

Momento de Retroalimentación Estadística

Ricardo Arriaga A01570553

Introducción y Resumen

Este análisis busca identificar cuáles son las condiciones que hacen a una persona del área de análisis de datos tenga un mejor sueldo. Para esto se respondieron tres preguntas que dan un panorama para comprender las opciones que un analista de datos tiene. Dichas preguntas con sus resultados son:

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

- En promedio 92893.06 USD

¿En qué países se ofrecen mejores salarios?

- RU Russian Federation 157500.0 USD

¿Qué tanto afecta los salarios dependiendo del radio de trabajo remoto?

- 0% - 50% remoto: -18.20465 %
- 50% - 100% remoto: 82.82617 %
- 0% - 100% remoto: 36.70695 %

Para la obtención de estos resultados se realizó una limpieza de datos eliminando variables que interfieran con el análisis y así poder buscar con datos estandarizados la información requerida.

Este análisis es útil ya que indica tendencias y opciones de trabajo para un aspirante o un actual trabajador del área de ciencia de datos. En especial creemos que la primera y tercera pregunta son las importantes ya que muestran el panorama actual en un mundo postpandemia, en donde el trabajo online comenzó a hacerse más común.

Análisis de los resultados

Calidad de datos

Para empezar, comprobemos la calidad de los datos buscando valores vacíos que puedan interferir en el futuro.

```

## ----- work_year -----
## NAs: 0
## ----- experience_level -----
## NAs: 0
## ----- employment_type -----
## NAs: 0
## ----- job_title -----
## NAs: 0
## ----- salary -----
## NAs: 0
## ----- salary_currency -----
## NAs: 0
## ----- salary_in_usd -----
## NAs: 0
## ----- employee_residence -----
## NAs: 0
## ----- remote_ratio -----
## NAs: 0
## ----- company_location -----
## NAs: 0
## ----- company_size -----
## NAs: 0

```

Medidas de tendencia central y dispersión

Para comenzar con el análisis de esta actividad abra que hacer una exploración preliminar de los datos de nuestro dataset. En esta exploración buscaremos datos que nos ayuden a ver la dispersión de los datos y entender tendencias que se puedan presentar más adelante. Para variables numéricas buscaremos el promedio, mediana, moda, desviación estándar, varianza y su mínimo y máximo valor y para variables categóricas buscaremos su moda y la distribución de las categorías.

```

## ----- work_year -----
## Promedio: 2021.405 Mediana: 2022 Moda: 2022
## Desviacion estandar: 0.692133 Varianza: 0.4790481
## Minimo: 2020 Maximo: 2022
##
## ----- salary -----
## Promedio: 324000.1 Mediana: 115000 Moda: 80000 100000
## Desviacion estandar: 1544357 Varianza: 2.38504e+12
## Minimo: 4000 Maximo: 30400000
##
## ----- salary_in_usd -----
## Promedio: 112297.9 Mediana: 101570 Moda: 100000
## Desviacion estandar: 70957.26 Varianza: 5034932663
## Minimo: 2859 Maximo: 600000
##
## ----- remote_ratio -----
## Promedio: 70.92257 Mediana: 100 Moda: 100

```

```

## Desviacion estandar: 40.70913   Varianza: 1657.233
## Minimo: 0   Maximo: 100

## ----- experience_level -----
## Moda: SE
##
##
## EN EX MI SE
## 88 26 213 280
##
## ----- employment_type -----
## Moda: FT
##
##
## CT FL FT PT
## 5 4 588 10
##
## ----- job_title -----
## Moda: Data Scientist
##

```

Titulo	Numero
3D Computer Vision Researcher	1
AI Scientist	7
Analytics Engineer	4
Applied Data Scientist	5
Applied Machine Learning Scientist	4
BI Data Analyst	6
Big Data Architect	1
Big Data Engineer	8
Business Data Analyst	5
Cloud Data Engineer	2
Computer Vision Engineer	6
Computer Vision Software Engineer	3
Data Analyst	97
Data Analytics Engineer	4
Data Analytics Lead	1
Data Analytics Manager	7
Data Architect	11
Data Engineer	132
Data Engineering Manager	5
Data Science Consultant	7
Data Science Engineer	3
Data Science Manager	12
Data Scientist	143
Data Specialist	1
Director of Data Engineering	2

Director of Data Science	7
ETL Developer	2
Finance Data Analyst	1
Financial Data Analyst	2
Head of Data	5
Head of Data Science	4
Head of Machine Learning	1
Lead Data Analyst	3
Lead Data Engineer	6
Lead Data Scientist	3
Lead Machine Learning Engineer	1
Machine Learning Developer	3
Machine Learning Engineer	41
Machine Learning Infrastructure Engineer	3
Machine Learning Manager	1
Machine Learning Scientist	8
Marketing Data Analyst	1
ML Engineer	6
NLP Engineer	1
Principal Data Analyst	2
Principal Data Engineer	3
Principal Data Scientist	7
Product Data Analyst	2
Research Scientist	16
Staff Data Scientist	1

```

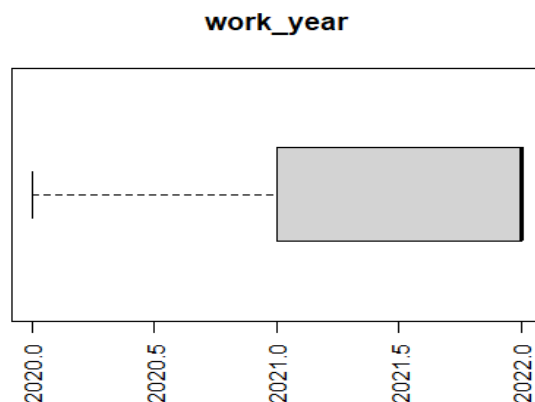
## ----- salary_currency -----
## Moda:  USD
##
##
## AUD BRL CAD CHF CLP CNY DKK EUR GBP HUF INR JPY MXN PLN SGD TRY USD
##  2  2 18  1  1  2  2 95 44  2 27  3  2  3  2  3 398
##
## ----- employee_residence -----
## Moda:  US
##
##
## AE  AR  AT  AU  BE  BG  BO  BR  CA  CH  CL  CN  CO  CZ  DE  DK  DZ
EE  ES  FR
##  3  1  3  3  2  1  1  6 29  1  1  1  1  1 25  2  1
1 15 18
## GB  GR  HK  HN  HR  HU  IE  IN  IQ  IR  IT  JE  JP  KE  LU  MD  MT
MX  MY  NG
## 44 13  1  1  1  2  1 30  1  1  4  1  7  1  1  1  1
2  1  2
## NL  NZ  PH  PK  PL  PR  PT  RO  RS  RU  SG  SI  TN  TR  UA  US  VN
##  5  1  1  6  4  1  6  2  1  4  2  2  1  3  1 332  3
##

```

```
## ----- company_location -----
## Moda:  US
##
##
##  AE  AS  AT  AU  BE  BR  CA  CH  CL  CN  CO  CZ  DE  DK  DZ  EE  ES
FR  GB  GR
##   3   1   4   3   2   3  30   2   1   2   1   2  28   3   1   1  14
15  47  11
##  HN  HR  HU  IE  IL  IN  IQ  IR  IT  JP  KE  LU  MD  MT  MX  MY  NG
NL  NZ  PK
##   1   1   1   1   1  24   1   1   2   6   1   3   1   1   3   1   2
4   1   3
##  PL  PT  RO  RU  SG  SI  TR  UA  US  VN
##   4   4   1   2   1   2   3   1 355   1
##
## ----- company_size -----
## Moda:  M
##
##
##   L   M   S
## 198 326  83
```

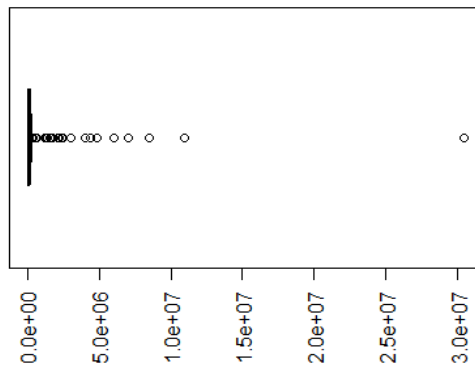
Medidas de distribución y medidas de posición

```
## ----- work_year -----
## Quartil 1:  2021   Quartil 3:  2022
```



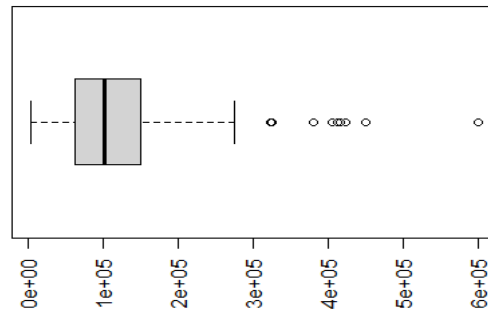
```
## ----- salary -----
## Quartil 1:  70000   Quartil 3: 165000
```

salary



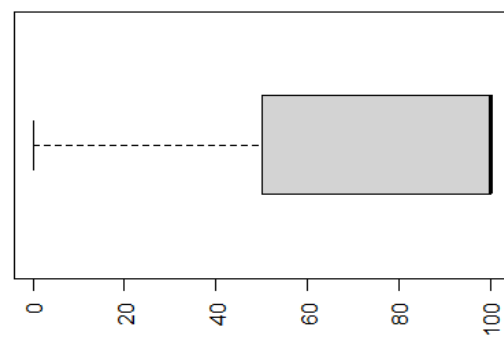
```
## ----- salary_in_usd -----  
## Quartil 1: 62726  Quartil 3: 150000
```

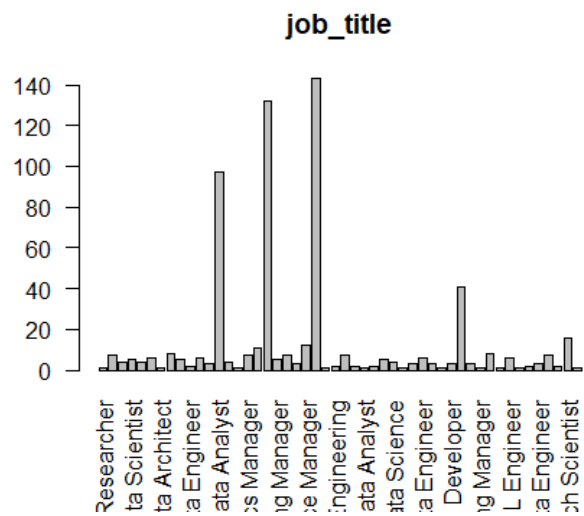
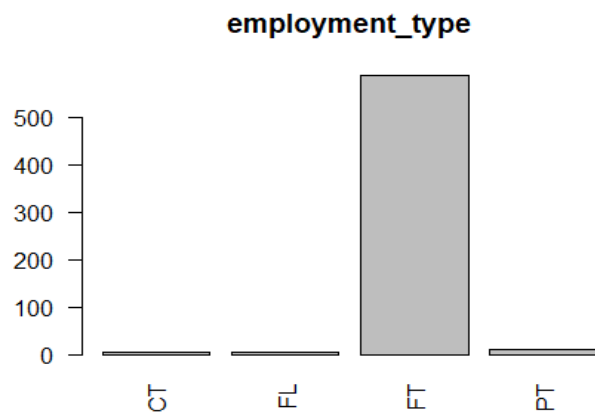
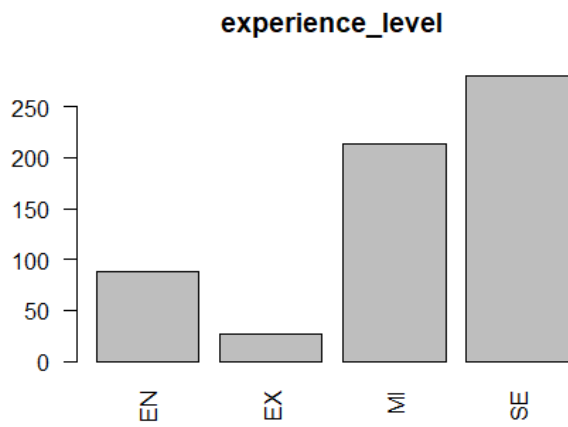
salary_in_usd

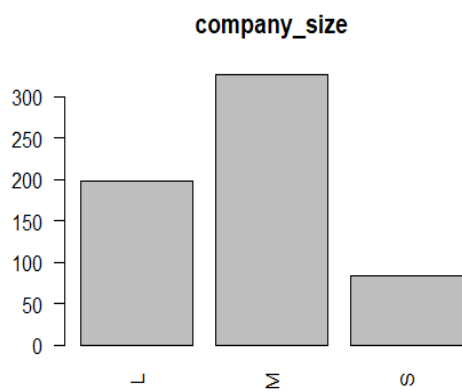
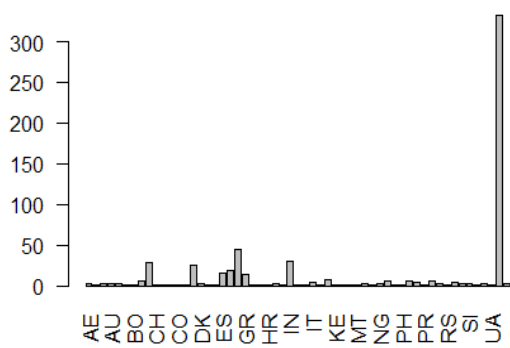
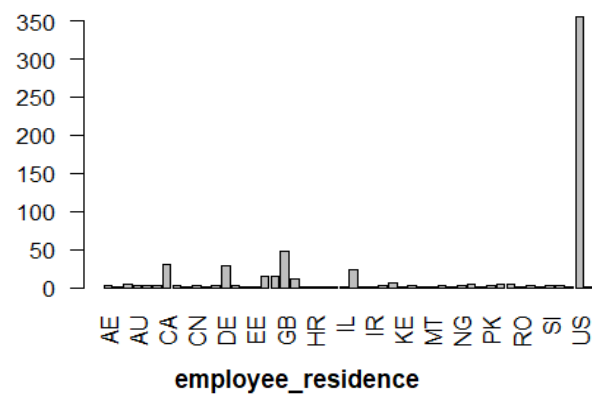
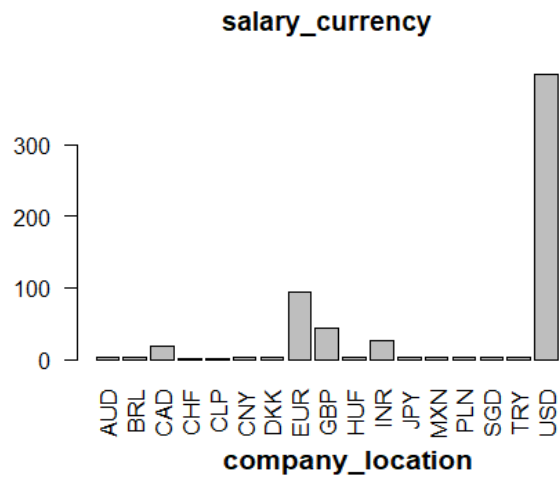


```
## ----- remote_ratio -----  
## Quartil 1: 50  Quartil 3: 100
```

remote_ratio





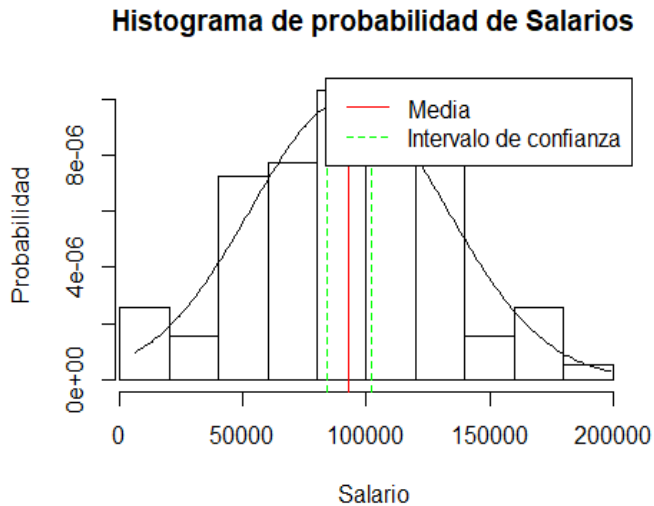


Preparación de datos

Para nuestras predicciones hemos decidido eliminar las columnas de salario y tipo de moneda, ya que buscamos mantener un estándar en cuanto a los salarios.

Pregunta 1

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



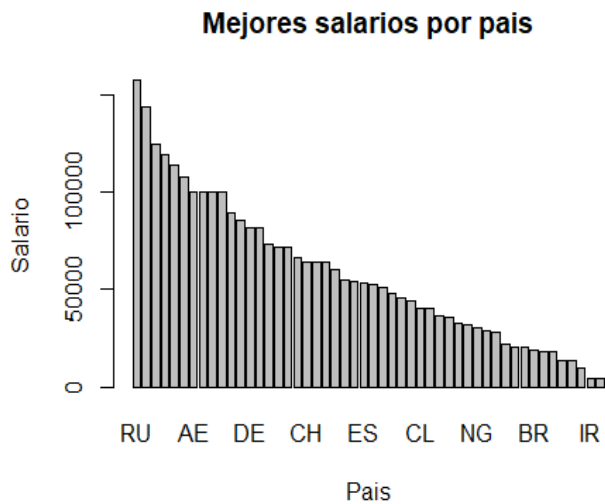
```
## En general, el salario al que puede aspirar un analista de datos es:
92893.06 USD
## La confianza en el salario al que puede aspirar un analista de datos
es:
## 83955.25 USD - 101830.9 USD
```

Pregunta 2

¿En qué países se ofrecen mejores salarios?

```
## En general, los mejores salarios son en:

##   Alias                Pais  Salario
## 1:   RU      Russian Federation 157500.0
## 2:   US United States of America 144055.3
## 3:   NZ          New Zealand 125000.0
## 4:   IL              Israel 119059.0
## 5:   JP              Japan 114127.3
```



Pregunta 3

¿Que tanto afecta los salarios dependiendo del radio de trabajo remoto?

```
## En general, el afecto de los salarios dependiendo del radio de trabajo
remoto es:
## 0% - 50% remoto: -18.20465 %
## 50% - 100% remoto: 82.82617 %
## 0% - 100% remoto: 36.70695 %
```

Con estos resultados podemos ver un comportamiento algo imprevisto en como aumentan los sueldos mientras más se trabaje desde casa, pero esto tiene una explicación. Los sueldos no suben entre más tiempo se pase fuera de la oficina, si no que las posiciones cambian. Los puestos de SE y EX (Senior y Executive) son los puestos en promedio mejor pagados y estos puestos han tendido a ir a puestos online.

Es por eso que vemos que si un empleado cambia a un sistema hibrido su sueldo cae ya que usualmente no va a cambiar a un puesto ejecutivo o senior, pero eso cambia cuando el empleado asciende a alguno de esos dos puestos su sueldo aumentará y podrá estar en un sistema más flexible. Eso lo podemos ver al subir del 50% al 100% de trabajo remoto y cuando se sube de 0% a 100% podemos ver también esta tendencia, aunque no todos ascienden a un puesto ejecutivo o senior.

Conclusion

En conclusión, podemos observar las 3 observaciones que se consideraron importantes para realizar este análisis y podemos dar una recomendación estudiada sobre como buscar un mejor sueldo dentro de la industria del análisis de datos.