Home Assignment

Informatics Practices

Data Handling Using Pandas

Q.1 Write a Pandas program to drop those rows from a given DataFrame in which specific columns have missing values

Test Data:

	ord_no	purch_amt	ord_date	customer_id
0	NaN	NaN	NaN	NaN
1	NaN	270.65	2012-09-10	3001.0
2	70002.0	65.26	NaN	3001.0
3	NaN	NaN	NaN	NaN
4	NaN	948.50	2012-09-10	3002.0
5	70005.0	2400.60	2012-07-27	3001.0
6	NaN	5760.00	2012-09-10	3001.0
7	70010.0	1983.43	2012-10-10	3004.0
8	70003.0	2480.40	2012-10-10	3003.0
9	70012.0	250.45	2012-06-27	3002.0
10	NaN	75.29	2012-08-17	3001.0
11	NaN	NaN	NaN	NaN

Q2. Write a Pandas program to detect missing values of a given DataFrame. Display True or False.

Q3. Write a Pandas program to identify the column(s) of a given DataFrame which have at least one missing value.

Q4. Write a Pandas program to identify the column(s) of a given DataFrame which have at least one missing value.

Q5. 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

Q6. 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

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

Q.9. Write a Pandas program to create a Pivot table with multiple indexes from a given Saledata.csv file.

Q10. Write a Pandas program to create a Pivot table and find the total sale amount region wise, manager wise.

Q11. Write a Pandas program to display the dimensions or shape of the SaleData dataset. Also extract the column names from the dataset.

Q12. Write a Pandas program to remove the duplicates from 'region' column of Salesdata dataset