

Pengantar Sains Data



SALES PREDICTION

Time Series Forecasting

GROUP 5

01

Komang Ryandhi S.
(5002211109)

01

Maulana Yusuf Ikhwan R.
(5002211167)

01

Radityo Ar Rasyid
(5002211034)

01

Maya Angelina
(5002211038)

DATASET SUMMARY

TRAIN.CSV

Memuat data **(date)** tanggal, **store_nbr** (toko penjualan), **family** (tipe produk), dan **onpromotion** yang akan memenuhi target **sales**

TEST.CSV

Data pengujian untuk memprediksi penjualan 15 hari kedepan

STORES.CSV

Store metadata, yang memuat **city**, **state**, **type**, and **cluster**

OIL.CSV

Harga minyak harian mencakup nilai selama kedua rentang waktu pada data train dan test

HOLIDAYS_EVENTS.CSV

Memuat data hari libur dan acara khusus

TRANSACTION.CSV

Data penjualan berdasarkan jenis toko (**store_nbr**) & tanggal (**date**)



TIME SERIES FORECASTING WITH XGBOOST DMATRIX

[Link Collab Kelompok Kami.](#)



[Link Kaggle Kelompok Kami](#)

kaggle



PREPROCESSING

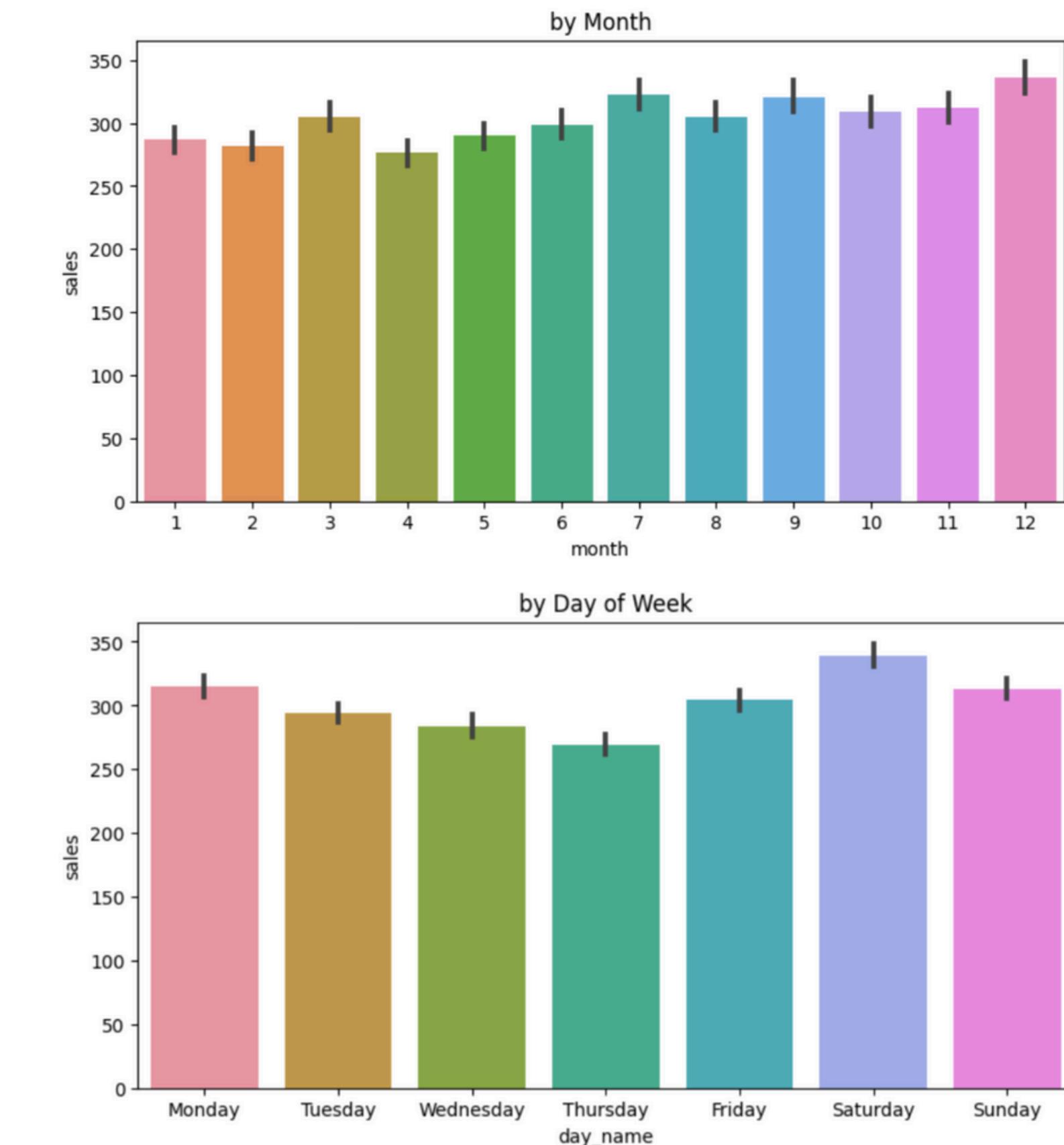
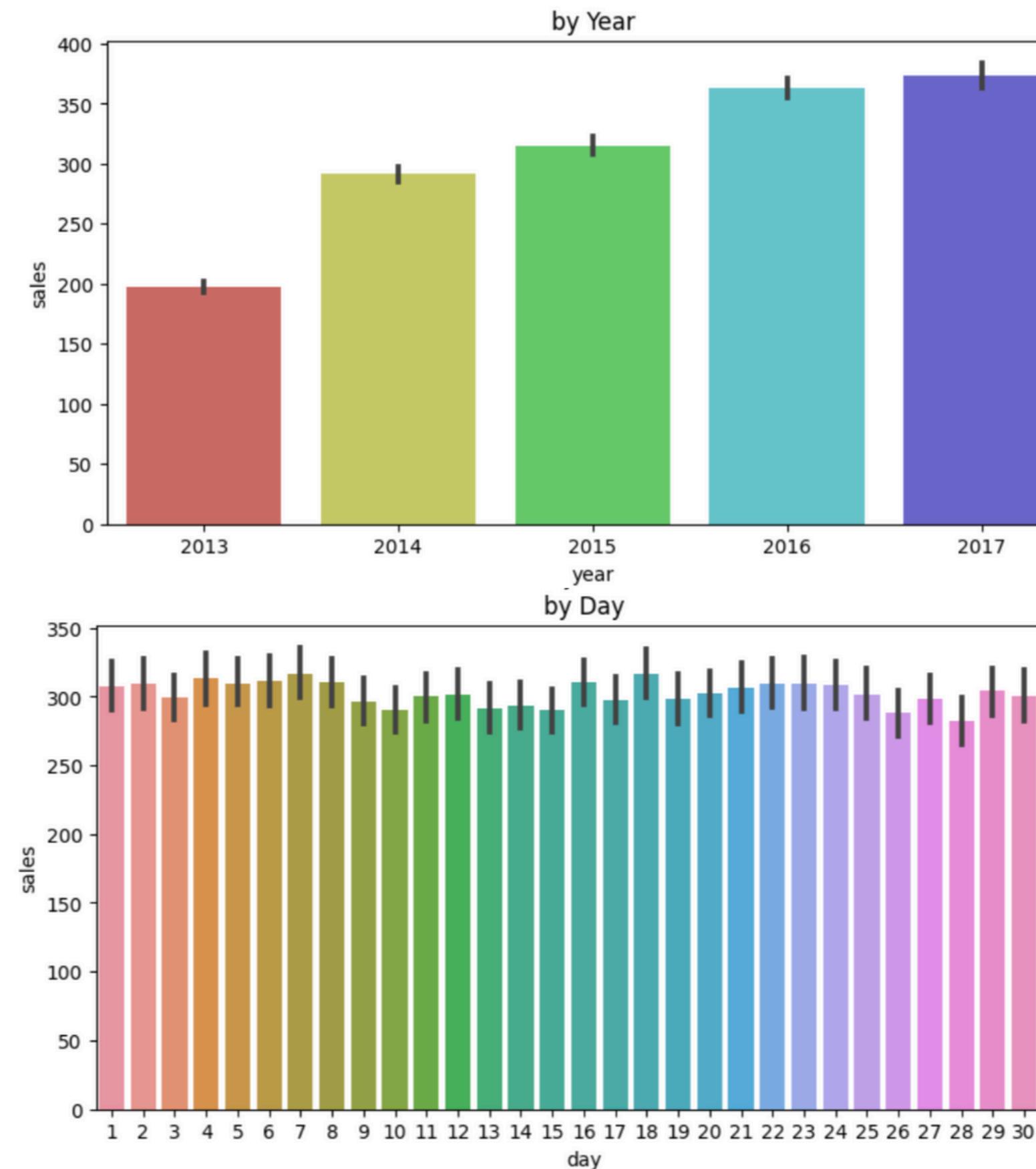
EXPLORATORY DATA ANALYSIS

Melalui proses Eksplorasi Data (EDA), Semua dataset telah digabungkan, kecuali data liburan. Dari penjualan hingga harga minyak, transaksi, dan informasi toko, integrasi dataset ini tentang pola penjualan dan faktor-faktor yang mempengaruhinya, untuk pengambilan keputusan.

	<code>id</code>	<code>date</code>	<code>store_nbr</code>	<code>onpromotion</code>	<code>dcoilwtico</code>	<code>transactions</code>	<code>city</code>	<code>state</code>	<code>type</code>	<code>cluster</code>	<code>holiday</code>	<code>year</code>	<code>month</code>	<code>day</code>	<code>day_name</code>	<code>new_family</code>	
0	3000888	2017-08-16	1	0	46.8		NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest
1	3000889	2017-08-16	1	0	46.8		NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest
2	3000890	2017-08-16	1	2	46.8		NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest
3	3000891	2017-08-16	1	20	46.8		NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	second
4	3000892	2017-08-16	1	0	46.8		NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest

RATA-RATA PENJUALAN

07



HASIL PENGAMATAN

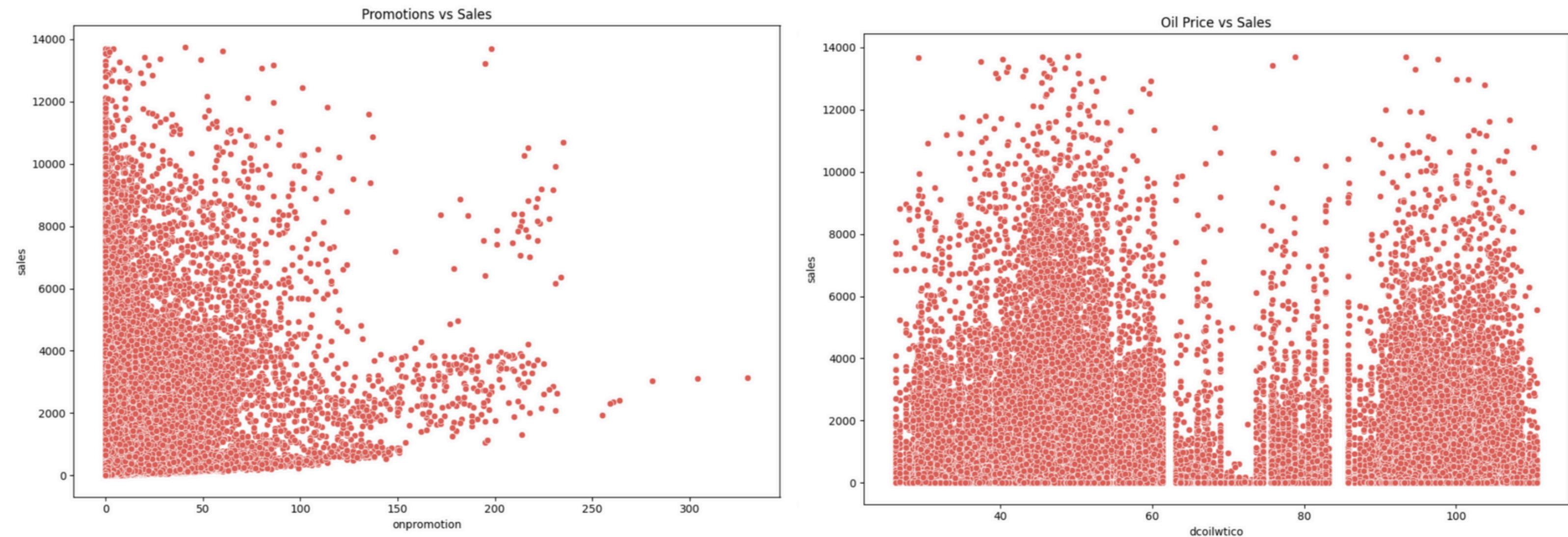
08

1. Penjualan terus meningkat dari 2013 sampai 2017
2. Penjualan lebih tinggi pada bulan Desember
3. Penjualan lebih tinggi pada awal bulan
4. Penjualan lebih tinggi pada hari libur



PENGARUH PROMOSI ATAU HARGA MINYAK TERHADAP SALES

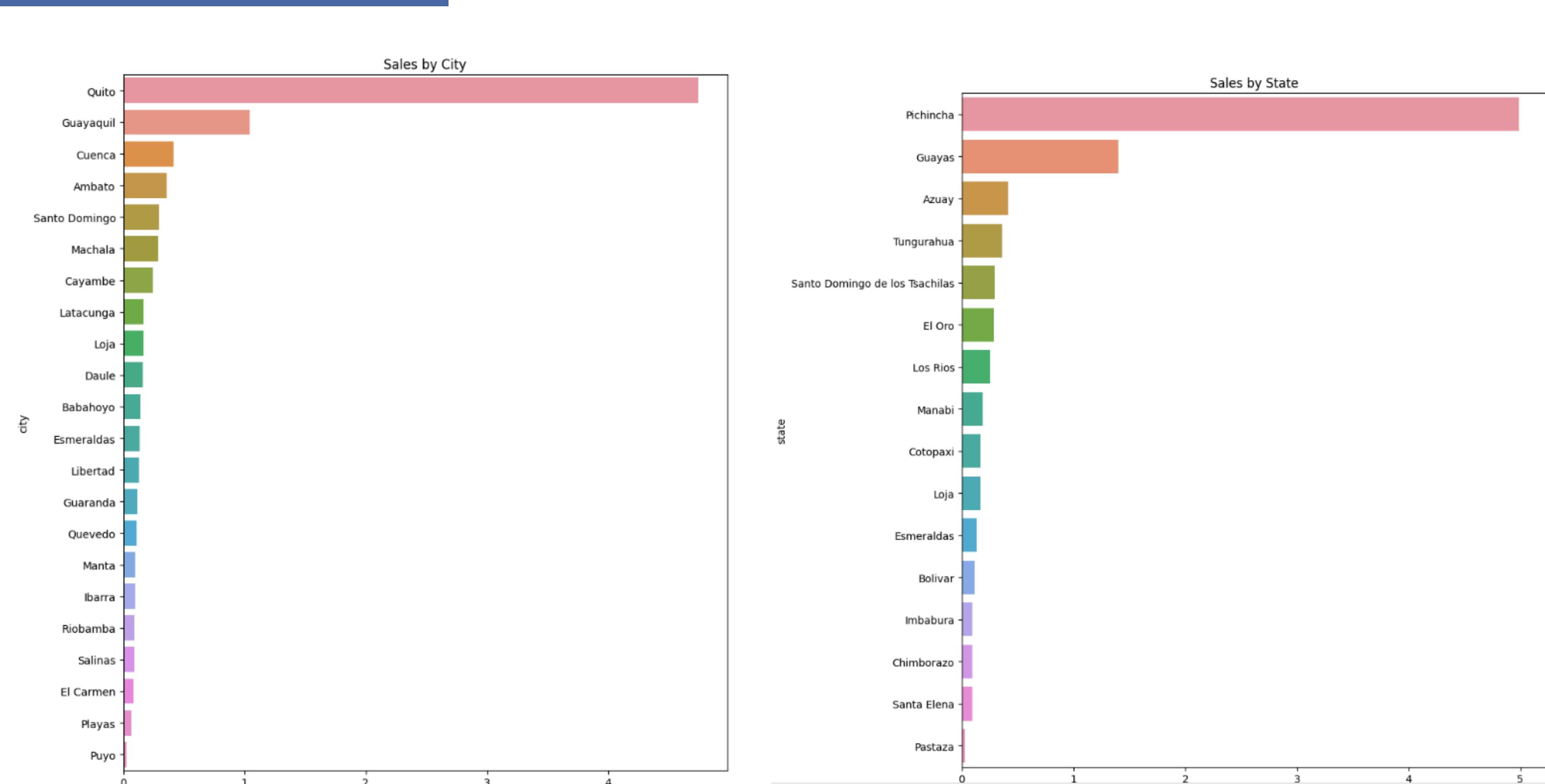
09



1. Ekuador adalah negara yang mayoritas komoditasnya minyak bumi. Harga minyak bumi kemungkinan besar berpengaruh pada penjualan
2. Promosi mempengaruhi banyaknya penjualan juga

PENJUALAN TERBANYAK

10





PREDICTIVE ANALYSIS

EVALUASI DENGAN PENGGABUNGAN DAN PEMROSESAN FITUR

12

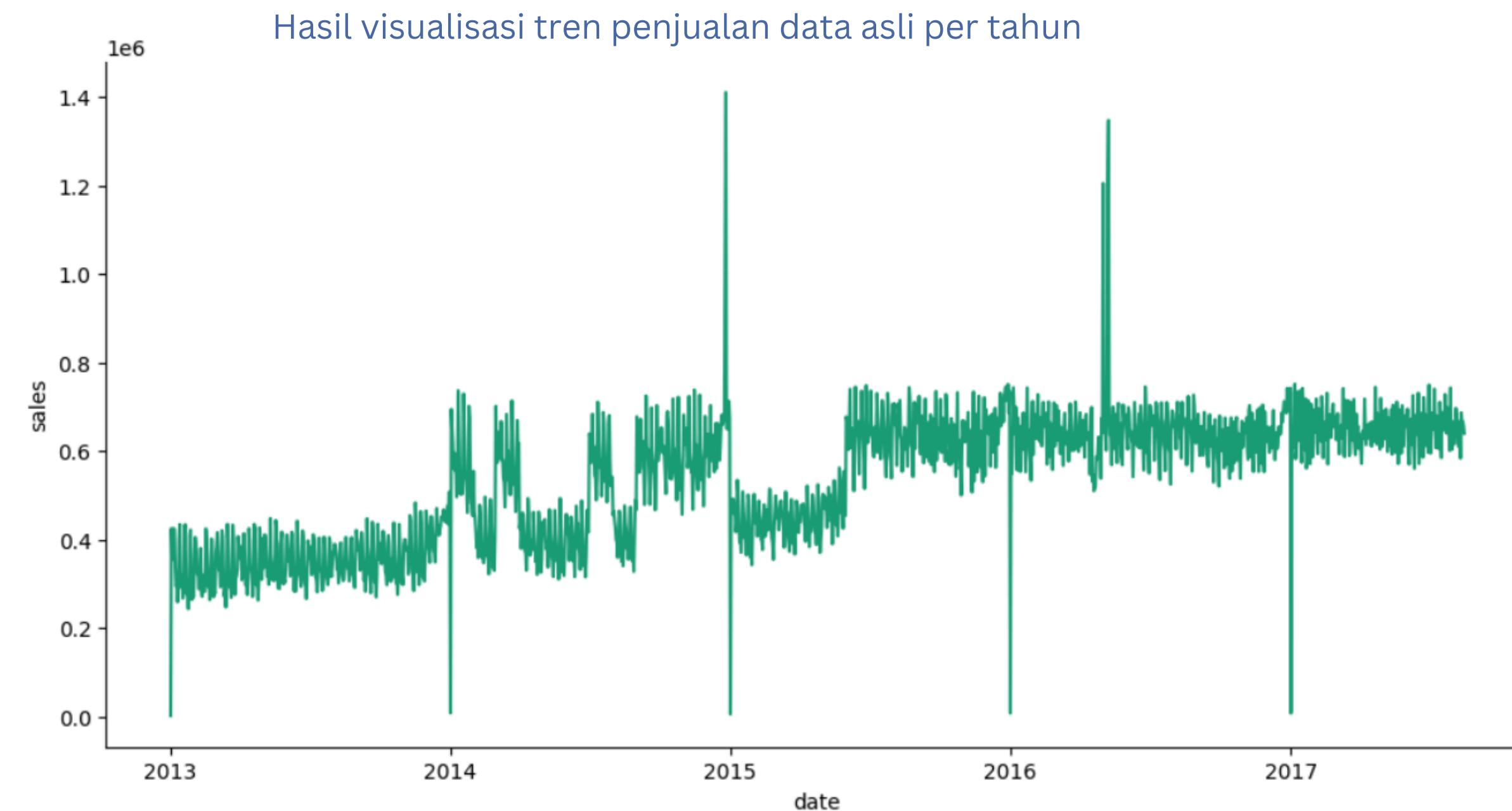
Melakukan persiapan data uji dengan menggabungkan beberapa DataFrames berdasarkan tanggal dan nomor toko, menambahkan fitur-fitur baru seperti harga minyak dan fitur tanggal, serta melakukan pemrosesan data.

	id	date	store_nbr	onpromotion	dcoilwtico	transactions	city	state	type	cluster	holiday	year	month	day	day_name	new_family
0	3000888	2017-08-16	1	0	46.8	NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest
1	3000889	2017-08-16	1	0	46.8	NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest
2	3000890	2017-08-16	1	2	46.8	NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest
3	3000891	2017-08-16	1	20	46.8	NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	second
4	3000892	2017-08-16	1	0	46.8	NaN	Quito	Pichincha	D	13	NaN	2017	8	16	Wednesday	rest

TREN PENJUALAN DARI WAKTU KE WAKTU

13

Secara keseluruhan, codingan ini menghasilkan plot dari data deret waktu 'sales' dengan sumbu x yang menunjukkan tanggal dan sumbu y yang menunjukkan nilai penjualan.



ROOT MEAN SQUARED LOG ERROR (RMSLE)

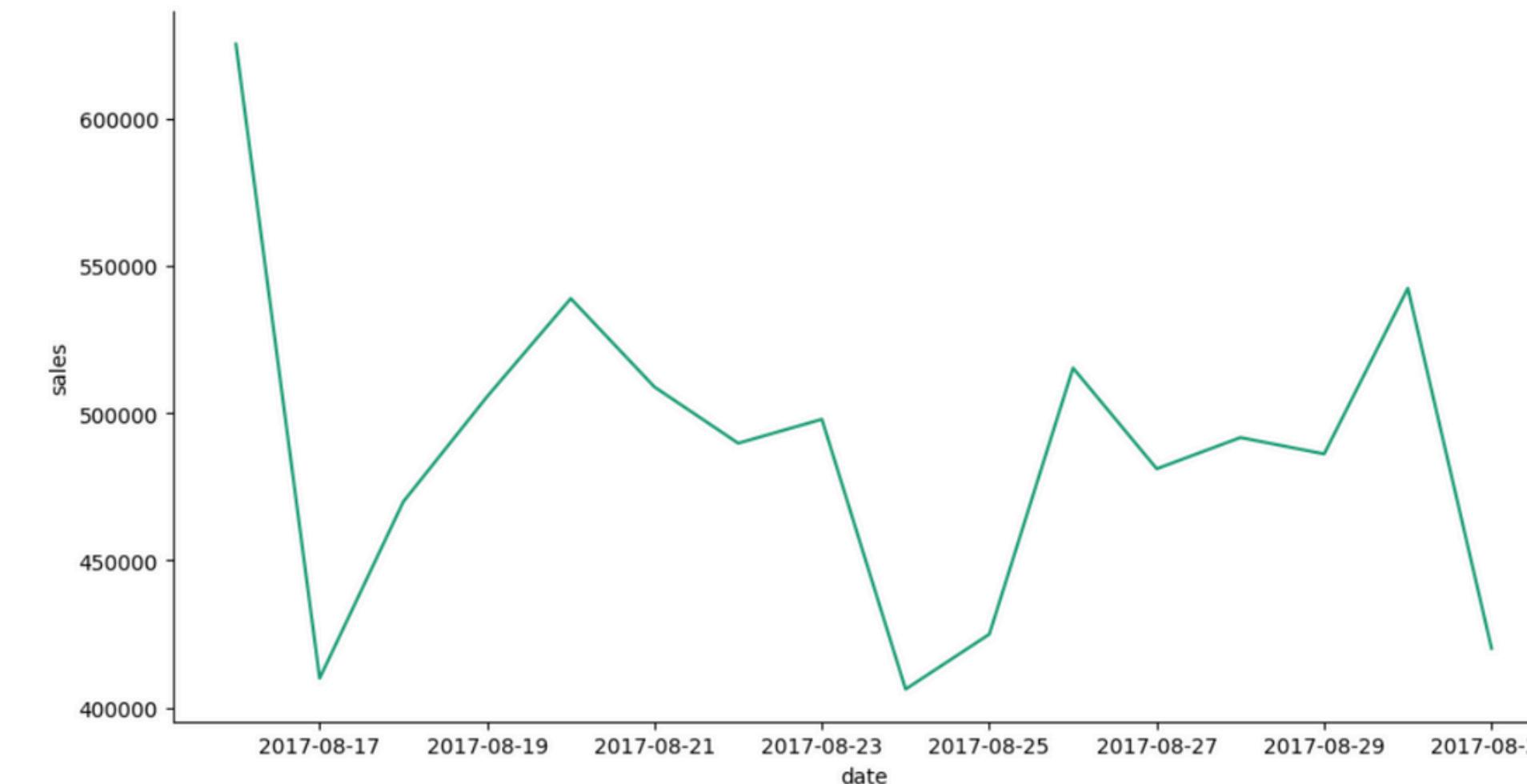
Menggunakan algoritma XGBoost untuk menghasilkan prediksi berdasarkan model regresi, kemudian mengukur kinerja model dengan Root Mean Squared Log Error (RMSLE) antara prediksi dan nilai sebenarnya.

```
[0]      train-rmse:670.36026    validation-rmse:675.33456
[248]     train-rmse:219.97495    validation-rmse:296.29460
RMSLE: 1.4651594766422502
```

VISUALISASI TREN PENJUALAN TERURUT

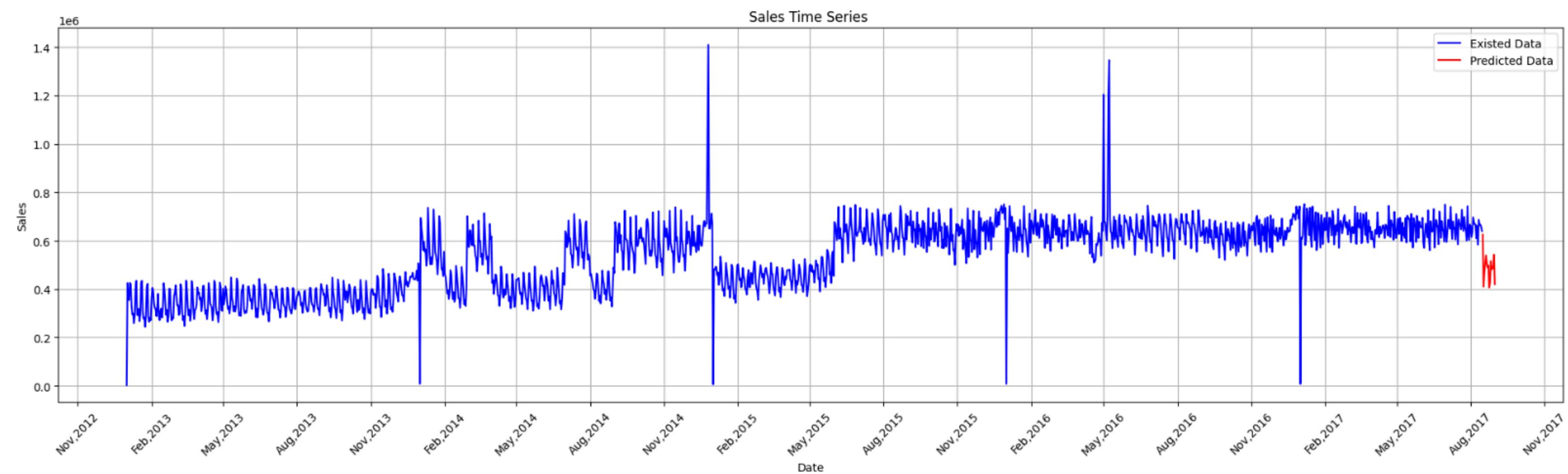
Ini adalah hasil prediksi untuk 15 hari kedepan yang dilakukan model XGBoost Reggresion

XGBoost digunakan untuk peramalan penjualan karena akurasinya yang tinggi dalam menangani kumpulan data besar dengan faktor eksternal, mengatasi masalah musim (seasonal) dan kelangkaan data(data scarcity secara efektif.



PENGARUH PROMOSI

Data berupa deret waktu (warna biru) digunakan untuk melatih model sehingga model dapat memprediksi penjualan 15 hari (warna merah).



THANK YOU
