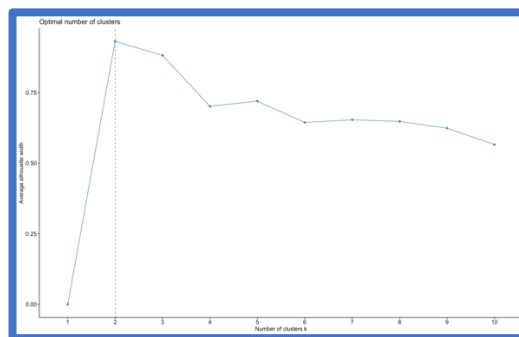
Berdasarkan hasil yang diperoleh dengan metode *silhoutte*, didapatkan hasil:



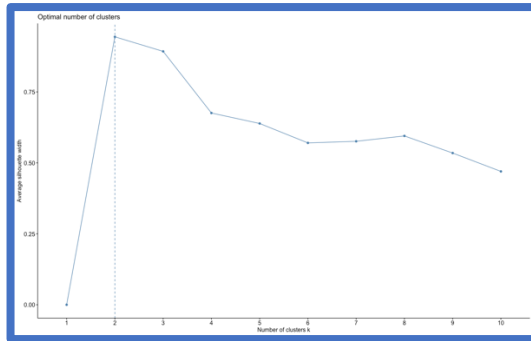
→ Jumlah cluster optimal untuk populasi di benua Asia adalah 2 cluster.



→ Jumlah cluster optimal untuk benua Asia berdasarkan pengguna internet tahun 2000 sebanyak 3 cluster.



→ Jumlah cluster optimal untuk benua Asia berdasarkan pengguna internet tahun 2021 (sampai 30 Juni) sebanyak 2 cluster.



➔ Jumlah cluster optimal untuk benua Asia berdasarkan populasi, pengguna internet tahun 2000 dan pengguna internet tahun 2021 didapatkan sebanyak 2 cluster.

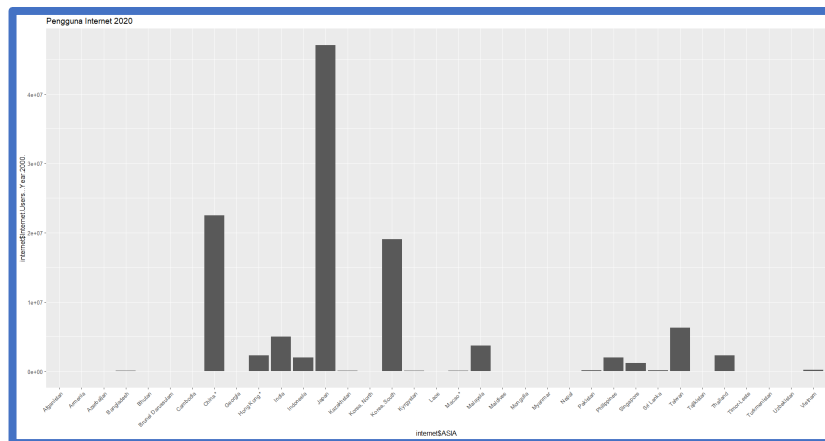
g. Negara-negara manakah yang berpindah kluster menuju kluster lebih baik untuk pengguna internetnya?

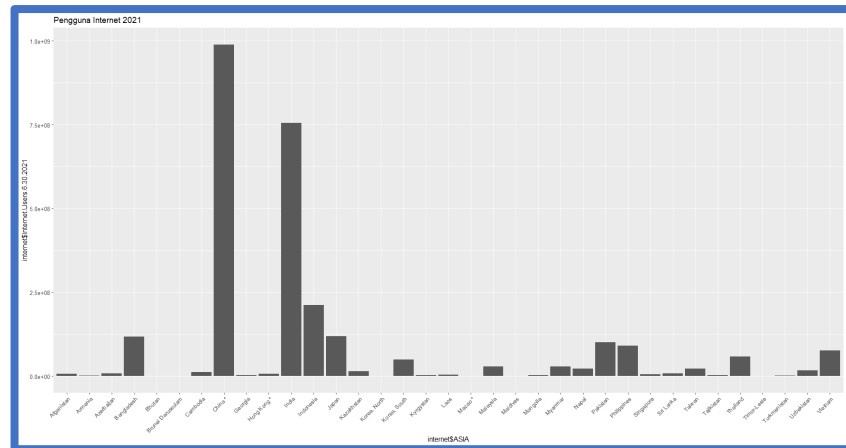
Jawab:

Negara yang berpindah cluster menuju cluster lebih baik yaitu india dimana negara tersebut memiliki perkembangan pengguna internet yang baik.

h. Negara manakah yang mempunyai percepatan pertumbuhan pengguna Internet tertinggi?

Jawab:





Berdasarkan *output* diatas dapat dilihat bahwa negara dengan percepatan pertumbuhan pengguna internet tertinggi adalah China.

- i. Bagaimana posisi Indonesia di Asia menurut pengguna internetnya? Apa insight yang muncul menurut anda terhadap hasil kluster yang anda buat?

Jawab:

Berdasarkan *output* visualisasi pada poin h, Indonesia pada tahun 2000 berada pada posisi dengan urutan ke- 9, namun ketika tahun 2021 Indonesia naik ke urutan ke-3 yaitu dengan 212.354.070 pengguna.

ASIA INTERNET USE, POPULATION STATISTICS DATA AND FACEBOOK DATA - MID-YEAR 2021						
ASIA	Population (2021 Est.)	Internet Users, (Year 2000)	Internet Users 30-JUNE-2021	Penetration % Population	Users % Asia	Facebook 30-JUNE-2021
Afganistan	39,835,428	1,000	7,337,489	18.4 %	0.3 %	4,417,500
Armenia	2,968,127	30,000	2,126,716	71.7 %	0.1 %	1,863,000
Azerbaijan	10,223,342	12,000	7,991,630	78.2 %	0.3 %	4,327,000
Bangladesh	166,303,498	100,000	117,310,000	70.5 %	4.2 %	48,800,000
Bhutan	779,898	500	538,499	69.0 %	0.0 %	524,200
Brunei Darussulam	441,532	30,000	461,600	104.5 %	0.0 %	461,600
Cambodia	16,946,438	6,000	12,444,000	73.4 %	0.4 %	12,444,000
China *	1,444,216,107	22,500,000	989,080,566	68.5 %	35.7 %	3,859,500
Georgia	3,979,765	20,000	3,222,000	81.0 %	0.1 %	3,222,000
Hong Kong *	7,552,810	2,283,000	6,698,252	88.7 %	0.2 %	6,165,000
India	1,393,409,038	5,000,000	755,820,000	54.2 %	27.3 %	416,600,000
Indonesia	276,361,783	2,000,000	212,354,070	76.8 %	7.7 %	176,500,000
Japan	125,570,000	47,080,000	118,626,672	94.5 %	4.3 %	55,840,000

Gambar 1. Pengguna Internet di benua Asia

