Filas: 1614

Columnas: 41

Empresa: 1980

EEUU

IA y machine learning

1. **Age**: La edad del empleado.

- cambiamos

- no hay nulos

- tipo de dato: object (str) à cambiar a int

- Count: 1614

- Unique: 54

- Top: 35

- Freq: 84

- hay números escritos con palabras 🡪 los cambiamos a números (thirty-two, fifty-eight, twenty-six, thirty-seven, thirty-one, thirty, fifty-two, fifty-five, thirty-six, forty-seven, twenty-four)

- cambiado tipo de dato a int64

- media: 36.9206

- std: 9.10

- min: 18

- max: 60

1. **Attrition**: Indica si el empleado ha dejado la empresa ("No" significa que no ha dejado la empresa y "Yes" significa que ha dejado la empresa).

- no hay nulos

- tipo de dato: object (str)

- Count: 1614

- Unique: 2

- Top: NO

- Freq: 1355

- No 🡪 1355

- Yes 🡪 259

3. **BusinessTravel**: Describe la frecuencia de los viajes relacionados con el trabajo del empleado (por ejemplo, "Travel\_Rarely" para raramente).

- hay NaN

- 772 nulos

- tipo de dato: object (str)

- Count: 842

- Unique: 3

- Top: Travel rarely

- Freq: 586

- ‘travel\_rarely’: 586

- ‘travel\_frecuently’: 165

- ‘non-travel’: 91

- cambiamos ‘non-travel’ por ‘non\_travel’

- cambiamos los Nan por ‘desconocido’

4. **DailyRate**: La tarifa diaria del empleado.

- sacamos $ para convertirlo a float

- 124 nulos

- tipo de dato: object (str) à float

- Count: 1614

- Unique: 849

- Top: nan$

- Freq: 124

- ‘nan’ a np.nan para pasar la columna a float64

- intentar cambiar los nulos con cálculos respecto al monthlyRate, o similar

5. **Department**: El departamento en el que trabaja el empleado (por ejemplo, "Research & Development", "Sales", etc.).

- 1312 nulos

- object

- Count: 302

- Unique: 3

- Top: Research & Development

- Freq: 196

- Research & Development: 196

- Sales: 910

- Human Resources: 15

- cambiamos:

- ‘Research & Development’ por ‘research\_development’

- ‘Sales’ por ‘sales’

- Human Resources por ‘human\_resources’

- nulos a ‘desconocido’

6. **DistanceFromHome**: La distancia desde el hogar del empleado hasta su lugar de trabajo.

- Datos negativos

- no nulos

- Int64

- media: 4.5273

- std: 14.59

- min: -49

- max: 29

- cambiamos negativos a positivos

- media: 11.1419

- std: 10.45

- min: 1

- max: 49

7. **Education**: Nivel de educación del empleado (generalmente en una escala del 1 al 5).

- no nulos

- int64

- media: 2.9250

- std: 1.02

- min: 1.0

- max: 5.

- count:

- 1: 180 eso

- 2: 314 fp

- 3: 621 bachillerato/college

- 4: 445 university degree

- 5: 54 phd

8. **EducationField**: El campo de educación del empleado.

- Nan

- 745 nulos

- object

- Count: 869

- Unique: 6

- Top: Life Sciences

- Freq: 349

- counts:

- Life sciences: 349

- Medical: 276

- Maketing: 104

- Technical degree: 69

- Other: 59

- Human Resources: 12

- cambiamos a minúsculas y un \_ en los espacios de los valores

9. **EmployeeCount**: Un contador que generalmente es 1 y se usa para contar empleados.

- no nulos

- int64

- media: 1.0000

- std: 0.0000

- min: 1.0

- max: 1.

- eliminamos columna (añadir id auto incremental)

10. **EmployeeNumber**: Un número de identificación único para el empleado.

- transformar a inbt? Completar vacíos (NaN)

- añadir columna auto incremental (id) y eliminar esta o no tenerla en cuenta

- 431nulos

- object

- Count: 1183

- Unique; 1079

- Top: 482

- Freq: 2

11. **EnvironmentSatisfaction**: Nivel de satisfacción del empleado en relación con su entorno de trabajo. Con valores que están comprendidos entre el 1 y el 4, siendo el 4 el nivel de máxima satisfacción.

- no nulos

- int64

- media: 4.2949

- std: 6.99

- min: 1.00

- max: 49.

- máximo debería ser 4, gestionar

12. **Gender**: El género del empleado. Donde 0 corresponde con "hombre" y 1 con "mujer".

- cambiar a string 🡪 0 por ‘hombre’ y 1 por 1 por ‘mujer

- no nulos

- int64

- min: 0

- max: 1.

13. **HourlyRate**: La tarifa por hora del empleado.

- no nulos

- object à cambia a int?

- Count: 1614

- Unique; 72

- Top: Not available

- Freq: 84

14. **JobInvolvement**: Nivel de implicación del empleado en su trabajo.

- no nulos

- int64

- media: 2.7397

- std: 0.71

- min: 1

- max: 4

- valoraciones del 0 al 4 🡪 correct

15. **JobLevel**: Nivel jerárquico del empleado en la empresa.

- no nulos

- int64

- media: 2.0681

- std: 1.10

- min: 1

- max: 5

16. **JobRole**: El rol o puesto de trabajo del empleado.

- tendremos que hacer un lower(), cada elemento escrito diferente

- no nulos

- object

- Count: 1614

- Unique:1579

- Top: Manager

- Freq: 4

17. **JobSatisfaction**: Nivel de satisfacción del empleado con su trabajo.

- no nulos

- int64

- media: 2.7385

- std: 1.10

- min: 1

- max: 4

- valoraciones del 0 al 4 🡪 correcto

18. **MaritalStatus**: El estado civil del empleado (por ejemplo, "Single", "Married", etc.).

- NaN

- 651 nulos

- object

- Count: 963

- Unique: 5

- Top: Married

- Freq: 404

19. **MonthlyIncome**: Ingresos mensuales del empleado.

- 843 nulos

- object

- Count: 771

- Unique: 668

- Top: 6347,0

- Freq: 4

20. **MonthlyRate**: Tasa mensual del empleado.

- no nulos

- int64

- media: 14284.4956

- std: 7110.41

- min: 2094.0

- max: 26999.0

21. **NumCompaniesWorked**: Número de compañías en las que el empleado ha trabajado.

- no nulos

- int64

- cambiar nombre de columna minúsculas

- media: 2.6734

- std: 2.50

- min: 0

- max: 9

22. **Over18**: Indica si el empleado es mayor de 18 años.

- (Y)

- Nan

- 901 nulos

- object

- Count: 713

- Unique: 1

- Top: Y

- Freq: 713

23. **OverTime**: Indica si el empleado trabaja horas extra ("Yes" para sí o "No" para no).

- 676 nulos

- object

- Count: 938

- Unique: 2

- Top: no

- Freq: 682

24. **PercentSalaryHike**: El porcentaje de aumento salarial del empleado.

- no nulos

- int64

- media: 15.1654

- std: 3.64

- min: 11

- max: 25

25. **PerformanceRating**: Calificación de rendimiento del empleado.

- decidir puntuación todas en int o todas en float

- 195 nulos

- object

- Count: 1419

- Unique: 2

- Top: 3,0

- Freq: 1205

26. **RelationshipSatisfaction**: Nivel de satisfacción en las relaciones interpersonales del empleado.

- no nulos

- int64

- media: 2.7044

- std: 1.07

- min: 1

- max: 4

- valoraciones del 0 al 4 🡪 correcto

27. **StandardHours**: Las horas estándar de trabajo.

- 1195 nulos

- object

- Count: 419

- Unique: 1

- Top: 80,0

- Freq: 419

28. **StockOptionLevel**: Nivel de opciones de compra de acciones del empleado.

- no nulos

- int64

- media: 2.7044

- std: 1.07

- min: 1

- max: 4

29. **TotalWorkingYears**: Total de años de experiencia laboral del empleado.

- 526 nulos

- object

- Count: 1088

- Unique: 40

- Top:10,0

- Freq:144

30. **TrainingTimesLastYear**: Número de veces que el empleado recibió capacitación el año pasado.

- no nulos

- int 64

- media: 2.8097

- std: 1.29

- min: 0

- max: 6

31. **WorkLifeBalance**: Equilibrio entre trabajo y vida personal del empleado.

- 108 nulos

- object

- nombre columna todo en mayúscula

- Count:1506

- Unique: 4

- Top: 3,0

- Freq: 913

32. **YearsAtCompany**: Años que el empleado ha trabajado en la empresa actual.

- no nulos

- int64

- media: 7.1325

- std: 6.12

- min: 0

- max: 40

33. **YearsInCurrentRole**: Años que el empleado ha estado en su puesto actual.

- 1580 nulos

- object

- Count: 34

- Unique: 10

- Top: 2,0

- Freq: 11

34. **YearsSinceLastPromotion**: Años desde la última promoción del empleado.

- no nulos

- int64

- media: 2.2459

- std: 3.23

- min: 0

- max: 15

35. **YearsWithCurrManager**: Años que el empleado ha estado bajo la supervisión del actual gerente.

- no nulos

- int64

- media: 4.2205

- std: 3.56

- min: 0

- max: 17

36. **SameAsMonthlyIncome**: Ingresos mensuales del empleado.

- comprobar para errores y eliminar

- 843 nulos

- object

- Count: 771

- Unique: 668

- Top: 6347,0

- Freq: 4

37. **DateBirth**: Año de nacimiento del empleado (teniendo en cuenta que los datos fueron recogidos en el 2023)

- no nulos

- int64 🡪 cambiar a datetime?

- media: 1986.0762

- std: 9.10

- min: 1963

- max: 2005

38. **Salary**: Salario de los empleados.

- todos son iguales, calcularlo a partir del mensual, diario, ….

- no nulos

- object

- Count: 1614

- Unique: 1

- Top: 1000000000$

- Freq: 1614

39. **RoleDepartament**: El departamento y el rol del empleado.

- usar el rol para completar errores del JobRole y eliminar (dejar solo departamento)

- 1312 nulos

- object

- Count: 302

- Unique: 301

- Top: Manager sales

- Freq: 2

40. **NumberChildren**: Número de hijos de los empleados

- todo nulos (eliminar columna)

- float64

41. **RemoteWork**: Si el empleado puede teletrabajar o no.

- False, 0 ,1, … decidir tipo de dato

- no nulos

- object

- Count: 1614

- Unique: 5

- Top: 1

- Freq: 360

Procedimiento:

- Estudiar dataframe: nulos, tipos de dato x columna, datos atípicos que llamen la atención

- cambiamos nombres de columna a minúsculas y con \_ entre palabras