

```
In [1]: # import dependencies
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sqlalchemy import create_engine
```

```
In [2]: engine = create_engine('postgres://postgres:fakepassword@localhost:5432/sql_h
w')
conn = engine.connect()
```

```
In [3]: # review employees table
employees_df = salaries_df = pd.read_sql("SELECT * FROM employees", conn)
employees_df.head()
```

Out[3]:

	emp_no	title_id	birth_date	first_name	last_name	sex	hire_date
0	473302	s0001	1953-07-25	Hideyuki	Zalocco	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]: # review the salary table
salaries_df = pd.read_sql("SELECT * FROM salaries", conn)
salaries_df.head()
```

Out[4]:

	emp_no	salary
0	10001	60117
1	10002	65828
2	10003	40006
3	10004	40054
4	10005	78228

```
In [5]: # merge employees table with salaries table
emp_salaries_df = pd.merge(employees_df, salaries_df, on="emp_no", how="inner"
)

#emp_salaries_df = emp_salaries_df.rename(columns = {"title_id_x": "title_i
d"})
emp_salaries_df.head()
```

Out[5]:

	emp_no	title_id	birth_date	first_name	last_name	sex	hire_date	salary
0	473302	s0001	1953-07-25	Hideyuki	Zalocco	M	1990-04-28	40000
1	475053	e0002	1954-11-18	Byong	Delgrande	F	1991-09-07	53422
2	57444	e0002	1958-01-30	Berry	Babb	F	1992-03-21	48973
3	421786	s0001	1957-09-28	Xiong	Verhoeff	M	1987-11-26	40000
4	282238	e0003	1952-10-28	Abdelkader	Baumann	F	1991-01-18	40000

```
In [6]: # check titles table
titles_df = pd.read_sql("SELECT * FROM titles", conn)
titles_df.head()
```

Out[6]:

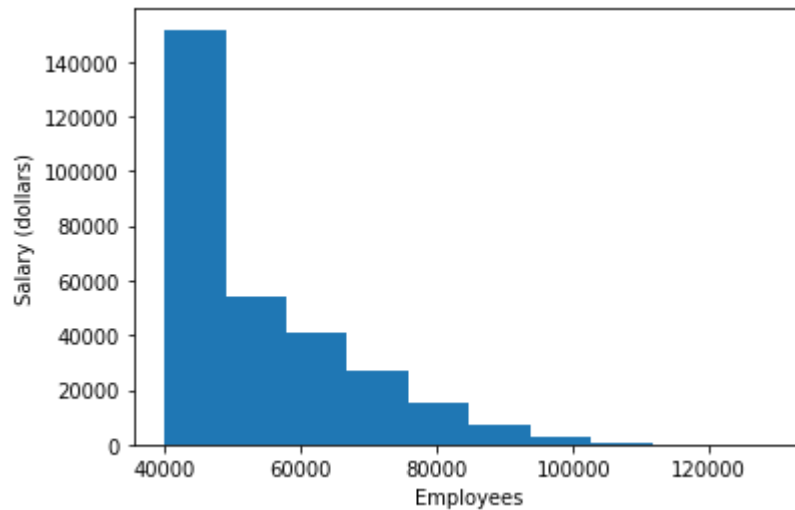
	title_id	title
0	s0001	Staff
1	s0002	Senior Staff
2	e0001	Assistant Engineer
3	e0002	Engineer
4	e0003	Senior Engineer

```
In [7]: # merge employees, salaries, and titles
emp_salaries_title_df = pd.merge(emp_salaries_df, titles_df, on="title_id", ho
w="inner")
emp_salaries_title_df.head()
```

Out[7]:

	emp_no	title_id	birth_date	first_name	last_name	sex	hire_date	salary	title
0	473302	s0001	1953-07-25	Hideyuki	Zalocco	M	1990-04-28	40000	Staff
1	421786	s0001	1957-09-28	Xiong	Verhoeff	M	1987-11-26	40000	Staff
2	273487	s0001	1957-04-14	Christoph	Parfitt	M	1991-06-28	56087	Staff
3	246449	s0001	1958-03-23	Subbu	Bultermann	F	1988-03-25	87084	Staff
4	48085	s0001	1964-01-19	Venkatesan	Gilg	M	1993-06-28	63016	Staff

```
In [10]: plt.hist(emp_salaries_title_df.salary)
plt.xlabel("Employees")
plt.ylabel("Salary (dollars)")
plt.show()
```



```
In [16]: # Find the average salary per title
title_grouped = round(emp_salaries_title_df.groupby('title').mean(),2)
title_grouped
```

Out[16]:

	emp_no	salary
title		
Assistant Engineer	252992.60	48564.43
Engineer	253654.35	48535.34
Manager	110780.83	51531.04
Senior Engineer	253038.75	48506.80
Senior Staff	254481.80	58550.17
Staff	253408.50	58465.38
Technique Leader	251811.43	48582.90

```
In [20]: # Clean up the dataframe
clean_title_grouped = title_grouped.drop(columns = 'emp_no')
clean_title_grouped
clean_title_grouped = clean_title_grouped.reset_index()
clean_title_grouped
```

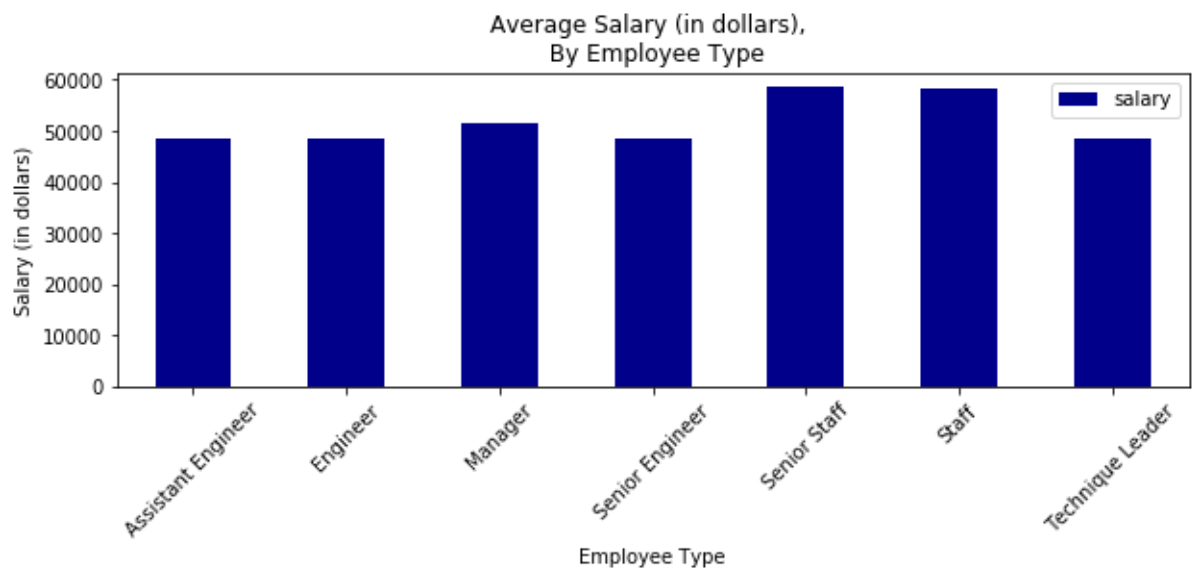
Out[20]:

	title	salary
0	Assistant Engineer	48564.43
1	Engineer	48535.34
2	Manager	51531.04
3	Senior Engineer	48506.80
4	Senior Staff	58550.17
5	Staff	58465.38
6	Technique Leader	48582.90

```
In [37]: # Use DataFrame.plot() in order to create a bar chart of the data
figure = clean_title_grouped.plot(kind="bar", facecolor="darkblue", figsize=(10,3))
figure.set_xticklabels(clean_title_grouped["title"], rotation=45)

# Set a title for the chart
plt.title("Average Salary (in dollars), \n By Employee Type")
plt.xlabel("Employee Type")
plt.ylabel("Salary (in dollars)")

plt.show()
plt.tight_layout()
```



<Figure size 432x288 with 0 Axes>

In [ ]: