Pandas

Creating Series in Pandas

```
import pandas as pd
List = [10,20,30,40,50]
s = pd.Series(List)
print (s)
```

Creating Series in Pandas with index

```
l = [1,2,3,4,5]
p = pd.Series(l)
print(p)
p1 = pd.Series(l,index=["aaa","bbb","ccc","ddd","eee"])
print(p1)
print(p1["ccc"]) # returns value at index 'ccc'
```

Simple series creation using Pandas with index and reindex

```
mys1 = pd.Series([1,2,3,4,5], index=['a','b','c','d','e'])
print(mys1)
print(mys1.reindex(index = ['b','a','c','d','e']))
```

Simple series converting pandas series to list as displaying type
```
ds = pd.Series([2, 4, 6, 8, 10])
print(ds)
print(type(ds))
d = ds.tolist()
print(d)
print(type(d))
```

Simple series arithmetic operations on pandas series

```python
ds1 = pd.Series([2, 4, 6, 8, 10])
ds2 = pd.Series([1, 3, 5, 7, 9])
ds = ds1 + ds2
print(ds)
ds = ds1 - ds2
print(ds)
ds = ds1 * ds2
print(ds)
ds = ds1 / ds2
print(ds)
```

compare the elements of the two Pandas Series

```python
ds1 = pd.Series([2, 4, 6, 8, 10])
ds2 = pd.Series([1, 3, 5, 7, 10])
print("Series1:")
print(ds1)
print("Series2:")
print(ds2)
print("Compare the elements of the said Series:")
print("Equals:")
print(ds1 == ds2)
print("Greater than:")
print(ds1 > ds2)
print("Less than:")
print(ds1 < ds2)
```

Creating Dictionary to series

```python
d1 = {'a': 100, 'b': 200, 'c':300, 'd':400, 'e':800}
```

```python
print("Original dictionary:")
print(d1)
s1 = pd.Series(d1)
print("Converted series:")
print(s1)
```

#Creating pandas series from numpy array
```python
Import numpy as np
a = np.array([10, 20, 30, 40, 50])
print("NumPy array:")
print(a)
s1 = pd.Series(a)
print("Converted Pandas series:")
print(s1)
```

Change data type of column Series

```python
s1 = pd.Series(['100', '200', 'python', '300.12', '400'])
print("Original Data Series:")
print(s1)
print("Change the said data type to numeric:")
s2 = pd.to_numeric(s1, errors='coerce')
print(s2)
```

Converting dataFrame Column to a series

```python
d = {'col1': [1, 2, 3, 4, 7, 11], 'col2': [4, 5, 6, 9, 5, 0], 'col3': [7, 5, 8, 12, 1,11]}
df = pd.DataFrame(data=d)
print("Original DataFrame")
print(df)
s1 = df.ix[:,0]
print("\n1st column as a Series:")
print(s1)
```

```python
print(type(s1))
```

#Sorting Series

```python
s = pd.Series(['100', '200', 'python', '300.12', '400'])
print("Original Data Series:")
print(s)
new_s = pd.Series(s).sort_values(ascending=False)
print(new_s)
```

#Appending data to series

```python
s = pd.Series(['100', '200', 'python', '300.12', '400'])
print(s)
new_s = s.append(pd.Series(['500', 'php']))
print(new_s)
```

#Slicing series based on criteria

```python
s = pd.Series([0, 1,2,3,4,5,6,7,8,9,10])
print(s)
n = 6
new_s = s[s < n]
print(new_s)
```

Series aggregate functions

```python
s = pd.Series(data = [1,2,3,4,5,6,7,8,9,5,3])
print(s.mean())
print(s.mode())
print(s.median())
print(s.max())
print(s.min())
```

```python
# Creating DataFrames in Pandas

myData = {
  "Temperature": [20, 30, 90],
  "Humidity": [90, 80, 75]
}
Print(mydata)
df = pd.DataFrame(myData)
print(df)
print(df.loc[2])


# Creating Data Frames in Pandas using lst

data = [['AAA','Admin',10000],['BBB','Manager',12000],['CCC',Clerk,1300]]
df = pd.DataFrame(data,columns=['Names','Designation','Salary'])
print (df)


#Creating DataFrames with Named index

Df1 = pd.DataFrame(data, index = ["row1", "row2", "row3"])


#Creating and using Aggregate functions is DataFrames :

d = {'Name':pd.Series(['AAA','BBB','CCC','DDD','EEE']),
  'Age':pd.Series([25,26,25,23,24]),
  'Salary' :pd.Series([4000, 4567, 3456,2345,2323])
}
d = pd.DataFrame(d)
print (d["Salary"].sum())
print (d["Salary"].mean())
print(d["Salary"].median())
print(d["Salary"].min())
print(d["Salary"].max())
print(d["Salary"].count())
```

```python
print(d["Salary"].std())
print(d["Salary"].var())
print(d["Salary"].prod())
print(d["Salary"].describe()) # summarize stats of data
print(d["Salary"].mode())
```

Importing files in Pandas

```python
df = pd.read_csv('data.csv')
print(df)
```

#Reading and retrieving files :

```python
Datafile  = pd.read_csv('c:/users/data.csv')
print(Datafile)
print(Datafile.to_string())
print(Datafile.head()) # print first five rows
print(Datafile.head(10))
print(Datafile.tail()) # print last 5 rows
print(Datafile.info())  provides complete information of dataframes
```

Removing Empty cells

```python
new_df = df.dropna()
print(new_df.to_string())
```

Removing Null Values

```python
df.dropna(inplace = True)
print(df.to_string())
```

Fill Null Values

```python
df.fillna(130, inplace = True)
print(df.to_string())
```

Handling duplicates

```python
print(df.duplicated())
df.drop_duplicates(inplace = True)
```

#Write a Pandas program to import given excel data

```python
import pandas as pd
df = pd.read_excel("xyz.xlsx')
print(df.head)
```

#Write a Pandas program to read specific columns from a given excel file

```python
import pandas as pd
import numpy as np
cols = [1, 2, 4]
df = pd.read_excel('test.xlsx', usecols=cols)
print(df)
```

#Write a Pandas program to find the sum, mean, max, min value of Salary

```python
Create excel file with emno, ename, desig, salary,deptno, comm
import pandas as pd
import numpy as np
df = pd.read_excel("test.xlsx')
print("Sum: ",df["column"].sum())
print("Mean: ",df["column"].mean())
print("Maximum: ",df["column"].max())
print("Minimum: ",df["column"].min())
```

#Write a Pandas program to import excel data into a dataframe and find details where criteria Ex: "salary" > 20000 any value

```python
import pandas as pd
```

```
import numpy as np
df = pd.read_excel(test.xlsx')
df[df["salary"] > 20000].head()
df.query('name  == ["hello"]').head()
```

Write a Pandas program to import excel data into a dataframe and sort data

```
import pandas as pd
import numpy as np
df = pd.read_excel(e.xlsx')
result = df.sort_values(column)
print(result)
```

#Inpoerting data from Excel in pandas and exporting to another excel

```
import pandas as pd
df1 = pd.read_excel('test1.xlsx',sheet_name=0)
df2 = pd.read_excel(test1,sheet_name=1)
df3 = pd.read_excel(test1,sheet_name=2)
df = pd.concat([df1, df2, df3])
df.to_excel('test2', index=False)
```

Exercise : For the given question use pandas to analyze

Given the excel file Order breakdown analyze data using pandas the following:

1. Retrieve all the records in an ascending order by product ID and create a separate file.
```

2. From the new file find the top ten profitable and lowest ten profitable categories.

3. Find the mean, median and mode of the sales and display.

4. Display the total quantity of sales.

5. Display all the sales which has crossed $800.