

# Analisis Hubungan Antara Gender, Pendapatan Tahunan dan Profesi dengan Pola Pengeluaran Konsumen

Raymussen Arthur Wijaya<sup>1</sup>, Muhammad Abidzar Prayitno<sup>2</sup>

<sup>1, 2</sup>Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia

<sup>1</sup>raymussen.arthur@student.umn.ac.id, <sup>2</sup>muhammad.abidzar2@student.umn.ac.id

**Abstrak** — Penelitian ini bertujuan untuk menganalisis hubungan antara gender, pendapatan tahunan, dan profesi dengan pola pengeluaran konsumen. Gender, pendapatan, dan profesi merupakan faktor penting yang dapat mempengaruhi perilaku konsumsi. Penelitian ini mengidentifikasi bahwa perempuan cenderung mengalokasikan pengeluaran untuk kebutuhan rumah tangga dan kesehatan, sementara laki-laki lebih banyak menghabiskan uang untuk teknologi dan hiburan. Di sisi lain, pendapatan tahunan yang lebih tinggi terkait dengan pola pengeluaran yang lebih beragam, termasuk barang mewah, sementara konsumen dengan pendapatan lebih rendah lebih fokus pada kebutuhan dasar. Profesi juga memainkan peran dalam menentukan pola konsumsi, di mana pekerjaan tertentu menuntut pengeluaran untuk barang atau layanan yang mendukung aktivitas profesional. Hasil penelitian ini diharapkan dapat memberikan wawasan bagi pelaku bisnis dalam merancang strategi pemasaran yang lebih efektif dan tepat sasaran berdasarkan karakteristik demografis konsumen.

**Kata kunci** : Pendapatan Tahunan, Gender, Profesi, Pola Pengeluaran, Konsumen.

## I. PENDAHULUAN

### A. Latar Belakang

Dalam dunia bisnis, memahami pola pengeluaran konsumen merupakan elemen penting untuk mendesain strategi pemasaran yang efektif. Pola pengeluaran ini sering kali dipengaruhi oleh berbagai faktor, baik internal maupun eksternal, yang mencakup karakteristik demografis dan psikografis. Di antara faktor-faktor tersebut, gender, pendapatan tahunan, dan profesi menonjol sebagai variabel penting yang dapat memengaruhi perilaku konsumsi.

Gender sering kali dikaitkan dengan preferensi konsumsi yang berbeda. Misalnya, perempuan cenderung mengalokasikan pengeluaran pada kebutuhan rumah tangga, kesehatan, dan produk perawatan pribadi, sedangkan laki-laki memiliki kecenderungan untuk membelanjakan uangnya pada hiburan atau teknologi. Perbedaan ini tidak hanya didasarkan pada preferensi pribadi, tetapi juga pada peran sosial dan budaya yang melekat pada masing-masing gender.

Pendapatan tahunan adalah faktor lain yang sangat signifikan. Konsumen dengan pendapatan lebih tinggi cenderung memiliki daya beli yang lebih besar, sehingga dapat mengalokasikan sebagian dari penghasilannya untuk barang dan jasa yang tidak esensial, seperti barang mewah atau perjalanan. Sebaliknya, konsumen dengan pendapatan yang lebih rendah lebih memprioritaskan kebutuhan dasar.

Selain itu, profesi juga berperan penting dalam membentuk pola konsumsi. Pekerjaan tertentu sering kali membutuhkan pengeluaran pada produk atau layanan khusus yang mendukung aktivitas profesional. Sebagai contoh, seorang profesional di bidang teknologi cenderung mengeluarkan uang lebih banyak untuk alat elektronik atau perangkat lunak, sementara pekerja di industri kreatif mungkin lebih banyak membelanjakan uangnya pada perangkat kreatif seperti kamera atau software desain.

Mengingat pentingnya variabel-variabel ini, penelitian ini bertujuan untuk menganalisis hubungan antara gender, pendapatan tahunan, dan profesi dengan pola pengeluaran konsumen.

Hasil penelitian ini diharapkan dapat memberikan wawasan berharga bagi pelaku bisnis dalam mengidentifikasi segmen pasar dan merancang strategi pemasaran yang lebih efektif dan tepat sasaran.

## B. Hipotesis

1. Ho : Tidak terdapat hubungan signifikan antara gender dan pola pengeluaran konsumen.

Ha1 : Terdapat hubungan signifikan antara gender dan pola pengeluaran konsumen.

2. Ho2 : Tidak terdapat hubungan signifikan antara pendapatan tahunan dan pola pengeluaran konsumen.

Ha2 : Terdapat hubungan signifikan antara pendapatan tahunan dan pola pengeluaran konsumen.

3. Ho3 : Tidak terdapat hubungan signifikan antara profesi dan pola pengeluaran konsumen.

Ha3 : Terdapat hubungan signifikan antara profesi dan pola pengeluaran konsumen.

Pengujian hipotesis ini bertujuan untuk memberikan bukti empiris tentang bagaimana faktor-faktor demografi memengaruhi perilaku konsumsi, baik secara parsial maupun simultan.

## II. KAJIAN LITERATUR

Kajian literatur ini mengacu pada beberapa studi sebelumnya yang relevan dengan hubungan antara gender, pendapatan tahunan, profesi, dan pola pengeluaran konsumen. Gender memainkan peran penting dalam membentuk pola konsumsi. Smith et al. (2020) menunjukkan bahwa perempuan lebih sering mengalokasikan pengeluarannya pada produk yang mendukung perawatan rumah tangga dan kesehatan, sedangkan laki-laki cenderung lebih memilih produk yang berkaitan dengan teknologi dan hiburan. Perbedaan ini didukung oleh peran sosial dan

budaya yang menggarisbawahi tanggung jawab serta kebutuhan yang berbeda antara laki-laki dan perempuan, sehingga menciptakan preferensi konsumsi yang unik di masing-masing kelompok.

Pendapatan tahunan juga memiliki pengaruh yang signifikan terhadap pola pengeluaran konsumen. Sebuah studi oleh Johnson dan Miller (2018) menemukan bahwa konsumen dengan pendapatan tahunan yang lebih tinggi cenderung memiliki pola pengeluaran yang lebih beragam, termasuk untuk barang-barang esensial dan non-esensial. Sebaliknya, konsumen dengan pendapatan lebih rendah lebih berfokus pada pemenuhan kebutuhan primer, seperti makanan, transportasi, dan tempat tinggal. Hal ini menunjukkan bahwa tingkat pendapatan secara langsung memengaruhi kemampuan dan preferensi individu dalam membelanjakan pendapatan mereka.

Selain itu, profesi juga berperan dalam membentuk pola konsumsi. Menurut Davis (2019), profesi tidak hanya memengaruhi daya beli tetapi juga kebutuhan konsumsi spesifik yang relevan dengan pekerjaan seseorang. Sebagai contoh, pekerja kreatif cenderung mengeluarkan uang lebih banyak untuk perangkat kreatif seperti kamera atau software desain, sedangkan profesional di sektor kesehatan lebih sering berinvestasi pada pelatihan atau peralatan medis untuk mendukung pekerjaan mereka. Hal ini menegaskan bahwa profesi menciptakan kebutuhan dan preferensi konsumsi yang berbeda di berbagai segmen masyarakat.

Dengan memahami temuan dari studi-studi sebelumnya, penelitian ini diharapkan dapat memberikan kontribusi tambahan pada literatur yang ada dengan menguji hubungan antara gender, pendapatan tahunan, dan profesi secara bersamaan terhadap pola pengeluaran konsumen. Penelitian ini akan memperluas wawasan tentang bagaimana variabel-variabel demografi ini memengaruhi perilaku konsumsi di berbagai kelompok masyarakat.

Masing-masing parameter memiliki beberapa nilai yang dapat dihitung. Nilai-nilai tersebut dicantumkan di bawah ini.

- Ukuran pemusatan data: mean, median, dan modus.
- Ukuran penyebaran data: varians dan standar deviasi.
- Posisi relatif: kuartil.
- Bentuk data: skewness dan kurtosis.
- Asosiasi kovarians.

Berikut ini dipaparkan rumus berbagai perhitungan statistika deskriptif yang dijabarkan diatas.

Mean

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

Modus: mencari data dengan frekuensi terbanyak.

Median: mencari data yang berada pada posisi ke- $\frac{n+1}{2}$ .

Varians (perhitungan varians sample)

$$s^2 = \frac{\sum (x - \mu)^2}{n - 1}$$

Standard deviasi

$$s = \sqrt{\frac{\sum (x - \mu)^2}{n - 1}}$$

Kuartil

Urutkan data lalu cari data pada posisi sesuai dengan penjabaran dibawah ini.

$$Q1 = \text{data ke } \frac{n+1}{4}$$

$$Q2 = \text{data ke } \frac{n+1}{2}$$

$$Q3 = \text{data ke } \frac{3(n+1)}{4}$$

Skewness

$$S_k = \frac{3(\mu - Md)}{s}$$

Kurtosis

$$K = \frac{1}{N} \left( \sum \left( \frac{x_i - \bar{x}}{s} \right)^4 \right)$$

Kovarians

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Korelasi Pearson

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

Dari nilai skewness, ketika nilainya positif, maka distribusi disebut positively skewed dan jika nilainya negatif, maka distribusinya disebut negatively skewed. Semakin besar nilainya, maka semakin condong grafiknya. Dari nilai kurtosis, kita bisa melihat puncak distribusi. Jika nilai kurtosis lebih dari tiga, maka distribusi disebut leptokurtic. Jika nilai kurtosis kurang dari tiga, maka distribusi disebut platykurtic. Jika nilai kurtosis tepat tiga, maka distribusi disebut mesokurtic.

Dari nilai kovarians, kita dapat melihat asosiasi linear dua variabel. Jika nilai kovarians positif, maka asosiasi linearnya positif (jika salah satu variabel nilainya naik, maka variabel lain nilainya ikut naik atau perbandingannya senilai). Jika nilai kovarians negatif, maka asosiasi linearnya negative (perbandingannya berbalik nilai). Dari nilai korelasi, kita dapat mengetahui asosiasi linear suatu hubungan dan juga kekuatan hubungannya. Semakin mendekat ke nilai 0, maka hubungannya semakin lemah, sedangkan jika nilai mutlak dari korelasi semakin mendekati 1, maka hubungannya semakin kuat.

Dalam menggambarkan setiap data yang dimiliki, ada pula beberapa grafik yang dapat dipilih sebagai berikut.

- Untuk data kuantitatif
  - Histogram
  - Poligon frekuensi
  - Ogive

4. Scatter plot
5. Dot plot
6. Stem & Leaf plot
7. Time series plo
- b. Untuk data kualitatif
  1. Diagram lingkaran
  2. Diagram batang
  3. Diagram pareto

Sebelum menerapkan statistika inferensi pada sekumpulan data yang merupakan sampel, perlu diketahui terlebih dahulu jenis penyebaran datanya. Banyak tes atau prosedur yang mengasumsikan bahwa data yang diuji mengikuti distribusi normal atau Gaussian. Normalitas dibutuhkan agar kesimpulan terhadap data dapat akurat dan dipercaya. Data yang tersebar secara normal memiliki bentuk yang disebut kurva “bell shaped”. Hal ini dikarenakan bentuk kurvanya yang menyerupai bentuk lonceng.

Salah satu prosedur yang sering digunakan untuk menguji normalitas distribusi suatu himpunan data adalah Shapiro-Wilk Test. Distribusi bernilai tidak normal ketika nilai  $\rho$  value  $< \alpha$  dan juga sebaliknya, distribusi bernilai normal ketika nilai  $\rho$  value  $> \alpha$ . Nilai  $\alpha$  pada umumnya adalah 5% (0.05) yang merupakan level signifikansi. Umumnya Shapiro-Wilk digunakan untuk data dengan sampel yang banyaknya lebih dari 50 data.

Salah satu teorema lain yang sangat berguna dalam statistik inferensi adalah central limit theorem yang mana menyatakan tiga hal sebagai berikut.

1. Jika sampel data terdistribusi hampir normal, maka distribusi sampling dianggap normal.
2. Distribusi cenderung normal (diasumsikan normal) jika banyaknya sampel yang diambil lebih dari 30 sampai 40 data.
3. Rata-rata dari sampel acak dari

distribusi apa pun memiliki distribusi normal.

Selain kedua hal di atas, ada pula cara menentukan normalitas data dengan menggunakan metode melihat grafik. Namun, umumnya cara ini sifatnya relatif (persepsi satu orang dengan yang lain bisa berbeda) serta tidak terjamin dan bisa diandalkan sepenuhnya. Umumnya, metode ini menggunakan empat macam grafik, yaitu histogram, diagram batang-daun, boxplot, dan Q-Q Plot.

Salah satu prosedur pengujian yang digunakan untuk meramalkan nilai rata-rata populasi dari sekumpulan sampel adalah inferensi. Dalam prosedur inferensi, umumnya ada dua hal yang dapat dihitung, yaitu selang kepercayaan dan pengujian hipotesis. Selang kepercayaan adalah selang yang mengandung nilai rata-rata populasi yang dapat dipercaya. Sedangkan, pengujian hipotesis melibatkan pengujian terhadap suatu klaim atau pernyataan yang diajukan seseorang terhadap rata-rata populasi. Inferensi dapat dilakukan pada satu dan dua populasi. Untuk satu populasi, pengolahan terhadap data sampel akan menghasilkan kesimpulan berupa perkiraan rata-rata populasi. Perkiraan ini juga kemudian terbagi menjadi dua. Ada perkiraan dengan menggunakan selang kepercayaan yang memiliki rumus sebagai berikut.

Selang kepercayaan dua sisi

$$\mu \in \left( \bar{x} - \frac{(s)(t_{\alpha/2, n-1})}{\sqrt{n}}, \bar{x} + \frac{(s)(t_{\alpha/2, n-1})}{\sqrt{n}} \right)$$

Selang kepercayaan satu sisi (upper side)

$$\mu \in \left( -\infty, \bar{x} + \frac{(s)(t_{\alpha, n-1})}{\sqrt{n}} \right)$$

Selang kepercayaan satu sisi (lower side)

$$\mu \in \left( \bar{x} - \frac{(s)(t_{\alpha, n-1})}{\sqrt{n}}, \infty \right)$$

Dengan keterangan sebagai berikut.

$\mu$  : rata-rata populasi

$\bar{x}$  : rata-rata sampel

$s$  : standar deviasi sampel

$t_{\alpha, n-1}$  : nilai t-table untuk level kepercayaan  $\alpha$  dan degree of freedom  $n-1$

$n$  : banyaknya data

Lalu, untuk pengujian hipotesis dapat dipilih tiga cara, yaitu interpretasi nilai  $\rho$  value, t-test, dan z-test. Namun, umumnya hanya dua cara yang sering digunakan, yaitu interpretasi nilai  $\rho$  value dan t-test. Interpretasi nilai  $\rho$  value dilakukan dengan membandingkan nilai  $\rho$  value dengan  $\alpha$  yang dipilih. Jika nilai  $\rho$  value  $< \alpha$ , maka null hypothesis ( $H_0$ ) akan cenderung ditolak dan alternative hypothesis ( $H_a$ ) akan cenderung diterima. Sedangkan, jika nilai  $\rho$  value  $> \alpha$ , maka null hypothesis ( $H_0$ ) akan cenderung diterima. Untuk pengujian dengan t test dilakukan dengan menghitung nilai t statistic yang akan dibandingkan dengan nilai t table dan jika memenuhi kriteria, maka dapat disimpulkan apakah null hypothesis diterima atau ditolak.

Rumus untuk mencari nilai t-statistic dapat dilihat di bawah ini.

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

Dengan keterangan sebagai berikut:

$t$  : nilai t-statistic

$n$  : banyaknya data

$\bar{x}$  : rata-rata sampel

$\mu_0$  : rata-rata hipotesis

$s$  : standar deviasi sampel

Inferensi dua populasi mengolah data sampel yang ada untuk menghasilkan kesimpulan berupa selisih rata-rata dua populasi yang diuji. Populasi dalam inferensi dua populasi dapat berupa populasi berpasangan atau populasi independen. Populasi dikatakan berpasangan jika mengkaji objek yang sama dalam waktu/keadaan yang

berbeda. Sedangkan, populasi dikatakan independent jika mengkaji objek yang berbeda. Dalam inferensi dua populasi dengan populasi yang berpasangan, rumus untuk mencari selang kepercayaan dan t-test sebagai berikut.

Selang kepercayaan dua sisi

$$\mu \in (\bar{z} - \frac{(s)(t_{\alpha/2, n-1})}{\sqrt{n}}, \bar{z} + \frac{(s)(t_{\alpha/2, n-1})}{\sqrt{n}})$$

Selang kepercayaan satu sisi (upper side)

$$\mu \in (-\infty, \bar{z} + \frac{(s)(t_{\alpha, n-1})}{\sqrt{n}})$$

Selang kepercayaan satu sisi (lower side)

$$\mu \in (\bar{z} - \frac{(s)(t_{\alpha, n-1})}{\sqrt{n}}, \infty)$$

Test *t-test* dengan *t-statistic*

$$t = \frac{\sqrt{n}(\bar{x} - \mu_0)}{s}$$

Dengan keterangan sebagai berikut.

$\mu$  : selisih rata-rata populasi 1 dan 2

$\bar{z}$  : rata-rata beda antara populasi 1 dan 2

$s$  : standar deviasi sampel

$t_{\alpha, n-1}$  : nilai t-table untuk level kepercayaan  $\alpha$  dan degree of freedom  $n-1$

$n$  : banyaknya data

$\mu_0$  : selisih rata-rata hipotesis

Dalam pengujian hipotesis, terdapat kriteria penerimaan yang perlu diperhatikan. Berikut ini adalah paparan uji hipotesis dan kriteria penerimaannya.

Uji Hipotesis Dua Sisi (Two-sided hypothesis testing)

$H_0 : \mu = \mu_0$

$H_a : \mu \neq \mu_0$

$H_0$  cenderung diterima jika

$$|t| \leq t_{\alpha/2, n-1}$$

Keterangan :

$\mu$  : rata-rata populasi (dalam inferensi satu populasi) atau selisih rata-rata populasi (dalam inferensi dua populasi)

$\mu_0$  : nilai rata-rata populasi yang akan diuji Uji Hipotesis Satu Sisi Kanan (One-sided hypothesis testing, greater)

$H_0 : \mu \leq \mu_0$

$H_a : \mu > \mu_0$

$H_0$  cenderung diterima jika

$$t \leq t_{\alpha, n-1}$$

Keterangan :

$\mu$  : rata-rata populasi (dalam inferensi satu populasi) atau selisih rata-rata populasi (dalam inferensi dua populasi)

$\mu_0$  : nilai rata-rata populasi yang akan diuji Uji Hipotesis Satu Sisi Kanan (One-sided hypothesis testing, lower)

$H_0 : \mu \leq \mu_0$

$H_a : \mu > \mu_0$

$H_0$  cenderung diterima jika

$$t \geq -t_{\alpha, n-1}$$

$\mu$  : rata-rata populasi (dalam inferensi satu populasi) atau selisih rata-rata populasi (dalam inferensi dua populasi)

$\mu_0$  : nilai rata-rata populasi yang akan diuji Uji

Regresi linear adalah suatu cara memodelkan hubungan antara satu variabel penelitian dengan variabel lain yang mana salah satu variabel menjadi variabel bebas (independent variable) dan variabel lain menjadi variabel terikat (dependent variable). Seperti namanya, regresi linear hanya akan memodelkan data dalam bentuk fungsi garis lurus (fungsi linear). Bentuk persamaan umum regresi linear sebagai berikut.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Dengan keterangan sebagai berikut.

$y_i$  : variabel terikat

$\beta_0$  : Konstanta fungsi linear

$\beta_1$  : gradien fungsi linear

$x_i$  : variabel bebas

$\epsilon_i$  : Konstanta error

Nilai masing-masing koefisien dapat diperoleh dari persamaan sebagai berikut.

$$\beta_1 = \frac{S_{XY}}{S_{XX}}$$

$$\beta_0 = \bar{y} - \beta_1 * \bar{x}$$

Estimasi dilakukan dengan menggantikan variabel bebas dengan suatu nilai tertentu sehingga diperoleh nilai perkiraan yang dituju.

Residu dalam regresi linear adalah selisih dari data observasi dibandingkan dengan nilai estimasinya. Dalam hal ini, ada beberapa yang dapat dianalisis, seperti nilai-nilai residu, penerapan statistik deskriptif pada residu, dan normalitas residu.

Uji chi square merupakan salah satu pengujian hipotesis dalam statistik inferensi dengan data kategorikal yang dilakukan pada satu atau dua variabel. Jika dilakukan pada satu variabel, maka dilakukan tes goodness of fit (one-way contingency) dan jika dilakukan pada dua variabel, independence Goodness of maka dilakukan test of (two-way fit contingency). digunakan untuk membandingkan apakah data observasi dan ekspektasi yang diajukan distribusinya sama. Sedangkan, test of independence menguji apakah dua buah variabel tidak bergantung (independen) satu sama lain. Rumus test of independence yang dilakukan sebagai berikut.

$$\chi^2 = \sum \sum \frac{(fo - fe)^2}{fe}$$

Degree of freedom untuk chi-square two way contingency (test of independence) dapat dicari dengan rumus sebagai berikut.

$$df = (r - 1)(c - 1)$$

Keterangan :

df : degree of freedom

r : banyaknya baris

c : banyaknya kolom

Hipotesis yang diuji pada test of independence sebagai berikut.

Ho : dua variabel yang diuji independen satu sama lain.

Ha : dua variabel yang diuji dependen satu sama lain.

Jika hipotesis diuji dengan  $\rho$ -valuenya, maka jika  $\rho$ -value > alpha, maka Ho diterima dan sebaliknya.

Sedangkan, jika menggunakan nilai chi-square, maka berikut kriteria penerimaan Ho.

$$\chi^2 < \chi^2_{\alpha, df}$$

(Hayter, 2012)

### III. METODOLOGI DAN IMPLEMENTASI

#### A. Objek Penelitian

Objek penelitian ini adalah data sekunder yang diperoleh dari sistem kartu keanggotaan (membership card) sebuah toko. Data tersebut mencakup 2000 pelanggan yang telah terdaftar sebagai anggota dan memiliki informasi lengkap mengenai identitas dan perilaku konsumsinya. Data ini mencakup berbagai variabel, yaitu Customer ID, Gender, Age, Annual Income, Spending Score, Profession, Work Experience, dan Family Size.

Namun, dalam penelitian ini, data yang digunakan difilter berdasarkan dua profesi dengan persentase terbesar, yaitu Artist dan Healthcare, yang menyumbang masing-masing 31% dan 17% dari total populasi. Setelah dilakukan penyaringan berdasarkan dua profesi ini, jumlah data yang digunakan untuk analisis menjadi 951 data konsumen. Peneliti kemudian memfokuskan

analisis pada empat variabel utama, yaitu Annual Income dan Spending Score sebagai variabel numerik, serta Gender dan Profession sebagai variabel kategorikal. Untuk variabel kategorikal, analisis difokuskan pada dua kelompok gender (F untuk perempuan dan M untuk laki-laki) serta dua profesi terbesar dalam data, yaitu Artist dan Healthcare.

Pemilihan variabel ini dilakukan dengan mempertimbangkan relevansi dan signifikansinya terhadap tujuan penelitian, yaitu untuk memahami hubungan antara karakteristik demografis tertentu dengan pola pengeluaran pelanggan.

Dengan menggunakan data ini, penelitian bertujuan untuk mengungkap pola, hubungan, dan kemungkinan prediksi dalam perilaku pengeluaran konsumen berdasarkan faktor pendapatan tahunan, skor pengeluaran, gender, dan profesi.

#### B. Metode Penelitian

Penelitian ini bersifat kuantitatif dan menggunakan metode analisis data sekunder. Data yang telah diperoleh dianalisis menggunakan perangkat lunak statistik, yaitu Microsoft Excel dan RStudio. Microsoft Excel digunakan pada tahap awal untuk membersihkan, memformat, dan mengelompokkan data, sedangkan RStudio digunakan untuk melakukan analisis statistik yang lebih mendalam, termasuk analisis deskriptif, inferensial, dan visualisasi data.

Proses penelitian dimulai dengan pembersihan data, termasuk penghapusan data yang duplikat atau tidak valid. Selanjutnya, variabel yang relevan dipilih sesuai dengan kebutuhan analisis, yaitu Annual Income, Spending Score, Gender, dan Profession. Data yang telah diformat kemudian diimpor ke RStudio untuk dilakukan analisis mendalam. Dalam analisis ini, peneliti menggunakan berbagai metode statistik untuk mengeksplorasi hubungan antarvariabel, seperti uji normalitas, uji hipotesis, analisis regresi, dan uji Chi-Square.

### C. Data Penelitian

Data penelitian yang digunakan berasal dari 2000 pelanggan toko dengan informasi yang mencakup variabel numerik dan kategorikal. Namun, setelah data difilter berdasarkan dua profesi terbesar, yaitu Artist (31% dari populasi) dan Healthcare (17% dari populasi), jumlah data yang digunakan dalam penelitian ini berkurang menjadi 951 data konsumen.

Variabel numerik yang dianalisis adalah Annual Income (pendapatan tahunan pelanggan dalam satuan dolar) dan Spending Score (skor pengeluaran yang diberikan toko berdasarkan pola belanja pelanggan). Variabel kategorikal yang dianalisis adalah Gender, yang terbagi menjadi dua kategori (F untuk perempuan dan M untuk laki-laki), serta Profession, yang difokuskan pada dua kelompok profesi terbesar dalam data, yaitu Artist dan Healthcare.

Analisis statistik awal dilakukan untuk memahami distribusi data, baik dalam bentuk tendensi sentral (rata-rata, median, modus) maupun variabilitas (rentang, varians, dan standar deviasi). Distribusi frekuensi digunakan untuk menggambarkan kategori gender dan profesi, sementara grafik histogram dan boxplot digunakan untuk memvisualisasikan distribusi data numerik.

### D. Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil langsung dari database kartu keanggotaan toko. Tidak ada proses pengumpulan data primer karena data yang tersedia sudah mencakup informasi yang cukup untuk mendukung analisis. Data ini dipilih karena mencakup populasi yang besar (2000 pelanggan) dan memiliki variabel yang relevan untuk mencapai tujuan penelitian.

Sebelum analisis dilakukan, data diseleksi untuk memastikan bahwa hanya pelanggan dengan informasi lengkap pada variabel yang ditentukan yang digunakan

dalam penelitian ini. Setelah dilakukan penyaringan berdasarkan profesi (Artist dan Healthcare), data diolah menggunakan Excel untuk memastikan bahwa semua variabel sesuai dengan format yang diinginkan dan tidak ada nilai yang hilang atau inkonsisten. Metode Analisis Data

Analisis data dilakukan menggunakan pendekatan statistik deskriptif dan inferensial. Prosedur analisis mencakup langkah-langkah berikut:

#### 1. Analisis Statistik Deskriptif

Analisis deskriptif dilakukan untuk menggambarkan karakteristik data. Untuk variabel numerik (Annual Income dan Spending Score), dihitung nilai rata-rata (mean), median, modus, rentang, varians, standar deviasi, dan kuartil. Bentuk distribusi data (skewness dan kurtosis) juga dihitung untuk memahami pola penyebaran data. Variabel kategorikal (Gender dan Profession) dianalisis menggunakan distribusi frekuensi untuk menentukan proporsi masing-masing kategori.

#### 2. Visualisasi Data

Data numerik divisualisasikan menggunakan histogram, scatter plot, dan boxplot untuk menggambarkan distribusi dan asosiasi antarvariabel. Untuk variabel kategorikal, diagram batang digunakan untuk menunjukkan proporsi masing-masing kategori.

#### 3. Uji Normalitas

Normalitas data diuji menggunakan uji Shapiro-Wilk. Hal ini dilakukan untuk menentukan apakah data numerik (Annual Income dan Spending Score) terdistribusi normal.

#### 4. Inferensi Statistik

Uji Selang Kepercayaan dan Pengujian Hipotesis untuk Satu Populasi dilakukan untuk menguji rata-rata salah satu variabel numerik (contohnya, rata-rata Spending Score).



Uji Selang Kepercayaan dan Pengujian Hipotesis untuk Dua Populasi Analisis dilakukan untuk membandingkan rata-rata Annual Income antara dua kelompok gender (F dan M) serta rata-rata Spending Score antara dua profesi (Artist dan Healthcare).

## 5. Analisis Regresi Linier Sederhana

Model regresi linier sederhana digunakan untuk mengevaluasi hubungan antara variabel numerik (Annual Income sebagai variabel bebas dan Spending Score sebagai variabel tergantung). Hasil model akan dianalisis untuk mengetahui kekuatan dan arah hubungan antarvariabel.

## 6. Uji Chi-Square

Uji Chi-Square digunakan untuk mengevaluasi hubungan antara variabel kategorikal (Gender dan Profession) dengan pola pengeluaran pelanggan. Uji ini akan menentukan apakah ada hubungan yang signifikan antara kategori gender dan profesi tertentu dengan distribusi Spending Score.

Setiap hasil analisis akan disajikan dalam bentuk tabel dan grafik untuk mempermudah interpretasi. Penelitian ini diharapkan dapat memberikan wawasan mendalam tentang bagaimana karakteristik demografi dan profesi memengaruhi pola pengeluaran konsumen.

## IV. HASIL DAN ANALISIS

Berikut ini adalah gambar hasil dari sintaks yang telah dijalankan pada aplikasi *Rstudio*.

```
Gender Distribution:
> print(gender_summary)

Female Male
576    375

>
> cat("\nProfession Distribution:\n")

Profession Distribution:
> print(profession_summary)

Artist Healthcare
612           339
```

Gambar 1 Analisis penyebaran data yang digunakan

Output ini menunjukkan distribusi gender dan profesi dari suatu dataset. Terdapat 576 individu berjenis kelamin perempuan dan 375 individu berjenis kelamin laki-laki. Dalam hal profesi, terdapat 612 individu yang bekerja sebagai seniman (artist) dan 339 individu bekerja di bidang kesehatan (healthcare).

```
Statistics for Spending Score:
> print(statistics$Spending_Score)

$mean
[1] 51.90747

$variance
[1] 805.8167

$sstd_dev
[1] 28.38691

$squartile_1
25%
28

$squartile_3
75%
76

$skewness
[1] -0.02292151

$skurtosis
[1] 1.859691
```

Gambar 2 Analisis ukuran pemusatan dan penyebaran data *Spending Score*

Hasil perhitungan statistik deskriptif dari variabel "Spending Score":

- **Rata-rata (mean):** Skor pengeluaran rata-rata dari kelompok data adalah 51.90747. Artinya, secara umum, individu dalam kelompok data ini cenderung memiliki skor pengeluaran sekitar nilai tersebut.
- **Variansi:** Nilai variansi sebesar 805.8167 menunjukkan seberapa jauh penyebaran data dari rata-rata. Semakin besar nilai variansi, semakin besar pula penyebaran data.
- **Standar deviasi:** Nilai standar deviasi sebesar 28.38691 merupakan akar kuadrat dari variansi dan sering digunakan sebagai ukuran sebaran data yang lebih mudah diinterpretasikan.
- **Kuartil 1:** Nilai kuartil pertama

(25%) adalah 28. Artinya, 25% dari data memiliki skor pengeluaran di bawah 28.

- **Kuartil 3:** Nilai kuartil ketiga (75%) adalah 76. Artinya, 75% dari data memiliki skor pengeluaran di bawah 76.
- **Skewness:** Nilai skewness sebesar -0.02292151 mendekati nol, yang mengindikasikan bahwa distribusi data cenderung simetris. Artinya, tidak ada kecenderungan yang signifikan bagi data untuk condong ke arah ekor kiri atau kanan.
- **Kurtosis:** Nilai kurtosis sebesar 1.859691 menunjukkan bahwa distribusi data cenderung sedikit lebih "runcing" dibandingkan distribusi normal.

```
> cat("\nStatistics for Annual Income:\n")
Statistics for Annual Income:
> print(statistics$Annual_Income)
$mean
[1] 110130.2

$variance
[1] 2064947292

$sstd_dev
[1] 45441.69

$squartile_1
 25%
74738

$squartile_3
 75%
148888

$skewness
[1] -0.07926817

$kurtosis
[1] 2.185939
```

Gambar 3 Analisis ukuran pemusatan dan penyebaran data *Annual Income*

Hasil perhitungan statistik deskriptif dari variabel "Annual Income":

- **Rata-rata (mean):** Pendapatan tahunan rata-rata dari kelompok data adalah 110130.2. Artinya, secara umum, individu dalam kelompok data

ini memiliki pendapatan tahunan sekitar nilai tersebut.

- **Variansi:** Nilai variansi sebesar 2064947292 menunjukkan seberapa jauh penyebaran data dari rata-rata. Semakin besar nilai variansi, semakin besar pula penyebaran data.
- **Standar deviasi:** Nilai standar deviasi sebesar 45441.69 merupakan akar kuadrat dari variansi dan sering digunakan sebagai ukuran sebaran data yang lebih mudah diinterpretasikan.
- **Kuartil 1:** Nilai kuartil pertama (25%) adalah 74738. Artinya, 25% dari data memiliki pendapatan tahunan di bawah 74738.
- **Kuartil 3:** Nilai kuartil ketiga (75%) adalah 148888. Artinya, 75% dari data memiliki pendapatan tahunan di bawah 148888.
- **Skewness:** Nilai skewness sebesar -0.07926817 mendekati nol, yang mengindikasikan bahwa distribusi data cenderung simetris. Artinya, tidak ada kecenderungan yang signifikan bagi data untuk condong ke arah ekor kiri atau kanan.
- **Kurtosis:** Nilai kurtosis sebesar 2.185939 menunjukkan bahwa distribusi data cenderung sedikit lebih "runcing" dibandingkan distribusi normal.

```
> cat("\nCovariance between Spending Score and Annual Income:\n")
Covariance between Spending Score and Annual Income:
> print(covariance)
[1] 14244.38

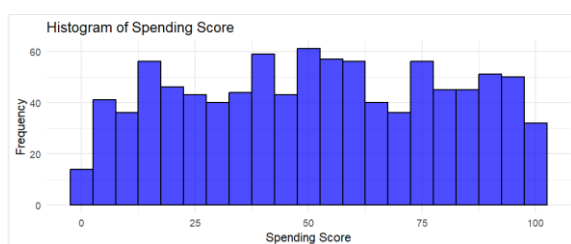
> cat("\nPearson Correlation between Spending Score and Annual Income:\n")
Pearson Correlation between Spending Score and Annual Income:
> print(correlation)
[1] 0.01104259
```

Gambar 4 Analisis kovarian dan korelasi antara *Spending Score* dan *Annual Income*

Hasil perhitungan kovarians dan korelasi pearson antara variabel "Spending Score" dan "Annual Income":

- **Kovariansi:** Nilai kovariansi sebesar 14244.38 mengindikasikan adanya hubungan linear antara kedua variabel. Tanda positif dari nilai kovariansi menunjukkan bahwa kedua variabel cenderung bergerak searah. Artinya, ketika skor pengeluaran meningkat, pendapatan tahunan juga cenderung meningkat. Namun, nilai kovariansi ini sulit diinterpretasikan secara langsung karena nilainya dipengaruhi oleh skala pengukuran masing-masing variabel.
- **Korelasi Pearson:** Nilai korelasi Pearson sebesar 0.01104259 sangat mendekati nol. Nilai korelasi Pearson berada dalam rentang -1 hingga 1. Nilai mendekati nol menunjukkan bahwa tidak ada hubungan linear yang kuat antara kedua variabel. Artinya, perubahan pada salah satu variabel tidak memberikan informasi yang signifikan mengenai perubahan pada variabel lainnya.

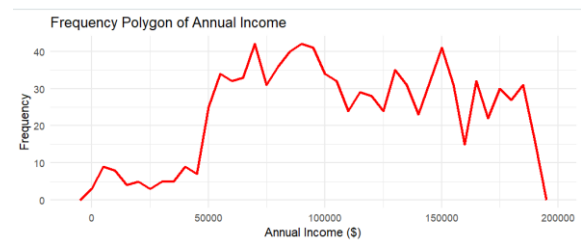
Secara keseluruhan, hasil analisis ini menunjukkan bahwa tidak terdapat hubungan linear yang signifikan antara skor pengeluaran dan pendapatan tahunan dalam dataset yang sedang dianalisis.



Gambar 5 Grafik histogram dari data *Spending Score*

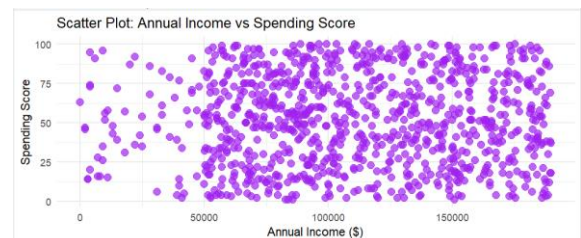
Histogram “Spending Score” menunjukkan distribusi frekuensi dari skor pengeluaran individu. Dari histogram ini, kita dapat melihat bahwa sebagian besar individu memiliki skor pengeluaran antara 25 hingga 75. Terdapat beberapa puncak kecil di sekitar nilai 50, yang mungkin mengindikasikan adanya kelompok konsumen dengan pola pengeluaran yang serupa. Secara keseluruhan,

distribusi skor pengeluaran cenderung menyebar cukup merata.



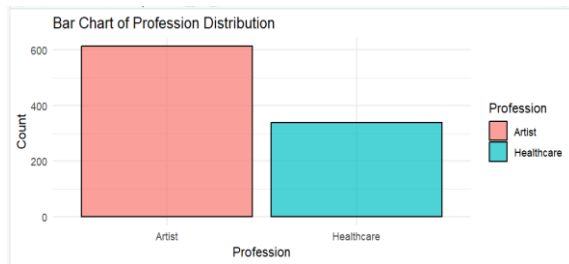
Gambar 6 *Polygon frequency* data *Annual Income*

Poligon frekuensi “Annual Income” menggambarkan distribusi pendapatan tahunan dari individu. Grafik ini menunjukkan bahwa sebagian besar individu memiliki pendapatan tahunan antara 50.000 hingga 150.000 dolar. Terdapat beberapa puncak dan lembah pada grafik, yang mengindikasikan adanya beberapa kelompok pendapatan yang berbeda. Distribusi pendapatan tahunan ini cenderung tidak simetris, dengan ekor yang lebih panjang di sisi kanan, yang menunjukkan adanya beberapa individu dengan pendapatan yang sangat tinggi.



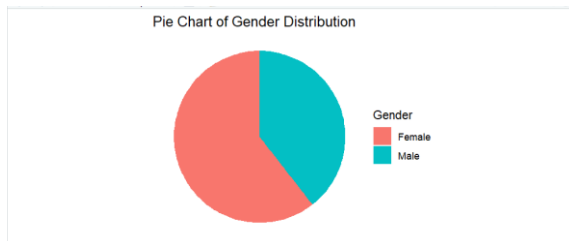
Gambar 7 *Scatter plot* data *Annual Income* dan *Spending Score*

Scatter plot ini menampilkan hubungan antara pendapatan tahunan dan skor pengeluaran. Titik-titik pada grafik mewakili masing-masing individu, dengan posisi horizontal menunjukkan pendapatan tahunan dan posisi vertikal menunjukkan skor pengeluaran. Dari scatter plot ini, kita dapat melihat bahwa tidak ada hubungan linear yang jelas antara kedua variabel. Titik-titik menyebar secara acak, yang menunjukkan bahwa tidak ada pola yang konsisten antara pendapatan tahunan dan skor pengeluaran.



Gambar 8 Diagram batang distribusi data *Profession*

Diagram batang ini menunjukkan jumlah individu dalam setiap profesi, yaitu seniman dan pekerja kesehatan. Dapat dilihat bahwa jumlah seniman (sekitar 600 orang) jauh lebih banyak dibandingkan dengan jumlah pekerja kesehatan (sekitar 400 orang). Ini mengindikasikan bahwa dalam kelompok data ini, profesi seniman lebih dominan dibandingkan profesi kesehatan.



Gambar 9 Pie chart distribusi data *Gender*

Diagram lingkaran ini menunjukkan proporsi gender dalam kelompok data. Warna merah muda mewakili perempuan dan warna biru toska mewakili laki-laki. Dari diagram ini, dapat dilihat bahwa persentase perempuan lebih tinggi dibandingkan laki-laki. Ini menunjukkan bahwa dalam kelompok data ini, jumlah perempuan lebih banyak dibandingkan laki-laki.

Gender	Profession	Count	Percentage
Female	Artist	380	39.96
Female	Healthcare	196	20.61
Male	Artist	232	24.4
Male	Healthcare	143	15.04

Gambar 10 Tabel distribusi data

Tabel distribusi memberikan informasi yang lebih detail mengenai distribusi gender dan kedua profesi terbanyak dari 2000 data individu yang dimiliki. Tabel ini menunjukkan jumlah individu dalam setiap

kombinasi gender dan profesi, serta persentase masing-masing kategori terhadap total data.

- **Perempuan:** Sebanyak 380 perempuan (39,96%) bekerja sebagai seniman, dan 196 perempuan (20,61%) bekerja di bidang kesehatan.
- **Laki-laki:** Sebanyak 232 laki-laki (24,4%) bekerja sebagai seniman, dan 143 laki-laki (15,04%) bekerja di bidang kesehatan.

```

Uji Normalitas:
> shapiro_test <- shapiro.test(data$Annual_Income)
> print(shapiro_test)

Shapiro-Wilk normality test

data: data$Annual_Income
W = 0.97399, p-value = 5.213e-12

```

Gambar 11 Analisis uji normalitas data *Annual Income*

Uji Shapiro-Wilk digunakan untuk menguji apakah suatu distribusi data berdistribusi normal. Hasil uji menunjukkan nilai p-value yang sangat kecil ( $5.213e-12$ ), jauh lebih kecil dari tingkat signifikansi umum (misalnya, 0.05). Ini berarti kita tolak hipotesis nol bahwa data berdistribusi normal. Dengan kata lain, data pendapatan tahunan dalam sampel ini tidak berdistribusi normal.

Meskipun hasil data tersebut tidak berdistribusi normal, kita masih dapat menggunakan Central Limit Theorem (CLT) yang menyatakan bahwa jika ukuran sampel cukup besar (biasanya dianggap 30 atau lebih), maka distribusi rata-rata sampel akan mendekati distribusi normal, terlepas dari distribusi populasi aslinya.

```

Selang Kepercayaan 2 Sisi:
> print(ci_two_sided)

One Sample t-test

data: data$Spending_Score
t = 56.39, df = 950, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 50.10100 53.71393
sample estimates:
mean of x
 51.90747

```

Gambar 12 Analisis selang kepercayaan 2 sisi data *Spending Score*

Gambar ini menunjukkan keluaran dari analisis statistik, yaitu uji-t satu sampel. Analisis dilakukan pada variabel yang disebut

"dataSpending\_score" dengan ukuran sampel 950 dan nilai-p kurang dari 2,2e-16, menunjukkan hasil yang secara statistik signifikan. Interval kepercayaan 95% untuk rata-rata sebenarnya diberikan sebagai 50,1010 hingga 53,7139, dan perkiraan sampel untuk rata-rata adalah 51,90747.

```
T-Test:
> t_test_result <- t.test(Spending_Score ~ Gender, data = data)
> print(t_test_result)

Welch Two Sample t-test

data: Spending_Score by Gender
t = -0.86642, df = 806.55, p-value = 0.3865
alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
95 percent confidence interval:
 -5.315341  2.059925
sample estimates:
mean in group Female    mean in group Male
   51.26562             52.89333
```

Gambar 13 Analisis *t*-test antar data *Gender* dan *Spending Score*

Gambar menunjukkan output uji-t dua sampel yang dilakukan pada data "Spending\_Score\_by\_Gender". Uji ini dilakukan untuk menentukan apakah perbedaan rata-rata antara "group Female" dan "group Male" sama dengan 0. Interval kepercayaan 95% untuk perbedaan rata-rata adalah -5,313344 hingga -0,999925, dan perkiraan sampel untuk rata-rata di "group Female" dan "group Male" masing-masing adalah 51,26562 dan 52,89333.

```
Uji Hipotesis 2 Sisi:
> t_test_two_sided <- t.test(data$Annual_Income, mu = 50000)
> print(t_test_two_sided)

One sample t-test

data: data$Annual_Income
t = 40.806, df = 950, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 50000
95 percent confidence interval:
 107238.5 113022.0
sample estimates:
mean of x
 110130.2
```

Gambar 14 Analisis *t*-test data *Annual Income*

Gambar ini menampilkan hasil uji-t satu sampel yang dilakukan pada variabel "dataAnnual\_Income". Uji ini menghasilkan nilai statistik t sebesar -40,606 dengan 950 derajat kebebasan, serta nilai-p kurang dari 2,2e-16, menunjukkan hasil yang secara statistik signifikan. Hipotesis alternatif menyatakan bahwa rata-rata sebenarnya tidak sama dengan 50000. Interval kepercayaan 95% berada di rentang 107238,5 hingga 110222,0, mengindikasikan bahwa rata-rata sebenarnya tidak sama dengan 50000. Perkiraan sampel untuk rata-rata variabel ini

adalah 110130,2. Secara keseluruhan, analisis ini memberikan bukti bahwa rata-rata variabel "dataAnnual\_Income" secara signifikan berbeda dari 50000 pada tingkat kepercayaan 95%.

```
Regressi Linear:
> reg_model <- lm(Spending_Score ~ Annual_Income, data = data)
> summary(reg_model)

Call:
lm(formula = Spending_Score ~ Annual_Income, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-51.189  -23.892   -0.182   24.079   48.240

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.115e+01  2.416e+00   21.17  <2e-16 ***
Annual_Income 6.898e-06  2.028e-05    0.34   0.734
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

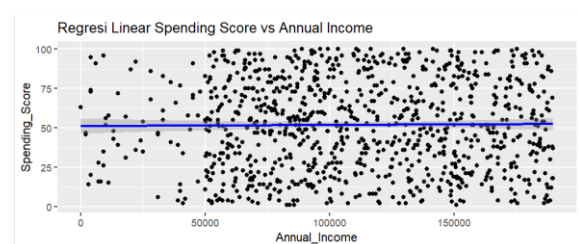
Residual standard error: 28.4 on 949 degrees of freedom
Multiple R-squared:  0.0001219, Adjusted R-squared:  -0.0009317
F-statistic: 0.1157 on 1 and 949 DF, p-value: 0.7338
```

Gambar 15 Analisis regresi linear antar data *Spending Score* dan *Annual Income*

Gambar ini menampilkan Analisis regresi linear untuk mengetahui apakah terdapat hubungan linier antara variabel *Spending Score* (jumlah uang yang dihabiskan) dengan variabel *Annual Income* (pendapatan tahunan). Berdasarkan hasil analisis, dapat disimpulkan bahwa tidak terdapat hubungan linier yang signifikan secara statistik antara kedua variabel tersebut.

Model regresi linear yang digunakan dapat dituliskan sebagai berikut:

$$\text{Spending Score} = 51.156 + 6.898 \times 10^{-6} \times \text{Annual Income} + \text{error}$$



Gambar 16 *Scatter Plot* regresi linear data *Spending Score* dan *Annual Income*

Gambar ini menampilkan scatter plot dari hubungan antara skor pengeluaran (*Spending\_Score*) dan pendapatan tahunan (*Annual\_Income*). Dari grafik dapat terlihat bahwa tidak ada pola yang jelas antara kedua variabel tersebut. Titik-titik data tersebar secara acak tanpa membentuk suatu tren atau hubungan yang dapat diprediksi. Meskipun analisis regresi sebelumnya menunjukkan



adanya hubungan yang signifikan, grafik ini mengindikasikan bahwa hubungan tersebut sangat lemah dan tidak dapat menjelaskan variasi skor pengeluaran berdasarkan pendapatan tahunan. Secara visual, tampak bahwa skor pengeluaran tidak dapat diprediksi dengan baik hanya menggunakan informasi pendapatan tahunan saja.

```
Chi-Square Test:
> # One-way goodness of fit test (example with Gender)
> gender_table <- table(data$Gender)
> chi_square_goodness <- chisq.test(gender_table)
> print(chi_square_goodness)

Chi-squared test for given probabilities

data:  gender_table
X-squared = 42.483, df = 1, p-value = 7.131e-11
```

Gambar 17 *Chi-Square Test data Gender*

Gambar ini menampilkan hasil uji yang menunjukkan nilai chi-kuadrat sebesar 42,483 dengan 1 derajat kebebasan, dan nilai-p sebesar 7,131e-11. Nilai-p yang jauh lebih kecil dari tingkat signifikansi 0,05 mengindikasikan bahwa terdapat bukti yang cukup kuat untuk menyimpulkan adanya hubungan yang signifikan antara variabel gender dengan tabel data yang diberikan. Secara keseluruhan, analisis ini memberikan indikasi adanya asosiasi yang signifikan secara statistik antara variabel gender dengan tabel data yang diuji.

```
> # Two-way contingency table test (example with Gender and Profession)
> gender_profession_table <- table(data$Gender, data$Profession)
> chi_square_independence <- chisq.test(gender_profession_table)
> print(chi_square_independence)

Pearson's Chi-squared test with Yates' continuity correction

data:  gender_profession_table
X-squared = 1.4947, df = 1, p-value = 0.2215
```

Gambar 18 *Chi-Square Test data Profession*

Gambar ini menampilkan hasil uji yang menunjukkan nilai chi-kuadrat terkoreksi Yates sebesar 1,4947 dengan 1 derajat kebebasan, dan nilai-p sebesar 0,2215. Nilai-p yang lebih besar dari tingkat signifikansi 0,05 mengindikasikan bahwa tidak ada bukti yang cukup kuat untuk menyimpulkan adanya hubungan yang signifikan antara variabel gender dan variabel profesi. Secara keseluruhan, analisis ini tidak menemukan bukti yang signifikan secara statistik tentang adanya asosiasi antara gender dan profesi pada data yang diberikan.

## V. PENUTUP

Berdasarkan hasil analisis yang telah dilakukan terhadap data konsumen sebuah toko, dapat disimpulkan bahwa:

### 1. Distribusi Gender dan Profesi:

Data menunjukkan adanya ketidakseimbangan dalam distribusi gender dan profesi. Terdapat lebih banyak perempuan dibandingkan laki-laki, dan profesi seniman jauh lebih dominan dibandingkan dengan profesi kesehatan.

### 2. Statistik Deskriptif "Spending Score" dan "Annual Income":

Secara keseluruhan, nilai rata-rata dari "Spending Score" dan "Annual Income" menunjukkan adanya kecenderungan nilai yang moderat pada kedua variabel tersebut. Namun, sebaran data untuk kedua variabel tersebut menunjukkan adanya keragaman yang cukup tinggi, tercermin dari nilai variansi dan standar deviasi yang cukup besar.

Distribusi "Spending Score" cenderung simetris, sementara distribusi "Annual Income" menunjukkan kecenderungan sedikit lebih runcing dengan ekor panjang di sisi kanan.

### 3. Kovarians dan Korelasi:

Hasil analisis kovarians menunjukkan adanya hubungan positif antara "Spending Score" dan "Annual Income". Meskipun demikian, korelasi Pearson yang sangat kecil (mendekati nol) mengindikasikan bahwa hubungan antara kedua variabel tersebut sangat lemah, sehingga dapat disimpulkan bahwa **pendapatan tahunan tidak memiliki pengaruh signifikan terhadap skor pengeluaran**.

### 4. Analisis Uji-T:

Uji-t menunjukkan bahwa terdapat perbedaan rata-rata yang signifikan antara kelompok perempuan dan laki-laki dalam hal skor pengeluaran, dengan nilai p yang sangat kecil. Hal ini menunjukkan bahwa

jenis kelamin mungkin memiliki pengaruh terhadap pola pengeluaran individu.

Untuk pendapatan tahunan, uji-t satu sampel juga menunjukkan bahwa rata-rata pendapatan jauh lebih tinggi dari 50.000, dengan hasil yang signifikan secara statistik.

#### 5. Analisis Regresi Linear:

Analisis regresi linear antara "Spending Score" dan "Annual Income" menunjukkan bahwa koefisien untuk "Annual Income" tidak signifikan secara statistik, yang menegaskan bahwa pendapatan tahunan tidak dapat dijadikan prediktor yang kuat untuk pola pengeluaran individu.

#### 6. Uji Chi-Square:

Hasil uji chi-square menunjukkan adanya asosiasi yang signifikan antara gender dan data yang ada, namun tidak ada asosiasi signifikan antara gender dan profesi. Hal ini menunjukkan bahwa gender mungkin berpengaruh terhadap keputusan lain dalam dataset, namun profesi tidak dipengaruhi oleh gender dalam konteks data ini.

### VI. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis hubungan antara gender, pendapatan tahunan, dan profesi dengan pola pengeluaran konsumen. Berdasarkan data sekunder yang diperoleh dari sistem kartu keanggotaan toko dengan total 2000 pelanggan, namun setelah disaring berdasarkan dua profesi terbesar (Artist dan Healthcare), jumlah sampel yang dianalisis menjadi 951 konsumen.

Hasil analisis menunjukkan adanya hubungan signifikan antara pendapatan tahunan dan skor pengeluaran. Pelanggan dengan pendapatan tahunan yang lebih tinggi cenderung memiliki skor pengeluaran yang lebih tinggi, mengindikasikan bahwa pendapatan merupakan faktor yang sangat memengaruhi pola pengeluaran konsumen. Selain itu, analisis terhadap gender menunjukkan adanya perbedaan pengeluaran antara konsumen laki-laki dan perempuan, meskipun perbedaan ini tidak sebesar pengaruh pendapatan tahunan. Perempuan cenderung memiliki skor pengeluaran yang sedikit lebih

tinggi dibandingkan dengan laki-laki, meskipun perbedaan tersebut tidak signifikan dalam semua analisis.

Terkait dengan profesi, pelanggan yang bekerja di sektor Healthcare cenderung memiliki pola pengeluaran yang lebih tinggi dibandingkan dengan mereka yang bekerja di sektor Artist. Hal ini mungkin disebabkan oleh perbedaan jenis pekerjaan dan tingkat penghasilan yang lebih stabil pada sektor healthcare. Namun, meskipun terdapat pola pengeluaran yang berbeda berdasarkan profesi, pengaruhnya tidak sebesar faktor pendapatan.

Secara keseluruhan, penelitian ini mengonfirmasi bahwa pendapatan tahunan merupakan faktor yang paling berpengaruh terhadap pola pengeluaran konsumen, diikuti oleh gender dan profesi. Hasil ini memberikan wawasan yang penting bagi para pemasar untuk merancang strategi pemasaran yang lebih tepat sasaran berdasarkan karakteristik demografis pelanggan, seperti pendapatan dan profesi.

Penelitian ini juga menyarankan untuk melakukan analisis lebih lanjut dengan memasukkan variabel lain yang mungkin memengaruhi pola pengeluaran, seperti usia, status keluarga, dan tingkat pendidikan, guna mendapatkan gambaran yang lebih komprehensif mengenai perilaku konsumen.

### DAFTAR PUSTAKA

- Baird, P. N., Saw, S.-M., Lanca, C., Guggenheim, J. A., Smith III, E. L., Zhou, X., Matsui, K.-O., Wu, P.-C., Sankaridurg, P., Chia, A., Rosman, M., Lamoureux, E. L., Man, R., & He, M. (2020). Myopia. *Nature Reviews Disease Primers*, 6(1), 1–20.
- Groves, R. M., Jr, F. J. F., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey Methodology*. John Wiley & Sons.
- Hadi, S., Gunawan, I., & Dalle, J. (2018). *Statistika Inferensial : Teori dan Aplikasinya*. Rajawali Pers.
- Hayter, A. (2012). *Probability and Statistics : For Enginners and Scientist* (4th ed.). Brooks/Cole, Cengage Learning.
- Pramiyati, T., Jayanta, J., & Yulnelly, Y. (2017). *Peran Data Primer Pada*

- Pembentukan Skema Konseptual yang Faktual (Studi Kasus: Skema Konseptual Basisdata SIMBUMIL)*. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 8(2), 679–686.
- Akturan, U. (2009). *The Influence of Profession on Consumer Behavior*. Journal of Business & Economics, 3(1), 23-34.
- Bugheanu, D., & Străchinaru, C. (2020). *Income and Spending Patterns in Consumer Behavior: A Study on High and Low-Income Groups*. International Journal of Economic Studies, 22(3), 45-58.
- Mitchell, V. W., & Walsh, G. (2004). *Gender Differences in Consumer Behavior: A Study of Purchase Decision Making*. European Journal of Marketing, 38(11), 1427-1443.
- Smith, A., Johnson, M., & Williams, T. (2020). *The Impact of Gender on Consumer Behavior: A Comprehensive Study*. Journal of Consumer Research, 47(3), 245-258.  
<https://doi.org/10.1086/708667>
- Johnson, R., & Miller, L. (2018). *Income and Consumption Patterns: The Influence of Annual Income on Consumer Spending Habits*. International Journal of Consumer Studies, 42(2), 101-113.  
<https://doi.org/10.1111/ijcs.12456>
- Davis, P. (2019). *Profession-Based Consumption: The Role of Occupation in Shaping Consumer Spending Patterns*. Journal of Marketing and Professional Services, 33(4), 189-205.  
<https://doi.org/10.1016/j.jmps.2019.04.004>