**Code**

**import** pandas **as** pd

data = pd.read\_csv(**"turnover.csv"**)

data.info()

print(data.head())

print(data.salary.unique())

*# Change the type of the "salary" column to categorical*

data.salary = data.salary.astype(**'category'**)

*# Provide the correct order of categories*

data.salary = data.salary.cat.reorder\_categories([**'low'**,**'medium'**,**'high'**])

*# Encode categories with integer values*

data.salary = data.salary.cat.codes

print(data.head())

departments = pd.get\_dummies(data.department)

print(departments)

departments = departments.drop(**"technical"**, axis = 1)

print(departments.head())

*# Get the total number of observations and save it*

n\_employees = len(data)

*# Print the number of employees who left/stayed*

print(data.churn.value\_counts())

*# Print the percentage of employees who left/stayed*

print(data.churn.value\_counts()/n\_employees\*100)

**import** matplotlib.pyplot **as** plt

**import** seaborn **as** sns

corr\_matrix = data.corr()

sns.heatmap(corr\_matrix)

plt.show()

**Output**

**Connected to pydev debugger (build 193.6015.41)**

**<class 'pandas.core.frame.DataFrame'>**

**RangeIndex: 14999 entries, 0 to 14998**

**Data columns (total 10 columns):**

**# Column Non-Null Count Dtype**

**--- ------ -------------- -----**

**0 satisfaction\_level 14999 non-null float64**

**1 last\_evaluation 14999 non-null float64**

**2 number\_project 14999 non-null int64**

**3 average\_montly\_hours 14999 non-null int64**

**4 time\_spend\_company 14999 non-null int64**

**5 Work\_accident 14999 non-null int64**

**6 churn 14999 non-null int64**

**7 promotion\_last\_5years 14999 non-null int64**

**8 department 14999 non-null object**

**9 salary 14999 non-null object**

**dtypes: float64(2), int64(6), object(2)**

**memory usage: 1.1+ MB**

**satisfaction\_level last\_evaluation ... department salary**

**0 0.38 0.53 ... sales low**

**1 0.80 0.86 ... sales medium**

**2 0.11 0.88 ... sales medium**

**3 0.72 0.87 ... sales low**

**4 0.37 0.52 ... sales low**

**[5 rows x 10 columns]**

**['low' 'medium' 'high']**

**satisfaction\_level last\_evaluation ... department salary**

**0 0.38 0.53 ... sales 0**

**1 0.80 0.86 ... sales 1**

**2 0.11 0.88 ... sales 1**

**3 0.72 0.87 ... sales 0**

**4 0.37 0.52 ... sales 0**

**[5 rows x 10 columns]**

**IT RandD accounting hr ... product\_mng sales support technical**

**0 0 0 0 0 ... 0 1 0 0**

**1 0 0 0 0 ... 0 1 0 0**

**2 0 0 0 0 ... 0 1 0 0**

**3 0 0 0 0 ... 0 1 0 0**

**4 0 0 0 0 ... 0 1 0 0**

**... .. ... ... .. ... ... ... ... ...**

**14994 0 0 0 0 ... 0 0 1 0**

**14995 0 0 0 0 ... 0 0 1 0**

**14996 0 0 0 0 ... 0 0 1 0**

**14997 0 0 0 0 ... 0 0 1 0**

**14998 0 0 0 0 ... 0 0 1 0**

**[14999 rows x 10 columns]**

**IT RandD accounting hr ... marketing product\_mng sales support**

**0 0 0 0 0 ... 0 0 1 0**

**1 0 0 0 0 ... 0 0 1 0**

**2 0 0 0 0 ... 0 0 1 0**

**3 0 0 0 0 ... 0 0 1 0**

**4 0 0 0 0 ... 0 0 1 0**

**[5 rows x 9 columns]**

**0 11428**

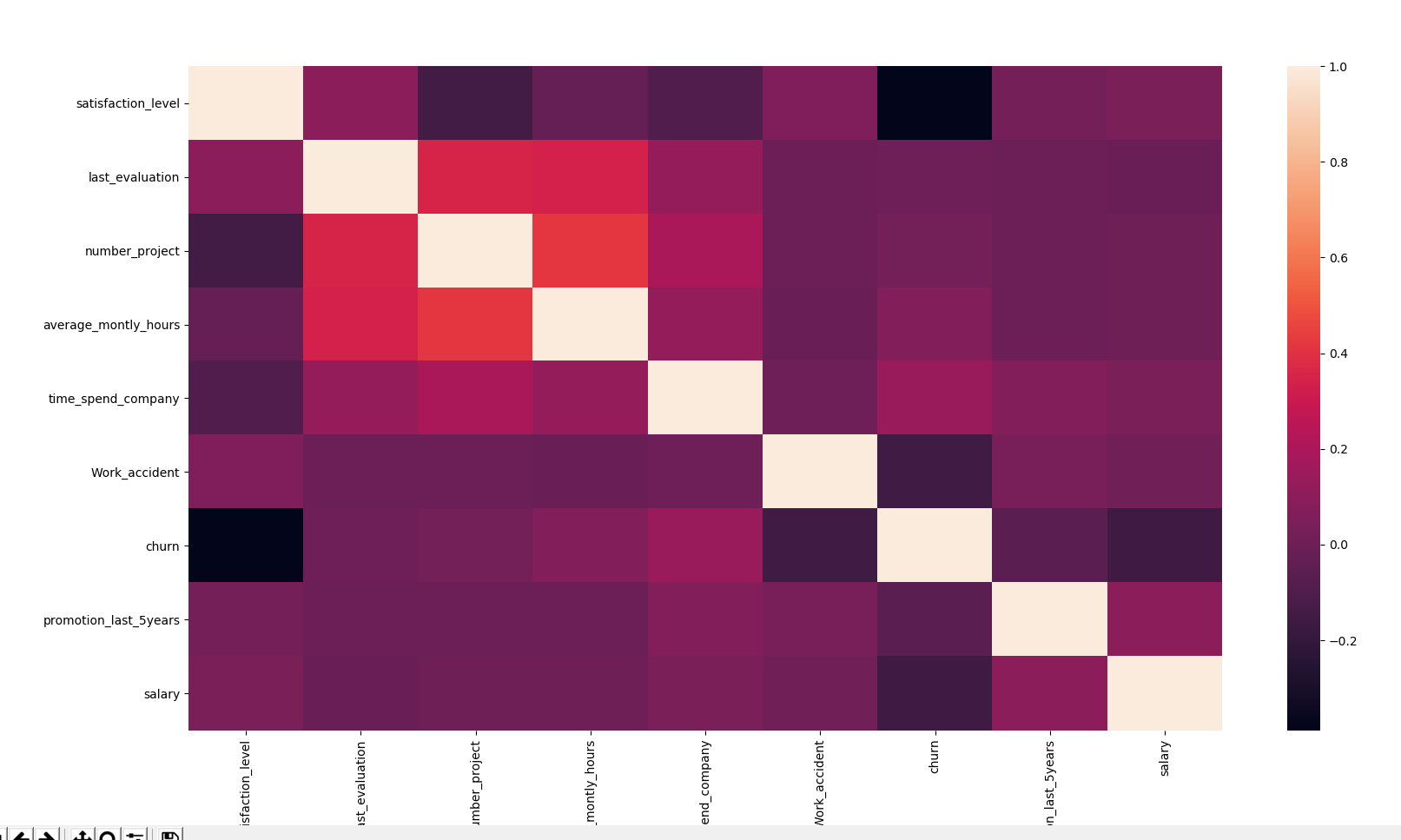
**1 3571**

**Name: churn, dtype: int64**

**0 76.191746**

**1 23.808254**

**Name: churn, dtype: float64**

****