|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| A close up of a sign  Description automatically generated | | | | | | | |
| Data Science Academy Bootcamp | | | | | | | |
| Final exam | | | | | | | |
| **Name,Surname:** | Shahin Mammadov | | |  | | Date: | 11/19/2022 |
| **Exam:** Data Science with SPSS | |  |  | | | | |
| Instructions: You will be given ***1 hour and 45 minutes*** to complete the examination. There are 2 parts and 3 types of questions including: True/False, Multiple choice and Practical.  **First part** will last ***45 minutes****,* and you must write answers directly on word document.  **Second part** is Practical question and will last ***60 minutes***. It is your responsibility to submit your exam answers electronically to [homework@dsa.az](mailto:homework@dsa.az) with **“SPSS Final Exam”** subject.  Before you leave examination, you **MUST** contact the mentor to confirm that you have submitted your examination.  Good Luck! | | | | | | | |
| PART 1: TRUE/FALSE | | | | | | | |
| Directions: Read each statement below carefully. Place a Checkmark on the “True” cell if you think a statement is TRUE, and on the “False” cell if you think the statement is FALSE. | | | | | | | |
|  | | | | | | | |
|  | | | | | True | False | |
| 1. In Diagnostic analytics, if the p-value >0.5, it means there is a significant relationship between variables. (3 points) | | | | |  | X | |
| 2. Spearman correlation evaluates the linear relationship between two ordinal variables. (1 point) | | | | |  | X | |
| 3. The possible responses to the question “how many shoes do you own” is continuous variable (2 points) | | | | |  | X | |
| 4. Skewness is 0.75 then this data is approximately normally distributed (2 points) | | | | |  | X | |
| 5. Antecedents are one or more IF choices. (3 points) | | | | | X |  | |
| 6. False negative is when actual class is yes but predicted class is no. (3 points) | | | | | X |  | |
| 7. Chi square correlation is used when the 2 variables are categorical (2 points) | | | | | X |  | |
| 8. K means node can detect optimal number of clusters while creating clustering model (3 points) | | | | | X |  | |
| 9. Type node must always be connected to data before any modeling process (1 point) | | | | | X |  | |
| 10. Field ops palette contains nodes that performs operations on columns of the data (2 points) | | | | | X |  | |

|  |
| --- |
| PART 1: MULTIPLE CHOICE |
| Directions: Read each question carefully and place a + next to the answer that best fits the question. |
|  |
| 01. Which two functions are used in the Select and Filler nodes to identify missing values? (2 points)  a) @BLANK @NULL b) @PREDICTED @INDEX c) @NULL @WHITE SPACE  02. You want to obtain a subset of data from a larger data set, with equally represented subgroups within the subset.  Which node would you use to accomplish this task? (1 point)  a) Analysis node b) Partition node c) Sample node  03. Which node is used to read data from a comma delimited text file? (1 point)  a) Statistics File b) Var. File c) Fixed File  04. You want to illustrate the strength of the relationship between blood pressure and prescription drug to which a patient responded. The blood pressure variable has been binned into ten values. Which node should be used? (2 points)  a) Analysis node b) Time plot node c) Web node  05. Given that the heights of male population is approximately normally distributed with a mean of 170 cm and standard deviation of 5 cm, the proportion of males with greater height of 185 cm is …… (3 points)  a) 2.5 % b) 0.3 % c) 0.15 %  06. An e-retailer conducting a data mining project has limited an initial study to approximately 30,000 customers who have registered on the site. There are still millions of records in the Web logs. The data miner wants to determine the frequency distribution of the age of their customers. The Age column is a continuous field. Which node would you use to accomplish this task?  (3 points)  a) Histogram node b) Data audit node c) Sample node  07. You want to create a Filter node to keep only a subset of the variables used in model building, based on predictor importance. Which menu in the model nugget browser provides this functionality? (1 point)  a) View b) Preview c) File  08. You have a data set with two different date fields (columns). Your analysis requires you to create a new field which is the number of calendar days elapsed between these two dates. Which node will you use to accomplish this task? (2 points)  a) Filter node b) Type node c) Derive node  09. You have a large amount of data from which you want to build a model. Although many of the records of data are complete, there are substantial amounts of records which contain missing data. The records containing incomplete information should be excluded from analysis. Which node will exclude the undesired records? (2 points)  a) Filter node b) Filler node c) Select node  10. Which palette contains the Data Audit node? (1 point)  a) Output b) Source c) Export |
|  |

www.dsa.az Bütün hüquqlar qorunur DATA SCIENCE ACADEMY 2