

Penilaian Risiko dalam Asuransi Kesehatan

Probability Course - Sekolah Data Pacmann

Azka Nur Afifah

Outline

- Pendahuluan
- Dataset
- Analisis Statistik Deskriptif
- Analisis Variabel Kategorik
- Analisis Variabel Kontinu
- Korelasi Variabel
- Uji Hipotesis
- Kesimpulan

Pendahuluan

Pendahuluan

- Dalam bisnis asuransi sangatlah penting untuk menilai risiko yang dibawa oleh pengguna asuransi agar perusahaan tetap bisa mendapatkan untung dari perlindungan yang diberikan
- Penilaian risiko dilakukan untuk menentukan besar premi atau tagihan yang harus dibayar pengguna tiap bulannya
- Dalam project ini kita ingin mengetahui variabel yang mempengaruhi besarnya tagihan pengguna

Dataset

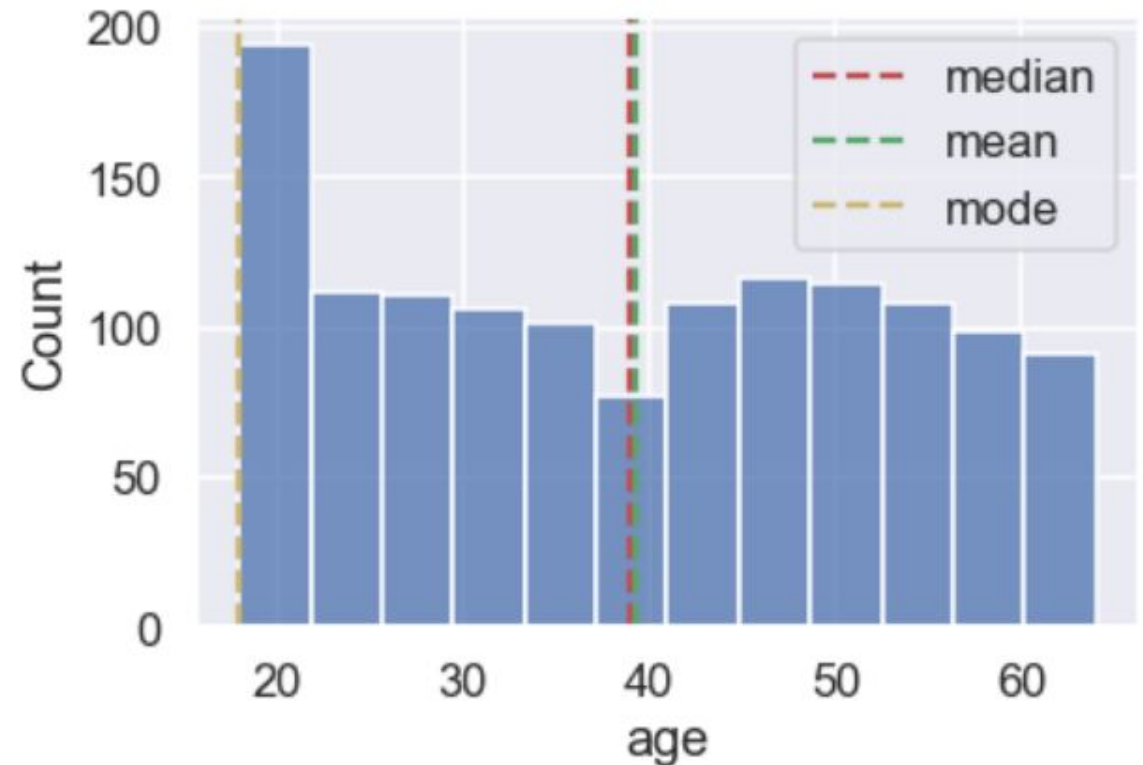
Dataset

- Dataset ini terdiri dari beberapa fitur
 - Umur
 - Jenis kelamin
 - BMI
 - Jumlah anak
 - Perokok/bukan
 - Region
 - Tagihan

Analisis Statistik Deskriptif

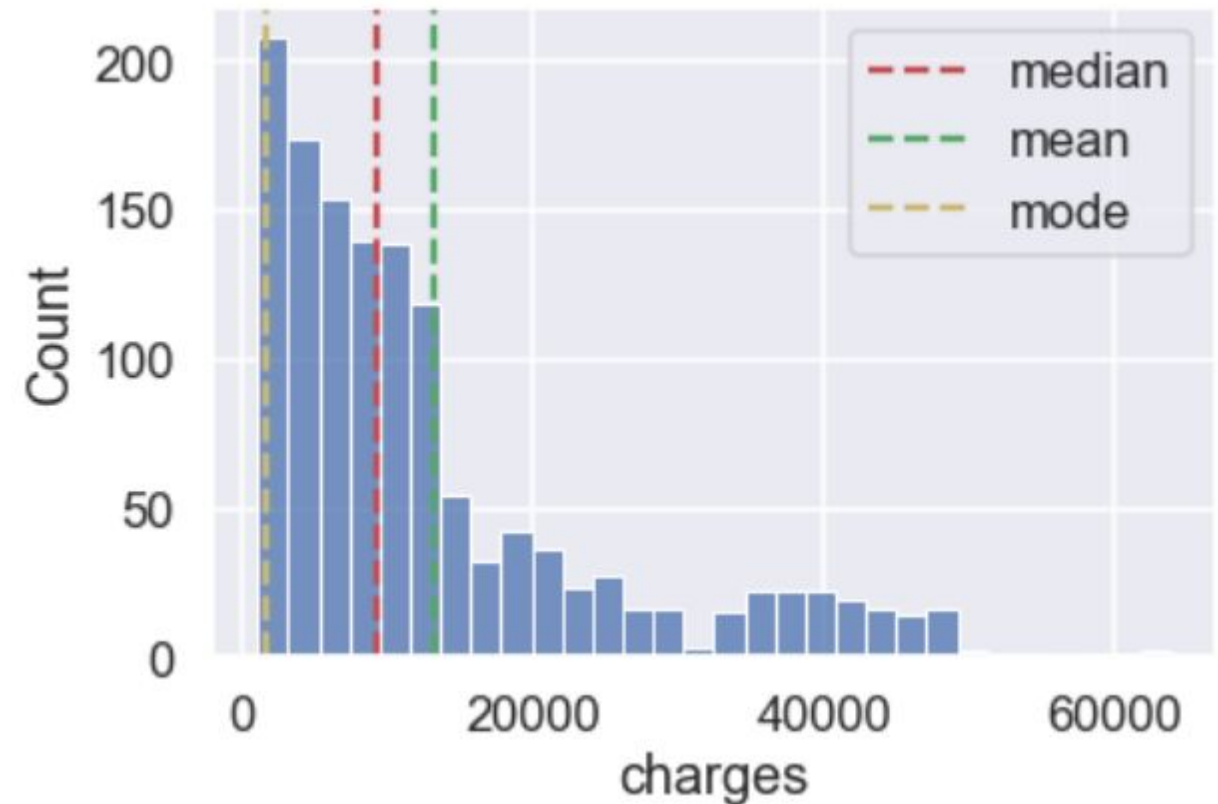
Nilai Tengah Usia

- Usia pengguna memiliki rata-rata dan median yang berdekatan yaitu masing-masing sebesar 39.2 dan 39
- Dengan membuat histogram dapat dilihat bahwa bentuk distribusi usia cenderung seragam kecuali di ujung kiri yaitu di usia 18 dan 19 tahun yang masing-masing memiliki frekuensi 69 dan 68



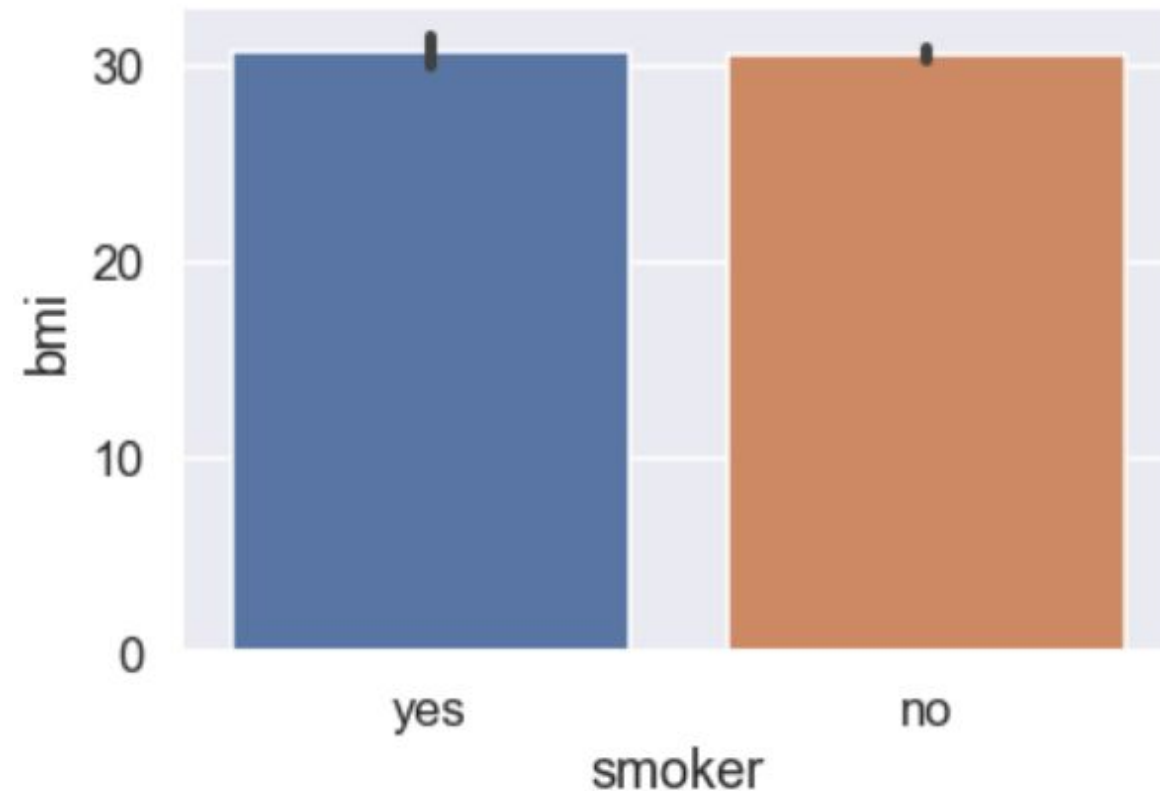
Nilai Tengah Tagihan

- Distribusi nilai tagihan memiliki ekor yang panjang di kanan
- Rata-rata tagihan adalah sebesar 13270, lebih besar dari nilai mediannya yaitu 9382
- Nilai modus tagihan jauh lebih kecil dibanding rata-rata dan median yaitu sebesar 1639



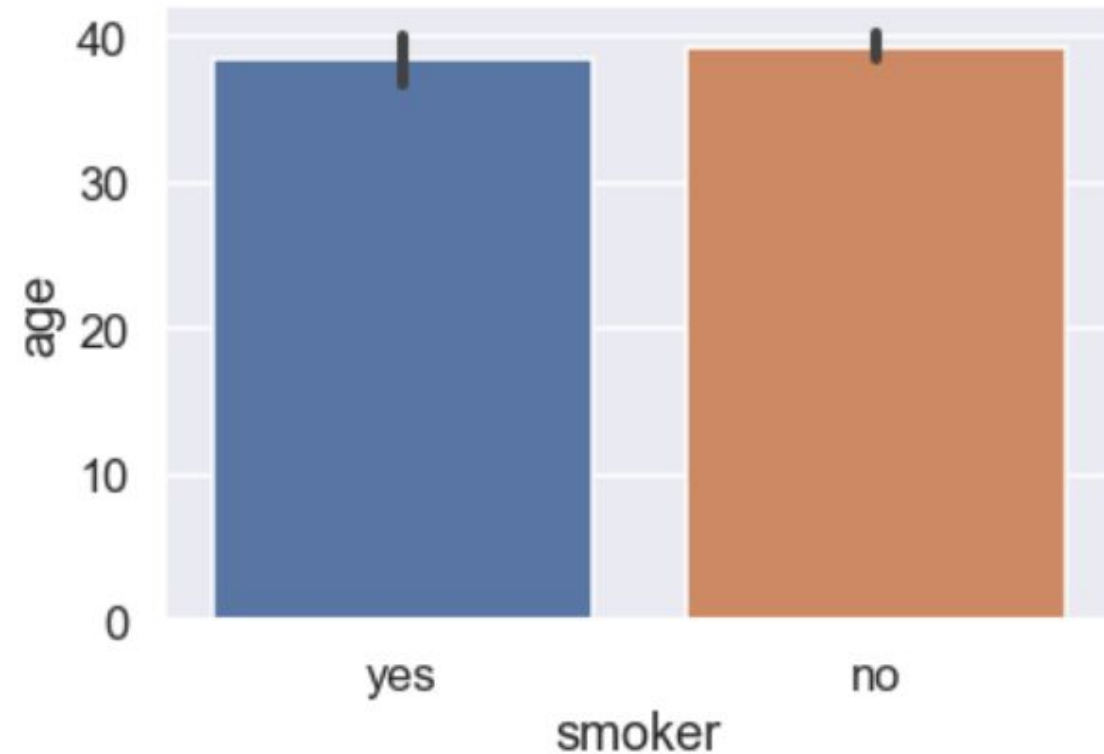
BMI Perokok dan Non-Perokok

- Pengguna perokok dan non-perokok memiliki **BMI yang cenderung sama** dengan BMI rata-rata seluruh pengguna, yaitu sebesar 30.66
- Perokok memiliki rata-rata BMI sebesar 30.71 sedangkan non-perokok memiliki rata-rata BMI sebesar 30.65



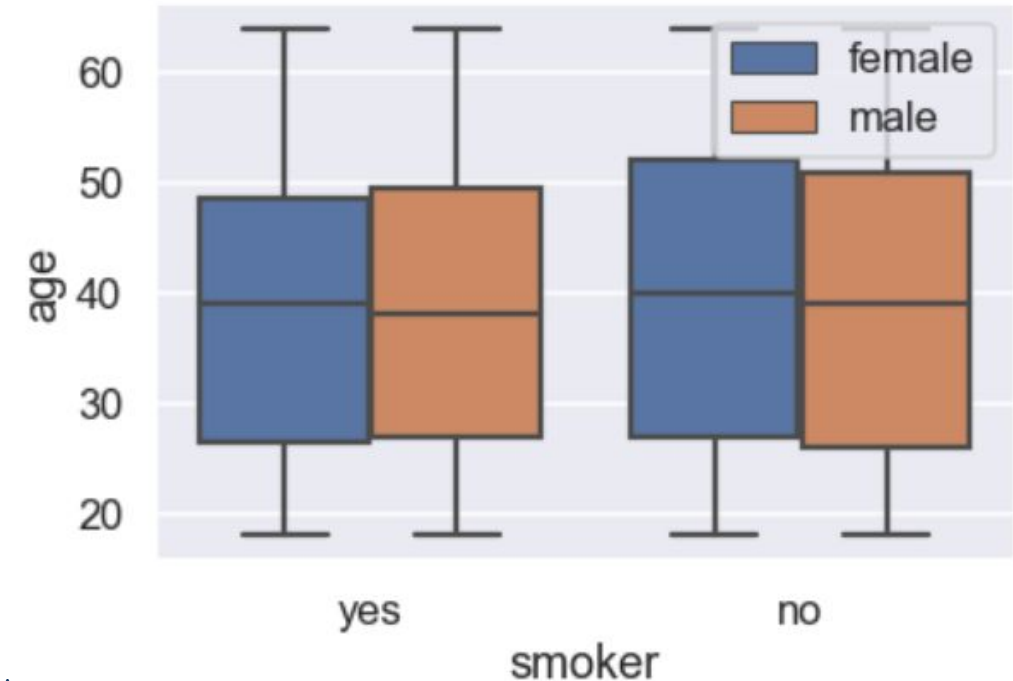
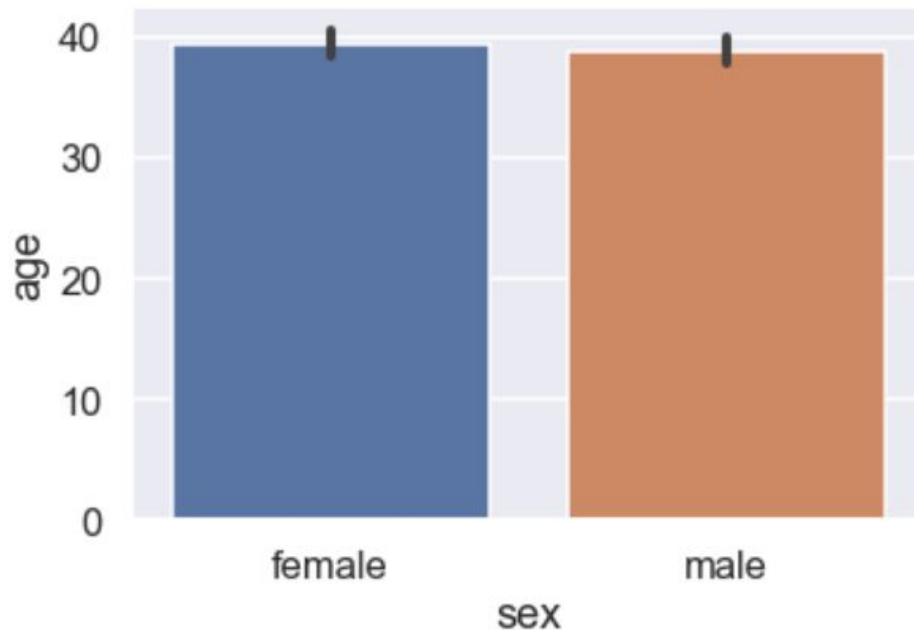
Usia Perokok dan Non-Perokok

- Pengguna perokok dan non-perokok memiliki rata-rata usia yang tidak jauh berbeda
- Rata-rata usia perokok adalah 38.51 sedangkan non-perokok 39.39



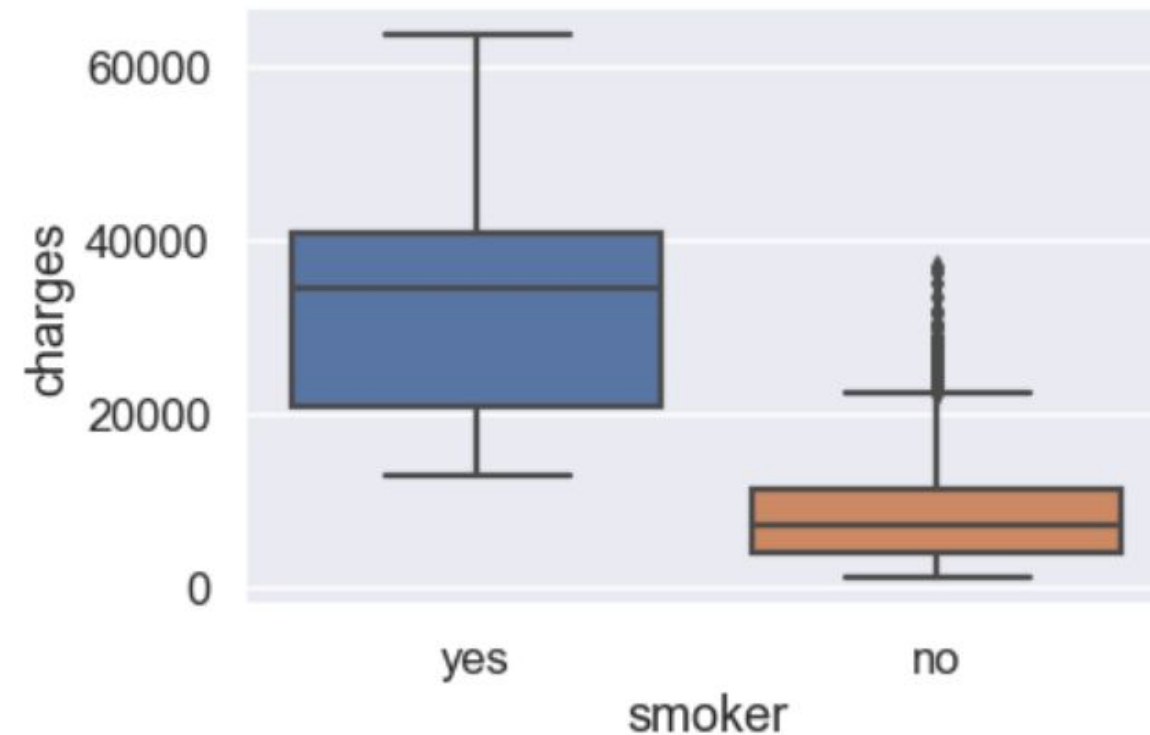
Usia Laki-Laki dan Perempuan

- Pengguna laki-laki dan perempuan memiliki rata-rata usia yang relatif sama, yaitu masing-masing 38.4 dan 38.6
- Di antara pengguna perokok dan non-perokok, masing-masing jenis kelamin memiliki rata-rata umur yang tidak berbeda jauh dengan sebaran yang relatif sama besar



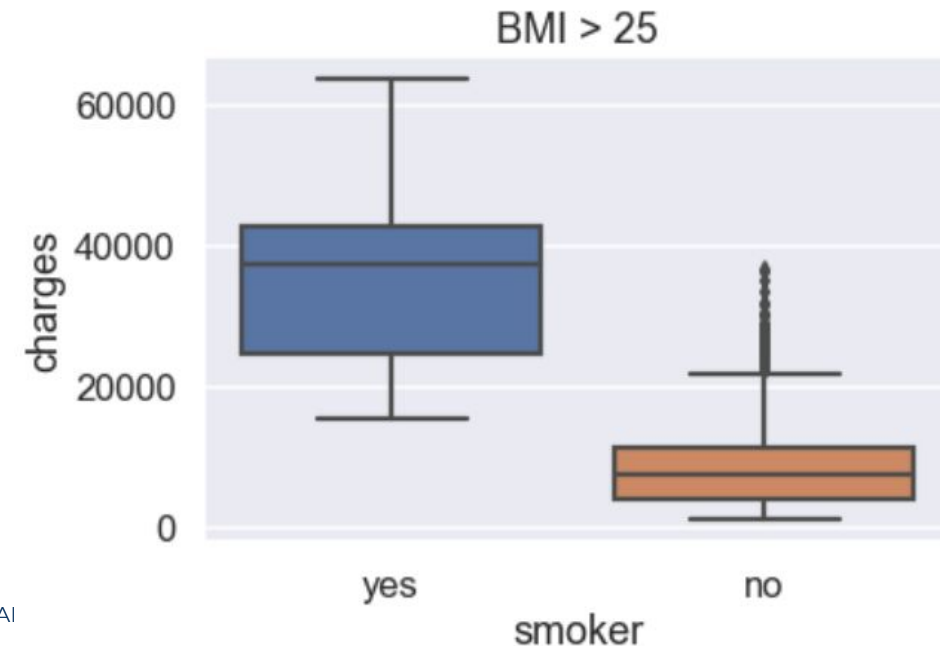
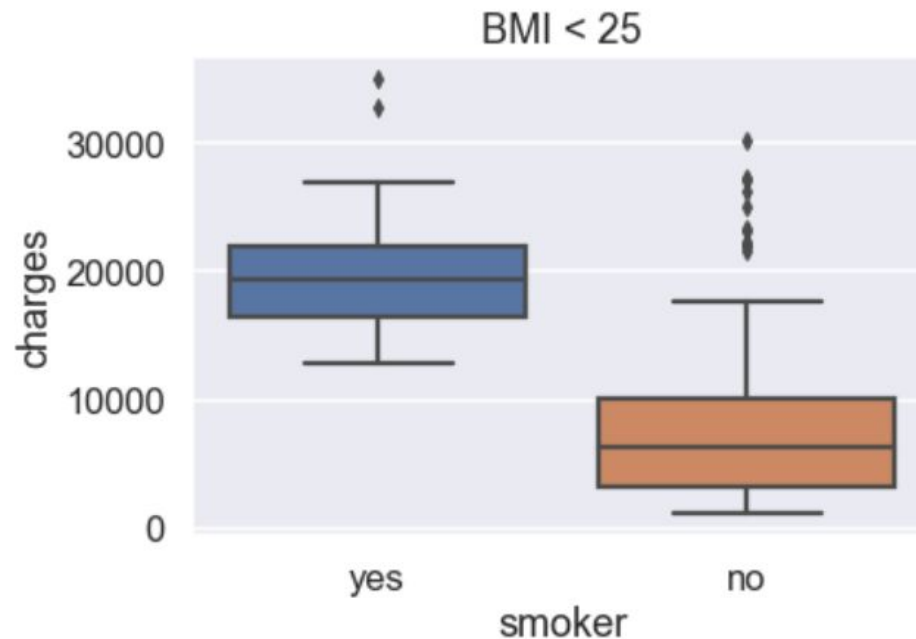
Tagihan Perokok dan Non-Perokok

- Perokok memiliki rata-rata tagihan asuransi hampir **4x non-perokok**
- Besar tagihan 75% non-perokok tidak jauh berbeda dengan nilai paling kecil dari tagihan pengguna perokok
- Perokok memiliki sebaran tagihan yang lebih besar dibandingkan non-perokok



Tagihan Pengguna dengan BMI < 25 dan BMI > 25

- Tagihan perokok juga lebih besar dari non-perokok baik untuk pengguna dengan BMI > 25 maupun pengguna dengan BMI < 25
- Pada pengguna dengan BMI < 25, pengguna perokok memiliki rata-rata tagihan 2.6 kali lebih besar dari non-perokok
- Pada pengguna dengan BMI > 25, pengguna perokok memiliki rata-rata tagihan 4 kali lebih besar dari non-perokok



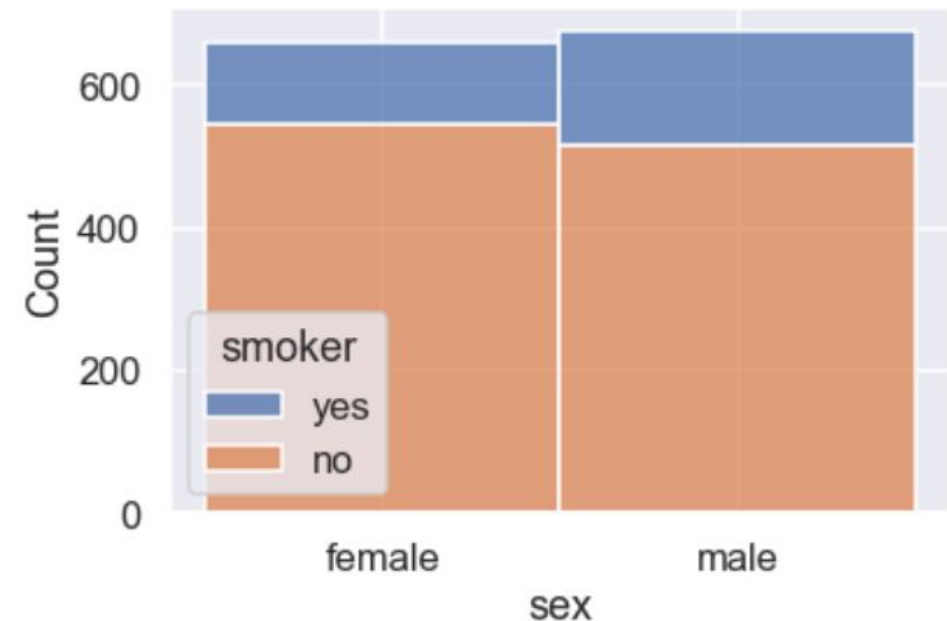
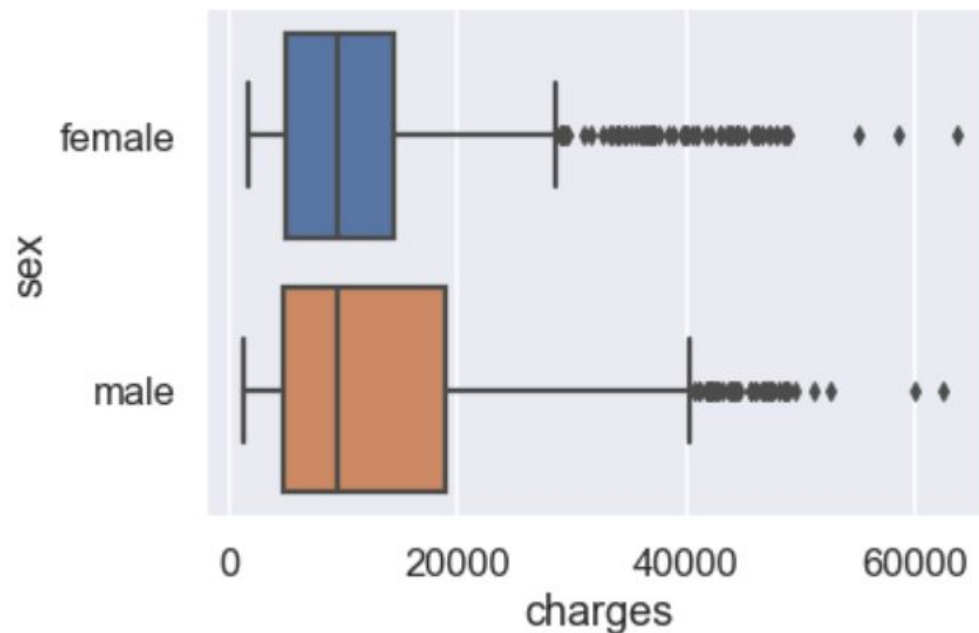
Analisis

- Dataset tagihan asuransi ini memiliki jumlah sampel sebanyak 1338 pengguna. Pengguna asuransi rata-rata berusia 39.2 dengan rata-rata umur yang serupa untuk pengguna laki-laki dan perempuan. **Rata-rata BMI** pengguna adalah sebesar 30.66 dan nilainya **tidak jauh berbeda untuk pengguna perokok dan non-perokok**. Nilai BMI ini [termasuk ke dalam kategori obesitas menurut CDC](#).
- Rata-rata usia untuk berbagai kategori pengguna tidak jauh berbeda. Rata-rata usia pengguna perokok dan non-perokok serta rata-rata usia laki-laki dan perempuan hampir serupa.
- Distribusi besar tagihan berekor panjang ke kanan. Nilai rata-rata tagihan **4 kali lebih besar untuk pengguna perokok dibanding non-perokok**. Hal yang sama ditemukan pada pengguna overweight. Perbedaan rata-rata tagihan pengguna perokok dan non-perokok pada pengguna non-overweight tidak sebesar yang ditemukan pada pengguna overweight. Hal ini mungkin menunjukkan tingginya BMI menyebabkan tingginya tagihan.

Analisis Variabel Kategorik

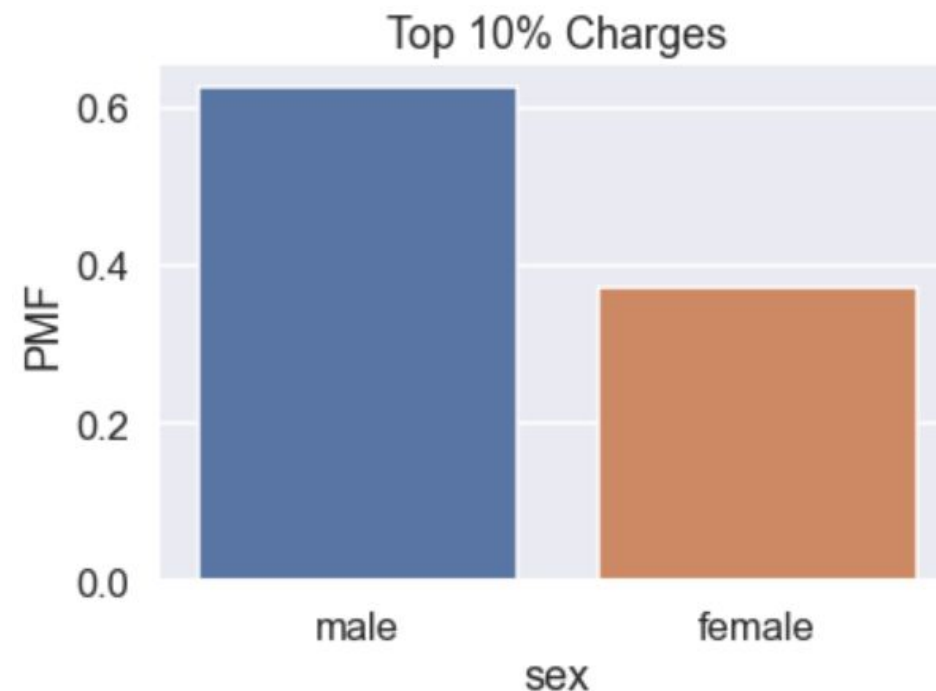
Jenis Kelamin dengan Tagihan Paling Tinggi

- Pengguna laki-laki memiliki rata-rata tagihan yang lebih tinggi dibanding perempuan dengan sebaran yang lebih besar pula
- Hal ini mungkin disebabkan **lebih banyak proporsi perokok di antara pengguna laki-laki (23.52%)** dari proporsi perokok di antara pengguna perempuan (17.52%).



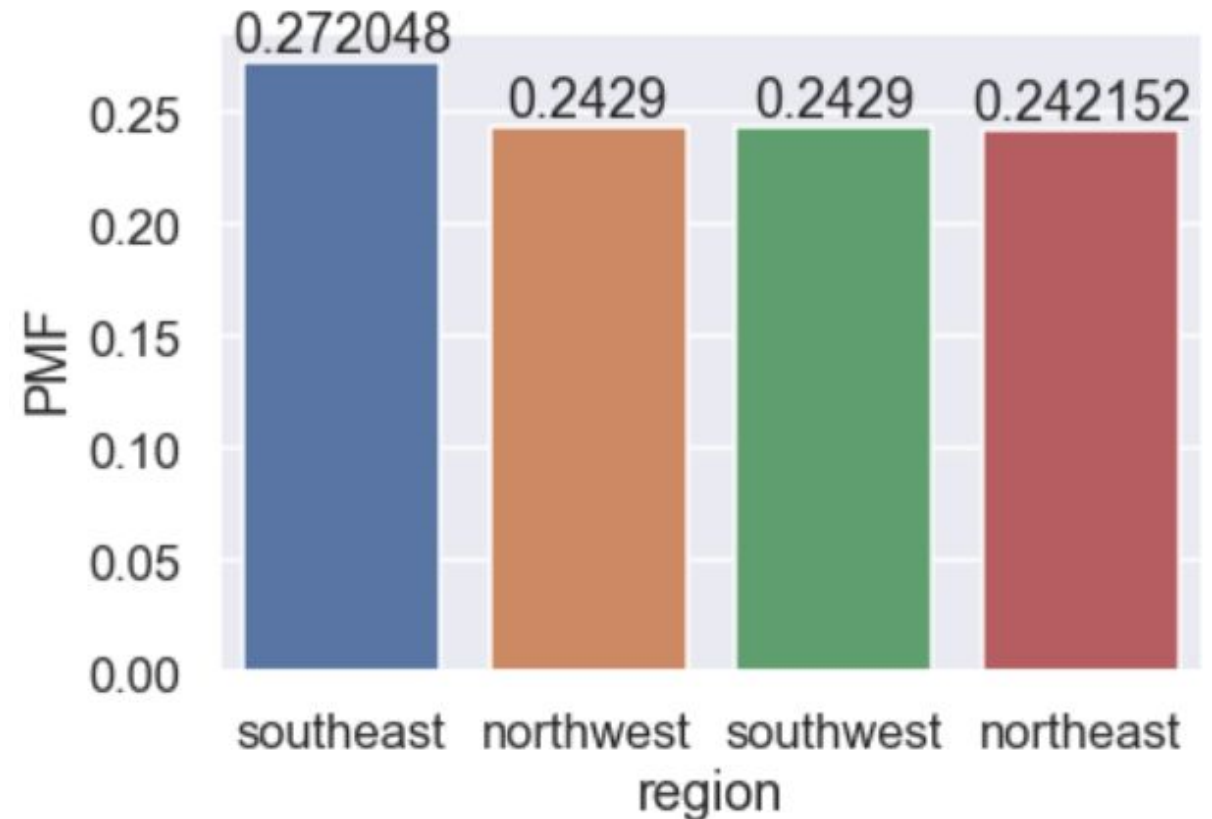
Jenis Kelamin dengan Tagihan Paling Tinggi

- Kita misalkan tagihan paling tinggi dengan mengambil 10% nilai teratas
- Pengguna laki-laki lebih mungkin memiliki tagihan paling tinggi dibandingkan perempuan dengan peluang sebesar 62.69%



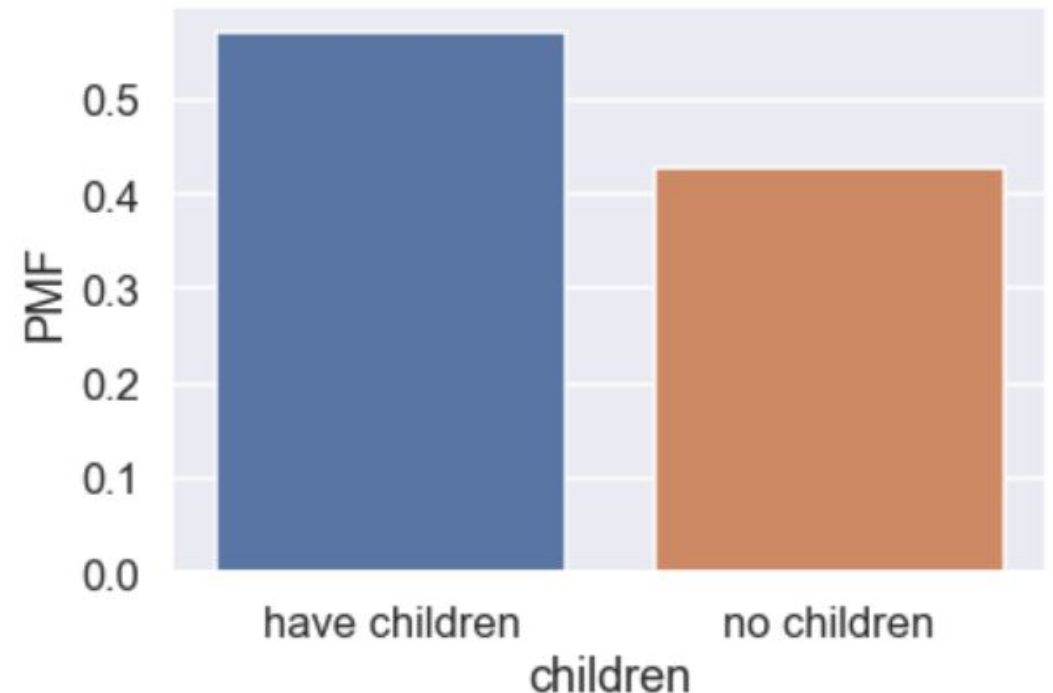
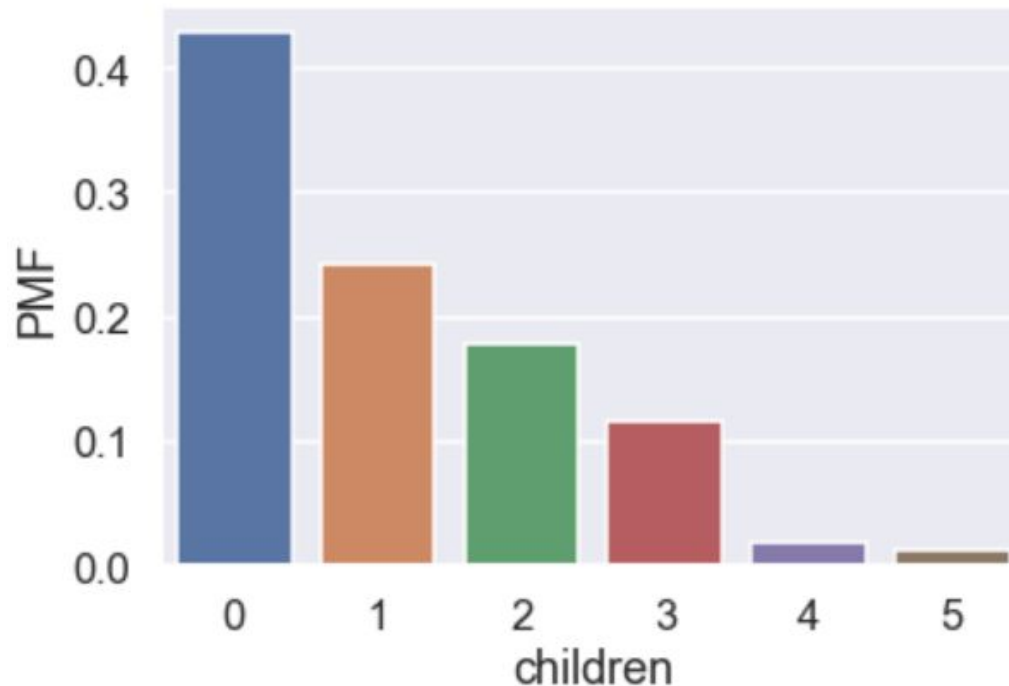
Proporsi Pengguna di Tiap Region

- Peluang pengguna asuransi berasal dari region Southeast adalah sebesar 27.2%, paling besar di antara semua region
- Peluang pengguna berasal dari Northwest, Southwest, dan Northeast adalah serupa, sekitar 24.2%



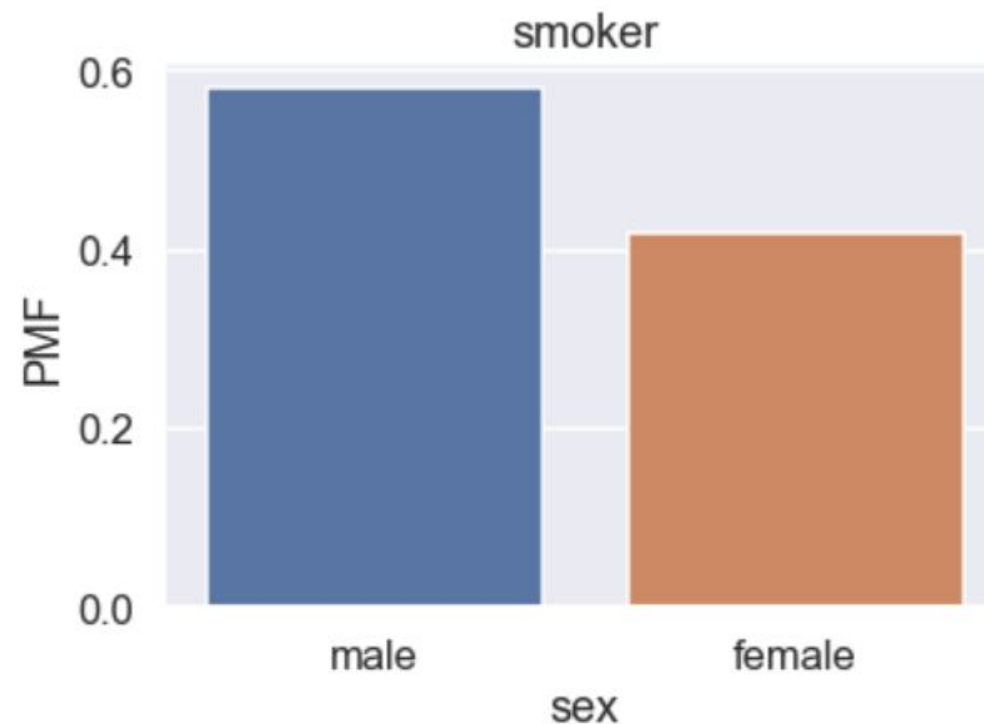
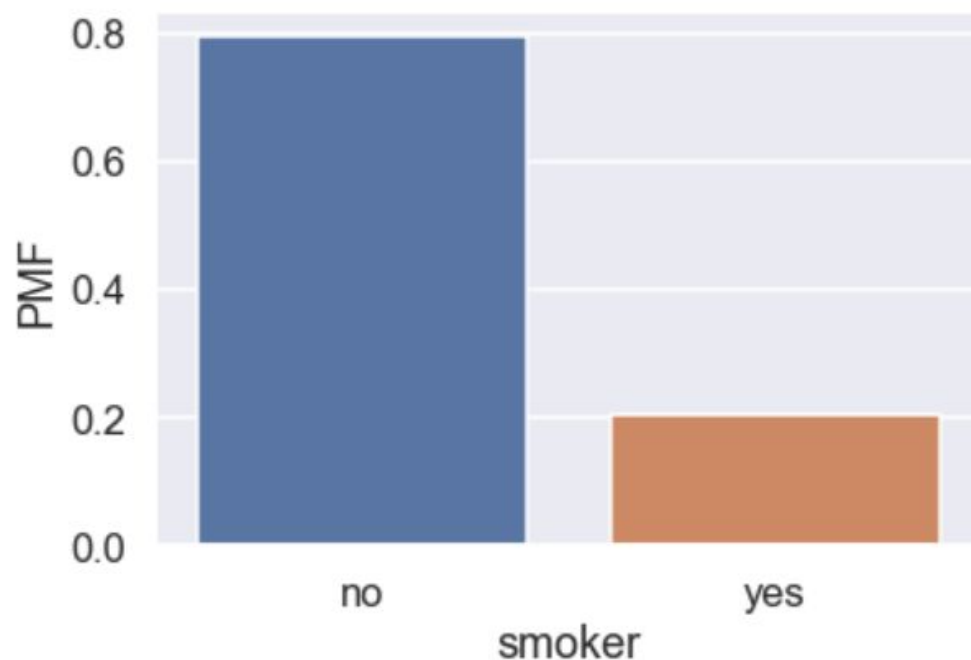
Proporsi Pengguna yang Memiliki Anak dan Tidak

- Orang yang tidak memiliki anak memiliki peluang yang lebih besar sebagai pengguna asuransi dibanding orang yang memiliki anak 1, 2, 3, 4, dan 5
- Di sisi lain, proporsi pengguna yang memiliki anak (1 atau lebih) lebih besar dibanding yang tidak memiliki anak.



Proporsi Perokok dan Non-Perokok

- Proporsi pengguna non-perokok hampir empat kali lebih besar dibandingkan perokok
- Peluang seorang pengguna adalah laki-laki jika diketahui dia adalah seorang perokok adalah sebesar 0.58, lebih besar dari perempuan



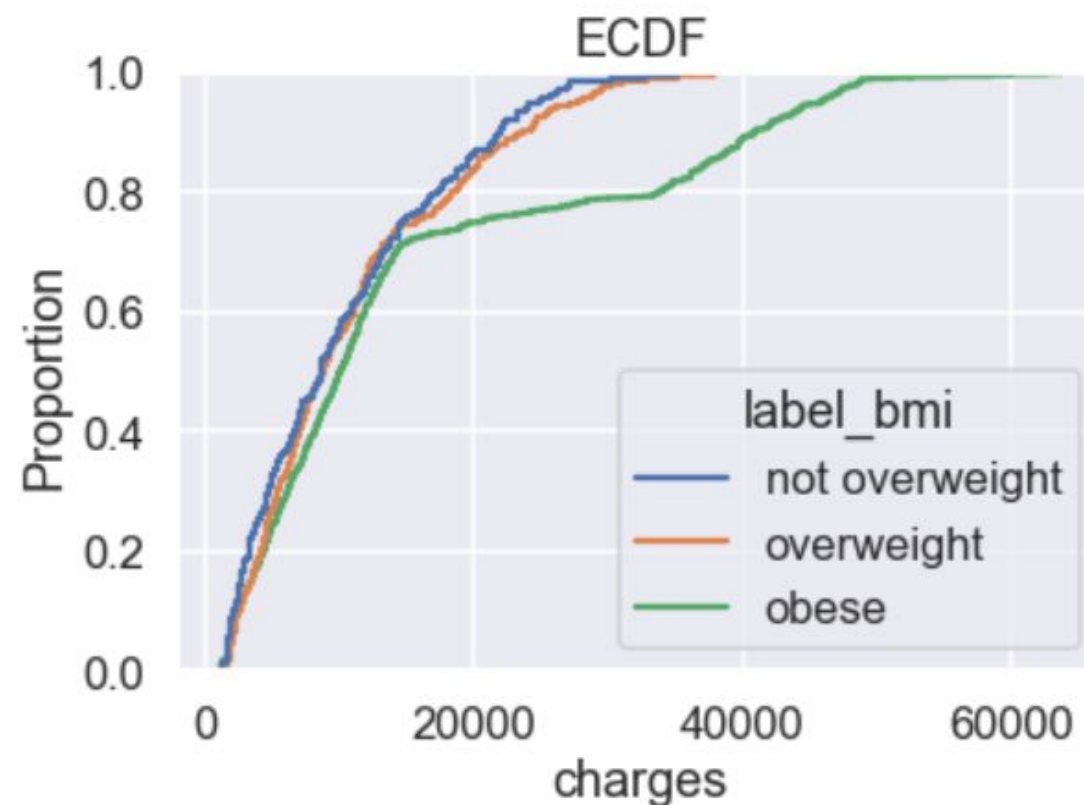
Analisis

- **Pengguna laki-laki berpeluang lebih besar untuk memiliki tagihan paling tinggi (top 10%)** dibanding perempuan. Hal ini dapat disebabkan proporsi perokok di antara pengguna laki-laki lebih besar dari proporsi perokok di antara perempuan.
- Proporsi pengguna asuransi di berbagai region relatif sama, tidak berbeda jauh kecuali untuk region Southeast yang memiliki proporsi sedikit lebih tinggi dibanding ketiga region lainnya.
- Proporsi non-perokok jauh lebih besar dari proporsi perokok. Seorang laki-laki memiliki peluang lebih besar sebagai pengguna asuransi jika diketahui dia adalah seorang perokok, dibandingkan perempuan.
- Kategori perokok atau bukan kemungkinan menjadi variabel penting yang menentukan besar tagihan.

Analisis Variabel Kontinu

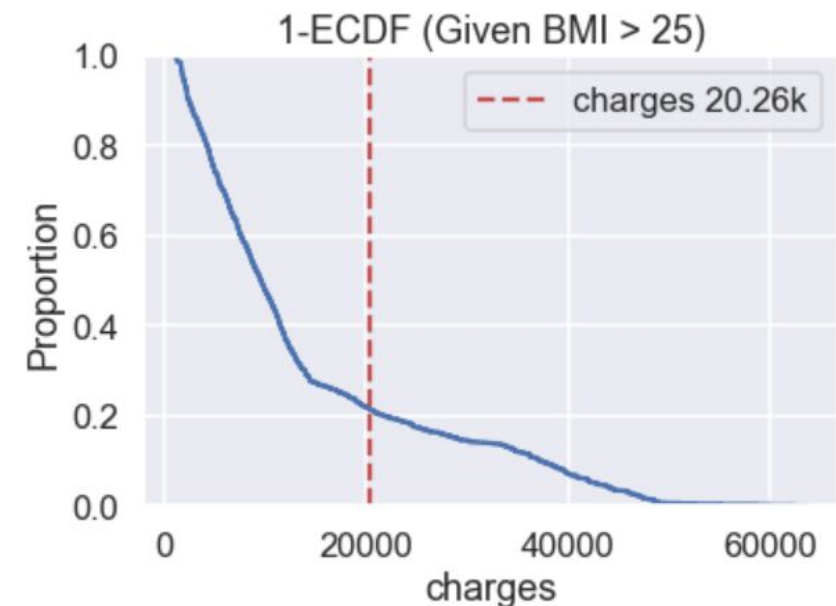
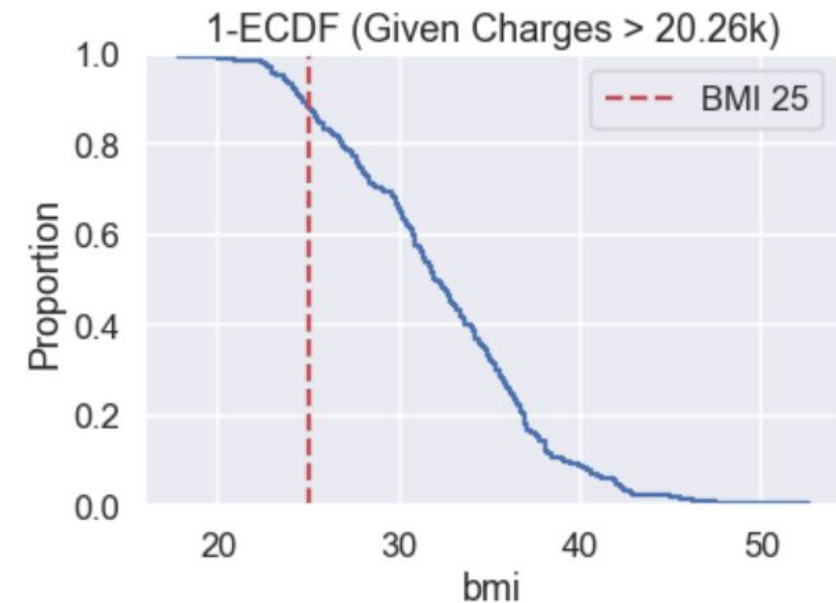
Peluang Tagihan untuk Berbagai Kategori BMI

- Tiap-tiap pengguna dibagi ke dalam beberapa kategori menurut BMI-nya
 - BMI di bawah 18.5 -> underweight
 - BMI antara 18.5 dan 25 -> healthy
 - BMI antara 25 dan 30 -> overweight
 - BMI di atas 30 -> obese
- Untuk penyederhanaan masalah, kita akan mengelompokkan BMI di bawah 25 menjadi kelompok “not overweight”
- Untuk pengguna yang **obesitas**, 20% pengguna memiliki tagihan sebesar **1.84 kali tagihan dari 80% pengguna yang overweight** dan hampir 2 kali tagihan pengguna yang memiliki BMI tidak overweight



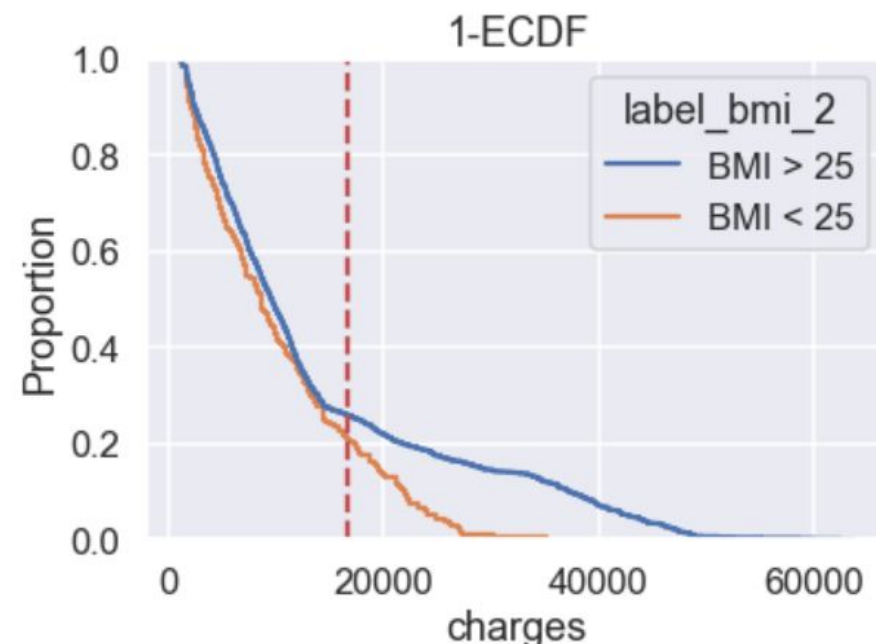
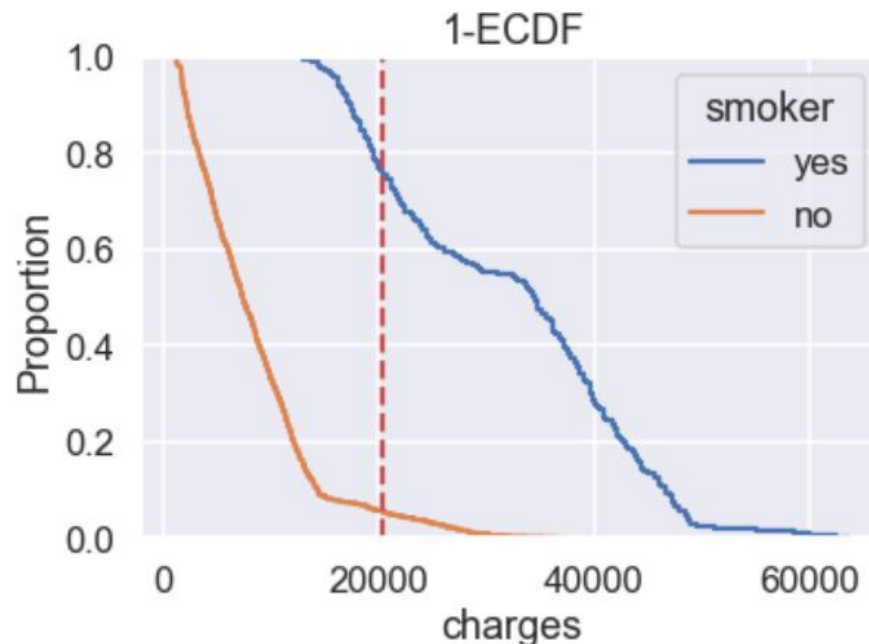
BMI vs Tagihan

- Peluang seorang pengguna memiliki BMI > 25 jika diketahui dia memiliki tagihan > 20.26k adalah 88.06%
- Peluang seorang pengguna memiliki tagihan > 20.26k jika diketahui dia memiliki BMI > 25 adalah 21.63%
- Seorang pengguna asuransi lebih mungkin memiliki BMI yang tinggi (> 25) jika diketahui dia memiliki tagihan yang tinggi (top 20% charges) dibanding peluang dia memiliki tagihan yang tinggi diketahui dia memiliki BMI yang tinggi



BMI vs Perokok

- Peluang pengguna dengan BMI > 25 memiliki tagihan > 20.26k adalah sebesar 21.63%
- Peluang pengguna perokok memiliki tagihan > 20.26k adalah sebesar 75.91%
- Pengguna memiliki peluang yang lebih besar untuk memiliki tagihan tinggi (top 20% charges) jika ia adalah seorang perokok dibanding seorang yang overweight



Analisis

- Untuk ketiga kategori BMI yaitu not overweight, overweight, dan obese, **pengguna kategori obese berpeluang paling besar untuk memiliki tagihan yang tinggi.** Sementara itu peluang pengguna kategori overweight untuk memiliki tagihan yang tinggi relatif sama besar dengan kategori not overweight
- Peluang seseorang untuk memiliki tagihan tinggi jika diketahui dia overweight tidak lebih besar dari peluangnya untuk memiliki BMI overweight jika diketahui dia memiliki tagihan tinggi
- Selain itu, seorang yang diketahui perokok lebih mungkin memiliki tagihan tinggi dibanding seseorang yang overweight
- Hal ini menunjukkan bahwa jika BMI dan kategori perokok dibandingkan, maka **kategori perokok/tidak lebih bisa menggambarkan peluang besarnya tagihan pengguna**

Korelasi Variabel

Korelasi

- Tidak ditemukan bukti yang cukup untuk mengatakan bahwa umur dan region pengguna berpengaruh pada besar tagihan
- Jenis kelamin laki-laki memiliki rata-rata tagihan yang lebih besar dari perempuan namun dapat dijelaskan oleh besarnya proporsi perokok di kalangan laki-laki dibandingkan proporsi perokok di kalangan perempuan
- BMI dan kategori perokok/tidak mempengaruhi besar tagihan
- Semakin besar BMI seorang pengguna, semakin besar peluang untuk memiliki tagihan yang besar
- Perokok memiliki peluang yang lebih besar dibanding non-perokok untuk memiliki tagihan yang besar
- Namun kategori perokok/tidak lebih mempengaruhi peluang pengguna untuk memiliki tagihan yang besar dibandingkan kategori BMI

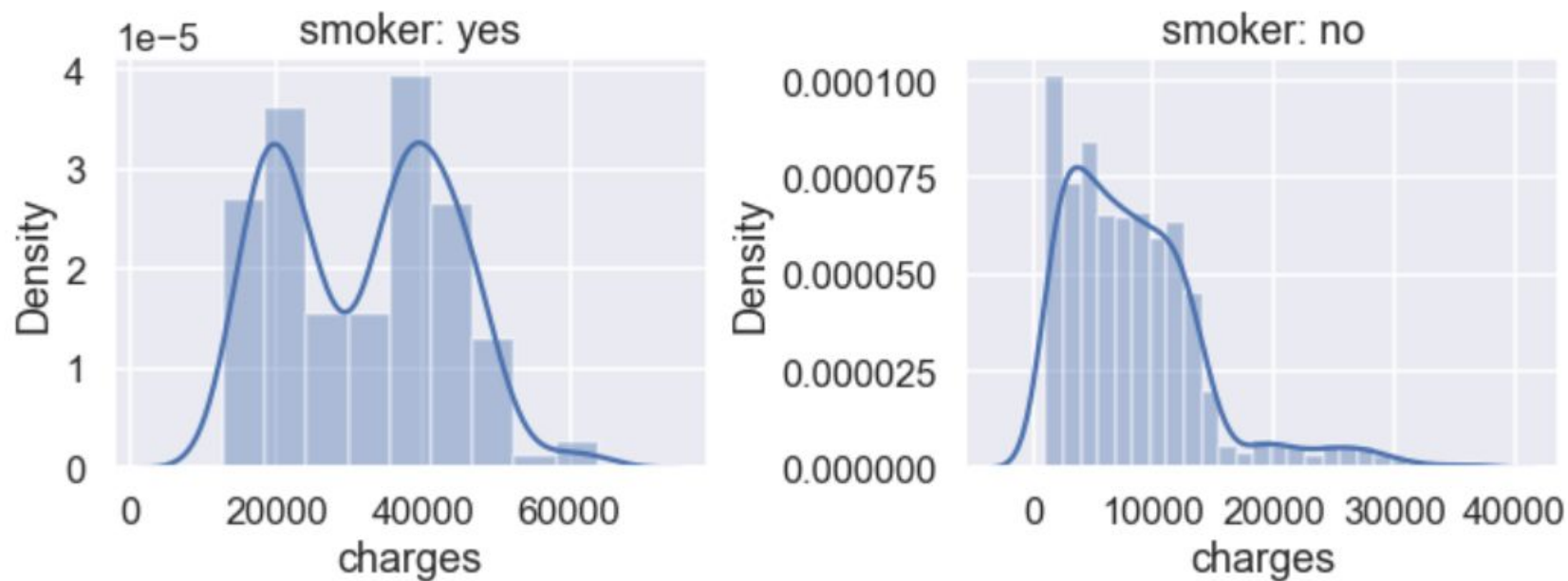
Uji Hipotesis

Hipotesis 1: Tagihan perokok lebih besar dari tagihan non-perokok

- Berdasarkan analisis sebelumnya, kita menemukan bahwa besar tagihan pengguna sangat dipengaruhi oleh apakah dia perokok atau bukan. Berdasarkan nilai rata-rata kedua kelompok, perokok memiliki tagihan yang lebih besar dibanding non-perokok.
- Dirumuskan:
 - H_0 : Rata-rata tagihan perokok lebih kecil sama dengan tagihan non-perokok
 - H_1 : Rata-rata tagihan perokok lebih besar dari tagihan non-perokok
- Ditentukan
 - $\alpha = 0.05$

Hipotesis 1: Tagihan perokok lebih besar dari tagihan non-perokok

- Diketahui:
 - Distribusi tagihan perokok dan non-perokok bukan distribusi normal
 - Standar deviasi dari populasi tidak diketahui
- Gunakan uji Mann-Whitney U



Hipotesis 1: Tagihan perokok lebih besar dari tagihan non-perokok

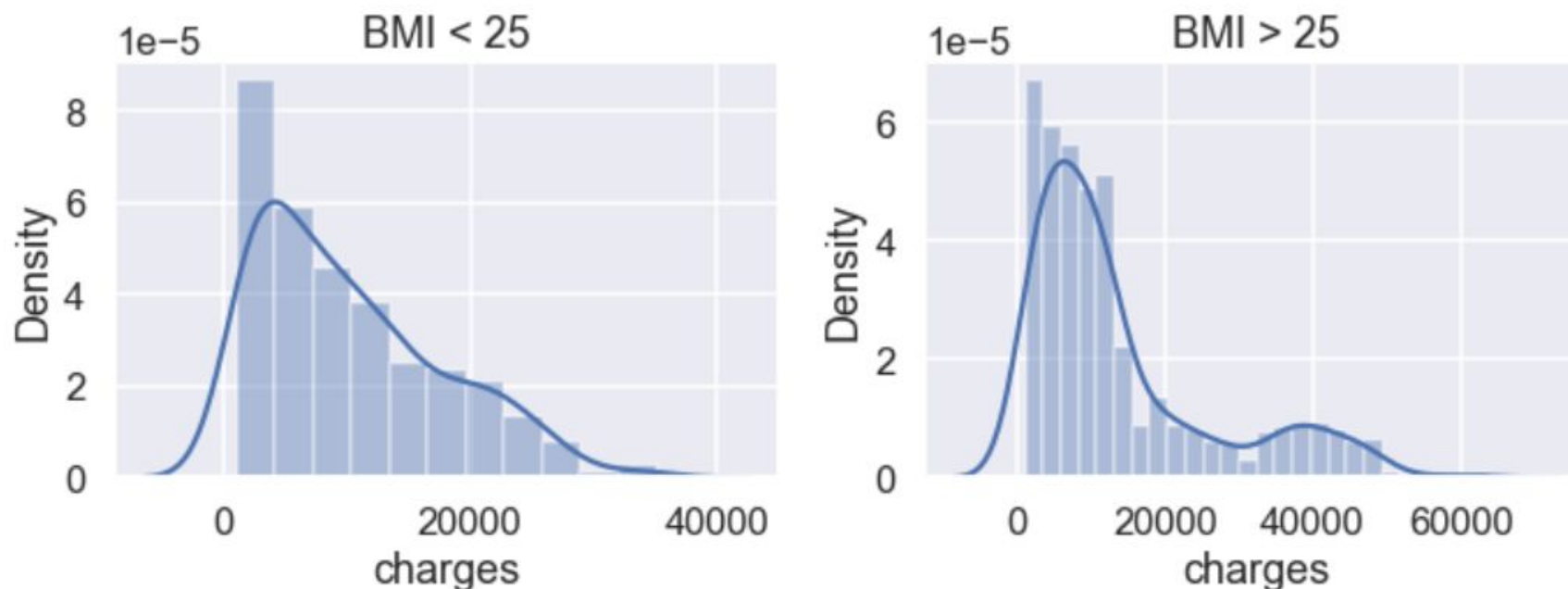
- Didapat p value = ~ 0.00000 (p value $< \alpha$)
- H_0 ditolak, H_1 diterima: Tagihan perokok lebih besar dari non-perokok

Hipotesis 2: Tagihan pengguna dengan BMI di atas 25 lebih besar daripada tagihan pengguna dengan BMI di bawah 25

- Berdasarkan analisis sebelumnya, kita menemukan bahwa besar tagihan pengguna dengan BMI > 25 lebih besar dari pengguna dengan BMI < 25
- Dirumuskan:
 - H_0 : Rata-rata tagihan pengguna dengan BMI > 25 lebih kecil sama dengan tagihan pengguna dengan BMI < 25
 - H_1 : Rata-rata tagihan pengguna dengan BMI > 25 lebih besar dari tagihan pengguna dengan BMI < 25
- Ditentukan
 - $\alpha = 0.05$

Hipotesis 2: Tagihan pengguna dengan BMI di atas 25 lebih besar daripada tagihan pengguna dengan BMI di bawah 25

- Diketahui:
 - Distribusi tagihan pengguna dengan BMI > 25 dan BMI < 25 bukan distribusi normal
 - Standar deviasi dari populasi tidak diketahui
- Gunakan uji Mann-Whitney U



Hipotesis 2: Tagihan pengguna dengan BMI di atas 25 lebih besar daripada tagihan pengguna dengan BMI di bawah 25

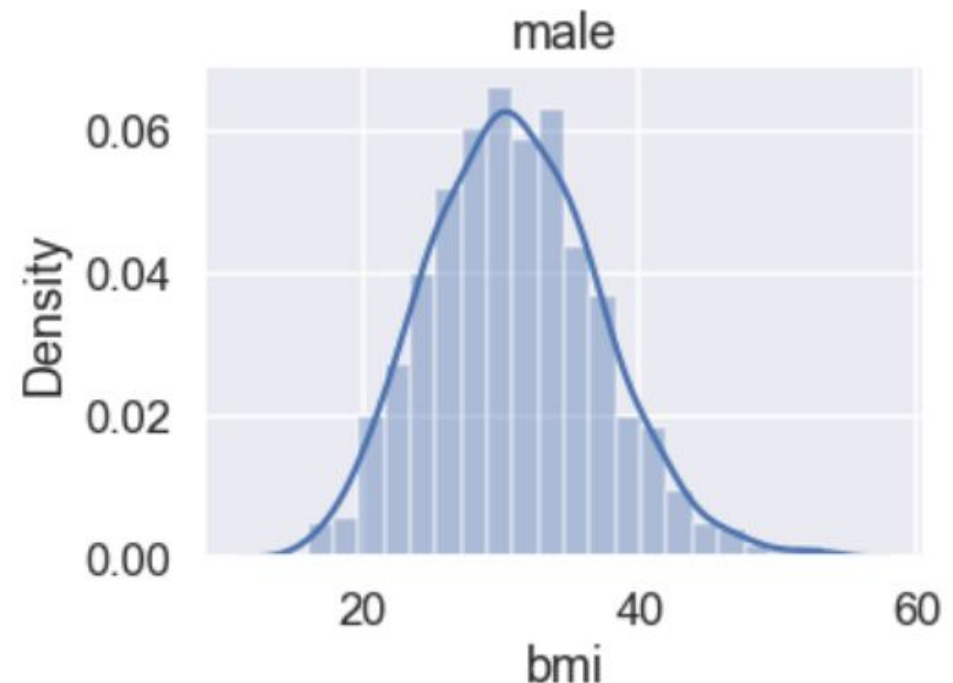
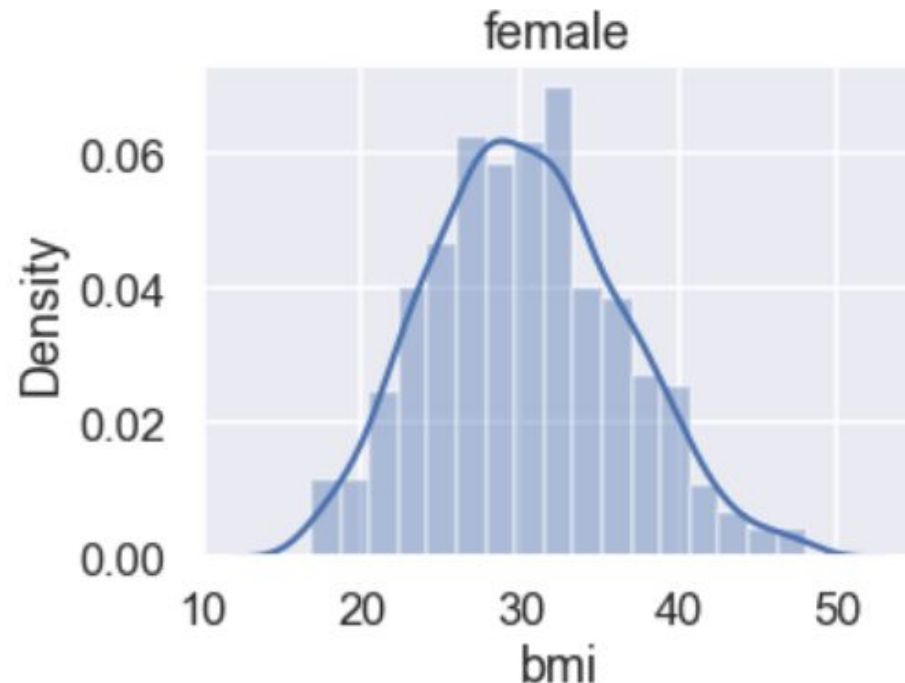
- Didapat p value = 0.00353 (p value < alpha)
- H_0 ditolak, H_1 diterima: Tagihan pengguna dengan BMI > 25 lebih besar dari pengguna dengan BMI < 25

Hipotesis 3: Rata-rata BMI perempuan sama dengan BMI laki-laki

- Berdasarkan analisis sebelumnya, kita menemukan bahwa rata-rata BMI perempuan tidak berbeda jauh dengan laki-laki
- Dirumuskan:
 - H_0 : Rata-rata BMI perempuan sama dengan rata-rata BMI laki-laki
 - H_1 : Rata-rata BMI perempuan tidak sama dengan rata-rata BMI laki-laki
- Ditentukan
 - $\alpha = 0.05$
-

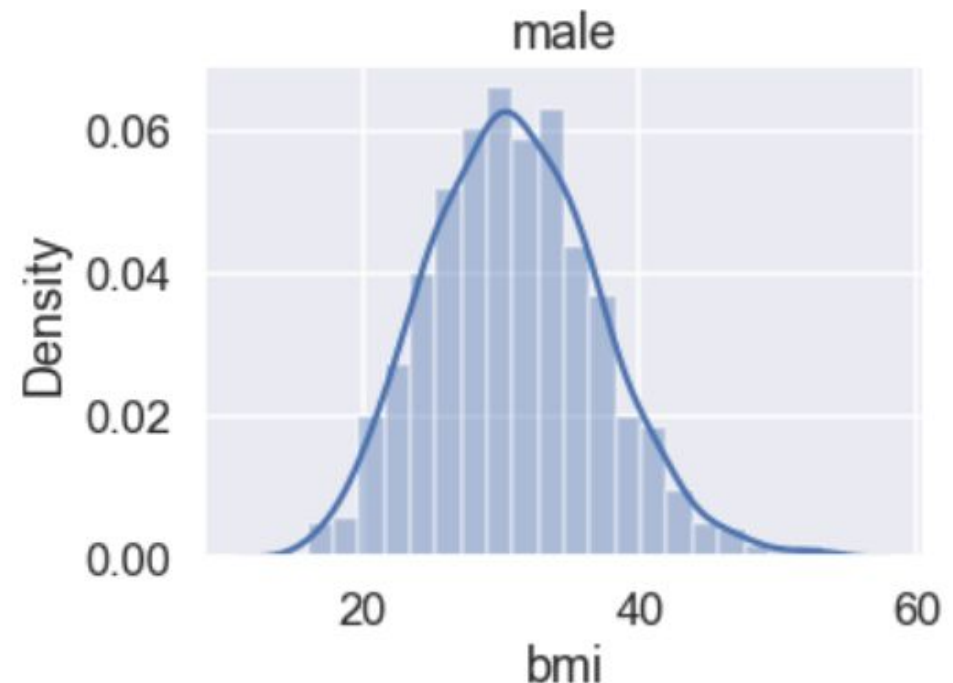
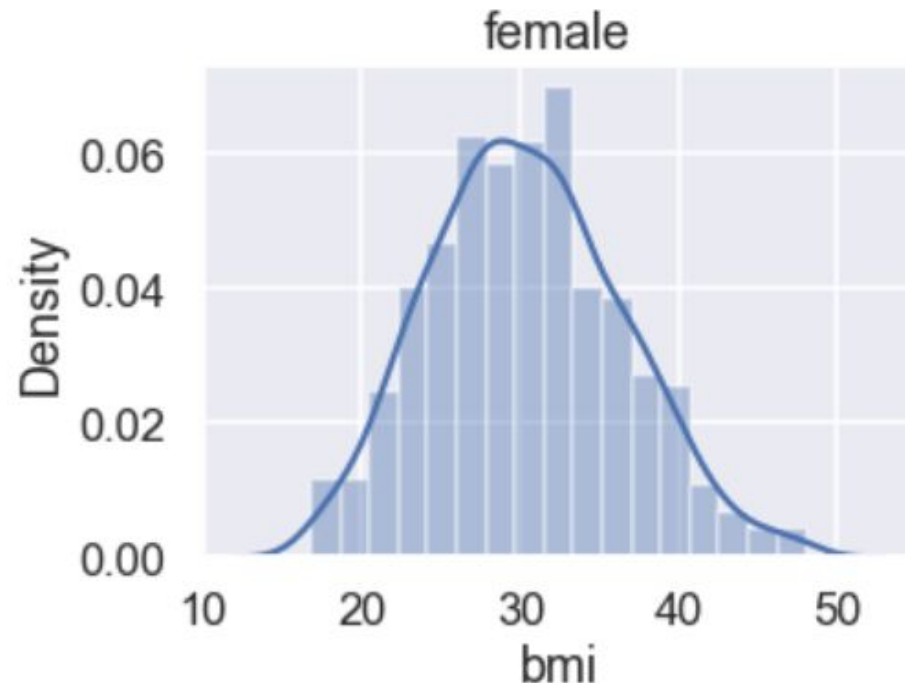
Hipotesis 3: Rata-rata BMI perempuan sama dengan BMI laki-laki

- Diketahui:
 - Distribusi BMI perempuan dan laki-laki adalah normal
 - Standar deviasi dari populasi tidak diketahui
- Gunakan T test



Hipotesis 3: Rata-rata BMI perempuan sama dengan BMI laki-laki

- p value = 0.08997 (p value > alpha)
- Tidak cukup bukti untuk menolak H_0 : Rata-rata BMI perempuan sama dengan rata-rata BMI laki-laki



Kesimpulan

Kesimpulan

- Terdapat beberapa variabel yang mempengaruhi besarnya tagihan asuransi, yaitu perokok/tidak, BMI, dan jenis kelamin
- Jenis kelamin mungkin dipengaruhi oleh proporsi perokok/tidak
- Kategori perokok/tidak lebih besar pengaruhnya untuk menentukan peluang besarnya tagihan asuransi dibanding BMI

Referensi

- <https://www.cdc.gov/obesity/basics/adult-defining.html>