dunnhumby

The dunnhumby **Customer Data Science Dictionary**

Data & Technology

> In recent years, increased computing power has enabled us to unlock the power of big data and to lift digital services on to the next level through sophisticated AI algorithms. But it is not just a success story of better hardware, but also clever software that unleashes the potential of improved computing capabilities.

Clickstream Data

Clickstream data shows how customers navigate through websites. This data can be used to develop science that improves the customer experience on those sites.

Unstructured Data

Not all data comes in a table.
Often the first step in a project is to structure data and give it meaning. e.g., making different sources of customer feedback comparable.

Spark and Hadoop

Spark and Hadoop are computing frameworks that allow the distributed storage and processing of large sets of data. This makes it possible to analyse and/or predict the behaviour of millions of customers in a short amount of time.

Scoring Engine

A lightspeed method to predict customer preferences, e.g., when we need to decide what products to recommend to a shopper based on the items currently in their basket.

Supervised Learning

> We use supervised learning whenever we are trying to predict a variable for which we have historical data. For example, we can predict which customers will redeem a coupon using redemption data from previous coupon campaigns.

Collaborative Filtering

A technique used to recommend products to customers based on other customers with similar likes and dislikes.

Propensity Modelling

Used to identify people that are likely to show a certain behaviour, e.g., we can identify people who are more or less likely to show interest in an offer.

Cross-Validation

The process of training and testing your model on different subsets of data. This enables you to build up a good estimate of how your machine learning model performs on unseen data (e.g., different customers) and ensures your model performs well not just with historical data but also when you use it to make predictions.

Overfitting

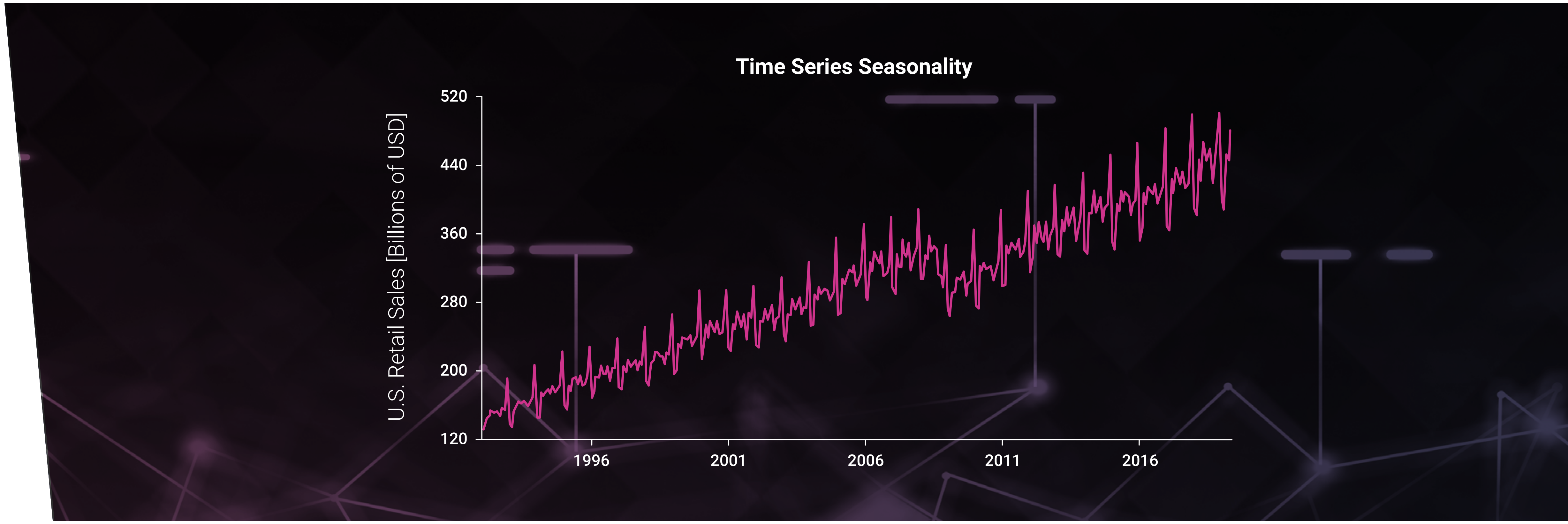
An overfitted model will explain the behaviour of customers in the historical training data well but fail to predict the behaviour of unseen customers in future datasets.

Bias

For example, when a model generally predicts that customers will spend more money on purchases than they actually do, it's probably biased.

Time Series Analysis (e.g., ARIMA)

This is when we analyse events that happen consecutively in time and are therefore dependent on each other. For example, a customer's purchase today might influence what they're going to buy next time they shop.



Unsupervised Learning

> Unsupervised Learning is used to predict a variable that has no ground truth. For example, clustering customers together as ‘loyal’ can lead to different results depending on what definition of loyalty is used.

Topic Modelling (e.g., LDA)

Identifying different missions in a customer’s basket and therefore explaining the context of their shopping (e.g., birthday shop, easy to prepare food, etc.).

Word2Vec

A technique used to identify related words and meanings of words in natural language data. This can, for example, be used to give structure to verbal customer feedback.

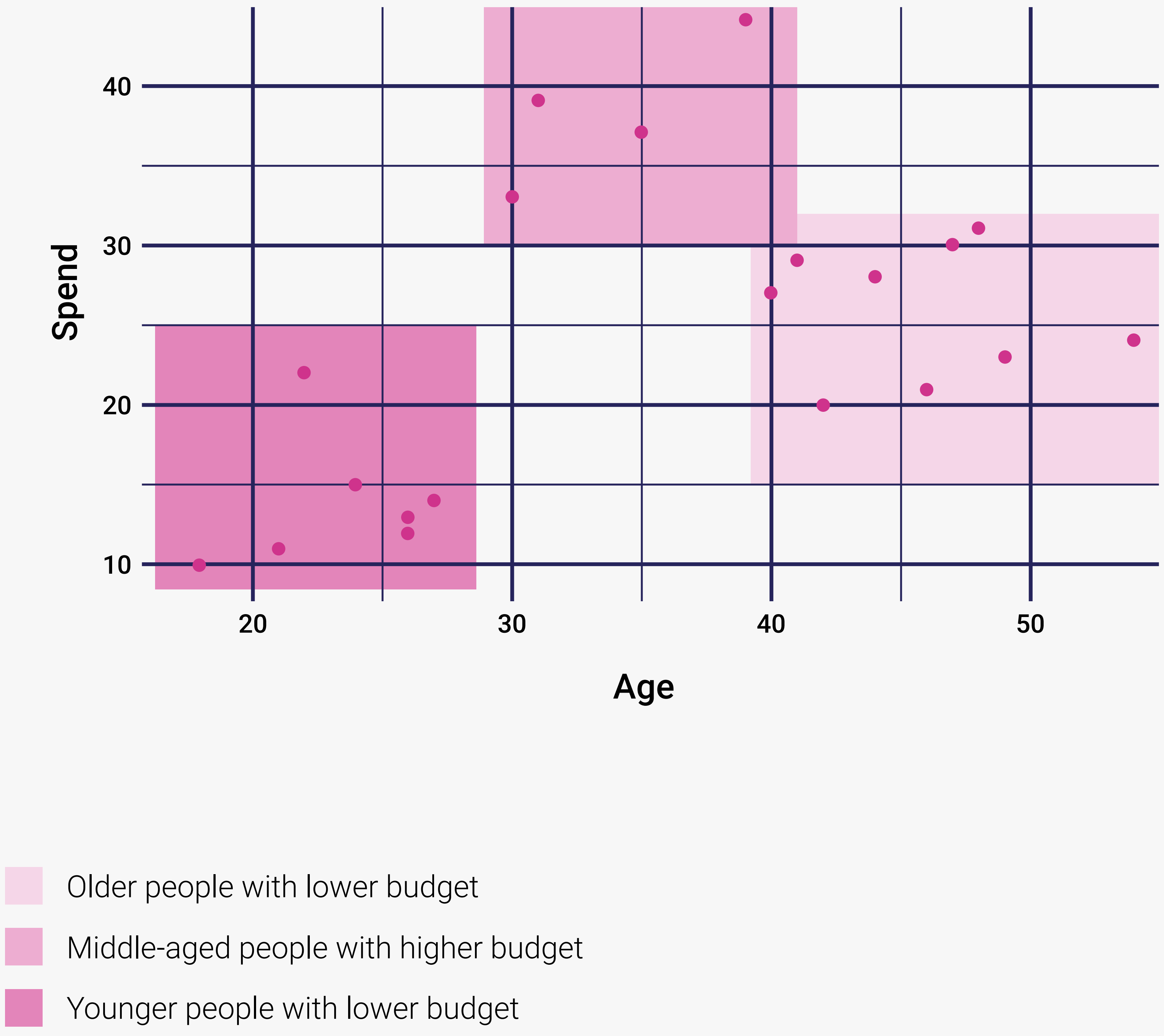
Clustering
(e.g., K-Means)

Clustering can be used to group together customers who exhibit similar shopping habits and preferences so we can start recognising how their needs are different from other groups.

Dimensionality Reduction
(e.g., PCA)

When we have too many potential factors that could explain consumer behaviour, we use dimensionality reduction to narrow it down and find the ones that are most important.

Clustering example



Simulations

> Simulations are often used to understand complex environments. For example, we create rules by which we produce (i.e., simulate) data artificially. Next, we test how well our rules capture what is going on in the real world by compared the artificial data to real world data.

Agent-based Modelling

A simulation of individual customer behaviour where one can change different environmental variables (e.g., prices) to see how customers would react to that change.

Gravity Modelling

Explaining how store locations of our clients and their competitors, together with general infrastructure, impact customers' decision to shop at a certain store.

Permutation Tests

To find out whether customers are making choices in a specific order, we compare their sequence of chosen items to randomly shuffled versions of that sequence. We can then check if the order of customers' choices contains specific patterns.

Deep Learning

Using artificial neural networks to model the relationship of variables is often referred to as Deep Learning. An artificial neural network model can be imagined as layers of different models, where one model's output is the input to the next model. Such a deep layered structure enables computers to learn complex relationships, such as recognising faces on photos.

Reinforcement Learning

A type of model used to capture long-term consequences of actions, such as the impact of sending coupon offers on long-term customer loyalty. Reinforcement Learning famously enabled computers to beat human champions in games such as Chess or Go.

Image Recognition

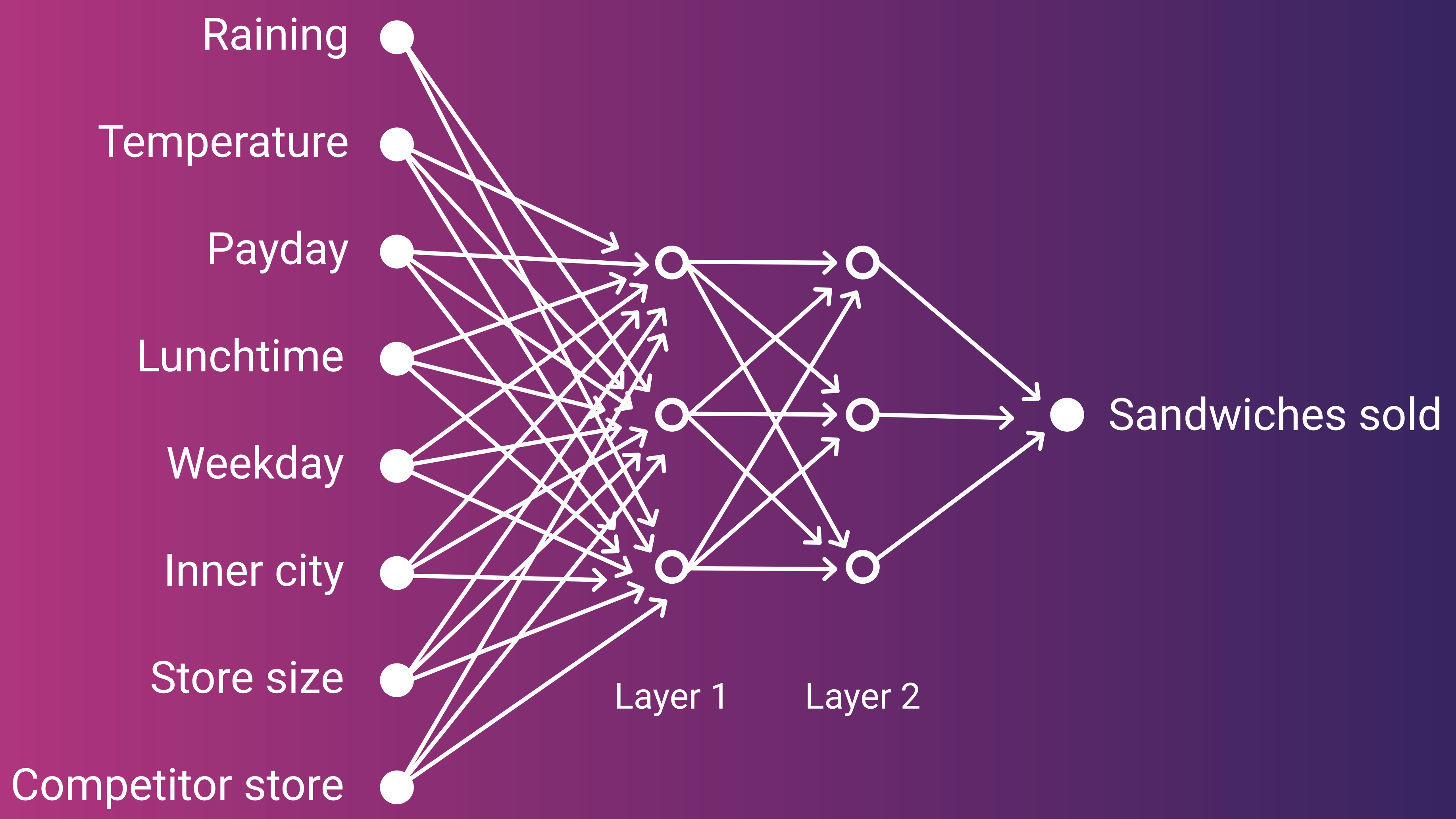
Deep Learning techniques can be used to process image data (e.g., product images) and teach computers to recognise related images. We can use this information to help customers find related products more easily.

Neural Nets

Artificial Neural Networks are useful for modelling the complexity of customer behaviour as they are able to capture how different pieces of information in the data relate to each other. In Artificial Neural Networks, the data moves through a chain of models that transform it until it is used to make a final prediction.

Example Neural Network

"Fully connected" directed graph of neurons



Statistical Theory

> Despite many exciting new algorithms that have been released since the advent of machine learning, there are still numerous techniques and principles that have been around for a long time and still provide best in class analytical solutions.

Shapley Values

Shapley values can tell us to what extent different factors that are working together, such as promotion, price or product range, are responsible for the customer behaviour we observe.

Causal Modelling

Correlation is not causation, or in other words: predicting what customers will do next is not equal to explaining what causes them to behave that way. Causal modelling seeks to make rigorous connections between statistical modelling and cause and effect.

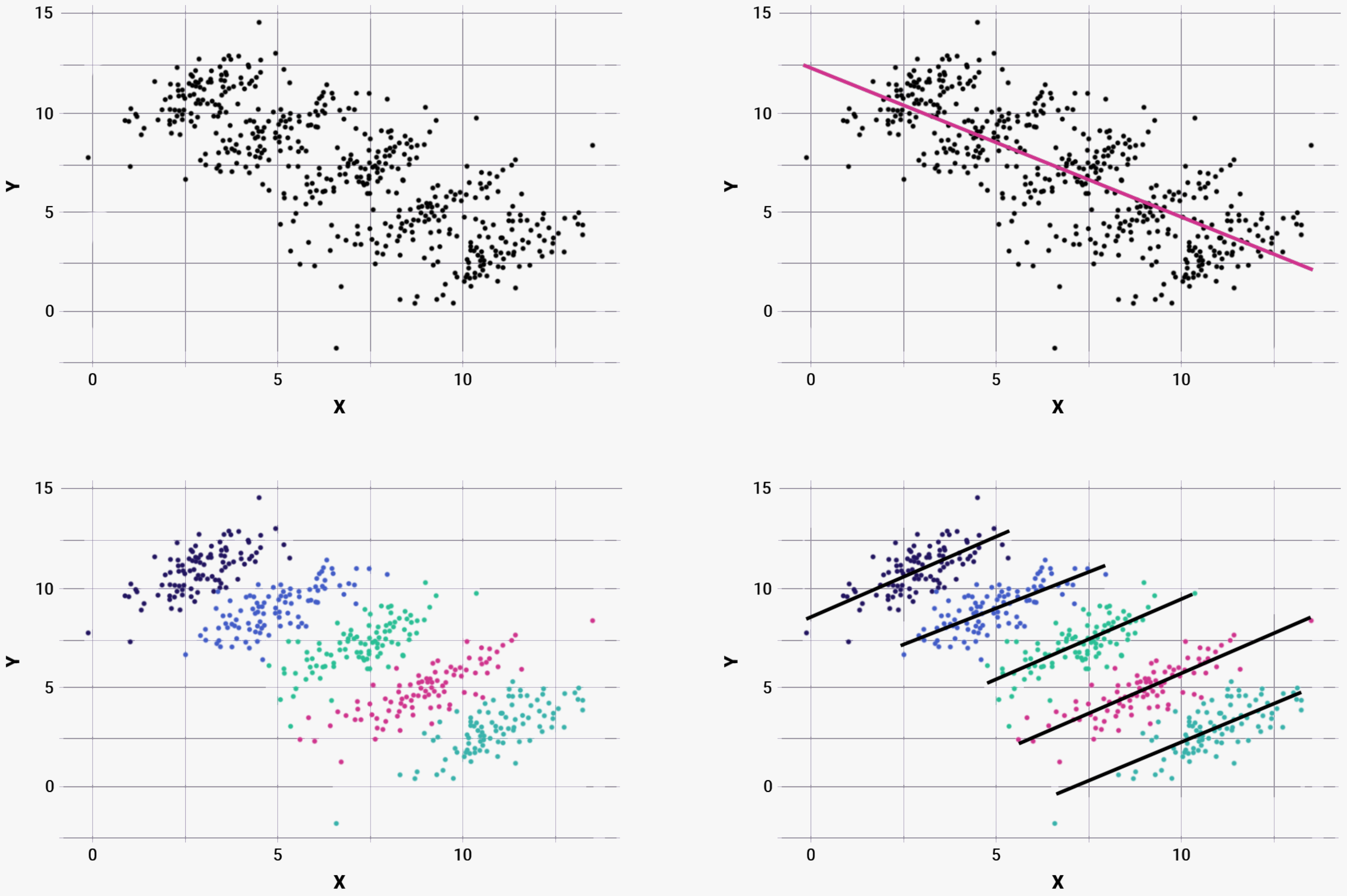
Simpson's Paradox

In Simpson's paradox, a pattern seen in customer groups vanishes or reverses if we combine those groups. In the figure, the red and blue lines imply being taller is advantageous in both age groups. However, the black line fitted on the combined group implies the reverse.

Hitchhiker's Paradox

The time between a product coming back into stock on a shelf and being bought by a customer is on average twice as long as one might expect, due to Hitchhiker's paradox.

Simpson's Paradox



dunnhumby THE WORLD'S FIRST CUSTOMER DATA SCIENCE PLATFORM

dunnhumby is the global leader in Customer Data Science, empowering businesses everywhere to compete and thrive in the modern data-driven economy. We always put the Customer First. Our mission: to enable businesses to grow and reimagine themselves by becoming advocates and champions for their Customers.

With deep heritage and expertise in retail — one of the world's most competitive markets, with a deluge of multi-dimensional data — dunnhumby today enables businesses all over the world, across industries, to be Customer First.

The dunnhumby Customer Science Platform is our unique mix of technology, software and consulting enabling businesses to increase revenue and profits by delivering exceptional experiences for their Customers – in-store, offline and online. dunnhumby employs over 2,000 experts in offices throughout Europe, Asia, Africa, and the Americas working for transformative, iconic brands such as Tesco, Coca-Cola, Meijer, Procter & Gamble, Raley's, L'Oreal and Monoprix.

Connect with us to start the conversation

dunnhumby.com