

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: a minute ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
In [17]: #import required Library
import pandas as pd

In [18]: #read the econ_purchase csv file Located at the data subfoloder into a dataframe name the dataframe dfEcon
dfEcon=pd.read_csv('data/Econ_purchase')
```

```
In [19]: #display the head of the data frame
dfEcon.head()
```

Out[19]:

	Address	Lot	AM or PM	Browser Info	Company	Credit Card	CC Exp Date	CC Security Code	CC Provider	Email	Job	IP Address
0	16629 Pace Camp Apt. 448 inAlexisborough, NE 77...	46	in	Opera/9.56. (X11; Linux x86_64; sl... Presto/2...	Martinez-Herman	6011929061123406	02/20	900	JCB 16 digit	pdunlap@yahoo.com	Scientist, product/process development	149.146.147.2
1	9374 Jasmine Spurs Suite 508 inSouth John, TN 8...	28	m	Opera/8.93. (Windows 98; Win 9x 4.90; en-US) Pr...	Fletcher, Richards and Whitaker	3337758169645356	11/18	561	Mastercard	anthony41@reed.com	Drilling engineer	15.160.41.
2	Unit 0065 Box 5052 inDPO AP 27450	94	VE	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Simpson, Williams and Pham	675957666125	08/19	699	JCB 16 digit	amymiller@morales-harrison.com	Customer service manager	132.207.160.
3	7780 Julia FordsinNew Stacy, WA 45798	36	vm	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_0 ...	Williams, Marshall and Buchanan	6011578504430710	02/24	384	Discover	brent16@olson-robinson.info	Drilling engineer	30.250.74.
4	23012 Munoz Drive Suite 337 inNew Cynthia, TX 5...	20	IE	Opera/9.58. (X11; Linux x86_64; it-IT) Presto/2...	Brown, Watson and Andrews	6011456623207998	10/25	678	Diners Club / Carte Blanche	christophenvrigh@gmail.com	Fine artist	24.140.33.

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 2 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

	Address	Lot	AM or PM	Info	Company	Credit Card	Exp Date	Security Code	Provider	Email	JOB	IP Address
0	16629 Pace Camp Apt. 448 inAlexisborough, NE 77...	46	in	Opera/9.56. (X11; Linux x86_64; sl... Presto/2...	Martinez-Herman	6011929061123406	02/20	900	JCB 16 digit	pdunlap@yahoo.com	Scientist, product/process development	149.146.147.2
1	9374 Jasmine Spurs Suite 508 inSouth John, TN 8...	28	m	Opera/8.93. (Windows 98; Win 9x 4.90; en-US) Pr...	Fletcher, Richards and Whitaker	3337758169645356	11/18	561	Mastercard	anthony41@reed.com	Drilling engineer	15.160.41.
2	Unit 0065 Box 5052 inDPO AP 27450	94	VE	Mozilla/5.0 (compatible; MSIE 9.0; Windows NT ...	Simpson, Williams and Pham	675957666125	08/19	699	JCB 16 digit	amymiller@morales-harrison.com	Customer service manager	132.207.160.
3	7780 Julia FordsinNew Stacy, WA 45798	36	vm	Mozilla/5.0 (Macintosh; Intel Mac OS X 10_8_0 ...	Williams, Marshall and Buchanan	6011578504430710	02/24	384	Discover	brent16@olson-robinson.info	Drilling engineer	30.250.74.
4	23012 Munoz Drive Suite 337 inNew Cynthia, TX 5...	20	IE	Opera/9.58. (X11; Linux x86_64; it-IT) Presto/2...	Brown, Watson and Andrews	6011456623207998	10/25	678	Diners Club / Carte Blanche	christophenvrigh@gmail.com	Fine artist	24.140.33.

How many rows and columns are there? (worth 10 points)

```
In [20]: #how many rows and columns are in that dataframe
```

71°F Clear 9:43 PM 9/22/2022

Pandas\_Lab - Jupyter Notebook x +

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 2 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [20]: #how many rows and columns are in that dataframe  
dfEcom.shape

Out[20]: (10000, 14)

Display the first 10 rows of only the job column (Worth 10 points)

In [21]: #display the first 10 rows of only the job column of the dfEcom data frame  
dfEcom['Job'].head(10)

Out[21]: 0 Scientist, product/process development  
1 Drilling engineer  
2 Customer service manager  
3 Drilling engineer  
4 Fine artist  
5 Fish farm manager  
6 Dancer  
7 Event organiser  
8 Financial manager  
9 Forensic scientist  
Name: Job, dtype: object

Display the first 10 rows of only the language column (worth 10 points)

In [22]: #Display the first 10 rows of only the language column  
dfEcom['Language'].head(10)

Out[22]: 0 el  
1 fr  
2 de  
3 es  
4 es  
5 ru  
6 de  
7 pt  
8 ru  
9 zh  
Name: Language, dtype: object

71°F Clear 9:43 PM 9/22/2022

Pandas\_Lab - Jupyter Notebook x +

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 2 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

In [22]: #Display the first 10 rows of only the language column  
dfEcom['Language'].head(10)

Out[22]: 0 el  
1 fr  
2 de  
3 es  
4 es  
5 ru  
6 de  
7 pt  
8 ru  
9 zh  
Name: Language, dtype: object

What is the average Purchase Price? (Worth 10 points)

In [115]: dfEcom['Purchase Price'].mean()

Out[115]: 50.34730200000025

What were the highest and lowest purchase prices? (worth 10 points)

In [116]: dfEcom['Purchase Price'].max()

Out[116]: 99.99

In [117]: dfEcom['Purchase Price'].min()

71°F Clear

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 2 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
In [116]: dfEcom['Purchase Price'].max()
Out[116]: 99.99

In [117]: dfEcom['Purchase Price'].min()
Out[117]: 0.0
```

How many people have French 'fr' as their Language of choice on the website? (worth 10 points)

```
In [23]: dfEcom['Language'].value_counts()['fr']
Out[23]: 1097
```

How many people have the job title of "Forensic scientist" ? (worth 10 points)

```
In [24]: dfEcom['Job'].value_counts()["Forensic scientist"]
Out[24]: 18
```

How many people work for "Simpson, Williams and Pham" company (worth 10 points)

```
In [52]: dfEcom[dfEcom['Company'] == "Simpson, Williams, Pham"]['Company'].count()
Out[52]: 0
```

71°F Clear

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 2 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
In [52]: dfEcom[dfEcom['Company'] == "Simpson, Williams, Pham"]['Company'].count()
Out[52]: 0
```

How many people made the purchase during the AM and how many people made the purchase during PM ? (worth 10 points)

*(Hint: Check out [value\\_counts\(\)](#))*

```
In [47]: dfEcom['AM or PM'].value_counts()
Out[47]: PM    5068
        AM    4932
        Name: AM or PM, dtype: int64
```

What are the 5 most common Languages? (worth 10 points)

```
In [49]: dfEcom['Language'].value_counts()[0:5]
Out[49]: de    1155
        ru    1155
        el    1137
        pt    1118
        en    1098
        Name: Language, dtype: int64
```

*\*Someone made a purchase that with the following credit card 6011929061123406\*, what was the Purchase Price for this transaction? (worth 10 points) \**

```
In [56]: dfEcom[dfEcom['Credit Card']==6011929061123406][['Purchase Price']]
```

71°F Clear



localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 3 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
In [64]: dfSal = pd.read_csv('data/Salaries.csv')
dfSal.head()
```

```
Out[64]:
```

	Id	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits	Year	Notes	Agency	Status
0	1	NATHANIEL FORD	GENERAL MANAGER-METROPOLITAN TRANSIT AUTHORITY	167411.18	0.00	400184.25	NaN	567595.43	567595.43	2011	NaN	San Francisco	NaN
1	2	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28	2011	NaN	San Francisco	NaN
2	3	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279.91	335279.91	2011	NaN	San Francisco	NaN
3	4	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343.61	332343.61	2011	NaN	San Francisco	NaN
4	5	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT (FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373.19	326373.19	2011	NaN	San Francisco	NaN

What is the highest amount of BasePay in the dataset? (worth 10 points)

```
In [67]: dfSal['BasePay'].max()
```

```
Out[67]: 319275.01
```

What is the name of lowest paid person (including benefits)? (worth 10 points)

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 3 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
In [67]: dfSal['BasePay'].max()
```

```
Out[67]: 319275.01
```

What is the name of lowest paid person (including benefits)? (worth 10 points)

```
In [68]: dfSal[dfSal['TotalPayBenefits'] == dfSal['TotalPayBenefits'].min()] [['EmployeeName']]
```

```
Out[68]:
```

EmployeeName
148653 Joe Lopez

What was the average (mean) TotalPayBenefits of all employees per year? (2011-2014) ? (worth 10 points)

```
In [69]: dfSal.groupby('Year').mean() [['TotalPayBenefits']]
```

```
Out[69]:
```

Year	TotalPayBenefits
2011	71744.103871
2012	100553.229232
2013	101440.519714
2014	100250.918884

Pandas\_Lab - Jupyter Notebook

localhost:8888/notebooks/Downloads/Pandas\_Lab\_Assignment/Pandas\_Lab\_Assignment/Pandas\_Lab.ipynb

jupyter Pandas\_Lab Last Checkpoint: 3 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

148653 Joe Lopez

What was the average (mean) TotalPayBenefits of all employees per year? (2011-2014) ? (worth 10 points)

```
In [69]: dfSal.groupby('Year').mean()[['TotalPayBenefits']]
```

Out[69]:

TotalPayBenefits	
Year	
2011	71744.103871
2012	100553.229232
2013	101440.519714
2014	100250.918884

How many unique agencies are there? (worth 10 points)

```
In [70]: dfSal['Agency'].nunique()
```

Out[70]: 1

In [ ]:

71°F Clear 9:44 PM 9/22/2022