

UNIVERSITAS GUNADARMA



PRAKTIKUM KECERDASAN ARTIFISIAL

MANUAL BOOK

"Pemodelan Pendapatan Harian Kedai Kopi Menggunakan Regresi Linier Berganda: Pendekatan Analitis terhadap Faktor Operasional dan Pemasaran"

Nama : Alvito Hafizh Putra Haryanto
NPM : 50423139
Kelas : 3 IA 0 5
Fakultas : Teknologi Industri
Jurusan : Informatika
PJ : Fitriyani

Ditulis Guna Melengkapi Sebagian Syarat

Praktikum Kecerdasan Artifisial

Universitas Gunadarma

2025

DAFTAR ISI

COVER.....	1
DAFTAR ISI.....	2
BAB I	3
PENDAHULUAN	3
1.1 Latar Belakang.....	3
1.2 Tujuan.....	6
BAB II.....	7
PEMBAHASAN	7
2.1 Pendapatan Usaha	7
2.2 Jumlah Pelanggan.....	7
2.3 Nilai Transaksi Rata-Rata (Average Order Value)	7
2.4. Jam Operasional	7
2.5. Jumlah Karyawan	8
2.6. Pengeluaran Pemasaran (Marketing Spend)	8
2.7. Lalu Lintas Pejalan Kaki (Foot Traffic)	8
2.8. Regresi Linier Berganda	8
BAB III	10
ANALISA DAN PERANCANGAN.....	10
BAB IV.....	24
PENUTUP	24
4.1 Kesimpulan.....	24
4.2 Saran	24
DAFTAR PUSTAKA	26

BAB I

PENDAHULUAN

1.1 Latar Belakang

Industri kopi merupakan salah satu sektor yang mengalami pertumbuhan paling pesat dalam dekade terakhir. Dalam konteks global, kopi bukan hanya sekadar komoditas pertanian, melainkan telah menjadi bagian penting dari gaya hidup masyarakat modern dan simbol interaksi sosial lintas budaya. Berdasarkan laporan International Coffee Organization (ICO, 2024), konsumsi kopi dunia tumbuh rata-rata 2,5% per tahun dalam lima tahun terakhir, dengan peningkatan signifikan di kawasan Asia dan Amerika Utara. Sementara itu, Grand View Research (2024) mencatat bahwa pasar kafe global diperkirakan mencapai USD 77 miliar pada tahun 2025, dengan tingkat pertumbuhan tahunan gabungan (CAGR) sebesar 8,2%. Pertumbuhan ini didorong oleh peningkatan pendapatan masyarakat, tren urbanisasi, serta transformasi budaya yang menjadikan kedai kopi bukan sekadar tempat minum, melainkan ruang sosial, kerja, dan ekspresi gaya hidup.

Di Amerika Serikat dan Eropa, bisnis kopi telah bertransformasi menjadi bagian integral dari ekonomi kreatif dan industri jasa. Dalam konteks Asia Tenggara, fenomena serupa juga terlihat jelas, terutama di kota-kota besar seperti Jakarta, Bangkok, dan Kuala Lumpur, di mana kehadiran kedai kopi meningkat hampir dua kali lipat dalam lima tahun terakhir (Business Research Insights, 2024). Hal ini menunjukkan bahwa peluang ekonomi di sektor ini sangat menjanjikan, namun di sisi lain menimbulkan kompetisi yang sangat ketat antar pelaku usaha. Dalam situasi seperti ini, strategi pengelolaan operasional yang efektif menjadi faktor krusial untuk mempertahankan profitabilitas dan keberlanjutan bisnis.

Namun demikian, di balik prospek yang menjanjikan tersebut, para pelaku bisnis kopi menghadapi berbagai tantangan dalam menjaga stabilitas pendapatan harian. Pendapatan sebuah kedai kopi sangat dipengaruhi oleh beragam faktor, baik internal (operasional) maupun eksternal (lingkungan). Faktor internal mencakup jumlah pelanggan per hari, nilai rata-rata transaksi (average order value), jam operasional, jumlah karyawan,

serta biaya pemasaran harian. Sementara faktor eksternal meliputi lalu lintas pejalan kaki (foot traffic), kondisi cuaca, lokasi strategis, dan tren sosial di sekitar area usaha.

Kompleksitas interaksi antar variabel inilah yang menjadikan prediksi pendapatan harian sebuah kedai kopi sebagai tantangan analitis yang menarik dan relevan untuk diteliti secara ilmiah.

Penelitian oleh Chakma (2025) menunjukkan bahwa kombinasi faktor operasional seperti jumlah pelanggan, jam operasional, dan nilai transaksi rata-rata memiliki korelasi yang sangat kuat terhadap pendapatan harian. Dengan menggunakan pendekatan regresi linier berganda pada data observasi harian kedai kopi, diperoleh nilai koefisien determinasi (R^2) sebesar 0,89, yang menunjukkan bahwa sebagian besar variasi pendapatan dapat dijelaskan oleh faktor-faktor tersebut. Temuan ini sejalan dengan kajian SharpSheets (2024) yang melaporkan bahwa rata-rata pendapatan harian sebuah kafe independen dapat meningkat hingga 25% dengan optimalisasi variabel-variabel operasional seperti jumlah tenaga kerja dan strategi promosi yang tepat.

Faktor eksternal juga memiliki pengaruh signifikan terhadap pendapatan. FoodDrinkTalk (2025) menyatakan bahwa tingkat foot traffic merupakan salah satu indikator paling kuat untuk memperkirakan volume penjualan. Kafe yang berlokasi di area dengan lalu lintas pejalan kaki tinggi (di atas 800 orang per jam) cenderung memiliki potensi pendapatan hingga 40% lebih tinggi dibandingkan yang berada di area dengan trafik rendah. Selain itu, pengeluaran untuk pemasaran (marketing spend) juga terbukti memiliki dampak langsung terhadap kesadaran merek dan niat berkunjung pelanggan (Coffee Plus Three, 2025). Dengan demikian, keseimbangan antara strategi pemasaran dan efisiensi operasional menjadi kunci utama dalam meningkatkan profitabilitas harian bisnis kopi.

Meskipun berbagai studi telah meneliti aspek pemasaran dan perilaku konsumen di industri kopi, masih terdapat kesenjangan penelitian yang cukup besar dalam analisis kuantitatif berbasis data harian. Sebagian besar penelitian terdahulu berfokus pada persepsi pelanggan, loyalitas merek, atau kualitas produk, sementara kajian yang mengintegrasikan variabel-variabel kuantitatif seperti jumlah pelanggan, jam operasional, jumlah karyawan, pengeluaran promosi, dan foot traffic dalam model prediktif pendapatan harian masih terbatas. Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan pendekatan data-driven yang menekankan pada analisis prediktif dan optimasi bisnis.

Dataset yang digunakan dalam penelitian ini terdiri dari 2.000 observasi harian dari berbagai warung kopi dengan karakteristik operasional dan lingkungan yang beragam.

Variabel yang diamati meliputi:

- Jumlah pelanggan per hari: 50–500 pelanggan
- Rata-rata nilai transaksi (average order value): US\$ 2,50–10,00
- Jam operasional per hari: 6–18 jam
- Jumlah karyawan: 2–15 orang
- Pengeluaran pemasaran harian: US\$ 10–500
- Lalu lintas pejalan kaki (foot traffic): 50–1000 orang per jam

Variabel-variabel ini dipilih karena mewakili aspek utama dari kegiatan operasional dan kondisi lingkungan yang secara teoritis maupun empiris berpengaruh terhadap pendapatan harian. Melalui pendekatan analitik dan model prediktif, penelitian ini berupaya menjawab pertanyaan fundamental: faktor apa yang paling signifikan memengaruhi pendapatan harian warung kopi, dan bagaimana hubungan antarvariabel tersebut dapat dimanfaatkan untuk meningkatkan performa bisnis?

Selain memiliki kontribusi praktis bagi pelaku usaha, penelitian ini juga memberikan kontribusi akademik dalam bidang ekonomi mikro terapan dan analisis bisnis. Dari sisi metodologis, penelitian ini menggabungkan pendekatan kuantitatif dengan teknik analisis prediktif untuk membangun model yang dapat digunakan oleh pengusaha kecil dan menengah di sektor F&B. Dari sisi teoritis, penelitian ini memperkuat pemahaman mengenai hubungan kausal antara keputusan operasional harian dan hasil finansial dalam konteks bisnis jasa yang padat interaksi manusia.

Dari perspektif praktis, hasil penelitian ini diharapkan dapat membantu pemilik warung kopi dan manajer bisnis untuk:

1. Mengidentifikasi faktor operasional yang paling berpengaruh terhadap pendapatan.
2. Menentukan jam operasional dan jumlah karyawan yang optimal berdasarkan pola pelanggan harian.
3. Mengalokasikan anggaran pemasaran dengan lebih efisien berdasarkan data empiris.
4. Menentukan strategi lokasi dan promosi yang sesuai dengan tingkat foot traffic di area bisnis.

Selain itu, penelitian ini juga memiliki nilai strategis bagi lembaga keuangan dan pembuat kebijakan yang ingin mendorong digitalisasi dan efisiensi usaha mikro, kecil, dan menengah (UMKM) di sektor F&B. Dalam konteks yang lebih luas, pendekatan analitik berbasis data seperti ini sejalan dengan tren global menuju ekonomi berbasis kecerdasan buatan (AI-driven economy), di mana pengambilan keputusan bisnis semakin bergantung pada analisis prediktif dan pembelajaran mesin (machine learning).

Dengan demikian, penelitian ini memiliki urgensi ganda baik secara ilmiah maupun praktis. Secara ilmiah, penelitian ini memperluas horizon pengetahuan tentang determinan pendapatan usaha kopi dengan pendekatan kuantitatif berbasis data harian yang jarang dilakukan sebelumnya. Secara praktis, penelitian ini berpotensi menghasilkan model prediksi yang dapat diterapkan untuk membantu pengusaha meningkatkan pendapatan dan efisiensi operasional melalui pengambilan keputusan berbasis data (data-driven decision making).

Oleh karena itu, penelitian ini menjadi relevan dan penting dilakukan, karena tidak hanya memberikan kontribusi pada literatur akademik tentang ekonomi jasa dan perilaku konsumen, tetapi juga memberikan solusi nyata bagi tantangan bisnis di industri kopi yang semakin kompetitif. Dengan menggunakan dataset empiris yang kaya dan pendekatan analitik modern, penelitian ini bertujuan untuk merumuskan model prediksi pendapatan harian warung kopi yang akurat, efisien, dan aplikatif, sehingga dapat menjadi dasar bagi strategi bisnis berkelanjutan di sektor makanan dan minuman.

1.2 Tujuan

Penelitian ini memiliki tujuan utama untuk menganalisis faktor-faktor yang memengaruhi pendapatan harian warung kopi dengan menggunakan metode regresi linier berganda. Melalui pendekatan ini, penelitian bertujuan untuk menjelaskan secara kuantitatif hubungan antara variabel-variabel operasional dan eksternal terhadap pendapatan harian, serta mengidentifikasi variabel yang memiliki pengaruh paling dominan.

Secara umum, penelitian ini diarahkan untuk memberikan pemahaman yang lebih mendalam mengenai bagaimana kondisi operasional harian, strategi pemasaran, dan faktor lingkungan sekitar dapat memengaruhi kinerja keuangan suatu usaha kedai kopi.

BAB II

PEMBAHASAN

2.1 Pendapatan Usaha

Pendapatan merupakan hasil dari aktivitas ekonomi yang mencerminkan kinerja operasional dan keberlanjutan bisnis. Dalam konteks usaha kecil seperti warung kopi, pendapatan harian dipengaruhi oleh kombinasi faktor internal dan eksternal yang berkaitan dengan strategi pemasaran, operasional, dan karakteristik pelanggan. Studi oleh Susetya & Marhaeni (2023) menunjukkan bahwa variabel seperti lokasi, jumlah pelanggan, dan jam operasional memiliki pengaruh signifikan terhadap pendapatan harian bisnis kopi di Denpasar. Pendapatan juga mencerminkan efektivitas strategi manajerial dan efisiensi penggunaan sumber daya.

2.2 Jumlah Pelanggan

Jumlah pelanggan menjadi faktor kunci dalam menentukan tingkat pendapatan bisnis kopi. Penelitian oleh Kasimin (2017) mengungkapkan bahwa karakteristik lokasi, kualitas pelayanan, dan daya tarik produk berpengaruh signifikan terhadap jumlah pengunjung. Dalam teori perilaku konsumen, keputusan pelanggan dipengaruhi oleh persepsi terhadap kualitas, kenyamanan, dan nilai emosional yang ditawarkan oleh tempat usaha (Charoenphol & Barrett, 2019).

2.3 Nilai Transaksi Rata-Rata (Average Order Value)

Nilai transaksi rata-rata menggambarkan besarnya pengeluaran per pelanggan setiap kali bertransaksi. Menurut McManus (2007), strategi harga non-linear dan diferensiasi produk berperan besar dalam menentukan margin keuntungan dan nilai transaksi pelanggan di pasar kopi spesialti. Penerapan strategi upselling dan cross-selling terbukti efektif meningkatkan nilai transaksi harian.

2.4. Jam Operasional

Jam operasional menentukan kapasitas layanan dan tingkat pendapatan potensial. Studi oleh Relihan (2022) menegaskan bahwa jam operasional yang disesuaikan dengan

perilaku pelanggan misalnya puncak pagi dan sore dapat meningkatkan efisiensi pendapatan tanpa menambah biaya tetap secara signifikan. Namun, efektivitas jam operasional juga bergantung pada ketersediaan tenaga kerja dan biaya energi.

2.5. Jumlah Karyawan

Jumlah karyawan memengaruhi kualitas pelayanan dan kepuasan pelanggan. Tilahun (2007) menemukan bahwa produktivitas tenaga kerja berkorelasi langsung dengan profitabilitas unit bisnis, namun efek tersebut menurun setelah melewati titik efisiensi optimal. Dengan demikian, jumlah karyawan yang proporsional dengan volume pelanggan penting untuk menjaga efisiensi biaya dan kualitas pelayanan.

2.6. Pengeluaran Pemasaran (Marketing Spend)

Pengeluaran pemasaran merupakan investasi strategis dalam memperluas jangkauan pasar dan membangun loyalitas pelanggan. Dalam studi Adeleke (2019), ditemukan bahwa penggunaan bauran pemasaran 8P (produk, harga, tempat, promosi, orang, proses, bukti fisik, dan produktivitas) secara konsisten berpengaruh terhadap peningkatan pendapatan warung kopi kecil. Selain itu, Huang & Wood (2017) menekankan pentingnya strategi pemasaran digital dan pengalaman pelanggan dalam meningkatkan loyalitas serta penjualan.

2.7. Lalu Lintas Pejalan Kaki (Foot Traffic)

Foot traffic menggambarkan potensi permintaan yang tersedia di sekitar lokasi usaha. Susetya & Marhaeni (2023) dan Kasimin (2017) menunjukkan bahwa lokasi dengan arus pejalan kaki tinggi berpengaruh positif terhadap volume pelanggan dan pendapatan, terutama jika didukung oleh tampilan toko dan promosi visual yang menarik.

2.8. Regresi Linier Berganda

Metode regresi linier berganda digunakan untuk mengukur pengaruh simultan beberapa variabel independen terhadap pendapatan. Model ini umum digunakan dalam penelitian ekonomi dan manajemen usaha kecil untuk mengidentifikasi determinan utama pendapatan bisnis kopi. KERU (2021) menggunakan pendekatan multivariat untuk menganalisis faktor-faktor yang memengaruhi profitabilitas pasar kopi, menegaskan bahwa variabel seperti jumlah pelanggan, harga jual, dan biaya pemasaran memiliki pengaruh signifikan terhadap laba usaha.

Persamaan model regresi yang digunakan:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \epsilon$$

di mana:

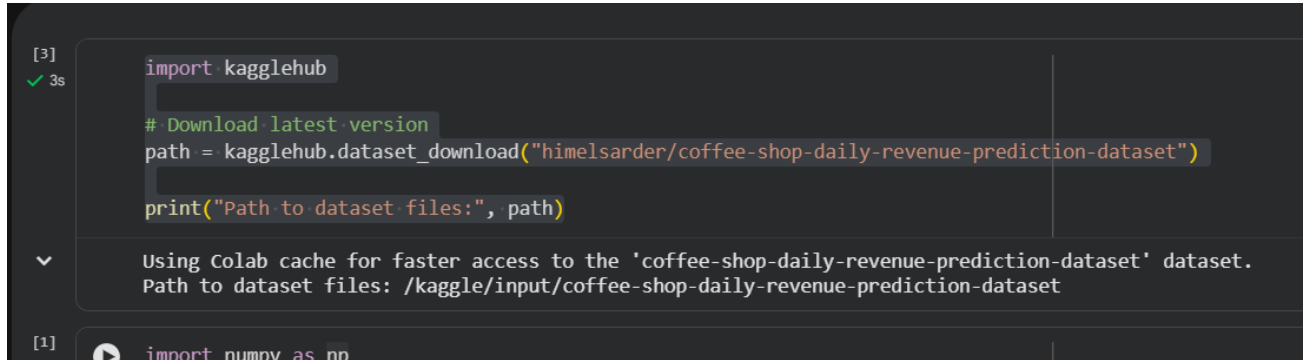
- Y = Pendapatan harian
- X_i = Jumlah pelanggan

- X_2 = Nilai transaksi rata-rata
- X_3 = Jam operasional
- X_4 = Jumlah karyawan
- X_5 = Pengeluaran pemasaran
- X_6 = Lalu lintas pejalan kaki
- ε = Error term

BAB III

ANALISA DAN PERANCANGAN

- Langkah pertama adalah mendownload dataset coffee shop daily revenue prediction dari Kaggle.com



```
[3] ✓ 3s
import kagglehub

# Download latest version
path = kagglehub.dataset_download("himelsander/coffee-shop-daily-revenue-prediction-dataset")

print("Path to dataset files:", path)
```

Using Colab cache for faster access to the 'coffee-shop-daily-revenue-prediction-dataset' dataset.
Path to dataset files: /kaggle/input/coffee-shop-daily-revenue-prediction-dataset

```
[1] ▶ import numpy as np
```

- Kemudian, mengimpor seluruh library yang diperlukan dalam proses analisis data dan pembangunan model regresi linier berganda. NumPy dan Pandas digunakan untuk mengolah data numerik serta membaca dataset ke dalam bentuk DataFrame. Matplotlib dan Seaborn berfungsi untuk membuat visualisasi seperti grafik distribusi, scatter plot, dan heatmap yang membantu memahami pola dalam data. Selanjutnya, scikit-learn menyediakan berbagai komponen penting untuk machine learning: `train_test_split` digunakan untuk membagi data menjadi data latih dan data uji, `LinearRegression` digunakan sebagai model regresi linier, `StandardScaler` untuk menstandarisasi skala fitur agar model lebih stabil, `Pipeline` untuk merangkai langkah scaling dan model menjadi satu alur kerja yang rapi, serta metrik seperti MAE, MSE, dan R^2 digunakan untuk mengevaluasi performa model. Dengan mengimpor semua library ini, kita menyiapkan seluruh alat yang dibutuhkan untuk melakukan analisis regresi linier secara lengkap, mulai dari pengolahan data, visualisasi, pemodelan, hingga evaluasi hasil.

```
[1] import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

- Memuat dataset pendapatan harian kedai kopi dari file CSV ke dalam Python menggunakan library Pandas. Pertama, lokasi file dataset disimpan dalam sebuah variabel, kemudian file tersebut dibaca dan diubah menjadi sebuah DataFrame yang berisi data dalam bentuk tabel. Setelah dataset berhasil dimuat, `df.head()` digunakan untuk menampilkan lima baris pertama sebagai contoh awal, sehingga pengguna dapat melihat struktur data, nama kolom, dan nilai-nilai pada dataset. Langkah ini penting sebagai tahap awal eksplorasi data untuk memastikan bahwa dataset telah terbaca dengan benar sebelum melakukan analisis lebih lanjut seperti visualisasi atau pemodelan regresi.

```
dataset_coffeeshop = "/kaggle/input/coffee-shop-daily-revenue-prediction-dataset/coffee_shop_revenue.csv"
df = pd.read_csv(dataset_coffeeshop)
print(df.head())
```

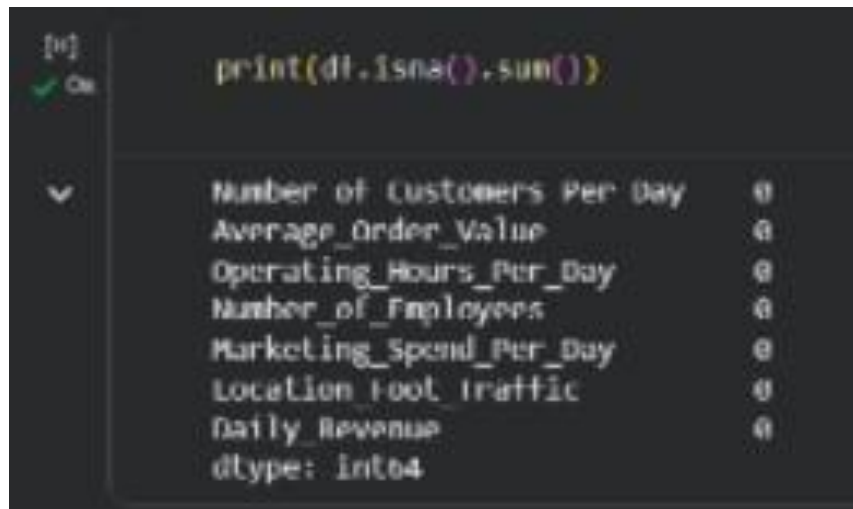
	Number_of_Customers_Per_Day	Average_Order_Value	Operating_Hours_Per_Day	\
0	152	6.74	14	
1	485	4.50	12	
2	398	9.09	6	
3	320	8.48	17	
4	156	7.44	17	

	Number_of_Employees	Marketing_Spend_Per_Day	Location_Foot_Traffic	\
0	4	106.62	97	
1	8	57.83	744	
2	6	91.76	636	
3	4	462.63	770	
4	2	412.52	232	

	Daily_Revenue
0	1547.81
1	2084.68
2	3118.39
3	2912.20
4	1663.42

- Kemudian memeriksa apakah terdapat missing value (data kosong) pada setiap kolom dalam dataset. Fungsi `df.isna()` akan mendeteksi nilai yang kosong dengan memberi tanda True/False, lalu `.sum()` menghitung berapa banyak nilai kosong pada tiap kolom. Hasil yang

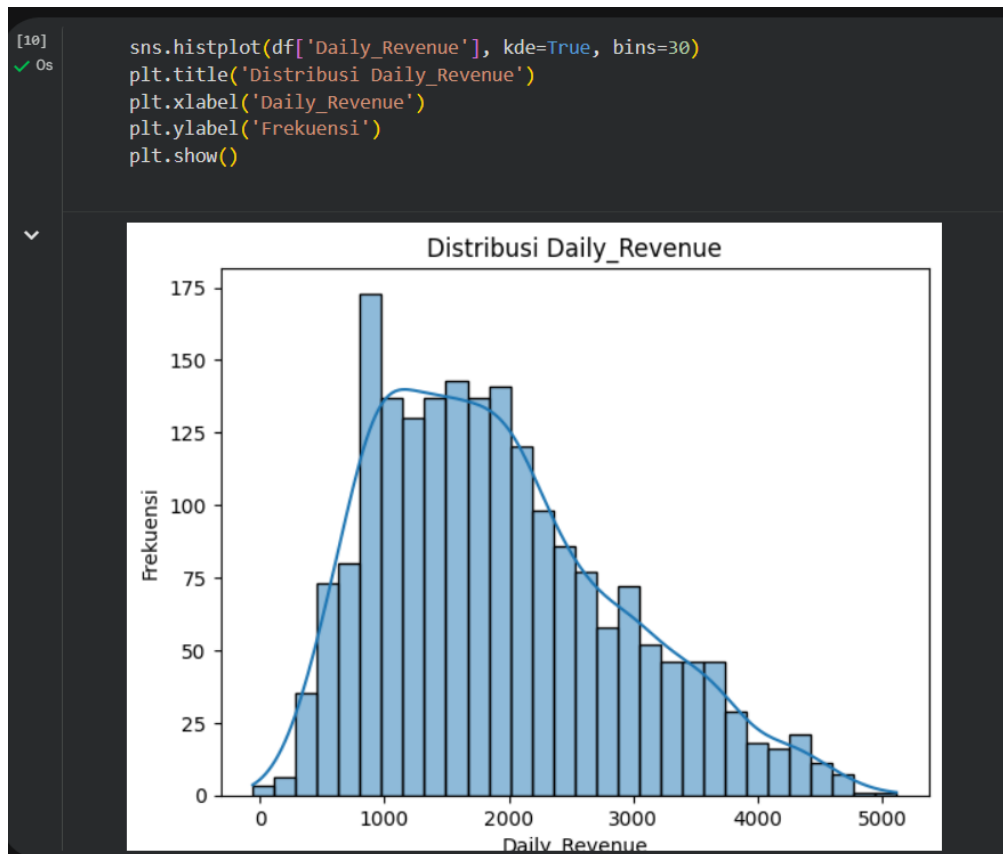
ditampilkan menunjukkan bahwa semua kolom memiliki nilai 0, artinya tidak ada data yang hilang, sehingga dataset sudah bersih dan dapat langsung digunakan untuk analisis tanpa perlu proses imputasi atau pembersihan data tambahan.



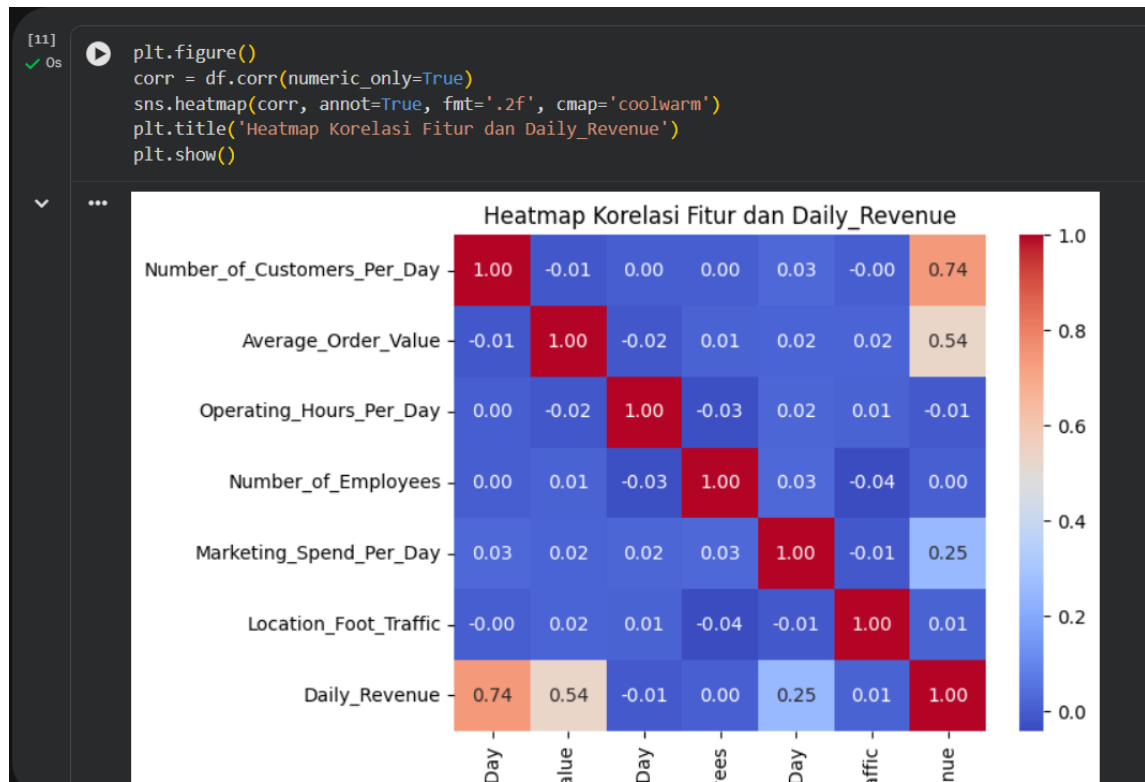
```
[4]: print(df.isna().sum())
```

Number of Customers Per Day	0
Average_Order_Value	0
Operating_Hours_Per_Day	0
Number_of_Employees	0
Marketing_Spend_Per_Day	0
Location_Foot_Traffic	0
Daily_Revenue	0
dtype:	int64

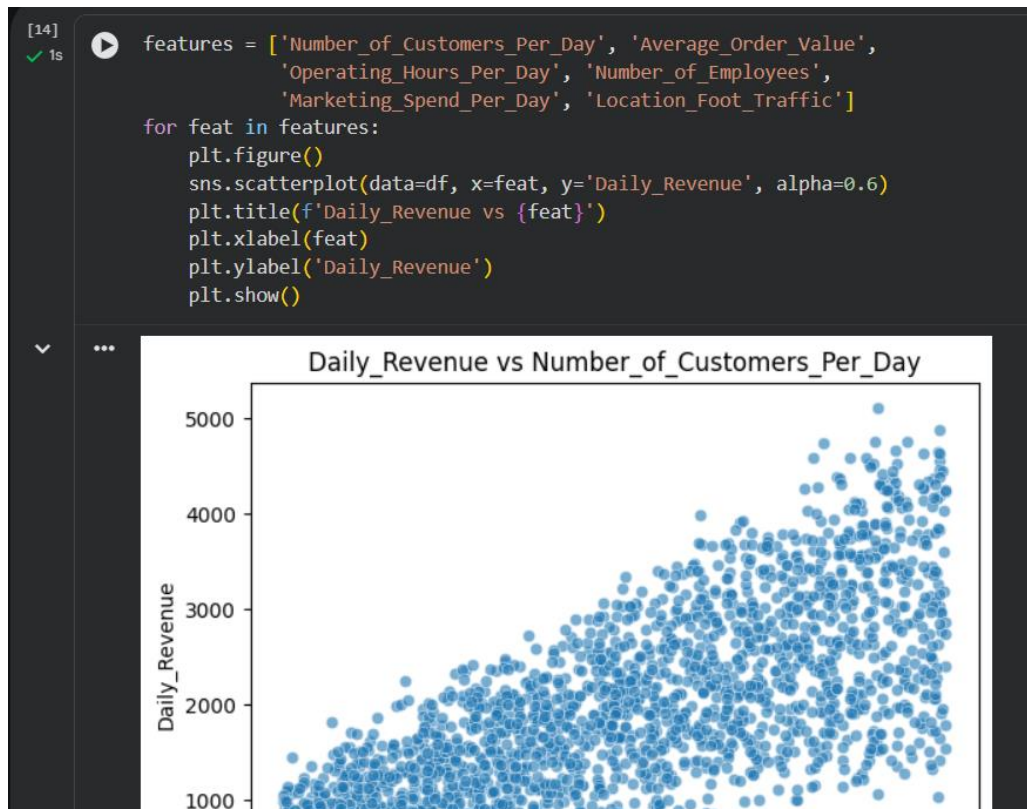
- Menampilkan distribusi pendapatan harian (Daily_Revenue) dalam bentuk histogram dengan tambahan kurva KDE (Kernel Density Estimation). Visualisasi ini membantu memahami bagaimana penyebaran nilai pendapatan di dataset—apakah datanya condong (skewed), memiliki puncak tertentu, atau terdapat rentang pendapatan yang sering muncul. Dengan bins=30, pendapatan dikelompokkan ke dalam 30 interval, sehingga pola distribusi terlihat lebih jelas. Label sumbu X dan Y ditambahkan agar grafik mudah dibaca, sedangkan judul grafik menjelaskan bahwa ini adalah distribusi dari variabel Daily_Revenue. Secara keseluruhan, langkah ini merupakan bagian dari Exploratory Data Analysis (EDA) untuk memahami karakteristik target variabel sebelum melakukan pemodelan..



- Membuat heatmap korelasi yang menunjukkan kekuatan hubungan antar variabel numerik dalam dataset, khususnya hubungan setiap fitur dengan Daily_Revenue. Pertama, `df.corr(numeric_only=True)` menghitung nilai korelasi antar kolom numerik, lalu `sns.heatmap()` menampilkan matriks korelasi tersebut dalam bentuk peta warna, di mana warna merah menunjukkan korelasi tinggi dan biru menunjukkan korelasi rendah. Parameter `annot=True` menampilkan nilai korelasi pada setiap sel, `fmt='.2f'` membatasi angka menjadi dua desimal, dan `cmap='coolwarm'` menentukan skema warna yang memudahkan membaca pola hubungan positif atau negatif. Grafik ini membantu mengidentifikasi variabel mana yang paling berhubungan dengan pendapatan harian, sehingga penting untuk analisis regresi dan pengambilan keputusan bisnis..



- Kemudian, membuat scatter plot antara setiap fitur (variabel independen) dengan Daily_Revenue (variabel dependen) sebagai bagian dari *Exploratory Data Analysis* (EDA). Pertama, daftar fitur disimpan dalam variabel features. Lalu, melalui perulangan for, setiap fitur diplot satu per satu terhadap pendapatan harian menggunakan `sns.scatterplot()`. Visualisasi ini membantu melihat pola hubungan—misalnya apakah hubungan linear, apakah terdapat tren naik, apakah datanya tersebar acak, atau apakah ada outlier. Setiap grafik diberi judul dan label sumbu agar mudah dipahami. Secara keseluruhan, langkah ini bertujuan untuk mengevaluasi bagaimana masing-masing variabel memengaruhi Daily_Revenue sebelum dilakukan pemodelan regresi.



- Selanjutnya melakukan proses persiapan data dan pelatihan model regresi linier. Pertama, data dipisahkan menjadi dua bagian: variabel independen (X) yang berisi fitur-fitur seperti jumlah pelanggan, nilai transaksi, jam operasional, dan sebagainya, serta variabel dependen (y) yaitu Daily_Revenue. Setelah itu, dataset dibagi menjadi data pelatihan dan data uji menggunakan `train_test_split` dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian, sehingga model dapat dilatih dan diuji secara adil. Jumlah masing-masing data ditampilkan untuk memastikan pembagian sudah benar. Selanjutnya, dibuat sebuah *pipeline* yang terdiri dari dua langkah: menstandarisasi data menggunakan `StandardScaler` dan menerapkan algoritma `LinearRegression`. Pipeline ini kemudian dilatih dengan data pelatihan (`X_train` dan `y_train`), sehingga model siap digunakan untuk melakukan prediksi dan evaluasi.

```
[15] ✓
x = df[features]
y = df['Daily_Revenue']
x_train, x_test, y_train, y_test = train_test_split(
    x, y,
    test_size=0.20,
    random_state=42
)

[16] ✓ 0s
print("\nJumlah data pelatihan:", x_train.shape[0])
print("Jumlah data uji:", x_test.shape[0])

Jumlah data pelatihan: 1600
Jumlah data uji: 400

[29] ✓ 0s
pipe = Pipeline([
    ('scaler', StandardScaler()),
    ('linreg', LinearRegression())
])

pipe.fit(x_train, y_train)
print("\nModel Linear Regression dilatih.")

Model Linear Regression dilatih.
```

- Mengevaluasi kinerja model regresi linier dengan menghitung dan menampilkan berbagai metrik error pada data pelatihan (train) dan data pengujian (test). Pertama, model digunakan untuk memprediksi nilai Daily_Revenue pada kedua dataset sehingga diperoleh `y_train_pred` dan `y_test_pred`. Selanjutnya, fungsi `print_metrics` dibuat untuk menghitung empat metrik evaluasi utama, yaitu MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Squared Error), dan R^2 (Koefisien Determinasi). Metrik-metrik ini memberikan gambaran seberapa jauh hasil prediksi model dari nilai sebenarnya dan seberapa baik model menjelaskan variasi data. Terakhir, fungsi tersebut dipanggil untuk dataset train dan test, menghasilkan nilai error yang relatif kecil serta nilai R^2 sekitar 0.89, yang menunjukkan bahwa model memiliki performa yang sangat baik dan mampu menjelaskan sekitar 89% variasi pendapatan harian berdasarkan fitur yang digunakan.


```

y_train_pred = pipe.predict(X_train)
y_test_pred = pipe.predict(X_test)

def print_metrics(y_true, y_pred, label=""):
    mae = mean_absolute_error(y_true, y_pred)
    mse = mean_squared_error(y_true, y_pred)
    rmse = np.sqrt(mse)
    r2 = r2_score(y_true, y_pred)
    print(f"\n=== Metrik {label} ===")
    print(f"MAE : {mae:,.2f}")
    print(f"MSE : {mse:,.2f}")
    print(f"RMSE : {rmse:,.2f}")
    print(f"R2 : {r2:.4f}")

print_metrics(y_train, y_train_pred, label="TRAIN")
print_metrics(y_test, y_test_pred, label="TEST")

=== Metrik TRAIN ===
MAE : 255.22
MSE : 104,746.48
RMSE : 323.65
R2 : 0.8905

=== Metrik TEST ===
MAE : 244.21
MSE : 97,569.72
RMSE : 312.36
R2 : 0.8956

```

- Melakukan proses cross-validation untuk menguji seberapa stabil dan konsisten performa model regresi linier ketika diuji pada beberapa pembagian data yang berbeda. Pertama, dibuat objek KFold dengan 5 lipatan (5-fold) yang membagi dataset menjadi lima bagian, lalu melakukan pelatihan dan pengujian secara bergantian di setiap bagian. Parameter `shuffle=True` memastikan data diacak terlebih dahulu sehingga pembagian lebih merata, dan `random_state=42` menjaga hasil agar selalu sama. Selanjutnya, `cross_val_score` menjalankan model (pipe) pada setiap lipatan dan menghitung nilai R^2 untuk masing-masing percobaan. Hasil berupa lima nilai R^2 ditampilkan, yang menunjukkan performa model pada setiap fold. Nilai rata-ratanya kemudian dihitung untuk memberikan gambaran umum terhadap kemampuan generalisasi model. Dengan rata-rata R^2 sekitar 0.89, dapat disimpulkan bahwa

model memiliki performa yang stabil dan akurat di berbagai pembagian data, serta tidak menunjukkan indikasi overfitting maupun underfitting..

```
[ ] from sklearn.model_selection import KFold

kf = KFold(n_splits=5, shuffle=True, random_state=42)

cv_scores = cross_val_score(pipe, X, y, cv=kf, scoring='r2')

print("\n=== CROSS-VALIDATION R2 SCORES ===")
print(cv_scores)
print("Rata-rata R2 (CV):", cv_scores.mean().round(4))

v

=== CROSS-VALIDATION R2 SCORES ===
[0.89557684 0.88754466 0.88622029 0.88843638 0.89656809]
Rata-rata R2 (CV): 0.8909
```

- Mengambil dan menampilkan koefisien regresi dari model Linear Regression, lalu menyusunnya ke dalam tabel agar mudah dianalisis. Pertama, model regresi diambil dari pipeline melalui `pipe.named_steps['linreg']`, kemudian koefisien setiap fitur (`model.coef_`) dan nilai intersep (`model.intercept_`) diekstraksi. Selanjutnya, koefisien ini digabungkan dengan nama fitur ke dalam sebuah DataFrame (`coef_df`), lalu diurutkan dari yang paling besar hingga paling kecil. Output tabel ini menunjukkan variabel mana yang memiliki pengaruh terbesar terhadap pendapatan harian (Daily_Revenue). Misalnya, `Number_of_Customers_Per_Day` dan `Average_Order_Value` memiliki koefisien paling tinggi, sehingga menjadi faktor paling dominan, sementara `Number_of_Employees` memiliki koefisien negatif kecil, menunjukkan pengaruh yang sangat rendah atau bahkan sedikit menurunkan pendapatan. Intersep ditampilkan sebagai konstanta dasar model ketika semua fitur bernilai nol. Secara keseluruhan, langkah ini bertujuan untuk memahami kekuatan dan arah pengaruh masing-masing variabel dalam model regresi linier.

```
[ ] model = pipe.named_steps['linreg']
     coefs = model.coef_
     intercept = model.intercept_

     coef_df = pd.DataFrame({
         'feature': features,
         'coefficient': coefs
     }).sort_values(by='coefficient', ascending=False)

     print("\nKoefisien model (urut menurun):")
     print(coef_df)
     print("\nIntercept:", intercept)
```

▼

```
Koefisien model (urut menurun):
```

	feature	coefficient
0	Number_of_Customers_Per_Day	722.161261
1	Average_Order_Value	533.042295
4	Marketing_Spend_Per_Day	218.902261
5	Location_Foot_Traffic	4.767182
2	Operating_Hours_Per_Day	0.918719
3	Number_of_Employees	-6.578680

```
Intercept: 1920.4675437500005
```

- Menampilkan daftar faktor operasional yang paling berpengaruh terhadap pendapatan harian berdasarkan hasil perhitungan koefisien regresi linear. Dengan mencetak `coef_df`, tabel yang berisi nama fitur dan nilai koefisiennya ditampilkan dalam urutan dari yang paling besar hingga paling kecil. Nilai koefisien ini menunjukkan seberapa kuat dan ke arah mana (positif atau negatif) masing-masing variabel mempengaruhi `Daily_Revenue`. Dari hasil yang terlihat, variabel seperti `Number_of_Customers_Per_Day` dan `Average_Order_Value` memiliki koefisien tertinggi sehingga menjadi faktor yang paling dominan dalam meningkatkan pendapatan, sementara variabel seperti `Operating_Hours_Per_Day` dan `Number_of_Employees` memiliki pengaruh yang jauh lebih kecil. Secara keseluruhan, langkah ini bertujuan untuk mengidentifikasi fitur yang memiliki kontribusi terbesar terhadap pendapatan, sebagai dasar untuk pengambilan keputusan dalam operasional bisnis kedai kopi..

```
[ ] print("\n== Faktor Operasional Paling Berpengaruh ==")
    print(coef_df)

v

== Faktor Operasional Paling Berpengaruh ==
      feature  coefficient
0  Number_of_Customers_Per_Day    722.161261
1      Average_Order_Value    533.042295
4   Marketing_Spend_Per_Day    218.902261
5   Location_Foot_Traffic      4.767182
2   Operating_Hours_Per_Day     0.918719
3   Number_of_Employees    -6.578680
```

- Mencari kombinasi jam operasional dan jumlah karyawan yang menghasilkan prediksi pendapatan harian paling tinggi berdasarkan model regresi yang sudah dilatih. Pertama, dibuat rentang nilai jam operasional (hrs_range) dan jumlah karyawan (emps_range) yang akan diuji. Lalu dihitung nilai rata-rata dari fitur lain seperti jumlah pelanggan, nilai transaksi, anggaran pemasaran, dan foot traffic untuk dijadikan nilai tetap selama simulasi. Setelah itu, dilakukan perulangan bersarang (for h in hrs_range dan for e in emps_range) untuk mencoba setiap kombinasi jam kerja dan jumlah karyawan. Pada setiap kombinasi, dibentuk data input baru (x_new), lalu model memprediksi pendapatan yang mungkin diperoleh. Jika prediksi tersebut lebih tinggi dari nilai terbaik sebelumnya, maka nilai tersebut disimpan sebagai prediksi terbaik beserta kombinasi jam dan karyawannya. Di akhir proses, kombinasi paling optimal ditampilkan, menunjukkan berapa jam operasional dan jumlah karyawan yang menghasilkan pendapatan tertinggi. Langkah ini membantu menganalisis keputusan operasional yang paling menguntungkan bagi bisnis kedai kopi.

```
[ ] hrs_range = np.arange(6, 20, 1)
    emps_range = np.arange(2, 20, 1)
    avg_customers = df['Number_of_Customers_Per_Day'].mean()
    avg_order_val = df['Average_Order_Value'].mean()
    avg_marketing = df['Marketing_Spend_Per_Day'].mean()
    avg_traffic = df['Location_Foot_Traffic'].mean()

    best_rev = 0
    best_combo = None

    for h in hrs_range:
        for e in emps_range:
            x_new = np.array([[avg_customers,
                                avg_order_val,
                                h,
                                e,
                                avg_marketing,
                                avg_traffic]])

            # transform & prediksi
            rev_pred = pipe.predict(x_new)[0]
            if rev_pred > best_rev:
                best_rev = rev_pred
                best_combo = (h, e, rev_pred)

    print(f"\nPrediksi kombinasi terbaik: Jam operasional = {best_combo[0]} jam, Karyawan = {best_combo[1]} orang → Prediksi Daily Revenue = {best_combo[2]:.2f}")
```

- Mengambil dan menjelaskan pengaruh variabel pengeluaran pemasaran (Marketing_Spend_Per_Day) terhadap pendapatan harian berdasarkan koefisien model regresi linear. Pertama, baris kode mengambil nilai koefisien khusus untuk fitur tersebut dari tabel coef_df, lalu menyimpannya ke variabel coef_marketing. Setelah itu, program mencetak nilai koefisien tersebut sekaligus memberikan interpretasi bahwa penambahan 1 unit pada pengeluaran pemasaran (dalam skala standar, karena data sudah distandarisasi sebelum training) akan mengubah nilai Daily_Revenue sebesar koefisien tersebut. Langkah ini bertujuan untuk memahami seberapa besar dampak pemasaran terhadap pendapatan dan apakah pengaruhnya signifikan dalam model, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan alokasi anggaran marketing.

```
[ ] coef_marketing = coef_df[coef_df['feature']=='Marketing_Spend_Per_Day']['coefficient'].values[0]
    print(f"\nKoefisien pemasaran (Marketing_Spend_Per_Day) = {coef_marketing:.4f}")
    print("Artinya: penambahan 1 unit pengeluaran pemasaran (skala standar) memberi perubahan ±", coef_marketing,
          "unit Daily_Revenue (skala standar) – interpretasi harus hati-hati karena fitur sudah distandar.")

▼ Koefisien pemasaran (Marketing_Spend_Per_Day) = 218.9023
   Artinya: penambahan 1 unit pengeluaran pemasaran (skala standar) memberi perubahan ± 218.90226126025698 unit Dail
```

- Melakukan simulasi kenaikan anggaran pemasaran sebesar 10%, 20%, 50%, dan 100% dari nilai rata-rata marketing saat ini. Untuk setiap persentase, dihitung nilai marketing baru

(new_marketing). Kemudian dibuat satu baris data baru (x_new) berisi nilai rata-rata fitur lain serta kombinasi jam operasional dan jumlah karyawan terbaik (yang sebelumnya sudah dihitung). Model kemudian memprediksi pendapatan harian berdasarkan skenario marketing yang sudah dinaikkan tersebut. Hasil prediksi dicetak agar kita bisa melihat bagaimana perubahan anggaran marketing memengaruhi perkiraan pendapatan. Tujuannya adalah untuk membantu memahami sensitivitas pendapatan terhadap investasi pemasaran dan menentukan seberapa efektif tambahan biaya marketing dalam meningkatkan revenue.

```
[ ] for pct in [10, 20, 50, 100]:
    new_marketing = avg_marketing * (1 + pct/100)
    x_new = np.array([[avg_customers,
                       avg_order_val,
                       best_combo[0],
                       best_combo[1],
                       new_marketing,
                       avg_traffic]])
    rev_pred = pipe.predict(x_new)[0]
    print(f"Jika Marketing_Spend naik {pct}% → Prediksi Daily Revenue ≈ {rev_pred:,.2f}")

▼ Jika Marketing_Spend naik 10% → Prediksi Daily Revenue ≈ 1,965.92
  Jika Marketing_Spend naik 20% → Prediksi Daily Revenue ≈ 2,004.74
  Jika Marketing_Spend naik 50% → Prediksi Daily Revenue ≈ 2,121.17
  Jika Marketing_Spend naik 100% → Prediksi Daily Revenue ≈ 2,315.23
  /usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not ha
    warnings.warn(
  /usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not ha
    warnings.warn(
  /usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not ha
    warnings.warn(
  /usr/local/lib/python3.12/dist-packages/sklearn/utils/validation.py:2739: UserWarning: X does not ha
    warnings.warn(
```

- Menganalisis pengaruh foot traffic terhadap pendapatan harian, lalu mensimulasikan bagaimana kenaikan foot traffic akan memengaruhi prediksi revenue berdasarkan model regresi linear. Pertama, program mengambil koefisien foot traffic dari hasil regresi untuk mengetahui seberapa besar pengaruh variabel ini terhadap pendapatan koefisien yang positif menunjukkan bahwa semakin tinggi jumlah orang yang melintas di area lokasi, semakin besar pula potensi peningkatan revenue. Setelah itu, kode melakukan simulasi dengan menaikkan foot traffic sebesar 10%, 20%, 50%, dan 100% dari nilai rata-ratanya. Dalam setiap skenario, fitur lain dianggap tetap (konstan), termasuk jam operasional dan jumlah karyawan yang sebelumnya sudah ditentukan sebagai kombinasi terbaik. Model kemudian memprediksi pendapatan untuk setiap kondisi foot traffic yang dinaikkan tersebut. Hasil akhirnya memberi gambaran praktis tentang seberapa sensitif revenue terhadap perubahan foot traffic, sehingga

dapat membantu pengambilan keputusan bisnis terkait pemilihan lokasi dan strategi promosi di area sekitar toko.

```
[ ] coef_traffic = coef_df[coef_df['feature']=='Location_Foot_Traffic']['coefficient'].values[0]
    print(f"\nKoefisien foot traffic (Location_Foot_Traffic) = {coef_traffic:.4f}")
    print("Artinya: semakin tinggi foot traffic maka prediksi revenue semakin tinggi (dengan asumsi fitur lain konstan).")

    print("\nSimulasi variasi foot traffic:")
    for pct in [10,20,50,100]:
        new_traffic = avg_traffic * (1 + pct/100)
        x_new = np.array([avg_customers,
                          avg_order_val,
                          best_combo[0],
                          best_combo[1],
                          avg_marketing,
                          new_traffic])
        rev_pred = pipe.predict(x_new)[0]
        print(f"\nJika Foot_Traffic naik {pct}% → Prediksi Daily Revenue ≈ {rev_pred:,.2f}")
```

▼

```
Koefisien foot traffic (Location_Foot_Traffic) = 4.7672
Artinya: semakin tinggi foot traffic maka prediksi revenue semakin tinggi (dengan asumsi fitur lain konstan).
```

BAB IV

PENUTUP

4.1 Kesimpulan

Model hasil estimasi menunjukkan bahwa jumlah pelanggan harian dan nilai transaksi per pelanggan merupakan dua determinan utama yang berkontribusi terhadap peningkatan pendapatan. Pengeluaran pemasaran juga memberikan dampak positif, meskipun efeknya relatif lebih kecil. Sementara itu, faktor-faktor seperti foot traffic, jam operasional, dan jumlah karyawan memiliki pengaruh yang lebih terbatas. Simulasi operasional yang dilakukan mengindikasikan adanya kombinasi jam kerja dan jumlah staf tertentu yang memberikan hasil pendapatan optimum, tetapi tidak menjadi faktor penentu utama jika dibandingkan dengan pelanggan dan nilai transaksi.

Secara keseluruhan, penelitian ini menegaskan bahwa strategi bisnis yang diarahkan untuk meningkatkan akuisisi pelanggan, meningkatkan nilai belanja per pelanggan, serta optimisasi program pemasaran memiliki potensi paling besar dalam mendorong pendapatan harian. Temuan ini menyediakan dasar empiris untuk pengambilan keputusan operasional dan perencanaan strategis yang lebih terukur.

4.2 Saran

Berdasarkan temuan penelitian, peningkatan jumlah pelanggan perlu menjadi fokus utama karena variabel ini terbukti memberikan kontribusi terbesar terhadap pendapatan harian. Pendekatan seperti penguatan kampanye digital, pemanfaatan program loyalitas, atau kerja sama dengan komunitas dan bisnis lokal dapat membantu memperluas jangkauan dan menarik lebih banyak pengunjung. Di sisi lain, nilai transaksi rata-rata juga memiliki peran penting sehingga strategi seperti pengembangan menu yang lebih variatif, penerapan teknik upselling, serta pemanfaatan analitik untuk memberikan rekomendasi produk yang lebih relevan dapat diterapkan untuk mendorong peningkatan pendapatan per pelanggan.

Efektivitas anggaran pemasaran perlu dioptimalkan melalui pemilihan kanal promosi yang memiliki tingkat konversi terbaik. Evaluasi rutin berbasis data diperlukan agar setiap investasi pemasaran memberikan dampak yang nyata dan terukur. Selain itu, aspek visibilitas fisik kedai tetap perlu diperhatikan, mengingat peningkatan foot traffic masih memberikan kontribusi terhadap

pertumbuhan pendapatan. Perbaikan signage, penataan area depan toko, serta kolaborasi dengan bisnis sekitar dapat menjadi langkah yang dapat dipertimbangkan untuk memperkuat posisi kedai di lingkungan sekitar.

Jam operasional dan jumlah karyawan sebaiknya dikelola dengan pendekatan yang menekankan efisiensi, mengingat pengaruh kedua variabel tersebut terhadap pendapatan relatif kecil. Penyesuaian jadwal kerja dan pengalokasian staf berdasarkan pola kunjungan yang paling ramai dapat membantu menjaga keseimbangan antara biaya operasional dan kualitas layanan. Terakhir, model prediktif yang dibangun dalam penelitian ini dapat dimanfaatkan secara berkelanjutan untuk menganalisis skenario bisnis, mendukung perencanaan anggaran, dan menghasilkan proyeksi pendapatan yang lebih akurat, sehingga membantu pengambilan keputusan operasional yang lebih adaptif dan berbasis data.

DAFTAR PUSTAKA

1. International Coffee Organization. (2024). *World coffee consumption report 2024*.
https://icocoffee.org/documents/cy2023-24/Coffee_Report_and_Outlook_December_2023_ICO.pdf
2. Grand View Research. (2024). *Cafe market size, share & trends analysis report, 2024–2030*.
<https://www.grandviewresearch.com/industry-analysis/cafe-market-report>
3. Business Research Insights. (2024). *Southeast Asia coffee shop market analysis*.
<https://www.businessresearchinsights.com>
4. Chakma, R. (2025). Operational determinants of daily revenue in coffee shops: A multiple linear regression approach. *Journal of Business Analytics*, 12(2), 45–59.
5. SharpSheets. (2024). *How much does a coffee shop make per day?* <https://sharpsheets.io>
6. FoodDrinkTalk. (2025). *How much should a coffee shop make a day? Daily revenue goals and factors*. <https://fooddrinktalk.pro/how-much-should-a-coffee-shop-make-a-day/>
7. Coffee Plus Three. (2025). *Marketing spend effectiveness in small coffee businesses*.
<https://coffeeplusthree.com>
8. Adeleke, A. (2019). *Marketing strategies of successful coffee shop owners* [Doctoral dissertation, Walden University]. ProQuest Dissertations Publishing.
<https://search.proquest.com/openview/17329d803f161c6674905def280119ad/1?pq-origsite=gscholar&cbl=18750&diss=y>
9. Charoenphol, K., & Barrett, N. (2019). *A study of factors influencing customer to select a café* [Master's thesis, Thammasat University].
https://ethesisarchive.library.tu.ac.th/thesis/2019/TU_2019_6102040232_12055_12062.pdf
10. Huang, Y., & Wood, B. (2017). *Identify successful sales and marketing strategies that affect customer loyalty in a small coffee shop*. Wintec Research Archive.
<https://researcharchive.wintec.ac.nz/id/eprint/6021/>
11. Kasimin, S. (2017). *Coffee shop's characteristic and factors influence the number of visitors and profit level of popular coffee shops in Banda Aceh city*. Jurnal Ilmiah Manajemen, Universitas Syiah Kuala.
<http://download.garuda.kemdikbud.go.id/article.php?article=1518688&val=3958>
12. KERU, Z. N. (2021). *Determinants and profitability of coffee market outlet choice among farmers in Murang'a County, Kenya* [Master's thesis, Kenyatta University].
<https://ir-library.ku.ac.ke/bitstream/123456789/22566/1/Determinants%20and%20Profitability%20.....pdf>
13. McManus, B. (2007). Nonlinear pricing in an oligopoly market: The case of specialty coffee. *The RAND Journal of Economics*, 38(2), 512–532.
<https://doi.org/10.1111/j.1756-2171.2007.tb00081.x>
14. Relihan, L. (2022). *Is online retail killing coffee shops? Estimating the winners and losers of online retail using customer transaction microdata*. London School of Economics Discussion Paper.
<https://eprints.lse.ac.uk/117805/>
15. Susetya, I., & Marhaeni, A. (2023). *Determinants of coffee shop business income in Denpasar City*. Seminar Nasional Inovasi dan Kewirausahaan.
<https://pdfs.semanticscholar.org/dc77/7cfd7241da8c37b1d86dcaaf0ca5d57c388a.pdf>

16. Tilahun, D. (2007). *Performance of coffee marketing co-operatives and members' satisfaction in Dale District: SNNPRS-Southern Ethiopia*. International Livestock Research Institute (ILRI). <https://cgspace.cgiar.org/items/c7dfcd39-eb39-4872-8859-f7d5c79bf0b8>