# **DATA SCIENCE CASE STUDY: ROOT CAUSE FOR LATE DELIVERIES**

## Problem Statement

The company is an online retailer that delivers customized apparel to customers in Europe. They specialize in 3 categories – Tops, Jackets & Vests, and Accessories – and customization typically means monogramming or stitching a custom logo on to their pre-manufactured products.

For the purpose of this case, you can think about their supply chain as linear with 4 steps following a customer’s order.

**Plan:** Assign a date the customer can expect to receive their order. This is automatically generated upon order confirmation according to complex, but imperfect rules. The plan does not take into account exactly where the order will be sourced or made.

**Source:** Pick the product out and drop it in to a working station so it is ready to be customized by hand. (*Note:* we will call this sourcing/supply, even though it’s really just moving the product between 2 adjacent facilities).

**Make:** Customize the product

**Deliver:** Transport the product to the customer using local carriers



The company has a problem: their deliver-to-plan performance is far below the online retail industry standard. Customers often receive their orders late and, as a result, are beginning to turn to competitive retailers. Competitors are changing the paradigm of digital fulfillment, and the company needs advanced analytics support to deliver better to their consumers.

**The business problem: how should the company address its late deliveries?** The individual functions in the supply chain have previously not had visibility into the customer impact of their processes. Your task is to prototype an analytical solution that will connect the supply chain functions and discover or alleviate root causes of late deliveries.

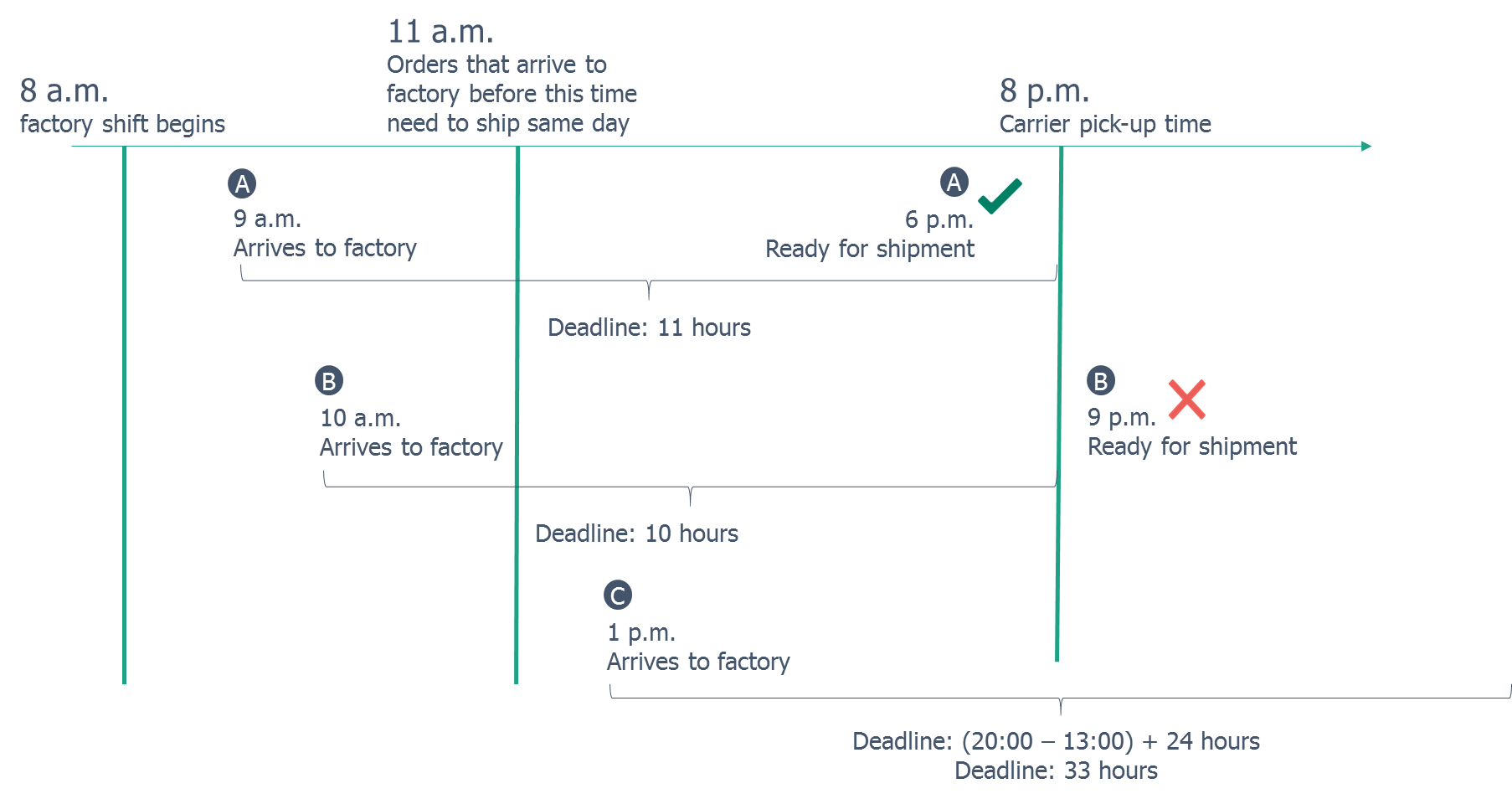
## Data Set

The company collects two types of data on orders placed by customers:

1. Order attributes: descriptors of the order, customer, or items within the order
2. Order milestones: timestamps reflecting when orders have cleared each process step, as well as the duration of slack time each order has to clear the step

The “duration of slack time each order has to clear the step” is the most complex part of the company’s supply chain. To illustrate, consider the process step between MAKE (customization) and DELIVER (carrier logistics). Typically, factories and distribution centers have set times during the day for carriers to pick up shipments; UPS and Fedex don’t continuously have trucks picking up every individual shipment when it is ready to go. Therefore, the deadline for an order to be ready is based on these cut-off times.

For example, consider 3 orders: A, B, and C. A and B arrive to the factory early in the morning, so their customization needs to be ready by the carrier pick-up time at 20:00. Order A has 11 hours; Order B has 10 hours. Order A makes it in time, while Order B does not. Order B is picked up the next day, and may end up late to the customer. The third order – Order C – arrives at the factory in the afternoon, so it doesn’t need to be ready for the carrier until the next day, 33 hours later.



**The SOURCE, MAKE, and DELIVER steps all have these deadlines. PLANNING does not.** Assume that if the order meets all its deadlines, but is still late, then it operated under a bad PLAN. Note that the deadlines are independent for each step; meaning, if an order is already late to the factory, it still gets sufficient time to be processed. A step won’t be penalized for an upstream delay.

**SOURCE deadlines are measured in minutes.**

**MAKE deadlines are measured in hours.**

**DELIVER deadlines are measured in days.**

Most of the order attribute fields are self-explanatory. Below is more detail on a few of them:

* *facility*: location where the customization occurs and shipment originates
* *shipping\_method* and *transit\_days*: the shipping method the customer requested. Next Day, 2-Day, and 3-Day take 1, 2, and 3 business days respectively; Ground is dependent on the customer location, and takes the amount of business days in *transit\_days* to reach the customer.
* *returned*: the order was ultimately returned; the timestamps reflect the initial delivery of the product to the consumer (not the return flow)
* *backordered*: an item in the order needs to be sourced from an off-premises supplier location, which usually takes longer. This exception does *not* get factored into the planned delivery date.

## Deliverables

* First, scope the problem. Use descriptive analytics to set context and find interesting performance trends in the supply chain and the company’s service to the consumer. The purpose of this is solely to inform the modeling and simplify the approach into a manageable proof-of-concept. For example – perhaps Sourcing in Germany seems to be a huge problem on certain days; how would the company use the other data to address it?
* Second, develop a proof of concept to address the business problem. Consider creating new features derived from the data set that may explain the environment of the supply chain, beyond the features of the individual orders. Many analytical approaches are possible with this data set; creativity and ingenuity are more important than model results.
* Finally, summarize your results, providing a brief slide deck (<10 slides) containing:

1. Context and scope explanation
2. Approach and methodology
3. Results
4. Adoption plan – how would the company operationalize the solution?

Please also provide all code (including comments) written for your case.