

EDA DENGAN DATA NYATA

Mata Kuliah Data Science

Dosen Pengampu : Vynska Amalia Permadi,S.Kom.,M.Kom.



Disusun oleh :

NAMA : ALIVI MILOVA

NIM : 123170062

KELAS : C

PROGRAM STUDI INFORMATIKA

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK INDUSTRI

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”

YOGYAKARTA

2020

ANALISIS DATA OBESITAS

Data yang diperoleh merupakan data dimana menjadi dasar-dasar untuk pengukuran atau penilaian suatu kondisi yang menyatakan seseorang itu tergolong obesitas atau tidak. Dan data berikut didapatkan dari salah satu website untuk mencari dataset yaitu Kaggle.

Data yang diperoleh terdapat 53392 baris dengan jumlah 33 variable yang ada didalamnya. Dan analisis data akan dilakukan dengan RStudio, setelah melakukan berbagai tahapan dan menginput data pada RStudio maka bisa dilakukan pencarian data seperti yang akan disajikan pada gambar-gambar dibawah ini. Kemudian dengan analisis data akan didapatkan hasil prediksi atau perkiraan untuk mengetahui siapa yang memiliki tingkat resiko obesitas tertinggi berdasarkan dari data yang sudah ada. Lihat dan perhatikan gambar-gambar di bawah ini



Gambar 1. Variable data berat

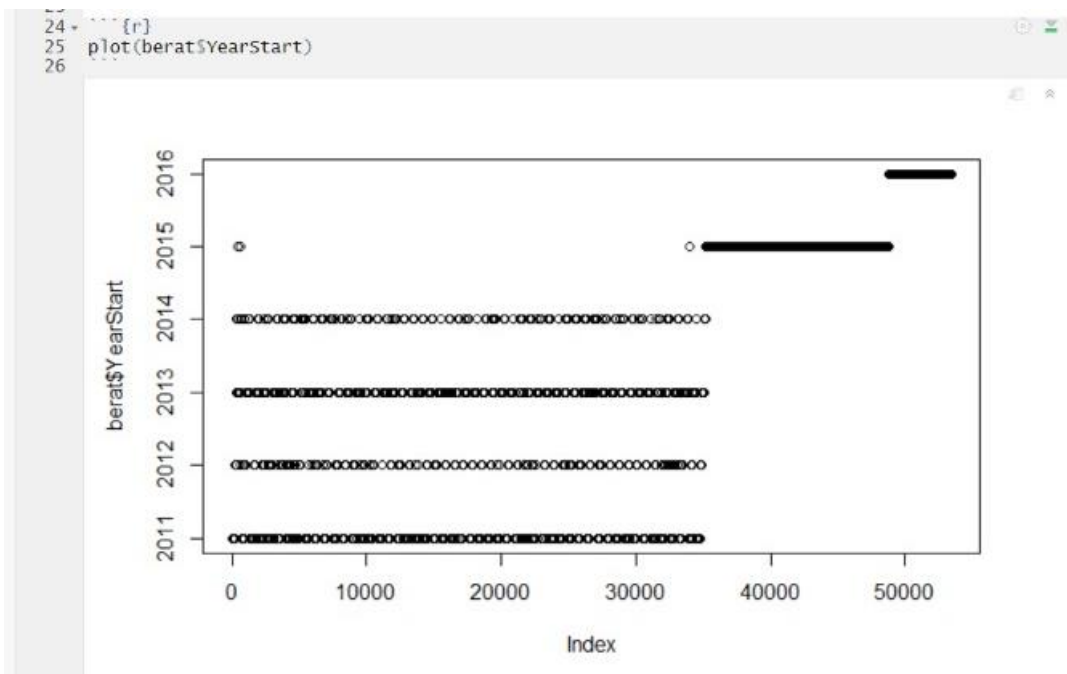
Pada Gambar 1. Variable data berat ditampilkan daftar nama variable yang ada pada data yang bernama berat. Dengan menuliskan kode pada R maka akan muncul nama variable yang ada pada data. kode pada R yang dimaksud adalah *names(nama_file)*

Setelah itu perhitungan akan bisa dilihat dengan menuliskan baris kode pada R yaitu *summary(nama_file)* maka hasilnya bisa dilihat seperti pada Gambar 2. Summary data berat, ditunjukkan hasil terkait type data yang ada pada data, jika data berupa string atau karakter maka tidak bisa menggunakan perhitungan summary, dan data akan otomatis terhitung jika data berupa angka, perhitungan mulai dari data minimal, maksimal, mean, median dan sebagainya. Untuk hasilnya bisa dilihat pada gambar yang mana menunjukkan bahwa tidak semua data bisa dihitung, lihat pada gambar dibawah ini

Topic	Question	Data_Value_Unit	Data_Value_Type	Data_Value
Data_Value_Alt				
Length:53392	Length:53392	Mode:logical	Length:53392	Min. : 0.90
Min. : 0.90				
Class :character	Class :character	NA's:53392	Class :character	1st Qu.:24.10
1st Qu.:24.10				
Mode :character	Mode :character		Mode :character	Median :30.70
Median :30.70				
				Mean :31.16
Mean :31.16				
				3rd Qu.:37.00
3rd Qu.:37.00				
				Max. :77.60
Max. :77.60				
				NA's :5046
NA's :5046				
Data_Value_Footnote_Symbol	Data_Value_Footnote	Low_Confidence_Limit	High_Confidence_Limit	
Sample_Size				
Length:53392	Length:53392	Min. : 0.30	Min. : 3.00	
Min. : 50				
Class :character	Class :character	1st Qu.:20.00	1st Qu.:28.20	
1st Qu.: 566				
Mode :character	Mode :character	Median :26.45	Median :35.60	
Median : 1209				
		Mean :26.89	Mean :35.99	
Mean : 3889				
		3rd Qu.:32.90	3rd Qu.:42.20	
3rd Qu.: 2519				
		Max. :69.50	Max. :87.70	
Max. :476876				
		NA's :5046	NA's :5046	
NA's :5046				

Gambar 2. Summary data berat

Setelah itu lihat data untuk berapa jumlah data pervariable berdasarkan data yang ada, lihat gambar berikut

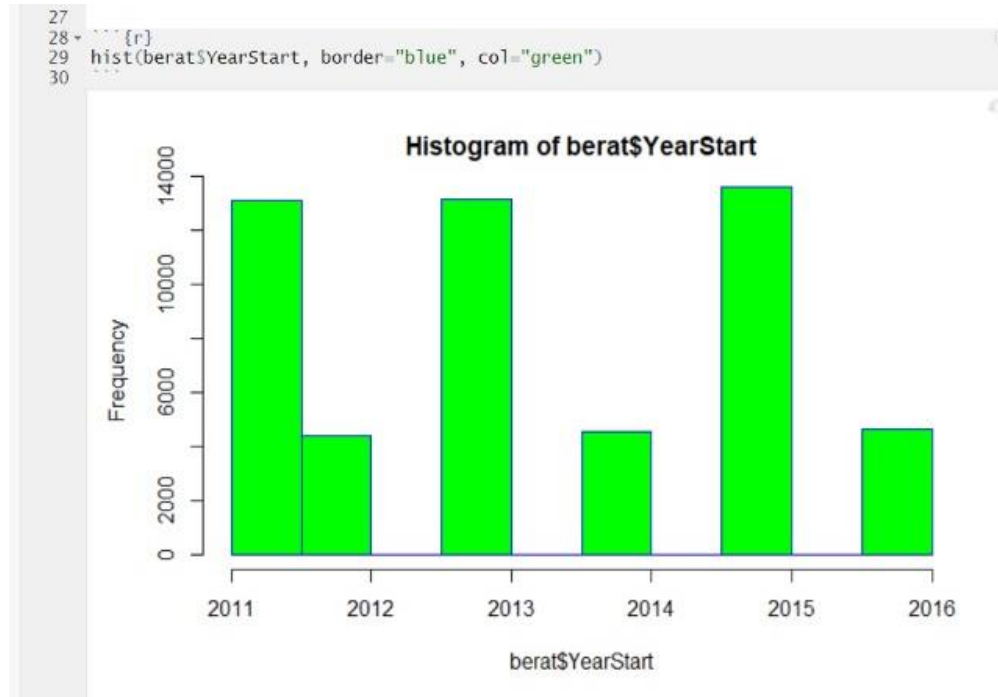


Gambar 3. Tahun Mulai

Berdasarkan Gambar 3. Tahun Mulai maka terlihat dari berapa jumlah orang yang terkena obesitas dilihat dari tahunnya. Dan bisa dilihat bahwa pada tahun 2011 hingga 2014

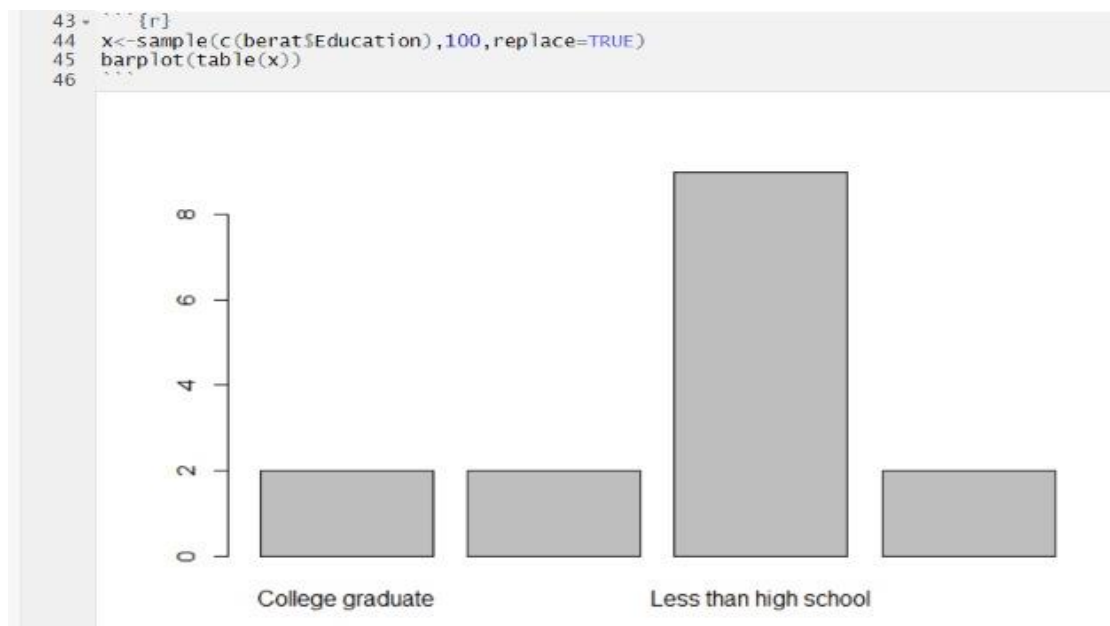
rata-rata berjumlah sama yaitu antara 0 sampai kurang dari 4000 tetapi pada tahun 2015 dan 2016 mengalami perubahan hingga tahun 2016 mencapai 5000.

Selanjutnya lihat gambar dibawah ini



Gambar 4.Histogram Tahun Mulai

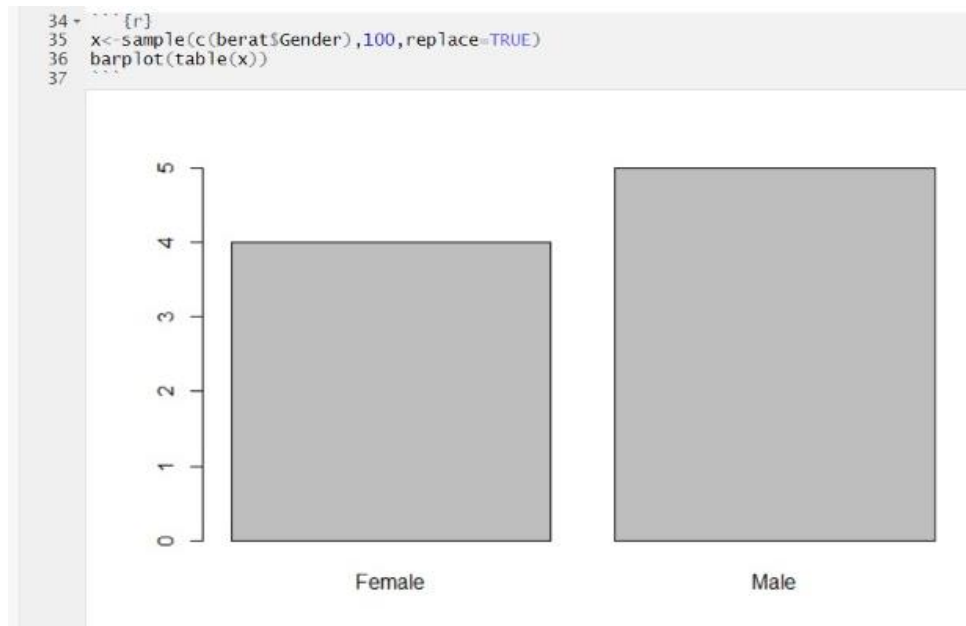
Dari variable yang sama yaitu YearStart di bandingkan dengan Gambar 3 maka Gambar 4 terlihat lebih spesifik dengan angka rentang angka yang lebih jelas. Kemudian lihat gambar selanjutnya



Gambar 5. Education

Gambar 5. Education merupakan hasil dari plot salah satu variable yaitu Pendidikan sehingga bisa dilihat bahwa angka tertinggi yang menempati posisi obesitas jika dibandingkan dengan College graduate dan pendidikan lainnya maka Less than high school menduduki peringkat tertinggi.

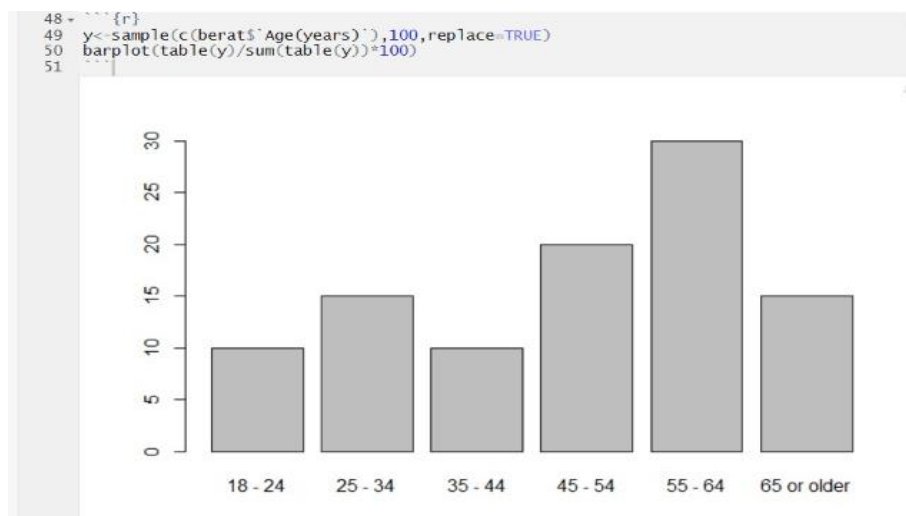
Selanjutnya bisa dilihat untuk gambar plot lainnya yaitu diambil berdasarkan variable gender atau jenis kelamin, maka hasil yang didapatkan adalah sebagai berikut



Gambar 6. Gender

Dari Gambar 6. Gender menunjukkan bahwa yang terkena obesitas antara perempuan tidak jauh berbeda tetapi untuk laki-laki lebih banyak dibandingkan dengan perempuan.

Kemudian Lihat untuk data umur yang terdapat pada data dan hasilnya terlihat seperti gambar dibawah ini



Gambar 7. Umur

Berdasarkan Gambar 7. Umur maka terlihat berapa jumlah orang yang terkena obesitas dan angka menunjukkan bahwa yang paling banyak terkena obesitas adalah orang dengan umur rentang antara 55 hingga 64 tahun dan terendah adalah dengan umur 33 sampai 44 tahun.

Dengan demikian terdapat 4 variable yang menjadi perbandingan yaitu tahun, pendidikan, jenis kelamin dan umur, sehingga bisa didapatkan kesimpulan berdasarkan data tersebut adalah perkiraan terkena obesitas tertinggi adalah laki-laki dengan umur rentang antara 55 sampai dengan 64 tahun kemudian jika dibandingkan dengan variable lain maka kemungkinan tertinggi yang terkena obesitas adalah laki-laki yang mana pendidikannya adalah less than high school.