











## A Trader as the Manager of Data

- Data Quality Accuracy & Balanced Representation
- Data Quantity to avoid discriminatory outcomes
- Data Relevance
- Longitudinal Data: Data is collected over a sufficient period to capture trends and changes.

## Input Data

## Managing Data

- Privacy & Security anonymization and storage
- Ownership, licensing and management of data vendors
- Data Annotation and Labelling – quality, consistency and verification
- Data Transformation and Preprocessing Imputation/Augmentation methods for missing data
- Outlier Detection

- · Portability & Bridging of Data
- Feature Engineering relevance, scaling/standardization for modelling
- Training, Validation and Testing to evaluate performance
- Cross Validation (K-fold) to avoid overfitting and improve the generalizability of the model

Portability of Data



Confidential - Not for Distribution





