Every day, Amazon ships millions of products globally. But how do they ensure the right goods reach the right regions at the lowest cost? Supply chain inefficiencies waste billions annually—and that’s where our project steps in. By modeling Amazon’s 2020 sales data as a network, we’ll pinpoint where to cut costs and how to boost sales through smarter logistics and marketing.

The data collected from Kaggle gives us product prices, quantities, and categories—but no regional or logistics details. To fill this gap, we’re generating synthetic regions, distribution centers, and customer groups. This hybrid approach lets us simulate a realistic supply chain network while staying grounded in actual sales behavior.

Our technical plan in divided into 3 segments:

1. Data Prep: using pandas to resolve biases like uneven regional sales distributions.
2. Network Design: Use NetworkX to model products, regions, distribution hubs as nodes and distribution paths and costs as edges.
3. Visualization: Building an intuitive flow diagram in Gephi revealing high-traffic routes and underperforming clusters.

Now, Darren will walk you through the size and structure of this network.