



Campus Irapuato-Salamanca | División de Ingenierías



UNIVERSIDAD DE  
GUANAJUATO

UNIVERSIDAD DE GUANAJUATO  
DIVISIÓN DE INGENIERÍAS, CAMPUS IRAPUATO-SALAMANCA

*Minería de Datos.*  
*Dr. Juan Carlos Gómez Carranza.*

# Práctica Intermedia.

Jonathan Joel Corona Ortega.  
NUA: 780840

Ingeniería en Sistemas Computacionales.

Fecha de Entrega:  
08 de mayo del 2020.

## DESCRIPCIÓN.

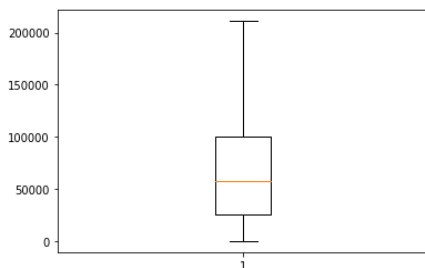
Hay un archivo CSV llamado `survey_results_public.csv` (comprimido como `survey_results_public.rar`) que contiene los datos de una encuesta recopilada de los usuarios de Stack OverFlow en 2019 con respecto a las siguientes diez variables (el nombre de los campos en el archivo están entre paréntesis): país (Country), nivel educativo (EdLevel), tipo de desarrollador (DevType), años de experiencia con codificación (YearsCode), salario anual en dólares estadounidenses (ConvertedComp), número promedio de horas de trabajo por semana (WorkWeekHrs), lenguaje de programación he / ella tiene experiencia con (LanguageWorkedWith), edad (Age), género (Gender) y etnia (Ethnicity). Hay datos para 88.883 usuarios.

Para algunas variables, los usuarios podrían responder con más de una respuesta, con las respuestas separadas con un ; en el archivo. Por ejemplo, en el lenguaje de programación con el que tiene experiencia, un usuario puede seleccionar al mismo tiempo C; C++; JavaScript; Python. En ese caso, para las estadísticas, el mismo usuario contará por cada idioma que elija. Lo mismo se aplica para cualquier otra variable que permita múltiples respuestas.

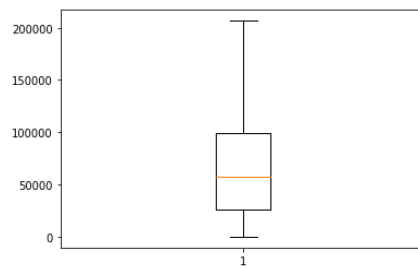
En otros casos, los usuarios pueden omitir una o varias respuestas, y en el archivo, podemos encontrar valores NaN o valores de cadena vacíos. En ese caso, para las estadísticas, esos valores deben ignorarse.

La práctica consiste en las siguientes pequeñas tareas de procesamiento y análisis de los datos contenidos en el archivo. Para cada tarea, debe escribir una función de Python como parte del archivo de código `practice.py`.

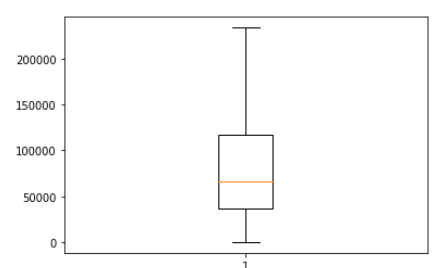
1. Calcule el resumen de cinco números, boxplot, la media y la desviación estándar para el salario anual por género.



Salario – Género.  
Man



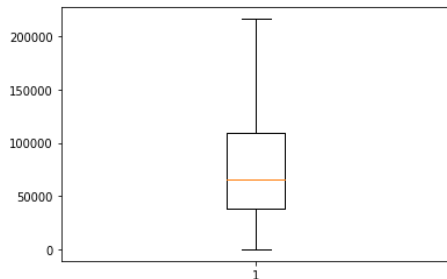
Salario – Género.  
Woman



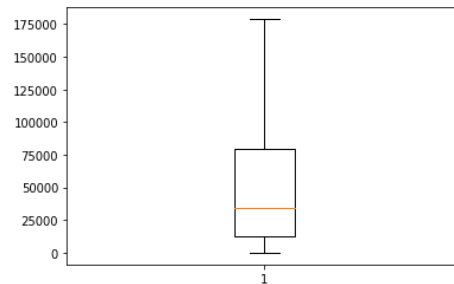
Salario – Género.  
Non-binary,  
genderqueer, or  
gender non-  
conforming

Como se puede observar en las gráficas de caja, los salarios anuales de cada género suelen estar en el mismo rango. Con unas pequeñas diferencias. Como se puede ver, las personas de género no binario suelen tener un mejor salario, que los hombres o las mujeres.

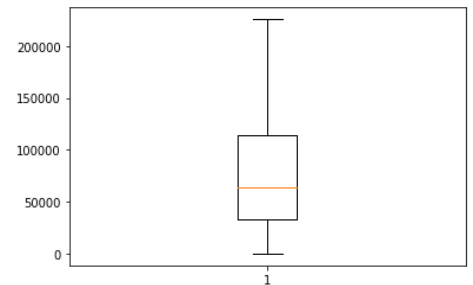
2. Calcule el resumen de cinco números, boxplot, la media y la desviación estándar del salario anual por grupo étnico.



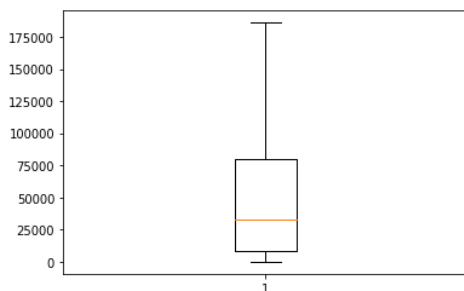
Salario – Grupo étnico.  
White or of European  
descent



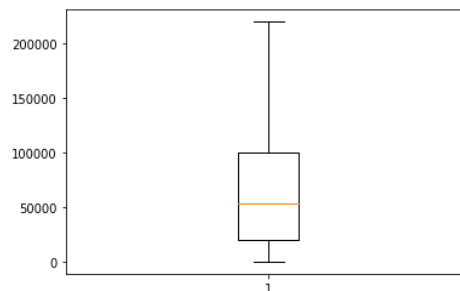
Salario – Grupo étnico.  
Middle Eastern



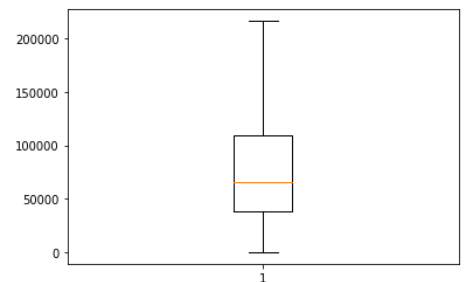
Salario – Grupo étnico.  
Biracial



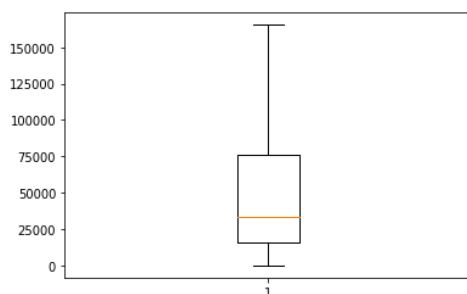
Salario – Grupo étnico.  
Black or of African  
descent



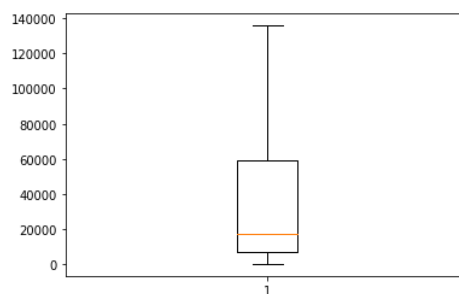
Salario – Grupo étnico.  
Multiracial



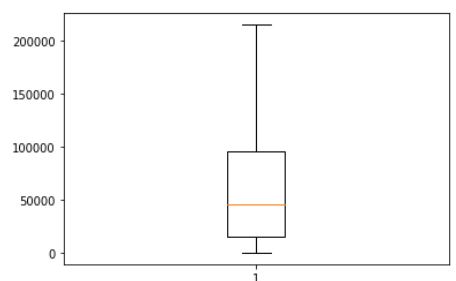
Salario – Grupo étnico.  
White or of European  
descent



Salario – Grupo étnico.  
Hispanic or Latino/Latina



Salario – Grupo étnico.  
South Asian

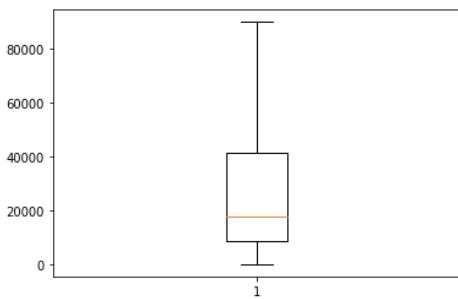


Salario – Grupo étnico.  
East Asian

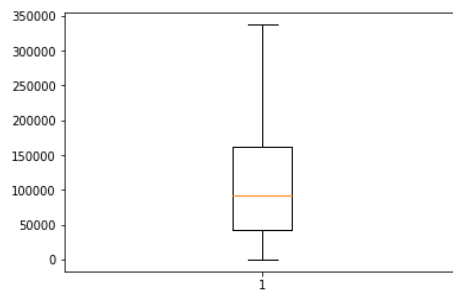
Como se puede observar en las gráficas, las personas blancas son las que suelen ganar más, que aquellas que no lo son. Por lo que puedo deducir que las personas blancas suelen tener un mejor puesto en las empresas de desarrollo y por ende mejores salarios.

Y las personas cuya descendencia proviene de lugares en vías de desarrollo suelen ganar un poco menos. En Asia oriental, por ejemplo, China, los trabajadores suelen ganar más que las del sur de Asia, como India; India y China suelen conocerse por su gran mano de obra barata, pero en la India se muestra más la sobre explotación.

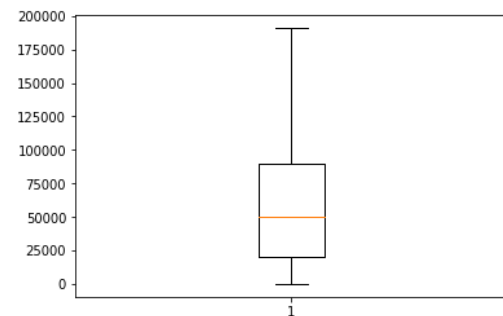
3. Calcule el resumen de cinco números, boxplot, la media y la desviación estándar para el salario anual por tipo de desarrollador.



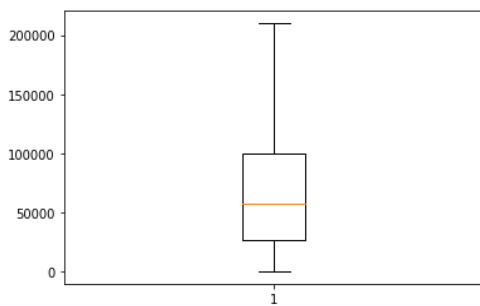
Salario – Nivel desarrollador.  
Student



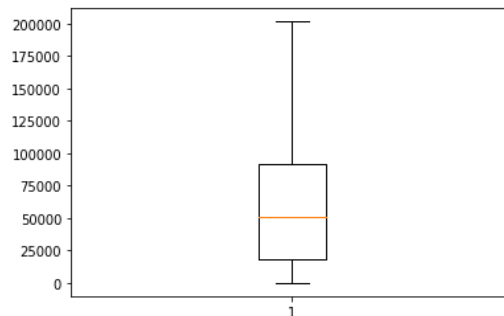
Salario – Nivel desarrollador.  
Senior executive/VP



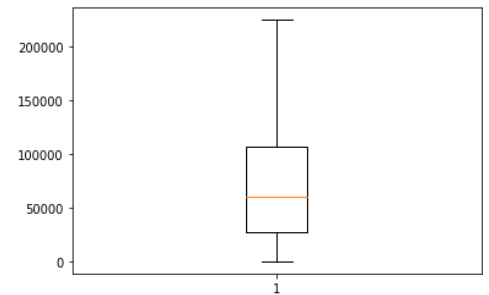
Salario – Nivel desarrollador.  
Educator



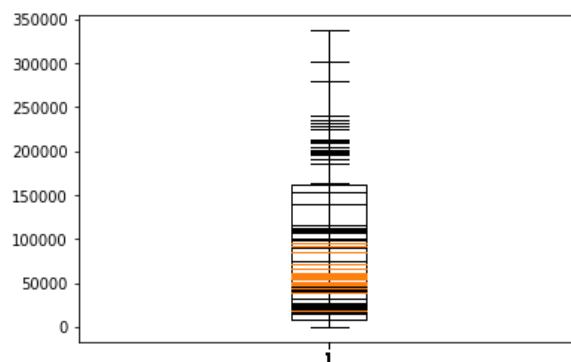
Salario – Nivel desarrollador.  
Developer, full-stack



Salario – Nivel desarrollador.  
Designer



Salario – Nivel desarrollador.  
Product Manager



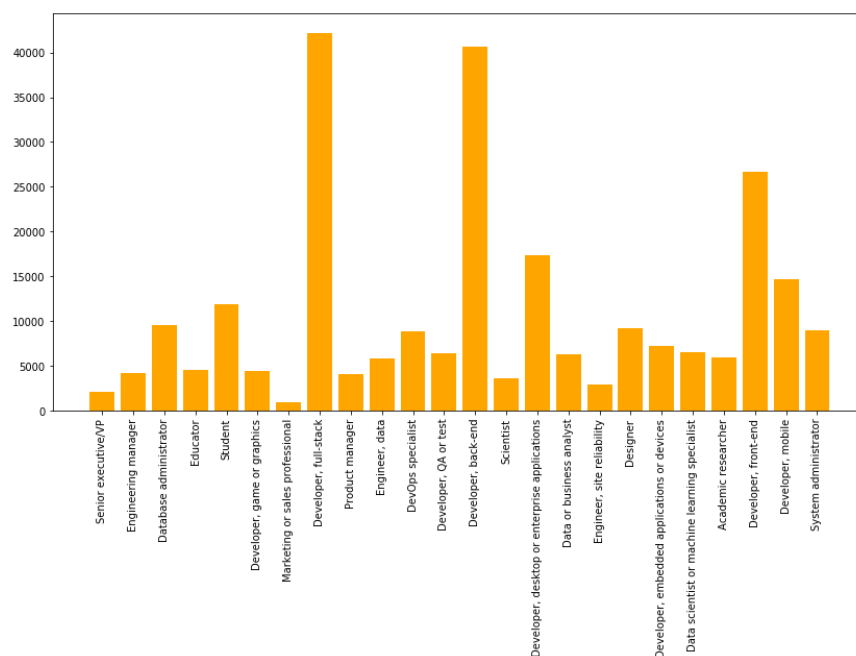
Como se puede observar en las gráficas hay mucha diferencia en los salarios anuales de cada tipo de desarrollador. En el área de educación, los desarrolladores suelen ganar menos, los estudiantes ganas menos debido que aun no cuentan con algún título, no tengan mucha experiencia o sean becarios.

Los desarrolladores que se especializan solo en un área o en muy pocas, suelen ganar menos en cambio a los que pueden trabajar en muchas áreas. También tener buenas habilidades te puede dar un mejor puesto y salario. En la última gráfica se puede notar la diferencia de salarios.

#### 4. Calcule la desviación media, media y estándar del salario anual por país.

De acuerdo con los resultados que se obtuvieron, los países en vías de desarrollo o de tercer mundo, como algunos países de África, de Oriente Medio o de Latinoamérica, son los que suelen remunerar muy poco a sus trabajadores en el área de desarrollo de software. Los países de primer mundo son los que mejor remuneran a sus trabajadores al año. Hay países que no se tomaron en cuenta para el análisis debido a que no se encontró la información suficiente para la prueba.

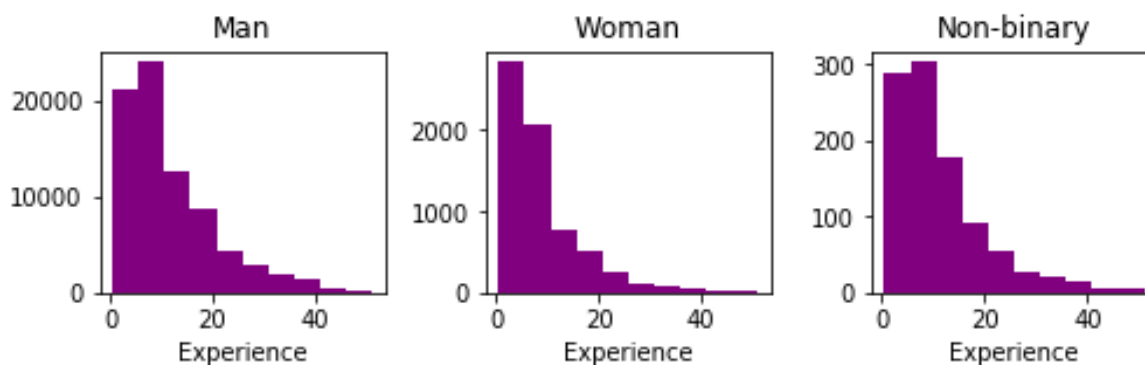
#### 5. Obtenga un diagrama de barras con las frecuencias de respuestas para cada tipo de desarrollador.



De acuerdo con la gráfica, los que hicieron la encuesta en su mayoría son desarrolladores full-stack o en seguida por casi nada los desarrolladores backend. Aunque podría decirse los full-stack están en los de backend y los de frontend, ya que en teoría pueden dominar ambas facetas. Los de menor frecuencia son aquellos

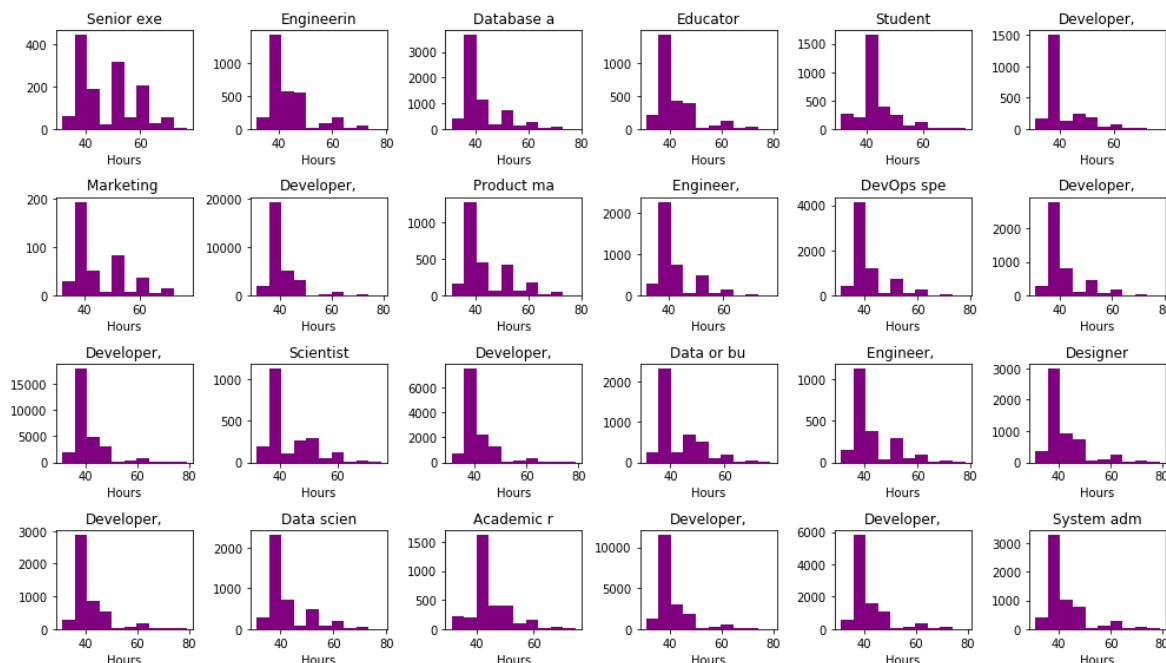
que a pesar de estar en la industria, no son desarrolladores de software como los diseñadores o producto manager, por lo que tal vez por eso son pocos en comparación a los demás.

- Trace histogramas con 10 bins para los años de experiencia con la codificación por género.



Como se observa en los histogramas, los hombres suelen tener más años de experiencia que las mujeres o las personas no binarias. A lo que mi parecer se debe debido a que en la actualidad las mujeres y la gente de otros géneros se están interesando en el área de la tecnología sin miedo a ser criticados debido a los estereotipos que suele haber. Con el pasar de los años se podrá notar una similitud en las tablas.

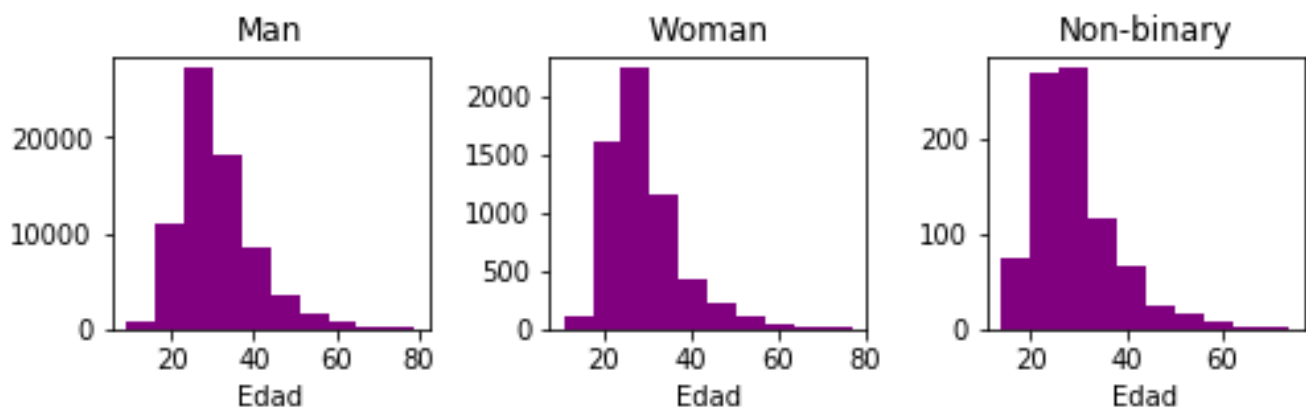
- Trace histogramas con 10 bins para el número promedio de horas de trabajo por semana, por tipo de desarrollador.



Como se observa en los histogramas, los senior executive a pesar de ser menos personas con esos cargos, son los que más horas suelen trabajar a la semana. Posiblemente debido a que se tienen que encargar de más áreas a parte del área de desarrollo de software.

Se puede notar en los demás casos que casi todos suelen tener las mismas horas de trabajo, algunos más, algunos pocos menos, también puede deberse a las horas extras que pueden hacer.

8. Trace histogramas con 10 bins para la edad por género.



De acuerdo con el histograma, la mayoría de las personas que realizaron la encuesta, se encuentran entre un rango de 20-40 años, donde podría decirse que la mayoría tiene más de 25 y menos de 35 años aproximadamente. En los tres histogramas se puede notar mucha similitud entre los géneros.

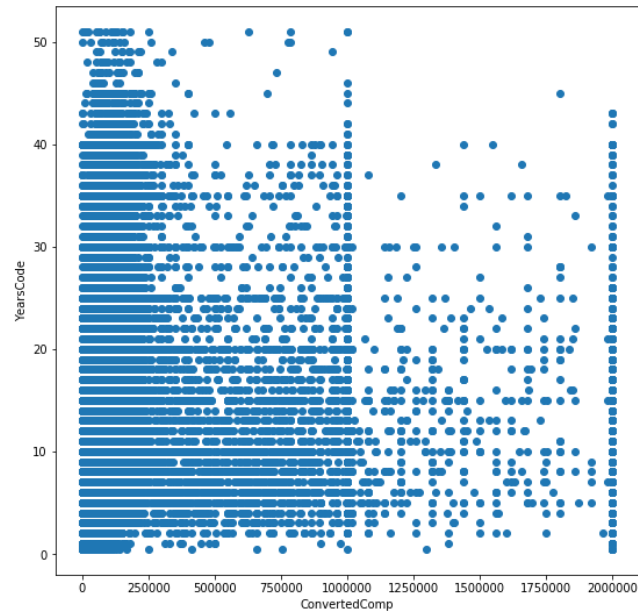
9. Calcule la mediana, la media y la desviación estándar de la edad por lenguaje de programación.

De acuerdo con los datos que se obtuvieron, las personas que trabajan con uno o varios lenguajes de programación están entre un rango de 25 y 30 años en promedio. Teniendo en cuenta que la gente más joven, dentro del rango suele trabajar con lenguajes un poco más nuevos y los mayores, suelen trabajar con lenguajes un poco más anticuados o por así decirlo, que salieron hace varios años.

10. Calcule la correlación entre años de experiencia y salario anual.

Correlation: 0.10600459758020611

Podemos observar una correlación positiva.

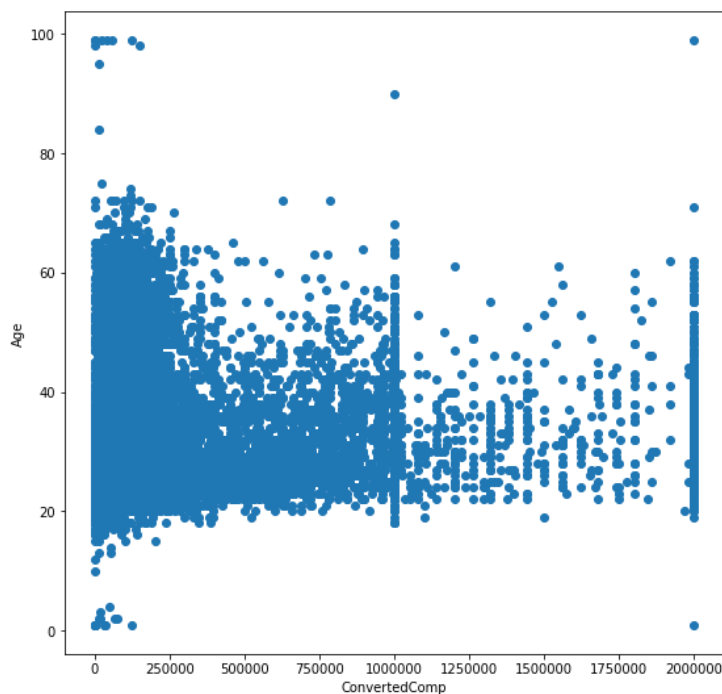


Como se observa, a pesar de los años de experiencia, muchas veces no te asegura un buen salario, puede depender mucho también a que área se está enfocando y cuantas personas hay con la misma experiencia que tú en ciertas áreas.

11. Calcule la correlación entre la edad y el salario anual.

Correlation: 0.10826846489974247

Podemos observar una correlación positiva.

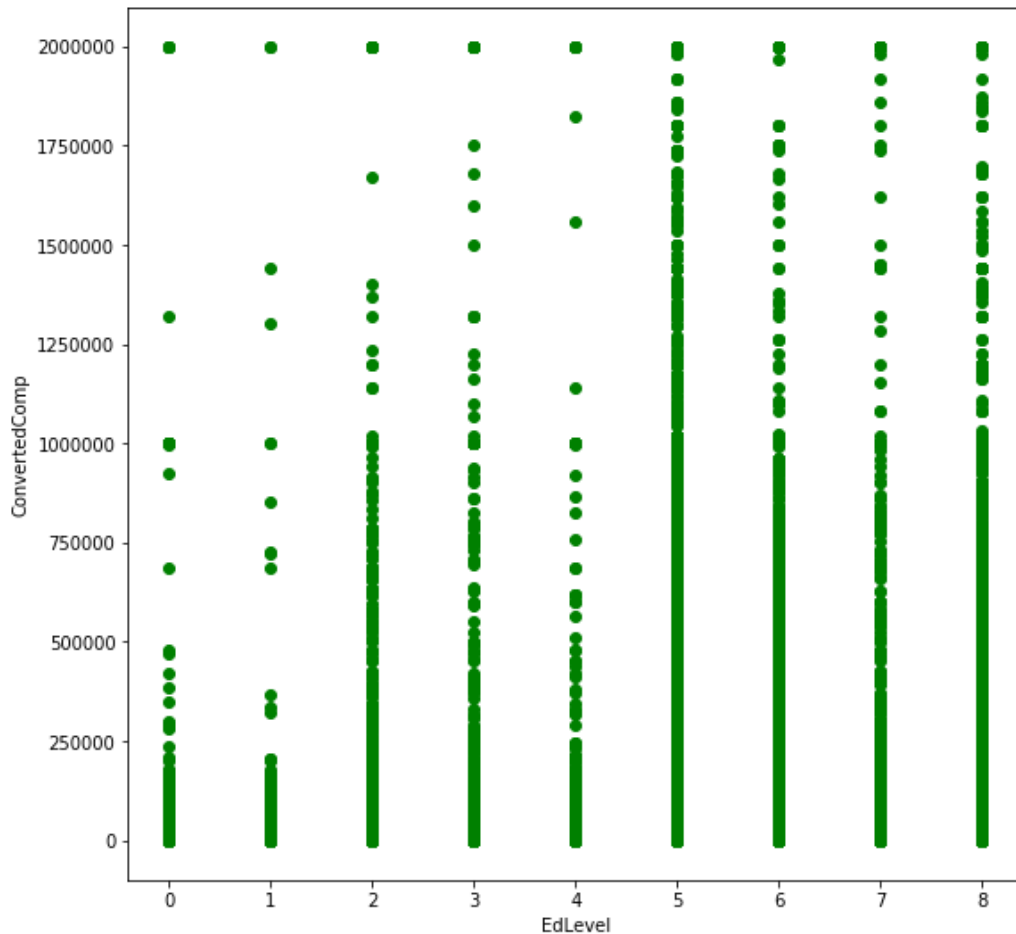




Como se puede observar ser joven no implicada que se te pagará poco, hay muchos factores que pueden atribuir al salario de cada persona.

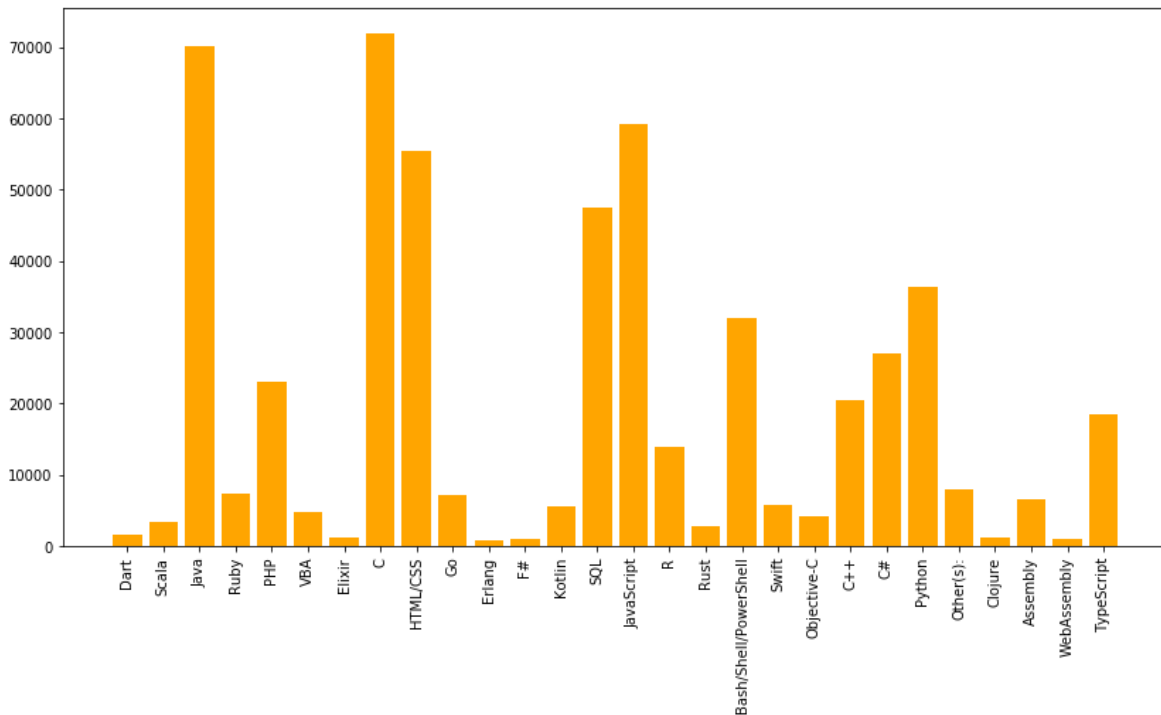
12. Calcule la correlación entre el nivel educativo y el salario anual. En este caso, reemplace la cadena del nivel educativo por un índice ordinal (por ejemplo, Primaria = 1, Secundaria = 2, etc.).

Correlation: 0.004743429243741032



Como se observan, en la mayoría de los casos, un mejor nivel educativo te beneficia al tener un mejor salario, pero no necesariamente tiene que ser así. Hay pocos casos, pero personas con bajo nivel educativo puede tener un salario similar o más alto que alguien con un mayor educación.

13. Obtain a bar plot with the frequencies of the different programming languages.



De acuerdo con las frecuencias, los lenguajes más usados con Java y C, los que suelen más usados para el desarrollo de software, detrás se encuentran los que se enfocan en el desarrollo web (una de las áreas más populares en la actualidad. Al final se encuentran los lenguajes más difíciles (ensamblador) o más nuevos (Dart).

## CONCLUSIÓN.

De acuerdo con las observaciones que se hicieron se pueden tomar los siguientes puntos:

- Los lenguajes de programación más antiguos siguen usando y los nuevos van mejorando y se usan cada vez más.
- El género no implica que salario puedas tener, pero tu etnia o el país en el que te encuentras puede afectar en la remuneración que puedes tener trabajando.
- Años de experiencia no es lo mismo a las habilidades que tienes; puedes tener muchos años trabajando en un área y ganar poco, puedes tener pocos años de experiencia, pero puedes tener muchas habilidades y herramientas.
- Las horas de trabajo pueden depender mucho de la persona y no tanto de la empresa en la que se trabaja. Es decir, puedes trabajar 8 horas diarias y terminar tu trabajo, o trabajar más debido que aún no lo terminas