

## PROYEK 2B - Kasih kredit ga ya?

Kelompok 8:

- Bryan Christopher Wijaya
- James Patrick Oentoro
- Noel Christevent Mandak
- Stefannus Christian
- Tiffany Sondakh

Link Tugas Pemrograman :

<https://datalore.jetbrains.com/notebook/dzMZsBqYsyz4CSIR6fN6a/BU8vZ43x3NAyhn3r1Ge0uy/>

1. Tulislah ringkasan singkat (maksimal 4 paragraf) untuk menceritakan proses statistika yang Anda lakukan, mulai dari mengekstrak informasi dari dataset hingga menghasilkan inferensi statistika.
2. Product Manager ingin Anda menjawab pertanyaan-pertanyaan berikut dengan justifikasi berupa inferensi statistika.
  - a. Berikan ulasan mengenai nilai rata-rata dan standar deviasi dari Income dan Car Price penduduk di kota tersebut secara keseluruhan. Berikan penafsiran terhadap hal ini.
  - b. Berikan sebuah batas bawah  $s$  sehingga dengan keyakinan 95%, rata-rata penghasilan keseluruhan kota tersebut nilainya lebih besar atau sama dengan  $s$ .
  - c. Bank mengkategorikan warga yang memiliki pendapatan  $> 60K$  USD per tahun sebagai '*calon nasabah yang prospektif*'. Berikan inferensi statistika untuk persentase banyaknya calon nasabah di kota tersebut secara keseluruhan.
  - d. Dari antara semua calon nasabah yang prospektif, terdapat kumpulan calon nasabah *yang akan diberikan promosi*, yaitu mereka yang sekiranya tertarik untuk memiliki mobil dengan harga di atas 15K USD. Jika diketahui kota tersebut memiliki populasi sebesar 1,2 juta penduduk, berikan estimasi statistika untuk rentang banyaknya calon nasabah yang akan diberikan promosi oleh pihak bank. Berikan penjelasan dan justifikasi statistiknya.

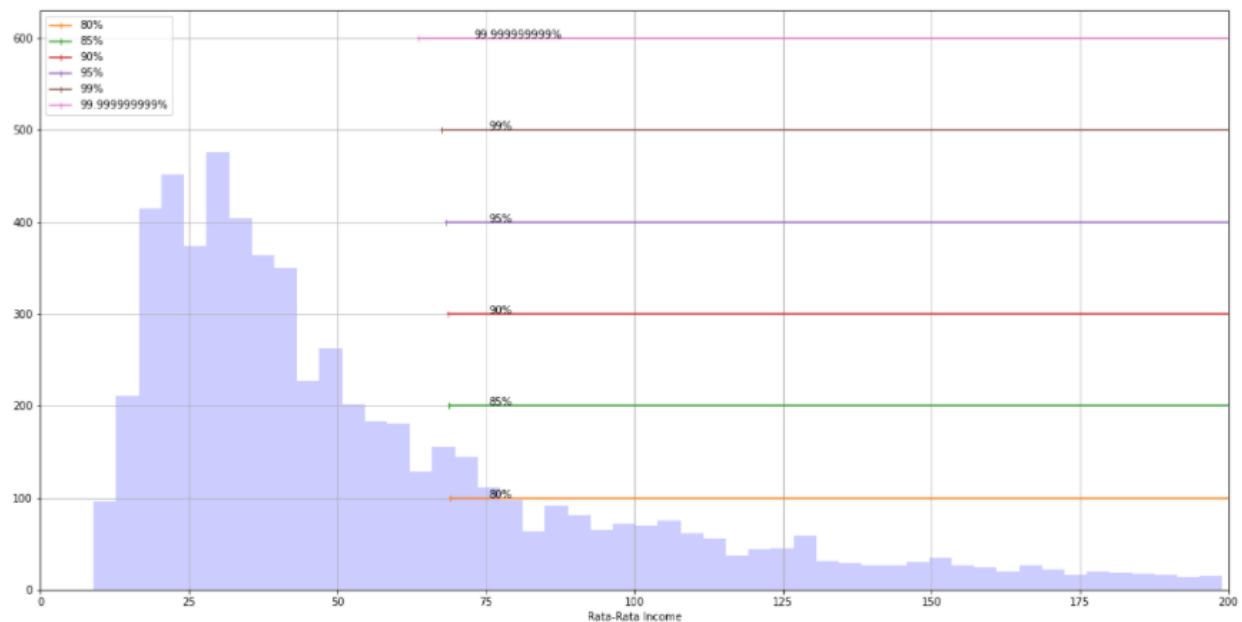
Pada laporan kali ini kami akan menganalisis data penghasilan dan harga mobil dari suatu kota untuk mengetahui gambaran rentang rata-rata penghasilan dan rata-rata harga mobil yang dimiliki warga disana sebagai tinjauan membuat produk Kredit Pemilihan Mobil yang tepat untuk warga kota tersebut. Data yang akan kami analisis adalah survei dari 6400 KK dengan 28 parameter. Namun, dalam kasus kali ini kami hanya akan mengambil parameter penghasilan pertahun (income) dan harga mobil (car price) yang dimiliki warga tersebut dalam satuan ribu dolar (USD).

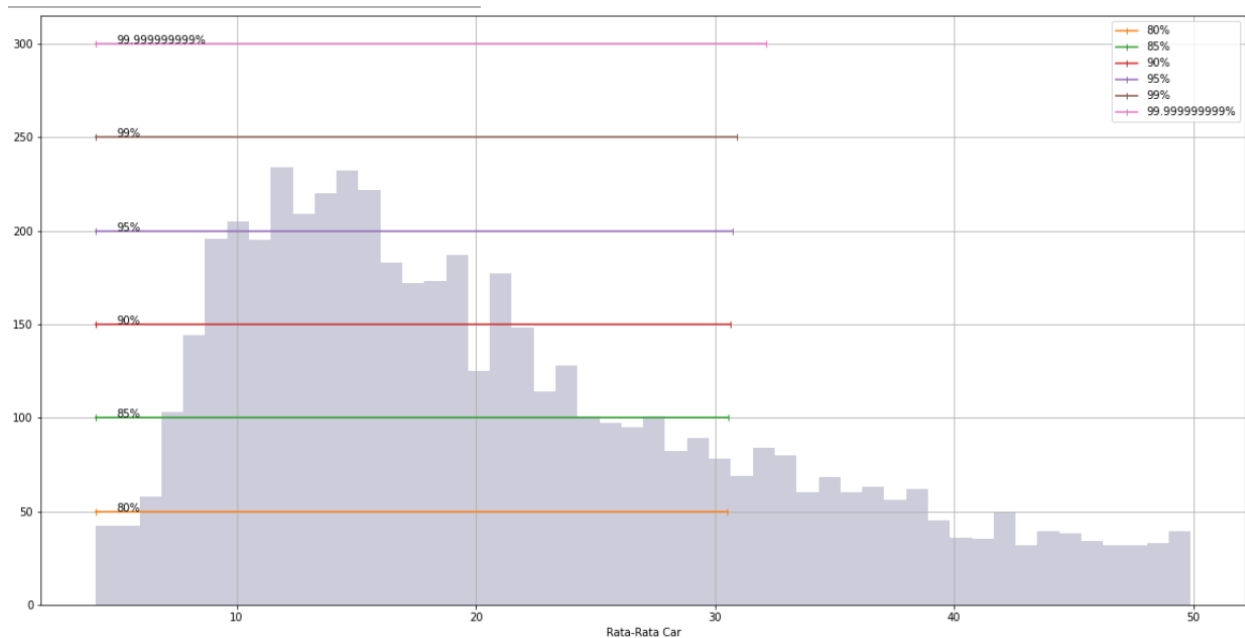
Perlu diingat bahwa data yang kita miliki merupakan data sampel dari populasi suatu kota. Maka dari itu, diperlukan suatu alat statistika yang dapat memberikan gambaran mengenai keseluruhan populasi dari data sampel yang ada. Untuk itu, kami menggunakan statistika inferensial berupa confidence interval, yaitu mengestimasi rentang nilai rata-rata dan

standar deviasi populasi dari sampel. Melalui analisa confidence interval, kami mendapatkan hasil sebagai berikut:

- Estimasi rata-rata penghasilan keseluruhan kota tersebut lebih besar atau sama dengan 68.28K USD dengan kepercayaan 99%.
- Estimasi rata-rata harga mobil lebih kecil atau sama dengan 30.9K USD dengan tingkat kepercayaan 99%.
- Estimasi rata-rata standar deviasi penghasilan kota tersebut ada pada rentang 76.66K hingga 79.37K dengan tingkat kepercayaan 95%.
- Estimasi rata-rata standar deviasi harga mobil kota tersebut ada pada rentang 21.62K hingga 22.38K dengan tingkat kepercayaan 95%.
- persentase calon nasabah prospektif ada di interval 34,5% sampai 36,8% dengan kepercayaan 95%.
- Rentang banyaknya nasabah yang mendapatkan promosi adalah 413.789 sampai 441.960 dengan kepercayaan 95%.

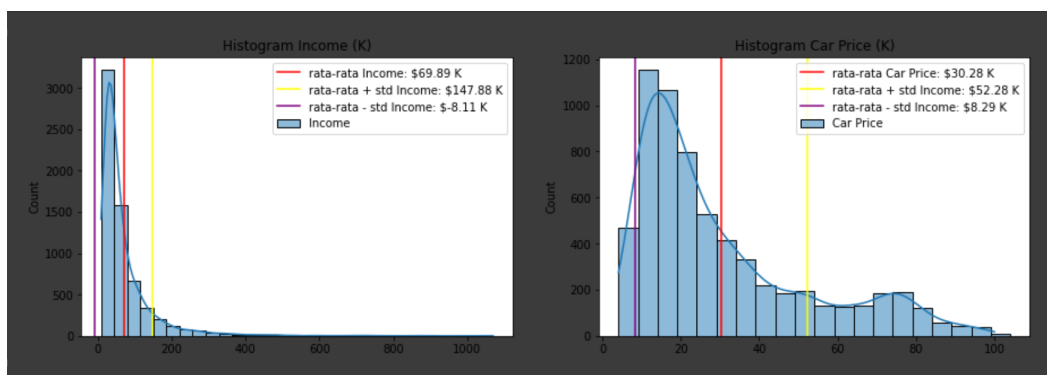
Kami memilih mengestimasi rata-rata penghasilan kota dan harga mobil dengan tingkat kepercayaan 99% karena melalui visualisasi dibawah, kami mendapatkan bahwa rentang tingkat kepercayaan 99% dan 95% tidak terlalu beda jauh.

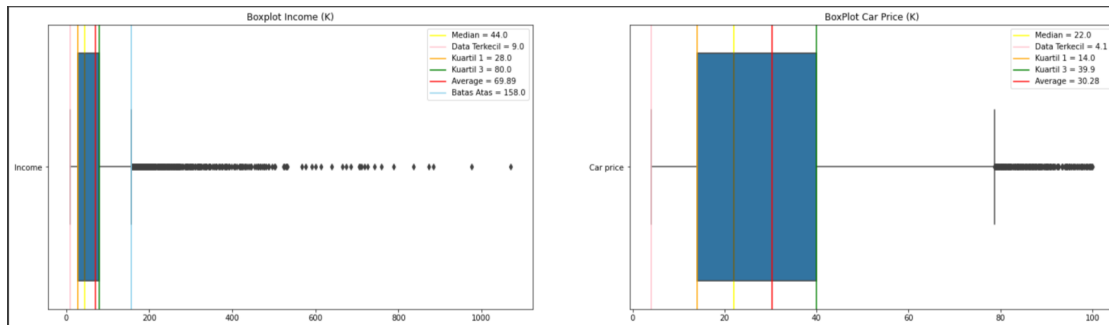




Dengan data batas bawah income sebesar 67.62 dan batas atas harga mobil sebesar 31. Maka kita dapat menyimpulkan bahwa worst case lama waktu pelunasan pembelian mobil adalah  $31/67.62 = 0.458$  tahun atau 5.5 bulan (dengan asumsi mereka menggunakan 100 persen pendapatan untuk melunasi mobil). Maka kita dapat dengan yakin memberikan produk cicilan untuk 5 - 6 bulan.

## Menjawab Pertanyaan:





- a. Dapat dilihat dari histogram diatas bahwa rata-rata dari pendapatan (income) adalah sedikit di bawah \$70K dollar yaitu \$69.89K dollar. Tetapi fitur income memiliki standar deviasi yang cukup tinggi yaitu \$77.99K dollar. Ini artinya bahwa rentang dari income sangat tinggi dan hal ini dapat dilihat dari histogram dan boxplot juga. Dapat dilihat dari histogram bahwa histogram berbentuk right skewed (positive skewed). Ini artinya bahwa rata-rata lebih besar dari median (dapat dilihat dari boxplot). Artinya juga bahwa kebanyakan titik data bertumpuk pada sebelah kiri dari histogram. Artinya bahwa lebih banyak orang yang income nya cenderung rendah dibandingkan dengan orang yang memiliki income yang tinggi. Dapat dilihat dari boxplot income bahwa data banyak tersebar pada rentang \$28K - \$80K tetapi akibat standar deviasi yang tinggi, terdapat banyak sekali data pencilan yang terletak di sebelah kanan dari boxplot. Artinya data - data pencilan ini jauh lebih besar dari rata - rata. Untuk *car price*, rata-rata nya adalah \$30.28K. Ini artinya bahwa kebanyakan warga kota tersebut memiliki mobil dengan harga tidak jauh dari \$30.28K. Standar deviasi dari fitur *car price* adalah \$21.998K (cukup tinggi). Kebanyakan dari titik-titik data *car price* ini tersebar pada rentang \$14K - \$39.9K. Ada juga warga yang memiliki harga mobil yang sangat murah yaitu \$4.1K tetapi ada juga warga yang memiliki mobil yang sangat mahal yaitu \$100K. Penafsiran yang bisa diambil adalah sebagai berikut. Standar deviasi yang tinggi menyebabkan banyaknya pencilan extreme yang tinggi pada fitur *income*. Dapat dilihat juga dari *boxplot car price* bahwa standar deviasi yang cukup tinggi menyebabkan banyak pencilan yang tinggi juga. Penafsiran terdapat hal ini adalah orang dengan *income* yang tinggi cenderung membeli mobil dengan harga yang tinggi juga dan orang yang memiliki income rata-rata cenderung membeli mobil yang sesuai range kemampuan nya yaitu sekitar \$30K (setengah dari rata-rata income).
- b. Rata-rata penghasilan keseluruhan kota tersebut diatas atau sama dengan 71.490 USD per tahun dengan kepercayaan 95%.
- c. Persentase calon nasabah prospektif dapat temukan dengan menggunakan estimasi indikator random variable.  
Misalkan Y adalah indikator nasabah prospektif. Y akan bernilai 1 jika nasabah memiliki penghasilan diatas 60 ribu USD dan bernilai 0 untuk sisanya.  
Kami mencari proporsi nasabah prospektif dengan menghitung rata-rata dari Y, yaitu 0,357. Kami juga membutuhkan standar deviasi untuk mencari estimasi, yaitu 0,479. Dengan melakukan statistika inferensial dari kedua sisi dengan tingkat kepercayaan

95%, didapatkan estimasi interval proporsi nasabah prospektif secara keseluruhan dari 0,345 sampai 0.368.

Jadi persentase calon nasabah prospektif ada di interval 34,5% sampai 36,8%.

- d. Setelah melakukan filter data sampel, ternyata jumlah nasabah dengan kriteria prospektif sama dengan nasabah kriteria nasabah yang menerima promosi. Jadi, estimasi interval proporsi dari nasabah prospektif juga sama dengan estimasi interval proporsi nasabah akan yang menerima promosi. Dengan demikian, jika diketahui populasi sebesar 1,2 juta penduduk, rentang banyaknya nasabah yang mendapatkan promosi adalah 413.789 sampai 441.960 dengan kepercayaan 95%.