# Customer Analytics in Python

## Notes

# Words of welcome

**Customer Analytics**: The first part of the course focuses on how to perform customer segmentation, using a hands-on approach. It involves the application of hierarchical and flat clustering techniques for dividing customers into groups. It also features applying the Principal Components Analysis (PCA) to reduce the dimensionality of the problem, as well as combining PCA and K-means for an even more professional customer segmentation.

**Purchase Analytics**: The second part of the course explores both the descriptive and predictive analysis of the purchase behaviour of customers, including models for purchase incidence, brand choice, and purchase quantity. Not only that, but it also covers the application of state-of-the-art deep learning techniques to make predictions using real-world data.

# The STP Framework

Segmentation

Targeting

Positioning

STP is a fundamental marketing framework. It can be applied to all areas of business and marketing activities.

**B2C**

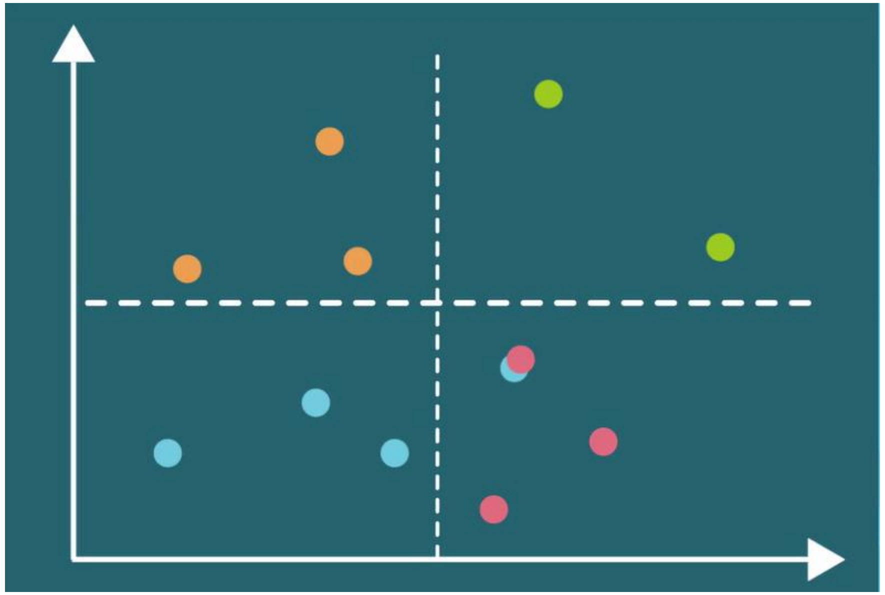
(business-to-customer)

The clients of our business are individuals rather than organizations.

The advantage of a B2C model in terms of data science is that we have much more data points.

# The STP Framework

## Segmentation



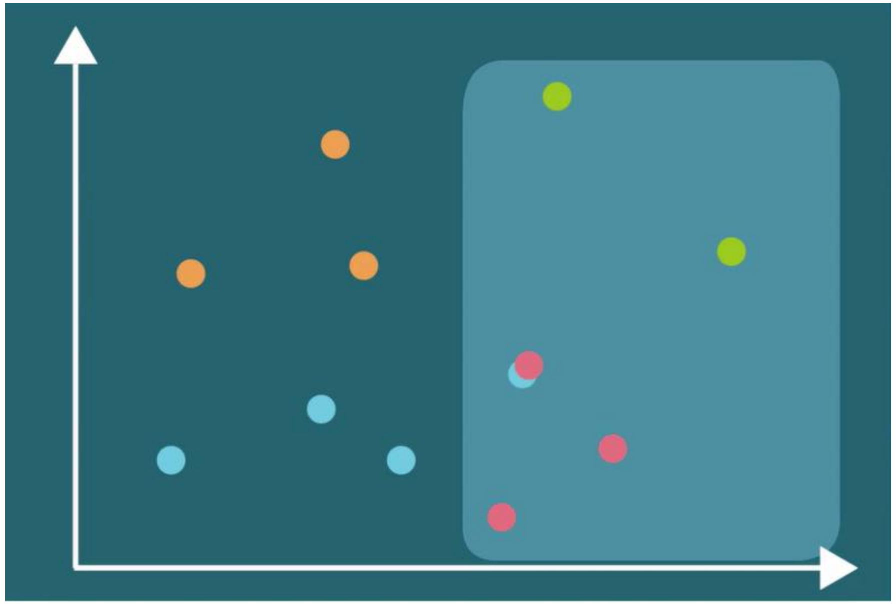
The process of dividing a population of customers into groups that share similar characteristics.

Observations within the same group would have comparable purchasing behavior.

Observations within the same group would respond similarly to different marketing activities.

# The STP Framework

## Targeting



The process of evaluating potential profits from each segment and deciding which segments to focus on.

Selecting ways to promote your products. You can ***target*** one segment on TV and another online.

Examining customers’ perception. *(Involves psychology and usually budget constraints).*

# The STP Framework

Positioning

Marketing Mix

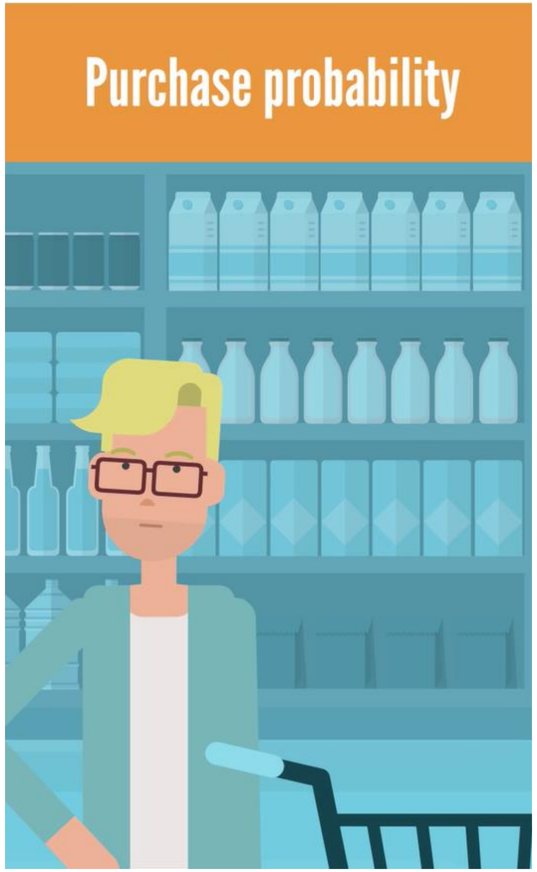
What product characteristics do the customers from a certain segment need?

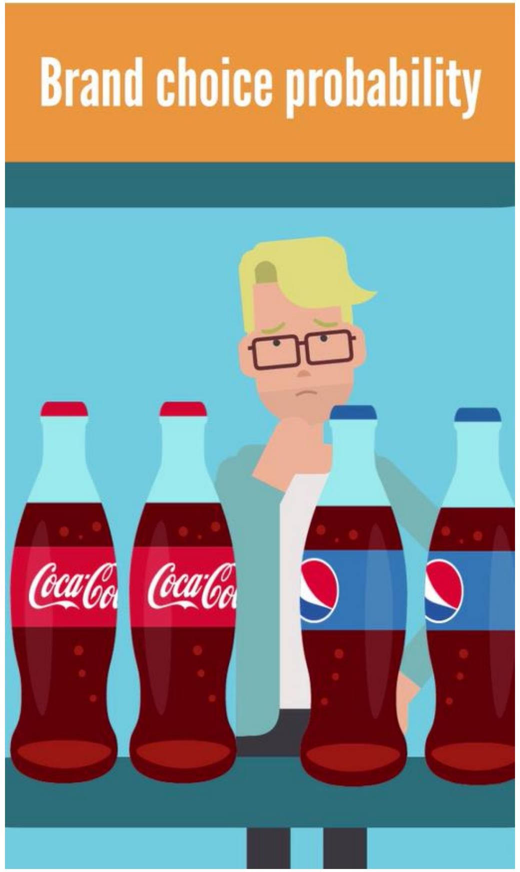
Shows how a product should be **presented** to the customers and through what **channel**.

In fact, this process is so important, that it has a framework of its own called: **Marketing Mix**.

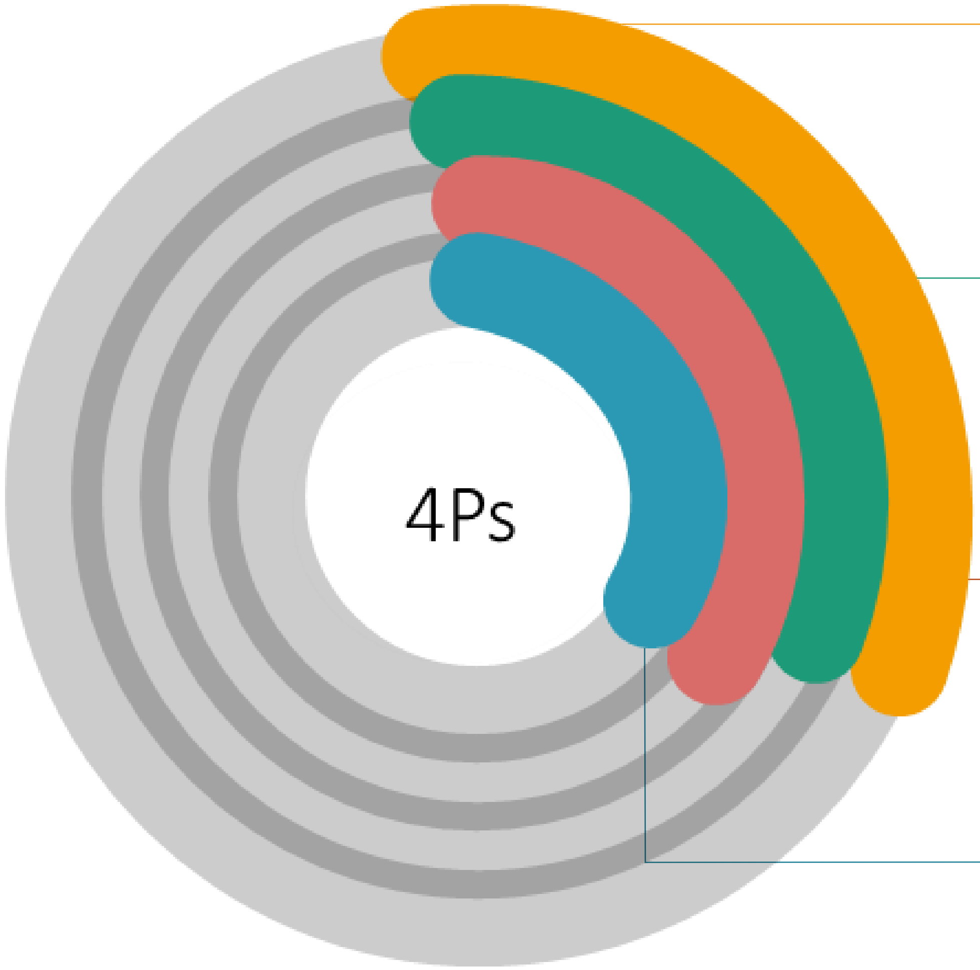
# Marketing Mix

Develop the *best* product or service and offer it at the *right price* through the *right channels*.




# Marketing Mix




**Product**
Product features;
Branding; Packaging

**Price**
Product cost;
Long term price changes

**Promotion**
Price reduction,
display and feature

**Place**
Distribution:
intensive, selective, exclusive

# Physical and Online Retailers

| Characteristics / Retailers | Physical | Online |
|---|---|---|
| Location | Many locations, returning customers. | One location, many customers. |
| Data | -- Fewer customers at a particular store, due to the physical restriction. | ++ More data points and more diverse customer information. |
| Returns | ++ In physical stores, customers can see the product itself. Returns are less likely. | -- Products are returned more often, as customers cannot see and test an item. |
| Purchase history | Gather information through loyalty cards. | Database with all past purchases of customers. |
| Brand choice | -- Unavailable in physical stores. We assume the customer has considered all competitor brands. | ++ We may have data for all products that the customer has looked at and which competing products a customer has considered. |
| Ratings and reviews | -- Unavailable in physical stores. | ++ Different items could be reviewed and rated (significant features for predictive modeling). |

# Price Elasticity

Price elasticity measures how a variable of interest changes when the price changes.

## Price Elasticity

$$E = \frac{\frac{\Delta Y}{Y}}{\frac{\Delta P}{P}}$$

Y: economic variable of interest
P: price

## Supply and Demand

*The cheaper the product
the higher the demand

$$Revenue_i = P_i * Q_i$$

Q: quantity
P: price

## Own Price Elasticity

Price elasticity with respect to the same product

## Cross Price Elasticity

Price elasticity with respect to another product

# K-means Clustering

01 Choose the number of clusters

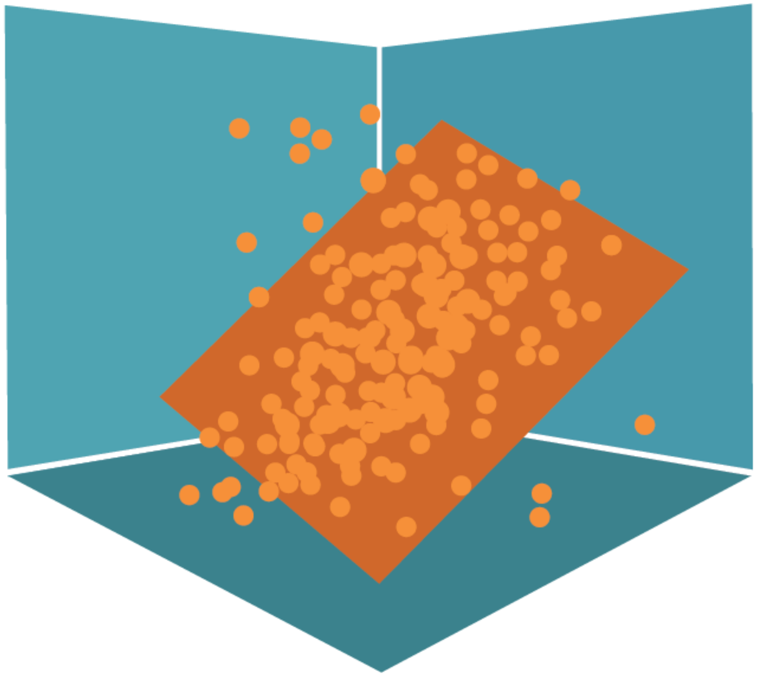
02 Specify cluster seeds

03 Calculate the centroid (geometrical center)

04 Repeat until the centroids stop changing

# Principal Components Analysis - PCA

The goal of PCA is to find the best possible subspace which explain most of the variance.
Most commonly it is used to reduce the dimensionality (number of features) of a problem.



**3D Plane**

**Dimensionality reduction**

**technique**

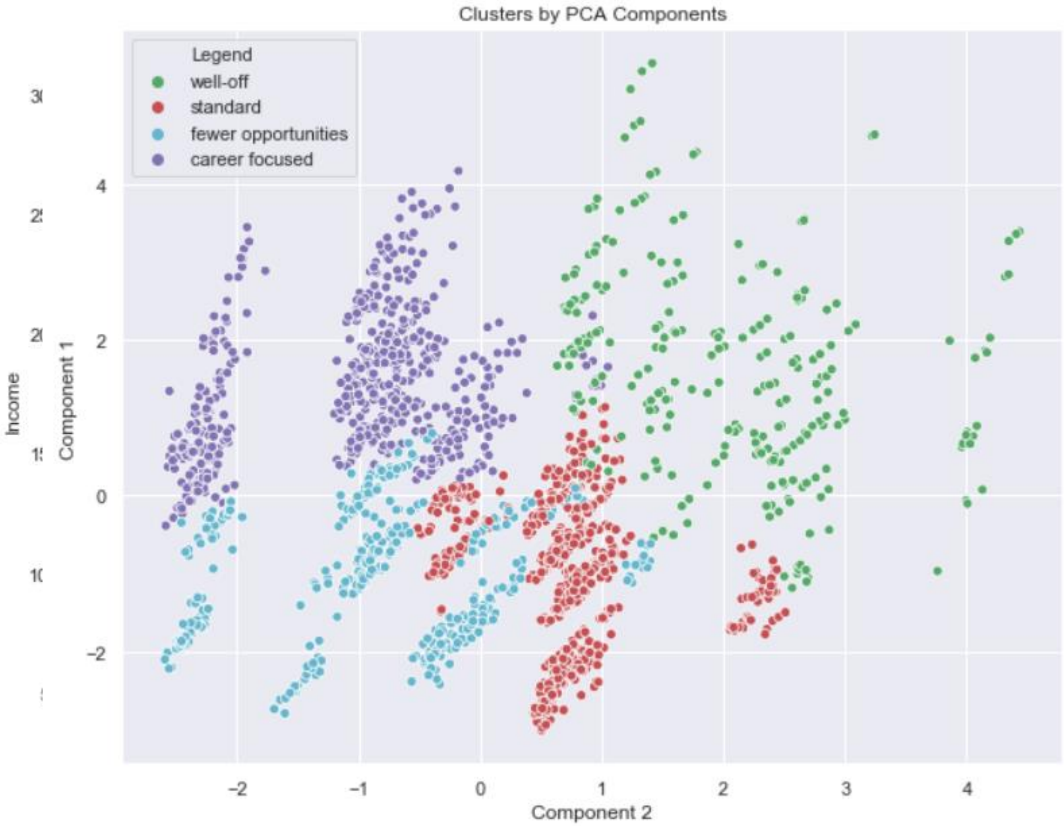
**2D Plane**

# K-means with PCA

01 Reduce Dimensionality with PCA

02 Perform K-means with PCA scores as features

03 Visualize and interpret clusters

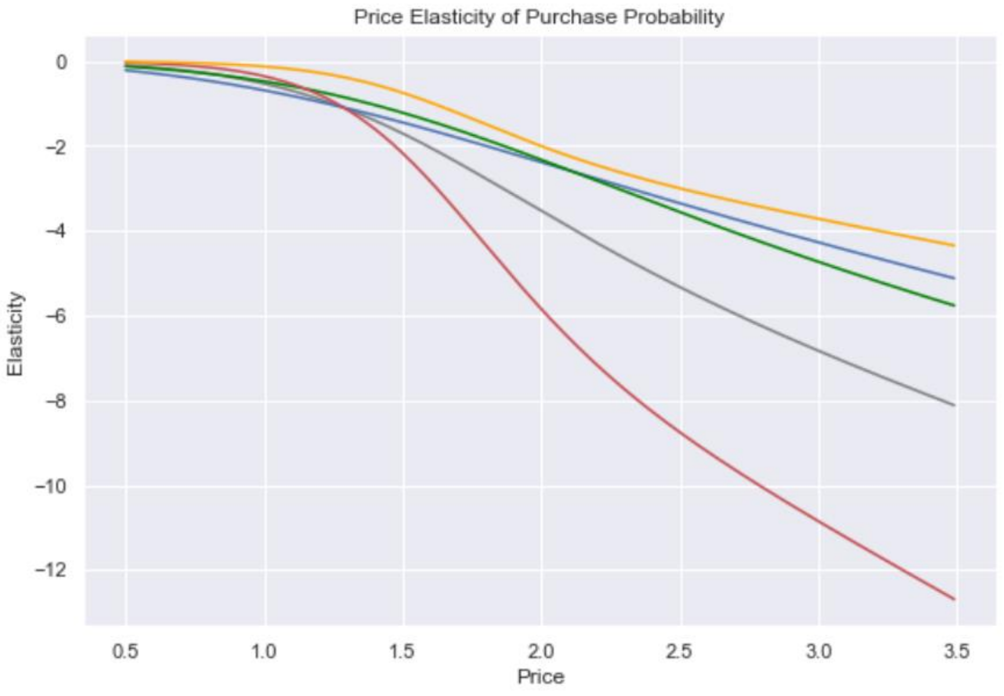


# Price Elasticity of Purchase Probability

Quantifies the change in probability of purchase of a product with a given change in its price

**Own-price elasticity of purchase probability**

$$E = beta * price * (1 - Pr(purchase))$$

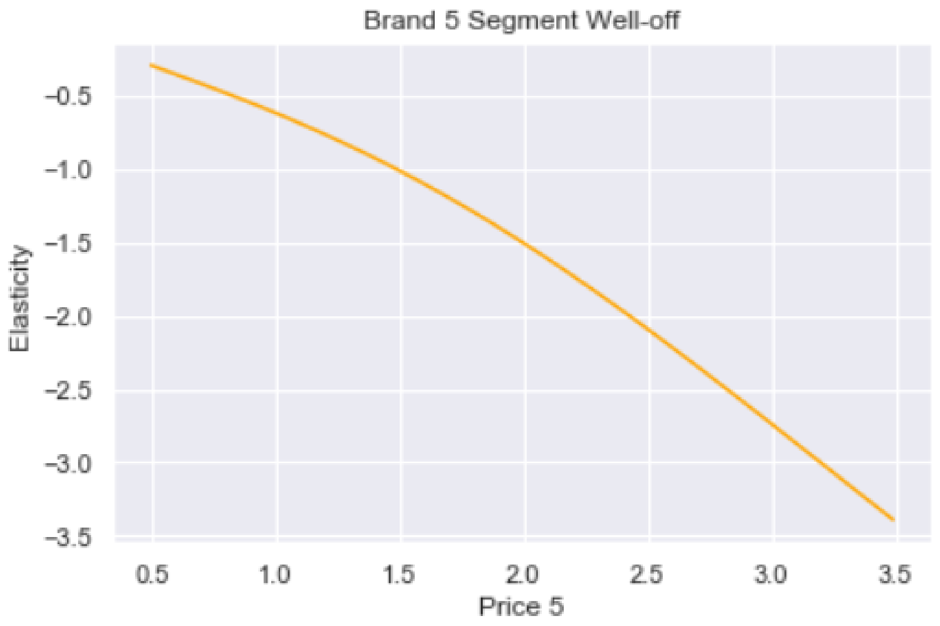


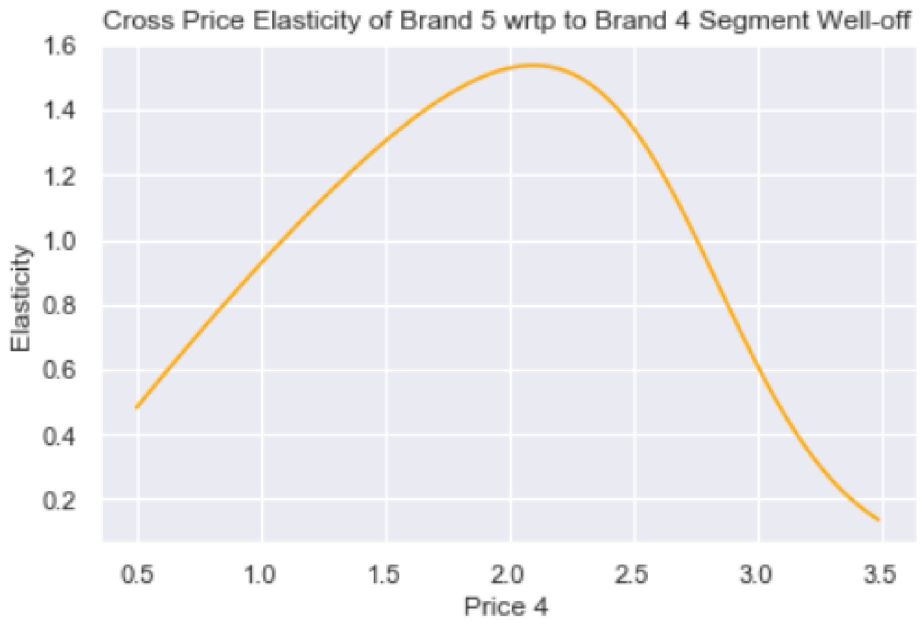
# Price Elasticity of Brand Choice Probability

Quantifies the change in probability of purchase of a product with a given change in a **competitor brand's** price

## Cross-price elasticity of brand choice

$$E = -beta(own\ price) * price(cross\ brand) * Pr(cross\ brand)$$



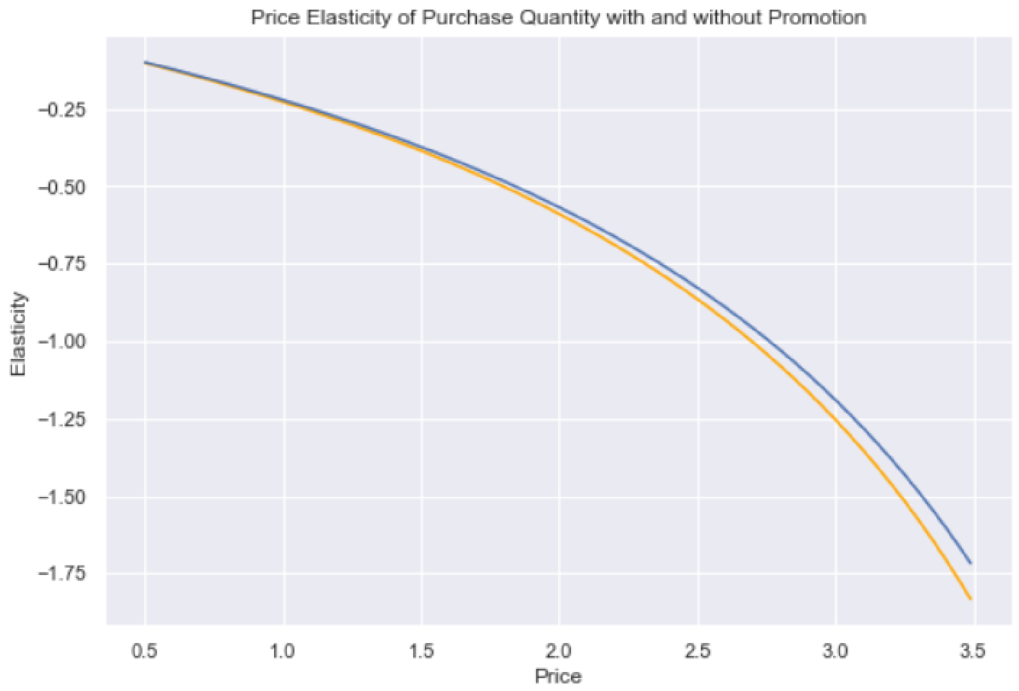
# Price Elasticity of Purchase Quantity

Quantifies the change in purchase quantity of a product with a given change in its price

**Price elasticity of purchase quantity**

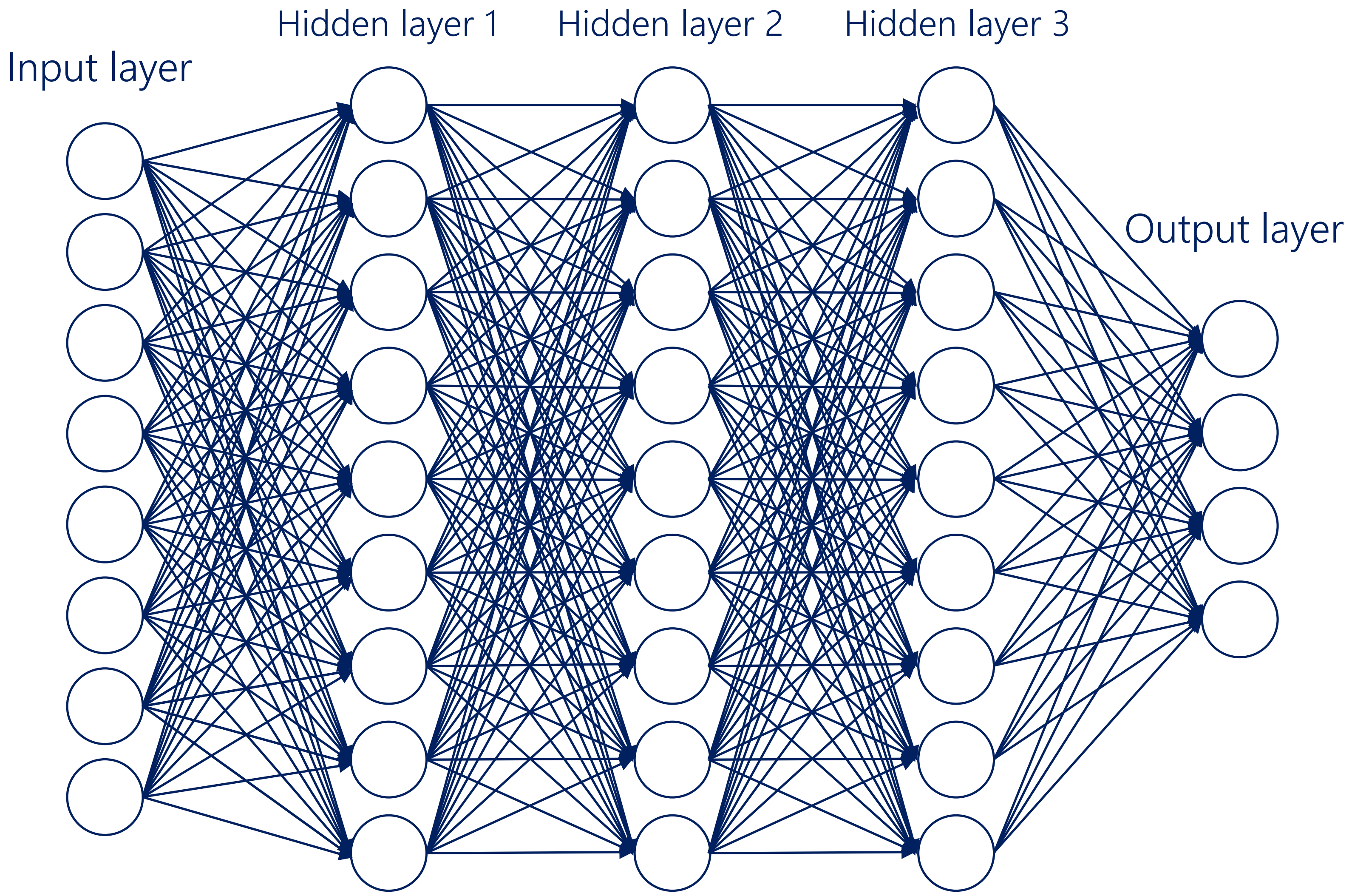
**Closest to price elasticity of demand*

$$E = beta * \frac{Price}{Quantity(purchase)}$$

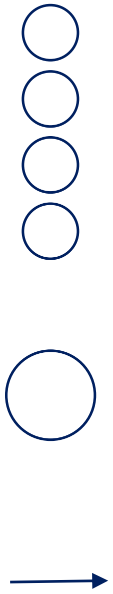


# A Deep Neural Network

This is a deep neural network (deep net) with 5 layers.




**How to read this diagram:**



A layer

A unit (a neuron)

Arrows represent mathematical transformations

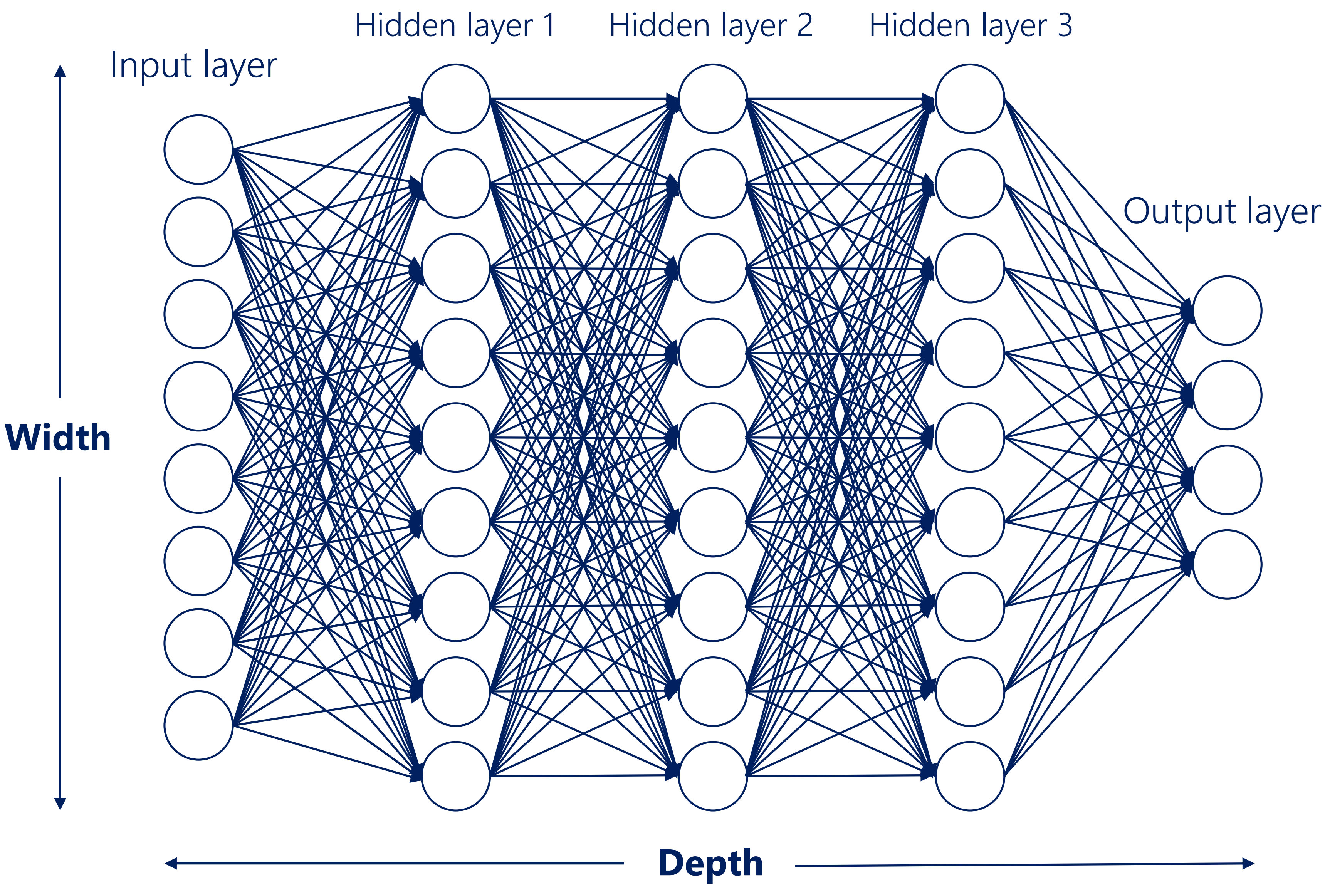
# A Deep Neural Network

The **width** of a layer is the number of units in that layer.

The **width** of the net is the number of units of the biggest layer.

The **depth** of the net is equal to the number of layers or the number of hidden layers. The term has different definitions. More often than not, we are interested in the number of hidden layers (as there are always input and output layers).

The width and the depth of the net are called **hyperparameters.** They are values we manually chose when creating the net.

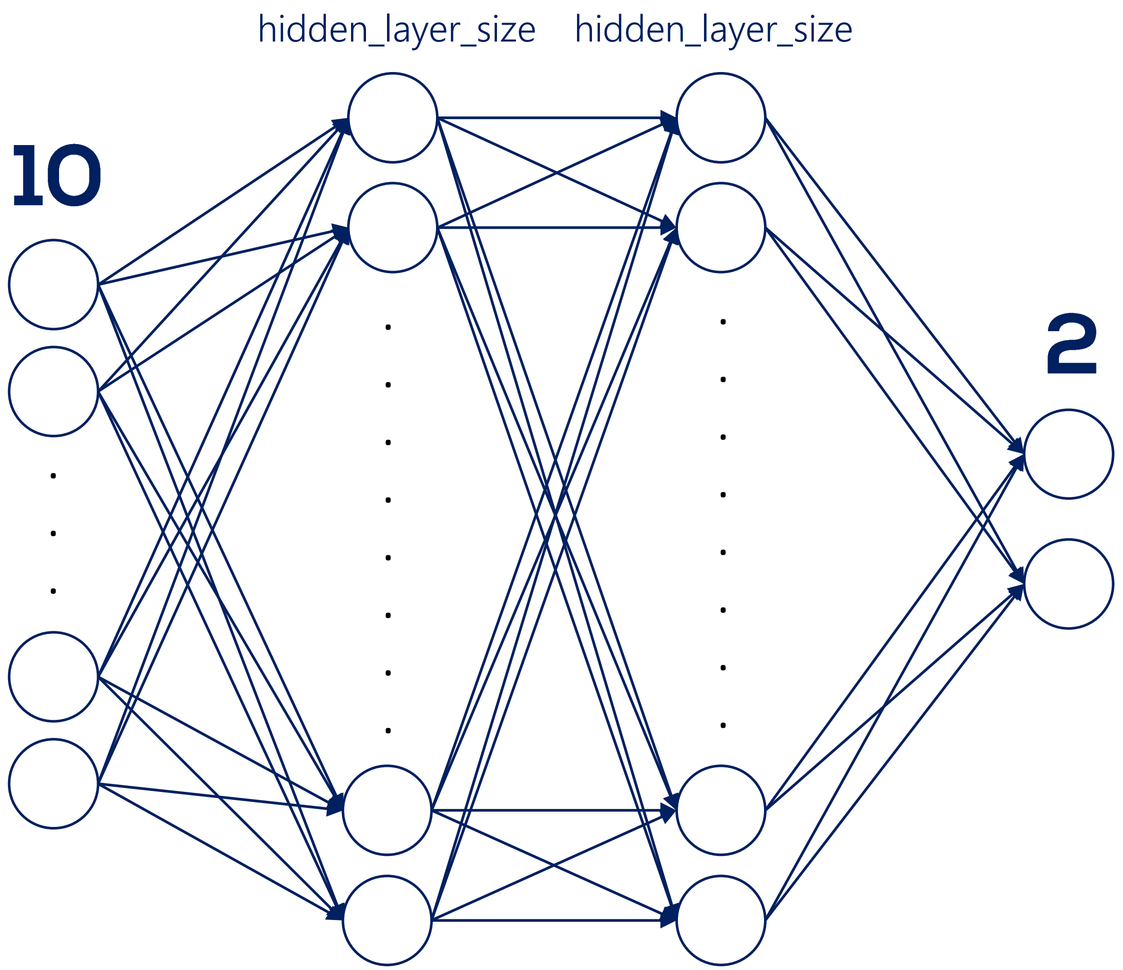


# The Business Case Deep Neural Net

Our neural network has 10 features; therefore the input layer size is 10.

The hidden layer size is a hyperparameter. We can adjust it during the learning process. In the lectures we start from 50 nodes, but its size could be any integer number.

The output layer contains the two possibilities for the targets (0 and 1), therefore it has a size of 2.




## How to approach similar problems?

1. Preprocess the data
    a. Balance the dataset
    b. Create train, validation and test sets
    c. Save the data in a tensor-friendly format
2. Train the model
    a. Outline the model (create or envision a diagram like the one on the right)
    b. Create the actual network and choose appropriate starting hyperparameters
    c. Optimize the model by fiddling with the hyperparameters
    d. Test the model
3. Save the model and deploy it where needed

# Price Elasticity of Purchase Probability

Y: Outcome (purchase probability)
P: Price
E: Elasticity

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Price + \beta_2 Promotion + \cdots)}}$$

*Y (purchase probability) is determined by a logistic regression*

Logistic regression coefficient of Price

Elasticity $\qquad E \qquad (1 - Y) * \beta_1 * P$ $\qquad$ Price

Purchase probability

# Price Elasticity of Purchase Probability

Y: Outcome (purchase probability)
P: Price
E: Elasticity

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 Price + \beta_2 Promotion + \cdots)}}$$

*Y (purchase probability) is determined by a logistic regression*

$$E = \frac{dY}{Y} \Big/ \frac{dP}{P} =$$ -> We start from the general price elasticity formula

$$= \frac{dY}{dP} * \frac{P}{Y} =$$ -> We simply rearrange the terms of the equation

$$= \frac{dY}{d(\beta_0 + \beta_1 Price + \beta_2 Promotion + \cdots)} * \frac{d(\beta_0 + \beta_1 Price + \beta_2 Promotion + \cdots)}{dP} * \frac{P}{Y} =$$ -> We apply the chain rule to determine the partial derivative of Y w.r.t. P, where the green expression is the interim function

$$= \frac{e^{-(\beta_0 + \beta_1 Price + \beta_2 Promotion + \cdots)}}{[1 + e^{-(\beta_0 + \beta_1 Price + \beta_2 Promotion + \cdots)}]^2} * \beta_1 * \frac{P}{Y} =$$ -> We find the partial derivative of Y, w.r.t. the green expression
-> We find the partial derivative of the green expression, w.r.t. P ($\beta_1$)

$$= Y * (1 - Y) * \beta_1 * \frac{P}{Y} =$$ -> We replace the value of Y from the logistic regression expression

$$= (1 - Y) * \beta_1 * P$$ -> We reach the final formula for the price elasticity of purchase probability

# Brand Choice Cross Price Elasticity

$E_{cross}$: Cross-price elasticity
$Y_i$ : Purchase probability of our own brand
$Y_j$ : Purchase probability of our competitor brand
$P_j$ : Price of competitor brand

$$Y_i = \frac{e^{(\beta_{0i}+\beta_{1i}Price_i+\beta_{2i}Promotion_i+\cdots)}}{\sum_{k=1}^{I} e^{(\beta_{0i}+\beta_{1i}Price_k+\beta_{2i}Promotion_k+\cdots)}}$$

*Y (purchase probability of a given brand i) is determined by a softmax function (multinomial logistic regression)*

Logistic regression coefficient of the price of **our** product

Purchase probability for competitor brand product

Cross-price Elasticity

$$E_{cross} = -\beta_{1i} * Pj * Y_j$$

Price of the competitor brand product

# Brand Choice Cross Price Elasticity

$E_{cross}$: Cross-price elasticity
$Y_i$ : Purchase probability of our own brand
$Y_j$ : Purchase probability of competitor brand
$P_j$ : Price of competitor brand

$$Y_i = \frac{e^{(\beta_{0i}+\beta_{1i}Price_i+\beta_{2i}Promotion_i+\cdots)}}{\sum_{k=1}^{I} e^{(\beta_{0i}+\beta_{1i}Price_k+\beta_{2i}Promotion_k+\cdots)}}$$

*Y (purchase probability of a given brand i) is determined by a softmax function (multinomial logistic regression)*

**For a softmax function:**
$\frac{\partial Y_i}{\partial x_j} = Y_i(\delta_{ij} - Y_j)$
where $\delta_{ij}$ is 1 if i=j, 0 otherwise*

$$E_{cross} = \frac{dYi}{Yi} \Big/ \frac{dP_j}{P_j} =$$ -> We start from the general price elasticity formula

$$= \frac{dY_i}{dP_j} * \frac{P_j}{Y_i} =$$ -> We simply rearrange the terms of the equation

$$= \frac{dY_i}{d(\beta_0+\beta_{1i}Price_j+\cdots)} * \frac{d(\beta_{0j}+\beta_{1ij}Price_j+\cdots)}{dP_j} * \frac{P_j}{Y_i} =$$ -> We apply the chain rule to determine the partial derivative of $Y_i$ w.r.t. $P_j$

$$= Y_i * (-Yj) * \beta_{1i} * \frac{P_j}{Y_i} =$$ -> We find the partial derivative of $Y_i$, w.r.t. the green expression (from the softmax derivative)
-> We find the partial derivative of the green expression, w.r.t. P ($\beta_{1i}$)

$$= -Yj * \beta_{1i} * P_j$$ -> We reach the final formula for the price elasticity of purchase probability

*This is the general case. The price elasticity of purchase probability (own brand) was a particular case.
If you are interested, you can practice by using this proof to derive the purchase probability elasticity