

Data Analytics with Python

Mini-Project Guidelines

VALUE ADDED COURSE

Dear students, consider the document as guidelines for completing the mini-project. This is considered as the final work for our VAC. Marks will be awarded based on the Genuity of the content and the quality of it.

GUIDELINES:

- Problem Statements for the mini-project is provided after the guidelines.
- This is a group task.
- Choose one problem statement.
- Explanation of the project for minimum of 15 lines is expected.
- Explanation is to be done in the markdown file in GitHub.
- Contents for the markdown file,
 - Introduction about the problem statement & data.
 - Explanation of the process and things carried out.
 - Inference.
 - Performance Metrics Explanation.
- Nomenclature for file names,
 - Mini-Project – **“Regression – Prediction of <Name of problem statement/data>”**.
- **Post** your works in your **GitHub profile** under the folder “Data Analytics” as separate sub-folder.
- Name the sub folder as **“Regression”** and in it upload the following files.
 - Python File/Notebook
 - Dashboard of the data in pdf
 - Explanation of the project along with group members name in the md file
- **One week** is provided to complete this mini-project. **Last date** will be **24th March 2022**.
- If required from students a one hour connect through GMeet can be arranged between 18th - 23rd March 2022 for clarification of doubts. Discussion to be requested & done through WhatsApp group.
- For any doubts or clarification, you can contact Pradish & Sandeep from Prag Robotics and also you can use the whatsapp group.

PROBLEM STATEMENTS:

STATEMENT 1: BODY FAT PREDICTION

Estimate the body fat percentage from about 250 men. Perform the following things,

- Perform descriptive analysis using statistics
- Dashboard using the data
- Prediction of body fat with performance metrics

DATASET ATTRIBUTES:

1. Density determined from underwater weighing
2. Percent body fat from Siri's (1956) equation
3. Age (years)
4. Weight (lbs)
5. Height (inches)
6. Neck circumference (cm)
7. Chest circumference (cm)
8. Abdomen 2 circumference (cm)
9. Hip circumference (cm)
10. Thigh circumference (cm)
11. Knee circumference (cm)
12. Ankle circumference (cm)
13. Biceps (extended) circumference (cm)
14. Forearm circumference (cm)
15. Wrist circumference (cm)

STATEMENT 2: DIAMOND PRICE

Predict the price of diamond from the attributes provided in the dataset. Perform the following things,

- Perform descriptive analysis using statistics
- Dashboard using the data
- Prediction of price with performance metrics

DATASET ATTRIBUTES:

1. price - price in US dollars (\\$326--\\$18,823)
2. carat - weight of the diamond (0.2--5.01)

3. cut - quality of the cut (Fair, Good, Very Good, Premium, Ideal)
4. colour - diamond colour, from J (worst) to D (best)
5. clarity - a measurement of how clear the diamond is (I1 (worst), SI2, SI1, VS2, VS1, VVS2, VVS1, IF (best))
6. x - length in mm (0--10.74)
7. y - width in mm (0--58.9)
8. z - depth in mm (0--31.8)
9. depth - total depth percentage = $z / \text{mean}(x, y) = 2 * z / (x + y)$ (43--79)
10. table - width of top of diamond relative to widest point (43--95)

STATEMENT 3: QUALITY OF RED WINE

Predict the quality of red wine based on the physicochemical tests done on it. Perform the following things,

- Perform descriptive analysis using statistics
- Dashboard using the data
- Prediction of quality with performance metrics

DATASET ATTRIBUTES:

1. fixed acidity
2. volatile acidity
3. citric acid
4. residual sugar
5. chlorides
6. free sulphur dioxide
7. total sulphur dioxide
8. density
9. pH
10. sulphates
11. alcohol
12. Output variable (based on sensory data): quality (score between 0 and 10)

All the datasets are provided in the following link,

LINK:

https://drive.google.com/drive/folders/1L2_twpEOSMB5L2RQ0FI51JojL8Y0Nhh8?usp=sharing

Relevant file names are provided for the datasets according to the problem statements, so choose the appropriate one to work.