Pandas

Import pandas as pd

Sal =Pd.read\_csv(‘salaries’.csv)

Sal.head()

Sal.info()

sal["BasePay"].describe()

sal["BasePay"].describe().transpose()

Sal[‘JobTitle’].unique()

Sal[‘JobTitle’].nunique()

Sal[‘Basepay’].max()

Sal[‘Basepay’].min()

Sal[‘Basepay’].mean()

Sal[‘jobtitle’].value\_count.head(5)

Sal.groupby(“Year”).mean() ## average salary per year

Sal.dropna() ## removes all rows and columns with NA values

Sal.dropna(axis=1) ## removes columns with NA

Sal.dropna(thresh= 2) ## will drop NA with min 2 NA in arrow

Sal.fillna(value =”AAA”)

Sal[“BasePay”].fillna(value = sal[“BasePay”].mean())

Pd.concat([df1,df2,df3])

Pd.concat([df1,df2,df3],axis=1) ## concat row wise

Pd.merge(df1,df2,how=’inner’,on=’key’) ## similar to joins in sql

Pd.merge(df1,df2,how=’inner’,on=[’key1’,’key2’])

Df1.join(df2)

Df1.join(df2,how=’outer’)