



United International University
Fall 2025 Trimester
Data Analytics Lab : Assignment 1
Topic: Data Preprocessing and EDA, Total Marks: 15

For the dataset given in the drive link, clean the dataset if found any missing data, outliers, duplicate data, formatting, inconsistency etc. After that find out the answers to the following EDA questions using python code in notebook:

1. Which category generates the highest total sales?
2. Within each category, which sub-category is the most profitable?
3. Which region has the highest average profit per order?
4. What percentage of total sales comes from each segment?
5. Identify the top 10 most profitable products.
6. Which ship mode has the fastest average delivery time?
7. Which city generates the highest total profit? Give the top 5.
8. For each region, what is the average discount offered, and how does it correlate with profit?
9. Which product category suffers the highest loss (negative profit)?
10. Is there a relationship between discount and quantity sold?
11. What are the top 5 customers by total purchase amount?
12. Which months have the highest number of orders? Show in a bar chart.
13. Compare profit margins between first-class and same-day shipping.
14. Which sub-categories are consistently unprofitable across regions?
15. Compute the year-over-year growth of total sales and show the plot.

Submission Guidelines: Complete the code in a notebook and run it so that the notebook contains the outputs. Then submit that notebook in the elms.

During the evaluation, you will be taken a viva. Prepare accordingly.

Submission Deadline: 11:55 pm, Saturday, 6th December 2025.

Note: Please DO NOT COPY solutions from anywhere (your friends, seniors, internet, etc.). Any form of plagiarism (irrespective of source or destination), will result in getting-100% marks in this assignment. You have to protect your code.