```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
pwd
```

```
'C:\\Users\\Alfredo\\Desktop\\Machine Learning'
```

```
ls
```

```
 El volumen de la unidad C es Windows
 El número de serie del volumen es: FC43-3B7A

 Directorio de C:\Users\Alfredo\Desktop\Machine Learning

19/04/2021  22:06    <DIR>          .
19/04/2021  22:06    <DIR>          ..
19/04/2021  20:49    <DIR>          .ipynb_checkpoints
03/07/2020  10:47             5,037 agrupacion por columna.ipynb
16/04/2021  20:34           185,160 cufflinks graficos .ipynb
14/04/2021  21:06             3,318 cufflinks.ipynb
27/07/2020  20:14             4,692 data frame en una tabla SQL.ipynb
28/07/2020  10:58           137,766 dataframe df.ipynb
02/07/2020  16:33             7,405 DataFrames modificacion de los nombres de las filas .ipynb
17/06/2020  10:23             7,468 diccionarios.ipynb
07/04/2021  10:14            14,378 ejercicio 1 matplotlib.ipynb
17/06/2020  10:23             1,911 Filter.ipynb
17/06/2020  10:30             1,743 funciones.ipynb
07/04/2021  21:03            33,836 Grafico de columna para tipo categoria seaborn.ipynb
03/04/2021  19:31           171,852 Graficos de Regresion.ipynb
08/04/2021  16:08            63,630 heat maps.ipynb
19/04/2021  22:06            31,598 linear regression.ipynb
07/04/2021  10:40            35,966 Matplotlib .ipynb
03/07/2020  12:24             8,413 merge con dataframe.ipynb
17/06/2020  10:29             2,468 numpy  con listas.ipynb
08/04/2021  16:08           201,001 pairgrid & facegrid.ipynb
07/04/2021  10:24            98,134 seaborn .ipynb
02/07/2020  10:34            13,249 seleccion de datos pandas.ipynb
08/01/2021  10:55            16,898 Untitled.ipynb
03/07/2020  12:24             1,385 Untitled1.ipynb
16/04/2021  20:15             2,429 Untitled10.ipynb
08/04/2021  13:05           555 Untitled2.ipynb
16/04/2021  22:01     3,656,909 Untitled3.ipynb
15/04/2021  18:08           555 Untitled4.ipynb
05/07/2021  15:56         1,895 Untitled5.ipynb
15/04/2021  18:32           853 Untitled6.ipynb
15/04/2021  21:06         72,264 Untitled7.ipynb
16/04/2021  20:01         1,398 Untitled9.ipynb
19/04/2021  21:13       726,209 USA_Housing.csv
02/07/2020  10:35         2,766 valores nulos.ipynb
              32 archivos      5,505,736 bytes
               3 dirs  354,873,909,184 bytes libres
```

```
casas = pd.read_csv('C:/Users/Alfredo/Desktop/Machine Learning/USA_Housing.csv')
```

```
casas
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058998e+06	9127 Elizabeth Stravenue\nDanielton, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFPO AE 09386
...	...	...	...	...	...	...	...
4995	60567.944140	7.830362	6.137356	3.46	22837.361035	1.060194e+06	USNS Williams\nFPO AP 30153-7653
4996	78491.275435	6.999135	6.576763	4.02	25616.115489	1.482618e+06	PSC 9258, Box 8489\nAPO AA 42991-3352
4997	63390.686886	7.250591	4.805081	2.13	33266.145490	1.030730e+06	4215 Tracy Garden Suite 076\nJohnsland, VA 01...
4998	68001.331235	5.534388	7.130144	5.44	42625.620156	1.198657e+06	USS Wallace\nFPO AE 73316
4999	65510.581804	5.992305	6.792336	4.07	46501.283803	1.298990e+06	37778 George Ridges Apt. 509\nEast Holly, NV 2...

```
5000 rows × 7 columns
```

```
casas.info
```

```
<bound method DataFrame.info of         Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0       79545.458574             5.682861                   7.009188
1       79248.642455             6.002900                   6.730821
2       61287.067179             5.865890                   8.512727
3       63345.240046             7.188236                   5.586729
4       59982.197226             5.040555                   7.839388
...              ...                  ...                        ...
4995    60567.944140             7.830362                   6.137356
4996    78491.275435             6.999135                   6.576763
4997    63390.686886             7.250591                   4.805081
4998    68001.331235             5.534388                   7.130144
4999    65510.581804             5.992305                   6.792336

      Avg. Area Number of Bedrooms  Area Population          Price  \
0                             4.09         23086.800503  1.059034e+06
1                             3.09         40173.072174  1.505891e+06
2                             5.13         36882.159400  1.058998e+06
3                             3.26         34310.242831  1.260617e+06
4                             4.23         26354.109472  6.309435e+05
...                            ...                  ...           ...
4995                          3.46         22837.361035  1.060194e+06
4996                          4.02         25616.115489  1.482618e+06
4997                          2.13         33266.145490  1.030730e+06
4998                          5.44         42625.620156  1.198657e+06
4999                          4.07         46501.283803  1.298990e+06

                                            Address
0      208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1      188 Johnson Views Suite 079\nLake Kathleen, CA...
2      9127 Elizabeth Stravenue\nDanielton, WI 06482...
3                   USS Barnett\nFPO AP 44820
4                   USNS Raymond\nFPO AE 09386
...                                                 ...
4995          USNS Williams\nFPO AP 30153-7653
4996             PSC 9258, Box 8489\nAPO AA 42991-3352
4997  4215 Tracy Garden Suite 076\nJohnsland, VA 01...
4998                   USS Wallace\nFPO AE 73316
4999  37778 George Ridges Apt. 509\nEast Holly, NV 2...

[5000 rows x 7 columns]>
```

```
casas.describe
```

```
<bound method NDFrame.describe of         Avg. Area Income  Avg. Area House Age  Avg. Area Number of Rooms  \
0       79545.458574             5.682861                   7.009188
1       79248.642455             6.002900                   6.730821
2       61287.067179             5.865890                   8.512727
3       63345.240046             7.188236                   5.586729
4       59982.197226             5.040555                   7.839388
...              ...                  ...                        ...
4995    60567.944140             7.830362                   6.137356
4996    78491.275435             6.999135                   6.576763
4997    63390.686886             7.250591                   4.805081
4998    68001.331235             5.534388                   7.130144
4999    65510.581804             5.992305                   6.792336

      Avg. Area Number of Bedrooms  Area Population          Price  \
0                             4.09         23086.800503  1.059034e+06
1                             3.09         40173.072174  1.505891e+06
2                             5.13         36882.159400  1.058998e+06
3                             3.26         34310.242831  1.260617e+06
4                             4.23         26354.109472  6.309435e+05
...                            ...                  ...           ...
4995                          3.46         22837.361035  1.060194e+06
4996                          4.02         25616.115489  1.482618e+06
4997                          2.13         33266.145490  1.030730e+06
4998                          5.44         42625.620156  1.198657e+06
4999                          4.07         46501.283803  1.298990e+06

                                            Address
0      208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1      188 Johnson Views Suite 079\nLake Kathleen, CA...
2      9127 Elizabeth Stravenue\nDanielton, WI 06482...
3                   USS Barnett\nFPO AP 44820
4                   USNS Raymond\nFPO AE 09386
...                                                 ...
4995          USNS Williams\nFPO AP 30153-7653
4996             PSC 9258, Box 8489\nAPO AA 42991-3352
4997  4215 Tracy Garden Suite 076\nJohnsland, VA 01...
4998                   USS Wallace\nFPO AE 73316
4999  37778 George Ridges Apt. 509\nEast Holly, NV 2...

[5000 rows x 7 columns]>
```

```
casas.columns
```

```
Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],
      dtype='object')
```

DISTRIBUCION DE LA MEDIA DE LOS DATOS

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

```
casas.head()
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058998e+06	9127 Elizabeth Stravenue\nDanielton, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFPO AE 09386

```
X= casas[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
       'Avg. Area Number of Bedrooms', 'Area Population']]
```

```
Y= casas['Price']
```

```
X_train, X_test, Y_train, Y_test =train_test_split(X, Y, test_size=0.3,random_state =42)
```

```
X_train
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population
1840	55245.337288	3.965745	8.961106	4.38	43557.843435
2115	62305.638407	6.490769	7.647362	3.19	47066.617420
4437	77345.472379	5.407515	8.243178	3.11	26706.911029
1146	45946.171039	6.385374	6.804131	3.18	29214.363551
2406	69350.793357	6.910415	8.288048	4.29	36779.958567
...	...	...	...	...	...
4426	76223.561256	6.371627	5.342217	2.42	30165.337445
466	56685.214442	6.958045	7.502115	3.38	43322.166854
3092	66195.337714	6.507971	6.611861	3.14	37288.823574
3772	58694.515017	7.394768	9.269453	4.32	49960.977236
860	61162.580254	5.896316	7.880521	6.04	36033.701431

```
3500 rows × 5 columns
```

```
lrm =  LinearRegression()
lrm.fit(X_train, Y_train)
```

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

EVALUACION DEL MODELO

```
from sklearn import metrics
```

```
prediciones = lrm.predict(X_test)
```

```
prediciones  #los datos de test se comparan con los valores de Y (datos de pruebas )
```

```
array([1308536.13592605, 1237122.72746462, 1243035.62817079, ...,
       1457119.79297226, 1483428.95369313, 1047510.59737291])
```

grafico de compracion

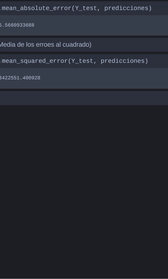
```
plt.scatter(Y_test, predicciones)
```

```
<matplotlib.collections.PathCollection at 0x211dd1da0c8>
```

histograma

```
sns.distplot(Y_test - predicciones)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x211dd695248>
```



series de metricas para evaluar el modelo de forma numerica

MAE (Mean absolute error) - Media del valor absoluto de los errores

```
metrics.mean_absolute_error(Y_test, predicciones)
```

```
81135.5660933688
```

MSE (Media de los errores al cuadrado)

```
metrics.mean_squared_error(Y_test, predicciones)
```

```
10068422551.409928
```