

Deserción de personal

ALONSO MEDINA

CODERHOUSE - DATA SCIENCE

PRIMERA PRE ENTREGA

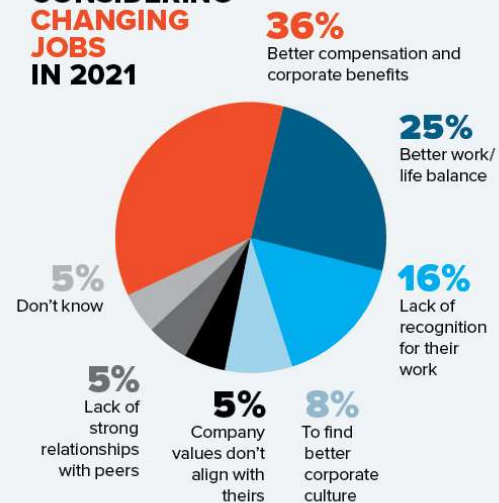
Problema

The Great Attrition or the Great Attraction?

A record number of employees are quitting or thinking about doing so. Organizations that take the time to learn why—and act—will have an edge in attracting and retaining talent



WHY EMPLOYEES ARE CONSIDERING CHANGING JOBS IN 2021



Source: *Engagement and Retention Report*, Achievers Workforce Institute, February 2021.

SE ESTIMA QUE EL COSTO DE LA DESERCIÓN EN EMPRESAS PRIVADAS ES ALREDEDOR DE 6 A 9 MESES DEL SALARIO DEL TRABAJADOR, SIN CONTAR EL COSTO OPORTUNIDAD DE NO TENER A UN EMPLEADO CAPACITADO PARA LOGRAR LOS OBJETIVOS DEL NEGOCIO.

Fuente de datos

IBM Employee Dataset

Data Code (7) Discussion (0)

▲ 53

New Notebook

test.csv (62.71 kB)


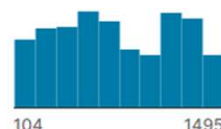



Detail Compact Column

10 of 34 columns

About this file

Train.csv

# Age	▲ BusinessTravel	# DailyRate	▲ Department	# DistanceFromHo...												
Age	Business Travel	Daily Rate	Department	Distance From Home												
	<table><tr><td>Travel_Rarely</td><td>72%</td></tr><tr><td>Travel_Frequently</td><td>17%</td></tr><tr><td>Other (43)</td><td>10%</td></tr></table>	Travel_Rarely	72%	Travel_Frequently	17%	Other (43)	10%		<table><tr><td>Research & Develo...</td><td>63%</td></tr><tr><td>Sales</td><td>31%</td></tr><tr><td>Other (25)</td><td>6%</td></tr></table>	Research & Develo...	63%	Sales	31%	Other (25)	6%	
Travel_Rarely	72%															
Travel_Frequently	17%															
Other (43)	10%															
Research & Develo...	63%															
Sales	31%															
Other (25)	6%															

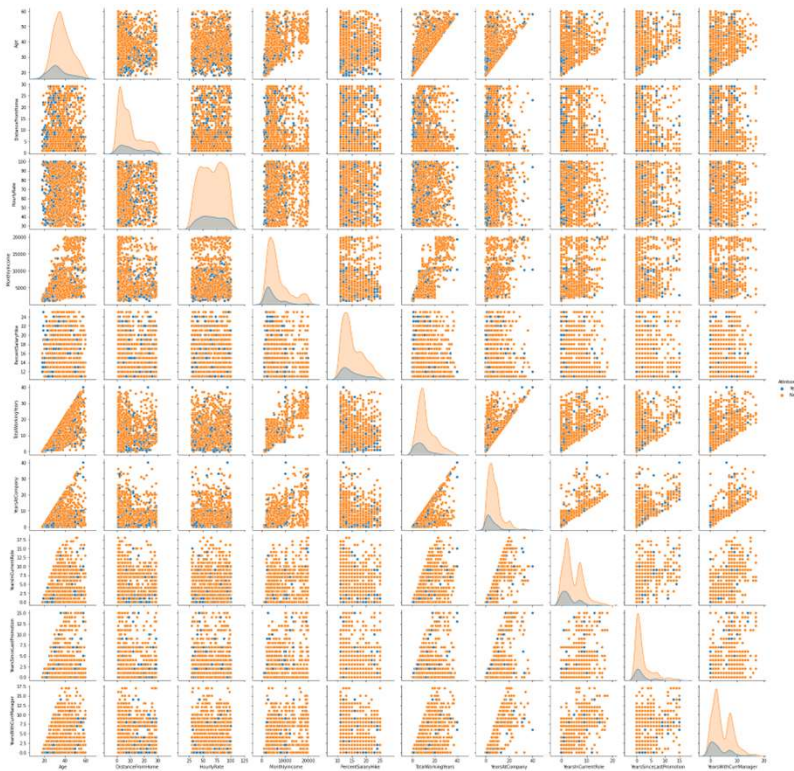
EL CONJUNTO DE DATOS UTILIZADO ES FICTICIO, FUE CREADO POR LOS CIENTÍFICOS DE DATOS DE IBM WATSON. TIENE 35 ATRIBUTOS DE ALREDEDOR DE 1470 EMPLEADOS.

Exploratory Data Analysis EDA

	count	mean	std	min	25%	50%	75%	max
Age	1470.0	36.923810	9.135373	18.0	30.00	36.0	43.00	60.0
DailyRate	1470.0	802.485714	403.509100	102.0	465.00	802.0	1157.00	1499.0
DistanceFromHome	1470.0	9.192517	8.106864	1.0	2.00	7.0	14.00	29.0
Education	1470.0	2.912925	1.024165	1.0	2.00	3.0	4.00	5.0
EmployeeCount	1470.0	1.000000	0.000000	1.0	1.00	1.0	1.00	1.0
EmployeeNumber	1470.0	1024.865306	602.024335	1.0	491.25	1020.5	1555.75	2068.0
EnvironmentSatisfaction	1470.0	2.721769	1.093082	1.0	2.00	3.0	4.00	4.0
HourlyRate	1470.0	65.891156	20.329428	30.0	48.00	66.0	83.75	100.0
JobInvolvement	1470.0	2.729932	0.711561	1.0	2.00	3.0	3.00	4.0
JobLevel	1470.0	2.063946	1.106940	1.0	1.00	2.0	3.00	5.0
JobSatisfaction	1470.0	2.728571	1.102846	1.0	2.00	3.0	4.00	4.0
MonthlyIncome	1470.0	6502.931293	4707.956783	1009.0	2911.00	4919.0	8379.00	19999.0
MonthlyRate	1470.0	14313.103401	7117.786044	2094.0	8047.00	14235.5	20461.50	26999.0

Exploraremos la data para identificar patrones y probar hipótesis. Posteriormente verificaremos estas suposiciones con la ayuda de estadística.

Data Wrangling (Munging)



En este análisis observaremos las variables que mayor correlación muestran, como por ejemplo:

- Edad
- Total de años en el trabajo
- Total de años en el puesto
- Nivel
- Lejanía del lugar de trabajo