

OPTIMIZING SALES STRATEGIES:

Leveraging Monthly Shopping Patterns and Voucher **Preferences**

- 2602104806 DIANDRA NATHANIA AUWLIA
- 2602152515 RATU FARADIBA ADIAZAHRA
- 2602108760 VIOLA





DAFTAR ISI

- 1. PENDAHULUAN
- 2. **STUDI LITERATUR**
- 3. METODE PROYEK
- 4. HASIL IMPLEMENTASI
- 5. **KESIMPULAN**
- 6. REFERENSI
- 7. LAMPIRAN



LATAR BELAKANG

Dalam era digital dan persaingan bisnis yang semakin ketat, pemahaman mendalam tentang perilaku pelanggan dan strategi penjualan efektif adalah kunci keberhasilan. Dengan teknologi dan data yang melimpah, kesempatan untuk memanfaatkan informasi guna meningkatkan strategi penjualan dan meraih keunggulan kompetitif terbuka lebar.

Proyek "Optimizing Sales Strategies: Leveraging Monthly Shopping Patterns and Voucher Preferences" dipilih untuk mengetahui preferensi penggunaan voucher berdasarkan customer behavior, dengan fokus pada faktor seperti jenis kelamin dan lokasi tempat tinggal. Analisis peak days pada 10 hari pertama setiap bulan juga penting untuk membantu perusahaan dalam menentukan stok dan pemberian voucher kepada pengguna berdasarkan data yang sudah dianalisis.





PENDAHULUAN

TUJUAN & MANFAAT

Kami memilih proyek ini untuk mengumpulkan **data perilaku belanja pelanggan** selama **10 hari pertama pasca penggajian bulanan** dan memahami preferensi penggunaan voucher. Dengan data ini, kami akan mengidentifikasi faktor-faktor yang memengaruhi keputusan pembelian, dengan manfaat **peningkatan efisiensi penjualan dan pengambilan keputusan** berbasis data untuk meningkatkan daya saing.

METODE

Metode yang digunakan dalam proyek ini adalah Descriptive Analytics dan Linear Regression Cross-validation, yang bertujuan untuk memahami tren, pola, dan hubungan dalam data dengan cara yang mudah dipahami yang melibatkan deskripsi data, filtering, grouping, visualisasi, dan interpretasi. Tujuan utama cross-validation adalah untuk menghindari overfitting (ketika model terlalu "memorisasi" data pelatihan dan tidak dapat menggeneralisasi dengan baik pada data baru) dan untuk mendapatkan perkiraan yang lebih stabil tentang kinerja model.





Referensi Dataset

Referensi dataset yang digunakan berjudul "EDA and Visualizing Shopping Dataset" oleh Jackson Divakar R. Link Referensi (Kaggle):

Dataset (file.csv)

<u>kaggle.com/code/jacksondivakarr/eda-and-visualizing-shopping-dataset/input</u>

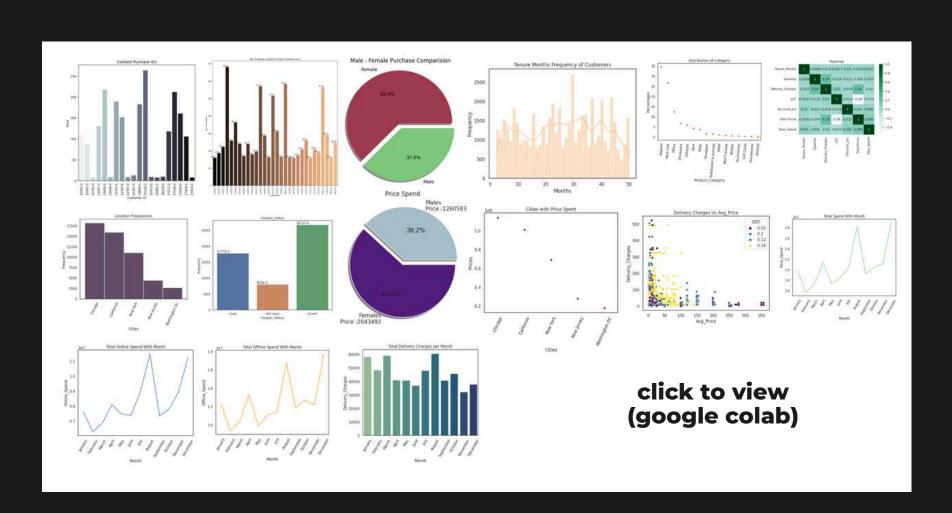
Code (Notebook)

<u>kaggle.com/code/jacksondivakarr/eda-and-visualizing-shopping-dataset/notebook</u>



STUDI LITERATUR

Hasil Analisis Referensi



Algoritma yang digunakan oleh referensi tersebut bertujuan untuk mengambil beberapa **analisis** antara lain:

- TOP 20 Costliest Purchase ID's,
- TOP 30 Popular Customer ID With Purchase Count
- Total Male-Female Purchase Comparison
- Tenure Months Frequency Of Customers
- Product Category Frequency Analysis
- Heatmap For Correlation
- Location Frequencies
- Coupon Status
- Male-Female Price Spent
- Cities With Price Spent
- Delivery Charges Vs Average Price
- Total Spend
- Online Spend
- Offline Spend
- Total Delivery Charges Per Month

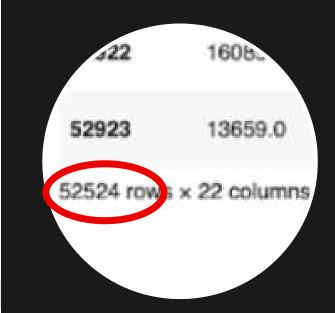




Filter data untuk hanya mencakup 10 hari pertama setiap bulan

```
[ ] data['Transaction_Date'] = pd.to_datetime(data['Transaction_Date'])
  data_filtered = data[data['Transaction_Date'].dt.day <= 10]</pre>
```

1.1	data																	
		CustomerID	Gender	Location	Tenure_Months	Transaction_ID	Transaction_Date	Product_SKU	Product_Description	Product_Category	Quantity		Coupon_Status	GST	Date	Offline_Spend	Online_Spend)	Mon
	0	17850.0	М	Chicago	12.0	16679.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainle	Nest-USA	1.0	***	Used	0.10	1/1/2019	4500.0	2424.50	
	1	17850.0	м	Chicago	12.0	16680.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainle	Nest-USA	1.0	(80)	Used	0.10	1/1/2019	4500.0	2424.50	
	2	17850.0	М	Chicago	12.0	16696.0	2019-01-01	GGOENEBQ078999	Nest Cam Outdoor Security Camera - USA	Nest-USA	2.0		Not Used	0.10	1/1/2019	4500.0	2424.50	
	3	17850.0	м	Chicago	12.0	16699.0	2019-01-01	GGOENEBQ079099	Nest Protect Smoke + CO White Battery Alarm- USA	Nest-USA	1.0	-	Clicked	0.10	1/1/2019	4500.0	2424.50	
	4	17850.0	М	Chicago	12.0	16700.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainle	Nest-USA	1.0	***	Clicked	0.10	1/1/2019	4500.0	2424.50	
	***	295	255	1085	100	596	800	lett	98	1990	- +0	100		100		200	0.00	
	52919	13155.0	F	California	8.0	22504.0	2019-03-10	GGOEGGCX056399	Gift Card - \$250.00	Gift Cards	1.0	-	Clicked	0.05	3/10/2019	2500.0	1294.22	
	52920	18077.0	м	Chicago	34.0	24250.0	2019-03-28	GGOEGGCX056299	Gift Card - \$25.00	Gift Cards	1.0	(444)	Used	0.05	3/28/2019	2000.0	1066.12	
	52921	16085.0	М	California	15.0	39991.0	2019-10-06	GGOEGOCD078399	Google Leather Perforated Journal	Notebooks & Journals	1.0	44	Clicked	0.05	10/6/2019	3000.0	2230.76	
	52922	16085.0	М	California	15.0	39991.0	2019-10-06	GGOEGOCR078499	Google Spiral Leather Journal	Notebooks & Journals	1.0	120	Used	0.05	10/6/2019	3000.0	2230.76	
	52923	13659.0	F	Chicago	8.0	39998.0	2019-10-06	GGOEGOCC077999	Google Spiral Journal with Pen	Notebooks & Journals	1.0	12	Not Used	0.05	10/6/2019	3000.0	2230.76	
- 8	52524 ro	ws × 22 column																-77



Data sebelum di filter menjadi 10 hari pertama di setiap bulan





Filter data untuk hanya mencakup 10 hari pertama setiap bulan

```
[ ] data['Transaction_Date'] = pd.to_datetime(data['Transaction_Date'])
  data_filtered = data[data['Transaction_Date'].dt.day <= 10]</pre>
```

data_fi	iltered																
	CustomerID	Gender	Location	Tenure_Months	Transaction_ID	Transaction_Date	Product_SKU	Product_Description	Product_Category	Quantity		Coupon_Status	GST	Date	Offline_Spend	Online_Spen	đ
0	17850.0	м	Chicago	12.0	16679.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainle	Nest-USA	1.0		Used	0.10	1/1/2019	4500.0	2424.5	0
1	17850.0	м	Chicago	12.0	16680.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainle	Nest-USA	1.0	122	Used	0.10	1/1/2019	4500.0	2424.5	0
2	17850.0	М	Chicago	12.0	16696.0	2019-01-01	GGOENEBQ078999	Nest Cam Outdoor Security Camera - USA	Nest-USA	2.0	344	Not Used	0.10	1/1/2019	4500.0	2424.5	0
3	17850.0	м	Chicago	12.0	16699.0	2019-01-01	GGOENEBQ079099	Nest Protect Smoke + CO White Battery Alarm- USA	Nest-USA	1.0	998	Clicked	0.10	1/1/2019	4500.0	2424.5	0
4	17850.0	М	Chicago	12.0	16700.0	2019-01-01	GGOENEBJ079499	Nest Learning Thermostat 3rd Gen-USA - Stainle	Nest-USA	1.0	- 11	Clicked	0.10	1/1/2019	4500.0	2424.5	0
	222	(iii)	222		4.0	1 (32)	- 1	72	Tana	122	1	724	- 55	44			nie.
52903	16367.0	М	California	20.0	27406.0	2019-05-06	GGOEWCKQ085457	Waze Pack of 9 Decal Set	Accessories	1.0		Used	0.10	5/6/2019	3000.0	2100.8	19
52919	13155.0	F	California	8.0	22504.0	2019-03-10	GGOEGGCX056399	Gift Card - \$250.00	Gift Cards	1.0	100	Clicked	0.05	3/10/2019	2500.0	1294.2	2
52921	16085.0	м	California	15.0	39991.0	2019-10-06	GGOEGOCD078399	Google Leather Perforated Journal	Notebooks & Journals	1.0	122	Clicked	0.05	10/6/2019	3000.0	2230.7	6
52922	16085.0	м	California	15.0	39991.0	2019-10-06	GGOEGOCR078499	Google Spiral Leather Journal	Notebooks & Journals	1.0	140	Used	0.05	10/6/2019	3000.0	2230.7	6
52923	13659.0	F	Chicago	8.0	39998.0	2019-10-06	GGOEGOCC077999	Google Spiral Journal with Pen	Notebooks & Journals	1.0	-	Not Used	0.05	10/6/2019	3000.0	2230.7	6
16865 rov	ws × 22 columns																

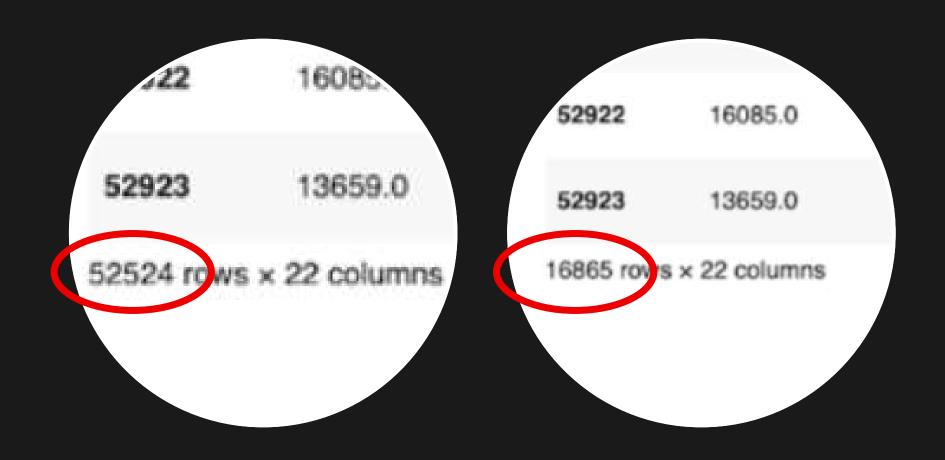


Data setelah di filter menjadi 10 hari pertama di setiap bulan





Filter data untuk hanya mencakup 10 hari pertama setiap bulan



Dari filter data disamping, dapat disimpulkan bahwa data awal yang berjumlah 52524 rows di filterisasi dan berkurang menjadi 16865 rows saja.

Tujuan dari adanya filter data ini adalah agar kita menganalisis data selama 10 hari saja di setiap bulannya. Alasan hanya menggunakan data pada periode tersebut karena merupakan periode dimana pekerja cenderung mendapatkan gaji mereka, sehingga bisa dikatakan sebagai peak days, dimana tingkat pembelian sedang melambung tinggi.



```
[ ] jppg = data_filtered.groupby(['Gender', 'Month'])['Quantity'].sum().reset_index()

sb.barplot(x='Month', y='Quantity', hue='Gender', data=jppg, palette='pastel')

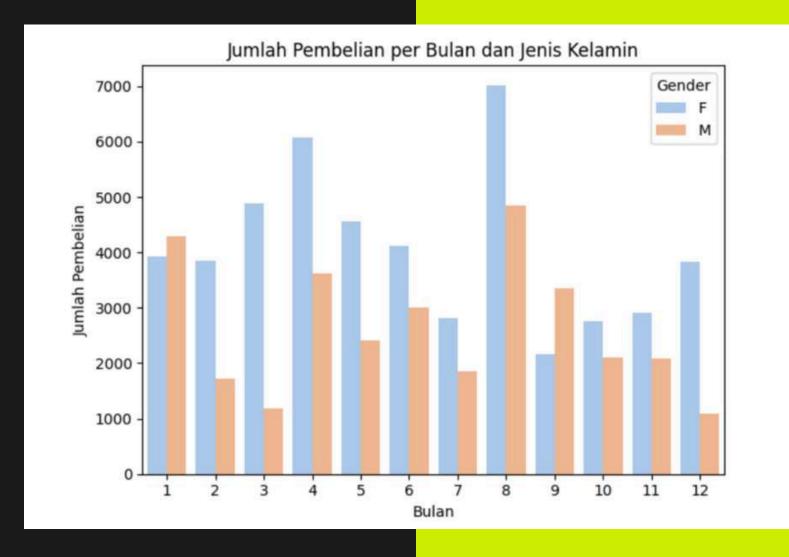
plt.title('Jumlah Pembelian per Bulan dan Jenis Kelamin')

plt.xlabel('Bulan')

plt.ylabel('Jumlah Pembelian')

plt.legend(title='Gender')

plt.show()
```





Menghitung Jumlah Pembelian

yang dilakukan oleh pria dan wanita selama periode tersebut

Dari data tersebut, dapat kita analisis bahwa **Female atau perempuan cenderung melakukan lebih banyak pembelian per bulannya.**

Hal ini ditunjukkan pada lebih tingginya angka jumlah pembelian oleh kelamin Female atau perempuan pada bulan 2, bulan 3, bulan 4, bulan 5, bulan 6, bulan 7, bulan 8, bulan 10, bulan 11, dan bulan 12. Sedangkan sisanya yaitu pada bulan 1 dan bulan 9, angka jumlah pembelian didominasi oleh Male atau pria. Namun perbandingannya adalah 10 : 2 dimana Female atau perempuan lebih banyak melakukan pembelian. Hal tersebut dapat memberi kesimpulan bahwa perempuan cenderung lebih banyak melakukan pembelian.

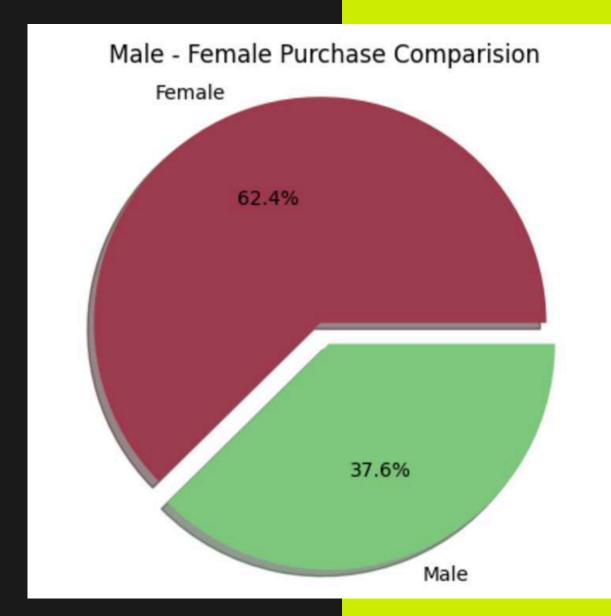


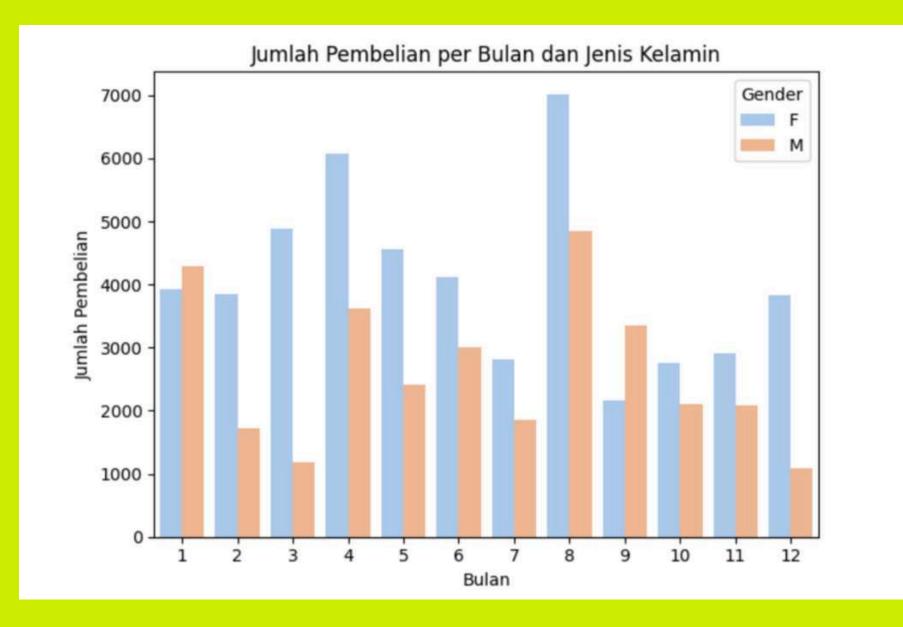
Perbandingan & Peningkatan Detail Analisis



sebelum

sesudah







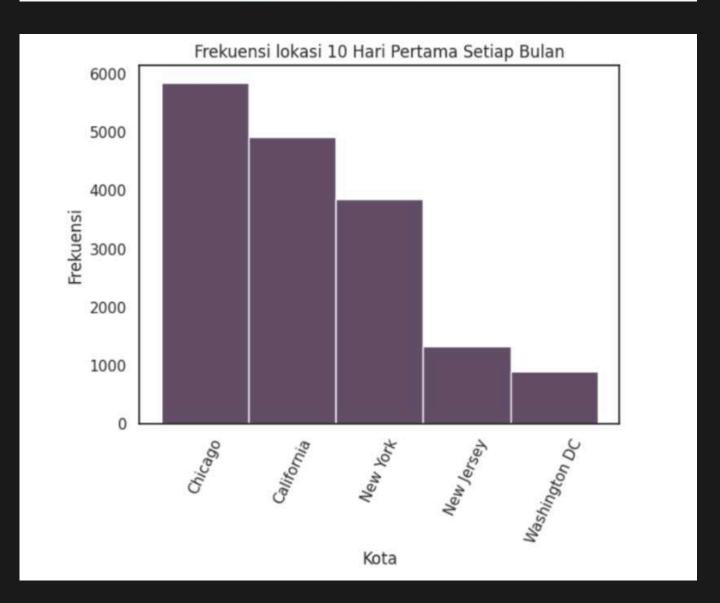


Menganalisis lokasi pembelian selama periode tersebut

Dari hasil visualisasi data tersebut, dapat kita analisis dan simpulkan bahwa di 10 hari pertama di setiap bulan, persebaran lokasi pembeli paling banyak berasal dari Chicago, dan kemudian disusul oleh California, New York, New Jersey, dan terakhir Washington DC.

Analisis frekuensi ini dapat memberikan asumsi bahwa Chicago merupakan kota dengan tingkat impulsif yang tinggi karena memiliki jumlah pembelian tertinggi. Hal ini dapat berguna bagi perusahaan untuk lebih banyak memberikan manfaat atau fasilitas seperti voucher atau kupon kepada user dengan domisili kota tersebut.

```
[ ] sb.histplot(data_filtered.Location,color=c2(0.5))
   plt.ylabel('Frekuensi')
   plt.xlabel('Kota')
   plt.xticks(rotation=65)
   plt.title('Frekuensi lokasi 10 Hari Pertama Setiap Bulan')
```



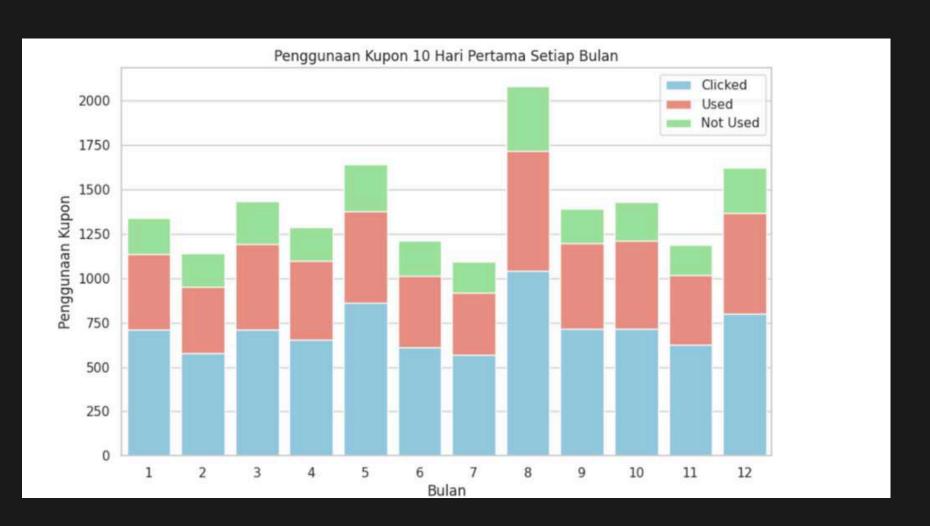




Menghitung jumlah penggunaan kupon selama periode tersebut

Penggunaan kupon meliputi parameter seperti clicked, used, dan not used.

Kode Tipe 1







Menghitung jumlah penggunaan kupon selama periode tersebut

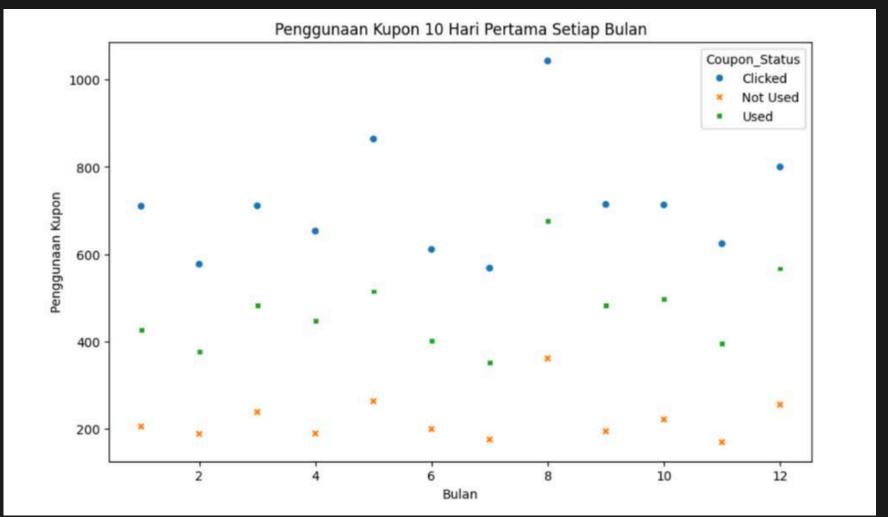
Penggunaan kupon meliputi parameter seperti clicked, used, dan not used.

Kode Tipe 2

```
[ ] coupon_usage = data_filtered.groupby(['Month', 'Coupon_Status']).size().unstack()
    plt.figure(figsize=(10, 6))
    sb.scatterplot(data=coupon_usage, marker='o')

plt.xlabel('Bulan')
    plt.ylabel('Penggunaan Kupon')
    plt.title('Penggunaan Kupon 10 Hari Pertama Setiap Bulan')

plt.show()
```





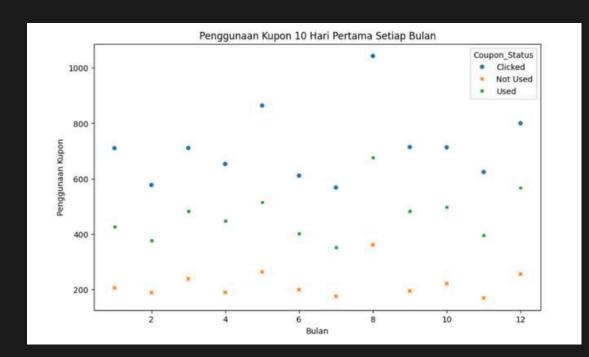


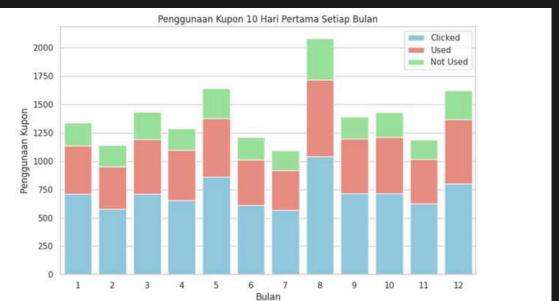
Menghitung jumlah penggunaan kupon selama periode tersebut

Penggunaan kupon meliputi parameter seperti clicked, used, dan not used.

Dari hasil data tersebut, dapat kita analisis bahwa penggunaan kupon atau coupon status cenderung pada 10 hari pertama di setiap bulan lebih banyak diklik atau clicked dibandingkan digunakan atau used dan tidak digunakan atau not used.

Hal ini dapat disebabkan oleh beberapa faktor seperti kurang relevannya penggunaan kupon dengan kebutuhan pengguna. Pada visualisasi data tersebut, yaitu dalam bentuk bar chart dan dot, keduanya menyimpulkan bahwa urutan penggunaan atau status kupon dari yang terbanyak adalah clicked, used, kemudian not used.



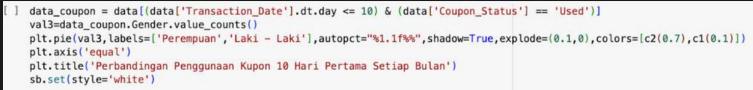






Menganalisis preferensi penggunaan kupon berdasarkan jenis kelamin

Dari analisis tersebut, dapat disimpulkan bahwa visualisasi data dalam bentuk pie chart tersebut menunjukkan perbandingan penggunaan kupon 10 hari pertama setiap bulan yang didominasi oleh jenis kelamin perempuan sebanyak 58.3% atau sekitar 9832 data dari dataset yang ada.







LINEAR REGRESSION

[] # Create some example data

```
np.random.seed(0)
     X = 2 * np.random.rand(100, 1)
    y = 4 + 3 * X + np.random.randn(100, 1)
[] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
[ ] model = LinearRegression()
     model.fit(X_train, y_train)
 ∓
     ▼ LinearRegression
     LinearRegression()
[ ] # Melakukan K-Fold Cross-Validation
     kf = KFold(n_splits=5, shuffle=True, random_state=42)
     mse_scorer = make_scorer(mean_squared_error)
[ ] # Mendapatkan hasil cross-validation
     cv_scores = cross_val_score(model, X_train, y_train, cv=kf, scoring=mse_scorer)
[ ] # Mencetak hasil cross-validation
     print(f"Cross-Validation MSE scores: {cv_scores}")
     print(f"Mean Cross-Validation MSE: {cv_scores.mean()}")
    print(f"Standard Deviation of Cross-Validation MSE: {cv_scores.std()}")
Tross-Validation MSE scores: [0.77419801 0.97646011 0.89300955 1.54408278 1.25879503]
    Mean Cross-Validation MSE: 1.0893090965014898
    Standard Deviation of Cross-Validation MSE: 0.2778877436226016
[ ] # Melatih model pada seluruh training set
     model.fit(X_train, y_train)
    # Melakukan prediksi pada test set
    y_pred = model.predict(X_test)
    # Evaluasi model pada test set
     mse_test = mean_squared_error(y_test, y_pred)
     r2_test = model.score(X_test, y_test)
     print(f"Mean Squared Error on Test Set: {mse_test}")
    print(f"R^2 Score on Test Set: {r2_test}")
→ Mean Squared Error on Test Set: 0.9177532469714291
    R^2 Score on Test Set: 0.7521157503858556
```

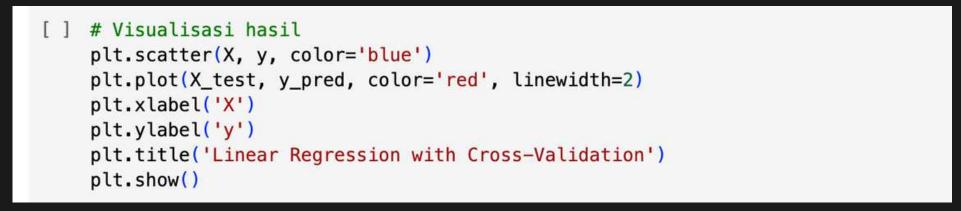


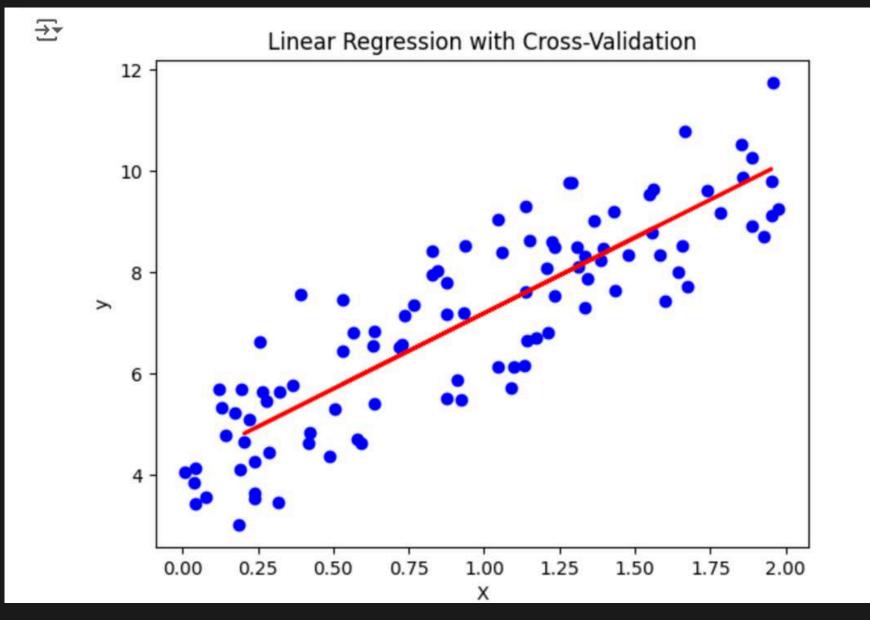
Mean Squared Error on Test Set: 0.9177532469714291
R^2 Score on Test Set: 0.6521157503858556



LINEAR REGRESSION







- Data Training: 13,492 data
- Data Validasi: 1,687 data
- Data Testing: 1,686 data



KESIMPULAN PROYEK



Hasil Implementasi

Hasil implementasi analisis pada dataset, ditemukan beberapa temuan penting:

01

02

03

04

05

Filter Data

Filter data dilakukan untuk hanya mencakup 10 hari pertama setiap bulan, yang merupakan periode dengan tingkat pembelian yang tinggi.

Jumlah Pembelian yang dilakukan oleh Pria dan Wanita

Perempuan cenderung melakukan lebih banyak pembelian daripada pria selama periode yang diteliti.

Jumlah Penggunaan Kupon

Penggunaan kupon cenderung lebih banyak dalam status "clicked" dibandingkan "used" atau "not used".

Penggunaan Kupon Berdasarkan Jenis Kelamin

Preferensi penggunaan kupon didominasi oleh jenis kelamin perempuan.

Lokasi Pembelian

Lokasi pembelian terbanyak berasal dari Chicago, menunjukkan tingkat impulsif yang tinggi dalam pembelian.



KESIMPULAN PROYEK





Peningkatan Hasil Analisis

Analisis ini lebih mendalam karena memfokuskan pada periode peak days, yaitu 10 hari pertama setiap bulan yang dianggap sebagai puncak aktivitas pembelian. Dengan mengambil data dari periode ini, analisis dapat mengamati tren dan pola pembelian yang lebih signifikan. Pemilihan data yang spesifik ini memungkinkan interpretasi yang lebih rinci, terutama terkait preferensi penggunaan kupon berdasarkan jenis kelamin dan lokasi pembelian.

Dengan demikian, fokus pada periode peak days memberikan dasar yang kuat untuk analisis yang lebih terperinci dan interpretasi yang lebih mendalam, yang pada akhirnya meningkatkan pemahaman tentang implikasi hasil analisis dalam konteks strategi penjualan.



REFERENSI



- <u>kaggle.com/code/jacksondivakarr/eda-and-visualizing-shopping-dataset/input</u>
- https://colab.research.google.com/drive/1lhGAgP0IT5SyuLniLexo0H5InOrbwF6H?usp=sharing
- <u>kaggle.com/code/jacksondivakarr/eda-and-visualizing-shopping-dataset/notebook</u>
- <u>kaggle.com/code/ibrahimelgmmal/online-shopping-analysis/notebook</u>
- <u>kaggle.com/code/ahmedismaiil/online-shopping-sales</u>
- https://www.jaspersoft.com/articles/what-is-descriptive-analytics



LAMPIRAN

SCREENSHOT HASIL ANALISIS



