

1-2

```
8 #t1
9
10 import pandas as pd
11 from datetime import datetime
12
13 df_emp = pd.read_csv('employees.csv', header=0, dtype={'phone1':str, 'phone2':str})
14 df_dep = pd.read_csv('departments.csv', header=0)
15
16 print (df_emp.describe())
17 print (df_emp.info())
18
19 print (df_emp['phone2'].isnull())
20
21 df = pd.merge(df_emp, df_dep, how='left', on='dep')
22
23 df.drop(labels='image', inplace=True, axis=1)
24
25
26 #t2
27
28 emp_count = df.shape[0]
29
30 m_count = sum(df.gender==0)
31 f_count = sum(df.gender==1)
32
33 m_pros = round(m_count / emp_count * 100, 1)
34 f_pros = round(f_count / emp_count * 100, 1)
35
36 sal_min = df['salary'].min()
37 sal_max = df['salary'].max()
38 sal_mean = round(df['salary'].mean(), 2)
39
40 sal_mean_tk = df[df['dname']=='Tuotekehitys']['salary'].mean()
41
42 count_no_phone2 = sum(df['phone2'].isnull())
43
44 df['age'] = (datetime.now() - pd.to_datetime(df['bdate'])) // TimeoutError()
45
46 bins=[]
47
48 for i in range(15,75,5):
49     bins.append(i)
50
51 labels = bins[1:]
52 #labels = bins.copy()
53
54 df['age_group'] = pd.cut(df['age'], bins=bins, labels=labels, right=True)
```

count_no_phone2	int	1	10
df	DataFrame	(15, 11)	Column names: id, fname, lname, salary, bdate, email, dep, phone1, pho
df_dep	DataFrame	(5, 2)	Column names: dep, dname
df_emp	DataFrame	(15, 11)	Column names: id, fname, lname, salary, bdate, email, dep, phone1, pho
emp_count	int	1	15
f_count	int	1	5
m_count	int	1	10
m_pros	float	1	33.3
sal_max	int64	1	10000
sal_mean	float64	1	3123.33
sal_mean_tk	float64	1	2787.5
sal_min	int64	1	2000

T3-4

```

1  import pandas as pd
2
3  df_tit_data = pd.read_csv('Titanic_data.csv', header=0)
4  df_tit_names = pd.read_csv('Titanic_names.csv', header=0)
5
6  print(df_tit_data.describe())
7  print(df_tit_data.info())
8
9  print(df_tit_names.describe())
10 print(df_tit_names.info())
11
12 print(df_tit_data.hist(bins = 4))
13
14 df = pd.merge(df_tit_data, df_tit_names, how='inner', on='id')

```

