

Los Salarios

Miguel Alejandro Salas Reyna A00827219

Resumen

El presente documento tiene como objetivo emplear técnicas de procesamiento de datos para el análisis estadístico y para la construcción de modelos a través de la base de datos que proporciona Kaggle para el análisis de salarios de especialistas en análisis de datos.

Introducción

La rápida generación de datos en los últimos años ha concluido en la demanda de especialistas que mediante estadística y programación puedan transformar información útil para las empresas. En la base de datos proporcionada por Kaggle podemos encontrar datos del salario, ubicación de la empresa, nivel de experiencia laboral, el tamaño de la empresa y el puesto de trabajo. La mayoría de los están directamente relacionados con el manejo de los datos en cualquiera de sus etapas de producción y en este reporte encontraremos la respuesta a preguntas como: ¿Cuál es el salario al que pueda aspirar un analista de datos?, ¿En qué países se ofrecen mejores salarios? o ¿Influye el nivel de experiencia en el salario?, entre otras.

Análisis de los resultados

Lo primero que se realizó fue un análisis exploratorio de los datos, identificar qué es lo que tenemos y qué podemos hacer con ellos. De esta manera podemos identificar las siguientes variables:

Int64Index: 607 entries, 0 to 606

Data columns (total 11 columns):

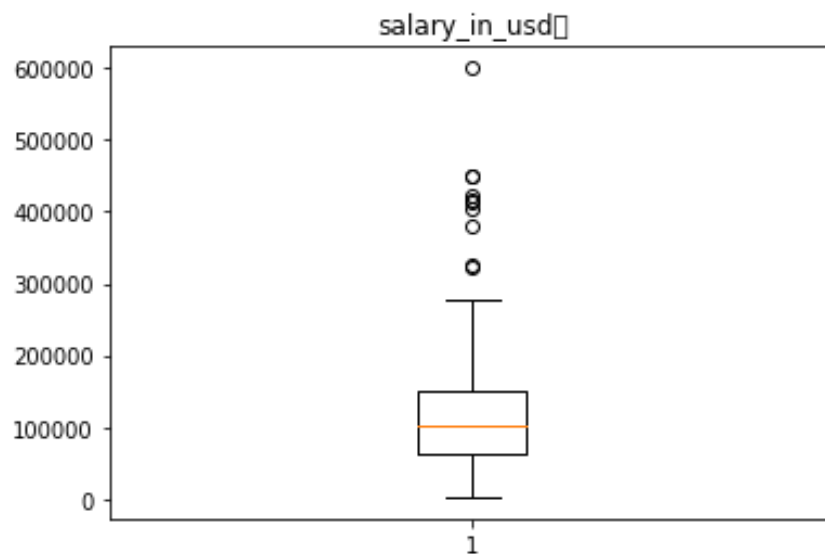
#	Column	Non-Null Count	Dtype
0	work_year	607 non-null	int64
1	experience_level	607 non-null	object
2	employment_type	607 non-null	object
3	job_title	607 non-null	object
4	salary	607 non-null	int64
5	salary_currency	607 non-null	object
6	salary_in_usd	607 non-null	int64
7	employee_residence	607 non-null	object
8	remote_ratio	607 non-null	int64
9	company_location	607 non-null	object
10	company_size	607 non-null	object

Podemos identificar 11 columnas, siendo las más que más destacan a simple vista el salario en dólares, la experiencia, el nombre del trabajo y la locación del mismo. Entre las variables numéricas podemos encontrar las correlaciones correspondientes:

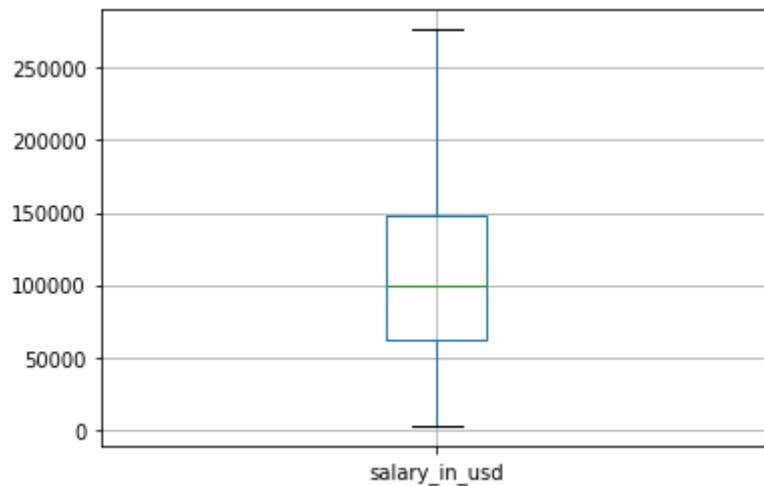
	work_year	salary	salary_in_usd	remote_ratio
work_year	1.000000	-0.087577	0.170493	0.076314
salary	-0.087577	1.000000	-0.083906	-0.014608
salary_in_usd	0.170493	-0.083906	1.000000	0.132122
remote_ratio	0.076314	-0.014608	0.132122	1.000000

Relativamente, no tienen altas correlaciones, las más destacadas son el año de trabajo y si es remoto para el caso del salario en dólares.

En la información proporcionada por las variables podemos encontrar que no hay valores nulos y que es una base de datos relativamente limpia por lo que no es necesario realizar un preprocesamiento extremo para la limpieza de la misma.



Por ejemplo, al analizar algunos boxplots podemos observar que realmente la cantidad de valores atípicos es mínima. En el caso de salarios en dólares podemos encontrar 10 valores atípicos.



Ahora, podemos encontrar la solución de algunas de las preguntas planteadas al inicio.

¿Cuál es el salario al que pueda aspirar un analista de datos?

Obteniendo el promedio de salarios en dólares, una buen referencia o estimación del salario al que puede aspirar un científico de datos es de \$107,168 dólares. Esto si nos basamos en una media de salarios.

¿En qué países se ofrecen mejores salarios?

De acuerdo a la media por país, Rusia encabeza la lista seguido por Estados Unidos.

```
company_location
AE      100000.000000
AS      18053.000000
AT       72920.750000
AU     108042.666667
BE       85699.000000
BR      18602.666667
CA      99823.733333
CH       64114.000000
CL       40038.000000
CN      71665.500000
CO       21844.000000
CZ       50937.000000
DE       81887.214286
DK       54386.333333
DZ     100000.000000
EE       32974.000000
ES       53060.142857
FR       63970.666667
```

Estadística para ciencia de datos y nombre de la concentración. Grupo 102.

GB	81583.042553
GR	52293.090909
HN	20000.000000
HR	45618.000000
HU	35735.000000
IE	71444.000000
IL	119059.000000
IN	28581.750000
IQ	100000.000000
IR	4000.000000
IT	36366.500000
JP	114127.333333
KE	9272.000000
LU	43942.666667
MD	18000.000000
MT	28369.000000
MX	32123.333333
MY	40000.000000
NG	30000.000000
NL	54945.750000
NZ	125000.000000
PK	13333.333333
PL	66082.500000
PT	47793.750000
RO	60000.000000
RU	157500.000000
SG	89294.000000
SI	63831.000000
TR	20096.666667
UA	13400.000000
US	136100.342029
VN	4000.000000

¿Influye el nivel de experiencia en el salario?

experience_level	
EN	61643.318182
EX	159963.318182
MI	82953.142857
SE	135797.263538

Sí influye el nivel de experiencia ya que conforme vaya subiendo de nivel aumenta el salario considerablemente.

Concluye

Los análisis exploratorios nos permiten encontrar información entre los datos, de esta manera podemos encontrar lo que está más allá de lo que simplemente se muestra y resolver cuestiones o dudas presentes. Con ello, podemos identificar

problemas planteados y encontrar soluciones. En este reporte se respondieron las tres preguntas planteadas y aunque algunas preguntas podrían encontrar solución por sentido común, no está de más comprobar lo que realmente dicen los datos porque en ocasiones nos podemos encontrar que la realidad es muy diferente.

Anexos

<https://github.com/miiguelsr/PortafolioAnalisis.git>