|  |
| --- |
|  |
| Informe  de Minería de Datos(PARTE 3) |
|  |
| Nombre : Alexander Blanco, Fernando Peña, Diego Silva, Manuel Gonzalez, Carlos Alvarez  Nombre del docente: Bignollo Vaccaro  09-07-2020 |

**Índice**

[**Introducción**](#_7x6iialcos4l)3

[**Desarrollo**](#_8ifepjjifhaf)3

[**Importación de pandas\_profilling**](#_slpfai5enigl)4

[**Definir X e Y**](#_rg7nhjna2wnv)5

[**Implementación de modelo de clasificación**](#_f7t4r36b36lv)6

[**Tabla de métrica de rendimiento**](#_zh8wpn1v0o9)6

[**Predicción de matrículas 2018**](#_vxb7czeam31n)7

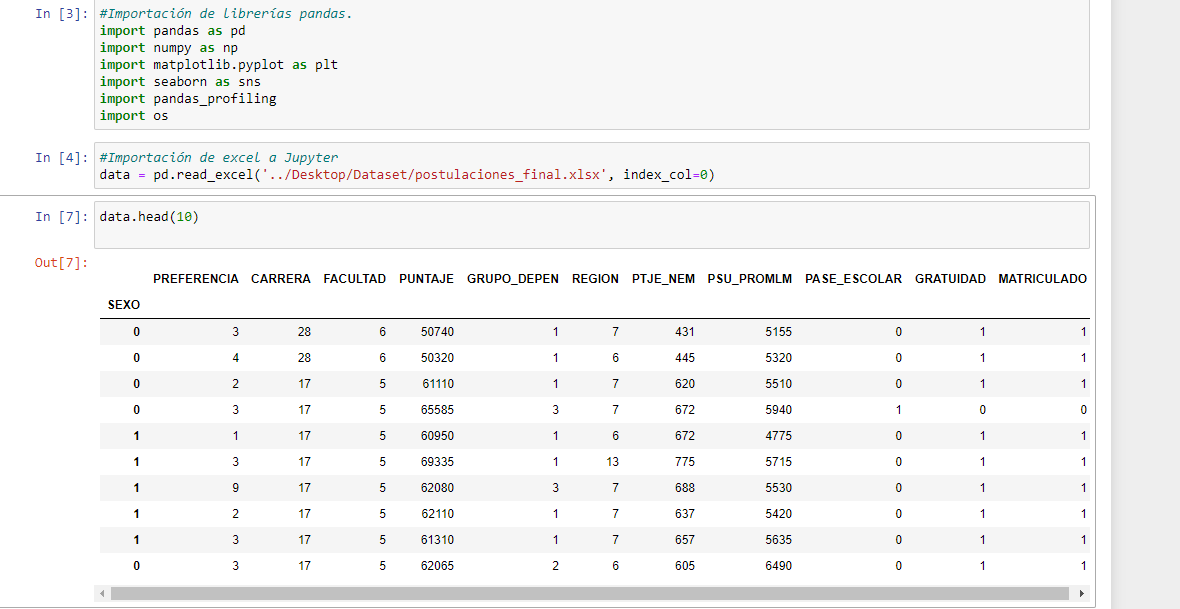
[**Rutina de patrones de K-Means Clustering**](#_uejtklx93sso)8

# Introducción

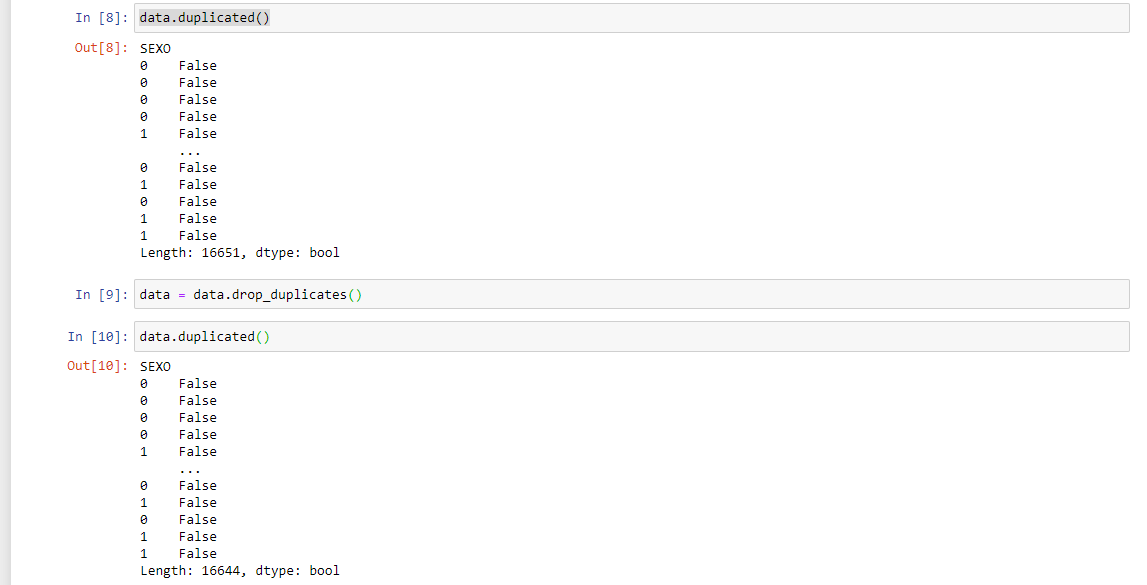
En el presente documento mostraremos imagenes de paso a paso donde nosotros damos en evidencia la limpia de los datos también en Desarrollo, documentación y análisis de los resultados de un programa con la aplicación de modelos de Machine Learning en Python, para extraer conocimiento sobre el dominio de las matricula efectivas de Universia-DUOC

# Desarrollo

Importamos el nuevo .xlsx de postulaciones\_final realizado en la evaluacion1



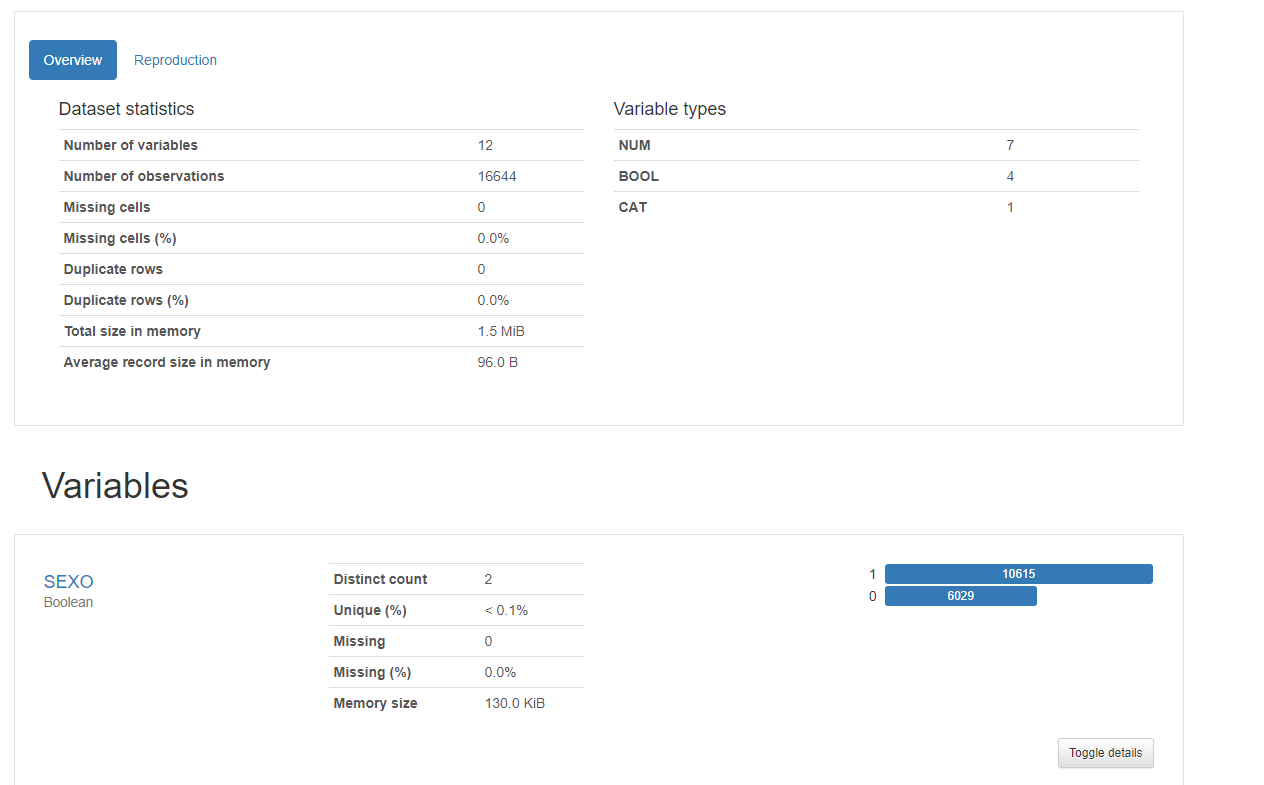
Luego realizamos una limpieza de los datos duplicados



# Importación de pandas\_profilling

Importación y visualización del reporte generado por pandas\_profilling





En el reporte arrojó que no se deben hacer cambios en el proyecto ya que la fase de limpieza de datos está realizada correctamente

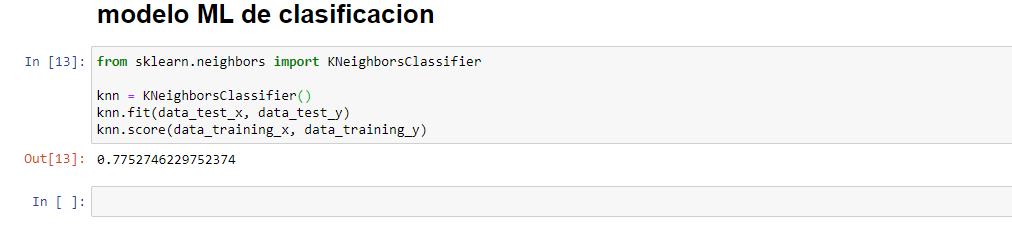
# Definir X e Y

Luego Se realiza la separación de las variables dependientes(X) de la variable dependiente (Y)

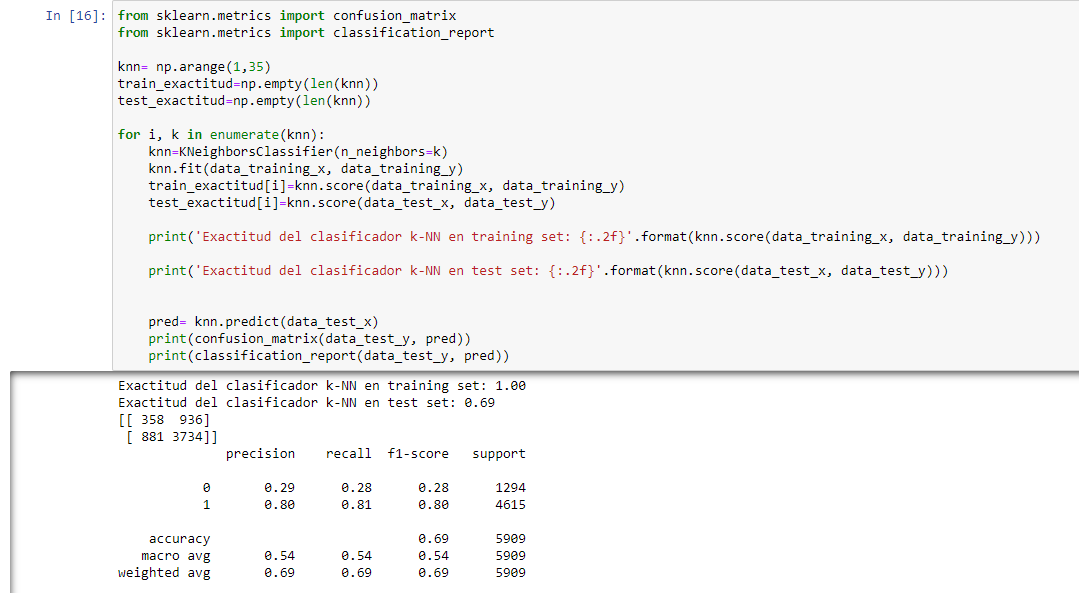


# Implementación de modelo de clasificación

Implementamos el modelo de Kneiboars Classifier ya que es un modelo de clasificación donde busca en las observaciones más cercanas a la que se está tratando de predecir y clasifica el punto de interés basado en la mayoría de datos que le rodean. los parámetros utilizados fueron “matriculado” ya que era el que mejor calzaba con la aplicación



# Tabla de métrica de rendimiento

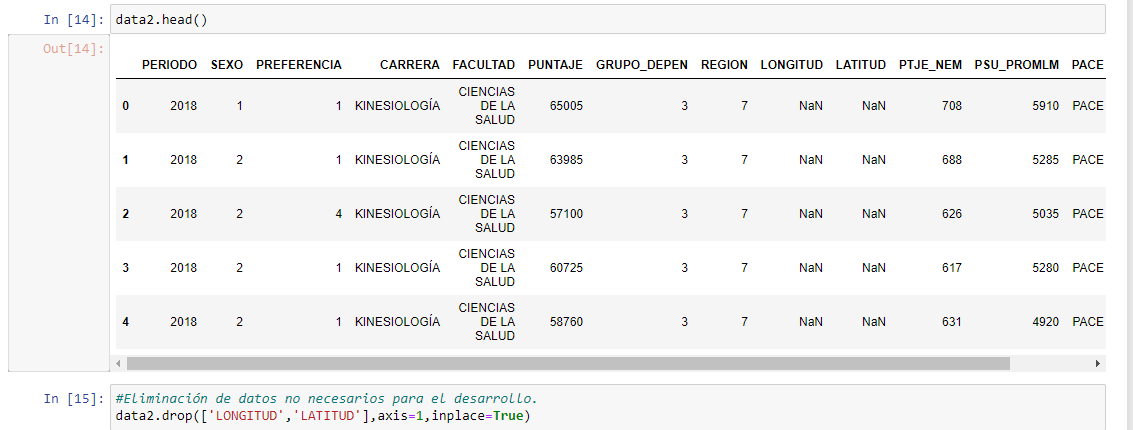


# Predicción de matrículas 2018

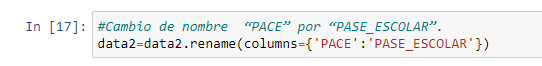
Primero se importa el dataframe postulaciones\_2018



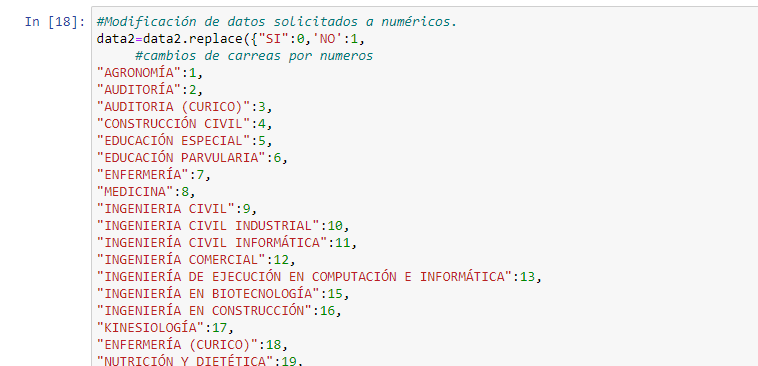
Luego se realiza la etapa de limpieza de datos del dataFrame postulaciones\_2018



se reemplaza el nombre de PACE a PASE\_ESCOLAR



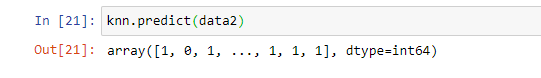
Se cambian los valores a datos numericos



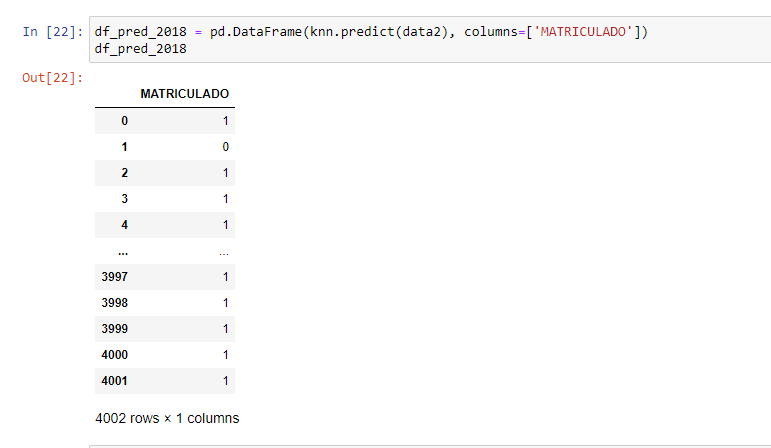
Luego se modifican los datos nulos y se reemplazan por 0



Luego se realiza la prediccion con knn.predict() del dataframe postulaciones\_2018



luego se realiza la predicion con la tabla ‘matriculado’

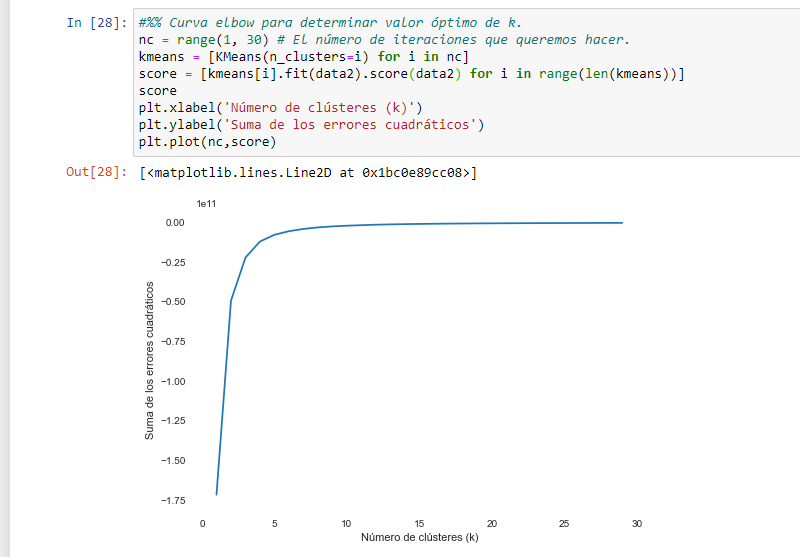


Luego exportamos a excel con el nombre de “predicciones\_matriculas\_2018”

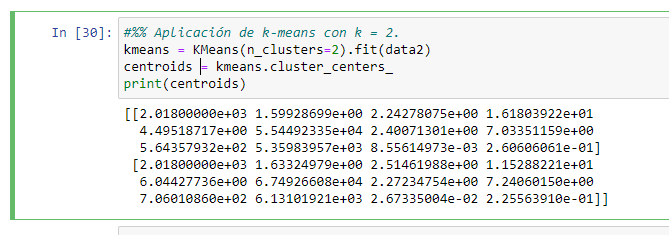


# Rutina de patrones de K-Means Clustering

se realiza una rutina que determina el valor optimo de K



Aplicación de k-means con k = 2



Como se puede observar a través de knneighboards se hace la aplicación con 2 clusters donde muestra la distribución de los vecinos más cercanos por lo cual sirve para estimar la función de densidad de las predictoras por cada clase