



DATA SCIENTIST

# PROPOSAL

Prepared For :

Epsilon Course

Made by: Ahmed Yasser

Address: 84 Gesr el Suez street



# Introduction

This project aims to conduct a comprehensive cleaning and analysis of the 2024 Data Science Sales dataset. Given that the dataset has been artificially altered to include common data issues, the project will serve as a demonstration of data preprocessing techniques and exploratory data analysis (EDA) to uncover insights within the sales domain of data science.

# OBJECTIVE

To perform detailed data cleaning to rectify artificially introduced issues, such as missing values, incorrect data types, outliers, and inconsistent entries.

To analyze the cleaned dataset to discover key trends, patterns, and correlations in the data science sales landscape for the year 2024.

To provide actionable insights that could be used to inform sales strategies in the data science industry.

# GOALS

## DATA CLEANING

Inspect the dataset for issues and anomalies.  
Perform cleaning steps including normalization,  
type conversion, deduplication, and outlier  
treatment.

Validate the integrity of the dataset post-  
cleaning.

## DATA ANALYSIS

- Descriptive statistics to summarize the data.
- Time-series analysis to understand sales trends over time.
- Segmentation analysis to compare sales across different categories such as regions, product types, and customer segments.
- Predictive analysis, if applicable, to forecast future sales trends.

## DATA VISUALIZATION

- Use of visual tools to represent findings, including bar charts, line graphs, heatmaps, and box plots.
- Creation of a dashboard summarizing the key metrics and trends for stakeholder review.

# CONCLUSION

- Use of visual tools to represent findings, including bar charts, line graphs, heatmaps, and box plots. Creation of a dashboard summarizing the key metrics and trends for stakeholder review.

**THANK YOU.**