Project - Applied Statistics

**Submission type**

:

Online url,online upload

**Due Date**

:

Jul 26, 11:00 PM

**Total Score**

:

60

**Available from**

:

Jul 10, 8:00 AM

**Description**

Dear Participant,

Please find below the details for Project. Kindly submit it before the deadline.

**Project Description -**[**Problem Statement - Statistical Learning.pdf**](https://olympus.greatlearning.in/courses/13187/files/916035/download?verifier=PPUog9OngMAGItvQSUQMLoPDZaULKQBqHmAUB0Al&wrap=1)

**Data** - [insurance.csv](https://olympus.greatlearning.in/courses/13187/files/916074/download?verifier=PrwQgP6mhBzjAT6qUdkGFHJyyTyr5W6DCqyAbGW7&wrap=1)

Apply the concepts and techniques you have learnt in the previous weeks and summarise your insights at the end.

You need to submit a python notebook in both **".ipynb and HTML"** format with proper comments.

Regards

Program Office

**Scoring guide (Rubric) - Applied Statistics Rubric**

| **Criteria** | **Points** |
| --- | --- |
| Import the necessary libraries | 2 |
| Read the data as a data frame | 2 |
| EDA - shape & type of Data | 4 |
| EDA - Checking the presence of missing values | 3 |
| EDA - 5 point summary of numerical attributes | 3 |
| EDA - Distribution of ‘bmi’, ‘age’ and ‘charges’ columns | 4 |
| EDA - Measure of skewness of ‘bmi’, ‘age’ and ‘charges’ columns | 2 |
| EDA - Checking the presence of outliers in ‘bmi’, ‘age’ and ‘charges columns | 4 |
| EDA - Distribution of categorical columns | 4 |
| EDA - Pair plot that includes all the columns of the data frame | 4 |
| Do charges of people who smoke differ significantly from the people who don't? | 7 |
| Does bmi of males differ significantly from that of females | 7 |
| Is the proportion of smokers significantly different in different genders | 7 |
| Is the distribution of bmi across women with no children, one child and two children, the same | 7 |
| Points | 60 |

**Submit Assignment**