Dataframe exercise!

In this exercise we will be using the SF Salaries Dataset from Kaggle:
https://www.kaggle.com/kaggle/sf-salaries
For your convenience you can find the csv file also here:
`http://bit.do/salaries-csv`
This data contains the names, job title, and compensation for San Francisco city employees on an annual basis from 2011 to 2014.

Import Pandas as pd and read salaries.csv as a dataframe called sal

Check the head of the dataframe

How many entries there are in the dataframe?

What is the average Basepay?

What is the job title of JOSEPH DRISCOLL ? (Use all caps)

What is the name of highest paid person (including benefits)?

How many unique job titles are there?

What are the top 3 most common jobs?

How many Job Titles were represented by only one person in 2013?

How many people have the word Chief in their job title?

Exercise

Download the file at:
`http://bit.do/train-csv`
This data set is related with a mortgage loan.

1-Load the data and make the Loan_ID the index_col

2-Try to filter values of a column based on conditions from another set of columns. For istance, get a list of the ApplicantIncome of all males who are graduated and does not got a loan.

3-Compare the ApplicantIncome of graduates males and not graduate males.

4-How can you check for missing values? Define a custom function that counts the nulls, and apply it to columns and rows

5-Now you must deal with the missing values…Somewhat arbitrarily set the null Gender, Married and Self_employed to the "mode" of that columns. Check the missing values again to confirm.

6- Pandas can be used to create MS Excel style "pivot tables". For instance, in this case, a key column is “LoanAmount” which has missing values.
We can impute it using mean amount of each ‘Gender’, ‘Married’ and ‘Self_Employed’ **group**. The mean ‘LoanAmount’ of each group can be determined with the pivot_table function: build it!

7-Use your brand new pivot table to replace null values in the LoanAmount column. You can use a for loop or a one-line instruction.

8-Use the crosstab function to get a feeling of the Credit History importance for the Loan Status

- https://www.analyticsvidhya.com/blog/2016/01/12-pandas-techniques-python-data-manipulation/
- https://www.udemy.com/python-for-data-science-and-machine-learning-bootcamp/
- https://github.com/yew1eb/DM-Competition-Getting-Started/blob/master/AV-loan-prediction/train.csv
- https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/learning-path-data-science-python/
- https://www.analyticsvidhya.com/blog/2016/01/guide-data-exploration/
- http://www.labri.fr/perso/nrougier/teaching/numpy.100/