Practice Exercise: EDA With Python

The following is a post-class exercise for practicing exploratory data analysis using Python.

Note: This is neither a graded assessment nor has any time restraints for completion.

Case Study Number & Title	4. Exploratory analysis of US salary data
Background Information	The dataset consists of variables that can be used to predict the salary of an individual in USA.
Problem Statement/ Business objectives	Understand the impact of different variables on the salary of an individual.
Data, Information for case analysis	Data is provided as a csv file. Below is the source and attribute information. Source link: https://www.kaggle.com/datasets/saumitgp/adult-income-prediction Data Description **Age:** Age of the individual **Workclass:** Type of job, if employed **Education**: Educational qualifications **Marital status:** Married or unmarried or divorced **Occupation**: Occupational information **Relationship:** Family status **Race**: Race of the individual **Sex**: Gender of the individual **Hours-per week:** The number of hours worked by the person **Native country:** Country to which the person belongs to **Income:** Margin of income an individual earns
Questions	1. Carry out missing value analysis on “Workclass” variable and treat them (if any) using appropriate methods. 2. Visualize the income-wise average working hours on the grounds of race.

	3. Display the income and education wise count of the individuals. 4. How is marital status and income group correlated? Which relationship status has highest number of >50k salaried individuals? 5. Does race have any impact on the working hours put in by the people?
Solution	A sample solution also provided with the dataset
Deliverables for Solution and Rubric	Non-graded assessment
Key Takeaways/Results	Exploring and analyzing data using Python and deriving meaningful insights.