Internship Data Engineering recruitment

# Introduction

Welcome to the final stage of our internship recruitment process! This final stage has two parts: a practical exercise, and an interview with questions about you, your experience, your academic background, etc.

The exercise gives you an opportunity to show how you would approach a real-world task. During the interview we would like you to present your solution of this exercise, explain what you've done, and show us anything of note regarding both the exercise and your solution.

There are no right or wrong answers, the aim here is to see your approach to solve the problem and for us to discuss the pros and cons during the interview. We will expect you to present for 10-15 minutes with questions along the way from us.

Do make sure you read through everything in this document before you get started.

# Scenario

The company Dunder Mifflin, situated in Spain, has requested our assistance in deriving conclusions from the dataset containing their performance metrics for the year 2023. We ask you to analyse the data, draw your own conclusions, and present your insights to them.

From your report, the company wants to understand where their revenue comes from, as well as their performance, both by department, location and date (see tasks below).

However, multiple project managers have been inputting their raw data into an csv file, which means there are data quality errors in the dataset that need to be fixed before analysis can start.

# Tasks

1. Clean the data
2. Transform the data (feel free to delete, add or change columns if appropriate).
3. Visualize the temporal evolution of sales (how sales progress in time).
4. Visualize sales by their location.
5. Visualize sales by product category.
6. Visualize any other insight you feel is interesting for the company
7. Prepare an explanation for your solution to the data quality issues and suggestions on how to improve them.
8. Prepare a presentation highlighting your insights from your analysis.

# Considerations

* You can cleanse and visualise the data with any tool you want: Excel, Python, R, ...
* You will have to present your result and explain the process you followed to get there. We expect you to justify your decisions and take advantage of this opportunity to showcase both your skills and knowledge as well as your communication skills.
* If you have any questions regarding the task, you can contact us at bcnacademy@fsp.co
* Take it easy, we are more interested in your explanation and how you justify your choices than in the result.

# Data Summary

|  |  |
| --- | --- |
| **Column Name** | **Column Description** |
| SalesId | Unique identifier for each sale transaction |
| SalesDate | Date of the sale |
| SalesChannel | Channel (Retail or Online) of the sale |
| SalesLocation | Location (city) of the sale |
| PostalCode | Postal code of the location of the sale |
| CustomerId | Unique identifier of the customer |
| CustomerName | Name of the customer |
| ProductId | Product identifier |
| ProductName | Name of the product |
| UnitPrice | Price of a single unit of the product |
| Quantity | Units of the product has the costumer purchased/returned |
| ProductCategory | Categorization of products |
| SalesRepresentativeId | Id of the salesperson |
| SalesStatus | Current status of the sale (Confirmed, Delivered, On-Site, Shipped) |