🚚 Real-World Case: Operational Intelligence in Urban Logistics

📦 Scenario

A food delivery platform operates in multiple urban regions. Lately, they've received complaints about delivery delays — but the causes are unclear. Your task is to investigate the issue, build a system to predict delivery time, and share actionable insights that help the Ops team respond more intelligently.

You are free to use any external tools, libraries, and GenAI assistants (e.g., ChatGPT, Copilot), but please disclose where and how you used them.

📘 Dataset

You'll use this Kaggle dataset, it contains features such as delivery distance, weather, traffic, order type, and time taken:

👉 https://www.kaggle.com/datasets/denkuznetz/food-delivery-time-prediction\

✅ Deliverables

You will have **1 week** to send your **GitHub repository link** to the person who shared this assignment with you. This section outlines the different parts of the challenge.

> 💡 The goal of this technical challenge is to assess your programming and development skills, as well as your statistical/analytical thinking in the context of solving a business question. Keep this in mind when structuring your repository, writing your code, and providing explanations for insights you uncover along the way.

PART I: SQL

1. Assume you also have access to the following tables:

```
-- Delivery-level data
deliveries (
      delivery_id VARCHAR,
      delivery_person_id VARCHAR,
      restaurant_area VARCHAR,
      customer_area VARCHAR,
      delivery_distance_km FLOAT,
      delivery_time_min INT,
      order_placed_at TIMESTAMP,
      weather_condition VARCHAR,
      traffic_condition VARCHAR,
      delivery_rating FLOAT
)
-- Delivery personnel metadata
delivery_persons (
      delivery_person_id INT,
      name VARCHAR,
      region VARCHAR,
```

```
      hired_date DATE,
      is_active BOOLEAN
)
-- Restaurant metadata
restaurants (
      restaurant_id VARCHAR,
      area VARCHAR,
      name VARCHAR,
      cuisine_type VARCHAR,
      avg_preparation_time_min FLOAT
)
-- Orders table
orders (
      order_id INT,
      delivery_id VARCHAR,
      restaurant_id VARCHAR,
      customer_id VARCHAR,
      order_value FLOAT,
      items_count INT
)
```

Design queries to answer the following questions, you can create hypothetical tables if you want.

1. Top 5 customer areas with highest average delivery time in the last 30 days.
2. Average delivery time per traffic condition, by restaurant area and cuisine type.
3. Top 10 delivery people with the fastest average delivery time, considering only those with at least 50 deliveries and who are still active.
4. The most profitable restaurant area in the last 3 months, defined as the area with the highest total order value.
5. Identify whether any delivery people show an increasing trend in average delivery time.

How to submit Part I within your GitHub repository:

- sql/sql_queries.sql: The queries you design for the questions above.
- sql/sql_insights.md: Any other analysis/questions you think are useful based on the business problem, and their respective queries if applicable.

PART II: Modeling

1. Exploration & Modeling

Below you fill find what we're looking for, but you are free to add more files/folders as you need it:

- model_pipeline/: This folder should contain code that runs the end-to-end process such as data preprocessing, training, and a predictive model for delivery time. Consider using software development practices that would accelerate deployment to production.

Show us your work in these reports:

- EDA_report.md: Key patterns, outliers, and assumptions you made.
- model_notes.md: Your modeling logic, metric choice, and tuning approach

- explainability.md: Insights from feature importance tools
- error_insights.md: Insights on when and why your model fails

2. Strategic Thinking & Communication

In strategic_reflections.md, answer the following questions

1. **Model Failure:** Your model underestimates delivery time on rainy days. Do you fix the model, the data, or the business expectations?
2. **Transferability:** The model performs well in Mumbai. It's now being deployed in São Paulo. How do you ensure generalization?
3. **GenAI Disclosure:** Generative AI tools are a great resource that can facilitate development, what parts of this project did you use GenAI tools for? How did you validate or modify their output?
4. **Your Signature Insight:** What's one non-obvious insight or decision you're proud of from this project?
5. **Going to Production:** How would you deploy your model to production? What other components would you need to include/develop in your codebase? Please be as detailed each step of the process, and feel free to showcase some of these components in the codebase.

PART III: Optional (extra credit!)

API Prototype

- Create (or at least articulate the steps) a basic endpoint that takes delivery details and returns an estimated delivery time.
- Include a README.md with setup and usage instructions.

You can add a next_steps.md file in your repository with your answer. Again, Part III is optional.