Universidad **Ricardo Palma**

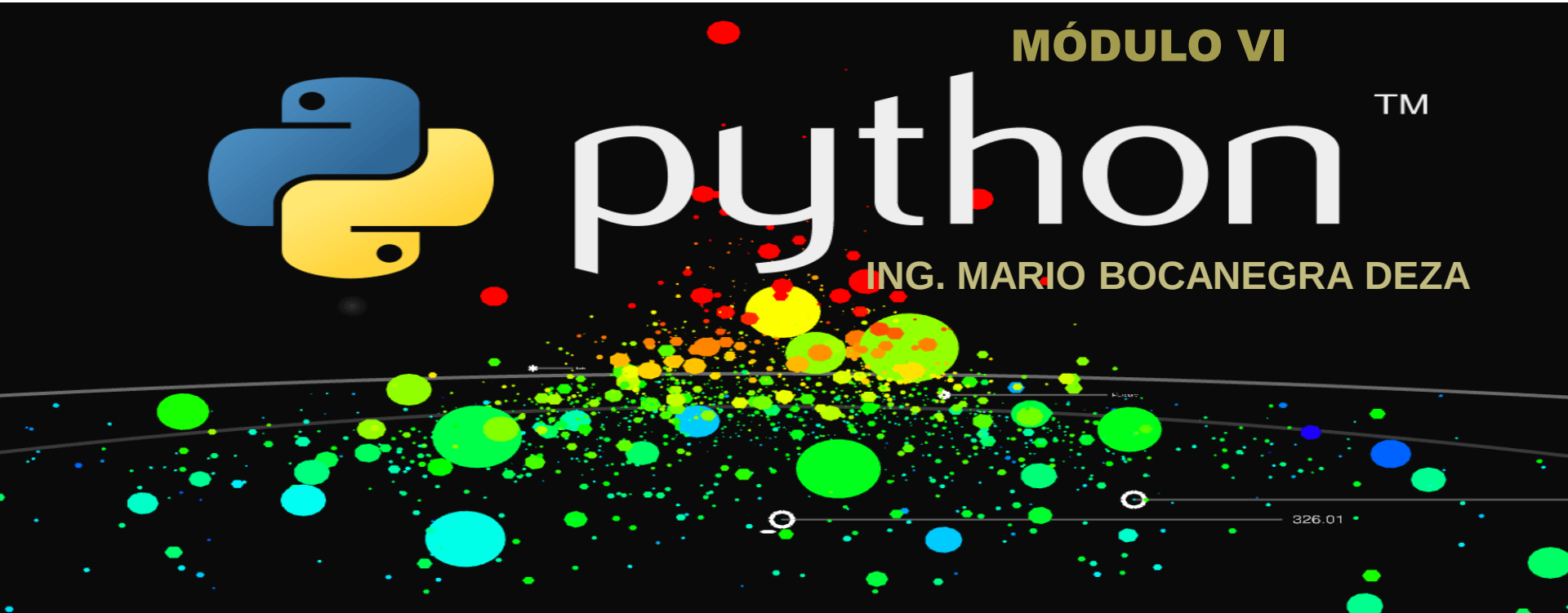
RECTORADO

Formamos seres humanos para una cultura de paz

Primer Programa de Especialización

INTRODUCCIÓN AL DATA SCIENCE

ANÁLISIS DE DATOS CON PYTHON



Contenido

- ✓Análisis Exploratorio : Estadística Descriptiva

DataSet

	rank	discipline	yrs.since.phd	yrs.service	sex	salary
0	Prof	B	19	18	Male	139750
1	Prof	B	20	16	Male	173200
2	AsstProf	B	4	3	Male	79750
3	Prof	B	45	39	Male	115000
4	Prof	B	40	41	Male	141500
5	AssocProf	B	6	6	Male	97000
6	Prof	B	30	23	Male	175000
7	Prof	B	45	45	Male	147765
8	Prof	B	21	20	Male	119250
9	Prof	B	18	18	Female	129000

Funciones de Estadística

`df.describe()` - Summary statistics for numerical columns

`df.mean()` - Returns the mean of all columns

`df.corr()` - Returns the correlation between columns in a DataFrame

`df.count()` - Returns the number of non-null values in each DataFrame column

`df.max()` - Returns the highest value in each column

`df.min()` - Returns the lowest value in each column

`df.median()` - Returns the median of each column

`df.std()` - Returns the standard deviation of each column

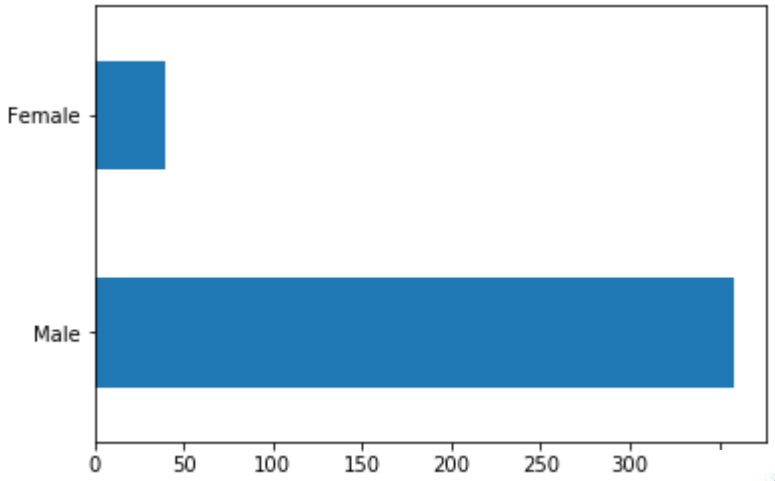
Gráficos de Exploración

Importar las librerías para trabajar con los gráficos

```
In [157]: import pandas as pd
          import matplotlib.pyplot as plt
          %matplotlib inline
          from pandas.tools.plotting import scatter_matrix
```

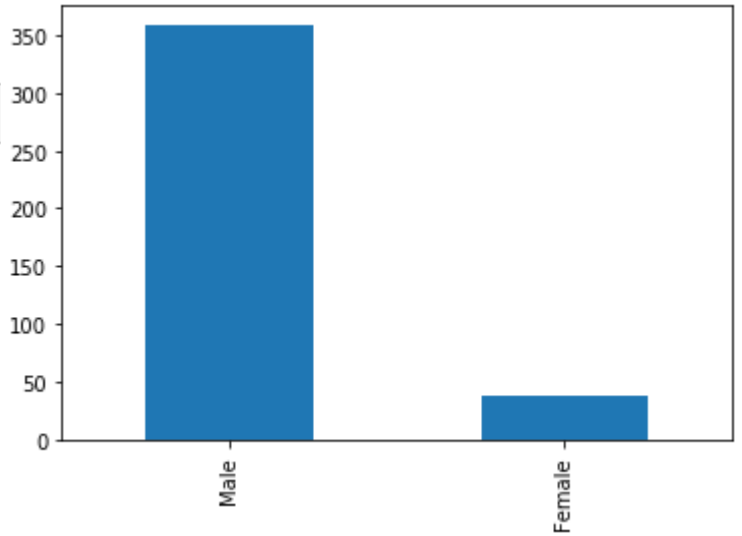
Gráficos de Exploración

```
In [6]: df['sex'].value_counts()

Out[6]: Male      358
        Female     39
        Name: sex, dtype: int64

In [23]: genero_tabla=df['sex'].value_counts()

In [24]: genero_tabla.plot(kind='barh',rot=0)
```

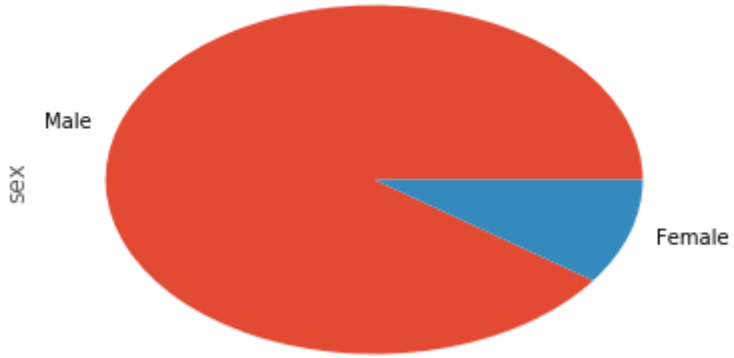


Gráficos de Exploración

```
In [26]: genero_tabla.plot.bar()
```

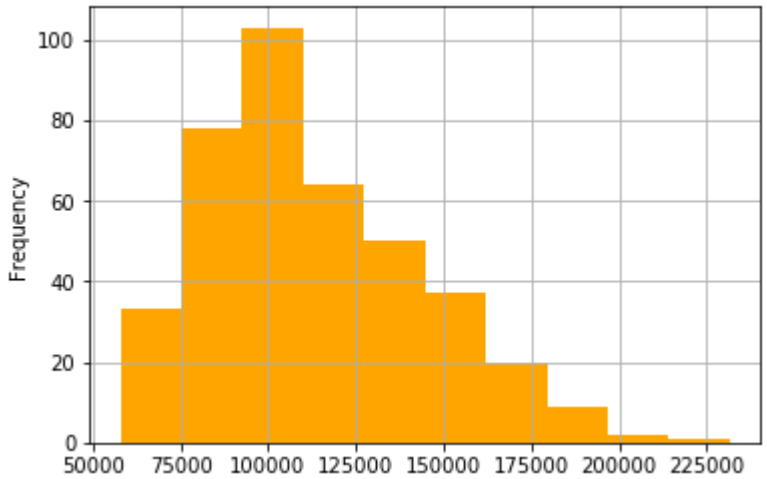



```
In [38]: genero_tabla.plot.pie()
```



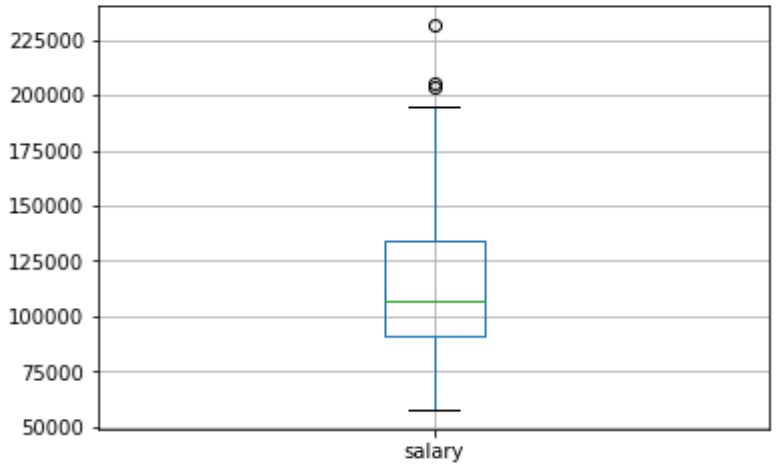
Gráficos de Exploración

```
data_salary=df['salary']
data_salary.plot.hist(color='orange',grid=True)
```

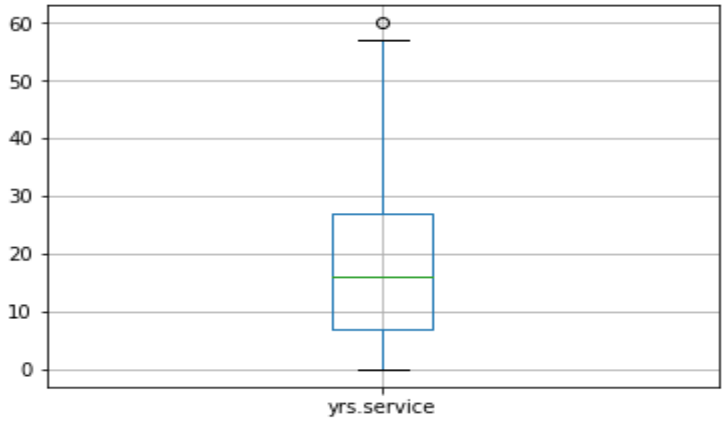


Gráficos de Exploración

```
In [168]: df[['salary']].boxplot()
```

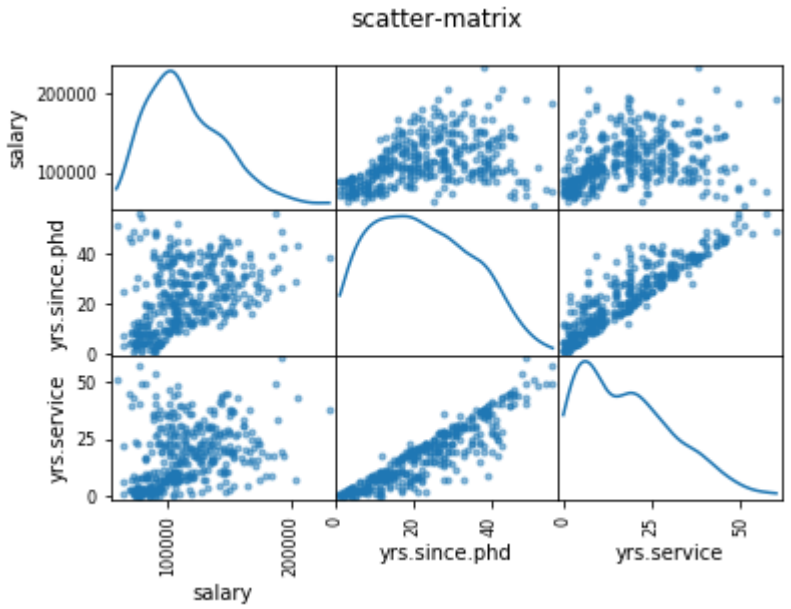



```
In [171]: df[['yrs.service']].boxplot()
```



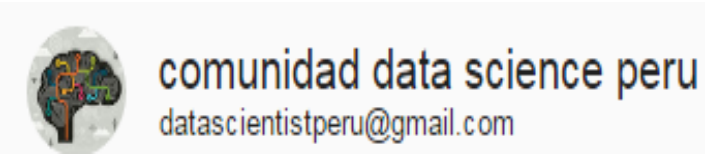
Gráficos de Exploración

```
In [165]: scatter_matrix(df[['salary','yrs.since.phd','yrs.service']], alpha=0.5,diagonal='kde')
          plt.suptitle('scatter-matrix')
          plt.show()
```




fanpage
facebook


Comunidad
Data Science
Perú
@comunidaddatascience


comunidad data science peru
datascientistperu@gmail.com