

employees_db

October 23, 2019

```
[1]: import pandas as pd
from sqlalchemy import create_engine
import matplotlib.pyplot as plt

[2]: engine = create_engine('postgresql+psycopg2://postgres:postgres@localhost/
    ↪employees_db')
conn = engine.connect()
salaries = pd.read_sql("SELECT * FROM salaries", conn)
salaries.head()

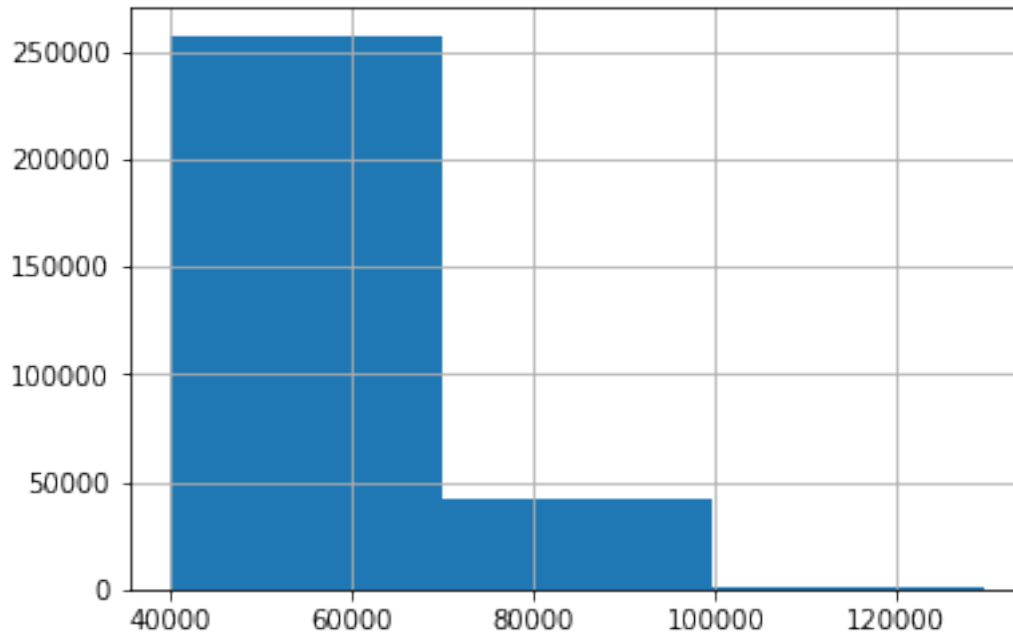
[2]:   emp_no      salary  from_date  to_date
0   10001  $60,117.00  1986-06-26  1987-06-26
1   10002  $65,828.00  1996-08-03  1997-08-03
2   10003  $40,006.00  1995-12-03  1996-12-02
3   10004  $40,054.00  1986-12-01  1987-12-01
4   10005  $78,228.00  1989-09-12  1990-09-12

[3]: salaries.dtypes

[3]: emp_no      int64
salary      object
from_date    object
to_date      object
dtype: object

[4]: employees = pd.read_sql("SELECT * FROM employees", conn)
departments = pd.read_sql("SELECT * FROM departments", conn)
dept_emp = pd.read_sql("SELECT * FROM dept_emp", conn)
dept_manager = pd.read_sql("SELECT * FROM dept_manager", conn)
titles = pd.read_sql("SELECT * FROM titles", conn)

[5]: #salaries.salary.value_counts()
#salaries.nlargest(50,['salary'])
#salaries.salary.count()
salaries['salary'] = salaries['salary'].replace('[\$,]', '', regex=True).
    ↪astype(float)
hist = salaries['salary'].hist(bins=3)
```



```
[6]: titles.head()
```

```
[6]:   emp_no      title  from_date  to_date
0  10001  Senior Engineer  1986-06-26  9999-01-01
1  10002      Staff  1996-08-03  9999-01-01
2  10003  Senior Engineer  1995-12-03  9999-01-01
3  10004      Engineer  1986-12-01  1995-12-01
4  10004  Senior Engineer  1995-12-01  9999-01-01
```

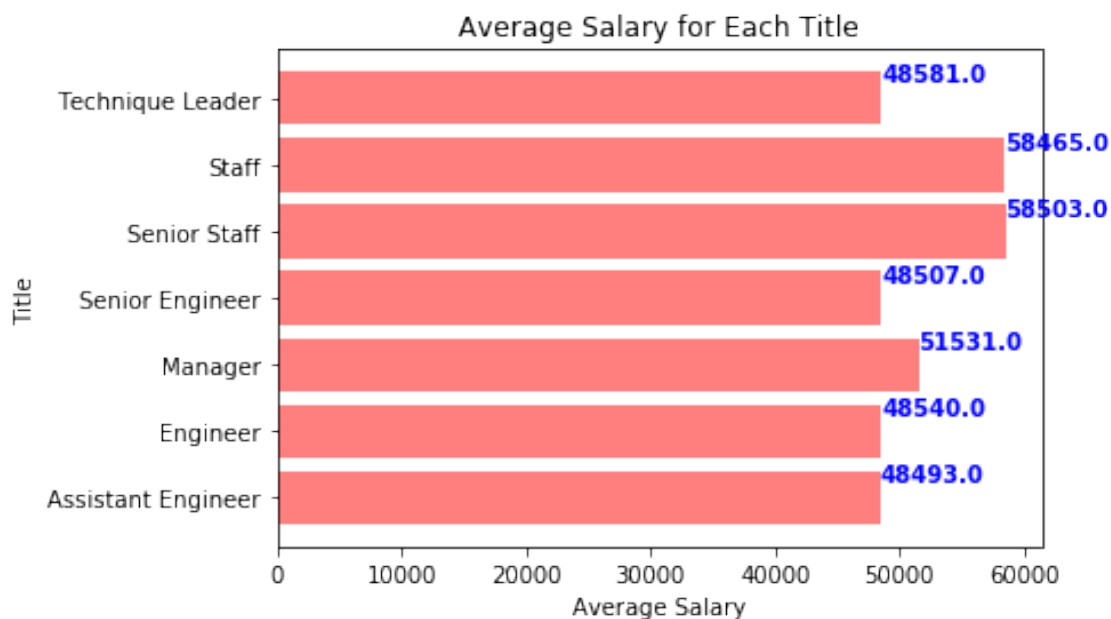
```
[20]: ts = pd.merge(salaries,titles,on='emp_no',how='outer')
ts[ts.columns[1]].replace('[\$,]', '', regex=True).astype(float)
#ts['title'].unique()
ts_mean = ts.groupby('title').mean()
ts_mean
ts_mean.reset_index(level=0, inplace=True)
ts_mean = ts_mean.round()
ts_mean
```

```
[20]:      title      emp_no  salary
0  Assistant Engineer  251495.0  48493.0
1      Engineer  252943.0  48540.0
2      Manager  110781.0  51531.0
3  Senior Engineer  253034.0  48507.0
4  Senior Staff  253423.0  58503.0
5      Staff  253399.0  58465.0
6  Technique Leader  251709.0  48581.0
```

```
[21]: #ts_mean['salary'] = ts_mean['salary'].map("${:.2f}".format)
ts_mean['salary']
ts_mean.dtypes
```

```
[21]: title      object
emp_no    float64
salary    float64
dtype: object
```

```
[23]: fig,ax = plt.subplots()
ax.barh(ts_mean['title'], ts_mean['salary'],color='r',alpha = 0.5, align = 'center')
plt.ylabel("Title")
plt.xlabel("Average Salary")
plt.title("Average Salary for Each Title")
for i,v in enumerate(ts_mean['salary']):
    ax.text(v, i + .25, str(v), color = 'blue', fontweight = 'bold')
plt.show()
```



```
[10]: ets = pd.merge(employees, ts, on = 'emp_no', how = "left")
ets.head()
```

```
[10]:  emp_no  birth_date first_name last_name gender  hire_date  salary  \
0   10001  1953-09-02    Georgi   Facello      M  1986-06-26  60117.0
1   10002  1964-06-02   Bezalel   Simmel      F  1985-11-21  65828.0
2   10003  1959-12-03     Parto   Bamford      M  1986-08-28  40006.0
3   10004  1954-05-01  Chirstian  Koblick      M  1986-12-01  40054.0
4   10004  1954-05-01  Chirstian  Koblick      M  1986-12-01  40054.0
```

	from_date_x	to_date_x		title	from_date_y	to_date_y
0	1986-06-26	1987-06-26	Senior	Engineer	1986-06-26	9999-01-01
1	1996-08-03	1997-08-03		Staff	1996-08-03	9999-01-01
2	1995-12-03	1996-12-02	Senior	Engineer	1995-12-03	9999-01-01
3	1986-12-01	1987-12-01		Engineer	1986-12-01	1995-12-01
4	1986-12-01	1987-12-01	Senior	Engineer	1995-12-01	9999-01-01

```
[11]: ets_junction = pd.merge(ets, dept_empt, on = 'emp_no', how = "left")
ets_junction.head()
```

```
[11]: emp_no birth_date first_name last_name gender hire_date salary \
0 10001 1953-09-02 Georgi Facello M 1986-06-26 60117.0
1 10002 1964-06-02 Bezalel Simmel F 1985-11-21 65828.0
2 10003 1959-12-03 Parto Bamford M 1986-08-28 40006.0
3 10004 1954-05-01 Chirstian Koblick M 1986-12-01 40054.0
4 10004 1954-05-01 Chirstian Koblick M 1986-12-01 40054.0
```

	from_date_x	to_date_x		title	from_date_y	to_date_y	dept_no	\
0	1986-06-26	1987-06-26	Senior	Engineer	1986-06-26	9999-01-01	d005	
1	1996-08-03	1997-08-03		Staff	1996-08-03	9999-01-01	d007	
2	1995-12-03	1996-12-02	Senior	Engineer	1995-12-03	9999-01-01	d004	
3	1986-12-01	1987-12-01		Engineer	1986-12-01	1995-12-01	d004	
4	1986-12-01	1987-12-01	Senior	Engineer	1995-12-01	9999-01-01	d004	

	from_date	to_date
0	1986-06-26	9999-01-01
1	1996-08-03	9999-01-01
2	1995-12-03	9999-01-01
3	1986-12-01	9999-01-01
4	1986-12-01	9999-01-01

```
[27]: etsd = pd.merge(ets_junction, departments, on = 'dept_no', how = 'left')
etsd_sort = etsd.sort_values(by = "emp_no",
                             axis=0,
                             ascending=True,
                             inplace=False,
                             kind='quicksort',
                             na_position='last')
etsd_sort.tail(100)
```

```
[27]: emp_no birth_date first_name last_name gender hire_date \
489803 499937 1955-07-28 Pantung Litzler F 1987-02-03
489805 499938 1960-06-30 JiYoung Binding F 1993-04-18
489804 499938 1960-06-30 JiYoung Binding F 1993-04-18
489806 499939 1952-02-22 Haldun Zaumen M 1988-11-27
489807 499940 1952-02-12 Yolla Auria M 1986-04-25
489808 499941 1956-10-16 Udaiprakash Hockney F 1994-02-04
489809 499942 1963-01-10 April Foolsday F 1997-02-10
```

489810	499943	1956-08-04	Ziyad	Baaz	M	1987-09-09
489811	499943	1956-08-04	Ziyad	Baaz	M	1987-09-09
489812	499944	1964-12-18	Bernd	Baumann	M	1988-07-03
489813	499945	1958-08-12	Sivanarayana	Haddadi	M	1987-03-02
489814	499945	1958-08-12	Sivanarayana	Haddadi	M	1987-03-02
489815	499946	1956-08-10	Barton	Soicher	F	1986-02-21
489816	499946	1956-08-10	Barton	Soicher	F	1986-02-21
489818	499947	1960-02-06	Conrado	Koyama	F	1989-02-19
489817	499947	1960-02-06	Conrado	Koyama	F	1989-02-19
489819	499948	1953-05-24	Cordelia	Paludetto	M	1993-01-28
489820	499949	1962-10-21	Krister	Stranks	M	1987-01-31
489821	499949	1962-10-21	Krister	Stranks	M	1987-01-31
489822	499950	1964-07-15	Weidon	Gente	F	1991-06-05
489823	499951	1962-08-17	Kasidit	Krohm	M	1985-03-17
489824	499952	1953-11-19	Aron	Lindqvist	M	1990-01-07
489825	499953	1960-09-04	Leszek	Pulkowski	M	1985-02-23
489826	499953	1960-09-04	Leszek	Pulkowski	M	1985-02-23
489827	499954	1961-09-24	Tuval	Tempesti	M	1990-08-09
489828	499954	1961-09-24	Tuval	Tempesti	M	1990-08-09
489829	499955	1961-10-23	Kankanahalli	Zucker	M	1993-03-06
489830	499955	1961-10-23	Kankanahalli	Zucker	M	1993-03-06
489831	499955	1961-10-23	Kankanahalli	Zucker	M	1993-03-06
489832	499955	1961-10-23	Kankanahalli	Zucker	M	1993-03-06
...
489873	499982	1954-08-25	Mohammed	Pleszkun	M	1986-02-21
489876	499983	1955-08-29	Uri	Juneja	F	1989-08-28
489874	499983	1955-08-29	Uri	Juneja	F	1989-08-28
489875	499983	1955-08-29	Uri	Juneja	F	1989-08-28
489877	499983	1955-08-29	Uri	Juneja	F	1989-08-28
489879	499984	1959-08-31	Kaijung	Rodham	M	1985-09-11
489878	499984	1959-08-31	Kaijung	Rodham	M	1985-09-11
489880	499985	1964-12-26	Gila	Lukaszewicz	M	1997-02-11
489881	499986	1952-07-22	Nathan	Ranta	F	1985-08-11
489882	499986	1952-07-22	Nathan	Ranta	F	1985-08-11
489883	499987	1961-09-05	Rimli	Dusink	F	1998-09-20
489884	499988	1962-09-28	Bangqing	Kleiser	F	1986-06-06
489885	499988	1962-09-28	Bangqing	Kleiser	F	1986-06-06
489886	499989	1954-05-26	Keiichiro	Lindqvist	M	1993-10-28
489887	499990	1963-11-03	Khaled	Kohling	M	1985-10-10
489888	499991	1962-02-26	Pohua	Sichman	F	1989-01-12
489891	499992	1960-10-12	Siamak	Salverda	F	1987-05-10
489892	499992	1960-10-12	Siamak	Salverda	F	1987-05-10
489889	499992	1960-10-12	Siamak	Salverda	F	1987-05-10
489890	499992	1960-10-12	Siamak	Salverda	F	1987-05-10
489893	499993	1963-06-04	DeForest	Mullainathan	M	1997-04-07
489894	499994	1952-02-26	Navin	Argence	F	1990-04-24
489895	499995	1958-09-24	Dekang	Lichtner	F	1993-01-12

489896	499996	1953-03-07	Zito	Baaz	M	1990-09-27
489897	499996	1953-03-07	Zito	Baaz	M	1990-09-27
489898	499997	1961-08-03	Berhard	Lenart	M	1986-04-21
489899	499997	1961-08-03	Berhard	Lenart	M	1986-04-21
489901	499998	1956-09-05	Patricia	Breugel	M	1993-10-13
489900	499998	1956-09-05	Patricia	Breugel	M	1993-10-13
489902	499999	1958-05-01	Sachin	Tsukuda	M	1997-11-30

	salary	from_date_x	to_date_x		title	from_date_y \
489803	40000.0	1987-02-03	1988-02-03	Senior Engineer	1993-02-02	
489805	44765.0	1999-10-07	2000-10-06	Staff	1999-10-07	
489804	44765.0	1999-10-07	2000-10-06	Staff	1999-10-07	
489806	62807.0	1999-12-25	2000-12-24	Staff	1999-12-25	
489807	43417.0	1986-04-25	1987-04-25	Senior Staff	1986-04-25	
489808	50799.0	1994-02-04	1995-02-04	Assistant Engineer	1994-02-04	
489809	40000.0	1998-03-28	1999-03-28	Technique Leader	1998-03-28	
489810	40000.0	1989-04-04	1990-04-04	Engineer	1989-04-04	
489811	40000.0	1989-04-04	1990-04-04	Senior Engineer	1997-04-03	
489812	40000.0	1992-09-10	1993-09-10	Senior Staff	1992-09-10	
489813	48097.0	1992-07-22	1993-07-22	Engineer	1992-07-22	
489814	48097.0	1992-07-22	1993-07-22	Senior Engineer	2001-07-22	
489815	45059.0	1993-12-10	1994-12-10	Engineer	1993-12-10	
489816	45059.0	1993-12-10	1994-12-10	Senior Engineer	2001-12-10	
489818	53325.0	1992-03-29	1992-07-09	Engineer	1992-03-29	
489817	53325.0	1992-03-29	1992-07-09	Engineer	1992-03-29	
489819	40000.0	1999-04-09	2000-04-08	Engineer	1999-04-09	
489820	54147.0	1991-07-17	1992-07-16	Engineer	1991-07-17	
489821	54147.0	1991-07-17	1992-07-16	Senior Engineer	1997-07-16	
489822	54788.0	1999-02-02	2000-02-02	Staff	1999-02-02	
489823	40000.0	1999-06-18	2000-06-17	Staff	1999-06-18	
489824	48431.0	1990-01-07	1991-01-07	Senior Engineer	1990-01-07	
489825	85507.0	1985-02-23	1986-02-23	Senior Staff	1994-02-23	
489826	85507.0	1985-02-23	1986-02-23	Staff	1985-02-23	
489827	40000.0	1990-08-09	1991-08-09	Engineer	1990-08-09	
489828	40000.0	1990-08-09	1991-08-09	Senior Engineer	1995-08-09	
489829	53698.0	1993-03-06	1994-03-06	Senior Staff	2000-03-06	
489830	53698.0	1993-03-06	1994-03-06	Senior Staff	2000-03-06	
489831	53698.0	1993-03-06	1994-03-06	Staff	1993-03-06	
489832	53698.0	1993-03-06	1994-03-06	Staff	1993-03-06	
...	
489873	40000.0	1999-04-08	2000-04-07	Staff	1999-04-08	
489876	40000.0	1995-07-15	1996-07-14	Staff	1995-07-15	
489874	40000.0	1995-07-15	1996-07-14	Senior Staff	2001-07-14	
489875	40000.0	1995-07-15	1996-07-14	Senior Staff	2001-07-14	
489877	40000.0	1995-07-15	1996-07-14	Staff	1995-07-15	
489879	40000.0	1989-07-30	1990-07-30	Senior Engineer	1994-07-30	
489878	40000.0	1989-07-30	1990-07-30	Engineer	1989-07-30	

489880	40000.0	1997-02-11	1998-02-11		Staff	1997-02-11
489881	91988.0	1985-08-11	1986-08-11	Senior	Staff	1992-08-11
489882	91988.0	1985-08-11	1986-08-11		Staff	1985-08-11
489883	52282.0	1999-12-21	2000-12-19		Staff	1999-12-21
489884	74890.0	1988-07-25	1989-07-25	Senior	Staff	1997-07-25
489885	74890.0	1988-07-25	1989-07-25		Staff	1988-07-25
489886	64183.0	1996-10-29	1997-10-29	Senior	Staff	1996-10-29
489887	40000.0	1999-10-28	2000-10-27		Engineer	1999-10-28
489888	43612.0	1997-02-11	1998-02-11		Staff	1997-02-11
489891	43005.0	1987-05-10	1988-05-09		Staff	1987-05-10
489892	43005.0	1987-05-10	1988-05-09		Staff	1987-05-10
489889	43005.0	1987-05-10	1988-05-09	Senior	Staff	1992-05-09
489890	43005.0	1987-05-10	1988-05-09	Senior	Staff	1992-05-09
489893	40000.0	1997-04-07	1998-04-07		Engineer	1997-04-07
489894	40000.0	1993-02-22	1993-10-27		Engineer	1993-02-22
489895	40000.0	1997-06-02	1998-06-02		Engineer	1997-06-02
489896	58058.0	1996-05-13	1997-05-13		Engineer	1996-05-13
489897	58058.0	1996-05-13	1997-05-13	Senior	Engineer	2002-05-13
489898	49597.0	1987-08-30	1988-08-29		Engineer	1987-08-30
489899	49597.0	1987-08-30	1988-08-29	Senior	Engineer	1992-08-29
489901	40000.0	1993-12-27	1994-12-27		Staff	1993-12-27
489900	40000.0	1993-12-27	1994-12-27	Senior	Staff	1998-12-27
489902	63707.0	1997-11-30	1998-11-30		Engineer	1997-11-30

	to_date_y	dept_no	from_date	to_date	dept_name
489803	9999-01-01	d005	1987-02-03	9999-01-01	Development
489805	9999-01-01	d009	1999-10-07	2000-02-17	Customer Service
489804	9999-01-01	d005	2000-02-17	9999-01-01	Development
489806	2001-03-03	d001	1999-12-25	2001-03-03	Marketing
489807	9999-01-01	d003	1986-04-25	9999-01-01	Human Resources
489808	9999-01-01	d005	1994-02-04	9999-01-01	Development
489809	9999-01-01	d005	1998-03-28	9999-01-01	Development
489810	1997-04-03	d005	1989-04-04	9999-01-01	Development
489811	9999-01-01	d005	1989-04-04	9999-01-01	Development
489812	9999-01-01	d008	1992-09-10	9999-01-01	Research
489813	2001-07-22	d004	1992-07-22	9999-01-01	Production
489814	9999-01-01	d004	1992-07-22	9999-01-01	Production
489815	2001-12-10	d004	1993-12-10	9999-01-01	Production
489816	9999-01-01	d004	1993-12-10	9999-01-01	Production
489818	1992-07-09	d009	1992-05-24	1992-07-09	Customer Service
489817	1992-07-09	d004	1992-03-29	1992-05-24	Production
489819	9999-01-01	d004	1999-04-09	9999-01-01	Production
489820	1997-07-16	d006	1991-07-17	9999-01-01	Quality Management
489821	9999-01-01	d006	1991-07-17	9999-01-01	Quality Management
489822	2000-02-23	d002	1999-02-02	2000-02-23	Finance
489823	9999-01-01	d003	1999-06-18	9999-01-01	Human Resources
489824	9999-01-01	d004	1990-01-07	9999-01-01	Production

489825	9999-01-01	d007	1985-02-23	9999-01-01	Sales
489826	1994-02-23	d007	1985-02-23	9999-01-01	Sales
489827	1995-08-09	d004	1990-08-09	9999-01-01	Production
489828	9999-01-01	d004	1990-08-09	9999-01-01	Production
489829	9999-01-01	d005	1996-10-02	9999-01-01	Development
489830	9999-01-01	d009	1993-03-06	1996-10-02	Customer Service
489831	2000-03-06	d005	1996-10-02	9999-01-01	Development
489832	2000-03-06	d009	1993-03-06	1996-10-02	Customer Service
...
489873	2000-05-16	d009	1999-04-08	2000-05-16	Customer Service
489876	2001-07-14	d004	2001-12-09	9999-01-01	Production
489874	9999-01-01	d004	2001-12-09	9999-01-01	Production
489875	9999-01-01	d009	1995-07-15	2001-12-09	Customer Service
489877	2001-07-14	d009	1995-07-15	2001-12-09	Customer Service
489879	9999-01-01	d004	1989-07-30	9999-01-01	Production
489878	1994-07-30	d004	1989-07-30	9999-01-01	Production
489880	9999-01-01	d008	1997-02-11	9999-01-01	Research
489881	9999-01-01	d007	1985-08-11	9999-01-01	Sales
489882	1992-08-11	d007	1985-08-11	9999-01-01	Sales
489883	9999-01-01	d007	1999-12-21	9999-01-01	Sales
489884	2001-10-09	d007	1988-07-25	2001-10-09	Sales
489885	1997-07-25	d007	1988-07-25	2001-10-09	Sales
489886	2001-03-07	d002	1996-10-29	2001-03-07	Finance
489887	9999-01-01	d005	1999-10-28	9999-01-01	Development
489888	9999-01-01	d009	1997-02-11	9999-01-01	Customer Service
489891	1992-05-09	d001	1995-05-31	9999-01-01	Marketing
489892	1992-05-09	d003	1987-05-10	1995-05-31	Human Resources
489889	9999-01-01	d001	1995-05-31	9999-01-01	Marketing
489890	9999-01-01	d003	1987-05-10	1995-05-31	Human Resources
489893	9999-01-01	d004	1997-04-07	9999-01-01	Production
489894	1993-10-27	d004	1993-02-22	1993-10-27	Production
489895	9999-01-01	d004	1997-06-02	9999-01-01	Production
489896	2002-05-13	d004	1996-05-13	9999-01-01	Production
489897	9999-01-01	d004	1996-05-13	9999-01-01	Production
489898	1992-08-29	d005	1987-08-30	9999-01-01	Development
489899	9999-01-01	d005	1987-08-30	9999-01-01	Development
489901	1998-12-27	d002	1993-12-27	9999-01-01	Finance
489900	9999-01-01	d002	1993-12-27	9999-01-01	Finance
489902	9999-01-01	d004	1997-11-30	9999-01-01	Production

[100 rows x 16 columns]

```
[ ]: # etsdm = pd.merge(etsd,dept_manager, on = 'dept_no', how = 'left')
      # etsdm.tail()
```

```
[30]: etsd_sort.loc[etsd['emp_no'] == '499942']
```



```
[30]: Empty DataFrame
      Columns: [emp_no, birth_date, first_name, last_name, gender, hire_date, salary,
      from_date_x, to_date_x, title, from_date_y, to_date_y, dept_no, from_date,
      to_date, dept_name]
      Index: []
```