



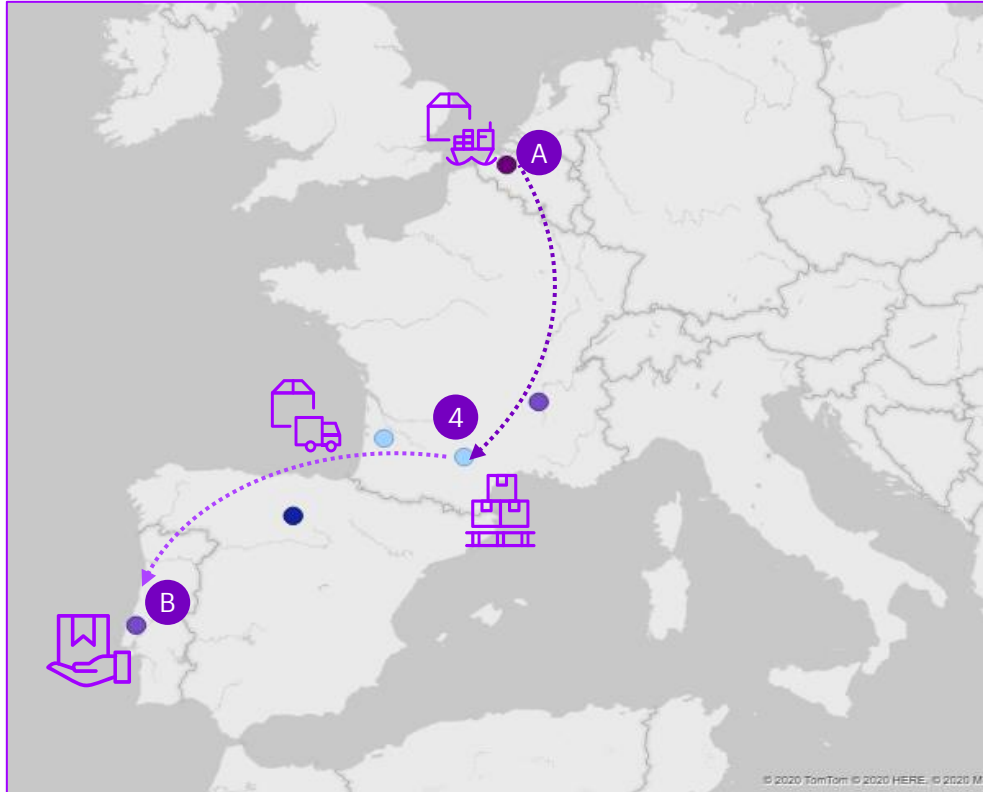
Applied Intelligence IBERIA COE

***A Global Hub of Data &
Analytics experts***

Challenge description

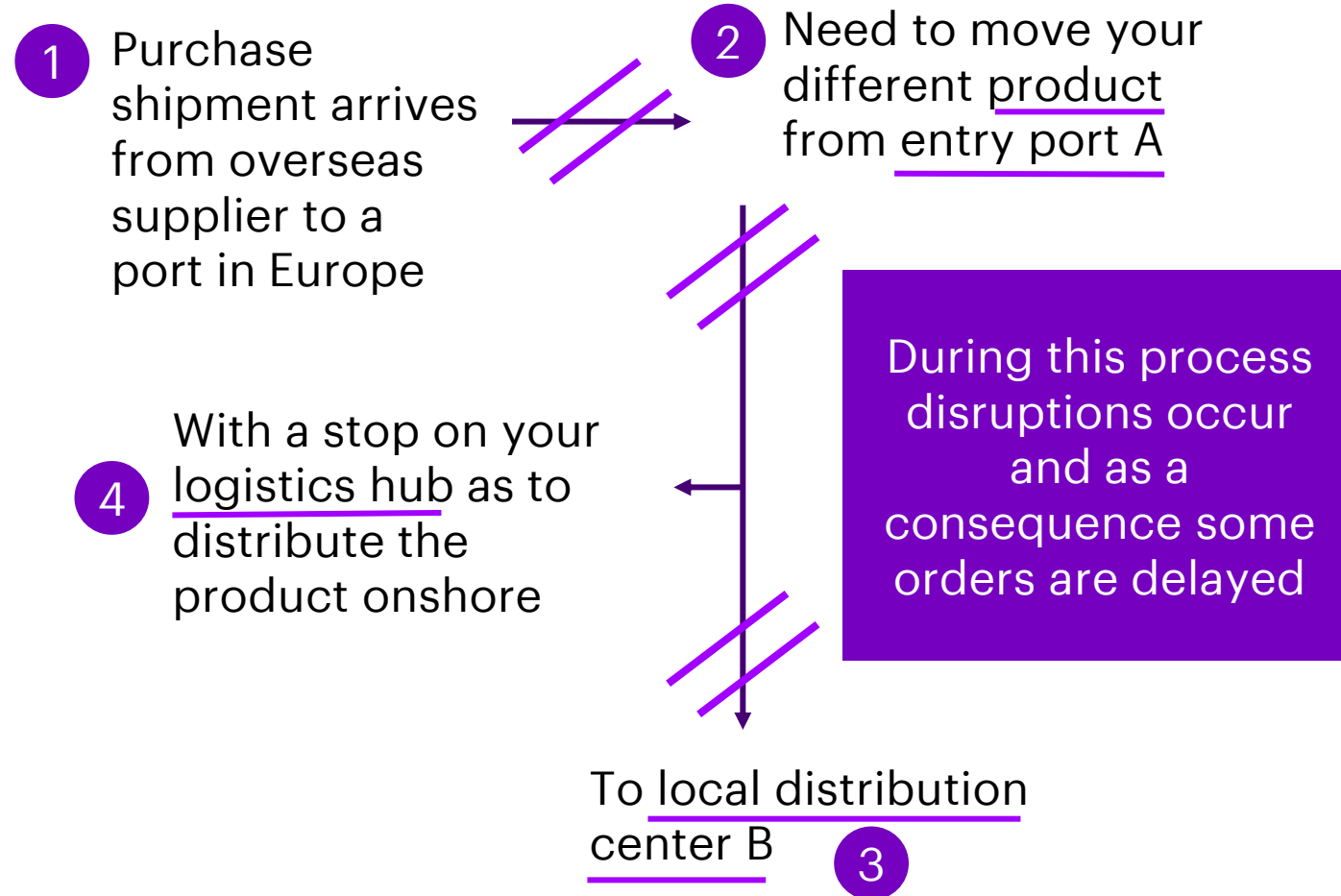
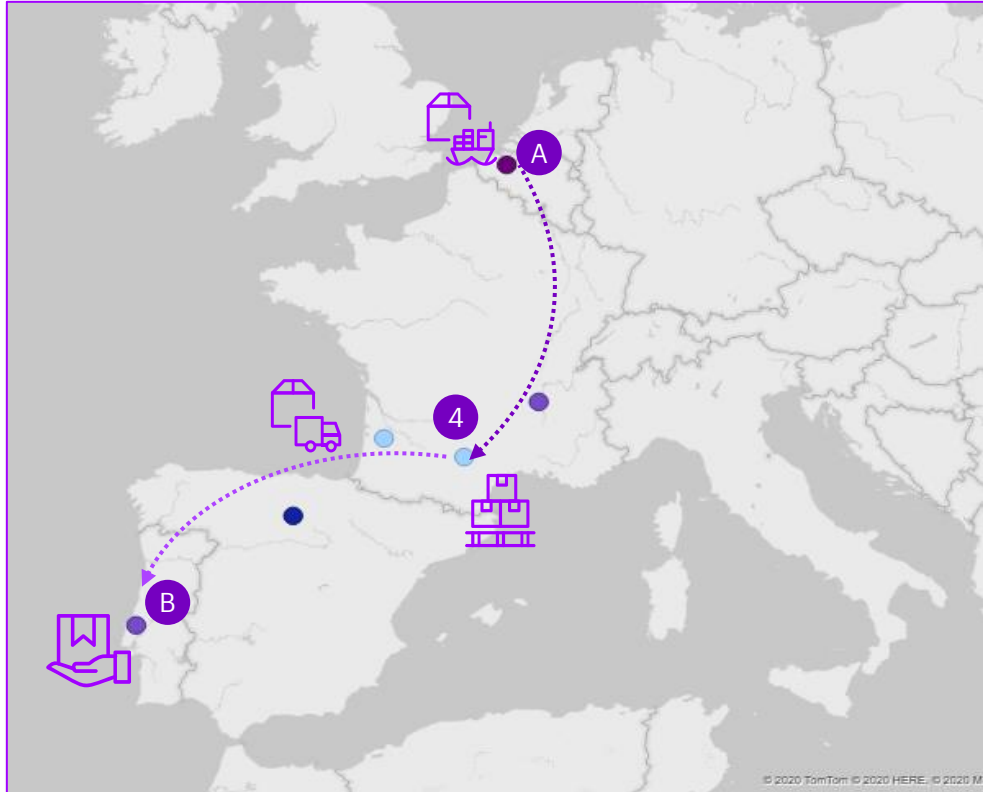
Context- Understanding Supply Chain data

Imagine you are a fashion retailer



Context- Understanding Supply Chain data

Imagine you are a fashion retailer



UPC – Datathon Event

Supply Chain Resiliency - Accenture



Challenge Description



Data Set



Wining Metrics

1

Understand main root-cause

affecting supply chain deliveries and **predicting product delays** based on historical inbound & outbound orders of the company

2

Provide data insights on historical data, giving visibility of **supply chain resiliency** to maximize customer service levels

3

Predictive model to get likelihood or order delays. Accuracy will be validated against a test set

FAQS



→ **What data sets are available that we can use?** Anonymized supply chain data model:

- **Historical transactional orders** from port to final customer
- **Product attributes** master data
- **Geographic coordinates** of cities involve in the supply chain.



→ **Where can I find the data?** All data sets are available in the GIT repository

→ **How the model will be tested?** A [Kaggle competition](https://www.kaggle.com/competitions/datathon-2022-upc-accenture) have been created where participants will create a live test model accuracy and compare it vs other teams

<https://www.kaggle.com/competitions/datathon-2022-upc-accenture>

What will we measure?

- **Business case** presentation outlining descriptive insights of main drivers' toward order delays.
How those insights can be translated to business actions and value proposition?
- **Model accuracy:** ROC Curve (AUC) evaluation metric for given test data set

What are the main deliveries we expect from participants?

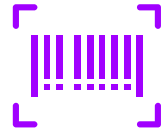
- PowerPoint (methodology, algorithms, main insights, and use case).
- Prediction on test set with likelihood of order delay
- Other business-oriented insights: sustainability, cost, etc.



Context- Understanding Challenge data sets



4 Datasets



+115K

Historical orders records



+35

European cities within supply chain



+700

Products

- **orders.csv**

Transactional historical data of the company supply chain inbound/outbound shipments

- **product_attributes.csv**

Master data of product unit weight

- **cities_data.csv**

Geographic coordinates of cities involve in the supply chain. Including distance between pair of cities

- **test.csv**

Same as *orders.csv* but variable 'late_order' has been truncated. This is the target variable

Metadata with data sources and columns descriptions available in [GIT readme](#)

order_id	origin_port	3pl	customs_procedure	logistic_hub	customer	product_id	units	late_order
a9c0f39d8f1c	Rotterdam	v_004	DTP	Liege	Valencia	1668362	420	TRUE
ddefb0983de4	Barcelona	v_004	CRF	Dusseldorf	Lyon	1688294	427	FALSE
6f06cd5c734c	Rotterdam	v_001	CRF	Hamburg	Madrid	1699425	437	FALSE
115df9e80041	Athens	v_002	CRF	Hamburg	Munich	1613321	479	TRUE
68c4c3cb1bb6	Rotterdam	v_002	DTP	Zaragoza	Bucharest	1682790	447	TRUE
c6e27c769c9a	Barcelona	v_004	CRF	Rome	Milan	1686764	510	FALSE
bba3abe196fd	Barcelona	v_002	CRF	Rome	Munich	1660563	430	FALSE

How to read a supply chain order

Product 1699425 is imported via **Rotterdam port** following a **CRF customs procedure** and transported to the **logistic hub of Hamburg** by the **third-party logistic (3PL) V_001**. Later, a **customer in Madrid** orders **473 units** of the product. **Order id:** f5e98cb29790 is the unique identifier tracking the end-to-end path of product movement through the supply chain



Want to go above and beyond?

Understanding shipment costs and sustainability scores

- Different factors intervene in shipment costs and emissions. The key factors that we will explore in this part of the challenge are three:
 - **Distance:** Route length affects both economic costs and emissions.
 - **Shipment weight:** Total weight of shipped goods affects both economic costs and emissions.
 - **Economics of scale:** As routes become longer and shipments bigger, the unit costs per unit shipped lowers. This only affects economic costs.
- For this part of the challenge, you will need to work with two modified versions of the dataset.
 - **Product weight class:** To simplify the analysis, the products have been segmented into weight class categories. Class 0 refers to the lighter products, class 9 refers to the heavier products.
 - **Two new variables in the cities dataset:** cost per unit shipped, c02 emissions per unit shipped. These are bound to route used and weight category of the order shipped.
 - **Important note!** To mimic the effects of economics of scale, the total shipment cost of an order is calculated as $n \cdot \sqrt{\text{units shipped} \cdot \text{cost per unit} / n}$. Further details in the documentation of the challenge.



Want to go above and beyond?

Understanding shipment costs and sustainability scores

This part of the challenge is free form. Ask yourself: What data insights can bring business value? Here are some ideas:

- Which are the orders that represent a bigger cost? And the most emissions?
- What actions can you take to bring down shipping costs while causing minimal disruption in your supply chain? And if we want to bring down emissions?
- If you were to add a new logistic hub, where would you place it? Does your recommendation change if you base it in costs vs emissions? How would you approach taking a decision based on both variables?

Argument all your decisions with data analysis! Good luck and happy coding! 😊