

# **Take Home Test Data Science**

## **Food Delivery Time Prediction**

By : Ragil Setyo Utomo

# **Deskripsi singkat**

Nama : Ragil Setyo Utomo

Minat : Pengolahan data, visualisasi , serta prediksi dengan machine learning untuk mendukung keputusan bisnis.

Latar belakang : Data science, terbiasa dengan data dan mengubahnya menjadi insight yang dipahami.

Tujuan : Membangun karier yang berdampak positif di perusahaan



# **Education dan Pengalaman**

Saya lulusan S-1 Brawijaya teknik Informatika dan sudah mengikuti bootcamp dibimbing untuk data science.

Saya belum mempunyai pengalaman kerja sebagai data science, tetapi pada saat bootcamp, saya sudah mengerjakan proyek untuk memberikan insight dan melakukan prediksi dengan machine learning.



**Dibimbing**

# **Overview Project Sebelumnya**

Tujuan proyek untuk mengetahui prediksi harga perumahan di Washington. Didapatkan hasil model terbaik dengan menggunakan XGboost Regressor :

- Rata-rata prediksi meleset (MAE) = \$ 66,794 per rumah
- Rata-rata akar kuadrat prediksi meleset (RMSE) = \$ 95,761 per rumah
- R2 Score= 80.00%
- Rata-rata jumlah harga rumah = \$ 487,456.90

# Project Background

Dalam industri layanan pesan-antar makanan (food delivery), waktu pengiriman merupakan salah satu faktor terpenting yang memengaruhi kepuasan pelanggan. Pelanggan cenderung memberikan ulasan negatif jika pesanan datang terlambat, sedangkan pengiriman yang cepat dapat meningkatkan loyalitas dan reputasi platform.

Namun, waktu pengiriman dipengaruhi oleh berbagai faktor seperti jarak pengantaran, kondisi cuaca, lalu lintas, waktu dalam sehari, jenis kendaraan kurir, serta kecepatan restoran dalam menyiapkan makanan.

Dengan meningkatnya volume pesanan setiap hari, perusahaan perlu memiliki model prediktif yang dapat memperkirakan waktu pengiriman secara akurat, sehingga dapat membantu dalam :

- Mengoptimalkan penjadwalan kurir
- Memberikan estimasi waktu yang realistik kepada pelanggan, dan
- Meningkatkan efisiensi operasional.

Oleh karena itu, analisis ini bertujuan untuk membangun model prediksi waktu pengantaran makanan menggunakan berbagai faktor yang memengaruhi proses pengiriman.

# Problem Statement

1

Ketidakpuasan  
pelanggan karena  
pengiriman terlambat

2

Ketidakefisienan  
operasional akibat  
pembagian tugas kurir  
yang tidak optimal.

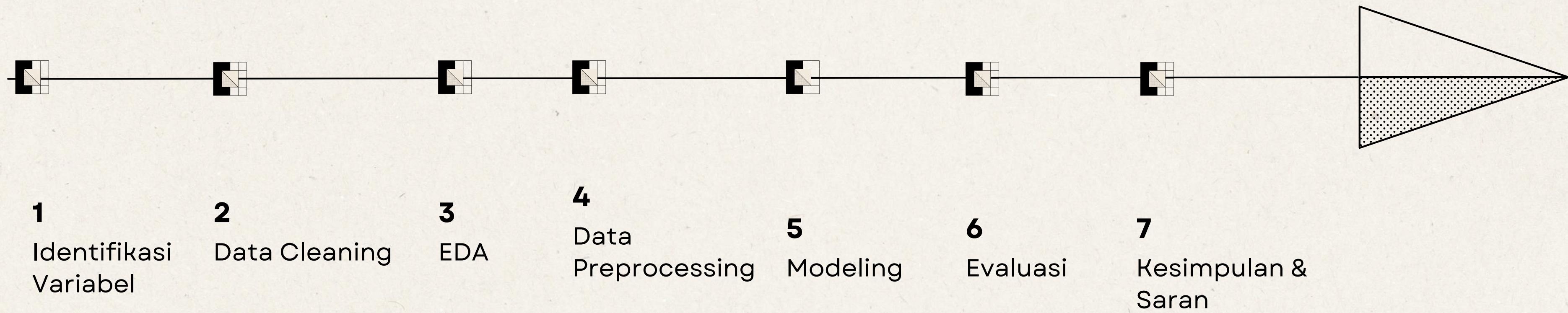
3

Bagaimana cara memprediksi  
waktu pengantaran makanan  
secara akurat berdasarkan  
faktor-faktor yang ada?

# Objective & Bussiness Question

Objective	<ol style="list-style-type: none"><li>1. Membangun model prediksi yang mampu memperkirakan waktu pengantaran (Delivery_Time_min) berdasarkan fitur-fitur yang relevan..</li><li>2. Memberikan insight bagi manajemen untuk meningkatkan efisiensi pengiriman dan pengalaman pelanggan.</li></ol>
Bussiness Question	<ol style="list-style-type: none"><li>1. Faktor utama apa yang paling memengaruhi waktu pengantaran makanan?</li><li>2. Apakah model prediksi yang dibangun dapat memberikan estimasi waktu pengantaran dengan tingkat kesalahan yang rendah?</li><li>3. Bagaimana hasil prediksi ini dapat dimanfaatkan untuk meningkatkan kepuasan pelanggan dan efisiensi operasional perusahaan?</li></ol>

# Alur Pemodelan



# Identifikasi Variabel

Penjelasan kolom yang ada :

- Order\_ID = Nomor unik untuk setiap pesanan makanan
- Distance\_km = Jarak antara restoran dan pelanggan dalam satuan kilometer. Semakin jauh jarak, biasanya waktu pengiriman lebih lama.
- Weather = Kondisi cuaca saat pengiriman, misalnya: Sunny, Rainy, Cloudy, Stormy, dll. Cuaca dapat memengaruhi waktu pengiriman.
- Traffic\_Level = Tingkat kepadatan lalu lintas, misalnya: Low, Medium, High. Lalu lintas padat biasanya memperlambat kurir.
- Time\_of\_Day = Waktu dalam sehari saat pesanan dikirim, seperti Morning, Afternoon, Evening, Night. Bisa berpengaruh terhadap lalu lintas dan waktu pengiriman.
- Vehicle\_Type = Jenis kendaraan yang digunakan oleh kurir,
- Preparation\_Time\_min = Waktu yang dibutuhkan restoran untuk menyiapkan makanan (dalam menit). Semakin lama waktu ini, semakin lama pula total waktu pengiriman.
- Courier\_Experience\_yrs = Lama pengalaman kurir dalam tahun. Kurir yang lebih berpengalaman biasanya bisa mengirim lebih efisien.
- Delivery\_Time\_min = Target/Output — total waktu pengiriman dari pemesanan hingga sampai ke pelanggan (dalam menit). Ini biasanya variabel yang ingin diprediksi.

Jumlah Baris : 1000

Jumlah kolom : 9

# **Data Cleaning**

Pengecekan Missing Value : Terdapat missing value

Pengecekan Duplikat data : Tidak ada duplikat pada dataset

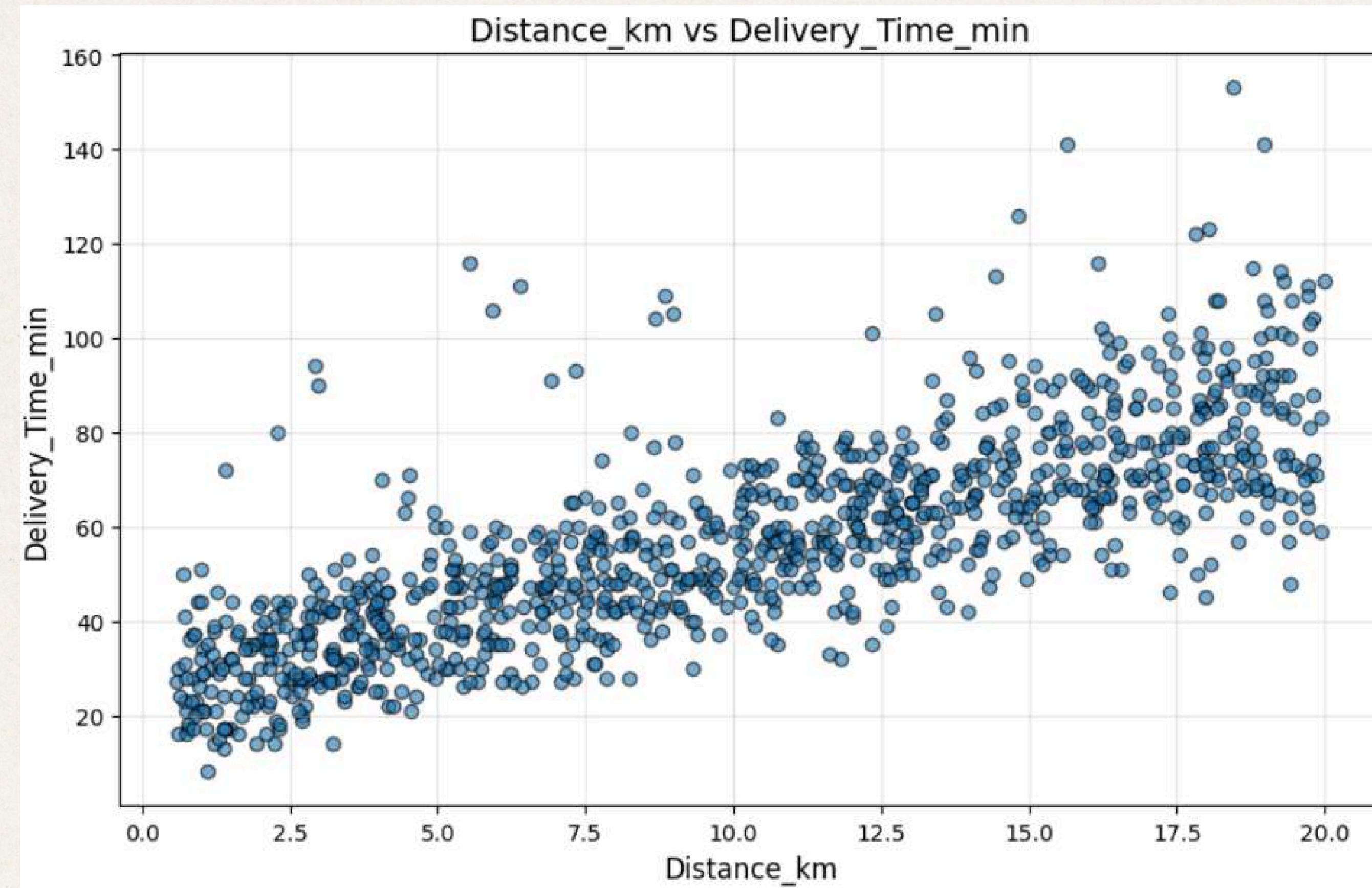
Pengecekan Outlier : Tidak terdapat outlier

# **Missing Value Handling**

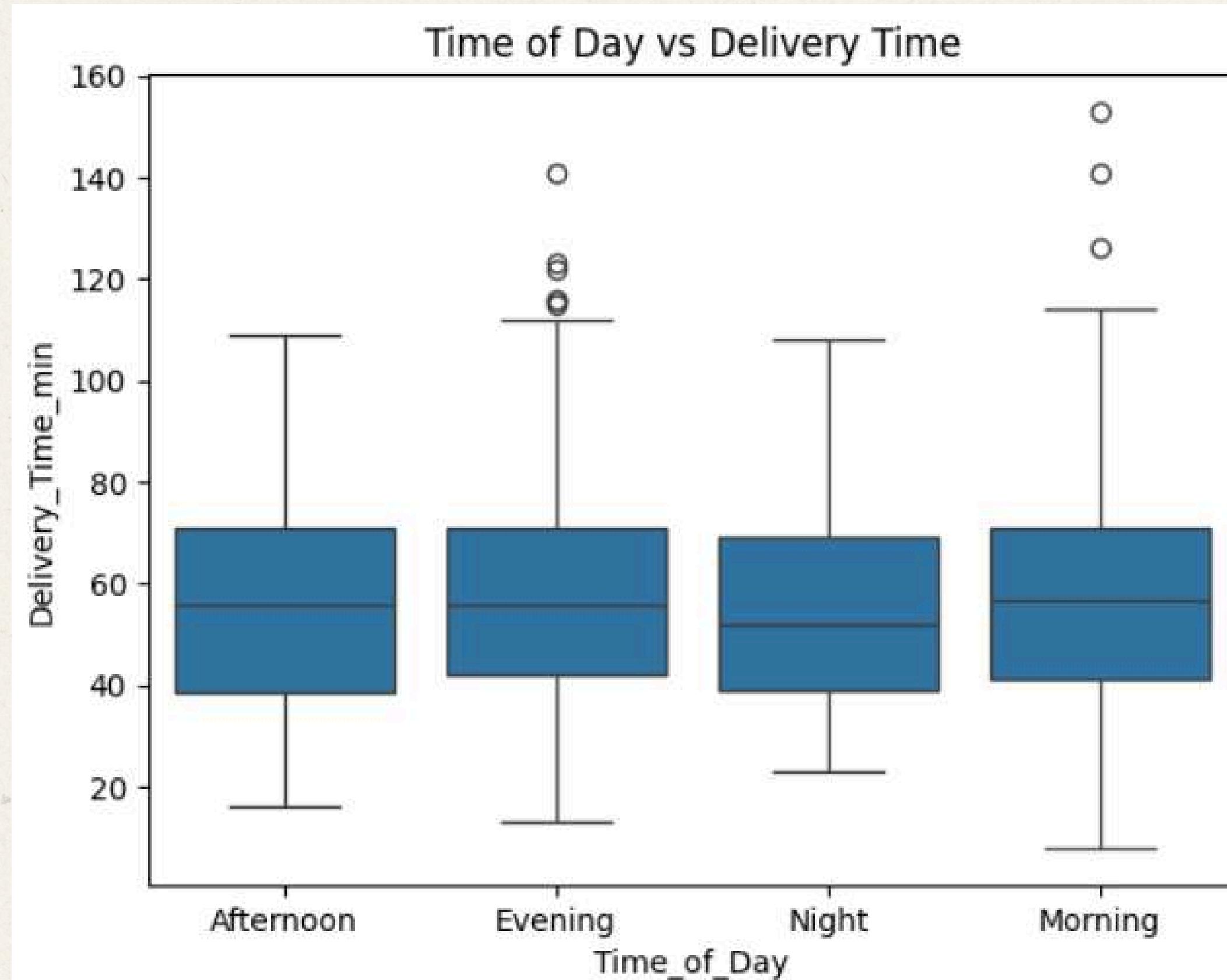
- Jumlah missing value tiap kolom = 30
- Pada kategorical : weather, Time\_of\_Day, Traffic\_Level akan dihandling dengan modus
- Pada numerical : Courier\_Experience\_yrs akan dihandling dengan median

# **Exploratory Data Analysis (EDA)**

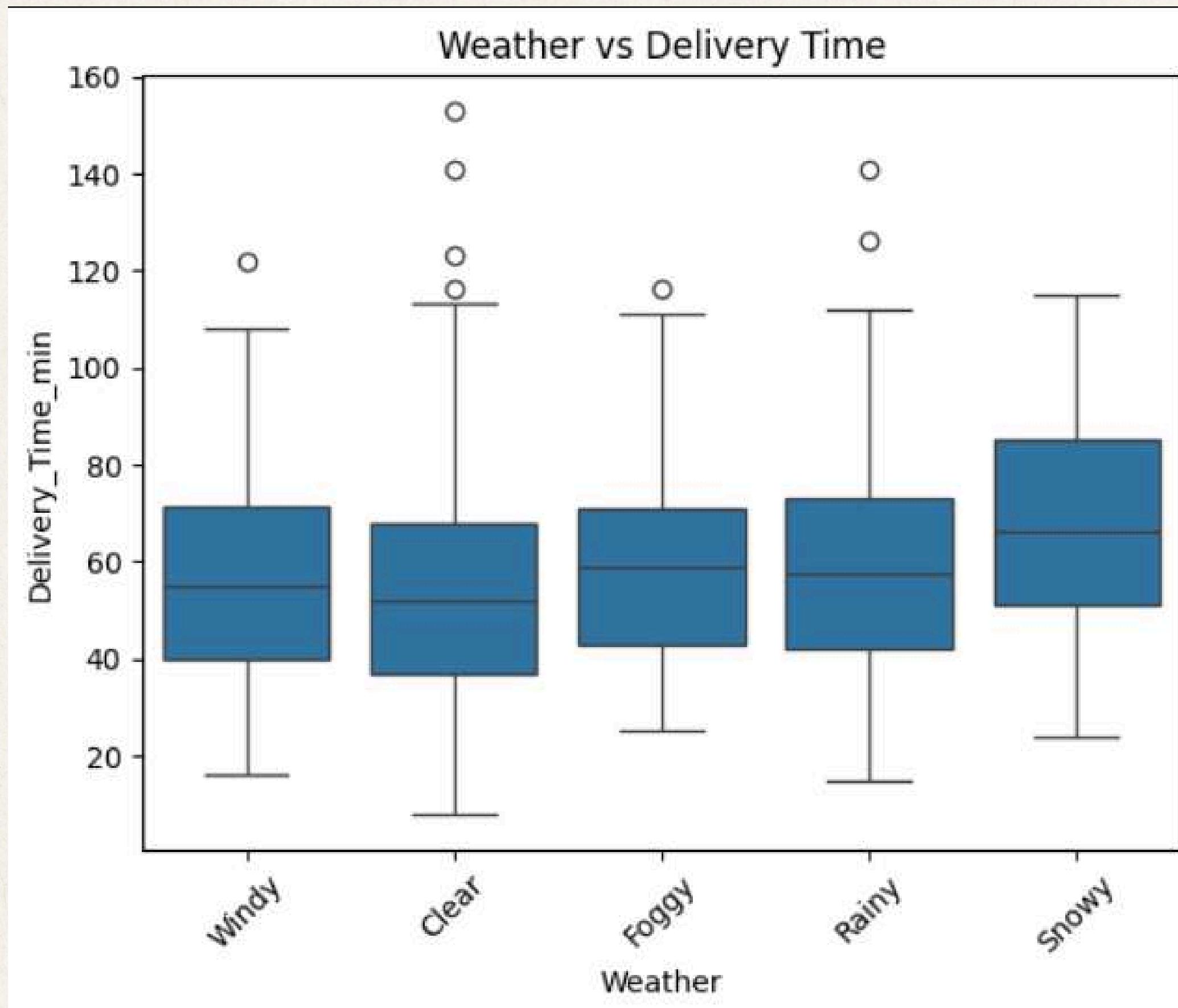
# Persebaran data jarak dengan waktu delivery



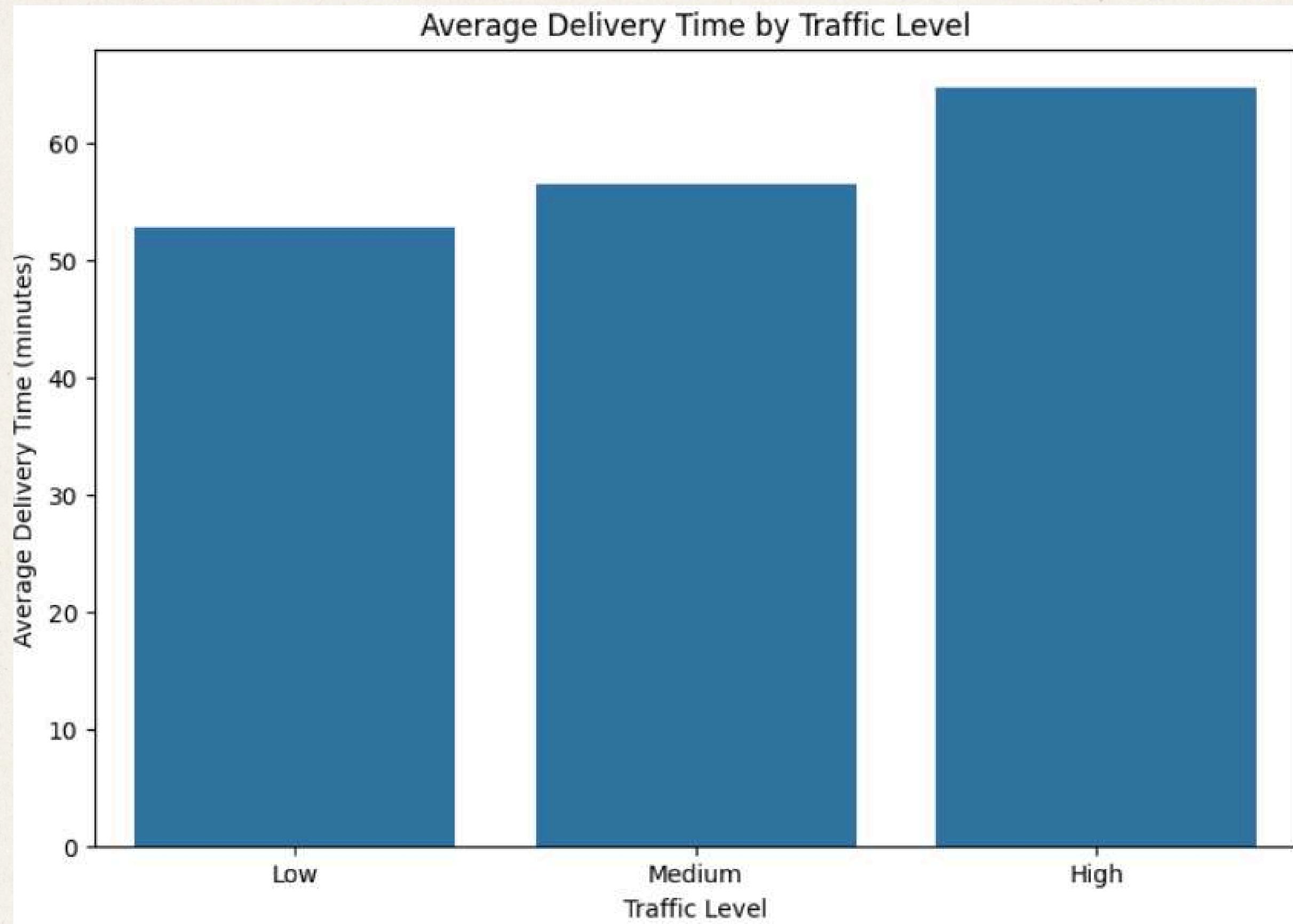
# Persebaran data waktu hari dengan waktu delivery



# Persebaran data cuaca dengan waktu delivery



# Distribusi Data waktu delivery berdasarkan traffic level



# **Data Preprocessing**

# **Split Dataset**

- Data Train = 70%, Test 30%
- Target prediksi = Delivery\_Time\_min

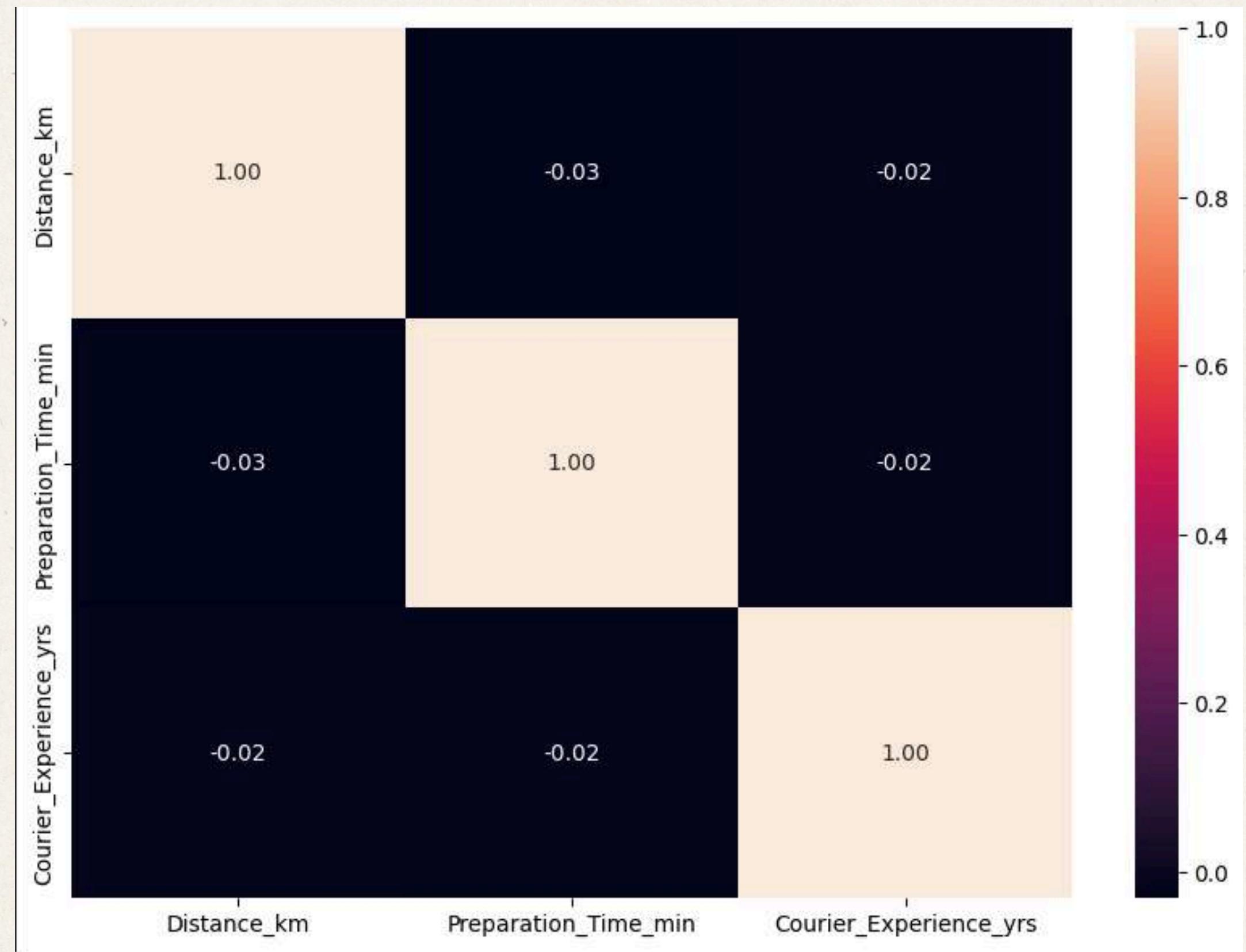
# Cek Korelasi Data

Tidak terdapat vif\_score yang tinggi.

	feature	vif_score
1	Distance_km	1.001229
2	Preparation_Time_min	1.001339
3	Courier_Experience_yrs	1.000628

# Pengecekan melalui Heatmap

Pada heatmap, diketahui tidak ada nilai korelasi yang tinggi.



# Data Preprocessing

- Encoding = One-hot Encoding
- Tidak melakukan normalisasi data karena tidak ada data yang rentang nilainya jauh.

	Order_ID	Distance_km	Preparation_Time_min	Courier_Experience_yrs	Delivery_Time_min
count	1000.000000	1000.000000	1000.000000	970.000000	1000.000000
mean	500.500000	10.059970	16.982000	4.579381	56.732000
std	288.819436	5.696656	7.204553	2.914394	22.070915
min	1.000000	0.590000	5.000000	0.000000	8.000000
25%	250.750000	5.105000	11.000000	2.000000	41.000000
50%	500.500000	10.190000	17.000000	5.000000	55.500000
75%	750.250000	15.017500	23.000000	7.000000	71.000000
max	1000.000000	19.990000	29.000000	9.000000	153.000000

# **Modeling**

## **Model 1**

Linear Regression

## **Model 2**

XgBoost Regressor

## **Model 3**

Random Forest Regressor

**Penjelasan Test model yang dilakukan :**

- Metrik yang digunakan adalah R2 Score, MAE, dan RSME
- Pemodelan akan dibandingkan dengan data train

# Evaluasi Model

Metrik/Model	Linear Regression Test/ <a href="#">Train</a>	XGBoost Regressor Test/ <a href="#">Train</a>	Random Forest Regressor Test/ <a href="#">Train</a>
R2 Score (Root Mean Squared Error)	<b>0.7898/0.7668</b>	<b>0.7673/0.8488</b>	<b>0.7652/0.9103</b>
MAE (Mean Absolute Error)	<b>6.49/6.56</b>	<b>6.92/5.56</b>	<b>7.02/4.09</b>
RMSE (Root Mean Squared Error)	<b>9.94/10.73</b>	<b>10.46/8.64</b>	<b>10.51/6.65</b>

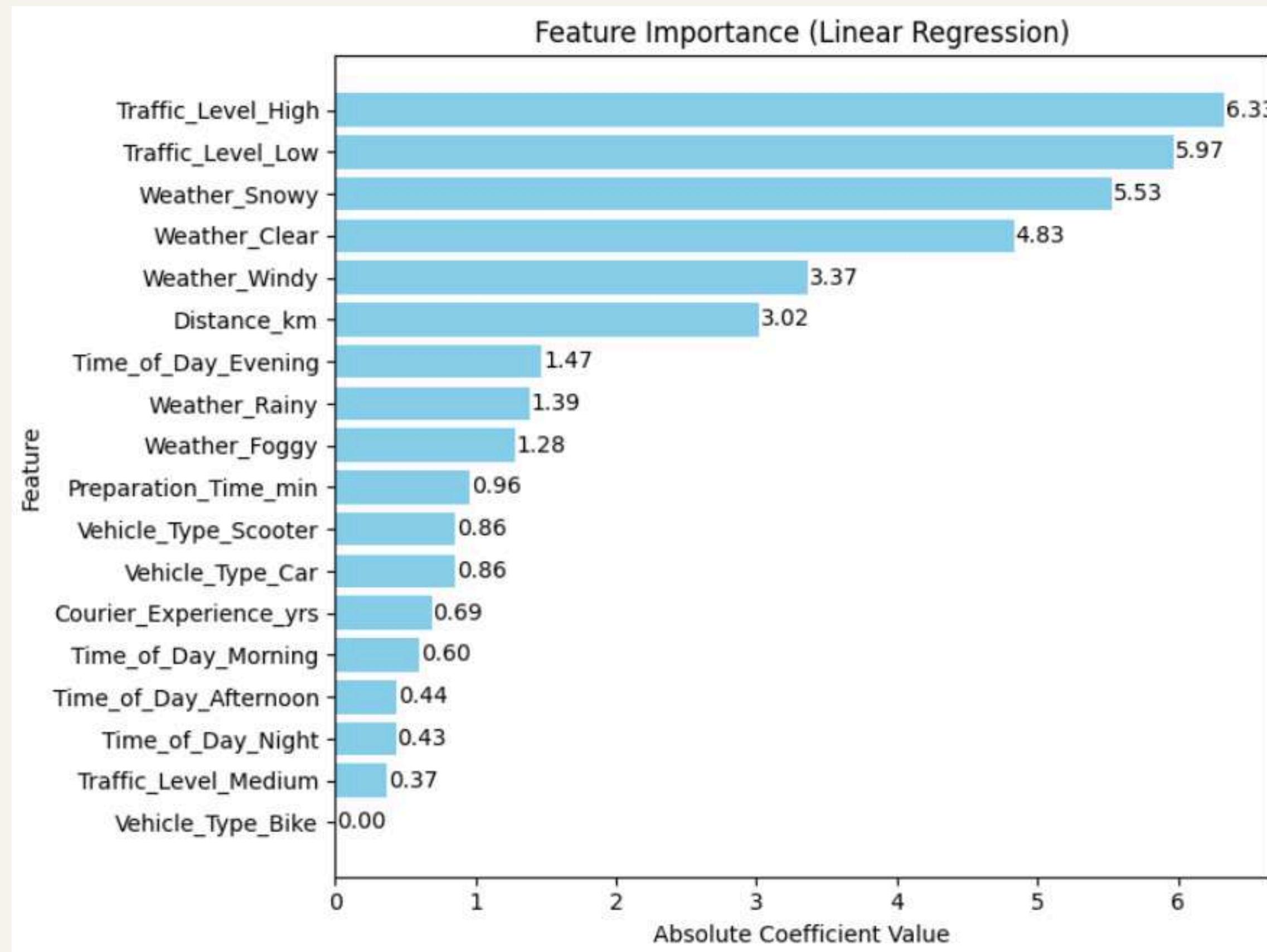
# Hasil Pemodelan

Linear Regression mempunyai nilai R2 Score lebih baik dari model lain, yaitu :

- R2 Score = 0.789 atau 79%
- MAE = 6,49 atau rata-rata kesalahan prediksi model = 6,49 menit
- RSME = 9,94 atau rata-rata deviasi prediksi model = 9.94 menit

# Feature Importance

Fitur Importance diambil dari model terbaik



# Kesimpulan

Faktor utama yang mempengaruhi nilai pada model Linear Regression adalah Traffic, terutama Traffic\_level\_high dengan nilai 6.33. Setelah traffic diikuti dengan weather.

Penjelasan nilai Metrik model terbaik (Linear Regression) :

- Rata-rata prediksi meleset (MAE) = 6,49 menit
- Rata-rata akar kuadrat prediksi meleset (RMSE) = 9.94 menit
- R2 Score= 0.789 atau 79%
- Rata-rata jumlah harga rumah = 56.73 menit

Dengan hasil ini, model dapat memberikan estimasi waktu pengantaran yang cukup akurat dan realistik, sehingga dapat digunakan untuk menampilkan estimasi waktu yang lebih tepat kepada pelanggan. Hal ini berpotensi meningkatkan kepuasan pelanggan karena mengurangi perbedaan antara estimasi dan waktu pengiriman aktual..

## **Saran**

- Bisa dilakukan penelitian lebih lanjut dengan menambahkan variabel eksternal seperti lokasi geografis, kepadatan area pengantaran, atau jumlah pesanan simultan per kurir untuk meningkatkan akurasi model.

# TERIMA KASIH

 GitHub Repository: [Link Github data](#)