

Analisis Dataset 'Take Me Out Indonesia'

Dharma Kurnia Septialoka (13514028), Joshua Atmadja (13514098)

Penjelasan Dataset

Dataset ini ialah kumpulan data dari sebuah event Speed Dating (atau dikenal di Indonesia sebagai Take Me Out). Data tersebut berisi tentang data peserta event tersebut dengan beberapa atribut yang substansial dan fundamental, seperti id peserta, id pasangan, penilaian pasangan terhadap 6 faktor ketertarikan (attractiveness, intelligence, sincerity, ambitiousness, fun, shared interests), kecocokan peserta dengan pasangannya, tingkatan suka peserta terhadap pasangannya, keputusan peserta untuk memilih seseorang di kontes tersebut, pernah tidaknya peserta menemui orang seperti itu, kemungkinan peserta berinteraksi dengan pasangan yang ditemuinya di kontes, kriteria pasangan ideal yang peserta cari, karir, pendapatan, ras, dan pendidikan peserta, pendapat peserta tentang penilaian orang lain terhadap dirinya, tujuan mengikuti acara, dan lainnya.

Eksplorasi 1a

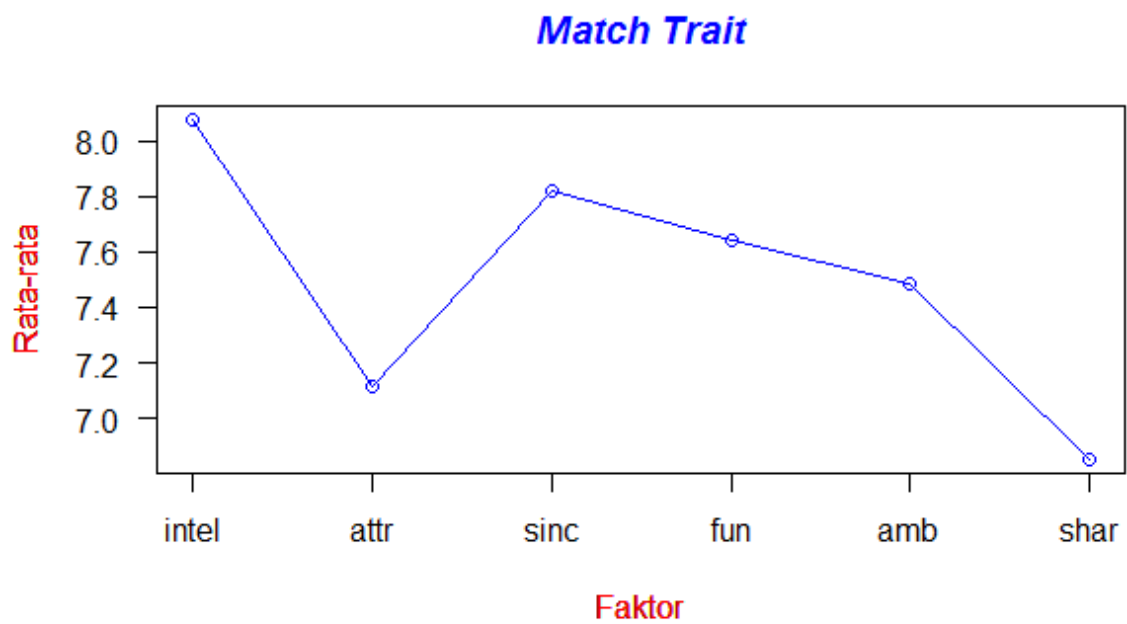
Eksplorasi pertama bagian pertama ialah membuktikan bahwa ketertarikan wanita pada pria cenderung pada kemampuan intelegensianya dari calon partnernya. Eksplorasi ini dilakukan dengan melakukan 3 analisis yaitu rata-rata penilaian faktor, perkembangan ketertarikan wanita kepada pria dari waktu ke waktu, dan korelasi antara kesesuaian dan kesukaan masing-masing dengan enam faktor yang ada.

Analisis Rata-rata Penilaian dari Enam Faktor

Analisis pertama ialah dengan menggunakan fungsi *subset*. Fungsi ini bertujuan untuk mengambil sebagian dari dataset sesuai dengan kriteria yang diinginkan. Ada 2 subset besar dari dataset yang diambil yaitu dataset dengan nilai atribut match = 1 (dataMatch) dan match = 0 (dataNoMatch). Dari dataMatch kemudian dipisahkan lagi menjadi dua subset yaitu dataMatch dengan nilai gender = 0 (dataFemale) dan dengan nilai gender = 1. Demikian juga untuk dataset mula-mula yang dipisahkan menjadi dataAllFemale dan dataAllMale; dataNoMatch dipisahkan menjadi dataNoFemale dan dataNoMale. Untuk eksplorasi bagian ini, yang dipakai adalah dataset bagian yang memiliki nilai atribut gender = 0, yaitu dataNoFemale, dataFemale, dataAllFemale.

Analisis ini mencari nilai rata-rata dari atribut-atribut yang menjadi faktor yaitu 'intel', 'attr', 'sinc', 'fun', 'amb', dan 'shar'. Mula-mula, nilai rata-rata dari enam atribut faktor tersebut dicari dari dataFemale. Secara berturut-turut, nilai rata-rata dari atribut tersebut ialah:

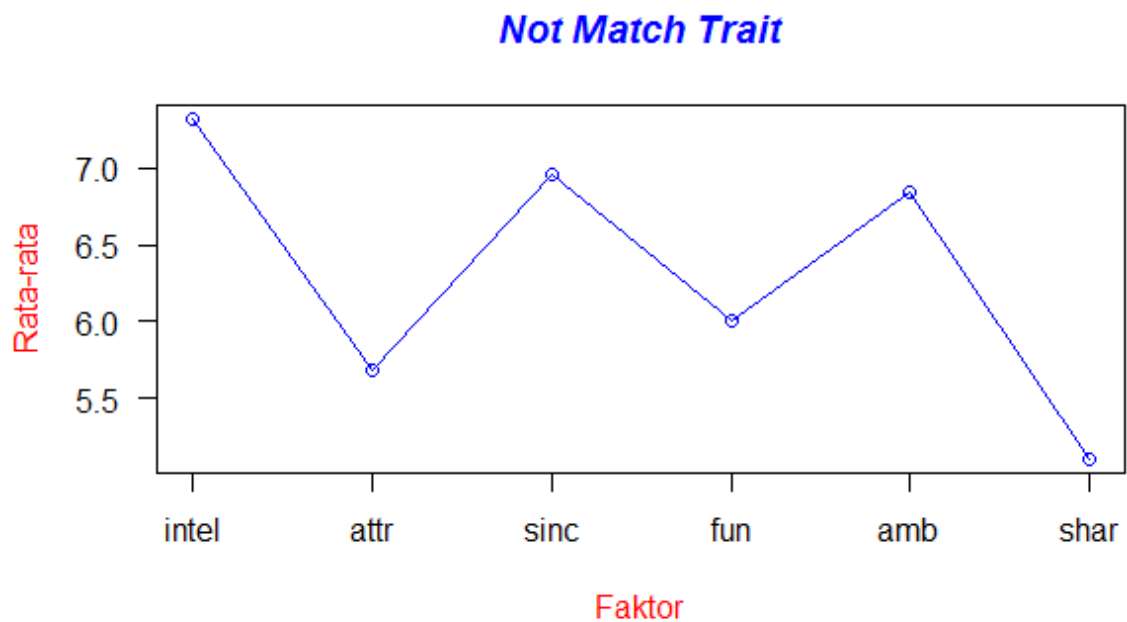
```
[1] 8.075073 7.118459 7.820700 7.644217 7.485271 6.854167
```



Gambar 1 Plot Garis Ketertarikan Perempuan terhadap Pasangannya dari dataFemale

Kemudian, nilai rata-rata dari enam atribut yang sama pada dataNoFemale ialah:

[1] 7.318860 5.676289 6.952939 5.999849 6.844665 5.104390



Gambar 2 Plot Garis Ketertarikan Perempuan terhadap Pasangannya dari dataNoFemale

Dari analisis cara pertama ini, sudah terbukti bahwa bawasannya memang baik wanita yang *match* maupun tidak cenderung tertarik pada pria dari intelejensianya.

Analisis Rata-rata Ketertarikan dari Waktu ke Waktu

Analisis kedua dilihat dari waktu ke waktu terhadap apa yang dicari rata-rata perempuan dari laki-laki mulai saat sign up (mendaftarkan diri), event berlangsung, follow up pertama, hingga follow up kedua. Atribut yang dianalisis pada saat ialah 'intel1_1', 'attr1_1', 'sinc1_1', 'fun1_1', 'amb1_1', dan 'shar1_1'. Untuk saat event berlangsung, sufiks atribut berubah dari 1_1 menjadi 1_s. Demikian juga saat follow up pertama, sufiks atribut berubah menjadi 1_2 serta saat follow up kedua, sufiks berubah menjadi 1_3. Terdapat total 24 atribut untuk dicari masing-masing rata-ratanya. Berikut ini adalah rata-rata ke-dua-puluh-empat atribut tersebut yang dikelompokkan pada enam vektor mulai dari 'intel', 'attr', 'sinc', 'fun', 'amb', dan 'shar' berturut-turut dengan sufiks '...All'.

```
[[1]]  
[1] 21.29801 18.73287 19.10053 19.61547
```

```
[[2]]  
[1] 17.76575 17.40720 21.68790 20.20858
```

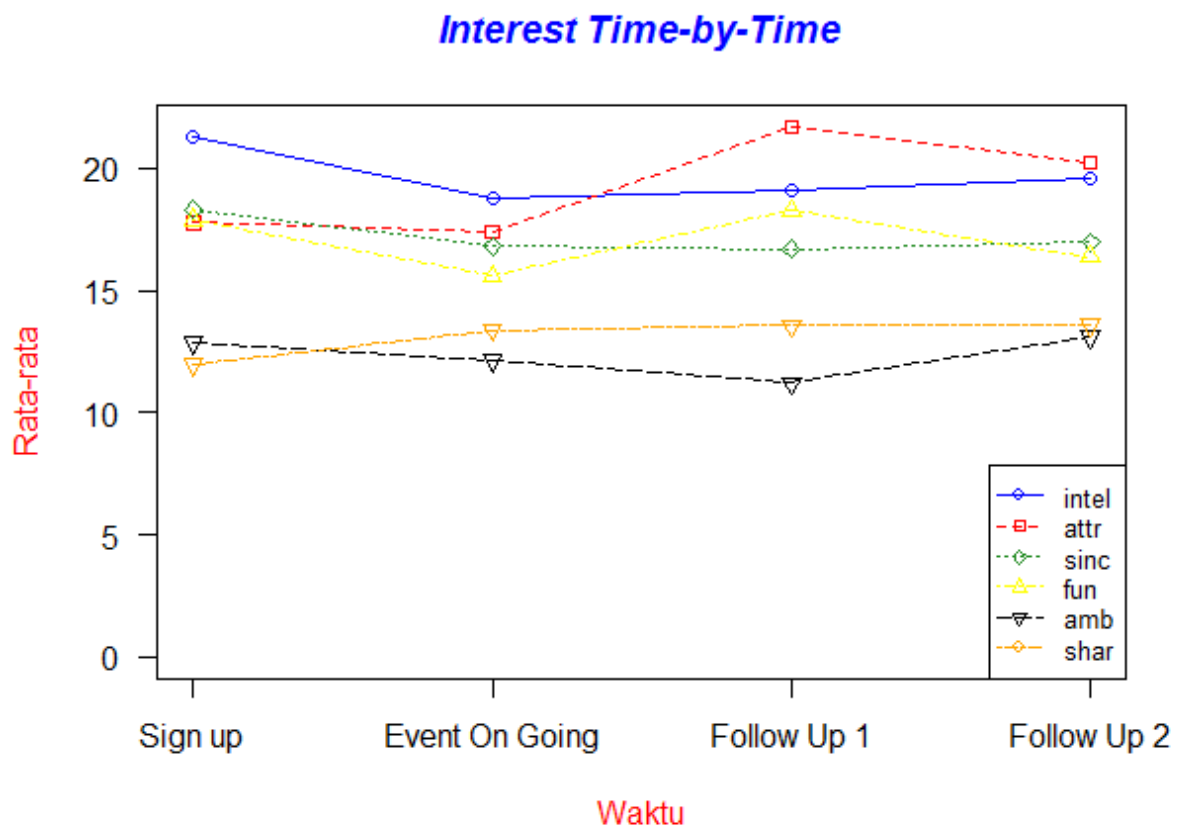
```
[[3]]  
[1] 18.31191 16.81470 16.69889 16.96268
```

```
[[4]]  
[1] 17.88524 15.58991 18.25232 16.36668
```

```
[[5]]  
[1] 12.85540 12.12460 11.20889 13.09922
```

```
[[6]]  
[1] 11.95756 13.36457 13.55671 13.57101
```

Urutan setiap vektor ialah rata-rata faktor saat sign up, saat event berlangsung, saat follow up pertama, dan saat follow up kedua. Rata-rata tersebut divisualisasi pada sebuah plot garis dimana X-axis-nya merupakan tahapan waktu.

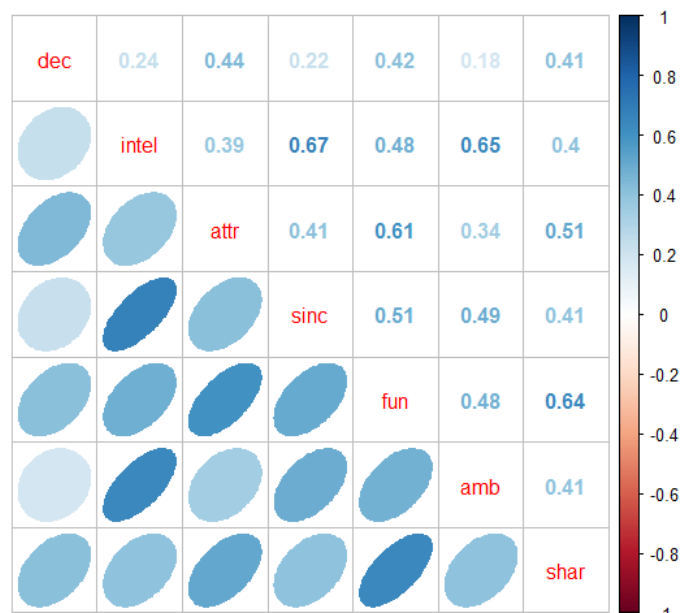


Gambar 3 Plot Garis Kriteria Laki-laki yang Dicari Perempuan dari Waktu ke Waktu

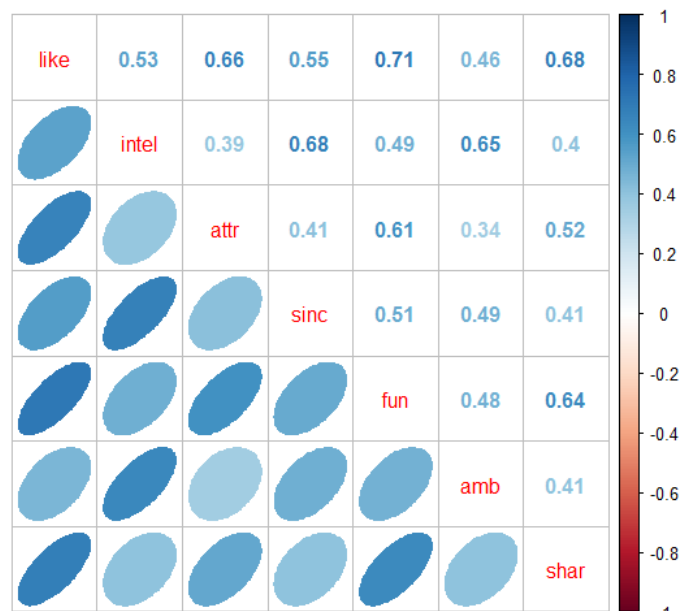
Terlihat dari analisis kedua bahwa rata-rata wanita memang pada awalnya cenderung mencari pria dengan melihat intelegensinya terlebih dahulu. Setelah mendapatkan pasangan yang sesuai – karena ini diambil dari dataFemale setelah di follow up– wanita cenderung mencari pria dari kemenarikannya. Terlihat bahwa saat event berlangsung, rata-rata wanita mulai mencari pria yang menarik dan tidak lagi yang intelegen.

Analisis Korelasi Enam Faktor

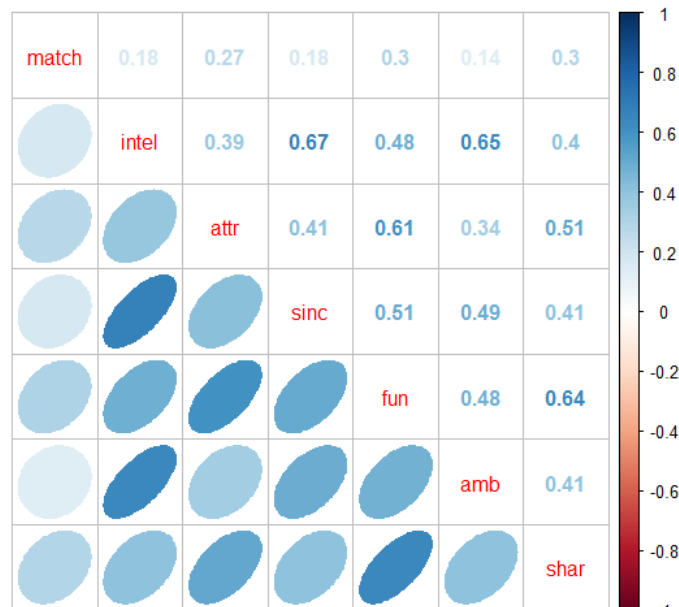
Analisis ketiga ialah dengan tabel korelasi enam faktor. Kali ini analisis dilakukan dengan menggunakan dataAllFemale. Terdapat tiga pengujian yaitu mencari korelasi 'dec', 'like', dan 'match'. Visualisasi grafis matriks korelasi ini dibantu dengan tools package yaitu *corrplot*. Berikut adalah visualisasi grafis dari tiga pengujian korelasi tersebut.



Gambar 4 Plot Korelasi Enam Faktor dengan atribut 'dec' dari dataAllFemale



Gambar 5 Plot Korelasi Enam Faktor dengan atribut 'like' dari dataAllFemale



Gambar 6 Plot Korelasi Enam Faktor dengan atribut 'match' dari dataAllFemale

Korelasi dengan atribut 'dec' yang menunjukkan apakah wanita ingin menyetujui untuk berkenan dengan pria paling tinggi berada pada nilai 0,44 (shared variance = 0,194) pada faktor kemenarikannya (attractiveness), sedangkan dengan atribut 'like' paling tinggi berada pada nilai 0,71 (shared variance = 0,504) pada faktor menyenangkan (fun) dan dengan atribut 'match' juga paling tinggi pada faktor menyenangkan pada nilai 0,3 (shared variance = 0,09). Semakin tinggi shared variance maka semakin tinggi juga korelasinya sehingga semakin dekat persebaran sampel data dan semakin banyak kemiripan sampel data. Terlihat bahwa korelasi dengan ketiga atribut pasti lebih dari 0 sebagai akibat dari nilai 'dec', 'like', dan 'match' yang biner (0 atau 1). Kecenderungan yang ada pada analisis ini ialah bahwa

1. wanita cenderung menyukai pria yang menyenangkan,
2. wanita cenderung memilih peserta kontes pria yang menarik, dan
3. wanita cenderung cocok dengan pria yang menyenangkan.

Kesimpulan

Terlihat bahwa tiga analisis di atas cenderung tidak mengerucut pada salah satu faktor. Akan tetapi, sebagian dari tiga analisis di atas membuktikan bahwa memang pada awalnya rata-rata wanita mencari pria dari intelegensinya, namun beberapa wanita justru cenderung menyukai pria yang menyenangkan dan cenderung memilih peserta kontes pria yang menarik baginya.

Eksplorasi 1b

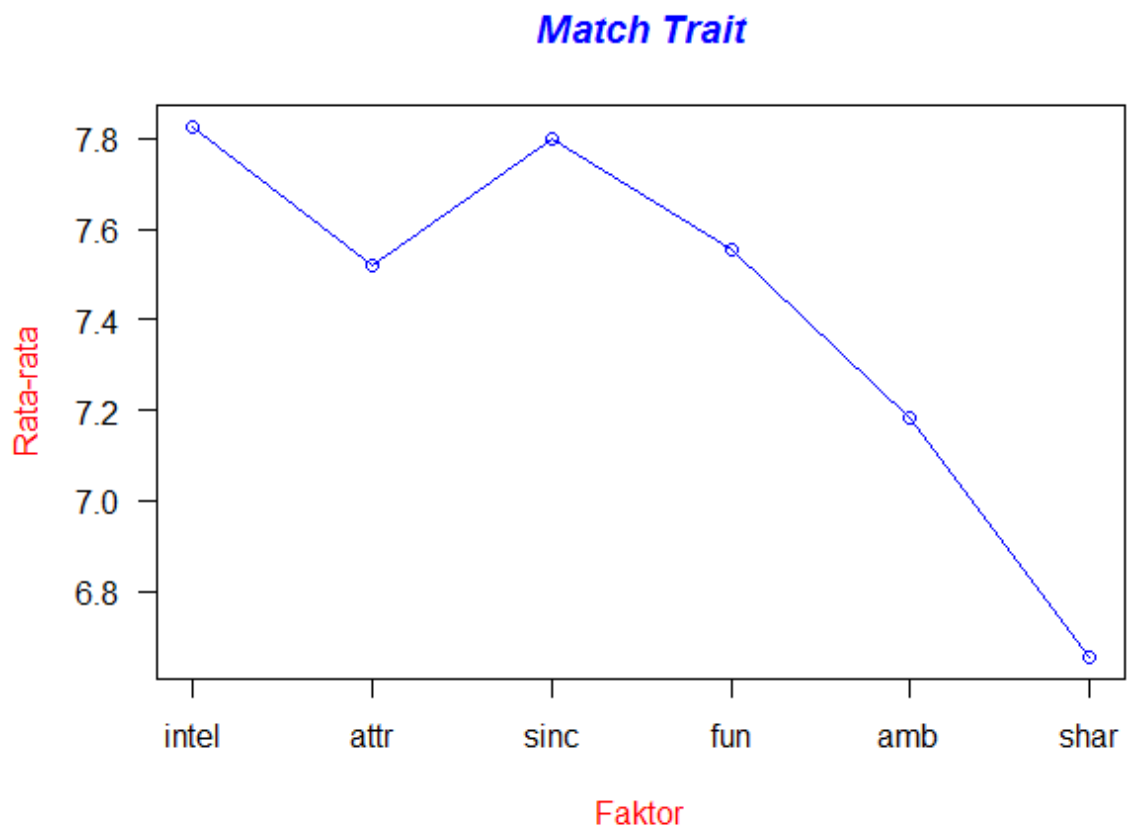
Eksplorasi pertama bagian kedua ialah membuktikan bahwa ketertarikan pria pada wanita cenderung pada kemenarikannya. Eksplorasi ini dilakukan sama seperti eksplorasi 1a.

Analisis Rata-rata Penilaian dari Enam Faktor

Analisis pertama untuk bagian ini akan memakai adalah dataset bagian yang memiliki nilai atribut gender = 1, yaitu dataNoMale, dataMale, dataMale.

Analisis ini mencari nilai rata-rata dari atribut-atribut yang menjadi faktor yaitu 'intel', 'attr', 'sinc', 'fun', 'amb', dan 'shar'. Mula-mula, nilai rata-rata dari enam atribut faktor tersebut dicari dari dataMale. Secara berturut-turut, nilai rata-rata dari atribut tersebut ialah:

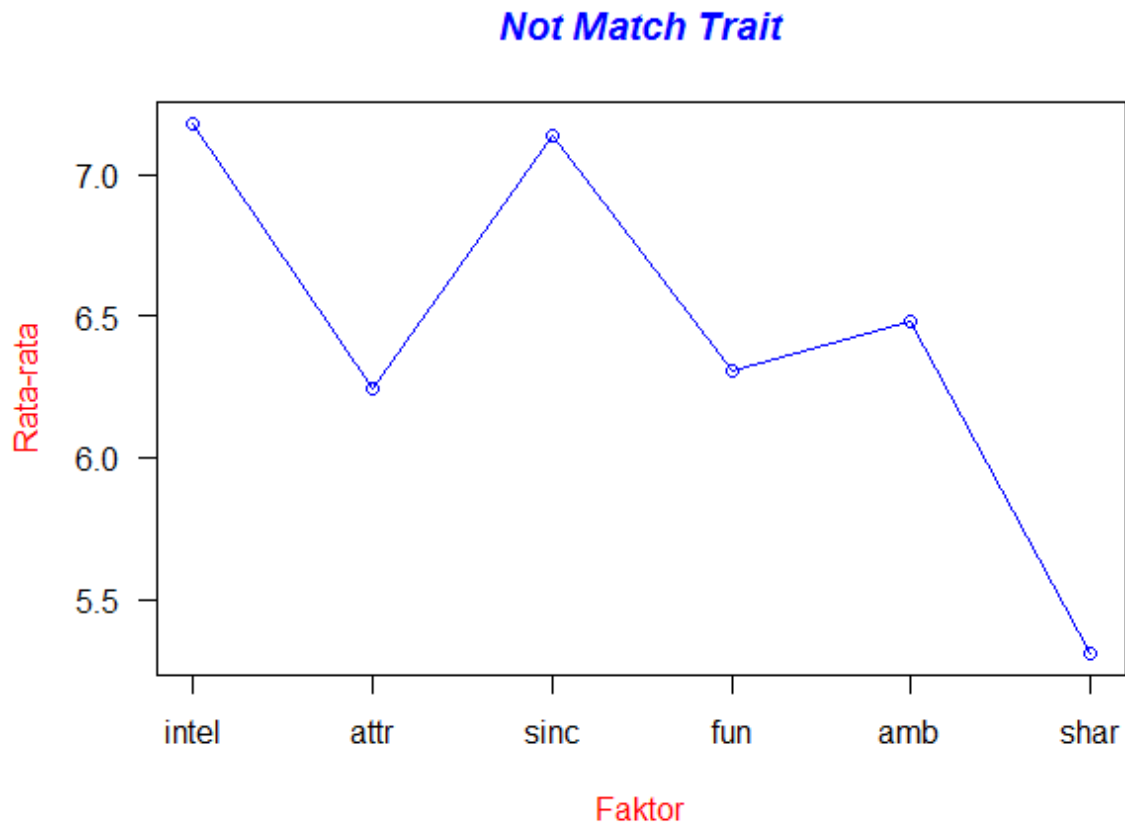
[1] 7.825292 7.520319 7.798246 7.552786 7.183409 6.654874



Gambar 7 Plot Garis Ketertarikan Laki-laki terhadap Pasangannya dari dataMale

Kemudian, nilai rata-rata dari enam atribut yang sama pada dataNoMale ialah:

```
[1] 7.181047 6.245417 7.139310 6.309268 6.483684 5.309260
```



Gambar 8 Plot Garis Ketertarikan Laki-laki terhadap Pasangannya dari dataNoMale

Dari analisis cara pertama ini, dapat disimpulkan bahwa rata-rata pria baik yang *match* maupun tidak cenderung tertarik pada wanita dari intelegensiannya.

Analisis Rata-rata Ketertarikan dari Waktu ke Waktu

Dengan menggunakan analisis yang sama dengan eksplorasi 1a, didapat total 24 atribut untuk dicari masing-masing rata-ratanya. Berikut ini adalah rata-rata ke-dua-puluh-empat atribut tersebut yang dikelompokkan pada enam vektor mulai dari 'intel', 'attr', 'sinc', 'fun', 'amb', dan 'shar' berturut-turut dengan sufiks '...All'.

```
[[1]]  
[1] 19.66988 16.69410 17.04939 18.86698
```

```
[[2]]  
[1] 28.07397 26.07268 30.18293 30.63091
```

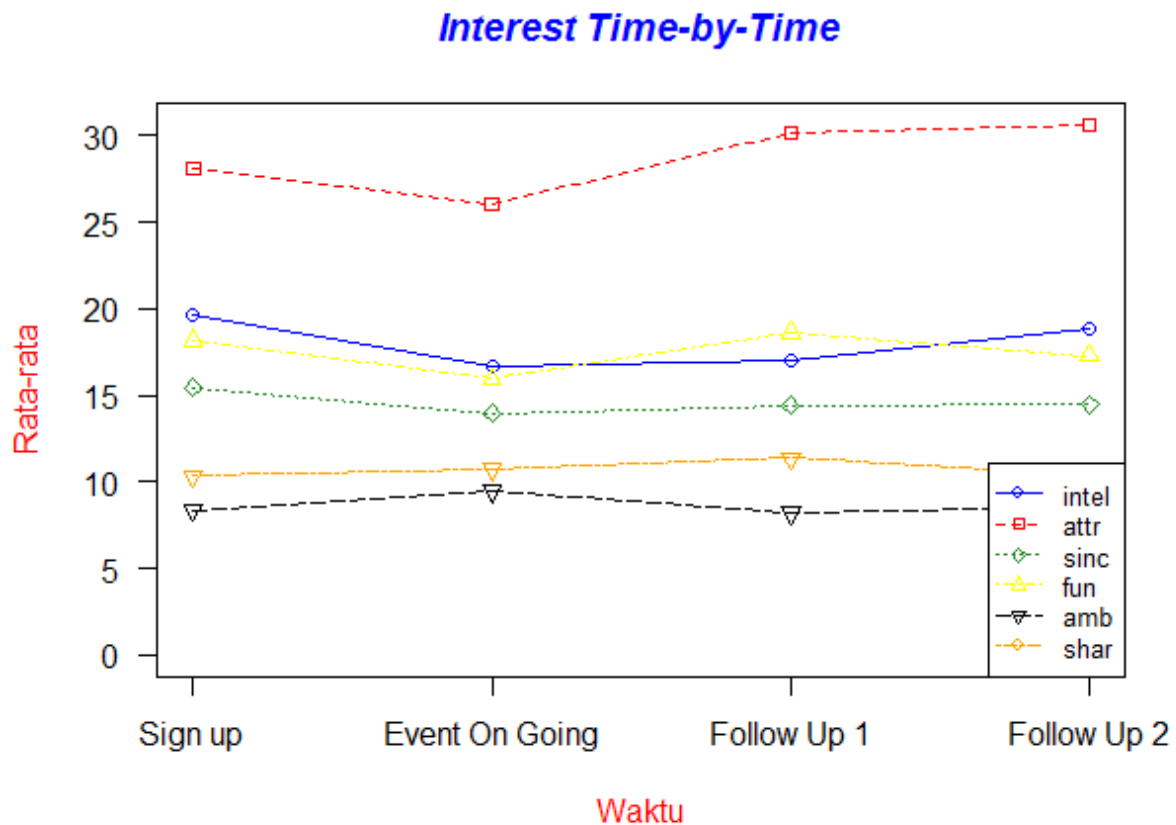
```
[[3]]  
[1] 15.46457 13.96988 14.39455 14.49821
```

```
[[4]]  
[1] 18.17088 15.94410 18.64699 17.29390
```

```
[[5]]  
[1] 8.386912 9.456519 8.201005 8.593929
```

```
[[6]]  
[1] 10.34821 10.70655 11.37550 10.44192
```


Urutan setiap vektor ialah rata-rata faktor saat sign up, saat event berlangsung, saat follow up pertama, dan saat follow up kedua. Rata-rata tersebut divisualisasi pada sebuah plot garis dimana X-axis-nya merupakan tahapan waktu.

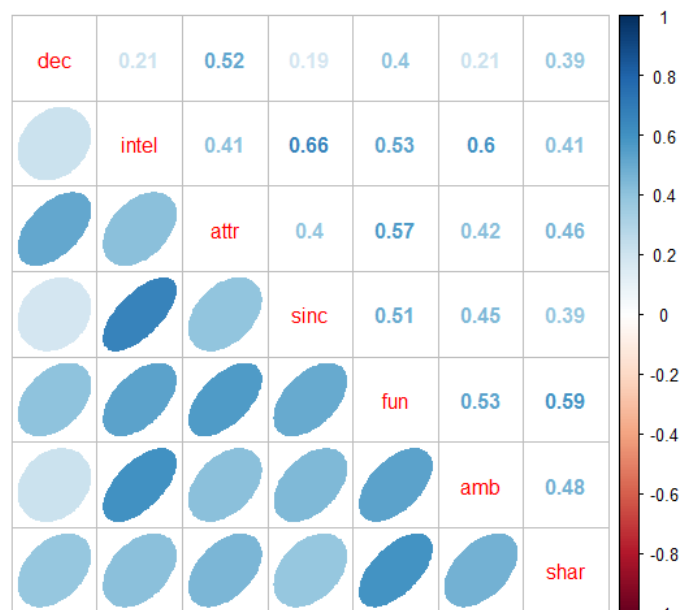


Gambar 9 Plot Garis Kriteria Perempuan yang Dicari Laki-laki dari Waktu ke Waktu

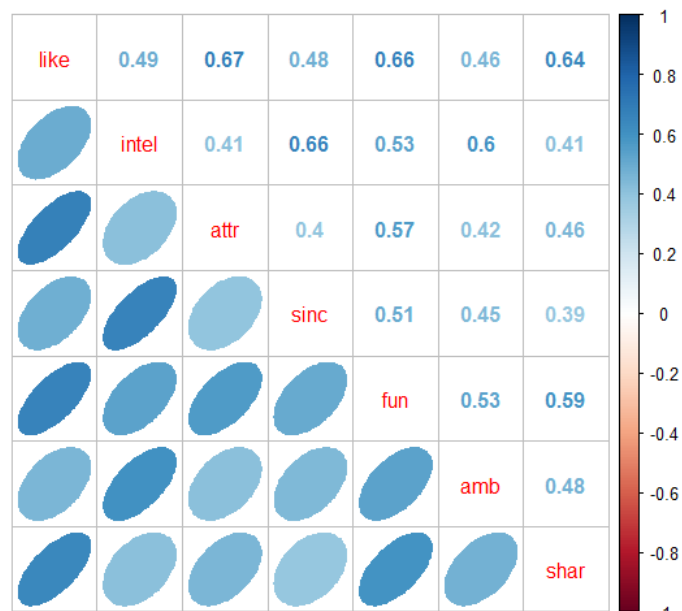
Terlihat dari analisis kedua bahwa rata-rata pria memang cenderung mencari wanita dengan melihat kemenarikannya. Terlihat bahwa sejak saat mendaftar hingga follow up kedua, rata-rata pria tetap mencari wanita yang menarik.

Analisis Korelasi Enam Faktor

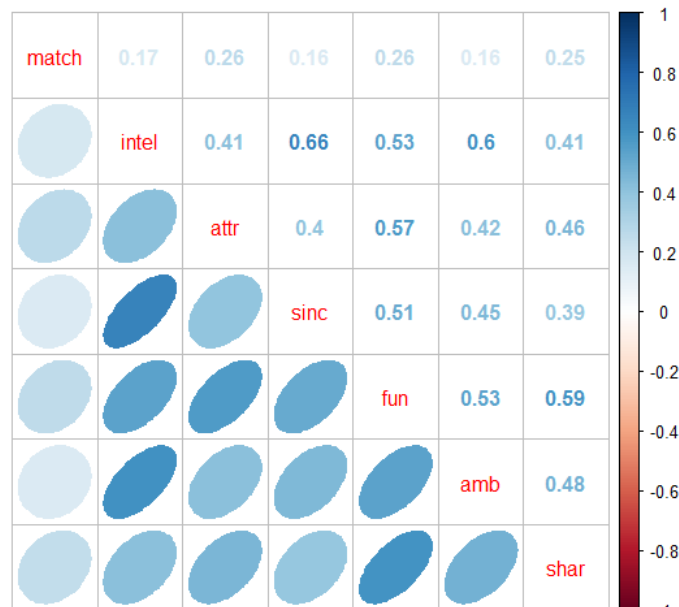
Masih menggunakan metode analisis dari eksplorasi 1a, kali ini analisis dilakukan dengan menggunakan dataAllMale. Terdapat tiga pengujian yaitu mencari korelasi 'dec', 'like', dan 'match'. Berikut adalah visualisasi grafis dari tiga pengujian korelasi tersebut.



Gambar 10 Plot Korelasi Enam Faktor dengan atribut 'dec' dari dataAllMale



Gambar 11 Plot Korelasi Enam Faktor dengan atribut 'like' dari dataAllMale



Gambar 12 Plot Korelasi Enam Faktor dengan atribut 'match' dari dataAllMale

Korelasi dengan atribut 'dec' yang menunjukkan apakah pria ingin menyetujui untuk berkencan dengan wanita paling tinggi berada pada nilai 0,52 (shared variance = 0,2704) pada faktor kemenarikannya (attractiveness). Demikian dengan atribut 'like' paling tinggi berada pada nilai 0,67 (shared variance = 0,4489) pada faktor kemenarikannya. Namun demikian, atribut 'match' memiliki dua faktor paling tinggi yaitu pada faktor menyenangkan dan faktor kemenarikannya yang bernilai 0,26 (shared variance = 0,0676). Kecenderungan yang ada pada analisis ini ialah bahwa

1. pria cenderung menyukai wanita dan memilih peserta kontes wanita dari faktor kemenarikannya, dan
2. pria cenderung sesuai dengan wanita yang unggul di faktor kemenarikan dan menyenangkan.

Kesimpulan

Terlihat bahwa tiga analisis di atas mengerucut pada satu faktor yaitu kemenarikannya meskipun analisis pertama gagal membuktikan hal tersebut. Meskipun demikian, sebagian analisis di atas membuktikan bahwa rata-rata pria mencari wanita dari faktor kemenarikannya.

Eksplorasi 2

Eksplorasi kedua ialah mendapatkan satu faktor dominan lain yang membuat seorang pria dapat tertarik pada wanita dalam waktu singkat berdasarkan dataset di atas, dan bagaimana wanita dapat tertarik ke pria. Eksplorasi ini menggunakan analisis dari eksplorasi 1a dan 1b dan terpusat pada analisis kedua dan ketiga.

Bagi wanita (pada eksplorasi 1a), seorang pria harus unggul dalam kemenarikannya. Namun dari analisis ketiga terlihat bahwa juga ada faktor 'fun' yang menyebabkan rata-rata wanita sesuai dengan pasangannya. Faktor 'fun' adalah faktor yang memiliki korelasi yang paling tinggi di korelasi enam faktor dengan atribut 'match', menyebabkannya memiliki keterhubungan linier lebih baik dibanding faktor lainnya. Analisis kedua juga membuktikan hal yang sama yakni dari saat event on going menuju follow up 1, faktor 'fun' menanjak naik melampaui faktor 'sinc' (kebaikan).

Sedangkan bagi pria (pada eksplorasi 1b), seorang wanita harus unggul dalam kemenarikannya. Namun dari analisis ketiga terlihat bahwa ada faktor 'fun' yang menyebabkan rata-rata pria sesuai dengan pasangannya. Faktor 'fun' adalah salah satu faktor yang memiliki korelasi yang paling tinggi di korelasi enam faktor dengan atribut 'match', menyebabkannya memiliki keterhubungan lanjut lebih baik dibanding faktor lainnya. Analisis kedua juga membuktikan demikian yakni dari saat event on going menuju follow up 1, faktor 'fun' meningkat naik melampaui faktor 'intel'.

Kesimpulan

Faktor dominan lain yang menjadi ketertarikan wanita kepada pria ialah faktor 'fun' atau faktor menyenangkan. Demikian juga dengan faktor dominan lain yang menjadi ketertarikan pria kepada wanita, yaitu faktor 'fun'.