

TUGAS 2

SELEKSI WARGA BASDAT

Oleh :

Ramos Janoah Hasudungan 13514089

Vitra Chandra 13514043

Kelompok : 3

I. PENJELASAN DATASET

Dataset yang didapat bernama Speed Dating Data. Data tersebut berisi tentang data seseorang yang mengikuti take me out. Pada data tersebut, ada banyak sekali atribut per record. Berikut adalah beberapa data yang

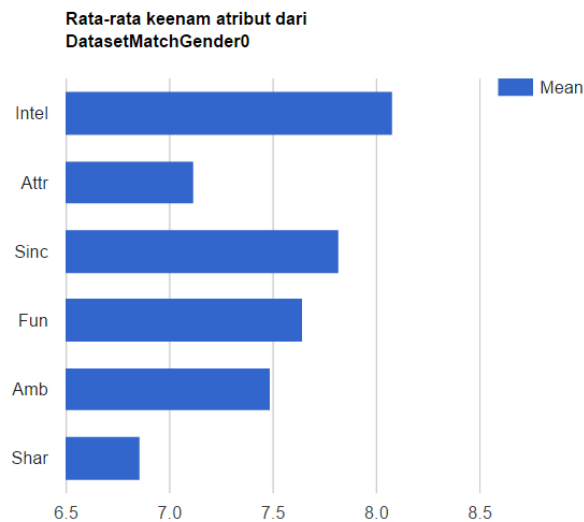
- Id dan iid dari partisipan.
- Id dari partner partisipan
- Pendapat partisipan terhadap 6 aspek pasangan (Attractive, Intelligence, Sincere, Ambitious, Funny, Shared Interest)
- Apakah partisipan 'match' dengan pasangannya
- Tingkat kesukaan partisipan dengan pasangannya
- Apa yang partisipan inginkan dari pasangannya, pada saat registrasi, pada saat event, pada saat follow-up pertama, dan follow-up kedua, dinilai dari 6 aspek.
- Karir partisipan
- Pendapatan partisipan
- Ras partisipan
- Pendidikan partisipan
- Apa yang partisipan pikirkan tentang pendapat orang lain terhadap dia dari 6 aspek.
- Target mengikuti acara Take Me Out
- Dsb...

Dari data tersebut, kami diminta untuk melakukan eksplorasi terhadap data tersebut dengan hipotesis tertentu.

II. ANALISIS

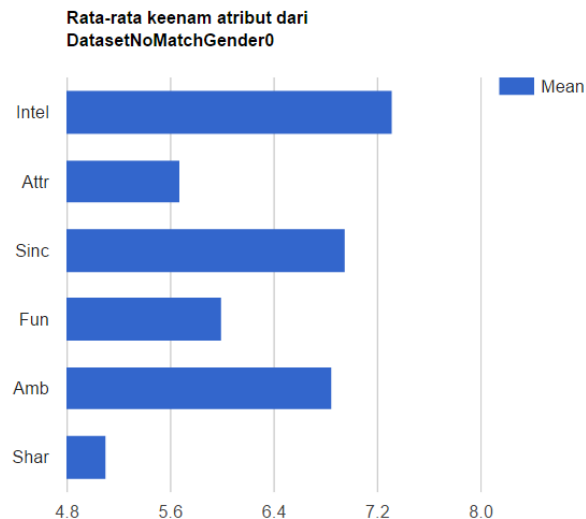
Pada analisis ini, saya menggunakan R, dan menggunakan IDE RStudio. Package-package yang digunakan untuk visualisasi adalah googleVis dan corrplot. Berikut adalah analisis dari pemecahan persoalan-persoalan yang diberikan.

1. **Membuktikan bahwa ketertarikan wanita pada pria cenderung pada kemampuan intelegensia (intelligence) dari calon partnernya dan pria kepada wanita pada fisiknya (attractive).**
 - a. Membuktikan bahwa ketertarikan wanita pada pria cenderung pada kemampuan intelegensia (intelligence) dari calon partnernya.
 - Analisis pertama : Melihat rata-rata dari 6 atribut yaitu intel, attr, sinc, fun, amb, shar dari subset data wanita (gender = 0) yang match (match = 1), dan rata-rata dari subset data wanita (gender = 0) yang tidak match (match = 0).



Data: data • Chart ID: BarChartIDe708b658db • googleVis-0.5.10
R version 3.3.0 (2016-05-03) • Google Terms of Use • Documentation and Data Policy

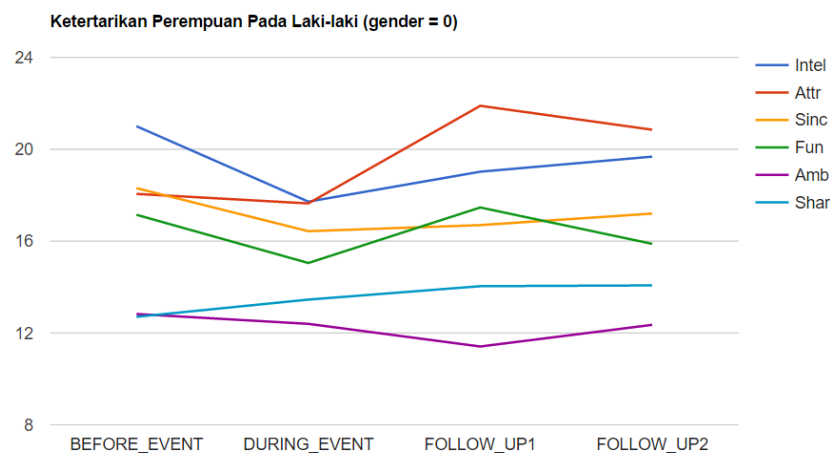
Gambar 1. Histogram dari ketertarikan perempuan terhadap partnernya dari subset data perempuan yang mengalami match.



Data: data • Chart ID: BarChartIDe701ce4783d • googleVis-0.5.10
R version 3.3.0 (2016-05-03) • Google Terms of Use • Documentation and Data Policy

Gambar 2. Histogram dari ketertarikan perempuan terhadap partnernya dari subset data perempuan yang tidak mengalami match.

- Analisis kedua : Melihat rata-rata dari ketertarikan wanita (gender = 0) pada saat registrasi (atribut dengan ujung 1_1), pada saat di tengah-tengah event (1_s), pada saat follow-up pertama (atribut dengan ujung 1_2), dan pada saat follow-up kedua (atribut dengan ujung 1_3).



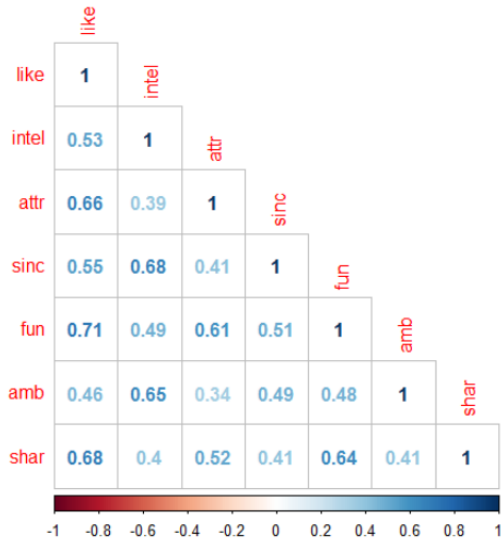
Data: data • Chart ID: LineChartIDe7015c62a71 • googleVis-0.5.10
R version 3.3.0 (2016-05-03) • Google Terms of Use • Documentation and Data Policy

Gambar 3. Line chart dari ketertarikan perempuan pada saat registrasi, saat event, saat follow-up pertama dan follow-up kedua.

- Analisis ketiga : Melihat korelasi nilai match dan like dengan 6 atribut yaitu intel, attr, sinc, fun, amb. Korelasi menggunakan fungsi `cor(...)`. Korelasi adalah suatu keterhubungan linear antara 2 variabel. Korelasi antara 2 variabel memiliki nilai

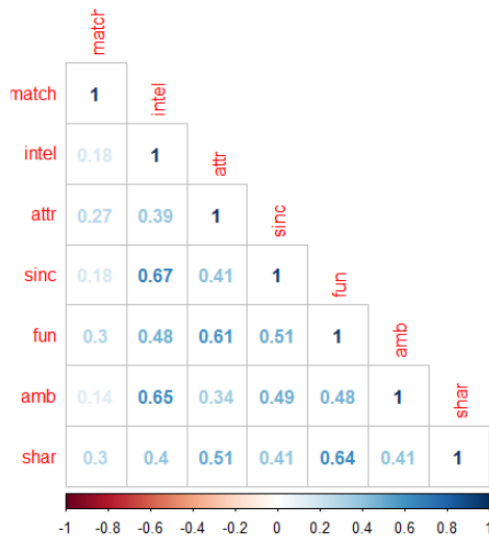
maksimal 1 dan nilai minimal -1. Nilai 1 menandakan bahwa korelasi kedua hubungan tersebut sangat kuat dan berbanding lurus, nilai 0 menandakan bahwa kedua variabel tersebut tidak terhubung sama sekali, dan nilai 1 menandakan bahwa kedua variabel tersebut terhubung namun secara terbalik.

Diagram Korelasi Penilaian 6 atribut oleh Perempuan Terhadap Pasangan Dengan Atribut 'like'



(a)

Diagram Korelasi Penilaian 6 atribut oleh Perempuan Terhadap Pasangan Dengan Atribut 'match'



(b)

Gambar 4. (a) Tabel korelasi penilaian 6 atribut dengan like. (b) Tabel korelasi penilaian 6 atribut dengan match.

Hasil Analisis dan Kesimpulan dari 1a

Atribut intel, attr, sinc, amb, fun, dan shar adalah atribut pada dataset yang menunjukkan penilaian terhadap pasangannya, dengan *range* 1 sampai 10.

Pada analisis pertama, faktor intelligence memang menjadi faktor paling besar pada dataset yang match, tetapi faktor tersebut juga paling besar pada dataset yang tidak match. Hal ini tidak dapat dijadikan bukti bahwa intelligence.

Pada analisis kedua, faktor intelligence juga tidak terlihat menjadi ketertarikan utama bagi perempuan secara stabil. Ketertarikan tersebut hanya ada pada awal saja (sebelum event), namun pada saat keberjalanannya, faktor intelligence tidak lagi menjadi faktor tertinggi.

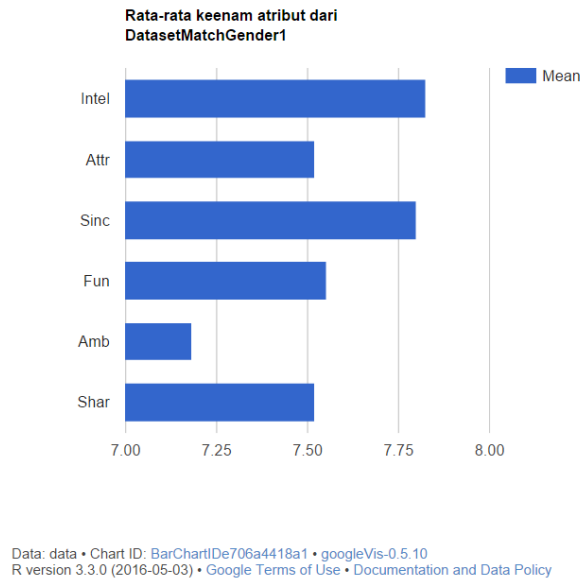
Pada analisis ketiga, kami melihat korelasi antara ketertarikan 6 penilaian wanita terhadap pria dengan nilai match dan nilai like. Dari tabel korelasi, dapat dilihat bahwa faktor intelligence juga bukan atribut yang memiliki nilai korelasi terbesar dengan match dan like.

Ketiga analisis di atas gagal untuk membuktikan bahwa intelegjensia adalah faktor dominan.

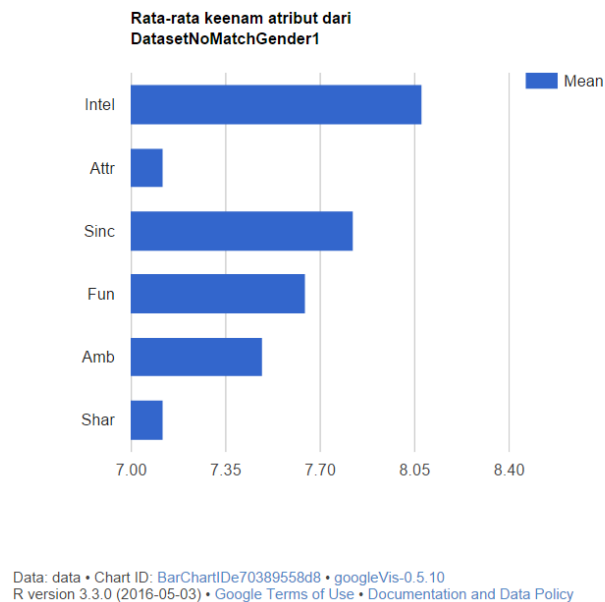
Kesimpulan:

Tidak terbukti bahwa ketertarikan wanita pada pria cenderung pada kemampuan intelegjensia (intelligence) dari calon partnernya.

- b. Membuktikan bahwa ketertarikan pria pada wanita cenderung pada kemenarikan (attractive) dari calon pertnernya.
- Analisis pertama : Melihat rata-rata dari 6 atribut yaitu intel, attr, sinc, fun, amb, shar dari subset data pria (gender = 1) yang match (match = 1), dan rata-rata dari subset data pria (gender = 1) yang tidak match (match = 0).



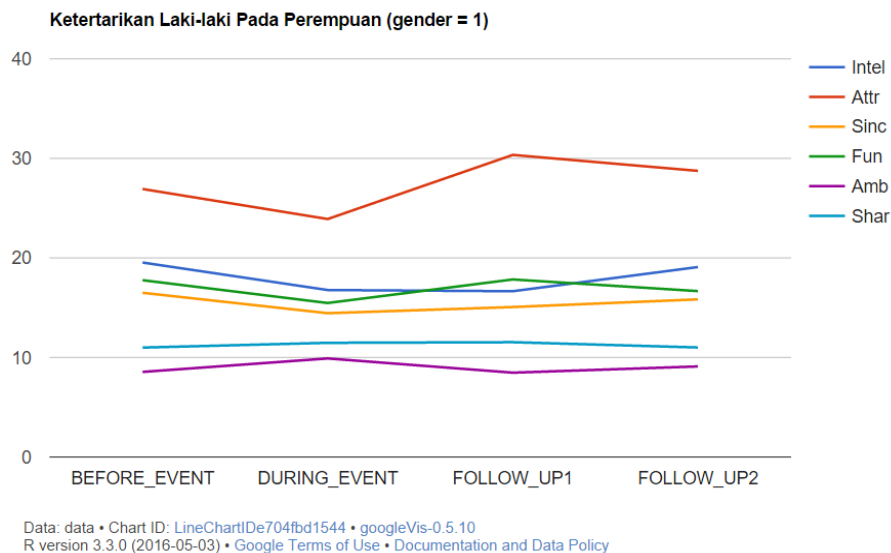
(a)



(b)

Gambar 4. (a) Tabel histogram penilaian 6 atribut dataset pria (penilaian wanita oleh pria) yang match. (b) Tabel histogram penilaian 6 terhadap dataset pria (penilaian wanita oleh pria) yang tidak match.

- Analisis kedua : Melihat rata-rata dari ketertarikan pria (gender = 1) pada saat registrasi (atribut dengan ujung 1_1), pada saat di tengah-tengah event (1_s), pada saat follow-up pertama (atribut dengan ujung 1_2), dan pada saat follow-up kedua (atribut dengan ujung 1_3).



Gambar 5. Line chart dari ketertarikan laki-laki pada saat registrasi, saat event, saat *follow-up* pertama dan *follow-up* kedua.

- Analisis ketiga : Melihat korelasi nilai match dan like dengan 6 atribut yaitu intel, attr, sinc, fun, amb. Korelasi menggunakan fungsi `cor(...)`. Korelasi adalah suatu keterhubungan linear antara 2 variabel. Korelasi antara 2 variabel memiliki nilai maksimal 1 dan nilai minimal -1. Nilai 1 menandakan bahwa korelasi kedua hubungan tersebut sangat kuat dan berbanding lurus, nilai 0 menandakan bahwa kedua variabel tersebut tidak terhubung sama sekali, dan nilai -1 menandakan bahwa kedua variabel tersebut terhubung namun secara terbalik.

Diagram Korelasi Penilaian 6 atribut oleh Laki-laki Terhadap Pasangan dengan Atribut 'like'

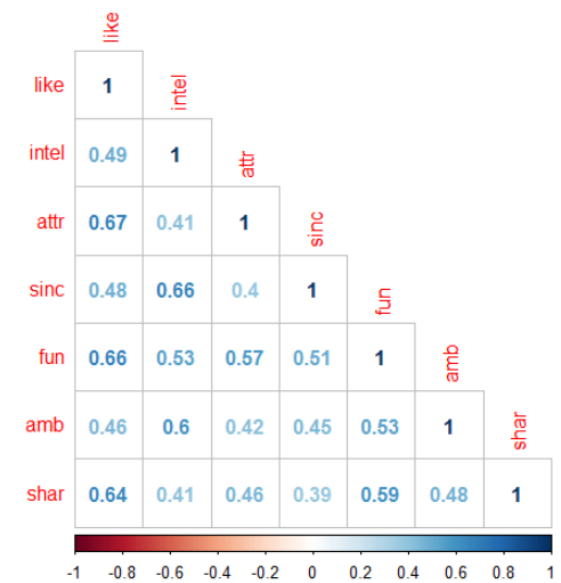
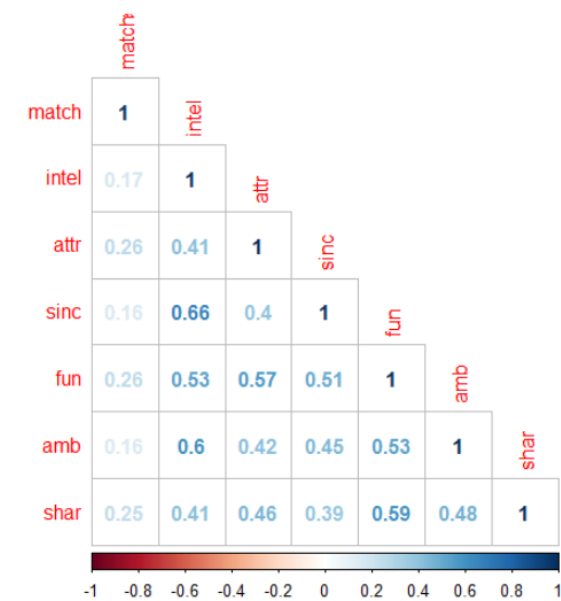


Diagram Korelasi Penilaian 6 atribut oleh Laki-laki Terhadap Pasangan dengan Atribut 'match'



Gambar 6. (a) Tabel korelasi penilaian 6 atribut dengan like. (b) Tabel korelasi penilaian 6 atribut dengan match.

Hasil Analisis dan Kesimpulan dari 1b

Atribut intel, attr, sinc, amb, fun, dan shar adalah atribut pada dataset yang menunjukkan penilaian terhadap pasangannya, dengan *range* 1 sampai 10.

Pada analisis pertama, faktor attractive memang tidak menjadi faktor paling besar, namun pada dataset yang tidak match, nilai attractive sangatlah rendah.

Pada analisis kedua, faktor attractive sangat terlihat memiliki bobot besar dari apa yang dicari oleh kaum laki-laki.

Pada analisis ketiga, kami melihat korelasi penilaian 6 atribut dari pasangan dengan nilai match dan like dengan enam atribut yang lain, dan memang nilai korelasi attractive dengan match dan like memiliki nilai yang paling besar.

Kesimpulan:

Terbukti bahwa ketertarikan pria pada wanita cenderung pada kemenarikan (attractive) dari calon partnernya.

2. Mendapatkan 1 (satu) faktor dominan lain yang membuat seorang pria dapat tertarik pada wanita dalam waktu singkat berdasarkan dataset tersebut, dan bagaimanawanita dapat tertarik ke pria (masing-masing satu).
 - a. Mendapatkan 1 (satu) faktor dominan lain yang membuat seorang wanita dapat tertarik pada pria dalam waktu singkat berdasarkan dataset tersebut.
 - Melihat analisis 1a yang ketiga, melihat dan menganalisis 2 tabel korelasi tersebut.

Hasil Analisis dan Kesimpulan dari 2a

Melihat tabel korelasi dari keenam atribut tersebut dengan match dan like, dapat dilihat bahwa faktor fun memiliki nilai korelasi terbesar dibandingkan keenam nilai lainnya. Dengan kata lain, dapat dikatakan bahwa faktor fun memiliki keterhubungan secara linear lebih baik dibandingkan 5 aspek lainnya.

Kesimpulan:

Satu faktor yang dominan yang menjadi ketertarikan wanita adalah faktor 'fun' atau sifat humoris.

- b. Mendapatkan 1 (satu) faktor dominan lain yang membuat seorang pria dapat tertarik pada wanita dalam waktu singkat berdasarkan dataset tersebut.
- Melihat analisis 1b yang ketiga, melihat dan menganalisis 2 tabel korelasi tersebut.

Hasil Analisis dan Kesimpulan dari 2a

Melihat tabel korelasi dari keenam atribut tersebut dengan match dan like, dapat dilihat bahwa faktor fun menjadi faktor terbesar kedua setelah attr

Kesimpulan:

Satu faktor yang dominan yang menjadi ketertarikan wanita adalah faktor 'fun' atau sifat humoris.
