

Salarios 2020



Titular de la materia: Mc. Fernando Castillo

Alumno: José Luis Toribio

Reporte Salarios

Salarios 2020

A continuación en el siguiente análisis sugerido por el prof.Fernando Castillo, utilizaremos el set de datos "Salarios 2020", proporcionado por la empresa "Software guru".

En el cual nos daremos a la tarea de crear un análisis entre los profesionistas que se dedican o bien tienen conocimientos en el área de DevOps, contra aquellos que conocen o bien dominan la tecnología llamada Dockers Containers, ya que consideramos de suma importancia que los profesionista que se dedican al área de DevOps, tengan conocimiento de Contenedores.

Como paso siguiente veremos que porcentaje de la comunidad que practica DevOps implementa Dockers Containers, así mismo analizaremos la media salarial, entre dichos segmentos.

como primer paso cargaremos nuestro set de datos para comenzar con nuestro análisis.

```
In [3]: import pandas as pd
```

```
In [33]: #vamos a cargar nuestro set de datos.
df = pd.read_csv('answers.csv')
df
```

Out[33]:

		id	ip_address	created	recaptcha_score	salarymx	salaryusd	extramx	extrausd	variation	english_num	...	ben_gy
0	2		189.205.154.128	2019-12-23 22:19:42.131242	NaN	28000	0	28	0	5	1	...	Na
1	3		189.205.154.128	2019-12-23 22:19:53.567686	NaN	26000	0	40000	0	4	2	...	Na
2	4	2806:2f0:9002:462c:348e:dbb6:d957:f16b		2019-12-23 22:24:04.580942	NaN	35000	0	18000	0	30	3	...	Na
3	5		189.205.154.128	2019-12-23 22:24:22.759488	NaN	17000	0	219000	0	0	1	...	
4	6		187.190.30.86	2019-12-23 22:26:46.171217	NaN	30000	0	10000	0	1	2	...	Na
...
2547	2650		201.159.102.34	2020-02-05 22:53:17.709502	NaN	46000	0	20000	0	22	3	...	Na
2548	2651	2806:107e:7:49d7:44fc:2254:e307:536e		2020-02-07 20:45:30.489241	NaN	28000	0	0	0	10	2	...	Na
2549	2652		148.240.67.129	2020-02-10 23:59:35.442983	NaN	25000	0	25000	0	25	2	...	Na
2550	2653	2806:268:403:88a4:a439:ed0a:13bf:e85e		2020-02-11 02:38:24.61969	NaN	24000	0	10000	0	0	3	...	Na
2551	2654	2806:106e:24:1f32:60eb:a482:1d18:9096		2020-02-12 16:56:04.093759	NaN	30000	0	15000	0	5	2	...	Na

2552 rows x 159 columns

como siguiente paso, procederemos a extraer el set de datos, donde si el campo : "act_devops" contiene un valor positivo(en otras palabras la letra Y). Y después trabajemos con el nuevo data frame

```
In [5]: #extraer los row que contienen el 'Y'. Lo que indica que son profesionistas que usan DevOps...
dfDevops = df[df.act_devops=='Y']
print("Tabla de datos")
dfDevops
```

```
Out[5]:
```

	id	ip_address	created	recaptcha_score	salarymx	salaryusd	extramx	extrausd	variation	english_num	...	ben_gy
2	4	2806:2f0:9002:462c:348e:dbb6:d957:f16b	2019-12-23 22:24:04.580942	NaN	35000	0	18000	0	30	3	...	Na
10	12	189.216.205.103	2019-12-23 23:04:39.046583	NaN	70000	0	0	0	0	3	...	Na
16	18	177.248.225.111	2019-12-23 23:41:48.487477	NaN	10000	0	6000	0	0	3	...	Na
35	37	2806:103e:b:70b1:3dc2:89f9:44f4:9142	2019-12-24 01:00:03.188644	NaN	70000	0	120000	0	0	3	...	Na
37	40	201.175.132.97	2019-12-24 01:03:43.81779	NaN	26000	0	18000	0	30	2	...	Na
...
2496	2596	189.216.140.186	2020-01-19 16:23:06.369485	NaN	15000	0	7500	0	0	3	...	Na
2497	2597	189.212.127.201	2020-01-19 16:29:35.891567	NaN	28000	0	25000	0	20	3	...	Na
2499	2599	2806:2f0:9005:f8c6:20a2:54dc:5033:ff7	2020-01-20 00:55:24.081861	NaN	35000	0	10000	0	0	2	...	Na
2502	2602	162.218.228.196	2020-01-20 18:43:55.474372	NaN	50000	0	40000	0	4	4	...	Na
2509	2609	200.56.77.165	2020-01-21 15:57:22.401189	NaN	62000	0	1050000	0	3	5	...	Na

202 rows x 159 columns

Ahora haremos una extracción de aquellos profesionistas que implementan DevOps pero no Docker Containers

```
In [29]: #Extraer información acerca de los profesionistas que no usan Dockers

dfNoDocker = dfDevops[dfDevops.infra_docker!='Y']
print("Total de profesionistas que no usan Dockers:",dfNoDocker.shape[0])

dfNoDocker

Total de profesionistas que no usan Dockers: 78
```

```
In [30]: #Extraer información acerca de los profesionistas que usan Dockers
dfSiDocker = dfDevops[dfDevops.infra_docker=='Y']
print("Total de profesionistas que usan Dockers:",dfSiDocker.shape[0])

dfSiDocker
```

Total de profesionistas que usan Dockers: 124

Out[30]:

	id	ip_address	created	recaptcha_score	salarymx	salaryusd	extramx	extrausd	variation	english_num	...	ben_gy
2	4	2806:2f0:9002:462c:348e:dbb6:d957:f16b	2019-12-23 22:24:04.580942	NaN	35000	0	18000	0	30	3	...	Na
10	12	189.216.205.103	2019-12-23 23:04:39.046583	NaN	70000	0	0	0	0	3	...	Na
16	18	177.248.225.111	2019-12-23 23:41:48.487477	NaN	10000	0	6000	0	0	3	...	Na
35	37	2806:103e:b:70b1:3dc2:89f9:44f4:9142	2019-12-24 01:00:03.188644	NaN	70000	0	120000	0	0	3	...	Na
37	40	201.175.132.97	2019-12-24 01:03:43.81779	NaN	26000	0	18000	0	30	2	...	Na
...
2472	2572	179.6.201.92	2020-01-17 13:38:53.222368	NaN	0	2400	0	0	1	2	...	Na
2492	2592	174.194.203.203	2020-01-18 20:52:30.435736	NaN	113000	0	200000	0	50	5	...	Na
2493	2593	2806:109f:16:52f5:767f:d632:4357:2c62	2020-01-19 04:26:12.223584	NaN	30000	0	15000	0	0	4	...	Na
2496	2596	189.216.140.186	2020-01-19 16:23:06.369485	NaN	15000	0	7500	0	0	3	...	Na
2499	2599	2806:2f0:9005:f8c6:20a2:54dc:5033:ff7	2020-01-20 00:55:24.081861	NaN	35000	0	10000	0	0	2	...	Na

124 rows x 159 columns

```
In [8]: import math
porcentajeSiDocker = (int(dfSiDocker.shape[0])*100)/dfDevops.shape[0]

print("Porcentaje de Profesionistas que Si emplean DevOps y Docker ",math.trunc(porcentajeSiDocker,"%"))
```

Porcentaje de Profesionistas que Si emplean DevOps y Docker 61 %

```
In [9]: import math
porcentajeNoDocker = 100-(math.trunc(porcentajeSiDocker))
print("Porcentaje de Profesionistas que No emplean DevOps y Docker ",porcentajeNoDocker,"%")
```

Porcentaje de Profesionistas que No emplean DevOps y Docker 39 %

A continuación mostraremos en una grafica de Barras el numero de los dos casos, tanto usan y no usan Docker Containers

```
In [16]: import matplotlib.pyplot as plt

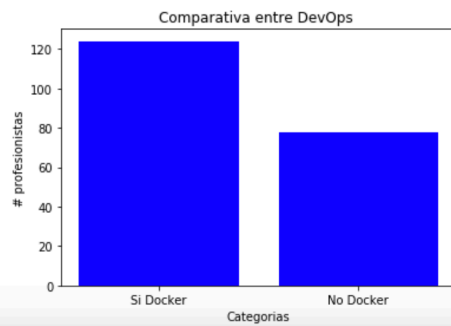
#definir las categorias en nuestro casa sera Si Docker y No Docker

categoryX= ['Si Docker', 'No Docker']
dataY = [dfSiDocker.shape[0],dfNoDocker.shape[0]]

plt.bar(categoryX, dataY, color="blue")

plt.xticks(categoryX)
plt.ylabel('# # profesionistas')
plt.xlabel('Categorias')

plt.title('Comparativa entre DevOps ')
plt.show()
```



Out[29]:

	id	ip_address	created	recaptcha_score	salarymx	salaryusd	extramx	extrausd	variation	english_num	...	ben_gyr
44	47	201.97.24.183	2019-12-24 01:39:06.795801	NaN	10000	0	0	0	1	2	...	Na
81	85	177.228.99.109	2019-12-24 17:09:04.111177	NaN	36000	0	0	0	25	3	...	Na
82	87	2806:2f0:9100:dbbc:e5e5:37ac:634f:dd54	2019-12-24 17:11:56.906353	NaN	34000	0	100000	0	5	2	...	Na
92	99	187.233.146.133	2019-12-24 18:37:58.05995	NaN	80000	0	960000	0	0	4	...	
102	109	201.175.156.63	2019-12-24 18:48:15.215682	NaN	40000	0	21000	0	1	4	...	Na
...
2363	2461	187.147.120.7	2020-01-14 21:22:42.701203	NaN	24000	0	0	0	10	1	...	Na
2404	2502	200.68.138.39	2020-01-15 05:45:29.106513	NaN	21000	0	30000	0	20	3	...	Na
2497	2597	189.212.127.201	2020-01-19 16:29:35.891567	NaN	28000	0	25000	0	20	3	...	Na
2502	2602	162.218.228.196	2020-01-20 18:43:55.474372	NaN	50000	0	40000	0	4	4	...	Na
2509	2609	200.56.77.165	2020-01-21 15:57:22.401189	NaN	62000	0	1050000	0	3	5	...	Na

78 rows x 159 columns

Ahora como podemos ver en la gráfica, nos muestra que es mayor el número de profesionistas en DevOps que tienen habilidades en Dockers, a diferencia de los que no. Lo cual ayuda a nuestro estudio a refirmar la importancia de Docker container como DevOps enginner.

Claro que ahora queda analizar la brecha salarial entre los dos grupos. Para el mejor desempeño de nuestro analisis convertiremos los salarios a la unidad de dólares, con la finalidad de estandarizar nuestros datos. Tomaremos el valor del dolar al día domingo 23 de 2020, 1 dolar = 19.04 MXN

```
In [28]: #ahora procederemos a convertir nuestro campo "salarymx a dolares", mecionar que usaremos dos campos,
#el anterior mencionado y ademas "salaryusd"
```

```
salarySiDocker = dfSiDocker[['act_devops','infra_docker','salarymx','salaryusd']]
#tabla con salarios de profesionistas que Si implementan Dockers

salarySiDockerUsd1 = (salarySiDocker['salarymx'].sum())/19.04
salarySiDockerUsd2 = salarySiDocker['salaryusd'].sum()

usdMeanSiDocker = (salarySiDockerUsd1 + salarySiDockerUsd2)/int(dfSiDocker.shape[0])

print("Promedio de salarios de profesionistas que usan Docker",usdMeanSiDocker)
```

Promedio de salarios de profesionistas que usan Docker 2967.0244815668207

```
In [15]: salaryNoDocker = dfNoDocker[['act_devops','infra_docker','salarymx','salaryusd']]
#tabla con salarios de profesionistas que No implementan Dockers

#tabla con salarios de profesionistas que Si implementan Dockers

salaryNoDockerUsd1 = (salaryNoDocker['salarymx'].sum())/19.04
salaryNoDockerUsd2 = salaryNoDocker['salaryusd'].sum()

usdMeanNoDocker = (salaryNoDockerUsd1 + salaryNoDockerUsd2)/int(dfNoDocker.shape[0])

print("Promedio de salarios de profesionistas que usan Docker",usdMeanNoDocker)
```

Promedio de salarios de profesionistas que usan Docker 2663.7807315233786

Ahora veremos en la siguiente Gráfica los promedios salariales de profesionistas en DevOps, que Si usan o No usan Docker.

```
In [27]: #definir las categorias en nuestro casa sera Si Docker y No Docker

data = {'Clase':['Si Docker','No Docker'],'Promedio Salarios (USD)':[usdMeanSiDocker,usdMeanNoDocker]}

dfSalaryMean = pd.DataFrame(data, columns=['Clase','Promedio Salarios (usd)'])

print("\n")
print(dfSalaryMean)
print("\n")

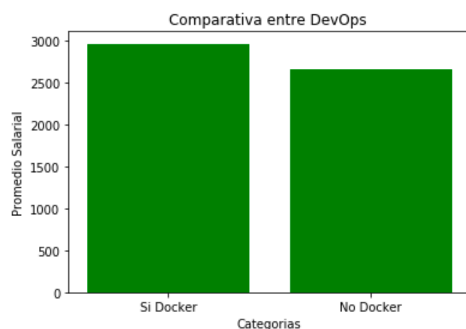
categoryX= ['Si Docker', 'No Docker']
dataY = [usdMeanSiDocker,usdMeanNoDocker]

plt.bar(categoryX, dataY, color="green")

plt.xticks(categoryX)
plt.ylabel('Promedio Salarial')
plt.xlabel('Categorias')

plt.title('Comparativa entre DevOps ')
plt.show()
```

	Clase	Promedio Salarios (usd)
0	Si Docker	NaN
1	No Docker	NaN



Como Pudimos observar en la anterior tabla y Gráfica, nuestra comparativa entre los promedios salariales de aquellos profesionistas que utilizan Dockers o no lo utilizan, es mínima, por lo que tenemos como resultado de nuestro estudio, que el saber o implementar Dockers siendo un DevOps enginner representa una diferencia mínima salarial.

Lo que nos da como resultado es que podríamos saber o no saber Dockers y aún así aspirar a un salario que oscile entre los \$2967.02(USD) o \$2663.78 (USD). Sin embargo en contraste pudimos notar que un 61% de los profesionistas involucrados en DevOps si implementan Dockers, lo cual nos indica que a pesar de tener un impacto mínimo en el salario, los profesionistas suelen aprender Contenedores.

