1. Write a Pandas program to split the following dataframe into groups based on school code. Also check the type of GroupBy object.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

Click me to see the sample solution

2. Write a Pandas program to split the following dataframe by school code and get mean, min, and max value of age for each school.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

3. Write a Pandas program to split the following given dataframe into groups based on school code and class.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

4. Write a Pandas program to split the following given dataframe into groups based on school code and cast grouping as a list.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

5. Write a Pandas program to split the following given dataframe into groups based on single column and multiple columns. Find the size of the grouped data.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

6. Write a Pandas program to split the following given dataframe into groups based on school code and call a specific group with the name of the group.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

Click me to see the sample solution

7. Write a Pandas program to split a dataset, group by one column and get mean, min, and max values by group. Using the following dataset find the mean, min, and max values of purchase amount (purch\_amt) group by customer id (customer\_id).

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 2012-10-05 3005 5002

1 70009 270.65 2012-09-10 3001 5005

2 70002 65.26 2012-10-05 3002 5001

3 70004 110.50 2012-08-17 3009 5003

4 70007 948.50 2012-09-10 3005 5002

5 70005 2400.60 2012-07-27 3007 5001

6 70008 5760.00 2012-09-10 3002 5001

7 70010 1983.43 2012-10-10 3004 5006

8 70003 2480.40 2012-10-10 3009 5003

9 70012 250.45 2012-06-27 3008 5002

10 70011 75.29 2012-08-17 3003 5007

11 70013 3045.60 2012-04-25 3002 5001

8. Write a Pandas program to split a dataset to group by two columns and count by each row.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 2012-10-05 3005 5002

1 70009 270.65 2012-09-10 3001 5005

2 70002 65.26 2012-10-05 3002 5001

3 70004 110.50 2012-08-17 3009 5003

4 70007 948.50 2012-09-10 3005 5002

5 70005 2400.60 2012-07-27 3007 5001

6 70008 5760.00 2012-09-10 3002 5001

7 70010 1983.43 2012-10-10 3004 5006

8 70003 2480.40 2012-10-10 3009 5003

9 70012 250.45 2012-06-27 3008 5002

10 70011 75.29 2012-08-17 3003 5007

11 70013 3045.60 2012-04-25 3002 5001

9. Write a Pandas program to split a dataset to group by two columns and then sort the aggregated results within the groups.

In the following dataset group on 'customer\_id', 'salesman\_id' and then sort sum of purch\_amt within the groups.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 2012-10-05 3005 5002

1 70009 270.65 2012-09-10 3001 5005

2 70002 65.26 2012-10-05 3002 5001

3 70004 110.50 2012-08-17 3009 5003

4 70007 948.50 2012-09-10 3005 5002

5 70005 2400.60 2012-07-27 3007 5001

6 70008 5760.00 2012-09-10 3002 5001

7 70010 1983.43 2012-10-10 3004 5006

8 70003 2480.40 2012-10-10 3009 5003

9 70012 250.45 2012-06-27 3008 5002

10 70011 75.29 2012-08-17 3003 5007

11 70013 3045.60 2012-04-25 3002 5001

10. Write a Pandas program to split the following dataframe into groups based on customer id and create a list of order date for each group.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 2012-10-05 3005 5002

1 70009 270.65 2012-09-10 3001 5005

2 70002 65.26 2012-10-05 3002 5001

3 70004 110.50 2012-08-17 3009 5003

4 70007 948.50 2012-09-10 3005 5002

5 70005 2400.60 2012-07-27 3007 5001

6 70008 5760.00 2012-09-10 3002 5001

7 70010 1983.43 2012-10-10 3004 5006

8 70003 2480.40 2012-10-10 3009 5003

9 70012 250.45 2012-06-27 3008 5002

10 70011 75.29 2012-08-17 3003 5007

11 70013 3045.60 2012-04-25 3002 5001

11. Write a Pandas program to split the following dataframe into groups and calculate monthly purchase amount.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 05-10-2012 3001 5002

1 70009 270.65 09-10-2012 3001 5005

2 70002 65.26 05-10-2012 3005 5001

3 70004 110.50 08-17-2012 3001 5003

4 70007 948.50 10-09-2012 3005 5002

5 70005 2400.60 07-27-2012 3001 5001

6 70008 5760.00 10-09-2012 3005 5001

7 70010 1983.43 10-10-2012 3001 5006

8 70003 2480.40 10-10-2012 3005 5003

9 70012 250.45 06-17-2012 3001 5002

10 70011 75.29 07-08-2012 3005 5007

11 70013 3045.60 04-25-2012 3005 5001

12. Write a Pandas program to split the following dataframe into groups, group by month and year based on order date and find the total purchase amount year wise, month wise.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 05-10-2012 3001 5002

1 70009 270.65 09-10-2012 3001 5005

2 70002 65.26 05-10-2012 3005 5001

3 70004 110.50 08-17-2012 3001 5003

4 70007 948.50 10-09-2012 3005 5002

5 70005 2400.60 07-27-2012 3001 5001

6 70008 5760.00 10-09-2012 3005 5001

7 70010 1983.43 10-10-2012 3001 5006

8 70003 2480.40 10-10-2012 3005 5003

9 70012 250.45 06-17-2012 3001 5002

10 70011 75.29 07-08-2012 3005 5007

11 70013 3045.60 04-25-2012 3005 5001

13. Write a Pandas program to split the following dataframe into groups based on first column and set other column values into a list of values.

Test Data:

X Y Z

0 10 10 22

1 10 15 20

2 10 11 18

3 20 20 20

4 30 21 13

5 30 12 10

6 10 14 0

14. Write a Pandas program to split the following dataframe into groups based on all columns and calculate Groupby value counts on the dataframe.

Test Data:

id type book

0 1 10 Math

1 2 15 English

2 1 11 Physics

3 1 20 Math

4 2 21 English

5 1 12 Physics

6 2 14 English

Output:

book English Math Physics id type

1 10 0 1 0

11 0 0 1

12 0 0 1

20 0 1 0

2 14 1 0 0

15 1 0 0

21 1 0 0

15. Write a Pandas program to split the following dataframe into groups and count unique values of 'value' column.

Test Data:

id value

0 1 a

1 1 a

2 2 b

3 3 None

4 3 a

5 4 a

6 4 None

7 4 b

Output:

value

a 3

b 2

16. Write a Pandas program to split a given dataframe into groups and list all the keys from the GroupBy object.

Test Data:

school\_code class name date\_Of\_Birth age height weight

S1 s001 V Alberto Franco 15/05/2002 12 173 35

S2 s002 V Gino Mcneill 17/05/2002 12 192 32

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30

S5 s002 V Gino Mcneill 11/05/2002 14 151 31

S6 s004 VI David Parkes 15/09/1997 12 159 32

17. Write a Pandas program to split a given dataframe into groups and create a new column with count from GroupBy.

Test Data:

book\_name book\_type book\_id

0 Book1 Math 1

1 Book2 Physics 2

2 Book3 Computer 3

3 Book4 Science 4

4 Book1 Math 1

5 Book2 Physics 2

6 Book3 Computer 3

7 Book5 English 5

18. Write a Pandas program to split a given dataframe into groups with bin counts.

Test Data:

ord\_no purch\_amt customer\_id sales\_id

0 70001 150.50 3005 5002

1 70009 270.65 3001 5003

2 70002 65.26 3002 5004

3 70004 110.50 3009 5003

4 70007 948.50 3005 5002

5 70005 2400.60 3007 5001

6 70008 5760.00 3002 5005

7 70010 1983.43 3004 5007

8 70003 2480.40 3009 5008

9 70012 250.45 3008 5004

10 70011 75.29 3003 5005

11 70013 3045.60 3002 5001

19. Write a Pandas program to split a given dataframe into groups with multiple aggregations.

Split the following given dataframe by school code, class and get mean, min, and max value of height and age for each value of the school.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

20. Write a Pandas program to split a given dataframe into groups and display target column as a list of unique values.

Test Data:

id type book

0 A 1 Math

1 A 1 Math

2 A 1 English

3 A 1 Physics

4 A 2 Math

5 A 2 English

6 B 1 Physics

7 B 1 English

8 B 1 Physics

9 B 2 English

10 B 2 English

Output:

List all unique values in a group:

id type book

0 A 1 Math,English,Physics

1 A 2 Math,English

2 B 1 Physics,English

3 B 2 English

21. Write a Pandas program to split the following dataframe into groups and calculate quarterly purchase amount. Go to the editor

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 05-10-2012 3001 5002

1 70009 270.65 09-10-2012 3001 5005

2 70002 65.26 05-10-2012 3005 5001

3 70004 110.50 08-17-2012 3001 5003

4 70007 948.50 10-09-2012 3005 5002

5 70005 2400.60 07-27-2012 3001 5001

6 70008 5760.00 10-09-2012 3005 5001

7 70010 1983.43 10-10-2012 3001 5006

8 70003 2480.40 10-10-2012 3005 5003

9 70012 250.45 06-17-2012 3001 5002

10 70011 75.29 07-08-2012 3005 5007

11 70013 3045.60 04-25-2012 3005 5001

22. Write a Pandas program to split the following dataframe into groups by school code and get mean, min, and max value of age with customized column name for each school.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

23. Write a Pandas program to split the following datasets into groups on customer id and calculate the number of customers starting with 'C', the list of all products and the difference of maximum purchase amount and minimum purchase amount.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 05-10-2012 C3001 5002

1 70009 270.65 09-10-2012 C3001 5005

2 70002 65.26 05-10-2012 D3005 5001

3 70004 110.50 08-17-2012 D3001 5003

4 70007 948.50 10-09-2012 C3005 5002

5 70005 2400.60 07-27-2012 D3001 5001

6 70008 5760.00 10-09-2012 C3005 5001

7 70010 1983.43 10-10-2012 D3001 5006

8 70003 2480.40 10-10-2012 D3005 5003

9 70012 250.45 06-17-2012 C3001 5002

10 70011 75.29 07-08-2012 D3005 5007

11 70013 3045.60 04-25-2012 D3005 5001

24. Write a Pandas program to split the following datasets into groups on customer\_id to summarize purch\_amt and calculate percentage of purch\_amt in each group.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 05-10-2012 3001 5002

1 70009 270.65 09-10-2012 3001 5005

2 70002 65.26 05-10-2012 3005 5001

3 70004 110.50 08-17-2012 3001 5003

4 70007 948.50 10-09-2012 3005 5002

5 70005 2400.60 07-27-2012 3001 5001

6 70008 5760.00 10-09-2012 3005 5001

7 70010 1983.43 10-10-2012 3001 5006

8 70003 2480.40 10-10-2012 3005 5003

9 70012 250.45 06-17-2012 3001 5002

10 70011 75.29 07-08-2012 3005 5007

11 70013 3045.60 04-25-2012 3005 5001

25. Write a Pandas program to split a dataset, group by one column and get mean, min, and max values by group, also change the column name of the aggregated metric. Using the following dataset find the mean, min, and max values of purchase amount (purch\_amt) group by customer id (customer\_id).

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

26. Write a Pandas program to split a given dataset, group by two columns and convert other columns of the dataframe into a dictionary with column header as key.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

27. Write a Pandas program to split a given dataset, group by one column and apply an aggregate function to few columns and another aggregate function to the rest of the columns of the dataframe.

Test Data:

salesman\_id sale\_jan sale\_feb sale\_mar sale\_apr sale\_may sale\_jun \

0 5002 150.50 250.50 150.50 150.50 130.50 150.50

1 5005 270.65 170.65 270.65 270.65 270.65 270.65

2 5001 65.26 15.26 65.26 95.26 65.26 45.26

3 5003 110.50 110.50 110.50 210.50 310.50 110.50

4 5002 948.50 598.50 948.50 948.50 948.50 948.50

5 5001 2400.60 1400.60 2400.60 2400.60 2400.60 3400.60

6 5001 1760.00 2760.00 5760.00 760.00 760.00 5760.00

7 5006 2983.43 1983.43 1983.43 1983.43 1983.43 983.43

8 5003 480.40 2480.40 2480.40 2480.40 2480.40 2480.40

9 5002 1250.45 250.45 250.45 250.45 250.45 250.45

10 5007 75.29 75.29 75.29 75.29 75.29 75.29

11 5001 1045.60 3045.60 3045.60 3045.60 3045.60 3045.60

sale\_jul sale\_aug sale\_sep sale\_oct sale\_nov sale\_dec

0 950.50 150.50 150.50 150.50 150.50 150.50

1 270.65 70.65 270.65 270.65 270.65 70.65

2 65.26 65.26 65.26 65.26 95.26 65.26

3 210.50 110.50 110.50 110.50 110.50 110.50

4 948.50 948.50 948.50 948.50 948.50 948.50

5 2400.60 400.60 200.60 2400.60 2400.60 2400.60

6 5760.00 5760.00 5760.00 5760.00 5760.00 5760.00

7 983.43 1983.43 1983.43 1983.43 1983.43 1983.43

8 2480.40 2480.40 2480.40 2480.40 2480.40 2480.40

9 250.45 250.45 250.45 250.45 250.45 250.45

10 75.29 75.29 75.29 75.29 75.29 75.29

11 3045.60 3045.60 3045.60 3045.60 3045.60 3045.60

28. Write a Pandas program to split a given dataset, group by one column and remove those groups if all the values of a specific columns are not available.

Test Data:

school class name date\_Of\_Birth age height weight address

S1 s001 V Alberto Franco 15/05/2002 12 173 35 street1

S2 s002 V Gino Mcneill 17/05/2002 12 192 32 street2

S3 s003 VI Ryan Parkes 16/02/1999 13 186 33 street3

S4 s001 VI Eesha Hinton 25/09/1998 13 167 30 street1

S5 s002 V Gino Mcneill 11/05/2002 14 151 31 street2

S6 s004 VI David Parkes 15/09/1997 12 159 32 street4

29. Write a Pandas program to split a given dataset using group by on specified column into two labels and ranges.

Split the group on 'salesman\_id',

Ranges:

1) (5001...5006)

2) (5007..5012)

Test Data:

salesman\_id sale\_jan

0 5001 150.50

1 5002 270.65

2 5003 65.26

3 5004 110.50

4 5005 948.50

5 5006 2400.60

6 5007 1760.00

7 5008 2983.43

8 5009 480.40

9 5010 1250.45

10 5011 75.29

11 5012 1045.60

30. Write a Pandas program to split the following dataset using group by on first column and aggregate over multiple lists on second column.

Test Data:

student\_id marks

0 S001 [88, 89, 90]

1 S001 [78, 81, 60]

2 S002 [84, 83, 91]

3 S002 [84, 88, 91]

4 S003 [90, 89, 92]

5 S003 [88, 59, 90]

Output:

student\_id

S001 [83.0, 85.0, 75.0]

S002 [84.0, 85.5, 91.0]

S003 [89.0, 74.0, 91.0]

31. Write a Pandas program to split the following dataset using group by on 'salesman\_id' and find the first order date for each group.

Test Data:

ord\_no purch\_amt ord\_date customer\_id salesman\_id

0 70001 150.50 2012-10-05 3005 5002

1 70009 270.65 2012-09-10 3001 5005

2 70002 65.26 2012-10-05 3002 5001

3 70004 110.50 2012-08-17 3009 5003

4 70007 948.50 2012-09-10 3005 5002

5 70005 2400.60 2012-07-27 3007 5001

6 70008 5760.00 2012-09-10 3002 5001

7 70010 1983.43 2012-10-10 3004 5004

8 70003 2480.40 2012-10-10 3009 5003

9 70012 250.45 2012-06-27 3008 5002

10 70011 75.29 2012-08-17 3003 5004

11 70013 3045.60 2012-04-25