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

TEST

Correct Answer

Answered in 167.31666666667 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
- [x] sqlite:///TMDB.db
- [ ] sqlconnect:///TMDB.db/
- [ ] TMDB.db:///sqlite/connect

Question 2/50

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

- [ ] film_id
- [x] movie_id
- [ ] movie_key
- [ ] title

Question 3/50

How many foreign keys does the “languagemap” table have? (1 mark)

- [ ] 0
- [ ] 1
- [x] 2
- [ ] 3

Question 4/50

What code would you use to set up a view of all movies that did not get released? (1 mark)

- [x] CREATE VIEW Not_Released AS SELECT * FROM movies WHERE release_status <> 'Released'
- [ ] CREATE VIEW Not_Released SELECT * FROM movies WHERE release_status != 'Released'
- [ ] VIEW Not_Released SELECT * FROM movies WHERE release_status <>

'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 release status columns from the view you created in the previous question? (1 mark)

- [ ] Select title, release_date, and release_status From View Not_Released
- [x] 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
- [x] 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
- [x] Interstellar
- [ ] Avatar

Question 8/50

How many movies do not have English as their original language? (2 marks)

- [ ] 315
- [x] 298
- [ ] 492
- [ ] 387

Question 9/50

How many movies in the database were produced by Pixar Animation Studios?(3 marks)

- [ ] 18

- [ ] 14
- [x] 16
- [ ] 20

Question 10/50

How many movies are in the database that are both a Romance and a Comedy? (4 marks)

- [ ] 595
- [ ] 373
- [ ] 262
- [x] 484

Question 11/50

What is the most popular action movie that has some German in it? (Hint: The German word for German is Deutsch) (3 marks)

- [x] Captain America: Civil War
- [ ] The Bourne Identity
- [ ] Quantum of Solace
- [ ] Mission: Impossible - Rogue Nation

Question 12/50

In how many movies did Tom Cruise portray the character Ethan Hunt? (3 marks)

- [x] 5
- [ ] 4
- [ ] 6
- [ ] 1

Question 13/50

How many times was the actress Cate Blanchett nominated for an Oscar? (3 marks)

- [ ] 2
- [ ] 4
- [ ] 5
- [x] 7

Question 14/50

How many movies contains at least one of the official South African Languages, Afrikaans or Zulu? (3 marks)

- [x] 8
- [ ] 10
- [ ] 12

- [ ] 15

Question 15/50

Which of the movies mentioned above is the most popular? (3 marks)

- [x] 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)

- [x] 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’

Question 17/50

What would be the code to insert a new genre called ‘Sport’ with an id of 10? (1 mark)

- [ ] INSERT genres (genre_id, genre_name) Values (10, 'Sport')
- [ ] INSERT (genre_id, genre_name) INTO genres SET VALUE (10, 'Sport')
- [x] INSERT INTO genres (genre_id, genre_name) Values (10, 'Sport')
- [ ] INSERT INTO genres (genre_id = 10, genre_name = ‘Sport’)

Question 18/50

You have just watched The Flintstones movie and did not find it very funny. What code would delete the entry that links The Flintstones to the Comedy genre? (2 marks)

- [ ] DELETE ENTRY FROM genremap WHERE genre_id = 35 and movie_id = 888
- [ ] DELETE FROM genremap (genre_id = 35, movie_id = 888)
- [ ] REMOVE ENTRY FROM genremap WHERE genre_id = 35 and movie_id = 888
- [x] DELETE FROM genremap WHERE genre_id = 35 and movie_id = 888

Question 19/50

What code will give me the 10 most recently released movies in the database? (1 mark)

- [x] 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 language table that could be used for the English names of the different languages? (1 mark)

- [ ] ALTER TABLE languages ADD language_english_name
- [x] 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

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)

- [x] True
- [ ] False

Question 22/50

Which of the following statements about numpy arrays is false? (1 mark)

- [x] 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)
- [x] 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
- [x] 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)
- [x] np.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]

- [ ] df.ix[4]
- [ ] df[5]
- [x] df.loc[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’].map(def lambda x: x / 10)
- [x] df[‘decades’] = df[‘Age’].apply(lambda x: x / 10)

Question 28/50

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

- [x] R. Mahrez
- [ ] Y. Brahimi
- [ ] M. Boulahia
- [ ] I. Messaoud

Question 29/50

Which back had the highest rating for ‘Sliding tackle’? (2 marks)

- [ ] R. Woodcock
- [ ] M. Hummels
- [ ] J. Boateng
- [x] Sergio Ramos

Question 30/50

Which preferred position type of England has on average the highest overall rating? (2 marks)

- [ ] Goalkeeper
- [ ] Forward
- [x] Back
- [ ] Midfielder

Question 31/50

Brazil’s forwards have a higher average overall rating than the backs (1 mark)

- [ ] True
- [x] False

Question 32/50

Which country has the oldest player? (2 marks)

- [x] England
- [ ] South Africa
- [ ] Mexico
- [ ] Egypt

Question 33/50

Which of the following attributes is on average the lowest for goalkeepers? (3 marks)

- [ ] GK handling
- [x] GK kicking
- [ ] GK reflexes
- [ ] GK diving

Question 34/50

Which preferred positions type has the most entries in this dataset? (3 marks)

- [ ] Back
- [x] Midfielder

- [ ] Goalkeeper
- [ ] Forward

Question 35/50

Which player from Portugal, that is younger than 25, has the highest overall rating? (3 marks)

- [ ] Andre Gomes
- [ ] Joao Mario
- [x] Bernardo Silva
- [ ] Dany Mota

Question 36/50

Visualise and Communicate (15 Questions, 25 Marks)

Questions 36 - 50

The practical questions of this section should be answered by using the HR_data.pbix file.

Which one of the following data sources can Power BI NOT connect to? (1 mark)

- [ ] SQL Database
- [x] DOCX file
- [ ] CSV file
- [ ] JSON file

Question 37/50

Which one of the following statements about DAX is incorrect? (1 mark)

- [ ] It cannot modify or insert data
- [x] It can be used to calculate rows
- [ ] It works on column values
- [ ] It can be used to create calculated columns and measures

Question 38/50

What does the following DAX measure do: (1 mark)

```
CALCULATE(COUNT(table1[col1]), table[col1] = “x”)
```

- [ ] Adds the rows in col1 to “x”
- [x] Counts the number of rows in col1 that is equal to “x”
- [ ] Multiples the number of rows in col1 with “x”
- [ ] Calculates the number of times col1 is not equal to “x”

Question 39/50

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
- [x] 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)

1. Rounding data
2. Changing the data type
3. Changing font size

- [x] 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

- [x] At least 5 different type of visuals per page
- [ ] Making it relevant to the audience

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)

- [x] 13
- [ ] 23
- [ ] 27
- [ ] 51

Question 43/50

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

(February = month 2)

- [ ] 19
- [ ] 16
- [ ] 1
- [x] 18

Question 44/50

Which month had the second-highest sum of absenteeism hours, starting from January? (2 marks)

(January = month 1)

- [x] July
- [ ] February
- [ ] January
- [ ] March

Question 45/50

How many employees are: (2 marks)

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

- [ ] 9
- [ ] 1
- [x] 8
- [ ] 11

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)

- [ ] 30
- [ ] 29
- [x] 10
- [ ] 28

Question 47/50

For Questions 47 to 50, use the column Day 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
- [x] 3
- [ ] 28

Question 48/50

How many absent days did the employee with the fourth most absent days record? (2 marks)

- [ ] 33
- [ ] 42
- [ ] 37

- [x] 39

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)

- [x] 43.59%
- [ ] 44.23%
- [ ] 42.25%
- [ ] 47.62%

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)

- [x] 95.59%
- [x] 95.03%
- [ ] 94.47%
- [ ] 93%