1. **My Chosen Domain: B2B Logistics**

I specifically chose the Logistics & Supply Chain domain because it’s at the very heart of what RoaDo does. Your company’s mission to be an “operating system for logistics” and to prevent “delays via AI Enabled Alerts” really stood out to me.

I wanted to use this assignment to show that I can analyze the exact kind of data that is relevant to your business. My goal is to find insights from real world Indian logistics data that a tech company like RoaDo could act on.

1. **The Dataset I’ll be Using**

**Source:** Kaggle(a primary repository for public, real word datasets)

**Dataset: Delhivery** Dataset (logistics and supply chain company that provides a full suite of services, from last mile delivery to warehousing)

**Link:** <https://www.kaggle.com/datasets/santanukundu/delhivery-dataset>

1. **What’s in the Data?**

**Trip Info:** trip\_uuid, route\_type(Full Truck Load/Carting)

**Route Info:** source\_name, destination\_name(eg; Anand\_VUNagar\_DC(Gujarat))

**The Core Metrics:** segment\_actual\_time(how long it really took), segment\_orsm\_time (how long it should have taken, according to a open sourse routing engine(orsm))

And many more, these are some key details from them.

1. **The Business Questions I Aim to Answer**

To guide my analysis, I’ve framed goals as set of key business questions. I’m not just looking for what happened, but why it happened.

1. What’s the actual gap between expected delivery times and real worls results? (I’ll calculate a delay metric for every single segment)
2. Do ‘FTL’(Full Truck Load) shipments experience worse delays than ‘Carting’ shipments?
3. Can we pinpoint specific bottleneck hubs or states that are biggest sources of delays?
4. What time of day or day of the week do the worst delays happen?
5. Do longer distance segments get delayed more often, or is the delay random?