

Docente: **Juan Carlos Gómez Carranza**Materia: **Minería de datos**Alumno: **José Alfredo Romero Gonzalez**

Practica Intermedia

Introducción

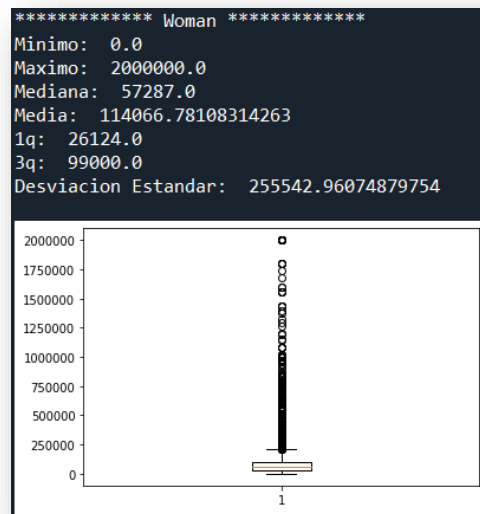
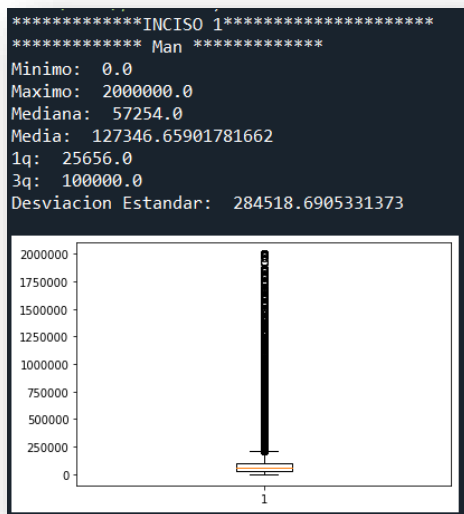
La minería de datos es el proceso de hallar anomalías, patrones y correlaciones en grandes conjuntos de datos para predecir resultados. Empleando una amplia variedad de técnicas, puede utilizar esta información para incrementar sus ingresos, recortar costos, mejorar sus relaciones con clientes, reducir riesgos y más.

El objetivo de esta práctica es analizar un conjunto de datos extraídos de una encuesta realizada a los usuarios de Stack OverFlow en 2019 con respecto a las siguientes diez variables: country (Country), educational level (EdLevel), developer type (DevType), years of experience with coding (YearsCode), annual salary in US dollars (ConvertedComp), average number of working hours per week (WorkWeekHrs), programming language he/she has experience with (LanguageWorkedWith), age (Age), gender (Gender) and ethnicity (Ethnicity). Teniendo un total de 88.883 usuarios encuestados.

Desarrollo

En esta parte hablare sobre algunos incisos donde pude ver resultados interesantes respecto al análisis que se le hizo.

Inciso 1

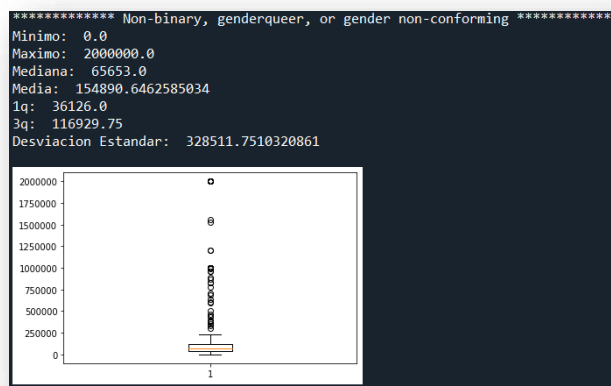




Docente: **Juan Carlos Gómez Carranza**

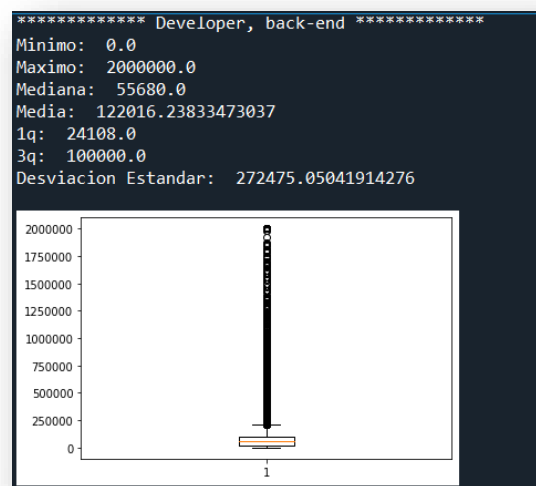
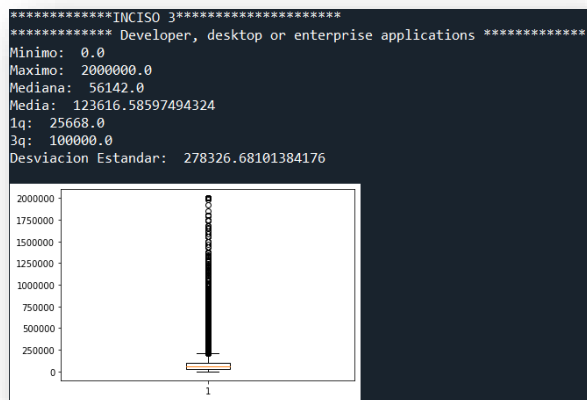
Materia: **Minería de datos**

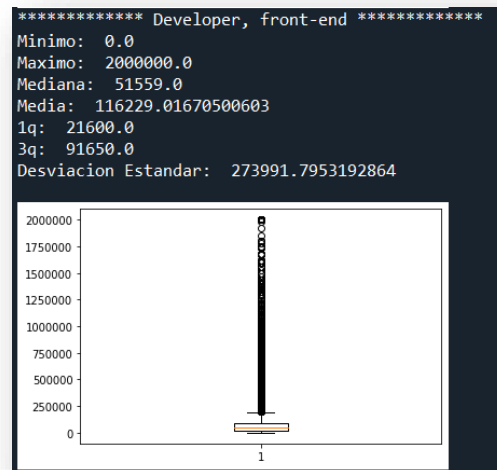
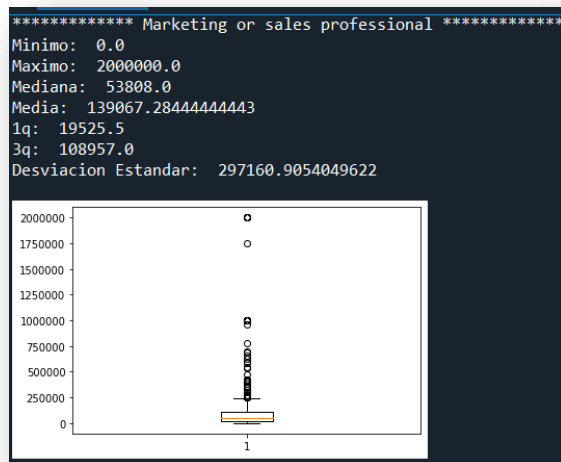
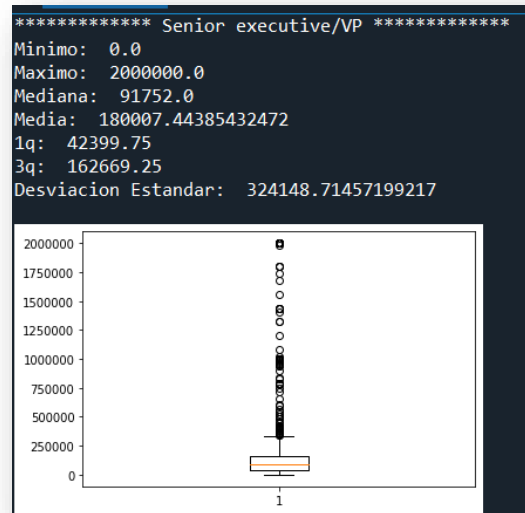
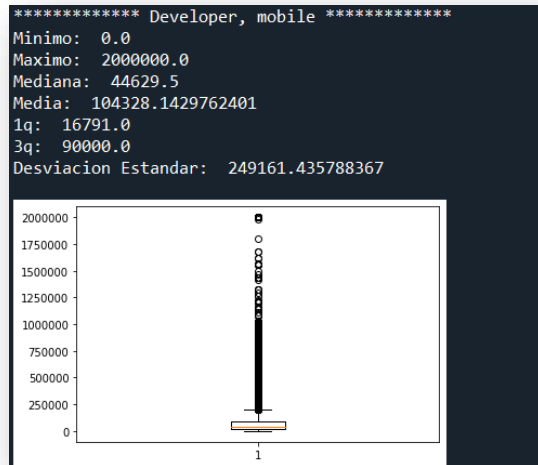
Alumno: **José Alfredo Romero Gonzalez**



Algo interesante en esta parte y en la mayoría de los datos donde desplegamos un boxplot, es que existen muchos outliers, estos los podemos trabajar utilizando la media cortada, en este caso no se hizo de esa manera, pero seria una opción, como vemos en la desviación estándar, los datos se alejan considerablemente respecto a la media, esto nos habla de que existe una gran diferente de salarios entre los empleados, otra cosa, es que en la mayoría de los usuarios encuestados son hombres.

Inciso 3





En esta parte, se puede observar, con respecto a la media, la diferencia de salarios que existen dependiendo del tipo de desarrollador, este puede ser un dato muy importante para aquellas personas que les interese saber cuanto gana un programador junior y saber si estudiar alguna carrera sobre programación o si ya esta empezando, saber que tipo de desarrollador le gustaría ser si piensa en lo económico, que obviamente si trabajamos es por el dinero.

Por ejemplo, podemos ver el boxplot de un desarrollador móvil, este tipo de desarrollador a veces es muy cotizado por el trabajo que se realiza en ella, y podemos ver que les va demasiado bien y puede ser una opción para alguien que aun no se decide por que tipo de desarrollador ser.

Docente: **Juan Carlos Gómez Carranza**Materia: **Minería de datos**Alumno: **José Alfredo Romero Gonzalez**

También podemos ver a los desarrolladores Senior executive, son los que ganan mas con referencia en la media, son el valor mas alto que dio el análisis.

Inciso 4

```
***** Guyana *****
Dont exists registers
***** Honduras *****
Mediana: 18022.0
Media: 21239.866666666665
Desviacion Estandar: 18929.222846757923
***** Qatar *****
Mediana: 79326.0
Media: 224692.666666666666
Desviacion Estandar: 347270.5181798318
***** Jamaica *****
Mediana: 13476.0
Media: 24536.090909090908
Desviacion Estandar: 19984.39803471491
***** Fiji *****
Mediana: 30529.5
Media: 30529.5
Desviacion Estandar: 9685.5
***** Haiti *****
Mediana: 8910.0
Media: 8910.0
Desviacion Estandar: 6606.0
```

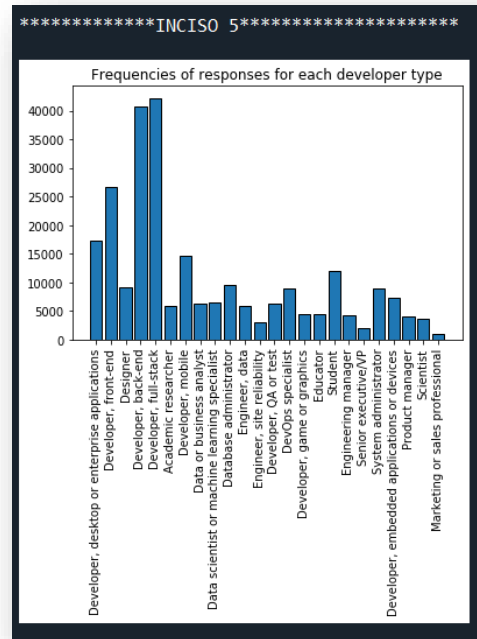
```
***** Mexico *****
Mediana: 16968.0
Media: 30372.46220302376
Desviacion Estandar: 100074.95296587024
***** Dominican Republic *****
Mediana: 16667.0
Media: 27355.147058823528
Desviacion Estandar: 37099.503982043585
***** Guatemala *****
Mediana: 20400.0
Media: 23677.627906976744
Desviacion Estandar: 14086.754118052782
***** Chile *****
Mediana: 27528.0
Media: 40438.90972222222
Desviacion Estandar: 55030.375370097136
***** Ecuador *****
Mediana: 21800.0
Media: 26596.10714285714
Desviacion Estandar: 22194.318077316362
***** Ghana *****
Mediana: 7212.0
Media: 16520.18181818182
Desviacion Estandar: 28514.85017198194
```

En el inciso 4, al hacer el análisis, pudimos ver que existen países como México que el salario esta considerablemente aceptable para un desarrollador con referencia en la media, pero, aun así, viendo la desviación estándar, algunos de sus salarios pueden elevarse considerablemente, pero existen países como Haití que tienen muy pocos desarrolladores y no son muy bien pagados.

Esto puede importante para personas que buscan un mejor salario y poder emigrar a algún otro país donde su trabajo sea mejor valorado y tengan más oportunidad.

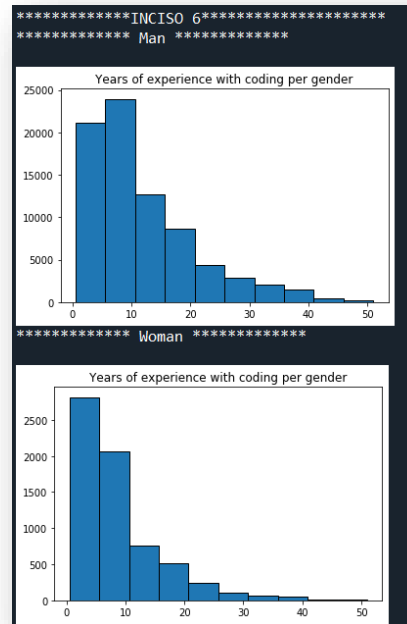
Docente: **Juan Carlos Gómez Carranza**Materia: **Minería de datos**Alumno: **José Alfredo Romero Gonzalez**

Inciso 5



Esta parte también del análisis también es muy interesante, ya que podemos ver claramente que la mayoría de los usuarios son desarrolladores full stack con más de 40,000 usuarios, lo cual significa que pueden realizar cualquier tipo de sistema que se les solicite, aunque estos no son los mejores pagados pero si tienen un buen salario. En contraparte, los desarrolladores de marketing o ventas son muy pocos, rondando por debajo de 5,000 usuarios encuestados.

También destaca que los desarrolladores Senior soy muy pocos, lo cual nos podría hablar de que para lograr ser un desarrollador Senior no es una tarea fácil.

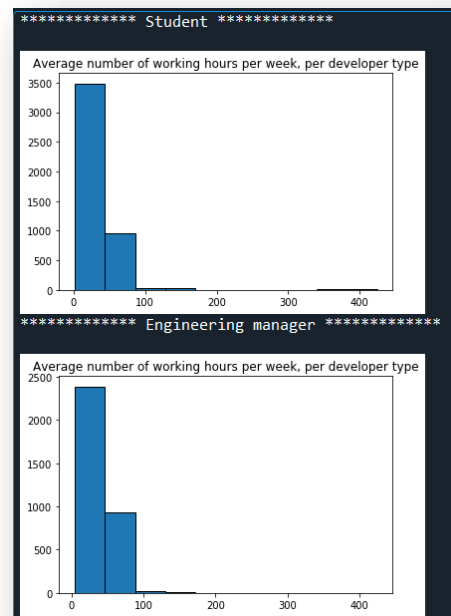
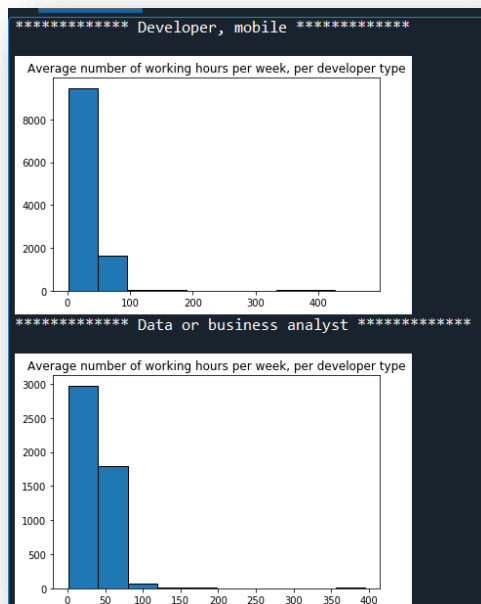
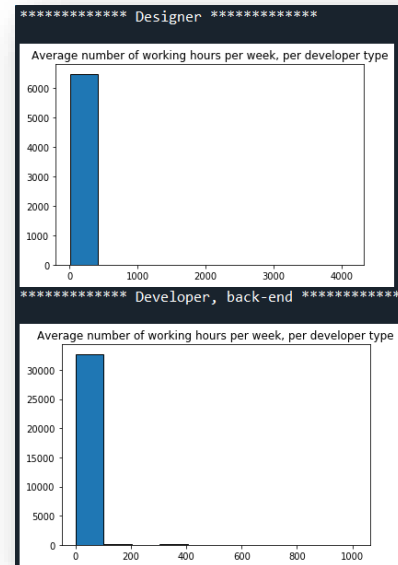
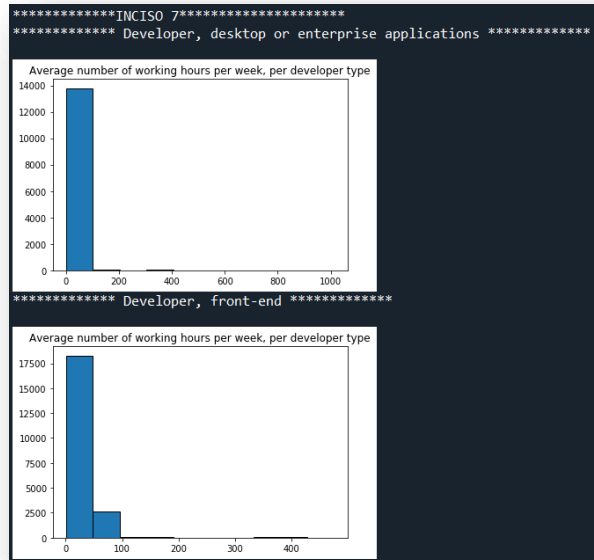
Docente: **Juan Carlos Gómez Carranza**Materia: **Minería de datos**Alumno: **José Alfredo Romero Gonzalez****Inciso 6**

Ya en el inciso 6, mostramos graficamente los años de experiencia programando por genero, como podemos ver, en los hombres tiene mas experiencia programando que ronda de los 5 a los 10 años, y las mujeres tienen experiencia programando que ronda de los 0 a los 5 años, lo cual nos podria decir que los hombres duran mas tiempo programando que las mujeres. De hecho, en la grafica de las mujeres se ve como de los 10 a los 15 años baja considerablemente el numero de usuarios y ya casi en los 40 años no hay mujeres desarrollando.

En la parte de los hombres, da un salto considerable entre los 0 y 5 a los 5 y 10, ya para los 10 y 15 baja pero son tan significativo como en el caso de las mujeres.



Inciso 7

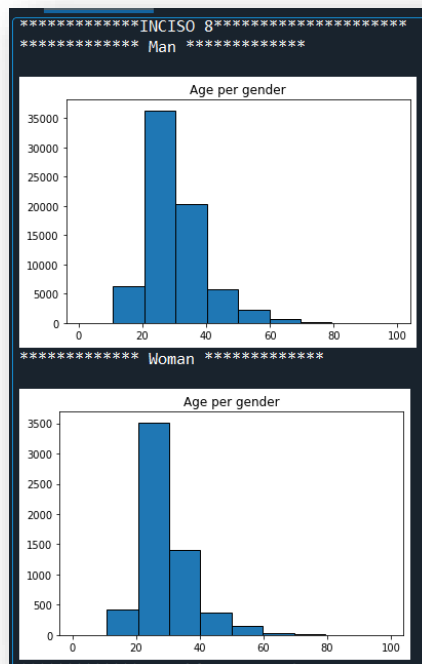


Otro analisis en el conjunto de datos interesante es el saber cuantas horas trabaja un desarrollador dependiendo de su puesto, por ejemplo, en las graficas mostradas podemos ver que los diseñadores llegan a trabajar bastantes horas, la grafica nos muestra que mas de 6000 diseñadores trabajan alrededor de 500 horas por promedio en la semana, lo cual es bastante.

Docente: **Juan Carlos Gómez Carranza**Materia: **Minería de datos**Alumno: **José Alfredo Romero Gonzalez**

Por otro lado, los desarrolladores frontend, con mas de 17500 llegan a trabajar en promedio 100 horas a la semana, obviamente hay quienes trabajan por abajo del promedio y quienes terminan haciendo hasta horas extras, esto dependiendo del tipo de proyecto que esten realizado y lo atrasado que puedan llevarlo.

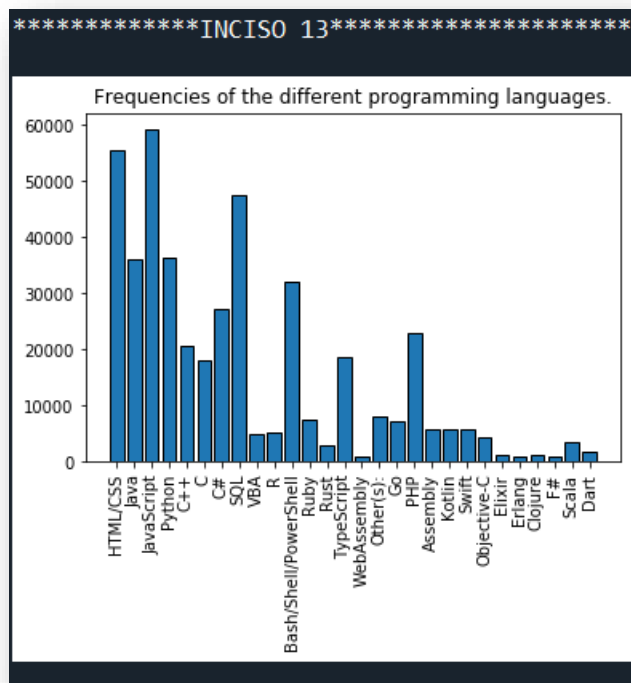
Inciso 8



En el inciso 8, mostramos graficas de las edades por genero de los usuarios, donde podemos observar que la mayoría de los desarrolladores tienen entre 20 y 30 años, tanto en hombres como en mujeres. Las graficas son muy similares, pero como lo dijimos anteriormente, hay mas hombres que mujeres desarrollando, hay una amplia diferencia, en la grafica de los hombres son poco mas de 35000 y en la grafica de mujeres poco mas de 3500, no mostramos el tercer género, pero son un poco menos que las mujeres.

Esto puede ser un detalle que las mujeres pueden tomar en cuenta, ya que casi no hay desarrolladores, un factor importante es que los hombres desde pequeños se involucran en este tipo de tecnología por videojuegos y cosas de esa índole, y las mujeres pueden que se involucren cuando ya están en una edad mas adolescente.

Seria interesante hacer una encuesta sobre eso, porque a los hombres les llama más la atención que a las mujeres.

Docente: **Juan Carlos Gómez Carranza**Materia: **Minería de datos**Alumno: **José Alfredo Romero Gonzalez****Inciso 13**

Esta probablemente sea la parte mas interesante de este trabajo, ya que se muestra la cantidad de usuarios que programan en los diferentes lenguajes de programación, por ejemplo, vemos que el lenguaje que usan mas usuarios es el de JavaScript, un lenguaje que ha ido agarrando cada vez mas fuerza, tiene casi 60000 usuarios.

El lenguaje que menos usuarios usan, son 3 probablemente según la gráfica, los cuales son WebAssembly, Erlang y F#, con menos de 5000 usuarios.

El lenguaje Python también ha agarrado cada vez más fuerza a través de los años desde su creación, para el años del 2019 contaba, en base a esta encuesta, con casi 40000 usuario, no dudamos que actualmente existan más desarrolladores que vayan iniciando en este.