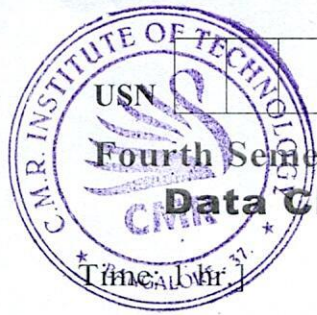




Question Paper Version : A

Fourth Semester B.E./B.Tech. Degree Examination, Dec.2024/Jan.2025 Data Cleaning & Preparation with Python Pandas

[Max. Marks: 50]

INSTRUCTIONS TO THE CANDIDATES

1. Answer all the fifty questions, each question carries one mark.
2. Use only **Black ball point pen** for writing / darkening the circles.
3. **For each question, after selecting your answer, darken the appropriate circle corresponding to the same question number on the OMR sheet.**
4. Darkening two circles for the same question makes the answer invalid.
5. **Damaging/overwriting, using whiteners** on the **OMR** sheets are strictly prohibited.

1. What is the primary data structure in Pandas for handling one – dimensional labeled data?
a) Array b) DataFrame c) Series d) Panel
2. Which of the following is not a data type that a Pandas series can have?
a) int 64 b) float 32 c) str d) complex
3. In Pandas DataFrame, what is used to uniquely identify each row?
a) Column b) Index c) Series d) Data
4. Which function is used to sort the values in a Pandas Data Frame along a particular axis?
a) Sorted_values () b) order () c) arrange () d) sort ()
5. What Pandas function is used for creating a copy of an existing DataFrame?
a) duplicate () b) copy () c) replicate () d) clone ()
6. How can you select a column named 'age' from a DataFrame df?
a) df['age'] b) df.loc ['age'] c) df.iloc ['age'] d) df.select ('age')
7. What is the purpose of using Boolean indexing in Pandas?
a) To filter data based on conditions
b) To perform bitwise operations on columns
c) To convert all values to Boolean type
d) To check if the Data Frame is empty.
8. Which of the following is used for selecting data by its label in Pandas?
a) iloc b) loc c) ix d) select

9. What is Multi Index used for in Pandas?
a) Sorting the DataFrame b) Handling missing values
c) Indexing with multiple levels d) Creating multiple DataFrames
10. What operation is used to combine two DataFrames along columns in Pandas?
a) merge () b) join () c) concat () d) combine ()
11. In Pandas, what function is used to reshape a DataFrame?
a) pivot () b) reshape () c) stack () d) transform ()
12. What does the Pivot_table () function in Pandas do?
a) Pivot columns based on their values
b) Create a summary table with aggregated values
c) Reshape the dataframe by stacking columns
d) Merge two Data Frames based on common columns.
13. How can you stack multiple DataFrames vertically in Pandas?
a) append () b) stack () c) merge () d) concatenate ()
14. Which Pandas function is used for string manipulation in a DataFrame column?
a) str_manipulate () b) text_process c) apply_str () d) str
15. What is the purpose of the str.contains () function in Pandas?
a) Check if a DataFrame contains specific values
b) Check if a string column contains a specific substring
c) Concatenate two string columns
d) Count the number of occurrences of a character in a string column.
16. How can you drop rows with missing values in a Pandas DataFrame?
a) dropha () b) remove_missing ()
c) drop_missing () d) delete_na ()
17. What function is used to fill missing values in a Pandas DataFrame?
a) fill_missing () b) complete () c) fill na () d) replace_na ()
18. What does the isna () function in Pandas return?
a) True for missing values, False for non-missing values
b) False for missing values, True for no – missing values
c) 1 for missing values, 0 for non-missing values
d) Non for missing values, a value for non-missing values
19. What is the first step in the process of grouping data in Pandas?
a) Splitting b) Aggregation c) Iterating d) Transforming
20. How can you select a specific group from a grouped DataFrame in Pandas?
a) group.select () b) group.get () c) get_group () d) select_group ()
21. What does the transform () function in Pandas do when applied to grouped data?
a) aggregates the data b) Transforms each group independently
c) Filters the data d) Sorts the data

22. What Pandas function is used to convert a column to a dateline format?
 a) to_datetime () b) convert_datetime ()
 c) date_format () d) datetime_convert ()
23. How can you resample time series data in pandas
 a) Resample() b) Sample() c) Time-resample() d) Time-Sample()
24. What does the diff() function do in pandas time series analysis?
 a) Calculates the difference between consecutive elements
 b) Filters out missing values
 c) Computes the mean difference
 d) Finds the first difference of the data
25. Which function is used to create plots in pandas?
 a) Plot() b) crentre_Plot() c) draw() d) Visunlize()
26. How can save a pandas plot to a file?
 a) save_Plot() b) export_Plot() c) savefig() d) Plot_save()
27. What is the purpose of the legend() function in pandas plotting?
 a) Adds a legend to the plot b) Hides the legend
 c) Changes the plot colors d) Adjusts the plot size
28. How can you read a large dataset in pandas in chunks?
 a) read_large() b) read_chunk()
 c) read_CSV_Chunk() d) read_in_chunks()
29. What is the purpose of the memory_usage parameter in the read_CSV()function?
 a) Controls the chunk size
 b) Filters the columns based on memory usage
 c) Displays memory usage information
 d) Optimizes memory usage when reading large files
30. What function can be used to detect and drop duplicate rows in a pandas DataFrame?
 a) find_dulicates() b) drop_duplicates()
 c) remove_duplicates() d) delete_duplicates()
31. How can you filter groups in a pandas DataFrame based on a custom condition?
 a) filter() b) group_filter() c) apply_filter() d) custom_filter()
32. What is the purpose of the nth()function in pandas when applied to grouped data?
 a) Retrieves the nth group in the DataFrame
 b) Returns the nth occurrence of each group
 c) Filters groups with less than n elements
 d) Computes the nth aggregation of each group
33. In Pandas, what is the difference between aggregation and transformation in the context of grouping?
 a) They are synonymous terms
 b) Aggregation combines groups, while transformation applies a function to each group
 c) Transformation combines groups, while aggregation applies a function to each group
 d) Aggregation and transformation are not related to grouping.
34. What is the purpose of the shift () function in Pandas time series analysis?
 a) Shifts the index by a specified number of periods
 b) Shifts the values by a specified number of periods
 c) Shifts the date format of a column
 d) Shifts the entire Data Frame to different time
35. How can you convert a Data Frames index to a date time index in Pandas?
 a) Convert_to_datetime () b) to_datetime_index ()
 c) Set_datetime_index () d) as_datetime_index ()
36. What does the rolling () function do in Pandas time series analysis?
 a) Rolls the DataFrame along the specified axis
 b) Computes a rolling mean or sum over a specified window
 c) Shifts the index based on a rolling window
 d) Performs a rolling transformation on each group
37. In Pandas, what is the purpose of the subplots parameter in the plot () function?
 a) Creates subplots for each column
 b) Subtracts one plot from another
 c) Specifies the number of subplots to be created
 d) Adjusts the spacing between subplots
38. How can you customize the color of a Pandas plot?
 a) Color () b) Set_Color
 c) Colormap () d) C parameter in the plot () function
39. What does the kind parameter in the plot () function specify?
 a) The type of plot to create b) The color of the plot
 c) The size of the plot d) The style of the plot
40. How can you handle memory errors when working with large datasets in Pandas?
 a) Increase system RAM b) Use the optimize_memory () function
 c) Adjust the chunk size while reading data d) Install a larger hard drive
41. What function can be used to randomly sample rows from a Pandas DataFrame?
 a) Sample () b) random_sample ()
 c) Select_random () d) Choose ()
42. What does the set_index () function do in Pandas?
 a) Resets the DataFrame index
 b) Sets a new index for the DataFrame
 c) Sorts the DataFrame by the specified index
 d) Creates a multi_index for the DataFrame.

43. Which Pandas function is used to calculate the correlation between columns in a DataFrame?
- Corr ()
 - Correlation ()
 - Calculate _Corr ()
 - Column _Correlation ()
44. How can you convert a DataFrame to a Numpy array in Pandas?
- to_array ()
 - as_array ()
 - Convert_to_array ()
 - Values ()
45. What is the purpose of the sample () function in Pandas?
- Returns a random sample of items from an axis of the object
 - Filters the DataFrame based on a specified condition
 - Selects a specific subset of columns
 - Samples values from a specified distribution
46. How can you drop a column named 'salary' from a Pandas Data Frame df?
- df.remove_column ('salary')
 - df.drop('salary', axis = 1)
 - df.delete_column ('salary')
 - df.exclude ('salary')
47. What is the purpose of the pivot () function in Pandas?
- Pivot columns based on their values
 - Reshape the DataFrame by stacking columns
 - Create a summary table with aggregated values
 - Merge two DataFrames based common columns
48. In Pandas, what function is used to handle categorical data efficiently?
- Categorical ()
 - handle_categorical
 - Category ()
 - Convert_Cateorical ()
49. How can you concatenate two Dataframes vertically in Pandas?
- df.add_rows ()
 - df.concat (axis = 0)
 - df.merge (axis = 1)
 - df.combine_first ()
50. What does the resample () function do in Pandas time series analysis?
- Resamples the DataFrame based on a specified rule
 - Computes the mean of each group
 - Converts the time index to a numeric index
 - Rearranges the DataFrame based on column values.
