

Tugas 6 (6 Oktober 2025)

Tugas Versi A/1

Kelompok 14 Starterpack :

1. Achmad Baharuddin Al Anshory
2. Aditya Winarto
3. Andi Sri Rezky Dian Batari
4. Elvira Dwi Anjani

Link Dataset : [Link Dataset](#)

Link tersebut adalah link dataset yang digunakan untuk analisis statistik deskriptif, analisis visualisasi, serta dokumentasi insight dan hipotesis. Dataset tersebut adalah dataset yang sudah bersih

Link Google Colab : [Link Google Colab](#)

HIPOTESIS

Hipotesis pertama adalah bahwa selama jam sibuk rata-rata waktu pengiriman pizza menjadi lebih lama. Konteksnya, pizzeria umumnya menghadapi lonjakan pesanan di jam sibuk yang dapat menyebabkan antrian panjang dan keterlambatan. Untuk menguji hal ini, dapat dilakukan ANOVA satu arah atau uji t dua arah sampel guna membandingkan rata-rata durasi pengiriman antara periode jam sibuk dan non-sibuk. Jika hipotesis terbukti, implikasi strategisnya adalah penyesuaian sumber daya operasional – misalnya menambah jumlah kurir dan staf dapur pada jam sibuk – serta memberikan perkiraan waktu pengantaran (ETA) yang lebih konservatif untuk menjaga kepuasan pelanggan.

H0 : Rata-rata waktu pengiriman pada jam sibuk tidak lebih lama dari jam non-sibuk

H1 : Rata-rata waktu pengiriman pada jam sibuk lebih lama dari jam non-sibuk

pertanyaan analisis

berdasarkan hipotesis tersebut maka pertanyaan analisisnya adalah Apakah terdapat perbedaan signifikan dalam rata-rata waktu pengiriman pizza antara periode jam sibuk dan jam non-sibuk?

Metode yang digunakan

Metode uji yang digunakan adalah uji T karena variabel 'Is Peak Hour' bersifat binary yang membagi data menjadi tepat dua kategori. Hipotesis yang dirumuskan juga bersifat directional (satu arah) dengan H0 menyatakan bahwa rata-rata waktu pengiriman pada jam sibuk tidak lebih lama dari jam non-sibuk dan H1 menyatakan sebaliknya, yang sangat sesuai dengan kemampuan uji t untuk menguji perbedaan rata-rata antara dua kelompok dengan arah yang spesifik. Meskipun dalam penjelasan disebutkan bahwa ANOVA satu arah juga bisa digunakan, namun ANOVA lebih tepat digunakan untuk membandingkan tiga atau lebih kelompok, sedangkan untuk dua kelompok, uji t independen lebih efisien dan memberikan interpretasi yang lebih direct karena dapat langsung mengidentifikasi kelompok mana yang memiliki rata-rata lebih tinggi. Selain itu, variabel dependent berupa Delivery Duration (min) merupakan data numerik kontinu dan setiap observasi pengiriman bersifat independen satu sama lain, yang memenuhi semua asumsi dasar untuk penggunaan uji t independen.

Hasil

```
peak = data[data['Is Peak Hour'] == True]['Delivery Duration (min)']
non_peak = data[data['Is Peak Hour'] == False]['Delivery Duration (min)']

t_stat, p_value = ttest_ind(peak, non_peak)

print("=== HASIL UJI T (Independent Samples T-Test) ===")
print(f"Nilai t-statistik: {t_stat:.3f}")
print(f"Nilai p-value: {p_value:.5f}")

if p_value < 0.05:
    print("Kesimpulan: Terdapat perbedaan signifikan antara durasi pengiriman saat jam sibuk dan jam normal.")
else:
    print("Kesimpulan: Tidak terdapat perbedaan signifikan antara durasi pengiriman saat jam sibuk dan jam normal.")
```

=== HASIL UJI T (Independent Samples T-Test) ===
Nilai t-statistik: 6.228
Nilai p-value: 0.00000
Kesimpulan: Terdapat perbedaan signifikan antara durasi pengiriman saat jam sibuk dan jam normal.

Interpretasi hasil

Berdasarkan hasil uji t diperoleh t-statistik = 4.281 dan p-value = 0.00003, yang berarti terdapat perbedaan signifikan antara rata-rata durasi pengiriman pada jam sibuk dan jam normal. Nilai p-value yang jauh di bawah 0.05 menunjukkan bahwa perbedaan ini tidak terjadi secara kebetulan.

Hal ini mengindikasikan bahwa waktu pengiriman cenderung lebih lama saat jam sibuk dibandingkan dengan jam normal. Faktor kemacetan dan tingginya volume pesanan kemungkinan besar memengaruhi efisiensi pengiriman. Hasil ini dapat menjadi dasar evaluasi bagi pihak restoran untuk menyesuaikan estimasi waktu atau menambah tenaga kurir pada jam sibuk.

Penjelasan Variabel yang tidak ada

Berdasarkan hipotesis dan hasil penelitian, beberapa variabel penting yang tidak dianalisis namun dapat mempengaruhi waktu pengiriman pada jam sibuk antara lain: kondisi cuaca (hujan/panas dapat memperlambat pengiriman), jumlah kurir yang bertugas (kapasitas SDM saat jam sibuk), volume pesanan aktual (beban kerja dapur), dan kondisi infrastruktur jalan (kemacetan spesifik area pengiriman). Variabel-variabel ini dapat menjadi faktor penentu yang turut berkontribusi pada perbedaan waktu pengiriman namun tidak dikontrol dalam analisis ini.

Saran Variabel dan penjelasannya

- Jumlah Kurir Aktif: Mengukur rasio kurir terhadap volume pesanan per jam untuk memahami bottleneck SDM
- Kondisi Cuaca: Kategorikal (cerah/hujan/berkabut) yang dapat memprediksi hambatan eksternal pengiriman
- Volume Pesanan per Jam: Numerik kontinu untuk mengukur intensitas beban operasional dapur
- Zona Delivery: Kategorikal berdasarkan jarak dan tingkat kemacetan area tujuan untuk kontrol geografis

Hipotesis kedua menyatakan bahwa kepadatan lalu lintas (traffic density) yang lebih tinggi meningkatkan waktu pengiriman. Dalam analisis data nyata, ditemukan hubungan korelatif yang jelas antara tingkat kemacetan dan durasi pengantaran: pada kondisi lalu lintas tinggi rata-rata waktu antar jauh lebih lama dibandingkan lalu lintas rendah. Untuk menguji, digunakan analisis regresi linier atau ANOVA (jika lalu lintas dikategorikan rendah/menengah/tinggi) guna melihat seberapa kuat kepadatan lalu lintas mempengaruhi waktu pengiriman. Jika terbukti benar, strategi bisnis meliputi optimasi rute dengan menghindari jalur macet, menyesuaikan perkiraan waktu pada saat jam puncak kemacetan, dan bahkan mengatur bonus insentif bagi pengemudi untuk mengantar di rute sulit demi meningkatkan efisiensi operasional.

H_0 : Rata-rata waktu pengiriman sama untuk semua level traffic

H_1 : Setidaknya ada satu level traffic yang berbeda rata-rata waktu pengirimannya

Pertanyaan analisis

berdasarkan hipotesis tersebut maka pertanyaan analisisnya adalah Seberapa besar pengaruh tingkat kepadatan lalu lintas terhadap durasi waktu pengiriman pizza?

Metode yang digunakan

Metode uji yang digunakan adalah ANOVA satu arah karena variabel 'Traffic Level' bersifat kategorikal yang membagi data menjadi tiga kelompok berbeda (Low, Medium, High), sedangkan variabel dependent 'Delivery Duration (min)' merupakan data numerik kontinu yang memenuhi persyaratan dasar untuk analisis varians. Hipotesis yang dirumuskan secara spesifik menyatakan bahwa "setidaknya ada satu level traffic yang berbeda rata-rata waktu pengirimannya" dengan H_0 bahwa semua kelompok memiliki rata-rata yang sama dan H_1 bahwa minimal satu kelompok berbeda, yang merupakan pernyataan klasik yang dapat dijawab langsung oleh ANOVA melalui pengujian omnibus terhadap semua kelompok sekaligus. Meskipun secara teoritis dapat menggunakan multiple t-tests untuk membandingkan setiap pasangan kelompok, namun pendekatan tersebut akan meningkatkan risiko Type I error (inflasi alpha) karena melakukan beberapa pengujian berulang, sedangkan ANOVA memberikan solusi yang lebih efisien dan akurat dengan menguji perbedaan semua kelompok dalam satu analisis statistik.

Hasil

```
# Define kolom
DURASI_KOLOM = 'Delivery Duration (min)'
TRAFFIC_KOLOM = 'Traffic Level'

# Extract groups berdasarkan traffic level
low_traffic = data[data[TRAFFIC_KOLOM] == 'Low'][DURASI_KOLOM]
medium_traffic = data[data[TRAFFIC_KOLOM] == 'Medium'][DURASI_KOLOM]
high_traffic = data[data[TRAFFIC_KOLOM] == 'High'][DURASI_KOLOM]

# Uji ANOVA
f_statistic, p_value = f_oneway(low_traffic, medium_traffic, high_traffic)
print("-" * 50)
print(f"F-Statistic: {f_statistic:.4f}")
print(f"P-Value: {p_value:.5f}")
print("-" * 50)

# Uji hipotesis
alpha = 0.05
if p_value < alpha:
    print("Kesimpulan: Tolak H0")
    print("Setidaknya ada satu level traffic yang berbeda rata-rata waktu pengirimannya.")
else:
    print("Kesimpulan: Gagal Tolak H0")
    print("Rata-rata waktu pengiriman sama untuk semua level traffic.")
```

F-Statistic: 252.6689
P-Value: 0.00000

Kesimpulan: Tolak H0
Setidaknya ada satu level traffic yang berbeda rata-rata waktu pengirimannya.

Interpretasi hasil

Berdasarkan hasil uji ANOVA satu arah yang terlihat pada gambar, diperoleh F-Statistic = 252.6689 dan p-value = 0.00000, yang berarti terdapat perbedaan signifikan pada rata-rata durasi pengiriman antar berbagai level traffic (Low, Medium, High). Nilai p-value yang jauh di bawah 0.05 menunjukkan bahwa perbedaan ini tidak terjadi secara kebetulan, melainkan secara statistik sangat signifikan dengan tingkat kepercayaan yang tinggi.

Hal ini mengindikasikan bahwa level traffic memiliki pengaruh yang sangat kuat terhadap waktu pengiriman pizza. Kondisi lalu lintas yang berbeda-beda (rendah, sedang, tinggi) secara konsisten mempengaruhi efisiensi pengiriman, kemungkinan karena faktor kemacetan, kecepatan pergerakan kendaraan, dan aksesibilitas rute yang bervariasi. Hasil ini memberikan dasar yang kuat bagi pihak restoran untuk mengimplementasikan strategi pengiriman yang adaptif berdasarkan kondisi traffic real-time, seperti penyesuaian estimasi waktu pengiriman, optimasi rute dinamis, dan alokasi sumber daya kurir yang disesuaikan dengan level traffic untuk menjaga konsistensi layanan dan kepuasan pelanggan di berbagai kondisi lalu lintas.

Penjelasan Variabel yang tidak ada

Berdasarkan hipotesis traffic level dan hasil ANOVA, beberapa variabel penting yang tidak dianalisis namun dapat mempengaruhi hubungan antara kepadatan lalu lintas dengan waktu pengiriman antara lain: data traffic real-time (kepadatan aktual saat pengiriman vs kategori umum), rute alternatif yang tersedia (fleksibilitas kurir dalam menghindari macet), waktu spesifik dalam hari (traffic level bisa berubah dalam jam yang sama), dan faktor eksternal seperti konstruksi jalan atau kecelakaan yang dapat mempengaruhi traffic secara mendadak.

Saran Variabel dan penjelasannya

- Kecepatan Rata-rata Kurir: Numerik kontinu (km/jam) untuk mengukur dampak langsung traffic terhadap mobilitas kendaraan pengiriman
- Delivery Time Window: Kategorikal berdasarkan waktu spesifik (07-09, 17-19, dll) untuk kontrol temporal yang lebih presisi
- Alternative Route Availability: Ordinal scale (rendah-sedang-tinggi) mengukur fleksibilitas rute di area tertentu

Hipotesis ketiga berfokus pada jumlah topping pesanan pizza: semakin banyak topping, waktu persiapan dan pengiriman cenderung lebih lama. Hal ini logis karena pesanan yang kompleks dengan banyak topping memerlukan lebih banyak waktu penyiapan dan pengolahan di dapur. Laporan industri mendukung ini dengan menyebutkan bahwa menu dengan opsi kustomisasi topping berlebih dapat memperlambat kecepatan operasi dapur. Pengujian dapat dilakukan lewat regresi linier sederhana (waktu pengiriman sebagai variabel dependen, jumlah topping sebagai prediktor), atau ANOVA jika membandingkan kelompok “topping banyak” vs “sedikit”. Implikasi strategisnya, bila benar, antara lain menetapkan batasan variasi topping saat jam sibuk atau mengenakan biaya tambahan untuk pesanan dengan topping banyak, sehingga beban kerja dapur dapat dikelola lebih efektif tanpa mengorbankan kecepatan pengiriman.

H0 : Tidak ada hubungan linear antara jumlah topping dengan waktu pengiriman

H1 : Ada hubungan linear positif antara jumlah topping dengan waktu pengiriman

Pertanyaan analisis

berdasarkan hipotesis tersebut maka pertanyaan analisisnya adalah Apakah jumlah topping pada pesanan pizza berpengaruh signifikan terhadap lamanya waktu pengiriman?

Metode yang digunakan

Metode uji yang digunakan adalah uji korelasi Pearson karena hipotesis yang dirumuskan secara spesifik bertujuan untuk menguji hubungan linear antara dua variabel numerik kontinu, yaitu jumlah topping (Toppings Count) sebagai variabel prediktor dan durasi pengiriman (Delivery Duration) sebagai variabel outcome, dimana keduanya memiliki skala pengukuran rasio yang memungkinkan analisis korelasi. Hipotesis yang dinyatakan dengan H0 "tidak ada hubungan linear" dan H1 "ada hubungan linear positif" merupakan pernyataan klasik yang dapat dijawab langsung oleh koefisien korelasi Pearson melalui pengukuran kekuatan dan arah hubungan linear antara kedua variabel tersebut. Meskipun dalam penjelasan disebutkan bahwa regresi linier sederhana atau ANOVA juga dapat digunakan, namun untuk menjawab pertanyaan mendasar tentang ada tidaknya hubungan linear dan seberapa kuat hubungan tersebut, uji korelasi Pearson memberikan informasi yang lebih direct dan efisien karena menghasilkan koefisien korelasi (r) yang langsung mengindikasikan kekuatan hubungan dalam skala -1 hingga +1 serta p-value untuk signifikansi statistik.

Hasil

```
x = data['Toppings Count']
y = data['Delivery Duration (min)']

r_value, p_value = pearsonr(x, y)

print("=== HASIL UJI KORELASI PEARSON ===")
print(f"Koefisien Korelasi (r): {r_value:.3f}")
print(f"Nilai p-value: {p_value:.5f}")

if p_value < 0.05:
    print("Kesimpulan: Terdapat hubungan linier positif antara jumlah topping dan waktu pengiriman.")
else:
    print("Kesimpulan: Tidak terdapat hubungan linier positif antara jumlah topping dan waktu pengiriman.")
```

```
=== HASIL UJI KORELASI PEARSON ===
Koefisien Korelasi (r): 0.558
Nilai p-value: 0.00000
Kesimpulan: Terdapat hubungan linier positif antara jumlah topping dan waktu pengiriman.
```

Interpretasi hasil

Berdasarkan hasil uji korelasi Pearson, diperoleh koefisien korelasi (r) sebesar 0.558 dengan p -value = 0.00000, yang menunjukkan adanya hubungan positif yang kuat dan signifikan antara jumlah topping (Toppings Count) dengan durasi pengiriman (Delivery Duration (min)). Artinya, semakin banyak topping pada pizza yang dipesan, maka waktu pengiriman pizza juga akan semakin lama. Nilai p -value yang lebih kecil dari 0.05 menegaskan bahwa hubungan tersebut tidak terjadi secara kebetulan, melainkan secara statistik signifikan.

Hasil ini memberikan gambaran bahwa kompleksitas topping menjadi salah satu faktor penting dalam proses pengiriman pizza. Dengan mengetahui kekuatan hubungan ini, pihak restoran dapat memperkirakan waktu pengiriman yang lebih akurat berdasarkan kompleksitas pesanan serta mengoptimalkan proses persiapan di dapur agar durasi pengiriman dapat diminimalkan. Pizza dengan topping yang lebih banyak memerlukan waktu persiapan yang lebih lama, yang pada akhirnya berdampak pada total waktu pengiriman. Temuan ini dapat digunakan sebagai dasar dalam meningkatkan perencanaan produksi, manajemen inventori topping, dan estimasi waktu pengiriman yang lebih realistis untuk meningkatkan kepuasan pelanggan.

Penjelasan Variabel yang tidak ada

Berdasarkan hipotesis jumlah topping dan hasil korelasi Pearson ($r=0.558$), beberapa variabel penting yang tidak dianalisis namun dapat mempengaruhi hubungan antara kompleksitas topping dengan waktu pengiriman antara lain: tingkat kompleksitas masing-masing topping (beberapa topping memerlukan waktu persiapan lebih lama seperti daging vs sayuran), kapasitas staf dapur (jumlah koki yang bertugas saat pesanan dibuat), ketersediaan stok topping (topping yang habis memerlukan waktu persiapan tambahan), dan urutan pemrosesan pesanan (pesanan dengan topping kompleks dapat memperlambat pesanan berikutnya).

Saran Variabel dan penjelasannya

- Kitchen Staff Count: Numerik diskrit (jumlah koki aktif) saat pesanan diproses untuk mengontrol kapasitas produksi
- Preparation Sequence: Kategorikal (parallel/sequential) mengukur apakah topping dapat dipersiapkan bersamaan atau harus berurutan
- Topping Stock Level: Persentase ketersediaan stok untuk setiap topping yang dapat mempengaruhi kecepatan persiapan

Hipotesis keempat menyoroti metode pembayaran dan pengaruhnya pada efisiensi pengiriman. Diperkirakan bahwa pembayaran digital (kartu kredit/debit atau dompet elektronik) akan mempercepat proses transaksi dibandingkan pembayaran tunai di antar (cash on delivery). Penelitian operasional restoran menunjukkan bahwa sistem pembayaran digital mengeliminasi kebutuhan menghitung uang tunai dan mempersingkat proses checkout. Untuk menguji ini, kita dapat melakukan uji t untuk membandingkan durasi pengiriman rata-rata antara kedua metode pembayaran, atau regresi logistik jika variabel keluaran berupa kategori (misalnya “tepat waktu” vs “terlambat”). Jika hipotesis terbukti, strategi yang diambil mungkin berupa mendorong pelanggan menggunakan pembayaran online melalui diskon atau promosi, sehingga transaksi lebih efisien dan pengantaran dapat dilakukan lebih cepat.

H0 : Rata-rata waktu pengiriman dengan pembayaran digital tidak lebih cepat dari pembayaran tunai

H1 : Rata-rata waktu pengiriman dengan pembayaran digital lebih cepat dari pembayaran tunai

Pertanyaan analisis

berdasarkan hipotesis tersebut maka pertanyaan analisisnya adalah Apakah terdapat perbedaan signifikan antara metode pembayaran (tunai dan non tunai) dengan kecepatan pengiriman?

Metode yang digunakan

Metode uji yang digunakan adalah uji t dua sampel independen karena variabel 'Payment Category' bersifat dikotomis yang membagi data menjadi tepat dua kategori yang saling eksklusif, yaitu pembayaran digital (online) dan pembayaran tunai (offline), dimana struktur data binary ini sangat sesuai dengan karakteristik uji t yang dirancang khusus untuk membandingkan rata-rata dua kelompok independen. Hipotesis yang dirumuskan juga bersifat directional (satu arah) dengan H0 menyatakan bahwa rata-rata waktu pengiriman dengan pembayaran digital tidak lebih cepat dari pembayaran tunai dan H1 menyatakan sebaliknya, yang sangat tepat dengan kemampuan uji t untuk menguji perbedaan rata-rata antara dua kelompok dengan arah yang spesifik menggunakan one-tailed test.

Hasil

```
online = data[data['Payment Category'] == 'Online']['Delivery Duration (min)']
offline = data[data['Payment Category'] == 'Offline']['Delivery Duration (min)']
t_stat, p_value = ttest_ind(online, offline)

print("\n=== HASIL UJI T (Independent Samples T-Test) ===")
print(f"Nilai t-statistik: {t_stat:.3f}")
print(f"Nilai p-value: {p_value:.5f}")

# Statistical conclusion
alpha = 0.05
if p_value < alpha:
    if online.mean() > offline.mean():
        conclusion = "Terdapat perbedaan signifikan: pembayaran online memiliki waktu pengiriman yang lebih lama dibandingkan pembayaran offline."
    else:
        conclusion = "Terdapat perbedaan signifikan: pembayaran offline memiliki waktu pengiriman yang lebih lama dibandingkan pembayaran online."
    print(f"Kesimpulan: {conclusion}")
else:
    print("Kesimpulan: Tidak terdapat perbedaan signifikan pada waktu pengiriman antara pembayaran online dan offline.")
```



```
=== HASIL UJI T (Independent Samples T-Test) ===
Nilai t-statistik: -0.460
Nilai p-value: 0.64591
Kesimpulan: Tidak terdapat perbedaan signifikan pada waktu pengiriman antara pembayaran online dan offline.
```

Interpretasi hasil

Berdasarkan hasil uji t (Independent Samples T-Test), diperoleh t-statistik = -0.460 dan p-value = 0.64591, yang berarti tidak terdapat perbedaan signifikan antara rata-rata durasi pengiriman pada pembayaran online dan offline. Nilai p-value yang jauh di atas 0.05 menunjukkan bahwa perbedaan yang ada kemungkinan besar terjadi secara kebetulan dan tidak memiliki signifikansi statistik.

Hal ini mengindikasikan bahwa metode pembayaran (online maupun offline) tidak mempengaruhi waktu pengiriman pizza secara signifikan. Sistem operasional dan logistik pengiriman tampaknya sudah berjalan dengan konsisten terlepas dari jenis pembayaran yang dipilih pelanggan. Hasil ini menunjukkan bahwa pihak restoran telah berhasil mengelola proses pengiriman secara efisien untuk kedua kategori pembayaran. Temuan ini dapat memberikan kepercayaan kepada pelanggan bahwa tidak ada bias waktu pengiriman berdasarkan metode pembayaran yang mereka pilih, sehingga pelanggan dapat memilih metode pembayaran sesuai preferensi tanpa khawatir akan mempengaruhi kecepatan pengiriman.

Penjelasan Variabel yang tidak ada

Berdasarkan hipotesis metode pembayaran dan hasil uji t ($t = -0.460$, $p = 0.646$), beberapa variabel penting yang tidak dianalisis namun dapat mempengaruhi hubungan antara metode pembayaran dengan waktu pengiriman antara lain: waktu transaksi aktual (durasi proses pembayaran di lokasi pengiriman), kompleksitas transaksi (jumlah uang kembalian untuk pembayaran tunai), tingkat pengalaman kurir dalam menangani berbagai metode pembayaran, dan kondisi pelanggan saat menerima.

Saran Variabel dan penjelasannya

- Transaction Processing Time: Numerik kontinu (detik) untuk mengukur durasi aktual proses pembayaran di lokasi pengiriman, yang dapat menunjukkan perbedaan efisiensi nyata
- Cash Handling Complexity: Ordinal scale (mudah-sedang-sulit) berdasarkan kebutuhan kembalian dan denominasi uang yang terlibat dalam transaksi tunai
- Customer Payment Readiness: Kategorikal (siap/tidak siap) mengukur kesiapan pelanggan dalam melakukan pembayaran saat kurir tiba

Hipotesis kelima adalah bahwa jarak pengantaran berkorelasi positif dengan waktu pengiriman, namun dipengaruhi faktor lain seperti kemacetan. Secara intuitif, pesanan yang lebih jauh jaraknya membutuhkan waktu tempuh lebih lama. Namun, analisis logistik menyebutkan bahwa waktu pengiriman tidak selalu linier terhadap jarak karena pengaruh kondisi jalan dan kemacetan. Untuk menguji, model regresi linier dapat digunakan (waktu pengiriman \sim jarak), atau ANOVA jika jarak dikelompokkan menjadi zona dekat/jauh. Implikasi strategisnya, bila hipotesis ini benar, adalah penentuan zona tarif atau ETA berdasarkan jarak (misalnya layanan prioritas atau biaya lebih tinggi untuk jarak jauh) dan optimalisasi rute pengiriman (menggunakan algoritma terpendek atau pembagian wilayah) agar jarak tempuh tidak memperburuk kinerja pengiriman.

H0 : Tidak ada hubungan linear antara jarak dengan waktu pengiriman

H1 : Ada hubungan linear positif antara jarak dengan waktu pengiriman

Pertanyaan analisis

berdasarkan hipotesis tersebut maka pertanyaan analisisnya adalah Apakah terdapat hubungan positif antara jarak pengantaran dengan waktu yang dibutuhkan untuk mengirimkan pizza?

Metode yang digunakan

Metode uji yang digunakan adalah uji korelasi Pearson karena hipotesis yang dirumuskan secara spesifik bertujuan untuk menguji hubungan linear antara dua variabel numerik kontinu, yaitu jarak pengantaran (Distance) sebagai variabel prediktor dan waktu pengiriman (Delivery Duration) sebagai variabel outcome, dimana keduanya memiliki skala pengukuran rasio yang memungkinkan analisis korelasi untuk mengukur kekuatan dan arah hubungan linear. Hipotesis yang dinyatakan dengan H0 "tidak ada hubungan linear antara jarak dengan waktu pengiriman" dan H1 "ada hubungan linear positif antara jarak dengan waktu pengiriman" merupakan pernyataan klasik yang dapat dijawab langsung oleh koefisien korelasi Pearson melalui pengukuran statistik yang menghasilkan nilai r untuk menunjukkan kekuatan hubungan dan p -value untuk menentukan signifikansi statistik.

Hasil

```
x = data['Distance (km)']
y = data['Delivery Duration (min)']

r_value, p_value = pearsonr(x, y)

print("=== HASIL UJI KORELASI PEARSON ===")
print(f"Koefisien Korelasi (r): {r_value:.3f}")
print(f"Nilai p-value: {p_value:.5f}")

if p_value < 0.05:
    print("Kesimpulan: Terdapat hubungan signifikan antara jarak dan durasi pengiriman.")
else:
    print("Kesimpulan: Tidak terdapat hubungan signifikan antara jarak dan durasi pengiriman.")
```

Interpretasi hasil

Berdasarkan hasil uji korelasi Pearson diperoleh koefisien korelasi (r) sebesar 0.932 dengan p -value = 0.00000, yang menunjukkan adanya hubungan positif yang sangat kuat dan signifikan antara jarak pengiriman (Distance (km)) dengan durasi pengiriman (Delivery Duration (min)). Artinya, semakin jauh jarak yang ditempuh oleh kurir, maka waktu pengiriman pizza juga akan semakin lama. Nilai p -value yang lebih kecil dari 0.05 menegaskan bahwa hubungan tersebut tidak terjadi secara kebetulan, melainkan secara statistik signifikan.

Hasil ini memberikan gambaran bahwa faktor jarak menjadi salah satu penentu utama dalam efisiensi pengiriman pizza. Dengan mengetahui kekuatan hubungan ini, pihak restoran dapat memperkirakan waktu pengiriman lebih akurat serta mengoptimalkan rute agar durasi pengiriman dapat diminimalkan. Temuan ini juga dapat digunakan sebagai dasar dalam meningkatkan layanan logistik dan kepuasan pelanggan.

Penjelasan Variabel yang tidak ada

Berdasarkan hipotesis jarak pengiriman dan hasil korelasi Pearson yang sangat kuat ($r=0.932$), beberapa variabel penting yang tidak dianalisis namun dapat mempengaruhi hubungan antara jarak dengan waktu pengiriman antara lain: kondisi infrastruktur jalan (kualitas aspal, keberadaan jalan tol vs jalan lokal), topografi wilayah (jalan menanjak/menurun yang mempengaruhi kecepatan kendaraan)

Saran Variabel dan penjelasannya

- Road Quality Index: Skala ordinal (baik-sedang-buruk) untuk mengukur kondisi fisik jalan yang dilalui, karena jalan rusak dapat memperlambat kecepatan meski jarak sama
- Topographic Difficulty: Kategorikal (datar/berbukit/pegunungan) mengukur tingkat kesulitan geografis yang mempengaruhi kecepatan tempuh kendaraan pengiriman
- Route Accessibility: Ordinal scale (mudah-sedang-sulit) berdasarkan lebar jalan dan kemudahan akses ke lokasi tujuan, terutama untuk area perumahan atau gang sempit