**JAIPURIA INSTITUTE OF MANAGEMENT, INDORE**

**PGDM**

**FOURTH TRIMESTER (Batch 2021-23)**

**END TERM EXAMINATION, NOV-2022**

|  |  |  |  |
| --- | --- | --- | --- |
| Course Name | Python for Business Analytics | Course Code | 40823 |
| Max. Time | **2 hours** | Max. Marks | **40** |

**INSTRUCTIONS:**

* *Please write the analysis in the jupyter notebook using multiline comments.*
* *Submit on one jupyter notebook file, with proper heading of each question*
* *For Question 4( part a and b) write the interferences in the jupyter notebook mentioning the question number.*

**Questions.1** The dataset contains the retrospective sample of males in a heart -disease high risk region of the western Cape, South Africa. There are roughly two controls per case of coronary heart disease (CHD). Many of the HD-positive men have undergone blood pressure reduction treatment and other programs to reduce their risk factors after their CHD event. (Use SASheart Dataset) **(15 Marks)**

sbp: Systolic blood pressure

tobacco: Commutative tobacco (kg)

idl: Low density lipoprotein cholesterol

adiposity

famhist- family history of heart disease (Present, Absent)

typea- Type A behavior

obesity

alcohol- Current alcohol consumption

age- Age at onset

chd- Response, coronary heart disease

1. List down the records present in the dataset? Explain the metadata information of the given dataset.
2. Explain with the help of suitable plot the number of the persons having CHD or not in comparison to they having family history of disease or not, if organisation would like to introduce some medicine for the good heart health, comment on their advertising strategy based on the above data.
3. Comment on correlation of age with sbp? Choose appropriate plot to show the relation, how App based workout organisation focus on age parameter to design some exercise modules.
4. Compare the distribution of tobacco consumption for persons having CHD and not having CHD. Can you interpret effect of tobacco consumption on having coronary heart disease?
5. Derive the new column called agegroup from age column where persons falling in different age ranges are categorized as below:

(0-15): young

(15-35): adults

(35-55): mid

(55-): old

 f) Compare distribution of *Idl* for different age groups. Comment

**Questions.2** A glints.com digital marketing company claims that a firm’s new customer acquisition increased by utilizing various digital marketing tools, as compared to firms that are using only conventional marketing strategies. In order to check the reliability, the Absolute Marketing Research firm undergoes a test to determine whether the increase in new customer acquisition with digital marketing intervention is different or not. Analyze your result using an appropriate plot in order to justify your result. **(10 Marks)**

**Questions.3** a)In order to attract more sales in terms of amount, True Retail Corporation designed the strategy. Please help them to design the code which helps the marketing people to devise the discount strategy on the basis of the amount of purchase. If the purchase amount is more than Rs. 10,000, they will get a 30% discount. If it is more than 8,000, they are eligible for a 20% discount. If they have spent around 5000, they will get a 10% discount. If the customer spends less than this amount, the customer is eligible for a thankyou note. **(5 Marks)**

b) A fabric manufacture would like to understand the proportion of defective fabrics they produce. The shop floor staff members have been stating that the percentage of defective is not more than 18%. They would like to test whether the claim made by their shop floor staff is correct. They picked up a random sample of 20 fabrics and found 16 defectives. Write a code to test the hypothesis that the percentage of defective components is less than 18%.

Sample of 20 [ 16.0, 16.0, 30.0, 37.0, 25.0, 19.0, 35.0, 27.0, 32.0, 34.0, 28.0, 24.0, 35.0, 24.0, 21.0, 32.0, 29.0, 24.0, 35.0, 28.0] **(5 Marks)**

**Questions.4 a)** The below mention histogram explains the sold price of a commodity, comment on the sold price. **(2.5 Marks)**



b) The below scatter plot chart is customer purchasing in terms of number of items (Horizontal axis) and price in (vertical axis). Comment on customer purchasing decisions **(2.5 Marks)**

