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

**import** sqlalchemy

**import** pprint

**import** psycopg2

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

**%matplotlib** inline

**from** sqlalchemy **import** create\_engine

In [9]:

*# connect to local database*

**from** sqlalchemy **import** create\_engine

engine **=** create\_engine('postgresql+psycopg2://postgres:postgres@localhost:5433/Homework')

connection **=** engine**.**connect()

In [12]:

departments **=** pd**.**read\_sql('select \* from departments', connection)

departments

employees **=** pd**.**read\_sql('select \* from employees', connection, parse\_dates**=**['birth\_date', 'hire\_date'])

employees**.**head()

*### Import the Departments table"*

departments **=** pd**.**read\_sql('select \* from departments', connection)

departments**.**head()

Out[19]:

|  | **dept\_no** | **dept\_name** |
| --- | --- | --- |
| **0** | d001 | Marketing |
| **1** | d002 | Finance |
| **2** | d003 | Human Resources |
| **3** | d004 | Production |
| **4** | d005 | Development |

In [21]:

*### Import the Salaries table*

salaries **=** pd**.**read\_sql('select \* from salaries', connection, parse\_dates**=**['from\_date', 'to\_date'])

salaries**.**head()

*### Import the Department Manager table*

dept\_manager **=** pd**.**read\_sql('select \* from dept\_manager', connection, parse\_dates**=**['from\_date', 'to\_date'])

dept\_manager**.**head()

Out[23]:

|  | **dept\_no** | **emp\_no** | **from\_date** | **to\_date** |
| --- | --- | --- | --- | --- |
| **0** | d001 | 110022 | 1985-01-01 | 1991-10-01 |
| **1** | d001 | 110039 | 1991-10-01 | NaT |
| **2** | d002 | 110085 | 1985-01-01 | 1989-12-17 |
| **3** | d002 | 110114 | 1989-12-17 | NaT |
| **4** | d003 | 110183 | 1985-01-01 | 1992-03-21 |

In [25]:

*###Replace null dates*

dept\_manager**.**to\_date **=** dept\_manager['to\_date']**.**fillna(pd**.**to\_datetime('2050-12-31'))

dept\_manager**.**head()

Out[25]:

|  | **dept\_no** | **emp\_no** | **from\_date** | **to\_date** |
| --- | --- | --- | --- | --- |
| **0** | d001 | 110022 | 1985-01-01 | 1991-10-01 |
| **1** | d001 | 110039 | 1991-10-01 | 2050-12-31 |
| **2** | d002 | 110085 | 1985-01-01 | 1989-12-17 |
| **3** | d002 | 110114 | 1989-12-17 | 2050-12-31 |
| **4** | d003 | 110183 | 1985-01-01 | 1992-03-21 |

In [27]:

*### Import the Titles table*

titles **=** pd**.**read\_sql('select \* from titles', connection, parse\_dates**=**['from\_date', 'to\_date'])

titles**.**head()

*#### Replace null dates*

titles**.**to\_date **=** titles['to\_date']**.**fillna(pd**.**to\_datetime('2050-12-31'))

titles**.**head()

Out[29]:

|  | **emp\_no** | **title** | **from\_date** | **to\_date** |
| --- | --- | --- | --- | --- |
| **0** | 10001 | Senior Engineer | 1986-06-26 | 2050-12-31 |
| **1** | 10002 | Staff | 1996-08-03 | 2050-12-31 |
| **2** | 10003 | Senior Engineer | 1995-12-03 | 2050-12-31 |
| **3** | 10004 | Engineer | 1986-12-01 | 1995-12-01 |
| **4** | 10004 | Senior Engineer | 1995-12-01 | 2050-12-31 |

In [30]:

*#### Create a merged dataframe of titles and salaries*

employee\_salaries **=** titles**.**merge(salaries, on**=**'emp\_no')

employee\_salaries**.**head()

Out[30]:

|  | **emp\_no** | **title** | **from\_date\_x** | **to\_date\_x** | **salary** | **from\_date\_y** | **to\_date\_y** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 10001 | Senior Engineer | 1986-06-26 | 2050-12-31 | 60117 | 1986-06-26 | 1987-06-26 |
| **1** | 10002 | Staff | 1996-08-03 | 2050-12-31 | 65828 | 1996-08-03 | 1997-08-03 |
| **2** | 10003 | Senior Engineer | 1995-12-03 | 2050-12-31 | 40006 | 1995-12-03 | 1996-12-02 |
| **3** | 10004 | Engineer | 1986-12-01 | 1995-12-01 | 40054 | 1986-12-01 | 1987-12-01 |
| **4** | 10004 | Senior Engineer | 1995-12-01 | 2050-12-31 | 40054 | 1986-12-01 | 1987-12-01 |

In [35]:

*### Creating Dataframe with emp\_no, title, salary*

employee\_salaries\_df **=** employee\_salaries[['emp\_no', 'title', 'salary']]

employee\_salaries\_df**.**head()

Out[35]:

|  | **emp\_no** | **title** | **salary** |
| --- | --- | --- | --- |
| **0** | 10001 | Senior Engineer | 60117 |
| **1** | 10002 | Staff | 65828 |
| **2** | 10003 | Senior Engineer | 40006 |
| **3** | 10004 | Engineer | 40054 |
| **4** | 10004 | Senior Engineer | 40054 |

In [37]:

employee\_salaries\_df**.**groupby('title')['salary']**.**mean()**.**round(2)

employee\_salaries\_df**.**head()

Icon

Description automatically generated with medium confidence