



Instituto Tecnológico
de Buenos Aires

PEREZ HOCHBERG AUGUSTO

ANÁLISIS PREDICTIVO

—



Introducción

Base de datos elegida:

"IBM HR Analytics Employee Attrition & Performance"

Link de la base

"Uncover the factors that lead to employee attrition and explore important questions such as 'show me a breakdown of distance from home by job role and attrition' or 'compare average monthly income by education and attrition'. This is a fictional data set created by IBM data scientists."

Objetivo:

- Realizar una limpieza, análisis y graficación de los resultados obtenidos.
- Explorar las relaciones entre las diferentes variables y ver qué podemos inferir de ellas.

Observaciones iniciales

- La base está compuesta por 1470 observaciones y con 35 variables (8 numéricas), de las cuales se eliminaron 5 por la falta de aporte al análisis.
- No hay datos faltantes dentro de la base de datos (NA).
- Hay outliers presentes en el dataset, principalmente en aquellas variables correspondientes a tiempo, bajo la definición de Valor Atípico Leve ($Q1 - 1.5 * IQR$ y $Q3 + 1.5 * IQR$)
- Un ejemplo al respecto de los outliers es la variable “MonthlyIncome”, que presenta un total del 7,75% de observaciones fuera de rango (114 observaciones).

Composición de la base

Attrition:

No	1233
Yes	237

Education:

1	170
2	282
3	572
4	398
5	48

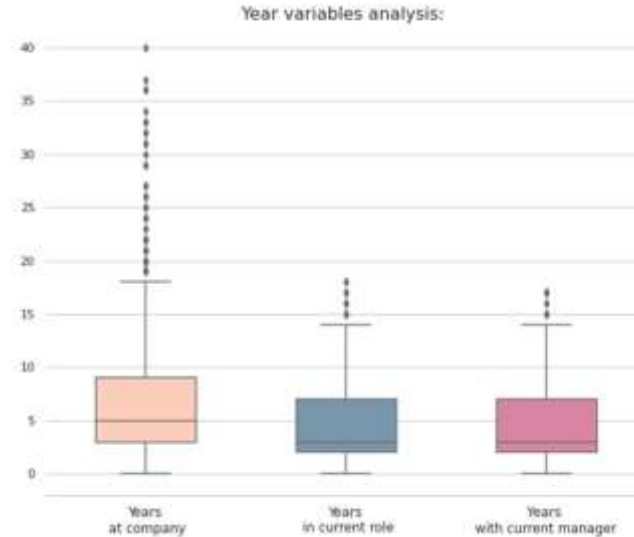
Job Level:

1	543
2	534
3	218
4	106
5	69

#	Column	Non-Null Count	Dtype
0	Attrition	1470 non-null	object
1	Age	1470 non-null	int64
2	Gender	1470 non-null	object
3	MaritalStatus	1470 non-null	object
4	NumCompaniesWorked	1470 non-null	int64
5	TotalWorkingYears	1470 non-null	int64
6	BusinessTravel	1470 non-null	object
7	DistanceFromHome	1470 non-null	int64
8	Department	1470 non-null	object
9	Education	1470 non-null	int64
10	EducationField	1470 non-null	object
11	HourlyRate	1470 non-null	int64
12	DailyRate	1470 non-null	int64
13	MonthlyRate	1470 non-null	int64
14	MonthlyIncome	1470 non-null	int64
15	EnvironmentSatisfaction	1470 non-null	int64
16	JobInvolvement	1470 non-null	int64
17	JobLevel	1470 non-null	int64
18	JobRole	1470 non-null	object
19	JobSatisfaction	1470 non-null	int64
20	OverTime	1470 non-null	object
21	PercentSalaryHike	1470 non-null	int64
22	PerformanceRating	1470 non-null	int64
23	RelationshipSatisfaction	1470 non-null	int64
24	TrainingTimesLastYear	1470 non-null	int64
25	WorkLifeBalance	1470 non-null	int64
26	YearsAtCompany	1470 non-null	int64
27	YearsInCurrentRole	1470 non-null	int64
28	YearsSinceLastPromotion	1470 non-null	int64
29	YearsWithCurrManager	1470 non-null	int64

Análisis outliers “Monthly Income” - Years

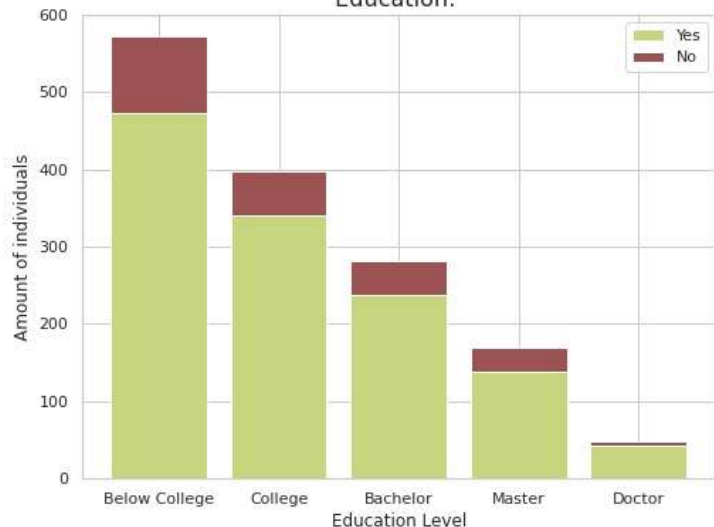
```
count    1470.000000
mean     6502.931293
std      4707.956783
min      1009.000000
25%      2911.000000
50%      4919.000000
75%      8379.000000
max      19999.000000
```



Educación y renuncia

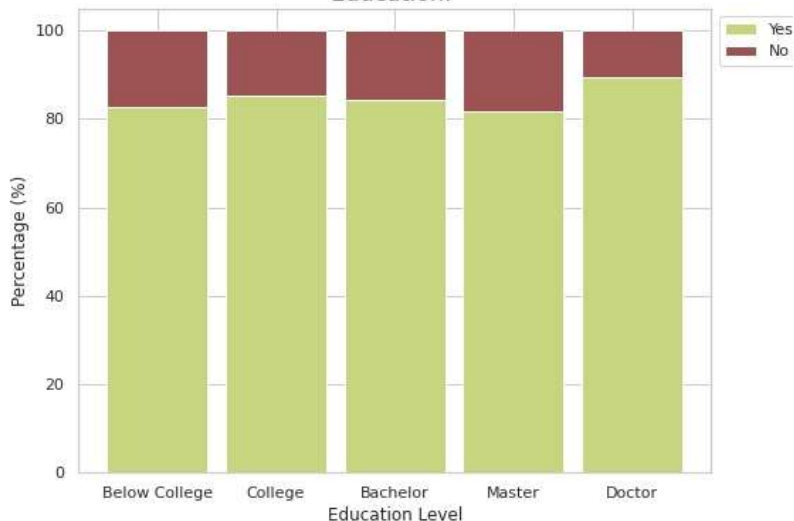
Valores absolutos:

Is the employee still working for us?
Education:

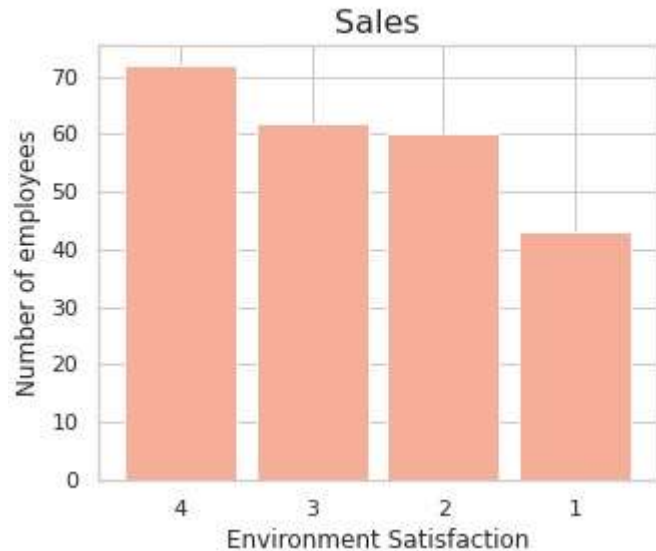


Valores relativos:

Is the employee still working for us?
Education:



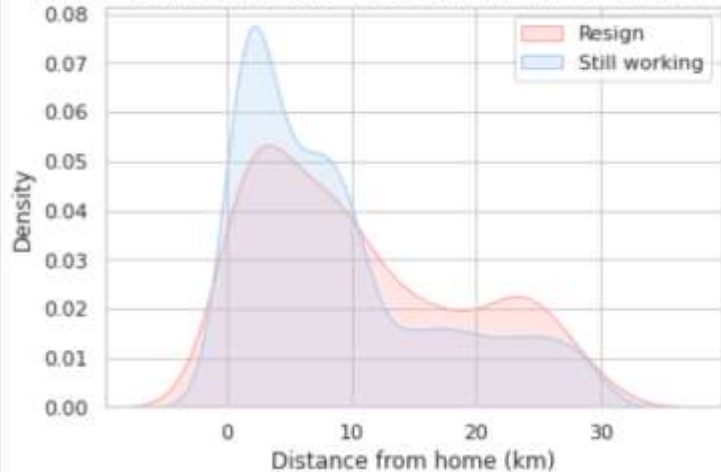
Satisfacción en el ambiente laboral y renuncias



EnvironmentSatisfaction	Department	
1	Human Resources	4
	Research & Development	43
	Sales	25
2	Human Resources	3
	Research & Development	18
	Sales	22
3	Human Resources	2
	Research & Development	38
	Sales	22
4	Human Resources	3
	Research & Development	34
	Sales	23

Transporte y personal

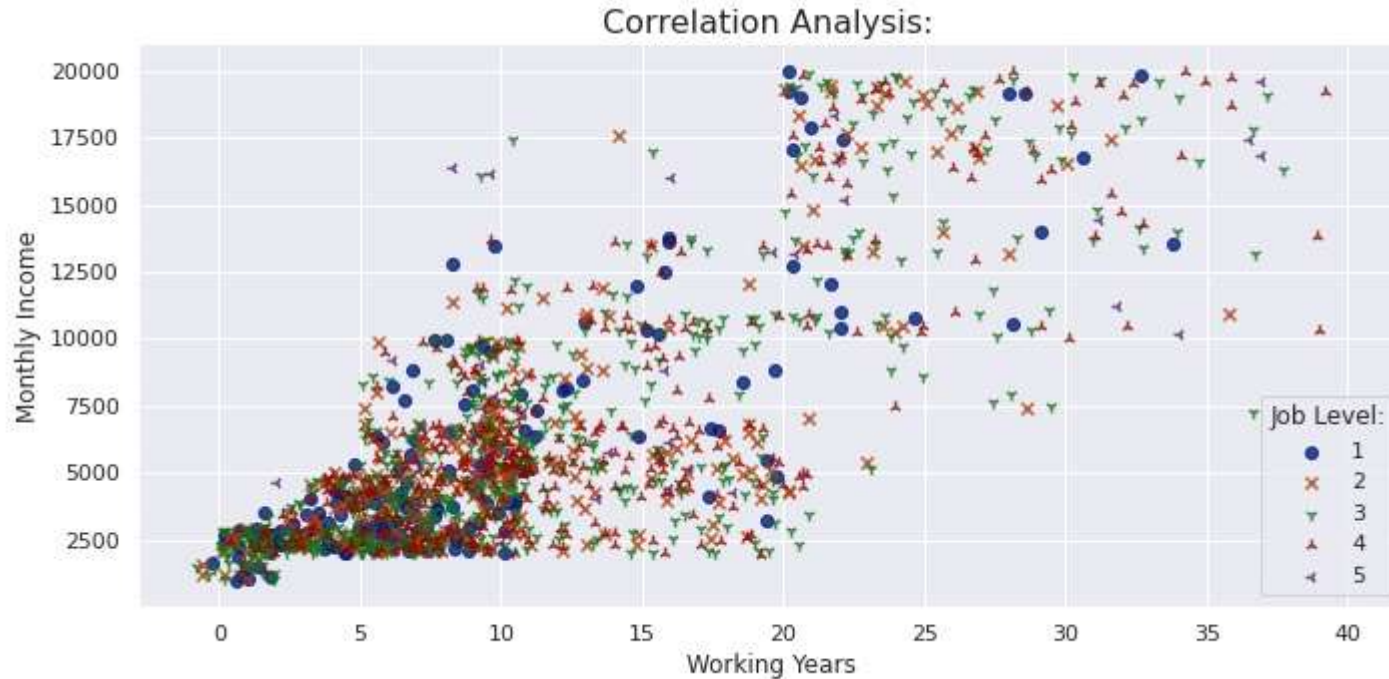
What is the distance between home and the office?



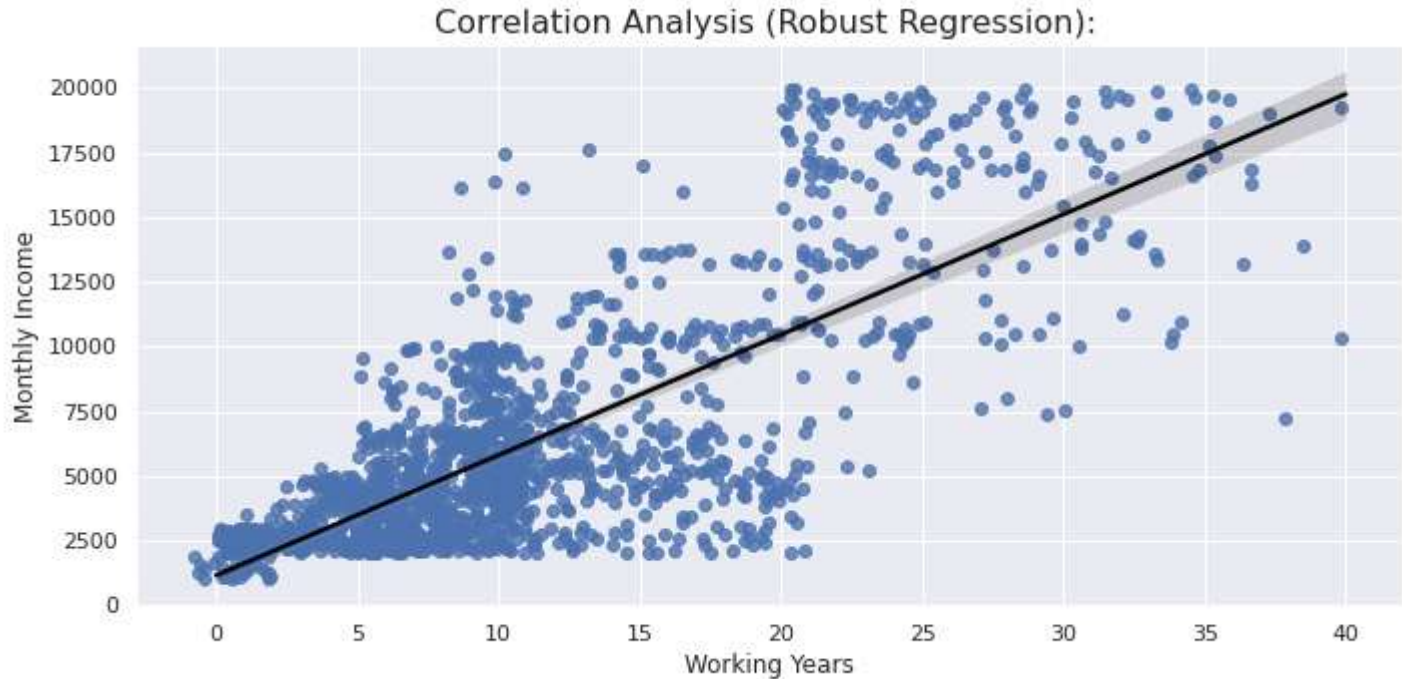
```
Data for those that have resigned:
Median: 9.0
count    237.000000
mean     10.632911
std      8.452525
min       1.000000
25%       3.000000
50%       9.000000
75%      17.000000
max      29.000000
```

```
Data for those that are still working with us:
Median: 7.0
count   1233.000000
mean     8.915653
std      8.012633
min       1.000000
25%       2.000000
50%       7.000000
75%      13.000000
max      29.000000
```


Análisis de correlación



Análisis de correlación



Conclusiones

- Las variables parecen tener poca correlación entre sí, a excepción de casos en específico.
- La educación recibida y la satisfacción de las personas dentro del ambiente laboral no parecen ser causas determinantes en el motivo de renuncia de las personas.
- El nivel educativo no demuestra, según el análisis en la correlación de las variables, ser un factor determinante en el ingreso recibido por los empleados.
- Sorpresivamente son más quienes renuncian teniendo que viajar al lugar de trabajo en pocas ocasiones.

BusinessTravel	Attrition	
Non-Travel	0	138
	1	12
Travel_Frequently	0	208
	1	69
Travel_Rarely	0	887
	1	156