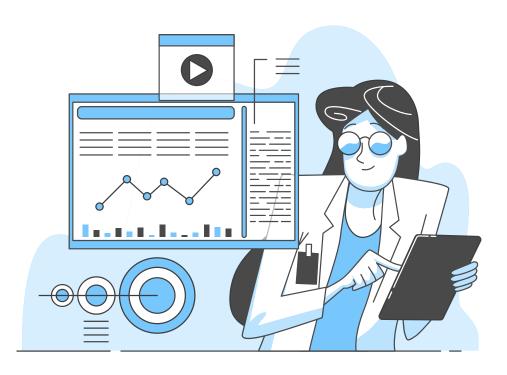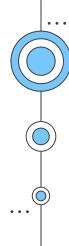# PREDICT MEAT PRODUCT CONSUMPTION

FINAL PROYECT
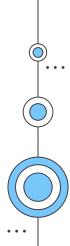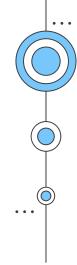## VARINI-BERRONDO

Table of Contents
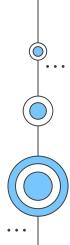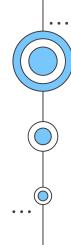
01
SUPERSTORE

Food retail store!

Discover the Universe of "Superstore"

For two years, we have painstakingly captured our clients' secrets and the data they revealed to us.
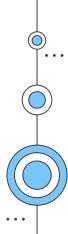
Dive into a world of demographic information where every detail helps us better understand our customers.
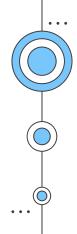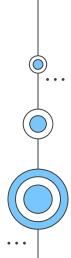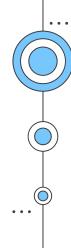
02

OBJECTIVE

What we want to achieve?

Our goal is clear: increase meat sales and decipher consumer behavior in the face of an exciting marketing campaign. How will we achieve it? Through the magic of data.
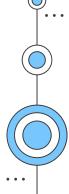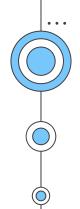
03

CONTEXT

What we have?
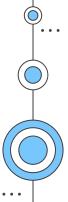
In the current commercial context, increasing the sale of meat can be based on:

- Diversification of the offer to satisfy customer preferences
- Identify factors that influence purchasing decisions.
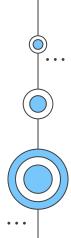
We will take an analytical approach that will examine all the factors that influence our meat eaters. We will explore educational levels, ages, genders, household types.
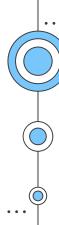
We will also closely analyze the behavior of those customers who prefer other foods. We will look for hidden patterns and trends that give us valuable insights to reposition our business strategies.

04
DATASET

The Dataset contains customer information, including age, education, marital status, income, and purchase history, used to predict customer response to a marketing campaign.

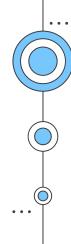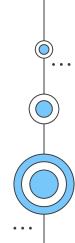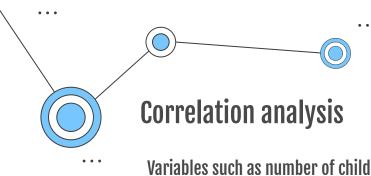| COLUMN	DESCRIPTION	DATATYPE
ID	Unique ID of each customer	Integer
Year_Birth	Age of the customer	Integer
Complain	1 if the customer complained in the last 2 years	String
Dt_Customer	date of customer's enrollment with the company	DateTime
Education	customer's level of education	String
Marital	customer's marital status	String
Kidhome	number of small children in customer's household	Integer
Teenhome	number of teenagers in customer's household	Integer
Income	customer's yearly household income	Decimal
MntFishProducts	the amount spent on fish products in the last 2 years	Integer
MntMeatProducts	the amount spent on meat products in the last 2 years	Integer
MntFruits	the amount spent on fruits products in the last 2 years	Integer
MntSweetProducts	amount spent on sweet products in the last 2 years	Integer
MntWines	the amount spent on wine products in the last 2 years	Integer
MntGoldProds	the amount spent on gold products in the last 2 years	Integer
NumDealsPurchases	number of purchases made with discount	Integer
NumCatalogPurchases	number of purchases made using catalog (buying goods to be shipped through the mail)	Integer
NumStorePurchases	number of purchases made directly in stores	Integer
NumWebPurchases	number of purchases made through the company's website	Integer
NumWebVisitsMonth	number of visits to company's website in the last month	Integer
Recency	number of days since the last purchase	Integer
Response	1 if customer accepted the offer in the last campaign, 0 otherwise	Integer
Complain	1 if the customer complained in the last 2 years	Integer
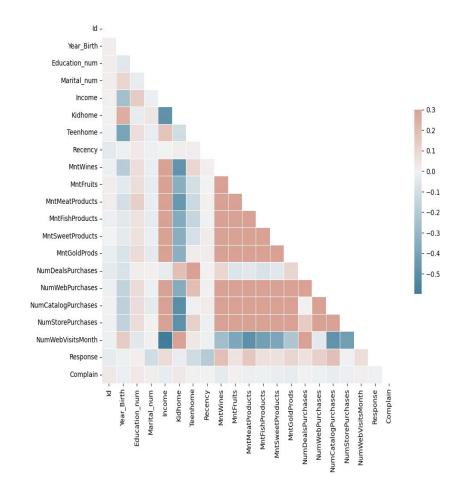
05

EXPLORATORY DATA ANALYSIS

Correlation analysis

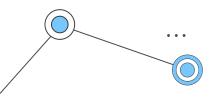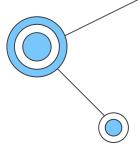Variables such as number of children, year of birth, and adolescents in the home would not show a correlation with our impact variables (type of consumption), so they would not be relevant to explain the purchasing behavior of our customers.
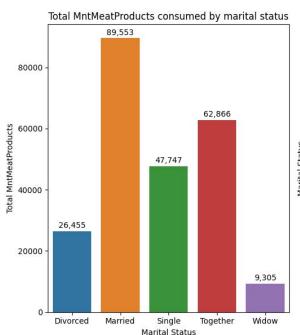
On the other hand, regarding the purchase or consultation channel, a strong correlation is observed, with a higher prevalence of certain products depending on the channel of origin.
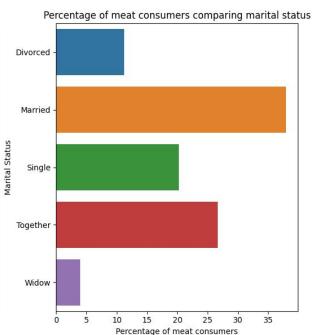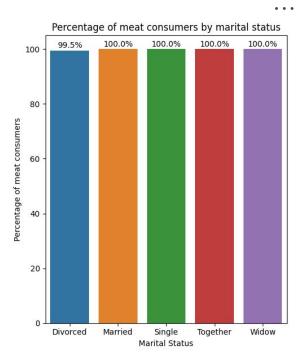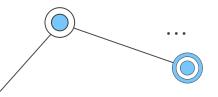
1 What are the marital states that consume more meat and its percentage?

Total MntMeatProducts consumed by marital status

89,553 — Married
62,866 — Together
47,747 — Single
26,455 — Divorced
9,305 — Widow

Percentage of meat consumers comparing marital status

Percentage of meat consumers by marital status

Divorced 99.5% | Married 100.0% | Single 100.0% | Together 100.0% | Widow 100.0%
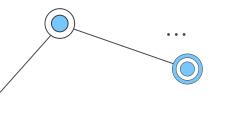
Marital Status

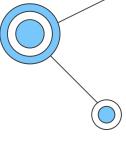It indicates that the majority of individuals across all marital statuses are meat consumers, with slight variations observed among the categories.
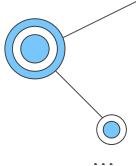
Also, It can be observed that Married individuals tend to consume the highest amount of meat products compared to other marital statuses, followed by Together and Single
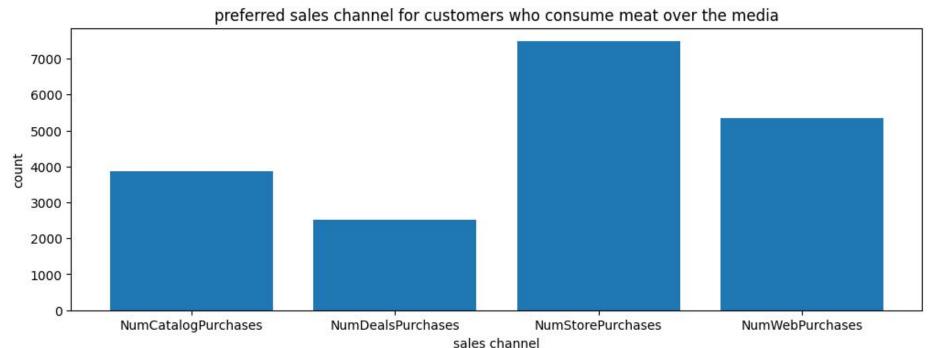
This suggests that marital status plays a role in determining meat consumption patterns.
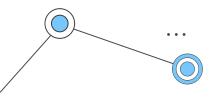
2 What is the preferred sales channel for customers who consume meat and are over the mean value?

preferred sales channel for customers who consume meat over the media
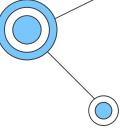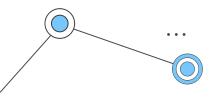
Sales channel
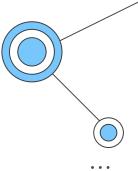
Among the identified sales channels, customers who consume meat above the average tend to make a relatively higher count of purchases through the NumStorePurchases and NumWebPurchases channels.
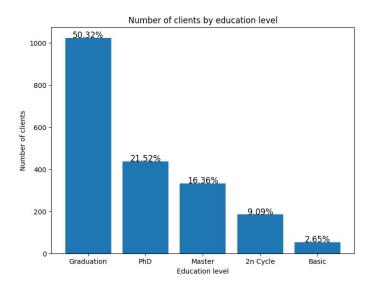
This implies that these channels are influential in driving meat product sales and should be targeted in marketing strategies.
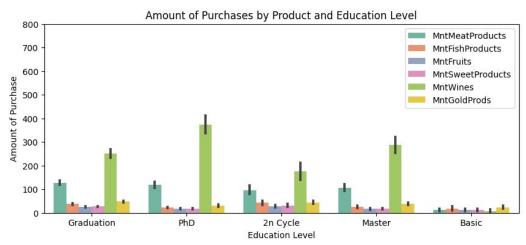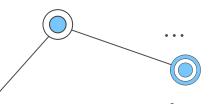
3 What kind of education do customers who consume meat have?

Number of clients by education level

Amount of Purchases by Product and Education Level
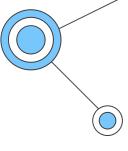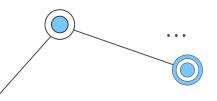
Education
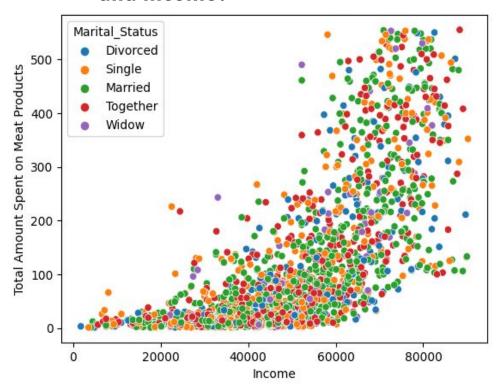
There is a clear difference between the different types of education, being Graduation the one with the highest number of client followed by PhD and Master, but we can see that expecting Basic all the other education types consume same percentage of meat.
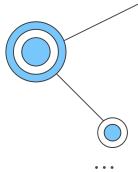
This indicates that education level may not have a significant impact on meat consumption preferences. We can also observe and separate the largest number of users who consume in the store for future analysis.
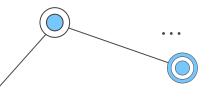
4 What relation is between meat consumption and income?
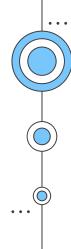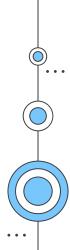
Income

First of all, income has a significant influence on the total amount spent on meat products. People with higher incomes tend to allocate a greater part of their resources to meat consumption. However, consumption is also observed in those with low incomes. Therefore. This indicates that meat consumption is not limited to a specific income group and that there are opportunities to attract customers of different economic levels.

06

MODELING

Trying to predict
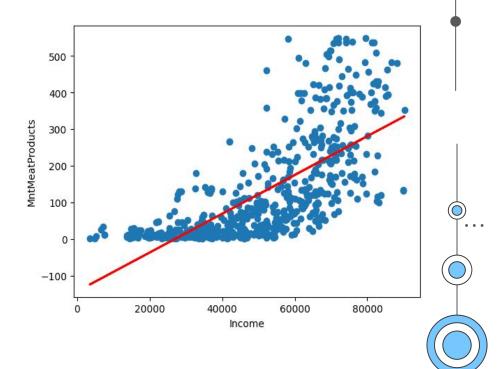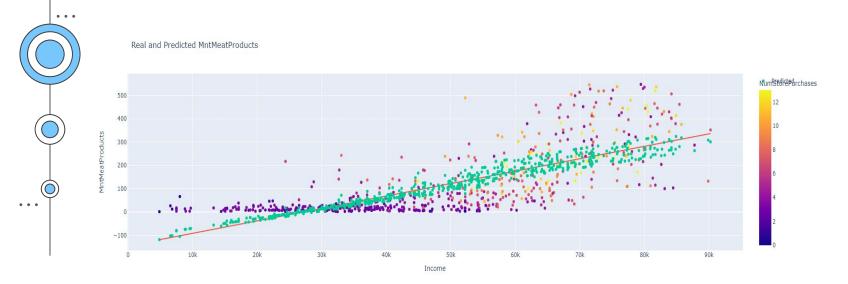
LINEAL REGRESSION

We are going to work with a linear regression model because it establishes a linear relationship between the dependent variable (meat consumption) and the independent variables (characteristics or predictor variables). If meat consumption is expected to increase or decrease proportionally to certain predictor variables, the linear regression model is appropriate.

The R-squared value is 0.50, which means that approximately 50% of the variance in the amount spent on meat products can be explained by the income variable.

However, it's important to note that the model's performance can be further improved, as there is still a significant amount of unexplained variance in the data (as indicated by the MSE and R-squared values).

Real and Predicted MntMeatProducts

By adding the variable 'NumStorePurchases', we achieved a higher R-squared value of 0.5424, indicating that it explains a larger proportion of the variance in meat product consumption compared to the first model's R-squared value of 0.50. Additionally, has a lower mean squared error (MSE) of 9165.63, indicating smaller prediction errors compared to the first model's MSE of 9992.53.
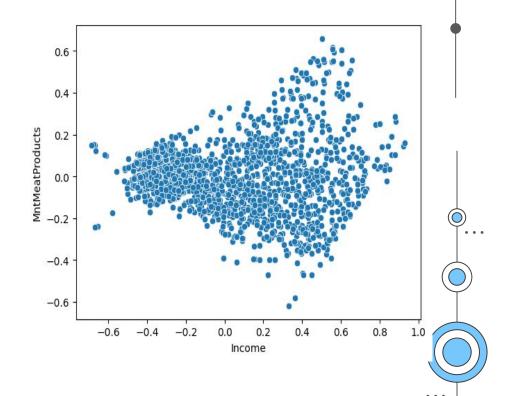
In conclusion, the second model, which incorporates both income and the number of store purchases as predictors, demonstrates superior performance in predicting meat product consumption compared to the first model that solely relied on income. The second model exhibits a higher R-squared value, lower MSE, and lower MAPE, indicating a better fit to the data and more accurate predictions. The inclusion of relevant additional variables in the modeling process can significantly improve the predictive power and accuracy of the model.
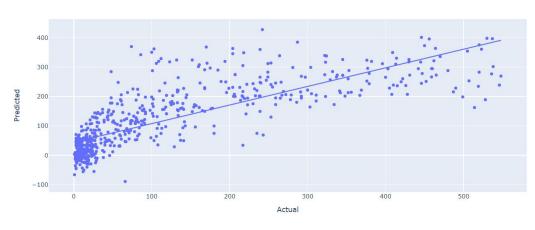
PCA (Principal Component Analysis)

The PCA results provide insights into the variance explained by each component and the cumulative variance explained by the selected number of components. This analysis aids in dimensionality reduction and understanding the relative importance of the components in capturing the variability within the data.

The results show that applying PCA with 2 components explains approximately 81.42% of the total variance of the data. The first principal component explains 67.9% of the total variance and the second principal component explains 13.51%.

Actual vs Predicted MntMeatProducts

The **R-squared value is 0.63.** This indicates that the model captures about 63% of the variability in the data, suggesting a moderate level of predictive power. The mean squared error (MSE) is 9165.63 The mean absolute percentage error (MAPE) is 2.71.

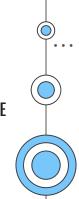Based on these results, we can conclude that the linear regression model, trained on the principal components obtained from PCA, demonstrates a reasonable level of predictive performance. The R-squared value indicates that the model captures a significant portion of the target variable's variability.

However, it is important to note that there is still room for improvement, as indicated by the moderate MSE and MAPE values. Further refinement of the model or exploration of alternative algorithms may lead to better predictive accuracy. Additionally, it is worth considering the interpretability of the transformed features derived from PCA and their impact on the model's predictions.

FEATURE ENGINEERING

Total Food Spending

EVALUATION MSE –
Test: 5736.74,
Train: 4921.14

Mean Absolute Error (MAE) – Test: 48.625 Root

Mean Squared Error (RMSE) – Test: 75.741

R-squared (R2) Score: 0.711

Family Size

EVALUATION MSE – Test: 5499.49, Train: 4677.34

Mean Absolute Error (MAE) – Test: 49.565

Root Mean Squared Error (RMSE) – Test: 74.159

R-squared (R2) Score: 0.723

POLYNOMINAL REGRESSION
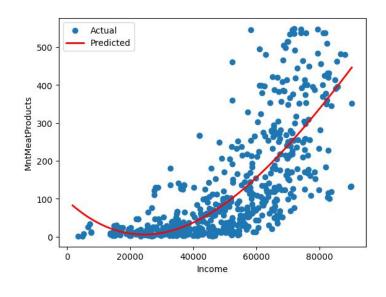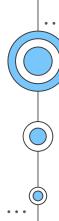
Comapring Linear vs Polynominal

LINEAR REGRESSION:
- Mean Absolute Error (MAE) – Test: 76.670
- Mean Squared Error (MSE) – Test: 9992.53
- Root Mean Squared Error (RMSE) – Test: 99.963
- R-squared (R2) Score: 0.497

POLYNOMIAL REGRESSION:
- Mean Absolute Error (MAE): 62.691
- Mean Squared Error (MSE): 8809.579
- Root Mean Squared Error (RMSE): 93.859
- R-squared (R2) Score: 0.556

The polynomial regression model out performs the linear regression model in MAE, MSE, RMSE and R2, indicating that the polynomial regression model has smaller errors in predicting the target variable and explains more variance in the target variable.
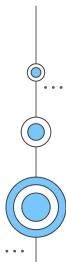
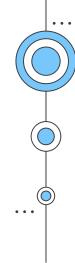# CLASSIFICATION BY MEANS OF DECISION TREE

Select features for training:
- Income
- MntFishProducts
- NumStorePurchases
- NumWebVisitsMonth

Results:
- Accuracy: 0.7659574468085106
- Recall: 0.6733322102425876
- Precision: 0.6653054264783093
- F1-score: 0.6658552694261265
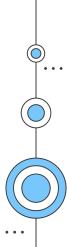- AUC: 0.8280771935848784

The results obtained from the Decision Tree Model show outstanding performance in terms of accuracy, sensitivity and precision, with a high ability to correctly classify customers with high potential to increase meat consumption. These findings suggest that the model is highly effective in identifying specific customer groups and targeting personalized marketing and promotional strategies to increase meat consumption.

07

CONCLUSIONS

Our insights

After conducting a thorough analysis to increase meat sales by predicting meat product consumption in response to a new marketing campaign, we have gained valuable insights
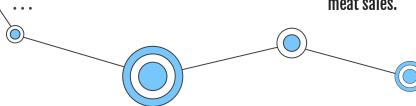
01 We observed that customers with *higher incomes* and who made more purchases through various channels, such as web and catalog, tend to consume more meat products.

02 Moreover, customers with *higher education levels* showed a greater preference for meat products.

03 We found positive correlations between meat product consumption and variables such as income, *number of purchases*, and educational level. These factors can significantly influence meat sales.
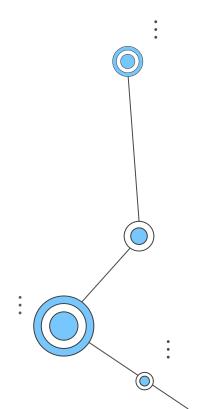
04 Using *linear regression models*, we predicted meat product consumption based on various customer attributes. We evaluated the models using metrics such as R-squared value, mean squared error (MSE), and mean absolute percentage error (MAPE). The results indicated that the models had *moderate predictive power*, capturing a significant portion of the variability in meat consumption.

05 After applying *polynomial regression* we observe a *better performance* in reducing the predicting errors but still a *moderate predictive power again*.

06 Decision Tree Model demonstrated outstanding performance in classification, with an accuracy, sensitivity and precision of 99.90%. This indicates that the model is highly effective in identifying customer groups with high potential to increase meat consumption.

Thanks!