



1. Which variant of SMOTE is most appropriate when you have a mixture of categorical and continuous variables? D
 - A. KMeansSMOTE
 - B. BorderlineSMOTE
 - C. SVM SMOTE
 - D. SMOTENC
 - E. SMOTE

2. Which fundamental part of the data ingestion process is concerned with to the phrase "bad data in equals bad data out"? B
 - A. Gather all relevant data from the sources of provided data
 - B. Implement several checks for quality assurance
 - C. Take the initial steps towards automation of the ingestion pipeline
 - D. All of the answers

3. What are the two steps that must be carried out if you want to iterate locally on a model then deploy it to the Watson Machine Learning (WML) service? D
 - A. Provision a WML service & Dockerize your model
 - B. Create a Python virtual environment & Dockerize your model
 - C. Provision cloud storage & Provision a WML service
 - D. Provision a WML service & create a Python virtual environment
 - E. Dockerize your model & Provision cloud storage

4. Which of the following is not a valid way to pass parameter to a Spark MLlib model? A



- A. Pass a pickle file when the model is declared
- B. Set the named parameters directly when the model is declared
- C. Pass a ParamMap to the fit() method of the model
- D. Create a ParamMap, update it then pass it to a fit() method of the model
- E. Pass a ParamMap to the fit() method of a pipeline

5. Which of the following statements is not a feature of the package imbalanced-learn imbalance? D

- A. Has a suite of over and under-sampling methods implemented
- B. Works with TensorFlow
- C. Has a number of tutorials to work from
- D. Has outlier detection algorithm packaged as part of library
- E. Works with scikit-learn pipelines

6. Which type of recommender system is readily available in Spark Machine learning? B

- A. Utility-based recommender systems
- B. Collaborative-based recommender systems
- C. Content-based recommender systems
- D. Hybrid recommender systems
- E. Demographics-based recommender systems

7. Which tool can help you scale your analysis with such massive data sources ? A

- A. Dask
- B. CSV application
- C. Visual Studio Code



D. Database

E. Pycharm

8. Which of the following statements describes the best strategy to address class imbalance? E
- A. If there is a lot of data just use under-sampling otherwise use outlier detection algorithms.
 - B. Determine the best variant of SMOTE by comparisons and use it.
 - C. Continue to collect data until you have balanced classes.
 - D. Use an outlier detection algorithm or SVM instead of a re-sampling technique.
 - E. Compare re-sampling approaches to a baseline and to detection algorithms.
-
9. Which of the following lists contains one or more references to a technology that is not a specific python package used to speed up and improve the performance of python code? C
- A. py-cuda, Cython
 - B. multiprocessing, mpi4py
 - C. subprocess, symmetric-multiprocessing
 - D. threading, ipyparallel
-
10. When printing the most representative words from each topic what best describes the insight we gain? B
- A. The top words in each topic correspond to the most frequently used words in the corpus
 - B. The topics are defined by their representative words and the document is a mixture of these topics
 - C. Documents have topics and the words describe the



corpus

D. The words make up the document and the topics describe the words

E. Topics are latent features and the top words describe the average document

-
11. Docker images are the basis of containers. It is possible to pull an image from the registry and ask the Docker client to run a container based on that image. Some images are official while many others are user defined. A

A. True

B. False

-
12. In the continuing AAVAIL streaming case study example, one of the data features that can be useful in answering questions about customer churn is the total number of streams that a customer has watched. Imagine that you are working with a dataset where 10% of customers are missing this feature. A good place to start would be to go back and see if it's possible to gather this information from the user logs, but assuming that this initiative is unsuccessful, you will have to decide what to do about this missing data. Which course of action is LEAST likely to be helpful in modeling churn? B

A. Replace the missing stream count with the mean stream count among users where this information is available.

B. Replace the missing stream count with a - 1 to flag that it is unknown for a given user.

C. Use the other features in the dataset in a model to



predict the missing stream counts.

D. None of the correct answer.

13. Cross-validation along with train-test splits are critical tools to help ensure our models are well calibrated for unseen data A

A. True

B. False

14. What important feature of the Watson NLU API do we have to ensure so that we can repeat an experiment in the future in a reproducible way? B

A. GitHub integration

B. Passing a version argument with each request

C. Each future version of the API is guaranteed to keep the same arguments

D. The NLU can be run locally

E. The exchanged JSON is guaranteed to keep the same format

15. Which types of programming tasks best describes what you are expected to already have some familiarity with before beginning this course? B

A. dashboarding, high performance computing, and code profiling

B. numeric computing, data munging, data visualization and data modeling

C. convex optimization, python programming, statistical programming

D. continuous integration, linear programming, and data exploration

- 16.

A



If you were to munge the data into a pandas.DataFrame which of the following would describe a reasonable goal for the cleaning process?

- A. customer on the rows and items like total _sales, name, most_bought on the columns
 - B. daily revenue on the rows items like customer_name and total on the columns
 - C. transactions on the rows and items like customer-_name and item _id on the columns
 - D. None of the correct answer
-

17. Which of the following does not describe a feature of the Watson Natural Language Understanding service? A

- A. Perform document classification tasks using a custom model built from text
 - B. Identify high-level concepts that aren't necessarily directly referenced in the text
 - C. Find people, places, events, and other types of entities mentioned in your content
 - D. Recognize when two entities are related, and identify the type of relation
 - E. Analyze the sentiment toward specific target phrases and the sentiment of the document as a whole
-

18. Which of the following is NOT normally a part of the EDA process D

- A. Visual summaries of the data
 - B. Connecting the data to the business opportunity
 - C. Communication to stakeholders
 - D. Predictive linear or logistic regression
-



-
19. In the context of the AI workflow presented in these materials which of the following is not an example of a valid feedback loop? D
- A. Trying different data transformations on a given model
 - B. Returning to the data collection stage from transformations to reduce the number of transforms
 - C. Performing EDA on the data after a model has been deployed and data have been logged
 - D. Moving from the business opportunity and data collection to model iteration
 - E. Returning to discuss the business opportunity after a model has been deployed
-
20. Which of the following model is not an example of an ensemble approach to learning? A
- A. Decision tree
 - B. Random forest
 - C. Boosting
 - D. Model stacking
 - E. Gradient boosting
-
21. You may only pass a pickle file to save your model in the your Watson Machine Learning library. B
- A. True
 - B. False
-
22. The three types of missingss discussed during this module were: E
- A. MRAR, MAR, MCAR
 - B. MNAR, MRAR, MCAR
 - C. MNAR, MAR, MARC



- D. MAR, MRAR, MCAR
- E. MCAR, MNAR, MAR

-
23. IBM Watson OpenScale only works with models that have been built on, and deployed by, IBM Watson Stuido. B
- A. True
 - B. False
-
24. Which types of data generally work well with sparse matrices? C
- A. word counts, time-series data
 - B. audio files, images
 - C. word counts, user-item matrix for recommendations
 - D. text data, audio files
-
25. Which of the following is not an example of a relevant question when tuning a NLP classification pipeline? E
- A. Should I use bag-of-words or a vector embedding representation?
 - B. Which stop words do I include?
 - C. Which n-grams do I include?
 - D. Should I use a TF or a tf-idf transformation?
 - E. Should I use RMSE or MAE as an evaluation metric?
-
26. A Kubernetes pod can contain multiple kubernetes deployments B
- A. True
 - B. False
-
27. Which of the following is not an example of a technique used for dimensionality reduction technique? E



- A. Latent Dirichlet allocation
- B. Non-negative matrix factorization
- C. Singular value decomposition
- D. Eigenvalue decomposition
- E. K-nearest neighbors
- F. Principal Components Analysis

-
28. To move through the AI enterprise workflow quickly, we used the concept of workflow templates. Which of the following aspects of the template contributes the least to the re-usability of these templates? C
- A. When they are used with a version control system
 - B. Ensuring that the model exists as a separate Python module to be called by Flask
 - C. Have an HTML endpoint for the Flask app to create dashboards
 - D. Use a template for a unit test suite
 - E. Ensure that data ingestion and data visualization code exists as a modules or scripts
-
29. Common sources of performance drift include: D
- A. Changes in customer demographics or user behavior after a model is trained
 - B. Incomplete, incorrect or unbalanced training data
 - C. Changes in versions of libraries or modules used in models
 - D. All of the answers
-
30. Which statement is the least true about using Jupyter notebooks in the context of EDA A
- A. They naturally lend themselves to version control systems



- B. They can be ported from one environment to another easily
- C. They are helpful because a mixture of code and markdown enables storytelling
- D. They are integrated with the plotting library matplotlib
- E. They are integrated with the data manipulation library pandas

-
31. For the vector $v = [2.0, -3.5, 5.1]$: A
Find the L1 norm of v
Return the result as a float
`import numpy as np`
`def calculate_l1_norm(v):`
 `#Code here?`
 `return L1`
A. `L1 = np.linalg.norm(v, ord=1)`
B. `L1 = np.linalg.norm(v, ord=1)`
C. `L1 = np.linalg.norm(v, ord=1)`
D. `L1 = norm(v, ord=1)`
-
32. Though the emphasis may change, which two elements are both essential and common to all three import num models we talked about? C
A. prediction, recommendation
B. data mining, data cleaning
C. resolve the business question, feedback loops
D. testing, model deployment
-
33. Which of the following is most concerned with ensuring deployed models scale well with added users? C
A. data scientist



- B. data analysts
- C. data engineer
- D. product manager

-
34. When you use Watson Services like Watson Natural Language Understanding via the Python SDK, what are the three items that need to be saved? These items are generally saved on a local machine and included in scripts and notebooks as imported variables. C
- A. service version, service API key, service JSON map
 - B. service URL, service JSON map, service API key
 - C. service API key, service version, service URL
 - D. service version, service IAMAuthenticator, service URL
 - E. service API key, service URL, service IAMAuthenticator
-
35. A Spark cluster is generally managed using a Docker container and a YAML file. B
- A. True
 - B. False
-
36. When training a custom classifier in Watson Visual Recognition the negative images should be: A
- A. As visually similar as possible to the positive images
 - B. Background images without the positive images
 - C. As random as possible to establish a background
 - D. Randomly generated from the positive images
 - E. Visually distinct from the positive images
-
37. Lets imagine there is a start-up that has a speech-to-text service that incorporates gestures and C



body language into its output. They offer annotated meeting reports as a product and customers are generally very

satisfied, but sales to new customers tend to be very slow to acquire. Which of the following business opportunities should be the highest priority?

- A. Develop and delivery new products to existing customers
- B. Develop new products and target new customers
- C. Use customer segmentation and or market analysis to help marketing with new customers
- D. Use customer segmentation and/or market analysis to move into a different market

38. What is Port Forwarding in Docker? C

- A. The process of creating a data file on the local computer that persists after the container is closed
- B. Sending one container to multiple nodes on a multicloud environment
- C. Allowing data to be passed in and out of a docker container and controlling which applications can do this
- D. Building one Docker image using another docker image as a template

39. Which of the following is least likely to be a use case for Docker containers? E

- A. Microservices: many loosely coupled and independently deployable services
- B. DevOps/Data Engineers use containers as a common platform for many teams
- C. Avoid install overhead with a Spark environment



- D. Hybrid, multi-cloud portability for a machine learning model
- E. Replacement for a standard virtual machine

-
40. Which of the following is an example of a data manipulation that is NOT considered reproducible research? D
- A. Saving classes and functions in a Python file to be called by Jupyter
 - B. Code blocks in Jupyter notebooks
 - C. Graphics, plots and other visualizations
 - D. Copy and paste actions in a spreadsheet
 - E. The use of proprietary tools to carry out research
-
41. It is reasonable to think of the commands in a Dockerfile as a step-by step recipe on how to build up a Docker A
- A. True
 - B. False
-
42. Your company is convinced it is time to change the nature of your company's core product and management has come to ask your advice. Which question DOES NOT exemplify scientific thinking in this situation? C
- A. Do we have any data like a corpus of customer feedback to support this decision?
 - B. Can we run an experiment like A/B testing to see if it helps support this decision?
 - C. Which members of leadership support this decision?
 - D. Have any other companies been successful making a comparable change?

43.

C



Processing the corpus with the provided lemmatize_document reduces the total number of tokens to what percentage of the original?

- A. 10-15%
- B. 20-35%
- C. 45-50%
- D. 70-75%
- E. 85-95%

44. It is critical in machine learning model deployment workflows to ensure close to 100% test coverage before production. B

- A. True
- B. False

45. The date and time of a transaction should never be logged because this will only reflect past transactions and is not relevant to the performance of predictions on an on-going basis. B

- A. True
- B. False

46. Which of the following use cases is the least appropriate use case for a convolutional neural network? C

- A. Image classification
- B. Image retrieval
- C. Image composition
- D. Object detection
- E. Image segmentation

47. Which of the following was not discussed as a tunable parameter of a neural network? A

- A. Hardware availability



- B. Activation functions: sigmoid, tanh, softmax, ReLU, leaky ReLU
 - C. Regularization techniques: weight decay, early stopping, dropout
 - D. Training method: Loss function, learning rate, batch size, number of epochs
 - E. Structure: the number of hidden layers, the number of nodes in each layer
-

48. Which of the following are three services provided by IBM Watson OpenScale? C

- A. Automatic CSS-HTML Generation, Hyperparameter Tuning, Performance Metrics
 - B. Compliance, Cross Regional Availability, Automatic Backup
 - C. Drift Monitoring, Fairness, Explainability
 - D. General Artificial Intelligence, Time Travel, World Domination
-

49. A decision tree classifier is useful as a model for the AAVAIL subscriber churn data. A

- A. True
 - B. False
-

50. When you worked on model deployment case study, which modification to the ALS algorithm had the largest effect on model performance? B

- A. The explicit training vs implicit training comparison
 - B. The lambda or regularization parameter
 - C. The epsilon or scale parameter
 - D. The I1 vs I2 comparison
-

51. B



To succeed in this course you are expected to be proficient in any one of the following: R, Python or Java.

- A. True
- B. False

52. Which of the following is the least accurate statement B
about the advantages of using process models in data science? Process models generally help by...

- A. avoiding unnecessary tangents
- B. speeding up the process of getting through the workflow
- C. minimizing the model selection process
- D. guiding effective time allocation

53. Design thinking is applied in other domains which A
helps make the task of communicating the AI workflow to those outside of data science easier.

- A. TRUE
- B. FALSE

54. Lets imagine there is a start-up that has a D
speech-to-text service that incorporates gestures and body language into its output. Which of the following products represents the most defensible business opportunity.

- A. Offer a service that hooks into streaming video and predicts the emotional state of people in the videos
- B. Create an app that allows job interviewers to get additional information about candidates
- C. Create a new and improved conferencing app
- D. Create a service that improves on existing audio



recognition systems as a richer interface to mobile devices

55. CSV files are one of the most commonly used file formats for data science because file input/output is easy they are plain-text, and they work well with commonly used spreadsheet tools. A
A. TRUE
B. FALSE
-
56. Which of the following DOES NOT represent a valid relational database to connector relationship? D
A. MySQL --> MySQL-python
B. PostgreSQL --> psycopg2/psycopg2
C. SQLite --> sqlite3
D. Berkeley DB --> bsddb
-
57. Sparse matrices from SciPy need to be transformed into a dense matrix before using scikit-learn's train-test-split function? B
A. TRUE
B. FALSE
-
58. The EDA process is decoupled from modeling and cannot be used to help estimate the time it will take to complete a modeling procedure B
A. TRUE
B. FALSE
-
59. Which of the following is the least valid statement when it comes to dashboards? D
A. Dashboards are an easy way to share summaries and findings



- B. Dashboards have interactive functionality that helps create a rich experience for the user
- C. Dashboards are generally used after several iterations of the AI workflow
- D. Dashboards are quick way to create portable simple plots
- E. Dashboards can be used to tell the story of investigative visualizations

60. Suppose that on average 2.5% of visitors to your website sign up for your newsletter. In a recent week, 2701 visitors out of a total of 108879 signed up. B

Using a binomial distribution. What is the probability that number of visitors who signed up is 2701 or fewer?

- A. 0.125
- B. 0.346
- C. 0.414
- D. 0.007
- E. 0.015

61. If there customer churn were quantified using a Poisson distribution, then a bootstrap could be used to quantify the uncertainty associated with the estimate. A

- A. TRUE
- B. FALSE

62. Which of the following is NOT an example of a valid strategy to deal with the multiple comparisons problem? C

- A. Benjamini/Hochberg correction based on False dis-



covery Rates

B. Create a null distribution using permutations to help provide context

C. Perform all comparisons then only keep the single test that performs the best

D. If appropriate use an alternative modeling framework like generalized linear models

E. Bonferroni Correction

63. The .fit_transform method corresponds to which scikit-learn interface(s)? B

A. Transformer, Estimator, Predictor

B. Transformer, Estimator

C. Estimator, Predictor

D. Transformer

E. Transformer, Predictor

64. A principal reason for emphasizing the use of pipelines in the AI workflow is to have a consistent platform that enables comparison of many variants of the workflow. A

A. TRUE

B. FALSE

65. Which of the following statements does not describe valid use case for dimensionality reduction? E

A. Principal components analysis to process images used in classification.

B. Non-negative matrix factorization to resolve topics from a corpus of words.

C. t-distributed stochastic neighbor embedding to visualize the results of a clustering algorithm.



- D. Using an ANOVA to select a subset of features
- E. Down-sampling of the majority class

-
66. tSNE is a reasonable alternative to PCA because it describes a wider variety of structures. However, it is still recommended to use another dimensionality reduction method, like PCA if the number of features is very high. A
- A. TRUE
 - B. FALSE
-
67. Which statement best describes why visualization of topics can have an impact on the business opportunity? A
- A. Because sharing with domain experts might enable topic-specific feature engineering
 - B. Because visual inspection can help choose the number of topics
 - C. Because we are able to see the top words with each topic
 - D. Because we are able to see the relative importance of each topic across the corpus
 - E. Because domain experts can visually inspect the validity of the topics
-
68. You are asked to build a recommendation engine for new products at an online retailer. Which of the following features is the least likely to be a protected attribute? Recall that a protected attribute is one that may contain privileged and unprivileged classes. E
- A. age
 - B. gender identity



- C. race
 - D. religion
 - E. purchase history
-

69. The API for the reweighting algorithm has a number of custom methods that will require you to consult the documentation to ensure appropriate use. A

- A. TRUE
 - B. FALSE
-

70. Which outlier detection method is known to work well on high-dimensional data? A

- A. Random Forests
 - B. Elliptic Envelope
 - C. One Class SVM
 - D. Isolation Forest
 - E. Local Outlier Factor
-

71. In an imbalanced dataset if the minority class represents less than 5% of all of the samples then outlier detection algorithms must be used in place of other supervised learning algorithms. B

- A. TRUE
 - B. FALSE
-

72. Which clustering method can be readily applied to graphs? B

- A. Gaussian mixture models
 - B. Spectral clustering
 - C. Affinity Propagation
 - D. k-means
 - E. Dirichlet process Gaussian mixture models
-



-
73. Which of the following clustering methods does not need to set the number of clusters? E
- A. Gaussian mixture models
 - B. Spectral clustering
 - C. MiniBatch k-means
 - D. k-means
 - E. Dirichlet process Gaussian mixture models
-
74. In the clustering case study, the suggested re-sampling methods drove a major improvement in model performance. A
- A. TRUE
 - B. FALSE
-
75. In the context of customer profiling and the AAVAIL data set it makes sense to first perform a dimension reduction technique like PCA before running the model through a clustering estimator. A
- A. TRUE
 - B. FALSE
-
76. If we have a situation where false positive is not as potentially costly as a false negative say flagging comments for manual review based on suspected unlawful activity, which of the following is the best approach to consider? E
- A. Only look at the f-score of the negative class for evaluation
 - B. Use recall as the evaluation metric
 - C. Use precision as the evaluation metric
 - D. Set beta to 0.5 in the fscore
 - E. Set beta to 2.0 in the fscore
-



-
77. All classifiers in scikit-learn do multi-class classification out-of-the-box. These classifiers can differ in their approach though (e.g one-vs-all or one-vs-one). A
- A. TRUE
- B. FALSE
-
78. Which of the following is not an example of a generalized linear model (GLM)? D
- A. ANOVA
- B. Multinomial regression
- C. Poisson regression
- D. KNN regression
- E. Logistic regression
-
79. Models in the generalized linear mixture model (GLMM) family like multilevel models are generally optimized using sophisticated techniques like MCMC sampling. A
- A. TRUE
- B. FALSE
-
80. Bagging and boosting ensemble methods both use only decision trees as base classifiers. The difference is in the bias and variance of the individual trees. B
- A. TRUE
- B. FALSE
-
81. Transfer learning is a recent advancement to come out of the field of reinforcement learning. B
- A. TRUE
- B. FALSE
-

82. B



The Watson Visual Recognition service can only be accessed using an API key via Python or curl.

- A. TRUE
- B. FALSE

83. A typical convolutional neural network is constructed using a combination of convolutional, pooling and dense layers.

- A. TRUE
- B. FALSE

84. What is the principal reason to create a virtual environment before creating your model locally?

- A. Because all models should have their own virtual environment
- B. Because it will ensure that the most recent packages are used
- C. Because virtual environments can be containerized easily
- D. Because the Python client for Watson Machine Learning has specific dependencies
- E. Because it ensures that locally created/trained models are compatible with Watson Machine Learning

85. Spark MLlib has several estimators and transformers, but it lacks basic tooling for natural language processing (NLP) like the ability to make term frequency-inverse document frequency (TF-IDF) transformations.

- A. TRUE
- B. FALSE

86. Which phrase most accurately reflects what is meant by the cold start problem in recommendation sys-



tems?

- A. When a new user is introduced and the recommendations are made on similarities
- B. When a new item is introduced and the recommendations are made on similarities
- C. When a new user is introduced and the recommendations are made on item popularity
- D. When a new user or new item is introduced and the recommendations depend on popularity
- E. When a new user or new item is introduced and the recommendations depend on similarities

87. Matrix factorization is commonly associated with collaborative filtering recommender systems. A

- A. TRUE
- B. FALSE

88. The Spark collaborative filtering implementation is built by default for explicit feedback, but it can be used with implicit feedback just as easily. A

- A. TRUE
- B. FALSE

89. In the context of the AI workflow presented in these materials which of the following is not an example of a valid feedback loop? D

- A. Trying different data transformations on a given model
- B. Returning to the data collection stage from transformations to reduce the number of transforms
- C. Performing EDA on the data after a model has been deployed and data have been logged



- D. Moving from the business opportunity and data collection to model iteration
- E. Returning to discuss the business opportunity after a model has been deployed

-
90. Which list contains one or more elements that were presented as non-essential when creating a logging system to monitor the performance of a machine learning model that has been deployed. A
- A. request_type, input data
 - B. model_version_number, timestamp
 - C. predictions/recommendations, timestamp
 - D. input_data_summary, runtime
-
91. What is the purpose of kubectl in kubernetes? D
- A. Automatic logging of requests and responses
 - B. A tool that makes it easy to run a single-node cluster locally
 - C. The primary node agent on each node, responsible for the processes running on that machine
 - D. The CLI for communicating with the kubernetes cluster
-
92. AAVAIL Management has described their needs to you in the language of business. How would you describe this project in the language of Data Science? D
- A. Recommender system
 - B. A/B Testing
 - C. Dimensionality Reduction
 - D. Regression Analysis
 - E. Unsupervised Learning
 - F. Classification Analysis
-



-
93. When you compiled the JSON files into a single DataFrame or NumPy array, about how many days did the entire range of dates span? C
- A. 400
 - B. 450
 - C. 500
 - D. 600
 - E. 650
-
94. Which country had the most total revenue when you summed across all purchases? B
- A. Singapore
 - B. United Kingdom
 - C. USA
 - D. EIRE
 - E. Germany
-
95. There are many established methods from the literature that may be used to analyze time-series data. Classical machine learning techniques can also be used, but which of the following is the most important caveat to keep in mind when employing this approach? E
- A. You cannot use scikit-learn to combine multiple predictions
 - B. The data cannot be feature engineered to help identify features that are needed by the model to detect trends
 - C. There is no simple way to do a train/test split of the data
 - D. Significant computational resources are generally required



E. A specialized strategy that uses recursion (for example) or multiple models must be employed for multi-point forecasting

96. When the data are pivoted and summarized across days which of the following features were the most difficult to include as a summarized or engineered feature? E
- A. revenue
 - B. stream_views
 - C. stream_ids
 - D. purchases
 - E. customer_id
-
97. The time-series plots using a day time interval for revenue, purchases and total_views revealed many similarities. Which part of the year indicated a trend for the highest level of activity? C
- A. January-February
 - B. April-May
 - C. November-December
 - D. June-July
 - E. September-October
-
98. Your Flask application needs to error check the input when making a prediction. Which features need to be provided (at a minimum) as input given the context of this business opportunity? B
- A. The previous three months of data and the target month from the year before
 - B. The target date and country
 - C. At least one year of revenue data



- D. The recent number of transactions and the number of views
 - E. Metadata associated with the streams
-

99. Unit testing and novelty detection provide an automated mechanism to monitor model performance. A
- A. TRUE
 - B. FALSE
-

100. To move through the AI enterprise workflow quickly, we used the concept of workflow templates. Which of the following aspects of the template contributes the least to the re-usability of these templates? C
- A. When they are used with a version control system
 - B. Ensuring that the model exists as a separate Python module to be called by Flask
 - C. Have an HTML endpoint for the Flask app to create dashboards
 - D. Use a template for a unit test suite
 - E. Ensure that data ingestion and data visualization code exists as a modules or scripts
-