



UNIVERSITAS ISLAM INDONESIA
FAKULTAS MIPA
Jl. Kaliurang Km. 14,5 Jogjakarta

MIDTERM EXAM ODD ACADEMIC YEAR 2020/2021

Course	: Data Intelligence
Study Program/Class	: Statistics/ A and B
Date	: Tuesday, 10 November 2020
Time	: 100:00 AM
Periode	: 100 minutes
Lecture	: Arum Handini Primandari, M.Sc.
Method	: Open Notebook/laptop

CPL : PPa (*Intelligence*) Mastering the concept of probability theory and statistics, mathematics, calculus, elementary linear algebra, statistical analysis methods and elementary computer programming

Problem in Short Answer (1-6) and True/False Question (1-5)
CO PPa2 Students are able to compile programming codes for theory of probability and statistics
Indicator: Students use python structure
Problem in Essay (1-3)
CO PPa3 Students are able to compose intelligent computer programming or data analysis
Indicator: Students employ python library

Instructions
<ol style="list-style-type: none">1. Login into browser using your UII student account2. Open Google Classroom Data Intelligent in a post of "<i>Ujian Tengah Semester</i>" (Midterm Exam).3. Solve the problems related to HRDataset stated in Google Form with python language.4. You will be asked to upload the python code (written in the Jupyter Notebook) that you used into given field in the google form.5. The given time includes the time for solving problems and submitting answer. The form will close automatically when the time is up.

A Glimpse of the Data

	Employee Name	Employee Number	MarriedID	MaritalStatusID	GenderID	EmpStatus_ID	DeptID	Perf_ScoreID	Age	Pay Rate	...	Date of Hire	Days Employed	Date of Termination	Reason For Termination
0	Brown, Mia	1103024456	1	1	0	1	1	3	30	28.50	...	2008-10-27	3317	NaT	N/A - still employec
1	LaRotonda, William	1106026572	0	2	1	1	1	3	34	23.00	...	2014-01-06	1420	NaT	N/A - still employec
2	Steans, Tyrone	1302053333	0	0	1	1	1	3	31	29.00	...	2014-09-29	1154	NaT	N/A - still employec
3	Howard, Estelle	1211050782	1	1	0	1	1	9	32	21.50	...	2015-02-16	58	2015-04-15	N/A - still employec
4	Singh, Nan	1307059817	0	0	0	1	1	9	30	16.56	...	2015-05-01	940	NaT	N/A - still employec
5	Smith, Leigh Ann	711007713	1	1	0	5	1	3	30	20.50	...	2011-09-26	730	2013-09-25	career change
6	LeBlanc, Brandon R	1102024115	1	1	1	1	1	3	33	55.00	...	2016-01-05	691	NaT	N/A - still employec
7	Quinn, Sean	1206043417	1	1	1	5	1	3	33	55.00	...	2011-02-21	1636	2015-08-15	career change
8	Boutwell, Bonalyn	1307060188	1	1	0	1	1	0	31	34.95	...	2015-02-16	1014	NaT	N/A - still employec

Problems

A. Short answer

- How many employees are registered in the HR dataset? (poin 5)
- How to display data for the first 100 employees which contain information of name, ID, age, and the average of salaries? (Poin 5)
- How to select column "Employment Status"? Write down all the possible code. (Poin 10)
- How to filter the active employees who have been working for at least 1,000 days? (Poin 5)
- Regex (Regular Expression) can be used as a wild card to find character contained in the data. The code to find the employee whose family names starting with "Trump" is `hrdata[hrdata["Employee Name"].str.contains(r'Trump (?! $) ')]`. By using regex, how many people are there in the department with the prefix "Admin"? (Poin 8)
- The function `groupby` is employed to group the data based on categories in a specific column. Example: `hrdata.groupby('Position').count()` will calculate the number of other variables based on the position grouping. By employing the group function, which department has the most employees? (Poin 8)

B. True/False Question

- There are exactly 25 unique positions in the HR dataset. (T/F) (Poin 5)
- The position that has the most employees in the HR dataset is IT. (T/F) (Poin 5)

3. There are 181 employees whose performance scores have been fully meet. (T/F) (Poin 5)
4. There is exactly one employee whose family name is the same as the US presidential candidate 2021. (T/F) (Poin 5)
5. There are more than 150 active employees who have worked for at least 1000 days. (T/F) (Poin 5)

C. Essay

1. How to get an array from the multiplication between column Age and Perf_ScoreID? Please write down the steps and the result. (Poin 10)
2. What information or data insight that you get from the HR dataset? Please write it in 3-5 sentences. (Poin 15)
3. Please provide the python code that you used to solve the whole problems in this midterm exam in html/pdf form by uploading it into the given field in google form. (Poin 5)

The verification of Midterm Exam, Odd Academic Year 2020/2021				
The suitability of the problems with CO	The Completeness of the problems information	Verifier	Verification date	Sign
Suitable/ Less suitable / Not Suitable	Complete/ Less Complete / Not Complete	Science cluster coordinator	5-Nov-2020	