

Lathan Persiapan Exam Modul 2

1. Jika kode di bawah ini dijalankan

```
np.random.randint(0, 50, 6).reshape(3, 2)
```

Output yang benar adalah ...

- a. `array([[0.11272585, 0.50231826], [0.95693598, 0.83532571], [0.24816881, 0.92556657]])`
 - b. `array([[0.11272585, 0.50231826, 0.95693598], [0.83532571, 0.24816881, 0.92556657]])`
 - c. `array([[29, 22], [12, 35], [23, 21]])`
 - d. `array([[29, 22, 12], [35, 23, 21]])`
2. `myArray = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])`
`print(myArray * np.array([[10],[10],[10]]))`

Output dari adalah ...

- a. `array([[10, 20, 30], [40, 50, 60], [70, 80, 90]])`
 - b. `array([[10, 20, 30], [4, 5, 6], [7, 8, 9]])`
 - c. `array([[10, 2, 3], [40, 5, 6], [70, 8, 9]])`
 - d. error
3. Diketahui 2 dataframe bernama df_X dan df_Y seperti di bawah ini.

df_X			
	a	b	c
0	1	6	11
1	2	7	12
2	3	8	13
3	4	9	14
4	5	10	15

df_Y		
	a	z
0	1	100
1	2	200
2	3	300
3	4	400

Manakah cara penggabungan dataframe df_X dan df_Y yang tepat untuk menghasilkan output:

	a	b	c	z
0	1	6	11	100
1	2	7	12	200
2	3	8	13	300
3	4	9	14	400

- `df_X.merge(df_Y)`
- `df_X.merge(df_Y, how = 'left')`
- `df_X.merge(df_Y, how = 'outer')`
- `df_X.join(df_Y)`

4. Dataframe df di bawah ini berisikan daftar nama mobil beserta spesifikasinya. Berikut ini adalah 5 baris pertamanya:

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130.0	3504	12.0	70	usa	chevrolet chevelle malibu
1	15.0	8	350.0	165.0	3693	11.5	70	usa	buick skylark 320
2	18.0	8	318.0	150.0	3436	11.0	70	usa	plymouth satellite
3	16.0	8	304.0	150.0	3433	12.0	70	usa	amc rebel sst
4	17.0	8	302.0	140.0	3449	10.5	70	usa	ford torino

Untuk menampilkan nama mobil yang memiliki horsepower paling besar, manakah kode yang benar?

- `df[df['horsepower'] == df['horsepower'].max()].loc['name']`
- `df['name'][df['horsepower'] == df['horsepower'].max()]`
- `df.iloc[df['horsepower'] == df['horsepower'].max()]['name']`
- `df.loc[df['horsepower'].max()]['name']`

5. Manakah pernyataan yang benar mengenai populasi dan sampel?

- Parameter adalah ringkasan numerik dari populasi
- Populasi adalah bagian dari sampel
- Pengambilan sampel tidak boleh acak
- Pemanfaatan sampel untuk menarik kesimpulan terhadap populasi dinamakan Statistika Deskriptif

6. `myNumber = [2,2,3,4,2,7,4,8,9,3,5]`

Manakah nilai deskriptif statistik yang tepat dari variable `myNumber`?

- Mean = 4.2
- Median = 7

- c. Standard Deviation = 2.78
- d. Q3 = 6

7. Manakah pernyataan yang tepat mengenai uji korelasi?

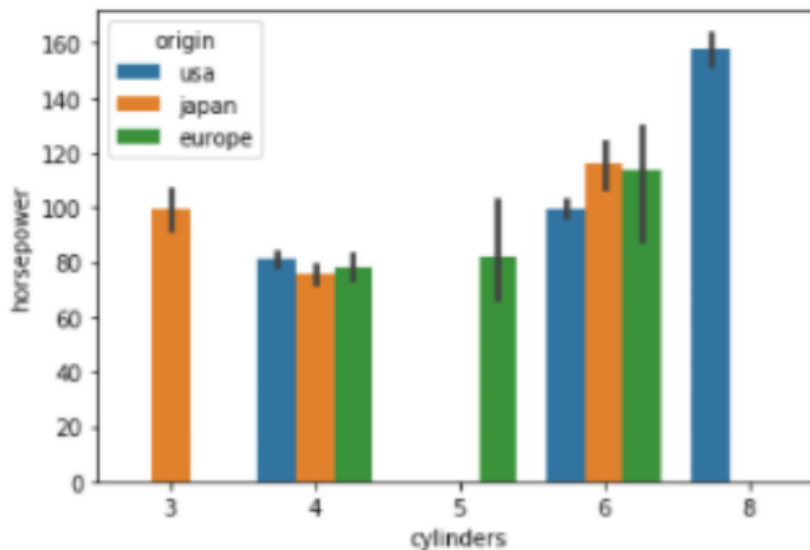
- a. Uji korelasi Spearman dapat digunakan untuk data numerik yang tidak terdistribusi normal
- b. Nilai korelasi positif berarti hubungan antara variabel yang diuji sangat kuat
- c. Uji korelasi Pearson dapat digunakan untuk 2 variabel kategorikal nominal
- d. Uji korelasi Spearman dapat digunakan untuk 2 variabel kategorikal nominal

8. Pilihlah visualisasi yang paling tepat!

	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
0	18.0	8	307.0	130.0	3504	12.0	70	usa	chevrolet chevelle malibu
1	15.0	8	350.0	165.0	3693	11.5	70	usa	buick skylark 320
2	18.0	8	318.0	150.0	3436	11.0	70	usa	plymouth satellite
3	16.0	8	304.0	150.0	3433	12.0	70	usa	amc rebel sst
4	17.0	8	302.0	140.0	3449	10.5	70	usa	ford torino

- a. Scatter Plot untuk variabel origin vs horsepower
- b. Bar Plot untuk variabel origin vs weight
- c. Box Plot untuk variabel name
- d. Pie plot untuk variabel displacement vs weight

9. Kode yang tepat untuk menampilkan visualisasi di bawah ini adalah ...

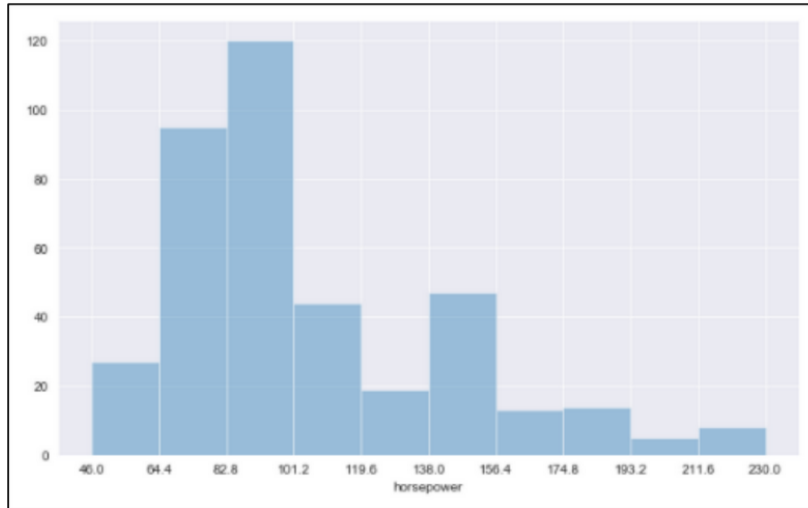


- a. `plt.plot(data=df, x='cylinders', y='horsepower', hue='origin')`
- b. `sns.countplot(data=df, x='cylinders', y='horsepower', hue='origin')`
- c. `sns.barplot(data=df, x=origin, y='horsepower',`

```
hue='cylinders')
```

```
d. sns.barplot(data=df, x='cylinders', y='horsepower',  
hue='origin')
```

10. Pernyataan yang benar mengenai visualisasi di bawah, kecuali ...



- a. Variabel Horsepower tidak terdistribusi normal
- b. Sumbu y (vertikal)menunjukkan frekuensi atau jumlah kemunculan data
- c. Mobil dengan horse power pada interval 82.8 - 101.2 yang muncul paling banyak
- d. Grafik di atas biasa disebut dengan count plot

11. Di bawah ini adalah hasil uji normalitas dari variabel 'weight':

```
from scipy.stats import normaltest
```

```
normaltest(df['weight'])
```

```
NormaltestResult(statistic=46.08618081354447,
```

```
pvalue=9.829082726891064e-11)
```

Pilihlah pernyataan yang tepat mengenai uji di atas

- a. Hasil uji normalitas di atas menunjukkan variabel 'weight' terdistribusi normal
- b. Hasil uji normalitas di atas menunjukkan variabel 'weight' terdistribusi tidak normal
- c. Method normaltest dari scipy di atas menggunakan metode Shapiro-Wilk
- d. Method normaltest dari scipy di atas menggunakan metode

Kolmogorov-Smirnov

12. Batas atas dari boxplot adalah

- a. $Q1 + (1.5 * IQR)$
- b. $Q1 - (1.5 * IQR)$
- c. $Q3 + (1.5 * IQR)$
- d. $Q2 + (1.5 * IQR)$

13. Tipe data yang bukan merupakan tipe data pada Tableau adalah ...

- a. Text values
- b. Date & Time values
- c. Numerical values
- d. Float values

14. Pernyataan yang tepat mengenai story dan dashboard pada Tableau adalah ...

- a. Story dan dashboard merupakan kumpulan dari worksheet
- b. Story bisa berisi urutan dari beberapa dashboard
- c. Dashboard digunakan untuk menampilkan banyak grafik atau data secara bersamaan
- d. Semua benar

15. Diketahui sebuah dataset berisikan data penjualan rumah secara online di Indonesia memiliki sebuah kolom/variabel bernama 'Harga_Rumah' yang berisi nilai harga dari tiap rumah yang dijual dan distribusi datanya tidak normal. Sebaiknya *measure of central tendency* yang digunakan adalah

- a. Mean
- b. Median
- c. Modus
- d. Standar deviasi

16. Anda memiliki Numpy Array sebagai berikut :

```
arr = np.array([[ 1, 2, 3, 4],  
                [ 6, 7, 8, 9],  
                [10, 11, 12, 13],  
                [14, 15, 16, 17]])
```

Code apa yg kita gunakan jika ingin mengakses/menghasilkan output :

```
array([[ 7, 8],
```

[11, 12]])

- a. arr[1:2 , 1:2]
- b. arr[2:3 , 2:3]
- c. arr[1:3 , 1:3]
- d. arr[2:4 , 2:4]

MySQL

employees

	emp_no	birth_date	first_name	last_name	gender	hire_date
▶	10001	1953-09-02	Georgi	Facello	M	1986-06-26
	10002	1964-06-02	Bezalel	Simmel	F	1985-11-21
	10003	1959-12-03	Parto	Bamford	M	1986-08-28
	10004	1954-05-01	Chirstian	Koblick	M	1986-12-01
	10005	1955-01-21	Kyoichi	Maliniak	M	1989-09-12

salaries

	emp_no	salary	from_date	to_date
▶	10001	60117	1986-06-26	1987-06-26
	10001	62102	1987-06-26	1988-06-25
	10001	66074	1988-06-25	1989-06-25
	10001	66596	1989-06-25	1990-06-25
	10001	66961	1990-06-25	1991-06-25

title

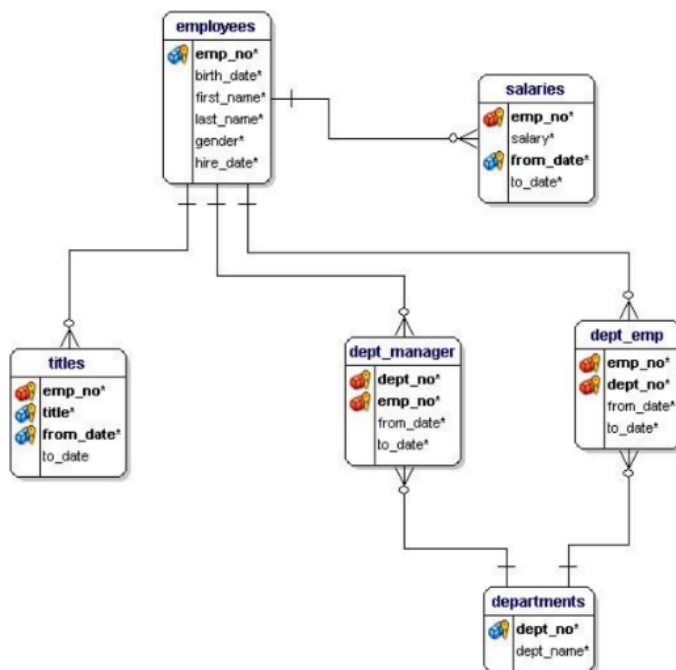
	emp_no	title	from_date	to_date
▶	10001	Senior Engineer	1986-06-26	9999-01-01
	10002	Staff	1996-08-03	9999-01-01
	10003	Senior Engineer	1995-12-03	9999-01-01
	10004	Engineer	1986-12-01	1995-12-01
	10004	Senior Engineer	1995-12-01	9999-01-01
	10005	Senior Staff	1996-09-12	9999-01-01

dept_emp

	emp_no	dept_no	from_date	to_date
▶	10001	d005	1986-06-26	9999-01-01
	10002	d007	1996-08-03	9999-01-01
	10003	d004	1995-12-03	9999-01-01
	10004	d004	1986-12-01	9999-01-01
	10005	d003	1989-09-12	9999-01-01

Keterangan:

- Pada table **employees**. Tiap 'emp_no' adalah karyawan yang berbeda. Tidak ada pengulangan karyawan yang sama pada tabel **employees**.
- Berdasarkan table **salaries**, tiap karyawan bisa memiliki gaji yang berbeda (lebih dari satu record) dikarenakan adanya perubahan gaji tiap tahunnya.
- Berdasarkan table **title**, tiap karyawan bisa memiliki title berbeda (lebih dari satu record) dikarenakan kemungkinan adanya kenaikan pangkat.
- Berdasarkan table **dept_emp**, tiap karyawan bisa memiliki dept_no (departemen) berbeda (lebih dari satu record) dikarenakan kemungkinan adanya pindah departemen.



17. Tampilkan daftar karyawan yang pernah bekerja di departemen yang sama dengan Heng Gideon (tidak harus di waktu yang sama)

a. `SELECT *`

`FROM employees E`

```
JOIN dept_emp D ON E.emp_no = D.emp_no  
WHERE dept_no ON  
  
    (SELECT dept_no  
      FROM employees E  
      JOIN dept_emp D ON E.emp_no = D.emp_no  
      WHERE E.first_name = 'Heng'  
      AND E.last_name = 'Gideon')  
  
;
```

b. SELECT *

```
FROM employees E  
JOIN dept_emp D ON E.emp_no = D.emp_no  
WHERE dept_no IN  
  
    (SELECT dept_no  
      FROM employees E  
      JOIN dept_emp D ON E.emp_no = D.emp_no  
      WHERE E.first_name = 'Heng'  
      AND E.last_name = 'Gideon')  
  
;
```

c. SELECT *

```
FROM employees E  
JOIN dept_emp D ON E.emp_no = D.emp_no  
WHERE dept_no IN  
  
    (SELECT dept_no  
      FROM employees E  
      JOIN dept_emp D ON E.emp_no = D.emp_no  
      WHERE E.first_name = 'Heng'
```



```
AND E.last_name = 'Giveon')
```

```
;
```

```
d. SELECT *
```

```
FROM employees E
```

```
JOIN dept_emp D ON E.emp_no = D.emp_no
```

```
WHERE dept_no IN
```

```
(SELECT dept_no
```

```
FROM employees E
```

```
JOIN dept_emp D ON E.emp_no = D.emp_no
```

```
WHERE E.first_name = 'Heng'
```

```
AND E.last_name = 'Giveon')
```

```
;
```

18. Tampilkan 3 karyawan dengan gaji tertinggi pada tahun kerja 2001 (from_date) sampai 2002 (to_date)

```
a. SELECT *
```

```
FROM employees E
```

```
JOIN salaries S IN E.emp_no = S.emp_no
```

```
WHERE from_date like '2001%' AND to_date like '2002%'
```

```
ORDER BY salary DESC
```

```
LIMIT 3
```

```
;
```

```
b. SELECT *
```

```
FROM employees E
```

```
JOIN salaries S ON E.emp_no = S.emp_no
```

```
WHERE from_date like '2001%' AND to_date like '2002%'
```

```
ORDER BY salary DESC
```

```
LIMIT 3
```

;

c. `SELECT *`
`FROM employees E`
`JOIN salaries S ON E.emp_no = S.emp_no`
`WHERE from_date like '2001%' AND to_date like '2002%'`
`ORDER BY salary DESC`
`LIMIT 3`
`;`

d. `SELECT *`
`FROM employees E`
`JOIN salaries S ON E.emp_no = S.emp_no`
`WHERE from_date like '2001%' AND to_date like '2002%'`
`ORDER BY employees ASC`
`LIMIT 3`
`;`

19. Jika query di bawah ini dijalankan, maka output yang ditampilkan adalah ... (jawablah dalam kalimat)

```
SELECT gender, COUNT(emp_no) FROM employees
WHERE year(hire_date) IN ('1990','1991') GROUP BY gender;
```

- a. Jumlah karyawan berdasarkan gender yang direkrut pada tahun 1990 atau 1991
 - b. Jumlah karyawan berdasarkan gender yang tidak direkrut pada tahun 1990 dan 1991
 - c. Jumlah karyawan berdasarkan gender yang direkrut pada tahun 1990 dan 1991
 - d. Jumlah gender berdasarkan karyawan yang direkrut pada tahun 1990 dan 1991
20. Perhatikan query di bawah ini. Isilah titik-titik untuk menampilkan rata-rata total gaji tiap karyawan (selama bekerja) yang dikelompokkan berdasarkan gender!

```
SELECT gender, AVG(Total_Salary)
```

```
FROM (....) as mySubquery
```

```
GROUP BY gender;
```

a. SELECT gender, AVG(Total_Salary)

```
FROM
```

```
(SELECT E.emp_no, E.gender, SUM(S.salary) is Total_Salary
```

```
FROM employees E JOIN salaries S ON E.emp_no = S.emp_no
```

```
GROUP BY emp_no) as mySubquery
```

```
GROUP BY gender
```

```
;
```

b. SELECT gender, AVG(Total_Salary)

```
FROM
```

```
(SELECT E.emp_no, E.gender, SUM(S.salary) as Total_Salary
```

```
FROM employees E JOIN salaries S IN E.emp_no = S.emp_no
```

```
GROUP BY emp_no) as mySubquery
```

```
GROUP BY gender
```

```
;
```

c. SELECT gender, AVG(Total_Salary)

```
FROM
```

```
(SELECT E.emp_no, E.gender, SUM(S.salary) as Total_Salary
```

```
FROM employees E JOIN salaries S ON E.emp_no = S.emp_no
```

```
GROUP BY emp_no) is mySubquery
```

```
GROUP BY gender
```

```
;
```

d. SELECT gender, AVG(Total_Salary)

```
FROM
```

```
(SELECT E.emp_no, E.gender, SUM(S.salary) as Total_Salary
```

```
FROM employees E JOIN salaries S ON E.emp_no = S.emp_no
```

```
GROUP BY emp_no) as mySubquery
```

```
GROUP BY gender
```

```
;
```

21.