localhost:8892/notebooks/SQL-Challenge/homework/SqlEmployees.ipynb#

Apps | Welcome to the Da... | GitHub | Stack Overflow - W... | Slack | Get Started |... | RutgersCodingBoot... | Video Library | Boot... | Welcome to opens... | Other bookmarks

jupyter SqlEmployees Last Checkpoint: 2 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [22]:
```
from config import username, password
import pandas as pd
import matplotlib.pyplot as plt
```

In [23]:
```
from sqlalchemy import create_engine
engine = create_engine(f'postgresql://{username}:{password}@localhost:5432/employees_db')
connection = engine.connect()
```

In [3]:
```
emp_df = pd.read_sql('select * from employees', connection)
emp_df.head()
```

Out[3]:

	emp_no	emp_title_id	birth_date	first_name	last_name	sex	hire_date
0	473302	s0001	1953-07-25	Hideyuki	Zallocco	M	1990-04-28
1	475053	e0002	1954-11-18	Byong	Delgrande	F	1991-09-07
2	57444	e0002	1958-01-30	Berry	Babb	F	1992-03-21
3	421786	s0001	1957-09-28	Xiong	Verhoeff	M	1987-11-26
4	282238	e0003	1952-10-28	Abdelkader	Baumann	F	1991-01-18

In [4]:
```
salary_df = pd.read_sql('select * from salaries', connection)
salary_df.head(2)
```

Out[4]:

	emp_no	salary
0	10001	60117
1	10002	65828

In [5]:
```
title_df = pd.read_sql('select * from titles', connection)
title_df
```

Out[5]:

	title_id	title
0	s0001	Staff
1	s0002	Senior Staff
2	e0001	Assistant Engineer
3	e0002	Engineer
4	e0003	Senior Engineer
5	e0004	Technique Leader
6	m0001	Manager

In [6]:
```
titles_df = title_df.rename(columns = {"title_id": "emp_title_id"})
titles_df
```

Out[6]:

	emp_title_id	title
0	s0001	Staff
1	s0002	Senior Staff
2	e0001	Assistant Engineer
3	e0002	Engineer
4	e0003	Senior Engineer
5	e0004	Technique Leader
6	m0001	Manager

In [7]:
```
EmpJoinTitles = pd.merge(titles_df, emp_df, on="emp_title_id", how = 'inner' )
EmpJoinTitles
```

Out[7]:

	emp_title_id	title	emp_no	birth_date	first_name	last_name	sex	hire_date
0	s0001	Staff	473302	1953-07-25	Hideyuki	Zallocco	M	1990-04-28
1	s0001	Staff	421786	1957-09-28	Xiong	Verhoeff	M	1987-11-26
2	s0001	Staff	273487	1957-04-14	Christoph	Parfitt	M	1991-06-28
300020	m0001	Manager	111133	1955-03-16	Hauke	Zhang	M	1986-12-30
300021	m0001	Manager	110085	1959-10-28	Ebru	Alpin	M	1985-01-01
300022	m0001	Manager	110386	1953-10-04	Shem	Kieras	M	1988-10-14
300023	m0001	Manager	111784	1956-06-14	Marjo	Giarratana	F	1988-02-12

300024 rows × 8 columns

In [8]:
```
EmpSal = pd.merge(EmpJoinTitles, salary_df, on="emp_no", how = 'inner' )
EmpSal
```

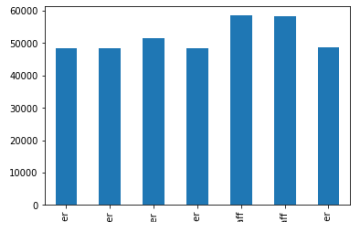
Out[8]:

	emp_title_id	title	emp_no	birth_date	first_name	last_name	sex	hire_date	salary
0	s0001	Staff	473302	1953-07-25	Hideyuki	Zallocco	M	1990-04-28	40000
1	s0001	Staff	421786	1957-09-28	Xiong	Verhoeff	M	1987-11-26	40000
2	s0001	Staff	273487	1957-04-14	Christoph	Parfitt	M	1991-06-28	56087
3	s0001	Staff	246449	1958-03-23	Subbu	Bultermann	F	1988-03-25	87084
4	s0001	Staff	48085	1964-01-19	Venkatesan	Gilg	M	1993-06-28	63016
...	...	...	...	...	...	...	...	...	...
300019	m0001	Manager	111400	1959-11-09	Arie	Staelin	M	1985-01-01	72446
300020	m0001	Manager	111133	1955-03-16	Hauke	Zhang	M	1986-12-30	71148
300021	m0001	Manager	110085	1959-10-28	Ebru	Alpin	M	1985-01-01	60026
300022	m0001	Manager	110386	1953-10-04	Shem	Kieras	M	1988-10-14	40000

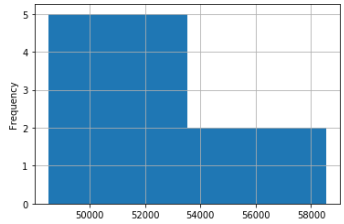
In [9]:
```
meansalbytitle = EmpSal.groupby("title").mean()["salary"].round(2)
meansalbytitle
```

Out[9]:
```
title
Assistant Engineer    48564.43
Engineer              48535.34
Manager               51531.04
Senior Engineer       48506.80
Senior Staff          58550.17
Staff                 58465.38
Technique Leader      48582.90
Name: salary, dtype: float64
```

In [27]:
```
meansalbytitle.plot.bar()
plt.savefig("../barplot")
```



In [28]:
```
meansalbytitle.plot.hist(bins=2)
plt.grid()
plt.savefig("../histogram")
plt.show()
```



In [ ]: