

1. Kode manakah yang akan menghasilkan output berikut:

```
array([[ 1.63988466, -0.88341116,  1.80529247,  1.14795166, -0.0193594 ,
        -0.56581239, -0.30226833],
       [-1.96708111,  0.16962762, -2.07738465, -1.49520289, -0.3129494 ,
         1.40844017,  0.59624262],
       [ 0.64175352,  0.54124008,  0.81040728, -0.46123178, -0.4377472 ,
         0.6484016 , -0.07050411],
       [-0.7519326 , -0.94000345,  1.3638632 , -1.04471599, -0.35552215,
        -0.44581391, -0.98469451],
       [ 0.67702117,  0.41516654,  0.36738315, -0.23842107,  0.0992968 ,
        -0.31953641, -0.09538029]])
```

Pilihlah jawaban yang tepat.

- a. `np.random.rand(7,5)`
- b. `np.random.rand(35).reshape(7,5)`
- c. `np.random.randn(35).reshape(5,7)`
- d. `np.random.randn(35,((7,5)))`
- e. `np.random.randint(0,100,((7,5)))`

2. Perhatikan array berikut:

```
array([[ 10,  20,  30,  40,  50],
       [ 60,  70,  80,  90, 100],
       [  6,   7,   8,   9,  10]])
```

Mana cara yang tepat untuk memanggil sub array berikut:

```
array([[ 80,  90, 100],
       [  8,   9,  10]])
```

- a. `arr[1:,2:]`
- b. `arr[1,2:]`
- c. `arr[2:][1:]`
- d. `arr[2:,1:]`
- e. `arr[1:][2:]`

3. Pada Pandas, metode di bawah ini bisa dipakai untuk menggabungkan 2 dataframe, kecuali:

Pilihlah jawaban yang tepat.

- a. Concat
- b. Extend
- c. Append
- d. Merge
- e. Join

4. Dataframe df berisikan data penumpang, asal kota dan umur. Berikut adalah 5 baris pertama dari dataframe df.

| | survived | pass | sex | age | sibsp | parch | fare | embarked | class | who | adult_male | deck | embark_town | alive | alone |
|---|----------|------|--------|------|-------|-------|---------|----------|-------|-------|------------|------|-------------|-------|-------|
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |

Manakah cara yang paling tepat untuk mencari jumlah asal kota (embark_town) unik yang terdaftar dalam dataframe df?

- a. `df['embark_town'].nunique()`
- b. `df['embark_town'].count()`
- c. `df.sort_values('embark_town', ascending=False)`
- d. `df.groupby('embark_town')`
- e. `df['embark_town'].unique()`

5. Dataframe df berisikan data penumpang, asal kota dan umur. Berikut adalah 5 baris pertama dari dataframe df.

| | survived | pass | sex | age | sibsp | parch | fare | embarked | class | who | adult_male | deck | embark_town | alive | alone |
|---|----------|------|--------|------|-------|-------|---------|----------|-------|-------|------------|------|-------------|-------|-------|
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |

Manakah cara yang tepat untuk mencari rata-rata biaya perjalanan (fare) dari penumpang yang berangkat dari Southampton?

- `df.loc['Southampton']['fare'].mean()`
- `df.iloc[df['embark_town']=='Southampton']['fare'].mean()`
- `df['fare'][df['embark_town']=='Southampton'].mean()`
- Semua benar
- Semua salah

6. Dataframe df berisikan data penumpang, asal kota dan umur. Berikut adalah 5 baris pertama dari dataframe df.

| | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult_male | deck | embark_town | alive | alone |
|---|----------|--------|--------|------|-------|-------|---------|----------|-------|-------|------------|------|-------------|-------|-------|
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |

Manakah penjelasan yang tepat untuk output dari kode berikut:

```
df.groupby('embark_town').mean().sort_values('fare', ascending=False)
['fare'].iloc[:3]
```

- 3 kota asal ('embark_town') dan rata-rata total biaya perjalanannya ('fare') yang memiliki total biaya perjalanannya ('fare') terendah
- 3 kota asal ('embark_town') yang memiliki total biaya perjalanannya ('fare') paling rendah
- 3 kota asal ('embark_town') dan rata-rata total biaya perjalanannya ('fare') yang memiliki total biaya perjalanannya ('fare') tertinggi
- 3 kota asal ('embark_town') yang memiliki total biaya perjalanannya ('fare') paling tinggi
- Rata-rata total biaya perjalanannya ('fare') berdasarkan kota asal ('embark_town')

7. Anda tertarik untuk melihat apakah biaya perjalanan ('fare') memiliki hubungan dengan umur penumpang ('age'). Berikut adalah hasil uji normalitas data anda:

| | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult_male | deck | embark_town | alive | alone |
|---|----------|--------|--------|------|-------|-------|---------|----------|-------|-------|------------|------|-------------|-------|-------|
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |

```
1 normaltest(df['age'])
```

```
NormaltestResult(statistic=53.46581133311897, pvalue=2.454972229498001e-12)
```

```
1 normaltest(df['fare'])
```

```
NormaltestResult(statistic=904.5872705326926, pvalue=3.727091062842147e-197)
```

Manakah cara yang paling tepat untuk melakukan uji korelasi antara 'fare' dan 'age'?

- `df['fare'].corr(df['age'], method='pearson')`
 - `df['fare'].corr(df['age'])`
 - `df.corr()[['fare', 'age']]`
 - `df[['fare', 'age']].corr('spearman')`
 - `df[['fare', 'age']].corr()`
8. Anda ingin memperlihatkan tren hubungan antara umur penumpang ('age') dan biaya perjalanan ('fare') dengan membuat plot. Pilihlah cara yang tepat untuk membuat plot yang sesuai!

| | survived | pclass | sex | age | sibsp | parch | fare | embarked | class | who | adult_male | deck | embark_town | alive | alone |
|---|----------|--------|--------|------|-------|-------|---------|----------|-------|-------|------------|------|-------------|-------|-------|
| 0 | 0 | 3 | male | 22.0 | 1 | 0 | 7.2500 | S | Third | man | True | NaN | Southampton | no | False |
| 1 | 1 | 1 | female | 38.0 | 1 | 0 | 71.2833 | C | First | woman | False | C | Cherbourg | yes | False |
| 2 | 1 | 3 | female | 26.0 | 0 | 0 | 7.9250 | S | Third | woman | False | NaN | Southampton | yes | True |
| 3 | 1 | 1 | female | 35.0 | 1 | 0 | 53.1000 | S | First | woman | False | C | Southampton | yes | False |
| 4 | 0 | 3 | male | 35.0 | 0 | 0 | 8.0500 | S | Third | man | True | NaN | Southampton | no | True |

```
1 normaltest(df['age'])
```

```
NormaltestResult(statistic=53.46581133311897, pvalue=2.454972229498001e-12)
```

```
1 normaltest(df['fare'])
```

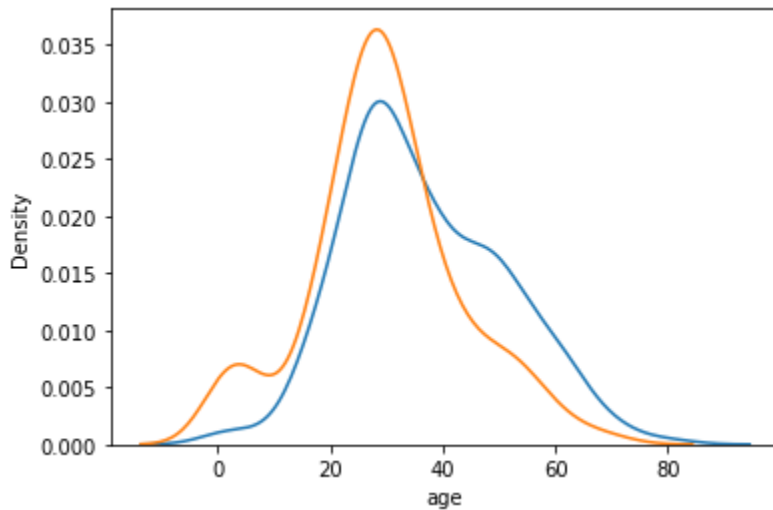
```
NormaltestResult(statistic=904.5872705326926, pvalue=3.727091062842147e-197)
```

Pilihlah jawaban yang tepat.

- a. `plt.lineplot(x=df['age'], y=df['fare'])`
- b. `sns.lineplot(data=df, x='age', y='fare')`
- c. `plt.scatterplot(x=df['age'], y=df['fare'])`
- d. `sns.scatter(x=df['age'], y=df['fare'])`
- e. `sns.scatterplot(x=df['age'], y=df['fare'])`

9. Perhatikan plot berikut ini:

Bagaimana cara yang tepat untuk membuat plot di atas? Dengan asumsi matplotlib.pyplot as plt dan seaborn as sns.



- a. `sns.lineplot(df[df['pclass']==1]['age'])`
`sns.lineplot(df[df['pclass']==2]['age'])`
`plt.show()`

- b.

```
plt. subplot(211)

sns.lineplot(df[df['pclass']==1]['age'])

plt. subplot(211)

sns.lineplot(df[df['pclass']==2]['age'])

plt.show()
```
- c.

```
sns.displot(df[df['pclass']==1]['age'])

sns.displot(df[df['pclass']==2]['age'])

plt.show()
```
- d.

```
sns.kdeplot(df[df['pclass']==1]['age'])

sns.kdeplot(df[df['pclass']==2]['age'])

plt.show()
```
- e.

```
plt.subplot(211)

sns.displot(df[df['pclass']==1]['age'])

plt.subplot(211)

sns.displot(df[df['pclass']==2]['age'])

plt.show()
```

10. Statistik deskriptif adalah jenis statistik yang menggambarkan kondisi data kita. Pilihlah pernyataan yang benar mengenai statistik deskriptif dari pernyataan-pernyataan berikut!, kecuali:

Pilihlah jawaban yang tepat.

- a. Statistik deskriptif mencakup analisis data melalui grafik, bagan, tabel, dan penghitungan berbagai ukuran deskriptif seperti rata-rata, variasi, persentil, dan lainnya.
- b. Ukuran Pusat Data atau Measures of Central Tendency. Ringkasan ini menggambarkan posisi sentral dari distribusi frekuensi untuk sekelompok data.

- c. Ukuran Penyebaran Data atau Measures of Spread. Ringkasan ini menggambarkan seberapa tersebar data kita.
- d. Apabila distribusi data anda miring, gunakan median sebagai ukuran pusat data
- e. Mean (rata-rata) paling sering digunakan sebagai ukuran pusat data kualitatif.

11. Pilihlah pernyataan yang salah tentang uji statistik dari pernyataan-pernyataan berikut!, kecuali:

Pilihlah jawaban yang tepat.

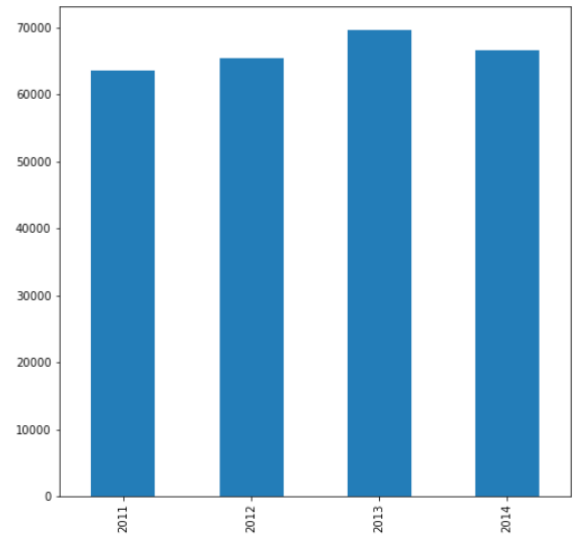
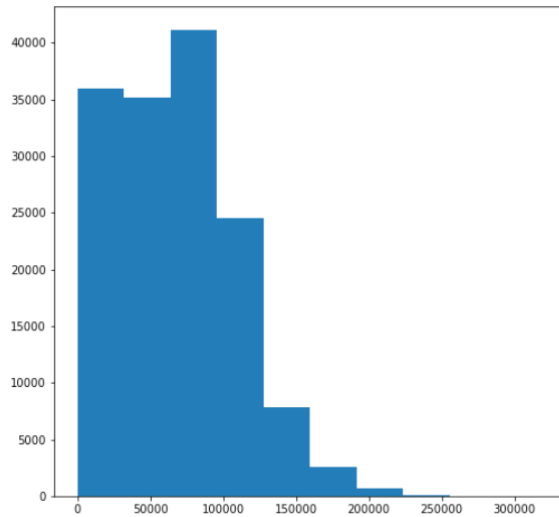
- a. Mann-whitney, Wilcoxon dan Kruskal-Walis merupakan contoh uji statistik parametrik.
- b. Uji statistic non-parametrik menggunakan mean sebagai ukuran pusat data
- c. Uji statistic parametrik digunakan ketika kita tidak memiliki cukup data
- d. Pearson, ANOVA dan t-test merupakan contoh uji statistic non parametrik.
- e. Ada dua jenis uji statistik, yaitu parametrik dan non-parametrik

12. Distribusi normal merupakan konsep yang penting dalam statistik. Pilihlah pernyataan yang benar tentang distribusi normal dari pernyataan-pernyataan berikut!, kecuali:

Pilihlah jawaban yang tepat.

- a. Semua jawaban salah
- b. Histogram dan boxplot dapat dipakai untuk melihat distribusi suatu data.
- c. Distribusi normal penting dalam algoritma Machine Learning.
- d. Jika data memiliki distribusi normal, maka nilai mean dan median akan berada di titik yang sama.
- e. Pada distribusi normal, persebaran data secara standar deviasi lebih banyak di dekat nilai rata-rata.

13. Perhatikan gambar berikut. Pilihlah pernyataan yang benar mengenai kedua gambar di bawah, kecuali:



Pilihlah jawaban yang tepat.

- Plot a adalah histogram, sementara plot b adalah barplot
- Plot a dan b adalah jenis plot yang bisa digunakan untuk menilai apakah data terdistribusi normal atau tidak.
- Plot a untuk visualisasi data numerik, sementara plot b untuk visualisasi data kategorikal.
- Pada plot b, setiap titik data ditampilkan sebagai bilah terpisah. Sementara pada plot a, titik data dikelompokkan dan dirender berdasarkan nilai bin. Seluruh rentang nilai data dibagi menjadi serangkaian interval yang tidak tumpang tindih.
- Plot b adalah jenis plot yang bisa digunakan untuk membandingkan berbagai kategori data, sedangkan plot a adalah jenis plot yang bisa digunakan untuk menilai apakah data terdistribusi normal atau tidak.

14. Dalam pembuatan visualisasi, perlu diperhatikan agar plot yang dibuat sesuai dengan data yang ada. Pilihlah pernyataan yang tepat tentang tipe plot di antara pernyataan berikut, kecuali:

Pilihlah jawaban yang tepat.

- Line plot dipakai untuk melihat trend suatu kolom data dari waktu ke waktu (time series)
- Histogram dan box plot adalah jenis plot yang bisa digunakan untuk menilai apakah data terdistribusi normal atau tidak.
- Barplot paling sesuai digunakan untuk melihat aggregate data numerik berdasarkan suatu kategori

- d. Boxplot digunakan untuk melihat sebaran data dan outliers dari suatu kolom berisikan data numerik
- e. Pie chart adalah grafik statistik melingkar, yang dibagi menjadi beberapa bagian untuk menggambarkan proporsi numerik

15. Pilihlah pernyataan yang benar mengenai jenis grafik pada tableau!, kecuali:

Pilihlah jawaban yang tepat.

- a. Tidak ada jawaban yang benar.
- b. Tidak ada jawaban yang salah.
- c. Pakailah line graph untuk melihat tren data berdasarkan waktu.
- d. Tableau bisa digunakan untuk membuat heatmap.
- e. Dual axis graph digunakan untuk membuat dua sumbu-y.

16. Pilihlah pernyataan yang tepat mengenai Story dan Dashboard pada tableau!, kecuali:

Pilihlah jawaban yang tepat.

- a. Semua pilihan benar.
- b. Baik story dan dashboard merupakan kumpulan dari worksheet.
- c. Pada story, visualisasi ditunjukkan secara berurutan untuk menyampaikan informasi, sedangkan pada dashboard, visualisasi ditunjukkan secara bersamaan agar dapat dibandingkan dan di-monitor.
- d. Ada 2 button yang berbeda untuk membuat Story dan Dashboard di Tableau.
- e. Semua pilihan salah.

No. 17 - 20 menggunakan database seperti di bawah ini:

Perhatikan database schema dan contoh sebagian isinya. Berikut keterangan tabelnya:

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

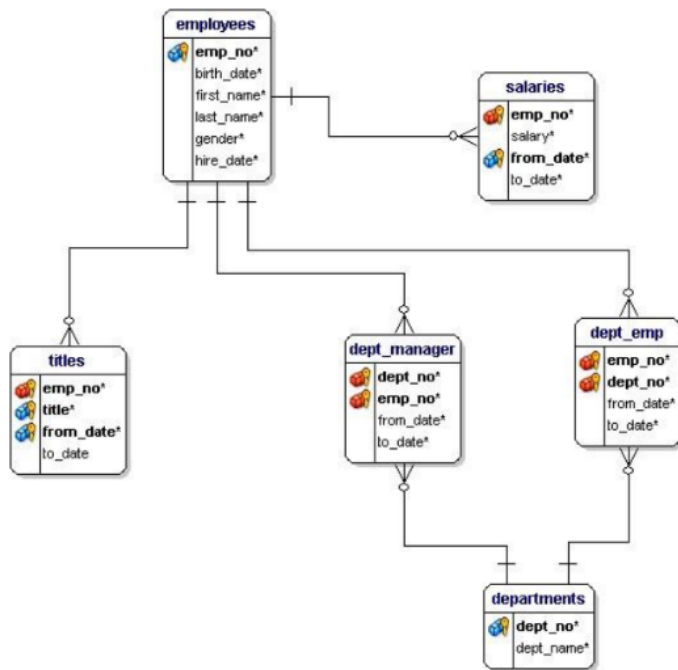
| | emp_no | title | from_date | to_date |
|---|--------|-----------------|------------|------------|
| ▶ | 10001 | Senior Engineer | 1986-06-26 | 9999-01-01 |
| | 10002 | Staff | 1996-08-03 | 9999-01-01 |
| | 10003 | Senior Engineer | 1995-12-03 | 9999-01-01 |
| | 10004 | Engineer | 1986-12-01 | 1995-12-01 |
| | 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)
- Tidak ada jawaban yang benar
 - Tidak ada jawaban yang salah
 - 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

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

WHERE E.first_name = 'Heng'

AND E.last_name = 'Gideon')

;

- 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')
;

```

```

e. 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
     LEFT 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 ON E.emp_no = S.emp_no
WHERE from_date like '2001%' AND to_date like '2002%'
ORDER BY salary ASC
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 ASC

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 salary DESC

LIMIT 3

;

e. Tidak ada jawaban yang benar

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 di antara tahun 1990 dan 1991.
- b. Jumlah karyawan berdasarkan gender yang direkrut sebelum tahun 1990 dan 1991.
- c. Jumlah karyawan berdasarkan gender yang direkrut sesudah tahun 1990 dan 1991.
- d. Jumlah karyawan berdasarkan gender yang direkrut pada tahun 1990 dan 1991.
- e. Tidak ada jawaban yang benar.

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 E.emp_no, E.gender, SUM(S.salary) as Total_Salary
FROM employees E FULL JOIN salaries S ON S.emp_no = E.emp_no
GROUP BY emp_no;
- b. 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;
- c. SELECT E.emp_no, E.gender, SUM(S.salary) as Total_Salary
FROM employees E OUTER JOIN salaries S ON E.emp_no = S.emp_no
GROUP BY emp_no;
- d. SELECT E.emp_no, E.gender, SUM(S.salary) as Total_Salary
FROM employees E OUTER JOIN salaries S ON S.emp_no = E.emp_no
GROUP BY emp_no;
- e. Tidak ada jawaban yang benar

21.

22.