



UNIVERSIDAD DE  
GUANAJUATO

# Reporte de práctica intermedia

*Universidad de Guanajuato*  
*Diego Eduardo Rosas*  
*González*

NUA: 908594  
Minería de Datos  
viernes, 8 de mayo de 2020

## Description

*The practice consists of the following small tasks of processing and analyzing of the data contained in the file. For each task, you must write a python function as part of the pratice.py code file.*

- 1. Compute the five-number summary, the boxplot, the mean, and the standard deviation for the annual salary per gender.*
- 2. Compute the five-number summary, the boxplot, the mean, and the standard deviation for the annual salary per ethnicity.*
- 3. Compute the five-number summary, the boxplot, the mean, and the standard deviation for the annual salary per developer type.*
- 4. Compute the median, mean and standard deviation of the annual salary per country.*
- 5. Obtain a bar plot with the frequencies of responses for each developer type.*
- 6. Plot histograms with 10 bins for the years of experience with coding per gender.*
- 7. Plot histograms with 10 bins for the average number of working hours per week, per developer type.*
- 8. Plot histograms with 10 bins for the age per gender.*
- 9. Compute the median, mean and standard deviation of the age per programming language.*
- 10. Compute the correlation between years of experience and annual salary.*
- 11. Compute the correlation between the age and the annual salary.*
- 12. Compute the correlation between educational level and annual salary. In this case, replace the string of the educational level by an ordinal index (e.g. Primary/elementary school = 1, Secondary school = 2, and so on).*
- 13. Obtain a bar plot with the frequencies of the different programming languages.*

# 1. Salario anual por género

No se asumió que el salario de 0 dólares estaba incorrecto debido a que no se sabe precisamente cuál es el salario mínimo que recibe un programador en cada país. Además, la cantidad de cero puede indicar que ese desarrollador es un practicante y por lo cual podría trabajar sin remuneración.

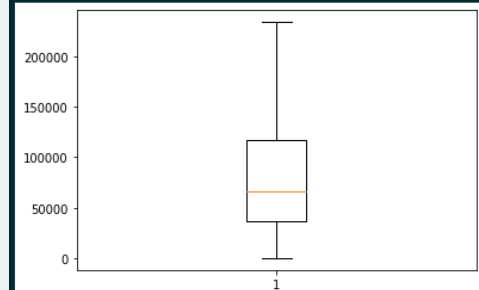
En el estudio hay tres tipos de géneros, hombres, mujeres y agrupados los no binarios y no género-conformes.

Los boxplot de los tres tienen una forma similar. Con la caja posicionada hacia abajo lo cuál quiere decir que la mayoría de los desarrolladores se encuentra en la posición media-baja en cuestión de salarios.

Usando la medida que menos varía con valores muy altos o bajos, es decir, la mediana, podemos observar que el grupo de los no binarios y no género-conformes es el que tiene los salarios más altos, le sigue el de las mujeres y al final el de los hombres, aunque entre estos dos hay muy poca diferencia.

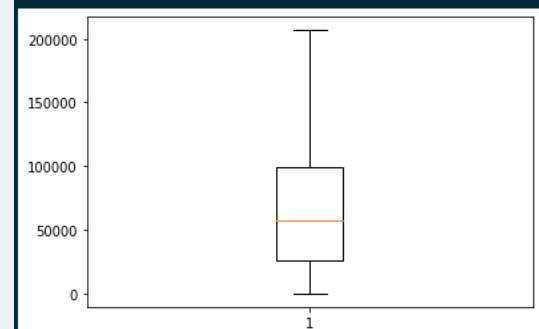
```
Non-binary, genderqueer, or gender non-conforming
Annual salary per gender:
Five number summary:
Min: 0.0
Max: 2000000.0
First quartile: 36042.0
Median: 65653.0
Third quartile: 117485.25

Mean: 154890.6462585034
Standard deviation 328791.4545744816
```



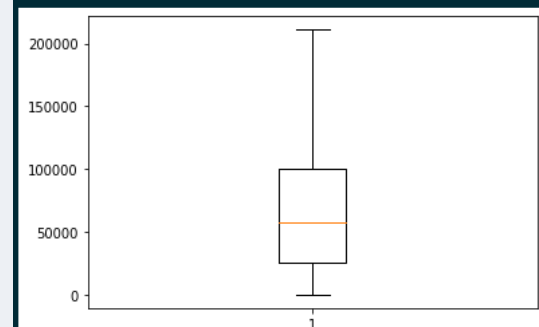
```
Woman
Annual salary per gender:
Five number summary:
Min: 0.0
Max: 2000000.0
First quartile: 26124.0
Median: 57287.0
Third quartile: 99000.0

Mean: 114066.78108314263
Standard deviation 255575.45397300526
```

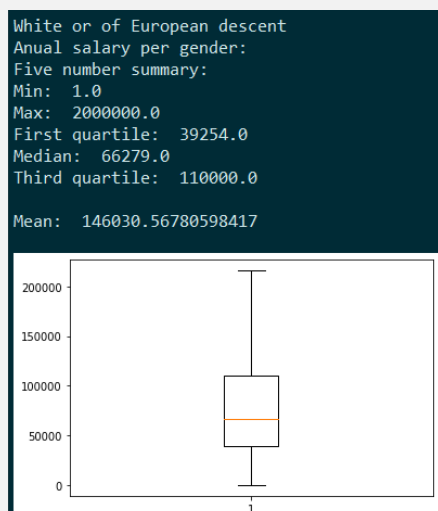
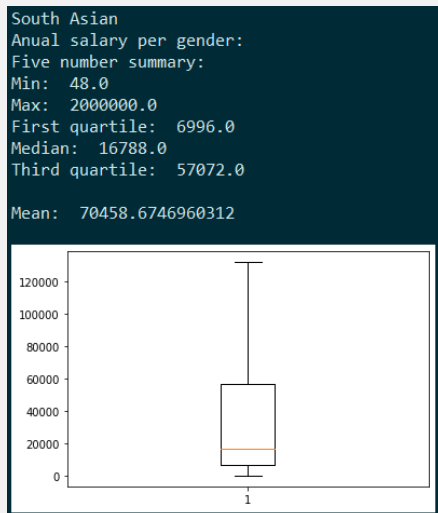


```
Man
Annual salary per gender:
Five number summary:
Min: 0.0
Max: 2000000.0
First quartile: 25656.0
Median: 57254.0
Third quartile: 100000.0

Mean: 127346.65901781662
Standard deviation 284521.4974200705
```



## 2. Análisis por etnia



Debido a la cantidad de etnias de las selecciones que había disponibles sólo nos fijaremos en la que tiene los mejores salarios y en la que tiene los salarios más bajos.

Utilizando como medida la mediana. El valor más alto se puede encontrar en el grupo de “White or of European descent” con una media de 66279.0 dólares.

Por el contrario, el valor más bajo está en el grupo “South Asian” con una media de 16788.0 dólares.

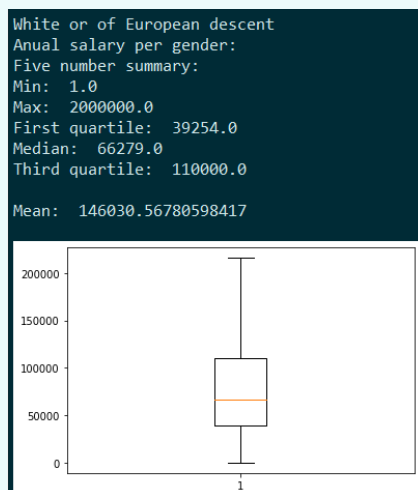
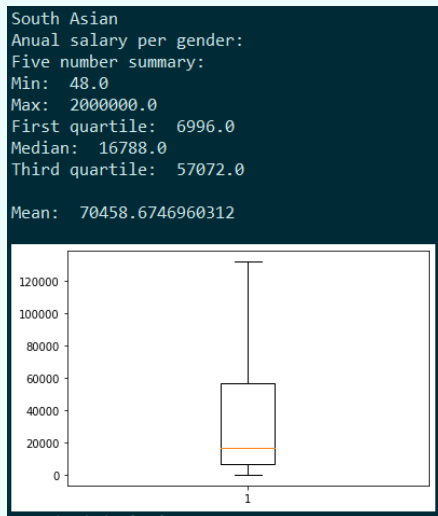
Los grupos de medio oriente y latinos también tenían sueldos más cercanos al valor de los del sur de asia que a los blancos o descendientes europeos.

Esto podría justificarse porque en su mayoría se localizan en Estados Unidos y Europa donde el salario es mayor sin embargo no hay que olvidar que esos países también tienen población afroamericana y asiática por lo cual debería de estar más equilibrados los valores.

## 3. Salario por tipo de desarrollador

Era previsible pero el tipo de desarrollador con el sueldo más alto es “Senior executive/VP” con 96000.0. Mientras que el tipo de desarrollador con la mediana de salario más bajo es “Student” con 18000.0. Se supondría que los años de experiencia son un fuerte factor a considerar en el salario que se recibe.

## 4. Análisis por país



Debido a la cantidad de etnias de las selecciones que había disponibles sólo nos fijaremos en la que tiene los mejores salarios y en la que tiene los salarios más bajos.

Utilizando como medida la mediana. El valor más alto se puede encontrar en el grupo de “White or of European descent” con una media de 66279.0 dólares.

Por el contrario, el valor más bajo está en el grupo “South Asian” con una media de 16788.0 dólares.

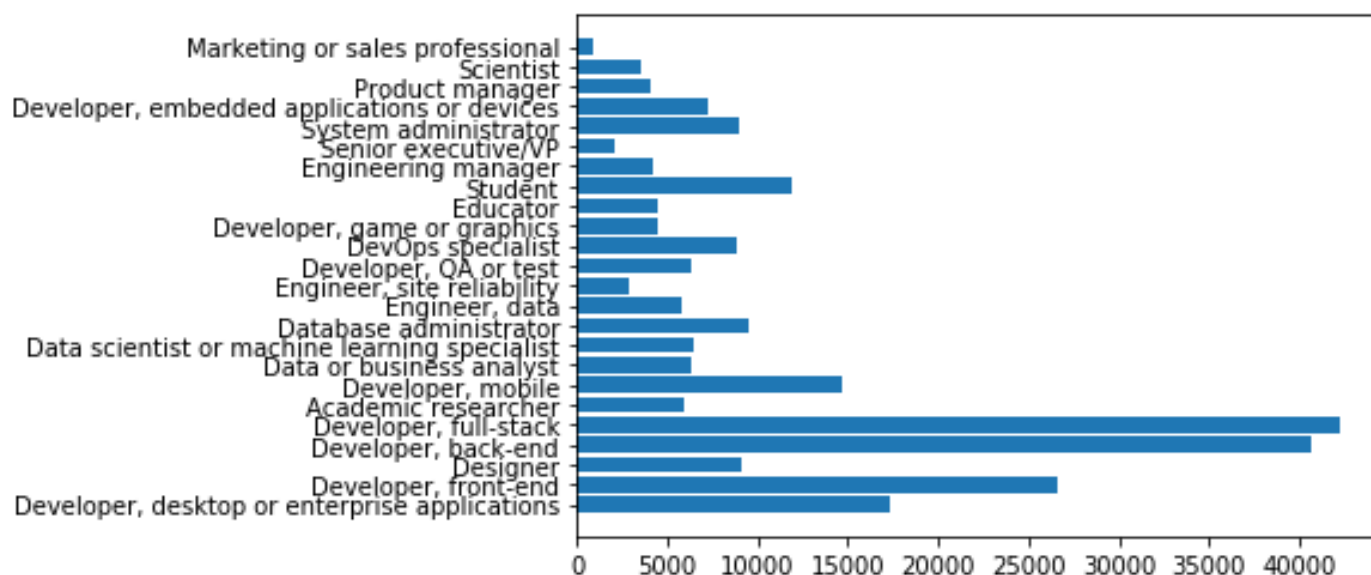
Los grupos de medio oriente y latinos también tenían sueldos más cercanos al valor de los del sur de asia que a los blancos o descendientes europeos.

Esto podría justificarse porque en su mayoría se localizan en Estados Unidos y Europa donde el salario es mayor sin embargo no hay que olvidar que esos países también tienen población afroamericana y asiática por lo cual debería de estar más equilibrados los valores.

## 5. Salario por tipo de desarrollador

Era previsible pero el tipo de desarrollador con el sueldo más alto es “Senior executive/VP” con 96000.0. Mientras que el tipo de desarrollador con la mediana de salario más bajo es “Student” con 18000.0. Se supondría que los años de experiencia son un fuerte factor a considerar en el salario que se recibe.

## 6. Frecuencia de respuestas por tipo de desarrollador



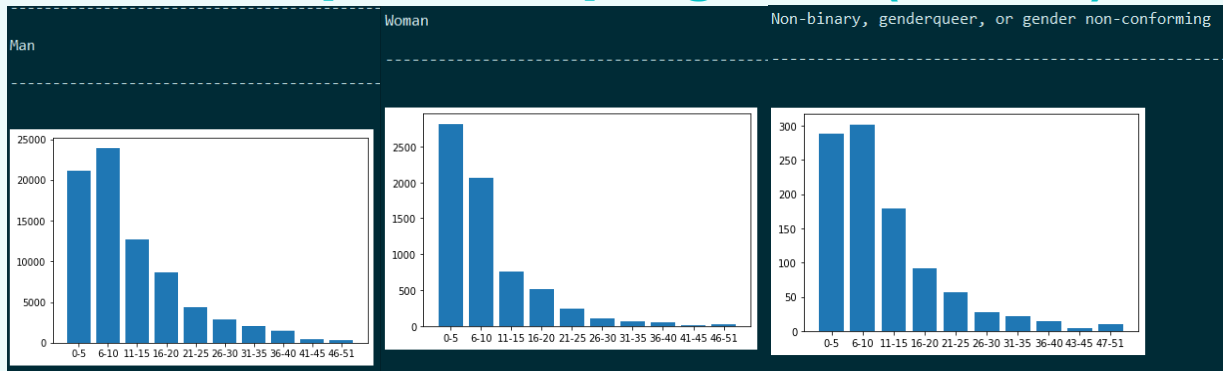
Con la tabla generada con nuestro script podemos observar que:

- Los tipos de desarrolladores más comunes son los del tipo “Developer, full-stack” y “Developer, back-end”
- El tipo de desarrollador menos común es “Marketing or sales professional”
- Hay una gran diferencia de proporcionalidad entre los dos primeros tipos y el resto de los tipos, el más cercano es “Developer, front-end”

No se sabe con certeza pero es posible que:

- El tipo menos común es Marketing y ventas debido a que la programación no está dentro de su campo aunque les puede ser útil en el análisis de datos importantes.
- La industria demanda programadores versátiles como los full-stack.

## 7. Años de experiencia por género (10 bins)



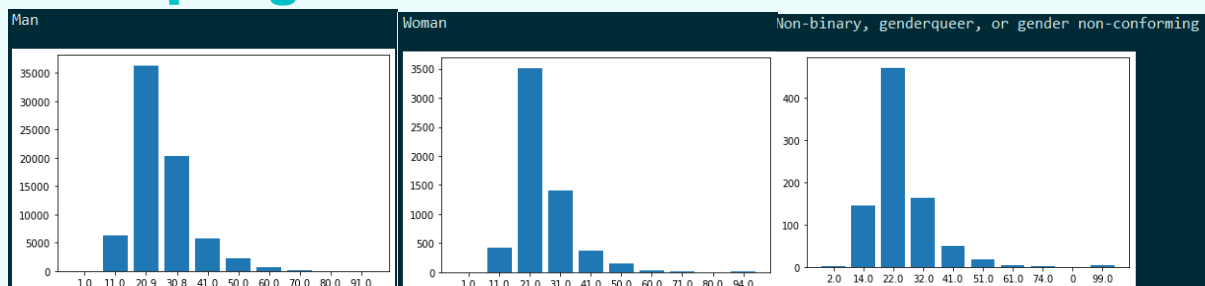
La tendencia indica que es más común que las mujeres tengan de 0 a 5 años de experiencia y los hombres tengan 6-10 años de experiencia.

## 8. Horas de trabajo semanales por tipo de desarrollador

En la mayoría de los tipos de desarrollador había una tendencia a la baja a encontrar personas con más horas de trabajo semanales que el resto. A excepción de estos tipos:

- Marketing or sales professional
- Scientist
- Developer, QA or test
- Senior executive/VP
- Engineer, site reliability

## 9. Edad por género



Podemos observar que entre todos los participantes de la encuesta el rango de edad más común fue de 20 a 30 años. La mayoría de los programadores son jóvenes adultos.

## 10. Correlación entre años de experiencia y salario anual.

Correlación = -0.0013187692522961042

La correlación entre estas dos variables es muy baja. Aunque por lógica se intuye que si debiera haber una correlación. Es probable que sea debido a la diversidad de los datos porque habíamos visto que los salarios anuales cambiaban mucho según la región y etnia. Un análisis por región y etnia sería más apropiado

## 11. Correlación entre edad y salario anual

Correlación = 0.10826846489974247

Al igual que en el ejemplo anterior nos encontramos con una correlación baja pero con un valor más alto que el anterior. Se volvió a verificar la función de correlación que se usaba y si es correcta. Por lo visto hay una muy ligera tendencia a que el salario de una persona aumente con su edad. Esto es probablemente debido a que una persona mayor tenga puestos que requieren más experiencia o por la antigüedad que genera y se ve reflejada en su salario

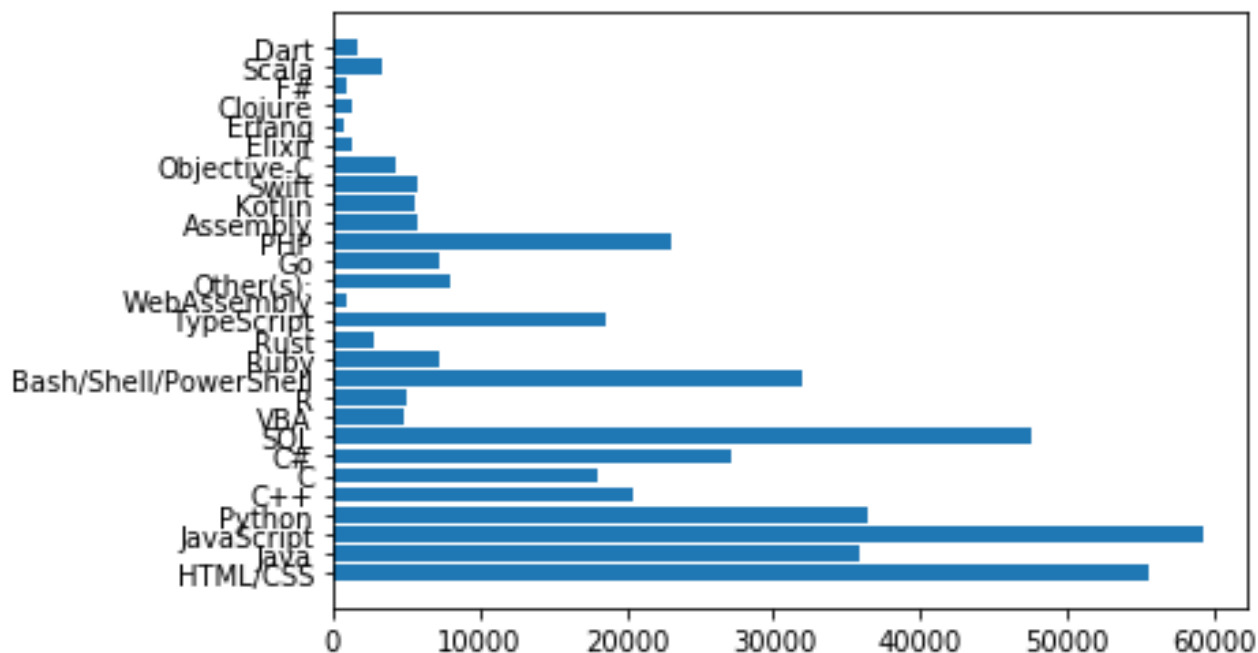
## 12. Correlación entre nivel de educación y salario anual

Correlación = 0.0031232203557014105

Esta correlación también es muy baja esto es probablemente a que ha habido una tendencia cada vez mayor de las empresas a contratar a personal sin estudios universitarios debido a que el conocimiento está disponible para todos en internet y es posible aprender a programar sin tener una educación formal.



## 13. Frecuencia de diversos lenguajes de programación



El lenguaje de marcado HTML y de estilo CSS son los más populares después del lenguaje javascript. El tercero más popular es SQL. Esto indica que hay una gran oferta de desarrolladores web. En el cuarto lugar se encuentran los lenguajes para hacer uso de la terminal en diversos sistemas operativos. Esto sumado al manejo de las bases de datos son los trabajos correspondientes a programadores back-end. Por lo cuál se pueden ver claramente estos dos tipos de programadores representados aquí. Esto concuerda con los análisis anteriores en donde se observó que los programadores full-stack eran los más numerosos.

## Conclusión

A través de este ejercicio de análisis puedo concluir que la población más común de los programadores son adultos jóvenes entre 20 y 30 años que se dedican al desarrollo full-stack con lenguajes de programación web. Estos programadores reciben aproximadamente la misma paga sin distinción de género, pero hay una brecha salarial ocasionada por la etnia a la que pertenecen en donde ser blanco o descendiente europeo te da una mejor paga que el resto. En su mayoría tienen pocos años de experiencia por lo cuál podemos ver que una gran parte se está iniciando en el mundo de la programación.