# Data Analysis & Data Engineering Portfolio

## Mohamed Wahban

Data Analyst & Data Engineer

# About Me

Passionate about data and driven by impact, I specialize in both Data Analysis and Data Engineering. I've worked on multiple end-to-end projects that involved cleaning, transforming, and analyzing data to extract insights, as well as designing efficient data pipelines and automation workflows.

With strong command of tools like Python, SQL, Excel, and Power BI, I've built dashboards, implemented ETL processes, and developed solutions that help streamline decision-making and optimize performance.

I approach every project with a problem-solving mindset, attention to detail, and a focus on scalability and clarity. Continuously learning and improving, I'm committed to delivering high-quality, data-driven solutions that make a real difference.

# Education

- Bachelor’s Degree in Computer Engineering
- Faculty of Engineering, Alexandria University
- Focused on core areas including programming, data structures & algorithms, computer systems, and applied mathematics, with a growing specialization in data analysis and engineering through coursework and hands-on projects.

# Skills and Expertise

1. **Programming Languages:**
   Python, Java, C, C++, R, HTML

2. **Data & Database Tools:**
   SQL, Microsoft SQL Server, MySQL, Excel, Pandas, NumPy

3. **Data Engineering & Workflow Tools:**
   Linux , Git, GitHub, Bash scripting, ETL Concepts, Data Pipelines , LaTeX

4. **Data Visualization & Analysis:**
   Power BI, Matplotlib, Seaborn, Jupyter Notebook, Excel Charts & Pivot Tables

5. **Machine Learning & AI :**
   Supervised Learning, Neural Networks, Model Evaluation

6. **Cloud & Big Data:**
   Azure Data Tools , Big Data Concepts

# Work Experience

## Data Engineering Trainee – Microsoft Track

- Engaged in practical training on data engineering tools and workflows, including ETL pipelines, data modeling, and cloud-based data solutions.
- Built foundational experience in handling structured data, automating data flows, and working within cloud environments under Microsoft standards.
- Worked on real-life scenarios to design data solutions that support analytical use cases.

## Advanced Data Analysis Trainee – NTI

- Participating in an advanced-level training track focused on in-depth data analysis, including real-world applications in business intelligence and data storytelling.
- Working with tools like Python, SQL, Excel, and Power BI to build full-cycle analysis projects and generate professional insights.
- Collaborating in a structured learning environment with hands-on assignments and case studies aligned with market needs.

## Freelance Data Analyst (Project-Based)

- Developed multiple data analysis and reporting solutions using Excel, Power BI, SQL, and Python.
- Created interactive dashboards and automated reports to visualize performance metrics and trends.
- Applied data cleaning and EDA techniques to deliver actionable insights for real or sample datasets in education, business, and performance monitoring domains.

# Offered Services

I help individuals and businesses turn raw data into clear, actionable insights through practical, results-oriented data solutions. Here's what I can offer:

## Data Cleaning & Preparation

Preparing datasets for analysis by handling missing data, fixing inconsistencies, and organizing structure using Excel, SQL, and Python (Pandas).

## Exploratory Data Analysis (EDA)

Analyzing datasets to uncover patterns, trends, and anomalies using Python (Pandas, basic visualization) and Excel tools.

## Interactive Dashboards & Reports

Building professional dashboards and automated reports in Power BI and Excel to support business insights and decision-making.

## Data Querying & Manipulation

Writing efficient SQL queries to extract, filter, join, and summarize data from structured databases.

## Automating Tasks with Python

Using Python scripts to streamline repetitive tasks and basic reporting processes.

## Insight Generation & Recommendations

Delivering clear insights based on data analysis and providing recommendations to improve outcomes or optimize performance.

# Projects

## PREMIER LEAGUE PLAYER DATA ANALYSIS

Dataset contains detailed statistics for all Premier League players.
Total features: 57 columns covering offensive, defensive, and passing metrics. Each row represents a player.
Objective: Apply all 4 types of analytics to extract actionable insights
Microsoft Excel – Calculated Fields, Charts, Trendlines.
Data cleaning & feature selection applied to reduce dimensionality.
Filtered essential columns based on questions and use cases.
Visualizations created for enhanced interpretability

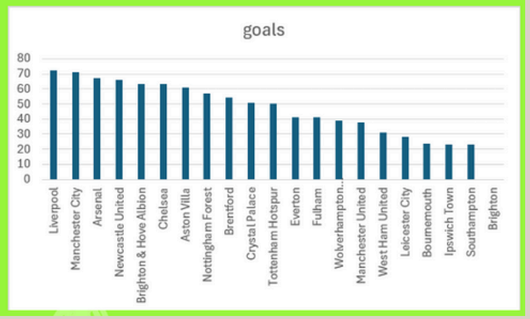
**The distribution of players by nationality**

This pie chart show that the most common nationality in league is England

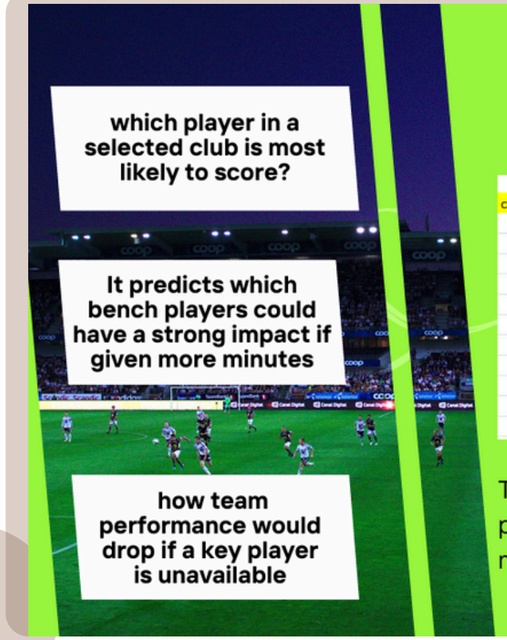

**The distribution of goal by position**

This pie chart show that the midfielders has more goals than forwards players


**Which club's players have scored the most goals?**




## DESCRIPTIVE INSIGHTS




## PREDICTIVE MODELING

| club | Liverpool | player | expected goals per 90 minutes | cumulative |
|---|---|---|---|---|
| | | Mohamed Salah | 0.77 | 0.77 |
| | | Cody Gakpo | 0.46 | 1.24 |
| | | Diogo Jota | 0.45 | 1.69 |
| | | Darwin Núñez | 0.40 | 2.09 |
| | | Harvey Elliott | 0.25 | 2.34 |
| | | Dominik Szoboszlai | 0.22 | 2.56 |
| | | Alexis Mac Allister | 0.17 | 2.73 |
| | | Curtis Jones | 0.16 | 2.89 |
| | | Trent Alexander-Arnold | 0.11 | 3.00 |
| | | Virgil van Dijk | 0.08 | 3.08 |
| | | Ibrahima Konaté | 0.04 | 3.12 |

This system will predict the number of goals expected for each player and so the number of goals expected per team in next matches by using the cumulative function

## SMART SCOUTING & PLAYER SELECTION SYSTEM

| | | | |
|---|---|---|---|
| | Position | GKP | |
| | Nationality | England | |
| | Club | Total | |
| Player Recommendation System | sort by | saves | |

| player | saves |
|---|---|
| Aaron Ramsdale | 125 |
| Jordan Pickford | 122 |
| Dean Henderson | 104 |
| Nick Pope | 88 |
| Fraser Forster | 29 |
| Alex McCarthy | 24 |
| Sam Johnstone | 23 |
| Christian Walton | 20 |
| Joe Lumley | 13 |
| Dan Bentley | 8 |

| | | | |
|---|---|---|---|
| | Position | MID | |
| | Nationality | Brazil | |
| | Club | Total | |
| Player Recommendation System | sort by | offensive contributions | |

| player | offensive contributions |
|---|---|
| Matheus Cunha | 21 |
| Gabriel Martinelli | 12 |
| Joelinton | 7 |
| Andreas Pereira | 6 |
| Bruno Guimarães | 5 |
| Lucas Paquetá | 4 |
| Matheus França | 1 |
| Casemiro | 1 |

**This enables coaches and scouts to make data-informed decisions when:**

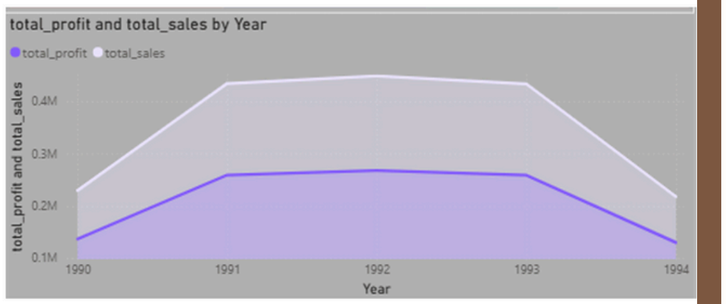
- **Identifying high-potential targets**
- **Filling squad gaps**
- **Replacing underperforming players**

# Projects

## Food Mart Retail Analysis Sales report

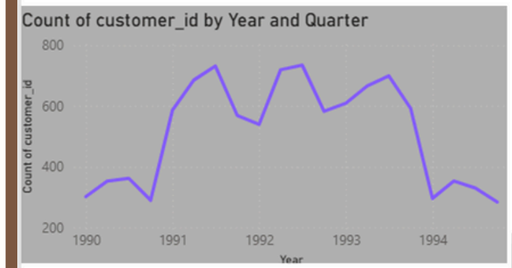
This project analyzes the Food Mart retail dataset, which includes customer demographics, sales, profits, and product returns. The objective is to extract actionable insights across: Customer Behavior Sales & Profit Trends Product Performance All analysis was done using Power BI, applying interactivity, storytelling, and clean visuals.

**food mart sales and profit over time (yearly and monthly)**

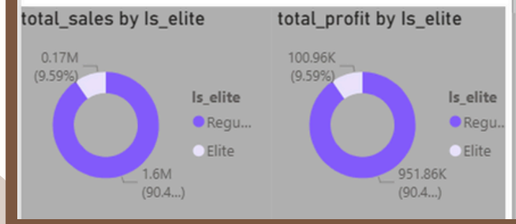



**we noticed that our sales are begging low and it raises for some year, and it becomes low again**

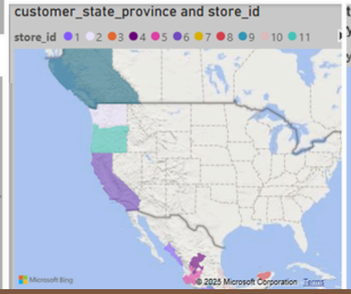
**The distribution of customer gained every year**

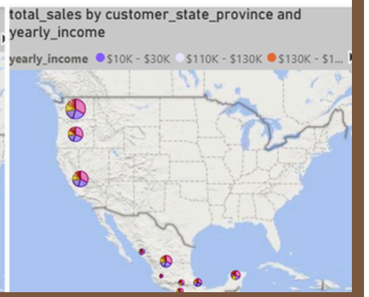



**The profitability of elite customers**




**The geographic distribution of customers and/ or Sales across different regions**




**key figures**

**Total Sales, Total Cost, Total Profit**

| 1.76M | 711.73K | 1.05M |
|---|---|---|
| total_sales | total_cost | total_profit |

**Profit Percentage**

147.92%
profit_percentage

**Total Returns**

8289
Total_Returns

**Number of Transactions**

270K
count_trans

**Brand Traded**

111
brand_traded


**This chart shows the top 10 Customers by total Purchases. it is clear that top customer called Ida Rodri purchased by 2.2k**

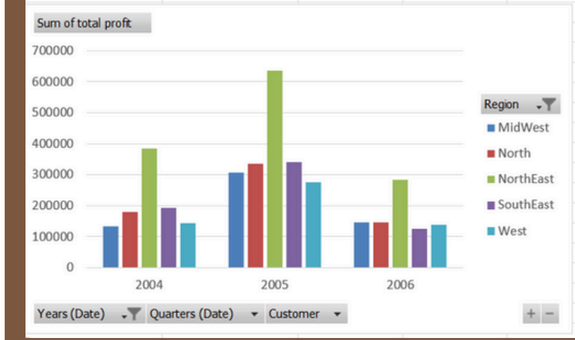
# Projects

## Sales Data Analysis – Amazon.com Products

In this analysis, we explored a dataset of over 2,000 sales transactions from Amazon.com, including detailed fields such as Region, Product, Sales Representative, COGS, Sales, Profit, and Date.
Using Pivot Tables and Excel Charts, we aimed to answer key business questions and extract valuable insights about performance across time, regions, and products.

### Profit per Region per Year

| Sum of total profit | Column Labels | | | | | |
|---|---|---|---|---|---|---|
| Row Labels | MidWest | North | NorthEast | SouthEast | West | Grand Total |
| 2004 | 133035.9949 | 179232.6414 | 385742.6644 | 192976.8509 | 144087.3935 | 1035075.545 |
| 2005 | 307311.1577 | 336691.4167 | 636655.7844 | 341629.0599 | 276176.948 | 1898464.367 |
| 2006 | 145531.1127 | 146350.8184 | 284780.5302 | 125226.8018 | 137432.2569 | 839321.5199 |
| Grand Total | 585878.2653 | 662274.8764 | 1307178.979 | 659832.7125 | 557696.5984 | 3772861.431 |




**Purpose:**

- To track the profitability trends of each region year by year.

**What it shows:**

- This table reveals how each region performed financiall across different years. It he identify growth trends, downturns, and regional consistency.

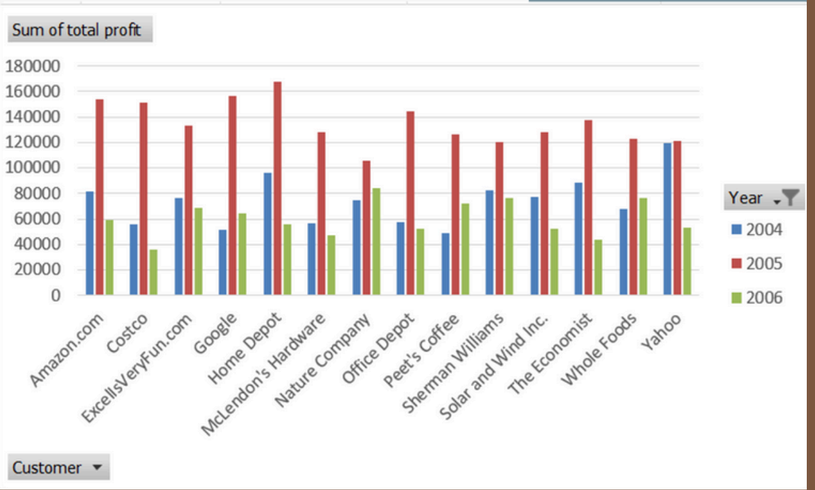
### Profit per Customer per Year

**Purpose:**

- To evaluate the value each customer brings yearly.

**What it shows:**

- This report helps us understand which customers consistently contribute to profit and which may be declining or peaking in certain years.




### Profit per Customer per region

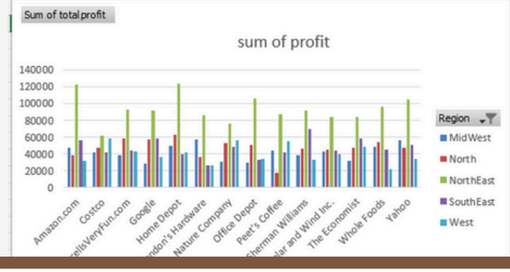
**Purpose:**

- To evaluate customer value from a regional perspective.

**What it shows:**

- Useful for identifying where each customer is most valuable, and whether they perform better in certain regions than others.

| Row Labels | MidWest | North | NorthEast | SouthEast | West | Grand |
|---|---|---|---|---|---|---|
| Amazon.com | 46969.03255 | 38527.3152 | 122180.522 | 56606.7567 | 31685.7906 | 29596 |
| Costco | 41553.90365 | 47066.697 | 61552.3041 | 42300.3125 | 58347.7275 | 25082 |
| ExcellsVeryFun.com | 38854.07275 | 58517.8887 | 93044.8246 | 44478.6786 | 43593.3216 | 27848 |
| Google | 28867.222 | 57451.5417 | 91786.6692 | 58254.4639 | 36234.455 | 27259 |
| Home Depot | 50117.8391 | 62792.6407 | 124328.136 | 39417.7594 | 42326.9401 | 31898 |
| McLendon's Hardware | 56913.73185 | 36315.742 | 85994.4403 | 26808.1519 | 26601.1746 | 23263 |
| Nature Company | 31310.93855 | 52998.8743 | 76597.5216 | 48122.9438 | 55804.2207 | 26483 |
| Office Depot | 29774.31635 | 50307.7159 | 106439.28 | 32698.8661 | 34327.9235 | 25354 |
| Peet's Coffee | 43939.8167 | 17929.8643 | 87794.0448 | 42368.1141 | 55275.4978 | 24730 |
| Sherman Williams | 38798.63455 | 46224.8521 | 91975.0524 | 69853.9372 | 32706.1657 | 27955 |
| Solar and Wind Inc. | 43184.4643 | 44774.1295 | 84336.2443 | 44464.3802 | 40239.1883 | 25699 |
| The Economist | 32080.00855 | 47493.8719 | 83661.9538 | 58219.6483 | 48517.709 | 26997 |
| Whole Foods | 48254.57135 | 54502.4909 | 96587.0515 | 45310.52 | 21534.3335 | 26618 |
| Yahoo | 56244.4906 | 47371.2527 | 104703.259 | 50928.18 | 34729.322 | 29397 |
| Grand Total | 586863.0429 | 662274.876 | 1310981.3 | 659832.713 | 561923.77 | 3781 |



# Achievement

Timmerman Industries

## DataCamp

- Cleaning Data with Generative AI
- Introduction to DAX in Power BI
- Introduction to SQL Server
- Introduction to Power BI
- Intermediate Python
- Introduction to Python
- Understanding Prompt Engineering

## Coursera

- Supervised Machine Learning: Regression and Classification
- What is Data Science?
- Open Source Software Development Methods

# Thank you for visiting, and I hope you find inspiration in my work!

## Mohamed Wahban

### Contact Information


- Alexandria, Egypt
- ( +20 ) 1010613314
- m.whban369@gmail.com
- Linkedin >> Mohamed Wahban
- GitHub >> HEBO-369