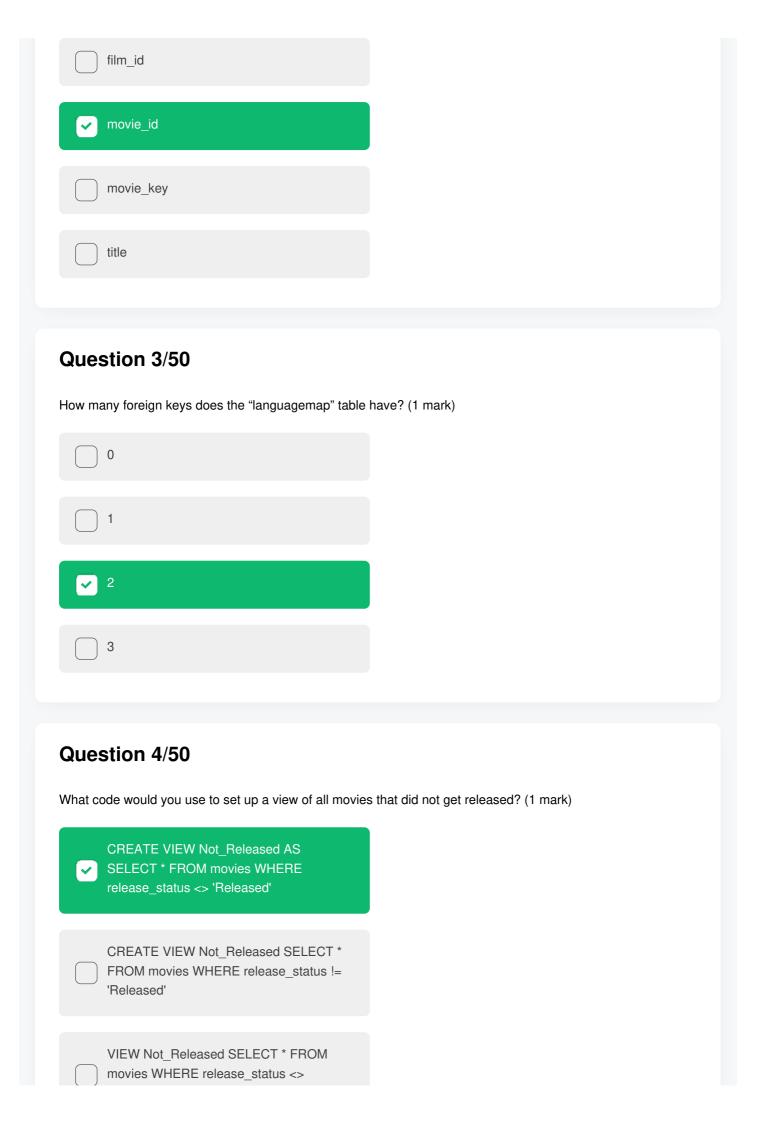
# Final Integrated Exam - Part 1: Fundamentals (Version : 1)

# **TEST** Correct Answer (L) Answered in 167.3166666667 Minutes Uploaded File: 1652217285-Final\_Integrated\_Exam\_-\_Part\_1:\_Fundamentals-5665-Eric\_Mbuthia.Zip Question 1/50 Total: 90 Marks Gather (20 Questions, 40 Marks) Questions 1 - 20 You should have access to a file called TMDB.db and a starter notebook. The first step to answering this set of questions will be to use the starter notebook provided and connect to the TMDB.db file to access the data. Based on that data, answer the following questions: What is the code you can use to connect to the TMBD database that is saved in the same location as the Jupyter notebook? (1 mark) sqlite:/TMDB.db sqlite:///TMDB.db sqlconnect:///TMDB.db/ TMDB.db:///sqlite/connect

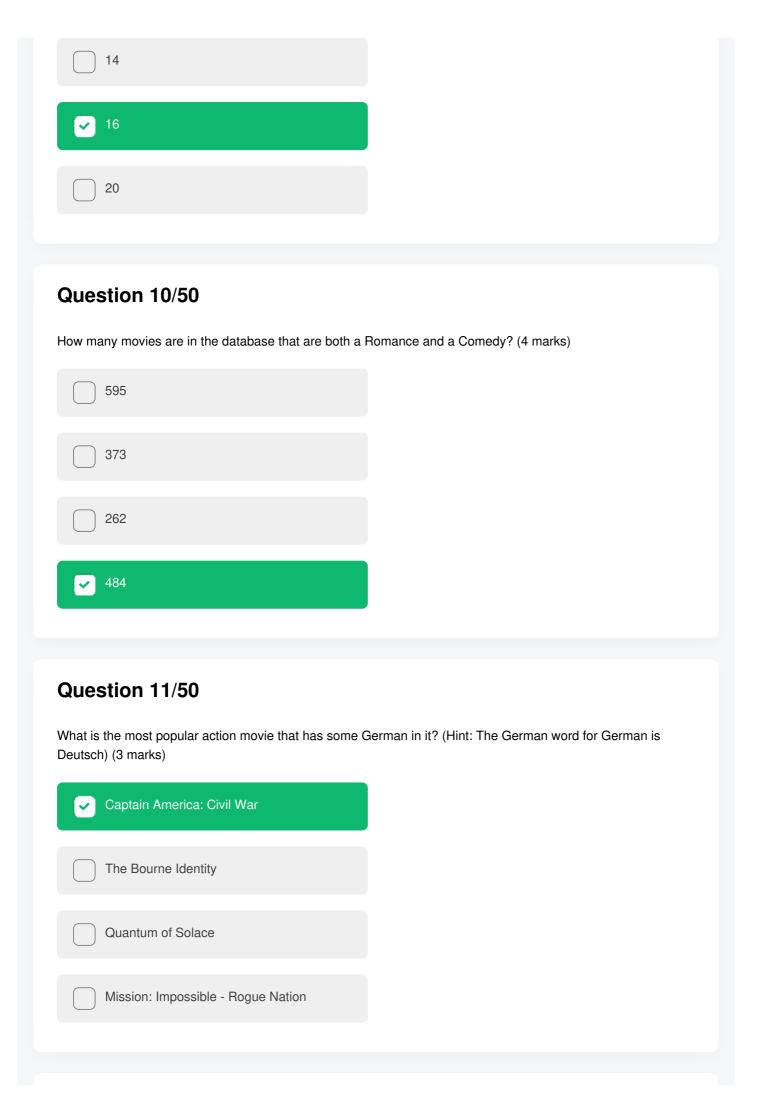
# Question 2/50

What is the primary key for the table "movies"? (1 mark)



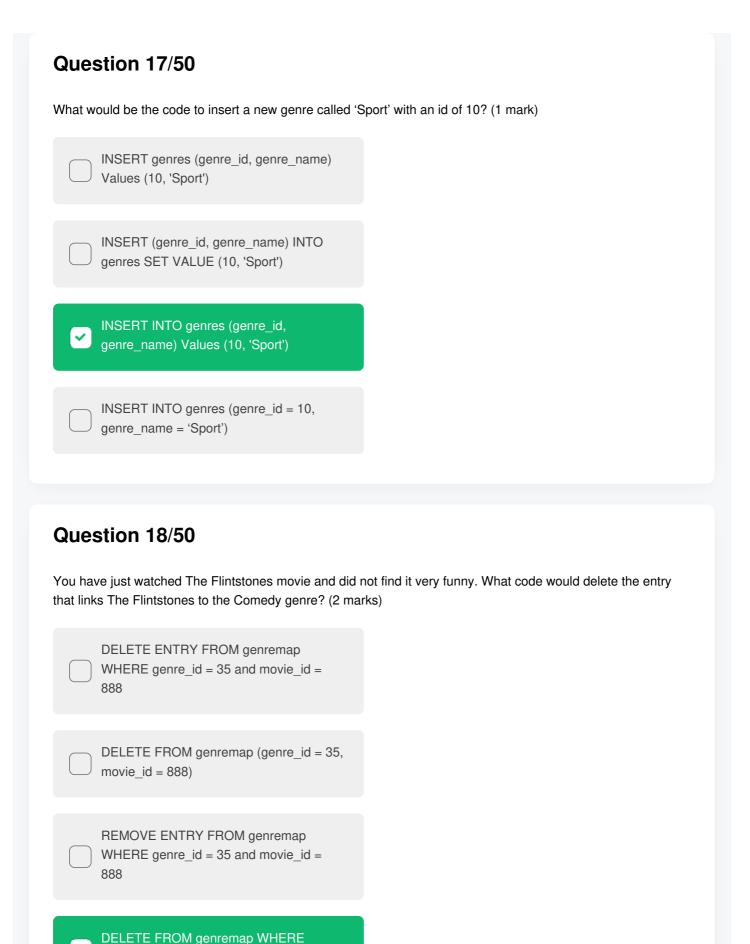
'Released'	
NEW VIEW Not_Released SELECT * FROM movies WHERE release_status <> 'Released'	
Question 5/50	
How would you select only the title, release date and releprevious question? (1 mark)	ease status columns from the view you created in the
Select title, release_date, and release_status From View Not_Released	
Select title, release_date, release_status From Not_Released	
Select title, release_date, release_status From Not_Released VIEW	
Select title, release_date, release_status From VIEW Not_Released	
Question 6/50	
How many movies exist that no longer use their original	titles? (2 marks)
24	
261	
74	
187	

Question 7/50
What is the most popular movie that was made after 01/01/2000 with a budget of more than \$100 000 000? (Hint: Use the popularity field in the Movies table. Larger numbers are more popular.) (3 marks)
The Dark Knight
Pirates of the Caribbean: The Curse of the Black Pearl
Interstellar
Avatar
Question 8/50
How many movies do not have English as their original language? (2 marks)
315
298
492
387
Question 9/50
How many movies in the database were produced by Pixar Animation Studios?(3 marks)
18



# Question 12/50 In how many movies did Tom Cruise portray the character Ethan Hunt? (3 marks) Question 13/50 How many times was the actress Cate Blanchett nominated for an Oscar? (3 marks) 2 5 Question 14/50 How many movies contains at least one of the official South African Languages, Afrikaans or Zulu? (3 marks) 10 12

15
Question 15/50
Which of the movies mentioned above is the most popular? (3 marks)
District 9
Blood Diamond
Gangster's Paradise: Jerusalema
Tsotsi
Question 16/50
What would be the code to change the name of the language with the 'zh' iso code in the "language" table to 'Chinese'? (1 mark)
UPDATE languages SET language_name = 'Chinese' WHERE iso_639_1 = 'zh'
UPDATE languages (language_name = 'Chinese') WHERE iso_639_1 = 'zh'
ALTER languages SET language_name =  'Chinese' WHERE iso_639_1 = 'zh'
MODIFY languages SET language_name = 'Chinese' WHERE iso_639_1 = 'zh'



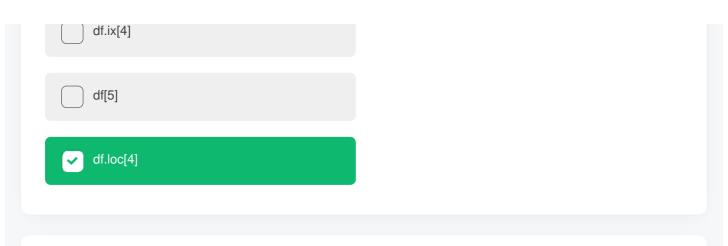
genre\_id = 35 and movie\_id = 888

what code will give me the 10 most recently released m	iovies in the database: (1 mark)
SELECT TOP(10) * FROM movies ORDER BY release_date DESC	
SELECT TOP(10) * FROM movies ORDER BY release_date ASC	
SELECT * FROM movies ORDER BY release_date ASC LIMIT 10	
SELECT * TOP(10) FROM movies ORDER BY release_date DESC	
Question 20/50  What code would you use to add a column to the langua different languages? (1 mark)	age table that could be used for the English names of the
ALTER TABLE languages ADD language_english_name	
ALTER TABLE languages ADD language_english_name varchar(50)	
UPDATE TABLE languages APPEND language_english_name varchar(50)	
UPDATE TABLE languages ADD language_english_name	
O	
Question 21/50	

Analyse (15 Questions, 25 Marks) Questions 21 - 35

The practical questions of this section should be answered by analysing the football_players.csv dataset – the recommended tool is Python (Pandas library).
In NumPy arrays: all data should be of the same type (1 mark)
True
False
Question 22/50
Which of the following statements about numpy arrays is false? (1 mark)
It is exactly the same as lists of lists
It can be modified (changed)
It can have more than 2 dimensions
It is indexed by a tuple of non-negative integers
Question 23/50
Which of the following statements about pandas dataframes is false? (1 mark)
Data types can differ between columns
The data type must be the same within a column
It can be modified (changed)
It usually consists of more than 2 dimensions

Question 24/50
How would you select the top left element in a 2-D numpy array a? (1 mark)
a[1,1]
a[:,1]
It is not possible
✓ a[0,0]
Question 25/50
How would you add a row of entries 92, 88 and 78 to a 2-D numpy array a with dimensions of 3x3? (1 mark)
np.append(a, 92, 88, 78, axis=0)
pnp.append(a, [[92, 88, 78]], axis=0)
np.append(a, [[92, 88, 78]], axis=1)
np.append(a, [92, 88, 78], axis=0)
Question 26/50
How would you select the 5th row of a Pandas DataFrame, df, with names as indices? (1 mark)
df[4]



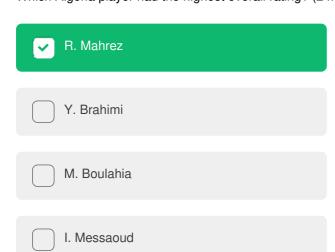
### Question 27/50

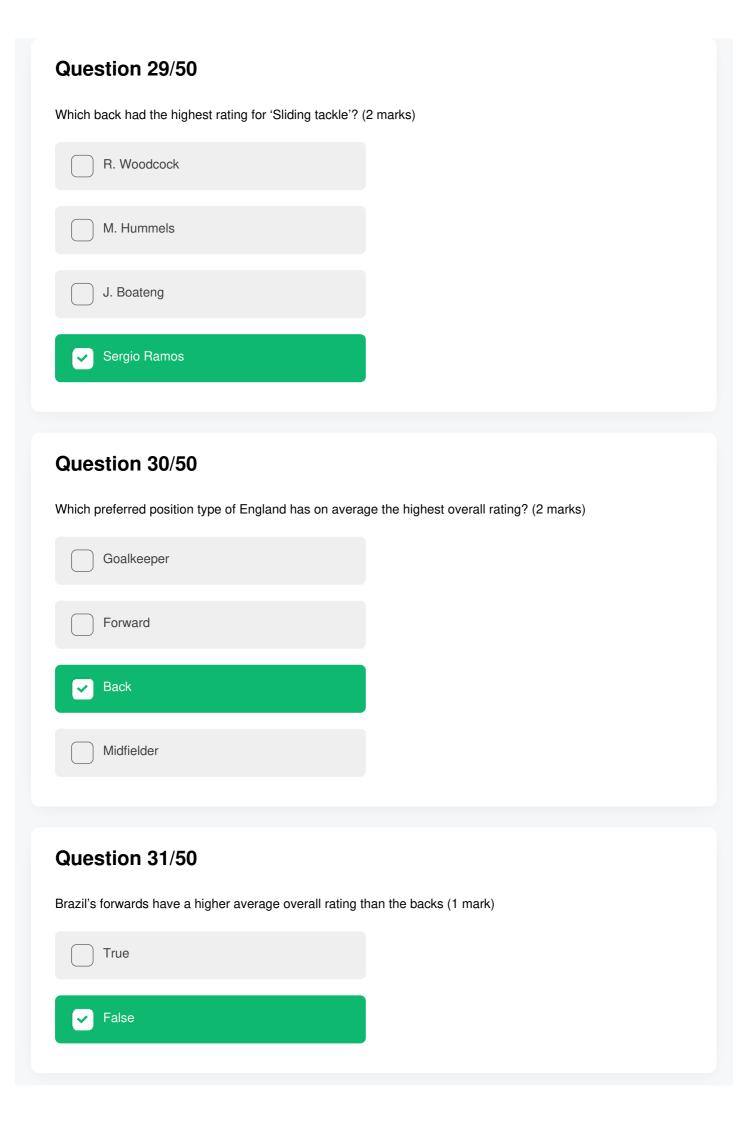
How would you create a new column to transform a column 'Age' from years to decades? (1 mark)

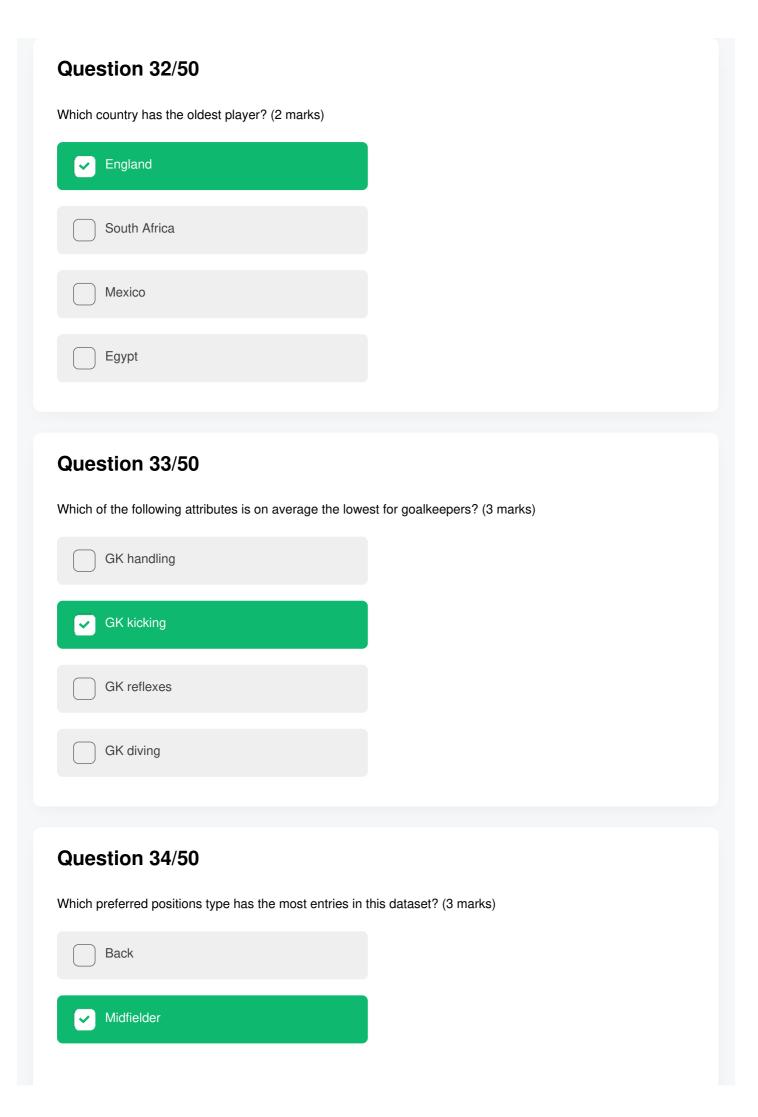
- decades = df['Age'].apply(lambda x: x / 10)
- df['decades'] = df.applymap(lambda x: x /
  10)
- df['decades'] = df['Age'].apply(lambda x: x
  / 10)

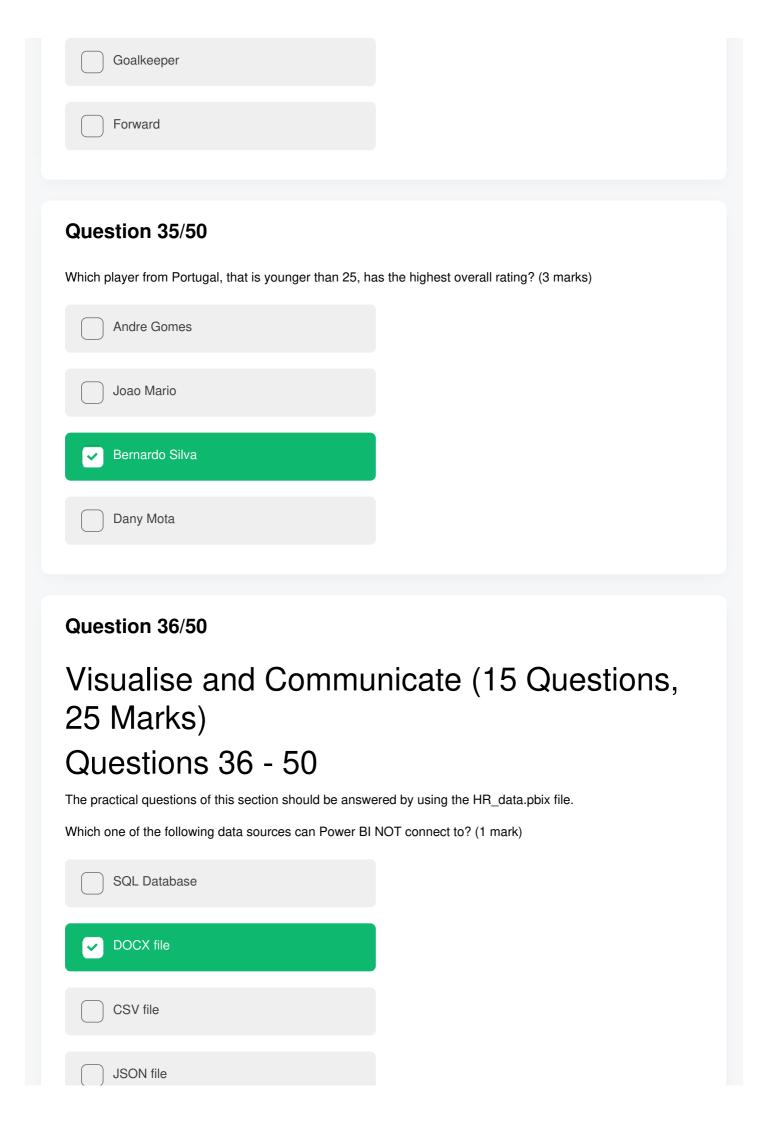
# Question 28/50

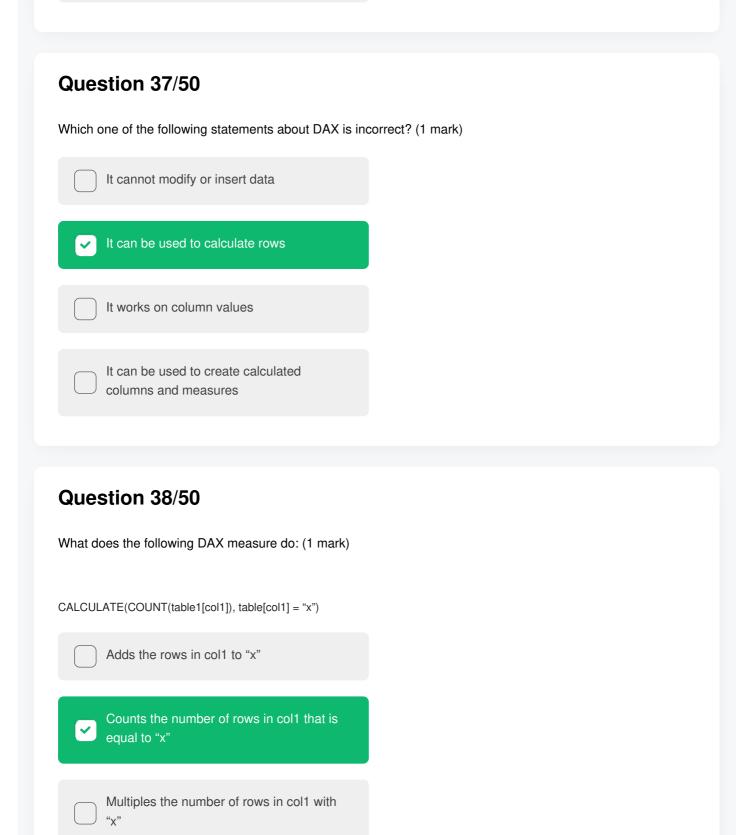
Which Algeria player had the highest overall rating? (2 marks)







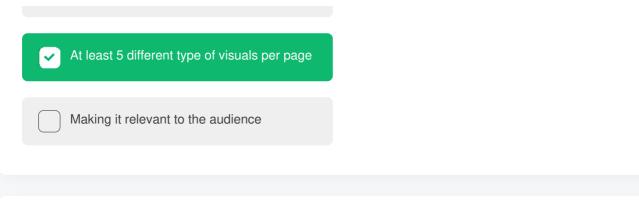




equal to "x"

Calculates the number of times col1 is not

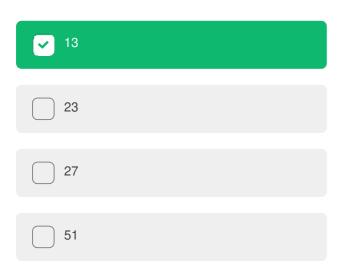
Which type of visual in Power BI would be best suited if we want to display data per category, when we have many categories with long names? (1 mark)
Column Chart
✓ Bar Chart
Line Chart
Pie Chart
Question 40/50
Which of the following 3 actions can be performed in the formatting section of a visual in Power BI? (1 mark)
<ol> <li>Rounding data</li> <li>Changing the data type</li> <li>Changing font size</li> </ol>
Only 1 and 3
Only 1
All of them
Only 2 and 3
Question 41/50
Which one of the following is not regarded as best practice when building a dashboard? (1 mark)
Don't clutter your dashboard
Group data logically



## Question 42/50

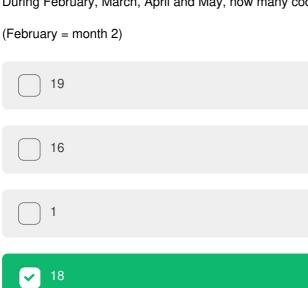
The next 9 questions (Question 42 - 50) are based on the HR Absenteeism dataset (HR\_data.pbix)

What code was the fourth most common reason for absenteeism? (2 marks)



### Question 43/50

During February, March, April and May, how many code 19 reports were recorded in the total? (2 marks)



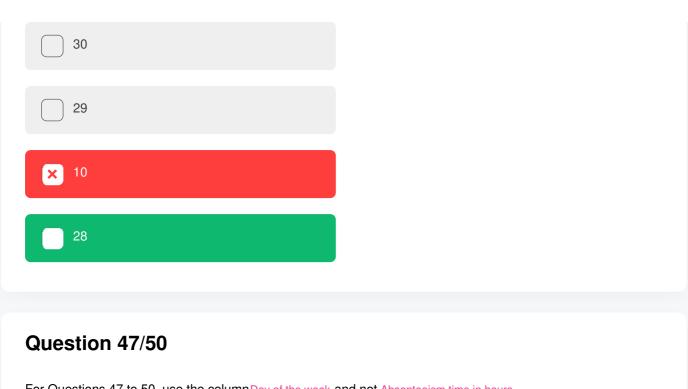
# 

# 

## Question 46/50

Which of the listed employees meet all of the following criteria? (2 marks)

- Social drinkers (yes=1)
- Not social smokers (no = 0)
- Overweight (BMI > 25)



# Question 47/50 For Questions 47 to 50, use the columnDay of the week and not Absenteeism time in hours. An absent day is any day with absenteeism, irrespective of the number of hours absent on a day. Which employee recorded the most number of absent days? (2 marks) 22 33

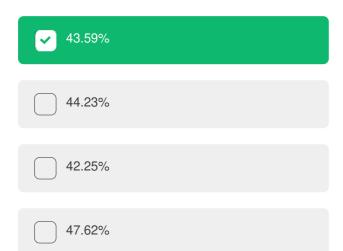
28

# Question 48/50 How many absent days did the employee with the fourth most absent days record? (2 marks) 33 42

# Question 49/50

What percentage of absenteeisms were on Mondays or Fridays for the employee with the fourth most absent days? (3 marks)

(Monday = 2 and Friday = 6)



# Question 50/50

What was the average hit rate during the month of December, for employees with more than 40 absenteeisms in total? (2 marks)

