CT College Dublin

**Assessment Cover Page**



| **Module Title:** | Strategic Thinking |
| --- | --- |
| **Assessment Title:** | Report 2 CA2 |
| **Lecturer Name:** | Neil Doyle |
| **Student Full Name:** | Daniela Maftei |
| **Student Number:** |  |
|  |  |
|  | sba24270 |
|  |  |
|  |  |
| **Assessment Due Date:** | 29/12/2024 |
| **Date of Submission:** | 17/12/2024 |



**Declaration**



Table of Contents

[**Introduction 2**](#_heading=h.vss196g43mui)

[**CRISP-DM Model 2**](#_heading=h.urfs13295dfh)

[Porter's Five Forces 3](#_heading=h.woettu1hclw2)

[**Business Objective 3**](#_heading=h.nh6p1iy27uj2)

[Key Questions to Answer 3](#_heading=h.115797sz5nqu)

[**Business Understanding 4**](#_heading=h.9zp06i4jwe3l)

[About the Dataset 4](#_heading=h.5kph66zqwpb)

[**Exploratory Data Analysis (EDA) 5**](#_heading=h.ptdpe33lhasl)

[**Data Visualization 5**](#_heading=h.ayj104m5ijae)

[**Correlation Analysis 5**](#_heading=h.b611k0lvz4oo)

[**Data Preparation 5**](#_heading=h.suu597qkud0g)

[**Data Cleaning 5**](#_heading=h.8txqkycbxhtg)

[**Data Transformation 5**](#_heading=h.z70a1h66nxf4)

[**Modeling 5**](#_heading=h.7bwtr4d9efsr)

[Customer Segmentation 5](#_heading=h.7aqwqfpbju0k)

[**Sales Forecasting 5**](#_heading=h.dut1q8c51am8)

[**Recommendation Systems 5**](#_heading=h.cc6qm6jefru1)

[**Findings and Recommendations 6**](#_heading=h.30j0zll)

[**Key Performance Indicators 10**](#_heading=h.im30cdtzgc7w)

[Total Revenue 10](#_heading=h.c7rq8zv79ixb)

[Average Amount Spent per Order 10](#_heading=h.c7rq8zv79ixb)

[Total Pizzas Sold 10](#_heading=h.c7rq8zv79ixb)

[Average Pizzas per Order 10](#_heading=h.6den6ba1trji)

[Total Orders 10](#_heading=h.c7rq8zv79ixb)

[Highest and Lowest Revenue Generator Categories 11](#_heading=h.twfi7x75dgrs)

[Most and Least Ordered Categories 11](#_heading=h.1qegj4vybg2i)

[Seasonal Analysis Statistics 11](#_heading=h.ubf96s8g3l34)

[Pizza Analysis Statistics 12](#_heading=h.w21tvaqvuof)

[Pizza with Highest and Lowest Price 12](#_heading=h.hadofealv6xc)

[Most Used Ingredients 12](#_heading=h.lon82c1h0bom)

[References 13](#_heading=h.3j2qqm3)

# Introduction

This assignment requires the compilation of a comprehensive analysis of a chosen dataset using a project management methodology. I have chosen the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology as showcased by Taylor, 2017.

# CRISP-DM Model

1. Business Understanding: Defining the business problem and its significance.
2. Data Understanding: Exploring the dataset's structure, quality, and initial insights.
3. Data Preparation: Cleaning, transforming, and preparing the data for analysis.
4. Modeling: Applying appropriate techniques to extract insights.
5. Evaluation: Assessing the model's performance and refining it as needed.
6. Deployment: Integrating the model into a real-world application.

Also acknowledging this is an open process prone to reevaluation, transformation and adaptation to Porter’s 5 forces.

# 

| **CRISP-DM Model Taylor, 2017** | **Porter’s Five Forces** |
| --- | --- |

### 

### Porter's Five Forces

Porter's Five Forces is a framework for analyzing industry competitiveness. Applying it to the pizza sales context can provide insights into the competitive landscape in terms of Threat of New Entrants, Barriers to Entry**,** Bargaining Power of Suppliers**,** Supplier Concentration, Bargaining Power of Buyers, Customer Concentration, Threat of Substitute Products, Substitute Availability, Competitive Rivalry and Intensity of Competition. The number of competitors, their market share, and their competitive strategies can impact market dynamics.

By analyzing these forces, we can identify potential threats and opportunities for our business. For example, if the threat of new entrants is low, the company can focus on differentiating its products and improving customer loyalty.

For the purpose of getting an understanding into the pizza restaurants market, I have examined a dataset pizza\_sales.csv from kaggle, that contains the transactional data for the year 2015 of the restaurant Plato's Pizza. The information extracted from this dataset will decide whether using such a business model to our restaurant will increase revenue and profits.

# Business Objective

To assess the feasibility of implementing a similar business model as Plato's Pizza to increase revenue and profits of our cafe restaurant business.

To achieve this, we can identify customer preferences by analyzing purchase history to tailor our offerings. As well as optimize staffing by scheduling staff based on peak hours and customer demand, and improving marketing strategies through personalized promotions.

## Key Questions to Answer

Based on the business objective and the available data, we can formulate the following questions in product popularity and sales trends:

* + What are the best-selling and worst-selling pizza items?
  + What is the most popular pizza size?
  + What is the busiest month of the year for pizza sales?
  + What are the busiest days of the week?
  + What are the peak hours for pizza orders?
  + What are the most popular pizza toppings?

By answering these questions, we can gain valuable insights into customer preferences, sales trends, and operational efficiency. These insights can be used to inform business decisions such as menu optimization, inventory management, and staffing schedules.

# Business Understanding

## 

## About the Dataset

This pizza\_sales dataset contains information about pizza sales. Each column represents a specific attribute related to the pizza sales. It’s made up of 48621 orders of pizza and customer transactions and make up 12 relevant features:

* order\_details\_id: a unique identifier for each pizza placed within each order
* order\_id: a unique identifier for each order placed
* pizza\_id: a unique key identifier that ties the pizza ordered to its details, like size and price
* quantity: the quantity ordered for each pizza of the same type and size
* order\_date: the date the order was placed
* order\_time: the time the order was placed
* unit\_price: the unit price of the pizza
* total\_price: unit\_price x quantity
* pizza\_size: the size of the pizza (Small, Medium, Large, X Large, or XX Large)
* pizza\_category: the unique key identifier that ties the pizza ordered to its details, like size and price
* pizza\_ingredients: toppings used for pizza
* pizza\_name: the name of each pizza in the menu

Tools used for Data Visualisation: I imported pandas, numpy, matplotlib, seaborn and plotly libraries and kept it simple in order to focus on the key findings.

**I** checked for missing values, outliers, and inconsistencies and assessed data types and format. Also verified data accuracy and completeness.

# Exploratory Data Analysis (EDA)

I used Descriptive Statistics and calculated measures like mean, median, mode, standard deviation.

# Data Visualization

Icreated visualizations (histograms, pie charts, box plots, scatter plots) to understand data distribution, relationships, and trends.

# Correlation Analysis

Attempted correlation analysis toidentify relationships between variables.

# Data Preparation

# Data Cleaning

To handle missing values (imputation or removal), address outliers (capping, flooring, or removal) and correct data inconsistencies and errors.

# Data Transformation

Normalization or standardization of numerical variables.

Looked into encoding categorical variables (label encoding) to potentially create new features.

# Modeling

## Customer Segmentation

Clustering techniques could be used (e.g., K-means, hierarchical clustering) to group customers based on similar behaviors.

# Sales Forecasting

I will apply time series forecasting models to predict future sales.

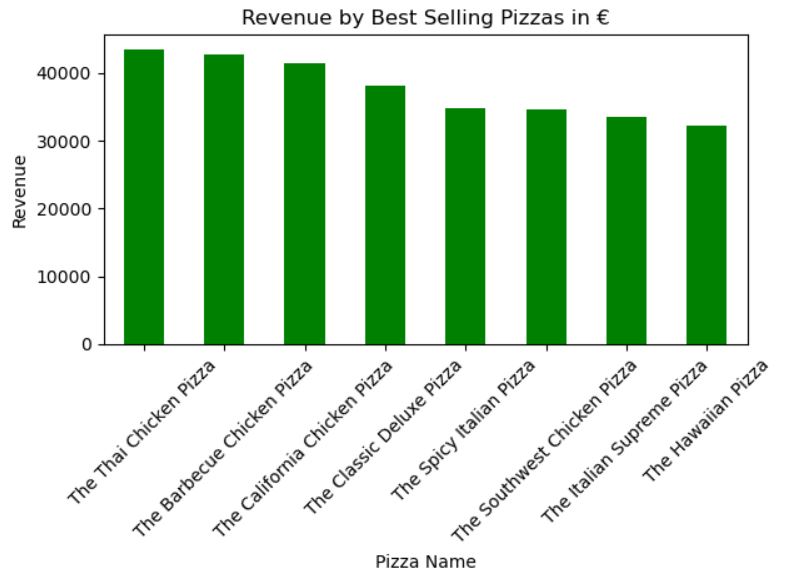
# Recommendation Systems

Implement collaborative filtering or content-based recommendation systems to suggest products or services.

# Findings and Recommendations

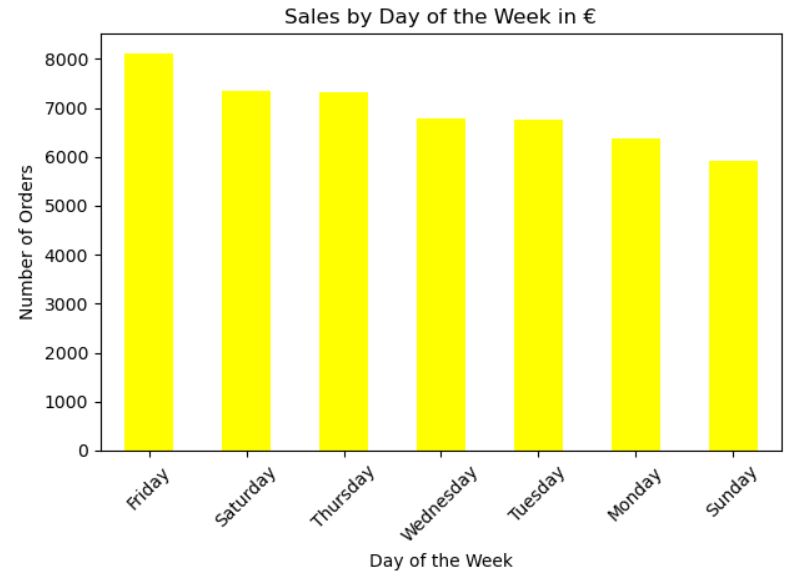
The average of 2.32 pizzas (>2) per order indicates the main customer base is made up of groups and families.

It looks like the best selling pizzas are The Thai Chicken Pizza, followed by The Barbeque Chicken Pizza, The California Chicken Pizza, The Classic Delux Pizza and the fifth best selling pizza is The Spicy Italian Pizza. Pizza preferences are usually trending everywhere. This is very helpful in compiling our own pizza menu. These five pizza recipes will definitely be on the menu. The Thai Chicken Pizza, despite limited availability, brings a high unit revenue, attracting a niche market willing to pay for its unique flavour.



The pizza size most sold is the Large one. We also know that the less revenue brings the XXL pizza. This helps in keeping the menu as simple and profitable as possible. For this reason we won’t be having XXL or XL pizza size in our menu. This is good to know because the largest sizes are the most unprofitable sizes to produce.

Looking at the days of the week we observe that the busiest day is Friday followed by Saturday and Thursday. This is common within the hospitality industry and is valuable in preparing with the right amount of staff and enough stock to be able to keep up with customers’ needs. We also notice that Sunday is the quietest day. It can also depend on the location, however this will also help in deciding whether it will be more cost efficient to close on Sundays.



In finding the busiest time of the day, I first tried with counts of 5, then 10, and I realised the sales are still high around lunch time, so I expanded to 30 counts for a more accurate result. Therefore the busiest time of the day is between 11:57 to 13:33. This is again a very common occurrence in the industry at lunch time. Considering we grouped the data by delivery\_time, allowing for the cooking time as well, we can estimate that we should be busiest between 11:45 to 13:30. This helps in being prepared with the right amount of staff and ingredients.

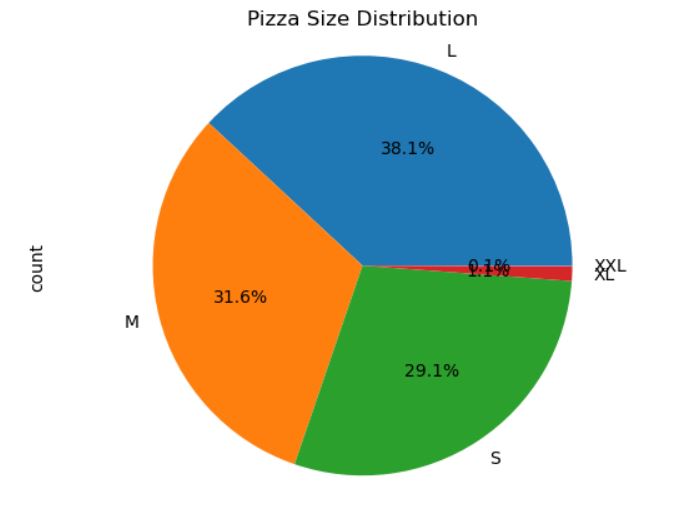


In terms of seasonal demand, the data shows that the busiest month is the month of July. This helps to better plan for staff annual leaves. We are interested in the least busy months. Those being October in the winter and June in the Summer.

In terms of the most popular pizza based on three categories: pizza\_name, pizza\_size and order\_id, the winner is The Big Meat Pizza in size S. The Big Meat Pizza's high demand shows a preference for its meaty toppings, even if it's not the highest revenue generator.

The lowest priced pizza is The Pepperoni Pizza at €9.75 and the highest priced pizza is The Greek Pizza at €35.95. The latter is quite expensive. It might be linked to the highly priced ingredients used and the extra time required to produce it. I will work on a more cost efficient way to source ingredients and produce it at a lower price. In terms of pricing I will rework some of the prices to fit our own business.

By calculating the pizza count by type, I produced pie charts for a better visualisation , and in the pizza size distribution it shows the huge gap between the better sold sizes L, M and S opposed to less sold XL and XXL sizes. Which makes it obvious to create a menu with the pizza sizes Large, Medium and Small only.



Garlic and Tomatoes are the most popular ingredients/ pizza toppings. Followed by Red Onions and Red Peppers. The most relevant toppings are the top 9. Knowing this can help optimize stock levels, reduce wastage and increase profits.

Correlation is a statistical measure that indicates the strength and direction of the linear relationship between two variables. A correlation coefficient ranges from -1 to 1 and the closer it is to -1 or 1 the bigger the correlation.



The Quantity and Unit Price correlation of 0.007142 is very low. This indicates a negligible linear relationship between the quantity of pizzas ordered and their unit price. This makes sense, as the quantity of pizzas ordered typically doesn't affect the individual price of each pizza.

Quantity and Total Price correlation of 0.541926 is a moderate positive correlation. This suggests that as the quantity of pizzas ordered increases, the total price of the order tends to increase as well. This is logical, as more pizzas mean a higher total cost but is still not relevant enough for future strategy.

Unit Price and Total Price of 0.836087 isa strong positive correlation. This implies that as the unit price of pizzas increases, the total price of the order also tends to increase significantly. This is intuitive, as higher-priced pizzas will naturally lead to a higher total cost, even if the quantity remains the same. However we can conclude that this dataset does not help in drawing relevant correlation insight.

To better understand these relationships, I created scatter plots for each pair of variables. The scatter plots would visually confirm the strength and direction of the correlations.

In conclusion, the correlation matrix provides valuable insights into the relationships between the variables. It helps us understand how changes in one variable might affect the others. This information can be useful for various purposes, such as forecasting sales, optimizing pricing strategies, and improving inventory management.

# Key Performance Indicators

| Total Revenue | €817,860 |
| --- | --- |
| Average Amount Spent per Order | €38.30 |
| Total Pizzas Sold | 49,574 |
| Average Pizzas per Order | 2.32 |
| Total Orders | 21,350 |

#### Highest and Lowest Revenue Generator Categories

| CATEGORY | HIGHEST | LOWEST |
| --- | --- | --- |
| Pizza | The Thai Chicken Pizza | The Brie Carre Pizza |
| Pizza Size | Large | XXL |
| Pizza Category | Classic | Veggie |

#### Most and Least Ordered Categories

| CATEGORY | MOST ORDERED | LEAST ORDERED |
| --- | --- | --- |
| Pizza | The Big Meat Pizza | The Brie Carre Pizza |
| Pizza Size | Small | XXL |
| Pizza Category | Classic | Chicken |

#### Seasonal Analysis Statistics

| TIME PERIOD | HIGHEST REVENUE | LOWEST REVENUE | MOST ORDERED | LEAST ORDERED |
| --- | --- | --- | --- | --- |
| Day of Week | Friday (Revenue - €136,073) | Sunday (Revenue - €99,203) | Friday (Orders - 3,538) | Sunday (Orders - 2624) |
| Month | July (Revenue - €72,557) | October (Revenue - €64,027) | July (Orders - 1,935) | October (Orders 1,646 |

#### Pizza Analysis Statistics

##### Pizza with Highest and Lowest Price

| CATEGORY | HIGHEST PRICE | LOWEST PRICE |
| --- | --- | --- |
| Pizza | The Pepperoni Pizza | The Greek Pizza |

##### Most Used Ingredients

| NAME | COUNT |
| --- | --- |
| Garlic | 27,422 |

# References

<https://www.kaggle.com/datasets/nextmillionaire/pizza-sales-dataset>, accessed 08/12/2024.

Faculty of Computing Level 8 HDIP Programmes 2024 - HDip in data Analytics for Business (HCI) - Sept Semester 1 StratThink\_HDipData\_Sept24\_HCI Week 2 - Data Analytics Life Cycle, Research, Citing and Referencing Presentation [Moodle course]. Retrieved from<https://moodle.cct.ie/course>

Home Faculty of Computing Level 8 HDIP Programmes 2024 - HDip in Data Analytics for Business (HCI) - Sept Semester 1 StratThink\_HDipData\_Sept24\_HCI, Class Notes, Retrieved from<https://moodle.cct.ie/course>

MüllerA.C. and Guido, S. (2017). Introduction to machine learning with Python : a guide for data scientists. Sebastopol, CA: O’Reilly Media, Inc., [Introduction to Machine Learning with Python](https://www.nrigroupindia.com/e-book/Introduction%20to%20Machine%20Learning%20with%20Python%20(%20PDFDrive.com%20)-min.pdf)

Correlation Matrix and Plot in Python - Exploratory Data Analysis - Hepatitis C Infection analysis, Bio-Resource, [Correlation Matrix and Plot in Python - Exploratory Data Analysis - Hepatitis C Infection analysis](https://www.youtube.com/watch?v=8Vq9MkTklCM&t=30s), YouTube video accessed 14/12/2024

Data Cleaning in Pandas | Python Pandas Tutorials, Alex The Analyst, YouTube video accessed 12/12/2024

Pizza Sales Analysis - Business/Data Understanding, Idris Alugo, [Pizza Sales Analysis - Data Extraction, Transformation and Loading](https://www.youtube.com/watch?v=JDQLv7KqqEg), YouTube video accessed 14/12/2024

Python and Pandas with Reuven Ler, [Using Markdown in Jupyter](https://www.youtube.com/watch?v=uVLzL5E-YBM), YouTube video accessed 11/12/2024

Python: univariate statistics, Mark Keith, [Python: univariate statistics](https://www.youtube.com/watch?v=78ut-S-QOEQ&list=PLe9UEU4oeAuV7RtCbL76hca5ELO_IELk4), YouTube video accessed 12/12/2024