**MACQUARIE UNIVERSITY**

**Macquarie Business School**

--------🙞🙞🕮🙜🙜--------



**BUSA 8000 Techniques in Business Analytics**

**Group 9 Report**

**Authors:**

**Task 1 - Khuat Son Tra Nguyen - 48144134**

**Task 2 - Trung Quoc Nguyen -**

**Task 3 - Thi Bich Van Pham -**

**Task 4 - Duc Huy Tran – 48477192**

**Sydney, May 2024**

# INTRODUCTION

Table of contents

[INTRODUCTION 1](#_Toc167651070)

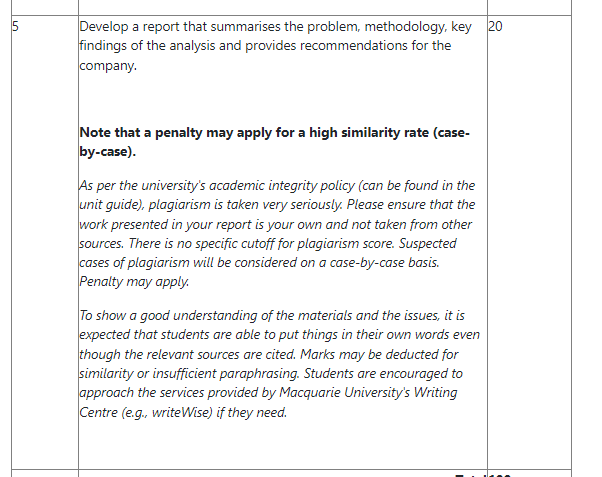
[Chapter 1: Data processing and finding 4](#_Toc167651071)

[1.1 Data cleaning and wrangling 4](#_Toc167651072)

[Chapter 2: Data Visualization 5](#_Toc167651073)

[Chapter 3: Predicting model and future sales 6](#_Toc167651074)

[Chapter 4: Summarization and Recommendation 7](#_Toc167651075)



## Chapter 1: Data Cleaning

## To provide informed recommendations for Dibs regarding their sales strategies, the initial and crucial stage is data preprocessing. The concept of “garbage in, garbage out” highlights that poor-quality data entry leads to unreliable data output (Kilkenny & Robinson, 2018). Moreover, effective sales strategies rely on accurate and timely sales forecasting, a process greatly enhanced by a well-prepared database (Pawar, 2023).

## **Objective:** To prepare for the subsequent stages, we aim to *develop a comprehensive monthly sales database* from raw sales data files for the years 2019, 2020, and 2021.

## The database will be meticulously cleaned to ensure it is error-free, consistent, and complete. The process involves a detailed examination of the raw data for each variable (as summarized below) to understand its characteristics. Subsequently, we will identify and address any issues associated with each variable. The table below summarizes the identified issues and the solutions implemented to resolve them for each variable in the dataset.

|  |  |  |
| --- | --- | --- |
| **Column** | **Issues** | **Implementations (in the following order)** |
| Order ID | None | None |
| Product | Incorrect Data Entries Noises, Missing Values | Replace with imputed values Remove noises and missing values |
| Quantity Ordered | Values = 0Formatting Issues | Remove any null valuesChange formatting to integer |
| Price Each | Incorrect Data EntriesFormatting Issues | Replace with imputed valuesChange formatting to integer |
| Order Date | Incorrect Data Entries Not Ideal for Data Manipulation  Noises, Missing Values | Replace with imputed values Create 5 new columns: ‘Day’, ‘Month’, ‘Year’, ‘Hour’ and ‘Minute’ Remove noise and missing values (if any) |
| Purchase Address | Not Ideal for Data ManipulationIncorrect Data Entries | Create 2 new columns: ‘City’ and ‘State’Replace with imputed values |

Finally, we will export a cleaned CSV file, "sales\_clean.csv," for other team members to proceed with the subsequent stages.

## Chapter 2: Data Visualization

## Chapter 3: Predicting model and future sales

## Chapter 4: Summarization and Recommendation