In [1]:
```
# Import all dependant libraries & connect to postgres database called pweletthac
import pandas as pd
import sqlalchemy
import psycopg2
import matplotlib as plt
%matplotlib inline
%matplotlib notebook
import matplotlib.pyplot as plt
import numpy as np
from sqlalchemy import create_engine
engine = create_engine('postgresql://postgres:password@localhost:5432/pewletthacl
connection = engine.connect()
```

In [2]:
```
# Run an SQL select all from employees table to get a peek into the data.
pd.read_sql_query('''SELECT * FROM employees;''', engine)
```

Out[2]:

	emp_no	birth_date	first_name	last_name	gender	hire_date
0	10001	1953-09-02	Georgi	Facello	M	1986-06-26
1	10002	1964-06-02	Bezalel	Simmel	F	1985-11-21
2	10003	1959-12-03	Parto	Bamford	M	1986-08-28
3	10004	1954-05-01	Chirstian	Koblick	M	1986-12-01
4	10005	1955-01-21	Kyoichi	Maliniak	M	1989-09-12
5	10006	1953-04-20	Anneke	Preusig	F	1989-06-02
6	10007	1957-05-23	Tzvetan	Zielinski	F	1989-02-10
7	10008	1958-02-19	Saniya	Kalloufi	M	1994-09-15
8	10009	1952-04-19	Sumant	Peac	F	1985-02-18
9	10010	1963-06-01	Duangkaew	Piveteau	F	1989-08-24
10	10011	1953-11-07	Mary	Sluis	F	1990-01-22
11	10012	1960-10-04	Patricio	Bridgland	M	1992-12-18
12	10013	1963-06-07	Eberhardt	Terkki	M	1985-10-20
13	10014	1956-02-12	Berni	Genin	M	1987-03-11
14	10015	1959-08-19	Guoxiang	Nooteboom	M	1987-07-02
15	10016	1961-05-02	Kazuhito	Cappelletti	M	1995-01-27
16	10017	1958-07-06	Cristinel	Bouloucos	F	1993-08-03
17	10018	1954-06-19	Kazuhide	Peha	F	1987-04-03
18	10019	1953-01-23	Lillian	Haddadi	M	1999-04-30
19	10020	1952-12-24	Mayuko	Warwick	M	1991-01-26
20	10021	1960-02-20	Ramzi	Erde	M	1988-02-10
21	10022	1952-07-08	Shahaf	Famili	M	1995-08-22
22	10023	1953-09-29	Bojan	Montemayor	F	1989-12-17
23	10024	1958-09-05	Suzette	Pettey	F	1997-05-19
24	10025	1958-10-31	Prasadram	Heyers	M	1987-08-17
25	10026	1953-04-03	Yongqiao	Berztiss	M	1995-03-20
26	10027	1962-07-10	Divier	Reistad	F	1989-07-07
27	10028	1963-11-26	Domenick	Tempesti	M	1991-10-22
28	10029	1956-12-13	Otmar	Herbst	M	1985-11-20
29	10030	1958-07-14	Elvis	Demeyer	M	1994-02-17
...	...	...	...	...	...	...
299994	499970	1963-03-25	Danai	Hedayat	M	1994-08-06
299995	499971	1963-12-28	Uwe	Uludag	M	1989-02-26

	emp_no	birth_date	first_name	last_name	gender	hire_date
299996	499972	1957-07-25	Katsuo	Leuchs	F	1989-11-23
299997	499973	1963-06-03	Lobel	Taubman	M	1994-02-01
299998	499974	1956-09-10	Shuichi	Piazza	F	1989-09-16
299999	499975	1952-11-09	Masali	Chorvat	M	1992-01-23
300000	499976	1963-08-20	Guozhong	Felder	M	1988-12-26
300001	499977	1956-06-05	Martial	Weisert	F	1996-09-17
300002	499978	1960-03-29	Chiranjit	Kuzuoka	M	1990-05-24
300003	499979	1962-10-29	Prasadram	Waleschkowski	M	1994-01-04
300004	499980	1959-06-28	Gino	Usery	M	1991-02-11
300005	499981	1955-01-02	Yunming	Mitina	F	1991-03-07
300006	499982	1954-08-25	Mohammed	Pleszkun	M	1986-02-21
300007	499983	1955-08-29	Uri	Juneja	F	1989-08-28
300008	499984	1959-08-31	Kaijung	Rodham	M	1985-09-11
300009	499985	1964-12-26	Gila	Lukaszewicz	M	1997-02-11
300010	499986	1952-07-22	Nathan	Ranta	F	1985-08-11
300011	499987	1961-09-05	Rimli	Dusink	F	1998-09-20
300012	499988	1962-09-28	Bangqing	Kleiser	F	1986-06-06
300013	499989	1954-05-26	Keiichiro	Lindqvist	M	1993-10-28
300014	499990	1963-11-03	Khaled	Kohling	M	1985-10-10
300015	499991	1962-02-26	Pohua	Sichman	F	1989-01-12
300016	499992	1960-10-12	Siamak	Salverda	F	1987-05-10
300017	499993	1963-06-04	DeForest	Mullainathan	M	1997-04-07
300018	499994	1952-02-26	Navin	Argence	F	1990-04-24
300019	499995	1958-09-24	Dekang	Lichtner	F	1993-01-12
300020	499996	1953-03-07	Zito	Baaz	M	1990-09-27
300021	499997	1961-08-03	Berhard	Lenart	M	1986-04-21
300022	499998	1956-09-05	Patricia	Breugel	M	1993-10-13
300023	499999	1958-05-01	Sachin	Tsukuda	M	1997-11-30

300024 rows × 6 columns

In [3]:
```
# create a dataframe called salaries to store the results from the SQL select al
df_salaries = pd.read_sql_query('''SELECT * FROM salaries;''', engine)
df_salaries
```

Out[3]:

	emp_no	salary	from_date	to_date
0	10001	60117.0	1986-06-26	1987-06-26
1	10002	65828.0	1996-08-03	1997-08-03
2	10003	40006.0	1995-12-03	1996-12-02
3	10004	40054.0	1986-12-01	1987-12-01
4	10005	78228.0	1989-09-12	1990-09-12
5	10006	40000.0	1990-08-05	1991-08-05
6	10007	56724.0	1989-02-10	1990-02-10
7	10008	46671.0	1998-03-11	1999-03-11
8	10009	60929.0	1985-02-18	1986-02-18
9	10010	72488.0	1996-11-24	1997-11-24
10	10011	42365.0	1990-01-22	1991-01-22
11	10012	40000.0	1992-12-18	1993-12-18
12	10013	40000.0	1985-10-20	1986-10-20
13	10014	46168.0	1993-12-29	1994-12-29
14	10015	40000.0	1992-09-19	1993-08-22
15	10016	70889.0	1998-02-11	1999-02-11
16	10017	71380.0	1993-08-03	1994-08-03
17	10018	55881.0	1987-04-03	1988-04-02
18	10019	44276.0	1999-04-30	2000-04-29
19	10020	40000.0	1997-12-30	1998-12-30
20	10021	55025.0	1988-02-10	1989-02-09
21	10022	40000.0	1999-09-03	2000-09-02
22	10023	47883.0	1999-09-27	2000-09-26
23	10024	83733.0	1998-06-14	1999-06-14
24	10025	40000.0	1987-08-17	1988-08-16
25	10026	47585.0	1995-03-20	1996-03-19
26	10027	40000.0	1995-04-02	1996-04-01
27	10028	48859.0	1991-10-22	1992-10-21
28	10029	63163.0	1991-09-18	1992-09-17
29	10030	66956.0	1994-02-17	1995-02-17
...	...	...	...	...
299994	499970	99187.0	1995-01-09	1996-01-09

	emp_no	salary	from_date	to_date
299995	499971	41007.0	1997-03-05	1998-03-05
299996	499972	41002.0	1989-11-23	1990-11-23
299997	499973	40000.0	1994-02-01	1995-02-01
299998	499974	40000.0	1996-12-11	1997-12-11
299999	499975	40000.0	1995-11-21	1996-11-20
300000	499976	81721.0	1988-12-26	1989-12-26
300001	499977	70216.0	1999-12-28	2000-12-27
300002	499978	40000.0	1996-07-09	1997-07-09
300003	499979	47933.0	1996-08-29	1997-08-29
300004	499980	90395.0	1991-02-11	1992-02-11
300005	499981	42720.0	1991-03-07	1992-03-06
300006	499982	40000.0	1999-04-08	2000-04-07
300007	499983	40000.0	1995-07-15	1996-07-14
300008	499984	40000.0	1989-07-30	1990-07-30
300009	499985	40000.0	1997-02-11	1998-02-11
300010	499986	91988.0	1985-08-11	1986-08-11
300011	499987	52282.0	1999-12-21	2000-12-19
300012	499988	74890.0	1988-07-25	1989-07-25
300013	499989	64183.0	1996-10-29	1997-10-29
300014	499990	40000.0	1999-10-28	2000-10-27
300015	499991	43612.0	1997-02-11	1998-02-11
300016	499992	43005.0	1987-05-10	1988-05-09
300017	499993	40000.0	1997-04-07	1998-04-07
300018	499994	40000.0	1993-02-22	1993-10-27
300019	499995	40000.0	1997-06-02	1998-06-02
300020	499996	58058.0	1996-05-13	1997-05-13
300021	499997	49597.0	1987-08-30	1988-08-29
300022	499998	40000.0	1993-12-27	1994-12-27
300023	499999	63707.0	1997-11-30	1998-11-30

300024 rows × 4 columns

In [4]:
```
# create a new dataframe and insert the title from titles database, calculate the
df_salaries2 = pd.read_sql_query('''select t.title as "Title", round(sum(salary),
join titles t on t.emp_no = s.emp_no
group by t.title;''', engine)
df_salaries2
```

Out[4]:

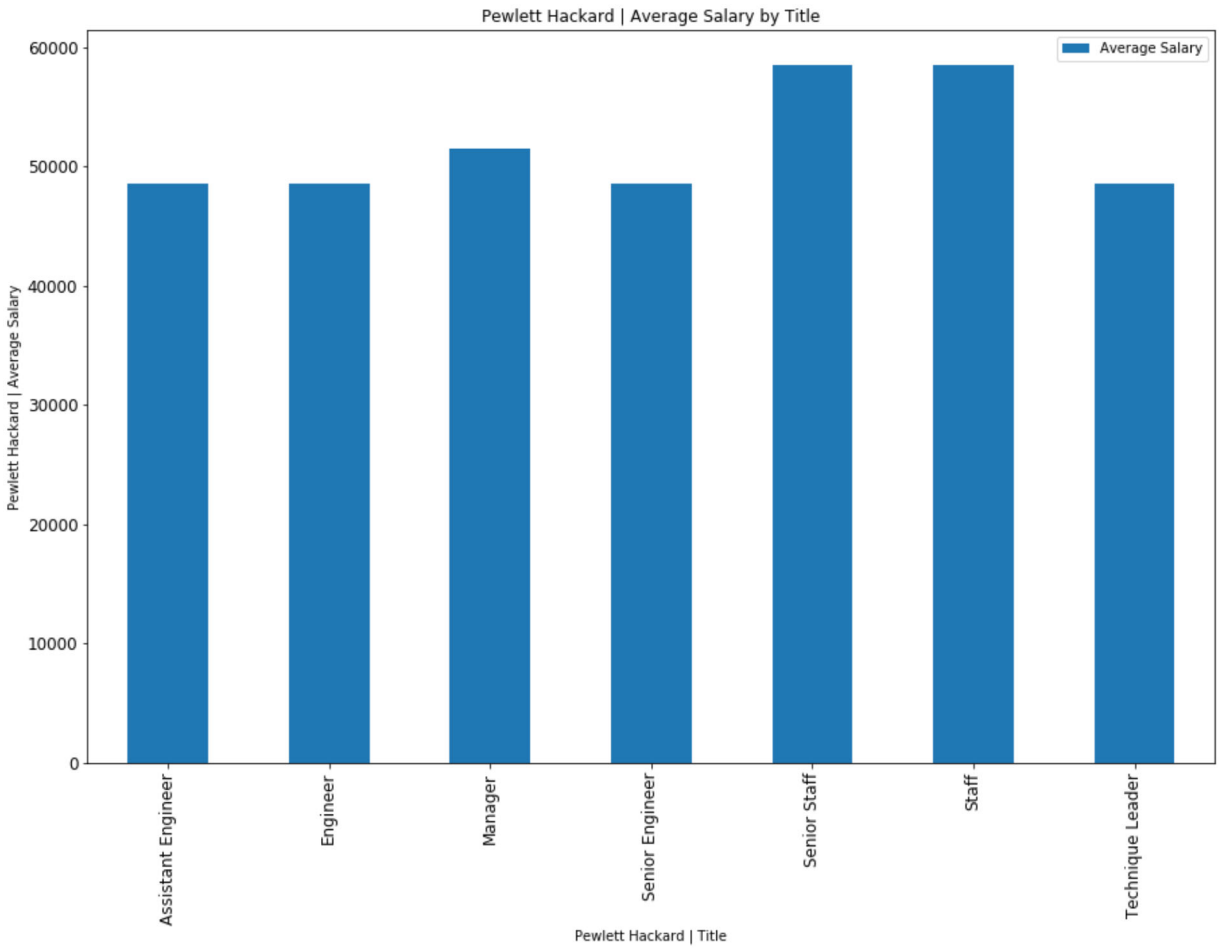
	Title	Average Salary
0	Assistant Engineer	48493.20
1	Engineer	48539.78
2	Manager	51531.04
3	Senior Engineer	48506.75
4	Senior Staff	58503.29
5	Staff	58465.27
6	Technique Leader	48580.51

In [5]:
```
# set the Index to title inplace by dropping the index.
# https://www.dataquest.io/blog/adding-axis-labels-to-plots-with-pandas-and-matpl

df_salaries2.set_index('Title',drop=True, inplace=True)
df_salaries2
```

Out[5]:

Title	Average Salary
Assistant Engineer	48493.20
Engineer	48539.78
Manager	51531.04
Senior Engineer	48506.75
Senior Staff	58503.29
Staff	58465.27
Technique Leader	48580.51

In [6]:
```
# plot dataframe for the average salary as a bar chart with the appropriate labe
df_salaries2.plot.bar()
ax = df_salaries2['Average Salary'].plot(kind='bar', title ="Pewlett Hackard | A
plt.xlabel("Pewlett Hackard | Title")
plt.ylabel("Pewlett Hackard | Average Salary")

plt.show()
```



In [ ]: