|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Part 1 – Questions  Duration: 10 minutes  Cannot navigate backwards | | | | | | | |
| # | Skill code | Question type (for available q-types refer to this [link](https://help.sabacloud.com/sabacloud/help-learning/topics/help-assessment-question-types-supported-by-tests.html)) | Question | Link to .png image | Question choices (if applicable) | Correct answer | Link to dataset (if applicable) |
| 8 | DS1-1 | Multiple choice | Which variable from below most likely comes from a normal distribution? | [n8\_0](https://worldbankgroup.sharepoint.com/:i:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/images_olc/n8_0.png?csf=1&web=1&e=qFBy0S) | -x  -y  -z | -x |  |
| 16 | DS1-4 | Multiple choice | What is the output of the following Python code?  x = [6, 28, 5, 26, 49]  print(x[1:-1]) |  | A) [6, 28, 5, 26, 49]  B) [28, 5, 26, 49]  C) [28, 5, 26]  D) [6, 28, 5, 26] | C |  |
| 18 | DS1-4 | All that apply | Suppose you have defined the following list in your Python environment:  x = ['a', 'b', 'c']  Which of the following syntaxes would make x equal to:  ['a', 'b', 'c', 'd'] |  | A) x.append('d')  B) x += ['d']  C) x = x.append('d')  D) x = ['d'] | A, B |  |
| 20 | DS1-4 | Multiple choice | What is the output of the following Python code? | [n20\_0](https://worldbankgroup.sharepoint.com/:i:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/images_olc/n20_0.png?csf=1&web=1&e=9AZMVW) | A) [True, True, False]  B) True  C) [False, False, True]  D) [False, False, True, True]  E) [False, False, True, False]  F) [True, True] | B |  |
| 24 | DS1-5 | Multiple choice | Suppose you have the following Pandas dataframe in your Python environment:      You create a grouped dataframe using the following code:    What are the columns that will be included in the resulting dataframe grouped\_df? | [n24\_0\_1](https://worldbankgroup.sharepoint.com/:i:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/images_olc/n24_0_1.png?csf=1&web=1&e=o2f1ff)  [n24\_0\_2](https://worldbankgroup.sharepoint.com/:i:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/images_olc/n24_0_2.png?csf=1&web=1&e=cGEt7K) | A) [1, 2, 3, 4]  B) crop, Price, Quantity  C) Price, Quantity  D) farmerid, Price, Quantity  E) farmerid, crop, Price, Quantity | C |  |
| 31 | DS1-6 | Fill in the blanks | To upload your local commits to a remote repository:  git \_\_\_\_\_\_\_\_\_\_ <remote\_name><branch\_name> |  |  | push |  |
| 42 | DS1-9 | True or False | According to the World Bank’s Personal Data Privacy:  Personal data shall not be protected by any technical or organizational safeguards and freely processed without any restrictions, regardless of authorization |  |  | False |  |
| 43 | DS1-9 | True or False | Restricted information is assigned a securtity classification based on the level of harm posed by unauthorized disclosure |  |  | True |  |
| 46 | DS1-10 | True or False | You were provided with AWS credentials. Since every time you want to access the data you need to provide them, it is a good idea to hard code them in the notebook you are using for your task |  |  | False |  |
| Part 2 – Task Based Questions  Duration: 110 minutes  Cannot navigate backwards  This part will be approved conditional on the review of the material submitted on the final link | | | | | | | |
| **Rubric for the candidate:**  You are provided with a dataset that stores information about different movies like their genre, budget, rating, etc. You are also provided with a second dataset that has the credits for the movies from the first dataset. Load the datasets and answer the following questions.  [Movies,](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h) [Credits](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5)  At the end of the exam, you will be requested to load your notebook/script with the code written to obtain the solutions. The code will be reviewed by the evaluator who will determine whether you pass the exam or not. | | | | | | | |
| 34 | DS1-7 | Fill in the blanks | There are \_\_\_\_\_\_\_\_\_\_ columns in movie dataset and \_\_\_\_\_\_\_\_\_\_ columns in credits dataset. |  |  | 20, 4 | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 35 | DS1-7 | Fill in the blanks | There are total \_\_\_\_\_\_\_\_\_\_ missing values in movie and \_\_\_\_\_\_\_\_\_\_ missing values in credit dataset |  |  | 6851, 0 | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 38 | DS1-7 | Fill in the blanks | What would be the best way to impute missing values in the following columns :  1- vote\_average: \_\_\_\_\_\_\_\_\_\_  2- popularity: \_\_\_\_\_\_\_\_\_\_  Choose between mean, median, mode Note: Provide your answer using lowercase |  |  | 1- mean  2-median | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 39 | DS1-7 | Fill in the blanks | The analysis will continue by only using the records that do not have any missing values in the following columns: budget, popularity, revenue, runtime, vote\_average, vote\_count.  The new data frame has \_\_\_\_ rows |  |  | 2527 | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 3 | DS1-1 | Fill in the blanks | What is the budget threshold or budget value (as an integer) below which 50% of the movie budgets are situated? |  |  | 14000000 | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h) [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 13 | DS1-3 | Fill in the blanks | Complete the following line of code in [a], [b] and [c] to calculate the mean budget by original\_language:    movies.[a] ('original\_language')[b].[c]() Note: Provide your answer using lowercase |  |  | a-groupby  b-budget  c-mean | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 14 | DS1-3 | Fill in the blanks | Which original language had the second highest mean budget expenditure? Note: Provide your answer using lowercase |  |  | ja | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 15 | DS1-3 | Fill in the blanks | The number of movies with the following characteristics is \_\_\_\_\_\_\_:  - Spent less than 500,000  - The runtime is above 120 minutes or below 100 minutes  -The original language is abbreviated as 'fr' or 'es' |  |  | 8 | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 10 | DS1-2 | All that apply | Which plot would you use to study the distribution of the budget variable? |  | A) scatter plot  B) boxplot  C) histogram  D) pie chart | B, C | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 28 | DS1-5 | Fill in the blanks | Combine the movie and credits datasets based on a shared attribute. Fill in the missing value for [a], [b], [c]  combined\_df = movies.[a](credits, left\_on = [b], right\_on = [c], how = ‘left’)  **Continue your analysis using the combined\_df**  Below you can find a description of each dataset’s columns. **The Credit dataset contains the following features:**  **movie\_id** - A unique identifier for each movie.  **Title -** The movie title  **cast** - The name of lead and supporting actors.  **crew** - The name of Director, Editor, Composer, Writer etc.  **The Movie dataset has the following features:**  **budget** - The budget in which the movie was made.  **genres** - The genres of the movie, Action, Comedy ,Thriller etc.  **homepage** - A link to the homepage of the movie.  **id** - This is in fact the movie\_id as in the first dataset.  **keywords** - The keywords or tags related to the movie.  **original\_language** - The language in which the movie was made.  **original\_title** - The title of the movie before translation or adaptation.  **overview** - A brief description of the movie.  **popularity** - A numeric quantity specifying the movie popularity.  **production\_companies** - The production house of the movie.  **production\_countries** - The country in which it was produced.  **release\_date** - The date on which it was released.  **revenue** - The worldwide revenue generated by the movie.  **runtime** - The running time of the movie in minutes.  **spoken\_languages** - languages in the movie  **status** - "Released" or "Rumored".  **tagline** - Movie's tagline.  **title** - Title of the movie.  **vote\_average** - average ratings the movie received.  **vote\_count** - the count of votes received. |  |  | [a]merge  [b]id  [c]movie\_id | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 37 | DS1-7 | Multiple choice | Find the total number of columns after combining the dataframes and by removing redundant columns. |  | A) 22  B) 23  C) 21  D) 24 | A | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 22 | DS1-4 | Multiple choice | Calculate the weighted rating (wr) for the movie  WeightedRating(WR)=[(v.R)/(v+m)] + [(m.C)/(v+m)]  here,  **v** is the number of votes for the movie  **m** is the minimum votes required to be listed in the chart  **R** is the vote\_average of the movie  **C** is the mean vote\_average across the whole report  For calculation of m, we will use 90th percentile as our cutoff. In other words, for a movie to feature in the charts, it must have more votes than at least 90% of the movies in the list.    List the five movies with the highest WR. |  | A) Inception, Forrest Gump, The Lord of the Rings: The Fellowship of the Ring, The Empire Strikes Back, Star Wars  B) The Shawshank Redemption, The Godfather, Interstellar, The Lord of the Rings: The Return of the King, The Empire Strikes Back  C) The Godfather, Inception, Forrest Gump, Interstellar, The Lord of the Rings: The Return of the King  D) Fight Club, Pulp Fiction, The Godfather, Inception, The Empire Strikes Back | B | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 12 | DS1-3 | All that apply | Create the correlation matrix of the dataset and select the appropriate option: |  | A) Vote count has the maximum correlation coefficient with popularity and revenue  B) Strength of relationship for vote average and id is similar to vote average and popularity  C) A correlation coefficient of 0.6 indicates a stronger relationship than a correlation coefficient of –0.7  D) If two variables have a low pearson correlation coefficient they can’t have any relationship with each other | A | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 29 | DS1-5 | Fill in the blank | What was the budget (integer) of the film with a crew of 150 |  |  | 78000000 | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |
| 41 | DS1-8 | All that apply | Which of the following file formats can be used to save the merged dataframe |  | A) png  B) tiff  C) csv  D) xlsx  E) doc | C, D | [M0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_0.csv?d=w3a1f4b08fce7466e851ecde5ee5f0eb8&csf=1&web=1&e=6m2f5h)  [C0](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_0.csv?d=w7c33fa7f879c4168897e76cbccf97195&csf=1&web=1&e=fC1Oj5) |