

## 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	1. Analyzing medical costs data for drawing inferences on charges incurred
Background Information	The dataset contains records of clients of an insurance provider, detailing out the medical charges incurred for the treatment(s) availed by that individual.
Problem Statement/ Business objectives	Analyze the medical costs data to discern the insurance company's operations and explore using Python to extract relevant insights.
Data, Information for case analysis	<p>Data is provided as a csv file. Below is the source and attribute information.</p> <p>Source link:  <a href="https://www.kaggle.com/datasets/mirichoi0218/insurance?resource=download">https://www.kaggle.com/datasets/mirichoi0218/insurance?resource=download</a></p> <p><u>Data Description</u></p> <p><b>age:</b> Age of primary beneficiary</p> <p><b>sex:</b> Insurance contractor gender; female, male</p> <p><b>bmi:</b> Body mass index providing an understanding of body; weights that are relatively high or low relative to height; objective index of body weight (kg/m<sup>2</sup>) using the ratio of height to weight, ideally 18.5 to 24.9</p> <p><b>children:</b> Number of children covered by health insurance / number of dependents</p> <p><b>smoker:</b> Whether the individual has smoking habit or not</p> <p><b>region:</b> The beneficiary's residential area in the US; northeast, southeast, southwest, northwest</p> <p><b>charges:</b> Individual medical costs billed by health insurance</p>
Questions	<ol style="list-style-type: none"> <li>Find out the region-wise average BMI for each gender.</li> <li>Create a "weight" column wherein the values based on BMI would be as follows: <ul style="list-style-type: none"> <li>&lt;18.5 – Underweight</li> <li>&gt;25 – Overweight</li> <li>Otherwise, Healthy weight</li> </ul> </li> </ol>

	<p>3. Which region has incurred the highest average charges?</p> <p>4. Visualize region wise average charges on the basis of smoking habit using suitable chart.</p> <p>5. Create age bins as shown and infer about the BMI based on it.</p> <ul style="list-style-type: none"> <li>• 18-29 – Young adult</li> <li>• 30-59 – Adult</li> <li>• 60 &amp; above – Senior</li> </ul>
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.