

```

import pandas as pd

from sklearn.decomposition import PCA

path='empleadosRETO.csv'
query=pd.read_csv(path)
EmpleadosAttrition=query
col_names=EmpleadosAttrition.columns
print(col_names)

print(EmpleadosAttrition[['EmployeeCount','EmployeeNumber','Over18','StandardHours']].head(10))

EmpleadosAttrition.drop(['EmployeeCount','EmployeeNumber','Over18','StandardHours'], axis=1, inplace=True)

print(EmpleadosAttrition['HiringDate'].str[-4:].sort_values(ascending=False))

EmpleadosAttrition['Year']=EmpleadosAttrition['HiringDate'].str[-4:].astype(int)

print(EmpleadosAttrition['Year'].sort_values(ascending=True))

print(EmpleadosAttrition['TotalWorkingYears'].head(15).sort_values(ascending=False))

EmpleadosAttrition['YearsAtCompany']=2018-EmpleadosAttrition['Year']

print(EmpleadosAttrition["YearsAtCompany"].head(35))

EmpleadosAttrition=EmpleadosAttrition.rename(columns={'DistanceFromHome':'DistanceFromHome_Km'})

print(EmpleadosAttrition['DistanceFromHome_Km'].head(10))

EmpleadosAttrition['DistanceFromHome']=EmpleadosAttrition['DistanceFromHome_Km'].str[:2].astype(int)

print(EmpleadosAttrition['DistanceFromHome'].head(10))

EmpleadosAttrition.drop(['Year','HiringDate','DistanceFromHome_Km'], axis=1, inplace=True)

print(EmpleadosAttrition.columns)

print(EmpleadosAttrition['Department'])

SueldoPromDepto=EmpleadosAttrition[['MonthlyIncome','Department']]

print(SueldoPromDepto)

SueldoPromedio=SueldoPromDepto.groupby('Department').mean()

```

```

print(SueldoPromedio)

print(EmpleadosAttrition['MonthlyIncome'])

EmpleadosAttrition['MonthlyIncome_Norm']=EmpleadosAttrition['MonthlyIncome']/EmpleadosAttrition['MonthlyIncome'].abs().max()

#print('\n',EmpleadosAttrition['MonthlyIncome_Norm'].head(20))

print('\n',round(EmpleadosAttrition['MonthlyIncome_Norm'], 4).head(40))

corr=EmpleadosAttrition[['BusinessTravel','Department','EducationField','Gender','JobRole','MaritalStatus','Attrition']].apply(lambda x: pd.factorize(x)[0]).head(40)

print(EmpleadosAttrition.columns)

print(EmpleadosAttrition.columns)

print(EmpleadosAttrition['Attrition'])

print(corr)

print('\nBusiness Travel\n',corr['BusinessTravel'].corr(corr['Attrition']))

print('\nDepartment\n',corr['Department'].corr(corr['Attrition']))

print('\nEducationField\n',corr['EducationField'].corr(corr['Attrition']))

print('\nGender\n',corr['Gender'].corr(corr['Attrition']))

print('\nJobRole\n',corr['JobRole'].corr(corr['Attrition']))

print('\nMaritalStatus\n',corr['MaritalStatus'].corr(corr['Attrition']))

EmpleadosAttritionFinal=EmpleadosAttrition.drop(['BusinessTravel','Department','MaritalStatus'],axis=1)

print(EmpleadosAttritionFinal.columns)

EmpleadosAttritionFinal.to_csv('EmpleadosAttritionFinal.csv',index=False)

pca=PCA()

pca.fit(corr)

print(pca.components_)

print(pca.explained_variance_)

print(pca.explained_variance_ratio_)

EmpleadosAttritionFinalPCA=pca.transform(corr)

print(EmpleadosAttritionFinalPCA)

```