

1	What is the output of the below code? import pandas as pd import numpy as np df=pd.DataFrame({"a":[1,2,np.nan,3,4],"b":[1,5,np.nan,2,1]}) df=df.drop_duplicates(subset="b") df.dropna() df.fillna(20,inplace=True) print(df.shape[0])
2	What is the output of the below code? import re s = "black, blue and brown" pattern = r'bl\w+\W' matches = re.findall(pattern,s) print(len(matches[0]))
3	In a box plot, the bottom line of the box represents which quartile?
4	import pandas as pd import numpy as np df=pd.DataFrame({"a":[1,2,np.nan,3,4],"b":[1,5,np.nan,2,1]}) df=df.drop_duplicates(subset="b",inplace=True) df.dropna(inplace=True) df.fillna(20,inplace=True) print(df.shape[0])
5	What is the output of the below code? import pandas as pd import numpy as np df=pd.DataFrame([[1,2,3,4,5],[2,1,3,4,5],[np.nan,np.nan,np.nan,np.nan,np.nan]]) df.dropna(thresh=3,axis=1,inplace=True) print(df.shape[1])
6	What is the output of the below code? import re text = "The code is ABC123XYZ and XYZ789." pattern = r'[A-Z]{3}\d{3}[\^s]{3}' result = re.findall(pattern, text) print(result)
7	What will be the output of the following Python function? re.findall("hello world", "hello")
8	What is the output of the below code? Import pandas as pd import numpy as np df=pd.DataFrame([[1,2,3,4,5],[2,1,3,4,5],[np.nan,np.nan,np.nan,np.nan,np.nan]]) df.drop(1,inplace=True) df=df.dropna() print(df[0].shape[0])
9	What is the output of the below code import re pattern = r'\d{3}' string = 'The price of the product is 1234 dollars.' match = re.findall(pattern, string) print(match[0])
10	What is the output of the below code? import re text = "Hello, how are you?" pattern = r'\w{3}\W+' result = re.sub(pattern, "###", text)

	(result)
11	<pre>import pandas as pd import numpy as np df=pd.DataFrame({'a':[1,2,np.nan,3,4],"b":[1,5,np.nan,2,1]}) df.drop_duplicates(subset="b",inplace=True) df.dropna(inplace=True) df.fillna(20,inplace=True) print(df.loc[0].shape)</pre>

<p>What is the output of the below code?</p> <pre>import re text = "The quick brown fox jumps over the lazy dog." result = re.findall(r'\w{4}'s', text) print(len(result))</pre>	
<pre>df=pd.DataFrame([[1,2,np.nan,3,5],[np.nan,5,7,9,1],[np.nan,3,np.nan,7,8],[np.nan,5,np.nan,7,10]] ,columns=["A","B","C","D","E"]) df.dropna(subset=["A","C"],how="all",inplace=True) df.drop_duplicates(subset=["A","C"],inplace=True) df.shape[0]</pre>	
<p>Data Description: This section details the features (columns) within the "train.csv" dataset: Survived: Indicates passenger survival (0 = No, 1 = Yes). Pclass: Passenger class (1 = 1st, 2 = 2nd, 3 = 3rd). Name: Passenger name. Sex: Passenger's gender (male or female). Age: Passenger's age in years. SibSp: Number of siblings/spouses aboard the Titanic with the passenger. Parch: Number of parents/children aboard the Titanic with the passenger. Ticket: Ticket number. Fare: Passenger fare paid. Cabin: Cabin number (may be missing). Embarked: Port of embarkation (C = Cherbourg, Q = Queenstown, S = Southampton).</p> <p>Data Analysis Tasks: The following tasks will be performed on the dataset:</p> <ol style="list-style-type: none"> 1. Give the basic information 2. Identify the number of missing values in each column 3. To drop the missing values in a given dataset permanently 4. Provide descriptive statistics for excluding numerical data 5. Calculate and display the correlation coefficients between numerical variables and visualize them using a heatmap with value. 6. Create a scatter plot to visualize the relationship between Fare and Survived. 7. Analyze and display the distribution of survival rates across Sex using a cross-tabulation table and a heatmap with value. 8. Explore how survival rates vary based on passenger class, fare, and age using parallel coordinates. 9. Analyze the relationship between age and fare using a regression plot. 10. Create a scatter matrix to visualize relationships between "Age," "Fare," and "Survived." 11. Determine the survival rates for males and females. 12. Identify outliers in various columns using boxplots: PassengerId, Survived, Pclass Age, SibSp, Parch, Fare 13. Remove the outlier in the dataset of the Fare column 14. After handling outliers in Fare column, repeat outlier detection using boxplots. 15. Passenger Categorization: <p>To analyze survival rates based on age groups, we will categorize passengers as follows:</p> <ul style="list-style-type: none"> • Children: Age <= 18, Adults: 18 < Age <= 30, Young Adults: 30 < Age <= 50, Seniors: Age > 50 <p>This categorization will be used to create a cross-tabulation table to analyze survival rates across these age groups.</p> <ol style="list-style-type: none"> 16. Calculate the total fare class wise 17. Draw scatter plot between age and fare 	

