|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | | | | | | | | | | | | | |  |
| # | 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 | Question choices (if applicable) | Correct answer | Link to dataset (if applicable) | 1. Comments for revision  2. Add 1 “\*” if your top choice | Question – Iteration 1 | Question choices (if applicable) – Iteration 1 | Correct answer – Iteration 1 | Link to dataset (if applicable) | Question – Iteration 2 | Question choices (if applicable) – Iteration 2 | Correct answer – Iteration 2 | Link to dataset (if applicable) | Question – Iteration 3 | Question choices (if applicable) – Iteration 3 | Correct answer – Iteration 3 | Link to dataset (if applicable) | Question – Iteration 4 | Question choices (if applicable) – Iteration 4 | Correct answer – Iteration 4 | Link to dataset (if applicable) |
| 1 | DS1-1 | Multiple choice | What summary statistic is the following Python code attempting to compute? | -Interquartile range  -quartile 75  -first quartile  -third quartile | -third quartile |  |  | What summary statistic is the following Python code attempting to compute? | -Interquartile range  -quartile 75  -first quartile  -third quart | first quartile |  | What summary statistic is the following Python code attempting to compute? | -Standard deviation  -Variance  -Mean  -Kurtosis | -Variance |  | What summary statistic is the following Python code attempting to compute? | -Mean  -Median  -Mode  -Range | -Median |  | What summary statistic is the following Python code attempting to compute? | -Standard Deviation  -Variance  -Kurtosis  -Mean | -Standard Deviation |  |
| 2 | DS1-1 | Multiple choice | From the following visualizations, which appears to contain outliers? | -a  -b  -c  -d | -d |  |  | From the following visualizations, which appears to contain outliers? | -a  -b  -c  -d | c |  | From the following visualizations, which appears to contain outliers? | -a  -b  -c  -d | b |  | From the following visualizations, which appears to contain outliers? | -a  -b  -c  -d | d |  | From the following visualizations, which appears to contain outliers? | -a  -b  -c  -d | a |  |
| 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) |  | What is the budget threshold or budget value (as an integer) below which 50% of the movie budgets are situated? |  | 40000000 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | What is the `vote\_average` for the 10th percentile? |  | 4.9 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | What is the vote\_average that 90% of the movies have a value greater than it? |  | 7.3 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | What is the vote\_average that 10% of the movies have a value lower than it? |  | 4.9 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 4 | DS1-1 | Multiple choice | The following Python code uses Numpy library to generate data of which probability distribution?  data = np.random.uniform(low, high, num\_samples) |  | c |  |  | The following Python code uses Numpy library to generate data of which probability distribution?  data = np.random.poisson(lam, num\_samples) |  | c |  | The following Python code uses Numpy library to generate data of which probability distribution?  data = np.random.normal(loc, scale, num\_samples) |  | a |  | The following Python code uses Numpy library to generate data of which probability distribution?  data = np.random.binomial(trials, success, num\_samples) |  | c |  | The following Python code uses Numpy library to generate data of which probability distribution?  data = np.random.logistic(loc, scale, num\_samples) |  | b |  |
| 5 | DS1-1 | All that apply | Given the distribution of variable x shown below, it is correct to state that: | - Around 50% of x is below 5  - Around 50% of x is above 5  - Around 75% of x is below 10 | - Around 50% of x is below 5  - Around 50% of x is above 5 |  |  | Given the distribution of variable x shown below, it is correct to state that: | - 50% of the datapoints are below 10  - The distribution does not have outliers  - 75% of the datapoints are below 12.75 (approximately)  - The distribution does not seemt o be skewed | - 50% of the datapoints are below 10  - The distribution does not seem to be skewed |  | Given the distribution of variable x shown below, it is correct to state that: | - Variable x does not reach values higher than 10  -There are no outliers in the distribution  - 25% of the data is below 3.6 (approximately)  - The distribution is skewed | - Variable x does not reach values higher than 10  -There are no outliers in the distribution  - 25% of the data is below 3.6 (approximately) |  | Given the distribution of variable x shown below, it is correct to state that: | - The mean of the distribution is 1  - The median of the distribution is 2  - The distribution is right skewed | - The distribution is right skewed |  | Given the distribution of variable x shown below, it is correct to state that: | - The distribution is left skewed  - 10 is the mode of the distribution  - 50% of the data is greater than 10 | - 50% of the data is greater than 10 - The distribution is left skewed |  |
| 6 | DS1-1 | Multiple choice | The following histogram shows the distribution of variable x. What would be the most appropriate label for axis y: | - Variable X  - Frequency of X  - Distribution of X | - Frequency of X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | DS1-1 | Multiple choice | Which variable has the lowest standard deviation? | - x  - y  - z | -y |  |  | Which variable has the highest standard deviation? | - x  - y  - z | -x |  | Which variable has the highest standard deviation? | - x  - y  - z | y |  | Which variable has the lowest standard deviation? | - x  - y  - z | -x |  | Which variable has a mid-standard deviation? | - x  - y  - z | -z |  |
| 8 | DS1-1 | Multiple choice | Which variable from below most likely comes from a normal distribution? | -x  -y  -z | -x |  |  | Which variable from below most likely comes from a uniform distribution? | -x  -y  -z | -z |  | Which variable from below most likely comes from a lognormal distribution? | -x  -y  -z | -y |  | Which variable from below most likely comes from a left-skewed distribution? | -x  -y  -z | -y |  | Which variable from below most likely comes from a bimodal distribution? | -x  -y  -z | -x |  |
| 9 | DS1-2 | True or False | It is useful to create a boxplot of the id feature from the dataset movies. |  | False | [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) |  | It is useful to create a scatter plot between budget and revenue from the dataset movies |  | True | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | It is useful to create a scatter plot between `id` and `budget` from the dataset movies |  | False | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | It is useful to create a scatter plot between ‘id’ and ‘vote\_average’ from the dataset movies? |  | False | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | It is useful to create a scatter plot between `budget` and `vote\_average` from the dataset movies? |  | True | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | Which plot would you use to study the distribution of the original\_language variable? | A) scatter plot  B) pie chart  C) heatmap | B | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Which plot would you use to study the relationship between budget and revenue? | A) scatter plot  B) boxplot  C) pie chart  D) histogram | A | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Which plot would you use to study the distribution of vote\_average? | A) scatter plot  B) boxplot  C) pie chart  D) histogram | A, D | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Which plot would you use to study the distribution of vote\_count? | A) scatter plot  B) boxplot  C) pie chart | B | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 11 | DS1-3 | All that apply | From the options below, which are exploratory data techniques? | - Hypothesis generation  - Outlier detection  - Pivot table  - Grid search  - Machine learning models  - Cross-validation | - Hypothesis generation  - Outlier detection  - Pivot table |  |  | From the options below, which are exploratory data techniques? | - Correlation analysis  - Data visualization  - Summary statistics  - Ensemble learning  - Regularization techniques  - Data processing | - Correlation analysis  - Data visualization  - Summary statistics |  | From the options below, which are exploratory data techniques? | -Summary statistics  - Hypothesis generation  - Pivot Table  - A/B testing  - Grid search  - Cross-validation | -Summary statistics  - Hypothesis generation  - Pivot Table |  | From the options below, which are exploratory data techniques? | -Data visualization  -Summary statistics  - Correlation analysis  -Data preprocessing  -A/B testing  -Data collection | -Data visualization  -Summary statistics  - Correlation analysis |  | From the options below, which are exploratory data techniques? | -Data visualization  - Correlation analysis  - Outlier detection  -Data preprocessing  -Data collection  -Machine learning models | -Data visualization  - Correlation analysis  - Outlier detection |  |
| 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) |  | Create the correlation matrix of the dataset and select the appropriate option: | A) Vote count has the maximum correlation coefficient with popularity and budget    B) A correlation coefficient of 0.6 indicates a stronger relationship than a correlation coefficient of –0.7    C) According to the correlation matrix, it is worth exploring the increase of budget to increase revenue | C | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Create the correlation matrix of the dataset and select the appropriate option: | A) Score is positive and highly correlated with budget  B) We should not be studying the correlation between id and the rest of the movies’ attributes  C) It seems promising to increase the budget in order to increase the score | B | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Create the correlation matrix of the dataset and select the appropriate option: | A) Score is positive and highly correlated with budget  B) There is a linear relationship between vote\_count and revenue  C) A correlation of -0.8 does not imply a relationship between two variables | B | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Create the correlation matrix of the dataset and select the appropriate option: | A) There is a linear relashionship between popularity and runtime  B) Score is positive and highly correlated with budget  C) There is a linear relationship between vote\_count and popularity | A | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | Complete the following line of code in [a], [b] and [c] to calculate the mean budget and mean vote\_average by original\_language:    movies.[a] ( [b] )[['budget', [c]]].mean()  Note: Provide your answer using lowercase |  | a-groupby  b-’original\_language’  c-’vote\_average' | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Complete the following line of code in [a] and [b] to calculate the mean budget by genre:    movies.[a] ('main\_genre')[b].mean()  Note: Provide your answer using lowercase |  | a -groupby  b-budget | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Complete the following line of code in [a] and [b] to calculate the mean budget by genre:    movies.[a] ([b])['budget'].mean()  Note: Provide your answer using lowercase |  | a-groupby  b-main\_genre | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Complete the following line of code in [a] and [b] to calculate the mean budget by genre and original\_language:    movies.[a] ([b])['budget'].mean()  Note: Provide your answer using lowercase |  | a-groupby  b- 'main\_genre', 'original\_language' | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | Which original language had the lowest average vote\_average?  Note: Provide your answer using lowercase |  | ro | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Which genre had the highest mean budget expenditure?  Note: Provide your answer using lowercase |  | Animation | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Which genre had the lowest mean budget expenditure?  Note: Provide your answer using lowercase |  | Foreign | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Which genre and language had the second highest mean budget expenditure? Genre: \_\_\_\_\_ and language \_\_\_\_\_\_  Note: Provide your answer using lowercase |  | Fantasy, ko | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | The number of movies with the following characteristics is \_\_\_\_\_\_\_:  - Spent more than 1,000,000  - The runtime is above 100 minutes or below 60 minutes  - The original language is abbreviated as 'fr' or 'es' |  | 23 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | The number of movies with the following characteristics is \_\_\_\_\_\_\_:  - Spent more than 500,000  - The main genre is Animation  - The runtime is above 120 minutes or below 110 minutes  - Theoriginal language is abbreviated as 'fr' or 'es' |  | 4 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | The number of movies with the following characteristics is \_\_\_\_\_\_\_:  - Spent more than 5,000,000  -The main genre is Fantasy  - Thevote\_average is above 7 or below 5  - IThe original language is abbreviated as 'en' or 'es' |  | 19 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | The number of movies with the following characteristics is \_\_\_\_\_\_\_:  - Spent more than 500,000  - The main genre is Crime  - IThe vote\_average is above 7 or below 5  - Theoriginal language is abbreviated as 'ko' or 'es' |  | 2 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 16 | DS1-4 | Multiple choice | What is the output of the following Python code?  x = [6, 28, 5, 26, 49]  print(x[1:-1]) | - [6, 28, 5, 26, 49]  - [28, 5, 26, 49]  - [28, 5, 26]  - [6, 28, 5, 26] | [28,5,26] — Indexing starts with zero, so the index 1 will start in the second item. The end index -1 means that the last index should be one less than the max index |  |  | What is the output of the following Python code?  x = [8, 9, 6, 10, 12, 14]  print(x[2:4]) | -[9, 6, 10]  -[6, 10]  -[6, 10, 12]  -[9, 6] | [6, 10] |  | What is the output of the following Python code?  x = [8, 9, 12, 14]  print(x[3:]) | -[14]  -[12, 14]  -[] | [14] |  | What is the output of the following Python code?  x = [4, 7, 29, 30, 18, 29]  print(x[3:-1]) | -[30,18]  -[29, 30, 18]  -[29, 30, 18, 29]  -[30, 18, 29] | [30, 18] |  | What is the output of the following Python code?  x = [4, 7, 2, 3, 60, 18, 29]  print(x[:-1]) | [4, 7, 2, 3, 60, 18]  -[4, 7, 2, 3, 60, 18, 29]  - [4, 7, 2, 3, 60] | [4, 7, 2, 3, 60, 18] |  |
| 17 | DS1-4 | Multiple choice | What is the output of the following Python code?  x = list(range(4))  print(x) | - [0, 1, 2, 3, 4]  - [0, 1, 2, 3]  - [4, 4, 4, 4, 4]  - [4, 4, 4, 4] | [0, 1, 2, 3] — Indexing starts with zero and is non-inclusive for the last term. |  |  | What is the output of the following Python code?  x = list(range(10, 15, 2))  print(x) | -[10, 12, 14, 15]  -[10, 12, 14, 16]  - [10, 12, 14] | [10, 12, 14] |  | What is the output of the following Python code?  x = list(range(5, 0, -1))  print(x) | -[0, 1, 2, 3, 4, 5]  -[5, 4, 3, 2, 1, 0]  -[0, 1, 2, 3, 4]  -[5, 4, 3, 2, 1] | [5, 4, 3, 2, 1] |  | What is the output of the following Python code?  x = list(range(7, 1, -3))  print(x) | -[1, 4, 7]  -[7, 4]  -[7, 4, 1]  -[4,7] | [7,4] |  | What is the output of the following Python code?  x = list(range(-6, -1, 2))  print(x) | -[-6, -4, -2]  -[-1, -3, -5]  -[-1, -3, -5, -6]  -[-6, -4, -2, -1] | [-6, -4, -2] |  |
| 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'] | I) x.append('d')  II) x += ['d']  III) x = x.append('d')  IV) x = ['d'] | I and II are correct. II: concatenates the list with the new list [``'``d``'``]. I: append() is an in-place method that modifies the list it is used on.    III and IV are incorrect. III: Since append() is an in-place operator it returns None and x is assigned to what append() returns, i.e. None. IV: Here x is assigned to only be the new list [``'``d``'``] |  |  | Suppose you have defined the following list in your Python environment:  x = [2, 3, 4]  Which of the following syntaxes would make x equal to:  [2, 3, 4, 5] | a- x.append(5)  b- x += 5  c- x = x.append('d')  d- x += [5] | a, d |  | Suppose you have defined the following list in your Python environment:  x = [2, 3, 4]  Which of the following syntaxes would make x equal to:  [1, 2, 3, 4] | a-x.insert(0, 1)  b-x.append(1, 0)  c- x = [1] + x  d- x = 1 + x | a,c |  | Suppose you have defined the following list in your Python environment:  x = [2, 3, 4]  Which of the following syntaxes would make x equal to:  [2, 3, 4, ‘a‘] | a-It is impossible, you cannot add a string to a list of integers  b x.append(‘a’)  c-x = x + [‘a’]  d-x+= ‘a’ | B, c, d |  | Suppose you have defined the following list in your Python environment:  x = [1, 2, 3, 4]  Which of the following syntaxes would make x equal to:  [1, 2, 3] | A) x -= 1 B) x -= [4]  C) x = x[:-1]  D) x = x[0:-1] | C, D |  |
| 19 | DS1-4 | Multiple choice | What would be the output of the following Python code?  print(type(3.0)) | - <class 'int'>  - <class 'float'>  - <class ‘bool’>  - 3  - <class 'str'>  - 3.0 | <class 'float'> — type() returns the type of the object not a value like ( 3 or 3.0 ). Even though the value 3.0 is the same value as the int 3 Python assigns any numeric variable with a decimal to a float. type('3') would have returned <class 'str'> |  |  | What would be the output of the following Python code?  print(type(5)) | - <class 'str'>  - <class 'int'>  - <class 'float'>  - 5  - <class ‘bool’> | <class 'int'> |  | What would be the output of the following Python code?  print(False\*1) | - <class ‘bool’>  - False  - 0  - True  - 1 | 0 |  | What would be the output of the following Python code?  print(type(‘5’)) | - <class 'int'>  - <class 'float'>  - ‘5’  - <class 'str'>  - 3 | <class 'str'> |  | What would be the output of the following Python code?  print(type(‘1’==1)) | - <class ‘bool’>  - False  - True  - <class 'int'> | - <class ‘bool’> |  |
| 20 | DS1-4 | Multiple choice | What is the output of the following Python code? | A) [True, True, False]  B) True  C) [False, False, True]  D) [False, False, True, True]  E) [False, False, True, False]  F) [True, True] | B |  |  | What is the output of the following Python code? | A)[‘a’, True, 1, 1.0]  B) a  C) [False, True, False, False | B |  | What is the output of the following Python code? | A) [2,3,4]  B) [2, 3, 4, 1]  C) None  D) [3, 4, 5] | B |  | What is the output of the following Python code? | A) (2, True)  B) (1, True)  C) [False, (2, True), (3, True), False] | A |  | What is the output of the following Python code? | A) None  B) {1: 2, 2: 'a', 3: 'a', 4: 2}  C) {'a':2, 2:1, 3:1, 4:2} | B |  |
| 21 | DS1-4 | Multiple choice | Given the code below:   How many loops will the code run? | - One  - None  - Infinite  - Ten | correct answer: Infinite. There is no way to stop the while loop since there is no action to change x from True to False. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 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) |  | Calculate the weighted rating (wr) for the movie  WeightedRating(WR)=[(v.R)/(v+m)] + [(m.C)/(v+m)]  where,  **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 95th percentile as our cutoff. In other words, for a movie to feature in the charts, it must have more votes than at least 95% 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) Fight Club, Pulp Fiction, The Godfather, Inception, The Empire Strikes Back    C) Inception, Forrest Gump, The Empire Strikes Back, Schindler's List, Whiplash    D) The Godfather, Inception, Forrest Gump, Interstellar, The Lord of the Rings: The Return of the King | A | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Calculate the weighted rating (wr) for the movie  WeightedRating(WR)=[(v.R)/(v+m)] + [(m.C)/(v+m)]  where,  **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 85th percentile as our cutoff. In other words, for a movie to feature in the charts, it must have more votes than at least 85% 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) Fight Club, Pulp Fiction, The Godfather, Inception, The Empire Strikes Back  D) The Godfather, Inception, Forrest Gump, Interstellar, The Lord of the Rings: The Return of the King | C | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Calculate the weighted rating (wr) for the movie  WeightedRating(WR)=[(v.R)/(v+m)] + [(m.C)/(v+m)]  where,  **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 85th percentile as our cutoff. In other words, for a movie to feature in the charts, it must have more votes than at least 85% of the movies in the list.    List the five movies with the highest WR. | A) Inception, Forrest Gump, The Empire Strikes Back, Schindler's List, Whiplash  B)Inception, Forrest Gump, The Lord of the Rings: The Fellowship of the Ring, The Empire Strikes Back, Star Wars  C) The Shawshank Redemption, The Godfather, Interstellar, The Lord of the Rings: The Return of the King, The Empire Strikes Back  D) The Godfather, Inception, Forrest Gump, Interstellar, The Lord of the Rings: The Return of the King | D | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Calculate the weighted rating (wr) for the movie  WeightedRating(WR)=[(v.R)/(v+m)] + [(m.C)/(v+m)]  where,  **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 85th percentile as our cutoff. In other words, for a movie to feature in the charts, it must have more votes than at least 85% of the movies in the list.    List the five movies with the highest WR. | A) Inception, Forrest Gump, The Empire Strikes Back, Schindler's List, Whiplash  B) The Godfather, Inception, Forrest Gump, Interstellar, The Lord of the Rings: The Return of the King  C) Fight Club, Pulp Fiction, The Godfather, Inception, The Empire Strikes Back  D) Inception, Forrest Gump, The Lord of the Rings: The Fellowship of the Ring, The Empire Strikes Back, Star Wars | B) | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 23 | DS1-5 | Multiple choice | Suppose you have the following Pandas dataframe in your Python environment:      What is the output of the following code? | - 11  - 9  - 18  - 20  - 4  - 5 | 18 — The code subsets the Quantity value of the second observation (index 1) among observations where crop is Maize or Onion |  |  | Suppose you have the following Pandas dataframe in your Python environment:    What is the output of the following code? | - Maize  - Sorghum  - Onion  - [‘Sorghum’, ‘Maize’, ‘Onion’]  - [‘Maize’, ‘Sorghum’, ‘Onion’] | Maize— The code is subsetting observations in rows 3, 2, 4 (in that order), and then getting the crop column of the first row in them |  | Suppose you have the following Pandas dataframe in your Python environment:    What is the output of the following code? | -489  -469  -[469. 12]  -’Sorghum’ | 469 |  | Suppose you have the following Pandas dataframe in your Python environment:    What is the output of the following code? | -[‘Maize’, ‘Sorghum’]  -’Maize’  -[‘Maize’, ‘Sorghum, ‘Onion’]  - [‘Maize’, ‘Sorghum, ‘Maize’] | [‘Maize’, ‘Sorghum’] |  | Suppose you have the following Pandas dataframe in your Python environment:    What is the output of the following code? | -[‘Maize’, ‘Sorghum’]  - [‘Maize’, ‘Sorghum, ‘Maize’]  - ['Onion', 'Sorghum', 'Maize'] | ['Onion', 'Sorghum', 'Maize'] |  |
| 24 | DS1-5 | Multiple choice | Suppose you have the below Pandas dataframe and code snippet. What are the columns that can be included in the resulting dataframe grouped\_df? | A) [1, 2, 3, 4]  B) crop, Price, Quantity  C) Price, Quantity  D) farmerid, Price, Quantity  E) farmerid, crop, Price, Quantity | C |  |  | Suppose you have the below Pandas dataframe and code snippet. How many rows will grouped\_df have? | A) 4  B) 5  C) 7  D) 8 | A |  | Suppose you have the below Pandas dataframe and code snippet. How many rows will grouped\_df have? | A) 4  B) 7  C) 3  D) 5 | B |  | Suppose you have the below Pandas dataframe and code snippet. How many index levels will grouped\_df have? | A) 1  B) 2  C) 4  D) 3 | B |  | Suppose you have the below Pandas dataframe and code snippet. What is the value for index Onion in grouped\_df? | A)[260, 190]  B)[190. 260]  C) 225  D)None | C |  |
| 25 | DS1-5 | Multiple choice | Which column should be selected as the index for this data frame: | - country\_name  - numeric\_code  - continent\_code | numeric\_code |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 26 | DS1-5 | Multiple choice | What should be the data type for the numeric\_code column from the dataset below: | - integer  - float  - string | - string |  |  | What should be the data type for the GDP\_trillon column from the dataset below: | - int  - float  - str  - object | -float |  | What should be the data type for the id column from the dataset below: | -int  -float  -boolean  -str | -str |  | What should be the data type for the has\_return column from the dataset below: | - int  -boolean  -float  -str | -boolean |  | What should be the data type for the price column from the dataset below: | -str  -boolean  -float  -int | int |  |
| 27 | DS1-5 | Multiple choice |  | -Left-join  -Right-join  -Inner-join  -Horizontal stack  -Vertical stack | -Right-join |  |  |  | -Left-join  -Right-join  -Inner-join  -Horizontal stack  -Vertical stack | -Horizontal stack |  |  | -Left-join  -Right-join  -Inner-join  -Horizontal stack  -Vertical stack | -Vertical stack |  |  | -Left-join  -Right-join  -Inner-join  -Horizontal stack  -Vertical stack | -Inner-join |  |  | -Left-join  -Right-join  -Inner-join  -Horizontal stack  -Vertical stack | -Left-join |  |
| 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. |  | Merge, id, 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) |  |  |  |  | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) |  |  |  | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) |  |  |  | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) |  |  |  | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | What was the vote average of the film with a crew of 168? |  | 6.9 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | What was the budget (integer) of the film with a crew of 151 |  | 85000000 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | What was the budget (integer) of the film with a crew of 130 |  | 30000000 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | What was the vote\_average of the film with a crew of 160 |  | 7.2 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 30 | DS1-6 | All that apply | From the options below, which are helpful ways comments can make code more reusable? | -Documentation of functions and methods  -Code reviews  -Error handling  -Cryptic abbreviations  -Overjustification  -Misleading comments | -Documentation of functions and methods  -Code reviews  -Error handling |  |  | From the options below, which are helpful ways comments can make code more reusable? | -Code review  -Version control  -Error handling  -Overjustification  -Duplicate comments  -Misleading comments | -Code review  -Version control  -Error handling |  | From the options below, which are helpful ways comments can make code more reusable? | -Explanation of complex logic  -Documentation of functions and methods  -TODO lists  -Redundancy  -Cryptic abbreviations  -Non-descriptive comments | -Explanation of complex logic  -Documentation of functions and methods  -TODO lists |  | From the options below, which are helpful ways comments can make code more reusable? | -Explanation of complex logic  -TODO lists  -Version control  -Redundancy  -Non-descriptive comments  -Duplicate comments | -Explanation of complex logic  -TODO lists  -Version control |  | From the options below, which are helpful ways comments can make code more reusable? | -Explanation of complex logic  -Code reviews  -Version control  -Cryptic abbreviations  -Non-descriptive comments  -Redundancy | -Explanation of complex logic  -Code reviews  -Version control |  |
| 31 | DS1-6 | Fill in the blanks | To upload your local commits to a remote repository:  git \_\_\_\_\_\_\_\_\_\_ <remote\_name><branch\_name> |  | push |  |  | To show the current condition of your repository, including modified files and staged changes:  git \_\_\_\_\_\_\_\_\_\_ |  | status |  | To initialize a new Git repository in your current directory:  git \_\_\_\_\_\_\_\_\_\_ |  | init |  | To fetch the latest changes from the specified branch on the remote repository and merge them into your current local branch:  git \_\_\_\_\_\_\_\_\_\_ <remote\_name><branch\_name> |  | pull |  | To Download a new repository from a remote server to your local machine:  git \_\_\_\_\_\_\_\_\_\_ clone <repo\_url> |  | clone |  |
| 32 | DS1-7 | Matching | Match the file format with the corresponding Python function:  a- pd.read\_csv(sep=’\t’)  b-open()  c-h5py.File()  d-pickle.load()  1-.txt  2-.h5  3-.pkl  4-.tsv |  | a-4  b-1  c-2  d-3 |  |  | Match the file format with the corresponding Python function:  a-open()  b-pd.read\_excel()  c-h5py.File()  d-pickle.load()  1-xlsx  2-json  3-pkl  4-h5 |  | a-2  b-1  c-4  d-3 |  | Match the file format with the corresponding Python function:  a-open()  b-pd.read\_excel()  c-pd.read\_sas()  d-sqlite3.connect()  1-.sas7bdat  2-.db  3-.xlsx  4-.txt |  | a-4  b-3  c-1  d-2 |  | Match the file format with the corresponding Python function:  a-Image.open()  b-pd.read\_csv(sep=’\t’)  c-open()  d-sqlite3.connect()  1-.jpg  2-.txt  3-.db  4-.tsv |  | a-1  b-4  c-2  d-3 |  | Match the file format with the corresponding Python function:  a-pd.read\_csv(sep=’\t’)  b-pd.read\_sas()  c-open()  d-sqlite3.connect()  1-.sas7bdat  2-.json  3-.db  4-.tsv |  | a-4  b-1  c-2  d-3 |  |
| 33 | DS1-7 | Multiple choice |  | -Dropping missing values  -Median fill  -Mean fill  -Backward fill  -Forward fill | -Dropping missing values |  |  |  | -Dropping missing values  -Median fill  -Mean fill  -Backward fill  -Forward fill | -Backward fill |  |  | -Dropping missing values  -Median fill  -Mean fill  -Backward fill  -Forward fill | -Median fill |  |  | -Dropping missing values  -Median fill  -Mean fill  -Backward fill  -Forward fill | -Mean fill |  |  | -Dropping missing values  -Median fill  -Mean fill  -Backward fill  -Forward fill | -Forward fill |  |
| 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) |  | There are \_\_\_\_\_\_\_\_\_\_ columns in movie dataset and \_\_\_\_\_\_\_\_\_\_ columns in credits dataset. |  | 21,3 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | There are \_\_\_\_\_\_\_\_\_\_ columns in movie dataset and \_\_\_\_\_\_\_\_\_\_ columns in credits dataset. |  | 22,3 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | There are \_\_\_\_\_\_\_\_\_\_ columns in movie dataset and \_\_\_\_\_\_\_\_\_\_ columns in credits dataset. |  | 21,3 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | There are \_\_\_\_\_\_\_\_\_\_ columns in movie dataset and \_\_\_\_\_\_\_\_\_\_ columns in credits dataset. |  | 22,3 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | There are total \_\_\_\_\_\_\_\_\_\_ missing values in movie and \_\_\_\_\_\_\_\_\_\_ missing values in credit dataset |  | 6502, 0 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | There are total \_\_\_\_\_\_\_\_\_\_ missing values in movie and \_\_\_\_\_\_\_\_\_\_ missing values in credit dataset |  | 4830, 0 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | There are total \_\_\_\_\_\_\_\_\_\_ missing values in movie and \_\_\_\_\_\_\_\_\_\_ missing values in credit dataset |  | 4290, 0 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | There are total \_\_\_\_\_\_\_\_\_\_ missing values in movie and \_\_\_\_\_\_\_\_\_\_ missing values in credit dataset |  | 4429, 0 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 36 | 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 |  |  | [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) |  |  |  |  | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) |  |  |  | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) |  |  |  | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) |  |  |  | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | Find the total number of columns after combining the dataframes and by removing redundant columns. | A)22  B) 23  C) 21  D) 24 | B | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Find the total number of columns after combining the dataframes and by removing redundant columns. | A)22  B) 23  C) 21  D) 24 | D | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Find the total number of rows after combining the dataframes and removing the rows that did not get any information about credits.  Remove these records from the analysis | A)4489  B)4480  C)4450  D)4459 | B | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Find the total number of rows after combining the dataframes and removing the rows that did not get any information about credits.  Remove these records from the analysis | A)4489  B)4480  C)4357  D)4365 | C | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | What would be the best way to impute missing values in the following columns :  1- runtime: \_\_\_\_\_\_\_\_\_\_  2- revenue: \_\_\_\_\_\_\_\_\_\_  Choose between mean, median, mode |  | 1- median  2-median | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | What would be the best way to impute missing values in the following columns :  1- runtime: \_\_\_\_\_\_\_\_\_\_  2- budget: \_\_\_\_\_\_\_\_\_\_  Choose between mean, median, mode |  | 1- median  2-median | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | What would be the best way to impute missing values in the following columns :  1- vote\_average: \_\_\_\_\_\_\_\_\_\_  2- vote\_count: \_\_\_\_\_\_\_\_\_\_  Choose between mean, median, mode |  | 1- mean  2-median | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | What would be the best way to impute missing values in the following columns :  1- budget: \_\_\_\_\_\_\_\_\_\_  2- popularity: \_\_\_\_\_\_\_\_\_\_  Choose between mean, median, mode |  | 1- median  2-median | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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) |  | 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 |  | 2734 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | 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 |  | 4022 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | 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 |  | 4489 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | 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 |  | 4365 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 40 | DS1-7 | Fill in the blanks | Find the total number of columns after combining the dataframes and by removing redundant columns. |  | 22 | [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) |  | Find the total number of columns after combining the dataframes and by removing redundant columns. |  | 23 | [M1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_1.csv?d=w2f4610778fbd4a93a3bc2fa1826d7ab0&csf=1&web=1&e=l6IkWp)  [C1](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_1.csv?d=w3a3aea5d9b9343ba846a3df577db3852&csf=1&web=1&e=zj2B74) | Find the total number of columns after combining the dataframes and by removing redundant columns. |  | 24 | [M2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_2.csv?d=w59cf9b4777884079a73ef5e33b47b6e5&csf=1&web=1&e=pQ9SiD)  [C2](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_2.csv?d=wd4b699f804ad4aa3adf5604517ec904e&csf=1&web=1&e=tMGGNb) | Find the total number of rows after combining the dataframes and removing the rows that did not get any information about credits. Remove these records from the analysis |  | 4480 | [M3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_3.csv?d=wafc95aab49dd477d8f87d52120d14004&csf=1&web=1&e=ehvJeq)  [C3](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_3.csv?d=w5c21e77f92e240dd879cc0066fcdd3c0&csf=1&web=1&e=O6ZKnv) | Find the total number of rows after combining the dataframes and removing the rows that did not get any information about credits. Remove these records from the analysis |  | 4357 | [M4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/movies_iter_4.csv?d=w8eb11c0ab8c7410d927e141ed6fcf940&csf=1&web=1&e=7nR2Tc)  [C4](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/credits_iter_4.csv?d=w137f1d92acf848ef806753ce2c88561e&csf=1&web=1&e=zR40Fl) |
| 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 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 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 |  |  | According to the World Bank’s Personal Data Privacy:  Personal data shall be transferred to third parties without regard of legitimacy and with no consideration for the protection of personal data |  | False |  | According to the World Bank’s Classification and Control Policy:  Strictly Confidential, Confidential and Official Use Only are the only information classification categories available |  | True |  | According to the World Bank’s Data Quality Assurance:  Verifiability, Transparency, and Relevance are Data Quality Principles |  | True |  | According to the World Bank’s Quality Assurance:  Data quality assurance is a shared responsibility of all staff involved in the production of Bank Development Data |  | True |  |
| 43 | DS1-9 | True or False | Restricted information is assigned a security classification based on the level of harm posed by unauthorized disclosure |  | True |  |  | Restricted information is assigned a security classification based on restrictions imposed by the Information Provider or Originator |  | True |  | Restricted information is assigned a security classification based on the document file extension |  | False |  | Restricted information is assigned a security classification based on the extension of the document |  | False |  | Restricted information is assigned a security classification based on the date the information was originated |  | False |  |
| 44 | DS1-9 | True or False | Access to information classified as "Official Use Only" must be restricted to the World Bank Group staff of the relevant World Bank Group entity |  | True |  |  | Documents classified as "Official Use Only" can be uploaded to a public GitHub repository |  | False |  | You received a document classified as “Confidential”. You consider that your colleague also needs to have access to this document to perform the task you are working on. You can share the document with your colleague. |  | True |  | You received a document classified as “Strictly” confidential. You consider that your colleague also needs to have access to this document to perform the task you are working on. You can share the document with your colleague. |  | False |  | Information classified as “Strictly Confidential” should not be forwarded to anyone without access, and must only be shared if there is explicit permission from the Information Provider or Originator. |  | True |  |
| 45 | DS1-9 | Multiple choice | Which classification would you give to a Mission Announcement Letter that informs a client government about an upcoming mission including information about the mission’s composition purpose and counterparts | -Official Use Only  -Confidential  -Strictly Confidential | Official Use Only |  |  | Which classification would you give to an internal report for a project in the Philippines that includes legal analysis provided by the World Bank’s legal department | -Official Use Only  -Confidential  -Strictly Confidential | -Official Use Only |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 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 |  |  | You were provided with AWS credentials. In order to keep them safe, it is a good idea to store them as environmental variables |  | True |  | It is not necessary to have disk encryption if your computer already has a password for logging in |  | False |  | If you lose your key to decrypt your disk, you lose all the information that was stored in it. |  | True |  | Your colleague has not received their AWS credentials yet. It is fine to share yours with her because you are working on the same project |  | False |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-1 | True or False | You want to study the differences in productivity of manufacturing workers in publicly vs. privately owned firms. You have access to a sample of 200 workers. 100 of them work for private firms, and their productivity is 74.3 parts per hour, with a standard deviation of 16 parts per hour. The remaining 100 have a productivity of 69.7 parts per hour and a standard deviation of 18 parts per hour. Assuming the unknown variances are equal, assess the truth/falsehood of the following statement:  The difference in mean productivity is significant at the 10% level. |  | True |  |  | You want to study the differences in productivity of manufacturacturing workers in publicly vs. privately owned firms. You have access to a sample 200 workers. 100 of them work for private firms, and their productivity is 74.3 parts per hour, with a standard deviation of 16 parts per hour. The remaining 100 have a productivity of 69.7 parts per hour and a standard deviation of 18 parts per hour. Assuming the unknown variances are equal, assess the truth/falsehood of the following statement:    The difference in mean productivities is signifficant at the 5% level |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-1 | Multiple Choice | You want to study the differences in productivity of manufacturing workers in publicly vs. privately owned firms. You have access to a sample of 200 workers. 100 of them work for private firms, and their productivity is 74.3 parts per hour, with a standard deviation of 16 parts per hour. The remaining 100 have a productivity of 69.7 parts per hour and a standard deviation of 18 parts per hour. Assuming the unknown variances are equal, what is the significance level? | 0.028  0.056  0.036  0.092  0.12 | .56 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-1 | True or False | Suppose H0 : μ = μo is rejected in favor of  H1 : μ <> μo at the α = 0.05 level of signiﬁcance.  Then, H0 will be rejected at the α = 0.01 level of signiﬁcance. |  | False |  |  | Suppose H0 : μ = μo is rejected in favor of H1 : μ <> μo at the α = 0.01 level of signiﬁcance.  Then, H0 will be rejected at the α = 0.05 level of signiﬁcance. |  | True |  | Suppose H0 : μ = μo is rejected in favor of  H1 : μ <> μo at the α = 0.05 level of signiﬁcance.  Then, H0 will be rejected at the α = 0.1 level of signiﬁcance. |  | True |  | Suppose H0 : μ = μo is rejected in favor of  H1 : μ <> μo at the α = 0.1 level of signiﬁcance.  Then, H0 will be rejected at the α = 0.05 level of signiﬁcance. |  | False |  | Suppose H0 : μ = μo is rejected in favor of  H1 : μ <> μo at the α = 0.1 level of signiﬁcance.  Then, H0 will be rejected at the α = 0.01 level of signiﬁcance. |  | False |  |
|  | DS2-1 | Multiple Choice | Given the following table, pick the option that best interprets the regression results | a) Estimates for the coefficients of variables POP and GNP have the highest probability of being significant  b) If the true parameters for GNP, POP, and GNPDEFL were zero, it is very likely that an OLS regression would render the estimates we see in this table  c) The model has low explanatory power for the variance of the dependent variable  d) The variance-covariance matrix is robust to multicollinearity | b |  |  | Given the following table, pick the option that best interprets the regression results | a) There is an 82.6% probability of the estimate for coefficient POP to be zero  b) This model indicates that every time TOTEMP variable increases ~15, GNPDELF increases 1  c) It is better to assess this model with the R-squared rather than the Adj R-squared  d) The estimate for the coefficient for ARMED variable is not statistically significant at a level of 0.05 | a |  | Given the following table, pick the option that best interprets the regression results | a) The estimate for coefficient for variable YEAR is not statistically significant at a level of 0.05 b)When the year increases 1 unit, TOTEMP increases ~ 1800 units c)The estimate for ARMED variable should not be considered because their p-value is lower than 0.01  d) It is better to assess this model with the R-squared rather than the Adj R-squared | b |  | Given the following table, pick the option that best interprets the regression results | a) It is better to assess this model with the R-squared rather than the Adj R-squared b) This model indicates that as UNEMP increases, TOTEMP increases c) The estimate for coefficient for variable YEAR is not statistically significant at a level of 0.05 d)The probability of the estimate for coefficient UNEMP to be zero is lower than 5% because zero is not contained in the confidence interval [0.025, 0.975] | d |  | Given the following table, pick the option that best interprets the regression results | a) This model is better assessed by Adj R-squared rather than R-squared  b) This model indicates that as UNEMP increases, TOTEMP increases c) The estimate for coefficient for variable YEAR is not statistically significant at a level of 0.05  d)There is a 68.7% probability that the estimate for coefficient for variable GNP to be zero | a |  |
|  | DS2-1 | Multiple Choice | Which of the following statements about Type I is true? | A) Type I error occurs when the null hypothesis is not rejected when it is true.  B) Type I error occurs when the null hypothesis is rejected when it is true.  C) Type I error occurs when the null hypothesis is not rejected when it is false. | B |  |  | Which of the following statements about Type II is true? | A) Type II error occurs when the null hypothesis is rejected when it is true.  B) Type II error occurs when the null hypothesis is rejected when it is true.  C) Type II error occurs when the null hypothesis is not rejected when it is false. | C |  | What is the p-value in hypothesis testing? | A) The probability of making a Type II error  B) The probability of observing the sample result, given that the null hypothesis is true  C) The probability of rejecting the null hypothesis when it is actually true  D) The probability of observing a test statistic at least as extreme as the one calculated, assuming the null hypothesis is true | D |  | What is the p-value in hypothesis testing? | A) The p-value is the smallest alpha at which we can reject H0  B) The probability of making a Type II error  C) The probability of making a Type I and II error | A |  |  |  |  |  |
|  | DS2-2 | Multiple choice | What is the primary purpose of differencing in time series analysis? | a) To smooth out irregularities in the data  b) To make the series stationary  c) To identify seasonal patterns  d) To eliminate outliers from the dataset | b) |  |  | What is the effect of over-differencing a time series? | a) It introduces unnecessary complexity and potential for overfitting  b) It leads to the loss of long-term trend information  c) It causes the series to become non-stationary  d) It enhances the detection of seasonal patterns | b) |  | How can you determine the appropriate order of differencing for a time series? | a) By repeatedly differencing until the autocorrelation function (ACF) shows no significant lags  b) By observing the seasonal decomposition of the time series  c) By minimizing the sum of squared errors (SSE)  d) By comparing the residuals of multiple models with different orders | a) |  | Which of the following tests is commonly used to check for stationarity after differencing a time series? | a) Dickey-Fuller test  b) Ljung-Box test  c) Granger causality test  d) Breusch-Godfrey test | A) |  | What happens if you apply first-order differencing to a time series that is already stationary? | a) The series will become stationary  b) The series will show signs of over-differencing  c) The series will remain unchanged  d) The series will exhibit periodicity | B) |  |
|  | DS2-2 | Multiple choice | Which of the following formulas represents a time series model in first differences? | A  B  C  D | a |  |  |  | a) It represents the current value of the series    b) It is the first difference of the lagged series    c) It captures the seasonal component    d) It models the noise in the series | b |  | What is the main difference between the models | a) The first model is for non-stationary series, while the second is for stationary series  b) The first model includes a seasonal component, while the second does not  c) The first model is an autoregressive model, while the second is a moving average model  d) The first model does not involve differencing, while the second involves first differencing | d |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Multiple choice | Given the following Autocorrelation and Partial Autocorrelation functions, which of the following would be a good model to use for fitting the data according to the Box-Jenkins method? | a) ARMA(1, 1) b) ARMA(0, 2) c) ARMA(0, 1) d) ARMA (2, 0) | d) |  |  | Given the following Autocorrelation and Partial Autocorrelation functions, which ARIMA model would be appropriate for fitting the data according to the Box-Jenkins method? | a) ARIMA(1, 1, 0)  b) ARIMA(0, 1, 1)  c) ARIMA(1, 0, 1)  d) ARIMA(2, 1, 0) | b |  | Considering the displayed ACF and PACF plots, which of the following ARMA models is likely to be the best fit according to Box-Jenkins methodology? | a) ARMA(1, 1)  b) ARMA(0, 2)  c) ARMA(2, 2)  d) ARMA(1, 0) | a) |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | True or False: | The following time series plot represents a trend component. |  | True |  |  | The following time series plot represents seasonality. |  | False |  | The following time series plot represents cyclicity |  | True |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Multiple choice | Given the following code, what type of model are we looking to estimate? | a) ARIMA(0,0,1)×(2, 1, 0)12 b) ARIMA(2, 1, 0)×(0, 0, 1)12 c) ARIMA(1, 1, 1)×(0, 0, 0)0 d) ARIMA(2, 1, 0)×(0, 1, 1)12 | b) |  |  | What type of ARIMA model is specified by the following code? | a) ARIMA(1,1,1)  × (1,1,0)12      b) ARIMA(0,1,1)  × (1,1,0)12      c) ARIMA(1,1,0)  × (1,1,1)12      d) ARIMA(1,1,1)  × (0,1,0)12 | a |  | Based on the code snippet below, what ARIMA model are we estimating? | a) ARIMA(2,1,1)  × (1,1,0)12    b) ARIMA(1,1,1)  × (2,0,1)12    c) ARIMA(2,1,0)  × (1,1,1)12    d) ARIMA(1,1,1)  × (1,1,0)12 | a |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Multiple choice | For a time series, we have conducted an augmented Dickey-Fuller test, and we get the following results. Which of the following options are valid interpretations of this result if we are working with a 5% level of confidence? | a) We cannot reject H0 at this significance level b) We may reject H0 at this significance level c) We consider that the time series does not contain a unit root d) We consider that the time series contains a unit root | a) and d) |  |  | After performing an augmented Dickey-Fuller test on a time series, we obtain the following results. Which interpretation is correct at a 5% significance level? | a) The null hypothesis cannot be rejected at this significance level  b) The null hypothesis can be rejected at this significance level  c) The time series is considered to have a unit root  d) The time series is considered to be stationary | A, c |  | Given the results of an augmented Dickey-Fuller test for a time series, what is the appropriate conclusion at the 5% confidence level? | a) Reject the null hypothesis at the 5% significance level  b) Fail to reject the null hypothesis at the 5% significance level  c) The time series has a unit root  d) The time series is stationary | b, c |  | An augmented Dickey-Fuller test was conducted, and the following results were obtained. What can be inferred from these results at the 5% significance level? | a) We reject the null hypothesis  b) We fail to reject the null hypothesis  c) The time series is non-stationary  d) The time series is stationary | B,c |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Create residuals of the values and check if they are normally distributed |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Given the results of Q1, take another differencing order to validate that the values are indeed stationary |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Create a 10-year moving average |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Given:  Calculate the following trend and seasonality components: |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Calculate predictions using Naive method: |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Calculate predictions using Average method: |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Calculate predictions using Drift method: |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Calculate predictions using SES method: |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Calculate the mean squared error for each of the previous model and pick the best performing |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-2 | Upload notebook | Forecast the next 10-years of population in the largest city. Does it pass the 20M mark? |  |  | tyler |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice | Which of the following graphs represents a ReLU activation function? |  | c) |  |  | Which of the following graphs represents a Sigmoid activation function? |  | a |  | Which of the following graphs represents a Tanh activation function? |  | b |  | Which of the following graphs represents a Leaky Relu activation function? |  | d |  |  |  |  |  |
|  | DS2-3 | Multiple choice | What is true about the following graphs and what they represent about regularization methods? | a) Both graphs represent L2 regularization with different lambda parameters b) The graph on the left represents L1 regularization and the graph on the right represents L2 regularization c) Both graphs represent special cases of Elastic Net regularization d) Neither represents a regularization method. Instead, they represent different approaches to gradient descent. e) b) and c) | e) |  |  | What is true about the following graphs and what they represent about regularization methods? | A) Both graphs illustrate the Lasso regression method, with different constraints.  B) The left graph illustrates Lasso regression, while the right graph illustrates Ridge regression.  C) Both graphs illustrate the Ridge regression method, with different constraints.  D) The left graph illustrates Ridge regression, while the right graph illustrates Lasso regression. | b |  | What is true about the following graphs and what they represent about regularization methods? | A) The left graph shows an L1 norm constraint, and the right graph shows an L2 norm constraint.  B) The left graph shows an L2 norm constraint, and the right graph shows an L1 norm constraint.  C) Both graphs show L1 norm constraints but applied differently.  D) Both graphs show L2 norm constraints but applied differently. | a |  | How do the constraints in the graphs affect the estimated coefficients 𝛽 ? | A) Lasso (left graph) tends to produce sparse solutions by setting some coefficients exactly to zero, while Ridge (right graph) shrinks coefficients but does not set them to zero.  B) Both Lasso and Ridge shrink coefficients without setting any coefficients to zero.  C) Ridge (left graph) tends to produce sparse solutions by setting some coefficients exactly to zero, while Lasso (right graph) shrinks coefficients but does not set them to zero.  D) Both Lasso and Ridge set all coefficients to zero. | A |  | What do the red ellipses and blue shapes represent in the context of regularization methods? | A) The red ellipses represent the constraint regions, while the blue shapes represent the contours of the residual sum of squares (RSS).  B) The blue shapes represent the constraint regions, while the red ellipses represent the contours of the residual sum of squares (RSS).  C) The red ellipses represent the solution paths of the coefficients, while the blue shapes represent the possible values of the residual sum of squares (RSS).  D) The blue shapes represent the solution paths of the coefficients, while the red ellipses represent the possible values of the residual sum of squares (RSS). | b |  |
|  | DS2-3 | Fill in the blanks | The following decision tree shows the salary earned by a baseball player using the RBI (runs batted in), putouts, years playing, walks, runs and hits.  How many terminal nodes does the following tree contain? |  | 12 |  |  | The following decision tree shows the salary earned by a baseball player using the RBI (runs batted in), putouts, years playing, walks, runs and hits.  Which is the salary of a player that had played for 3 years, has 63 RBIs, has 60 Walks, 50 runs, 100 putouts and 50 hits? |  | 5.394 |  | The following decision tree shows the salary earned by a baseball player using the RBI (runs batted in), putouts, years playing, walks, runs and hits.  In how many parts is the space being divided? |  | 12 |  | The following decision tree shows the salary earned by a baseball player using the RBI (runs batted in), putouts, years playing, walks, runs and hits.  Which is the salary of a player that had played for 10 years, has 105 RBIs, 63 walks, 98 runs, 101 hits and 78 putouts? |  | 6.407 |  | The following decision tree shows the salary earned by a baseball player using the RBI (runs batted in), putouts, years playing, walks, runs and hits.  Which is the salary of a player that had played for 15 years, has 101 RBIs, 35 walks, 107 runs, 91 hits and 87 putouts? |  | 5.571 |  |
|  | DS2-3 | Multiple choice | In decision trees, what is the purpose of pruning, and how does it help improve model performance? | A) Pruning removes noisy data points from the training set, improving the decision tree's accuracy. B) Pruning reduces the complexity of the decision tree by removing unnecessary branches, preventing overfitting and improving generalization to unseen data. C) Pruning adjusts the learning rate of the decision tree algorithm, enabling faster convergence to the optimal solution. D) Pruning increases the depth of the decision tree, allowing it to capture more complex patterns in the data. | B) |  |  | You have a dataset to predict school choice based on explanatory variables like tuition, number of students, rate, etc. Before applying machine learning algorithms to this dataset, how would you ensure that all features contribute equally to the model and are on a similar scale, considering the different measurement units they represent? | A) Apply a logarithmic transformation to all numerical features    B) Convert the ‘budget’ feature to categorical bins based on arbitrary ranges (e.g. low, medium, high)    C) Convert the ‘budget’ feature to categorical bins based on arbitrary ranges (e.g. low, medium, high)    D) Standardize the numerical features using Z-score scaling | D |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice | After preparing our dataset for predicting weighted ratings based on specific features, what is the final shape of the dataset following these steps?    **Step 1** : x variables would be   * budget * genres * vote\_count * vote\_average   **Step 2** : Create a new column named 'top\_genres' by extracting the first value from each genre.  **Step 3** : Eliminate redundant columns like 'genres' and 'score'. (prefilled in colab)  **Step 4** :Perform encoding based on variable types to achieve the most meaningful representation.  What is the resulting shape of this refined dataset? | -(4803, 23)  -(4803, 4)  -(4803, 10) |  | [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) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Fill in the blanks | You are provided with a train and a test dataset with 13 variables. M is the dependent variable and the rest of them are the independent ones we want to use to predict the behavior of M.  Using the Python package sklearn, create a random forest with 100 trees, a random state of 0 and 6 as the maximum number of variables to pick the best split in each node.  The test MSE for this model is (provide integer part only) |  | 20 | [train\_0.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_0.csv?d=wc450737b3fa6407bad530cc7ae98ea83&csf=1&web=1&e=xxJWhI)  [test\_0.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_0.csv?d=w90a51726b39d42f99d768321f0732fd8&csf=1&web=1&e=hd6n6w) |  | You are provided with a dataset to predict whether a client will subscribe to term deposit or not. Considering that all the variables are of interest for predicting the dependent variable y, use the Python package sklearn, create a random forest with 100 trees, a random state of 42 and 6 as the maximum number of variables to pick the best split in each node.  The test MSE for this model is (Provide integer part only): |  | 0 | [train\_1.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_1.csv?d=w68a3e9a70a844cb08152a621dc62ded4&csf=1&web=1&e=jNzMjJ)  [test\_1.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_1.csv?d=w734f7986d5ec4d43a6d33d8b37b1a148&csf=1&web=1&e=Vzk3ZB) | You are provided with a train and a test dataset with 13 variables. target is the dependent variable and the rest of them are the independent ones we want to use to predict the behavior of target. Using the Python package sklearn, create a random forest with 95 trees, a random state of 56 and 5 as the maximum number of variables to pick the best split in each node. The test MSE for this model is (provide integer part only): |  | 11712584 | [train\_2.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_2.csv?d=wdd7c398227aa4f19beba14db4482721c&csf=1&web=1&e=lFdKFt)  [test\_2.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_2.csv?d=w803e446405c6436f9ccf8b9193903a70&csf=1&web=1&e=HWETDo) | You are provided with a train and a test dataset with 15 variables. target is the dependent variable and the rest of them are the independent ones we want to use to predict the behavior of target. Using the Python package sklearn, create a random forest with 57 trees, a random state of 1 and 7 as the maximum number of variables to pick the best split in each node. The test MSE for this model is (provide integer part only): |  | 46792148 | [train\_3.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_3.csv?d=wde48a269e50b441983d36e80a3c212af&csf=1&web=1&e=oTr6SE)  [test\_3.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_3.csv?d=w04451e32172f45c1b16ddce6cf352e8f&csf=1&web=1&e=1P0F6T) | You are provided with a train and a test dataset with 15 variables. target is the dependent variable and the rest of them are the independent ones we want to use to predict the behavior of target. Using the Python package sklearn, create a random forest with 102 trees, a random state of 3 and 6 as the maximum number of variables to pick the best split in each node. The test MSE for this model is (provide integer part only): |  | 68094 | [train\_4.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_4.csv?d=w5a286110637d4bc7a42b78f3201a6569&csf=1&web=1&e=2YQWEL)  [test\_4.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_4.csv?d=wb9d3d0d5778744259df83576197604ec&csf=1&web=1&e=aAUeHQ) |
|  | DS2-3 | Fill in the blanks | \_\_\_\_\_\_ is the variable with the greatest feature importance (Use capital letters for your answer) |  | L | [train\_0.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_0.csv?d=wc450737b3fa6407bad530cc7ae98ea83&csf=1&web=1&e=xxJWhI)  [test\_0.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_0.csv?d=w90a51726b39d42f99d768321f0732fd8&csf=1&web=1&e=hd6n6w) |  | \_\_\_\_ is the variable with the greatest feature importance (Use capital letters for your answer) |  | DURATION | [train\_1.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_1.csv?d=w68a3e9a70a844cb08152a621dc62ded4&csf=1&web=1&e=jNzMjJ)  [test\_1.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_1.csv?d=w734f7986d5ec4d43a6d33d8b37b1a148&csf=1&web=1&e=Vzk3ZB) | \_\_\_\_ is the variable with the greatest feature importance (Use capital letters for your answer) |  | FEATURE8 | [train\_2.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_2.csv?d=wdd7c398227aa4f19beba14db4482721c&csf=1&web=1&e=lFdKFt)  [test\_2.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_2.csv?d=w803e446405c6436f9ccf8b9193903a70&csf=1&web=1&e=HWETDo) | \_\_\_\_ is the variable with the greatest feature importance (Use capital letters for your answer) |  | FEATURE2 | [train\_3.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_3.csv?d=wde48a269e50b441983d36e80a3c212af&csf=1&web=1&e=oTr6SE)  [test\_3.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_3.csv?d=w04451e32172f45c1b16ddce6cf352e8f&csf=1&web=1&e=1P0F6T) | \_\_\_\_ is the variable with the greatest feature importance (Use capital letters for your answer) |  | FEATURE2 | [train\_4.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_4.csv?d=w5a286110637d4bc7a42b78f3201a6569&csf=1&web=1&e=2YQWEL)  [test\_4.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/test_4.csv?d=wb9d3d0d5778744259df83576197604ec&csf=1&web=1&e=aAUeHQ) |
|  | DS2-3 | Fill in the blanks | Using the train dataset generate a linear regression model to predict variable M. Consider that the rest of the available variables might have explanatory powder for predicting M.  What is the estimate for the coefficient for variable A (provide 2 decimal)? \_\_\_\_\_\_ |  | -0.12 | [train\_0.csv](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_0.csv?d=wc450737b3fa6407bad530cc7ae98ea83&csf=1&web=1&e=xxJWhI) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice | How much of the variance present in M is being explained by this model? | 0.761  90.73  2.48e-98  2095. | 0.761 | [train](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train.csv?d=w7f1c5dbe6cfc47f5ae938e329205a9d2&csf=1&web=1&e=Teu2t7) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | True or False | J is not statistically significant for alpha equals 0.05 |  | False | [train](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train.csv?d=w7f1c5dbe6cfc47f5ae938e329205a9d2&csf=1&web=1&e=Teu2t7) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Fill in the blanks | What is the 95% confidence interval for D (2 decimals, truncated): left bound: (a)\_\_\_\_\_\_\_ right bound: (b)\_\_\_\_\_\_\_ |  | (a) 0.67  (b) 4.66 | [train](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train.csv?d=w7f1c5dbe6cfc47f5ae938e329205a9d2&csf=1&web=1&e=Teu2t7) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Fill in the blanks | You are provided with a subset of World Development Indicators for a subset of countries. You are requested to calculate the PCA (using standardized version of the variables) and find out how many PCAs you will need to explain at least, 60% of the variance in the data.  Use sklearn package and a random state of 0 The number of PCAs to explain at least 60% of the variance in the dataset is \_\_\_\_ |  | 3 | [pca](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/data_for_pca.csv?d=w61851167f5ee4c00b6ce57648a22ec10&csf=1&web=1&e=uLa5ck) |  | You want to reduce its dimensionality using Principal Component Analysis (PCA). Your goal is to find the minimum number of principal components that can capture at least 91% of the total variance in the dataset. |  | 21 | [train\_balanced](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_balanced.csv?d=w182a06d1c8ea4b329df99f29c8cc9b86&csf=1&web=1&e=vhLzr2) |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice | What are the first 4 PCAs scores for Argentina? | a) [ 4.32333729, 0.28848278, -0.64094878, 1.57629295]  b) [-0.59162819, -0.74478837, 1.89450388, 0.1892842 ]  c) [-0.48524025, -0.46921854, 0.50755109, -0.70924373]  d) [ 0.66612359, -0.32830627, 0.11618753, -0.06549968] | c | [pca](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/data_for_pca.csv?d=w61851167f5ee4c00b6ce57648a22ec10&csf=1&web=1&e=uLa5ck) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice | Cluster the data with K-means and the PCA results as features. Use sklearn, random state 0 and 3 clusters.  We can say that Chile (CHL) is in the same group as: | a) UZB b) GNQ c) NAM d)ZAF | a) | [pca](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/data_for_pca.csv?d=w61851167f5ee4c00b6ce57648a22ec10&csf=1&web=1&e=uLa5ck) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | All that apply | We are trying to predict whether the client has subscribed to term deposit? Which modelling algorithm could be used to solve this problem. | a) Random Forest Regressor  b) Random Forest Classifier  c) K nearest neighbour  d) neural networks  e) logistic regression | b,c,de | anjali |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | All that apply | Is the dataset balanced? If not, what kind of problems might occur due to imbalance? | A) The dataset is balanced and accuracy would be the right metric to evaluate the performance of the model  B) The model may achieve high accuracy on the majority class but poor accuracy on the minority class.  C) The model will always perform equally well on both the majority and minority classes.  D) Evaluation metrics like accuracy may not provide a clear picture of the model's performance.  E) The model may struggle to generalize well to new, unseen data. | b,d,e |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | All that apply | How can you solve the imbalance in the given dataset ? | A) Oversampling the minority class using techniques like Random Oversampling or SMOTE (Synthetic Minority Over-sampling Technique).  B) Undersampling the majority class by randomly removing instances to balance the class distribution.  C) Adjusting the class weights during training to give more importance to the minority class.  D) Applying data normalization techniques to scale the features.  E) Using a smaller learning rate during model training. | a,b,c |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | ??? | Read the balanced train dataset. Mention the variables using the following encodings | 1) OneHotEncoding  2) LabelEncoding  3) Sin Cos Encoding | Answer missing | [train\_balanced](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train_balanced.csv?d=w182a06d1c8ea4b329df99f29c8cc9b86&csf=1&web=1&e=vhLzr2) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice |  | a) 20  b) 23  c) 30  d) 21 | d |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | Multiple choice |  | Options are missing |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-3 | All that apply | Based on the classification report, which of the following is correct | A) The model has successfully learned the underlying patterns in the data and generalizes well.    B) The validation set may not be representative of the true data distribution or is too similar to the training set.    C) The binary classification problem could be relatively simple, and the available features are highly informative.    D) The model has been underfit, and increasing its complexity will likely improve its performance. | a,b,c |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Multiple choice | What type of iterable will the following function output? | a) A string b) A list c) A generator d) A tuple | c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Multiple choice | In Python, what is the key difference between a list comprehension and a generator expression, and when might you prefer one over the other? | a) List comprehension is more memory-efficient than generator expression, suitable for smaller datasets. b) List comprehension constructs the entire list in memory, while generator expressions produce elements on-the-fly, conserving memory, which is advantageous for larger datasets and infinite sequences. c) List comprehension is used for one-dimensional data structures, while generator expressions are used for multidimensional data structures. d) List comprehension is parallelizable, while generator expressions are not, making them suitable for distributed computing tasks. | b) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Multiple choice | In the context of parallel computing in R, what is the primary advantage of using the foreach package over other parallelization approaches? | a) foreach provides lower-level control over parallel execution, allowing for finer optimization.  b) foreach is specifically designed for distributed computing across multiple machines in a cluster environment.  c) foreach abstracts away the complexity of parallel programming by allowing users to parallelize loops across multiple processors or nodes with minimal code changes.  d) foreach offers better compatibility with GPU acceleration compared to other parallelization frameworks. | c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Multiple choice | In the context of RESTful APIs, what is the significance of HTTP status codes, and how are they used in error handling? | a) HTTP status codes are optional and are primarily used for logging purposes rather than error handling.  b) HTTP status codes are standardized across all APIs and have the same meanings universally.  c) HTTP status codes indicate the success or failure of API requests and provide additional context about the nature of errors encountered, facilitating robust error handling and recovery mechanisms.  d) HTTP status codes are specific to each API and may vary in meaning and interpretation. | c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Multiple choice | What is the dimension of the following NumPy array? | - 2  - 5  - 4  - 6  - 3 | 3 — We have a list of numbers [5, 5] (first dimension), repeated twice (second dimension), and that array of lists is also repeated twice (third dimension).    <class 'int'> is the type of the items added to the array, int32 is the numpy type of what the array contains, and <class 'numpy.ndarray'> is the type of the variable my\_array. 5 is the value of each item in the array. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Multiple choice | Suppose you have a list of 2-value tuples in your Python environment. It would be on this format, but with a much larger number of tuples. I.e. len(my\_list) > 100000    You want to find the second value of the pair where the first value is equal to a variable named some\_value. For this, you write the following function:  Assume that the first value across every pair in my\_list will always be unique and that some\_value is equal to one of the first values across the pairs in my\_list, so that the search will always find a unique match.    Which of the following options would be the most time-efficient alternative to this function independently of the position of some\_value in my\_list? | - Using list comprehension instead of looping in the function. As in:    - Transforming my\_list into a tuple and iterating through it instead of the list. As in:    -Transforming “my\_list” to a dictionary and using key-value lookup. As in:      - Transforming my\_list to a Pandas dataframe and using value filtering with some\_value to get the second value of the same row. As in: | Transforming my\_list into a dictionary — The single-value lookup options (dictionary and Pandas) will always be more efficient than the iteration operations (list comprehension and tuples). From the first, using a built-in data structure (dictionaries) will be more efficient than calling a structure from an external library (Pandas dataframe). |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-4 | Fill in the blanks | What is the fourth element of the resulting list of the code below: |  | 2\_text |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-5 | Multiple choice | Explain the concept of API versioning and its importance in maintaining backward compatibility and supporting long-term API stability. | a) API versioning entails assigning unique version identifiers to different releases of an API, allowing developers to control when and how to adopt changes while ensuring existing integrations remain unaffected.  b) API versioning involves updating API documentation regularly to reflect changes in functionality and usage.  c) API versioning is unnecessary as APIs are inherently backward compatible by design.  d) API versioning refers to the practice of releasing major updates to APIs without considering backward compatibility, enabling rapid iteration and innovation. | a) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-5 | Multiple choice | You were provided with the url below that let you query Brazil population for 2006. Modify the url accordingly to query Brazil’s population from 2000 to 2022.  url = <http://api.worldbank.org/v2/country/br/indicator/SP.POP.TOTL?date=2006&format=json>  Which of the following is the standard deviation of the obtained sample of population (integer part)? | a) 12,287,013  b)13,589,208  c)12,745,397  d)13,934,749 | 12,287,013 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-6 |  | ??? |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-7 | Multiple choice | Which of the following visualization techniques is useful for illustrating the goodness of fit of a model for classification tasks? | a) Scatter plot b) ROC/AUC Graph c) Confusion Matrix d) Bar plot e) Principal Component Analysis f) Gaussian Kernel fit | b) and c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-7 | Multiple choice | Of the curves in the following graph, which one represents the model with the best fit? | a) 1 b) 2 c) 3 d) 4 | a) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-7 | Upload notebook | You need to explain to a non-technical audience how a random forest works. Plot one of the trees of the random forest with a depth of 2. for the train data.  Hint: The command plot\_tree will be useful for this task |  |  | [train](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/train.csv?d=w7f1c5dbe6cfc47f5ae938e329205a9d2&csf=1&web=1&e=Teu2t7) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-7 | Upload notebook | You need to explain how your PCA model explains the variance of the data. Generate the plot needed for this task |  |  | [pca](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/data_for_pca.csv?d=w61851167f5ee4c00b6ce57648a22ec10&csf=1&web=1&e=uLa5ck) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-7 | Upload notebook | Plot your results from the clustering done before. In the X axis plot PCA1 and in the y axis plot PCA2. Color the points using the labels obtained. |  |  | [pca](https://worldbankgroup.sharepoint.com/:x:/r/teams/DevelopmentDataPartnershipCommunity-WBGroup/Shared%20Documents/Projects/Data%20Lab/Certifications/data_science_skills/datasets/data_for_pca.csv?d=w61851167f5ee4c00b6ce57648a22ec10&csf=1&web=1&e=uLa5ck) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-8 | Multiple choice | What is a common use case for employing heatmaps in data visualization? | a) Visualizing spatial relationships between geographic entities b) Representing temporal trends and patterns over time c) Highlighting patterns of concentration within a grid of data values d) Showing hierarchical structures within data | c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-8 | Multiple choice | What is the name of the programming paradigm that allows for dynamic changes in data analysis through user input like in the example below? | a) Reactive programming b) Imperative programming c) Functional programming d) Object oriented programming | a) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-8 | Multiple choice | What role do reactive programming paradigms play in interactive dashboards built with Shiny and Dash? | a) Reactive programming enables the creation of interactive user interfaces by automatically updating dashboard components in response to user inputs or changes in underlying data b) Reactive programming is not applicable to interactive dashboards and is primarily used for backend development.  c) Reactive programming is used for data preprocessing and feature engineering tasks in interactive dashboards. d) Reactive programming is especially useful for event-driven programming and orchestration of workflows in big-data contexts | a) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS2-9 |  | ??? |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-1 | Multiple choice | Which of the following graphs represents a ReLU activation function? |  | c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-1 | Multiple choice |  | a) To increase the model's capacity  b) To reduce overfitting by randomly dropping neurons during training  c) To speed up training by reducing the number of epochs  d) To improve the interpretability of the model's predictions | b) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-1 | Multiple choice |  | a) The process of transferring data from one domain to another  b) The process of transferring neural network layers from one model to another  c) The process of transferring knowledge from one pre-trained model to another model for a different task  d) The process of transferring weights from one layer to another within the same model | c) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-1 | Multiple choice | What is tokenization in natural language processing? | a) The process of converting text into a sequence of integers  b) The process of converting text into a sequence of tokens or words  c) The process of identifying parts of speech in a sentence  d) The process of transforming text into binary format | b) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-1 | Multiple choice | What is the purpose of stemming or lemmatization in NLP? | a) To reduce words to their base or root form  b) To remove stop words from the text  c) To identify named entities in the text  d) To convert text into lowercase format | a) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-1 | Multiple choice | What is the Transformer architecture commonly used in NLP? | a) A type of recurrent neural network (RNN)  b) A type of convolutional neural network (CNN)  c) A type of graph neural network (GNN)  d) A type of self-attention mechanism-based neural network | d) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | How do you make HTTP requests to AWS services using Python? | - Using the requests library  - Using the boto3 library  - Using the aws-sdk library  - Using the urllib library | Using the boto3 library |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What Python library is commonly used to interact with AWS services programmatically? | - boto3  - aws-sdk  - aws-python  - botobot | boto3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What is the command to install the boto3 library using pip? | - pip install boto3  - pip install aws-sdk  - pip install aws-python  - pip install botobot | pip install boto3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | How do you configure AWS credentials in your Python script to interact with AWS services? | - By setting environment variables  - By passing credentials as parameters to AWS service functions  - By storing credentials in a configuration file  - By hardcoding credentials directly into the script | By setting environment variables |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | How do you make HTTP requests to Google Cloud services using Python? | - Using the requests library  - Using the google-cloud library  - Using the gcloud library  - Using the urllib library | Using the google-cloud library |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What Python library is commonly used to interact with Google Cloud services programmatically? | - google-cloud  - gcloud  - cloud-python  - google-sdk | google-cloud |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What is the HTTP method commonly used to retrieve data from an API? | - GET  - POST  - PUT  - DELETE | GET |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What Python library is commonly used to make HTTP requests to APIs? | - requests  - urllib  - httplib  - http.client | requests |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What is the typical format of data exchanged between a client and an API? | - JSON (JavaScript Object Notation)  - XML (eXtensible Markup Language)  - CSV (Comma-Separated Values)  - HTML (Hypertext Markup Language) | JSON (JavaScript Object Notation) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-2 | Multiple choice | What AWS service can be used to store and retrieve any amount of data at any time, anywhere on the web? | - Amazon S3  - Amazon RDS  - Amazon DynamoDB  - Amazon Redshift | Amazon S3 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | What does CI/CD stand for? | Continuous Integration/Continuous Deployment  Centralized Integration/Continuous Deployment  Continuous Improvement/Continuous Deployment  Continuous Integration/Continuous Delivery | Continuous Integration/Continuous Deployment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | Which of the following is a key benefit of CI/CD in data science projects? | Reduced testing  Faster delivery of features  Increased technical debt  Decreased automation | Faster delivery of features |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | What tool is commonly used for automating CI/CD pipelines in data science projects? | Jenkins  Docker  Kubernetes  Ansible | Jenkins |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | How can version control systems like Git be integrated into CI/CD pipelines for data science projects? | By automatically triggering builds on new commits  By storing large datasets  By managing production deployments  By running machine learning algorithms | By automatically triggering builds on new commits |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | Which section of the GitHub Actions YAML file is used to specify the events that will trigger the workflow? | triggers  on  events  schedule | on |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | What is the name of the YAML file where GitHub Actions workflows are defined? | - .travis.yml  - workflow.yml  - action.yml  - .github/workflows/main.yml | .github/workflows/main.yml |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | How do you execute a Python script in a GitHub Actions workflow? | - run: python script.py  - python: script.py  - script: python script.py  - steps: python script.py | script: python script.py |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | DS3-4 | Multiple choice | What is the correct syntax to schedule a workflow to run every day at 10:00 AM UTC? | a) schedule:  - cron: "0 10 \* \* \*"  b) schedule:  - every: "0 10 \* \* \*"  c) schedule:  - daily: "10:00 AM"  d) schedule:  - time: "10:00 AM" | a) schedule:  - cron: "0 10 \* \* \*" |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |